

# **Frontiers in Biomechanics**

# Frontiers in Biomechanics

Edited by

G.W. Schmid-Schönbein

S.L-Y. Woo

B.W. Zweifach

With 136 Figures



Springer-Verlag  
New York Berlin Heidelberg Tokyo

**G.W. SCHMID-SCHÖNBEIN** Department of Applied Mechanics and  
Engineering Sciences—Bioengineering, University of California, San Diego,  
La Jolla, California 92093, U.S.A.

**S.L-Y. Woo** Department of Surgery, Division of Orthopedics and  
Rehabilitation, University of California, San Diego, La Jolla, California  
92093, U.S.A.

**B.W. ZWEIFACH** Department of Applied Mechanics and Engineering  
Sciences—Bioengineering, University of California, San Diego, La Jolla,  
California 92093, U.S.A.

Cover art by Venus K. Sabin. Used with permission.

Library of Congress Cataloging in Publication Data

Main entry under title:

Frontiers in biomechanics.

Bibliography: p.

Includes index.

1. Biomechanics—Congresses. 2. Human mechanics—Congresses. 3. Fung, Y.C. I. Fung, Y.C. (Yuan-cheng), 1919— . II. Schmid-Schönbein, G.W. III. Woo, Savio L-Y. IV. Zweifach, Benjamin W. (Benjamin William), 1910— QP303.F69 1985 612.014 85-17157

© 1986 by Springer-Verlag New York Inc.

Softcover reprint of the hardcover 1st edition 1986

All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

While the advice and information in this book are believed to be true and accurate at the date of going to press, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Typeset by Asco Trade Typesetting Ltd., Hong Kong.

9 8 7 6 5 4 3 2 1

ISBN-13: 978-1-4612-9337-8 e-ISBN-13: 978-1-4612-4866-8

DOI 10.1007/978-1-4612-4866-8

# Foreword

Biomechanics is concerned with the response of living matter to forces, and its study has taken long strides in recent years. In the past two decades, biomechanics has brought improved understanding of normal and pathophysiology of organisms at molecular, cellular, and organ levels; it has helped developing medical diagnostic and treatment procedures; it has guided the design and manufacturing of prosthesis and instruments; it has suggested the means for improving human performance in the workplace, sports, and space; it has made us understand trauma in war and in peace. Looking toward the future, we see many more areas of possible development such as:

- reduction in heart diseases and atherosclerosis
- improved vascular assist and replacement devices, including a permanent artificial heart
- enhanced oxygen transport in the lung
- understanding and control of growth and changes
- mechanics of neuromuscular control and robotics
- prevention of joint degeneration
- permanent total joint replacements
- prevention of low back pain
- workplace designs to enhance productivity
- ambulation systems for the handicapped
- fully implantable hearing aids
- improved understanding of the mechanisms for permanent disability injuries
- identification of factors such as alcohol use and disease influence on impact tolerance
- improved cellular bioreactor designs
- mechanics of DNA and its application in biotechnology.\*

Obviously, the attainment of these prospects will greatly improve the quality of human life and reduce the costs of living.

---

\* This list is from a report by the U.S. National Committee on Biomechanics, April, 1985.

Looking at the frontiers of science is always great fun. This is what this book offers. I derived a tremendous pleasure reading it. Waves of excitement went through me. Again and again the ideas here generated in my mind beautiful vistas of the future and immediate plans for new investigations to be made.

These descriptions of the scenes at the frontiers of biomechanics are stories of considerable accomplishment. Wang Kuo-wei has analyzed the course of any significant accomplishment. He asserts that any person making a significant contribution must have gone through three stages: the stage of seeing the prospects, the stage of hard work and frustration, and the stage of sudden and unexpected success. He culled three stanzas from three famous Chinese poems to describe the moods of these stages,\* and Donald Hall and Chia-shun Yih<sup>†</sup> translated them into English as follows:

## I

Last night the west wind rising  
stripped leaves from the green tree.  
I climbed alone to the top floor  
of my house, and looked up the road  
as far as the horizon.

## II

For love of her, I grew pale;  
I turned thin and haggard.  
For love of her, the belt  
hung loose on my robe.  
... Yet I regretted nothing.

## III

I looked for her in the crowd  
a hundred, a thousand times,  
then suddenly, as I turned  
my head, I glimpsed her  
in the dim light, in half-darkness.

Every author who has contributed to this book recognizes these moods. Each scientific discipline has its own life cycles, too, and each cycle goes through the stages Wang alluded to. At what stage is biomechanics? I think it is mostly in the first two, and the evidences can be seen in this book. Every chapter

\* Wang Kuo-wei (1877–1927) was a famous archaeologist, historian, and poet. These remarks were given in his book “On Living Poetry” (王國維, 人間詞話, 開明書店, p. 16). The first stanza is the work of Yen Shu (991–1055, 晏殊, 蝶戀花), the second is the work of Liu Yung (990–1050?, 柳永, 凤棲梧), and the third is the work of Hsin Chi-hsi (1140–1207, 辛棄疾, 青玉案).

<sup>†</sup> Donald Hall is an American poet. Chia-shun Yih is a fluid dynamicist and mathematician. These translations were read at a banquet in the Hyatt Regency Hotel, Mission Bay, San Diego, on 22 July 1984.

opens up a perspective to a larger field, suggesting many possible applications to engineering, industry, and the health of people and animals. Most of them have not yet arrived at the state of perfection as described by Dirichlet or Gauss's work: finished sculpture on which all chisel marks have been polished off. And this is probably good for students, who need the chisel marks to guide them.

This book is written by people I know and to whom I am affectionately attached. It is impossible to describe the deep emotion of the individual who stands at the receiving end of the dedication from them. I thank one and all of them.

Y.C. Fung

# Preface

Advances in every aspect of scientific knowledge during the past decade have been catalyzed not so much through the development of sophisticated instrumentation, but through the marriage and interaction between hitherto separate basic sciences. Nowhere is this situation better exemplified than with respect to the enormous impact that the application of the tools and methods of applied mechanics has had on the life sciences. The creation of the new discipline of biomechanics has been formalized through the development of new models and constitutive equations. This approach has added new dimensions to applied mechanics in the exploration of structure–function relationships in living materials.

A recent symposium in honor of the 65th birthday of Professor Y.C. Fung was timely not only because it dealt with innovative and provocative aspects of biomechanics, but also because it exemplified the substantial advances made possible by an interdisciplinary approach. This volume, a compendium of the symposium entitled “Frontiers of Applied Mechanics and Biomechanics” held at the University of California, San Diego, in July 1984, covers a spectrum of topics at the core of modern biomechanics research. It is clear that an interdisciplinary approach involves not only the transcription of established methodology from one field to another, but also the formulation of new field equations dealing with the unique properties of biological materials and systems.

In arranging the topics to be included in this volume, primary consideration was given to the suitability of the coverage and contents for graduate student use in biomechanics. A major objective of the book is to provide sufficient background material, terminology, specific formulations, and literature references to guide the reader in further exploration into biomechanics research. No attempt is made to be all-inclusive, but rather to develop representative areas that reflect the contributions of the Fung school of thinking as a basis for the directions in which biomechanics is moving.

The choice of the honoree, Professor Y.C. Fung, was especially propitious, not only because he so uniquely exemplifies the insatiable curiosity and singularity of purpose so necessary for the emergence of biomechanics as a

distinctive discipline, but also because he has been so instrumental in opening up the field and providing many of the basic analytical tools needed. Interdisciplinary activity requires the strong personal commitment and vision of key individuals whose insight and guidance serve as a catalyst. Y.C. Fung is one of these rare individuals. His impact on the field of biomechanics has been decisive, and his leadership has stimulated students and colleagues to join him in the endeavor of biomechanics. This volume is dedicated to him as a token of the esteem and admiration in which he is held by the entire scientific community. He has made the long road ahead much brighter and smoother.

G.W. Schmid-Schönbein  
S.L-Y. Woo  
B.W. Zweifach

# Contents

## Part I Molecules and Cells

1. Membrane Mechanics and Cell Adhesion E.A. EVANS	3
2. New Birth To Runaway Solitons T.Y. WU	18
3. Leukocyte Protopodia G.W. SCHMID-SCHÖNBEIN, R. SKALAK, S. CHIEN	33

## Part II Biological Tissues

4. Biomechanics of Smooth Muscle J.M. PRICE	51
5. Smooth Muscle Rheology: In Search of a Specimen H.Y.-L. CHEN, E.F. OWENS JR.	62
6. The Phenomenology of the Heart Muscle J.G. PINTO	77
7. Determination of Material Properties of Biological Tissues: Pericardium P.H. CHEW, S.L. ZEGER, F.C.P. YIN	86
8. <i>In-Vivo</i> Measurement of Regional Strains in Myocardium L.K. WALDMAN	98
9. Residual Stress in Arteries C.J. CHUONG, Y.C. FUNG	117

10. On the Structural Origin of the Quasilinear Viscoelastic Behavior of Tissues Y. LANIR	130
11. Fibrous Skin Mechanics: Superstructure and New Problems S.A. WAINWRIGHT	137
12. Mechanism of Osmosis: Hulett's Versus Lewis' View of Altered Solvent in Solution H.T. HAMMEL, A.R. HARGENS	142

### Part III Orthopedics

13. A Finite Deformation Theory for Nonlinearly Permeable Soft Hydrated Biological Tissues V.C. MOW, M.K. KWAN, W.M. LAI, M.H. HOLMES	153
14. Biomechanics of Tendons and Ligaments S.L-Y. WOO	180
15. Bone Mechanics: From Tissue Mechanical Properties to an Assessment of Structural Behavior W.C. HAYES	196
16. Biomechanics of the Lumbar Spine A.I. KING, K.-H. YANG	210
17. Biomechanics of the Human Gait E.Y.S. CHAO	225
18. Biomechanics of Bone-Implant Interactions R. HUISKES	245
19. Mechanics of Healing Soft Tissue Wounds A. VIIDIK, F. GOTTRUP	263

### Part IV Circulation

20. Biomechanics of the Microcirculation B.W. ZWEIFACH	283
21. Biomechanics of Capillary Blood Flow R. SKALAK, N. ÖZKAYA, T.W. SECOMB	299

22. The Static Versus the Dynamic State of the Microvasculature M. INTAGLIETTA	314
23. Vascular Dynamics and the Endothelium R.M. NEREM, M.J. LEVESQUE, M. SATO	324
24. Self-excited Oscillation of a Collapsible Tube Conveying Fluid Y. MATSUZAKI	342
 <b>Part V The Lung</b>	 351
25. Microvascular Hematocrit of the Lung J.S. LEE	353
26. Pulmonary Blood Flow in the Cat: Correlation Between Theory and Experiment R.T. YEN, S.S. SOBIN	365
27. Respiratory Dynamics—Computer Simulation Y. SEGUCHI, Y.C. FUNG, T. ISHIDA	377
Index	393

# Contributors

**E.Y.S. CHAO** Director, Biomechanics Research, The Mayo Clinic, Rochester, Minnesota 55901, U.S.A.

**H.Y.-L. CHEN** School of Engineering Science and Mechanics, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A.

**P.H. CHEW** Department of Physiology, The Johns Hopkins Medical Institutions, Baltimore, Maryland 21205, U.S.A.

**S. CHIEN** Department of Physiology, College of Physicians and Surgeons, Columbia University, New York, New York 10032, U.S.A.

**C.J. CHUONG** Biomedical Engineering Program, University of Texas at Arlington, Arlington, Texas 76019, U.S.A.

**E.A. EVANS** Department of Pathology, University of British Columbia, Vancouver, B.C., V6T 1W5, Canada.

**Y.C. FUNG** Department of Applied Mechanics and Engineering Sciences—Bioengineering, University of California, San Diego, La Jolla, California 92093, U.S.A.

**F. GOTTRUP** Department of Connective Tissue Biology, Institute of Anatomy, University of Aarhus, University Park, DK-800 Aarhus C, Denmark.

**H.T. HAMMEL** Scripps Institute of Oceanography, University of California, San Diego, La Jolla, California 92093, U.S.A.

**A.R. HARGENS** Department of Surgery, Veteran's Administration Medical Center, La Jolla, California 92037, U.S.A.

**W.C. HAYES** Department of Orthopedic Surgery, Beth Israel Hospital, Harvard Medical School, Boston, Massachusetts 02215, U.S.A.

M.H. HOLMES Department of Mathematics, Rensselaer Polytechnic Institute, Troy, New York 12181, U.S.A.

R. HUISKES Laboratory of Experimental Orthopedics, Biomechanics Section, University of Nijmegen, 6500 HB Nijmegen, The Netherlands.

M. INTAGLIETTA Department of Applied Mechanics and Engineering Sciences —Bioengineering, University of California, San Diego, La Jolla, California 92093, U.S.A.

T. ISHIDA Department of Systems Engineering, Kobe University, Rokkodai Nada, Kobe 657, Japan.

A.I. KING Biomechanics Research Center, Wayne State University, Detroit, Michigan 48202, U.S.A.

M.K. KWAN Department of Mechanical Engineering, Aeronautical Engineering and Mechanics, Rensselaer Polytechnic Institute, Troy, New York 12181, U.S.A.

W.M. LAI Department of Mechanical Engineering, Aeronautical Engineering and Mechanics, Rensselaer Polytechnic Institute, Troy, New York 12181, U.S.A.

Y. LANIR Department of Biomedical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel.

J.S. LEE Division of Biomedical Engineering, University of Virginia Medical Center, Charlottesville, Virginia 22908, U.S.A.

M.J. LEVESQUE Physiological Fluid Mechanics Laboratory, Department of Mechanical Engineering, University of Houston, Houston, Texas 77004, U.S.A.

Y. MATSUZAKI Department of Aeronautical Engineering, Nagoya University, Chikusa-ku, Nagoya, Japan 464.

V.C. Mow Department of Mechanical Engineering, Aeronautical Engineering and Mechanics, Rensselaer Polytechnic Institute, Troy, New York 12181, U.S.A.

R.M. NEREM Physiological Fluid Mechanics Laboratory, Department of Mechanical Engineering, University of Houston, Houston, Texas 77004, U.S.A.

E.F. OWENS JR. School of Engineering Science and Mechanics, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A.

N. ÖZKAYA Bioengineering Institute, Department of Civil Engineering and Engineering Mechanics, Seeley W. Mudd Building, Columbia University, New York, New York 10027, U.S.A.

J.G. PINTO Department of Mechanical Engineering, San Diego State University, San Diego, California 92182, U.S.A.

J.M. PRICE Department of Physiology, University of South Florida, College of Medicine, Tampa, Florida 33612, U.S.A.

M. SATO Physiological Fluid Mechanics Laboratory, Department of Mechanical Engineering, University of Houston, Houston, Texas 77004, U.S.A.

G.W. SCHMID-SCHÖNBEIN Department of Applied Mechanics and Engineering Sciences—Bioengineering, University of California, San Diego, La Jolla, California 92093, U.S.A.

T.W. SECOMB Department of Physiology, University of Arizona, Tucson, Arizona 85724, U.S.A.

Y. SEGUCHI Department of Systems Engineering, Kobe University, Rokkodai Nada, Kobe 657, Japan.

R. SKALAK Bioengineering Institute, Department of Civil Engineering and Engineering Mechanics, Seeley W. Mudd Building, Columbia University, New York, New York 10027, U.S.A.

S.S. SOBIN American Heart Association—Greater Los Angeles Affiliates, Cardiovascular Research Laboratory, University of Southern California, and Los Angeles County, University of Southern California Medical Center, Los Angeles, California 90033, U.S.A.

A. VIIDIK Department of Connective Tissue Biology, Institute of Anatomy, University of Aarhus, DK-800 Aarhus C, Denmark.

S.A. WAINWRIGHT Department of Zoology, Duke University, Durham, North Carolina 27706, U.S.A.

L.K. WALDMAN Department of Medicine, Cardiology, University of California, San Diego, La Jolla, California 92093, U.S.A.

S.L-Y. Woo Department of Surgery, Division of Orthopedics and Rehabilitation, University of California, San Diego, La Jolla, California 92093, U.S.A.

T.Y. WU Engineering Science Department, California Institute of Technology, Pasadena, California 91125, U.S.A.

**K.-H. YANG** Biomechanics Research Center, Wayne State University, Detroit, Michigan 48202, U.S.A.

**R.T. YEN** Department of Applied Mechanics and Engineering Sciences—Bioengineering, University of California, San Diego, La Jolla, California 92093, U.S.A.

**F.C.P. YIN** Department of Medicine, The Johns Hopkins Medical Institutions, Baltimore, Maryland 21205, U.S.A.

**S. ZEGER** Department of Biostatistics, The Johns Hopkins Medical Institutions, Baltimore, Maryland 21205, U.S.A.

**B.W. ZWEIFACH** Department of Applied Mechanics and Engineering Sciences—Bioengineering, University of California, San Diego, La Jolla, California 92093, U.S.A.

# **Part I Molecules and Cells**

# 1

# Membrane Mechanics and Cell Adhesion

E.A. Evans

## 1. Introduction

Membrane–membrane recognition and adhesion are prominent processes in biology. In general, these events involve specific molecular binding and cross-bridging reactions. Formation of adhesive contact induces tensions in the membranes that ultimately limit the extent of contact. Subsequent separation by physical force also creates stresses that are transmitted through the membrane to the contact zone. The general impression in biology is that the factors that influence cell adhesion simply alter the chemical attraction between opposing membrane surfaces, and that the attraction is the sole determinant of the process. However, closer inspection shows that other physical factors play equally important roles in adhesion–separation processes. Most obvious are the external forces that act on cell or vesicle bodies produced by shear stresses in convecting suspensions. Less obvious, but intrinsically significant, is the mechanical rigidity of the cell or vesicle because it directly regulates the adhesion process. Simply stated: “rigid” bodies have limited ability to adhere (unless they fit together perfectly), whereas “flaccid” or easily deformed bodies can form large contact areas with little energy expense. In this chapter, a brief outline of the mechanical properties of membranes will be given first, followed by a description of the detailed mechanics of cell membrane–membrane adhesion. The latter will be considered for two situations: (1) where the cross-bridging forces are continuously distributed as normal stress; and (2) where the cross-bridging forces are present at discrete receptor sites. These two cases represent a wide range of adhesion phenomena in biology, ranging from blood cell rouleau formation and aggregation in high molecular-weight polymers and blood plasma to the agglutination of cells via discrete molecular cross-bridges produced by plant lectins and monoclonal antibodies.

## 2. Membrane Mechanical Properties

Early experimenters concluded that cell membranes are composite materials made up of a bimolecular layer of amphiphilic molecules (phospholipids, etc.), plus additional materials presumed to be proteins that support the surface

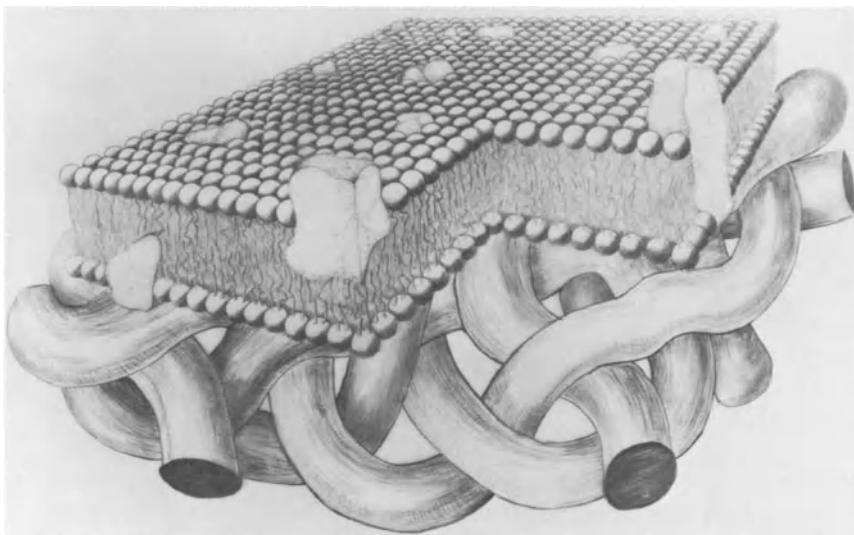


FIGURE 1.1. Schematic view of cell membranes as composite, lamellar structures. Reprinted with permission from *Mechanics and Thermodynamics of Biomembranes*, by E.A. Evans and R. Skalak (1980). Copyright CRC Press, Inc., Boca Raton, Florida.

liquid structure (Danielli and Davson, 1935; Singer and Nicholson, 1972). This general picture of the cell membrane ultrastructure has been confirmed and refined in recent years by advances in electron microscopy, biochemistry, and other microbiological sciences. Figure 1.1 shows a schematic illustration of the surface liquid bilayer supported by a protein "cytoskeleton." This anisotropic, or lamellar, configuration of thin membrane structures is peculiar to the preferential assembly of the molecular amphiphiles where the rate of exchange of membrane molecules with the adjacent aqueous phases is negligibly slow. Because it is only a few molecules thick, a biological membrane can only be considered as a continuum in the two dimensions that characterize the surface. As such, continuum properties represent the material structure over a scale of fractions of a micron on the surface. Because of the microscopic size of cells, measurement of forces and deformations are extremely difficult and often imprecise. Consequently, progress in membrane mechanics has not come easily.

Experiments designed to determine mechanical properties of biological membranes have been carried out for more than 50 years, beginning with tests of sea urchin eggs (Cole, 1932) and nucleated red blood cells (Norris, 1939). The next attempts at measurement of membrane elastic moduli were based on red cell pressurization experiments (Katchalsky et al, 1960; Rand, 1964; Rand and Burton, 1964); although these experiments were innovative, they failed to provide an adequate model for membrane constitutive behavior. Fung (1966) was the first to seriously consider mechanical equilibrium and deformation of

cell membranes from the viewpoint of thin shell theory. He pointed out the significance of bending and extensional stiffnesses in regard to changes in shape of red blood cells; he recognized that cell deformation would require either stretch of the membrane surface or buckling and folding. Further sophistication in the material science of biological membranes was provided by Fung and co-workers (1968) through the introduction of large deformation theory into the analytical approach. With these developments, the foundations had been laid for quantitative analysis of mechanical deformations of cell membranes. Since that time, many experiments have been designed and carried out to directly measure the material properties of cell membranes and membranes of synthetic phospholipid bilayer vesicles (Hochmuth and Mochadas, 1972; Evans, 1973; Skalak et al, 1973; Evans and LaCelle, 1975; Evans and Hochmuth, 1976; Chien, 1978; Hochmuth, 1979; Waugh and Evans, 1979; Evans, 1983).

Constitutive relations for elastic and viscous behavior of membranes have been derived from general principles of mechanics and thermodynamics that are consistent with the biochemical and ultrastructural developments (see Evans and Skalak, 1980, for details). The simplest rheological model for membrane behavior is represented by elastic and viscous coefficients for three independent modes of membrane deformation and rate of deformation (illustrated in Fig. 1.2): (1) membrane area expansion or condensation; (2) membrane extension (shear) without change in surface area; and (3) membrane bending or folding. This simple model has been used satisfactorily to correlate material properties with direct micromechanical tests. Measurement of the static coefficients of rigidity (elastic moduli) for each mode of deformation has resulted in a hierarchy of membrane stiffnesses, which are presented in Table 1.1. The most significant observation is that biological membranes greatly resist area dilation but can be easily deformed by extension or bending if the membrane surface area or density is not required to change. Since the internal volumes of cells or vesicles are held constant by the osmotic strength of the contents, it is apparent that the surface area and volume establish extrinsic geometric constraints for deformation, (i.e., constant surface area and constant volume). In addition to the static rigidity of a membrane capsule, there is dynamic resistance to deformation. The rate of deformation is limited by viscous dissipation in the cell membrane and cytoplasm. Measurement and analyses of viscous elastic recovery of cell membranes after extension have shown that dissipation in the membrane greatly exceeds that in the cytoplasm. The results for surface viscosity are also shown in Table 1.1. Even more complicated rheological behavior (such as plastic yield and flow, creep and relaxation, etc.) have been studied (Evans and Skalak, 1980).

In an adhesion or separation process, both elastic and viscous stresses oppose adhesive forces. For stationary (time-independent) configurations, only elastic stresses are involved in the mechanical equilibrium in competition with the forces of attraction. Viscous stresses limit the rate of contact formation or separation, but not the extent. Viscous forces are significant in the

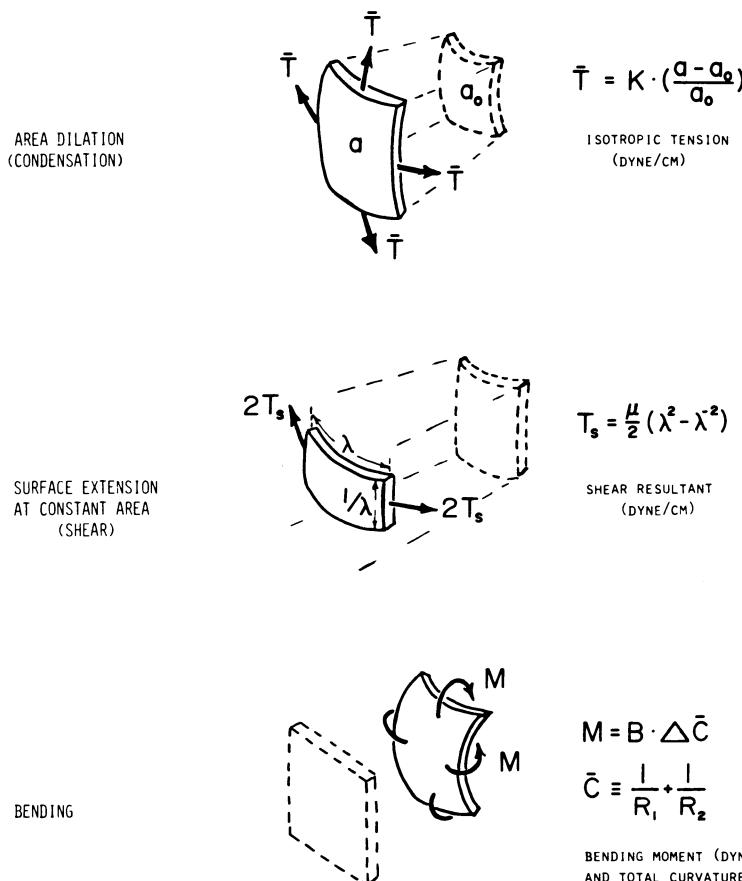


FIGURE 1.2. Schematic of the independent modes of membrane deformation: (top) *area dilation*; (middle) *shear or extension at constant area*; and (bottom) *bending*. Also shown are the force and moment resultants associated with each mode of deformation plus the first-order elastic proportionality. From Evans and Parsegian (1983) with permission from the New York Academy of Sciences.

TABLE 1.1. Membrane surface, elastic properties.

	Red blood cells	Lipid bilayer vesicles
$K_T$ (dyn/cm)	$4 \times 10^2$	$10^2$
$\mu$ (dyn/cm)	$10^{-1}$	0
$B$ (dyn/cm)	$10^{-12}$	$10^{-12}$
Viscous Properties		
$\eta^a$ (dyn-s/cm)	$10^{-3}-10^{-2}$	$10^{-6}-10^{-5}$

<sup>a</sup>Surface extensional or shear viscosity.

kinetic balance of aggregation–disaggregation processes in suspension. In the mechanical analyses of membrane–membrane adhesion and separation to follow, only the static mechanical equilibrium is considered, which provides the lower bound to the stresses associated with separation of contact and the upper bound for the stresses associated with contact formation.

### 3. Mechanics of Membrane–Membrane Adhesion and Separation: Continuum of Molecular Cross-Bridges

The analysis to be outlined here considers cross-bridging and adhesion forces as finite range interactions, distributed continuously over the surface (Evans, 1984a). Thus, at equilibrium, there are two membrane regions: (1) a free (unbridged) zone where the membranes are not subject to attractive forces; and (2) an adherent (bridged) zone where the membranes are held together by attractive stresses. This is illustrated in Figure 1.3. The membrane is treated as an elastic continuum in which the attractive stresses are assumed to act normal to the membrane surface. The approach is to analyze the membrane mechanics for each zone separately and then to require continuity of the solutions at the interface between the two zones. In general, membrane curvature proximal to the contact is quite large for an arc normal to the edge of the contact zone; by comparison, the curvature of an arc tangent to (or concentric with) the contact zone is very small. Thus, the problem need only be considered in the meridional plane normal to the edge of the contact and depends only on the curvilinear coordinates ( $s, \theta$ ) of the membrane, as illustrated in Figure 1.3. The membrane supports the intensive forces shown in Figure 1.3: a principal tension,  $T_m$ , and a transverse shear,  $Q_m$  (Evans and Skalak, 1980). The transverse shear is the resultant of bending stresses (moments) in the membrane that are localized to the sharp bend adjacent to and within the adherent zone. In the macroscopic region away from the adherent zone, it is assumed that the membrane is a plane structure under uniform tension,  $T_m^0$ ; here, the membrane forms a macroscopic (observable) contact angle of  $\theta_o$  with respect to the other surface.

The local mechanical equilibrium of the membrane in the free, unbridged region is given by the following equations (Evans and Skalak, 1980): the balance of forces tangent to the surface,

$$\frac{dT_m}{ds} - Q_m \cdot K_m = 0; \quad (1.1)$$

next, the balance of forces normal to the surface,

$$T_m \cdot K_m + \frac{dQ_m}{ds} = 0, \quad (1.2)$$

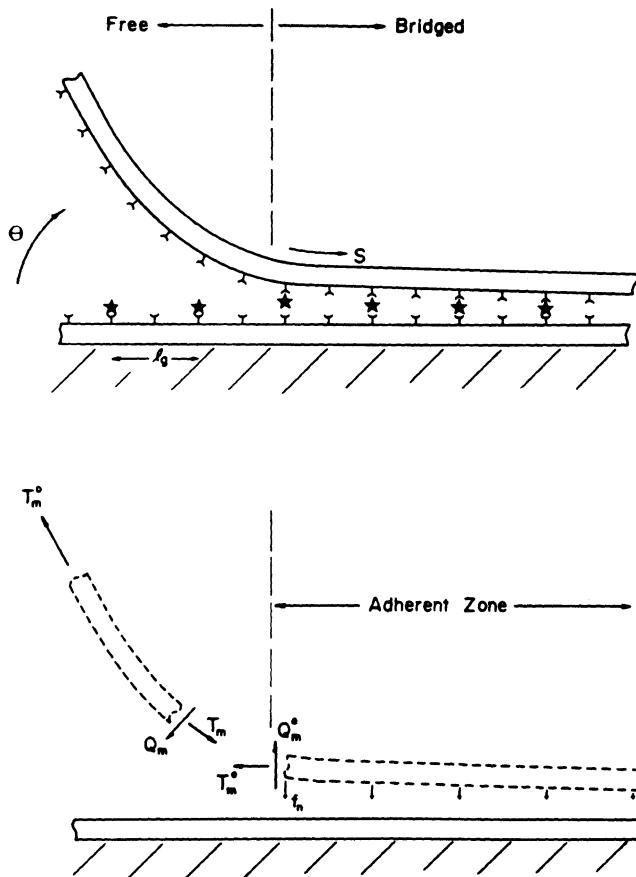


FIGURE 1.3. Schematic illustration of the adherent (bridged) and free (unbridged) zones proximal to the edge of the contact zone. The discrete forces that arise from cross-bridges between membrane receptors are illustrated, as well as the intensive forces supported by the membrane surface itself. In addition to the attractive forces,  $f_n$ , the membrane supports a principal tension,  $T_m$ , which acts tangent to the plane of the membrane, and a transverse shear,  $Q_m$ , which acts normal to the membrane surface.

where  $K_m$  is the local curvature of the meridional arc. In order to solve these two equations for the unknowns ( $T_m$ ,  $Q_m$ ,  $K_m$ ), the elastic constitutive behavior for the membrane must be introduced. In this region of high curvature, the bending or curvature elasticity determines the membrane shape. The elastic constitutive relation for membrane-bending yields the relation that the transverse shear is proportional to the gradient of the curvature,

$$Q_m = -B \cdot \frac{dK_m}{ds}. \quad (1.3)$$

Solution to these equations is easily obtained as a function of curvilinear angle,  $\theta$ , and the macroscopic tension applied to the membrane in the region far from the contact,

$$\begin{aligned} T_m &= T_m^\circ \cdot \cos(\theta_o - \theta) \\ Q_m &= T_m^\circ \cdot \sin(\theta_o - \theta) \\ (K_m)^2 &= \left( \frac{2 \times T_m^\circ}{B} \right) [1 - \cos(\theta_o - \theta)] + (K_m^\circ)^2. \end{aligned} \quad (1.4)$$

It is apparent that the transverse shear and membrane curvature increase to maxima that depend on the microscopic contact angle,  $\theta^*$ , at the edge of the contact zone. The tension, transverse shear, curvature, and microscopic contact angle must be continuous with the solution for these variables in the adherent (bridged) zone.

Turning now to the adherent zone, the equations of mechanical equilibrium in this region include the attractive normal stress,  $\sigma_n$ , illustrated in Figure 1.3. Hence, the local balance of normal forces becomes

$$T_m \cdot K_m + \frac{dQ_m}{ds} = -\sigma_n. \quad (1.5)$$

The tangential force balance remains the same as in Equation (3). At this point, a constitutive relation is needed for the normal stress in terms of displacement of the membranes from equilibrium (planar) contact. The displacement of the bond from equilibrium is represented by the variable  $\zeta$ ; the cross-bridge force is approximated by the linear relation given by the strength of the bond,  $\hat{f}_n$ , divided by the bond length,  $l_b$ , multiplied by  $\zeta$ .

In the limit that the cross-bridges can be considered as a continuous distribution, the normal stress is modeled by the product of the surface density of cross-bridges,  $\tilde{n}$ , times the force–displacement relation. For this model, the work per unit area for formation or separation of necessary cross-bridges is given by

$$\gamma \equiv (\tilde{n} \cdot \hat{f}_n \cdot l_b)/2 \quad (1.6)$$

where  $\gamma$  is the adhesion energy. In terms of the adhesion energy, the normal stress relation can be written as

$$\sigma_n = (2\gamma/l_b^2) \cdot \zeta \quad (1.7)$$

For membrane angles (measured relative to the equilibrium contact plane) that are less than about  $30^\circ$ , the angle and curvature can be well approximated by the first and second spatial derivatives of the displacement from the equilibrium plane. With the constitutive relations plus these derivatives, the local balance of normal forces, Equation (1.5), becomes

$$-T_m \cdot \frac{d^2\zeta}{ds^2} + B \cdot \frac{d^4\zeta}{ds^4} = -\left(\frac{2\gamma}{l_b^2}\right) \cdot \zeta, \quad (1.8)$$

which, in general, must be solved simultaneously with the balance of tangential forces, Equation (1.1). The tension term makes the Equation (1.8) nonlinear, but is a higher order (smaller) term in comparison to the bending stress term. Consequently, the variation in tension can be neglected, and the tension can be assumed constant just inside the contact zone as a first-order approximation; this will allow Equation (1.8) to be solved analytically. The solution is the membrane displacement from equilibrium in the contact zone,

$$\zeta = e^{-\alpha_1 \times s} \cdot [C_1 \cdot \sin(\alpha_1 \cdot s) + C_2 \cdot \cos(\alpha_2 \cdot s)], \quad (1.9)$$

where the spatial variations are governed by the parameters  $\alpha_1$  and  $\alpha_2$ ,

$$\alpha_1 = [t + \theta_\alpha^2]^{1/2}/l_b, \quad \alpha_2 = [-t + \theta_\alpha^2]^{1/2}/l_b$$

and

$$\theta_\alpha \equiv (\gamma \cdot l_b^2/2B)^{1/4}, \quad t \equiv T_m \cdot l_b^2/4B.$$

The displacement is maximum at the edge of the contact zone and decreases exponentially with distance into the contact. The effective width of the “boundary layer” where molecular bonds are stretched near the edge of the contact zone is given by

$$\delta \sim l_b/\theta_\alpha = (2 \cdot B \cdot l_b^2/\gamma)^{1/4}. \quad (1.10)$$

It is apparent that the width,  $\delta$ , will be large when the membrane is “stiff” (i.e., large bending modulus) and will be small when the adhesion energy is strong (i.e., strong attractive forces between the membranes).

From the contour given by Equation (1.9), the microscopic contact angle, curvature, and transverse shear at the edge of the contact zone are specified by the first, second, and third spatial derivatives respectively,

$$\begin{aligned} \theta^* &= \alpha_1 \cdot C_2 - \alpha_2 \cdot C_1, & K_m^* &= (\alpha_2^2 - \alpha_1^2) \cdot C_2 + 2\alpha_1 \cdot \alpha_2 \cdot C_1 \\ Q_m^*/B &= [(\alpha_2^2 - \alpha_1^2) \cdot \alpha_1 + 2\alpha_1 \cdot \alpha_2^2] \cdot C_2 + [(\alpha_2^2 - \alpha_1^2) \cdot \alpha_2 + 2\alpha_1^2 \cdot \alpha_2] \cdot C_1. \end{aligned} \quad (1.11)$$

The displacement at the edge of the contact zone is equal to the maximum bond length,  $l_b$ ; thus there are only two unknown variables,  $\theta^*$  and  $C_1$ . These are determined by the requirements of continuity with the solution previously derived for the free unbridged zone.

The continuity at the edge of the contact zone of the solutions derived for the free (unbridged) and adherent (bridged) regions can only be satisfied for specific values of the macroscopic tension applied to the membrane,  $T_m^\circ$ , and the microscopic contact angle,  $\theta^*$ . Specific values for these variables are obtained as a function of the dimensionless parameter,  $\theta_\alpha$ , which represents the ratio of adhesion to bending energies. The ideal “Young” equation predicts a simple relation for macroscopic tension in terms of the macroscopic contact angle,  $\theta_o$ ,

$$T_m^\circ/\gamma = (1 - \cos \theta_o).$$

The macroscopic tension determined by the continuity requirements at the edge of the contact zone show that the classical Young equation is valid for finite-range molecular cross-bridging forces and nonzero microscopic contact angles. The results for the microscopic contact angle as a function of the ratio of adhesion to bending energies show that the microscopic contact angle is appreciable when the strength of adhesion is large or the bond length is large; however, when the membrane is “stiff,” the microscopic contact angle approaches zero. For shallow macroscopic contact angles (e.g., 30°), the microscopic contact angle approaches the macroscopic value as the ratio of adhesion to bending energies becomes large. When the microscopic contact angle becomes large, the width of the boundary layer,  $\delta$ , is small and there are few bonds stretched at the edge of the contact zone.

#### 4. Mechanics of Membrane–Membrane Adhesion and Separation: Discrete, Kinetically Trapped Molecular Cross-Bridges

In the previous section, the mechanics of membrane–membrane adhesion were examined for the case where the adhesive cross-bridging forces between the membranes are distributed continuously over the surfaces but with a finite extent of interaction away from each surface. For the continuum model, the tension necessary to oppose spreading of a membrane–membrane contact is equal to the minimum level of tension required to separate adherent membranes. Also, the macroscopic tension applied to the membrane in the free region away from the contact is given by the classical Young equation, which relates the free-energy reduction per unit area for formation of planar contact to the membrane tension and macroscopic contact angle. The continuum model has been shown to be a valid representation for adhesion and separation of synthetic, phospholipid bilayer membrane vesicles that were allowed to aggregate, either via van der Waal’s attraction (Evans and Metcalfe, 1984a) or in high-molecular weight glucose polymer (dextran) solutions (Evans and Metcalfe, 1984b). It also appears to be an appropriate model for red cell rouleaux formation (Buxbaum et al, 1982; Skalak et al, 1981).

In contrast with the classical model, experiments often show that there is little or no tendency for membrane–membrane contact to spread, even though the tension required to subsequently separate the contact is very large (Evans and Leung, 1984). The deviation of the level of tension induced in the membranes by contact formation from that required to separate the contact appears to be due to the sparse distribution of strong molecular cross-bridges. Thus, the purpose of this section is to consider the mechanics of adhesion and separation for discrete molecular cross-bridges (Evans, 1984b). The approach is to determine the membrane contour that minimizes the total free energy (membrane elastic energy of deformation plus cross-bridge energies) in the

contact zone. Two disparate values of the macroscopic tension are found: (1) the level of tension that stresses the first cross-bridge at the edge of the contact zone to the near breaking point; and (2) the reduced level of tension that will just permit the membranes to approach sufficiently for the next cross-bridge to form adjacent to the contact zone. Deviation between these two tensions can be very large and is shown to depend strongly on the surface density of cross-bridges.

As in the previous development, the analysis here considers the cross-bridging (adhesion) forces as finite-range interactions, and the membrane is treated as an elastic continuum. In the adherent zone, the total free energy functional includes membrane elastic deformation plus cross-bridge energies. The functional is minimized in conjunction with the work required to displace forces at the boundaries; the variation is taken with respect to the parameters that characterize the membrane contour in the contact zone. The same linear approximation to the intermolecular force,  $f_n$ , is employed here as used previously. The contribution to the total free energy from extension of the cross-bridges is given by the approximation

$$F_{CB} = \frac{1}{2}(\hat{f}_n \cdot l_b) \sum_{i=0,1}^N \zeta_i^2, \quad (1.12)$$

where  $N$  is the number of cross-bridges involved in the adhesive contact, and the product,  $F_b = \hat{f}_n \cdot l_b / 2$ , is the free energy change (work) that results from formation(−) or breakage(+) of a single cross-bridge. The initial value for the sum of cross-bridge energies ( $i = 1$  or 2) is determined by one of two situations: (1)  $i = 1$  when a cross-bridge is maximally stretched at the edge of the contact zone (i.e., separation of the adherent contact), (2)  $i = 2$  when a cross-bridge is about to be formed at the edge of the contact (given by the lattice position  $i = 1$ )—i.e., minimal spreading of the contact.

The free energy functional for elastic deformation of the membrane is given by

$$F_{ED} = \left(\frac{B \cdot l}{2} g\right) \int_0^{N \cdot l_g} (K_m - K_m^\circ)^2 ds + l_g \cdot \int_0^{N \cdot l_g} (T_m \cdot \varepsilon) ds, \quad (1.13)$$

where the first term is the bending or curvature elastic energy and the second term is the work for in-plane extension of the membrane (Evans and Skalak, 1980). The principal curvature of the contour is  $K_m$ ;  $\varepsilon$  is the measure of strain along the contour (i.e., the fractional extension of membrane elements). The free energy equations for membrane deformation and cross-bridge extension, Equations (1.12) and (1.13), represent the total energy of a strip of membrane that has a width of  $l_g$ . The parameter  $l_g$  is the average distance between cross-bridge sites. In the region of high curvature that characterizes the membrane contour close to and within the adherent zone, membrane bending energy determines the shape of the contour, so the in-plane (extensional) elasticity will be neglected. As such, the membrane tension acts as a Lagrange multiplier in Equation (1.13) with the auxiliary requirement that the local strain approach

zero. The result is that the density of cross-bridge sites remains uniform and constant.

Mechanical equilibrium is defined by the expression that the variation in total free energy is equal to the variation in work required to displace forces at the boundaries. The virtual work required to displace forces at the boundaries reduces to the virtual displacement of the macroscopic tension  $T_m^\circ$  supported by the membrane in the planar region away from the contact zone,

$$\delta W_B = T_m^\circ \cdot \delta s_0, \quad (1.14)$$

where  $\delta s_0$  is the virtual displacement of the membrane in the direction tangent to the surface. Equation (1.14) presumes that the membrane contour in the free (unbridged) zone is prescribed and that the membrane surface is inextensible.

The contour is defined by displacements ( $z, r$ ) from initial positions in the membrane ( $z$ , normal to the plane of the adhesive contact;  $r$ , parallel to the plane of the contact). Solutions to the variational expression for mechanical equilibrium are restricted to the class of contours that are locally inextensible;

$$\frac{dr}{ds} \left[ 1 + \frac{1}{2} \left( \frac{dr}{ds} \right) \right] = -\frac{1}{2} \left( \frac{dz}{ds} \right)^2. \quad (1.15)$$

Hence lateral displacements are determined by the normal displacements, and a single function,  $z(s)$ , completely specifies the contour. For this class of contours, the variational statement of equilibrium reduces to

$$0 = \left( \frac{\hat{f}_n \cdot l_b}{B} \right) \sum_{i=0,1}^N z_i \cdot \delta z_i + l_g \cdot \int_0^{N \cdot l_g} (K_m - K_m^\circ) \cdot \delta K_m \cdot ds - T_m^\circ \cdot \delta s_0. \quad (1.16)$$

The method of solution utilizes a Newton–Raphson iterative solution to Equation (1.11) in which the displacement function,  $z(s)$ , is a piecewise continuous function with continuous first and second derivatives.

The continuity requirements at the intersection with the unbridged zone yield specific values of the macroscopic tension for each macroscopic contact angle: (1) an upper value that represents the minimum level of tension required to separate the adhesive contact; and (2) a lower value that characterizes the maximum level of tension that will just allow the contact to spread. Figure 1.4 shows the two levels of tension determined for a macroscopic contact angle of 90° plus two ratios,  $l_g/l_b$ , of cross-bridge spacing to bond length, 1 : 1 and 10 : 1. The tension values depend on the dimensionless parameter,  $\theta_\alpha$ ,

$$\theta_\alpha \equiv (F_b \cdot l_b^2 / 2 \cdot B \cdot l_g^2)^{1/4},$$

which represents the ratio of adhesion to bending (deformation) energies. As expected, the two levels of tension approach the ideal value from the Young equation when the density of cross-bridge sites becomes large.

There is a specific membrane contour and microscopic contact angle associated with each equilibrium level of tension. Figure 1.5 shows the contours in the adherent zone determined for two ratios,  $l_g/l_b$ , of cross-bridge spacing

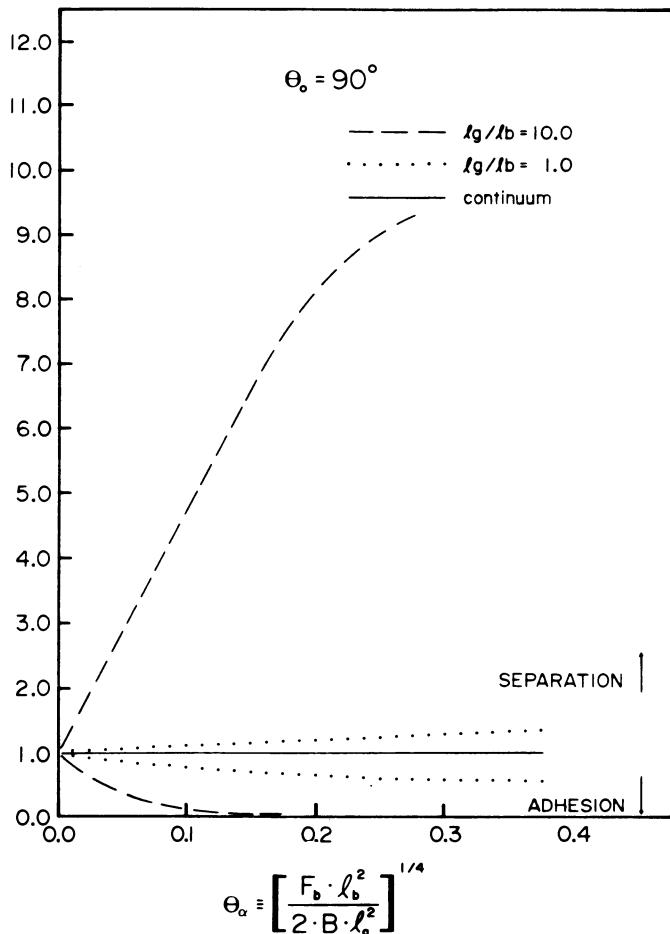


FIGURE 1.4. The macroscopic tension applied to the membrane in a plane region away from the contact zone, normalized by the adhesion energy per unit area, is plotted versus the dimensionless parameter that represents the ratio of adhesion to bending (deformation) energies. These results are for the specific case where the macroscopic contact angle is  $90^\circ$  and for two ratios,  $l_g/l_b$ , of cross-bridge spacing to bond length, 1 : 1 and 10 : 1. For each cross-bridge density ( $l_g/l_b$ ), two levels of tension are predicted: (1) an upper value that represents the minimum level of tension required to separate the adhesive contact; and (2) a lower value that characterizes the maximum level of tension that will just allow the contact to spread. Even for moderately dense cross-bridges ( $l_g/l_b = 1$ ), the tension for separation deviates from the level of tension that will allow the contact to spread. The solid line is the level of tension predicted from the classical Young equation, in which there is no deviation between the minimum tension required to separate adhesive contact and the tension that will allow the contact to spread.

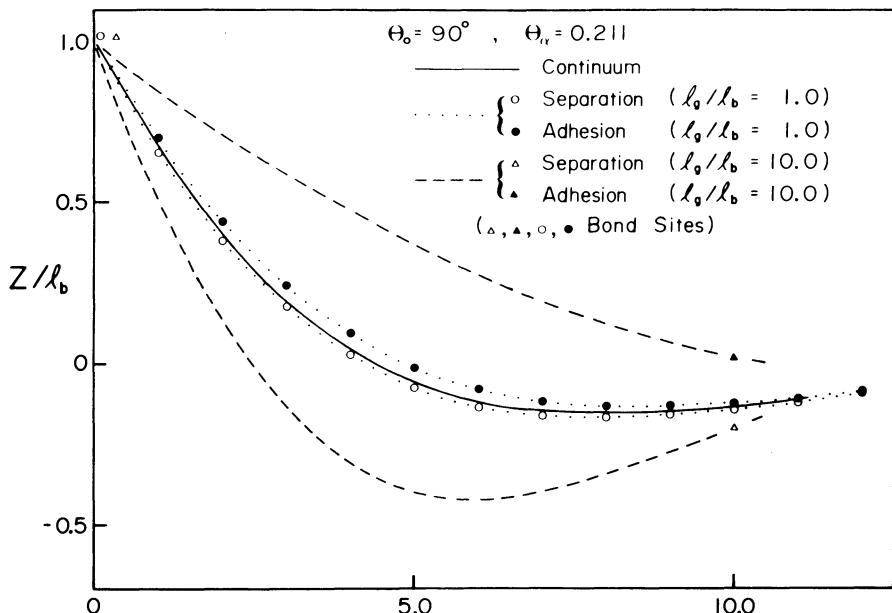


FIGURE 1.5. Membrane contours are shown for two ratios,  $l_g/l_b$ , of cross-bridge spacing to bond length, 1 : 1 and 10 : 1, and a specific value of the parameter for adhesion to bending energy ratio, 0.211, and a macroscopic contact angle of  $90^\circ$ . Also shown is the contour (solid line) derived previously from the continuum model, in which the cross-bridges were assumed to be infinitely dense (1). For relatively dense populations of cross-bridges, the contours deviate only slightly from that predicted by the continuum model. However, for low densities (i.e.,  $l_g/l_b = 10$ ), there is significant deviation of the contours from the continuum solution. Note: there are no bonds formed at the first receptor site (i.e., no solid circle and no star) for the situation where the contact is about to spread to the next site. The displacement axis is greatly enlarged in comparison to the axis for the curvilinear distance along the membrane.

to bond length, 1 : 1 and 10 : 1, and a specific value of the ratio of adhesion to bending energies. Also shown is the contour derived previously from the continuum model in which the cross-bridges are assumed to be infinitely dense. Figure 1.6 shows the progressive increase in microscopic contact angle at the edge of the contact zone as the ratio of adhesion to bending energies is increased. It is apparent that the contours (one for separation and the other for spreading of the contact) and microscopic contact angles deviate only slightly from the continuum solution when the cross-bridges are relatively dense (i.e.,  $l_g/l_b = 1$ ) whereas the contours and angles deviate appreciably from the continuum solution when the cross-bridge density is low (i.e.,  $l_g/l_b = 10$ ). It is also apparent that when the cross-bridge density is low, only the first cross-bridge is significantly stressed. Also, when the sites are far apart, the tension must be reduced to nearly zero in order to permit the next cross-bridge to form; thus,

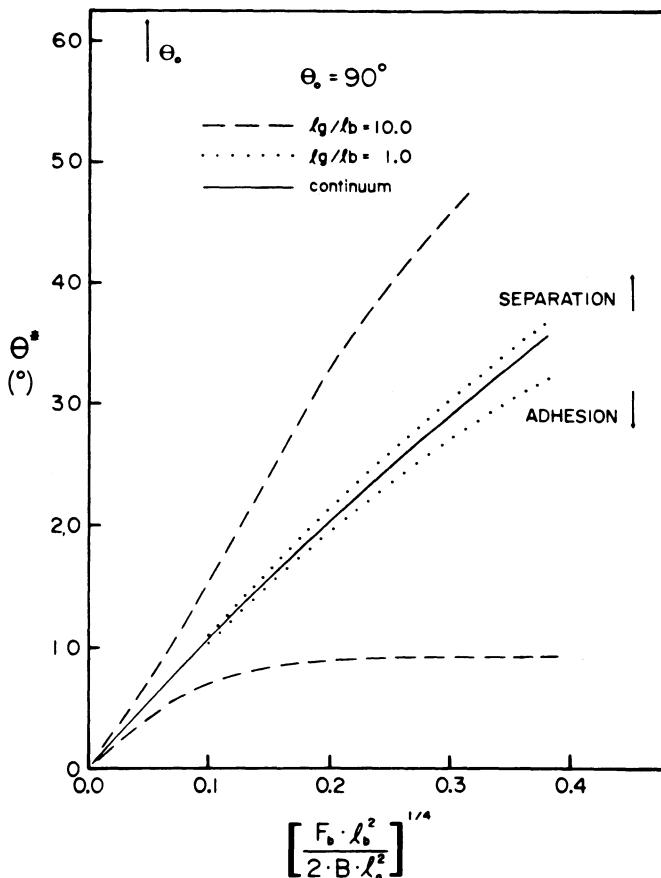


FIGURE 1.6. Values for the microscopic contact angle at the edge of the contact zone are plotted as a function of the parameter that represents the ratio of adhesion to bending energies for a specific value of the macroscopic contact angle of  $90^\circ$ . The solid line is the microscopic contact angle derived from the continuum solution in which the cross-bridges are assumed to be infinitely dense. There are two broken curves plotted for each ratio,  $l_g/l_b$ , of cross-bridge spacing to bond length, 1 : 1 and 10 : 1. The upper broken curves represent the situation where the contact is separated, and the lower curves are appropriate to the situation where the contact is about to spread.

there is little or no tendency for the contact to spread unless the surfaces are forced together. Since nearly all of the stress is taken by the first cross-bridge site at the edge of the contact zone when the density is low, the tension required to separate the contact greatly exceeds the value anticipated from the Young equation. When the cross-bridge density is low, the level of tension required to separate the contact approaches the value predicted by the Young equation multiplied by the ratio of the cross-bridge spacing to bond length.

## Acknowledgment

This work was supported in part by National Institutes of Health Grant HL 26965 and Canadian MRC Grant MT 7477.

## References

- Buxbaum, K., Evans, E.A., and Brooks, D.E. 1982. *Biochemistry* 21:3235–3239.
- Chien, S., Jan, K.-M., Skalak, R., and Tözeren, A. 1978. *Biophys. J.* 24:463–487.
- Cole, K.S. 1932. *J. Cell. Comp. Physiol.* 1:1–9.
- Danielli, J.F., and Davson, H. 1935. *J. Cell Comp. Physiol.* 5:495–508.
- Evans, E.A. 1973. *Biophys. J.* 13:941–954.
- \_\_\_\_\_. 1983. *Biophys. J.* 43:27–30.
- \_\_\_\_\_. 1985a. *Biophys. J.* 48:175–183.
- \_\_\_\_\_. 1985b. *Biophys. J.* 48:185–192.
- Evans, E.A., and Hochmuth, R.M. 1976a. *Biophys. J.* 16:1–11.
- Evans, E.A., and Hochmuth, R.M. 1976b. *Biophys. J.* 16:13–26.
- Evans, E.A., and LaCelle, P.L. 1975. *Blood* 45:29–43.
- Evans, E.A., and Leung, A. 1984. *J. Cell Biol.* 98:1201–1208.
- Evans, E.A., and Metcalfe, M. 1984a. *Biophys. J.* 46:423–426.
- Evans, E.A., and Metcalfe, M. 1984b. *Biophys. J.* 45:715–720.
- Evans, E.A., and Parsegian, V.A. 1983. In *Surfaces phenomena and hemorheology: theoretical, experimental and clinical aspects*, Copley and Seaman, eds. *Annals of N.Y. Acad. Sci.* 416:13–33.
- Evans, E.A., and Skalak, R. 1980. *Mechanics and thermodynamics of biomembranes*. CRC Press, Inc., Boca Raton, Florida. 254 pp.
- Fung, Y.C. 1966. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* 25:1761.
- Fung, Y.C., and Tong, P. 1968. *Biophys. J.* 8:175–189.
- Hochmuth, R.M., and Mohandas, N. 1972. *J. Biomech.* 5:501–509.
- Hochmuth, R.M., Worthy, P.R., and Evans, E.A. 1979. *Biophys. J.* 26:101–114.
- Katchalsky, A., Kedem, O., Klibansky, C., and DeVries, A. 1960. In *Flow properties of blood and other biological systems*., A.L. Copley and G. Stainsby, eds., Pergamon Press, New York.
- Norris, C.H. 1939. *J. Cell. Comp. Physiol.* 14:117–133.
- Rand, R.P. 1964. *Biophys. J.* 4:303–316.
- Rand, R.P., and Burton, A.C. 1964. *Biophys. J.* 4:115–135.
- Singer, S.J., and Nicolson, G.L. 1972. *Science* 175:720.
- Skalak, R., Tözeren, A., Zarda, R.P., and Chien, S. 1973. *Biophys. J.* 13:245–264.
- Skalak, R., Zarda, P.R., Jan, K.-M., and Chien, S. 1981. *Biophys. J.* 35:771–782.
- Waugh, R., and Evans, E.A. 1979. *Biophys. J.* 26:113–115.

# 2

# New Birth to Runaway Solitons

T.Y. WU

## 1. Introduction

A new frontier in biology and biomechanics is emerging. Its scope appears much broader than has been covered to date in the discipline, since the new viewpoint stresses first the molecular level as the basis for studies of biological phenomena involving cells and organisms. In this new field, enthusiastic workers from interrelated disciplines believe that, to improve our understanding of essential properties of living matter and to boost our ability to make valuable applications, the objective is perhaps best achieved by investigating the basic biological processes involving DNA, enzymes, and other participating molecules *in vivo*. It has been contended that, although the classic model of the DNA helix by James D. Watson and Francis Crick has been indeed invaluable, the structural models can only tell part of the story because they have been presented without life in them. The real DNA is rich in movement. It bends, vibrates, and resonates, and with such movements it goes on transporting and distributing biochemical energy that is needed for perpetuating life. To provide a concrete case germinated from the seed ideas conceived in "molecular biomechanics," the school of thought led by Davydov (1971, 1979) is recapitulated here, in Section 3, with the main result elucidating the significant role that can be played by solitary waves in bioenergetics.

This recent development certainly was not a chance encounter, for it has come at the height of blooming activities in the research of nonlinear waves. Various phenomena of solitons (more generally speaking, very long waves with lengths large compared to the transverse dimensions of the medium in which they propagate) have been found (mostly in the past two decades) to appear in many scientific fields, ranging from classical fluids, plasma and solid state physics, mechanics of polymer and crystal lattice, nonlinear optics and elementary particle physics, to bioenergetics and neurophysics. These major advances have been truly colorful, in both physical and mathematical contexts. New knowledge obtained of the essential properties of solitary waves from these studies, as summarized in Section 2, can help provide effective tools for future research.

However, our survey also reveals that the new research results have been carried out mostly for closed systems for which the generation of new solitary waves is not being considered. In the many different fields listed above, long waves can be generated to propagate and evolve within their own systems by certain forcing excitations according to the same or like principle. The primary objective of this study is to make a preliminary attempt to explore this common principle.

The specific case of study presented here is of the generation and evolution of long waves in shallow water as the result of a moving surface pressure or a moving boundary along the bottom of the water layer. The long-wave model adopted for the current research is the one recently given by Wu (1980, 1981). A series of numerical examples is presented, along with related experiments to bring out salient features of the nonlinear, dispersive waves so generated and evolving. A new finding of particular interest is that, after the turn-on of a two-dimensional disturbance, which is otherwise kept fixed in strength and velocity within a desired transcritical range, new solitary waves are generated, one after another, almost periodically, each to surge ahead in turn to form an upstream moving train of “runaway solitons.” This is a beautiful example illustrating that steady disturbances do not necessarily always result in steady responses.

Whether similar phenomena can manifest in other disciplinary fields, in biological molecules in particular, remains to be seen; and this will be impossible without further careful research.

## 2. Solitary Waves in Shallow Water—A Preliminary

In 1834, John Scott Russell (1838, 1845) was the first to observe the “singular and beautiful phenomenon” that a “wave of a large solitary elevation” formed, surged ahead free from a boat, when the boat drawn by a pair of horses suddenly stopped. Pursuing the runaway solitary wave on horseback, Russell found that the wave continued its own course along the channel without change of form and speed. This keen observation and Russell’s subsequent pioneering experimental studies brought many key features of solitary waves to the fore and attracted much scientific interest for some time. It drew such eminent work as that from Boussinesq (1871, 1872, 1877) and from Rayleigh (1876), who pointed out that a well-balanced interplay between the nonlinear and the dispersive effects is essential to a solitary wave’s maintaining its permanent form. Some 60 years after Russell’s discovery, Korteweg and deVries (1895) derived the equation bearing their names (recently known as the KdV equation) for study of unidirectional long water waves. These early contributions have helped in opening an important chapter of hydrodynamics.

In the past two decades, interest in long wave phenomena has been intense; it can be traced to the discovery of Zabusky and Kruskal (1965), who showed that solitary waves invariably emerge in the asymptotic solution of the KdV

equation; there are entities in these waves that they called solitons. Dynamic development of the field has ensued, partly because soliton phenomena have been discovered in various physical disciplines, and partly because soliton-bearing or soliton-like solutions are found to exist in many other mathematical models, including the cubic Schrödinger equation, the sine-Gordon equation, the Born-Infeld equation, and some of the equations to be discussed here. For these recent advances, we refer to the literature and reviews by, e.g., Whitham (1974), Scott et al (1973), Miles (1980), Bullough and Caudrey (1980), and Dodd et al (1982).

Before we address new issues and explore new directions for future research, it is useful to give a brief summary of the basic properties of solitary waves and solitons as observed experimentally and/or found from the KdV model.

1. A unidirectional free solitary wave moving in a uniform medium (e.g., water of uniform depth) maintains a permanent form and a constant phase velocity. Transfer of mass, momentum, and energy remains localized with the wave.
2. The KdV equation (with normalization),

$$\eta_t + \eta_x + \frac{3}{2}\eta\eta_x + \frac{1}{6}\eta_{xxx} = 0, \quad (2.1)$$

has for solitons a one-parameter family of solutions:

$$\eta = a \operatorname{sech}^2 \left\{ \left( \frac{3}{4}a \right)^{1/4} (x - ct - x_0) \right\}, \quad (2.2)$$

$$c^2 = g(h + a), \quad (2.3)$$

where the subscripts denote partial differentiations, and in the expression for the phase velocity  $c$ , the gravity constant  $g$  and the undisturbed water depth  $h$ , both normalized to unity in the KdV equation, have been restored for clarity.

3. For the KdV equation, there are an infinite number of conservation laws of the form (Miura et al, 1968):

$$\mathcal{E}_t + \mathcal{F}_x = 0 \quad (2.4)$$

where  $\mathcal{E}$  is a density function and  $\mathcal{F}$  the corresponding flux function.

4. Solitons of different amplitudes, and hence of different speeds, pass through one another without any permanent change in form and amplitude, but suffer slight phase shifts as the only traceable mark of having undergone the nonlinear interaction (Lax, 1968).
5. Inverse scattering method (ISM) is a nonlinear transformation that reduces the nonlinear problem to a linear calculation of exact solution of the initial value problem (with initial data of compact support) for the KdV equation.
6. The KdV equation is a completely integrable Hamiltonian system (Zakharov and Faddeev 1972).
7. A free solitary wave in a uniform channel is stable with respect to any moderate disturbance that is admissible to a closed system.

It is of significance to note that the foregoing list of basic properties of solitary waves and solitons of the KdV family, though common to many other theoretical models, pertain primarily, if not exclusively, to a closed system—that is, a system not open to exertion by an external forcing agency and hence free of momentum source or energy source. To foster further development of the central field, I think it will be helpful to classify nonlinear wave problems into two classes:

1. Closed systems—defined as those systems in which waves propagate like free waves with their energy conserved according to a conservation law,

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{F}}{\partial x} = 0, \quad (2.5)$$

where now  $\mathcal{E}$  is explicitly the energy density and  $\mathcal{F}$  is the associated energy flux. (For the higher dimensions,  $\mathcal{F}_x$  is simply replaced by the divergence of an  $n$ -vector  $\mathcal{F}$ ,  $n = 2$  or 3.)

2. Open systems—defined as those systems in which new waves can be continually generated by some extraneous forcing disturbances so that the energy conservation equation becomes inhomogeneous,

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{F}}{\partial x} = \dot{W}, \quad (2.6)$$

where  $\dot{W}$  is the rate of working on the system by a specific external forcing function.

An interesting special case of this general relationship is the process characterized by a steady state of energy balance between the rate of energy release from an external source and the rate of energy consumption by the system. A good example is given by the problem of a burning candle. Diffusion of heat down the candle maintains the wax supply to the flame, where it burns to provide the heat in turn (see Scott 1977), giving the energy balance

$$P = uE, \quad (2.7)$$

where  $P$  is the power required for maintaining the flame,  $E$  is the chemical energy stored in unit length of the candle, and  $u$  is the velocity at which the flame propagates as a nonlinear wave.

Just as in the class of closed systems, where nonlinear waves have exhibited interesting, even surprising features, the class of nonlinear waves in open systems can be expected to supply us with no less opportunity to find stimulating and intriguing phenomena that will require new understanding of the underlying physical mechanisms. For instance, one may even wonder if a steady disturbance imposed on an initially unperturbed, nonlinear open system must necessarily result, under all circumstances, in asymptotically steady responses. In other words, will all the transient motions that are generated

after an initial “turn-on” of the disturbance always fall off asymptotically in time? Before we return to these questions, we shall next describe a soliton phenomenon related to bioenergetics, which may be expected to emerge as a new frontier common to both applied mechanics and biomechanics.

### 3. Solitary Waves in Bioenergetics

In the 1973 Seminar of the U.S. Academy of Science (see Green 1973), questions were raised concerning (1) whether there exists a crisis in modern bioenergetics, and (2) if so, how to avoid it. The basis of the discussion can be interpreted to include the following contention.

In all biological phenomena (except for photosynthesis), the primary energy source comes from the hydrolysis of adenosine triphosphate (ATP) molecules. This energy unit is about 0.5 eV, which is only 20 times the thermal energy at physiological temperatures and is too weak to excite the electronic states of molecules. This argument has been used to stress the importance of vibrational excitation of separate groups of atoms in protein molecules, such as the amide-1 vibration, or the valence vibration of the C=O double bond in peptide groups since it requires an energy of only about 0.21 eV. The electric dipole moment of these vibrations is found to be large enough to excite resonance, due to dipole-dipole interactions, of a protein molecule. Such vibrational excitation has been investigated by Davydov (1971, 1979), Davydov and Kislukh (1976), Hyman et al (1981), Scott (1981), and Adey (1983), which we recapitulate in the following.

For one-dimensional molecular systems, vibration of the C=O double bond in peptide groups of a protein molecule occurs in the presence of light absorption and luminescence (the optical mode). Photons will propagate through the system from one group to another; the quantum dynamics of this propagation involves dispersive effects. In addition, on much longer spatiotemporal scales, longitudinal elastic waves (the acoustic mode) will propagate down the molecular chain with phase velocity  $C_p$ . Davydov's nonlinear analysis shows that the optical mode (that is, the internal molecular excitations that he called excitons) and the acoustic mode interact with each other with a characteristic energy. The resulting excitation propagates down the protein molecule as a localized and dynamically stable wave, which he called *soliton* and which is characterized by the Hamiltonian composition:

$$H_{sol} = H_{ex} + H_{ph} + H_{int} \quad (2.8)$$

(see Davydov 1979). The property of the exciton component  $H_{ex}$  is described for the probability wave function by the plane wave

$$\psi_0(x, t) = A \exp[i(kx - \Omega(k)t)], \quad (2.9)$$

where  $k$  is a wave vector directed along the molecular chain, such that  $kl \ll 1$ ,  $l$  being the equilibrium distance between the molecules, and  $\Omega(k)$  the frequency

of excitation. This plane wave has the energy

$$\hbar\Omega(k) = E(0) + \hbar^2 k^2 / 2m_{ex}, \quad (2.10)$$

where  $\hbar$  is Planck's constant,  $m_{ex} = \hbar^2 / 2Jl^2$  is the effective mass of excitation corresponding to energy  $J$  of the resonance interaction between molecules and  $E(0)$  is the internal energy. The energy–frequency formula (2.10) is in fact also the dispersion relationship, which relates the frequency,  $\Omega$ , of each mode to its proper wavelength ( $2\pi/k$ ). A specific excitation is a localized excitation, say initially distributed over a segment  $l_0$  of the chain, and can be described by the wave packet

$$\psi_{ex}(x, t) = \int_{k_0 - \Delta k}^{k_0 + \Delta k} A(k) e^{i(kx - \Omega(k)t)} dk, \quad (2.11)$$

where  $\Delta k = \pi / 2l_0$ . This exciton will propagate along the chain with the group velocity

$$C_g(k_0) = \frac{\partial \Omega(k_0)}{\partial k_0} = \frac{\hbar k_0}{m_{ex}}, \quad (2.12)$$

with its segmental length increasing in time  $\tau$  from  $l_0$  to  $l_\tau$ , where

$$l_\tau = l_0 + (\pi \hbar \tau / l_0 m_{ex}). \quad (2.13)$$

Localized excitons are therefore dispersive since the energy of exciton of shorter wavelengths (larger  $k$ ) will travel faster than that of longer waves. Further, it has been found that the interactions between the exciton mode and the acoustic mode results in a deceleration of the excitons due to phonon production (Cherenkov radiation), and these effects are nonlinear. Thus, it is the balanced interplay of the dispersive and nonlinear effects that constitutes the key mechanism for generating solitary waves (solitons) in molecular systems.

The resulting soliton has been analyzed by Davydov and co-workers (1971–1979) as a Hamiltonian system of quantized lattice. In the limit of a continuum approximation, Davydov derived the following chain equations:

$$\left[ i\hbar \frac{\partial}{\partial t} - (\Lambda - 2\chi\rho) + Jl_0^2 \frac{\partial^2}{\partial x^2} \right] a(x, t) = 0, \quad (2.14)$$

$$\left( \frac{\partial^2}{\partial t^2} - C_p^2 \frac{\partial^2}{\partial x^2} \right) \rho + \frac{2\chi l_0^2}{M} \frac{\partial^2}{\partial x^2} |a|^2 = 0, \quad (2.15)$$

where  $a(x, t)$  is the continuum limit of the (time-dependent) probability wave function amplitude of the soliton,  $\rho(x, t) = -\partial \bar{u}(x, t) / \partial x$ ,  $\bar{u}(x, t)$  is the continuum limit of the mean value of displacements of molecules from their respective equilibrium positions. Further,  $\Lambda$  is a characteristic energy involving the excitation energy, chain deformation energy, and resonance interaction energy;  $\chi$  is an exciton–phonon interaction coefficient;  $l_0$  is the lattice spacing;  $M$  is the mass of a molecule; and  $C_p$  is the longitudinal sound speed in the chain,  $C_p = l_0 \sqrt{K/M}$ , with  $K$  denoting spring constant.

Equation (2.14) is a nonlinear Schrödinger equation for  $a(x, t)$ , and (2.15) is a wave equation for  $\rho$ , with a forcing term that is quadratic in  $a(x, t)$ . They admit the following exact solution:

$$a(x, t) = \sqrt{\frac{\mu}{2}} \exp \left\{ i \left[ \frac{\hbar V}{2Jl^2}(x - x_0) - \frac{E_v t}{\hbar} \right] \right\} \operatorname{sech} [\mu(x - x_0 - Vt)], \quad (2.16)$$

$$\rho(x, t) = \frac{2\chi}{(1 - s^2)K} |a|^2 = \frac{\mu\chi}{(1 - s^2)K} \operatorname{sech}^2 \mu(x - x_0 - Vt), \quad (2.17)$$

where

$$\mu = \chi^2 / [(1 - s^2)KJl^2], \quad s = V/C_p \quad (0 < s < 1). \quad (2.18)$$

This distribution satisfies the normalization that the integral of  $|a|^2$  from  $x = -\infty$  to  $+\infty$  is unity.

Thus, for all values of the wave energy, the resulting soliton has the  $\operatorname{sech}^2 \mu(x - Vt)$  distribution for its probability wave function and travels always at a subcritical speed ( $0 < V/C_p < 1$ ) with respect to the longitudinal sound velocity. For a soliton traveling at a high subcritical speed ( $0 < (1 - s^2) \ll 1$ ) along a soft chain (small  $K$ ), the intramolecular excitation associated with molecular displacements can be very strong, hence to become highly localized (large  $\mu$ ) and to be able to carry a great soliton energy.

These pioneering studies have cast light on some important roles that can be played by solitary waves in transporting biochemical energy through a biological system. At the molecular level, for instance, solitons may be able to transfer along an  $\alpha$ -helical protein molecule the energy of valence vibration C=O of the peptide groups without exciting the thermal motion, hence with little dissipation. The same concept is being applied by Davydov (1979) to explain the contraction mechanism of animal muscles at the molecular level, and by Scott and Luzadar (1979), Adey (1983), and others to study nonlinear electrodynamics in cell membrane.

#### 4. Runaway Solitons

We now proceed to extend our scope from the closed system to an open one and to concentrate our investigation on the new features of generation and evolution of long waves by a continuous forcing excitation (an external agency). To fix this idea, we consider the generation and propagation of three-dimensional long gravity waves of finite amplitude that can propagate in two horizontal dimensions  $\mathbf{r} = (x, y)$  in a layer of water whose initial free surface, when unperturbed at time  $t = 0$ , is at  $z = 0$  and whose floor is at a prescribed depth  $z = -h(\mathbf{r}, t)$ . The forcing functions responsible for generating waves will include the free-surface pressure distribution  $p_0(\mathbf{r}, t)$  acting over the displaced water surface at  $z = \zeta(\mathbf{r}, t)$  and the unsteady movement of the water floor given by  $z = -h(\mathbf{r}, t)$ . The surface pressure may be used to represent disturbances of

meterological and naval architectural nature, such as in applications to air-cushion vehicle and ships, while the floor movement can simulate tsunamiogenic disturbances of the ocean floor. In our exposition, these forcing functions will be confined to such a type that the resulting waves will have lengths,  $\lambda$ , primarily large compared with the characteristic water depth,  $h_0$ , and will have typical amplitude,  $a$ , small relative to  $h_0$ , i.e.,

$$\alpha = a/h_0 \ll 1, \quad \varepsilon = h_0/\lambda \ll 1. \quad (2.19)$$

(The condition of  $\varepsilon \ll 1$  holds for long waves by definition.) Further, we shall concentrate our attention to the case when the magnitude of  $\alpha$  relative to  $\varepsilon^2$ , known as the Ursell number, is of order unity,

$$Ur = \alpha/\varepsilon^2 = a\lambda^2/h_0^3 = O(1). \quad (2.20)$$

The gravity waves satisfying conditions (2.19) and (2.20) are said to belong to the “Boussinesq class.” It signifies the case in which the nonlinear effects causing the waves to steepen forward, at a rate proportional to  $\alpha$ , are comparable in importance with the opposite tendency, due to the dispersive effects by which waves of longer lengths travel at greater velocities by a margin proportional to  $\varepsilon^2$ .

For the theoretical model capable of representing appropriately the long waves generated in shallow water by this general class of forcing functions, we adopt the one recently developed by Wu (1979, 1981—his equations [41] and [42]):

$$\zeta_t + \nabla \cdot [(h + \zeta) \nabla \bar{\phi}] = -h_t + \nabla \cdot \left\{ \left[ \frac{h}{2} (h_t + \nabla \cdot (h \nabla \bar{\phi})) - \frac{h^2}{3} \nabla^2 \bar{\phi} \right] \nabla h \right\}, \quad (2.21)$$

$$(\bar{\phi})_t + \frac{1}{2} (\nabla \bar{\phi})^2 + g\zeta + \frac{1}{\rho} p_0 = \frac{h}{2} \frac{\partial}{\partial t} [h_t + \nabla \cdot (h \nabla \bar{\phi})] - \frac{h^2}{6} \nabla^2 \bar{\phi}_t. \quad (2.22)$$

Here,  $\nabla$  represents the two-dimensional vector operator  $\nabla \equiv \partial/\partial \mathbf{r} = (\partial/\partial x, \partial/\partial y)$  with respect to the position vector  $\mathbf{r} = (x, y)$  in the horizontal plane, and  $\bar{\phi}$  denotes the layer-mean value of the original velocity potential  $\phi(x, y, z, t) = \phi(\mathbf{r}, z, t)$  as defined by

$$\bar{\phi}(\mathbf{r}, t) = \frac{1}{\eta} \int_{-h}^{\zeta} \phi(\mathbf{r}, z, t) dz \quad (\eta = h + \zeta). \quad (2.23)$$

The velocity potential  $\phi$  and its layer mean  $\bar{\phi}$  are related by

$$\phi - \bar{\phi} = -(z + \frac{1}{2}h)[h_t + \nabla \cdot (h \nabla \bar{\phi})] - \frac{1}{2}(z^2 - \frac{1}{3}h^2) \nabla^2 \bar{\phi}. \quad (2.24)$$

From this relation, one can readily deduce the velocity distribution as  $(\varphi_x, \varphi_y, \varphi_z)$  and the pressure field from the Bernoulli equation,

$$\frac{1}{\rho} p = -gz - \phi_t - \frac{1}{2}(\varphi_x^2 + \varphi_y^2 + \varphi_z^2). \quad (2.25)$$

In the sequel, the fluid density,  $\rho$ , and the gravitational constant,  $g$ , are normalized to unity and reinstated whenever needed for clarification.

The above set of basic equations (2.21) and (2.22) may be regarded as forming generalized Boussinesq class, in view of the new features that the medium is now inhomogeneous (due to the spatial and temporal variation of  $h$ ), there are external forcing functions, and there is the added dimension of wave propagation. In the special case of homogeneous motion, that is with  $h = \text{constant}$  and  $p_0 = 0$ , Equations (2.21) and (2.22) reduce to the classic Boussinesq equations, which admit solitons (either right- or left-running) as exact solutions. For the general case, however, it is possible to obtain solutions of Equations (2.21) and (2.22) by numerical computation.

A numerical method has been developed by Wu and Wu (1982) and subsequently improved (for increased accuracy) by Lee (1985) for the case of two-dimensional initial boundary problems. In this case, Equations (2.21) and (2.22) are solved by employing the modified Euler's predictor-corrector method in advancing time and the central difference approximation for the spatial derivatives. A mixed implicit-explicit scheme is adopted for the forward difference computation of  $\zeta$  and  $\bar{\phi}$ , the implicit part being incorporated in order to achieve the desired numerical stability and accuracy with relatively large time step  $\Delta t$ . A simple but highly effective open boundary condition for the downstream and upstream boundaries has proven successful, and as a result, the region of computation can be taken as relatively small. We present below some typical numerical results obtained by this method.

An interesting example is given by the following initial value problem of a traveling surface pressure with the distribution

$$p_0 = p_{om} \frac{1}{2} \left[ 1 - \cos \left( 2\pi \frac{x + Ut}{L} \right) \right] \quad (0 < x + Ut < L, t > 0) \quad (2.26)$$

with

$$p_{om} = 0.1, \quad h = 1, \quad L = 2, \quad F_h = U/\sqrt{gh} = U = 1.0, \quad (2.27)$$

and  $p_{om} = 0$  elsewhere for  $t > 0$  as well as for all  $x, t < 0$ . And at  $t = 0$ , the water was at rest. The transient motion resulting from the application of this  $p_0$  moving at the critical speed was computed by using the present numerical method over the region  $-30 < x + Ut < 40$ , with  $\Delta x = 0.2$  and  $\Delta t = 0.2$ . The numerical result for the free surface elevation  $\zeta$  is shown in Figure 2.1, here plotted with reference to the moving pressure disturbance, marked by two vertical lines for its extent, over the time period (normalized vs.  $L/\sqrt{gh}$ ) in  $0 < t < 80$ .

A new feature of particular interest is that, after the surface pressure is exerted and kept moving at the critical speed for a sufficiently long time, a solitary wave emerges just ahead of the disturbance, and eventually surges away from the disturbance to propagate ahead as a free solitary wave, an entity we call a "runaway soliton." This is followed by another new solitary

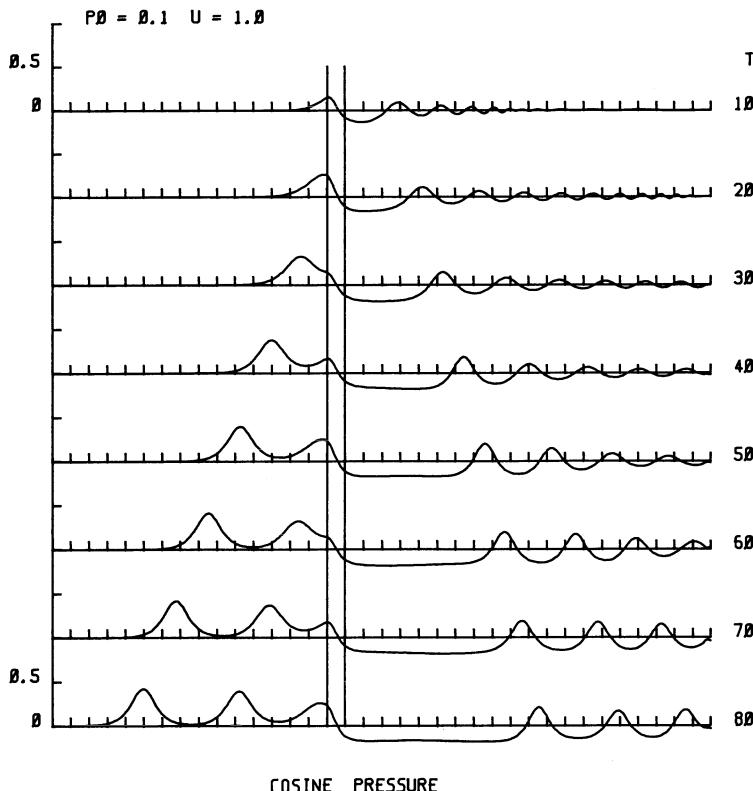


FIGURE 2.1. The free solitary waves generated ahead and cnoidal-like waves produced behind the free-surface pressure distribution specified by Equations (2.26) and (2.27).

wave that is being generated and subsequently “born”; this process seems to continue almost periodically and indefinitely. There are nearly three solitons having run away at the end of this computation period. The increasing phase velocity of the runaway solitons and the decreasing phase velocity of the trailing cnoidal-like waves with increasing time both result from the nonlinear and dispersive effects, as can be reflected by the ever-prolonging region of depression (in water depth) immediately behind the disturbance. We point out here that the rate of growth in magnitude of the runaway solitons is in this case very slight, much weaker than shown in the earlier calculation by Wu and Wu (1982, figure 11); this difference can be attributed to the improved accuracy of the present numerical code over its previous version.

In Figure 2.2 we present the result for  $\zeta$  generated by a surface pressure similar to Equation (2.26), except now with

$$p_0 = 0.15, \quad h = 1, \quad L = 2, \quad U = 1.0. \quad (2.28)$$

Associated with this stronger disturbance, the runaway solitons and the

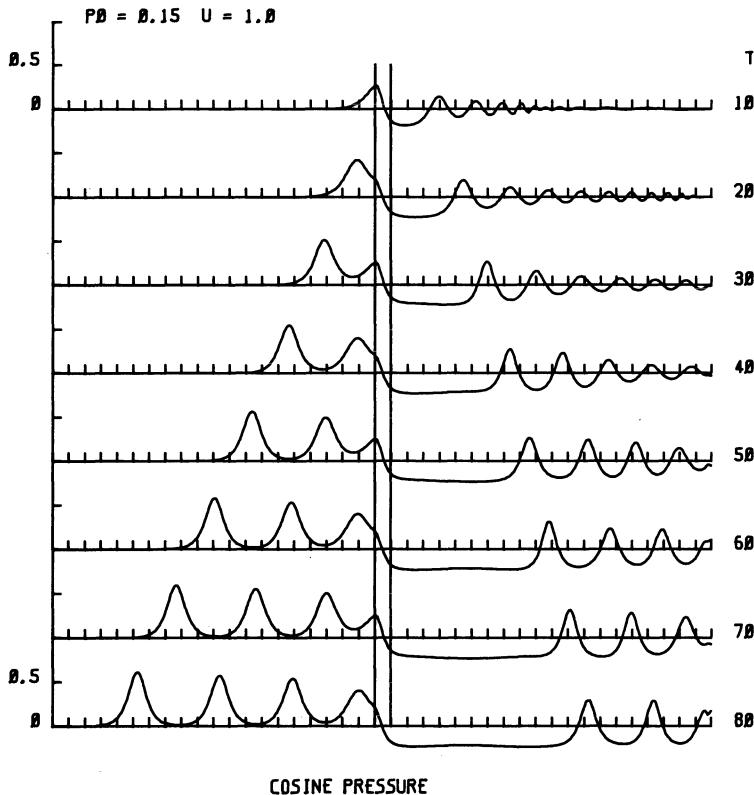


FIGURE 2.2. The free solitary waves generated ahead and cnoidal-like waves produced behind the free-surface pressure distribution specified by Equations (2.26) and (2.28).

trailing wave train both become stronger, and the period for soliton generation becomes shorter, as a direct comparison with Figure 2.1 shows.

The wave resistance,  $D_w$ , experienced by the surface pressure (per unit width) has the coefficient

$$C_{Dw} = D_w / \rho g h L = -\frac{1}{\rho g h L} \int_0^L p_0(x, t) \frac{\partial \zeta}{\partial x} dx. \quad (2.29)$$

And the coefficient of rate of working by  $p_0$  is

$$C_{(\dot{W}/U)} = \frac{\dot{W}}{\rho g h L U} = -\frac{1}{\rho g h L U} \int_0^L p_0(x, t) \frac{\partial \zeta}{\partial t} dx. \quad (2.30)$$

The result of  $C_{Dw}$  and  $C_{(\dot{W}/U)}$  corresponding to Equations (2.26) and (2.27) is given in Figure 2.3 and that for Equations (2.26) and (2.28), in Figure 2.4, in which the coefficient  $C_{(\dot{W}'/U)} = C_{(\dot{W}/U)} - C_D$  is also shown. The wave resistance is thus seen to vary considerably during the course when each soliton is generated and runaway. If the period of  $C_{Dw}$  variations is used to indicate the "gestation period" of soliton generation, we find that this provides very

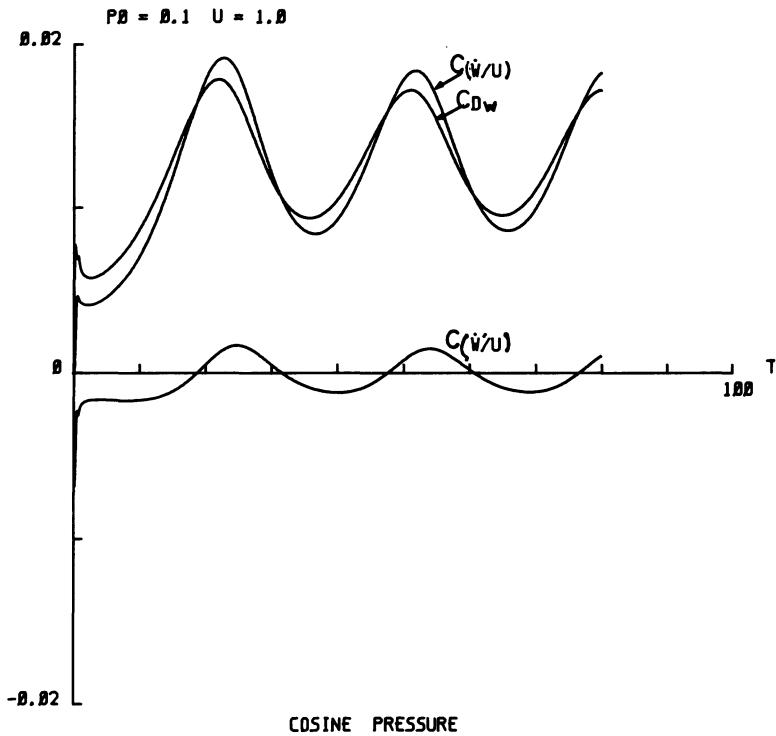


FIGURE 2.3. The wave resistance coefficient  $C_{Dw}$  corresponding to the motion given in Figure 2.1.  $C_{(W'/U)}$  is the coefficient of rate of working by the surface pressure  $p_0$ ;  $C_{(W'/U)}$  is the difference between  $C_{(W'/U)}$  and  $C_{Dw}$ .

uniform readings of gestation period in succession, which is  $T = 30$  for  $p_{om} = 0.1$  and is reduced to  $T = 18$  when  $p_{om}$  is increased to 0.15, while the drag coefficient is overall higher for greater  $p_{om}$ .

This phenomenon of “new birth to runaway solitons” is found to manifest over a range of the depth Froude number,  $F_h = U/\sqrt{gh}$ , from as low as  $F_h = 0.2$  to about  $F_h = 1.2$ , depending on the form and extent of  $p_0$  distribution. At low subcritical speeds ( $0.3 < F_h < 0.7$ ), the solitons generated are relatively weak, but definitely discernible both numerically and experimentally, while the period of generation is relatively short. The periodically generated solitons gain in magnitude with increasing  $F_h$ , whereas the trailing wave train becomes gradually weaker and the period of generation becomes somewhat more prolonged, especially when  $F_h$  increases beyond 1. As  $F_h$  reaches a certain value of about 1.2, the leading soliton was found experimentally to break into a turbulent bore and numerically to evanesce for still higher values of  $h$ .

Generation of solitons has also been investigated for the case when the forcing disturbance is provided by a curved bump moving along the bottom of a water layer. It has been found, both numerically and experimentally by Lee (1985) that the free solitary waves generated ahead and the cnoidal-like waves

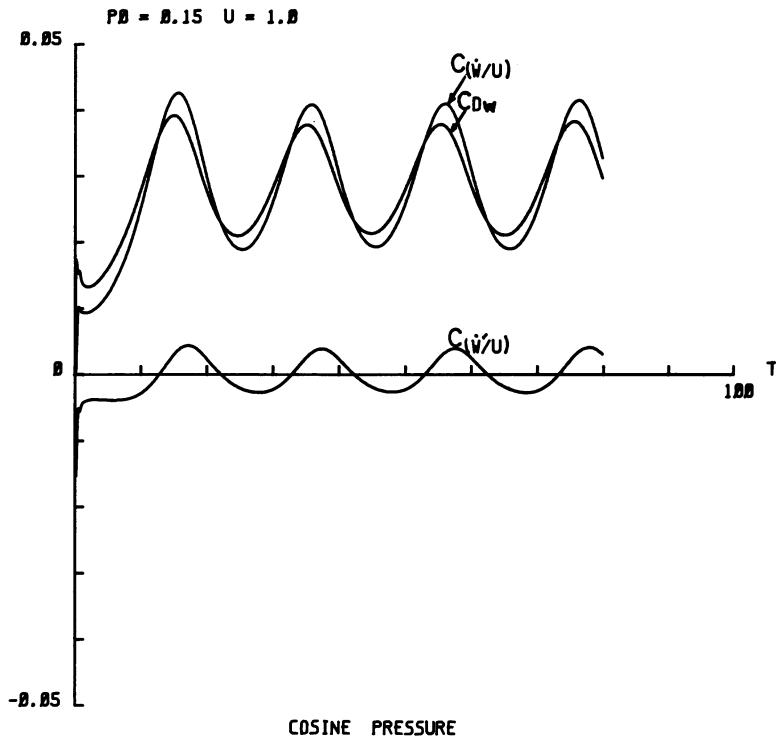


FIGURE 2.4. The wave resistance coefficient  $C_{Dw}$  corresponding to the motion given in Figure 2.2.

trailing behind the moving bump closely resemble those produced by a surface pressure moving at the same transcritical speed. This problem has also been investigated by Ertekin (1984).

In regard to excitations of higher dimensions, we mention the interesting results of the towing tank experiment with ship models in very shallow water reported by Huang et al (1982a, b). Within a transcritical range of the depth Froude number, they found that approximately two-dimensional solitary waves spanning across the tank (8 feet wide) were generated, one after another, to move down the tank ahead of the model. The longer the run, the more of such waves appear. We thus note that, while the forcing agencies are different in these two cases, the resulting solitary waves nevertheless share the same main features.

## 5. Forward Outlook

Having experienced the newly discovered phenomenon just described, we may wonder what impact it will exert on various other soliton-bearing phenomena and what will be its ramifications. It should not be difficult to attain a scientific

view on such prospects if we draw out the physical quantities playing the analogous roles in these different subjects and note that the generation of solitary waves in shallow water by external forcing excitations would not be possible without a well-balanced interplay by both the nonlinear and dispersive effects. This point is now qualitatively clear, albeit the criteria of its manifestation is not yet precise.

On the other hand, the very recent developments in bioenergetics and neurophysics have contributed valuable data to indicate that nonlinear and dispersive waves like solitons can play some essential roles leading to understanding of biochemical processes in physical terms. It is a most brilliant and crucial idea, I believe, to usher in biodynamics and to assign it a leading role in the play directed to investigate the activities of biological molecules. For its future development, we may serve well to coin the term *molecular biodynamics*. There cannot be the slightest doubt that, along the frontiers bordering biology, physics, and engineering lie some most exciting and rewarding problems. Success in working along those frontiers may ultimately lead to an understanding of essential properties and mechanisms of living matter.

Along this new frontier, major advances in such new directions as exemplified by the present study can be richly benefited by our mentor and friend, Professor Y.C. Fung, to whom we owe our warmest gratitude for his enlightening teaching, voluminous writings, inspiring encouragement, and above all, friendship. And, of course, we continue looking to him for advice and leadership.

## Acknowledgments

I take great pleasure in expressing my appreciation to Professor Geert Schmid-Schönbein for his assuring encouragement, to Dr. George Yates for having interesting discussions on various interrelating subjects, and to S.J. Lee for letting me quote his research results.

Studies from this research group are supported jointly by ONR Contract N00014-82-K-0443, NR 062-737, and NSF Grant MEA-8118429. I wish to thank Dr. Choung M. Lee and Dr. George K. Lea for their scientific interest in this general subject.

## References

- Adey, W.R., ed. 1983. Nonlinear, nonequilibrium aspects of electromagnetic field interactions at cell membranes. Proc. Intern. Conf. on Nonlinear Electrodynamics in Biological Systems. Held at Loma Linda, California, 6–9 June 1983. Plenum Press, New York.
- Boussinesq, M.J. 1871. *Acad. Sci. Paris, Comptes Rendus* 73:256–60.
- . 1872. *J. Math. Pures Appl.* 17(2):55–108.
- . 1877. Mém. présentés par divers Savants à L'Acad. Sci. Inst. France 23:1–680; 24:1–64.

- Bullough, R.K., and Caudrey, P.J. eds. 1980. *Topics Curr. Phys.* 17. Springer Verlag, Berlin.
- Davydov, A.S. 1971. *Theory of molecular excitons*. Plenum Press, New York.
- . 1979. *Physica Scripta* 20: 387–394.
- Davydov, A.S., and Kislyukha, N.I. 1976. *Zh. Eksp. Teor. Fiz.* 71: 2090; *Sov. Phys. JETP* 44: 571.
- Dodd, R.K., Eilbeck, J.C., Gibbon, J.D., and Morris, H.C. 1982. *Solitons and nonlinear wave equations*, Academic Press, New York.
- Ertekin, R.C. 1984. *Soliton generation by moving disturbances in shallow water: theory, computation and experiment*. Ph.D. dissertation, University of California, Berkeley.
- Green, D.E. 1973. *Science* 181: 583.
- Huang, De-Bo, Sibul, O.J., and Wehausen, J.V. 1982a. *Ships in very shallow water*. Festkolloquium—Dedication to Professor Karl Wieghardt. March 1982, Institut für Schiffbau der Universität Hamburg.
- Huang, De-Bo, Sibul, O.J., Webster, W.C., Wehausen, J.V., Wu, De-Ming, and Wu, T.Y. 1982b. Proc. Conf. on Behavior of Ships in Restricted Waters, vol. 2, 26.1–26.12. Held at Varna, Bulgaria, 11–13 Nov. 1982.
- Hyman, J.M., McLaughlin, D.W., and Scott, A.C. 1981. *Physica D* 30: 23.
- Korteweg, D.J., and deVries, G. 1895. *Phil. Mag.* 39: 422–443.
- Lax, P.D. 1968. *Comm. Pure Appl. Math.* 21, 467–490.
- Lee, S.J. 1985. Generation of long waterwaves by moving disturbances. Ph.D. Thesis. California Institute of Technology, Pasadena.
- Miles, J.W. 1980. *Ann. Rev. Fluid Mech.* 12: 11–43, Annual Reviews Inc., Palo Alto, California.
- Miura, R.M., Gardner, C.S., and Kruskal, M.D. 1968. *J. Math. Phys.* 9, 1204–1209.
- Rayleigh, Lord. 1876. *Phil. Mag.* 1: 257–279. *Sci. Pap.* 1: 251–271.
- Russell, J.S. 1838. Report of the Committee on Waves. Pages 417–496 in Rep. Meet. Brit. Assoc. Adv. Sci. 7th, Liverpool, 1837. John Murray, London.
- . 1845. Pages 311–390 in Rep. Meet. Brit. Assoc. Adv. Sci. 14th, York, 1844. John Murray, London.
- Scott, A.C. 1981. *Phys. Letters* 86A: 60.
- . 1977. *Neurophysics*. John Wiley & Sons, New York.
- Scott, A.C., Chu, F.Y.F., and McLaughlin, D.W. 1973. *Proc. IEEE* 61: 1443–1483.
- Scott, A.C., and Luzader, S.D. 1979. *Physica Scripta* 20: 395–401.
- Whitham, G.B. 1974. *Linear and nonlinear waves*. John Wiley & Sons, New York.
- Wu, T.Y. 1980. Pages 110–149 in *Tsunamis—Proc. National Science Foundation Workshop*, L.S. Hwang and Y.K. Lee, eds. Tetra Tech, Inc., Pasadena, California.
- . 1981. *Proc. ASCE, J. Eng. Mech. Div.* 107: EM3, 501–522.
- Wu, De-Ming, and Wu, T.Y. 1982. Pages 103–125 in *Proc. 14th Sympos. on Naval Hydrodynamics*. Held at University of Michigan, Ann Arbor, 23–27 August 1982. National Academy Press, Washington, D.C.
- Zabusky, N.J., and Kruskal, M.D. 1965. *Phys. Rev. Lett.* 15: 240–243.
- Zakharov, V.B., and Faddeev, L.D. 1972. *Funct. Anal. Appl.* 5: 280–287.

# 3

## Leukocyte Protopodia

G.W. SCHMID-SCHÖNBEIN, R. SKALAK, and S. CHIEN

### 1. Introduction

One of the important functions of the leukocytes in the circulation is to participate in the body's defense mechanisms against intrusion of foreign organisms or materials. A way to demonstrate this is by means of a local laser heat injury in the rabbit ear chamber (Grant, 1974). After the diffusion of inflammatory substances from the burn towards adjacent blood vessels, leukocytes adhere to the endothelium (Mayrovitz et al, 1980) and emigrate into tissue towards the site of the injury. In the bloodstream leukocytes are normally carried along passively by hydrodynamic forces, whereas their locomotion through the tissue is active. The cells create projections of the cytoplasmic membrane and cytoplasm beyond the perimeter of the undeformed shapes, and then achieve a displacement by attachment to and detachment from the substrate accompanied by relaxation. Figure 3.1 shows a protopodium of a neutrophilic leukocyte projected underneath the endothelial cells. This process is the first step in the emigration of the leukocyte into the tissue.

A key event in cell motility is the formation of protopodia. This is a general phenomenon in biology, as fundamental as muscle contraction, that was well recognized by early investigators (Mast, 1926). It is observed in leukocytes, platelets, endothelial cells (Majno et al, 1969), connective tissue cells, and microorganisms. It is the basis for amoeboid migration, phagocytosis, and many other cell functions.

Several mechanisms have been proposed for protopodia formation. Most investigators conclude that a phase transition from a "gel" to a "sol" or vice versa is involved (DeBruyn, 1946; Taylor and Fechheimer, 1983; Stossel, 1982). But these ideas have not been translated into a continuum mechanics theory that explains the deformation of the cell and the shapes observed. There is a need for a continuum mechanical model that can describe the properties and motion of each part of the cell and synthesize them into descriptions of the deformation and locomotion of the cell as a whole. In the following sections, a particular proposal advanced by Schmid-Schönbein and Skalak (1984) and Skalak et al (1984) is further illustrated and developed.

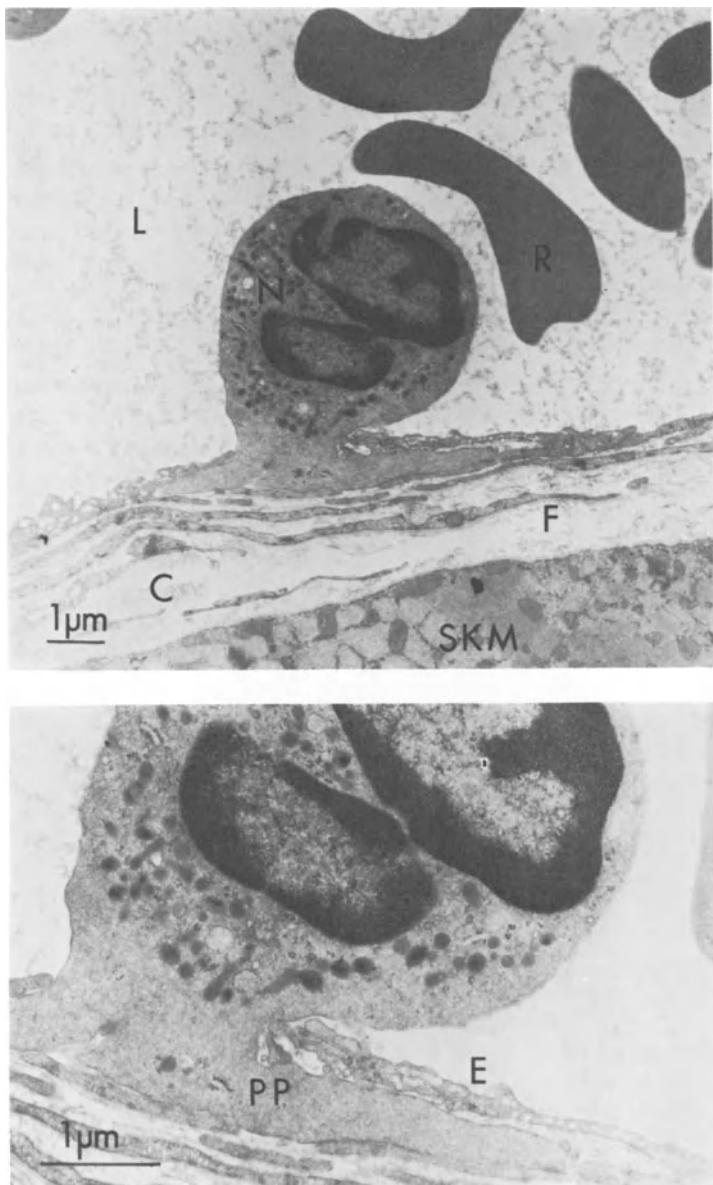


FIGURE 3.1. Transmission electron microscopic image of a neutrophilic leukocyte projecting a protopod between (top) and below (bottom) two endothelial cells in a venule of the rat spinotrapezius muscle. Note that the protopod (PP) does not contain any cell organelles. (L = lumen, R = red cell, N = neutrophil, C = collagen fibers, F = fibroblast cell, SKM = skeletal muscle cell, E = endothelial cell).

## 2. The Cell Membrane

Professor Y.C. Fung was among the first investigators to recognize the importance of the cell membrane to cellular mechanics (Fung, 1981). Biological membranes exhibit a stiff resistance to an increase of their surface area. Quantitatively, this feature is expressed in terms of the areal compressibility modulus  $K$ . For a membrane element with initial unstressed area  $A_0$ ,

$$K = \frac{1}{T_0} \frac{\Delta A}{A_0}, \quad (3.1)$$

and  $\Delta A$  is the change of area under the isotropic membrane tension  $T_0$ . For red cell membrane,  $K$  is of the order of 450 dyn/cm (Evans and Waugh, 1977) which is several orders of magnitude larger than typical shear or Young's modulus of cell membrane. For the leukocyte plasma membrane,  $K$  is about 636 dyn/cm (Schmid-Schönbein, 1984). If the cell membrane is forced to increase its surface area by increasing isotropic tension, it ruptures after 2% or 3% increase in area. This is characteristic of the lipid membranes of all cells. This constraint on the plasma membrane area limits the degree of spreading of leukocytes during active migration, in phagocytosis and also during passive spreading on a blood smear (Schmid-Schönbein et al, 1984). In isotonic media, the leukocytes have an excess membrane above that needed to enclose their cytoplasm in the form of a smooth sphere. The excess area is found in the numerous surface folds (Schmid-Schönbein et al, 1980) and may be 80–100% of the smooth sphere area required to enclose the cell volume. The cells may increase their plasma membrane area by incorporation of membrane from internal organelles, of membrane during endocytosis, or by metabolic production, but this is a relatively slow process.

## 3. The Cytoplasm

The major part of the leukocyte volume is occupied by a cell matrix that includes an extensive actin fiber network. Biochemical analyses of cell matrix extracts show another abundant macromolecule, an actin-binding protein, that crosslinks actin. Other molecules, like gelsolin and acumentin, may be able to influence the actin filament length (Stossel, 1982). Importantly, small amounts of myosin are found primarily in the immediate submembrane cortical space (Rikihisa and Misuno, 1977) and in regions where the cells form protopodia (Stendahl et al, 1980). Leukocyte cell extracts have been reported to gel under *in-vitro* conditions and form a fibrous form of actin gel (*F*-actin) (Kasai et al, 1960; Maruyama et al, 1974; Brotschi et al, 1978). The gelation process requires calcium ( $\text{Ca}^{++}$ ) and is reversible. Protopods in various situations stain heavily for *F*-actin, more strongly than the rest of the cell (Sullivan and Mandell, 1983; Fechheimer and Zigmoid, 1983).

#### 4. Cell Shapes During Protopod Formation

After phlebotomy, most leukocytes have no protopods, indicating that the circulating leukocytes are in the passive state. If a blood sample is left standing for about 1 hour in a test tube, protopods begin to form on the leukocytes spontaneously. Figure 3.2 shows a time sequence of protopod formation as

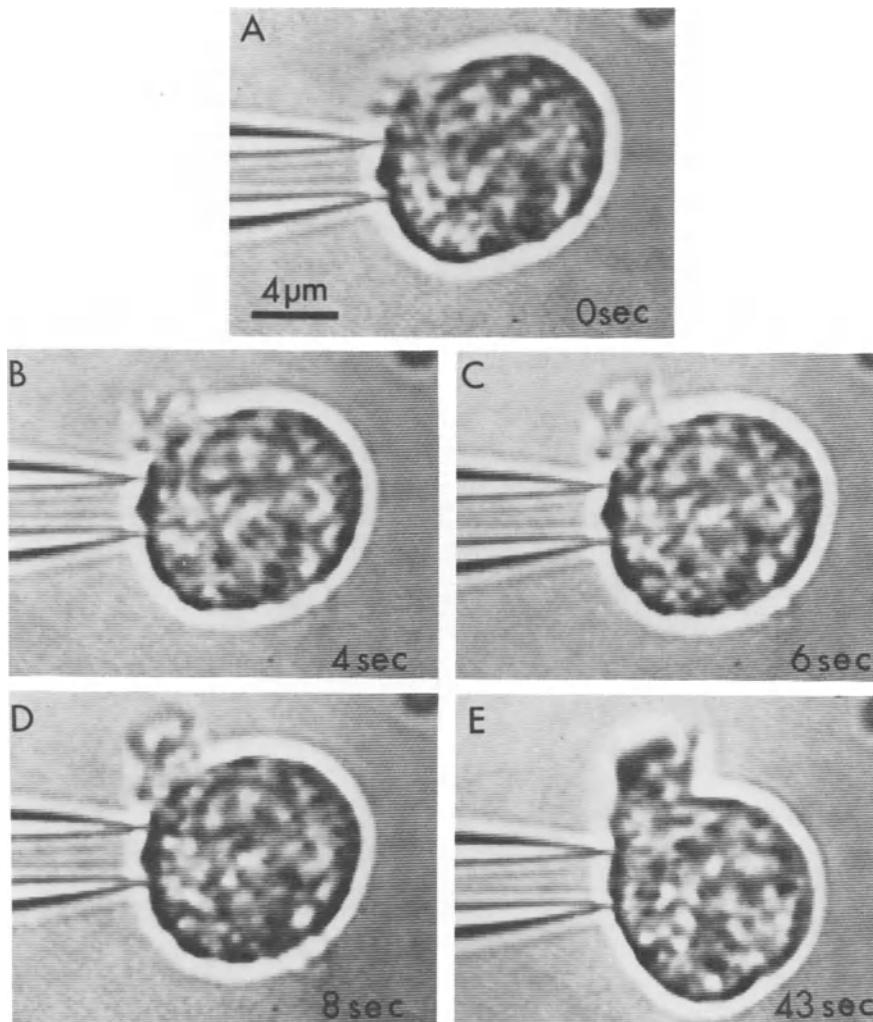


FIGURE 3.2. Sequence of frames photographed from a television screen during the projection of a protopod in human neutrophil. The tip region of the protopod moves as a rigid body outward without any further changes of shape being detectable at this magnification. The cell is in free suspension except that it is prevented from drifting away by gentle aspiration to a holding pipette (for methods, see Schmid-Schönbein et al, 1982).

seen on a video tape of a white blood cell in free suspension. Protopod projection and retraction is a cyclic process (Fig. 3.3) in which formation and retraction occur successively at different points of the cell's surface. The protopodia appear to develop from existing surface folds (Fig. 3.4a, b), and the

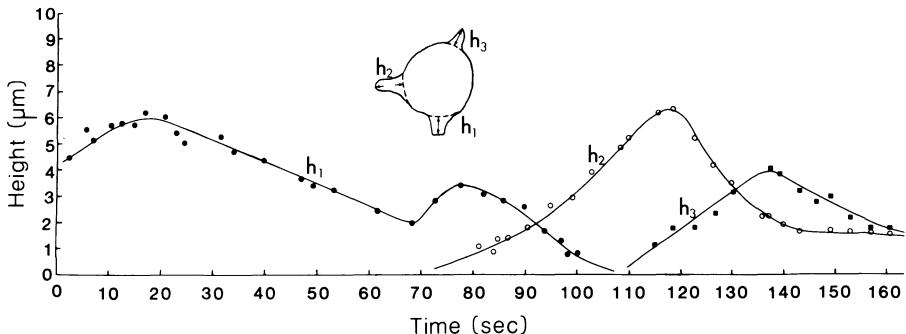


FIGURE 3.3. Time course of protopod projection in human neutrophil. The cell is in free suspension without adhesion to a substrate.

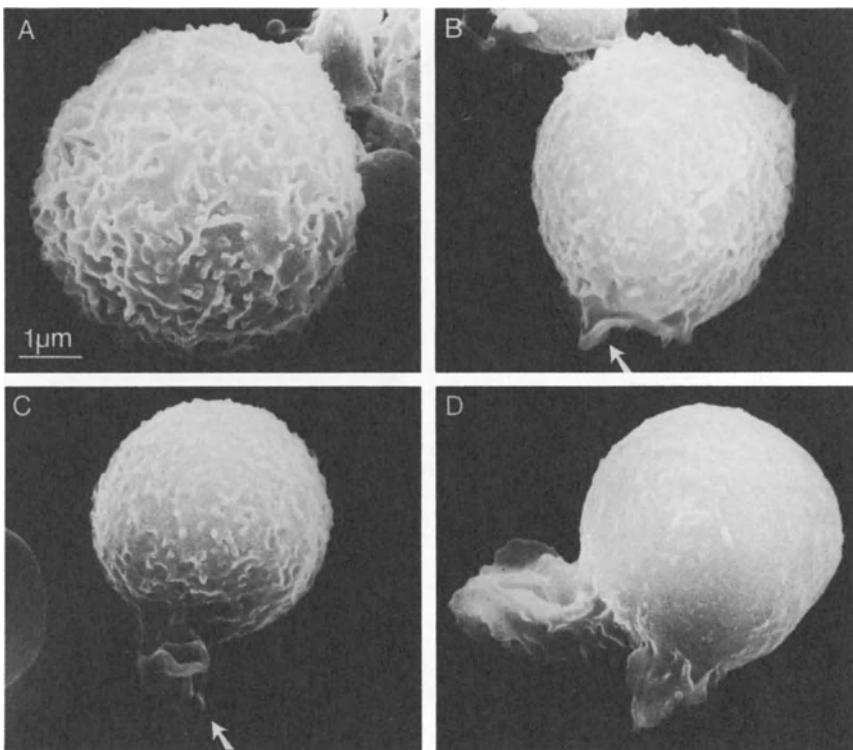


FIGURE 3.4. Scanning electron microscopic images of human neutrophils in homologous plasma. The cells were fixed at an early stage of protopod formation (about one hour after phlebotomy) and prepared for microscopy as described by Schmid-Schönbein et al (1980).

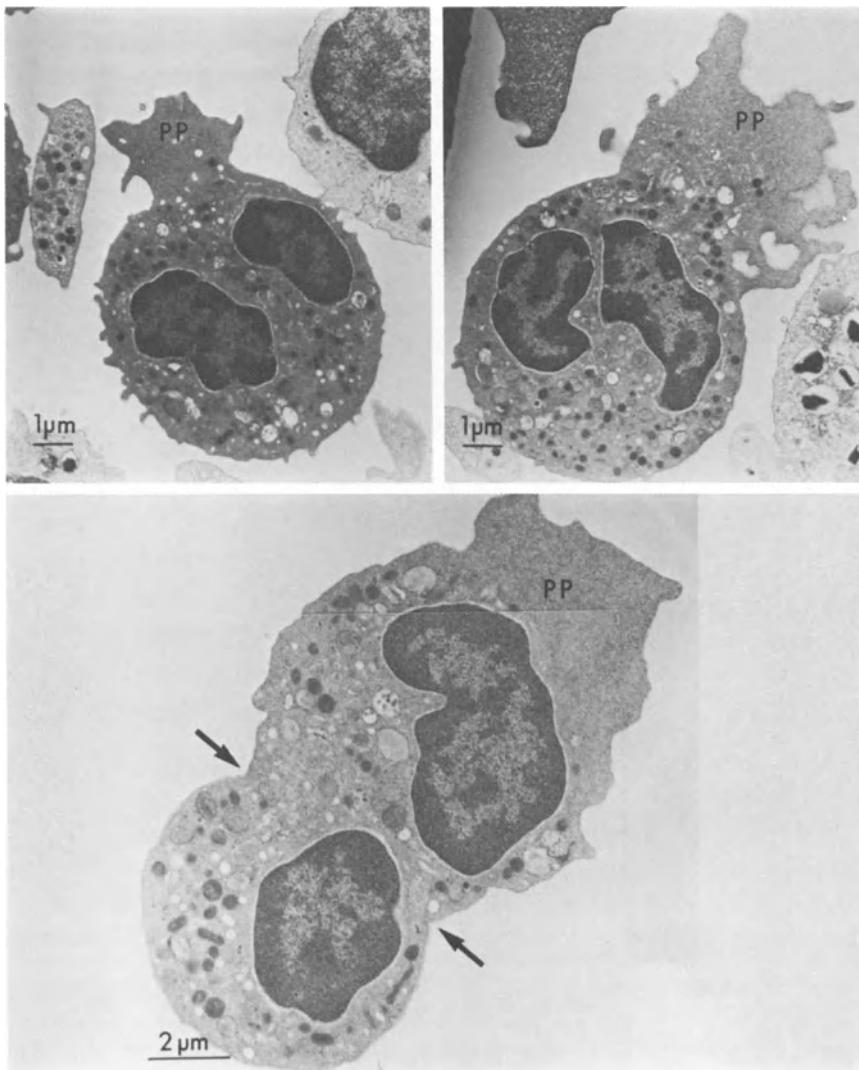


FIGURE 3.5. Transmission electron micrograph of human neutrophilic leukocytes during projection of a protopod (PP).

cell membrane unfolds in the vicinity of a protopod, as seen in Figure 3.4d. Different protopodia may exist in the formation and retraction phases at the same time in a single cell. All protopodia on intact cells are covered by cell membrane, and newly forming protopodia contain no cell organelles (Fig. 3.5). The cytoplasm in protopodia shows a fine fibrous structure oriented in the direction of the projection, which implies a preferred orientation of the actin

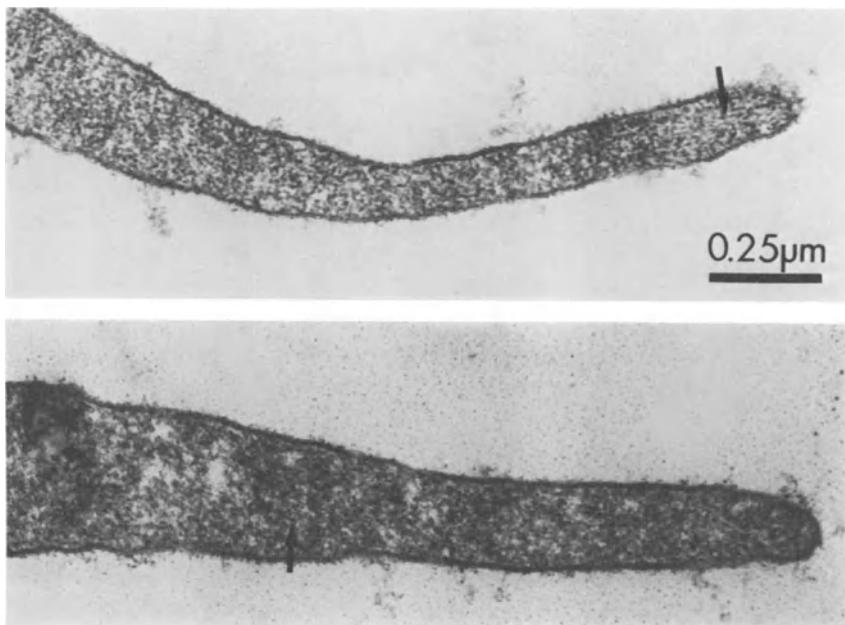


FIGURE 3.6. Transmission electron micrograph of a single protopod with fibrous structure (arrows) parallel to the membrane in the direction of the projection.

molecules in the gelled matrix (Fig. 3.6). The typical protopod growth velocity relative to the centroid of the cell at room temperature ( $22^{\circ}\text{C}$ ) is of the order of  $5 \mu\text{m}/\text{min}$ , indicating the relative slowness of the process in general. Only a few actin molecules need to be assembled into a polymer per second in order to achieve this longitudinal speed. For example, at a typical molecular length of  $650 \text{ \AA}$  only 1.3 monomers per second in each fibre need to assemble to achieve such a protopod speed. Some authors (Stossel, 1982) have proposed that actin exists in the ungelled cell in a larger fragmented form consisting of several monomers. In this case, even lower alignment rates would suffice.

Electron micrographs of cortical cytoplasm after critical point drying show a three-dimensional anisotropic network of fibers (Wolosewick and Porter, 1979), suggesting that besides polymerization, some alignment of the cytoskeleton network may also play an important role for the protopod shape changes. On neutrophils, the shape of the protopod is initially sheet-like, which also implies orientation (Fig. 3.4b, d). As several sheets join together, the base of the protopod widens (Fig. 3.4c, d).

In some experiments on neutrophils, a characteristic neck originates at the base of the protopod and travels along the main cell body (Fig. 3.7) at a speed of about  $5.6 \mu\text{m}/\text{min}$ . In contrast to the protopod, the neck has a more or less axisymmetric shape at all times (Fig. 3.8), and there may be two or more necks

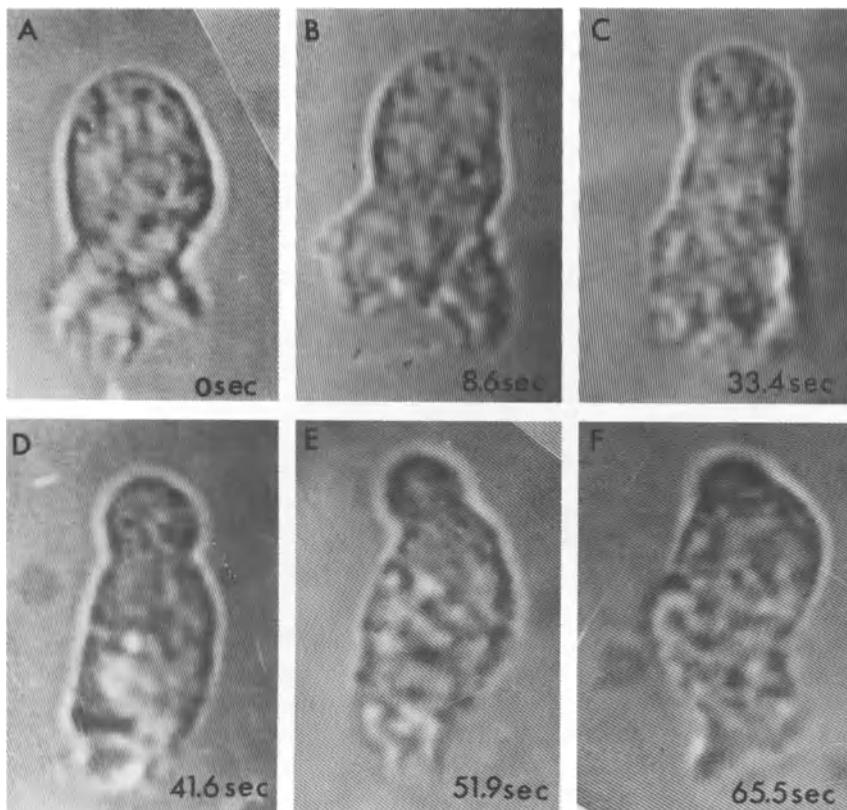


FIGURE 3.7. Time course of neck formation in a freely suspended neutrophil. The neck originates at the root of the protopod and travels with a speed of about  $5.6 \mu\text{m}/\text{min}$  along the cell body. As the neck approaches the larger diameter, a general narrowing of the cell body can be observed (frame C). Note that two necks are present. From Schmid-Schönbein et al (1982) with permission of the publisher.

present at one instant. With standard transmission electron microscopy, no specialized ultrastructural feature is visible in the neck region (Fig. 3.5, arrows). There is no extensive granule exclusion in the neck region, although the closest approach of the granules to the plasma membrane is usually a distance about equal to their own radius. Necks have been observed on cells migrating in the tissue (Mast, 1926), and they may aid the cell in slipping through the endothelial junctions (Fig. 3.1).

Protopod formation continues in most white blood cells in free suspension for many hours. Over a period of time, a cell may develop more than one protopod and may not recover to its undeformed spherical configuration after severe distortion by protopod formation, resulting in highly irregular cell shapes.

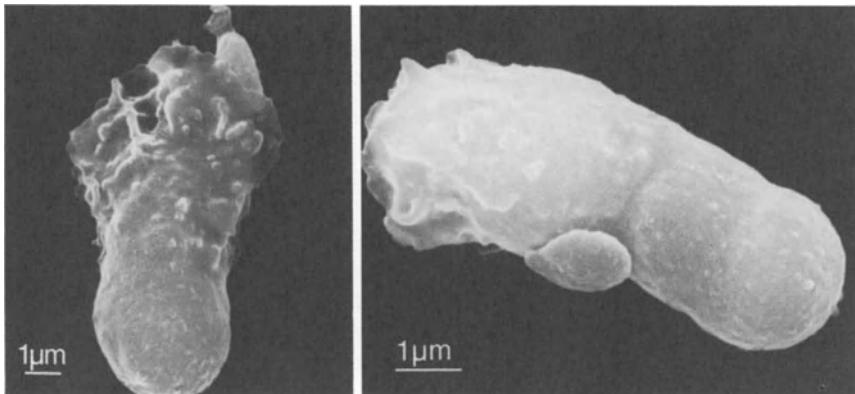


FIGURE 3.8. Scanning electron micrographs of human neutrophil during neck formation. The schematic (Fig. 3.10) shows the position of the polymerization interface.

## 5. Rheological Properties of Passive Leukocytes

The viscoelasticity of leukocytes may be conveniently demonstrated by means of the micropipette aspiration experiment. The unpolymerized cytoplasm exhibits, for small strains, properties that may be described by a three-element standard solid body (Schmid-Schönbein et al, 1981):

$$\sigma'_{ij} + \frac{\mu}{k_2} \dot{\sigma}'_{ij} = k_1 e'_{ij} + \mu \left( 1 + \frac{k_1}{k_2} \right) \dot{e}'_{ij}. \quad (3.2)$$

$\sigma'_{ij}$  is the Cauchy stress deviator;  $\dot{\sigma}'_{ij} = \partial \sigma'_{ij} / \partial t$  is the derivative; and  $e'_{ij}$  is the strain deviator according to Fung's notation (Fung, 1977). In the absence of any flux across the membrane, the cytoplasm is incompressible and hence

$$e_{kk} = e_{11} + e_{22} + e_{33} = 0, \quad (3.3)$$

where  $e_{ij}$  is the strain tensor. Then  $e'_{ij} = e_{ij} - \delta_{ij} e_{kk} = e_{ij}$ . In Equation (3.2)  $k_1$ ,  $k_2$ , and  $\mu$  are material coefficients and are functions of temperature, pH, osmolarity of suspending medium, and other biochemical factors (Chien et al, 1984). The white cells in the passive state can undergo large strains, requiring an extension of Equation (3.2) to a large strain formulation and recognition of the influence of the nucleus and other internal structures.

## 6. Rheological Properties of Active Leukocytes

Micropipette aspiration experiments have indicated that cells with protopods exhibit heterogeneous rheological properties. Aspiration of the main cell body away from the protopod with a step pressure leads to the usual viscoelastic

creep predicted by Equation (3.2) for small strains. On the other hand, if portions of the protopod are aspirated, a much smaller deformation without creep is observed at the same aspiration pressure (Schmid-Schönbein et al, 1982). Once the protopod melts, the usual viscoelastic creep of passive cells is observed. The gelation and stiffening is compatible with the exclusion of granules in the protopods that cannot readily diffuse into a solid gel. The stiffening of the cytoplasm can also be demonstrated in other ways. If leukocytes with protopods are compressed to pancakes on a hematological blood film, their protopods remain projected (Fig. 3.9) in spite of a large membrane tension tending to flatten the protopod (Schmid-Schönbein et al, 1984). When an entire leukocyte with protopod is aspirated into a micropipette, it can be observed that the protopod may remain essentially undeformed, whereas the rest of the cell shows a large deformation and is locally compressed by the protopod. This attests in a dramatic way to the stiffness of the protopods.

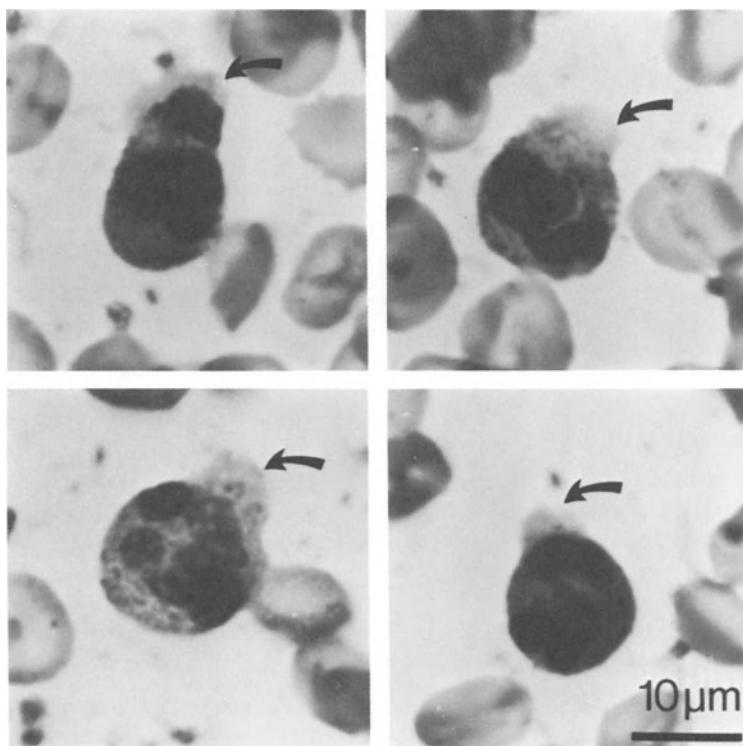


FIGURE 3.9. Light micrographs of neutrophilic leukocyte on a hematological blood film after compression to a pancake shape. The neutrophils are forming protopods at the time they are compressed. Note the projection with the typical granule-free cytoplasm (arrows). The protopods were not rounded out in spite of large isotropic membrane tension in such pancakes during compression.

## 7. The Interface

In the presence of a phase transition such as observed at the base of the protopod, the cell may conveniently be divided into two regions: (1) the region with the unpolymerized cytoplasm—which is governed by Equation (3.2)—and (2) the region consisting of the protopod. The two regions are separated by an interface that is the site of the polymerization activity.

Assuming isotropic linear elastic properties for the protopod, the constitutive equation for region (2) may be expressed in the form:

$$\sigma_{ij} = 2G(e_{ij} - e_{ij}^o) - p\delta_{ij}. \quad (3.4)$$

$G$  is a shear modulus for the polymerized region,  $p$  is the hydrostatic pressure, and  $\delta_{ij}$  is the Kronecker symbol. The term  $e_{ij}^o$  represents a strain due to change of volume and shape of the actin matrix along the interface between the two regions resulting from the polymerization process. It is the mathematical representation of the active contractile process.

In principal coordinates,  $e_{ij}^o$  can be expressed in terms of stretch ratios  $\lambda_i$ . These are related to the Cauchy strain  $e_{ij}^o$  by

$$e_{ii}^o = \lambda_i - 1, \quad i = 1, 2, 3, \text{ no summation.}$$

Further,  $\lambda_i$  for the actin matrix may be divided into two parts (Schmid-Schönbein and Skalak, 1984). One part is the volumetric strain by exclusion of the granules ( $\lambda_i^v$ ) and the other is due to alignment of the actin molecules ( $\lambda_i^p$ ) such that

$$\lambda_i = \lambda_i^v \lambda_i^p, \quad i = 1, 2, 3, \text{ no summation.} \quad (3.5)$$

Here  $\lambda_1^p \lambda_2^p \lambda_3^p = 1$ , so  $\lambda_i^p$  represents a pure shear deformation. The polymerization strain  $e_{ij}^o$  is an empirical quantity in the continuum theory that needs to be established from experiments. It represents the driving mechanism for the deformation.

The equation of motion in both regions of the cell reduces to

$$\frac{\partial \sigma_{ij}}{\partial x_j} = 0, \quad (3.6)$$

since inertia and gravitational forces are negligible in most cases. Here,  $x_i$  are the Cartesian coordinates in a fixed spatial frame of reference.

At any instant of time the base of the protopod is in contact with the unpolymerized cytoplasm. Thus if  $T_i^{(1)}$  is the traction of material region (1) on (2), and  $T_i^{(2)}$  is the traction of (2) on (1) along the interface in equilibrium,

$$T_i^{(1)} = -T_i^{(2)}. \quad (3.7)$$

The traction on the cytoplasm depends on membrane tension and the stresses exerted by the exterior fluid, such as hydrostatic pressure in free suspension or adhesive stresses at points of contact with a substrate.

The position  $X_i(t)$  of the interface is determined by the diffusion and reaction

of the trigger molecule and its antagonist. If  $c$  is the concentration of the trigger molecule (possibly  $\text{Ca}^{++}$ ), then

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x_j} D_c \frac{\partial c}{\partial x_j} + v_k \frac{\partial c}{\partial x_k}. \quad (3.8)$$

$D_c$  is the diffusion coefficient of  $\text{Ca}^{++}$  and  $v_k$  the velocity of the cytoplasm. At the interface,  $\text{Ca}^{++}$  is assumed to be absorbed into the polymerizing actin matrix at a concentration  $A_c$ . The velocity of the interface  $V_i(t) = dX_i/dt$  is then

$$V_i = \frac{1}{A_c} \left( D_c^2 \frac{\partial c^{(2)}}{\partial x_i} - D_c^1 \frac{\partial c^{(1)}}{\partial x_i} \right). \quad (3.9)$$

Equations (3.2) through (3.9) have to be satisfied simultaneously during protopod formation, starting from an appropriate set of initial and boundary conditions. The localized nature of the protopods in the early phases suggest that the initial influx of  $\text{Ca}^{++}$  across the membrane is localized. Different distributions of the calcium gates lead to different shapes of the protopods. At the beginning of the motion, the interface builds up at the membrane and then moves inward.

The notion of the moving interface also leads to a tenable explanation for the neck formation (Fig. 3.7, 3.8). This is a matter of the actin alignment and the shape of the interface. As the interface originates at the cell membrane, it forms a closed surface. Because of the alignment of the actin molecules, there exists a strain  $\lambda_z^p > 1$  with outward projection of cytoplasm, where  $r, \theta, z$  are taken as a local coordinate system on the interface (Fig. 3.10). Since  $\lambda_z^p$  represents a pure shear deformation, and if one assumes that  $\lambda_r^p = \lambda_\theta^p$ , then

$$\lambda_\theta^p = \frac{1}{\lambda_z^{p2}}. \quad (3.10)$$

Thus the lateral contraction  $\lambda_z^p > 1$  in the interface leads to a circumferential shortening  $\lambda_\theta^p < 1$ . If the interface has reached a point where the nucleus and other internal portions are penetrating the interface without polymerizing, a ring-shaped interface with an internal opening is formed. This leads to a circumferential stress  $\sigma_{\theta\theta}$  (or hoop stress), which tends to be self-centering and

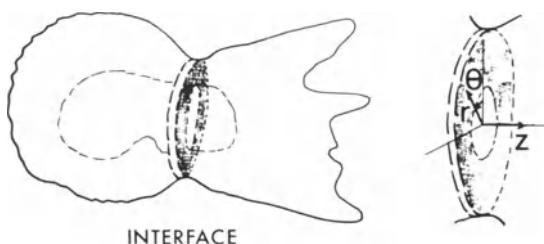


FIGURE 3.10. Schematic of the interface during neck formation.

yields axisymmetric necks. The degree of necking is a matter of which portion of the nucleus has reached the interface. In the case of the neutrophils, the nucleus is multilobed, so that deep and shallow necks can be observed depending on whether the polymerizing cortex can locally compress the nucleus. Neck formation may be similar to the furrows and contractile rings in cell division.

Although it has been recognized by early workers (Heilbrunn and Daugherty, 1932) that  $\text{Ca}^{++}$  can stiffen plasma gels and that it is important in muscle contraction, there is today no agreement that  $\text{Ca}^{++}$  is the rate-determining trigger molecule for the actin-myosin polymerization reaction (Pozzan et al, 1983). Also, the origin and chemical kinetics of the antagonist causing depolymerization are currently unresolved, although it is expected that high-energy phosphates play an important role. Both problems need to be investigated in greater detail.

## 8. Protopod Formation and Adhesion

Many physiological processes involving leukocytes require both adhesion and protopod formation. Migration of a cell on a substrate is only possible if the cell creates points or regions of attachment and if these are phased appropriately in time with points of detachment (Sullivan and Mandell, 1983). Phagocytosis of bacteria requires strong attachment of the white cell membrane to the bacterium, whereas rapid rolling of a leukocyte on the venous endothelium indicates a weak but not zero adhesion. Thus it appears that the strength of attachment can be modulated by biochemical agents generated in an inflammatory reaction or released by a bacterium. Some evidence for this exists in the case of a synthetic chemotactic agent, *f*-Met-Leu-Phe (*f*-MLP) (Schiffman et al, 1975), which causes monotonically increasing strength of adhesion with increasing concentrations (Fehr and Dahinden, 1979). At high concentrations of *f*-MLP, the leukocyte spreads on its substrate with strong adhesion, and even newly forming protopods cannot generate enough force to detach the membrane from its attachment area. This seems to be the condition most favorable for phagocytosis. With decreasing concentrations of *f*-MLP, the strength of adhesion seems to reduce monotonically. Little is known quantitatively about these relationships. A monotonically increasing adhesion force favors a directed motion of the cell, such as seen in chemotaxis (Carter, 1965). If the cell is located in a field that, on the average, exerts a higher force of attachment at one end than the other, then even in the presence of randomly oriented protopod projections, the cell will inevitably move towards the direction of stronger attachments. This occurs because the cell detaches at the point of weakest attachment to its substrate when a new protopod is projected. It appears to be a possible mechanism to ensure that a neutrophil finds its target in the tissue, (i.e., an inflammatory site or a bacterium). The *f*-MLP has another effect: it induces protopod formation. When added to a blood sample,

even in low concentrations, the cells in free suspension form protopods spontaneously. The *f*-MLP membrane attachment sites may not be uniformly distributed over the cell surface, which may cause projection of protopods in preferred directions. The action of *f*-MLP on the cell membrane thus serves a double function. On the one hand, it may open the calcium gates and allow this or other trigger molecules to enter the cell matrix, and on the other hand, it may provide a molecular substrate for attachment of the cell membrane to other surfaces.

## Acknowledgment

Supported by USPHS HL-10881 and HL-16851.

## References

- Brotschi, E.A., Hartwig, J.H., and Stossel, T.P. 1978. *Biol. Chem.* 253:8988–8993.
- Carter, S.B. 1965. *Nature* 208:1183–1187.
- Chien, S., Schmid-Schönbein, G.W., Sung, K.-L.P., Schmalzer, E.A., and Skalak, R. 1984. Pages 19–51 in *White cell mechanics: basic science and clinical aspects*, H.J. Meiselman, M.A. Lichtman, P.L. LaCelle, eds. Alan R. Liss, New York.
- De Bruyn, P.P.H. 1946. *Anatomical Rec.* 95:177–191.
- Evans, E.A., and Waugh, R. 1977. *Biophys. J.* 20:307–313.
- Fechheimer M., and Zigmoid, S.H. 1983. *Cell Motility* 3:349–361.
- Fehr, J., and Dahinden, C. 1979. *J. Clin. Invest.* 64:8–16.
- Fung, Y.C. 1977. *A first course in continuum mechanics*. 2d ed. Prentice-Hall, Englewood Cliffs, N. J.
- Fung, Y.C. 1981. *Biomechanics: mechanical properties of living tissues*. Chapter 4. Springer-Verlag, New York.
- Grant L. 1974. Page 342 in *The inflammatory process*. 2d ed., vol. 1, B.W. Zweifach, L. Grant, and R.T. McCluskey, eds. Academic Press, New York.
- Heilbrunn, L.V., and Daugherty, K. 1932. *Physiol. Zool.* 5:254–274.
- Kasai, M., Kawashima, H., and Oosama, F. 1960. *J. Polym. Sci.* 44:51–69.
- Majno, G., Shea, S.M., and Leventhal, M. 1969. *J. Cell Biol.* 42:647–672.
- Maruyama, K., Kaibara, M., and Fukada, E. 1974. *Biochim. Biophys. Acta* 136: 169–182.
- Mast, S.O. 1926. *J. Morph. Physiol.* 41:347–425.
- Mayrovitz, N.H., Tuma, R.F., and Wiedeman, M.P. 1980. *Microvascular Res.* 20: 264–274.
- Pozzan, T., Lew, D.P., Wallheim, C.B., and Tsian, R.Y. 1983. *Science* 221:1413–1415.
- Rikihisa, Y., and Misuno, D. 1977. *Exp. Cell Res.* 110:87–92.
- Schiffmann, E., Corcoran, B.A., and Wahl, S.M. 1975. *Proc. Natl. Acad. Sci. USA*, 72:1059–1062.
- Schmid-Schönbein, G.W. 1985. Leukocyte rheology. In *Handbook of Bioengineering*, R. Skalak and S. Chien, eds. McGraw Hill, New York, in press.
- Schmid-Schönbein, G.W., Jan, K.-M., Skalak, R., and Chien, S. 1984. *Biorheology* 21:676–781.

- Schmid-Schönbein, G.W., and Skalak, R. 1984. *J. Biomech. Eng.* 106:10–18.
- Schmid-Schönbein, G.W., Shih, Y.Y., and Chien, S. 1980. *Blood* 56:866–875.
- Schmid-Schönbein, G.W., Sung, K.L.P., Tözeren, H., Skalak, R., and Chien, S. 1981. *Biophys. J.* 36:243–256.
- Schmid-Schönbein, G.W., Skalak, R., Sung, K.L.P., and Chien, S. 1982. Pages 21–31 in: *White blood cells: morphology and rheology as related to function*, U. Bagge, G.V.R. Born, and G. Gaehrtgens, eds. Martinus Nijhoff The Hague, Netherlands.
- Skalak, R., Chien, S., and Schmid-Schönbein, G.W. 1984. Pages 3–18 in *White cell mechanics: basic science and clinical aspects*, H.J. Meiselman, M.A. Lichtman, P.L. LaCelle, eds. Alan R. Liss, New York.
- Stendahl, O.I., Hartwig, J.H., Brotschi, E.A., and Stossel, T.P. 1980. *J. Cell Biol.* 84:215–224.
- Stossel, T.P. 1982. *Phil. Trans. R. Soc. Lond. B*, 229:275–289.
- Sullivan, J.A., and Mandell, G.L. 1983. *Cell Motility* 3:31–46.
- Taylor, D.L., and Fechheimer, M. 1982. *Phil. Trans. R. Soc. Lond. B*, 299:185–197.
- Wolosewick, J.J., and Porter K.R. 1979. *J. Cell Biol.* 82:114–139.

## **Part II   Biological Tissues**

# 4

## Biomechanics of Smooth Muscle

J.M. PRICE

### 1. Introduction

Studies on the mechanics of smooth muscle contraction have borrowed many of the concepts that have been established for striated muscle. There are similarities between smooth muscle and striated muscle, yet there are important differences in the contractile filaments, their mechanism of action, and the excitation-contraction coupling mechanism that must be accounted for if a constitutive equation is to be obtained for smooth muscle (Burnstock, 1970; Murphy, 1976, 1979, 1982; Johansson and Somlyo, 1980; Price et al, 1977).

The ability of skeletal muscle to produce a contractile force and displacement when stimulated and to behave as a noncontractile tissue when not stimulated has led to the separation of force into active and passive components. A model of muscle to describe the active and passive states was originally proposed by A.V. Hill in 1938 (Hill, 1938, 1970). Hill's model is valid if the contractile element is freely extensible, and is described by Hill's hyperbolic equation if the series and parallel elements are elastic and if the muscle is composed of identical sarcomeres in parallel and series. None of these contingencies is true, even for skeletal muscle (Fung, 1981). Since Hill's model is for tetanized muscle, the description of a single twitch must include the time dependence of the contractile element with respect to stimulation. This time dependence is known as the active state and has been widely studied in heart muscle by various methods (Fung, 1970; Pinto and Fung, 1973; Pinto et al, 1975).

The assumption that the contractile element is freely extensible is necessary for separation of the force into passive and active components. Implicit in this is a uniqueness for the constitutive equation of the parallel elastic element (i.e., independent of the active state of the contractile element) (Fung et al, 1979; Tanaka and Fung, 1974; Yin and Fung, 1971). However, it can be shown that the division of force between parallel and contractile elements and the division of extension between contractile and series elements are not unique but are arbitrarily divided (Fung, 1981). Thus, experimentally obtained mechanical properties reflect a model that cannot be uniquely defined and are not intrinsic properties of the muscle.

An example of non-unique properties is demonstrated in arresting the spontaneous contractions of the taenia coli smooth muscle. Price, Patitucci and Fung (1979) have shown that suppression of spontaneous activity by different methods leads to different mechanical behavior of the muscle.

In skeletal muscle, the contractile response follows the "all or none" law because depolarization of the cell membrane is sufficient to generate an action potential and cause the contractile mechanism to be fully activated.

Smooth muscle cells of the taenia coli are called unitary because spontaneous contractions are propagated from one cell to many cells. In contrast, individual cells found in the wall of many arteries depend on the extracellular concentration of norepinephrine, epinephrine, potassium, and other agonists to initiate contraction. With norepinephrine, a graded depolarization without action potentials can occur and a tonic contraction is usually maintained. Hence, the active state is dependent on the concentration of chemical agonist.

The relationship between active force and agonist concentration in smooth muscle is known as the dose-response relationship (Tallarida et al, 1974). It is widely used in determining the sensitivity of smooth muscle to various agonists. In addition, the active force is known to vary with length. This relationship is well documented for smooth muscle. It has been assumed that the dose-response relationship and the length-active force relationship are independent of each other. Recent studies have shown that the shape of the length-active tension relationship of striated muscle at short lengths can be altered by caffeine, by stripping away the cell membrane, or by the pattern of electrical stimulation (Jewell, 1977). Thus, length alters activation as well as filament overlap. In this chapter, length-dependent activation in vascular smooth muscle and its relation to the dose-response relationship will be discussed. The hypothesis to be tested is that if the activation process in smooth muscle is length dependent then the dose-response relationship must be length dependent.

## 2. Methods

### 2.1. Preparation of Vessels

The dog anterior tibial artery was used in all experiments. A 5.0-cm segment was excised and placed in a dissection bath containing a physiological salt solution (PSS) at 37°C. It was bubbled with a mixture of 95% O<sub>2</sub> and 5% CO<sub>2</sub>. The adventitia was dissected from the vessel and cuts were made perpendicular to the longitudinal axis of the vessel. The cuts were approximately 2.0 mm apart in the middle portion of the segment. Ring width is the distance between the two cuts. Two stainless steel wires (.33 mm diameter) were inserted through the lumen of the ring. Each wire formed a triangle, with the corner opposite the side of the ring (base) attached to a supporting hook. The rings equilibrated in the bath for one hour at a 1.0 gram preload. A continuous flow of fresh PSS at

37°C was provided. The pH was maintained at 7.4 by bubbling with a 95% O<sub>2</sub> and 5% CO<sub>2</sub> gas mixture.

## 2.2. Experimental Apparatus

Distance between the wires in the vessel was adjusted by a micrometer. Force was measured by a Statham UC2 transducer. The apparatus has been described previously (Price and Davis, 1981; Price et al., 1981). The internal circumference, media thickness, and width of the ring were measured on-line with a noncontact electronic video caliper. In dose-response experiments the magnification factor was 38X. In length-tension experiments this was improved to 72X. A digital meter provided a readout of the distance that was precise to 10  $\mu\text{m}$ .

The length around the inside edge of the ring is analogous to the inside circumference when it is in a cylindrical shape. Therefore, the term *inside circumference* and *length* are used interchangeably. Length was measured as twice the distance from the lower edge of the upper wire in the lumen to the upper edge of the lower wire, plus the wire circumference and two wire diameters. In order to avoid end effects, media thickness and width of the ring were measured at a point half the distance between the supporting wires.

## 2.3. Length-Tension Experiments

The initial length for force in the resting vessel ring ( $L_0$ ) was determined first by shortening until force was zero. Then the ring was lengthened until force was first detected by the force transducer. Internal circumference at this position was measured as  $L_0$ . The ring was then stretched in increments of  $L_0$ . After each stretch the viscoelastic force was allowed to decrease to a steady state value before the ring was stimulated (approximately 15 min). The arterial ring was stimulated with the agonist norepinephrine (NE) or potassium. After stimulation the ring was allowed to relax to the resting force that was present before stimulation. Length was increased until the additional force resulting from stimulation decreased in two successive contractions.

## 2.4. Dose-Response Experiments

In dose-response experiments,  $L_{max}$  was determined first. This was done using electrical stimulation at parameters (frequency, duration, and voltage) for maximum response. The following protocol was used to obtain the dose-response curve at each length. The vessels were stimulated once with 10<sup>-6</sup> mol NE and allowed to relax in a normal PSS. Cumulative responses were obtained by increasing bath concentration from 10<sup>-8</sup> mol in specific increments such that the logarithm of concentration increased by log [5] and was continued until peak active force did not increase with increasing concentrations.

Then the vessels were returned to normal PSS and allowed to relax to resting force. Dose-response curves were obtained in different experiments at specific lengths relative to  $L_{max}$ .

Dose-response experiments with high potassium ( $K^+$ ) concentration were done in order at  $L_{max}$ , 0.70  $L_{max}$ , and 1.15  $L_{max}$ . The effect of time on the potassium dose-response relationship was determined in a separate group of vessels by repeating the experiment three times at  $L_{max}$ . Extracellular  $K^+$  was increased with  $K_2SO_4$ , which replaced KCL in the PSS. Cumulative responses to  $K^+$  were obtained by increasing the concentration from 5 m mol to 10, 20, 30, 45, 60, 90, and 120 m mol.

## 2.5. Data Analysis

We have shown (Price et al, 1982) that no decrease in force of the resting arteriolar tibial artery occurs when normal PSS is replaced by a calcium-free PSS. Accordingly, we define the additional force developed after stimulation as the active force. Stress was computed as force divided by the cross-sectional area of the media at  $L_{max}$ . The area was measured directly with the video caliper.

A best-fit curve to the data from a single dose-response experiment was obtained by a linear regression to the logarithmic form of Hill's equation (Price et al, 1981), or  $R = C^n/(K + C^n)$ , where  $R$  is the peak active force at a specific dose divided by the maximum active force in a dose-response experiment,  $C$  is the concentration of agonist, and  $K$  and  $n$  are constants. The values of  $ED_{50}$  ( $R = .5$ ), and the concentration for 10% of maximum active force ( $ED_{10}$  or  $R = .1$ ), were calculated from Hill's equation using the values of  $n$  and  $K$  that were determined from fitting a line to the logarithmic form of the equation.

Data are presented as group-mean  $\pm SEM$ . The two-tailed Student's *t*-test was used to determine statistically significant differences in the mean values at the 5% probability level.

## 3. Experimental Findings and Discussion

In a previous study of rabbit aorta (Tallarida et al, 1974) the dose-response curve with norepinephrine was shown to be different at two different preloads. Although the effect of agonist concentration on length-tension curves in vascular smooth muscle has not been previously studied, experiments on heart muscle (Huntsman and Stewart, 1977; Jewell, 1977; Lakatta and Jewell, 1977) have clearly shown a shift in the length-tension curve with a change in the degree of activation (contractility). Experimental manipulation of length was found to affect the contractile state of the muscle as well as the degree of overlap of contractile filaments. The current view of heart muscle is that length and contractility cannot be regarded as independent regulators of muscle

force. Similar results have been found in skeletal muscle (Jewell, 1977). The work on heart muscle has demonstrated that  $L_{max}$  shifts to longer lengths when active force is reduced, and towards shorter lengths when it is increased (Huntsman and Stewart, 1977). Our experiments show that the dose-response relationship of vascular smooth muscle depends on muscle length and the length-active force relationship is influenced by the concentration of agonist.

### 3.1. The Effect of Agonist Concentration on the Length-Tension Relationship

Length-tension curves with NE stimulation at  $10^{-5}$  mol concentration and  $10^{-6}$  mol concentration are shown in Figure 4.1 (see Price, 1984). The response is given as relative force so that the data are similar to a dose-response curve. The length ( $L$ ) is normalized to initial length for resting force ( $L_0$ ). If the length-tension relationship is independent of the concentration of agonist, then the two normalized curves in Figure 4.1 should be superimposed on one another. However, these curves are not the same. Specifically, the relative force of  $10^{-6}$  mol is significantly lower than the relative force for  $10^{-5}$  mol at all lengths less than 320%  $L_0$ . In addition, the length for maximum active force ( $L_{max}$ ) and the initial length for an active response ( $L_i$ ) are significantly longer for  $10^{-6}$  mol concentration than for the  $10^{-5}$  mol concentration.

Repeated length-tension curves obtained with  $10^{-6}$  mol concentration of

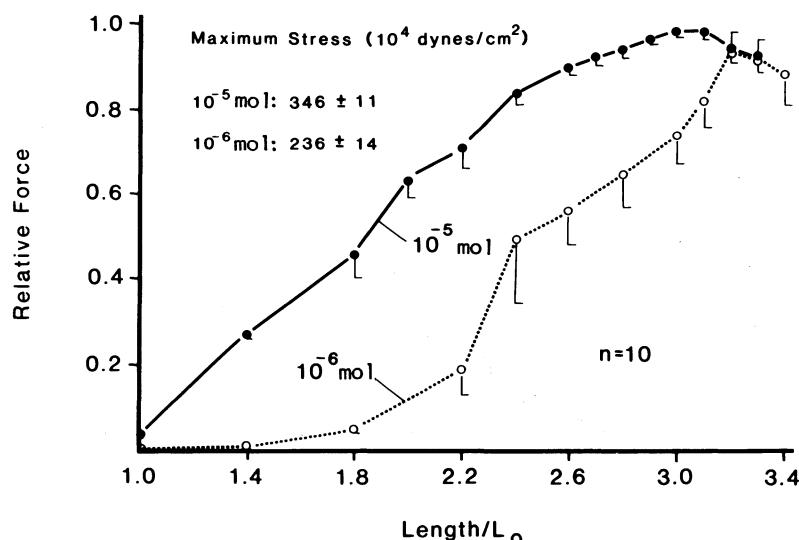


FIGURE 4.1. Length-tension relationship with  $10^{-5}$  mol and  $10^{-6}$  mol norepinephrine stimulation. The normalized force is significantly higher with  $10^{-5}$  mol than with  $10^{-6}$  mol at lengths less than 320%  $L_0$ . From Price (1984) by permission.

NE were essentially the same (Price, 1984). Although the responses tended to be slightly reduced when the second length-tension curve was obtained from each arterial ring, the difference was not significant for this group of vessels. There was no significant difference in  $L_{max}$  or  $L_i$  for repeated curves. We have also obtained repeated length-tension curves with  $10^{-5}$  mol NE. There was no significant difference in the first and second curve (Price, 1984).

Thus, differences in the length-tension curve are present and are characterized by a significantly lower normalized force for a low concentration of agonist as compared to the normalized force for a high concentration of agonist. The term normalized force represents the peak active force at each length divided by the maximum active force in a length-tension curve. In addition, the initial lengths for an active response ( $L_i$ ) and  $L_{max}$  were found to be significantly longer for a low concentration as compared to  $L_i$  and  $L_{max}$  for a high concentration of the stimulant.

### 3.2. The Effect of Length on the Norepinephrine Dose-Response Relationship

Dose-response curves with norepinephrine stimulation at different lengths relative to  $L_{max}$  are shown in Figure 4.2 (see Price et al, 1981). The response is calculated as the peak active force at a specific concentration divided by the maximum active force in a dose-response experiment. The order in which the dose-response curves were obtained was  $L_{max}$ ,  $0.85 L_{max}$ ,  $0.70 L_{max}$ , and  $1.15 L_{max}$ .

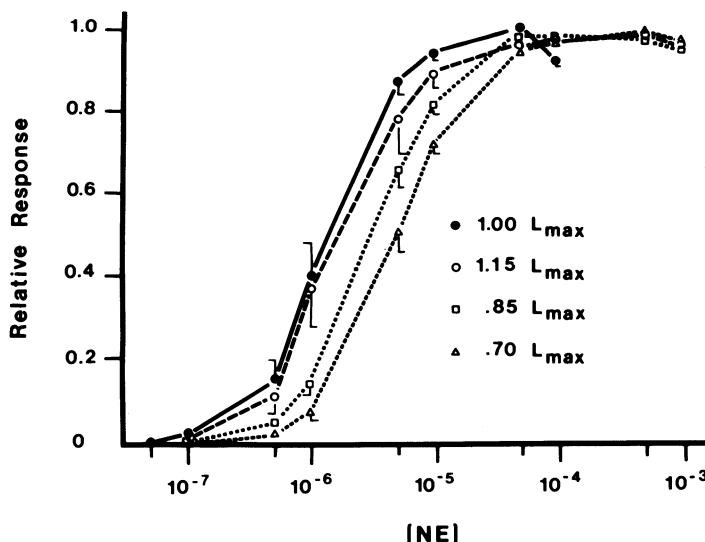


FIGURE 4.2. Dose-response curves at 1.00, 0.85, 0.70 and 1.15  $L_{max}$  with norepinephrine stimulation. Change in the  $ED_{50}$  and  $ED_{10}$  was significant for each change in length. From Price et al (1981) by permission.

$L_{max}$ . The curve at  $L_{max}$  appears to be well to the left of the curve at  $0.70 L_{max}$ . The shift of the curve to the right with decreasing length ( $1.00 L_{max}$  to  $0.85 L_{max}$  to  $0.70 L_{max}$ ) was reversed when the direction of length change was reversed and length was increased from  $0.70 L_{max}$  to  $1.15 L_{max}$ . The  $ED_{50}$ , maximum active stress in a dose-response experiment or  $S_{max}$ , and the preload (resting stress at each length) are significantly different at each length.

Dose-response curves with norepinephrine stimulation repeated at  $L_{max}$  were found to be superimposed on each other, and no significant change was found for the  $ED_{50}$ ,  $ED_{10}$ , and  $S_{max}$  (Price et al, 1981). We have also shown that a random order of lengths will give the same results as a set order of lengths (Price et al, 1981).

Dose-response curves with NE stimulation from arterial rings that were obtained at  $0.70 L_{max}$ ,  $L_{max}$  and  $1.30 L_{max}$  are shown in Figure 4.3 (see Price et al, 1983).  $ED_{50}$  and  $ED_{10}$  significantly decreased from  $0.70 L_{max}$  to  $1.00 L_{max}$ .  $ED_{50}$  and  $ED_{10}$  continued to decrease significantly with further stretching from  $1.00 L_{max}$  to  $1.30 L_{max}$ . This is in agreement with separate series of experiments that were performed at  $1.00$ ,  $1.15$ , and  $1.30 L_{max}$  (Price et al, 1983). The increase in absolute length and resting stress was significant for both changes of length. Maximum active force and maximum active stress increased significantly from  $0.70$  to  $1.00 L_{max}$  and decreased significantly from  $1.00 L_{max}$  to  $1.30 L_{max}$ . These results clearly show that sensitivity of vascular smooth muscle continually increases with stretch and does not have a maximum value at the length for maximum force.

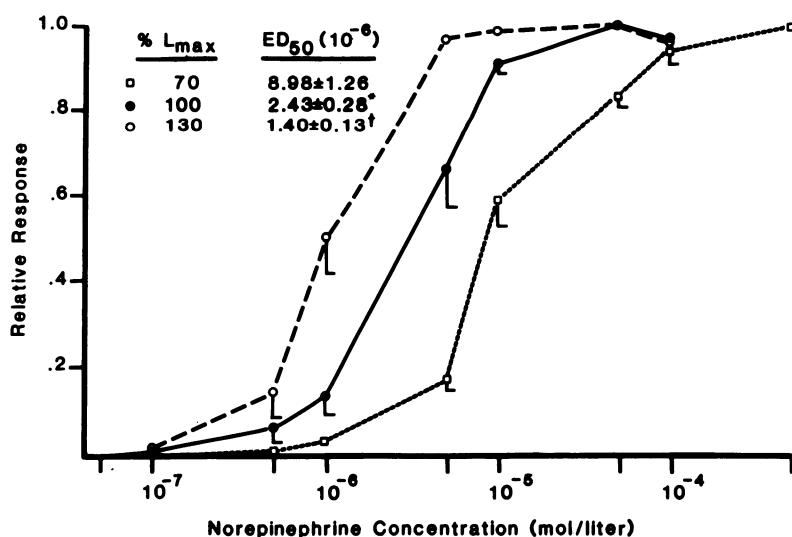


FIGURE 4.3. Comparison of dose-response relationship at a length below  $L_{max}$  to the relationship at a length above  $L_{max}$ . Change in  $ED_{50}$  and  $ED_{10}$  was significant for each change in length. From Price et al (1983) by permission.

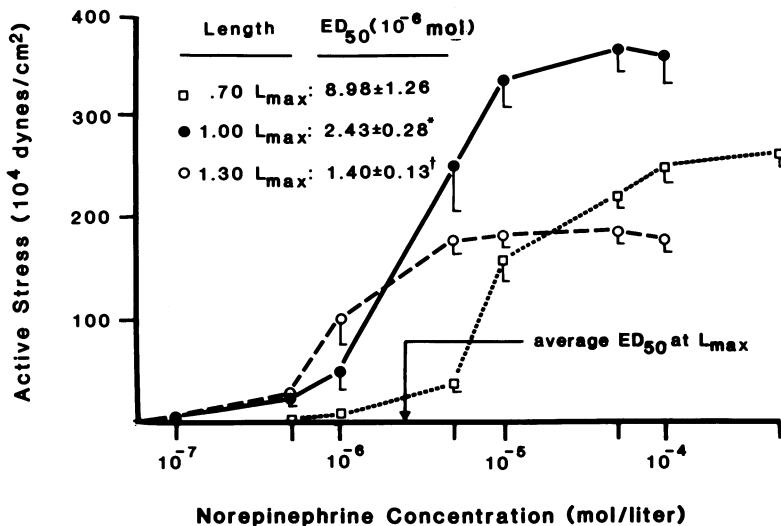


FIGURE 4.4. Dose-response relationship at different lengths, with the response given as active stress. Data are from same vessels as in Figure 4.3. From Price et al (1983) by permission.

It should be noted that the assumption is made here that muscle length is proportional to specimen length under all of the experimental conditions. All length-tension curves displayed a peak for maximum active force. In contrast, the  $ED_{50}$  was shown to be a monotonically decreasing function of length. This suggests that the change in  $ED_{50}$  with length does not share the same mechanism as the change in active force with length.

The dose-response curves of Figure 4.3 have been replotted in Figure 4.4, where the response is given as the absolute active stress rather than as relative force. Figure 4.4 suggests that the length-tension relationship may depend on the concentration of agonist. When the concentration of NE (for example,  $10^{-5}$  mol) is greater than the  $ED_{50}$  at  $L_{max}$ , the active stress is higher at  $L_{max}$  than at  $0.70 L_{max}$  or  $1.30 L_{max}$ . When the concentration (for example,  $10^{-6}$  mol) is less than the  $ED_{50}$  at  $L_{max}$ , the active stress is higher at  $L_{max}$  than at  $0.70 L_{max}$  but is less at  $L_{max}$  than at  $1.30 L_{max}$ . Thus, the dose-response data predict that the length-tension curves of Figure 4.1 should depend on agonist concentration.

The change in normalized length-tension curves with a change in concentration may be of significance in the *in vivo* behavior of arteries and their control of blood flow. Previous work (Meyers and Honig, 1969) has shown the magnitude of vasoconstriction to be a linear function of the resistance prior to stimulation in the vasculature of the dog hind limb. It was hypothesized that change in the initial value of resistance could impose a different degree of stretch on the vascular muscle, which would alter its capacity to respond through the length-tension relationship. Thus, the effect of a concentration-

dependent length-tension relationship may be to modify its role in controlling the resistance to blood flow. For example, if the concentration is high ( $10^{-5}$  mol), then with an increase in length the active force increases to a maximum value and in turn decreases with further stretch. In situations where the concentration is low ( $10^{-6}$  mol), the active force continually increases with length over the same range of lengths. Conversely, if the artery constricts and muscle length becomes shorter, the depressive effect of a shorter length on active force is stronger when the NE concentration is low. This interaction is demonstrated by examining the ratio of active stress for a high concentration vs. a low concentration at various lengths in Figure 4.1. At 300%  $L_0$  this ratio is approximately 2.0. At 180%  $L_0$  this ratio is approximately 9.0. At 140%  $L_0$  this active stress for  $10^{-6}$  mol NE is virtually zero, whereas the active stress for  $10^{-5}$  mol NE is approximately  $59.0 \times 10^4$  dynes/cm<sup>2</sup>. This behavior is in agreement with Speden's study (1973) on cylindrical pressurized segments of the rabbit ear artery. He showed that the effect of varying adrenaline concentration was to modify the radius at which active tension first began to be developed.

Gore (1972) concluded that arterioles are more responsive than arteries because their wall stress is closer to an optimal wall stress for vasoconstriction. Since the wall stress in arteries is normally greater than optimal, they cannot develop their full constrictive potential. He indicated that his results probably reflect the well-known length-tension properties of smooth muscle (Cox, 1976; Dobrin, 1973, 1978). Gore used an iontophoretic application of constant supramaximal doses of norepinephrine so that these results correspond to our data at high concentrations. At high concentrations, the active force decreases as length is increased above the optimal length. In experiments where a low concentration is used, the active force continues to increase with length over the same range of lengths, and attains a maximum response at a significantly longer length. These findings suggest that arteries may be more responsive than arterioles at low concentrations of norepinephrine.

### 3.3. The Effect of Length on the Potassium Dose-Response Relationship

Dose-response curves with potassium (K<sup>+</sup>) stimulation that are obtained at different lengths relative to  $L_{max}$  have similar results as norepinephrine (see Price et al, 1981). As shown in Figure 4.5, the curves are shifted to the right with decreasing length ( $L_{max}$  to 0.70  $L_{max}$ ), which was reversed when the direction of length change was reversed (increased from 0.70  $L_{max}$  to 1.15  $L_{max}$ ). The  $ED_{50}$  was significantly different from  $L_{max}$  to 0.70  $L_{max}$  and from 0.70  $L_{max}$  to 1.15  $L_{max}$ . We have also shown that the  $ED_{50}$  of repeated dose-response experiments at  $L_{max}$  with K<sup>+</sup> is not significantly changed (Price et al, 1981).

The fact that similar results were obtained with norepinephrine and with potassium stimulation shows that a norepinephrine receptor is not required for a length-dependent  $ED_{50}$ . In view of the studies on heart muscle (Jewell,

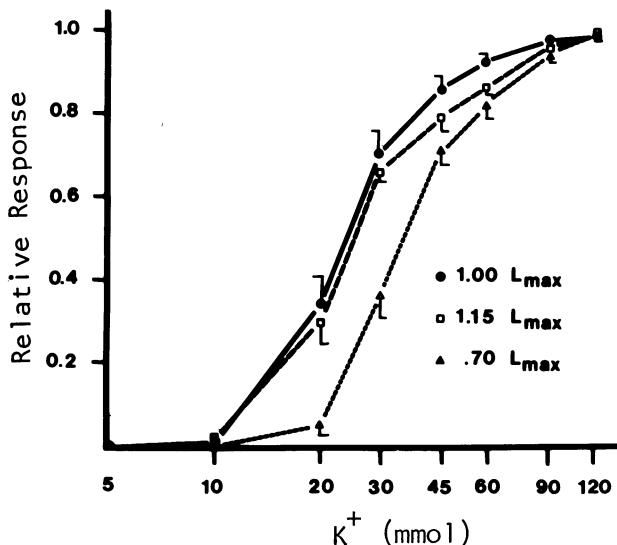


FIGURE 4.5. Dose-response relationship at 1.00, 0.70, and 1.15  $L_{max}$  with potassium stimulation. Change in  $ED_{50}$  and  $ED_{10}$  was significant for each change in length. From Price et al (1981) by permission.

1977), which show that release and/or binding of calcium to the contractile filaments could be the mechanism for a change in the length-tension curves with a change in contractile state, a similar mechanism may be responsible for the behavior of smooth muscle.

Although the data discussed in this chapter emphasize some new characteristics of smooth muscle behavior, the maximum active stress is consistent with measurements from strips and pressurized segments of arteries in previous work. Specifically, the average maximum active stress from dose-response experiments at  $L_{max}$  with norepinephrine was  $380 \times 10^4$  dynes/cm<sup>2</sup>.

## Acknowledgments

These studies were done in collaboration with Dr. D.L. Davis, Department of Physiology, College of Medicine, University of South Florida, Tampa. Technical support was given by T. Moher and E. Knauss. The research was supported by NIH grant # HL21103 and by the American Heart Association, Florida Affiliate.

## References

- Burnstock, G. 1970. Pages 1-69 in *Smooth muscle*. E. Bülbürg et al, Arnold, London.
- Cox, R.H. 1976. *Am. J. Physiol.* 230:462-470.
- Dobrin, P.B. 1973. *Am. J. Physiol.* 225:664-670.

- \_\_\_\_\_. 1978. *Physiol. Rev.* 58:397–460.
- Fung, Y.C. 1970. *J. Biomech.* 3:381–404.
- \_\_\_\_\_. 1981. *Biomechanics: mechanical properties of living tissues*. Springer-Verlag, New York.
- Fung, Y.C., Fronek, K., and Patitucci, P. 1979. *Am. J. Physiol.* 237:H620–H631.
- Gore, R.W. 1972. *Am. J. Physiol.* 222:82–91.
- Hill, A.V. 1938. *Proc. Royal Soc. London Ser. B* 126:136–195.
- \_\_\_\_\_. 1970. *First and last experiments in muscle mechanics*. Cambridge University Press, England, U.K.
- Huntsman, L.L., and Stewart, D.K. 1977. *Circ. Res.* 40:366–371.
- Jewell, B.R. 1977. *Circ. Res.* 40:221–230.
- Johansson, B., and Somlyo, A.P. 1980. In *Handbook of physiology*, sec. 2, Cardiovascular system, vol. 2, *Vascular smooth muscle*. D.F. Bohr, A.O. Somlyo, and H.V. Sparks, Jr., eds. American Physiology Society, Bethesda, Maryland.
- Lakatta, E.G., and Jewell, B.R. 1977. *Circ. Res.* 40:251–257.
- Meyers, H.A., and Honig C.R. 1969. *Am. J. Physiol.* 216:1429–1436.
- Murphy, R.A. 1976. *Blood Vessels* 13:1–23.
- \_\_\_\_\_. 1979. *Ann. Rev. Physiol.* 41:737–748.
- \_\_\_\_\_. 1982. *Hypertension* 4:Suppl II, 3–7.
- Pinto, J.G., Price, J.M., Fung, Y.C., and Mead, E. 1975. *J. Appl. Physiol.* 39:863–867.
- Pinto, J.G., and Fung, Y.C. 1973. *J. Biomech.* 6:617–630.
- Price, J.M. 1984. *Annals Biomed. Eng.* 12:481–496.
- Price, J.M., Patitucci, P.J., and Fung, Y.C. 1977. *Am. J. Physiol.* 233:C47–C55.
- \_\_\_\_\_. 1984. *Am. J. Physiol.* 236:C211–C220.
- Price, J.M., and Davis, D.L. 1981. *Blood Vessels* 18:75–85.
- Price, J.M., Davis, D.L., and Knauss, E.B. 1981. *Am. J. Physiol.* 241:(*Heart Circ. Physiol.* 10) H557–H563.
- \_\_\_\_\_. 1983. *Am. J. Physiol.* 245:(*Heart Circ. Physiol.* 14) H379–H384.
- Price, J.M., Davis, D.L., and Baker, C.H. 1982. *Blood Vessels* 19:203–216.
- Speden, R.N. 1973. *J. Physiol.* 229:361–381.
- Tallarida, R.J., Sevy, R.W., Harakal, C., Bendrick, J., and Faust, R. 1974. *Arch. Int. Pharmacodyn.* 210:67–74.
- Tanaka, T.T., and Fung, Y.C. 1974. *J. Biomech.* 7:357–370.
- Yin, F.C.P., and Fung, Y.C. 1971. *Am. J. Physiol.* 221:1484–1493.

# 5

# Smooth Muscle Rheology: In Search of a Specimen

H.Y.-L. CHEN and E.F. OWENS JR.

## 1. Introduction

Smooth muscle can be considered a general term for several types of tissue, each specially suited to particular functions. The challenge of the rheology of smooth muscle lies in the fact that it is “activity and relaxation intertwined” (Brady, 1977). Furthermore, the majority of smooth muscle exists as a component of an intricate composite, of which the mechanical properties are determined to a significant extent by the other components, such as collagen and elastin, and their configurations in the tissue.

The activity of smooth muscle can also add difficulty to mechanical testing. The ability of many visceral smooth muscles to contract spontaneously masks the passive tension. In addition, some smooth muscles are known to exhibit a myogenic response to stretch.

Tracheal smooth muscle is one tissue that has been found to be particularly well suited for mechanical studies. According to N.L. Stephens (1975):

- (a) All the fibers run parallel to each other in the transverse axis of the trachea.
- (b) More than 75% of the tissue is muscle, as assessed by histologic (Stephens and Wrogeaman, 1970) techniques.
- (c) The muscle under normal circumstances exhibits no spontaneous rhythmic contractile or electrical activity (Stephens and Kroeger, 1970) and does not manifest a myogenic response on the application of quick stretch (Stephens and Kroeger, 1970). These features permit recording of steady resting tensions in the muscle. The latter is necessary in the evaluation of active responses.
- (d) The muscle can be stimulated with single shocks to produce twitches or tetanized with either ac or dc currents. Thus studies of active state and force–velocity relationships are facilitated.
- (e) The summit of the active tension curve is such that between 85% to 110% of standard muscle length,  $L_0$ , there is no significant change in active tension. This is a desirable feature in deriving force–velocity curves, as pointed out by Brady (1965). Standard length, or optimal length, or  $L_0$  are synonymous terms and refer to that length at which maximum active tension is elicited.
- (f) Resting tension at the muscle length used for study is very small.

A firm groundwork in mechanical studies has been laid by Stephens, following the classical approach of Hill and Sonnenblick. A considerable number of biochemical and pharmacological studies have been made on tracheal smooth muscle, many of which use mechanical datum as an index of the effect of pharmacological agents. A phenomenological model of tracheal smooth muscle mechanical properties could serve as a tool for the interpretation of biochemical and pharmacological results.

The contractile properties have been characterized by force–velocity relationships under normal (Stephens, Kroeger, and Mehta 1969) and hypoxic conditions (Stephens and Kroeger, 1970). The physical properties of the series elastic component (SEC) have been studied using quick-stretch and force–velocity measurement techniques (Stephens and Kromer, 1971).

One study of the effects of temperature on force generation (Stephens et al, 1977) showed not only the expected change in the rate of active force generation, but also an increase in the stiffness of the SEC with decreased temperature.

The length–tension relationship of the unstimulated muscle has been used to describe the physical properties of the parallel elastic component. It was noted that, following an increase in length, the resting tension did not remain constant, but decreased with time, reaching a steady state usually within two minutes (Stephens, 1975).

Noting that there are sufficient differences among different smooth muscles, we can see that tracheal smooth muscle presents an ideal specimen for looking into the common features of smooth muscle rheology and, above all, for applying the methodology and the mentality that have been taught and demonstrated to us.

## 2. The Experimental System

The requirements that smooth muscle puts on the testing apparatus are unique. Forces need to be applied in the range from 50 mg to 50 g, with resolution of 10 mg; specimen length has to be changed from 0.1 mm to 10 mm with resolution of 10  $\mu\text{m}$ ; and temperature needs to be adjusted from 5°C to 40°C, with a resolution of 0.05°C. The data acquisition requires programmable linear time, with temporal resolution of 10 ms; logarithmic time regimen up to five decades; and a multichanneled vector of data that includes time, position, force, and temperature. Furthermore, the system should be software-oriented to provide easy access to a variety of loading regimens as well as their combinations in a preprogrammed sequence. The system should be quiet when the specimen is either at rest or at a programmed loading regimen, without instability or vibration that might stimulate muscle activity. Also, the viability of the muscle should be maintained.

Another major requirement is that the system can be operated effectively yet simply. Features such as data file transmission, storage, and management, and data reduction and presentation are needed.

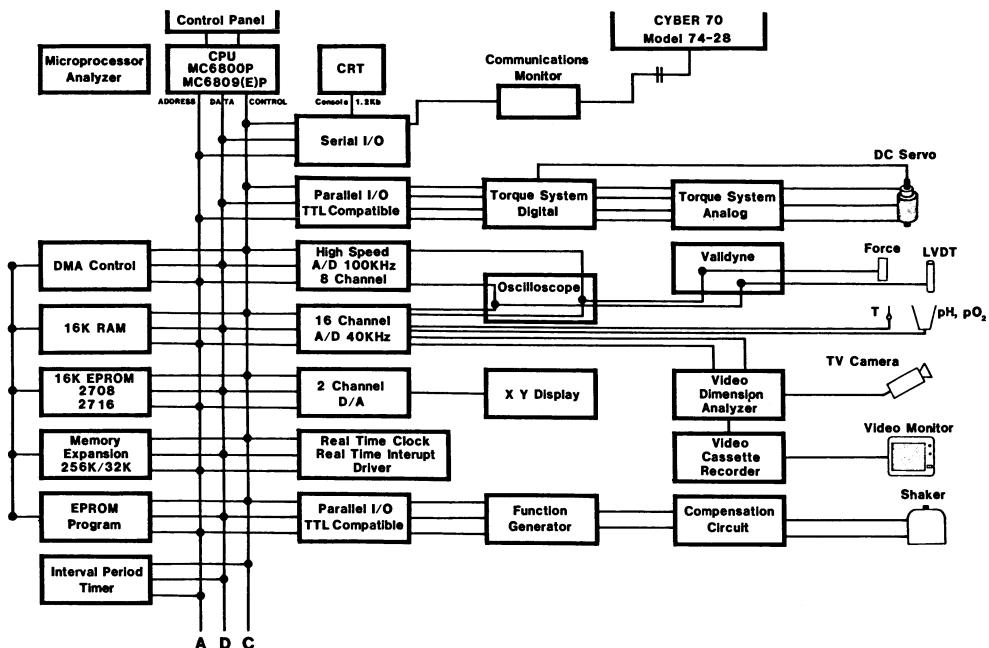


FIGURE 5.1. A schematic of the Alphatron.

A system that was designed with these criteria, the Alphatron, has been constructed in our laboratory and used for the test of smooth muscle. A report of the instrument was recently published (Chen et al, 1985). A schematic of the system is shown in Figure 5.1.

### 3. Reference Length

For mechanical studies, a reliable reference length is needed with which deformations can be compared. One reference length, referred to in the literature as  $L_{max}$ , or  $L_0$ , is determined experimentally as the length at which stimulation produces the maximum isometric tension. It represents the intrinsic optimal configuration of the muscle, independent of the geometry, and as such would be a good choice for a reference length.

In tracheal smooth muscle, however, "the summit of the active tension curve is such that between 85% to 110% of standard muscle length  $L_0$  there is no significant change in the active tension" (Stephens, 1975). There is possibly a 25% error in the determination of  $L_{max}$ . An uncertainty of this order of magnitude may render a reference length so determined unsuitable as a rheological parameter.

In studies of elasticity, the unstressed state is usually used as the reference length. Biological soft tissues have an ill-defined unstressed state. Therefore efforts have been directed toward defining the unstressed state as closely as possible. One such method uses the length of the tissue when hanging in solution, straightened by a 7.5 mg weight (Yin and Fung, 1971). At first glance, it would seem that this method does not take into account variations in the specimen length or thickness, and no error estimate is possible.

A method has been demonstrated for assessing the unstressed state of mesentery (Chen, 1973). The unstressed state is taken as the average of the lengths of the specimen when under slight compression (i.e., slightly buckled, and at the first sign of tension). The error, at most half of the difference of the buckled and slightly stressed lengths was reported to be 0.02 mm or 0.2% for the shortest specimen.

In a third method, the muscle is lengthened in successive increments of 0.2 mm until the first increase in force is indicated by the recording apparatus (Sparks and Bohr, 1962). The precise reference length, hence, depends on the sensitivity of the force-detecting apparatus used.

Our method is a two-point interpolation, with the first tension increase measured by the force transducer. A force change of 0.04 g is detectable. The length is increased from the slightly buckled length in increments of 0.127 mm. A length increase of 0.127 mm is the smallest length change that will produce a confidently measurable force change for our system. The reference length, denoted  $L_r$ , is the average of the lengths before and after the single movement that produces a sustained force change.

For a specimen length of 10 mm, this could entail as much as 0.6% error, and the force at  $L_r$  could be anywhere between zero and 30 mg. Even though the method of Yin and Fung (1971) may seem to have drawbacks, the straightening load is very small, the state of stress is known precisely, and the error is well within the range afforded by any method currently used.

#### 4. Viability

The mechanical tests are performed *in vitro*, so that the viability of the test specimen needs to be verified and maintained. If the muscle contracts vigorously when stimulated, it is considered viable.

At the beginning of the experiment, the length of the specimen is adjusted in increments of 0.127 mm until one gram of tension is indicated. As the muscle relaxes, the length is periodically increased to maintain the tension. The muscle is allowed to equilibrate for two hours at this length in continuously oxygenated Krebs-Ringer bicarbonate-buffered (KRB) solution at a designated constant test temperature.

The equilibrated length of the specimen, called  $L_a$ , is recorded and used for viability tests. This length was recommended by Dr. N.L. Stephens as appro-

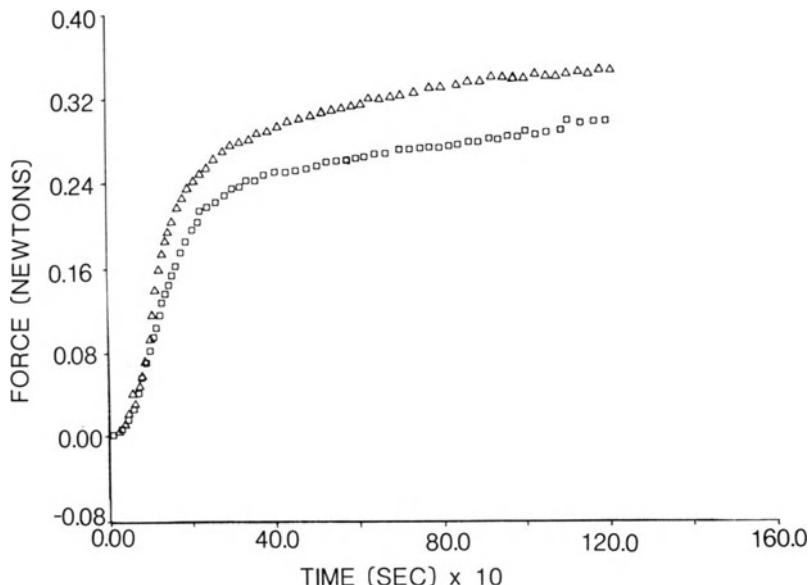


FIGURE 5.2. Isometric force versus time for a viability check. The triangular symbols ( $\Delta$ ) represent the isometric force development before smooth muscle testing, the squares ( $\square$ ) after the smooth muscle testing about 12–16 hr later.

priate for equilibration and isometric viability tests since the length of tracheal smooth muscle of these approximate dimensions under one gram of tension is near the optimal length for activation.

With the muscle still at  $L_a$ , the active state is induced by injecting Carbachol into the muscle chamber, to provide a Carbachol concentration of  $10^{-7}$  mole in the bathing solution. The time course of the contractile force is recorded (Figure 5.2).

This procedure establishes the initial viability of the specimen and sets a criterion, namely force generation, for subsequent viability tests. The viability of the specimen is again checked at the end of an experiment using the same conditions, (i.e., equilibration, the equilibrated length, temperature, stimulation, and bathing solution). If the force generated is at least 85% of that in the initial viability test, the data are considered valid.

## 5. The Passive State

The passive state is often defined as the unstimulated state. In tissues that exhibit spontaneous contractions, there may be no quiescent unstimulated state. Measures have been taken to circumvent this problem by functionally inactivating the contractile system. Such methods include the use of myo-

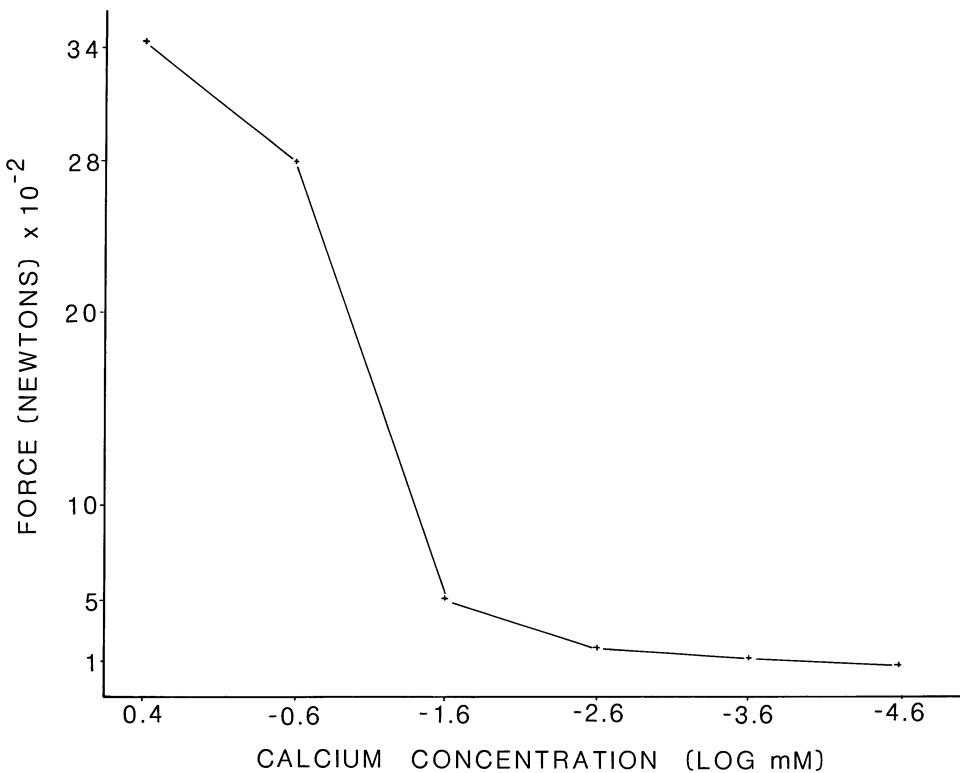


FIGURE 5.3. Isometric force generation as a function of calcium concentration in the bathing solution.

inhibitory substances, low temperatures, and removal of calcium from the tissue (Murphy, 1976; Lowy and Mulvany, 1973).

Even though tracheal smooth muscle does not exhibit any spontaneous activity, the unstimulated state may not be completely free of active tension. It was observed that the tension in the muscle in the low-calcium state at a given length is much less than the tension in the muscle at the same length in normal KRB solution.

For this study we approximate the passive state by the low-calcium state. The low-calcium state is achieved by equilibrating the specimen in calcium-free KRB solution. As shown in Figure 5.3, a test of the tissue response to Carbachol stimulation for varying calcium concentrations in the bathing medium showed the contractile force at a calcium concentration of  $2.5 \times 10^{-5}$  m mol to be less than 3% of that in normal KRB (pCa 0.4 m mol). The muscle seems able to tolerate the low-calcium state for long periods of time without apparent damage, since, when the tissue is returned to the normal KRB solution after several hours in the calcium-free solution, it is able to contract with nearly as much vigor as before the low-calcium state. For tests in the

passive state, the muscle and muscle chamber are rinsed with a 2.0 m mol EGTA KRB solution before the calcium-free solution is introduced into the chamber. The EGTA solution is not used during the tests in the passive state because the tissue, it seems, cannot tolerate this solution for long periods. The so-called calcium-free solution is not completely free of calcium because of impurities in the reagents and deionized water used and also because some calcium may be leached out of the tissue.

For tests that require long periods of time in the passive state, the low-calcium state is easier to use than the repeated administration of quickly metabolized drugs, such as epinephrine.

As we view the smooth muscle specimen as a polymeric material, the effects of temperature are twofold. The low temperatures used to inhibit muscle activity may not only affect the rate of metabolism of the muscle but may also change the rheology of the component polymers. The passive properties of the tissue at low temperature may not be applicable to the tissue at physiological temperatures.

Two categories of tests are performed on the tissue in the passive state:

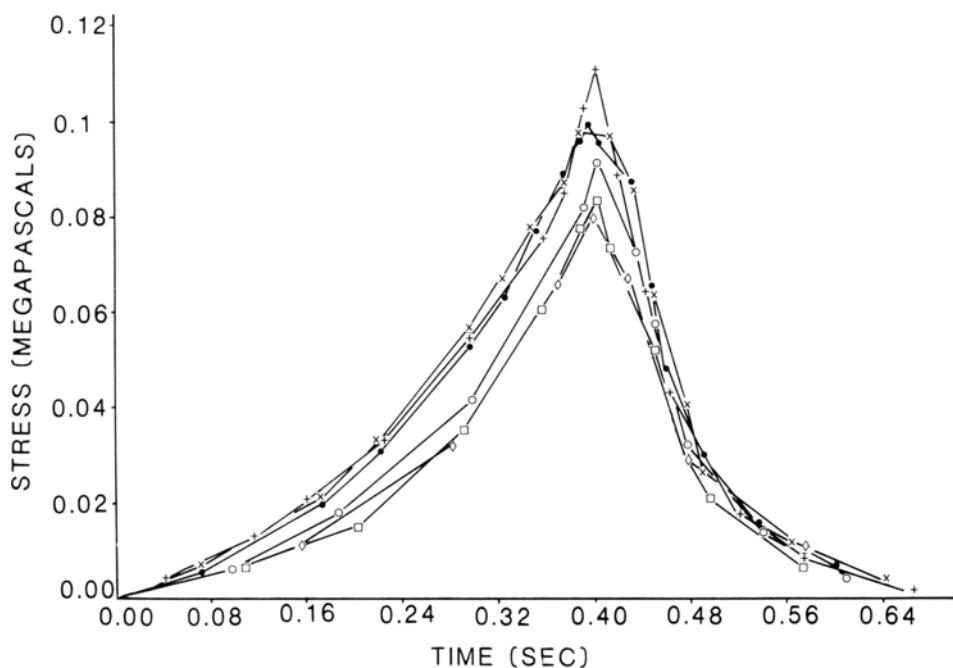


FIGURE 5.4. Stress versus time for a succession of cyclic elongation tests with decreasing time intervals between tests and with the same rate and magnitude of elongation. The first test is indicated by the dots (·). The other curves show the stress for different rest times between tests. Rest times are: 15 minutes (+), 10 minutes (x), 3 minutes (○), 1 minute (□), and 30 seconds (◊), respectively.

tests exploring the stress-strain-history relationship—such as the steady-state length-tension test, the relaxation test, and the cyclic elongation—and tests exploring the dependence of the passive behavior on temperature.

## 6. Reproducibility

Since a series of tests will be performed on the same tissue sample, it is important to determine how the tissue response is affected by previous tests. Usually preconditioning is used to achieve reproducibility. In a preconditioning procedure, several cycles of the intended test are performed at short intervals until the response becomes reproducible. One unique feature of tracheal smooth muscle is that it seems to recover completely after a test, so that reproducibility could be achieved without preconditioning.

In order to test the repeatability of the passive behavior, a series of the same test was performed at decreasing intervals of rest. Figures 5.4 and 5.5 show the results of a succession of cyclic elongation tests. The results of a succession of relaxation tests is shown in Figure 5.6. It was found that the tissue will

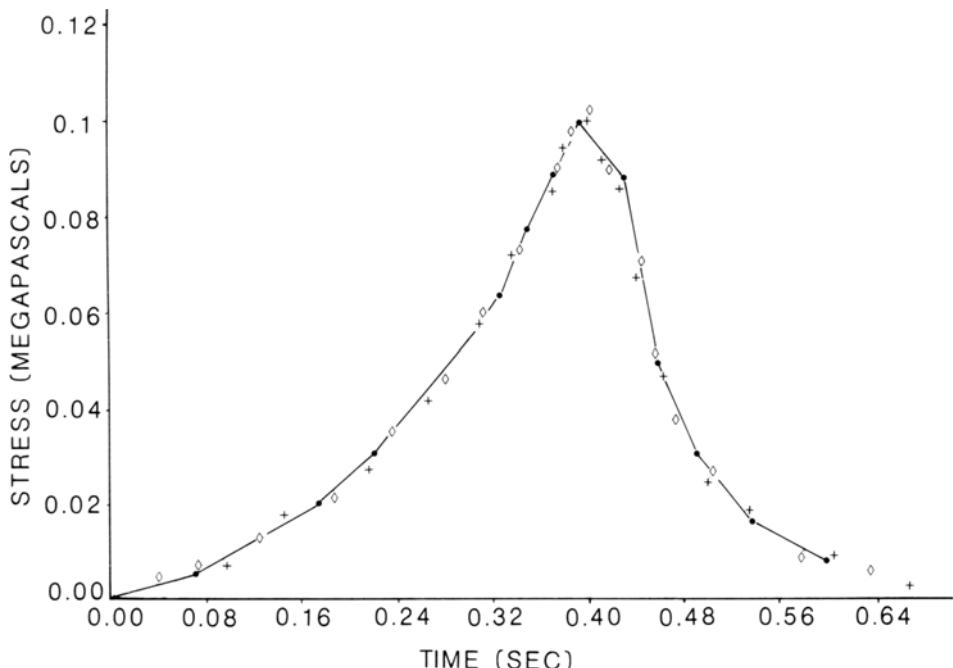


FIGURE 5.5. Another test was performed 15 minutes (+) after the sixth test in Figure 5.4, and was followed 9 minutes later by the final elongation test. The results are shown in comparison to the initial test (●) performed.

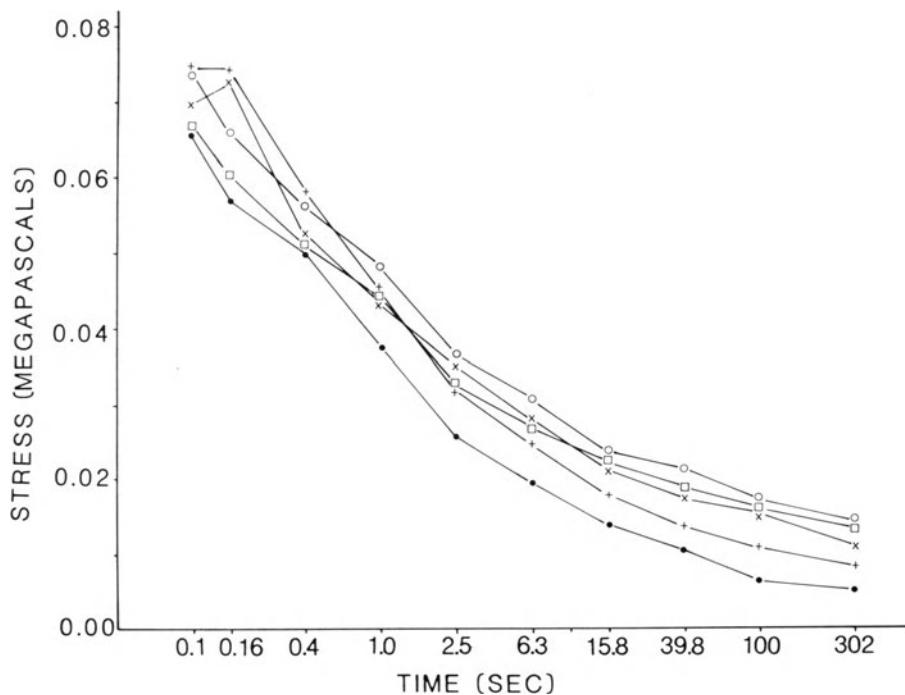


FIGURE 5.6. Stress verses time on a logarithmic scale for a succession of relaxation tests. The first test is represented by the dots (·). The intervals between successive tests were: 40 minutes (+), 24 minutes (x), 20 minutes (○), and 10 minutes (□).

recover completely and the results are reproducible as long as sufficient time is allowed between tests: 15–30 minutes after relaxation tests, and 5–10 minutes after cyclic elongation, depending on the extent of elongation.

Elongations of more than 40% of  $L$ , at rapid rates seem to damage the tissue because, shortly after several of these tests, there is a slow rise in the tension at  $L_r$ , even in the calcium-free solution. The rise in tension has been observed to exceed the previously recorded near-zero tension at  $L_r$  by as much as one-half gram weight, without showing any tendency to level off after three hours.

## 7. The Viscoelastic Behavior

A relaxation with normalized force verses logarithmic time is shown in Figure 5.7. It is noted that the stress relaxation is very rapid, with a well-defined steady-state residual tension, usually between 5 and 10% of the recorded initial tension, reached within 200 seconds. The muscle has also been shown to recover completely after elongation. It seems plausible to consider the tracheal

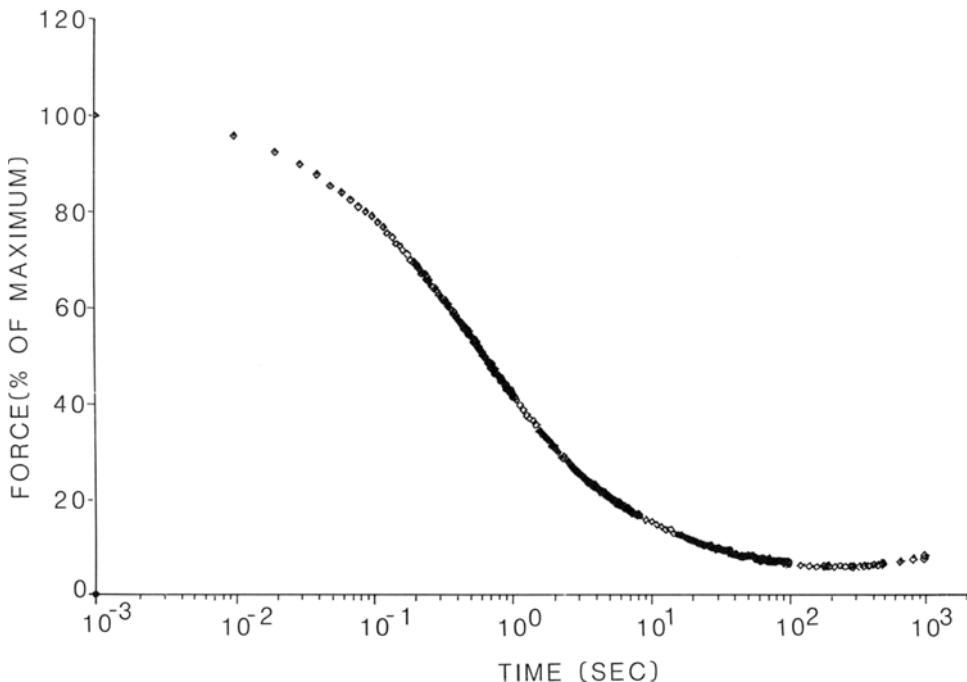


FIGURE 5.7. Normalized force versus time on a logarithmic scale for a typical relaxation test.

smooth muscle, in the passive state, as a viscoelastic solid. A temperature change from 37° to 27°C does not appear to have an appreciable effect on the time scale of stress relaxation, and relaxation has also been found to be indifferent to the extent of elongation.

The elastic response was studied by two tests: the cyclic elongation and the steady-state length-tension test. Both the steady-state length-tension relationship, as shown in Figure 5.8, and the hysteresis, as shown in Figure 5.9, show a nonlinear relationship between stress and stretch ratio. A mathematical model is needed to describe the elastic response. It appears to be either an exponential relationship, similar to that found for the ureter (Yin and Fung, 1971) and the mesentery (Fung, 1967), or a power law relationship as found in the taenia coli smooth muscle (Price et al., 1979).

Since stress relaxation is quite rapid in this material, it may interfere significantly with the stress increase during elongation. As a consequence, the apparent stress-strain relation is rate-dependent, and, it may not be advisable to approximate the instantaneous elastic response by the apparent stress-strain relationship. However, since there is a steady-state residual stress obtainable within a relatively short time, the steady-state elastic response could be used as the kernel in the formulation of a viscoelastic model.

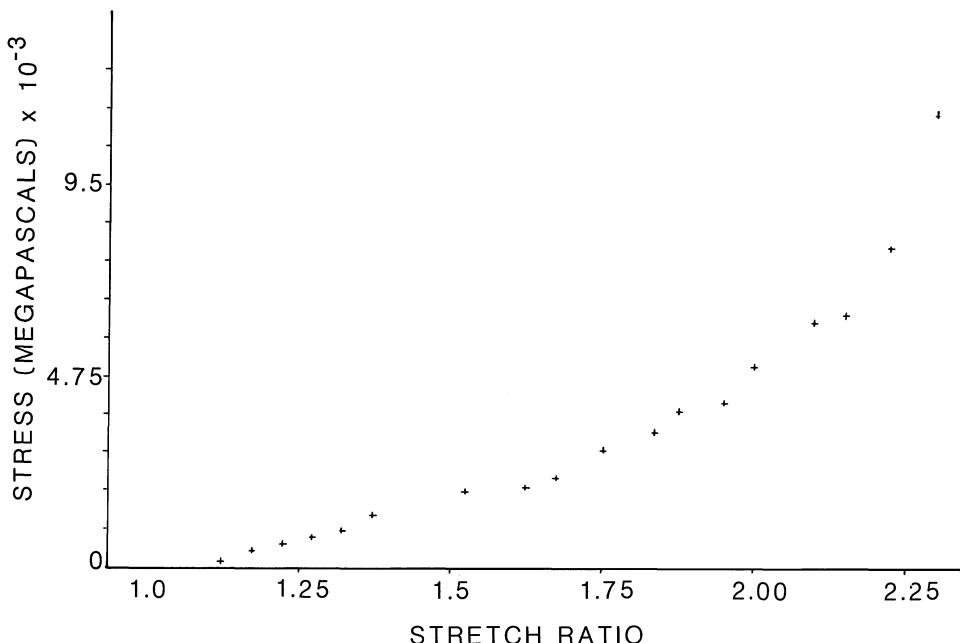


FIGURE 5.8. Stress verses stretch ratio for a steady-state length-tension test.

## 8. Temperature Effect

Tracheal smooth muscle can be considered as a polymeric system composed of collagen, elastin, and myofilaments in a network, filled with amorphous ground substance.

The rheology of most polymeric materials depends on temperature, and we have observed a temperature effect on the rheology of the smooth muscle in the passive state.

One of the tests on the effect of temperature on the passive behavior of the tracheal smooth muscle is the measurement of the residual isometric tension during a change in the temperature. Figure 5.10 shows the time course of the changes of stress and temperature for such an experiment.

At the start of the test, the specimen in the passive state was lengthened to a stretch ratio of 1.30. After the stress relaxation was complete, the temperature was lowered from 37°C to 17°C, and back to 37°C, continuously over a period of two hours. There was a slight reduction in the force as the temperature decreased, but as the temperature was increased, the tension rose almost exponentially, quickly exceeding the resting tension previously recorded at 37°C by almost sevenfold.

Apparently the increase in tension was produced by some active mechanism induced by the change of temperature. Activity induced by the change of

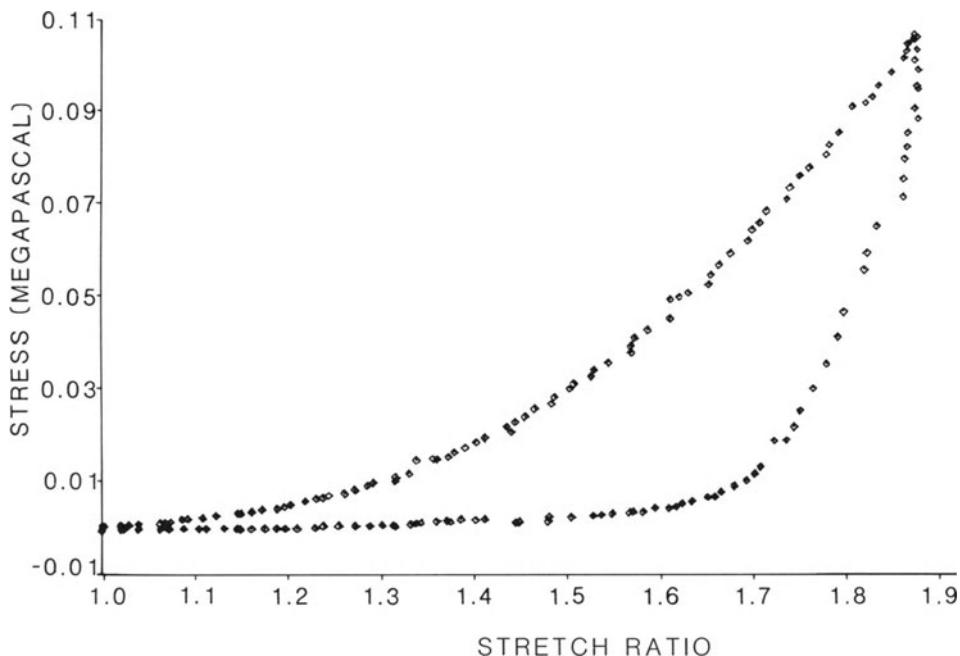


FIGURE 5.9. Stress verses stretch ratio for a cyclic elongation test.

temperature had been reported for intestinal smooth muscle in a review by Prosser (1974), and the effects of change of temperature on the membrane potential in smooth muscles has been documented by Magaribuchi et al (1973).

Attempts were made to remove additional calcium from the muscle with chelating agents, but the chelating agents themselves produced other activation artifacts. Epinephrine had no effect on the temperature-induced activity.

On the assumption that the muscle activity was induced by the combined effect of the change of temperature and tension, the temperature effect was, instead, investigated by allowing the specimen to equilibrate at  $L_0$  at various temperatures before performing the same series of relaxation and cyclic elongation tests. Thus the effects of temperature and the effects of the change of temperature are separated. As the temperature decreased from 37°C to 27°C, the stiffness of the tissue, as shown in Figure 5.11, decreased by  $54\% \pm 8\%$ . This follows the same trend of temperature dependence found in resting striated muscle (Hill, 1952).

From the network theory of rubber, the elastic constant,  $G$ , is represented as a function of the absolute temperature,  $T$ , by the equation (Treloar, 1967):

$$G = NkT, \quad (5.1)$$

where  $N$  is the number of molecular chains per unit volume, and  $k$  is the

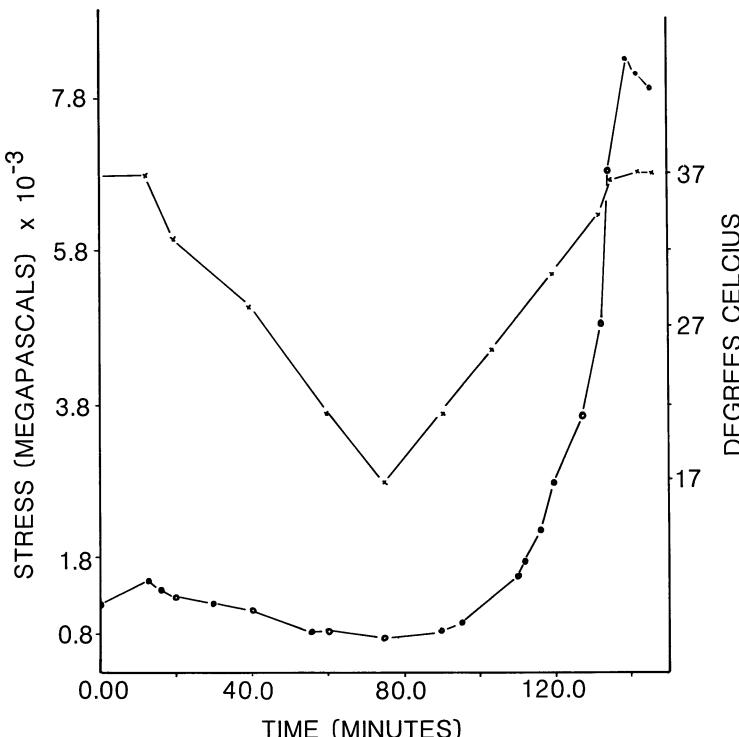


FIGURE 5.10. The effect of the change of temperature on the stress in a fully relaxed specimen. The circles represent stress with respect to the left axis, with temperature indicated by the symbol  $\times$ . Note the small decrease in stress during the decrease in temperature from 37°C to 27°C.

Boltzmann constant. The temperature change from 37°C to 27°C represents a 3.33% change in stiffness and could not explain the 54% change found in the stiffness of the smooth muscle. It has been suggested that the phenomenological theory of rubber elasticity should be more applicable to living tissues (Fung, 1972).

One of the challenges is to build a unified theoretical model to explain both time- and temperature-dependence.

## 9. Closure

A smooth muscle specimen has been found that has a one-dimensional structural geometry; that exhibits no spontaneous rhythmic contractile activity, nor manifests myogenic activity; and that can be easily obtained, kept viable during long experiments, and rendered passive. An experimental system has

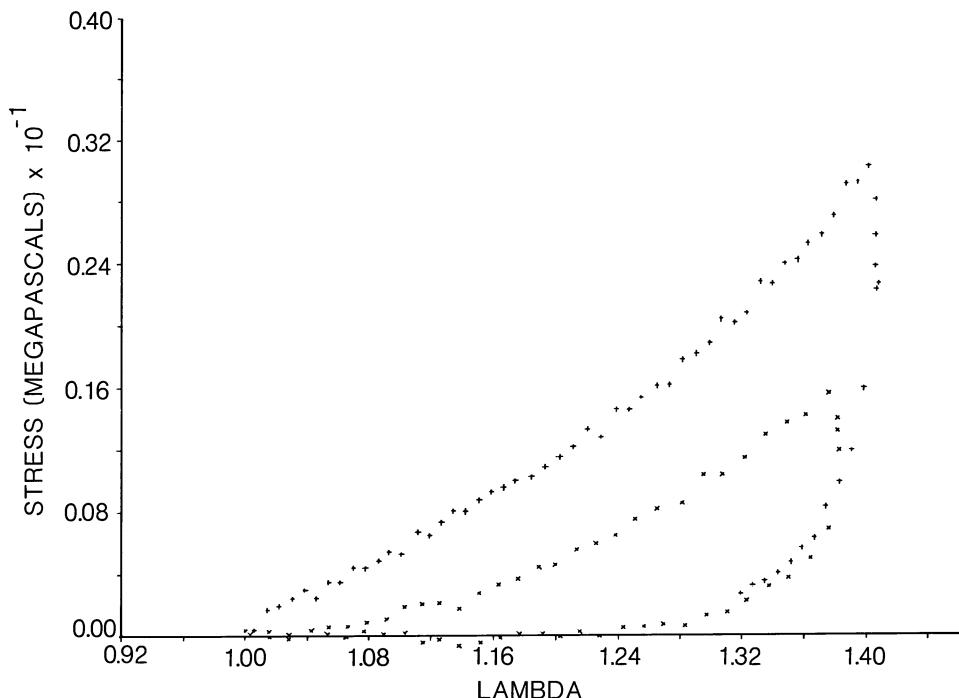


FIGURE 5.11. Stress verses stretch ratio for cyclic elongation tests to 40%  $L_r$  on the same specimen at different temperatures. The test at 37°C is denoted by +, and  $\times$  denotes the same test at 27°C.

been constructed with the characteristics needed for the study of the mechanical properties of the smooth muscle. It has been observed that the passive properties of tracheal smooth muscle are time- and temperature-dependent.

The classical model of muscle results from two dichotomies. Based upon observation, the first dichotomy gives the muscle a passive element and an active element. The second dichotomy further subdivides the active element into a contractile element and a series element, the contractile element being responsible for the length change and the series element being responsible for the generation of force. Clearly, the second dichotomy is one of convenience, since it is not based on either theoretical requirements or observations. The rapid relaxation of the tracheal muscle in the passive state indicates that it may not fit into this classical three-element framework. The smooth muscle behavior is activity- and viscoelasticity-intertwined since during activation, it is simultaneously in rapid relaxation. A clear-cut active state might be difficult to define. Hence the promises and the challenges would appear to lie in the interpretation and analysis of the data and in the formulation of a model to describe smooth muscle behavior in both the passive and active states.

## References

- Brady, A.J. 1977. Pers. Commun.
- Chen, H.Y.-L. 1973. Ph.D. dissertation, University of California, San Diego.
- Chen, H.Y.-L., Liu, Z.-Y., and Lee, J.L. 1985. *Proc. First China-Japan-USA Conf. on Biomechanics*. Science Press, Beijing, China.
- Fung, Y.C. 1967. *Am. J. Physiol.* 213:1532-1544.
- . 1972. Pages 181-208 in *Biomechanics: its foundations and objectives*, Y.C. Fung, N. Perrone, and M. Anliker, eds. Prentice-Hall, Englewood Cliffs, N.J.
- Hill, A.V. 1952. *Proc. Royal Soc.* 139(B):464-497.
- Lowy, J., and Mulvaney, J.M. 1973. *Acta Physiol. Scand.* 88:123-136.
- Magaribuchi, T., Ito, Y., and Kuriyama, H. 1973. *J. General Physiol.* 61:323-341.
- Murphy, R.A. 1976. *Blood Vessels* 13:1-23.
- Price, J.M., Patitucci, P., and Fung, Y.C. 1979. *Am. J. Physiol.* 236(5):C211-C220.
- Prosser, C.L. 1974. *Rev. Physiol.* 36:503-535.
- Sparks, H.V., and Bohr, D.R. 1964. *Am. J. Physiol.* 22(5):835-840.
- Stephens, N.L. 1975. In *Methods in pharmacology*, vol. 3, E.E. Daniel and D.M. Paton, eds. Plenum Press, New York.
- Stephens, N.L., and Kroeger, E. 1970. *J. Appl. Physiol.* 28(5):630-635.
- Stephens, N.L., Kroeger, E., and Mehta, J.A. 1969. *J. Appl. Physiol.* 26(6):685-692.
- Stephens, N.L., and Kromer, U. 1971. *Am. J. Physiol.* 220:1890-95.
- Stephens, N.L., Cardinal, R., and Simmons, B. 1977. *Amer. J. Physiol.* 233(3):C92-C98.
- Treloar, L.R.G. 1967. *The physics of rubber elasticity*, 2d ed., Oxford University Press, Oxford, England, U.K.
- Yin, F.C.P., and Fung, Y.C. 1971. *Am. J. Physiol.* 221(5):1484-1493.

# 6

# The Phenomenology of the Heart Muscle

J.G. PINTO

## 1. Introduction

This chapter discusses the philosophy of developing a constitutive equation for cardiac muscle based on uniaxial tests performed on papillary muscles. We begin with a background description of cardiac mechanics to highlight the methodological and procedural difficulties encountered in intact heart mechanics. The papillary muscle model is suggested as an alternative. Theoretical and experimental aspects requiring consideration in *in-vitro* papillary muscle mechanics are presented. The chapter concludes with the presentation of two constitutive equations; one describes the diastolic (passive) behavior of muscle, and the other describes its systolic (active or contractile) behavior.

## 2. Background

Cardiac mechanics aims to explain, in quantitative terms, the mechanical performance of the heart organ. From a clinical standpoint, this is a useful and significant objective. One can identify basically two approaches in the quantification of cardiac performance, and these may be termed as a direct approach and an indirect approach. In the direct method, the clinician attempts to measure the patient's blood pressure and its variation during the cardiac cycle. He may, in addition, measure the size and shape of heart, the cardiac output, the ejection velocity of blood, the systolic time duration, and other such parameters, to form an estimate of the "health" of the heart. This is the preferred method in current use. In the indirect method, effort is made to estimate the "health" using mechanical analyses based on fewer and more specific measurements. The former is an *ad hoc* experimental approach, while the latter aims to be of general validity based on principles of continuum mechanics. This analytical-experimental approach is still under development and, as such, is not yet widely used clinically. Its development depends on reasonably complete descriptions of the geometry, structure, and material (muscle) behavior of the heart organ.

The work of several investigators hitherto has led to impressive advances in the mapping of the dynamic geometry of the heart (Sandler and Alderman, 1974). Also, details on the architecture of the organ have been determined (Streeter and Hanna, 1973; Guasp, 1973). A description of the muscle material based on constitutive considerations has not been fully established at the present time. Considerable effort has been devoted to this subject by a number of researchers, beginning with the landmark paper by Fung on the quasilinear viscoelastic behavior of soft biological tissues (Fung, 1972). Ideally, a three-dimensional muscle law is required to facilitate the mechanical analysis of the heart organ. Such a law, however, is unavailable because of a number of difficulties. For example, interpretation of data from intact heart experiments is complicated by geometrical, structural, and material complexities. Additional aspects, such as nonuniform organ contraction, interaction between heart chambers, tethering influence of aorta, influence of the pericardium, influence of anesthetic agents, status of health, etc., all contribute to the difficulty of generating reliable data and interpreting it unambiguously and quantitatively.

A reasonable alternative to the problem is to first characterize the uniaxial behavior of a muscle bundle, and then use this data to make deductions on the behavior of the intact ventricle or heart. The transition from the uniaxial muscle model to the three-dimensional heart model must follow well-established principles of continuum mechanics. A method of analysis such as the finite-element method is suitable for heart analysis and has already been tried by many (Janz and Grim, 1972; Pao et al, 1974). The success of the method depends on the accuracy of the constitutive law used and, for this reason, the predictions of earlier analyses have been less than reliable. Clearly, in order to make progress, one needs a comprehensive description of muscle behavior. Such description appears feasible using the results of uniaxial tests performed on muscle bundles such as the papillary muscle isolated from the heart.

The virtue of an isolated muscle preparation lies in its simplicity. With simplicity, the hope arises that one can describe the constitutive equation in greater detail. Despite the apparent simplicity, however, a number of difficulties pertaining to methods and procedures enter the picture and need due attention. When a constitutive equation is obtained, it can be used as a building block to develop the analysis of the heart. Since the muscle law deduced from *in-vitro* experiments may be affected by the nonphysiological trauma suffered by the specimen during its preparation and testing, the results of analysis must be subjected to experimental verification. Until isolated muscle data is successfully applied in the analysis of intact heart and results of analysis compared satisfactorily with *in-vivo* data, we cannot claim success for the biomechanical approach. Professor Yuan-Cheng Fung, through his writings, lectures, and discussions, has profoundly influenced the progress of biomechanics. He has impressed upon his colleagues and students the importance of an analytical-cum-experimental approach to problems in physi-

ology. This chapter on the phenomenology of heart muscle highlights the spirit of Fung's approach.

### 3. The Phenomenology of Papillary Muscle

The general philosophy of phenomenology embodies statements on the relationship between cause and effect. In problems of tissue mechanics, the phenomenological approach aims to predict the behavior of the tissue under generally complex stress states using the experimental observations made under particularly simple and well-controlled (duplicated) conditions. In this respect, the uniaxial tests are the simplest. The papillary muscles that exist naturally within the heart are suitable for *in-vitro* laboratory testing. They are attractive because they are easier to extract with minimal damage. Strips of myocardium, although potential candidates, are unsuitable owing to their complex structure and geometry. Additionally, the strips suffer trauma and damage during resection. In comparison, papillary muscle is superior because specimens of slender size and uniform cross-sectional area (< 1 mm diameter) can be found (Pinto, 1980). Accordingly, their *in-vitro* survival is not jeopardized. Furthermore, their structure and geometry facilitate easier testing and cleaner interpretation of resulting test data.

### 4. Relevance of the Papillary Muscle Model to Intact Heart Analysis

Although the papillary muscle is an attractive test subject, as discussed above, the constitutive law based on its *in-vitro* biomechanical response may not be directly applicable in the analysis of the ventricular behavior. This qualification arises because *in-vitro* experiments are less than physiological, and hence the constitutive parameters derived may be inaccurate. Secondly, the papillary muscle parameters may not be representative of the myocardium (i.e., ventricular walls, septum, etc.). In view of this, only the functional forms of the equation can be used for analysis of the ventricles. Such extension of the equation appears justifiable because of a number of similarities in the mechanical behavior of *in-vitro* muscle bundles and that of the intact ventricle—for example: the tension-length feature of muscle vs pressure-volume (Starling) behavior of the ventricle, the tension-time course of muscle vs pressure-time course of the ventricle, the creep and creep recovery of muscle (Pinto and Patitucci, 1977) vs creep and creep recovery of the ventricle (Papadimitriou et al, 1974) etc. These similarities suggest that the functional form of the equation describing the mechanics of a muscle bundle may be similar to the equation describing the mechanics of the ventricle. By using this form of the equation to process data from specific *in-vivo* intact heart experiments, one can evaluate the unknown parameters appearing in the equation. The results

of other perturbations can be checked or compared with those predicted through analysis performed using the constitutive equation.

## 5. Theoretical Considerations: Concept of Linear Fading Memory

In the phenomenological approach to testing a material, one examines the relationship between cause and effect. Stated alternately, a relationship between cause and effect is sought in the form of a transfer function. The above statements can be cast in mathematical form as follows:

Let  $x(t)$  represent the history of input and  $y(t)$  be the corresponding output. Now suppose

$$\begin{aligned} x(t) &= 0 \text{ for } t < 0 \\ &= x(0) \text{ for } t \geq 0. \end{aligned} \quad (6.1)$$

Thus we write

$$y(t) = \phi(t)x(0) \text{ for } t \geq 0. \quad (6.2)$$

$\phi(t)$  in Equation (6.2) is generally a decreasing function of time and is known as the fading memory function.

A special case of (6.2) is the case of a perfectly elastic material, where  $\phi(t)$  is a constant independent of  $t$  and is replaced by Young's modulus  $E$ . We recognize the familiar equation

$$\sigma = E\varepsilon. \quad (6.3)$$

$E$  being constant, the material described by Equation (6.3) is said to have a perfect memory.

In the general case where  $x(t)$  is an arbitrary function of time,  $y(t)$  is determined by superposition if the cause and effect are linearly related. This is done as follows. We compute the change in  $y(t)$  for an incremental change in  $x(t)$  of magnitude  $\frac{dx}{d\tau} d\tau$  occurring at  $t = \tau$  as

$$dy(t - \tau) = \phi(t - \tau) \frac{dx(\tau)}{d\tau} d\tau. \quad (6.4)$$

By the principle of superposition for linear memory, the response  $y$  can be determined by integrating Equation (6.4):

$$y(t) = \phi(t)x(0) + \int_0^t \phi(t - \tau) \frac{dx(\tau)}{d\tau} d\tau. \quad (6.5)$$

In applying the above theme and the concept of linear fading memory to muscle, the problem boils down to identification of expressions to represent input and output functions such that they are uniquely and linearly related.

Expressions meeting these criteria were first hypothesized by Fung for soft biological tissues (Fung, 1972) and were later verified experimentally in the case of passive cardiac muscle (Pinto and Patitucci, 1980).

## 6. Experimental Considerations

Since the constitutive equation of a material is the description of a cause-and-effect relationship, the "effect" part for a specified "cause" can be obtained only through experiments. The simplest experiment is the uniaxial test, in which a biological specimen of "cylindrical" shape is prepared and held in a testing machine. The testing machine is usually equipped with a chamber containing a physiological solution of proper chemical composition to help preserve the biological viability of the tissue. The analysis of the recorded data is performed according to a theoretical framework, with the unification theme as a criterion.

The above scheme of experimentation was applied to papillary muscle specimens harvested from rabbit, cat, dog, and pig hearts in formulating its constitutive equation. Generally, the animal is anesthetized with an overdose of a sedative such as sodium pentathol or ketamine hydrochloride. Animals such as rabbits can be stunned to avoid the effect of sedatives on heart muscle. Heart muscle response is very sensitive to oxygen tension in the tissue. Accordingly, specimen preparation must be done while it is kept immersed in the physiological solution, which is bubbled continuously with 95% O<sub>2</sub> and 5% CO<sub>2</sub>. Carbon dioxide is necessary to maintain the pH balance of the solution at a value of 7.4. An oxygen tension of 600 mm Hg usually suffices to keep the tissue metabolism intact if the specimen chosen is thin (i.e., the diameter does not exceed 1 mm). For details in specimen preparation, the reader may refer to the literature (Pinto and Win, 1977; or Pinto and Patitucci, 1977). Recently, a detailed and more successful protocol for heart muscle preparation has been devised by P. Patitucci in Dr. Fung's laboratory (personal communication).

## 7. Experimental Difficulties in Papillary Muscle Research

During the course of research spanning several years, many aspects contributing to difficulties became evident. These are described briefly to highlight their importance in future research.

### 7.1. Attachment Artifacts

To apply the material testing procedures, the specimen needs to be held in a testing machine. Thread knot and wire hooks, tweezers, miniature clamps, etc., are some of the common means by which this is accomplished. These methods of support cause large deformation and mechanical damage of the tissue at the point of support. The effect of this damage on the normal response, and the

contribution of end-fixtures on the recorded data, is unknown and uncertain. The effect can be large, particularly when sudden perturbations are imposed on the specimen. Kreuger and Pollack (1975) have called attention to the effects of end conditions on papillary muscle test results. Additionally, non-uniform strain distribution caused by clamping and/or the natural taper in the specimen can cause significant uncertainties in experimental results (Pinto, 1977, 1978, 1980).

## 7.2. The Testing Machines

There is good evidence in the literature that shows how machine performance can influence the measurements. For example, Edman and Nilsson (1972) found different shapes for the force–velocity curve of muscle depending on whether input perturbations were properly damped. In this respect, it is necessary to choose testing machines, transducers, data recording devices, etc., with adequate frequency response. The common material testing machines usually found in engineering laboratories are not suitable for dynamic testing of biological tissues.

## 7.3. The Testing Procedure

As yet, no standardized method of testing papillary muscle has been developed. A lack of standard procedure has allowed different researchers to follow different procedures. For example, some investigators use the isometric–isotonic transition method (Sonnenblick 1962) to study the force–velocity behavior of the muscle. This method is gentle on the specimen, but has the drawback of acquiring information at different time instances in the twitch (i.e., at different active states of the muscle). To overcome this difficulty, other investigators use the quick-stretch or quick-release method. But this method is very abrupt and can be injurious to the muscle specimen (Bodem and Sonnenblick, 1975). In addition, abrupt perturbations in specimens of inhomogeneous structure and non-uniform cross-sectional area can cause nonuniform patterns of wave propagation. These waves may modify the “normal” response of the tissue or may cause corruption of recorded data.

## 7.4. Size and Shape of Specimen

Clearly, for uniaxial tests, a cylindrically shaped specimen is ideal. The need to minimize the artifact of end-fixtures suggests that we need to choose long specimens. Since the *in-vitro* viability of the specimen depends on the diffusion of nutrients and O<sub>2</sub>, we need specimens of small (<1 mm) diameter. A tailor-made specimen fitting all these criteria is generally not available in species that are commonly used in laboratories. Rabbit and cat hearts have only short specimens (generally <6 mm); specimens from dog hearts are generally bulky (diameter >1.5 mm). Common pigs in the weight range of 20–35 pounds yield

acceptable specimens about 80% of the time. It suffices to say that it is necessary to pay attention to the size and shape of the specimen in choosing one for mechanical testing.

## 8. A Constitutive Law for the Passive Heart Muscle

As stated earlier in this chapter, a constitutive law with features of uniqueness and generality was hypothesized by Fung in the later 1960s for soft biological tissues. It is popularly known as the quasilinear viscoelastic law. Its theoretical development can be found in Fung's recent text book *Biomechanics: Mechanical Properties of Living Tissues* (Fung, 1981). Assuming that heart muscle is a passive (i.e., noncontracting) material during the diastolic (filling) phase of the heart cycle, we can adopt the framework of the quasilinear law to describe the heart muscle. This framework can be written as

$$T(\lambda, \theta, t) = \int_0^t G(t - \tau) \frac{\partial T^e}{\partial \lambda}(\lambda, \theta) \frac{\partial \lambda}{\partial \tau} d\tau. \quad (6.6)$$

In the above expression,  $T$  represents the Lagrangian stress as a function of stretch ratio  $\lambda$ , temperature  $\theta$  and time  $t$ .  $G(t)$  is the so-called reduced relaxation function determined from relaxation experiments. Experiments performed on papillary muscle show that  $G(t)$  is a function only of time and that it is independent of  $\theta$  and  $\lambda$ .  $T^e(\lambda, \theta)$  is the "elastic" response and is a hypothetical quantity denoting tensile stress generated instantly in the tissue when a step stretch ( $\lambda$ ) is imposed on the specimen maintained at a temperature  $\theta$ . Due to the inability to attain an instantaneous length change in practice,  $T^e(\lambda, \theta)$  cannot be measured in the laboratory. However, the papillary muscle, like most biological tissues, displays insensitivity to the rate of strain (Pinto and Fung, 1973), and consequently, the term  $\frac{dT^e}{d\lambda}$  in Equation (6.6) can be justifiably approximated by a relationship of the type

$$\frac{dT^e}{d\lambda} \approx \frac{dT}{d\lambda} = \alpha(T + \beta). \quad (6.7)$$

$\alpha$  and  $\beta$  are constants that have been determined for rabbit papillary muscles from stress-strain experiments at different temperatures. The term  $\frac{d\lambda}{d\tau}$  represents the input perturbation and can take any of the forms that can be executed and controlled in the laboratory (e.g., ramp, step, sinusoid, etc.). The model shown in Equation (6.6) was subjected to partial experimental verification (Pinto and Patitucci, 1980). With relaxation experiments, a fading memory function  $G(t)$  was devised and is given by

$$G(t) = 1 + C[E_1(t/\tau_2) - E_1(t/\tau_1)]/(1 + C \ln(\tau_2/\tau_1)). \quad (6.8)$$

The expression  $E_1(t)$  in Equation (6.8) is the exponential integral function. The parameters  $C$ ,  $\tau_1$ ,  $\tau_2$  are determined from relaxation test results. The  $G(t)$  function, by virtue of its length and temperature independency, permits the application of linear viscoelasticity theory to the nonlinear cardiac muscle and allows the unification of several mechanical features under one formulation. Further verification that includes creep characteristics of the muscle remains to be done.

## 9. A Constitutive Law for the Contracting Heart Muscle

Following the unification theme presented earlier, this author proposed a constitutive law to describe the contractile behavior of the cardiac muscle bundle (Pinto, 1983). This proposal, has since been verified experimentally using papillary muscle specimen from pig hearts (Pinto, 1985).

In formulating this law, three principal features were taken into account: (1) the slow onset of contraction following stimulation; (2) the time feature of the twitch; and (3) length vs active tension of the muscle. The phenomenology incorporating the above features can be written as

$$F_{\mathcal{A}}(\lambda, t) = A(\lambda)t^{\gamma}e^{-\delta t}. \quad (6.9)$$

In (6.9),  $F_{\mathcal{A}}(\lambda, t)$  represents the active force as a function of nondimensional length ( $\lambda$ ) and time  $t$  after stimulus.  $A(\lambda)$  describes a factor related to the length-dependent amplitude of the twitch. It can be further described using the length-tension behavior, which at the present time has been approximated by two straight lines, one describing the ascending limb and the other, the descending limb. The parameters  $\gamma$  and  $\delta$  are introduced as constitutive parameters.  $\gamma$  accounts for contraction delay and  $\delta$  the inotropy (activation process) of the muscle. Since the time course of contraction is influenced by temperature, stimulation frequency, ionic composition of the perfusate, and many biochemical and pharmacological agents, the parameter  $\delta$  is not a single constant. Instead,

$$\delta = \delta \text{ (temperature, frequency, pH, chemical composition...)} \quad (6.10)$$

The formalism contained in Equation (6.9) interestingly describes well-known features of the muscle, such as the inverse force-velocity relationship, corresponding shifts dependent on length, shifts in force-time character depending on inotropy, and isometric, isotonic transient response.

The model presented above, besides being simple, unified, and general, carries a conceptual advantage over classical models, which describe the muscle in terms of “spring-and-damper” analogs. The drawback of these analogs is that they have influenced the procedure of experiments and data processing. Accordingly, different investigators have found different results, thus causing much doubt and debate in the literature. Our model overcomes

these difficulties. This model is a simple, explicit mathematical relationship incorporating the well-known interrelationships of force-length-time and contraction velocity of the muscle in the systolic state. Furthermore, the form of the equation can be readily extended to describe the pressure-volume-time and velocity of the intact ventricle. It is necessary to point out here that the above claims are made on the basis of the capability of the model to simulate well-known features of the muscle/ventricle as described in the literature. A comprehensive experimental validation remains to be done and is currently underway.

## Acknowledgments

Financial support for this work is received from USPHS Grant HL 29779-01. Ms. Susan Glover cheerfully typed the manuscript.

## References

- Bodem, R., and Sonnenblick, E.H. 1975. *Circ. Res.* 34:214-225.
- Edman, K.A.P., and Nilsson, E. 1972. *Acta. Physiol. Scand.* 72:205-219.
- Fung, Y.C. 1972. Pages 181-208 in *Biomechanics: its foundations and objectives*, Y.C. Fung, N. Perrone, and M. Anliker, eds. Prentice-Hall, Englewood Cliffs, N.J.
- \_\_\_\_\_. 1981. Pages 226-228 in *Biomechanics: mechanical properties of living tissues*. Springer-Verlag, New York.
- Guasp, F.T. 1973. *The cardiac muscle*. The Juan March Foundation.
- Janz, R.F., and Grim, A.F. 1972. *Circ. Res.* 24:339-349.
- Kreuger, J.W., and Pollack, G.H. 1975. *J. Physiol.* 251:627-643.
- Pao, Y.C., Ritman, E.L., and Wood, E.H. 1974. *J. Biomech.* 7:469-477.
- Papadimitriou, J.M., Hopkins, B.E., and Taylor, R.R. 1974. *Circ. Res.* 35:127-135.
- Pinto, J.G. 1978. *Biorheology* 15:511-522.
- \_\_\_\_\_. 1980. *J. Biomech. Eng.* 102(1), 61-66.
- \_\_\_\_\_. 1984. *Proc. First U.S.-China-Japan Mtg. on Biomechanics*, Science Press, Peking.
- Pinto, J.G., and Fung, Y.C. 1973. *J. Biomech.* 6:597-616.
- Pinto, J.G., and Patitucci, P.J. 1977. *Am. J. Physiol.* 232(6):H553-H563.
- \_\_\_\_\_. 1980. *J. Biomech. Eng.* 102(1):57-61.
- Pinto, J.G., and Win R. 1977. *Am. J. Physiol.* 223(3):H410-H416.
- Sandler, H., and Alderman, E. 1974. *Circ. Res.* 34:1-8.
- Sonnenblick, E.H. 1962. *Am. J. Physiol.* 202:931-939.
- Streeter, D.D., Jr., and Hanna, W.T. 1973. *Circ. Res.* 33:639-655.

# Determination of Material Properties of Biological Tissues: Pericardium

P.H. CHEW, S.L. ZEGER, and F.C.P. YIN

## 1. Introduction

The constitutive relationship of a material—that is, the three-dimensional relationship between its stresses and strains—uniquely defines its properties. Whereas the constitutive relationships of most engineering materials are well documented, this is not the case for soft tissues. The scarcity of data is due, in part, to the relative newness of the field of biomechanics and, to a greater degree, to the extreme difficulty of performing three-dimensional tests in soft tissues, with their large deformations, nonlinear properties, and susceptibility to deterioration.

Most existing data on soft tissue properties have been obtained under either uniaxial or biaxial loading conditions (Doyle and Dobrin, 1971; Fung et al, 1979; Hudetz, 1979; Manak, 1980; Parmley et al, 1973; Rabkin and Hsu, 1975; Sonnenblick, 1974; Spurgeon et al, 1977; Tong and Fung, 1976; Vaishnav et al, 1972; Vito and Hickey, 1980; Wiegner and Bing, 1981). Uniaxial studies are relatively easy to perform, and interpretation of the loads and deformations is rather straightforward. Uniaxial data cannot, however, provide a unique description of the tissue's constitutive properties because there are too many degrees of freedom (Crisp, 1972; Fung, 1973; Moriarity, 1980). While the full three-dimensional constitutive properties also cannot be deduced from biaxial data alone, they can be inferred when biaxial data are combined with results from another type of test, such as torsion or shear loading (Crisp, 1972; Fung, 1973). Thus, while biaxial data alone do not describe the tissue properties fully, they are a step in that direction.

Before biaxial tests can be interpreted meaningfully, some problems with quantification of biaxial test results need to be overcome. Some of these considerations have been discussed previously in studies of biaxial loading in skin and vascular tissue (Fung et al, 1979; Tong and Fung, 1976). Biaxial tests are more complex than uniaxial tests, primarily because in biaxial tests one can load the tissue in any number of ways. For example, Fung showed that when a specimen was subjected to successive replications of the same loading procedure, or to slightly different loading procedures, the coefficients of the strain-

energy function often varied considerably, sometimes even changing signs from one replication to another. In addition, the corresponding coefficients of individual specimens varied widely, making interpretation of the results difficult. The sources of the variabilities were not clearly identified, although instability of the nonlinear algorithms used to estimate the coefficients, the particular loading procedures, and real differences among specimens were mentioned as contributing sources. No method was proposed, however, to assess these sources of uncertainty in the coefficients. Thus, identification and quantification of some of the sources of variability are a necessary preliminary to meaningful interpretation of these test results.

This chapter briefly describes our approach to studying the mechanical properties under biaxial loading conditions of one class of soft tissues—the pericardium. We have restricted our attention to the pericardium because its membrane-like structure lends itself to biaxial testing. The experimental methods we employed are similar to those used previously for myocardial sheets (Demer and Yin, 1983), except that the use of an improved test apparatus has enabled closer control over the biaxial loading conditions. In particular, we describe here our data analysis approach, which consists of combining nonlinear estimation techniques with stringent statistical criteria. Our aim was to be able to estimate more reliably the coefficients of exponential pseudo-strain-energy functions similar to those proposed for other biological tissues (Fung, 1973). Using this approach, we feel that we can now better assess both the suitability of a particular strain-energy function, and some of the sources of variability of that function.

## 2. Methods

### 2.1. Data Acquisition

Pericardial sheets of approximately  $4 \times 4$  cm from the area overlying the anterior wall of the left ventricle were obtained from anesthetized mongrel dogs. Before the specimen was removed from the animal, the head-to-tail and transverse directions were marked. The specimen was prepared for mounting in the test apparatus by suturing the four edges in trampoline fashion, the previously marked directions being aligned along the two axes of the apparatus. The specimen was floated in a 37°C physiologic solution bubbled with 95% oxygen and 5% carbon dioxide to maintain pH at 7.4, for the duration of the test. The apparatus consists of two independent servocontrolled linear motors that allow control of the force, length, or velocity in two orthogonal directions. The motors drive two pairs of carriages that are coupled to two force transducers. The specimen is coupled to the carriages by four arrays of threads. Orthogonal surface dimensions in the central third of the specimen are measured by a noncontact video method, using two cameras and a beam splitter (Yin et al, 1972; Demer and Yin, 1983). The outer edge dimensions were

measured with a pair of LVDTs. To minimize viscous effects in these studies, we slowly loaded and unloaded the tissue at constant rate, one complete cycle being completed in about 100 seconds.

We wished to examine, specifically, the effects of different loading procedures. Therefore, we subjected this specimen to three different simultaneous biaxial cycles, each composed of loading and unloading with the ratio of the orthogonal strain components (henceforth called the *loading protocol*) constant during each cycle. The choice of the protocol was dictated by our desire not to exceed 500 grams total force in any direction. Each protocol was replicated three times. In another specimen, we assessed the combined effects of time and loading protocol by performing 10 replications of each of the three protocols in randomly selected order. Just before the data for any specific protocol were recorded, the tissue was preconditioned for several cycles. This involved cyclically loading and unloading at the selected strain rate, over the entire strain range of that particular protocol until the force-deformation curves were reproducible.

## 2.2. Quantification of Biaxial Test Results

The forces measured at the edges of the tissue, and also the dimensions of the outer edge and central portion, were recorded on analog tape for offline analysis. The thicknesses of the tissues were estimated by water displacement at the end of the study. The force and dimension data were digitized and converted to Kirchoff stresses and Green's strains in the principal directions of loading in the central region. For these calculations, the tissue was assumed to be incompressible. The data we report are those obtained during the loading portion of each cycle only. Analysis of the unloading portions would follow the same procedure.

We began by assuming, as suggested by Fung (1973), that the stress-strain properties could be described by an exponential pseudo-strain-energy function in terms of the principal Green's strain components. We initially tried the 7-coefficient strain-energy function that was proposed for skin (Tong and Fung, 1976). This function is

$$W = \alpha_x E_x^2 + \alpha_y E_y^2 + 2\alpha_{xy} E_x E_y + 0.5C \exp[A_x E_x^2 + A_y E_y^2 + 2A_{xy} E_x E_y] \quad (7.1)$$

where the  $E$ s are the Green's strains,  $C$ ,  $\alpha$ s, and  $A$ s are the coefficients obtained by fitting the stress-strain relationships  $S_x = \partial W / \partial E_x$  and  $S_y = \partial W / \partial E_y$  to the experimentally determined biaxial stress-strain data. As discussed below, however, we found that this function was overparameterized for each specimen. Consequently, we used a slightly simpler function of the form

$$W = 0.5B_1(E_x^2 + E_y^2) + B_2 E_x E_y + 0.5C(\exp[A_x E_x^2 + A_y E_y^2] - 1) \quad (7.2)$$

where  $B$ s, and  $C$  are the numerical parameters to be determined.

The goal of the quantification was to obtain estimates of the parameters of the strain-energy function and to assess the uncertainty of these estimates. Bivariate nonlinear estimation techniques and nonparametric statistical techniques for assessing uncertainty were used. An outline of the approach is as follows:

1. Based on criteria such as symmetry in the strains, positive definiteness, etc. (Fung, 1973), we first selected a strain-energy function.
2. For each data set, we ascertained qualitatively whether or not the selected strain-energy function was appropriately parameterized—i.e., that it was neither over- nor underparameterized. An overparameterized function will fit the data well but makes interpretation difficult because the coefficients may vary widely over replications. In addition, we have found that, for different starting values, an overparameterized function will often converge to different values of the coefficients. We have been able to identify overparameterized or colinear models by examining the correlation matrix of the estimated parameters. Overparameterization is characterized by a nearly singular matrix. Specific criteria for assessing colinearity are discussed by Belsley et al (1980). An underparameterized model will fail to represent the observed data adequately. The residuals from such models display systematic patterns. Thus, our goal in choosing a model was to obtain the best fit to the data without introducing problems of colinearity. Only those data for which the final function was suitably parameterized were retained for further analysis.
3. The next step was to assess the variability of the parameter estimates obtained and the resulting strain-energy function for each loading protocol. If the residuals are approximately independent and follow a Gaussian distribution, classical statistical methods (Draper and Smith, 1981) can be used. Because our data often violate the independence assumption, we have adopted a technique known as *bootstrapping* (Efron, 1980; Diaconis and Efron, 1983). The idea of bootstrapping is to mimic the process that generated the original data, thereby obtaining a bootstrap replica that has similar statistical properties. This procedure is repeated many times to obtain many independent data sets. Each of the generated data sets is then analyzed in the same way as the original one, and the variability in the resulting coefficients and strain-energy function can be assessed. The key step is to determine a procedure for generating a set of stress-strain data that are stochastically similar to the original set.

In our study, we repeated each loading a number of times. We assumed that the observed stresses for a single loading could be represented as the sum of the actual stress—a smooth error representing systematic deviations of the model from reality, and rough or random errors mainly representing experimental noise. We created a pool of smooth error curves composed of smoothed versions of the residuals from each of the protocols. A pool of rough errors was composed of the residuals after the smooth part was removed. A bootstrap stress-strain curve was then generated by adding a randomly selected set of smooth residuals and a randomly selected set of rough residuals (with replace-

ment) to the predicted values obtained from fitting the model to the original data.

4. The final step was to interpret the results. In addition to examining the variability in parameter estimates across specimens, or across different loading protocols within a specimen, it is also important, we feel, to examine the strain-energy function itself, since it incorporates all of the parameters. For assessing the load protocol dependence of the function, we used each of the bootstrapped sets of coefficients from the three protocols to calculate the corresponding strain energies. The choice of the strains at which to make the comparison was arbitrary. We elected to compare identical pairs of strains for each protocol, even though that strain pair may not actually have been attained experimentally. We used a low and a high level of equal  $x$  and  $y$  strains that were equivalent to the lowest and highest vector sum, respectively, of the strains attained at the conclusion of the three loading protocols. For the actual comparison, we have defined an approximately 95% confidence interval for a given set of strain-energies as the smallest set of values that include 95% of the bootstrapped replicates. If, at identical strains for the different loading protocols, any of the confidence intervals did not overlap, we concluded that there was loading history dependence.

### 3. Results and Discussion

Examination of the results from the 10 randomly selected replications for each loading protocol revealed no discernible difference in either the stress-strain curves or the estimated coefficients. This finding implies that, despite the several-hour duration of the test and despite differing boundary conditions, if the tissue is properly preconditioned, its mechanical properties are stable. Therefore, for conciseness, most of our data here pertain to only three replications of the loading protocol.

Obviously, a concise format for biaxial stress-strain data facilitates objective comparisons among specimens, or among multiple test results for a single specimen. Selection of a strain-energy function is a necessary first step in finding a concise format. The choice of an appropriate function has been the subject of considerable discussion (Blatz et al, 1969; Decraemer et al, 1980; Dehoff, 1978; Demiray, 1972; Fung, 1973; Fung et al, 1979; Gou, 1970; Manak, 1980; Synder, 1972; Tong and Fung, 1976; Vaishnav, 1980; Yin et al, 1972) and is constrained by the necessity for the function to be symmetrical, positive definite, to fit the data well over a well-defined range, and to be easily interpretable. The approach we have used follows that proposed previously (Fung, 1973; Tong and Fung, 1976; Fung et al, 1979) in which an exponential function in quadratic powers of the strain components is used. Fung has argued persuasively that the exponential form is more suitable than a polynomial. This is not to say, however, that another form of the function may not be as well, or even better suited for describing the data. How to use the

data to guide choice of the strain-energy function is a fascinating subject but is beyond the scope of this chapter.

We began with the 7-coefficient function proposed by Fung (1973). As we will show below, however, this function is overparameterized for these data. One of the difficulties that arises when an overparameterized function is used to fit the data is illustrated in Figure 7.1. The left panels show the values of the two exponential coefficients of Equation (7.1) for three replications of the three loading protocols. In contrast, when we used the better parameterized function, Equation (7.2), to fit the same data, we found much more consistent results within each protocol (Figure 7.1, right panels). It is clear that interpretation of the values of the coefficients when an overparameterized function are used would be difficult, and could be misleading. A function with fewer parameters avoids this problem—but before it is used, we need to examine how well it fits the data.

The most straightforward way to examine the goodness of fit is to see how closely the observed stresses correspond to those predicted. Representative stress-strain curves in both the  $x$  and  $y$  directions for each of the three loading protocols are shown in Figure 7.2. Only one of the replications is shown for each protocol, since the data from successive trials are almost superimposable on the first. The symbols are the actual experimental data, and the solid lines are the predicted stress-strain relationships, obtained using Equation (7.2) to fit the data. Figure 7.2 shows qualitatively that this simpler function provides a reasonably good fit to the observed biaxial stress-strain data and captures its sense.

A more quantitative feel for the goodness of fit is obtained, however, when one examines the residuals. An example of this is shown in Figure 7.3, where the composite  $x$ -direction stress residual, as well as its smooth and rough components for one of the loading protocols, are shown. The magnitudes of the smooth residuals are small, but there is a consistent pattern across the strain range. This consistent pattern was also seen in the  $y$ -direction stress residual. Indeed, this consistent misfit to the data has been seen in every data set we examined. In addition to this consistent pattern of the residuals, it should be noted that their variability also increases with increasing strains. Because the magnitude of this variability is small, however, our assessment of the uncertainty in the coefficient or in the strain energy is essentially unchanged whether or not we take this variability into account.

The slight but systematic variation of the residuals over the strain range indicates that the strain-energy function does not fit the data perfectly. At this point we have two alternatives: one is to try to find another function that may fit the data better; the other is to use alternative methods to estimate the variability due to noise in the data or the numerical algorithm in spite of the imperfect fit to the data. Despite the shortcoming of this 5-parameter function, we feel that the fit is reasonable, the results are consistent within each protocol, and the function is sufficiently concise to allow easy interpretation. Therefore, we have chosen to use this function and to obtain some estimate of the

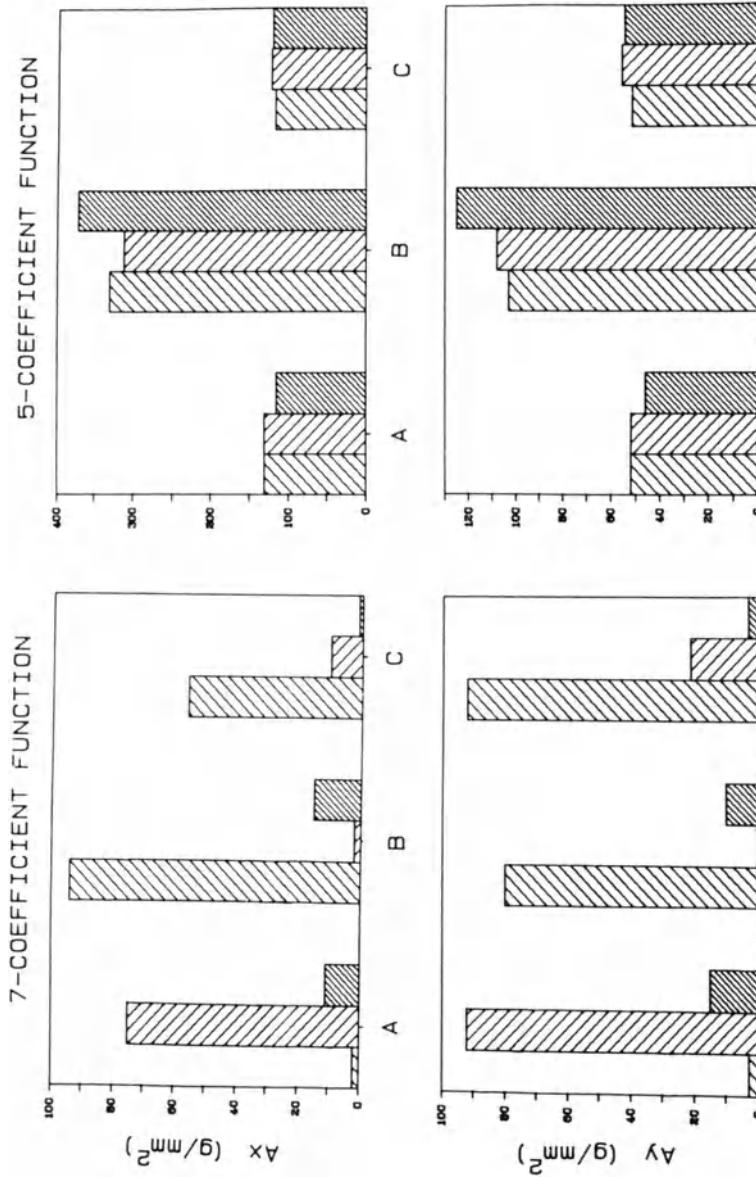


FIGURE 7.1. Values of the two exponential coefficients for the 7-parameter function in replications of each of the loading protocols. The protocols are indexed by the ratio of the  $y$ -direction to the  $x$ -direction strains. For this specimen, the ratios are: group A, 2.488; group B, 1.108; group C, 4.475.

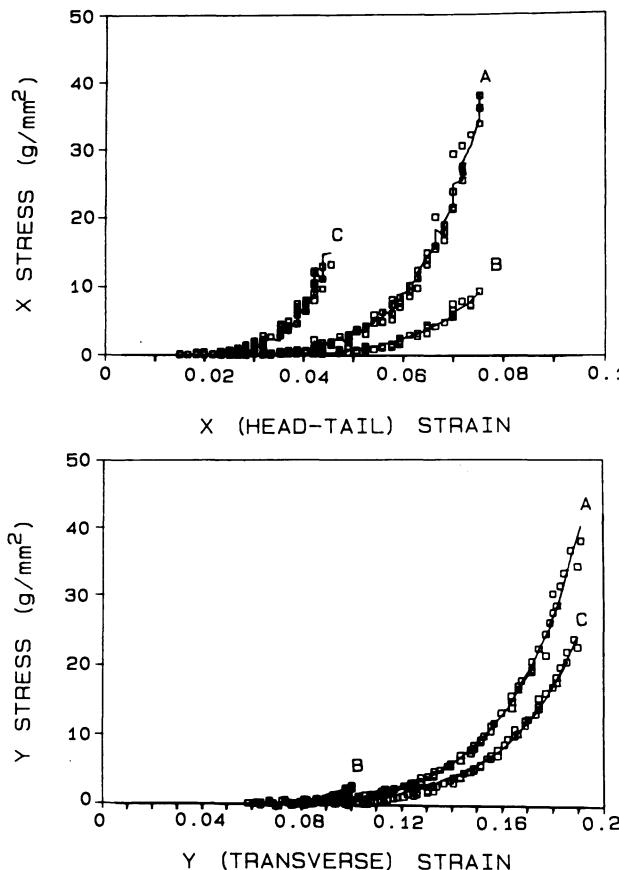


FIGURE 7.2. The stress-strain relations in the  $x$  and  $y$  directions for one replication of the three protocols labeled with the same nomenclature as in Figure 7.1. The symbols are the observed data, and the lines are the stresses predicted using the 5-parameter strain-energy function of Equation (7.2).

reliability of its individual coefficients, as well as of the overall strain energy. Since the residuals are not independent and identically distributed Gaussian variates, classical parametric statistical methods cannot be used to estimate the variability. Rather, we have selected the nonparametric bootstrapping technique for this purpose. We are still working on other methods of using the data to help select another function.

The data in Figure 7.4 illustrate the results of the bootstrapping technique. The strain energy calculated from the estimated coefficients for the three loading protocols is illustrated. The strain energy for all three protocols is evaluated at identical strain pairs (i.e., low strain of 0.0566 [left panel] and high strain of 0.1014 [right panel]). The groups of symbols are the results of the

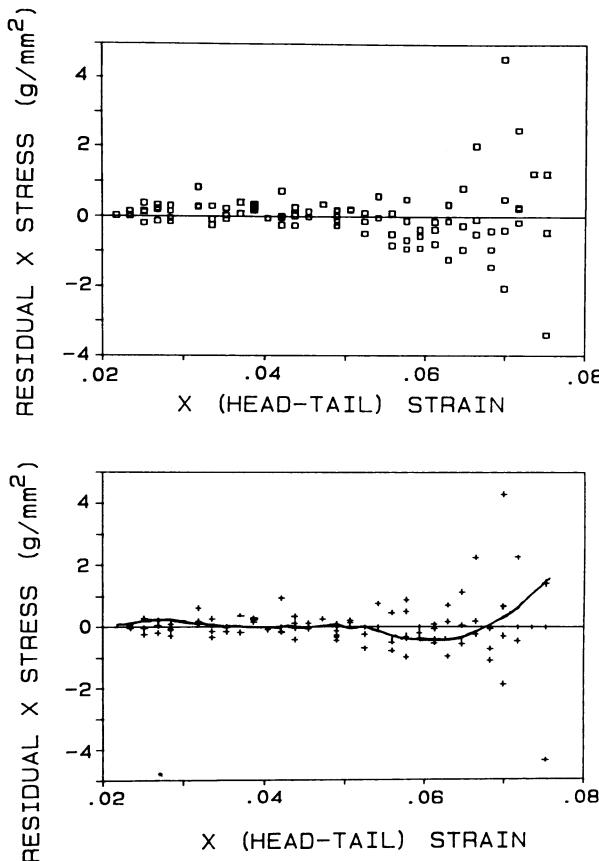


FIGURE 7.3. The stress residuals (ordinate) in the  $x$  directions as a function of the ordinal values of the strains (abscissa). The upper panel illustrates the total residuals. The lower panel illustrates the components of the residuals; the crosses are the random errors, and the lines are the smooth errors representing systematic deviation of the strain-energy model from reality.

bootstrapping performed 90 times for each loading protocol. The tight clustering of the values for each of the different loading protocols indicates that there is very little variability attributable to experimental or numerical noise. The confidence intervals for protocol *B* at the higher strains do not overlap those for the other protocols. Thus, the use of this strain-energy function to fit the data shows that there is strain history dependence of the tissue that cannot be attributed to either experimental or numerical variability.

The finding that the coefficients of the pericardium depend upon the exact history of loading is not surprising, since we know that pericardial tissue, like most other biological tissue, is viscoelastic. This finding, however, complicates comparisons among different specimens. Either several overlapping loading

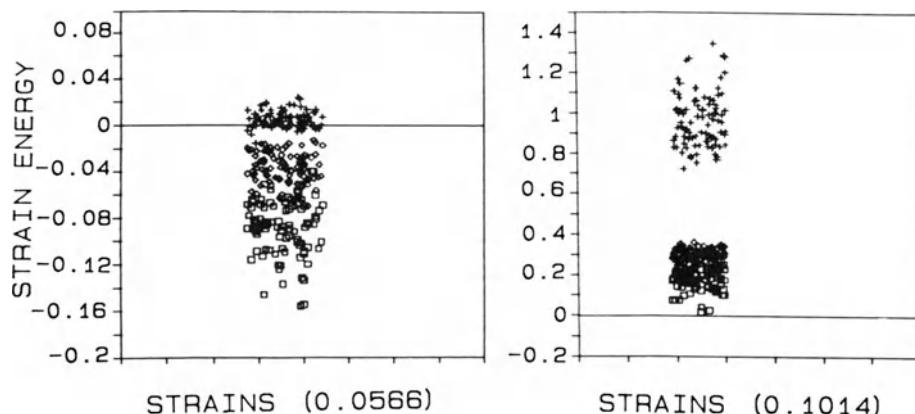


FIGURE 7.4. Estimates of the strain energy calculated from the 90 bootstrapped coefficients at two different levels of equal  $x$  and  $y$  strains for the three loading protocols. The variability in the strain-energy due to noise in the data or numerical algorithm is indicated by the range of each group of symbols ( $\square$  = protocol A;  $+$  = protocol B;  $\diamond$  = protocol C).

histories need to be performed for all the specimens, or all specimens need to be tested with the same loading history. Depending upon the specimen, this latter may or may not be feasible. Finally, because there may be loading history dependence, we must be careful in making physiologic extrapolations, because at this point we do not know the loading history that the pericardium is subjected to *in situ*.

Examination of the two estimated exponential parameters ( $A_x$  and  $A_y$ ) for the original data, prior to bootstrapping, for all three loading protocols are shown in Table 7.1. In every case, the value of  $A_x$  was greater than  $A_y$ . When bootstrapping was performed to obtain estimates of the variability of the individual coefficients, we found that the coefficients were statistically distinct within each loading protocol. Thus, we can also conclude reliably that the tissue is anisotropic, with a greater stiffness in the  $x$  than in the  $y$  direction. That is, this tissue is stiffer in the head-to-tail than the transverse direction. This finding is consistent with the results of other workers using different techniques (Wiegner and Bing, 1981). We must point out, however, that this assessment of relative stiffness is only qualitative, since it is based primarily upon the relative magnitudes of the two coefficients. The stiffness in each direction (the respective second derivatives of the strain-energy function) is an algebraic expression that is dependent upon the absolute strain levels and thus varies for different strain and stress levels.

Finally, we should point out some limitations of the 5-coefficient function that we have selected for describing the pericardial data. This function assumes that the tissue is isotropic in the low-strain region. This is not a severe limitation, since the tissue is so highly compliant that the very low stresses

TABLE 7.1. Coefficients of the strain-energy function

$$W = 0.5B_1(E_x^2 + E_y^2) + B_{xy}E_xE_y + 0.5C \exp[A_xE_x^2 + A_yE_y^2]$$

for three replications of three different biaxial loading protocols.

Protocol	$A_x$	$A_y$	$B_l$	$B_{xy}$	$C$
<i>A</i>	131.602	52.969	-14.637	-19.922	.299
	131.140	52.269	-15.188	-19.711	.312
	118.295	47.937	-20.945	-24.649	.425
<i>B</i>	320.673	101.683	-17.560	3.238	.0247
	398.509	105.381	-21.918	3.995	.0291
	341.166	117.982	-21.704	5.671	.0184
<i>C</i>	115.564	52.236	-36.168	-5.206	.382
	121.321	56.046	-32.329	-3.140	.299
	118.941	55.177	-34.495	-3.569	.316

generated in this region are just at the resolution of our force transducer and accurate determination of anisotropy would be difficult. The function also has no cross-coupling term between the two strains in the exponential term. This implies that, for loading protocols with strain along one axis held at zero while the opposite direction is stretched, the stress in the unstretched direction must increase linearly with strain. Whether the tissue behaves in this manner is difficult to check experimentally, however, because this type of loading protocol is quite difficult to perform. Finally, at the very low strain regions, the calculated strain energy is negative. This arises because of some unavoidable noise in the experimental data that led to the estimated negative coefficients. This is just a matter of an additive constant and could have been avoided if all negative stress values were eliminated. Nevertheless, we elected to include all of the data. Despite its limitations, this particular function is still useful for describing the mechanical properties of the tissue in a concise manner.

In conclusion, this chapter covers our method of obtaining, experimentally, biaxial stress-strain data for a simple tissue. We have shown that a 5-parameter combined exponential-polynomial strain-energy function provides reasonable fit to these data, and we have described a nonparametric technique to estimate the variability of the coefficients of this function. We can reliably conclude that (1) the pericardial properties are stiffer in the head-tail than the transverse directions; and (2) the properties depend upon the loading procedure imposed. Some physiologic implications of our findings are also discussed.

## References

- Belsley, D., Kuh, E., and Welsch, R. 1980. *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons, New York.  
 Blatz, P.J., Chu, B.M., and Wayland, H. 1969. *Trans. Soc. Rheology* 13:83-102.

- Crisp, J.D.C. 1972. Pages 141–179 in *Biomechanics: its foundations and objectives*, Y.C. Fung, N. Perrone, and M. Anliker, eds. Prentice-Hall, Englewood Cliffs, N.J.
- Decraemer, W.F., Maes, M.A., and Vanhuyse, V.J. 1980. *J. Biomech.* 13:463–468.
- Dehoff, P.H. 1978. *J. Biomech.* 11:35–40.
- Demer, L.L., and Yin, F.C.P. 1983. *J. Physiol.* 339:615–630.
- Demiray, H. 1972. *J. Biomech.* 5:309–314.
- Diaconis, P., and Efron, B. 1983. *Scientific American* 250:116–128.
- Doyle, J.M., and Dobrin, P.B. 1971. *Microvasc. Res.* 3:400–415.
- Draper, N., and Smith, H. 1981. Pages 458–529 in *Applied regression analysis*, 2d ed. John Wiley & Sons, New York.
- Efron, B. 1979. *SIAM Rev.* 21:460–480.
- Fung, Y.C. 1973. *Biorheology* 10:139–155.
- Fung, Y.C., Fronek, K., and Patitucci, P. 1979. *Am. J. Physiol.* 237:H620–H631.
- Gou, P.F. 1970. *J. Biomech.* 3:547–550.
- Hudetz, A.G. 1979. *J. Biomech.* 12:651–655.
- Manak, J.J. 1980. *J. Biomech.* 13:637–646.
- Moriarity T.F. 1980. *Circ. Res.* 46:321–331.
- Parmley, W.W., Chuck, L., Kivowitz, C., Matloff, J.M., and Swan, H.J.C. 1973. *Am. J. Cardiol.* 32:889–894.
- Rabkin S.W., and Hsu, P.H. 1975. *Am. J. Physiol.* 229:896–900.
- Snyder, R.W. 1972. *J. Biomech.* 5:601–606.
- Sonnenblick, E.H. 1974. *Am. J. Physiol.* 207:1330–1338.
- Spurgeon, H.A., Thorne, P.R., Yin, F.C.P., Shock, N.W., and Weisfeldt, M.L. 1977. *Am. J. Physiol.* 232:H373–380.
- Tong, P., and Fung, Y.C. 1976. *J. Biomech.* 9:649–657.
- Vaishnav, R.N. 1980. *Biorheology* 17:219–226.
- Vaishnav, R.N., Young, J.T., Janicki, J.S., and Patel, D.J. 1972. *Biophys. J.* 12:1008–1027.
- Vito, R. 1973. *J. Biomech.* 6:561–564.
- Vito, R.P., and Hickey, J. 1980. *J. Biomech.* 13:951–957.
- Wiegner, A.W., and Bing, O.H.L. 1981. *Circ. Res.* 49:807–814.
- Yin, F.C.P., Tompkins, W.R., Peterson, K.L., and Intaglietta, M. 1972. *IEEE. Trans. Biomed. Eng.* BME-19:376–381.

# 8

# *In-Vivo Measurement of Regional Strains in Myocardium*

L.K. WALDMAN

## 1. Introduction

The motion of the heart wall and its relationship to cardiac function have been the subject of intensive investigation for many years. While large rigid body motions (translations and rotations) have been observed in myocardium, it is the relative motion (i.e., the deformation or strain) that may best be associated with the muscle's contractile abilities. These strains accompany the large internal stresses and forces that react against the blood in order to raise its pressure and eject it from the chamber. During contraction, the normal left ventricular wall undergoes large and complex patterns of deformation that have been partly characterized by both global and local dimensional measurements. The motion of the left ventricle is often identified with the shape changes of an axisymmetric membrane or shell. Thus, early studies in cardiac mechanics focused on variations in global cardiac dimensions such as chamber volumes and areas, major and minor semi-axes, and circumferential shortening and shortening velocities as measured by changes in ventricular radius (Rushmer et al, 1956; Hawthorne, 1961; Mitchell et al, 1969; Rankin et al, 1976).

While the geometry of the epicardial surface might suggest that cardiac dynamics can be understood in terms of simple model geometries, the complex internal structure of this thick-walled organ belies such simplifications. The left ventricle is formed from a syncitium of muscle fibers wound in a spiral about the cavity (Torrent-Guasp, 1973; Streeter et al, 1969). The directions of these fibers projected on the epicardial tangent plane have been shown to vary continuously from epicardium to endocardium by as much as 100 degrees or more (Streeter, 1979). In addition, the fibers are embedded in a complex weave of heavily cross-linked collagen (McClain, 1974; Medugorac, 1982; Borg and Caulfield, 1981). These structural complexities are accompanied by non-uniform function. Thus, recently, more localized measurements have revealed substantial variation in cardiac deformation and regional function at different locations in normal myocardium (Ingels et al, 1971, 1975, 1981; LeWinter et al, 1975; Lew and LeWinter, 1983; Freeman et al, 1985; Myers et al, 1984; Lew and

LeWinter, 1984). Alterations in cardiac function under conditions of ischemia (Tennant and Wiggers, 1935; Theroux et al, 1977; Edwards et al, 1981; Gallagher et al, 1982a, b; Hess et al, 1983; Osakada et al, 1983) and hypertrophy (Sasayama et al, 1977; Carew and Covell, 1978; Badke and Covell, 1979; LeWinter et al, 1980; Sasayama et al, 1981) have been correlated with changes in these deformation patterns. Indeed, the "holosystolic bulge" observed in ischemia, the subendocardial localization of infarcts, and the redistribution of transmural blood flow under pathological conditions provide further motivation to fully quantify the accompanying patterns of local deformation. In the present article, we focus on regional measurement of strain or deformation. After discussing prior measurements of segmental shortening, wall thickening, shear angles, and shear displacements, concepts from continuum mechanics are reviewed, with particular emphasis on the kinematics of finite deformation; the algorithms needed to calculate two-dimensional finite strains from measured coordinate data are discussed in detail; and the three-dimensional generalizations needed to examine the deformation of muscle volumes are outlined briefly (Waldman et al, 1985). Subsequently, examples of *in vivo* three-dimensional finite strains are displayed and discussed in the context of previous measurements.

Regional myocardial deformation during systole has frequently been quantified with techniques that measure uniaxial length changes at one or more sites in the ventricle. However, little information is available concerning variations in function at different depths beneath a given epicardial site. Considering the large thickness-to-radius ratio of the left ventricle and its complex fiber-winding pattern, such variations might be expected. Ultrasonic dimension gauges have been used to quantify segmental shortening at a number of ventricular sites (LeWinter et al, 1975; Freeman et al, 1985; Lew and LeWinter, 1984). These studies indicate that endocardial strain exceeds epicardial strain and that shortening may vary at different ventricular sites. A few attempts have been made to measure average shearing motions (Feigl and Fry, 1964a; Osakada et al, 1980) and variations in transmural thickening (Myers et al, 1984). These studies have revealed small average systolic shear and wall thickening that is greater near the endocardium than the epicardium. In the studies of Dieudonne (1969a, b) a triad of strain gauges (strain rosette) and attached pins inserted into the myocardium was used to measure two-dimensional, infinitesimal strains at a number of depths beneath the epicardium and in a number of positions on the left ventricle. His results indicate greater shortening strain at increasing depths beneath the epicardium. Apparently, Dieudonne was the first to calculate two-dimensional principal strains from his strain data. In another method, needles were placed across the heart wall; their relative motions were followed with an electromagnetic inductive technique in an effort to elucidate transmural deformation (Prinzen et al, 1984). These data are interpreted to show that endocardial shortening can be predicted from epicardial shortening. Yet another method has been to place radiopaque markers in the myocardium and to track their motion with high-

speed biplane cineradiography. With this technique, two-dimensional finite deformation has been quantified on the epicardium (Meier et al, 1980, 1982), and transmural deformation gradients have been measured from columns of markers implanted across the left ventricular free wall (Fenton et al, 1976, 1978).

However, these methods have serious disadvantages that may preclude their use in elucidating the transmural variation of function even in normal myocardium. Despite high temporal and spatial resolution, measurements made with ultrasonic dimension gauges depend on the orientation of the devices. In order to remove the arbitrariness with which pairs of dimension gauges are positioned, more elaborate arrays of these devices are required to measure shortening, thickening, and shearing motions simultaneously. Even if this were possible, it is doubtful that enough crystals could be implanted locally to reveal transmural variations in deformation without miniaturization. Electrical leads (wires) attached to these devices always present difficulties. Both the needle implantation method (Prinzen et al, 1984) and the strain rosette (Dieudonne, 1969a, b) are incapable of measuring thickening and transverse shearing motions that accompany shortening. Furthermore, the needles and the pins attached to the rosette may constrain the surrounding muscle to move without any allowance for localized in-plane or transverse shearing and without allowing for the possibility that the direction of principal shortening varies with depth. Also, the pins may measure only average strains from the epicardium to a given depth. Moreover, assumptions concerning the relationship between motion on the epicardial surface and transmural deformation may be inappropriate and can only be tested through a thorough study of transmural deformation. While Fenton and co-workers (1976, 1978) undertook such a study and successfully implanted columns of markers, the deformation gradient tensor they calculated from their coordinate data is non-symmetric and was obtained in a somewhat arbitrary coordinate system. The lack of symmetry forced the retention of nine arbitrary components instead of the six needed for a complete description and precluded a calculation of eigenvalues and eigenvectors (i.e., the principal strains and axes of deformation that quantify local deformation most simply and in a manner independent of the reference coordinates). Thus, none of these methods is capable of measuring arbitrary three-dimensional finite deformations at a number of locations across the wall thickness.

In recent studies in this laboratory (Waldman et al, 1985), columns of lead markers have been implanted in a single ventricular site as done previously (Fenton et al, 1976, 1978). However, a new approach is used to calculate strains from coordinate data whereby finite strains, principal strains, and the directions of the principal axes of deformation can be computed. The results of studies in normal dogs indicate that substantial in-plane and transverse shears often accompany normal strains measured in a cardiac coordinate system (Meier et al, 1980). The strains and their principal values tend to increase with depth beneath the epicardium. Furthermore, while the orientation of the

principal axis of shortening does vary somewhat with depth, it does not vary nearly as much as the corresponding fiber angle does (Streeter et al, 1969), suggesting that anatomic links between myofibers may be an important mechanism in transmural deformation.

## 2. Strain Analysis

In the experiments (Waldman et al, 1985) mongrel dogs are anesthetized with pentobarbital. A fourth or fifth interspace bilateral thoracotomy is performed. The heart is placed in a pericardial cradle, and columns of four to six 1-mm lead beads are implanted in the anterior free wall of the left ventricle with an intercolumn spacing of 1 cm or less. High-speed biplane cineradiography (16 mm, 120 frames/s) is performed in order to visualize the radiopaque markers. The film of each two-dimensional view (lateral and anterior-posterior) is projected onto a digitizing pad (Houston Instruments: HIPAD) in order to obtain the coordinates of the centroids of the marker shadows. Subsequently, the three-dimensional coordinates are reconstructed for strain analysis. The coordinate data of any four noncoplanar markers are used to calculate a set of six independent finite strains by either a new method that uses all six edges of a tetrahedron, or by a three-dimensional generalization of Meier's surface method, which uses any three edges. The results are identical. These strains are calculated with respect to a Cartesian cardiac coordinate system at end diastole (Meier et al, 1980). In Figure 8.1 these reference coordinates are identified. In the remainder of this section, the kinematics of finite deformation are reviewed (Truesdell, 1966; Fung, 1965). Details of finite strain calculations by both methods are described in the two-dimensional case. The three-dimensional generalizations of these methods are outlined.

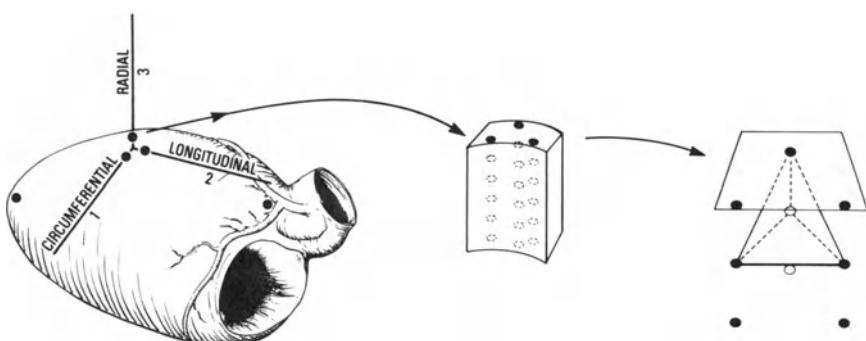


FIGURE 8.1. A schematic diagram of the left ventricle showing the five reference markers used to calculate a cardiac coordinate system. In addition, the locations of three columns of lead markers are shown; a typical tetrahedron formed from four of the markers is identified.

Finally, questions concerning material compressibility, kinematic nonlinearity, coordinate transformation, and homogeneity are addressed.

## 2.1. The Kinematics of Finite Deformation

Consider a material point  $P$  in the myocardium that is identified by the position vector of the centroid of an implanted lead marker (Figure 8.1), with coordinate components  $a_i$  in a reference coordinate system (Figure 8.2). At some later time, the same material point  $P$  may move to a new location,  $Q$ , and is now identified by the new position vector with components  $x_i$ , in the same reference coordinate system. Components of the displacement vector,  $u_i$ , indicate the translation of the material point from  $P$  to  $Q$ . Thus,

$$x_i = x_i(a_i) = u_i + a_i. \quad (8.1)$$

In the one-dimensional case that has been applied in many unidirectional studies of myocardial deformation, we have  $x_1 = x_1(a_1)$ . By the chain rule for functions of a single variable, an ordinary derivative suffices to map, say,  $da_1$ , to  $du_1$ ; i.e.,

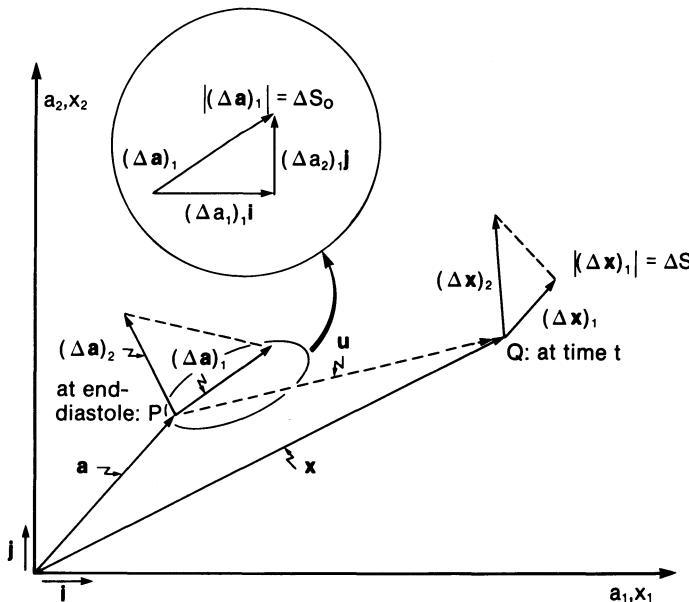


FIGURE 8.2. The vector geometry of two-dimensional deformation. A material point  $P$  identified by the vector  $\mathbf{a}$  in a reference coordinate system (at end diastole) undergoes the displacement  $\mathbf{u}$  and is identified by the vector  $\mathbf{x}$  at a later time. Small but finite vectors (directed material line segments)  $\Delta\mathbf{a}$  and  $\Delta\mathbf{x}$  are associated with the local continuum in the “original” and “deformed” configurations, respectively.

$$du_1 = \frac{du_1}{da_1} da_1.$$

For unidirectional motions, this type of mapping might suffice, but for arbitrary three-dimensional motions in which  $x_i$  is a function of three variables ( $x_i = x_i(a_1, a_2, a_3)$ ), the chain rule for partial derivatives must be introduced. We belabor this point because it provides motivation for our first use of the concept of tensors. Writing the derivative in finite form, an idea that is very useful in the three-dimensional theory, as we shall soon see, we have (without subscripts):

$$\frac{du}{da} \approx \frac{\Delta u}{\Delta a} = \frac{l - l_0}{l_0}$$

where  $l_0$  is the initial length and  $l$  is the length at the time in question. Note that

$$\frac{l - l_0}{l_0}$$

is the familiar fractional shortening commonly employed to describe ventricular function.

In the higher dimensional case, we have (with the summation convention)

$$dx_i = \frac{\partial x_i}{\partial a_j} da_j = F_{ij} da_j \quad (8.2)$$

and

$$du_i = \frac{\partial u_i}{\partial a_j} da_j = G_{ij} da_j. \quad (8.3)$$

Here, the derivative of a vector with respect to another vector yields a tensor mapping such as  $F_{ij}$ , the displacement gradient tensor, or  $G_{ij}$ , the deformation gradient tensor. In general, both  $F_{ij}$  and  $G_{ij}$  are nonsymmetric so that they have nine independent components in 3-space. In mechanics the directed infinitesimal line segments,  $da_i$ ,  $dx_i$ , and  $du_i$ , are called contravariant vectors or tensors of rank 1. In the mathematics of modern differential geometry, they are most properly called differential forms, and in this case, one-forms (Flanders, 1963). Note that a tensor is a generalization of a vector, i.e., we need three numbers to identify an arbitrary vector in 3-space, while nine numbers are needed to identify a tensor of rank 2 in the same space. In finite deformation mechanics, a careful distinction must be made between material configurations at different times due to the large relative motions. Thus, for example,

$$\frac{\partial u_1}{\partial a_1} \neq \frac{\partial u_1}{\partial x_1},$$

and we are careful to identify mappings made with respect to the starting configuration (e.g.,  $\partial/\partial a_j$ ). Only in the infinitesimal theory are these mappings

interchangeable, but as we shall soon see, the infinitesimal approach is inadequate in cardiac mechanics.

Differentiating Equation (8.1) with respect to the vector  $a_i$  yields the simple relationship between the displacement gradient and deformation gradient tensors, i.e.,

$$\frac{\partial x_i}{\partial a_j} = \frac{\partial u_i}{\partial a_j} + I_{ij} \quad (8.4)$$

or

$$F_{ij} = G_{ij} + I_{ij}$$

so that the off-diagonal terms of both tensors are identical while the diagonal elements differ by the additive constant one.  $I_{ij}$  is the kronecker symbol.

Now, an arbitrary infinitesimal length can always be computed from its coordinate components, so that in the starting configuration we have (Figure 8.2)

$$ds_0^2 = da_i da_i, \quad (8.5)$$

and at a later time we have

$$ds^2 = dx_i dx_i. \quad (8.6)$$

Substituting Equation (8.2) into Equation (8.6) and subtracting Equation (8.5) from the results yields the classical quadratic form that defines the strain tensor—i.e.,

$$ds^2 - ds_0^2 = 2E_{ij} da_i da_j, \quad (8.7)$$

in which

$$E_{ij} = E_{ij}(a_i, t) = \frac{1}{2}(F_{ki} F_{kj} - I_{ij}). \quad (8.8)$$

Here, the symmetric strain tensor,  $E_{ij}$ , is related to the displacement gradient tensor,  $F_{ij}$ . Substituting Equation (8.4) into Equation (8.8) relates  $E_{ij}$  to the deformation gradient tensor,  $G_{ij}$ , showing us how to convert the measurements of Fenton and co-workers (1976, 1978) to finite strains—i.e.,

$$E_{ij} = \frac{1}{2}(G_{ij} + G_{ji} + G_{ki} G_{kj}). \quad (8.9)$$

On surfaces (two-dimensional theory) the strain tensor  $E_{ij}$  consists of three independent components—two normal strains and one shear strain. In a cardiac coordinate system (Meier, 1980), circumferential strain ( $E_{11}$ ), longitudinal strain ( $E_{22}$ ), and shear in the plane of the epicardium ( $E_{12}$ ) may occur. However, for a three-dimensional continuum, six independent components may exist. Thus, radial strain ( $E_{33}$ ) and two transverse shear strains ( $E_{13}$  and  $E_{23}$ ) may accompany the surface strains. The most complex deformation at a given point can always be decomposed into three relative length changes and three angle changes. These dimensional changes are easily related to the finite strains described above.

## 2.2 Details of Two-Dimensional Strain Calculations Using Displacement Gradients

Meier and co-workers (1980) computed stretch ratios directly from a finite version of Equation (8.2). That is, they wrote Equation (8.2) in the form

$$\Delta x_i = F_{ij} \Delta a_j; \quad (8.10)$$

and with a knowledge of  $\Delta a_i$  in some starting configuration and  $\Delta x_i$  in a later configuration, they compute the surface tensor,  $F_{ij}$ , by solving two  $2 \times 2$  linear algebra problems from the two directed line segments connecting three points (surface markers) on the epicardium (Figure 8.2). Here, the displacement gradients can be written explicitly in terms of the components of the differential vectors as follows:

$$F_{11} = [(\Delta x_1)_1 (\Delta a_2)_2 - (\Delta x_1)_2 (\Delta a_2)_1] / \Phi$$

$$F_{12} = [(\Delta x_1)_2 (\Delta a_1)_1 - (\Delta x_1)_1 (\Delta a_1)_2] / \Phi$$

$$F_{21} = [(\Delta x_2)_1 (\Delta a_2)_2 - (\Delta x_2)_2 (\Delta a_2)_1] / \Phi$$

$$F_{22} = [(\Delta x_2)_2 (\Delta a_1)_1 - (\Delta x_2)_1 (\Delta a_1)_2] / \Phi$$

in which

$$\Phi = (\Delta a_1)_1 (\Delta a_2)_2 - (\Delta a_1)_2 (\Delta a_2)_1.$$

Once they have  $F_{ij}$  they decompose it into a rotation and a stretch (Truesdell, 1966) and compute the principal values of the stretch matrix which, of course, is symmetric. Alternatively, the finite strains can be related to the deformation gradients, i.e., by premultiplying  $F_{ij}$  by its transpose,  $F_{ij}$ , the strain tensor,  $E_{ij}$ , can be computed directly from Equation (8.8). This method is easily generalized to three dimensions in a volume rather than two dimensions on a surface by adding a fourth nearby (but noncoplanar) marker (point). Then, three  $3 \times 3$  linear algebra problems must be solved to compute  $F_{ij}$ , and again Equation (8.8) yields  $E_{ij}$ .

## 2.3. Details of Two-Dimensional Strain Calculations by a Direct Method

An alternative method computes  $E_{ij}$  directly from Equation (8.7) (Waldman et al, 1985). This technique is elaborated in the two-dimensional case for clarity. Here, all three line segments are utilized (Figure 8.2). First, we write Equation (8.7) in finite form,

$$\begin{aligned} \Delta s^2 - \Delta s_0^2 &= 2E_{ij}\Delta a_i\Delta a_j, \\ &= 2(E_{11}\Delta a_1^2 + 2E_{12}\Delta a_1\Delta a_2 + E_{22}\Delta a_2^2). \end{aligned} \quad (8.7)'$$

Equation (8.7)' can be rewritten as a  $3 \times 3$  linear algebra problem in the standard form

$$[A]\{x\} = \{b\}, \quad (8.11)$$

in which

$$[A] = \begin{bmatrix} (2\Delta a_1^2)_1(4\Delta a_1 \Delta a_2)_1(2\Delta a_2^2)_1 \\ (2\Delta a_1^2)_2(4\Delta a_1 \Delta a_2)_2(2\Delta a_2^2)_2 \\ (2\Delta a_1^2)_3(4\Delta a_1 \Delta a_2)_3(2\Delta a_2^2)_3 \end{bmatrix},$$

$$\{x\} = \begin{Bmatrix} E_{11} \\ E_{12} \\ E_{22} \end{Bmatrix}, \quad \{b\} = \begin{Bmatrix} (\Delta s^2 - \Delta s_0^2)_1 \\ (\Delta s^2 - \Delta s_0^2)_2 \\ (\Delta s^2 - \Delta s_0^2)_3 \end{Bmatrix}.$$

When generalized to three dimensions, the matrix  $[A]$  becomes a  $6 \times 6$  square matrix, the vector  $\{b\}$  becomes a  $6 \times 1$  column vector, and the six independent strains are written as a  $6 \times 1$  column vector instead of a  $3 \times 3$  symmetric strain tensor.

After obtaining the strain tensor  $E_{ij}$ , an algebraic eigenvalue problem is solved to compute the principal strains and axes of deformation. With the three principal strains,  $E_i$ , in hand, the stretch ratios,  $\lambda_i$ , are easily calculated (Fung, 1967), i.e.,

$$\lambda_i = \sqrt{2E_i + 1}.$$

If the tetrahedron of muscle in question does not change volume during deformation, then

$$\lambda_1 \lambda_2 \lambda_3 = 1.$$

On the other hand, for a compressible material, the product of the stretch ratios during deformation is less than one, e.g., a value of 0.9 indicates a volume reduction of 10%.

## 2.4. Kinematic Nonlinearity in Finite Deformation

Finite strains are kinematically nonlinear. For example, Equation (8.9) relates the finite radial strain to elements of the deformation gradient tensor as follows:

$$E_{33} = \frac{\partial u_3}{\partial a_3} + \frac{1}{2} \left[ \left( \frac{\partial u_1}{\partial a_3} \right)^2 + \left( \frac{\partial u_2}{\partial a_3} \right)^2 + \left( \frac{\partial u_3}{\partial a_3} \right)^2 \right],$$

so that an infinitesimal radial strain having a value of +0.4 is in error by at least -16% and probably more in the presence of transverse shear. We see that finite radial strain is equivalent to infinitesimal radial strain augmented by three terms, one of which is the infinitesimal strain squared, while the other two are actually contributions from the shearing elements of the deformation gradient tensor. Notice that the nonlinear terms in this equation are quadratic (always positive), so that positive normal and principal strains are always underestimated, while negative ones are always overestimated by the in-

infinitesimal theory. This rule does not necessarily apply to shear strains, for which nonlinear terms are products of different elements of the deformation gradient tensor; e.g., one of the finite transverse shears is

$$E_{23} = \frac{1}{2} \left[ \frac{\partial u_2}{\partial a_3} + \frac{\partial u_3}{\partial a_2} + \left( \frac{\partial u_1}{\partial a_2} \frac{\partial u_1}{\partial a_3} + \frac{\partial u_2}{\partial a_2} \frac{\partial u_2}{\partial a_3} + \frac{\partial u_3}{\partial a_2} \frac{\partial u_3}{\partial a_3} \right) \right].$$

Here, several of the diagonal components of the deformation gradient tensor that are normal strains in the infinitesimal theory contribute to finite shear.

## 2.5. Coordinate Transformation and Material Homogeneity

As mentioned previously,  $da_i$  is a contravariant vector; and as such, it obeys the transformation laws of tensors of rank 1 (also called vectors). Therefore, rather than transforming the tensor  $E_{ij}$  to the cardiac coordinate system after computing the strains, we can transform  $da_i$  to cardiac coordinates first. Then the strains are automatically calculated with respect to cardiac coordinates. Given the transformation matrix,  $\beta_{ij}$ , and again writing our differential elements in finite form, we have

$$\Delta a'_i = \beta_{ij} \Delta a_j. \quad (8.12)$$

A brief comment on our use of assumptions concerning spatial homogeneity is in order. Homogeneity is actually assumed in two ways. First, along any small material length we assume homogeneity as soon as we write the strain equations (e.g. Equation (8.7)') in finite rather than infinitesimal form. Second, when we use three or six small line segments to compute a single strain tensor in a small area or volume, homogeneity is assumed within that area or volume.

## 3. Examples of *In-Vivo* Three-Dimensional Finite Strains

In Figures 8.3 through 8.7, typical finite strains observed in canine myocardium are displayed. To illustrate time series of finite strains occurring during two consecutive contractions in a subendocardial tetrahedron (centroid depth of 9.5 mm beneath the local epicardium), the normal strains (Figure 8.3), shear strains (Figure 8.4), and accompanying principal strains (Figure 8.5) are shown. These data have been smoothed with a three-point moving average to eliminate some of the digitizing error. Each time series is initialized at end diastole, as determined from a simultaneous pressure tracing. Notice the repeatability of these measurements from one contraction to the next. Circumferential strain ( $E_{11}$ ) typically reaches systolic values of  $-0.20$  at this depth, while corresponding radial strains ( $E_{33}$ ) may peak as high as  $+0.40$  (Figure 8.3). Significant shear strains may often accompany normal strains (Figure 8.4). Here, transverse shear at end systole is  $0.10$  or greater. Once the normal and shear strains are computed, the principal strains are easily calculated (Figure 8.5). Observe the similarity of these results to the normal strains.

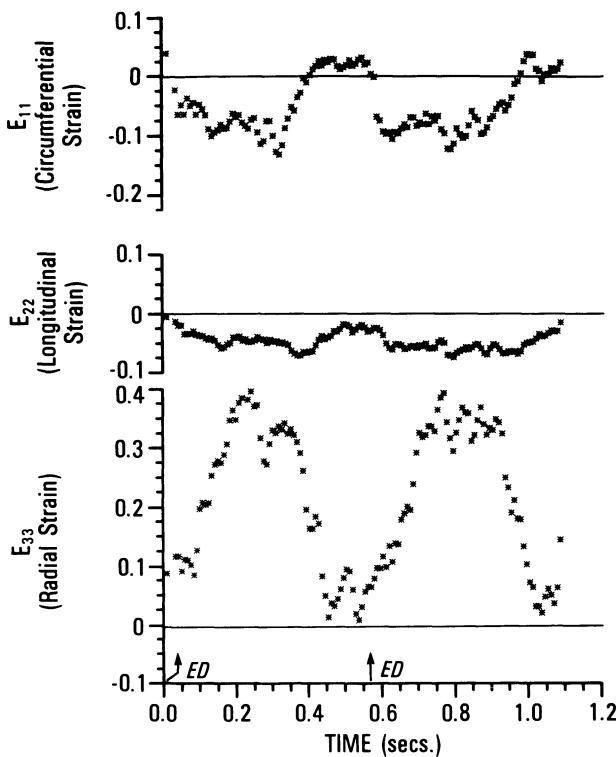


FIGURE 8.3. The normal strains as a function of time during two consecutive contractions from a sub-endocardial tetrahedron whose centroid is located at a depth of 9.5 mm. beneath the local epicardium.  $ED$  = end diastole.

Although this is often the case, it should be remembered that the principal strains may be acting in directions that are considerably different from the normal strains. Moreover, the principal strains of shortening and thickening are usually greater than the normal strains because they reflect the presence of and are augmented by shear. Given the principal strains, the product of the stretch ratios can be calculated (see Section 2.3). This function is usually between 0.9 and 1.0, indicating slight compressibility of these small muscle volumes. In Figure 8.6, end systolic principal strains are plotted as a function of depth from the local epicardium by selecting a number of tetrahedra across the wall thickness for analysis. Observe the linearly decreasing trend of principal shortening and the monotonically increasing trend of principal thickening with depth. These findings have been consistent at similar ventricular sites in a number of studies (Waldman et al, 1985). Along with the principal strains, we compute eigenvectors that indicate the directions of the principal axes of deformation with respect to the cardiac coordinates. In Figure 8.7, the angle between the direction of principal shortening and the circumferential coordi-

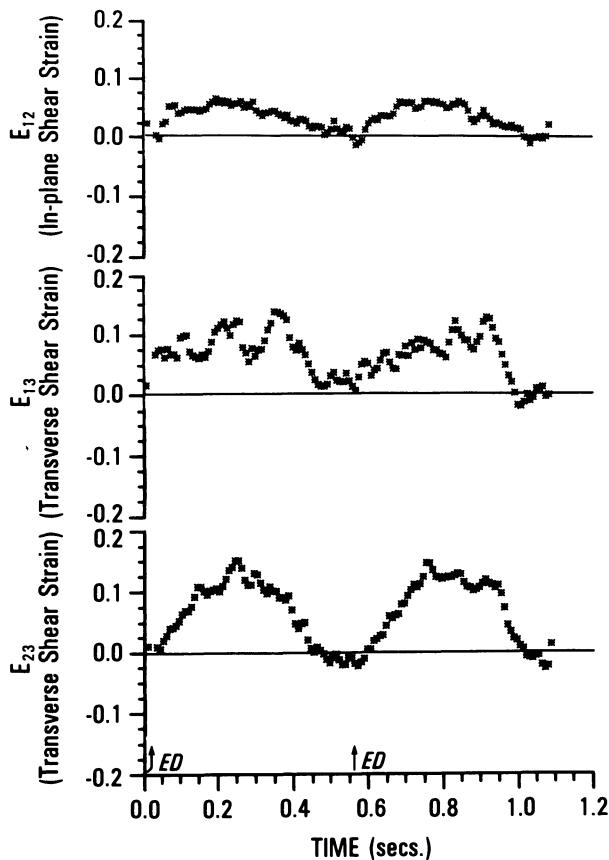


FIGURE 8.4. Shear strains that accompany the normal strains shown in Figure 8.3.

nate axis projected on the epicardial tangent plane is shown as a function of depth. Notice that this angle only changes by about 30 degrees.

#### 4. Discussion

Aspects of the strains calculated in our recent studies (Waldman et al, 1985) may be compared with previously acquired measurements. In particular, the substantial increase in shortening at midwall over that observed near the epicardium (Dieudonne, 1969a, b; LeWinter et al, 1975) is verified by both our normal strain and principal shortening data. Furthermore, evidence for the presence of circumferential strain near the epicardium (Freeman et al, 1985) where fibers have a more longitudinal orientation (Streeter et al, 1969) is corroborated. However, circumferential strain does not always dominate longitudinal strain at any location. In fact, our results can be split into two

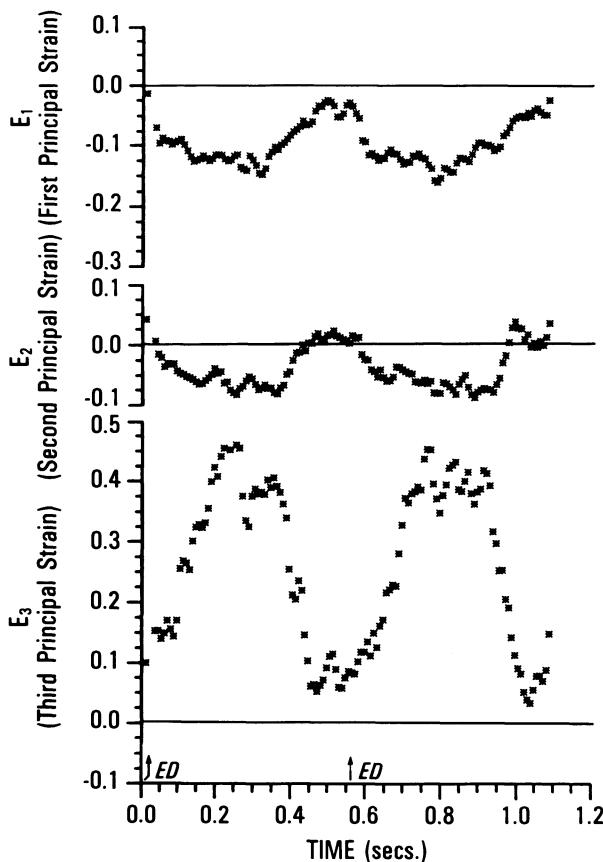


FIGURE 8.5. Principal strains as a function of time during two consecutive contractions. These strains were calculated from the normal and shear strains appearing in Figures 8.3 and 8.4, respectively.

groups. In the first group, hoop shortening dominates across the entire wall thickness; while in the second, longitudinal shortening may be comparable to hoop shortening everywhere. These findings provide strong motivation to calculate principal shortening, which is much more consistent and which is uninfluenced by the orientation of the reference coordinates.

Because the heart wall ejects blood by thickening a great deal as it shortens, detailed information concerning the transmural variation of thickening is relevant. Average wall thickening measured previously (Feigl and Fry, 1964b; Ross et al, 1967; Gallagher et al, 1980; Osakada et al, 1980) is comparable to our radial strains at midwall ( $\approx 0.20$ ), but radial strain near the endocardium may be larger ( $\approx 0.30$ ) and the corresponding principal thickening even larger still ( $\approx 0.40$ ). Again, the need to measure principal data is emphasized, and in this case (thickening), the strains are so large that nonlinear effects must be

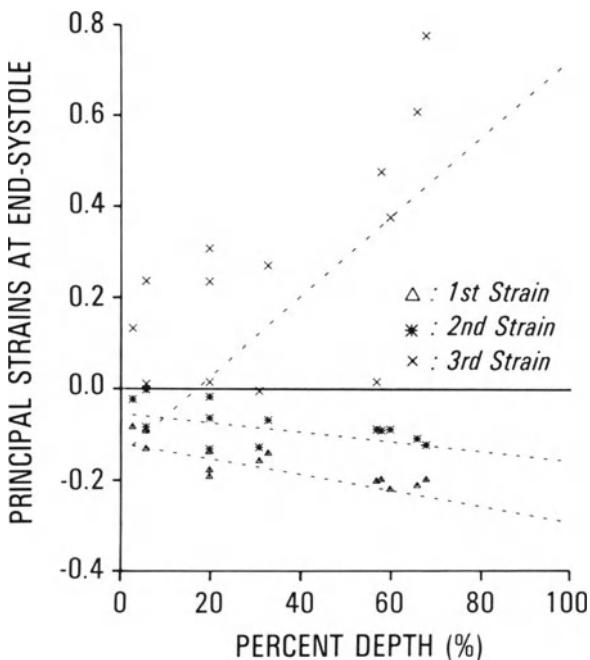


FIGURE 8.6. End systolic principal strains as a function of percent depth from the epicardium (100% is the endocardium).

included (see the discussion of kinematic nonlinearity under Strain Analysis). While ultrasonic dimension gauges have measured greater average wall thickening in the endocardial half than in the epicardial half of the wall (Myers et al, 1984), our recent results indicate a substantial variation in radial strain as well as principal thickening with depth. In fact, in some preparations, virtually no thickening is observed near the epicardium, while strains in the range of +0.4 to +0.5 occur near the endocardium. Perhaps the only other method capable of acquiring these data is gated pulsed doppler, but if the third principal axis ever tips away from the beam direction (surface normal) at all, the doppler method will underestimate the full principal thickening at that depth. Furthermore, the doppler method may only measure average thickening from the epicardium to a given depth.

Attempts have been made to measure average transverse shears across the heart wall (Osakada et al, 1980) and at two depths beneath the epicardium (Feigl and Fry, 1964a). In the latter study, the two shear angles measured were the two components of transverse shear. Moreover, referring all strains measured at a given time to the same starting configuration (say, end diastole), the measurements of Feigl and Fry indicate an angle change of 0.05 radians during isovolumetric contraction and that this shear is *maintained* during ejection. Nevertheless, this shear is small ( $e_{\text{shear}} \approx \frac{1}{2} \tan 0.05 = 2-3\%$ ). In the study of

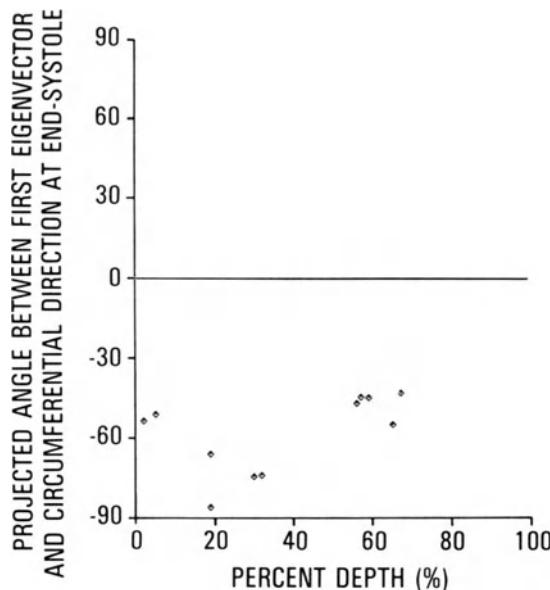


FIGURE 8.7. The angle between the principal axis of shortening at end systole and the circumferential coordinate direction projected on the epicardial tangent plane as a function of percent depth.

Osakada and co-workers, the average transmural shear displacement in the 2–3 plane was measured. This measurement can be converted to a shear angle and subsequently to an infinitesimal transverse shear by dividing by the thickness. For a longitudinal displacement of 0.6 mm over a wall thickness of 9.6 mm, we have

$$e_{23} \approx \frac{1}{2} \tan \frac{0.6}{9.6} = 3\%.$$

Again, a small shear strain results. Finite transverse shears vary substantially across the heart wall (Waldman et al, 1985). In some cases shears change sign with increasing depth, so that a transmural average might be negligible, masking the presence of local shears. Frequently, epicardial transverse shear strains are found to be small (as are the other strain components)—in agreement with Feigl and Fry—but transverse shears at midwall and deeper are often substantial ( $\approx 0.10$ ) and sometimes comparable to the shortening strains there. Excluding nonlinear effects, a shear strain of 0.10 is equivalent to a shear angle of 0.20 radians, which is between 4 and 10 times the magnitudes measured previously. Transverse shear may be evidence for the contribution of the radial component of fiber direction (or pitch angle) to deformation (i.e., the nonconcentric geometry of overlying fibers and the interaction between layers due to tethering may result in deformation patterns that are truly three-

dimensional). This observation argues against the use of axisymmetric membrane or shell hypotheses in the analysis of local wall motion and regional function.

By definition, the presence of shear indicates that the initial coordinate system (in our case, the cardiac coordinate system at end diastole) is not canonical. Despite the physiological relevance of these coordinates, they are still somewhat arbitrary. First, due to the possibility of residual "active" tone at end diastole and the possible presence of initial passive stresses at that end diastolic pressure (external loading), the end diastolic configuration may be substantially strained. In fact, even if active tone and the transmural pressure were negligible at end diastole, the myocardium may have evolved to be a continuum with residual stresses (see Chapter 9). Therefore, the strains we calculate might best be called relative strains rather than, say, Lagrangian or Green's strains except for the fact that relative tensors are already defined differently in the mathematical literature. Recall that Lagrangian implies an original configuration that is unstrained, and that the continuum returns to this configuration in the absence of loading. Another possible reason why some of the shear strains may be artifactual is that any nonconcentricity in the arrangement of muscle fibers at increasing depths beneath the epicardium may make the orientation of the chosen epicardial tangent plane less meaningful near the endocardium of this thick-walled structure. Perhaps, a midwall reference system would be more meaningful. In spite of these drawbacks, substantial shears have sometimes been measured even near the epicardium. Inherent arbitrariness in any set of starting coordinates provides strong motivation to calculate eigenvalues and eigenvectors (canonical results), but the possibility that even principal strains at end diastole may not be negligible cannot be overlooked.

Principal strains and their accompanying axes provide a complete description of local deformation at a given time and location. While time series of the strain components for one or two heartbeats are presently acquired only with great labor, end systolic trends are obtained easily, since only end diastolic and end systolic coordinate data are needed. The transmural trends in principal strains at end systole, as mentioned in the results, are consistent and indicate that normal myocardium deforms more at increasing depths beneath the epicardium. The linearity of transmural variation of the principal strain of shortening ( $E_1$ ) is a striking result. The interpretation of the principal axes results is less obvious. Why does the first principal direction vary so little across the heart wall, and why is it not necessarily circumferential at midwall where the fiber direction probably is? One possible explanation is that because significant shortening and thickening do occur epicardially to midwall, a substantial volume of muscle above midwall can be expected to contract along a direction that is -40 to -60 degrees from circumferential in the plane of the epicardium. Therefore, if this muscle acts in concert with deeper muscle due to mutual attachments (tethering), we might expect shortening to occur along directions that are vector sums of adjacent fiber directions. An important

extension of our current protocol will be to acquire fiber angle data in the same preparations in which strains are measured, i.e., fiber angles in the implantation region.

The fact that the product of the stretch ratios is close to one, but that it is often slightly less during systole (albeit often within measuring accuracy of one), indicates the possibility of some loss of coronary blood volume during contraction, although the possibility that the muscle itself is somewhat compressible cannot be ruled out. Improved calibration, along with smaller markers, should provide further support for our findings concerning transmural variations in wall thickening, transverse shear, and slight compressibility. Finally, additional studies are needed in which the implantation region is enlarged and moved. Markers must be implanted in a variety of positions in the left ventricle to study the possible spatial variation of strain. Larger tetrahedra will be needed to establish the extent of "local" homogeneity and to provide adequate tissue volumes for blood flow measurement (radioactive microspheres) in an ischemic protocol.

## Acknowledgments

I gratefully acknowledge Professor Y.C. Fung for his simple yet brilliant proposal at the beginning of the transmural deformation project and for suggesting my participation. I thank Professor J.W. Covell for his interest and support in this project. I would also like to acknowledge the assistance of Charles M. Thrasher in technical design, and I thank Richard S. Pavelec for surgical expertise and technical assistance.

This project is supported by the following grants: Transmural Deformation PHS HL 32583, Ischemic Heart Disease SCOR PHS HL 17682, and Research Training in Cardiovascular Physiology and Pharmacology PHS HL 07444.

## References

- Badke, F.R., and Covell, J.W. 1979. *Circ. Res.* 45:420–428.
- Borg, T.K., and Caulfield, J.B. 1981. *Fed. Proc.* 40:2037–2041.
- Carew, T.E., and Covell, J.W. 1978. *Am. J. Cardiol.* 42:82–88.
- Dieudonne, J.M. 1969a. *J. Physiol. (Paris)* 61:199–218.
- \_\_\_\_\_. 1969b. *J. Physiol. (Paris)* 61:305–330.
- Edwards, C.H., Rankin, J.S., McHale, P.A., Ling, D., and Anderson, R.W. 1981. *Am. J. Physiol.* 240:H413–H420.
- Feigl, E.O., and Fry, D.L. 1964a. *Circ. Res.* 14:536–540.
- \_\_\_\_\_. 1964b. *Circ. Res.* 14:541–545.
- Fenton, T.R., Klassen, G.A., and Outerbridge, J.S. 1976. *Intern. Conf. Med. Biol. Eng.* 714–715.
- Fenton, T.R., Cherry, J.M., and Klassen, G.A. 1978. *Am. J. Physiol.* 235:H523–H530.

- Flanders, H. 1963. *Differential forms with applications to the physical sciences*. Academic Press, New York.
- Freeman, G., LeWinter, M.M., Engler, R.E., and Covell, J.W. 1985. *Circ. Res.* 56:31–39.
- Fung, Y.C. 1965. *Foundations of solid mechanics*. Prentice-Hall, Englewood Cliffs, N.J.
- \_\_\_\_\_. 1967. *Am. J. Physiol.* 213:1532–1544.
- Gallagher, K.P., Kumada, T., Koziol, J.A., McKown, M.D., Kemper, W.S., and Ross, J.Jr. 1980. *Circulation* 62:1266–1274.
- Gallagher, K.P., Osakada, G., Hess, O.M., Koziol, J.A., Kemper, W.S., and Ross, J.Jr. 1982a. *Circ. Res.* 50:352–359.
- Gallagher, K.P., Osakada, G., Matsuzaki, M., Kemper, W.S., and Ross, J.Jr. 1982b. *Am. J. Physiol.* 242(5):H698–H707.
- Hawthorne, E.W. 1961. *Circ. Res.* 9:110–119.
- Hess, O.M., Osakada, G., Lavelle, J.F., Gallagher, K.P., Kemper, W.S., and Ross, J.Jr. 1983. *Circ. Res.* 52:387–400.
- Ingels, N.B., Jr., Daughters, G.T.II, and Davies, S.R. 1971. *J. Biomech.* 4:541–550.
- Ingels, N.B., Jr., Daughters, G.T.II, Stinson, E.B., and Alderman, E.L. 1975. *Circulation* 52:859–867.
- \_\_\_\_\_. 1981. *J. Biomech.* 14(4):221–233.
- Lew, W., and LeWinter, M.M. 1983. *Am. J. Physiol.* 245:H741–H748.
- \_\_\_\_\_. 1984. (Submitted for publication).
- LeWinter, M.M., Kent, R.S., Kroener, J.M., Carew, T.E., and Covell, J.W. 1975. *Circ. Res.* 37:191–199.
- LeWinter, M.M., Engler, R.L., and Karliner, J.S. 1980. *Am. J. Physiol.* 238:H126–H133.
- McClain, P.E. 1974. *J. Biol. Chem.* 249:2303–2311.
- Medugorac, I. 1982. *Basic Res. Cardiol.* 77:589–598.
- Meier, G.D., Ziskin, M.C., and Bove, A.A. 1982. *Am. J. Physiol.* 243:H1–H11.
- Meier, G.D., Ziskin, M.C., Santamore, W.P., and Bove, A.A. 1980. *IEEE Trans. Biomed. Eng.* 27:319–329.
- Mitchell, J.H., Wildenthal, K., and Mullins, C.B. 1969. *Fed. Proc.* 28(4):1334–1343.
- Myers, J.H., Stirling, M.C., Choy, M., and Gallagher, K.P. 1984. *Proc. 37th ACEMB*, Los Angeles.
- Osakada, G., Sasayama, S., Kawai, C., Hirakawa, A., Kemper, W.S., Franklin, D., and Ross, J., Jr. 1980. *Circ. Res.* 47:173–181.
- Osakada, G., Hess, O.M., Gallagher, K.P., Kemper, W.S., and Ross, J., Jr. 1983. *Am. J. Cardiol.* (in press).
- Prinzen, F.W., Arts, T., Van Der Vusse, G.J., and Reneman, R.S. 1984. *J. Biomech.* 17:801–811.
- Rankin, J.S., McHale, P.A., Arentger, L.E., Liey, D., Greenfield, J.C., and Anderson, R.W. 1976. *Circ. Res.* 39:304–313.
- Ross, J., Jr., Sonnenblick, E.H., Covell, J.W., Kaiser, G.A., and Spiro, D. 1967. *Circ. Res.* 21:409–421.
- Rushmer, R.F., Franklin, D. L., and Ellis, R.M. 1956. *Circ. Res.* 4:684–688.
- Sasayama, S., Franklin, D.L., Ross, J., Jr. 1977. *Am. J. Physiol.* 232:H418–H425.
- Sasayama, S., Gallagher, K.P., Kemper, W.S., Franklin, D.L., and Ross, J., Jr. 1981. *Am. J. Physiol.* 240:H293–H299.

- Streeter, D.D. 1979. Pages 61–112 in *Handbook of physiology*, section 2: The cardiovascular system, vol 1: the heart, R.M. Berne, ed. Bethesda, Maryland, Waverly Press.
- Streeter, D.D., Spotnitz, H.M., Patel, D.J., Ross, J., Jr., and Sonnenblick, E.H. 1969. *Circ. Res.* 24:339–347.
- Tennant, R., and Wiggers, C.J. 1935. *Am. J. Physiol.* 112:351–361.
- Theroux, P., Ross, J., Jr., Franklin, D.L., Covell, J.W., Bloor, C.M., and Sasayama, S. 1977. *Circ. Res.* 4:158–165.
- Torrent-Guasp, F. 1973. *The cardiac muscle*. Juan March Foundation, Barcelona, Spain.
- Truesdell, C. 1966. *The elements of continuum mechanics*. Springer-Verlag, New York.
- Waldman, L.K., Fung, Y.C., and Covell, J.W. 1985. *Circ. Res.* 57:152–163.

# 9

## Residual Stress in Arteries

C.J. CHUONG and Y.C. FUNG

### 1. Introduction

Knowledge of stress distributions in the arterial wall is important for many reasons. In the study of the propagation of pulse waves, one must know the incremental modulus of the elasticity that changes with the stress level. In the study of circulation control, the action of the vascular smooth muscle, which depends on its local stress level (see review in Fung, 1984), must be evaluated. In the study of atherogenesis, one must know the stress distribution in the vessel wall because the tensile and shear stress can alter the local wall permeability and pressure gradient which is the force that drives the fluid in or out of the vessel wall (Chuong and Fung, 1983). Accurate evaluation of stress distributions in the arterial wall is therefore an important step toward a better understanding of various physiological functions and pathological mechanisms associated with the circulatory system.

In the study of blood vessels, the unloaded tube (the one without transmural pressure and axial stretch) is commonly taken as the reference state in which the stress is zero everywhere. However, when an unloaded artery is cut longitudinally, the vessel wall will spring open, with its cross-section becoming a sector (Figure 9.1). The opening angle of the vessel wall is time-dependent after the sudden relief of the initial residual stress. This shows the existence of residual stress in the unloaded tube state; thus the unloaded tube is not stress-free.

An arterial vessel at the stress-free state is, of course, unphysiological. However, it is important to identify this state. When we use pseudo-elasticity (Fung, 1981) to characterize the arterial wall or any other biological soft tissues, we need a stress-free state, the geometry of which is used as the reference state for strain measurements. Correspondingly, we also want to define stresses with respect to this same reference state so that we can relate stresses to strains easily. Any residual stress in the unloaded tube will certainly affect the stress and strain distributions in the tube wall due to actual external loadings of physiological range. In a linearized system, in which a material obeying Hooke's law is subjected to infinitesimal deformation, the final state of

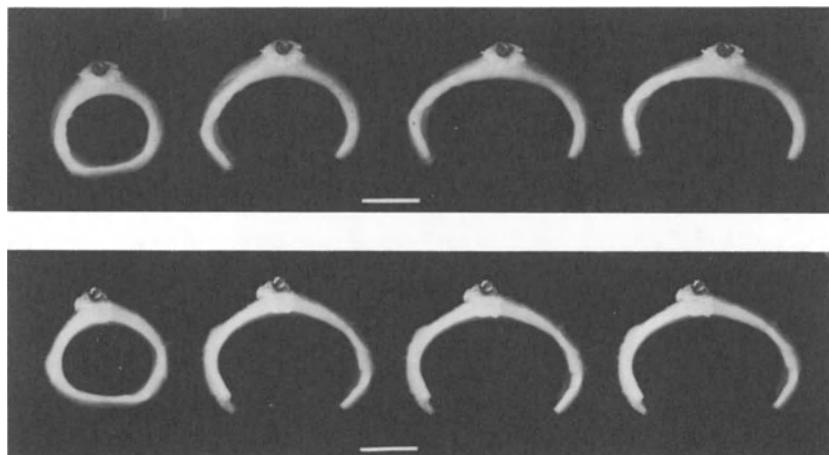


FIGURE 9.1. The cross-sectional shape of the thoracic aorta of the cat. Upper row: vessel soaked in normal saline. Left to right: (1) before cut, (2) 15 s after cut, (3) 15 min after cut, (4) 30 min after cut. Lower row: Normal saline + 2 mg. papaverine/cc. Left to right: same time sequence, before cut, 15 s, 15 min, 30 min after cut. Photograph by courtesy of Paul Patitucci. Bar length: 2 mm. The artery segment was glued to a small pin to support it vertically for photography. From Fung (1984) *Biodynamics: circulation*. Reproduced by permission.

stress is simply the sum of the residual stress and the additional stress due to external loadings. In a nonlinear system, such as the artery whose stress-strain relationship is nonlinear and whose deformation is so large that the strain-displacement relationship is also nonlinear, the principle of superposition does not apply, and the analysis of additional stresses due to external loading must take the residual stresses into account.

In this article we shall show: (1) how the geometry of a stress-free reference state can be determined for arteries from the opened up stress-free configuration, (2) how the stress-strain relationship of the arterial wall can be determined from experimental data in the presence of the residual stress, (3) how the residual strain and residual stress at the unloaded tube state can be evaluated, and (4) how the residual stress affects the stress distribution in the arterial wall due to loadings at the physiological range.

## 2. Method

### 2.1. Determination of the Stress-Free State

We shall assume that the artery is a circular cylindrical tube with a wall material that is homogeneous and cylindrically orthotropic (Patel and Fry, 1969). Under this hypothesis, upon the relief of the residual stress from the

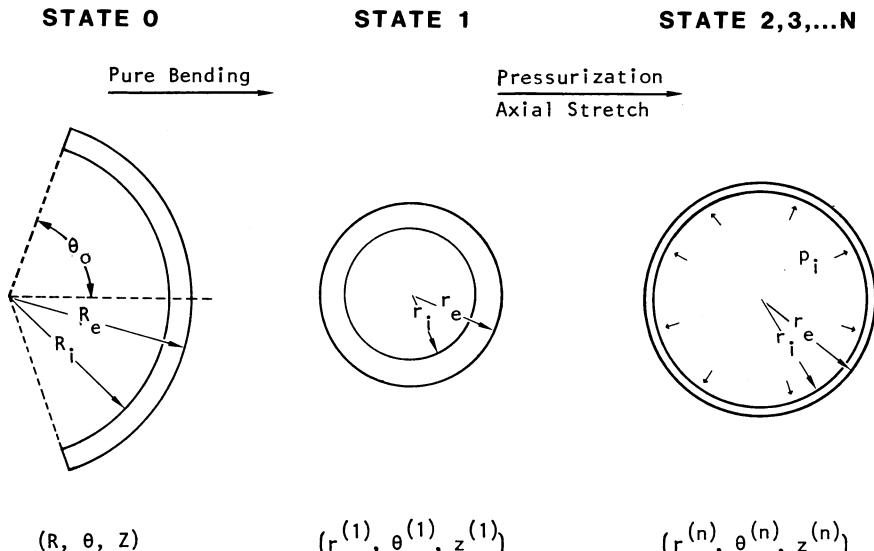


FIGURE 9.2. The cross-sectional representations of arteries at the stress-free reference state 0, the unloaded tube state 1, and the sequential loaded states under transmural pressure and axial stretch.

unloaded tube state, the vessel wall should become a sector of constant curvature and thickness. The noncircular configuration in Figure 9.1 can be due to the adventitia tethering, experimental handling difficulties, or other violations of the assumptions. For arterial elasticity studies, it is important to have an accurate geometric description of the stress-free reference state. In the following, we shall propose a method to determine *the effective radii* and *the effective angle* for the stress-free reference state from the opened up configuration of Figure 9.1.

As shown in Figure 9.2, we shall call the stress-free reference state state 0, the unloaded tube state as state 1, and the subsequent loaded states state 2, 3, ... N. With cylindrical polar coordinates, a material point is denoted as  $(R, \Theta, Z)$  in the state 0 and  $(r, \theta, z)$  in the deformed states 1, 2, 3, ... N. The subscripts *i* and *e* indicate, respectively, the inner and external wall radii at various states.  $\Theta_0$  represents half of the angle of the stress-free arterial wall at state 0.

The angle  $\Theta_0$  and the inner and external wall radii for state 0 and state 1 can be determined by measuring the fiber lengths at the internal and external surfaces directly from the photo in Figure 9.1. For the unloaded tube state 1, it is obvious that we have

$$2\pi r_i = l_i, \quad 2\pi r_e = l_e \quad (9.1)$$

for the inner and external surfaces, respectively, where  $l$  denotes the measured fiber length. For the stress-free reference state, we write

$$2\Theta_0 R_i = L_i, \quad 2\Theta_0 R_e = L_e \quad (9.2)$$

for the inner and external surfaces, respectively, where  $L$  denotes the fiber length measurements at this state. The determination of  $r_i$  and  $r_e$  for the unloaded tube state 1 is straightforward. However, we need another equation for the stress-free state 0, since there are three unknowns,  $\Theta_0$ ,  $R_i$ , and  $R_e$  in Equation (9.2). The incompressibility condition of the vessel wall (Chuong and Fung, 1984) provides the third equation:

$$\Theta_0(R_e^2 - R_i^2) = \pi\lambda_z(r_e^2 - r_i^2) \quad (9.3)$$

where  $\lambda_z$  is the stretch ratio in the axial direction and is to be measured directly. By solving Equations (9.2) and (9.3) the geometric description of state 0 is determined and is to be used as the reference state for strain and stress measurements.

## 2.2. Kinematics

With the geometry of the reference state 0 determined, the deformation of a thick-walled artery under transmural pressure and longitudinal tethering can be described by the following expressions (see Figure 9.2):

$$r = r(R), \quad \theta = (\pi/\Theta_0)\Theta, \quad z = z(Z) \quad (9.4)$$

for radial, circumferential, and longitudinal directions, respectively. Here, the cylindrical polar coordinate system is used. The principal stretch ratios in the radial, circumferential, and longitudinal directions are, respectively,

$$\lambda_r = \frac{\partial r}{\partial R}, \quad \lambda_\theta = \left( \frac{\pi}{\Theta_0} \right) \frac{r}{R}, \quad \lambda_z = \frac{\partial z}{\partial Z}. \quad (9.5)$$

With incompressibility condition  $\lambda_r \lambda_\theta \lambda_z = 1$ , the first of Equation (9.4) can be rewritten as

$$r = \sqrt{r_e^2 - \frac{\Theta_0}{\pi\lambda_z}(R_e^2 - R^2)}. \quad (9.6)$$

## 2.3. Stress–Strain Relationship

The stress–strain relationship is obtained through the use of a pseudo-strain-energy function of the exponential type (Fung, 1981)

$$\rho_0 W = \frac{c}{2} \exp Q \quad (9.7)$$

where

$$Q = b_1 E_\theta^2 + b_2 E_z^2 + b_3 E_r^2 + 2b_4 E_\theta E_z + 2b_5 E_z E_r + 2b_6 E_r E_\theta.$$

$W$  represents the pseudo-strain energy per unit mass,  $\rho_0$  is the mass density of the vessel wall, so that  $\rho_0 W$  is the strain energy per unit volume in the

undeformed reference state 0. The constants  $c, b_1, b_2, b_3, b_4, b_5, b_6$  characterize the wall material.  $E_\theta, E_z$  and  $E_r$  are Green's strain components in the circumferential, longitudinal, and radial directions, respectively. They are related to the principal stretch ratios of Equation (9.5) by

$$E_i = \frac{1}{2}(\lambda_i^2 - 1) \quad (i = r, \theta, z). \quad (9.8)$$

The material is assumed to be incompressible. This constraint is added to the strain-energy function through a Lagrangian multiplier  $H$  to yield

$$\rho_0 W^* = \rho_0 W + \frac{H}{2}[(1 + 2E_\theta)(1 + 2E_z)(1 + 2E_r) - 1]. \quad (9.9)$$

It is known that  $H$  has the significance of a hydrostatic pressure. If  $X_\alpha$  denotes the coordinates of a material particle at the reference state 0, and  $x_i$  denotes that at the deformed state, then the Cauchy stress components can be obtained from:

$$\sigma_{ij} = \frac{\rho}{\rho_0} \frac{\partial x_j}{\partial X_\alpha} \frac{\partial x_i}{\partial X_\beta} \frac{\partial}{\partial E_{\beta\alpha}} \rho_0 W^* \quad (i, j, \alpha, \beta = r, \theta, z) \quad (9.10)$$

where  $\rho, \rho_0$  denote the densities of the material in the deformed and undeformed states, respectively. Summation convention is used in these expressions. Introduction of Equations (9.5), (9.8) and (9.9) into Equation (9.10) yields

$$\begin{aligned} \sigma_\theta &= c(1 + 2E_\theta)[b_1 E_\theta + b_4 E_z + b_6 E_r]e^Q + H \\ \sigma_z &= c(1 + 2E_z)[b_4 E_\theta + b_2 E_z + b_5 E_r]e^Q + H \\ \sigma_r &= c(1 + 2E_r)[b_6 E_\theta + b_5 E_z + b_3 E_r]e^Q + H. \end{aligned} \quad (9.11)$$

## 2.4. Equation of Equilibrium and Boundary Conditions

The problem of a prestrained (or prestressed) thick-walled artery under transmural pressure and longitudinal tethering force can be solved by substituting Equation (9.11) into the equation of equilibrium

$$\frac{\partial \sigma_r}{\partial r} + \frac{\sigma_r - \sigma_\theta}{r} = 0 \quad (9.12)$$

and the boundary conditions.

The boundary conditions are as follows:

1. On the external surface  $r = r_e$ , the extravascular pressure is zero. On the inner surface  $r = r_i$ , an internal pressure  $p_i$  is imposed. Application of these two conditions yields

$$\begin{aligned} p_i &= \int_{r_e}^{r_i} c \{(1 + 2E_r)[b_6 E_\theta + b_5 E_z + b_3 E_r] \\ &\quad - (1 + 2E_\theta)[b_1 E_\theta + b_4 E_z + b_6 E_r]\} e^Q \frac{dr}{r}. \end{aligned} \quad (9.13)$$

It includes the deformed but unloaded tube state 1, at which  $p_i$  has a zero value.

2. On the ends of a blood vessel segment, an external force  $F$  acts. For static equilibrium  $F + p_i \pi r_i^2$  equals the integral of  $\sigma_z$  over the vessel wall cross-section; hence

$$F + p_i \pi r_i^2 = 2\pi \int_{r_i}^{r_e} \sigma_z r dr. \quad (9.14)$$

Use of Equations (9.11) and (9.13) in (9.14) yields

$$\begin{aligned} F = 2\pi c \int_{r_i}^{r_e} r e^\Theta & [(1 + 2E_z)(b_4 E_\theta + b_2 E_z + b_5 E_r) \\ & - \frac{1}{2}(1 + 2E_r)(b_6 E_\theta + b_5 E_z + b_3 E_r) \\ & - \frac{1}{2}(1 + 2E_\theta)(b_1 E_\theta + b_4 E_z + b_6 E_r)] dr. \end{aligned} \quad (9.15)$$

Equations (9.13) and (9.15) are two integral equations from which we can determine the material constants with the knowledge of  $p_i$ ,  $F$ ,  $r_e$ ,  $r_i$  and  $E_\theta$ ,  $E_z$ ,  $E_r$ , i.e., intraluminal pressure, longitudinal stretch force, external radius, internal radius, and distribution of strain components, respectively. Once the material constants are determined, we can evaluate the residual stress at tube state 1 and the stress distribution at loaded states with residual stress (or residual strain) taken into consideration.

### 3. Results and Discussion

#### 3.1. Determination of the Geometry for Reference State 0

To demonstrate the method in determining the geometry of the stress-free reference state 0, an example is given. The values of  $l_i$ ,  $l_e$ ,  $L_i$ , and  $L_e$  from the top row of Figure 9.1 (the tube state 1 and the opened up state 0 at the right) are 8.75, 12.5, 9.75, and 11.25 mm, respectively. By assuming  $\lambda_z = 1$ , the use of Equation (9.2) and (9.3) yields  $R_e = 4.52$  mm,  $R_i = 3.92$  mm, and  $\Theta_0 = 71.4$  degrees for the geometry of the reconstructed state 0 (see the sketch in Figure 9.2). By this method, we identify the stress-free reference state for the artery of Figure 9.1.

#### 3.2. Residual Strain and Stress at the Unloaded Tube State 1

With the geometry of the stress-free state determined, the residual strain at the unloaded tube can be obtained. This is shown in Figure 9.3(a) in terms of principal stretch ratios: fibers at the inner wall are shortened, and fibers at the outer wall are elongated. The unloaded tube is in a state of pure bending. One might ask where the neutral surface is where the circumferential fiber of the

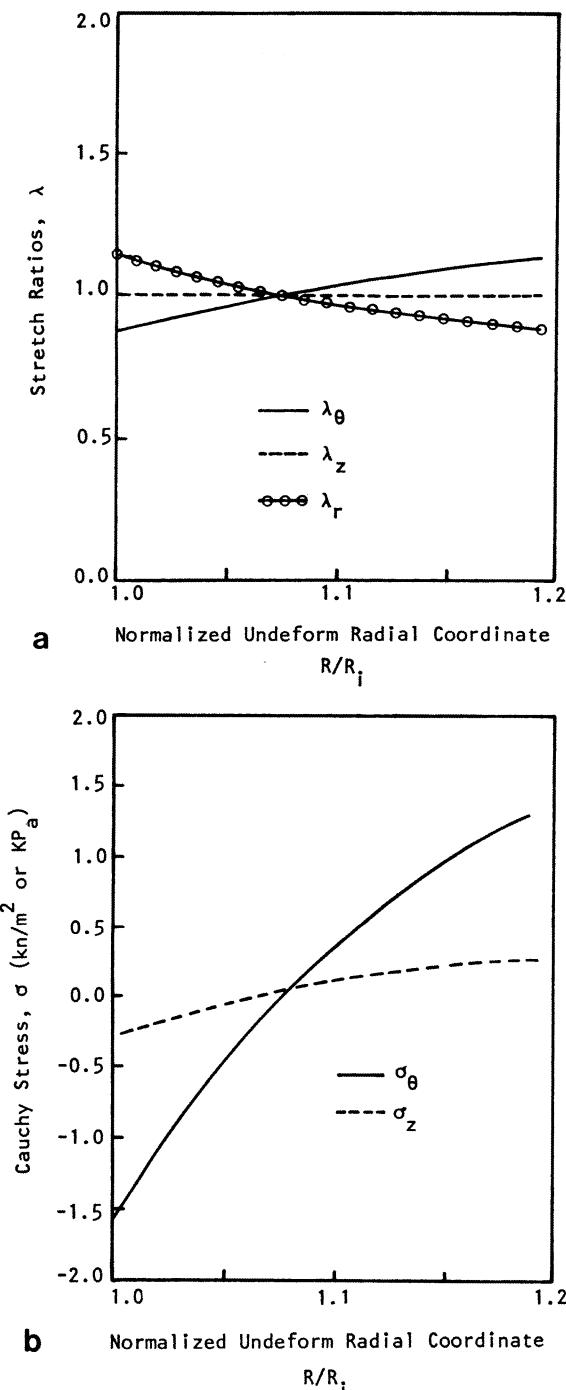


FIGURE 9.3. Residual strain and stress distribution through the thickness of a rabbit thoracic artery at the unloaded tube state. (a) Distribution of principal stretch ratios  $\lambda_\theta$ ,  $\lambda_z$ ,  $\lambda_r$ . (b) Distribution of residual stresses  $\sigma_\theta$ ,  $\sigma_z$ .

arterial vessel is not strained at all. Its location can be obtained by substituting Equation (9.6) into the second of Equation (9.5) when  $\lambda_0 = 1$ . In this same example, it falls at 43% of the thickness measured from the inner wall.

To illustrate how to evaluate the residual stress at state 1 and the stress-strain distributions at the loaded states, we shall use the raw data of Exp: 71 on rabbit thoracic artery from Fung, Fronek, and Patitucci (1979). We assume the residual strain in the artery of Exp: 71 to be distributed like that of the artery in Figure 9.1. Experimentally, increasing pressure and varying axial tethering force were applied on the rabbit thoracic artery step by step, and changes of the arterial diameter and length were recorded. Following the formulation in the previous section, we solve Equations (9.13) and (9.15) to obtain the material constants. They are:

$$\begin{aligned} c &= 22.40 \text{ kPa} \\ b_1 &= 1.0672 & b_2 &= 0.4775 & b_3 &= 0.0499 \\ b_4 &= 0.0903 & b_5 &= 0.0585 & b_6 &= 0.0042 \end{aligned}$$

With the material constants obtained, we can evaluate stress distribution at various states. Figure 9.3(b) shows the residual stress at the tube state 1 when there is no external loading. In the circumferential direction, the inner wall is under compression of 1.5 kPa, and the outer wall is under tension of 1.3 kPa. The distribution across the wall thickness is approximately linear. At this state, the internal pressure is zero. For static equilibrium, the integration of the total circumferential components force across the wall thickness is zero. The net effect is a pure bending.

### 3.3. Strain Distributions at Loaded States

Figures 9.4(a), 9.5(a), and 9.6(a) show the distributions of principal stretch ratios of the vessel wall for  $p_i = 60, 120$ , and  $160 \text{ mmHg}$  ( $\sim 8, 16$ , and  $21.3 \text{ kPa}$ ), respectively. The corresponding longitudinal stretch ratios are 1.542, 1.691, and 1.696. The use of the opened up tube as the zero-stress-zero-strain state greatly reduces the degrees of the strain developed due to pressurization and axial tethering force.

In a previous computation (Chuong and Fung, 1983), using the hypothesis that the unloaded tube is stress-free, we obtained large strain at the inner wall and large stress concentrations. In the present work, we identify the opened up tube as the stress-free state, and defined the strains with respect to this state. The effect of the residual strain in the unloaded tube results in a less nonuniform strain distribution across the vessel wall in loaded states.

### 3.4. Stress Distributions at Loaded States

Figures 9.4(b), 9.5(b), and 9.6(b) show the stress distributions of the vessel wall for  $p_i = 60, 120$ , and  $160 \text{ mmHg}$ , respectively. The corresponding longitudinal stretch ratios are 1.542, 1.691, and 1.696. It appears that the longitudinal stress

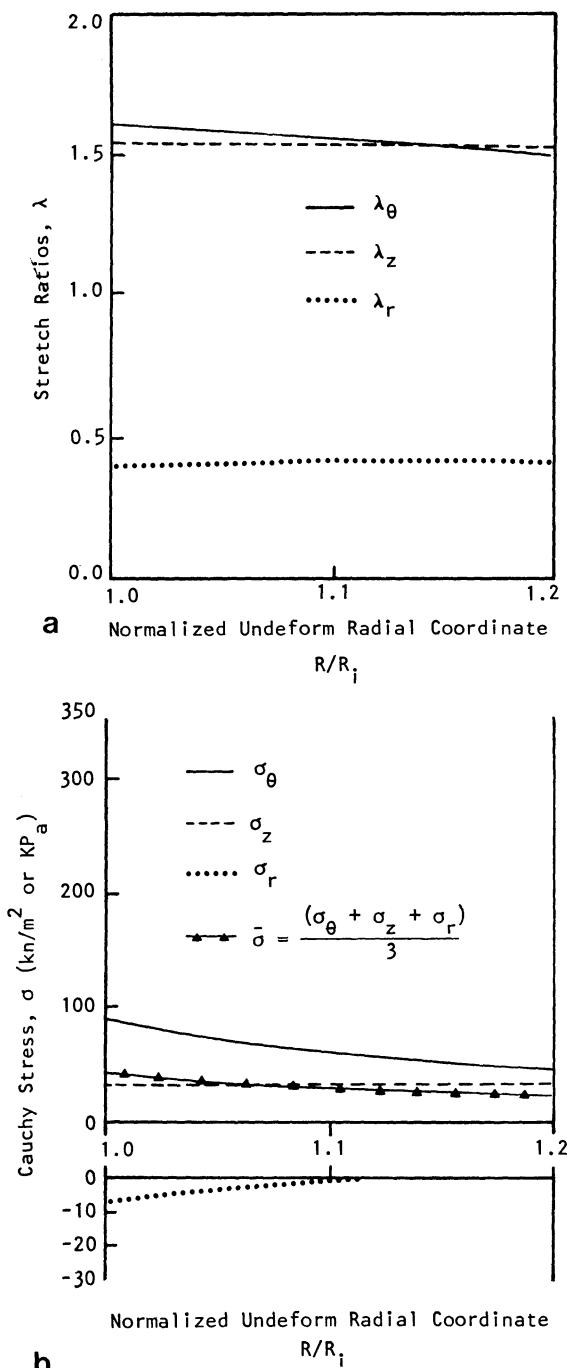


FIGURE 9.4. Strain and stress distribution through the thickness of a rabbit thoracic artery at  $p_i = 60 \text{ mmHg}$  ( $\sim 8 \text{ kPa}$ ),  $\lambda_z = 1.542$ . (a) Distributions of principal stretch ratios  $\lambda_\theta$ ,  $\lambda_z$ ,  $\lambda_r$ . (b) Distributions of Cauchy stress  $\sigma_\theta$ ,  $\sigma_z$ ,  $\sigma_r$  and the mean stress  $\bar{\sigma}$ .

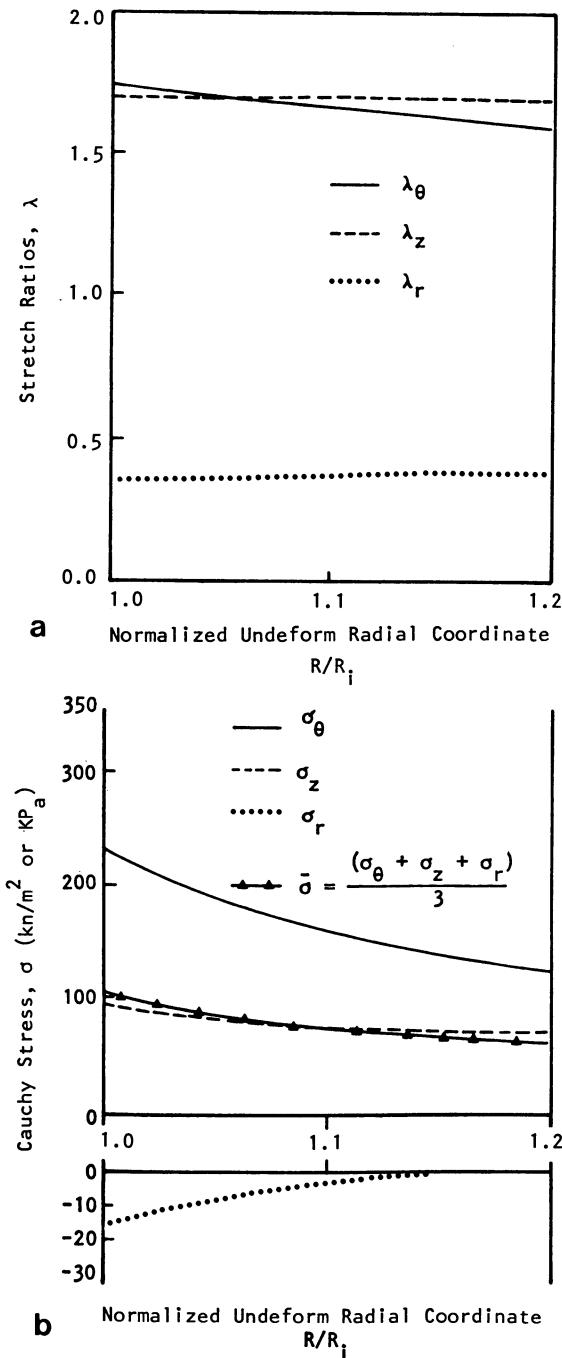


FIGURE 9.5. Strain and stress distribution through the thickness of a rabbit thoracic artery at  $p_i = 120 \text{ mmHg}$  ( $\sim 16 \text{ kPa}$ ),  $\lambda_z = 1.691$ . (a) Distributions of principal stretch ratios  $\lambda_\theta$ ,  $\lambda_z$ ,  $\lambda_r$ . (b) Distributions of Cauchy stress  $\sigma_\theta$ ,  $\sigma_z$ ,  $\sigma_r$  and the mean stress  $\bar{\sigma}$ .

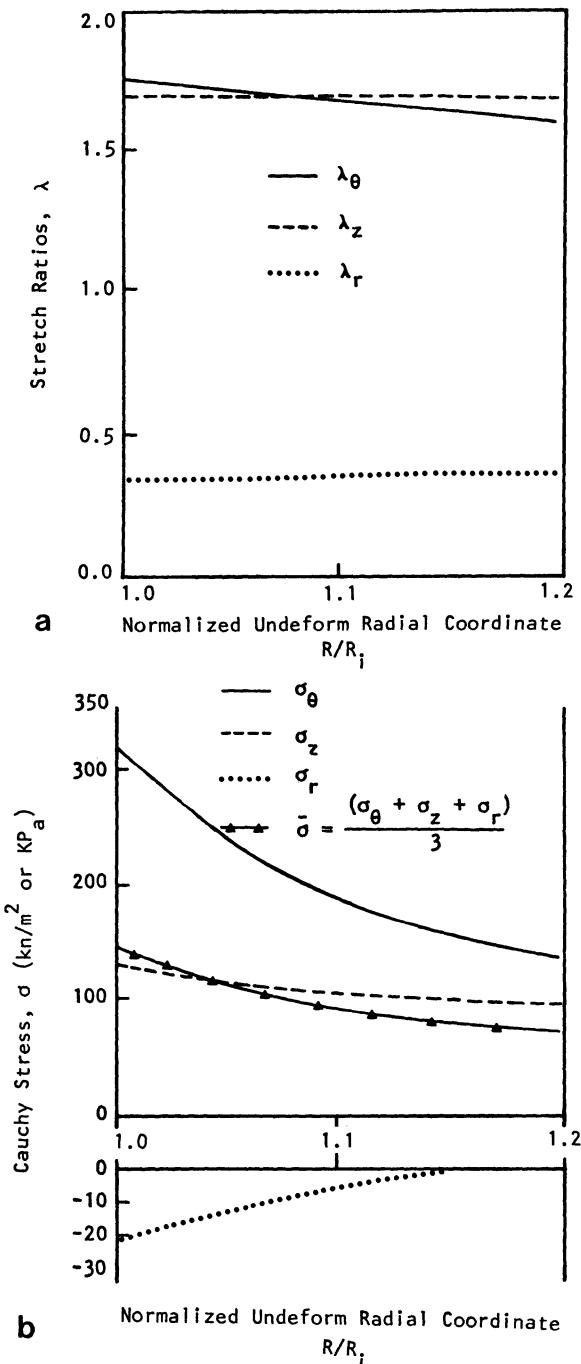


FIGURE 9.6. Strain and stress distribution through the thickness of a rabbit thoracic artery at  $p_i = 160 \text{ mmHg}$  ( $\sim 21.3 \text{ kPa}$ ),  $\lambda_z = 1.696$ . (a) Distributions of principal stretch ratios  $\lambda_\theta$ ,  $\lambda_z$ ,  $\lambda_r$ . (b) Distributions of Cauchy stress  $\sigma_\theta$ ,  $\sigma_z$ ,  $\sigma_r$  and the mean stress  $\bar{\sigma}$ .

component  $\sigma_z$  is consistently uniform across most of the wall thickness, even when  $p_i = 160$  mmHg. At 160 mmHg, the longitudinal stress  $\sigma_z$  is 120 kPa at the inner wall and 100 kPa at the outer wall, with changes occurring primarily at the inner layer. For the circumferential stress  $\sigma_\theta$ , the non-uniform distribution is clear. At  $p_i = 120$  mmHg, it is 230 kPa at the inner wall and 120 kPa at the outer wall. At  $p_i = 160$  mmHg, it becomes 320 kPa at the inner wall and 130 kPa at the outer wall.

The distribution of the hydrostatic pressure (the mean stress) is also plotted in these figures. The gradient of the hydrostatic pressure is a driving force for fluid transport across the arterial wall. At 60 mmHg, the distribution of hydrostatic pressure is rather uniform throughout the wall thickness. At 120 mmHg, it is 100 kPa at the inner wall and 60 kPa at the outer wall. At 160 mmHg, it becomes 140 kPa at the inner wall and 75 kPa at the outer wall. This suggests that, as the pressure varies from 120 to 160 mmHg, the driving force for fluid movement across the vessel wall can be significantly changed at the inner layer.

In our earlier work (Chuong and Fung, 1983), in which the unloaded tube is assumed to be stress-free, very large stress concentration at the inner wall was obtained (e.g., at 120 mmHg the circumferential stress at the inner wall was 6.5 times larger than the average value across the vessel wall). In the present work, we removed this hypothesis. With respect to an experimentally identified stress-free state, we have evaluated residual stress and strain at the unloaded tube state. The residual stress (or the residual strain) in the unloaded tube is demonstrated to be significant in reducing the high stress concentration and high stress gradient at the vessel wall due to loadings in the physiological range. For example, at 120 mmHg, the circumferential stress at the inner wall is calculated to be (Figure 9.5(b)) 1.42 times larger than the average value across the vessel wall (in contrast to 6.5 under the previous hypothesis). Thus, there is no doubt that the residual stress is an important factor to consider.

#### 4. Summary

In this chapter, a method is proposed to identify the stress-free state from experimental observation of longitudinally sliced vessel specimen. Stress and strain are referred to this state. A formulation is outlined to evaluate the residual stress and strain in unloaded blood vessels as well as stress and strain in vessels under pressurization and axial stretch.

The presence of the prestress leads to a reduction of the stress concentration at the inner wall of the tube and therefore plays an important role in vessel wall mechanics.

#### References

- Chuong, C.J., and Fung, Y.C. 1983. *J. Biomech. Eng.* 105:268–274.  
———. 1984. *J. Biomech.* 17:35–40.

- Fung, Y.C. 1981. *Biomechanics: mechanical properties of living tissues*. Springer-Verlag, New York.
- \_\_\_\_\_. 1984. *Bio dynamics: circulation*. Springer-Verlag, New York.
- Fung, Y.C., Fronek, K., and Patitucci, P. 1979. *Am. J. Physiol.* 237:H620–H631.
- Green, A.E., and Adkins, J.E. 1970. *Large elastic deformations*, 2d ed. Clarendon Press, Oxford, England, U.K.
- Patel, D.J., and Fry, D.L. 1969. *Circ. Res.* 24:1–8.
- Patitucci, P. 1982. Computer programs for fitting pseudo-strain energy functions to soft tissue experimental stress and strain data. Digital Equipment Users Society, DECUS, No. 11-548, Malboro, Mass.

# 10

## On the Structural Origin of the Quasilinear Viscoelastic Behavior of Tissues

Y. LANIR

### 1. Introduction

Since its introduction by Fung (1972), the law of quasilinear viscoelasticity (QLV) was applied for several tissues including, among others, tendons (Haut and Little, 1972), ligaments (Geiger et al, 1983; Woo et al, 1981), articular cartilage under tension (Woo et al, 1980), heart valves (Rousseau et al, 1983), papillary muscle (Pinto and Patitucci, 1980), and smooth muscle (Price et al, 1977). Fung based the QLV law on the observation that, in several tissues, the time-varying response to a step change in uniaxial stretch (from  $\lambda = 1.0$  to  $\lambda > 1.0$ ), can be expressed by:

$$\sigma(\lambda, t) = G(t) \cdot T^{(e)}(\lambda), \quad G(0) = 1, \quad (10.1)$$

where  $G(t)$  is the reduced relaxation function and  $T^{(e)}(\lambda)$  is the elastic (immediate) response.

By superposition of infinitesimal steps of stretch, Equation (10.1) leads to the quasilinear hereditary integral as follows:

$$T(t) = \int_0^t G(t - \tau) \cdot \dot{\lambda}(\tau) \cdot (dT^{(e)}/d\lambda) d\tau, \quad (10.2)$$

where  $T(t)$  is the time-varying Lagrangian stress.

The importance of the QLV law of Equation (10.2) is that, due to its mathematical similarity to the linear hereditary integral, it is possible to utilize the powerful mathematical tools of linear viscoelasticity in the analysis of the viscoelastic behavior of tissues.

Following studies of the correlates between structure and mechanical behavior of tissues (e.g., Lanir, 1980), the question arises whether the QLV behavior of tissues may be accounted for by their morphological architecture. In the following article, this possibility will be examined.

A careful study of the tendon's fiber architecture during stretch (Viidik, 1972) has demonstrated that, with increasing stretch, a gradually increasing number of originally wavy (crimped) fibers become straight and subsequently stretch. This correlates with an increase in the tendon's stiffness with strain.

A similar yet more complex response has been observed in tissues in which the collagen fibers are biaxially disposed: the wavy fibers become gradually straight upon tissue stretch (Chu et al, 1972), but at the same time, they gradually rotate toward the direction of stretch in uniaxial tests (Broom, 1978; Brown, 1973).

These structural features have been incorporated into stochastic models for the tissues' overall mechanical response (e.g., Lanir, 1983). A summary of this approach is presented first. Next, it is determined if and under which circumstances the structural models predict quasilinear viscoelastic response. We shall start with the uniaxial case of the tendon. Later, the general three-dimensional case is studied.

## 2. The Uniaxial Case

For clarity, we shall first consider the tendon as being elastic. The tendon consists of a parallel array of wavy collagen fibers. The function  $D(x)$  expresses the density distribution of the ensemble waviness, where the variable  $x$  stands for the straightening stretch ratio  $\lambda_s$ . The percentage of fibers that become straight at stretch ratio between  $x$  and  $x + dx$  is  $D(x)dx$ . Under an overall stretch  $\lambda > x$ , these fibers are stretched by  $\lambda/x$ , so that their contribution to the overall stress is  $q \cdot f(\lambda/x)D(x)dx$ , where  $q$  is the fiber's volume fraction in the tendon and  $f$  is the Lagrangian stress in a stretched fiber, expressed in terms of its true stretch ratio. The overall elastic Lagrangian stress  $T^{(e)}$  is thus given by:

$$T^{(e)}(\lambda) = \int_{1.0}^{\lambda} q \cdot f(\lambda/x) \cdot D(x) \cdot dx, \quad (10.3)$$

where it is assumed that at  $\lambda = 1.0$ , all the fibers are wavy; this is in agreement with experimental observations (Viidik, 1972).

The collagen fibers of the elastic tendon will be considered as linear elastic in tension. This is a reasonable assumption since their true strain is rather small (2–4%). In addition, it has been shown (Viidik, 1972; Abrahams, 1967; Rigby et al, 1959) that at high overall strain, when all the fibers are stretched, the stress–strain relations are linear. Under compression, the fibers crimp and can support no load. Hence the fiber's stress is given by:

$$\begin{aligned} f(\lambda_t) &= K(\lambda_t - 1) = K \cdot (\lambda - x)/x & \lambda > 1.0 \\ f(\lambda_t) &= 0 & \lambda \leq 1.0 \end{aligned} \quad (10.4)$$

where  $K$  is a material parameter. Introducing Equation (10.4) to (10.3) yields the stress–strain relations for elastic tendons under tension as follows:

$$T^{(e)}(\lambda) = \int_{1.0}^{\lambda} q \cdot K \cdot (\lambda - x) \cdot D(x)/x \cdot dx. \quad (10.5)$$

The slope  $dT^{(e)}/d\lambda$  can be obtained via Liebnitz rule, and we get:

$$dT^{(e)}/d\lambda = K \cdot q \cdot \int_{1.0}^{\lambda} D(x)/x \cdot dx. \quad (10.6)$$

Real tendons are, however, viscoelastic. The stress is a function of the entire strain history. But the basic principle still holds: the viscoelastic response of the whole tendon reflects the sum of contributions of all its components. The tendons' components are viscoelastic collagen fibers and a fluid-like ground substance. The latter probably plays a negligible role in tensile deformations. The viscoelastic response of the collagen fibers is complex since it is not symmetric in stretch vs. contraction. This results from the fact that the fibers can sustain no compressive loading. If we consider an example of stretched fibers that are quickly contracted, then each fiber will contract in accordance with the overall deformation only up to a point where its tensile load diminishes to zero, and from this point on it will fold and further contract at a slow rate under no load, and in the mode that is in general different from that of a stretched fiber.

The following discussion is thus restricted to deformations in which the stretch rate  $\dot{\lambda}(t)$  is non-negative, i.e., the stretch ratio is either constant or increases with time. This restriction guarantees that a single relaxation mode governs the tissue's response, since the response of contracting fibers is not present. Tests with non-negative stretch rate are common and include stress relaxation and creep tests as well as the loading phases of constant rate of stretch and constant rate of loading tests. But periodic tests are not included since they contain unloading periods.

The time-varying response of the collagen fibers is taken to be linear viscoelastic, for reasons similar to the ones discussed above (i.e., their low true strain and the tendon's linear response when all the fibers are straight). In addition, if the fibers were nonlinear viscoelastic, it is unlikely that their combined viscoelastic response would be quasilinear.

The elastic part of the fiber's behavior, which is the immediate response to a step change of stretch, is expressed by Equations (10.5) and (10.6). The full viscoelastic behavior of a single straight fiber is expressed by the Boltzman hereditary integral:

$$f(\lambda_t, t) = K \cdot \int_0^t G(t - \tau) \cdot \dot{\lambda}_t(\tau) \cdot d\tau, \quad (10.7)$$

where  $G(T)$  is the reduced relaxation function and  $G(0) = 1.0$  and  $\lambda_t = \lambda/x$ . In Equation (10.7), the history of  $\lambda(t)$  starts at  $t = 0$ , where  $\lambda(0) = 1.0$ . By summing up the contributions of all fibers, we obtain the overall tendon's stress as follows:

$$T(\lambda, t) = \int_{1.0}^{\lambda(t)} q \cdot D(x) \left[ K \cdot \int_0^t G(t - \tau) \cdot \dot{\lambda}(\tau)/x \cdot d\tau \right] \cdot dx \quad \dot{\lambda}(t) \geq 0. \quad (10.8)$$

The order of integration can be changed and we get:

$$T(\lambda, t) = \int_0^t G(t - \tau) \cdot \dot{\lambda}(\tau) \left[ K \cdot q \cdot \int_{1.0}^{\lambda(\tau)} D(x)/x \cdot dx \right] d\tau, \quad (10.9)$$

which in view of Equation (10.6) can be written as:

$$T(\lambda, t) = \int_0^t G(t - \tau) \cdot \dot{\lambda}(\tau) [dT^{(e)}/d\lambda] d\tau, \quad \dot{\lambda}(\tau) \geq 0. \quad (10.10)$$

Equation (10.10) is Fung's quasilinear viscoelastic law. It is thus seen that, under non-negative strain rate, the quasilinear viscoelastic behavior of tendons can be considered a consequence of their microstructure.

### 3. The Three-Dimensional Case

Agreement between structural models and the law of quasilinear viscoelasticity may not be restricted to cases in which the collagen fibers are uniaxially disposed. In fact, multidimensional structural considerations may offer a simple way to generalize the law to the three-dimensional situation.

If the distribution density of the fiber orientation in a tissue is expressed by a density distribution  $R(\phi, \theta)$  where  $\phi$  and  $\theta$  are the spherical polar angles, then the total strain energy ( $W$ ) of the fiber ensemble is given by:

$$W = \int_{\phi} \int_{\theta} q \cdot R(\phi, \theta) \cdot w(\varepsilon) \cdot d\theta \cdot d\phi, \quad (10.11)$$

where  $q$  is the fiber volume fraction in the tissue and  $w$  is the fiber's uniaxial strain energy, which is a function of its axial strain  $\varepsilon$ . In the three dimensional case it is preferred to use the strain  $\varepsilon$  rather than the stretch ratio  $\lambda$ , since only the former is a tensor component which can be readily transformed between different coordinate systems. The integrals in Equation (10.11) are performed over the whole range of possible orientations  $0 \leq \theta \leq \pi$ ,  $0 \leq \phi \leq \pi/2$ . The fiber's strain  $\varepsilon$  depends on its initial orientation  $(\phi, \theta)$  and on the tissue's overall strain  $e_{ij}$  (Lanir, 1983). As an example, if  $e_{ij}$  is Green's strain tensor, then in the case of triaxial stretch expressed by  $e_{ij} = e_i \delta_{ij}$  ( $i$  not summed), the transformation from the overall strain  $e_{ij}$  to the fiber's axial strain  $\varepsilon$  is:

$$\varepsilon = (e_1 \cdot \sin^2 \phi \cdot \cos^2 \theta + e_2 \cdot \sin^2 \phi \cdot \sin^2 \theta + e_3 \cdot \cos^2 \phi). \quad (10.12)$$

If the fibers are elastic, then under general deformation, the elastic Kirchoff's stress component  $-{}^{(e)}S^{ij}$  can be derived from  $W$  as follows:

$${}^{(e)}S^{ij} = \partial W / \partial e_{ij}, \quad (10.13)$$

where the usually small contribution of the ground substance (through hydrostatic pressure) in tensile deformations is disregarded. From Equations (10.11) and (10.13) and use of chain rule differentiation, we get:

$${}^{(e)}S^{ij} = \int_{\phi} \int_{\theta} R(\phi, \theta) \cdot S^{(e)}(\varepsilon) \cdot (\partial \varepsilon / \partial e_{ij}) \cdot d\theta \cdot d\phi, \quad (10.14)$$

where  $\partial \varepsilon / \partial e_{ij}$  is derived from transformations like Equation (10.12) and is a function of the orientation alone;  $S^{(e)}(\varepsilon)$  is the elastic stress in an ensemble of parallel crimped fibers with density  $q$  and is given by an expression analogous to Equation (10.3), namely:

$$S^{(e)}(\varepsilon) = \int_0^{\varepsilon} q \cdot \bar{f}(\varepsilon_t) \cdot \bar{D}(y) \cdot dy \quad (10.15)$$

where  $\bar{f}(\varepsilon_t)$  is the axial Kirchhoff stress in the fiber expressed in terms of its true strain and  $\bar{D}(y)$  is the density distribution of the fiber's waviness, where  $y$  stands for straightening strain. Since  $\varepsilon = (\lambda^2 - 1)/2$ , it can be readily shown that  $\varepsilon_t = (\varepsilon - y)/2y + 1$ .

For later use, the term  $\partial^{(e)}S^{ij}/\partial e_{kl}$  is derived from Equation (14) by chain differentiation:

$$\partial^{(e)}S/\partial e_{kl} = \int_{\phi} \int_{\theta} R(\phi, \theta) \cdot (\partial S^{(e)}/\partial \varepsilon) \cdot (\partial \varepsilon / \partial e_{kl}) \cdot (\partial \varepsilon / \partial e_{ij}) \cdot d\theta \cdot d\phi, \quad (10.16)$$

where  $dS^{(e)}/d\varepsilon$  is related to  $\bar{f}(\varepsilon_t)$  through Equation (10.15). If the fibers are linear elastic with respect to the strain  $\varepsilon$  (i.e.,  $\bar{f}(\varepsilon_t) = K \cdot \varepsilon_t$ ), then:

$$dS^{(e)}/d\varepsilon = Kq \int_0^{\varepsilon} (\bar{D}(y)/2y + 1) dy. \quad (10.17)$$

If the fibers are linear viscoelastic, then under conditions of non-negative strain rate ( $\dot{\varepsilon}(t) \geq 0$ ), the function  $\bar{f}(\varepsilon_t)$  in Equation (10.15) is replaced by its viscoelastic counterpart  $\bar{f}(\varepsilon_t, t)$  given by:

$$\bar{f}(\varepsilon_t, t) = K \cdot \int_0^t G(t - \tau) \cdot \dot{\varepsilon}_t(\tau) \cdot d\tau. \quad (10.18)$$

Since  $\dot{\varepsilon}_t(\tau) = \dot{\varepsilon}(\tau)/2y + 1$ , then by substituting Equation (10.18) in Equation (10.15), changing the order of integration and using Equation (10.17), we get:

$$S(\varepsilon, t) = \int_0^t G(t - \tau) \cdot \dot{\varepsilon}(\tau) \cdot (dS^{(e)}/d\varepsilon) \cdot d\tau. \quad (10.19)$$

To obtain the viscoelastic response of the whole tissue, we replace the term  $S^{(e)}(\varepsilon)$  in Equation (10.14) by  $S(\varepsilon, t)$  of Equation (10.19). Thus:

$$S^{ij}(t) = \int_{\phi} \int_{\theta} R(\phi, \theta) \cdot (\partial \varepsilon / \partial e_{ij}) \cdot \left[ \int_0^t G(t - \tau) \cdot \dot{\varepsilon}(\tau) \cdot (dS^{(e)}/d\varepsilon) \cdot d\tau \right] \cdot d\theta \cdot d\phi. \quad (10.20)$$

Since  $\dot{\varepsilon}(\tau) = (\partial \varepsilon / \partial e_{kl}) \cdot \dot{e}_{kl}(\tau)$  ( $k, l$  summed), then by changing the order of integration in Equation (10.20) and using Equation (10.16), we get:

$$S^{ij}(t) = \int_0^t G(t-\tau) \dot{e}_{kl}(\tau) \cdot (\partial^{(e)} S^{ij}/\partial e_{kl}) \cdot d\tau, \quad (k, l \text{ summed}), \quad \dot{e}(t) \geq 0. \quad (10.21)$$

Equation (10.21) is formally similar to Fung's QLV law (10.10) and can be regarded as its three-dimensional generalization. The integrand for each component  $S^{ij}(t)$  involves six independent derivatives  $\partial^{(e)} S^{ij}/\partial e_{kl}$ , which are defined in Equations (10.16), (10.17), and (10.19) in terms of the tissue's structure—its orientation  $R(\phi, \theta)$ , waviness  $D(y)$  and fiber density  $q$ —and the mechanical properties of its fibers ( $G(t)$  and  $k$ ). Alternatively, the derivatives  $\partial^{(e)} S^{ij}/\partial e_{kl}$  can be determined experimentally without reference to the tissue's structure by measuring the immediate response of the stress component  $S^{ij}$  to a small change of the strain component  $e_{kl}$ .

## 4. Discussion and Conclusion

The general rheological characterization of nonlinear viscoelastic materials like soft tissues is very complex. Several attempts have been made to develop theories that are simpler and usable. One such simplification is Fung's QLV law. Another line of approach is structural theories, which have been developed for tendons, skin, and other flat tissues and for fibrous tissues in general. They have been applied in their various forms for tendons in the elastic case (Lanir, 1978), uniaxial tests of tympanic membrane (Decraemer et al, 1980), skin (Wijn, 1980), and skin and pericardium (Shoemaker, 1984).

Since the structural theories consider the physical environment in the deforming tissue, it is of interest to examine if and under what circumstances they conform with other theories, notably the QLV law.

Structural considerations show that tissues consisting of one type of linear viscoelastic and wavy fibers are expected to follow Fung's QLV law under deformations in which the strain rate is non-negative. This includes stress relaxation and creep tests and the loading phases of cyclic tests. In the unloading phase, the tissue's response is expected to differ from the QLV law owing to the anticipated difference between the viscoelastic responses of stretched vs. contracting crimped fibers. Under these restrictions, the structural theory predicts that the quasilinear viscoelastic law is valid not only under uniaxial stretch—Equation (10.10)—but under three-dimensional deformations as well. In fact, by considering the microstructure, it was possible to derive a precise expression for the three-dimensional QLV law—Equation (10.21).

It may be argued that the QLV behavior of tissues could be accounted for by QLV behavior of the fibers themselves, and that the tissues' microstructure may augment this effect but not be the sole reason for it. It is difficult to examine this point directly from experiments, since isolated fibers are not

readily available. Yet the stress-strain relationships of whole tendons are known to be linear when all the fibers are stretched. This would indicate that the fibers themselves are linear and that the nonlinear aspects of the tissue stem primarily, if not solely, from their dynamic microstructure.

Most tissues have more than one type of fiber, and each may have a different mechanical response (e.g., elastin and collagen). If the relative contribution of each fiber network remains constant with strain, then all the networks can be replaced by an equivalent single network. The tissue's response is then expected to obey the QLV law. If, however, this is not the case (e.g., the skin in which elastin is predominant at low strains and collagen at high strains), then the mode of behavior shifts from one to the other. The QLV law is not expected to hold for such tissues, and a composite nonlinear viscoelastic approach (e.g., Lanir, 1983) must be employed. The skin was found to behave in a nonlinear viscoelastic manner (Lanir and Fung, 1974; Lanir, 1976), thus supporting this conclusion.

## References

- Abrahams, M. 1967. *Med. Biol. Eng.* 5:433–443.
- Broom, N.D. 1978. *J. Thorac. Cardiovasc. Surg.* 75:212–1130.
- Brown, I.A. 1973. *Brit. J. Dermatol.* 89:383–393.
- Chu, B.M., Frasher, W.G., and Wayland, H. 1972. *Annals Biomed. Eng.* 1:182–203.
- Decraemer, W.F. 1980. *J. Biomechanics* 13:559–564.
- Fung, Y.C. 1972. Pages 181–208 in *Biomechanics: its foundation and objectives*, Y.C. Fung, N. Perrone, and M. Anliker, eds. Prentice-Hall, Englewood Cliffs, N.J.
- Geiger, D., Trevisan, D., and Oddou, C. 1983. *Proc. Europ. Mech. Colloq.* 170:10.
- Haut, R.G., and Little, R.W. 1972. *J. Biomech.* 5:423–430.
- Lanir, Y. 1976. *Annals Biomed. Eng.* 4:250–270.
- \_\_\_\_\_. 1978. *J. Bioeng.* 2:119–128.
- \_\_\_\_\_. 1980. ASME Trans., *J. Biomech. Eng.* 102:332–339.
- \_\_\_\_\_. 1983. *J. Biomech.* 16:1–12.
- Lanir, Y., and Fung, Y.C. 1974. *J. Biomech.* 7:171–182.
- Pinto, J.G., and Patitucci, P.J. 1980. ASME Trans., *J. Biomech. Eng.* 102:57–61.
- Price, J.M., Patitucci, P.J., and Fung, Y.C. 1977. *Am. J. Physiol. Cell Physiol.* 2: 047–055.
- Rigby, B.J., Hirai, R., Spikes, J.D., and Eyring, H. 1959. *J. Gen. Physiol.* 43:265–283.
- Rousseau, E.P.M., Sauren, A.A.H.J., van Haut, M.C. and van Steenhoven, A.A. 1983. *J. Biomech.* 16:339–348.
- Shoemaker, P.A. 1984. Ph.D. dissertation, University of California, San Diego.
- Viidik, A. 1972. *Ztschr. Anat. Entw.-gesch.* 136:204–212.
- Wijn, P.F.F. 1980. Ph.D. thesis, Katholieke University, Nijmegen.
- Woo, S.L-Y., Simon, B.R., Kuei, S.C., and Akeson, W.H. 1980. ASME Trans., *J. Biomech. Eng.* 102:85–90.
- Woo, S.L-Y., Gomez, M.A., and W.H. Akeson, 1981. ASME Trans., *J. Biomech. Eng.* 103:292–298.

# 11

## Fibrous Skin Mechanics: Superstructure and New Problems

S.A. WAINWRIGHT

### 1. Introduction

We divide the living world into levels of organization—atoms, molecules, cells, tissues, organs, organisms—that constitute a structural hierarchy. Most biomechanics research is done by identifying a functional problem in one of these levels and then analyzing structure and function at that level and the levels immediately below. Conclusions are then drawn about the mechanisms permitting and controlling the function. The argument is made here that not only may structure at higher levels also control function, as has been shown for capillary vessels in mammalian mesentery (Fung et al, 1966), but also that function at higher levels is completely different and is often a more challenging and rewarding goal for research.

Shark skin will serve as an example of how structural features at the levels of macromolecules (collagen), tissues (skin), and organs (the skin formed into a cylinder) interact to produce the higher level function of swimming. Then, to show the importance of qualitative thinking in early stages of new biomechanical research areas, the problem of whale swimming is introduced.

### 2. The Nature of Skin: Substructure and Properties

“Skin” is a small word that alludes to a wide variety of materials across the animal kingdom, most of which will not concern us here. What many skins have in common is that, at first approximation, they can be described as two-dimensional structures: the third dimension, thickness, while finite, is less than 1% of the orthogonal dimensions. Skin does not occur in one-dimensional strands like tendon. Nor does it, like the bone in vertebral bodies, occur as three-dimensional lumps. We shall call the category of two-dimensional sheets of pliant material *membranes*.

In any membrane, a large, forced extension in one dimension will cause a large forceful contraction normal to it in the plane of the membrane. The assumption that there is a negligible change of volume is justified for all

examples discussed here. This expression of the Poisson ratio is observed every time one stretches skin, whether it is still attached to the body or isolated from the body in a testing rig.

In an intact cat, for example, the skin is both attached to the underlying tissues and continuous around the cat's body. Both of these conditions make the skin stiffer on the cat than a strip of cat skin is in a uniaxial testing rig. The stiffness that is most interesting here is the resistance to change in area, the area elastic modulus (Fung, 1981).

Skin of familiar higher vertebrates is not only two-dimensional and highly extensible, but it is also heterogeneous and often anisotropic at the microscopic level of structure. Skin is a complex composite material whose structural proteins keratin, collagen, and elastin are present in fibers that are surrounded by a microscopically homogeneous matrix of amorphous proteins, glycoproteins, proteoglycans in an aqueous solution of electrolytes, and sometimes, as in whale blubber, lipids. In the skin of fishes, collagen fibers occur in layers parallel to the skin's surface. The size, orientation, and spacing of these fibers and their relative volume fraction varies from layer to layer in each species, and it also varies from place to place over a single animal's body.

This means that skin must be thought of as fiber-reinforced. The stiffness of skin in any direction in the plane of the skin will be proportional to the maximum stiffness parallel to the fiber array times the cosine of the angle between the array and the force axis. When a skin with a biaxial array of fibers, such as the skin of sharks (Wainwright et al, 1978) and all other fishes, is pulled along a line bisecting the angle made by the two sets of fibers, the stress-strain curve is *J*-shaped: modulus is low at low strains and becomes disproportionately higher with large strains over 20%. This important property of fish skin therefore arises from the orientation of fibers in the skin—its substructure. The property results in a safety factor for the skin: the more the skin is stretched, the stiffer it becomes and the less likely it is to tear. Such safety factors are surely among the features that lead to the survival of the species.

### 3. Superstructure: Skin as a Component of the Body

The cat's body, its tail, and each of its legs is roughly cylindrical. So the skin of each of these parts is a membrane formed into a thin-walled tube or hollow cylinder. This body-level, three-dimensional form has implications for the mechanical properties of the skin which, in turn, control properties of the flexing body. All of these tissue- and body-level implications are enhanced if we add an additional mechanical condition: pressure in the cylinder. Since worms of many phyla (Clark, 1964) and sharks (Wainwright et al, 1978) are pressurized cylinders with highly oriented, crossed helical arrays of collagen fibers reinforcing their skin, they are excellent examples to illustrate these implications.

Due only to the geometry of form, the average stress in the skin of a pressurized cylinder in the circumferential direction,  $S_c$ , is twice the average

stress in the longitudinal direction,  $S_L$ :

$$S_C = 2S_L \quad (11.1)$$

This means that any increase in internal pressure will tend to produce an aneurysm in the cylinder. Engineers in their wisdom reinforce pressurized cylinders by wrapping them helically with high-modulus fibers. Worms, sharks, cells in plant stems, and mammalian arteries are similarly reinforced by helically wound, high-modulus fibers (Wainwright et al, 1976). The fiber winding angle will determine exactly how the circumferential and longitudinal stresses will be apportioned. The most important function of the helical fiber array in worms, fishes, and arteries is that it allows these bodies to flex without kinking and wrinkling that would spoil their hydrodynamic surfaces. Note that this function of skin, a tissue, occurs at the much higher level of the organism behaving in its environment.

When sharks swim, their internal body pressure varies with swimming speed (Wainwright et al, 1978). At rest, an 87-cm-long lemon shark (*Negaprion brevirostris*) has a subdermal pressure of 11 kPa above the ambient hydrostatic pressure. When swimming at slow, cruising speed, the pressure rises to 28 kPa, and at fast swimming, pressure reaches peaks of 200 kPa. This is equivalent to the 30 psi in your steel-belted radial car tires.

A further implication of the pressurized cylinder into which the shark skin is formed is that the skin stress is equal to the product of the internal pressure,  $p$ , and the radius,  $r$ , of the cylinder divided by the thickness,  $t$ , of the cylinder wall:

$$S_C = pr/t. \quad (11.2)$$

In a fiber-wound, pressurized cylinder such as the skin of a shark, an increase in pressure caused ultimately by muscular action of blood and other fluids will cause the skin stress to rise considerably without much change in  $r$ ,  $t$ , or skin strain. Thus, since skin stress varies with pressure, so does modulus. The shark's skin, while remaining flexible and undergoing the necessary changes in shape to accommodate the bending of the body, becomes stiffer (increased area expansion modulus) and can more efficiently transmit forces exerted on it by the great swimming muscles. The force of these muscles is thus transmitted from the muscles to the tail via the skin, which has an important tendon-like role in locomotion.

As the longitudinal muscle on one side of the fish contracts and bends the fish's body to that side, the muscle, in preserving constant volume, bulges radially with great force. This forceful radial expansion of contracting muscle does no useful work in the familiar locomotor systems of running lizards, birds, and mammals. But in the shark, it appears that this forceful radial work on the concave side of the fish helps the longitudinal muscles to shorten the skin on that side. In this way, the cylindrical shape of sharks, a body-level feature, and the collagen array in the skin, a tissue-level feature, allow use of the radial work of bulging muscles, an organ-level function, that we higher vertebrates appear to waste in our locomotor systems.

By not being satisfied with studying simply uniaxial tensile properties of

skin, we have come to a whole new view of how sharks swim. Let's see what this approach has taught us as we try to apply it to the problem of how whales swim.

#### 4. Whale Design: Ideas Versus Equations

The great baleen whales swim as sharks do, by wagging their tails to generate thrust. Like sharks, they also have a backbone and a layer of crossed helically wound collagen fibers surrounding the body between the muscle and the overlying blubber and skin. In whales, this collagenous structure is called an *aponeurosis*. There the similarity ends. Sharks wag their tails side to side, while whales wag them up and down—an unimportant difference for the present analysis. The locomotor muscle in sharks is segmentally organized and attaches to the skin at 1- to 2-cm intervals all along the body. Locomotor muscles in the whale run for long distances from their origins behind the head to their insertions at the base of the tail flukes, and yet they bend only the posterior fourth or fifth of the body length. Finally, and most amazingly, only whales have blubber.

The standard notion of blubber is that it resembles beef fat, that is a fatty tissue used to store metabolic energy for the migrations of thousands of miles that these animals make without feeding and to insulate these warm-bodied animals from the infinite heat sink of the sea. Although published accounts (Sokolov, 1960) clearly show that blubber has some collagen fibers in it, this has not yet attracted the attention of physiologists.

We find that the volume fraction of collagen in whale blubber may be as high as 50% and that it is organized in large tendonous strands and layers in a pattern that has yet to be described but that encloses the whole swimming body (Wainwright and Orton, unpublished). Blubber is a thick, anisotropic wall around the cylindrical whale's body (Melville, 1851).

The problem is this: What role(s) do the aponeurosis and the blubber play in the swimming of whales? At this stage of the study, there is no clear answer and there is no constitutive equation that will help us. We must keep watching the kinematics of swimming whales and describing global fiber arrays in the blubber, keep looking for the connections of muscles to the aponeurosis, and keep wracking our brains and picking the brains of engineers for ideas about how complex structures undergo large deformations. We must do this in order to decide which properties we will want to measure. Surely the identification of the interacting mechanisms that produce locomotion is as important a step as the precise quantification of the stress-strain relationship of a single material. When we have accomplished that, then constitutive equations will be the crucial step in the biomechanical analysis (Fung, 1984). Meanwhile, we must be free to think broadly and qualitatively, or we will never arrive at the point where rigorous quantitative analysis will be possible. It should be noted that the whale problem is a metaphor for every scientific investigation.

## References

- Clark, R.B. 1964. *Dynamics in Metazoan evolution*. Oxford University Press, New York.
- Fung, Y.C. 1981. *Biomechanics: mechanical properties of living tissues*. Springer-Verlag, New York.
- \_\_\_\_\_. 1984. *Am. Zool.* 24:13–22.
- Fung, Y.C., Zweifach, B.W., and Intaglietta, M. 1966. *Circ. Res.* 19:441–461.
- Melville, H. 1851. *Moby Dick*, chap. 86. Richard Bentley, London.
- Sokolov, V.E. 1960. *Zool. Zhurnal*. 39:307–308.
- Wainwright, S.A., Biggs, W.D., Currey, J.D., and Gosline, J.M. 1976. *Mechanical design in organisms*. Edward Arnold, London.
- Wainwright, S.A., Vosburgh, F., and Hebrank, J.H. 1978. *Science* 202:747–749.

# 12

## Mechanism of Osmosis: Hulett's Versus Lewis' View of Altered Solvent in Solution

H.T. HAMMEL and A.R. HARGENS

### 1. Introduction

When solute is dissolved in solvent, many properties of the solvent are altered by the solute. The four colligative properties and all partial molar properties of the solution solvent differ from the same properties of pure solvent.

Two views were proposed in the first decade of the 20th century to account for the altered state of the solvent in the solution. Historically, the first and now-forgotten view was proposed by Hulett in 1903, and the familiar view was proposed by Lewis in 1908. Hulett proposed that the solute enhanced the tension in the solution solvent, whereas Lewis proposed that the solute diminished the solvent activity. We shall note that Lewis' view was cleverly contrived. Yet, we may ask, since Hulett's view is based on a physical property of the solvent and can be derived by rigorous thermodynamic argument, why contrive any other view of the altered state of the solvent in the solution, especially if one is led to deny a valid physical view?

### 2. Chemical Potential of Pure Solvent

First, let us review some elementary thermodynamic statements about pure solvent that will be needed to comprehend both views. An important property of pure solvent is its Gibbs molar free energy, and this is known as its chemical potential. Its symbol is:

$\mu_1^{l_0}(T, p^l)$  when applied to pure liquid solvent and  $\mu_1^{v_0}(T, p^v)$  when applied to pure solvent vapor. A change in temperature and a change in pressure applied to the pure liquid changes its chemical potential an amount  $d\mu_1^{l_0} = -\bar{S}_1^{l_0} dT + \bar{V}_1^{l_0} dp^l$ , where  $\bar{S}_1^{l_0}(T, p^l)$  and  $\bar{V}_1^{l_0}(T, p^l)$  are the molar entropy and molar volume of pure liquid and both are functions of  $T$  and  $p^l$ . Likewise, for the vapor phase, a change in temperature and a change in the vapor pressure changes the chemical potential of the vapor an amount  $d\mu_1^{v_0} = -\bar{S}_1^{v_0} dT + \bar{V}_1^{v_0} dp^v$ , where  $\bar{S}_1^{v_0}$  and  $\bar{V}_1^{v_0}$  are the molar entropy and molar volume of the pure vapor.

If, as illustrated in Figure 12.1A and B, the pressure applied to pure sol-

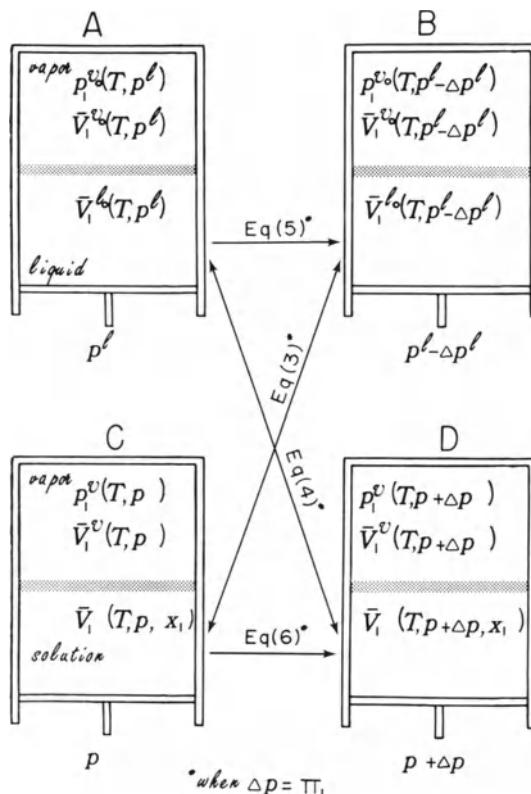


FIGURE 12.1. Equilibrium vapor pressures and molar volumes altered at constant temperature by a change of hydrostatic pressure applied to pure liquid solvent (A to B) and separately, to a solution (C to D). Hydrostatic pressure is applied by a piston at the base of each chamber. A rigid separator between the vapor phase and liquid (A and B) and solution (C and D) phases allows passage of vapor solvent, but not liquid solvent. A and B: Variation of molar volumes of pure liquid and vapor solvent and of vapor pressure as a function of pressure applied to the liquid from  $p^l$  to  $p^l - \Delta p^l$ . C and D: Variation of partial molar volumes of solution solvent and its vapor and of its vapor pressure as a function of pressure applied to the solution from  $p$  to  $p + \Delta p$ . See text for Equations (12.3), (12.4), (12.5) and (12.6).

vent is decreased slowly from  $p^l$  to  $p^l - \Delta p^l$ , the chemical potentials of the pure liquid and vapor must decrease an equal amount, since they remain in equilibrium—i.e.,  $d\mu_1^{l_0} = d\mu_1^v$ . Thus at constant  $T$ ,

$$\int_{\mu_1^{l_0}(T, p^l)}^{\mu_1^{l_0}(T, p^l - \Delta p^l)} d\mu_1^{l_0} \Big|_T = \int_{p^l}^{p^l - \Delta p^l} \bar{V}_1^{l_0} dp^l = \int_{p_1^v(T, p^l)}^{p_1^v(T, p^l - \Delta p^l)} \bar{V}_1^v dp^v. \quad (12.1)$$

The liquid is compressible, and it expands with increasing temperature so that  $\bar{V}_1^{l_0}$  is a complicated function of  $T$  and  $p^l$ . Likewise, the relationship between

$\bar{V}_1^{v_0}$  and  $p^v$  is complicated. Nevertheless, Equation (12.1) is an exact thermodynamic statement, the right-hand equality being the Poynting relation.

### 3. Chemical Potential of Solution Solvent

The chemical potential of the solution solvent is also a function of  $T$  and applied pressure, as well as a function of the mole fraction of solvent in the solution,  $\mu_1(T, p, x_1)$  where  $x_1 = \frac{n_1}{n_1 + n_2}$  for  $n_1$  moles of solution solvent and  $n_2$  moles of solution solute. A change in each parameter changes the chemical potential of the solvent an amount  $d\mu_1 = -\bar{S}_1 dT + \bar{V}_1 dp + \frac{\partial\mu_1}{\partial x_1} dx_1$ , where  $\bar{S}_1$  is the partial molar entropy and  $\bar{V}_1$  is the partial molar volume of the solution solvent. As illustrated in Figure 12.1C and D, the pressure applied to the solution increases from  $p$  to  $p + \Delta p$  at constant  $T$  as the equilibrium vapor pressure of the solution solvent increases from  $p_1^v(T, p)$  to  $p_1^v(T, p + \Delta p)$ . Thus

$$\int d\mu_1 \Big|_T = \int_p^{p+\Delta p} \bar{V}_1 dp = \int_{p_1^v(T, p)}^{p_1^v(T, p+\Delta p)} \bar{V}_1^v dp_1^v. \quad (12.2)$$

Again,  $\bar{V}_1(T, p, x_1)$  is a complicated function of  $T$ ,  $p$  and  $x_1$ , and the exact relationship between  $\bar{V}_1^v$  and  $p_1^v$  is also complicated.

### 4. Hulett's View of the Altered Solution Solvent

$\Pi_1$  is the osmotic pressure of the solvent in the solution. It is the difference in pressures applied to the solution and pure solvent such that the chemical potentials of the solution solvent and pure solvent are equal, i.e.

$$\mu_1(T, p, x_1) = \mu_1^0(T, p - \Pi_1, 1), \quad (12.3)$$

and

$$\mu_1(T, p + \Pi_1, x_1) = \mu_1^0(T, p, 1). \quad (12.4)$$

Integrating Equation (12.1) from  $p$  to  $p - \Pi_1$  at constant  $T$ ,

$$\mu_1^0(T, p - \Pi_1, 1) - \mu_1^0(T, p, 1) = \int_p^{p-\Pi_1} \bar{V}_1^0 dp. \quad (12.5)$$

Integrating Equation (12.2) from  $p$  to  $p + \Pi_1$  at constant  $T$ ,

$$\mu_1(T, p + \Pi_1, x_1) - \mu_1(T, p, x_1) = \int_p^{p+\Pi_1} \bar{V}_1 dp. \quad (12.6)$$

Combining Equations (12.3) and (12.4) with Equations (12.5) and (12.6) gives

$$\int_p^{p+\Pi_1} \bar{V}_1 dp = \int_{p-\Pi_1}^p \bar{V}_1^0 dp. \quad (12.7)$$

Both the solution solvent and pure solvent are compressible, so that  $\bar{V}_1$  and  $\bar{V}_1^0$  decrease as the applied pressure increases by  $\Pi_1$ . Although  $\bar{V}_1$  and  $\bar{V}_1^0$  are not the same at the same  $T$  and  $p$ , Equation (12.7) requires that they be equal when the pressures applied to the solution and pure solvent differ by  $\Pi_1$ . Thus,

$$\bar{V}_1(T, p + \Pi_1, x_1) = \bar{V}_1^0(T, p, 1) \quad (12.8)$$

and

$$\bar{V}_1(T, p, x_1) = \bar{V}_1^0(T, p - \Pi_1, 1).$$

We can also conclude that every partial molar quantity of the solution solvent  $\bar{Z}_1$  equals the molar quantity of the pure solvent  $\bar{Z}_1^0$  at  $T$  when the pressure applied to the solution exceeds the pressure applied to the pure solvent by  $\Pi_1$ , i.e.,

$$\bar{Z}_1(T, p + \Pi_1, x_1) = \bar{Z}_1^0(T, p, 1) \quad (12.9)$$

and

$$\bar{Z}_1(T, p, x_1) = \bar{Z}_1^0(T, p - \Pi_1, 1).$$

The chemical potentials of the solution solvent and pure solvent at the same  $T$  and  $p$  differ by an amount obtained by adding Equation (12.3) and (12.5);

$$\begin{aligned} \mu_1(T, p, x_1) - \mu_1^0(T, p, 1) &= \int_p^{p-\Pi_1} \bar{V}_1^0 dp, \\ \text{and} \qquad \qquad \qquad &= \int_{p_1^0(T, p)}^{p_1^0(T, p - \Pi_1)} \bar{V}_1^0 dp_1^v, \end{aligned} \quad (12.10)$$

according to Equation (12.1).

Hulett's view was that the solute enhances the tension in the solvent by an amount  $\Pi_1$  and thereby lowers the vapor pressure of the solution solvent. The formal thermodynamic argument presented here and elsewhere (Hammel, 1983, 1984a, 1984b), rigorously validates Hulett's view of the altered state of the solvent in the solution.

## 5. Hulett's Gedanken Experiment

Actually, Hulett arrived at his point of view by considering an experiment depicted in Figure 12.2. Hulett recognized that solvent vapor decreases with  $z$  above the pure liquid solvent surface in a gravity field. The law governing the distribution of vapor is a fundamental law of statistical mechanics known as Boltzman's distribution law;

$$N_1(z) = N_1(0) \exp - \left[ \frac{[U_1(z) - U_1(0)]}{kT} \right], \quad (12.11)$$

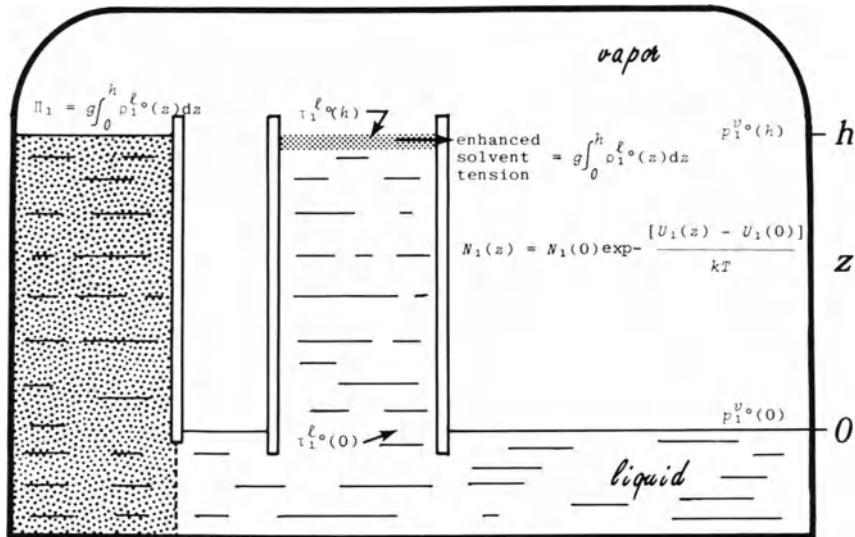


FIGURE 12.2. Hulett's Gedanken experiment, comparing the vapor pressures of the solvent as a function of  $z$  above and in the pure liquid solvent (upright cylinder at center, separated from vapor at top by a matrix) and in the solution solvent (upright cylinder at left, separated from pure liquid solvent at bottom by a membrane permeable to solvent).  $T$  is constant and the same throughout the system.  $N_1(z)$  is the number of vapor molecules in unit cross-section  $dz$  at  $z$ , and  $U_1(z)$  is their energy.

$$-[U_1(z) - U_1(0)]L \simeq \mu_1^{v_0}(z) - \mu_1^{v_0}(0)$$

where  $L$  is Avagadro's constant. Thus

$$RT \ln \frac{N_1(z)}{N_1(0)} \simeq \mu_1^{v_0}(z) - \mu_1^{v_0}(0),$$

and according to the second part of Equation (12.10),

$$RT \int_{N_1(0)}^{N_1(z)} \frac{dN_1}{N_1} \simeq \int_{p_1^{v_0}(0)}^{p_1^{v_0}(z)} \bar{V}_1^{v_0} dp_1^v.$$

Also,

$$\mu_1^{l_0}(h) - \mu_1^{l_0}(0) = \int_{p_1^{v_0}(0)}^{p_1^{v_0}(h) - \Pi_1} \bar{V}_1^{l_0} dp_1^l,$$

where  $\Pi_1 = g \int_0^h \rho_1^l(T, z) dz$  and  $\rho_1^l$  is the density of the liquid solvent.

where  $N_1(z)$  is the number of vapor molecules at  $z$ , and  $U_1(z)$  is the potential energy of each molecule at  $z$  above 0. Hulett also knew that the Poynting relation, Equation 12.1, determined exactly how the vapor pressure of a column of pure solvent liquid must decrease as the tension in the liquid increases with  $z$  from  $\tau_1^l(0)$  to  $\tau_1^l(h)$ . The pure solvent liquid is held at  $h$  by a matrix that passes vapor but does not allow the vapor–liquid interface to pass.

The tension in the liquid at  $h$ , due to the weight of a unit area of the liquid column, is  $g \int_0^h \rho_1^{lo}(z) dz$ , where  $\rho_1^{lo}(z)$  is the density of pure solvent liquid and is a function of  $z$  since all liquids are compressible. The tension in the pure liquid is also a function of  $T$  and applied pressure, where the applied pressure in Figure 12.2 (upright cylinder at center) is the sum of the vapor pressure and the hydrostatic pressure. At  $z = 0$ , the hydrostatic pressure of the pure liquid is 0, so the applied pressure is only the vapor pressure at 0. Thus, at  $z = 0$ ,  $\tau_1^{lo}(0) = \tau_1^{lo}(T, 0) - p_1^{vo}(0)$  where  $\tau_1^{lo}(T, 0)$  is the internal tension in pure liquid solvent at  $T$  and applied pressure 0. At  $z = h$ , the hydrostatic pressure is  $-g \int_0^h \rho_1^{lo}(z) dz$ , so the applied pressure is the vapor pressure plus this hydrostatic pressure and  $\tau_1^{lo}(h) = \tau_1^{lo}(T, 0) + g \int_0^h \rho_1^{lo}(z) dz - p_1^{vo}(h)$ .

Finally, Hulett considered a solution to be separated from pure solvent liquid by a membrane permeable only to the solvent (left column, Figure 12.2). The solute concentration of this solution is such that the solution attained an equilibrium height  $h$ . The osmotic pressure of the solution solvent,  $\Pi_1$ , is the difference in pressures applied to the solution and the pure solvent such that, at the same  $T$ , the solution solvent and pure solvent are in equilibrium—i.e., they have the same chemical potential. The pressure applied to the solution at  $h$  in Figure 12.2 is  $p_1^{vo}(h)$ , whereas the pressure applied to the pure solvent at  $h$  is  $p_1^{vo}(h) - g \int_0^h \rho_1^{lo}(z) dz$ . Since the former pressure exceeds the latter pressure by  $g \int_0^h \rho_1^{lo}(z) dz$ , and since the solution solvent and pure solvent are in equilibrium at  $h$ , then it follows that  $\Pi_1 = g \int_0^h \rho_1^{lo}(z) dz$ .

Since the vapor pressure of the solution solvent at every  $z$  had to be exactly the same as the vapor pressure of pure liquid at the same  $z$  and had also to match the vapor pressure of the vapor above the pure liquid at every  $z$ , Hulett concluded that solvent in the solution column was altered by the solute exactly as if the tension in the solution solvent increased, as does pure solvent in the pure liquid column. Thus, the vapor pressure of solution solvent above the free surface of the solution is identical with all vapor pressures at the same height.

## 6. Lewis' View of the Altered Solution Solvent

Lewis defined an ideal solution as one whose solvent has three properties:

$$p_1^{vo} \equiv RT/\bar{V}_1^{vo} \quad (\text{Ideal Gas Law}) \quad (12.12)$$

$$p_1^v/p_1^{vo} \equiv x_1 \quad (\text{Raoult's Law}) \quad (12.13)$$

$$a_1 \equiv x_1, \quad (12.14)$$

where  $a_1$  is the activity of the solvent in the ideal solution. Combining these definitions with the second part of Equation (12.10) gives

$$\mu_1(T, p, x_1) - \mu_1^o(T, p, 1) = RT \ln a_1. \quad (12.15)$$

Too often, Equation (12.15) is perceived as a fundamental thermodynamic statement. It isn't. It applies only to the solvent of a solution that does not exist. To render Equation (12.15) applicable to the solvent in a real solution, Lewis had to define a dimensionless factor  $\gamma_1$ , the activity coefficient of the solution solvent, such that

$$a_1 \equiv \gamma_1 x_1. \quad (12.16)$$

Lewis also defined another term, the fugacity of the solvent, and designated  $F_1$  as the fugacity of the solution solvent and  $F_1^0$  as the fugacity of the pure solvent. These fugacities are related to the solution solvent activity as follows:

$$a_1 \equiv F_1/F_1^0 \equiv \gamma_1 x_1. \quad (12.17)$$

If one wishes, one may determine the magnitude of  $\gamma_1$  by combining Equations (12.10) and (12.15). Thus

$$RT \ln \gamma_1 x_1 = \int_p^{p-\Pi_1} \bar{V}_1^0 dp. \quad (12.18)$$

This is the essence of Lewis' concept of solvent activity. Its long usage has conferred upon it some convenience and acceptability. However, the activity concept is acceptable only as long as it does not masquerade as an explanation of the lowered chemical potential of the solution solvent. The lowered chemical potential must be attributed to the enhanced tension in the solution solvent. To deny this is to be grossly misled by the activity concept.

## 7. Summary

Physical chemists and biophysicists generally accept Lewis' theory for osmotic pressure, that solute reduces solvent activity on the solution side of a membrane. Historically, however, Hulett's view that solute enhances the tension or negative pressure of solvent in solution predated the Lewis theory. Our purpose is to compare Lewis' and Hulett's proposals in order to determine which theory provides the better physical model to explain osmotic pressure  $\Pi_1$  and other colligative properties.

In order for Lewis to develop an exact thermodynamic relationship between the difference in chemical potentials of pure solvent and solvent in a real solution, he had to invoke a dimensionless activity coefficient  $\gamma_1$  and mole fraction of solvent  $x_1$ . Thus, the Lewis theory provides no explanation of the lowered chemical potential of solvent within a solution. On the other hand, Hulett's view that solute enhances solvent tension by  $\Pi_1$  and lowers the vapor pressure of solvent  $p_1^v$  within a solution provides a rigorous explanation of the lowered chemical potential:

$$\mu_1(T, p, x_1) - \mu_1^0(T, p, 1) = \int_p^{p-\Pi_1} \bar{V}_1^0 dp = \int_{p_1^v(T, p)}^{p_1^v(T, p-\Pi_1)} \bar{V}_1^v dp_1^v$$

where  $\mu_1^0$  and  $\mu_1$  are chemical potentials of pure solvent and solvent in solution, respectively; and  $\bar{V}_1^0$  and  $\bar{V}_1^{v_0}$  are molar volumes of pure solvent liquid and vapor, respectively. Hulett arrived at his point of view by considering the Poynting equation and a Gedanken experiment in which solvent vapor is distributed over a solution according to the Boltzman Law. Thus, the solvent in a solution is altered exactly as pure solvent under tension. Similarly, vapor pressure of solvent above the surface of the solution is identical with all vapor pressures at the same height. These considerations are likewise applied to solutes and their enhanced tension when coupled with solvent in solution. (Hammel, 1984b).

## Acknowledgments

Professor Y.C. Fung was among the first to recognize the validity of the rediscovery by Scholander et al (1965, 1971) of Hulett's view of the altered solution solvent. This work was supported in part by NSF Grant PCM 82-12072.

## References

- Hammel, H.T. 1983. *Phys. Chem. Liq.* 13:25–36.  
\_\_\_\_\_. 1984a. *Phys. Chem. Liq.* 14:171–179.  
\_\_\_\_\_. 1984b. *Phys. Chem. Liq.* 13:255–268.  
Hulett, G.A. 1903. *Z. Phys. Chem.* 42:353–368.  
Lewis, G.N. 1908. *J. Am. Chem. Soc.* 30:668–683.  
Scholander, P.F. 1971. *Microvasc. Res.* 3:215–232.  
Scholander, P.F., Hammel, H.T., Bradstreet, E.D., and Hemmingsen, E.A. 1965. *Science* 148:339–346.

## **Part III Orthopedics**

# 13

## A Finite Deformation Theory for Nonlinearly Permeable Soft Hydrated Biological Tissues

V.C. MOW, M.K. KWAN, W.M. LAI, and M.H. HOLMES

### 1. Introduction

Our work on the biomechanics of diarthrodial joints began in 1969 with the publication of the paper “The Role of Lubrication in Biomechanical Joints” (Mow, 1969), in which we followed Fung’s (1968) urging that, in order “to put biomechanics on a firm foundation, the first important task for the future is to obtain the stress–strain–history relations for biological tissues.” At that time, there was almost no real appreciation of the constitutive laws governing the deformational behaviors of nature’s bearing materials—i.e., cartilage, synovial fluid, or, for that matter, any of the ligaments and tendons surrounding a joint—so that many contradictory lubrication theories (MacConaill, 1932, 1951; Charnley, 1959; McCutchen and Lewis, 1959; Linn, 1968; Walker et al, 1968) and, in retrospect, erroneous models were proposed to describe the mechanism involved in diarthrodial joint lubrication. Thus, we began our attempt to provide the required “firm foundation” by developing valid constitutive equations to describe these bearing materials prior to any further nonproductive speculations on diarthrodial joint lubrication. In order to gain some insight into the motivations for our approach to constitutive modeling of articular cartilage, a brief historical account of this effort is required. (In this chapter, we treat only the subject of modeling the deformational behavior of cartilage and other hydrated soft connective tissues.) The reader is referred to a review chapter by Armstrong and Mow (1980) and review papers by Mow and Lai (1979, 1980) for discussions on diarthrodial joint lubrication mechanisms.

### 2. Historical Background

In the 1920s, Bär (1926) and Göcke (1927) began compressive indentation studies of articular cartilage on the joint surface. They chose the indentation experiment because the anatomical curvature of the surfaces and the thinness of the tissue made other types of experiments very difficult to perform. Hirsch (1944) followed by applying the Hertz solution for contacting elastic bodies to

determine the “Young’s modulus” of articular cartilage, basing the justification of this linearly elastic constitutive assumption on Gildemeister’s (1914) investigation of contacting gels. Because Bär, Göcke, and Hirsch did not perform their tests in physiological solutions, artifactual evaporation effects occurred that led Elmore et al (1963) to investigate the nature of the “imperfect elasticity” of articular cartilage. The latter investigators found that loss of fluid mass due to evaporation prevented the tissue from full recovery, thus explaining the cause of imperfect elasticity observed by the earlier investigators and providing the first well-documented evidence that interstitial fluid is important to the deformational process.

Subsequent to these attempts, Sokoloff (1963, 1966) used the indentation method, based on Hirsch’s 1944 publication, to measure the Young’s modulus E for cartilage from the relation

$$E = \frac{P}{2.67w_0a}, \quad (13.1)$$

where P is the constant applied load,  $w_0$  is the depth of penetration, and a is the radius of the plane-ended cylindrical indenter. This formula comes from the solution of the punch problem for an isotropic, linearly elastic medium of infinite depth, where the result yields  $w_0$  for the penetration

$$w_0 = \frac{P(1 - v)}{4\mu a}, \quad (13.2)$$

where  $\mu$  is the shear modulus. By assuming cartilage to be incompressible (i.e., Poisson’s ratio  $v = \frac{1}{2}$  and Young’s modulus  $E = 3\mu$ ) and solving for E, Equation (13.1) results from Equation (13.2). The analysis for a plane-ended and spherical-ended indenter indenting a layer of linearly elastic compressible material attached to a rigid foundation, which models the cartilage–bone system, was published by Hayes et al (1972). For the case of a plane-ended indenter, these investigators found that the Young’s modulus is given by

$$E = \frac{P(1 - v^2)}{2w_0a\kappa(a/h, v)}, \quad (13.3)$$

where the function  $\kappa$  comes from a solution of the integral equation, which depends strongly on  $v$  as well as on the aspect ratio ( $a/h$  = radius of the indenter/cartilage thickness). Almost simultaneously, Kempson et al (1971) adopted an empirical approach by using Waters’ (1965) formula for indentation of thin sheets of rubber, where again  $v = \frac{1}{2}$  was assumed. Waters’ empirical formula is of the same form as Equation (13.3). Kempson chose the deformation at 2 s to determine the “two-second modulus” using this approach.

The hypothesis that cartilage is purely elastic applies, at best, only to the initial instant after the application of load and at equilibrium. In every in-

dentation test, after the sudden application of load, a rapid compression takes place, followed by a slow creep process toward equilibrium. We now know that creep can occur for over 20,000 seconds before equilibrium is reached. Thus, one is at a loss to ascribe any real meaning to the "Young's modulus" or the "two-second modulus," determined using either Equation (13.1) or Equation (13.3) by all these investigators. Although this shortcoming was elegantly noted in the Zarek and Edwards (1964) and Edwards (1967) studies on cartilage permeability and confined-compression creep, which suggested that creeping of cartilage may be due to fluid exudation, this line of investigation was not pursued. Instead, a number of investigators in the 1970s used various linear viscoelastic constitutive laws to describe the indentation and the confined-compression creep behavior of cartilage (Hayes and Mockros, 1971; Coletti et al, 1972; Parsons and Black, 1977, 1979). The major difficulty with this approach is that it ignores the real nature of these tissues. Articular cartilage and other soft connective tissues are highly hydrated, containing between 70% and 90% water by wet weight. The movement of this fluid through the interstitial space should be expected to influence strongly the deformational behavior of these tissues.

During the 1960s, reports on measurements of the permeability of cartilage and water movement through cartilage began to appear. These studies included those by McCutchen (1962), Edwards (1967), and Linn and Sokoloff (1965). In the first two instances, direct permeability measurements were made—i.e., a pressure differential was applied across a flat specimen; in the latter case, a constant compressive load was applied onto the tissue. McCutchen and Edwards obtained permeability coefficients of the order of  $10^{-16} \text{ m}^4/\text{N}\cdot\text{s}$  using a  $\Delta P \sim 0.1$  megapascals (MPa) and 0.5 MPa, respectively. Linn and Sokoloff did not measure permeability, but rather the residual fluid content after a sustained period of compressive loading. Later, permeability measurements were extended by Maroudas (1968, 1975a, b), Mulholland et al (1975), and Mansour and Mow (1976). Maroudas showed that permeability is inversely related to the fixed-charge density within the tissue—i.e., the proteoglycan content and degree of tissue hydration—and varies with depth, while Mulholland et al demonstrated that cartilage permeability might be anisotropic. Mansour and Mow (1976) showed that cartilage permeability decreases with increasing pressure differential (pressures as high as 2.8 MPa were used) and with the compressive strain used to clamp the specimen into the permeability apparatus. All these studies show that: (1) articular cartilage is permeable, (2) frictional drag associated with interstitial fluid flow might exert an important influence on the load-deformational behaviors of these tissues, and (3) fluid transport through the interstitium might be nonlinearly coupled to the deformational response of the tissue in compression. With this background, we began our development of a multiphasic theory to describe the fluid flow and mechanical properties of articular cartilage and other hydrated biological tissues.

### 3. Biphasic Model for Articular Cartilage

Articular cartilage is the dense connective tissue, 1–5 mm thick, that covers the bone (Figure 13.1). It is a multiphasic, nonlinearly permeable, viscoelastic material, consisting of two principal phases: a solid phase, composed primarily of collagen arranged in a specific fibrillar network (Ghadially, 1978) and proteoglycans in the aggregated form (Muir, 1980; Mow et al, 1984b), and a movable interstitial fluid phase, approximately 78% by wet weight, which is predominantly water (Linn and Sokoloff, 1965; Venn and Maroudas, 1977; Muir, 1980; Armstrong and Mow, 1982; Torzilli et al, 1982) (Figure 13.2). We also know that the movement of this fluid through the porous matrix and the consequent frictional interaction between the two phases controls the compressive “viscoelastic” behavior of the tissue as a whole and is the predominant mechanism of energy dissipation (Mow et al, 1980; Holmes et al, 1984a, b, c; Mow et al, 1984a). These fundamental and simple concepts may now be summarized coherently to describe the stress-strain-history relationship for articular cartilage or for any other biphasic tissue (meniscus, nasal cartilage, ligaments, etc.) of the body. Thus, the objective of this chapter is to present a brief summary of our current understanding of this biphasic material and some recent advances utilizing finite-deformation theory to describe the deformational behavior of the nonlinearly permeable porous solid organic matrix.

#### 3.1. Continuity Equations

The development of the biphasic theory\* for articular cartilage was begun by Torzilli and Mow (1976a, b) and extended by Mow and Lai (1979) and Mow et al (1980), based on the multiphasic mixture theory of Craine et al (1970) and Bowen (1976).† Our model depicts the tissue as a soft, porous, and permeable elastic solid filled with water, with both phases assumed to be intrinsically incompressible. For such a binary mixture, the continuity equation is

$$\operatorname{div} \mathbf{v}^f + \alpha \operatorname{div} \mathbf{v}^s + \alpha(\mathbf{v}^s - \mathbf{v}^f) \cdot \operatorname{grad} \ln \rho^s = 0, \quad (13.4)$$

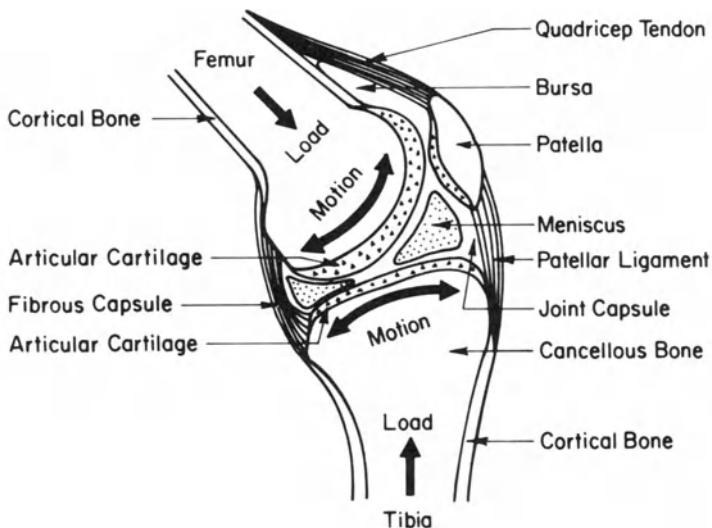
where

$$\alpha = \frac{\rho^s(t)\rho_T^f(0)}{\rho^f(t)\rho_T^s(0)} = \frac{dV^s(t)}{dV^f(t)} \quad (13.5)$$

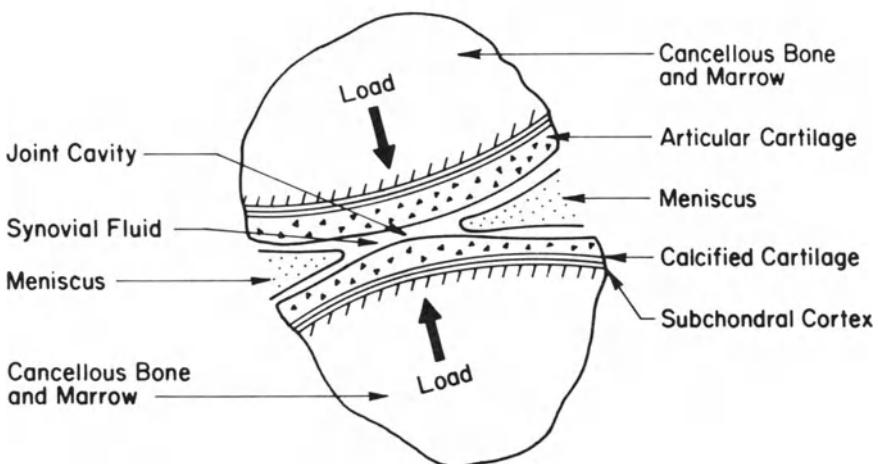
---

\* It should be pointed out that, even though a two-phase model is adequate for the description of the flow and deformation resulting from mechanical loading, a more involved model of articular cartilage would be multiphasic, taking into account the influence of the mobile electrolytes on the charged collagen–proteoglycan solid matrix (Myers and Mow, 1983; Myers et al, 1984).

† While poroelastic theories have been developed by Biot for soil, modern mixture theory is preferred because it is more consistent—especially when nonlinear permeability and finite-deformational effects occur.

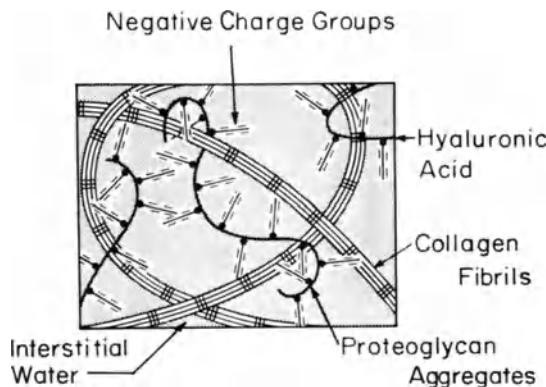
**a**

Human Knee Joint

**b**

Load Bearing Region

FIGURE 13.1. (a) Diagrammatic representation of a human knee joint, showing the active loads, degree of motion, and the major load-bearing elements—articular cartilage, meniscus, and synovial fluid within the joint—forming a magnificent lubrication system. (b) Enlarged view of the load-bearing region within the knee, showing interaction between articular cartilage, meniscus, and synovial fluid. During knee function, 50–80% of the total compression load is transmitted through the meniscus.



**FIGURE 13.2.** Schematic representation of articular cartilage depicting a porous, composite organic solid matrix swollen with water, ~78% by wet weight. The major organic components of the solid matrix are collagen fibrils, ~60% by dry weight, and proteoglycan aggregates, ~30% by dry weight. Electrostatic and mechanical interactions between proteoglycans and collagen determine the strength, cohesiveness, and material properties of the solid matrix.

is defined as the “solid content.” In these equations,  $v^{f,s}$  are the velocities of the fluid and solid phases,  $\rho^{f,s}(t)$  are the apparent densities (i.e., per total volume), and  $\rho_T^{f,s}(0)$  are the initial true densities (i.e., per constituent volume), which are assumed to be homogeneous. Equation (13.4) may also be re-written as

$$\operatorname{div}[(1 + \alpha)^{-1}(v^f + \alpha v^s)] = 0. \quad (13.6)$$

Details of derivation for Equations (13.4) through (13.6) may be found in the appendix. Equation (13.4) or (13.6) describes the principle of conservation of mass for a mixture of a fluid phase and a solid phase in which each phase is intrinsically incompressible (e.g., water by itself is intrinsically incompressible). Equation (13.6) has a simple interpretation. Since  $1/(1 + \alpha) = dV^f(t)/dV(t)$  is the porosity and  $\alpha/(1 + \alpha) = dV^s(t)/dV(t)$  is the solidity,  $\operatorname{div}(v^f/(1 + \alpha))$  is the net loss of fluid volume,  $\Delta(dV^f(t))$ , per unit mixture volume ( $dV(t)$ ), and  $\operatorname{div}(\alpha v^s/(1 + \alpha))$  is the net loss of solid volume,  $\Delta(dV^s(t))$ , per unit mixture volume. Obviously, the total loss of volume must be zero since both phases are incompressible. For finite deformation, it can be shown (see Appendix) that the solid content  $\alpha$  satisfies the equation

$$\alpha = \frac{\alpha_0}{(1 + \alpha_0)\sqrt{\det \mathbf{B}} - \alpha_0}, \quad (13.7)$$

where  $\alpha_0 = dV^s(0)/dV^f(0)$  is the initial solid content and  $\mathbf{B}$  is the left Cauchy-Green deformation tensor (see Equation (13.9a) below).

### 3.2. Constitutive Equations for Finite Deformation

For a binary mixture in which both phases of the mixture are intrinsically incompressible, it can be shown that the apparent stress  $\sigma^s$  of the solid phase is given by

$$\sigma^s = -\alpha p \mathbf{I} + \rho^s (A^s - A) \mathbf{I} + \rho \mathbf{F}^s \left( \frac{\partial A}{\partial \mathbf{F}^s} \right)^T \quad (13.8a)$$

and the apparent stress  $\sigma^f$  of the fluid phase is given by

$$\sigma^f = -p \mathbf{I} - \rho^s (A^s - A) \mathbf{I} - \rho \rho^f \left( \frac{\partial A}{\partial \rho^f} \right) \mathbf{I}. \quad (13.8b)$$

Here  $\mathbf{F}^s$  is the deformation gradient for the solid matrix,  $A$  is the mixture's Helmholtz free-energy-per-unit-mixture mass (defined in terms of the Helmholtz free energy for each of the constituents  $A^{f,s}$  for the fluid and solid phases, respectively:  $\rho A = \rho^f A^f + \rho^s A^s$ ),  $\rho$  is the total density given by  $\rho = \rho^s + \rho^f$ , and  $p$  is the interactive fluid pressure arising from the incompressibility condition. For binary mixtures, the coupling of motion between the fluid phase and the solid phase involves a diffusive drag that occurs as an equal and opposite pair of body forces in the linear momentum equations governing each of the phases. This diffusive drag force is given by

$$\begin{aligned} -\boldsymbol{\pi}^s = \boldsymbol{\pi}^f &= \rho^s \frac{\partial A}{\partial \rho^f} \operatorname{grad} \rho^f - \rho^f \frac{\partial A}{\partial \mathbf{F}^s} [\operatorname{grad} \mathbf{F}^s] - \alpha p \operatorname{grad} \ln \rho^s \\ &\quad + \operatorname{grad} [\rho^s (A^s - A)] + \mathbf{K} (\mathbf{v}^s - \mathbf{v}^f), \end{aligned} \quad (13.8c)$$

where  $\mathbf{K}$  is the diffusive drag coefficient of relative motion.

An alternative representation of the binary mixture constitutive equations, Equations (13.8a) and (13.8b), may be obtained by expressing the Helmholtz free energies in terms of the left Cauchy–Green deformation tensor  $\mathbf{B}$  given by

$$\mathbf{B} = \mathbf{F}^s (\mathbf{F}^s)^T. \quad (13.9a)$$

By the chain rule of differentiation for tensors, the partial derivative in Equation (13.8a) may be expressed as

$$\rho \mathbf{F}^s \left( \frac{\partial A}{\partial \mathbf{F}^s} \right)^T = \rho \mathbf{B} \left[ \frac{\partial A}{\partial \mathbf{B}} + \left( \frac{\partial A}{\partial \mathbf{B}} \right)^T \right], \quad (13.9b)$$

or by expressing Equation (13.8a) and (13.8b) in terms of  $A^s$  and  $A^f$ , it can be easily show that

$$\sigma^s = -\alpha p \mathbf{I} + 2\mathbf{B} \left\{ \rho^s \frac{\partial A^s}{\partial \mathbf{B}} + \rho^f \frac{\partial A^f}{\partial \mathbf{B}} \right\}, \quad (13.10a)$$

$$\sigma^f = -p \mathbf{I} - \rho^f \left\{ \rho^s \frac{\partial A^s}{\partial \rho^f} + \rho^f \frac{\partial A^f}{\partial \rho^f} \right\} \mathbf{I}. \quad (13.10b)$$

These equations are the incompressible counterpart of, but otherwise identical to, those derived by Bowen (1976) for his compressible “porous media model.” Bowen (1980) derived a set of equations for incompressible binary mixtures that are identical to these equations. In this symmetric form, we see that, for binary mixtures, the Helmholtz free energy of each phase contributes to the stress acting on each phase of the mixture in an expected manner similar to a single-phase continuous medium.

For soft hydrated tissues such as articular cartilage and meniscus, it is reasonable to expect an isotropic response in compression (Mow et al, 1980; Lai et al, 1981; Armstrong and Mow, 1982). In this case, the dependence of the Helmholtz free-energy function on  $\mathbf{B}$  reduces to a dependence on the three invariants of  $\mathbf{B}$ :  $J_1$ ,  $J_2$ , and  $J_3$  where

$$J_1 = \text{tr } \mathbf{B}, \quad J_2 = \frac{1}{2}[(\text{tr } \mathbf{B})^2 - \text{tr}(\mathbf{B}^2)], \quad J_3 = \det \mathbf{B}. \quad (13.11)$$

Further, by assuming that  $A^s = A^s(J_1, J_2, J_3)$  and  $A^f = A^f(J_1, J_2, J_3)$  only, we obtain from Equations (13.8a), (13.8b), and (13.9b):

$$\boldsymbol{\sigma}^s = -\alpha p \mathbf{I} + \rho^s(A^s - A)\mathbf{I} + 2\rho \left\{ \left( J_2 \frac{\partial A}{\partial J_2} + J_3 \frac{\partial A}{\partial J_3} \right) \mathbf{I} + \frac{\partial A}{\partial J_1} \mathbf{B} - J_3 \frac{\partial A}{\partial J_2} \mathbf{B}^{-1} \right\}, \quad (13.12a)$$

$$\boldsymbol{\sigma}^f = -p \mathbf{I} - \rho^s(A^s - A)\mathbf{I}, \quad (13.12b)$$

$$-\boldsymbol{\pi}^s = \boldsymbol{\pi}^f = -\alpha p \text{grad} \ln \rho^s + \text{grad}[\rho^s(A^s - A)] + \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f). \quad (13.12c)^*$$

### 3.3. Equations of Motion

For the binary mixture, each phase must satisfy an equation of motion:

$$\rho^{f,s} \mathbf{a}^{f,s} = \text{div } \boldsymbol{\sigma}^{f,s} + \rho^{f,s} \mathbf{b}^{f,s} + \boldsymbol{\pi}^{f,s}. \quad (13.13)$$

Here  $\mathbf{b}^{f,s}$  are the body forces per unit mass, and  $\mathbf{a}^{f,s}$  are the accelerations for the fluid and solid phases, respectively. Under quasi-static conditions, and neglecting the body forces  $\mathbf{b}^{f,s}$ , we have

$$\text{div } \boldsymbol{\sigma}^{f,s} + \boldsymbol{\pi}^{f,s} = \mathbf{0}. \quad (13.14)$$

Substituting Equations (13.12) into Equation (13.14), we obtain the governing equations for our biphasic material:

$$\begin{aligned} & -\text{grad}(\alpha p) + \alpha p \text{grad} \ln \rho^s - \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) \\ & + \text{div} 2\rho \left\{ \left[ J_2 \frac{\partial A}{\partial J_2} + J_3 \frac{\partial A}{\partial J_3} \right] \mathbf{I} + \frac{\partial A}{\partial J_1} \mathbf{B} - J_3 \frac{\partial A}{\partial J_2} \mathbf{B}^{-1} \right\} = \mathbf{0} \end{aligned} \quad (13.15a)$$

for the solid phase and

---

\* To simplify calculation, the term involving  $[\text{grad } \mathbf{F}^s]$  in  $\boldsymbol{\pi}^s$  and  $\boldsymbol{\pi}^f$  is neglected.

$$-\operatorname{grad} p - \alpha p \operatorname{grad} \ln \rho^s + \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) = \mathbf{0} \quad (13.15b)$$

for the fluid phase. Note that for homogeneous true densities, i.e.,  $\rho_T^{f,s}(0)$  being constants, the following relationship between  $\alpha$  and  $\rho^s$  exists:

$$\operatorname{grad} \alpha = \alpha(1 + \alpha) \operatorname{grad} \ln \rho^s. \quad (13.16)$$

Using Equation (13.16), Equations (13.15a) and (13.15b) may be rewritten as

$$\begin{aligned} & -\frac{\alpha}{1 + \alpha} \operatorname{grad}[(1 + \alpha)p] - \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) \\ & + \operatorname{div} 2\rho \left[ \left( J_2 \frac{\partial A}{\partial J_2} + J_3 \frac{\partial A}{\partial J_3} \right) \mathbf{I} + \frac{\partial A}{\partial J_1} \mathbf{B} - J_3 \frac{\partial A}{\partial J_2} \mathbf{B}^{-1} \right] = \mathbf{0}, \end{aligned} \quad (13.17a)$$

$$-\frac{1}{1 + \alpha} \operatorname{grad}[(1 + \alpha)p] + \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) = \mathbf{0}. \quad (13.17b)$$

The term  $(1 + \alpha)p$  represents the true fluid pressure acting inside the mixture. Eliminating  $(1 + \alpha)p$  from Equations (13.17a) and (13.17b), we obtain one governing equation for finite deformation of an isotropic biphasic material:

$$\operatorname{div} 2\rho \left\{ \left[ J_2 \frac{\partial A}{\partial J_2} + J_3 \frac{\partial A}{\partial J_3} \right] \mathbf{I} + \frac{\partial A}{\partial J_1} \mathbf{B} - J_3 \frac{\partial A}{\partial J_2} \mathbf{B}^{-1} \right\} - (1 + \alpha) \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) = \mathbf{0}. \quad (13.18)$$

### 3.4. Infinitesimal-Strain Theory

Under strictly controlled experimental conditions, we may prescribe the strain field to be small. Under these conditions, it would be convenient to have the infinitesimal-strain theory corresponding to Equation (13.18). For this case, the deformation tensor  $\mathbf{B}$  and other variables reduce to

$$\begin{aligned} \mathbf{B} &\approx \mathbf{I} + 2\mathbf{e}, \quad \mathbf{B}^{-1} \approx \mathbf{I} - 2\mathbf{e}, \quad J_1 \approx 3 + 2e, \\ J_2 &\approx 3 + 4e, \quad J_3 \approx 1 + 2e, \quad \alpha \approx \alpha_0 [1 - (1 + \alpha_0)e], \\ \rho^s &\approx \rho_0^s(1 - e), \quad \rho \approx \rho_0(1 - C_0e), \\ C_0 &= \alpha_0(1 - d_0)/(\alpha_0 + d_0) \quad \text{and} \quad d_0 = \rho_T^f(0)/\rho_T^s(0), \end{aligned} \quad (13.19)$$

where  $\mathbf{e}$  is the infinitesimal-strain tensor of the solid matrix and  $e$  is the dilatation given by trace  $\mathbf{e}$ . By inserting Equation (13.19) into Equations (13.12a), (13.12b), and (13.12c) we obtain the constitutive equations for this biphasic material corresponding to the infinitesimal-strain theory

$$\boldsymbol{\sigma}^s = -\alpha p \mathbf{I} + \lambda_s e \mathbf{I} + 2\mu_s \mathbf{e}, \quad (13.20a)$$

$$\boldsymbol{\sigma}^f = -p \mathbf{I}, \quad (13.20b)$$

and

$$-\boldsymbol{\pi}^s = \boldsymbol{\pi}^f = -\alpha p \operatorname{grad} \ln \rho^s + \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f), \quad (13.20c)$$

where  $\lambda_s$  and  $\mu_s$  are the Lamé coefficients for the elastic solid matrix. Similarly, the equation of motion, Equation (13.18), yields

$$\operatorname{div}(\lambda_s e \mathbf{I} + 2\mu_s \mathbf{e}) - (1 + \alpha) \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) = \mathbf{0}. \quad (13.21)$$

For the case of constant  $\alpha_0$ , the governing equation associated with Equation (13.21) becomes

$$\operatorname{div}(\lambda_s e \mathbf{I} + 2\mu_s \mathbf{e}) - (1 + \alpha_0) \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f) = \mathbf{0}. \quad (13.22)$$

This equation may be easily derived from the assumed simplified constitutive equations for the solid matrix:

$$\boldsymbol{\sigma}^s = -\alpha_0 p \mathbf{I} + \lambda_s e \mathbf{I} + 2\mu_s \mathbf{e}, \quad (13.23a)$$

for the interstitial fluid:

$$\boldsymbol{\sigma}^f = -p \mathbf{I}, \quad (13.23b)$$

and for the diffusive drag body force:

$$-\boldsymbol{\pi}^s = \boldsymbol{\pi}^f = \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f). \quad (13.23c)$$

These equations define what is generally known as the KLM biphasic theory for articular cartilage (Mow and Lai, 1980), and they have been used to describe the deformational behavior of other hydrated collagenous tissues as well (Mow and Schoonbeck, 1982; Favenesi et al, 1983).

## 4. Nonlinear Problems for Cartilage in Compression

We now utilize this theory to describe two different classes of nonlinear material behavior. The first to be examined are the nonlinearities arising from the strain-dependent permeability effects under infinitesimal-strain conditions (Mansour and Mow, 1976; Lai and Mow, 1980; Mow et al, 1984a). The second to be examined are the nonlinearities associated with the finite deformation of the solid matrix (Mow et al, 1980; Lai et al, 1981; Armstrong et al, 1979). Thus, in this chapter, we treat each of the two types of nonlinearities of Equation (13.18) separately. A complete treatment of Equation (13.18), where both nonlinearities occur simultaneously, needs further scrutiny and is currently being pursued in our laboratory. The merit of artificially separating these nonlinearities, however, is that each may be examined for its effects on the deformational response of the tissue. Under strictly controlled material-testing conditions, it is possible to produce experimentally the first class of nonlinear effects (see next section).

### 4.1. Nonlinear Permeability Effects—Creep

In compression, the time-dependent behavior of this type of permeable hydrated tissue is dominated by the friction interaction between the fluid phase

and the solid phase as they flow relative to each other. The experiment we have used to quantify the effect of fluid flow on the compressive viscoelastic property of such tissues as articular cartilage is the uniaxial confined-compression test. In this experiment, an osteochondral plug is confined along its lateral boundary and compressed by a rigid, porous free-draining filter at the surface. The prescribed loading on the filter has included creep tests (Mow et al, 1984a; Armstrong and Mow, 1982), stress-relaxation tests (Lai et al, 1981; Holmes et al, 1985), and low-frequency oscillatory tests (Holmes et al, 1984b, c). This confined configuration is used because the theoretical solution of a one-dimensional model approximates the test very closely. We have found this to be an accurate and repeatable method for determining the material coefficients of cartilage and other permeable hydrated tissues.

The macroscopic measure of the diffusive resistance of the fluid flow through the porous solid matrix is the permeability. Lai and Mow (1980) have shown that the permeability ( $k$ ) of a permeable media, under the condition of slow flow, is related to the diffusive drag coefficient ( $K$ ) by

$$k = \frac{1}{(1 + \alpha_0)^2 K}, \quad (13.24)$$

where  $\alpha_0$  is the initial solid content, which is a constant. The variation of the permeability ( $k$ ) with strain was first measured by Mansour and Mow (1976); later, this nonlinearity was determined by Lai and Mow (1980) to be

$$k = k_0 \exp(eM), \quad (13.25)$$

where  $e$  is the dilatation of the solid phase, and  $k_0$  and  $M$  are intrinsic material parameters. For normal tissues,  $M$  may range up to 20, even at small compressive strains. This strong nonlinearity results in a significant decrease in the permeability, even with small compression. The KLM biphasic theory for the uniaxial confined-compression tests, incorporating this nonlinear permeability effect, reduces to the following *nonlinear diffusion equation* for the axial displacement  $u(z, t)$  of the solid phase (Holmes, 1984; Mow et al, 1984a):

$$H_A \frac{\partial^2 u}{\partial z^2} = \frac{1}{k_0} \exp\left(-M \frac{\partial u}{\partial z}\right) \frac{\partial u}{\partial t}, \quad 0 < z < h, \quad (13.26)$$

where  $H_A$  is the aggregate elastic modulus given by  $\lambda_s + 2\mu_s$  and  $h$  is the thickness of the tissue. The associated rate of energy dissipation is given by

$$\Phi(z, t) = \left(\frac{\partial u}{\partial t}\right)^2 \frac{1}{k_0} \exp\left(-M \frac{\partial u}{\partial z}\right). \quad (13.27)$$

In this model, the only energy dissipation is that caused by the diffusive drag( $\pi^{f,s}$ ) of relative motion. We see that, in addition to the energy dissipation increase with velocity, Equation (13.27) shows that energy dissipation increases exponentially with compressive strain. This clearly illustrates the important role of the nonlinear fluid-solid interaction taking place within the tissue.

Equation (13.26) is obviously very difficult to solve in closed form; therefore, it is necessary to use either asymptotic or numerical approximations of the solution. To illustrate the accuracy of the relatively simple form of our asymptotic solutions, let us consider the case of a confined-compression creep test where a stress  $F_0$  is applied to the articular surface via a rigid, porous free-draining filter. The appropriate boundary conditions are

$$H_A \frac{\partial u}{\partial z}(0, t) = -F_0 \quad \text{for } t > 0, \quad (13.28a)$$

$$p = 0, \quad (13.28b)$$

and at the cartilage–bone interface we assume

$$u(h, t) = 0. \quad (13.28c)$$

The initial condition is

$$u(z, 0) = 0. \quad (13.28d)$$

In the first few moments ( $\sim 100$  s) after the application of the load, the deformation diffuses from the articular surface to the cartilage–bone interface ( $z = h$ ). Until it effectively reaches this interface, we may assume  $0 < z < \infty$  for the domain of Equation (13.26), rather than  $0 < z < h$  (Holmes, 1984). In this case, there is a similarity solution for Equations (13.26) and (13.28) given by

$$u(z, t) = \varepsilon_0 \sqrt{k_0 H_A t} f\left(\frac{z}{\sqrt{k_0 H_A t}}\right), \quad (13.29)$$

where  $f(s)$  satisfies

$$2f''(s) = \exp(-\varepsilon_0 M f')(f - sf'), \quad 0 < s < \infty \quad (13.30a)$$

with

$$f'(0) = -1, \quad f(\infty) = 0, \quad (13.30b)$$

and

$$\varepsilon_0 = \frac{F_0}{H_A}. \quad (13.31)$$

Although Equations (13.30a) and (13.30b) are nonlinear, they form a set of ordinary differential equations that can be readily solved by numerical methods. More importantly, Equation (13.29) shows that the surface displacement is given by

$$u(0, t) = \beta_0 \sqrt{t}, \quad 0 \leq t \ll \frac{h^2}{k_0 H_A}, \quad (13.32a)$$

where

$$\beta_0 = f(0) F_0 \sqrt{\frac{k_0}{H_A}}. \quad (13.32b)$$

The dependence of  $\beta_0$  on the material parameters is known, except for  $M$ , which requires the solution of Equations (13.30a) and (13.30b). An asymptotic solution of this problem can be found for the case of a small strain  $\varepsilon_0$  by expanding the solution as follows:

$$f(s) \sim f_0(s) + \varepsilon_0 M f_1(s) + \dots \quad (13.33)$$

Inserting this into Equation (13.30a), equating like powers of  $\varepsilon_0$ , and solving the resulting linear problems, Holmes (1984) found that

$$f_0(s) = \frac{2}{\sqrt{\pi}} \left[ e^{-s^2/4} - \frac{1}{2}s\sqrt{\pi} \operatorname{erfc}\left(\frac{s}{2}\right) \right], \quad (13.34a)$$

$$f_1(s) = \left[ \frac{1}{2} \operatorname{erfc}\left(\frac{s}{2}\right) - \frac{2+\pi}{2\pi} \right] f_0(s), \quad (13.34b)$$

and

$$f(0) \sim \frac{2}{\sqrt{\pi}} \left( 1 - \frac{\varepsilon_0 M}{\pi} \right). \quad (13.34c)$$

Thus the early time response of the surface displacement in confined-compression creep is now completely determined. The accuracy of this approximation is shown in Figure 13.3, which was obtained by regressing our

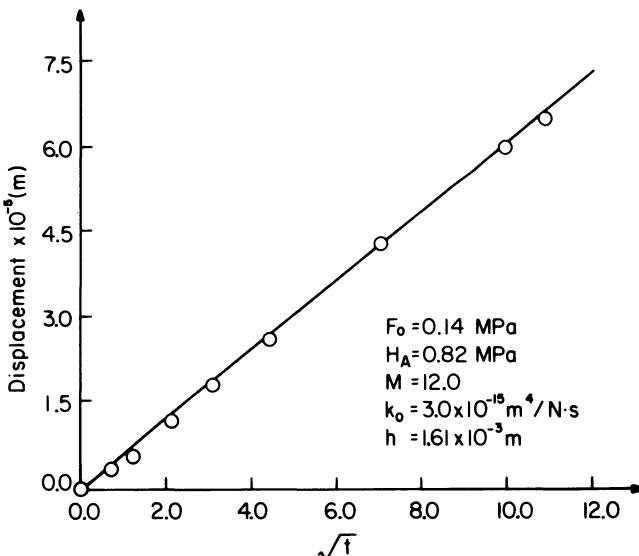


FIGURE 13.3. Creep displacement of a bovine patellofemoral articular cartilage sample under confined compression versus  $\sqrt{t}$  for small time. The regression yields  $M$  and  $k_0$ , while  $H_A$  is determined by the equilibrium displacement at 10,000 s. The values for  $M$  and  $k_0$  are similar to those obtained from a confined-compression stress-relaxation test.

asymptotic solution, Equations (13.32a) and (13.32b), with a corresponding experimental creep result on a bovine patello femoral cartilage sample. Clearly this is a good numerical fit, with all the values of the material constants close to those obtained by other methods.

#### 4.2. Nonlinear Permeability Effects—Steady Cyclic Excitation

Another way to determine the material coefficients, and one that is recommended by Fung (1981), is by using an oscillatory loading condition. We have performed these tests to evaluate the energy dissipation capacity of the tissue (Holmes et al, 1984c). In these experiments, we apply an offset compressive strain of 10% and then precondition the plug (Fung, 1981) prior to our oscillatory test. After the sample has equilibrated in our testing apparatus, a prescribed periodic surface displacement, sinusoidal or sawtooth, is imposed (Holmes et al, 1984c). The hysteresis loop, obtained from a test on a sample of bovine articular cartilage equilibrated in water, corresponding to a sawtooth displacement input with a frequency of  $10^{-3}$  and a peak-to-peak amplitude of 2% strain is shown in Figure 13.4. The agreement between theory and experiment is seen to be remarkably good at this frequency. The area within this

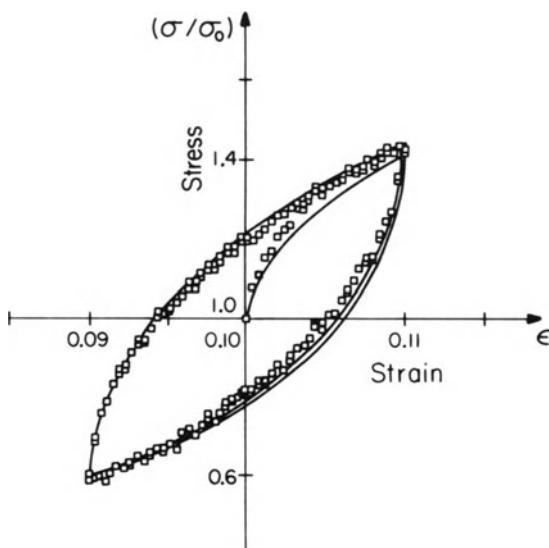


FIGURE 13.4. Energy dissipation associated with a 2% peak-to-peak sawtooth displacement about a 10% d.c. offset strain. For this sample, total energy dissipated in one cycle is  $1.69 \text{ mol} \times 10^{-5} \text{ J}$ . Thus, at low frequencies ( $10^{-3} \text{ Hz}$ ), this nonlinear permeability fluid flow dissipation mechanism is the predominant mechanism of energy dissipation in cartilage.

hysteresis loop is indicative of the total energy dissipated by the tissue within one loading cycle. For the case of a sinusoidal loading function, our asymptotic solution also agrees very well with our experimental results (Figure 13.5) The corresponding theoretical values obtained from a numerical solution of Equation (13.26) are also shown in Figure 13.5. The theoretical result shows a dependence of the energy dissipation on the driving frequency. Figure 13.6 shows the variation of energy dissipation with frequency for tissues that have been equilibrated in de-ionized water and 0.15 mol NaCl. The value of energy dissipation for the sample shown in Figure 13.4 is for bovine cartilage equilibrated in water. Our results show that energy dissipation increases for the lower frequencies ( $f < 10^{-2}$  Hz) because the fluid velocities in the tissue increase with frequency. However, apparently due to the existence of an intense boundary layer in the deformation field within the tissue at higher strain rates (Holmes et al, 1985), the energy dissipation decreases for the higher frequencies ( $f > 10^{-2}$  Hz), thus deviating from our infinitesimal-strain theory prediction. The increase in the energy dissipation with decreasing salt concentration might be due to a decrease of the charge-shielding effect of the counter-ion concentration in the interstitium. As the concentration of NaCl decreases, proteoglycans within the tissues expand (Myers et al, 1984). This might actually cause an increase of  $\alpha$ , thus a decrease of the porosity of the tissue. From Equation (13.24), decrease of the permeability will occur with increasing  $\alpha$ , thus increasing the energy dissipation.

We have shown that the solid matrices of hydrated cartilaginous tissue are very soft,  $H_A = O(1.0 \text{ MPa})$ , and that the permeability of these hydrated tissues is also very low,  $k = O(10^{-16} \text{ m}^4/\text{N} \cdot \text{s})$ . With moderately high strain rates (e.g.,  $> 10\%$  in 100 s) or frequency of excitation (e.g.,  $> 0.1$  Hz) and moderately high applied tractions (e.g.,  $> 0.2$  MPa), very large local strains will exist. Thus, the infinitesimal-strain theory is really very restrictive—prohibiting the use of the biphasic theory for any loading condition resembling those found under physiological situations. For this, we have developed our finite-deformation biphasic theory.

## 5. Finite-Deformation Theory—Helmholtz Energy Functions

Equations (13.12a), (13.12b), and (13.12c) define the general stress-strain law for a binary mixture of an incompressible, isotropic, nonlinearly elastic solid and an incompressible fluid in terms of the Helmholtz free-energy functions  $A$ ,  $A^s$ ,  $A^f$ . In this section we specialize two choices for  $A$ :

$$\begin{aligned} \rho_0 A = & -\frac{3\lambda_s + 2\mu_s}{4} J_1 + \frac{\lambda_s + 2\mu_s}{8} J_1^2 - \frac{\mu_s}{2} J_2 + l_s \left\{ -a_1 J_1 - a_2 J_1^2 - a_3 J_2 \right. \\ & \left. + \frac{1}{2} \beta^2 (6J_1 - J_1^2) + \exp[a_1(J_1 - 3) + a_2(J_1^2 - 9) + a_3(J_2 - 3)] \right\}, \end{aligned} \quad (13.35)$$

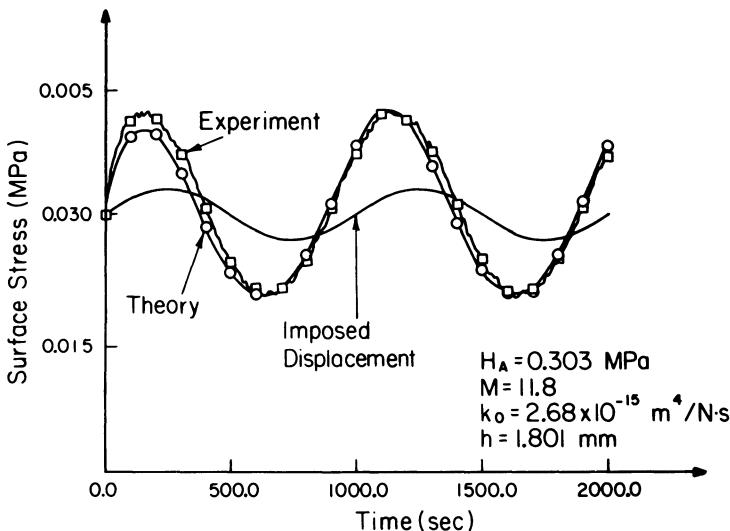


FIGURE 13.5. Comparison of our theoretical prediction of surface stress (○) and experimental data (□) for a bovine articular cartilage specimen subjected to a low frequency ( $10^{-3}$  Hz) sinusoidal surface displacement about a 10% d.c. offset strain. The phase lag between imposed displacement and stress is  $\sim 30^\circ$ . The intrinsic permeability coefficients  $k_0$  and  $M$  were obtained by a nonlinear regression analysis.

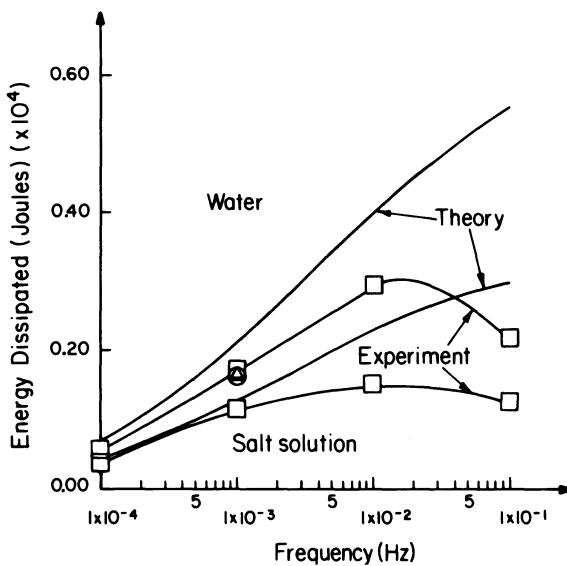


FIGURE 13.6. Variation with frequency of energy dissipation associated with sinusoidal excitation of the tissues equilibrated in de-ionized water and 0.15 mol NaCl.  $\Delta$  denotes value of energy dissipation corresponding to the sawtooth case shown in Figure 13.4.

where  $\beta = a_1 + 6a_2 + 2a_3$ , and

$$\rho_0 A = \frac{1}{8} J_3^{-1} \{(3\lambda_s + 2\mu_s)J_1 + (\lambda_s + 2\mu_s)J_1^2 - (9\lambda_s + 10\mu_s)J_2\}. \quad (13.36)$$

In addition, for simplicity, we will assume  $A = A^f = A^s$ . Equation (13.35) is similar to that suggested by Fung (1981) to describe the tensile behavior of soft biological materials. We note that for infinitesimal strains, both forms of the Helmholtz free-energy function yield the exact form of infinitesimal-stress-strain relationships as in Equations (13.20) or (13.21).

### 5.1. Finite-Deformation Equilibrium Stress–Strain Relations

Nonlinear equilibrium compression stress–strain behaviors have been reported (Sokoloff, 1966; Eisenfeld et al, 1978, Mow et al, 1980) and observed in many of our laboratory results. In general, with increasing compressive strain under uniaxial confined compression conditions, the solid matrix of articular cartilage, and possibly other hydrated tissues as well, stiffens, i.e., becomes nonlinear, for compressive strains beyond 20%. Under uniaxial confined-compression conditions, the kinematics of deformation are given by

$$\lambda_1 = \lambda_2 = 1, \quad 0 < \lambda_3 \leq 1, \quad (13.37a)$$

$$J_1 = \lambda_3^2 + 2, \quad J_2 = 2\lambda_3^2 + 1, \quad J_3 = \lambda_3^2, \quad (13.37b)$$

where now  $x_3$  is the direction of compression and  $\lambda_i$  are the stretches in the  $x_i$  direction. At equilibrium,  $p = 0$ . Thus, the equilibrium stress-strain relations corresponding to Equations (13.35) and (13.36), respectively, are

$$\begin{aligned} \sigma_{zz}^s = & 2\lambda_3[1 + d_1(\lambda_3 - 1)] \left\{ \frac{\lambda_s + 2\mu_s}{4}(\lambda_3^2 - 1) \right. \\ & + l_s[-2a_2(\lambda_3^2 - 1) - \beta - \beta^2(\lambda_3^2 - 1)] \\ & \left. + l_s[2a_2(\lambda_3^2 - 1) + \beta] \cdot \exp[\beta(\lambda_3^2 - 1) + a_2(\lambda_3^2 - 1)^2] \right\} \end{aligned} \quad (13.38)$$

and

$$\sigma_{zz}^s = (\lambda_s + 2\mu_s) \frac{[1 + d_1(\lambda_3 - 1)]}{4\lambda_3} \left[ \lambda_3^2 - \frac{1}{\lambda_3^2} \right], \quad (13.39)$$

where  $d_1 = d_0(\alpha_0 + 1)/(\alpha_0 + d_0)$ . These are obtained by simple differentiation by inserting Equations (13.35) and (13.36) into Equation (13.12a). Equations (13.38) and (13.39) may be used to describe our finite-deformation data. Figures 13.7a and b were obtained from nonlinear regression analysis using Equations (13.38) and (13.39) to fit our experimental data. In both cases, we see that the curve fits are very good within the range of stresses and strains used in our experiments. Shown in these figures are the values of the coefficients that provide the best least-squares fit of the data. The coefficients for the finite-deformation theory are listed. Figures 13.8a and b show the two theoretical

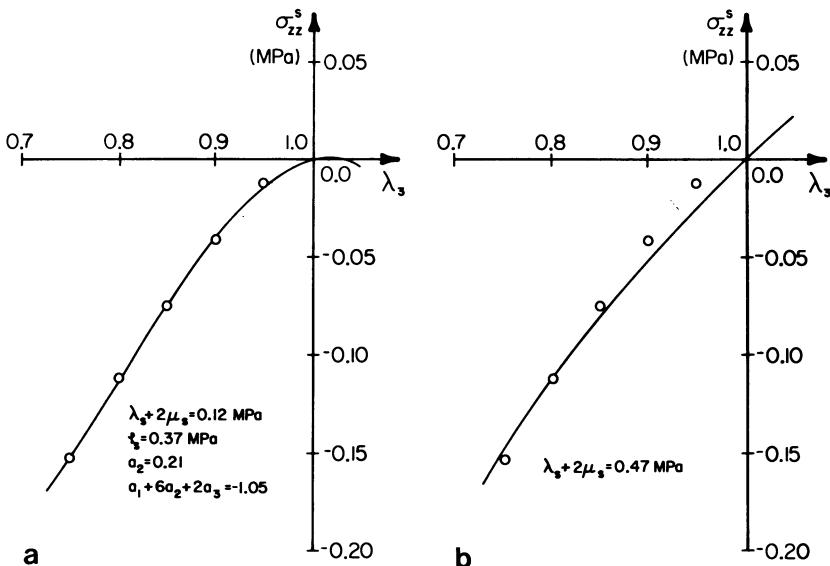


FIGURE 13.7. Data are mean values of five sets of data obtained from confined compression of bovine patellofemoral articular cartilage at equilibrium: (a) Nonlinear regression of mean value data using Equation (13.38). This  $\sigma_{zz}^s(\lambda_3)$  is particularly accurate for the compressive strains used. (b) Nonlinear regression of mean value data using Equation (13.39). This  $\sigma_{zz}^s(\lambda_3)$  deviates from data for  $\lambda_3 \approx 1$  but is physically more reasonable for  $\lambda_3 \rightarrow 0$  or  $\lambda_3 > 1$ .

stress-strain behaviors for  $0 < \lambda_3 < \infty$  as predicted by Equations (13.38) and (13.39). For stretches  $\lambda_3 \rightarrow 0$  and  $\lambda_3 \gg 1$ , Equation (13.38) fails to predict a physically reasonable monotonic behavior. This violates the convexity condition for finite-deformation theory (Truesdell and Toupin, 1963). The stress-strain relationship given by Equation (13.39) appears to be able to predict physically reasonable results for the entire range of stretch:  $0 < \lambda_3 < \infty$ . However, both Helmholtz free-energy forms, Equations (13.35) and (13.36), yield reasonable stress-strain results if the strains are not too large. Equations (13.36) and (13.39) appear to be a better choice if very large strains are anticipated.

## 5.2. Finite-Deformation Confined-Compression Creep Behavior

In order to consider the time-dependent compressive creep response of soft hydrated connective tissue under finite-deformation conditions, we need to revert to Equations (13.12a), (13.12b), and (13.12c), where  $p \neq 0$ . In this section, we will use only Equation (13.36) since it predicts a reasonable equilibrium confined-compression response over the entire range of  $\lambda_3$ . By inserting Equation (13.36) into Equation (13.12), we find

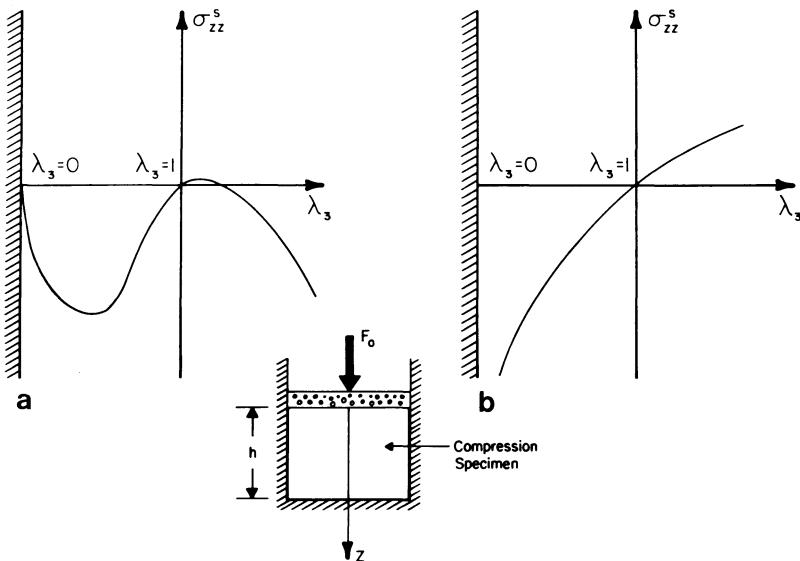


FIGURE 13.8. (a) Behavior of  $\sigma_{zz}^s$  defined by Equation (13.38) for the set of numerical values obtained from Figure 13.7a. The failure to predict nonmonotonicity for  $\lambda_3 \rightarrow 0$  and  $\lambda_3 > 1$  is a drawback. (b) The stress  $\sigma_{zz}^s$  defined by Equation (13.39) predicts physically reasonable responses for the complete range of deformation.

$$\begin{aligned}\boldsymbol{\sigma}^s = -\alpha p \mathbf{I} + \frac{1}{4} \frac{\rho}{\rho_0} J_3^{-1} \{ & -[(3\lambda_s + 2\mu_s)J_1 + (\lambda_s + 2\mu_s)J_1^2] \mathbf{I} \\ & + [3\lambda_s + 2\mu_s + 2(\lambda_s + 2\mu_s)J_1] \mathbf{B} + (9\lambda_s + 10\mu_s)J_3 \mathbf{B}^{-1} \},\end{aligned}\quad (13.40a)$$

$$\boldsymbol{\sigma}^f = -p \mathbf{I}, \quad (13.40b)$$

$$-\boldsymbol{\pi}^s = \boldsymbol{\pi}^f = -\alpha p \operatorname{grad} \ln \rho^s + \mathbf{K}(\mathbf{v}^s - \mathbf{v}^f), \quad (13.40c)$$

$$\frac{\rho}{\rho_0} = \frac{1}{\sqrt{\det \mathbf{B}}} \{ 1 + d_1(\sqrt{\det \mathbf{B}} - 1) \}, \quad (13.40d)$$

where  $\alpha$  is given by Equation (13.7).

The confined-compression problem considered here corresponds to the problem we solved earlier, defined by Equations (13.26) to (13.28), except that now we are considering the case of finite deformation with constant permeability. For this case, we introduce a set of material coordinates  $(X, Y, Z)$  so that the motion  $(x, y, z)$  may be defined by

$$x = X, \quad y = Y, \quad z = Z + u(Z, t), \quad (13.41)$$

where  $u(Z, t)$  is the uniaxial displacement of interest. Integrating the continuity equation, Equation (13.6), under the uniaxial confined-compression condition,

and using boundary conditions Equation (13.28c) and  $v^f(h, t) = 0$ , we obtain

$$v_Z^f = -\alpha v_Z^s. \quad (13.42)$$

Now, either by substitution of the Helmholtz free-energy equation, Equation (13.36), and continuity equation, Equation (13.42), into the governing equation, Equation (13.18), or by substitution of the stress-strain relations, Equation (13.40), and continuity equation, Equation (13.42), into the equations of motion and eliminating  $p$ , we obtain the Eulerian or spatial equations of motion:

$$\frac{\lambda_s + 2\mu_s}{4} \frac{\partial}{\partial Z} \left\{ [1 + d_1(\lambda_3 - 1)] \left( \lambda_3 - \frac{1}{\lambda_3^3} \right) \right\} - (1 + \alpha)^2 K v_Z^s = 0. \quad (13.43)$$

It is more convenient, when dealing with finite deformations, to formulate the problems in the Lagrangian or material coordinates used. Expressing  $\lambda_3$  and  $v_Z^s$  in terms of displacement  $u(Z, t)$  we have

$$\lambda_3 = 1 + \frac{\partial u}{\partial Z}, \quad (13.44a)$$

$$v_Z^s = \frac{Du}{Dt} = \frac{\partial u}{\partial t} \Big|_{\mathbf{x}} \quad (13.44b)$$

$$\frac{\partial}{\partial Z} (\ ) = \left( 1 + \frac{\partial u}{\partial Z} \right)^{-1} \frac{\partial}{\partial Z} (\ ). \quad (13.44c)$$

Substituting Equations (13.44a), (13.44b), and, (13.44c) into Equation (13.43) we obtain

$$\frac{\lambda_s + 2\mu_s}{4} \left( 1 + \frac{\partial u}{\partial Z} \right)^{-1} \frac{\partial}{\partial Z} \left[ \left( 1 + d_1 \frac{\partial u}{\partial Z} \right) \frac{\left( 1 + \frac{\partial u}{\partial Z} \right)^4 - 1}{\left( 1 + \frac{\partial u}{\partial Z} \right)^3} \right] = \frac{1}{k} \frac{\partial u}{\partial t}. \quad (13.45)$$

Equation (13.45) is now to be solved subject to the following boundary and initial conditions:

$$\sigma_{ZZ}^s(0, t) = \frac{\lambda_s + 2\mu_s}{4} \left( 1 + d_1 \frac{\partial u}{\partial Z} \right) \frac{\left( 1 + \frac{\partial u}{\partial Z} \right)^4 - 1}{\left( 1 + \frac{\partial u}{\partial Z} \right)^3} = -F_0, \quad t > 0, \quad (13.46a)$$

$$p = 0, \quad (13.46b)$$

$$u(h, t) = 0, \quad (13.46c)$$

$$u(h, 0) = 0. \quad (13.46d)$$

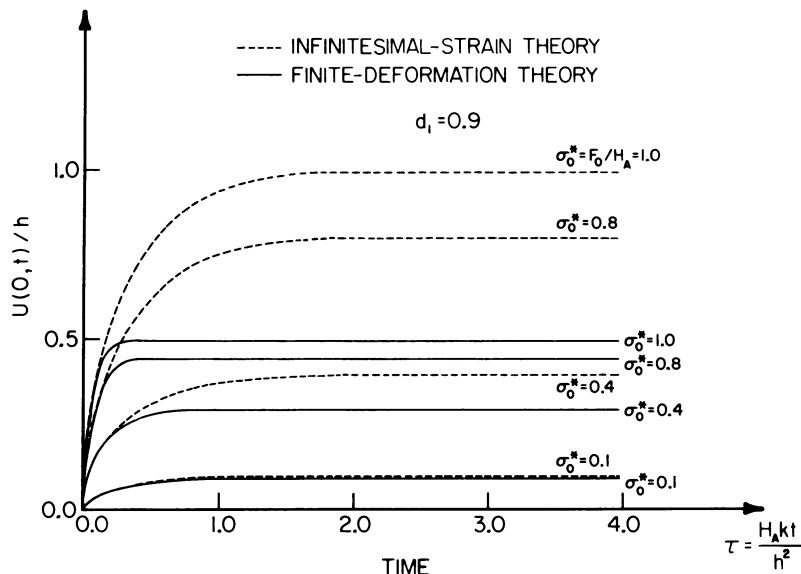


FIGURE 13.9. Predictions of confined-compression creep of a biphasic material using infinitesimal-strain theory and finite-deformation theory. For  $\sigma_0^* = 1.0$ , the linear infinitesimal strain theory predicts 100% compression—an obvious impossibility. For  $\sigma_0^* = 0.1$ , the infinitesimal-strain theory yields a good approximation to the finite-deformation theory. For  $\sigma_0^* = 0.4$ , the infinitesimal-strain theory is clearly inaccurate. The parameter  $d_1 = d_0(\alpha_0 + 1)/(\alpha_0 + d_0) = \rho_f/\rho_0$ .

It can be readily seen that Equations (13.45) and (13.46a) are the finite-deformation analogs to Equations (13.26) and (13.28a).

This set of differential systems has been solved numerically using a finite-element scheme, and the results are shown in Figure 13.9, where the dimensionless creep response  $[u(0,t)/h]$  is plotted against the dimensionless time  $\tau = H_A kt / h^2$  with  $\sigma_0^* = F_0 / H_A$  used as the parameter of the problem ranging from 0.1 to 1.0. Shown also are the predicted creep responses from the infinitesimal-strain theory, with  $\alpha_0$  and  $k$  being constants. The finite-deformation theory predicts much lower values for displacements at high loads. For the case of  $\sigma_0^* = 1.0$ , the linear infinitesimal theory predicts an equilibrium compressive strain of 100%, which is an obvious impossibility, while the finite-deformation theory predicts a strain less than 50%. Finally, the finite-deformation theory predicts a higher rate of creep at small times at high loads, reaching equilibrium much faster than creep at low loads. This might provide a very easy way to experimentally verify our finite-deformation theory. Earlier creep and stress-relaxation studies have always had slight deviations between theory and experiment. With the present theory, especially when both classes of nonlinearities are included, these discrepancies might be cleared up.

## 6. Conclusions

We have developed a set of constitutive equations to describe a class of biological materials. These materials are biphasic with the frictional drag of fluid flow through the porous, permeable, solid matrix dominating the viscoelastic creep and stress-relaxation behavior of the tissue in compression. The linear biphasic theory has been used successfully, as a first-order approximation, to describe the compressive creep and stress-relaxation behavior of articular cartilage, meniscus, and nasal septum. A number of experiments have also been performed in which the nonlinear permeability effects dominate the creep and stress-relaxation behaviors of these tissues. Now, with our nonlinear finite-deformation theory at hand, this theory may be applied to any set of material-testing conditions in the finite-strain range, and the properties determined this way might provide for a better correlation with biochemical composition and tissue microstructure than do the current results. Finally, the finite-deformation theory can now be used in situations simulating the actual highly loaded diarthrodial joint articulation conditions and to solve, for example, the joint lubrication problem.

## Acknowledgments

This research was supported by NIADDK grants AM19094 and AM26440, NSF grant MEA 8211968, and the Clark and Crossan Endowment at Rensselaer Polytechnic Institute. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily represent the views of the National Science Foundation.

## Appendix: Continuity Equation for a Binary Mixture of an Incompressible Solid and an Incompressible Fluid

Consider a control volume  $V_c$  bounded by the surface  $S_c$  with a unit outward normal vector  $\mathbf{n}$ . The rate of mass inflow of the solid phase into  $V_c$  is

$$\text{mass influx} = - \iint_{S_c} \rho^s \mathbf{v}^s \cdot \mathbf{n} dS, \quad (13.A.1)$$

where  $\rho^s$  is the “apparent density” defined by the solid mass per total volume and  $\mathbf{v}^s$  is the velocity of the solid phase. In the absence of sources and sinks, the rate of change of mass of the solid phase within  $V_c$  is

$$\text{rate of change of mass} = \frac{\partial}{\partial t} \iiint_{V_c} \rho^s dV. \quad (13.A.2)$$

Equating the right-hand sides of Equations (13.A.1) and (13.A.2) and using the divergence theorem, we obtain

$$\iiint_{V_c} \left[ \operatorname{div}(\rho^s \mathbf{v}^s) + \frac{\partial \rho^s}{\partial t} \right] dV = 0$$

or

$$\frac{\partial \rho^s}{\partial t} + \operatorname{div}(\rho^s \mathbf{v}^s) = 0 \quad (13.A.3)$$

for the continuity equation of the solid phase. Similarly, the continuity equation for the fluid phase is given by

$$\frac{\partial \rho^f}{\partial t} + \operatorname{div}(\rho^f \mathbf{v}^f) = 0, \quad (13.A.4)$$

where  $\rho^f$  and  $\mathbf{v}^f$  are the apparent density and velocity of the fluid phase.

The coupling of the two phases in the mixture is provided by the axiom that, in the mixture, the total volume,  $dV(t)$ , of an element of the mixture is given by the sum of the volume occupied by each phase of the mixture, i.e.,  $dV^f(t)$  and  $dV^s(t)$

$$dV(t) = dV^f(t) + dV^s(t). \quad (13.A.5)$$

By defining the apparent density and true density of the fluid phase and the solid phase to be

$$\rho^f(t) = \frac{dm^f}{dV(t)} \text{ (apparent)} \quad \rho^s(t) = \frac{dm^s}{dV(t)} \text{ (apparent)}, \quad (13.A.6a)$$

$$\rho_T^f(t) = \frac{dm^f}{dV^f(t)} \text{ (true)} \quad \rho_T^s(t) = \frac{dm^s}{dV^s(t)} \text{ (true)}, \quad (13.A.6b)$$

where  $dm^f$  and  $dm^s$  are the masses of the fluid and solid phases in  $dV^f(t)$  and  $dV^s(t)$ , respectively, Equation (13.A.5) may be written as

$$\rho_T^s(t)\rho_T^f(t) = \rho^s(t)\rho_T^f(t) + \rho^f(t)\rho_T^s(t). \quad (13.A.7)$$

For intrinsically incompressible phases defined by

$$\rho_T^s(t) \equiv \rho_T^s(0) \quad \text{and} \quad \rho_T^f(t) \equiv \rho_T^f(0), \quad (13.A.8)$$

Equation (13.A.7) becomes

$$\rho_T^s(0)\rho_T^f(0) = \rho^s(t)\rho_T^f(0) + \rho^f(t)\rho_T^s(0). \quad (13.A.9)$$

Differentiating Equation (13.A.9) with respect to time yields:

$$\frac{\partial \rho^f(t)}{\partial t} = -\frac{\rho_T^f(0)}{\rho_T^s(0)} \frac{\partial \rho^s(t)}{\partial t}. \quad (13.A.10a)$$

Further, if the initial true densities are homogeneous, taking the gradient of Equation (13.A.9) yields

$$\operatorname{grad} \rho^f = -\frac{\rho_T^f(0)}{\rho_T^s(0)} \operatorname{grad} \rho^s. \quad (13.A.10b)$$

By combining Equations (13.A.3), (13.A.4), and (13.A.10a), we obtain a simple relationship for the continuity equation of the fluid–solid mixture:

$$\operatorname{div}(\rho^f \mathbf{v}^f) + \frac{\rho_T^f(0)}{\rho_T^s(0)} \operatorname{div}(\rho^s \mathbf{v}^s) = 0. \quad (13.A.11)$$

Using the well-known vector identity

$$\operatorname{div}(\gamma \mathbf{a}) = \operatorname{grad} \gamma \cdot \mathbf{a} + \gamma \operatorname{div} \mathbf{a}, \quad (13.A.12)$$

where  $\gamma$  is an arbitrary scalar and  $\mathbf{a}$  is an arbitrary vector, Equation (13.A.10b) and Equation (13.A.11) may be combined to yield.

$$\operatorname{div} \mathbf{v}^f + \alpha \operatorname{div} \mathbf{v}^s + \alpha(\mathbf{v}^s - \mathbf{v}^f) \cdot \operatorname{grad} \ln \rho^s = 0, \quad (13.A.13)$$

where

$$\alpha \equiv \frac{\rho_T^f(0)}{\rho_T^s(0)} \times \frac{\rho^s(t)}{\rho^f(t)} = \frac{dV^s(t)}{dV^f(t)} \equiv \text{solid content} \quad (13.A.14)$$

where Equation (13.A.8) has been used. Equations (13.A.13) and (13.A.14) define the continuity equations Equation (13.4) and (13.5) of the text for the incompressible fluid–solid mixture. The physical meaning of each term of Equation (13.A.13) becomes clearer if we insert the relations

$$\rho^f(t) = \frac{\rho_T^f(0)}{1 + \alpha} \quad \text{and} \quad \rho^s(t) = \frac{\alpha \rho_T^s(0)}{1 + \alpha} \quad (13.A.15)$$

into Equation (13.A.11). This results in an alternate and sometimes more convenient form for the continuity equation of the mixture given by

$$\operatorname{div} \left[ \frac{1}{1 + \alpha} (\mathbf{v}^f + \alpha \mathbf{v}^s) \right] = 0, \quad (13.A.16)$$

where each term of the left side of Equation (13.A.16) represents the volume flux per unit mixture volume (Equation (13.6) of the text).

Finally, for finite deformation, the volumetric ratio  $\alpha = dV^s(t)/dV^f(t)$ , i.e., the solid content, can change appreciably. To relate this change to the deformation, let us first consider the relationship

$$dV^s(t) = dV^s(0) \quad (13.A.17)$$

for an incompressible solid. By virtue of the definition of  $\alpha$  from Equation (13.A.14), this is identical to the relationship

$$\frac{\alpha}{1 + \alpha} dV = \frac{\alpha_0}{1 + \alpha_0} dV_0. \quad (13.A.18)$$

Since the volume change is related to the Jacobian of transformation describing the motion of the solid matrix, i.e.,  $dV = (\sqrt{\det \mathbf{B}}) dV_0$ , Equation (13.A.18) may be expressed as

$$\frac{\alpha}{1 + \alpha} dV = \frac{\alpha_0}{1 + \alpha_0} \frac{dV}{\sqrt{\det \mathbf{B}}}, \quad (13.A.19)$$

where  $\mathbf{B}$  is the left Cauchy–Green deformation tensor. Solving Equation (13.A.19) for  $\alpha$  yields

$$\alpha = \frac{\alpha_0}{(1 + \alpha_0)\sqrt{\det \mathbf{B}} - \alpha_0} \quad (13.A.20)$$

(Equation (13.7) in the text). Thus, for finite deformation of the solid matrix, the relative amount of fluid within the solid matrix will change in accordance to Equation (13.A.20).

## References

- Armstrong, C.G., and Mow, V.C. 1980. Pages 223–232 in *Scientific foundations: orthopaedics and traumatology*, R. Owen, J. Goodfellow, and P. Bullough, eds. William Heinemann Medical Books, Ltd., London.
- \_\_\_\_\_. 1982. *J. Bone Joint Surg.* 64A:88–94.
- Armstrong, C.G., Bahrani, A.S., and Gardner, D.L. 1979. *J. Bone Joint Surg.* 61A: 744–755.
- Bär, E. 1926. *Arch. Entwicklungsmech. Org.* 108:739–760.
- Bowen, R.M. 1976. Pages 1–127 in *Continuum physics*. Vol 3, A.E. Eringen, ed. Academic Press, New York.
- \_\_\_\_\_. 1980. *Int. J. Eng. Sci.* 18:1129–1148.
- Charnley, J. 1959. Pages 12–22 in *Sympos. on biomech.* Institution of Mechanical Engineers, London.
- Coletti, J.M., Akeson, W.H., and Woo, S.L-Y. 1972. *J. Bone Joint Surg.* 54A:147–160.
- Craine, R.E., Green, A.E., and Naghdi, P.M. 1970. *Q. J. Mech. Appl. Math.* 23: 171–184.
- Edwards, J. 1967. *Proc. Inst. Mech. Eng. (London)* 181:16–24.
- Eisenfeld, J., Mow, V.C., and Lipshitz, H. 1978. *J. Math. Biosci.* 39:97–111.
- Elmore, S.M., Sokoloff, L., Norris, G., and Carmeci, P. 1963. *J. Appl. Physiol.* 18:393–396.
- Favenesi, J.A., Shaffer, J.C., and Mow, V.C. 1983. *Trans. Orth. Res. Soc.* 8:57.
- Fung, Y.C.B. 1968. *Appl. Mech. Rev.* 21:1–20.
- Fung, Y.C.B. 1981. *Biomechanics: mechanical properties of living tissues*. Springer-Verlag, New York.
- Ghadially, F.N. 1978. Pages 105–176 in *The joints and synovial fluid*. Vol. 1, L. Sokoloff, ed. Academic Press, New York.
- Gildemeister, M. 1914. *Z. Biol.* 63:175–190.
- Göcke, E. 1927. *Verh. Dtsch. Orth. Ges.* 22:130–147.
- Hayes, W.C., and Mockros, L.F. 1971. *J. Appl. Physiol.* 31:562–568.

- Hayes, W.C., Keer, L.M., Herrmann, G., and Mockros, L.F. 1972. *J. Biomech.* 5: 541–551.
- Hirsch, C. 1944. *Acta Chir. Scand.* 90(83): 1–106.
- Holmes, M.H. 1984. *SIAM J. Appl. Math.* 44: 545–556.
- Holmes, M.H., Lai, W.M., and Mow, V.C. 1984a. In *Tissue nutrition and viability*, A. Hargens, ed. Springer-Verlag, New York.
- Holmes, M.H., Hou, J.S., Lai, W.M., and Mow, V.C. 1984b. *Proc. Europ. Soc. Biomech.*: 175.
- Holmes, M.H., Hou, J.S., and Mow, V.C. 1984c. In *Bioeng. Sympos., Trans.*, R.L. Spilker, ed. ASME, New Orleans.
- Holmes, M.H., Lai, W.M., and Mow, V.C. 1985. *J. Biomech. Eng.*, in press.
- Kempson, G.E., Freeman, M.A.R., and Swanson, S.A.V. 1971. *J. Biomech.* 4: 239–250.
- Lai, W.M., and Mow, V.C. 1980. *Biorheology* 17: 111–123.
- Lai, W.M., Mow, V.C., and Roth, V. 1981. *J. Biomech. Eng.* 103: 221–231.
- Linn, F.C. 1968. *J. Biomech.* 1: 193–205.
- Linn, F.C., and Sokoloff, L. 1965. *Arthritis Rheum.* 8: 481–494.
- MacConaill, M.A. 1932. *J. Anatomy* 66: 210–227.
- \_\_\_\_\_. 1951. *J. Bone Joint Surg.* 33B: 251–257.
- McCutchen, C.W. 1962. *Wear* 5: 1–17.
- McCutchen, C.W., and Lewis, P.R. 1959. *Nature* 184: 1284–1285.
- Mansour, J.M., and Mow, V.C. 1976. *J. Bone Joint Surg.* 58A: 509–516.
- Maroudas, A. 1968. *Biophys. J.* 8: 575–595.
- \_\_\_\_\_. 1975a. *Biorheology* 12: 233–248.
- \_\_\_\_\_. 1975b. *Ann. Rheum. Dis.* 34, Suppl.: 77–81.
- Mow, V.C. 1969. *J. Lubr. Technol. Trans. ASME* 91: 320–329.
- Mow, V.C., and Lai, W.M. 1979. Pages 247–288 in *Ann. Rev. Fluid Mech.* Vol. 11, M. Van Dyke, ed. Annual Reviews Inc., Palo Alto, California.
- \_\_\_\_\_. 1980. *SIAM Rev.* 22: 275–317.
- Mow, V.C., and Schoonbeck, J.M. 1982. *Trans. Orth. Res. Soc.* 7: 209.
- Mow, V.C., Kuei S.C., Lai, W.M., and Armstrong, C.G. 1980. *J. Biomech. Eng.* 102: 73–84.
- Mow, V.C., Holmes, M.H., and Lai, W.M. 1984a. *J. Biomech.* 17: 377–394.
- Mow, V.C., Mak, A.F., Lai, W.M., Rosenberg L.C., and Tang, L-H. 1984b. *J. Biomech.* 17: 325–338.
- Muir, I.H.M. 1980. Pages 27–94 in *The joints and synovial fluid*. Vol. 2, L. Sokoloff, ed. Academic Press, New York.
- Mulholland, R., Millington, P.F., and Manners, J. 1975. *Ann. Rheum. Dis.* 34, Suppl.: 104–107.
- Myers, E.R., and Mow, V.C. 1983. Pages 313–342 in *Cartilage*. Vol. 1, B.K. Hall, ed. Academic Press, New York.
- Myers, E.R., Lai, W.M., and Mow, V.C. 1984. *J. Biomech. Eng.* 106: 151–158.
- Parsons, J.R., and Black, J. 1977. *J. Biomech.* 10: 21–30.
- \_\_\_\_\_. 1979. *J. Biomech.* 12: 765–773.
- Sokoloff, L. 1963. *Science* 141: 1055–1056.
- \_\_\_\_\_. 1966. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* 25: 1089–1095.
- Torzilli, P.A., and Mow, V.C. 1976a. *J. Biomech.* 9: 541–552.
- \_\_\_\_\_. 1976b. *J. Biomech.* 9: 587–606.
- Torzilli, P.A., Rose, D.E., and Dethmers, D.A. 1982. *Biorheology* 19: 519–537.
- Truesdell, C., and Toupin, R. 1963. *Arch. Rational Mech. Anal.* 12: 1–33.

- Venn, M., and Maroudas, A. 1977. *Ann. Rheum. Dis.* 36:121–129.
- Walker, P.S., Dowson, D., Longfield, M.D., and Wright, V. 1968. *Ann. Rheum. Dis.* 27:512–520.
- Waters, N.E. 1965. *Br. J. Appl. Phys.* 16:1387–1392.
- Zarek, J.M., and Edwards, J. 1964. Pages 187–203 in *Biomechanics and related bioengineering topics*, R.M. Kenedi, ed. Pergamon Press, Cambridge, England, U.K.

# 14

## Biomechanics of Tendons and Ligaments

S.L-Y. Woo

### 1. Introduction

The common approaches used in describing the constitutive relationships for soft biological tissues have been: (1) to directly measure the stresses ( $\sigma$ ) and strains ( $\epsilon$ ) based on unidimensional force and deformation (in tension, bending, or shear) to form a simple stress-strain relationship, or (2) to use a continuum approach in which a pseudo-strain-energy density function is utilized. The following equations exemplify the latter approach:  $W = W(\epsilon_{ij})$  where  $i, j = 1, 2, 3$ ; thus,  $\sigma_{ij} = \partial W / \partial \epsilon_{ij}$  can be written to represent a generalized three-dimensional formulation of the state of stress (Fung, 1973). For tendons and ligaments, uniaxial tensile tests and the small-strain theory have been used to describe their mechanical and structural properties (Viidik, 1973, 1979; Tipton et al, 1975; Woo, 1982). Since the major function of these tissues is to support very large muscle and joint forces with minimal deformation, their parallel-fibered organization is designed uniquely to connect muscle to bone or bone to bone, respectively.

Typical stress-strain curves for tendons, ligaments, and skin are shown in Figure 14.1. The factors contributing to the large variation in mechanical properties between these soft tissues are the organization and/or orientation of the collagen fibers, as well as the percentages of various constituent materials. It is well known that skin has a random and loose network of Type I collagen fibers. Thus, during deformation, the largest percentage of strain acts to realign and to recruit the collagen fibers into the direction of the load. Tendons and ligaments also consist predominantly of the same Type I collagen fibers, but the fibers are oriented in parallel along the tissue length, and thus require little or no realignment during stretch.

Large differences exist in the contents of constituent materials, (i.e., the percentage of water, collagen, elastin, proteoglycans (PGs), and other non-collagenous proteins) between various soft tissues. It can be seen from Table 14.1 that water occupies the largest percentage of weight (and space) in all tissues (60–80%), while collagen is the major solid component. On a fat-free, dry weight basis, the percentage of collagen in ligaments and tendons ranges

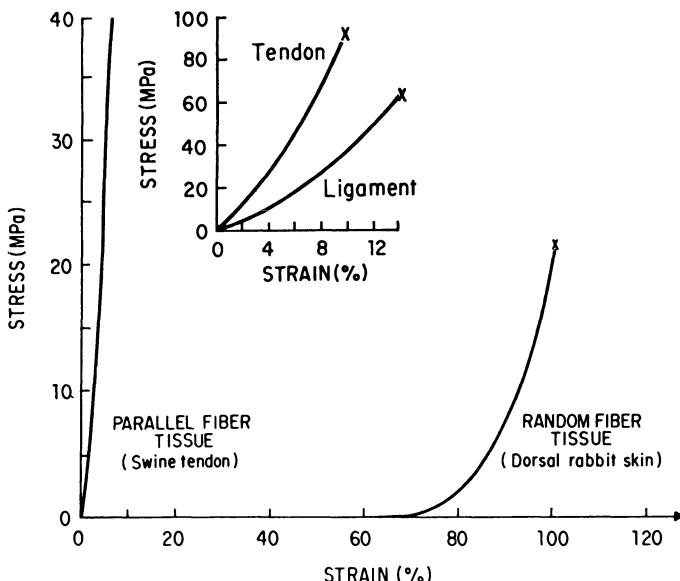


FIGURE 14.1. Typical stress-strain curves for tendons, ligaments, and skin.

from 75–80%. Elastin, another solid protein with stiffness significantly lower than that of collagen, also exists in tendons and ligaments in much smaller amounts. The last major solid components of soft tissues are PGs, which are noncollagenous macromolecules interwoven with collagen fibers and elastin. These hydrophylic and large macromolecules contribute only minimally to the tissues' elasticity, but, because of their enormous size, play a significant role in the time- and history-dependent viscoelastic properties in tendons and ligaments (Rosenberg, 1975).

Despite the apparent morphologic simplicity, the reported mechanical properties for tendons and ligaments in the literature have been widely different (Abrahams, 1967; Yamada, 1970; Kennedy et al, 1976; Noyes, 1977; Butler et al, 1978; Viidik, 1979; Dorlot et al, 1980; Woo, 1982). For example, the ultimate strain values for tendons can range from 9% to over 30%; those for the ligaments range from 12% to over 50%. These variations are partly due to age, species, and type of specimens studied. But, there are additional, perhaps more important variables that would contribute to the reported differences. These variables may include: (1) error in cross-sectional area measurements; (2) clamping of the test specimens, resulting in premature failures; (3) poor length-width (aspect) ratio for uniform stress during testing; (4) poor techniques in measuring tissue strain; and (5) improper test environments (i.e. temperature, pH, drying of specimens, etc). To illustrate, soft tissues usually slip within the clamps during tensile stretching; thus, the strain measurements using the clamp-to-clamp deformation (a common practice in reporting tissue strain)

**TABLE 14.1.** Biochemical constituents of soft tissues.

Tissue	(% of total wt) Water	(% of fat-free dry wt)		
		Collagen	Elastin	PGs
Skin	60–65	65–70	5–10	1.5–2
Aorta	60–70	25–30	40–50	2–2.5
Tendon	65–70	75–80	<3	1–1.5
Ligament				
cruciate	65–70	75–80	<5	2.5–3
collateral	65–70	75–80	<5	1–1.5
Articular cartilage	70–80	60–65	Trace	10–15

are in error. Also, ligaments are usually too short (poor aspect ratio), and thus, cannot (and should not) be tested in their isolated state. But using the bone–ligament–bone preparation, despite its obvious advantages, can result in other errors, such as nonuniform strains between and along the ligament substance, large variations in deformation between ligament substance and its insertion to bone, and complications in defining its original length. In an effort to minimize some of these experimental errors, improvements have been made for studying the properties of tendons and ligaments. Some of the current techniques are elaborated in this chapter.

## 2. Quasi-Static Mechanical Properties

To determine the stress–strain relationships of ligaments and tendons, accuracies in cross-sectional area measurements, appropriate test specimen preparation, and clamping techniques are important. Detailed methodologies for these aspects can be found in Woo (1982) and are thus not elaborated upon here. To measure the tensile behaviors of ligaments and tendons, an experimental apparatus as shown in Figure 14.2 has been employed. During a uniaxial tensile test, the tensile load is typically measured by a load cell, while the tensile strain is determined by a video dimensional analyzer (VDA) system. The VDA system permits accurate measurements of the tissue strain within the gauge length marks (Woo et al, 1976a). Such a system has many advantages over traditional methods, including: (1) the VDA system can measure tissue strains continuously and automatically without requiring external attachments such as transducers or clip gauges to the soft tissue specimen; (2) the tissue strains measured have no error contribution from the specimen slippage at the clamps, or the nonuniformity from the complexities, such as at the tendon–muscle or ligament–bone junctions; (3) the test information is videotape-recorded and can be used for repeated tissue strain analyses, and also for the determination of strain variation along a tendon or ligament

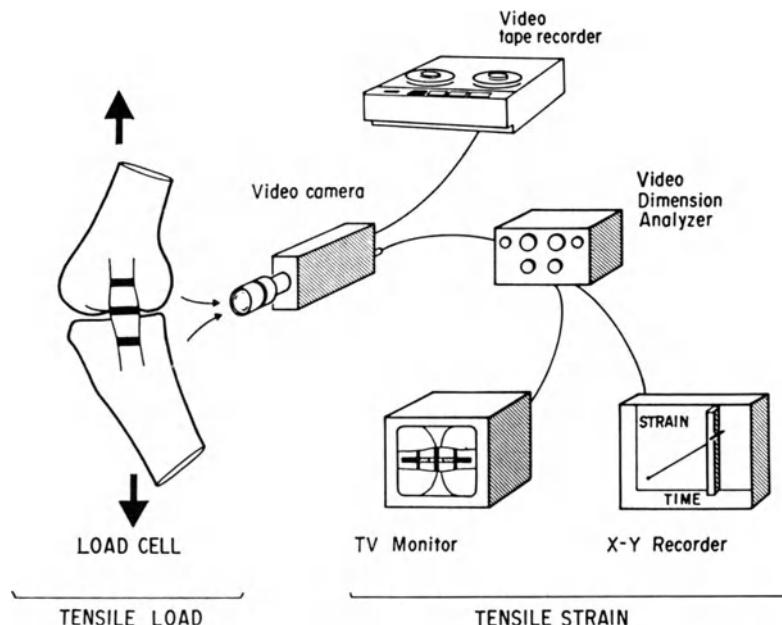


FIGURE 14.2. A schematic diagram detailing the apparatus used to acquire the tensile characteristics of soft tissues. A video dimensional analyzer system was used to determine the tensile strains along the tissue substance. By using the sample cross-sectional area (within the gauge length marks) stress-strain curves may be obtained.

substance (Woo et al, 1983); and (4) this system determines the strain with little or no human error and is much less laborious when compared to other techniques used. The VDA system described is a modification of those originally developed by Yin et al (1972). The frequency response of the system is linear up to 20 Hz (recently it has been improved to 120 Hz), the resolution is better than 0.2% of full scale, and the linearity is better than 1% of full scale.

The apparatus shown in Figure 14.2 has been used to study the mechanical properties of a wide range of soft tissues, including articular cartilage, skin, and various ligaments and tendons (Woo et al, 1979; Woo, 1982). The system's accuracy further permits the differentiation of mechanical properties between similar soft tissues (e.g. digital flexor tendons and their compensatory digital extensor tendons) as well as the variation of the properties along the ligament substance (see Figure 14.3). In addition, the changes of tensile properties of the medial collateral ligament (MCL)-bone complex simultaneously with the mechanical properties of the MCL substance of rabbits during stages of maturity have been evaluated (Woo et al, 1985). It can be seen from Figure 14.4 that there is a continuous change in the properties of MCL between the younger and more mature animals. The major changes, however, appear to occur during the epiphyseal closure and at the ligament insertions to bone.

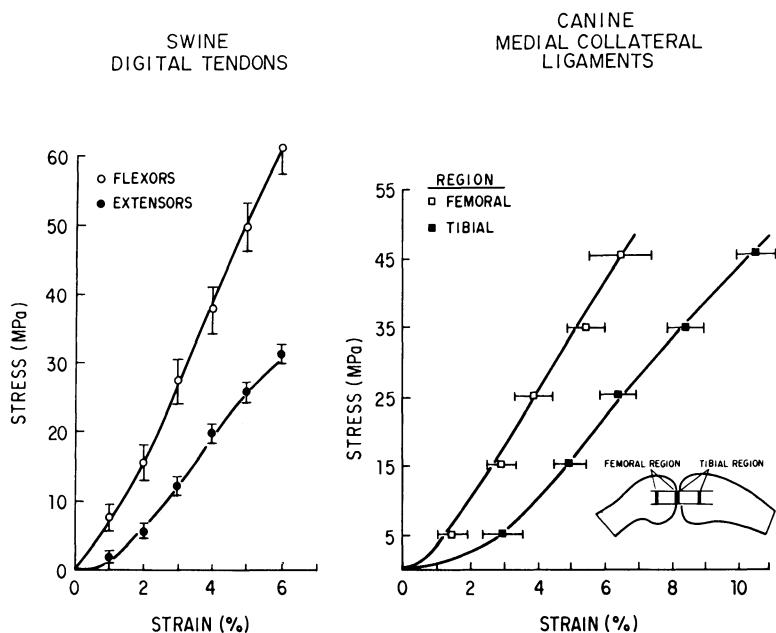


FIGURE 14.3. Mechanical properties of swine digital tendons and canine medial collateral ligaments. Diagram (left) illustrates the differences in  $\sigma$ - $\epsilon$  behavior between digital flexor and extensor tendons, and diagram (right) details the variation in properties along the length of the ligament substance.

### 3. Time- and History-Dependent Viscoelastic Properties

Several advanced theories of viscoelasticity are available to describe the time- and history-dependent viscoelastic properties of soft tissues. The theories include that of Lianis (1970); that of Bernstein, Kearsley, and Zapas (BKZ theory) (1963); and the quasi-linear viscoelastic (QLV) theory of Fung (1972). The QLV theory has been the most popular and has been successfully applied on rabbit mesenteries (Chen, 1973), on aorta and aortic tree (Tanaka and Fung, 1974), on parallel fiber tissues (Haut and Little, 1972; Jenkins and Little, 1974), on papillary muscle (Pinto and Patitucci, 1980), and on smooth muscle (Price et al, 1977). The QLV theory has been applied to articular cartilage and medial collateral ligaments (Woo et al, 1980a, 1981a; Simon et al, 1984).

The QLV theory is relatively simple to use. The theory assumes that the stress relaxation function can be written as:

$$\sigma[\epsilon(t); t] = G(t) * \sigma^e(\epsilon) \quad (14.1)$$

where  $\sigma^e(\epsilon)$  is the nonlinear “elastic response” (function of  $\epsilon$  only), and  $G(t) = \sigma(t)/\sigma(0)$  is the reduced relaxation function (function of  $t$  only). The stress at time  $t$ ,  $\sigma(t)$ , is thus the convolution integral of the  $G(t)$ , and  $\sigma^e(\epsilon)$ , or:

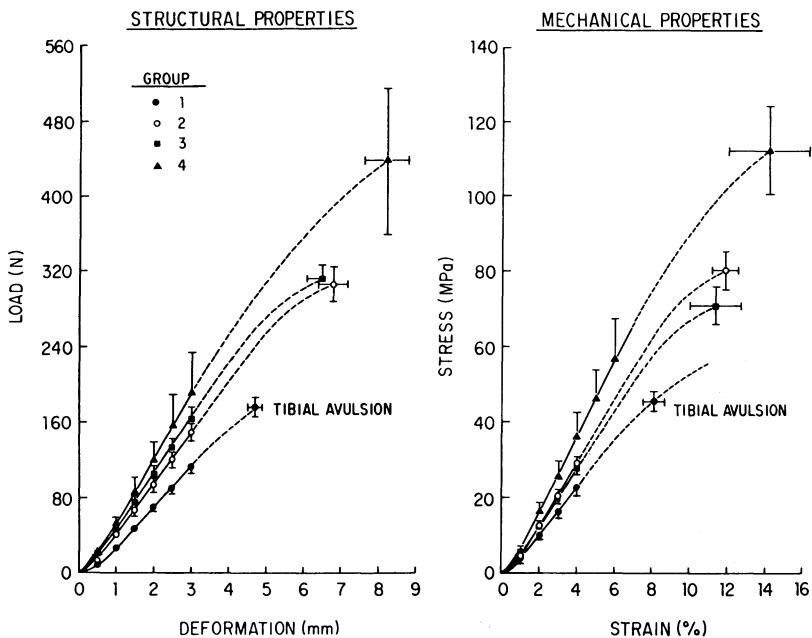


FIGURE 14.4. Curves showing the structural and mechanical properties of the rabbit medial collateral ligament–bone complex at different stages of maturity (Group 1: 4–5 months with open epiphysis; Group 2: 6–7 months with closed epiphysis; Group 3: 12–15 months; and Group 4: over 40 months).

$$\sigma(t) = \int_0^t G(t-\tau) \frac{\partial \sigma^\epsilon(\epsilon)}{\partial \epsilon} \frac{\partial \epsilon}{\partial \tau} d\tau. \quad (14.2)$$

Therefore, once  $G(t)$ ,  $\sigma^\epsilon(\epsilon)$  and the strain history  $\epsilon(t)$  are known, the time- and history-dependent stress,  $\sigma(t)$ , is completely described by Equation (14.2).

With exceptionally high strain rates (in the order of 1,000% per second) (Haut, 1983), soft tissues are insensitive to strain rates for several decades. In this case, Fung (1972) further proposed a generalized reduced relaxation function as follows:

$$G(t) = \left[ 1 + \int_0^\infty S(\tau) e^{-t/\tau} d\tau \right] / \left[ 1 + \int_0^\infty S(\tau) d\tau \right] \quad (14.3)$$

where  $S(\tau)$  is a continuous spectrum and has the following special form:

$$S(\tau) = \begin{cases} \frac{C}{\tau} & \text{for } \tau_1 \leq \tau \leq \tau_2 \\ 0 & \text{for } \tau < \tau_1, \tau > \tau_2. \end{cases} \quad (14.4)$$

Equation (14.3) can then be rewritten as:

$$G(t) = \frac{[1 + C\{E_1(t/\tau_2) - E_1(t/\tau_1)\}]}{[1 + C \ln(\tau_2/\tau_1)]} \quad (14.5)$$

where the exponential integral,  $E_1(y) = \int_0^\infty \frac{e^{-t}}{t} dt$ . Constants  $C$ ,  $\tau_1$  and  $\tau_2$  can be determined by solving three simultaneous equations obtained from  $dG/d(\ln t)$ ,  $G(\infty)$  and  $G(t)$ , where  $\tau_1 \ll t \ll \tau_2$ .

For the nonlinear elastic response,  $\sigma^e(\varepsilon)$ , an exponential expression, i.e.:

$$\sigma^e = A[\exp(Be) - 1] \quad (14.6)$$

is chosen. The constants  $A$  and  $B$  can be determined by performing a quasi-static tensile test at a constant strain rate,  $\alpha (= \Delta\varepsilon/\Delta t)$ . Using Equation (14.6), Equation (14.2) can be rewritten as

$$\sigma(t) = AB\alpha \int_0^t G(t-\tau)e^{\alpha B\tau} d\tau. \quad (14.7)$$

With  $\sigma(t)$ , and  $G(t)$  known, the constants  $A$  and  $B$  from Equation (14.7) can then be determined by using a nonlinear least-square curve-fitting procedure.

As an example, the viscoelastic properties of the canine MCL have been determined using the QLV method described. The MCL is rapidly stretched (to 2.5% strain at a rate of 10%/s) and then allowed to stress-relax up to 16 hours. From this stress relaxation test,  $dG/d(\ln t)$ ,  $G(\infty)$  (represented by  $t = 16$  hrs), and  $G(t)$  where  $t = 120$  s are known. From Equation (14.5), three simultaneous equations are obtained, and thus, the three unknown constants  $C$ ,  $\tau_1$  and  $\tau_2$  are determined. Their values are 0.099 s, 0.29 s, and  $1.99 \times 10^5$  s, respectively. Thus, the reduced relaxation function for MCL can be rewritten following Equation (14.5) to be

$$G(t) = 0.749 - 0.042 \ln t. \quad (14.8)$$

A separate constant strain rate ( $\alpha = 0.01\%/\text{s}$ ) experiment of canine MCL has been done. From the history of  $G(t)$ ,  $\sigma(t)$ , and  $\alpha$ , constants  $A$  and  $B$  in Equation (14.7) can be obtained by a nonlinear curve-fit procedure (described in detail in Woo et al, 1981a). Thus, the elastic response of the MCL as written in Equation (14.6) is reduced to:

$$\sigma^e = 0.193[\exp(161\varepsilon) - 1]. \quad (14.9)$$

Since  $G(t)$  and  $\sigma^e(\varepsilon)$  are obtained from two independent sets of experiments, the results obtained in Equations (14.8) and (14.9) must be confirmed by an additional set of experimental data. To do this, cyclic stretching of MCL between two strain levels, 1.5 and 2.5%, for 10 cycles at a strain rate of 0.1%/s were performed; the resulting peak and valley stresses are shown in Figure 14.5. Using  $G(t)$  from Equation (14.8) and  $\sigma^e(\varepsilon)$  from Equation (14.9) and the known strain rate  $\partial\varepsilon/\partial t = 0.1\%/\text{s}$ , the peak and valley stresses,  $\sigma(t)$ ,  $\sigma(t)$  at the respective times,  $t_s$  for peak and valley strains are calculated by using Equation (14.2). As can be seen from Figure 14.5, the calculated stresses using the

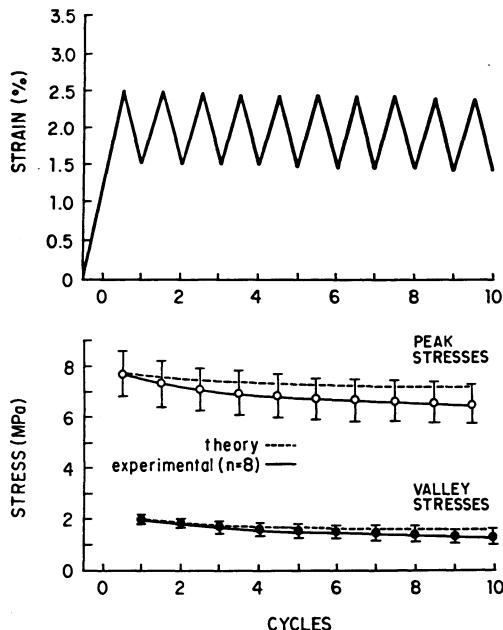


FIGURE 14.5. Curves showing the experimental and theoretically (QLV theory) predicted peak and valley stresses for a canine medial collateral ligament subjected to cyclic testing at low strain levels.

QLV theory match well with those obtained by the experiments. Thus it is concluded that the  $G(t)$  and the nonlinear  $\sigma^*(\epsilon)$  determined for the canine MCL substance are reasonable and can be appropriately used to describe its time- and history-dependent viscoelastic properties at the strain levels studied. Recently, the sensitivity of the reduced relaxation function has been further evaluated by Sauren and Rousseau (1983). These authors find that constant  $C$  is the most important parameter to describe the viscous effects, while constants  $\tau_1$  and  $\tau_2$  govern the "fast" and "slow" viscous phenomena, respectively.

#### 4. Homeostasis—Modification of Mechanical Properties in Vivo

In spite of the inert appearance of ligaments and tendons, there are significant homeostatic changes in *in-vivo* mechanical properties (Woo et al, 1976b). Conservative treatment of ligament injuries by cast or splint immobilization has been demonstrated to cause severe joint contractures, especially in elderly patients. In the laboratory, immobilization of animal knee joints has been shown to significantly compromise the properties of the ligament and the

bone-ligament-bone complex in dogs (Akeson et al, 1968), rabbits (Akeson et al, 1973; Woo et al, 1975), and primates (Noyes et al, 1977). Following nine weeks of immobilization, the structural properties (i.e., the linear slope, ultimate load, and energy-absorbing capabilities) of the rabbit's MCL-bone complex during tensile failure tests were reduced to approximately a third of that of the contralateral nonimmobilized control (see Figure 14.6). Also, the mechanical properties of the mid-MCL substance were significantly inferior.

It is also interesting to observe the recovery rates of MCL and MCL-bone complex following remobilization. After nine weeks of remobilization, the

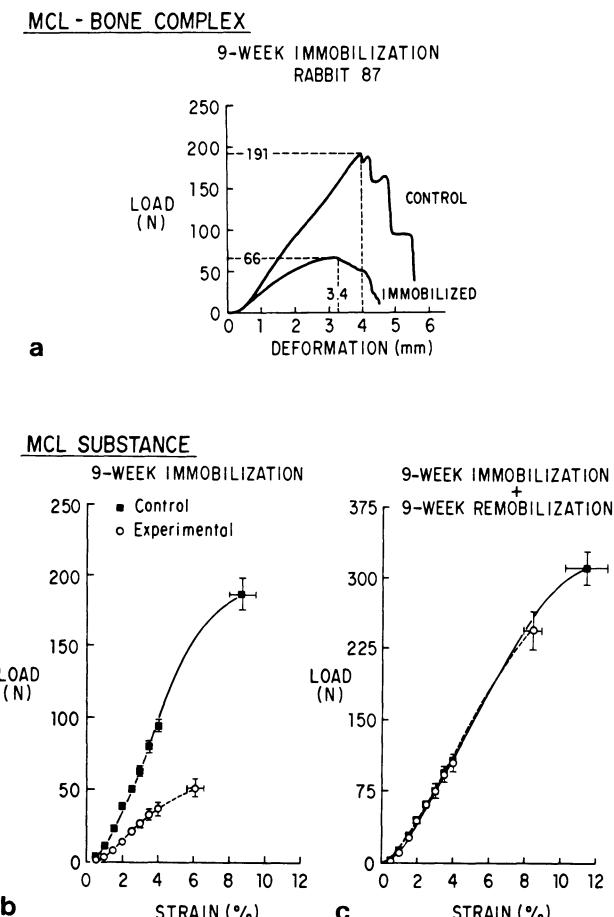


FIGURE 14.6. Diagram (a) shows the difference in structural properties of control (right) rabbit medial collateral ligament-bone complex and that for the experimental (left), which was immobilized for 9 weeks. Diagrams (b) and (c) show the mechanical properties of the MCL substances between the control and immobilized as well as between the control and immobilized plus remobilized.

structural properties of the experimental rabbit MCL–bone complex recover somewhat, but continue to be inferior compared to nonimmobilized controls. (The maximum load and energy absorbed at failure are 76% and 62%, respectively, of control values). These results are in agreement with Noyes et al (1977). These investigators showed that primates subjected to eight weeks of knee immobilization followed by 5 and 12 months of reconditioning recover less than 80% and 90%, respectively, of their original anterior cruciate ligament–bone strength. Of interest in our findings, however, is that the mechanical properties of the ligament (MCL) substance recover almost completely in the physiological functional range. The stress–strain curves of MCLs from immobilized experimental and normal active control knees are identical (see Figure 14.6).

These findings suggest that it is necessary to differentiate the mechanical properties of the ligament substance from that of the structural characteristics of the bone–ligament functional unit. While the time course for recovery at the bone–ligament junction is a slow one, the same may not be true for the function of ligament substance in the physiological range (up to 5% ligament strain). The aforementioned changes in biomechanical properties following immobilization and remobilization are concomitant with changes in morphological appearance of the ligaments, as well as the biochemical constituents of the ligaments (collagen mass, collagen turnover, and collagen cross-linking). The readers are referred to the recent review article on interdisciplinary studies of ligaments for detailed information (Akeson et al, 1984).

Excellent work on the structural characteristics of rat and canine bone–ligament complexes and their responses to a range of physical activities has been published by Tipton et al (1975). Woo et al (1981b) subjected miniature swine to exercise training programs at a schedule of 40 km/wk of running at moderate speeds (6–8 km/hr) for 3 and 12 months. The animals' heart rates during training were monitored to be 215–260 beats per minute, which correspond to 65–80% of their maximum heart rates. Matching sedentary animals were used as controls. At sacrifice, the digital extensor and flexor tendons of the two middle toes of the swine were excised and studied. The details of specimen preparations and test procedures are described in Woo et al (1980b, 1981b). It can be seen from Figure 14.7 that short-term (3 months) exercise has little or no effect on the mechanical properties of tendons, whereas the long-term (12 months) exercise has a positive effect on the extensor tendons but not on the flexor tendons. The tensile strength of the extensor tendons from the exercised animals was 22% higher than that of the controls. Tendon hypertrophy was also observed, along with the corresponding increase in tendon structural properties (Woo et al, 1980b).

In the case of flexor tendons, exercise training can alter the strength of tendon insertion to bone. The load to failure for the flexor–bone complex from the exercise group was superior to the control group—6% in the short-term exercised animals, and 19% in the long-term exercised animals ( $p < 0.005$ ). It is interesting to note that exercise has no effect on the flexor tendon substance.

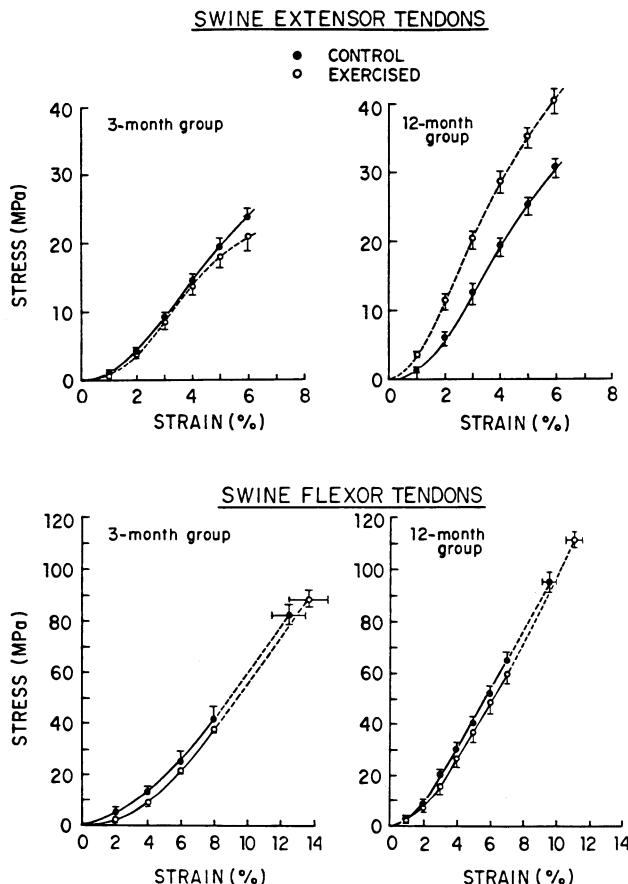


FIGURE 14.7. Curves comparing the effects of shorter (3 months) and longer (12 months) duration of exercise training on the mechanical properties of the digital extensor and flexor tendons.

There is no flexor tendon hypertrophy, probably due to the fact that tendon hypertrophy would impede its gliding within the sheath. Further, flexor tendons inherently have high strength (as tendon substance failure is not seen clinically) and are not influenced by a moderate increase in levels of stress and motion. Biochemical analysis of the flexor tendons confirm this hypothesis as the collagen concentration (on a wet weight basis) is  $28.6 \pm 0.2\%$ , which is significantly higher than that for the extensor tendon ( $23.0 \pm 0.4\%$ ) (Woo et al, 1981b).

Based on the *in-vivo* studies of tissue homeostasis, it should be possible to hypothesize a highly nonlinear relationship (Woo et al, 1976b) between stress and motion, and between tissue properties and mass for ligaments and tendons (Figure 14.8). With stress and motion deprivation, a rapid reduction of

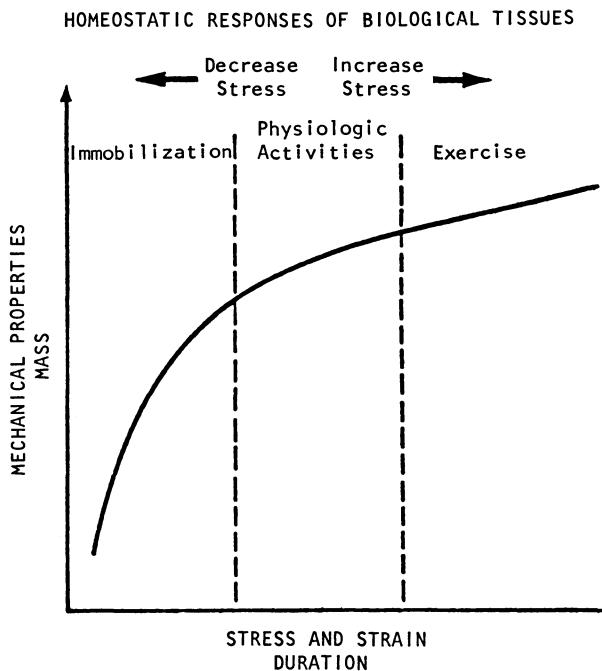


FIGURE 14.8. A hypothetical curve demonstrating the relationship between soft tissue nonlinear mechanical properties (and/or mass) and the level (and/or duration) of stress or stain it has been subjected to.

tissue properties and/or mass will occur, since the slope of the curve in this region is the steepest. Hence a short period of nine weeks of immobilization of the rabbit knee results in a profound reduction in ligament properties. On the contrary, with increased levels of stress and motion, the resultant changes of tissue properties and/or mass may be small, as the slope of the curve in this region is relatively flat. Hence the changes in swine tendon properties following exercise training are negligible at 3 months, and become significant only after 12 months of training. It should be noted that the proposed curve is only qualitative, and the goal is to illustrate a highly nonlinear functional relationship. Quantitative data to substantiate or to refute such a hypothesis will undoubtedly follow.

## 5. Role of Biomechanics in Treatment of Tissue Injuries

Although ligaments and tendons are frequently subjected to traumatic injuries, prognosis with current clinical management remains uncertain. For example, in treatment of flexor digitorum profundus tendon lacerations by

immobilization, healing is often compromised by the formation of dense adhesions between the synovial sheaths and the repair site, leading to limited tendon excursion and poor digital function. Clinical and experimental efforts to improve the understanding of digital tendon healing mechanisms have begun in recent years. It has been demonstrated that immobilization following surgical repair is deleterious to the tendon healing process, while the use of early passive mobilization can enhance the cellular activity and remodeling at the repair site and within the sheath. Biomechanical measurements of the angular rotation of the digits (i.e., excursion of the healed tendons) from the mobilized group are consistently and significantly higher than those from the immobilized group for all time periods (up to 12 weeks) studied (Woo et al, 1981c; Gelberman et al, 1982). Further, the strength and stiffness of healed flexor tendons from the mobilized groups are significantly higher.

Even less is known about ligament healing following trauma. In spite of the very high frequency of ligamentous injuries, no uniform clinical opinion exists with regard to the treatment regimen, specifically on repair vs. nonrepair and on immobilization vs. early motion. In the experimental animal studies, only limited data on the structural properties of the bone-healing ligament complex are available (Tipton et al, 1970; Vailas et al, 1981).

The lacerated canine MCL has been studied by Woo et al (1984). Methods of treatment and rehabilitation included: MCL repaired and knee immobilized (Group 1); MCL repaired and knee immobilized, plus knee remobilized (Group 2); and MCL not repaired and knee not immobilized (Group 3). At six weeks, the healed MCL was tested for stability (varus-valgus knee laxity) and tensile properties. For the motion groups (Groups 2 and 3), the knees had the least joint laxity (physiologic), and the mechanical properties of the healed ligament substance were superior (Figure 14.9). These preliminary results indicate that prolonged immobilization may not be beneficial to the MCL healing.

At present, the Class I and II MCL injuries have been advocated by most clinicians to be treated conservatively without surgical intervention. However, based on experimental studies, this consensus may not produce the most satisfactory long-term clinical results. After the mid-substance rabbit MCL is divided, the ligament can heal uneventfully, but via scar formation (Frank et al, 1983). The properties of the healed MCL (scar) are persistently inferior to that of the contralateral, control MCL up to 40 weeks post-injury, despite the compensatory increases in scar mass and size. The mechanical properties of the healing mid-ligament substance appear to plateau at about 14 weeks post-injury.

Thus it is easy to visualize that biomechanics plays an essential role in evaluating the effectiveness of the various available treatments of tendon and ligament injuries. The examples given here are only limited cases. Most of the therapeutic manipulation programs currently used during rehabilitation of ligament and tendon injuries lack an experimentally established basis. In ad-

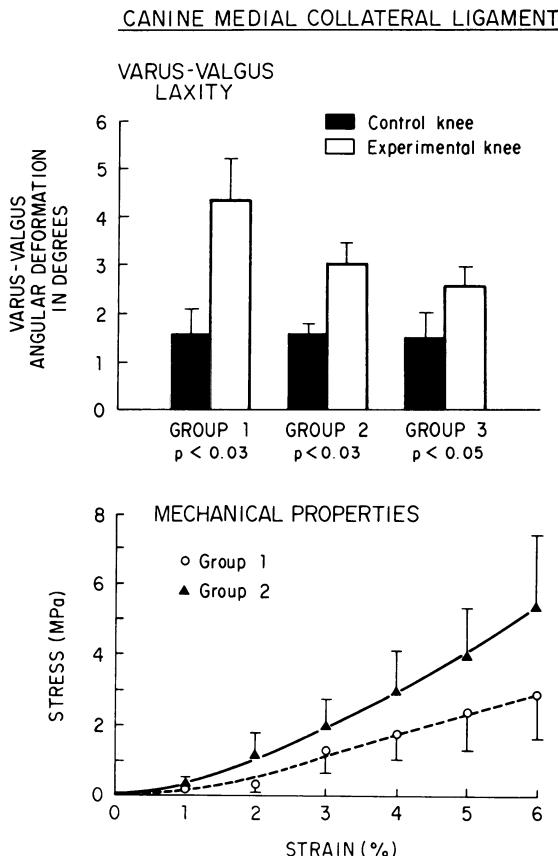


FIGURE 14.9. Prolonged immobilization (6 weeks) of the sutured canine medial collateral ligament (Group 1) is shown to result in significantly higher varus-valgus knee joint laxity and **inferior** mechanical properties of the ligament repair site as compared to ligaments with or without repair, but with earlier mobilization (Groups 2 and 3, respectively).

dition, autograft, allograft, and synthetic graft materials are being used clinically as tendon and ligament replacements without conclusive scientific data demonstrating their efficacy. It will require significant efforts in biomechanical research to gain a better understanding of the physiology of normal and injured ligaments and tendons, including not only those mentioned in this chapter, but the hundreds of other tendons and ligaments in the human body. Much remains to be done. It is hoped that, eventually, relevant information can be derived from good interdisciplinary studies and can be provided to clinicians so that the prevention and treatment of ligament and tendon injuries can be made on a sound fundamental and scientific basis.

## Acknowledgments

The author wishes to thank his many collaborators as well as his laboratory staff in the Division of Orthopaedic Surgery at the University of California, San Diego, and the San Diego Veterans Administration Medical Center. Support from NIH grants AM14918 and AM34264, the RR&D of the Veterans Administration, and the Bone and Joint Disease Foundation is gratefully acknowledged.

## References

- Abrahams, M. 1967. *Med. Biol. Eng. Comput.* 5:433–443.
- Akeson, W.H., Amiel, D., and LaViolette, D. 1968. *Exp. Gerontol.* 3:289–301.
- Akeson, W.H., Woo, S.L-Y., Amiel, D., Coutts, R.D. and Daniel, D. 1973. *Clin. Orth.* 93:356–362.
- Akeson, W.H., Woo, S.L-Y., Amiel, D., and Frank, C.B. 1984. Pages 93–148 in *Rehabilitation of the injured knee*. *Proc. Amer. Acad. Orth. Surg.*, L.Y. Hunter and F.J. Funk, eds. C.V. Mosby, St. Louis.
- Bernstein, B., Kearsley, E.A., and Zapas, L.S. 1963. *Trans. Soc. Rheology* 7:391–410.
- Bulter, D.L., Noyes, F.R., and Grood, E.S. 1978. Pages 279–314 in *CRC handbook of engineering in medicine and biology*. Section B, *Instruments and measurements*. Vol. 1, Fineberg et al, eds. CRC Press, Inc., Palm Beach, Florida.
- Chen, H.Y-L. 1973. Ph.D. thesis, University of California, San Diego.
- Dorlot, J-M., Ait ba sidi, M., and Tremblay, G.M. 1980. *J. Biomech. Eng.* 102:190–193.
- Frank, C.B., Woo, S.L-Y., and Amiel, D., 1983. *Am. J. Sports Med.* 11:379–389.
- Fung, Y.C. 1972. Pages 181–208 in *Biomechanics: its foundations and objectives*. Y.C. Fung, N. Perrone, and M. Anliker, eds. Prentice-Hall, Englewood, Cliffs, N.J.
- Fung, Y.C. 1973. *Biorheology* 10:139–155.
- Gelberman, R.H., Woo, S.L-Y., and Lothringer, K.S. 1982. *J. Hand Surg.* 7:170–175.
- Haut, R.C. 1983. *J. Biomech. Eng.* 105:296–299.
- Haut, R.C., and Little, R.W. 1972. *J. Biomech.* 5:423–430.
- Jenkins, R.B., and Little, R.W. 1974. *J. Biomech.* 7:397–402.
- Kennedy, J.C., Hawkins, R.J., and Willis. 1976. *J. Bone Joint Surg.* 58:350–355.
- Lianis, G. 1970. *Int. J. Nonlinear Mech.* 5:23–34.
- Noyes, F.R. 1977. *Clin. Orth.* 123:210–242.
- Pinto, J.G., and Patitucci, P.J. 1980. *J. Biomech. Eng.* 102:57–61.
- Price, J.M., Patitucci, P.J., and Fung, Y.C. 1977. *Am. J. Physiol (Cell Physiol.)* 2:C47–C55.
- Rosenberg, L.C. 1975. Pages 105–124 in *Dynamics of connective tissue macromolecules*. Burleigh and Poole, eds. American Elsevier, North Holland, Amsterdam.
- Sauren, H.J., and Rousseau, E.P.M. 1983. *J. Biomech. Eng.* 105:92–95.
- Simon, B.R., Coats, R.S., and Woo, S.L-Y. 1984. *J. Biomech. Eng.* 106:159–164.
- Tanaka, T.T., and Fung, Y.C. 1974. *J. Biomech.* 7:357–370.
- Tipton, C.M., James, S.L., and Mergner, W. 1970. *Am. J. Physiol.* 218:894–901.
- Tipton, C.M., Matthes, D., and Maynard, J.A. 1975. *Med. Sci. Sports Exerc.* 7:165–175.

- Vailas, A.C., Tipton, C.M., and Matthes, R.D. 1981. *Connect. Tissue Res.* 9:25–31.
- Viidik, A. 1973. *Int. rev. connect. tissue res.* Hall and Jackson, eds. Academic Press, New York, 6:127–209.
- \_\_\_\_\_. 1979. Pages 75–113 in *Progress in biomechanics*, N. Akkas, ed. Sijthoff and Noordhoff, Alphen Aaw den Rijn.
- Woo, S.L-Y., Matthews, J.V., Akeson, W.H., Amiel, D., and Convery, F.R. 1975. *Arthritis Rheum.* 18:257–264.
- Woo, S.L-Y., Akeson, W.H., and Jemmott, G.F. 1976a. *J. Biomech.* 9:785–791.
- Woo, S.L-Y., Kuei, S.C., Amiel, D., Gomez, M.A., Hayes, W.C., White, F.C., and Akeson, W.H. 1981. *J. Bone Joint Surg.* 63A:780–787.
- Woo, S.L-Y., Lubcock, P., Gomez, M.A., Jemmott, G.E., Kuei, S.C., and Akeson, W.H. 1979. *J. Biomech.* 12:437–496.
- Woo, S.L-Y., Simon, B.R., and Kuei, S.C. 1980a. *J. Biomech. Eng.* 102:85–90.
- Woo, S.L-Y., Ritter, M.A., and Amiel, D. 1980b. *Connect. Tissue Res.* 7:177–183.
- Woo, S.L-Y., Gomez, M.A., and Akeson, W.H. 1981a. *J. Biomech. Eng.* 103:293–298.
- Woo, S.L-Y., Gomez, M.A., and Amiel, D. 1981b. *J. Biomech. Eng.* 103:51–56.
- Woo, S.L-Y., Gelberman, R.H., and Cobb, N.G. 1981c. *Acta Orth. Scand.* 52:615–622.
- Woo, S.L-Y. 1982. *Biorheology* 191:385–396.
- Woo, S.L-Y., Gomez, M.A., and Seguchi, Y. 1983. *J. Orth. Res.* 1:22–29.
- Woo, S.L-Y., Gomez, M.A., Woo, Y-K., Sato, S., and Akeson, W.H. 1984. *Proc. 30th Ann. Orth. Res. Soc.* 9:108.
- Woo, S.L-Y., Orlando, C.A., Frank, C.B., Gomez, M.A. and Akeson, W.H. 1985. *Proc. 31st Ann. Orth. Res. Soc.* 10:37.
- Yamada, H. 1970. Pages 80–81 in *Strength of biological materials*, F.G. Evans, ed. Williams and Wilkins Co., Baltimore.
- Yin, F.C.T., Tompkins, W.R., Peterson, K.L., and Intaglietta, M. 1972. *IEEE Trans. Biomed. Eng.* 19:376–381.

# 15

## Bone Mechanics: From Tissue Mechanical Properties to an Assessment of Structural Behavior

W.C. HAYES

### 1. Introduction

Bone is the primary structural element of the human body. Bone is also unique among structural materials in that it is self-repairing and can alter its properties and configuration in response to changes in mechanical demand (Hayes and Snyder, 1981; Woo et al, 1981). Bone density changes are commonly observed with aging, after periods of disuse, or after chronic exercise (Jowsey, 1977; Lane and Vigorita, 1983; Woo et al, 1981).

Furthermore, certain skeletal conditions such as osteoporosis (Avioli, 1983; Jowsey, 1977) can seriously compromise the structural integrity of the skeleton. The associated reduction in bone mass gradually increases vulnerability to fracture, particularly of the femoral neck and vertebrae (Jowsey, 1977; Kelsey et al, 1978; Lane and Vigorita, 1983). The frequency of osteoporosis in the United States is well recognized, with 50% of women age 45 or older exhibiting radiographic evidence of osteoporosis of the lumbar spine (Kelsey et al, 1978). From 16–18 million women in the United States have a significant degree of vertebral atrophy, and over 4 million people of age 50 years or older have osteoporosis severe enough to cause vertebral fractures.

Osteoporosis is clearly a serious national health problem that can be expected to increase along with the average age of the population. However, despite advances in understanding its patho-physiology and major efforts to develop improved diagnostic tools, the identification of individual patients exhibiting increased fracture risk remains an elusive goal. In fact, in many clinical settings, the occurrence of spontaneous vertebral fractures provides the only conclusive evidence of disease. Although such fractures are the major clinical sequelae of osteoporosis, the biomechanics of osteoporotic bone have received little attention. Thus the structural consequences of known levels of bone loss are not well quantified.

One of the difficulties is that, except in rare instances, the accuracies of clinical procedures used in the diagnosis of osteoporosis (photon absorptiometry, computed tomography) have been evaluated only against other procedures that are themselves subject to significant measurement error (histomorphometry, neutron activation). On occasion, physical parameters such

as ash weight or apparent density are employed, but these are seldom related to the biomechanical characteristics of the measurement site or to the structural behavior of skeletal regions that exhibit high fracture risk. In light of the national health impact of osteoporosis (manifested primarily through increased rates of fracture) and the potential risk associated with many forms of therapy, this failure to focus on the biomechanical aspects of the disease represents a major challenge.

The objective of this chapter is to provide an introduction to the mechanical behavior of bone as a material and of bones as living structures. This dual objective is addressed by approaching the biomechanics of bone at two levels. At one level, we can perform standardized mechanical tests on uniform specimens and determine the material properties of the tissue. We can also test the structural behavior of a whole bone (which represents an organized construct of bone tissues in a complex geometric arrangement). It is important to realize that bone fracture represents failure of bone tissue at the material level and of the whole bone at the structural level. Predicting this fracture behavior requires a critical analytical step involving calculation of the internal stresses in the whole bone. This step involves some knowledge of the geometric features of the whole bone and of the loads being applied.

Fortunately, the biological nature and metabolic activity of bone does not, at least in the short term, change the fact that bone is a structural material whose function is to allow the various activities of daily living. As a structural material, bone is subject to the same engineering laws that govern the behavior of nonbiologic materials. This is a great convenience, since it allows us to ignore many of the complex biological activities taking place in bone. This assumption does not hold over the long term, however, since changes in bone metabolism, disease, and the normal aging process can result in important changes in the material properties of bone tissue and in the structural and geometric features of whole bones. Under these conditions, the structural behavior of whole bones cannot be expected to remain constant.

This chapter first presents a brief review of some of the most important data on the material properties of cortical and trabecular bone. With this background, the structural performance of a particular region of the skeleton (the lumbar vertebrae) is examined. The specific goal is to develop valid fracture risk predictors for this region based on computed tomography (CT). The hope is also to contribute to an improved dialogue between those interested in the biological and patho-physiological aspects of metabolic bone disease and those interested in the structural behavior of bone.

## 2. Mechanical Properties of Bone Tissue

### 2.1. Cortical Bone

To determine the mechanical properties of bone tissue, we test small, uniform specimens that are loaded under well-defined conditions. This general ap-

proach to materials testing has allowed documentation of cortical bone material properties in tension, compression, bending, and torsion (Burstein et al, 1972; Burstein et al, 1976; Evans, 1973; Hayes, 1978; Hayes and Carter, 1979; Reilly and Burstein, 1974; Reilly et al, 1974; Wright and Hayes, 1976; Wright, 1980).

Several factors influence the material properties of cortical bone. The properties are dependent upon the rate at which the bone tissue is loaded. Materials such as bone in which the stress-strain characteristics and strength properties are dependent upon the applied strain rate are said to be *viscoelastic* (or time-dependent materials). However, this rate dependency is relatively weak (Wright and Hayes, 1976). The elastic modulus and ultimate strength of bone are approximately proportional to the strain rate raised to the 0.06 power. Thus, over a very wide range of strain rates, the ultimate tensile strength increases by a factor of three, and the modulus increases by about a factor of two (Wright and Hayes, 1976).

The stress-strain behavior of cortical bone is also strongly dependent upon the orientation of the bone microstructure with respect to the direction of loading. Several investigators have demonstrated that cortical bone is stronger and stiffer in the longitudinal direction (direction of predominant osteon orientation) than in the transverse direction. Materials such as bone, whose elastic and strength properties are dependent upon the direction of applied loading, are said to be *anisotropic* materials. The viscoelastic, anisotropic nature of cortical bone distinguishes it as a complex material. Because of these characteristics, one must specify the strain rate and the direction of applied loading when describing material behavior.

Ultimate strengths of adult femoral cortical bone under various modes of loading in both longitudinal and transverse directions (Reilly and Burstein, 1974; Reilly et al, 1974) are summarized in Table 15.1. These results indicate that the material strength of bone tissue is dependent upon the type of loading

TABLE 15.1. Ultimate strength of adult femoral cortical bone.<sup>a</sup>

Loading mode	1 Ultimate strength
<b>Longitudinal</b>	
Tension	133 MPa
Compression	193 MPa
Shear	68 MPa
<b>Transverse</b>	
Tension	51 MPa
Compression	133 MPa

<sup>a</sup>Mean values from Reilly and Burstein (1974). Age span of population, 19–80 years. 1 GPa (gigapascal) = 1000 MPa.

TABLE 15.2. Modulus values for adult femoral cortical bone.<sup>a</sup>

	1
Longitudinal	17.0 GPa
Transverse	11.5 GPa
Shear	3.3 GPa

<sup>a</sup>Mean values from Reilly and Burstein (1975). Age span of population, 19–80 years. 1 GPa = 1000 MPa.

as well as on the loading direction. The compressive strength is greater than the tensile strength in both longitudinal and transverse directions. Transverse specimens are weaker than longitudinal specimens in both tension and compression.

The shear properties (determined by torsion tests about the longitudinal axis and reflecting shear stresses along transverse and longitudinal planes) is about one-third of the compressive strength. The modulus values for adult femoral cortical bone are shown in Table 15.2 (Carter and Spengler, 1978; Reilly and Burstein, 1974; Reilly et al, 1974). The longitudinal elastic modulus is approximately 50% greater than the transverse elastic modulus. The shear modulus for torsion about the longitudinal axis is about one-fifth the longitudinal modulus.

The material properties of cortical bone also decline with age (Burstein et al, 1976). Both tensile strength and modulus decrease about 2% per decade over the age range 20–90. Thus, the ultimate tensile strength declines from 140 megapascals (MPa) in the third decade to 120 MPa in the ninth decade. Over the same period, the elastic modulus decreases from 17 gigapascals (GPa) to 15.6 GPa (Burstein et al, 1976) (Figure 15.1).

## 2.2. Material Properties of Trabecular Bone

The major physical difference between trabecular bone and cortical bone is the increased porosity exhibited by trabecular bone. This porosity is reflected by measuring the apparent density (i.e., the mass of bone tissue present divided by the bulk volume of the test specimen). In the human skeleton, the apparent density of trabecular bone ranges from approximately 0.1 g/cm<sup>3</sup> to 1 g/cm<sup>3</sup>. The apparent density of cortical bone is about 1.8 g/cm<sup>3</sup>. A specimen with an apparent density of 0.2 g/cm<sup>3</sup> has a porosity of approximately 90%.

Apparent density has a profound influence on the compressive stress-strain behavior of trabecular bone (Carter and Hayes, 1976; Carter and Hayes, 1977). These stress-strain properties are similar to the compressive behavior of many porous engineering materials that can be used to absorb energy on impact.

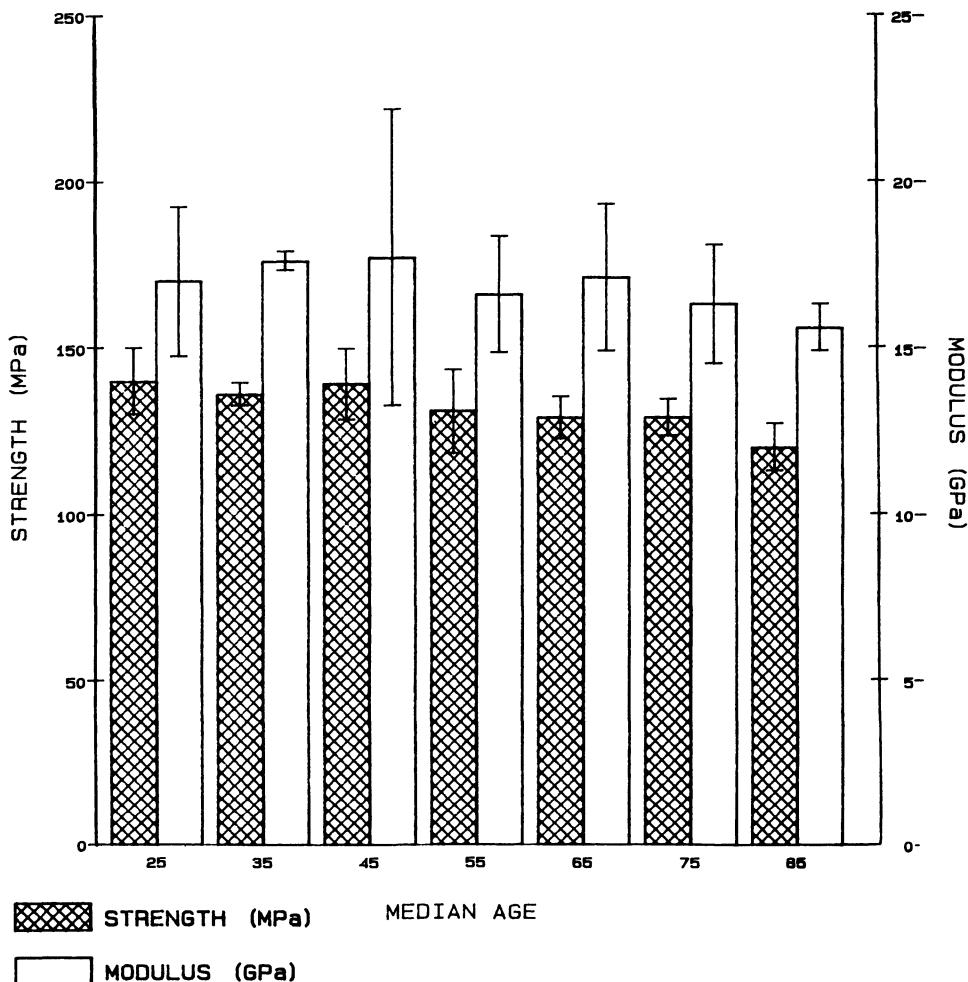


FIGURE 15.1. Tensile properties vs. age for human femoral cortical bone.

Both the compressive strength and the compressive modulus of trabecular bone (Carter and Hayes, 1976; Carter and Hayes, 1977) are markedly influenced by the apparent density of the tissue (Figure 15.2). These data include cortical bone with an apparent density of approximately  $1.8 \text{ g/cm}^3$  as well as trabecular bone specimens representing a wide range of apparent densities. These relationships indicate that the compressive strength of all bone tissue in the skeleton is approximately proportional to the square of the apparent density (Carter and Hayes, 1976; Carter and Hayes, 1977). The elastic modulus of all bone tissue is approximately proportional to the cube of the apparent density. Although these relationships were initially derived from compression

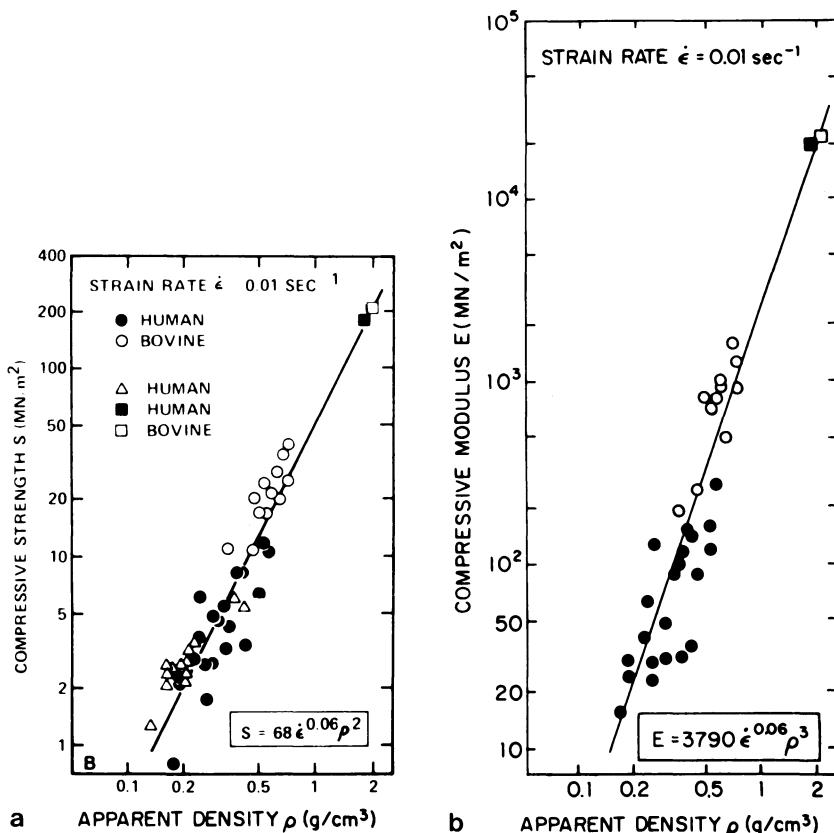


FIGURE 15.2. (a) Compressive strength vs. apparent density of compact and trabecular bone. (b) Compressive modulus vs. apparent density of compact and trabecular bone. Filled circles represent human and open circles represent bovine trabecular bone. Triangles represent fresh human compact bone. Filled squares represent embalmed human compact bone; open squares represent fresh bovine compact bone (from Carter and Hayes, 1977).

tests, tensile tests of trabecular bone indicate that its strength in tension is approximately the same as that in compression (Carter et al, 1980). In addition, the elastic modulus of trabecular bone is approximately the same in both tensile and compressive loading.

These relationships between mechanical properties and the apparent density of bone tissue are of profound physiological and biomechanical importance. First, they indicate that bone tissue can generate large changes in modulus and strength by small changes in bone apparent density. Conversely, subtle changes in bone apparent density result in large differences in strength and modulus. Note that bone density changes are usually not evident radio-

graphically until the bone density has been reduced by 30–50%. The power-law relationships shown in Figure 15.2 indicate that such reductions in bone density result in nearly order-of-magnitude reductions in bone stiffness and strength.

### 2.3. Metaphyseal Shell

The structural contributions of the thin metaphyseal shell represents a major uncertainty in understanding increases in fracture risk associated with osteoporosis. Almost all previous studies of metaphyseal regions have assumed material property values for these regions that are close to the values for diaphyseal cortical bone described above. In addition, shell thicknesses ranging from 0.5 to 2.0 mm have been assumed in most previous studies of the structural behavior of metaphyseal regions. Consequently, most workers have concluded that the metaphyseal shell plays a major structural role in joint loading and in the resistance of vertebral bodies to fracture. To provide direct measurements of the mechanical properties of the metaphyseal shell, we conducted experiments on small specimens removed from the proximal tibial metaphysis.

Cadaveric proximal tibiae were harvested from fresh cadavers (age 59–90, mean age 74) and small (9.5-mm) plugs containing the shell and underlying trabecular bone were removed. Trabecular bone was carefully ground away to produce a thin disk of constant thickness. Specimens were prepared from: (1) proximal metaphysis (0–3 cm distal from the tibial plateau), (2) distal metaphysis (3–5 cm), and (3) the proximal diaphysis (5–10 cm). Specimens were tested as simply supported, centrally loaded, circular plates; the modulus, yield strength, and ultimate strength were determined by plate theory. The validity of the method was assessed by using thin aluminum disks of similar geometry.

The values for yield strength and modulus of the metaphyseal shell for these regions are shown in Table 15.3. These values indicate that the modulus and yield strengths in the metaphyseal regions are far less than has been previously assumed. In the most proximal region, the yield strength (Table 15.3) is less than half the tensile strength of diaphyseal cortical bone (Table 15.1). The modulus is about one-tenth the modulus value for cortical bone of the shaft.

TABLE 15.3. Yield strength and modulus of metaphyseal shell.<sup>a</sup>

Region	1 n	2 Elastic modulus (MPa)	3 n	4 Yield stress (MPa)
Proximal metaphysis	12	1434.13 (461.06)	12	58.18 (32.91)
Distal metaphysis	28	5038.79 (980.02)	24	105.67 (18.19)
Proximal diaphysis	39	7056.32 (873.08)	28	121.33 (13.09)

<sup>a</sup> Mean values (95% Confidence Interval).

Our modulus values of  $7.18 \pm 0.87$  GPa in the proximal diaphysis are lower than the reported longitudinal values for cortical bone, but are in good agreement with data in the circumferential direction.

These data indicate that previously assumed modulus values for the metaphyseal shell are high by a factor of from 10 to 20. The findings also suggest that the axial rigidity of the metaphyseal shell may be more than an order of magnitude less than has been assumed. These findings have strong significance for the structural contributions of these regions, as we shall see below in a study of failure characteristics of lumbar vertebrae.

### 3. Structural Behavior of Lumbar Vertebrae

Fracture occurs when the local stresses exceed the ultimate strength of the bone in that region. The imposed stresses can be either cyclic, resulting in fatigue failure, or traumatic, resulting in catastrophic failure. Bone fracture can therefore be viewed as an event that is initiated at the material level and then affects the load-bearing capacity of the bone at the structural level. Any attempt to predict the structural behavior of a skeletal region must therefore reflect both the heterogeneous material properties of the region and the complex geometric arrangement of each type of bone.

Current research underway in our laboratory for predicted vertebral fracture risk using noninvasive techniques (McBroon et al, 1984) provides a practical example for the application of some of these biomechanical principles. Our objectives are to establish the accuracy of computed tomography (CT) as a measure of vertebral trabecular bone apparent density and to determine whether noninvasive techniques can be used to identify vertebrae "at risk" for the development of spontaneous fractures.

#### 3.1. Prediction of Vertebral Fracture Risk

Intact lumbar spines (T11/L5) were harvested within 12 hours of death in a series of necropsy specimens (age 63–99, mean age 78). Lumbar vertebrae L1 and L3 were designated as target vertebrae so that the endplates and disc immediately above and below these vertebrae could be incorporated in the mechanical testing procedures. Computed tomographic scans were obtained with a General Electric 8800 scanner with the spine immersed in a water bath to improve CT resolution and to avoid scatter. A phantom with chambers filled with water, ethyl alcohol, and known concentrations of dipotassium hydrogen phosphate ( $K_2HPO_4$ ) (5, 10, 20, 30 g percent) was utilized to calibrate and monitor the accuracy of the scanner (Genant et al, 1981). For each scan, CT values (Hounsfield units) were obtained from each chamber of the phantom, and from six preselected locations within the vertebral body. An average CT value for the interior trabecular bone was then determined by

using the cursor to outline a central region, positioned two pixels from the vertebral cortex and from the entrance of the basivertebral vein. By using a linear regression of the CT values for water and the solutions for  $K_2HPO_4$ , a corrected CT value for each measurement was calculated (Genant et al, 1981).

For mechanical testing, each target vertebra was removed together with the proximal and distal endplates and discs. The posterior elements were removed by division of the pedicles. The endplates above and below the target vertebra were embedded in a test rig so that load was applied perpendicular to the endplates. Uniaxial compressive loads were applied using an electrohydraulic materials testing machine at a constant deformation rate of 0.01 mm/s. Testing was discontinued immediately upon an abrupt change in the slope of the load deformation curve with a loss of load-bearing capacity. After testing, the discs were removed and the endplates of the target vertebra examined for fracture. After removal of the endplates, the vertebral body was machined under continuous irrigation to form six cylinders of trabecular bone. The cylinders measured 9.5 mm in diameter and 10–18 mm in height and were composed of the trabecular bone from which CT values had previously been determined. Hydrated weights of these specimens were calculated on an analytical balance, and apparent density was calculated by standard methods (Carter and Hayes, 1976; Carter and Hayes, 1977).

From these data, computed tomography appears to provide a very accurate quantitative measurement of bone apparent density. Once the CT values are corrected in terms of the reference calibration scale, there is a highly significant positive correlation ( $R^2 = 0.89; p < 0.001$ ) between apparent density and corrected CT number. These calibration procedures are essential, however. Despite frequent internal calibrations of the scanner so that the CT value of water was set at 0, we found variations from scan to scan of up to 10 Hounsfield units for the density of water. This correlation of corrected CT value to apparent density did not vary between spines or between individual vertebrae.

As we showed earlier (Carter and Hayes, 1976; Carter and Hayes, 1977), the compressive strength of trabecular bone is proportional to the apparent density squared. The most important determinants of the compressive strength of a region of trabecular bone are thus its apparent density and its cross-sectional area. Plotting vertebral failure load against the average apparent density squared, multiplied by the cross-sectional area (Figure 15.3), results in highly significant correlations with vertebral failure load ( $R^2 = 0.81, p < 0.001$ ). This indicates that noninvasive determinations of vertebral density and cross-sectional area can be used with precision to predict vertebral failure load under these test conditions. These results thus suggest a simple but reliable noninvasive method for predicting vertebral fracture risk *in vivo*.

Surprisingly, the relationship between vertebral failure load and the noninvasive determinations of trabecular density and area is relatively simple, indicating that axial compressive stresses in the interior trabecular bone are an important feature of vertebral loading. We do not have to resort to more sophisticated models to represent the relative contributions of cortical and

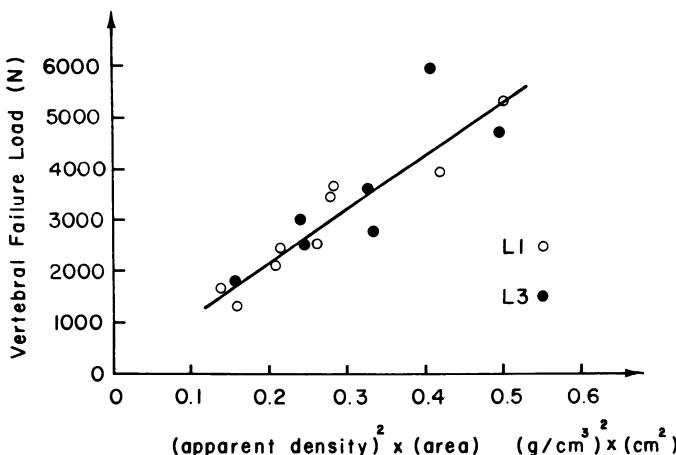


FIGURE 15.3. Correlation between vertebral failure load and average apparent density squared multiplied by cross-sectional area of trabecular bone within vertebral body ( $R^2 = 0.81, p < 0.0001$ ) (from McBroom et al, 1985).

trabecular bone, as might have been assumed from previous studies (Atkinson, 1966; Bradley et al, 1978; Kazarian and Graves, 1977; Rockoff et al, 1975). It remains to be determined from retrospective or prospective clinical studies whether this simple model will provide a useful tool for identifying individual patients at increased risk for spontaneous vertebral fracture.

### 3.2. Vertebral Cortical Shell

These findings would also suggest that, under these loading conditions, the thin shell surrounding the trabecular bone in the vertebral centrum makes only a modest contribution to the structural strength of the vertebral body. However, investigations of the distribution of vertebral body load have resulted in considerable controversy (Bartley et al, 1966; Bell et al, 1967; Rockoff et al, 1975; Evans, 1973). In the most recent of these studies, Rockoff et al, (1975) reported that the cortex contributes 45–75% of the peak strength of the vertebral body under compressive loading.

We assessed the relative contribution of the cortical shell and trabecular bone to the strength of the vertebral body by using 40 L1 and L3 vertebrae, 20 from the above experiments and 20 additional vertebrae excised from 10 additional spines. The 40 vertebrae were divided into two groups of L1, L3 specimen pairs. To determine if the compressive failure loads from the same spine were similar, the cortex from the first and third lumbar vertebrae from 10 spines was removed by painstaking manual sanding, with frequent water picking to prevent inadvertant damage to the underlying trabecular bone. The importance of the cortex to the vertebral body was assessed by comparing the

TABLE 15.4. Cortical shell contribution to overall strength of vertebral body.<sup>a</sup>

Vertebral level	1 n	2 Failure load (N)	3% Difference
L1 (intact)	10	3160 (891)	
L3 (intact)	10	3385 (995)	-6.6%
L1 (w/o cortex)	10	2771 (737)	-12.8%
L3 (intact)	10	3176 (618)	

<sup>a</sup>Mean values (95% confidence interval).

difference in strength of L1 and L3 for the 10 spines from which the cortex of L1 had been removed.

From these experiments, the cortical shell appears to contribute very little to the overall strength of the vertebral body (Table 15.4). For vertebral bodies tested intact, the average failure load for L1 was 3160 N ( $n = 10$ , 95% Confidence Interval (C.I.) = 891 N) and for L3 was 3385 N ( $n = 10$ , 95% C.I. = 995 N). Thus, with the vertebral cortex intact, the average failure load of L1 was 6.6% less than L3 ( $n = 10$ , 95% C.I. = 1.9%). In the 10 additional spines, for L1 vertebrae tested without cortex, the average failure load for L3 vertebral bodies was again determined with the cortex intact and found to be 3176 N ( $n = 10$ , 95% C.I. = 618 N). A grouped comparison showed that the two groups of intact L3 vertebrae were not significantly different ( $p > .05$ ), indicating that the two groups of vertebral specimens could be assumed to be from the same population. In the second set of 10 spines, the mean percent difference in strength between L3 and L1 (without cortex) was 12.8% ( $n = 10$ , 95% C.I. = 11.4). When this percent difference is compared by unpaired t-test with the percent difference in strength of L1 and L3 for the 10 spines with an intact L1 cortex, this difference is significant ( $p < 0.05$ ), but far less than the 40–75% reduction in strength found by Rockoff et al (1975).

Thus, our results show that, with appropriate calibration, single-energy quantitative computed tomography can be used to predict the local apparent density of trabecular bone. The average apparent density of the interior trabecular bone can be used to predict the failure load of vertebral bodies, with posterior elements removed in response to uniaxial compressive loads. Multiplying the squared apparent density of the vertebral trabecular bone by the area ( $p^2 A$ ) provides a highly significant predictor of vertebral failure load ( $R^2 = 0.81$ ,  $p < .0001$ ) under these *in-vitro* test conditions. It remains to be demonstrated from retrospective or prospective clinical studies whether this simple parameter will provide a useful tool for identifying individual patients at increased risk for spontaneous vertebral fractures.

## 4. Summary

We have shown here that the structural behavior of bone can be approached on both a material level and a structural level. The material level is investigated by preparing specimens of regular geometry and determining their mechanical properties. Structural behavior, on the other hand, includes contributions from both the mechanical properties (at the material level) and the effects of complex bone geometries.

At the material level, cortical bone is both viscoelastic and anisotropic. For trabecular bone, the most important parameter controlling mechanical properties is the apparent density. Strength is related to the square of the apparent density and modulus to the cube of the apparent density. These power-law relations hold for all bone throughout the skeleton and indicate the remarkable sensitivity of bone to changes in porosity. The metaphyseal shell exhibits mechanical properties far less than those of cortical bone, indicating that the contribution of this material to the structural behavior of metaphyseal regions is far less than has been commonly assumed.

These approaches can also be applied to the development of noninvasive, biomechanical predictors of fracture risk for vertebral bodies. We have shown that quantitative computed tomographic scans can be used with suitable precision to predict the apparent density of trabecular bone in the vertebral centrum. These predictions can in turn be used to estimate the compressive strength of trabecular bone in the vertebral centrum and thereby to predict failure loads for the vertebral body. The strong agreement between the predicted failure loads and those actually measured *in vitro* suggests that similar methods can be used to develop noninvasive predictors of fracture risk for other regions of the skeleton. The results also suggest that the structural contribution of the vertebral cortical shell is far less than has been previously assumed, a finding that is in apparent agreement with our direct measurements of mechanical properties of the metaphyseal shell.

The vertebral fracture risk example suggests the exciting possibility of predicting (and therefore assessing) the *in-vivo* structural capacity of regions of the skeleton. By using available data on the mechanical properties of bone tissues, a simple biomechanical analysis could be used to develop CT-based, noninvasive procedures for the prediction of fracture risk in the vertebral body. While only simplified loading situations were used, the predictive capability of this analysis is sufficiently precise to suggest both testing these predictors in clinical settings and extending the approach to other vertebral loading modes.

It would also be of great interest to extend these analytical predictors to skeletal regions, such as the proximal femur, for which an increased fracture risk is both observed clinically and is of particularly dire consequence. However, the geometric, material, and loading complexities of such regions require that even more sophisticated analytical techniques (such as the finite-element method) be used for stress prediction. Such analytical techniques are available,

however, and these methods, coupled with improved experimental data on the material properties of metaphyseal bone in the proximal femur, could be used to provide accurate predictors of fracture risk in individual patients. The value of such predictors would be of obvious significance in guiding the assessment and treatment of osteoporosis.

## Acknowledgments

We gratefully acknowledge the support of RCDA AM 00749 and of the Angus Foundation.

## References

- Atkinson, P.J. 1966. *Calc. Tiss. Res.* 1:24-32.
- Avioli, L.V., ed. 1983. *The osteoporotic syndrome: detection, prevention and treatment.* Grune and Stratton, New York.
- Bartley, M.H., Arnold, J.S., Haslam, R.K., and Jee, W.S.S. 1966. *J. Gerontol.* 21:517-521.
- Bell, G.H., Dunbar, O., and Beck, J.S. 1967. *Cal. Tiss. Res.* 1:75-86.
- Bradley, J.G., Huang, K.K., and Ledley, R.S. 1978. *Radiology* 128:103.
- Burstein, A.H., Currey, J., Frankel, V.H., and Reilly, D.T. 1972. *J. Biomech.* 5:35-44.
- Burstein, A.H., Reilly, D.T., and Martens, M. 1976. *J. Bone Joint Surg.* 58-A:82-86.
- Carter, D.R., and Hayes, W.C. 1976. *Science* 194:1174-1176.
- \_\_\_\_\_. 1977. *J. Bone Joint Surg.* 49-A:954-962.
- Carter, D.R., Schwab, G.H., and Spengler, D.M. 1980. *Acta Orth. Scand.* 51:733-741.
- Evans, F.G. 1973. *Mechanical properties of bone.* Charles Thomas Co., Springfield, Illinois.
- Genant, H.K., Boyd, D., Rosenfeld, O., Abols, Y., and Cann, C.E. 1981. Pages 121-149 in *Non-invasive measurement of bone mass and their clinical application*, S.H. Cohen, ed. CRC Press, Inc., Boca Raton, Florida.
- Hayes, W.C. 1978. Pages 333-372 in *CRC handbook of engineering in medicine and biology*. Section B, *Instruments and measurement*. Vol. 1, A. Burstein and E. Bahniuk, and E. Bahniuk, eds. CRC Press, Inc., Cleveland, Ohio.
- Hayes, W.C., and Carter, D.R. 1979. Pages 263-300 in *Skeletal research: an experimental approach*, D. Simmons and A.S. Kunan, eds., Academic Press, New York.
- Hayes, W.C., and Snyder, B.D. 1981. Pages 43-68 in *Sympos. on mechanical properties of bone*. Amer. Soc. Mech. Eng., Boulder, Colorado.
- Jowsey, J. 1977. *Metabolic diseases of bone.* W.B. Saunders Co., Philadelphia.
- Kazarian, L.E., and Graves, G.A. 1977. *Spine* 2:1-14.
- Kelsey, J.L., Pastides, H., and Bisbee, G.E. 1978. *Musculoskeletal disorders: their frequency of occurrence and their impact on the population of the United States.* Prodist, New York.
- Lane, J.M., and Vigorita, V.J. 1983. *J. Bone Joint Surg.* 65-A:274-278.
- McBroom, R.J., Hayes, W.C., Edwards, W.T., Goldberg, R.P., and White, A.A. 1984. *J. Bone Joint Surg.*, in press.
- Reilly, D.T., and Burstein, A.H. 1974. *J. Bone Joint Surg.* 56-A:1001-1022.
- Reilly, D.T., Burstein, A.H., and Frankel, V.H. 1974. *J. Biomech.* 7:271-275.

- Rockoff, S.D., Sweet, E., and Bluestein, J. 1975. *Calc. Tiss. Res.* 3:163-175.
- Woo, S.L-Y., Kuei, S.C., Amiel, D., Hayes, W.C., White, F.C., and Akeson, W.H. 1981. *J. Bone Joint Surg.* 63-A:780-786.
- Wright, T.M. 1980. Pages 252-258. *Scientific foundations of orthopaedics*, O. Goodfellow, and P. Bullough, eds. William Heinemann Medical Books, Ltd., London.
- Wright, T.M., and Hayes, W.C. 1976. *Med. Biol. Eng.* 14:671-680.

# 16

## Biomechanics of the Lumbar Spine

A.I. KING and K.-H. YANG

### 1. Introduction

Although the anatomy of the human spine is not drastically different from that of quadrupedal vertebrates, its role in supporting the head and torso is obviously not the same. The erect posture of a biped imposes a gravitational compressive load and a forward flexion moment on the vertebral column. It is hypothesized that muscular action or muscle tone is used to maintain this posture since a ligamentous spinal specimen is quite incapable of bearing any eccentric load. The action of the back muscles is analogous to that of cables of a pylon that is strengthened by the tension in the cables. The load-bearing system is made more complex by the curvatures found in the thoracolumbar spine and by the nature of the intervertebral joints. Thus, the biomechanics of spinal support and function can be understood only with a detailed knowledge of its anatomy.

There is considerable interest in the biomechanics of the lumbar spine because of the high incidence of low back pain (LBP) in Western society. White (1982) indicated that 80% of the population in this society will have a significant complaint of back pain at one time or another in their lives and that the cost of treatment and compensation is in excess of \$14 billion per year, not including loss of productivity. Despite its prevalence, the causes of LBP are difficult to identify. It is frequently called idiopathic because the cause is unknown or indeterminate. However, there is sufficient evidence to associate LBP with mechanical agents, and biomechanical analyses need to be made in order to find mechanical sources of LBP.

The aim of this chapter is to present a finite-element model of a spinal motion segment that studies the response of the spine under load. It utilizes experimental data and results from previous work for its formulation and for deriving the material properties needed by the model. An attempt was also made to use model predictions as a hypothesis for LBP.

## 2. Anatomy of the Lumbar Spine

The spine consists of 24 bones, called vertebrae, that are linked together by intervertebral discs and ligaments to form a column-like structure. It is located in the torso near the posterior border along the midline (midsagittal plane). The first seven vertebra below the head form the cervical spine, which usually has a lordotic curve (convex anteriorly). Below the cervical spine is the thoracic spine, made up of 12 vertebrae. Its curvature is kyphotic (convex posteriorly). The last five vertebrae form the lumbar spine, which is also a lordotic curve. A lateral view of the spine is shown in Figure 16.1.

In general, each vertebra is made up of an anterior body, which is almost cylindrical in shape, and posterior elements consisting of the arch (pedicles),

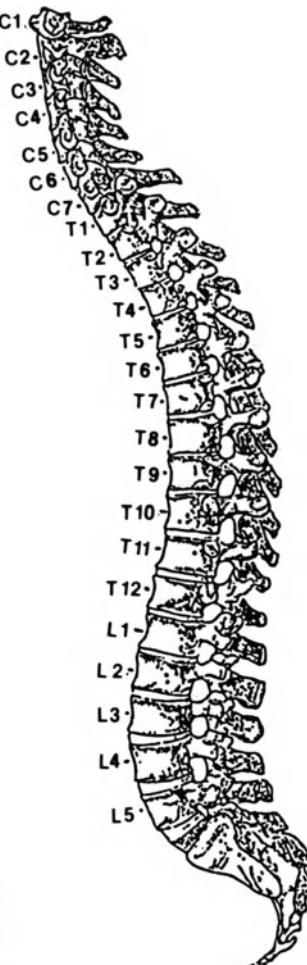


FIGURE 16.1. Lateral view of the human vertebral column (Reproduced with permission from Gray's Anatomy, 1973).



FIGURE 16.2. Side and rear view of a lumbar vertebra.

the laminae and facet joints, and the processes. Figure 16.2 shows the lateral and rear view of a lumbar vertebra. The body has a thin layer of cortical (compact) bone around its circumference and is mainly composed of spongy bone capped superiorly and inferiorly by cartilaginous endplates. The sides of the body are not straight but are concave inwards, forming a waist at the center. The pedicles are located on each side of the posterior body, projecting rearward from the superior half. They form the sides of the spinal canal that surround and protect the spinal cord. The posterior wall of the canal consists of two laminae that are flat, quadrilaterally shaped bones. At the junction between the pedicles and the laminae are the four articular facets. They are shown in the rear view. The inferior facet of a vertebra articulates with the superior facet of the vertebra below it, forming a synovial joint. For lumbar vertebrae, the orientation of the facet surfaces are almost vertical and are generally curved when viewed from above.

The vertebrae are joined together by soft tissue—anteriorly by ligaments and intervertebral discs, and posteriorly by ligaments and facet joint capsules. Intervertebral discs are cartilaginous tissue consisting mainly of collagen, proteoglycans, and water. The two main regions of the disc are the annulus fibrosus and the nucleus propulsus. The former is a ring of collagen fibers found in layers having different fiber orientations. In the nucleus, there is a higher proteoglycans content, resulting in a gel-like appearance of the propulsus. The anatomy and function of the disc are dependent on its stage of degeneration. The number and size of collagen fibrils increase with age, and the distinguishing features of the nucleus disappear as age transforms the entire disc into fibro-cartilage. It is always safe to assume that normal, healthy discs in adults are the exception rather than the rule. More details can be found in Peacock (1952) and Buckwalter (1982).

The articular facets are enclosed by a joint capsule and appear to allow the spine to flex freely but to act as motion-limiters in extension. The joint surfaces are lined with articular cartilage and are lubricated by synovial fluid. Three

spinal ligaments run the entire length of the spine; these are the anterior and posterior longitudinal ligaments and the supraspinous ligament. The ligamentum flavum is a strong band that connects adjacent laminae along the posterior aspect of the spinal canal. The interspinous ligament is a thin membrane that fills the space between adjacent spinous processes.

There are also many muscles attached to the spine. Most of them are extensors that are used to maintain an erect posture and to allow the performance of lifting tasks. These back muscles have been divided into two groups by Gray (1973): the superficial stratum, called the transversocostal and splenius group, and the deeper stratum, called the transversospinal group. The reader is referred to a standard text in human anatomy such as Gray (1973) for a detailed description of these muscles.

### 3. A Review of the Literature

This review is concerned with the biomechanics of load bearing as it relates to the lumbar spine, the modeling of the lumbar spine, and the possible mechanical causes of LBP. It is evident from the complex anatomy of the lumbar spine that it is not a simple column supporting an axial compressive load. The multisegmented curved structure has two load paths, one through the bodies and discs and the other via the articular facets. King (1984) reviewed the literature dealing with the existence of facet loads, beginning with studies on the vertical accelerator at Wayne State University. Yang and King (1984) provided conclusive evidence that the facet joints were load-bearing elements of the lumbar spine and that the mechanism whereby the load was transmitted was the bottoming out of the tip of the inferior facet on the pars interarticularis of vertebra below. The spine is thus composed of a stack of three-legged stools, one leg of which (the disc) bears more load than the other two (the facets). Loads were applied to a motion segment with eccentricities of up to 10 mm anterior and posterior to the center of the vertebral body. A specially designed load cell was inserted into the vertebral body to measure disc load. It was found that facet load increased from 17% to 32% of the total load for a 20-mm rearward movement of the applied load. These data do not include the effects of muscle tension because the cadaveric specimens used were unable to simulate this action. However, in any lifting activity, flexion moment is apt to be rather large and, since extensor muscles have a small moment arm, the forces generated by the muscles to balance this moment are large. Schultz and Andersson (1981) have developed a model to compute spinal load and muscle forces across a single transverse section of the torso. The forces in the extensors due to a 40-Newton (N) weight held at a distance of 400 mm from the spine are estimated to be as high as 664 N. Experimental verification of muscle loads is still beyond the state of the art of current measurement technology. Electromyographic (EMG) data can be used only to indicate the level of activity of a given muscle and provide a qualitative estimate of the force generated in muscle.

Mathematical models can be used to simulate muscular action. Spinal models that have this feature include those proposed by Tennyson and King (1977), Pontius and Liu (1976), and Belytschko and Williams (1982). These are models of the entire column or a portion of the column and do not treat the mechanism of load transmission in detail. To achieve this aim, it is necessary to develop a detailed model of a motion segment of the spine, consisting of two vertebrae and an intervening disc. Yang et al (1983) proposed this model but did not simulate muscle action. This feature has been added to the model and is presented in this chapter along with a detailed description of the model. The rationale for the inclusion of muscle forces in the model is their contribution to the loading of the facets. The significance of facet load as a possible source of LBP cannot be overemphasized. It was shown by Yang and King (1984) that an overloaded facet joint results in rearward rotation of the inferior facets and stretching of the capsular ligaments. Pedersen et al (1956) demonstrated the rich innervation of these ligaments, and Wyke (1982) was able to cause the mechano-receptors to fire by stretching the joint capsule. It was thus hypothesized that the facet joint could be a source of LBP not related to disc rupture or pressure of a protruding disc on a spinal nerve. In fact, there is a wealth of clinical evidence of facet pain. Ghormley (1933) coined the term *facet syndrome*, implying that the facets might be one of the sites for LBP. Other studies that implicate the facet joints include the work of Lewin (1964), Shealy (1976), and Mooney and Robertson (1976).

#### 4. Material Properties of Spinal Tissues

Mechanical properties of each component of the motion segment are important for the proper simulation of the load-bearing role of the spine. To date, all mechanical tests were carried out *in vitro* with the exception of the intradiscal pressure measurements made *in vivo* by Nachemson and Morris (1964). Reviews of material properties were written by Rolander (1966) and Schultz (1974).

For the vertebral body, Rockoff et al (1969) estimated that the cortical shell supported 25 to 55% of the total applied compressive load. The elastic modulus (Young's modulus) is about 12,000 MPa (Carter and Hayes, 1977). Yamada (1970) reported an elastic modulus of 90 MPa for cancellous bone, while Carter and Hayes (1977) obtained a value of 211 MPa at a strain rate of 10/s. The compressive strength of the whole vertebra is about 4,000 to 7,000 N. Sonoda (1962) observed that the vertebral body fractured before failure of the disc.

In a young healthy disc, nucleus pressure was found to be hydrostatic, and annular pressures were about the same as those in the nucleus. Other mechanical properties include the modulus of its annulus fibers. Their approximate value of 4 MPa was measured by Wu and Yao (1976), who performed tests without regard to the orientation of the fibers. It was felt that the modulus is

TABLE 16.1. Material properties used in the model.

Description	Modulus MPa (psi)	Poisson ratio	Thickness mm (in)
Centrum—trabecular bone	207 (30,000)	0.2	—
Disc—nucleus (bulk modulus)	2,255 (327,000)	—	—
Disc—annulus	39 (5,700)	0.45	—
Laminae, facets, pedicles	11,032 (1,600,000)	0.25	—
Endplate—peripheral	12,480 (1,810,000)	0.28	0.800 (0.032)
Endplate—central	12,480 (1,810,000)	0.28	0.500 (0.02)
Vertebral body—cortex	11,032 (1,600,000)	0.25	0.640 (0.25)
Ant. and post. long. lig.	21 (3,000)	0.25	1.020 (0.04)
Processes	12,480 (1,810,000)	0.28	1.270 (0.05)
Ligament stiffness	40 N/mm		
Facet joint stiffness	1,300 N/mm		

much higher for the entire disc, and a value of 39 MPa was used in the model.

For the posterior elements, Yang and King (1984) found the stiffness of the facet joints to be nonlinear and to be an order of magnitude higher in compression than in tension. An average value of 1,300 N/mm was used for compressive response. Cyron et al (1976) observed that the pedicle–body junction was subjected to the highest strain during compressive loading.

Mechanical properties of the spinal ligaments are rather sparse. The main difficulties are to identify the spinal ligaments precisely and to measure their cross-sectional area. For the anterior longitudinal ligament, Shah et al (1978) reported a mean value of 12 MPa, while Tkaczuk (1968) found the modulus to range from 23 to 19 MPa for ages 26 to 64 years. For subhuman primates, Little et al (1983) reported a mean value of 38 MPa. Shah's (1978) data for posterior ligaments were high compared to Tkaczuk's (1968) results of 21 to 18 MPa for the same age range stated above. The corresponding subhuman primate value obtained by Little et al (1983) is 54 MPa. Panjabi et al (1984) found that ligament response was nonlinear and reported their results as load-deflection curves because the cross section of the ligaments was irregular and difficult to measure. For deformations of less than 1 mm, the stiffness of the ligaments posterior to the vertebral body averaged 40 N/mm.

The values of material properties used in this model are tabulated in Table 16.1. They represent the best estimate of the average value and are used with the knowledge that biological tissues can have a wide variation in physical properties among subjects.

## 5. Finite-Element Model of a Lumbar Motion Segment

Finite-element models of a single vertebra have been proposed by Hakim and King (1979) and by Balasubramanian et al (1979). Finite-element models of a vertebral joint have been proposed by Belytschko et al (1974), Kulak et al

(1976), Lin et al (1978), and Shirazi-Adl et al (1983). An attempt was made by Koogle et al (1979) to model a motion segment using the basic model developed by Hakim and King (1979). Spring elements were used to simulate the disc, but no results were presented. The model described in this chapter is that of a bilaterally symmetric motion segment, consisting of two lumbar vertebrae and a disc. The geometry of each vertebra was similar to that used by Hakim and King (1979), including the complete simulation of the posterior elements. It is till a linear three-dimensional model, but the number of nodes and elements were reduced to minimize computational cost. Figure 16.3 shows the side and top view of the finite-element model. Each vertebra was made up

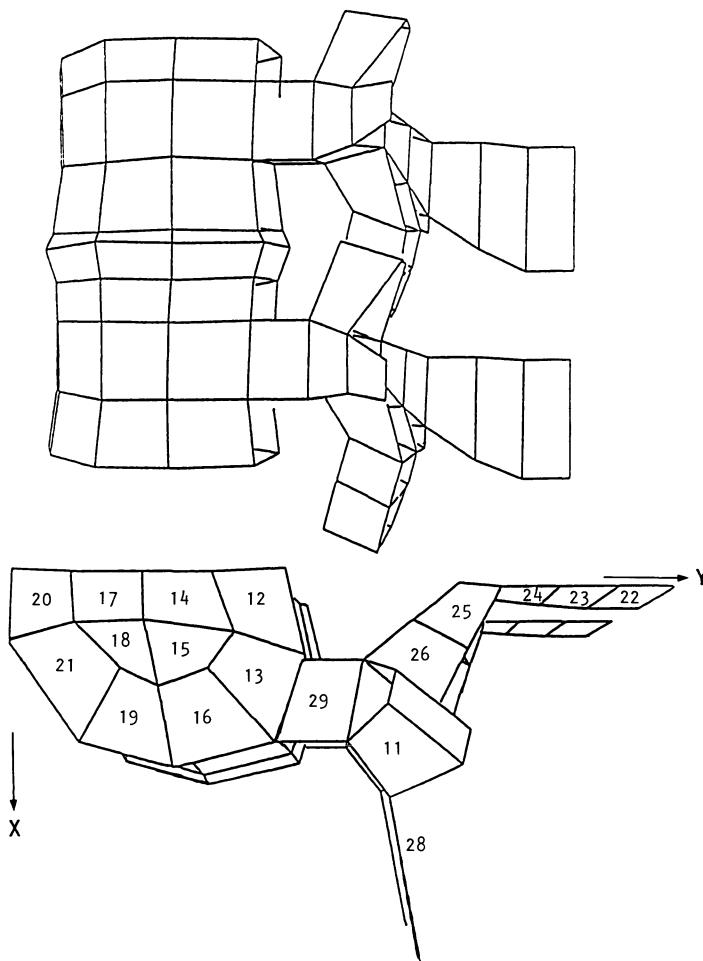


FIGURE 16.3. Side and top view of a finite element model of a lumbar motion segment.

of three layers of brick elements, while the disc consisted of two layers.

The finite-element code used was a version of FEAP (Finite Element Analysis Program). The three types of elements selected from its element library were the three-dimensional 8-noded hexahedral (brick) elements, thick/thin shell elements, and elastic bar elements. The brick elements were used to simulate either an elastic continuum or an inviscid compressible fluid. Elastic elements were used for cancellous bone of the centrum, cortical bone of the posterior elements (pedicles, laminae, and facets), and the annulus fibrosus of the disc. The nucleus propulsus was represented by fluid brick elements that assumed the bulk modulus of water (2,255 MPa). The vertebral endplates, the cortex of the vertebral bodies, and the processes were simulated by the shell element. Those shell elements covering the disc were given a lower modulus to simulate the anterior and posterior longitudinal ligaments. The resistance of the facet joints was modeled by five bar or spring elements capable of responding to flexion-extension and anterior-posterior shear between the two vertebrae. To provide the model with the capability of withstanding an eccentric load, a rigid plate consisting of 10 brick elements was placed on top of the motion segment. The applied load was placed anterior to the motion segment, and the counter-balancing muscle force was located just posterior to the spinous process. The rigid plate was constrained to move axially in the vertical direction to prevent excessive bending of the motion segment that can occur if it were also allowed to rotate. It was further assumed that the muscles most effective in providing spinal extension span the lumbar spine, with their origin and insertion points on the pelvis and rib cage, respectively. A total of 268 nodes were used to generate 207 elements. A breakdown of the various elements is shown in Table 16.2.

TABLE 16.2. List of elements used in the model.

<b>8-Noded elements</b>	
Vertebral body, spongy bone	60
Disc nucleus	8
Disc annulus	12
Laminae, facets, pedicles	18
Extension bar	10
<b>4-Noded elements</b>	
Cortical bone of vertebra	36
Endplate	40
Ant. long. lig.	4
Post. long. lig.	2
Spinous and transverse processes	8
<b>2-Noded elements</b>	
Facet joint	5
Ligaments	4
<b>Total number of elements</b>	<b>207</b>

## 6. Load Case Studies

The intervertebral joint is subjected to loads far in excess of the weight of the torso, due to muscular action. Nachemson (1981) estimated the load at L3 to be about 1,300 N, or three times the weight of the torso if the subject carried a 20-kg mass in his hands and the torso had a 20-degree forward bend. Schultz and Andersson (1981) found that large muscle forces were generated for relatively light loads carried in front of the body. The model used was a linear optimization scheme that computed individual muscle forces, assuming that the vertebral pivot point for flexion and extension was at the geometric

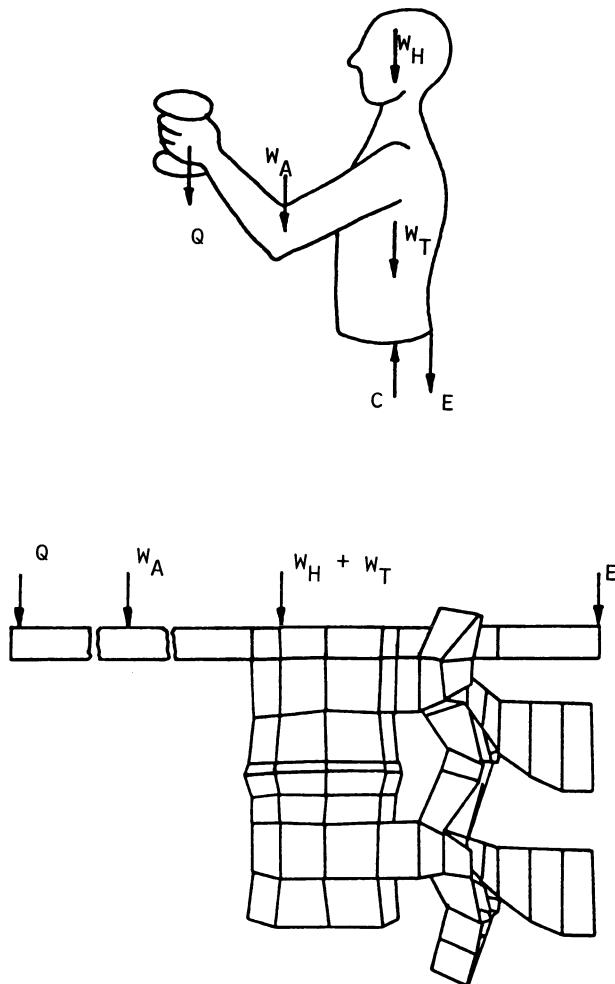


FIGURE 16.4. Loading condition simulated by the model.

**TABLE 16.3.** Table of load cases.

2 Pivot points
Geometric center of disc
Center of spinal canal
5 Loading conditions
30 N.m flexion
15 N.m flexion
Zero net moment
15 N.m extension
30 N.m extension
2 disc conditions
Normal disc
Degenerated disc
Total of 20 load cases

center of the disc. This instantaneous center has been measured by several investigators for different types of spinal motion. There is no agreement regarding its actual locus.

In this study, the lifting activity simulated was similar to that used by Schultz and Andersson (1981). The subject was assumed to carry a 40-N weight in his hands 400 mm in front of the center of the disc. The necessary muscle force to maintain static equilibrium was computed by assuming the pivot point to be either at the disc center or at the center of the spinal canal. Figure 16.4 shows a drawing of the lifting activity and the equivalent loading condition on the motion segment. The values used for the weight of the arms, torso, and head were the same as those provided by Schultz and Andersson (1981). From this static equilibrium load condition, other load cases were developed by changing the muscle force to produce a net moment of either 15 or 30 N.m about the motion segment. A net extension moment means the application of a larger muscle force to further extend the spine. For these five cases, there were two pivot points and two conditions of the disc—normal or degenerated. In the latter case, its modulus was reduced to 30% of the original value to simulate a loss in disc height. The 20 load cases are summarized in Table 16.3.

## 7. Theoretical Results

As in any finite-element model, the output files contain a large amount of information. Selected response variables that may have a direct bearing on LBP are discussed in this chapter. They include facet load, disc bulge, nucleus pressure, and stresses in the annulus fibrosus. It was observed that these parameters increased with an increase in extensor muscle force. Furthermore, they were higher if the pivot point was assumed to be located in the center of

TABLE 16.4. Model predictions of facet load and nucleus pressure pivoting about geometrical center of disc.

Net moment (N.m)	Facet load (N)		Nucleus pressure (MPa)	
	Normal disc	Degenerated disc	Normal disc	Degenerated disc
30 Flexion	54	78	0.37	0.58
15 Flexion	84	126	0.57	0.90
0	122	174	0.78	1.22
15 Extension	157	222	0.98	1.54
30 Extension	191	271	1.18	1.86

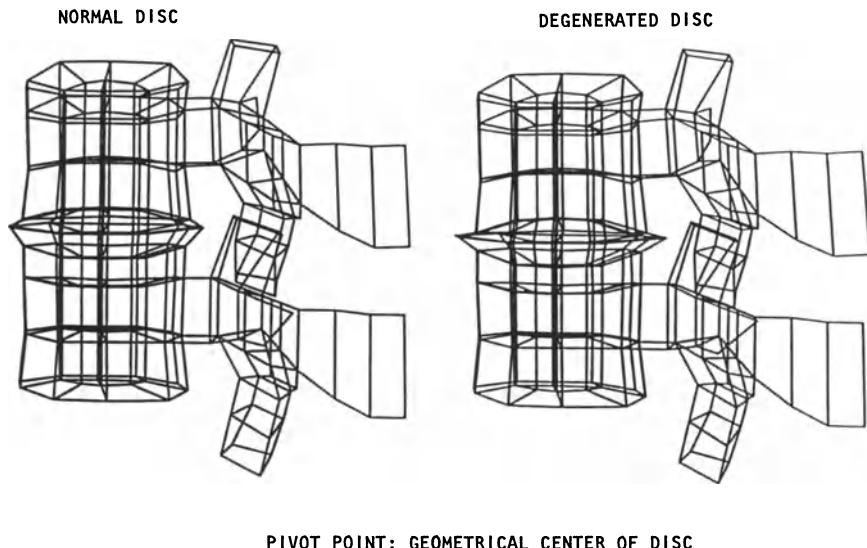
TABLE 16.5. Model predictions of facet load and nucleus pressure pivoting about center of spinal canal.

Net moment (N.m)	Facet load (N)		Nucleus pressure (MPa)	
	Normal disc	Degenerated disc	Normal disc	Degenerated disc
30 Flexion	89	127	0.58	0.91
15 Flexion	143	203	0.90	1.42
0	198	279	1.22	1.92
15 Extension	252	356	1.54	2.43
30 Extension	306	432	1.86	2.93

the spinal canal. The predicted facet loads and nucleus pressures for normal and degenerated discs are shown in Tables 16.4 and 16.5 for pivot points at the center of the disc and canal, respectively. The location of the pivot point had a profound influence on these parameters, as evidenced by a 65% increase in facet load and a 56% increase in nucleus pressure. When the extension moment or applied muscle force was increased, both parameters were higher. The equilibrium facet load was about 200 N for a normal spine, loaded by a 40-N weight at an eccentricity of 0.4 m. It increased by about 47% due to disc degeneration. The predicted pressures were within the range of pressures reported by Nachemson and Morris (1964) (0.75–1.50 MPa). In degenerated discs, the pressure increase was about 57%.

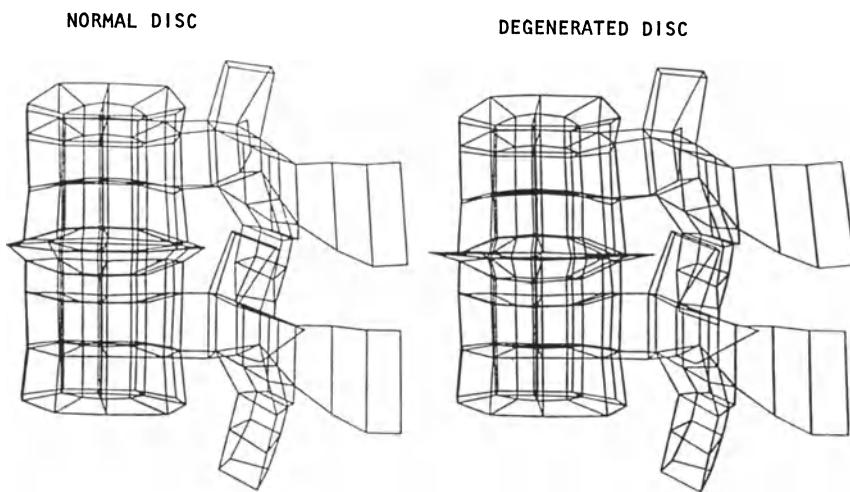
Disc bulge for normal and degenerated discs is shown in Figures 16.5 and 16.6 for the two assumed pivot points. In these lateral views, the bulge was magnified 40 times to demonstrate the substantial increase in disc bulge as a result of degeneration. It was also higher as a result of moving the pivot point rearward.

Annular stresses in a normal disc are given in Table 16.6 for the two pivot points. There was an increase of 54% due to the rearward location of the pivot point. The stresses were also higher if there was a net extension moment on the spine. The distribution of stress around the disc was a function of applied muscle loads. In the normal disc, the posterior stresses increased at a slightly higher rate than did the anterior stresses. The trend was reversed for the degenerated disc.



PIVOT POINT: GEOMETRICAL CENTER OF DISC

FIGURE 16.5. Disc bulge predicted by the model for normal and degenerated disc (bulge is magnified 40 times) pivoting at center of disc.



PIVOT POINT: CENTER OF SPINAL CANAL

FIGURE 16.6. Disc bulge predicted by the model for normal and degenerated disc (bulge is magnified 40 times) pivoting at center of spinal canal.

Muscle force also caused an increase in cortical strain around the vertebral body. The anterior strains were slightly higher than those along the posterior aspect for all loading cases. Maximum stresses occurred on the cortex at the junction of the pedicles with the vertebral body.

TABLE 16.6. Annulus stresses for normal disc at disc center and spinal canal center pivot points.

Net moment (N.m)	Center of disc			Center of spinal canal		
	Anterior	Lateral	Posterior	Anterior	Lateral	Posterior
30 Flexion	0.25 MPa	0.22 MPa	0.21 MPa	0.37 MPa	0.34 MPa	0.35 MPa
15 Flexion	0.36 MPa	0.33 MPa	0.34 MPa	0.55 MPa	0.53 MPa	0.55 MPa
0	0.48 MPa	0.46 MPa	0.47 MPa	0.74 MPa	0.74 MPa	0.76 MPa
15 Extension	0.60 MPa	0.57 MPa	0.60 MPa	0.93 MPa	0.90 MPa	0.97 MPa
30 Extension	0.72 MPa	0.69 MPa	0.73 MPa	1.11 MPa	1.09 MPa	1.17 MPa

## 8. Discussion and Conclusions

The load cases studied represented one static equilibrium case and four conditions of imbalance caused by a change in the input muscle force. Another way of looking at this is to assume a change in the load carried by the subject. A change in moment of 15 N.m is equivalent to a change in load of 37.5 N for a 400-mm moment arm. Thus, the model predicted a facet load of about 143 N with no load in the hands and with the arms extended forward slightly. It can be inferred that large facet loads are generated when one is engaged in heavy lifting activities. Conversely, constant extensor muscle tension placing the lumbar spine in lordosis can also result in facet overload and stretching of the capsular ligaments.

Disc degeneration was simulated by lowering the elastic modulus of the disc. The principal effect is that of a loss in disc height. The model predicted an increase in facet load due to this effect. The trend is reasonable, but the actual increase could be different if the loss of disc height was simulated without a change in modulus. Similarly, the other results associated with disc degeneration could also be different.

The manner in which the extensor muscles were simulated was based on the fact that the superficial muscle group spans more than one vertebra and is capable of generating effective extension moments to counteract the loads carried in front of the body. In general, the deep extensors are smaller and span a single disc space. The net extension moment and facet load due to these muscles are therefore quite small.

The top endplate was constrained to translate in the vertical direction. This condition was needed to control the degree of flexion of the motion segment and is a simulation of the performance of lifting tasks with minimal flexion of the lumbar spine. If heavy lifts with extreme flexion are involved, the rotational constraint needs to be relaxed. The facets were represented by spring elements that had identical responses in tension and compression. This is not a true simulation of the actual facet joint, which is relatively stiff in compression and weak in tension. A contact model can be introduced to improve the simulation of facet joint function. That is, a facet load would be generated during com-

pression of the joint and there would be no tensile facet load. Another improvement that can be made is the introduction of nonlinear material properties for spinal tissues, if such data are available.

The conclusions of this study are:

1. Muscular action has a profound effect on spinal compression in general and the facet joint in particular. If facet overload is a mechanical cause of LBP, the model can be used to assess the changes in facet load for various activities and levels of muscle activation.
2. An increase in extensor muscle action resulted in an increase in facet load, nucleus pressure, disc bulge, and annulus stresses.
3. Disc degeneration simulated by a loss of disc height resulted in higher facet loads, nucleus pressures, and disc bulge. Annulus stresses were lower.
4. Maximum stresses were found at the junction of the pedicles and the vertebral body, on the cortical surface.

## References

- Balasubramanian, K., Ranu, H.S., and King, A.I. 1979. *J. Biomech.* 12:813-823.
- Belytschko, T.B., Kulak, R.F., and Schultz, A.B. 1974. *J. Biomech.* 7:277-285.
- Belytschko, T.B., and Williams, J.L. 1982. Pages 35-38 in *Advance in bioengineering*, ASME, New York. L. Thibault, ed.
- Buckwalter, J.A. 1982. Pages 108-143 in *American Acad. Orth. Surg.* A.A. White and S.L. Gordon, eds. C.V. Mosby Co., St. Louis.
- Carter, D.R., and Hayes, W.C. 1977. *J. Bone Joint Surg.* 59A:954-962.
- Cyron, B.M., Hutton, W.C., and Troup, J.D.G. 1976. *J. Bone Joint Surg.* 58B:462-466.
- Ghormley, R.K. 1933. *J. Am. Med. Assoc.* 101:1773.
- Gray, H. 1973. *Anatomy of the human body*. 29th ed., C.M. Goss, ed. Lea and Febiger, Philadelphia.
- Hakim, H.S., and King, A.I. 1979. *J. Biomech.* 12:277-292.
- King, A.I. 1984. Pages 191-226 in *The biomechanics of impact trauma*. B. Aldman and A. Chapon, eds. Elsevier Science Publ., Amsterdam.
- Koogler, T.A., Swenson, L.W., Jr., and Piziali, R.L. 1979. Pages 65-68 in *Advances in bioengineering*, M.D. Wells, ed. ASME, New York.
- Kulak, R.F., Belytschko, T.B., and Schultz, A.B. 1976. *J. Biomech.* 9:377-386.
- Lewin, T. 1964. *Acta Orth Scand. Suppl.* 73:1-112.
- Lin, H.S., Liu, Y.K., Ray, G., and Nikravesh, P. 1978. *J. Biomech.* 11:1-14.
- Little, R.W., Hubbard, R.P., and Slonim, A.R. 1983. Final Report no. AFAMRL-TR-83-0005. Wright-Patterson Air Force Base, Dayton, Ohio.
- Mooney, V., and Robertson, J. 1976. *Clin. Orth. Rel. Res.* 115:149-156.
- Nachemson, A. 1981. *Spine* 6:93-97.
- Nachemson, A., and Morris, J.M. 1964. *J. Bone Joint Surg.* 46A:1077-1092.
- Nachemson, A., and Evans, J.G. 1968. *J. Biomech.* 1:211.
- Panjabi, M., Jorneus, L., and Greenstein, G. 1984. *Trans. 30th Ann. Mtg. Orth. Res. Soc.* 9:112. Atlanta, Georgia.
- Peacock, A. 1952. *J. Anatomy* 86:162-179.

- Pederson, H.S., Blunck, C.F.J., and Gardner, E. 1956. *J. Bone Joint Surg.* 38A:377–391.
- Pontius, U.R., and Liu, Y.K. 1976. Pages 21–30 in *Mathematical modeling biodynamic response to impact*. SAE Publication SP-412, Warrendale.
- Rockoff, S., Sweet, E., and Bleustein, J. 1969. *Calc. Tiss. Res.* 3:163–175.
- Rolander, S.D., 1966. *Acta Ortho Scand. Suppl.* 90.
- Schultz, A.B. 1974. *Appl. mech. rev.*: 1487–1497.
- Schultz, A.B., and Andersson, G.B.J. 1981. *Spine* 6:76–82.
- Shah, J.S., Hampson, W.G.J., and Jayson, M.I.V. 1978. *J. Bone Joint Surg.* 60B:246–251.
- Shealy, C.N. 1976. *Clin. Orth. Rel. Res.* 115:157–164.
- Shirazi-Adl, S.A., Shrivastava, S.C., and Ahmed, A.M. 1983. *1983 Biomech. Sympos.* 141–144.
- Sonoda, T. 1962. English summary. *J. Kyoto Pref. Med. Univ.* 71:659–702.
- Tennyson, S.A., and King, A.I. 1977. *Proc. First Int. Conf. Mathematical Modeling*, vol. 2:977–985. University of Missouri-Rolla.
- Tkaczuk, H. 1968. *Acta Orth Scand. Suppl.* 115.
- White, A.A. 1982. Pages 1–2 in *Sympos. on idiopathic low back pain*, A.A. White and S.L. Gordon, eds. C.V. Mosby Co., St. Louis.
- Wu, H.C., and Yao, R.F. 1976. *J. Biomech.* 9:1–7.
- Wyke, B. 1982. Pages 97–107 in *Sympos. on idiopathic low back pain*, A.A. White and S.L. Gordon, eds. C.V. Mosby Co., St. Louis.
- Yamada, H. 1970. *Strength of biological materials*, F.G. Evans, ed. Williams and Wilkins Co., Baltimore.
- Yang, K.H., and King, A.I. 1984. *Spine* 9:557–565.
- Yang, K.H., Khalil, T., Tzeng, C.R., and King, A.I. 1983. Pages 137–140 in *1983 Biomech. Sympos.*

# 17

## Biomechanics of the Human Gait

E.Y.S. CHAO

### 1. Introduction

Humans perform physical activities in a variety of ways, depending upon individual anatomical structure, physiological functions, psychomotor control patterns, and associated pathological or chronological changes. In addition, many of our physical activities are task-oriented, varying drastically due to personal habit, training, and motivation. Assuming that our limb segments are compound pendulums with multiple degrees of freedom, the infinite possibilities for achieving a given task make objective analysis and modeling of human physical activities an intractable problem in modern biomechanics. However, human locomotion represents a unique subset of such activities that can be studied with an acceptable degree of accuracy if the limb segments to be included are limited to only the lower extremities. This specific branch of biomechanics is defined as *gait analysis*—a quantitative description of human locomotion.

The historical development of gait analysis since the pre-Renaissance era has been thoroughly reviewed (Contini and Drillis, 1954; Steindler, 1953; Pauwels, 1948). The first systematic perusal of this branch of science that utilized the tools of classical mechanics was pioneered by Braune and Fisher (1895). Development of modern techniques to measure human motion has enabled gait analysis to be performed with unprecedented sophistication and objectivity. However, to a great extent, this is till a phenomenological science, restricted to description and synthesis of the motion and forces involved, with very little neuromuscular control theory to explain the cause and variability of such biological events. It has been well stated by Bernstein (1967) that the human locomotion process of a living morphological subject is of inexhaustible complexity, and one can only set a primary goal of observation and description of this process as closely as possible.

The scope of gait analysis can be divided into three broad categories:

1. kinematic study of limb and joint motion
2. foot/ground reaction force and contact pressure analysis
3. internal joint and muscle force prediction

These analyses require basic knowledge of human musculoskeletal anatomy in a quantitative sense, motion- and force-measuring instruments, three-dimensional rigid-body kinematic theory, kinetics of structures engaged in periodic motion, and the specific data-retrieving, reduction, and presentation techniques. In the entire field of biomechanics, gait analysis represents a broad segment of the discipline in which classical mechanics concepts are utilized to their fullest extent. Results generated from such analysis also serve as fundamental input data for connective tissue and bone biomechanical research. The importance of gait analysis is therefore obvious, since it is the science that investigates the effect of internal and external forces on human lower limbs in movement—a general definition of biomechanics in a broader sense.

As in most biological sciences, the potential for current gait analysis methods to generate relevant applications will not be limited by a lack of knowledge of the underlying neurophysiological theory governing this complex phenomenon. However, it is important to understand the precise limitation of gait analysis as performed by the present measurement techniques, analysis methods, and modeling simplifications, so that proper application of the results can be justified. In addition, future research and development in gait analysis can also be directed with clearer perspectives. This chapter attempts to achieve this goal, with more emphasis on the mechanistic aspects of gait analysis.

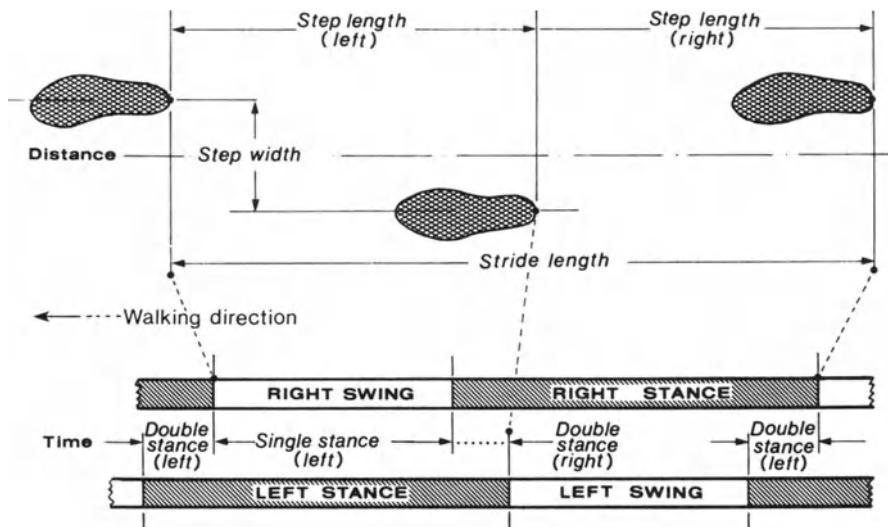


FIGURE 17.1. A schematic diagram illustrating the typical temporal distance factors in gait analysis. Not shown here is the foot angle, which is the angle between the midline of the foot or shoe and the line of progression. Note that the single stance period of one leg is the same as the swing phase of the contralateral leg.

## 2. Joint Motion and Temporal Distance Factors

Temporal distance gait factors can be measured by using foot switches and an instrumented walkway (Chao et al, 1980; Chao et al, 1983). These factors are defined based on the basis of a footprint diagram (Figure 17.1). Since gait is a periodic motion, only a typical stride is illustrated. To compensate for anthropometric variations, stride and step lengths are normalized against the lower extremity length (LEL). Cadence is defined as the number of strides,  $\text{min}^{-1}$ , and walking speed is expressed as  $\text{m min}^{-1}$ . Among normal subjects, left-right symmetry can be assumed in level walking. These parameters serve as the fundamental characteristics of human gait based on which joint motion, foot/ground reaction, and joint and muscle forces can be discussed in a more discrete manner.

In order to describe three-dimensional joint motion in the lower extremity during gait, localized coordinate axes must be defined on the basis of underlying skeletal structures and properly placed surface markers attached directly on the skin (Figure 17.2). Once the spatial locations of these markers and their

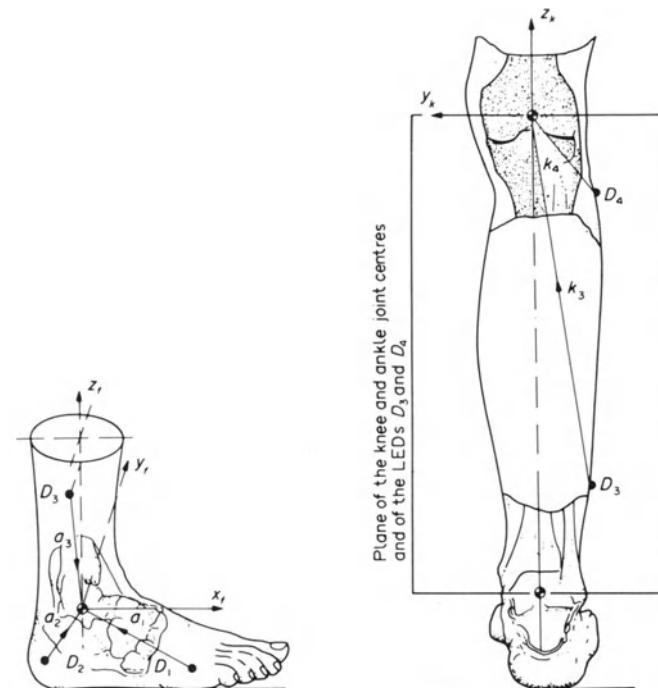


FIGURE 17.2. Definition of the reference coordinate system for the foot ( $X_f, Y_f, Z_f$ ), knee ( $X_k, Y_k, Z_k$ ), and surface markers ( $D_1, D_2, D_3, D_4$ ) used to locate these axes. Light-emitting diodes (LED) are usually used as the markers to improve point location resolution. From Röhrle, Scholten, Sigolotto, and Sollbach (1984), by permission.

fixed relationships with the bone orientation are known, the relative joint motion following the classical kinematic theory can be calculated. Assuming that the amount of translation is small, joint rotation can be defined by a set of Eulerian angles that can be determined from the following equations:

$$\begin{aligned}\theta &= \sin^{-1}(-\bar{i} \cdot \bar{K}) - \pi/2 < \theta < \pi/2 \\ \psi &= \sin^{-1}(\bar{j} \cdot \bar{K}/\cos \theta) \\ \phi &= \sin^{-1}(\bar{i} \cdot \bar{j}/\cos \theta),\end{aligned}\quad (17.1)$$

where  $\bar{i}, \bar{j}, \bar{k}$  and  $\bar{I}, \bar{J}, \bar{K}$  are the unit vectors along the coordinate axes fixed to the distal and proximal bone segments of a joint, respectively. Anatomically, these angles correspond to the clinical definition of joint rotation. However, for the purpose of identifying the instantaneous axes of rotation for these motion components, a gyroscopic system (Chao, 1980; Grood and Suntay, 1983) using two fixed axes on each bone segment and a floating axis, mutually perpendicular to the fixed axes, can be used to define the same angles (Figure 17.3). Based on such rotational axes, the sequence of motion becomes independent and can be easily measured by using an extraskeletal linkage system (triaxial goniometer). These angles (commonly defined as the Bryant or Cardan angles) are the same as the Eulerian angles determined using Equation (17.1), where the orthogonal reference axis system for each bone can be monitored during gait analysis. A typical rotational pattern for the knee joint during normal gait is illustrated, with discretized measurements in order to allow parametric analysis, in Figure 17.4. Since joint motion in gait is periodic, the analog angular patterns can be expressed in finite Fourier series to facilitate data storage, display, and analysis (Chao et al, 1983; Schneider and Chao, 1983).

The commonly used experimental methods for joint motion measurement include:

1. cinematography and stroboscopic photography
2. triaxial electrogoniometry
3. ultrasound
4. optoelectronics
5. accelerometry

A comparative description of these methods and minor variations in each version were previously described, with a discussion of the advantages and disadvantages associated with each (Chao, 1978; Woltring, 1984; Cappozzo, 1984). Besides electrogoniometry, all others require external markers that can be directly or indirectly related to the reference axes in each bone of the limb. This is a difficult procedure and does not always easily ensure reliability and accuracy. On the other hand, attachment of the goniometer system also requires alignment of the measurement potentiometer axis with the fixed and floating axes of a joint. Sometimes, when measurement axes and the joint axes are not colinear, mathematical correction is required to avoid effects of cross-talk artifacts (Chao et al, 1970).

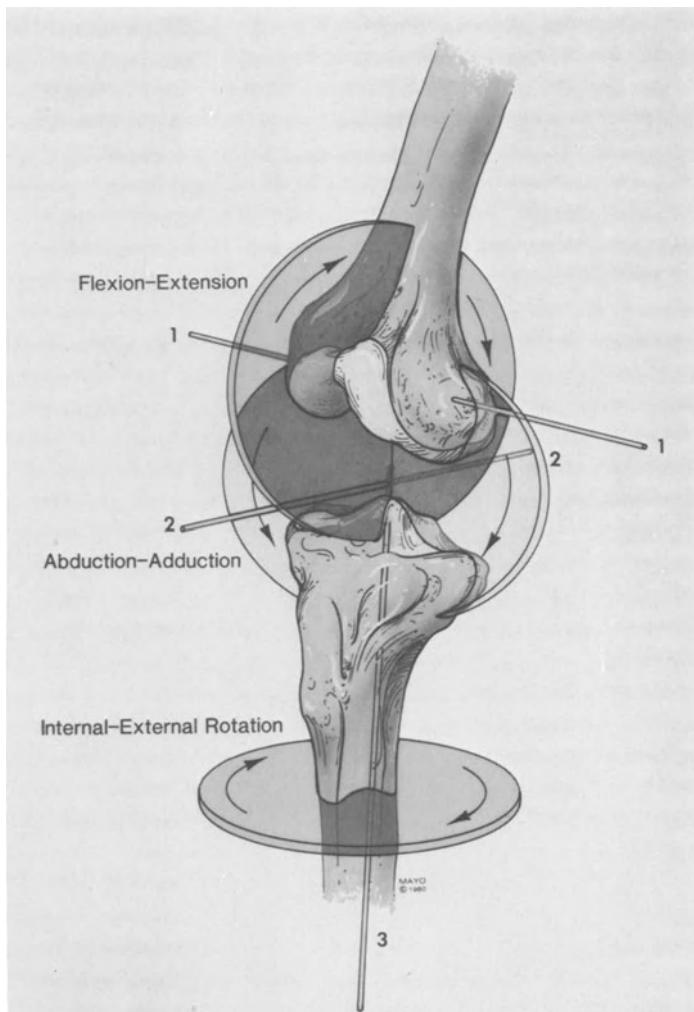


FIGURE 17.3. Axes and planes of rotation for the knee joint following the gyroscopic system. Axis 1–1 is fixed to the distal femur used to define the flexion–extension motion, and axis 3–3 is fixed to the tibia along its anatomical axis, describing the internal–external rotation. Axis 2–2 (floating axis) is orthogonal to both axes 1–1 and 3–3 and is used to measure the abduction and adduction motion. These three axes are mutually orthogonal when the knee joint is in its neutral position (zero rotation).

None of the methods is believed to have the resolution to measure joint translation. In the assessment of joint instability, static testing methods using loading instruments or x-ray assessment are required. All joint motions recorded in gait are classified as gross motion and do not reflect the true articulating joint surface motion. In analyzing natural or prosthetic joint

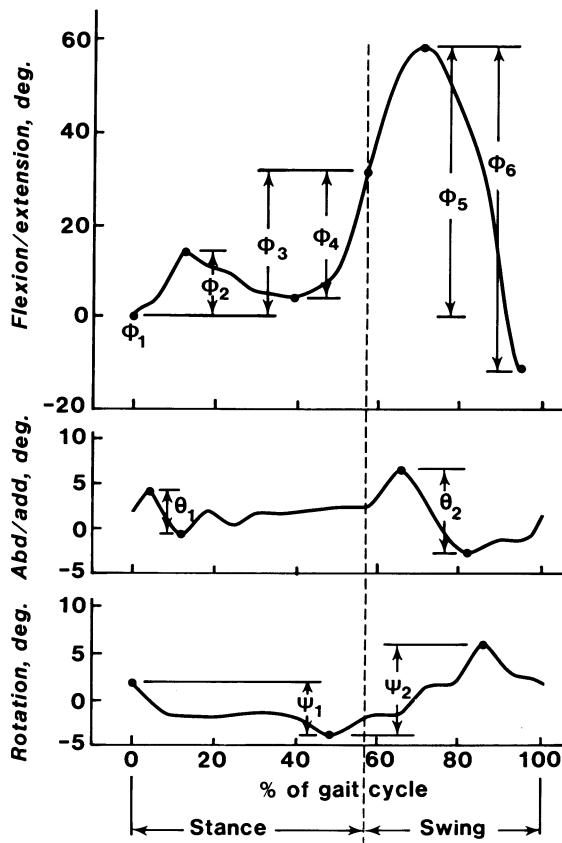


FIGURE 17.4. Typical knee joint motion pattern for a complete gait cycle. Key parameters are identified to facilitate statistical analysis. These graphs were reproduced from fitting the analog raw data by finite Fourier series.

performance, gross angular measurements must be transferred to the joint surface motion (Kinzel et al, 1972; Wismans et al, 1980).

Finally, a generalized three-dimensional joint motion model including six degrees of freedom can be formulated using the screw displacement concept based on the same limb-segment marker locations in space. The formulation can be derived based on the unique Rodrigues' vector  $\bar{\rho}$  (Rodrigues, 1840; Chao and An, 1982):

$$\bar{\rho} = \tan(\alpha/2)\bar{\lambda} \quad 0 \leq \alpha \leq \pi, \quad (17.2)$$

where  $\alpha$  is the screw rotation, and  $\bar{\lambda}$  is a unit vector along  $\bar{\rho}$ .

$$\alpha = \cos^{-1}\{(\text{tr}[R] - 1)/2\}, \quad (17.3)$$

where  $[R]$  is a  $3 \times 3$  rotational matrix expressed in terms of the directional cosines of the coordinate axes between the proximal and distal reference

system of a joint. The Eulerian angles expressing the joint rotation can also be expressed in terms of the Rodrigues vector components

$$\theta = \sin^{-1} \{2(\rho_z \rho_x + \rho_y)/Q\} - \pi/2 < \theta < \pi/2 \quad (17.4)$$

$$\phi = \sin^{-1} \{-2(\rho_y \rho_z - \rho_x)/Q \cos \theta\} - \pi/2 \leq \phi \leq \pi/2 \quad (17.5)$$

$$\psi = \sin^{-1} \{-2(\rho_x \rho_y - \rho_z)/Q \cos \theta\} - \pi/2 \leq \psi \leq \pi/2, \quad (17.6)$$

where

$$Q = 1 + \rho_x^2 + \rho_y^2 + \rho_z^2.$$

When any other joint angle of rotation exceeds its range, as defined by the above equations, a "gimbal-lock" situation will occur that will require reassignment of motion axes or use of a special technique to resolve the singularity problem. Generally, human gait analysis will not face such a problem because the lower extremity joints rarely exceed the motion limitation assigned to the rotational axis system defined.

### 3. Foot/Ground Reaction Force and Pressure

Reactive forces between the foot and the ground are measured by a piezoelectric force plate capable of recording the time history of the vertical, fore-aft, and mediolateral components throughout the stance phase of gait. These forces are normalized against the body weight of the test subject and the gait cycle time in order to minimize anthropometric variation. Discrete measurements along the analog patterns are identified to facilitate parametric analysis (Figure 17.5). In addition, the resultant force application center (centroid of the foot/ground contact areas) is also determined in reference to the force-plate coordinate system. A twisting torque between the foot and ground can also be quantitated, but the value is less reliable because of the inherent error in the instrument measurement and data reduction process.

Foot/ground reaction-force patterns can be expressed in Fourier series to facilitate phasic analysis, as well as to allow easy data storage (Schneider and Chao, 1983). The maximum number of harmonics required to represent the reaction force pattern is 12, but for mathematical modeling purposes, the first 4 harmonics are sufficient to represent the analog waveforms. Expressing the joint motion and ground reaction-force patterns in Fourier series can also simplify the averaging process when large numbers of sample curves with varying periods and amplitudes are involved.

The foot/ground reaction-force components can also be displayed in a phase diagram to analyze the gross and minute variations in normal and pathological subjects (Yoshida, 1982). This can be accomplished by plotting all three components of the reaction force as a vector in three-dimensional space, using the force plate as a reference. A two-dimensional vector diagram using only the vertical and the fore-aft reaction force components is more commonly used, since the magnitude of the mediolateral reaction force is generally

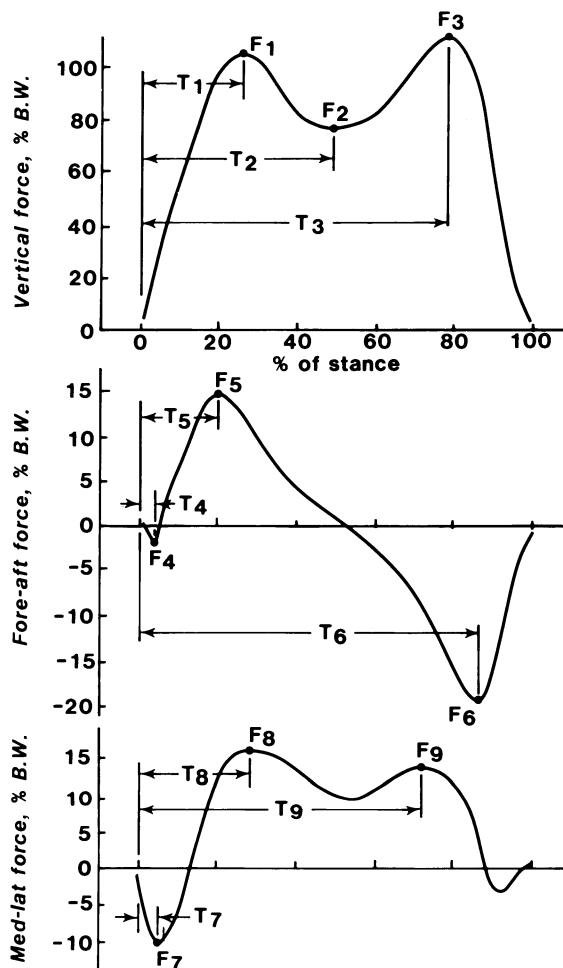
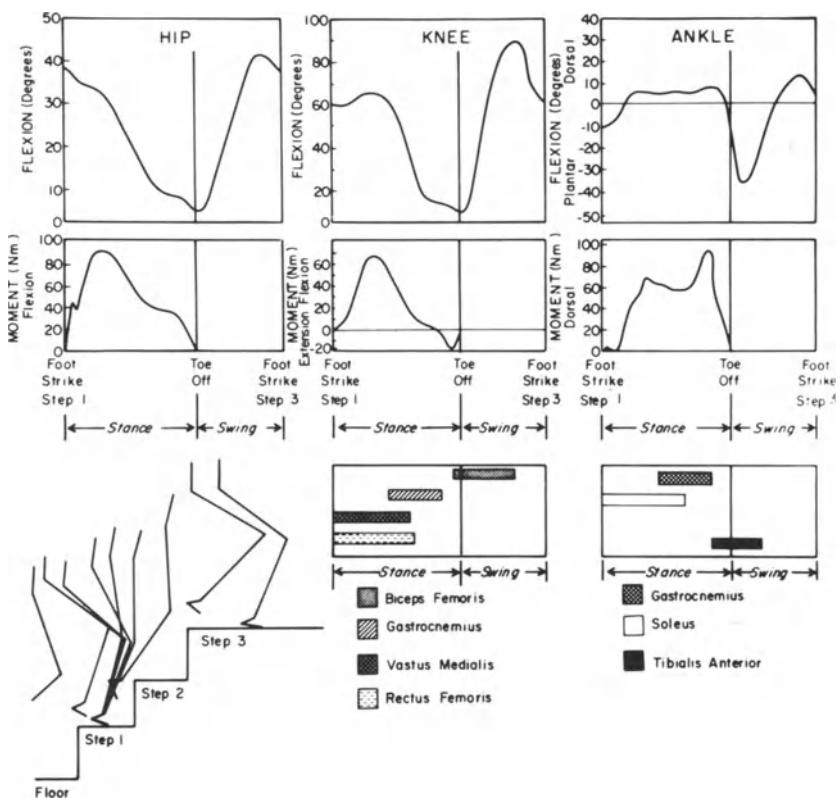


FIGURE 17.5. Typical ground reaction force diagram of a normal subject during the stance phase of level walking. Key force magnitudes ( $F$ ) and time instances ( $T$ ) are identified for the purpose of performing parametric analysis. These graphs were reconstructed by fitting the raw data with finite Fourier series.

small. However, such a data display does not take into account the lower extremity joint position during the stance phase of gait. Consequently, the external moment acting at the joint generated by the foot/ground reaction-force vector cannot be assessed.

For this reason, the foot/ground reaction-force vector has been extended to determine its moment with respect to the joint center of motion for the purpose of studying leg stability during level walking and stair-climbing activities (Andriacchi et al, 1980). The joint center of motion in the lower extremity during gait is monitored on the basis of surface markers attached to the limb



**FIGURE 17.6.** Sagittal-plane flexion-extension movement of the hip and knee and plantar flexion-dorsiflexion movement of the ankle; moments about these joints; and phasic activities of the knee and ankle muscles in one limb of a subject ascending from Step 1 to Step 3. From Andriacchi, Andersson, Fermier, Stern, and Galante (1980), by permission.

segments and using the force-plate coordinate axes as references. The mass center of each limb segment and its mass moment of inertia are also determined so that they can be included in the calculation of the net external moment generated by the foot/ground reaction-force vector. Knowing the joint motion and limb position during gait, one can determine the external moment diagram during stance phase in level walking and stair-climbing (Figure 17.6). Such analysis treats the joint muscle and constraint forces as internal forces, thus avoiding the difficulties of solving the indeterminate problem involving too many unknown variables.

The foot/ground reaction force is a result of the actual contact pressure in different regions of the foot at any time. Therefore, it does not provide the information concerning the area of contact, nor does it reflect the intensity of contact pressure in localized regions of the foot. Various techniques can be used to determine the foot/ground contact areas as well as pressure intensity

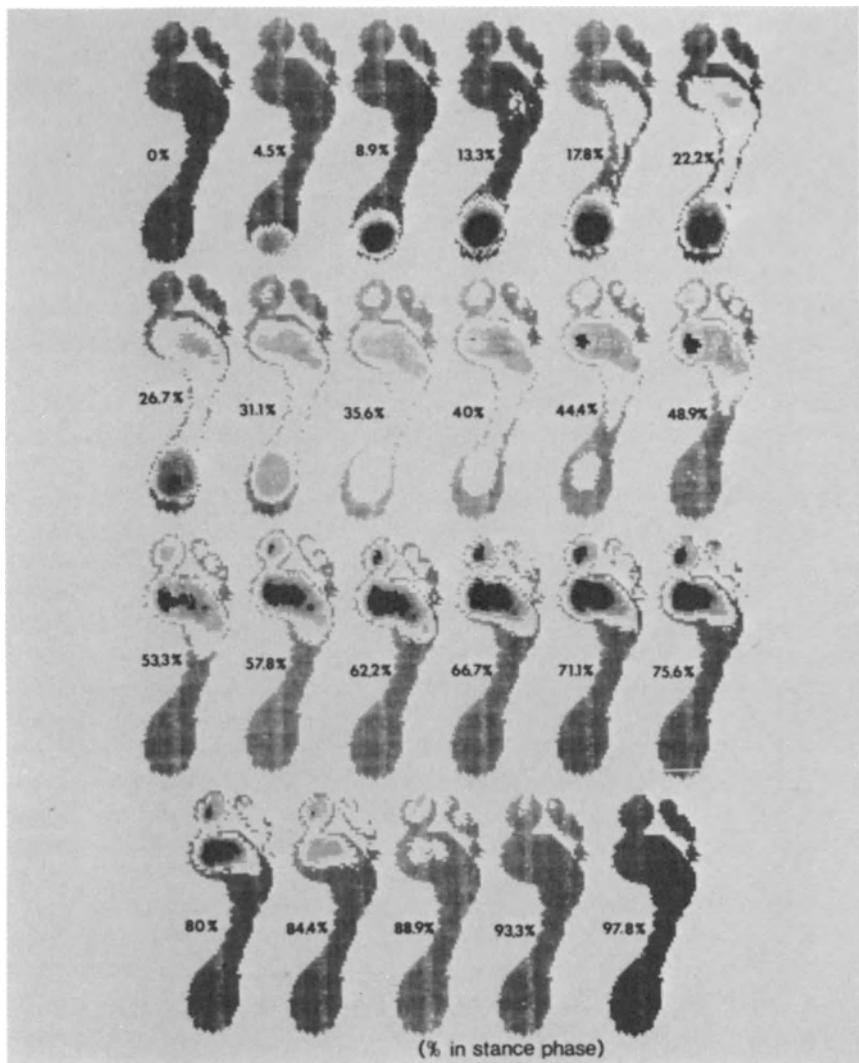


FIGURE 17.7. Foot/ground contact area and pressure distribution during gait: (a) Chronological sequence of foot/ground contact pattern throughout entire stance phase of gait cycle. The entire outline of the foot is presented to allow identification of pressure zone and intensity; (b) History of foot/ground contact area and pressure distribution according to prescaled grades. Eight zones were used to identify the anatomical locations of the pressure contact area, and these zones can be reduced to six for simplification purposes.

during the stance phase of gait. Among these, a light-refraction method (Betts et al, 1980) was developed that uses a transparent force plate with a floodlight projecting through the entire plate and a video camera underneath the plate to record the reflection pattern of the foot when it contacts the plate. In order to

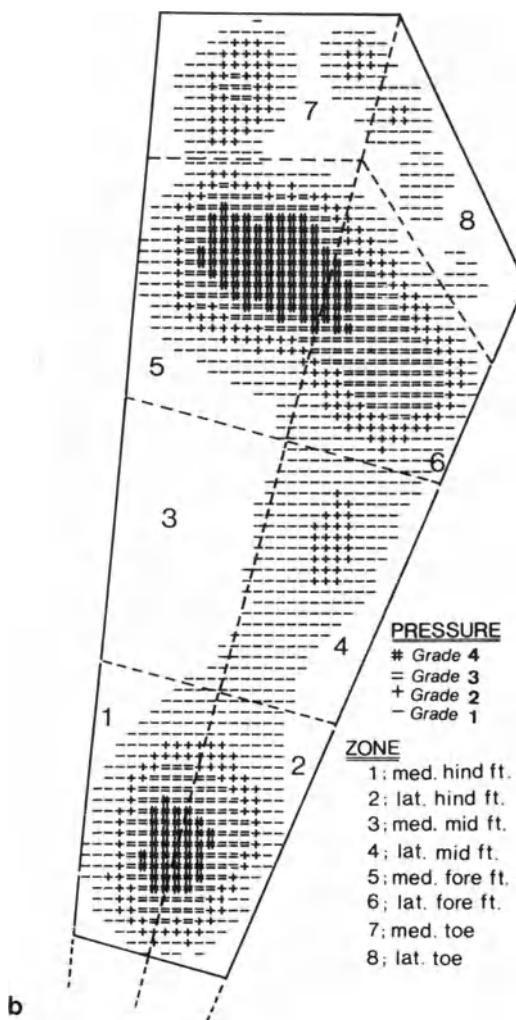


FIGURE 17.7. (continued)

increase the resolution of the reflection pattern in different gray levels, a silicone rubber mat is used between the foot and the force plate. Gray level quantification is grouped into five grades to correlate with pressure intensity. A scale factor is developed that is based on the resultant force magnitude (vertical component) and the contact area (rated according to a geometric ratio of different grades of light intensity) so that the absolute pressure in each contact area can be quantitated in  $\text{N}/\text{cm}^2$ .

The foot is divided into eight regions by a special graphics package, so that the pressure distribution in each region can be quantitated throughout the stance phase of gait (Figure 17.7). This method provides direct assessment of

the pressure distribution on the sole of the foot during gait. Such information cannot be generated by using the foot/ground reaction-force vector alone. The addition of this evaluation technique has increased the proficiency and scope of gait analysis associated with foot and shoe function.

#### 4. Joint and Muscle Force Analysis

When the musculoskeletal joint system in the lower extremity is engaged in gait function, three types of internal forces are involved: muscle or tendon forces, joint surface contact or shear forces, and the capsuloligamentous constraint forces. In addition, the external forces due to gravity and inertia are also important in the analysis. For the convenience of modeling and computation, joint contact and ligamentous forces are combined to form the joint constraint forces and moments. Thus, the governing equations of motion can be derived from either the Newtonian or Lagrangian formulations. In Newtonian formulation, free-body diagrams of the leg segments are obtained, with muscle forces expressed either explicitly or implicitly. The equation of motion can be written in the following generalized form:

$$\sum_{i=1}^n \bar{F}_i = m\ddot{\bar{r}}_c \quad (17.7)$$

$$\sum_{i=1}^n \bar{M}_{ic} = \dot{\bar{H}}_c, \quad (17.8)$$

where

$\bar{F}_i$  = force vectors acting on the free body,

$m$  = mass of the segment,

$\bar{r}_c$  = position vector of the segment mass center,

$\bar{M}_{ic}$  = moment for  $\bar{F}_i$  about the mass center,

$\bar{H}_c$  = angular momentum about mass center, and

$$\bar{H}_c = \{I\} \cdot \bar{\omega}, \quad (17.9)$$

where

$\{I\}$  = inertia tensor of the limb segment, and

$\bar{\omega}$  = instantaneous angular velocity vector.

In Langrangian formulation, a set of generalized coordinates,  $q_1, q_2, \dots, q_n$  representing joint relative angular position are required for the musculoskeletal system to be analyzed without constructing the free-body diagram. Consequently, all internal forces occurring at the joint and muscle do not appear in the equations of motion explicitly. Instead, a set of generalized forces ( $\Phi_i$ ) is required that represents the moments at the joints caused by muscle contractions. The equilibrium equations have the form,

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = \Phi_i \quad i = 1, 2, \dots, n, \quad (17.10)$$

where

$L$  = Langrangian, which is the difference between the kinetic and potential energies of the system.

With proper initial conditions for the generalized coordinates during each cycle of gait, the above nonlinear equations can be solved numerically. However, the incomplete and imprecise kinematic data of limb movement during gait and the extremely complex anatomical geometry and inertial properties of the musculoskeletal systems make the force analysis in gait one of the most challenging tasks in biomechanics. In order to obtain practical results, assumptions and simplifications in modeling and computation are inevitable.

In addition to the aforementioned difficulties, the force analysis problem must be solved in a different manner, since the internal forces or the generalized forces (moment at the joint produced by muscle contraction) are not known and must be solved in an inverse manner by measuring the kinematic data of the limb segment during gait. This problem was defined as the “Inverse Dynamic Problem” (Chao and Rim, 1973) that constitutes unique characteristics in biomechanics involving musculoskeletal joint dynamics. There are two methods of solving the inverse dynamic problems.

The first method is to experimentally measure the displacement data of limb segments and numerically differentiate them to obtain the corresponding velocity and acceleration data involving limb segments in motion. By substituting such kinematic data in the original equations of motion, a set of algebraic equations containing only the unknown muscle and joint constraint forces can be formed. Diagrammatically, this method can be illustrated, as in Figure 17.8. This has been a standard method in gait analysis when muscle and joint forces are to be determined (Paul, 1964; Morrison, 1968; Seireg and Arvikar, 1975; Crowninshield and Brand, 1981; Röhrlé et al, 1984). However, numerical differentiation tends to magnify inherent experimental error in displacement data, which causes great concern regarding the reliability of the final results obtained.

For this reason, the second method uses an iterative scheme to determine the joint moments that will minimize the total energy in gait or attempt to match the measured and calculated displacement data in gait analysis (Beckett and Chang, 1968; Chow and Jacobsen, 1971; Chao and Rim, 1973). Although this method cannot be used to determine the specific values of muscle and joint constraint forces, useful information in terms of joint moment appears to be valuable in assessing functional effects of therapeutic and rehabilitation procedures (Andriacchi et al, 1980). In addition, the formulation will make the equilibrium equations determinate, since muscle redundancy becomes intrinsic.

In Newtonian formulation, unknown muscle and joint constraint forces

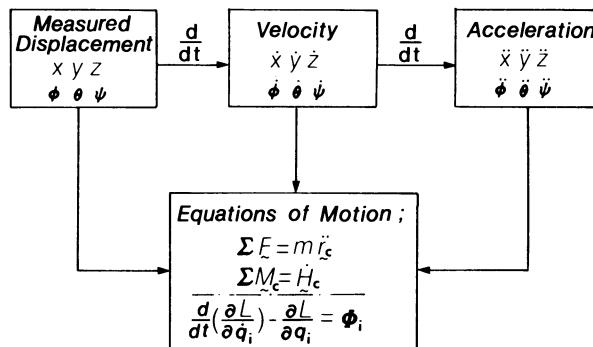


FIGURE 17.8. Diagram illustrating the solution process for the Inverse Dynamic Problem in gait analysis. Displacement information must be differentiated twice, and both the Newtonian and Lagrangian formulations can be used to derive the equations of motion describing gait.

exceed the number of governing equations of motion and constraint relationships, thus making the problem “indeterminate.” Multiple antagonistic muscles are found to be active for the purpose to maintain joint stability and control reflexes that make the system redundant and difficult to solve uniquely. An optimization algorithm is usually used to resolve this problem (Seireg and Arvikar, 1975; Crowninshield, 1978; Pedotti et al, 1978; Hardt, 1978; Röhrle et al, 1984). The general formulation of the optimization method to solve the indeterminate problem in gait analysis can be summarized:

$$\text{Minimize } J = f(X_1, X_2, \dots, X_n) \quad (17.11)$$

$$\text{Subject to } g_j(X_1, Y_2, \dots, X_n) = 0, \quad j = 1, 2, \dots, m \quad (17.12)$$

$$\text{and } b_i \leq X_i \leq a_i, (b_i \geq 0), \quad i = 1, 2, \dots, n \quad (17.13)$$

where  $J$  is an optimal criterion (cost function) that can be linear or nonlinear. The function  $g$  represents the equations of motion and other equality constraint relationships based on anatomy and on neuromuscular control theory governing the characteristics of gait. The  $X_i$  stands for the independent variables, which are the unknown joint and muscle forces. These variables are also subject to inequality constraint, which could be converted to equality constraints. Solving the linear or nonlinear optimization problems presents no great difficulties, as many programs are available to provide fast convergence to the local minimum in the solution domain. However, selection of proper optimal criteria lacks physiological justification. Hence, the results obtained can be regarded only as hypothetical at best.

The determination of joint and muscle forces in gait analysis is an unsolved problem in modern biomechanics. Aside from the difficulties in modeling, formulation, and solution, determinations of limb center of mass and inertial properties add more complexities and uncertainty to the analysis. Above all,

the computational process is usually long and tedious, requiring modern computer facilities. In practical applications of gait analysis results, these factors must be carefully considered.

## 5. Basic Science and Clinical Applications

Many fundamental research studies concerning the physiology and biomechanics of the musculoskeletal joint system must rely upon kinematic and kinetic data obtained during limb segment functions. In lower extremities, gait analysis becomes the logical choice, since it represents the essential function to serve as a common focus for many other related investigations. Because gait is also the standard function to evaluate the normality of leg joints and the foot, it has long been utilized as a tool in clinical examination.

### 5.1. Joint and Limb Motion During Gait

The effect of walking velocity on joint motion variation has been studied to gain further understanding of neuromuscular control principles and the natural compensatory mechanism in the lower extremities (Pedotti et al, 1978; Andriacchi et al, 1977; Sutherland et al, 1980; Schneider et al, 1983). Since joint motion and temporal distance factors in gait are relatively easy to measure with a short turn-around time, such information has been used to assess functional improvement after therapeutic and rehabilitative treatments (Chao et al, 1980; Laughman et al, 1980). With the aid of a special biofeedback instrument, lower extremity amputee gait training can be automated using joint motion and temporal distance measurements as control signals (Laughman et al, 1982). Normative data involving gait can be categorized according to sex and age to serve as the reference base for comparative studies (Chao et al, 1983).

To improve the efficiency and reliability of the gait data for various applications, important parameters are identified using well-established statistical methods. The repeatability of gait data can be tested using the variance ratio so that the most representative pattern can be selected for the final analysis (Hershler and Milner, 1978). Discriminate analysis and principal component analysis are used to identify the key parameters to assist in differentiating normal and abnormal motion results. To facilitate final result presentation, performance and symmetry indices are developed to provide a grading scale based on weighted combinations of the key gait parameters (Chao et al, 1980).

### 5.2. Foot/Ground Reaction Force and Pressure Distribution

Gait data produced by the force plate appear to be most reliable because of the high resolution of the instrument used and the low error margin involved. However, since ground reaction force and foot pressure distribution represent

focal information, they cannot be used to accurately assess remote joint and limb function during gait. Therefore, their application would be more suitable for functional assessment of the foot and ankle afflicted by local pathology, treatment, and footwear (Katoh et al, 1983; Yoshida, 1982; Yoshida et al, 1983; Merkel et al, 1983; Fogel et al, 1982). When the foot/ground reaction-force vector is used to determine the external moment applied to proximal joints of the leg, reliability of the results generated must be carefully examined, as the calculation will involve limb kinematic, inertia, and pressure location data that tend to be less accurate. However, the ground reaction-force vector has been demonstrated as a potential tool for relevant clinical application (Andriacchi et al, 1982; Simon et al, 1983).

Although foot/ground pressure measurement has been developed for clinical application, the extent of utilization is rather limited as a result of the inefficient data reduction and analysis (Cavanagh and Michiyoshi, 1980; Duckworth et al, 1982; Lord, 1981; Nicol and Henning, 1978; Arcan and Brull, 1976). In addition, calibration of absolute pressure magnitude in the reflecting-light method still lacks reproducibility. These technical problems are certainly solvable and when resolved, foot/ground pressure measurement can be quite reliable and useful to study foot-related problems in medicine, sports, and shoe design.

### 5.3. Muscle and Joint Force in Lower Extremity During Gait

Forces within the joints serve as fundamental input data to examine the biomechanical performance of articulating cartilage, bone remodeling under load, and ligamentous stability. In the design and application of prosthetic implants for fracture treatment or joint and ligament replacement, it is essential to know what magnitudes of internal force are being imposed upon these artificial devices in order to prevent failure through fracture, loosening, and nonphysiological tissue remodeling. In certain abnormal cases, an excessive amount of force can occur at the joint that may lead to premature failure of the related biological tissues or prosthetic devices; this must be detected and prevented. Consequently, quantitative analysis of joint and muscle forces during gait must be performed.

Determination of internal muscle and joint forces is an extremely complex analytical and experimental process. When results are obtained, no method is available to validate the data. It is therefore desirable to simplify the method for internal force determination.

Judging from the joint force pattern obtained on normal subjects, a striking similarity between the joint force pattern and the foot/ground reaction forces is observed (Figure 17.9). This similarity could be utilized to estimate joint forces when the individual patient's ground reaction-force pattern is available. When muscle force data are required, quantitative EMG measurement can also be used to provide indirect assessment of the monitored muscles on a relative basis (An et al, 1983; Schultz et al, 1983). This philosophy is

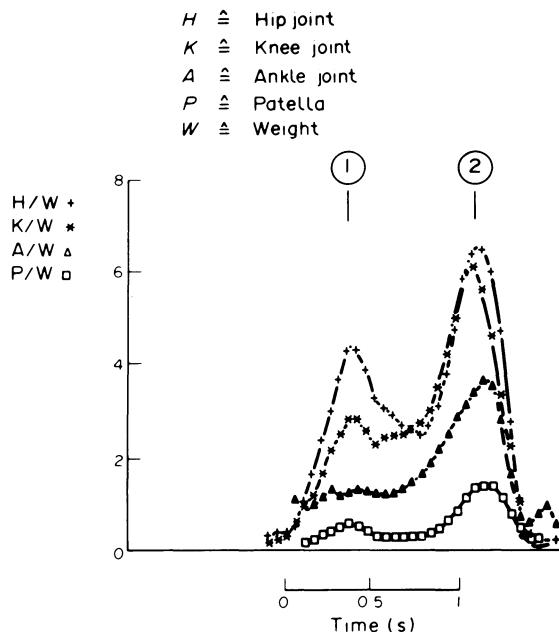


FIGURE 17.9. Typical plot of joint resultant forces of the lower extremities expressed in terms of body weight ( $W$ ) during stance phase of gait at walking velocity of  $1.3 \text{ m/s}^{-1}$ . The time-dependent patterns of joint forces are similar to those of the foot/ground reaction forces. The “1” and “2” numbers mark the locations of maximum resultant forces that will vary according to walking velocity.

important to rationalize the value of routine force evaluation during gait when the application is limited to only a single case subjected to sequential assessments.

## 6. Future Perspectives

There seems to be a large gap between practical application of gait analysis and its basic research development. The main reason for this deficiency may have been a lack of communication among different disciplines. The absence of mutually acceptable definitions of joint motion and internal forces could have been a key contributing factor. A set of understandable and correct terminologies related to the mechanistic and physiological aspects of lower limb motion in gait would be extremely helpful.

The techniques of human evaluation in gait have reached a degree of sophistication at which improvement will always be limited by the inherent difficulties associated with the complex anatomical systems being analyzed. Future efforts in this respect must be devoted to the development of effective

data analysis schemes that take into account the variation and degree of reliability of the data generated, so that relevant applications can be established in medicine and related fields. It is important to recognize the limitations in interpreting physical measurements of biological systems. Objective gait analysis should always be regarded as a supplemental examination that may be quite useful to reconfirm clinical diagnosis and assist treatment decisions.

In the determination of muscle and joint forces during gait, the experimental and computational procedures are too complex and costly for routine application in the medical field. Effort must be devoted in the future to simplify the method of analysis so that useful information in legible form can be communicated to the physicians. From the research point of view, theoretically predicted joint forces must be validated experimentally using instrumented implants (Rydell, 1966; Kilvington and Goodman, 1981) or performing animal experiments where *in-vivo* measurements can be obtained. Development of normative models for joint and muscle force calculation is also needed so that various analytical methods can be evaluated for benchmark comparison. More effort is needed to develop and incorporate neuromuscular control theory into the gait analysis model to more rationally formulate the redundancy problem.

The future for gait analysis application in various clinical fields is extremely promising, though it faces challenges similar to those encountered 15 years ago by computerized electrocardiogram (ECG) analysis. No one questions the value of objective ECG analysis today, as final diagnosis and treatment decisions still depend on clinical judgment and medical examination. Gait analysis should be viewed in the same light. It is futile to argue the relevancy of gait analysis; more time and effort should be spent to identify its proper role in clinical application. Without such effort, its real value may never be realized.

## References

- An, K.N., Cooney, W.P., Chao, E.Y., Askew, L.J., and Daube, J.R. 1983. *J. Appl. Physiol.: Respirat. Environ. Exercise Physiol.* 54(3):714–719.
- Andriacchi, T.P., Ogle, J.A., and Galante, J.O. 1977. *J. Biomech.* 10:261–268.
- Andriacchi, T.P., Andersson, G.B.J., Fermier, R.W., Stern, D., and Galante, J.O. 1980. *J. Bone Joint Surg.* 62[A]:749–757.
- Andriacchi, T.P., Galante, J.O., and Fermier, R.W. 1982. *J. Bone Joint Surg.* 64[A]:1328–1335.
- Arcan, M., and Brull, M.A. 1976. *J. Biomech.* 9:453–457.
- Beckett, R.E., and Chang, K. 1968. *J. Biomech* 1:147–159.
- Bernstein, N.A. 1967. *The coordination and regulation of movements*. Pergamon Press, Ltd., Oxford, England, U.K.
- Betts, R.P., Duckworth, T., and Austin, I.G. 1980. *J. Med. Eng. & Tech.* 4(3), 136–142.
- Braune, C.W., and Fisher, O. 1895. *Internationale abhandlungen der mathematisch physisch königlich en sächsischen gesellschaft für wissenschaften* 21:151–324.
- Cappozzo, A. 1984. Pages 27–50 in *Human movement science*. Vol. 3, H.T.A. Whiting, ed. Elsevier Science Publishers B.V., North-Holland.

- Cavanagh, P., and Michiyoshi, A.E. 1980. *J. Biomech.* 13:69–75.
- Chao, E.Y. 1978. Pages 385–411 in *CRC handbook of bioengineering in medicine and biology. Experimental methods for biomechanical measurement of joint kinematics*. Vol. 1. CRC Press, Inc., Palm Beach, Florida.
- \_\_\_\_\_. 1980. *J. Biomech.* 13:989–1006.
- Chao, E.Y., Rim, K., Smidt, G.L., and Johnston, R.C. 1979. *J. Biomech.* 3:459–471.
- Chao, E.Y., and Rim, K. 1973. *J. Biomech.* 6:497–510.
- Chao, E.Y., Laughman, R.K., and Stauffer, R.N. 1980. *Arch. Orth. Traum. Surg.* 97:309–317.
- Chao, E.Y., and An, K.N. 1982. In *Biomechanics: principles and applications*, R. Huiskes, D. Van Campen, and J. De Wijn, eds. Martinus Nijhoff, The Hague, Netherlands.
- Chao, E.Y., Laughman, R.K., Schneider, E., and Stauffer, R.N. 1983. *J. Biomech.* 16:219–233.
- Chow, C.K., and Jacobsen, D.H. 1971. *J. Math. Biosc.* 10:239–306.
- Contini, R., and Drillis, R. 1954. *Appl. Mech. Rev.* 7:49–52.
- Crowninshield, R.D. 1978. *ASME J. Biomech. Eng.* 100:88–92.
- Crowninshield, R.D., and Brand, R.A. 1981. *J. Biomech.* 14:793–801.
- Duckworth, T., Betts, R.P., Franks, C.I., and Burke, J. 1982. *Foot & Ankle* 3(3):130–141.
- Fogel, G.R., Katoh, Y., Rand, J.A., and Chao, E.Y. 1982. *Foot & Ankle* 3(2):105–113.
- Grood, E.S., and Suntay, W.J. 1983. *J. Biomech. Eng.* 105:136–144.
- Hardt, D.E. 1978. *J. Biomech. Eng.* 100:72–78.
- Hershler, C., and Milner, M. 1978. *IEEE Trans. Biomed. Eng.* BME-25:413.
- Katoh, Y., Chao, E.Y., Laughman, R.K., Schneider, E., and Morrey, B.F. 1983. *Clin. Orth.* 177:23–33.
- Kilvington, M., and Goodman, R.M.F. 1981. *Eng. in Med.* 10:175–187.
- Kinzel, G.L., Hillberry, B.M., Hall, A.S., Jr., Sickle, V., and Harvey, W.M. 1972. *J. Biomech.* 5:283–193.
- Laughman, R.K., Carr, T.A., Chao, E.Y., and Youdas, J.W. 1980. *Am. J. Sports Med.* 8(6):425–431.
- Laughman, R.K., Schneider, E., Bogard, S., and Chao, E.Y. 1982. In *IFAC control aspects of prosthetics and orthotics*, R.M. Cambell, ed. Pergamon Press, New York.
- Lord, M. 1981. *J. Biomed. Eng.* 3:91–99.
- Merkel, K.D., Katoh, Y., Johnson, E.W., Jr., and Chao, E.Y. 1983. *Foot & Ankle* 3(4):189–196.
- Morrison, J.B. 1968. *Biomed. Eng.* 1:164–170.
- Nicol, K., and Hennig, E.M. 1978. Pages 374–380 in *Measurement of pressure distribution by means of flexible, large-surface mat*. VI-A University Park Press, Baltimore.
- Paul, J.P. 1964. In *Biomechanics and related bio-engineering topics*, R.M. Kenedi, ed. Pergamon Press, Oxford, England, U.K.
- Pauwels, F. 1948. *Z. Anat. Entwgesch.* 114:1–2.
- Pedotti, A., Krishman, V.V., and Stark, L. 1978. *Math. Biosci.* 38:57–76.
- Rodrigues, O. 1840. *J. de Mathematiques Pures et Appliquees, 1st Series* 5:380–440.
- Röhrle, H., Scholten, R., Sigolotto, C., and Sollbach, W. 1984. *J. Biomech.* 17:409–424.
- Rydell, N.W. 1966. *Acta Orth. Scand. Suppl.* 88.
- Schneider, E., and Chao, E.Y. 1983. *J. Biomech.* 16:591–601.

- Schneider, E., Kozumi, K., Garcia, T.J., Laughman, R.K., Stauffer, R.N., and Chao, E.Y. 1983. *Trans. 29th Ann. Mtg.: Orth. Res. Soc.* 8:152 and *Orth. Trans.* 7:304, Summer.
- Seireg, A., and Arvikar, R.J. 1975. *J. Biomech.* 8:89–102.
- Schultz, A., Haderspeck, K., Warwick, D., and Portillo, D. 1983. *J. Orth. Res.* 1(1):7791.
- Simon, S.R., Trieshmann, H.W., Burdett, R.G., Ewald, F.C., and Sledge, C.B. 1983. *J. Bone Joint Surg.* 65[A]:605.
- Steindler, A. 1953. *J. Bone Joint Surg.* 35[A]:540–542.
- Sutherland, D.H., Olshen, R., Cooper, L., and Woo, S.L-Y. 1980. *J. Bone Joint Surg.* 62[A]:336–353.
- Wismans, J., Veldpaus, F., Janssen, J., Husan, A., and Struben, P. 1980. *J. Biomech.* 13:677–685.
- Woltring, H.J. 1984. Pages 35–63 in *Human movement science*. Vol. 1, H.T.A. Whiting, Elsevier Science Publishers B.V., North-Holland.
- Yoshida, K. 1982. *Mie Univ. Med. J.* 26(3), 296–308, December.
- Yoshida, K., Iwakura, H., and Inoue, F. 1983. *Scand. J. Rehab. Med.* 15:133–140.

# 18

## Biomechanics of Bone–Implant Interactions

R. HUISKES

### 1. Introduction

#### 1.1. Human Joint Replacement

The possibilities for artificial reconstruction of diseased human joints have improved tremendously since the early 1960s, and the numbers of joint replacements in orthopedic surgery have grown almost exponentially. For the hip and the knee in particular, a multitude of different artificial joint designs are available, usually made of special metal alloys in combination with several kinds of plastics (Walker, 1977). Roughly speaking, artificial joint components can be categorized into intramedullary fixated ones, where a stem is fixed into the medullary canal of a long bone, and surface replacements, which are more or less fixed against the (spongy) bone at the joint site (Figure 18.1).

Whereas some prostheses are wedged or screwed into the bones to obtain a good fixation, usually some sort of fixation medium is applied. Most popular is acrylic cement (polymethylmethacrylate), a filling agent introduced in a viscous phase to adapt the smooth implant to the irregular bone (Charnley, 1970). Since recently, ready-made profiled surfaces or porous coatings are utilized, too, in closely fitting prostheses that allow bone to grow in and gradually grip the implant firmly (Galante, 1983).

An artificial joint has two distinct general functions: first, to restore the natural kinematic characteristics of the joint, and second, to transfer the joint forces to the bone. The second function must be fulfilled without causing pain or failure, for the longest possible period of time, a requirement closely related to the stress patterns occurring in the different materials and at their connecting planes (interfaces) as a result of the load transfer mechanism.

The purpose of this chapter is to discuss the load transfer function, in particular in (intramedullary fixated) hip prostheses, and to investigate the relations between structural parameters and stress patterns and their effects on possible failure mechanisms.

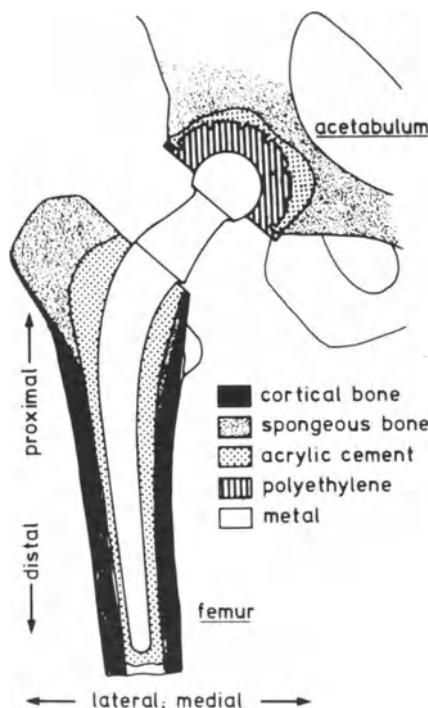


FIGURE 18.1. Schematic representation of a cemented Charnley total hip replacement. The femoral component can be categorized as an intramedullary prosthesis, the acetabular cup as a surface prosthesis (reproduced from Huiskes, 1979).

## 1.2. Prosthetic Loosening

Several kinds of biological and mechanical phenomena may enhance failure of a bone–prosthesis composite; these include infections, component fractures, excessive wear, massive bone resorption, and aseptic (noninfectious) loosening. The latter is considered to be the most important complication, limiting the life span of replacements in general.

The key region for implant loosening is the implant–bone interface. Since implant materials do not adhere to bone in a chemical sense, the interface connection relies on mechanical interlock (e.g., acrylic cement in the trabecular voids of bone, or bone growth into porous layers). We define *interface loosening* of two connecting surfaces as the loss of capacity, locally, to transfer tensile stresses and shear stresses (other than friction). In this sense, certain parts of implant–bone interfaces will be loose from the beginning, if only at a microlevel. It is quite possible that local interface loosening (failure initiation) propagates to an extent that the whole implant–bone connection becomes loose in the mechanical sense, not long after the operation. However, that does not necessarily mean that the reconstruction has failed.

*Clinical loosening*, on the other hand, is associated with pain, motion, and/or subsidence of the prosthesis relative to bone, and will generally imply removal and revision of the replacement. *Roentgenological loosening*, another term popular in orthopedics, implies that, although the patient may be without complaints, a relatively thick and/or extensive gap between implant and bone becomes visible on x rays, often combined with subsidence of the replacement (e.g., Stauffer, 1982). It has been shown in animal experiments and evaluations of material removed from patients that the gap or “radiolucent line” visible on x rays is filled with (soft) fibrous tissue. This layer may be of an aggressive nature, in the sense that it contains a relatively large amount of cells with a known capacity to resorb bone (Goldring et al, 1983).

The effects that potentially play a role in the development of interface loosening, radiological loosening, and subsequent clinical loosening have been subject to much research and speculation, in particular with respect to cemented prostheses. Examples are bone reactions to chemical components in acrylic cement and the heat of polymerizing (e.g., Feith, 1975; Huiskes, 1979), effects of a disturbed vascularity (Rhinelander et al, 1979), and the migration of wear particles along the cement–bone interface (Willert and Semlitsch, 1976).

Normal bone is continuously remodeling. Hence, the implant–bone interface is not a static physical entity, but changes its characteristics over time. This process may be influenced by interface stress or strain patterns, in the same way that normal, intact bone is able to adapt its structure and shape to an altered loading environment (e.g., Cowin et al, 1984).

These biological and mechanical stimuli may act to reduce the strength of the implant–bone connection in cemented and cementless prostheses alike, if only locally. High interface stresses will then provoke interface loosening. Once local failure is initiated, failure propagation occurs, introducing (micro) motions of the implant relative to bone. It is quite probable that these (micro) motions provoke bone resorption and the progressive development of the fibrous tissue membrane seen in radiological loosening (Perren et al, 1975).

### 1.3. Stress Analyses of Bone–Prosthesis Structures

The stress analysis procedure involves the following aspects: (1) description of the geometry of the structure, (2) determination of the constitutive equations governing the mechanical behavior of the materials, and the values of relevant parameters, (3) an adequate method of solution, (4) determination of the boundary loads and displacements, (5) the actual calculations, (6) creation and execution of verification experiments, (7) the evaluation of model precision and validity, and (8) interpretation of results as seen in the light of the original problem (Fung, 1981).

In biomechanics of joint replacement, virtually all these aspects are highly complicated when compared to analyses of engineering structures. Bone geometry cannot be described accurately by mathematical functions. Cortical

bone is approximately linear elastic, transversely isotropic, and relatively homogeneous, but spongeous bone is highly anisotropic and nonhomogeneous in its mechanical properties.

Another difficulty is the modeling of joint and muscle forces. Although estimates of these forces in particular human functions (e.g., walking, stair climbing) are available in the literature (e.g., Paul et al, 1972; Röhrle et al, 1984), they are highly variable in both magnitudes and orientations, and general worst-case values in patients have not been firmly established. In addition, it is possible that high loading rates occur to such an extent that dynamic effects should be accounted for.

Additional complicating factors are the physical properties and constitutive equations for material connections (interfaces), the inaccessibility of these structures to measure strains at the most vital locations, and the interpretation of model results relative to the clinical reality.

The final, but probably most significant and principle problem is the high variability of the structural characteristics. This is due not only to the natural variety in bone shapes and properties, but also to the arbitrary and imprecise nature of the implantation procedures.

In view of these complicating factors, contemporary models for structural analyses of implant structures must be necessity be highly abstract and relatively simple, suited to study general effects on a relative basis, rather than to predict precise stress values (Huiskes and Chao, 1983).

The function of a model in science in general, whether theoretical or experimental, has been well expressed by Hooft (1979) in the following paragraph (translated):

To a scientist a mathematical model is as a stethoscope to a physician or a trowel to a bricklayer. The model is an image of reality, an abstraction, which touches upon reality in such a manner as to render some if its characteristics visible. It is the expedient relating complex systems to more trivial and simple systems, which last then represent the former. The development of models enables us to compensate for the limits of our imagination and descriptive abilities; by refraining from details and concentrating on the essentials, the complex reality becomes surveyable, controllable and comprehensible.

Evidently, a model can fail to give the desired information by being both too detailed and too coarse in describing the relevant aspects of a complex system. Modeling compromises are inevitable and must be carefully tailored to suit the objectives of the analysis. Evaluating the mechanical behavior of two specific prosthetic designs on a relative basis, for example, requires different model characteristics than does evaluating the effects of design parameters in general. Quite often models are developed simply with the purpose of obtaining an initial general idea about the behavior of a structure and to find guidelines for subsequent experimentation and analysis. Of course, models can be refined when more experimental data become available.

An example of such a development, concerning the mechanical behavior of

femoral hip prostheses, is presented in the next sections. The examples are mainly taken from the author's previous work. References for other relevant investigations may be found in Huiskes and Chao (1983) and in Huiskes (1984).

## 2. Load Transfer in Intramedullary Hip Prostheses

### 2.1. Principles and General Effects

A typical characteristic of load transfer in intramedullary fixated prostheses is "stress shielding" or "stress protection" of the bone by the prosthetic stem. It is based on the fact that the intramedullary stem carries a part of the load, which would otherwise have been carried by the bone alone ("load sharing"). The result is that the bone stresses are reduced after implantation of the prosthesis. This is nicely illustrated in Figure 18.2, where the lower part shows stresses at the bone surface as calculated from a strain-gauge experiment on a cadaver bone, intact and after subsequent implantation of two different prostheses,

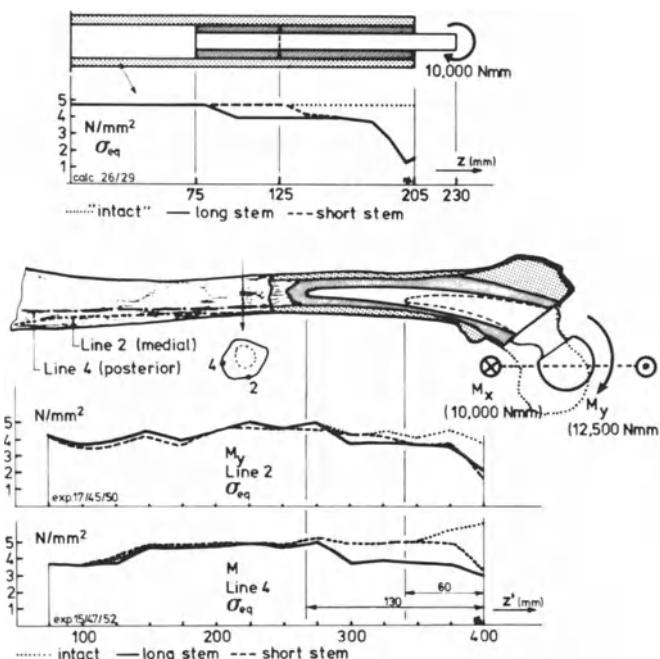


FIGURE 18.2. Von Mises stresses on the surface of a femur, before and after subsequent implantation of two prostheses. As calculated in a simplified (axisymmetric) FEM model (upper part) and as measured during *in-vitro* strain-gauge experiments, for bending in two planes (reproduced from Huiskes, 1979).

fixated with acrylic cement (Huiskes et al, 1981). The upper part shows the same effect in a simple (axisymmetric) Finite Element Method (FEM) model (Huiskes, 1979).

It is widely accepted that “stress shielding” is responsible for “disuse osteoporosis” in the cortex, a process in which cortical bone gradually resorbs locally and becomes more porous than usual, and for “calcar resorption,” which is massive resorption of bone in the medial, proximal region. Both are often seen in clinical cases, and the philosophy behind the stress-shielding hypothesis is based on the assumed adaption of bone to reduced strains (Cowin et al, 1984). It is feasible that disuse osteoporosis is actually a process of cortical thinning (Woo, 1981), hence, a process of surface remodeling, as is calcar resorption.

The simplified (axisymmetric) model shown in Figure 18.2 is quite useful for studying the principles of the load transfer mechanism in general, the more so since it allows for the application of analytical, “closed-form” methods (Huiskes, 1979; Huiskes et al, 1982; Huiskes, 1984). These methods are based on beam-on-elastic-foundation theory, and assume that the bone and stem, separately, behave in accordance with linear elastic beam theory, whereas the intermediate cement layer (the stiffness of which is an order of magnitude smaller than bone and two orders of magnitude smaller than metal) is assumed to act as a continuous row of linear springs.

The application of this theory allows the distribution of the principal internal loading variables to be approximated as functions of the external load and the most important structural parameters. An example for bending ( $M_0$ ) is shown schematically in Figure 18.3 (compare the upper part of Figure 18.2). In this case, the internal loading variables  $M_b(z)$  (Nmm, the bending moment carried by the bone),  $M_s(z)$  (Nmm, the bending moment in the stem), and  $p(z)$

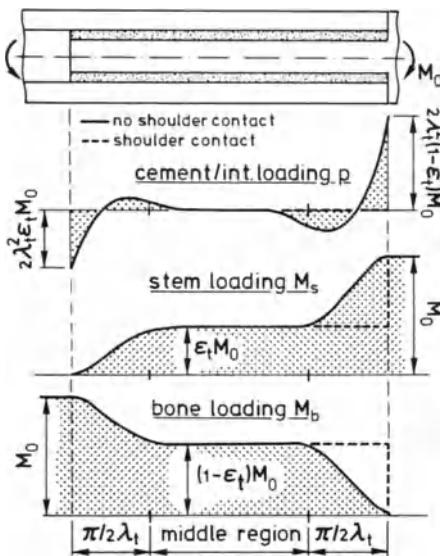


FIGURE 18.3. Schematic representation of the load transfer mechanism in pure bending, with and without (idealized) shoulder contact, in the case that the stem is relatively long (compare Figure 18.4a). The lumped transverse force  $p(z)$ , the stem ( $M_s(z)$ ), and bone ( $M_b(z)$ ) internal bending moments are shown as functions of the longitudinal coordinate  $z$ . The distribution depends on the lumped structural parameters  $\lambda_t$ , the fixation exponent, and  $\varepsilon_t$ , the relative stem rigidity (reproduced from Huiskes, 1984).

(N/mm, the lumped transverse load transferred from cement to bone, through the cement layer and the interfaces) are shown ( $z$  is the longitudinal coordinate). As is evident from this figure, the curves depend on the structural parameters  $\varepsilon_t$  (the relative stem flexural rigidity) and  $\lambda_t$  ( $\text{mm}^{-1}$ , the fixation exponent) only. The former can be evaluated from the flexural rigidities of the stem and the bone (elastic modulus  $x$  second moment of inertia), the latter depends on the transverse stiffness of the cement layer, and the flexural rigidities of stem and bone (Figure 18.4a).

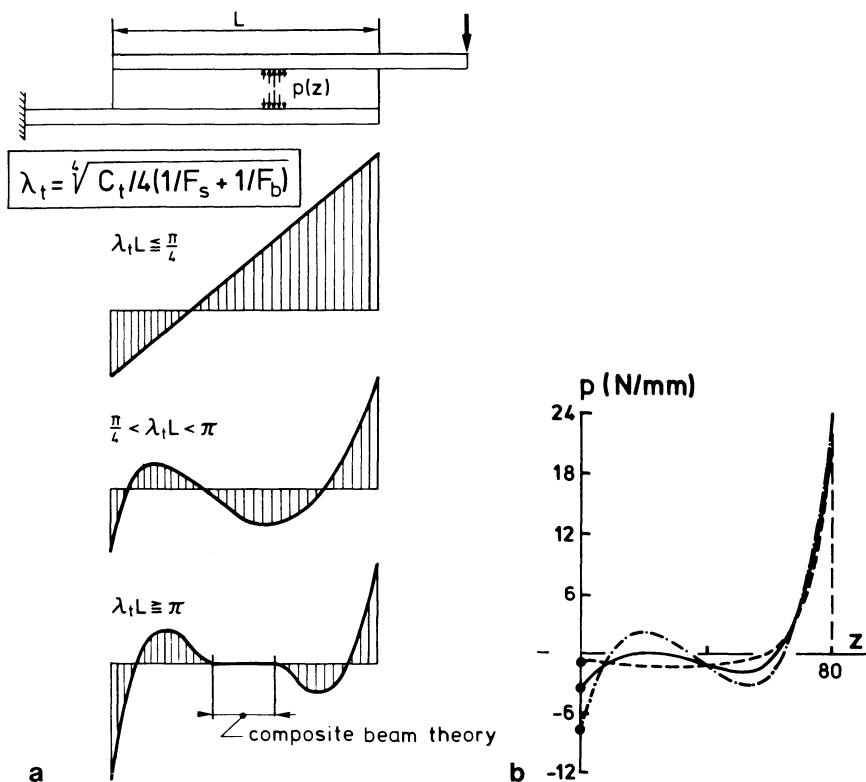


FIGURE 18.4. (a) The load transfer mechanism in transverse force loading and bending is characterized by the fixation exponent  $\lambda_t$  ( $\text{mm}^{-1}$ ), which depends on the flexural rigidities of the stem ( $F_s$ ,  $\text{Nmm}^2$ ) and the bone ( $F_b$ ,  $\text{Nmm}^2$ ), and on the transverse stiffness of the cement mantle ( $C_t$ ,  $\text{N/mm}^2$ ).  $C_t$  can be evaluated from the cement geometry, elastic modulus, and Poisson's ratio. The curves show the (schematic) distributions of the lumped transverse load  $p(z)$ , transferred through the cement mantle and over the interfaces, for the different value ranges of  $\lambda_t L$ . Comparable characteristics can be found for axial loading and torsion (reproduced from Huiskes, 1984). (b) Distribution of the lumped transverse load  $p(z)$  in the case of a straight stem (dash-point-dash curve, compare Figure 18.4a) and two tapered stems, taper angles  $2.8^\circ$  (continuous curve) and  $5.7^\circ$  (dashed curve). All stems are 80 mm long and 10 mm thick at  $z = 40$  mm (reproduced from Huiskes, 1979).

The internal bending moments  $M_s(z)$  and  $M_b(z)$  can be related to the stem and bone bending stresses, using linear elastic beam theory. The lumped transverse load  $p(z)$  is related to the transverse (radial) stresses  $\sigma_r$  and the tangential shear stresses  $\tau_{r\phi}$  in the cement layer and at the interfaces by

$$p(z) = \int_0^{2\pi} \{\sigma_r(\phi, z) \cos \phi - \tau_{r\phi}(\phi, z) \sin \phi\} r d\phi. \quad (18.1)$$

Generally speaking, the curves in Figure 18.3 disclose that the load is transferred in two regions, a proximal and a distal one. The stiffer the stem, the more load it carries in the middle region. This increases the amount of load transferred at the distal side and reduces the amount transferred at the proximal side.

The lengths of the load transfer regions can be approximated by  $\pi/2\lambda_t$ . Hence, if the stem length is reduced to  $L < \pi/\lambda_t$ , an interaction between proximal and distal load transfer will occur, as shown in Figure 18.4a. However, the general trends in the effects of the structural parameters on the internal loading variables remain the same until the stem length approaches  $\pi/4\lambda_t$ .

Evidently, this simple model can be used only to illustrate the principles of the load transfer mechanism in general. If the stem has a tapered shape (as most prosthetic stems do), numerical tools must be applied to obtain solutions, and the results cannot be expressed in the form of readily interpretable formulas.

Figure 18.4b shows a comparison of transverse load transferred by a straight stem and two tapered ones. These results have been obtained with an FEM model in which again the beam-on-elastic-foundation characteristics are assumed (Huiskes, 1979). The effect of the tapers, apparently, is that they smooth the load transfer at the distal side in particular.

## 2.2. Effects in Specific Prostheses

FEM meshes of two-dimensional models representing four different designs of femoral hip prostheses are shown in Figure 18.5 (Huiskes and Vroemen, 1985). One of these relies only marginally on cement fixation (Figure 18.5a), and one is completely uncemented (Figure 18.5d). The models are of nonuniform thickness and have "sideplates" to account for the three-dimensional structural integrity of the bone (Huiskes and Chao, 1983). The geometry of the bones is equal in all cases.

Figure 18.6 shows some stress results, assuming a unit hip-joint bending moment of 1 Nmm. The reason for choosing this load in a comparison of designs is that the stress patterns are highly sensitive to the hip-joint force orientation and point of application, an effect that can shade the significance of particular stem design characteristics in the load transfer mechanism. The effects of a bending moment are invariant to translations or rotations within the same plane.

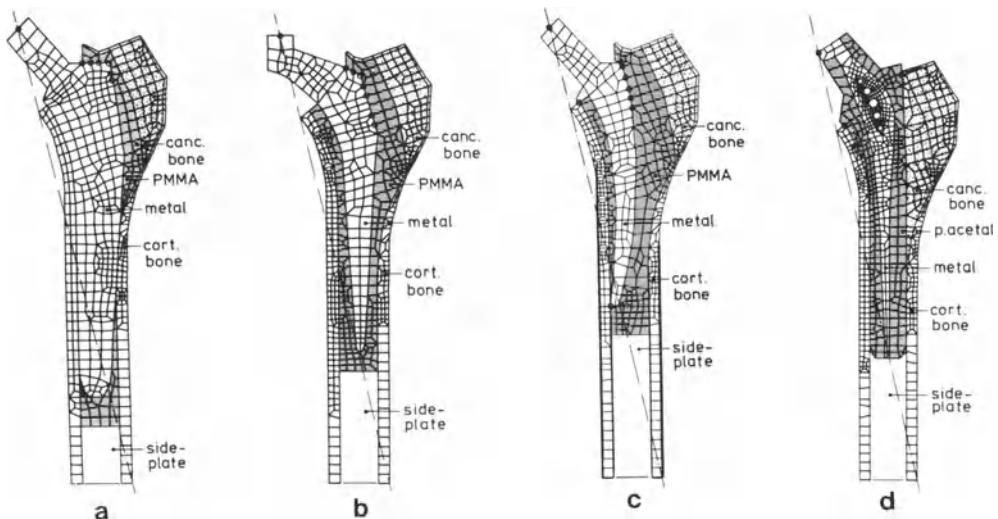


FIGURE 18.5. Element meshes (front plates only) for two-dimentional FEM models to evaluate four different prosthetic designs of femoral hip replacement, coded as (a) MSS, (b) EXE, (c) MCS, (d) ISO (Huiskes and Vroemen, 1985).

The stresses on the outside femoral surface (Figure 18.6a) qualitatively reproduce those found in the strain-gauge experiments and the simplified model (Figure 18.2). Obviously, the heavy stem (MSS) displays the most serious stress-shielding behavior. The stress curves of the other three, more flexible prostheses are not far apart. It is attractive to conclude from these curves that the effects of these three latter stems are approximately equal. However, this is far from the truth, for the following reason.

Referring again to the simplified model, described by beam-on-elastic-foundation theory, it can be shown (Huiskes, 1979) that the internal bending moments of the stem ( $M_s(z)$ ) and the bone ( $M_b(z)$ ), and the lumped transverse load  $p(z)$  are interrelated by

$$\frac{d^2 M_s}{dz^2} + p = 0 \quad \text{and} \quad \frac{d^2 M_b}{dz^2} - p = 0, \quad (18.2)$$

even when the stem and the bone are of nonuniform cross section. Hence, using  $M_s + M_b = M_o$ ,

$$p(z) = \frac{1}{2} M_o - \frac{d^2 M_b}{dz^2}, \quad (18.3)$$

or, in other words, the transverse load transferred from stem to bone depends on the second derivative, rather than the actual value of the bone bending moment. This also predicts that strain-gauge measurements on the outside

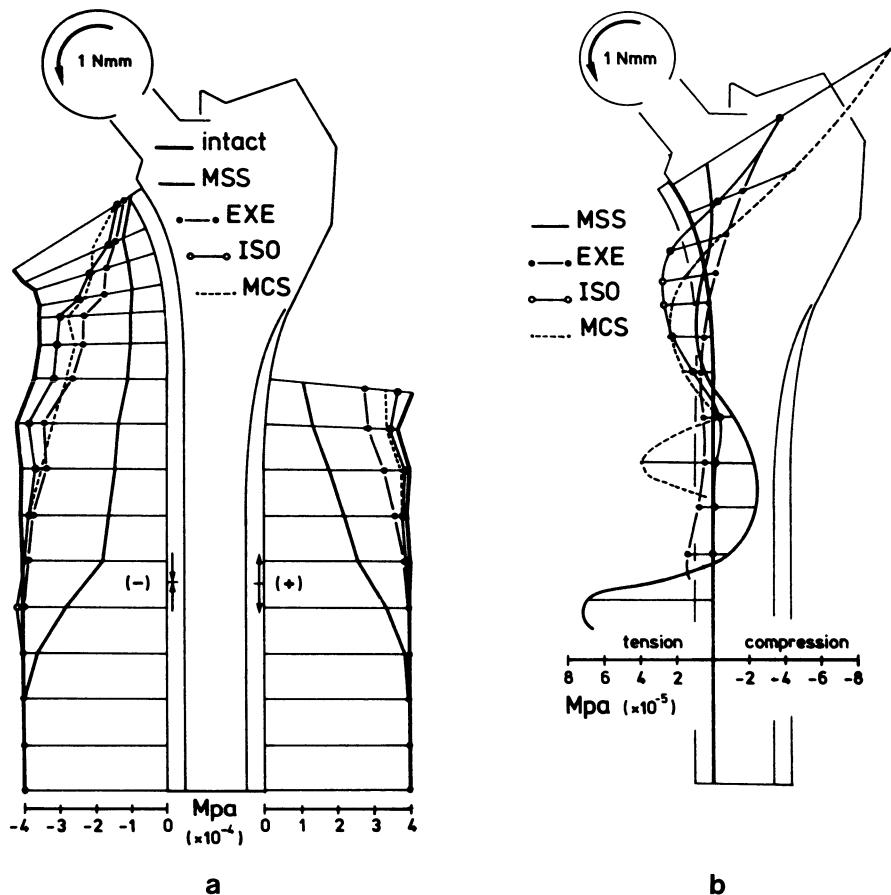


FIGURE 18.6. Comparison of stresses as calculated with the FEM models shown in Figure 18.5. (a) Longitudinal (bending) stresses at the bone surface; (b) normal direct stresses at the medial implant–bone (inner cortex) interface (Huiskes and Vroemen, 1985).

bone surface cannot give precise quantitative information about cement and implant–bone interface stresses.

These principles are reflected in the comparison of implant–bone interface stresses for the four stem designs, which are vastly different (Figure 18.6b). The heavy stem (MSS) generates high peak stresses at the distal side in particular, whereas the more flexible stems generate high stresses on the proximal side, in accordance with the principles described earlier. The taper effects in the different stem designs are obvious if Figures 18.6b and 18.4b are compared. The effects of the heavy stem (MSS) closely resemble those of the straight stem in the simplified model, whereas the effects of the other prostheses are compa-

able to those of the tapered stems, in particular where it concerns the EXE design.

It can be concluded from this brief discussion that the load transfer mechanism in intramedullary hip prostheses, and in particular the stresses that are generated at the implant bone interfaces, are highly susceptible to the details of stem designs. It is also evident that compromises are inevitable in prosthetic design, e.g., proximal versus distal load transfer, proximal load transfer versus stress-shielding, etc. Criteria for these compromises must be derived from the question of what the stresses actually do to bone, which is discussed in the next section.

### 3. Implant–Bone Interactions

The examples in the previous section nicely illustrate the load transfer mechanism with respect to intramedullary prostheses in general, and the actual stress patterns in particular designs, in idealized circumstances. It is questionable whether three-dimensional FEM models can provide additional information, in view of their complexity, as long as the same basic assumptions with respect to constitutive material behavior are applied. Evidently, three-dimensional models are very useful, already in this stage, to provide reference data bases for simpler models (e.g., Rohlmann et al, 1983).

One of these basic assumptions, affecting the actual stress patterns in particular, is the rigid bond at material connections, such as the metal–cement interface. Normal direct and shear stress patterns at the metal–cement interface of the MCS prosthesis, as predicted by the FEM model discussed in the previous section, are shown in Figure 18.7a. If we compare the values of these stresses, for a hip-joint force of about three times body weight, to the strength of this connection (Ahmed et al, 1984), it is evident that there is hardly a safety factor involved, and loosening of this interface is likely; probably initialized at the medial, distal stem tip, where tension peaks occur in combination with shear peak stresses. Also, in view of experimental and clinical evidence, it is likely that cement–metal interfaces loosen as a rule, relatively shortly after load bearing is resumed postoperatively.

Several investigators have shown that loosening of this interface, although not affecting the basic principles of the load transfer mechanism, has a drastic influence on the actual stress patterns within the cement mantle and at the interfaces (Huiskes and Schouten, 1980; Hampton et al, 1981; Klever et al, 1983), an example of which is presented in Figure 18.7b.

Hence, when actual stress values are to be compared to mechanical failure criteria in cemented bone–prostheses structures, it is questionable whether linear models, assuming bonded interfaces, give information relevant to the actual, clinical situation. It is even possible that nonlinear interface behavior must be assumed when actual designs are compared on a relative basis.

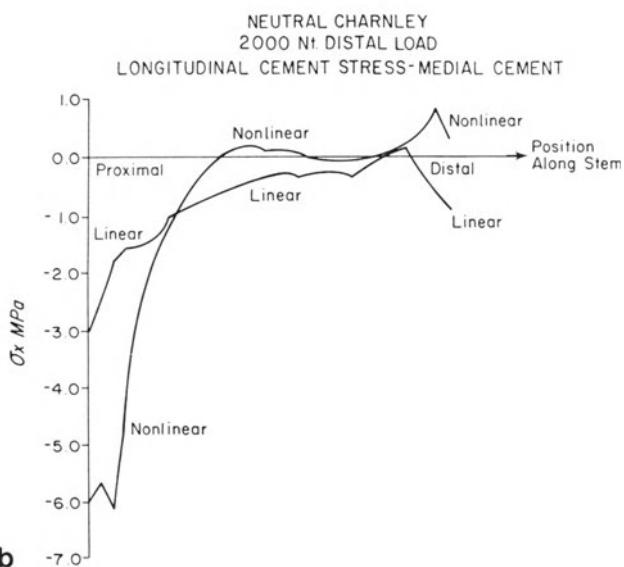
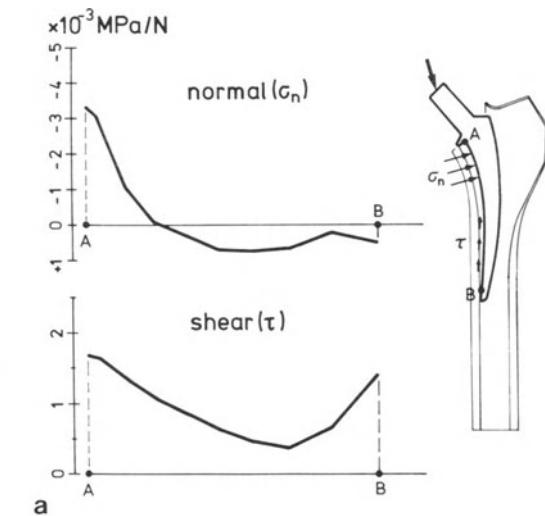


FIGURE 18.7. Stem-cement interface stresses, (a) normal direct and shear, at the medial side, relative to the magnitude of the hip-joint force, as calculated in the FEM model for the MCS prosthesis (note the correspondence of the normal stress pattern with the tapered stem curve of Figure 18.4b). (b) Medial cement stresses as calculated with a two-dimensional FEM model, simulating a Charnley femoral hip replacement (compare Figure 18.1), assuming a rigidly bonded stem-cement interface (linear), and a loose stem-cement interface (nonlinear). Note the stress increases in the latter case (Reproduced from Hampton, 1981).

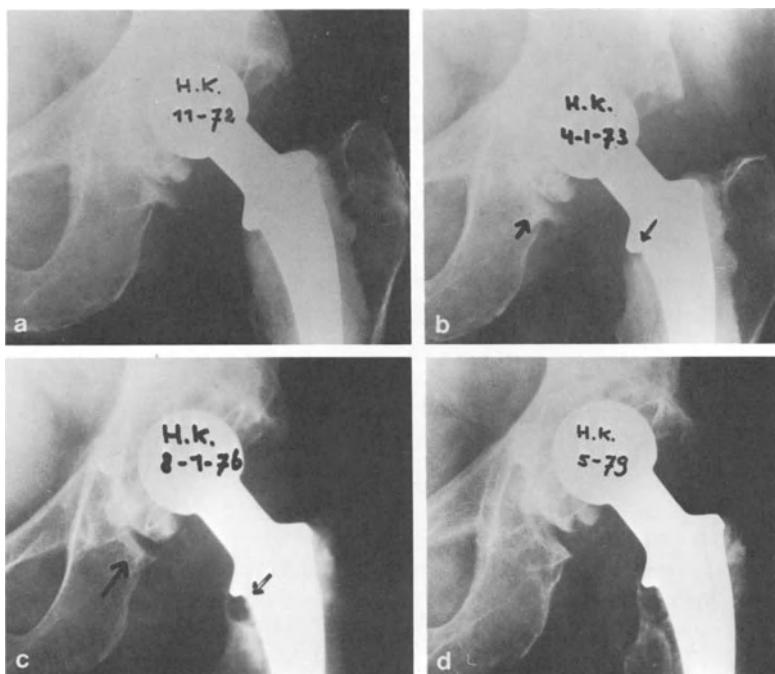


FIGURE 18.8. X ray showing gradual resorption of the calcar (the proximal, medial bone region). (a) directly postoperative, (b) after two months, (c) after 38 months, (d) after 78 months (reproduced from Huiskes, 1979).

Although time-consuming for a computer, this is definitely feasible with the sophisticated numerical tools recently developed.

Unfortunately, however, cement–stem interface loosening is only one aspect of nonlinear behavior. A more complicated role is played by the implant–bone interface and by the biological reactions of bone, in cemented and cementless prostheses alike. Examples, already discussed, are disuse osteoporosis and calcar resorption. Both may be *caused* by stress (or strain) patterns in bone, but they also (and most certainly) have an *effect* on the stress patterns.

An example of calcar resorption, as seen on x rays, is shown in Figure 18.8. Evidently, such a massive disappearance of bone leads to significant loss of support, the extent of which is illustrated in Figure 18.9, showing shear stresses at the implant–bone interface for two of the prostheses previously discussed, before and after calcar resorption. The effects on the interface stress distribution are qualitatively different. The heavy MSS stem has a much better “fail-safe” design in this case than the flexible, strongly tapered EXE stem. On the other hand, however, if calcar resorption is indeed caused by stress-shielding, then it is more likely to occur in association with the MSS design than with the EXE design (compare Figure 18.6a).

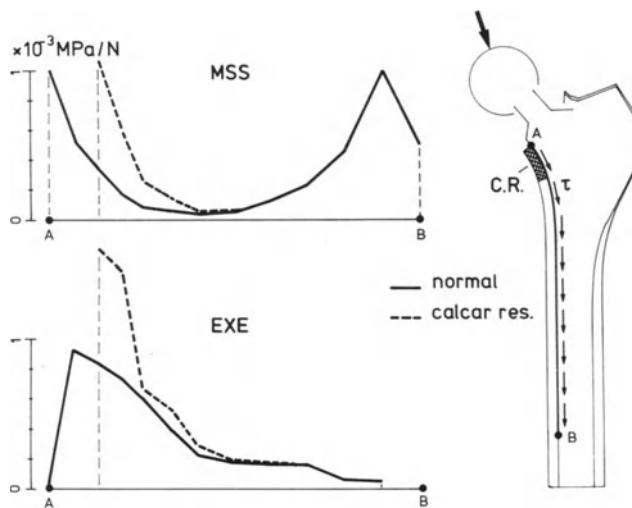


FIGURE 18.9. The effects of calcar resorption on the shear stresses at the medial implant–bone interface for the MSS design and the EXE design, as predicted by the FEM models (compare Figure 18.5).

An example of implant–bone interface remodeling is presented in Figure 18.10a, showing bone ingrowth patterns in a local piece of a profiled implant, six months postoperatively. Figure 18.10b shows stress patterns evaluated with the FEM (local two-dimensional models), simulating subsequent stages of bone ingrowth, from direct postoperative to full ingrowth (Huiskes and Nunamaker, 1984). Evidently, the local stress patterns at the implant–bone interface change drastically through local bone remodeling, and in this case reduce the chances for interface failure. Comparing the stress patterns to the ingrowth patterns, it can even be hypothesized that the ingrowth mechanism is enhanced by the stress patterns. Conversely, as was shown in the same investigation, in regions where the initial, postoperative interface stress peaks (Figure 18.10b, top) are excessively high, no ingrowth occurs, but bone resorbs instead.

Finally, consider Figure 18.11a, showing x rays of the so-called iso-elastic prosthesis directly postoperative and after four years. The results concern an early design of this (cementless) prosthesis, with a flexible metal core embedded in polyacetal. A problem with this design was early loosening, evident in particular at the upper part of the medial implant–bone interface (Figure 18.11a). Because of this problem, a stiffer core was introduced. Both designs were simulated with the two-dimensional models previously discussed (Figure 18.5d), and an example of results—shear stresses at the implant–bone interface—is shown in Figure 18.11b. Evident is the shear stress peak in the case of the early, flexible core, precisely where the problems were seen clinically. It

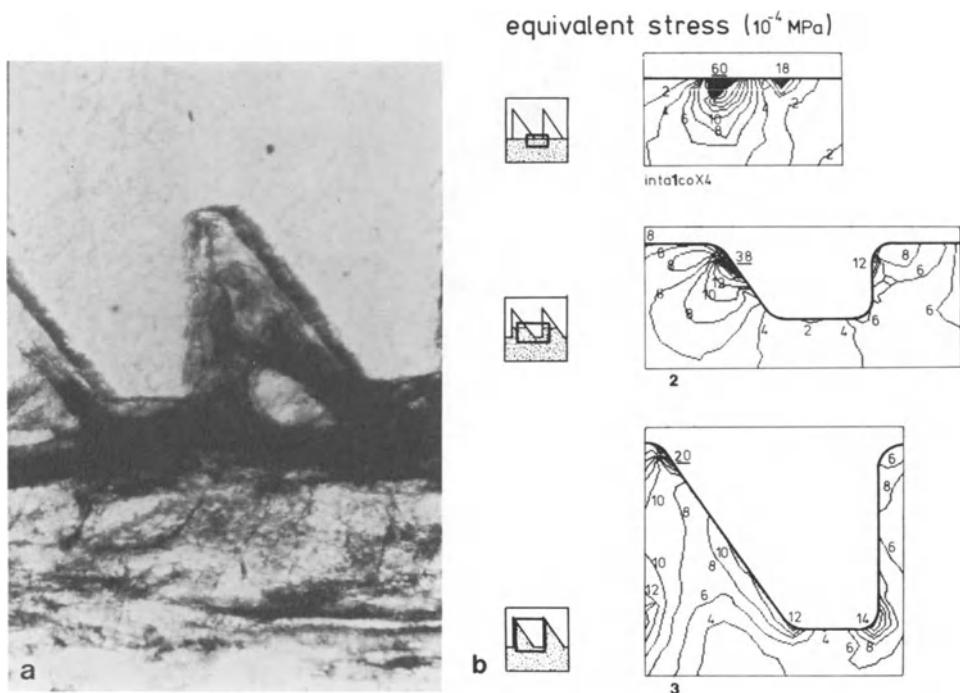


FIGURE 18.10. (a) Bone ingrowth patterns into a profiled intramedullary plug, with grooves of approximately 1.5 mm at the outside, six months postoperatively. (b) Von Mises stress patterns in local bone as calculated for a two-dimensional FEM model of the local implant–bone interface, shown in Figure 18.9a, simulating gradual bone ingrowth (reproduced from Huiskes and Nunamaker, 1984).

can now be argued that the eventual clinical problem is a result of failure *initiation* at the interface through excessive interface stresses. However, it is also feasible that interface failure occurs anyway, but that the real problem is in failure *propagation* through micromotions and subsequent bone resorption. And of course, the extent of the micromotions also depends on the structural characteristics of the prosthetic design.

Combined mechanical and histological investigations of other prosthetic types have also suggested the probability that failure propagation, rather than initiation, is the important failure criterion for prosthetic designs (Huiskes et al, 1984).

In summary, implant and bone truly interact mechanically through phenomena such as osteoporosis, massive bone resorption, interface remodeling, and resorption, which are likely to be induced, or at least influenced, by stress (strain) patterns and relative motions and, in their turn, cause changes in the stress and motion patterns. Mechanical analyses of prosthetic designs and failure predictions can only be complete if these effects are taken into account

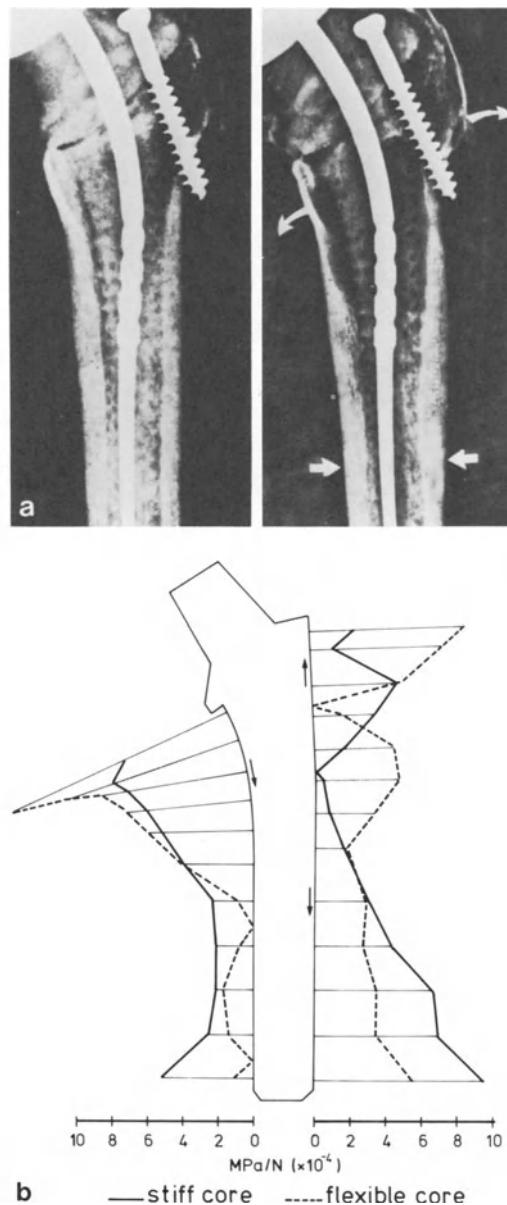


FIGURE 18.11. (a) X rays showing proximal loosening of an early design of the so-called iso-elastic prosthesis; the flexible metal core was later replaced by a stiffer one. Left: initial situation, Right: after four years (reproduced from Morscher, 1983). (b) Shear stresses at the implant–bone interface, as evaluated for the ISO design, comparing a rigid metal core (Figure 18.5d) to a flexible one (Figure 18.11a).

by introducing feedback loops into the mathematical descriptions of the bone properties, tying stresses to biological reactions.

Although very little is known about mechanical induced bone reactions in a quantitative sense, modeling efforts in this direction should be enhanced. As stated by Fung (1981) in a slightly different context:

The most serious frustration to a biomechanics worker is usually the lack of information about the constitutive equations of living tissues. Without the constitutive laws, no analysis can be done. On the other hand, without the solution of boundary value problems the constitutive laws cannot be determined. Thus, we are in a situation in which serious analyses (usually quite difficult because of nonlinearity) have to be done for hypothetical materials, in the hope that experiments will yield the desired agreement. If no agreement is obtained, new analyses based on a different starting point would become necessary. (p.13)

Constitutive equations for strain-induced remodeling in normal bones, both for internal remodeling (e.g., disuse osteoporosis) and external remodeling (volume increase or decrease) were derived from a theory of adaptive elasticity (Cowin, 1981). To describe surface remodeling, for example, it was assumed that the rate of outward surface movement in a given point is linearly proportional to the difference between the local strain state and a local reference strain state at which no remodeling occurs (Cowin, 1981):

$$u = c_{ij}\{E_{ij} - E_{ij}^0\}, \quad (18.4)$$

in which the summation convention applies.  $E_{ij}$  is the local strain tensor,  $E_{ij}^0$  the local reference strain tensor, and  $c_{ij}$  are remodeling coefficients yet to be determined. If the remodeling surface is a free surface (normal bone), the indexes in the above equation may range over 1, 2. Fung (1981), however, suggested the use of:

$$u = c_{ij}\{E_{ij}^2 - (E_{ij}^0)^2\}, \quad (18.5)$$

where the indexes range over 1, 2, 3; first, in order to include remodeling due to surface stresses caused by implants, and second, to incorporate the hypothesis that tensile stress and compressive stress have the same effects.

Although it is questionable whether implant–bone interface tension plays the same role as compression in interface–bone resorption, and although these equations do not account for all the phenomena discussed previously, they may be good starting points for realistic analyses of biomechanical implant–bone interactions.

## References

- Ahmed, A.M., Raab, S., and Miller, J.E. 1984. *J. Ortho. Res.* 2:105–118.
- Charnley, J. 1970. Acrylic cement in orthopedic surgery. E. and S. Livingstone, London.
- Cowin, S.C. 1981. Pages 193–210 in S.C. Cowin, ed. *Mechanical properties of bone*. AMD, vol. 45, ASME, New York.

- Cowin, S.C., Lanyon, L.E., and Rodan, G., eds. 1984. Proc. Conf. on Functional Adaption in Bone Tissue. *Calc. Tiss. Int.* 36:s1–s161.
- Feith, R. 1975. *Acta Orth. Scand. Suppl.*: 161.
- Fung, Y.C. 1981. *Biomechanics: mechanical properties of living tissues*. Springer-Verlag, New York.
- Galante, J.O., ed. 1983. *Clin. Orth. Rel. Res.* 176:2–108.
- Goldring, S.R., Schiller, A.L., Roelke, M., Rourke, C.M., O'Neill D.A., and Harris, W.H. 1983. *J. Bone Joint Surg.* 65[A]:575–584.
- Hampton, S.J. 1981. Ph.D. thesis, University of Illinois, Chicago.
- Hampton, S.J., Andriacchi, T.P., Draganich, L.F., and Galante, J.O. 1981. Page 144 in *Proc. 27th Ann. Mtg. Orth. Res. Soc.*
- Hooft, J.P. 1979. *De Ingenieur* 91:289.
- Huiskes, R. 1979. *Acta Orth. Scand. Suppl.*: 185.
- . 1984. Pages 121–162 in P. Ducheyne and G.W. Hastings, eds. *Functional behavior of orthopaedic biomaterials*. Vol. 2, *Applications*. CRC Press, Inc., Boca Raton, Florida.
- Huiskes, R., and Schouten, R.Y. 1980. Pages 213–217 in V.C. Mow, ed. *1980 Advances in bioengineering*. ASME, New York.
- Huiskes, R., Janssen, J.D., and Slooff, T.J. 1981. Pages 211–234 in S.C. Cowin ed. *Mechanical properties of bone*. AMD, vol. 45, ASME, New York.
- . 1982. Pages 313–343 in R.H. Gallagher, B.R. Simon, P.C. Johnson, and J.F. Gross, eds. *Finite elements in biomechanics*. John Wiley & Sons, New York.
- Huiskes R., and Chao, E.Y. 1983. *J. Biomech.* 16:385–409.
- Huiskes, R., and Nunamaker, D. 1984. *Calc. Tiss. Int.* 36:s110–s117.
- Huiskes, R., Strens, P., van Heck, J., and Slooff, T.J. 1985. *Acta Orth. Scand.*, in press.
- Huiskes, R., and Vroemen, W. 1985. In *Proc. 31st Ann. Mtg., Orth. Res. Soc.* Held at Las Vegas, Nevada, 21–24 Jan.
- Klever, F.J., Klumpert, R., Grootenboer, H.J., van Campen, D.H., Panly, T., and Matthiass, H.H. 1983. Pages 55–58 in S.L. Woo and R.E. Mates, eds. *1983 Biomech. Sypos.*, AMD, vol 56, New York.
- Morscher, E. 1983. Pages 188–195 in E. Morscher, ed. *Die zementlose Fixation von Hüftenprothesen*. Springer-Verlag, Berlin.
- Paul, J.P., Hughes, J., and Kenedi, R.M. 1972. In Y.C. Fung, V. Perrone, and M. Anliker, eds. *Biomechanics: its foundation and objectives*. Prentice-Hall, Englewood Cliffs, N.J.
- Perren, S.M., Ganz, R., and Rueter, A. 1975. *Med. Orth. Tech.* 95:6–10.
- Rhinelander, F.W., Nelson, C.L., Stewart, R.D., and Stewart, C.L. 1979. *Clin. Orth. Rel. Res.* 141:74–89.
- Rohrle, H., Scholten, R., Sigolotto, C., Sollbach, W., and Kellner, H. 1984. *J. Biomech.* 17:409–424.
- Rohlmann, A., Moessner, U., Bergmann, G., and Koelbel, R. 1983. *J. Biomech.* 16:727–742.
- Stauffer, R.N. 1982. *J. Bone Joint Surg.* 64[A]:983–990.
- Walker, P.S. 1977. *Human joints and their artificial replacements*. Charles Thomas Co., Springfield, Illinois.
- Willert, H.G., and Semlitsch, M. 1976. Pages 325–346 in N. Schaldach, and D. Hohmann, eds. *Advances in hip and knee-joint technology*. Springer Verlag, Berlin.
- Woo, S.L-Y. 1981. Pages 107–130 in S.C. Cowin, ed. *Mechanical properties of bone*. AMD, vol. 45, ASME, New York.

# 19

## Mechanics of Healing Soft Tissue Wounds

A.V. VIIDIK and F. GOTTRUP

### 1. Biology of the Healing Soft Tissue Wound

Before dealing with the mechanical behavior of healing wounds, this chapter will discuss key aspects of the biological phenomena that form the basis for development of physical strength in healing wounds. These phenomena and the factors capable of modifying them form a necessary part of the basis for the design of adequate protocols for the mechanical testing of healing wounds.

#### 1.1. The Basic Phenomena

The primary goal of the wound-healing process after injury to soft tissue is the restoration of continuity between the wound edges. This takes place in mammals, most often regardless of the type of original tissue, by bridging the gap with a scar of collagenous connective tissue. Regeneration of the original tissues is, however, contrary to common belief, not exceptional (Glynn, 1981). The regeneration of epithelia and endothelia is well known. Regenerative capability to varying extents has been shown to be present in kidney, liver, and striated muscle as well as in the central and peripheral nervous systems. The healing by scar formation can thus not immediately be taken for granted in the design of an experimental model.

The healing process can be divided into three histological phases, one gradually merging into the next: (1) inflammation, (2) fibroplasia, and (3) maturation. A “cleaning up” process prepares for the formation of granulation tissue, which is subsequently replaced by fibrous scar tissue. Most of our knowledge of these processes is derived from studies on tissues containing predominantly connective tissues (e.g., skin).

##### 1.1.1. THE INFLAMMATORY PHASE

This, the earliest phase after injury (also called the lag phase), takes place during the first three to four days of healing. The tissue defect is filled with clotted blood, tissue debris, and fluid; the strength of the wound is derived only

from fibrin in the clot and is not sufficient to keep the tissue edges together. Sutures or some other mechanical means must be used to keep the wound closed.

The classical signs for inflammation (redness, swelling, heat, and pain) are seen, though less pronouncedly than in a local infection. Dilatation of vessels and increased vessel wall permeability results initially in local edema and migration of polymorphonuclear cells, which primarily provide the defense against bacteria. Macrophages increase in number later in this phase; they remove the debris and, furthermore, play an important role in the regulation of the healing process. Vessel proliferation starts also, and toward the end of this phase, fibroblasts derived from undifferentiated perivascular mesenchymal cells begin to migrate into the wound area.

#### 1.1.2. THE FIBROPLASIA PHASE

The strength of the closed wound (e.g., a linear skin wound sutured immediately after incision) increases rapidly during this phase as "granulation tissue" is formed to repair the defect. This tissue, named after the reddish, granular surface seen in an open wound, contains endothelial cells, fibroblasts, and macrophages (derived from blood monocytes) as well as connective tissue elements. The endothelial cells participate in the angiogenesis and the fibroblasts secrete collagen and ground substance components.

While some collagenous elements can be seen already in the previous phase (after two to three days of healing), the peak period of collagen synthesis occurs for most tissues around the fifth to seventh day of healing. The collagen fibrils are laid down in a random pattern and, in the beginning, possess little mechanical strength. Gradually a more systematic pattern emerges and strength is gained, when the collagen in the microfibrils is stabilized by crosslinking and the fibrils are assembled into fibers. The diameter of the fibrils also increases; at least during the maturation process in normal skin such an increase is paralleled by that of mechanical strength.

During this phase, old damaged collagen is broken down enzymatically, as are some of the newly formed elements. The "biochemically active zone" encompasses tissue up to more than 5 mm from the incision line (Adamsons et al, 1966; Danielsen and Fogdestam, 1981; Gottrup, 1981a). After about two to three weeks of healing, the deposition and removal of connective tissue elements reach a steady state (Danielsen and Gottrup, 1981), while the number of fibroblasts and macrophages decreases together with the degree of vascularization in the area. The process merges into its third phase. On the other hand, if the wound edges are not properly adapted, this granulation tissue-forming (or proliferative) phase is prolonged considerably.

#### 1.1.3. THE MATURATION PHASE

The length of this phase is not known and extends frequently over periods of time much longer than a year. The scar tissue consists now of dense connective

tissue: collagen fibers dominate, while the amount of ground substance is considerably smaller than in granulation tissue and regular loose connective tissue. The turnover of collagen is still higher than in normal, dense connective tissue, and the mechanical strength of the tissue continues to increase, though less rapidly than during the phase of fibroplasia. This increase is explained, in the absence of an increase of the collagen content, by remodeling of the fiber architecture and continuing maturation of the fibrils. The gradual shift in collagen from Type III to Type I predominance might also contribute to the development of strength (similar to that in the ontogenetic development in skin).

The remodeling process takes place partly as a response to mechanical (stretching) stimuli; the development of strength in a totally immobilized wound is somewhat retarded (Hunt, 1979). The regularity of the fiber structure in normal dense connective tissue (e.g., skin) is, however, never regained, although the fiber pattern is to some extent aligned in the direction of the stresses present or applied experimentally.

## 1.2. Factors Influencing the Healing Process

### 1.2.1. NEGATIVE FACTORS

Factors that impair the healing process are well documented by numerous experimental as well as clinical studies (e.g., Hunt, 1979; Zederfelt, 1980). On the other hand, most of the literature describing means to improve wound healing does not deal with stimulation of the normal healing process but with restoration to normal of a disturbed mechanism. The inflammatory reaction as well as the fibroplasia can be affected.

A prolonged inflammatory reaction will prolong the lag phase and thereby retard the development of wound strength. This is most frequently caused by infections (often combined with excessive amounts of devitalized tissue) or irritating foreign bodies; these two factors do not add to but amplify each other. The experimental investigator should have in mind that certain suture materials (e.g., silk) are especially likely to cause such reactions and that further tissue damage should be avoided as much as possible. Other causes for prolongation of the inflammatory reaction are diabetes mellitus, immunodeficiency syndromes, and treatment with anti-inflammatory drugs (e.g., corticosteroids and immunosuppressants). A prolonged mild inflammation during the later phases of wound healing, on the other hand, tends to augment the formation of "granulation tissue."

Impaired fibroplasia, following a normal inflammatory phase, is most often caused by some local or systemic metabolic deficiency. The nutrition to and removal of metabolic waste from the wound area is dependent on the local circulation; hypoxia and/or accumulation of compounds like lactate caused by circulatory insufficiency create an adverse local environment, which delays the healing (Niinikoski, 1980) (for a discussion of the opposite effect of hypoxia on

the inflammatory phase, see Section 1.2.2). Hemorrhage delays wound healing by decreasing the circulating blood volume, but not by the anemia it causes; the same mechanism is true for the impairment of wound healing following extensive trauma to another part of the body. Increased oxygen tension counters these adverse effects. Hypoproteinemia will affect wound healing only when severe enough to produce generalized edema, while a lesser degree is sufficient when there is a concomitant infection. Ascorbic acid deficiency impairs the synthesis of collagen, and thereby wound healing, in the fibroplasia phase. Zinc deficiency is another negative factor. Wound healing is, contrary to "clinical experience," not impaired in old age (Holm-Pedersen, 1973; Gottrup, 1981b).

### 1.2.2. REGULATING FACTORS

The first step of healing, after bleeding from damaged arterioles and capillaries has ceased, is dilatation and increased permeability of venules. The results in local edema and migration primarily of polymorphonuclear cells into the tissue (for discussion of factors eliciting these processes, see Muller, 1981). The inflammatory process is thus established 12 to 24 hours after injury.

The next step in the healing process involves several factors that have regulatory functions. The first factors to stimulate the healing process are the *platelet-derived growth factor* (PDGF) (Ross and Vogel, 1978) and some not yet characterized blood plasma factor (present in "plasma-derived serum," or PDS) (Vogel et al, 1980), which stimulate mitotic activity in fibroblasts and endothelial cells. The growth factor, which probably plays the key role during the fibroplasia phase, is the *macrophage-derived growth factor* (MDGF) (Leibovich and Ross, 1975, 1976). The macrophage further stimulates angiogenesis (by *wound angiogenetic factor*, WAF), eliminates debris, and attracts additional macrophages, probably by chemotaxis. Its role with regard to synthesis of collagen is less clear (both stimulation and inhibition have been reported).

The random scatter of cells seen in the late stage of the inflammatory phase is then organized into the characteristic "wound-healing module" (Hunt and Van Winkle, 1979; Hunt et al, 1984): the macrophages line the injured surfaces with a layer of less mature fibroblasts beneath them. Still deeper, more mature fibroblasts are mitotically active and, here, angiogenesis (vessel budding) is also seen. The next layer contains newly synthesized collagen, the fibers of which orient themselves towards the collagen meshwork in the neighboring "intact" tissue, which is part of the "biochemically active zone." Such "wound-healing modules" are seen most distinctly in the granulation tissue formed in open wounds.

These "modules" move from both sides towards the wound line: the fibroplasia phase of healing is completed when the defect is eliminated. They are self-regulating from a functional point of view, the question being: what mechanisms initiate and terminate their activity? Angiogenesis and the activ-

ity of macrophages, which secrete increased amounts of WAF and MDGF, are primarily stimulated by the hypoxia always present in the wound area early in healing (Knighton et al, 1984). The proliferation stimuli are turned off when the tissue oxygen supply from newly developed vessels increases. The collagen synthesis, however, with the fibroblasts in an optimal microenvironment, proceeds until the defect is filled and continues thereafter at a level adequate to meet the requirements of the remodeling process during the maturation phase. The stimuli regulating this, the last part of the wound-healing process, are unknown.

Although there are today no means to accelerate normal healing (except by delayed primary closure, as discussed later, in Section 3.2.2), the further purification and characterization of the above-mentioned factors promises to provide means to enhance the development of mechanical strength in healing wounds. Such intervention would benefit the treatment of wounds in slow-healing tissues as well as in patients with impaired healing due to concomitant disease.

## 2. Designs for Mechanical Testing

Some methodological problems are universal for the testing of soft biological tissues, while others are specific for the analysis of healing wounds. This section focuses on the latter group of problems; for an overview of general technical considerations, see, for example, Viidik (1979).

### 2.1. *In-vivo* and *In-situ* Testing

From a biological point of view, *in-vivo* testing should provide the most relevant and reliable information, while *in-situ* testing should be the next best option. These approaches have, however, severe technical limitations that often render the value or interpretation of the results obtained questionable. The difference between these two approaches would be important only if rapid postmortem changes, not so far reported in the literature, were to be anticipated. These two approaches employ either blow-out or tear-apart techniques, while conventional tensile strength testing techniques are used for the *in-vitro* approach.

#### 2.1.1. THE BLOW-OUT TECHNIQUE

This technique can be used for measuring the strength of healing wounds in hollow organs (Chlumsky, 1899; Irvin, 1981) as well as in the abdominal wall, with the abdominal cavity being the “hollow organ” (Harvey, 1929; DeHaan et al, 1974). A segment of the organ (e.g., a part of the intestinal tract) with a healing wound is sealed off and inflated (with fluid or air) until the pressure to burst the weakest point in the specimen is reached (Figure 19.1a). This method

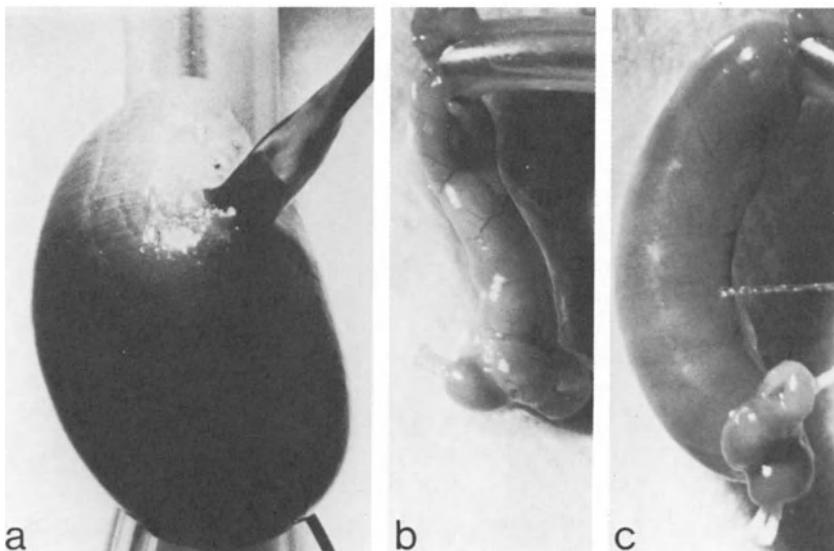


FIGURE 19.1. The blow-out technique. (a) A rat stomach with a healing wound is filled with fluid until it ruptures at the weakest point in the wall. (b) A fairly cylindrical segment of rat small intestine is sealed off proximally and distally and filled with fluid until a minimum pressure can be recorded. (c) The pressure in the specimen shown in (b) is increased until it bursts at its weakest point, which is at the mesenteric attachment; note the change in geometry.

measures the bursting strength in the testing situation, which approximates the physiological distribution of forces in the organ wall to the extent possible in the laboratory.

Attempts have been made to increase the accuracy of this measure by calculating the circular wall tension according to La Place's law (Nelsen and Anders, 1966; Irvin, 1981). This transformation would require a cylindrical tube. With few exceptions, like blood vessels and other ducts, the hollow organs in the body are less regular in their geometry. A transversal wound, for example, an anastomosis, would disturb a ring of the segment tested. A longitudinal wound would disturb whatever circularity there was in the organ before wounding. No other parameters than bursting strength can, therefore, be measured with accuracy. This is true also for a reasonably cylindrical organ, for example, the small intestine, for which the attachment of the mesenterium creates a geometry-changing constraint (Figure 19.1b and c) that changes its shape in an unpredictable (and unmeasurable) way with increasing pressure.

#### 2.1.2. THE TEAR-APART TECHNIQUE

This technique has most often been used for studying healing skin wounds. Hooks are pierced through the skin (Botsford, 1941; Sandblom et al, 1953) on

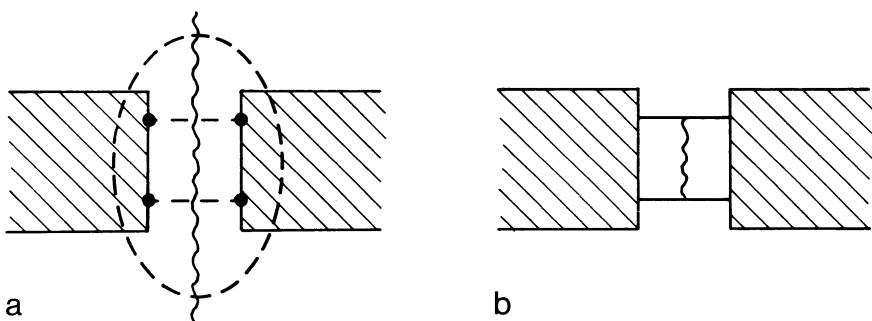


FIGURE 19.2. (a) Schematic drawing of a wound (indicated by the wavy line) fixed for *in situ* testing with the tear-apart technique. The solid circles indicate hooks pierced through the tissue; the hatched rectangles indicate the device generating the force. The dashed ellipse indicates the probable extent of tissue influencing the measurement, while the dashed straight lines and (b) indicate the clear delineation of a strip specimen tested by the *in-vitro* technique (Modified from Fogdestam and Gottrup, 1980).

either side of the linear wound; they are forced apart by some mechanical device and the separation force is recorded. This force (frequently mislabeled as breaking strength) is not dependent only on the mechanical properties of the wound but also on those of surrounding tissue, around and beneath the hooks (Figure 19.2) (see Fogdestam and Gottrup, 1980). These interferences are probably negligible during early wound healing, when the strength of the wound is minimal and the danger for tissue damage during preparation of strip specimens for *in-vitro* testing is pronounced. The best technique for testing skin wounds during early healing would be to adapt the technique of Wijn (1980) for measuring mechanical properties of skin *in vivo*. Stubs, fastened to a materials testing machine, are glued to both sides of the wound line and the load deformation relationship is recorded until separation.

## 2.2. *In-vitro* Testing

### 2.2.1. THE TESTING OF STRIP SPECIMENS

These techniques measure the mechanical properties of wound tissue and/or adjacent tissue, not the functional behavior of the healing wound *in situ*. With the exception of the van Noort technique (see Section 2.2.2), materials testing machines of the Instron tensile tester type are used, and the strip-shaped specimens are usually cut with their long axes perpendicular to the wound line (Figure 19.2b). The equipment enables the recording of a number of parameters, besides breaking and/or tensile strength. A comprehensive analysis should include the stress-strain curve from the relaxed state until separation, the “elastic stiffness” of the linear part of this curve, maximum tensile strength and strain at that point, as well as energy absorption.

The first difficulty in processing the data is to determine the cross-sectional area of the strip specimen, which is not homogeneous throughout its length. This area, if measured at or near the wound line during early healing, is enlarged by edema and contains few stress-bearing elements. Stress should then, to be meaningful from a biological point of view, rather be expressed as force per unit cross-sectional area at the time of wound infliction or per unit collagen content in a standard volume of intact tissue (Holm-Pedersen and Viidik, 1972). These principles enable calculations of whole stress-strain curves; the second option should be preferred in long-time studies with concomitant maturation changes in the intact tissue. Stress should be expressed in percent of that for strip specimens of intact tissue with the same dimensions, when relative healing rates in different tissues are analyzed (see Section 3.1); the disadvantage in this case is that only a few parameters, not the whole curve, can be studied. The same percentage values can be calculated from force (load) values; for healing wounds in tissues consisting primarily of nonconnective tissue components, complete load-strain curves should be calculated (Fogdestam and Gottrup, 1980).

The second difficulty can arise when defining the specimen length segment on which to calculate strain. It can be avoided by selecting a short jaw space (less than 4 mm) in order to measure deformation only in the wound area (Fogdestam, 1981; Gottrup, 1980; Holm-Pedersen and Viidik, 1972). Information will, however, be lost about the weakest point in the "biochemically active zone" if the strength of the healing wound exceeds that of the "intact" surrounding tissue (Gottrup, 1981c). In such cases, well-defined segments of long strip specimens should be analyzed in a video dimensional analyzer (Woo, 1982) with the window set for different marked segments in consecutive reruns of the recording.

The inherent weakness in using cut strip specimens is that, except for the rare case of almost parallel-fibered tissue, some fibers will be cut and thus not be stress-bearing during testing, especially when rather long jaw spaces are used (Viidik, 1978, 1979). This effect is pronounced for intact tissues with regular arrangement of fiber meshwork, while it is negligible during early wound healing, when the fiber arrangement is almost random (Table 19.1) (Nilsson, 1982a). For comparative purposes, such as estimation of relative healing speed, this systematic error is minor compared with the inherent limitations of previously discussed techniques.

### 2.2.2. THE TESTING OF MEMBRANE SPECIMENS

Some of the drawbacks of using strip-shaped specimens could in theory be avoided by testing a circular specimen containing the wound, safely away from the circular edge. Such a technique has been described by van Noort et al (1978). The healing wound with surrounding tissue is mounted as a membrane sealing a chamber, the pressure in which is increased until the membrane ruptures. It functions thus like a blow-out technique *in vitro*. This method is

TABLE 19.1. Mechanical parameters for strip specimens of rabbit lateral abdominal wall.

	Tensile strength	Relative energy absorption	Maximum strain	Elastic stiffness
<b>Intact strip<sup>a</sup></b>				
2 mm wide	1.0	1.0	1.0	1.0
4 mm wide	1.6	2.6	1.7	1.1
8 mm wide	2.1	4.8	2.5	2.1
<b>Intact strip<sup>b</sup></b>				
10 mm long	1.0	1.0	1.0	1.0
20 mm long	1.0	0.7	0.7	1.2
<b>Wound strip<sup>c</sup></b>				
4 mm wide	1.0	1.0	1.0	1.0
8 mm wide	1.0	1.0	1.1	1.4

Note: Parameters were calculated from data in Nilsson, 1982a.

<sup>a</sup>Relative to 2 mm wide, 10-mm-long strips.

<sup>b</sup>Relative to 10 mm long, 4-mm-wide strips.

<sup>c</sup>Relative to 4 mm wide, 10-mm-long strips, 6th day.

basically the same as that used by Dick (1951) for measuring skin rheology and has the same limitations (for discussion, see Viidik, 1973). It attempts also to apply membrane stress analysis to membranes containing abdominal wall wounds; this could yield a more detailed description but would require accurate measures for the radius of the membrane curvature and thickness for each pressure increment. These membranes are, however, anisotropic and inhomogeneous and neither have constant thickness nor expand spherically. It is doubtful whether the more detailed data recorded by this method compared to the blow-out method allow more comprehensive interpretation when the necessary approximations are taken into account.

### 3. Functional Properties of Healing Soft Tissue Wounds

#### 3.1. Comparison of Healing in Different Soft Tissues

Formation of a scar consisting of dense connective tissue is the most common outcome when an incisional wound heals in a soft tissue. The basic properties of this scar probably do not vary significantly from one tissue to another. One factor that could modulate the healing process is the degree of vascularization. The functional properties of tissue—scar—tissue complexes, on the other hand, vary considerably, depending on the properties of the original tissue. The healing rate, when related to the mechanical properties of the original tissue, is therefore very different from tissue to tissue. This is shown in Figure 19.3, where the tensile strengths of healing wounds in different tissues are calculated as percentages of the strengths of the respective intact tissues. It should especially be noted that the relative healing rate varies from very rapid to

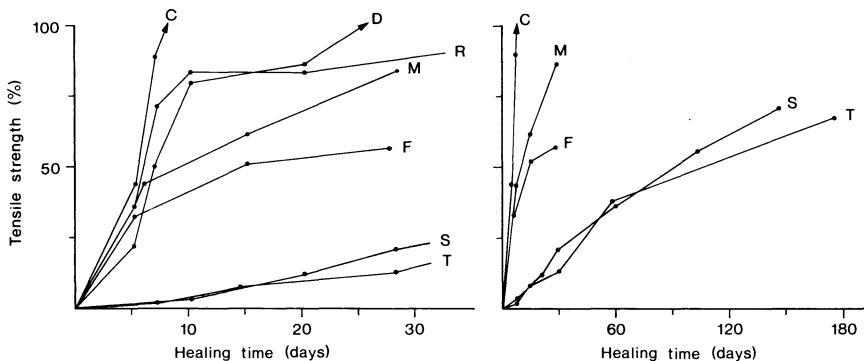


FIGURE 19.3. Relative healing rates for linear incisional wounds in different tissues, the tensile strength being calculated in percent of that of the respective intact tissues: C: Corpus ventriculi (rat), D: Duodenum (rat), R: Rumen ventriculi (rat), M: Lateral abdominal wall muscle (rabbit), F: Fascia (linea alba, rabbit), S: Skin (dorsal thoracic, rat), and T: Tendon (peroneus brevis, rabbit). (Data calculated for C, D, and from Gottrup, 1980; for M and F, from Nilsson, 1982c; for S, from Fogdestam, 1980 and Holm-Pedersen and Viidik, 1972; for T, from Hirsch, 1974).

TABLE 19.2. Tensile strength and relative energy absorption for healing wounds at day 20 in various tissues (expressed in units of values for intact tissues).

Tissue	Tensile strength	Relative energy absorption
Skin	0.12	0.05
Tendon	0.10 <sup>a</sup>	NA <sup>b</sup>
Fascia	0.55 <sup>c</sup>	0.55 <sup>c</sup>
Muscle	0.76 <sup>c</sup>	0.63 <sup>c</sup>
Duodenum	0.86	0.93
Rumen ventriculi	0.83	0.67
Corpus ventriculi	1.29	0.86

Note: Strength and absorption were calculated for skin, from Fogdestam (1980); for tendons, from Hirsch (1974); for fascia and muscle, from Nilsson (1982c); and for duodenum and ventriculus, from Gottrup (1980).

<sup>a</sup>Extrapolated from values for 14 and 28 days of healing.

<sup>b</sup>Data not available.

<sup>c</sup>Extrapolated from values for 15 and 28 days of healing.

rather slow. The more informative parameter, relative energy absorption, which measures the energy required to break the wound, shows a somewhat different pattern (Table 19.2). It can be seen from the comparison in this table that the relative energy absorption (up to the point of rupture in the wound or adjacent tissue) cannot be predicted from tensile strength data; although there

is a certain linear correlation between these two parameters, it cannot be used for extrapolation from one tissue to another.

### 3.2. Healing in Specific Tissues

#### 3.2.1. TENDONS AND LIGAMENTS

These tissues are composed of dense, parallel-fibered connective tissue, in which fibers of Type I collagen dominate. They possess the highest tensile strength of all soft tissues in the body, since the collagen is the strongest fibrous protein and since these fibers are arranged in parallel with each other and in the direction of the muscle forces acting upon them. It could therefore be anticipated that any tissue with a geometric configuration inferior to this arrangement—e.g., the less regular fiber architecture in scar tissue—would have lower tensile strength. This is also demonstrated by the development of strength in healing tendon wounds (Figure 19.3). The scar tissue, which always and especially during early healing exhibits some “unscheduled” waviness and random pattern (Frank et al, 1983; Hutton and Ferris, 1984), does not achieve a strength comparable to that of intact tendons even after prolonged periods of healing. The same is true for ligament wounds, for which Frank et al (1983) recorded a strength amounting to about 60% of that of intact ligaments after 98 days of healing and 70% after 280 days. There is also, after this considerable healing time, a substantial amount of Type III collagen present, and the highly organized fiber pattern of normal tissue has not been fully restored.

Fibroblasts, derived from undifferentiated perivascular mesenchymal cells, are in most tissues the main “wound healers” with respect to connective tissue component synthesis; the mature fibrocytes are considered to play no significant role in this process. There is, however, still a controversy regarding the cell types participating in tendon wound healing, especially the role of cells in the surrounding sheath and thereby the possible desirability of adhesion formation (for discussion, see Hutton and Ferris, 1984). None of the hypotheses, however, rule out that the main “wound healers” are derived from perivascular mesenchymal cells, since all the implicated tissues contain small blood vessels.

The main role of tendons is to transmit forces and that of ligaments to resist them. Attention should, therefore, be focused on the influence of stresses and immobilization on the development of mechanical strength in healing wounds in these tissues. This is emphasized by the observation (Hutton and Ferris, 1984) that the tissue elements become successively more organized after the initial phases of healing following partial tenotomy. The fibroblast-like cells begin to orient in the direction of the force line after 14 days of healing; this process seems to be completed after eight weeks, when the cells resemble fibrocytes. Early weight bearing during healing after repaired tenotomy, however, results in rupture in most cases—but also in increased strength later on, if the continuity is not broken. Tensile strength increase during healing is enhanced by controlled passive motion; the strength (ultimate load) increases to

35% of the value for intact tendons after 84 days, while complete immobilization results in an increase of only 21% (Woo et al, 1981). The rate is enhanced already during the first weeks after surgery (Gelbermann et al, 1982).

While the immobilization of a healing tendon does not prevent some tension derived from muscle contraction, a ligament bridging an immobilized joint is subjected to no stresses at all. Vailas et al (1981) reported that the separation force for healing medial collateral ligaments was 111% of that for intact ligaments after six weeks of exercise (following two weeks in cast), while eight weeks in cast resulted in an force increase of only 54%. However, they also reported that the total collagen content of the ligament decreased after immobilization (from 1.00 to 0.84) and increased after exercise (to 1.97). Using these relative values for cross-sectional area, the separation "stress" can be recalculated to 56% and 64%, respectively, figures in the same order of magnitude as reported for tendons. The improvement after exercise seems thus to be derived here mainly from an excessive deposition of collagen in the scar area.

### 3.2.2. SKIN

A large number of wound-healing studies have used linear incision wounds in the skin, i.e., dermis, (for review, see Holm-Pedersen, 1973, and Fogdestam, 1980) as the model, especially as a general model for wound healing and the elucidation of the factors influencing the healing rate. The description of the classical stages of wound healing (see Section 1.1) is thus derived mainly from skin wound-healing experiments. It could be anticipated, when taking into account the three-dimensional meshwork of collagen fibers and the higher degree of vascularization in dermis, compared to tendons, that the rate of wound strength increase would be considerably higher. This is, however, not the case (see Figure 19.3). The key point is how the newly formed collagen fibers connect with those at the wound edges; this is still an open question.

The stress-strain curve for the scar tissue is the classical one for soft connective tissues, being sigmoid with a more or less linear part between the toe part and the leveling off when the tissue starts to rupture. This curve is always lower and to the right of that for intact skin; the extensibility is higher and the ultimate strength is lower (Viidik, 1975; Holm-Pedersen and Viidik, in preparation). The reports claiming that scar tissue ultimately becomes stiffer have used control tissue from growing animals taken at the time of wounding and not when sampling wounds for testing, thus disregarding that skin becomes stiffer with maturation and aging.

A number of factors impair wound healing (see Section 1.2.1 and, for review, Zederfeldt, 1980), while only a delay of primary closure accelerates the gain of mechanical strength (Fogdestam, 1981). With this technique, the wounds are kept open (under cover of surgical dressing) for three to six days before suturing, enabling the formation of granulation tissue surfaces that "glue" together (as well as preventing infection). The strength of a 10-day-old wound, which has been open initially for 3 days, is 26% higher than that of one closed

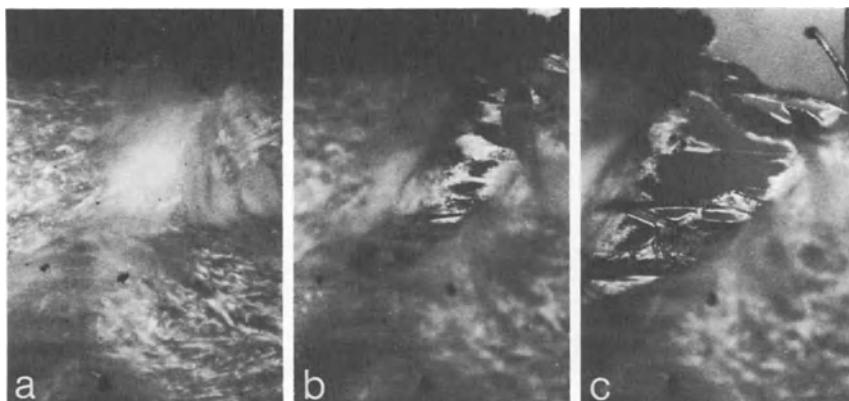


FIGURE 19.4. (a) A three-week-old scar in skin; the tissue is cut perpendicularly to the scar, with epidermis at top. Tension is applied from left to right. (b) The scar starts to rupture. (c) Most of the scar has broken; a few connective tissue fiber bundles still stretch across the gap (From background material to Holm-Pedersen and Viidik, 1972).

immediately after incision. This difference is 52% 60 days after wounding and still significant after 120 days of healing. The factor(s) eliciting this acceleration are not known; the influence of another wound can be ruled out since the results are similar in paired and nonpaired experiments.

The failure pattern for a skin wound is shown in Figure 19.4. It is more abrupt in early stages of wound healing and less gradual in older animals than in young ones at the same stage of healing (Holm-Pedersen and Viidik, 1972). The failure occurs in the wound itself during early healing and later on, when the wound/scar tissue has gained strength but the skin–scar–skin complex is weaker than intact tissue, probably in the border zone between the scar and the adjacent tissue. This zone, which is biochemically active even after 60 days of healing, is about 5 mm (Danielsen and Fogdestam, 1981).

### 3.2.3. FASCIA AND MUSCLE

Abdominal surgical procedures always create incisional wounds through fascia and muscle. Wound rupture in the early postoperative phase and hernia formation later on are not infrequent complications. Various techniques, with emphasis on incision types (with reference to differences in muscle pull in the postoperative phase) as well as suture materials and techniques, have therefore been tried in order to optimize this healing process (for review, see Nilsson, 1983).

Wounds tend to gain relative strength more rapidly when the incision is placed in muscular areas of the abdominal wall than in fascia (e.g., linea alba) (compare Figure 19.3). A somewhat elastic suture material (synthetic poly-

mers) that permits a more yielding fixation of the wound than stainless steel sutures seems to stimulate the gain of mechanical strength (Nilsson, 1982b). While a certain tension (or motion) thus promotes wound healing, both excessive tension (Nilsson, 1982b) and relaxation (Thorngate and Ferguson, 1958) seem to impair wound healing from a mechanical point of view. The gain of strength is further enhanced when each muscle layer in the abdominal wall is sutured separately (Nilsson, 1981).

The extent of the biochemically active zone around this type of wound is not known. The occurrence of ruptures outside the visually discernible wound area already after 15 days of healing and at lower strength than that of intact tissue (Nilsson, 1982c) suggests that a rather pronounced zone is present.

### 3.2.4. DIGESTIVE TRACT

While wound rupture in the abdominal wall is a serious but easily recognized complication to surgery, an intestinal anastomosis insufficiency is more insidious, though not less serious. It is thus essential for clinical practitioners to have information from experimental investigations, which can analyze the influence of various factors on the healing process separately. Most investigations in this field have used the blow-out technique, which yields no other information than the bursting pressure of a geometrically complicated structure (see Section 2.1.1 and Figure 19.1). The design of experiments on these tissues with the strip specimen technique is, however, not without complications.

The weakest point in a digestive tract strip specimen, cut at a right angle to its healing wound, moves rapidly from the wound line (Figure 19.5a and b) into the biochemically active zone around it and then into intact tissue (Figure 19.5c and d). This zone stretches 3–4 mm lateral to the incision line (Gottrup, 1981a). The healing, measured as gain of mechanical strength relative to that of the intact tissue, is very rapid in these tissues (compare Figure 19.3). More than 50% of the ruptures occur outside the wound area (i.e., more than 4 mm lateral

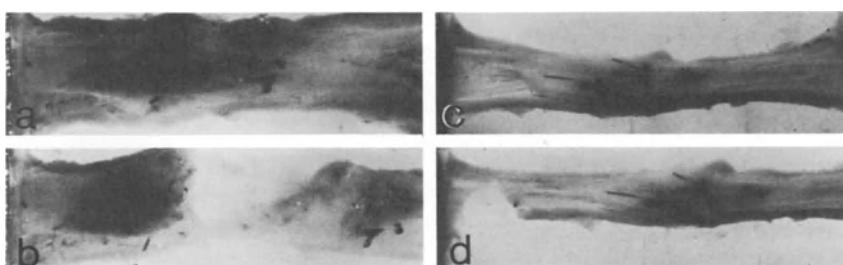


FIGURE 19.5. Same as Figure 19.4, with healing scar in corpus ventriculi. (a) and (b) show the scar tissue rupturing after five days of healing. (c) and (d) show that the rupture occurs outside the scar area (in the middle, with cut suture ends showing), i.e., the scar is stronger than the intact tissue after 40 days. (Modified from Gottrup, 1981c).

to the incision line) already after 10 days of healing (Figure 19.5c and d) (Gottrup, 1981c).

The experimental protocol must thus be designed so that either only the wound tissue (the specimen clamped with a jaw space of about 2–3 mm) or the whole complex consisting of the wound, adjacent tissue, and intact tissue (jaw space of at least 10 mm) is included into the strip specimen. In the first-mentioned alternative, information is derived regarding the wound/scar tissue only. Both approaches should be used, if data about the functional properties of the healing organ are needed. The weakest point, as recorded with the last-mentioned alternative (compare Figure 19.5), is consistent with that found with the blow-out technique (compare Figure 19.1a; Gottrup and Viidik, to be published).

### 3.2.5. OTHER TISSUES

The information on the mechanics of healing wounds in other tissues is meager. The healing in urinary bladder is rapid, as the strength of intact tissue is reached already after 14 days (Hastings et al, 1975); this tissue has a low collagen concentration like that of the digestive tract. Most research on healing in blood vessels has been focused on intimal regeneration, while little is known about healing in the media and adventitia, the stress-resisting parts of the vessel wall (Leaper, 1984).

## 4. Perspectives for Future Research

Today, we have reasonably detailed knowledge of the mechanical properties of healing wounds in a number of soft tissues, while information is lacking on that in other tissues, most notably blood vessels. Future experimental and clinical wound-healing research will probably focus on the nature of the regulating factors, their structure and mechanisms of action. When available for experimental and clinical use they will accelerate the healing of wounds but most probably not alter the qualitative mechanical properties of healing wounds. Our knowledge of these properties is, however, not complete today.

Our most detailed knowledge concerns the breaking and/or tensile strength of the healing wounds, and we know somewhat less about the stress-strain and load-strain behaviors of these dynamic tissues. On the other hand, there is very little information available about their mechanical properties in the range, which is probably the one of interest for normal functional behavior of the intact tissue–wound tissue–intact tissue complexes, i.e., the toe parts of the curves. The analytical approach of Fung (1972), utilizing the linear relationship between stress and corresponding “tangent modulus,” should be employed (for discussion, see also Viidik, 1980). This approach has been proved to be a powerful tool for the characterization of other “fragile” tissues (Viidik et al, 1982).

## Acknowledgments

The research of the authors and their co-workers reported here was supported by the Danish Medical Research Council, grants no. 512-213, 512-715, 512-10126, 512-20046, 12-1657, and by funds from the Medical Schools of the Universities of Aarhus (Denmark) and Göteborg (Sweden).

This chapter is based on, besides the literature cited above, the doctoral theses of Poul Holm-Pedersen (1973), Ingemar Fogdestam (1980), Tove Nilsson (1983), and Finn Gottrup (1984) from this department.

## References

- Adamsons, R.J., Musco, F., and Enqvist, I.F. 1966. *Surg. Gynecol. Obstet.* 123: 515-521.
- Botsford, T.W. 1941. *Surg. Gynecol. Obstet.* 72: 690-697.
- Chlumsky, V. 1899. *Beitr. Klin. Chir.* 25: 539-600.
- Danielsen, C.C., and Fogdestam, I. 1981. *J. Surg. Res.* 31: 210-217.
- Danielsen, C.C., and Gottrup, F. 1981. *Eur. Surg. Res.* 13: 194-201.
- DeHaan, B.B., Ellis, H., and Wilks, M. 1974. *Surg. Gynecol. Obstet.* 138: 693-700.
- Dick, J.C. 1951. *J. Physiol. (London)* 112: 102-113.
- Fogdestam, I. 1980. *Delayed primary closure.* Thesis, University of Göteborg, Göteborg.
- \_\_\_\_\_. 1981. *Surg. Gynecol. Obstet.* 153: 191-199.
- Fogdestam, I., and Gottrup, F. 1980. Pages 363-371 in *Biology of collagen*, A. Viidik and J. Vuust, eds. Academic Press, London.
- Frank, C., Woo, S.L-Y., Amiel, D., Harwood, F., Gomex, M., and Akeson, W.H. 1983. *Am. J. Sports. Med.* 11: 379-389.
- Fung, Y.C. 1972. Pages 181-208 in *Biomechanics: its foundations and objectives*, Y.C. Fung, N. Perrone, and M. Anliker, eds. Prentice-Hall, Englewood Cliffs, N.J.
- Gelbermann, R.H., Woo, S.L-Y., Lothringer, K., Akeson, W.H., and Amiel, D. 1982. *J. Hand Surg.* 7: 170-175.
- Glynn, L.E., ed. 1981. *Tissue repair and regeneration.* Elsevier, North-Holland, Amsterdam.
- Gottrup, F. 1980. *Am. J. Surg.* 140: 296-301.
- \_\_\_\_\_. 1981a. *Am. J. Surg.* 141: 222-227.
- \_\_\_\_\_. 1981b. *Acta Chir. Scand.* 147: 363-369.
- \_\_\_\_\_. 1981c. *Am. J. Surg.* 141: 706-711.
- \_\_\_\_\_. 1984. *Danish Med. Bull.* 31: 32-48.
- Harvey, S.C. 1929. *Arch. Surg.* 18: 1227-1240.
- Hastings, J.C., Van Winkle, Jr., W., Barker, E., Hines, D., and Nichols, W. 1975. *Surg. Gynecol. Obstet.* 140: 933-937.
- Hirsch, G. 1974. *Acta Orthop. Scand. Suppl.* 153.
- Holm-Pedersen, P. 1973. *Studies on healing capacity in young and old individuals.* Munksgaard, Copenhagen.
- Holm-Pedersen, P., and Viidik, A. 1972. *Scand. J. Plast. Reconstr. Surg.* 6: 24-35.
- Hunt, T.K. 1979. Pages 68-168 in *Fundamentals of wound management*, T.K. Hunt and J.E. Dunphy, eds. Appleton-Century-Crofts, New York.
- Hunt, T.K., and Van Winkle, Jr., W. 1979. Pages 2-67 in *Fundamentals of wound*

- management, T.K. Hunt and J.E. Dunphy, eds. Appleton-Century-Crofts, New York.
- Hunt, T.K., Knighton, D.R., Thakral, K.K., Andrews, W., and Michaeli, D. 1984. Pages 3–19 in *Soft and hard tissue repair*, T.K. Hunt, R.B. Heppenstall, E. Pines, and D. Rovee, eds. Praeger, New York.
- Hutton, P., and Ferris, B. 1984. Pages 286–296 in *Wound healing for surgeons*, T.E. Bucknall and H. Ellis, eds. Ballière Tindall, London.
- Irvin, T.T. 1981. *Wound healing: principles and practice*. Chapman and Hall, London.
- Knighton, D.R., Oredsson, S., Banda, M., and Hunt, T.K. 1984. Pages 41–49 in *Soft and hard tissue repair*, T.K. Hunt, R.B. Heppenstall, E. Pines, and D. Rovee, eds. Praeger, New York.
- Leaper, D.J. 1984. Pages 221–240 in *Wound healing for surgeons*, T.E. Bucknall and H. Ellis, eds. Ballière Tindall, London.
- Leibovich, S.J., and Ross, R. 1975. *Am. J. Pathol.* 78: 71–100.
- \_\_\_\_\_. 1976. *Am. J. Pathol.* 84: 501–503.
- Muller, H.K. 1981. Pages 145–175 in *Tissue repair and regeneration*, L.E. Glynn, ed. Elsevier, North-Holland, Amsterdam.
- Nelsen, T.S., and Anders, C.J. 1966. *Arch. Surg.* 93: 309–314.
- Niinikoski, J. 1980. *Med. Biol.* 58: 303–309.
- Nilsson, T. 1981. *Acta Chir. Scand.* 147: 399–403.
- \_\_\_\_\_. 1982a. *J. Biomech.* 15: 131–135.
- \_\_\_\_\_. 1982b. *Scand. J. Plast. Reconstr. Surg.* 16, 101–105.
- \_\_\_\_\_. 1982c. *Acta Chir. Scand.* 148: 251–256.
- \_\_\_\_\_. 1983. *Danish Med. Bull.* 30: 394–407.
- Ross, R., and Vogel, A. 1978. *Cell* 14: 203–210.
- Sandblom, P., Petersen, P., and Muren, A. 1953. *Acta Chir. Scand.* 105: 252–257.
- Thorngate, S., and Ferguson, D.J. 1958. *Surgery* 44: 619–624.
- Vailas, A.C., Tipton, C.M., Matthes, R.D., and Gart, M. 1981. *Connect. Tissue Res.* 9: 25–31.
- van Noort, R., Black, M.M., Greaney, M.G., and Irvin, T. 1978. *Eng. in Med.* 7: 217–221.
- Viidik, A. 1973. Pages 157–189 in *Front. Matrix Biol.*, Vol. 1. L. Robert, and B. Robert, eds. Karger, Basel.
- \_\_\_\_\_. 1975. Pages 155–169 in *Annals of Estonian Medical Association* 1975, J. Kaude, ed. Eesti Post, Malmö.
- \_\_\_\_\_. 1978. *Verh. Anat. Ges.* 72: 75–89.
- \_\_\_\_\_. 1979. Pages 75–113 in *Progress in biomechanics*, N. Akkas, ed. Sijthoff & Noordhoff, Alphen aan den Rijn.
- \_\_\_\_\_. 1980. Pages 237–255 in *Biology of collagen*, A. Viidik and J. Vuust, eds. Academic Press, London.
- Viidik, A., Danielsen, C.C., and Oxlund, H. 1982. *Biorheology* 19: 437–451.
- Vogel, A., Ross, R., and Raines, E. 1980. *J. Cell Biol.* 85: 377–385.
- Wijn, P. 1980. *The Alinear viscoelastic properties of human skin in vivo for small deformations*. Thesis, Katholieke Universiteit Nijmegen, Nijmegen.
- Woo, S.L-Y. 1982. *Biorheology* 19: 385–396.
- Woo, S.L-Y., Gelbermann, R.H., Cobb, N.G., Amiel, D., Lothringer, K., and Akeson, W.H. 1981. *Acta Orthop. Scand.* 52: 615–622.
- Zederfeldt, B. 1980. Pages 347–362 in *Biology of collagen*, A. Viidik and J. Vuust, eds. Academic Press, London.

## Part IV Circulation

# 20

## Biomechanics of the Microcirculation

B.W. ZWEIFACH

### 1. Network Design

The microvascular network consists of a succession of small arteries that branch to give rise to smaller continuations until vessels are reached with diameters approximating the dimension of red blood cells, the so-called capillaries. The myriad of capillaries interconnect freely and then become confluent to form progressively wider vessels, the small veins. This hierarchy of vessels from the feeder arterioles through to the analogous venules are collectively referred to as the microcirculatory system. Most models of such networks are depicted as a series of regular bifurcations, despite the fact that numerous departures are encountered in representative microvascular networks, such as in skeletal muscle or mesentery.

Considerable attention has been given to the effect that network deployment—branch orders, vessel diameter, length, number of branchings—has on hemodynamic variables such as pressure and flow distribution (Intaglietta and Zweifach, 1974; Zweifach and Lipowsky, 1984). The presence of numerous side-arm branches has made it necessary to consider the restrictive effect of entry conditions on pressure and flow attenuation at such junctional configurations. The heterogeneity of the microvascular population, particularly with regard to parent-to-daughter diameter ratios, results in an uneven partition of the red blood cell stream at branch points. The variable reduction in red cell hematocrit in the terminal ramifications of the arterioles is determined by physical factors that in turn influence the pressure gradient through the network.

Mathematical analyses for these physical variables have been developed for networks with varying degrees of complexity, beginning with the almost idealized deployment in the omentum, to the more densely arrayed network in the intestinal mesentery and the complex architecture of skeletal muscle beds.

#### 1.1. Patterned Versus Random Networks

The network of capillaries that makes up the terminal vascular bed appears to have a random distribution in many tissues. In skeletal muscle or in cardiac

muscle, because of the parallel alignment of the muscle fibers, the capillaries themselves have a regular arrangement with reasonably uniform intercapillary spacing. Other tissues fall in between these two extremes.

On the basis of the organization of the microcirculation in several mesenteric tissues, Zweifach (1957, 1974) proposed that two physical features underlie the basic functional attributes of the microcirculation in general: (1) The majority of capillaries are distributed as side-arm branchings along the course of the terminal end arteriole, which becomes progressively narrower until it is of capillary dimensions. This structural alignment allows these arterioles to serve as centrally located distributive channels for the capillary network proper. (2) Smooth muscle cells become less conspicuous along the terminal extensions of the arterioles, but the muscular coat terminates at the junctional portion of the side branches to form muscular sphincters. The furthermost extensions of the arterioles, the capillaries are nonmuscular, endothelial tubes.

A substantial percentage of the precapillary vessels are distributed as side branches of the terminal arterioles and not as dichotomizations of the parent trunk. These side-arm offshoots (about  $10\text{--}15 \mu\text{m}$  wide) are operationally in parallel with the main arteriolar-to-venular pathway. Since the junctional configurations are endowed with smooth muscle, they can function as controls for the distribution of blood into groups of capillaries. In most tissues, each side branch breaks up into four to six separate capillary-sized vessels.

Several basic approaches have emerged in an attempt to elucidate the relationships between microvascular pressure and flow distributions and network topography: (1) lumped-parameter models that describe microhemodynamics in terms of fundamental microvascular divisions, i.e., arterioles, capillaries, and venules; (2) distributive geometric models that rely on idealizations of a specific network configuration; and (3) discrete-network analyses of specific topographical patterns. These studies, varying in both complexity and sophistication, have been used to examine microhemodynamics in relation to the data acquired by intravital microscopy.

## 1.2. Lumped Parameter Models

Granger and Shepherd (1973) employed a greatly simplified representation of the anatomical features of the microvascular network (arterioles, precapillary sphincters, capillaries, and venules) to examine the feedback mechanisms regulating arteriolar resistance, capillary density, and intracellular oxygen tensions in the surrounding parenchymal tissue. By incorporating into the model the mechanisms that may modulate resistance through arteriolar vasodilation or recruitment of additional in-parallel capillaries, the model demonstrates the relationship between decreased oxygen levels in the tissue and the attendant demand for increased network perfusion.

Models can also be used to determine the vascular components involved in the response to specific microvascular regulatory maneuvers. Nellis and Zweifach (1977) have characterized the microvasculature in terms of an anal-

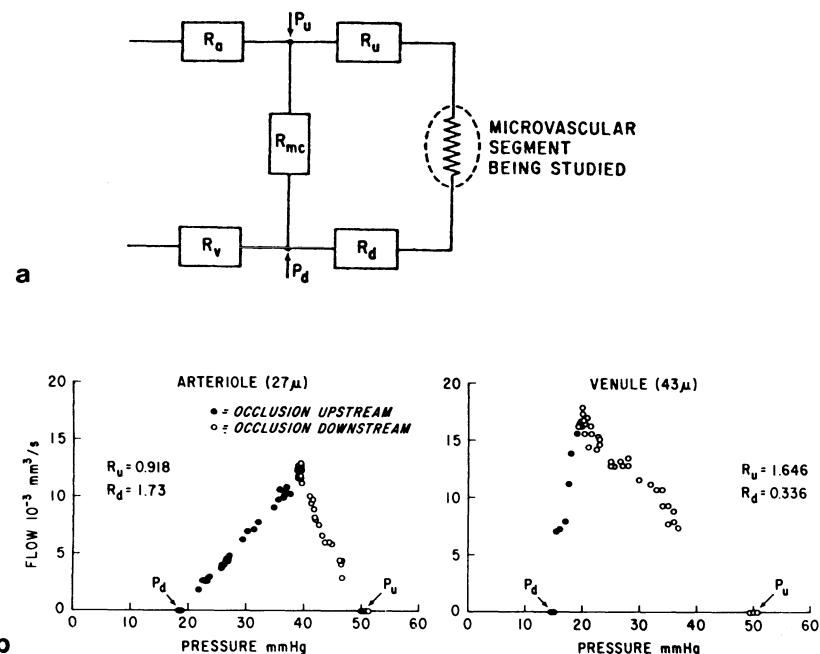


FIGURE 20.1. (a) Model of a microvascular segment including a variable resistance created by an occluding needle.  $P_u$  and  $P_d$  are upstream and downstream source pressures.  $R_u$  and  $R_d$  are equivalent resistances for the vascular segments between the two source pressure points.  $R_a$  and  $R_v$  are the resistances for the arterial feeding and venous effluent vessels.  $R_{mc}$  is the fraction of the total resistance for the network between  $a$  and  $v$ . (b) Pressure ( $P$ ) and flow ( $Q$ ) were monitored during the rapid occlusion, first upstream and then downstream. Log-log plot of arteriolar and venular resistance values vs. flow. Exponential nature of  $R$  distribution is clearly shown (from Nellis and Zweifach, 1977, by permission of the American Heart Association, Inc.).

ogous electrical circuit composed of six generalized resistance segments (Figure 20.1a). This procedure utilized the analogy between the mechanical manipulation of flow (vessel occlusion) and the creation of an open circuit in an electrical network. From the flow-versus-pressure slopes for individual microvessel data (see Figure 20.1b), it is possible to determine the effective upstream and downstream resistances on either side of an occlusion probe that was used to interrupt flow.

In physical terms, the upstream and downstream resistances reflect the relative deployment of a given microvessel with respect to its contiguous vessels. For example, at the level of the true capillaries with numerous parallel pathways, occlusion of any capillary has a relatively minor effect on intravascular pressure and hence upstream and downstream resistances are fairly large. However, in larger microvessels that have fewer effectively parallel neighbors (e.g., the 40–60 μm arterioles), occlusion of the vessel yields rela-

tively large increases in pressure proximal to the occluding probe. Similarly, large decreases in pressure distal to the probe reflect the fact that a reduction in arteriolar pressure toward venous values has occurred.

### 1.3. Distributive Models

Analysis of microhemodynamics by formulating detailed network models has received considerably more attention than other techniques. The general approach has been to characterize the branching pattern of the network in terms of a schematic representation of the overall hierarchy of microvessels or of specific portions of a given network. The resistance of an individual microvessel of circular cross section may be estimated from the Poiseuille-Hagen equation (Hagen, 1839; Poiseuille, 1840). Realistic representations of the distribution of geometric parameters affecting microhemodynamics have been examined for rabbit omentum (Intaglietta and Zweifach, 1971). In this study the network was characterized by the distribution of hydraulic hindrance from arteriole to venule as a succession of serial resistances. Each network is composed of a specified number of parallel vessels that make up the successive branching orders in the network.

### 1.4. Discrete Network Models

In contrast to the preceding models of the microvasculature, several studies have performed network computations based on comparatively exact network topography (Lipowsky and Zweifach, 1974, 1977; Smaje et al, 1970; Gore and Bohlev, 1977). In this approach the techniques of electrical network analysis are used to compute the distribution of pressure and flow throughout a discrete network of microvessels (Davis and Bohlev, 1985). Measurements of the lengths and diameters of all microvessels in the field permitted the schematic description of the network, in which each junction or branch point (node) is assigned a number. Each microvessel was thus characterized as an element of the network running from node  $i$  to node  $j$ , with constant diameter and length.

The distribution of intravascular pressures throughout the interior of the network was computed by using the electrical engineering technique of node-voltage analysis. This procedure applies the laws of mass-flow conservation at each node point, which dictate that the sum of all flows into a node (taken as positive) and all flows exiting a node (taken as negative) must equal zero,  $\Sigma Q_{ij} = 0$ .

The expression of hemodynamic processes on the scale of the entire network has been studied in both the arterioles and venules of mesentery (Lipowsky et al, 1981) and the capillary bifurcations of cremaster muscle (Klitzman and Johnson, 1982). These studies examined the conservation of red cell flux at bifurcations as a function of the relative flows in the daughter branches. Specifically, conservation of volumetric flow at a bifurcation must satisfy the

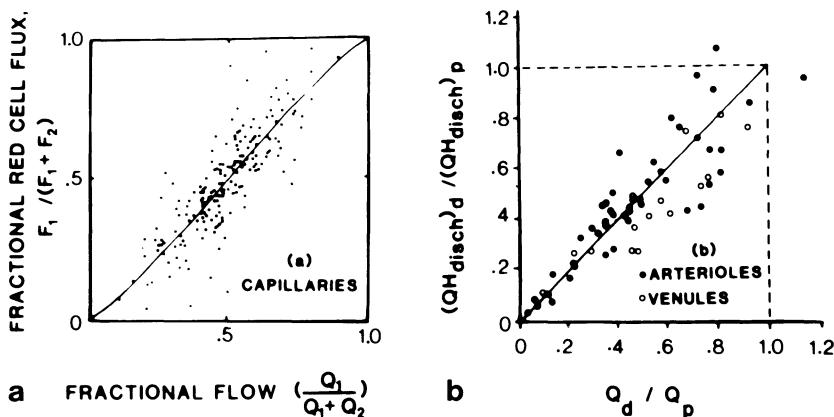


FIGURE 20.2. Red cell flux at bifurcation. (a) Ratio of red cell flux from daughter to parent vs. bulk flow in capillaries of cremaster muscle. (b) Volumetric flux ratio as product of bulk flow and discharge hematocrit between daughter ( $d$ ) and parent ( $p$ ) vs. ratio of their bulk flows of cells plus plasma ( $Q_d/Q_p$ ), in arterioles of mesentery (from Zweifach and Lipowsky, 1984).

following relationships for bulk flow of cells plus plasma in the daughter and parent vessels,  $Q_d$  and  $Q_p$ , respectively, as well as for the volumetric flux of red cells,  $Q_{RBC}$ :

$$Q_{d1} + Q_{d2} = Q_p, \quad (20.1)$$

and therefore

$$Q_{RBC_{d1}} + Q_{RBC_{d2}} = Q_{RBC_p}, \quad (20.2)$$

where the volumetric flux of red cells is given by the product of bulk flow and the discharge hematocrit ( $QH_D$ ) for each branch. Figure 20.2 presents representative plots of these ratios of red cell flux versus bulk flow from parent to daughter branches for the capillaries of the cremaster network in Figure 20.2a, and the arterioles and venules of the mesenteric network in Figure 20.2b. The data indicate that in cases where parent vessels feed daughters of comparable size, the effects of plasma skimming do not contribute in any substantial way to a nonuniform distribution of blood throughout successive microvascular divisions.

Network pressure gradients delineate the specific location of the major sites of resistance to flow throughout the ensemble of microvessels. For example, the mesentery shows maximum network resistance in the precapillary microvessels ranging in diameter from 15–35  $\mu\text{m}$ , which is consistent with observations on the arcaded modular structure of this network (Frasher and Wayland, 1972). The omentum, however, shows a maximum pressure gradient at the level of the true capillaries consistent with its nearly dichotomous branching pattern (Intaglietta and dePlomb, 1973). In skeletal muscle beds, a maximum fall in pressure occurs in the region of the small arterioles that run transverse

to the muscle fibers prior to the final ramification at the true capillaries (Smaje et al, 1970; Eriksson and Myrhage, 1972; Fronek and Zweifach, 1974).

### 1.5. Volumetric Flow Rates

Although the arterio-venous (AV) distributions of pressure ( $P$ ) and mean pressure ( $\bar{P}$ ) reflect the relative deployment of network resistances, it is the AV distribution of flow that gives insight into the relative number of microvessels that are effectively in parallel at each division to carry off the total flow of the network. Figure 20.3 presents the AV distributions of intravascular pressures and red cell velocities versus luminal diameter. The latter relationship may be

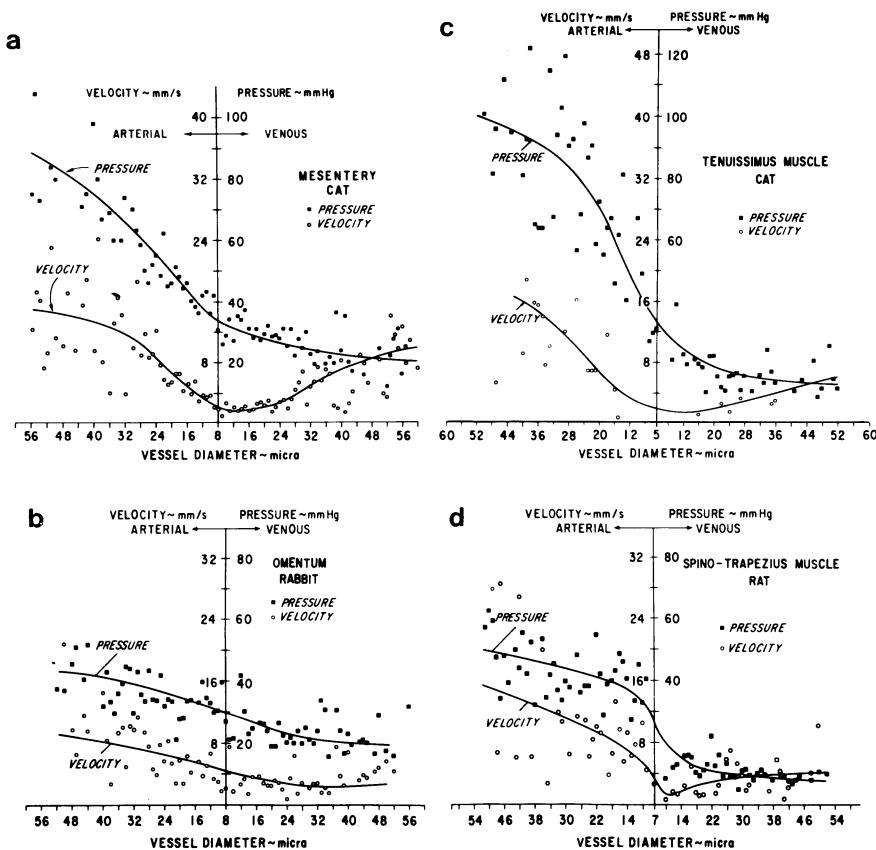


FIGURE 20.3. Arterio-venous distribution of intravascular pressure and red cell velocities as a function of microvessel luminal diameter. Velocity profiles are largely determined by topography of network. Range of pressure variations is greater than that for velocity. (a) Mesentery (Zweifach and Lipowsky, 1977); (b) Omentum (Zweifach and Lipowsky, 1977, by permission of the American Heart Association, Inc.); (c) Tenuissimus (Fronek and Zweifach, 1974); (d) Spinotrapezius (Zweifach et al, 1981).

interpreted as an index of position in the overall hierarchy of microvessels in mesentery (Zweifach and Lipowsky, 1977), omentum (Zweifach and Lipowsky, 1977), spinotrapezius muscle (Zweifach et al, 1981), and tenuissimus muscle (Fronek and Zweifach, 1975). It is evident that each network has a distribution pattern dictated by its topography, overall AV pressure drops, flows, and properties of blood rheology.

One can compute the bulk volumetric flow as  $Q = \bar{V}\pi d^2/4$  and plot its AV distribution. In light of the limited variation of red cell velocity relative to  $d^2$  it would appear that microvessel cross-sectional area dominates the AV distribution of flow, thus yielding a parabolic distribution. The minimum flow appears to fall in the postcapillary microvessels immediately following the true capillaries.

When one calculates the slopes of the smoothed data for the change in flow relative to the successive decrease in vessel diameter,  $d_Q/d_d$ , the commonalities in flow apportionment become more striking. It is significant that, although branching patterns may vary considerably, the rates at which flow varies with respect to the diameters of the constituent vessels of the network are similar for each tissue. The similarities of trends such as these have been analyzed in the context of an optimal network design (Mayrovitz and Roy, 1981) based on the principle of minimum work (Murray, 1926). The hypothesis was advanced that the branching pattern of the microvascular network may be configured to facilitate perfusion of the true capillaries for the least amount of energy expended in overcoming vascular resistance (Mayrovitz and Roy, 1981).

## 1.6. Branch Points and Bifurcations

Variability of hemodynamic parameters throughout the microvascular network has also been examined in light of the flow behavior of blood at a branching site. Most analyses of branch flow processes have centered around bifurcations. In this context, three major characteristics of branch hemodynamics have been explored: (1) attenuation of intravascular pressure, (2) intermittency of flow, and (3) hematocrit variations.

*In-vivo* data are available for the reduction of intravascular pressure at bifurcations in the network for parent arterioles ranging from 50 to 20  $\mu\text{m}$  in diameter (Zweifach, 1974). The ratios of pressure in the parent vessels ( $P$ ) to that in the daughter vessel ( $d$ ) were found to increase substantially with the diminishing ratio of their respective diameters ( $d_d/d_p$ ), rising from 5% to 40% as  $d_d/d_p$  decreased from 0.9 to 0.1. Detailed information on this process has been derived from theoretical analyses (Lipowsky and Zweifach, 1974). It has been shown that the specific geometry of the orifice of the side-arm branch may greatly influence the pressure drop incurred as the flow enters the branch. Irregularities at the orifice may obstruct flow up to 80% and produce a pressure drop on the order of 20 to 30 times greater than for entry sites without a constriction.

Detailed studies of blood cell distributions at bifurcations where branch diameters are comparable to those of capillaries have emphasized the roles of both the balance of hydrodynamic forces and blood cell deformability (both erythrocytes and leukocytes) as determinants of cell concentrations in daughter branches (Schmid-Schönbein et al, 1980; Yen and Fung, 1977). The net hydrodynamic forces that steer a cell into a particular side-arm branch have been shown to be a direct function of the relative ratio of the branch flow (i.e., the sum of the two daughter branch flows) to the parent flow (Yen and Fung, 1978). Observations of the rabbit ear chamber *in situ* (Schmid-Schönbein et al, 1980) have shown that in parent vessels with sparse cell concentrations, the position of the blood cell (either erythrocyte or leukocyte) with respect to the dividing streamline (i.e., radial eccentricity of the cell position) is the principal determinant of which branch the cell may enter. The position of the dividing streamline is in turn affected by the ratio of daughter-to-parent vessel bulk flow.

Attempts to evaluate microvascular pressure-flow relationships in terms of the relative contributions of rheological and vascular factors have formed a recurrent theme in microvascular research. Drawing on a broader base of microcirculatory experimental observations, Lampert (1955) proposed that the summated effects of rheological and topographical factors may be described by defining the total resistance ( $R$ ) to blood flow for a given organ or region of tissue as the product of a geometrical hindrance ( $Z$ ) and effective blood viscosity ( $\eta$ ) such that  $R = Z\eta$ . However, a major difficulty has been to decipher the topography of the constituent microvessels in light of their serial or parallel deployment. Although the hindrance of  $n$  serial elements is computed easily from the electrical analogue as  $Z = nZ_i$  and for parallel vessels as  $Z = 1/\Sigma(1/Z_i)$ , it is not possible to represent actual networks in terms of serial or parallel microvessels in all but simple deployments, such as in omentum and mesentery.

## 2. Blood-Tissue Exchange

Materials carried by the bloodstream have access to the cells of the body only across the rich network of thin-walled capillaries. Exchange of materials between the blood and tissue compartments occurs for the most part by diffusion, as a result of concentration differences on the two sides of the capillary barrier. Diffusion of water molecules in and out of the bloodstream occurs relatively freely in both directions, resulting in considerable exchange but no net volume flow. On the other hand, a bulk transfer of fluid in either direction will be initiated when there is a shift in balance between two opposing forces, the hydrostatic and protein osmotic pressure as conceived by Starling (1896). The operation of this homeostatic balance is well supported by a broad array of experimental evidence.

If one accepts the validity of the Starling constitutive equation for fluid

exchange and attempts a rigorous solution of the problem, he must use the numerical values that represent the underlying assumptions of the formulation. One of the difficulties encountered is that the equation is applicable to a situation in which two fluid compartments are separated by a membrane. In blood-tissue exchange, the compartment immediately surrounding the capillary vessels is a complex gel. This creates a formidable difficulty in attempting to characterize and to obtain measurements from the fluid in the tissue matrix that would be applicable to the Starling relationship.

The basic exchange flows presumed to exist are schematically depicted in Figure 20.4. It is apparent that for the analytical solution of such a problem, the statistical and dynamic properties of the tissue barrier can be established only on the basis of perturbations recorded outside the tissue proper—that is, measurements in the blood and terminal lymph compartments.

To circumvent some of these difficulties and to provide a description of the phenomena of exchange in terms that can be readily defined and measured,

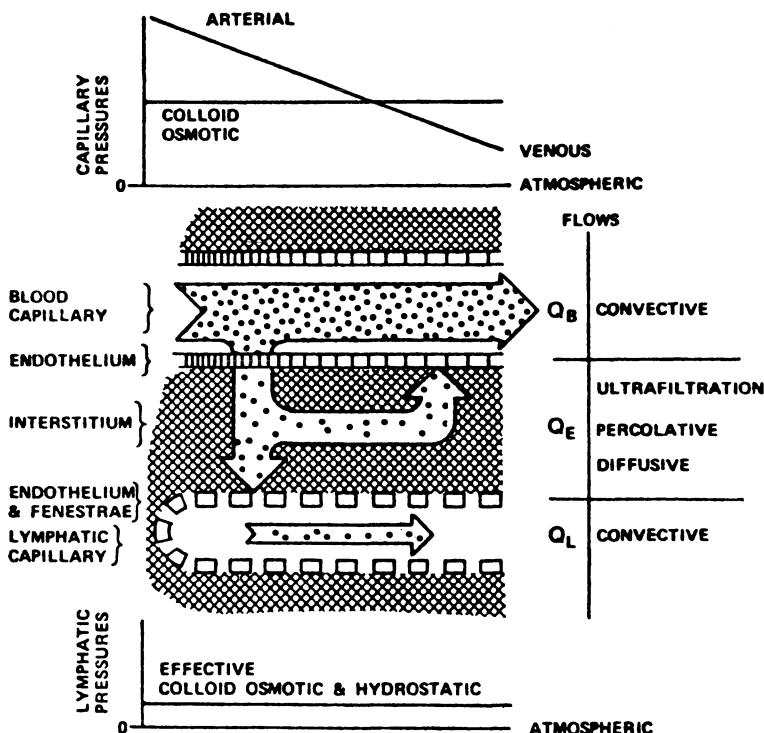


FIGURE 20.4. Diagrammatic representation of the Starling mechanism of fluid balance. Fluid filtered at the arterial portion of the microcirculation is reabsorbed at the venous end, where plasma colloid osmotic pressure is higher than the local blood hydraulic pressure. Lymph flow,  $Q_L$ , is a consequence of an imbalance between filtration and absorption flows (from Intaglietta and Zweifach, 1974).

the Starling hypothesis can be modified and formulated for the effects that take place between the blood and lymph compartments. This configuration is particularly desirable because these compartments are filled with fluids whose properties can be determined directly with sensors that establish continuity with the phase whose property is being measured.

When such an approach is followed, the exchange process can, in fact, be regarded as a classical diffusive and convective transport between two fluid-filled compartments separated by a membrane (or barrier), and can be analyzed without detailed information on the flows and conditions of the fluid within the membrane proper. This approach has been particularly successful in the formulation of the theory of membranes by Kedem and Katchalsky (1958), which is based on the study of phenomena occurring in the compartments outside the membrane, where measurements can readily be made, but which does not consider phenomena within the membrane.

## 2.1. The Hydrodynamic Conductivity of Single Vessels

The measurement of capillary permeability to fluid, and more specifically the rate at which fluid passes through the capillary wall under the influence of pressure and osmotic gradients, was pioneered by Landis (see review by Landis and Pappenheimer, 1963) through the development of a micro-occlusion technique.

In the method devised by Landis, a selected capillary is occluded by compression of a portion of its length (usually about one to two diameters) with a glass microtool until blood flow stops. When this occurs, any subsequent sustained motion of the red cells trapped in the occluded capillary must be due to the net exchange of fluid with the surrounding tissue. If it is assumed that the red cells move with the velocity of the fluid they displace, they can be utilized as markers, and the rate at which two red cells approach each other or are separated becomes a measure of the loss or gain of fluid.

Intaglietta and Zweifach (1974) utilized a somewhat different approach based on the formulation of Starling's hypothesis, expressed by Equation (20.1).

$$m^\circ = k_f(p_c - \pi_{pl} - p_t - \pi_t), \quad (20.3)$$

where  $m^\circ$  is fluid movement in  $\mu^3/\mu^2/s$ ,  $k_f$  is the filtration coefficient,  $\pi_{pl}$  is plasma colloid osmotic pressure (COP),  $P_t$  is tissue hydrostatic pressure,  $P_c$  is capillary hydrostatic pressure, and  $\pi_t$  is tissue COP which can then be used to evaluate  $k_f$ . In effect, when fluid exchange is measured first under control conditions, and then again after the blood plasma colloid osmotic pressure has been changed, both capillary pressure and capillary permeability can be determined simultaneously, if it is assumed that tissue effects are negligible.

In Figure 20.5 a graphic representation is presented in which a comparison is made of the rate of fluid movement with the COP at different times during the experimental run. Separate points are indicated for arterial and venous

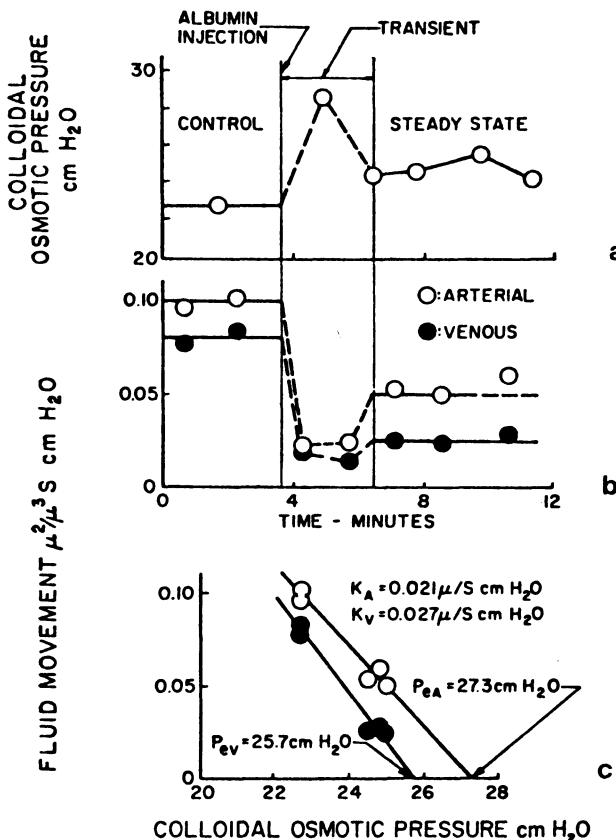


FIGURE 20.5. (a) Osmotic pressures and (b) Fluid movement. The COP values after albumin transient were used to calculate filtration constant and net pressure. (c) COP and fluid movement are compared separately in arterial and venous segments. Lines connecting points represent filtration constant  $K$ . The intercept with the zero fluid movement axis indicates the net or effective arterial pressure (○) and venous pressure (●). Note that slopes of two lines are different, indicating different  $K$  values for arterial and venous portions of the vessels. Effective capillary pressure on arterial ( $P_{ea}$ ) and venous ( $P_{ev}$ ) sides of occlusion are shown (from Zweifach and Intaglietta, 1968).

measurements. The line connecting the open circles intercepts the abscissa at the point where fluid movement is zero and where pressure on the arterial side of the microneedle and COP should be equal. The slope of the line reflects the filtration constant  $k_f$  of the particular vessels involved. As might be anticipated, the pressures on the arterial and venous sides of the occlusion are different. The arterial pressure here averaged some 27 cm H<sub>2</sub>O, while the venous pressure was close to 26 cm H<sub>2</sub>O, a differential of only 1 cm H<sub>2</sub>O across the two ends of this particular capillary. In addition, it can be seen that

the slopes of the two lines were somewhat different. The sharper slope on the venous side indicates a greater filtration constant. These values are indicated in the upper right-hand corner of the graph, with the arterial filtration constant being calculated as .021 and the venous constant as  $.027 \mu^3/\mu^2 \text{ s cm H}_2\text{O}$  difference.

Lee et al (1971) made an analysis of the transcapillary fluid movement during micro-occlusion, in terms of a model that accounts for hematocrit and plasma concentration during filtration. This model was utilized to analyze high-resolution studies of the red cell trajectory in the rabbit omentum. With this approach it was possible to calculate the effective capillary pressure and the filtration coefficient from a single occlusion. The values obtained corroborated previous findings.

In the rat, Smaje et al (1970) found that it was not practical to shift blood plasma COP and to measure this shift by means of serial blood samples. Changes of filtration rate in occluded capillaries were therefore induced by manipulating the COP of the bathing solution. The same technique, when applied to the rabbit omentum, provided values of the filtration coefficient within the range previously found by the intravascular osmotic pressure transient. When applied to the capillaries in muscle, the method of osmotic transients in the bathing solution showed that the hydrodynamic filtration coefficient was of the order of  $0.001 \mu\text{m/s cm H}_2\text{O}$ , which is lower than the corresponding value for omentum and mesentery by a factor of 10. When the suffixes (1) and (2) are used to denote two different concentrations of albumin in the bathing solution, then:

$$K = \frac{m_1^\circ - m_2^\circ}{\pi_{t1} - \pi_{t2}}. \quad (20.4)$$

A plot of fluid movement against differences in external COP could therefore be used to determine "K" (slope of the line in Figure 20.6).

The relative magnitude of the intravascular and extravascular tissue fluxes can be determined by comparing flow and exchange in a single capillary. As a first-order approximation, blood flow in capillaries ( $Q_v$ ) can be assumed to follow Poiseuille's law, expressed by

$$Q_v = \frac{\pi d_c^4 \Delta P}{128 \eta L_c}, \quad (20.5)$$

when  $\Delta P$  is the pressure drop,  $\eta$  is the viscosity of blood,  $L_c$  is capillary length, and  $d_c$  is the capillary diameter. It is assumed that the capillaries have uniform diameters and that the length is the average for a particular tissue. Under *in-vivo* conditions, the hematocrit in the peripheral network is comparatively low, and the viscosity of blood in the capillaries can be assumed to be that of blood with a corresponding hematocrit, without incurring errors of greater than one order of magnitude. For simplicity, the assumption can also be made that the viscosity is Newtonian, as previously indicated. Exchange flow ( $Q_E$ ) can be calculated from the Landis-Zweifach relation:

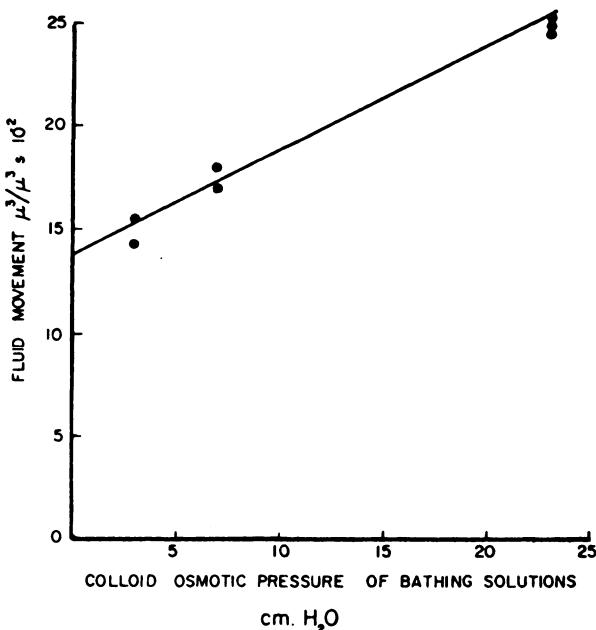


FIGURE 20.6. The diameter of the capillary and the red cell velocity following micro-occlusion are used to calculate fluid movement ( $\mu^3/\mu^2$  capillary surface/s). This is plotted against COP of bathing solution. The slope gives a new  $K$  value of 0.0065  $\mu^3/\mu^2/\text{s cm H}_2\text{O}$  (from Smaje et al, 1970).

$$Q_E = \frac{1}{2} \pi k_f d_c L_c \Delta P_F, \quad (20.6)$$

where  $k_f$  is the filtration coefficient and  $\Delta P_F$  is the average net driving pressure for exchange, and  $d_c$  and  $L_c$  are the diameter and length of the capillary.

It was found that  $\Delta P$  is about four times  $\Delta P_F$ , so that the ratio between the two flows, termed the exchange perfusion parameter,  $R_{E/P}$ , by Intaglietta and Zweifach (1971), becomes

$$R_{E/P} = \frac{Q_E}{Q_V} = 16 \frac{K_F \eta \cdot L_c^2}{d_c^3}. \quad (20.7)$$

This dimensionless parameter makes possible a relatively simple comparison of the factors that determine blood flow and fluid exchange and is expressed as the ratio of exchange flow to blood flow. It provides a measure of dilution of plasma protein as a consequence of the fluid exchange process. Given that  $R_{E/P}$  is always smaller than  $10^{-3}$ , it can be concluded that the effect of the exchange process on concentration of plasma protein is negligible. Considerations of this kind reveal that the interstitium cannot be considered as a closed system; it is actually an open system, continuously irrigated (bathed) by a percolative flux that originates in the exchange microvessels and is collected by the terminal lymphatics.

## 2.2. Perfusion Models

The availability of both flow and precise geometry measurements for several tissues makes it possible to compare micro- and macroscopic phenomena in terms of tissue perfusion—that is, the intravascular fluid flow in the tissue as a whole. Such a comparison is particularly important in studies where the level of tissue perfusion, as a function of different input–output pressure relations, is used as a means of determining the site of action of physiological and pharmacological stimuli.

A comparatively simple model of tissue perfusion can be constructed by assuming that parameters such as flow, velocity, capillary diameter and spacing, and capillary length have a Gaussian distribution. The average flow through a single capillary can then be related to a hypothetical tissue cylinder supplied by that capillary. Several assumptions are implicit: first, that a given volume of blood passes only once through a capillary as it is moved from the arterial to the venous portion of the vasculature; second, that each capillary is associated with a volume of tissue that quantitatively is described by a cylinder whose diameter is that of the intercapillary spacing, and whose length is the average capillary length; third, that the perfusion of the above cylinder of tissue is determined by the flow rate of blood through the capillary  $Q_B$ , which is obtained from the product of the capillary cross section and the average flow velocity,  $\bar{v}$  (cm/s), according to the relation

$$Q_B = 60 \frac{\pi}{4} d_c^2 \bar{v}, \quad (20.8)$$

where  $d_c$  is the average capillary diameter, and 60 is the factor that converts flow into milliliters per minute (density of blood is assumed to be 1).

The amount of tissue,  $W_p$  (assumed to have a density of 1), perfused by the above flow is given by

$$W_p = \frac{\pi}{4} d_c^2 L_c, \quad (20.9)$$

where  $d_s$  is the capillary spacing, and  $L_c$  is the average capillary length. The ratio between Equations (20.8) and (20.9) represents the tissue perfusion at the microscopic level and, when multiplied by 100, provides the value of tissue perfusion,  $Q_{mt}$ , in the usual units of milliliters per minute per 100 g of tissue:

$$Q_{mt} = 6000 \frac{d_c^2 \bar{v}}{d_s L_c}. \quad (20.10)$$

Measurements of both diameter and length are necessary for the construction of a representative model of the microcirculation. Previous studies (Zweifach and Lipowsky, 1984) focused on the total cross-sectional area at various levels; however, this value cannot properly be related to either capacitance or hydraulic resistance and exchange surface unless the actual length–diameter ratio is known.

The accuracy with which this type of model depicts the flow distribution in a bed is dependent on the actual distribution of diameters in a given network of microvessels. The diameter parameter ( $d^4$ ) has a profound effect on the hydraulic resistance of the circuit and therefore on the pressure distribution. In tissues where specialized vessels such as preferential or thoroughfare channels exist, in the sense defined by Zweifach (1957), the calculated distribution of hindrance may change.

### 3. Summary

The mechanisms for control and regulation of blood flow in the microcirculation operate principally through alterations in the resistance to blood flow in individual microvessels, as well as through specific functional segments of the network. For the flow of a Newtonian fluid (i.e., one with constant viscosity) the relationship between pressure, flow, and geometry (vessel length and luminal diameter) can be depicted in terms of the classic Poiseuille-Hagen relationship,  $Q(\pi/128)\Delta Pd^4/l$  (Hagen, 1839; Poiseuille, 1840). In the microvascular network for flow with low Reynolds number ( $\sim 10^{-3} - 10^{-2}$ ), it is not very difficult to apply this relationship to an ensemble of interconnected tubes to delineate the distribution of pressure and flow throughout successive divisions. Departures from such ideal fluid behavior, however, and complexities in network geometry result in nonlinear pressure-flow relationships that obscure the relative roles of blood rheology and network topography as determinants of microvascular perfusion. The network topography encompasses the number of vessels at a specific position in the network available to carry off the total flow of the network, their relative deployment in terms of an arrangement of serial and/or parallel vessels, and the configuration of their cross sections. In the framework of the pressure gradients that propel blood of a specific viscosity through the network, it is the branching pattern of the array that dictates the fraction of the overall arterio-venous pressure drop across any local segment. It is conceivable, if the relationship between pressure gradient, vessel geometry, and blood viscosity is nonlinear, that the branching pattern may interact with the rheological determinants of resistance in the blood to produce rheological equilibrium throughout the network.

### References

- Davis, M.J., and Gore, R.W. 1985. *Am. J. Physiol.* 249:H174–H187.
- Eriksson, G., and Myrhage, R. 1972. *Acta Physiol. Scand.* 86:211–222.
- Frasher, W.G., and Wayland, H. 1972. *Microvasc. Res.* 4:62–76.
- Fronek, K., Zweifach, B.W. 1974. *Angiolgia* 3:35–39.
- \_\_\_\_\_. 1975. *Am. J. Physiol.* 228:791–796.
- Gore, R.W., and Bohlev, H.G. 1977. *Am. J. Physiol.* 242:H268–H287.
- Granger, H.J., and Shepherd, A.P. 1973. *Microvasc. Res.* 5:49–72.

- Hagen, G. 1839. *Ann. Phys. Chem.* 46:423–442.
- Intaglietta, M., and de Plomb, E.P. 1973. *Microvasc. Res.* 6:153–168.
- Intaglietta, M., and Zweifach, B.W. 1971. *Circ. Res.* 28:593–600.
- \_\_\_\_\_. 1974. Pages 111–159 in *Advances in biological and medical physics*. Vol. 15. Academic Press, New York.
- Kedem, O., and Katchalsky, A. 1958. *Biochim. Biophys. Acta* 27:229–246.
- Klitzman, B., and Johnson, P.C. 1982. *Am. J. Physiol.* 242:H211–H219.
- Lamport, H. 1955. Pages 595–599 in *Textbook of physiology*, 17th ed., J.F. Fulton, ed. W.B. Saundersco, Philadelphia.
- Landis, E.M., and Pappenheimer, J.R. 1963. Pages 961–1034 in *Handbook of physiology*. Section 2, *Circulation*. Vol. II. Am. Physiol. Soc., Washington, D.C.
- Lee, J.S., Smaje, L.H., and Zweifach, B.W. 1971. *Circ. Res.* 28:358–370.
- Lipowsky, H.H., and Zweifach, B.W. 1974. *Microvasc. Res.* 7:73–83.
- \_\_\_\_\_. 1977. *Microvasc. Res.* 14:345–361.
- Lipowsky, H.H., Rofe, S., Tannenbaum, L., Firrell, J.C., Usami, S., and Chien, S. 1981. *Microvasc. Res.* 21:249–250.
- Mayrovitz, H.N., and Roy, J. 1981. *Microvasc. Res.* 21:251.
- Murray, C.D. 1926. *Proc. Nat. Acad. Sci.* 12:207–213.
- Nellis, S.H., and Zweifach, B.W. 1977. *Circ. Res.* 40:546–556.
- Poiseuille, J.L.M. 1840. *C.R. Acad. Sci.* 11:961–967.
- Schmid-Schönbein, G.W., Skalak, R., Usami, S., and Chien, S. 1980. *Microvasc. Res.* 19:18–44.
- Smaje, L., Zweifach, B.W., and Intaglietta, M. 1970. *Microvasc. Res.* 2:96–110.
- Starling, E.H. 1896. *J. Physiol. (London)* 19:312–326.
- Yen, R.T., and Fung, Y.C. 1977. *J. Appl. Physiol: Respirat. Environ. Exercise Physiol.* 42:578–586.
- \_\_\_\_\_. 1978. *Am. J. Physiol.* 235:H251–H257.
- Zweifach, B.W. 1957. *Am. J. Med.* 23:684–696.
- \_\_\_\_\_. 1974. *Circ. Res.* 34:843–857.
- Zweifach, B.W., and Intaglietta, M. 1968. *Microvasc. Res.* 1:83–101.
- Zweifach, B.W., Kovalcheck, S., Delano, F., and Chen, P. 1981. *Hypertension* 3:601–614.
- Zweifach, B.W., and Lipowsky, H.H. 1977. *Circ. Res.* 41:381–390.
- \_\_\_\_\_. 1984. Chapter 7 in *Handbook of physiology*. Vol. III. *The Cardiovascular system. Microcirculation*, E.M. Renkin and C.C. Michel, eds. Am. Physiol. Soc., Washington, D.C.

# 21

## Biomechanics of Capillary Blood Flow

R. SKALAK, N. ÖZKAYA, and T.W. SECOMB

### 1. Introduction

The quantitative study of capillary flow dates back to the work of Poiseuille (1840). Because of the difficulties with the clotting of blood, Poiseuille used homogeneous fluids in his famous tests, which established empirically that the discharge through a capillary varies directly with the pressure drop per unit length and the fourth power of the diameter. The formula, which is usually called Poiseuille's Law, is:

$$Q = P\pi D^4/128 \mu L, \quad (21.1)$$

where  $Q$  is the discharge;  $P$  is the pressure drop in the length of the capillary,  $L$ ;  $D$  is the diameter of the capillary; and  $\mu$  is the fluid viscosity. Pouiseuille's Law does not apply directly to capillary blood flow because the blood cells are of the same order of magnitude in diameter as the capillaries themselves. Equation (21.1) nevertheless, is useful to define an apparent viscosity of blood in capillaries. If  $Q$ ,  $P$ ,  $D$ , and  $L$  are all measured, then Equation (21.1) may be solved for the apparent viscosity,  $\mu_a$ . The ratio of  $\mu_a/\mu$ , where  $\mu$  is the viscosity of the suspending fluid, is defined as the relative apparent viscosity,  $\eta$ .

The fact that red blood cells nearly fill the capillary lumen was first observed by Leeuwenhoek in 1688, but the modeling of capillary blood flow containing discrete particles was first proposed in the papers of Prothero and Burton in 1961. Since then, a number of theoretical solutions have been given for axisymmetric particles including spheres, ellipsoids, cylinders, rigid blood cell shapes, rouleaux, and elastic blood cells. A more limited treatment has been given to spheres that are not located axisymmetrically, in which case they rotate, and to so-called tank-treading motions of red blood cells, in which the membrane rotates steadily about the interior although not in a spherical shape. In the sections below, each of the principal categories of the above models are described in more detail.

## 2. Lubrication Theory Models

The basic mechanical properties of red blood cells are well established (Fung, 1966; Skalak, 1976). A thin membrane surrounds the cytoplasm, which is considered to be an incompressible and Newtonian fluid. The membrane exhibits viscoelastic properties, but under steady conditions, only the elastic deformations need be considered. When a red blood cell is subjected to steady external forces, it responds by assuming a configuration in which the external forces are balanced by stresses induced within the membrane and cytoplasm. The elastic shear modulus of the membrane is several orders of magnitude lower than the modulus in dilation (isotropic stretching). Also, bending resistance is not important unless very small radii of curvature are involved. Large isotropic tensions may be generated by small increments in membrane area. Consequently, when highly stressed, the cell will tend to a state in which the dominant membrane stress is isotropic tension. Membrane shear and bending forces become increasingly important at lower cell velocities.

An important advance in modeling the rheology of flow in narrow capillaries was made by Lighthill (1968) and Barnard et al (1968), who applied Reynolds' lubrication theory to calculate the fluid flow between the cells and the vessel wall. In lubrication theory, the equations of fluid motion are simplified by neglecting inertia and by the assumption that the thickness of the gap between the cell and the wall is small compared to the length of the gap. Lubrication theory has been shown to be sufficiently accurate for spherical particles whose diameter is about 90% or more of the capillary diameter (Skalak et al, 1972).

Both the low-velocity and high-velocity regimes may be investigated by using lubrication theory. Consider the flow of an axisymmetric particle (consisting of a membrane filled with an incompressible fluid), through a fluid-filled circular cylindrical tube of radius,  $a$ . For axisymmetric geometry, the fluid flow in the gap and the mechanical equilibrium of the elastic membrane are governed by a system of nonlinear ordinary differential equations. In cylindrical coordinates, the independent variable is the axial distance,  $z$ . The dependent variables are:  $b(z)$ , the local radial coordinate of the membrane-fluid interface;  $\theta(z)$ , the angle between the normal to the membrane and the symmetry axis;  $k(z)$ , the membrane curvature;  $p(z)$ , the fluid pressure in the lubrication layer relative to the internal fluid pressure;  $q(z)$ , the shear force per unit length; and  $t(z)$ , the membrane tension where  $k(z)$ ,  $q(z)$ , and  $t(z)$  are components in a plane containing the axis. Then the governing equations are:

$$\frac{db}{dz} = \frac{1}{\tan \theta}, \quad (21.2)$$

$$\frac{d\theta}{dz} = \frac{k}{\sin \theta}, \quad (21.3)$$

$$\frac{dk}{dz} = \frac{\cos \theta}{b^2} - \frac{k}{b \tan \theta} + \frac{q}{B}, \quad (21.4)$$

$$\frac{dq}{dz} = \frac{kt}{\sin \theta} + \frac{1}{b} [t - \kappa(\lambda^2 - \lambda^{-2})] + \frac{p}{\sin \theta} - \frac{q}{b \tan \theta}, \quad (21.5)$$

$$\begin{aligned} \frac{dp}{dz} = & - \left\{ aq_0 - u_0 \left[ \frac{a^2}{2} + \frac{a^2 - b^2}{4 \ln(b/a)} \right] \right\} \left[ \frac{a^2 - b^2}{16\mu} \right]^{-1} \\ & \times \left[ a^2 + b^2 + \frac{a^2 - b^2}{\ln(b/a)} \right]^{-1}, \end{aligned} \quad (21.6)$$

$$\frac{dt}{dz} = - \frac{\kappa(\lambda^2 - \lambda^{-2})}{b \tan \theta} - \frac{kq - \tau}{\sin \theta}, \quad (21.7)$$

where

$$\tau = \left\{ \frac{1}{4} \frac{dp}{dz} \left[ \frac{a^2 - b^2}{b \ln(b/a)} + 2b \right] - \frac{\mu u_0}{b \ln(b/a)} \right\} \sin \theta, \quad (21.8)$$

$u_0$  is the velocity of the particle,  $q_0$  is the leakback per unit circumference,  $\mu$  is the suspending fluid viscosity,  $\lambda$  is the extension ratio for stretch at constant area, and  $\kappa$  and  $B$  are the membrane elastic moduli of shear and bending, respectively. Here, the elastic response of the cell membrane to shear and bending has been modeled according to Evans and Skalak (1980), with the assumption of negligible membrane dilation. The unstressed shape of the membrane is taken as a sphere with the same surface area as the cell, although other unstressed shapes may be modeled by including an additional term in Equation (21.4). The net force on the particle must be zero since only neutrally buoyant particles are considered. This is the so-called zero-drag condition. The other boundary conditions are that  $b$  and  $\theta$  vanish on the axis, and that surface area and volume are prescribed.

Depending on the additional assumptions made, a number of different approaches are available for solving Equations (21.2)–(21.7). If rigid particles are assumed, then only the lubrication equation (21.6) and zero drag are required. This case is discussed in Section 2.1, below. For flexible particles, additional equations must be retained, as is discussed in Sections 2.2 and 2.3. The full problem of flexible particles has also been treated by the finite-element method (Zarda et al, 1977; Skalak and Tözeren, 1980).

## 2.1. Rigid Particle Models

Studies of flow of rigid particles in circular cylindrical tubes allow exact computations of certain effects. The pressure drop with a rigid particle present is always greater than that for the suspending fluid alone at the same discharge. An axisymmetric particle moves more quickly than the average velocity of the suspending fluid. This means that, relative to the particle, there is a leakback

of suspending fluid. The instantaneous hematocrit (volumetric concentration) of cells in a capillary is always less than that measured in a stationary reservoir in which the discharge of the capillary is collected. This is the Fahraeus effect. There is also a Fahraeus–Lindqvist (1931) effect, which refers to the apparent viscosity in small blood vessels (diameter of about 100  $\mu\text{m}$ ). At the capillary level, the Fahraeus–Lindqvist effect is reversed. That is, the apparent viscosity will increase as the capillary diameter decreases, assuming the cells remain axisymmetric.

An interesting result established by the study of a line of spherical particles in capillaries (Wang and Skalak, 1969; Chen and Skalak, 1970) is that the interaction between neighboring particles is negligible if the particles are spaced more than about one diameter of the capillary apart. This conclusion may also be drawn from the work of Lew and Fung (1968, 1969, 1970). It follows that, at low hematocrits, the additional pressure drop due to the presence of the particle is proportional to the hematocrit. At hematocrits above about 30%, an interaction between the particles reduces the additional pressure drop per particle (Skalak et al, 1972). The details of the flow between two particles were also studied by Lew and Fung (1968).

The pressure distribution in the neighborhood of a solid rigid particle is of interest as an indicator of the stresses to which a flexible particle will be subjected. In Figure 21.1, the pressure distributions for a series of particles consisting of cylindrical portions with ellipsoidal ends are shown. For this model, closed-form solutions are also available (Özkaya and Skalak, 1983). It can be seen that a hemispherical end leads to a high-pressure zone near the front end of the particle and a reduced-pressure zone near the rear end. This pressure distribution suggests that a flexible particle will be compressed at the front and expanded at the rear, as is observed in actual capillary flow.

The assumption of a rigid particle is not realistic for red blood cells, but in any steady flow when the cell shape does not change with time, the flow is the same as for a rigid particle of the same shape.

## 2.2. Flexible Membrane Models

For flexible particles, it is necessary to solve for the shape of the particle as well as the fluid flow. However, Equations (21.2)–(21.7) can be simplified in certain special cases. If bending resistance is neglected, the membrane cannot support a transverse shear force, so  $q = 0$ . Then Equation (21.4) is redundant and Equation (21.5) reduces to an algebraic equation for  $k$ , so that the order of equations reduces by two. The system becomes singular when  $t = 0$ , and allows the possibility of a cusp. Such a cusp may form at the trailing edge of the cell, where the membrane tension is at minimum. The further simplification of neglecting membrane shear elasticity can be achieved by setting  $\kappa = 0$ . These approximations reduce the number of equations and unknowns to four, which can be solved numerically. The model, which extends the approach of Lin et al (1973), and some of the results are given by Secomb and Gross (1983).

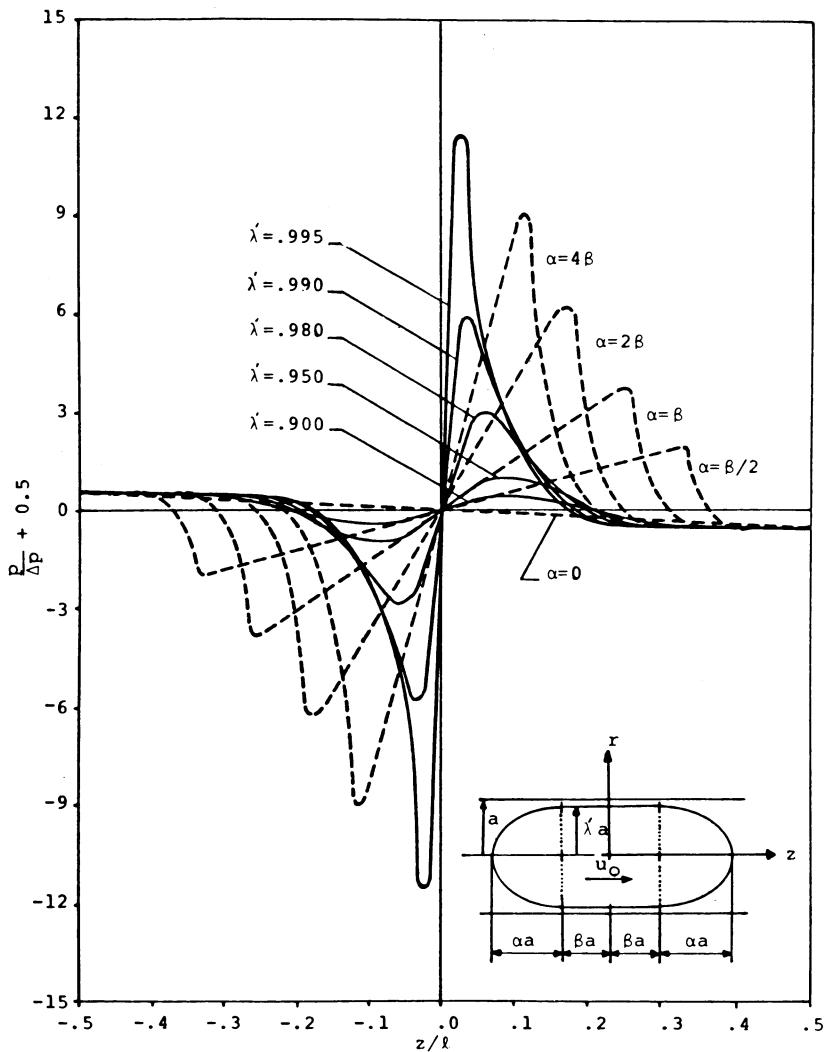


FIGURE 21.1. Fluid pressure distribution along the surface of a set of neutrally buoyant rigid particles. The pressure is normalized with the total pressure drop,  $\Delta p$ , and the axial distance is normalized with the particle length,  $l = 2(\alpha + \beta)a$ . ( $\lambda' a$ ,  $\beta a$  are the diameter and length of the cylindrical portion of the particles that have spheroidal ends of semi-axes  $\alpha a$  and  $\lambda' a$ .) Solid lines are for spheres ( $\alpha = \lambda'$ ,  $\beta = 0$ ), and dashed lines are for particles with  $\lambda' = 0.995$  and various  $\alpha/\beta$ .

This model predicts an apparent viscosity that is independent of flow rate and is appropriate at moderate to high flow rates (above about 1 mm/s) where the viscous stresses are high and elastic stresses are negligible. The predictions of apparent viscosity and Fahraeus effect in vessels with diameters from 3  $\mu\text{m}$

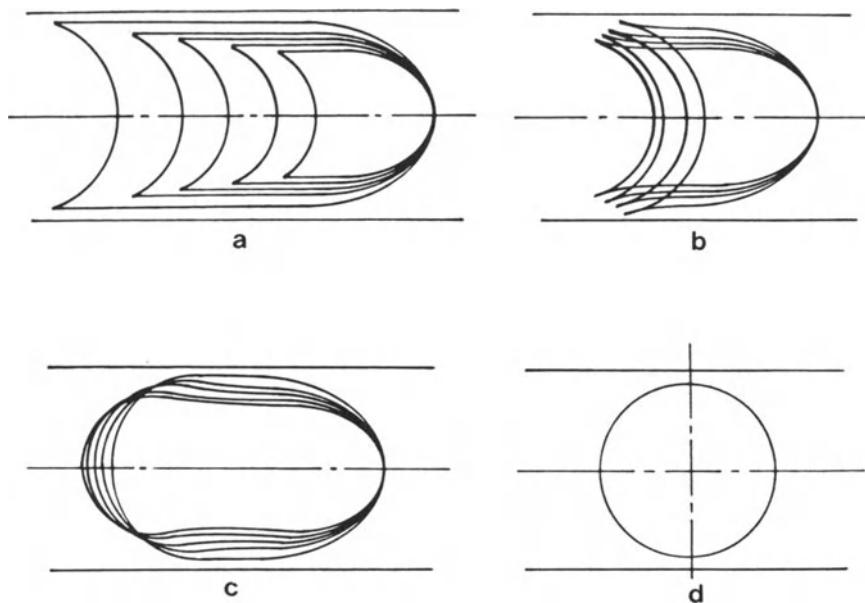
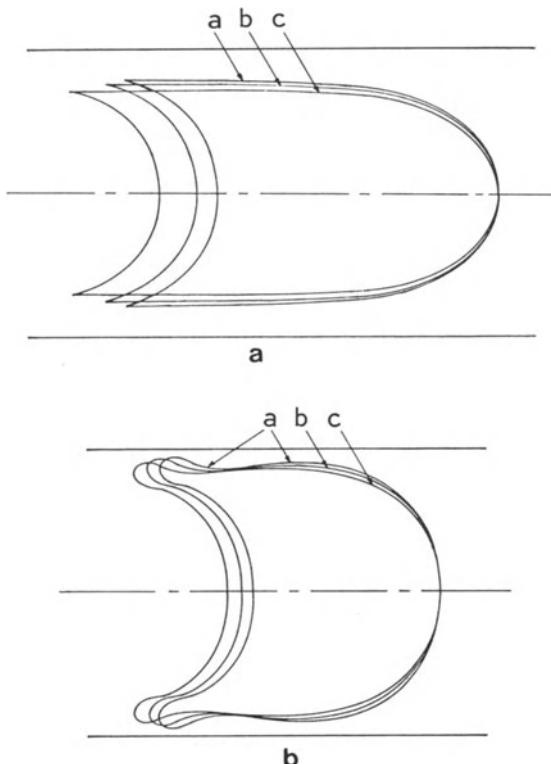


FIGURE 21.2. Neutrally buoyant membrane particles whose shapes are computed using the isotropic tension model. (a) Long cylindrical particles, axial lengths of which are determined by the point where the tension in the membrane and the pressure difference across the membrane vanish. (b) Parachute-shaped particles, the rear ends of which are under finite compression. (c) Particles whose shapes are determined by continuous solutions of the governing equations. (d) A highly inflated spherical particle. The rear ends of class (a) and (b) particles are approximated by spherical segments. In class (a) particles, the rear membrane is not under stress and its shape is arbitrary. However, the shapes shown in (a) and (b) have the correct sphericity index,  $S = \frac{4\pi}{A} \left[ \frac{3V}{4\pi} \right]^{2/3}$ , for red blood cells ( $S = 0.694$ ). Cases (c) and (d) have  $S \geq 0.694$ .

to  $6 \mu\text{m}$  agree well with available experimental data (Gaehtgens, 1980). Through varying the input geometrical parameters, a number of different types of cell shapes may be obtained. Some examples are shown in Figure 21.2.

### 2.3. Elastic Membrane Models

The flow of elastic red blood cells in vessels of diameter  $7.4 \mu\text{m}$  or more has been studied by Zarda et al (1977) and by Skalak and Tözeren (1980), using finite elements to calculate the fluid flow and the cell shape. In narrower vessels, the fluid gap becomes small and a lubrication theory approach is appropriate. If bending resistance is neglected, the effect of membrane shear elasticity may be incorporated as already described. The computed cells again



**FIGURE 21.3.** Shapes of axisymmetric red blood cells passing along a vessel of diameter  $6 \mu\text{m}$ , computed using lubrication theory. Cell volume is  $90 \mu\text{m}^3$ , surface area is  $135 \mu\text{m}^2$ . Motion of cell is from left to right. (a) Results of model including shear elasticity: a:  $\mu_0 = 0.125 \text{ cm/s}$ ; b:  $\mu_0 = 0.25 \text{ cm/s}$ ; c:  $\mu_0 \rightarrow \infty$  which corresponds to the isotropic surface tension model. (b) Results of model including both shear and bending elasticity: a:  $\mu_0 = 0.005 \text{ cm/s}$ ; b:  $\mu_0 = 0.01 \text{ cm/s}$ ; c:  $\mu_0 = 0.02 \text{ cm/s}$ .

have cusps at the trailing edge, and the shape depends on the flow rate, broadening with decreasing cell velocity (Figure 21.3a).

To include both shear and bending elasticity in a lubrication model, it is necessary to solve the full system, Equations (21.2)–(21.7), numerically. Because of ill conditioning, a multiple-shooting method is used to solve the boundary value problem. Computed cell shapes in a  $6\text{-}\mu\text{m}$  vessel are shown in Figure 21.3b. Parameter values are as follows  $\mu = 1 \text{ cP}$ ;  $\kappa = 0.005 \text{ dyn/cm}$ ; and  $B = 10^{-12} \text{ dyn}\cdot\text{cm}$ . Again, the cell more completely fills the lumen as velocity decreases.

Changes in cell shape result in changes in apparent viscosity, which increases as flow rate decreases. Results of several theoretical and experimental studies

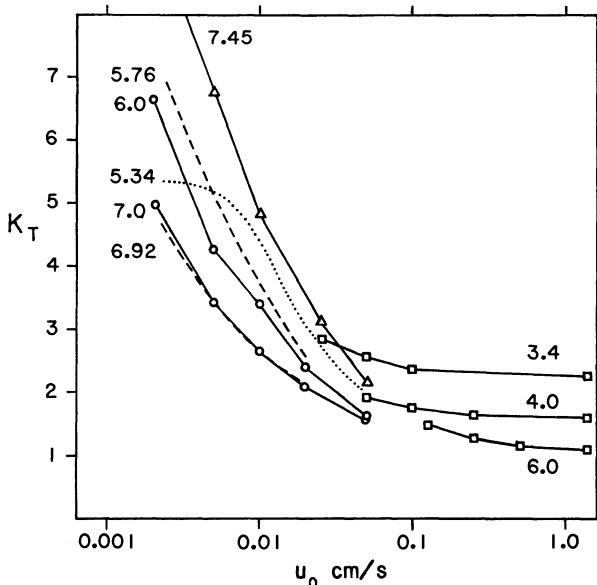


FIGURE 21.4. Selected theoretical and experimental results for apparent viscosity (expressed in terms of  $K_T$ ) in microvessels of various diameters, as a function of cell velocity  $u_0$ . The labels on the curves give vessel diameter in  $\mu\text{m}$ . Theoretical results: —□—, lubrication model including shear elasticity but neglecting bending; —○—, lubrication model including both shear and bending elasticity; —△—, finite-element results of Skalak and Tözeren (1980). Experimental results: ---, results of Lee and Fung (1969) using macroscopic model cells; ...., results of Lingard (1979) using red blood cells.

demonstrating this effect are summarized in Figure 21.4. They are expressed in terms of apparent intrinsic viscosity  $K_T$ , defined by setting

$$\eta = 1 + H_T K_T, \quad (21.9)$$

where  $\eta$  is the apparent viscosity relative to plasma and  $H_T$  is the tube hematocrit. Lee and Fung (1969b) used rubber model cells in their study. In plotting their data, we have rescaled vessel diameters according to cell dimensions, and velocities via the dimensionless group  $\mu u_0/M$ , where  $M$  is the membrane elastic modulus for uniaxial in-plane stress. The values of intrinsic viscosity  $K_T$  predicted by Skalak and Tözeren (1980) appear relatively high compared to the other data, perhaps because the undeformed cell shape was chosen to be a disk coaxial with the vessel. If the capillary diameter is less than that of the undeformed cell shape, the axisymmetric assumption leads to a complete blockage of the lumen as the velocity falls to zero. *In vivo*, it appears that the cells are most often observed with edge-on orientation, which may result in a lower  $K_T$  value. A reduction in the bending stiffness,  $B$ , would also lead to lower  $K_T$  values. The assumption of an unstressed shape of the

membrane is also an influential factor at low velocities. The assumption of an unstressed membrane of spherical form leads to cell shape at zero velocity that has a smaller radius than the normal red blood cell for the elastic parameters assumed above.

### 3. Tank-treading in Capillaries

Most studies of capillary blood flow have assumed axisymmetric geometry, as in the results outlined above. *In vivo*, the geometry of red blood cells is not always axisymmetric (Skalak and Brånenmark, 1969; Bagge et al, 1980). Observations of red cells in glass capillaries (Gaehtgens, 1981; Gaehtgens and Schmid-Schönbien, 1982) have shown that asymmetrically placed red cells may exhibit a tank-treading motion of the membrane similar to that observed in a shear flow (Fischer et al, 1978). A two-dimensional formulation of a capillary tank-treading motion has been developed by Secomb and Skalak (1982), assuming the cell membrane to be perfectly flexible, but inextensible, as shown in Figure 21.5. The flow in the gaps of each side of the cell is assumed to be governed by lubrication theory, and the tension in the membrane is developed by the shear stresses applied by the fluid. In axes moving with the cell, the membrane rotates steadily at the velocity  $u_1$ . The tension in the membrane is assumed to fall to zero at the rear face of the cell. The fluid inside of the cell also undergoes rotation and shear. The results of the analysis suggest that the pressure drop for the asymmetric cell at a given mean velocity may be less than that for an axisymmetric case. As the cell diameter increases, the tank-treading velocity  $u_1$  approaches the forward velocity of the cell so that, on the long side, the cell approaches rolling on the wall (a true tank-treading motion). While these two-dimensional results provide some insights, a three-dimensional analysis of asymmetric cell shapes is needed for more complete understanding.

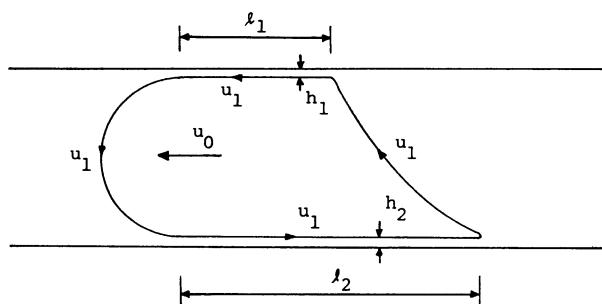


FIGURE 21.5. Two-dimensional model for a tightly fitting, asymmetric cell in a capillary. The cell moves with constant shape to the left at velocity  $u_0$ . The membrane moves at constant speed  $u_1$  relative to axes moving with the cell. The gap spacing  $h_1$  on the shorter side  $l_1$  is found to be generally less than the gap  $h_2$  on the longer side  $l_2$  so that the shear stresses applied to the outside and inside of the membrane are in balance.

In microvessels with diameters in the order of 12  $\mu\text{m}$  or more, cells often travel in two rows each of which is offset from the centerline of the vessel. The two rows form a zipper-like pattern (Gaehtgens et al, 1980). In such arrays, each cell may also undergo a tank-treading motion, but no analytic solutions of such cases have been developed.

#### 4. Hematocrit Fluctuations and Train Formation

When the spacing between cells in a microvessel falls much below one vessel diameter, the cells interact in terms of altered streamline patterns and pressure drop (Skalak and Tözeren, 1980). In a narrow capillary, this can occur when red cells collect behind a slowly moving larger cell, such as a white cell, and a “train” of very high hematocrit is formed (Schmid-Schönbein et al, 1980). In slightly larger microvessels, of diameter about 8  $\mu\text{m}$  or more, the distance between cells is less than the vessel diameter, even at moderate hematocrits. In either case, the effect of the cell-to-cell interactions is to increase the proportion of the lumen occupied by the red cells, and to decrease their mean velocity relative to the bulk flow. The Fahraeus effect is then hematocrit-dependent, decreasing with increasing hematocrit (Barbee and Cokelet, 1971; Albrecht et al, 1979).

The reduction in dynamic vessel hematocrit ( $H_T$ ) relative to the discharge hematocrit ( $H_D$ ) associated with the Fahraeus effect occurs because the mean velocity of red blood cells ( $u_c$ ) exceeds the mean bulk velocity ( $\bar{u}$ ). Under steady conditions, the hematocrit dependence of the Fahraeus effect for vessels of a given diameter may be represented by:

$$\frac{H_D}{H_T} = \frac{u_c}{\bar{u}} = f(H_T). \quad (21.10)$$

The behavior of the function  $f(H_T)$  is not known in general, but for vessels of diameter around 10  $\mu\text{m}$ , suitable data are given by Albrecht et al (1979). Their results, for vessels of diameter of 9.5  $\mu\text{m}$  and 11  $\mu\text{m}$ , are fitted adequately by the single curve (Secomb and Gross, 1984):

$$f(H_T) = \frac{1 + \beta}{1 + \beta H_T}, \text{ where } \beta = 0.7. \quad (21.11)$$

This equation can be derived from the well-known stacked coins model, with the modification that the central core region, which moves with uniform speed, has a cross-sectional area that varies linearly with  $H_T$ .

Hematocrit fluctuations, including train formation, may be described by a one-dimensional continuum model (Secomb and Gross, 1984). The hematocrit at time  $t$  at a point  $z$  along the vessel is denoted by  $H_T(z, t)$ , and the volume flux of red blood cells is denoted by  $Q(z, t)$ . Then the condition for conservation of red cells is:

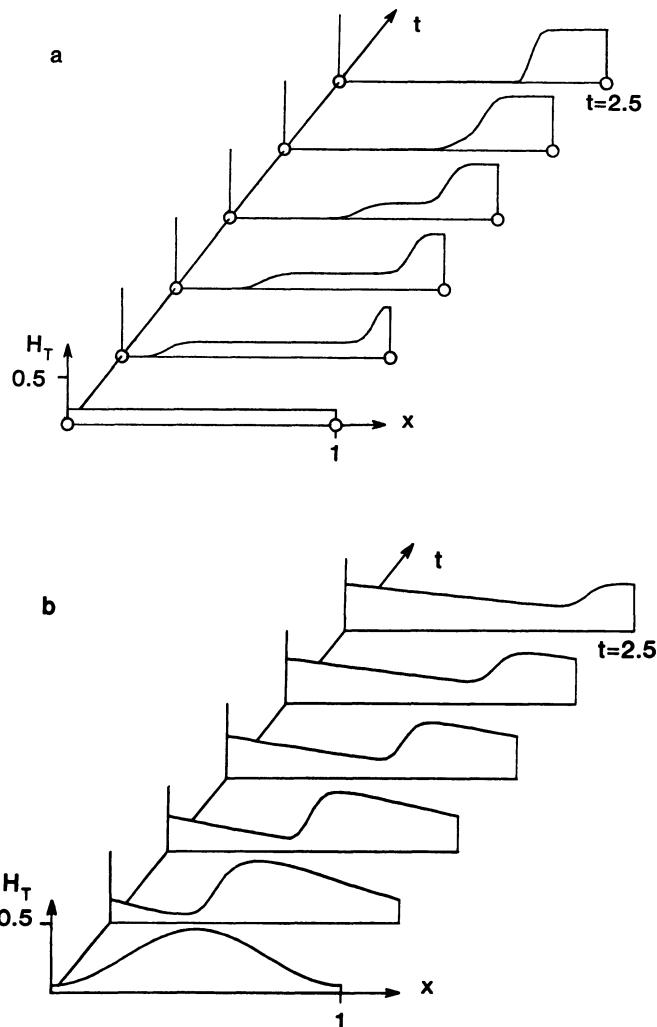


FIGURE 21.6. Numerical results illustrating periodic hematocrit fluctuations. (a) Formation of a train in a vessel containing white blood cells (o). (b) Evolution into a ramp-like waveform and decay due to diffusion of an initially sinusoidal hematocrit fluctuation.

$$\frac{\partial Q}{\partial z} + A \frac{\partial H_T}{\partial t} = 0, \quad (21.12)$$

where  $A$  is the cross-sectional area of the vessel. The dependence of red cell flux on hematocrit is modeled by:

$$Q = A \bar{u} f(H_T) - AD \frac{\partial H_T}{\partial z}. \quad (21.13)$$

The first term gives the flux in a region of uniform hematocrit, according to Equation (21.10), and the second term represents a diffusive effect. Gaehtgens et al (1980) showed that red cells in multiple-file flow oscillate relative to one another. Where there is a hematocrit gradient, the result is a net flux in the direction of decreasing hematocrit. From the data of Gaehtgens et al (1980), an estimate for  $D$  in a 10- $\mu\text{m}$  tube is  $4 \times 10^{-5} \text{ cm}^2/\text{s}$ .

Combining Equations (21.12) and (21.13) gives:

$$\frac{\partial H_T}{\partial t} + \bar{u} \frac{\partial}{\partial z} [H_T f(H_T)] - D \frac{\partial^2 H_T}{\partial z^2} = 0. \quad (21.14)$$

This one-dimensional wave equation is well known in other contexts, such as vehicular traffic flow, and has been studied extensively (Lighthill and Whitham, 1955). Its wavelike solutions are known as “kinematic” waves, to distinguish them from solutions of the classical wave equation.

When the diffusive effect is relatively small, Equation (21.14) has solutions in which regions of uniform hematocrit are connected by zones with rapid hematocrit jumps. For example, suppose that a single white blood cell moving with velocity  $u_w$  enters a vessel in a region of otherwise undisturbed uniform hematocrit  $H_0$ , in which red cells move with velocity  $u_0$ . As the white cell progresses, a train of red cells accumulates behind it, with hematocrit  $H_t$  and moving with mean velocity  $u_w$ . The tail end of the train moves with velocity  $u_b$ . Ahead of the white cell, there is a “plasma gap” with zero hematocrit. The conservation of the volume of cells entering and leaving a small length of the vessel containing the tail end of the train gives a condition analogous to a “shock” condition of aerodynamics:

$$(u_0 - u_b)H_0 = (u_t - u_b)H_t. \quad (21.15)$$

From Equations (21.10) and (21.11), it follows that:

$$H_t = \beta^{-1} \left[ \frac{(1 + \beta)\bar{u}}{u_w} - 1 \right], \quad (21.16)$$

and

$$u_b = \frac{u_w}{(1 + \beta H_0)}. \quad (21.17)$$

According to this theory, train hematocrit is independent of train length, but varies inversely with the speed of the white cell.

To study hematocrit fluctuations in more detail, a numerical scheme was developed for Equation (21.14). For simplicity, the hematocrit distribution was assumed to be periodic with distance along the vessel. Typical time and length scales are one second and one millimeter. Figure 21.5a shows the formation of a train, with a finite diffusive effect, and Figure 21.5b shows the distortion into a ramp-like waveform and eventual decay of an initially sinusoidal hematocrit waveform entering the vessel. The latter case may arise *in vivo* when vasomotion occurs in an arteriolar network, modulating the feed hematocrit (Johnson et al, 1971).

## 5. Closure

The above paragraphs have indicated some of the theoretical solutions and experimental results that apply to capillary blood flow in vessels that are more or less cylindrical in cross section. In the living circulation there are other phenomena of interest, some of which will be only briefly mentioned here. One important aspect is the behavior of a suspension of cells at a bifurcation of a small blood vessel. When particles suspended in fluid are small compared to the diameters of the vessels involved, the division of the bulk flow and of the suspended material will be the same percentage of each constituent. But if the particles are of the same order of magnitude as the vessel diameters, then all of the particles may be diverted into only one of the daughter branches, depending on the eccentricity of the particles approaching the junction and the discharge ratio in each of the daughter branches. These phenomena have been explored experimentally and theoretically to some extent by Fung (1980), Schmid-Schönbein et al (1980), and Vawter et al (1974).

Another aspect of blood flow that is very interesting and important is the flow through organs such as the lung, liver, and spleen, in which the channels are not long cylindrical tubes but have a special geometry in each organ. In the case of the lung, extensive analysis has been developed by Fung et al (1967, 1968, 1969, 1971, 1972, 1973, 1977, 1982, 1983). The elasticity of the pulmonary capillary walls and their interaction with the air pressure in the alveoli form a more complex system than the capillaries in a tissue without air spaces. This subject is reviewed in detail by Fung (1984). In the case of the sinusoids of the liver and the spleen, very little analysis or experimental data on the deformations of the blood cells and the pressure drop involved are available. These highly irregular geometries are not yet analyzed, but the study of flow through fine pores may shed some light on the phenomena here (Skalak et al, 1983). Here, as in the other topics discussed in this review, biomechanical analyses enhance understanding of the functioning of biological systems in terms of mechanical properties of the constituent cells and tissues.

## Acknowledgment

This research was supported by NIH Grants HL-17427 and HL-16851.

## References

- Albrecht, K.H., Gaehtgens, P., Pries, A., and Heuser, M. 1979. *Microvasc. Res.* 18:33-47.
- Bagge, U., and Bränemark, P.-I. 1969. *Blood Cells* 6:231-236.
- Barbee, J.H., and Cokelet, G.R. 1971. *Microvasc. Res.* 3:6-16.
- Barnard, A.C.L., Lopez, L., and Hellums, J.D. 1968. *Microvasc Res* 1:23-34.
- Chen, T.C., and Skalak, R. 1970. *Appl. Sci. Res.* 22:403-441.
- Evans, E.A., and Skalak, R. 1980. *Mechanics and thermodynamics of biomembranes*. CRC Press, Inc., Boca Raton, Florida.

- Fahraeus, R., and Lindqvist, T. 1931. *Am. J. Physiol.* 96:562–568.
- Fischer, T.M., Stöhr-Liesen, M., and Schmid-Schönbein, H. 1978. *Science* 202:894–896.
- Fung, Y.C. 1966. Pages 151–166 in *Biomechanics*. Y.C. Fung, ed. ASME, New York.
- \_\_\_\_\_. 1968. Pages 310–320 in *Biomedical sciences instrumentation*. Plenum Press, New York.
- \_\_\_\_\_. 1969. *J. Biomech.* 2(4):353–373.
- \_\_\_\_\_. 1973. *Microvasc. Res.* 5:34–48.
- \_\_\_\_\_. 1982. *Biorheology* 19:79–94.
- \_\_\_\_\_. 1984. *Biodynamics: circulation*. Springer-Verlag, New York.
- Fung, Y.C., and Sabin, S.S. 1967. *Proc. Ann. Conf. Eng. Med. Bio.* 9:97.
- \_\_\_\_\_. 1969. *J. Appl. Physiol.* 26(4):472–488.
- \_\_\_\_\_. 1972. *Circ. Res.* 30:470–490.
- \_\_\_\_\_. 1977. Pages 267–358 in *Bioengineering aspects of lung biology*, L.W. West, ed. Marcel Dekker, New York.
- Fung, Y.C., and Zweifach, B.W. 1971. Pages 189–210 in *Ann. Rev. Fluid Mech.*, vol. 3 Van Dyke and Vincenti, eds.
- Fung Y.C., Zweifach, B.W., and Intaglietta, M. 1966. *Circ. Res.* 19:441–461.
- Gaehtgens, P. 1980. *Biorheology* 17:183–189.
- \_\_\_\_\_. 1981. Pages 257–275 in *The rheology of blood, blood vessels, and associated tissues*, N.H.C. Hwang and D.R. Gross, eds. Sijhoff and Noordhoff, Amsterdam.
- Gaehtgens, P., Duhressen, C., and Albrecht, K.H. 1980. *Blood Cells* 6:799–812.
- Gaehtgens, P., and Schmid-Schönbein, H. 1982. *Naturwissenschaften* 69:294–296.
- Johnson, P.C., Blaschke, J., Burton, K.S., and Dial, J.H. 1971. *Am. J. Physiol.* 221:105–112.
- Lee, J.S., and Fung, Y.C. 1969a. *J. Fluid Mech.* 37(4):657–670.
- \_\_\_\_\_. 1969b. *Microvasc. Res.* 1(3):221–243.
- \_\_\_\_\_. 1971. *Microvasc. Res.* 3:272–287.
- Leeuwenhoek, A.V. 1688. *On the circulation of the blood*. Latin texts, 65th letter to the Royal Society.
- Lew, H.S., and Fung, Y.C. 1968. *Biorheology* 6:109–119.
- \_\_\_\_\_. 1969. *J. Biomech.* 2:105–119.
- \_\_\_\_\_. 1970. *Biophys. J.* 10(1):80–99.
- Lighthill, M.J. 1968. *J. Fluid Mech.* 34:113–143.
- Lighthill, M.J., and Whitham, G.B. 1955. *Proc. Roy. Soc. Lond. A* 229:281–345.
- Lin, K.L., Lopez, L., and Hellums, J.D. 1973. *Microvasc. Res.* 5:7–19.
- Lingard, P.S. 1979. *Microvasc. Res.* 17:272–289.
- Özkaya, N., and Skalak, R. 1983. Pages 9–10 in *1983 Advances in bioengineering*. ASME, New York.
- Poiseuille, J.L.M. 1840. *Academia des Sciences, Comptes Rendes* 11:961–967, 1041–1048.
- Prothero, J., and Burton, A.C. 1961. *Biophys. J.* 2:213–222.
- Schmid-Schönbein, G.W., Skalak, R., Usami, S., and Chien, S. 1980a. *Microvasc. Res.* 19: 18–44.
- \_\_\_\_\_. 1980b. *Microvasc. Res.* 19:45–70.
- Secomb, T.W., and Gross, J.F. 1983. *Int. J. Microcirc.: Clin. Exp.* 2:229–240.
- \_\_\_\_\_. 1984. In *Progress in microcirculation research II*, F.C. Courtice, D.G. Garlick, and M.A. Perry, eds. University of New South Wales, Sydney.
- Secomb, T.W., and Skalak, R. 1982. *Microvasc. Res.* 24:194–203.

- Skalak, R. 1976. Pages 53–70 in *Microcirculation*, vol. 1. J. Grayson and W. Zingg, eds. Plenum Press, New York.
- Skalak, R., and Bränemark, P-I. 1969. *Science* 164: 717–719.
- Skalak, R., Chen, P.H., and Chien, S. 1972. *Biorheology* 9: 67–82.
- Skalak, R., Impelluso, T., Schmalzer, E.A., and Chien, S. 1983. *Biorheology* 20: 41–56.
- Skalak, R., and Tözeren, H., 1980. Pages 17–40 in *Mathematics of microcirculation phenomena*, J.F. Gross and A.S. Popel, eds. Raven Press, New York.
- Tong, P., and Fung, Y.C. 1971. *J. Appl. Mech.* 38: 721–728.
- Vawter, D., Fung, Y.C., and Zweifach, B.W. 1974. *Microvasc. Res.* 8: 44–52.
- Wang, H., and Skalak, R. 1969. *J. Fluid Mech.* 38: 75–96.
- Yen, R.T., and Fung, Y.C. 1977. *J. Appl. Physiol.* 42(4): 578–586.
- Zarda, P.R., Chien, S., and Skalak, R. 1977. Pages 65–82 in *Computational methods for fluid–solid interaction problems*, T. Belyschko and T.L. Geers, eds. ASME, New York.
- Zhuang, F.Y., and Fung Y.C. 1982. Pages 43–45 in *1982 Advances in bioengineering*. ASME, New York.

# The Static Versus the Dynamic State of the Microvasculature

M. INTAGLIETTA

## 1. Introduction

The raison d'être of the microcirculatory system is the adequate supply of oxygen-carrying red blood cells to the most minute interstices of an organ. The rate at which this process takes place is determined in part by the static mechanical properties of the microvasculature, including spatial architectural relationships, and fluid mechanical phenomena. These properties do not have precise and unique magnitudes but are variable and exhibit a statistical distribution that implies a local variability of blood flow; consequently, in the static microvasculature, some given microscopic domains can be more adequately perfused than others; therefore, as a consequence of this inhomogeneity, localized microscopic regions where cellular survival is marginal can be present in the tissues. A fundamental question is whether these static mechanical characteristics have a real physiological significance in terms of maintaining tissue homeostasis in normal conditions, since the range of local adjustment of the blood supply that these mechanisms allow for is, at best, limited.

When the microcirculation is observed *in vivo* and in conditions that closely approximate the normal physiological state, it becomes apparent that a variety of fail-safe active mechanisms that minimize the danger of localized ischemia are operational. One of these is spontaneous vasomotor activity at the level of individual arterioles, which due to their quasi-periodic dilation and contraction provide a continuous time-dependent redistribution of blood in the tissue, causing alternating periods of localized hyperperfusion and ischemia, a process that increases the number of microscopic zones exposed to adequate perfusion. This phenomenon, in the past, has not been studied on a systematic basis, probably because of the fact that anesthesia, tissue exposure, and acute surgical procedures tend to make it erratic and difficult to reproduce. In fact, in some preparations, anesthesia causes vasomotion to stop altogether (Colantuoni et al, 1984a).

The concept that the circulatory system is active at the level of the individual blood vessels was present since the inception of the study of the micro- and

macrovasculature. Funk and Intaglietta, (1983), summarized the pertinent literature, which begins with the studies by Jones (1852) in the venules of the bat wing, Schiff (1854) in small arteries of the rabbit ear, and Gunning (1857) in the web of the frog.

Activity at the individual vessel level has been repeatedly observed (Krogh, 1929) as periodic changes of red blood cell flow caused by corresponding changes in arteriolar diameters. It is readily evident in tissues that are observed intact, such as the bat wing (Nicoll and Webb, 1946; Wiedeman, 1957; Wiederhielm and Weston, 1973), and the human nailfold (Fagrell et al, 1980) where red blood cell velocity in groups of capillaries is observed to vary periodically at 6 to 10 cycles per minute and synchronously in groups of 10 to 15 capillaries, suggesting that the origin of this activity in this tissue is at the level of the supply arterioles. Vasomotion is evident in skeletal muscle when physiologic conditions comparable to the normal *in vivo* state are maintained. Arfors et al (1975) and Tuma et al (1977) showed this in rabbit tenuissimus muscle when surgical trauma was minimized and when the  $pO_2$  of the superfusing solution was kept at 5 mmHg. Flow in single capillaries in the isolated mesentery loop (Johnson et al, 1971) showed periodicity at 10 cycles per minute. In some instances, the phenomenon was widespread, while in other cases the surrounding vessels maintained steady flow; therefore, both precapillary sphincters and arteriolar activity may cause the effect in this tissue.

Vasomotion as a phenomenon is well documented; however, its significance is controversial and in fact, some investigators believe that it is an abnormal process. In this context, we should note that smooth muscles of many organ systems normally display rhythmic contractile behavior, where the closest analogue is the vasomotion noted in the lymphatic vessels. In fact, these conduits exhibit a rate of contraction dependent on the distention and rate of wall deformation (Ohhashi et al, 1980) and on the presence of groups of pacemaker cells located downstream from the lymphatic valves (as proposed by Pfuhl and Wiegand, 1940, and as tentatively identified by our group as the bifurcations in the microvasculature).

## 2. The Biological Models for the Study of *In-Vivo* Conditions

There is a famous physics premise that the method of observation should not disturb the system being observed. The study of the microcirculation is particularly sensitive to this tenet since it is a delicate living system, which operates at a microscopic level, the normal state of which is easily disrupted when we attempt to render its components accessible to observation and measurement.

Ideally we would study the microcirculation in the intact condition, and in man. This is now possible through the development of video capillaroscopy (Fagrell et al, 1980), which gives data on the microcirculation of the skin.

Measurements are made from video tape recordings of capillaries rendered visible through the microscope by applying a drop of oil on the skin. Velocity of red blood cells is obtained by the dual video window densitometer method, which superposes two densitometric windows over the video image of the capillary under study. Blood flow velocity is measured continuously by computing the delay time to maximum cross-correlation between identical optical signals at the two windows generated by blood cells and plasma gaps. The relative hematocrit is deduced from the continuous video densitometric output of one window. Relative calibration is accomplished by assigning zero hematocrit to the output window in the absence of red cells, and 100% to the maximum output over a period of five minutes. Investigations in normal persons show that quasi-periodic fluctuations of velocity and relative hematocrit are present in most capillaries studied. The frequency is of the order of 5 to 10 cycles per minute, and groups of up to 10 to 15 capillaries are synchronous. The changes of flow are almost invariably accompanied by changes in hematocrit.

These studies show that vasomotion changes the proportion between red cells and plasma. The cause for this effect is that, in the smallest arterioles, the diameter of the vessels can narrow to the extent that erythrocytes are prevented from passing into the capillary. Prior to closure of the arteriole, there is sometimes a flow of plasma such that the capillary is empty of blood cells when flow stops. In the vicinity of the active arterioles, the decrease in velocity and hematocrit will be simultaneous. At some distance downstream, however, the blood in the microvasculature will contain varying local hematocrits. Hematocrit and velocity fluctuations will be progressively further apart as the distance between the point of observation and the point of occlusion increases.

The rhythmic variations of the flow pattern in the human nailfold capillaries suggests the presence of related rhythmic variations in the caliber of the arterioles that control the flow of these capillaries. These vessels, however, are not usually accessible to observation in humans. Furthermore, the capillary flow pattern is of limited value in attempting to quantitate the phenomenon, since it is the summation of the upstream and downstream activity of the vessels.

These considerations have lead us to implement the so-called chamber technique developed by Clark and Clark (1943), and more recently adapted to rats and hamsters by Pappenfuss et al (1979), and Endrich et al (1980). These preparations allow observation of the microcirculation without anesthesia and after healing has mitigated the effects of surgery. The preparations exhibit a consistent pattern of vasomotion when the following conditions are met: (1) there is no bleeding during surgery, and animals that exhibit microbleedings after day 2 are considered pathologic; (2) there is no macroscopic sign of inflammation such as neovascularization, vasodilation, or pus at the points of attachment of the chamber; (3) edema, which is a sign of disturbed vascular permeability and beginning inflammation, is absent; and (4) there are no

leukocytes rolling along venular vessel walls, which is an early sign of impaired physiological condition.

### 3. The Quantitative Characterization of Vasomotion

Spontaneous vasomotion in the hamster skin fold preparation consists in the rhythmic contraction and relaxation of arteriolar vessels of this vascular network. The phenomenon is accurately measured by means of the video technique for image shearing developed by Intaglietta and Tompkins (1973), on the basis of the concept proposed originally by Baez (1966). Continuous diameter records show that vasomotion has a characteristic frequency pattern that is associated with vessel size. Small arteries in the range of 70 to 100  $\mu\text{m}$  exhibit an amplitude of diameter variation of the order of 10 to 20% of mean diameter, and dominant frequencies in the range of 1 to 3 cycles per minute. The smallest arterioles, which have diameters of 8 to 10  $\mu\text{m}$  when they are maximally dilated, are capable of constricting to total closure, i.e. 100% amplitude of diameter variation, with frequencies ranging from 5 to 12 cycles per minute, as described by Colantuoni et al (1984b). Vasomotion, in terms of diameter changes, is not present in the capillary circulation as a consequence of the absence of smooth muscle in these vessels. The anatomical (Hammersen, 1980) and experimental (Tymly et al, 1984) evidence for the concept of spontaneous capillary "contractility" appears to be scarce and restricted to occasional changes in single endothelial cells.

The quantitative measurement of vasomotion constitutes a special mathematical problem, because the activity lacks uniformity, and periods of well-defined and regular patterns of oscillatory behavior are followed by episodes of apparently random vessel wall motion. In this situation, the standard power spectrum techniques such as Fourier analysis fail to give an adequate measure of the time-dependent properties of the activity, because this kind of analysis requires explicit assumptions about the nature of the data outside of the period of observation. Typical assumptions are that it is zero, or that it is a repetition of that contained in the interval under scrutiny, which restricts the use of the Fourier method to phenomena for which the records are long.

A quantitative description of vasomotion can be obtained by means of the Prony Spectral Line Estimator (PSLE) method, which consists in fitting the experimental records to a restricted number of sinusoidal waveforms by a means of a least-squares approximation technique, as described by Kay and Marple (1981). This procedure requires no assumption about the nature of the data outside of the record of interest, and assumes that the phenomenon under study is the consequence of the activity of a limited number of oscillators that are not necessarily harmonically related.

The PSLE method models the signal to be analyzed with a function  $X_{..}$ , which is the sum of  $M$  sinusoids of amplitude  $A_j$ , frequency  $f_j$ , and phase  $\theta_j$ ,

where the signal is known at  $N$  discrete, equally spaced intervals  $\Delta t$ , in such a fashion that:

$$X_n = X(N\Delta t) = \sum_{j=1}^M A_j \cos(2\pi N f_j \Delta t + \theta_j), \quad (22.1)$$

and in complex form:

$$X_n = \sum_{j=1}^M C_j \phi_j^n + C_j^* \phi_j^{*n}, \quad (22.2)$$

where

$$C_j = A_j \exp \frac{1}{2} i \theta_j; \phi_j = \exp(2\pi i f_j \Delta t), \quad (22.3)$$

where \* denotes complex conjugate. This system is nonlinear in the unknown parameter  $\phi_j$ . However, a linear equation can be obtained by introducing a new set of parameters  $a_j$ , that are the coefficients of the polynomial:

$$\sum_{j=0}^{2M} a_j \xi^{2M-j} = \prod_{j=1}^M (\xi - \phi_j)(\xi - \phi_j^*), \quad (22.4)$$

which yields the linear set of equations for  $a_j$ :

$$\sum_{n=0}^{2M} a_{2M-n} X_{k+n} = 0, \text{ where } k = 1, 2, \dots, N - 2M. \quad (22.5)$$

There are  $M$  unknown  $a_j$ s (22.5), and since in general,  $N > 3M$ , the system of equations (22.5) can be solved by the method of the least-squares error. The  $\phi_j$ s are then found by solving the polynomial equation (4.22), and the problem of finding the  $C_j$ s is again linear and solvable using the least-square method.

The basic feature of this technique for fitting sinusoids to a given data set is that nonlinearities are in the polynomial equation (22.4), and the least-square fitting is performed on the linear portions of the system, for which various well-developed algorithms are available. It should be noted that solutions yield the  $\phi_j$ s, which in our case, allow us to calculate the velocity of propagation of the various frequency components of vasomotion. According to this model, vasomotion is characterized by means of a limited number of sinusoids whose sum best fits the given data. The method consists in calculating the Prony spectrum of frequencies, amplitudes, and phases, and then computing the correlation coefficient between the waveform reconstructed with the data from the Prony spectrum and the original data. A first-order approximation contains one sinusoid, and the correlation coefficient is computed as each additional sinusoid is added to the spectrum. If the correlation coefficient increases as a new sinusoid is found, then this solution is kept and the prior solution is discarded. The procedure is continued by adding successive solutions and noting the correlation. If the correlation is high enough, and greater than some acceptable level, a solution has been found. If the maximum correlation is less than the acceptable level, then the signal cannot be quantitatively described by this method, as the assumption of narrow-

bandedness has been violated. As a rule, vasomotion records of two to three minutes, containing 200 to 300 individual data points, show frequency spectra with three to eight separate frequencies, which when utilized to reconstruct the original waveform, fit the data with a correlation coefficient of the order of .85 to .90, which is significant with  $p \gg 0.001$ .

The application of this type of mathematical modeling has tended to confirm the visual observation that this phenomenon originates at the bifurcations of the blood vessels, where these locations assume the function of pacemakers for this activity. This is evidenced by the fact that continuous measurements of diameter fluctuations along a blood vessel show that a given contraction episode appears first in the neighborhood of the bifurcation and later at the distal locations. The fact that the activity is propagated implies the existence of a velocity of propagation, as well as the attenuation of this effect. Since the contraction and dilatation of the vessel travels along the vessel in a wave-like fashion, where the effects are attenuated in proportion to the distance from the source (Intaglietta, 1984), the pattern of diameter changes at any given location is due to the effects that originate at different bifurcations, can be quite complex in appearance, and will be different from place to place.

#### 4. The Origin of Vasomotion

The most uniform patterns of vasomotion are seen in the small arteries and in the smallest arterioles. This suggests that vasomotion is the resultant of the interaction of factors that have a distribution that is vessel size-dependent. It is quite probable that one of these is blood pressure, which is high and close to systemic values in the smaller arteries. The distending effect due to this pressure should effectively overcome the stimuli propagated from distal locations that have been attenuated in proportion to the distance traveled.

A different circumstance should be operational at the level of the terminal arterioles. First, the cessation of smooth muscle involvement in the vessel wall in the capillary microvasculature determines that no signals travel from this system upwards into the terminal arterioles. Thus, the terminal arterioles would be primarily susceptible to the activity of their own pacemakers. Second, these vessels possess the greatest amount of smooth muscle per unit cross-sectional area. Both factors should cause a vigorous and well-defined activity, which in fact corresponds to the findings by Colantuoni et al (1984b).

The principal organic change that occurs between the arteriolar and capillary circulation is the disappearance of the smooth muscle component from the vessel wall, which suggests that this activity is directly related to a property of this cellular element. Many other factors support this concept. Mechanical experiments carried out with smooth muscle preparation from small subcutaneous arteries exhibit spontaneous rhythmic contractile activity and smooth

muscle cells exhibit self-oscillations in its electrical properties, as shown by Funaki and Bohr (1964). Furthermore, Siegel et al (1980) have recently shown that these oscillations are the consequence of the allosteric qualities of the enzyme phosphofructokinase.

The velocity of propagation of vasomotion is in the range of 0.5 to 1.5 mm/s, which is low for transmission through a neurogenic pathway. Furthermore, the frequencies of oscillation at different bifurcations are neither synchronized nor harmonically related, both of which factors suggest the absence of a centralized timing mechanism. The widespread effect of anesthesia, which depresses vasomotion, and in the hamster skin fold eliminates it, can be interpreted in terms of the known direct effects of this process on smooth muscle.

An oxygen-related metabolic feedback has also been proposed by Guyton et al (1964). According to this concept, smooth muscle tone is directly proportioned to the level of extravascular oxygen tension. The rationale for this mode of control is based on the competition between the mechanism that attempts to spare tissue from the toxic effect of high oxygen tension, and that which senses metabolic requirements and causes the arterial microcirculation to dilate in response to ischemia. Oscillation in this kind of a system is due to the lag between cause and effect due to the slow speed of the transmission of signals in the tissue by the diffusion process. The possibility that a tissue-mediated pathway may be involved in vasomotion is also suggested by the coupling of activity that can be seen between paired arterioles and venules.

The preceding discussion highlights the variety of factors that give rise to, and can influence the oscillatory behavior of microvessels. This activity appears to be principally due to the property of smooth muscle cells, which in possessing an oscillatory behavior at the level of single elements, have the potential of becoming pacemaker-like groups of cells through frequency entrainment as proposed by Siegel et al (1980). In fact, Folkow (1964) had proposed the existence of this type of functional entity when he analyzed the myogenic response and stated: "The smooth muscle cells in the vascular sections may be said to form a sort of functional network upon which excitatory and inhibitory influences exert their modulating effects."

## 5. The Physiological Consequence of Vasomotion

The consistent and rhythmic variation of diameter directly affects local peripheral vascular resistance, as discussed by Funk et al (1983). The extent to which the local effect is reflected in the systemic conditions and ultimately blood pressure is difficult to extrapolate because of the probable random distribution of the contraction and dilation episodes in the microcirculation viewed as whole. The fact remains that any given microvessel that is active—i.e., that rhythmically contracts and dilates—possesses a substantially lower hydraulic resistance than its inert counterpart, which has the same average

diameter. We can therefore hypothesize that a decrease of vasomotion, per se, in the absence of any decrease in the average diameter of the blood vessel will determine higher blood pressure. As an example, the aging process, with its concomitant stiffening of the arterial conduits, should lead to decrease in the amplitude of vasomotion, and the consequent progressive rise of blood pressure.

A more subtle, but equally significant effect is determined by the local changes in blood pressure that are due to the activity in the terminal arterioles when they undergo complete closure during portions of the vasomotion cycle. In fact, the periodic strangulation of the terminal arteriole should cause capillary blood pressure to equilibrate with that in the venous exit of the given microcirculatory network, thus enhancing fluid reabsorption from the tissue into the intravascular compartment. This effect was analyzed in detail by Intaglietta (1981), who concluded that vasomotion constitutes an additional mechanism for the extraction of fluid from the tissue, and therefore a safety barrier against edema. Since anesthesia causes the cessation of vasomotion, it is not surprising that edema is often noted as a post-operative syndrome.

In normal physiological conditions, the periodic contraction and relaxation of the terminal arterioles causes a continuous redistribution of blood flow, shifting the supply of oxygenated red cells alternately between the different capillary networks. An important feature of this process is that the narrowing of the terminal arterioles affects the plasma layer, causing red blood cells to travel through the microcirculation in the form of waves rather than as a continuous flux as shown by Fagrell et al (1980). A propitious consequence of this effect is that the high driving pressure, i.e., the high flow rate, is synchronous with the high hematocrit phase, which also corresponds to the higher viscosity in the flowing blood (Intaglietta and Gross, 1981). In this fashion, vasomotion appears to cause a time-dependent hemodilution, which optimizes the use of the pressure gradient available for the motion of blood in the microcirculation.

The presence of periodic changes of volume of arterioles in tissue suggests that there could be concomitant fluid shifts, since a circumferential change in the arteries must displace an equivalent amount of tissue and fluid. The net amount of tissue and fluid displaced between the maximum and minimum diameter excursions is readily calculated (Intaglietta and Gross, 1982).

This model shows that the somewhat abrupt displacement of fluid that accompanies the contraction and relaxation of arterioles during the vaso-motor cycle can shift fluid into the lymphatic drainage system, where this fluid would be prevented from returning to the tissue compartment by lymphatic valves, as demonstrated by Hargens and Zweifach (1977), and by the proposed valve-like structures in the endothelium of the terminal sac (Leak, 1970). Morphological studies in skeletal muscle by Skalak et al (1984) also show that lymphatic vessels run parallel to the active arterioles, and therefore vasomotion induces periodic volume changes in these conduits, giving rise to a peristaltic mechanism for the transport of lymph.

We do not know at present the extent to which vasomotion and related phenomena are fundamental for the survival of the organism in normal conditions. Vasomotion affects in a fundamental way the three principal homeostatic processes, namely peripheral vascular resistance, (and therefore systemic blood pressure), tissue fluid balance, and tissue blood flow distribution. Therefore, a disruption of vasomotion, or its gradual change from the physiological condition, should have a potentially negative outcome for the well-being of the organism. Given the distributed nature of the phenomenon, it is likely that such a process would have effects that manifest themselves on a protracted and long-term basis.

In terms of pathophysiological states, vasomotion can become a highly significant factor in ischemic disease, when the blood flow is insufficient in the presence of an arterial supply that is operational in a static system and that has reached the limit of its dilatation capacity. In this condition, the reappearance of vasomotion will cause a substantial redistribution of blood flow, directing concentrated packets of oxygenated red blood cells to various parts of the microcirculation on an alternate basis. A key feature of this phenomenon is that the blood flow does not necessarily have to increase; however, a beneficial effect is obtained as a consequence of the better redistribution of red blood cells, which then can reach portions of the microvasculature that were previously inaccessible.

The effect of vasomotion on fluid balance and the enhancement of fluid absorption would also be beneficial in those lesions accompanied by edema, where the compression caused by the high tissue fluid pressure restricts capillary diameter and therefore capillary flow. Again, in this condition, the reestablishment of vasomotion would decrease edema, and improve capillary flow and blood redistribution. A paradoxical feature of the effect of vasomotion is that we can hypothesize that in ischemic tissue, when the dilatory capacity of the circulation is exhausted, it is possible that conditions can be ameliorated through the appearance of constrictory events.

## References

- Arfors, K.-E., Bergqvist, D., Intaglietta, M., and Westergren, B. 1975. *Uppsala J. Med. Sci.* 80:27–33.
- Baez, S. 1966. *J. Appl. Physiol.* 21:299.
- Clark, E.R., and Clark, E.L. 1943. *Am. J. Anat.* 73:215–250.
- Colantuoni, A., Bertuglia, S., and Intaglietta, M. 1984a. *Int. J. Microcirc. Clin. Exp.* 3:13–28.
- \_\_\_\_\_. 1984b. *Am. J. Physiol.* 246 (Heart Circ. Physiol. 15):H508–H517.
- Endrich, B., Asaishi, K., Goetz, A., and Messmer, K. 1980. *Res. Exp. Med.* 177: 125–134.
- Fagrell, B., Intaglietta, M., and Ostergren, J. 1980. *Microvasc. Res.* 20:327–335.
- Folkow, B. 1964. *Circ. Res. Suppl.* 14–15:I279–I285.
- Funaki, S., and Bohr, D.F. 1964. *Nature* 203:192–194.

- Funk, W., Endrich, B., Messmer, K., and Intaglietta, M. 1983. *Int. J. Microcirc. Clin. Exp.* 2:11–25.
- Funk, W., and Intaglietta, M. 1983. *Prog. Appl. Microcirc.* 3:66–82.
- Gunning, W.M. 1857. *Utrecht.* Quoted from Riegel, F. 1871. *Pflügers Arch. ges Physiol.* 4:350–434.
- Guyton, A.C., Ross, J.M., Carrier, O., Jr., and Walker, J.R. 1964. *Circ. Res. Suppl.* 1 14–15:I60–I69.
- Hammersen, F. 1980. *Adv. Microcirc.* 9:95–134.
- Hargens, A.R., and Zweifach, B.W. 1977. *Amer. J. Physiol.* 233:H57–H65.
- Intaglietta, M. 1981. *Microvasc. Res.* 21:153–164.
- \_\_\_\_\_. 1984. Pages 5–14 *Angiodynamik und Angiopathie.* W. Zuckschwerdt Verlag München, K. Messmer (Heidelberg). ed.
- Intaglietta, M., and Gross, J.F. 1981. *Microcirculation* 1:161–176.
- \_\_\_\_\_. 1982. *Int. J. Microc. Clin. Exp.* 1:55–65.
- Intaglietta, M., and Tompkins, W.R. 1973. *Microvasc. Res.* 5:309–312.
- Johnson, P.C., Blaschke, J., Burton, K.S., and Dial, J.H. 1971. *Amer. J. Physiol.* 221:105–111.
- Jones, T.W. 1852. *Phil. Trans. R. Soc.* 142:131–136.
- Kay, S.M., and Marple, S.L., Jr. 1981. *Proc. IEEE* 69:1380–1419.
- Krogh, A. 1929. *Anatomy and physiology of the capillaries.* Yale University Press, New Haven, Connecticut.
- Leak, L.V. 1970. *Microvasc. Res.* 2:361–391.
- Nicoll, P.A., and Webb, R.L. 1946. *Ann. N.Y. Acad. Sci.* 46:697–732.
- Ohhashi, T., Azuma, T., and Sakaguchi, M. 1980. *Amer. J. Physiol.* 239:H88–H95.
- Papenfuss, H.D., Gross, J.F., Intaglietta, M., and Treese, F.A. 1979. *Microvasc. Res.* 18:311–318.
- Pfshul, W., and Wiegand, W. 1940. *Z. Mikroskop. Anat. Forsch.* 47:117–136.
- Schiff. 1854. *Arch. Physiol. Hellk. (Leipzig)* 13:523–527.
- Siegel, C., Eblling, B.J., and Mofer, H.W. 1980. *Ber Bunsenges Phys. Chem.* 84:403–406.
- Skalak, T.C., Schmid-Schönbein, G.W., and Zweifach, B.W. 1984. *Microvasc. Res.* 28:95–112.
- Tuma, R.F., Lindbom, L., and Arfors, K.-E. 1977. *Amer. J. Physiol.* 233:289–294.
- Tyml, K., Weigelt, H., and Lubbers, D.W. 1984. *Microvasc. Res.* 27:135–151.
- Wiedeman, M.P. 1957. *Circ. Res.* 5:641–644.
- Wiederhielm, C.A., and Weston, B.V. 1973. *Amer. J. Physiol.* 225:992–996.

# 23

## Vascular Dynamics and the Endothelium

R.M. NEREM, M.J. LEVESQUE, and M. SATO

### 1. Introduction

As the heart rhythmically beats and ejects its periodic output, there are a wide range of phenomena that take place as the pressure pulse propagates through the vascular system. Initially, the interest in vascular dynamics was simply on this phenomenon of wave propagation, and workers such as Young, Frank, Moens, and Korteweg are associated with this earliest phase of research on vascular dynamics. As part of any attempt to explain pulse wave propagation, however, it became necessary to study the elastic and viscoelastic nature of the arterial wall as well as the variation of such properties throughout the various segments of the arterial tree. With such studies, and during the 1950s and 1960s, our understanding of the development of pressure and flow in the systemic circulation (i.e., vascular dynamics at a system level) made considerable progress. The names associated with this period include Womersley, McDonald, Taylor, Fung, and Bergel. These investigations have been reviewed by Fung (1981, 1984) and McDonald (1974), including contributions from more recent studies.

Starting in the late 1960s, there was a surge of interest in the study of the more detailed fluid dynamic characteristics of blood flow in the major mammalian arteries. This activity was the direct result of the possible role of hemodynamics and, in a broader sense, fluid mechanics in atherosclerosis. Because there is a very specific pattern to the disease, i.e., a localization of lesions, it became important to determine the temporal and spatial pattern of blood flow in large arteries. Thus, investigations into local, regional vascular dynamics considered such questions as how to characterize the instantaneous blood velocity profiles, whether there is turbulence, what the nature is of secondary motions that may be present, and whether there is flow separation. In addition, if one considers the hemodynamic force imposed on the wall to have both a normal pressure component and a tangential wall shear stress component, what is the distribution of the latter? These are not easy questions to answer because the flows of interest are pulsatile, pass through vessels that are geometrically complex and vary individualistic from subject to subject,

and in terms of *in-vivo* studies, in some cases are of limited access. Still, our knowledge has grown enormously (Bergel, 1972; Caro et al, 1978; Nerem, 1981; Schettler et al, 1983; and Fung, 1984).

This interest in hemodynamics as a factor in atherosclerosis, particularly its genesis, has also raised questions about the influence of hemodynamic-related events on the arterial wall itself, and here the focus has turned to the endothelium. This monolayer of cells, which makes up the inner lining of arteries, is not only a blood-compatible container, but represents a selective permeability barrier to molecules being transported and has synthetic, metabolic capabilities. It also is the interface between the flowing blood and the wall. As such, it is believed to be the mediator of any hemodynamic effect, and there have been a number of laboratories that have turned their attention to the study of the influence of flow on the behavior of the endothelium and the dynamics of the endothelial cells of which it is composed.

This represents a new focus for our interest in vascular dynamics. From what started as an interest in system vascular dynamics in the last century, to the interest in local, regional vascular dynamics stimulated by the possible role of hemodynamics in atherogenesis, we now are beginning to look at the microscopic, cellular level of vascular dynamics, that level associated with the endothelium and vascular endothelial cells.

It is this new, emerging area of study—vascular dynamics at the level of the endothelial cell—that forms the subject of this chapter. In this, it is the mechanics of the endothelium, including the influence of mechanical forces, that is the focus of our attention. With the advent of mammalian cell culture systems, it now is possible to study *in vitro* the behavior of vascular endothelial cells exposed to various types of flow conditions, and as will be seen in the next section, some rather striking results are being obtained. These, of course, must be taken in the context of what we know about the *in-vivo* endothelium, and such studies will be described in the third section of this chapter. Finally, speculations on the role of the endothelium and the dynamics of endothelial cells in the initiation or genesis of atherosclerosis are discussed.

## 2. *In-vitro* Cell Culture Studies

Although the first investigation employing cultured cells to study the effects of flow used kidney cells, for the most part it has been vascular endothelial cells that have been the focus of these studies. The most commonly used cell type is the bovine aortic endothelial cell, and the culture of such cells has been described by Gimbrone (1976). These cells, in the presence of a nutrient, can be grown as a monolayer in a petri dish to the point of confluence, where contact inhibition seems to arrest their growth. If one detaches some of the cells, for example with trypsin, and then replates or reattaches the cells at a lower density (cells/area) in a new dish, a new generation can be started. This can be repeated, a process called multiple passage, and most of the studies on

the effects of a fluid-mechanically imposed shear stress to be discussed here have used between fifth and tenth passage cells.

The first major effort in this area was that of Dewey et al (1981), where cultured monolayers of bovine aortic endothelia cells were placed in a cone-plate apparatus that produced a uniform fluid shear stress on their cell samples. They showed that subconfluent endothelial cultures continuously exposed to 1–5 dynes/cm<sup>2</sup> shear proliferated at a rate comparable to that of static cultures and reached the same saturation density ( $\cong 10^5$  cells/cm<sup>2</sup>). When exposed to a laminar shear stress of 5–10 dynes/cm<sup>2</sup>, confluent monolayers underwent a time-dependent change in cell shape, becoming more elongated and oriented with the flow. Furthermore, regeneration of linear “wounds” in confluent monolayers appeared to be influenced by the direction of the applied shear force.

This group has also looked at the organization of the endothelial cytoskeleton using bovine aortic endothelial cells under static, no flow, and shear stress conditions *in vitro* (White et al, 1982). Cytoskeletal elements, such as stress fibers made up of linear arrays of the “contractile” proteins actin, myosin, and other “accessory” proteins, are thought to play a role in various basic cellular functions, including maintenance of cell shape, cell attachment to its substrate, and cellular contraction. Cultured cells under static, no-flow conditions exhibited a random stress fiber orientation. However, after exposure to 8 dynes/cm<sup>2</sup> for 72 hours, these cells became aligned with the flow direction, and their cytoskeletal structure underwent a dramatic reorganization, with stress fibers appearing to be much more prominent and showing orientation with the major cell axis and with the direction of flow. The organization of the extracellular matrix beneath the endothelial monolayer also appeared to be influenced; collagen on a coverslip demonstrated a basket-weave pattern for unstressed samples and an oriented fibrillar pattern for endothelial cells exposed to shear stress.

Our own laboratory efforts also have been directed mostly at the use of bovine aortic endothelial cells, with various projects now in progress. In one of these, confluent cultured endothelial cells are being exposed to known levels of shear stress using a parallel-plate, channel-flow device. Such cells, exposed to a constant shear stress, are compared with a similar, but control population of cells in Figure 23.1. From the results obtained, there is a clear indication of the presence of a shear stress effect on orientation, but one that is substrate-dependent. For example, using a glass substrate with only a one-hour exposure to a shear stress of 8 dynes/cm<sup>2</sup>, there is almost a total orientation of the cells with the direction of flow. This is illustrated in Figure 23.2, where the shape index defined as  $4\pi A/P^2$  (where  $A$  is cell area and  $P$  is cell perimeter) is shown for each cell in the population sample as a function of angle of orientation with the flow. Also shown is the least-squares fit to the data, and it can be seen that the more oriented a cell is with the flow, the more elongated it also is, as evidenced by a lower shape index (a larger shape index corresponds to a more rounded cell, with a circle having a value of 1.0; a lower

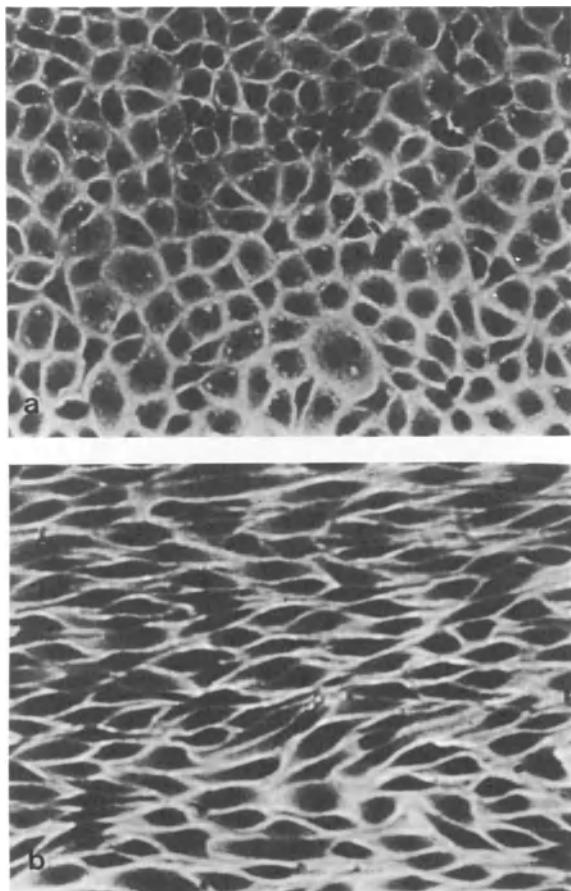


FIGURE 23.1. Light micrographs of a Giemsa stained confluent cultured monolayer of bovine aortic endothelial cells on a Thermanox plastic substrate: (a) Static, no-flow conditions, (b) Exposed to a shear stress of 85 dynes/cm<sup>2</sup> for a period of 24 hours.

index corresponds to a more elongated cell, with a straight line having a value of 0.) This is to be contrasted with endothelial cells on a plastic substrate (Thermanox cover slips) exposed to a shear stress of 85 dynes/cm<sup>2</sup> where there was virtually no tendency to orient the cells in the direction of the flow after one hour, even though the shear was an order of magnitude higher. However, cells grown on the same plastic substrate and also exposed to a shear stress of 85 dynes/cm<sup>2</sup>, but for an extended period of 24 hours, again showed a strong sense of orientation. Here again, and as can be seen in Figure 23.3, there was a correlation such that the more oriented a cell was with the flow, the more elongated it was also.

Others have also documented the effect of flow on the orientation of

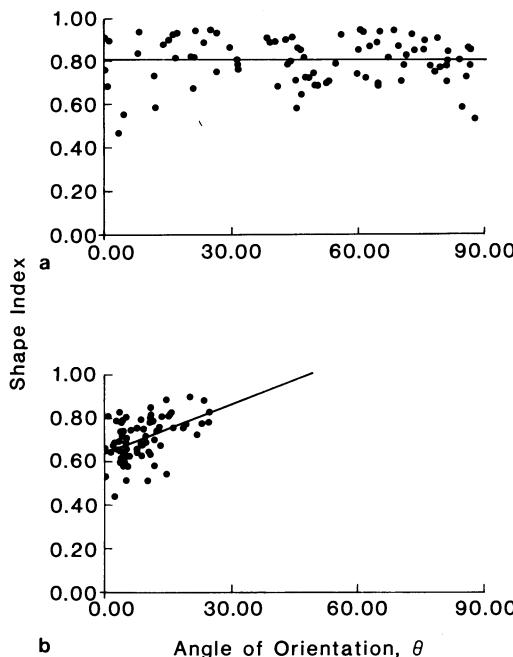


FIGURE 23.2. Correlation of cell shape index with angle of cell orientation relative to the flow for a confluent cultured monolayer of bovine aortic endothelial cells on a glass substrate: (a) Static, no-flow condition (mean angle of orientation =  $47.6^\circ$ , mean shape index = 0.803). (b) Exposure to a shear stress of 8 dynes/cm<sup>2</sup> for one hour (mean angle of orientation =  $9.1^\circ$ , mean shape index = 0.707).

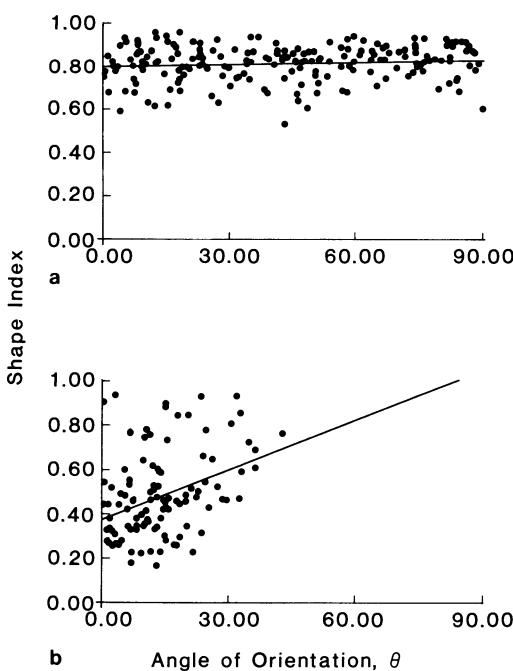


FIGURE 23.3. Correlation of cell shape index versus cell angle of orientation relative to the flow for a confluent cultured monolayer of bovine aortic endothelial cells on a Thermanox plastic substrate: (a) Exposed to a shear stress of 85 dynes/cm<sup>2</sup> for one hour (mean angle of orientation =  $42.7^\circ$ , mean shape index = 0.816). (b) Exposed to a shear stress of 85 dynes/cm<sup>2</sup> for 24 hours (mean angle of orientation =  $14.0^\circ$ , mean shape index = 0.479).

cultured endothelial cells. In addition, data are beginning to accumulate on the influence of shear stress on cell function. For example, Dewey (1984) has reported data obtained with his co-workers on endocytosis using the soluble enzyme tracer molecule horseradish peroxidase (HRP). In a series of experiments, endocytosis rates have been measured for endothelial cells, with and without shear, and with changing shear. Where endothelial cells were exposed to two different levels of steady stress, 0 and 8 dynes/cm<sup>2</sup>, for 48 hours, the endocytosis rates were equal, i.e., independent of the level of shear stress. However, a change in the shear stress, either from 0 to 8 dynes/cm<sup>2</sup> or 8 dynes/cm<sup>2</sup> to 0, produced an endocytosis rate that was double that of cells exposed to a steady stress. In a companion experiment, the shear stress was alternated between 1 and 8 dynes/cm<sup>2</sup> every 15 minutes over a two-hour period. Again, the endocytosis rate was twice that of static, no-flow controls (1 dyne/cm<sup>2</sup> is not sufficient to produce endothelial orientation).

To investigate the response of endothelial cells to an oscillating shear stress as occurs *in vivo*, the cone axis was tilted relative to the axis of rotation, thus producing a "wobble" in the plate. The shear oscillated at 1 Hz, with a peak-to-peak amplitude of 10 dynes/cm<sup>2</sup>, and with a mean of 8 dynes/cm<sup>2</sup>. Over a wide range of combinations of different shear stress levels and times of application, there was no observable increase of uptake from the level of static controls. This perhaps suggests that, at least for HRP, temporal effects are important, but not on a time scale characteristic of the normal heart beat, i.e., the time associated with the oscillating shear stress in a large artery.

There also have been studies of prostacyclin (PGI<sub>2</sub>) production by cultured endothelial cells, but with somewhat different results. *In vitro* PGI<sub>2</sub> production was measured under no-flow conditions, steady shear stress, and pulsatile shear stress (Frangos et al, 1983). Cultured human umbilical vein endothelial cells were placed in a parallel plate flow chamber where pulsatility was introduced by periodically varying the circuit resistance. At the onset of a steady shear stress of 16 dynes/cm<sup>2</sup>, PGI<sub>2</sub> production was 2.0 ng/min/10<sup>6</sup> cells. This leveled off to 0.14 ng/min/10<sup>6</sup> cells after 1.5 hours. However, upon superimposing a pulsatile shear stress at a frequency of 1 Hz and an amplitude of 10% of the mean value, PGI<sub>2</sub> production rose quickly to 6.2 ng/min/10<sup>6</sup> cells and continued to increase over the next few hours. This is in contrast with no-flow conditions, where production was 0.004 ng/min/10<sup>6</sup>. This, taken together with the earlier noted results, suggests that, depending on the molecule and the particular cell process of interest, there may be a wide variety of time scales associated with the influence of a fluid-mechanically imposed shear stress on cell function.

A different technique for studying the response of cultured endothelial cells to transient mechanical phenomena is now underway in our own laboratory. This involves exposing cells to an oscillatory pressure field. In a pilot study, cells grown on polystyrene were exposed to different levels of static or oscillatory pressures at frequencies of 1 and 2 Hz. In general, the endothelial cells

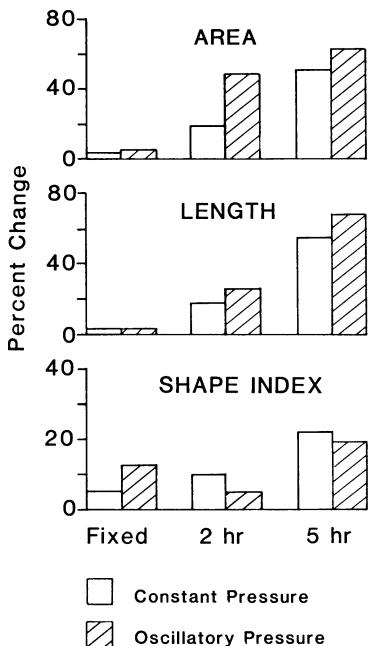


FIGURE 23.4. The percent change in area, length, and shape index with time for a confluent cultured monolayer of bovine aortic endothelial cells on a Thermanox plastic substrate exposed to a constant pressure of 100 mmHg and to an oscillatory pressure at a frequency of 1 Hz with a mean pressure level of 100 mmHg and a pulse pressure of 40 mmHg. Control denotes cells fixed at pressure at zero time; percent change relative to controls at zero pressure.

responded to an increase in pressure by changing their geometry, as can be seen in Figure 23.4. As an example, endothelial cells exposed to a static pressure of 100 mmHg for five hours showed a significant change in surface area, perimeter, length, and shape index, with the response being dependent on the pressure level, duration of exposure, and degree of confluence. The results in Figure 23.4 show no significant difference between cells exposed to an oscillatory pressure versus those exposed to a constant pressure. However, with an oscillatory pressure, the endothelial cells tended to lose their degree of confluence and were characterized by enhanced migration and increased mitotic activity. It is of interest that, in general, observations of cultured endothelial cells undergoing a change in mechanical environment, either through flow or pressure, demonstrate enhanced activity as measured by such variables as migration, mitosis, HRP endocytosis, and PGI<sub>2</sub> production.

These data suggest that fluid mechanical forces play an important role in determining cellular structure and function, both of which are important elements in maintaining vascular integrity. Furthermore, the morphology of an endothelial cell and its response to its mechanical environment may depend on the cell membrane's mechanical properties. Thus, experiments recently have been initiated in our laboratory to study endothelial cell membrane deformation using a micropipette technique (Fung, 1981; Evans and Skalak, 1979). In brief, a small micropipette (inside diameter ~3  $\mu\text{m}$ ) is attached to a pressure reservoir. The tip of the pipette is brought close to a cell which has been detached from its substrate. When the pipette makes contact with the

cell, the pressure reservoir is lowered to create a negative pressure on the cell membrane, and the membrane is then pulled into the pipette. The elongation of the membrane inside the pipette in relation to pressure is recorded, where the relationship between these is given by

$$\Delta p = (\mu/Rp)[(2L/Rp - 1) + \ln(2L/Rp)]. \quad (23.1)$$

Here  $\Delta p$  = recorded pressure,  $Rp$  = pipette radius,  $L$  = elongation inside the pipette, and  $\mu$  = shear modulus. Thus, the elastic shear modulus ( $\mu$ ) is determined by correlating the observed elongation length versus the aspirating pressure. Preliminary results indicate that isolated endothelial cells exhibit values for the measure shear modulus that range from 0.05 to 0.2 dyne/cm for trypsin-detached cells and from 0.1 to 0.3 dyne/cm for mechanically detached cells. It appears that the shear modulus of an isolated endothelial cell is an order of magnitude greater than the shear modulus of a red blood cell, and it is likely that these should be considered measurements of an effective membrane shear modulus (i.e., one that includes the influence of the cytoskeletal structure and the cytoplasm on the "apparent" membrane characteristics). We also suspect that the scatter in our results may be dependent on the metabolic state of the cells (i.e., phase in cell cycle).

Other experiments are in progress that are focusing on membrane deformation of isolated cells in relation to passage number, degree of confluence, and exposure to shear stress. With regard to the latter, Figure 23.5 presents data on the effect of a steady shear stress on a confluent layer of endothelial

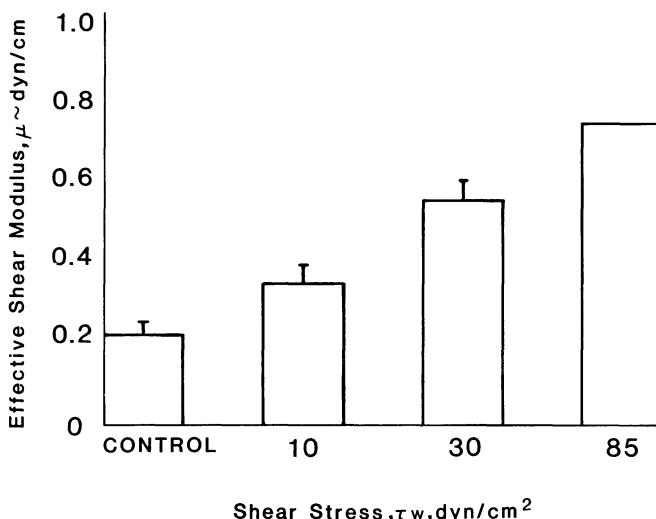


FIGURE 23.5. The effective shear modulus of a confluent cultured monolayer of bovine aortic endothelial cells on a Thermanox plastic substrate exposed for 24 hours to a shear stress of up to 85 dynes/cm<sup>2</sup>.

cells on a plastic substrate for three different levels of shear stress and for 24 hours of exposure to stress. As may be seen, with increasing shear stress, there is a significant increase in the value of the effective shear modulus. These experiments indicate that, for low shear stress levels and short exposure times, there is little effect of a fluid-imposed shear stress on the mechanical properties of an endothelial cell. However, at longer exposure times and for higher stress levels there is a very significant increase as measured by the cell mechanical properties.

A rather fascinating observation has been as follows. A cell, which has been exposed to shear and then is trypsin-detached, becomes spherical, and thus for such cells, it is the membrane properties, i.e. the shear modulus, of a spherical endothelial cell which is measured. The same is true for a mechanically-detached endothelial cell exposed to either zero or very low shear. However, for an endothelial cell exposed to high shear stress levels for a long period of time and which is subsequently mechanically detached, the cell retains its elongated shape. Although any explanation of this phenomenon is speculative at this point in time, it is known that, when fibroblasts are detached, they also take on a spherical shape and cytoskeletal structure which is characterized by a submembranous microfilament region. This is in contrast to a more diffuse, microfilament bundle structure when attached to a substrate. It thus may be that endothelial cells which have never been exposed to a fluid-imposed shear stress, upon detachment, have a cytoskeletal structure which may be characterized as submembranous in nature; while endothelial cells exposed to shear stress, particularly for long exposure times, adapt by modifying their cytoskeletal structure so as to enhance cell stiffness and whereby, even upon detachment, retain a diffuse character such that the elongated shape of a cell can be maintained.

### 3. Studies of the Arterial Endothelium

Studies using cultured endothelial cells have opened new avenues and are providing new results. However, it is important to note that such monolayers are not a replicate of the arterial endothelium, and it is this that is of interest in terms of a possible influence of hemodynamic-related events in the genesis of atherosclerosis. Early studies of the influence of hemodynamics on the endothelium, largely of an *in-vitro* nature and using excised tissue, focused on transendothelial transport. Fry's results (1968, 1969) suggested that the wall shear stress influences the rate of transport of macromolecules between blood and the arterial wall, and the results of Caro and Nerem (1973) using  $^{14}\text{C}$ -4-cholesterol have shown that it is not an effect of wall shear on diffusion boundary layer-controlled transport, but one considerably more subtle in terms of an effect of shear stress on the properties of the arterial wall. Similar steady flow results have been obtained using  $^{131}\text{I}$ -albumin, and data from both radioactively labeled albumin and cholesterol experiments indicate that it is

not only the mean component of arterial pressure and flow, but also the oscillatory nature of the phenomenon that has an important influence on the blood-arterial wall transport of macromolecules (Nerem, 1981). These results suggest that, at low levels of wall shear stress, there is at best a very weak shear dependence; however, at shear stress levels greater than 50 dynes/cm<sup>2</sup>, there is a much stronger dependence of the transport rate on shear stress. More recent results confirm this, but raise a question as to whether excised vessels, particularly at such shear stress levels, have an intact endothelium.

Vaishnav et al (1983) have attempted to quantify the strength of the endothelium in the canine aorta. In these experiments, jets of physiological saline were impinged normally on the endothelium of freshly excised segments of canine middle descending thoracic aortas. A characteristic type of annular lesion was found that showed that the endothelium could withstand large normal stresses where the jet impinged, but was easily eroded by the shear stress resulting from the jet efflux. These results also demonstrate that the erosion stress,  $\tau_e$ , i.e. the arterial shear stress required to cause endothelial denudation, is strongly dependent on the duration of exposure and on environmental factors such as temperature. For short durations,  $\tau_e$  increases to the order of several thousand dynes/cm<sup>2</sup>, and it thus appears that peak physiological wall shear stress levels are far below that required to cause denudation and that the influence of fluid mechanics *in vivo* must occur in the presence of a intact endothelium.

Below the level of the erosion stress required for denudation, it appears that endothelial cell geometry and orientation is influenced by hemodynamic forces. Endothelial cells tend to align themselves with the flow field and have an elongated teardrop configuration. The cell nucleus and the bulk of the cell cytoplasm are distal to the midpoint of the longitudinal axis of the cell, and endothelial cells about orifices of branches are often polygonal in outline. Thus the shape and orientation of endothelial cells appear to be dependent on the flowing blood, with cells being more elongated in regions of higher shear.

Our own studies of arterial endothelium have focused on the effects of pressure and of wall shear stress. Using the vascular casting technique, rabbit endothelial cell geometry and orientation have been studied under conditions corresponding to acutely elevated pressure and in regions of hemodynamic interest. These studies, although largely conducted using the rabbit as an animal model, have looked at various animals ranging from the rat to the dog. The data obtained provide striking evidence that the *in-vivo* endothelium is intact, i.e., no regions of denudation were found, and that in general the shape index has a range from 0.2 to 0.6 for physiologic conditions. Studies of rabbit abdominal aortic endothelium indicate that, over a pressure range from 65 to 150 mmHg, there is a significant increase in shape index—a rounding of the endothelial cells with increased pressure. Whether this would also be true for chronic hypertensive conditions and whether this is due to a change in cell shape or to an adjustment in cell junctional overlap is not known.

We have also examined endothelial cell patterns in the region of intercostal

ostia, as determined from casts of rabbit aortae, and these have been shown to be very much suggestive of the complicated flow patterns that one might expect to be present on the basis of fluid mechanic considerations (Nerem et al, 1981). To be more specific about the shape of aortic endothelial cells in a region of branching, comments on the endothelial cell pattern are shown in Figure 23.6 with some scanning electron micrographs of cell patterns from the rabbit intercostal orifice included as examples. As is noted, as one approaches the branch point from the proximal side, the cells on the surface from which the branch emanates will become increasingly more rounded, indicating the possibility of a decreasing wall shear stress. On the leading edge of the flow divider, the cells will also be of a generally round shape; however, here and in contrast to other locations, the pattern will be highly disorganized, demonstrating no particular preferred orientation, and presumably indicative of a stagnation region. On the aortic flow divider surface immediately distal to the leading edge, there is a region a few cell lengths long in which the cells are highly elongated, particularly when compared to the shorter, more rounded cells proximal to the branch point. This suggests a comparatively higher shear stress on the distal side of the branch than on the proximal side. For the coeliac artery, which is a large branch vessel, there also is a region of more rounded cells on the aortic surface opposite the flow divider; and in the region of the iliac bifurcation, endothelial cells appear to be more rounded on the outer, lateral walls and more elongated on the flow divider.

In the analysis of these patterns, it is the mean flow that is of interest since cell culture studies suggest that the orientation of cells would not be responsive to a time scale as short as that associated with pulsatile blood flow. From this perspective, regions of high and low wall shear stress, flow stagnation, and nonaxial streamline patterns are all suggested by the orientation and geometry of the endothelial cells. It thus may well be that, whereas aerodynamicists must "tuft" wings with small lengths of thread in order to visualize the pattern of flow present, the body has provided us with a natural marker or indicator of detailed features of blood flow and perhaps even the level of wall shear stress.

In a more recent experiment in our laboratory, aortic stenosis has been produced in the dog by banding the abdominal aorta. Using the vascular casting technique, endothelial cell geometry and orientation have been determined as a function of the position relative to the stenosis, and these results are presented in Figure 23.7. As may be seen, far proximal to the stenosis, the shape index is normal and equal to that for control conditions. However, as one enters into the stenosis where the cross-sectional area decreases, the cells become more elongated and achieve their greatest elongation (i.e., smallest shape index) right at the throat of the stenosis where the shear stress would be expected to be the highest. Immediately downstream of the throat, one finds very round endothelial cells. This is presumably a region where there is a jet and the flow is separated from the wall. However, as the jet grows and ultimately reattaches to the wall, there is a gradual return of the shape index

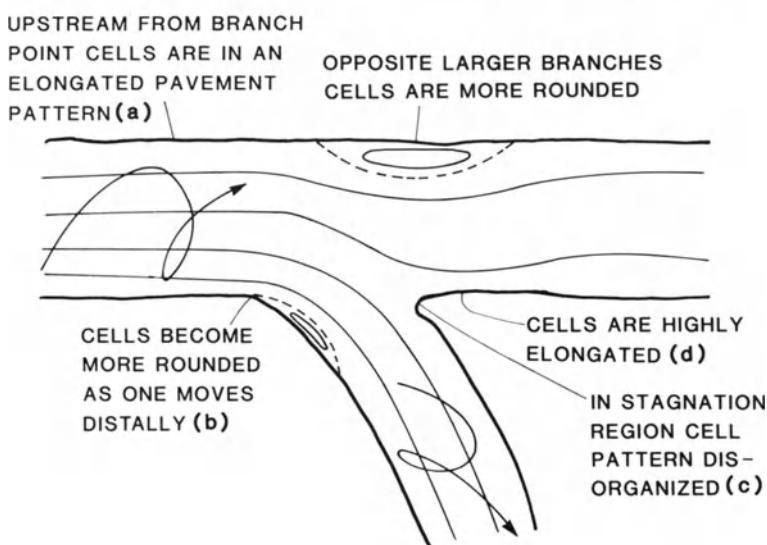
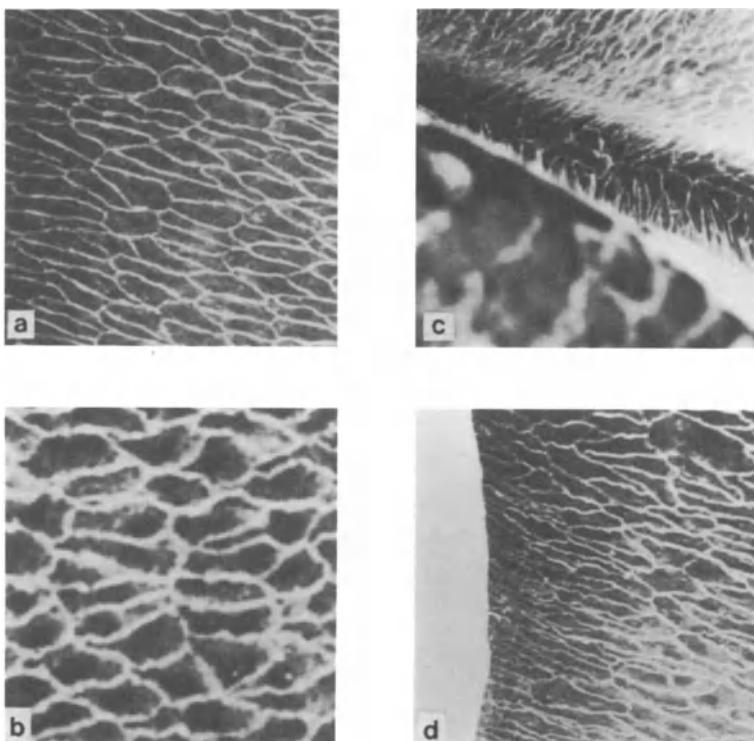


FIGURE 23.6. Illustration of an aortic flow in the vicinity of a branch vessel with selected notations on endothelial cell characteristics and with example scanning electron micrographs from positions around the rabbit aorta intercostal orifice as indicated.

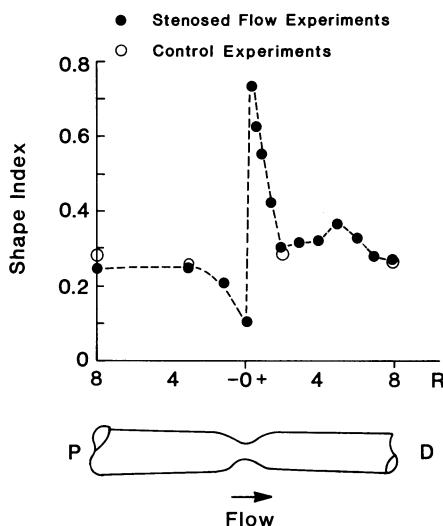


FIGURE 23.7. Endothelial shape index as a function of position along a dog aortic stenosis for both stenosed flow conditions and control, unstenosed aortae.

to control conditions. This then represents an experiment where, by changing the geometry of the aorta to produce a specific hemodynamic situation, the resulting endothelial cell pattern is consistent with what one would expect based on the fluid dynamics of a stenosed flow.

Based on the cell culture studies as well as the above, there seems to be a clear influence of fluid mechanical forces on the geometry and orientation of endothelial cells. But what is the mechanism by which any fluid-mechanical effect on the biology of the endothelium becomes manifest? Is it a hemodynamic-related effect on transendothelial transport? Is it an effect on endothelial endocytosis? Is it a response of the cell to mechanical stimuli? How are endothelial cell function and transendothelial permeability related to one another? If it is a transport problem, what are the transport mechanisms? Assuming that the intercellular junctions are too tight to allow for the passage of high-molecular-weight macromolecules, how do lipids and other such molecules move across the endothelium from blood into the arterial wall? One possibility is some vesicular process in which material at the lumen side of the endothelial cell is taken into the cell by an invagination of the cell membrane. The resulting vesicle then detaches and through Brownian motion moves across the cell, and at the other side, the vesicle attaches and the particle is released. The presence of vesicles in endothelium have been observed (Simionescu et al, 1974); however, Weinbaum's (1983) calculations suggest that vesicular transport may not be the answer.

Weinbaum, in fact, has investigated a number of possibilities for the movement of molecules across the endothelial interface. Of particular interest has been his theoretical calculations of the effect of a damaged cell and the influence of the concomitant "leaky" junctions on the enhancement of macromolecule transport. These results indicate that as the number of damaged cells

increases from 1 in 1,000 to 5 in 1,000, the flux or uptake ratio increases by 67 percent. This suggests that the endothelium indeed can have a rate-controlling effect on the transport of macromolecules, as had earlier been indicated by the experiments of Caro and Nerem (1973), but one that is related to cell damage or turnover and death (i.e., the life expectancy of a cell) and as such, possibly to cell function.

In fact, an extremely important question is what controls the life expectancy of a cell. What is the aging process of a cell? The data of Caplan and Schwartz (1973) suggest that endothelial cell turnover rate may be different in regions that are believed to be different in shear. These results indicate that in regions of more rounded cells, which we now believe to be associated with low shear, there will be enhanced endothelial cell turnover.

In summary, there thus is a body of data indicating that endothelial cell geometry and orientation are strongly influenced by hemodynamic forces. If the functioning of a cell is influenced by its shape, then this could be a means whereby a hemodynamic effect becomes manifest. Furthermore, there may be additional influences on the biology of the endothelium that may lend themselves to any interpretation of the relationship between cell function and cell shape.

#### 4. Vascular Geometry, the Endothelium, and the Localization of Atherogenesis

It is appropriate at this point to examine the relationship of endothelial changes to the pattern of early lesions found in atherosclerosis and in the context of our knowledge of arterial fluid mechanics. Although our knowledge is too limited—perhaps the interactions are too complicated—to do this in a general way, some tentative speculations are offered here.

First, as already noted, it appears that the endothelium is intact, with desquamation occurring without denudation. The endothelial cells are of a teardrop shape, aligned with the direction of the blood flow, and more elongated in regions of high shear and more rounded in regions of low shear. When vessels branch, the flow patterns may be quite complicated, and the cells will not necessarily be aligned with the axis of the blood vessel. Furthermore, cells will be more rounded in some regions than in others, and this may be of some importance, since the recent paper by Repin et al (1984) suggests that atherosclerosis in humans is more likely to occur in regions having more rounded endothelial cells. The shape of the endothelial cells will be related to local flow conditions, as already noted; however, there will be other influences on cell geometry. For example, using the rat, we have studied the influence on arterial endothelial cell morphology of hypercholesterolemia, diabetes, and the combined effect of diabetes and hypercholesterolemia. The results obtained show that induced diabetes (streptozotocin injection) and diet-induced hypercholesterolemia (2% cholesterol diet) have an effect on cell morphology.

Endothelial cells from diabetic and hypercholesterolemic rats had a smaller surface area, and were shorter and rounded when compared to controls. The effect was more pronounced when diabetic rats were subjected to a hypercholesterolemic diet. There we observed a greater decrease in surface area and a larger increase in the shape index of the cells toward a significant rounding.

Although there is limited information on the pattern of early lesions for spontaneous disease, the existence of a general pattern indicating low shear regions to be of higher predilection has been confirmed in general in studies of the adult human (Caro et al, 1971; Nerem and Levesque, 1983). It again is also consistent with the suggestion that regions of more rounded cells have a much higher predilection for atherogenesis. More rounded cells do appear to be associated with regions of low shear, and thus one can hypothesize the following sequence of events as being intrinsic to the process of atherogenesis. In regions of low shear, which appear to have a higher predilection for the disease, the endothelium may be characterized as having comparatively rounded cells. These cells possibly have larger gaps around their endothelial cell borders that may make possible and enhance lipid transport. These cells also appear to have a shorter lifetime (i.e., a higher cell turnover rate) which also may be associated with the influence of hemodynamic forces or other related events on cell function or dysfunction. During cell turnover, the associated intercellular junctions may become leaky, allowing for an enhanced influx of lipids and other macromolecules. Thus, regions of low shear and the concomitant rounder endothelial cells, either because of a larger intercellular gap or a higher cell turnover rate, demonstrate a greater permeability. It is this enhanced permeability and the associated increased influx of lipid macromolecules that we believe to be an essential ingredient in the initiation of the disease atherosclerosis.

However, in addition to the multitude of questions relating to any hemodynamic influence on endothelial cells, there are two cautionary notes. First, the pattern of lesions in diet-induced atherosclerosis in animals is in general not in agreement with the concept of low shear regions having an enhanced predilection for atherogenesis. The same may be said for the pattern of early lesions in fetuses, neonates, and infants. Although this can possibly be explained by the fact that the lipid environment, and thus the disease, is not the same as in the adult human case, this remains to be further investigated. Secondly, although the evidence surveyed here provides an indictment of low shear regions as having the highest predilection for early lesions, it must be reemphasized that our knowledge is still too incomplete to say that with any certainty. In fact, whether it is shear, or rapid changes in shear as reflected in the shear pulse amplitude, or some other related hemodynamic event remains to be demonstrated in a more definitive fashion. Studies of cultured endothelial cells, as noted earlier, suggest that it is cells undergoing a change in mechanical environment that demonstrate enhanced activity as measured by such variables as migration, mitosis, HRP endocytosis, and PGI<sub>2</sub> production.

If this is the case, then perhaps regions characterized as low shear should be looked at as regions of greatest alternating shear stress. This would be true because, for a given flow pulsatility, there would be a greater possibility of flow reversal—and thus the presence of an oscillating or alternating shear stress—in a region of low mean shear stress than in a region of a high mean shear stress. What we do not know are the mechanisms by which cell function is influenced by such a mechanical event and what the time constants are for the various cellular processes that may be of interest. We also do not know the answer to such basic questions as why cell turnover is different in regions of low shear versus high shear—i.e., what determines the lifetime of a cell?

Finally, much of the evidence for a hemodynamic involvement in atherosclerosis can be construed to be in fact geometry-related. This thus suggests that vascular geometry, and the way it influences vascular dynamics and the local detailed flow properties, may be the primary determinant of a hemodynamic effect in the initiation of atherosclerosis. Under such a hypothesis, whatever affects vascular geometry would alter the local, detailed flow characteristics and correspondingly influence endothelial morphology and function—i.e., it would alter vascular dynamics at all levels, including that of the endothelial cell. For an individual subject, that geometry might be inherited, as with a family history or genetic effect, and as such one's geometry may in and of itself be a risk factor. Alternatively, a subject's vascular geometry may have been significantly altered by a risk factor such as hypertension. However, whatever the effect, the primary determinant of the pattern of atherosclerosis would be vascular geometry, and whether or not an individual has an abnormal predilection for the disease would depend not only on biochemically related factors, but also on the exact nature of the geometry, for example, branching pattern and degree of curvature and/or tortuosity, the resulting detailed hemodynamic characteristics, and what happens at the level of an endothelial cell.

The rather subtle influence of hemodynamics on the endothelium that is hypothesized here is not meant to totally rule out a more traumatic effect such as endothelial cell denudation; however, such gross injury would appear to be nonphysiologic and it is not likely to occur under normal flow conditions. It is also not meant to rule out a role for thrombosis in atherosclerosis. In fact, it is believed that thrombosis is a very important factor in the later stages of the disease. It similarly is not meant to deny the importance of platelets, since transient platelet—endothelium encounters could possibly release platelet factors and stimulate smooth muscle cell proliferation.

A few years ago, the prevailing view was that if one could produce a statistically significant picture of the pattern of the disease and a correspondingly statistically significant picture of the flow pattern, then one could draw conclusions on the relationship of hemodynamics to atherosclerosis. However, it now appears that it is not the normal pattern of disease (in the sense of mean data representing the statistically normal situation) nor the normal flow pattern that are of interest. Rather, it is deviations from the normal that are

important, and it is in this context that vascular geometry and vascular dynamics may be important. From this point of view, vascular geometry, whether inherited or as altered by the influence of a factor like elevated blood pressure, would determine, at least in part, the vascular dynamics of the system, in terms of local, detailed hemodynamics, and at the level of the endothelium. The latter, together with other biochemically related factors, and when deviating from the normal, may represent the subtle injury that we believe is the genesis of atherosclerosis.

## Acknowledgments

The authors are appreciative of the support provided by the National Science Foundation through Grant MEA-8200344 and the National Institutes of Health through Grant HL-26890.

## References

- Bergel, D.H., ed. 1972. *Cardiovascular fluid dynamics*. Vols. 1 and 2. Academic Press, New York.
- Caplan, B.A., and Schwartz, C.J. 1973. *Atherosclerosis* 17:401–417.
- Caro, C.G., and Nerem, R.M. 1973. *Circ. Res.* 32:197–205.
- Caro, C.G., Pedley, T.J., Schroter, R.C., and Seed, W.A. 1978. *The mechanics of the circulation*. Oxford University Press, Oxford, England, U.K.
- Dewey, C.F., Jr. 1984. *ASME J. Biomech. Eng.* 106:31–35.
- Dewey, C.F., Jr., Bussolari, S.R., Gimbrone, M.A., Jr., and Davies, P.F. 1981. *ASME J. Biomech. Eng.* 103:177–185.
- Evans, E.A., and Skalak, R. 1979. *Mechanics and thermodynamics of biomembranes. CRC critical reviews in bioengineering*. Vol. 3, issues 3 and 4. CRC Press, Inc., Boca Raton, Florida.
- Frangos, J.A., McIntire, L.V., Eskin, S.G. and Ives, C.L. (1985) Effects of flow on prostacyclin production by cultured human endothelial cells. *Science* 227:1477–1479.
- Fry, D.L. 1968. *Circ. Res.* 22:165–197.
- \_\_\_\_\_. 1969 *Circ. Res.* 24:93–108.
- Fung, Y.C. 1981. *Biomechanics*. Springer-Verlag, New York.
- \_\_\_\_\_. *Biodynamics: circulation*. Springer-Verlag, New York.
- Gimbrone, M.A., Jr. 1976. Pages 1–28 in *Progress in hemostasis and thrombosis*. Vol. 3, T.H. Spaet, ed. Grune and Stratton, New York.
- McDonald, D.A. 1974. *Blood flow in arteries*. 2d d. Edward Arnold, London.
- Nerem, R.M. 1981. Pages 719–835 in *Structure and function of the circulation*, C.J. Schwartz and N. Werthessen, eds. Plenum Press, New York.
- Nerem, R.M., and Levesque, M.J. 1983. Pages 26–37 in *Fluid dynamics as a localizing factor in atherosclerosis*. Symposium proceedings. G. Schettler, et al, eds. Springer-Verlag, Heidelberg.
- Nerem, R.M., Levesque, M.J., and Cornhill, J.F. 1981. *ASME J. Biomech. Eng.* 103:172–176.

- Repin, V.S., Dolgov, V.V., Zaikina, O.E., Novikov, J.D., Antonov, A.S., Nikolaeva, M.A., and Smironov, V.N. 1984. *Atherosclerosis* 50:35-52.
- Schettler, G., Neren, R.M., Schmid-Schönbein, H., Mörl, H., and Diehm, C., eds. 1983. *Fluid dynamics as a localizing factor in atherosclerosis*, Springer-Verlag, Heidelberg.
- Simionescu, M., Simionescu, N., and Palade, G.E. 1974. *J. Cell Biol.* 60:128-152.
- Vaishnav, R.N., Patel, D.J., Atabek, H.B., Deshpande, M.D., Plowman, F., and Vossoughi, J. 1983. *ASME J. Biomech. Eng.* 105:77-83.
- Weinbaum, S. 1983. Pages 92-114 in Ann. of New York Academy of Sciences: Surface Phenomena in Hemorheology: Their Theoretical, Experimental and Clinical Aspects. Vol. 416. A.L. Copley, GVF Simon, eds.
- White, C.E., Fujiwara, K., Shefton, E.J., Dewey, C.F., Jr., and Gimbrone, M.A., Jr. (1982) *Federation Proc.* 41:321 (abst.).

# 24

## Self-excited Oscillation of a Collapsible Tube Conveying Fluid

Y. MATSUZAKI

### 1. Introduction

This chapter reviews recent developments in research on the oscillation of collapsible tubes and their internal flow. The oscillations under consideration are not those caused by pulsation, but rather the self-excited ones that are induced by interaction between the overall tube deformation and fluid pressure acting on the internal wall of the tube. Only if certain critical conditions are satisfied, will the tube wall and flow lose their stability to vibrate together with finite amplitudes. Such an oscillatory phenomenon also takes place in external flows. Fluttering of flags in the wind is one of the examples found in everyday life. In engineering, airplane wings and chemical plant pipings, etc., must be designed to avoid the occurrence of oscillation due to aero- or hydroelastic interaction because it causes immediate destruction of airplanes and imperils normal operation of plants. In aeronautics, especially, a large number of theoretical and experimental investigations in aeroelasticity have been made since the end of World War I, and these investigations are actively continuing. Several complete books are currently available. Among them, *Theory of Aeroelasticity* by Fung (1955, 1969) is one of the “bibles” for aeronautical engineers and researchers.

In physiological and medical areas, similar oscillatory instability may occur between biological tubes and fluids. For example, the interactions between blood vessels and blood, airways and air, and urethras and urine, are considered to be responsible for Korotokoff sound in arterial blood vessels, and flow limitation during forced expiration and micturition, respectively. Such biological tubes are very flexible and are even completely collapsible. In addition, the biological flows usually have much lower Reynolds numbers than flows in engineering pipes, so that the effect of viscosity prevails in the former. Biological phenomena may therefore become much more complex than those encountered in engineering.

## 2. Oscillation During Heart Surgery

Before focusing on recent studies, we shall introduce a typical example that is observable in a critical area of clinical treatment: oscillations of the venae cavae and blood flow during heart bypass surgery. These oscillations were reported by a group of surgeons of Tokyo Women's Medical College (Tsuji et al, 1978). In order to withdraw the blood for gas exchange in an oxygenator, two rigid cannulas penetrating the wall of the right ventricle were inserted into the superior and inferior venae cavae (SVC and IVC). The blood stream from SVC and IVC joined at a plastic junction pipe and flowed into a long, flexible, but noncollapsible tube connecting to the oxygenator. The oxygenated blood was sent back into a femoral artery by a roller pump. During the operation, flow from SVC and IVC was often controlled by means of suction pressure and constriction by a clamp placed on the tube downstream of the junction pipe, while the rolling rate of the roller pump was fixed. With increasing suction pressure, the flow rate was increased to a point at which the venae cavae walls and flow began to vibrate. Further increase in the suction pressure resulted in the growth of oscillatory amplitudes but no variance in the mean flow rate. The flow rates in SVC and IVC,  $Q_S$  and  $Q_I$ , were measured. Figure 24.1 presents a record of the flow rates  $Q_S$  and  $Q_I$  during an operation on a three-year-old girl. Limit-cycle oscillations of the flow rates are clearly seen in the first third of the record. The frequency of oscillation was 2.6 Hz. In regions A

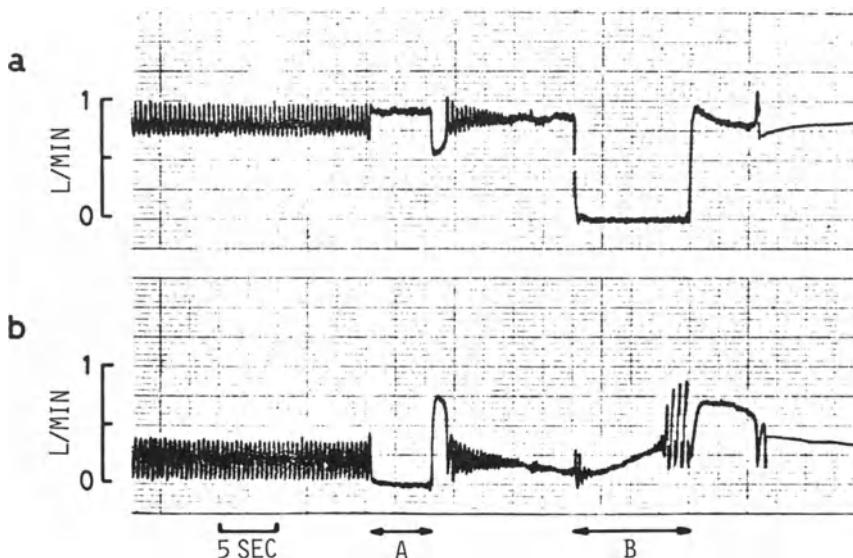


FIGURE 24.1. A record of the blood flow rates during heart surgery. (a)  $Q_S$ . (b)  $Q_I$ . From Matsuzaki (1980), by courtesy of Dr. T. Tsuji.

and B, IVC and SVC were, respectively, squeezed in order to examine the effect of flow blockage. In region A, where  $Q_1$  was set to zero,  $Q_s$  went up to a higher value instantaneously and the oscillation disappeared. On the other hand, when the flow was blocked in SVC,  $Q_1$  increased gradually to a point and then started to vibrate more violently, as shown in region B. In order to stop this vibration, the blockage was removed at the end of region B. Then the suction pressure was decreased before the oscillation was established again, as in the first third of the record. In addition, the rolling rate of the pump was increased in order to ensure a required blood flow without oscillation.

According to the report by Tsuji et al, 22 of 52 patients experienced the oscillations, which continued for 3 to 36 minutes (11 minutes on an average) during extracorporeal perfusion of 35 to 157 minutes. The patients having long durations of oscillation suffered from severe hemolysis. In addition, the renal function was considered to degenerate because the longer the vibration continued, the lower the mean urine rate became during the vibration. It was therefore concluded that the oscillation must be avoided in order to perform the operation safely. In order to maintain a sufficient flow rate without unfavorable oscillation, it is essential to understand what kind of mechanism underlies this oscillatory phenomenon.

### 3. Experiments on Rubber Tubes Conveying Water

In order to systematically examine the static and dynamic behavior of biological tubes conveying fluids, experimental studies were performed by many investigators using elastic tube models and water. Among them are Conrad (1969), Conrad et al (1978), Katz et al (1969), Griffiths (1977), Bertram (1982) and Ohba et al (1984). The experiments carried out by Ohba and his co-workers were the most extensive and revealed that tube-flow characteristics were more complicated than had been shown in the previous studies. Figure 24.2 depicts a schematic diagram of their experimental setup, which is similar to those used by other investigators. A collapsible tube was horizontally mounted between two rigid pipes in an airtight box where the external static pressure  $p_e$  acting on the tube was controlled. Measurements were made on the flow rates and static pressures upstream and downstream  $Q_1, Q_2, p_1$  and

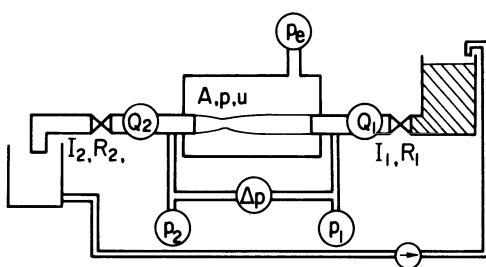


FIGURE 24.2. A schematic diagram of the experimental setup.

$p_2$ , and the pressure drop  $\Delta p (= p_1 - p_2)$  along the tube. The pressures  $p_1$ ,  $p_2$ , and  $p_e$  were referred to the atmospheric pressure. The models used were a Penrose tube of 0.15 mm thickness and a natural rubber tube of 1.2 mm thickness that spanned 160 mm between the rigid pipes. The external diameter was 6 mm for both tubes. The flow exit system consisted of a flowmeter, a pressure tap, and rigid pipes. As the very end of the outlet pipe was in the atmosphere, a positive value for  $p_2$  represented a decrease in the static pressure of the flow along the exit system.

In order to examine the effect of peripheral (downstream) condition on the tube-flow characteristics, three outlet pipes, called types a, b and c, were used. The pipe length and inner diameter were, respectively, 1600 mm and 4 mm for type a, 870 mm and 6 mm for type b, and 150 mm and 6 mm for type c. The relationships between  $p_2$  and  $Q$  are illustrated in Figure 24.3. When a pipe of type a or b was used,  $p_2$  increased with increasing  $Q$ . For the case of type c, however, the characteristics were quite opposite:  $p_2$  decreased as  $Q$  increased, so that  $p_2$  was negative. Because the external pressure on the tube was kept nonnegative, i.e.,  $p_e \geq 0$ , the transmural pressure  $p_2 - p_e$  at the downstream end of the flexible tube was negative. This corresponded to the fact that the tube equipped with the pipe of type c remained at least partially buckled over the entire flow range measured.

Figure 24.4 shows the time-averaged  $\Delta p$ - $Q$  characteristics of the natural rubber tube with the three outlet pipes for various external pressures  $p_e$ . Solid curves represent theoretical  $\Delta p$ - $Q$  relationships that were obtained when a rigid tube of the same internal diameter was assumed to replace the collapsible tube. In the  $p_e$ -constant measurement except for  $p_e = 0$ , the oscillation occurred in the flow ranges that are demarcated by a set of two arrows, as shown in Figure 24.4a and b, or in the ranges that are higher than the flow rate indicated by an arrow, as shown in Figure 24.4c.

The  $\Delta p$ - $Q$  characteristics plotted in Figure 24.4a are essentially the same as

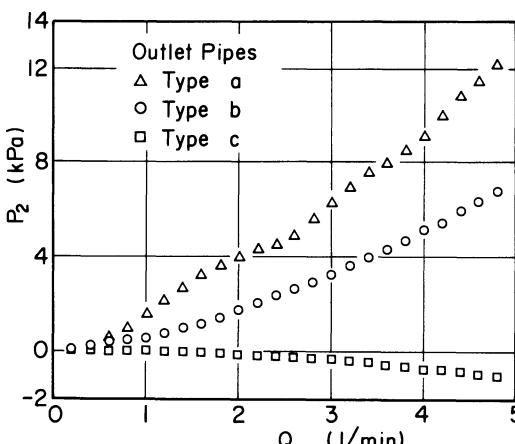


FIGURE 24.3. Peripheral pressure  $p_2$  vs. flow rate  $Q$  for three different types of outlet pipes. Redrawn from Ohba et al (1984), by permission.

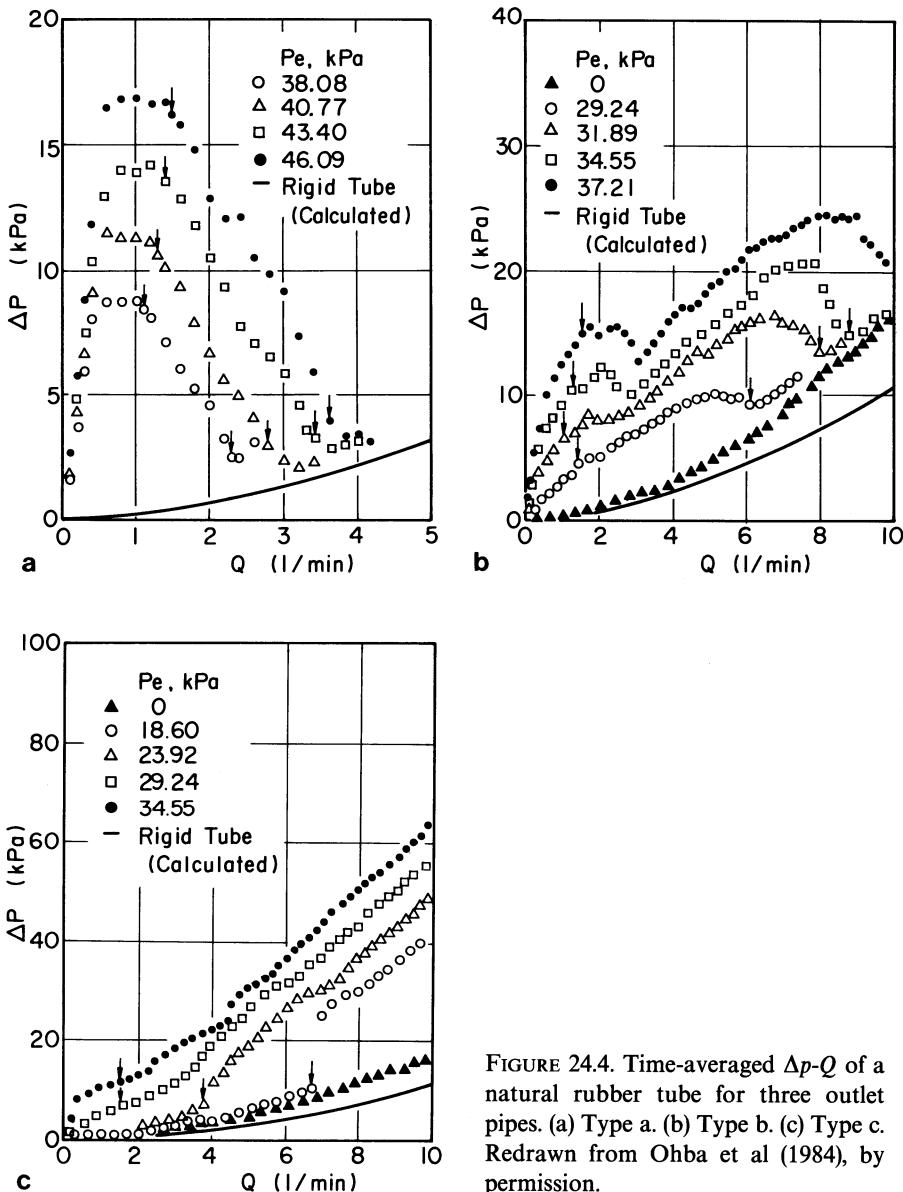


FIGURE 24.4. Time-averaged  $\Delta p$ - $Q$  of a natural rubber tube for three outlet pipes. (a) Type a. (b) Type b. (c) Type c. Redrawn from Ohba et al (1984), by permission.

those reported by Conrad (1969) and Katz et al (1969). The oscillation took place in the so-called *negative* slope region, where  $d(\Delta p)/dQ < 0$ . However, the oscillatory characteristic aspects illustrated in Figure 24.4b and c are different from those observed in most of the previous experiments. In other words, the vibration was also induced in the *positive* slope. The occurrence of oscillation in the positive slope region was also reported by Bertram (1981). It is very clear that the peripheral conditions have a strong influence on the static and

dynamic behaviors of the tube and flow. The significance of peripheral conditions was first pointed out by Conrad et al (1978). As for the Penrose tube, the characteristics were similar to the corresponding ones of the natural tube, except in a very low  $Q$  range, where no detectable flow occurred unless  $\Delta p$  exceeded a certain critical value. This phenomenon is considered to be the same as found by Conrad et al (1980) using a canine femoral artery.

In addition, Ohba et al (1984) made detailed measurements of the relationship between the flow rate and frequency of oscillation, and also distinguished (by using high-speed cameras) two different modes of oscillation: one for lower flow rates and the other for higher flow rates. To the author's knowledge, a paper by Weaver and Paidoussis (1977) was the first to describe two types of oscillations of a tube conveying flow. The descriptions of the two oscillatory patterns by Weaver et al, Bertram, and Ohba et al do not necessarily seem to agree with each other. This suggests that there may be several different mechanical factors that cause the oscillatory instabilities.

#### 4. Theoretical Analysis

As seen in the preceding sections, the vibration characteristics observed *in vivo* and *in vitro* are extremely complicated. Prerequisite to a theoretical investigation of this hydroelastic unstable phenomenon are three subanalyses on finite-amplitude vibration of a buckled tube, three-dimensional unsteady viscous flow in an oscillating tube, and stability of nonlinear oscillation. Each of the three represents an interesting but difficult problem of applied mechanics. Especially, the first two mentioned, the structural and fluid-mechanical problems, have not fully been investigated so that much remains to be done both theoretically and experimentally. So far, only simple theoretical hydroelastic models have been presented in the study of the characteristic features of the initiation and continuation of the self-excited vibration.

Here, we will briefly mention a lumped-parameter model analysis by Pedley (1980), and Bertram and Pedley (1982) that appears to be successful in simulating some experimental observations. The collapsible finite-length tube conveying fluid is reduced to a one-degree-of-freedom system. The single variable selected is the cross-sectional area,  $A(t)$ , at the throat of the collapsed tube. It is assumed that the elastic property of the whole tube is expressible in terms of a simple relationship between  $A$  and the transmural pressure  $p - p_e$ :

$$p - p_e = P(A), \quad (24.1)$$

where  $p$  is the pressure of the flow at the throat (Figure 24.2).

Pedley (1980) asserted that during oscillation the flow volume,  $V$ , changes mainly upstream to the throat and scarcely downstream, and assumed that the flow rate through the throat,  $uA$ , is equal to the flow rate downstream,  $Q_2 = u_2 A_2$ , and that the equation of conservation of flow mass at the upstream segment is expressible as

$$Q - uA = \dot{V} = l\dot{A}_v, \quad (24.2)$$

where  $Q = u_1 A_1$ , and  $l$  and  $A_v (= (A_1 + A)/2)$  are, respectively, the length and average cross-sectional area of the upstream segment. At the upstream convergent segment, the energy balance is considered, including the effects of the viscous resistance and fluid inertia. The resistance is approximated by the Poiseuille law, and the inertia term is associated with the rate of change of the average velocity  $u_v$ , which corresponds to  $A_v$ , where  $u_v A_v = (Q + Q_2)/2$ . It is well known that a divergent flow is easily separated and becomes unsteady even though the divergent angle is not large. One of the most simplified assumptions of the analysis is that the pressure loss at the downstream divergent segment is predicted by the momentum equation for the steady flow:

$$p - p_2 = -\rho(A_2/A - 1)u_2^2 \quad \text{for } A < A_2, \quad (24.3)$$

where  $\rho$  is the density of fluid. As for the flow in the inlet and outlet rigid pipes, the momentum equations are formulated, taking into account the flow inertias  $I_1$  and  $I_2$  and resistances  $R_1$  and  $R_2$ .

These relationships are put all together to obtain a third-order differential equation with respect to the representative cross-sectional area,  $A$ . The numerical analysis was made in two ways: a stability examination using Routh's criterion with respect to a small disturbed motion about the static equilibrium cross-sectional area,  $A_0$ , and a direct time-wise numerical integration of the differential equation. The numerical result for a moderate value of  $R_2$  gave a static  $\Delta p$ - $Q$  curve similar to those shown in Figure 24.4a, and the limit-cycle solution existed in both the positive and negative slope ranges. When  $R_2$  was set to a small value, the flow characteristics predicted appeared to correspond to those illustrated in Figure 24.4c. The conclusion reached is that major mechanical factors responsible for the oscillation are the energy dissipation due to flow separation at the divergent segment and the properties of flow in the inlet and outlet rigid pipes.

As for a cylindrical shell model, several investigators predicted the occurrence of postbuckling oscillation with the aid of cylindrical shell equations and potential flow theories, i.e., inviscid flow assumptions. It has been shown, however, that no limit-cycle solution is predicted to exist when the damping and geometrical nonlinearity of the elastic wall are also taken into account in the analysis based on the inviscid flow approximation (Matsuzaki, 1985). As indicated in the lumped-parameter model analysis, the energy loss due to viscosity in the collapsed tube is the most essential factor for the oscillatory instability.

## 5. Flow Separation in the Divergent Segment

The flow characteristics in the inlet and outlet pipes can be evaluated quantitatively. However, there is neither general theory nor experimental data for quantitative prediction of viscous flow through an oscillating collapsed tube. In the divergent segment, the flow may separate to become randomly un-

steady, although the inlet flow is steady. No flow characteristics are fully expressible in terms of the simple formula given by Equation (24.3). According to the experiment on flow in a two-dimensional divergent channel of fixed rigid walls (Matsuzaki and Fung, 1976), the static pressure downstream increases with increasing divergence angle as long as the flow remains unseparated. On the other hand, in separated flow, the time-averaged static pressure decreases as the divergence angle increases. This means that the effect of change in divergence angle on the pressure loss is opposite between separated and unseparated flows. If the walls of the channel are forced to vibrate, then a phase lag occurs between the wall motion and pressure response. Since the phase lag generally plays a significant role in dynamic instability, it is important to investigate the dynamic relationship between the wall motion and pressure and to incorporate it into the analysis instead of using Equation (24.3).

## 6. Conclusion

The static and dynamic behaviors of the collapsible tubes that convey fluid, observed *in vivo* and *in vitro*, are very complicated. As for the self-excited oscillation, qualitative understanding of some basic mechanisms may be obtained by using a theoretical analysis based on a lumped-parameter representation of the tube-flow system. Although many theoretical and experimental investigations have been performed, mentioning has been made only of recent publications together with closely related papers. For further study, the textbook by Pedley (1980) is recommended.

## References

- Bertram, C.D. 1982. *J. Biomech.* 15(3):223–224.
- Bertram, C.D., and Pedley, T.J. 1982. *J. Biomech.* 15(1):39–50.
- Conrad, W.A. 1969. *IEEE Trans. Biomed. Eng.* 16(4):284–295.
- Conrad, W.A., Cohen, M.L., and McQueen, D.M. 1978. *Med. Biol. Eng. Comput.* 16(3):211–214.
- Conrad, W.A., McQueen, D.M., and Yellin, E.L. 1980. *Med. Biol. Eng. Comput.* 18(7):419–426.
- Fung, Y.C. 1955. *An introduction to the theory of aeroelasticity*. John Wiley & Sons, New York. Also 1969, rev. ed. Dover, New York.
- Griffiths, D.J. 1977. *Med. Biol. Eng. Comput.* 15(7):357–362.
- Katz, A.I., Chen, Y., and Moreno, A.H. 1969. *Biophys. J.* 9:1261–1279.
- Matsuzaki, Y. 1980. Pages 226–243 in *Cardiovascular dynamics and basic measurements* (in Japanese), H. Okino, M. Sugawara, and H. Matsuo, eds. Kodansha Scientific, Tokyo.
- . 1985. chapter 15 in *Encyclopedia of fluid mechanics*, vol. 1. N.P. Cheremisinoff, ed. Gulf Publishing, Houston.
- Matsuzaki, Y., and Fung, Y.C. 1976. *J. Appl. Mech.* 98(2):227–231.

- Ohba, K., Yoneyama, N., Shimanaka, Y., and Maeda, H. 1984. *Tech. Reports Kansai Univ.* 25:1–13.
- Pedley, T.J. 1980. *The fluid mechanics of large blood vessels*, chapt. 6. Cambridge University Press, Cambridge, England, U.K.
- Tsuji, T., Nakajima, K., Takeuchi, Y., Inoue, K., Shiroma, K., Koyama, Y., Tokuchi, K., Yoshikawa, T., and Suma, K. 1978. *Artificial organs* (in Japanese) 7(2):435–438.
- Weaver, D.S., and Paidoussis, M.P. 1977. *J. Sound Vibration* 50(1):117–132.

## **Part V   The Lung**

# 25

## Microvascular Hematocrit of the Lung

J.S. LEE

### 1. Introduction

Microvascular or capillary hematocrit plays an important role in determining the flow resistance and oxygen uptake in pulmonary capillaries. We shall review in this chapter the methods to quantify the microvascular hematocrit of the lung and to provide a critique on the measurement methods or their interpretation. The methods are: morphometric measurement of capillary hematocrit, volume dilution of tagged red blood cells (RBC) and albumin in the lung, dilutions of RBC and plasma indicators in pulmonary blood flows, dilution of density indicators from pulmonary capillaries, and the Fahraeus effect in small tubes. Finally, the physiological significance of a capillary hematocrit ( $H_c$ ) lower than the arterial hematocrit ( $H_a$ ) is discussed.

### 2. Morphometric Measurements of Capillary Hematocrit

A direct method to quantify the capillary hematocrit is to measure morphometrically the amount of red blood cells and plasma in the capillary. In the morphometric studies of Weibel's group, canine lungs were instilled with a 2.5% solution of buffered glutaraldehyde at a pressure 22 torr above the trachea. The heart and lung were removed from the dog and submerged in fixative for two hours. From each lobe, a slice of lung about 1 to 2 mm thick was resected and diced into small cubes of 2 to 3 mm<sup>3</sup>. Sampling from all five lobes of the lung, the researchers selected a total of 20 to 40 blocks for embedding. Eighty-eight random electron micrographs were analyzed. A lattice of  $P_t$  test points was superimposed on the section as shown in Figure 25.1. If, for each micrograph,  $P_c$  of these points were found to be located with the blood in the capillaries, then the average of  $P_c/P_t$  from all these two-dimensional micrographs was, according to the Delesse principle, an unbiased estimate of the fraction of capillary blood volume, a quantity describing a three-dimensional structure. Let  $P_r$  be the number of points falling on red blood cells for each micrograph. Then the capillary hematocrit,  $H_c$ , can be

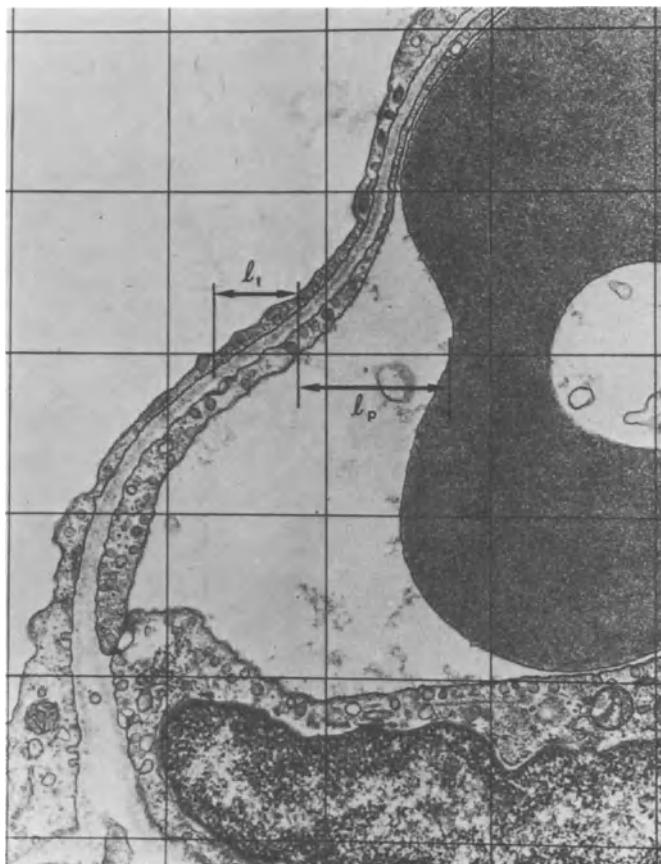


FIGURE 25.1. A portion of the micrograph with a lattice of points superimposed onto the alveolar structure for the determination of capillary hematocrit. From Weibel (1971), and courtesy of Dr. E.R. Weibel and *Respiration Physiology*.

estimated as  $\overline{Pr}/\overline{Pc}$ . Siegwart et al (1971), using this point count method, reported that the capillary hematocrit was 50%.

The blood can flow from arteries into capillaries during the fixation process. To investigate this factor, Bur and his co-workers (personal communication) stopped the circulation by tying the base of the heart and then instilled the lung of a rat with glutaraldehyde. From their morphometric estimation, they reported a capillary hematocrit of 39%. In contrast, the arterial hematocrit in the left ventricle of other rats was 44%. The capillary hematocrit is 88% of the arterial hematocrit. (The rat lungs, without their circulation being stopped before instillation, had a hematocrit of 52%).

A rapid freezing technique was used by Vreim and Staub (1973) for quantifying the pulmonary capillary volume. Liquid propane was poured onto the

lungs continuously until they were frozen solid. Sample blocks were cut from the upper and lower lobes for uniform representation of the entire lung. After embedding, sections were cut from tissue within 5 mm of pleural surface. The alveolar surface area was measured by a linear method, and the alveolar wall thickness, with an image-splitting micrometer. The volume of the alveolar wall, calculated as one half the product of the surface area and the thickness, was found to be 8.1 ml. By the point count method, the RBC fraction in alveolar wall was 19%.

These measurements yield an RBC volume of 1.54 ml ( $= 8.1 \text{ ml} \times 19\%$ ). Using the measured arterial hematocrit of 34% as the capillary hematocrit, Vreim and Staub (1973) estimated the capillary blood volume to be 4.5 ml ( $= 1.54 \text{ ml}/34\%$ ). Considering the alveolar wall as an alveolar sheet having a vascular space tissue ratio (VSTR) of 0.91 as measured by Sabin et al (1970), Vreim and Staub obtained a capillary blood volume of 7.4 ml ( $= 8.1 \text{ ml} \times 0.91$ ). Thus they suggested that 61% ( $= 4.5 \text{ ml}/7.4 \text{ ml}$ ) of the capillary space was filled with arterial blood. The other 39% was stationary and free of RBC.

A different interpretation of their data should be considered. It is possible that all the capillary blood volume (7.4 ml) was filled with flowing blood of a lower capillary hematocrit as the lung was frozen. In reference to the measured RBC volume, we estimate that the capillary hematocrit is 21% ( $= 1.54 \text{ ml}/7.4 \text{ ml}$ ). This capillary hematocrit is 61% ( $= 21\%/34\%$ ) of the arterial hematocrit. It is noted that the capillary hematocrit can be identified simply as the ratio of RBC fraction and VSTR.

Using the indirect method of carbon monoxide uptake, Vreim and Staub (1973) reported that the blood volume of pulmonary capillaries was 3.5 ml. Because of the affinity of carbon monoxide with hemoglobin, one should use the rate constant of red blood cells instead of blood to calculate the uptake of carbon monoxide in pulmonary capillaries. Then the indirect method predicts the RBC volume in the capillaries which, for their experiment, is 1.2 ml ( $= 3.5 \text{ ml} \times 34\%$ ). To fill the capillaries partially with arterial blood or fully with blood of low hematocrit becomes irrelevant to the comparison between the direct and indirect estimate of RBC volume. A larger direct estimate (1.54 ml) is explained by methodological limitations.

One way to examine the error in the morphometric estimate of capillary hematocrit is to evaluate the potential errors inherent in each method. First, one needs to arrest the capillary blood flow right at the beginning of the fixation process. With no additional red blood cells or plasma coming into the capillary, the capillary hematocrit can be identified as the tube hematocrit correspondent to the flowing condition. The rapid freezing technique may be a good process. Second, systemic error could be introduced by dimensional changes of tissue during the preparation. Freeze-drying produced a 6% linear shrinkage and hence 19% volume shrinkage of the tissue. If the relative shrinkage is assumed to be uniform for plasma, red blood cells, and the tissue (although we do not know if it is), then the dimensional change will not affect the estimation of RBC fraction, VSTR, and hence capillary hematocrit. Third-

ly, the angle of sectioning is not always perpendicular to the alveolar wall. The alveolar wall thickness may be overestimated. On the other hand, the estimation of the fraction of RBC volume is not affected, according to the Delesse principle, by the angle of sectioning if the samples are randomly selected.

### 3. Volume Dilution of Tagged Red Blood Cells and Albumin

A volume dilution procedure was used by Aarseth et al (1978) to measure the whole-lung hematocrit. The isolated lungs of rabbit were perfused with homologous blood until the perfusion pressure and preparation weight had stabilized. Then RBC tagged with  $Cr^{51}$  and albumin with  $I^{125}$  were added to the blood as the indicators of RBC and plasma. To assure a good mixing on the one hand, while minimizing the loss of these tracers on the other, the pulmonary artery and the left atrial appendage were closed eight minutes after the injection of indicators. The lungs were immersed in liquid nitrogen, crushed, and homogenized. By comparing the radioactivity in the lung tissue with that in samples of the blood perfusate, Aarseth et al (1978) estimated that the plasma volume in the lung was 9.3 ml and the RBC volume was 4.5 ml. These measurements yielded a whole-lung hematocrit of 32%. In contrast, the hematocrit of the blood perfusate was 35%. The whole-lung hematocrit,  $H_w$ , was 93% of the arterial one.

To estimate the capillary hematocrit from this volume dilution measurement, we consider the pulmonary vasculature as one formed by three compartments: an arterial, a microvascular, and a venous one. Due to large vessel size for the arterial and venous compartments, we assume their hematocrit to be  $H_a$ . For such a model and the definition of  $H_w$ , we have

$$H_c/H_a = 1 - (1 - H_w/H_a)V/V_c, \quad (25.1)$$

where  $V$  is the vascular volume and  $V_c$  is the microvascular volume. Horsfield and Gordon (1981) reported that the capillary volume of the human lung is 47% of the vascular volume. If we regard the microvascular compartment as one formed by the capillaries, we find from the whole-lung hematocrit measured by Aarseth et al (1978) and Equation 25.1 that the capillary hematocrit becomes 30%, which is 85% of the arterial hematocrit.

This volume dilution method measures the blood volume containing the tagged indicators and is subject to the condition that they are uniformly distributed in the blood volume. Aarseth et al (1978) found in two lungs that the hemoglobin content predicted an RBC volume 10% larger than that of tagged RBC, and Gibson et al (1946) found the former is 20% smaller than the latter, indicating a uniform distribution of tagged RBC. Tagged albumin leaks through the pulmonary endothelium into the tissue. The protocol used by Aarseth et al (1978) may have minimal leakage of albumin while providing a time long enough for mixing of indicators with circulating blood.

Use of the three-compartment model to estimate the capillary hematocrit from the whole-lung hematocrit, assuming the microvascular volume to be the capillary volume, is reasonable for the lung. According to the Fahraeus effect, arteries and veins smaller than 100 to 200  $\mu\text{m}$  are likely to have a tube hematocrit lower than the arterial, but higher than the capillary hematocrit. To assess their contributions, we assume that the difference between their hematocrit and the arterial one is one half of  $Ha - Hc$ . Horsfield and Gordon (1981) reported that these vessels contain 3.8% of the total vascular volume. For these values, we find that the hematocrit ratio,  $Hc/Ha$ , is underestimated by 1%. Another uncertainty may result from the estimation of the fractional volume of the capillaries. If it occupies a larger fraction, say 55%, the hematocrit ratio increases to 87%, 2% above the previous estimate.

A different procedure for volume dilution was used by Gibson et al (1946) to measure the hematocrit in minute vessels. Radioactive RBC and albumin were injected into the circulation of the dog. One hour later, the dog was sacrificed and the lung was isolated and cut up to obtain samples devoid of large vessels. By counting the radioactivities in the samples, they determined that the volume of plasma in the minute vessels of the lung was 20 ml and the volume of circulating RBC was 9.9 ml. The lung had a weight of 99 g, and the arterial hematocrit was 42%. The fractional volume of these vessels was smaller than that of capillaries as measured by Horsfield and Gordon (1981). Since no attempt was made in Gibson's protocol to prevent blood from oozing out of the cut samples, the large volume reduction of the capillaries could flush clear the original blood content in the minute arteries and veins of the samples and fill them with the capillary blood. Thus we can take the hematocrit from the study of Gibson et al (1946) as the capillary hematocrit to obtain an  $Hc/Ha$  of 78%.

#### 4. Dilution of Indicators in Pulmonary Blood Flows

One indicator dilution technique for measuring the RBC and plasma volumes is exemplified by the study of Aarseth et al (1978). The lungs were perfused through an arterial cannula inserted into the pulmonary artery and a venous one to the left atrium. Bolus injections of  $I^{125}$  tagged albumin and  $Cr^{51}$  tagged RBC into the arterial cannula were carried out. The outflows of the indicators were collected to eliminate their recirculation. Part of the outflow was sampled every 1.16 seconds for the determination of the dilution curves. Figure 25.2 shows the indicator dilution curves derived from the measurements of the radioactivity of outflow blood samples. The mean transit time from the site of injection to that of the sampling is calculated from the following formula:

$$\bar{t} = \sum_{i=1}^n (dpm_i \times t_i) / \sum_{i=1}^n dpm_i, \quad (25.2)$$

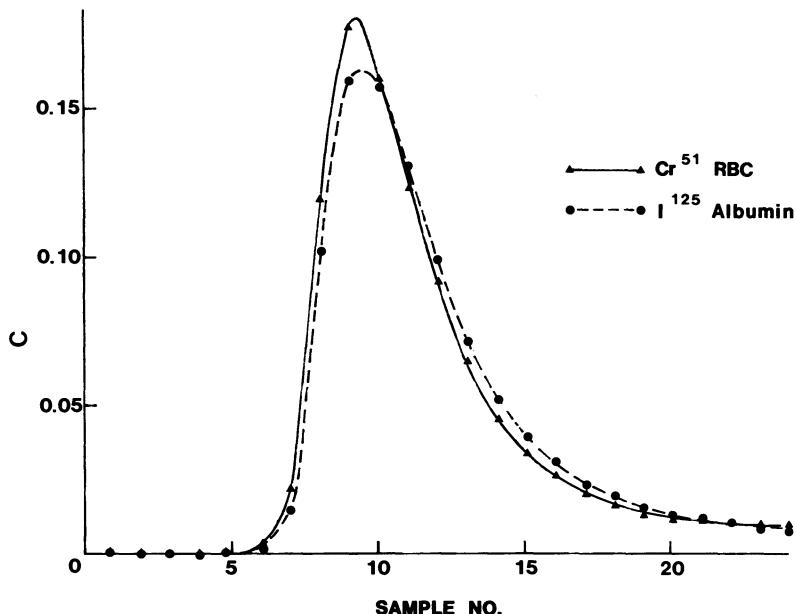


FIGURE 25.2. The indicator dilution curves obtained for an isolated perfused lung of rabbit.  $t = 0$  corresponded to the time the indicator bolus was injected to the pulmonary artery. Each sample was collected over a period of 1.16 seconds. The venous blood flow was collected and was not recirculated. These data from the study of Aarseth et al (1978) were normalized to give a unit area. Courtesy of *Acta Physiol. Scand.* and Drs. Aarseth, Waaler, and Nicolaysen.

where  $dpm_i$  is the disintegrations per minute counted from the  $i$ th sample and the summation starts from the time of injection and sums over  $n$  (about 55) samples. The downslope of the indicator curve was corrected by fitting it with a mono- or bi-exponential decay curve. The mean transit time over the injection cannula and the sampling cannula was independently determined and removed from the calculation of the mean transit time. Multiplying the mean transit time of the RBC with the total RBC flow, Aarseth et al (1978) determined the RBC volume from the pulmonary artery to the left atrium to be 5.6 ml. The product of the plasma mean transit time and the total plasma flow yields a plasma volume of 14.8 ml.

The total blood volume (20.4 ml) is larger than that based on the volume dilution technique (13.8 ml) because the former includes the volume of the pulmonary artery and left atrium, and the flow in the lung used in the indicator dilution study was higher (personal communication with G. Nicolaysen). Subtracting the volume of the pulmonary artery and left atrium, estimated to be 10%, we obtain a whole-lung hematocrit of 27%, which is 91% of  $Ha$ . Using the three-compartment model with a microvascular fraction of 47% and Equation 25.1, we find that the capillary hematocrit is 81% of  $Ha$ . (From the

data of another experimental sequence of Aarseth et al (1978), the capillary hematocrit was estimated to be 86% of  $Ha$ .

Because of the weighting factor  $t, i$  in the numerator of Equation 25.2, the calculation of the mean transit time depends more on the portion of data from large times. For *in-vivo* experiments, the way to correct the recirculation of the indicators at large times is very important. Rapaport et al (1956) used different procedures to estimate the blood volume and found their difference could be as large as 40%. To simulate the decay portion of the dilution curve with a monoexponential function for the case of no recirculation, Aarseth et al (1978) showed that the median value of the mean transit time was 10% smaller than that calculated from Equation 25.2 with the measured data. The loss of the indicator may also affect the accuracy in calculating the mean transit time. A comparison on the mass of indicator injected with the one determined from all outflow may help to evaluate the loss of indicator. For the *in-vivo* experiments performed by Dow et al (1946), 10% of tagged RBC and 20% of tagged albumin were lost, some of them through the bronchial circulation. The errors associated with the three-compartment model remain the same as discussed previously. This indicator dilution method excludes the portion of the vascular volume having no flow.

## 5. Dilution of Density Indicator from Pulmonary Capillaries

Blood is a mixture of red blood cells and plasma. The density of blood  $\rho_b$  having a hematocrit  $H$  is related to the RBC density  $\rho_r$  and plasma density  $\rho_p$  by:

$$\rho_b = \rho_p + (\rho_r - \rho_p)H. \quad (25.3)$$

When we inject a volume ( $Vi$ ) of RBC concentrate ( $H = 100\%$ ) into the pulmonary artery, the venous outflow exhibits a transient density increase. From this measurement and Equation 25.3, we can determine the hematocrit change in venous blood,  $Hv$ . In reference to the arterial hematocrit  $Ha$ , which is maintained constant, we regard the hematocrit difference  $Hv - Ha$  as the concentration of an “indicator” in the theory of indicator dilution. It can be shown from the consideration of mass conservation that the mass of indicator injected is  $Vi(1 - Ha)$ . For the injection of a plasma bolus of zero hematocrit, we expect to observe a transient reduction in hematocrit. For such a case, the mass of indicator injected ( $-ViHa$ ) and the indicator concentration in venous blood  $Hv - Ha$  are negative.

Let us apply this concept to the three-compartment model of the lung mentioned previously. When the microvascular volume is to decrease from a volume  $Vc$  to  $Vc'$  without changing the capillary hematocrit, a volume of blood ( $Vc - Vc'$ ) having the capillary hematocrit is released to the blood flow. This is equivalent to the injection of this amount of indicator,  $(Vc - Vc')(Hc - Ha)$ ,

to the capillary out flow. The mass of indicator leaving the pulmonary vein is given by

$$Vr = \int_0^T Qv(Hv - Ha) dt. \quad (25.4)$$

The mass conservation of indicator leads to this equation:

$$Vr = (Ha - Hc)(Vc' - Vc). \quad (25.5)$$

Letting the change of the entire pulmonary vascular volume be  $(V' - V)$ , we can normalize the equation above to be

$$Vr/(Ha(V' - V)) = (1 - Hc/Ha)(Vc' - Vc)/(V' - V). \quad (25.6)$$

Lee et al (1985) have used isolated perfused lobes of dog lungs to quantify the release fraction given on the right hand side of Equation 25.6. When we lowered the arterial and venous pressure perfusing the lobe, we observed a transient decrease in the density of venous blood. See Figure 25.3 (Lee et al. 1985). Converting this density to  $Hv$ , we calculated  $Vr$  according to Equation 25.4. The measured weight reduction of the lobe was used to identify the change in vascular volume,  $V' - V$ . We found that the value of  $Vr(Ha(V' - V))$  was  $0.10 \pm 0.03$ . It was  $0.11 \pm 0.02$  for the case of pressure elevation. A Student's t-test showed no statistically significant difference between the released fraction for the two pressure adjustments.

Suppose the change in the microvascular volume was 55% of the change in whole vascular volume during the pressure elevation. From Equation 25.6, we found that the capillary hematocrit was 80% of  $Ha$ .

We obtained the volume  $Vr$  by integrating the dilution curve for the area and then calculated the change in capillary blood volume. To determine the volume from conventional indicator dilution methods, the concentration is weighted by the time (see Equation 25.2). Thus the density dilution method

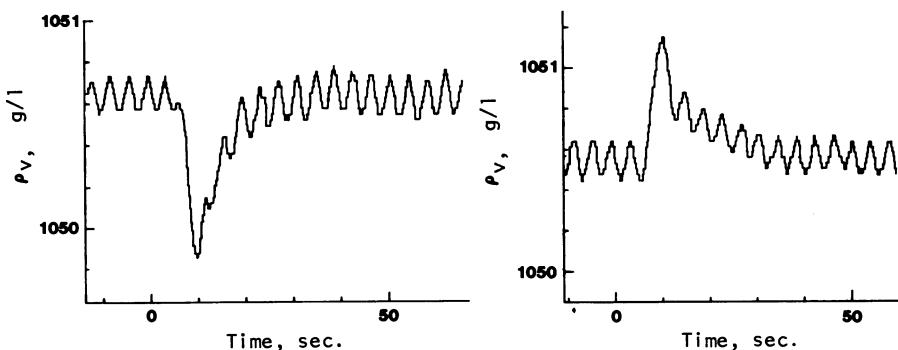


FIGURE 25.3. The transient density change in venous blood due to a lowering (left) and an elevation (right) of vascular pressures perfusing an isolated left lower lobe of dog, adapted from Lee et al (1985).

need not be so stringent in the determination of the small concentration at large times. In terms of calculating the capillary hematocrit, we still need to use a three-compartment model and to assume a microvascular volumetric fraction. Because of the change in density ( $\rho_v - \rho_a$ ) or hematocrit ( $Hv - Ha$ ) is small, the validity of the measured data depends strongly on our capability to maintain a constant arterial hematocrit.

## 6. The Fahraeus Effect in Small Circular Tubes

To assess the hematocrit in the pulmonary capillaries in a different perspective, we shall review the Fahraeus effect in blood flow through small circular tubes. Albrecht et al (1978) used glass tubes as a model for capillaries. They counted the number of RBC over a section of the tube with a volume  $V$  and a length  $L$ . Then the RBC volume ( $V_r$ ) was taken as the product of this number and the average volume of individual RBC. The ratio,  $V_r/V$ , was the capillary or tube hematocrit,  $Hc$ . Two photosensors were used to detect the delay or mean transit time for red blood cells ( $Tr$ ) and a dye bolus ( $Tp$ ) to travel across the distance separating the sensors,  $L$ . The mean velocity of RBC ( $Ur$ ) was  $L/Tr$  and that of plasma ( $Up$ ) was  $L/Tp$ .

Let  $Qr$ ,  $Qp$ , and  $Q$  be the RBC, plasma, and blood flows through the tube, respectively, and  $Ub$  be the mean velocity of the blood. The hematocrit of blood discharged from the tube is identified here as the arterial hematocrit  $Ha$ . Relating  $V$ ,  $V_r$ , and plasma volume ( $Vp$ ) in that section of tube to the flows, mean transit times, and the tube and discharge hematocrits leads to these identities:

$$V_r = HcV = HcQL/Ub = QrTr = QHaL/Ur \quad (25.7)$$

$$Vp = (1 - Hc)V = QpTp = Q(1 - Ha)L/Up. \quad (25.8)$$

From these relations, we have

$$Ha/(1 - Ha) = (Ur/Up)Hc/(1 - Hc) \quad (25.9)$$

and

$$Ur/Ub = Ha/Hc. \quad (25.10)$$

With the measurements of  $Ur$ ,  $Up$ , and  $Hc$ , one calculates the discharge hematocrit  $Ha$  from Equation 25.9 for the blood flowing in the tubes. Inter- or extrapolating the data on  $Hc$  of Albrecht et al (1978) for a discharge hematocrit of 45%, the dependence of  $Hc/Ha$  on the tube diameter,  $D$ , is depicted in Figure 25.4.

One view of the microvasculature of the lung is to consider it as a network of elliptical capillaries having a minor axis of  $5\text{ }\mu\text{m}$  and a major axis of  $10\text{ }\mu\text{m}$ . To make use of the previous results on the Fahraeus effect, one may equate the pulmonary capillary with a circular tube having a diameter of  $7\text{ }\mu\text{m}$ . This tube

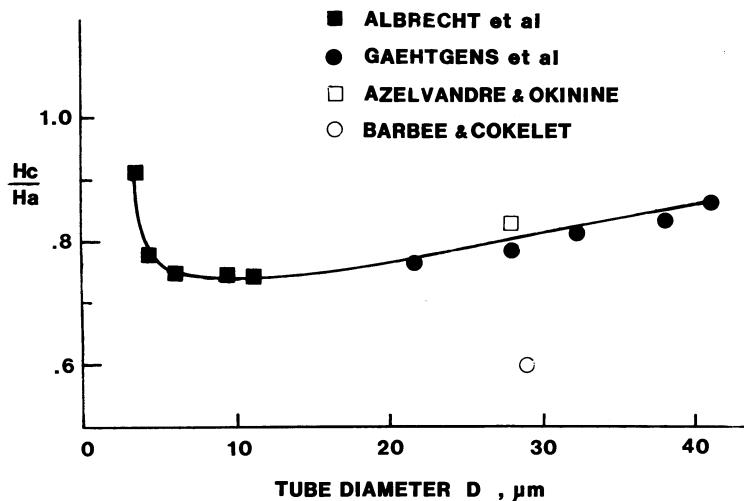


FIGURE 25.4. The dependence of the ratio of capillary and discharge hematocrit,  $H_c/H_a$ , on the capillary diameter  $D$ . Data from Albrecht et al (1978) are interpolated or extrapolated to find one correspondent to a discharge hematocrit of 45%. Other data were quoted in their study.

has the same cross-sectional area as the elliptical capillary. Having a major axis larger than the diameter of biconcave RBC, the capillary may allow RBC to move more freely than a circular capillary of the same area. An equivalent circular diameter larger than  $7 \mu\text{m}$  should also be considered. If the highly distensible pulmonary capillaries are to have an equivalent diameter within 7 to  $15 \mu\text{m}$ , we can regard, as suggested from Figure 25.4, that its  $H_c$  is about 74% of  $H_a$ . Only when the alveolar capillaries are compressed or expanded beyond this diameter range is the hematocrit ratio larger and approaching unity.

## 7. Physiological Significance of a Low Capillary Hematocrit

For the diversified methods of indicator dilution and the study on the Fahraeus effect, the pulmonary capillary hematocrit is estimated to range from 74% to 85% of the arterial hematocrit. The morphometric estimates of  $H_c/H_a$  yield a wider range of 60% to 120%. Roughton and Forster (1957) used a carbon monoxide method to determine the capillary blood volume in the lung. One direct consequence of a low capillary hematocrit, say 80% of  $H_a$ , is to increase their volume determination by 25%.

A low microvascular hematocrit affects the diffusion capacity of oxygen in

the lung in two aspects. First it shortens the transit time of RBC through pulmonary capillaries. For an  $Hc$  80% of the arterial  $Ha$ , the apparent viscosity of blood is lowered by 16% as estimated from Barbee and Cokelet's study (1971) on the Fahraeus-Linqvist effect. To assure a discharge hematocrit the same as the arterial one, the RBC move faster than the blood by 25% (Equation 25.10). Combining these two effects, we estimate that the transit time of RBC through the capillary is 67% of the transit time if the capillary hematocrit were to assume the arterial value. Secondly, a lower hematocrit corresponds to a thicker plasma layer for the oxygen to diffuse to RBC. In the sheet flow model of Fung and Sabin (1969), the biconcave RBC may prefer to flow with its face parallel to the sheet and stay in the central portion of the sheet. Consequently, it may be reasonable to simplify the two-phase flow in pulmonary capillaries as a sheet flow formed by one central RBC layer and two plasma layers that are enclosed by two alveolar membranes. For  $Hc = 36\%$  and  $Ha = 45\%$ , the thickness of the plasma layer is increased by 16% over the case with the arterial hematocrit. (The simplicity of the sheet flow model allows us to construct this layer model for analyzing the effect of reaction kinetics between oxygen and hemoglobin on the oxygen uptake in pulmonary blood flow.) Under hypoxia and at heavy exercise, the two factors considered above reduce the oxygen uptake by RBC to a degree that they leave the capillaries with an oxygen tension lower than that of alveolar air.

In summary, the studies reviewed here suggest that the microvascular hematocrit in the lung is about 80% of the arterial hematocrit. A lower hematocrit is important to the blood flow and oxygen uptake in pulmonary capillaries.

## Acknowledgments

This research was supported in part by National Heart, Lung, and Blood Institute HL 23769.

## References

- Aarseth, P., Waaler, B.A., and Nicolaysen, G. 1978. *Acta Physiol. Scand.* 103:165–172.
- Albrecht, K.H., Gaehtgens, P., Pries, A., and Heuser, M. 1979. *Microvasc. Res.* 18:33–47.
- Azelvandre, F., and Oiknine, C. 1976. *Biorheology* 13:315–355.
- Barbee, J.M., and Cokelet, G.R. 1971. *Microvasc. Res.* 3:17–21.
- Dow, P., Hahn, P.F., and Hamilton, W.F. 1946. *Am. J. Physiol.* 147:493–499.
- Fung, Y.C., and Sabin, S.S. 1969. *J. Appl. Physiol.* 26:472–488.
- Gibson, J.G., II, Seligman, A.M., Peacock, W.C., Aub, J.C., Fine, J., and Evans, R.D., 1946. *J. Clin. Invest.* 25:848–857.
- Horsfield, K., and Gordon, W.I. 1981. *Lung* 159:211–218.
- Lee, J.S., Lee, L.P., Villafane, M., and Evans, M.V. 1985. *Microvasc. Res.* 30:222–234.

- Rapaport, E., Kuida, H., Haynes, F.W., and Dexter, L. 1956. *Am. J. Physiol.* 185: 127-132.
- Roughton, F.J.W., and Forster, R.E. 1957. *J. Appl. Physiol.* 11:290-302.
- Siegwart, B., Gehr, P., Gil, J., and Weibel, E.R. 1971. *Resp. Physiol.* 12:141-159.
- Sobin, S.S., Tremer, H.M., and Fung, Y.C. 1970. *Circ. Res.* 26:397-414.
- Vreim, C.E., and Staub, N.C. 1973. *J. Appl. Physiol.* 34:452-459.
- Weibel, E.R. 1971. *Resp. Physiol.* 11:54-75.

# Pulmonary Blood Flow in the Cat: Correlation Between Theory and Experiment

R.T. YEN and S.S. SOBIN

## 1. Introduction

Knowledge of the various factors controlling pulmonary blood flow is important in human health and disease, and today our knowledge and understanding of those mechanisms is incomplete. Precise experimental data required to fill in the gaps is best obtained when theoretical understanding of the biomechanics precedes and accompanies the experimental approach. For this physicomathematical approach to a study of the pulmonary circulation, we require (1) detailed description of vascular geometry, (2) measured elasticity of blood vessels, and (3) rheology of blood in blood vessels. With this information, together with the basic physical laws and appropriate boundary conditions, differential equations can be written and the problems solved either analytically or numerically. Although theoretical analyses can be equally applied to the human and the animal lung, as will be noted below all the necessary data cannot be appropriately obtained from the human lung, and we will use the feline for specifically stated reasons.

Although alveoli are considered the gas exchange unit of the lung, alveoli exist because of early postnatal partitioning of air spaces by the capillary network and its supporting connective tissue of collagen and elastin. These pulmonary microvessels were first seen in the frog lung by Malpighi (1661), whose observations were extended by Stephen Hales in 1731. The extensive literature was reviewed by Fenn and Rahn (1964). In recent times, the interalveolar capillary network in the human lung has been categorized as an interconnected system of wedged cylindrical tubes by Weibel (1963), and described as a microvascular sheet in the cat by Fung and Sabin (1969), as well as in other mammals, including man (personal communication, Sabin and Fung). The many extensive studies of the interalveolar wall have been critically reviewed by Weibel and Gil (1977) from the point of view of gas exchange.

The larger human pulmonary arteries were studied morphometrically by Singhal et al (1973) and Horsfield (1978). Using resin casts of the human arterial tree, they measured length, diameter, and branching order of vessels using the Strahler (1957) method of counting. The volumes of all the measured

branches were added and gave a cumulated volume of 151 ml for the human pulmonary arterial system. Corresponding measurements of the human pulmonary venous tree were made by Horsfield and Gordon (1981); similar data for the venous trees of other mammals were not found.

In addition to morphologic data, distensibility data of all generations of pulmonary vessels is needed because pressure-flow behavior of blood vessels is also a function of the distensibility of blood vessels, and pulmonary vessels are quite distensible. The data here are incomplete and scattered for all species except the cat; the papers of Sabin et al (1972, 1978) and recent papers of Yen et al (1980, 1981; see also for a literature review) now supply a set of distensibility data for all generations of the pulmonary blood vessels of the cat.

The present status of the data base for pulmonary blood flow analysis may be summarized briefly as follows: For humans, known data include vessel branching pattern and dimensions of the arterial and venous trees, capillary wall structure, and incomplete distensibility data of large vessels; data do not exist for the distensibility of small arteries and veins, arterioles, and venules. For the dog, there are large amounts of physiological data, but again there are no available data for details of vessel branching, dimensions, and distensibility of blood vessels. For the cat, known data include capillary network geometry, distensibility of arteries, arterioles, capillaries, venules, and veins, but data for the vascular morphometry of the cat have not been reported. Thus we do not have a complete set of data on any species, animal or man.

This is a severe handicap for a comparison between physiological theory and experiment. It is logical to obtain one complete set of data about one species. The cat is the animal of choice because only the vessel morphometric data are missing. Recently we have obtained a complete set of data on the vascular geometry of the cat lung. The morphometric data of the pulmonary arterial and venous tree of the cat were studied using silicone elastomer casts, and these results are published separately (Yen et al, 1983, 1984). Based on these data, we carry out theoretical analysis of the pulmonary circulation of the cat. In the calculations, the sheet flow theory of Fung and Sabin (1969) is used for the pulmonary capillary blood flow and an analogous "fifth power" law is used for flow in the arteries and veins. In the theoretical calculations of the flow in zone 2 conditions, an assumption is made in the earlier paper of Fung and Sabin (1972) that the site of flow limitation must be located at junctions of capillaries and venules. Recently Fung et al (1983) verified experimentally that the pulmonary veins do not collapse when the local blood pressure falls below airway pressure. Thus the theoretical calculations that take into account the detailed anatomical and elasticity data can be checked against experimental results. The comparison is presented below.

## 2. Theoretical Calculations

Detailed theoretical calculations are presented in Zhuang et al (1983). A brief description follows.

## 2.1. Pulmonary Capillary Blood Flow

The “sheet-flow” model proposed by Fung and Sabin (1969) is used to obtain the pressure–flow relationship in the alveolar capillary bed. Equations (26.1) and (26.2) below are from Fung et al (1972). From the sheet-flow theory and measurement of sheet elasticity, when the local microvascular pressure exceeds the alveolar gas pressure, the sheet thickness ( $h$ ) varies linearly with the “transmural pressure”  $p - p_A$ ; thus,

$$h = h_0 + \alpha(p - p_A), \quad (26.1)$$

where  $h$  denotes the sheet thickness,  $p$  is the blood pressure,  $p_A$  is the airway pressure,  $h_0$  is the capillary sheet thickness when transmural pressure is zero, and  $\alpha$  is the compliance or elastic constant. Using Equation (26.1), together with the equation of motion of the fluid, the equation of continuity, and the boundary conditions, the following equation relating the flow ( $\dot{Q}$ ) with the pressures is obtained:

$$\begin{aligned} \dot{Q} &= \frac{SA}{4\mu kf\bar{L}^2\alpha} \{h_{art}^4 - h_{ven}^4\} \\ &= \frac{SA}{4\mu kf\bar{L}^2\alpha} \{[h_0 + \alpha(p_{art} - p_A)]^4 - [h_0 + \alpha(p_{ven} - p_A)]^4\}. \end{aligned} \quad (26.2)$$

This is known as the fourth-power law. In the formula above,  $\dot{Q}$  is the volume flow rate,  $h_{art}$  is the sheet thickness at the arteriole where blood enters into the sheet,  $h_{ven}$  is the sheet thickness at the venule where blood exits from the sheet,  $A$  is the capillary area,  $S$  is the “vascular space tissue ratio,” which is the fraction of the alveolar sheet occupied by the blood, and  $\mu$  is the apparent viscosity of blood in the capillary sheet.  $K$  is a dimensionless factor that depends on the ratio of the thickness to width of the capillary, and has a numerical value of about 12;  $f$  is a geometric friction factor that is a function of the ratio of the sheet thickness to post diameter and several other flow parameters.  $\bar{L}$  is the average length of blood pathway between the inlet and outlet. Sabin et al (1980) found the average path length  $\bar{L} = 556 \pm 285$  Standard Deviation (S.D.)  $\mu\text{m}$ . Zhuang et al (1985) found that the total alveolar surface area of the lung of the cat is  $A = 0.87\text{--}1.27 \text{ m}^2$ . The apparent viscosity  $\mu$  of blood depends on the hematocrit. Yen and Fung (1973) showed that the apparent viscosity of blood in the pulmonary capillary sheet is 1.92 cp when the hematocrit is 30%, although the precise hematocrit of the pulmonary microvascular sheet is not known for any flow regimen.

## 2.2. Blood Flow in Arteries and Veins

Pulmonary arteries and veins were found to behave like elastic tubes whose diameter  $D$  changes with transmural pressure  $\Delta p$  according to a linear law (Yen et al, 1980; Yen and Foppiano, 1981):

$$D = D_0[1 + \alpha\Delta p]. \quad (26.3)$$

Here  $D_0$  is the diameter when the transmural pressure is zero,  $\alpha$  is the compliance constant, and  $\Delta p$  is the transmural pressure. If the flow is assumed to be Poiseuillean, and the fluid is Newtonian with viscosity  $\mu$ , then the pressure gradient is related to the flow rate by the equation

$$\frac{dp}{dx} = \frac{128\mu}{\pi D^4} \dot{Q}. \quad (26.4)$$

On substituting Equation (26.3) into Equation (26.4) and integrating, we obtain the following formula for the flow rate in a tube of length  $L$ :

$$\begin{aligned} \dot{Q} &= \frac{\pi}{640\mu\alpha D_0 L} \{ [D(0)]^5 - [D(L)]^5 \} \\ &= \frac{\pi}{640\mu\alpha D_0 L} \{ [D_0(1 + \alpha\Delta p_{entry})]^5 - [D_0(1 + \alpha\Delta p_{exit})]^5 \}. \end{aligned} \quad (26.5)$$

This formula is analogous to Equation (26.2) and is known as the fifth-power law. If losses at the entry and exit sections at points of bifurcation or convergence of vessels are ignored, then a repeated application of Equations (26.5) and (26.2) will synthesize the flow in different segments into that of the whole lung.

### 2.3. Data Base and Predictions

Tables 26.1 and 26.2 list the morphometric and elastic data of the cat's right lung. These morphometric data (Yen et al, 1983, 1984) were obtained from cats of average weight 3.5 kg, with airway pressure of 10 cm H<sub>2</sub>O and pleural

TABLE 26.1. Morphometric and elastic data of the pulmonary arteries of the cat (right lung).

Order	Number of branches $N_n$	Diameter $D_{on}$ (cm)	Length $L_n$ (cm)	Apparent viscosity coeffic. $\mu_n$ (cp)	Compliance $10^{-5}$ (dyne/cm <sup>2</sup> ) <sup>-1</sup>
1	300358	0.0024	0.0116	2.5	1.928
2	97519	0.0044	0.0262	3.0	1.928
3	31662	0.0073	0.0433	3.5	1.928
4	9736	0.0122	0.0810	4.0	1.928
5	2925	0.0192	0.1510	4.0	1.122
6	774	0.0352	0.2720	4.0	0.796
7	202	0.0533	0.4600	4.0	0.714
8	49	0.0875	0.8190	4.0	1.122
9	12	0.1519	1.4260	4.0	2.663
10	4	0.2486	1.1870	4.0	2.663
11	1	0.5080	2.5000	4.0	2.663

Note: The morphometric data are measured at transpulmonary pressure  $p_A - p_{PL} = 10$  cm H<sub>2</sub>O.

TABLE 26.2. Morphometric and elastic data of the pulmonary veins of the cat (right lung).

Order	Number of branches $N_n$	Diameter $D_m$ (cm)	Length $L_n$ (cm)	Apparent viscosity coeffic. $\mu_n$ (cp)	Compliance		Compliance $P_{pl} = -17 \text{ cm H}_2\text{O}$	Compliance $P_{pl} = -23 \text{ cm H}_2\text{O}$
					$10^{-5} \text{ (dyne/cm}^2\text{)}^{-1}$	$10^{-5} \text{ (dyne/cm}^2\text{)}^{-1}$		
1	282/33	0.0025	0.0086	2.5	1.928	1.928	1.928	1.928
2	86/241	0.0046	0.0247	3.0	1.928	1.928	1.928	1.928
3	26/306	0.0077	0.0496	3.5	1.928	1.928	1.928	1.928
4	80/24	0.0159	0.1545	4.0	2.148	2.680	2.300	
5	23/48	0.0293	0.2380	4.0	1.567	1.878	1.388	
6	65/6	0.0484	0.3810	4.0	1.037	0.988	0.951	
7	17/1	0.0719	0.4950	4.0	1.037	0.988	0.951	
8	46	0.1119	0.7610	4.0	0.749	0.635	0.855	
9	13	0.1859	1.5120	4.0	0.749	0.635	0.855	
10	4	0.3240	1.9240	4.0	0.749	0.635	0.855	
11	1	0.4840	2.5000	4.0	0.749	0.635	0.855	

Note: The morphometric data are measured at transpulmonary pressure  $p_A - p_{PL} = 10 \text{ cm H}_2\text{O}$ .

pressure of zero (atmospheric). From these data,  $D_0$  is computed according to Equation (26.3). The compliance data  $\alpha$  for arteries have been determined at  $p_A = 0$  and  $p_{PL} = -10$  and are listed in Table 26.1. For veins,  $\alpha$  values have been determined at  $p_A = 0$  and three values of  $p_{PL}$ ,  $-7$ ,  $-17$ , and  $-23$  cm H<sub>2</sub>O, and are listed in Table 26.2. The apparent viscosity values listed in Tables 26.1 and 26.2 are estimated under the hypothesis that the hematocrit varies from about 45% in larger vessels with diameters greater than 100  $\mu\text{m}$  (or order numbers 4–11) to about 30% in the capillaries. At a hematocrit of 45% in large vessels, an apparent viscosity of 4.0 cp is assumed. The apparent viscosity of blood in small vessels with order numbers 1–3 is obtained by linear interpolation.

To analyze blood flow, the morphological data of pulmonary arterial and venous vessels of the cat (Yen et al, 1983, 1984) are used. The Strahler (1957) system of counting is used to describe the branching pattern. The statistical data show that the cat lung has 11 orders of branching, and that the branching ratio (ratio of the total number of vessels of order  $n$  to that of order  $n + 1$ ) is 3.58 for the arterial tree and 3.52 for the venous tree. For one of the calculations, we consider flow in a branch of order  $n + 1$  to be divided equally into  $B_n$  daughter vessels of order  $n$ ,  $B_n$  being the branching ratio of order  $n$ . In another example, an asymmetric bifurcation pattern is assumed. The result depends, of course, on the degree of asymmetry.

### 3. Experiments

Two different experiments on cats were performed. The data obtained were used to check the theoretical predictions.

#### 3.1. Experiments on Resistance to Blood Flow

Six cats ranging in weight from 1.8 to 3.0 kg were anesthetized intraperitoneally with sodium pentobarbital, followed by exposure of a vein and administration of heparin for anticoagulation. The trachea was exposed and cannulated and the lung mechanically ventilated. The chest was opened in midline, exposing the heart and lungs. The aorta and superior and inferior cavae were ligated, and by appropriate cannulations of the pulmonary artery and left atrium, a simple isolated lung preparation was obtained, as shown in Figure 26.1. The pulmonary artery was attached to a constant-flow Harvard perfusion pump. The left atrial pressure was controlled by a Starling resistor, whereas the airway pressure was controlled via an air pressure source and monitored on a water manometer. In each experiment, the lung was perfused with Macrodex<sup>®</sup> (6% dextran-70 in normal saline, with papaverine 1 mg per ml for smooth muscle relaxation) at a constant flow rate while airway pressure was held constant at +7, +17, or +23 cm H<sub>2</sub>O pressure and the pulmonary venous pressure was fixed at 3 cm H<sub>2</sub>O. Thus, since  $p_a > p_A > p_v$ , this es-

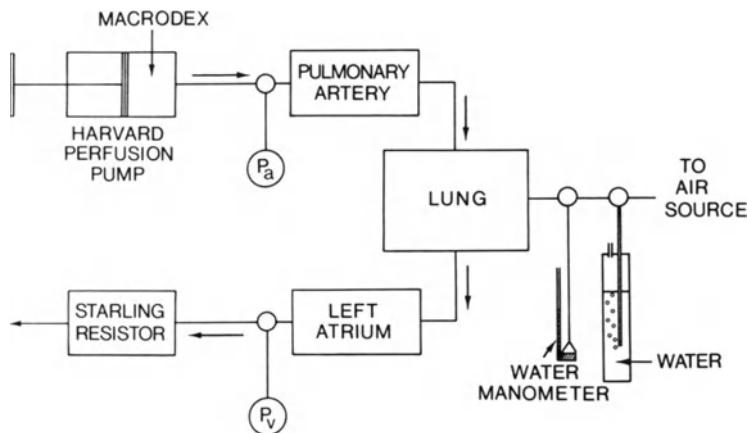


FIGURE 26.1. Schematic diagram of the experiments on resistance to blood flow.

tablishes a zone 2 condition of the lung. Pulmonary arterial pressure and left atrial pressure were continuously measured.

### 3.2. Experiments on Flow Limitation

The preparation, including solutions, was the same as that described in the preceding section. Figure 26.2 shows a schematic diagram of the experiment. The lung was perfused at a constant pulmonary arterial pressure of either 20

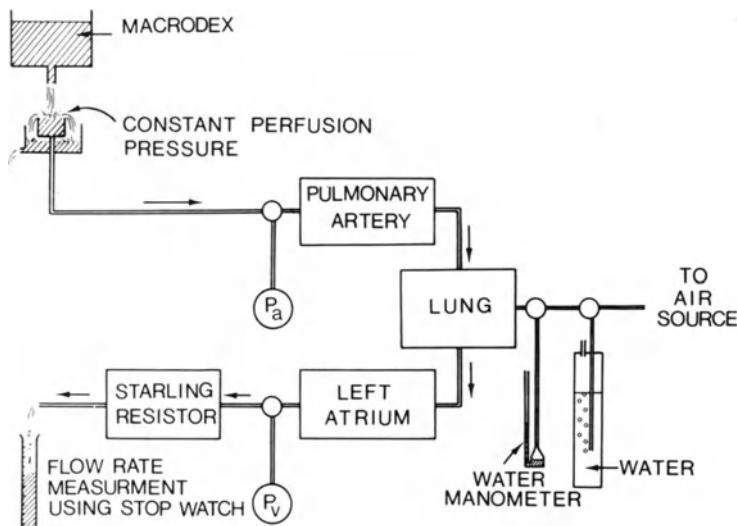


FIGURE 26.2. Schematic diagram of the experiments on flow limitation.

or 25 cm H<sub>2</sub>O from an open reservoir, with the airway pressure controlled at 10 or 15 cm H<sub>2</sub>O and the left atrial pressure varied from -6 to +20 cm H<sub>2</sub>O. The flow rate was obtained in each experiment by measuring a known amount of fluid exiting from the left atrium in a given time period after the flow reached a steady state.

#### 4. Results

Figure 26.3 shows the results of experiments on resistance to blood flow. The relationship between flow rate per kilogram body weight and pulmonary arterial pressure for fixed left atrial pressure and airway pressure under a zone 2 condition is shown. Solid curves represent the theoretical calculations, while the crosses and triangles represent the experimental data. The pressure-flow relationship is nonlinear. The flow rate increases more rapidly than the pulmonary arterial pressure. The figure also shows that, for positive inflation, flow decreases with increasing transpulmonary (airway) pressure, for the same pulmonary arterial pressure. The experimental results compare well with theoretical calculations.

Figures 26.4 and 26.5 show the relationship between flow and variable pulmonary venous pressure with fixed arterial pressures and alveolar pressures. In the case shown in Figure 26.4, the arterial pressure was kept at 20 cm

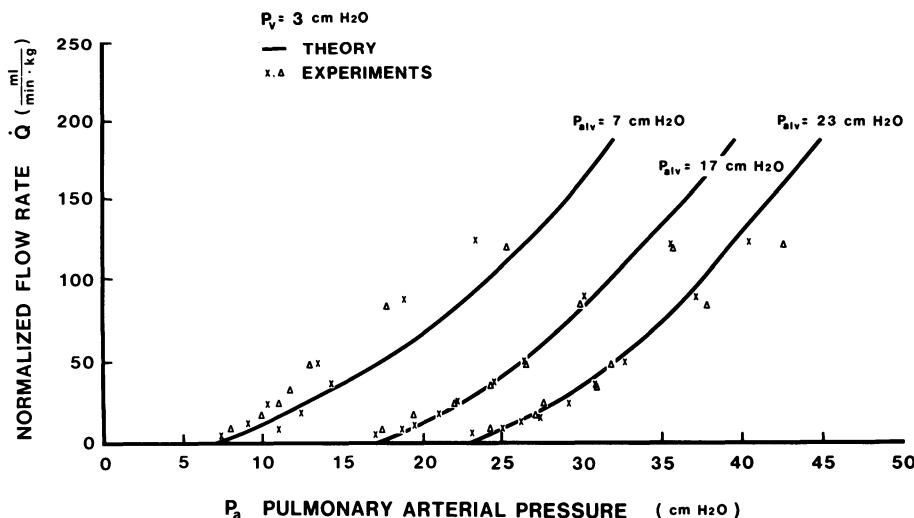


FIGURE 26.3. Comparison between flow rate per kilogram body weight and pulmonary arterial pressure for fixed left atrial pressure and airway pressure under zone 2 condition. In the experiments, three airway pressures were used (+ 7, + 17, + 23 cm H<sub>2</sub>O) while the left atrial pressure was held at 3 cm H<sub>2</sub>O. Solid curves represent the theoretical calculations; crosses represent the data obtained from experiments.

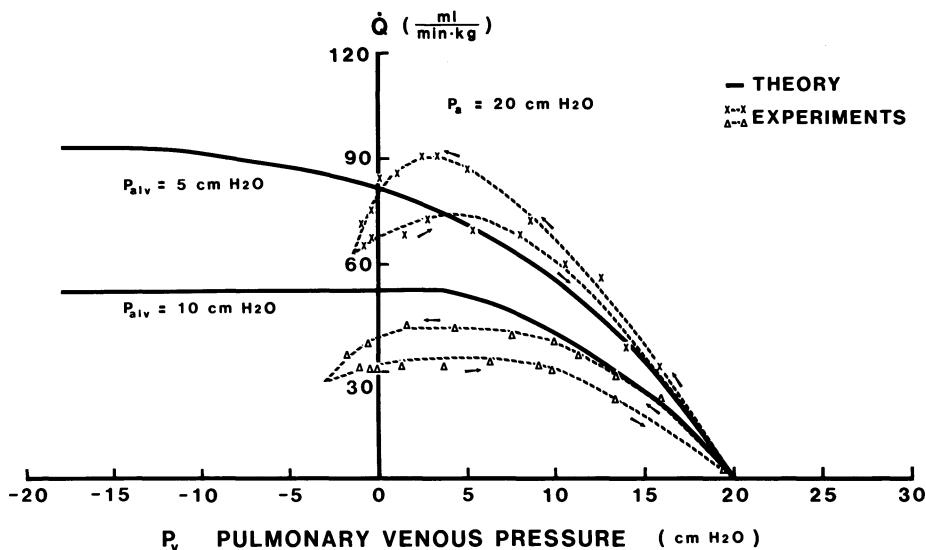


FIGURE 26.4. Comparison between flow and variable pulmonary venous pressure with fixed arterial pressures and alveolar pressures. In this figure, the arterial pressure was kept at 20 cm H<sub>2</sub>O, while the alveolar pressure was kept at 5 or 10 cm H<sub>2</sub>O. Solid curves refer to theoretical results. Broken curves refer to experimental results.

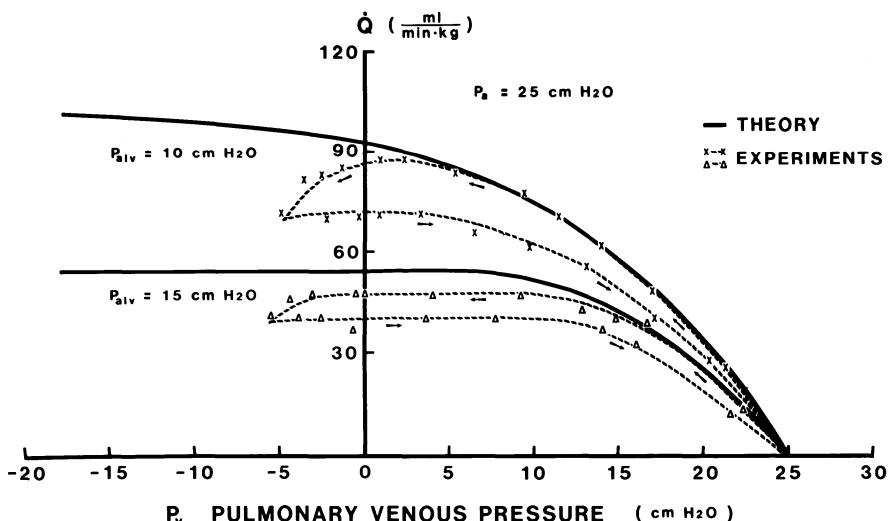


FIGURE 26.5. Comparison between flow and variable pulmonary venous pressure with fixed arterial pressures and alveolar pressures. In this figure, the arterial pressure was kept at 25 cm H<sub>2</sub>O, and the alveolar pressure was held constant at 10 and 15 cm H<sub>2</sub>O. Solid curves refer to theoretical results. Broken curves refer to experimental results.

$\text{H}_2\text{O}$  while the alveolar pressure was kept at 5 or 10 cm  $\text{H}_2\text{O}$ . For the case in Figure 26.5, the arterial pressure was kept at 25 cm  $\text{H}_2\text{O}$  and the alveolar pressure was held constant at 10 or 15 cm  $\text{H}_2\text{O}$ . In the theoretical calculation, the sluicing gates are assumed to lie at the exit ports of the capillary sheets, as was proposed by Fung (1972) under the hypothesis that the pulmonary venules and veins will not collapse because of the tethering effect from the interalveolar septa that attach to the vessel walls. The solid curves are theoretical results; broken curves are the experimental results. As can be seen, the flow reaches a maximum as the left atrial pressure is decreased continuously, while the pulmonary arterial and airway pressures are kept constant. When the flow rate reaches a maximum, further decreasing of the left atrial pressure will not result in an increase of flow rate. This is the so-called waterfall phenomenon. It is demonstrated clearly in both the theoretical curves and experiments. The experiments, however, show the existence of a hysteresis loop when the left atrial pressure first increases and then decreases. The hysteresis phenomenon was not predicted in the theory presented by Fung et al (1983).

In the course of the experiments, we found that if the lung was perfused with Macrodex without the addition of papaverine, the flow rate was about two-fifths of the value with papaverine added. From the curves of Figure 26.1, we can estimate the cardiac output under normal physiological conditions. Take the case in which  $p_{alv} = 7$ ,  $p_v = 3$ , and  $p_a = 20$  cm  $\text{H}_2\text{O}$ . Figure 26.1 yields a flow of  $88 \text{ ml/min} \times \text{kg}$  for the right lung of the cat. Thus the cardiac output is  $176 \text{ ml/min} \times \text{kg}$ , a value agreeing quite well with that given by Weiner et al (1967).

## 5. Discussion

These experiments are an excellent example of the value of a biomedical physicomathematical approach to a physiological problem. The analysis of control parameters of pulmonary blood flow requires minimal blood vessel geometry, blood vessel elasticity, and microvascular rheology. For the cat, only the latter had not been determined experimentally, and it was necessary to make certain assumptions about the microvascular sheet hematocrit and apparent blood viscosity. Under appropriate experimental conditions, these could be obtained. Such data are critically needed.

The overall experimental results compare well to theory, especially when normalized blood flow rate is considered as a function of pulmonary artery pressure. With the assumption that pulmonary blood flow is equivalent to the cardiac output, these values are close to the physiological values in the literature for the cat. (Cardiac output is considered to be pulmonary blood flow less coronary artery flow.)

The observation that in the absence of papaverine, a powerful vasodilator, the blood flow rate measured by Macrodex perfusion is much reduced over

that with papaverine, is of interest. Since the cardiac output with papaverine is equivalent to the physiological normal value, it is highly suggestive that in the normal physiological state, the pulmonary vasculature is not tonically constricted. This is in accord with the known physiological data in the terrestrial mammal.

There are differences between theory and experiment, as illustrated in Figures 24.4 and 24.5: (1) the experimental maximal flow rate is slightly less than predicted by theory; (2) hysteresis is present in the pressure-flow relationship when left atrial pressure first increases, then decreases; (3) there is a minor but definite decrease in pulmonary blood flow below the maximum as pulmonary venous pressure continues to decrease below alveolar pulmonary pressure; (4) the value of pulmonary venous pressure at which flow limitation is first reached is experimentally equivalent to alveolar pressure, although theory predicts a lower pulmonary venous pressure.

In addition to comparison between theoretical and experimental results in these experiments, it is feasible to compare our theoretical calculations with experimental data of others. Recently, Nagasaka et al (1984), using micropuncture on the superficial or pleural vessels of the cat lung, has provided data for comparison with our theoretical calculations. Table 26.3 shows the comparison. The theoretical data in Table 26.3 refer to the case in which  $p_a = 16.9$ ,  $p_A = 7.0$ ,  $p_v = 10.0 \text{ cm H}_2\text{O}$ , for a zone 3 condition. Our theoretical pressure is calculated at the exit end of a venular vessel of order 2, with a diameter of  $46 \mu\text{m}$ . From the table, we see that our theoretical results compare quite well with the experimental results of Nagasaka et al (1984). This is surprising since the microvessels studied by Nagasaka et al were pleural vessels. None of the values cited here for the cat pulmonary circulation for geometry and elasticity are for pleural vessels. The answer is not obvious, although the experimental model and the theoretical model may be different.

This is the first time a theoretical calculation of pulmonary circulation of a whole lung of an animal is made on the basis of a complete set of experimental data on the morphology of the vascular tree and elasticity of blood vessels of all orders. In general, theoretical results compare well with experimental results, but a few discrepancies exist that suggest further work is needed to fully correlate the theoretical and the experimental models.

TABLE 26.3. Comparison of theoretical results with experimental results of Nagasaka.

Method	Vessels	Pressure in zone 3 cdn. (cm H <sub>2</sub> O)	Pulmonary artery (cm H <sub>2</sub> O)	Alveolar pressure (cm H <sub>2</sub> O)	Left atrium (cm H <sub>2</sub> O)	Reference
Micro- puncture	Venules (30–50 $\mu\text{m}$ )	$11.5 \pm 1.1$	$17.8 \pm 1.1$	8	$8.9 \pm 0.4$	Nagasaka (1984)
From theory	Order 2 (46 $\mu\text{m}$ )	11.9	16.9	7	10.0	This report, see text

## Acknowledgments

This investigation was supported by the U.S. Public Health Service through Grants NIH HL-00854, HL-26647, HL-07089, and by the National Science Foundation through Grant NSF CME 79-10560 at UCSD and HL-11152 at U.S.C.

## References

- Fenn, W.O., and Rahn, H., eds. 1964. In *Handbook of physiology*. Section 3, *Respiration*, vol. I and II. Am. Physiol. Soc., Washington, D.C.
- Fung, Y.C., and Sabin, S.S. 1969. *J. Appl. Physiol.: Respir. Environ. Exercise Physiol.* 26(4):472-488.
- \_\_\_\_\_. 1972. *Circ. Res.* 30:440-450.
- Fung, Y.C., Sabin, S.S., Tremer, H., Yen, R.T., and Ho, H.H. 1983. *J. Appl. Physiol.: Respir. Environ. Exercise Physiol.* 54:1538-1549.
- Horsfield, K. 1978. *Circ. Res.* 42:593-597.
- Horsfield, K., and Gordon, W.I. 1981. *Lung* 159:211-218.
- Malpighi, M. 1661. *De Pulmonibus*, J. Young, transl. 1930. *Proc. Royal Soc. Med.* 23:1.
- Nagasaki, Y., Bhattacharya, J., Nanjo, S., Gropper, M.A., and Staub, N.C. 1984. *Circ. Res.* 54:90-95.
- Singhal, S., Henderson, R., Horsfield, K., Harding, K., and Cumming, G. 1973. *Circ. Res.* 33:190-197.
- Sabin, S.S., Fung, Y.C., Tremer, H.M., and Rosenquist, T.H. 1972. *Circ. Res.* 30:440-450.
- \_\_\_\_\_. 1980. *Microvasc. Res.* 19:217-233.
- Sabin, S.S., Lindal, R.G., Fung, Y.C., and Tremer, H.M. 1978. *Microvasc. Res.* 15:57-68.
- Strahler, A.N. 1957. *Trans. Am. Geophys. Union* 38:913-920.
- Weibel, E.R. 1963. *Morphometry of the human lung*. Springer-Verlag, Berlin.
- Weibel, E.R., and Gil, J. 1977. In *Bioengineering aspects of the lung*, J.B West, ed. Marcel Dekker, New York.
- Weiner, D.E., Verrier, R.L., Miller, D.T., and Lefer, A.M. 1967. *Am. J. Physiol.* 213:473-476.
- Yen, R.T., and Foppiano, L. 1981. *J. Biomech. Eng. Trans. ASME* 103:38-42.
- Yen, R.T., and Fung, Y.C. 1973. *J. Appl. Physiol.: Respir. Environ. Exercise Physiol.* 35:510-517.
- Yen, R.T., Fung, Y.C., and Bingham, N. 1980. *J. Biomech. Eng. Trans. ASME* 102:170-177.
- Yen, R.T., Zhuang, F.Y., Fung, Y.C., Ho, H.H., Tremer, H., and Sabin, S.S. 1983. *J. Appl. Physiol.: Respir. Environ. Exercise Physiol.* 55:236-242.
- \_\_\_\_\_. 1984. *J. Biomech. Eng. Trans. ASME* 106:131-136.
- Zhuang, F.Y., Fung, Y.C., and Yen, R.T. 1983. *J. Appl. Physiol: Respir. Environ. Exercise Physiol.* 55:1341-1348.
- Zhuang, F.Y., Yen, R.T., Fung, Y.C., and Sabin, S.S. 1985. *Microvasc. Res.* In press.

# Respiratory Dynamics—Computer Simulation

Y. SEGUCHI, Y.C. FUNG, AND T. ISHIDA

## 1. Introduction

The modeling of respiratory dynamics has been attempted for many years. The state of the art, however, is not yet responsive to the needs for understanding all aspects of circulation and ventilation.

Generally, an academic discipline tends to shift from the black box approach to analysis with explicit details. Rapid initial advances can often be made by the black box approach, but more practical applications usually require explicit details. A scientific discipline has its own life cycle also.

Many necessary data on the morphology and function of the respiratory system are still unknown, although much has been accumulated. As the data base grows, one can look into the ventilation system in greater detail. The more we do, the more complex the system will appear to be. Not only does the system tend to become larger, but also the interaction between subsystems becomes significant; and often for a living system, the isolation of a subsystem tends to be less meaningful. Usually the differential equations can be solved only for a simple isolated system, having become too complex for the complete system. In such cases, the alternative approach of computer simulation becomes more attractive.

System simulation is a powerful tool to meet the new demands of practical applications. The mathematical model can be as specific as the state-of-the-art data base permits. The results will be quantitative. A good simulation can give us insight into the nature of the system, and may lead us to recognize and develop new fields of research (see Vichenevetsky, 1976).

## 2. Guidelines for Model Building

The term *computer simulation* has been used in the literature for many years, but there seems to be no precise definition. If all the structural details of a system are known, then simulation means the building of a model that behaves like the original, and any convenient procedure may be used. However, all

details of a living system are not known, hence guidelines for computer simulation must be specified. Our guidelines are the following:

1. Oversimplification should be avoided. The characteristics of living systems have no simple mechanical or electrical analogs. Hence analog circuits of electrical components should be avoided even if the merit of simplification seems to be significant from a computational point of view.
2. The model should be built in such a way that subsystems, sub-subsystems, etc. can be assembled easily. It is from this point of view that we caution against simplified electric analogs because a careless introduction of analog may make it difficult to include the hierachial concept of systems. Non-linearities arising from material property and large deformation, and unusual changes in system structure and morphology are important to the behavior of living systems over a wide macro-micro spectrum and should be simulated.
3. The morphology or anatomy should be described in as much detail as possible in the system model.
4. Description of the system by differential equations with initial and boundary conditions should be done in mathematically consistent form. The number of equations should not be too few or too numerous. Nonconservative features of the system should be carefully preserved. Variational approach is often very useful in mathematical formulation and in obtaining the solutions.
5. If the behavior of a complex system must be described by using continuous variables, one may find that the only feasible method of solution is by discretization and computer simulation. Variational formulation is again useful for discretization. Since the solution depends on the computational procedure, it is important to have the formulation match the computational technique.
6. A successful system simulation may include analog components of black boxes, as well as discretized differential equations. The solution can be supplemented by the classical solutions and experimental data.

### 3. Preparation for Model Building

Although the lung does not exist alone, one may still consider it in isolation in order to initiate a simulation, as an excised lung cut away from the surrounding organs such as the heart, chest wall, diaphragm, abdomen, larynx, and the mouth. This isolated lung is embedded in a larger system through the pleura. Moreover, from the functional point of view, the lung is closely related to both ventilation (external respiration) and circulation (internal respiration), and thus it also interacts with the systemic arteries, capillaries, and veins. The relationship between ventilation and circulation is regarded as that of primal and complement. The situation of the lung within the whole function of life is

so complex that a decoupling between the primal and the complement is desirable by assuming that the circulatory function is maintained in constant or normal state. Moreover, another decoupling can be done between mechanical and nonmechanical aspects of ventilation. In this paper, we shall consider only the mechanical aspect of ventilation, i.e., breathing. For the last decoupling, we assume again that the nonmechanical factor of ventilation is kept constant or normal.

Even after all these preparations, the system is still very complex, as is shown below.

## Morphology

The lung is divided into two main subsystems: the airway tree and the lung parenchyma. The upper and lower airways are distinguished at the lobular level, and the tree system encompasses up to 24 generations, the last being the alveoli. The part of the airway tree distal to the respiratory bronchioles is considered as a part of the lung parenchyma. The airway tree is completely surrounded by the parenchyma except for a part of the upper trachea. The airway system has an asymmetric dichotomous structure the morphology and dimensions of which are widely scattered for each individual. For the human lung, the dichotomous model proposed by Weibel (1963) can be considered as a standard because of its completeness of data.

The lung parenchyma consists of the alveolar ducts, alveolar sacs, alveoli, and their walls, the interalveolar septa. Since the characteristic dimension of the parenchyma microstructure is of the order of 0.1 mm, the parenchyma may be regarded as a continuum in problems in which the characteristic dimensions are much larger, say, on the order of 1 cm or more.

## Mechanical Behavior of Materials

Material properties determine system behavior. The constitutive equations of many tissues, however, are still unknown. The properties of the lung parenchyma as a continuum has been studied, and a constitutive equation has been proposed (Fung, 1975). The constitutive equation includes the effect of surface tension on the interalveolar septa. The surface tension has a large hysteresis when the lung is subjected to cyclic loading. On the other hand, the mechanical properties of the airway, which is pseudoelastic, are still largely unknown.

## Interaction of Tissues and Gas

As Fung (1981) pointed out, the deformation of airway can have significant influence on airway dynamics under certain circumstances. The equations of motion and continuity of gas through the airway should be formulated for large deformation of the airway.

The gas in the airway is compressible and viscous. Energy is dissipated in the flow by viscous stresses, by turbulence, by boundary layer growth at points of bifurcation in the airway network, and by the generation of secondary flow and vortices. The resistance at branching points cannot be neglected in the equation of motion. The relations proposed by Kunkle et al 1969 and used by Pardaens et al 1972 are considered to be sufficiently realistic for simulation.

The excised lung and its environment interact mechanically through the pleura. Fung (1975) shows that the pleural pressure  $P_{pl}$  equilibrates with the alveolar gas pressure  $P_{al}$  minus the stress  $\sigma$  in the parenchyma and the product  $N_{pl}\kappa_{pl}$ , where  $N_{pl}$  is the membrane tension in the visceral pleura, and  $\kappa_{pl}$  is the curvature of pleural membrane. Thus,

$$P_{pl} = P_{al} - \sigma - N_{pl}\kappa_{pl}. \quad (27.1)$$

The curvature term could be disregarded if the lung volume is sufficiently small ( $N_{pl}$  small). The alveolar gas pressure influences the tree because the transmural pressure of the airway is the gas pressure in the airway minus the alveolar gas pressure and the parenchyma stress  $\sigma$ .

## 4. Assumptions for Simulation

The information mentioned above is insufficient for a complete simulation of the lung. There are many other important parameters for which information is lacking, and whose values have to be assumed. We list them below:

1. The airway tube of each generation is assumed to be straight, taperless, and of circular cross section. The length-to-diameter ratio of each segment is sufficiently large so that the governing differential equations of aerodynamics can be appropriately simplified as one-dimensional. These simplifications naturally lead to another assumption of axisymmetric flow and stress distribution in the airway wall. Moreover, in the airway wall, the radial and axial stresses are assumed to be negligibly small. The last assumption might be incorrect, but it is more reasonable than neglecting the longitudinal deformation, because the airway deformation cannot be disregarded (Hoppin et al, 1977).
2. The interaction between the parenchyma and the airways is still not entirely clear, especially for smaller airways. The thickness of the airway wall is difficult to identify. Data on the distensibility of airways embedded in parenchyma are scarce, especially for humans.
3. The detailed geometry of neighboring organs is ignored in the simulation. The lung parenchyma is assumed to be uniform, homogeneous, and isotropic in all lobes.
4. The effect of the gravitational force is disregarded except when the stress distribution in the lung is studied or when the effect of posture is considered.
5. The natural state, defined as the state of zero stress and zero strain, must be known before the deformation under load can be computed. Because of the

possible existence of residual stresses, the no-load condition may not be the natural state. Hence the natural state is a simulation parameter. For the lung, the residual stress is unknown, and the natural state is yet to be determined.

## 5. Illustrations of a Computer Model of Simulated Lung

The configuration of the model shown in Figures 27.1 and 27.2 is based on the assumptions named above. Figure 27.1 shows the structural relation between various subsystems. Figures 27.2a–c show the functional interactions among subsystems.

The mathematical expression of the system is rather complicated. Some formulas are given in Figures 27.2a–c. A brief summary is given below. Further details are given in Seguchi et al (1984).

On denoting the mean velocity of the flow averaged over each cross section of the airway by  $v$  and the mean pressure by  $P$ , the approximate “one-dimensional” equations of motion and continuity can be written, respectively, as

$$\dot{v} + P_x/\rho + 8\mu v/\rho r^2 = 0 \quad (27.2)$$

$$(\rho A)_t + (\rho Av)_x = 0 \quad (27.3)$$

where  $\rho$  is the density of the gas,  $r$  is the radius of the tube,  $\mu$  is the coefficient of viscosity of the gas, and the subscripts  $x$  and  $t$  indicate partial differentiation with respect to  $x$  and  $t$ , respectively. A dot over a variable denotes a material

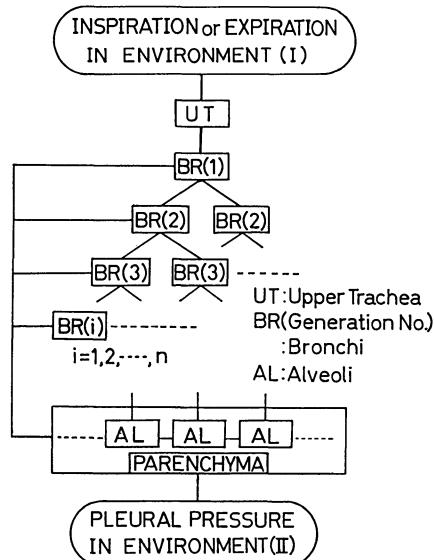
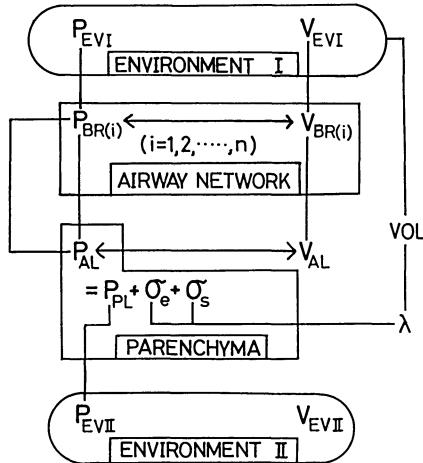


FIGURE 27.1. Block diagram of structural configuration of the system.

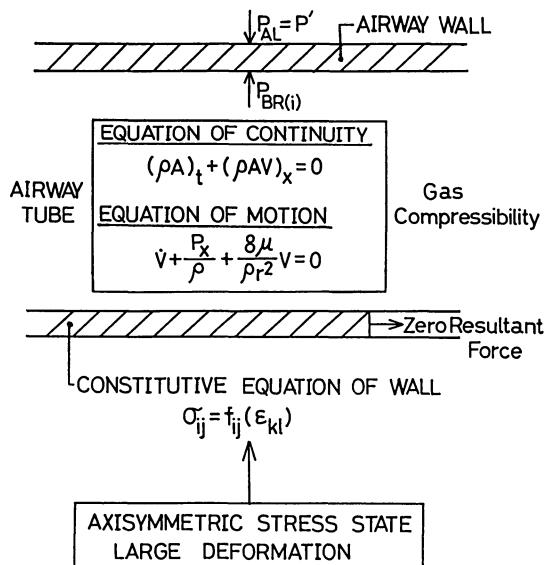


SUFFIX

**STATE VARIABLES**

- P: Pressure
- V: Flow Velocity
- EVI: Environment I
- EVII: Environment II
- PL: Pleural
- $\sigma_s$ : Surface Tension
- $\sigma_e$ : Elastic Tissue Stress
- VOL: Lung Volume
- $\lambda$ : Stretch Ratio

a



b

$\rho$ : Density

$\mu$ : Coefficient of Viscosity

A : Cross Sectional Area

FIGURE 27.2. (a) Block diagram of functional configuration of the system. (b) Configuration of airway subsystem. (c) Configuration of parenchyma subsystem.

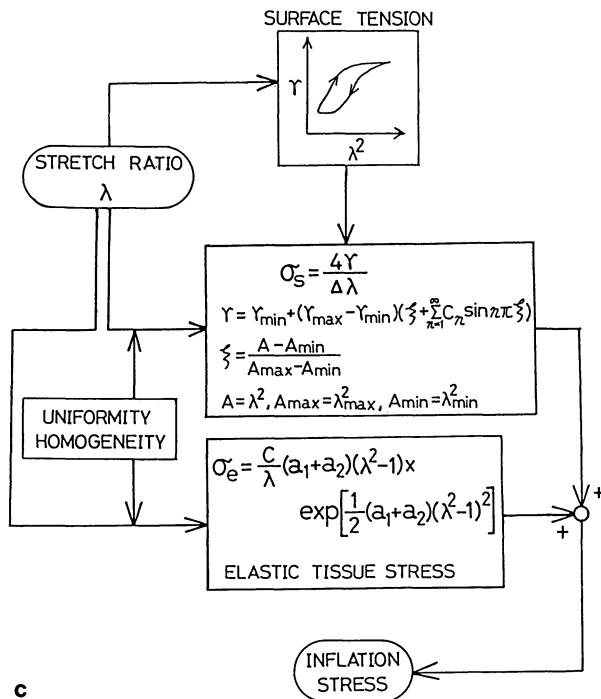


FIGURE 27.2 (c) continued.

derivative, e.g.,  $\dot{v} = v_t + vv_x$ . Carrying out the differentiation in Equation (27.3), regrouping the terms and dividing through with  $\rho A$ , we obtain

$$\dot{A}/A + \dot{\rho}/\rho + v_x = 0 \quad (27.4a)$$

This can be written as

$$2\dot{\epsilon}_\theta + v_x + \dot{P}/K = 0, \quad (27.4b)$$

if  $\dot{A}/A$  is expressed in terms of the rate of change of the circumferential strain of the tube wall as seen by the moving fluid,  $\dot{\epsilon}_\theta$ , and  $\dot{\rho}/\rho$  is expressed in terms of the rate of change of the pressure divided by the bulk modulus of elasticity of the fluid,  $K$ , so that

$$\dot{A}/A = 2\dot{\epsilon}_\theta, \quad \dot{\rho}/\rho = \dot{P}/K. \quad (27.5)$$

Equation (27.4b) is a form of the equation of continuity given by Streeter et al (1967). The strain  $\epsilon_\theta$  can be expressed in terms of the stresses in the tube wall through the constitutive equation of the tube. The stresses in the tube wall are related to the transmural pressure by statics. Use of these relations in Equation (27.4b) will reduce the equation of continuity into one that relates  $P$  and  $v$  and their derivatives.

The final governing equations thus obtained are discretized by using the

finite-element technique. A summary of the procedure is as follows: First, the system of differential equations is transformed into a variational principle, as the conditions for the existence of a stationary value of a certain integral that is a functional of  $p(x, t)$  and  $v(x, t)$ . However, since the force system is non-conservative, and the original system of equations are non-self-adjoint, the existence of such a stationary value integral is not guaranteed. Seguchi et al (1984) have obtained such a functional by extending the method of Lagrangian multipliers. This method is a modification of the optimization technique, which transforms a problem of optimization with constraints into an unconstrained optimization. The Lagrangian multipliers are the variables of an adjoint system. It turns out that the primal system is independent of the adjoint system, while the converse is not true. Since our problem is highly nonlinear, the process is made simple by quasi-linearization. It is natural that the scheme leads to iterative computation. The number of the state variables is increased by this method. Whether the stationary value of the functional is a maximum or a minimum is uncertain. A rigorous theory of the method remains a challenge, but the application to some test examples has been successful. In an investigation of the stability of a column subjected to a force that is always tangential to the column (the so-called follower force), Prasad and Herrmann (1969, 1972) used a similar integral and showed that the method is more advantageous than the Galerkin's method and its extension, such as the weighted residual method.

Next, each generation or branch of the airway is considered as an element or is subdivided into finite elements. The state variables are evaluated at the nodal points and extended by interpolation polynomials to be continuous functions. These state variables are substituted into the functional to obtain the matrix description of the element behavior. Terms expressing the energy loss at the points of bifurcation of the airway tree and the sudden change of kinetic energy due to sudden change in cross-sectional area at these bifurcation points are added to the assemblage. The conditions for assembling the adjoint systems are automatically decided by the condition for the existence of a stationary value of the functional. In the example for which results are given in Figures 27.3–27.7, below, the number of degrees of freedom of the primal system is 48 in the case of symmetric dichotomy. Any irregular dichotomy can be dealt with, but the degrees of freedom increase as the number of irregular dichotomy increases. Derivatives with respect to time are replaced by finite differences and are integrated by the conventional numerical integration scheme.

Initial and boundary conditions must be specified for the state variables. Any initial states can be studied. The problem becomes well posed if two of four state variables are specified at the boundaries. Examples are:

1. The pleural pressure and the pressure at the inlet are specified. In normal breathing, the pressure at the mouth is atmospheric; the flow is controlled by the pleural pressure.

2. The velocity of flow and the pressure are specified at the inlet. In this case, the problem is reduced to the estimation of the pleural pressure. Once the pleural pressure is estimated, the solution follows case 1, above. The solution can be obtained by iteration.

## 6. Case Studies

We did several case studies by the proposed simulation technique. In the simulations, the parenchymal stress  $\sigma$  is given as

$$\begin{aligned}\sigma &= \sigma_e + \sigma_s, \quad \sigma_s = 4\gamma/\Delta\lambda, \\ \sigma_e &= (2/\lambda)c(a_1 + a_4)e \exp[a_1 + a_2 + 2a_4]e^2, \\ e &= (\lambda^2 - 1)/2,\end{aligned}\tag{27.6}$$

where  $\sigma_e$  and  $\sigma_s$ , respectively, the stress due to the tissue elasticity and that due to surface tension,  $\lambda$  is the stretch ratio,  $\gamma$  is the surface tension (force/length), and  $\Delta$  is the average diameter of the alveoli. The assumed values of the parameters are

$$\begin{aligned}c &= 19.13 \text{ g/cm}, \quad \Delta = 0.01 \text{ cm}, \\ a_1 = a_2 &= 0.6137, \quad \text{and} \quad a_4 = 0.4235.\end{aligned}$$

These numbers, unfortunately, refer to the dog. They were collected from the papers of Hoppin et al (1975) and Lee et al (1975). The surface tension data were taken from Fung (1975). The airway elasticity was simplified to be linear, and the elastic moduli were assumed rather arbitrarily because of a paucity of reliable data—see the papers of Croteau et al (1961), Hughes et al (1972), Tisi et al (1975), and Nakagawa et al (1978). Airway wall thickness is still a simulation parameter, and here it is set as 10% of airway radius. Both symmetric and asymmetric dichotomies were simulated, and normal and obstructive airways with varying stiffness were studied. The morphological model of man given by Weibel (at 75% total lung capacity) was selected as the natural state.

The results of the simulation calculations for breathing at normal frequency are shown in Figures 27.3, 27.4, and 27.5. In these figures, the airway dichotomy is symmetric, the pressure at the mouth is atmospheric, and the pressure in the pleura is varied, as shown in the figures. Figure 27.3a shows the variations of the pressures, flow, and volume over one cycle when the breathing frequency is 0.5 Hz. Figure 27.3b shows the corresponding variables at 1 Hz. Figure 27.4 shows the frequency dependence, or the Bode diagrams, of respiration. Four cases are shown. The “reference” case is the one having the elastic moduli of the airway wall listed in the box for different generations. Case A is for a “stiff” lung with  $E$  values 100 times larger than those of the reference lung. Case B is for a “flexible” lung with  $E$  values one tenth those of the reference lung. Case C is for a lung with obstructive disease, the  $E$  values

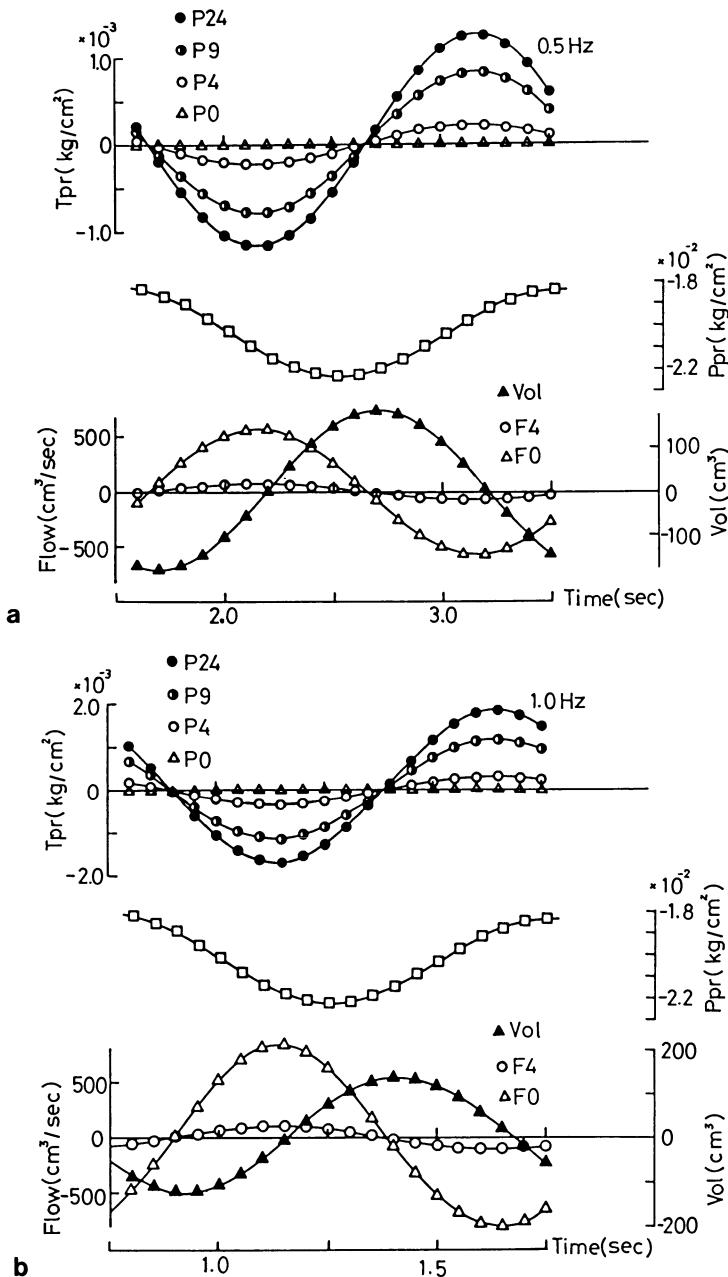


FIGURE 27.3. (a) A typical cycle of respiration at 0.5 Hz. Oscillations of the transmural pressures ( $T_{pr}$  = airway pressure – alveolar gas pressure + parenchyma stress  $\sigma$ ), flow rates in typical generations ( $F_4, F_0$ ), and the lung volume (vol) are shown. Pleural pressure ( $P_{pr}$ ) was controlled. (b) A typical cycle of respiration at 1.0 Hz. Oscillations of the transmural pressure ( $T_{pr}$ ), flow rates in typical generations ( $F_4, F_0$ ), and the lung volume (vol) are shown. Pleural pressure ( $P_{pr}$ ) was controlled.

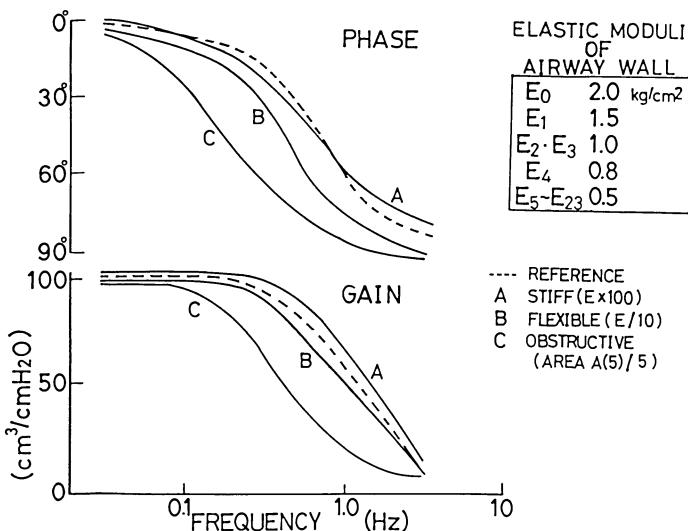


FIGURE 27.4. Frequency dependence of respiration, or the Bode diagrams, showing the change in amplitude and phase relationship between the flow at the mouth and the pleural pressure. The elastic moduli of the airway wall at different generations are listed in the figure. The dotted curve is labeled as the “reference” case, which has the  $E$  values in the box. Case A is for a “stiff” lung with  $E$  values 100 times greater than those listed in the box. Case B is for a “flexible” lung with  $E$  values one-tenth those of the reference lung. In case C, the  $E$  values are those of the reference lung, but the cross-sectional area of the airway of the fifth generation is reduced to 20% of the normal cross-sectional area.

of which are those of the reference lung, but the cross-sectional area of the airway of the fifth generation is reduced to 20% of the normal cross-sectional area. It is seen that a decrease in elastic modulus decreases the gain (amplitude of flow at the same pleural pressure fluctuation) at a given frequency.

Figure 27.5 shows the volume versus pleural pressure loops in cyclic motion for the “reference” lung, the “flexible” lung, and the “obstructive” case, corresponding to the cases in Figure 27.4. Large hysteresis loops are seen. The size of the hysteresis depends on the lung stiffness and the breathing frequency.

Forced expiration characteristics are considered to be sensitive to lung disease. (Menkes, 1981). If the velocity of flow at the mouth is measured and used as a boundary value, we can compute the pleural pressure in forced expiration by a method of iteration. Figure 27.6 shows several examples of forced expiration of normal and obstructive lungs. The “normal” lung has the same parameters as the reference lung in Figure 27.4. The obstructive lung A(2)/5 has its airway of the second generation reduced in cross section to one-fifth of the normal. Similarly, case A(5)/5 has its airway of the fifth generation reduced to one-fifth of the normal; A(7)/5 has the seventh generation reduced. In this calculation, the same estimated pleural pressure was used in all cases. It

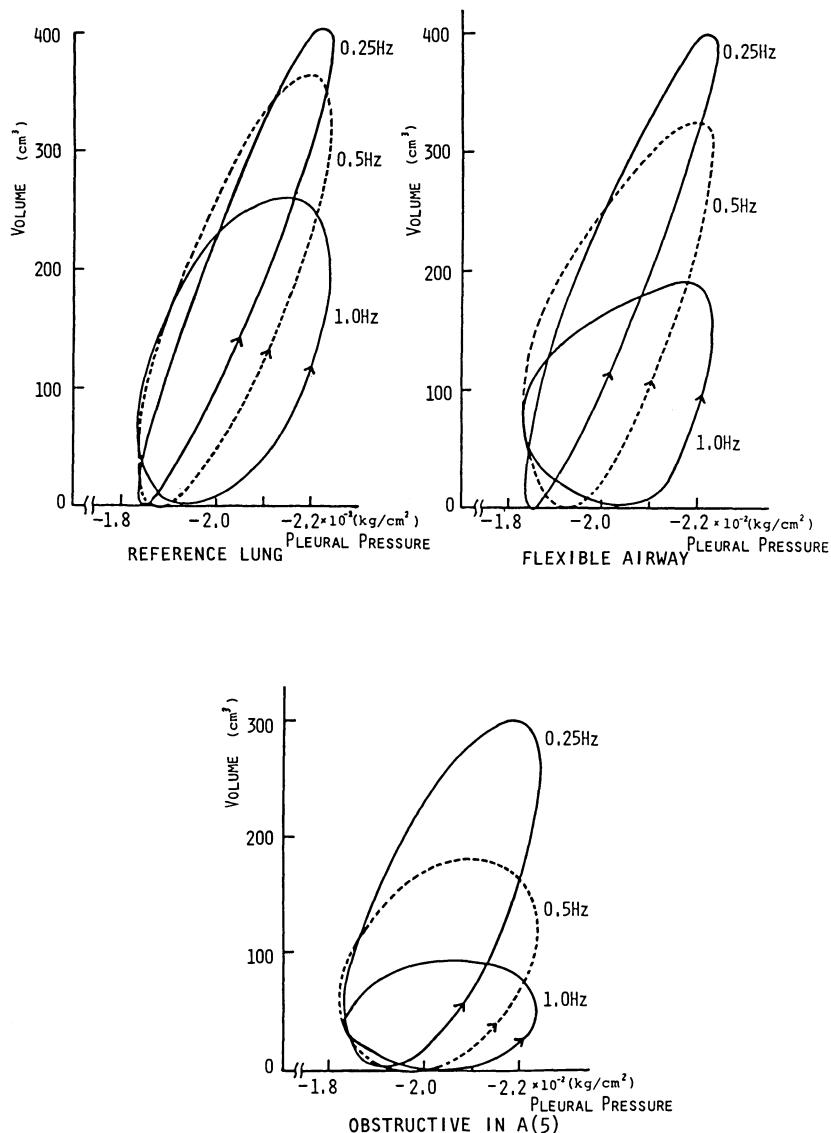
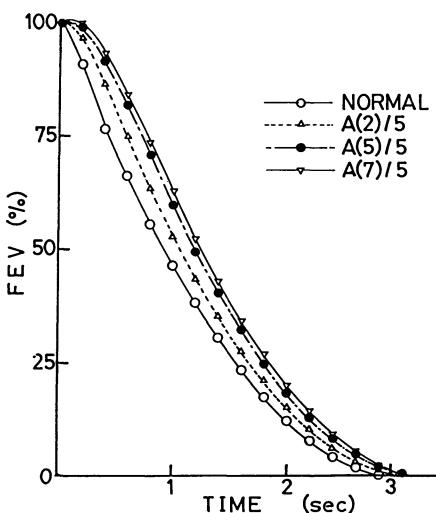


FIGURE 27.5. Hysteresis loops of lung volume versus pleural pressure for a lung with symmetric dichotomy. The cases denoted as reference, flexible, and obstructive, are defined by the same sets of parameters shown in Figure 27.4.

**FIGURE 27.6.** Simulation of forced expiration. The lung volume under forced expiration (FEV) is plotted as a function of time for five cases.  $A(n)/5$  means that the airway of  $n$ th generation is reduced by obstruction to 20% of the normal cross-sectional area.



is seen that obstruction reduces the flow so that the time it takes to reach the same lung volume becomes longer with obstruction. It is seen also that obstruction in the fifth and seventh generation of airway has a larger effect than obstruction of the second generation.

Finally, we present in Figure 27.7 a study of high-frequency ventilation that is attracting increasing interest. The subject is more complex than the simple respiratory dynamics because attention should be focused on  $O_2-CO_2$  exchange due to diffusion and convection. But we shall limit ourselves to looking at the flow dynamics during inspiration and expiration at higher frequencies. Figure 27.7 shows the results of breathing at frequencies of 5, 15, and 25 Hz under the same sinusoidal pleural pressure. The maximum and minimum transmural pressures in each generation of airway are plotted against the generation number. The steeper-than-normal pressure distribution along the airway in higher frequency is observed.

These examples illustrate the principle, but not the practical applications. The more case studies we make, the further we would like to modify and extend. At the same time, we notice the lack of morphological and functional data on many important tissues and organs. Undoubtedly, the simulation model will be improved from time to time as the data base grows. Thus the detailed description here might not be important, but the concept proposed might be. This is the reason that the description was so brief and so qualitative. We hope that in the near future the simulation technique will become an important methodology for physiology and medicine and that it will be applied to practical clinical problems.

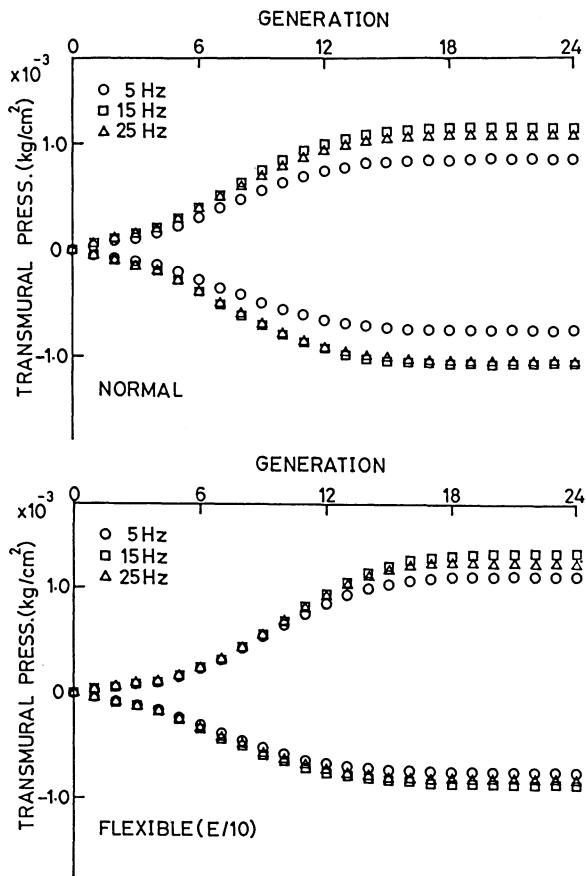


FIGURE 27.7. Simulation of high-frequency ventilation. The maximum and minimum transmural pressure in each generation is plotted as a function of the generation number at frequencies of 5 Hz, 15 Hz, and 25 Hz. "Flexible" means that compliance of airway wall is ten times the normal.

## Acknowledgment

The technical assistance and computer implementation by Messrs. Y. Ichioka, H. Maki, and H. Nagatani are gratefully acknowledged.

## References

- Croteau, J.R., and Cook, C.D. 1961. *J. Appl. Physiol.* 16(1): 170-172.
- Fung, Y.C. 1975. *Circ. Res.* 37: 481-496.
- \_\_\_\_\_. 1981. *J. Biomech. Eng.* 103: 91-96.

- Hopkin, F.G., Jr., Hughes, J.M.B., and Mead, J. 1977. *J. Appl. Physiol.* 42(5): 773–781.
- Hughes, J.M.B., Hopkin, F.G., Jr., and Mead, J. 1972. *J. Appl. Physiol.* 32(1): 25–35.
- Kunkle, J., Wilson, S.D., and Cota, R.A. 1969. *Compressed gas handbook*. NASA-SP-3045. U.S. Govt. Printing Office, Washington, D.C.
- Lee, G., and Frankus, A. 1975. *Biophys. J.* 15: 481–493.
- Menkes, H., Cohen, B., Permutt, S., Beatty, T., and Shelhamer, J. 1981. *Annals Biomed. Eng.* 9: 501–511.
- Nakagawa, T., Seguchi, Y., and Fung, Y.C. 1978. *Proc. 3rd Intern. Cong. of Biorheology* La Jolla, California.: 65.
- Pardaens, J., Van De Woestijne, K.P., and Clement, J. 1972. *J. Appl. Physiol.* 33: 479–490.
- Prasad, S.N., and Herrmann, G. 1969. *Int. J. Solids Struct.* 5: 727–735.
- \_\_\_\_\_. 1972. *Int. J. Solids Struct.* 8: 29–40.
- Seguchi, Y., Fung, Y.C., and Maki, H. 1984. *Biomech. in China, Japan, and U.S.A.*, Science Press, Beijing: 224–239.
- Streeter, V.L., and Wylie, E.B. 1967. *Hydraulic transients*. McGraw-Hill, New York.
- Tisi, G.M., Minh, V.D., and Friedman, P.J. 1975. *J. Appl. Physiol.* 39: 23–29.
- Vichenevetsky, P. 1976. *Simulation of systems*, L. Dekker, ed. Elsevier, North-Holland, Amsterdam.
- Weibel, E.R. 1963. *Morphometry of human lung*. Academic Press, New York.

# Index

- Adhesion, 3
- Adhesion energy, 9
- Airways, 380
- Alveoli, pulmonary, 365
- Aorta, 182
- Arteries
  - endothelial, 332
  - residual stress, 117
  - stress distribution, 122–128
  - unstressed state, 118
- Articular cartilage, 153–177
- Artificial joint, 245
- Atherosclerosis, 325, 337
  
- B**iaxial loading
  - pericardium, 86
- Bifurcation
  - arterial, 334
  - microvascular, 289
- Biphasic model for cartilage, 156
- Blood flow
  - capillary, 299–311
  - microvascular, 283–298, 314–322
  - pulmonary, 365
- Boltzman's distribution, 145
- Bone, 196–208
  - density, 194, 204
  - implants, 245
  - mechanical properties, 197–203
- Bovine aortic endothelium, 326
  
- C**apillary, 283, 299
  - flow, 302
- hematocrit, 308, 353, 362
- permeability, 292
- pulmonary, 367
- Cardiac coordinate system, 101
- Cell culture, 325
- Cervical spine, *see* Spine
- Collagen
  - fiber density, 131
  - viscoelasticity, 132
- Collapsible tube, 344
- Computer tomography, 197
- Constitutive equation
  - active heart muscle, 84
  - articular cartilage, 156
  - ligaments, 184–187
  - passive pericardium, 88
- Contact pressure, 225
- Cortical bone, 197
- Cross bridges, 7
- Cytoskeleton, 326
  
- D**ensity, 359
- Diarthrodial joint, 153
- Digestive tract
  - wound repair, 276
- Disc, *see* Lumbar disc
- Distensibility, blood vessels, 366
  
- Endocytosis, 329
- Endothelium, 324
  - cytoskeleton, 326
  - mechanical properties, 330
- Erythrocyte, *see* Red blood cell

- Exchange, blood-tissue, 290  
 Exchange-perfusion parameter, 295
- Fading memory model, 80  
 Fahraeus effect, 361  
**Fascia**  
 wound repair, 275  
 Femoral bone, 199  
 Fibroblast, 273  
 Fibroplasia, 264  
**Finite element analysis**  
 cell membrane, 304  
 hip prosthesis, 252  
 lumbar spine, 215  
 respiration, 384  
**Flow in collapsible tube**, 342  
**Flow limitation**, 371  
**Foot reactive force**, 231  
**Fung's quasilinear viscoelastic law**, *see Quasilinear viscoelastic law*
- Gait**, 225–242  
**Granulation tissue**, 264  
**Ground reaction force**, 231
- Heart muscle**, 77, 86, 98  
 constitutive equation, 84, 88  
 finite strain, 101  
 principal strains, 107  
 regional deformation, 99  
**Hematocrit**, 308  
 capillary, 308, 353, 362  
 pulmonary, 325  
**High frequency ventilation**, 390  
**Hip prosthesis**, 249  
**Hulett's theory for osmotic pressure**, 145
- Implant**, 246  
**Indicator-dilution**, 357  
**Inflammation**, 263  
**Interface loosening**, 246  
**Inverse dynamic problem**, 237
- Joint motion**, 227  
**Joint replacement**, 245
- Knee joint**, 228  
**Korteweg-deVries equation**, 19–20
- Leukocyte**  
 adhesion, 45  
 capillary motion, 308  
 contraction ring, 40, 44  
 cytoplasm, 35  
 membrane, 35  
 properties, 41  
 protopodia, 33  
 wound healing, 264  
**Lewis theory for osmotic pressure**, 148  
**Ligaments**, 180–193  
 healing rates, 273  
 immobilization, 187  
 injury, 191  
 medial-collateral ligament, 183  
**Lower back pain**, 210  
**Lubrication theory in capillary flow**,
- Lumbar disc**, 219–222  
**Lumbar spine**, *see Spine*  
**Lumbar vertebrae**, 203, 211  
**Lung**  
 blood flow, 367  
 hematocrit, 353
- Membrane**  
 adhesion, 7  
 area dilatation, 6, 35  
 crossbridges, 7–16  
 deformation, 6  
 properties, 6  
 separation, 13–16  
**Mesenchymal cell**, 273  
**Metaphyseal shell**, 202  
**Microcirculation**, 283, 299, 314, 353  
 pulmonary, 367  
**Morphometry**, 353  
**Muscle**, *see also Smooth muscle or Heart muscle*  
 wound healing, 275
- Networks**  
 microvascular, 286

- Oscillation**, 343  
**Osmosis**, 142  
**Osteoporosis**, 196
- Papillary muscle**, 79  
**Penrose tube**, 345  
**Perfusion**, microvascular, 296  
**Permeability**, 292  
**Pleura**, 380  
**Preconditioning**  
  smooth muscle, 69  
**Prony spectral analysis**, 317  
**Prosthesis**, joint, 246  
**Protopodia**, 33  
**Pseudostrain energy function**  
  arteries, 120  
  pericardium, 87  
**Pulmonary**  
  blood flow, 365  
  morphology, 368, 380  
  pressure-flow relationship, 294, 299,  
    344, 372
- Quasilinear viscoelastic law**, 130  
  heart muscle, 83  
  ligaments, 184
- Raoult's law**, 147  
**Red blood cell**, 299  
  deformation in capillaries, 304  
  membrane lipid structure, 4  
  membrane properties, 6  
**Regional strains**, 98  
**Residual stress**, 117  
**Resistance**, pulmonary, 370  
**Respiration**, 377  
**Runaway solitions**, 24
- Shape index**, endothelium, 326  
**Sheet flow**, 366  
**Simulation**, 377  
**Skin**, 137, 181  
  wound healing, 274
- Smooth muscle**  
  dose-response, 56–60  
  isometric force, 67  
  length-tension relationship, 55  
  reference length, 64  
  viscoelastic properties, 70, 71
- Solitions**, 18  
**Spine**, 196, 210  
  material properties, 214  
**Starling relationship**, 291  
**Strain**  
  myocardium, 98  
**Stress-strain relationship**, 80, 83, 88  
  arterial wall, 120  
  bone, 197  
  cartilage, 162  
  erythrocyte, 3  
  leukocyte, 41  
  tendon, 184
- Taemia coli**, 52  
**Tank-treading**, 307  
**Temperature effect on**  
  smooth muscle, 74  
**Tendon**, 180, 182  
  healing rates, 273  
**Thoracic spine**, *see* Spine  
**Tibial bone**, 202  
**Trabecular bone**, 199–202  
**Tracheal smooth muscle**, 62–76
- Van derWaals attraction**, 11  
**Vapor pressure**, 146  
**Vasomotion**, 314  
**VDA**, *see* Video dimension analyzer  
**Vertebral body**, 212  
**Vertebral disc**, 215  
**Video dimension analyzer**, 183  
**Viscoelasticity**, 131  
  for heart, 83  
  for leukocyte, 41  
  for tendon, 184
- Water-fall**, 374  
**White blood cell**, *see* Leukocyte  
**Wound healing**, 263

# Related Titles from Springer-Verlag

*"Will surely be a standard text"\*\**

## Biomechanics

**Mechanical Properties of Living Tissues** Y.C. Fung

*"Professor Fung has consolidated in a clearly and concisely written textbook much of the existing mechanical descriptions of biologically derived fluids and tissues . . . For those of you interested in analysis of the mechanical properties of biologically derived tissues, this book is a must buy. Professor Fung has organized, consolidated, and unified this difficult field in an elegant and erudite textbook."*

—**Medical Physics**

*"A connected account of much of the progress that has been made in recent years . . . This is a book which will surely be a standard text for some time."*

—\***Journal of Biomechanical Engineering**

### Contents

Introduction: A Sketch of the History and Scope of the Field. The Meaning of the Constitutive Equation. The Flow Properties of Blood. Red Blood Cells and Their Deformability. The Rheology of Blood in Microvessels. Bio-viscoelastic Fluids. Bio-viscoelastic Solids. Mechanical Properties of Blood Vessels. Skeletal Muscle. Heart Muscle. Smooth Muscles. Bone and Cartilage.

1981/433 pp./224 illus./Cloth  
ISBN 0-387-90427-7

## Biodynamics: Circulation

**Y.C. Fung**, University of California, La Jolla

This advanced text is a concise, up-to-date presentation of the mechanics of blood circulation. It is the second of a unique three-volume series on biomechanics, written by a leader in the field of bioengineering. Using mathematics to bring anatomy, rheology, physiology, and engineering into a unified approach, **Biodynamics: Circulation** covers fundamental mechanics as well as physical principles of circulation, heart, arteries, veins, microcirculation, and pulmonary blood flow. The basic equations of fluid and solid mechanics are given in the Appendix. Theory and application receive equal attention, and the reader will appreciate the many well-designed exercises that illuminate the subject matter.

**Biodynamics: Circulation** includes a comprehensive guide to the literature. The outstanding illustrations, thorough introduction of qualitative concepts, and definitive mathematical approach—as well as the exercises—make Y.C. Fung's treatment of biomechanics easy to grasp, exciting, and above all useful to the reader.

### Contents

Physical Principles of Circulation. The Heart. Blood Flow in Arteries. The Veins. Microcirculation. Blood Flow in the Lung. Appendix: Basic Field Equations.

1984/404 pp./189 illus./Cloth  
ISBN 0-387-90867-6

To order, or for more information, write: Springer-Verlag New York, Inc.,  
175 Fifth Avenue, New York, New York 10010



Springer-Verlag  
New York Berlin Heidelberg Tokyo