

# product engineering

MOLECULAR STRUCTURE  
and PROPERTIES

**James Wei**

CD ROM INCLUDED



## Product Engineering

Topics in Chemical Engineering: A Series of Textbooks and Monographs

*Series Editor*

Keith E. Gubbins, North Carolina State University

*Associate Editors*

Mark A. Barteau, University of Delaware

Douglas A. Lauffenburger, MIT

Manfred Morari, ETH

W. Harmon Ray, University of Wisconsin

William B. Russel, Princeton University

Chimowitz, *Introduction to Critical Phenomena in Fluids*

Deen, *Analysis of Transport Phenomena*

Doraiswamy, *Organic Synthesis Engineering*

Floudas, *Nonlinear and Mixed Integer Optimization: Fundamentals and Applications*

Friedlander, *Smoke, Dust, and Haze, Second Edition*

Fuller, *Optical Rheometry of Complex Fluids*

Harrison, Todd, Rudge, and Petrides, *Bioseparations Science and Engineering*

Larson, *The Structure and Rheology of Complex Fluids*

Lauffenburger and Linderman, *Receptors: Models for Binding, Trafficking, and Signalling*

Morrison, *Understanding Rheology*

Ogunnaike and Ray, *Process Dynamics, Modeling, and Control*

Pearson, *Discrete-Time Dynamic Models*

Phan-Thein and Kim, *Microstructures in Elastic Media*

Pozrikidis, *An Introduction to Theoretical and Computational Fluid Dynamics*

Pozrikidis, *Numerical Computation in Science and Engineering*

Schmidt, *The Engineering of Chemical Reactions*

Varma and Morbidelli, *Mathematical Methods in Chemical Engineering*

Wei, *Product Engineering: Molecular Structure and Properties*

# ■ Product Engineering: Molecular Structure and Properties

James Wei

Princeton University

**OXFORD**  
UNIVERSITY PRESS

2007



Oxford University Press, Inc., publishes works that further  
Oxford University's objective of excellence in research, scholarship,  
and education.

Oxford New York  
Auckland Cape Town Dar es Salaam Delhi Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in  
Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2007 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.  
198 Madison Avenue, New York, New York 10016

[www.oup.com](http://www.oup.com)

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
electronic, mechanical, photocopying, recording, or otherwise,  
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data  
Wei, James, 1930-  
Product engineering : molecular structure and properties / James Wei.  
p. cm. — (Topics in chemical engineering)  
ISBN-13 978-0-19-515917-2  
ISBN 0-19-515917-9  
1. Chemical engineering. 2. Production engineering. I. Title. II. Topics in  
chemical engineering (Oxford University Press)  
TP155.W378 2005  
660—dc22 2005017098

9 8 7 6 5 4 3 2 1  
Printed in the United States of America  
on acid-free paper

This book is dedicated to my beautiful grandchildren:

Kevin, Cheryl, Cecily, Stacey,

Ryan, Aislyn, Natalie, Vivien

*This page intentionally left blank*

## Preface

Engineers make useful things for people, and chemical engineers do it with chemistry. The two main tasks are what we should make, or product engineering, and how we should make it, or process engineering. Both topics are important if we are to make with skill what people want. The current chemical engineering curriculum concentrates on the manufacturing process, which leads to efficient production, reduced cost, and improved safety. However, new and improved products are needed periodically to rejuvenate the industry, and to help customers lead better lives. There was a time that consumers waited eagerly for the introduction of miraculous new chemical products that transformed their lives, such as celluloid, nylon, penicillin, synthetic rubber, Teflon, and Kevlar. But, for the past decade, information technology has held center stage and captured the attention of the public, with new products such as the personal computer, cellular telephone, word processors, spreadsheets, and the Internet. Ambitious new engineering graduates always seek opportunities to demonstrate their prowess, to innovate products with high growth and profit and to avoid stagnant commodities such as sulfuric acid and common salt. The business landscape is littered with the wreckage of once-successful companies that focused exclusively on cost cutting and neglected the development of new products.

It is sometimes said that chemical engineers wait for the chemists to invent new products, and then are summoned to manufacture them in quantity with economy and safety. Indeed, in many companies, the most ambitious new chemical engineering graduates often gravitate toward the process departments, as their education and curriculum seem more relevant to solving process problems. But neither is there anything in the curriculum of chemists that teaches them how to design products. Thomas Midgley, Jr. invented both tetraethyl lead and Freon, two of the most celebrated products in the 20th century, but he only had a bachelor's degree in mechanical engineering, and he had to learn on his own the chemistry of structure–property relations. This exclusive concentration on processes also cuts off the chemical engineers from exposure to the

desires and fears of society, and hinders their development into fully fledged policy makers and organizational leaders.

Many chemical engineers currently work in industry and academia on solving product engineering problems, and their work would be much more productive if they were given appropriate toolboxes of theories that are generally applicable, powerful new computer hardware and software, and triumphant cases of historical developments to inspire and show the winning ways. In the last few years, a number of universities in the United States and Europe have started courses and degree programs in product engineering. There are many different approaches to the course contents, depending on the resources of these pioneering teachers. An effective method is to introduce the subject by the historic case method. This shows the patterns of past successes, which methods were fruitful, and which paths were barren. A second method is to provide useful tools in the search for materials that possess the desired properties, as well as ways to modify materials to improve their properties. A third method of teaching is by doing a product design project, perhaps in parallel with the traditional process design project, which must take into consideration market needs and safety and environmental impacts.

This book takes the approach that these three methods all teach valuable lessons, and is thus divided into three parts. Part I describes inspiring historic cases of product innovations, with a focus on the creative product engineering work involved. It is largely descriptive, and provides justifications for the analytical tools and synthesis efforts to follow. Part II covers molecular structure–property relations, which provides the analytical tools to search for materials with desired properties, as well as ways to modify materials to improve their properties. The subject of molecular structure–property relations has been tremendously enhanced by powerful new computer methods. Part III challenges the future product engineers to understand the design goals, and to satisfy both market demand and public acceptance. The culmination is a design project for the students to exercise creativity, and to make trade-offs in synthesizing numerous elements together to make a successful product: What is the market demand for a product and what properties should it have? What may be the product composition and manufacturing technology? What are the safety and environmental problems from extraction to final disposal? What may be the financial reward for launching this product?

New products are urgently needed by the public and the chemical processing industries; and perhaps the understanding and skills of molecular structure and properties, combined with modern computer software and the Internet, would help to give a boost to creativity. In the terminology of Thomas Kuhn, a paradigm is a galaxy of concepts and tools that can solve important problems, becomes widely taught in universities, and generates meaningful research challenges. Perhaps product engineering and molecular structure–property relations can follow the advances of unit operations and transport phenomena, and become the Third Paradigm of Chemical Engineering.

Shortly after the publication of the Amundson Report on the future of chemical engineering in 1989, Edward Cussler and I discussed the need for a book on products. Subsequently, he became the pioneer and coauthored the first modern book in product design. Neither of us has ever designed a product that became a winner in the marketplace. It is customary in our profession that those who write design books have never designed a chemical plant or an oil refinery, but those who have done so do not write. I am indebted to many friends who have participated in the design of successful

products and who are willing to share their experiences: Harold Chung told me about the exciting story of guided-missile fuel, Mauricio Futran and San Kiang told me about the history and development of taxol, and Bob Langer told me about controlled drug delivery.

I acknowledge the gracious permissions by the Nobel Foundation, the Chemical Heritage Foundation, the University of Pennsylvania Library, Cambridge University Press, and John Wiley to reproduce the photographs and figures.

*This page intentionally left blank*

# Contents

## Part I

### Product Innovations

#### 1 Product Exploration and Discovery, 3

- 1.1 Freon CFC, 1928, 5
- 1.2 Market-Pull, Search for Technology, 12
- 1.3 Technology-Push, Search for Market, 24

#### 2 Product Development to Business, 34

- 2.1 Nylon: Centralized Development, 35
- 2.2 Penicillin: Distributed Development, 38
- 2.3 Taxol: Distributed Development, 41
- 2.4 Tasks and Skills in Product Innovation, 45

## Part II

### Molecular Structure–Property Relations

#### 3 Search Challenges and Methods, 53

- 3.1 Forward Search from Material to Property, 55
- 3.2 Reverse Search from Properties to Material, 59

- 3.3 Searches from Technology to Markets, 61
- 3.4 Literature and Databases, 62
- 4 Theory and Quantitative Predictions, 72
  - 4.1 Single Molecules, 72
  - 4.2 Statistical Thermodynamics, 105
  - 4.3 Gases, Solids, and Liquids, 124
- 5 Predictions by Correlations, 153
  - 5.1 Populations, Predictors, and Functions, 153
  - 5.2 Regression with a Single Predictor, 161
  - 5.3 Regression with Several Predictors, 174
  - 5.4 Group Contribution Methods, 182
  - 5.5 Mixtures and Composite Materials, 188
- 6 Estimation by Associations and Trends, 199
  - 6.1 Empirical Structure–Property Relations, 200
  - 6.2 Structure Effects and Ambient Influences on Properties, 203
  - 6.3 Risk Assessment, 211
- 7 Random Searches, 229
  - 7.1 Strategies of Random Searches, 229
  - 7.2 Combinatorial Chemistry, 234
- 8 Research Frontiers, 237
  - 8.1 Synthesis and Discovery, 237
  - 8.2 Measurement and Discovery, 238
  - 8.3 Understanding of Molecular Structure–Property Relationships, 239
  - 8.4 Prediction of Molecular Structure–Property Relationships, 240
- Part III**
  - Product Engineering**
- 9 Product Marketing, 245
  - 9.1 The Sellers: The Chemical Processing Industries, 246
  - 9.2 The Buyers, 254
  - 9.3 Marketing Gasoline: A Mature Product, 268
  - 9.4 Marketing Ink for Inkjet Printers, 277

- 10 Safety, Health, and Environment, 284**
  - 10.1 The Responsibilities of the Chemical Processing Industries, 285
  - 10.2 Health and Safety Problems, 290
  - 10.3 Environmental Problems, 296
  - 10.4 Green Engineering, 299
- 11 Product Design, 308**
  - 11.1 Bottled Drinking Water: Fluid Compound, 310
  - 11.2 Missile Fuel: Homogeneous Fluid Solution, 316
  - 11.3 Adsorbents for Separation: Solid Material, 320
  - 11.4 Controlled Drug Release: Composite Solid, 325
  - 11.5 Evaluation, Selection, and Business Plan, 329
- 12 Product Innovation Opportunities, 340**
  - 12.1 Market-Pull Opportunities, 341
  - 12.2 Technology-Push Opportunities, 345
- Appendix, 351**
- Index, 357**

*This page intentionally left blank*

# Part I

## Product Innovations

Great product innovations change the world, improve the way we live, and prepare the foundations to further innovations. How does one distinguish between a good innovation and a great innovation? The commonly accepted criteria of a great innovation include one or more of the following aspects:

1. It addresses and solves an important need of the world and improves lives.
2. It is widely used by many people over a long time.
3. It advances the state of science and technology, leading to other great inventions.

The evolution of an innovative product can be divided into four phases: the initial exploration–discovery–invention phase, creating a product concept; the development–design phase, making plans and blueprints to specific products and manufacturing processes; the commercialization phase, implementing the blueprints to make and to sell the products for the market; and the business and maintenance phase, monitoring outside responses to the new products and making continuous product improvements. The work of innovation is carried out by a relay team, staffed by people with different expertise in different phases of the race. The inventor who started the race may leave the relay team soon, but the race cannot be won without the other team members and their creative contributions and problem solving in all four phases. The press and historians often concentrate on the excitements of the exploration–discovery phase, and neglect the critical roles of the follow up. Out of the thousands of inventions each year, only a handful manage to find resourceful and effective champions to arrange the finance and oversee the transformation of an idea into reality; that is, a widely used product innovation.

In Part I we examine a number of great historical product innovations. Students of product engineering should study the successful paths taken in the past, learn the problems encountered and solved, and examine the effective methods used, as well as the failures that should be avoided in the future. Chapter 1 is concerned principally with the exploration–discovery phase of the innovation. Chapter 2 is concerned with the tasks involved in the development, planning, commercialization, and business phases of the innovation.

*This page intentionally left blank*

# Product Exploration and Discovery

## **1.1 Freon CFC, 1928**

## **1.2 Market-Pull, Search for Technology**

- Modifying current products
  - Search for materials not currently used
  - Creation of new synthetic material
- 1.3 Technology-Push, Search for Market**
- Adapt “platform” technology to new markets
  - Invention of new technologies

## **References**

## **Further Reading**

## **Websites**

## **Discussion Topics**

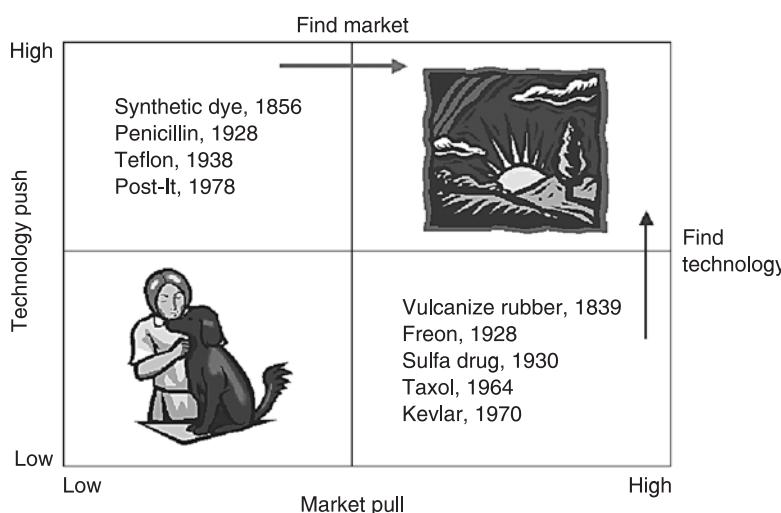
## **Exercises**

It is sometimes said that “necessity is the mother of invention.” Many product innovations have begun with the observation and recognition that many people are in need of a new or improved product, and investigators then looked for a technology that would produce such a product to satisfy this market need. Investigators may examine current products to find what properties need improvement and whether these properties can be modified; for example, raw rubber is brittle when cold and is sticky when hot, whereas vulcanized rubber, which is used to make tires and gaskets, remains flexible whether hot or cold. Investigators may take the more ambitious approach of looking for materials that are not currently used for a particular product to see whether they have better properties to offer: for example, the use of ether as an anesthetic relieved the pain from surgery and childbirth that people were subjected to previously. The more ambitious investigator would attempt to create new synthetic materials to suit a particular market: Freon, a chlorofluorocarbon (CFC), was invented to make a safe refrigerant for home refrigerators. These are called the Market-Pull products, or market looking for a technology.

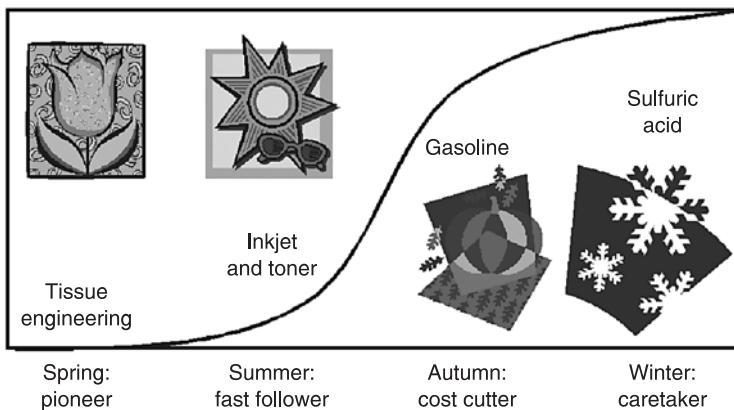
Another frequent innovation path begins with a technology that is dormant or underutilized, followed by the search for new markets. When Freon was established as a safe

refrigerant, it became the platform for new markets, such as air conditioning, aerosol propellants, and computer cleaning. Some technologies began as accidental discoveries when investigators were looking for something else, or were driven by curiosity. Penicillin is one of the best known examples of serendipity, of making unexpected discoveries. The most ambitious paths start from planned explorations to create a new technology, followed by the quest for a place in the market. Carothers created the field of synthetic condensation polymerization, and DuPont decided that this method could be used to make nylon fibers to replace silk stockings. These are called the Technology-Push products, or technology looking for a market. Figure 1.1 shows some historic innovations that began from a high market need with low technical identification, with examples from vulcanized rubber to Kevlar, as well as innovations that began from a high technology capability with low market identification, including synthetic dyes to Post-It notes.

We often give the greatest admiration to pioneering innovations; that is, the bold introduction of a totally new technology or the creation of a new market, one with great potential combined with great risk. This type of innovation can be likened to the spring season. When a new market has already been created by a new product, the task of the fast-follower is to come up with a better, or me-too, product that usually has less potential and less risk; this can be compared to summer. A mature product in an established market can be compared to autumn, where there may be limited opportunities to improve the existing products by a new formulation or a new process to lower costs. A declining product can be compared to winter, where caretakers take rearguard maintenance efforts to prevent obsolescence. Figure 1.2 shows a product life cycle, with tissue engineering as an example of spring, ink for inkjet printers as an example of summer, gasoline as an example of autumn, and sulfuric acid as an example of winter.



**Figure 1.1** Historic innovations that began as Market-Pull and Technology-Push identifications



**Figure 1.2** Product life cycle divided into spring, summer, autumn and winter

### 1.1 Freon CFC, 1928

Thomas Midgley, Jr. invented chlorofluorocarbons (CFCs), or Freon, as a safe refrigerant for the home refrigerator, on assignment from his boss Charles Kettering. He set out with a specific purpose in mind, and he bent all his intellectual powers to that quest (McGrayne and Sharon 2001).

The preservation of food was one of mankind's first problems. When our ancestors came down from the trees in East Africa 2 million years ago, they changed their diet from gathered plants and fruits to hunted small animals. When they managed a kill, they could eat the meat for a few days by hiding the kill from scavengers, like hyenas and vultures, but they could not keep it for long from bacteria and maggots which would turn the meat putrid. Joseph helped the Egyptians to survive 7 years of famine by storing 7 years' harvest of dry grains. The American Indians would preserve an abundant salmon catch in the fall by drying, to tide them over until the spring.

The traditional methods of food preservation included drying (such as for fish and grains), which discourages bacteria growth. There are many other ways to decrease the availability of water in foods, such as by the salting of fish, the sugar-curing of hams, and the acid-pickling of vegetables. Another method is to lower the temperature, such as by packing fish in ice, to decrease the speed of bacteria growth.

The evaporation of liquids, such as water or liquid ammonia, removes the faster molecules and cools down the liquid that remains. This would also make the air more saturated with water, so this process cannot operate for long without bringing in fresh, dry air. The vapor compression refrigerator operates on the principle of capturing and condensing the evaporated vapor, by mechanical compression and cooling, so as to regenerate the liquid for the next round of evaporation. In 1748, William Cullen of Glasgow was reputed to be the inventor of the first vapor compression refrigerator, and the evaporating liquid refrigerant was diethyl ether. As we all know, the inhalation of ether will cause unconsciousness, and eventual death, so that it would be prudent to

locate an ice-making machine in a remote industrial area away from homes, and to have the product ice delivered to homes and offices on horse-drawn carts.

In 1859, Ferdinand Carré of France used ammonia as the refrigerant. This was not much better, as ammonia can leak from the refrigerator and has a bad odor and irritating properties. In 1928, there were some 2.5 million home refrigerators in the United States, and the refrigerants used were: nitrous oxide (boiling point  $-88^{\circ}\text{C}$ ), ammonia ( $-33^{\circ}\text{C}$ ), sulfur dioxide ( $-10^{\circ}\text{C}$ ), methyl amine ( $-6.7^{\circ}\text{C}$ ), butane ( $-0.5^{\circ}\text{C}$ ), ether ( $34.6^{\circ}\text{C}$ ), and chloroform ( $61^{\circ}\text{C}$ ). It is dangerous to have these compounds in the average person's kitchen. A Cleveland hospital had a leaking refrigerator, and more than 100 people died as a consequence. The newspapers were clamoring for a law to stop the "killer refrigerators."

In 1928, Charles Kettering of General Motors asked Thomas Midgley, Jr., to develop a nontoxic and nonflammable refrigerant (Midgley 1937, Wilson 1937). Midgley did a systematic and logical analysis of the property requirements of such a refrigerant, studied the data available in the literature of existing refrigerants, considered molecular structures that had the potential for such properties, synthesized the compound and tested it to determine that it was nonflammable and nontoxic, all in 3 days!

Thomas Midgley, Jr. (figure 1.3), was born 1889 in Pennsylvania. He received a bachelor's degree in mechanical engineering from Cornell in 1911. He appeared to have received no formal education in chemistry or chemical engineering. In 1921, he invented tetraethyl lead (TEL), which was used to boost the octane number of gasoline—a revolutionary invention that had tremendous impact for half a century. One of the consequences of TEL use is the by-product of lead aerosol from the tailpipe of automobiles, which can be a health hazard to breathing and to skin exposure. Besides the safe refrigerant, he had numerous other inventions. Owing to the prestige his inventions brought him, Thomas Midgley became president of the American Chemical Society in 1946. He died by accident in 1944.



**Figure 1.3** Thomas Midgley, Jr.  
1889–1944

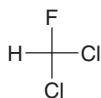
Let us follow Thomas Midgley's journey of discovery of his revolutionary refrigerant. When he received the 1937 Perkin Medal for his development of both antiknock motor fuels and safe refrigerants, he wrote an article "From the Periodic Table to Production" (Midgley 1937) that described how he did it.

I was in the laboratory one morning and called Kettering in Detroit about something of minor importance. After we had finished this discussion, he said: "Midge, I was talking with Lester Keilholtz last night and we came to the conclusion that the refrigeration industry needs a new refrigerant if they ever expect to get anywhere." What was wanted was obvious—a nontoxic, nonflammable refrigerant. I expressed myself as very doubtful that we would be able to find a single substance suited to the task, but that there might be some hope of greatly reducing the existing hazards by using mixtures where nonflammable but toxic materials would be mixed with nontoxic but flammable compounds, to give a mixture substantially nonflammable and considerably less toxic than the refrigerants then commonly employed.

The desired combination of properties was a boiling point between 0 and  $-40^{\circ}\text{C}$ , stability, nontoxicity, and nonflammability. The *International Critical Tables* gave us a partial summary of the volatile organic compounds, and gave the mistaken information that carbon tetrafluoride boiled at  $-15^{\circ}\text{C}$ . This struck us in the face and started us thinking about fluorine. No one could doubt at that time that it was terribly toxic, probably too toxic to use even with isobutane. Perhaps we could add some chlorine compound to it with beneficial results. Henne suggested chlorofluorides as a class to be investigated further. Recognizing that the *International Critical Tables* list was very incomplete, I decided to bring into play the periodic table. Perhaps volatility could be related to it in some way. It takes but a fraction of a second to see that this is true. The elements on the right-hand side are the only ones which make compounds sufficiently volatile for the purpose in hand. Volatile compounds of boron, silicon, phosphorous, arsenic, antimony, bismuth, selenium, tellurium, and iodine are all too unstable and toxic to consider. The inert gases are too low in boiling points. Flammability decreases from left to right, and toxicity (in general) decreases from the heavy elements at the bottom to the lighter elements at the top. These two desiderata focus on fluorine. It was an exciting deduction. Seemingly no one previously had considered it possible that fluorine might be nontoxic in some of its compounds. This possibility had certainly been disregarded by the refrigeration engineers. Everything looked right except that old fear of hydrofluoric acid burns. As it turns out, hydrofluoric acid burns are a special case, and gaseous hydrofluoric acid is less toxic than hydrochloric acid, but we did not know it that afternoon.

Carbon tetrafluoride seemed rather hard to make. And then how could dichlorodifluoromethane boil at  $-20^{\circ}\text{C}$ , and carbon tetrafluoride at  $-15^{\circ}\text{C}$ ? Plotting of boiling points, hunting for data, corrections, slide rules, log paper, eraser dirt, pencil shavings, and all the rest of the paraphernalia that takes the place of tea leaves and crystal spheres in the life of the scientific clairvoyant, were brought into play. We decided that carbon tetrafluoride boiled at about  $-136^{\circ}\text{C}$ , or else it was a very special kind of substance. (Not long after this, a publication on the subject appeared. Carbon tetrafluoride boils at  $-128^{\circ}\text{C}$ , not  $-15^{\circ}\text{C}$ .) Feeling pretty certain at the time that  $-15^{\circ}\text{C}$  was wrong and that it was a sizable research problem to make carbon tetrafluoride, we selected dichloromonofluoromethane as the starting point for experimentation.

Five one-ounce bottles of antimony trifluoride arrived. One was taken at random, and a few grams of dichloromonofluoromethane were prepared. A guinea pig was placed under a bell jar with it and much to the surprise of the physician in charge, did not suddenly gasp and die. In fact it was not even irritated. We then took another bottle and made a few more grams and tried it again. This time the animal did what the physician expected, and died. We repeated again but this time we smelled the material first, and the answer was phosgene. A simple caustic wash was all that was needed to make it perfectly safe. Then we examined the two remaining bottles of antimony trifluoride, and they were not pure. Of five bottles marked "antimony trifluoride," only one had really contained good material. We had chosen that one by accident for our first trial. Had we chosen any one of the other four, the animal would have died as expected by very one else in the world except ourselves. We would have given up on what would then have seemed a "bum hunch."



We will fill in the gaps in his description of the product requirements and design considerations, the search process, and the aftermath. The appropriate range of boiling point temperatures is from 0 °C for a simple refrigerator to –40 °C for a commercial deep freezer. Let us consider the best operating pressure ranges. A unit that operates under tens of atmospheres in pressure would require very powerful pumps, as well as heavy pipes and seals, which can be hazardous and would not be economical for home use. A unit that operates under high vacuum would occupy very large volumes; any leaks in the system would invite moisture and air to enter and contaminate the refrigerant. It is better to let the refrigerant leak out, than to let air and moisture leak in. A satisfactory refrigerant should boil between –40 and 0 °C, and be designed for more than 1 atm pressure.

The refrigerant that leaked should be harmless in an average home. It should not be toxic or obnoxious, such as sulfur dioxide and ammonia are; nor should it be flammable, such as propane and butane are; nor should it be corrosive to metals or remove paint from the furniture. A healthy heat of evaporation would require that it should be operated far below its critical temperature where the distinction between gas and liquid disappears, and there is no heat of evaporation. Actually, there are many other requirements that have not yet been stated. For instance, glycerin would not be a suitable refrigerant even if it satisfied all the stated requirements, as it is very viscous and difficult to pump even at room temperature, and it would be like molasses at freezing temperatures. It is relatively easy to do an acute toxicity test with a massive dose, such as putting a mouse in a bell jar with a beaker of such refrigerant, and see if the mouse dies within 10 s. It is much more difficult to find out whether there are long-term effects of exposure at very low doses, such as 1 ppm in the air for 30 years, which may lead to birth defects at the rate of three additional cases out of a thousand.

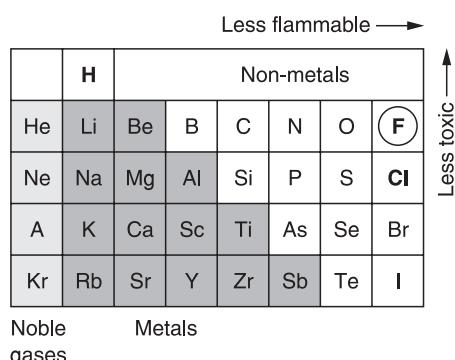
This product should also be economically acceptable. Ammonia and sulfur dioxide are cheap on a dollars per pound basis. We may set a target of \$1 per pound for the refrigerant, but that is only a small fraction of the total cost of home refrigeration.

The major costs involved in a home refrigerator are the capital costs of buying the compressor, the heating and cooling coils, the circulation pipes, and the expansion valves, as well as the yearly operating cost of electricity and maintenance.

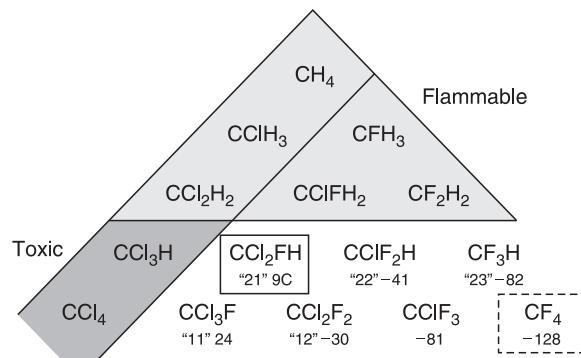
Thomas Midgley executed a marvelously rational and efficient search, which will serve as a model for future product engineers. He did not examine the current products to see whether they could be modified and improved. He did not follow the Thomas Edison system of a massive random search of available material. He asked what material can possibly have this set of properties, regardless of whether they are currently available or not. What are all the known compounds that have boiling points between  $-40$  and  $0$   $^{\circ}\text{C}$ . In his journal, Midgely said that he depended on the *International Critical Tables*, which he found to be very incomplete, as well as inaccurate. It would be much more difficult to find a list of nonflammable and nontoxic materials.

Midgley formulated his plan on the periodic table, from which all things are to be made. Figure 1.4 shows the periodic table and his analysis, which suggested that he should limit his search to the upper right-hand corner of non-metals. He reasoned that the noble gases of helium to radon are certainly very inert, but they boil between  $-268.9$  and  $-61.8$   $^{\circ}\text{C}$ , which is much too low. The metals and their compounds are mostly solids, with the exception of mercury, which boils at  $356.9$   $^{\circ}\text{C}$ . That leaves the nonmetal elements of groups 14 to 17. Midgley noticed that toxicity decreases on moving from the higher atomic weight elements in period 4 to the lower atomic weight elements in period 2: nitrogen is less toxic than phosphorus, which, in turn, is less toxic than arsenic; oxygen is less toxic than sulfur, which, in turn, is less toxic than selenium and tellurium; and chlorine is less toxic than bromine. When he moved from groups 14 to 17, he found decreasing flammability: ammonia is less flammable than methane; and water is less flammable than ammonia. He reasoned that the most desirable element is fluorine, despite the reputation of hydrofluoric acid being a corrosive fluid.

He began by considering carbon tetrafluoride ( $\text{CF}_4$ ). Figure 1.5 shows all 15 compounds that can be derived from methane by replacing hydrogen with chlorine or fluorine. When he excluded the upper part of the triangle as flammable and the left side as toxic, he was left with the lower right-hand side of the triangle. With his set of incomplete and somewhat erroneous information, he reasoned that, if he disregarded the reported boiling point of  $\text{CF}_4$  as wrong, the boiling points of these seven compounds should increase from lower right to upper left. Chlorine has the effect of



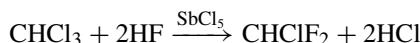
**Figure 1.4** Midgley's analysis of refrigerant candidates, based on Mendeleyev's periodic table



**Figure 1.5** Triangular diagram of CFCs based on methane, showing toxic and flammable regions, as well as their commercial names and their normal boiling points in °C

greatly increasing the boiling point. His team settled on dichloromonofluoromethane,  $\text{CHCl}_2\text{F}$  (which is now called the CFC-21), as the starting point; he synthesized it and found it to be nontoxic, and what he was looking for.

A research lead is only the beginning of product development, as Ketterling and General Motors had to find a suitable manufacturing process, secure supplies, build a plant, and establish marketing organizations. We have a product only when it is available commercially in a useful form, and the public must be informed that it is available in the store in large and reliable quantity, and at affordable prices. The commercial production proceeds by the reaction between chloroform and hydrogen fluoride, bubbling through molten antimony pentachloride as catalyst. This takes place in a reactor at 1 atm and 60 °C for 24 h.



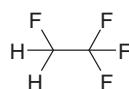
The product is washed with water and sodium hydroxide to remove the hydrochloric acid, and then distilled to separate the different forms of CFC.

Thomas Midgley was also a showman, and when the new product was announced to the world he made a demonstration of his perfect refrigerant at a meeting of the American Chemical Society. He sat on a podium, with a flask of CFC and a lit candle in front of him. He swallowed a mouthful of CFC to show that it was nontoxic and then he blew out the candle to show that it was nonflammable, to thunderous applause. This product created prosperity for many chemists and chemical engineers, who found employment and satisfying careers, handsome profits for investors, and satisfaction for householders, who could enjoy their convenient and safe home refrigerators.

Freon 21 ( $\text{CHCl}_2\text{F}$ ) has a boiling point of 9 °C and is suitable as a refrigerant; Freon 11 ( $\text{CCl}_3\text{F}$ ) is heavier, has a boiling point of 24 °C, and is suitable for air conditioning; and Freon 12 ( $\text{CCl}_2\text{F}_2$ ) has a boiling point of -30°C and is suitable both for refrigerators and deep freezers. General Motor's Frigidaire Division chose Freon 12 when developing the first refrigerators in 1931. CFCs are inert and safe, and became widely used and admired, and hailed as a miracle. They formed the basis of the new industry of air conditioning, by removing heat and moisture from homes, offices,

automobiles, trains, hospitals, restaurants, concert halls, and theaters. This contributed to the great benefit of the comfort of living in warm climates, which is a luxury that even emperors and kings in the past had not possessed. It also made cities such as Houston, Atlanta, Calcutta, and Singapore far more livable. Owing to its inert properties, it was also used to propel aerosols for hairspray, to clean electronic equipment and computers, and to extinguish flames in tight quarters, such as in the Apollo space capsules. In 1988, 2.5 billion pounds of CFCs were consumed. The pattern of use was refrigerants 30%, blowing agents 28%, cleaning agents 19%, and aerosols 19% (Glas 1989).

Midgley did not anticipate that future generations would demand that their refrigerants should have many additional properties, as he had never heard of the ozone layer in the stratosphere, nor of global warming and greenhouse gases. Product engineers face an ever-changing moving target that cannot be totally anticipated at the beginning. Consider the case of ozone, which is produced in the tropical stratosphere by ultraviolet irradiation of oxygen. This stratospheric ozone drifts around the world at an elevation of 20 to 40 km, protecting the Earth from deadly levels of ultraviolet radiation. In 1970, Sherwood Rowland and Mario Molina published a paper showing that, because of their exceptional stability, CFCs accumulate in the atmosphere when released, rather than degrade. When these CFCs rise to the stratosphere they decompose under the action of ultraviolet light to form chlorine radicals, which catalyze the destruction of ozone. This creates a hole in the ozone layer over Antarctica each spring, which may endanger humans and wildlife in Patagonia, New Zealand, and other southern lands. In the same year, Lovelock detected atmospheric concentrations of CFC-11, which rose to 0.6 parts per billion in 1985. The world conference of 1987 in Montreal agreed upon a protocol on the banning of CFCs, and the search was on for a replacement for CFCs that had one additional property, that is of not harming the stratospheric ozone layer. This turned out to be a very difficult task, as Thomas Midgley had done his analysis well. When the additional requirement of no chlorine atoms is added, there remain only two hydrofluorocarbons in figure 1.5, namely  $\text{CF}_3\text{H}$  and  $\text{CFH}_4$ , but these have boiling points that are too low to be useful. The answer is to use an HFC with a higher boiling point, one that is based on ethane instead of methane. The most suitable substitute now is HFC 134a, or  $\text{CFH}_2\text{--CF}_3$ , even though it is considerably more expensive and less energy efficient than the CFC that it replaced. To encourage the public to convert, the tax of \$0.60 per pound on CFC-12 rose to \$7.35 per pound in 1997.



However, the HFC 134a product had an uncertain future shortly after it was introduced. It was anticipated that, after many years of heavy use, some other unforeseen side effects would be discovered and the race would be on again to find yet another substitute. When DuPont first introduced HFC 134a, the company tried in vain to secure an assurance from the regulators that they would be able to recoup their plant investments before the product would once again be declared unacceptable. In fact, HFC 134a is a significant greenhouse gas: 1 mol of this gas has a greater potential to cause global warming than 10,000 mol of carbon dioxide. Will HFC 134a be outlawed

in a very short time, and what material can conceivably replace it in the evaporation-compression refrigerator? Shall we return to butane for the home refrigerator, and risk fires in the home? Will we abandon this concept of evaporation-compression entirely in favor of other methods of cooling? It may be said that “today’s solutions are tomorrow’s problems,” and the work of the product engineers is never done.

## 1.2 Market-Pull, Search for Technology

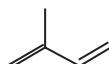
In this mode of discovery, the path starts from the identification of a market that is not well served by the current products available, or the more ambitious path of the identification of a potential market that presently does not exist. The task for the investigator is to discover interesting technology leads, and to design a product that targets this market segment. Perhaps we can modify a current product to make it better, perhaps we can find a neglected material that is not currently used, or perhaps we can create a new synthetic material.

### Modifying Current Products

In a modern corporation with an established research and development department, the normal mode of product development begins with the realization that a current product has some undesirable properties, and modifications are needed to improve these properties. The product engineer is given a specific assignment on which properties to improve, and then searches among the available methods to modify the product. Let us look at a number of historical examples of the modification of existing products.

#### *Vulcanization of Rubber—Goodyear, 1839*

The use of rubber for sports balls and for waterproofing fabric had been noted by Christopher Columbus in his voyages to America, but the results had not been satisfactory (Asimov 1989). Natural rubber is a polymer of the isoprene molecule ( $C_5H_8$ )<sub>n</sub> with molecular weight in the order of 200,000, and consists of thousands of chains with short side chains.



Rubber becomes brittle in cold weather and tacky in hot weather, and it is odorous and perishable. It also has very low tensile strength and low resistance to abrasion. One of the major advances in the improvement of rubber was in the discovery by Charles Macintosh in Scotland in 1820 that coal-tar naphtha is a cheap and effective solvent for rubber. He placed a solution of rubber and naphtha between two fabrics, and in so doing he covered up the sticky or brittle surfaces that had been common in earlier single-texture garments treated with rubber. Macintosh patented the process in 1823. These double-textured waterproof cloaks, which were first introduced to the public in 1824, have been known ever since as mackintoshes.

The American inventor Charles Goodyear spent 5 years doing trial-and-error experiments looking for an additive for rubber, being encouraged by the fact that hides are improved by tanning and that iron is converted to steel by adding a small quantity of carbon (Wolf 1939, Korman 2002). In 1839, he discovered that rubber, when mixed with sulfur and heated, becomes vulcanized, remaining elastic in cold weather and avoiding tackiness in hot weather. The formula was 100 parts rubber plus 5 parts sulfur, followed by heating to 141 °F for 3–4 h. The molecular foundation of this process was not understood until many decades later, when it was found that disulfide bonds formed between strands of rubber chains, turning one-dimensional raw rubber chains into a cross-linked three-dimensional network of chains. However, he was unsuccessful as a businessman, was imprisoned in Paris for debt in 1855, and died disappointed in New York in 1860. After his death, John B. Dunlop of Britain founded the tire industry by patenting and developing pneumatic tires for bicycles and tricycles in 1888. There is nothing better than rubber for tires for transportation: for automobiles, buses, trucks, and airplane landing gear. When the source of natural rubber from Malaysia was cut off in World War II, the United States had to develop a crash program to produce synthetic rubber as a replacement for such uses.

### *Celluloid—Hyatt, 1870*

Ivory taken from the tusks of the elephant is prized for its beauty, durability, and suitability for carving (Morris 1986). Tusks from African elephants average about 6 ft in length, and weigh about 50 lbs, whereas tusks from Asian elephants are somewhat smaller. Ivory was once used for billiard balls and piano keys, as well as many personal articles, such as combs and brush handles. Then, in 1838, cellulose nitrate was discovered; this is highly inflammable and heat sensitive material that cannot be molded at high temperature into products. In England in 1845, Parkes invented pyroxylin as a product of cellulose dinitrate with approximately 11% nitrogen by weight. He tried various solvents and found camphor, incorporated at 2–20%, to be an effective plasticizer that enabled processing at 65–80 °C. He exhibited a range of his products in 1862 at the International Exhibition, including combs and billiard balls. He established a company in 1866, but the company failed in 1868. The material was flammable. He was obsessed with keeping the price low; hence, he used only cheap material.

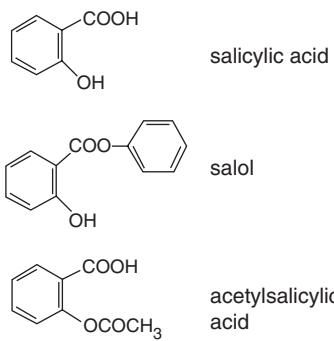
The American John Hyatt was attracted by a prize offer to develop an alternative to ivory for billiard balls. He combined nitrocellulose, camphor, and alcohol, heated the mixture under pressure to make it pliable for molding, and allowed it to harden under normal atmospheric pressure. Celluloid is a tough material, with great tensile strength, and is resistant to water, oil, and dilute acid. It can be made at low cost in a variety of colors. Celluloid was used to stiffen collars, for laminated safety glasses, spectacle frames, piano keys, and photographic films (Billmeyer 1971). In 1870, he and his brother set up a company that became very successful. In 1887, Hannibal Goodwin first used celluloid roll film as a base for photographic emulsion. In 1888, George Eastman began mass production of celluloid roll films for still photography in Rochester, N.Y. It paved the foundation for the invention of motion pictures by Thomas Edison in the same year.

*Aspirin—Hoffman, 1898*

Aspirin is one of the most important drugs used as an antipyretic (lower fever) and as an analgesic (relief pain) (Korolkovas 1988). An old method to treat typhoid patients is to lower their fevers by immersion in cold baths. Then, Carl Buss administered salicylic acid obtained from willow bark to typhoid patients, and published his findings in 1875. This compound is strongly acidic, which makes it disagreeable and irritating, and leads to vomiting. Nevertheless, it became a very important drug, and methods were sought to improve its properties. In 1883, von Nencki reacted phenol with salicylic acid to produce salol, which is insoluble and passes unconverted through the stomach. In the alkaline environment of the small intestine, it hydrolyzes back to phenol and salicylic acid, and became a popular substitute for salicylic acid as an antipyretic, despite the unpleasant by-product of phenol in the intestine.

The chemist Felix Hoffman of Bayer laboratories was asked by his father to seek a more acceptable drug. He searched through the literature on salicylic acid derivatives and came across acetylsalicylic acid, which was synthesized earlier in 1853. The compound was tested by his colleagues successfully on animals, and in 1898 Bayer sent samples to a number of hospitals for trial, and received enthusiastic responses. Acetylsalicylic acid is not very soluble in water, and also passes through the mouth and stomach without change, and is hydrolyzed in the intestine into salicylic acid and acetic acid, which is less damaging than phenol. As Bayer was unable to patent this 45-year-old compound, it patented the processes used in large-scale manufacturing instead, and bestowed the proprietary name of “aspirin” on the compound.

The formulas of salicylic acid, salol, and acetylsalicylic acid are given below to show the structural modifications that turned an unsatisfactory product into a much better product.

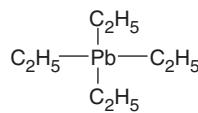
*Tetraethyl Lead as Gasoline Additive—Midgley, 1921*

We have already encountered Thomas Midgley, Jr., who came from an inventive family. His father was an inventor in woven wires, bicycles, and detachable tires. His mother's father invented the inserted-tooth saw (McGrayne 2001). Midgley graduated from mechanical engineering at Cornell University in 1911. He did the minimum amount of work with which he could get by in most of his courses, so that he could concentrate

on a few things that were really of interest to him. He joined the Dayton (Ohio) Engineering Laboratory Company in 1916, which was acquired by General Motors, and worked with Charles Kettering.

The challenge in engine research is in obtaining the highest power output per engine volume and the highest thermal efficiency per gallon of fuel. In an assignment to get more power out of a Delco engine, he found that the compression ratio of a gasoline-fueled engine can be raised to give a fairly good power output and efficiency; however, severe knocking and even cracked cylinder heads resulted if kerosene is used as fuel (Midgley 1937, Wilson 1937). Midgley showed that the knock was due to a rapid rise in pressure before ignition, when the cylinder is near top dead center. He erroneously reasoned that the defect of kerosene is due to its lower volatility, and sought to dye the kerosene so that the droplets would absorb radiant heat and vaporize sooner. It was suggested later that he would have known that this theory was untenable if he had done some calculations, but he felt that it would be easier to do the experiment than the calculations. He went to the stockroom in search of oil-soluble dyes and found some iodine. He discovered that, indeed, the knocking was eliminated in a moderately high compression engine when a substantial quantity of iodine is dissolved in kerosene. Then he tried a dozen different dyes without success; on the other hand, when he added colorless ethyl iodide to the gasoline, he found that it worked very well indeed. Thus, it is the chemical properties of iodine rather than its color which stopped knocking.

For the next few years he studied every branch of chemistry to help him to explain his observations, and to make new compounds for trial as antiknock agents. He also discovered that the various pure hydrocarbons in gasoline differed a great deal in their knocking characteristics. He tried many additive agents, including ethyl iodide, aniline, selenium oxychloride, diethyl telluride, and hundreds of others. Diethyl telluride was the best antiknock agent found up to the time of the discovery of TEL. It had a very decided disadvantage of a strong garlic-like odor, which penetrated every pore of the human body and affected everything and everyone that came in contact. He became ostracized by friends and family for being extremely smelly.



Three years after he began his search, in 1921, he consulted the periodic table of elements and found that all the known good antiknock compounds contain heavy metals from the lower right-hand side of the table. He decided to try lead, as he knew that TEL was soluble in oil. When a small amount of TEL was added to gasoline, the wild knocking stopped. But the combustion product of TEL is lead oxide, which deposits on spark plugs, exhaust valves, and suchlike parts. He eventually found a solution to this problem by including ethyl bromide with the TEL in the gasoline. This reacts with lead to form lead bromide, a compound with a high vapor pressure. This invention brought him fame and fortune, and it also gave the world access to high-compression gasoline engines to produce more power per volume and higher efficiency. The heavily leaded gasoline gave American and British fighters in World War II the ability to climb faster, and they became more successful than German fighters in dogfights. This new

application also led to a serious shortage of bromine, which was solved by recovery from sea water.

Lead aerosol in the air is poisonous to breathe, especially for young children. Many people called for the abolition of lead in gasoline. In the 1970s, the photochemical smog in California was attributed to unburned hydrocarbons and carbon monoxide from automobile tailpipes, and the best solution was the catalytic converter which works with finely divided platinum particles deposited on alumina monoliths. When leaded gasoline is used, these platinum atoms would be quickly covered by a barrage of lead aerosols. This finally led to the abolishment of TEL as a gasoline additive.

### Search for Materials Not Currently Used

Instead of finding ways to modify an existing product to increase the satisfaction of current customers, a more ambitious approach is to look for materials not currently used, but which have many superior properties.

#### *Ether as Anesthetic—Morton, 1846*

The pain of childbirth and surgery had been with mankind from the beginning, and was considered a biblical curse for disobedience to God (Fenster 2001). Nevertheless, various methods have been devised for relief. Homer wrote of “nepenthe,” which was probably a narcotic such as cannabis or opium, and Arabian physicians used opium and henbane. British naval surgeons used rum for sailors before emergency amputations on board ships after a battle. In 1799, Humphrey Davy inhaled nitrous oxide and discovered its anesthetic properties, but the implications of his findings for surgery were ignored. By the early 1840s, parties had become fashionable in Britain and the United States at which nitrous oxide contained in bladders were passed around and inhaled for its soporific effect. It was soon found that ether, which could be carried much more conveniently in small bottles, was equally potent. In the United States, several young dentists and doctors experimented independently with the use of nitrous oxide or ether to dull the pain of tooth extractions and other minor operations.

The first actual successful surgical use of an anesthetic was by the surgeon Crawford Long in Georgia in 1842, who was asked by friends to obtain nitrous oxide for inhalation at a party they were planning. Next morning, Long and his friends found several bruises, but none of them could recollect having felt any pain during the party. He realized that this may be the answer to painless surgery, and he placed a towel soaked in ether on the face of a patient. As soon as his patient was insensible, Long proceeded to excise a large growth on his neck, but the youth felt no pain. Since his findings were not published until 1849, he did not receive the recognition that he deserved. This reminds one that the father and uncle of Marco Polo went to China before him, but they did not publish either.

Historians credit the first public demonstration of true surgical anesthesia to William Morton, an American dentist (Fenster 2001). In 1844, Horace Wells, who was a dentist in Hartford, Connecticut, attended a public lecture demonstration of the effects of laughing gas. One member of the audience had volunteered to inhale the nitrous oxide and then accidentally gashed his leg but felt no pain. Wells asked William Morton to

arrange a demonstration of anesthesia in the presence of America's leading surgeons in the Massachusetts General Hospital, but the demonstration was a failure because insufficient anesthetic was administered and the effort was ridiculed. Morton began his own experiments to find a stronger method, and experimented on his household pets in 1846. He visited a chemist friend Charles Jackson, who suggested ether, which was tested on animals, then on a patient in his office. The first successful public demonstration of surgical anesthesia was held before a large gathering in the domed amphitheater of the Massachusetts General Hospital on 16 October, 1846. He administered ether to a patient, and the surgeon John Warren removed a tumor on the lower jaw. Since ether was not a new material, he could not get a patent on the invention, and both Jackson and Wells challenged Morton's claims. A few weeks after Morton's demonstration, ether was used during a leg amputation at University College Hospital in London. Queen Victoria gave official royal sanction to anesthetics, as she accepted chloroform from her physician when giving birth to her eighth child in 1853.

Early anesthetics had unpleasant side effects, often causing vomiting on recovery. The patients must be given sufficient doses to produce unconsciousness and full muscle relaxation, but not so much that they paralyzed the breathing center of the brain. This is called the safety margin, and needed to be made larger to give the physicians more confidence to handle the unpredictable responses of different individuals, and to scaling for different ages and weights. The formulas of nitrous oxide ( $N_2O$ ), ethyl ether ( $C_2H_5-O-C_2H_5$ ), and chloroform ( $CHCl_3$ ) do not share much structural similarity; this observation does not help researchers to search for other products with similar or better properties. The early administration used simple devices consisting of glass or metal containers, and sponges soaked in ether or chloroform, which allowed no control of dosage. Even today, anesthesiologists are subject to more malpractice lawsuits than most physicians.

### *Incandescent Lamp—Thomas Edison, 1879*

Sir Humphrey Davy discovered in 1811 that an electrical arc can be passed between two poles separated by an air gap to produce light (Asimov 1989). In 1841, experimental arc lights were installed as public lighting along the Place de la Concorde in Paris. Electric arc lights were not practical, as the electrodes burned out too quickly. James Prescott Joule theorized that electrical current through a resistant conductor would glow white hot, and turn heat energy into luminous energy. However, the presence of oxygen caused the filament to burn. Swan was the first person to construct an electric light bulb to exclude oxygen, but he had trouble maintaining a vacuum in his bulb.

In the period from 1878 to 1880, Edison and his associates worked on at least 3,000 different ideas to develop an efficient incandescent lamp (Israel 1998). He housed a filament in a glass vacuum bulb. By January 1879, at his laboratory in Menlo Park, New Jersey, Edison built his first high-resistance incandescent electric light through a thin platinum filament. The lamp burned for only a few short hours. He tested thousands of other materials to use for the filament. He even thought about using tungsten, but the tools available at that time were inadequate for the purpose. He turned to carbonizing materials from every plant imaginable, including bay wood, boxwood, hickory, cedar, flax, and bamboo. He even contacted biologists, who sent him plant fibers from places

in the tropics. He said, “Before I got through, I tested no fewer than 6,000 vegetable growths, and ransacked the world for the most suitable filament material.” Edison decided to try a carbonized cotton thread filament, which burned out 15 h later. Thomas Edison solved this problem on 21 October 1879 with a carbon filament, housed in a glass vacuum light bulb that glowed continuously for 40 h. By 1800, he had produced a 16 W bulb that could last 1500 h, and he began to market his new invention.

### *Chlorination of Drinking Water, Chicago, 1908*

Barbara Tuchman (in 1978) wrote in *A Distant Mirror: the Calamitous 14th Century* that “Of all the characteristics in which the medieval age differs from the modern, none is so striking as the comparative absence of interest in children.... Owing to the high infant mortality of the time, estimated at one or two in three, the investment of love in a young child may have been so unrewarding that it was suppressed.” Adam Smith (in 1776) wrote in *The Wealth of Nations* that “in some places one-half of the children born died before they are four years of age; in many other places before they are seven: and in almost all places before they are nine or ten. ....”

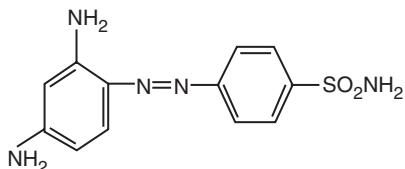
The germ theory of disease was developed by Robert Koch in 1876 and Louis Pasteur in 1880; and the greatest cause of death, and especially of infant mortality, was infectious diseases from contaminated food and from water, which contains bacteria, viruses, amoebae, and parasitic worms. Death by water was not limited to the poor and ignorant, either: Prince Albert, the husband of Queen Victoria, died in 1861 of typhoid through drinking Windsor Castle water; and Pyotr Ilyich Tchaikovsky died in 1893 from drinking a glass of unboiled water during a cholera epidemic in St Petersburg. Even today the drinking water in third-world countries is still deadly, as clean municipal drinking water is not available: in the 1990s there were cholera epidemics in Ethiopia, Sudan, and Peru. And after a man-made disaster such as war, or after a natural disaster such as earthquake or hurricane, the first relief effort is to restore electric power and clean water supply. The best methods of killing bacteria are high temperature, acidity, ozone, and sunlight; filtration also proved effective at reducing incidences of cholera in Hamburg as early as 1892.

Disinfection of drinking water with chlorine was under way in 1908 in Chicago and Jersey City, and by 1910 it was being used in several major U.S. and Canadian cities. Abel Wolman of Johns Hopkins University and the Maryland Department of Health were given credit for perfecting a formula for purifying water with chlorine in 1915 (National Academy of Engineering 2001). The resulting epidemiological transition led to an initial fall in death rate without a change in the birth rate, which led to great population increases. Fortunately, in the next phase, when families realized that it was no longer necessary to bear three children in order to have one surviving child, the birth rate decreased.

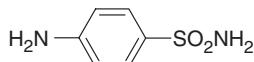
### *Sulfa Drugs—Gerhard Domagk, 1930*

Gerhard Domagk was a pathologist and bacteriologist inspired by the ideas of Paul Ehrlich, and he began testing numerous newly developed dyes for their possible effects against various infections (Korolkovas 1988). He was encouraged by the fact that some

dyes preferentially adhere to bacteria and not to mammalian cells, which suggested that if they were modified to be toxic they would selectively kill bacteria but would not harm mammalian tissue. He joined the firm I. G. Farben in 1927, and had a very well equipped research laboratory. He screened dyes on mice inoculated with a highly virulent strain of streptococcus, which causes tonsillitis, scarlet fever, permanent deafness, and fatal meningitis. The virulence of the strain that he used can be gauged by the fact that 100% of the mice died within 4 days of inoculation. He tested gold compounds, acridines, and azo dyes. His first success was with gold compounds, but they often led to kidney damage, which prevented their administration in the sufficiently high dosages required to cure streptococcal infections in patients. In 1932, he tested the azo dye protosil synthesized by his chemist colleagues, which is a brilliant red dye, and found that it cured mice with streptococcus.



In 1935, his daughter Hildegarde developed a severe septicemia after pricking her finger with a needle, and her life was saved by protosil rubrum. It turned out that the azo link in this red dye was cleaved in the patient's body to form colorless 4-aminobenzene sulfonamide, which is really the active principle. This led to the synthesis of more than 1000 sulfa compounds as derivatives that may have had even better properties, by substitution of one of the hydrogen atoms on the  $-\text{SO}_2\text{NH}_2$  group, including sulfathiazole and sulfapyridine, and led to the beginning of the sulfa drugs.



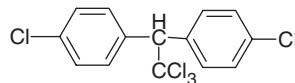
This sulfonamide was previously synthesized back in 1908, but the chemist and his colleagues who synthesized it did not discover its wonderful antibacterial properties. As a consequence, Domagk was unable to secure a patent for his invention, but he was awarded the Nobel Prize for Medicine in 1939 for discovering the therapeutic values of the compound. He was unable to receive it in person because of the Nazi policy of the German government, and finally received it after the war in 1947.

### *Dichlorodiphenyltrichloroethane—Müller, 1939*

Paul Hermann Müller received a degree in chemistry and worked for the J. R. Geigy Corporation, which later became part of Novartis (McGrayne 2001). Geigy specialized in dyestuffs for woolens. Chemists at the company discovered a chlorinated hydrocarbon compound that protected woolens from clothes moths, but it was a stomach poison. Geigy then searched for other insecticides that killed other pests. Natural insecticides made from plants include pyrethrum from chrysanthemum, rotenone from a tropical

vine, and nicotine from tobacco; they are expensive and not persistent, and easily destroyed by heat and light. Another insecticide was lead arsenate, which is dangerous to humans. The German company I. G. Farben patented a family of chemicals that includes organophosphate insecticides.

When Müller became interested in insecticides, he decided to concentrate on a contact poison, which does not depend on how the insect eats and has the potential to become a universal insecticide. He decided to work on chlorinated hydrocarbons, and screened their effectiveness on houseflies in a cubic-meter glass chamber. His early experiments showed that compounds with the group  $\text{CH}_2\text{Cl}$  had some insecticide effects. He found a 1934 paper on diphenyltrichloroethane, which was found to be poisonous to the flies in his glass box. By 1939 he had screened 349 compounds, and his 350th compound was dichlorodiphenyltrichloroethane (DDT). Spraying DDT on the flies in his glass cage, Müller saw them fall helplessly onto their backs in 10 min. His cage remained poisonous for weeks and killed any fly that touched its walls. DDT works slowly but has a sure kill with tremendous persistence. It is a cheap contact poison without objectionable odors, it is stable in air and light, it is effective in extremely small doses, and it dissolves poorly in water, so that warm-blooded organisms absorb only traces of it.



DDT enters an insect by dissolving the thin layer of fatty substances that repel water from the waxy outer skin and then paralyzes vital nerve centers. Müller was disappointed to learn that he was not the first discoverer of DDT: 65 years earlier, a graduate student Othmer Zeidler had synthesized the compound as part of his chemistry doctoral thesis. Zeidler described many of DDT's properties and developed the method used to make it commercially, but he did not discover its insecticide powers. When Geigy took out the basic Swiss patent in 1940, it was not for the composition matter, but for its use as an insecticide. This began the era of synthetic chemical pesticides.

DDT was tested for its ability to kill lice, which carried the typhus bacteria that had decimated Napoleon's Grand Army, and killed 3 million people after World War I. In 1941, Geigy reported its discovery to the governments of the United States, Britain, and Germany. DDT was used heavily in the Pacific to kill mosquitoes, which are carriers of malaria. General Douglas MacArthur said that unless malaria could be controlled, he would need three divisions in order to have one combat-ready division, as one division of men would be hospitalized with malaria, and another division would be recuperating from it. Along with penicillin and atomic energy, DDT emerged from World War II as one of the wonders of modern science. In 1948, Müller received the Nobel Prize for Medicine.

Environmental alarm about DDT began in 1945, and Rachel Carson published *Silent Spring* in 1962. DDT accumulates in the fat deposit of animals, and becomes concentrated in the top of the food chain, such as in eagles. DDT was banned by the U.S. Environmental Protection Agency (EPA) shortly after it was established in 1971.

### *Taxol—Wall, 1966*

Taxol was the most triumphant new anticancer drug developed in the last few decades. It resulted from a drive to screen thousands of synthesized compounds and plant extracts, financed by the National Cancer Institute (Georg et al. 1994, Suffness 1995, Goodman and Walsh 2001). Taxol is extracted from the bark of the Pacific yew tree, or *Taxus brevifoli*, found in the Pacific Northwest, and has been proven effective against intractable breast and ovarian cancers. It is now produced by semi-synthetic methods from the needles of the common ornamental tree English Yew, or *Taxus baccata*. We will discuss taxol in greater detail in Chapter 2.

### **Creation of New Synthetic Material**

Midgley modified gasoline with TEL, but he took an even more courageous route when create CFCs, which were new materials that were not based on existing refrigerants. The synthesis of a new class of materials with unusual properties is the most challenging of the assignments, and holds the promise of opening new doors to worlds that have never been visited and conquered before. There are many other illustrious historic examples.

### *Salvarsan—Ehrlich, 1909*

Syphilis was reputed to be brought to Europe by the crew of Christopher Columbus, and was devastating for many centuries (Korolkovas 1988). As early as 1786, Thomas Fowler investigated the use of arsenic compounds in the treatment of malaria, sleeping sickness, and syphilis. Numerous preparations were tried with compounds of mercury and arsenic, and the difficult requirement was for a safety margin, between a dose that was high enough to be effective and yet low enough to be safe.

Since the discovery of synthetic dye by Perkin and the establishment of the synthetic organic chemistry industry, a wide variety of synthetic dyes provided a set of choices for staining microbes so that they could be readily examined under the microscope. It is reasonable to suppose that if a dye binds selectively to a microbe much more than to mammalian tissues, then it may be possible to find a dye that selectively harms the microbe and spares the mammal. Paul Ehrlich (figure 1.6) was a medical student who did research in the distribution of foreign substances in the body, and he was particularly interested in the influence of chemical structures of different types of molecule in live animals. He observed that acidic dyes with the sulfonic acid function, used by dye manufacturers to enhance water solubility, were unable to penetrate into the brain or fat tissues.

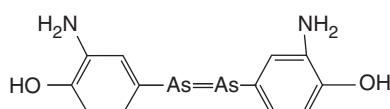
He was made director of a new government laboratory in 1899 in Frankfurt, where he began to experiment in the synthesis of new substances not necessarily found in nature that could kill parasites or inhibit their growth without damaging the mammal. In 1902, Laveran and Mesnil of the Pasteur Institute found a method to infect mice with trypanosomes, and studied the effects of Fowler's Solution on them. They discovered that subcutaneous injection of the sodium salt of arsenous acid can cause the rapid disappearance of the parasite trypanosomes from the blood of mice and rats, but the parasites reappeared within a few days to cause the death of the animals. Ehrlich began



**Figure 1.6** Paul Ehrlich, discoverer of Salvarsan, 1854–1915

his search for an effective arsenic preparation, and had a cooperative arrangement with the Cassella Dye Works (later became part of Hoechst), which supplied him with various dyes that he requested. He turned his attention to the protozoa spirochete *Treponema pallidum*, discovered in 1905 to be the causal organism of syphilis.

Ehrlich believed that a molecule can be constructed to be a specific agent against a parasite without major injury to the host. He observed that methylene blue can selectively stain certain parasites in preference to host tissues. By 1907, Ehrlich had synthesized and tested more than 600 arsenic compounds, and he took out a patent on preparation 606 but stopped further investigations. In 1908, Ehrlich received the Nobel Prize for Medicine for his other discoveries made earlier, but his best work was yet to come. The Japanese scientist Sahachiro Hata joined his group as his assistant, retested the whole series of arsenic compounds, and found 606 to be very effective. They persuaded the firm Farbwerke-Hoechst to produce this drug, which was given the official name of arsphenamine. By 1910, they had treated 10,000 cases of syphilis, and they named the drug Salvarsan, the “magic bullet.” It was the most effective cure for syphilis until the development of penicillin.



Paul Ehrlich is regarded as the father of chemotherapy, as he had made the first totally synthetic drug that was effective and ushered in the modern pharmaceutical industry.

*Corfam—DuPont, 1949*

DuPont had been interested for a long time in developing a substitute for shoe leather (Hounshell and Smith 1988). Leather has many weaknesses, such as nonuniformity from irregularly shaped hides, becoming stiff after wetting and drying, and requiring high maintenance through frequent polishing. At that time, the search for artificial leather was encouraged by an erroneous forecast of a coming shortage of leather. The past attempts to make artificial leather were centered on coating fabrics with nitrocellulose or polyvinyl chloride, which created nonporous patent leather materials. In 1949, a new approach was discovered, based on coating porous nonwoven fabrics. Between 1950 and 1955, three separate departments of Textile Fibers, Films, and Fabrics and Finishes put 25 researchers on the development. Fabrics and Finishes became the lead developer in 1955, and worked out a composition based on a nonwoven Dacron substrate with a porous polyurethane binder and a porous outer coating. By 1958, process development began, as well as serious market analysis.

The market research analysis concluded that a satisfactory product, which could be sold at leather-equivalent prices to the shoemakers, would find an adequate market. In 1962, 15,000 pairs of shoes were distributed for testing, and the results showed that 8% of the Corfam wearers complained of some discomfort, compared with 3% and 24% for leather and vinyl, respectively. Since an overly complicated mechanical process to manufacture Corfam was developed, its price had to be set equivalent to that of finer grades of shoe leather, which led to the decision to market Corfam as a premium product for top-quality shoes. DuPont felt that Corfam in some ways was better than leather, because Corfam was unaffected by moisture, weighed a third less, kept its luster, and did not have to be broken in. In the fall of 1963, 30 high-quality shoe manufacturers that had contracts with DuPont exhibited Corfam shoes at the National Shoe Fair.

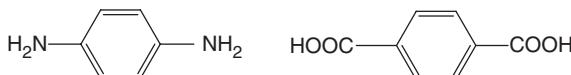
Over the next 5 years, shoe retailers sold 75 million pairs of Corfam shoes, but losses piled up until they reached \$70 million. Corfam encountered more retailer and consumer resistance than expected, and the main problem encountered was foot discomfort. Since Corfam shoes did not stretch, and it was essential to have a proper fit, the solution was to buy a larger size shoe than usual, something that fashion-conscious ladies refused to do. At the same time, European shoes made from many different styles of leather became very popular, and DuPont was unable to change the product to keep up with the latest fashions. The imported shoes pushed Corfam out of the market of expensive shoes into cheaper shoes, where it had to compete with vinyl, which is not as durable. While durability may be important in utilitarian shoes such as army, police, and fieldwork, it is not important in fashion shoes. Owing to a series of miscalculations on matching technology capabilities to market needs, Corfam walked into a trap that it was unable to escape from. Finally, DuPont gave up on Corfam and sold the plant and process to Poland in 1971. Corfam became known as “DuPont’s Edsel,” and led to a reorganization of DuPont’s new product strategy.

*Kevlar—DuPont, 1970*

DuPont in the early 1960s had the goal of developing a fiber with the heat-resistance of asbestos and the stiffness of glass (Hounshell and Smith 1988, Tanner et al. 1988).

By 1965, Stephanie Kwolek discovered that *para*-aminobenzoic acid could be polymerized into rigid rods, and solubilized to yield a spinnable polymer. In a dilute solution, flexible molecules like nylon have a low level of entanglement. At higher concentrations, the random coils become highly entangled, so that spinning and drawing would lead to only partially extended chains. The more rigid polymers like *para*-aramid cannot form random coils, because their movement is restricted. At a critical concentration, they are forced to align parallel to each other in a randomly oriented liquid crystalline domain. Under shear, as in a spinneret capillary, the liquid crystalline domains orient in the direction of the deformation. This leads to fully extended chains in fibers and the observed high strength and high modulus. It can be compared to a log jam in a river, where the packed logs become aligned as they are rafted and flow down the river.

The best candidate for a strong fiber was the polymer made from *p*-phenylenediamine and terephthalic acid, and it eventually became the basis for “Kevlar.” It was very difficult to spin Kevlar into fibers, until the discovery that it forms a crystalline complex with sulfuric acid at a ratio of 1 mol Kevlar to 5 mol sulfuric acid. This enabled fiber spinning at high polymer concentrations. After spinning, it is necessary to get rid of the sulfuric acid by reaction with lime, to produce 7 lbs of gypsum per pound of fiber. By 1972, they completed a 1 million pound market development plant, and by 1982 they reached full commercialization with a 45 million pound plant.



Some of the best uses of Kevlar are in aircraft composites and flak jackets for military and police uses, which require high tensile strength and high modulus per unit of weight. It also has the property of damage tolerance, as it has a ductile compressive failure mode, which is very different from the brittle failure of carbon fibers.

### 1.3 Technology-Push, Search for Market

In this mode of innovation, investigators have a technology that they believe has potential that is underutilized, and are looking for the first application or for more applications. A “platform technology” is a technology successful in one or more markets, and has potential applications in many more markets. Sometimes, an existing technology is dormant and has never been used. The most exciting stories concern the discovery of new technologies, either as an accidental discovery, or as a conscious plan.

#### Adapt “Platform” Technology to New Markets

Now that we have a technology to make a family of interesting materials, after this material has been used successfully for one product, how do we find more applications in other fields for this material?

### *Refrigerant to Air Conditioning and Other Uses for Freon, 1930*

Freon was invented by Thomas Midgley, primarily in response to the need for a safe refrigerant in 1928. After it proved to be nontoxic and nonflammable, it became known as the miracle chemical that is safe for use in the average home. This inert property made it extremely popular, and it was adopted for numerous other uses. Air conditioning is obviously a closely related application, but the temperature relevant for room comfort is closer to 75 °F than to the 40 °F for refrigerators and the 0 °F for freezers. In fact, drying of the humid air may bring as much comfort as cooling, as cooler air cannot carry as much moisture as warmer air. Therefore, a higher boiling-point CFC would be preferred, with more chlorine atoms per carbon to increase the boiling point. The first window air conditioner unit was marketed by Philco in 1938, and the first air-conditioned automobile was engineered by Packard in 1939 (National Academy of Engineering 2001).

There are many Freon applications that ignored its property of evaporative cooling, depend only on its ability to evaporate and propel aerosols, for cosmetics and personal hygiene, without creating a noxious gas in its wake. It will also generate small bubbles that are suitable for making polymer foams, such as seat cushions, mattresses, automobile seats, and insulations. Finally, its inertness alone is sufficient for it to be used to clean a computer by total immersion and shaking, to remove accumulated dust and adhering particles. It is also used for fire extinguishing, especially in confined spaces such as in space capsules.

### *Sandpaper to Mending and Recording Tapes*

The 3M company started out as the Minnesota Mining and Manufacturing company, and had a much admired technology of gluing sand grains on to kraft paper to make sandpapers (Nayak and Ketteringham 1986). This technology of adhering solid particles securely to surfaces became the platform technology for numerous other products, and gave 3M the reputation as a most innovative company. One of the early products was Scotch tape, which was used to mend books in homes and offices, as well as to hold broken toys together. This technology was also used to glue iron oxide on to Mylar films, which became the magnetic tape used for sound and movie recording, as well as being the workhorse of memory storage in the electronic age. The various forms of computer magnetic disks are also based on this technology, whether they are called floppy disks, hard disks, Zips or Jazzes.

### *Botox to Remove Skin Wrinkles*

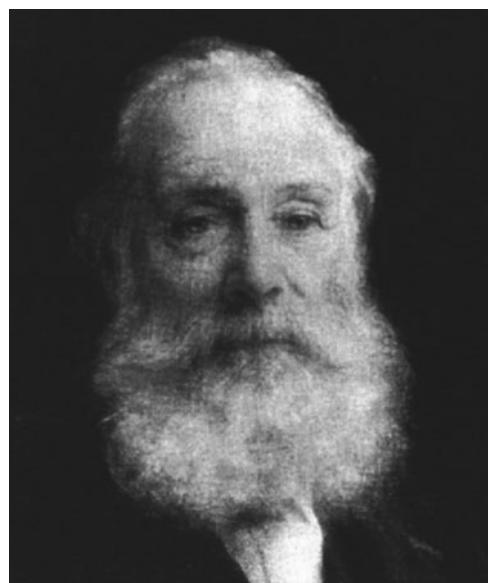
Botulin, which is produced by the bacterium *Clostridium botulinum*, is among the most powerful poisons and can cause death in a very small dose of 0.0003 µg/kg. It acts by paralyzing the muscles, so that breathing stops. In recent years, it was discovered that it can also be used to paralyze the muscles that cause skin wrinkles, and so has found use in the facial cosmetics industry for improving the appearances of aged people.

### Invention of New Technologies

The starting points of many of the greatest inventions were based on the discovery of a new technology, whether planned or accidental. An observant investigator may make an accidental observation that a natural or a man-made material has unusual properties that could have novel applications, searched and found an application for that material, and introduced that material into a new market. A diligent investigator may work very hard to invent a new technology. The discovery of condensation polymerization by Carothers was planned, but the discovery of penicillin by Fleming was accidental.

#### *Synthetic Dye—Perkin, 1856*

In 1853, William Perkin was 15 years old and entered the Royal College of Chemistry in London to study under August von Hofmann as his laboratory assistant (Garfield 2000). He undertook the synthesis of quinine for the treatment of malaria, which had been a leading cause of death in many continents, particularly in tropical countries. Over the Easter vacation in 1856, he worked at his home laboratory to oxidize aniline from coal-tar with potassium dichromate, and he produced instead a black precipitate. Treating it with ethyl alcohol, he dipped a piece of silk in it, and found a wonderful purple color. This was the beginning of the modern industrial dye industry, and his discovery was named aniline purple, or mauve. At that time, all the dyes had natural origins, from plants, animals, and minerals. The ancient dye Tyrian purple was obtained from the head sack of murex mollusks. It was used to dye the robes of Roman emperors as a symbol of authority, and the Roman senators were allowed to wear only one stripe of purple. In 1856, Perkin (figure 1.7) obtained a patent for manufacturing the dye,



**Figure 1.7** William Perkin, discoverer and manufacturer of synthetic mauve, 1838–1907

and the next year, with the aid of his father and his brother Thomas, he set up an aniline manufacturing plant.

The conservative Scottish textile manufacturers were slow to adapt to this new dye, and the leadership of innovation went to French textile manufacturers by default. When Empress Eugenie of France wore mauve in public, it impressed Queen Victoria to wear mauve at her daughter's wedding, so there was a wave of enthusiasm for synthetic dye. Perkin became very successful in commerce, and gave up business for chemistry research later in life. He was knighted in 1906, at the 50-year anniversary of his discovery, and died in 1907. Mauve had a short commercial life, as its early success led to further experimentations that produced many better dyes, particularly in Germany. This marked the beginning of the synthetic dye industry, which was the first modern chemical industry.

### *Penicillin—Fleming, 1928*

Alexander Fleming (figure 1.8) was a Scottish bacteriologist who worked at St Mary's Hospital Medical School of London University. In 1928, Fleming was working with *staphylococcus* bacteria, and was growing colonies in Petri dishes in his musty, dusty, and none too tidy laboratory. After a vacation, he observed in a Petri dish that colonies of the bacterium *Staphylococcus aureus* failed to grow in rings around areas that had been accidentally contaminated by the green mold *Penicillium notatum*. He found a substance in the mold that prevented growth of the bacteria even when it was diluted



**Figure 1.8** Alexander Fleming, discoverer of penicillin, 1881–1955

800 times, which he called penicillin. Penicillin works by inhibiting the bacterial enzymes responsible for cell-wall synthesis, and activating other enzymes to break down the organisms' protective walls. It is not effective against microorganisms that do not produce cell walls, nor is it threatening to mammalian cells. He did not have collaborators skilled in chemistry, and was unable to grow and purify a sufficient quantity of penicillin to make clinical trials. We will discuss the subsequent developments of penicillin in Chapter 2.

### *Nylon—Carothers, 1938*

Wallace Hume Carothers (figure 1.9) was born in Burlington, Iowa, in 1896, and he entered Capital City Commercial College in Des Moines in 1914 to study accountancy and secretarial administration (Morris 1986, Hermes 1996). Later, he received his bachelor's degree from Tarkio College of Missouri in 1920, and his master's degree at the University of Illinois in 1921. He returned to Illinois to receive his doctorate in chemistry with Roger Adams in 1924, and went to teach at Harvard, before he was recruited by DuPont in 1928. He pursued his fundamental studies on the nature of polymers, and the relation between structure and properties. He established the principles of condensation polymerization, between compounds with two active ends, like ethylene glycol and dibromoethane, involving the removal of water as a by-product of reaction. The molecular weight of the polymers remained below 5000 until he introduced a molecular still that operates at low pressure to remove water, and shifted the equilibrium in favor of higher molecular weights.

Carothers decided to concentrate on the development of polyamides, as silk is also a polyamide (Houshell and Smith 1988). The best polymer should have a high enough melting point to be washed and ironed, but not too high to allow the polymer to be drawn and spun before it would decompose. The raw materials adipic acid and hexamethylenediamine could be produced from benzene, which was plentiful, so it was the 6,6'-polyamide (nylon-66) that first went into production in 1939. Carothers committed



**Figure 1.9** Wallace Carothers,  
discoverer of nylon,  
1896–1937

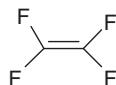
suicide in a Philadelphia hotel in 1937, after spells of deep depression. We will discuss the invention of nylon in more detail in Chapter 2.



### *Teflon—Plunkett, 1938*

In the 1930s, A. L. Henne and Thomas Midgley brought Freon 11 and Freon 12 to DuPont at the Jackson Laboratory in Deepwater, N.J. DuPont was looking for another effective refrigerant that it could sell, and the assignment was given to Roy J. Plunkett, who was aged 27 after completing his doctorate at Ohio State University (Funderburg 2000). He tried to react tetrafluoroethylene (TFE), a gas at room temperature, with hydrochloric acid. He and his assistant Jack Rebok prepared 100 lbs of TFE and stored it in pressure cylinders, to be dispensed as needed, and they kept the canisters in dry ice to prevent an explosion. On the morning of 6 April 1938, Rebok opened the valve on the TFE container and observed that nothing came out, but there was nothing wrong with the valve. They weighed the cylinders and discovered that their weights had not changed. Then they removed the valve and turned the canister upside down, shook it, and some flecks of white powder floated out. They sawed open several of the storage canisters and found that their interior walls were lined with a smooth, waxy white coating.

The resulting material had a most interesting set of properties. They noted: “It is thermoplastic, melts at a temperature approaching red heat, and boils away. It burns without residue; the decomposition products etch glasses.” It was insoluble in cold and hot water, acetone, ether, petroleum, alcohol, pyridine, toluene, ethyl acetate, concentrated sulfuric acid, glacial acetic acid, nitrobenzene, isoamyl alcohol, *ortho*-dichlorobenzene, sodium hydroxide, and concentrated nitric acid. It did not char or melt when exposed to a soldering iron or an electric arc. Moisture did not cause it to rot or swell, prolonged exposure to sunlight did not degrade it, and it was impervious to mold and fungus. He filed for a patent in 1939, which was granted in 1941.



The next stage of development went to polymer chemists and development engineers, as the expertise of Roy Plunkett was really in fluorine chemistry. The first great application was in the separation of the isotope U-235 from U-238 by gaseous diffusion of UF<sub>6</sub> to make atomic bombs, as the gas uranium hexafluoride was exceedingly corrosive and destroyed conventional gaskets and seals. PTFE was just what was needed to form the diffusion membrane, as it was not attacked by fluorine. When peace returned, PTFE registered the trademark of Teflon in 1944.

Teflon can withstand temperatures as high as 1000 °F, and it begins to soften at 620 °F, which can be a problem in heated kitchen stovetops. The bonding of Teflon to metallic surfaces was accomplished by the French engineer Marc Gregoire, who etched

the surface of aluminum with acid to create microscopically pitted surfaces, covered the surface with Teflon powder, and heated it to just below the melting point to interlock with the aluminum surface. His wife Colette had the idea to use this technology to coat her cooking pans, and a patent was granted in 1954. This led to the sale of 3 million Teflon-coated frying pans in 1958, but it was not sold in the United States until Thomas Hardie obtained a supply from Gregoire and introduced it in 1960 in Macy's in New York and Neiman Marcus in Dallas. Teflon is also used to insulate fabrics, and the Gore-Tex outdoor clothing is based on this material, which makes it proof against water penetration yet allows air to pass. Its inert properties led to use in pacemakers, dentures, medical sutures, artificial body parts, printed circuits, and thousands of other manufactured products. Ronald Reagan was named the "Teflon president," because his detractors could not make any charges that stuck to him.

### *Post-It—Silver and Fry, 1964*

3M is an adhesive company, which began by gluing abrasives on paper to produce sandpaper (Nayak and Ketteringham 1986). It became famous as the producer of Scotch tape, as well as the magnetic tape for sound recording and memory storage. The scientists at 3M were always looking for even stronger adhesives, and they believed that the ultimate adhesive is one that forms an unbreakable bond. Their goal was better glues, not weaker glues.

In 1964, Spencer Silver a chemist at 3M, was searching for polymers to make adhesives. Archer, Daniels, Midland (ADM) had developed a family of monomers that had promise. He was under the 3M research culture that scientists can use up to 15% of their time pursuing interests outside their primary assignments. He understood that science is one part meticulous calculation and one part "fooling around." He mixed up the ADM monomer at the wrong proportion, and ended up with a material that was not "aggressively" adhesive. It would produce "tack" between two surfaces, but it would not bond tightly to them. This material was more "cohesive" than "adhesive," as it clings to its own molecules better than it clings to any other molecules. Subsequent research would show that the acrylate copolymers form a suspension of microspheres, so that the adhesive has a pebbled surface instead of the flat surface of permanent tapes.

Silver's contribution is more than just the synthesis of this material, as he observed its unusual properties, and he persisted to convince others to try his creation. For many years, this weak glue had no perceptible application, and it was a solution looking in vain for a problem. In 1973, Silver visited Arthur Fry from another research group in 3M, and gave him samples of his weak glue. One day in 1974, Fry was singing in the choir of the North Presbyterian Church, and was in the habit of marking his place in musical scores with little slips of paper. He put Silver's adhesive on little pieces of paper, and so invented the better bookmark 10 years after the discovery of this weak glue.

## References

- Asimov, I. 1989. *Asimov's Chronology of Science and Discovery*. New York: HarperCollins.  
 Billmeyer, F. W. 1971. *Textbook of Polymer Science*. New York: Wiley-Interscience.

- Fenster, J. M. 2001. *Ether Day*. New York: HarperCollins.
- Funderburg, A. C. 2000. Making Teflon Stick. *American Heritage of Invention & Technology* 16(1): 10–20.
- Garfield, S. 2000. *Mauve: How One Man Invented a Colour that Changed the World*. London: Faber and Faber.
- Georg, G. I., T. T. Chen, I. Ojima, and D. M. Vyas, eds. 1995. *Taxane Anticancer Agents: Basic Science and Current Status*. ACS Symposium Series. Washington, DC: American Chemical Society.
- Glas, J. P. 1989. Protecting the Ozone Layer: A Perspective from Industry. In *Technology and Environment*, Ausabel, J. H. and H. E. Sladovich, eds. Washington, DC: National Academy Press; 137–158.
- Goodman, J. and V. Walsh. 2001. *The Story of Taxol: Nature and Politics in the Pursuit of an Anti-Cancer Drug*. Cambridge: Cambridge University Press.
- Hermes, M. 1996. *Enough for One Lifetime: Wallace Carothers*. Washington, DC: American Chemical Society.
- Hounshell, D. A. and J. K. Smith. 1988. *Science and Corporate Strategy: DuPont R&D 1902–1980*. Cambridge: Cambridge University Press.
- Israel, P. 1998. *Edison: A Life of Invention*. New York: John Wiley & Sons.
- Korman, R. 2002. *The Goodyear Story*. San Francisco: Encounter Books.
- Korolkovas, A. 1988. *Essentials of Medicinal Chemistry*. New York: John Wiley.
- McGrayne, S. B. 2001. *Prometheans in the Lab: Chemistry and the Making of the Modern World*. New York: McGraw-Hill.
- Midgley, T. Jr. 1937. From the Periodic Table to Production. *Industrial and Engineering Chemistry* 29(2): 241–244.
- Morris, P. J. T. 1986. *Polymer Pioneers*. Philadelphia, PA: Chemical Heritage Foundation.
- National Academy of Engineering. 2001. *Greatest Engineering Achievements of the 20th Century*. Washington, DC: National Academy Press.
- Nayak, P. R. and J. M. Ketteringham. 1986. *Breakthroughs*. New York: Rawson Associates.
- Suffness, M., ed. 1995. *Taxol: Science and Applications*. Boca Raton, FL: CRC Press.
- Tanner, D., J. A. Fitzgerald, and B. R. Philips. 1988. *The Kevlar Story*. DuPont presentation at Advanced Material Conference, Hotel DuPont, Wilmington, DE.
- Wilson, R. E. 1937. Perkin Medal: the Medalist. *Industrial and Engineering Chemistry* 29(2): 239–241.
- Wolf, R. F. 1939. *India Rubber Man: The Story of Charles Goodyear*. Caldwell, ID: Caxton Printers.

## Further Reading

- Bowden, M. E. 1997. *Chemical Achievers*. Philadelphia: Chemical Heritage Foundation.
- Bowden, M. E. and J. K. Smith. 1994. *American Chemical Enterprise*. Philadelphia, PA: Chemical Heritage Foundation.
- Duerr, M. G. 1966. *The Commercial Development of New Products*. New York: The Conference Board.
- Elder, A. L. 1970. *The History of Penicillin Production*. Chemical Engineering Progress Symposium Series Vol. 66. New York: American Institute of Chemical Engineers.
- Guthrie, V. B. ed. 1960. *Petroleum Products Handbook*. New York: McGraw-Hill.
- Gwynne, P. and G. Heebner. 2003. Drug Discovery and Biotechnology Trends. *Science* 299: 915–921.
- Kirk-Othmer Encyclopedia of Chemical Technology. 2001. New York: John Wiley.

- Landau, R., B. Achilladelis, and A. Scriabine, eds. 1999. *Pharmaceutical Innovations*. Philadelphia, PA: Chemical Heritage Press.
- Lorsch, J. W. 1965. *Product Innovation and Organization*. New York: Macmillan.
- Nobel e-Museum. 2002. *Nobel Lectures*. Stockholm: Nobel Foundation.
- Raber, L. R. 2001. Scale-up of a Miracle Drug. *Chemical & Engineering News* (30 July): 63–64.
- Rothberg, R. R. 1976. *Corporate Strategy and Product Innovation*. New York: Macmillan.
- Smil, V. 2001. *Enriching the Earth: Fritz Haber, Carl Bosch and the Transformation of World Food Production*. Cambridge, MA: MIT Press.
- Sneader, W. 1985. *Drug Discovery: The Evolution of Modern Medicines*. New York: John Wiley.
- Ullmann's Encyclopedia of Industrial Chemistry*. 2001. Weinheim: VCH.
- Wintermantel, K. 1999. *Process and Product Engineering*. Chemical Engineering Science No. 54. Oxford: Pergamon Press: 1601–1620.
- Yang, R. T. 2003. *Adsorbents: Fundamentals and Applications*. New York: Wiley-Interscience.

## Websites

- <http://www.google.com>  
[http://www.invent.org/hall\\_of\\_fame](http://www.invent.org/hall_of_fame) Inventors Hall of Fame  
<http://www.chemistry.org/milestones> American Chemical Society  
<http://www.greatachievements.org/> National Academy of Engineering  
<http://www.nobel.se/> Nobel e-Museum  
<http://www.chemheritage.org> Chemical Heritage Foundation of Philadelphia  
<http://www.greatachievements.org/greatachievements/> National Academy of Engineering

## Discussion Topics

1. Consider the three suggested criteria to rate the greatness of an invention, given at the beginning of Part I. How would you rank the relative importance of these criteria, and why? Would you like to suggest other criteria of greatness, and rank them together with the others?
2. Who were the lucky inventors who were in the right place at the right time; and who were the intellectual giants who did their inventions with unique scientific insight and skill? Is one type more worthy of admiration than the other, and why?
3. Some discoverers, like Alexander Fleming with penicillin, abandoned the development to others, as he had no organization to back him. Some discoverers, like Thomas Midgley and Roy Plunkett, were happy to discover great research leads and turned over the problem of developing commercial products to others in the company. William Perkin went all the way from research discovery to building his own company, and carried out the development and commercialization work. How do you view the contributions of these three types of discoverer, and how does the world reward them?
4. In the product life cycle, would it be accurate to associate the synthesis of new material with spring, the search for new material as summer, and the modification of current product as autumn?

5. What are the greatest needs of mankind for new chemical product inventions, and what are plausible technology leads or hunches on where they can be found? Would you like to pursue this when you graduate?
6. If your goal is to become a great inventor of a chemical product, how would you prepare yourself and what knowledge should you try to acquire in the next 5 to 10 years?

## Exercises

1. What would you regard as the greatest inventions in the twentieth century, and what are the greatest *chemical* inventions in the twentieth century? Please state the set of criteria of greatness that you believe in, and compare them with your great inventions.
2. Do a literature research on your top two *chemical* product innovations, read up on the inventors, the circumstances leading to their invention, whether they led the commercialization, the impact of the invention in changing many lives the world, and in opening doors to other inventions.
3. For the arduous task of innovations, the inventors were driven by a number of motivations: intellectual curiosity, fame and fortune, desire to help others, and so on. Compare and rank these motivations in your estimation and admiration.

## Product Development to Business

### **2.1 Nylon: Centralized Development**

Exploration–discovery phase  
Development phase  
Commercialization and business phases

### **2.2 Penicillin: Distributed Development**

Exploration–discovery phase  
Development phase  
Commercialization–business phases

### **2.3 Taxol: Distributed Development**

Exploration–discovery phase  
Development phase  
Commercialization and business phases

### **2.4 Tasks and Skills in Product Innovation**

**References**  
**Further Reading**  
**Discussion Topics**

Many historians and journalists conclude an exciting story of innovation at the discovery phase, when there emerges a plausible idea of a new product. However, for most innovative products, the exciting story has barely begun. Usually, many years of hard work still lie ahead, to modify and adapt the technology to the needs of the marketplace, and to find solutions to many problems. Sometimes, the development work is accomplished mainly inside a single organization, such as in the development of nylon by DuPont; at other times, the development work is accomplished by many units under a loose network of cooperation, such as in the development of penicillin and of taxol by numerous companies and organizations.

The innovation project will not go forth unless sponsors can be found and persuaded that the product can be manufactured from available technology and raw materials, accepted in the marketplace as superior in quality to other products at a competitive price, acceptable in safety and environmental concerns, and make a handsome profit for the manufacturer. The critical go-ahead signal is in obtaining finance to pay for the capital cost of land, building, and equipment, as well as the working capital cost of hiring management and staff, buying raw materials, and paying for utility and transportation. Other major landmarks are the start-up of the first manufacturing plant and the initial product offer in the market. If the product is to remain viable in the

marketplace for many years, then there must be vigilance in monitoring the reactions of the customers, as well as that of the competition and the government, and taking appropriate actions.

The development stories of numerous products are not well recorded or are completely lost. We are grateful that there are a few stories that have received attention from the innovators who took pains to record their stories.

## 2.1 Nylon: Centralized Development

The book *Science and Corporate Strategy: DuPont R&D, 1902–1980* by Hounshell and Smith (1988) presents a well-documented story of the long journey of nylon from discovery to the marketplace. A large number of people were involved. The effort received many lucky breaks, as well as entering and abandoning numerous blind alleys, and requiring dedication and determination when things looked bleak. The journey can be divided into a number of phases that overlap in time: the exploration–discovery phase, the development phase, and the commercialization–business phases.

### Exploration–Discovery Phase

The story began in December 1926, when the director of the Chemistry Department, Charles Stine, submitted to the Executive Committee of DuPont a short memorandum entitled “Pure Science Work,” which represented a radical new approach to establishing or to discovering new scientific facts, instead of the traditional policy in industry of applying previously established scientific facts to practical problems. Stine had discussed the idea with other leading research people, including Arthur D. Little the consultant, C. E. K. Mees of Eastman Kodak, Whitney of General Electric, Edward B. Craft of the Bell Telephone Laboratory, and Professor Emmet Reid of Johns Hopkins University. After he received approval and funding from his management, he went about recruiting outstanding research chemists, but by February 1928 he had recruited only one man, a 31-year-old instructor from Harvard named Wallace Hume Carothers.

The nature of polymers was still being debated, and the Nobel Prize in Chemistry for Hermann Staudinger came decades later in 1953, for his proposal in the early 1920s that polymers are long chains of monomers held together by ordinary chemical bonds. Carothers supported this point of view, and proposed a scheme to prove it. He proposed to build long-chain molecules one step at a time, by carrying out well-understood chemical reactions. His scheme was to react dihydric alcohol molecules with –OH groups on both ends, together with diacid molecules with acid groups on both ends, to make a polyester of potentially unlimited length. He and Julian Hill, a member of his research group, were able to make chains with a molecular weight of 5000 to 6000, and he called them “condensation polymers.” Water, which forms as a by-product of this condensation reaction, must be removed to push the equilibrium towards a higher polymer. He remembered hearing at a conference about a molecular still, which operates under vacuum, to remove water; and so they built one. In April 1930, they heated a 16-carbon diacid with a three-carbon dihydric alcohol, removed the product from the

still, and found a molten polyester with a molecular weight of 12,000. They made an unexpected discovery: after cooling, this fiber could be stretched or “cold drawn” into very strong fibers. However, they found that this 3–16 fiber was not suitable as a textile fiber because it melted below 100 °C, was partially soluble in dry-cleaning solvents, and was sensitive to swelling in water. The discovery was made in April 1930, and the long journey of development and commercialization would take 10 years before nylon became generally available for sale as stockings.

This all-synthetic fiber has different properties from the semi-synthetic rayon, as it retains nearly all its strength when wet and has an elasticity that only silk can match. Their next move was to switch from dihydric alcohols to diamines, which react with a diacid to form a polyamide, as they knew that simple amides have higher melting temperatures than the corresponding simple esters. However, the desirable properties for a textile fiber of high melting point and low solubility in water and solvents would also make it difficult to spin into fibers before knitting into fabrics. The work was halted in the middle of 1933, due to the inability to find ways of making progress, until it was taken up again early in 1934 by his assistant Donald D. Coffman. In March 1934, Coffman drew a fiber from 4 g of polymer that he had made, and heated it in a bath at 200 °C, just above its melting point, and drew a lustrous filament of polyamide fiber. That was the first nylon fiber. Not long after that, Carothers went into an unusually severe depression and was hospitalized for a while.

### Development Phase

From the day that Hill pulled the first polyester fiber, DuPont focused its energies on creating a synthetic fiber, instead of polymer resins or rubber. An early decision was made to concentrate only on full-fashioned silk hosiery. A silk stocking contains only 10 g of resin, but a wool sweater consumes much more resin and would require a much larger plant investment. At that time, \$70 million worth of silk went into stockings each year, which were knitted into eight pairs of stockings per American consumer per year. The tasks of development and commercialization were formidable. The technology to manufacture the intermediate chemicals and the technology of spinning the polymer into fibers had to be developed. The numerous small textile mills that knit silk into stockings were incapable of adapting to knitting nylon, as they did not have the staff and the technical capabilities to do new product development, and they were not used to large-scale advertising to educate and persuade customers. DuPont was able to undertake the development, as it had the necessary resources and connections: its Ammonia Department had experience with high-pressure catalytic reactions, and its Rayon Department had experience in manufacturing and selling to the textile business.

Carothers collected much physical, chemical, and mechanical data on the fibers, but he did not know how the data correlated with “ultimate practical behavior” and performance to the satisfaction of consumers. The only way to know whether nylon would find favor in the marketplace was to make fabric samples and to prepare garments to be worn and to be maintained. In the summer of 1934, his group also had to determine which combination of diacid and diamine would be the best, within the range from 2 to 10 carbon atoms. The initial focus was on the 5–10 polymer; enough polymers were

made to spin into fibers and made into fabrics and then sent to the Rayon Department for testing. It is easy to polymerize and spin 5–10, but it melted at 190 °C, so that it could not withstand practical ironing temperatures. Gerald J. Berchet prepared nylon 6–6 in February 1935, which would become the commercial product. This nylon 6–6 melts at over 250 °C, which is excellent. It is even more important that the raw material comes from benzene, which in turn comes from easy to obtain coal tar or petroleum refining. They still had to solve the problem of the commercial production of the two starting materials, namely adipic acid and hexamethylenediamine. But the difficulty was greater in the spinning process, which could be carried out either from a solution, or from a melt at above 260 °C. The fine filaments had to be twisted together, as is done with silk, in bundles of 20 or 30 to make a textile fiber. Then it had to be tested on standard textile machinery and put through such commercial finishing processes as washing and dyeing.

The Chemical Department made the decision to develop nylon 6–6 in July 1935. The process to make hexamethylenediamine required a multistep synthesis, involving: hydrogenating benzene to make cyclohexane, which was oxidized to make a mixture of cyclohexanone and cyclohexanol, which was oxidized to adipic acid (one of the components of nylon 6–6). Half of the adipic acid was catalytically reacted to make adiponitrile, which was hydrogenated to make hexamethylenediamine (the other component of nylon 6–6). Several batches of the order of a few pounds of intermediates were needed to do the polymerization reaction. Melt spinning procedures at a temperature of about 260 °C were developed. By the summer of 1936, DuPont became confident that they could make nylon yarn at a cost of 80¢ per pound in a plant of 8 million pounds capacity, which could be sold at \$2.00 per pound at roughly the price of silk. They decided to expand the manufacturing capacity from 2 lbs to 100 lbs a day to improve confidence in the process, and to provide material for extensive field testing. In February 1937, a batch of yarn was taken for a knitting test at the Union Manufacturing Company in Maryland; the results were unsatisfactory, as the product did not come off the spools easily, snagged at knitting machines, and looked like a wrinkled mess with an unpleasant gray color.

In the meantime, the mental condition of Carothers deteriorated inexorably. He was elected to the National Academy of Sciences in 1936. But he had another breakdown in the summer of 1936, and finally, in April 1937, he committed suicide with cyanide in a Philadelphia hotel room. At Christmas 1937, the Van Raalte Mill in New Jersey finally turned out "full-fashioned hosiery with excellent appearance and free of defects." They distributed 56 hosiery and lingerie garments to wives of the men on the nylon project. The majority of the women thought that the stockings were very durable, but they wrinkled too easily and were too lustrous and slippery.

### Commercialization and Business Phases

The Executive Committee of DuPont authorized a pilot plant on January 1938, which took 6 months to build and another 6 months to provide sufficient data to help in the construction of a commercial plant. In October 1938, the Executive Committee allocated \$8.5 million to build a 4 million pound a year plant at Seaford Delaware, which began operation in January 1940. Charles Stine announced this new product

in October 1938 at the New York World's Fair to 3000 members of a women's club. He said:

This textile fiber is the first man-made organic textile fiber prepared wholly from new material from the mineral kingdom. . . . Though wholly fabricated from such common raw material as coal, water, and air, nylon can be fashioned into filaments as strong as steel, as fine as spider's web, yet more elastic than any of the common natural fibers.

Nylon finally became available to the general public in May 1940. Ten years had passed from initial discovery to full commercialization. It was a tremendous effort, even for a company with the resources of DuPont, and the R&D cost was \$4.3 million. During World War II, DuPont nylon production went up to 25 million pounds a year, and was used to make parachutes, airplane tire cords, and glider tow ropes. DuPont resumed selling nylon for stockings after the war.

## 2.2 Penicillin: Distributed Development

The development of penicillin into one of the most influential antibiotics of the twentieth century followed an entirely different and accidental path from that of Freon, which was a purposeful search with a definite goal in mind (the need for a nonflammable and nontoxic refrigerant), and developed within a single organization. In his Nobel Lecture in December 1945, Alexander Fleming said (Nobel e-museum 2002):

It arose simply from a fortunate occurrence which happened when I was working on a purely academic bacteriological problem which had nothing to do with antagonism, or molds, or antiseptics, or antibiotics. In my first publication I might have claimed that I had come to the conclusion, as a result of serious study of the literature and deep thought, that valuable antibacterial substances were made by molds and that I set out to investigate the problem. That would have been untrue and I preferred to tell the truth that penicillin started as a chance observation. My only merit is that I did not neglect the observation and that I pursued the subject as a bacteriologist. My publication in 1929 was the starting-point of the work of others who developed penicillin, especially in the chemical field. . . . We tried to concentrate penicillin, but it is easily destroyed, and to all intents and purposes we failed. We were bacteriologists—not chemists—and our relatively simple procedures were unavailing.

### Exploration–Discovery Phase

It began as an accidental discovery and observation that lay dormant for many years, as the discoverer lacked the means and connection to develop a product with sufficient quantity and purity to be used in clinical use. It laid dormant until another team appeared on the horizon, and it took World War II to provide the motivation and the resources critical to its development before penicillin became the life saver for millions.

Penicillin, the first and still one of the most widely used antibiotic agents, is derived from the *Penicillium* mold. Alexander Fleming was a Scottish bacteriologist who took his degree at St Mary's Hospital Medical School of London University and continued his research in the Royal Army Medical Corps in World War I. He was familiar with the harsh chemical antiseptics used to clean wounds, which often did too much damage to body tissues but not enough damage to the microbes. In 1928, Fleming was working with *Staphylococcus* bacteria, and was growing colonies in Petri dishes in his musty, dusty, and none too tidy laboratory. After a vacation, he observed in a Petri dish that colonies of the bacterium *Staphylococcus aureus* failed to grow in rings around areas that had been accidentally contaminated by the green mold *Penicillium notatum*. He found a substance in the mold that prevented growth of the bacteria even when it was diluted 800 times, which he called penicillin. Penicillin works by inhibiting the bacterial enzymes responsible for cell-wall synthesis, and activating other enzymes to break down the organism's protective walls. It is not effective against microorganisms that do not produce cell walls, nor is it threatening to mammalian cells. He was aware of the significance of his discovery, but he lacked the necessary chemical knowledge and equipment to isolate and identify the active compound involved. Pure penicillin is very unstable, and he was only able to make solutions of crude material to work with. He was thus unable to obtain a sufficient quantity of penicillin to test its therapeutic value for use on animals or humans.

### Development Phase

The work lay dormant for 12 years, until the German fire-bombing of London in World War II, when an Oxford group began searching for new antibacterial drugs for burns treatment. Howard Florey was an Australian pathologist who studied medicine at Adelaide and Oxford and became a professor of pathology at Oxford in 1935. Ernst Chain was a German-born biochemist who studied chemistry and physiology in Berlin and fled to England to work with Florey at Oxford. They were also encouraged by the news of the success of the new sulfa drugs of Gerhard Domagk. In 1939, they studied a number of antibacterial substances and decided to concentrate on the penicillin mold, making an injectable form of the drug. They demonstrated its therapeutic effectiveness on mice, and then on humans suffering from *Staphylococcus*. But it was clear that wartime England did not have the resources to develop large-scale industrial production of penicillin.

In the summer of 1941, Florey came to the United States and sought help from the National Research Council, which sent him to the Northern Regional Research Laboratory at Peoria, Illinois (Elder 1970). This was a laboratory devoted to finding new industrial uses for surplus farm commodities, and had a new fermentation division. In the beginning, the fermentation scheme in solid surface culture produced penicillin at 3 units/mL, with a unit containing about 0.6 µg of pure sodium penicillin G. The staff at Northern Regional switched the growth medium to liquid corn steep-lactose, and the production rate increased to 100 units/mL. They realized that different strains of mold may produce different quantities of penicillin, and so began a worldwide search for high-yield molds coming from far-ranging places, such as Chunking, Calcutta, and North Africa. It is with some irony that, after screening thousands of strains

of *Penicillium* molds, the best strain found was from a moldy cantaloupe in a Peoria fruit market, and the production rate increased to 1500 units/mL.

The U.S. Office of Scientific Research and Development under Vannevar Bush took an interest in this project. Howard Florey needed 1 kg of penicillin for clinical trials. A consortium of pharmaceutical companies was summoned, including Abbott, Lederle, Merck, Pfizer, and Squibb, in December 1941 and given the challenge of industrial-scale production. Large-scale production received a big boost when the fermentation vessels were changed from the surface of Petri dishes to submerged fermentation in bottles. Merck and Pfizer scheduled experiments in 100,000 1 L bottles, each containing 200 mL of medium, mounted on shake tables to stir the fermentation and to aerate the broth. This was considered a big improvement at that time, but the next breakthrough was the development of deep tank fermentation in 10,000 gallon tanks, which is much more efficient and is still the reactor of choice today.

The fermentation broth typically contains 20–30 mg/L of antibiotics, which is to say 30 parts per billion, and must be extracted into concentrated form using solvent extraction. The solvent extraction method was developed by Shell Oil and by Podbielniack and is based on the principle that penicillin is hydrolyzed in aqueous medium to  $H^+$  and  $RCOO^-$  ions. Thus, equilibrium in an acidic medium (i.e., one with low pH or high  $H^+$  concentration) is favored by the neutral  $RCOOH$  form, whereas equilibrium in an alkaline medium (i.e., one with high pH or low  $H^+$  concentration) is favored by the  $RCOO^-$  ionic form. The neutral form is more soluble in an organic medium, and the ionic form is more soluble in an aqueous medium. Thus, with amyl acetate as the organic solvent the partition coefficient of penicillin between solvent and water is about 100 at pH 3 and about 1 at pH 6. In the industrial process, the aqueous broth was acidified to pH 3 for the extraction into the organic solvent, and alkalized to a pH 6 for reverse extraction back into an aqueous medium.

Natural penicillin is fragile and breaks down in the presence of impurities and enzymes, having a half-life of just 2.5 h even when cooled to 0 °C. The method of choice for stabilizing penicillin for storage and shipment turned out to be the freeze-drying method. Penicillin is frozen in trays at –30 °C; it is then placed in a chamber at a pressure of 0.1–0.6 Torr to sublimate the ice crystals into water vapor and removed.

### Commercialization–Business Phases

A glimpse of the progress in production can be seen with the following figures:

1943	1.7 billion units/month
1944	138.0 billion units/month
1945	570.0 billion units/month

In the human body, natural penicillin breaks down under acidic conditions as it passes through the stomach, so semi-synthetic penicillin was developed that is more acid stable and can be given in oral medication. Penicillin has significant side effects, in causing allergic or hypersensitivity reactions, including skin rashes, hives, swelling, and anaphylaxis or allergic shock. The War Production Board took a great interest in the project, and ample stock became available for the D-Day invasion of Europe

on 6 June 1944. Penicillin has been proven to be the most effective medicine against throat infections, pneumonia, spinal meningitis, gas gangrene, diphtheria, syphilis, and gonorrhea.

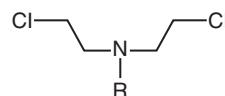
Fleming, Florey, and Chain received the Nobel Prize in Physiology or Medicine in 1945. Florey said in 1949, “Too high tribute cannot be paid to the enterprise and energy with which the American manufacturing firms tackled the large-scale production of the drug. Had it not been for their efforts, there would certainly not have been sufficient penicillin by D-Day in Normandy in 1944 to treat all severe casualties, both British and American.” (Nobel e-Museum 2002). Sometimes there is a very long journey from initial observation and discovery, to significant production and purification, to proof of medical efficacy and safety, to production in quantity, and finally to the marketplace.

### 2.3 Taxol: Distributed Development

There are several books on the history of the development of taxol, which is one of the most remarkable stories in product development. In fact, it inspired the 1992 motion picture *Medicine Man*, starring Sean Connery as a research botanist looking for a cancer cure in the Brazilian rain forest. For a time, it became a moral drama pitting the needs of patients of intractable ovarian and breast cancer against the passions of environmentalists to preserve an obscure Pacific yew tree. Suffness and Wall are two of the principals in this story, and they wrote (1995): “It [Taxol] is not an obvious winner till the very end, and there were a number of times till the very end when it seemed highly likely that it would not be put into development at all, or that once it had been accepted, it would be dropped.” More than 30 years passed between the discovery of taxol, with its potential as an anticancer drug, and its approval by the Food and Drug Administration (FDA) for clinical use.

#### Exploration–Discovery Phase

The discovery of penicillin and its successful application in World War II inspired the “antibiotic era,” and a broad search for other cures for infectious diseases. Cancer has a totally different cause, as it arises through the malignant mutation of normal cells instead of from the actions of bacterial or other outside organisms. Penicillin destroys the bacteria cell walls, but not the mammalian cell membranes. Unless a drug could be found that could tell the difference between a normal cell and a cancer cell, then it was not clear that there would be an effective cancer drug, that is until the first report by Goodman in 1946 that nitrogen mustard, developed as a war gas, was an effective chemotherapeutic for human leukemia.



This led to cooperation between the Army Chemical Warfare Service and Sloan–Kettering Cancer Institute “to concentrate on the organization of industrial techniques

for cancer research." The screening program lasted from 1947 to 1955, when they tested 20,000 synthetic and natural compounds; they standardized the test by testing with only the sarcoma 180 mouse tumor, but the results were not impressive. Another impetus was the 1955 discovery from Eli Lilly and the University of Western Ontario that, whereas the leaves of the Madagascar periwinkle were not effective in treating diabetes as hoped, they were, however, effective against rat leukemia, which led eventually to the vinca alkaloid drugs.

Pressure for a national cancer drug screening program led in 1953 to a Congressional directive to the National Cancer Institute (NCI) to organize a Cancer Chemotherapy program. The newly created Cancer Chemotherapy National Service Center in 1955 set up a screening program for antitumor agents in the synthetic and fermentation products with known chemical structures. In 1960, the program was extended to natural plant and animal products, in which the structures were unknown. The organic chemist Jonathan Hartwell, who had an interest in folk medicine, took a leadership role in the screening program. He drew up an interagency agreement with Bob Perdue of the U.S. Department of Agriculture (USDA) to collect over 35,000 plant species to test for anticancer activity between 1960 and 1980.

Perdue and Hartwell argued that selective searching, of targeting a specific promising family, would lead to higher number of hits, but it would not discover substances possessing entirely new structures and mechanisms of activity. Thus, random selection of plants for screening was their main collection method. NCI contractors handled most of the work, and the crude materials were sent to the Wisconsin Alumni Research Foundation for plant extraction, and then to the Research Triangle Institute for preclinical evaluations. In addition to collecting new plant samples, attempts were also made to secure access to existing collections of plant extracts made for other purposes.

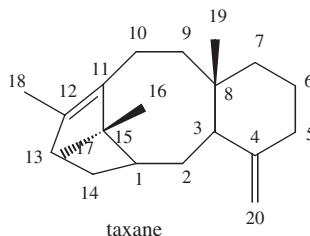
In August 1962, USDA botanist Arthur S. Barclay collected 650 plant samples in California, Washington, and Oregon, including barks, twigs, leaves, and fruits of the rare Pacific yew *Taxus brevifolia* in Washington State. Yew has an ancient reputation as sacred to the Celts, and a symbol of immortality; in fact, Julius Caesar mentioned poisoning by yew in his book *Gallic War*. Pacific yew *T. brevifolia* is a very slow-growing conifer found principally in the understory of old-growth forests in the Pacific Northwest, towered over by Douglas firs and hemlocks. Bulk plant samples were dried and shipped to the NCI contract extraction laboratory at the Wisconsin Alumni Research Foundation, and were tested against KB cells in Bethesda, Maryland, April 1964, and was found in a quick in vitro test in Petri dishes to be cytotoxic. The first in vivo test of the extract was done in April 1964 against the L1210 leukemia at Microbiological Associates, and proved toxic, but the second test showed no activity.

Monroe Wall of the USDA Eastern Regional Research Laboratory had been collecting plants for steroids that are oxygenated at positions 11 or 12, which could be converted into cortisone and related compounds. The search for a cancer cure led Wall to move to the Research Triangle Institute in 1960, to work on the isolation of plant-derived antitumor agents. Wall requested specimens of plants showing KB activity, and worked with a 30 lbs shipment of bark in late 1964. The procedure included extraction by ethanol, followed by concentration and partition between water and an organic solvent. He found that fractions from this extract were active in vivo for mice with

P-1534 leukemia, the Walker 256 carcinosarcoma, and the P388 leukemia. He also worked on isolation by extraction of 12 kg of Pacific yew bark with ethanol, followed by partition of the ethanol extract between chloroform and water. In his first publication in 1967, about 0.5 g of taxol was isolated from 12 kg of air-dried stem and bark from *T. brevifolia*, and the yield was about 0.004%, or 40 ppm.

### Development Phase

In 1971, Wall determined the structure of taxol with colleagues in X-ray crystallography and nuclear magnetic resonance. Taxol belongs to a family of compounds based on the taxane molecule,  $C_{20}H_{32}$ , a diterpene as the condensation of four terpene molecules (2-methyl-1,3-butadiene). Taxane is based on the fusion of an 8-ring with two 6-rings of carbon. The official name of taxol is paclitaxel, with a formula of  $C_{47}H_{51}NO_{14}$ , a molecular weight of 853.92, and a melting point of 220–222 °C. It is practically insoluble in water.



The development efforts of taxol slowed down due to difficulties with four problems, and could not become a candidate for development till they were solved:

- Its activity against L1210 leukemia in rodents and against solid tumors was only modest.
- It requires a high dose of 10–20 mg/(kg day) to be effective.
- It has a very low aqueous solubility of less than 0.01 mg/mL, or 10 ppm by weight, so a way must be found to increase its solubility for clinical administration.
- It was difficult and expensive to isolate taxol on a large scale, as *Taxus* leaves contain only 20–70 ppm taxol, and the bark contains 70–400 ppm taxol. The tree *T. brevifolia* is found only in a narrow geographic belt in the Pacific Northwest, and grows to 60–75 ft tall with a diameter of 18–30 inches in 100 years. One has to kill and strip the bark from six trees to obtain enough taxol to treat one patient. There must be a demonstration of adequate supply, sufficient for the coming animal toxicology test and the Phase I clinical test. A little calculation will show that, to treat a 70 kg person, there is need of a daily dose of 1.4 g, leading to

$$\frac{70 \text{ kg}}{\text{person} \cdot \text{day}} \times \frac{20 \text{ mg}}{\text{kg} \cdot \text{day}} \times \frac{10^6 \text{ mg bark}}{100 \text{ mg taxol}} = \frac{14 \text{ kg bark}}{\text{person} \cdot \text{day}}$$

It languished on the shelf of NCI during 1971–1974. Fortunately, in 1974, taxol was clinically found to be effective against the B16 melanoma, so it was finally selected for development. The formulation problem to increase solubility was solved by using a Cremophor–ethanol EL surfactant (a modified castor oil), which made a stable enough emulsion. A major issue in scale-up was to determine which species of *Taxus* plant material possessed the dual properties of high taxol content and adequate abundance. An extensive survey of a variety of species and geographic studies determined that Pacific yew was the best source of taxol. The Polyscience Company had the contract to acquire 10,000 lbs of Pacific yew bark, to yield 500 g of drug-quality taxol, which is equivalent to a yield of about 100 ppm.

A scientific breakthrough in understanding came in 1979, when Susan B. Horwitz and coworkers in Albert Einstein College of Medicine in Bronx, N.Y., discovered the mechanism of taxol action. Their study found a unique mechanism for the antitumor activity of taxol that involved cell microtubules, which play a key role in mitosis, for the maintenance of cell shape, cell motility, and intracellular transport. A substance that interferes with microtubules can disrupt cell growth and function. This understanding gave a firm theoretical foundation to the understanding of the anticancer mechanism of taxol, and created greater respect and interest in the drug.

In 1982, animal toxicology studies were completed, and taxol was approved by the NCI for an Investigational New Drug Application filing, which was approved by the FDA. In 1984, Phase I clinical trials began on a few terminally ill patients at seven clinical sites. The aim was to determine the maximum tolerated dose in humans, and to discover any dose-limiting toxicity. Regardless of its excellent activity in model tumor systems, clinical trials were often delayed owing to short supplies of the drug. At first, the drug produced some serious hypersensitivity reactions. By slowing the rate of infusion or premedicating patients with antihistamines and steroids, such reactions were averted. The Phase II clinical trials on breast cancer were on a far bigger scale than Phase I, in which drug effectiveness was tested on a large number of patients. An additional Phase II clinical trial in refractory ovarian cancer began in 1989.

Since the bottleneck on the clinical trial for taxol was in the supply, the NCI turned in 1989 to industrial partners, and issued a request for a Cooperative Research and Development Agreement, and selected Bristol-Myers Squibb (BMS) in 1991 as the partner in taxol development. BMS would take responsibility for the short-term supply of taxol, and NCI would sponsor research to deal with long-term supply. BMS collected 750,000 lbs of dried *T. brevifolia* bark from 38,000 trees during the 1991 growing season, sufficient to yield 25 kg of pure taxol to treat about 12,000 cancer patients. Hauser Chemical Research of Boulder, Colorado, was overseeing collection of yew bark and processing of the bark to extract taxol. BMS prepared the final dosage formulation and delivered it to the NCI for use in clinical studies.

Some 60,000 U.S. women die of ovarian and breast cancer each year. Each patient would need 2 g of taxol, which means 120 kg of taxol per year. This would mean 800,000 *Taxus* trees per year, assuming a yield of 1500 ppm. Since six century-old trees had to be killed to treat each patient, the destruction or extinction of these trees aroused the anger of many environmentalists. To ensure the future supply of trees, the Weyerhaeuser company was asked to grow more than 500,000 yew plants as starter material for future production. In 1990, a large-scale clinical Phase III trial began on ovarian cancer, and the supply of taxol for these crucial clinical trials became

extremely acute. Fortunately, the clinical results were outstanding at the M. D. Anderson Cancer Center, and confirmed at Johns Hopkins and at Memorial Sloan-Kettering.

In 1992, a New Drug Application for taxol was filed by BMS. The FDA approved taxol for the treatment of refractory ovarian cancer 6 months after filing. In December 1992, the FDA approved taxol for marketing. It had been 30 years since the first collection of Pacific yew for testing at the Research Triangle Institute. In 1993, taxol was marketed by BMS. In 1994, a supplemental FDA approval was issued for taxol in treatment of metastatic breast cancer.

### Commercialization and Business Phases

The supply of taxol must not depend solely on the Pacific yew, and a number of alternate sources were studied. Other species of tree were sought, and cell culture was investigated. Many people were involved in making synthetic taxol, which included total synthesis from small molecules, such as camphor, and semi-synthesis from readily available precursors. The semi-synthesis of taxol was accomplished by the extraction of the intermediate 10-DAB from the needles of *Taxus baccata*, the English yew, which is a common ornamental shrub, with a yield that is six to eight times higher. A great advantage of this new source is that the collection of needles does not require killing the tree. BMS announced in 1993 that it had licensed the method of semi-synthesis from Bob Holton of Florida State University, and it will cease to acquire raw material for taxol from forests. This put an end to the dramatic and emotional battle between cancer patients and environmentalists.

Taxol has received considerable competitive challenge from taxotere, officially named docetaxel, a patented taxol analog developed by Rhône-Poulenc, based on the 1981 first successful semi-synthesis by Gueritte-Voeglein from 10-DAB. In 1987, Rhône-Poulenc selected taxotere for clinical development, and it is now used clinically in the United States.

Taxol is the showcase for the billions spent by the NCI over many decades. But there has not yet been a dramatic decrease in the death rate from cancer in the United States. Future challenges include discovering whether some derivatives of taxol would be even more effective than taxol, more easy to administer, and could be made at lower cost. The total synthesis of taxol also gives us the tool to investigate numerous alternatives. Another challenge is whether we can go beyond ovarian and breast cancer to treat the big killers of lung and stomach cancer.

## 2.4 Tasks and Skills in Product Innovation

Product innovation involves a team effort with people of different skills and knowledge. Technology and marketing are two principal skills required, as a successful product is designed to be manufactured to be sold to customers. Many other skills also play critical roles, such as management, finance, accounting, government relations, and public relations. The task requirements change in different phases of product innovation, and the product engineers play leading roles at times and supporting roles at other times.

It would be useful for the product engineers to understand the nature and contributions of the other people on the team, especially the marketing people, who are their principal partners. The information that engineers work with is from authoritative textbooks, derived from fundamental and unchanging laws of nature, such as Newton's law of motion and the second law of thermodynamics, and supplemented by precise measurements by dedicated and highly motivated scientists and engineers. In contrast, the information gathered in marketing is collected from surveys of the poorly informed opinion of hundreds and thousands of consumers, who give only quick answers that are changeable, inaccurate, and possibly contradictory. To derive useful information from such a bazaar is an imprecise art, in marked contrast to getting a boiling point from a handbook.

Phase 1 of exploration-discovery is principally a collaboration between the product engineers and the marketers. For the market-pull approach, the marketers start the dialog and the product engineers respond, such as the case of Charles Kettering and Thomas Midgley with Freon. Several technologies are suggested, samples made and tested, and the two partners keep each other informed as results pour in. An evaluation of the results may lead to revision of the technology, sample collection and testing, and another round of evaluation. Eventually, they arrive at one or more good candidates to recommend for development, or they abandon the project. The major tasks of this process are shown in table 2.1, where the involvements of product engineers are marked by  $\times$  and the leading roles of product engineers are marked by  $\times\times$ . For the technology-push approach, the product engineers start the dialog and the marketers respond. Several markets are suggested and tested, and the results may lead to the

**Table 2.1 Phase 1, exploration-discovery phase**

<i>Market-pull tasks</i>	<i>Product engineering tasks</i>	<i>Tech-push task</i>	<i>Product engineering tasks</i>
Identify and study customers and needs, specify properties required		Identify and study technology, develop capabilities, study properties	$\times\times$
Identify and study competitive products		Identify and study competitive technologies	$\times\times$
Generate suggestions and test suitable technology	$\times\times$	Generate suggestions of suitable market that would appreciate properties	
Collect and synthesize samples	$\times\times$	Make samples	
Lab testing of properties	$\times\times$		
Compare required properties with sample properties	$\times$	Market testing of customer appreciations	
Recommend candidate product	$\times$	Compare suggested with sample appreciations	
		Recommend candidate product	$\times$

$\times$ : product engineering involved;  $\times\times$ : product engineering plays leading role.

**Table 2.2 Phase 2, development phase**

<i>Product development</i>	<i>Product engineering tasks</i>	<i>Process development</i>	<i>Product engineering tasks</i>
Identify and study best market niche to develop		Identify and study best process to develop	×
Generate alternate product concepts and property requirements		Generate alternate process concepts, raw material, equipment	
Make and test structural modifications and formulation to improve properties	xx	Make and test manufacturing process, raw material, equipment	
Lab test samples for properties	xx	Pilot plant for efficacy	
Customer test samples for satisfaction		Make samples	
Evaluate potential	x	Evaluate potential	
Recommend commercialization	x	Recommend commercialization	×

×: product engineering involved; xx: product engineering play leading role.

generation or modification of the markets. After a number of recycles, the emergence of a good candidate may lead to a recommendation for development.

Phase 2 of development normally consists of a dual track of product development and process development. On the side of product development, the product engineers play dominant roles. The product engineers work with the marketers in product development, in fine tuning the technology to make the product more suitable for the marketplace, which is shown in table 2.2. It is necessary to have small samples for laboratory testing of physical-chemical properties, and it is necessary to have larger samples for customer testing of appreciation and satisfaction. When the results come back, there may be several iterations to generate a number of candidate products. The candidates are then compared and evaluated, leading to the recommendation of commercialization with one or more products. On the side of process development, the product engineers play supporting roles. They brief the process engineers at the beginning, on the product and the properties required, take samples from the process engineers for laboratory and customer testing, and participate in evaluating the effectiveness of process design.

Table 2.3 shows that in Phase 3 of commercialization, and in Phase 4 of business and renewal, the product engineers play supporting roles in many tasks. Product engineers return to a dominant role in creating the next generation of products. No product can last long without renewal, as there would be customer complaints, environmental concerns, and the competition may come up with a better or less-expensive product. The main task for product engineers is in keeping in touch with the outside world, and continuously improving the current product to suit changing needs in the world.

**Table 2.3 Phases 3 and 4, commercialization and business renewal phases**

Commercialization	Product engineering tasks	Business renewal	Product engineering tasks
Finance		Manufacturing quality	×
Plant location and construction	×	Sales	
Raw material	×	Transportation	
Process and equipment	×	Customer monitoring	×
Utility		Government and environmental monitoring	×
Transportation		Competition monitoring	×
Disposal, safety, environment	×	Next generation of product	xx
Marketing, sales, advertising, distribution	×		

×: product engineering involved; xx: product engineering play leading role.

## References

- Elder, A. L. 1970. *The History of Penicillin Production*. Chemical Engineering Progress Symposium Series Vol. 66, New York: American Institute of Chemical Engineers.
- Hounshell, D. A. and J. K. Smith. 1988. *Science and Corporate Strategy: DuPont R&D 1902–1980*. Cambridge: Cambridge University Press.
- Nobel e-Museum. 2002. *Nobel Lectures*. Stockholm: Nobel Foundation. <http://www.nobel.se/>.
- Suffness, M. and M. E. Wall. 1995. Discovery and Development of Taxol. In *Taxol: Science and Applications*, Suffness, M., ed., Boca Raton, FL: CRC Press.

## Further Reading

- Borman, S. 1991. Scientists Mobilize to Increase Supply of Anticancer Drug Taxol. *Chemical & Engineering News* (2 September): 11–18.
- Budavari, S. et al. eds. 1996. *The Merck Manual*. Whitehouse Station, NJ: Merck & Co.
- Chemical & Engineering News*. 2002. Facts and Figures for the Chemical Industry (June).
- Cooper, R. G. 1986. *Winning at New Products*. Reading, MA: Addison-Wesley.
- George, G. I. et al. eds. 1995. *Taxane Anticancer Agents: Basic Science and Current Status*. ACS Symposium Series 583, Washington, DC: American Chemical Society.
- Goodman, J. and V. Walsh. 2001. *The Story of Taxol: Nature and Politics in the Pursuit of an Anti-Cancer Drug*. Cambridge: Cambridge University Press.
- McGuire, W. P. 2001. Developments from a Beneficial Bark. *Science* 292: 1073–1074.
- Merck Research Laboratories. 1989. *The Merck Index*. 12th edition. Whitehouse Station, NJ: Merck & Company.
- Petousis, N. H. 2002. Anatomy of Process Research and Development. In *From Bench to Pilot Plant: Process Research in the Pharmaceutical Industry*, M. Nafissi, J. A. Ragan and K. M. DeVries, eds. Washington, DC: American Chemical Society.

- Rowinsky, E. K. and R. C. Donehower. 1991. Taxol: Twenty Years Later, the Story Unfolds. *Journal of the National Cancer Institute* 83: 1778–1781.
- Silverman, R. B. 1992. *The Organic Chemistry of Drug Design and Drug Action*. San Diego, CA: Academic Press.
- Statistical Abstract of the US. 2002. Washington, DC: US Census Bureau.
- Thayer, A. M. 2000. Busting Down a Blockbuster Drug. *Chemical & Engineering News* (6 November): 20–21.
- Urban, G. L. and J. R. Hauser. 1993. *Design and Marketing of New Products*. Englewood Cliffs, NJ: Prentice-Hall.
- Yevich, J. 1991. Drug Development: from Discovery to Marketing. In *A Textbook of Drug Design and Development*, Krogsgaard-Larsen, P. and H. Bundgaard, eds. Chur, Switzerland: Harwood Academic Publishers: 606–630.

## Discussion Topics

1. What knowledge and skills are the most valuable for investigators in the exploration–discovery phase, in the development phase, in the commercialization phase, and in the widespread use phase? Are these topics already in your curriculum? If not, then you need to take electives on your own among courses taught in your university, or you have to acquire them from a business school or a company with a good track record of developing new products?
2. Which of the knowledge and skills for product development are common to those used in process development, and which are not common? Does product development require more soft “people skill” in addition to hard “scientific skill,” in comparison with process development?
3. Which set of challenges is more exciting to you in your future career, product or process development? Can you compare the rewards and satisfactions of these two sets of challenges?

*This page intentionally left blank*

## Part II

# Molecular Structure–Property Relations

Part I gave an overall view of historic product innovation cases and the work of the product engineers in searching for information and in creating solutions to challenging problems. The discovery of CFCs was motivated by a market need, and the major search questions were: What are the properties required of a safe refrigerant for home refrigerators? Which substances or mixtures would have these properties? How do we modify and formulate the material to arrive at a satisfying product to suit the market needs? What are the best ways to make them? The discovery of nylon was motivated by the invention of the technology of condensation polymerization, and the major search questions were: What classes of materials can be made with this technology? What are the interesting properties possessed by these materials? What is the potential marketplace demand for the products that can be thus fabricated?

Part II deals with the subject of molecular structure–property relations. It addresses many of the search problems raised in the development of new products, and forms the intellectual core of the science behind product engineering.

*This page intentionally left blank*

## Search Challenges and Methods

### **3.1 Forward Search from Material to Property**

- Literature search
- Prediction from theory
- Prediction by correlation
- Estimation by associations and trends
- Synthesis, collection, measurements

### **3.2 Reverse Search from Properties to Material**

- Literature and database search

Forward searches plus interpolation  
Random Searches

### **3.3 Searches from Technology to Markets**

### **3.4 Literature and Databases**

- Printed database
- Electronic databases

#### **References**

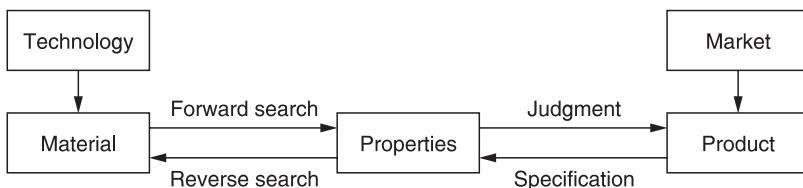
#### **Further Reading**

#### **Exercises**

A successful product must have the following elements: a market and customers with needs for a product that is available in quantity and at suitable prices, with a set of properties that are suitable for the application, containing appropriate material that can be produced by a suitable technology (figure 3.1).

The design of a successful product for the marketplace can be described as the creative synthesis of many elements together, with optimization and harmony. The product innovators usually start with some of the required elements, but other key elements are missing and have to be found: for example, when we needed a good refrigerant but did not know what material would have the right properties, or when we found an interesting nonstick polymer but did not know what products would benefit from it and what markets would appreciate it. At later stages, we need to find ways to optimize the elements to make a better product, such as by learning to make nylon from the raw materials adipic acid and hexamethylenediamine, and by increasing the solubility of taxol by emulsifying with castor oil. Thus, a successful design involves many searches for missing elements, as well as for ways to improve existing elements.

The search from a material to its properties is called the forward search, since handbooks and tables of properties are organized and listed by the materials, so one



**Figure 3.1** The design process for bringing a successful product to the marketplace

looks up the boiling points and the flammability of the compounds by their names. The invention of nylon by Carothers and associates at DuPont in 1928 can be represented by the set of arrows from the left to the right in figure 3.1. Carothers discovered the science and technology of condensation polymerization, which has the capability to make polyester material in possession of properties that are promising but not entirely suitable, as well as the capability to make many other materials. They decided that the product should be a silken fiber, suitable for the market of ladies stockings. From this “lead compound,” they searched to find ways to modify the structure to obtain material with superior and desirable properties that were fine-tuned for the stocking market.

The search from properties to a material that has them is called the reverse search, since handbooks are not organized this way. The search for a safe refrigerant by Midgley and associates in 1928, which resulted in the discovery of CFCs, can be represented by the set of arrows from the right to the left in figure 3.1. Kettering told Midgley that there was a market demand for a safe refrigerant and left it to Midgley to specify quantitatively all the parameters of the properties required for this product. Midgley decided that it should have a boiling point of  $-30$  to  $0$   $^{\circ}\text{C}$ , and should be nonflammable and nontoxic. The next challenge was to search for substances that had these properties, to make modifications for greater convenience in their use, and to find methods to make them on a large scale at a competitive cost.

The most important product design parameters can be grouped into three sets:

1. *Structure parameters.* For a single compound, the structure parameters include: the proportion of atoms and their connectivity, the geometric and energetic parameters of bonds, angles, and conformation, and the electronic parameters of electron distribution and polarization. For multicomponent systems of solutions, microstructural material, and composite material, the additional structure parameters include: the proportion of the various components, and the relations of their phases as solutions, colloids, or composite solids.
2. *Property parameters.* The physical property parameters include: state of matter, phase equilibrium, thermal, mechanical, optical, and electromagnetic properties. The chemical property parameters include: preparation, reactivity, reactants and products, kinetics, flash point, and explosion limit. The biological property parameters include: toxicity, physiological and pharmaceutical effects, nutrition value, odor, and taste.
3. *Market and environment parameters.* The market parameters include: application areas, sales volume, price, and growth potential. The environment parameters include: environmental dispersion, transportation and transformation, areas of concentration, safety, health, and environmental impact.

### 3.1 Forward Search from Material to Property

The forward search starts from a newly discovered or currently underutilized technology that can be used to make a variety of materials, or can serve as a starting point for improved technologies. The investigators need to discover what material can be made by this technology, the properties of these materials, and what useful products can be designed and made from them. Let us represent the material structure by a vector  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ , which stands for parameter values such as bond length, bond angle, conformation, dipole moments, etcetera. Let us represent the properties by the vector  $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$ , which represents a set of properties such as boiling points and densities. The task of forward search is to find the function  $\mathbf{y} = f(\mathbf{x})$ . The “domain” of  $\mathbf{x}$  is the set of all the material structures of interest, and the “range” of  $\mathbf{y}$  is the set of all the property values that these materials can assume. Sometimes the forward search function takes the form of  $\mathbf{y} = f(\mathbf{y}')$  where  $\mathbf{y}'$  is a set of properties that are more readily available than  $\mathbf{y}$ , such as using the boiling points at 1 atm to predict the critical temperature. Some of the forward search questions that we seek answers to include:

1. Find the structure parameters  $\mathbf{x}$  and the properties  $\mathbf{y}$  of a given compound at room temperature and pressure. What are the properties of the compound at elevated or lowered temperatures and pressures, as well as of other environmental variables, such as electromagnetic field and solar radiation, so that this search can be written as  $\mathbf{y} = f(\mathbf{x}, T, P)$ ?
2. When the lead compound is subjected to various chemical and physical modifications, what are the various families of derivative materials that can be made with this technology, where the derived structure can be written as  $\mathbf{x} + \Delta\mathbf{x}$ ?
3. What are the properties of these derivative materials,  $\mathbf{y} + \Delta\mathbf{y}$ ? The derivative  $\Delta\mathbf{y}/\Delta\mathbf{x}$  is of particular interest, as it represents the property modifications resulting from structure modifications. This knowledge would be useful in exploring ways to improve product properties.

The study of the function  $\mathbf{y} = f(\mathbf{x})$  starts from measurements of the properties of many compounds, and subsequent compilations into tables and databases. An analysis of these empirical observations can lead to useful associations and trends, and generalizations that may have predictive power. When a sufficiently large and systematic database has been accumulated, researchers will try to find correlations between a property  $\mathbf{y}$  and “predictors,” which are parameters relating to molecular structure  $\mathbf{x}$ , or other more easily available properties,  $\mathbf{y}'$ .

Product engineers need to be familiar with the search methodologies outlined in the following sections.

#### Literature Search

The first step in the forward search is to see whether we can find the desired property of the material in a database, which may be a textbook, a handbook, a research journal, or, increasingly, an electronic resource. This method is, in principle, the fastest and least expensive. Properties of materials are measured and first published in research

journals, after having passed a review process. Subsequently, these primary publications are further reviewed and placed in secondary publications of handbooks and textbooks. This system of vetting is designed to weed out unsubstantiated claims, and to reconcile the inconsistencies of different authors. Some databases, such as the Beilstein, elect to list all the measured values in the research literature, instead of a single recommended value for the boiling point of a compound. There are also many printed and electronic sources that have not been vetted by independent reviewers, and so should be used with caution.

The forward search starts from the name of a chemical compound, proceeds to finding its molecular structure, and then its physical and chemical properties, such as the boiling point, melting point, density, etcetera, in a handbook. Many databases for single compounds are also organized by classes and families of similar structures. Fluid solutions represent the next level of complexity. For the most important fluids, such as water, air, and some refrigerants, we can find extensive tables for the thermal properties of mixtures. For complex fluids, such as paint and emulsion, which are difficult to characterize and to reproduce, specialized books and journals should be consulted. The properties of some crystalline solids can be found, but usually not for multicrystal composite and amorphous solids.

Biological and environmental properties are generally not systematically measured and tabulated, until the recent compilation activities sponsored by the U.S. EPA, and by the Occupational Safety and Health Administration (OSHA). Properties such as prices and concentrations regarded as safe are changeable and can only be found in newspapers and journals, such as the *New York Times* and *Chemical Marketing Reporter*.

### Prediction from Theory

When the desired properties are not found in available databases, the next move might be to see whether they can be estimated from theory. This is becoming easier due to the advances of the last 80 years in computational chemistry and the availability of fast computers and of improved software. It is possible to calculate the structure of a small molecule from theory, which yields a wealth of structural information on bond angles and energies, on charges and dipole moments, etcetera. There are also a few physical properties that are so thoroughly understood that they can be computed from theory, including the translational and rotational heat capacities of diatomic molecules. There is a much larger class of molecular properties that can be computed by molecular modeling of the behavior of assemblies of molecules.

The computational methods for the structure of a molecule are divided into: the ab initio, the semi-empirical, and the molecular mechanics methods.

#### *The Ab Initio Method*

The ab initio method begins by solving the Schrödinger equation for the orbitals of electrons around a molecule, using as little simplification and approximation as is practical. This exact method is available only for small molecules with few atoms. The Schrödinger equation for a system with only one nucleus and a single electron

is given by

$$H\psi = E\psi$$

$$H = -\frac{\hbar^2}{8\pi^2m} \left( \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \right) - \frac{Ze^2}{r}\psi$$

The solutions are the wave functions  $\psi$ , called the atomic orbitals. The molecules are built from the atomic orbitals, and the most stable configuration is the one with the lowest energy.

### *Semi-empirical Methods*

The semi-empirical methods ignore all the core electrons and consider only the outer shell of valence electrons. Many other types of simplification are also made to make the computation feasible in a reasonable length of time. These simplifications make the results less accurate, but make it possible to study larger molecules.

### *Molecular Mechanics*

Molecular mechanics is the modeling of the molecule not as electrons obeying quantum mechanics, but as a number of atoms connected together by flexible bonds, which move according to Newtonian mechanics and empirical “force fields.” For instance, in a simple diatomic molecule like  $H_2$ , the length of the bond between the two atoms is assumed to have an energy equal to

$$E_{\text{bonds}} = \frac{k}{2}(l - l_0)^2$$

This describes a parabolic energy well, and the lowest energy is attained for the normal length of the bond at  $l_0$ , which is 0.74 Å. The energy of the molecule is the sum of all these bond length, bond angle, bond torsion, dipole, and van der Waals energies. The most stable configuration of the molecule is that one with the lowest energy. The parameter values of these force fields are taken from quantum mechanical calculations of similar molecules, and from observations such as spectroscopy.

### *Prediction by Correlation*

Quantitative correlations can be constructed when there is a sufficiently large and systematic database of the properties  $y$  and of the predictors  $x$ . The domain of the correlation is the set of material with a similar structure, where a correlation can be expected to be reasonable correct, such as the boiling points for the normal paraffin. The predictor can be a discrete variable, such as the number of carbons on an alkane and the group position of an element; it can also be a continuous variable, such as temperature, pressure, and mole percent in a solution. There may be only one or several predictors in a correlation, and the function can be linear or nonlinear.

The most reliable choice for the  $x$  parameters and for the functions is based on theoretical understanding of the cause and effect between the parameter and the property, and it is more risky to use parameters and functions based on experience without theoretical justifications. When there is no appropriate theory to rely on, the simplest method is a linear function with only one predictor, such as  $y = c_0 + c_1x$ , or a nonlinear function such as  $y = c_0 \exp(-c_1x)$ . When there are several predictors, the simplest function is a multilinear combination  $y = c_0 + c_1x_1 + c_2x_2 + \dots$ . It is more reliable to interpolate within the domain with experimental data than it is to extrapolate beyond the domain.

The group contribution method (GCM) begins with the observation that a functional group such as Br<sup>-</sup> increases the boiling points of many organic compounds much more than the group Cl<sup>-</sup>, which in turn has a greater effect than the group F<sup>-</sup>. The GCM decomposes a molecule into a number of “groups,” such as H<sup>-</sup>, K<sup>-</sup>, CH<sub>3</sub><sup>-</sup>, NH<sub>2</sub><sup>-</sup>, OH<sup>-</sup>, Cl<sup>-</sup>, C<sub>6</sub>H<sub>5</sub><sup>-</sup>, etcetera. The property of a molecule, such as the boiling point, is assumed to be the sum of the contributions of each type of group, so that  $T_b = 198 + \sum \Delta_b$ . The method of quantitative structure-activity relations (QSAR) can be considered a branch of this effort, except that the predictors can be structural parameters such as the connectivity among the atoms.

### Estimation by Associations and Trends

Situations are often encountered where there are some fragmentary experimental observations to provide for suggestions and speculations, but not enough for a systematic and quantitative study. For instance, we have the toxicity information for many polychlorinated compounds that are used as pesticides and for electrical transformers, but many more species have never been synthesized or measured; their properties are not easily organized into a series with toxicity increasing with the number of chlorine atoms. There are also many situations where the diverse structural variations do not lend themselves to quantitative treatment. For instance, there are seven isomers of the heptane molecule: one has a straight seven-carbon backbone, three have a single side branch, four have two side branches, and one has three side branches. It is a challenge to create a quantitative formulation that can be generalized to any number of carbon atoms on the paraffin chain.

One tries to study and discover patterns of behavior in a domain where data are available, to form hypotheses of associations and trends, and to export them to a larger domain where there are no data. For instance, the water solubility at 25 °C for straight-chain hydrocarbons with six carbons is in the order



We can make the hypothesis that this ranking is also valid for a paraffin with any number of carbon atoms, and the even more bold hypothesis that this ranking is valid for any organic compound. This hypothesis might suggest fruitful areas that should be investigated further.

## Synthesis, Collection, Measurements

After all the theoretical and empirical estimation work is done, the only way to make sure that a substance actually has the property estimated is to synthesize the material and to make experimental measurements. This method is much more costly in time and money, and should be undertaken after a great deal of library and theoretical searches have already been done. This process of synthesis and measurements also has the effect of enlarging the database and improving understanding in the domain relevant to this development purpose.

### 3.2 Reverse Search from Properties to Material

The motivation of the reverse search is often to serve a market need that is not satisfied by currently available products. The investigators begin with a study of the marketplace and the business opportunities in introducing a new product by talking to the customers, equipment builders, government regulatory agencies, and experts. Then a hypothetical material is proposed with a given set of required properties, and the problem is where to find such a material.

The starting point of a reverse search is often from a lead product that is currently on the market but in need of improvements, with a given set of properties that should be modified. The search questions include:

1. What physical and chemical modifications of the lead compound have the best potential to improve its properties in the desired directions  $\Delta y$ ? The modified properties can be written as  $y + \Delta y$ , and the modified structure can be written as  $x + \Delta x$ . The derivative  $\Delta x/\Delta y$  represents the structure change required to achieve a particular property change.
2. What other types of material would have a set of properties in the range of the desired properties? For instance, which class of compounds smells even better than musk, and where would we find a cure of lung cancer? This is symbolically written as  $x = g(y)$ .

The reverse search is the most often encountered search question in product engineering, but there are few organized and convenient search engines.

## Literature and Database Search

The printed databases are usually not designed and organized for a reverse search, as tables of densities and boiling points are normally listed according to the types and names of compounds. It is a laborious task to go through a handbook to find all the metal elements that melt between 400 and 500 °C. Many electronic databases, such as CD-ROM and Web sources, have a “search” capability, and this permits a search from a set of properties to a list of the substances that have them. For a single property such as boiling point, the database can sort according to this variable and display all the substances from the lowest to the highest boiling points, which makes it easy to find all compounds that boil between –40 and –39 °C. For several variables, this search would take the form of a Boolean search with “AND” and “OR” between criteria,

such as “boiling point between  $-40$  and  $-39$ ” AND “density between  $1.1$  and  $1.2$ .” For the currently available databases, the reverse search capability is available only for some of the most common physical parameters, such as boiling points and densities, and not available for the less common parameters such as toxicity and global warming potential.

### Forward Searches Plus Interpolation

For most properties that can be predicted from theory, the usual formulation is the forward search from structure to property in the form  $y = f(x)$ . However, a planned series of forward searches can be used to simulate a reverse search. For instance, in response to the need for a reverse search for the sweetest substances we restructure the search by compiling the sweetness of existing substances by their structures, which would facilitate a number of forward searches from structure to sweetness, and then use the structures of the sweetest substances to suggest areas for future explorations. In figure 3.2, three structures  $x_1$ ,  $x_2$ , and  $x_3$  are mapped to three property values  $y_1$ ,  $y_2$ , and  $y_3$ . When the desired property  $y^0$  is contained in a triangle of the  $y$  space, it may be surmised that the relevant structure  $x^0$  is also contained in the corresponding triangle in the  $x$  space. In fact, if  $y^0$  can be given as a linear combination of  $y_1$ ,  $y_2$ , and  $y_3$ , in the form of  $y^0 = c_1y_1 + c_2y_2 + c_3y_3$ , then one can estimate that the desired structure  $x^0$  can be approximated by the linear combination of  $x^0 = c_1x_1 + c_2x_2 + c_3x_3$ .

Thus, we execute three forward searches plus interpolation, which has the effect of a reverse search.

An association or a trend can also be used in the forward or the reverse directions. The Midgley search for a safe refrigerant relied on the qualitative trend that, among the set of simple compounds, the compound tends to be less flammable when one moves from the elements of group 14 to group 17, and they tend to be less toxic when one moves from the elements of period 5 to period 2 in the periodic table. Analogy is another method often used. If one makes the empirical observation that a reduction of bond saturation by removing hydrogen increases the water solubility of a hydrocarbon with six carbon atoms, then one may also try to decrease bond saturation of another

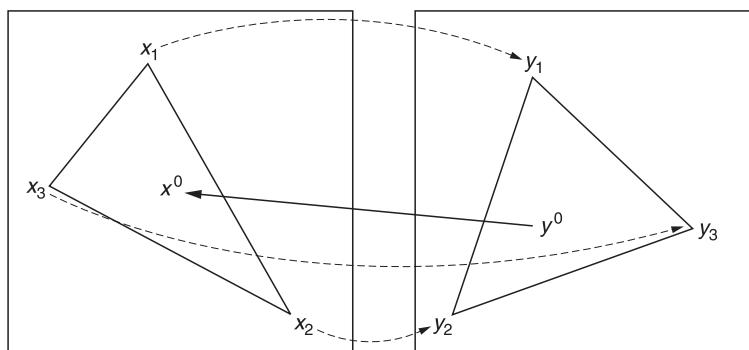


Figure 3.2 Mapping structures to property values

organic compound in order to increase water solubility. Associations and trends are sources of suggestions on fruitful areas that deserve further investigations.

### Random Searches

The random search is usually the last resort when there are no good leads on where to start and what the fruitful directions to explore are. This approach is costly in time and money, and success is not certain, so that a strong motivation and a good budget are required. When a compound with unusual properties is discovered, such as when the morphine structure was first identified, synthetic chemists descended with gusto to make many derivatives, which is to say the creation of a set of random  $\Delta x$ , so that nearly all possible derivatives were made and tested for the property changes  $\Delta y$ . Paul Ehrlich had a set of arsenic compounds that had therapeutic properties for syphilis and unacceptable toxic properties, so he searched for better arsenic compounds by changing their structures more or less by random. Thomas Edison had a carbon filament that worked reasonably well as a lamp filament, and he tried many other sources of carbon for his incandescent lamp. These are incremental random searches, by making small perturbations from a promising starting point, with a high probability of finding something interesting.

A search is much more speculative when bold leaps and “wildcat” searches in uncharted territories are undertaken. The initiative may be based on rumors of uncertain validity or an inspired hunch. There are no solid reasons to believe that the target can be found in this area; thus, there are no explorers and one can purchase property rights at very low costs. Success is unlikely, but when the improbable takes place we would be crowned with glory and profit.

### 3.3 Searches from Technology to Markets

When we have a powerful technology that can make a material with interesting properties, and we can keep out competition for a while with unique technology and patents, we ask which market will appreciate a material with these properties. An ideal market is also one where the potential buyers are numerous and affluent, where this material will help them to fulfill one or more significant needs, where there are no very satisfactory competitive products, where there are no threats to safety and the environment, and where it can be manufactured and make a good profit. What is the search engine that can systematically discover many or all of the major potential markets? Let us look at the historic cases mentioned in Chapters 1 and 2.

There were a number of adaptation cases that can be classified as fairly obvious, as the inventors and their management had experience with the market to be served. The concept of adapting nylon to women's hosiery was relatively obvious for DuPont, as they had been involved for many years in the manufacturing and marketing of rayon used as textile fibers. The high price and small quantity needed for a pair of hose also fits in well for a novel technology that has never been mass produced in the past. The concept of using penicillin for antibiotics would also be relatively obvious to Arthur Fleming, as he worked in a hospital and was in the army medical services.

The adaptation of CFCs for air conditioning to reduce moisture and temperature would not require a large leap in creativity, although the adaptation of CFCs for blowing air and polymer foam, for cleaning computers, and for fire extinguishers is less obvious. William Perkin had no experience with the dye industry, so it is to his credit that he recognized the potential market for mauve. The gluing of particles to surfaces to make mending tapes and memory recording surfaces are in the borderline area of being less obvious.

There are many more examples of new technologies that required many years of wandering in the wilderness before they found appropriate markets. People would exclaim in amazement at the discovery that the deadly poison botulin could be used to erase skin wrinkles and be reborn as botox. Roy Plunkett could not have anticipated the many uses of Teflon used for diffusion barriers in making the atomic bomb; the DuPont company could not anticipate the nonstick frying pans, nor the Gor-Tex sports garments. Spencer Silver could not find a use for his weak glue until his friend Arthur Fry used it to keep his place in conducting music. We have many powerful and wonderful technologies waiting for the discovery of a suitable market, such as buckminsterfullerene and electrically conducting polymers. At this moment, we do not have systematic and effective search engines from technology and material to marketplaces.

### 3.4 Literature and Databases

When we need the properties of a chemical compound, such as the boiling point of benzene, the fastest and least expensive method is a forward search from the structure to the properties by consulting a database. The sources of such experimental information are first published in primary research journals, and then pass through professional editors and panels to make their way to secondary textbooks and handbooks.

#### Printed Database

These are the traditional passive databases of books, handbooks, reference books, journals, and catalogs. They have passed the inspection of editors of committees, and are generally reliable and accurate.

#### *Books, Handbooks, and References*

A general reference often consulted today for the physical and chemical properties of common chemicals is *Lange's Handbook of Chemistry* (Dean 1999), which lists many chemical compounds and their most important properties. It is organized into separate chapters of "Physical constants of organic molecules" with 4300 compounds and "Physical constants of inorganic molecules," and lists each compound alphabetically by name. Some of these properties are very sensitive to temperature, but less sensitive to pressure, and they are listed as tables, or more compactly as equations of the form  $f(T)$ ; for example, liquid heats of evaporation, heat capacities of multi-atom gases, vapor pressures over liquids, liquid and solid solubilities in liquids, and liquid viscosities. Some of these properties are sensitive both to temperature and pressure,

and are listed as tables or as equations of the form  $f(T, P)$ ; for example, gas densities and gas solubilities in water. There is much less data available for mixtures, except for the most important ones, as the variations and combinations are very large.

Other important handbooks include *Chemical Properties Handbook* (Yaws 1999) and the *CRC Handbook of Chemistry and Physics* (Weast and Lide 1989); *Tables of Physical and Chemical Constants* (Kaye and Laby 1986) is a more compact handbook of physical and chemical data. One should be on guard that sometimes, when experimental results are not available, the editors may list estimated values in a handbook, which are of less certain accuracy. A printed handbook normally has only limited reverse search capability, of going from a set of properties to the structures that have these properties.

Specialized properties that are not covered in these standard databases can be found in specialized books. The properties of food can be found in *Physical Properties of Foods* (Peleg and Bagley 1983). The properties of many petroleum products can be found in *Petroleum Products Handbook* (Guthrie 1960). The *Merck Index* (1996) lists chemicals, drugs, and biologicals.

The toxic properties of chemicals can be found in a reference such as *Hazardous Chemicals Desk Reference* (Sax and Lewis 1987), which lists a number of chemicals alphabetically by name. For instance, “*n*-butane” is classified (by the Department of Transportation) as a flammable gas, is moderately toxic via inhalation, causes drowsiness, is an asphyxiant, poses very dangerous fire hazard when exposed to heat, flame, or oxidizers, and is highly explosive when exposed to flame. Information about safety and environment is also provided and updated by websites maintained by government agencies, such as the U.S. EPA, the National Institute of Occupational Safety and Health, and the National Safety Data Sheet.

A very handy sourcebook for data on social, economical, and political information is the *Statistical Abstract of the United States*, which is published by the U.S. Department of Commerce through the Census Bureau. It contains information on: population, vital statistics, health and nutrition; education, law enforcement, geography and environment, parks and travel; elections, state and local government finances, federal government finances; national defense, social insurance; labor force and employment, income and wealth, prices; banking and finance, business enterprises; communications and information technology, energy, science and technology, transportation by land, air and water; agriculture, natural resources, construction and housing, manufacturing; domestic trade and services, foreign commerce; comparative international statistics.

### *Journals and Catalogs*

The research journals are the primary place where research results are first published. The prices of compounds change rapidly, and are to be found in journals instead of books. The *New York Times* finance page lists a number of commonly traded metals (aluminum, antimony, copper, gold, iron, lead, mercury, platinum, silver, zinc), food (corn, soya, wheat, rice, sugar), fuels (fuel oil, gasoline, natural gas), and textile (cotton, wool). The *Chemical Market Report* is a weekly journal that lists the prices of many chemical substances in a supplement, which depend strongly on purity and intended use.

The current prices of fine chemicals can be found in the Sigma-Aldrich *Handbook of Fine Chemicals and Laboratory Equipment* catalog, and in Fluka's *Laboratory Chemicals and Analytical Reagents*.

### Electronic Databases

These are the modern active databases that have interactive capabilities, and are often searchable for a set of properties. The floppy disks and CD-ROMs have definite dates on them, and do become obsolete with time; the Internet databases can, in principle, be updated at very frequent intervals. A note of caution about Internet databases is that some of them have not been vetted for accuracy and can be unreliable.

### *Spreadsheets and Databases*

A simple spreadsheet, such as Microsoft Excel, can serve as the foundation of a database that has forward and reverse search capabilities. For instance, a table of normal alkanes, together with their densities, boiling points, and melting points, can serve as the starting point. If we want to know all the “normal paraffins that boil between 0 and 40 °C,” all we have to do is to do a “sort” operation on the boiling-point column and obtain the result that the only paraffin that is in the range is normal heptane with a boiling point of 36.1 °C. For the more advanced Boolean search of normal alkanes that “boil between 0 and 40 °C” AND “melt between –40 and 0 °C,” it would be a far more laborious task in a spreadsheet.

A database such as Microsoft Access would be able to do this Boolean search with ease. The Filter by Form, the Filter by Selection, and the Advanced filter/sort function allow the user to specify: “density >0.6 AND <0.7” or “bp >–40 AND <0.” This search yields three results:

Compound	B.p. (°C)	Density (g/mL)
Cyclopropane	–33	0.619
Methylamine	–6.5	0.655
Propylene	–48	0.691

### *CD-ROMs and Floppy Disks*

Many publishers make available their databases on floppy disks or CD-ROMs, which are often found inside the back cover of the book. An example is the *Properties of Organic Compounds* (Lide and Milne 1999), which is a CD-ROM list of 27,500 known organic compounds and their formulas, molecular weights, melting points, boiling points, densities, refractive indexes, colors, solubility scales, vapor pressures, and threshold limit values (TLVs) that workers can be exposed to in a normal 8 h workday and 40 h workweek. This reference gives only a limited number of properties, and does not go into flammability or toxicity. Solubility in water is given only as a

qualitative scale: 1 = insoluble; 2 = slightly soluble; 3 = soluble; 4 = very soluble; 5 = miscible; 6 = decomposes. It can be searched by the following variables:

Name	Variable
MF	Molecular formula
CASRN	Chemical Abstract Service Registered Number
BRN	Beilstein Registered Number
MERCK	Merck Number
BP	Boiling point, °C
MP	Melting point, °C
DENS	Density in g/cm <sup>3</sup>
MW	Molecular weight
IR	Infrared peaks, wavenumbers
UV	Ultraviolet peaks, nm
Raman	Raman peaks, cm <sup>-1</sup>
HNMR	Hydrogen NMR shift, δ ppm
CNMR	Carbon chemical shift, δ ppm
MS	Mass spectrum, the most abundant and the parent peaks

One can also do a reverse search, such as for all compounds that boil between  $-40$  and  $0$  °C, to speed up the search a great deal. Each of these variables can be searched according to the following criteria:

- Equal to (*value*)
- Less than (*value*)
- Greater than (*value*)
- Between (*lower limit*) and (*upper limit*)

If Thomas Midgley had today's tools in his search for a safe refrigerant, then he might start his search for all the known compounds that satisfy the conditions

$$[0 > \text{bp} > -40 \text{ } ^\circ\text{C}] \text{ at 1 atm}$$

This would have results in 77 hits: from 1,2-[propadiene, 1,1,2,2-tetrafluoro-] with a boiling point of  $-38$  °C, to [propane, 1,1,1,2,3,3-heptafluoro-2-(trifluoromethyl)-] with a boiling point of  $0$  °C. Out of 77 hits, 49 of them contain elements that include B, Si, N, P, As, O, S, Cl, Br, and I, and perhaps too toxic to be considered as refrigerants seriously. There are 11 hits that are hydrocarbons, such as butane, which would be too flammable to be considered. Perhaps we would eliminate the six hits that have double or triple bonds, as they tend to be less stable and could polymerize. The remaining ones are all hydrofluorocarbons (HFCs) without chlorine, and the prime candidates are C<sub>2</sub>H<sub>2</sub>F<sub>4</sub>, C<sub>3</sub>H<sub>3</sub>F<sub>5</sub>, and C<sub>4</sub>F<sub>10</sub>.

*Internet Databases*

A most convenient Internet source of chemical data is the *Chemistry WebBook* of the National Institute of Science and Technology (NIST; <http://webbook.nist.gov/chemistry>). Finding data in the *WebBook* can be done by one of the 14 types of direct search for chemical species:

- formula search
- chemical name search
- CAS registry number search
- ionization energy search
- electron affinity search
- acidity search
- appearance energy product
- vibration energy search
- electronic energy level search
- applet-based structure search
- file-based structure search
- structure class search
- molecular weight search
- author search.

The types of data available include:

- gas-phase thermochemistry data (enthalpy of formation, molar entropy, ideal gas heat capacity)
- condensed-phase thermochemistry (liquid and solid enthalpy of formation, molar entropy, heat capacity)
- phase change (normal boiling point, normal fusion point, critical temperature, enthalpies of vaporization, fusion, and sublimation, entropies of vaporization, fusion, and sublimation, Antoine equation vapor pressure parameters)
- reaction thermochemistry (enthalpy of combustion, properties of specific reactions)
- gas-phase ion energetics (enthalpy, entropy and Gibbs free energy of reaction)
- ion clustering
- gas-phase IR spectra
- mass spectra
- UV-visible spectra
- vibrational and electronic spectra
- constants of diatomic molecules (vibration and rotation parameters)
- Henry's law data (solubility of gases in water).

The data can also be displayed as a spreadsheet or data table, or as a graphical display, if the browser is Java capable. In addition, much more extensive information is available for selected fluids:

water; He, Ne, Ar, Kr, Xe; H<sub>2</sub>, D<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, F<sub>2</sub>, CO, CO<sub>2</sub>; NH<sub>3</sub>, NF<sub>3</sub>; CH<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>3</sub>H<sub>8</sub>, C<sub>3</sub>H<sub>6</sub>, butane, isobutane, pentane, hexane, heptane; CH<sub>2</sub>F<sub>2</sub>, CHClF<sub>2</sub>, CH<sub>2</sub>F<sub>2</sub>; CHCl<sub>2</sub>-CF<sub>3</sub>, C<sub>2</sub>HF<sub>5</sub>, CF<sub>3</sub>-CH<sub>2</sub>F, R143a, R152a.

The data available includes the temperature and pressure dependence of these fluids for:

density, specific volume, heat capacity at constant pressure, heat capacity at constant volume, enthalpy, internal energy, entropy, and the speed of sound.

Beilstein is a much more comprehensive database, marketed under the name of *MDL Crossfire*, and requires a subscription fee for access. It contains far more organic substances and lists many more properties. The search commands include:

- text search for key words
- structure editor to draw structure
- tabular search by fields
- bibliographic search by citations, journals
- identification data, by name, formula
- physical data, by melting point, boiling point, pK
- reaction data, by starting material, product, yield, solvent
- solubility data, by solubility, Pow, Henry's law, solution behavior
- spectral data, NMR, ESR, IR, UV, MS.

It has extensive capability for forward search and reverse search from properties to substances. The properties available include:

Chemical data

reactions, purifications

Physical data, single component

structure and energy, bond distance, angles  
physical state

crystals—melting point, density  
liquid—boiling points  
gas—critical temperature, vapor pressure  
others—surface tension

transport—viscosity, diffusion, thermal conductivity  
thermochemical—enthalpy, heat capacity  
safety—flash point, explosive limit

Physical data, multicomponent

solution, solubility, Henry's law constant, partition water-octanol  
liquid-liquid systems, liquid-vapor, liquid-solid

Pharmacological and ecological data.

A tabular search using the Fact Editor for the conditions: [bp > -40] AND [bp < 0] yields 462 hits in 15 s. Since most of the hits contain heteroatoms other than fluorine, it would be better to reissue the search command for compounds that do not contain any nitrogen: [bp > -40] AND [bp < 0] NOT [“\*N\*”], which yields 317 hits in 156 s. The asterisk indicates the wild card that stands for any number of characters of any kind, and the “ ” is for a text string. When we refine it further to exclude chlorine

as well, we issue the search command: [bp > -40] AND [bp < 0] NOT [“\*N\*”] NOT [“\*Cl\*”], which yields 247 hits in 204 s.

At this moment, the NIST *Chemistry WebBook* and the Beilstein are the two most useful search tools for molecular structure and property relations. There are many more types of properties available in the research literature of interest only to a smaller set of readers; these will not be listed in the general-purpose handbooks, but are to be found in specialized books and journals. Our ability to search for such properties in the research literature on the Internet is increasing rapidly, and students should practice doing such searches. Many of the required properties have not been tabulated in a searchable electronic database; for example, flammability, toxicity, ozone hole potential, and greenhouse gas potential. Information on health, safety, and the environment that is publicly available on websites is listed in the references.

## References

- Dean, J. A., ed. 1999. *Lange’s Handbook of Chemistry*. 15th edition. New York: McGraw-Hill.
- Guthrie, V. B., ed. 1960. *Petroleum Products Handbook*. New York: McGraw-Hill.
- Kaye, G. W. G. and T. H. Laby. 1986. *Tables of Physical and Chemical Constants*, 16th edition. London: Longman.
- Merck Index: An Encyclopedia of Chemicals, Drugs & Biologicals*. 1996. 12th edition. Merck & Co.
- Peleg, M. and E. B. Bagley, eds. 1983. *Physical Properties of Food*. Westport, CT.: AVI Publishing Company.
- Sax, N. I. and R. J. Lewis. 1987. *Hazardous Chemicals Desk Reference*. New York: Van Nostrand Reinhold.
- Statistical Abstract of the United States*. 2003. Washington, DC: U.S. Government Printing Office.
- Weast, R. C. and D. R. Lide, eds. 1989. *CRC Handbook of Chemistry and Physics: A Ready Reference Book of Chemical and Physical Data*. 70th edition. CRC Press.
- Yaws, C. L., ed. 1999. *Chemical Properties Handbook*. New York: McGraw-Hill.

## Further Reading

### Books and Journals

#### General Database

- Damm, H. C., P. K. Besch, and A. J. Goldwyn. 1966. *The Handbook of Biochemistry and Biophysics*. Cleveland: The World Publishing Company.

#### Journals and Catalogs

- Aldrich. *Handbook of Fine Chemicals and Laboratory Equipment, 2000–2001*.
- Chemical Market Reporters*. Current prices of commodities, metals, fuels, and chemicals.
- Fluka. *Laboratory Chemicals and Analytical Agents, 1999–2000*.
- New York Times*. Financial page on the prices of commodities, metals, etcetera.

## Electronic Databases

### *Floppy Disks and CD-ROMs*

- Beilstein/CrossFire. 1999. Properties of organic chemicals. Frankfurt, Germany.  
“Cranium”, Molecular Knowledge, Bedford, New Hampshire, 1998. Group contribution methods (GCM) for estimation of range of molecular properties.  
Lide, D. R. and G. W. A. Milne, eds. 1996. *Properties of Organic Compounds*. CD-ROM, CRC Press. 27,000 compounds, b.p., m.p., density, spectroscopic data.

### *Websites*

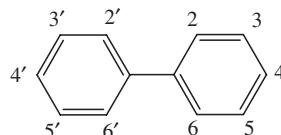
- ChemFinder Online: <http://ChemFinder.camsoft.com>  
Columbia University chemistry department:  
<http://www.columbia.edu/itc/chemistry/orgolab/cheminfo.html>  
EPA, Integrated Risk Information System: <http://www.epa.gov/iriswebp/iris/index.html>  
Material Safety Data Sheet (MSDS): [www.camd.lsu.edu/msds/jsssearch.htm](http://www.camd.lsu.edu/msds/jsssearch.htm)  
National Institute of Occupational Safety & Health (NIOSH), Pocket Guide to Chemical Hazards: [www.cdc.gov/niosh/npg/pgdstart.html](http://www.cdc.gov/niosh/npg/pgdstart.html)  
National Institute of Science and Technology (formerly National Bureau of Standards) properties of chemicals: <http://webbook.nist.gov/chemistry>.  
Occupational Safety & Health Administration (OSHA): [www.osha.gov](http://www.osha.gov)  
Statistical Abstract of the U.S.: <http://www.census.gov/statab/>.

## Exercises

This set exercises concerns research on finding relevant information, as well as on preliminary organization. Start with simple search engines, then switch to more specialized search engines, and then go to the original references when necessary. Document all the sources that you have looked into and note any information you cannot find.

1. Do a forward search from materials to properties. Find the boiling points and melting points of 1-alcohols from  $C_1$  to  $C_{20}$ .
  - (a) Make a table for these properties.
  - (b) Make a plot of these properties versus the number of carbons  $N_C$ .
  - (c) Compare this plot with the normal paraffins.
2. Make a table of the densities, boiling points, and melting points of the first 54 elements.
  - (a) How many of these elements melt below room temperature? In fact let us divide all the elements into sets with the melting point bin of  $[-273 \ 0 \ 500 \ 1000 \ 1500 \ 2000 \ 2500 \ 3000 \ 3500]$ . Report the population distribution of elements within each bin.

- (b) For these 54 elements, is there a good correlation between melting point and boiling point?
- (c) For these 54 elements, is there a good correlation between density and melting point?
3. There are two forms for butane  $C_4H_{10}$ : normal butane and isobutane. There are three forms for pentane,  $C_5H_{12}$ : pentane, 2-methylbutane, and 2,2-dimethylpropane.
- (a) How many forms are there for hexane, for heptane, and for octane?
- (b) Find all the melting and boiling points of paraffins containing from one to eight carbons, and organize them into a table according to whether they are: (i) straight chain, (ii) with one side branch (the longest chain is assumed to be the main chain), (iii) with two side branches, (iv) with three side branches, (v) with four side branches.
- (c) We observe the general pattern that the boiling and melting points do increase with the number of carbon atoms. Is there a general pattern associated with the number of side chains on the boiling point and melting point?
4. Polychlorinated biphenyls (PCBs) were once used extensively to fill electrical equipment, such as transformers and relays, but they are now banned because they are toxic when discharged to the environment. The basic biphenyl  $C_{12}H_{10}$  is two benzene rings connected by a single bridge.



The base biphenyl has 10 hydrogen atoms: four of them (2, 6, 2', 6') are near the bridge and are named the  $\alpha$  positions, four of them (3, 5, 3', 5') are away from the bridge and are named the  $\beta$  positions, and two of them (4, 4') are opposite the bridge and are named the  $\gamma$  positions. There are three forms for mono-chloro biphenyl, depending on the position of the chlorine atom, and they each have different properties. The PCBs are a set of 210 compounds with 0 to 10 chlorine atoms in a molecule. Monsanto made them under the trade name Aroclor, and they came in a number of mixtures: the light Aroclor 1221 is basically mono- and di-chloro, the medium Aroclor 1248 is principally tri-, tetra- and penta-chloro, and the heavy Aroclor 1260 is principally hexa- and hepta-chloro. We are concerned with the transportation and concentration of these chemicals, particularly in fish and humans, which are controlled by the three parameters of vapor pressure (breathe in air), solubility in water (drink water), and octanol–water partitions (accumulation in fat).

- (a) Find these relevant transportation and concentration parameters for biphenyl.

- (b) Find the parameters of the three mono-chloro biphenyls, and point out the differences among the three forms of  $\alpha$ ,  $\beta$ , and  $\gamma$ .
  - (c) How many forms of di-chlorinated biphenyls are there? Name them, and find as many of their parameter values as you can.
  - (d) How many forms of tri-chlorinated biphenyls are there? Name them, and find as many of their parameter values as you can.
  - (e) Which type of PCB seems the most dangerous, and what is the nature of the threat?
5. Do a reverse search from properties to materials. What are all the hydrocarbons with boiling points between 0 and 5 °C, and melting points below –20 °C?
6. Let us repeat the journey taken by Thomas Midgley in his discovery of CFCs, in the light of superior modern databases and search engines.
- (a) What are all the compounds that boil between 0 and 2 °C when the pressure is 760 Torr?
  - (b) Which of the compounds in (a) are nonflammable and nontoxic?
  - (c) Which ones have no chlorine to react with ozone in the stratosphere?
  - (d) Which ones are not greenhouse gases?
7. Let us imagine a dialog within the corporate headquarters of Lockheed. CEO: Our attack fighters require fuselage material of the highest melting point and the lowest density. What should we use? By the way, the Defense Secretary does not worry about price.

Chief Engineer: Carbon is the best deal and lead is the worst deal.

CEO: Do you recommend a carbon fuselage then?

Chief Engineer: There are actually other required properties.

Question for students: What is a complete set of property specifications, besides melting point and density? What procedure would you use to search for the best material, which databases would you apply the procedure to, and what is the best idea that you can come up with?

# 4

## Theory and Quantitative Predictions

### 4.1 Single Molecules

- Quantum mechanics
  - Molecular mechanics
  - Predictable properties
- 4.2 Statistical Thermodynamics**
- Ensembles of molecules
  - Molecular dynamics and Monte Carlo
  - Entropy, disorder, and stability

### 4.3 Gases, Solids, and Liquids

- Ideal gases
  - Real gases
  - Solids
  - Liquids
- References**
- Further Reading**
- Exercises**

The purpose of this chapter is to review the theories of molecular structure and property relations, to discuss computational methods for prediction of molecular structure and properties, and to discuss some of the properties that can be predicted by computations.

### 4.1 Single Molecules

Quantum mechanics is the foundation of molecular structure and properties. The position and energy of the electrons around a molecule are determined by solving the Schrödinger equation for a given set of positions of the nuclei of the atoms. There is a lot of powerful and effective computer software that can be used to calculate many of the properties of an isolated single molecule, especially at zero absolute temperature. The starting point is the construction of the sketch of a molecule by connecting atoms with the appropriate bonds. This qualitative sketch does not need accurate values for the bond lengths and angles. To set up the computation, the investigator specifies one of three computation methods: *ab initio*, semi-empirical, or molecular mechanics. The first and second methods are based on quantum mechanics about a model of the

molecule as a number of negatively charged electrons surrounding a collection of positively charged nuclei. The third option of molecular mechanics is based on classical Newtonian mechanics about a model of the molecule as a number of mechanical bonds linking the atoms together, and these bonds can be stretched and bent according to empirical force fields.

When either the Schrödinger equation or the Newtonian equation is solved with the initial spatial distribution of nuclei, in what is called the single-point determination, the binding energy of the molecule is obtained. If we make random perturbations of the positions of the various atoms, and repeat the single-point calculations, we can map the energy levels of the molecule in a neighborhood. The most stable or equilibrium position of the molecule is the one with the lowest energy in the neighborhood, and the search for this equilibrium position of the atoms is called geometry optimization.

## Quantum Mechanics

The most rigorous and accurate method of calculation is the *ab initio* method, which is also the most demanding in computational time and resources, so that it is most often used for smaller molecules.

### *Single-electron Nucleus*

Let us begin by considering a nucleus with a single electron, such as the hydrogen atom, or the singly charged helium ion, or the doubly charged beryllium ion. The behavior of this electron in space and time is described by the Schrödinger equation, which is a partial differential equation of the form

$$H\Psi(x, y, z) = E\Psi(x, y, z)$$

where  $H$  is the Hamiltonian operator of the kinetic and potential energies of the molecule between a single nucleus and an electron,  $\psi$  is a solution to the equation, which is called the atomic orbital of an electron, and  $E$  is the energy of the electron. The simplest case is a hydrogen atom where there is only one electron:

$$\begin{aligned} H\Psi &= \left[ -\frac{\hbar^2}{8\pi^2m} \nabla^2 + V(x, y, z) \right] \Psi \\ &= -\frac{\hbar^2}{8\pi^2m} \left( \frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2} \right) - \frac{e^2}{r} \Psi \end{aligned} \quad (4.1)$$

where  $\hbar$  is the Plank's constant ( $6.63 \times 10^{-34}$  J s),  $m$  is the mass of the atomic nucleus,  $r$  is the distance between the electron and the nucleus, and  $e$  is the charge on the electron. The first term in the equation represents the kinetic energy of the electron, and the second term represents the potential energy of the electron, which is determined by the electrostatic force between the electron and the nucleus.

There are many solutions to the Schrödinger equation, and each solution is called an atomic orbital  $\Psi$ , with an energy  $E$ , and has a spatial distribution characterized by four quantum numbers:

The principal quantum number  $n = 1, 2, 3, \dots$

The orbital shape quantum number  $l = 0, 1, 2, \dots, n - 1$

The orbital orientation quantum number  $m = 0, \pm 1, \pm 2, \pm 3, \dots, \pm l$

The spin quantum number  $s = +1/2, -1/2$

The principal quantum number  $n$  is the most important determinant of the radius and energy of the electron atomic orbital. The orbital shape quantum number  $l$  determines the shape of the atomic orbital. When  $l = 1$ , the atomic orbital is called an s orbital; there are two s orbitals for each value of  $n$ , and they are spherically symmetric in space around the nucleus. When  $l = 2$ , the orbitals are called the p orbitals; there are six p orbitals, and they have a dumbbell shape of two lobes that are diametrically opposed. When  $l = 3$  and 4, we have 10 d orbitals and 14 f orbitals. The orbital orientation quantum number  $m$  controls the orientation of the orbitals. For the simplest system of a single electron in a hydrogen atom, the most stable wave function 1s has the following form:

$$\Psi(1s) = \frac{1}{\sqrt{\pi}} \left( \frac{1}{a_0} \right)^{3/2} \exp(-r/a_0) \quad (4.2)$$

For an atom or ion with a nuclear charge of  $Z$ , the radius of maximum probability and the energy of the single electron depend only on the principal quantum number  $n$ :

$$r_n \cong \frac{n^2}{Z} 0.529 \text{ \AA}$$

$$E_n \cong -\frac{Z^2}{n^2} 13.6 \text{ eV}$$

$E$  is the first ionization energy required to remove the one electron from the nucleus to infinite distance. The energy increases when the single electron of hydrogen is promoted from the base level of  $n = 1$  to higher orbitals, so there is a very strong tendency for the electron to fall back to the most stable position of  $n = 1$ .

For the hydrogen atom, the energy value at the lowest state 1 is  $-2.178 \times 10^{-18}$  J, and alternately as  $-1311.2$  kJ/mol when multiplied by Avogadro's number, in wavenumbers as  $109.94$  cm $^{-1}$ , and also as  $13.59$  eV. The energy of the electron is proportional to  $1/n^2$ , represented in table 4.1.

The energy gap to promote an electron from  $n = 1$  to  $n = 2$  is  $10.19$  eV, or  $983.4$  kJ/mol.

**Table 4.1 Variation of electron energy with *n***

<i>n</i>	$J \times 10^{18}$	$kJ/mol$	$I/cm$	$eV$
$\infty$	0	0	0	0
5	-0.087	-52.4	-4.39	0.54
4	-0.136	-81.9	-6.87	0.85
3	-0.242	-145.7	-12.22	1.51
2	-0.545	-327.8	-27.48	3.40
1	-2.178	-1311.2	-109.94	13.59

### Multi-electron Atoms

The helium atom has two electrons, so that the potential energy is the sum of two attraction energies between the single nucleus and the two electrons, and the repulsion energy between the two electrons. Each electron is also shielded from the nucleus by the other electron, and feels less than the full attractive forces of  $Z = 2$  charges from the nucleus. There is no precise solution of this “three-body problem,” involving one positively charged nucleus and two negatively charged electrons. The effect of shielding can be approximated by introducing an effective charge  $Z_e$ , which is smaller than  $Z$  due to shielding, and would result in a larger radius and a smaller energy well. The appropriate value for the effective charge is usually obtained from empirical data of X-ray crystal diffraction and spectroscopy for energy transition, instead of theory. There is more shielding for atoms with larger atomic numbers  $Z$  and more shells  $n$ , so that the outer or valence electrons would feel even less of the full electrostatic attraction from the nucleus.

For an atom with many electrons, the first electron fills the lowest energy orbital, and the second electron fills the next lowest energy orbital, and so forth. For a one-electron atom or ion, the energy depends only on  $n$ , the principal quantum number; but for a many-electron atom or ion, the value of  $l$  also plays a role in the energy. The order of atomic orbital energy is given by

$$1s < 2s < 2p < 3s < 3p < 4s < 3d < 4p < 5s < 4d < 5p < 6s < 4f \sim 5d < 6p \text{ etc.}$$

Note the reversal of order for  $4s < 3d$ , and of  $5s < 4d$ , and of  $6s < 4f \sim 5d$ . The helium atom has two  $1s$  electrons, which completes the first shell, and the convention is to write  $\text{He} = 1s^2$ . The neon atom has in addition two  $2s$  and six  $2p$  electrons, which completes the second shell, and the convention is to write  $\text{Ne} = [\text{He}]2s^22p^6$ . The argon atom is given as  $\text{Ar} = [\text{Ne}]3s^23p^6$ , the krypton atom as  $\text{Kr} = [\text{Ar}]3d^{10}4s^24p^6$ , the xenon atom is written as  $\text{Xe} = [\text{Kr}]4d^{10}5s^25p^6$ , and the radon atom as  $\text{Rn} = [\text{Xe}]4f^{14}5d^{10}6s^26p^6$ . For an atom like carbon, we have six electrons distributed as  $\text{C} = [\text{He}]2s^22p^2$ . The filled shells of an atom, such as  $[\text{Ar}]$  in  $\text{Kr}$ , are called the core electrons, and the electrons in the unfilled shell are called the valence electrons.

The size of an atom or molecule is important in numerous physical phenomena and properties, and its value depends on the circumstances. Table 4.2 gives the orbital radii of the atoms compiled by Karplus and Porter (1970), which are the radii of the

Table 4.2 Geometric and electronic properties of atoms

Element	At. no.	At. wt	Radius (Å) <sup>a</sup>		Ioniz. pot. (eV) <sup>a</sup>	Electroneg. <sup>b</sup>	Radius			
			Crystal	Orbital			VdW <sup>c</sup>	Covalent <sup>c</sup>	Ionic <sup>c</sup>	Metallic
Hydrogen	H	1	1.008	0.25	0.529	13.595	2.10	1.2	37	154
Helium	He	2	4.003	1.53	0.291	24.580	0.00	32	0	
Lithium	Li	3	6.940	1.45	1.586	5.390	1.00	152	68	1.225
Beryllium	Be	4	9.010	1.05	1.040	9.320	1.50	113	153	0.889
Boron	B	5	10.820	0.85	0.776	8.296	2.00	88	35	0.81
Carbon	C	6	12.010	0.70	0.620	11.264	2.50	77	260	
Nitrogen	N	7	14.008	0.65	0.521	14.540	3.00	1.5	70	171
Oxygen	O	8	16.000	0.60	0.450	13.614	3.50	1.4	66	140
Fluorine	F	9	19.000	0.50	0.396	17.420	4.00	1.35	64	136
Neon	Ne	10	20.180	1.60	0.354	21.229	0.00	69	0	
Sodium	Na	11	22.997	1.80	1.713	5.138	0.90	186	95	1.572
Magnesium	Mg	12	24.320	1.50	1.279	7.644	1.20	160	65	1.364
Aluminum	Al	13	26.970	1.25	1.312	5.984	1.50	143	51	1.248
Silicon	Si	14	28.060	1.10	1.068	8.149	1.80	117	41	1.173
Phosphorous	P	15	30.980	1.00	0.919	11.000	2.10	1.9	110	212
Sulfur	S	16	32.066	1.00	0.810	10.357	2.50	1.85	104	184
Chlorine	Cl	17	35.460	1.00	0.725	13.010	3.00	1.8	99	181
Argon	Ar	18	39.940	1.92	0.659	15.755	0.00	97	154	
Potassium	K	19	39.096	2.20	2.162	4.339	0.80	227	133	2.025
Calcium	Ca	20	40.080	1.80	1.690	6.111	1.00	197	118	1.736
Scandium	Sc	21	45.100	1.60	1.570	6.560	1.30		81	1.439
Titanium	Ti	22	47.900	1.40	1.477	6.830	1.50		68	1.324
Vanadium	V	23	50.950	1.35	1.401	6.740	1.60	0	59	1.224

/

Chromium	Cr	24	52.010	1.40	1.453	6.760	1.60	0	89	1.176
Manganese	Mn	25	54.930	1.40	1.278	7.432	1.50	0	80	1.171
Iron	Fe	26	55.850	1.40	1.227	7.896	1.80	0	76	1.165
Cobalt	Co	27	58.940	1.35	1.181	7.860	1.90	0	72	1.162
Nickel	Ni	28	58.690	1.35	1.139	7.633	1.90	0	69	1.154
Copper	Cu	29	63.540	1.35	1.191	7.723	1.90	0	96	1.173
Zinc	Zn	30	65.380	1.35	1.065	9.391	1.60	0	74	1.249
Gallium	Ga	31	69.720	1.30	1.254	6.000	1.60	122	62	1.245
Germanium	Ge	32	72.600	1.25	1.090	8.130	1.80	122	53	1.223
Arsenic	As	33	74.900	1.15	1.001	10.000	2.00	2	121	222
Selenium	Se	34	78.960	1.15	0.918	9.750	2.40	2	117	198
Bromine	Br	35	79.920	1.15	0.851	11.840	2.80	1.95	114	195
Krypton	Kr	36	83.700	1.97	0.795	13.996	0.00	110	0	
/										
Rubidium	Rb	37	85.480	2.35	2.287	4.176	0.80	247	148	2.16
Strontium	Sr	38	87.630	2.00	1.836	5.692	1.00	215	113	1.914
Yttrium	Y	39	88.910	1.80	1.693	6.600	0.00	0	0	1.616
Zirconium	Zr	40	91.220	1.55	1.593	6.950	1.40	0	80	1.454
Niobium	Nb	41	92.910	1.45	1.589	6.770	0.00	0	0	1.342
Molybdenum	Mo	42	95.950	1.45	1.520	7.180	1.80	0	70	1.296
Technetium	Tc	43	98.000	1.35	1.391		0.00	0	0	1.271
Ruthenium	Ru	44	101.700	1.30	1.410	7.500	2.20	0	68	1.246
Rhodium	Rh	45	102.910	1.35	1.364	7.700	2.20	0	0	1.252
Palladium	Pd	46	106.700	1.40	0.567	8.330	2.20	0	98	1.283
Silver	Ag	47	107.880	1.60	1.286	7.574	1.90	0	126	1.339
Cadmium	Cd	48	112.410	1.55	1.184	8.991	1.70	0	114	1.413
Indium	In	49	114.820	1.55	1.382	5.785	0.00	0	0	1.493
Tin	Sn	50	118.700	1.45	1.240	7.332	1.80	163	71	1.399

Continued

Table 4.2 — Continued

Element	At. no.	At. wt	Radius (Å) <sup>a</sup>		Radius					
			Crystal	Orbital	Ioniz. pot. (eV) <sup>a</sup>	Electroneg. <sup>b</sup>	VdW <sup>c</sup>	Covalent <sup>c</sup>	Ionic <sup>c</sup>	Metallic
Antimony	Sb	51	121.760	1.45	1.193	8.640	1.90	2.2	141	245
Tellurium	Te	52	127.610	1.40	1.111	9.010	2.10	2.2	143	221
Iodine	I	53	126.920	1.40	1.044	10.440	2.50	2.15	127	216
Xenon	Xe	54	131.300	2.60	0.986	12.127	0.00	130	0	
D	Cesium	55	132.910	2.66	2.518	3.893	0.70	0	167	2.35
	Barium	56	137.400	2.15	2.060	5.210	0.90	0	153	1.981
	Lanthanum	57	138.920	1.95	1.915	5.610	1.00	0	115	1.69
	Cerium	58	132.910	1.85	1.978	6.910				1.442
	Praseodymium	59	137.400	1.85	1.942	5.760				1.343
	Neodymium	60	144.270	1.85	1.912	6.310				1.304
	Promethium	61	147.000	1.85	1.882					1.283
	Samarium	62	150.430	1.85	1.854	5.600				1.26
	Europium	63	159.200	1.85	1.826	5.670				1.265
	Gadolinium	64	156.900	1.75	1.713	6.160				1.295
	Terbium	65	159.200	1.75	1.775	6.740				1.336
	Dysprosium	66	162.460	1.75	1.750	6.820				1.44
	Holmium	67	164.940	1.75	1.727					
	Erbium	68	167.200	1.75	1.703					
	Thulium	69	169.400	1.75	1.681					
	Ytterbium	70	173.040	1.75	1.658	6.200				
	Lutetium	71	174.990	1.55	1.553	5.000				
Hafnium	Hf	72	178.490	1.45	1.476	5.500	0.00	0	0	0
Tantalum	Ta	73	180.950	1.35	1.413	7.880	0.00	0	0	0

Tungsten	W	74	183.920	1.35	1.360	7.980	1.70	0	0	74
Rhenium	Re	75	186.310	1.35	1.310	7.870	0.00	0	0	68
Osmium	Os	76	190.200	1.30	1.266	8.700	2.20	0	0	76
Iridium	Ir	77	193.100	1.35	1.227	9.200	2.20	0	0	67
Platinum	Pt	78	195.230	1.35	1.221	8.960	2.20	0	0	80
Gold	Au	79	197.200	1.35	1.187	9.223	2.40	0	0	137
Mercury	Hg	80	200.610	1.50	1.126	10.434	1.90	0	0	110
Thallium	Tl	81	204.380	1.90	1.319	6.106	0.00	0	0	0
Lead	Pb	82	207.210	1.80	1.215	7.415	1.90	0	0	84
Bismuth	Bi	83	209.000	1.60	1.295	7.287	1.90	0	0	20
Polonium	Po	84	209.000	1.90	1.212	8.430	0.00	0	0	0
Astatine	At	85	210.000		1.146		0.00	0	0	0
Radon	Rn	86	222.000		1.090	10.746	0.00	0	0	0
Francium	Fr	87	223.000		2.447					
Radium	Ra	88	226.050	2.50	2.042	5.277	0.00	0	0	143
Actinium	Ac	89	227.020	1.95	1.895	6.900				
Thorium	Th	90	232.040	1.80	1.788					
Protactinium	Pa	91	231.040	1.80	1.804					
Uranium	U	92	238.070	1.75	1.775		1.40	0	0	0
Neptunium	Np	93	237.050	1.75	1.741					
Plutonium	Pu	94	244.000	1.75	1.784					

<sup>a</sup>Crystal radius, orbital radius, and ionization potential from Karplus and Porter (1970).

<sup>b</sup>Electronegativity from Pauling (1970).

<sup>c</sup>Van der Waals, covalent, and ionic radii from Dean (1999).

principal maxima of the electronic charge densities of the outermost atomic orbitals. The largest orbital radii are for the alkali metals, which have only single s electrons in the outer orbital, reaching 2.5 Å for cesium. Within each period, the orbital radii contract when the atomic number increases, as the increase in nuclear charges pulls the orbitals tighter. The minimal value of 0.29 Å is reached by helium. The orbital radii of atoms are also shown in figure 4.1.

The first ionization potential is the energy required to pull the first electron from the outer orbital into space, and is given in table 4.2 and figure 4.2. It is seen that the required energy is lower for the metallic elements, and reaches a minimum at 3.9 eV for cesium; it is higher for the nonmetallic elements, and reaches a maximum of 13.6 eV for hydrogen and 24.6 eV for helium.

### Molecular Orbitals

For molecules with more than one electron, precise solutions become even more difficult and time consuming, and additional approximations are sought. The simplest molecule is that of hydrogen, where there are two nuclei A and B, and two electrons 1 and 2. The potential energy of the system is the sum of six electrostatic terms: the four attractive terms between A-1, A-2, B-1, and B-2, and the two repulsive terms between A-B and 1-2. We seek solutions to the Schrödinger equation of this hydrogen molecule, and the solution is assumed to be a linear combination of the products of the atomic orbitals, of nucleus A associated with electron 1 multiplied by nucleus B associated with electron 2, plus nucleus A associated with electron 2 multiplied by

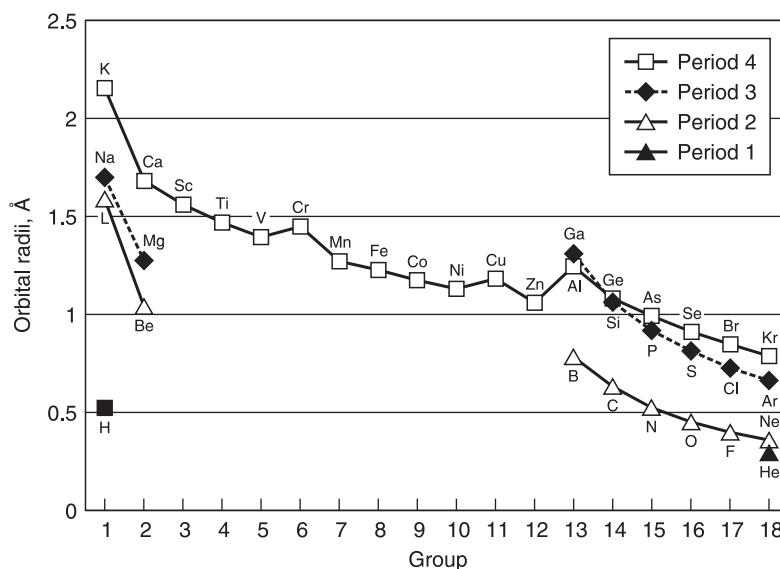


Figure 4.1 The orbital radii of elements

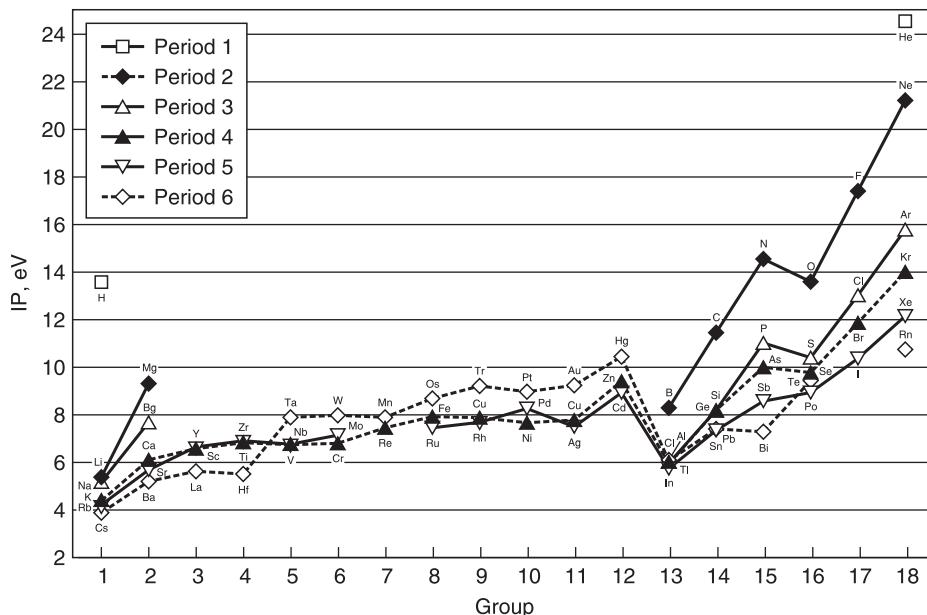


Figure 4.2 The first ionization potential of elements

nucleus B associated with electron 1:

$$\Psi = a_1\varphi_A(1)\varphi_B(2) + a_2\varphi_A(2)\varphi_B(1) \quad (4.3)$$

The parameters  $a_1$  and  $a_2$  are arbitrary parameters to be solved by minimizing the molecular energy. The solutions are called the molecular orbitals, which are responsible for bonding atoms A and B together. The equilibrium distance between A and B and the bonding energy are the most important results from this calculation.

For larger molecules of greater complexity, the computational efforts become excessive, so that even more approximations are needed to make progress. The *semi-empirical* quantum mechanical computation is a collection of methods that ignores all the core electrons and considers only the valence electrons, incorporates some parameters derived from experiments, uses a number of parameters from previously evaluated integrals, and ignores other terms in the Hamiltonian, which tends to be small from experience. There is a large and evolving set of practical methods under this umbrella, such as extended Hückel (considering only the  $\pi$  orbitals in molecules), the neglect of differential overlap (NDO), the intermediate neglect of differential overlap (INDO), the modified intermediate neglect of differential overlap (MINDO), and the Austin Model 1 (AM1) and PM3.

The two atoms in the hydrogen molecule are identical, so we have a covalent bond with equal sharing of the electrons by the two nuclei. When the two atoms are not identical, such as in HCl, we have a partially ionic bond with unequal sharing of electrons—which are clustered more around the Cl nucleus than the H nucleus.

## Molecular Mechanics

For a large and complex molecule with many electrons, the time and effort involved in a quantum mechanical calculation may become prohibitive; thus, we turn to *molecular mechanics* computations, which are based on a model of atoms as point masses connected to each other by a flexible scaffold of bonds to form molecules (Goodman 1998). These atoms move according to a set of force fields among the atoms by Newtonian classical mechanics. The information on force fields is derived from theoretical results on similar substances, and from experimental observations. It is thus generally not as accurate as the quantum mechanical methods. The quantum mechanical methods are concerned with the electrons, and can provide information on the electronic distributions (such as the shape and spatial distribution of the electrons, the dipole and quadrupole moments, and the ionization energies); however, the molecular mechanical methods deal with balls and springs, and can provide information on structure and energy, but cannot provide electronic information.

### Bonding Force Fields

The simplest case of a diatomic molecule A—B is modeled as two spheres connected by a spring, and the equilibrium bond length between them is given by  $r_0$ , for instance hydrogen has an equilibrium bond length of 0.74 Å. If the spring is stretched or compressed, the molecular energy is increased according to the quadratic equation

$$E_r = k(r - r_0)^2 \quad (4.4)$$

A triatomic molecule A—B—C is modeled as three spheres connected by two springs; and in addition, the bond angle  $\theta$  between A—B and B—C has an equilibrium value of  $\theta_0$ . For instance, water has an equilibrium bond angle of 104.5°, so that if it is bent to another value the molecular energy is increased according to the quadratic equation

$$E_\theta = k_\theta(\theta - \theta_0)^2 \quad (4.5)$$

For a four-atom molecule A—B—C—D, the plane formed by A—B—C may or may not lie in the plane formed by B—C—D, and the angle between these two planes is called the torsion or dihedral angle  $\tau$ . For instance, in ethane, the sequence H—C—C—H has an equilibrium torsion angle of 60° or 180°. When the torsion angle is not at the equilibrium value, the energy of the molecule is increased by

$$E_\tau = \frac{V_1}{2} (1 + \cos \tau) + \frac{V_2}{2} (1 - \cos 2\tau) + \frac{V_3}{2} (1 + \cos 3\tau) \quad (4.6)$$

The lowest energy of ethane H<sub>3</sub>C—CH<sub>3</sub> is the “staggered” form, with the equilibrium torsion angle of 60°.



These methods require a number of force fields, with geometric parameters such as  $r_0$ ,  $\theta_0$ , and  $\tau_0$ , which are determined by experimental values of X-ray diffraction, and with energy parameters such as  $k_r$ ,  $k_\theta$ , and  $k_\tau$  which are determined by infrared spectroscopy. Sometimes these parameters are taken from other molecules that have already been investigated and are considered similar. These simple energy wells are sufficiently accurate only when the deviation from equilibrium positions is small, but more complicated energy wells or force-field functions are used for large deviations.

### *Non-bonding Force Fields within a Molecule and between Molecules*

Attractive and repulsive van der Waals forces also exist between atoms from the same molecule that are not bonded together, such as the steric hindrance of the two methyl groups in ethane or *o*-xylene. Repulsion between two atoms occurs when they are pushed too closely together, according to the Pauli principle. Hydrogen bonds can form between atoms and functional groups within the same molecule, and one of the most famous examples is in the double helix of DNA molecules, responsible for holding the adenine–thymine and the cytosine–guanine groups together. When two molecules are pushed or “docked” together, they also exhibit attractive and repulsive forces.

The major forms of van der Waals forces between molecules that are not bonded together are the permanent dipole–dipole interaction, the dispersion-induced temporary dipole interaction, and the hydrogen bond. They are short-range forces that operate only when two atoms or molecules are in close proximity. The Lennard–Jones potential of 6–12 is a model of this potential field:

$$U = 4E \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (4.7)$$

where  $\sigma$  is a distance parameter and  $U$  is the energy, which reaches a minimum with the value of  $E$  at the distance of  $r = 2^{1/6}\sigma$ . See figure 4.3, where the energy is zero at  $r/\sigma = 1$ . This potential gives rise to a force on the particles with the value of  $F(r) = -dU/dr$ , which is attractive at  $r/\sigma > 1$  and repulsive at  $r/\sigma < 1$ .

**Dipole–dipole forces** The simplest model of a molecule with a dipole moment consists of a positive charge  $+q$  and a negative charge  $-q$ , separated by a distance  $L$ , and the dipole moment is defined as  $\mu = qL$ . The unit of the dipole moment is the Debye, which is equal to  $3.336 \times 10^{-30}$  C m. Only asymmetrical molecules can have dipole moments. The dipole moments of some important molecules are listed in table 4.3.

When two dipoles approach each other, the most stable configuration with the lowest energy is when they orient themselves in a linear head-to-tail way, so that the positively charged end of molecule 1 is close to the negatively charged end of molecule 2, creating an attractive force between them. The least stable form is a linear head-to-head orientation when the two positive ends are together. The interaction force depends on the distance and the angle between them. At absolute zero temperature, the two dipoles are lined up in a straight line head to tail, but at any other temperature the random thermal forces disrupt this line up, so that the resulting energy is given by an

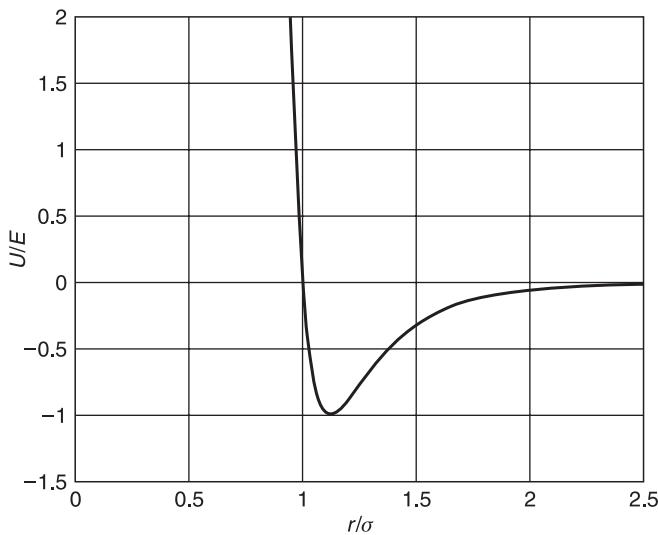


Figure 4.3 The Lennard-Jones potential

Table 4.3 Dipole moments of some important molecules

Dipole moment (Debye)	
He	0
Ne	0
Ar	0
Kr	0
Xe	0
N <sub>2</sub>	0
CH <sub>4</sub>	0
C <sub>2</sub> H <sub>6</sub>	0
C <sub>3</sub> H <sub>8</sub>	0.01
CO	0.12
CO <sub>2</sub>	0
HCl	1.04
NH <sub>3</sub>	1.51
H <sub>2</sub> O	1.85
C <sub>6</sub> H <sub>5</sub> NO <sub>2</sub>	4.2

averaging of many configurations, which results in

$$u(r) = -\frac{\mu_1^2 \mu_2^2}{3(4\pi\epsilon_0)^2 k T r^6} \quad (4.8)$$

This is the dipole-dipole interaction energy, often termed the orientation or Keesom interaction. Notice that it depends on the product of the squares of both dipole moments, but is inversely proportional to distance to the sixth power. This is a very short-range

**Table 4.4** Polarizabilities ( $\text{\AA}^3$ ) of some molecules

He	0.20	$\text{H}_2$	0.81	$\text{CH}_4$	2.6
Ar	1.63	$\text{O}_2$	1.60	$\text{C}_2\text{H}_6$	4.5
Xe	4.0	$\text{H}_2\text{O}$	1.48	$\text{C}_6\text{H}_6$	10.3
		$\text{CO}_2$	2.6	$\text{CCl}_4$	10.5
		$\text{NH}_3$	2.3		
		$\text{HCl}$	2.6		

force that operates only when the two molecules are quite close. When the distance is increased by 10%, the attractive force declines to 56%; and when the distance is doubled, the attractive force declines to 1.6%. The energy is also inversely proportional to the temperature  $T$ , so that this energy decreases at high temperature due to increased thermal random motions.

*London dispersion or induced dipole forces* Any atom or molecule can become polarized when placed in an electrical field  $E$ , and the induced dipole moment is proportional to its polarizability factor  $\alpha$  by

$$\mu_{\text{ind}} = \alpha E \quad (4.9)$$

For a nonpolar molecule, the polarizability arises from the displacement of its negatively charged electron cloud relative to the positively charged nucleus. The polarizabilities of some molecules are given in table 4.4.

When two unlike atoms approach each other, the dispersion energy induced between them is given by

$$u(r) = -\frac{3}{2} \frac{\alpha_{\text{o1}} \alpha_{\text{o2}}}{(2\pi\epsilon_0)^2 r^6} \left( \frac{I_1 I_2}{I_1 + I_2} \right) \quad (4.10)$$

where the  $I_i$  is the first ionization potential of the atom, which measures the energy required to detach an electron from the atom. The dispersion force is also proportional to the product of the two polarizabilities  $\alpha$ , and to the distance  $r^{-6}$ , but is independent of temperature.

*Hydrogen bond forces* The third major type of van der Waals force is the hydrogen bond, which is found in compounds that are rich in hydrogen attached in  $-\text{OH}$ ,  $-\text{NH}$  or  $\text{HF}$  groups of atoms. The strength of hydrogen bonds is around 10 to 40 kJ/mol, which is stronger than a typical dispersion force at 1 kJ/mol, but much less than a covalent or ionic bond at 500 kJ/mol. Water, ammonia, and hydrogen fluoride are examples of materials that have strong hydrogen bonds. They are also found in the alcohols, carboxylic acids, primary and secondary amines, and amides. Molecules with O, N, and F that are not attached to hydrogen atoms do not form hydrogen bonds. For instance, ethers, aldehydes, ketones, esters, tertiary amines, and carbon tetrachloride do not have hydrogen bonds. The hydrogen bonds have very strong directional qualities to them, so that a linear hydrogen bond of three atoms  $\text{H}-\text{O}\cdots\text{H}$  with an angle of  $180^\circ$  is the strongest, and seldom does a hydrogen bond form when the angle is smaller than  $120^\circ$ .

**Table 4.5 The effect of hydrogen bonds**

	<i>MW</i>	<i>M.p., °C</i>	<i>B.p., °C</i>	<i>Dipole moment</i>	<i>Hydrogen bond</i>
<chem>C2H6</chem>	30	-183	-89	0	No
<chem>CH3CHO</chem>	44	-123	20	2.5	No
<chem>CH3OH</chem>	32	-98	64	1.7	Yes

The importance of these hydrogen bonds can be demonstrated by the comparisons of melting points and boiling points between a neutral paraffin, a polar aldehyde, and an alcohol with hydrogen bonds given in table 4.5.

**Electrostatic forces** An electrically neutral molecule can carry charged atoms or groups, for instance in HCl, the more electronegative chlorine draws a fraction of an electron away from the less electronegative hydrogen, and the result is a polar molecule. The attraction between positive and negative charges, and the repulsion between like charges, follow Coulomb's law:

$$U = \frac{q_1 q_2}{4\pi \epsilon r} \quad (4.11)$$

where  $q_i$  is the electrical charge on the  $i$ th particle,  $\epsilon$  is the dielectric constant of the medium,  $r$  is the distance between the centers of these particles, and  $U$  is the potential between them.

The Coulombic electrostatic force is a very long-range force. When the distance is doubled, the force decreases to 50%; when a molecule is removed to 10 times the normal distance, the attractive force declines to 10%.

The molecular mechanics energy of the molecule is the sum of all these energies:

$$E_{\text{MM}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsion}} + E_{\text{vdWaals}} + E_{\text{electrostatic}} \quad (4.12)$$

and the most stable configuration of the molecule is the one with the lowest energy.

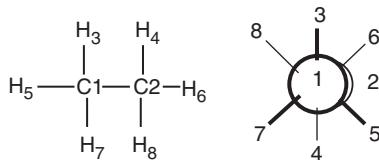
### Predictable Properties

At the conclusion of a geometric optimization calculation, we have the equilibrium positions of all the atomic nuclei, as well as the overall electron density distributed in space ( $x$ ,  $y$ ,  $z$ ). Many important properties, especially for an isolated single molecule at absolute zero temperature, can be obtained by solving the quantum mechanical or the molecular mechanical equations. Only the former method can produce electronic properties, such as electron distributions and dipole moments, but both methods can produce structural and energy properties.

The structure of the molecule is the most important information in the prediction of properties. One method of representation is to specify the ( $x$ ,  $y$ ,  $z$ ) coordinates of all the atomic nuclei. Another method, which is more compact and useful, is to list the positions of the atoms by the internal coordinate method called the  $Z$  matrix. For instance, this is given for ethane in table 4.6.

**Table 4.6** Z matrix for ethane

No.	Element	Connects to atom	Length, Å	Connects to atom	Angle, deg	Connects to atom	Torsion angle, deg
1	C						
2	C	1	1.54				
3	H	1	1.0	2	109.5		
4	H	2	1.0	1	109.5	3	180
5	H	1	1.0	2	109.5	4	60
6	H	2	1.0	1	109.5	5	60
7	H	1	1.0	2	109.5	6	180
8	H	2	1.0	1	109.5	7	-60



We are interested in the properties of the molecule on several scales, from the shortest bond scale to the intermediate submolecular scale, and finally to the molecular scale. The structure with the shortest scale is a single bond between two atoms A—B, and we are interested in its equilibrium length, as well as the energy required to break the bond and move the two fragments A and B to large distances from each other. We also need to know whether the bond is covalent or ionic. After mapping out the energy contours of the molecule around the equilibrium length of  $l_0$ , as well as the neighborhood of  $l_0 + \delta l$ , we can compute the steepness of the energy well, and thus its flexibility, as well as the vibration frequency. We also need the bond angle when there are three atoms bonded together as A—B—C, and the torsion angle when there are four atoms bonded together as A—B—C—D.

In the medium scale are various substructures of a molecule such as delocalized bonds, rings and fused rings, and their conformations—whether they are planar or shaped like a boat or a chair. Large biological molecules have secondary and tertiary structures, such as the  $\alpha$  helix and  $\beta$  sheets of proteins, and the double helix of DNA molecules.

The structure of the entire molecule includes: the connectivity of the atoms or topology, the size and shape of the molecule, the dipole and quadrupole moments, the symmetry group, the thermal degrees of freedom, including rotation and vibration, and symmetry. The outer surface structures of a molecule are particularly important in its interactions or docking with another molecule. The bonding of a substrate to a binding site depends on the geometric conformations of protrusions and cavities, and of surface charges and hydrogen bond sites.

### Bond-scale Properties

**Ionic and covalent bonds** An ionic bond is formed when one atom gives up an electron partially or entirely to the other atom. This is more or less the case with

**Table 4.7** Electronegativity scale of elements

	1	2	13	14	15	16	17							
1	H	2.1												
2	Li	1.0	Be	1.5	B	2.0	C	2.5	N	3.0	O	3.5	F	4.0
3	Na	0.9	Mg	1.2	Al	1.5	Si	1.8	P	2.1	S	2.5	Cl	3.0
4	K	0.8	Ca	1.0	Ga	1.6	Ge	1.8	As	2.0	Se	2.4	Br	2.8
5	Ru	0.8	Sr	1.0	In	1.7	Sn	1.8	Sb	1.9	Te	2.1	I	2.5
6	Cs	0.7	Ba	0.9	Tl	1.8	Pb	1.9	Bi	1.9	Po	2.0	At	2.2

the molecule  $\text{Cs}^+\text{F}^-$ . The electronegativity scale is the tendency of an atom to gain an electron and form a negatively charged ion. We follow the Linus Pauling nomenclature and values, given in table 4.7. The element fluorine is the most electronegative element with a value of 4.0, and cesium is the least electronegative element with a value of 0.7, for a maximum difference of 3.3. The C—H bond is very slightly ionic with a difference of 0.4, with the carbon atom being more negative than the hydrogen atom. Pauling suggested that the percent ionic character of a bond is related to the difference in electronegativity of the two atoms, and he proposed that

$$\text{Ionic character (\%)} = [1 - \exp(-\Delta x^2/4)] \times 100$$

where  $\Delta x$  is the difference in electronegativity of the two atoms. Table 4.8 and in figure 4.4 show the results of this computation.

The Cs—F bond is among the most polarized, with a higher than 90% ionic character. A molecule with ionic bonds has a tendency to manifest a permanent dipole moment; but a symmetrical molecule, such as the linear molecule  $\text{O}=\text{C}=\text{O}$  has no dipole moment as the effects of one ionic bond are canceled by an equal and opposite ionic bond. The bent molecule  $\text{H}-\text{O}-\text{H}$  would exhibit a dipole moment, as the two vectors of  $\text{H}-\text{O}$  are at an angle of  $104.5^\circ$ .

In a covalent bond, there is an equal sharing of the electrons from the two separate atoms. Examples are with the molecules  $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{O}_2$ ,  $\text{F}_2$ ,  $\text{Cl}_2$ ,  $\text{Br}_2$  and  $\text{I}_2$ . Since the

**Table 4.8** Ionic character as a function of electronegativity difference

$\Delta x$	<i>Ionic character (%)</i>
0	0
0.5	6
1.0	22
1.5	43
2.0	63
2.5	79
3.0	89
3.5	95

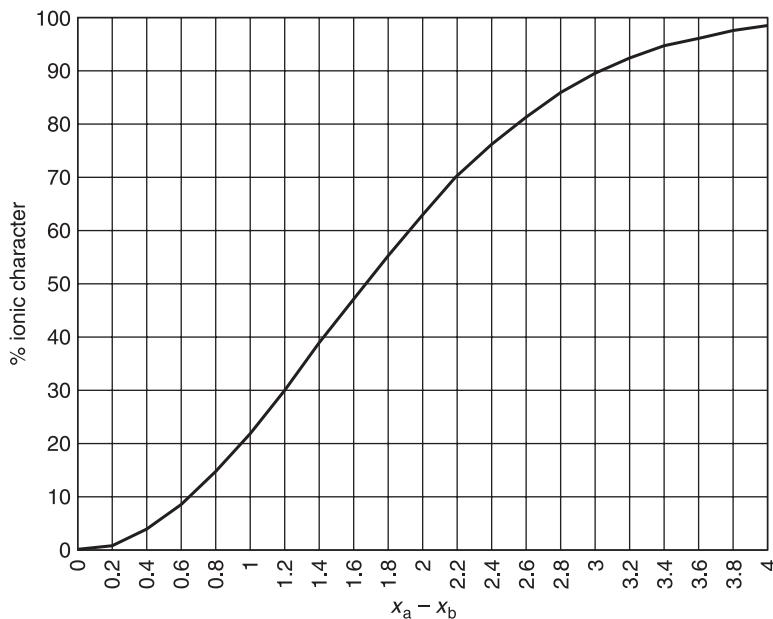


Figure 4.4 Percentage ionic bond and the electronegativity difference of elements

two atoms are the same, there is no tendency for one atom to gain or lose electrons with respect to the other. A molecule entirely bonded together with covalent bonds would not have a permanent dipole moment. The C—H bond has less than 6% ionic character, and is practically a covalent bond.

The covalent radii between two identical atoms of the most important elements are given in table 4.2. For the period 2 nonmetallic elements of C, N, O and F, the single bond radius is about 0.70 Å. The most important carbon bonds in organic chemistry are based on the hybridization of the s orbital with the p orbitals to form hybrid bonds. Methane, CH<sub>4</sub>, is built on the tetrahedral sp<sup>3</sup> orbitals formed by one s with all three p orbitals, and has a bond angle of 109.5° for each sequence H—C—H. Ethylene is built on the trigonal sp<sup>2</sup> orbitals formed by one s with two p orbitals, and the sequence H—C=C has an angle of 120°. Acetylene is built on the straight sp orbitals formed by one s with one p orbital, and the sequence H—C≡C has an angle of 180°. The double bond is shorter and more energetic than the single bond, and the triple bond is even shorter and more energetic than the double bond (table 4.9).

The covalent radius between identical atoms also decreases within a period when the group number is increased, due to the larger nuclear charges exerting more attraction on the electrons (table 4.10).

Pauling also suggested that the single bond length between two unlike atoms is approximately given by the arithmetic average of the two bonds  $d_{AB} = (d_{AA} + d_{BB})/2$ .

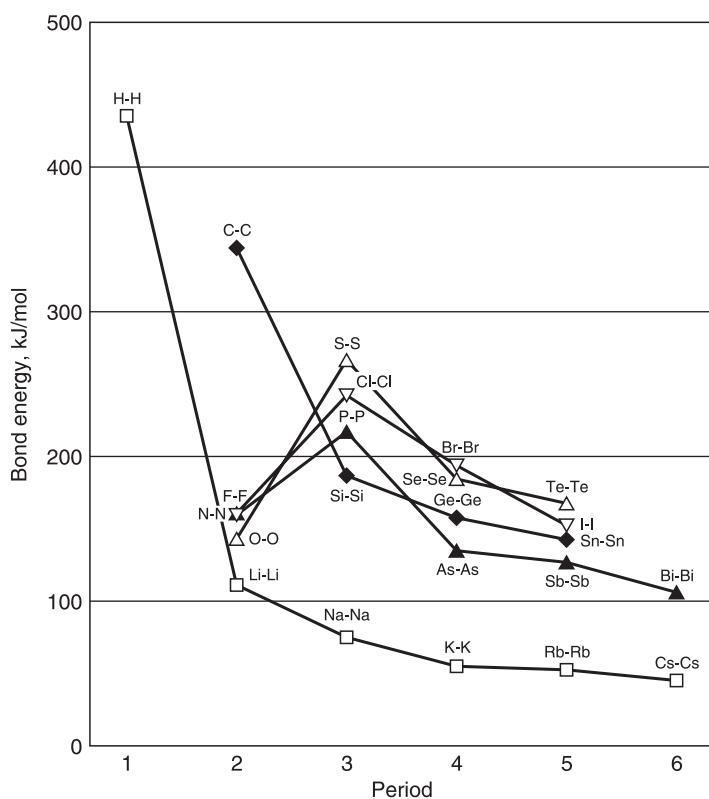
The energies of the covalent homoatomic X—X bonds are shown in figure 4.5, where the heavier atoms make weaker bonds. The energies of the covalent heteroatomic C—X and H—X bonds also decrease with the period, as shown in figure 4.6.

**Table 4.9 Carbon–carbon bond properties**

	Length, Å	Energy, kcal/mol	Hybrid orbital	Bond angle, deg
C—C	1.542	83.1	sp <sup>3</sup>	109.5
C=C	1.334	147	sp <sup>2</sup>	120
C≡C	1.206	194	sp	180

**Table 4.10 Covalent radius change within a period**

	C	N	O	F
Single	0.772	0.74	0.74	0.72
Double	0.667	0.62	0.62	
Triple	0.603	0.55		



**Figure 4.5** Covalent bond energy of X—X type

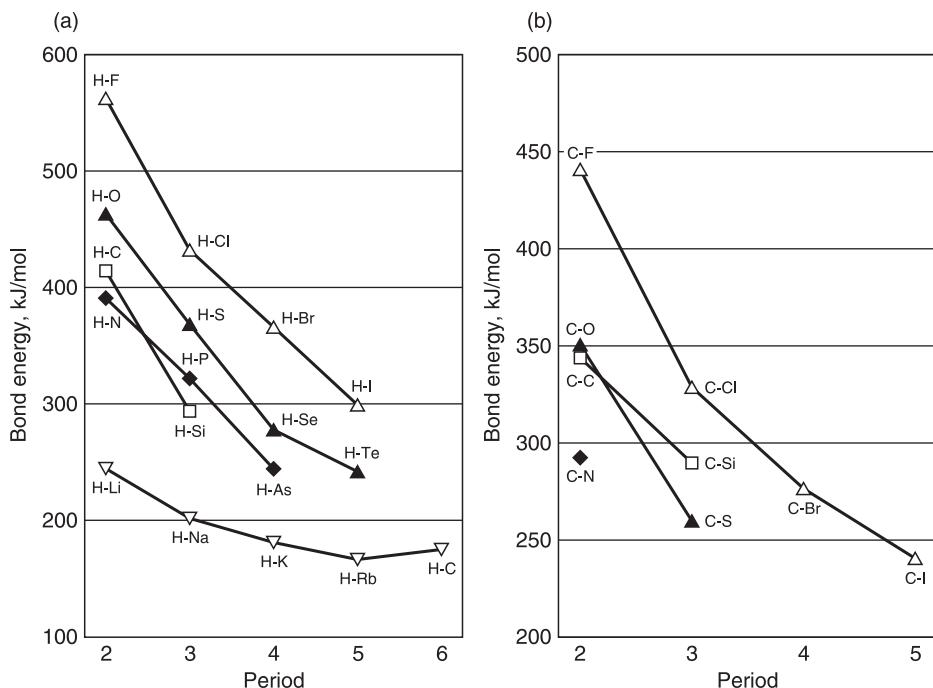


Figure 4.6 Covalent bond energy of (a) H—X and (b) C—X types

Pauling suggested that, in the process of tearing apart A—A and B—B and then forming two A—B bonds, there may be no loss or gain of energy unless there is a strong difference in electronegativity. In that case, there is an additional contribution due to the ionic component to the covalent bond. He suggested the formula

$$E_{ab} \text{ (kJ/mol)} = 100 (x_a - x_b)^2 + \sqrt{E_{aa} E_{bb}} \quad (4.13)$$

where  $x$  is the electronegativity of the element. For instance, the H—O bond energy is predicted to be  $100(3.5 - 2.1)^2 + (139 \times 436)^{0.5} = 196 + 246 = 442$ . These values are reasonably close to the actual value of 462 kJ/mol, or 110.4 kcal/mol. The strongest H—X and C—X bonds are from period 2 for C to F. When you increase in group, such as from C—N to C—O, then to C—F, the bond energy generally increases. Since F is the most electronegative of all the elements, the H—F and the C—F bonds have the largest differences in electronegativity and the largest bond energies.

*Bond length, angle, and energy* The length of a bond reflects both the sizes of the two atoms and the strength of the bonding between them. The length of a covalent bond is approximately equal to the sum of the two covalent radii of the atoms involved, and the shorter bonds tend to be stronger. The covalent radii for homoatomic bonds of the smaller elements are given in table 4.11.

**Table 4.11** Covalent radii (pm) within homoatomic bonds

	<i>Single</i>	<i>Double</i>	<i>Triple</i>		<i>Single</i>	<i>Double</i>	<i>Triple</i>		<i>Single</i>	<i>Double</i>		<i>Single</i>
C	77.2	66.7	60.3	N	70	60	55	O	66	55	F	30
Si	177	107	100	P	110	100	93	S	104	94	Cl	99
				As	121	111		Se	117	107	Br	114
								Te	137		I	133

These covalent bond lengths are reasonably constant among molecules, as the paraffin C—C bond usually has a length of 154 pm, the olefin C=C double bond has a length of 134 pm, and the acetylenic triple bond has a length of 120 pm. The C—H bond is 109 pm in a paraffin and 105 pm in an acetylene.

The angle  $\theta$  between the A—B and the B—C bonds can vary between a highly strained 60° in cyclopropane, to a bent 104.5° in H<sub>2</sub>O, to a straight 180° in acetylene and CO<sub>2</sub>. The bent structure of the water molecule can be explained as due to repulsion between the two hydrogen atoms, which pushes the bond angle from the natural 90° between two 2p orbitals to 104.45°. An alternate explanation is that the architecture of water is built around a tetrahedron with angles of 109.5°. Oxygen and the two H atoms share eight valence electrons among them, as four of them are used for bonding between the two H atoms and the one O atom. This leaves four valence electrons that are not used in the bonding, and they form two “lone pairs” that take up the other two branches of the tetrahedron. The simplest approach to the prediction of molecular geometry is called the valence-shell electron-pair repulsion (VSEPR) method, which states that the bonding electron pairs and the lone pairs of an atom tend to adopt a spatial arrangement to minimize electron-pair repulsion around that atom.

As one proceeds from water to the next few periods, the bond lengths of H—O, H—S, H—Se, to H—Te gradually become larger, so that there is less steric hindrance between the two —H atoms, and consequently the bond angles are closer to 90°. An even stronger repulsion is found in NH<sub>3</sub>, which has an angle of 107.3°. As the nitrogen and three hydrogen atoms share eight valence electrons among them, and six are used in bonding between three H and N, there is only one lone pair left. As the period increases from NH<sub>3</sub>, to PH<sub>3</sub>, to SbH<sub>3</sub>, the angle declines to 91.3° (table 4.12).

**Table 4.12** Bond parameter changes across periods

	<i>Bond length, pm</i>	<i>Bond energy, kJ</i>	<i>Angle, deg</i>		<i>Bond length, pm</i>	<i>Bond energy, kJ</i>	<i>Angle, deg</i>
H <sub>2</sub> O	95.8	498	104.45	NH <sub>3</sub>	103.4	435	107.3
H <sub>2</sub> S	133.3	344	92.2	PH <sub>3</sub>	142	343	93.3
H <sub>2</sub> Se		305	91.0	AsH <sub>3</sub>	152		91.8
H <sub>2</sub> Te		268	89.5	SbH <sub>3</sub>			91.3

**Table 4.13 Properties of hybrid bonds**

Example	Bond	Bond length, Å	Bond energy, kcal	Bond angle, deg	Bond	Bond length	Bond energy
sp <sup>3</sup> ethane	C—C	1.54	88	109.47	C—H	1.11	98
sp <sup>2</sup> ethylene	C=C	1.34	167	120	C—H	1.07	104
sp acetylene	C≡C	1.20	230	180	C—H	1.06	113

For the acetylene molecule, the sp hybrid orbital is formed by combining one 2s orbital and one 2p orbital to form two sp orbitals. Each C atom uses one sp orbital to bond to one 1s from H, and the other sp orbital to bond to the other C. The C≡C triple bond is formed by the addition of two  $\pi$  bonds by the overlap of two 2p orbitals from each C. Some of the properties of these hybrid bonds are shown in table 4.13.

A complete architectural description of ammonia needs to include the torsion angle  $\tau$ , the angle formed between the H<sup>1</sup>—N—H<sup>2</sup> plane and the H<sup>2</sup>—N—H<sup>3</sup> plane, which has a value of 119.14°.

### *Submolecular-scale Properties*

Some bonds are not local affairs between two atoms, but are extended over several atoms. Familiar examples are the resonance double bonds in butadiene, H<sub>2</sub>=CH—CH=CH<sub>2</sub>. In benzene, the electrons on each C are hybridized to form three sp<sup>2</sup> orbitals. Each C atom uses two sp<sup>2</sup> orbitals to make  $\sigma$  bonds to the neighboring carbon atoms, and the remaining sp<sup>2</sup> orbital to make one  $\sigma$  bond with the H atom. In addition, each C uses the remaining 2p orbital to form a ring of six delocalized  $\pi$  orbitals. For graphite, there is an extended covalent bonding in a two-dimensional plane; and for diamond, there is an extended covalent bonding in three-dimensional space. We shall see that, in metals, there is a sea of delocalized electrons that bind the metal together.

The cyclohexane molecule can exist in several conformations. The lowest energy is the chair conformation, where four carbon atoms form a plane, where one end carbon atom is above the plane and the other end carbon atom is below the plane. The higher energy is the boat conformation, where both end carbon atoms are above the plane.

Many molecules do not have rigid structures, and are flexible at higher temperatures. The rigidity is controlled by the steepness of the energy wells and the spring constants around the equilibrium bond lengths, bond angles, and bond torsion angles. Ethane can be in a number of configurations or conformations, and the “eclipsed” configuration (torsion angle  $\tau$  = 0 or 120°) has an energy that is 2.9 kcal/mol higher than the more stable “staggered” configuration ( $\tau$  = 60 or 180°); see figure 4.7. When we consider the central C—C bond of the butane molecule, we find that there are two relatively stable configurations: the “anti” ( $\tau$  = 180°) arrangement of the two methyl groups has an energy of 0.9 kcal/mol lower than the “gauche” ( $\tau$  = 60°) configuration, and these two configurations are separated by an energy barrier of 3.6 kcal/mol. At a temperature of absolute zero, the molecules will be entirely at the minimum energy configuration.

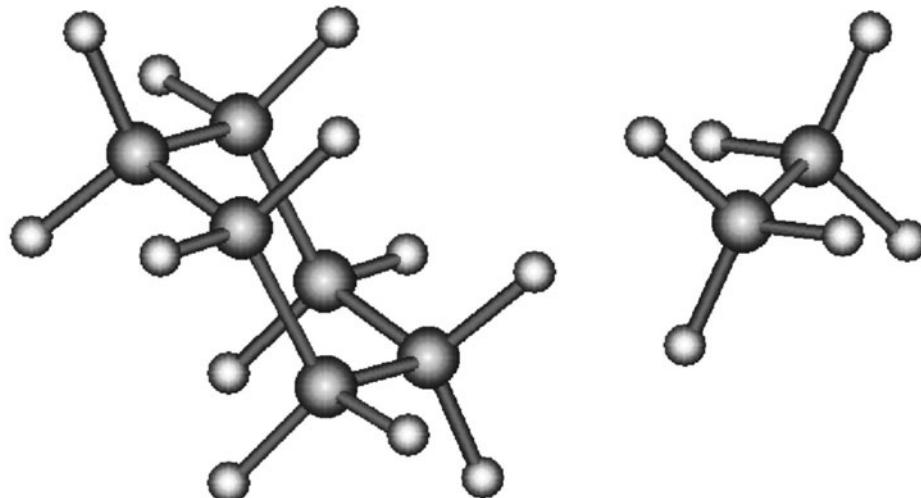


Figure 4.7 Conformations of cyclohexane in chair form and of ethane in staggered form

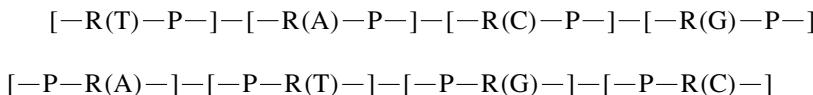
But, at the room temperature, we find that about 20% of the molecules have been promoted to the gauche configuration with higher energies.

For a long molecule, such as eicosane  $C_{20}H_{22}$ , with 19 C—C bonds, the lowest energy form is the extended configuration where all the torsion angles are at  $180^\circ$ , and all the C—C bonds are in the same plane, in the so-called “all trans” configuration. Since the energy difference between the anti and the gauche forms is only of the order of  $kT$  at room temperature, we would find a significant fraction of the eicosane molecules in one of the folded structures. The molecule of cyclopropane is rigid, but cyclobutane has two equivalent configurations. The molecule cyclopentane at the lowest energy form has four carbon atoms on the same plane, and the fifth carbon is not in the plane. The cyclohexane molecule has the most stable configuration in the chair form; this is shown in figure 4.7, in addition to the less stable configurations of the twisted chair and the boat.

The  $\alpha$  helix is a spiral coil in a protein molecule connecting consecutive amino acids, with approximately 3.6 to 3.7 amino acid residues per turn of the helix, held together by the hydrogen bonds between the N—H groups of one amino acid and the O atoms of another amino acid. The backbone of the protein is formed by the repeat sequence of  $[-C(O=)-N(H)-C(R)-]$ , where the side groups are enclosed in the parentheses. The formation of the alpha helix proceeds when the  $=O$  group of the first amino acid is hydrogen bonded to the N(H) group of the fifth amino acid, and the  $=O$  group of each  $j$ th amino acid is hydrogen bonded to each  $(j + 4)$ th N(H) group.

The DNA molecule is a double strand of nucleic acids, each consisting of a sequence of nucleosides, which are paired ribose sugar and phosphoric acid, held together in sequences of  $[-R(X)-P-]$ , where X is one of four bases of thymine, cytosine, adenine, or guanine. Thymine can bond to adenine by two hydrogen bonds, and cytosine can bond to guanine by three hydrogen bonds. The two strands of DNA molecules run

in opposite directions, and are bonded together by the hydrogen bonds between T and A and between C and G.



### *Molecule-scale Properties*

The thermochemistry parameters are some of the most useful results of quantum computation. The important features of the overall architecture of a molecule include the connectivity or topology, flexibility, vibration modes and frequencies, shape, moments of inertia, rotational energies and frequencies, and the overall symmetry. The shape of a molecule plays an important part in many of its properties. Some of the most interesting and easy to describe are the long and narrow rod-like eicosane,  $C_{20}H_{44}$ , the disk-shaped coronene,  $C_{24}H_{12}$ , and the sphere-shaped methane,  $CH_4$ , adamantine,  $C_{10}H_{16}$ , and buckminsterfullerene,  $C_{60}$ .

*Connectivity or topology* Many molecules have the architecture of a single chain, where all the atoms are connected together in a linear manner, such as in the normal paraffins. Many more have branching from the backbone, such as the iso-paraffins. The starch molecule is very highly branched, so that there is no identifiable backbone. These are also acyclic structures, so that if you cut any bond the molecule will part as two separate molecules.

Cyclohexane and benzene are examples of the monocyclic molecules, so that if you cut a bond, you will have a chain molecule instead of two molecules. Naphthalene is a bicyclic compound, and anthracene is a tricyclic compound, and it will take several cuts to reduce the molecule to a chain. The insulin molecule is made of two chains of length 21 and 30, held together by two disulfide bonds S—S between the cystine groups. This topological feature of connectivity plays an important role in the properties of molecules. For instance, the thermoplastic polymers, such as polyethylene, consist of individual chains and melt at high temperatures, but the thermosetting polymers, such as Bakelite, consist of cross-linked chains and decompose without melting at high temperatures.

*Thermal motions* A molecule has three translational degrees of freedom. Let us consider a system of  $M$  ideal monatomic gas molecules in a cubic box kept at a constant temperature. For a very dilute gas, where the molecules do not interact with one another, the quantum mechanical solution is a number of wave functions with three quantum numbers,  $n_x$ ,  $n_y$ , and  $n_z$ , for the translational energies in three dimensions. The energy of a molecule in a cubic box with side length  $a$  is given by

$$E_{n_x, n_y, n_z} = \frac{h^2}{8\pi^2 m a^2} (n_x^2 + n_y^2 + n_z^2) \quad (4.14)$$

The energy is quantized by the three quantum numbers, but the energy gap between quantum states is so small that, for most purposes, the energy is a continuous function of speed.

A polyatomic molecule structure has flexibility, and vibrates when the temperature is above absolute zero. A molecule with  $n$  nuclei has  $3n$  degrees of freedom, which represents motion in the  $x-y-z$  directions, including the three degrees of translational freedom. A linear molecule has two degrees of rotational freedom, but a nonlinear molecule has three degrees of rotational freedom. The vibrational degrees of freedom are thus  $3n - 5$  for a linear molecule, and  $3n - 6$  for a nonlinear molecule. Each degree of freedom has a vibrational frequency of  $\nu_m$ , which is given by  $\nu_m = (1/r_0)\sqrt{K/\mu}$ , where  $r_0$  is the bond length,  $K$  is the stiffness of the spring, and  $\mu = m_1m_2/(m_1 + m_2)$  is the reduced mass of the two atoms. The vibration energy is given by

$$E_m = h\nu_m\left(m + \frac{1}{2}\right) \quad m = 0, 1, 2, \dots \quad (4.15)$$

Even at the lowest energy state, where  $m = 0$ , there is an energy of vibration called the zero-point energy. There is only one vibrational state at each value of  $m$ , and the energies of the sequence are evenly spaced (table 4.14).

Changes in the vibrational energies can usually be detected in infrared spectroscopy in the wavenumber range 200 to 3500  $\text{cm}^{-1}$ . The vibrational temperature is  $\Theta_{\text{vib}} = h\nu/k$ , where  $k$  is the Boltzmann constant, or the gas constant divided by the Avogadro's number. When the temperature is less than the vibrational temperature, this degree of freedom is not fully activated; 50% activation is reached when  $T = 0.34\Theta$ .

The analysis of vibration spectra proceeds by the use of normal modes. For instance, the vibration of a nonlinear water molecule has three degrees of freedom, which can be represented as three normal modes. The first mode is a symmetric stretch at 3586  $\text{cm}^{-1}$ , where the O atom moves up and the two H atoms move away from the O atom; the second is an asymmetric stretch at 3725  $\text{cm}^{-1}$ , where one H atom draws closer to the O atom but the other H atom pulls away; and the third is a bending moment at 1595  $\text{cm}^{-1}$ , where the O atom moves down and the two H atoms move up and away diagonally. The linear  $\text{CO}_2$  molecule has four normal modes of vibration. The first is a symmetric stretch, which is inactive in the infrared, where the two O atoms move away from the central C atom; the second is an asymmetric stretch at 2335  $\text{cm}^{-1}$ , where both O atoms move right while the C atom moves left; and the third and fourth together constitute a doubly degenerate bending motion at 663  $\text{cm}^{-1}$ , where both O atoms move forward and the C atom moves backward, or both O atoms move upward and the C atom moves downward.

**Table 4.14** Vibration energy sequence versus  $m$

Quantum number, $m$	0	1	2	3	4
Energy, $m + 1/2$	1/2	3/2	5/2	7/2	9/2
Degeneracy	1	1	1	1	1

**Table 4.15** Rotation energy sequence versus  $J$ 

Quantum number, $J$	0	1	2	3	4
Energy, $J(J + 1)$	0	2	6	12	20
Degeneracy, $2J + 1$	1	3	5	7	9

The energy of rotation is quantized with the quantum number  $J$ , and there are more than one configuration for each quantum number  $J$ , which is known as the degeneracy  $g_J$ :

$$E_J = \frac{\hbar^2}{8\pi^2 I} J(J + 1) \quad J = 0, 1, 2, \dots \quad (4.16)$$

$$g_J = 2J + 1$$

where the moment of inertia  $I = \sum \mu R^2$ . The spectra of rotation differ from the spectra of vibration, as both the energy and the degeneracy increase with the quantum number  $J$  (table 4.15).

A diatomic molecule has only two rotational degrees of freedom with the same energy, since rotation along the bond axis has essentially zero moment of inertia and there is not enough thermal energy to activate it at any temperature. A linear molecule with  $n$  atoms also has two degrees of freedom, but a nonlinear molecule with  $n$  atoms has three degrees of freedom. Changes in the rotational energies can usually be detected in microwave spectroscopy in the wavenumber range 0.05 to 10  $\text{cm}^{-1}$ . Often, the energy is expressed in terms of a wavenumber  $\tilde{B} = h/8\pi^2 Ic$ , where  $c$  is the velocity of light. The rotational temperature is  $\Theta_{\text{rot}} = \hbar^2/8\pi^2 kI$ .

Some values of rotation and vibration wavenumbers and temperatures of small molecules, and the degeneracy in parentheses, are listed in table 4.16. One can see

**Table 4.16** Rotation and vibration wavenumbers and temperatures for some small molecules

	<i>Rotation wavenumber, <math>\text{cm}^{-1}</math></i>	<i>Rotation temp., K</i>	<i>Vibration wavenumber, <math>\text{cm}^{-1}</math></i>	<i>Vibration temp., K</i>
$\text{H}_2$	59.32(2)	85.3(2)	4401	6332
$\text{HCl}$	10.44(2)	15.02(2)	2886	4227
$\text{Cl}_2$	0.24(2)	0.35(2)	554	805
$\text{Br}_2$	0.0819(2)	0.12(2)	323	463
$\text{O}_2$	1.438(2)	2.07(2)	1556	2256
$\text{N}_2$	1.990(2)	2.88(2)	2330	3374
$\text{CO}$	1.923(2)	2.77(2)	2143	3103
$\text{CO}_2$	0.390(2)	0.56(2)	2335, 1313, 663(2)	3360, 1890, 954(2)
$\text{H}_2\text{O}$	27.9, 14.6, 9.31	40.1, 21.0, 13.4	3725, 3586, 1595	5360, 5160, 2290
$\text{NH}_3$	6.20, 9.45(2)	8.9, 13.6(2)	3336, 945.2, 3392(2), 1619(2)	4800, 1360, 4880(2), 2330(2)
$\text{CH}_4$	5.27(3)	7.6(3)	2898, 1515(2), 3002(3), 1300(3)	660, 310(2), 1230(3), 450(3)

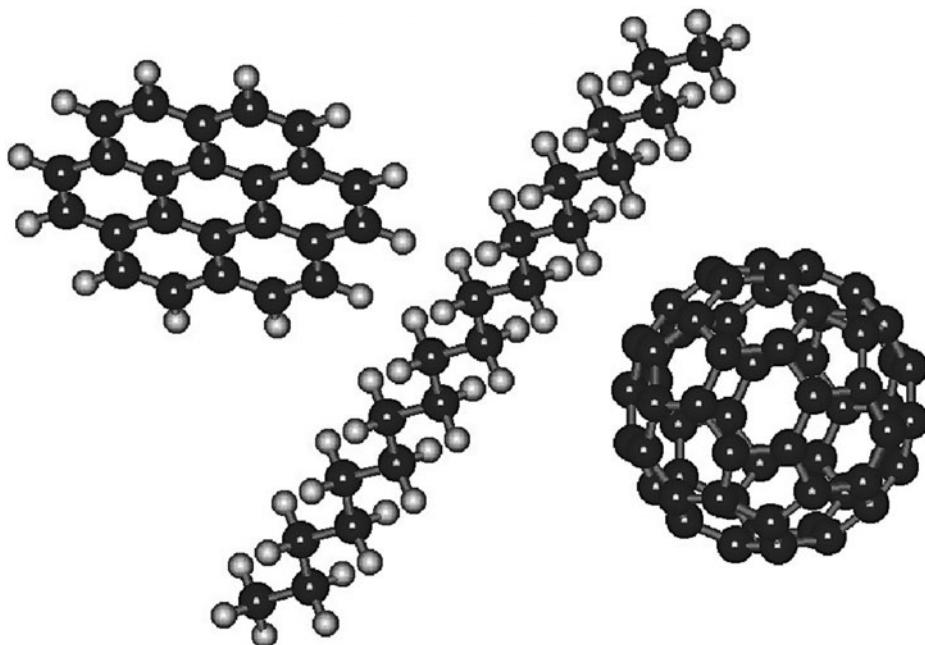


Figure 4.8 Coronene as a disk, eicosane as a rod, and Buckminsterfullerene as a sphere

that the vibrational temperatures are much higher than the rotational temperatures, and small molecules have higher temperatures than large molecules.

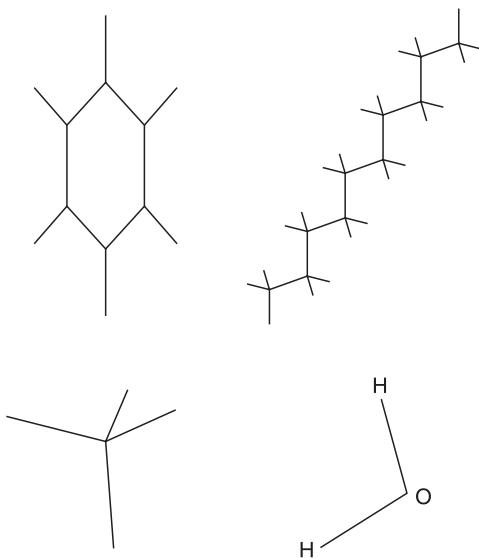
For a nonlinear molecule, there are three moments of inertia that are perpendicular to each other. We use the convention that  $I_1 \leq I_2 \leq I_3$ . The rotational geometry of a molecule belongs to one of three types, which are shown in figures 4.8 and 4.9.

Spherical top: three equal moments of inertia; for example, the spherical molecule methane,  $\text{CH}_4$ , has three equal moments of inertia  $I_1 = I_2 = I_3$  of [3.347, 3.347, 3.347]. Other examples include helium, 2,2-dimethylpropane (neopentane), 3,3-diethylpentane, and buckminsterfullerene.

Symmetric top: two equal moments of inertia. These can be pancake shaped:  $I_1 = I_2 \leq I_3$ ; for example, benzene [89, 89, 178] and coronene with two smaller moments of inertia along axes parallel to the plane, and a larger moment of inertia along the axis perpendicular to the plane. Or they can be rod shaped:  $I_1 \leq I_2 = I_3$ ; for example, decane [61, 1985, 2014] and eicosane have one small moment of inertia along the axis and two larger moments of inertia perpendicular to the axis.

Asymmetric top: irregular shape; for example,  $\text{H}_2\text{O}$  [0.716, 1.348, 2.665] has three unequal moments of inertia,  $I_1 \leq I_2 \leq I_3$ .

*Symmetry* The word symmetry means the same measure, which denotes harmony and beauty of the parts. It also plays a very important role in molecular architecture and material properties. The study of molecular symmetry is through a branch of group theory, that is, the point groups of rotations that leave one point



**Figure 4.9** Spherical top (methane), symmetric tops (benzene, decane), asymmetric top (water), clockwise from lower left

unchanged. A group is a set of operations, including the identity operator **E**, which does not change anything. For each operation, there is an inverse operation. The product of two operations is another operation in the group. There are four fundamental operations in a point group: rotation, reflection, inversion, and rotation-reflection.

*Rotation* around an axis. Has the symbol  $C_n$ , where  $n = 2$  means a twofold rotation by  $180^\circ$ ,  $n = 3$  means a threefold rotation by  $120^\circ$ ,  $n = 4$  is a fourfold rotation by  $90^\circ$ , and  $n = 6$  is a sixfold rotation by  $60^\circ$ .

*Reflection* about a plane. Has the symbol  $\sigma$ , and we distinguish between  $\sigma_h$  which is a reflection about a plane perpendicular to the principal rotation axis, and  $\sigma_v$  which is a reflection about one of the planes in the rotation axis.

*Inversion* through a central point. Has the symbol **i**.

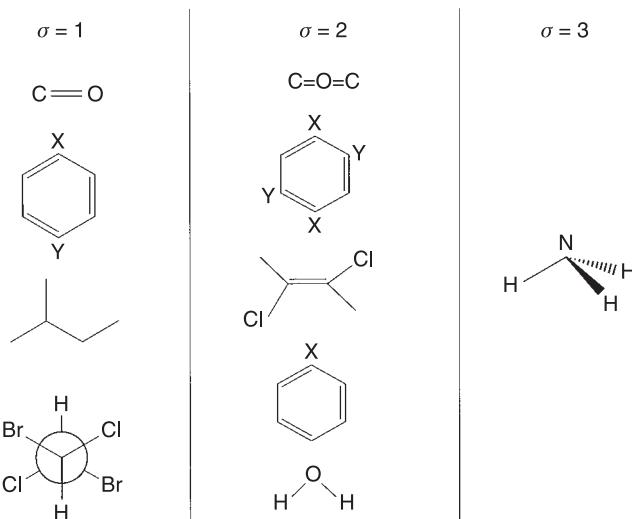
*Rotation-reflection*. Has the symbol  $S_n$ , which is a rotation followed by a reflection in the plane perpendicular to the axis of rotation.

The *identity* operator leaves everything in the original position, and has the symbol **E**.

Table 4.17 lists the symmetry point groups that contain the most interesting molecules. The order of a group is the number of operations, including rotations, reflections, inversion, and the identity operator. The symmetry number  $\sigma$  is the number of proper rotation operations that does not involve reflection or inversion. It is unfortunate that it shares the same symbol with the operation of reflection. The dipole moment is possible for only five of these low-symmetry groups. Figure 4.10 shows the low-symmetry molecules with  $\sigma = 1, 2$ , and  $3$ . The molecules of  $\sigma = 1$  that do not have rotational symmetry form the largest set, especially among larger molecules. Among the molecules with a rotational symmetry number of  $2$ , water and chlorobenzene are examples of  $C_{2v}$  with rotational symmetry around an axis lying in the plane of the paper; 1,4-dichloro-2,5-dibromobenzene is an example of  $C_{2h}$  around an axis through

Table 4.17 Symmetry groups of molecules

Group	Order	$\sigma$	Dipole moments	Examples
$C_1$	1	1	Yes	$\text{CHFCIBr}$
$C_s$	2	1	Yes	$\text{CH}_2=\text{CHCl}$ , 1-chloro-4-bromobenzene
$C_i$	2	1		$\text{XYZC}-\text{CXYZ}$
$C_{\infty v}$	$\infty$	1	Yes	$\text{HCl}$
$C_{2v}$	4	2	Yes	$\text{H}_2\text{O}$
$C_{2h}$	4	2		1,4-dichloro-2,5-dibromobenzene
$D_{\infty h}$	$\infty$	2		$\text{H}_2$
$C_{3v}$	6	3	Yes	$\text{NH}_3$
$D_{2h}$	8	4		$\text{C}_2\text{H}_4$ , <i>p</i> -xylene
$D_{3d}$	12	6		$\text{C}_2\text{H}_6$ , cyclohexane
$D_{3h}$	12	6		cyclo- $\text{C}_3\text{H}_6$ , 1,3,5-trichlorobenzene
$D_{4h}$	16	8		cyclo- $\text{C}_4\text{H}_8$
$D_{6h}$	24	12		$\text{C}_6\text{H}_6$
$T_d$	24	12		$\text{CH}_4$
$O_h$	48	24		$\text{SF}_6$
$I_h$	120	60		$\text{C}_{60}$

Figure 4.10 Low-symmetry molecules with  $\sigma = 1$ , 2, and 3

the center of the aromatic ring perpendicular to the plane of the paper; and  $\text{C}=\text{O}=\text{C}$  is an example of  $D_{\infty h}$  with a rotation axis perpendicular to the main axis. Ammonia is an example of the threefold symmetry of  $C_{3v}$ . Figure 4.11 shows the medium-symmetry molecules with  $\sigma = 4$ , 6, and 8. *p*-Xylene and ethylene have three rotational axes of  $D_{2h}$ : one axis through the two X atoms, one axis perpendicular to the two X atoms,

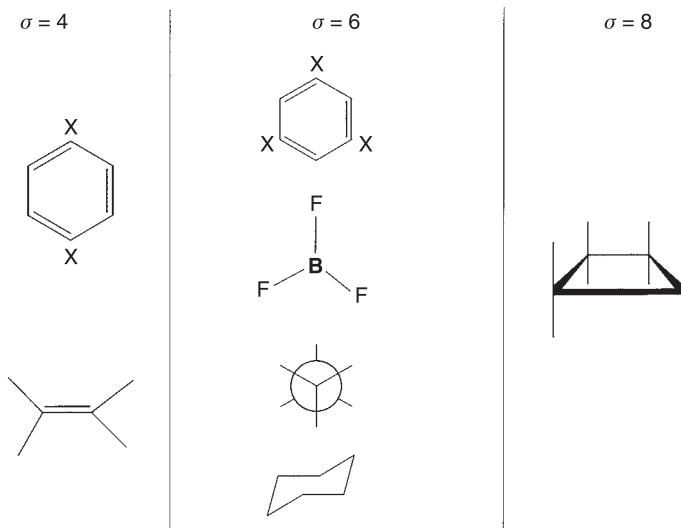


Figure 4.11 Medium-symmetry molecules with  $\sigma = 4$ , 6, and 8

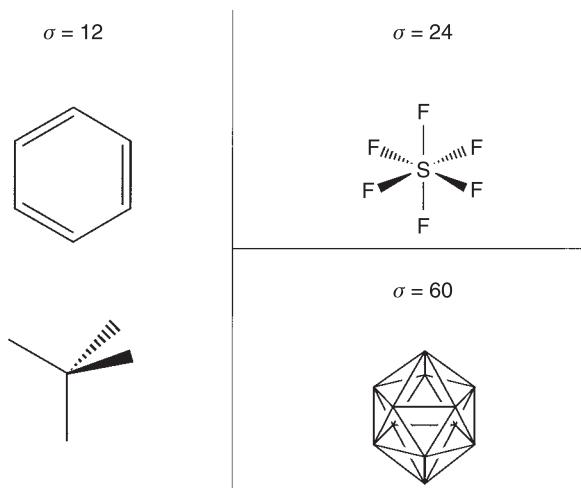


Figure 4.12 High-symmetry molecules with  $\sigma = 12$ , 24, and 60

and one axis perpendicular to the plane through the center of the aromatic ring. Cyclobutane is shown in the average planar form rather than the puckered form. Figure 4.12 shows the highest symmetry molecules with  $\sigma = 12$ , 24, and 60.

### Dipole Moments

The dipole moments of hydrogen halides decrease with the period of the halogen element; the increase in the bond length  $l$  is overpowered by the decrease

Table 4.18 Dipole moment variation with period

	<i>Dipole</i>		<i>Dipole</i>		<i>Dipole</i>
HF	1.826	NH <sub>3</sub>	1.471	<i>o</i> -Dichlorobenzene	2.50
HCl	1.109	PH <sub>3</sub>	0.574	<i>m</i> -Dichlorobenzene	1.72
HBr	0.827	HCN	2.93	<i>p</i> -Dichlorobenzene	0
HI	0.448			1,1-Dichloroethylene	1.34
H <sub>2</sub> O	1.84			<i>cis</i> -1,2-Dichloroethylene	1.90
H <sub>2</sub> S	0.97			<i>trans</i> -1,2-Dichloroethylene	0
H <sub>2</sub> Se	0.24				

in the charge  $q$ . The two chlorine atoms are  $60^\circ$  apart in *o*-dichlorobenzene, and pulling in roughly the same direction; the two chlorine atoms are  $120^\circ$  apart in *m*-dichlorobenzene and partially cancel each other; but the two chlorine atoms are opposed in *p*-dichlorobenzene and cancel each other totally. The same phenomenon is observed in *trans*-1,2-dichloroethylene, as the two charges of the chlorine atoms cancel each other (table 4.18).

**Docking molecules** When two molecules are far apart, the attractive and repulsive forces between them depend only on the distance; but when they come closer, the orientation and geometric fit between them become increasingly important. With the exception of electrostatic forces, most of the forces between molecules are short range in nature; thus, surface structures, such as the topography of elevations and depressions and the positions of the hydrogen bond sites, become controlling. The docking of molecules to achieve maximum adhesion involves positioning and rotating, which is critical in nonbonding adhesion, such as between an enzyme and a substrate and between a drug and a receptor site. The subject of drug design centers on the question of the shape and the configuration of the receptor site, and how to design a drug that would adhere sufficiently to it. Computational chemistry has become an important tool in drug discovery, particularly on the binding of drugs or pathogens to receptor sites. Molecular sieve adsorption and catalysis are also based on the premise that some molecules have the right shape to access narrow cavities, whereas other molecules with very similar properties are excluded.

The biochemical reactions in life are catalyzed by enzymes, which speed up the reaction and guide them in specific directions. Enzymes are proteins, and they are responsible for reactions such as breaking down food into energy and monomers (proteins into amino acids, starches into simple sugars, fats into fatty acids), synthesizing the monomers into tissues and organs, and generating energy to flex muscle and to pump the heart. They can be very specific: the enzyme urease only catalyzes the breakdown of urea; and the D-amino acid oxidase only oxidizes the D-amino acid, not the L-amino acid.

The actions of enzymes as catalysts depend on their three-dimensional conformation, or how the protein is folded into a three-dimensional object. A protein is said to be “denatured” if its three-dimensional conformation is altered, such as by heat or mechanical stirring, and is no longer biochemically active as a catalyst.

In 1972, Christian Anfinsen received the Nobel Prize in Chemistry for his principle of "thermodynamic determinism," that the amino acid sequence alone determines the three-dimensional structure. In the normal physiological state, the three-dimensional structure of a protein is the one in which the Gibbs free energy of the whole system is at a minimum. Many protein molecules have segments that are folded into a single-stranded  $\alpha$ -helix, which is held together by the hydrogen bonds between coils, called a secondary structure.

For example, acetylcholine,  $(\text{CH}_3)_3\text{N}-\text{CH}_2\text{CH}_2-\text{O}-\text{CO}-\text{CH}_3$ , is a neurotransmitter to skeletal muscles to make them contract. It is postulated that the nerve receptor to acetylcholine has a cavity with an anionic group to house the  $\text{N}^+$  end of the molecule, and another cavity to house the  $-\text{COO}-$  end, and the distance between these two cavities is 5 Å. Thus, for a protein to bind to the receptor, the optimal distance covered by the  $\text{N}-\text{CH}_2\text{CH}_2\text{O}-\text{C}$  chain is 5 Å. When the chain length is longer or shorter, the catalytic activity rapidly decreases to almost zero. Then consider the estrogen estradiol, which is secreted in the ovary to control the menstrual and reproductive cycles. It has the familiar structure of four fused rings that is common to cholesterol and the male hormone testosterone. The synthetic estrogen stilbestrol has only two isolated benzene rings connected by a diethyl-ethylene molecule, but has five times the activity of estradiol. It is postulated that the whole structure is not needed for receptor binding.

Morphine was known to the Sumerians and Egyptians as a very valuable pain killer in medicine; it is also a powerful narcotic that is habit forming. A great deal of effort has been spent in finding a derivative of morphine that has its good properties, and less of its bad properties. Then, in 1944, demerol and methadone were discovered, which are even more powerful than morphine, but bear no structural relation to morphine. The leading theory now is that both morphine and methadone have the same T-shape, and may fit the receptor.

Zeolites are crystalline aluminosilicates of alkali or alkaline earth elements, and are represented by the chemical composition  $\text{M}_{x/n}[(\text{AlO}_2)_x(\text{SiO}_2)_y]\cdot z\text{H}_2\text{O}$ , where  $x$  and  $y$  are integers with  $y/x \geq 1$ ,  $n$  is the valence of cation  $\text{M}$ , and  $z$  is the number of water molecules in each unit cell. The primary structural units of zeolites are the tetrahedra of silicon and aluminum,  $\text{SiO}_4$  and  $\text{AlO}_4$ . These tetrahedra can be assembled in regular three-dimensional crystalline frameworks that are found in nature, or can be synthesized in the laboratory. The molecular sieves are highly porous crystals of silica and alumina, and have molecular size pores that permit the passage of some molecules but not other molecules, depending on their molecular diameters. The main types of molecular sieve with industrial importance have pores or channels with 8, 10 and 12 atoms.

A zeolites have eight ring channels with 4.1 Å diameter, admit water and normal paraffins, but not iso-paraffins. Figure 4.13 shows that normal hexane has a smaller diameter than 2,3-dimethylbutane and can enter the narrow channels of A zeolites.

ZSM-5 with 10 ring channels of  $5.3 \text{ \AA} \times 5.6 \text{ \AA}$  diameter, admit benzene, toluene, *p*-xylene and mono-branched paraffins, but not double-branched paraffins or *m*- and *o*-xylenes. Figure 4.14 shows that *p*-xylene has a much smaller diameter than the *o*- and *m*-xylenes, and can enter the narrow channels of ZSM-5.

Faujasite X and Y, with 12 ring channels of 7.4 Å diameter, admit *o*- and *m*-xylenes, but not tri-isopropyl benzene.

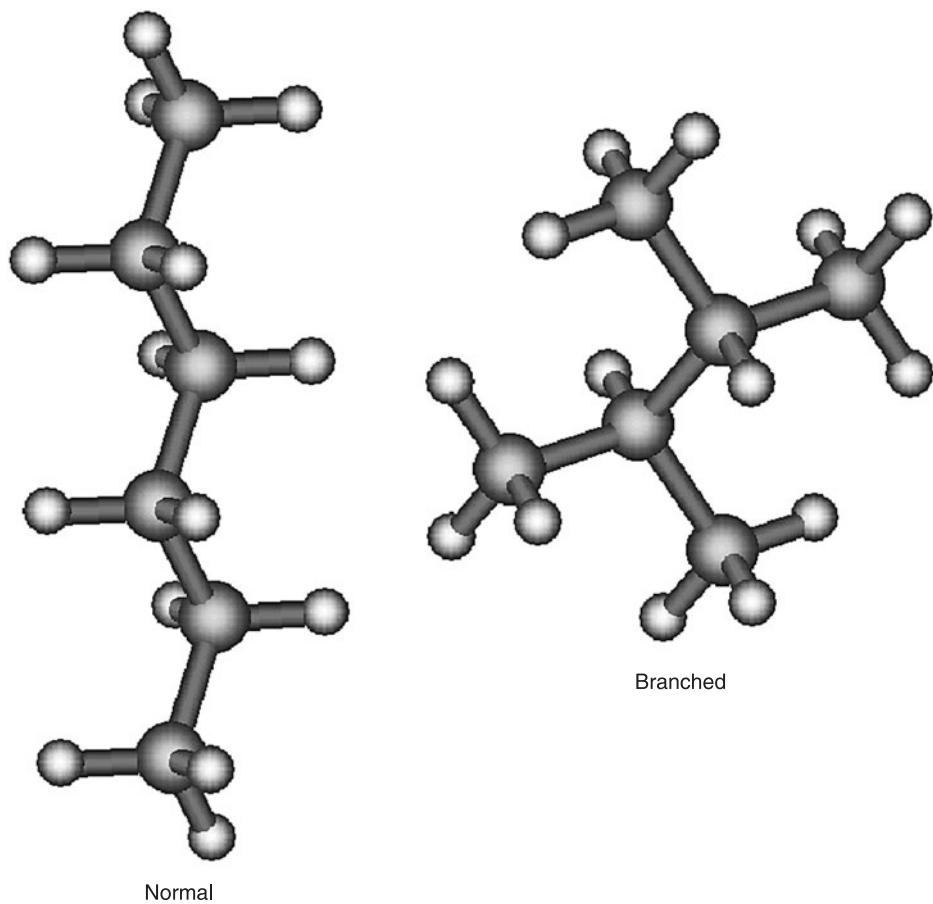


Figure 4.13 Shapes of normal and branched paraffins

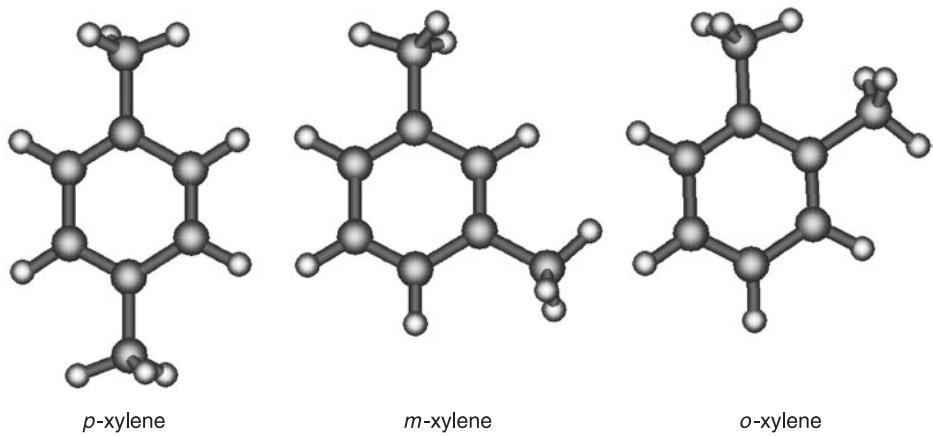


Figure 4.14 Shapes of *p*-, *m*- and *o*-xylene

## 4.2 Statistical Thermodynamics

The previous section is concerned mostly with an isolated molecule and its properties. When we have an ensemble of molecules in equilibrium with a particular volume and temperature, we find that the molecules are distributed among different energies and we are interested mostly in their average properties, which is the concern of statistical mechanics.

### Ensembles of Molecules

If each molecule in an ensemble of molecules can exist in various quantized states with energies  $E_1, E_2, E_3, \dots$ , then the probability  $p_j$  that a molecule will be found in the state with energy  $E_j$  is given by the Boltzmann distribution:

$$p_j \propto \exp(-E_j/kT) \quad (4.17)$$

where  $k$  is the Boltzmann constant, which is equal to the gas constant  $R$  divided by the Avogadro number, and  $T$  is the absolute temperature. The sum of all these distributions is called the partition function, which plays a central role in statistical mechanics:

$$Q = \sum_i \exp\left(-\frac{E_i}{kT}\right) \quad (4.18)$$

Since the sum of all the probabilities must equal one, the normalization constant for the probabilities is  $1/Q$ , so that the absolute value of the probability is

$$p_j = \frac{\exp(-E_j/kT)}{Q}$$

Figure 4.15 shows the Boltzmann distribution for several values of  $kT/E$  for a system where the states have evenly spaced energies. At low temperatures, most of the molecules can be found at the lowest energy states, with energy level equal to zero. When the temperature is increased, more and more molecules are promoted to higher energy states. When a molecule has several degrees of freedom, such as translations, rotations, and vibrations, each has its own quantum states and partition functions, and then the overall partition function is a product of all these separate partition functions:

$$Q = q_{\text{trans}} q_{\text{rot}} q_{\text{vib}} q_{\text{electronic}}$$

Many thermodynamic functions can be derived from the partition function of the canonical ensemble by a weighted average, or by differentiation of the partition function. For instance, the average energy of the ensemble can be given by a weighted

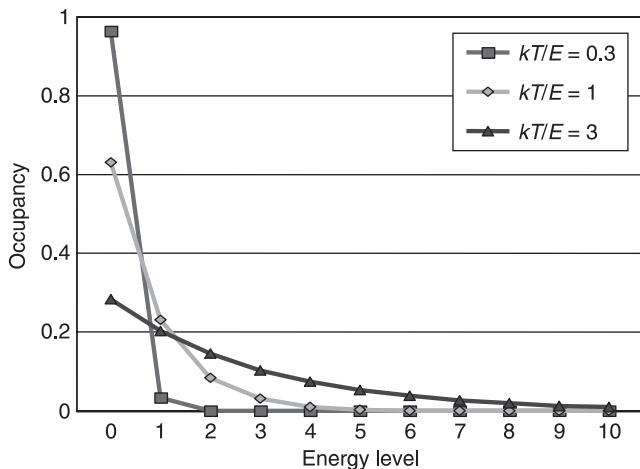


Figure 4.15 Boltzmann distribution with evenly spaced energy steps

average or by the alternative form by taking a partial derivative of  $Q$  by  $T$ :

$$\bar{E} = \sum p_j E_j = kT^2 \left( \frac{\partial \ln Q}{\partial T} \right) \quad (4.19)$$

The heat capacity at constant volume is by definition

$$C_v = \left( \frac{\partial \bar{E}}{\partial T} \right) \quad (4.20)$$

The Helmholtz free energy is

$$A = -kT \ln Q \quad (4.21)$$

The chemical potential of a component is given by

$$\mu_i = \left( \frac{\partial A}{\partial N_i} \right) = -kT \left( \frac{\partial \ln Q}{\partial N_i} \right) \quad (4.22)$$

The pressure of the system is given by

$$P = - \left( \frac{\partial A}{\partial V} \right) = kT \left( \frac{\partial \ln Q}{\partial V} \right)_{N,T} \quad (4.23)$$

The entropy  $S(N, V, T)$  is defined to be  $S = k \ln Q$ , which can be shown to be

$$S = -k \sum p_i \ln p_i$$

$$S = -\left(\frac{\partial A}{\partial T}\right) = kT \left(\frac{\partial \ln Q}{\partial T}\right) + k \ln Q \quad (4.24)$$

Let us consider a system of  $M$  ideal monatomic gas molecules in a cubic box kept at a constant temperature  $T$ . For a very dilute gas, where the molecules do not interact with one another, the quantum mechanical solution is a number of electronic wave functions with three quantum numbers  $n_x$ ,  $n_y$ , and  $n_z$  for the translational energies in three dimensions. The energy of a molecule for a set of quantum numbers, the observed average energy, and the heat capacity at constant volume are given by

$$E_{n_x, n_y, n_z} = \frac{h^2}{8\pi^2 m a^2} (n_x^2 + n_y^2 + n_z^2)$$

$$\bar{E} = \sum p_j (N, V, T) E_j (N, V, T) = U = \frac{3}{2} RT$$

$$C_v = \frac{\partial \bar{E}}{\partial T} = \frac{3}{2} R$$

The value of  $C_v$  should be zero at absolute zero temperature, which is required by the third law of thermodynamics, but the translational energy has a very small quantum step, so that it is fully activated even at very low temperatures, and  $C_v = 3R/2$  is the fully activated value.

In a diatomic molecule with rotation energy and rotation quantum number  $J$ , there is an energy proportional to  $J(J + 1)$ , and there are  $2J + 1$  configurations (referred to as degeneracy). The probability of finding a molecule in energy state  $J$  becomes

$$p_J = (2J + 1) \exp\left[-\frac{J(J + 1)\Theta_{\text{rot}}}{kT}\right] / q_{\text{rot}}$$

$$q_{\text{rot}} = \frac{1}{\sigma} \sum_{J=0}^{\infty} (2J + 1) \exp(-E_J/kT) \quad (4.25)$$

Here,  $\Theta_{\text{rot}} = h^2/8\pi^2Ik$  is the characteristic rotation temperature, where  $I$  is the moment of inertia of the molecule, and  $\sigma$  is the symmetry number. A monatomic molecule has a vanishing moment of inertia  $I$ , and thus an infinite characteristic temperature  $\Theta$ , so it has no rotation thermal energy. The value of  $I$  for a diatomic molecule is  $I = \mu R^2$ , where  $R$  is the distance between the atoms and the equivalent mass  $\mu = m_1 m_2 / (m_1 + m_2)$ . The hydrogen molecule has the smallest moment of inertia, and a characteristic temperature of 85 K. Therefore, the temperature dependence of the heat capacity for the hydrogen molecule is important only in the very low range of  $T < \Theta$ , and is more than 92% activated at temperatures higher than 85 K. All the other multiaatomic molecules have higher moments of inertia, and thus lower characteristic temperatures. Thus, for all practical temperatures, the rotation energy of a molecule is

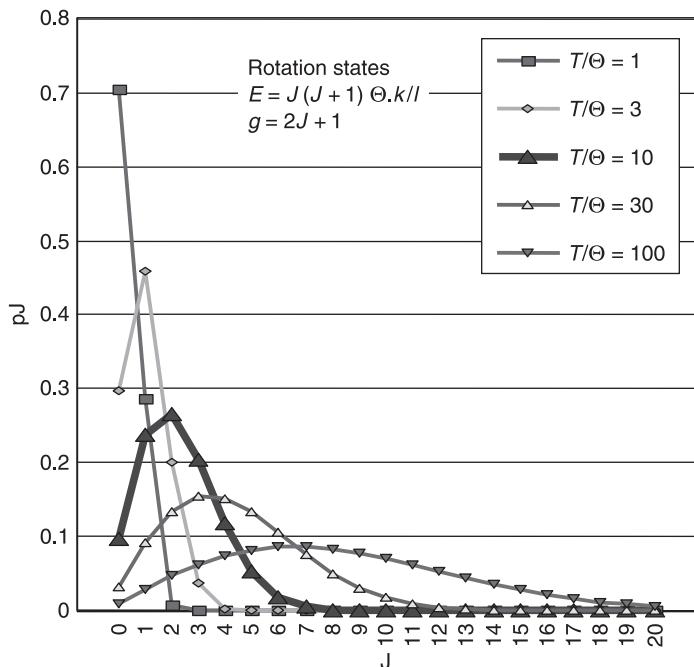


Figure 4.16 Boltzmann distribution of rotation states

equal to the fully activated value of  $R$  for a linear molecule, and  $3R/2$  for a nonlinear molecule. Here, the symmetry number of a molecule has the value  $\sigma = 1$  when the two atoms are not the same, such as HCl, but  $\sigma = 2$  when the two atoms are the same, such as  $\text{Cl}_2$ . When the temperature is much higher than the  $\Theta_{\text{rot}}$ , the value of  $q_{\text{rot}} = T/\sigma\Theta_{\text{rot}}$ . Figure 4.16 shows this distribution for a variety of values for  $T/\Theta_{\text{rot}}$  from 1 to 100. This shows that as the value of  $T/\Theta_{\text{rot}}$  increases, the distribution shifts to the higher  $J$  states with higher energies. Thus, a symmetrical molecule has a lower entropy by  $\Delta S = R \ln(\sigma)$  than a similar asymmetrical molecule. This difference is not more than 5–10% of the entropy of a gas at 25 °C, but it can be twice as large as the entropy of fusion, such as for methane and neopentane.

In a system of molecules with vibrational degrees of freedom, each degree of freedom has a vibrational frequency  $\nu_i$ . The characteristic vibration temperature  $\Theta_{\text{vib},i} = h\nu_i/k$  and the partition function for each vibration frequency is

$$q_i = \frac{\exp(-\Theta_{\text{vib},i}/2T)}{[1 - \exp(-\Theta_{\text{vib},i}/T)]} \quad (4.26)$$

$$C_v = R \left( \frac{\Theta_{\text{vib}}}{T} \right)^2 \frac{e^{-\Theta/T}}{(1 - e^{-\Theta/T})^2}$$

The value of the characteristic temperature is proportional to the vibration frequency  $\nu$ , which is proportional to  $(1/R)\sqrt{K/\mu}$ , where  $K$  is the stiffness of the bond as a spring.

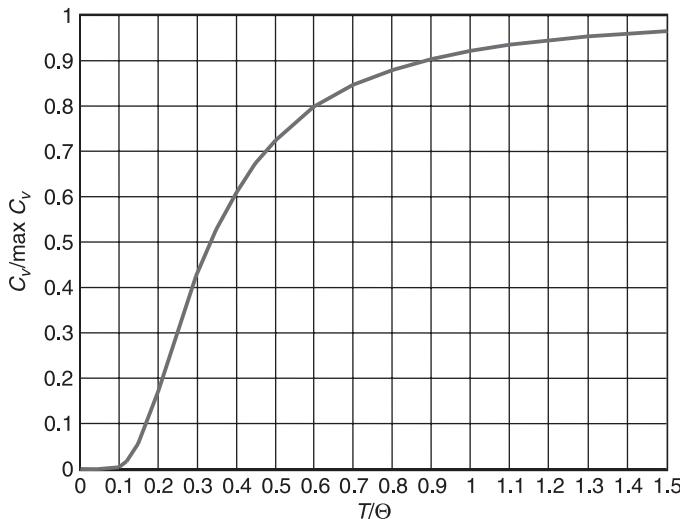


Figure 4.17 Vibration heat capacity in  $R/2$  and  $T/\Theta$

Thus, a soft and heavy material has low vibrational frequencies, and low characteristic temperatures, so they are fully activated even at low temperatures; for instance, lead has a characteristic temperature of 88 K. However, a brittle and light material would have a higher characteristic temperature and be fully activated only when the temperature is very high; for instance, diamond has a characteristic temperature of 1860 K. Figure 4.17 is a plot of  $C_v/C_{v \text{ max}}$  and temperature in an S-shaped curve, so that  $C_v = 0$  when  $T/\Theta$  is close to zero, and  $C_v$  rises to the full value of  $R/2$  for each degree of freedom when  $T/\Theta$  goes to infinity. The value of  $C_v$  is 50% activated when  $T/\Theta = 0.335$ , and 92% activated when  $T/\Theta = 1$ .

For molecules with three or more atoms, the number of vibrational frequencies is  $3n - 5$  for a linear molecule and  $3n - 6$  for a nonlinear molecule. The partition function for all vibration frequencies is the product of all these  $q_i$ , and the heat capacity for all vibration frequencies is the sum of their individual contributions:

$$q_{\text{vib}} = \prod_j \left( \frac{e^{-\Theta_j/2T}}{1 - e^{-\Theta_j/T}} \right) \quad (4.27)$$

$$C_v = R \sum_j \left( \frac{\Theta_j}{T} \right)^2 \frac{e^{-\Theta_j/T}}{(1 - e^{-\Theta_j/T})^2}$$

The heat capacity of a molecule is equal to the sum of all the contributions from translation, rotation, vibration, and electronic degrees of freedom (table 4.19).

### Molecular Dynamics and Monte Carlo

The examples in the previous section refer to simple ideal gases, where quantum mechanics can be used to compute the wave functions of each solution of the

Table 4.19 Constituents of heat capacity

Phase	Atoms per molecule	Translation	Rotation	Vibration	Total degrees of freedom	$C_v$ max
Gas	1	3	0	0	3	$3R/2$
Gas	2	3	2	1	6	$3R$
Gas	$n$ , linear	3	2	$3n - 5$	$3n$	$3nR/2$
Gas	$n$ , nonlinear	3	3	$3n - 6$	$3n$	$3nR/2$
Solid	$n$	0	0	$6n$	$6n$	$3nR$

Schrödinger equation. When the molecule is too large and difficult for quantum mechanical calculations, or the molecule interacts with many other molecules or an external field, we turn to the methods of molecular mechanics with empirical force fields. We compute and obtain numerical values of the partition functions, instead of precise formulas. The computation of thermodynamic properties proceeds by using a number of techniques, of which the most prominent are the molecular dynamics and the Monte Carlo methods.

We begin by considering an assembly of  $N$  particles and marking the initial momentum and position of each molecule as the set  $(\mathbf{p}^N(0), \mathbf{r}^N(0))$ , where  $\mathbf{p}^N(0)$  is a vector of the initial  $(p_x, p_y, p_z)$  momentum of every particle and  $\mathbf{r}^N(0)$  is a vector of the initial  $(x, y, z)$  position of every particle. Thus, for a system with  $N$  monatomic molecules, the number of variables would be  $6N$ . The equation of motion is usually written in the Hamiltonian form by defining the Hamiltonian as the sum of the potential and the kinetic energies of the particles:

$$H = U(r_1, r_2, \dots, r_{6N}) + \frac{1}{2} \left( \frac{p_1^2}{m_1} + \dots \right) \quad (4.28)$$

The variable  $p_1$  is the momentum of the particle, which is equal to  $m_1 v_1$ . The time rates of change of these parameters  $r$  and  $p$  follow Newton's law of motion, and are given by

$$\begin{aligned} \frac{dp_i}{dt} &= f_i = - \left( \frac{\partial H}{\partial r_i} \right) \quad i = 1, 2, 3, \dots, 6N \\ \frac{dr_i}{dt} &= v_i = \frac{p_i}{m_i} = \left( \frac{\partial H}{\partial p_i} \right) \end{aligned} \quad (4.29)$$

The force term  $f_i$  consists of all the forces that attract and repulse particle  $i$ , including all the bonding and the nonbonding forces plus external fields, and is equal to the negative of the partial derivative  $-\partial U/\partial r_i$ . The solution to this set of  $6N$  differential equations can be considered as the path of the system through time as  $(\mathbf{p}^N(t), \mathbf{r}^N(t))$ . The system may be far away from equilibrium at the beginning, but, as time goes by, the system should evolve gradually towards an equilibrium. The value of a thermodynamic variable of this system, such as  $A(\mathbf{p}^N(t), \mathbf{r}^N(t))$ , will undergo short-term fluctuations

in the start-up phase, followed by a long-term approach to equilibrium. From any arbitrary starting point, the system should settle down to the same equilibrium phase after a sufficiently long period of time. Then the time average of this function in the equilibrium phase should have the same value as the experimental measurements:

$$A_{\text{avg}} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \quad (4.30)$$

The practical difficulties of such a calculation are enormous, especially if  $N$  is on the order of Avogadro's number, that is,  $\sim 6 \times 10^{23}$ . Boltzmann and Gibbs invented the methods of statistical mechanics by replacing this time average of a single system with time with an ensemble average of a large number of replications of this system at the same instance of time. This is known as the ergodic hypothesis:

$$\langle A \rangle = \iint A(p^N, r^N) \rho(p^N, r^N) dp^N dr^N \quad (4.31)$$

Here,  $\langle A \rangle$  represents the ensemble average of  $A$ , and  $\rho$  is the probability density of the state represented by  $(p^N, r^N)$ . Under conditions of constant number of particles, volume and temperature, the probability density is the Boltzmann distribution

$$\rho(p^N, r^N) = \frac{\exp[-E(p^N, r^N)/kT]}{Q} \quad (4.32)$$

The molecular dynamics method is based on the time evolution of the path  $(p^N(t), r^N(t))$  for each particle to feel the attractions and repulsions from all other particles, following Newton's law of motion. The simplest case is a dilute gas following the "hard sphere" force field, where there is no interaction between molecules except during brief moments of collision. The particles move in straight lines at constant velocities, until collisions take place. For a more advanced model, the force fields between two particles may follow the Lennard-Jones 6-12 potential, or any other potential, which exerts forces between molecules even between collisions.

Consider  $N$  particles in a cubic box of volume  $V$ , so the energy of particle  $j$  is the sum of all the potential energies exerted by the other  $N - 1$  particles, and the potential energy of the entire system is the sum of all the particle energies:

$$\begin{aligned} \varepsilon_i &= \sum_{j \neq i}^N r_{ij} \varepsilon(r_{ij}) + \frac{mv_i^2}{2} \\ E &= \frac{1}{2} \sum_{i=1}^N \varepsilon_i = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N r_{ij} \varepsilon(r_{ij}) + \sum_{i=1}^N \frac{mv_i^2}{2} \end{aligned} \quad (4.33)$$

The factor 1/2 in front of the double summation is to take care of double counting. All collisions are considered to be completely elastic. To integrate forward to the next position, the computer replaces continuous time with small time steps for integration,

and the step size may be 1 fs, or  $10^{-15}$  s, and the length of the entire run may be 100,000 time steps, to simulate a run of 100 ps, or  $100 \times 10^{-12}$  s.

Diffusion can be modeled as a random walk in three dimensions, and the value of the diffusion coefficient can be computed by the correlation formula

$$D = \frac{1}{3} \lim_{t \rightarrow \infty} \frac{\langle |r(t) - r(0)|^2 \rangle}{2t} \quad (4.34)$$

The parameter  $\langle |r(t) - r(0)|^2 \rangle$  is the ensemble average of the square of the distances between the initial and the later positions for each particle in the system, and the factor of 3 takes care of the three-dimensional nature of the random walk. Given an ensemble of  $M$  particles, the correlation function of properties  $r(t)$  and  $r(0)$  is given by

$$\langle |r(t) - r(0)|^2 \rangle = \frac{1}{M} \sum_{i=1}^M \{ [x_i(t) - x_i(0)]^2 + [y_i(t) - y_i(0)]^2 + [z_i(t) - z_i(0)]^2 \}$$

The transport coefficients of diffusion, heat conductivity, and viscosity can all be computed by the method of correlation functions.

The Monte Carlo method also starts from an initial configuration of the positions, but does not consider the momenta ( $\mathbf{r}^N(0)$ ). Next, a succession of configurations, kept at a constant temperature  $T$ , are computed as a Markov process:

$$(\mathbf{r}^N(0)), (\mathbf{r}^N(1)), (\mathbf{r}^N(2)), (\mathbf{r}^N(3)), (\mathbf{r}^N(4)), \dots, (\mathbf{r}^N(j)), (\mathbf{r}^N(j+1)), \dots$$

In a Markov process, the evolution of the configuration depends randomly only on the immediate previous configuration, and not on earlier history. The configuration ( $\mathbf{r}^N(j+1)$ ) is assumed to be a random perturbation of the previous configuration ( $\mathbf{r}^N(j)$ ). For instance, in an assembly of  $N$  particles in a box with volume  $V$ , one of the particles is chosen at random to take a random walk in some direction by a small step, and the energy of the new configuration is calculated. To ensure that the ensemble would be consistent with the Boltzmann distribution, a procedure was introduced by Metropolis to screen each new configuration generated to see whether it should be rejected or admitted. The criterion is based on the Boltzmann factor,  $\exp(-U(r^N)/kT)$ , where we evaluate  $U$  as the potential energy alone instead of the sum of the potential and kinetic energies. When a new random configuration has a lower energy than the previous configuration, it is accepted; when the new configuration has a higher energy than the previous configuration, then a random number  $x$  between zero and one is generated, and the new configuration will be accepted if  $\exp(-\Delta U/kT) < x$  and rejected otherwise. Thus, a new configuration with a higher energy is not simply eliminated, but has a lower probability of success. This procedure will ensure that the ensemble evolves toward lower energy states, which are more probable than higher energy states, and eventually to the lowest energy ensemble consistent with the Boltzmann distribution. The equilibrium stage may be reached in something like 100,000 configurations. After reaching the equilibrium stage, the ensemble average of a property is computed

by taking an additional  $M$  steps:

$$\langle A \rangle = \frac{1}{M} \sum_{i=1}^M A(r_i^N) \quad (4.35)$$

The equation of state for a gas can be computed by approaching an ensemble with a minimum in potential energy. The pressure is computed by supplementing the perfect gas pressure with the intermolecular potentials

$$P = \frac{RT}{V} + \frac{\sum \sum r_{ij} f(r_{ij})}{3V}$$

The phase equilibrium between a liquid and a gas can be computed by the Gibbs ensemble Monte Carlo method. We create two boxes, where the first box represents the dense phase and the second one represents the dilute phase. Each particle in the boxes experiences a Lennard–Jones potential from all the other particles. Three types of motion will be conducted at random: the first one is particle translational movement in each box, the second one is moving a small volume from one box and adding to the other box, the third one is removing a particle from one box and inserting in the other box. After many such moves, the two boxes reach equilibrium with one another, with the same temperature and pressure, and we can compute their densities.

These two methods are different and are usually employed to calculate different properties. Molecular dynamics has a time-dependent component, and is better at calculating transport properties, such as viscosity, heat conductivity, and diffusivity. Monte Carlo methods do not contain information on kinetic energy. It is used more in the lattice model of polymers, protein structure conformation, and in the Gibbs ensemble for phase equilibrium.

### Entropy, Disorder, and Stability

In thermodynamics, entropy change is defined in a reversible process as

$$dS = \delta q_{\text{rev}}/T \quad (4.36)$$

The third law of thermodynamics states that, for a perfect crystal at absolute zero temperature, the value of entropy is zero. The entropy of a molecule at other temperatures can be computed from the heat capacities and heats of phase changes using

$$S(T) = \int_0^T \frac{C_p(T')}{T'} dT' + \sum \frac{\Delta H_{\text{tr}}}{T_{\text{tr}}}$$

The second term sums all the heats of phase transition divided by the temperatures of phase transition.

In statistical mechanics, Boltzmann defined entropy of a system as

$$S = k \ln W \quad (4.37)$$

where  $W$  is the number of possible arrangements for the system. Let us consider a box with  $m$  cells and  $n$  identical balls, and each cell can be empty or be occupied by one ball. The number of ways to place the first ball is  $m$ , and the number of ways to place the second ball is  $m - 1$ . Since it is arbitrary which ball to call first and which ball to call second, the number of combinations is really  $m(m - 1)/2$ . When we put in more and more balls, the number of combinations is the binomial number

$$W = \binom{m}{n} = \frac{m!}{(m - n)!n!} \quad (4.38)$$

For example for a box with four cells containing four balls, there is only one configuration and  $W = 1$ ,  $S = 0$ . When the box expands further to 64 cells,  $W = 64!/(60!4!) = 635,376$  and  $S/k = 13.36$ . When the box expands further to a very large number  $m$ , we use Stirling's approximation to calculate the number of combinations:

$$\ln(m!) \cong (m + 1/2) \ln(m) - m \cong m \ln(m) - m \cong m \ln(m) \quad (4.39)$$

The first formula is more accurate, and is good to 1.0% when  $m > 35$ , but the second formula is less accurate, and is good to 2.9% when  $m = 35$ , but the third formula has an error of 1.9% even at Avogadro's number of  $\sim 6 \times 10^{23}$ .

When we use Stirling's formula for the binomial number, we have

$$\begin{aligned} W &= \frac{m!}{(m - n)!n!} \\ \ln W &= m \ln m - m - (m - n) \ln(m - n) + m - n - n \ln(n) + n \\ &= m \ln(m) - (m - n) \ln(m - n) - n \ln(n) \\ &= -m \left[ \frac{m - n}{m} \ln \left( \frac{m - n}{m} \right) + \frac{n}{m} \ln \left( \frac{n}{m} \right) \right] \\ &= -m [x_1 \ln(x_1) + x_2 \ln(x_2)] \end{aligned} \quad (4.40)$$

Here, we designate  $x_1$  as the volume fraction of molecules and  $x_2$  as the volume fraction of holes; then, we have the entropy of mixing of balls and holes as

$$S/k = -m[x_1 \ln(x_1) + x_2 \ln(x_2)] \quad (4.41)$$

This formula for entropy is pretty accurate when  $m = 100$ . For instance in  $m = 100$  and  $n = 50$ , the value of  $\ln W = 66.78$ , but the formula of  $-100[0.5 \ln 0.5 + 0.5 \ln 0.5] = 69.31$ , which is 3.8% off.

In an alternate method, the entropy of an ensemble can be calculated by the partition function  $Q$ :

$$S = k \ln Q + kT \left( \frac{\partial \ln Q}{\partial T} \right) N, V$$

*Molecular entropies* For a perfect monoatomic gas, there is only translational motion. According to quantum mechanics, the translational energy of molecules in a box is quantized and the size of the quantum is proportional to the reciprocal of the atomic weight. Heavier gases have smaller gaps and the number of states available and degeneracies are greater.

$$E(n_x, n_y, n_z) = \frac{h^2 (n_x^2 + n_y^2 + n_z^2)}{8ma^2} \quad n_x, n_y, n_z = 1, 2, 3, \dots$$

$$g = \frac{(2\pi ekTm)^{3/2} V}{h^3} \quad (4.42)$$

The Sackur–Tetrode equation was derived for the molar entropy for a monatomic gas to be

$$S \text{ (J/(K mol))} = \frac{3}{2} R \ln M + \frac{3}{2} R \ln T + R \ln V + 11.11$$

$$S \text{ (J/(K mol))} = \frac{3}{2} R \ln M + \frac{5}{2} R \ln T - R \ln P - 9.69$$

Here,  $M$  is the molecular weight of the molecule,  $V$  is in L/mol and  $P$  is in atmospheres. The second equation is obtained from the first equation by substitution of the perfect gas law  $V = RT/P$ . The thermal energy of translation in three-dimensional space is  $3R/2$ , which is associated with the terms of  $T$  and  $M$  in the entropy equation. Heavier atoms have larger entropies than lighter atoms, as higher  $M$  leads to smaller steps for quantum jumps and, thus, a higher density of states. In comparison, the experimental entropies of the noble gases at 25 °C and 1 atm are shown in figure 4.18; they form a very satisfactory straight line with the form of  $S = 109 + 12.5 \ln(M)$ .

For a gas with more than one atom, rotations and vibrations are added to the translational motions. Figure 4.19 shows the entropies of the binary gases H<sub>2</sub>, N<sub>2</sub>, and O<sub>2</sub> and of the halogen elements; they are less than twice the entropies of the monatomic elements at the same value of  $\ln(M)$ , and they fit the equation  $S = 109 + 27.1 \ln(M)$ . This is due to the fact that 1 mol of diatomic gas is much less disordered in comparison with 2 mol of monatomic gas. Consider placing two identical balls in a 10 × 10 grid, where the number of configurations is  $W = (100 \times 99)/2$  or 4950. If the two balls are linked together, then the first ball still has 100 places to fill, but the second ball can take only four places, so  $W = (100 \times 4)/2 = 200$ . In fact, if the first ball is on one of the 32 cells on the edge of the grid, then the second ball has only three cells to fill; and if the first ball is on one of the four cells on the corner of the grid, then the second ball has only two cells to fill. So the total number of configurations is reduced to  $W = 180$  (see table 4.20).

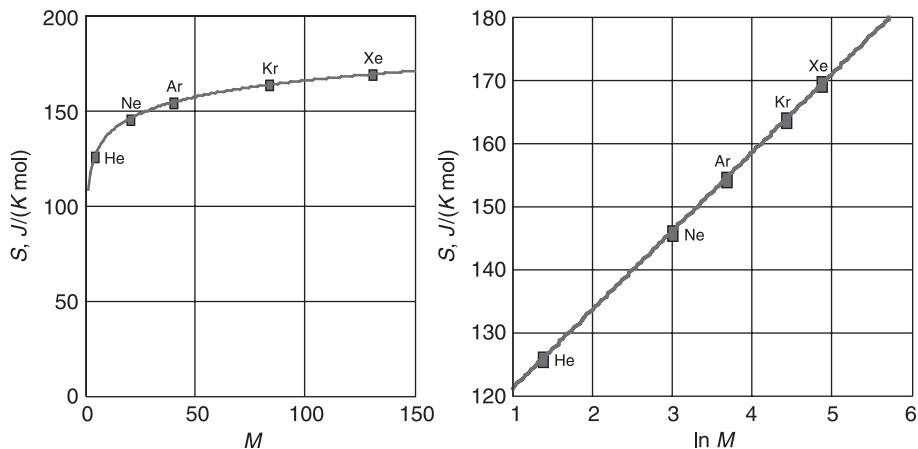


Figure 4.18 Entropy of monatomic molecules: noble gases

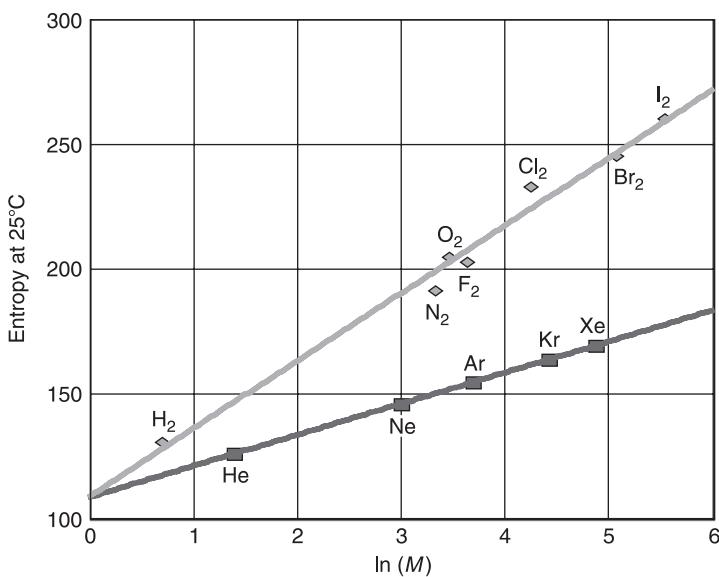
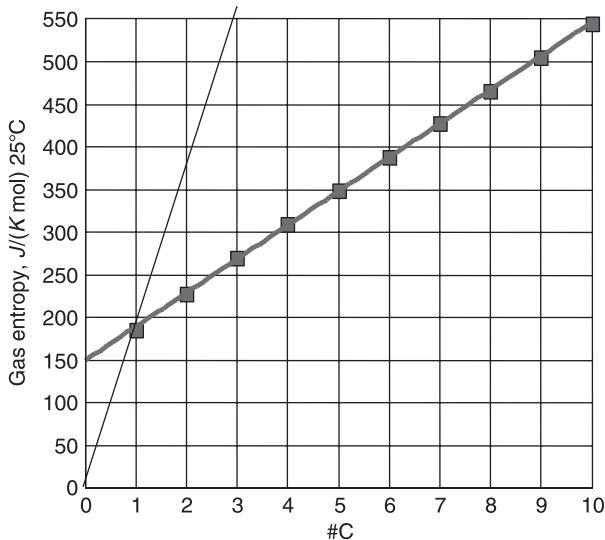


Figure 4.19 Entropy of diatomic molecules

Linking balls together makes the entropy go down, as the result is more orderly, despite the additional energetic motions of vibration and rotation. Figure 4.20 shows the entropies of normal paraffins, which shows that polymerization gives a polymer of much lower entropy than an equivalent number of monomers, so that  $S(C_nH_{2n+2}) \ll nS(CH_4)$ .

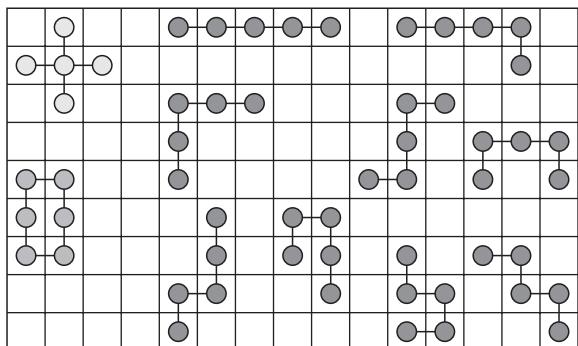
**Table 4.20 Configuration calculation results**

No. of balls	<i>W</i> for independent balls	<i>W</i> , linked flexible hinges	<i>W</i> , linked rod-like hinges
1	100	100	100
2	4,950	180	180
3	161,700	484	160
4	3,921,225	1,616	140

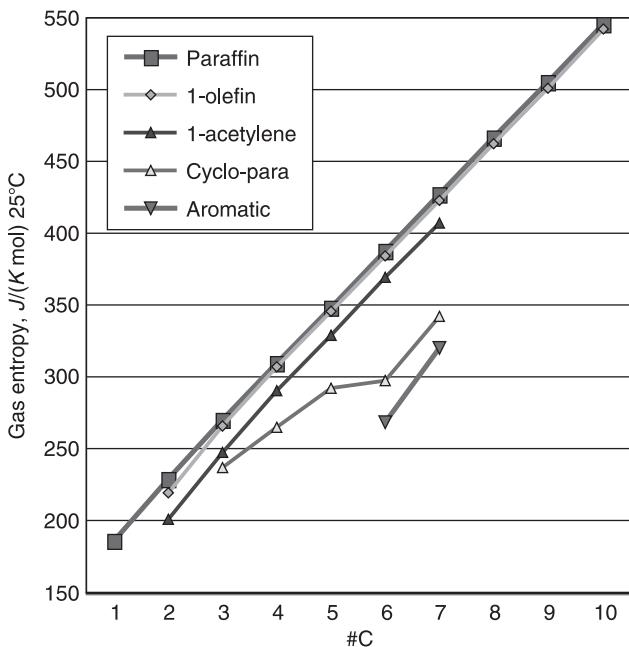
**Figure 4.20** Entropy of normal paraffin gases at 25 °C

Flexibility is another source of disorder and entropy. When we chain three or more balls together, we distinguish between a flexible chain and a rigid straight rod. The rod in the  $10 \times 10$  grid can have either a horizontal or a vertical orientation. There are eight ways to place a three-ball horizontal rod on row 1, and there are 10 rows, so that there are 80 horizontal configurations; since there are also 80 vertical configurations, the total configuration of rods is 160. Figure 4.21 shows that, on a two-dimensional grid, *n*-pentane has many more configurations than neopentane or cyclohexane. The data in figure 4.22 show that the flexible normal paraffins have much higher entropies than the rigid aromatics, whereas the 1-olefins, 1-acetylenes, and the cycloparaffins fall in between. Let us also consider the set of six-carbon hydrocarbons (table 4.21).

Hexane is expected to be mostly in the all-trans configuration, but can be rotated to more energetic configurations with some gauche configurations at sufficiently high temperatures. When hexane loses two H atoms to become 1-hexene, the entropy decreases by 3.8 J/(K mol), which is mainly due to the fact that the double bond does not rotate, whereas the single bond rotates with a small energy barrier. When 1-hexene cyclizes



**Figure 4.21** Configurations of chain branching and flexibility



**Figure 4.22** Entropies of hydrocarbons at 25 °C: paraffins, olefins, alkynes, cyclo-paraffins, and aromatics

**Table 4.21 Entropy values of six-carbon hydrocarbons**

<i>S, experimental, J/(K mol)</i>	
Hexane	388.4
1-Hexene	384.6
Cyclohexane	298.3
Benzene	269.2

**Table 4.22 Symmetry versus entropy for the pentanes**

Molecule	Symmetry $\sigma$	$W$ , 2-D configurational	$S$ at 298 K, experimental, $J/(K \text{ mol})$
Pentane	2	9	349.45
Isopentane	1	2	343.63
Neopentane	12	1	305.89

into cyclohexane, there is a bigger drop of entropy by 86 J/(K mol), which is mostly due to the fact that the ring is less flexible than the chain. However, cyclohexane is capable of several conformations, as the most stable is the chair form, in comparison with the less stable boat and twisted boat forms. Benzene has only one configuration, and the lowest entropy. So, flexible molecules have more configurations, and have higher entropies than rigid molecules.

Symmetry leads to more order, which in turn leads to a decrease in entropy. Consider the three isomers pentane, isopentane, and neopentane, with symmetry numbers  $\sigma = 2$ , 1 and 12, respectively (table 4.22). When placed in a two-dimensional lattice, neopentane can be represented as a cross, and there is only one configuration possible,  $W = 1$ . Isopentane can have  $W = 2$ , but the flexible pentane has nine configurations. The experimental entropies at 25 °C support the expectation that pentane should have a somewhat higher entropy than isopentane. The even more rigid neopentane may be expected to have a smaller entropy due to the lower flexibility; however, the higher symmetry contributes to an additional large drop of entropy according to the formula  $\Delta S = -R \ln(\sigma) = 20.7$ .

*Volume change and phase transition* There is another formulation of the binomial number that is used for the entropy change due to volume expansion. From equation (40), we obtain in the limit that when  $m \gg n$

$$S = k \ln (W) \cong nk \ln(m/n) = R \ln(V)$$

The parameter  $V$  is the volume per mole, or  $m/n$ . Thus, in an isothermal process, when a mole of solid sublimates to make a perfect gas, the entropy changes with the volume according to

$$\Delta S = R \ln(V_2/V_1) \quad (4.43)$$

For a perfect gas not in an isothermal process, we have

$$\Delta S = C_v R \ln(T_2/T_1) + R \ln(V_2/V_1) \quad \text{at constant pressure}$$

or

$$\Delta S = C_p R \ln(T_2/T_1) - R \ln(P_2/P_1) \quad \text{at constant volume}$$

Table 4.23 Entropy versus volume expansion

Mixing or volume expansion (%)	$\Delta S, \text{J}/(\text{K mol})$
0	0
10	2.70
20	4.16
30	5.08

In the phase transition of condensation from gas to liquid, the biggest source of entropy change is that of volume change. If there is a volume change of a factor of 1000, then the entropy change is of the order of  $R \ln(1000) = 57.4 \text{ J}/(\text{K mol})$ . This is 78% of the entropy change of vaporization of argon, at  $73.8 \text{ J}/(\text{K mol})$ . In the phase transition of melting from solid to liquid, the sources of entropy increase include a slight volume expansion, a loss of orientation order, and a loss of positional rigidity. The volume expansion, which may be as high as 30%, creates holes in the liquid. Since the volume expansion is small, we need to use the more precise mixing equation. If we designate  $x_1$  as the fractional volume of molecules and  $x_2$  as the fractional volume of holes, we have the entropy of mixing of holes as

$$S/R = -[x_1 \ln(x_1) + x_2 \ln(x_2)] \quad (4.44)$$

See table 4.23; this relation is shown in figure 4.23.

In the crystalline solid, the molecules are frozen from the possibility of free rotation, so all the molecules are oriented in the same direction; but they are randomly oriented in the liquid. Figure 4.24 shows a representation of each molecule as a ball with a handle attached, so that all the handles are pointed upwards in a solid with crystal order. In the melt, there is an unfreezing of the rotational degree, so the molecules have random orientations. This rotational disorder gives rise to an entropy term, which is related to  $\sigma$ , the number of rotation configurations allowed or the symmetry number. A symmetrical molecule is more ordered, and has a lower entropy than an asymmetrical molecule by  $R \ln(\sigma)$ . Let us consider the molecule 1,2,4-trimethylbenzene; this does not have any rotational symmetry, so there is only one configuration and  $\sigma = 1$ . In comparison, 1,3,5-trimethylbenzene has a symmetry number of 6, so the entropy of melting is reduced by  $\Delta S = -R \ln(6)$ , or an entropy of  $14.9 \text{ J}/(\text{K mol})$  lower. Thus, symmetric molecules have smaller entropies in the melt than equivalent asymmetrical molecules. It will be shown later that, as a consequence, symmetrical molecules tend to have much higher melting points.

*Solution of long-chain molecules* When two liquids mix to form a mixture, the entropy change is similar to that of the volume expansion, as long as the solute molecules have the same size as the solvent molecules and are randomly distributed. But when the solute forms long-chain molecules, the correct method of calculating the entropy was given by Flory. First consider a lattice model where the solvent and the solute molecules have the same volume. Let  $n_1$  and  $n_2$  be the number of solvent and

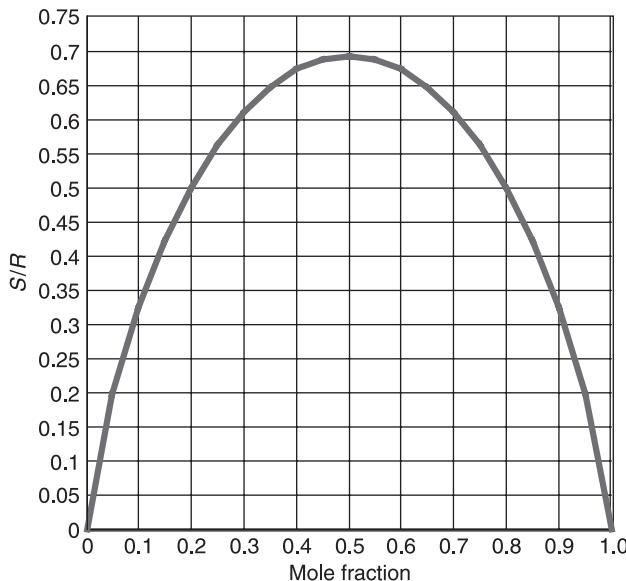


Figure 4.23 Entropy of mixing

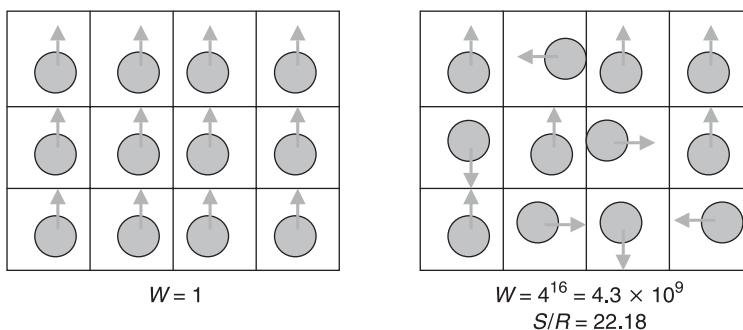


Figure 4.24 Conformations of orientation in liquids and solids

solute molecules, let  $v_1$  and  $v_2$  be the volumes that they occupy, and let  $f_1$  and  $f_2$  be the volume fractions. The entropy of the system is given by

$$\begin{aligned}
 f_1 &= \frac{v_1}{v_1 + v_2} \\
 f_2 &= \frac{v_2}{v_1 + v_2} \\
 S &= -k(n_1 \ln f_1 + n_2 \ln f_2)
 \end{aligned} \tag{4.45}$$

Now consider a solvent with small molecules and the solute has polymerized into large molecules that each occupy  $X$  squares. The number of solvent molecules and the squares that they occupy remain as  $n_1$  and  $v_1$ , but the number of solute molecules is reduced to  $n_2/X$ , but the squares that they occupy remain  $v_2$ . The entropy is given by

$$S = -k \left[ n_1 \ln f_1 + \frac{n_2}{X} \ln f_2 \right] \quad (4.46)$$

So that when 100 solute monomer molecules are dissolved in a lattice of 1000 squares, the entropy is given by:  $n_1 = 900$ ,  $n_2 = 100$ ;  $v_1 = 900$ ,  $v_2 = 100$ ;  $f_1 = 0.9$ ,  $f_2 = 0.1$ ;  $S/k = 325.1$ . But when the 100 solute monomer molecules are polymerized into 10 polymer molecules, and  $X = 10$ , the entropy is given by:  $n_1 = 900$ ,  $n_2 = 10$ ;  $v_1 = 900$ ,  $v_2 = 100$ ;  $f_1 = 0.9$ ,  $f_2 = 0.1$ ;  $S/k = 117.8$ .

*Stability and equilibrium* The structural factors that lead to low and high entropies in a molecule and in a system are summarized in table 4.24.

Entropy is central to stability and equilibrium. In a closed adiabatic system, any spontaneous change should lead to an increase in entropy, but the change may also lead to a change in the temperature. For an isothermal system, we require that, in a spontaneous change, the free energy  $G = H - TS$  be minimized. This means that for a spontaneous isothermal change

$$\Delta G = \Delta H - T \Delta S < 0 \quad \text{or} \quad \Delta H < T \Delta S$$

The most favorable change is the one that increases entropy while decreasing enthalpy. Let us now consider the changes shown in table 4.25.

Figure 4.25 shows a quadrant divided into four regions:

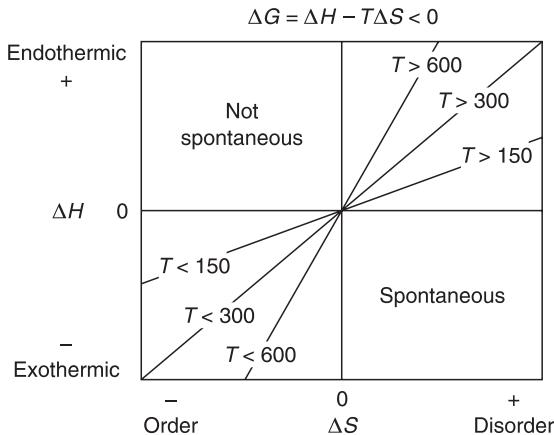
- Case (i). The lower right quadrant, where  $\Delta H < 0$  and  $\Delta S > 0$ , is always favorable at any temperature. An example is the dissolution of gaseous  $\text{H}_2\text{SO}_4$  into water, which is exothermic and the mixture has more entropy.

**Table 4.24 Summary of structural factors leading to low- and high-entropy molecules**

	<i>Low entropy</i>	<i>High entropy</i>
Organization	Order	Disorder
Molecular	Simple Atoms chained together Low mass Symmetric molecule Rigid structure High energy gap	Complex Individual separated atoms High mass Asymmetric molecule Flexible structure Low energy gap
External conditions	Low temperature Low volume High pressure	High temperature High volume Low pressure
Aggregation phase	Pure compound Solid	Mixture, solution Liquid, vapor

**Table 4.25 Entropy change versus enthalpy change effects**

$\Delta S < 0$	<i>Driven by <math>\Delta S &gt; 0</math></i>	
$\Delta H > 0$	Case (iv). Always unfavorable	Case (ii). Only when $T > \Delta H/\Delta S$
Driven by $\Delta H < 0$	Case (iii). Only when $T < \Delta H/\Delta S$	Case (i). Always favorable

**Figure 4.25** Spontaneous events, and their  $\Delta G$  and  $\Delta S$ 

- Case (ii). The upper right quadrant, where  $\Delta H > 0$ ,  $\Delta S > 0$ , is favorable only when the temperature is high enough, so that  $T > \Delta H/\Delta S$ . Examples are ice melting to water, water evaporating to steam, polymers cracking to monomers, roasting  $\text{CaCO}_3$  to make  $\text{CO}_2$  and  $\text{CaO}$ .
- Case (iii). The lower left quadrant, where  $\Delta H < 0$ ,  $\Delta S < 0$ , is favorable only when the temperature is low enough, so that  $T < \Delta H/\Delta S$ . Examples are condensation of steam into water, freezing of water into ice, monomers polymerizing into polymers,  $\text{CO}_2$  absorbs into  $\text{CaO}$ .
- Case (iv). The upper left quadrant, where  $\Delta H > 0$ ,  $\Delta S < 0$ , is always unfavorable at any temperature. An example is a solution of  $\text{H}_2\text{SO}_4$  spontaneously separating into acid gas and liquid water.

### *Predictable Properties*

There is no general-purpose or user-friendly software available for statistical mechanical computations, in comparison with the availability of useful and convenient tools to do single molecule quantum mechanical computations. There are many properties that can be computed, but users usually have to write their own computer programs to do each computation, so that it is a small research project instead of a convenient tool for product engineers. Westmoreland and Panagiotopoulos (2004) said that the availability

of force fields is basic to molecular simulation, or computational statistical mechanics. At this moment, most force fields have been formulated for isolated molecules and for liquids near room temperature. Progress in this field has been rapid, and it is expected that user-friendly software will be developed in the next 5 years to speed up the work of development and design engineers.

We list here some of the most successful quantitative predictions:

- ideal gas—heats of formation, entropy, heat capacity
- real gas—equation of state, critical properties
- liquid–vapor phase equilibrium, vapor pressure, Henry’s law constant
- normal boiling point
- densities
- surface tension
- viscosity, diffusion coefficient.

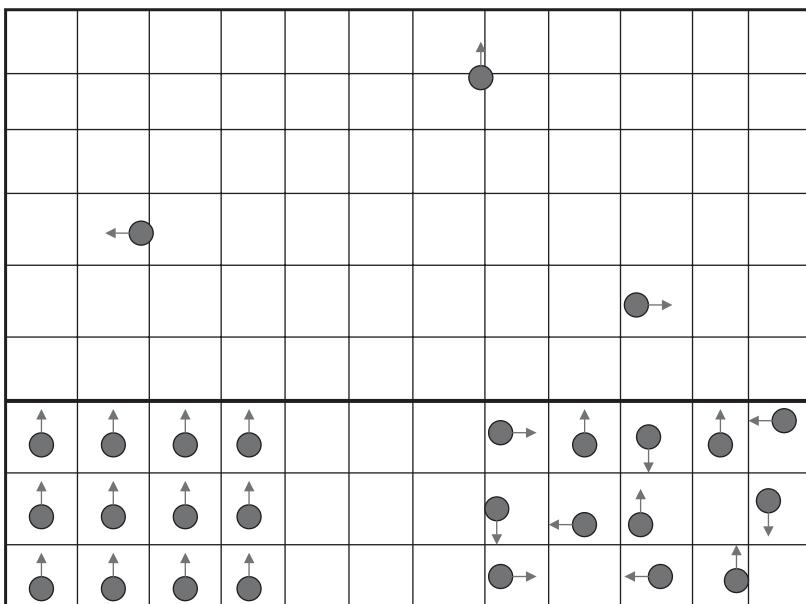
For the less ambitious goal of predicting qualitative trends upon molecular structure modifications, there are many more success stories. Some of the properties that have been used in industry to predict trends are:

- solution-phase thermochemistry
- kinetics and activity coefficients
- mixture properties
- entropy.

### 4.3 Gases, Solids, and Liquids

The macroscopic properties of the three states of matter can be modeled as ensembles of molecules, and their interactions are described by intermolecular potentials or force fields. These theories lead to the understanding of properties such as the thermodynamic and transport properties, vapor pressure, and critical constants. The ideal gas is characterized by a group of molecules that are “hard spheres” far apart, and they exert forces on each other only during brief periods of collisions. The real gases experience intermolecular forces, such as the van der Waals forces, so that molecules exert forces on each other even when they are not in collision. The liquids and solids are characterized by molecules that are constantly in contact and exerting forces on each other.

Figure 4.26 shows a cell model of the three phases. Gas in the upper region has a very low density and the molecules are free to fly around. When the vapor condenses into a liquid (shown lower right), the density is greatly increased so that there is very little free volume space; the molecules have limited ability to move around, and they have random orientation; that is, they can rotate and point in random directions. When the liquid freezes into a solid (shown lower left), the density is slightly increased to eliminate the void space, the molecules have assigned positions and are not free to move around, and there is now an orientation order; that is, they cannot rotate freely and they all point at the same direction.



**Figure 4.26** Cell model of gases, liquids, and solids

### Ideal Gases

Air at room temperature and pressure consists of 99.9% void and 0.1% molecules of nitrogen and oxygen. In such a dilute gas, each individual molecule is free to travel at great speed without interference, except during brief moments when it undertakes a collision with another molecule or with the container walls. The intermolecular attractive and repulsive forces are assumed in the “hard sphere” model to be zero when two molecules are not in contact, but they rise to infinite repulsion upon contact. This model is applicable when the gas density is low, encountered at low pressure and high temperature. This model predicts that, even at very low temperature and high pressure, the ideal gas does not condense into a liquid and eventually a solid.

The ideal gas law states that

$$PV = RT \quad (4.47)$$

where  $V$  is the volume per mole of gas. This law was derived empirically from Boyle’s law, which states that, for a fixed temperature, the product of  $P$  and  $V$  is a constant, and from Charles’ law, which states that, at a fixed pressure,  $V$  is proportional to the absolute temperature  $T$ . This law is generally valid when the temperature is much above the critical temperature and when the pressure is much below the critical pressure. A mole of air at 0 °C and 1 atm in pressure occupies 22.4 L. If an average nitrogen or oxygen molecule is a sphere of 3.6 Å in diameter, then an Avogadro’s number of such molecules would occupy only 0.015 L, which is about 0.06% of the total space.

A cubic container with a volume of 22.4 L would have edges of 28 cm, or 0.28 m, in length. If the molecules are evenly spaced in such a three-dimensional array, then the average distances between molecules would be 33 Å, which is nine times the molecular diameter.

The molecules of a dilute gas are sufficiently far from each other so that each would execute free flights with a very high velocity during periods between collisions. The mean velocity of a gaseous molecule is given by

$$\bar{u} = \sqrt{\frac{8kT}{\pi m}}$$

where  $m$  is the molecular weight of the molecule,  $T$  is the absolute temperature, and  $k$  is the Boltzmann constant, which is equal to the gas constant  $R$  divided by Avogadro's number.

For oxygen at room temperature,  $\bar{u} = 460$  m/s. The root-mean-square velocity is given by  $u_{\text{rms}} = \sqrt{3kT/m} = 1.085\bar{u}$ . The average number of molecules striking the container wall per square meter of surface area per second is given by  $Z = n\bar{u}/4$ , where  $n$  is the number density of molecules per cubic meter. The pressure generated on the wall is equal to the rate of collision multiplied by the change of moment of the gas molecule from  $+mu$  to  $-mu$ , so that  $P = 2muZ = (1/2)nmu^2$ . Thus, for air at 0 °C and 1 atm in pressure, the number density is  $n = 0.268 \times 10^{23}$  molecules/liter, and the number of collisions per square meter of surface is  $1.2 \times 10^{22}$  collisions per second.

The velocity distribution of the molecules was derived by Maxwell:

$$f(u) = 4\pi u^2 \left( \frac{m}{2\pi kT} \right)^{3/2} \exp \left( -\frac{mu^2}{2kT} \right) \quad (4.48)$$

This is shown in figure 4.27 for three different values of the parameter  $T/m$ . For a fixed value of  $m$ , higher temperature will shift the distribution to higher velocities; for a fixed  $T$ , higher mass will shift the distribution to lower velocities.

The ideal gas law is consistent with the assumption that molecules are point masses with no volume. However, the hard sphere model assumes that molecules are spheres with a finite diameters  $\sigma$  and introduces a central concept of the “mean free path”  $\lambda$ , which is the mean distance traveled by a molecule before collision with another molecule. It is given by

$$\lambda = \frac{1}{\sqrt{2}\pi\sigma^2 n} \quad (4.49)$$

where  $\sigma$  is the effective diameter of the molecule and  $n$  is the number of molecules per volume. The interval of time between collisions  $\tau$  is given by  $\lambda/u$ , and the binary collision rate of a given molecule with other molecules is given by  $F = u/\lambda$ , with a typical value for room air being about  $4 \times 10^9$ /s. The total number of binary collisions in a mole of gas at room temperature and pressure is then given by Avogadro's number multiplied by  $u/\lambda$ , and divided by 2 to avoid double counting, which is  $12 \times 10^{32}$ /s.

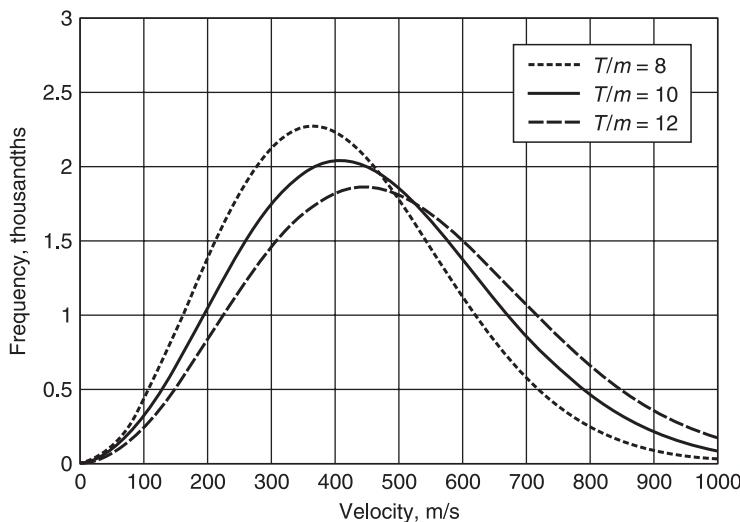


Figure 4.27 Maxwell velocity distribution of ideal gases

Table 4.26 Values of average velocity  $u$ , molecular diameter  $\sigma$ , mean free path  $\lambda$ , and average time between collisions  $\tau$  for some gas molecules

Gas	$m$	$u, \text{m/s}$	$\sigma, \text{nm}$	$\lambda, \text{nm}$	$\tau, \text{ps}$
H <sub>2</sub>	2	1914	0.292	98.8	52
He	4	1354	0.258	126.5	93
CH <sub>4</sub>	16	676	0.382	57.8	85
N <sub>2</sub>	28	511	0.368	62.9	121
O <sub>2</sub>	32	478	0.343	71.3	149
Cl <sub>2</sub>	71	321	0.412	49.6	154
C <sub>6</sub> H <sub>6</sub>	78	306	0.527	30.8	99

The values of some of these parameters at room temperature and pressure are given in table 4.26. These values are obtained from measurements of viscosity, thermal conductivity, diffusion, and from deviations from the perfect gas law.

### Real Gases

The perfect gas law applies when the gaseous molecules are far apart and have very little interaction with each other. At 0 °C and 1 atm, the air molecules have an average diameter of 3.6 Å and are at an average distance of 33 Å from their nearest neighbor. When the density is increased, the molecules draw closer to each other and the ideal

gas gradually breaks down. Let us introduce the compressibility parameter  $Z$ :

$$PV = ZRT \quad (4.50)$$

This parameter is equal to one when the perfect gas law applies. When the pressure is increased to 10 atm, the predicted perfect gas volume is reduced to 2.24 L and the average distance between the molecules shrinks to 15.5 Å; when the pressure is increased to 100 atm, the predicted volume is reduced to 0.224 L and the average distance shrinks further to 7.2 Å; when the pressure is further increased to 1000 atm, the volume predicted by the perfect gas law is 0.0224 L and the average distance shrinks further to 3.3 Å. By then, the intermolecular distance is roughly equal to the molecular diameters. As the molecules come together, the attractive forces are first felt, which manifests itself by a drop from the ideal gas volume, which is represented by  $Z < 1$ . With further compression, the molecules begin to touch, and the repulsive forces become dominant, which manifests itself as a resistance to further volume reduction and  $Z > 1$ .

Johannes van der Waals developed his famous equation of state by the introduction of both the attractive and the repulsive forces between the molecules. First he postulated that the gas behaves as if there is an additional internal pressure to augment the external applied pressure, which is based on the mutual attraction of molecules; since the density of molecules is proportional to  $1/V$ , the intensity of the binary attractive force would be proportional to  $1/V^2$ . Then he postulated that when the measured total volume begins to approach the volume occupied by the real gaseous molecules, the free volume is obtained by subtracting the molecular volume from the measured volume. Then he introduced the parameter  $a$ , which represents an attractive force responsible for the internal pressure, and the parameter  $b$ , which represents the volume taken by the molecules. He arrived at

$$(P + a/V^2)(V - b) = RT \quad (4.51)$$

The van der Waals constants for many molecules are tabulated in units of  $(\text{L}^2 \text{ atm})/\text{mol}^2$  for  $a$ , which is associated with intermolecular attraction; the unit for  $b$  is  $\text{L/mol}$ , and is associated with molecular volume. Some of the values of  $a$ ,  $b$ ,  $T_c$ ,  $P_c$ ,  $V_c$  are given in table 4.27.

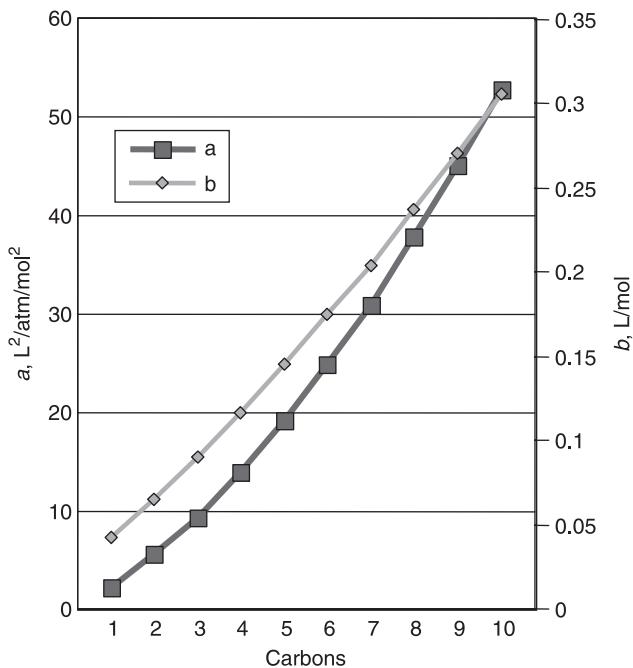
The values of  $a$  and  $b$  are shown in figure 4.28 for the homologous series of the normal paraffins from methane to decane. There is a smooth increase of both  $a$  and  $b$  with the number of carbon atoms, which is slightly accelerating from a linear relation, showing that the attractive forces grow much faster than the molecular weight, and in fact the best correlation is with the number of carbon atoms raised to the power 1.37. The critical parameters  $T_c$ ,  $P_c$ , and  $V_c$  of the normal paraffins are shown in figure 4.29. The critical temperatures show a slight saturation effect at high atomic weights, and the critical pressure of methane shows an anomaly.

The van der Waals equation can be reformulated thus:

$$Z = \frac{PV}{RT} = \frac{V}{V - b} - \frac{a}{VRT} \quad (4.52)$$

**Table 4.27 van der Waals' and critical parameters of gas molecules**

Molecule	$a, (L^2 \text{ atm})/mol^2$	$b, L/mol$	$T_c, K$	$P_c, \text{ atm}$	$V_c, cm^3/mol$
He	0.0341	0.0236	5.2	2.28	57.3
Ne	0.2102	0.01709	44.4	26.53	41.7
Ar	1.355	0.03201	150.86	48.98	74.6
Kr	2.325	0.0396	209.35	55.02	91.2
Xe	4.194	0.0516	278.74	58.4	118
Rn	6.660	0.0624	377.4	63.0	140
$H_2$	0.2484	0.0265	33.18	13.13	64.2
$O_2$	1.3820	0.0319	154.58	50.43	73.4
$N_2$	1.3900	0.0391	126.1	33.94	90.1
$CO_2$	3.592	0.04267	304.1	73.8	93.9
$NH_3$	4.170	0.03707	405.5	113.5	72.5
NO	1.340	0.02789	180	648	57.7
$H_2O$	5.464	0.03049	647.3	221.2	57.1
$CH_4$	2.300	0.0430	190.58	46.04	99.3
$C_2H_6$	5.57	0.0650	305.42	48.8	147.9
$C_3H_8$	9.385	0.0904	369.8	42.5	202.9
$C_5H_{12}$	19.13	0.1449	469.65	33.69	312.3
$C_{10}H_{22}$	52.88	0.3051	618.45	21.23	603.1



**Figure 4.28** van der Waals parameters of normal paraffins

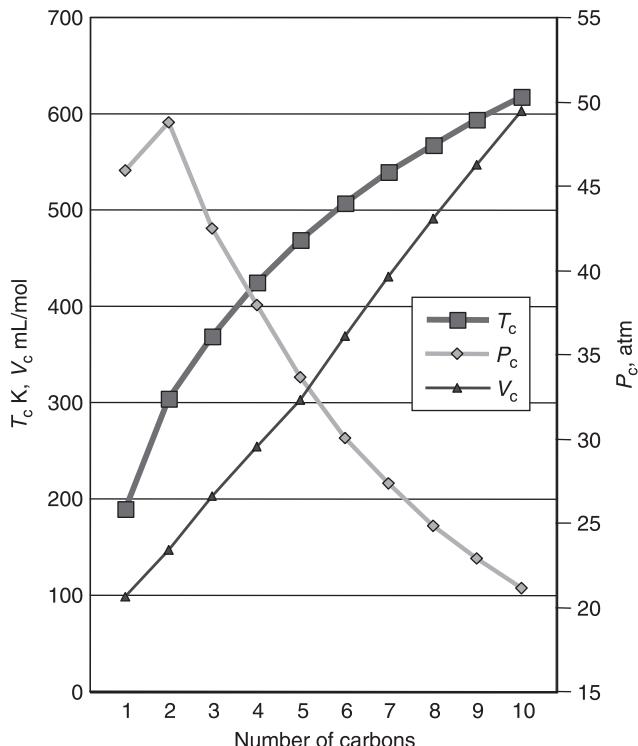


Figure 4.29 Critical parameters of normal paraffins

Let us examine the value of  $Z$  under different conditions. The first term is always greater than one, which represents the repulsion term making the volume greater than the ideal gas volume; and the second term reduces the value of  $Z$ , which represents the attraction term. At a fixed value of  $T$  above the critical temperature, compression will cause  $V$  to decrease so that  $Z$  will drop below one, and further compression will cause  $V$  to decrease even more so that  $Z$  will rise above one. When the temperature is at or below the critical temperature, compression will eventually cause the gas to condense into a liquid at or above the critical pressure  $P_c$ . The relations between the critical constants and the values of van der Waals'  $a$  and  $b$  are

$$V_c = 3b$$

$$P_c = \frac{a}{27b^2}$$

$$T_c = \frac{8a}{27Rb} \quad (4.53)$$

$$Z_c = \frac{P_c V_c}{RT_c} = \frac{3}{8}$$

According to the van der Waals equation of state, the value of compressibility at the critical point should be  $3/8 = 0.375$ . When does a real gas depart significantly from an ideal gas? We can write equation (4.9) as the reduced equation of state, with the reduced temperatures, pressures, and volumes:  $T_r = T/T_c$ ,  $P_r = P/P_c$ ,  $V_r = V/V_c$ . Then, all gases would have the same equation of state in the form of reduced parameters:

$$\left( P_r + \frac{3}{V_r^2} \right) \left( V_r - \frac{1}{3} \right) = \frac{8}{3} T_r \quad (4.54)$$

The compressibility  $Z$  is now given by

$$P_r V_r = Z \frac{8}{3} T_r$$

The assumption known as the law of corresponding states asserts that the compressibility factor  $Z$  should be a function only of the reduced temperature  $T_r$  and the reduced pressure  $P_r$ , which is approximately correct for many real gases. It is seen that, for a van der Waal gas, the minimum value of  $V_r = 1/3$ , which can be achieved only at infinite pressure. From the equation of state written for the reduced temperature and pressure, we can derive the equivalent formula of compressibility as

$$Z = \frac{V_r}{V_r - 1/3} - \frac{9/8}{V_r T_r}$$

This function breaks down when  $V_r$  approaches  $1/3$  as the value of  $Z$  goes to infinity, and when  $T_r \leq 0.85$  as the value of  $Z$  becomes negative around  $V_r \approx 0.65$ . When  $T \geq 27/8$ , the value of  $Z \geq 1$ . It is easy to calculate  $Z$  as a function of  $V_r$  and  $T_r$ . However, it is usually more convenient to have  $Z$  as a function of the reduced pressure and the reduced temperature, such as the graph in figure 4.30. The region to the upper left is the dilute gas region where  $Z \approx 1$ . The region to the lower right is the dense gas region where  $Z > 2$ , and where the gas volume is much greater than that predicted by ideal gas. The lower left region enclosed by the parametric curve where  $Z = 1$  has the value of  $Z < 1$ . It is also bounded by  $T_r \leq 27/8$ , or 3.375, from above, by  $P_r \leq 20/3$ , or 6.667, from the right. This is a region dominated by attractive forces, so that the volume is smaller than the ideal gas volume; and there may be more than one solution  $V_r$  to a given value of  $T_r$  and  $P_r$ . For instance, when  $T_r = 0.9$  and  $P_r = 0.6$ , there are two solutions:  $V_r = 1.0$  and  $V_r = 2.7$ .

The van der Waals equation is not a particularly accurate tool for prediction of compressibility  $Z$ , but it is the first theory to illuminate the nature of the attractive and repulsive forces that lead to departure from the perfect gas law. There are many more accurate equations of state that use more parameters, including the Benedict–Webb–Rubin equation, the Redlich–Kwong equation, and the Peng–Robinson equation. The compressibility factor can also be expanded into the virial form

$$Z = \frac{PV}{RT} = 1 + \frac{B(T)}{V} + \frac{C(T)}{V^2} + \frac{D(T)}{V^3} + \dots \quad (4.55)$$

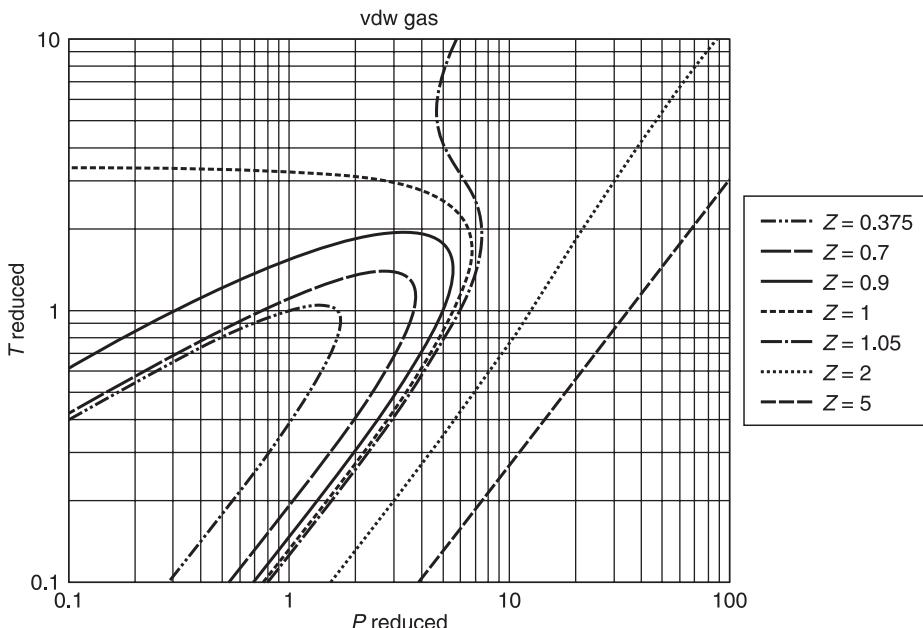


Figure 4.30 Compressibility  $Z$  of van der Waals gases, versus  $T_r$  and  $P_r$

When we make a plot of  $Z$  versus  $1/V$ , we expect that the value of  $Z = 1$  when  $1/V = 0$  or at infinite volume. The first virial coefficient  $B(T)$  is the slope of the line at  $1/V = 0$ .  $B(T)$  is usually negative at very low temperatures due to the attractive forces. But when the temperature increases,  $B(T)$  will turn positive, and the temperature at which  $B$  becomes zero is called the “Boyle temperature,” since that is the temperature where Boyle’s law applies exactly (at infinite volume).

### Solids

Solids and liquids are condensed matter that are held together by cohesive forces among the molecules. A measure of the strength of the cohesive force is the melting point. Table 4.28 and figure 4.31 show the melting points of the elements, which bears a general resemblance to the density, but with major exceptions. Osmium and iridium have the highest densities among metals at more than 22 g/mL, but they do not have the highest melting points; these belong to tungsten and rhenium, at more 3400 °C. On the other hand, the low-density nonmetallic elements of carbon and boron have exceptionally high melting points.

The density of a solid also depends on how they are packed together. The coordination number is the number of nearest neighbors to an atom. The most densely packed configurations are the face-centered cubic and the hexagonal close-packed with a coordination number of 12; and the least densely packed is the tetrahedral configuration with a coordination number of 4. The most important configurations are shown in figure 4.32, and their packing volume fractions are table 4.29.

**Table 4.28 Melting points (°C) of elements**

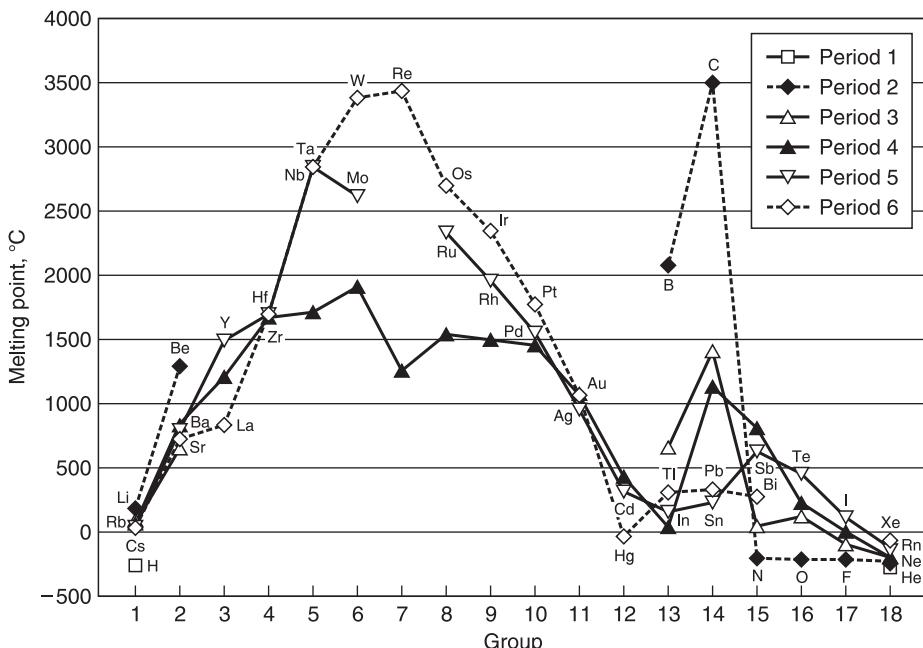


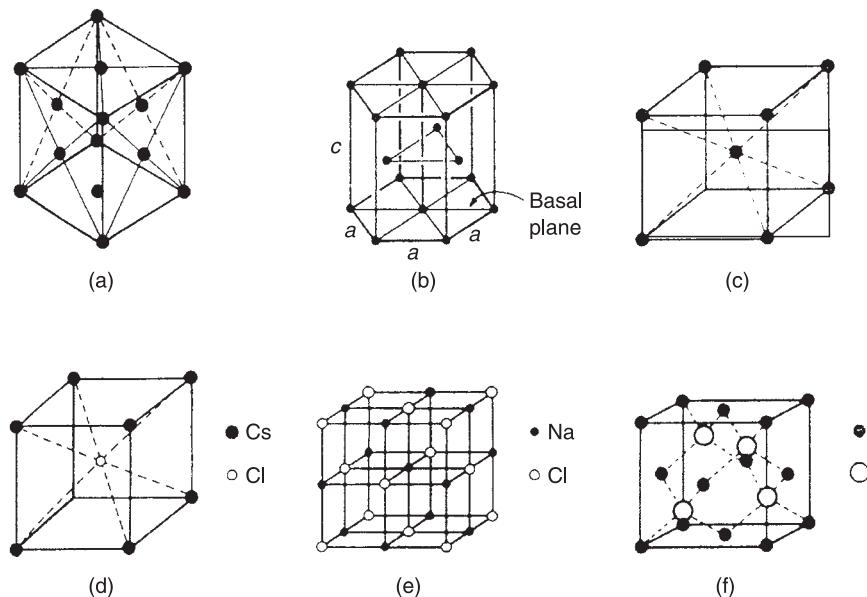
Figure 4.31 Melting points of elements

Condensed matter can be classified by the nature of the forces that hold it together: ionic solids; covalent solids; metallic solids; molecular solids.

### *Ionic Solids*

An ionic solid is a giant molecule held together by the coulombic electrostatic attraction forces between the positive and negative ions. The charge intensities between the ions depend on the differences in the electronegativity of the atoms, so any ionic bond has a partial covalent character. Examples are NaCl, MgO, KF, CaO, CaF<sub>2</sub>, TiO<sub>2</sub>. Since an A—B bond in a salt has both ionic and covalent characters, there is a gradation between these two types of bond—from the strongest ionic bond of group 1–17, salts such as KF, to the very strong ionic bond of group 2–16 salts, such as MgO, to the mildly ionic bond of group 3–15 salts, such as GaAs, to the pure covalent bonds of the group 14–14 compounds, such as diamond (table 4.30).

Table 4.31 shows the melting and boiling points of some elements and compounds. The ionic bond energies are frequently strong, as ionic compounds tend to have high melting and boiling temperatures. The simple cubic lattice is the most stable form for Na<sup>+</sup>Cl<sup>-</sup>, where each Na<sup>+</sup> has six neighboring Cl<sup>-</sup> ions. Figure 4.32 shows the NaCl crystal with coordination number of 6, where each Na<sup>+</sup> ion is attached equally to six Cl<sup>-</sup> ions, and individual molecules cannot be distinguished. Since two positive charges must not be placed next to each other to create repulsion, this rules out the



**Figure 4.32** Crystal structures of (a) f.c.c. ( $Z = 12$ ), (b) h.c.p. ( $Z = 12$ ), (c,d) b.c.c. ( $Z = 8$ ), (e) simple cubic ( $Z = 6$ ), and (f) tetrahedral ( $Z = 4$ )

**Table 4.29** Packing volume fractions for various types of atomic packing

Type	Coordination number	Packing fraction	Example
Tetrahedral	4	0.34	Diamond
Simple cubic	6	0.52	NaCl
Body-centered cubic	8	0.68	CsCl, K, Fe
Face-centered cubic	12	0.74	Ni, Au, Ar
Hexagonal close-packed	12	0.74	Be, Cd, Mg

**Table 4.30** Electronegativity differences between salts of different groups

Character	Salt of groups	Example	Electronegativity differences
Ionic	1–17	KF	4.0 – 0.8 = 3.2
	2–16	MgO	3.5 – 1.2 = 2.3
	3–15	GaAs	2.0 – 1.6 = 0.4
Covalent	14–14	Diamond	2.5 – 2.5 = 0.0

possibility of a highly packed configuration with the very high coordination of 12. The ionic bonds are not flexible, so that salt is brittle and it has little resistance to fracture from tension or shear. The cations and the anions generally do not have the same diameter, so that the most stable crystal structure is often controlled by the ratios of the diameters of the two ions. CsCl has the body-centered cubic structure,

**Table 4.31** Phase transition temperatures and densities of ionic, covalent, metallic, and molecular liquids and solids

Chemical	M.p., °C	B.p., °C	Density		
			Liquid	Solid	T <sub>m</sub> /T <sub>b</sub>
<i>Ionic</i>					
NaCl	800.0	1413.0		2.16	0.64
KF	880.0	1500.0		2.48	0.65
SiO <sub>2</sub>	1710.0	2230.0		2.32	0.79
Al <sub>2</sub> O <sub>3</sub>	1999.0	2210.0		3.99	0.92
CaO	2570.0	2850.0		3.32	0.91
MgO	2800.0	3600.0		3.65	0.79
<i>Covalent</i>					
Ge	958.5	2700.0		5.36	0.41
Si	1412.0	2600.0		2.40	0.59
C diamond	3500.0	4200.0		3.51	0.84
<i>Metallic</i>					
Hg	-38.8	356.9	13.60		0.37
K	62.3	760.0		0.86	0.32
Al	660.0	1800.0		2.70	0.45
Ag	961.8	2164.0		10.52	0.51
Fe	1535.0	3000.0		7.90	0.55
Os	2700.0	5300.0		22.48	0.53
W	3370.0	5900.0		19.30	0.59
<i>Molecular</i>					
He	-272.2	-268.9			0.21
Rn	-71.0	-61.8			0.96
H <sub>2</sub>	-259.1	-252.7			0.68
Cl <sub>2</sub>	-101.6	-34.6			0.72
S <sub>8</sub>	119.0	444.6		1.96	0.55
H <sub>2</sub> O	0.0	100.0	1.000	0.91	0.73
HF	-80.0	20.0			0.66
NH <sub>3</sub>	-80.0	-35.0			0.81
CH <sub>4</sub>	-182.4	-161.0			0.81
C <sub>3</sub> H <sub>8</sub>	-165.0	-91.0			0.59
C <sub>4</sub> H <sub>10</sub>	-138.3	-0.5	0.579		0.49
C <sub>6</sub> H <sub>6</sub>	5.5	80.1	0.879		0.79
C <sub>10</sub> H <sub>22</sub>	-29.7	174.0	0.730		0.54
C <sub>6</sub> Cl <sub>6</sub>	220.0	330.0		2.044	0.82

where each chlorine ion has eight cesium ions as neighbors, which is also shown in figure 4.32.

### Covalent Solids

A crystal of covalent solid can be considered as a single molecule held together as an extensive network of covalent bonds, with no individually distinguishable small molecules. Examples are graphite as multiple layers of two-dimensional networks,

diamond as a three-dimensional network, silicon and germanium, and various oxides, sulfides, and nitrides. The cohesive forces here are among the strongest in nature, and table 4.31 shows that the melting and boiling temperatures can be very high. The carbon atoms in graphite are held together by  $sp^2$  orbitals which are directed at an angle of  $120^\circ$ , so that the coordination number is three in each two-dimensional sheet. Different sheets are held together by weaker van der Waals forces. The carbon atoms in diamond are held together by  $sp^3$  orbitals, which are directed at an angle of  $109.5^\circ$ , so that the coordination number is 4. Figure 4.32 shows a tetragonal crystal of diamond with a coordination number of 4. The covalent solids have low coordination, leading to open structure and low density. The strong covalent bonds of these materials make them strong in resisting compression and shear, but these bonds have such strong preferred orientation that they cannot be bent, so that they can fail by fracture. Figure 4.31 shows that there can be a very large decrease in melting point with the period, which can be attributed to the decline in cohesive energies and the increasing metalloid character of the bonds at high atomic weights.

### *Metallic Solids*

The metallic crystal can be considered to be a single molecule with a network of positive ions held together by a sea of detached electrons. Examples are sodium, silver, and tungsten. Table 4.31 shows that they have the highest densities, and they often have very high melting and boiling points, with the exception of mercury, which is liquid at room temperature. This type of solid bonding is not very sensitive to direction, so that the coordination can be high. They also have sufficient flexibility in bond angles, which allows permanent deformation, and makes metals ductile and malleable. There is a highly developed theory of metals due to Drude and Sommerfeld, and many of the properties of the metal can be derived from this model. The electron gas has an average velocity that is much higher than gaseous velocities at the same temperature, due to the very small mass of the electron; it also has a mean free path, based on collision with the network of positive ions, that is of the order of magnitude of  $3 \text{ \AA}$ . Both electrical and thermal conductivities are carried out by this electron gas. Figure 4.31 shows that groups 5 and 6 have the highest melting points, which increase with the period; groups 1 and 12 have the lowest melting points, and have little trend with the period.

### *Molecular Solids*

The molecular solid is made of individually distinguishable molecules. Each molecule is internally glued together by the short and strong intramolecular covalent or ionic bonds; however, the ensemble of molecules are glued together by the long and weak intermolecular van der Waals forces. Examples are noble gases (He, Ne, Ar), halogens ( $\text{F}_2$ ,  $\text{Cl}_2$ ,  $\text{Br}_2$ ,  $\text{I}_2$ ), oxygen and nitrogen, sulfur as an  $\text{S}_8$  ring, phosphorous as a  $\text{P}_4$  tetrahedron, and the hydrocarbons and organic compounds. Table 4.31 shows that the molecular solids have the lowest melting points and densities, due to the weak van der Waals forces that are easily disrupted by thermal forces. The Lennard-Jones 6–12 potential is a central force that does not depend on angles or directions, so these solids

tend to have the largest number of near neighbors, that is, high coordination of 12. Figure 4.32 shows two crystals with the highest coordination number of 12: the face-centered cubic and the hexagonal close-packed. The densities of molecular solids tend to be low, both because of the light elements involved and because of the long van der Waals bonds. The manner in which the molecules “dock” to each other plays a critical role in the packing efficiencies, which controls properties, including the density and the boiling points.

One of the consequences of close packing in solids and liquids is much higher densities in comparison with gases; for instance, ice and water have densities that are a thousand times higher than water vapor at room pressure. Another consequence is that solids and liquids have much lower compressibility, so that the density is not sensitive to the pressure. The bulk modulus  $B$  is defined as  $B = -\Delta P/(\Delta V/V)$ , which has the units of pressure. This parameter measures the fractional volumetric response of a material, when pressure is applied to all faces of the material at the same time.

The molecules in a solid are locked in their lattice positions without the ability to move about or to rotate in position, but the liquid molecules have much greater mobility. This is partly due to the slight volume expansion when solids melt into liquids, which creates vacancies or “free volume holes” and promotes mobility. A solid resists tension as well as shear, so that the solid springs back to its original shape and volume when a temporary stress is removed; a liquid in the same circumstances will undertake lasting deformation. When a solid rod of length  $L$  and cross-sectional area  $A$  is pulled by a tension force  $F$ , it will be elongated and the Young’s modulus or elasticity modulus is defined as  $E = -(F/A)/(\Delta L/L)$ . When a material is subjected to a shearing force, the shear modulus or rigidity modulus  $G$  is defined as  $G = -F/A\theta$ , where  $\theta$  is the angle of deflection of a rectangular object. A liquid does not have a Young’s and shear modulus, so it conforms to the shape of the container, but a solid has these moduli and does not conform. However, liquids do resist shear, as the speed of response is related to the viscosity, so that it requires an expenditure of energy to move liquid molecules past each other.

## Liquids

Our theoretical understanding of the liquid state is not as good as the gaseous or the solid states. The best approaches today are to model the liquid state either as a disordered solid state or as a dense gas state. In the first approach, the liquid is treated as a disordered solid, with less organization and rigidity, and is slightly less dense; see figure 4.26. The nature of the cohesive forces in liquid bonding is also by molecular, covalent, ionic, and metallic forces; however, the orientation character of the covalent and ionic bonds is weakened by random thermal forces. The important geometric parameters of melting are the fractional volume expansion upon melting, or  $\Delta V/V$ , which tracks the amount of holes or “free volume” created in the liquid, as well as the expansion in cell size. This free volume leads to a less cohesive energy or an increase in enthalpy, which is unstable; this is balanced by an increase in the entropy and disorder of the system, which is stabilizing. The system is stable when the

**Table 4.32** Volume expansions on melting

	$T_m, K$	$\Delta V/V, \%$
Ar	83.8	14.3
N <sub>2</sub>	63.1	8.4
CH <sub>4</sub>	90.7	8.2
C <sub>6</sub> H <sub>6</sub>	278.8	13.3
C <sub>6</sub> H <sub>14</sub>	177.8	11.8
C <sub>8</sub> H <sub>18</sub>	216.3	14.8
KCl	1043	17.3
LiF	1121	29.4
Hg	234.3	3.6
Na	371	2.6

free energy is at a minimum,  $G = H - TS$ . Some values for volume expansion from melting are listed in table 4.32.

Some of the most important consequences of this free volume are a lower coordination number and a much higher mobility for the molecules. A typical coordination number in the liquid is 8 to 11, which is smaller than the 12 encountered in molecular and metallic solids. Another fundamental consequence in melting is the loss of rigidity in lattice position and rotational orientation. The packing is not regular and precise, so that the distances between the molecules are comparable to the solid state, but the bond angles are flexible and much less resistant to shearing motion. In a solid NaCl crystal, the distances of the six nearest Cl<sup>-</sup> ions to an arbitrary Na<sup>+</sup> ion are precisely one unit. The distances of the 12 nearest Na<sup>+</sup> ions are precisely  $\sqrt{2}$  units. In the molten state, the distances are no longer precise, but can be described by a statistical “radial distribution function,” which resembles the solid-state distribution in the short range, but converges to the average distribution in the long range. Shear forces would lead to permanent deformation, where the speed of deformation is related to the viscosity. The translational motion of a molecule is highly constrained by the closeness of neighboring molecules, which can be represented as vibrations in a force field. Diffusion takes place by the random motion of molecules into the holes, which may involve overcoming an activation energy barrier by thermal motion.

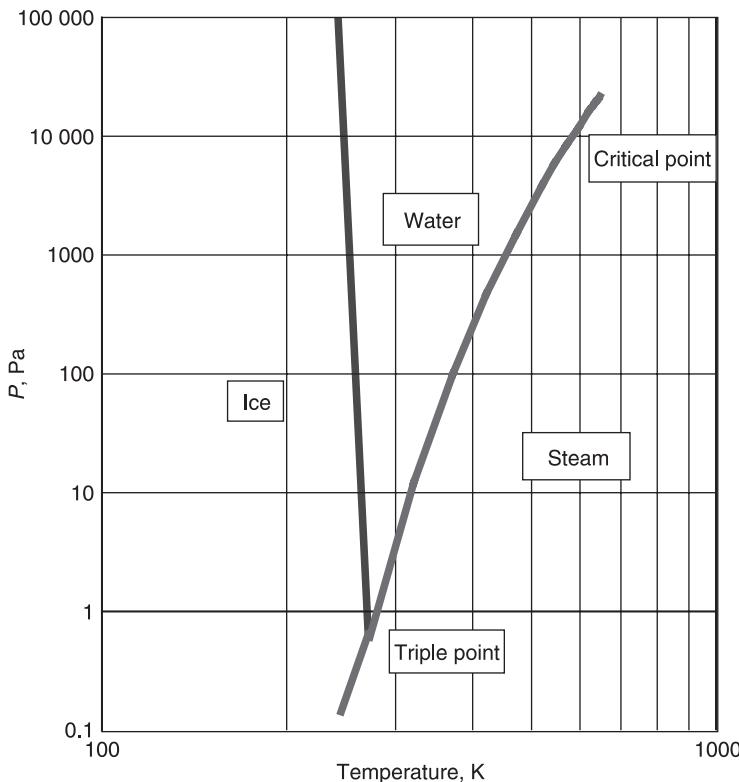
The other approach to liquid-state theory is to treat liquids as dense gases typical of high pressures, except that they have much higher densities and lower compressibility. In the van der Waals equation format, the value of  $(V - b)/V$  representing the free volume would be of the order of 10%. The mean free path of free flight between collisions becomes much shorter, and is comparable to the molecular diameters.

### Phase Transition and the Critical Region

The liquid state exists only below the critical point pressure and above the triple point pressure. When a vapor below the triple point pressure is cooled down, we encounter a discontinuous and abrupt phase change to solid; but, above the critical point pressure, a cooled vapor turns into the supercritical state where the properties of the fluid

Table 4.33 Phase-change characteristics for water

	$T, K$	$P, kPa$	$V_g, m^3/kg$	$V_L, m^3/kg$
Critical	647	22,089	0.003	0.003
Boiling, 1 atm	373	101	1.67	0.001
Freezing	273	0.6	206	0.001
Triple	273	0.6	206	0.001

Figure 4.33 Phase diagram of water,  $P$  versus  $T$ 

change continuously without phase change. Take the case of water, whose phase change characteristics are listed in table 4.33.

Let us review three different representations of the phase diagram of water. Figure 4.33 shows the  $P$  versus  $T$  diagram with the critical point on the upper right-hand corner, at a critical pressure of 22,089 kPa and critical temperature of 647 K; the supercritical region is above this point, where there is no distinction between liquid and gas. The triple point is at the bottom, where all three phases of steam–water–ice can coexist. The solid region is to the left, the liquid region is in the middle, and the vapor region is to the right. When we turn to Figure 4.34(a), we see a plot of  $P$

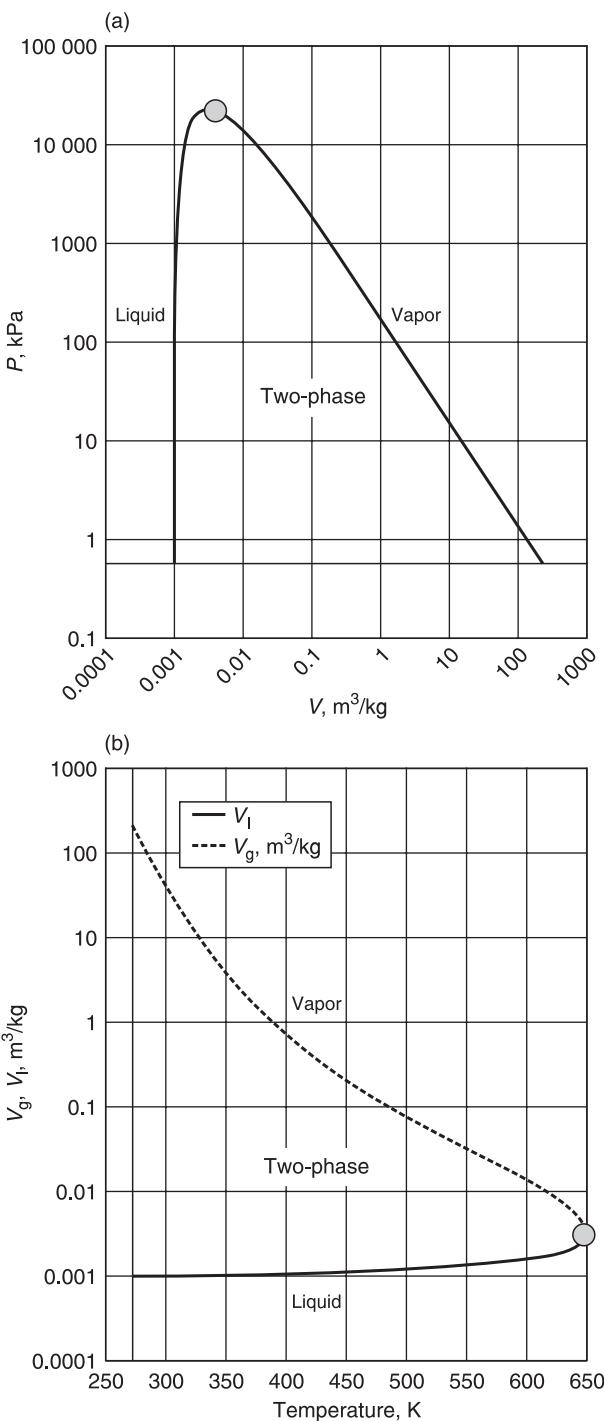


Figure 4.34 Phase diagram of water: (a)  $P$  versus  $V$ ; (b)  $V$  versus  $T$

**Table 4.34 Critical states and triple points for some gases**

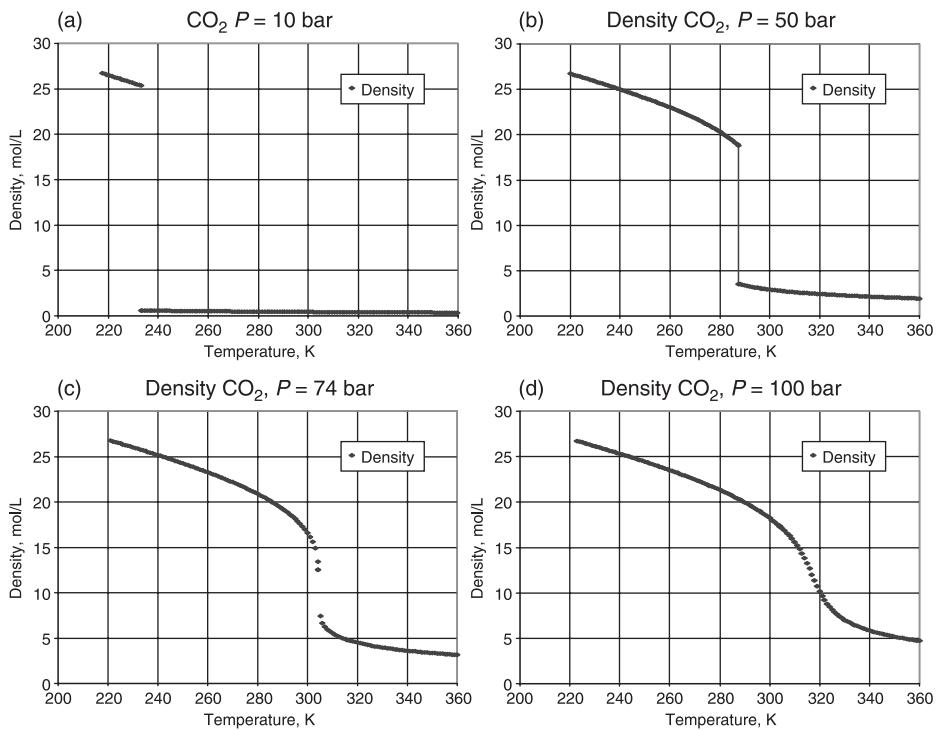
	$T_c, K$	$P_c, atm$	$T_{tr}, K$	$P_{tr}, atm$
He	5.19	2.27	2.17	0.051
$H_2$	33	13	14	0.072
$CO_2$	304	74	216	5.1
$H_2O$	647	221	273	0.006

versus  $V$ . We find the critical point at the top as the maximum point of the vapor-liquid equilibrium curve, a two-phase region at pressures below the critical pressure, and the triple point becomes the horizontal line at the bottom. Figure 4.34(b) gives a plot of  $V$  versus  $T$ , with the critical point on the lower right-hand corner of the vapor-liquid equilibrium curve. We have ice to the left of the diagram when  $T < 273$  K.

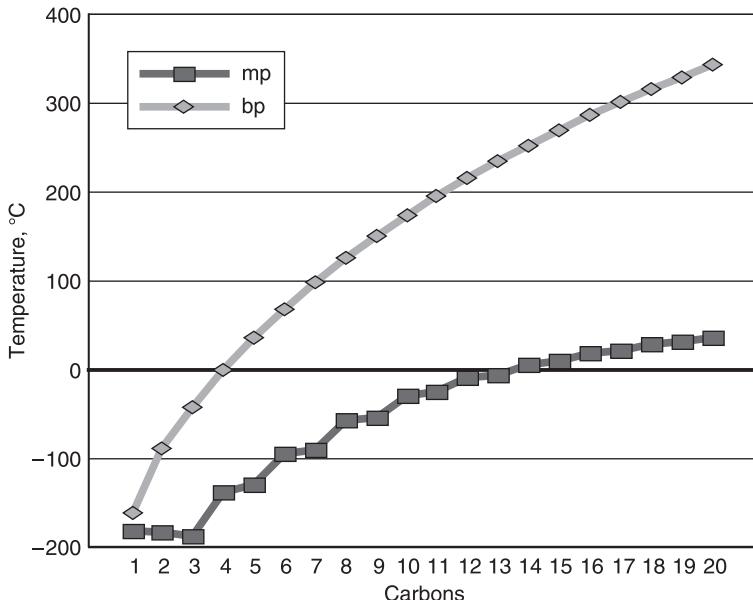
When we look at the critical states and triple points of other gases, we find the situation shown in table 4.34. The liquid phase exists only when the pressure is between the critical and the triple-point pressures. If we cool down hydrogen, helium or water at room temperature and pressure, we will get liquids before we get solids. But if we cool down  $CO_2$  from room temperature and pressure, we get dry ice rather than liquid carbonic; to obtain liquid carbon dioxide we have to raise the pressure to at least 5.1 atm to exceed the triple-point pressure. The melting point is not as sensitive to the pressure as the boiling point, which is stated usually for a room pressure of 1 atm, which prevails at sea level on Earth and not in Colorado or the Himalayas.

The phase change between liquid water and vapor below the critical pressure is called a first-order phase transition, which produces an abrupt jump change in many properties such as density, heat capacity, and enthalpy. Figure 4.35 shows that when carbon dioxide at 10 bar is cooled down from 360 K to 220 K, the density undergoes an abrupt change at 230 K when the vapor condenses into a liquid. When carbon dioxide is at 50 bar, the abrupt phase change takes place at 284 K and the change in density is just as abrupt but the gap becomes smaller. At the critical pressure of 74 bar, there is a continuous change of the density, but the slope is at a maximum at the critical temperature of 304 K, as the difference between vapor and liquid is becoming blurred. At the supercritical pressure of 100 bar, the density change is continuous and smooth, but the slope is highest near 330 K. When a liquid approaches the critical state, its properties pass into those of the gas imperceptibly.

The phase transition from solid to liquid is a more complex phenomenon than the phase transition from liquid to gas. Figure 4.36 and table 4.35 show the series of normal paraffins; these make a smooth progression in boiling point with the number of carbon atoms; however, the melting points do not follow such a simple pattern, as the normal order of higher melting points for more carbon atoms is reversed from methane to propane. There is also a curious even-odd pattern of the melting points, where the gap between an even and the subsequent odd number of carbon atoms is much smaller than the gap between an odd with the subsequent even number of carbon atoms. The simplicity of the boiling points is in large contrast to the complexity of the melting points, which is related to the symmetry of the molecules. The free energy of melting



**Figure 4.35** Phase diagram of carbon dioxide, density versus  $T$  at (a)  $P = 10 \text{ bar}$ , (b)  $P = 50 \text{ bar}$ , (c)  $P = 74 \text{ bar}$ , and (d)  $P = 100 \text{ bar}$



**Figure 4.36** Phase transition temperatures of normal paraffins

Table 4.35 Phase transition temperatures of normal paraffins

		<i>MW</i>	<i>M.p., °C</i>	<i>B.p., °C</i>
Methane	CH <sub>4</sub>	16	-182.4	-161.7
Ethane	C <sub>2</sub> H <sub>6</sub>	30	-182.7	-88.6
Propane	C <sub>3</sub> H <sub>8</sub>	44	-187.6	-42.2
Butane	C <sub>4</sub> H <sub>10</sub>	58	-138.2	-0.5
Pentane	C <sub>5</sub> H <sub>12</sub>	72	-129.7	36.1
Hexane	C <sub>6</sub> H <sub>14</sub>	86	-94	68.7
Heptane	C <sub>7</sub> H <sub>16</sub>	100	-90.5	98.4
Octane	C <sub>8</sub> H <sub>18</sub>	114	-56.8	125.6
Nonane	C <sub>9</sub> H <sub>20</sub>	128	-53.7	150.7
Decane	C <sub>10</sub> H <sub>22</sub>	142	-29.7	174
Undecane	C <sub>11</sub> H <sub>24</sub>	156	-25.5	196
Dodecane	C <sub>12</sub> H <sub>26</sub>	170	-9.5	216.4
Tridecane	C <sub>13</sub> H <sub>28</sub>	184	-5.4	235.5
Tetradecane	C <sub>14</sub> H <sub>30</sub>	198	5.5	252.5
Pentadecane	C <sub>15</sub> H <sub>32</sub>	212	10	270.7
Hexadecane	C <sub>16</sub> H <sub>34</sub>	226	18.2	286.9
Heptadecane	C <sub>17</sub> H <sub>36</sub>	240	22	302.2
Octadecane	C <sub>18</sub> H <sub>38</sub>	254	28.2	316.8
Nonadecane	C <sub>19</sub> H <sub>40</sub>	268	32.2	330
Eicosane	C <sub>20</sub> H <sub>42</sub>	282	36.5	343.8

at the melting point is zero, so that the melting temperature is a ratio of the enthalpy of melting to the entropy of melting:

$$\Delta F_m = F_L - F_s = \Delta H_m - T_m \Delta S_m = 0$$

$$T_m = \frac{\Delta H_m}{\Delta S_m} = \frac{H_L - H_s}{S_L - S_s} \quad (4.56)$$

Consider a molecule with a smaller symmetry number of  $\sigma_1$  and an isomer molecule with a larger symmetry number of  $\sigma_2$ ; the liquid entropy of the less symmetrical molecule is higher than the more symmetrical molecule by  $R \ln(\sigma_2/\sigma_1)$ . As a consequence, the more symmetric molecule has a smaller entropy of melting,  $\Delta S_{m,2} = \Delta S_{m,1} - R \ln(\sigma_2/\sigma_1)$ , which would lead to a higher melting temperature—if symmetry does not have a big effect on the enthalpy of melting or the entropy of the solid.

There are three forms of pentane: the straight-chain normal pentane ( $\sigma = 2$ ), the branched isopentane ( $\sigma = 1$ ), and the doubly branched neopentane ( $\sigma = 12$ ) in the shape of a tetrahedron. Figure 4.37 shows that normal pentane has the highest density and boiling point, which can be attributed to the superior packing efficiency of linear molecules; isopentane, in comparison, has lower density and lower melting and boiling points. Neopentane does not follow the usual rules; it has the lowest

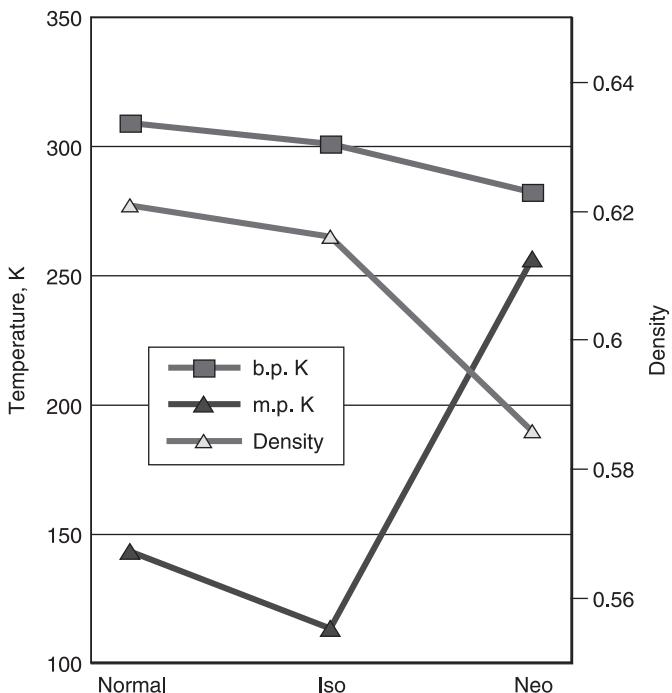


Figure 4.37 Phase transition temperatures of pentanes

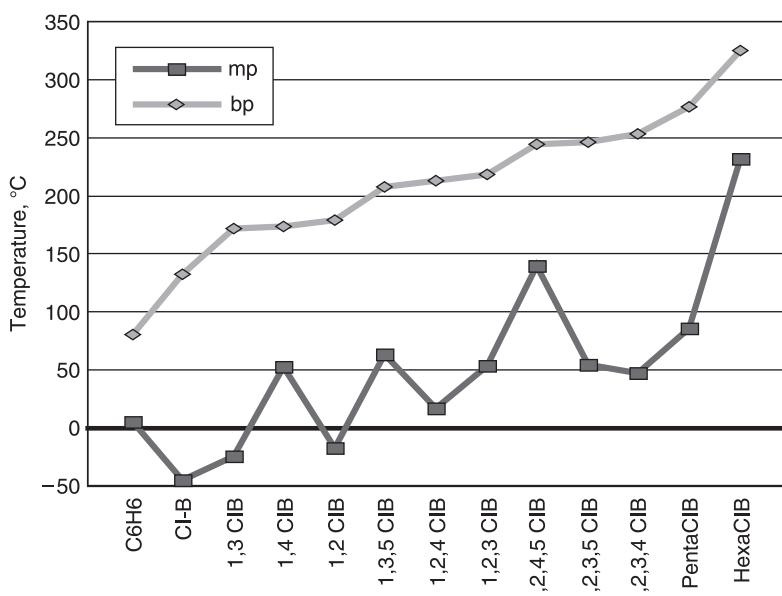


Figure 4.38 Phase transition temperatures of polychlorinated benzenes

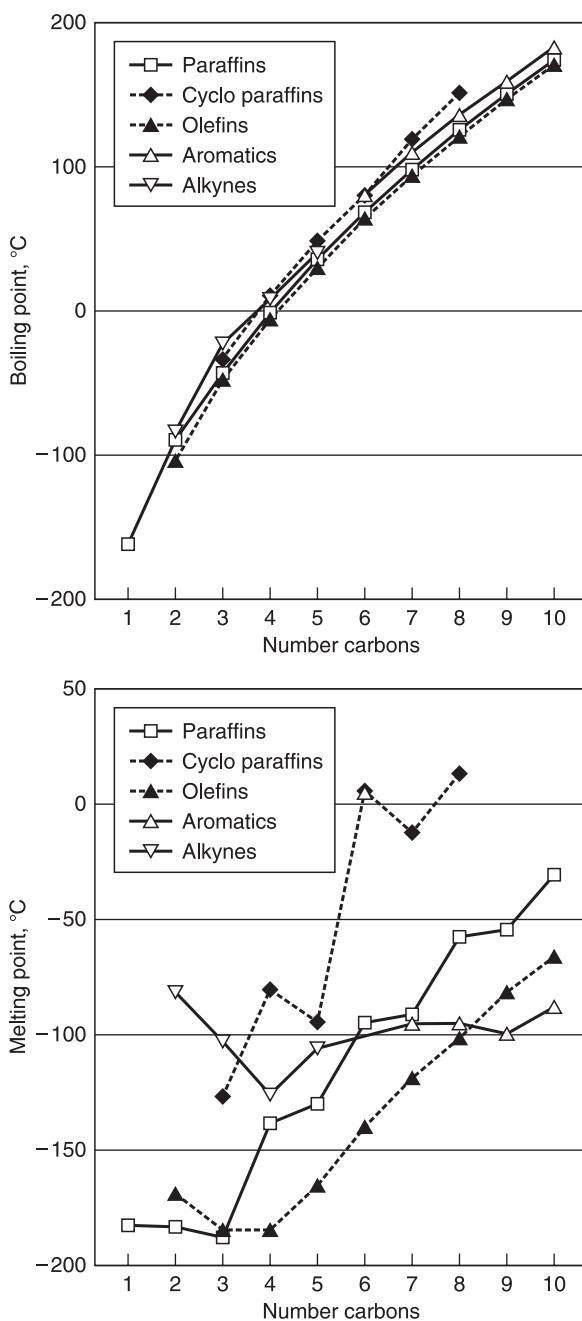


Figure 4.39 Phase transition temperatures of hydrocarbon series

density and boiling point, which indicates less efficient packing and lower cohesive forces, but by far the highest melting point—which is due to its high symmetry. When we consider the polychlorinated benzenes, there is only one form each for benzene, monochlorobenzene, pentachlorobenzene, and hexachlorobenzene; but there are three forms each for dichlorobenzene, trichlorobenzene, and tetrachlorobenzene, depending on the relative positions of the substitution groups. Figure 4.38 shows that the boiling points are largely dependent only on the number of chlorine atoms on the benzene ring, but the melting points are strikingly higher for the symmetrical molecules of benzene ( $\sigma = 12$ ), 1,4-dichlorobenzene ( $\sigma = 4$ ), 1,3,5-trichlorobenzene ( $\sigma = 6$ ), 1,2,4,5-tetrachlorobenzene ( $\sigma = 4$ ), and hexachlorobenzene ( $\sigma = 12$ ).

Figure 4.39 shows the melting and boiling points of five series of hydrocarbon compounds plotted against the number of carbon atoms. The left side of figure 4.39 shows that the boiling points of these five series of hydrocarbons depend largely only on the number of carbon atoms, but the right side of figure 4.39 shows that the melting points have a high degree of complexity. An example is the low-symmetry cyclopentane ( $\sigma = 2$ ), which has a much lower melting point than the high-symmetry cyclohexane ( $\sigma = 6$ ).

## References

- Dean J. A., ed., 1999. *Lange's Handbook of Chemistry*, London: McGraw-Hill.
- Goodman, J. M. 1998. *Chemical Applications of Molecular Modeling*. Cambridge: Royal Society of Chemistry.
- Karplus, M. and R. N. Porter. 1970. *Atoms and Molecules*. Menlo Park, CA: W. A. Benjamin.
- Pauling, L. 1960. *The Nature of the Chemical Bond*. Ithaca, NY: Cornell University Press.
- Westmoreland, P. R. and A. Z. Panagiotopoulos. 2004. Molecular-modeling methods and use for product and process design. In *Sixth International Conference on Foundations of Computer-aided Process Design*, Floudas C. A. and R. Agrawal eds; pp. 83–95.

## Further Reading

- Ashcroft, N. W. and N. D. Mermin. 1976. *Solid State Physics*. New York: Harcourt Brace.
- Assael, M. J., J. P. M. Trusler, and T. F. Tsolakis. 1996. *Thermophysical Properties of Fluids*. London: Imperial College Press.
- Bondi, A. 1968. *Physical Properties of Molecular Crystals, Liquids, and Glasses*. New York: John Wiley.
- Chaikin, P. M. and T. C. Lubensky. 1995. *Principles of Condensed Matter Physics*. Cambridge: Cambridge University Press.
- Cotton, F. A. 1990. *Chemical Applications of Group Theory*. 3rd edition. New York: John Wiley.
- DeKock, R. L. and H. B. Gray. 1980. *Chemical Structure and Bonding*. Menlo Park, CA: Benjamin/Cummings.
- Di Podesta, M. 1996. *Understanding the Properties of Matter*. London: University College London Press.
- Doucet, J.-P. and J. Weber. 1996. *Computer-Aided Molecular Design*. New York: Academic Press, 1996.

- Eliel, E. L. and S. H. Wilen. 1994. *Stereochemistry of Organic Compounds*. New York: John Wiley.
- Hamermesh, M. 1962. *Group Theory and its Application to Physical Problems*. New York: Dover.
- Hargittai, I. and M. Hargittai. 1995. *Symmetry Through the Eyes of a Chemist*. New York: Plenum Press.
- Hildebrand, J. H. and R. L. Scott. 1962. *Regular Solutions*. Englewood Cliffs, NJ: Prentice-Hall.
- Hill, T. L. 1960. *Introduction to Statistical Thermodynamics*. Reading, MA: Addison-Wesley.
- Israelachvili, J. N. 1991. *Intermolecular and Surface Forces*. London: Academic Press.
- Jeans, J. H. 1954. *The Dynamic Theory of Gases*. New York: Dover.
- Kittel, C. 1966. *Introduction to Solid State Physics*. New York: John Wiley.
- Kohler, F. 1972. *The Liquid State*. Verlag Chemie.
- Ladd, M. F. C. 1979. *Structure and Bonding in Solid State Chemistry*. New York: John Wiley.
- Leach, A. R. 1996. *Molecular Modeling: Principles and Applications*. Harlow: Longman.
- Lomont, J. S. 1959. *Applications of Finite Groups*. New York: Academic Press.
- Margenau, H. and N. R. Kestner. 1969. *Theory of Intermolecular Forces*. London: Pergamon.
- McQuarrie, D. A. and J. D. Simon. 1999. *Molecular Thermodynamics*. Sausalito, CA: University Science Books.
- McWeeny, R. 1979. *Coulson's Valence*. Oxford: Oxford University Press.
- Pauling, L. 1970. *General Chemistry*. New York: Dover.
- Pauling, L. and R. Hayward. 1964. *The Architecture of Molecules*. San Francisco, CA: W. H. Freeman.
- Smart, L. and E. Moore. 1992. *Solid State Chemistry*. 2nd edition. Cheltenham: Stanley Thornes.
- Streitwieser, A. and P. H. Owens. 1973. *Orbital and Electron Density Diagrams*. New York: Macmillan.
- Tabor, D. 1991. *Gases, Liquids and Solids*. Cambridge: Cambridge University Press.
- Ubbelohde, A. R. 1978. *The Molten State of Matter: Melting and Crystal Structure*. New York: John Wiley.
- Walton, A. J. 1983. *Three Phases of Matter*. Oxford: Clarendon Oxford University Press.
- Weyl, H. 1952. *Symmetry*. Princeton, NJ: Princeton University Press.

## Exercises

There are many computational chemistry software packages available that enable students to make and study the properties of molecules, by quantum mechanics and molecular mechanics methods. These include Gaussian, HyperChem, Spartan, and Biosym. The first four homework problems assume that the student has one of these software packages.

1. Make the following molecules with a computational chemistry software: pentane, isopentane (2-methylbutane), and neopentane (2,2-dimethylpropane). Note: do not forget to do geometry optimization!
  - (a) Print a picture of these three molecules, in your most informative display.
  - (b) Measure their bond lengths of H—C and C—C.

- (c) Measure their bond angles of H—C—H and H—C—C.  
(d) Measure their bond torsions of H—C—C—H and H—C—C—C.  
(e) When you put each molecule in a minimum shoe box, what are the length, the width, and the height?  
(f) What is the diameter of the smallest circular channel that they can pass through?  
(g) Calculate their heats of formation, and check against handbook values.
2. Make the following molecules: *o*-dichlorobenzene, *m*-dichlorobenzene, and *p*-dichlorobenzene.
- (a) Print a picture of their total electron densities.  
(b) What are their lengths, widths, and heights? What is the smallest diameter channel that they can pass through?  
(c) Calculate their moments of inertia, and note the positions of their axes.  
(d) Calculate their dipole moments, and note their directions.  
(e) Calculate their molecular point groups.
3. A molecular sieve is a solid that has holes or channels of molecular sizes. We have a sieve with 5.0 Å diameter channels and another one with 7.0 Å diameters.
- (a) Can the three pentane molecules in (1) get through these two channels?  
(b) Can the three dichlorobenzene molecules in (2) get through these two channels?
4. Provide a graph of the structure, the symmetry group, and the symmetry number of the following molecules: (a) chlorobenzene, (b) 1,3,5-trichlorobenzene, (c) 1,3-dichloro-5-bromobenzene, (d) chloroethane, (e) chloromethane, (f) dichloromethane, (g) anthracene, (h) phenanthrene, (i) adamantine, (j) coronene.
5. The transport coefficients of simple gas molecules are related to the mean free paths by the following formulas

$$\mu = \frac{1}{3} \rho \bar{u} \lambda$$
$$k = \frac{1}{3} \rho C v \bar{u} \lambda$$
$$D = \frac{1}{3} \bar{u} \lambda$$

- (a) Calculate the mean free paths of pure hydrogen and that of oxygen in air at room temperature and pressure.  
(b) Calculate the viscosity, thermal conductivity, and diffusivity of pure hydrogen and of oxygen in air, according to their mean free paths at room temperature and pressure. Compare with the handbook values of viscosity, conductivity, and diffusivity, and see how accurate they are.

- (c) Repeat the calculation of mean free paths and transport coefficients with the more complicated molecules of water and  $\text{SO}_2$ .
- (d) Repeat the calculations of mean free paths and transport coefficients with the more complicated molecules of hexane and benzene.
6. After the qualitative success of the van der Waals equation in explaining the properties of real gases, people noticed that it is not always accurate and so sought improvements. Dieterici came up with the formulation

$$P(V - b) = RT \exp\left(-\frac{a}{VRT}\right)$$

with the consequences that

$$V_c = 2b$$

$$P_c = \frac{a}{4e^2 b^2}$$

$$T_c = \frac{a}{4Rb}$$

$$Z_c = \frac{P_c V_c}{RT_c} = \frac{2}{e^2}$$

Take the data from table 4.27, and use the values of  $a$  and  $b$  to calculate the predicted values of  $V_c$ ,  $P_c$ ,  $T_c$  and  $Z_c$ .

- (a) Compare the accuracy of the Dieterici formula with the van der Waals formula for  $V_c$ .
- (b) Compare the accuracy for  $P_c$ .
- (c) Compare the accuracy for  $T_c$ .
- (d) Compare the accuracy for  $Z_c$ .
7. You are an engineer working for a developer of car engines for alternate fuels, and you have been given the assignment of packing gaseous hydrogen and methane into gasoline tanks at elevated pressures. You are pretty sure that, at room temperature and pressure, hydrogen and methane follow the ideal gas law, so that you can compute how many grams of gases you can pack into a 20-gallon tank.
- (a) If gasoline has a density of 0.75 g/mL, how many grams of gasoline can you pack into a 20-gallon tank?
- (b) When the pressure is increased to 10 atm, what do you estimate to be the compressibility  $Z$  for hydrogen and for methane, and consequently how many grams of each gas can you pack into a 20-gallon tank?
- (c) Repeat the calculation in (b) when the pressure is increased to 100 atm.
- (d) Repeat the calculation when the pressure is increased to 1000 atm. How does that compare with the packing effectiveness of gasoline?

8. Propane, acetaldehyde, and ethanol are approximately equal in their molecular weights, but they differ a great deal in their properties.

	<i>Propane, C<sub>3</sub>H<sub>8</sub></i>	<i>Acetaldehyde, CH<sub>3</sub>CHO</i>	<i>Ethanol, C<sub>2</sub>H<sub>5</sub>OH</i>
MW	44	44	46
No. atoms	11	7	9
No. valence electrons	20	18	20
No. electrons	26	24	26
Van der Waals <i>a</i>	8.664	11.37	12.56
Van der Waals <i>b</i>	0.0844	0.08695	0.0871
B.p., °C	−42.4	20.1	78.4
M.p., °C	−187.6	−123	−114
Dipole moment, Debye	0	2.69	1.70

- (a) It is suggested that ethanol is more cohesive than acetaldehyde, which in turn is more cohesive than propane. Can you provide theoretical reasoning to support this suggestion, or can you refute this reasoning? Analyze the data to see whether this set of properties is consistent with this suggestion.
- (b) Given your line of reasoning, what would you predict to be the comparative properties of formic acid, HCOOH?
9. The adhesive forces between molecules are related to the distances between them when they are pushed close to each other.
- (a) When you push two molecules of normal pentane together for the closest approach, what is the shortest distance between a carbon atom in the first molecule to a carbon atom in the second molecule? How many atoms in the first molecule have a similar short distance to one or more atoms in the second molecule?
- (b) When you push two molecules of isopentane (2-methylbutane) together, how many carbon atoms can be in close contact?
- (c) Repeat for neopentane (2,2-dimethylpropane).
10. Rank the following compounds in their entropies, from the lowest to the highest, and provide reasons for your ranking: normal hexane, 2-methylpentane, 3-methylpentane, 2,2-dimethylbutane, 2,3-dimethylbutane.
11. We will learn more about heat capacities after some calculations.

Let us do the rotation energy for H<sub>2</sub>, HCl, and Cl<sub>2</sub>. The moment of inertia is  $I = \mu r^2$  where  $\mu = (m_1 m_2) / (m_1 + m_2)$ , and the characteristic temperature of rotation is  $\Theta_r = h^2 / 8\pi^2 kI$ .

- (a) Compute these values and see how they compare with the experimental values.

Molecule	$r, \text{nm}$	$\Theta_r, \text{K, experimental}$
$\text{H}_2$	0.0742	85
$\text{HCl}$	0.1275	15
$\text{Cl}_2$	0.1988	0.35

Hint:  $h$  is the Planck's constant,  $k$  is the Boltzmann constant, or the gas constant  $R$  divided by Avogadro's number, and  $J = (\text{kg m}^2)/\text{s}^2$ .

- (b) Consider the molecules  $\text{H}_2$ ,  $\text{H}_2\text{O}$ ,  $\text{NH}_3$ , and  $\text{CH}_4$ . What are the contributions to their heat capacities  $C_v$  from translation, rotation, and vibration? What are the values of their heat capacities at very low temperatures and at very high temperatures?
12. Let us compute the  $C_v$  of the molecules  $\text{H}_2$ ,  $\text{HCl}$ , and  $\text{Cl}_2$  from 0 to 100 K. Quantum mechanics shows that for the vibration contributions towards heat capacity

$$\frac{C_v}{R} = \frac{y^2 e^{-y}}{(1 - e^{-y})^2}$$

where  $y = \Theta/T$ . Plot  $C_v$  versus  $T$  in the range of 0 to 1000 K for these three gases.

13. Let us continue with the example about solubility, by putting balls in a checkerboard with 25 squares.

For  $m = 0$  to 10 balls, make a list of the values of:  $W$ , number of configurations in putting  $m$  balls in 25 squares;  $S$ ,  $k \ln W$ , the entropy;  $U$ , the energy of the system,  $-(1 - x)^2 E$ ;  $x$ , the mole fraction of solute,  $m/25$ ;  $A$ , the Helmholtz free energy,  $U - ST$ .

For  $E/kT = 10, 15, 20, 25, 30$ , what is the value of  $m$  when  $A$  reaches a minimum; that is, what is the solubility  $x = m/25$ ? It is an even better idea to make a plot of  $x$  versus  $kT/E$ , so we can see more clearly how  $x$  depends on  $T$ .

# Predictions by Correlations

## 5.1 Populations, Predictors, and Functions

Populations: families of similar substances

Predictors for a given property

Functional forms

## 5.2 Regression with a Single Predictor

Simple linear regression

Nonlinear regression

## 5.3 Regression with Several Predictors

Multilinear regressions

Multipredictor nonlinear correlations

Quantitative property–property relationships and quantitative structure–property relationships

## 5.4 Group Contribution Methods

Method of Joback

Method of Hansch and Leo

## 5.5 Mixtures and Composite Materials

Fluid mixtures

Composite material

### References

### Further Reading

### Websites

### Exercises

## 5.1 Populations, Predictors, and Functions

After searching the literature and making predictions based on theory without getting sufficient satisfactory results, the next move would be to make estimates. We need the property  $y$  of substances  $p_i$  from a population  $P$  that has not been investigated and reported in the literature. Fortunately, there exists a subset  $S$  of  $P$  that has been investigated, and we have the values for the property  $y$ . For instance, we may want the boiling points of all the hydrocarbons, but we have only the boiling points of the normal paraffins from 1 to 20 carbon atoms. Can we use this piece of information on normal paraffins to estimate the boiling points for the rest of the hydrocarbon population? How much effort would be involved and how accurate would the results be?

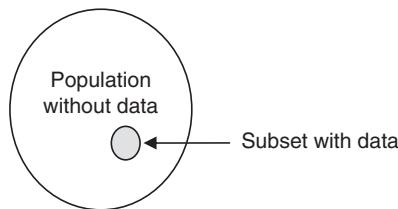


Table 5.1 Numbers of paraffin and olefin isomers

Carbons	Paraffin isomers	Olefin isomers
1	1	0
2	1	1
3	1	2
4	2	4
5	3	6
6	5	17
7	9	
8	18	
9	35	
10	75	
11	159	
12	355	
13	802	
14	1,858	
15	4,347	
20	366,319	
25	36,797,588	
30	4,111,846,763	
40	62,491,178,805,831	

The number of isomers of paraffin is very large; see table 5.1. We see that the iso-paraffins are not as well investigated as the normal paraffins. We have the boiling points of all three isomers of pentane, but not the 75 isomers of decane. It is inevitable that we have to resort to estimations.

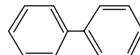
When we have obtained a good correlation for normal paraffins, we would naturally want to know if we can extend this to the branched paraffins, and onward to the population of all the saturated hydrocarbons (by including the cyclic paraffins), and onward to the population of all hydrocarbons (by including olefins, acetylenes, and aromatic compounds), and then onward to the population of all organic compounds (by including compounds with heteroatoms, such as O, N, Cl). A correlation that applies accurately to a larger domain is more useful than one that works only for a smaller domain.

Another example is polychlorinated biphenyls (PCBs), which have 10 hydrogen atoms that can be substituted by chlorine atoms. There are three types of site: the four  $\alpha$  sites near the bridge between the two phenyl fragments, the four  $\beta$  sites farther

**Table 5.2 Numbers of PCB isomers versus Cl content**

No. of Cl	No. of isomers
0	1
1	3
2	12
3	24
4	42
5	46
6	42
7	24
8	12
9	3
10	1

away from the bridge, and the two  $\gamma$  sites that are the farthest away from the bridge. The number of isomers is shown in table 5.2.



We need many of their properties for safety and environmental protection, such as vapor pressure, solubility in water, and partition between oil and water. We have the data for all the PCBs with 0, 1, 2 or 10 chlorine atoms, but we have only some of the data for PCBs with three to nine chlorine atoms.

A correlation is based on the proposition that, for a particular population  $S$ , the property  $y$  is related to one or several independent variables, often called “predictors”  $(x_1, x_2, \dots, x_k)$ ; the correlation is based on a particular function  $y = f(x_1, x_2, \dots, x_k)$  and a set of adjustable parameters  $(c_0, c_1, c_2, \dots)$ . The optimal parameter values for  $c_i$  are extracted from the data, which minimizes the differences between data and predictions. The principal usefulness of a regression formula, which has been proven valid only in  $S$ , is in its success in estimating the property in the bigger population of  $P$  (table 5.3). Some investigators would immediately proceed to export this correlation from the training set  $S$  to the larger population  $P$  without further assurances that it would still work.

A more conservative approach would be to split the set  $S$  into two subsets  $S_a$  and  $S_b$ , (i) using  $S_a$  to derive parameter values and (ii) testing the accuracy of the correlation on  $S_b$  to see it still works. The only way to tell whether the extrapolation of this formula will be successful is to make subsequent measurements in selected regions in  $P$  and to compare with predictions. A good fit would increase confidence in the correlation,

**Table 5.3 Properties of  $S$  and  $P$**

	Subset $S$	Population $P$
Number of substances	$n$	$m \gg n$
Data on property $y$	Have	Have not
Data on predictors $x_1, \dots, x_k$	Have	Have
Accuracy of correlation and parameter values	Established	Hopeful

and a poor fit should involve fine tuning the correlation, leading to another round of prediction and validation, in accordance with the scientific method.

The best predictors and functions have theoretical justifications. The boiling point is related to the attractive or cohesive forces between molecules, which hold the liquid together in opposition to the random thermal forces of  $kT$  that disrupt this cohesion. A good predictor should be related to the cohesive intermolecular forces, which may be a part of the molecular structure, such as:  $N_C$  the number of carbon atoms,  $N_H$  the number of hydrogen atoms, or MW the molecular weight; or it may be a related property that is more readily available, such as the polarizability  $\alpha$ , the dipole moment  $\mu$ , the strength of the hydrogen bond, the melting point  $T_m$ , etcetera. Sometimes a predictor is ad hoc, or chosen without theoretical guidance, and is justified only on its pragmatic utility. In an ideal situation, the functional form is suggested by theory. When this is not available or inconvenient, we explore simple empirical forms, such as linear, logarithmic, exponential, polynomial, and others.

When a linear one-predictor correlation is satisfactory, we may have discovered the major cause for this property, with no more than minor contributions from other factors. When the results are not satisfactory, there are many possible reasons and, thus, many approaches to improving the results, such as: using another predictor, using more than one predictor, and using a different functional form. A robust and highly useful correlation should have one or more of the following desirable properties:

1. *Cause-and-effect explanation.* When a correlation uses only a few predictors that have physical-chemical significance, and this can explain a large quantity of data from a wide area accurately, it leads to confidence that we have discovered a scientific cause-and-effect explanation of the property  $y$ , thus deepening our understanding of the physics and chemistry of the phenomenon.
2. *Extrapolation.* The correlation can be successfully extrapolated from the smaller sample  $S$ , to the larger population  $P$ , where  $m \gg n$ . Accurate estimation of the property in the larger population  $P$  without the cost of time and resources would greatly help the innovation of new products.
3. *Compact information storage and retrieval.* A good correlation leads to compact storage of information in the form of equations and coefficient values, which is more convenient for rapid retrieval than tables of numbers or charts and graphs. This property, which is valuable in a forward search from substance to property, is even more valuable in a reverse search from property to substances.

A correlation that is lacking in the first attribute of cause-and-effect explanation can still be a useful tool in interpolation, and in incremental extrapolation to areas not far from sample  $S$ ; but since it does not make contributions to the understanding of physics and chemistry, it would inspire less confidence when used to extrapolate in areas far from  $S$ .

### Populations: Families of Similar Substances

When we examine the melting points of all the elements, we find that they fall into different families, such as noble gases, alkali metals, and halogens. The noble gases of group 18 have very low melting points, which increases with the period in the

periodic table; the halogens of group 17 and the oxygen elements of group 16 behave in a similar manner, but with much higher melting points. However, the alkali elements of group 1 and the carbon elements of group 16 have the reverse trend, of lower melting points at higher periods. Before we can perform correlations of melting points with the period, it is necessary first to separate the elements into different families, and to perform correlations on each family separately.

It is also not practical to expect that a single equation will correlate simultaneously the properties of NaCl with hexane, since there are too many differences between them. When we study the properties within a homologous series, such as the normal paraffins that differ only in the number of  $-\text{CH}_2-$  units, we isolate only one parameter to study. In many cases, the molecular properties change smoothly with the number of units, which can be used as the predictor. The one-parameter populations set out in table 5.4 are frequently used.

### Predictors for a Given Property

The most important step in a structure–property correlation is to discover the set of relevant structural parameters that control the property, as well as any related and readily available property parameters. The identification of these parameters may be suggested by theoretical understanding or by empirical observations of experimental measurements. There are some generally useful methods to discover the independent variables  $[x_i]$  for a specific property in a set of molecules, such as the boiling points of the normal paraffins.

### *Structure Parameters Suggested by Theory*

We begin by asking what is our theoretical knowledge of the origin and cause of boiling points. We know that molecules are held together by the intermolecular forces. For organic molecules, these are the van der Waals forces that are dependent on the polarizability  $\alpha$ , the dipole moment  $\mu$ , the hydrogen bond of OH, NH, and FH groups, and the intermolecular distances  $r$ . The permanent dipole moment is due to electrical charges on the atoms, the distances between these charges, and to molecular symmetry. The polarizability is due to the number of loosely bound valence electrons that can be distorted by an external electrical field to create instantaneous dipole moments. The most important molecular structural parameters can be divided into the following three types:

- Electronic parameters include the number of electrons, the number of valence electrons in the outer orbit, the orbital quantum numbers  $n$ , the azimuthal quantum number or spdf status, electron radius and energy, polarizability, dipole moment, quadrupole moment, and first ionization energy.
- Steric parameters include bond length, bond angle, bond torsion, conformation, topology or connection among atoms in the molecule, symmetry, moment of inertia, molecular diameter, molecular surface area, molecular volume, and molecular shape.
- Hydrophobic parameters include charge distribution, electron density, electronegativity, dipole moment, and hydrogen bond.

**Table 5.4** Frequently used one-parameter populations*Homologous series with the same units, adding repeat units of  $-CH_2-$  or benzene rings*

Normal paraffins	$CH_4, C_2H_6, C_3H_8, \dots, C_nH_{2n+2}$
Cycloparaffins	$C_3H_6, C_4H_8, C_5H_{10}, \dots$
Olefins	$C_2H_4, C_3H_6, 1-C_4H_8, \dots, 1-C_nH_{2n}$
Alcohols	$CH_3OH, C_2H_5OH, 1-C_3H_7OH, \dots$
Aromatics	Benzene, naphthalene, anthracene

*Isomer series with rearrangement of units*

Branching	<i>n</i> -Pentane, isopentane, neopentane
Positional	<i>o</i> -Xylene, <i>m</i> -xylene, <i>p</i> -xylene
Stereo	<i>cis</i> -1,2-Difluoroethylene, <i>trans</i> -1,2-difluoroethylene

*Bond saturation and cyclization series*

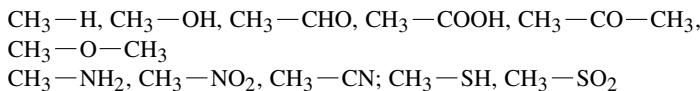
Two-carbon chain	Ethane, ethylene, acetylene
Six-carbon chain	Hexane, cyclohexane, cyclohexene, cyclohexadiene, benzene
Two cycles	Decalin, tetralin, naphthalene

*Element substitution series, from the same group but from different periods in the periodic table*

Group 1	LiF, NaF, KF, RuF, CsF
Group 11	CuCl, AgCl, AuCl
Group 14	$CH_4, SiH_4, GeH_4, SnH_4, PbH_4$
Group 15	$NH_3, PH_3, AsH_3, SbH_3, BiH_3$
Group 16	$CH_3OH, CH_3SH, CH_3SeH, CH_3TeH$
Group 17	$C_6H_5F, C_6H_5Cl, C_6H_5Br, C_6H_5I$
Group 18	He, Ne, Ar, Kr, Xe, Rn

*Element substitution from the same period, but from different groups in the periodic table*

Period 2	$CH_4, NH_3, H_2O, HF$
Period 3	$SiH_4, PH_3, H_2S, HCl$
Period 4	$GeH_4, AsH_3, H_2Se, HBr$
Period 5	$SnH_4, SbH_3, H_2Te, HI$

*Functional group substitution series*

This is a large menu to choose from, and we need to discover which are the principal controlling variables and which are the subsidiary contributing variables. When we consider the boiling points of the noble gases, we have fewer parameters to choose from. When we consider the paraffins, we have the additional factors of covalent bond lengths and angles, branching, and bond torsion. When we consider the boiling points of the hydrochlorofluorocarbons (HCFCs), we have yet another set of factors, including

number of chlorine and fluorine atoms, dipole moment, and hydrogen bond. Sometimes, the correlations created are called the quantitative structure–property relations (QSPRs).

### *Other Closely Related Properties*

A strong cohesive force, which is responsible for a large boiling point, would also tend to give rise to a larger van der Waals constant  $a$ , a larger critical temperature  $T_c$ , a larger heat of vaporization, a larger refractive index for light, and a larger surface tension. If these properties are more readily available in the literature, then they may be good candidates for predictors. Sometimes, these correlations are called the quantitative property–property relations (QPPRs).

The density of a liquid depends on the elements involved, the lengths of the bonds, and on the efficiency of packing. Within the normal paraffins, the hydrogen-to-carbon ratio and the chain length are expected to be predictors. Methane has a high ratio of 4/1 between the lighter hydrogen and the heavier carbon; as the number of carbon atoms increases, this ratio decreases to  $(n + 2)/n$  and approaches two for large  $n$ . There are a relatively longer intermolecular distance between two methane molecules, which are controlled by the weak van der Waals forces, compared with the shorter intramolecular C–C bonds in a higher paraffin, which are controlled by the strong covalent forces. So pentane should have a higher density than methane. The linear pentane would pack together in a more efficient manner than the once-branched 2-methylbutane and the twice-branched 2,2-dimethylpropane, but we are unable to calculate their effects on the density precisely. Thus, we expect that a higher paraffin with more carbon atoms will tend to have a larger density, and a linear paraffin would tend to have higher densities than a highly branched paraffin.

The melting point is more complex than the boiling point, as it also involves the transition from a highly ordered crystalline solid to a disordered liquid with a large entropy of melting. A highly symmetrical molecule, like benzene, has a much smaller entropy in the liquid phase and, consequently, a smaller entropy of melting than a less symmetrical molecule, like toluene; this gives rise to a higher melting point according to the thermodynamic equation  $T_m = \Delta H_m / \Delta S_m$ .

The chemical and biological reactions of a compound with other compounds involve a change of free energy, and a reaction is more spontaneous and probable if the products have much lower free energy than the reactants. Thus, the difference in the free energy of formation between the reactants and the products is likely to be related to reactivity. The speed of the reaction is related to the activation energy of the reaction, and often involves a promotion of an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). Thus, the energy difference between the LUMO and the HOMO is expected to be an indicator of reactivity.

### *Empirical Observations and Intuitive Hypothesis*

When we do not understand the cause and effect involved in a property, we have no guidance from theory on the relevant structural or related property parameters. This is

more often the case with biological and environmental properties, due to the complex interactions of many physical-chemical factors. We lack knowledge of the exact physiological and chemical nature of the smell of perfumes and putrid substances, and the flavor of food and spices. We make progress by examining a group of compounds with different smells, and try to discover whether there is any consistent pattern, such as “small and spherical molecules tend to smell like camphor.” Then we propose parameters to characterize “spherical” and “small” and do a correlation. We discard the parameters that do not correlate well, and keep the ones that do correlate well.

Often, the correlation is not good, and we need to search for the “hidden variable” that we have not yet discovered. But when we find a good correlation, it could prove useful in the reverse search for other untried compounds that may have higher or lower camphor smell, even if we do not understand the mechanism of how it works. There is always the hope that, if we know which parameters are important to smell, we may generate one or more hypotheses on the nature of camphor smell; this would be followed by predictions and experiments that could lead us to future understanding.

When we are truly clueless, we can nevertheless rely on intuition to propose an ad hoc set of structural and related property parameters for the correlation. We may be lucky and find the hidden variable by chance, and we may be inspired. An example is the topological index, which describes how carbon atoms are connected together, and was proposed in the hope that it would correlate a large range of molecular properties.

### Functional Forms

A correlation begins with the hypothesis that for a property such as boiling point, within a certain population such as the noble gases, and using the period of the element in the periodic table as the predictor, the appropriate correlation function is of the form

$$y = f(x) \quad y \in P$$

The predictor  $x$  can be a discrete variable, such as the number of carbon atoms, or a continuous variable, such as pressure. The simplest functional form of correlation is the linear form:  $y = a + bx$ . This form has two parameters  $a$  and  $b$ , which are to be determined by comparison with the data. This form becomes even simpler if the intercept  $a = 0$ .

All linear functions are simply alike, and all nonlinear functions are different in numerous ways. Perhaps the next level of complication is monotonic increasing functions, where the sign of the first derivative  $dy/dx$  is always positive, so the function is always rising in value. An example is  $y = x^n$  where  $n > 0$ . It would be useful here to distinguish between the convex functions, where the second derivative  $d^2y/dx^2$  is negative (when  $n < 1$ ) and concave functions, where the second derivative is positive (when  $n > 1$ ). The exponential function  $y = e^{nx}$  where  $n > 0$  is concave; the logarithm function  $y = \log(x)$  is convex.

The monotonic decreasing functions have a first derivative that is always negative, so that the function is always decreasing with  $x$ . Examples of the monotonic increasing functions that are concave include  $y = ax^n$ , where  $n$  is negative, and  $y = e^{nx}$ , where  $n$  is negative; an example of a monotonic decreasing convex function is  $y = (a - x)^n$  where  $0 < n < 1$ .

The polynomials are often used for correlations, and are often used for minor modifications for functions that are basically linear:

$$y = a + bx + cx^2 + dx^3 + \dots$$

However, they can have very exciting properties when the higher terms become large, as they can have multiple maxima and minima. For instance, take  $y = a + bx + cx^2$ , which is dominated by the constant term  $a$  at very small values of  $x$ . But at sufficient large values of  $x$ , say  $x > a/b$ , the linear term begins to be more important than the constant term; and at even larger values of  $x$ , such as  $x > \sqrt{a/c}$  and  $x > b/c$ , the quadratic term takes over. At the value of  $x = -b/2c$ , the first derivative  $dy/dx = 0$ , so that the function is either at a maximum or a minimum. If the function of interest is required by theory to be monotonic, then it may be hazardous to extrapolate outside of a limited range with a polynomial.

When the functional form of the correlation is suggested by theory, there is a great deal more confidence that the correlation can be extrapolated into regions of  $P$  that have no experimental data, and can be used for other families of compounds other than the training set  $S$ . Examples of theory-suggested functional forms include the van der Waals equation of state for gases, the Langmuir isotherm for adsorption and catalysis, and the Clausius–Clapeyron equation for the vapor pressure of liquids.

## 5.2 Regression with a Single Predictor

Let us construct a correlation for the property  $y$  of members of a sample population  $S$ , relating them to a single predictor  $x$ . We would like to know whether  $x$  and  $y$  are very closely related, so that a knowledge of  $x$  enables one to calculate  $y$  with reasonable accuracy. One use of such a correlation is easy retrieval of the values of  $y$  within  $S$ ; but it is much more useful if this correlation established in  $S$  can be extrapolated to a bigger population  $P$  where the property  $y$  has not been measured.

### Simple Linear Regression

Let us consider the boiling points of the noble gases from helium to xenon, and ask what is a good predictor. Figure 5.1 shows a plot of their boiling points against their period in the periodic table. Theory tells us that the boiling point is related to a competition between the disruptive random thermal energy of  $kT$  and the cohesive van der Waals forces holding molecules together in a condensed state. The cohesive forces that hold the noble gases molecules are of the London dispersion type, which is proportional to  $\alpha^2 I/r^6$ :  $\alpha$  is the polarizability of the molecules,  $I$  is the first ionization potential, and  $r$  is the distances between the molecules. All three parameters depend on the period in the periodic table, as well as the number of electrons, the number of valence electrons, and the atomic weight. Table 5.5 lists the boiling points and the values of these predictors for the noble gases. The atomic radii of the elements would serve as approximations of the distances between molecules in the liquid phase.

It is seen that the boiling points and most of the candidates for a predictor are monotonic increasing with the period  $P$ , with the exception of the ionization energy,

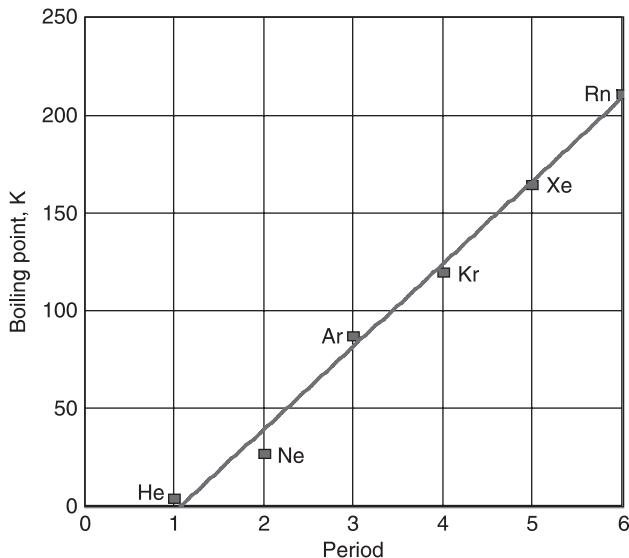


Figure 5.1 Boiling points of noble gases compared with linear regression

Table 5.5 Boiling points and predictors of noble gases<sup>a</sup>

	$T_b, K$	$P$	AN	AW	$V$	$\alpha$	$I, eV$	$r, \text{\AA}$
Helium	4.165	1	2	4.003	2	0.2051	24.587	1.53
Neon	27.05	2	10	20.183	8	0.395	21.564	1.60
Argon	87.23	3	18	39.944	8	1.64	15.759	1.91
Krypton	119.88	4	36	83.7	18	2.48	13.999	2.0
Xenon	165.06	5	54	131.3	18	4.04	12.130	2.2
Radon	211.1	6	86	222	32		10.748	

<sup>a</sup> $T_b$ : boiling point;  $P$ : period; AN: atomic number; AW: atomic weight;  $V$ : number of valence electrons;  $\alpha$ : polarizability;  $I$ : first ionization energy;  $r$ : atomic radius.

which is monotonic decreasing. We can try all these related variables one at a time and see which one does the best job of correlation, such as measured by the largest value for  $R^2$ .

Let us propose a correlation for the boiling points of the six noble gases of the form

$$y' = C_0 + C_1x$$

$$y = y' + \varepsilon = C_0 + C_1x + \varepsilon$$

where  $y$  (K) is the boiling point,  $x$  is the period in the periodic table,  $\varepsilon$  is the error or  $(y - y')$ , and  $y'$  is the estimate of  $y$ , which is equal to  $C_0 + C_1x$ .

This function has one predictor  $x$  and two arbitrary coefficients  $C_0$  and  $C_1$ . Since the noble gases produce  $n = 6$  sets of data and there are two arbitrary coefficients, the remaining degrees of freedom is  $6 - 2 = 4$ . The principle of linear regression is to minimize the sum of the squares of the error term

$$\sum (C_0 + C_1 x_i - y_i)^2 = \sum (y'_i - y_i)^2 = \sum \varepsilon_i^2$$

We will need the means, or average values, of  $\langle x \rangle$  and  $\langle y \rangle$ , given by the sums divided by the sample sizes; we also require the variances  $V(x)$  and  $V(y)$ , the standard deviations  $\text{sd}(x)$  and  $\text{sd}(y)$ , the covariance  $V(x, y)$ , and the correlation coefficient  $r$ .

$$\begin{aligned}\langle x \rangle &= \frac{\sum x_i}{n}, \quad \langle y \rangle = \frac{\sum y_i}{n} \\ V(x) &= \frac{\sum (x_i - \langle x \rangle)^2}{n}, \quad \text{sd}(x) = \sqrt{V(x)} \\ V(y) &= \frac{\sum (y_i - \langle y \rangle)^2}{n}, \quad \text{sd}(y) = \sqrt{V(y)} \\ \text{Cov}(x, y) &= \frac{\sum (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{n} \\ r &= \frac{\text{Cov}(x, y)}{\text{sd}(x)\text{sd}(y)}\end{aligned}$$

A simple linear regression is very easily computed with an ordinary computer spreadsheet program such as Lotus 1-2-3 or Microsoft Excel. The method involves calculating first the average values of  $x$  and  $y$ , then the values of  $C_0$  and  $C_1$ , and then the estimated value of  $y'_i$ :

$$C_1 = \frac{\text{Cov}(x, y)}{V(x)} = \frac{\sum (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sum (x_i - \langle x \rangle)^2}$$

$$C_0 = \langle y \rangle - C_1 \langle x \rangle$$

$$y' = C_0 + C_1 x$$

For the series above, the result for the boiling points of the noble gases are  $y' = -49.3 + 43.8x$ , where  $y'$  is the estimated value of  $T_b$  and  $x$  is the period  $P$ . The goodness of fit is usually evaluated by calculating the square correlation function  $R^2$ , and the standard deviation  $\text{sd}$  of the error term  $(y - y')$ :

$$\begin{aligned}R^2 &= 1 - \frac{\sum (y'_i - y_i)^2}{\sum (y_i - \langle y \rangle)^2} = \frac{[\text{Cov}(x, y)]^2}{V(x)V(y)} \\ \text{sd}(y - y') &= \sqrt{\frac{\sum (y'_i - y_i)^2}{\text{df}}}\end{aligned}$$

The coefficient of determination  $R^2$  measures the goodness of the regression, and varies between zero for no linear correlation and unity for perfect linear correlation. It is also a measure of the covariance divided by the two variances of  $x$  and  $y$ . The formula for the standard deviation of the error term ( $y - y'$ ) uses the degree of freedom instead of sample size as denominator; this is  $n - 2$  here, as there are  $n$  data points and two arbitrary coefficients. The value of  $R^2 = 0.992$ ; this is quite good, as it is sometimes interpreted as “the correlation is 99.2% satisfactory.” However, this is not such a great feat, as we began with a very small sampling of only six data points with two arbitrary coefficients. The value of the standard deviation of ( $y - y'$ ) is  $sd(y - y') = 8.0$ . If the errors follow a normal distribution, then we can expect that there is a 31.8% probability of finding an error term ( $y_i - y'_i$ ) greater than one standard deviation, or 8.0 K, that there is a 4.6% probability of finding an error term greater than two standard deviations, or 16.0 K, and that there is a 0.4% probability of finding an error term greater than three standard deviations, or 24.0 K. The result can be summarized as

$$T'_b = -45.7 + 42.3P \quad sd(y - y') = 8.01, R^2 = 0.992,$$

$$sd(C_0) = 8.0, sd(C_1) = 1.9$$

An examination of the graph of  $T_b$  versus  $P$  for the six noble gases shown in figure 5.1 reveals that the points do not lie on a straight line, so that no simple straight line can hope to go through all of them. The standard deviation of error is 8.0 K, which is much larger than the actual boiling point of helium at 4.2 K, and the predicted value of  $T_b$  of helium takes on the unreasonable value of -3.4 K. There are several possible causes for the poor results and remedies:

1. Find another predictor that would do a better job.
2. Use the method of weighted least squares and attach different weights according to the importance of the individual error terms  $\varepsilon_i$ .
3. Find a nonlinear function  $T_b = f(P)$ , such as logarithmic or polynomial.
4. Discover and add one or more “hidden variables” that have a strong influence on the value of  $y$ .

When we use the other predictors to perform all the one-parameter linear correlations, we found the correlation results listed in table 5.6. All these predictors are highly correlated with each other. The parameter with the best theoretical justification

**Table 5.6** One-parameter linear correlations for the predictors

Predictor	$C_0$	$C_1$	$R^2$	$sd(y - y')$
$\alpha^2 I/r^6$	2.92	91.57	0.9923	6.7
Period $P$	-45.7	42.3	0.9919	8.0
Polarizability $\alpha$	8.58	41.2	0.9729	12.6
Atomic number	18.5	2.44	0.9387	22.0
Atomic weight	24.8	0.929	0.9197	25.2
Valence electrons	2.09	7.00	0.8857	30.1

is  $\alpha^2 I/r^6$ , and it is slightly better than the period  $P$ ; however, this is achieved with only five sample points, as there are no data on the polarizability of radon.

The method of “weighted least squares” can be used to avoid the unsatisfactory negative boiling point predicted for helium, by minimizing the following sum

$$\text{minimize } \sum w_i \varepsilon_i^2 = \sum w_i (C_0 + C_1 x_i - y_i)^2 = \sum w_i (y'_i - y_i)^2$$

where  $w_i$  is a set of weights with positive values. For instance, if  $w_i = 1/y_i^2$ , then we minimize the sum of the percentage errors. The computations for the values of  $C_0$  and  $C_1$  are

$$C_0 = \frac{\sum w_i x_i^2 \sum w_i y_i - \sum w_i x_i \sum w_i x_i y_i}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2}$$

$$C_1 = \frac{\sum w_i x_i y_i \sum w_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2}$$

In the example of the boiling points of the six noble gases, the weight of  $w_i = 1/y_i^2$  will result in the correlation

$$T'_b = -30.0 + 34.1P$$

The standard deviation of the error has increased from 8.1 to 18.5 K, but the percentage error has decreased from 90.5 to 21.1%. The most important advantage of this method is that the predicted value of the boiling point of helium is 4.01 K, in the positive range.

This linear correlation of  $T'_b = -45.7 + 42.3P$  is constructed here to illustrate the method, but it has limited utility except to predict the boiling point of the next noble gas with  $P = 7$ , which is estimated here to be 257.3 K, or  $-15.8^\circ\text{C}$ .

## Nonlinear Regression

If the linear regression is not satisfactory, even after trying out a number of predictors, then the remedy may be to find another functional form. We distinguish between “nonlinear trend” and “scatter.” In the former case, the data of  $y$  change smoothly with the independent variable  $x$ , and there is a good cluster along a nonlinear trend, which shows that we have found the right independent variable but that we do not have the right functional form  $y = f(x)$ . On the other hand, a wide scatter in the data would indicate that changing the functional form would not fix matters, and we would need to look for hidden variables and add them as predictors.

There are an infinite number of nonlinear functions that can be proposed, depending on the shape of the data. The most common functions used for regression in the region of  $x > 0$  include those listed in table 5.7. Some of the more popular functional forms are shown in figure 5.2. A function should be chosen with a shape that matches the general shape of the data.

Table 5.7 Common functions used for regression

	<i>Convex, <math>d^2y/dx^2 &gt; 0</math></i>	<i>Concave, <math>d^2y/dx^2 &lt; 0</math></i>
Monotonic increasing, $dy/dx > 0$	$x^n, n > 0$ $e^{ax}, a > 0$	$x^n, 1 > n > 0$ $x/(a+x), a > 0$ $\log(x)$
Monotonic decreasing, $dy/dx < 0$	$e^{-ax}, a > 0$ $x^{-n}, n > 0$ $(1-x)/(a+x), a > 0$	$(1-x)/(a-x), a > 0$

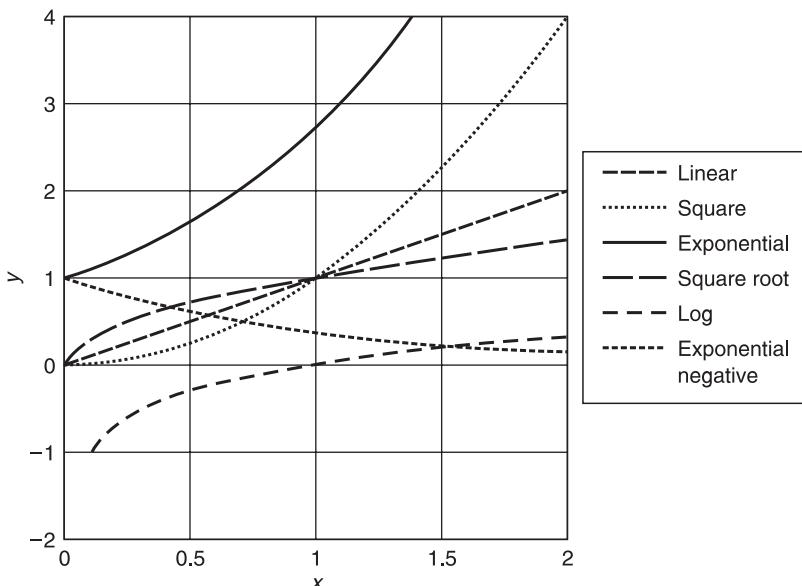


Figure 5.2 Simple nonlinear functions

The polynomial functions are approximately linear in a small region near  $x = 0$ , but they can have several minima and maxima, when the derivative  $dy/dx = 0$  in regions far from  $x = 0$ :

$$y = C_0 + C_1x + C_2x^2 + C_3x^3 + \cdots + C_kx^k$$

### Boiling Points of the Noble Gases

Let us consider as  $y$  the boiling points of the noble gases and then use the atomic number as the predictor  $x$ . This set of data is monotonic increasing, but without a clear

convex or concave character. A little computation will show these correlations with increasing accuracy:

exponential	$y = 16.2 \exp(0.0375x)$	$R^2 = 0.631$
logarithm	$y = -62.6 + 55.3 \ln(x)$	$R^2 = 0.894$
linear	$y = 18.5 + 2.45x$	$R^2 = 0.939$
power	$y = 2.37x^{1.07}$	$R^2 = 0.965$
quadratic	$y = -3.29 + 4.35x - 0.022x^2$	$R^2 = 0.983$
cubic	$y = -10.2 + 5.59x - 0.062x^2 + 0.00047x^3$	$R^2 = 0.986$

The exponential and logarithm functions have clear convex and concave characters, and do not do well in a nearly linear function. The last three functions provide minor corrections on the linear function, and do quite well. It should be pointed out that the first four functions have two arbitrary parameters each, the quadratic has three and the cubic has four. We expect that when the number of parameters increases, we can correlate ever more complicated data. For the cubic equation with three parameters and only six data points, there remain only three degrees of freedom. It is often said that “Give me three parameters, and I can fit an elephant,” so that it is a greater achievement to fit complicated data with as few parameters as possible, in the spirit of what is called the “Occam’s Razor.”

### *Boiling Points of Normal Paraffins*

Let us consider the boiling points of normal paraffins in degrees Celsius at 1 atm pressure: from methane to ethane, propane, on to the 20-carbon member of *n*-eicosane. This is clearly a monotonic increasing and concave function. In this family of straight chain paraffin’s, each member adds a —CH<sub>2</sub>—group to the previous member, which results in an increase in their melting points and boiling points, shown in table 5.8. As the molecules increase in the number of carbon atoms, their boiling points increase steadily and smoothly. When you add a —CH<sub>2</sub>—group to methane to form ethane, the result is a 73.1 °C increase of the boiling point; when you continue to add —CH<sub>2</sub>—groups to ethane to form propane, *n*-butane, and finally to *n*-eicosane, the effect of each additional group to the boiling point gradually drops to 13.8 °C. What is the boiling point of *n*-C<sub>21</sub>, the next paraffin with 21 carbon atoms? If you estimate that it would be 617 + 14 = 631 °C, it may be close enough for a number of engineering design purposes; but if you want to go into commercial production, you should first prove it with synthesis and measurement.

Let us start with a linear regression formula with two parameters:

$$y'_i = C_0 + C_1 x_i$$

where  $y'_i$  is the estimated boiling point and  $x_i$  is the number of carbon atoms in normal paraffin,  $C_0$  and  $C_1$  are empirical fitting parameters. We found that for the boiling points of the normal paraffin with the number of carbons, the best least square

Table 5.8 Boiling and melting points of normal paraffins

	$N_C$	$MW$	$M.p., ^\circ C$	$\Delta(m.p.), ^\circ C$	$B.p., ^\circ C$	$\Delta(b.p.), ^\circ C$
$CH_4$	1	16	-182.4		-161.7	
$C_2H_6$	2	30	-182.7	-0.3	-88.6	73.1
$C_3H_8$	3	44	-187.6	-4.9	-42.2	46.4
$C_4H_{10}$	4	58	-138.2	49.4	-0.5	41.7
$C_5H_{12}$	5	72	-129.7	8.5	36.1	36.6
$C_6H_{14}$	6	86	-94.0	35.7	68.7	32.6
$C_7H_{16}$	7	100	-90.5	3.5	98.4	29.7
$C_8H_{18}$	8	114	-56.8	33.7	125.6	27.2
$C_9H_{20}$	9	128	-53.7	3.1	150.7	25.1
$C_{10}H_{22}$	10	142	-29.7	24.0	174.0	23.3
$C_{11}H_{24}$	11	156	-25.5	4.2	196.0	22.0
$C_{12}H_{26}$	12	170	-9.5	16.0	216.4	20.4
$C_{13}H_{28}$	13	184	-5.4	4.1	235.5	19.1
$C_{14}H_{30}$	14	198	5.5	10.9	252.5	17.0
$C_{15}H_{32}$	15	212	10.0	4.5	270.7	18.2
$C_{16}H_{34}$	16	226	18.2	8.2	286.9	16.2
$C_{17}H_{36}$	17	240	22.0	3.8	302.2	15.3
$C_{18}H_{38}$	18	254	28.2	6.2	316.8	14.6
$C_{19}H_{40}$	19	268	32.2	4.0	330.0	13.2
$C_{20}H_{42}$	20	282	36.5	4.3	343.8	13.8

fit is with:

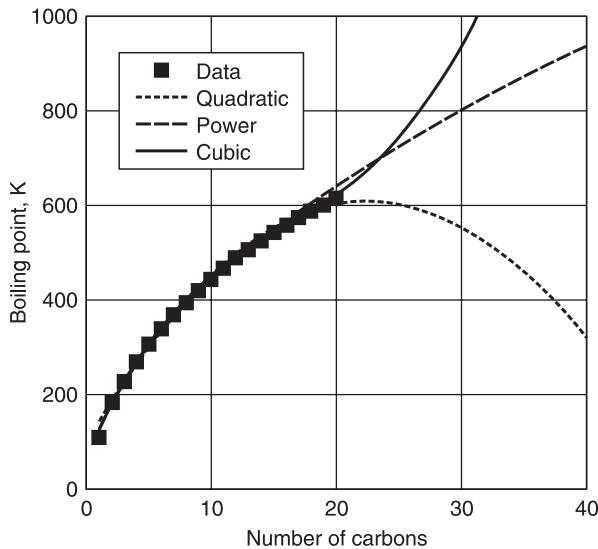
$$T'_b = -100.7 + 24.42x \quad \text{sd}(y - y') = 32.4, R^2 = 0.9546$$

This is not a very satisfactory fit even when the value of  $R^2 = 0.9546$ , as the standard error of boiling point estimation is  $32.4^\circ C$ . The maximum error in estimation is with methane, where the error is  $85.3^\circ C$ . The curve of experimental  $T_b$  versus  $N_C$  has a negative second derivative and makes a downward arch, and is not well represented by a straight line, especially at the two ends and at the middle.

Polynomial regression is frequently used, and for a set of data that is slightly curved, a small quadratic term could correct for the small deviation from a linear equation. However, polynomials are not necessarily monotonic, and a quadratic function can have a maximum or minimum when the coefficients  $C_0$  and  $C_1$  do not have the same sign and  $dy/dx = 0$ . Let us consider a quadratic polynomial regression for the boiling points of the normal paraffin. The regression shows

$$T'_b = -176.5 + 45.1x - 0.98x^2 \quad \text{sd}(y - y') = 10.4, R^2 = 0.9956$$

The value of  $R^2 = 0.9956$  has improved over the linear value of 0.9546, the standard deviation of  $10.4^\circ C$  has also improved over the linear value of  $32.4^\circ C$ , and the maximum error for methane has been reduced to  $29.2^\circ C$ , but it may still be not good



**Figure 5.3** Boiling points of normal paraffins with quadratic, cubic, and power regressions

enough. Furthermore, the quadratic equation is very treacherous when you try to use it for extrapolation, as it reaches a maximum at  $x = 23$  and begins to decline at even higher values of  $x$ ; see figure 5.3. When we go to a cubic equation, the regression shows

$$T'_b = -207.1 + 60.7x - 2.789x^2 + 0.0573x^3 \quad \text{sd}(y - y') = 4.98, R^2 = 0.9990$$

The improvement in  $R^2$  from 0.9956 to 0.9990 and the improvement of the standard deviation of  $y$  from 10.5 to 5.0 are both significant, and the maximum error for methane is now 12.6 °C. The standard deviation is often a more critical test of the goodness of fit than  $R^2$ . The cubic equation is even more treacherous when it is used for extrapolation, as the second derivative  $d^2y/dx^2$  changes sign from negative to positive at  $x = 16$ , and the value of  $y$  accelerates upwards after that. Figure 5.3 shows the consequences of trying to extrapolate the quadratic and cubic correlations to higher atomic numbers up to 40. We find that the polynomial quadratic and cubic equations behave wildly outside of the training set range. A polynomial correlation should not be extrapolated to regions beyond the training set.

In fact, it is better to use either the power function or the logarithm function, both being convex functions. When the boiling points of the 20 normal paraffins are regressed against a power function of the form  $y = ax^b$ , the regression is simplified by taking the logarithm on both sides to obtain  $\ln(y) = \ln(a) + b \ln(x)$ . This is an easy linear regression, and the results are

$$T'_b = 123.28x^{0.551} \quad \text{sd}(y - y') = 11.6, R^2 = 0.994$$

This is a very good result that is comparable in accuracy to the quadratic equation, while using only two fitting coefficients; another advantage is superiority in extrapolation, as it does not have the unfortunate habit of polynomials in going through maxima and minima. See figure 5.3. The exponential function is totally unsatisfactory, as it is convex and curves up instead of concave and curving down.

Let us examine the utility of such a regression for extrapolation from sample  $S$ , where data are available, to a larger population  $P$ , where data are not available. We can split the population  $P$  into two subsets: a training set  $T$  and a validation set  $V$ . Since greater accuracy can be expected in interpolation than in extrapolation, the training set should be designed to spread over a wide area instead of being clustered together in one neighborhood only.

Let us consider the dispersed training set  $T$  to be normal paraffins with 5, 10, 15, and 20 carbon atoms only. The regression with four data points and three parameters results in

$$T'_b = -129.8 + 36.60x - 0.648x^2 \quad \text{sd}(y - y') = 3.92, R^2 = 0.9997$$

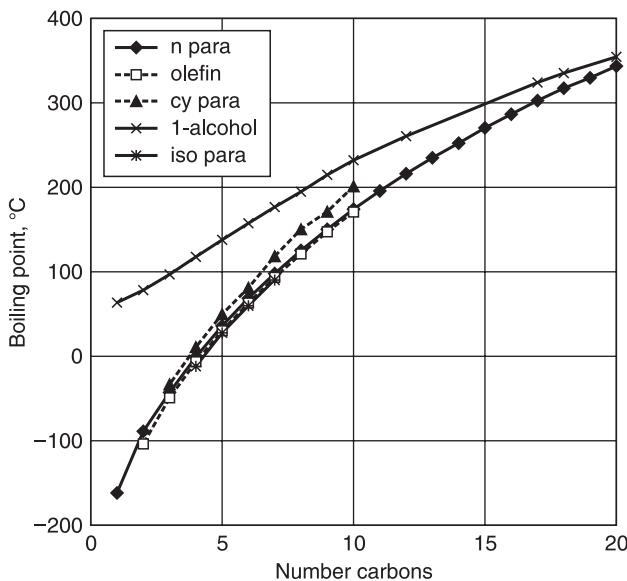
When this correlation is used to predict all 20 normal paraffins, the resulting  $\text{SD}(y - y')$  has grown to be  $16.5\text{ }^\circ\text{C}$ . The biggest error is in methane, which is in error by  $68\text{ }^\circ\text{C}$ .

What degree of success would we encounter when we extrapolate this equation trained in normal paraffins to other families of substances, such as the iso-paraffins, the cyclo-paraffins, the olefins, and even the 1-alcohols? The iso-paraffins and the olefins have lower boiling points than the normal paraffins with the same number of carbon atoms, the cyclo-paraffins have higher boiling points, and the 1-alcohols have much higher boiling points. The shapes of these curves are very similar, which can be seen in figure 5.4, so that the predictors and functional forms can be exported with slight adjustments of the parameter values.

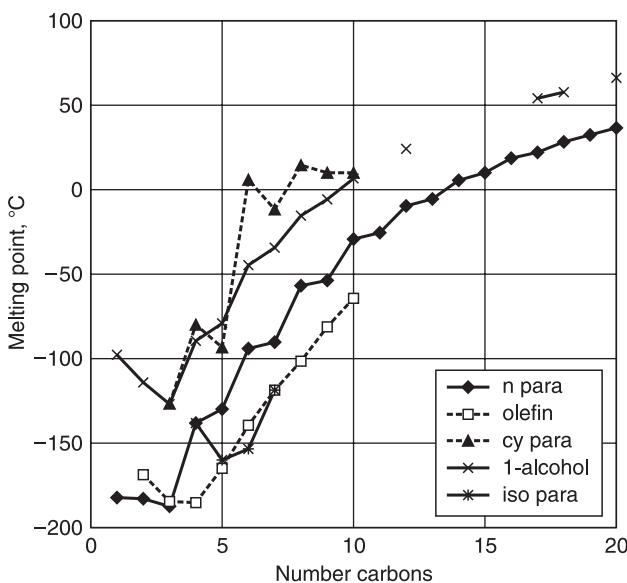
The melting points do not increase smoothly and show reversals and odd-even carbon number banding. Table 5.8 shows that there is a  $-0.3\text{ }^\circ\text{C}$  decrease in the melting point from methane to ethane, followed by a further decrease of  $-4.9\text{ }^\circ\text{C}$  from ethane to propane. As we continue to add  $-\text{CH}_2-$  groups, the increments on melting point gradually become positive, and then fluctuate wildly between  $5$  and  $50\text{ }^\circ\text{C}$ . There is a larger increment of melting points from odd to even carbon numbers, and a smaller increment from even to odd carbon numbers. The melting points are influenced by the molecular symmetry and by the crystalline structures, and these are the hidden variables that will be treated later. Figure 5.5 shows that the correlations of melting points for these series generally have the same shape as that of normal paraffins, as well as the wild gyrations, so that extrapolations would be much less accurate.

### Vapor Pressure of Water

Let us consider the vapor pressure of water as a continuous function of temperature for the range of  $0$  to  $100\text{ }^\circ\text{C}$ . These data are also monotonic increasing and convex. Let us use as training set the data from  $20$  to  $80\text{ }^\circ\text{C}$  at  $10\text{ }^\circ\text{C}$  intervals, which is a set of seven points. Let us propose a quadratic equation, and the regression result of



**Figure 5.4** Boiling points of the series of normal paraffins, iso-paraffins, cyclo-paraffins, olefins and 1-alcohols



**Figure 5.5** Melting points of the series in figure 5.4

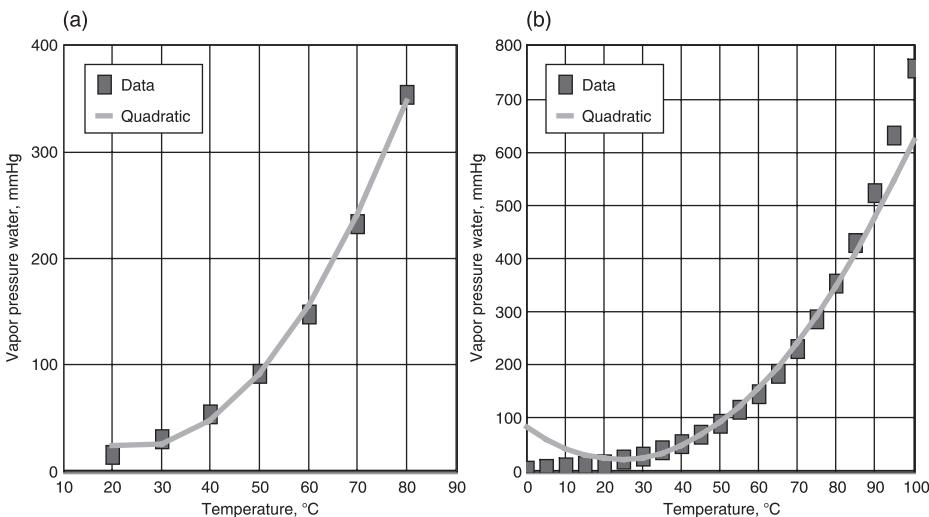


Figure 5.6 Vapor pressure of water versus temperature; (a) training set; (b) full range

vapor pressure in mmHg is

$$V_p(T) = 9273.1 - 62.2T + 0.1046T^2 \quad \text{sd}(P - P') = 8.5, R^2 = 0.9968$$

The degrees of freedom are  $7 - 3 = 4$ . The standard deviation of 8.5 mmHg is very acceptable at a temperature of 80 °C when the vapor pressure is 355.1 mmHg, but it is very poor at a temperature of 20 °C when the vapor pressure is 17.5 mmHg. When this formula trained in the range of 20 to 80 °C is extrapolated to the 0 to 100 °C range, we find that it breaks down, as shown in figure 5.6. The standard error of prediction for the full range is now increased to 43.8 mmHg. It predicts that there is a minimum vapor pressure at 25 °C; it overpredicts at lower temperatures, such as a value of 83 mmHg at 0 °C instead of the experimental value of 4.6 mmHg; and it underpredicts at higher temperatures, such as a value of 622 mmHg at 100 °C instead of the experimental 760 mmHg. The exponential functions do a comparable job with only two parameters:

$$V_p(T) = 6.26 \exp(0.0506T) \quad \text{sd}(P - P') = 34.2, R^2 = 0.9908$$

plus this is monotonic and does not behave wildly outside of the training set range.

### Frequently Used Nonlinear Functions

Let us consider some frequently used nonlinear functions. The Langmuir isotherm and the Michaelis–Menten kinetics are of the form

$$y = \frac{ax}{b + x}$$

which may appear at first to be difficult to do a regression analysis with. This function behaves as a straight line at very small values of  $x$  with a slope of  $a/b$ , and reaches the asymptotic value of  $y = a$  when  $x$  is very large. However, if we take the reciprocal of both sides of the equation we end up with

$$\frac{1}{y} = \frac{b+x}{ax} = \frac{1}{a} + \frac{b}{ax}$$

which is in the form of a linear equation for  $1/y$  with an intercept of  $1/a$  and a slope of  $b/a$ . This method is also called the Burk–Lineweaver plot.

The Clausius–Clapeyron equation relates the vapor pressure of a liquid in the following form:

$$\begin{aligned}\ln\left(\frac{p}{p_0}\right) &= -\frac{\Delta H_v}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right) \\ \ln(P) &= \left(\frac{\Delta H_v}{R}\frac{1}{T_0} + \ln p_0\right) - \left(\frac{\Delta H_v}{R}\right)\frac{1}{T}\end{aligned}$$

This is of the simple linear form of  $y = C_0 + C_1x$  with  $y = \ln(P)$  and  $x = 1/T$ . Since the heat of vaporization is itself a function of  $T$ , it is inconvenient that the parameters  $C_0$  and  $C_1$  are independent of temperature.

The three-parameter Antoine equation has found more frequent use in correlations:

$$\ln(P) = A - B/(T + C)$$

In this equation, the parameter  $B$  is analogous to  $\Delta H_v/R$ , and the parameter  $C$  gives flexibility to deal with nonideal behavior. An equation with even more terms is the one adopted in Yaws:

$$\ln(P) = A + B/T + C \log T + DT + ET^2$$

Next, consider the heat of vaporization, which is often given in the form

$$\Delta H_v = A \left(1 - \frac{T}{T_{\text{cr}}}\right)^n$$

Here,  $T_{\text{cr}}$  is the critical temperature and  $n$  is an exponent that can have the value of 0.25 to 0.6. The heat of vaporization vanishes when  $T = T_{\text{cr}}$ . Surface tension has a function of the same form, except that  $n$  varies from 1.0 to 1.5.

Then consider the density of a liquid, which is sometimes correlated in the form

$$\rho = AB^{(1-T/T_{\text{c}})^n}$$

This time  $n$  may vary from 0.13 to 0.30. Finally, solubility in water is sometimes given in the following terms:

$$\log S = A + \frac{B}{T} + \frac{C}{T^2}$$

where  $B$  is usually a negative number.

### 5.3 Regression with Several Predictors

There are many occasions when there are more than one cause behind a property, so that two or more predictors would be required to explain the phenomenon adequately.

#### Multilinear Regressions

The simplest function with several predictors is a multi-linear function of the form

$$y = c_0 + c_1x_1 + c_2x_2 + \cdots + c_kx_k$$

The property can be dominated by only one predictor, while the other predictors provide minor corrections; there are also times when several predictors have comparable influences.

#### *Boiling Points of the Hydrochlorofluorocarbons*

Consider the melting and boiling points of the 15 members of the family of one-carbon HCFCs, shown in table 5.9 and figure 5.7. The boiling points of the series  $\text{CCl}_4$ ,  $\text{CCl}_3\text{F}$ ,  $\text{CCl}_2\text{F}_2$ ,  $\text{CClF}_3$ , and  $\text{CF}_4$  behave approximately in a linear manner. The series from  $\text{CH}_4$  to  $\text{CCl}_4$  begins with a large difference between  $\text{CH}_4$  and  $\text{CH}_3\text{Cl}$ , and ends with a small difference between  $\text{CHCl}_3$  and  $\text{CCl}_4$ . The series from  $\text{CH}_4$  to  $\text{CF}_4$  has a maximum in  $\text{CH}_2\text{F}_2$ . When we do a single-parameter regression for the entire population with  $N_{\text{Cl}}$ , we obtain

$$T'_b = 179.8 + 46.3N_{\text{Cl}} \quad \text{sd}(T_b - T'_b) = 33.4, R^2 = 0.7757$$

This is shown in figure 5.8. When we consider  $N_{\text{F}}$  or  $\mu$  as the single predictor, the regression results are

$$T'_b = 279.5 - 28.4N_{\text{F}} \quad \text{sd}(T_b - T'_b) = 59.3, R^2 = 0.2922$$

$$T'_b = 225.1 + 15.5\mu \quad \text{sd}(T_b - T'_b) = 69.4, R^2 = 0.0288$$

Thus, if only one predictor is to be used, the number of chlorine atoms  $N_{\text{Cl}}$  is the best single predictor and the dipole moment is the worst single predictor. If we are not satisfied with a standard error ( $y - y'$ ) of  $33.4^{\circ}\text{C}$ , then we can escalate to

Table 5.9 Properties of the 15 CFCs with one carbon atom

Substance	B.p., K	M.p., K	Dipole moment $\mu$	Symmetry number $\sigma$
CCl <sub>4</sub>	349.9	250.1	0	12
CCl <sub>3</sub> F	296.8	162.0	0.45	3
CCl <sub>2</sub> F <sub>2</sub>	243.3	115.1	0.51	2
CClF <sub>3</sub>	191.7	92.1	0.51	3
CF <sub>4</sub>	145.1	89.6	0	12
CHF <sub>3</sub>	191.0	118.0	1.65	3
CH <sub>2</sub> F <sub>2</sub>	221.5	137.1	1.96	2
CH <sub>3</sub> F	194.7	131.3	1.85	3
CH <sub>4</sub>	111.6	90.7	0	12
CH <sub>3</sub> Cl	249.1	175.4	1.87	3
CH <sub>2</sub> Cl <sub>2</sub>	313.1	178.0	1.60	2
CHCl <sub>3</sub>	334.2	209.5	1.01	3
CH <sub>2</sub> ClF	264.0	140.1	1.82	1
CHClF <sub>2</sub>	232.4	115.4	1.42	1
CHCl <sub>2</sub> F	282.0	138.1	1.29	1
SD(y), SD(x)	65.6	44.0	0.72	4.0

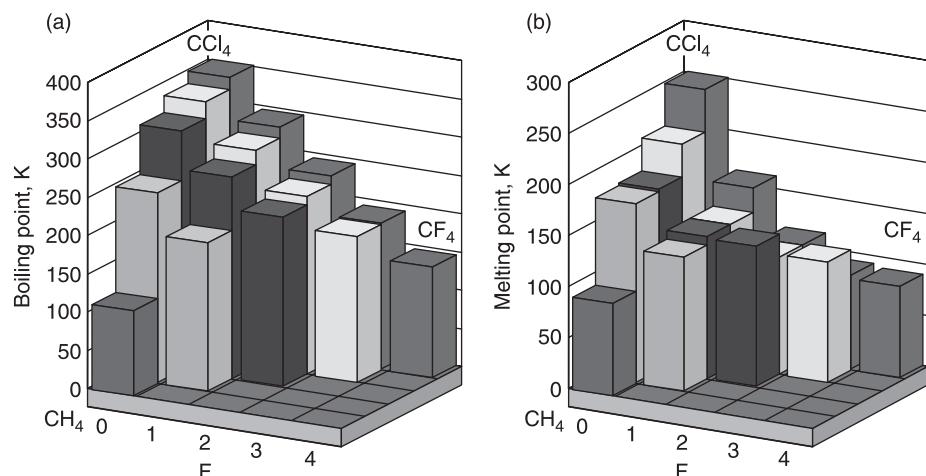


Figure 5.7 (a) Boiling and (b) melting points of one-carbon CFCs

a two-parameter regression. One may expect that the best combination would be using the top two single predictors, but the results show that the best result comes from using the number of chlorine atoms and the dipole moment, which results in a standard error of 11.5 °C. The reason lies with the five molecules from CH<sub>4</sub> to CF<sub>4</sub> without chlorine atoms, which cannot be resolved by the addition of the number of fluorine atoms, but can be resolved by their dipole moments. In fact, when we move

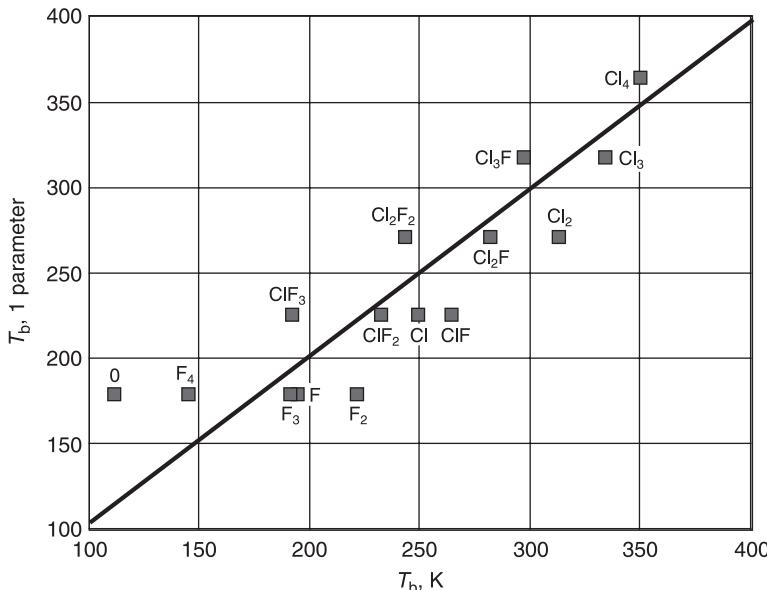


Figure 5.8 Boiling points of one-carbon CFCs; correlation with one parameter:  $N_{\text{Cl}}$

to a three-parameter regression, the gain in standard error to  $11.4^{\circ}\text{C}$  is very little. The results of this correlation are shown in figure 5.9. It is seen in the three-predictor correlation, the number of chlorine atoms and the dipole moment, are far more influential, but the substitution of fluorine atoms for hydrogen atoms has only a small influence (see table 5.10).

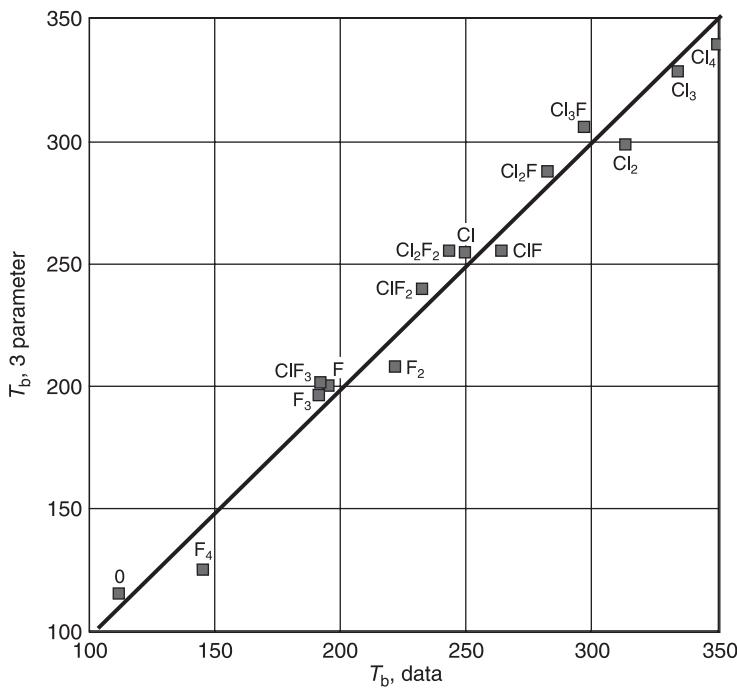
In the three-parameter correlation, the influence of each predictor on the property can be measured by multiplying its coefficient by its standard deviation (see table 5.11).

So the boiling points of this data set can be attributed much more to the number of chlorine atoms than to the dipole moment, with very little influence from the number of fluorine atoms. We are now using three parameters to explain 15 data points, for  $15 - 3 = 12$  degrees of freedom, and we have a standard error of  $11.4^{\circ}\text{C}$ . If we want a standard deviation of less than  $1^{\circ}\text{C}$ , then this is not good enough, and we would have to dig deeper for more hidden variables.

Let us try to extrapolate this correlation to the available experimental data on some of the two-carbon HCFCs, based on ethane instead of methane, which is shown in table 5.12. There are new considerations, as tetrafluoroethane can have two different forms: one with three fluorine atoms on one carbon and one fluorine atom on the other carbon, and the other with two fluorine atoms on each carbon. Nevertheless, the boiling points of the 19 substances satisfy the correlation formula of

$$T'_b = 184.4 + 42.8N_{\text{Cl}} + 22.2\mu - 2.6N_{\text{F}} \quad R^2 = 0.980$$

This is shown in figure 5.10. The value of the constant term is increased from 116 to 184 in recognition of the increase of carbon atoms from one to two, and the rest of



**Figure 5.9** Boiling points of one-carbon CFCs; correlation with three parameters:  $N_{\text{Cl}}$ ,  $\mu$ , and  $N_{\text{F}}$

**Table 5.10 Effect of number of predictors**

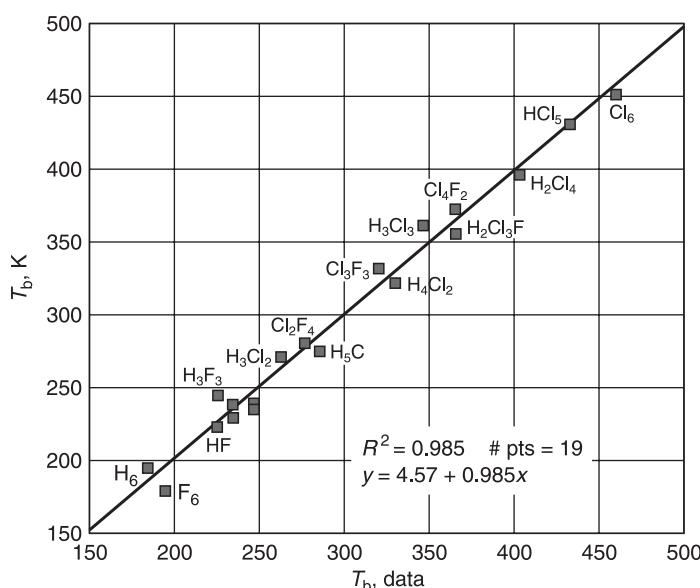
Predictors	$C_0$	$C_{\text{Cl}}$	$C_F$	$C_\mu$	$R^2$	$sd(T_b - T'_b)$
One	179.8	46.3			0.7757	33.4
	279.5		-28.4		0.2922	59.3
	225.1			15.5	0.0288	69.4
Two	193.9	42.8	-7.0		0.7891	33.7
	124.8	53.5		42.6	0.9754	11.5
	269.8		-27.7	8.2	0.3001	61.3
Three	116.3	55.3	3.1	44.4	0.9777	11.4

**Table 5.11 Influence of predictor for three-parameter correlation**

Predictor	$C_i$	$sd(x_i)$	$C_i sd(x_i)$
$N_{\text{Cl}}$	55.3	1.247	69.7
$\mu$	44.3	0.723	32.2
$N_{\text{F}}$	3.1	1.247	3.0

**Table 5.12 Properties of CFCs with two carbon atoms**

Molecule	Dipole	$T_m, K$	$T_b, K$
$C_2F_6$	0	172.4	194.9
$C_2ClF_5$	0.64	173.7	234.0
$C_2Cl_2F_4$	0.53	179.2	276.9
$C_2Cl_3F_3$	0.77	238.2	320.8
$C_2Cl_4F_2$	0.54	299.2	366.0
$C_2Cl_6$	0	459.9	460.0
$C_2HF_5$	1.89	170.2	225.2
$C_2HCl_5$	1.01	244.2	433.0
$C_2H_2F_4$	2.32	172.2	247.2
$C_2H_2FCl_3$	1.59	180.0	366.0
$C_2H_2Cl_4$	1.44	202.9	403.6
$C_2H_3F_3$	2.59	161.9	225.8
$C_2H_3ClF_2$	1.74	142.4	263.2
$C_2H_3Cl_3$	1.74	242.8	347.2
$C_2H_4F_2$	2.27	156.2	247.4
$C_2H_4Cl_2$	1.86	176.2	330.4
$C_2H_5F_2$	1.69	129.9	235.5
$C_2H_5Cl$	1.69	136.8	285.4
$C_2H_6$	0	90.4	184.6



**Figure 5.10** Boiling points of two-carbon CFCs; correlations with three parameters

the coefficients are roughly of the same order of magnitude and the same ranking in relative importance. This result encourages us to extrapolate this result further to the HCFCs based on propane.

### *Melting Points of Hydrochlorofluorocarbons*

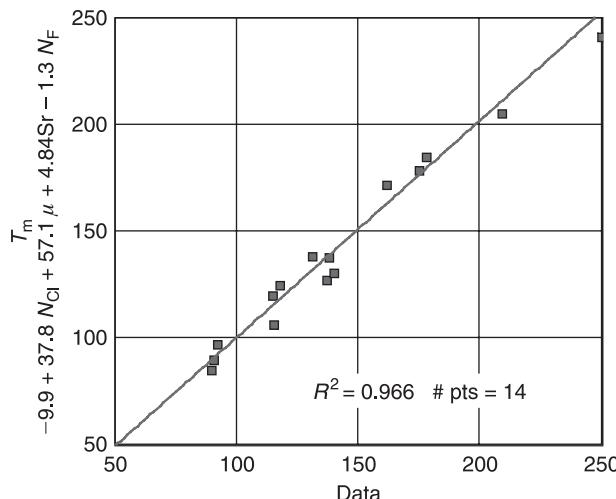
Figure 5.7 shows that the behavior of the melting points bears some resemblance to the boiling points, as well as having many differences. The three parameters used for boiling points are not sufficient for the melting points of the 15 CFCs, as the regression results is

$$T'_m = 104.9 + 25.6N_{Cl} - 10.1N_F + 16.3\mu \quad sd = 24.1, R^2 = 0.779$$

The phenomenon of melting point also involves the entropy of melting, which has a much smaller value for a symmetrical molecule (such as methane, with a symmetry number of 12) than with an asymmetrical molecule (such as  $CH_2ClF$ , with a symmetry number of 1). When we introduce the fourth parameter of the entropy of rotational symmetry given by  $R \ln(\sigma)$ , the regression results is

$$T'_m = -9.9 + 37.9N_{Cl} + 57.1\mu + 40.3\ln(\sigma) - 1.2N_F \quad sd = 8.4, R^2 = 0.976$$

The results are also shown in figure 5.11. Symmetry ranks just behind the number of chlorine atoms and the dipole moment in importance, but the number of fluorine



**Figure 5.11** Melting points of one-carbon CFC; correlation with four parameters:  $N_{Cl}$ ,  $\mu$ ,  $N_F$ , and  $\sigma$

atoms remains unimportant. In fact, when we delete  $N_F$  in favor of a three-parameter regression, we obtain

$$T'_m = -15.6 + 38.8N_{\text{Cl}} + 58.7\mu + 41.2\ln\sigma \quad \text{sd} = 8.1, R^2 = 0.975$$

The melting points in the data set can be explained as equally due to the influence of  $N_{\text{Cl}}$ , the dipole moment, and the molecular symmetry, but with very little influence from  $N_F$ .

### Multipredictor Nonlinear Correlations

When there is a nonlinear function involving more than one predictor, in general there is no simple method for extracting parameter values from a given set of data. In some functions, such as

$$A = \alpha B^\beta C^\gamma$$

$$\ln A = \ln \alpha + \beta \ln B + \gamma \ln C$$

it is possible by taking logarithms on both sides to reformulate the equation into a standard multilinear form, to separate and extract the parameter values of  $\alpha$ ,  $\beta$ , and  $\gamma$ .

For a gas that follows the van der Waals equation of state, the extraction of the values of the parameters  $a$  and  $b$  from a set of experimental data is facilitated by reformulating the equation to obtain

$$P = \left( \frac{R}{V - b} \right) T - \frac{a}{V^2}$$

One has to design the experiment to take a set of data designed to facilitate the task of parameter extraction. If a set of data is taken under constant volume conditions, and the pressure is plotted against the temperature, then there will be an intercept of  $-a/V^2$  and a slope of  $R/(V - b)$ . The van der Waals equation of state is the simplest of the equations of state beyond the perfect gas law, and the task of extracting parameter values from experimental data for the more complicated equations of state would require more ingenuity. The Redlich-Kwong equation has two parameters,  $A$  and  $B$ :

$$P = \frac{RT}{V - B} - \frac{A}{\sqrt{T}V(V + B)}$$

The Peng-Robinson equation has two parameters,  $\alpha$  and  $\beta$ :

$$P = \frac{RT}{V - \beta} - \frac{\alpha}{V(V + \beta) + \beta(V - \beta)}$$

And the Benedict–Webb–Rubin equation has many more parameters:

$$P = \frac{RT}{V} + \frac{1}{V^2} \left( RTB_0 - A_0 - \frac{C_0}{T^2} \right) + \frac{1}{V^3} (RT_b - a) \\ + \frac{\alpha a}{V^6} + \frac{c}{T^2 V^3} \left( 1 + \frac{\gamma}{V^2} \right) \exp \left( -\frac{\gamma}{V^2} \right)$$

These are semi-empirical equations of state that are formulated to describe experimental data accurately, instead of conforming to theoretical descriptions of molecular behavior, and each parameter does not necessarily have a physical interpretation.

### Quantitative Property–Property Relationships and Quantitative Structure–Property Relationships

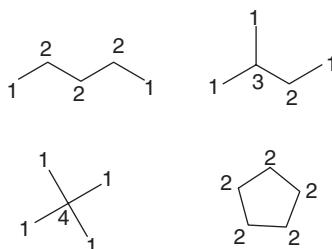
A QPPR is a correlation between a property  $y$  and one or more molecular properties  $\{p_1, p_2, \dots, p_m\}$ , where each property  $p$  is already known:  $y = f(p_1, p_2, \dots, p_m)$ . Examples are estimations of the boiling point from a knowledge of the critical temperature and pressure.

A QSPR is a correlation between a property  $y$  and one or more molecular descriptors  $\{x_1, x_2, \dots, x_m\}$ , where a descriptor can be computed from the structure of the molecule:

$$y = f(x_1, x_2, \dots, x_m)$$

This method is similar to the group contribution method (GCM) when the descriptor is the number of functional groups in a molecule. More often in the literature, the descriptors are based on the connectivity of the atoms in a highly elaborate manner. There are frequently no theoretical justifications to connect a topological index to a property, so they should be extrapolated with caution and vigilance.

An example is the method proposed by Kier and Hall, which usually considers only the C atoms and excludes the H atoms from consideration. Take the four isomers of pentane below, which are connected very differently. Each molecule can be described by a connection number  $\delta$  for each of the carbon atoms, 1 for a primary carbon, 2 for a secondary carbon, etc.



Kier and Hall proposed a series of connectivity indexes. The zeroth index sums the reciprocal of the square roots of the connection numbers of each individual atom.

For the isopentane molecule, the zeroth index of connectivity is given by

$$^0\chi = \sum (\delta_i)^{-0.5} = \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{1}} = 4.284$$

The first index considers each pair of adjacent carbon atoms, and sums the reciprocal of the square roots of their products. There are four sets of adjacent carbon atoms, and their connection numbers are (1,3), (3,2), (2,1), and (3,1). The first index for isopentane is

$$^1\chi = \sum (\delta_i \delta_j)^{-0.5} = \frac{1}{\sqrt{1.3}} + \frac{1}{\sqrt{3.2}} + \frac{1}{\sqrt{2.1}} + \frac{1}{\sqrt{3.1}} = 2.270$$

The second connection index considers three adjacent molecules at a time, and there are four choices: (1,3,2), (3,2,1), (1,3,1), and (1,3,2). The second-order index has the value of

$$^2\chi = \sum (\delta_i \delta_j \delta_k)^{-0.5} = \frac{1}{\sqrt{1.1.3}} + \frac{3}{\sqrt{1.2.3}} = 1.802$$

There are claims that these structural parameters do a very good job in correlating molecular properties, such as molar refraction of light, water solubility, molar volume, and chromatographic retention data. This method has also been applied to complex biological and toxicological properties, including flavor threshold, anesthetics, hallucinogenic potency, monoamine oxidase inhibition, taste, etcetera. The authors made no theoretical justifications for these topological indexes, and it remains to be seen how successful any method without theoretical justification can be in making bold extrapolations far from the sample base and training set.

## 5.4 Group Contribution Methods

In the last few sections, each population or family of compounds requires a separate regression formula, such as noble gases, normal paraffins, HCFCs, etcetera. There is another type of method that attempts to deal with many populations of organic compounds at one stroke. The GCM parses, or divides up, a molecule into a number of functional groups. A group can be a single atom, such as Cl—, or a fragment that often stays together, such as —CH<sub>3</sub>. The GCM assumes that some properties of the molecule can be estimated as the sums of the contributions from each group. In the simplest form, we assume that a particular property is linearly additive:

$$f = f_0 + n_1 f_1 + n_2 f_2 + \cdots + n_m f_m$$

where  $f$  is the property of the entire molecule,  $f_0$  is a constant,  $n_i$  is the number of times that the  $i$ th group appears in the molecule, and  $f_i$  is the contribution of the  $i$ th group to the property.

This method is certainly valid when the property in question is the molecular weight, which is precisely the sum of the contributions from each atom according to the atomic weights and number of occurrences. For some properties, this method can be reasonably effective. But this method does not work very well if the property is not additive, or if the groups have strong interactions in a way that is either synergistic or antagonistic.

### Method of Joback

Joback's method is an improved version of earlier methods, and attempts to estimate many properties of all the organic compounds. Table 5.13 shows the recommended Joback coefficients to build up the boiling and melting points of a compound, according to the formula

$$T_b(\text{K}) = 198 + \sum \Delta T_{b_i}$$

$$T_f(\text{K}) = 122 + \sum \Delta T_{f_i}$$

For example, normal pentane has two end groups of  $\text{CH}_3$ —and three middle groups of  $-\text{CH}_2-$ , so that its boiling point is estimated to be  $198 + (2 \times 23.6) + (3 \times 22.9) = 313.9$  K, which compares with an experimental value of 309.2 K for a

**Table 5.13** Joback coefficients for  $T_b$  and  $T_f$

Group	$\Delta T_b$	$\Delta T_f$	Group	$\Delta T_b$	$\Delta T_f$
$-\text{CH}_3$	23.58	-5.1	$-\text{OH}$ (alcohol)	92.88	44.45
$>\text{CH}_2$	22.88	11.27	$-\text{OH}$ (phenol)	76.34	82.83
$>\text{CH}-$	21.74	12.64	$-\text{O}-$	22.42	22.23
$>\text{C}<$	18.25	46.43	$-\text{O}-(\text{r})$	31.22	23.05
$=\text{CH}_2$	18.18	-4.32	$>\text{C}=\text{O}$	76.75	61.2
$=\text{CH}-$	24.96	8.73	$>\text{C}=\text{O}(\text{r})$	94.97	75.97
$=\text{C}<$	24.14	11.14	$\text{O}=\text{CH}-$ (aldehyde)	72.24	36.9
$=\text{C}=$	26.15	17.78	$-\text{COOH}$ (acid)	169.09	155.5
$\equiv\text{CH}$	9.2	-11.18	$-\text{COO}-$ (ester)	81.1	53.6
$\equiv\text{C}-$	27.38	64.32	$=\text{O}$ (except above)	-10.5	2.08
$-\text{CH}_2-$ (ring)	27.15	7.75			
$>\text{CH}-$ (r)	21.78	19.88	$\text{NH}_2$	73.23	66.89
$>\text{C}<$ (r)	21.32	60.15	$>\text{NH}$	50.17	52.66
$=\text{CH}-$ (r)	26.73	8.13	$>\text{NH}(\text{r})$	52.82	101.51
$=\text{C}<$ (r)	31.01	37.02	$>\text{N}-$	11.74	48.84
			$=\text{N}=$	74.6	0
$-\text{F}$	-0.03	-15.78	$=\text{N}=(\text{r})$	57.55	68.4
$-\text{Cl}$	38.13	13.55	$-\text{CN}$	125.66	59.89
$-\text{Br}$	66.86	43.43	$-\text{NO}_2$	152.54	127.24
$-\text{I}$	93.84	41.69	$-\text{SH}$	63.56	20.09
			$-\text{S}-$	68.78	34.4
			$-\text{S}-(\text{r})$	52.1	79.93

small error of 4.7 K. Neopentane, or 2,2-dimethylpropane, has four end groups of  $\text{CH}_3$ —and one quarternary group of  $>\text{C}<$ , so that its boiling point is estimated to be  $198 + (4 \times 23.6) + (1 \times 18.3) = 310.7$  K, which compares with an experimental value of 282.7 K for a larger error of 28.0 K.

An examination of table 5.13 shows that the effects of substituting the  $-\text{H}$  group with any group will increase the boiling point, except for  $-\text{F}$ . The hydrocarbon groups add electrons, and thus increase the dispersion forces. Since the group contributions  $\Delta T_b$  are ranked in the order



this method predicts that, among isomers with the same number of carbon atoms, the more highly branched isomers will have lower boiling points. The halogen group of elements creates dipole moments. Fluorine is an anomaly, as it forms hydrogen bonds, but has essentially no effect on boiling points, and in fact decreases melting points. The rest of the halogen atoms do not form hydrogen bonds, but do increase the polarity of the molecule. Some of the oxygen- and nitrogen-containing groups have the ability to form hydrogen bonds, such as  $-\text{OH}$  and  $-\text{NH}_2$ , and have the largest effect on raising the boiling and melting points. The oxygen and nitrogen compounds that do not form hydrogen bonds, such as  $-\text{O}-$  and  $>\text{N}-$ , have much smaller  $\Delta T_b$ .

The melting points do not behave in parallel with the boiling points, and the group contributions  $\Delta T_f$  are ranked in the reverse order:



So this method predicts that the more highly branched isomers will have higher melting points, despite their having lower boiling points. This can be interpreted as a manifestation of the generally higher symmetry of the  $>\text{C}<$  molecules.

GCMs can also be nonlinear; for instance, Joback recommended the following for the critical parameters  $T_c$ ,  $P_c$  and  $V_c$ :

$$T_c = \frac{T_b}{0.584 + 0.965 \sum \Delta T_i - (\sum \Delta T_i)^2}$$

$$P_c = \frac{1}{(0.113 + 0.0032n_A - \sum \Delta P_i)^2}$$

$$V_c = 17.5 + \sum \Delta V_i$$

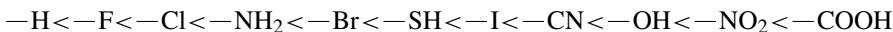
where  $n_A$  is the number of atoms in the molecule and  $\Delta T_i$ ,  $\Delta P_i$ , and  $\Delta V_i$  are the group contribution parameters for each group towards temperature, pressure, and volume. These second-order terms make it possible to account partially for adjustments due to interactions between groups.

The GCM is a quick and ready general method to make estimations in a larger population where there is little data, and thus cannot be expected to be as highly accurate as a special correlation tuned to a particular smaller population. It is also very

useful as an estimator of the differential effects of substituting one group by another. Caution in using this method is warranted under the following conditions:

1. There is no unique method to parse, or to take apart a molecule into component parts; this may lead to different estimates for the same molecule.
2. The method takes no notice of the substitution positions, so that 1-chloropropane and 2-chloropropane are predicted to have the same boiling points.
3. The repeated addition of a group often leads to diminished return. The boiling temperature increment in adding a single chlorine to propane is 78.6 °C, but the second chlorine adds only 32.8 °C more. In the boiling points of the series of normal paraffins, the addition of a —CH<sub>2</sub>— group to methane brings an increase of 73.5 °C to the boiling point of ethane and the second group brings only 47 °C, with the effects of each additional group gradually diminishing to 14 °C at *n*-eicosane.
4. Two strong groups that are adjacent may interact in a synergistic or antagonistic manner, leading to a strong proximity effect. For instance, there are four dichloropropanes, where the proximity of the two chlorine atoms leads to lower boiling points. The highest boiling point belongs to 1,3-dichloropropane, where the two chlorine atoms are the farthest apart; the next highest belongs to 1,2-dichloropropane; and the lowest boiling points belong to 2,2-dichloropropane, where the two chlorine atoms are on the same carbon atom (see table 5.14). However, 2,2-dichloropropane has the unique distinction of possessing both the lowest boiling point and the highest melting point of the four dichloropropanes, due to the fact that it has the highest symmetry.

Figure 5.12 shows the effects of the substitution of 10 different function groups on the boiling points and melting points of methane; the plot is ranked by the boiling points from the lowest for H— to the highest for —COOH.



The melting points do not follow the same ranking, as the —SH and the —OH substitutes have abnormally low melting points. When the same comparison is made for substitutions on ethane, the boiling points have the same ranking as on methane; but for the melting points, it is the —NH<sub>2</sub> substitute that has an abnormally high melting point.

**Table 5.14 Properties of the chloropropanes**

	B.p., °C	M.p., °C
Chloropropane	−42.1	−187.7
2-Chloropropane	36.5	−117
1-Chloropropane	46.4	−122.8
2,2-Dichloropropane	69.3	−33.8
1,1-Dichloropropane	88.1	
1,2-Dichloropropane	96.4	−100.4
1,3-Dichloropropane	120.9	−99.5

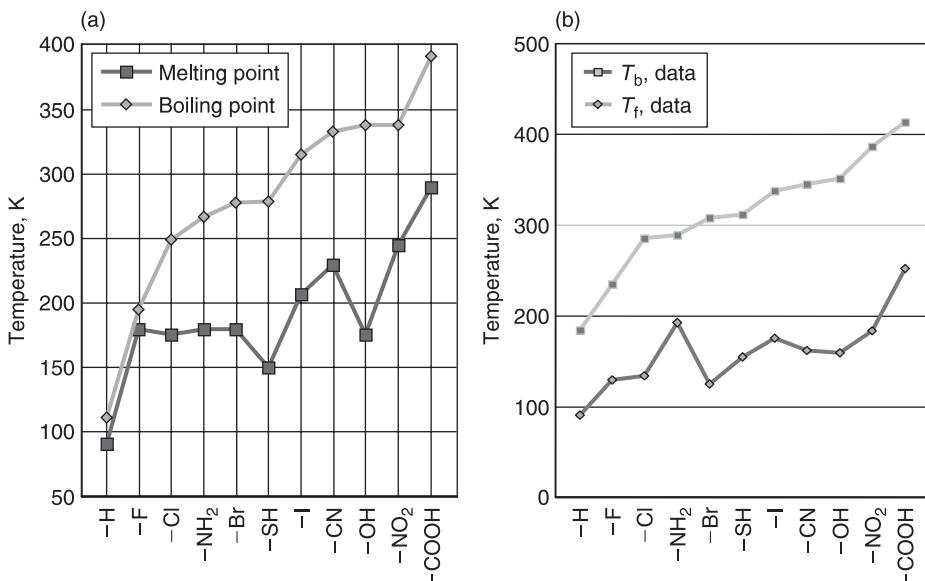


Figure 5.12 Melting and boiling points of compounds with substitution groups: (a) methyl compounds; (b) ethyl compounds

Figure 5.13 shows the substitution effects on benzene; once again, the boiling points are easy to predict and the melting points are difficult to predict.

The Joback method works quite well in methane substitutes for boiling points, except for the  $-H$ , the  $-F$  and  $-NO_2$  molecules; see figure 5.14. However, the melting points are much less satisfactory, as they depend not only on the properties of the groups added, but also on the architecture and the shapes of the molecules.

### Method of Hansch and Leo

Rekker introduced a method in 1977 to estimate the partition coefficient  $\log P$ , or  $\log K_{ow}$ , of a solute between 1-octanol and water, which is important in biological activities, environmental fates of chemicals, and drug design. A large value of  $\log P$  indicates that the compound prefers oil to water, and thus accumulates in fat and oil. This method also parses a molecule into a number of “fragments,” and each fragment is given a coefficient; the predicted value is given as a linear sum:

$$\log P = \sum a_n f_n + \sum b_m F_m$$

where  $a$  is the number of occurrences of fragment  $n$  with coefficient  $f$  and  $b$  is the occurrences of correction factors  $F$ . The first formulation of this method has only the first summation, which did not work well enough because there are numerous deviations. In the second formulation, a number of correction factors, such as for the ortho/para substitution position on benzene and for the multiple substitutions of  $-OH$ ,

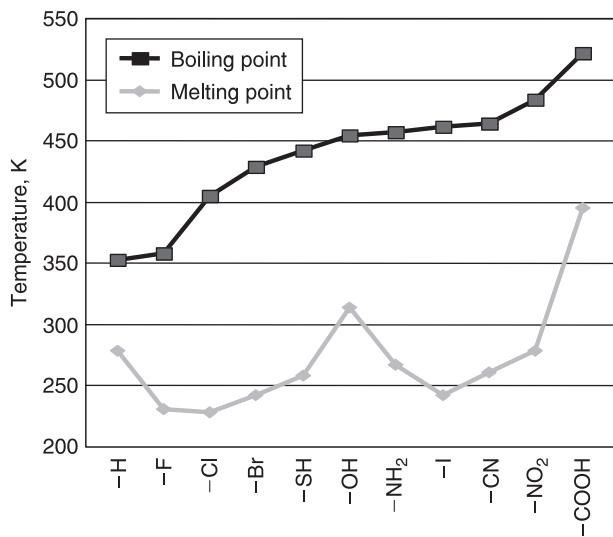


Figure 5.13 Melting and boiling points of phenyl compounds with substitution groups

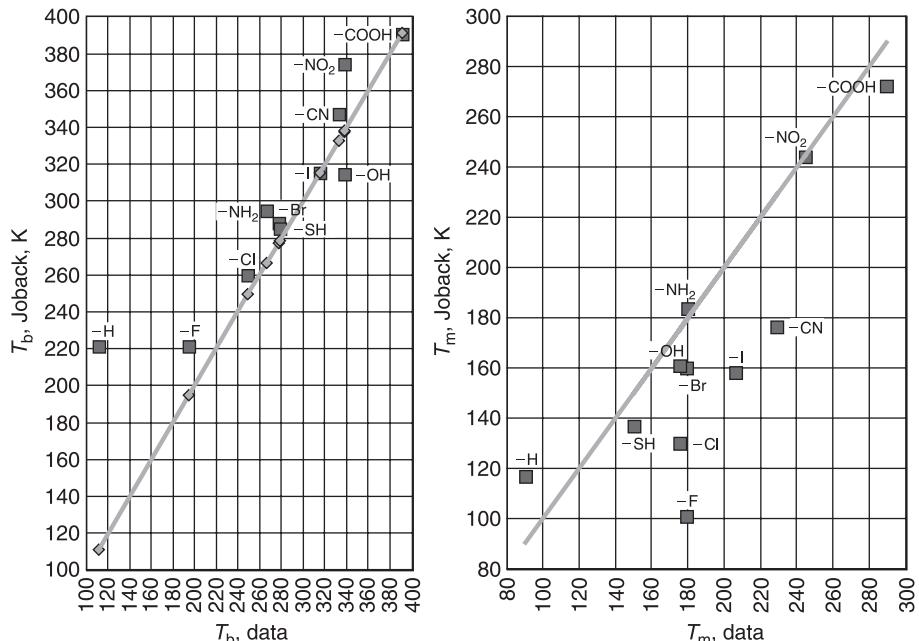


Figure 5.14 Comparing the experimental data and the correlations for melting and boiling points of the methyl compounds in figure 5.12

**Table 5.15** Hansch structural groups and group contributions for estimating octanol-water partition coefficients

	<i>Aliphatic</i>	<i>Aromatic</i>
—N<	−1.8323	−0.5262
—NH—	−1.4962	
—NH <sub>2</sub>	−1.4148	−0.9170
—OH	−1.4086	−0.4800
—O—	−1.2566	−0.0423
—C(O)H	−0.9422	−0.2830
—C≡N	−0.9218	−0.4530
—NO <sub>2</sub>	−0.8132	−0.1823
—C(O)OH	−0.6895	−0.1186
—S—	−0.4045	0.4082
—F	−0.0031	0.2004
—Cl	0.3102	0.6445
—Br	0.3997	0.8900
—I	0.8146	1.1670
Constant	0.2290	
≡C	0.1334	
>C<	0.2676	
=C—arom		0.2940
>CH—	0.3614	
=C<	0.3836	
olefin		
—CH <sub>2</sub> —	0.4911	
—CH <sub>3</sub>	0.5473	

were introduced to give better correlations. The presence of hydrophilic groups with polar and hydrogen bond potentials, such as —NH<sub>2</sub> and —OH, would lower the value of  $\log P$ ; and groups with a hydrophobic or oleophilic nonpolar hydrocarbon group would increase the value of  $\log P$ , such as —C<sub>6</sub>H<sub>5</sub> and —C<sub>6</sub>H<sub>11</sub>.

The approach of Hansch and Leo (1995) uses a small number of fragment values derived from very accurate partition coefficient measurements of a relatively small number of compounds, and requires a large number of correction factors. Some of the coefficient values are given in table 5.15. This approach has been designed for an automatic computer program that will do all the coefficients and corrections, called the CLOG. Hansch and Leo reported that for 7500 compounds tested, the correlation has a standard error of 0.336 and an  $R^2 = 0.978$ .

## 5.5 Mixtures and Composite Materials

It is often possible to predict with accuracy many properties of ideal solutions, such as dilute gas mixtures, as well as liquid mixtures of closely related substances such as pentane and hexane. On the other hand, liquid mixtures of substances with different

structures are usually far from ideal, and require specialized knowledge and experience. We will consider only the simplest cases here.

### Fluid Mixtures

When two chemicals are mixed, an ideal solution is defined when the properties of the mixture are a linear function of the properties of the two chemicals. The simplest relation is in the form of linear relations:

$$\langle Q \rangle = \sum x_i Q_i$$

where  $x_i$  is the weight or mole fraction of  $i$  in the mixture, and  $\sum x_i = 1$ ,  $Q_i$  is the property of the molecule  $i$ , and  $\langle Q \rangle$  is the overall property of the mixture.

This is certainly true of the simple properties, such as the molecular weight of a mixture. For the van der Waals equation of state, the parameter  $b$  stands for the excluded volume due to the molecule, which suggests that the linear additive relation, or the arithmetic mean, may be appropriate for an ideal mixture:

$$\langle b \rangle = \sum x_i b_i$$

However, parameter  $a$ , which stands for the interaction energy between two molecules, is proportional to the product of their polarizabilities and to the product of their dipole moments, and would not be linearly additive.

When there is only a slight deviation from the linear, it may be appropriate to use a quadratic equation to fit the data. An interaction energy  $Q_{ij}$  can be introduced for each pair of molecules  $i$  and  $j$ , and the interaction energy so generated to be  $x_i x_j Q_{ij}$ . The total quadratic contributions would sum to

$$\langle Q \rangle = \sum \sum x_i x_j Q_{ij} = x_1^2 Q_1 + 2x_1 x_2 Q_{12} + x_2^2 Q_2$$

There are two special cases that are worth special mention. For the case that the interaction term is the arithmetic mean, then

$$Q_{ij} = \frac{Q_i + Q_j}{2}$$

$$\langle Q \rangle = \sum \sum x_i x_j \frac{Q_i + Q_j}{2} = \sum x_i Q_i$$

The double summation simplifies into a single summation that is identical to the linear case. For the special case when the interaction term is the geometric mean, then

$$Q_{ij} = \sqrt{Q_i Q_j}$$

$$\langle Q \rangle = \sum \sum x_i x_j \sqrt{Q_i Q_j} = \left( \sum x_i \sqrt{Q_i} \right)^2 < \sum x_i Q_i$$

For a binary system, this geometric mean is always smaller than the arithmetic mean. It may be expected that for substances that are very similar, such as *n*-butane and *n*-pentane, the arithmetic mean would work for *b* and the geometric mean would work for *a*. For substances that are not very similar, it would be necessary to introduce adjustment parameters to be determined by regression, such as

$$Q_{ij} = \frac{k_{ij}(Q_i + Q_j)}{2}$$

$$Q_{ij} = k'_{ij}\sqrt{Q_i Q_j}$$

When the value of the adjustment parameters  $k_{ij} > 1$ , the two components are said to be synergistic, so that the resulting property is higher than the arithmetic average. When the value of  $k_{ij} < 1$ , the two components are said to be antagonistic, so that the resulting property is lower than the arithmetic average. For similar substances, such as propylene–propane and butane–pentane, the value of  $k_{ij} \approx 1$ . For estimation of the critical temperature, the values of  $k_{ij}$  are significantly smaller for dissimilar substances, such as HCl–propane and H<sub>2</sub>S–ethane. The values can also be significantly larger than one when there is a large size disparity, such as with methane–naphthalene and nitrogen–octane.

Figure 5.15 shows a number of mixing results including the arithmetic and the geometric averages, as well as the results with adjustment parameters for  $k = 0.1, 0.5, 1.5$  to 2. When  $k = 2$ , note that the strong synergistic mixture property can be higher

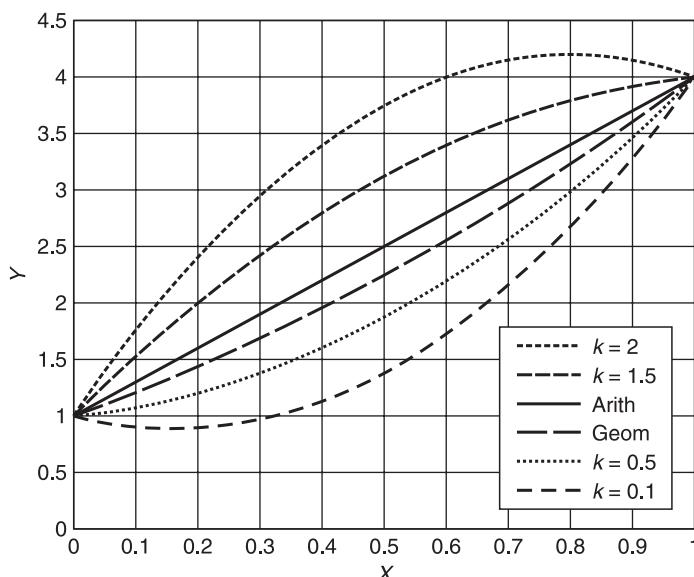


Figure 5.15 Mixture properties modeled by quadratic interaction parameters *k*

than either component; and when  $k = 0.1$ , the antagonistic mixture property can be lower than either component.

Another type of interaction is given by the form

$$\langle Q \rangle = \left( \sum x_i Q_i^m \right)^{1/m}$$

For all values of  $m > 1$ , the function  $Q$  is convex and the mixture property is synergistic; and for all values of  $1 > m > 0$ , the function  $Q$  is concave and the mixture property is antagonistic, with the geometric mean as a special case where  $m = 1/2$ . For instance, the octane number of gasoline often follows the rule

$$\langle Q \rangle = \left( \sum x_i Q_i^2 \right)^{1/2} \geq \sum x_i Q_i$$

So that a 50/50 mixture of two components with the octane numbers of 60 and 100 would have a mixture octane number of 82.5 instead of 80.0. The extra 2.5 octane number is known as an “octane bonus.”

The efficacy of an additive, such as tetraethyl lead on the octane number of gasoline, shows a concave curve so that the first cubic centimeter of additive would achieve much more than the second cubic centimeter, and the third cubic centimeter would do even less. This can also be described by the term “diminished return,” which is described by  $\partial^2 f / \partial x^2 < 0$ , which is the characteristic of a function with a diminishing value of  $\partial f / \partial x$  with increase in  $x$ .

It is advantageous to do mixing when the property is a convex function of the composition, so that the mixture property is higher than the arithmetic average; such functions have a second derivative  $d^2 y / dx^2 < 0$ , such as the function  $y = ax^{1/2}$ . This is a case of Jensen’s inequality, which states that, for a convex function, a chord between any two points is always below the curve; that is, the property of the average is always higher than the average of the two properties. On the other hand, it is not advantageous to do mixing when the property is a concave function of the composition, so that the mixture property is less than the arithmetic average.

For an ideal solution, where there is no heat of solution, Raoult’s law applies to the partial pressures of both components:

$$P_1 = x_1 P_1^\circ$$

$$P_2 = x_2 P_2^\circ$$

where  $P^\circ$  is the vapor pressure of the pure substance, and the total pressure is equal to the sum of the partial pressures. Any heat of solution will be added to the energy binding the liquid phase, and an endothermic heat of solution leads to  $\Delta\varepsilon > 0$ , which means a weaker binding between 1 and 2. The chemical potential and partial pressure

are given by

$$\mu_2 = -kT \frac{\partial G}{\partial N_2} = \mu_2^\circ + kT \ln x_2 + (1 - x_2)^2 \Delta \varepsilon$$

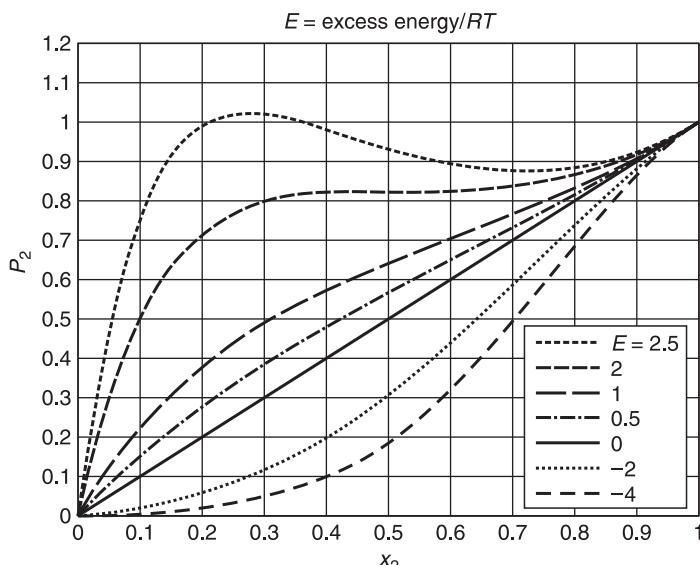
$$\frac{P_2}{P_2^\circ} = x_2 \exp \left[ \frac{(1 - x_2)^2 \Delta \varepsilon}{kT} \right]$$

When  $\Delta \varepsilon/kT = 0$ , such as in a mixture of pentane and hexane, Raoult's law for partial pressure will apply precisely. When  $\Delta \varepsilon/kT > 0$ , the 1–2 hetero-bonds are weaker than the average of the 1–1 and the 2–2 homo-bonds, such as between water and acetone, and we have a positive deviation to elevate the partial pressure. When  $\Delta \varepsilon/kT < 0$ , the 1–2 bonds are stronger, such as between acetone and chloroform, and we have a negative deviation to depress the partial pressure. The value of  $P_2$  versus  $x_2$  for a range of values of  $E/RT$  is given in figure 5.16.

When the value of  $x_2$  is very small, the expression becomes

$$P_2 = x_2 P_2^\circ \exp(\Delta \varepsilon/kT)$$

This is a form of Henry's law, where the Henry's law constant is simply the product of the vapor pressure of solute 2 multiplied by the energy correction term  $\exp(\Delta \varepsilon/kT)$ . For oleophilic substances, which do not like water, mixing is an endothermic process and the term  $\Delta \varepsilon/kT$  is positive, so that Henry's law constant is much higher than the vapor pressure of the pure substance.



**Figure 5.16** Partial pressure of solution, Raoult's law solution, and regular solutions with a range of values for  $\Delta \varepsilon/RT$

### Composite Material

A multicomponent solid material has many more degrees of freedom in arrangement than an isotropic and homogeneous fluid. The thermal conductivity of a composite material, formed by the lamination of sheets of two components with different thermal conductivities, is a well-analyzed system. When heat is flowing parallel to the sheet surfaces, the composite thermal conductivity is given by linear additivity of conductivities

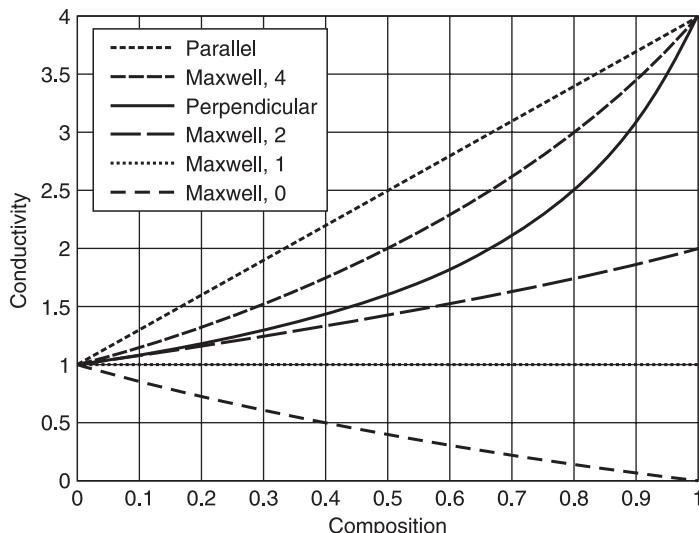
$$Q = x_1 Q_1 + x_2 Q_2$$

However, if heat flows perpendicular to the sheet surfaces, then the thermal conductivity of the composite is given by linear additivity of resistances:

$$1/Q = x_1/Q_1 + x_2/Q_2$$

Thus, the perpendicular conductivity is always less than the parallel conductivity. If the second component is a number of spheres embedded in a matrix of the first component, then the composite conductivity is given by the Maxwell equation when the volume fraction of spheres is very small:

$$\langle K \rangle = \frac{(2K + K') + 2(K' - K)x}{(2K + K') - (K' - K)x} K$$



**Figure 5.17** Heat conductivities of solids modeled by the series, the parallel, and the Maxwell formulation of spheres embedded in a matrix

where  $\langle K \rangle$  is the average thermal conductivity of the composite material,  $K$  is the thermal conductivity of the continuous matrix phase,  $K'$  is the thermal conductivity of the isolated spheres, and  $x$  is the volume fraction of the spheres.

Figure 5.17 gives the overall conductivity in the parallel and the perpendicular flows, as well as the Maxwell result when the matrix has unit conductivity and the balls have conductivities  $K = 4, 2, 1$  and  $0$ . It is seen that when  $x$  approaches zero,  $\langle K \rangle$  approaches  $K$ ; and when  $x$  approaches one,  $\langle K \rangle$  approaches  $K'$ . For the Maxwell composite solid, the values above  $x = 0.6$  have no significance, as the spheres will be in contact with each other.

## References

Hansch, C. and A. Leo. 1995. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. Washington, DC: American Chemical Society.

## Further Reading

- Allen, D. T. and D. R. Shonnard. 2002. *Green Engineering: Environmentally Conscious Design of Chemical Processes*. Upper Saddle River, NJ: Prentice-Hall.
- Assael, M. J., J. P. M. Trusler, and T. F. Tsolakis. 1996. *Thermophysical Properties of Fluids*. London: Imperial College Press.
- Benson, S. W. 1976. *Thermochemical Kinetics*. New York: John Wiley.
- Bondi, A. 1968. *Physical Properties of Molecular Crystals, Liquids and Gases*. New York: John Wiley.
- Carslaw, H. S. and J. C. Jaeger. 1959. *Conduction of Heat in Solids*. London: Oxford University Press.
- Charton, M. ed. 1999. *Advances in Quantitative Structure-Property Relationships*, vol. 2. Stamford, CT: JAI Press.
- Doucet, J.-P. and J. Weber. 1996. *Computer-Aided Molecular Design: Theory and Applications*. San Diego, CA: Academic Press.
- Fujita, T. 1995. *QSAR and Drug Design*. Amsterdam: Elsevier.
- Goodman, J. M. 1998. *Chemical Applications of Molecular Modeling*. London: Royal Society of Chemistry.
- Kier, L. B. and L. H. Hall. 1999. *Molecular Structure Description: The Electropotential State*. New York: Academic Press.
- Leach, A. R. 1996. *Molecular Modeling*. London: Longman.
- Livingstone, D. 1995. *Data Analysis for Chemists*. Oxford: Oxford University Press.
- Lyman, W. J., W. F. Reehl, and D. H. Rosenblatt. 1990. *Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds*. Washington, DC: American Chemical Society.
- Prausnitz, J. M., R. N. Lichthaler, and E. G. de Azevedo. 1986. *Molecular Thermodynamics of Fluid-Phase Equilibria*. Upper Saddle River, NJ: Prentice-Hall.
- Reid, R. C., J. M. Prausnitz, and B. E. Poling. 1987. *The Properties of Gases and Liquids*. New York: McGraw-Hill.
- Reinhard, M. and A. Drefahl. 1999. *Handbook for Estimating Physicochemical Properties of Organic Compounds*. New York: John Wiley.
- Thijssen, J. M. 1999. *Computational Physics*. Cambridge: Cambridge University Press.

## Websites

EPA Estimation Program Interface (EPI) Suite. Estimation for octanol–water partition coefficient  $\log K_{\text{ow}}$ , Henry’s law constant, melting point, boiling point, vapor pressure, water solubility. Requires SMILES notation with help guide.  
<http://www.epa.gov/oppt/exposure/docs/episuite.htm>.

## Exercises

1. Develop correlations for the van der Waals attraction term  $a$  and the volume term  $b$ , using the molecular weight as the predictor.
  - (a) Find the values of  $a$  and  $b$  for the following families of compounds: noble gases, period 1 to 6; normal paraffins, 1 to 20 carbon atoms; 1-chloro paraffins, 1 to 10 carbon atoms; 1-alcohols, 1 to 10 carbon atoms.
  - (b) Make graphs of van der Waals  $a$  (y axis) versus MW (x axis). How do these four families differ in their intercepts and slopes? Which chemical has the largest slope (most molecular weight dependent), and which chemical has the smallest slope?
  - (c) Make one-parameter linear correlations and compute their  $R^2$  and SD values.
2. Make a correlation for the viscosity of R134a in  $\mu\text{Pa s}$ , from  $-50$  to  $+100$   $^{\circ}\text{C}$  in  $10$   $^{\circ}\text{C}$  intervals. (Hint: a possible source of data is NIST Chembook.)
  - (a) Make a linear regression in the form  $\mu = C_0 + C_1 T$ . Obtain the coefficients  $C_0$  and  $C_1$ , as well as  $R^2$ ,  $\text{SD}(\mu - \mu')$ , the maximum error, and the maximum percentage error.
  - (b) Make a more accurate quadratic regression in the form  $\mu = C_0 + C_1 T + C_2 T^2$ . Compare the improvements of the quadratic correlation on the values of  $R^2$ , the standard error, the maximum error, and the maximum percentage error.
  - (c) Make an exponential regression in the form  $\mu = C_0 \exp(C_1 T)$ . How does it compare with the two previous correlations?
3. The correlations of liquid viscosities with structural and ambient parameters pose difficult problems. Yaws recommended the following equation for the viscosity at  $25$   $^{\circ}\text{C}$  for paraffins:

$$\log_{10}(\mu) = A + B/T + CT + DT^2 \quad T \text{ in kelvin}$$

The recommended parameter values, and the temperature range of validity, are:

- (a) Make plots of viscosity versus temperature in the range  $-50$   $^{\circ}\text{C}$  (coldest winter in North America) to  $250$   $^{\circ}\text{C}$  (hot engine) for these six compounds.

- (b) If the  $DT^2$  term is eliminated, how much difference would it make to the results?

	<i>A</i>	<i>B</i>	$10^2C$	$10^5D$	<i>Min. T</i>	<i>Max. T</i>
$C_6$	-5.0715	655.36	1.2349	-1.5042	178	507
$C_8$	-5.9245	888.09	1.2955	-1.3596	216	569
$C_{10}$	-6.0716	1017.7	1.2247	-1.1892	243	618
$C_{12}$	-7.0687	1253.0	1.3735	-1.2215	262	658
$C_{14}$	-7.8717	1446.7	1.4940	-1.2495	260	692
$C_{16}$	-8.1894	1557.0	1.5270	-1.2371	291	721

4. The Joule–Thompson coefficient measures the temperature change when a gas is expanded into a vacuum. For the refrigerant R134a, we have the following values:

<i>Temperature, °C</i>	<i>Joule–Thomson coefficient <math>\times 10^5</math>, K/torr</i>
-50	-4.19
-40	-3.92
-30	-3.61
-20	-3.24
-10	-2.79
0	-2.26
10	-1.60
20	-0.770
30	0.297
40	1.71
50	3.66
60	6.48
70	10.9
80	18.6
90	35.9
100	109.3

- (a) Make a plot of the Joule–Thompson coefficient versus temperature. What functional form would you suggest that would require the least number of coefficients to achieve the highest value of  $R^2$ ?
- (b) Make the correlation; compute the values of  $R^2$ , standard error, maximum error, and maximum percentage error.

5. The group contribution method, which is conveniently packaged in the software Cranium, can be useful when experimental data and correlations are not available. Let us consider the properties of pentane, isopentane, and neopentane.
- Estimate the vapor pressures (torr) of these three pentanes from 0 to 40 °C with 5 °C intervals, and make a plot.
  - Estimate the melting points (K) and boiling points (K) of these three pentanes and compare them with experimental values. What are the errors and the percentage errors in these estimates, and comment on their accuracy?
  - Estimate and compare with data on the heat of melting (J/mol), heat of vaporization, critical temperature, critical pressure (atm) and critical volume (L/mol) of pentane and neopentane. What are the percentage errors in these estimates, and are you impressed by the accuracy?
  - Estimate the heat capacity (J/(K mol)) of isopentane from 373 to 473 K with 10 K intervals and make a plot. Do the results look reasonable to you?
6. Mustard gas is the common name for bis(2-chloroethyl)thioether. It was used in World War I to kill people, and it can penetrate the rubber in gas masks. Its formula is  $\text{Cl}-(\text{CH}_2)(\text{CH}_2)-\text{S}-(\text{CH}_2)(\text{CH}_2)-\text{Cl}$ .
- Make the structure in Cranium, and estimate its molecular weight, melting point, and boiling point. Make a note of the estimation technique used.
  - Check against the experimental data from a reliable source. How well does the method work?
  - What are the vapor pressures (mmHg) of mustard gas at 0, 10, 20, 30, and 40 °C? Why is this information needed, and by whom?
7. DDT is the celebrated insecticide that was widely used to control mosquitoes and malaria, and denounced by Rachel Carson in *Silent Spring*. Make the structure, and estimate its MW, m.p. and b.p., and check on the accuracy of these properties.
8. There is only one form each of benzene, monochlorobenzene, pentachlorobenzene, and hexachlorobenzene, but there are three forms each of dichlorobenzene, trichlorobenzene, and tetrachlorobenzene. The following is a list of these 13 forms and some of their properties:

	<i>Cl</i>	<i>Dipole, D</i>	<i>Symmetry</i>	<i>B.p., K</i>	<i>M.p., K</i>
Benzene	0	0.00	12	353.3	228.0
Mono-Cl	1	1.04	2	404.9	178.6
1,2	2	2.14	2	453.7	256.0
1,3	2	1.38	2	446.2	248.4
1,4	2	0.00	4	447.3	326.1

*continued*

	<i>Cl</i>	<i>Dipole, D</i>	<i>Symmetry</i>	<i>B.p., K</i>	<i>M.p., K</i>
1,2,3	3	2.33	2	491.7	323.8
1,2,4	3	1.26	1	486.7	290.1
1,3,5	3	0.00	6	481.2	336.3
1,2,3,4	4	1.90	2	527.2	320.0
1,2,3,5	4	0.65	2	519.2	323.9
1,2,4,5	4	0.00	4	516.2	412.2
Penta-Cl	5	0.88	2	550.2	357.7
Hexa-Cl	6	0.00	12	597.7	501.3

- (a) Make a plot of their boiling points and melting points against the number of chlorine atoms. How good is the simple linear regression for the boiling point? How good is the regression for the melting point?
- (b) Would the boiling point correlation be much better if we include the dipole moment in a multilinear regression?
- (c) Would the melting point correlation be much better if we include the dipole moment and the  $\log(\text{symmetry})$  in a multilinear regression?

# 6

## Estimation by Associations and Trends

### **6.1 Empirical Structure–Property Relations**

Toxicity parameters  
Environmental parameters

### **6.2 Structure Effects and Ambient Influences on Properties**

**References**

Solubility in water  
Solid and liquid density

**Further Reading**

### **6.3 Risk Assessment**

**Websites**

Safety parameters

**Exercises**

Product engineers are often challenged to estimate:

1. the properties of new or obscure classes of material that have not been extensively studied before, or have never been made before;
2. the property changes on a material due to physical and chemical modifications;
3. the properties under a variety of ambient conditions, such as temperatures, pressures, electromagnetic fields, and prolonged environmental exposures and weathering.

When theoretical understanding is insufficient and quantitative correlations are not available, we can often make useful qualitative estimations by using fragmentary empirical structure–property relations. The principal tools are observations of associations and trends, which are often the only methods available in biological, health, safety, and environmental properties.

Association is based on analogy, or the probabilistic assumption that “similar substances have similar properties.” Does  $x^\circ$ , the material in question, resemble a known substance  $x_1$  or a class of substances  $\{x_1, x_2, \dots, x_i, \dots\}$  that have known properties and structure? If we can find such a match, then we can use it to make qualitative property estimates for  $x^\circ$ . We know that metals are nearly always solids at room temperature, but mercury is an exception. Trend is based on the study of the variation of properties among the substances  $\{x_1, x_2, \dots, x_i, \dots\}$  and how they depend on structure variations,

which amounts to qualitative and empirical structure–property relations. For example, we know that a larger molecule tends to have higher melting and boiling points, but chlorobenzene has lower melting point than benzene.

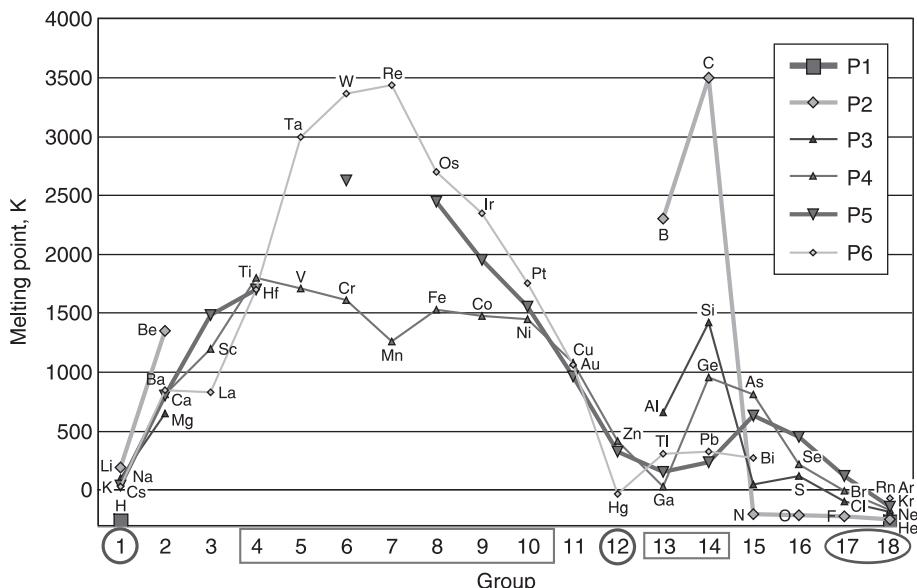
## 6.1 Empirical Structure–Property Relations

Thomas Midgley used qualitative trend analysis to help him to discover the chlorofluorocarbons as refrigerants. He did not have a well-organized database, nor did he have an established method of analysis and prediction, and his method was ad hoc and empirical. What he did know was that sulfur dioxide and ammonia were in use as refrigerants. He looked at the periodic table of nonmetallic elements and observed the trend that small compounds of these elements tend to become less flammable when we move from the left to the right of the table, which is certainly true in the sequence CH<sub>4</sub>, NH<sub>3</sub>, H<sub>2</sub>O, and HF. The small compounds also tend to become less toxic when moving from the bottom to the top, which is certainly true in the sequence H<sub>2</sub>Te, H<sub>2</sub>Se, H<sub>2</sub>S, and H<sub>2</sub>O. Thus, he concentrated his subsequent work on fluorine compounds, even when he did not have quantitative data or reliable estimations of toxicity and flammability.

C	N	O	F
Si	P	S	Cl
As	Se	Br	
Te	I		

The compound CF<sub>4</sub> can be made by replacing all the H atoms of methane by fluorine, but it has a boiling point of  $-128\text{ }^{\circ}\text{C}$ , which is much too low for consideration. Another observable trend is that the boiling points increase with the number of chlorine atoms (see figure 1.3), so all he had to do was to move from CF<sub>4</sub> to the left of the diagram and add one or two chlorine atoms. Midgley's analysis is a good example of an empirical structure–property relation that is based on observations instead of theory, and it amounts to a hypothesis that is probably valid but which may breakdown, especially when it is extrapolated to a distant region. Since there are no firm theoretical grounds, the prudent thing to do is to use this hypothesis to suggest areas that merit further investigation, which is what Midgley did. It can also serve as the beginning of a cycle of the scientific method: observations, hypotheses, predictions, further experimentations, confirmations, or revisions of the hypothesis, new predictions, etcetera.

The method begins with collecting all available information on substances  $\{x_1, x_2, \dots, x_i, \dots\}$  that are similar to the substance in question  $x^{\circ}$ . These should be the substances that have the same basic structure as  $x^{\circ}$ , but with some small structure differences, such as the number of repeated units, substitution of atoms by other atoms or small groups, rearrangement of the positions of substitution, rearrangement of the skeleton of the molecule, etcetera. The next move would be to subdivide the set  $S$  into structure-based subsets  $\{S_a, S_b, \dots\}$  and to observe the differences among their structures and their property values. One then proposes empirical structure–property relations; for instance: “the members of family  $S_a$  tend to have higher values of  $y$  than members of family  $S_b$ , and within a family  $S$ , the property  $y$  may be positively



**Figure 6.1** The melting points of the elements, arranged in lines by their periods, plotted against the group number

correlated with predictor  $x$ .” The utility of these empirical structure–property relations is in whether they generate useful information, and whether they can be used to make predictions that turn out to be accurate.

Metals are usually hard and difficult to melt; this suggests that a newly discovered metal is also expected to be hard and difficult to melt. Figure 6.1 shows the melting points of the elements arranged by groups and periods. Indeed, the metals of groups 5 to 9 have the highest melting points and are generally above 1500 °C, whereas the metals of the other groups have melting points below 1500 °C. This association of metals with the properties of hard and high melting points does not always work, as sodium is soft and mercury is liquid at room temperature. The melting points of the group 5 to 9 metals increase with the period; but this trend does not always work, as the melting points of the group 1 and 2 metals decrease with the period. Thus, empirical structure–property relations can be used to suggest associations and trends, but a prudent investigator must follow up with experiments to validate the range of applicability, or to revise the relations.

The reverse expectation, that two sets of materials with similar properties should have similar structures, is another hypothesis that has utility, as well as limitations and exceptions. Boron and carbon have melting points above 2000 °C, but they are not metals. Another example is that not all sweeteners have the same structure. The sugar sucrose is a disaccharide that can be hydrolyzed into the monosaccharides of glucose and fructose. Glucose and fructose have the same chemical formula of  $C_6H_{12}O_6$ , but fructose forms a five-atom ring and is two to three times sweeter than glucose, which forms a six-atom ring.

There are three well-known synthetic substitutes to sugar: saccharine, cyclamate, and aspartame; all have their own advantages and disadvantages. They were discovered by accident, and their structures are totally different from sucrose. Saccharin,  $C_7H_5NO_3S$ , was discovered by Ira Remsen in 1879, and has a fused ring structure of benzene with a five-member hetero-ring containing S and N. Cyclamate,  $C_6H_{13}NO_3S$ , was discovered in 1937 and has a cyclohexane ring connected to an  $NHSO_3Na$  group. Aspartame,  $C_{14}H_{18}N_2O_5$ , was discovered in 1965 and has a chain containing a primary and a secondary amine group, plus a benzene ring on the side. All these substances have the ability to produce a sweet sensation on the taste buds, but they do not have the same structure. We do not know the common structural element that gives rise to the sweet-tasting properties of these chemicals. We may also conclude that since similar properties can arise from widely different structures, there may be other classes of structures out there that have even better properties.

The smell of musk is important in perfumes and cosmetics, and is obtained from the glands of small animals. There are several types of synthetic musk: the aromatic musk is built on dinitrobenzene; the steroid musk is built around cholesterol as four staggered fused rings; and the macrocyclic musk is built on a 15-member carbon ring. Aside from the ability to elicit a pungent sensation in the nostrils and an emotional response of masculinity, we have not found the common structure responsible for these sensations.

The *Merck Manual* lists 17 items under the therapeutic category of “Anesthetics (inhalation)” that produce a loss of feeling or sensation, including cyclopropane, ethylene, ethyl ether, halopropane, nitrous oxide, and xenon. A leading theory to reconcile this structural diversity and functional similarity is in their high values of  $\log P$ , or high partition coefficients into the lipid layers of nerve cell endings. However, this is not a necessary and sufficient condition for anesthesia, as not all compounds with high  $\log P$  are inhalation anesthetics. The *Merck Manual* also gives 68 items under “Anesthetic (local),” which include drugs such as benzocaine, cocaine, ethyl chloride, lidocaine, methyl chloride, phenol, procaine, and salicylic alcohol. These items can be separated into distinct families and methods of action. Methyl and ethyl chlorides are sprayed on the skin, and their evaporation cools the tissue to make it less sensitive to pain. Benzocaine and procaine are amino benzoic acid esters, lidocaine is a phenyl acetamide, and cocaine is an ester between benzoic acid and cyclo-octane carboxylic acid.

Among the category “Analgesic (non-narcotic),” whose members produce insensitivity to pain, there are 133 items, including acetylsalicylic acid (aspirin,  $C_9H_8O_4$ ), acetaminophen (Tylenol,  $C_8H_9NO_2$ ), and ibuprofen (Advil,  $C_{13}H_{18}O_2$ ), all of which have a benzene ring. Aspirin is made by adding an acetyl group on salicylic acid, Tylenol is a phenyl acetamide, Advil is a benzeneacetic acid. There are 73 entries under “Analgesic (narcotic);” these compounds produce a stupor or arrested activity and include codeine, morphine, and methadone. The first two are opiates derived from opium, sharing the same structure with four fused rings, a quaternary carbon atom, aromatic rings, and several oxygen and nitrogen atoms. However, methadone is not derived from morphine, it being built around an entirely different structure with no fused rings.

An empirical structure–property relation is based on observations in a set of fragmentary and incomplete data, instead of being based on a well-established theory of wide applicability and tested extensively. It is a tentative hypothesis designed to explain

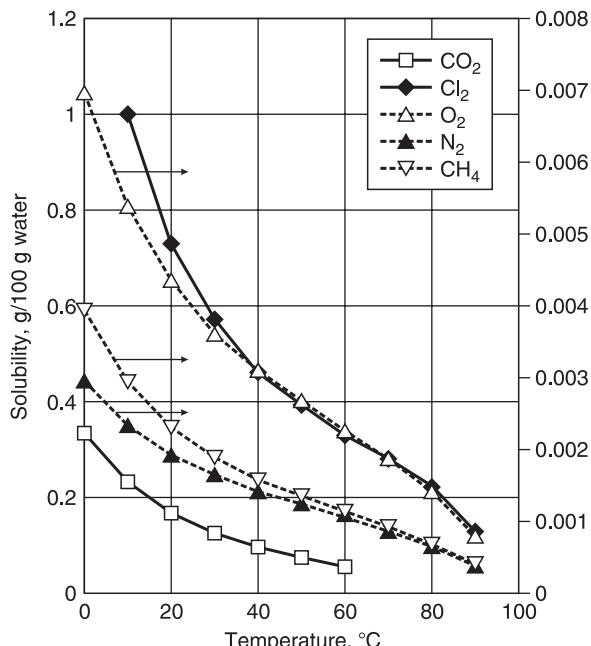
a few observations in a local neighborhood, which may or may not be valid when extrapolated to another region. The associations and trends derived can be useful to provide interim answers that can be useful to sort out and rank the potentials for further investigations among a set of candidates. Since there are no firm theoretical grounds, the prudent thing to do is to use this hypothesis to rank a number of candidates, which should then be followed up with measurements of their properties; this procedure also adds to the existing database and would eventually lead to improvements to the understanding and confidence in prediction.

## 6.2 Structure Effects and Ambient Influences on Properties

We give here a number of examples of empirical structure–property relations, and how they can be used to predict the effects of structure modifications and ambient conditions on properties.

### Solubility in Water

Theory leads us to expect that substances that are similar to water would be more soluble in water, and the process of solution needs a decrease in free energy, which means either a large decrease in enthalpy or a large increase in entropy. Figure 6.2 shows the effect of temperature on the water solubility of a number of gases. The reactive gases  $\text{CO}_2$  and  $\text{Cl}_2$  at a pressure of 1 atm generally have a higher solubility in water,



**Figure 6.2** The water solubility of gases with temperature at 1 atm. The scale for  $\text{CO}_2$  and  $\text{Cl}_2$  is to the left, and the scale for  $\text{O}_2$ ,  $\text{N}_2$ , and  $\text{CH}_4$  is to the right

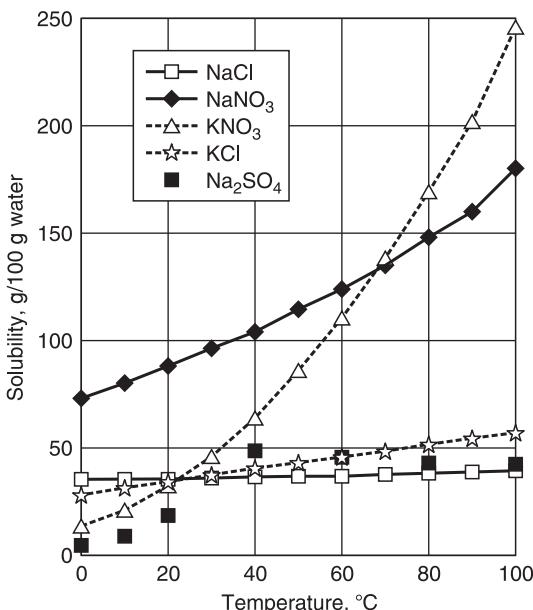


Figure 6.3 The water solubility of salts with temperature

whereas the inert oxygen, nitrogen and methane have much lower solubilities. They all show a strong decline of solubility with temperature, and it is expected that, at the water boiling point of 100 °C, the solubility of all gases should fall to zero. Figure 6.3 shows that the inorganic salts have the opposite tendency of higher solubility at higher temperatures: the solubilities of nitrates are very sensitive to temperature, the chlorides are insensitive, and the sulfates can rise and fall with a maximum. It is difficult, for example, to use these observations to construct a trend to predict the solubility of PCBs in the hot water discharged from a power plant and whether this solubility would increase or decrease upon meeting the cool river or lake.

Figure 6.4 shows that when a hydrogen atom in benzene has been replaced with a functional group, if that group is hydrophilic it would cause an increase in the water solubility and a hydrophobic group would cause a decrease in the water solubility. In fact, their modifications are ranked in the order



It is expected that the hydrophilic groups would make the compound less oleophilic, and thus decrease the octanol–water partition coefficient, or  $\log P$ , which is also demonstrated in figure 6.4. When multiple functional groups are added, there are often interactions among them, so that the results are not simply additive. Figure 6.5 shows that each addition of chlorine to benzene from one to three would depress the water solubility of these benzene derivatives by a factor of six to eight, but the totally chlorinated hexachlorobenzene has a water solubility that is six orders of magnitude smaller than benzene. The partition coefficients show a very consistent pattern in parallel with

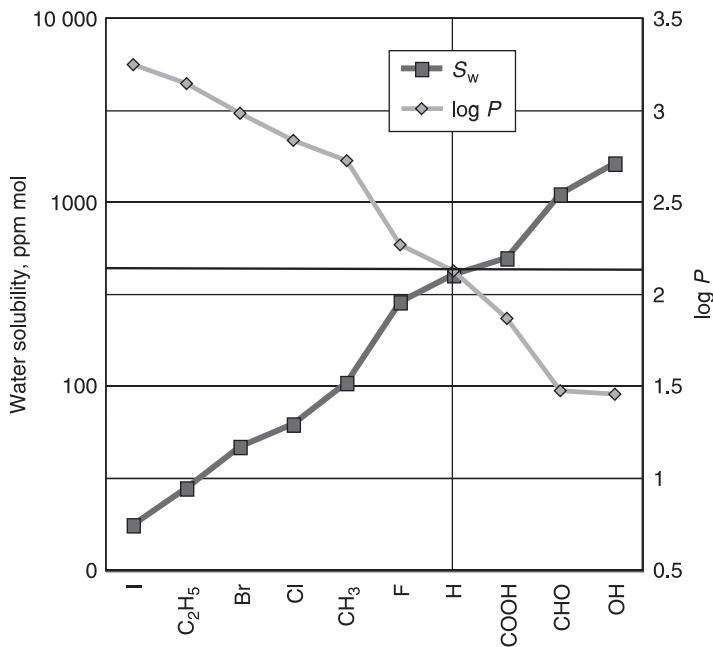


Figure 6.4 The water solubility and log partition coefficients of benzene derivatives

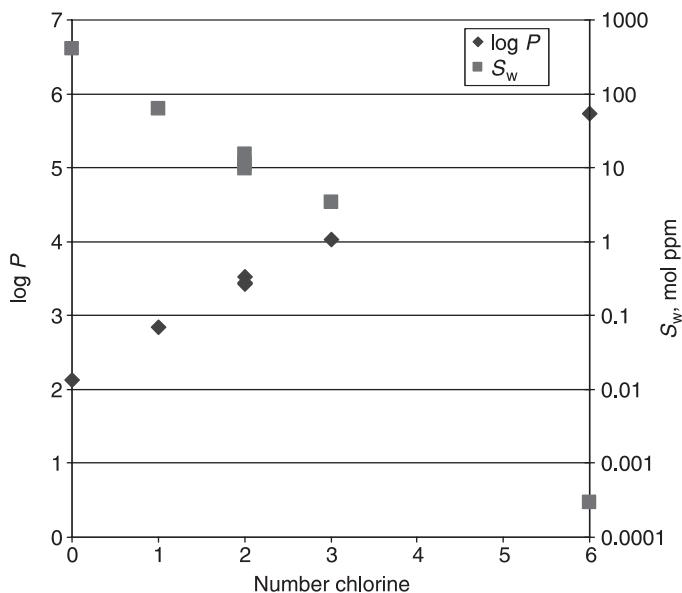


Figure 6.5 The water solubility and log partition coefficients of polychlorinated benzenes

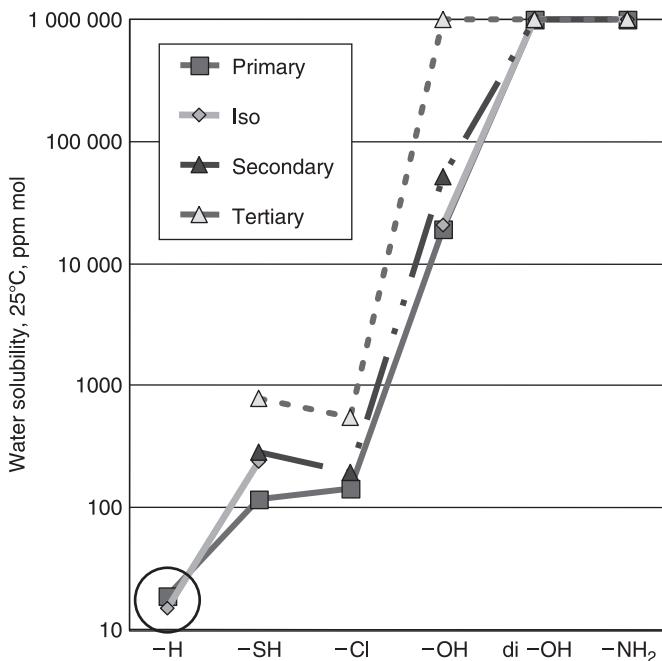


Figure 6.6 The water solubility of butane derivatives by substitution groups

the water solubility; that is, multiple chlorine atoms make a compound increasingly oleophilic.

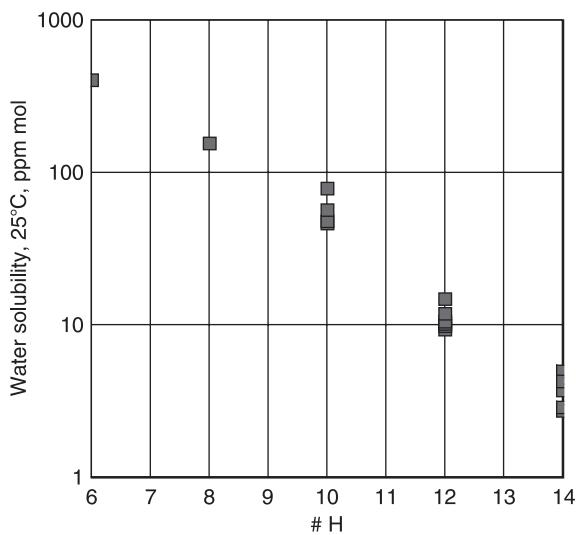
Figure 6.6 shows that the substitution positions of the  $-\text{SH}$ ,  $-\text{Cl}$ , and  $-\text{OH}$  groups on butane have a significant role in modifying water solubility, and are ranked in the order



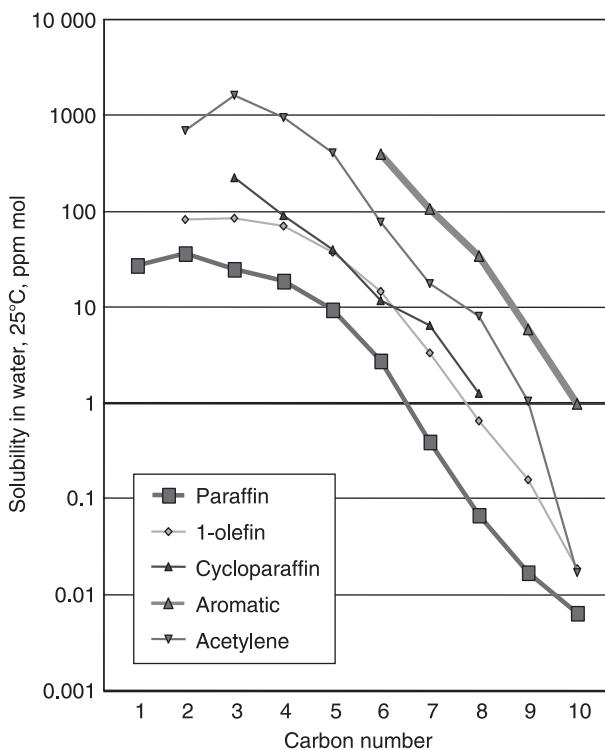
However, all the derivatives with two  $-\text{OH}$  and with  $-\text{NH}_2$  are completely miscible with water.

Hydrocarbon bond saturation and cyclization also play roles in water solubility. Figure 6.7 shows that, among the six-carbon hydrocarbons, the various forms of hexane,  $\text{C}_6\text{H}_{14}$ , have the lowest solubility, and the hexenes and cyclohexane with the formula  $\text{C}_6\text{H}_{12}$  have three times the solubility. Fewer hydrogen atoms consistently lead to higher solubilities, and benzene has one hundred times the water solubility of normal and iso-hexanes.

The length of a hydrocarbon molecule also has a depression effect on water solubility. Figure 6.8 shows the series of paraffins, 1-olefins, cycloparaffins, aromatics, and acetylenes. The longer hydrocarbon molecules with more  $-\text{CH}_2-$  groups have much lower water solubility, so that each additional  $-\text{CH}_2-$  group drops the water solubility by a factor of almost five.



**Figure 6.7** The water solubility of C6 hydrocarbons by degrees of hydrogen saturation



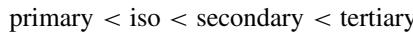
**Figure 6.8** The water solubility of five series of hydrocarbons by number of carbon atoms

A summary of this small roster of empirical structure–property relations on water solubility shows:

- Solubility of low molecular weight gas goes down with temperature, but inorganic salt solubility goes up with temperature;
- When a functional group is substituted for H in benzene, the solubility changes with the ranking



- multiple substitution groups in benzene affect solubility in roughly geometric series instead of arithmetic series;
- the position of substitution groups in butane affects the solubility with the ranking

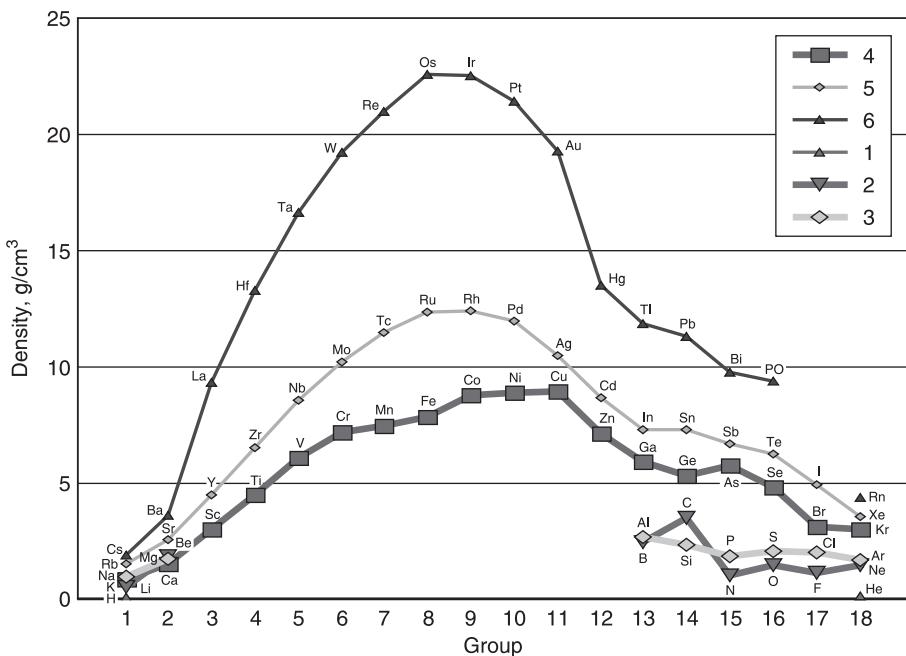


- removing H atoms and creating double bonds or cyclization in hydrocarbons with six carbons increases solubility;
- adding  $-\text{CH}_2-$  groups to many hydrocarbon series depresses solubility.

This set of working hypotheses can be used to guide explorations on how to make structure changes in other chemical species to influence water solubility. However, its reliability has to be tested, especially when extrapolated to materials that are far from these simple organic compounds. The verification also enlarges the database, and may lead to revisions of the hypothesis.

### Solid and Liquid Density

The density of condensed matter is determined by three factors: the density of the elements involved, the lengths of the bonds between the atoms, and the packing efficiency among molecules. The densities of the elements are shown in figure 6.9. There is a general tendency for the elemental density to increase from the lower periods to the higher periods: the metallic elements in groups 3 to 13 have higher densities than the alkali and alkaline earth metals of groups 1 and 2 and the metalloid and nonmetallic elements of groups 14 to 18. There are notable exceptions to the rule, as carbon has a higher density than silicon, both in the diamond and in the graphite structures. Table 6.1 shows that the halogen atoms show increasing densities with the period; but, among the alkali elements, the density of potassium is at a local minimum at  $0.89 \text{ g/cm}^3$ , which is lower than either of its neighbors sodium and rubidium. Part of this curious local minimum trend lies in the fact that the density is a ratio of the atomic weight divided by the atomic volume, and they both increase with the period, but at different rates. The densities of the alkali–halogen salts (1–17) are at a maximum with  $\text{CsI}$  at  $4.51 \text{ g/cm}^3$ , as expected. However, the densities reach a distinct minimum at potassium chloride:  $1.988 \text{ g/cm}^3$ . In fact, all the densities of any  $\text{K}-\text{X}$  salt are lower than the corresponding  $\text{Li}-\text{X}$ ,  $\text{Na}-\text{X}$ ,  $\text{Rb}-\text{X}$  and  $\text{Cs}-\text{X}$  salts; this is consistent with the fact that metallic potassium has the lowest density among all the alkali elements except



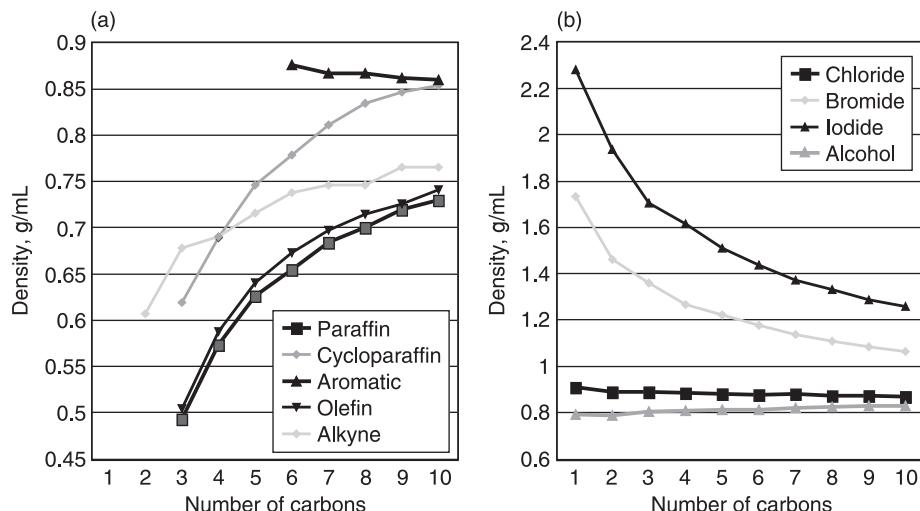
**Figure 6.9** The density of the elements, arranged in lines by their period, and plotted against group number

**Table 6.1 Densities of metal and alkali-halogen salts**

Element	Metal	F	Cl	Br	I
Li	0.534	1.667	2.98	3.102	4.63
Na	0.968	2.295	2.068	3.464	4.061
K	0.89	2.79	2.163	3.205	3.667
Rb	1.532	2.48	1.988	2.75	3.13
Cs	1.879	2.88	2.76	3.35	3.55

for lithium. The density of any Y—Cl salt also tends to be lower than the corresponding Y—F, Y—Br, and Y—I salts.

The influence of bond lengths on densities is more pronounced in organic compounds. The nonbonded low-energy van der Waals radii are longer than the three bonded high-energy radii: ionic, covalent and metallic. Since decane has many more intermolecular covalent bonds than methane, it also has a much higher density; see figure 6.10. The carbon–carbon bond lengths are ranked in the order of: single bond > double bond > triple bond, so that we should anticipate that paraffins have lower densities than olefins, which in turn have lower densities than the alkynes.



**Figure 6.10** Liquid density and the number of carbon atoms: (a) five hydrocarbon series; (b) four polar compound series

The aromatic bonds are also shorter than the single bonds, and cyclic aromatics are more compact than acyclic paraffins; thus, benzene has a higher density than hexane.

The influence of packing efficiency can be an important contributory factor, as carbon has a density of  $3.51\text{ g/cm}^3$  in the diamond form, but only  $2.25\text{ g/cm}^3$  in the graphite form. When we pack hard spheres in the most important configurations, we have the situation shown in table 6.2.

Let us examine the densities of organic compounds shown in figure 6.10. The compounds that have the heavy heteroatoms of bromine and iodine are much denser than the hydrocarbons, which are generally less than  $1\text{ g/cm}^3$ . The compounds that have the lighter heteroatoms of chlorine, oxygen, and nitrogen are intermediate in density; the hydrocarbons are of low density and the more saturated hydrocarbons are of very low density. The effect of adding  $-\text{CH}_2-$  groups is very different for these widely different sets: the heavy heteroatom series have declining density with  $N_{\text{C}}$ , the light heteroatom and the aromatic series have nearly constant densities,

**Table 6.2** Hard sphere packing

Type	Coordination	Packing efficiency	Example
Tetrahedral	4	0.510	Diamond
Simple cubic	6	0.523	NaCl
Body-centered cubic	8	0.680	CsCl, K, Fe
Face-centered cubic	12	0.740	Ni, Au, Ar
Hexagonal close-packed	12	0.740	Be, Cd, Mg

and the other hydrocarbons have rising densities. We can summarize these tendencies as follows:

- When the functional group is heavy, density decreases with  $N_C$ ; when the functional group is H, density increases with  $N_C$ .
- Dehydrogenation increases density.
- Cyclization increases density.

### 6.3 Risk Assessment

Product engineers need to evaluate the potential impact of a new or modified product, as well as the associated raw materials and by-products, concerning health, safety, and the environment. In the case of PVC pipes and polyurethane foams, the most hazardous substances involved are the raw materials of vinyl chloride and phosgene, rather than the final products. The hazardous raw materials and by-products are encountered mainly during transportation and within a chemical plant, and it is possible to design and install high-security barriers and controls under the watchful eyes of competent engineers to protect the workers and to prevent release to the neighborhood. However, hazardous products are distributed to the public, who have no special training and equipment, and so must be designed to make them harmless.

The availability of data on the health and safety aspects of chemicals has depended on the accumulation of cases of deplorable accidents leading to illnesses and death. The 1970 establishment of the EPA and the Occupational Safety and Health Administration (OSHA) led to large-scale and systematic investigations on the health, safety, and environmental hazards posed by chemicals. Even today, we have available data only for notorious and highly regulated substances on toxicology, flame and explosion, smog formation, water pollution, stratospheric ozone destruction potential, and global warming capacity. A few substances have been measured adequately and declared safe enough for use with caution, but the vast majority of substances are in the limbo of uncertainty. This is similar to the situation facing the judge and jury in a criminal trial in Scotland, who may declare three different verdicts about a defendant: guilty, not guilty, or guilt not proven. A defendant in the third category of limbo does not go to jail, but neither has he been cleared sufficiently that he can hold his head high.

During the exploration and discovery phase of a product innovation, when a proposed new or modified product is not on the list of hazardous substances, it is in limbo and the product engineers should proceed cautiously into the product development phase. The next move should be to examine whether the proposed product can be associated with any of the substances on the list, which would serve as a warning that there is a higher probability of hazard. When the product becomes very promising during the development phase, the product engineers must follow up with synthesizing the product and making laboratory and field tests. If some of the raw materials and by-products are on the hazardous lists, then prompt discussions with the manufacturing department should take place to provide for adequate planning of design and control. On the other hand, if the proposed new product itself is on the list of hazardous substances, then it can still be used by the industrial or military markets, but the only strategy for the

civilian market would be to modify or abandon this product in favor of more promising leads.

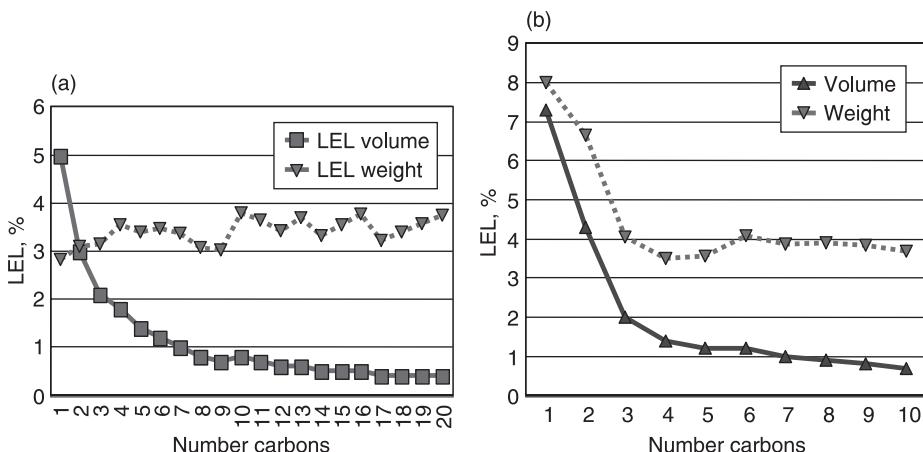
In this relatively young field of research, the experimental and theoretical results lend themselves better to associations and trends analysis than to quantitative correlations and theoretical predictions. Empirical structure–property relations should be regarded as temporary props that can be useful, but should be modified if serious exceptions are found. They should be regarded as a stage in the evolution from ignorance to empiricism, and then into knowledge and reliable theory, and should be supplanted when better ideas arrive.

### Safety Parameters

Some of the more important lists of hazardous substances controlled by the OSHA begin with safety from flame and explosion.

#### *Lower Explosion Limit*

When a combustible substance is mixed with air, the mixture will explode only when it is neither too rich nor too lean. The lower explosion limit (LEL) is the minimum volume percent of the substance in air with flammability, which is separated from the upper explosion limit (UEL) by the explosive concentration range. The tabulations in handbooks are based on experimental data, and sometimes derived from estimation methods based on the elemental composition of the fuel as  $C_mH_xO_y$ . Figure 6.11 shows the LEL for the series of normal paraffins and of 1-alcohols versus the number of carbon atoms. There are two ways to plot the results, which show that, for paraffins, the volume percent shows a steeply declining trend, but the weight percent shows a mildly increasing trend. One may conclude that a smaller volume percent of higher paraffin



**Figure 6.11** The LELs of (a) paraffins and (b) alcohols, expressed in volume percent and in weight percent.

is needed for explosion, in comparison with methane; but one can also conclude that the same weight percent of higher paraffin is needed for explosion, in comparison with methane. Qualitatively, the 1-alcohols behave in a similar manner, and it takes 3–4 wt% to reach the LEL.

### Flash Point

The flash point  $T_f$  represents the temperature at which a liquid gives off enough vapor to burn when exposed to an external ignition source or spark. The tabulations in handbooks are based on both experimental data and estimated values. Figure 6.12 shows that the paraffins lighter than the gasoline range, with less than 10 carbon atoms, burn easily at room temperature, but kerosene and higher paraffins do not. The alcohols require much higher temperatures to ignite, which is why brandy or vodka at room temperature, which are only 50% alcohol, cannot burn to make Cherries Jubilee without being heated up first. The critical requirement appears to be a temperature that is sufficiently high to generate a vapor pressure to exceed the LEL. At a room temperature of 25 °C (77 °F), octane has a vapor pressure of 1.86 kPa and nonane of 0.57 kPa, which is in the range of the LEL; but decane, with a vapor pressure of 0.17 kPa, is below the range of the LEL.

### Toxicity Parameters

The available regulatory quantitative measures are published by the government agencies with the responsibility of safeguarding health and safety. The FDA was established in 1938, to inspect food for humans and animals, drugs, medical devices, and cosmetics. The EPA was established in 1970, to protect human health and the environment, with

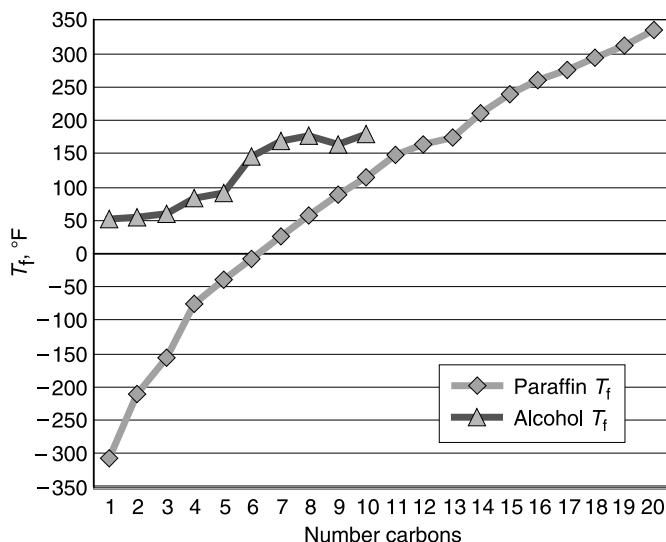


Figure 6.12 The flash point (°F) of paraffins and alcohols

jurisdiction over pesticides, industrial chemicals, drinking water supplies, hazardous waste control, and toxic pollutants in water and air. The OSHA was also established in 1970, to safeguard workplace exposure of workers. The Consumer Product Safety Commission (CPSC) was established in 1972 to examine products that are toxic, corrosive, combustible, radioactive, and pressure generating.

### *Lethal Dose 50*

The LD<sub>50</sub> is the amount of a material ingested at one time that causes the death of 50% of the test animals in a group. This is a measure of short-term poisoning potential, or acute toxicity, of a material. It is usually expressed in milligrams per kilogram of body weight, and designated as LD<sub>50</sub> (oral), LD<sub>50</sub> (skin), LD<sub>50</sub> (i.v.) according to the different routes of entry to a body. Different values are found for different animals, so that they should be designated as Oral LD<sub>50</sub> (mouse) and Dermal LD<sub>50</sub> (rat). Table 6.3 lists the tremendous range of some of the most important materials that have been measured, where many of the most toxic materials are plant alkaloids, such as

**Table 6.3 Lethal dose for 50% of animals  
(source: Sunshine 1969)**

Chemical	LD <sub>50</sub> , mg/kg	Animal
Sulfapyridine	27,000	Mouse
Succinic acid	8,500	Rat
Bromacil	5,200	Rat
Malathion	2,500	Rat
Salicylic acid	1,300	Rabbit
Imipramine HCl	682	Rat
Lidocaine	520	Mouse
Chlordane	395	Rat
Mirex	306	Rat
DDT	285	Rat
Morphine	250	Mouse
Caffeine	200	Rat
Heptachlor	130	Rat
Kepone	129	Rat
Lindane	125	Rat
Aldrin	67	Rat
Dieldrin	60	Rat
Warfarin	58	Rat
Nicotine	55	Rat
Amphetamine	38	Mouse
Endrin	25	Rat
Dimethyl sulfoxide	20	Rat
Strychnine	16	Rat
LSD	16	Rat
Parathion	4	Rat

**Table 6.4 The six toxicity classes of Gosselin, Smith, and Hodge (source: Loomis 1978)**

Class	Dose	For 70 kg person
6 Super-toxic	<5 mg/kg	7 drops
5 Extremely toxic	5–50 mg/kg	4 mL (7 drops to a teaspoon)
4 Very toxic	50–500 mg/kg	30 mL (teaspoon to fluid ounce)
3 Moderately toxic	0.5–5 g/kg	30–600 mL (fluid ounce to pint)
2 Slightly toxic	5–15 g/kg	600–1200 mL (pint to quart)
1 Practically nontoxic	above 15 g/kg	>1200 mL (>1 quart)

strychnine and the multi-chlorinated hydrocarbons such as parathion. The six toxicity classes according to Gosselin, Smith, and Hodge are shown in table 6.4.

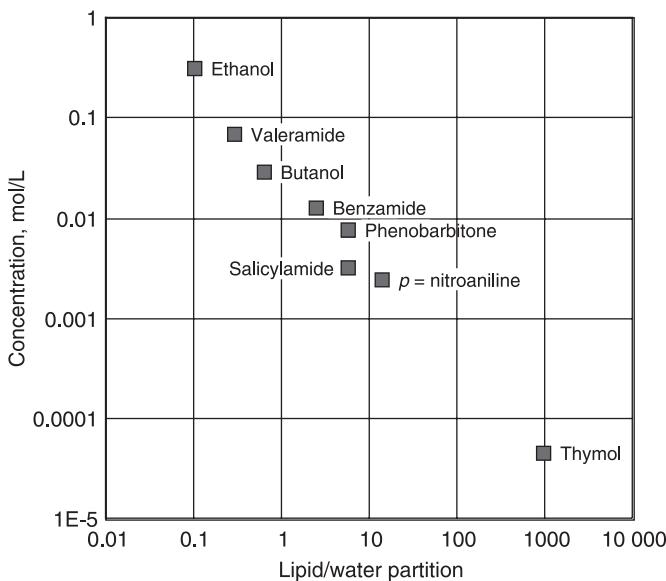
Parathion is super-toxic, amphetamine is extremely toxic, chlordane is very toxic, malathion is moderately toxic, succinic acid is slightly toxic, and finally sulfapyridine is practically nontoxic.

A large sample size of animals is required to reduce the uncertainty level of a measurement, which is of the order of  $1/\sqrt{N}$ , where  $N$  is the number of animals tested; thus, 100 animals will give approximately a 10% accuracy, and 10,000 animals will be needed to give a 1% accuracy. Such tests are expensive and time consuming, and will be undertaken only when there is an urgent need for the information, so it would not be done routinely for millions of substances. The substances that are not tested and reported here are in limbo, and not necessarily harmless. The substances that have been tested for LD<sub>50</sub> are mainly pesticides, drugs, and dangerous substances. A class of the most dangerous substances is alkaloids with plant origin, such as strychnine, nicotine, caffeine, and morphine. Another class of very dangerous substances is the highly chlorinated pesticides with 6 to 12 chlorine atoms, including endrin, aldrin, dieldrin, heptachlor, chlordane, and mirex. Organophosphorus esters are another dangerous class of pesticides that contain phosphorus, and include the nerve gas sarin.

Figure 6.13 shows the concentrations of organic compounds in water that are sufficient to immobilize tadpoles, and their values of log  $P$  or the partition coefficient between octanol and water. Thymol, with a log  $P$  of 1000, is 10,000 times more toxic than ethanol, which has a log  $P$  of around 0.3. A plausible explanation of this phenomenon lies in the structure of the membranes of cells, which are made of two layers of lipid, or fat, so that a more fat-soluble substance would find it easier to penetrate the cell and to cause damage.

#### *Threshold Limit Value for Breathing Air*

The TLV is given for exposure limits in air to safeguard the health of the workplace. It was established by the American Conference of Governmental Industrial Hygienists and is given in ppm by volume (ppmv) or milligrams per cubic meter of air. Then there is the permissible exposure limit (PEL), which was established by the OSHA. Table 6.5 lists some of the most prominent chemicals, from the extremely dangerous



**Figure 6.13** Chemical concentrations sufficient to immobilize tadpoles, and their lipid/water partition coefficients

**Table 6.5** TLV as time-weighted average for a normal average 8 h work day (source: Yaws 1999)

Name	Mol. formula	MW, g/mol	TLV, ppm volume
Hexachlorobenzene	C <sub>6</sub> Cl <sub>6</sub>	284.8	0.016
Phosphorus	P	30.98	0.02
Iodine	I <sub>2</sub>	253.8	0.1
1,2-Dinitrobenzene	C <sub>6</sub> H <sub>4</sub> N <sub>2</sub> O <sub>4</sub>	168.1	0.15
1,3-Dinitrobenzene	C <sub>6</sub> H <sub>4</sub> N <sub>2</sub> O <sub>4</sub>	168.1	0.15
1,4-Dinitrobenzene	C <sub>6</sub> H <sub>4</sub> N <sub>2</sub> O <sub>4</sub>	168.1	0.15
Chlorine	Cl <sub>2</sub>	70.91	0.5
Benzene	C <sub>6</sub> H <sub>6</sub>	78.1	0.5
Ethanethiol	C <sub>2</sub> H <sub>5</sub> SH	62.1	0.5
Fluorine	F <sub>2</sub>	38	1.0
Methyl bromide	CH <sub>3</sub> Br	94.9	1.0
Sulfuric acid	H <sub>2</sub> SO <sub>4</sub>	98.1	1.8
Sulfur dioxide	SO <sub>2</sub>	64.1	2.0
Nitric acid	HNO <sub>3</sub>	63	2.0
Methyl iodide	CH <sub>3</sub> I	141.9	2.0
1-Hexanol	C <sub>6</sub> H <sub>13</sub> OH	102.18	2.0
Hydrogen bromide	HBr	80.9	3.0
Hydrogen fluoride	HF	20	3.0
Nitrogen dioxide	NO <sub>2</sub>	46	3.0
Hydrocyanic acid	HCN	27	4.7

*continued*

**Table 6.5—Cont'd**

Name	Mol. formula	MW, g/mol	TLV, ppm volume
Ethyl bromide	C <sub>2</sub> H <sub>5</sub> Br	109	5.0
Ethyl amine	C <sub>2</sub> H <sub>5</sub> NH <sub>2</sub>	45.1	5.0
Carbon tetrachloride	CCl <sub>4</sub>	154	5.0
1,2,4-Trichlorobenzene	C <sub>6</sub> H <sub>3</sub> Cl <sub>3</sub>	181.5	5.0
Hydrogen chloride	HCl	36.5	5.0
Hydrogen sulfide	H <sub>2</sub> S	34.1	10.0
1,4-Dichlorobenzene	C <sub>6</sub> H <sub>4</sub> Cl <sub>2</sub>	147.1	10.0
Dichlorofluoromethane	CHCl <sub>2</sub> F	102.9	10.0
Naphthalene	C <sub>10</sub> H <sub>8</sub>	128.2	10.0
Chloroform	CHCl <sub>3</sub>	119.5	10.0
Chlorobenzene	C <sub>6</sub> H <sub>5</sub> Cl	112.6	10.0
Methyl nitrate	CH <sub>3</sub> NO <sub>3</sub>	77	20.0
Nitric oxide	NO	30	25.0
1,2,4-Trimethylbenzene	C <sub>9</sub> H <sub>12</sub>	120.2	25.0
1,2-Dichlorobenzene	C <sub>6</sub> H <sub>4</sub> Cl <sub>2</sub>	147.1	25.0
1,3,5-Trimethylbenzene	C <sub>9</sub> H <sub>12</sub>	120.2	25.0
Carbon monoxide	CO	28	25.0
1,2,3-Trimethylbenzene	C <sub>9</sub> H <sub>12</sub>	120.2	25.0
Ammonia	NH <sub>3</sub>	17	25.0
1-Hexene	C <sub>6</sub> H <sub>12</sub>	84.2	30.0
Toluene	C <sub>7</sub> H <sub>8</sub>	92.1	50.0
Nitrous oxide	N <sub>2</sub> O	44	50.0
Dichloromethane	CH <sub>2</sub> Cl <sub>2</sub>	85	50.0
Chloromethane	CH <sub>3</sub> Cl	50.5	50.0
1,2-Dimethylbenzene	C <sub>8</sub> H <sub>10</sub>	106.2	100.0
1,3-Dimethylbenzene	C <sub>8</sub> H <sub>10</sub>	106.2	100.0
1,4-Dimethylbenzene	C <sub>8</sub> H <sub>10</sub>	106.2	100.0
Nitroethane	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	75.1	100.0
Ethyl chloride	C <sub>2</sub> H <sub>5</sub> Cl	64.5	100.0
2-Methyl-2-propanol	C <sub>4</sub> H <sub>9</sub> OH	74.12	100.0
2-Methyl-1-propanol	C <sub>4</sub> H <sub>9</sub> OH	74.12	100.0
Nonane	C <sub>9</sub> H <sub>20</sub>	128	200.0
Methanol	CH <sub>3</sub> OH	32	200.0
1-Propanol	C <sub>3</sub> H <sub>7</sub> OH	60.1	200.0
Octane	C <sub>8</sub> H <sub>18</sub>	114	300.0
Cyclohexane	C <sub>6</sub> H <sub>12</sub>	84.2	300.0
Heptane	C <sub>7</sub> H <sub>16</sub>	100	400.0
2-Propanol	C <sub>3</sub> H <sub>7</sub> OH	60	400.0
Hexane	C <sub>6</sub> H <sub>14</sub>	86	500.0
Neopentane	C <sub>5</sub> H <sub>12</sub>	72	600.0
Isopentane	C <sub>5</sub> H <sub>12</sub>	72	600.0
<i>n</i> -Pentane	C <sub>5</sub> H <sub>12</sub>	72	600.0
Cyclopentadiene	C <sub>5</sub> H <sub>10</sub>	70.1	600.0
Butane	C <sub>4</sub> H <sub>10</sub>	58	800.0
Ethanol	C <sub>2</sub> H <sub>5</sub> OH	46	1000
Propane	C <sub>3</sub> H <sub>8</sub>	44	1000
Dichlorodifluoromethane	CCl <sub>2</sub> F <sub>2</sub>	121	1000
Chlorodifluoromethane	CHClF <sub>2</sub>	86.5	1000
Carbon dioxide	CO <sub>2</sub>	44	5000

**Table 6.6 TLV values for elements**

Element	TLV	
	ppmv	mg/m <sup>3</sup>
Beryllium	–	0.002
Cadmium	–	0.01
Arsenic	–	0.01
Cobalt	–	0.02
Mercury	–	0.025
Tellurium	–	0.1
Phosphorus (yellow)	0.02	0.1
Selenium	–	0.2
Bromine	0.1	0.66
Iodine	0.1	1
Platinum	–	1
Chlorine	0.5	1.5
Fluorine	1	1.6
Carbon	–	2
Silicon	–	10

hexachlorobenzene at 0.016 ppmv to carbon dioxide at 5000 ppmv (or 0.5 vol.%). All gases are harmful when present in large quantities, and even argon at very high concentrations can cause asphyxiation by excluding oxygen. Other influential guides include the recommended exposure limit (REL) published by the National Institute of Occupation Safety and Health (NIOSH), and the maximum concentration value (MAK) published by the Federal Republic of Germany.

A low value of TLV indicates a more toxic chemical than a high value, so that some authors prefer to work with the reciprocal 1/TLV as a measure of the toxicity. Some of the TLV values for elements, from the most toxic to the relatively benign are listed in table 6.6.

Listed as “asphyxiants” are the much more inert gases of argon, helium, hydrogen, neon, and nitrogen. Many elements are missing from this table, either because they are harmless or because they have not been measured. The values listed in table 6.6 are for the elements, but some of them have compounds that are much more toxic than the parent element. For instance, platinum metal is toxic at 1 mg/m<sup>3</sup>, but soluble platinum salts are toxic at 0.002 mg/m<sup>3</sup>.

#### *Maximum Contaminant Level*

The drinking water recommendations were made in the 1994 National Primary Drinking Standards and Health Advisories of the Office of Water, of the U.S. EPA. The most important parameters are:

- The maximum contaminant level (MCL) and maximum contaminant level goal (MCLG), which are given in mg/L; MCL is enforced and MCLG is not.
- RfD, or reference dose health advisories, are given in mg/(kg day).

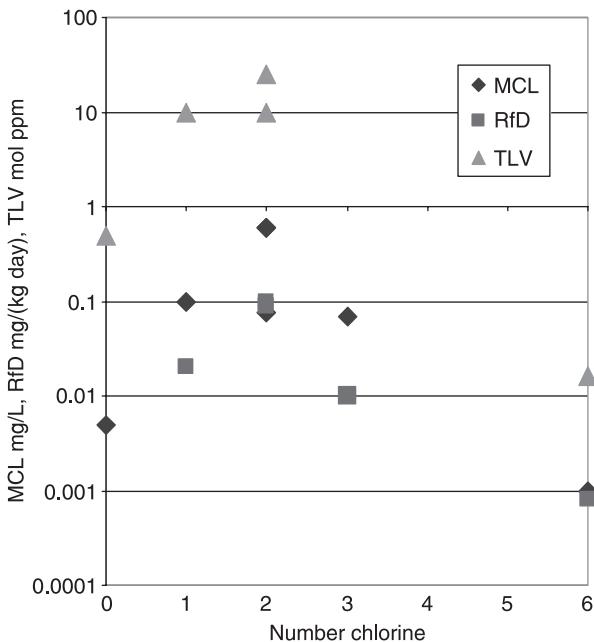
Table 6.7 lists the drinking water standards of a number of substances. Among the chlorinated hydrocarbons, it may be expected that a higher chlorine content would be associated with a greater toxicity. However, this is not always the case. The toxicities of the series of chlorinated benzenes are listed in table 6.8. This is also shown in

**Table 6.7 Drinking water standards: MCL and RfD (source: Casaretti and Doull)**

	<i>MCL, mg/L</i>	<i>RfD, mg/(kg day)</i>
<i>Elements</i>		
P, white		2E-005
Thorium	0.002	7E-005
Mercury	0.002	0.0003
Antimony	0.0006	0.0004
Cadmium	0.005	0.0005
Beryllium	0.004	0.0050
Selenium	0.05	0.0050
Silver		0.0050
Nickel	0.1	0.0200
Barium	2	0.0700
Chlorine	4	0.1000
Zinc		0.3000
Strontium		0.6000
Arsenic	0.05	
Chromium	0.1	
Copper	1.3	
<i>Organics</i>		
Aldrin		0.00003
Dieldrin		0.00005
Heptachlor		0.00005
Chlordane	0.0020	0.00006
CCl <sub>4</sub>		0.00007
Chloroform	0.1000	
Lindane		0.00030
Vinyl chloride		0.00050
Carbon tetrachloride	0.0050	0.00070
Benzene	0.0050	
Hexachlorobenzene		0.00080
Naphthalene		0.00400
Trichlorobenzene	0.0700	0.00600
CHCl <sub>3</sub>		0.01000
Chlorobenzene		0.02000
Malathione		0.02000
Dichlorobenzene	0.6000	0.09000
Styrene		0.20000
Toluene		0.20000
Phenol		0.60000
Xylene		2.00000

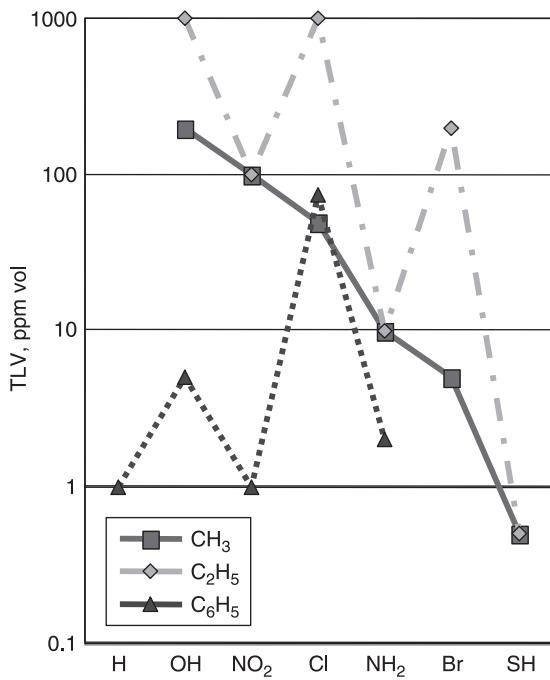
**Table 6.8** Toxicity of chlorinated benzenes

	$N_{Cl}$	$MCL, \text{mg/L}$	$RfD, \text{mg/(kg day)}$	$TLV \text{ mol ppm}$
Benzene	0	0.0050		0.5
Chlorobenzene	1	0.10	0.020	10
<i>o</i> -, <i>m</i> -Dichlorobenzene	2	0.600	0.090	25
<i>p</i> -Dichlorobenzene	2	0.075	0.1	10
Trichlorobenzene	3	0.070	0.006	
Hexachlorobenzene	6	0.001	0.0008	0.016



**Figure 6.14** Toxicity of polychlorinated benzenes, measured as MCL, RfD, and TLV

figure 6.14 as a “volcanic curve,” where the lowest toxicity is found among the dichlorobenzenes in the middle, and the highest toxicity belongs to the two ends of benzene and hexachlorobenzene. The upward curve from zero to two chlorine atoms is contrary to the expectation that more chlorine should not lead to less toxicity; but for the downward curve from two to six chlorine atoms, the expected trend that more chlorine atoms lead to more damage is satisfied. The major damage from these chlorinated benzenes takes place by two separate forms of action: the volatile benzene enters through the lungs, and this leads to bone marrow damage; the much less volatile hexachlorobenzene enters through the mouth and skin, and tends to accumulate in fat tissues. It should be emphasized that these are regulatory values, which can be different between the United States and other nations, and they do change with time as new research results arrive. For instance, the TLV for vinyl chloride was rated as 500 ppm by the OSHA in 1971, until the effect of PVC on liver cancer was discovered in chemical plants in the 1970s; the limit is now set as low as 1 ppm.



**Figure 6.15** TLV of methyl, ethyl, and phenyl groups by substitution groups

The toxicity of the paraffins is altered with the substitution of functional groups; see figure 6.15. The ranking of the toxicity of the methyl derivatives is



However, when we compare the toxicity of the ethyl derivatives, we find that they are generally less toxic, but the ranking among the derivatives does not follow the same rule as the methane derivatives. Benzene is very toxic in its own right, becoming less toxic with the addition of functional groups, and it does not follow the same trend as the methyl or the ethyl derivatives. This graph points to the difficulty of extrapolating the trends found in one series of compounds to another series, even when they are very closely related.

### Environmental Parameters

#### *Ozone Destroying Potential*

Ozone (O<sub>3</sub>) is formed in the tropical stratosphere, around 12 to 30 miles above the ground, where solar radiation is intense; it then migrates to the polar regions. The O<sub>3</sub> concentration can be as high as 10 ppm in the stratosphere; there, it absorbs a large part of the harmful ultraviolet radiation from the sun, thereby protecting life on Earth. CFCs are volatile and persist in the lower atmosphere (the troposphere) because of their inert nature: they resist chemical degradation reactions. It is estimated that

they have a lifetime of 65 to 110 years in the atmosphere. However, in time these CFCs migrate up to the stratosphere, where they break down under the influence of solar radiation to form chlorine radicals,  $\text{Cl}^{\cdot}$ , which have the property of catalyzing the decomposition of ozone back to oxygen. Bromine radicals have even stronger catalytic effects than chlorine radicals, but the fluorine radicals have no such effect. The ozone destroying potential values of chemicals are influenced both by their lifetime in the atmosphere and by their catalytic properties, and are estimated by the World Meteorological Organization.

### *Global Warming Potential of Greenhouse Gases*

The cause of global warming from man-made atmospheric gases also stems from the same factor of atmospheric persistence, as well as their ability to absorb infrared light radiating from Earth to space. Thermal radiation from a radiant body has a distribution of wavelengths, with the maximum value given by Wien's law:

$$\lambda_{\text{max}} \text{ } (\mu\text{m}) = \frac{2898}{T \text{ (K)}}$$

The surface of the sun is at 5800 K, so the maximum radiation is at a wavelength of 0.50  $\mu\text{m}$ , which is in the visible range between red light of 0.7  $\mu\text{m}$  and blue light of 0.4  $\mu\text{m}$ . The average surface temperature of the Earth is about 15 °C (288 K), so the maximum wavelength is 10  $\mu\text{m}$ , which is in the infrared range. Radiation from Earth is partly absorbed by the atmosphere when it contains gases that absorb in the range of 4 to 40  $\mu\text{m}$ , which is infrared with a wavenumber of 2500 to 250  $\text{cm}^{-1}$ . Many gases have vibration frequencies in this range. For instance the following bands ( $\text{cm}^{-1}$ ) are found:

methane:	2940	1350	1320	1270		
benzene:	3080	3030	1965	1815	1480	1035

The global warming potential of a gas depends both on its atmospheric persistence and on the absorption of radiation in this band. So, for a substance not to become a greenhouse gas, it should either break down quickly in the atmosphere, or it should have little or no infrared absorption peaks. Table 6.9 gives the global warming potentials for a number of greenhouse gases, using carbon dioxide as the benchmark; this was reported by the Intergovernmental Panel on Climate Change, which is an organization of the United Nations Environment Program.

There are very few substances in table 6.9 to generate associations and trends. When a substance is being considered for a product, inertness in the atmosphere can indicate a long lifetime, and absorption lines in the range of 0.4 to 0.7  $\mu\text{m}$  would be indications of potential problems.

### References

- Intergovernmental Panel on Climate Change (IPCC). 2001. *Third Assessment Report—Climate Change 2001*. United Nations Climate Program.
- Loomis, T. A. 1978. *Essentials of Toxicology*. Philadelphia, PA: Lea & Febiger.

**Table 6.9 Global warming potentials (GWP) relative to constant future CO<sub>2</sub> concentration (source: IPCC 2001)**

Species	Formula	Lifetime, years	Forcing per unit mass	GWP, 100 years
Carbon dioxide	CO <sub>2</sub>	120	1	1
Methane	CH <sub>4</sub>	12.2	58	21
Nitrous oxide	N <sub>2</sub> O	120	206	310
CFC-11	CFCl <sub>3</sub>	50	3970	4000
CFC-12	CF <sub>2</sub> Cl <sub>2</sub>	102	5750	8500
CFC-113	CF <sub>2</sub> CICFCl <sub>2</sub>	85	3692	5000
CFC-114	CF <sub>2</sub> CICF <sub>2</sub> Cl	300	4685	9300
HCFC-22	CF <sub>2</sub> HCl	13.3	5440	1700
HCFC-141b	CH <sub>3</sub> CFCl <sub>2</sub>	9.4	2898	630
HCFC-142b	CH <sub>3</sub> CF <sub>2</sub> Cl	18.4	4446	2000
HFC-134a	CF <sub>3</sub> CH <sub>2</sub> F	14.6	4129	1300
HFC-32	CH <sub>2</sub> F <sub>2</sub>	5.6	5240	650
H-1301	CF <sub>3</sub> Br	65	4724	5600
Methyl chloroform	CH <sub>3</sub> CCl <sub>3</sub>	5.4	913	110
Carbon tetrachloride	CCl <sub>4</sub>	42	1627	1400

Sunshine, I. 1969. *Handbook of Analytical Toxicology*. Cleveland, OH: Chemical Rubber Co.  
 Yaws, C. L. 1999. *Chemical Properties Handbook*. New York: McGraw-Hill.

## Further Reading

- Albert, A. 1973. *Selective Toxicity: The Physico-Chemical Basis of Therapy*. 5th edition. London: Chapman and Hall.
- Allen, D. T. and D. R. Shonnard. 2002. *Green Engineering: Environmentally Conscious design of Chemical Processes*. Upper Saddle River, NJ: Prentice Hall PTR.
- Goldstein, A., L. Aronow, and S. M. Kalman. 1974. *Principles of Drug Action: The Basis of Pharmacology*. New York: John Wiley.
- Haley, T. J., and W. O. Berndt. 1987. *Handbook of Toxicology*. New York: Hemisphere Publishing.
- Hansch, C. and A. Leo. 1995. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. Washington, DC: American Chemical Society.
- Houghton, J. T., G. J. Jenkins, and J. J. Ephraums, eds. 1990. *Climate Change: The IPCC Scientific Assessment*. Cambridge: Cambridge University Press.
- Klaassen, C. D. 1975. *Casarett & Doull's Toxicology: The Basic Science of Poisons*. 5th edition. New York: McGraw-Hill.
- Korolkovas, A. 1970. *Essentials of Molecular Pharmacology—Background for Drug Design*. New York: Wiley-Interscience.
- Korolkovas, A. 1988. *Essentials of Medicinal Chemistry*. 2nd edition. New York: Wiley-Interscience.

- Lyman, W. J., W. F. Reehl, and D. H. Rosenblatt. 1990. *Handbook of Chemical Property Estimation Methods*. Washington, DC: American Chemical Society.
- Reinhard, M. and A. Drefahl. 1999. *Handbook for Estimating Physicochemical Properties of Organic Compounds*. New York: John Wiley.
- Sax, N. I. and R. J. Lewis. 1987. *Hazardous Chemicals Desk Reference*. New York: van Nostrand.

## Websites

### Public Domain

EPA ECOSAR, ecological structure activity relationships.

<http://www.epa.gov/opptintr/newchems/21ecosar.htm>.

EPA EPI Suite, physical/chemical property and environmental fate estimation models.

<http://www.epa.gov/oppt/exposure/docs/episuite/htm>.

The hazardous substances data bank (HSDB), toxicology data file of the National Library of Medicine. <http://chem.sis.nlm.nih.gov/hsdb>.

Integrated risk information system (IRIS), database of the U.S. EPA.

<http://www.epa.gov/iriswebp/iris/index/html>.

NIOSH Pocket Guide to Chemical Hazards, National Institute of Occupational Safety and Health. <http://www.cdc.gov/niosh/npg/pgdstart.html>.

TOXNET, National Library of Medicine, cluster of databases on toxicology, hazardous chemicals. <http://toxnet.nlm.nih.gov>.

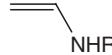
### Requiring Subscription

Knovel, scientific and engineering databases, including Perry's Chemical Engineers Handbook, Chemical Properties Handbook, Lange's Handbook of Chemistry. <http://www.knovel.com/Chemnetbase>, Combined Chemical Dictionary (CCD), Dictionary of Organic Compounds, Dictionary of Inorganic and Organometallic Compounds, CRC Handbook of Chemistry & Physics. <http://www.chemnetbase.com>

## Exercises

1. Estimate the densities, melting points, and boiling points of a number of compounds by association and trend. You may use the properties of any compound from handbooks and electronic sources, except the assigned compound of course.

- (a) The enamines are vinyl amines with the formula  $\text{CH}_2=\text{CHN}(\text{HR})$



Let R=H, so that the compound is  $\text{CH}_2=\text{CHNH}_2$  or ethyleneimine. We need its properties. (Hint. Does this molecule resemble any other molecules with known properties, such as ethylamine, with searchable properties?)

How does this molecule differ from ethylamine, so you can make appropriate adjustments to refine your property estimates?)

- (b) Ethylphosphine has the formula  $\text{CH}_3-\text{CH}_2-\text{PH}_2$ . Estimate its properties, and explain your method of estimation.
- (c) Dimethyl sulfoxide is an important solvent and has pharmaceutical properties. Its formula is  $\text{CH}_3-\text{S}(=\text{O})-\text{CH}_3$ . Estimate its density, boiling point, and melting point.



2. We have limited data on the PCBs, where information is available only on the biphenyls with no more than a few chlorine atoms. We find it necessary to provide answers to pollution arising from PCBs with five to six carbon atoms. How do we use this limited information to construct a plausible estimate of their vapor pressures, water solubilities, and partition coefficients between octanol and water?
3. It is well known that the boiling points of normal paraffins increase with the number of carbon atoms. What patterns can we observe about how the structure arrangements affect the boiling points of paraffins. Here are the data of the boiling points for paraffins with one to seven carbon atoms:

	Carbons	Branches	B.p., K
Methane	1	0	111.66
Ethane	2	0	184.55
Propane	3	0	231.11
<i>n</i> -Butane	4	0	272.65
Isobutane	4	1	261.43
<i>n</i> -Pentane	5	0	309.22
Isopentane	5	1	300.99
Neopentane	5	2	282.65
<i>n</i> -Hexane	6	0	341.88
2-Methylpentane	6	1	333.41
3-Methylpentane	6	1	336.42
2,2-Dimethylbutane	6	2	322.88
2,3-Dimethylbutane	6	2	331.13
<i>n</i> -Heptane	7	0	371.58
3-Ethylpentane	7	1	366.62
2-Methylhexane	7	1	363.20

*continued*

	<i>Carbons</i>	<i>Branches</i>	<i>B.p., K</i>
3-Methylhexane	7	1	365.00
2,2-Dimethylpentane	7	2	352.34
2,3-Dimethylpentane	7	2	362.93
2,4-Dimethylpentane	7	2	353.64
3,3-Dimethylpentane	7	2	359.21
2,2,3-Trimethylbutane	7	3	354.03

- (a) Among the four-carbon paraffins, *n*-butane has a higher boiling point than isobutane (2-methylpropane), which has one branch at the second carbon position. Would this be a trend that prevails through this series, so that the normal paraffin always has higher boiling points than the branched paraffin?
- (b) The attachment position of the branch also seems to make a difference. Among the five-carbon paraffins, 3-methylpentane has a higher boiling point than 2-methylpentane. Does this trend persist at higher carbons?
- (c) Among the five-carbon paraffins, the doubly branched neopentane (2,2-dimethylpropane) has a lower boiling point than the single-branched isopentane. Would this be a persistent trend so that all double-branched paraffins have lower boiling points than single-branched paraffins?
- (d) Among the doubly branched isomers, what is the influence of branch position, and how consistent is that?
- (e) What can you say about the boiling points of triple-branched paraffins?
4. The olefins have two more degrees of freedom in structure than paraffins: the position of the double bond and the *cis/trans* arrangement at each double bond. The data from two to six carbons are given below.

	<i>Carbons</i>	<i>Double double-bond position</i>	<i>Number branch</i>	<i>B.p., K</i>
Ethylene	2	1	0	169.47
Propylene	3	1	0	225.43
1-Butene	4	1	0	266.90
<i>cis</i> -2-Butene	4	2	0	276.87
<i>trans</i> -2-Butene	4	2	0	274.03
Isobutene	4	1	1	266.25
1-Pentene	5	1	0	303.11
<i>cis</i> -2-Pentene	5	2	0	310.08

*continued*

	<i>Carbons</i>	<i>Double double-bond position</i>	<i>Number branch</i>	<i>B.p., K</i>
<i>trans</i> -2-Pentene	5	2	0	309.49
2-Methyl-1-butene	5	1	1	304.30
3-Methyl-1-butene	5	1	1	293.21
2-Methyl-2-butene	5	2	1	311.71
1-Hexene	6	1	0	336.63
<i>cis</i> -2-Hexene	6	2	0	342.03
<i>trans</i> -2-Hexene	6	2	0	341.02
<i>cis</i> -3-Hexene	6	3	0	339.60
<i>trans</i> -3-Hexene	6	3	0	340.24
2-Ethyl-1-butene	6	1	1	337.82
2-Methyl-1-pentene	6	1	1	335.25
3-Methyl-1-pentene	6	1	1	327.33
4-Methyl-1-pentene	6	1	1	327.01
<i>trans</i> -3-Methyl-2-pentene	6	2	1	343.60
2-Methyl-2-pentene	6	2	1	340.45
3-Methyl- <i>cis</i> -2-pentene	6	2	1	340.85
4-Methyl- <i>cis</i> -2-pentene	6	2	1	329.53
4-Methyl- <i>trans</i> -2-pentene	6	2	1	331.75
2,3-Dimethyl-1-butene	6	1	2	328.76
3,3-Dimethyl-1-butene	6	1	2	314.40
2,3-Dimethyl-2-butene	6	2	2	346.35

- Let us start with the four-carbon olefins. What pattern can you detect on the influence of the position of the double bond? Can this pattern be used to predict the trend in the five- and six-carbon olefins?
- Among the four-carbon olefins, what pattern can be seen in the difference between the *cis* and the *trans* structure? Can this pattern be used to predict the trends in the five- and six-carbon olefins?
- Is the trend among the double-branched olefins consistent with the trend among the double-branched paraffins in Exercise (3)?
- Make predictions on olefins with seven carbons, and search the literature for such data to see how good your predictions are.

5. Consider the construction of a new biosphere, which is totally enclosed with soil, water, air, plants, and fish. The new biosphere is 100 ft long by 60 ft wide, by 30 ft high. It contains a pond of 50 ft by 40 ft by 5 ft deep, which is stocked with 100 lbs of fish. Then we spill 1000 kg of benzene in this enclosure near the edge of the pond, and would like to know where it can be found after a while. The enclosure is kept at 25 °C, and we assume that the benzene saturations in air and water reached 20%. The fish is 10% fat.
- (a) What are the concentrations in weight percent of benzene in air, in water, and in the fish? How many grams of the benzene can be found in air, in water, in fish, and in the land?
  - (b) Repeat the calculations for spilling 1000 kg of PCB with three chlorine atoms.
  - (c) Repeat the calculations for spilling 1000 kg or methyl *tert*-butylether.

# 7

## Random Searches

### 7.1 Strategies of Random Searches

- Starting point and promising directions
- Convergent or cluster strategy
- Divergent or wildcat strategy

### 7.2 Combinatorial Chemistry

#### Further Reading

The reverse search starts from a set of desired properties and asks for substances that possess them. Theoretical knowledge and past experience should be relied upon to suggest where to look, since it is the fastest and least expensive approach. When theoretical knowledge and past experience have been exhausted, then random searches may be the only way to make progress, if the problem is sufficiently important and there is enough budget and patience. Table 7.1 compares some of the requirements and the pros and cons of the guided search and the random search

The best strategy on how to spend resources of time and money most efficiently can be considered a problem in operations research, under the topic of “optimal resource allocation.” The best way to use the limited resources of money and time effectively may be a mixed strategy, with some guided and some random searches.

### 7.1 Strategies of Random Searches

Even a random search has to start somewhere. At the beginning, there should be a plan on what territories to cover and how to cover them. The plan can be deterministic, which is completely planned out in advance and executed accordingly. The plan can also be adaptive: after the arrival of each batch of results and preliminary evaluations, the plan would evolve to take advantage of the new information and understanding gained.

**Table 7.1 Comparison of guided and random searches**

<i>Guided search (e.g., freon)</i>	<i>Random search (e.g., taxol)</i>
Depends on a large database and reliable theory	Does not require much database and theory
Requires a few creative experts	Must be carried out by a large enough crew; needs much money, time, and luck
Experts need much education and creativity	Crew does not need very much education and creativity
Concentrates only on a few target areas	Casts a wide net over a large territory
More quick and sure of results, when you know what you are doing and where you are going	Less certain of results, and may be very costly in time and resources
Strong guidance can work as blinders to prevent discoveries of new territories, especially when theory is wrong or not comprehensive enough	Can stumble upon unexpected findings and open brand new territories not related to previous experience or theory

### Starting Point and Promising Directions

Even a random search must begin at a starting point and stake out the most promising directions for initial explorations. In most cases, there is a lead compound that has some of the desired properties, but which is deficient in others, and serves as the starting point of the random search to find better compounds in this neighborhood. The historic cases in section 1.2 involve the modification of an existing product, such as vulcanizing raw rubber and adding an acetyl group to salicylic acid. One explores around the lead compound by using small amounts of additives, blending with other material, changing processing conditions and temperature, and changing structure by chemical reactions. Such explorations lead to the making and discovery of many interesting compounds in this neighborhood, which are collected in a library to serve as the basis to design a range of products for the needs of different customers and purposes. This library and alternate designs can also serve for legal claims of exclusive intellectual property and patent coverage.

Ehrlich knew of several arsenic compounds that attacked the syphilis microbe, but which were too toxic to humans, so he and his assistants synthesized many derivatives and measured their properties. Thomas Edison searched for a fiber that would serve in incandescent lamps. He began with a carbon fiber, but he was not sure which plant material would give the most illumination for the least energy expenditure, and last the longest time. The development of penicillin began with a chance mold that landed on a Petri dish in London, but a strain of mold that gives the highest yield of penicillin was needed, and the worldwide search produced a lasting winner in a moldy cantaloupe in Peoria, Illinois.

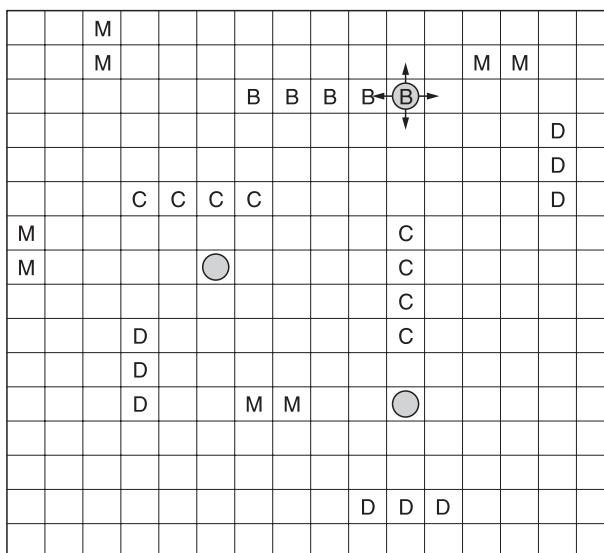
There are also cases where there is not a good lead compound, so that the searchers and their financial backers require a very large motivation to undertake an extensive random search, which covers many territories and will be long and expensive, and ultimately rely on good luck. Such is the case of the search for taxol that began in 1964, which was not confined to a particular plant species or geographic region, but involved the search of numerous categories of natural material. Some 35,000 plant species were collected, extracted, and then tested *in vitro* for their ability to kill cancer cells.

### Convergent or Cluster Strategy

If nature had arranged the best targets in clusters in the exploration region, then the best strategy would be to concentrate on finding a first hit and converge or cluster the subsequent explorations around that first hit. However, if nature had arranged the best targets to be randomly distributed in the exploration region, then one might take the opposite strategy that continues to emphasize random explorations.

Consider the classic board game Battleship, which is played by two players on a board of  $16 \times 16$  squares (figure 7.1). Each player has a fleet of ships: one battleship with a length of five squares, two cruisers with a length of four squares, three destroyers with a length of three squares, and four minesweepers with a length of two squares. Each player hides his board from the other player, and places their 10 ships at various parts of the board, which cover  $1 \times 5 + 2 \times 4 + 3 \times 3 + 4 \times 2 = 30$  squares out of the total of 256 squares. Each player then tries to guess where the opponent's ships are located, and takes turns shooting, say one or three shots per turn. The recipient of the shots tells the shooter which squares have been hit, and then takes a turn at returning the shooting. The objective is to sink all the opponent's ships.

A player has three shots in each turn. In the “exploration phase” at the beginning, the player has no idea where any of the enemy ships are located, and may wish to spread the shots widely, in the hope that one of the shots will make a lucky hit on a ship somewhere. The probability of a hit is  $30/256$  or 12%. The distances between consecutive shots should be random, and large enough to cover a large territory of the board. Once a ship has been located, the game goes into the “concentration phase” to cluster future shots on one of the four adjacent squares around the hit. If the first hit was at the head or tail of a ship, the probability of success at the next attack



**Figure 7.1** The game of Battleship, where the targets are a fleet of 10 ships

rises to 25%; if the first hit was at the middle of a ship, the success probability for the next attack has risen to 50%. After two hits in series have been achieved, the next shot should either be to the head or to the tail of this series, and the probability remains at 50%. Eventually the ship is sunk, and the game returns to the exploration phase, and future shots are spread randomly and widely in the hope of finding other ships.

This is a “heuristic” or opportunistic strategy, of continuing to take available data to revise the strategy and to discover better solutions to the problem. These ad hoc rules are put together for a localized problem, and are probably not applicable elsewhere. In fact, if a better set of rules shows up, then the initial rules should be put aside. This set of strategies is also based on the assumption that the game is about finding a fleet of long and short ships. If the game has been modified to Battleship2, where a player has 30 torpedo boats of single squares to be placed randomly on the board, then there is no advantage to cluster shots around a hit. If the player does not know at the start whether the game is Battleship or Battleship2, then the heuristic strategy may be first to test the hypothesis that your opponent has many long ships, which can be confirmed or refuted later when sufficient data arrive.

The exploration for new oil wells follows a similar set of strategic considerations, where the historic statistical average is about one successful well for 10 wells drilled. The geographic distribution of oil is determined by nature, and is not subject to the arbitrary decisions of a game maker. Veteran oilmen say that major oil fields usually run in packs, such as in the fabulous Persian Gulf with the super-giant oil fields of Burgan in Kuwait, Ghawar in Saudi Arabia, Gach Saran in Iran, and Abu Dhabi; see figure 7.2. Other giant concentrations include Texas–Oklahoma, Venezuela, and western Siberia. The two main types of oil exploration are incremental

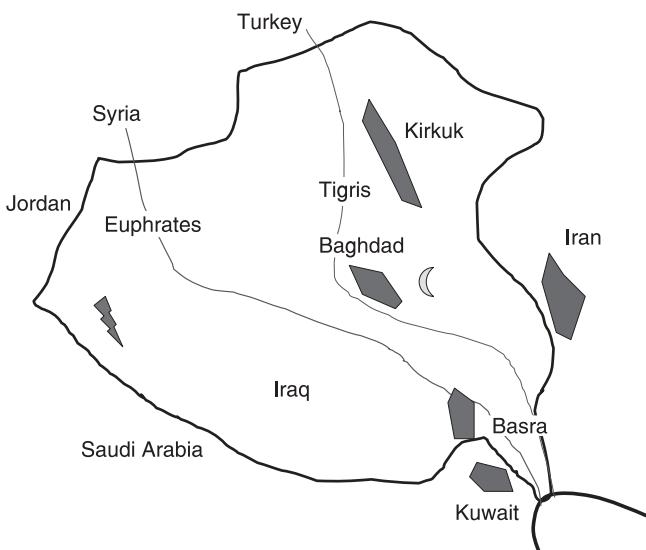
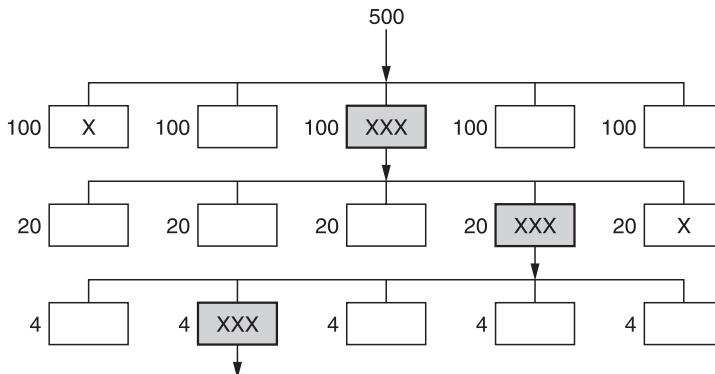


Figure 7.2 Wildcat and incremental oil exploration in the Persian Gulf

wells, which are drilled around the edges of known giant oil fields, and wildcat wells, which are drilled in the wilderness with no previous history of successful wells. The incremental drill is more likely to produce some positive results, but it is an obvious extension of known results that does not bring much glory or profit to the driller, as the cost of the land is already high in anticipation. The wildcat strategy is less likely to produce results, but would bring tremendous fortune and glory to the pioneer who has the courage and conviction to lead the charge in cheap land. Most oil companies play a mixed strategy of drilling many incremental wells balanced with a few wildcat wells.

Modern strategy for drug search resembles the classical Battleship game and oil exploration. A favorite starting point is an existing product or a new discovery that has some of the desired properties but which is deficient in other critical properties. In the drug discovery literature, this starting point is called the “lead compound,” and the subsequent search around it is called “lead optimization.” The next move is to decide which part of the compound is to remain invariant and which parts are to be modified and substituted. When the narcotic properties of morphine were discovered, synthetic organic chemists made a large set of closely related derivatives and tested them to see whether they may have had similar or better properties; this resulted in the incremental drugs codeine and heroin. Thus, a small army of researchers will make incremental explorations around the lead compound, to generate a library of derivative compounds, based on the belief that closely related molecular structures would lead to closely clustered properties.

A form of the cluster strategy is the convergent search. This is illustrated in figure 7.3, where the search region is divided into five families of closely related substances. If, during the first round of random sampling, there are more hits in family 3 than any of the other families, then the next round of random sampling will concentrate in this family, which is further divided into five subfamilies. During the second round of random sampling, if there are more hits in subfamily 4, then future sampling will concentrate on this subfamily.



**Figure 7.3** Convergent search that concentrates future searches on currently most productive regions

### Divergent or Wildcat Strategy

There are times when the opposite strategy is warranted. When the province around the lead compound has been exhaustively explored, which can be compared to the sinking of a battleship or the exhaustion of an oil province, it is time to move on to find other provinces. If nature had arranged the best targets randomly without cluster, then we would be perpetually in the exploration phase of the Battleship game. When an investigator has been excluded from the most promising provinces, due to politics or ownership by other investigators, the only way to participate may be to find a new province. Finally, when the investigators have special theoretical knowledge or experimental skill that is not widely shared, they may concentrate on the fields where they have a special advantage. There is a story about a person who was looking for a lost key under a street lamp at midnight, who said "I am looking here simply because there is light."

Some investigators maintain a deterministic strategy that is not adaptive, by keeping a broad search even when the preliminary results show that some regions are more productive than others. They argue that by concentrating exclusively on initial successes one could give up the opportunity to find new families that are much more important than the current fad. This virgin territory approach is appealing to investigators who want to become either very rich and famous, at the risk of being laughed at, or going broke. The successful practitioners of this craft are highly honored for their courage and vision as pioneers.

The drug discovery people in 1940 thought that they had thoroughly explored the regions around morphine, and had no promising leads to a new or better narcotic painkiller. Then, suddenly, in an unexpected event, methadone was discovered, which has many of the properties of morphine, but has a totally different and unrelated molecular structure. If the drug discoverers had investigated only around the successful hit of morphine, they would have totally missed finding other major classes of compounds that may have very different and useful properties. Wildcatting is one of the most valuable methods of discovering new provinces, which keeps the discovery process from becoming routine and dull.

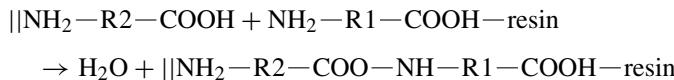
For a research organization that has a budget of time and resources to distribute, it is often prudent to make a balance between incremental and wildcat searches. The former would keep the organization in business by earning a steady stream of small fees, and the latter has the opportunity of making one huge blockbuster. Both approaches are valuable for society, and each organization decides what is the optimal mix.

## 7.2 Combinatorial Chemistry

Combinatorial chemistry is a rapidly expanding field that is particularly prominent in drug discovery and high-performance materials. It carries high hopes, but the number of commercial successes has been modest so far.

Combinatorial chemistry began with the work of Bruce Merrifield in 1963 to develop a solid-phase peptide synthesizer machine, which was recognized by the awarding of the Nobel Prize in Chemistry in 1984. All proteins are linear sequences

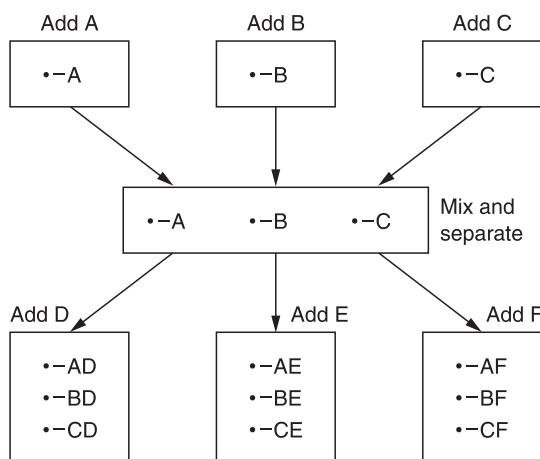
of 20 amino acids, and their synthesis involves numerous reaction steps followed by separation steps, adding one unit at a time with a low efficiency at each step. Thus, a multi-unit protein was prohibitively costly to produce in time and effort. His contribution was to anchor the acid end  $-\text{COOH}$  of the first amino acid on a polystyrene bead and expose the amino end  $-\text{NH}_2$ . Then he added a solution of the second amino acid, which had an exposed  $-\text{COOH}$  end and a protected  $||\text{NH}_2$  end. The two exposed ends of the two amino acids reacted to create a peptide bond between them:



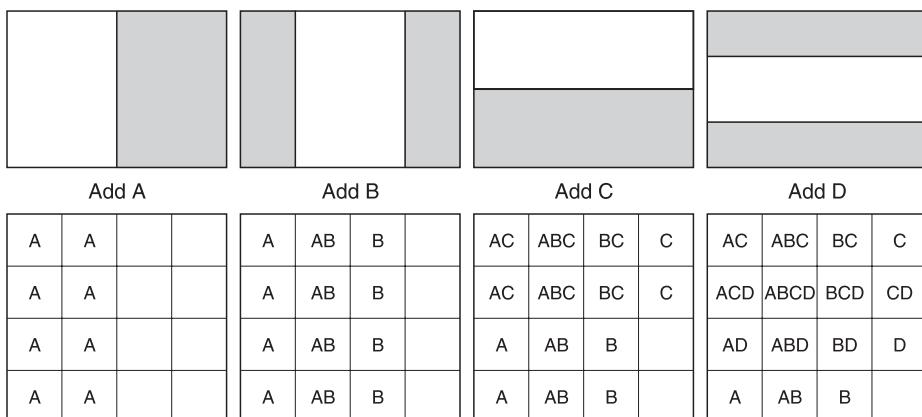
Then he removed the protection on the  $||\text{NH}_2$  end of the dipeptide amino and reacted with a solution of the third amino acid that was protected on the amino end. This sequence of steps was repeated for as many times as needed to make a protein with many amino acid sequences.

Modern combinatorial chemistry involves a number of steps that begin with the creation of a library of molecules that are closely related in structure. The library can be created in two ways: (a) parallel synthesis, which is simultaneous synthesis of numerous products in separate discrete reaction vessels; (b) combinatorial synthesis, of numerous reactions within one single reaction vessel followed by separations. The initial successes in parallel synthesis have been in solid peptide synthesis of proteins, which was based on Merrifield's solid-phase peptide synthesis.

The mix-and-split method was a combinatorial synthesis pioneered by Furka in 1988. The method is illustrated in figure 7.4 with three reaction vessels, each of which contains large quantities of small resin beads. The first vessel is treated with amino acid A to start a peptide, the second vessel with B, and the third vessel with C. Then the beads from the three vessels are mixed together and divided equally into three vessels. In the second generation of processing, the first vessel is treated with amino acid D, the second vessel with E, and the third vessel with F. Now we have a library



**Figure 7.4** The mix-and-split method of combinatorial chemistry



**Figure 7.5** The two-dimensional optical masks

of nine different peptides. Another important advance was pioneered by Fodor in 1991: a two-dimensional array of light-sensitive pins is covered with an optical mask so that only some of the pins are activated when the array is exposed to light; then, an amino acid A becomes attached to the activated pins. Different masks are used in subsequent stages of activation and attachment. Figure 7.5 shows that, after a sequence of four masks, 16 combinations of four amino acids are created.

After a library of molecules has been created, the peptides need to be measured for their properties. This is often done by automated analytical machines. The design of the library, which may have a million items, should be organized for easy positional tracking of the structure and properties of each compound.

### Further Reading

- Czarnik, A. W. and S. H. DeWitt, eds. 1997. *A Practical Guide to Combinatorial Chemistry*. Washington, DC: American Chemical Society.
- Daellenbach, H. G., J. A. George, and D. C. McNickle. 1983. *Introduction to Operations Research Techniques*. Boston, MA: Allyn and Bacon.
- Gordon, E. M. and J. F. Kerwin, eds. 1998. *Combinatorial Chemistry and Molecular Diversity in Drug Discovery*. New York: Wiley-Liss.

## Research Frontiers

### **8.1 Synthesis and Discovery**

### **8.2 Measurement and Discovery**

### **8.3 Understanding of Molecular Structure–Property Relationships**

### **8.4 Prediction of Molecular Structure–Property Relationships**

There are many challenging intellectual opportunities for the research community to create new knowledge in molecular structure–property relations, and to enlarge the toolbox of product engineering, that promise to inspire and accelerate new product innovations.

#### **8.1 Synthesis and Discovery**

There is a suspicion that there are inexhaustible families of material structures waiting to be discovered and to be used. Periodically, this suspicion is confirmed by the announcement of yet another family of interesting materials.

In the 1970s, the synthetic zeolites were heralded as a new class of compounds with a microstructure of pores with molecular sizes. These became very useful in making adsorbents and catalysts. This discovery also led to the invention of many other families of microporous materials. Then came the broader concept of self-assembly of material, namely of organic and inorganic compounds that organize themselves into intricate patterns, now counted within the umbrella of nanotechnology. Two of the recent Nobel Prizes in chemistry offer proof that this field of new synthesis is full of possibilities and honor. We once thought that carbon existed only in the graphite, diamond, and amorphous forms. Now we know that not only can it exist in the buckminsterfullerene form of  $C_{60}$ , but also in many other related forms that are spherical and cylindrical tubules. The notion of an electrically conducting polymer was not seriously considered until the invention of the electrically conducting polyacetylenes by Heeger, McDiarmid, and Shirakawa.

The synthesis of new material can also be divided into the twin paths of incremental synthesis of the derivatives of known structures and the wildcat synthesis of totally unrelated structures. The incremental approach is used when a material with interesting properties is discovered, and research chemists will swarm around it to make every conceivable derivative to see whether they can enlarge the menu to choose from, and to ensure that the new province is well explored. The opposite wildcat approach seeks new and exciting families of material that would not be found by adhering exclusively to known provinces.

## 8.2 Measurement and Discovery

The discovery of interesting natural materials not recognized before is a continuing theme of romance. The field of ethno-botanists features scientists traveling to remote parts of the globe and immersing themselves in local cultures to find native plants and medicines that are worthy of extraction, purification, and testing. Their method includes poking around tropical rainforests, interviewing shamans and nature healers, and listening to stories about miraculous cures of deadly diseases upon touching a toad or swallowing some bark of a tree. They collect the plants described, test their properties in applications, and extract the active ingredients. The *Materia Medica* is an ancient Chinese lore of medicine derived from herbal, animal, and mineral sources, and has been studied to discover the wisdom of the ages that is worth exploiting. Gordon Cragg of the natural product branch of the National Cancer Institute said that “Chemists at the bench could never dream up what nature provides.” Taxol is just one of the more recent examples of this long tradition of finding useful things from nature.

Synthetic chemists have given us a staggering number of compounds that have been made and registered in the main registries, such as the Chemical Abstract Service and the Beilstein; and there are many more compounds that have not been registered. For example, the number of registered compounds in January 2003 numbers 4,418,658 in the Chemical Abstract Service and 9,188,962 in the Beilstein Registry Numbers, and they are growing every day. The technique of combinatorial chemistry has further accelerated the expansion of the number of unregistered materials beyond belief.

The best-known materials are the small molecules of everyday encounter and commercial importance. Even for these compounds, we have measured only a few of the most common properties, such as melting point, boiling point, and density, but very few other properties. There is a very high probability that hidden among them are many materials that have outstanding properties, with the potential to solve many of our application problems. Both acetylsalicylic acid and protosil were extracted and synthesized many years before Hoffman and Domagk discovered their analgesic and antibiotic properties. It is not practical to measure all the properties of all the materials. We usually only make measurements of targeted properties in the neighborhoods of promising materials, which is the equivalent of rounding up the usual suspects; we also need the complementary approach to make random wildcat testing of nontargeted properties in unpromising materials that are not under suspicion. This measurement task takes aim at a moving target, as new molecules are being synthesized every day.

We need the compilation of much large databases, and vetting to eliminate the unreliable data, and to publish databases with more substances with more properties.

This would create a tremendous boost to the work of the product innovators, as well as to the work of researchers who seek to develop theory and correlations between molecular structure and properties. The researchers in product innovations also need better search engines, in the form of databases that are designed and compiled to be reverse searchable, so that one can state a set of desired properties and find a set of materials that have them. This would address the modern Thomas Midgley problem, of finding the set of all compounds that boil between  $-30$  and  $0$   $^{\circ}\text{C}$ , that are nonflammable and nontoxic, that do not harm the stratospheric ozone layer, and that do not cause global warming.

### 8.3 Understanding of Molecular Structure–Property Relationships

To make innovations in products, we need the ability to predict the properties that have not yet been measured, as well as to predict the properties of materials that have never been synthesized. We also need to know which structure modifications would lead to what property changes. Our knowledge and understanding about the relation between structure and their properties is generally not well developed. We do not understand the causes of many of the material properties, so that we are unable to predict how to modify the structures to make them more useful.

Chapter 3 mentioned only a few properties that we really understand and can predict with confidence, such as the heat capacity of simple gases. There are much larger numbers of properties that we can deal with only through empirical correlations with adjustable parameters, such as the boiling points of normal paraffins. There are even larger numbers of properties that we can only explain qualitatively in terms of their structure, such as their melting points. There are many more properties where we do not even have qualitative explanations, such as the sweet taste of sugar and the attractive aroma of musk. These are challenges to the research community, to increase our knowledge and our ability to make predictions.

We are the beneficiaries of the information technology revolution, which makes the storage and retrieval of information vastly more convenient. In the past, we stated a theory about the relation between structure and properties and could not go much further. Computational physics and chemistry have produced many new and powerful tools that enable us to make predictions with reasonable speed and accuracy, without making unacceptable simplifying assumptions. Now, even new students can call on a vast quantity of data to test a theory, to make some predictions, and to compare theory with data. In the early days of this revolution, the emphasis was on calculating structural information such as bond length and energy, bond angle and conformation, ionization potential, etcetera. In recent years, there has been an increase in applications to the properties of condensed matter and on phase equilibria. The advances so far have been swifter in the area of hard condensed matter, concerning solids bonded together by the strong forces of covalent, ionic, and metallic bonds. The advance has been slower in the area soft condensed matter, which concerns molecular solids and liquids that also have the soft bonding of van der Waals forces between molecules. The physical and chemical properties of a many-particle system can be studied by the promising methods of molecular dynamics and by Monte Carlo methods. In contrast to

the availability of general purpose quantum mechanics software, there is not yet such user-friendly software that the product designers and the students can use to predict properties.

Organic molecules are much more complicated than inorganic molecules, and biological molecules are even more complex, as they contain many more atoms bonded together in a hierarchy of structures and substructures. One of the early structure discoveries is the  $\alpha$  helix in proteins by Linus Pauling, which is held together by hydrogen bonds between hydrogen situated between O and N atoms. The most fundamental discovery of modern biology was the double helix of DNA molecules by Crick and Watson, which was inspired by the  $\alpha$  helix. The enzymes are proteins that are the biological catalysts responsible for all biological activities, from producing energy by oxidizing sugar, to turning an embryo into a human being. The enzymatic activity of an enzyme protein is very specifically designed to work on only one or a few substrate molecules, which depends on the binding of the enzyme to the substrate. The binding forces are weak and of short range, so the two molecules bind together only if there is a geometric lock-and-key fit between them. Thus, the prediction of the folding of the protein chain into a three-dimensional object is one of the central problems of molecular biology. The energy of docking two molecules is critical to the success of an enzyme, and this is a very important field of modern research, especially for flexible biological molecules at room or body temperatures.

#### 8.4 Prediction of Molecular Structure-Property Relationships

The most important tool in the arsenal of the product innovators is the ability to make predictions on which structure would lead to what properties, as well as what structure modifications would lead to what property modifications. The reverse research from a given set of properties to material that has these properties is even more important in creating new products and in modifying existing products. In most cases of molecular properties, it is more realistic to depend on empirical correlations between structure and properties. Forward and reverse searches are currently only available for simple physical-chemical properties, such as boiling points and densities; such a facility is still not available for biological properties, such as narcotic and antibiotic activities. The development of such search engines would have a tremendous impact on the productivity of product innovators.

The odor of ambergris, urine, sandalwood, and musk are believed to be related to their distinct structures, and small structural modifications can influence decisively the odor quality and potency. There are three different classes of musk that are not related in structure:

- macrocyclic musk
- aromatic musk
- steroid musk

The existence of these three types of material with totally different structures, and yet the same physiological functions, suggests that it is highly probable that there are some

other families of compounds that also possess the same function. If we have a hint on how these diverse structures give rise to the sensation of musk, we would be in a much better position to find the other families of compounds. In the case of perfume and smell, the objective should perhaps be termed structure-activity relations (SARs), instead of the more lofty title of QSAR.

*This page intentionally left blank*

## **Part III**

### **Product Engineering**

An innovative product represents the creative synthesis of technology capabilities with market needs. The marketers are in charge of marketing aspects, and the product engineers are in charge of technology aspects. Part III is concerned with the tasks of the product engineers in dealing with the outside world, and in designing the products. Chapter 9 describes the methods and limitations of gathering marketing information, and Chapter 10 describes the principal concerns of the public for product safety and environmental impact. Product design is discussed in Chapter 11, which concerns the creative synthesis of one or more products that comprehend all the elements of market needs, technology capabilities, safety and environmental concerns, and profit making for the sponsors. Chapter 12 points out some of the most prominent opportunities for new product designs, either from society needs that are not well satisfied, or from technological successes that are not fully exploited.

*This page intentionally left blank*

# Product Marketing

## **9.1 The Sellers: The Chemical Processing Industries**

Main sellers in the chemical processing industries  
Product prices and volumes

## **9.2 The Buyers**

Market segmentation  
Buyers of consumer products  
Buyers of business products  
Buyers of government products  
Buyers of export products

## **9.3 Marketing Gasoline: A Mature Product**

Customers' needs

Required properties

Market size, growth, and price

Opportunities and threats, incentives for a new fuel

## **9.4 Marketing Ink for Inkjet Printers**

Customers and needs  
Properties required  
Market size and prices  
Opportunities

### **References**

### **Further Reading**

### **Discussion Topics**

### **Exercises**

A marketer should follow the maxim of the 4th century BC strategist Sun-zi, who said “Know self, know opponents, hundred battles, hundred victories.” We are the chemical processing industries (CPI), which is a collection of firms that manufacture and sell a range of products that involve chemistry and employ many chemical engineers. The buyers are consumers, businesses, governments, and foreigners. When we consider selling a product to a buyer, we pay particular attention to profitable and growing markets where our product has a relative advantage over competition. It takes a bold pioneer to introduce a new product that requires the creation of a new market.

## 9.1 The Sellers: The Chemical Processing Industries

Let us study the sellers of chemical products, which are collectively called the CPI. These manufacturers are skilled in the use of chemical reactions and separations to make their products, and they employ many chemical engineers and chemists, often in highly responsible positions. Many of the firms in the CPI are also our suppliers of raw materials and intermediates, our customers for our products, and our competition in making and selling their products.

### Main Sellers in the Chemical Processing Industries

The *Statistical Abstract of the United States* is published annually by the U.S. Census Bureau, which groups all the economic activities in the United States into 11 divisions by the Standard Industrial Classification (SIC). The manufacturing division is divided into 20 sections designated by two-digit numbers. The manufacturers that involve chemistry intensively are listed in table 9.1, by two 2-digit numbers, such as: 20 Food, 28 Chemicals, and 29 Petroleum Refining. The table lists the number of establishments, the number of employees and value of shipment in 1996. The SIC 28,

**Table 9.1 The CPI in the United States**

SIC	Sector	Establishments, $10^3$	Employees, $10^3$	Shipments, $10^9$ \$	Shipments/ employee, $10^3$ \$	Shipments/ establishment, $10^6$ \$
20	Food	20.9	1567	461.3	294	22.07
26	Paper	6.5	631	160.7	255	24.72
28	Chemical	12.4	824	367.4	446	29.63
29	Petroleum	2.1	106	174.3	1644	83.00
30	Rubber	16.6	1018	150.5	148	9.07
31	Leather	1.8	77	9.3	121	5.17
32	Stone	16.4	520	82.4	158	5.02
	All CPI	76.7	4743	1405.9	296	18.33
281	Ind. inorg.		80	27.7	346	
282	Plastic		115	59.6	518	
283	Drug		213	86.5	406	
284	Soap		126	51.8	411	
285	Paint		51	18.3	359	
286	Ind. org.		126	75.7	601	
287	Agr. chem.		37	23.4	632	
289	Misc.		76	24.5	322	
28	All chemical	12.4	824	367.5	446	29.64
36	Electronics	17.1	1582	348.6	220	20.39
3674	Semiconductors	1.1	199	78.5	394	71.36
3691	Storage battery	0.14	23	4.4	191	31.43

“Chemical and Allied Products,” is the basic supplying industry to the other sectors. Table 9.1 also gives the subdivision of SIC 28 into three-digit subsectors, such as: 281 Industrial Inorganics, 283 Drugs, and 286 Industrial Organics. The subsectors of 281 and 286 form the core of the Chemical and Allied Products, as they provide raw material and intermediates for the rest of the subsectors, such as 282 Plastics and 287 Agricultural chemicals. The U.S. Census Bureau is in the process of changing their classification scheme to the North American Industrial Classification System (NAICS), but the historic and many of the current economic data are available only in the SIC scheme.

The major companies in each of these SICs can be found in the Fortune 500 tabulations, which are produced by *Fortune* magazine and are available on the Web. The large food companies of SIC 20 include food processors with names like Sara Lee, Heinz, Kellogg and General Mills. They buy food from the farms and agricultural complexes, and process them into familiar consumer brands of cakes, ketchups, corn flakes, and breakfast food. The largest forest and paper companies of SIC 26 are International Paper, Georgia-Pacific and Weyerhauser, who grow trees in forests and produce lumber for construction and wood pulp for paper. The largest Industrial Chemicals with SIC 281 and 286 are DuPont, Dow, PPG, and Rohm and Haas, which have large research budgets and produce numerous chemical products. The largest pharmaceuticals in SIC 283 are Merck, Johnson and Johnson, Pfizer, and Bristol-Myers Squibb, which have even larger research expenditures and rates of new product innovations. Examples of Household and Personal Products, SIC 284, include Procter and Gamble, Kimberly-Clark, Colgate-Palmolive, who produce soap, cosmetics, and toiletry goods. SIC 29 is Petroleum Refining, with some of the largest corporations on Earth, such as ExxonMobil, Shell, BP, and Chevron-Texaco. They are often vertically integrated companies that start from exploration and production of oil and gas from under the ground, transportation by tankers and pipelines, refining, marketing, and makers of petrochemicals. The major rubber, SIC 30, companies are Goodyear, Goodrich, and Michelin, who make rubber tires, gaskets, and many other elastomers.

In recent years, the electronics industry (SIC 36) has emerged as a major employer of chemical engineering graduates. It has 17,100 establishments, \$348 billion in sales, and 1.58 million employees. The employment of chemical engineers is particularly important in the SIC 3674 Semiconductor Manufacturing, in firms such as Intel, Texas Instruments, and Motorola. The growth of single crystals of silicon from a melt, the etching of silicon surfaces by fluorine, and chemical deposition to lay down thin films all involve chemical reactions and require the attention of chemical engineers. The initial placements of chemical engineering bachelor's degree holders of the last 10 years are given in figure 9.1, and the doctor's degree holders are given in figure 9.2 and table 9.2. It is especially striking that, in the first placement of doctorate recipients, Electronics has surpassed Chemicals, Bio-pharmaceutics, and Fuels as the top employer. The rise of the importance of this industry of Electronic, Photonic, and Recording Materials and Devices to chemical engineers was predicted by the Amundson Report (1988), and this industry should be added to the roster of the CPI.

The products manufactured and marketing by the core industrial chemical firms in SIC 281 and 286 tend to be simple compounds and materials sold to other manufacturers, to be incorporated by further processing into products for the final

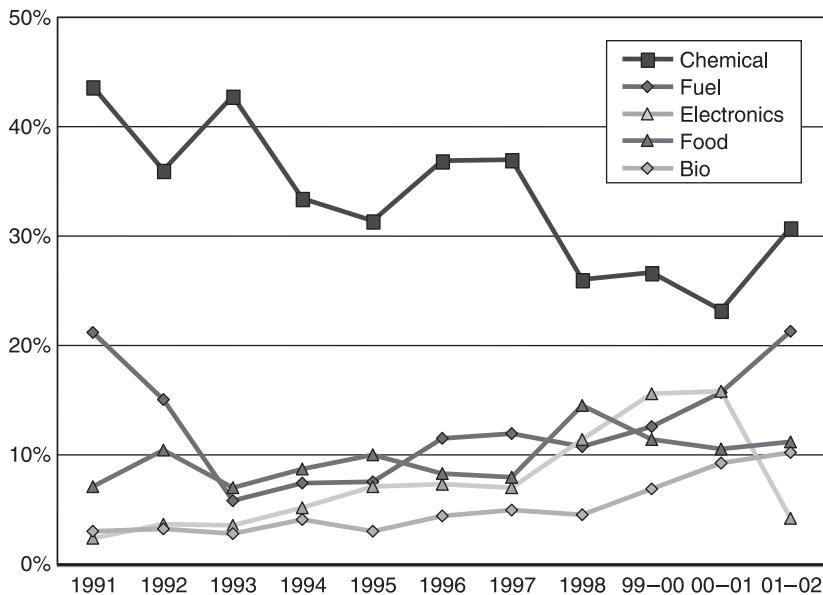


Figure 9.1 First placement of chemical engineering bachelors graduates, 1991–2001

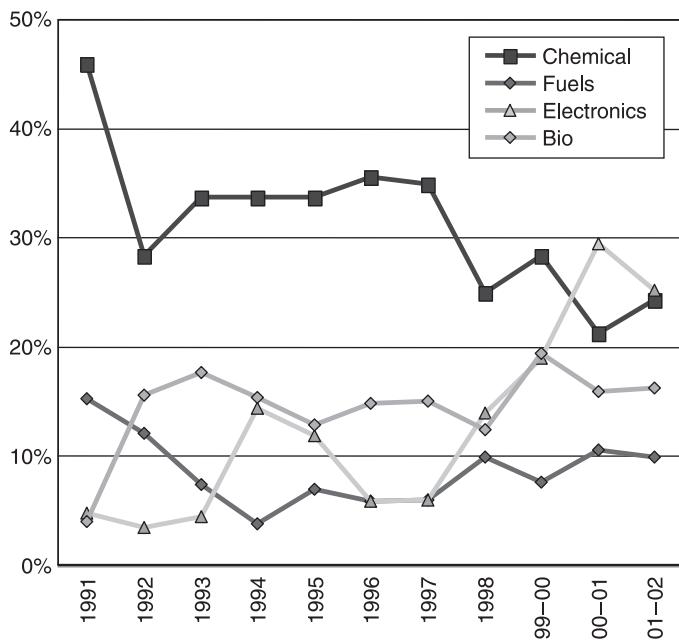


Figure 9.2 First placement of chemical engineering doctorate graduates, 1991–2001

**Table 9.2 First placement of chemical engineering graduates, in percent of graduates (AIChE Survey 2001)**

Industry	SIC	BS, %	PhD, %	Examples
Chemical	281, 286	23	21	DuPont, Dow, Rohm & Haas
Electronics, semiconductors	36, 3674	16	30	Intel, Texas Instruments, Motorola
Fuel	29	16	11	ExxonMobil, Shell, BP, Chevron-Texaco
Services	1629, 871	15	10	Bechtel, Halliburton, Aspen
Food/consumer	20	11	4	Heinz, General Mills, Procter & Gamble, Colgate-Palmolive
Biopharm	283	9	16	Merck, Pfizer, Bristol-Myers Squibb
Others		4	4	Morgan Stanley, Merrill Lynch
Material	30	3	3	Goodyear, Corning, Alcoa
Paper	26	2	2	International Paper, Georgia-Pacific, Weyerhauser

consumers, both individual consumers and federal and local governments. Table 9.3 gives their volume in tons and pounds, as well as their growth rate from 1990 to 2000. Whereas plastics and organic chemicals outputs are still growing faster than the gross national product, the inorganic chemicals and fertilizers are doing less well.

Dow sells vinyl chloride to Goodrich as raw material for polyvinyl chloride, which is later sold to Baxter to make catheters for glucose and blood transfusions, which is sold to your local hospitals. Another example of a chain of intermediate sales is ExxonMobil selling aromatics to DuPont, which makes nylon and sells to WestpointStevens, which makes fabric and sells to Ralph Lauren, which makes garments and sell to Macys, which sells to the final consumers. The most authoritative study of these relations is done by the Input-Output account for the U.S. economy, from the Survey of Current Business published by the Bureau of Economic Analysis of the U.S. Department of Commerce. Table 9.4 lists the percentage of direct sales to consumers in the manufacturing of the main industry groups of the CPI. It varies from a high of 76% for Soap and Toiletries and 66% for Drugs, to 1% for Industrial Chemicals and 0% for the Plastics Industries. The public has much more familiarity with the producers of consumer goods, and appreciates their relevance, and generally has more tolerant relations with them. The public has little contact with the producers of business and government goods, are not familiar with why these products are necessary to their well-being, and are hostile to their reputation for being dangerous and unfriendly to the environment. The public tends to be more patient with lake foam caused by the discharge of their own laundry detergents than with PCB toxicity in the river discharged from electrical transformers used by utility companies.

**Table 9.3 Largest volume products of the CPI**

<i>SIC</i>	<i>Category</i>	<i>Amount</i>	<i>Growth, 1990–2000, %</i>
281	<i>Inorganic Chemicals</i> ( $10^3$ tons)		
	sulfuric acid	43,639	–0.2
	ammonia	16,550	–0.3
	chlorine	13,225	1.2
	phosphoric acid	12,895	0.7
	sodium hydroxide	12,102	0.1
282	<i>Plastics</i> ( $10^6$ lbs)		
	polypropylene	15,439	6.4
	polyethylene, HD	15,423	6.5
	polyvinyl chloride	14,366	4.7
	polyethylene, LLD	9,907	9.8
	polyethylene, LD	7,987	1.0
	polystyrene	6,597	2.8
282	<i>Synthetic Fiber</i> ( $10^6$ lbs)		
	polyester	3,869	1.9
	olefin	3,185	5.8
	nylon	2,607	–0.2
	acetate rayon	349	–3.6
	acrylic	339	–3.9
285	<i>Paint</i> ( $10^6$ gallons)		
	architectural	668	1.8
	original equipment	449	2.9
	special purpose	189	–0.4
286	<i>Organic Chemicals</i> ( $10^6$ lbs)		
	ethylene	55,397	4.3
	propylene	31,825	3.8
	ethylene dichloride	21,850	4.7
	urea	15,322	–0.7
	ethylbenzene	13,156	4.6
	styrene	11,916	4.0
287	<i>Fertilizer</i> ( $10^6$ lbs)		
	ammonia	13,889	–0.4
	nitrogen solution	10,523	0.0
	diammonium phosphate	12,450	–0.6
	phosphoric acid	11,333	0.3
	potassium chloride	1,495	–4.4
30	<i>Synthetic Rubber</i> ( $10^3$ tons)		
	styrene—butadiene	874	1.0
	butadiene	605	3.4

**Table 9.4 Personal consumption as percentage of production output**

<i>SIC</i>	<i>Industry</i>	<i>Personal consumption, 10<sup>9</sup> \$</i>	<i>Total output, 10<sup>9</sup> \$</i>	<i>Direct sales to persons, %</i>
20	Food	201.20	329.6	61.0
26	Paper	11.90	80.9	14.7
281, 286	Industrial chemical	0.98	89.9	1.1
287	Agricultural chemical	0.78	13.3	5.9
282	Plastics		43.4	0.0
283	Drugs	23.90	35.9	66.6
284	Soap	25.00	32.9	76.0
285	Paint	0.19	12.2	1.6
29	Petroleum refining	60.20	137.6	43.8
30	Rubber	11.70	86.8	13.5
32	Glass	1.50	16.3	9.2
32	Stone, clay	2.70	43.3	6.2

### Product Prices and Volumes

Chemical products can be divided into commodities and specialties. A commodity is typically in the autumn or winter of the product life cycle, where there is little product innovation and patent protection, where there are many producers making undifferentiated products in large volumes (say more than 1 billion pounds per year), and selling them at low prices (say less than \$0.50/lb). The commodities tend to be capital intensive and have low profit margins, and the manufacturing companies compete among themselves mainly on their process efficiency, which leads to lower prices and higher reliability. The marketing departments would characterize their products by technical specifications and chemical purity. Typical examples would be gasoline, polyethylene, sodium chloride, ammonia, and others that you can find in table 9.3. The traditional chemical engineering curriculum emphasizes process engineering, which is the best preparation for such careers.

A specialty chemical is often in the spring or summer of the product life cycle, where there is a great deal more product innovation and patent protection, where there are few producers producing differentiated products in small volumes (say less than 1 million pounds per year), and high prices (such as more than \$10/lb). They tend to be research-and-development intensive and have high profit margins, and the companies compete mainly on their innovations in product properties, timely introduction to the marketplace, and performance that increases customer satisfaction. The marketing department would tout the performance of such products, such as resolution and permanence in inkjet printer ink and the effectiveness of curing breast and ovarian cancer in taxol. The chemical engineering curriculum should increase their efforts to prepare students to be useful in specialty chemical companies.

There is a very wide range of prices for products. Consider the examples listed in table 9.5. The listing shows a 10 orders of magnitude difference in price from tap water to diamond. The price per pound of diamond is 10,000 times higher than inkjet ink,

**Table 9.5** Price range for products

Product	Price, \$	Unit	\$/pound
Diamond	1500	2 carat	4,400,000
Platinum	285	troy ounce	7,690
Gold	300	troy ounce	4,370
Chanel No. 5	390	fluid ounce	3,900
Inkjet printer ink	40	40 mL	453
Laser jet toner	100	1 lb 12 oz	57
Extra virgin olive oil	56	Quart	38.50
Whisky	19.95	Quart	22.17
Wine	8.50	750 mL	5.13
Disposable diaper	22.08	96 pieces	3.29
Perrier water	18.00	24 bottles	0.97
Spring water	0.75	Quart	0.36
Gasoline	1.60	Gallon	0.25
Soya bean	5.05	Bushel	0.065
Sand	6.80	Ton	0.0034
Tap water	21.84	1000 ft <sup>3</sup>	0.0004

which is a 1000 times higher than spring water, which is in turn 1000 times higher than tap water. It is well known that when the price goes up, the amount of purchase will go down. This is known as the price elasticity of a product. In general, we anticipate that the volume of sale  $Q$  is determined by the ability of the product to satisfy the needs of the customer, as well as three external monetary factors  $Q = f(P, P'', I)$ . Here,  $Q$  is the number of units sold,  $P$  is the unit price of the product,  $P''$  is the unit price of the competitive product, and  $I$  is the income of the consumer. We expect that the volume of sales will go up when the price goes down, when the price of the competitive product goes up, and when the income of the consumer goes up. The most commonly used model of this function is given by

$$Q = aP^{-b}P''^cI^d$$

where  $a$  is a scaling constant and  $b, c$ , and  $d$  are the elasticity constants.

The definition of price elasticity  $E$  is

$$E = b = -\frac{P}{Q} \frac{dQ}{dP} \quad (9.1)$$

When  $E < 1$ , the consequence of a price increase is a small volume decrease: for instance, when  $P$  goes up 1.0% and the quantity sold  $Q$  goes down by 0.5%. We say that the product is price inelastic, which is the case for an indispensable product without a suitable substitute that one has to buy regardless of cost, or is a product of negligible cost, such as salt. When  $E > 1$ , the consequence of a price increase is a large volume decrease: for instance, when  $P$  goes up by 1.0% and  $Q$  goes down by 2.0%. We say that the product is price elastic, which is the case for discretionary luxury

goods, such as champagne and caviar, and for goods that have ready substitutions. A sudden price increase may catch the consumers unprepared, so that the short-term sales volume remains relatively constant. But, in the long term, the consumers would find suitable substitutes for the product, or learn to do without it, so that the sales would gradually fall and the long-term elasticity would always be larger than the short term elasticity.

The revenue received in the sale of the product is

$$\text{Revenue} = QP = aP^{1-E} \quad (9.2)$$

Thus, when  $E < 1$ , the revenue of a price-inelastic product increases with the price, so it may be advantageous to increase the price. But when  $E > 1$ , the revenue of a price-elastic product decreases with the price, and it may be advantageous to decrease the price. When  $E = 1$ , we have a price-neutral product, where the sales revenue is independent of the price.

The concept of price elasticity of demand is sometimes also applied to a set of products at the same instance of time, such as by making a log-log plot of the prices of the various types of plastic versus their volumes. Figure 9.3 shows the largest volume chemicals produced and sold, divided into categories of minerals, inorganic chemicals, organic chemicals, and plastics. The volume of chemicals output is reported in *Chemical and Engineering News* for 1998. The prices depend on the degree of

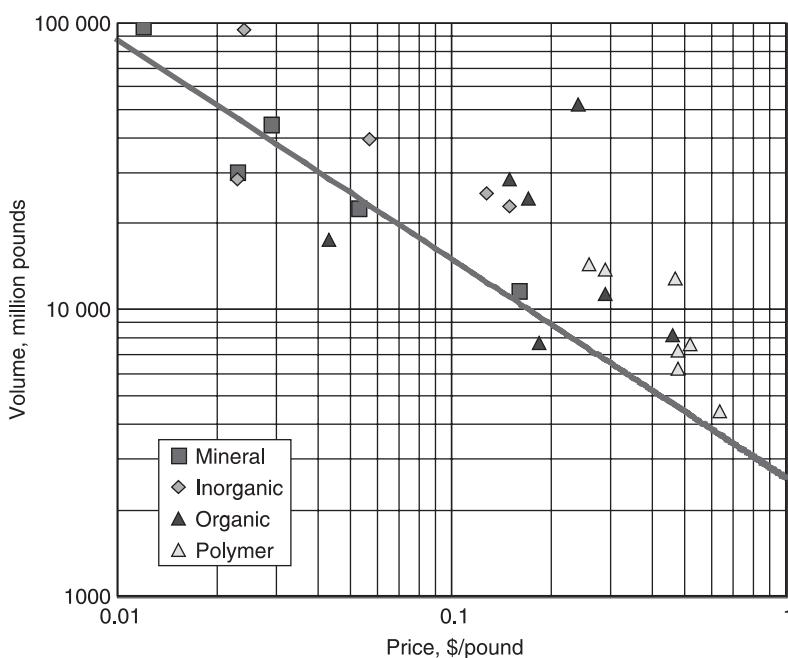


Figure 9.3 Price–volume relations of chemicals

purity, and are obtained from *Chemical Market Reporter*, 24 January 2000. Figure 9.3 shows a scatter with a general downward slope. In this set of data, the value of  $E$  is approximately  $-0.77$ , which is somewhat inelastic.

It is also well known that when the income of the consumer goes up, the purchases will tend to go up as well. This is known as the income elasticity, which can be described by a similar equation:

$$E' = d = \frac{I}{Q} \frac{dQ}{dI}$$

where  $E'$  is the coefficient of income elasticity and  $d$  is a constant. When  $E' < 1$ , the product is income inelastic, so that a 1% increase in price would lead to less than 1% increase in consumption, such as low-quality goods that hold less appeal when your income goes up. When  $E' > 1$ , the product is income elastic, so that a 1% increase in income would lead to more than 1% increase in consumption, such as a swimming pool or a Mercedes sports car, which takes a larger percentage of your budget when your income goes up. Products that are income elastic do better in a booming economy, and products that are income inelastic do better in a falling economy.

The volume of a product is also related to  $P''$ , the price of competitive or synergistic products. The cross-elasticity of volume to a change in the price of a competitive product is

$$E'' = c = \frac{P''}{Q} \frac{dQ}{dP''}$$

The parameter  $c$  is the cross-elasticity constant. One expects that if the competition drops its price, then the sale of our product would go down; so  $c$  is negative if the consumers regard them as substitutes. If Texaco reduces the price of gasoline, then Exxon should expect a reduced gasoline sale unless it follows suit and reduces price as well. However, the sale of Exxon jet fuel would not be affected, as a Boeing 747 cannot use gasoline. On the other hand,  $c$  would be positive if the price of a synergistic product goes down, such as a drop of gasoline price benefiting the business of Disneyland and Yellowstone National Park.

## 9.2 The Buyers

The marketing of a product should begin from an understanding of the customers, their reasons for buying, their characteristics and habits, and their budget. Since the customers have other ways to satisfy their needs, there must also be an assessment of all the competing products in the market, and their strong and weak points.

### Market Segmentation

The main customers of three major sectors of chemical products are summarized in table 9.6 as a percentage of total sales. Drugs are oriented more towards the average

**Table 9.6 Main customers of three major sectors of chemical products**

Buyers	<i>Industrial chemicals, %</i>	<i>Petroleum refining, %</i>	<i>Drugs, %</i>
Consumers	1	41	57
Industry	80	48	28
Government	5	7	8
Export	14	4	7

individual buyers of consumer products, industrial chemicals are oriented more towards the professional buyers of business and government, and petroleum refining has a more balanced customer basis. Export amounts to 4 to 14% of domestic productions of chemicals, and is increasing in importance in a world of growing globalization, with increasing population and wealth in developing countries.

We have to put faces on the buyers before we can understand their needs and their means. The final consumers of goods and services are the individuals and households, and the other three sets of buyers are acting on behalf of the final consumers. DuPont buys oil and gas from Exxon to make fibers, which are sold to textile manufacturers and eventually to the public; the state government of New Jersey buys asphalt and cement to pave roads for the public; and the Asian importers buy perfume and soap for retail stores, and eventually for the public. The sale of a product depends on “understanding the customer’s customers.”

A marketer should understand the great difference between individual buyers and professional buyers. There are 300 million individuals in the United States, who buy in small quantities, have no special technical education, have not studied the products in depth with sophisticated instruments, have little time for comparison shopping, and have very little short-term influence on the sellers (such as on the price and the design of the product). They buy a product brand based on limited information on price and quality; their buying decisions depend on familiarity and past experience, advertising and promotions featuring entertainment and sports celebrities, endorsement by respected friends and family members, and sometimes consumer reports. The sellers reach the consumers by announcements and advertisements in newspapers, from radio and television, from billboards, and from other mass media. When the consumer is not satisfied with the product, the consumer’s complaints will have to compete with those of numerous other consumers. The expression “caveat emptor,” or “let the buyer beware,” shows that information and power usually reside with a giant corporation instead of an average consumer.

On the other hand, the domestic and international buyers of business and government products are hundreds of hard-nosed professional buyers, who buy in large quantities at regular intervals or by long-term contracts, who have access to technical experts with instruments to measure the quality and performance of the products, have the time and motivation to do comparison shopping, and can exert enormous influence on the sellers, including of prices and terms of payment, of product design

and finishing, and in terms of delivery. They also base their buying decisions on price and quality, but they often know more about the quality and performance of the product than the seller. A new product is introduced to buyers by traveling salesmen who visit the buyers, by establishing exhibition booths at new product fairs, and by advertisements in trade journals. When the consumer is not satisfied with the product, the consumer often has tremendous leverage, so that the seller must try very hard to provide satisfactory technical support, to solve consumer problems, and to modify the product. We should use the modified expression "caveat vendor" to describe the relation between a small supplier of gasoline additives and a major oil company as the buyer. Federal and state governments are often the largest buyers of them all, with enormous bargaining power, but they are under the watchful eyes of Congress and watchdogs, and their purchases normally require competitive bids from many vendors.

In order to keep track of the health and purchasing power of the main customers, *Chemical and Engineering News* keeps track of the economic indicators of the nation's economy, as well as a number of industries:

- Gross domestic product
- Personal consumption expenditures
- Crops, acres planted (SIC 01)
- Housing starts (SIC 15)
- Textile mill products (SIC 22)
- Apparel production (SIC 23)
- Furniture and fixtures (SIC 25)
- Paper and products (SIC 26)
- Plastic products (SIC 282)
- Tires product (SIC 30)
- Aluminum production (SIC 33)
- Basic steel and mill products (SIC 33)
- Household appliances (SIC 36)
- Motor vehicle product (SIC 37)

Each firm has a list of their most important customers, which they keep track of with great attention. Sellers of jet fuels must be aware of the economic health of the airlines, which depends on the budgets of the business and leisure travelers, which in turn depends on business profits and individual incomes, as well as of accident rates and wars. Sellers of fertilizers must be aware of droughts and locusts in farming regions, and of public opinion on the safety of genetically modified food.

### Buyers of Consumer Products

There are 300 million Americans, with enormous differences in what they want to buy and what they can afford. A new consumer product should have a primary target aimed at a particular segment of the population, who are more likely to want this product, can afford to pay, and are receptive to new product ideas. Before we tackle the segmented market, let us take a look at the whole population.

### *Demographics of the Buyers*

Every 10 years, the Census Bureau takes a census of the American population, and the past results show:

Census 1980 226,542,199

Census 1990 248,718,301

Census 2000 274,634,000

The natural rate of increase is the difference between the birth rate, which is 14.2 per 1000 population (or annual birth of 3.9 million babies), and the death rate of 8.8 per 1000 population (or 2.4 million people), for a net increase of 5.4 per 1000. But the United States also has a net migration rate of 3.0 per 1000 (or 0.8 million people), so this adds up to an annual increase rate of 8.4 per 1000. In a globalized world, the seller is also concerned with the populations of major world regions, their buying powers, and their needs.

Table 9.7 shows that there were 270 million Americans in 1998, and the number is increasing at a rate of 0.84% each year. The bulge of population of age 30–44 is known as the “baby boomers,” better shown in figure 9.4, as people born after the

**Table 9.7 U.S. population in five-year grouping, 1998**

Age group	All, thousands	Male	Female	Ratio of male/100 female
Total	270,299	132,046	138,253	95.51
<5	18,966	9,696	9,270	104.60
5–9	19,921	10,195	9,726	104.82
10–14	19,242	9,855	9,387	104.99
15–19	19,539	10,046	9,493	105.83
20–24	17,674	8,996	8,678	103.66
25–29	18,588	9,247	9,341	98.99
30–34	20,186	10,007	10,179	98.31
35–39	22,626	11,256	11,370	99.00
40–44	21,894	10,845	11,049	98.15
45–49	18,859	9,252	9,607	96.30
50–54	15,726	7,648	8,078	94.68
55–59	12,407	5,956	6,451	92.33
60–64	10,269	4,849	5,420	89.46
65–69	9,593	4,393	5,200	84.48
70–74	8,802	3,857	4,945	78.00
75–79	7,218	2,997	4,221	71.00
80–84	4,734	1,764	2,970	59.39
85–89	2,556	814	1,742	46.73
90–94	1,117	294	823	35.72
95–99	324	70	254	27.56
>100	57	10	47	21.28
Median	35.2	34.1	36.3	

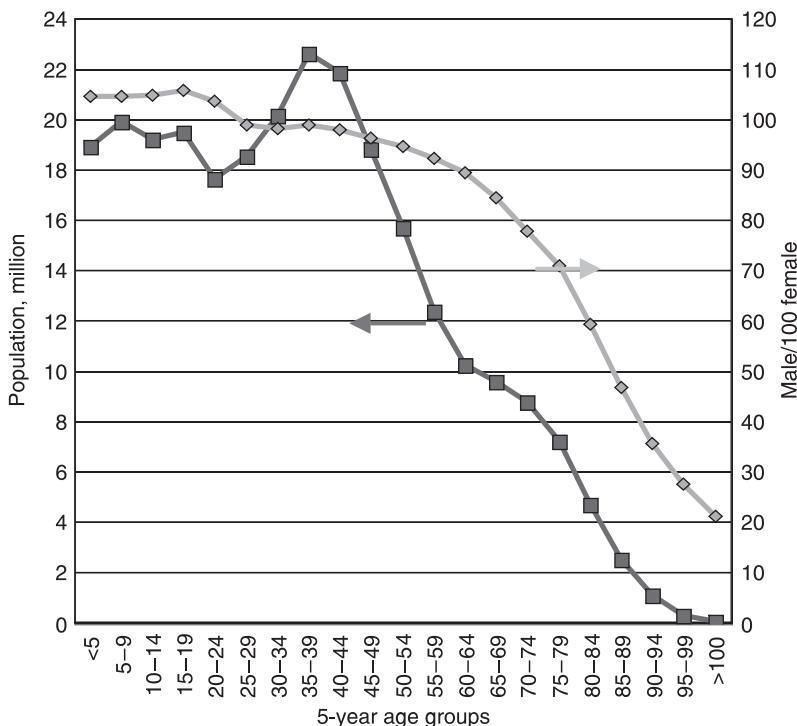


Figure 9.4 The U.S. population 1998 by five-year age groups and male/100 female ratios

end of the World War II. When this boom began, business was good for baby foods and diapers, and the elementary and secondary schools were crowded in the 1960s and 1970s. This was followed in the 1980s by a bust in school age population and the closing of many schools; but in recent years we have another baby boom, as the echo of the children born to the first baby boomers, which created the need to build many new schools.

It is also worth pointing out from table 9.7 that there are significantly fewer men than women in the group of age above 65. There is a marked shift of the male/100 female ratio with age, which is 104 at birth, but continues to drop with age until it becomes 100 at age 25, and 50 at age 85. This is due to the higher death rate of males than females at any age, and particularly in the teenage to young adult age of 15–24, where a male is three times as likely to die, in comparison with a female, from the violent causes of accidents, homicides, and suicides.

The official estimate in 1998 also includes a breakdown by race (table 9.8): Note that Hispanic is a special ethnic designation, and can be of any race. America is becoming less white, as most racial minorities are experiencing population increases faster than the whites. There is also the rise of the multi-racial population in America, which will be included in future population classifications as a separate category. All these characteristics and changes in the population are important, as marketers need to study the consumption preferences of each segment of the population.

**Table 9.8 U.S. population breakdown by race, 1998**

	<i>Thousands</i>	<i>%</i>
Total population	270,299	100.0
White	223,001	82.5
Black	34,431	12.7
American Indian	2,360	0.9
Asian	10,507	3.9
Hispanic	30,250	11.2

For the marketing of each product, the marketers should identify and concentrate on a target set of highest potential buyers. Despite the pejorative tone sometimes given to “profiling,” the marketers must know the different needs, buying habits, and means of buyers from different demographic segments:

- Age, life cycle: young single, married with no children, married with children, empty nest, alone and retired.
- Gender: male, female.
- Income, class, education: inherited wealth, professional and managerial, average middle class, working class, working poor, welfare.
- Religion, race.

Buyers from various geographic segments also have different needs, buying habits and means:

- Regional: Northeast, Midwest, Deep South, Mountain, Pacific Coast.
- Residential density: urban, suburban, country.
- Climate: tropical, temperate, arctic, desert, mountain, beaches.

Let us consider the market for diapers. The U.S. Statistical Abstract, Section 1 Population, gives the following infant and young children population numbers by age for 1998:

Age (years)	< 1	1	2	3	4
Population (10 <sup>3</sup> )	3776	3748	3750	3797	3895

So there are about 11 million babies before diaper training. Consultation with individual mothers, or a consumer panel of mothers, produces an estimate of the number of diaper changes per day; newborn babies use small diapers and have many changes per day, but 3-year-olds use larger diapers and have fewer changes. Let us pick a medium value, such as five changes per day, then multiply by 365 to obtain 1825 diaper changes per year, which translates to 20 billion diaper changes per year for 11 million babies! If a name brand disposable baby diaper sells for 25 cents, then the cost per baby is \$456 per year, which creates a \$5.0 billion per year business. A trip to the local Wal-mart or Kmart will show that the house-brand costs about 15 cents a piece, which is about half the cost of a name brand.

**Table 9.9 “Middle series” projection of resident population**

Year	Population, $10^6$	Age 0–4, $10^3$	Age 65–74, $10^3$	Age $\geq 75$ , $10^3$
1995	260	17,799	18,930	14,834
2000	274	16,898	18,243	16,639
2005	285	16,611	18,410	17,864
2010	297	16,899	21,039	18,323

The future population of this age group has been forecasted. The “middle series” projection of the resident population is shown in table 9.9. So we are anticipating a slight drop in the baby population in the next 10 years. But the aged population above age 65 is expected to be growing faster than the total population, at a rate of slightly more than 1.5% per year. This coming demographic trend is driven by longer life expectations from better nutrition and health care. It is anticipated that baby products do not have as much growth potential as do geriatric products for a growing aging population.

### *Consumer Budget*

Table 9.10 shows the distribution of U.S. household incomes, which limits what each household can afford to buy. Let us consider baby diapers. A mother has a number of options to consider: cloth diapers delivered to the home and retrieved after use, cloth diapers that are home laundered, disposable paper diapers for the garbage can after use, and the rural poor may let the babies run around with bare bottoms. The poor families buy less goods, and need to concentrate purchases on housing and basic food, such as bread and potatoes. The expenditure of \$456 for disposal diapers in a budget of \$75,000 is 0.6%, which is not a serious problem; but for families with less than \$10,000 income per year, this sum becomes 4.5%, which represents a much bigger burden, and less expensive and affordable options would be chosen.

Table 9.11 shows the shopping basket of an average shopper. How does each household construct its shopping basket? According to Abraham Maslow (1987), there

**Table 9.10 Distribution of U.S. household income, 1997**

Income level, $10^3$ \$	Thousand	%
< 10	11,296	11.0
10–15	8,326	8.1
15–25	15,244	14.9
25–35	13,586	13.3
35–50	16,698	16.3
50–75	18,531	18.1
> 75	18,847	18.4
Sum	102,528	100.0

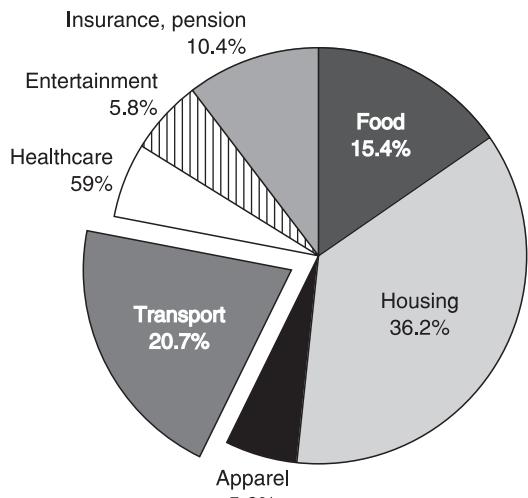
**Table 9.11 Average U.S. household consumption**

	1990	1994	1997	Ratio 1997/1990	Growth, % p.a.
Consumer units, thousands	96,968	102,210	105,576	1.09	0.9
Average expenditure (\$/year)/unit	28,381	31,751	34,819	1.23	2.3
food	4,296	4,411	4,801	1.12	1.2
housing	8,703	10,106	11,272	1.30	2.9
apparel	1,619	1,944	1,729	1.07	0.7
transport	5,120	6,044	6,457	1.26	2.6
car purchase	2,129	2,725	2,736	1.29	2.8
gasoline, oil	1,047	986	1,098	1.05	0.5
other car cost			2,230		
public transport			393		
health care	1,480	1,755	1,841	1.24	2.5
entertainment	1,422	1,567	1,813	1.27	2.7
insurance, pension	2,593	2,957	3,223	1.24	2.4

Note: "other car cost" include finance, maintenance, insurance, license, parking.

is a "hierarchy of needs," and the most basic human need is physiological survival through food and drink. When people have sufficient income to satisfy these needs, they go on to the next level of comfort with shelter and clothing. When people acquire more means, often the next needs are love and friendship, to be followed by amusement and knowledge. For the more successful and affluent people, the next needs may be security and liberty, culture and refinement, esteem and fame, and then influence and power. For a few rare individuals, fulfillment of a mission, such as saving the world from ruin or rescuing souls from hell, would be the highest level on the hierarchy. The Maslow scheme would predict that people will put first priority of their resources for food, and then for shelter and clothing. Actually, there is nothing more precious to life than air, but it is free. Food and drink may be the most basic to survival, but, for the average-income American family, the food bill (15%) is less than half of the housing bill (36%), and less than the transport bill (21%); see figure 9.5. This is a matter of supply and demand, represent nature's relative abundance, the superior productivity of the food industry, and consumer preferences. An intelligent and well-informed consumer has a budget based on his or her income, and switches the last dollar of purchases between housing to food, according to what seems to give the greater satisfaction or utility. An economist will say that an equilibrium is reached when the marginal utility (incremental satisfaction) of each purchase is the same as the marginal utility of any other purchase, so there is no point in switching a dollar from one purchase to another. Some consumers prefer more food dollars, and some consumers prefer more housing dollars, but the collective actions of one hundred million households determine the equilibrium for the nation.

The purchasing budget of a household depend on national economic and social conditions, such as the increase in average personal income, the unemployment rate,



**Figure 9.5** The U.S. personal consumptions by categories, 1998

the Dow Jones stock Index, and the Consumer Confidence Level. The purchase of expensive capital goods such as housing and cars, and new plants and equipments, are also sensitive to the interest rate involved in borrowing money.

### Buyers of Business Products

Table 9.12 lists the main customers of the three major segments of the CPI. The biggest customers of Industrial Chemicals are other Industrial Chemicals companies, which buy more than 20% of all the products, and the next biggest industrial customer is Plastics, which buy more than 15%. Export is a major source of sales at more than 16%. The various sectors of the CPI, such as Rubber, Soap, and Paper, are among the best customers of Industrial Chemicals. The federal and state governments buy 4.5%, but direct consumption by individuals is minimal at 1.0%.

On the other hand, Drug-sector companies have personal consumption as the largest buyer, at 66% of sales, and Health Services (hospitals and clinics), at 17.8%, as the second largest. Selling to other drug companies accounts for 10.6%; and governments and export buy 10.3% and 8.1%, respectively. In many other industries, such as Soap, the buying decision is made by a buyer who is often also a consumer. In the Pharmaceutical Industry, however, the consumers are the patients; so, the decision makers are the physicians who decide on what drugs should be bought, whether they should be generic or name brand, and how they should be used. A growing influence on drug buying decisions is the presence of the federal Medicaid and private health insurance companies, who have enormous influence on what drugs they will pay for and how much to pay. Despite the high percentage of sales to individuals, the marketing and advertising plans of pharmaceutical companies concentrate on the physicians, keeping them informed and persuaded that their products are the most effective and least costly. They also need continuing relations with the health maintenance organizations and the social security agencies.

Table 9.12 Major customers of three industries

Buyer	SIC	Percentage to customer		
		Industrial chemicals	Drug	Petroleum refining
Farming	1	0.1	0.6	0.2
Construction	15	1.4		8.1
Food	20	1.7		0.3
Paper	26	3.8		0.4
Industrial chem.	281	20.2		0.7
Agriculture chem.	287	1.6		
Plastics	282	15.0		0.1
Drug	283	0.8	10.6	
Soap	284	3.3		
Paint	285	2.6		
Petroleum refining	29	2.0		7.2
Rubber	30	4.3		
Glass	32	1.1		
Stone, clay	32	1.4		
Prim. iron and steel	33	1.9		
Local transport	41			1.5
Motor transport	42			4.2
Water transport	44			0.5
Air transport	45			6.3
Pipeline	46			
Utility	49			2.3
Trade	50			3.5
Auto repair	75			2.7
Health service	80		17.8	
Government	91	1.2		3.5
<i>All intermediate sales</i>		88.5	32.6	51.2
Personal		1.1	66.6	43.8
Export		16.2	8.1	4.4
Import		-11.9	-21.2	-9.7
Federal government		2.1	2.2	2.3
State local government		2.4	8.1	5.7
<i>All net final sales</i>		11.5	67.1	48.8
Total output		100.0	100.0	100.0

The petroleum refining companies have an intermediate position, since 44% of the products are sold directly to individuals, as gasoline and lube oil at gasoline stations. The other main customers are the four transportation industries taking 12% of products as jet fuel for airplanes, bunker oil for ships, and kerosene for buses and trucks. The Utility-sector buys heavy fuel oil to generate heat and electricity. The heavy sale to the Construction industry is both for fuel and for asphalt for paving roads and for roofs.

It would be useful to look deeper into individual classes of compounds, and learn about the buyers and how they use these compounds. A source of useful information can be found in a reference such as the *Kirk-Othmer Encyclopedia of Chemical Technology* (2001). The main consumer of phosphate rocks is the fertilizer industry, and they in turn sell to the farms and the large agro-business firms. The marketer keeps in touch with the farmers and their fortunes: is there a drought in the Midwest breadbasket, a freeze in the Florida orange groves, a crop failure in North Korea, or south of the Sahara, and thus a demand for the export of U.S. food, is a significant fraction of food grown in the United States exported abroad, and is that increasing, is the U.S. government continuing to subsidize surplus grain by selling abroad at a great discount, or giving them away? Who buys lime, and for what purpose? Since so many people are buying a lot of sulfuric acid over a long period of time, it must be very useful to them. What can possibly be the utility of something that is so corrosive and dangerous? Who buys ammonia, fibers, and rubber? Are they prospering?

The main manufacturers and sellers of zeolite molecular sieves are UOP, Grace, and PQ. According to *Kirk-Othmer*, the sales in 1991 in the United States were as follows:

Use	Volume (tons)	Price
Detergent builders	182,000	\$0.60/kg
Catalytic reaction	86,000	\$0.85
Adsorbent-separation	30,000	\$2.29

The biggest detergent companies are Procter & Gamble, Colgate-Palmolive, and others in SIC 284. The main purpose of buying zeolite is as a replacement for phosphates in detergents, which are added in laundry detergent powers needed in hard-water areas, to complex with the calcium and magnesium ions and free the soap to function. The phosphates have been blamed for much of the eutrophication, as they fertilize once sterile lakes and ponds to create algal blooms, creating bad smells and ecological ills. A sodium form of zeolite A can form the ion-exchange function of replacing sodium with calcium and magnesium in the wash, which is an alternate way to make the detergent function. When Procter & Gamble shop for a supplier of zeolites, the sellers need to know who are the buyers and what are their educational backgrounds and work experience, and what laboratory measuring instruments do they have. The sellers need to know what types of product specifications would impress them, how to bring these specifications back to the office, and how to help the design team in coming up with new products that would impress these buyers. Our sellers also need to know what are the competitive products produced by other firms, and what are their strong and weak points, so that our sellers can point out the advantages of our product.

The biggest buyers of catalytic cracking zeolites are the oil refiners, ExxonMobil, ChevronTexaco, Shell, and BP. Zeolites are also used to dewax lube oil by removing the normal paraffin, and in producing aromatic compounds such as *p*-xylene. The buyers and the research departments behind them are some of the most highly educated chemists and chemical engineers in the country, and have written award-winning papers and books. The seller must approach them only after thorough homework and

preparation to answer tough questions. The seller needs to know what are the objectives of the buyers, which is the right zeolite for the purpose they have in mind, what are the age and condition of the catalytic crackers of the customer, what are the latest advances in reaction engineering and catalysis, what is the current trend of crude oil feedstock, what are the marketing shifts in refinery output requirements, whether your catalyst can squeeze out 1 or 2% more isoparaffin than competitive products, and whether your catalyst will decrease the benzene yield and help in environmental regulations.

Who are the biggest buyers of adsorbents and desiccants, and why do they need them? Some customers are interested in purification by removing trace contaminants from a stream of gas or liquid, such as the removal of hydrogen sulfide and mercaptans from natural gas, and the removal of organic pollutants from water. Some customers are interested in bulk separation of two or more commercially valuable streams of products, such as the separation of linear paraffins (biodegradable) from branched and cyclic isomers (not biodegradable) in the manufacture of detergents, or the separation of *p*-xylene, which is the raw material for making terephthalic acid as the precursor to polyester fibers, from a mixture of C8 aromatic compounds. Our seller needs to know how our product compares with that of the main competitors, and in what properties and performance metrics our product is superior, and where it is inferior.

### Buyers of Government Products

The federal government buys many products related to national defense and security, health and medicine, transportation, and environmental protection. The state and local governments buy products related to highways and bridges, police and fire, sewage and solid waste, and in elementary and secondary schools. They buy a range of products related to these functions. The government buyers are also professional buyers, but have much lower income than the business goods buyers. They are constantly in the public eye, and they need to have an open and transparent process so that the normal method of awarding a purchase contract is through open bidding among many suppliers.

Let us consider the fuel of the Tomahawk missile that provided the main punch for the first week of the Iraq War of 2003. It was launched by the air force from an airplane, or by the navy from a submarine, a missile destroyer or cruiser. It is 20 ft long, and has a fuel tank of 150 gallons packed tightly in the fuselage between the bomb in front and the jet engine behind. It has a normal range of 690 miles, when it is loaded with the fuel of convention JP-4 or JP-5 hydrocarbon. Now the goal is to reach Baghdad from a safe distance that is in no danger of being retaliated against. The shortest distance from the Persian Gulf is 400 miles, the shortest distance from the Mediterranean Sea is 600 miles, and the shortest distance from the Red Sea is 700 miles. Is it possible to replace the \$5.00/gallon of JP-4 by an exotic \$50.00/gallon fuel that will stretch the range to 1000 miles?

There are really two separate markets here, as the navy specifies a flash point of 60 °C, but the air force specifies 38 °C. The buyer is a thoroughly professional team of engineers, scientists, accountants, and managers. They also have the backing of a very large research and development organization, and numerous consultants from

universities and major weapons manufacturers. They operate test facilities such as the Naval Air Warfare Center in China Lake, California. The major sellers of jet fuel, the oil companies such as ExxonMobil, do not have the full-scale missiles and test ranges to prove the effectiveness of a new jet fuel. The development and testing of a new fuel has to be managed as a joint effort, of small-scale laboratory testing at the oil company, and full-scale field testing at China Lake. The volume and growth of the market in the navy and air force depend on orders from the Secretary of Defense and the mood of Congress, which in turn depends on world tension and the anticipated missions.

### Buyers of Export Products

The best targets for export are goods in which we have a cost and availability advantage, and the customers are prosperous and growing in demand. In international trade, the United States as a whole exported \$780.4 billion worth of goods and imported \$1,216.9 billion, for a deficit of \$436.5 billion. Export represents from 4 to 14% of the chemicals produced in the United States, and these percentages are generally expected to keep on increasing as the world becomes more globalized in the future.

The U.S. chemical industry had a very favorable balance of trade for many years, earning some \$20 billion a year during the first half of the 1990s, but this surplus had turned into a deficit of \$3.2 billion by 2002; see figure 9.6. We still enjoy a healthy surplus of exports over imports with Canada, Latin America, Australia, Africa, and Asia; but we run a large deficit with Europe, particularly with Germany, Netherlands, France, and the United Kingdom, as shown in table 9.13 and in figure 9.7.

We are particularly weak in international competition on organic chemicals and medicine. See table 9.14 for 2004 data. This reflects a number of factors, including

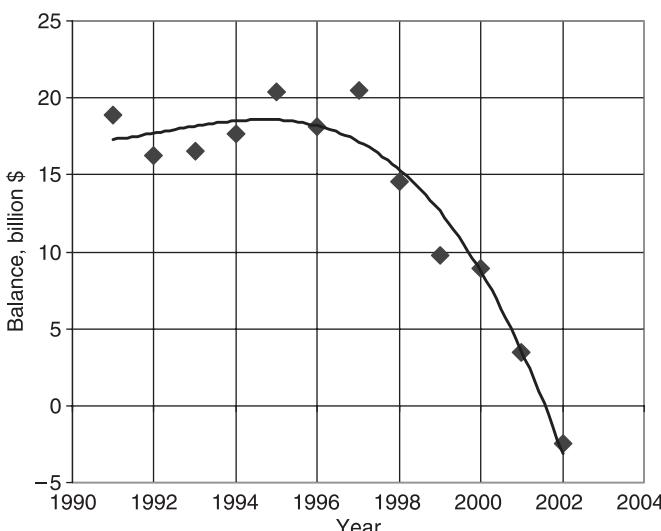
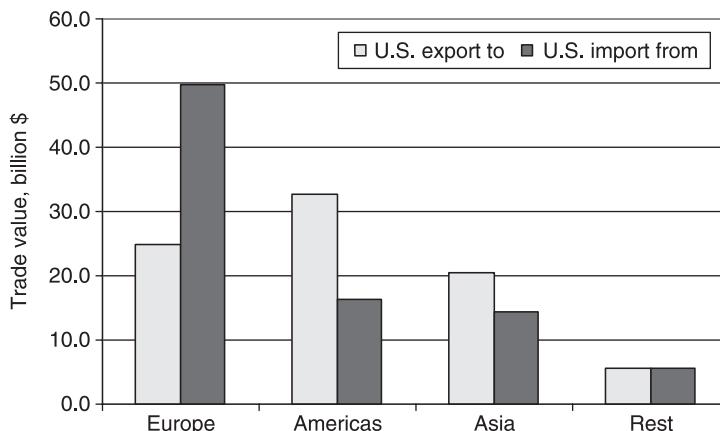


Figure 9.6 U.S. chemical trade balance, 1991–2002

**Table 9.13 U.S. trade figures with the world**

Geographic region	U.S. export to, 10 <sup>9</sup> \$	U.S. import from, 10 <sup>9</sup> \$	Balance, 10 <sup>9</sup> \$
Europe	24.8	49.8	-25.0
Americas	32.8	16.4	16.4
Asia	20.4	14.4	6.0
Rest of the World	5.5	5.5	0.0
Sum	83.5	86.1	-2.6

**Figure 9.7** World chemical trade by regions**Table 9.14 U.S. trade by product type with the world**

Product type	US export, 10 <sup>9</sup> \$	US import, 10 <sup>9</sup> \$	Export-import, 10 <sup>9</sup> \$
Organic chemicals	25.0	35.5	-10.5
Plastics	24.8	14.1	10.7
Medicine	23.4	34.7	-11.3
Inorganic chemicals	5.9	7.9	-2.0
Dye	4.6	2.7	1.9
Cosmetics	7.4	6.8	0.6
Fertilizer	2.6	2.4	0.2

the strength of the glorious history of fundamental and applied chemistry research in western Europe, which has been maintained in recent years. The bright American export areas are plastics and dyes; plastics are made from petrochemicals, where the United States used to enjoy a cost advantage compared with Europe and Japan due to our domestic oil and gas production, but this is now largely gone. The raw material cost advantage now belongs to the Persian Gulf and other oil-producing nations, classified here as Rest of the World.

The fastest growing economies in the world today are in the Far East, and particularly in China, which is widely expected to become the growth engine of the world. China is now the world's largest importer of polyolefins and a major importer of advanced specialty chemicals. The polyolefins consist of low-pressure linear low-density polyethylene (LLDPE), low-density polyethylene (LDPE), high-pressure polyethylene (HDPE), polypropylene (PP), and higher polymers of higher olefins, such as poly(1-butene). They are used mainly for pipes, toys, filaments, films, molding, and cable resins. The raw materials are olefins that are largely derived from cracking of hydrocarbons in petrochemical plants, which should be conveniently located for easy access from oil and gas fields by pipelines or tankers. Such plants are very capital intensive, as they need to be very large to have economy of scale. At this moment, the United States has advantages over China in the production of polyolefins due to the past investments in plants and access to raw material. This market is already crowded with U.S. and European sellers, so that it is a big and growing market, but there are many competitors.

The buyers from China are either large chemical companies or traders representing many smaller chemical companies. They have similar characteristics to domestic business buyers, but they speak a different language and have different customs and procedures. The sellers with foreign language abilities and past experiences as students, travelers, or residents definitely have an advantage in understanding the buyers.

### 9.3 Marketing Gasoline: A Mature Product

Gasoline is a mature product that is familiar to most people. It is sold in enormous quantities each year in the United States, and at relatively low prices in comparison with bottled drinking water. It is a very mature product that has been important for a century, since the rise of the gasoline buggy, and has been improved and reformulated numerous times in the past. There have been few product innovations in recent years that would enable a company to capture a significant market share from other gasoline manufacturers. There has also been little new science and technology on the horizon waiting to be exploited that has the potential to increase the quality and to decrease the cost of gasoline. However, there are threats to gasoline from environmental and safety problems, and from increasing government regulations, which can perhaps be turned into opportunities for innovations.

The study of gasoline is facilitated by the easy availability of market information. Let us consider the initial planning process for a new gasoline. A marketer should guard against what is called the "marketing myopia" by Theodore Levitt (1975), which is to think too narrowly about the business. The customers of gasoline are really interested in transportation fuel, and not necessarily in gasoline. An oil company that is single minded on marketing gasoline can become vulnerable to a competitor coming up with a new type of transportation fuel that is even more satisfactory to the consumer. Gasoline marketers need to think of themselves as not simply in the gasoline business, but as a part of the bigger transportation fuel business.

## Customers' Needs

Let us consider the market as transportation fuel for light vehicles, characterized by small engines and light fuels, and the buyers are individual owners and drivers, as well as business and government units that own motor vehicles. Most Americans depend on private cars to go from home to school or work, to shop for groceries and clothing, to visit doctors, and to go to baseball games, to go on dates, and to attend weddings, to visit grandmothers and to visit Disneyland. The U.S. Census (1990) showed that, of the 92 million households surveyed, only 11% had no cars, 34% had one car, and 55% had two or more cars. In a major city with a good system of public transportation, such as New York City, with a dense network of buses and subways, a person can get around without owning a car, but they will often hail a taxi. Venice is an exceptional city that has a network of canals that is traveled by gondolas, but the motorboats are usually powered by gasoline. In the American suburbs and rural areas, where the population is sparsely distributed, the only practical ways to get around are walking, bicycling, riding, and being pulled by horses and other animals—and the motor vehicle.

In the year 1998, there were 270.3 million Americans and 182 million registered driver's licenses (*Statistical Abstract of the United States* 2001). There were 137 million cars, 3.8 million motorcycles, 77 million trucks and 698,000 buses. Now a car needs a source of energy in order to move. One such source of energy is gasoline, of which Americans consume 72 billion gallons, so that an average car consumes 21.4 miles/gallon and 548 gallons/year. The power to move an automobile can, in principle, be provided by the environment, such as solar and wind energies, which are feeble and not dependable. It can also be provided by a connection to a central source, such as by an overhead electrical cable for a trolley or a live third rail for a subway, but that requires a large capital investment and makes the route quite inflexible. An automobile is nearly always provided by a source of stored energy. The forms of stored energy include: the mechanical energy of a coiled spring, or a rotating flywheel with a large moment of inertia; the stored electrical energy of a battery; and finally the stored chemical energy in fuels, such as gasoline and diesel fuel.

Since a car has limited space and weight-carrying capacity, it is important that the energy is stored compactly; it is also important that the energy can be released quickly, for rapid acceleration, such as when passing another car on the highway or for climbing a hill under heavy load. The efficiency of a stored energy can be measured volumetrically in watt-hours per cubic meter, or gravimetrically in watt-hours per kilogram. The rate of energy release can be measured in watts. A high storage efficiency allows a car to go for a long distance, such as 400 miles, without stopping to refuel. Except for nuclear energy for submarines, the most compact storage capacity combined with a rapid release rate is found in the chemical fuels that burn in air. A hydrocarbon such as octane has a combustion energy of 11.6 kcal/g and a density of around 0.7 g/ml, which works out to 8.1 kcal/mL. When converted to the units of watt-hours, gasoline has energy densities of 9400 W h/L and 13,400 W h/kg. Oxygen required by the combustion is provided by the air, and does not need to be stored in the car. This enables a 20-gallon tank of gasoline to propel a car 400 miles before the need to refuel. The oxygenated fuels, such as methanol and ethanol, are usually used

in mixtures, such as M85 (85% methanol and 15% gasoline) or E85 (85% ethanol and 15% gasoline). Oils of vegetable (such as peanut) and animal (such as whale) origin have been considered for motor fuels, but have the problems of inadequate supply and high cost. The oxygenated fuels can be considered chemically as partially burned hydrocarbons, and thus have much lower heats of combustion. A lead-acid battery has energy densities of about 100 W h/L and 65 W h/kg, and a lithium battery has densities that may be four times higher. The batteries would have to be a hundred times better before it can challenge gasoline in energy density.

Alternate fuel cars are being studied and subsidized intensively. In the year 1999, the United States employed the following number of alternate fuel cars: liquefied petroleum gas (LPG), 274,000; compressed natural gas (CNG), 96,017; M85, 21,829; E85, 17,892; electricity, 6481; liquefied natural gas (LNG), 1517; neat methanol, 378; E95, 14. A convenient fuel should be a liquid for easy pumping at the outdoor temperatures usually encountered, which may be below 0 °C in a northern Canadian winter, or above 40 °C in a desert. The light hydrocarbons, such as methane to butane, are vapors at 1 atm pressure in normal driving temperature ranges, and their gaseous volumes are a thousand times larger than liquids. Thus, compact storage can be attained only at higher pressures and at lower temperatures, such as in LPG, CNG, and LNG. Hydrogen can be stored in compression, or adsorbed in suitable packing material, or generated in the car from catalytic reactions of liquid fuels. Compressed cylinders require heavy walls, and pose leakage problems that might lead to fire and explosion risks, and are usually banned in confined areas, such as in harbor tunnels.

Solid fuels such as wood, charcoal, coal, and biomass have been used occasionally for motor vehicles at times of scarcity. They are more suitable for large-scale stationary power plants under the watchful eyes of competent engineers, as they are too difficult to feed into the furnaces, and the fuels are too full of sulfur and ash that would corrode and foul the light engines.

The Otto-cycle spark-ignition internal combustion engine cars require the vaporization of a significant fraction of the fuel in the combustion chamber, which requires a highly volatile gasoline, with hydrocarbon molecules in the range of 5 to 12 carbon atoms. The diesel engine has a much higher compression ratio, which makes the air-fuel mixture hotter, and works well with the higher boiling point diesel fuel. Kerosene has a higher gravimetric or volumetric heating value, and is quite popular in Europe. However, the diesel engine is much heavier, and also produces objectionable smoke, smell, and noise.

There is a dark side to gasoline and the automobile culture that poses a significant threat to its future. The United States sees 6.7 million motor vehicle accidents each year, which averages to one person out of 40, resulting in 42,000 deaths. The cars are also a major source of air pollution, contributing carbon monoxide, hydrocarbons, and nitrogen oxides to the urban air. The Clean Air Act was passed to regulate these emissions, and the most successful technology to cope with these problems is the catalytic converter, which is more than 90% efficient. The urban air in many major cities is still below desired standards, and there is public discussion on further tightening of the regulations. Most of the solutions appear to involve changing the engine, and there are few demonstrable results to show that further fuel changes can make a big difference.

Table 9.15 Comparison of fuels for motor vehicles

	<i>Gravimetric heat of combustion, kcal/g</i>	<i>Density, g/mL at 100 atm</i>	<i>Volumetric heat of combustion, kcal/mL</i>	<i>Moles CO<sub>2</sub> per 1000 kcal</i>
H <sub>2</sub>	34.2	0.0082	0.280	0
CH <sub>4</sub>	13.2	0.0655	0.865	4.7
Hexane	11.6	0.659	7.64	6.0
Benzene	10.1	0.879	8.88	7.6
Ethanol	7.4	0.789	5.84	5.9
Methanol	5.7	0.792	4.51	5.5
Lead-acid battery	0.056		0.086	0

The electrical vehicle has often been considered a cleaner alternative to the gasoline engine, but the electrical battery has a very low energy density. The biggest challenge to the gasoline engine is probably the fuel cell, which is under intensive research. The best fuel for the fuel cell would be hydrogen generated from light hydrocarbons, from methane to gasoline. Carbon dioxide is an inevitable emission product of hydrocarbon combustion. It is a greenhouse gas, and its emission has been linked to global warming. One of the measures of the global warming potentials of a fuel is the moles of carbon dioxide emitted per unit of energy released. Since carbon has a lower heat of combustion per gram than hydrogen, a fuel that has a high ratio of H/C would rank higher in energy per weight and lower in carbon dioxide emission. This leads us to hydrogen and methane as the best fuels, but both are gases that are difficult to package, which ranks low in energy per volume. Consider table 9.15, which compares liquid fuels with gaseous fuels compressed to 100 atm at room temperature. In this comparison, the very superior gravimetric energy densities of hydrogen and methane translate to very low volumetric energy densities, even at a pressure of 100 atm. Benzene has a high density and is seen as the best fuel from the point of volumetric heat of combustion; thus, for a fixed gasoline tank size, it would store the most energy. However, it also produces the most carbon dioxide per energy released and it tends to produce sooty flames. We should not assume that the use of hydrogen in motor vehicles would eliminate emitting carbon dioxide, since the source of hydrogen is seldom electrolysis of water by nuclear or hydro-power, but is most likely to be coal or hydrocarbon gasification that produces carbon dioxide in an industrial plant. Similarly, the manufacturing of ethanol and methanol from biomass usually involves intensive farming, which depends on the use of fertilizers, farm machinery, drying, fermentation, and distillation, with a large expenditure of fossil fuels.

### Required Properties

Gasoline is used to propel cars with the spark-ignition Otto-cycle internal combustion engine. Since it is burned to make heat, which drives the pistons of the engine, a high heat of combustion would be valued. Gasoline engines tend to knock under heavy load, such as when climbing a hill or engaging in rapid acceleration, which makes the engine

**Table 9.16 Gasoline property specifications**

Octane number (Research + Motor)/2	87
Reid vapor pressure, max.	9.0 lbs/in <sup>2</sup>
Distillation	
10% max.	140 °F
50% min.	170 °F
50% max.	240 °F
90% max.	365 °F
Lead	None
Oxygen	1.8–2.2%
Olefins, max.	6.0%
Aromatics, max.	25%
Benzene, max.	1%
Sulfur, max.	40 ppm

lose power or even destroy itself. A high-powered engine has a higher compression ratio, and needs even higher octane numbers. Gasoline should also be sufficiently volatile in the winter, so that the engine starts easily in the morning, but it should be sufficiently nonvolatile in the summer, so that it would not vaporize in the fuel line to make bubbles, which would lead to a vapor lock. Gasoline also needs to be clean and not make deposits that gum up the valves or corrode the engine.

The properties of gasoline are specified and regulated by the American Society of Testing Materials (ASTM) (Guthrie 1960). For instance, the property specifications may be as shown in table 9.16. There are two octane numbers: the Research method is measured at an engine speed of 600 rpm, and the Motor method is measured at 900 rpm. The octane numbers listed at the gasoline pump are the average of the two. A higher octane number gasoline can be achieved chemically by changing straight-chain paraffins (such as *n*-heptane) into branch-chained paraffins (*iso*-heptane), by dehydrogenating paraffins into olefins and aromatics, and by additives such as tetraethyl lead and methyl *tert*-butyl ether (MTBE).

Gasoline volatility is measured by the distillation curve, which specifies the relation between distillation temperature and percentage of gasoline evaporated. The ASTM specifies five standard brands of gasoline from A to E, depending on the expected ambient temperature encountered. For the C gasoline, expected to operate at an ambient temperature between 4 and 36 °C, the specification includes the values of T10 (for 10% of evaporation) at 140 °F (60 °C), T50 at 241 °F (116 °C), T90 at 365 °F (185 °C), and an endpoint for total evaporation at 437 °F (225 °C). A gasoline needs a light fraction, so that the engine would have enough gasoline vapor for a “cold start” on a morning in northern winter, and it should not have too large a light fraction, which could lead to the formation of bubbles in the gasoline line on a hot day and the unpleasant experience of vapor lock. It should not have a very heavy end point, so that a large fraction of gasoline would not burn in the engine, and accumulate and degrade the lubricating oil. The C gasoline also specifies a Reid vapor pressure of 9 lbs/in<sup>2</sup> at 100 °F, equivalent to 79 kPa or 0.79 atm, which is a measure of the maximum allowable pressure in the gasoline tank. The purpose of the Reid vapor pressure is to ensure that, when the

gasoline cap is opened for refueling at the gasoline pump, there is not a dangerous rush of vapor from the tank, leading to fire and explosion hazards and to pollution of the atmosphere.

Cleanliness refers to the low concentration of gum, sludge, and deposits. There should be no suspended solids or water in emulsion to block the flow of gasoline. Sulfurous combustion products are an air pollutant that causes acid rain, as well as being corrosive to the engine, and so sulfur is regulated to be below 40 ppm, with special attention to minimizing the smelly mercaptan component. The gasoline should not promote rust in metals. It should be clear and bright in appearance, and have no phosphorous additives, which have been used in the past for detergents; it should also have no lead, which is a legacy of using tetraethyl lead as octane number boosters, and is an undesirable pollutant in the atmosphere. Gasoline is often required by law to contain a minimum amount of oxygen, due to the successful lobbying of corn farmers from the Middle West used to make alcohol by fermentation. Aromatics are deficient in hydrogen, and have a tendency for sooty flames, and benzene is a known carcinogen.

### Market Size, Growth, and Price

Gasoline is a commodity sold at enormous quantity and low price. However, the private automobile is the second most expensive purchase in an average family in America, ranking after the purchase of a house. The *Statistical Abstract of the United States* showed (1997) that there were 129.8 million automobiles registered, plus 77.3 million trucks, 0.7 million buses and 3.8 million motorcycles. We had 182 million registered drivers, who are served by 126,000 gasoline stations, selling 68.9 billion gallons of gasoline per year. Assuming an average cost of \$1.50 per gallon, this amounts to \$92.0 billion a year.

The size of the market can be given by an approximate model

$$129 \times 10^6 \text{ cars} \times \frac{12,000 \text{ miles}}{\text{cars} \times \text{year}} \times \frac{\text{gallons}}{21 \text{ miles}} = \frac{73.7 \times 10^9 \text{ gallons}}{\text{year}} \quad (9.3)$$

According to this model, the number of gallons per year consumed will grow in proportion to the number of cars, to the annual miles driven per car, and inversely to the miles per gallon. The American population is increasing at the rate of approximately 1.0% per year, and the per capita income is increasing at approximately 3.0% per year. The rise of the use of gas-guzzling sports utility vehicle (SUVs) would decrease miles per gallon, but concerns of the environment and global warming could force legislation to increase the required miles per gallon. All considered, there is not a large potential to increase gasoline consumption in the United States, in comparison with emerging nations with faster population and economic growth rates. Economists define a market to be mature if it increases no faster than the increase in the gross national product, as a product increases in sales only when there are more people and when they become richer.

A typical gallon of gasoline sells for \$3.00 per gallon, with the premium gasoline (Octane Number 93) commanding about a \$0.20 premium over the regular gasoline

(Octane Number 87). The major contributors to the gasoline price are: costs of crude oil and other raw materials, transportation, refining, marketing, and federal and state taxes. The major uncertainty in the cost of gasoline is the price of the crude oil, which has varied in recent years from \$15 to more than \$40 per barrel (42 gallons). The Organization of Petroleum Exporting Countries (OPEC) has an enormous influence on the price of oil, and political instability has sent the prices soaring several times in the last few decades. There is a good deal of variation in gasoline prices at the pump, as the recognized brand names like Exxon and Shell sell for more than local brands, and high-tax states, such as Connecticut, have much higher gasoline prices than low-tax states, such as Wyoming and Alaska.

Let us take a look at the buyers and what they can afford. The U.S. Statistical Abstract showed that, in the year 1997, there were 105 million households, with a median income of \$37,005. Table 9.11 shows that, in the year 1997, the average household had a total expenditure of \$34,819 per year, of which the average expenditure in transportation was \$6,457 per year. But the average expenditure in gasoline and motor oil was only \$1,098; the cost of car purchase was \$2,725, the rest of the car cost (insurance, license) was \$2,230. So the family car had a fixed cost of some \$4,966 per year even if it stayed in the garage, and only \$1,098 of variable cost that depends on how much driving is done. Thus, for an average household, the transportation cost is a very high 20.7% of the total family expenditure (figure 9.5), but the gasoline cost is only of the order of 3%.

Do consumers buy more gasoline if their incomes go up and/or if the price goes down? This is called the income elasticity (percentage increase in consumption/percentage increase in income) and the price elasticity (percentage increase in consumption/percentage decrease in price). The statistics in table 9.11 show that, from 1990 to 1997, the average consumer household increased their spending from \$28,381 to \$34,819 (by 22.7%), but gasoline sales had increased only from \$1,047 to \$1,098 (by 4.9%). That is, after receiving a raise, we are more likely to feel good by buying a bigger car (elasticity 1.3) than by spending more on gasoline (elasticity 0.2). Since gasoline does not take up a large fraction of total expenditure, and there are few options, such as not going to work or buying another type of fuel, the price elasticity is quite small in the short run. However, if a price increase persists for a long time, then consumers can gradually switch to smaller and more fuel-efficient cars, so the long-run elasticity is likely to be much higher than the short-run elasticity. In the period 1990–1997, the average household expenditure went up by 2.3% per year, but housing, car purchase, health care, and entertainment went up faster than the average; only food, clothing, and gasoline grew slower than this average.

Table 9.17 shows the U.S. household vehicle expenses in 1988, arranged by household income. It is seen that, as the household income increases from below \$10,000 to above \$75,000, the gallons purchased per household more than doubled, from 650 to 1400 gallons/year, but the burden of the fuel purchase as a percentage of household income drops from 8.4 to 1.6%. The income elasticity can be computed to be

$$\frac{1400 - 650}{650} \frac{10}{75 - 10} = 0.18$$

**Table 9.17 U.S. vehicle expenses and household income, 1988**

<i>Income, 10<sup>3</sup> \$</i>		<i>Households, 10<sup>6</sup></i>	<i>Cars, 10<sup>6</sup></i>	<i>Miles driven, 10<sup>9</sup></i>	<i>Gallons consumed, 10<sup>9</sup></i>	<i>Gasoline cost, 10<sup>9</sup> \$</i>	<i>Car/house</i>	<i>Cost/house</i>	<i>Fuel/income, %</i>
<i>Household</i>	<i>Avg</i>								
< 10	7.5	10.2	13.3	111	6.6	6.4	1.30	627.45	8.37
10–15	12.5	11.8	17.0	153	9	8.8	1.44	745.76	5.97
15–20	17.5	8.3	13.3	132	7.3	7.2	1.60	867.47	4.96
20–25	22.5	8.6	15.1	154	8.5	8.3	1.76	965.12	4.29
25–35	30.0	16.0	29.2	297	16.4	16.1	1.83	1006.25	3.35
35–50	42.5	12.9	28.3	313	16.5	16.3	2.19	1263.57	2.97
50–75	62.5	8.8	20.5	226	11.8	11.6	2.33	1318.18	2.11
> 75	90.0	4.5	10.8	124	6.3	6.3	2.40	1400.00	1.56

For such income-inelastic goods like gasoline, a 1% increase in income only leads to a 0.2% increase in dollars spent, and the rest of the increase in income went to other goods or savings. Gasoline occupies a higher fraction of the family budget for low-income families, but is not as important for high-income families.

### Opportunities and Threats, Incentives for a New Fuel

Gasoline is a mature product in the autumn of its product cycle: it does not have the growth potential of innovative products in the summer of their product cycle, like inkjet printer ink and lithium-ion batteries for laptop computers and cellular phones. Some of the greatest historic innovations in gasoline included: the introduction of thermal and catalytic cracking to increase the yield of gasoline from crude oil, the introduction of tetraethyl lead and catalytic reforming to improve the octane number, and the introduction of detergents and hydro-processing to improve its cleanliness. Gasoline is now an undifferentiated commodity, as the consumers perceive very little differences among brands from different companies. Gasoline became important a hundred years ago, when the gasoline buggy became popular, and it has been reformulated many times. The American public still has a love affair with the sports and other prestige cars, and they remain fascinated by styling, and have developed a growing fascination with SUVs in recent years. A person on the street can readily tell the difference between a Mercedes convertible, which makes the heart beat faster, and a Toyota sedan, which gives a sense of comfort and dependability. In comparison, gasoline is something to fill the tank when it is empty, and it is difficult to say much about the difference between Exxon and Shell gasoline.

In each community, there are usually a number of major gasoline brands such as Exxon-Mobil and Chevron-Texaco, as well as a number of local brands, such as Coastal and Merit. The local brands are usually sold at 5 cents a gallon lower than the majors, and are perceived to have a lower quality. There are usually three separate grades of gasoline: the Regular with an octane number of 87, the Premium with an octane number of 89, and a Super with an octane number of 93. What can the marketer do to persuade the consumers to choose our gasoline, and why should a consumer buy Exxon instead of Shell? There are people who prefer to buy from a particular favored gasoline station, since it has a convenient location, a friendly owner, provides extra services, such as wiping the windshield, has a clean bathroom, and has a tie-in with a convenience store that sells soft drinks. There are people who prefer to buy from a recognized name brand that is highly advertised, with the reputation of quality and dependability. Many more would buy the lowest priced brand, as they could not tell the differences among them.

A marketer seeks to discover the customer's "hot button," which is a burning desire that leads a customer to buy. Does the new gasoline have a special detergent that keeps the engine cleaner? Would this new gasoline strike a blow for ecology? Is this gasoline 5 cents cheaper per gallon? The conservative approach is to tinker with the current gasoline and make it a little better, and the ambitious approach is to come up with a revolutionary new product. However, there does not seem to be any significant new technology on the way to make major product improvements.

There have been very few examples in recent years of product differentiation in the marketing of gasoline. The additive MTBE was thought to be a good idea, as it increases the octane number and adds oxygen to gasoline without the corrosive problems of methanol. It turned out that MTBE was leaking out of storage tanks and began contaminating underground water, and so it is being phased out. Thus, there is an opportunity to work on a substitute for MBTE that has its good qualities without its bad qualities. Do we have any technology lead on what that might be? If we are successful, how long can we keep the competition from copying our idea?

## 9.4 Marketing Ink for Inkjet Printers

Whereas gasoline is sold at large volume and low cost per pound, the ink for inkjet printers is sold in small volume at high cost per pound. It is a relatively new product that is riding on the crest of the information revolution, with great potential for growth and profit. Let us consider the marketing information needed in preparation for the introduction of a new product.

### Customers and Needs

Let us define the market as printing supplies, and the buyers are individuals, businesses, and governments. The need is for rapid and accurate records that could be stored for many years without deterioration. An older generation of typewriters based on striking an inked ribbon with a mechanical typeface is noisy and slow, and they do not have the flexibility of a large selection of typefaces to handle foreign languages and mathematical equations; also, it is impossible to make changes in a typed page or to correct mistakes. The computer word processor combined with the laser printer is an enormous advance that makes correction easy, is quiet and fast, and makes very sharp prints. It is a relatively expensive piece of equipment, and there are still not many popular color laser printers to choose from.

The inkjet printer is relatively inexpensive as a piece of equipment, and is currently the printer of choice in families and small businesses. It is not as fast, and the print is not as crisp as the laser printer, but it is quite acceptable; it excels in making color charts and printing photographs from digital cameras and scanners. The buyers of inkjet printers and ink are individuals, as well as business and government. The individual buyer connects the printer to a home computer to print out black-and-white texts of letters, email and Web messages, inventories, diaries, and receipts. People who have their offices at home would also use the printer for reports with charts and graphics. The more ambitious buyers would print colored greeting and birthday cards and party decorations; increasingly it is used to print photographs that can be edited to be bright and colorful, with wrinkles and moles removed, as well as cropped to remove distracting details. The buyers usually have some familiarity with operating computers and technology, and have sufficient income to afford a little refinement. Individual inkjet printers are preferred in small business units for cash registers, ticket counters, and warehouses, to print out receipts, tickets, and inventories. Large establishments

**Table 9.18 Printer catalog survey**

Printer and ink	Price, \$	Pages/min	
		Black-and-white	color
<i>Inkjet printers</i>			
HP Deskjet 3420	79.00	10	8
HP Deskjet 3820	99.99	12	10
HP Deskjet 5550	149.99	17	12
HP Photosmart 7150	179.00		
HP #45 black ink	29.99		
HP #78 tri-color ink	54.99		
<i>Laserjet printers</i>			
HP Laserjet 4100	1,099.00	25	
HP Laserjet 9000	2,999.00	50	
HP toner cartridge for Laserjet 4000	128.99		

often prefer centralized network laser printers that have greater speed and economy of scale.

A survey of a printer catalog (Google) from the summer of 2002 is shown in table 9.18. At the high end are the dye-sublimation printers that require three to four toner cartridges, but they make small nonoverlapping dots that are not particularly good for photographs. Digital photography is one of the fastest growing markets, and is capturing market share at the expense of conventional cameras and film-based photographs. A fairly sharp photograph requires a minimum of 200 dots per inch, so that an 8 in.  $\times$  10 in. requires an image of  $1600 \times 2000 = 3.3$  mega-pixels. This amounts to a resolution of 0.13 mm, or 130  $\mu$ m, for each dot. A photograph for medical diagnosis should have much higher resolution, such as 1440 dots per inch, so that the resolution is 18  $\mu$ m. Some photographs also serve archival purposes, so they should resist fading for 50 years or longer.

### Properties Required

The ink for inkjet printers is essentially colored water, and the current formulation of ink is about 75% water, plus dye and glue (Kang 1991, Le 1998, Pond 1999). The ink should flow readily from the inkwell to the nozzle, and then to the paper, without dripping or clogging. The viscosity of current inkjet ink is from 2 to 8 cP. The droplets formed should be very uniform in size, and should be as small as possible to increase the resolution. When the droplet hits the paper, the droplet should penetrate the fibers without splatter, which would increase the diameter of the dot on the printed text or picture. The surface tension of current ink is around 45 to 50 dynes/cm, which is much lower than water at 70 dynes/cm. It is believed that if the surface tension is raised from 45 to 65 dynes/cm, the spread (or ratio of diameter of spot on paper/diameter of droplet) would decrease from 3.3 to 1.1. Much higher resolutions

are required for the applications of aerial military photography and medical imaging. The goal of 1440 dots per inch of resolution requires that the diameters of the droplets are no larger than 17  $\mu\text{m}$  in diameter, or a volume of 32 pL, with no satellite drops.

The color should be fast; that is, it should not fade with heat or sunlight. This is easier to achieve with inorganic dyes, which tend to be insoluble granules that may have the disadvantage of clogging up the jets. Organic dyes are particularly susceptible to the effects of ultraviolet light, and would fade after long exposure. The competition for digital photographs is still the time-honored Kodachrome, which has a much higher resolution and produces a baby or wedding picture that is still enjoyable after 50 years.

### Market Size and Prices

This is a small-volume business in comparison with gasoline, with a very high price per weight or volume and a tremendous growth potential. It is hard to obtain marketing data, but it must be huge and growing. The price is currently about \$40 for a cartridge of 40 mL, which is \$1.00/g and is more than 2000 times higher than gasoline at \$1.50/gallon. The more relevant measure is the ink cost per page. Since a printed page varies in the amount of ink used, it is difficult to design and execute experiments to perform a fair and meaningful comparison.

The major brands are marketed by Hewlett Packard, Cannon, Epson, and others. There are also house brands that sell for less. The spent ink cartridges are removed from the printers and discarded, and new ones inserted into the printers. The spent cartridges contain much valuable plastics and metals, but are too small in quantity to warrant a recycle effort. There are also manufacturers who sell the ink separately, to be injected by hypodermic needles into the spent cartridges. The major manufacturers always warn the consumers not to use these low cost alternatives, as they may be inferior in quality and give poor performance due to lack of compatibility.

### Opportunities

There are huge opportunities for innovation. Digital photographs may be in the process of totally driving out the traditional film-based photography with silver halide. The amateur photographer has total control of picture cropping to concentrate on smiling faces, and to get rid of nonessential and annoying background areas in the picture, to adjust the contrast of light and shadow and the balance of color. Some of the main opportunities to improve the current ink are to increase the resolution to match Kodachrome and to increase the life of the print to enable archiving for 50 years. Can inorganic pigments be made very small and into stable emulsions, and can organic dyes be made to be stable against ultraviolet light and heat? If this business is very successful for many years, then there would be an accumulation of spent ink cartridges that will have to be reckoned with. So, the design of recycled ink cartridges would be a challenge in the future.

## References

- Guthrie, V. B., ed. 1960. *Petroleum Products Handbook*. New York: McGraw-Hill.
- Kang, H. R. 1991. Water based ink-jet ink. *Journal of Image Science* (May–June): 179–188, 189–194, 195–201.
- Kirk-Othmer Encyclopedia of Chemical Technology*. 2001. New York: Wiley.
- Le, H. P. 1998. Progress and trends in ink-jet printing technology. *Journal of Image Science and Technology* 42(1): 49–62.
- Levitt, T. 1975. Marketing myopia. *Harvard Business Review* (September–October).
- Maslow, A. 1987. *Motivation and Personality*. New York: Harper & Row.
- Pond, S. 1998. *Inkjet Technology and Product Development Strategies*. Carlsbad, CA: Torrey Pines Research.
- Statistical Abstract of the United States: the National Data Book*, 121st edition, 2001, Washington, DC: U.S. Census Bureau.

## Further Reading

- Bangs, D. 2002. *Business Planning Guide*.
- Barrer, R. M. 1978. *Zeolites and Clay Minerals as Sorbents and Molecular Sieves*. London: Academic Press.
- Blankenship, A. B., G. E. Breen, and A. Dutka. 1998. *State of the Art Marketing Research*. Chicago, IL: NTC Business Books.
- Bureau of Economic Analysis. 1994. *Benchmark Input–Output Accounts of the U.S.* 1987.
- U.S. Department of Commerce. Washington, DC: Government Printing Office.
- Bursk, E. C. and S. A. Greyser. 1975. *Cases in Marketing Management*. Englewood Cliffs, NJ: Prentice-Hall.
- Chung, H. S. et al. 1999. Recent developments in high-energy density liquid hydrocarbon fuels. *Energy & Fuels* 13(2): 644–649.
- Coodger, E. and R. Vere. 1985. *Aviation Fuels Technology*. London: Macmillan.
- Coordinating Research Council, Inc. 1984. *Handbook of Aviation Fuel Properties*. Warrendale, PA: Society of Automotive Engineers.
- Couper, J. R., O. T. Beasley, and W. R. Penney. 2001. *The Chemical Processing Industries Infrastructure*. New York: Marcel Dekker.
- Dahl, J. E., S. G. Liu, and R. M. K. Carlson. 2003. Isolation and structure of higher diamondoids, nanometer-sized diamond molecules. *Science* 299: 96–99.
- Edwards, D. et al. 1997. Large porous particles for pulmonary drug delivery. *Science*. 276: 1868–1871.
- Flank, W. H. ed. 1980. *Adsorption and Ion Exchange with Synthetic Zeolites*. ACS Symposium Series vol. 135. Washington, DC: American Chemical Society.
- Flank, W. H. and T. E. Whyte, eds. 1988. *Perspectives in Molecular Sieve Science*. ACS Symposium Series vol. 368. Washington, DC: American Chemical Society.
- Goldenberg, J. and D. Mazursky. 2002. *Creativity in Product Innovation*. Cambridge: Cambridge University Press.
- Hansen, H. L. 1967. *Marketing: Text, Techniques, and Cases*. Chicago, IL: Richard D. Irwin.
- Hanson, E. 1999. *Recent Progress in Ink Jet Technologies II*. The Society for Imaging Science and Technology.
- Kremer, T. 2001. Printing enters the jet age. *Invention & Technology* 17: 18–27.
- Lake, M. 2000. The inkjet color image, born in a rainbow of tiny drops. *New York Times* (May 4): G8.

- Lee, J. 2001. Beyond the leaky faucet: dissecting the complex drop. *New York Times* (March 6): F4.
- Marchand, A. P. 2003. Diamondoid hydrocarbons: delving into nature's bounty. *Science* 299: 52–53.
- National Research Council. 1988. *Frontiers in Chemical Engineering: Research Needs and Opportunities*. Washington, DC: National Academy Press.
- Nishiguchi, T., ed. 1996. *Managing Product Development*. Oxford University Press.
- North American Industrial Classification System*. 1997. Washington, DC: National Technical Information Service.
- Samuelson, P. A. 1961. *Economics*. New York: McGraw Hill.
- Sargeant, A. and D. C. West. 2001. *Direct and Interactive Marketing*. Oxford: Oxford University Press.
- Sperling, P. 1989. *Alternative Transportation Fuels*. Quorum Books.
- Standard Industrial Classifications*. 1987. Office of Management and Budget. Distributed by National Technical Information Service, Washington, DC.
- Statistical Abstract of the United States: the National Data Book*. 119th edition. 1999. Washington, D.C.: U.S. Census Bureau.
- Stigler, G. J. 1966. *The Theory of Price*. New York: Macmillan.

## Discussion Topics

1. Which market segments in the chemical industries are fastest in growth, and which segments are declining? Some of us prefer to work on products for a market segment that is above average in growth, but some of us prefer market segments that are below average in growth. What may be the reasons for these preferences, and where do you fit in?
2. Chemicals seldom win the Oscar Award for the best *leading role* in the market; they usually aspire to win the best *supporting roles* in the success of other goods with starring roles. For instance, gasoline and lubricating oil play supporting roles to the sports car, which holds center stage in attention and affection. Paint and glue play supporting roles to a house, which is the most important financial investment of an average family. For each of the following basic needs of mankind, name as many chemicals in supporting roles that you can:

- Housing (such as house and apartment)
- Household operations (such as stove, refrigerator, washer, heater, air conditioner, kitchen sink)
- Grooming (soap, cosmetics, shampoo)
- Food (bread, milk, meat, fish, vegetables)
- Clothing (suits, overcoats, shirts, dresses, ties, shoes)
- Transportation (airplanes, ships, cars, bicycles, buses, trucks)
- Information (books, newspaper, radio, television, Internet)
- Education (school, university)
- Entertainment (movies, music, sports)
- Health (doctors, hospitals, drugs)
- Government (police, judges, roads and bridges, sewage, army, space).

- How many Americans are of college age, from 18 to 21? What fraction of them goes to college? What fraction of college students takes Economics 101? How many introductory economics textbooks are sold in the United States every year? (Suggestion: start from the U.S. Statistical Abstract, and add your judgment and estimations.)
- How many gallons of gasoline are sold per year in the United States? Before you look this up in the U.S. Statistical Abstract, do an order of magnitude estimate based on common sense. (a) How many people live in the United States? (b) How many cars do we own per person? (c) How many miles a year is an average car driven? (d) What is the average gas mileage in miles per gallon? After you have made the estimation, look up and compare with the correct answer in the U.S. Statistical Abstract.

## Exercises

- The common table salt, NaCl, is a very versatile material that has many uses, including: (a) flavoring and preserving food, (b) animal forage in agriculture, (c) highway use to melt ice and snow, (d) raw material to make chlorine and caustic soda, (e) water softening. For each of these uses, list:
  - the properties of NaCl that are essential to the success.
  - the substitutes to NaCl for each use, if there are any;
  - the drawbacks to NaCl, from the consumer point of view, and from the environmental point of view;
  - what can be done to improve on the weaknesses of NaCl.
- Pick three products of the CPI from a different SIC. Do a literature research on the market of each product. Who are the major buyers? How does the product perform for the customers? What are the fortunes of the customers in recent years? What is the current price and price trend? Where are these products in the life cycle? Are there any new opportunities or threats to the prosperity of these three chemicals? (Suggestion: start with the *Kirk-Othmer Encyclopedia of Chemical Technology*, the *Ullmann Encyclopedia*, and the *Chemical Market Reporter*.)
- Some people think that the digital camera is a luxury good with a high price elasticity, such as  $E = 2$  so that  $Q = cP^{-2}$ . When the price goes up, the quantity sold would go down; so, when you take a derivative you obtain the formula

$$\frac{dQ}{dP} = -2cP^{-3}$$

When the price goes up by 1%, the percentage change in volume is then given by

$$\frac{P}{Q} \frac{dQ}{dP} = \frac{P}{cP^{-2}} (-2cP^{-3}) = -2$$

So when the price goes up by 1%, what happens to the percentage change in revenue? Derive a formula relating the percentage change in revenue from 1% change in price.

# 10

## Safety, Health, and Environment

### **10.1 The Responsibilities of the Chemical Processing Industries**

Third-party accountability  
Changing laws and public perceptions  
Cradle-to-grave stewardship responsibilities

### **10.2 Health and Safety Problems**

Flame and explosion hazards  
Toxic chemicals in plants  
Food, drink, and air

### **10.3 Environmental Problems**

Pollutant emission  
Pollutant transport in the environment

Pollutant transformation in the environment

Pollutant accumulation  
Damage and harm

### **10.4 Green Engineering**

Remedial clean-up  
Green process engineering  
Green product engineering

#### **Reference**

#### **Further Reading**

#### **Websites**

#### **Discussion Topics**

#### **Exercises**

There was a time when the primary concerns of the product engineers were the needs of the buyers and consumers of their products, and they concentrated most of their efforts on design and manufacture of products with the goal of meeting their requirements and approval. A product will certainly fail if it does not have the steady and continued confidence of the consumers. The CPI, just like any other industry at that time, were accountable mostly to the consumers. This was a two-party transaction, where the buyer gave the seller money in return for a satisfactory product.

This form of two-party transaction is no longer valid, as there are many bystanders whose welfare can be damaged in the transaction and their welfare must also be safeguarded. The manufacturing and marketing of chemical products is now a multiparty transaction, as the public and the governments have forcefully placed themselves into part of the bargain. The product engineers must become cradle-to-grave stewards of their products, and must solve many safety and environmental problems: from the extraction of raw material from farms and mines, to transportation on land and over

water, to manufacturing in plants, to use in the customers hands, and finally to recycle back to nature.

Before you begin to design the product, you need to focus your attention on the customer and their needs, but you also need to focus your attention on the potential hazards that the product poses to the safety and health of your workforces and the neighbors, and to the environment. You need some familiarity with the history of past mistakes, with the current government regulations, and with methods to deal with these potential problems.

## 10.1 The Responsibilities of the Chemical Processing Industries

### Third-party Accountability

Over a long period of history, commercial transactions were primarily a two-party affair, between the seller and the buyer. In the last few decades, a third party has forcefully entered into the transaction, based on concerns about safety and the environment, and on public opinion and government regulations. This is shown in figure 10.1. The third party consists of:

- The *employees*, particularly the workers at the plants that would be exposed to the hazards to safety and health created by the raw material, the intermediates, and the finished products and by-products. A plant that is unsafe to workers would not be able to recruit the most capable staff, would not be able to manufacture the best products, and would encourage the formation of militant labor unions and punitive regulations. It is in the self interest of the manufacturers to make sure that the workplace is safe and to adopt the best practice. There is also a recent tendency for manufacturing to move to regions and nations where plant safety rules are less rigorously posted and enforced. This is at best a short-term solution, as when these regions and nations become more affluent, they would demand the same protection as the most advanced industrial nations.
- The general *public* and the local neighbors who might be exposed to hazardous substances that come over the plant fences in the air as vapors and aerosols, spilled on the ground in traffic accidents, discharged into the water supply, and disposed of as solid wastes. The environment represents a collection of concerns

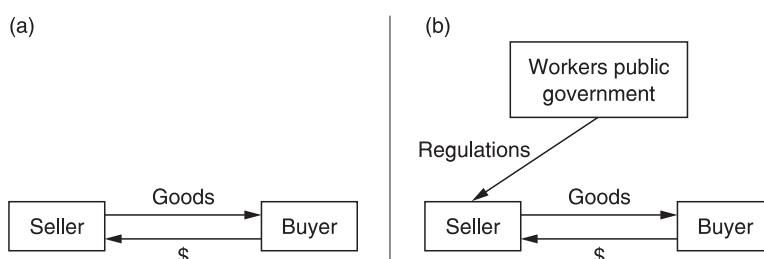


Figure 10.1 (a) Two-party and (b) multiparty transactions

that goes beyond human life and health, which includes economic and natural losses: clean air, acid rain, contamination of water in rivers and lakes, accumulation of solid wastes, impact on animals and plants, reduction of wildlife biodiversity, ozone hole in the stratosphere, and global warming. A plume of pollution can travel thousands of miles and can become a global problem across national boundaries and continents.

- The *governments*, which represent and enforce the interests of the public, can develop adversarial relations with the chemical industry. The principal U.S. government guardian on behalf of the employees is the OSHA, which establishes standards of exposure for workers, such as maximum acceptable air concentrations in a plant averaged over an 8-h day. The American Conference of Government Industrial Hygienists (ACGIH) is another organization that produces guidelines. The U.S. government guardian on behalf of the public is the EPA, which establishes standards of air concentrations in the cities, drinking water standards, and many others. The state of California has its own peculiar air pollution problems, and has regulations that are more stringent than the federal regulations. And industrially developed countries, such as Germany and Japan, often have their own standards and regulations.

### Changing Laws and Public Perceptions

The chemical industries need to obey the law, which is not fixed but keeps evolving with research discoveries, world events, and public opinion. There are numerous federal, local, and international laws about the responsibilities of chemical manufacturers. In addition, an injured party can sue and collect damages from the manufacturers about stated and implied guarantees that were not met, even if the role played by a chemical manufacturer is only peripheral. There are several basic laws that must be part of any product planning and design:

- 1970 Occupational Safety and Health Act
- 1970 Clean Air Act
- 1972 Clean Water Act
- 1972 Federal Insecticide, Fungicide, and Rodenticide Act
- 1976 Toxic Substance Control Act
- 1976 Resource Conservation and Recovery Act
- 1980 Comprehensive Environmental Response, Compensation, Liability Act

These safety and environmental standards change over the course of time, as new scientific discoveries are made every month on the harmful effects of introducing new chemical substances into the environment. For instance, in the 1920s, Thomas Midgley knew very little about the hazards of tetraethyl lead, and he certainly would not have heard about the ozone hole. If a product is designed to be inert and harmless, it may become very popular and used in very large quantities over a long period of time, and it becomes increasingly more probable that it and its decomposition products will accumulate somewhere on Earth and cause damage

to some species, so that making the product acceptable to the public is often a moving target that evolves with time. Sometimes, today's solutions become tomorrow's problems.

The law of the land usually follows the public's perception and opinion. To a significant fraction of the population of industrial nations, the chemical industry has a negative image as producers of dangerous toxic substances, and as being bent on profit pursuit with scant care for public welfare. Indeed, there are a number of well-publicized incidents that show that this perception is not totally without foundation. Furthermore, there is often a wide difference between the public perception of the safety and environmental effects of chemicals on the one hand, and the scientific theory and statistical evidence on the other hand. The public and the press react strongly to dramatic major accidents, such as airplane crashes, oil tanker spills, or nuclear plant accidents. On the other hand, there is no clamor to curb the 40,000 annual deaths by automobiles each year. The leading causes of death in the United States have been led for a long time by heart disease and cancer, as shown in table 10.1. Yet the EPA would like to control exposure to toxic substances to levels that pose a very small incremental lifetime cancer risk. Product engineers should work to safeguard public health, even if less than 1 in 100,000 cases of cancer deaths can be attributed to exposure to chemicals.

Industry rebuttals are often ineffective when they are made without considering the major difference between public tolerance of the risks of voluntary and elective activities, such as motorcycling and smoking, and the intolerance of the risks of involuntary activities, such as from drinking water and eating seafood. These relative differences are shown in table 10.2. It is not sufficient to say that our products represent only minuscule increments to the other risks that the public tolerates readily, as voluntary risks are readily accepted. In a recent case, a customer was scalded by a cup of hot coffee at McDonalds, and she sued and won despite the fact that she also has a responsibility to exercise caution either at home or at a restaurant.

**Table 10.1 Leading Causes of Death in the United States, 1992**

<i>Cause</i>	<i>Annual deaths (10<sup>3</sup>)</i>	<i>Percent</i>
Cardiovascular (heart)	720	33
Cancer	521	24
Cerebrovascular (strokes)	144	7
Pulmonary (bronchitis, etc.)	91	4
Pneumonia and influenza	76	3
Diabetes	50	2
Non-motor vehicle accident	48	2
Motor vehicle accidents	42	2
HIV/AIDS	34	1.6
Suicide	30	1.4
Homicide	27	1.2
All other causes	394	18
Total annual deaths	2177	100

**Table 10.2 Annual risk of death associate with voluntary and involuntary activities**

Activity/exposure	Annual deaths per 100,000 persons
Motorcycling	2000
Smoking, all causes	300
Smoking (cancer)	120
Hang gliding	80
Coal mining	63
Farming	36
Motor vehicles	24
Chlorinated drinking water (from chloroform)	0.8
Four tablespoons peanut butter per day (from aflatoxin)	0.8
A 3 ounce charcoal broiled steak per day (from polynuclear aromatic hydrocarbons)	0.5

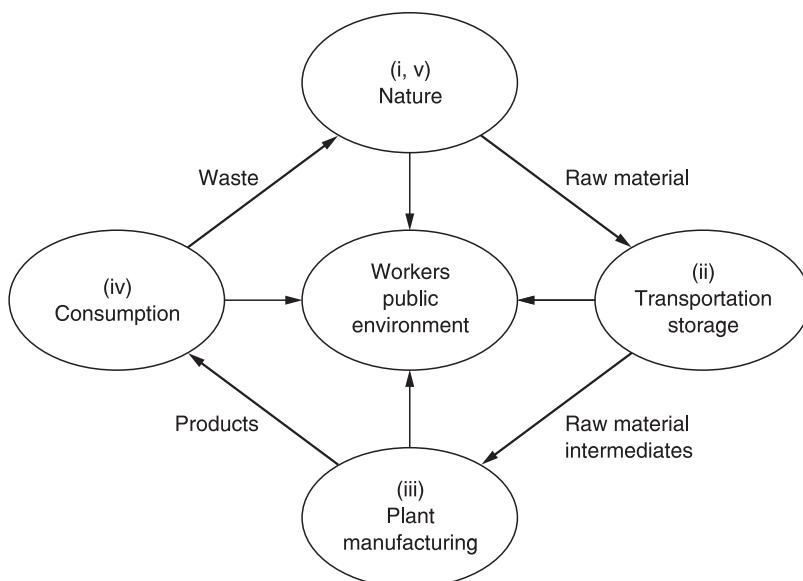
Risk assessment has become a dynamic process, in the sense that what was considered safe at one time can be considered as unsafe later. When we quantify risks, we need to distinguish among parameters that represent eternal truths and other parameters that represent temporal knowledge and considerations:

- Physico-chemical property parameters, such as boiling points, lower explosion limits, and flash temperatures, are well-defined parameters. Some of them have been measured accurately and can be found in databases that have been refereed and vetted. Sometimes, they have not yet been measured accurately, and there exist only tentative values in a database. But once they have been measured and authenticated, these parameter values should be archived and should not change with time.
- Biomedical parameters, such as LD<sub>50</sub>, are well defined, but they are derived from the studies of different populations and represent statistical probabilities of animals and people. These numbers often change from one study to another, when different populations are used; and species can develop resistances after long exposure. There is also no definitive method to relate and scale-up animal parameters to human parameters, or to scale among people of different ages and states of health; so, they are more indicative than authoritative.
- Legislated and regulatory parameters, such as the TLV and the PEL, are frequently changed with the publication of new research results and new government regulations. The value of TLV for vinyl chloride was 500 ppm in 1970. In 1973, Maltoni reported that rats exposed to 250 ppm of vinyl chloride developed tumors. In 1974, B. F. Goodrich announced that three workers died of angiosarcoma of the liver in the Louisville Kentucky PVC plant. After a public hearing and legislative decisions, the vinyl chloride TLV was dropped all the way to 1 ppm in 1980. Legislated parameters are also likely to be different for each government, and there can be a significant difference between European and American requirements.

### Cradle-to-Grave Stewardship Responsibilities

Some manufacturers take the high road, making public announcements that they intend to assume the responsibility of the cradle-to-grave stewardship of all products manufactured and sold. A product begins its journey through the product environmental cycle, shown in figure 10.2, passing through the following stages:

1. *Extraction* of raw materials from natural resources, such as mineral ores from the ground, oil and gas from wells, harvests from farms and forests, kelp and bromine from seawater, and nitrogen from the air, can create many problems. An active or abandoned mine or oilfield can be a very ugly sight and dangerous place, and can create discharges that are harmful. The very act of creating access roads and pipelines can cause destruction of natural habitats. Natural resources are limited, and extensive extraction can cause exhaustion and extinction of plants and animals.
2. *Transportation* of raw materials to depots and plants, by oil tankers, natural gas pipelines, barges, trains, and trucks, can create hazards. Another set of transportation hazards is posed by chemical intermediates (produced in one plant and transported to another plant for further processing) and industrial wastes. Sometimes, they travel past crowded urban areas, pristine beaches, and harbor tunnels. The hazards of oil spills on beaches by tankers are well known, and so are the instances of ruptures and breaks in oil and natural gas pipelines, which produce fire and pollution hazards. Trains and trucks can overturn in centers of habitation, and can lead to flames and explosions, as well as toxic hazards to the community. Toxic chemicals are barred from being trucked in tunnels under harbors and rivers. The *storage* of raw material and intermediates at plants or



**Figure 10.2** Cradle-to-grave responsibilities

depots, such as LPG tanks, can pose hazards to communities by leakage and explosion.

3. *Manufacturing* of the products from the raw materials and intermediates is done in industrial plants, where flammability, explosion, and toxic hazards exist for the workers and their neighbors. Plant discharges into the air and of liquid and solid wastes on to the land and into the water pose another set of potential hazards.
4. *Customer* consumption of the product, either right away or after storing for a period of time, also poses potential hazards to the user, as well as to other people in the neighborhood.
5. *Disposal* of the remnants after the product has passed its useful life should be done safely, perhaps to be reused or recycled, or broken down and returned to its origin. A product such as Freon may contain components designed to be safe and inert, and thereby would not degrade in the environment and persist and accumulate to cause long-term problems.

## 10.2 Health and Safety Problems

We address here the problem of human exposure to toxic and dangerous chemicals, including the workers in chemical plants and the public in homes or at work.

Plant workers are exposed to the raw materials, intermediates, products, by-products, and waste discharge. Regulations for the protection of plant workers are under the jurisdiction of the OSHA, and supported by the research arm of National Institute of Occupational Safety and Health (NIOSH). They issue a number of enforceable guidelines for safety, and they also issue advisories that are suggested but not required. We concentrate here on hazards that are specific to chemicals: flammability, and toxicity through breathing the air and by skin exposure.

The protection of the public in the safety of food, drinks, and drugs is entrusted to agencies such as the FDA and the EPA. They also issue many documents on what is not safe and the suggested procedures on how to handle and to dispose of them.

### Flame and Explosion Hazards

Fires and explosions are life-threatening hazards that have devastated many chemical plants. A measure of the hazard of flammability for a chemical can be quantified by the explosion limits and by the flash point. A mixture of air and chemical is explosive only within certain concentration ranges, which is familiar in the gasoline engine: a mixture that is too rich (lower than 11 parts air to 1 part fuel) or too lean (higher than 15 parts air to 1 part fuel) would not burn. When there is a leakage of a chemical into air, the LEL is attained sooner, and is the more critical for safety than the UEL (chapter 6). Figure 10.3 shows the LEL of paraffins and 1-alcohols, which are strongly decreasing functions of the number of carbon atoms  $N_c$ . Thus, methane will explode at 5% by volume in air, but decane will explode at 0.6%; considering that decane has a molecular weight that is almost nine times higher than methane, the LEL stated in weight percentages is approximately the same. The alcohols are slightly less explosive, and their LEL numbers are slightly higher than the paraffins.

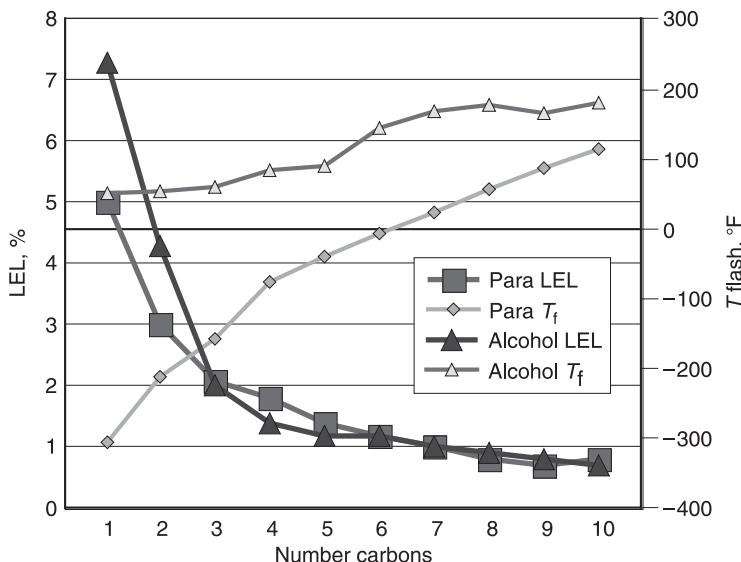


Figure 10.3 Lower explosion limits and flash temperatures of paraffins and alcohols

We also measure fire hazards by the flash temperature, which is the temperature at which a liquid will catch fire when exposed to oxygen and a flame. It is a function of the vapor pressure at the temperature, as well as the LEL. The small paraffin molecules in natural gas, LPG, and gasoline with up to  $N_c = 8$  have flash temperatures below room temperature, which makes them dangerous in storage and use in the home or plant. It also makes gasoline useful in spark-ignition automobile engines. Kerosene and fuel oil contain the higher paraffins with  $N_c > 12$ , which have higher flash temperatures and represent less of a fire hazard. An arsonist might pour gasoline on a house to set it on fire, but cannot use diesel or jet fuel for the same purpose. The alcohols have much higher flash temperatures, as they are less volatile, but they are easily set on fire up to  $N_c = 5$ . The vapor pressure of ethanol in brandy is less than half of that in pure ethanol, so that if you want to flambé your Cherries Jubilee, you need to warm your brandy before trying to light it.

One of the best known explosion disasters took place in Flixborough, England, in 1974. Nypro Limited manufactured 70,000 tons/year of caprolactam as intermediate for the manufacturing of nylon. This is done by air oxidation of cyclohexane to cyclohexanol, with the help of a number of catalysts in the reactors. At that time, cracks developed in the reactor combined with pipe rupture, which released 30 tons of cyclohexane in a cloud. It was ignited by an unknown source and exploded, which resulted in 28 deaths and 36 injured, and the fire burned for 10 days. This disaster was also devastating to the future of the company.

### Toxic Chemicals in Plants

The most basic concept in toxicology is that of the lethal dose. Paracelsus (1493–1541) said that “All substances are poisons, there is none which is not a poison. The right dose

differentiates a poison from a remedy.” Different people react to chemicals differently, and there is a wide distribution of responses. The major routes of entry into the human body are drinking water, eating food and other substances, inhalation of air into the lung, and exposure through the skin.

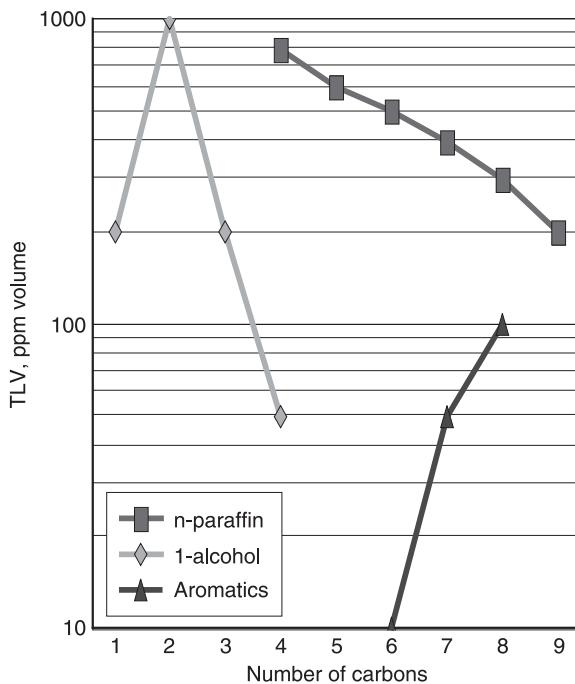
Respiration into the lung alveoli is the most important hazard for plant workers. The average lung has 300 million alveoli with a surface area of about  $70\text{ m}^2$ , which is designed for rapid gas exchange. This alveolar surface area is about 40 times larger than the external skin area of a person. There has been a systematic study of the major air pollutants associated with the protection of industrial workers, as monitored by the OSHA and the ACGIH, and there are three well-documented measurements and databases:

- OSHA PEL. The permissible exposure limit is measured on a normal 8 h day over a 40 h per week for an average worker, called the time-weighted average (TWA). Other important parameters are: “CL,” which indicates the ceiling limit that must not be exceeded, and “Pk,” which designates the maximum peak value allowed, which is above the ceiling value.
- ACGIH TLV. The threshold limit value is measured using the same TWA method, over an 8 h/day and a 40 h/week. The parameter “sk” designates the material penetrates the skin, so that skin contact should be avoided more than breathing; “STEL” indicates a short-term exposure limit, which is averaged over 15 min; “BEI,” indicates biological exposure indices, which are the warning levels in the samples taken from the tissues, fluids, or exhaled air of exposed workers.
- DFG MAK. This is the standard established by the German Research Society, designated as workplace hazard potentials. The MAK values are revised annually.

The paraffins and 1-alcohols are relatively low-risk compounds. When we make a comparison of their TLV values versus the number of carbon atoms  $N_c$ , we find that paraffins from methane to propane are not considered toxic, but the paraffins from butane to nonane are increasingly more toxic with  $N_c$ , which is shown in figure 10.4. The 1-alcohols make a curious volcanic curve, starting from the toxic methanol to the relatively harmless ethanol, and the trend is downwards from propanol to butanol. The aromatics are much more toxic, but the lower molecular weight benzene is more toxic than toluene, which is more toxic than the higher molecular weight ethyl benzene.

The other major route of chemical hazards to plant workers is penetration through the skin, which has a surface area of about  $1.8\text{ m}^2$  for a male adult, and is lined with a tough horny layer of densely packed cells without nuclei. So the skin is reasonably impervious to the penetration of chemical substances, and penetration depends on the lipid solubility of the compound, where the diffusion speed is inversely related to molecular weight. Skin penetration hazard becomes relatively more important than breathing for high boiling point liquids, which have lower vapor pressures.

One of the most notorious cases of industrial disaster took place in 1984 in Bhopal, India. A plant with a license from the Union Carbide company was making methyl isocyanate (MIC),  $\text{CH}_3\text{NCO}$ , which is an intermediate for the manufacture of pesticides. MIC boils at  $39^\circ\text{C}$ , and the vapor density is heavier than air and very toxic by inhalation and skin absorption. The maximum allowable air concentration is 0.02 ppm by volume over 8 h. MIC also reacts with water and produces heat, which must be removed to prevent boiling over. On that day in 1984, the cooling system failed during



**Figure 10.4** TLVs of paraffins, alcohols, and aromatics

a labor dispute, and water in the tank reacted with MIC. This led to overheating, and 25 tons of vapor escaped. The gas went beyond the plant gates to reach the local village, killed 2000 residents and injured 20,000 others. The resulting ill will and bad publicity continues to this day, over two decades later.

### Food, Drink, and Air

A different set of hazards threatens the safety and health of customers and the public, of poisoning by drinking water and eating objects. The FDA has jurisdiction over food and drugs, and the EPA has jurisdiction over drinking water.

The concept of  $LD_{50}$  is defined as the dosage given at one time, where death occurs to half the recipients (involving various types of animal) of the specified dose. The parameter  $TD_{50}$  is designed to measure toxic effects registered by half of the recipients of the specific dose. The lethal dose depends on the route of administration, so that it is important to distinguish between  $LD_{50}$  (oral) and  $LD_{50}$  (i.v.). There is a very wide range for  $LD_{50}$ , with typical numbers as follows:

Substance	Ethanol	NaCl	Nicotine	Dioxin	Botulin toxin
LD <sub>50</sub> (mg/kg body weight)	10,000	4000	1	0.001	0.00001

Different animals have different susceptibilities, so that it becomes necessary to distinguish between LD<sub>50</sub> (mouse) and LD<sub>50</sub> (rat). For instance, in the tests for the toxicity of ethanol, small animals are found to be much more resistant than humans:

<i>Animal</i>	<i>Rat</i>	<i>Mouse</i>	<i>Guinea pig</i>	<i>Human</i>
LD <sub>50</sub> , mg/kg	12,500	8,000	5,500	3,500

There are no general reliable methods to scale up from a mouse to a rat, and to a person; nor is there a reliable method to scale down from a 70 kg male adult to an infant. A faster and less expensive toxicity test is the Ames test, performed on bacteria to test potentials for carcinogenesis, but the scale-up to humans is even more imprecise.

Acute toxic damage means that an animal succumbs in less than a day. Chronic toxic damages measures long-term hazards, over months and years (e.g., measuring long-term teratogenic effects on malformation of a fetus). For prolonged exposures, the parameter LC<sub>50</sub> is used to measure the concentration where death occurs to half of the recipients. Some authors use a toxicity rating chart, which gives the probable oral lethal dose for humans:

1. Practically nontoxic	>15 g/kg	Health Hazard 1
2. Slightly toxic	5–15 g/kg	Health Hazard 1
3. Moderately toxic	0.5–5 g/kg	Health Hazard 2
4. Very toxic	50–500 mg/kg	Health Hazard 3
5. Extremely toxic	5–50 mg/kg	Health Hazard 3
6. Super-toxic	<5 mg/kg	Health Hazard 3

The EPA controls the regulation of poisonous substances in the drinking water and in food, and provides an Integrated Risk Information System (IRIS) database for assessment in their *Exposure Factors Handbook*. They published drinking water standards in 1994 in terms of MCL in mg/L and as a health advisory RfD in mg/(kg day) of intake (Chapter 6). Some typical values are given in table 10.3, which shows a tremendous range of five orders of magnitude between the mildly toxic xylene and the extremely toxic white phosphorus.

One of the best known cases of a massive public disaster took place in the 1950s in the fishing village of Minamata in Japan. The Chisso company began production of acetaldehyde in this village. Acetaldehyde is an intermediate to make ethyl acetate, an important paint solvent. Mercury was used as a catalyst in the oxidation of propane and butane. When waste mercury was discharged into Minamata Bay, it was converted to methyl mercury, which is fat-soluble and much more toxic than inorganic mercury and binds to nerve tissues. In 1956, there was an epidemic of fish and cat deaths, which also led to a human health disaster, since the main diet of this village was fish. The symptoms were of a degenerative nerve disease, which began with numbing of limbs and lips, slurred speech, constricted vision, brain damage, and coma. Many children were born with defects, which were graphically documented by photographers. We must learn from past mistakes.

**Table 10.3 Drinking water standards: MCL and RfD.**  
 (Source: Klaassen 1996)

	<i>MCL, mg/L</i>	<i>RfD, mg/(kg day)</i>
<i>Elements</i>		
Phosphorus, white		0.00002
Thorium	0.002	0.00005
Mercury	0.002	0.0003
Antimony	0.0006	0.0004
Cadmium	0.005	0.0005
Beryllium	0.004	0.0050
Selenium	0.05	0.0050
Silver		0.0050
Nickel	0.1	0.0200
Barium	2	0.0700
Chlorine	4	0.1000
Zinc		0.3000
Strontium		0.6000
Arsenic	0.05	
Chromium	0.1	
Copper	1.3	
<i>Organics</i>		
Aldrin		0.00003
Dieldrin		0.00005
Heptachlor		0.00005
Chlordane	0.0020	0.00006
CCl <sub>4</sub>		0.00007
Chloroform	0.1000	
Lindane		0.00030
Vinyl chloride		0.00050
Carbon tetrachloride	0.0050	0.00070
Benzene	0.0050	
Hexachlorobenzene		0.00080
Naphthalene		0.00400
Trichlorobenzene	0.0700	0.00600
CHCl <sub>3</sub>		0.01000
Chlorobenzene		0.02000
Malathione		0.02000
Dichlorobenzene	0.6000	0.09000
Styrene		0.20000
Toluene		0.20000
Phenol		0.60000
Xylene		2.00000

### 10.3 Environmental Problems

We consider here all the objectionable alterations of nature that go beyond human health and safety. The problems are often divided into regional problems, which are limited to specific regions, and global problems, which affect the entire world.

The regional problems include:

- air quality—lower visibility, odor, irritating gases, particulates;
- water quality—turbidity, color, odor, chemical and biological oxygen demand, temperature, acidity, eutropication, algae bloom, foam;
- solid problems—accumulation, landfill, recycle, hazard;
- ecology—species extinction, biodiversity.

The global problems include:

- ozone concentration—depletion in stratosphere, ozone hole in Antarctica;
- global warming—greenhouse gas emission.

The release of man-made materials into the environment can have numerous harmful effects, as well as beneficial ones, but the cautious and conservative approach is to assume that any change in the environment could be potentially harmful unless proven otherwise. Actually, there is no way to prove that all the changes from the introduction of a product can be investigated within a given time period and are beneficial or neutral; so, in effect, a strict application of this principle would mean a total ban on all product innovations and many existing products.

There are numerous parameters to characterize the environmental effects of a chemical, once it is released either by accident and leakage or by disposal after the end of the usefulness of a product. We need to follow the journey of the chemicals released.

#### Pollutant Emission

We begin from the sources of emission, the substances released, the duration and rate of discharge, and the methods of discharge. Possible sources of emission begin with extraction from nature in mines or farms, from transportation and storage facilities, from manufacturing plants, from users distributed over a large geographic area, and from disposal sites. A plant must keep an inventory of what substances are discharged and whether they are continuous or intermittent. There would be gaseous discharges of flue gases from furnaces, emissions from process units, and leakages from the various plants. There would be liquid discharges of hot water from cooling coils, as well as of waste water with dissolved chemicals and suspended solids and liquids. There would also be solid wastes and discharges. Transportation, customer use, and final disposal also represent sources of emissions.

#### Pollutant Transport in the Environment

There are many modes on how a chemical is transported in the environment. The most important mechanism is the natural flow of air, which is the most rapid method of dispersal and reaches all points on Earth. This is most important for gases and

vapors, as well as for aerosols of solid or liquid droplets in air. A gas with a low boiling point and high volatility will be carried by the wind currents, and a vapor heavier than air will sink to the floor in a building or the ground in the open. The next most important dispersal mechanism is travel by water, including municipal and storm sewers, precipitation run-off, ground water, rivers, lakes, and ocean circulations in the environment. The compounds so carried are soluble in water, fine colloidal particles in suspension, and small particles in mechanical suspension. A liquid or solid with low solubility will float or sink in water, depending on its density. Some of the most important determining parameters on the ranges and speed of transport are:

- melting point, which determines when the material is a solid or a liquid;
- boiling point, which determines when the material is a liquid or gas;
- density of gas, which determines accumulation on ceilings or floors in buildings;
- density for liquid or solid, which determines floating or sinking in water;
- diffusivity in gas or liquid, which determines the local speed of spreading.

### Pollutant Transformation in the Environment

Many chemical and physical transformations can take place for a chemical in the environment, leading to products of transformation. Some of the most important physical and chemical transformations and their parameters are:

- oxidation, which is the reaction with oxygen in air or water and the production of oxidation products;
- hydrolysis, which is the reaction with water and the production of hydrolysis products;
- photolysis, which is reaction with the help of ultraviolet radiation from sunlight and the production of photolysis products;
- biological reactions, which are produced by biological agents, such as bacteria, and lead to biological products of reaction.

A material that degrades in nature gives rise to a set of reaction products. In some happy cases, these reaction products blend back harmlessly into nature. In some unhappy cases, these reaction products, called secondary pollutants, are far more toxic than the parent primary products. A chemical that does not degrade in nature would accumulate and create problems.

### Pollutant Accumulation

A chemical released into the environment may be inert and accumulate, in places such as in air, in water or on land, and in the bodies of plants and animals. A material will distribute among various phases according to thermodynamic affinity leading to equilibrium. The process is called bioconcentration if there is a high ratio of concentration in biological organisms, in comparison with concentration in the environment. A fat-soluble substance would be much more highly concentrated in predatory animals at the top of the food chain; for example, a chemical in low concentrations in insects accumulates into higher concentrations in fish, which in turn accumulates to even higher concentrations in bald eagles that eat the fish. A material with a high octanol/water

partition coefficient would accumulate in fats, such as being stored in animals and plants. A material with a high vapor pressure would evaporate from solid and liquid, and a material with a high Henry's law constant would evaporate from water solution, and be found mostly in the air. A material with a high solubility in water would be found mostly dissolved in water. Sometimes, there are "sinks" in nature where some material accumulates; for example, carbon dioxide may be absorbed by mollusks to make their shells and eventually turn into limestone after they die and are laid down in deep beds.

### Damage and Harm

The products and by-products of industry, and the transformed secondary products, can cause harm by being toxic to men and animals and by creating environmental effects. A material can become a local or global environmental problem, depending on the speed of transport, the transformation activity, and its persistence.

Regional examples of harm include the following:

- The air quality in crowded urban areas often has a man-made origin, arising from combustion for heat, power, and transportation or as emissions from plants. The principal concerns are often the sulfur compounds of  $\text{SO}_2$  and biological  $\text{H}_2\text{S}$ , with bad odor and making acid rain that attacks limestone and metals. Particulate matter, or dust, contributes to respiratory discomfort and harms visibility. The carbon compound CO is toxic, and  $\text{CO}_2$  is a greenhouse gas. The nitrogen oxides and tropospheric ozone contribute to urban photochemical smog. The critical parameter for volatile organic chemicals is the vapor pressure of the substance at room temperature, so that a material with a high vapor pressure is more likely to exceed the TLV and cause harm. The more reactive olefins, diolefins, and aromatics are much more objectionable than the less reactive paraffins. And a heavy vapor is more likely to sink to the floor and harm infants and animals.
- Water quality is essential to health, comfort, and a sense of well-being. We are concerned with the odor, taste, and color in water. When the nutritional elements of nitrogen and phosphorus are released to water, eutrophication can take place; as a result, plants bloom in the once pristine northern lakes that were once lacking in these elements. Suspended solids cloud the visibility in water. Dissolved compounds may be oxidized in water and consume the dissolved oxygen, meaning that there is less available for aquatic life; tests called the chemical oxygen demand and the biological oxygen demand are used to indicate how much of the dissolved oxygen is required to oxidize the organic contaminants and therefore how much might be available for use by aquatic life. The desired oxygen content of water to sustain fish life is 5 mg/L, but active fish, such as trout and salmon, require more oxygen than placid fish, such as catfish. The solubility of oxygen declines rapidly with temperature, so that even if the stream is well aerated, a temperature of 25 °C has enough dissolved oxygen to sustain fish life, but a temperature of 40 °C would be marginal. Organic compounds (such as pesticides and polynuclear aromatics) and heavy metals (such as copper and chromium) are also dangerous.

- Solid waste discharges from chemical plants can represent very large problems, especially from paper mills, plastics plants, and food processing plants. Some materials do not decompose in the environment, and can become burdens when they accumulate. Some polymers have backbones that degrade in nature, such as aliphatic polyesters and polyvinyl alcohols; others do not, such as polyethylene and polystyrene.

Global examples concern material with a fast transport speed and long-term persistence:

- Ozone hole in the stratosphere. The CFC family of refrigerants and air conditioning fluids is designed to be very inert and stable. They persist in the atmosphere and travel to the upper stratosphere, 30–50 km above the Earth, where they are decomposed by ultraviolet rays from the sunlight to form chlorine radicals. The chlorine radicals catalyze the decomposition of ozone  $O_3$ , which protects life on Earth by absorbing much of the harmful solar ultraviolet light. There is now an ozone hole over the Antarctica each spring, which makes human and animal life in the Southern Hemisphere more susceptible to harmful ultraviolet light from the sun.
- Global warming. The global warming potentials of a greenhouse gas depend on light absorption in the visible and infrared regions, and special concerns are reserved for the ubiquitous  $CO_2$ , which is produced in the largest volume as the result of combustion, as well as farming and land use. Some critics believe that global warming can be tied into rising sea levels, inundation of low-lying coastal areas and islands, changes in ocean circulation, and changes in weather patterns.

## 10.4 Green Engineering

When Rachel Carson's book *Silent Spring* was first published in 1962, the initial reaction from the chemical industry was to deny that there was a problem. Gradually, the more enlightened companies realized that the problem was not going to go away, and a more positive approach was required for their own survival and long-term prosperity. In the last four decades, a great deal of progress has been made to make the chemical processing industries more responsive to the basic needs of society for safety, health, and environmental protection. The initial methods of green engineering began with the quick response of remedial clean-ups, which may be compared with the bandage approach to repair damage already done. The more thoughtful approach is to modify the process methods, which is more effective and has greater potential for better solutions. In the long run, product modifications and redesign have the best potential for avoiding problems from the start. Table 10.4 summarizes the examples given below.

### Remedial Clean-Up

The first positive step is the remedial action of cleaning up the environment that has already been damaged by the release of chemicals. This includes the restorations following the disastrous accidents of mercury in Minamata Bay, the Flixborough explosion of a vapor cloud, the Bhopal leaking of methyl isocyanate, and the Valdez oil tanker spill.

**Table 10.4 Examples of green engineering solutions**

<i>Hazards in product environmental cycle</i>	<i>Remedial clean-up</i>	<i>Green process engineering</i>	<i>Green product engineering</i>
Extraction	Restoration of mines, oilfields, forests	Taxol source switch from Pacific yew bark to English yew needles	Lamp oil from whales to petroleum, billiard balls from elephants to celluloid
Transportation	Clean up after oil tanker spills, and truck overturns	Double-hull bottom in oil tanker	
Manufacture	Restoration after plant fire and explosion	PVC plant reduce vinyl chloride concentration in air	Automobile paint from solvent-based to fused polymers
Use		Safety caps on medicine bottles	White paint from lead oxide to titanium oxide; breast implants from silicone oil to salt water
Disposal	Recycle programs, restore landfill sites		Detergents from branched sulfonates to linear sulfonates; refrigerant from CFCs to HFCs; gasoline octane additive from tetraethyl lead to oxygenates

These are very expensive actions, in paying for the damage to injured people and properties and in repairing lost plants and equipment, along with the accompanying loss of public and government good will and confidence. There are also numerous cases of less publicized events of restoring abandoned mines and oilfields, and of cleaning out landfill sites. These programs should be taken as first steps, leading to the more constructive efforts to move upstream to modify the processes and practices of the manufacturing technologies to make them inherently safer and more environmentally friendly.

#### *Extraction from Nature*

Forests after intensive logging are usually unsightly and frequently dangerous. Without healthy roots to hold the soil and banks, the land may be subject to erosion after storms, leading to mud slides. Planting of new trees and repairing of eroded land can restore a logged-out forest, to create a source of timber and paper for the future. Exhausted and abandoned mines and mining structures are ugly and pose cave-in dangers to wandering people and animals, and often produce acid discharges to endanger the land and water. It is important to restore these former extraction sites to how they were before.

### *Transport and Manufacturing*

The most famous remedial case is that of the 1989 Exxon Valdez oil tanker, which ran aground in Prince William Sound, Alaska, and dumped 11 million gallons of crude oil along 1300 miles of shoreline, causing tremendous damage to natural life and to the beaches. A large crew of professionals and volunteers cleaned animals that were smeared with crude oil, erected booms to keep large oil slicks from reaching the shore, spread detergents to disperse smaller oil patches, and bulldozed beach sand to contain the remainder. Total restoration is probably beyond reach, but the beaches and wildlife of Alaska are no longer attracting photographers looking for bad spots. Oil spills from ocean tankers continue to dominate the headlines. There is no substitute to oil as a raw material for transportation fuels and for petrochemicals, and the only solution appears to be tighter regulations plus enforcement. Frequently, the tankers are operated by small firms with inexperienced crew and poorly maintained ships, chartered in Third World countries, and not under the supervision and responsibility of well-organized and financed multinational corporations. International agreements on corporate responsibilities of the well-financed corporations that chartered these ships must be reached, to make sure that the clean-up will be made with adequate resources.

### *Disposal*

The main destinations of solid wastes are landfill and recycling. There has been little progress in the technology of landfill, but there have been much more imaginative approaches to recycle. Recycling can mean remanufacturing, such as shredding old newspapers to make paper for another newspaper or book; it can also mean recycling to basic material, such as to depolymerize a discarded resin into monomers for synthesis into new resins. The platinum contained in automobile catalytic converters is deposited on wash coats of alumina, which adhere on monoliths that are contained in stainless cans. There is a cottage industry of scavengers who go into automobile graveyards to remove the cans and then extract the platinum by using acids and alkalis for reuse in new catalytic converters. The high price of \$760/ounce for platinum makes this industry profitable, so that there is not much discarded platinum, in comparison with the less profitable industry of recycling lower priced paper and plastics. The Iraq War of 2003 has also created a wasteland of broken-down military equipment and ammunition, and the clean up is both dangerous and necessary if Iraq is to be restored to the family of nations.

### **Green Process Engineering**

A slower, but more effective method is green engineering. This begins by keeping the product exactly as before, but makes improvements in the manufacturing process: in the material supplies and equipment, and in manufacturing equipment and practices. The quickest and least expensive way to accomplish process improvements is to do a tailpipe clean up; that is, to treat discharge streams by installing air filters, water scrubbers, and absorption equipment. Next comes modifying the process equipment to make it safer and to use processing procedures that are more tightly controlled. Even more expensive and time consuming are changing the reaction and separation

processing steps, using less dangerous raw materials and intermediates, and installing equipment that is inherently safer. In the long run, the better engineering is to design and build it right the first time, so there will be little or no remedial clean-up to do. Let us cite a number of historical examples of successes, arranged by the hazards posed in the product environmental cycle, from extraction to transport, from manufacturing to use, and then to disposal.

### *Extraction from Nature*

The raw materials of chemical manufacturing are extracted from air, water, and mines, and from the organic nature of plants and animals. This process can lead to disturbances of nature and the extinction of species. The solution may be in finding an alternate source of raw material that does not cause as much disturbance, or a better method of extraction that is not as brutal. Taxol was extracted from the bark of the Pacific yew trees, which grow in only a few geographic areas and take 100 years to reach maturity. The manufacturer has to strip the bark and kill about six trees in order to extract enough taxol to treat one patient. This sets up a conflict between the needs of the patients with advanced breast and ovarian cancer and the desires of the environmentalists on preserving these ancient trees. A solution was sought to find a totally synthetic or a semi-synthetic route to taxol. The most successful solution now is to use the plentiful English yew needles as raw material, without stripping the bark from the Pacific yew trees and killing them.

### *Transportation and Manufacturing*

Many modern oil tankers are now equipped with a double-hull structure. Thus, if the outer hull strikes a rock and is penetrated, the inner hull can continue to hold the oil without spilling. Many dangerous intermediates, such as phosgene used in the production of polyurethane foams for cushions, are now manufactured on a site right next to the polymerization facilities, so it is no longer necessary to make it in one plant and ship it a long distance to another plant. Polyvinyl chloride is important in making water supply pipes, garden hoses, laboratory tubing, medical catheters for intravenous infusion of medicine and nutrients, shower curtains, diaper covers, and raincoats. Vinyl chloride monomer is the basic raw material in making PVC. In the 1940s, vinyl chloride vapor was considered to be safe at 100 ppm in the plant air, but a series of medical discoveries changed the requirement down to 1 ppm. The task of tightening the equipment and procedures to reduce the plant concentration by a factor of 100 was thought to be impossible at the beginning, and would lead to a total shutdown of the production of the very much needed PVC products. But with the dedication and creativity of many process engineers, this draconian reduction in plant air concentration was accomplished within a year.

### **Green Product Engineering**

The greatest potentials of improvement are in green product engineering, in redesigning the product or replacing it with a better product, which would make less impact on

safety and the environment. The best thing to do is to design a product that is inherently safe and friendly to the environment. This is the most satisfying thing to do, to plan the whole product strategy from the beginning, to pay attention to any step that has the potential of creating dangers, and to do all the planning in multi-scales: from molecules at the nanometer scale, to process equipments at the meter scale, to plants on a 100 m scale, and to the global environment on a 10,000 km scale. Instead of trying to fix what was broken, the planners and designers of a new product have the greater challenge of a clean slate to write on, and to do it right the first time.

### *Extraction*

The best billiard balls were made from elephant tusks, which are beautiful and have good bounce. Ivory was also used for combs, buttons, collar stays for shirts, cigarette holders, chess pieces, and many other ornamental articles. Ivory is obtained by hunting and killing elephants to extract their teeth, which is especially valuable from old and mature males. The extinction of elephants was avoided when John Wesley invented celluloid by combining nitrocellulose with camphor.

The best oil for illumination in lamps was whale oil, which is clean and provides a bright flame without smoke. However, the winning of whale oil involves hunting down sperm whales, to extract the spermaceti oil in their heads. The demand for light for a growing world population cannot be supported by hunting down this meager supply, and the solution was in the drilling of the first oil well by Edwin Drake to obtain kerosene, which spared the sperm whales from extinction.

### *Manufacture*

Automobile bodies used to have highly polished enamel paint coats designed to resist sunlight and weather for many years. This finish was made by dissolving enamel in an organic solvent, such as amyl acetate, which then evaporates to form the enamel coat. The evaporated solvent poses both a problem of hazard to plant workers and of pollution of urban air. Solvent-less automobile body coating is a successful substitution, whereby a switch was made to the new product of solvent-less paint, which entails the spraying and adhering of polymer and binder particles on the steel surfaces and then fusing this to form durable coats.

### *Use*

The most important ingredient of a white paint used to be lead oxide. This was objectionable, as it could be swallowed by infants and lead to brain damage. Other white powders with strong covering properties were needed, and the current material of choice is titanium dioxide. Many painters still prefer lead oxide-based paint, as it has a more desirable warmer shade of white and it dries much faster than titanium oxide-based paint. When a mistaken stroke is made by a painter, lead paint that dries quickly can be covered up by another stroke, but titanium paint needs 3 days to dry before it can be covered up by another stroke. A solution is available in quick-drying additives to the titanium paint.

Polyvinyl chloride can be shaped into clear, soft flexible tubing and sheets with the aid of using plasticizers such as diethylhexyl phthalate (DEHP). Polyvinyl chloride resins are used for children's toys, automobile seat covers, and for catheters used for intravenous transfusion of blood and nutrients in hospitals and homes. When DEHP was suspected of leaching out of the products, and of being a carcinogen, substitutions were developed with polyvinylacetate and polyethylene. However, these substitutes are not yet totally satisfactory, as they cannot be steam sterilized, and they are not as clear and flexible as PVC.

Surgeons' gloves are made of natural rubber, and are worn very tight to preserve the feeling of touch in the fingers required in delicate surgery. To slip their hands into these tight gloves, a powder is required to help in sliding their hands in. Natural rubber contains some compounds that can cause allergies, which are absorbed into the corn starch powder; so that when the surgeons remove their hands from the gloves, the corn starch powder flies out into the operating room air, and can cause allergies to operating room attendants and the patient. A new solution is needed for these natural products, perhaps with synthetic rubber and with nonabsorbing powder.

Silicone oils have very desirable visco-elastic properties, and are very stable. They have been used for many years for breast implants, as a cosmetic to improve appearance. There are many lawsuits charging that silicone oil can leak out of the containing sacks and cause adverse physiological reactions. The current customers either abandon the breast implant or use salt water implants, which do not have the same desirable visco-elastic properties.

Dry cleaning does not use water, which swells the fabrics and causes shrinkage and wrinkling. The dry-cleaning establishments use either petroleum fractions or chlorinated solvents to remove oil and grease from garments. The petroleum fractions are flammable and cause urban air pollution when they escape from the shops; the chlorinated solvents, mainly tetrachloroethylene, are not flammable, but they have significant toxic properties. A better dry-cleaning fluid is needed.

### *Disposal*

The 3M company invented Scotchgard as a coating on carpets and fabrics to prevent soiling, and it has been a very successful product used also for microwave popcorn bags, fast-food wrappers, semiconductor coatings, and airplane hydraulic fluids. It is based on perfluoro-octanyl sulfonate (PFOS), which was produced at 10 million pounds per year, leading to a \$300 million business. In the year 2000, the 3M company announced that this product would be withdrawn from the marketplace completely by 2002, despite the fact that there had been no death or illness linked with it, and there is no lawsuit pending. The decision was based on the fact that they found PFOS at up to 6 ppm in the blood of minks and eagles, which is 10 times lower than the laboratory-determined toxic concentrations on rats and monkeys. 3M decided that the safety factor of 10 is not good enough, considering that there is a great deal of species variability, so they are now working to develop non-fluorine-based alternatives.

The household laundry detergents were reformulated to become more biodegradable in the 1960s. The principal detergents used in households were alkylbenzene sulfonates,  $R-O-SO_3^- Na^+$ , where the alkyl group R has 9 to 17 carbon atoms and can have many

side branches. It was made by propylene alkylation of benzene using the Friedel-Crafts reaction, which makes a product with a high degree of branching,  $R-C_6H_4-SO_3Na$ . It turns out that such detergents are not completely biodegradable in nature, as the bacterial digestion of these alkyl chains proceeds by the enzyme  $\beta$ -amylase, which splits off two carbon chains at a time, until it reaches a branch where it is unable to proceed any further. A pile of undigested hydrocarbons remained, which accumulated to form unsightly suds and foams in many lakes and rivers. The problem was solved by removing the cyclic and branched paraffins, by using molecular sieve adsorbants with narrow channels that absorb the narrower normal paraffins but not the wider branched and cyclo-paraffins; the resulting linear alkylbenzene sulfonates are totally digested by the bacterial enzymes.

The CFCs used as refrigerants and air-conditioning fluids are stable in the atmosphere after they are released; they subsequently travel to the stratosphere to release chlorine radicals in the sunlight, which catalyze the destruction of the ozone needed to protect Earth from the sun's ultraviolet rays. The solution was to replace the CFCs by HFCs, which do not contain any chlorine atoms. Tetraethyl lead was once the leading additive in gasoline and was used to boost the octane number, but it posed a danger to the development of the brain in the young when ingested. It was replaced by increasing the fraction of olefins and aromatics in gasoline, which also later became objectionable as air pollutants. Then these were replaced by oxygenates, such as MTBE, which also became objectionable, as it is partially soluble in water and can contaminate ground water. A better substitute is needed.

Lubricating oil in automobiles needs detergents to keep the engine clean from deposits from the products of combustion. The detergent is normally a metal soap of an organic acid, which would produce an ash of metal oxides in the combustion process. This ash becomes an air pollutant, so that ashless detergents have to be developed. The textile finishing processes employ metal salts in organic complexes to protect the outdoor fabric from microbial degradation, usually with metals such as copper, chromium, zinc, cadmium, mercury, lead, and nickel. The effluents from textile plants cause water pollution. Alternate antimicrobials for outdoor fabrics are sought.

## Reference

Klaassen, C. D., ed. 1996. *Casarett and Doull's Toxicology*. New York: McGraw-Hill.

## Further Reading

- Albert, A. 1973. Structure-activity relations. In *Selective Toxicity: the Physico-Chemical Basis of Therapy*. London: Chapman and Hall; chapter 7.
- Allen, D. T. and D. R. Shonnard. 2001. *Green Engineering: Environmentally Conscious Design of Chemical Processes*. Upper Saddle River, NJ: Prentice-Hall.
- Davis, M. L. and D. A. Cornwell. 1991. *Introduction to Environmental Engineering*. New York: McGraw-Hill.
- Ehrlich, P. R., A. H. Ehrlich, and J. P. Holden. 1977. *Ecoscience: Population, Resources, Environment*. San Francisco, CA: W. H. Freeman.

- Goldstein, A., L. Aranow, and S. M. Kalman. 1974. Structure-activity (chapter 1); Drug absorption and distribution (chapter 2). In *Principles of Drug Action: The Basis of Pharmacology*. New York: Wiley.
- Kim, W. B., U. A. Joshi, and J. S. Lee. 2004. Making polycarbonate without phosgene. *Industrial & Engineering Chemistry Research* 43(9): 1897-1914.
- Kirk-Othmer Encyclopedia of Chemical Technology. 1994. New York: Wiley-Interscience.
- Klaassen, C. D., ed. *Cassarett & Doull's Toxicology, the Basic Science of Poisons*. 5th edition. New York: Macmillan.
- Louvar, J. F. and B. D. Louvar. 1998. *Health and Environmental Risk Analysis*. Upper Saddle River, NJ: Prentice-Hall.
- Lyman, W. J., W. F. Reehl, and D. H. Rosenblatt. 1990. *Handbook of Chemical Property Estimation Methods*. Washington, DC: American Chemical Society.
- Masters, G. M. 1998. *Introduction to Environmental Engineering and Science*. Upper Saddle River, NJ: Prentice-Hall.
- Nader, R., R. Brownstein, and J. Richard, eds. 1981. *Who's Poisoning America*. San Francisco, CA: Sierra Club Books.
- Patty's Industrial Hygiene and Toxicology. New York: Wiley.
- Sax, N. I. and R. J. Lewis, 1987. *Hazardous Chemicals Desk Reference*. New York: van Nostrand Reinhold.
- Schneider, S. H. 1989. *Global Warming*. San Francisco, CA: Sierra Club Books.
- Seinfeld, J. H. and S. N. Pandis. 1998. *Atmospheric Chemistry and Physics*. New York: Wiley.
- Ullmann's Encyclopedia of Industrial Chemistry*, vols. B7 and B8, Environmental Protection and Industrial Safety. Weinheim: VCH.
- Verschueren, K. 1996. *Handbook of Environmental Data on Organic Chemicals*. New York: van Nostrand Reinhold.
- Yen, T. F. 1999. *Environmental Chemistry*: vol. 4A, *Essentials of Chemistry for Engineering Practice*; vol. 4B, *Chemical Principles for Environmental Processes*. Upper Saddle River NJ: Prentice-Hall.

## Websites

- Environmental Protection Agency (IRIS) Integrated Risk Information System. Online.  
<http://www.epa.gov/ncea/iris.htm>
- Google. <http://www.google.com>
- Green Peace. <http://greenpeaceusa.org>
- National Institute of Occupational Safety and Health (NIOSH), Pocket Guide to Chemical Hazards. Online. Available: <http://www.cdc.gov/niosh/database.htm>
- National Library of Medicine, Hazardous Substance Data Bank. Online. Available:  
<http://chem.sis.nlm.nih.gov/hsdb>

## Discussion Topics

1. Which leading causes of death in the United States, listed in table 10.1, can be linked with certainty to products of the CPI? Which causes can be linked plausibly to the CPI? How would you estimate the drop in annual death rate if the CPI ceased to exist, and is this a large percentage change?
2. In your mind, what are the greatest evils in society today, and list them in the order of urgency and priority, as you have a limited budget to devote to

these causes? How would you inform and lobby your congressmen and president to prioritize federal funds for these causes? Some candidates include:

- terrorists from Iraq and Afghanistan, nuclear weapons in Pakistan, Iran, North Korea;
  - conflicts between Israeli and Arabs, Hindus and Moslems, Irish Catholics and Protestants;
  - civil and human rights violations in Sudan, Liberia, Saudi Arabia, Tibet;
  - clean drinking water for children in Bombay, hunger in North Korea, HIV and AIDS in Minneapolis;
  - global warming, ozone hole, vanishing Amazon rain forests, extinction of Texas horned toads;
  - air pollution in Los Angeles, Calcutta, Cairo, Tokyo, Mexico City;
  - unemployment in Silicon Valley, Moscow, Nairobi;
  - SARS outbreak, West Nile disease, Lyme's disease, return of malaria epidemics.
3. Some thinkers, such as Professor Milton Friedman of the University of Chicago, believe that corporations should be responsible only to the stockholders, and have no other responsibilities except obeying the law of the land. Taking his opinion into consideration, how much should a chemical company do to protect health, safety, and the environment? How much is too little or too much, and what is just right?

## Exercises

Let us return to three of the companies of the CPI that you encountered in Chapter 9: DuPont, ExxonMobil, and Merck.

1. What are their attitudes and reputations towards protection of the health, safety, and environment? How do they discuss their policies in their annual reports? Do they report pending lawsuits and settlements?
2. What are their recent successes in green engineering solutions, which they boasted about in their annual reports? Refer to table 10.4 and classify these green engineering solutions according to the hazards posed in product environmental cycles, and whether they are remedial clean-up, green process engineering, or green product engineering.

# 11

## Product Design

### **11.1 Bottled Drinking Water: Fluid Compound**

Identifying market and technology  
Generating alternate concepts  
Investigations and solutions  
Evaluation and selection  
Detailed design

### **11.2 Missile Fuel: Homogeneous Fluid Solution**

Identification of customers and technology  
Alternate product ideas  
Investigations and solutions  
Evaluations and selection

### **11.3 Adsorbents for Separation: Solid Material**

Identify market and technology  
Generate alternate product concepts

Investigations and solutions  
Evaluation and selection

### **11.4 Controlled Drug Release: Composite Solid**

Identify market and technology  
Generate alternate concepts  
Investigations and solutions  
Evaluation and selection

### **11.5 Evaluation, Selection, and Business Plan**

Multifactor optimization  
Finance, return, and risk  
Business plan

#### **References**

#### **Further Reading**

#### **Design Project**

Project selection  
Project research and analysis  
Project reporting

A good design is both art and science. Creativity is needed to come up with a product concept that has either not been considered before or has been considered and abandoned. Discipline is needed to investigate and evaluate all the details to make sure that they fit together, which is an exercise in multiscale optimization: the product structure is in the molecular scale of nanometers to micrometers, the process equipment is in centimeter to meter scales, marketing and finance are played in the meter to hundred kilometer scales, and environmental concerns are on a global scale.

The work of product innovation usually involves a discovery phase followed by a development phase, which was described at the end of Chapter 2. Some of the work involved can be described as:

1. identifying customer needs and technological capabilities;
2. generating a number of broad ideas and concepts for products and processes;
3. investigating and solving problems encountered in customer needs and plant operations;
4. evaluating alternate concepts, and selecting winning designs for commercialization;
5. drawing detailed designs and plans for the products and the processes.

A design team is critical to the success of the innovation work, involving people from marketing and research, plus critical inputs from manufacturing and finance. In the narrow sense, the design team is concerned with the work in (5), in drawing detailed designs and plans for the products and the processes. In the broad sense, the work of design begins in (2) by generating concepts, and makes significant contributions in (4) when evaluating alternate concepts. The decision to commercialize involves the commitment of major financial and manpower resources, and usually involves a presentation of a business plan at a high-level meeting, such as with the executive committee of a company.

Different authors have come up with a variety of ideas on the ideal organization and phases of the innovation work. Ulrich and Eppinger (2000) suggested a template that may be more appropriate for mechanical engineering products that do not require research breakthroughs. Cussler and Moggridge (2001) suggested a four-step schedule involving needs, ideas, selection, and manufacture. The best template for a particular innovation may be best modeled after successful historic innovation cases that are fairly similar to the present case.

A product is very seldom a single pure compound, and often consists of a formulation with several ingredients: the principal ingredients provide most of the functionality of the product, the auxiliary ingredients provide the supportive functions, and the unintentional impurities must be controlled. These different ingredients of formulation may exist in a single fluid phase, or coexist as several phases in a complex fluid or a solid, or as a composite solid. As the ingredients and phases increase, the designers have more and more design parameters to specify and to control. The simplest product is a single compound, such as table salt, where the principal ingredient is NaCl, together with possible added iodide salts and impurities coming from the sources of brine or rock salt. The bottled drinking water has the principal ingredient H<sub>2</sub>O, which will be discussed in more detail in section 11.2.

A homogeneous fluid solution, such as gasoline, perfume, and missile fuel, is the next in complexity. Gasoline is a blend of different hydrocarbon streams to provide the volatility and octane number appropriate to the applicable climate and engine compression; additives are used to increase the octane number, to provide the required oxygen content, and to improve the detergency to keep sparkplugs and valves clean; the impurities to control include sulfur, benzene, other aromatics, olefins, and traces of lead. A perfume consists of a blend of natural products, such as flowers and musk extracted by pressing or by solution with alcohols and hydrocarbons, as well as chemicals such as benzoic acids and ketones. The formulation includes a top note or high

volatility components that attract immediate attention, a middle note or medium volatility components that maintain interest, and finally a base note that lingers for hours after the lighter notes have dissipated. The missile fuels enable the delivery of bombs and weapons at a distance, and will be discussed in detail in section 11.3.

A complex fluid may contain several phases, such as whipped cream, mayonnaise, milk, and paint do. Paint consists of binders or resin, a pigment and inert extenders, a solvent, and additives. The binders form a coherent film that adheres to the wall and forms a barrier; the pigments imparts the desired color and refractive index, and the extenders provide low-cost bulk volume; the solvent keeps the binders and pigment particles from clumping together, and provides fluidity during application; and the additives include thickeners, biocides, driers, and de-foamers. Examples of solid products includes chalk, talcum powder, and industrial absorbents. Consider the zeolites, which are molecular sieves that have pores and channels with diameters comparable to many important chemicals. They are used to separate molecules that are closely related in structure and boiling points, by admitting and excluding molecules of different diameters. Examples of effective separation methods include separating the normal paraffins from the iso-paraffins, as normal paraffins in the all-trans configuration have a smaller diameter than the iso-paraffins, which is important in the production of the biodegradable detergents needed to prevent the accumulation of household detergents discharged into lakes and rivers. The three xylenes have almost identical boiling points, but *p*-xylene is needed for the manufacture of terephthalic acid, leading to polyester resins and fibers. The most effective separation method uses a zeolite with the appropriate channel diameter, as *p*-xylene has a smaller diameter than the *o*- and *m*-xylenes.

Composite solids include: high-strength fiber-reinforced polymers for golf clubs and tennis rackets; flexible barriers, such as GoreTex jackets, that repel water droplets, but which permit the passage of air; and controlled drug-release devices, which will be discussed in section 11.4.

## 11.1 Bottled Drinking Water: Fluid Compound

Let us start from the simplest product of bottled drinking water, which consists of a single fluid compound in possession of all the principal functionality required, plus minor amounts of unintended impurities and intended additives. The number of design variables that is available to the design engineers is very small.

### Identifying Market and Technology

Let us consider bottled drinking water for distribution in the Northeast region of the United States. Human physiological need for water is about two quarts per day. The primary market is for customers who are trendy urbanites and suburbanites who have comfortable incomes, and are not happy with drinking tap water, which costs less than 0.1 cents per gallon. Other customers live in isolated communities without municipal water, and depend on their own wells or natural streams. Municipal tap water is manufactured from streams and wells for many purposes, including such large-volume and mundane needs as washing dishes and sprinkling the lawn, so the designers of municipal water do not make the overkill to make sure that every drop of the water is

fit for the drinking pleasure of urban sophisticates. The market of bottled drinking water is dependent on the economic prosperity of the buyers, as it is not an essential item like bread. The buyer with reduced circumstances can first downgrade from premier bottled drinking water to basic bottled drinking water, and then to boiling tap water.

The safety standards of drinking water are regulated by civil authorities. The U.S. EPA published the “National Primary Drinking Water Regulations” (table 11.1), which are the legally enforceable standards that apply to public water systems. The primary standards protect public health by limiting the concentration levels of contaminants in drinking water: microorganisms such as *Escherichia coli* and viruses, disinfectants such as chlorine, disinfection by-products such as chlorites, inorganic chemicals such as arsenic and lead, organic chemicals such as benzene and PCBs, and radionuclides

**Table 11.1 National Primary Drinking Water Regulations, US EPA 2002**

Contaminants	Maximum contaminant level	Potential health effects
<i>Microorganisms</i>		
Coliforms total	More than 5.0% samples total coliform-positive in a month	Used to indicate presence of other harmful bacteria
Turbidity	Below 5 nephelometric turbidity units (NTU)	Association with disease-causing microorganisms such as viruses, parasites, bacteria
Viruses (enteric)	Require 99.99% removal/inactivation by disinfection and filtration	Gastrointestinal illness, diarrhea, vomiting, cramps
<i>Disinfectants</i>		
Chlorine	Maximum residual disinfectant level (MRDL) 4.0 mg/L	Eye/nose irritation, stomach discomfort
<i>Inorganic chemicals</i>		
Arsenic	0.010 mg/L	Skin damage, circulatory system
Cadmium	0.005 mg/L	Kidney damage
Copper	1.3 mg/L	Corrosion of household plumbing
Fluoride	4.0 mg/L	Bone pain and tenderness
Lead	0.015 mg/L	Infants delay physical and mental development
Mercury	10 mg/L	Kidney damage
Selenium	0.05 mg/L	Hair or fingernail loss, numbness in fingers or toes
<i>Organic chemicals</i>		
Benzene	0.005 mg/L	Anemia, decrease in blood platelet
Chlordane	0.002 mg/L	Liver or nervous system
Dioxin	0.00000003 mg/L	Reproductive difficulties
PCB	0.0005 mg/L	Skin change, thymus gland, immune deficiency

such as  $\alpha$  and  $\beta$  particles. The EPA also publish the “National Secondary Drinking Water Regulations,” which are non-enforceable guidelines, to recommend limits to substances that may cause cosmetic effects (such as skin or tooth discoloration) and esthetic effects (such as taste, odor, color). The EPA recommends them, but the states have the option to adopt them, or not, as enforceable standards.

Common drinking water problems (EPA 2001) that are noticeable to consumers include:

- Unpleasant feeling. Hard water gives the symptom of soap curd and scum in bathtubs, and is caused by calcium and magnesium salts.
- Bad smell. Water may have a musty, earthy, or wood smell. Rotten egg odor is often associated with tarnish in silverware, and is caused by dissolved hydrogen sulfide gas and sulfate-reducing bacteria. Detergent odor and foam is caused by seepage of septic discharge into underground water supply. Gasoline or oil smell is caused by leaks in fuel oil tanks or gasoline tanks. Methane gas is caused by decaying organics. Phenol smell is caused by industrial waste seeping into water supplies.
- Unpleasant taste. Salty or brackish taste is caused by high sodium content. Metallic taste is caused by acidic water (pH 3.0–5.5), and heavy concentration of iron, lead, and copper.
- Bad appearance. Turbidity is caused by suspended matter, which may contain disease-causing microorganisms. Discoloration that leaves stains can be caused by dissolved or colloidal iron, tannin, or humic acid from decaying vegetation, and excessive salt content.

Many local waterworks and private wells violate these government regulations, and create noticeable problems, thus creating a market for customers who can afford bottled water. A trip to an upscale food market, such as Wegmans in Princeton, New Jersey, will show that bottled drinking water is a thriving market that is highly segmented by price and prestige.

At the high end are the imported name brand spring waters that are heavily advertised, such as Perrier, Evian, San Pellegrino, and Ferrarelle. Evian is not carbonated, and advertises itself as “Natural spring water from the French Alps, naturally filtered through layers of glacial sand, gushes within the heart of the French Alps to reach us in its original purity, untouched by man, perfected by nature.” Perrier advertises that it is carbonated by nature. The Ferrarelle from Italy is advertised as “Natural spring water from the ancient volcanic basin south of Rome, fortified with gas from the spring to produce small bubbles, to give a smooth and pleasurable drink.” Less well-known and advertised brands include: Fonteviva from Italy, Voss from Norway, São Cristóvão is *água de nascente* or spring water from Serra Montemuro from northern Portugal, and Fiji Natural Artesian is from Fiji Island in the South Pacific. They sold for around \$2.00 per quart at the Wegmans Market in 2003.

Next in price and prestige are the domestic name brand spring waters. Saratoga has long been known for its spring in New York frequented by aristocratic families, but it is not heavily advertised and promoted. The Poland Spring is from Maine, and is heavily promoted in the Northeast. Other brands on the shelves include: Deer Park spring water from the Appalachian Mountains of Pennsylvania, “where you will discover crisp, clean air and spring”; Calistoga has its origin in geysers in upper Napa Valley,

California; Minere is advertised as “natural mineral water, 100% natural, contains valuable minerals your body needs everyday.” Their appeal is that nature is the designer, and chemical engineers are not involved, and it is evident that the chemical compositions are not prominently promoted as the main sales appeal. Undefined minerals in the water are often mentioned as positive aspects in the advertisement, although it is seldom clear which minerals at what concentrations are supposed to be beneficial to the customers. There are also special-occasion waters, such as for the inauguration of Carol T. Christ as the 10th president of Smith College in 2002; the water served was commissioned natural spring water from Willington, Connecticut, which does not have the glamour of a pristine gushing mountain stream or an enchanted fountain far from human habitation. Saratoga sells for \$2.00, and the rest sell for around \$1.00 per quart.

The next group has a reverse snob appeal of being manufactured with up-to-date technology, and are essentially cleaned-up municipal tap water. The bottled water Crystal advertises itself as “ozone treated,” with no mention of a source in a mountain spring. Desanni advertises that it is “filtered for purity, using state of the art treatment by reverse osmosis, and enhanced with minerals for a pure fresh taste.” Deja Blue is advertised as “using state of the art purification systems including reverse osmosis, carbon filtration, and ozonation,” and it is produced in Plano, Texas. These words are in very small print, and are not emphasized in the advertisement, so that most consumers would not be aware of their origins. Surprisingly, they sell for around \$0.88 per quart.

Finally, there are the house brands of each supermarket, with very little description of their source and treatment. Wegman’s Spring Water comes from Concord, New York, and sells for \$0.17 per quart, which is 10 times less than the premium imported water. The source may be municipal or well water, with very little claim to high-tech processing, and is the lowest cost option to tap water.

### Generating Alternate Concepts

What is a good business opportunity in bottled drinking water? How do we design a bottled water where chemical engineering skills would be appreciated, and where there is a market niche? The designers need to identify a market segment for the product, the customers, and the performance required. Each water design has a set of physical and chemical properties that can be measured in the laboratory, but it may be more important to anticipate customer pleasure, which means a set of more open-ended and intangible performance parameters.

The imported and domestic name brand drinking waters compete on factors such as perceived sophistication and snob appeal, and clever advertising, but little need for chemical engineering. A premium bottled water has to make the user feel good to justify \$2 or more per bottle; therefore, it has to look good, smell good, taste good, and make the user feel exclusive and sophisticated. We may consider the water from an even more remote and pristine source, such as from the ice and snow of South Pole and call it the “Polar Express,” or from the eternal ice from the Himalayas, which can be called the “Shangri-La.” Another approach is to fortify the water with

minerals that are perceived to be beneficial and tasty, even if there is no common agreement on what they are. The technology of obtaining Antarctic and Himalayan water may involve strenuous expeditions in hostile environment and difficult transportation chains.

The highly engineered waters compete on perceived technical superiority and purity, and have a sufficiently high price to warrant consideration. The \$1.00 processed water seems to be a reasonable target for less fussy drinkers, which requires chemical engineering skills and commands sufficient price to make some profit. We may start from the local tap or ground water, and see what is wrong with it and how we can improve on its properties. Among the high-tech purification methods, which ones or combinations would give the most pleasing product, or the least expensive product? The \$0.20 water with very little processing may be just right for the people who do not have reliable municipal water, and who are more concerned with health than with appearing to be sophisticated. A local supermarket chain may be persuaded that there is a market for an economy product, manufactured from tap water that sells for \$0.20 per quart, as well as a medium-priced product of greater specification that sells for \$1.00 per quart.

Taking the target to be an esthetically pleasing water that does not contain too much objectionable impurities, the designers start with the local tap water, and set a series of targets for the reduction of objectionable properties. There are to be separate targets for color, odor, chloride, sulfate, and hardness. We may start by putting a consumer panel together and ask them what is objectionable about the local tap water, and what they may be willing to pay to solve these problems. The performance requirement should certainly include the absence of anything that is noticeable and esthetically not pleasing to a fussy drinker, including discoloration, noticeable smell, and unpleasant taste to meet or exceed the Secondary Regulations. We may use the Evian or the Ferrarelle as a reference; but they are very different, as the ion concentrations in a "mineral water" are set by the rocks that they percolated through, and this is different for each spring. Since the word "mineral" has a certain cachet to it, the optimal concentrations of calcium and magnesium are not necessarily the lowest concentrations, and even sulfur can be turned into an asset. However, pathogenic microbes and toxic metals must be much below regulations.

### Investigations and Solutions

The sources of water include mountain streams, natural spring water, drilled well water, flowing water from rivers and lakes, and the municipal tap water. They need to be analyzed for their physical properties, as well as chemical and biological impurities. The principal methods of treatment to improve the water include:

- filtration, reverse osmosis
- chlorination
- activated charcoal treatment
- ion exchange, water softening
- soda ash or citric acid to raise or lower the pH
- ozone treatment, ultraviolet light treatment.

Serious scouting should be made on what are the treatment methods used by other water companies, or suggested in trade publications or meetings, and suggestions by university professors and independent researchers.

The first level of treatment, with sand filters and chlorination to remove suspended matters and disinfection of pathogens, may be good enough for the low-cost water. The removal of discoloration and bad smell is accomplished by activated charcoal absorption. Ozone and ultraviolet treatments are much more expensive for the removal of microbes and organic matter, and should be considered only when necessary to solve a technical problem, or to satisfy an advertisement need. Reverse osmosis is the most effective method used to recover clean water from brackish water, and to remove inorganic minerals such as sodium, copper, iron, and zinc. The removal of calcium and magnesium ions can be accomplished by the method of ion exchange with sodium, which would also increase the sodium concentration, and could cause objections. Different levels of treatment require a variety of costs, and can produce different levels of customer satisfaction.

The R&D people make sample batches of these designer waters, or acquire them from outside contractors. Laboratory tests are then performed to see whether these sample batches indeed have the chemical compositions and biological qualities that we targeted. Consumer panels are established and tested to determine whether these prototypes are indeed more pleasing than the original stock, and whether further processing would be needed. This panel should not simply be staffed by anyone who can be persuaded to spend time with us, but should be people who resemble the potential customers, who are knowledgeable and are fussy about the water that they drink. We ask them to rate these designer waters, against name brand spring waters Evian and Perrier, competing high-tech and low-tech processed waters, and against tap water. We ask them whether they like our designer waters in comparison, whether there is room for improvement, and what they would be willing to pay.

### Evaluation and Selection

When we receive these test results from laboratory and field testing, we reconsider the entire slate of designer waters to see whether there are some other options that we should explore. When we are finally satisfied with our investigations, or run out of time and budget, we begin the process of evaluations and assessment of all of these options, and see whether there are one or more options to recommend. The candidate should have the properties of: a potential market niche with defined customers, a defined process technology to make the product, and a financial plan that would make a suitable profit.

We may end up recommending a formulation for a low-priced “drinking water” under the house brand of the local supermarket, which is manufactured with simple separation steps from tap water. It will certainly be esthetically more pleasing than tap water, with or without carbon dioxide to give it the fizz, and it will have less objectionable appearance and smell than tap water. It is aimed at a market niche of people who are willing to pay a little more than tap water to ensure health, but not enough for Evian and Perrier. We will discuss the quantitative methods of evaluation and selection in section 11.6, and the presentation of the recommendations as a business plan in section 11.7.

## Detailed Design

The detailed process design is familiar to students of chemical engineering, and includes specifying: the source of the raw material water; the equipment to be used, such as filtration, reverse osmosis, charcoal absorption, ozone treatment, ion exchanger, and pumps; the processing conditions, such as flow rates and temperatures; and the plant flow sheet. The detailed product design plan for this simplest of products includes the composition of this bottled water, with special attention to the concentrations of compounds such as sodium and carbon dioxide, suspended matter, and microbes, with special emphasis on the appearance and smell.

### 11.2 Missile Fuel: Homogeneous Fluid Solution

The product that is a homogeneous fluid solution has different components to perform various functional roles. The product designer has available more design variables to manipulate, and needs to balance their proportions to achieve the desired customer performance and company profit. Let us consider the design of a guided-missile fuel intended for the Tomahawk, which was used prominently in the opening phase of the invasion of Iraq in 2002.

#### Identification of Customers and Technology

The customers of missile fuel are the navy and the air force, who use them during a war to hit enemy targets from a launching platform that is at a safe distance to minimize retaliation. The navy and the air force are giant customers, with very deep pockets and have an excellent staff of researchers and test sites for measurements of fuel properties and test sites for rocket firing that are unsurpassed. In the civilian market, a fuel with a 10% improvement in performance may command a premium of 5% in price. But in the military market, a fuel that can make a fighter fly 10% faster than the enemy fighter gives an edge in maneuvering that often means the difference between victory and death; the life of the pilot and the capital cost of the fighter are worth a premium of many-fold increase in fuel price. The same can be said for a missile fuel that can increase the range, so that the missile can reach more remote targets without the danger of being hit by enemy fire. The purchase of missile fuel is increased during anticipated or actual war, and is decreased during long peace. It is unthinkable not to give our soldiers the best weapons and increase their chance of survival, so the sale of missile fuel is relatively indifferent to the national economy or federal budget deficits.

The launch platforms for the navy are ships, such as guided missile launchers and cruisers that are sailing or anchored at a safe distance from the possibly hostile shore. The Tomahawk has a cruise range that is nominally 690 miles, which can reach targets near the shorelines of accessible and navigable waters, such as Baghdad, Teheran, and Pyongyang; but, during times of war, passage through narrow straits may be difficult or prohibited, so that the Black Sea and the Persian Gulf may not be available.

The Tomahawk has a length of 20 ft and weighs 2650 lbs, equipped with a turbojet engine with a speed of 550 m.p.h. and a fuel tank of 150 gallons. In a long and slender body, the front section is the guidance instrumentation, the next section is the high explosive bomb, followed by the fuel tank, and the last section is the turbojet engine.

The capital cost of a missile is around \$0.5 million. It is very desirable to extend the range from 690 miles, to increase the ship-to-shore distance, to reach even more remote targets, and to provide overlapping fire from several shores. The fuel should have the highest possible heat of combustion. Ships are rocking and pitching platforms that are frequently working in tropical waters, and the danger of a spill leading to disastrous fire and explosion must be guarded against. The fuel needs to have a flash point that is much higher than any temperature possibly encountered in the Red Sea during the summer.

There are many deep inland targets that are beyond the reach of the cruise naval missile, such as Timbuktu, Denver, Winnepeg, and Urumqi. The air force B-2 can fly 6000 miles from a base in Missouri, the B-52 can fly 3600 miles from the British island of Diego Garcia in the Indian Ocean, and the B-1 can fly 1200 miles from Oman. Such intercontinental flights are carried out at much higher altitudes to avoid detection and interception. The air force can place six air-launched cruise missiles, such as the AGM-86, on the B-52 bombers. This increases the bombers' ability to penetrate and to overwhelm enemy defenses. This air-to-surface missile is also 20 ft long, with a wing span of 12 ft and a diameter of 25 in, and weighs 3150 lbs. It has a similar speed of 550 m.p.h., and carries a price tag of a \$1 million. Since the temperature drops to  $-60^{\circ}\text{F}$  at above the tropopause of 30,000 ft elevation, the fuel should not have a high viscosity and should not solidify at these temperatures. Thus, the air force jet fuel is expected to have lower viscosity and lower freezing point, but is more relaxed about flash point than the navy jet fuel.

The early missile fuels were the JP-4 and JP-5 refined petroleum distillates used in turbine-powered aircrafts. The cost range is around \$2.00 per gallon, so that a tank of fuel costs \$300, which is a tiny fraction of the hardware cost.

### Alternate Product Ideas

Using the JP-4 derived from petroleum kerosene as the starting point, one concept is to investigate other petroleum hydrocarbons, perhaps fortified with additives. It is well understood that the heats of combustion of hydrocarbons are higher with compounds that are richer in hydrogen in ratio to carbon. Thus, in paraffins with a molecular formula of  $\text{C}_n\text{H}_{2n+2}$ , the highest H/C ratio is found to be four in methane and three in ethane, which is too volatile. The heats of combustion of the six-carbon compounds, measured in joules per gram, are shown in figure 11.1, and ranked thus:

hexane > hexene > cyclohexane > decalin > tetralin > benzene > naphthalene

On the other hand, the Tomahawk has a fixed tank of 150 gallons, so the appropriate measure of performance is not the gravimetric heat of combustion measured in joules per gram, but the volumetric heat of combustion measured in joules per liter. The cyclic compounds with a ring have much higher densities than the acyclic linear or branched compounds in joules per liter, and the compounds with two rings are even higher than the one-ring compounds, see figure 11.2. The ranking in volumetric heats of combustion is almost the inverse:

naphthalene > tetralin > decalin > benzene > cyclohexane > hexene > hexane

Thus, a better concept is to screen the heaviest compounds, or to add components to make the fuel heavier. Since two-ring compounds are better than one-ring compounds,

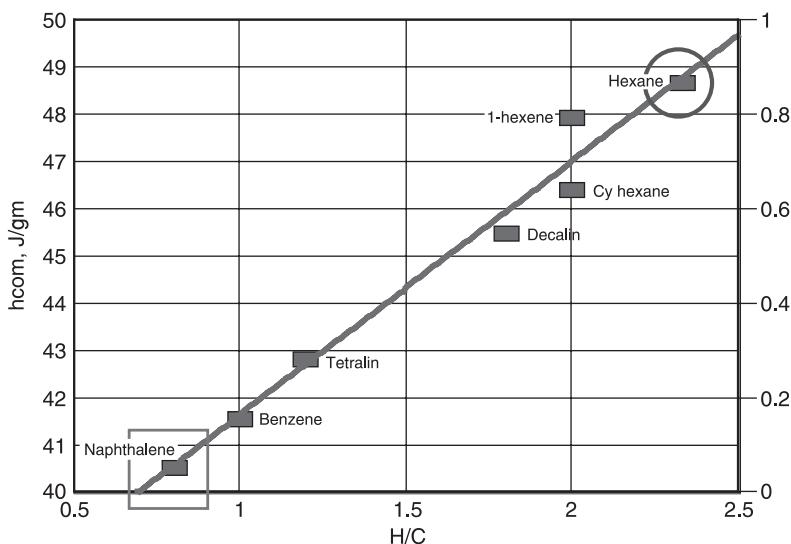


Figure 11.1 Gravimetric heating value and H/C ratio

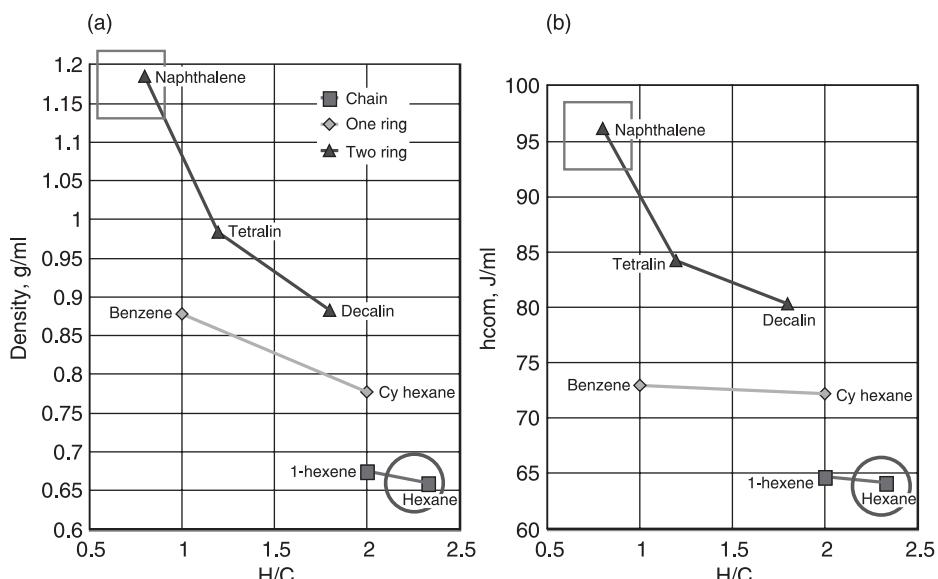
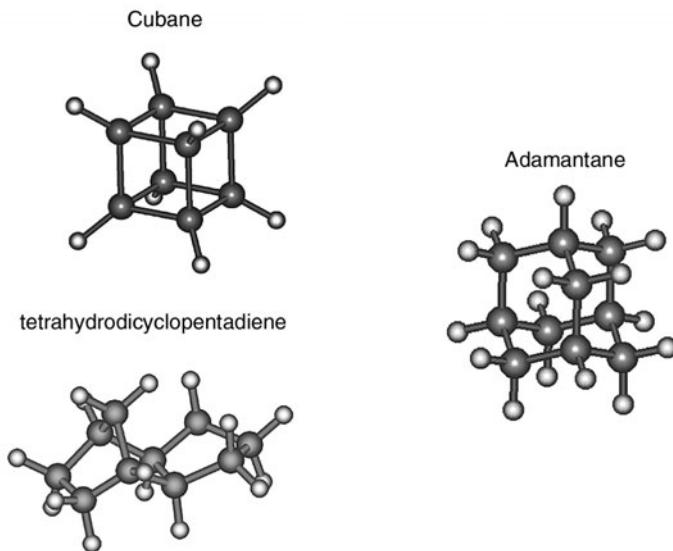


Figure 11.2 (a) Density and (b) volumetric heating values versus H/C ratio



**Figure 11.3** Molecular structure of the multicyclic compounds cubane, tetrahydronaphthalene and adamantane

it would be logical to consider multiple-ring compounds, such as cubane, tetrahydronaphthalene and adamantane. Their multi-ring structures are shown in figure 11.3. Another approach is to add combustible metallic powders to the fuel, such as boron and aluminum.

### Investigations and Solutions

After a number of investigations on these product concepts, useful data are collected and compared (Chung et al. 1999). The fuel RJ-4 was synthesized by hydrogenating di(methylcyclopentadiene) to yield a mixture of *endo*- and *exo*-tetrahydronaphthalene. It has 16% higher energy content than JP-4 or JP-5, and was used in the navy Talos missiles. The *endo*-dicyclopentadiene is used as the starting material with hydrogenation to yield JP-10, which is an improvement and became a standard fuel. A three-component blend based on JP-10, with added methylcyclohexane to increase volatility and RJ-5 to boost energy content, was another option at a high cost.

The idea of adding metallic powder ran into a number of problems. It is difficult to keep the metallic powder suspended and uniformly distributed in the fluid, so that jelled fuels were developed that are ready to fire at any moment in time and place. The metallic powder also exerts a tendency to erode valves and nozzles. Sometimes the combustion is incomplete, and produces solid exhaust particles.

Cubane,  $C_8H_8$ , is an exotic compound that is difficult to make in large quantities. On the other hand, two natural sources of adamantanes have been discovered: in gas fields offshore in Mobile Bay in Louisiana and in Alberta, Canada. The adamantanes come as

**Table 11.2 Properties of selected missile fuels**

	<i>JP-4 kerosene base</i>	<i>JP-10 THD CPD base</i>	<i>RF-1 adamantane base</i>	<i>RF-4 tri-adamantane base</i>
Average carbons	9.5	10	12	18
H/C ratio	2.0	1.60	1.50	1.33
Density, g/mL	0.77	0.94	0.92	1.09
M.p., °C	< -58	< -79	-54	-40
Flash point, °C	-29	54	74	> 126
Viscosity, cP	4.5 at -40 °C	19 at -40 °C	22 at -18 °C 180 at -54 °C	4538 at -15 °C 99999 at -54 °C
MBTU/gallon	118	142	140	159

adamantane ( $C_{10}H_{16}$ ), diamantane ( $C_{14}H_{20}$ ), triamantane ( $C_{18}H_{24}$ ), and their methyl- and ethyl-substituted counterparts. They are complex mixtures of caged, diamond-like polycyclo-alkane molecules. These compounds have very high heats of combustion, but they also have very high melting points and consequently very high viscosities as well. Diamantane has a melting point of 240 °C; this is due to its high degree of symmetry, which is equivalent to methane or diamond. When the symmetry is spoiled by the addition of a few methyl groups, the melting point drops to something more like -54 °C. They are suitable as blending feedstock for high-energy fuels.

### Evaluations and Selection

Table 11.2 shows a comparison of four types of fuel: the traditional fuel JP-4, which is based on petroleum distillate forms the starting point; the standard fuel JP-10, which is based on tetrahydromethylcyclopentadiene; the new fuels RF-1, based on adamantane; and RF-4, based on tri[-ad]amantane. The RF-4 has a volumetric heat of combustion that is 35% higher than the JP-4, with a potential to increase the range of the Tomahawk from 690 to 930 miles. It also has a very low melting point of -40 °C and a high flash point of 126 °C, but it also has a very high viscosity. It would be appropriate to blend it with other components that have lower viscosities. This material is obtained as condensate liquid in a natural gas field, which needs to be separated and purified, as well as alkylated to lower its melting point. In view of its very superior volumetric heat of combustion, it is suggested that it is worth a premium price of \$50 per gallon.

### 11.3 Adsorbents for Separation: Solid Material

When the product is a solid, such as titanium dioxide used as an additive to paint and to paper to improve opacity and covering, the descriptive parameters also include the crystalline and amorphous nature of the solid, the average diameter of the crystals,

and the relation between the amorphous and the crystalline phases. When the solid has a high degree of porosity, such as a silica gel or a zeolite, the descriptive parameters must also include the porosity, the surface area per volume, the average pore diameter, and the distribution of pore diameters, etcetera.

Let us consider a zeolite adsorbent for selective separation in an important oil refinery stream that contains aromatic compounds.

### Identify Market and Technology

Petroleum refineries produce a stream of valuable aromatic compounds called the BTX, or benzene–toluene–xylenes (Ruthven 1984). The C<sub>8</sub> compounds can be easily separated from the C<sub>6</sub> and C<sub>7</sub> compounds by distillation, and consist of ethyl benzene, *o*-xylene, *m*-xylene, and *p*-xylene. Ethyl benzene is the starting material for styrene, which is used to make polystyrene; *p*-xylene is oxidized to make terephthalic acid, and then condensed with ethylene glycol to make polyester for fibers and films. The buyers of *p*-xylene are the manufacturers of terephthalic acid, such as BP-Amoco, who in turn sell to the fiber manufacturers such as DuPont and Dow. These are big and sophisticated companies that have strong research and engineering capabilities, and are used to have multiple suppliers. The eventual consumers of adsorbents are the public who consider polyester as one of the choices in fabric and garments, in competition with other synthetic and natural fibers. Their purchases are also dependent on personal income and prosperity. In times of recession, it is always possible for a consumer to downgrade to cheaper fibers and to wear old clothes for a longer period of time before new purchases.

The process to separate the four C<sub>8</sub> aromatics by distillation is very difficult because of their closeness in boiling points (table 11.3), so that only *o*-xylene can be separated by distillation. *p*-Xylene has a unique high melting point, and can be separated by cryogenic crystallization, but this is an expensive process that requires refrigeration. What is desired is an economic separation process that singles out *p*-xylene among these four compounds.

The principal types of industrial adsorbent can be divided into amorphous and the crystalline types. The former includes activated carbon, silica gel, and activated alumina; the latter includes zeolites and their aluminum phosphate, AlPO<sub>4</sub> (or ALPO), analogs. Yang (2003) wrote that, since the invention of synthetic zeolites in 1959, adsorption has become a key separation tool in the chemical, petrochemical, and pharmaceutical industries. Adsorptive separation of different molecules can be achieved by three mechanisms: equilibrium adsorption differences, diffusion kinetics differences,

**Table 11.3 Boiling and melting points of C<sub>8</sub> aromatics**

	T <sub>b</sub> , °C	T <sub>m</sub> , °C
Ethyl benzene	136.2	-95.0
<i>p</i> -Xylene	138.4	13.0
<i>m</i> -Xylene	139.1	-47.9
<i>o</i> -Xylene	144.4	-25.2

and steric effects from the molecular sieve properties of zeolites. An example of the first method is the separation of oxygen from nitrogen in air by adsorption with Na-X and Ca-A zeolites, where the Henry's law constant of nitrogen is four times higher than that of oxygen. An example of the second principle is that oxygen diffuses 30 times faster than nitrogen in carbon molecular sieves. The third principle of steric separation is unique to zeolites, as they possess pores that are comparable in diameters to that of many organic molecules. The slight difference in molecular diameters between *p*-xylene and *o*-xylene is sufficient to admit the former and exclude the latter in the pores of ZSM-5.

The zeolites are nanoporous crystalline aluminosilicates, with a formula such as  $\text{Na}_3(\text{AlO}_2)_3(\text{SiO}_2)_{93}$ . The crystal framework consists of an assemblage of  $\text{SiO}_4$  and  $\text{AlO}_4$  tetrahedra, to form an open crystal lattices containing pores or channels of molecular dimension into which guest molecules can penetrate. Since the microporous structure is determined by the crystal lattice, it is precisely uniform with no distribution of pore sizes, and it is thus different from silica gel and activated charcoal. The ring sizes of commercially important zeolites are defined by the number of O atoms, and their pore openings can usually be described by a major and a minor diameter. The large-pore X and Y faujasites and mordenites have pore openings of 12-member rings and pore sizes of around 0.74 nm. The medium-pore ZSM-5 has 10-member rings and channel openings of 0.56 nm. The small-pore A zeolite has pore openings of eight-member channels and pore sizes around 0.41 nm. There is a large amount of literature on the crystal structure of these materials (Meier and Olson 1987). Another class of these nanoporous materials is the ALPOs, with molecular size rings.

The UOP Molex process began in 1962 (*Kirk-Othmer Encyclopedia of Chemical Technology* 2002), and used the eight-ring 5A molecular sieve to separate the narrower normal paraffins in the range of  $\text{C}_{10}$  to  $\text{C}_{15}$  from the branched and cyclic paraffins. This is critical in obtaining the normal paraffins needed to make alkylbenzene sulfonates for detergents, as the bacteria in the environment are able to digest the normal paraffins completely, but not the branched paraffins, so that suds build up from discharges from home washers. The past decades have shown an explosion in the development of new nanoporous materials: mesoporous molecular sieves, zeolites, pillared clays, sol-gel-derived metal oxides, and new carbon materials.

We seek a shape-selective adsorbent, or a molecular sieve, that would discriminate among these four  $\text{C}_8$  aromatic compounds, and would preferentially adsorb *p*-xylene. A measure of the goodness of the adsorbent is the selectivity, defined to be

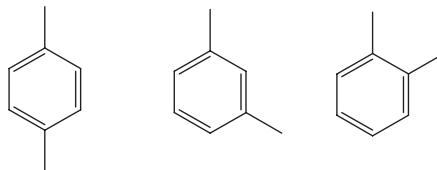
$$\alpha = \frac{y_A}{y_B} \frac{x_B}{x_A}$$

where A and B are the two components to be separated and  $x$  and  $y$  are the equilibrium concentrations in the liquid and adsorbed phases. Polymer-grade *p*-xylene needs to be more than 99% pure. Actually, there are many other requirements for the adsorbent. The ideal adsorbent should operate close to room temperature and pressure, which would make the process more economical. The adsorption should be easily reversible, so that *p*-xylene can be desorbed easily and be recovered without any change. This is done in one of three forms: temperature swing, which employs low temperature for adsorption and high temperature for desorption; pressure swing, which employs high pressure for

adsorption and low pressure for desorption; and displacement, which employs another gas that is intermediate in adsorption strength. The process of repeated adsorption and desorption usually leads to deactivation, often caused by the deposition of debris and coke, and needs to be regenerated. The candidates of our quest are the 10-ring zeolites, with large pores that are appropriate for separating the aromatic compounds, and the 12-ring zeolites with even larger pores.

### Generate Alternate Product Concepts

The three xylenes have different configurations for the attachment of the two methyl groups to the benzene ring: *p*-xylene has the two groups in a straight line, which gives it a smaller diameter of 0.63 nm, in comparison with the staggered methyl groups in *m*- and *o*-xylene, which lead to larger diameters of 0.68 nm. If we can find a porous solid that has a pore diameter smaller enough for the rapid diffusion of *p*-xylene, but either slower or prohibited transport for the other two xylenes, then we may be able to accomplish our separation goals.



There are many 10-ring zeolites: the channel cross-section can be nearly circular, such as with ZSM-5, where the major and minor diameters are 0.56 and 0.54 nm; they can be nearly elliptical, such as Heulandite, where the two diameters are 0.76 and 0.30 nm; and they can be approximately rectangular, such as the Laumontite, where the diameters are 0.53 and 0.40 nm. There are even more 12-ring zeolites to choose from. The faujasite X and Y have cross-sections that are nearly circular, with diameters of 0.74 nm. The ratio of Si to Al is a critical parameter, as the trivalent Al requires a cation such as Na<sup>+</sup>, but the tetravalent Si does not. There is also a choice of the cation involved, from the monovalent H<sup>+</sup>, Na<sup>+</sup>, and K<sup>+</sup>, to the divalent Mg<sup>2+</sup>, Ca<sup>2+</sup>, Sr<sup>2+</sup>, and Ba<sup>2+</sup>.

### Investigations and Solutions

When the molecular diameter approaches the channel opening of a ZSM-5 zeolite, the diffusion coefficient can drop by nine orders of magnitude between normal hexane and 1,3,5-trimethylbenzene, and can drop by three orders of magnitude between *p*-xylene and *o*-xylene (table 11.4).

It would seem that the xylene diameters, from 0.63 to 0.68 nm, are too large to enter the 10-ring ZSM-5 pores, with diameter 0.56 nm. On the other hand, the 12-ring faujasites X and Y have diameters of 0.74 nm, which is too large to discriminate among the xylenes. Actually, a molecule that is somewhat larger than the channel can nevertheless diffuse through the channel slowly, since the molecules are not hard

Table 11.4 Molecule and ZSM-5 data

Molecule	Critical diameter, nm	Diffusion coefficient in ZSM-5, $\text{cm}^2/\text{s}$
Normal hexane	0.44	$2 \times 10^{-4}$
3-Methylpentane	0.53	$5 \times 10^{-5}$
2,2-Dimethylbutane	0.61	$9 \times 10^{-8}$
<i>p</i> -Xylene	0.63	$1 \times 10^{-7}$
<i>o</i> -Xylene	0.69	$5 \times 10^{-11}$
1,3,5-Trimethylbenzene	0.77	$8 \times 10^{-13}$

spheres and the van der Waals repulsion can be overcome with an increase in energy; the motions of vibration and bending can also temporarily reduce the diameter of the molecule. It is also possible to narrow the channels by a number of techniques, such as by replacing the  $\text{H}^+$  and  $\text{Na}^+$  ions with more bulky cations, such as  $\text{Sr}^{2+}$  and  $\text{Ba}^{2+}$ , as well as adding adsorbed molecules, such as phosphorus, magnesium, boron, silicon, antimony oxide, and controlled coking. These measures will reduce the diffusivities of all the xylenes, but the retardation is even greater for the wider *o*- and *m*-xylenes. Very aggressive channel narrowing would lead to higher selectivity, but also lower diffusivity, which would make the separation process slower.

Pressure swing adsorption requires an adsorption that is very pressure sensitive, and temperature swing adsorption requires an adsorption that is very temperature sensitive. Displacement adsorption requires another gas with an equilibrium adsorption that is between the strengths of the aromatic components to be separated, as well as being inert and easily separated from the components in another column. After a number of cycles of adsorption-desorption, the zeolite ages due to the accumulation of debris that blocks the pores and reduces efficiency; this then has to be regenerated by purging or heating, or even burning. An ideal adsorbent ages slowly and can be completely regenerated with ease.

The zeolite crystals are in the form of fine powders, which would cause a very high pressure drop in a packed bed. They have to be formed into granules of approximately 3 mm in diameter, by using clay binders, such as kaolinite and montmorillonite. The methods consist of pelletization with binders under pressure into short cylinders, wet extrusion with a fluid into continuous cylinders, and granulation by rolling with binders into spheres. They also need to be dehydrated and calcined to remove volatile components before use.

### Evaluation and Selection

The commercial process Parex, developed by the UOP company, uses faujacite X in the forms of Sr and BaX, with a composition of  $\text{Sr}_{21}\text{Ba}_{22}(\text{AlO}_2)_{86}(\text{SiO}_2)_{106}$  for a ratio of  $\text{Si}/\text{Al} = 1.23$ , as well as faujacite Y in the form of Na and K, with a composition of  $\text{Na}_{56}(\text{AlO}_2)_{56}(\text{SiO}_2)_{136}$  for a ratio of  $\text{Si}/\text{Al} = 2.43$ . The 10-ring ZSM-5 is also used in commercial separations; it has a composition of  $\text{Na}_3(\text{AlO}_2)_3(\text{SiO}_2)_{93}$  for a ratio of  $\text{Si}/\text{Al} = 31$ . They produce a *p*-xylene purity of 98–99%.

## 11.4 Controlled Drug Release: Composite Solid

When the product is a composite solid material, with different components carrying out separate functions, the product designer has the most design parameters to manage. Familiar examples are the fiber-reinforced polymers used to make tennis rackets and golf clubs, and the coated fabric GoreTex that sheds water droplets and lets air breathe through. Then there are the numerous applications for biodegradable polymers in the biomedical field.

### Identify Market and Technology

Consider a diabetes patient who needs a steady level of insulin in the bloodstream, enough to lower the concentration of blood sugar, especially just after a meal, but not too low so as to lead to shock and coma. The patient can look forward to many decades of three painful and inconvenient injections per day, which requires a very disciplined and highly motivated attitude of the patient. It would be desirable to have drug capsules that release the drug slowly over a period of days or months, so that the patient does not have to endure constant pain, and does not need good memory and discipline for unpleasant routines. The rate of release must be uniform in time, and it must not depend on physical state (i.e., whether exercising or resting), the mental state (i.e., whether stressed or relaxed), or seasonal state (i.e., whether summer or winter). And there must be no danger of accidental release of large amounts of drug in a short time, which would have disastrous consequences.

Consider a cancer patient who needs chemotherapy after the removal of a tumor to suppress stray cancer cells that are not removed in surgery. Since cancer cells are body cells that have gone into uncontrolled growth, most cancer drugs operate by attacking all fast-growing cells. The drug is needed at a specified concentration only in the local neighborhood of the tumor, but not circulated in blood to attack hair follicles and cause loss of hair, to attack bone marrow and cause anemia, or to attack stomach and intestine linings and cause pain. Shortly after the drug is administered, care must be taken so that the drug concentration does not become so high as to cause excessive collateral damage to the bone marrow, to the stomach lining, or to cause the hair to fall out. After a while, the drug is metabolized or excreted from the body, so that the drug concentration may rapidly drop so low as to be ineffective. In an ideal situation, the desirable outcome is a steady and local concentration of the drug, below the maximum of damaging concentration and above the minimum of effective concentration, over a long enough period of time to secure the cure. This outcome can be secured if the patient is to receive numerous small injections that are closely spaced in time, which is inconvenient and painful. It would be desirable to have capsules that slowly release the drug locally.

There are many other biomedical applications for a biodegradable polymer. When a wound created by trauma or surgery is stitched together, and healing takes place over a period of days or weeks, a nondegradable suture has to be surgically removed at the end, which means another trip to see the surgeon at additional discomfort and cost. However, if an absorbable suture is used, which degrades into small fragments that are absorbed by the body and metabolized, then the advantages to the patient can

be considerable. Orthopedic surgery to repair or prevent deformities requires the use of temporary material, such as tendon and cartilage taken from the patient, or synthetic biodegradable polymers. In growing new tissues and organs in tissue engineering, a scaffold of polymer is erected and seeded with tissue cells, and the polymer degrades with time while the cells grow to fill the void. Other examples include: agricultural micro-encapsulation for slow release of fertilizers and insecticides; support against erosion of freshly cultivated hillsides; and discardable wrappers for packaging (Fan and Singh 1989).

There are a number of major ways to deliver a drug to a local target in the body where it is needed, such as to a tumor in the brain. The systemic methods include intravenous injection to reach the bloodstream, swallowing by mouth to reach the stomach and intestines, and inhalation of aerosols to reach the lungs. The more local methods include: intramuscular injection, subcutaneous injection, and skin patch. A hypodermic syringe driven by an electric motor has been proposed and tested in the past, but this incurs the risk of an accidental massive discharge of the drug into the body, which would have disastrous consequences. A skin patch, such as the Nicoderm for the delivery of nicotine substitute, is acceptable when the dosage does not have to be precise, as diffusion through the skin is not quantitative and reproducible among different patients. A polymer depot implanted under the skin can be effective if the rate of drug delivery is controlled by diffusion through a barrier, or controlled by the rate of swelling of the polymer, which expands and releases the drug. The most effective method today is to embed the drug in a polymer matrix that spontaneously biodegrades in the body by enzyme reactions, so releasing the drug slowly (Langer 1980, 1994, Langer and Peppas 1983).

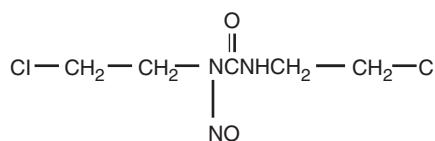
Let us consider polymers that degrade slowly when placed in the human body. Most synthetic polymers are designed to last a long time in the environment, and it is an advantage when users desire stability on the shelf and a long useful life, but it is a problem when they are discarded after use and accumulate as garbage in the environment. On the other hand, natural polymers always degrade in the environment, which involves bond scission in the backbone of the polymer into fragments with lower molecular weights and from a number of reactive attacks, including oxidation, ultraviolet radiation, hydrolysis, and biological agents. In the human body, only biological attack by enzymes is relevant.

### Generate Alternate Concepts

Langer (1994) discussed three types of controlled release system. The first type is a porous pellet manufactured by dissolving a biocompatible polymer in an appropriate solvent, and the protein is added in powder form. The resulting mixture can be cast in a mold and dried. When the pellet is placed in water, the molecules trapped within the polymer matrix are released gradually. In another formulation, a mixture of the drug and polymer powder below the glass transition temperature is made, and then compressed and sintered above the glass transition temperature. The second type involves biodegradable polymers, which eliminates the need to surgically remove the drug-depleted device. It works by the hydrolytic degradation of the surface of the polymer, and releasing the drug. The third type involves a pulsatile release system, such as a polymeric matrix that contains drug powder and magnetic beads. The release of drug

is accelerated by the application of an oscillating external magnetic field, such as at a frequency of 10 Hz.

The biodegradable implant has the drug embedded in a polymer matrix, with a particular geometry and size. The drug may be a small molecule ( $MW < 600$ ) with a high diffusivity coefficient in water, or it may be a protein or polysaccharide with Dalton number up to  $2 \times 10^6$ , with negligible diffusion coefficient. Let us consider the drug to be carmustine, at the rate of 3 mg/day over 20 days. The drug carmustine has the formula  $C_8H_9Cl_2N_3O_2$ , a molecular weight of 214, and the structural formula is



It has antitumor activities, but is also anticipated to be a carcinogen. It is used in conjunction with brain cancer surgery of malignant glioma to suppress recurrence. Following an intravenous infusion of carmustine at doses ranging from 30 to  $170 \text{ mg/m}^2$  (the surface area of a person in meters squared is approximately given by  $\sqrt{\text{Height (in)} \times \text{Weight (lb)} / 3131}$ ), the average terminal half-life of the drug in the body is only 22 min.

Purely hydrocarbon-based nonpolar polymers have very low rates of degradation by enzyme attack at body temperature, and polar groups of O and N are needed. A number of materials have been approved for biodegradable biomedical devices. Glycolic acid ( $\text{HO}-\text{CH}_2-\text{COOH}$ ) and L-lactic acid ( $\text{HO}-\text{CH}(\text{CH}_3)-\text{COOH}$ ) copolymers are well-established raw materials in the manufacture of synthetic biodegradable surgical sutures, clips, staples, and meshes. They have uniform and reproducible mechanical properties and cause minimal tissue reaction, and they do not have to be removed. The glycolides erode at a rate that is much faster than the lactides (Kim 2002), so that a set of copolymers with different proportions can be fashioned that would produce a large range of degradation and release rates, from a few days to several months. Langer (1994) suggested that polyanhydrides of sebamic acid (SA),  $\text{HOOC}-(\text{CH}_2)_8-\text{COOH}$ , can be copolymerized with polyphenoxy carboxypropane (PPCP), so that different proportions of the more biodegradable SA to the more resistant PPCP in copolymerization give rise to a range of release rates from 1 day to 3 years.

The device can be a flat disk that is implanted under the skin, or many microspheres that can be injected into the body cavities. The device must not cause adverse reaction from the body, such as bio-incompatibility and inflammation, and the polymer should disintegrate into harmless products that are easily absorbed into the body's metabolism.

### Investigations and Solutions

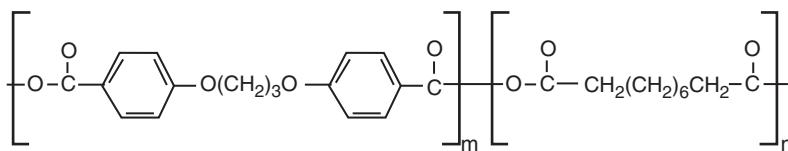
The requirement of 3 mg/day of carmustine over 20 days means a load of up to 60 mg in a device. Most formulations would suggest no higher than a 10% load of drug in the polymer matrix, so that the entire drug depot would be approximately 0.6 mL in volume. This is a very modest size disk, of about  $0.1 \times 2.5 \times 2.5 \text{ cm}^3$ . Or this amount of drug can be packaged into numerous injectable microspheres of  $300 \mu\text{m}$

in diameter or 0.014 mm<sup>3</sup> in volume, so we need 43,000 microspheres. Some of the information necessary to design the composition of this copolymer can be found in Langer (1994). For 79% SA with PCPP, it takes a little less than 2 weeks to degrade 100% of the polymer. Since we are interested in a 14-week release time, the appropriate concentration of SA is around 49%.

Making prototypes and laboratory testing of polymer disks are projects that require limited cost and time duration. Certification by the FDA, on the other hand, is a long drawn out and costly process, where animal tests are followed by three phases of clinical trials, which have been described elsewhere (e.g. Suffness 1995). When the results are assessed and evaluated, a brand new product that costs more than \$100 million can also run into many unforeseen problems, which makes many financiers very cautious. One of the biggest unknowns is who would pay for this costly new form of medication, and whether the medical insurance companies and Medicare would approve payment. This is the reason why so many information technology products, such as digital cameras and spreadsheets, are launched quickly, as they require much less capital to start and do not require FDA clearance.

### Evaluation and Selection

One device design is offered by the Guilford Company of Baltimore, namely the product GLIADEL Wafer, which was approved by the FDA in February 2003 (Guilford Pharmaceuticals 2003). It is indicated or suggested in newly diagnosed high-grade malignant glioma patients as an adjunct to surgery and radiation, for recurrent glioblastoma multiforme. It is capable of delivering chemotherapy directly to the site of a brain cancer, bypassing the blood-brain barrier and minimizing drug exposure to other areas of the body. It is a small, white to off-white dime-sized wafer approximately 1.45 cm in diameter and 1 mm thick. Each wafer contains 192.3 mg of a biodegradable polyanhydride copolymer and 7.7 mg of carmustine. Up to eight wafers can be implanted in the cavity created when a surgeon removes a brain tumor. They slowly dissolve there, releasing BCNU directly to the tumor site in high concentrations, while minimizing drug exposure to other areas of the body. The copolymer, polifeprosan 20, consists of poly[bis(*p*-carboxyphenoxy)propane : sebacic acid] in a 20:80 molar ratio and is used to control the local delivery of carmustine. Carmustine is homogeneously distributed in the copolymer matrix. The structural formula for polifeprosan 20 is



Ratio m:n = 20:80; random copolymer

The current price of a pouch with eight wafers is \$9600.

Now let us do a bit of reverse engineering, and take this design apart and see why it makes sense, which would be helpful in designing other devices. No reason is given for the formulation of eight wafers, each with the thickness of 1 mm and 1.45 cm

diameter, with a volume of 0.21 mL, a weight of 0.2 g, and an external surface area of 330 mm<sup>2</sup>. The drug loading in the polymer is 7.7 mg out of 200 mg, or 3.9%. Collectively, these eight wafers have a total volume of 1.68 mL, with a weight of 1.6 g and a surface area of 2642 mm<sup>2</sup>. The rate of surface-initiated rate of bio-erosion is approximately 1 mm on each surface over 20 days, or 25  $\mu\text{m}/\text{day}$ .

One advantage of multiple wafers is that each cavity created in brain surgery has a different shape and size, and there should be some flexibility to distribute the appropriate number of wafers over disturbed areas where surviving tumor cells may be lurking. A single large wafer that contains all the drug would have the same thickness of 1 mm and diameter of 4.1 cm, which may be too large for many cavities. If the formulation is distributed into 16 smaller wafers, then the diameter is 1 cm.

No. of wafers	1	2	4	8	16
Diameter, mm	41.0	29.0	20.5	14.5	10.3

Would 40,000 microspheres with a diameter of 0.3 mm be a better idea? One advantage of the thin disk geometry is that, as bio-erosion proceeds, the external surface area remains unchanged, so that the rate of drug release is constant with time. When a sphere is bio-eroded, the surface area decreases as the square of the diameter  $D^2$ ; so, in principle, there would be a declining rate of release with time. On the other hand, microspheres have the advantage that it is possible to distribute them even more uniformly, into even very small crevices where 1 cm disks may have difficulty fitting in. There are many other situations where microspheres would be preferred, especially by injection into cavities that do not need to be surgically opened.

## 11.5 Evaluation, Selection, and Business Plan

After we have done enough development work and the testing results are in, we need to compare our best product designs against the best products on the market. We need to assess the strong and weak points of our best designs, and to estimate whether they would win a significant share of the market in competition. This requires a financial analysis to see whether our product will earn a good enough profit to justify the investment, and what the major risks involved are. The final decision is usually made after we present a business plan to the decision makers, and make recommendations to proceed with commercialization or to withdraw.

### Multifactor Optimization

We begin by listing the attributes of our product designs and make comparisons. The factors of a product design may be divided into:

- Market factors—ability to solve important customer problems, quality and price advantages over competing products, sufficiently large and growing market sales volume.

- Manufacturing factors—secure raw material supply, reliable manufacturing process, availability of engineers and labor, low-cost and trouble-free manufacturing.
- Safety and environmental factors—absence of, or controllable, negative impacts, no current and pending negative regulations, reliable plans to control safety and environmental problems.
- Organizational factors—synergism with other products of the firm, taking advantage of internal strengths of the firm, diversify the risks of the total firm investment portfolio.
- Economic factors—low requirement for investment capital, high net present value, high internal rate of return on investment, low risks from uncertainties.

The first four factors are qualitative, and one then ranks all the product designs against each of these factors. When a new product design is found to be superior to all alternative designs, as well as current products on all factors, it would of course be declared the winner. Usually, there is not a single dominant product, and each design has some good points and some bad points. For instance, in biodegradable cancer implants, the design that appears to be the most effective in preventing the recurrence of cancer would be the disks, as they have a long time horizon and flat delivery rate, which is better than the microspheres with a falling rates of delivery, and is more convenient and less painful than multiple injections. The design that is the most convenient to use to reach all shapes and sizes of crevice would be the microspheres. There is little doubt that intravenous injections would be the lowest in cost, even when we do not have all the information. When there is a trade-off among factors, how do we pick a winner? We may assign degrees of satisfaction from 1 to 10 to each factor, and compute their average degree of satisfaction. If some factors are considered to be more important than others, such as effectiveness in saving lives may be more important than convenience and cost, or avoiding lawsuits to be more important than higher profit, then the method of weighted average is used. This involves assigning weights on each factor and computing the overall average score of each design.

These factors can also be organized as a multiscale optimization problem by their operations, which is seen in table 11.5. The chemistry of manufacturing is mostly concerned with the nanometer to micrometer scales of molecules and colloids, and process engineering is mostly concerned with the millimeter to meter scales of components and equipment. The systems of plants and markets operate in the kilometer scale, and the global environment operates in the thousand kilometer scale.

**Table 11.5 Multi-scale integration in design**

Scale	Components, objects, and systems
Molecular, nm	Atoms, molecules, bonds
Colloidal, $\mu\text{m}$	Phases, size distributions, surfactants
Components, mm	Pellets, granules, pores
Equipment, m	Trays, beds, heat exchangers, pumps, reactors, separators
Systemic, km	Plant, refinery, firm, market, finance
Global, $10^3$ km	Atmosphere, ocean, land, biosphere

## Finance, Return, and Risk

The most influential and determining factors in decision making are the quantitative financial analyses, which are used to declare the winner as the best use of the firm's assets. The criteria used most often are the net present value, NPV, and the internal rate of return, IRR. A financial plan begins with a time horizon, such as 5 years, and the forecast of a number of parameters of expenditures and incomes for each of the years, on:

- Investments—land, buildings, plant, equipment, working capital.
- Market—units sold, price per unit, gross revenue.
- Manufacturing—raw material, utility, labor, maintenance.
- General—sales, administration, general expenses.

Working capital is the difference between the current asset (inventory of raw materials and finished products, receivables or goods sold but payment not yet collected) and the current liability (payables or goods received but not yet paid for).

These parameters are used to construct a cash flow analysis, which contains items including:

- Sales, namely the units sold multiplied by price per unit.
- Cost of goods sold (COGS), which includes raw materials, utilities, and labor.
- Selling, administration, and general expenses (SAGE), which includes office expenses, marketing, and advertising expenses.
- Tax, namely federal business income tax and state and local taxes.
- Investment, namely the money spent on land, buildings, and equipment at the beginning, and the liquidation or disposal value at the end.

We may consider all the operations of a start-up company, or the incremental operations of a project attached to an existing company. Let us consider the example in table 11.6.

Is this a good investment? Let us introduce the basic concept that money on hand is worth more than future money, or the time value of money. The net present value NPV is defined to be

$$NPV = P_0 + \frac{P_1}{1+r} + \frac{P_2}{(1+r)^2} + \frac{P_3}{(1+r)^3} + \cdots + \frac{P_n}{(1+r)^n}$$

**Table 11.6 Example cash-flow analysis**

Year	0	1	2	3	4	5	Sum
Sales	0	100	150	250	200	150	850
COGS	0	-30	-45	-90	-70	-45	-280
SAGE	0	-20	-25	-30	-30	-30	-135
Tax	0	-8	-12	-15	-10	-8	-53
Investment	-350	0	0	0	0	200	-150
Net cash flow	-350	42	68	115	90	267	232

$P_i$  is the cash flow for the year  $i$ , and  $r$  is the “relevant discount rate,” which can be thought of either as the interest of borrowing money to start the company, or as the return from an alternate opportunity for a safe investment. This equation considers that \$1 next year is worth only  $1/(1 + r)$  dollars this year. The value of the NPV is the sum of all the cash flows for the  $n$  year, discounted by the power of compound interest. We give here the values of NPV for a number of values of  $r$ :

Interest rate $r$	0	0.05	0.10	0.15	0.20	0.25
NPV	232	128	53	-2	-42	-72

When  $r = 0$ , we take the position that future money is worth just as much as current money, so that NPV is equal to the sum of the cash flows. When  $r = 0.05$ , or a discount rate of 5%, the NPV equals 128; a positive NPV means that this investment is better than putting our investment of 350 into a safe alternate investment to earn 5% interest. When  $r = 0.25$ ,  $NPV = -72$ ; a negative NPV means we are better off investing our 350 into a safe alternate investment to earn 25% interest. The NPV for each investment is computed and compared, and the investment with the highest NPV has the most merit. Note that NPV is sensitive to the value of  $r$ , so that the winner when interest rates are low need not be the winner when interest rates are high.

The interest rate when  $NPV = 0$ , which occurs at approximately  $r = 0.15$ , is called the internal rate of return IRR. This investment has the same merit as putting our investment into an alternate venture that pays 15% interest. When an investment A has a higher IRR than investment B, it would normally rank higher to the people holding the financial purse strings. Most companies have a fixed investment budget each year, on new ventures that it can finance; they approve the proposed project that has the highest IRR, and then go down the list to projects with lower IRRs until the budget is exhausted, or when the next project falls below some cut-off value that the company considers worthwhile, such as 12%.

The forecasts of sales and expenses above are only educated guesses, and life is full of unexpected uncertainties. Prudence requires that we should take into consideration a number of risk issues, such as a sudden rise in the cost of crude oil, a labor strike at the loading dock, a change in prime interest rate by the central bank, the introduction of a me-too product by the competition, or the product is found to have environmental problems. According to each scenario, we modify the forecast of some of the key parameters; then we recompute the NPV and the IRR to see whether there is a strong negative impact, and whether the firm can survive the worst-case scenario. We also calculate the standard deviation  $\sigma$  of the dependence of the IRR on this set of scenarios. The example given in table 11.7 concerns the introduction of a new pesticide to an export market in an underdeveloped country, which has a reasonable return of 15% in a “normal” scenario. A number of alternate scenarios are then proposed, together with the probabilities of their occurrence, and we examine their impacts, from the favorable currency devaluation with a return of 22% to the worse-case scenario of environmental problems, which leads to a return of only 7%. With this set of scenarios, the average return is reduced to 14.8%, and the standard deviation is 2.2%.

Another consideration is for the firm to seek an increase in investment diversity to reduce risks for the entire portfolio of investments, so as not to put all the eggs in

**Table 11.7 Rate of return  $r$  for various scenarios**

Scenario	$r, \%$	Probability weight, %
Normal	15	85
Severe drought	10	3
Dock strike	13	5
Currency devaluation	22	4
Environmental problem	7	3
Average	14.79	
sd	2.19	

one basket. It would be a plus when a new project has a low or even a negative covariance to the current set of investments. If regional conflicts and currency devaluations will cause some investments to go down, then we desire some new investments that would go up under the same circumstances. The  $\beta$  coefficient of a proposed investment in comparison with the ongoing projects under a package of scenarios is defined to be

$$\beta_i = \frac{\text{Cov}(R_i, R_{\text{og}})}{\text{Var}(R_{\text{og}})}$$

where  $R_i$  is the return of a proposed investment  $i$  and  $R_{\text{og}}$  is the return of the ongoing projects. When  $\beta_i = 1$ , this investment moves in the same manner as the ongoing projects, so it provides no diversification of risks; but when  $\beta_i = -1$ , this investment moves in the opposite direction to the market, which provides maximum diversification of risks.

One also pays attention and guards against the possibility of a negative cash flow for a particular year, which means we would not have enough cash on hand to pay the current debts. In that case we must increase the working capital for that year, which works as insurance against bankruptcy but has a negative effect on IRR.

## Business Plan

The business plan is a formal proposal for action, to be presented to the management and the finance department, to persuade them to authorize the project. The business plan should summarize the reasons behind this proposal, the actions proposed, together with the financial and other consequences. There are many possible formats for the business plan for a particular product, but the essential features are outlined here (Bangs 2002):

### *Front Matter*

Cover sheet: name of business, principal officers, address and telephone number

Executive summary: a few pages to summarize the material covered

Table of contents

### *Description of the Business*

- A. The business: proprietorship, partnership, or corporation
- B. Products and services offered: benefits to customers, product differentiation
- C. Market: target segment, growth and profit, pricing, plans to expand or change
- D. Location of business: location, demographic, and market
- E. Competition: direct and indirect competitors, assessment of their strengths, product differentiation
- F. Manufacturing and operations: plant, process, equipment, raw material, labor
- F. Risks and opportunities: the most important risks and opportunities
- G. Management: relevant background and experience, team, strengths, and weaknesses
- H. Personnel: key individuals, hiring and training

### *Financial Analysis*

- A. Sources and uses of funds: equity, bond, loan, credit; land, equipment, raw material inventory, product inventory, working capital
- B. Capital equipment list
- C. Balance sheet: current and fixed assets, current and long-term liabilities, net worth
- D. Break-even analysis: sales volume needed to cover fixed and variable costs
- E. Income projections (profit and loss statements) for multiple years: sales, fixed costs (mortgage, loan interest, depreciation, rent, taxes, insurance, advertising, salaries, benefits), variable costs (cost of material, hourly labor, supplies, maintenance)
- F. Cash flow projection, NPV, IRR
- G. Variance and sensitivity analysis

### *Back Matter*

Supporting documents: personal resumes, letters of reference, job descriptions, contracts, legal documents

## **References**

- Bangs, D. H. 2002. *Business Planning Guide*. Chicago, IL: Dearborn Financial Publishing.
- Chung, H. S., C. S. H. Chen, R. A. Kremer, and J. R. Boulton. 1999. Recent developments in high energy density liquid hydrocarbon fuels. *Energy & Fuels* 13(3): 641–649.
- Cussler, E. L. and G. D. Moggridge. 2001. *Chemical Product Design*. Cambridge: Cambridge University Press.
- Environmental Protection Agency (EPA). 2001. *Introduction to the Clean Water Act* <http://www.epa.gov/watertrain/cwa/right14.htm>.
- Fan, L. T. and S. K. Singh. 1989. *Controlled Release: A Quantitative Treatment*. Berlin: Springer-Verlag.
- Guilford Pharmaceuticals. 2003. <http://www.guilfordpharm.com> and <http://www.gliadel.com>.
- Kim, Y. H. 2002. An overview on biodegradable polymers in biomedical application. In *ICS-UNIDO Edp EGM*, Trieste, Italy.

- Kirk-Othmer *Encyclopedia of Chemical Technology*. 5th edition. 2002. New York: John Wiley & Sons.
- Langer, R. 1980. Polymeric delivery system for controlled drug release. *Chemical Engineering Communications* 6: 1–48.
- Langer, R. 1994. Polymer systems for controlled release of macromolecules, immobilized enzyme medical bioreactors, and tissue engineering. In *Advances in Chemical Engineering*, vol. 19. San Diego: Academic Press; 1–50.
- Langer, R. and N. Peppas. 1983. Chemical and physical structure of polymers as carriers for controlled release of bioactive agents: a review. *Journal of Macromolecular Science: Reviews in Macromolecular Chemistry and Physics* C23: 61–126.
- Meier, W. M. and D. H. Olson. 1987. *Atlas of Zeolite Structure Types*. London: Butterworths.
- Ruthven, D. M. 1984. *Principles of Adsorption and Adsorption Processes*. New York: John Wiley & Sons.
- Suffness, M. ed. 1995. *Taxol: Science and Applications*. Boca Raton, FL: CRC Press.
- Ulrich, T. U. and S. D. Eppinger. 2000. *Product Design and Development*. Boston, MA: Irwin McGraw-Hill.
- Yang, R. T. 2003. *Adsorbents: Fundamentals and Applications*. Hoboken, NJ: Wiley-Interscience.

## Further Reading

- Anderson, J. R. and K. C. Pratt. 1985. *Introduction to Characterization and Testing of Catalysts*. New York: Academic Press.
- Ariens, E. J. ed. 1971. *Drug Design*, vol. 1. New York: Academic Press.
- Biegler, L. T., I. E. Grossman, and A. W. Westerberg. 1997. *Systematic Methods of Chemical Process Design*. Upper Saddle River, NJ: Prentice Hall.
- Bogucki, P. 1999. *The Origin of Human Society*. Malden, MA: Blackwell.
- Bordaz, J. 1970. *Tools of the Old and New Stone Age*. Garden City, NY: The Natural History Press.
- Bronikowski, R. J. 1986. *Managing the Engineering Design Function*. New York: Van Nostrand Reinhold.
- Cotton, F. A. and G. Wilkinson. 1972. *Advanced Inorganic Chemistry: A Comprehensive Text*. New York: Interscience Publishers.
- Daellenbach, H. G., J. A. George, and D. C. McNickle. 1983. *Introduction to Operations Research Techniques*. Boston, MA: Allyn and Bacon.
- Douglas, J. M. 1988. *Conceptual Design of Chemical Processes*. New York: McGraw-Hill.
- Edwards, D. A. et al. 1997. Large porous particles for pulmonary drug delivery. *Science* 276: 1868–1871.
- French, M. 1994. *Invention and Evolution: Design in Nature and Engineering*. Cambridge: Cambridge University Press.
- French, M. J. 1985. *Engineering Design: Conceptual Stage*. Berlin: Springer-Verlag.
- Goodman, J. and V. Walsh. 2001. *The Story of Taxol: Nature and Politics in the Pursuit of an Anti-Cancer Drug*. Cambridge: Cambridge University Press.
- Henry, C. 2001. Structure-based drug design. *Chemical and Engineering News* (June 4): 69–78.
- Kent, J. A., ed. 1962. *Riegel's Industrial Chemistry*. New York: Reinhold Publishing.
- Klamann, D. 1984. *Lubricants and Related Products*, Weinheim: Verlag Chemie.
- Konstandt, F. 1985. *Organic Coatings: Properties and Evaluation*. New York: Chemical Publishing Co.

- Lance, L. L. et al., eds. 2002. *Drug Information Handbook: for the Allied Health Professional*, Hudson, OH: Lexi-Comp and the American Pharmaceutical Association.
- Larson, R G. 1999. *The Structure and Rheology of Complex Fluids*. New York: Oxford University Press.
- McAvoy, J., ed. 1975. *Catalysts for the Control of Automotive Pollutants*. Advances in Chemistry Series 143. Washington, DC: American Chemical Society.
- Merck & Co., Inc. 1996. *The Merck Index*. Whitehouse Station, NJ: Merck Research Laboratories.
- Morgans, W. M. 1982. *Outlines of Paint Technology*, vol. 1, *Materials*, vol. 2, *Finished Products*. New York: Wiley.
- Pahl, G and W. Beitz. 1984. *Engineering Design*. Berlin: Springer-Verlag.
- Peters, M. S. and K. D. Timmerhaus. 1968. *Plant Design and Economics for Chemical Engineers*. New York: McGraw-Hill.
- Petroski, H. 1996. *Invention by Design: How Engineers Get from Thought to Thing*. Boston, MA: Harvard University Press.
- Physicians Desk Reference*. 1993. Medical Economics Data.
- Ross, S. A., R. W. Westerfield, and J. Jaffe. 1996. *Corporate Finance*. Chicago, IL: Irwin.
- Rudnick, L. R. and R. L. Shubkin. 1999. *Synthetic Lubricants and High-Performance Functional Fluids*. New York: Marcel Dekker.
- Schnabel, W. 1981. *Polymer Degradation: Principles and Practical Applications*. Munich: Carl Hanser Verlag.
- Seider, W D., J. D. Seader, and D. R. Lewin. 1991. *Process Design Principles: Synthesis, Analysis and Evaluation*. New York: Wiley.
- Shreve, R. N. 1967. *Chemical Process Industries*, third edition. New York: McGraw-Hill.
- Silverman, R. B. 1992. *The Organic Chemistry of Drug Design and Action*. San Diego, CA: Academic Press.
- Stiles, A. B. 1983. *Catalyst Manufacture: Laboratory and Commercial Preparations*. New York: Marcel Dekker.
- Straubinger, R. M. 1995. Biopharmaceutics of Paclitaxel (taxol): formulation, activity, and pharmacokinetics. In *Taxol: Science and Applications*, M. Suffness, ed., Boca Raton, FL: CRC Press; 237–258.
- Straubinger, R. M. A. Sharma, U. S. Sharma, and S. V. Balasubramanian. 1995. Pharmacology and antitumor effects of novel Paclitaxel formulations. In *Taxane Anticancer Agents*, George G. I. et al., eds. ACS Symposium Series 583. Washington, DC: American Chemical Society; 111–123.
- Suh, N. P. 1990. *The Principles of Design*. Oxford: Oxford University Press.
- Trimm, D. L. 1980. *Design of Industrial Catalysts*. Amsterdam: Elsevier.
- Wei, J. and E. R. Becker, 1975. The optimal distribution of catalytic material on support layers in automotive catalysis. In *Catalysis for the Control of Automotive Pollutants* J. E. McAvoy, ed. Advances in Chemistry Series 143. Washington, DC: American Chemical Society; chapter 11.
- Wei, J. 1975. Catalysis for motor vehicle emissions. *Advances in Catalysis* 24: 57–129.
- Wei, J. 1987. Towards the design of hydrodemetallation catalysts. In *Catalyst Design: Progress and Perspectives*, L. L. Hegedus, ed., New York: Wiley.

## Design Project

The design project is a great learning experience that should be taken by the students as the capstone to the course, where they can synthesize the information and analytical

tools to produce a creative design. It is also an opportunity for the students to work in a team, work in a budget of time and resources, and develop organization and persuasive skills in oral and written communications. When there are enough students in the class, a good model is to divide the work in a team with four members: the CEO, the market VP, the manufacturing VP, and the finance VP.

### Project Selection

The initial concept of a new product project can be inspired by an unsatisfied market demand, or an underutilized technological capability. Each student who observes everyday life is familiar with a consumer product that serves a market with some success, but has defects that needs improvement. It may be inconvenient to use by the consumer, such as a Tupperware box to save food for the refrigerator, but is not transparent so one has to remove the lids of many boxes to find the one with the peas. Can we design a box that is transparent, while keeping all the desirable properties of the current Tupperware? It may be a disposable baby diaper that requires change every 4 h, and can we design one that would last 6 to 8 h? We are using salt to melt ice and snow on the roads in the winter, but consider the bad features: the ice melt is much colder to step into, the salty runoff makes potholes on the roads and kills vegetation, so what is a better idea? We have many new power-hungry electronic gadgets, such as the laptop computer, the cellular phone, the digital camera, the musical jukebox, and the global positioning system, all powered by a heavy battery that has a small capacity.

A good design should match a market opportunity with one or more potential technological solutions that can be investigated and explored. For instance, in the case of a clear Tupperware box, what plastic materials are currently in use, and what other plastic materials available are transparent. What are the reasons that some plastics are transparent and some are opaque, and what is the role of the molecular structure and what is the role of blending and additives? Can we modify the current material to make it transparent, and yet retain the other desired properties (what are they)? Can we find a different material that would do a better job?

An observant student has also heard of many new and old technologies that seemed powerful and promising, but have not found much application. The buckminsterfullerene and the related carbon tubules seem like marvelous materials with unmatched and fascinating properties, but they are not used in the marketplace. Neither are the electrically conducting polyacetylenes, which hold the promise of a moldable conductor that can be made at low temperature. Almost every professor of chemical engineering and chemistry has numerous research results that are not used in the marketplace today. A visit to their offices and discussions with their research staff may result in several suggestions that are worth further investigations.

A good design project should match the technological capability with several potential market applications, which can be investigated and explored. Which market would appreciate a very strong and light fiber, and would it be in golf clubs and tennis rackets, in flak jackets for soldiers and police?

Innovation involves coming up with an idea that is not obvious, or to resurrect an idea that has been considered and discarded before. To find an innovative idea,

the student or the team should make many observations of everyday life and the literature, consult with many people, and explore many ideas.

### Project Research and Analysis

The business plan can serve as a suggestion on the material that should be gathered, analyzed, and organized. A design project may involve mostly library search and computation, but a laboratory component would be splendid. The following aspects should be covered:

#### *Marketing Analysis*

- Who are the customers, and why are these products needed? How do the customers use these products to enhance their lives? What properties do these products require?
- What products are currently serving this market, what are their properties and prices, what are their strong and weak points?
- What properties of current products are most unsatisfactory, so that improvements are sorely needed to make these products more successful?
- Are these products under regulatory threat for safety or environmental problems?
- What is a good opportunity for a new or improved product?

#### *Technology Analysis*

- What is the technology that is either new or underutilized, what are new things that we can make, and what properties might the products have?
- What are the potential applications for these new technologies, and what markets would appreciate these properties?
- How are the current products made, using what raw materials and intermediates, and employing what processes in their manufacturing?
- Are there problems with a shortage of the raw materials, or with a hazardous process?
- What technologies can we use to solve the market needs? How can we modify current products to improve their most relevant properties? Where can we find better material?

#### *Economics and Evaluations*

- Compare your several design concepts with the best existing products in the market.
- Marketing: why would the customers prefer your proposed products over current products, because of superior properties or lower prices?
- Manufacturing: would your proposed products be easier to make, less costly, or use more readily available material?
- Safety and environment: would your product have less negative impact on safety and the environment?

- Finance: make projections of yearly cash flow on costs and revenues for your design, compute its net present value and internal rate of return. How do the proposed products rank in project return and security, and in corporate risk diversification?

### *Recommendation*

- Do you recommend that we proceed to commercialization immediately? Or do you recommend a period of product development to solve important problems?
- What are the remaining problems on marketing, manufacturing, and the environment that remain to be solved? How much time and money should we budget in order to solve these problems?

### Project Reporting

There should be an oral presentation towards the end of the semester to the management (which is your classmates, the professor, and the invitees) on your business plan for a proposed new product: on the innovative product idea, on market and technology analysis, on the financial and environmental analysis, and on your recommended action. Be prepared to answer searching questions from your management, and to hear constructive comments intended to improve your research, your analysis, your recommendation, and your presentation skills. If time permits, it would be useful to take all the comments and constructive suggestions into consideration, and make a second oral presentation. The product design project concludes with the submission of a written report, which would be in parallel with the traditional process design project.

# 12

## Product Innovation Opportunities

### 12.1 Market-Pull Opportunities

- World market mega-trends
- Batteries for the information, biomedical, and vehicle technologies
- Artificial organs and tissue engineering
- Biodegradable polymers
- Blood substitutes

Electrically conducting polymers—Heeger, MacDiarmid, and Shirakawa  
Fullerene—Curl, Kroto, and Smalley  
Combinatorial chemistry

Genomics

#### References

#### Further Reading

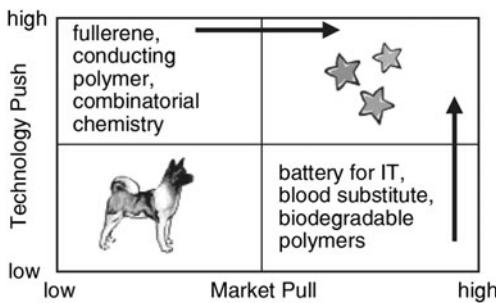
### 12.2 Technology-Push Opportunities

- Major scientific and technological advances

There is an endless list of exciting new products that are waiting to be invented, developed, and introduced to the marketplace, which would bring honor and profit to the innovators. We will discuss in this chapter some of the methods to identify the most exciting challenges and some of the most tantalizing targets.

What is a good target? An investigator should begin by making a self-examination on what are the internal competitive strengths of this individual or organization as the starting point. The internal competitive strength may be the ability to make keen observations on a market segment and its unsatisfied needs, plus the ability to connect with potential technological solutions; it may be the possession of a unique technological capability, or the observation of an underutilized technology. A good product development project should also have a large and profitable market that is growing, should have barriers of entry (such as patents) against imitators for a period of time, and should not have safety or environmental problems.

In Chapter 1, we discussed historic product innovations, both motivated by external market-pull and by internal technology-push motivations. Figure 12.1 shows a quadrant that was popularized by the Boston Consulting Group as the “stars and dogs” chart, and later popularized by Donald Stokes as Pasteur’s quadrant. The horizontal axis shows the strength of external market-pull, and the vertical axis shows the level of



**Figure 12.1** Product opportunities from market-pull and from technology-push

internal technology-push. The lower right quadrant represents market needs that are not matched by adequate technology, and the upper left quadrant represents technical capabilities that are not matched by adequate market demand. The mission of product innovation is to find ways to go from these two quadrants to the upper right quadrant, which is the promised land combining high external need with high internal capabilities, and a wonderful place to do business. We ignore the dog in the lower left quadrant.

## 12.1 Market-Pull Opportunities

Marketing people pay attention to major market trends in the world, as well as individual segments of opportunity. The normal assignment of a research and development organization in a manufacturing company is innovations in response to market needs in the current product areas. Companies seldom encourage their researchers to investigate leads outside of their main product lines, as the results would probably not be synergistic with existing manufacturing and marketing capabilities. Independent research organizations and university researchers have more freedom to widen their investigations, especially if they are financed by government and foundation funding. When they find significant research discoveries, they need to persuade an organization to finance the expensive process development phase, leading to manufacturing and marketing. We list here some examples of enticing targets that are much in the news.

### World Market Mega-trends

At the beginning of the 21st century, some of the world market mega-trends include:

- Growing population of senior citizens in all the more-developed countries, due to better health, longer life, and lower birth rates.
- Economic growth in East and South Asia, at a rate that is much higher than the more-developed countries in North America, western Europe and Japan.
- Continued growth of population and urbanization in less-developed countries, to constitute a larger fraction of humanity and to create new mega-cities.

- Greater globalization of trade, manufacturing, marketing, and labor migration.
- Greater realization of environmental degradation and need for protection, especially in the more-developed countries.

Each of these trends creates new market opportunities and alters the importance of existing markets. For instance, the aging of the population in the more-developed countries creates a greater geriatric market for things used by the aged, including devices such as walking canes, reading glasses, and hearing aids, as well as medicine used to deal with degenerative diseases such as diabetes, hypertension, cardiovascular problems, and cancer. The economies of the more-developed countries have also been shifting steadily from the primary industries, which is the winning of resources from nature, such as farming, fishing, forestry, and mining, to the secondary industries, which are manufacturing and construction. Most of the developed nations have gone further into the “post-industrial” phase, to shifting the bulk of their labor force and gross national product into the tertiary or service industries. The market for goods to the service market will continue to gain in importance compared with the market for goods to the agriculture or manufacturing markets.

The economic growth in Asia would create new markets that may have distinctly different buying habits, which needs study and changed product design. The other major determinant of buying power is the gross national product per capita, which has also been growing. The best opportunity may be in products with income elasticity much greater than unity, so that their demand will increase rapidly when the customers become more affluent. For instance, there has been a startling change in world diet, towards more meat and fish raised by industrial farming methods. Between 1980 and 1998 the consumption of meat and fish increased from 200 to 260 lbs per person each year in North America, from 80 to 180 lbs in Europe, and from 20 to 60 lbs in developing countries. Goods that were considered luxury items in the past will become more commonplace in the future.

The growing populations in the less-developed countries will be more concerned for a long time with the goods of physiological needs and basic necessities. The world population, which stands now at 6.2 billion, quadrupled in the 20th century. In 1900, 86% of the world population lived in rural areas, but only 53% are still in rural areas in 2000, and the rural percentage is expected to decrease steadily. The future world population is dependent on the lifetime fertility rate of an average woman, which should be 2.1 if the world population is to remain unchanged. This ratio was as high as five or six a few decades ago in nations such as Brazil, Egypt, India, and Mexico. As a result of increasing urbanization (which is associated with declining family size) and with the AIDS epidemic in Africa, the world population is expected to grow each year at slightly higher than 1.2%, which still produce 77 million extra people per year. About 97% of the growth will be in less-developed nations, and is dominated by six nations: India, China, Pakistan, Nigeria, Bangladesh, and Indonesia. In the meantime, the fertility rate in western Europe, such as Italy, has fallen to 1.2 children per woman, which is not sufficient to maintain the population at the current level. The rapid aging of the population and labor shortage are other consequences of the low birth rate.

Globalization forces a corporation to look beyond its own national border for finance, technology, raw materials, manufacturing, labor, and marketing. China and India used to have socialist isolationist economies that forbade or discouraged foreign

imports of goods and investment, but they have aggressively switched to globalization with spectacular results. If globalization continues to expand, the study of international market needs must be increased for an exporting nation to stay in the race. There are a few holdouts, such as North Korea, Burma, and Cuba, who do not participate in globalization, and are mired in poverty. Globalization, as a force that destroys traditional means of livelihood and dilutes national culture, has also been resisted by militant groups in many countries, who have made appeals to close the door to cheap foreign goods in order to protect native jobs. This resistance has also spread to America in recent years, as the scandal of “exporting American jobs,” particularly to East Asia.

Environmental concerns began in the richer nations after the basic needs have been satisfied, and is beginning to embrace the poorer nations that are still struggling to satisfy basic needs. There has also been a great increase in waste products from human civilization each year, from gases that go into the atmosphere, liquids that go into the ground and rivers, and solids that go into incinerators and landfills. These trends have profound implications on the redesign of products for marketing. The desire to protect the environment has also led to the development of new markets, such as organically grown food without the use of fertilizers, new HFC refrigerants without chlorine, and new technologies that emit less carbon dioxide and other greenhouse gases.

### Batteries for the Information, Biomedical, and Vehicle Technologies

The weakest link in mobile information technology is the availability of compact and long-lasting electrical supplies. A laptop computer can run on a battery for no more than 2–3 h, which would not survive a coast-to-coast flight. A PowerPoint presentation cannot be projected onto a screen without an AC power outlet. Cellular phones, digital cameras, and musical jukeboxes cannot be used for a long time away from a power outlet. But whereas Moore’s law suggests that communications technologies double in effectiveness every 18 months, the advances in battery technology move at a much slower pace.

Many life-saving and life-enhancing biomedical devices require electricity for their operation. This poses a problem if the patient is to move around and lead a more or less normal life. Some 80,000 new patients each year require implantable heart pacers to keep the heart beating in regular rhythm; it has a very modest power requirement and is currently equipped with an implantable battery to power it for 6 years. A failure of the battery would mean a surgical operation to extract the old battery and to insert a new one. The much more power-hungry implantable mechanical heart is becoming a reality, and it comes with an external rechargeable battery pack that may last 2 h, so that the patient can be free to walk around the room without being constantly plugged into an AC circuit. However, an outing for a movie or a walk in a public garden would require a battery cart with wheels (Altman 2001).

Even more power hungry are moving vehicles, such as wheelchairs, golf carts, and carts in poorly ventilated spaces, such as in warehouses, mines, and ship holds. The all-electric automobile that can travel for 300 miles at 60 m.p.h without recharging would be the greatest prize.

Batteries are divided into primary batteries (used only once) and secondary batteries (rechargeable). The estimated 1989 U.S. market was \$2.3 billion for primary

batteries and \$4.3 billion for secondary batteries. A battery's energy content can be given in the volumetric parameter of watt-hours/liter or in the gravimetric parameter of watt-hours/kilogram. A standard primary alkaline battery may have a capacity of 130 Wh/kg, and a lithium-iodine battery for heart pacers may have a capacity of 290 Wh/kg. The rechargeable lead-acid battery for automobiles has a much lower capacity of only 35 Wh/kg, and a nickel-cadmium has even lower capacity of 28 Wh/kg. Furthermore, many rechargeable batteries must be totally drained before recharging, and cannot be used while being recharged. The lithium-ion battery is a great improvement over the nickel-cadmium battery, but its storage capacity of 200 Wh/kg is still an order of magnitude lower than what we desire (Brodd 1992). The theoretical maximum capacity of a lithium-ion battery could be as high as 4400 Wh/kg, which is much higher than gasoline. This is an urgently needed, and potentially very lucrative market.

### Artificial Organs and Tissue Engineering

When an organ is defective or worn, life is threatened or lessened. There is a long history of man-made replacements and of auxiliary equipment, including: crutches for the lame, glasses for the near-sighted, silver fillings for teeth, and titanium hip joints. Tissue engineering is a new approach based on growing replacements from human cells. A scaffold of biodegradable polymers can be seeded with growing cells, so that the cells will grow to fill the space when the polymer degrades. There are some 450,000 surgical grafts each year to mend broken or damaged bones, using bone from elsewhere in the body. However, the quality and quantity of bones that one can harvest is not sufficient. Researchers are now growing replacement tissues to repair damaged organs, such as the liver and heart. There is more optimism for the near-term target of structural tissues, like bones and cartilage, which are not as complex and multifunctional as the liver and heart. Bone engineers erect a scaffold by inserting a matrix of special material, such as collagen and hydroxyapatite, into gaps in bones (Langer 1994, Service 2000).

### Biodegradable Polymers

Some synthetic polymers are designed to last a long time in the environment, and this stability can be both an advantage and a problem. Natural polymers, however, always degrade in the environment, and this involves bond scission in the backbone of the polymer to fragments with lower molecular weights. There are many modes of degradation, including chemical attack by acids and bases, thermal degradation at elevated temperatures, biological attacks by microbes and enzymes, mechanical degradation by shear forces, light-induced degradation by ultraviolet or visible light, and high-energy radiation attack from X-ray and particle radiation.

The biodegradable properties of polymers can lend themselves to biomedical applications, including: wound management, in absorbable sutures and staples; orthopedic repair, in tendon and cartilage materials; drug delivery systems; and tissue engineering. When a wound created by trauma or surgery is stitched together, and healing takes place over a period of days or weeks, a nondegradable suture has to be surgically removed, which adds to the discomfort and cost of surgery. However, if the suture degrades into small fragments, and is absorbed as a result metabolism by the body, then the

advantages to the patient can be considerable. One of the most successful controlled drug-delivery systems is by surgically inserting a disc or pellets that contain the active drug embedded in a polymer matrix. As the polymer slowly disintegrates, the drug is released continuously and evenly over a period of time. Biodegradable polymers are also used in tissue engineering, in agricultural micro-encapsulation for slow release of fertilizers and insecticides, in support against erosion of freshly cultivated hillsides, and in discardable wrappers for packaging.

A number of materials have been approved for biomedical devices. Glycolic acid ( $\text{HO}-\text{CH}_2-\text{COOH}$ ) and L-lactic acid ( $\text{HO}-\text{CH}(\text{CH}_3)-\text{COOH}$ ) copolymers are well-established raw materials in the manufacture of synthetic biodegradable surgical sutures, clips, staples, and meshes. They have uniform and reproducible mechanical properties and cause minimal tissue reaction, and they do not have to be removed. The glycolides erode at a rate that is much faster than the lactides (Kim 2002), so that a copolymer with different proportions can be fashioned that would produce a large range of degradation and release rates from a few days to several months. Langer (1994) suggested that polyanhydrides of sebamic acid (SA),  $\text{HOOC}-(\text{CH}_2)_8-\text{COOH}$ , can be copolymerized with polyphenoxy carboxypropane (PPCP), so that different proportions of the more biodegradable SA to the more resistant PPCP in copolymerization give a range of release rates from 1 day to 3 years.

### Blood Substitutes

There is usually not enough blood available for transfusion, which is needed when a patient has lost a good deal of blood. There are the problems of matching Rh factor and of the ABO blood types, there are concerns about AIDS contamination, and the cost of blood is \$100 per half pint. The red blood cells are the carriers of oxygen, and they should bind up to 100% saturation with oxygen when the tension is about 100 mmHg in the lung, and they should release almost all the oxygen to the tissues when the oxygen tension is below 10 mmHg. The red blood cells have diameters of about 9  $\mu\text{m}$ , which is larger than a typical capillary with a diameter of 7  $\mu\text{m}$ ; thus, if there is a clogging of the blood vessels, it is difficult for the red blood cells to squeeze through. Blood substitutes should equal or improve on these properties of real blood. Other requirements for a substitute are that it can be stored at room temperature for years without refrigeration, and that it can fight infections.

Some of the research activities in blood substitutes use cow's blood without cells, which can be obtained from slaughterhouses. Another idea is perfluorohydrocarbons, which have a good solubility for oxygen and carbon dioxide, are free from immune-response problems, have no granules and easily go around partially clogged arterioles, and are metabolized in the body after a period of time (Feder 2001).

## 12.2 Technology-Push Opportunities

New scientific discoveries and new technologies are being made and accumulate every day, forming a formidable set of opportunities waiting for creative product ideas and market applications. Some of them are tantalizing scientific curiosities with no obvious

applications, even many decades after their discoveries. Some find applications after many marketing opportunities have been considered and a host of difficult development problems have been solved. Since this strategy of new product innovation leads from internal technological strength, it should be taken by investigators and organizations that are confident that they possess the appropriate scientific knowledge and technological capabilities.

The term platform technology is often used for a technology that has been successful in one application and can be extended to other applications after minor modifications. This creates another set of opportunities, in finding marketing opportunities for currently underutilized technologies.

### Major Scientific and Technological Advances

New scientific discoveries and technological advances have consistently been driving forces for change in the last few decades. The dominant scientific discoveries in the last few decades have been in genetic biology, new materials, and in cosmology. The greatest technological advances have been in the information technologies, in computing and telecommunications. Many authors think that information technology is the greatest force for change; this is sometimes called the Third Wave, which accelerated around 1960 and has roared into full gale in the 21st century. This has been compared to the First Wave in 8000 BC, when agriculture and land became the sources of wealth, and the Second Wave in AD 1750, when industrialization and machinery became the main creators of wealth. Other authors give more homage to the biological revolution, which began with the discovery of the structure of DNA in 1953, and continues to today with the completion of the Human Genome Project, which is waiting to produce great future wealth. A significant difference between these two advances is in the easy implementation of innovations in computers and the Internet, in contrast to the need for approval of new drugs in costly and time-consuming trials supervised by the FDA.

Advanced materials represent another field of rapid advances, and the field of nanotechnology has received a great deal of attention. These fields are producing new understandings and materials that hold the promise of new products. Besides these giant waves, there are numerous smaller advances that do not capture the headlines, but which have great potential for applications. A few of the notable advances will be singled out for discussion below.

#### Electrically Conducting Polymers—Heeger, MacDiarmid, and Shirakawa

Metals have a sea of delocalized electrons that are not bound to specific nuclei, and these are responsible for both their high electrical and heat conductivities. Organic compounds do not ordinarily have such delocalized electrons, and are poor electrical and heat conductors. In fact, they are used to coat electrical wires as insulators. If polymers can be made to conduct, then they could be of a great deal of utility, as they are light, can be processed at low temperatures and pressures, and the raw material does not depend on metal mining. However, the possibility of a conducting polymer did not arise till very recent times.

Hideki Shirakawa of the University of Tsukuba was interested in making polymers of polyacetylene. In the early 1970s, a graduate student in his laboratory prepared a new form of polyacetylene when he mistakenly added 1000 times more catalyst to the reaction mixture than the recipe called for. The film of all-*trans*-polyacetylene produced in the reaction reflected light like aluminum foils, which was not the dark material they expected. Alan G. MacDiarmid was a professor at the University of Pennsylvania, who met Shirakawa in Tokyo and heard about his discovery. MacDiarmid invited Shirakawa to visit Penn to work with him and his colleague Alan J. Heeger. Shirakawa arrived at Penn in 1977, and the three researchers tried to dope the polyacetylene with iodine; the resulting polymer film became golden in color, and its conductivity increased more than a billion times to  $10^5$  S/m. By comparison, Teflon has a conductivity of  $10^{-16}$  S/m, and silver and copper have conductivities of  $10^8$  S/m (Nobel eMuseum 2003). When polyacetylene is exposed to iodine, it is oxidized and the polymer chain loses an electron, leaving a hole, and the electron moves to form the counter-ion  $I_3^-$ . For this work, they received the 2000 Nobel Prize in Chemistry.

The first conducting polymer could not be processed, as it was unstable in air and was sensitive to humidity. Work is under way to improve the properties to make useable products. To date, the best electrically conducting polymer is still much below copper in conductivity. There is hope that their properties can be improved to the point that they would make electronic devices, anti-static substances for photographic films, shields for computer screens against electromagnetic radiation, and “smart” windows that can exclude sunlight. Semiconducting polymers have recently been developed into light-emitting diodes, solar cells, and displays in mobile telephones and mini-format television screens. It remains to be seen whether a useful product will result from this discovery.

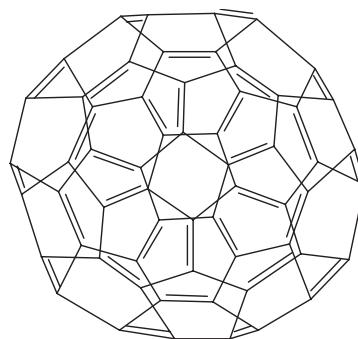
### Fullerene—Curl, Kroto, and Smalley

Everyone “knew” that there are precisely three forms for the element carbon: amorphous carbon, as in charcoal; crystalline graphite, which is packed in hexagonal sheets; and crystalline diamond, which is packed in three-dimensional tetrahedral networks.

Harold Kroto of the University of Sussex was active in microwave spectroscopy, and he discovered a number of interesting spectrum lines in carbon-rich giant stars. He described them as a kind of long-chained molecule of only carbon and nitrogen, and he also found them in interstellar gas clouds. Richard E. Smalley of Rice University was doing research in cluster chemistry, of aggregates of atoms or molecules. He designed and built a special laser-supersonic cluster beam apparatus able to vaporize almost any known material into a plasma of atoms, and studied the design and distribution of the clusters. He was particularly interested in metal clusters, and he worked with a colleague Robert F. Curl at Rice, whose background was in microwave and infrared spectroscopy. Since Kroto knew Curl, he learned that it was possible to use Smalley’s instrument to study the vaporization and cluster formation of carbon to simulate the hot parts of stellar atmosphere (Nobel eMuseum 2003).

In September 1985, Kroto arrived in Smalley’s laboratory to start the experiments on carbon vaporization together with Curl. They vaporized carbon by directing an intense pulse of laser light at a carbon surface. The released carbon atoms were mixed

with a stream of helium gas and combined to form clusters of a few to hundreds of atoms. The gas was then led into a vacuum chamber, where it expanded and was cooled into a vacuum chamber to a few degrees above absolute zero. They also discovered methods to influence the size distribution of the carbon clusters, and they found that the molecular weight of 60 occurred quite prominently, as well as the number 70, in the mass spectrometer. Instead of a long carbon chain, they speculated that the C<sub>60</sub> cluster could be a truncated icosahedron with the structure of a soccer ball, similar to the geodesic dome invented by the American architect R. Buckminster Fuller. This is a uniquely symmetrical molecule in three dimensions that is both beautiful and satisfying. For this work, they received the 1996 Nobel Prize for Chemistry.



A whole new chemistry has been developed around this discovery, and the unusual properties have given rise to suggestions that it could be made into products for a superconducting material, a three-dimensional polymer, new catalysts, new materials with unusual electrical and optical properties and very high mechanical strength, sensors, nanotubes, nanowires, and so on. At this moment, there are, as yet, no products based on the fullerene on the market.

### Combinatorial Chemistry

The simultaneous synthesis of many different but related compounds in a reactor vessel creates a library of compounds, and the automatic analysis of many of their properties carries the promise of rapid and inexpensive search for materials with unusual and superior properties. At this moment, there is still no big success story of an outstanding new product in the market, nor reports of profit.

### Genomics

Genomics is the complete understanding of all the locations and sequences in the tens of thousands of genes in humans. What are the functions associated with these sequences? What proteins are made from these gene sequences, what are their one-dimensional strings and three-dimensional architectures when folded appropriately? Can we use this knowledge to diagnose diseases, and to make material to cure these diseases?

## References

- Altman, L. K. 2001. Self-contained mechanical heart throbs for first time in a human. *New York Times*, 4 July, A10.
- Brodd, R. 1992. Battery. In *Kirk-Othmer Encyclopedia of Chemical Technology*. 5th edition, New York: John Wiley & Sons.
- Feder, B. J. 2001. Scalpel! Clamp! ... Hemopure? *New York Times*, 4 March; 15.
- Kim, Y. H. 2002. An overview on biodegradable polymers in biomedical application. In *ICS-UNIDO Edp EGM*, Trieste, Italy.
- Langer, R. 1994. Polymer systems for controlled release of macromolecules, immobilized enzyme medical bioreactors, and tissue engineering. In *Advances in Chemical Engineering*, vol. 19. San Diego, CA: Academic Press; 1–50.
- Nobel eMuseum. 2003. <http://www.nobel.se/chemistry/laureates>.
- Service, R. F. 2000. Tissue engineers build new bone. *Science* 289: 1498–1500.

## Further Reading

- Campbell, A. M. 2003. *Discovery of Genomics, Proteomics, and Bioinformatics*. San Francisco, CA: Benjamin Cummings.
- Cantor, C. 1999. *Genomics: the science and technology behind the Human Genome Project*. New York: Wiley.
- Czarnik, A. 1997. *A Practical Guide to Combinatorial Chemistry*. Washington, DC: American Chemical Society.
- Dagani, R. 2002. Electrifying plastics. *Chemical & Engineering News* (16 October): 4–5.
- Dalton, A. B. et al. 2003. Super-tough carbon-nanotube fibers. *Nature* 423: 803.
- Fan, L. T. and S. K. Singh. 1989. *Controlled Release: A Quantitative Treatment*. Berlin: Springer-Verlag.
- Gallagher, D. G. 2001. For the errant heart, a chip that packs a wallop. *New York Times*, 16 August; G7.
- Gross, R. A. and B. Kalra. 2002. Biodegradable polymers for the environment. *Science* 297: 803–810.
- Henry, C. M. 2000. Special delivery. *Chemical & Engineering News*. (18 September): 49–65.
- Henry, C. M. 2002. Drug delivery. *Chemical & Engineering News* (26 August): 39–47.
- Jain, R. 1994. Barriers to drug delivery in solid tumors. *Scientific American* 271: 58–65.
- Morrison, P. and P. Morrison. 1982. *Powers of Ten: About the Relative Sizes of Things in the Universe*. New York: Scientific American Library.
- National Research Council. 2002. *Beyond the Molecular Frontier: Challenges for Chemistry and Chemical Engineering*. Washington, DC: The National Academies Press ([www.nap.edu](http://www.nap.edu)).
- Reich, R. 1991. *The Work of Nations: Preparing Ourselves for 21st Century Capitalism*. New York: Vintage Books.
- Robertson, J., T. A. Friedmann, D. B. Geohegan, D. E. Luzzi, and R. S. Ruoff, eds. 2001. *Nanotubes, Fullerenes, Nanostructured and Disordered Carbon: Symposium Held April 17–20, 2001, San Francisco, California, U.S.A.* Materials Research Society Symposia Proceedings, vol. 675. Warrendale, PA: Material Research Society.
- Schnabel, W. 1981. *Polymer Degradation: Principles and Practical Applications*. Munich: Carl Hanser Verlag.
- Simon, J. L. and H. Kahn, eds. 1984. *The Resourceful Earth*. Oxford: Blackwell.

- Sipe, J., C. A. Kelley, and L. A. McNicol, eds. 2002. *Reparative Medicine: Growing Tissues and Organs*. Annals of the New York Academy of Sciences, vol. 961. New York: New York Academy of Sciences.
- Stokes, D. E. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovations*. Washington, DC: Brookings Institute Press.
- Toffler, A. 1980. *The Third Wave*. New York: William Morrow.
- Toffler, A. 1990. *Power Shift: Knowledge, Wealth and Violence at the Edge of the 21st Century*. New York: Bantam Books.

# Appendix

## A.1 Basic Constants and Conversion Factors

Basic constants

Conversion factors

Physical and chemical properties  
Environmental and safety properties  
Medical properties

## A.2 Periodic Table of the Elements

## A.3 Public Domain Databases

General, population, social, economic data

## A.4 Software for the Course

## A.1 Basic Constants and Conversion Factors

### Basic Constants

Constant	Symbol	Value
Speed of light	$c$	299,800 km/s
Electronic charge	$e$	$4.803 \times 10^{-10}$ esu
Avogadro's number	$N_a$	$6.023 \times 10^{23}$ molecules/mol
Electron mass	$m_a$	$5.486 \times 10^{-4}$ amu (g/ $N_a$ )
Proton mass	$m_p$	1.007 amu
Neutron mass	$m_n$	1.008 amu
Faraday constant	$F$	96,487.0 coulomb/mol
Planck constant	$h$	$6.626 \times 10^{-34}$ J·sec
Bohr radius	$a_0$	$5.291 \times 10^{-11}$ m
Gas constant	$R$	8.314 J/(K mol)
Boltzmann constant	$k$	$1.381 \times 10^{-23}$ J/K
Gravitational constant	$G$	$6.670 \times 10^{-8}$ N m <sup>2</sup> /g

## Conversion Factors

<i>From</i>	<i>To</i>	<i>Multiply by</i>
Atmosphere	bar	1.013
Atmosphere	inch of mercury	29.92
Atmosphere	pascal	101,325.0
Atmosphere	pound per square inch	14.695
Atmosphere	torr	760
BTU	calorie	251.996
BTU	joule	1055
Calorie	BTU	1/252
Calorie	joule	4.184
Electronvolt	joule	$1.602 \times 10^{-19}$
Gallon	fluid ounce	128
Gallon	liter	3.78
Gas constant	calorie/(mol K)	1.987
Gas constant	J/(mol K)	8.314
Gas constant	liter/(atm/mol K)	0.082057
Mile	kilometer	1.609
Pound	gram	453.59
Radian	degree	57.295 (180/ $\pi$ )

## A.2 Periodic Table of the Elements

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H 1.01																	2 He 4.00
2	3 Li 6.94	4 Be 9.012											5 B 10.81	6 C 12.01	7 N 14.01	8 O 16.00	9 F 18.99	10 Ne 20.2
3	11 Na 22.99	12 Mg 24.31											13 Al 26.98	14 Si 28.08	15 P 30.97	16 S 32.06	17 Cl 35.45	18 Ar 39.9
4	19 K 39.10	20 Ca 40.08	21 Sc 44.96	22 Ti 47.88	23 V 50.94	24 Cr 51.99	25 Mn 54.94	26 Fe 55.85	27 Co 58.93	28 Ni 58.69	29 Cu 63.54	30 Zn 65.39	31 Ga 69.72	32 Ge 72.59	33 As 74.92	34 Se 78.96	35 Br 79.90	36 Kr 83.8
5	37 Rb 85.47	38 Sr 87.62	39 Y 88.91	40 Zr 91.22	41 Nb 92.91	42 Mo 95.94	43 Tc 98	44 Ru 101.1	45 Rh 102.9	46 Pd 106.2	47 Ag 107.9	48 Cd 112.4	49 In 114.8	50 Sn 118.7	51 Sb 121.8	52 Te 127.6	53 I 126.9	54 Xe 131.2
6	55 Cs 132.9	56 Ba 137.3	57 La▲ 138.9	72 Hf 178.5	73 Ta 180.9	74 W 183.9	75 Re 186.2	76 Os 190.2	77 Ir 192.2	78 Pt 195.1	79 Au 196.9	80 Hg 200.6	81 Tl 204.4	82 Pb 207.2	83 Bi 208.9	84 Po 209	85 At 210	86 Rn 222
7	87 Fr 223	88 Ra 226.0	89 Ac◆ 227.0															

Lanthanides ▲ and actinides ◆:

58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

### A.3 Public Domain Databases

#### General, Population, Social, Economic Data

Census Bureau. U.S. Statistical Abstract	<a href="http://www.census.gov/statab/www">http://www.census.gov/statab/www</a>
<i>CIA World Fact Book</i>	<a href="http://www.cia.gov/cia/publications/factbook">http://www.cia.gov/cia/publications/factbook</a>
World Bank	<a href="http://www.worldbank.org/data">http://www.worldbank.org/data</a>
United Nations	<a href="http://unstats.un.org">http://unstats.un.org</a>
Nobel eMuseum	<a href="http://nobelprize.org">http://nobelprize.org</a>
National Academy of Engineering	<a href="http://www.nae.edu">http://www.nae.edu</a>

#### Physical and Chemical Properties

<i>NIST Chembook</i> . National Institute of Science and Technology, physical–chemical properties	<a href="http://webbook.nist.gov/chemistry">http://webbook.nist.gov/chemistry</a>
---	---

#### Environmental and Safety Properties

EPA. Environmental Protection Agency. EPA suite of database and software	<a href="http://www.epa.gov/oppt">http://www.epa.gov/oppt</a>
NIOSH. National Institute of Occupational Safety and Health	<a href="http://www.cdc.gov/niosh">http://www.cdc.gov/niosh</a>
OSHA. Occupational Safety and Health Agency	<a href="http://www.osha.gov/">http://www.osha.gov/</a>
TOXNET. National Institute of Health, National Library of Medicine, toxicology information	<a href="http://www.toxnet.nlm.nih.gov/chemistry">http://www.toxnet.nlm.nih.gov/chemistry</a>

#### Medical Properties

Mayo Clinic	<a href="http://www.mayoclinic.com">http://www.mayoclinic.com</a>
<i>Merck Manual</i>	<a href="http://www.merck.com/mmhe">http://www.merck.com/mmhe</a>

## A.4 Software for the Course

1. Making spreadsheet, graphs of data, statistical analysis: Lotus 1-2-3; Excel.
2. Quantum mechanical computation of molecular structure: HyperChem, molecular structure; Spartan; Gaussian; Cache.
3. Kinetic theory of molecular assembly: atomic microscope, molecular assemblies.
4. Property estimation methods, group contribution method: Cranium; EPA Suite; EPA ECOSA.

*This page intentionally left blank*

# Index

ab initio method 56  
accountability 285  
ACGIH *see* American Conference of Government Industrial Hygienists  
adamantane 319  
adsorbents 320–324  
age group 257, 258, 260  
aging 342  
air force 316  
alternate concepts 309, 313, 326  
American Conference of Government Industrial Hygienists (ACGIH) 286  
analgesics  
  narcotic 202  
  non-narcotic 202  
anesthetics 202  
  local 202  
artificial organs 244  
aspirin 14  
associations 58  
atoms, properties 76–79  
average values 163

balance of trade 266–267  
batteries 343–344  
Battleship 231  
Beilstein 67  
benzene–toluene–xylenes (BTX) 321  
Bhopal 292–293  
biodegradable polymers 344–345

birth rate 257  
blood substitute 345  
Boltzmann distribution 105, 108  
bond lengths 209  
bond scale 87  
Boolean search 59  
Boston Consulting Group 340  
botox 25  
bottled drinking water 310, 316  
BTX *see* benzene–toluene–xylenes  
buckminsterfullerene 62, 98, 237  
business and government goods 249, 255  
business plan 333–334

carmustine 327  
Carothers, W.H. 28, 35–37  
cash-flow analysis 331–333  
celluloid 13  
CFCs *see* chlorofluorocarbons  
Chain, E. 39  
characteristic temperature 107  
chemical and allied products 247  
chemical processing industries 246–254  
*Chemistry WebBook* 66  
chlorination of drinking water 18  
chlorofluorocarbons (CFCs) 200  
Clausius–Clapeyron 173  
Clean Air Act 286  
Clean Water Act 286  
cluster strategy 231

coefficient of determination 164  
 COGS *see* cost of goods sold  
 combinatorial chemistry 234–236, 348  
 combinatorial synthesis 235  
 commercialization  
     nylon 37  
     penicillin 40  
     taxol 45  
 commodity 251  
 complex fluid 310  
 composite material 193  
 composite solids 310, 325  
 compressibility 131  
 condensation polymers 35  
 connectivity indexes 181  
 consumer goods 240  
 consumer products 256  
 Consumer Product Safety Commission (CPSC) 214  
 controlled drug release 325–329  
 converge or cluster 231  
 convergent search 233  
 coordination number 132  
 Corfam 23  
 corn steep-lactose 39  
 correlation 57, 155  
 correlation coefficient 163  
 correlation function 163  
 cost of goods sold (COGS) 331  
 Coulombic electrostatic force 86  
 covalent solids 136  
 covariance 333  
 CPSC *see* Consumer Product Safety Commission  
 cradle-to-grave stewardship 289–290  
 critical constants 130  
 cross-elasticity 254  
 databases 62, 64  
 DDT 19  
 deep tank fermentation 40  
 degeneracy 107  
 degrees of freedom 167  
 demographics 257–260  
 density 208–211  
 derivative 55  
 design project 336–339  
 deterministic strategy 234  
 development, nylon 36  
 diabetes 325  
 diffusion 112

dipole 83  
 dipole moments 83, 101–103  
 disorder 113  
 divergent strategy 234  
 docking molecules 102  
 DuPont 35  
 economic growth 342  
 electrically conducting polymers 346–347  
 electronic parameters 157  
 electronics industry 247  
 electrostatic forces 86  
 emission, pollutant 296  
 empirical structure–property relations 200  
 energy densities 269, 270  
 energy of mixing 192  
 entropy 107, 113  
 environment 343  
 Environmental Protection Agency (EPA) 20, 56, 63, 211, 286, 287, 290, 311  
 EPA *see* Environmental Protection Agency  
 ether 16  
 evaluation 329  
 explosions 290–291  
 extrapolation 156, 169–170  
 Exxon Valdez 301  
 FDA *see* Food and Drug Administration  
 fertility rate 342  
 flame and explosion 290–291  
 flash point 213  
 flash temperature 291  
 Fleming, A. 27, 38  
 Flixborough 291  
 Florey, H. 39  
 fluid mixtures 189  
 Food and Drug Administration (FDA) 41, 213, 290, 293  
 force fields 82  
 forward search 53, 55  
 freeze drying, penicillin 40  
 freon 5, 15  
 Fry, A. 30  
 fullerene 347–348  
 functional form 160  
 functional group 158  
 gasoline 256, 263, 268–277  
 octane number 272  
 volatility 272

- genomics 348  
 global warming 296, 299  
 global warming potential 222  
 globalization 342–343  
 green process engineering 301–302  
 green product engineering 302–306  
 greenhouse gases 222, 296  
 group contribution methods 58, 181  
 guided search 229, 230
- Hansch and Leo's method 186  
 HCFC *see* hydrochlorofluorocarbons  
 heat capacity 106  
 Helmholtz free energy 106  
 Henry's law 192  
 heteroatoms 210  
 heuristic strategy 232  
 hierarchy of needs 261  
 HOMO (highest occupied molecular orbit) 159  
 homogeneous fluid solution 309, 316  
 homologous series 157–158  
 hot button 276  
 household incomes 260  
 Human Genome Project 346  
 hydrochlorofluorocarbons (HCFC) 158, 174  
 hydrogen bond 85  
 hydrophobic parameters 157
- ideal gas 125  
 incandescent lamp 17  
 income elasticity 254, 274  
 individual buyers 255  
 ink, for inkjet printers 277–279
  - surface tension 278
  - viscosity 278
 inkjet printers 277  
 internal rate of return (IRR) 331  
 international buyers 255  
 interpolation 60  
 ionic character 88  
 ionic solids 134  
 IRR *see* internal rate of return  
 isomer series 158  
 ivory 303
- Joback's method 183
- kerosene 317  
 Kevlar 24
- Lange's Handbook of Chemistry* 62  
 Langmuir isotherm 161  
 LD<sub>50</sub> *see* lethal dose 50 214, 293  
 lead compound 55, 230, 233  
 lead optimization 233  
 LEL *see* lower explosion limit  
 lethal dose 50 (LD<sub>50</sub>) 214, 293  
 linear function 58  
 linear regression 161–165  
 literature search 55, 62  
 lower explosion limit (LEL) 212, 288, 291  
 LUMO (lowest unoccupied molecular orbit) 159
- major brands 279  
 male/100 females ratio 258  
 market and environment parameters 54  
 market segmentation 254–256  
 market-pull 340, 341–345  
 material structure 55  
 maximum contaminant level (MCL) 218, 294, 295  
 Maxwell equation 193  
 MCL *see* maximum contaminant level  
 mean free path 126  
 melt spinning, nylon 37  
 metallic solids 137  
 methadone 234  
 methyl *tert*-butyl ether (MTBE) 272  
 microsphere 327–330  
 microtubules 44  
 Midgley, T., Jr 5, 6, 200  
 Minamata Bay 294  
 missile fuel 265, 316–320  
 mix-and-split 235  
 mixtures 189  
 molecular dynamics 109  
 molecular mechanics 57, 82  
 molecular orbitals 80  
 molecular solids 137  
 molecular structure–property relationships 239–241  
 monomers 35  
 Monte Carlo 109  
 morphine 233, 234  
 motor vehicle accidents 270  
 MTBE *see* methyl *tert*-butyl ether  
 multilinear function 58  
 multilinear regressions 174  
 musk 202, 240

- navy 316  
 net present value (NPV) 331  
 nitrogen mustard 41  
 nonlinear function 58, 160  
 nonlinear regression 165–174  
 normal paraffins 154  
 Northern Regional Research Laboratory 39  
*NPV* *see* net present value 331  
 nylon 28, 35–38  
 development 36  
 research–discovery 35
- Occam's Razor 167  
 Occupational Safety and Health  
 Administration (OSHA) 56, 286, 292  
 octane number 272  
 oil exploration 233  
 opportunistic strategy 232  
 optical masks 236  
 orbital radii 75–80  
 OSHA *see* Occupational Safety and Health  
 Administration  
 ozone destroying potential 211  
 ozone hole 299
- Pacific yew 42  
 packing efficiency 208  
 parallel synthesis 235  
 partition coefficient 215  
 partition function 105  
 PCBs *see* polychlorinated biphenyls 154  
 penicillin 27, 38–41  
 Perkin, W. 26  
 permissible exposure limit (PEL) 292  
 platform technology 346  
 Plunkett, R. 29  
 polarizability 85  
 pollutant emission 296  
 polyacetylenes 237  
 polyamide 36  
 polychlorinated biphenyls (PCBs) 154  
 polyester 35  
 polynomial regression 168  
 polynomials 161, 166  
 populations 156  
 Post-It 30  
 prediction from theory 56  
 predictors 155, 157  
 price elasticity 252, 274  
 product innovation 45  
 product life cycle 251
- professional buyers 255, 265  
 profiling 259  
 promising directions 230  
*Properties of Organic Compounds* 64  
 property parameters 54  
 pure compound 309
- quantitative structure–activity relations 58  
 quantitative property–property relations  
 (QPPR) 159, 181  
 quantitative structure–property relations  
 (QSPR) 159, 181  
 quantum mechanics 73
- random search 61, 229–236  
 Raoult's law 191  
 real gases 127  
 reduced parameters 131  
 reference dose (RfD) 218  
 refrigerants 5  
 regression  
 linear 161–165  
 multilinear 174  
 nonlinear 165–174  
 polynomial 168  
 remedial clean-up 299, 300, 302  
 research–discovery, nylon 35  
 reverse search 54  
 RfD *see* reference dose 218
- Sackur–Tetrode 115  
 safety 211, 212  
 SAGE *see* selling, administration and general expenses  
 Salvarsan 21  
 Schrödinger's equation 73  
 searchable 59  
 self-assembly 237  
 selling, administration and general expenses  
 (SAGE) 331  
 semi-empirical methods 57  
 semi-synthesis, taxol 45  
 Silver, S. 30  
 solubility 203–208  
 solvent extraction 40  
 specialty chemical 251  
 stability 113  
 standard deviation 163, 164  
 Standard Industrial Classification (SIC) 146  
*Staphylococcus* 39  
 starting point 230

- statistical thermodynamics 105  
steric parameters 157  
structure parameters 54  
submolecular scale 93  
substitution effects 186  
sugar, synthetic substitutes 202  
sulfa drugs 18, 39  
surface tension, of inkjet ink 278  
surfactant 44  
symmetry 98–100, 157, 179  
synthetic dye 26
- taxol 21, 41–45, 302  
technology-push 341, 345–348  
Teflon 29  
tetraethyl lead 14  
thermal motions 95  
threshold limit value (TLV) 288, 292, 298  
air 215  
Tomahawk missile 265, 316, 320  
topological index 160, 181
- toxic chemicals, in plants 291–293  
toxicity 213–221  
Toxic Substance Control Act 286  
trends 58  
triple point 139, 142
- van der Waal constants 128  
van der Waal forces 83, 161  
variance 163  
vinyl chloride 288, 302  
viscosity, of inkjet ink 278  
volatility, of gasoline 272  
vulcanization of rubber 12
- Wall, M. 42  
weighted least squares 165  
wildcat search strategy 61, 230  
world mega-trends 341
- Z matrix 86–87  
zeolites 14, 237, 264–265