

Jim Burns

PhD Thesis:
Studies on complex enzyme systems

Thesis presented for the Degree of Doctor of Philosophy of the
University of Edinburgh in the Faculty of Science. May, 1971.

Transcribed to L^AT_EX by Herbert M. Sauro February, 2023, based on a photocopy
of the thesis obtained in 1988.

Contents

I.	Non-quantitative nature of present system theory	7
II.	Some quantitative approaches	9
III.	Enzyme kinetics	10
IV.	Steady state enzyme kinetics	10
V.	Multi-Enzyme systems	11
VI.	The problem	12
VII.	Role of computer in relation to quantitative experiments	13
VIII.	General strategy for quantitative approach.	15
IX.	Theoretical treatment of metabolic systems	16
1	Real Organisms and the Choice of Kinetic Systems	19
1.	Evaluating the stationary state	22
2.	The limit cycle	24
3.	The use of rate equations	26
4.	Unimolecular systems	28
5.	The straight chain	28
6.	The sensitivity coefficient	31
7.	The network	34
8.	The electrical analogue	35
9.	Stability	38
10.	Nonlinear systems	41
2	Mathematical Formulation of Complex Systems	45
1.	Subsystem and complete system	46
2.	Network of general enzymically catalysed transformations	47
3.	Necessity for linear independence of the functions	51
4.	Independent fluxes in a system	52

5.	Introduction of genotypic parameters	53
6.	Non-enzymic transformations	55
7.	Expanding systems	55
8.	Stability of networks of general reactions	56
9.	Control by repression and induction of gene activity	56
10.	Complete growing systems	59
3	Sensitivity Analysis	61
1.	Sensitivity coefficients	63
2.	Operational meaning of the coefficient	63
3.	Formal relations of different measures	65
4.	Block coefficients	66
4.1	Coefficient with respect to an enzyme	69
4.2	Coefficient with respect to turnover number	69
4.3	Coefficient with respect to gene dosage	69
5.	Measurement of coefficients in a ‘real’ system	70
6.	Summation properties of block coefficients	71
7.	Elasticities of complete flux expressions	72
8.	Relative Importance of two successive enzymes in a straight chain	73
9.	Relation between the coefficients of all blocks affected a given pool	75
10.	System coefficients and a given independent flux	77
11.	Relations between the system coefficients	79
12.	Application of theorems relating system coefficients to elasticities	80
13.	General limitations on coefficient values	87
14.	Pool sensitivities in a chain of saturable enzymes	87
15.	Pool sensitivities in a chain of linear enzymes	89
16.	Flux coefficients	90
17.	‘Diagnostics’ for rate control in non-linear systems	91
18.	Pools attenuate with metabolic distance	94
19.	In vivo destination of elasticities	97
20.	Modulation to determine elasticities and coefficients	102
21.	Determination of $C_{\alpha_j}^{F_1 S_i}$ for a system	103
22.	Optimal allocation for growth	104

23.	Straight chain of linear enzymes	105
24.	Optimal solution given alternative food sources	107
4	Computer Methods For Studying The Steady State Behaviour Of Theoretical	109
1.	Multi-Enzyme systems	109
2.	The Analogue computer	110
3.	Digital computer systems	110
4.	Design of suitable simulation system	113
5.	Implementation on the KDF 9	116
6.	Editor (precompiler)	116
7.	The simulation language	117
8.	Loading and running a system	121
9.	Loading	121
10.	Details of step	123
11.	Running	124
12.	The simulation program	125
13.	Working section	126
14.	Numerical aspects of locating the s.s. and estimating sensitivity coefficients	127
5	DOMINANCE EPISTASIS AND SELECTION	129
1.	Dominance and epistasis in a straight chain of enzymes	130
2.	Dominance	132
3.	Epistasis	135
4.	Implications of study for response to selection	136
5.	The phenotype/genotype relation for more complex systems	140
6.	Dominance, when the allelic substitution has only a small effect on pools.	141
7.	Epilogue	145

Introduction

Advances in biochemistry and molecular genetics have been concerned with gene action, protein synthesis, and protein structure. Thus we now have a fairly detailed picture of how protein structures are determined, both as regards specific configuration and as regards the mechanism and energetics of their syntheses. Furthermore by a combination of genetic and biochemical investigations in the area usually described as 'metabolism' we are obtaining an ever widening picture of the relationships between metabolic steps, often displayed as metabolic maps which show in a qualitative way the flow of synthetic and degradative pathways (anabolism and catabolism). Although this picture is by no means complete there is a general feeling that the major structure of living systems, at this level, is known. Broadly speaking the genetic specification, i.e. the base sequence of the DNA, determines the parameters (quantity, turnover numbers and Michaelis constants) of the enzymes which catalyse the biochemical transformations of the metabolic network. What is much less understood, however, is how such specifications affect the quantitative outcome of the multiple interactions within a metabolic system which, in the end, determines the measurable properties of the whole system or organism. What is, therefore, required is a "quantitative biochemistry" as well as the relation of this to quantitative genetics, whose approach has been both historically and methodologically quite different from molecular genetics. This thesis is an attempt to relate the three fields of molecular genetics, metabolic biochemistry and quantitative genetics.

I. Non-quantitative nature of present system theory

During the past few decades the joint application of genetics and biochemistry to relatively simple organisms such as viruses, bacteria and syncytia has led to an increasing elucidation of their details. There has been great success in discovering system components, isolating them from homogenates, studying their properties *in vitro* and defining their relationship to the complete system. We now possess a fairly universal classification of system components into such types as genes, messenger RNA, ribosomes, proteins and a knowledge of their respective functions within the cell. The detailed action of these subunits may vary across strains and species but at least at the level of cell biology the general picture is fairly clear. For some types of organism such as *E.coli*, yeast and *Neurospora* a great deal is already known. Thus it is claimed that about 40% of all metabolic activity has been

accounted for in *E. coli* K12 (J.D. Watson, 1965).

It seems then that, as this detail is added to, we shall, at the level of cell biology, be increasingly in the situation of, viewing the organism in terms of a sort of chemical engineer's block diagram. The engineer's blocks now represent the separate genetically specified transformational activities (enzyme catalysis) taking place in metabolism and the engineer's system variables become the various chemical concentrations being transformed one to another in the cell. Indeed, the cell has been viewed in just this way for a long time, the great achievement being to define and characterise what all the separate 'blocks' are. Thus a large class of data is explained for a given cell type by a transformational block diagram, tentatively proposed for that cell. Such data is essentially qualitative such as, for example, whether a given enzymic activity is present or absent in the extract of a mutant or whether the mutant can grow or not on a minimal medium with a certain supplement. The block diagram explains the data insofar as it allows qualitative prediction of the effects of hypothesised large changes to blocks within the diagram.

Thus we have a reasonably well established 'qualitative system theory' of cell metabolism which is consistent with a large body of qualitative data and this theory is similar in general outline over a wide range of cell types.

More recently there has been considerable interest in extending transformational diagrams by adding to them a 'control level' showing which system variables have a strong effect on the rate of any given transformation, although in this case the details have been found to vary considerably over different cell types. These studies have involved the consideration of a less qualitative type of data, for example, the measurement of low, intermediate and high levels of an enzyme in extracts of cells grown under various conditions. Nevertheless, even with the inclusion of a control level, the present system theory of cell metabolism has remained essentially qualitative in nature. Unlike the engineer, the biologist is not at present easily able to use his block diagram to predict the quantitative effects of substituting a slightly altered form of one of his blocks (enzymes) or of altering the level of an external metabolite.

What will be the form of a quantitative theory? Clearly with the existing system theory one worked from a block diagram, itself constituting the theory, and used a form of logical argument to predict the consequences of any particular lesion. With a quantitative approach it will be necessary to attach to each block in the diagram a formula of some sort giving the rate of its transformational activity. This formula being in terms of the system variables which strongly affect the block and of the genetically specified parameters which describe it.

Although any change in a block or environmental parameter will directly affect the rates of only a few activities these will affect system variables which will affect other activities and so on. The quantitative response of any system property is thus the result of a complex balance between the various activities underlying the system. In a quantitative theory prediction will thus require a calculation which allows fully for the detailed behaviour of the separate blocks as well as for the block diagram which was sufficient for a qualitative

theory.

Just as the qualitative system theory aimed to explain a wide range of qualitative data, so any quantitative extension must now include a much larger body of quantitative data which, although it may not conflict with the transformational and control diagram, is by no means explained by it.

II. Some quantitative approaches

I shall consider first some studies which have avoided explanation of particular data and have asked instead what general properties may arise out of commonly occurring metabolic mechanisms. Thus Goodwin (1963) focussed attention on possible dynamic consequences of the feedback loops found to control enzyme quantity in a variety of cells, conjecturing that they may be in a state of spontaneous oscillation. He has further considered cellular dynamics from the point of view of there being a number of such oscillators, weakly coupled, within each cell and treated this by the methods of statistical mechanics.

Following on this and also because of the clear experimental findings concerning metabolic oscillation there has been considerable discussion of the conditions under which such oscillations could occur. J.S. Griffiths (1968a, 1968b) has considered oscillators of the Goodwin type and defined more clearly the necessary conditions, his work suggesting that rather special assumptions are required. The problem of stability has also been considered by C. Walter (1968a, 1968b). A more abstract treatment is offered by S.A. Kauffman (1969) who has considered the cell as a system in which genes have only two states, on and off, the state of a gene in the next instant of time being determined by the state of a small number of genes controlling it at the previous instant. This abstraction of a problem, not unlike Goodwin's coupled oscillators, has enabled him to consider large networks with extensive coupling and to demonstrate that a number of interesting properties are possessed by such systems. Thus it appears that a network of 'genes' connected at random and with randomly assigned boolean response function to their inputs will for the case where each gene has two inputs, almost always possess only a few distinct cycles of activity into which it rapidly falls when released from an arbitrary initial condition. Furthermore these cycles of activity are surprisingly short, involving often only a very few 'states' of the system before repeating. This simple cyclic or even steady state behaviour is of some interest arising as it does in systems which at first sight might be expected to generate extremely complex dynamical behaviour. A further value of Kauffman's approach is that he relates cell metabolic theory to the already much studied theory of 'finite automata'.

These approaches are, however, highly abstracted formulations remote from quantitatively ascertainable data and are, therefore, not suitable to make the link between the areas with which this thesis is concerned. More relevant to our purposes are the classical biochemical approaches as follows.

III. Enzyme kinetics

Here there exists an extensive literature, see for example "The Enzymes" (1959), the bulk of which is concerned with studies, both theoretical and experimental, on single enzymes. The basis of these studies is usually that the enzyme can exist as a number of alternative 'complexes' with its substrates, products and inhibitors and that interconversion between these is governed by the usual chemical law of mass-action.

This gives rise to a set of 1st order differential equations involving a number of hypothetical enzymic intermediates and the necessary rate constants to describe their mass-action interconversions. The data to which this theory is related is largely derived from *in vitro* studies, preferably with purified enzymes, and consists of the concentrations of substrates, products, and other substances at various times after the introduction of the enzyme. Only rarely is the concentration of an enzymic intermediate monitored directly and again the early rapid reaction phase is not normally observed. Consequently for most metabolic enzymes complicated kinetic schemes should be viewed with suspicion as also their rate constants, particularly when one considers that the conditions *in vivo* may be very different from that of the purified enzyme in isolation.

IV. Steady state enzyme kinetics

This considers the situation when substrates and products are held at a steady concentration and asks what is their respective rate of removal or production by the enzyme. Bimolecular reactions are usually considered to take place only between enzymic intermediates and metabolites, protein interactions not normally being included. Under these circumstances a set of simultaneous equations can be written down which define the steady state levels of the various enzymic complexes and since these equations are linear in terms of the enzyme complexes it is possible to solve them explicitly. This yields the overall 'rate' of the reaction as an algebraic expression involving the steady levels of the metabolites and the various rate constants characteristic of the protein. King and Altman (1956) have provided an elegant method for writing down the 'rate expression' directly from a diagram of the kinetic scheme. The results of applying this systematically to a variety of kinetic schemes has been presented by W.W. Cleland (1963) who also advocates that the expressions be written using Michaelis constants and other enzyme kinetic constants which can be defined operationally, rather than with elementary rate constants. The simplest expression appearing in Cleland's paper being for the well known case of an enzyme catalysed reaction having one substrate and one product (H.L. Segal, 1959), when the rate expression takes the form

$$\text{FLUX} = \frac{ET/M (S_1 - S_2/K_E)}{1 + S_1/M + S_2/M}.$$

where E is the concentration of the enzyme.

Here M , M' are Michaelis constants which are usually obtained directly from an assay for the forward and reverse reactions respectively as could T the turnover number corresponding to 'maximum velocity' in the forward direction. K_E is the equilibrium constant for the reaction $S_1 \rightarrow S_2$. This equation reduces to the well known rate expression due to Briggs & Haldane (1925) when S_2 is set at zero.

This flux is a good approximation to the instantaneous value, even when the pools are varying slowly (Richards and Pring, 1968). We shall make use of equations of this type in later chapters. The rate expressions for even slightly more complicated kinetic schemes can become very cumbersome. Consequently several workers, for example R.O. Harst (1968) have resorted to programming the King Altman method so that these more complicated rate equations can be produced automatically on a digital computer from a description of their kinetic schemes.

V. Multi-Enzyme systems

Perhaps the most consistent attempt at a quantitative treatment of a specific multi-enzyme system has been in relation to the glycolytic pathway, started by Chance in 1951. Since that time B. Chance, D. Garfinkel, J. Higgins and other workers have continued to develop experimental, theoretical and computer techniques - recently reviewed by D. Garfinkel (1970). Their reason for becoming involved in computer techniques is that the mathematical representation of even a simplified multi-enzyme system involves a large number of non-linear differential equations which cannot be handled by existing analytic techniques and have consequently to be solved by 'simulation' on a digital or analogue computer. Their main interest has been in a study of the transient effects in the glycolytic pathway based on work with mouse ascites cells (B. Chance et al., 1960) and with heart and liver, (Garfinkel et al. 1968).

Long computation times are found in these systems, mainly due to the commonly encountered situation that the molar concentration of enzymic intermediates is much lower than that of metabolic intermediates. This produces high frequency components in the D.E. and the step length of the numerical integration procedure used has to be sufficiently small to cope with these. As a result considerable amounts of computer time can easily be expended in representing only a little 'real' time in the biochemical system (Garfinkel, 1965).

Consequently an effort has been made to improve the methods of integration employed in simulation programmes so that they are more suitable for handling the 'stiff' equations commonly encountered (Cooper, 1970). Recently (Curtis, A., unpublished data, 1970) the predictor-corrector method of Gear (1968) has been employed on a simulation problem, previously run using the method of E. Chance (1969), with a claimed improvement in speed of 100 times.

Another approach to the problem of computer efficiency has been to use the rate-equations of steady state enzyme kinetics which were discussed earlier, thus removing the differential

equations corresponding to enzymic intermediates, which are mainly responsible for the stiffness encountered in biochemical systems. These rate equation representations are now, of course, only an approximation to the 'true' dynamics but this will not normally be a serious discrepancy (Rhoads and Pring, 1968).

A problem which has received considerable attention is that of how to 'describe' the biochemical system to the computer and the more recent programs of Gerfinkel (1968) and of E. Chance (1969) will accept a form of biochemical-language which is meaningful to the experimental biochemist. For programmes employing mainly rate-equation methods the use of a biochemical language is not so straightforward but, (Rhoads and Pring, 1968), methods have been devised to generate these automatically and link them into the commonly employed simulator programs (Garfinkel, 1968).

Some unexpected results have arisen on the experimental side. Thus Chance et al. (1964) find sustained oscillations, apparently arising at the level of intermediary metabolism. This has stimulated theorist's, as for example Achs and Garfinkel (1968) to construct oscillatory models. Again Garfinkel (1969) remarks that very few enzymes, working *in situ*, behave as they do in isolation.

All of the quantitative work just mentioned assumes implicitly that the cell can be regarded as chemically homogeneous or 'well mixed', although it is possible to include a limited type of inhomogeneity.

VI. The problem

The purpose of this thesis is to develop a suitable theoretical framework together with relevant computer techniques, which it is hoped will illuminate the behaviour of genetical systems. By genetical systems is meant an organism (or its analogue) or a population of organisms which is specified by a set of initial parameters (genes and environment) and into which we can introduce variation. Such variation will be introduced either by considering alternative functional alleles at one or more genetic loci, or by changes in the environment.

An alternative allele will be represented by a quantitative change in the specification of a catalytically active molecular species (enzyme) while the environmental variation will be represented by a quantitative change in the concentration of one or more external sources of metabolites.

The essence of genetic methodology is the measurement of difference between organisms or populations. Which measurements are taken is arbitrary, but an analysis of systemic behaviour requires that the functional dependence of any measure is analysed in terms of the effect of the imposed variation on the system. Our task, therefore, is to investigate how a metabolically organised system responds to variation in the parameters, which define it.

The measures in which we are interested are mainly time independent averages of some sort and are usually of a biochemical nature. Some examples would be concentration of a

metabolite, quantity of an enzyme in a cell, net-flux in a metabolic pathway. The averages being obtained by sampling from a large exponential growing population, of asynchronous cells, from a single syncytium in exponential growth or from particular tissues of adult metazoa. Consequently we can argue (Chapter I) that a suitably extended ‘steady state’ treatment of metabolism is the appropriate development from qualitative system theory. This steady state treatment must, however, include bimolecularity, inhibition of enzyme activity, repression and induction of enzyme quantity, saturation properties of enzymic reactions, and networks of arbitrary complexity since these are essential aspects of any real system and cannot be ignored without sacrificing important features of quantitative behaviour. It should be stressed that although much has been written on the theory of the steady state this literature has dealt almost exclusively with ‘linear’ systems leaving out of consideration the non-linear phenomena mentioned above.

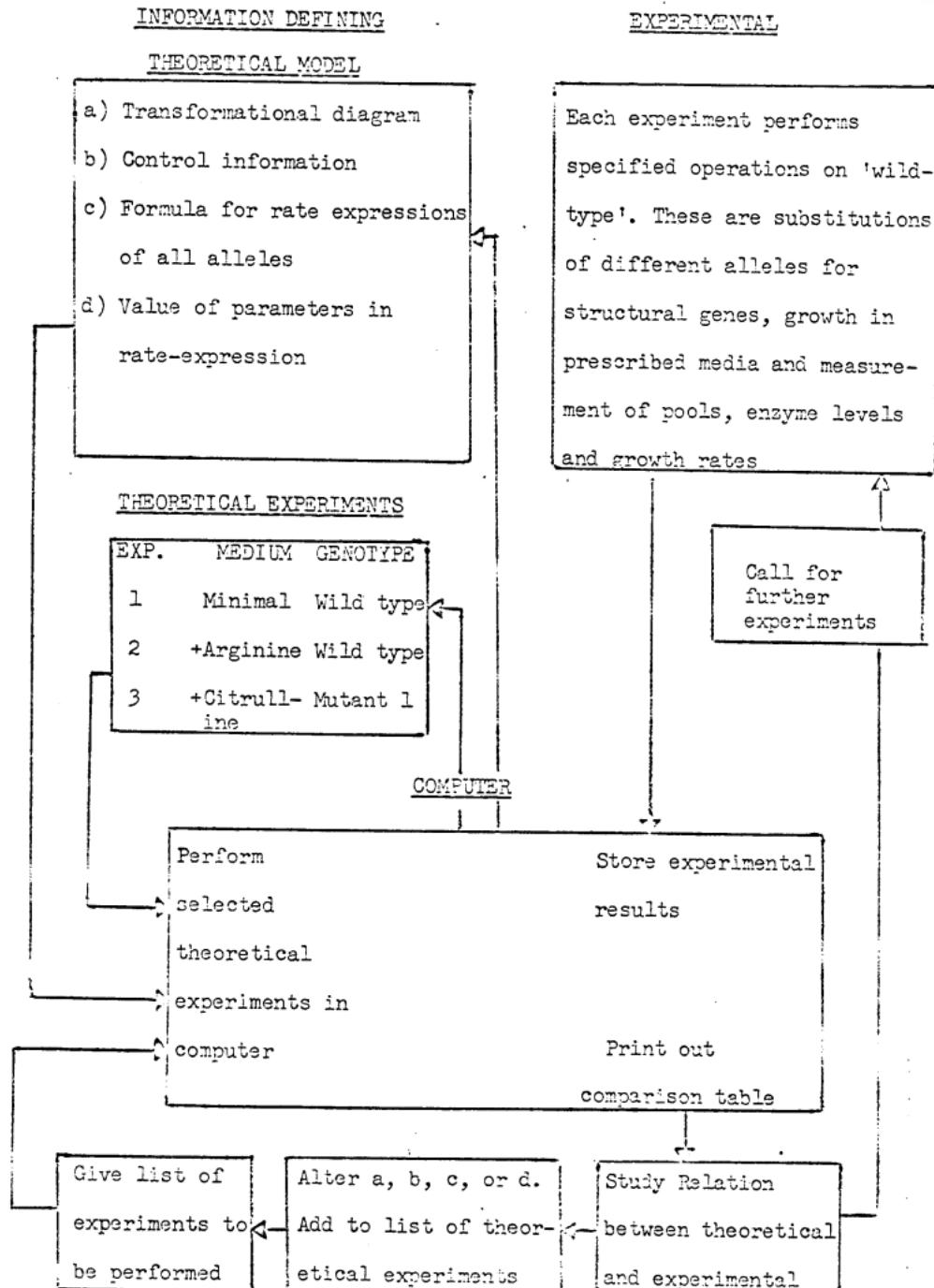
VII. Role of computer in relation to quantitative experiments

Let us consider how one would set about establishing such a quantitative model within an area of metabolism where the block structure and the control structure has been tentatively established by means of qualitative information.

I have in mind particularly the sort of work which is being carried out in laboratories studying pathways in structurally simple organisms, but it is by no means restricted to these. Essentially the procedure is to measure a number of system variables, metabolic pools, enzyme levels, flux in pathway, both for the wild type organism and for strains with various quantitatively altered forms of enzymes. These ‘quantitative’ mutants may be obtained as revertants from known mutants, affecting structural loci, which abolished enzyme activity, or may be discovered as natural variants. Measurements are taken both on minimal medium and on medium containing various supplements.

Thus a fairly large amount of data is gathered to which the proposed quantitative model must conform. In addition there will be some rough information, gained from in vitro studies, about rate expressions for the wild type and mutant enzymes which will be helpful in formulating the initial theoretical model. Similarly, this kind of information may become available from natural or domestic animal populations. The facing diagram indicates an example of the necessary procedure.

There are several interesting features of this process which should be noted. Firstly, it can be seen that the computer plays an essential role. Without it the theoretical model could not be evaluated against the facts. Secondly, the process is not to be thought of as simply adjusting the parameters of a fixed model to fit the data by, for instance, hill climbing on some least squares procedure. Rather the process is more ‘scientific’ than ‘statistical’. Thus not only the parameters but also the block diagram, the assumptions about control or the form of the rate expressions are all subject to alteration as the model is studied in relation to data.



A good example of this scientific situation and of the integral role of the computer is that the computer can be used to decide a quantitative question that often arises. Which of several experiments is likely to discriminate most sharply between currently competing hypotheses about the model?

Thirdly, it seems that the building up of a quantitative model is a more interactive process than was that for the qualitative model. Thus when we adjust our assumption about one part of the model to meet some new experiments we may disturb existing agreement with previous experiments. The computer is again invaluable here as it enables us to retest agreement with all previous experiments every time we adjust the model.

If this process of interaction between biological data and computer model proves successful the result would be twofold:

(1) We would be in possession of a model which consistently reflects the qualitative and quantitative results of the experimental data. This means that our confidence in having the "correct" model is strengthened. (2) We would have at our disposal a predictive tool which enables us to investigate novel situations which are either difficult, lengthy, or expensive to perform on biological material. This process is usually described as 'simulation' and forms the goal of many investigations (and has been successfully applied in many fields of engineering). In the next section we shall argue, however, that this may fail on grounds of feasibility and, more importantly, that the general properties of quantitative systems may be used to answer questions of theoretical importance beyond the problem of the simulation of a particular organism (Kacser and Burns, 1968).

VIII. General strategy for quantitative approach.

At this point we must take stock of the general position in which quantitative cell metabolism is at the moment.

Firstly, if the aim is the setting up of an exact-quantitative model of metabolism in some cell type, we must first argue the value of having such a model. Often the effort required to establish the model will be out of all proportion to its value. Secondly, the feasibility of setting up a comprehensive model of metabolism is doubtful at the present time, even for *E. coli* and certainly for all other cell types. Thus in most cases not even the necessary qualitative knowledge has been obtained. A more fundamental reason, however, is for instance, that the form of the rate expression of many enzymes, even *in vitro*, is still controversial and appears likely to remain so.

The strategy advocated here is that we should concentrate on acquiring the ability to incorporate quantitative information into a predominantly qualitative picture, and not to worry too much about details. As the need and opportunity arises this ability can be applied to various situations. Clearly just as the block diagram has been gradually extended over a period of many years so a qualitative model can gradually be turned into a quantitative model as the necessary reliable data on system disturbance and system response becomes

available.

IX. Theoretical treatment of metabolic systems

We have above painted a rather gloomy picture of the immediate prospects for a comprehensive quantitative approach based on real data and have also questioned its general usefulness, whilst pointing out the relation of computational and theoretical techniques to it.

However the same techniques can, just as easily, be applied to the construction and investigation, either by analysis or numerical exploration, of abstract metabolic systems expressed in the form of mathematical models. These abstract systems being chosen to possess structure, mechanisms and parametric values which are broadly similar to those of real systems. Many features of the quantitative response of real systems to parametric variation are not critically dependant on the exact parametric values obtaining. Such features, including the common occurrence of intra and inter locus interaction, the remarkable buffering ability against parametric change (Kacser, 1963) and so on, clearly arise mainly from the overall structure of the system rather than its precise parametric values. Accordingly we shall argue that it is legitimate to study such quantitative features, of systemic response to parametric variation, for suitably chosen abstract systems *per se* and this is in fact a major aspect of the thesis.

In principle this can be done by the numerical investigation of a wide variety of abstract systems over a wide variety of parameter values but in practice such *in numero* experiments are seldom informative (A. Robertson, 1967) owing to the large number of parameters involved in even the simplest model. Because of this considerable attention has been paid, whenever possible, to the algebraic investigation of system behaviour, both in a general way and, in more detail, for simple systems such as a chain of enzymes. In these studies it has proved useful to introduce the 'sensitivity coefficient' (Kacser and Burns, 1968; Burns, 1968) as an index of response and to show how these coefficients are related to the quantitative behaviour of the individual parts of a system. The theoretical properties of these coefficients can then be used to discuss features of parametric response, such as 'buffering', in a general way. A similar coefficient has been used by J. Higgins (1963) although in a somewhat different context.

Arising from such a treatment it may be possible to consider populations of biochemical systems which differ from each other in many parameters. This is the implicit assumption of quantitative genetics, but the usual theoretical method (e.g. Fraser, 1962, 1967) of handling metric characters in population problems has been to set up arbitrary genotype-phenotype relationships. A quantitative biochemical treatment, however, considerably restricts and defines the relationships insofar as 'additivity', 'epistasis', and 'dominance' must not be postulated but will (or will not) arise from the underlying metabolic structure of the organisms. We shall consider a number of aspects in this manner for a simple abstract system and

will suggest methods for investigating more complex systems by numerical means (Burns, 1970).

1

Real Organisms and the Choice of Kinetic Systems

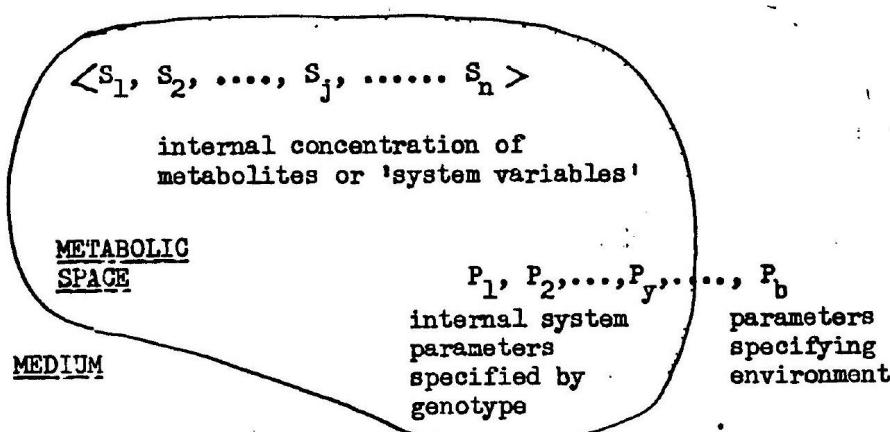
In the introduction it was stated that a major aim would be to provide a suitable theoretical framework, in which to consider how variability in an organism at the genetic (or information storage) level will be expressed at the level of measurements made on the ‘system’ which is the organism. The purpose of this chapter is to show precisely what is meant by this, to provide such a framework, to consider properties of unimolecular systems in relation to it and finally to introduce a number of general concepts in connection with these systems. In Chapter II consideration will be given to the formulation, within this scheme, of more complex metabolic systems which can more reasonably stand as analogues for real organisms.

Before proceeding to consider specific types of systems it is important to make clear the general point of view which is to be adopted. Genes act by specifying the precise properties of functional units, usually proteins, which themselves act in many cases by catalysing specific reactions within a metabolic network of some sort. Accordingly many aspects of the general gene-action problem will be reflected in a more limited and abstract problem, namely that of how enzymes interact within a multi-enzyme system to determine measurable properties of the system, as for example the concentration of a given metabolite or the metabolic flux at a point in the network. Whilst no apology will be made for this abstraction from the general problem, the genetic systems to which the limited treatment is most relevant will clearly be simple unicellular or syncytial organisms growing in a well defined culture medium. For complex multicellular organisms in a complex environment it will be less obviously relevant, although even here it may be applicable to the more biochemical

type of character.

Given then that a first attempt to understand something of the quantitative aspects of gene action may be reasonably limited to a consideration of multi-enzyme systems, two further tactical questions arise. The first of these is concerned with the quantitative importance of the spatial organisation of metabolism within the cell. There is ample evidence that such organisation occurs, as for example at the high level of cell organelles such as mitochondria, and, at a lower level, where it has been suggested that several enzymes in a pathway are physically associated in a particle (Wagner and Bergquist, 1963) thus promoting the passage of substrates from one enzyme molecule to another. However, there is reason to suppose that this organisation, whilst certainly important, is not absolutely crucial to the cell metabolism. Evidence for this is that mutants which block the production of various intermediates can be completely relieved by relatively low concentrations of the intermediate in the external medium. Whilst in principle it would be desirable to include spatial organisation in a treatment of multi-enzyme systems, this would be impossibly difficult, both because of a lack of reliable information about spatial organisation and because it would involve considering chemical concentration as dependent on position and subject to diffusion equations, a great extension of an already complex problem. Certainly, if we have to compromise between either a multi-enzyme system containing many enzymes related by a specific transformational and control diagram, but neglecting spatial organisation, or a system containing only a few enzymes and allowing for spatial organisation, the first alternative must be taken. This is because gene action and interaction are likely to depend crucially on the fact that the 'points of action' of the different genes are embedded in a metabolic network. To dispense with this network would thus be seriously to distort the very aspect of the system which is under consideration, that is gene action.

The second tactical question is concerned with the relative importance to be attached to what might be called the "long term" and the "transient" behaviour of multi-enzyme systems. A general 'open' multi-enzyme is represented below. This exists within a fixed volume and allows passage of metabolites between the surrounding medium and the inner 'metabolic space', across the surface of the volume.



The surrounding space is assumed to be sufficiently large and well mixed that concentrations of metabolites in it can be considered as constants in the ‘environment’ of the metabolising volume. Since perfect mixing is also assumed internally the state of the system will be specified at any moment by the instantaneous molar concentrations (S_1, \dots, S_n) of the n variable metabolites. These, together with the fixed parameters (P_1, \dots, P_b), determine the change which will ensue in the system at the next moment. This situation is expressed mathematically by saying that the ‘equations of motion’ for the system are the set of n 1st ordinary differential equations (D.E.).

$$\frac{d(S_i)}{dt} = \dot{S}_i = \varphi_i(S_1, \dots, S_n, P_1, \dots, P_b) \quad (1.1)$$

The parameters P include the constant concentration of environmental metabolites, kinetic constants belonging to genetically defined proteins, and other determinants such as the volume and surface area of the metabolising space. For the present the precise nature of the nonlinear functions $\varphi_i(\dots)$ may be left open, although it is clear that they will depend on expressions giving the instantaneous rates of the various metabolic transformation both within the system and across its surface. The respective contributions of genotype and environment to the properties of an organism are thus represented in the abstract problem as the way in which the solution of (1.1) depends on P_1, P_2, \dots, P_b .

‘Long term’ means that interest is primarily in the solution of (1.1) a long time after its release from initial conditions S_i^0 . Such a solution may approach a stationary steady state, when a set of values S_i arise such that all the \dot{S}_i become zero, or alternatively, a steady-state, in with the S_i arise such that S_i do not become time independent but a trajectory arises such that the system repeats its motion at regular intervals, T , satisfying the conditions

$$\int_{\alpha}^{\alpha+T} \dot{S}_i dt = 0$$

where α is an arbitrary large time. Such steady state trajectories are clearly completely characterised if the values of the S_i are known at just one point on the trajectory. A third possible long term solution is one which is neither stationary nor steady state.

It seems clear, in the context of this discussion, that attention should be concentrated on the long term solutions (1.1). Such solutions do not depend precisely on arbitrary initial conditions, but only on the system’s genetically specified set of proteins together with the concentration of chemicals external to the system. If there are alternative solutions the S_i^0 will determine which one is reached. In principle the way is now clear to study the dependence of the long term solutions of (1.1) on the parameters P_1, P_2, \dots , such dependence being representative of the more general dependence of the phenotype on the genotype and the environment for real genetic systems (organisms).

1. Evaluating the stationary state

In practice, however, even such abstract studies cannot be carried out by mathematical manipulation of (1.1) but, owing to the large number of system variables S_i and the non-linearity of the functions $\varphi_i()$, must be done numerically with the aid of a computer only on the simplest assumptions, namely that all $\varphi_i()$ are linear functions of the S_i , can an explicit algebraic stationary steady state solution of (1.1) be obtained and such linear models, whilst of considerable interest, do not properly allow for basic aspects of metabolic organisation like bimolecularity and feedback inhibition.

With regard to obtaining the long term solution of (1.1) by numerical means and for specific values of the parameters $P_1, P_2 \dots$, there are two distinct situations, that in which the solution is a stationary one, and that where it is a non-stationary steady state or limit cycle. In both cases our interest lies not only in the solution but also in the response or sensitivity of this solution to the parameters P , that is in the influence of variation of genotype and environment on the phenotype. This response will appear naturally in the form of a matrix since it is the response of the phenotype, represented by a group of measurements on the system, to the genotype represented by a group of system parameters.

For the case of a system possessing a stationary solution the equations to be solved are, from (1.1), the n non-linear simultaneous equations

$$\varphi_i(S_1, S_2, \dots, P_1, P_2, \dots) = 0 \quad (1.2)$$

Suppose we possess a concentration vector S_i such that each equation differs from zero by only a small deviation d_i then

$$d_i = \varphi_i(S_1, S_2 \dots) \quad (1.3)$$

The problem of finding the stationary solution \bar{S}_i is that of how to adjust the S_i in (1.3) so that all d_i approach zero.

Let δS_i be the necessary adjustment to S_i so that $\bar{S}_i = S_i + \delta S_i$. Now, since \bar{S}_i is the exact solution, $\varphi_i(\bar{S}_1, \bar{S}_2, \dots) = 0$

or

$$\varphi_i(S_1 + \delta S_1, S_2 + \delta S_2, \dots) = 0$$

which, ignoring terms above 1st degree in a Taylor expansion, gives

$$\varphi_i(S_1, S_2, \dots) + \sum_{j=1,n} \frac{\partial \varphi_i}{\partial S_j} \times \delta S_j = 0$$

In vector and matrix notation and using (1.3) this becomes

$$d_i \simeq \left[\frac{\partial d_i}{\partial S_j} \right] \cdot \delta S_j \quad (1.4)$$

An approximate δS_i is thus given by solving (1.4)

$$\delta S_j = - \left[\frac{\partial d_i}{\partial S_j} \right]^{-1} \cdot d_i \quad (1.5)$$

The numerical solution of (1.2) then consists in progressively correcting the existing S_1 using (1.3) and (1.5) until the deviations d_i are sufficiently close to zero. This ‘iterative’ process will be returned to in CH. IV when the numerical and computer aspects will be discussed.

The matrix $\left[\frac{\partial d_i}{\partial S_j} \right]$ or $\left[\frac{\partial \varphi_i}{\partial S_j} \right]$, evaluated at the stationary solution, is also involved in finding the sensitivity of this solution to small changes in P_1, P_2 , etc. In this case if the original solution was \bar{S}_i the new solution, $\bar{S}_i + \delta \bar{S}_i$ is one corresponding to, say, the change $P_1 \rightarrow P_1 + \delta P_1$, and we have the stationary condition (1.2) to satisfy for both the old and the new solution.

$$\text{Thus } \psi_i (\bar{S}_1, \bar{S}_2, \dots, P_1, P_2, P_3 \dots) = 0, \text{ for } i = 1 \text{ to } N$$

$$\text{and also } \psi_i (\bar{S}_1 + \delta \bar{S}_1, \bar{S}_2 + \delta \bar{S}_2, \dots, P_1 + \delta P_1, P_2, P_3) = 0, \text{ for } i = 1 \text{ to } N$$

Using these two conditions and assuming δP_1 sufficiently small for higher terms in the

Taylor expansion to be ignored we get

$$\begin{aligned} \frac{\partial \varphi_i}{\partial P_1} \cdot \delta P_1 + \sum \frac{\partial \varphi_i}{\partial S_j} \cdot \delta \bar{S}_j &= 0 \\ \frac{\delta \bar{S}_i}{\delta P_1} &= - \left[\frac{\partial \varphi_i}{\partial S_j} \right]^{-1} \cdot \frac{\partial \varphi_i}{\partial P_1} \end{aligned} \quad (1.6)$$

The relation (1.6) shows how stationary values of the system variables \bar{S}_i will respond to a small change in a parameter and can be used as a basis for considering how a general phenotype ‘measure’ made on a stationary system will respond. Such a measure, which reflects one aspect of the phenotype, is conceived as being the result, \bar{M} , of a particular measuring operation carried out upon the stationary system. \bar{M} will depend upon the variables defining the stationary state and possibly some system parameters so that in general

$$\bar{M} = \bar{M} (\bar{S}_1, \dots, \bar{S}_n, P_1, \dots, P_b) \quad (1.7)$$

Here the notation $\bar{M}(\dots)$ merely spans, the “function of certain arguments, whose value is \bar{M} ”. The response of \bar{M} to a particular parameter, say P_1 , is conveniently defined as the ‘sensitivity coefficient’ $C_{P_1}^{\bar{M}}$ which compares the fractional response in \bar{M} to the small fractional disturbance in P_1 which produced it. This concept is developed further in CH. III but for the moment we can write

$$C_{P_1}^{\bar{M}} = \frac{P_1}{\bar{M}} \frac{\partial \bar{M}}{\partial P_1}$$

Using (1.6) and (1.7) we therefore get

$$\mathbf{C}_{P_1}^{\bar{M}} = \frac{1}{\bar{M}} \frac{\partial \bar{M}}{\partial S_j} \times \left[\frac{\partial \varphi_i}{\partial S_j} \right]^{-1} \times P_1 \frac{\partial \varphi_i}{\partial P_1} \quad (1.8)$$

More generally the phenotype is described by a group of such measures $\bar{M}_1, \dots, \bar{M}_x, \dots, \bar{M}_a$ and the genotype and environment by a group of system parameters subject to variation $P_1, \dots, P_y, \dots, P_b$. The sensitivity of the phenotype to the genotype and the environment will then be displayed in the matrix of coefficients $\left[C_{P_y}^{\bar{M}_X} \right]$, where

$$\left[C_{P_y}^{\bar{M}_X} \right] = \left[\frac{1}{\bar{M}_X} \frac{\partial \bar{M}_X}{\partial S_j} \right] \times \left[\frac{\partial \varphi_i}{\partial S_j} \right]^{-1} \times \left[P_y \frac{\partial \varphi_j}{\partial P_y} \right] \quad (1.9)$$

The matrix $\left[\frac{\partial \varphi_i}{\partial S_j} \right]$ which appears in (1.7) and (1.9) is in fact also the matrix which underlies the dynamical behaviour of the system when close to the stationary state.

Thus suppose the system to be close to a stationary state so that $S_i = \bar{S}_i + X_i$ where X_i is small. Inserting this in (1.1) and ignoring terms above 1st order in the expansion we get

$$\dot{X}_j = \left[\frac{\partial \varphi_i}{\partial S_j} \right]_{S_j=\bar{S}_j} \times X_j \quad (1.10)$$

thus the ‘dynamical matrix’, D , of the system is just $\left[\frac{\partial \varphi_i}{\partial S_j} \right]_{S_j=\bar{S}_j}$.

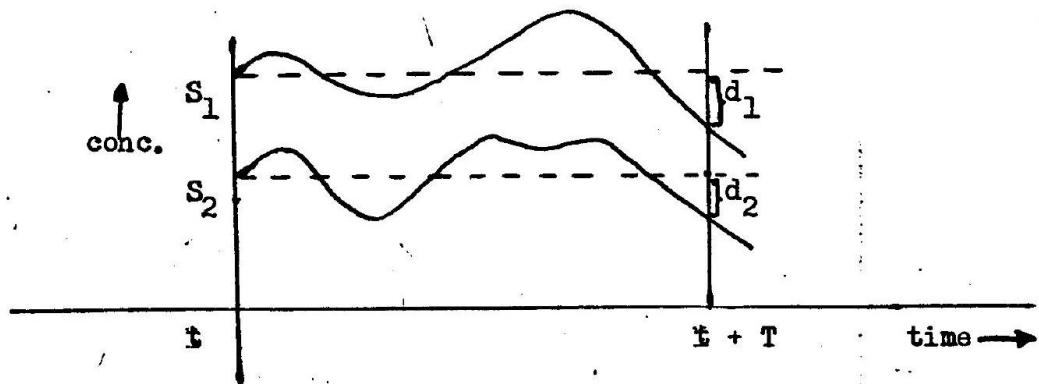
The stability or otherwise of the stationary solution will then depend on whether the ‘eigenvalues’ of D all have negative real parts.

2. The limit cycle

Can similar techniques be used when a system does not possess a stationary steady state? Let us consider a system possessing a limit cycle type of ‘long term’ solution, when it may be possible to use such a technique. Suppose that we have a trajectory close to an exact solution and defined by a concentration vector S_i and period T then the deviations defined by

$$d_i = \int_t^{t+T} \varphi_i(S_1, S_2, \dots) \cdot dt \quad (1.11)$$

correspond to the deviations of (1.3) in the stationary case. The integration referred to in (1.11) is performed on the set of D.E. (1.1), starting at time t with the proposed vector S_i . The meaning of these residues is illustrated below.



The problem of finding an exact solution is, as before, to manipulate the approximate solution until the d_i are sufficiently close to zero. This can be done by adjusting all but one of the S_i and also the approximate period T , parameters which can be designated ξ_1 to ξ_n . The holding constant of one of the S_i merely serves to define the arbitrary time, t , which is being used to characterise the final limit cycle. The small corrections to be applied to the existing values of the parameters are calculated by the relation

$$\delta\xi_j = -d_i \left/ \left[\frac{\partial d_i}{\partial \xi_j} \right] \right. \quad (1.12)$$

which is analogous to relation (1.5), and a similar ‘iteration’ is performed using (1.12) and (1.11). The matrix $\left[\frac{\partial d_i}{\partial \xi_j} \right]$ can, as before, also be used to discover the sensitivity of measures connected with the limit cycle to the parameters P_1, P_2, \dots, P_b .

There is, however, an important practical difference between the two cases just discussed in that the evaluation of the d_i for the limit cycle case involves, even on the best assumptions, many hundreds of times more computer operations than does the stationary case. This is because the integration of the system of D.E. (1.1) over the period T involves many hundreds of evaluations of the S_i , each of which corresponds to one d_i evaluation in the stationary state computation.

If the period T is large or the equations difficult to integrate the factor may be even worse and the difficulty is further compounded by the requirement that the d_i have to be calculated to a high accuracy if the numerical differentiation involved in estimating the matrix $\left[\frac{\partial d_i}{\partial \xi_j} \right]$ of (1.12) is to be reliable. It appears, therefore, that whilst the stationary type of computation proves quite feasible for reasonably large networks, say less than forty enzymes, it becomes totally impracticable for networks having limit cycle solutions. This conclusion applies a fortiori to systems whose ‘long term’ solution, although regular in some sense, cannot be described by limit cycles. In these circumstances, remembering that an interact in gene action involves networks of some size and complexity, consideration can only be given to systems possessing stable stationary state solutions. This limitation seems unavoidable, whether it turns out to be serious will depend on how frequently non stationary long term behaviour is encountered in metabolic systems under study.

3. The use of rate equations

Accepting then the considerable simplification that only stationary solutions of homogeneous multi-enzyme systems will be considered it is hoped that such solutions will still represent important aspects of the genotype-phenotype relation. This hope rests mainly on the strategy that minimal compromise will subsequently be made in representing the structure of metabolic networks, the structure clearly being critical to any attempt at representing gene-action.

The nature of the system variables S_i has deliberately been left open until this point. A reasonably complete set of such variables for a metabolic system has usually been taken as a set of concentrations of intermediary metabolites together with a set of concentrations of enzymic intermediates. The set of dynamical functions $\varphi(\dots)$ are then written down, using the 'Guldberg-Waage' law of mass-action, and used as a basis for further analysis, as for example in Pring (1967).

However, when interest is centred, as here, on the properties of the stationary state it is possible to avoid considering enzymic intermediates by using the fact that when the complete system is at a stationary state so will be the enzymic intermediates belonging to each enzyme. The 'net flux' carried by each enzyme at the stationary state will then be given by the King-Altman rate expression of steady state enzyme kinetics as discussed in the introduction. Thus the number of system variables is reduced to the number of intermediary metabolites and expressions $\varphi_i(\dots)$ for their rates of change become sums and differences of rate expressions. The expressions $\varphi_i(\dots)$ made up

General theoretical aspects of multi-enzyme systems

$\langle p_1, \dots, p_y, \dots, p_b \rangle$: set of genetic and environmental parameters type P

$\langle s_1, \dots, s_j, \dots, s_n \rangle$: set of system variables, type S.

$\langle \varphi_1, \dots, \varphi_i, \dots, \varphi_n \rangle$: set of functions, $\varphi(S, P)$, specifying 'motion' of system ,

$\langle \bar{s}_1, \dots, \bar{s}_j, \dots, \bar{s}_n \rangle$: set of stationary values for system variables, satisfying $\varphi(\bar{S}, P) = 0$

$\langle \bar{M}_1, \dots, \bar{M}_{x_1}, \dots, \bar{M}_a \rangle$: set of functions, $\bar{M}(\bar{S}, P)$, giving result of different measuring operations on stationary state and collectively characterising the phenotype.

Numerical operations

i) Evaluate matrix $[D'] = \left[\frac{\partial \varphi_i}{\partial S_j} \right]$ at present values of S_j by numerical approximation.

ii) Use it to find \bar{S} by means of the ‘iteration’

$$\begin{aligned} 1: \delta S_j &= [D']^{-1} \cdot \varphi_i \\ S_j &\leftarrow S_j + \delta S_j \\ \rightarrow (\text{goto}) 1 &\text{ unless all } \varphi_i \simeq 0 \end{aligned}$$

Having thus obtained \bar{S} and $D = \left[\frac{\partial \varphi_i}{\partial S_j} \text{ at } S = \bar{S} \right]$

iii) Compute the phenotype/genotype, environment) sensitivity matrix using the relation

$$\left[C_{P_y}^{\bar{M}_X} \right] = \left[\frac{1}{M_X} \frac{\partial M_X}{\partial S_j} \right] \times [D]^{-1} \times \left[\frac{P_y \partial \varphi_i}{\partial P_y} \right]$$

iv) Characterise the dynamics of the system close to the stationary state by extracting eigenvalues of $[D]$.

in this way will only approximate to the ‘true’ dynamics of the system, but Pring (1967) has shown that this approximation may be a very good one. However, the stationary state and its properties will not be in any way affected by this simplification and the dimension of a given system will be reduced by a considerable factor. For example a system with 20 metabolites and 25 enzymes each requiring, on the average, 4 enzymic intermediates to describe its kinetics, would have a dimension of 120 system variables. In this case a reduction in dimension of 6 fold would result from the use of rate expressions in making up the functions φ_i for the 20 intermediate metabolites, properties of the stationary state would be identical with the ‘true’ system and even the dynamics would be a good approximation.

Accordingly, unless otherwise stated, system variables will be taken to be the set of intermediary metabolite concentration. The set of dynamical functions φ_i attached to them will then be an approximation to the true dynamics, involving the use of rate expressions, which nevertheless leaves the properties of the stationary state unaltered.

Our general theoretical and computational framework for multienzyme systems which possess a stationary solution is now clear. This is summarised on p.14 opposite, where it can be seen that the set of dynamical functions $\varphi_1, \dots, \varphi_n$ and the matrix $[D] = \left[\frac{\partial \varphi_i}{\partial S_j} \right]$ play an important role. The remainder of this chapter will consider the particular situation when the φ_1 represent systems of unimolecular reactions. Chapter II will then be concerned with progressively showing how more general systems, having the necessary realistic detail, may be formulated in terms of suitable sets of functions φ_i , themselves based on rate expressions.

The remainder of this chapter will consider the particular situation when the φ_1 represent

systems of unimolecular reactions. Chapter II will then be concerned with progressively showing how more general systems, having the necessary realistic detail, may be formulated in terms of suitable sets of functions φ_i , themselves based on rate expressions.

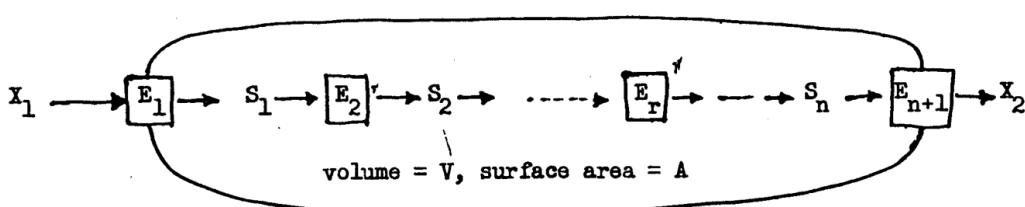
4. Unimolecular systems

These systems, often arising as a first approach to more realistic ones, are of interest because in certain cases explicit solutions for the stationary state, or S.S. in terms of the parameters can be obtained. In addition it proves possible to consider, in a general way, whether or not the S.S. of unimolecular systems are stable or not. Although these explicit solutions and stability properties apply strictly only to unimolecular systems they provide a useful analytic background against which numerical solution of the more general systems of CH II can be evaluated.

For the present unimolecular systems are considered to have a fixed volume, V , and to have enzyme quantities prescribed parametrically (genetically). Later, in CH II, these restrictions will be relaxed somewhat to include ‘growing systems’, when $\dot{V} \neq 0$, and ‘enzyme regulation’ in which enzyme quantity is not parametrically prescribed.

5. The straight chain

A start will be made with the simplest unimolecular system, that of the ‘straight chain of linear reactions’, and this will be considered in some detail. Such a system, consisting of n variable metabolites and $m (= n + 1)$ linear enzymes, is represented in the following diagram as taking place in a fixed volume V and converting an external metabolite, X_1 , to another, X_2 .



By ‘linear’ is meant that the rate-expressions for the separate reactions of type, $A \rightleftharpoons B$, are of the form.

$$f = \rho(A - B/K) \quad (1.13a)$$

where f is the net flux from A to B per unit quantity of enzyme, K is the equilibrium constant for the reaction and ρ is a kinetic constant. In fact every enzyme catalyzed reaction will have a non-linear rate expression, since saturation can always occur and (1.13a)

is merely an approximation which holds provided the concentrations, A and B , remain within certain limits. For example the simple non linear rate-expression for a unimolecular reaction mentioned in the introduction was of the form

$$f = \frac{T}{M} \left(A - \frac{B}{K} \right) / \left(1 + \frac{A}{M} + \frac{B}{M} \right) \quad (1.13b)$$

This approximates to the linear form (1.13a) provided $A \ll M$ and $B \ll M$, the constant ρ then equals $\frac{T}{M}$.

In the diagram E_1, E_2, \dots are used both as ‘names’ for the enzymes and also to refer to their concentrations, similarly for the metabolites S_1, S_2, \dots . Provided that the metabolites are well mixed the enzymes can be distributed non uniformly in the volume in which case E_r refers to the average concentration. Transport enzymes such as E_1, E_{n+1} will have to be situated, not necessarily uniformly, on the surface but their ‘linear’ rate expressions will be formally similar. Even if transport is non-enzymic and therefore depends on the surface area, A , there will be an effective ‘concentration of area’, A/V , which will be genetically (parametrically) defined and will play a similar role to enzyme concentration.

Considering the rate of change of the 1st metabolite in the whole volume

$$\frac{d(V_X S_1)}{dt} = (E_1 \times V) \rho_1 \left(X_1 - \frac{S_1}{K_1} \right) - (E_2 \cdot V) \rho_2 \left(S_1 - \frac{S_2}{K_2} \right)$$

or

$$\dot{S}_1 = E_1 \rho_1 \left(X_1 - \frac{S_1}{K_1} \right) - E_2 \rho_2 \left(S_1 - \frac{S_2}{K_2} \right) - \left(\frac{1}{V} \dot{V} \right) S_1 \quad (1.13)$$

The products $E_r \rho_r$ etc. can be written as v_r . Where v_r means the activity of the r th enzyme per unit volume of the system when the substrate has unit and the product zero concentration. Writing F_1, F_2 , as, the net fluxes/unit volume of the system and remembering that $\left(\frac{1}{V} \dot{V} \right)$ is zero, the set of dynamical function $\varphi_1, \dots, \varphi_n$ are given by

$$\begin{aligned} \varphi_1 &= \dot{S}_1 = v_1 \left(X_1 - \frac{S_1}{K_1} \right) - v_2 \left(S_1 - \frac{S_2}{K_2} \right) = F_1 - F_2 \\ \varphi_2 &= \dot{S}_2 = v_2 \left(S_1 - \frac{S_2}{K_2} \right) - v_3 \left(S_2 - \frac{S_3}{K_3} \right) = F_2 - F_3 \\ &\vdots \\ \varphi_n &= \dot{S}_n = v_n \left(S_{n-1} - \frac{S_n}{K_n} \right) - v_{n+1} \left(S_n - \frac{X_2}{K_{n+1}} \right) = F_n - F_{n-1} \end{aligned} \quad (1.15a)$$

The conditions, $\varphi_i = 0$, for a stationary solution give rise to a set of n linear simultaneous equations for the n unknown stationary values \bar{S}_i and these equations could be solved; because of their simple structure, by straightforward algebraic elimination. However, they

will be treated in a more general way both to illustrate the methods just described and to facilitate extension later to the "non-linear straight chain" where algebraic elimination is not possible.

The equations (1.15a) are equivalent to the 'band matrix' equation

$$\begin{bmatrix} -\left(\frac{v_1}{K_1} + v_2\right), \frac{v_2}{K_2} \\ v_2, -\left(\frac{v_2}{K_2} + v_3\right), \frac{v_3}{K_3} \\ v_3, -\left(\frac{v_3}{K_3} + v_4\right), \dots \\ \vdots \\ v_n, -\left(\frac{v_n}{K_n} + v_{n+1}\right) \end{bmatrix} \begin{bmatrix} \bar{S}_1 \\ \bar{S}_2 \\ \vdots \\ \bar{S}_n \end{bmatrix} = \begin{bmatrix} -v_1 X_1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \frac{v_{n+1}}{K_{n+1}} \end{bmatrix} \quad (1.15b)$$

Hence there is just one S.S. which furthermore turns out to be completely stable in the sense that the dynamics of the approach to this S.S. is a sum of decaying exponentials, (Denbigh, Hicks and Page, 1947). This latter property arises because the dynamical matrix $[D]$, which in this case is identical to the band matrix above, possesses special features. Since these features are of a general nature they will be dealt with later. in this chapter.

However, inverting the matrix gives little insight into how the \bar{S}_i , or indeed important measures of the system such as the pathway flux, \bar{F} , depend on the parameters.. This can be shown as follows. At the S.S. all the separate fluxes are equal to the pathway flux so that:

$$\begin{aligned} \bar{F} &= \bar{F}_1 = v_1 \left(\bar{X}_1 - \frac{\bar{S}_1}{\bar{K}_1} \right) \\ \bar{F} &= \bar{F}_2 = v_2 \left(\bar{S}_1 - \frac{\bar{S}_2}{\bar{K}_2} \right) \\ \bar{F} &= \bar{F}_{n+1} = v_{n+1} \left(\bar{S}_n - \frac{\bar{X}_2}{\bar{K}_{n+1}} \right) \end{aligned} \quad (1.16)$$

If \bar{F} is the correct pathway flux there are $n + 1$ linear equations to solve for only n unknowns, $\bar{S}_1 \dots \bar{S}_n$. This means that the coefficients must be linearly dependent and implies that their $(n + 1)$ determinant is zero.

$$\mathbf{A}_x = \begin{vmatrix} \bar{F} - v_1 x_1, & \frac{v_1}{K_1}, & 0, & 0, \\ \bar{F} & - v_2, & \frac{v_2}{K_2}, & \\ \vdots & \ddots & \ddots & \\ \bar{F} + \frac{v_{n+1}}{K_{n+1}} \cdot x_2, & & & \end{vmatrix} \quad \dots(1.17)$$

Since \bar{F} occurs only in the 1st column this is a simple linear equation in \bar{F} and when expanded gives on simplifying this yields

$$\bar{F} = \begin{vmatrix} v_1 x_1 & & & \\ \vdots & & & \\ -\frac{v_{n+1}}{K_{n+1}} & x_2 & & \\ & & & \end{vmatrix}$$

$$\bar{F} = X_1 - \frac{X_2}{K_1 \cdot K_2 \cdots K_{n+1}} \Big/ \left(\frac{1}{v_1} + \frac{1}{K_1 v_2} + \frac{1}{K_1 K_2 v_3} + \cdots + \frac{1}{K_1 K_2 \cdots K_n v_{n+1}} \right) \quad (1.18)$$

Having found \bar{F} it is a simple matter to compute S_1, S_2 etc. using successively the equations of (1.16).

6. The sensitivity coefficient

The relation (1.18), allows us to see explicitly exactly how the measure \bar{F} depends on the parameters v_1, v_2, \dots . In this case the flux controlling effect of the separate enzymes in, the pathway will be measured by the set of sensitivity coefficients $C_{v_i}^{\bar{F}}$ and an expression can be found, employing the usual definition.

Thus

$$\begin{aligned} C_{v_i}^{\bar{F}} &= \frac{v_i}{\bar{F}} \cdot \frac{\partial \bar{F}}{\partial v_i} = v_i \frac{\partial \log(\bar{F})}{\partial v_i} \\ &= v_i \left(-1 \left/ \frac{1}{v_1} + \frac{1}{K_1 v_2} + \dots \right. \right) - \frac{1}{K_1 K_2 \dots v_i^2} \\ C_{v_i}^{\bar{F}} &= \frac{1}{K_1 \dots K_{i-1} v_i} \left/ \left(\frac{1}{v_1} + \frac{1}{K_1 v_2} + \dots \right) \right. \end{aligned} \quad (1.19)$$

The two properties, $0 < C_{v_i}^F < 1$ and $\sum_{i=1..n} C_{v_i}^F = 1$ immediately follow.

Since (1.19) is of the form, $C_{v_i}^F = \frac{a_i}{\sum a_i}$, it is clear that the separate coefficients are all positive and less than unity and that in addition the sum of the separate coefficients is unity.

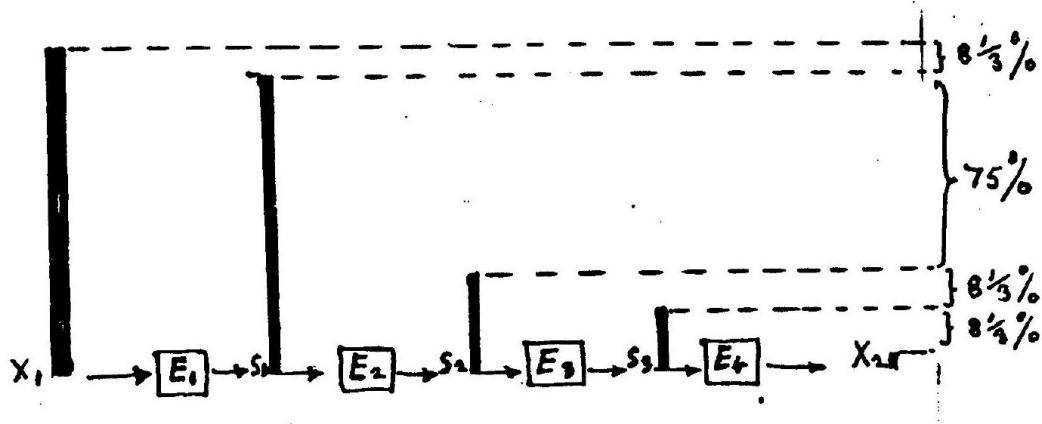
This leads to the important result that if any enzyme is found to have a control coefficient of approximately one it must be the controlling or ‘throttling’ enzyme in the sequence since the sum of all the other control coefficients will be small. For example a system having four reactions all with $K_E = 1$, three catalysed by enzymes of activity v and one of them, say the 2nd, by an enzyme of activity $v/9$ would be expected to have rate-control mainly from the second enzyme.

Broadly how strong this effect is can be seen from substituting in (1.19),

$$C_{v_2}^{\bar{F}} = \frac{1}{v/9} \left/ \left(\frac{1}{v} + \frac{1}{v/9} + \frac{1}{v} + \frac{1}{v} \right) \right. = \frac{9}{12} = 75\%$$

Thus of the total 100% control available 75% is ‘owned’ by the 2nd enzyme and 8.3% by each of the others.

In a real situation, we may wish to know which is the limiting enzyme in a sequence but find it inconvenient or impossible to estimate the response of \bar{F} to a modulation in v_i for several enzymes. In these circumstances it would be useful if some other ‘measure’ were diagnostic of rate control. Such a ‘diagnostic’ might be related to the pattern of the pools. Thus for the previous example, the pool levels are shown in the following diagram if X_1 is taken as 100 and X_2 as zero.



In this case the main drop is indeed across the 2nd enzyme and in fact it is true that the difference between pools on either side of an enzyme is a measure of its relative rate controlling effect. Will this be so for arbitrary $X_1 X_2$ and when $K_1 K_2$ etc are not unity?

From (1.19)

$$C_{v_i}^F \propto \frac{1}{K_1 K_2 \dots v_i}$$

and also the set of equation (1.16) can be written as

$$\frac{1}{v_1} = \frac{X_1}{\bar{F}} \cdot \left(\frac{X_1}{X_1} - \frac{S_1}{K_1 X_1} \right) = \frac{X_1}{\bar{F}} Q_1$$

$$\frac{1}{K_1 v_2} = \frac{X_1}{\bar{F}} \left(\frac{S_1}{K_1 X_1} - \frac{S_2}{K_1 K_2 X_1} \right) = \frac{X_1}{\bar{F}} Q_2$$

so that the factors $Q_1, Q_2 \dots$ are proportional to the coefficients of the respective enzymes. The meaning of these factors is apparent when it is noted that $K_1 X_1, K_1 K_2 X_1, \dots$ are the values $, S_1^*, S_2^*$, which S_1, S_2 would take if the flux in the pathway were blocked and equilibrium were established along the chain. The factors Q_1, Q_2, \dots etc are then just

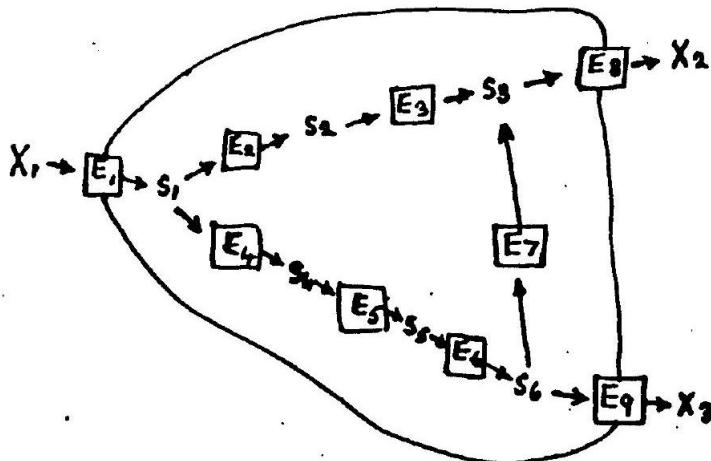
$$\left(1 - \frac{\bar{S}_1}{S_1^*} \right), \left(\frac{\bar{S}_1}{S_1^*} - \frac{\bar{S}_2}{S_2^*} \right), \dots \text{etc.}$$

In principle then an experiment could be done which first observed the levels of \bar{S}_1, \bar{S}_2 etc. then observed the values S_1^*, S_2^* , which resulted from blocking the pathway. Scaling the \bar{S}_i by their own equilibrium values, S_i^* , would result in a set of numbers, the difference between successive pairs yielding $Q_1, Q_2 \dots$. One advantage of such a diagnostic would be that it only involves ratios of concentrations of single pools and hence does not depend on the metabolites being uniformly extractable.

7. The network

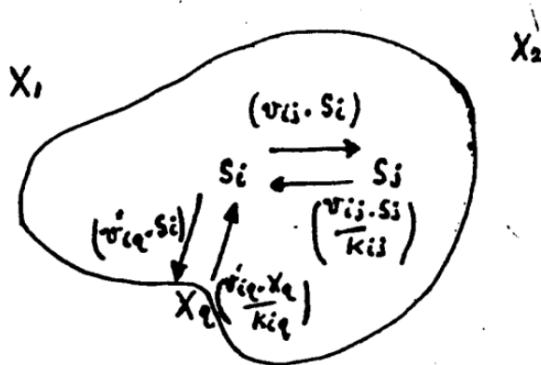
The system just dealt with at some length had a number of relatively simple properties. Thus it was completely stable, possessed only one stationary state, the sum of rate control over the enzymes was unity and finally the pool pattern gave a simple 'diagnostic' of which step was exerting control. In addition there was a simple expression for the pathway flux as a function of the parameters. It will be of interest to see to what extent these properties are retained as more realistic features are introduced into unimolecular systems. Thus we can consider a more general network of such linear unimolecular reactions, or consider the case where saturation effects are incorporated into the rate expressions.

The general network of linear unimolecular reactions allows for more than two external metabolites and also for any amount of branching and convergence of pathways. A typical network containing nine enzymes is shown below and exemplifies the possible features.



Of particular interest is that it is now possible to have a cycle of reactions in the sense that molecules leaving S_1 , along one path may return on another.

The general network is allowed for in the scheme below which shows the exchange between any two arbitrary internal metabolites S_i, S_j and also between any internal metabolite S_1 and external X_q . The net flux, used previously, is here considered as being made up of the +ve and -ve parts of the linear rate expression (1.13a).



The constant $K_{i,j}$ is the equilibrium constant for the reaction $S_i \rightleftharpoons S_j$ and is defined as $\left(\frac{S_j}{S_i}\right)$ eq. If all the reactions are truly unimolecular then, when more than one route exists between two metabolites, the product of the K_E s must be the same for both routes (Levis, 1925). For example in the nine enzyme network above $K_{1,2} \cdot K_{2,3} = K_{1,4} \cdot K_{4,5} \cdot K_{5,6} \cdot K_{6,3}$. The dynamical equations for the general network are written down in a similar way to those for the straight chain and are

$$\varphi_i = \dot{S}_i = - \sum_{\text{all } j \neq i} v_{ij} \left(S_i - \frac{S_j}{K_{i,j}} \right) - \sum_q v'_{iq} \left(S_i - \frac{X_q}{K'_{iq}} \right) \quad (1.20)$$

Since the network is truly unimolecular there will be a unique product of equilibrium constants between X_1 (say) and any other metabolite, S_i or X_q . Let these products be denoted by η_i and η'_q .

8. The electrical analogue

It will now be shown that if each S_i and X_q is scaled by its appropriate η the equations (1.20) will assume a particularly simple form from which certain observations follow. These scaled variables are $\theta_i = \frac{S_i}{\eta_i}$ and $\theta'_q = \frac{X_q}{\eta'_q}$ and the equations (1.20) may be grouped as follows:

$$\eta_i \times \left(\frac{\dot{S}_i}{\eta_i} \right) = - \sum v_{ij} \cdot \eta_i \left(\frac{S_i}{\eta_i} - \frac{S_j}{K_{i,j} \eta_i} \right) - \sum \dots$$

then noting that, by definition, $X_{i,j} \eta_i = \eta_j$ we have

$$\eta_i \dot{\theta}_i = - \sum \eta_i v_{ij} (\theta_i - \theta_j) - \sum \eta'_i v'_{iq} (\theta_i - \theta'_q)$$

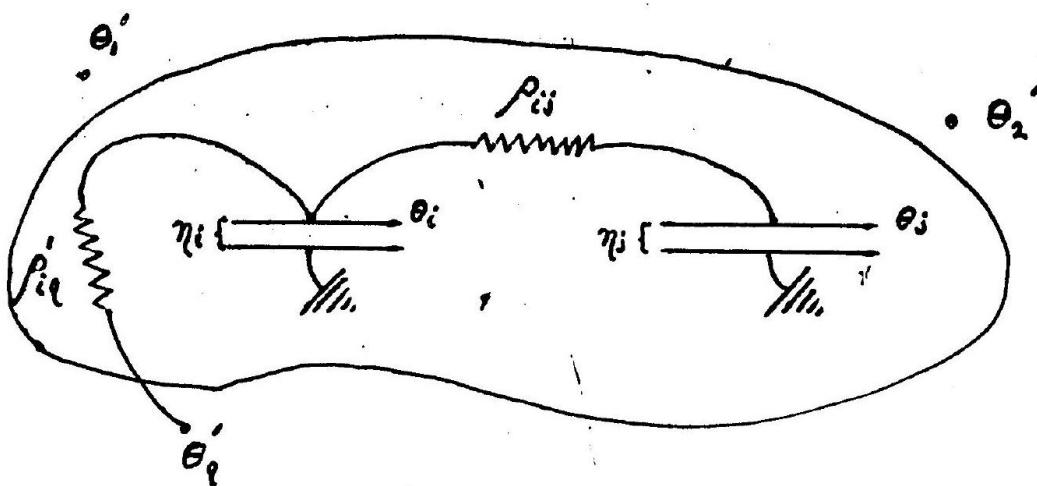
finally, introducing $\rho_{ij} = \eta_i v_{ij} = \eta_i K_{ij} \cdot \frac{v_{ij}}{K_{ij}} = \eta_j \cdot v_{ji} = \rho_{ji}$

$$\rho_{ij} = \eta_i v_{ij} = \eta_i K_{ij} \cdot \frac{v_{ij}}{K_{ij}} = \eta_j \cdot v_{ji} = \rho_{ji}$$

The equations (1.20) assume the form

$$\eta_i \dot{\theta}_i = - \sum_{j \neq i} \rho_{ij} (\theta_i - \theta_j) - \sum_q \rho'_{iq} (\theta_i - \theta_q^i) \quad (1.21)$$

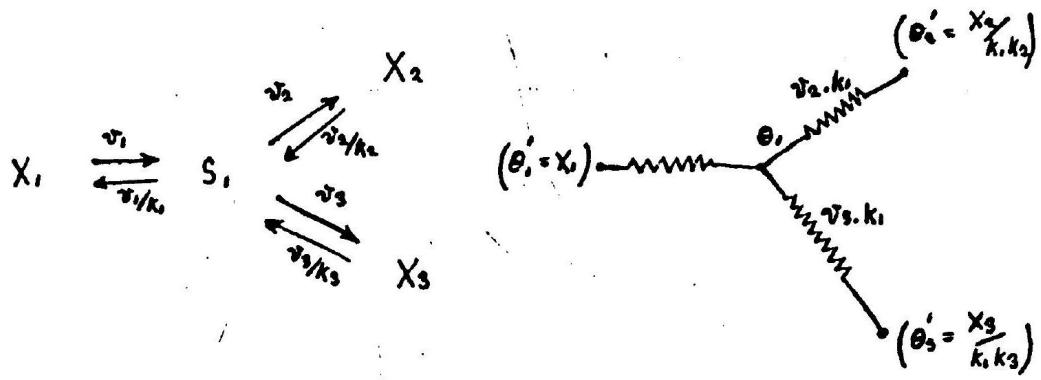
A set of interconnected earthed capacitors will be described by exactly the same set of equations when some of the capacitor is are connected via conductances ρ'_{iq} to fixed voltages, θ_q' , the i th capacitor having capacitance η_i and the conductance between the i th and j th capacitor being ρ_{ij} . Thus the general enzyme network shown previously is, in this sense, exactly equivalent to the general electric network following.



The rule of equivalence is that at the i th node of the chemical network should be placed a capacitor whose value, η_i , is the product of the equilibrium constants between X_1 and the node, and that each enzyme having activity v_{ij} radiating from this node should be placed by a conductance of value $\eta_i v_{ij}$. Then the fluxes in the enzyme network will be numerically identical with the currents in the electrical system, and the voltages θ_i at the nodes will be related to the metabolic pools by $S_i = \eta_i \cdot \theta_i$. There is no confusion about the values to be given conductances, according to which node they are calculated from, since $\eta_i v_{ij} = \eta_j v_{ji}$. When interest is centred on the stationary state the values assigned to the capacitors are unimportant and the equivalent system is merely the network of conductances.

This equivalence principle makes clear why the straight chain of reactions had simple properties, since no matter how many conductances are placed in series they still behave as one conductance which can be found by a simple rule for compounding them. For more complicated networks the stationary fluxes can easily be found by applying the known Kirchoff laws.

As an example the general branched pathway will be considered. To start with consider a system with only three reactions and where X_1 is converted via a branching system into X_2 and X_3 . The two equivalent systems are shown below



The fixed voltages θ' are found from the fixed chemical concentration by the appropriate scaling. The sum of the current leaving the branch point must be zero, so that

$$(\theta_1 - X_1) + \left(\theta_1 - \frac{X_2}{K_1 K_2} \right) \cdot v_2 K_1 + \left(\theta_1 - \frac{X_3}{K_1 K_3} \right) \cdot v_3 \cdot K_1 = 0$$

$$\text{or } \theta_1 = \frac{X_1 v_1 + X_2 v_2 / K_2 + X_3 v_3 / K_3}{v_1 + v_2 K_1 + v_3 K_1}$$

hence, for example, the flux in the 1st enzyme being numerically identical to the electrical current is

$$\bar{F}_1 = v_1 (X_1 - \theta_1) = v_1 \left(\frac{X_1 (v_2 K_1 + v_3 K_1) - X_2 v_2 / K_2 - X_3 v_3 / K_3}{v_1 + K_1 v_2 + K_1 v_3} \right)$$

If each pathway had contained several enzymes it would only have been necessary to replace the conductances $v_1, K_1 v_2, K_1 v_3$ by the overall conductance possessed by, each pathway. These are found by the rule for summing a sequence of conductors which is that for example the four conductances where



behave as one conductance C where:

$$1/C = 1/C_1 + 1/C_2 + 1/C_3 + 1/C_4$$

When applied to a straight chain this immediately yields an equivalent single conductance of value

$$C = 1 / \frac{1}{v_1} + \frac{1}{K_1 v_2} + \frac{1}{K_1 K_2 v_3} + \dots$$

an expression which was involved in the flux through a straight chain studied earlier.

9. Stability

The property of dynamical stability is still possessed by the general network of reactions just discussed and this can be proved in a very general way. Thus in the equivalent electrical network the rate of dissipation of energy in a single conductance is current x voltage or $\rho_{ij} (\theta_i - \theta_j) \times (\theta_i - \theta_j)$ and the total rate of dissipation, L , in the network is given by

$$L = \sum \rho_{ij} (\theta_i - \theta_j)^2 + \sum \rho'_{iq} (\theta_i - \theta'_q)^2 \quad (1.22)$$

where the summation is over all conductors. It can easily be verified that

$$\frac{\partial L}{\partial \theta_1} = 2 \sum_{j \neq i} \rho_{ij} (\theta_i - \theta_j) + 2 \sum \dot{\rho}'_{iq} (\theta_i - \theta'_q)$$

or using (1.21)

$$\frac{\partial L}{\partial \theta_i} = -2\eta_i \dot{\theta}_i \quad (1.23)$$

Now when the system is in motion

$$\dot{L} = \sum \left(\frac{\partial L}{\partial \theta_i} \right) \cdot \dot{\theta}_i$$

or from (1.23)

$$\dot{L} = -2 \sum \eta_i (\dot{\theta}_i)^2 \quad (1.24)$$

Inspection of (1.22) and (1.24) shows that throughout any motion $L \geq 0$ and $\dot{L} < 0$. Which implies that the dynamical system, when, released from any initial condition, can only move towards a stable stationary state. The results just established, equivalence to an electrical network and complete stability, apply to networks of linear unimolecular reactions and depend on the property that the product of equilibrium constants along different paths connecting the same metabolites is identical. This condition will also be satisfied in many networks of interest, which include ‘pseudo unimolecular’ reactions provided these do not appear within any ‘cycle’ of reactions.

By a ‘pseudo unimolecular’ reaction is meant, for example, one of the form $A + B \rightleftharpoons C + D$ but where B and D can be considered as parameters B' and D' , that is as constant concentrations external to the dynamical system. In this case the true thermodynamic eq. constant $K_E = \left(\frac{C \times D'}{A \times B'} \right)$ but since B' , and D' are constant these will be an effective equilibrium constant K'_E for the pseudo-unimolecular reaction $A \rightleftharpoons C$. This constant is given by $K'_E = K_E \times \frac{B'}{D'}$ and clearly is subject to no thermodynamic constraint. $\frac{B'}{D'}$ being set at arbitrary level.

The general pseudo-unimolecular network having linear rate expressions has no equivalent electrical network and correspondingly does not have the same simple solution and stability properties. However, even for this case there is a stability property which might be said to depend on the ‘conservation of molecules’, found in all unimolecular systems. Thus in the equations (1.20), for the general system, if the molar concentrations are measured as differences S'_i from the stationary values \bar{S}_i the equation of motion will be

$$\begin{bmatrix} \dot{S}'_1 \\ \dot{S}'_2 \\ \vdots \end{bmatrix} = [D] \begin{bmatrix} S'_1 \\ S'_2 \\ \vdots \end{bmatrix} \quad (1.25)$$

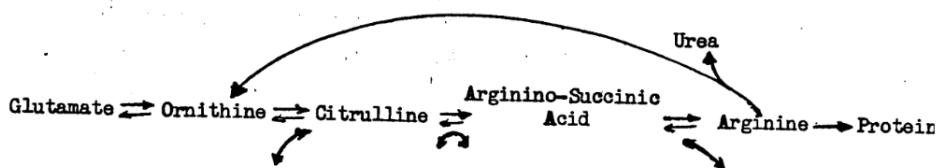
where, for this linear system, the dynamical matrix D has constant elements. A simple argument shows that these elements satisfy certain conditions. Suppose that all $S'_i = 0$ except one, say S'_3 , which is +ve then clearly, from the metabolic diagram and a consideration of the rate expression, \dot{S}'_3 must be -ve and all other \dot{S}'_i , re or zero, depending on whether they receive molecules from S_3 or not. Furthermore, since all molecules leaving S_3 must arrive at other intermediate metabolites or leave the system, the sum of \dot{S}'_3 and all other \dot{S}'_i must be ≤ 0 . Hence the matrix D will have the form

$$[D] \equiv \begin{bmatrix} - & - & + \\ + & - & - \\ - & - & - \end{bmatrix} \quad \dots(1.26)$$

That is, -ve diagonal and all other elements +ve or zero. In addition D will satisfy the condition that in each column the sum of the elements is less than or equal to zero. These are sufficient conditions (HEARON, 1963) to ensure the stability of the system in the sense that although complex eigenvalues may occur their real parts will be -ve. This means that such systems may have a tendency to oscillate but that these oscillations will always be decaying.

In the case of networks not containing pseudo-unimolecular reactions within loops, it is always possible by suitably scaling the dynamical variables to symmetrise the matrix D , since this corresponds to the condition $\rho_{ij} = \rho_{ji}$ in (1.21). This further property of symmetrisability is sufficient to ensure that all eigenvalues of D are real (Hearon, 1963), so that such systems move towards their stationary solutions as a sum of decaying exponential terms. The straight chain mentioned earlier, whether it is truly unimolecular or not, is an example of this.

An example of a network, in which we are interested, containing potentially oscillatory cycles is the arginine cycle, indicated below.

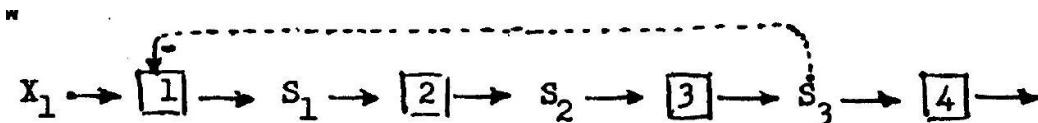


Here several of the reactions within the cycle involve other metabolites. Even if these are treated as constant, so that the system is unimolecular, the fact that the product of equilibrium constants around different sides of the loop is not the same cannot be ignored. However, the general property proved above indicates that when the arginine system is approximated by linear rate-expression it will have just one stable solution, possibly arrived at via decaying oscillations.

The situation when the rate expressions for reactions in a unimolecular network are non-linear is of interest. In this case there will still be a set of equations similar to (1.25) for movement about the stationary state, but they will only apply for small disturbances from the \bar{S}_i and the elements of D will become the $\left[\frac{\partial \varphi_i}{\partial S_j} \right]_{S_j = \bar{S}_j}$, discussed earlier.

However, the argument that (D) had the general form (1.26) and that its columns satisfied the required inequalities assumed only ‘conservation of molecules’, which of course is still true, and that the rate expressions were restricted to have certain properties. The required properties are that they should involve only their own substrates and products and that when the level of a metabolite is increased its net rate of removal by a given enzyme should respond positively in the algebraic sense. Such properties are fairly general and will certainly be possessed by the simple rate-expression of (1.13b) allowing for saturation. Thus it is still possible to state, subject to this restriction, that the general network of non-linear unimolecular reactions will be stable, with decaying oscillations, in the neighbourhood of a stationary solution. For example if the arginine cycle were represented by nonlinear rate-expressions it would still be stable. However for true unimolecular networks with cycles the introduction of a saturable rate expression within a cycle could alter the type of stable dynamical behaviour, since the matrix [D] will no longer be symmetrisable.

With the introduction of non-linear rate expressions it becomes possible to represent ‘end product inhibition’ and this is indicated below.



The rate expression for the 1st reaction, in this case, involves S_3 . Under these circumstances the system may become unstable since there are now off-diagonal -ve elements in [D]. Increasing the ‘strength’ of the feedback corresponds to increasing the magnitude of the off-diagonal element and at some point the previously stable system will become unstable.

However, because of the inherent stability of the initial system, a considerable feedback signal can be tolerated without instability. The stability of such systems has been treated subject to a number of simplifying assumptions (Walter, 1969).

10. Nonlinear systems

Turning now to the problem of locating the stationary state for a non linear unimolecular system it is clear that there will be no simple algebraic method since we now have a set of simultaneous non linear equations. The simplest case to consider is that of a straight chain of reactions where each enzyme has a rate-expression of the form (1.13b). This situation corresponds to the sequence of linear reactions discussed on p. 16 except that the fluxes per unit volume of the system become

$$F_i = \frac{v_i}{M_i} \left(S_{i-1} - \frac{S_i}{K_i} \right) / \left(1 + \frac{S_{i-1}}{M_i} + \frac{S_i}{M_i} \right) \quad (1.27)$$

Where v_i is now the maximum velocity of the i th enzyme in the direction $S_{i-1} \rightarrow S_i$ and for unit volume of the system v_i will be proportional to the concentration E_i of the enzyme. The constants M_i and M_i are the appropriate Michaelis constants and K_i is the 'effective' equilibrium constant for the step. Once again there are $n + 1$ linear equations in $S_1, S_2 \dots S_n$ expressing the fact that at the stationary state the $n + 1$ separate fluxes are all equal to the pathway flux \bar{F} . These are obtained by multiplying up the denominator in expressions of the form (1.27) and yield.

$$\frac{\bar{F}}{v_1} + \left(\frac{\bar{F}}{v_1 M_1} - \frac{1}{M_1} \right) \cdot S_1 + \left(\frac{\bar{F}}{v_1 M_1} + \frac{1}{K_1 M_1} \right) \cdot S_2 = 0$$

$$\frac{\bar{F}}{v_2} + \left(\frac{\bar{F}}{v_2 M_2} - \frac{1}{M_2} \right) \cdot S_2 + \left(\frac{\bar{F}}{v_2 M_2} + \frac{1}{K_2 M_2} \right) \cdot S_3 = 0 \quad \dots (1.28).$$

$$\begin{array}{c} \frac{\bar{F}}{v_3} \\ \vdots \\ \vdots \\ \vdots \end{array} = 0$$

As before there are $(n + 1)$ equations to determine only n unknowns that the determinant of the coefficients must be zero or,

$$\begin{array}{c}
 \frac{\bar{F}}{v_1} + \left(\frac{\bar{F}}{v_1 n_1} - \frac{1}{n_1} \right), \quad \frac{\bar{F}}{v_1 n_1} + \frac{1}{k_1 M_1}, \\
 \dots \\
 \frac{\bar{F}}{v_2} + \left(\frac{\bar{F}}{v_2 n_2} - \frac{1}{M_2} \right), \quad \frac{\bar{F}}{v_2 n_2} + \frac{1}{k_2 M_2}, \\
 \dots \\
 \frac{\bar{F}}{v_3} + \left(\frac{\bar{F}}{v_3 n_3} - \frac{1}{M_3} \right), \quad \frac{\bar{F}}{v_3 n_3} + \frac{1}{k_3 M_3}, \\
 \vdots \\
 \dots \\
 \frac{\bar{F}}{v_{2+1}} + \left(\frac{\bar{F}}{v_{2+1} n_{2+1}} - \frac{1}{M_{2+1}} \right), \quad \frac{\bar{F}}{v_{2+1} n_{2+1}} + \frac{1}{k_{2+1} M_{2+1}}
 \end{array}
 = 0$$

The pathway flux, \bar{F} , must satisfy the determinantal equation, but \bar{F} now occurs in all elements of the determinant whereas formerly, in the linear case, it was present only in the first column. Consequently the equation of which \bar{F} is a solution is now a polynomial in \bar{F} of order $n + 1$. Thus for two enzymes \bar{F} is one of the roots of a quadratic equation and for 10 enzymes it is one of the 10 roots of the 10th order equation. In fact it turns out that in this system there is never more than one \bar{F} satisfying the condition that all S_i should be real and +ve. One consequence of \bar{F} being the solution of a high order equation is that there will no longer be an explicit formula, corresponding to (1.18) of the linear case, giving \bar{F} in terms of the parameters. Nor is it any longer the case that the overall action of a sequence of enzymes is equivalent to one enzyme. It appears then that when non-linear rate expressions are introduced, even in the simple case of a straight chain, there is no alternative but to study possible to investigate the 'non-linear straight chain' problem algebraically with respect to the question of 'diagnosing' from the pool pattern which steps in the chain are rate controlling. This question will be returned to in CH II, at the moment it will be show that even though there is no explicit expression for the flux it is still true that $\sum C_i^{\bar{F}} = 1$. This can easily be seen by noting that in the functional relationship existing between \bar{F} and v_i .

$$\bar{F} = \bar{F}(v_1, v_2, \dots, v_{2+1})$$

Thus simultaneous changes in the v_i will cause a small change in \bar{F} according to

$$\delta \bar{F} = \sum \frac{\partial \bar{F}}{\partial v_i} \cdot \delta v_i \quad (1.30)$$

Now inspection of the determinant (1.28) shows that if all v_i are increased by the same factor $(1 + \alpha)$ then $\bar{F}(1 + \alpha)$ will be the new solution. So that changes $\alpha v_1, \alpha v_2, \dots$ produce a change $\alpha \bar{F}$ in the flux. When α is small this can be substituted in (1.30) and

$$\alpha \bar{F} = \sum_i \frac{\partial \bar{F}}{\partial v_i} \cdot \alpha v_i$$

or

$$1 = \sum \frac{v_i}{\bar{F}} \cdot \frac{\partial \bar{F}}{\partial v_i} = \sum C_{v_i}^{\bar{F}} \quad (1.31)$$

Thus it appears that the summation property of rate control coefficients, originally proved for the linear case, may be a fairly general one.

2

Mathematical Formulation of Complex Systems

In the last chapter, after a general theoretical section, unimolecular systems were considered in some detail. It was shown that, even for such unimolecular systems, explicit solutions for the stationary state could not be obtained, except for the hypothetical ‘completely linear’ case. However, it was seen that the concepts of ‘rate-control’ and ‘diagnostic’ could give a useful insight into the workings of systems, even when there was no explicit solution. For the more complex systems, whose representation is to be considered in this chapter, the necessity of understanding a system by considering its response to parametric variation is even greater, since very little in the way of analysis can be achieved.

When the variation is differential the response of the system will be expressed naturally by a ‘sensitivity coefficient matrix’ and in CH. III the theoretical basis of this will be developed with particular reference to how it is related to the properties of individual enzymes and how it can be estimated experimentally.

The response to finite parameter variation can only be studied by means of numerical experiments performed on mathematical models. In CH. IV a computer system capable of flexibly performing such ‘in numero’ experiments on a variety of general systems is discussed.

The contents of this chapter will, then, be concerned with the problem of formulating complex metabolic systems. in the mathematical terms suitable to computer analysis. After discussing the distribution between a ‘subsystem’ and a ‘complete’ system and initially restricting consideration to subsystems, the network of genial reactions is introduced.

This lifts the previous restriction to unimolecular reactions and in addition allows the representation, via suitably complex rate expressions, of ‘allosteric’ enzymes, where substances can influence the rates of reactions from which they are metabolically remote. Next it will be shown that, for systems where it is wished to consider the effects of substituting alternative alleles for structural genes, a new type of genotypic parameter can be introduced by extending the idea of a ‘rate expression’. The section ends by considering how the effects of ‘growth’ and ‘enzyme regulation’ may be represented for subsystems.

Finally an approach is made to the problem of formulating ‘complete systems’, which gives consideration to the allocation of resources by differential gene activity.

1. Subsystem and complete system

An dynamical system of the type which was discussed in CH.I (1.1) involves the assumption that some metabolites, denoted by $S_1, S_2 \dots$, are within the system and are variables, whereas others $X_1, X_2 \dots$, are outside the system and have the status of parameters. The assignment of any metabolite to one of these two classes is clearly a quantitative judgement and has up till now been made on the reasonable basis that the medium had a relatively large volume and approximated to an infinite source or sink, thus rendering the concentration environmental metabolites almost constant. However the assignment is clearly arbitrary and cannot de applied in certain situations, where it becomes necessary to widen the system so as to include at least some ‘environmental metabolites as variables, for example if it is wished to consider ‘growth in a chemostat.

Even when the environmental metabolic boundary can be accepted a difficult problem arises because of the necessity of reducing the large number, many thousands, of metabolites which are involved in a ‘real’ system. One way in which this reduction can be performed, for our stationary state studies, is by eliminating system variables which are not directly of interest, a technique which has already been used to eliminate unwanted enzymic intermediates. However, this was possible only because, on the assumption that protein-protein interaction was negligible, the necessary eliminations involved only linear equations in the enzymic intermediates and were already fully treated under the subject of steady state enzyme kinetics. The equivalent technique at the level of intermediary metabolites is not in general possible because non-linear equations are involved. For example it might involve eliminating metabolites within a sequence of enzymes and replacing the sequence by a single ‘enzyme’ with a complex rate expression, a task which, as was seen in CH I, has no explicit algebraic solution. Hence, in general, no such exact reduction can be made. It may, however, prove possible to find closed algebraic expressions which are a reasonably good description of pathway behaviour, provided that the remaining system variables lie within certain ranges.

In these circumstances it is necessary to view the problem of reduction in a flexible way; ‘what is a reasonable reduction in one situation may be quite unreasonable in another. One

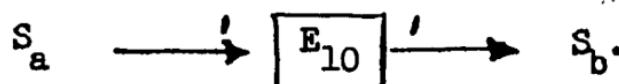
extreme is the situation when the parameter changes being studied are likely to have large effects throughout the whole metabolic system, even though the varied parameters may themselves all be within one pathway. Changes which, for example, markedly affected growth rate would seem to be of this type. In this case there is no escape from considering some representation of the ‘complete’ system. The representation being, perhaps, detailed only in the region of the parameter changes, it being sufficient elsewhere to represent the functioning of whole areas of metabolism, as for example the glycolytic pathway, in some approximate way. Some consideration will be given to the rather difficult problem of such complete systems at the end of this chapter.

Another way in which the desired reduction can be made is to consider the area of metabolism in which parameter variation is to be made as a ‘sub-system’. This involves arguing that internal metabolites which are ‘remote’ from the sub-system will remain more or less constant and can be formally treated as ‘external’ metabolites.

The criteria for deciding what constitutes such a sub-system are not simple. The correct decision is clearly a quantitative matter and will depend, *inter alia*, on the size of the parameter variations. In general, the larger the variation the larger will have to be the sub-system. Despite the difficulty in defining exactly what is meant by a subsystem, in this sense, there is no alternative but to use the concept.

2. Network of general enzymically catalysed transformations

Until now we have considered systems of effectively unimolecular reactions, where the transformations, taking place within a fixed volume, are represented in a system diagram by steps of the type:

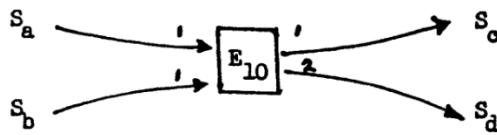


This means that the 10th enzyme catalyses a unimolecular conversion from S_a to S_b in the sense that for every molecule of S_a removed one of S_b is produced. The net-flux/unit quantity of enzyme, taken as +ve in the arrowed direction, was given by a rate-expression which assumes enzymic intermediates to be at a stationary state with respect to S_a and S_b and which takes fully into account that the reaction $S_a \rightarrow S_b$ can be reversible.

This can easily be generalised to allow for enzymically catalysed reactions of arbitrary complexity. For example the reaction



is symbolised by

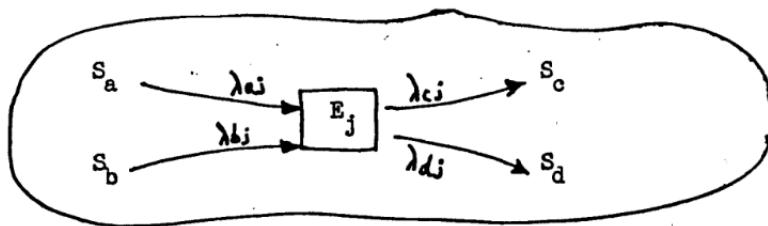


and has a rate-expression/unit quantity of enzyme giving the net flux towards, say, the metabolite S_c of f_{10} ($S_a, S_b, S_c, S_d, \dots$). Clearly the corresponding rate expressions for the metabolites S_a, S_b, S_d are $-f_{10}, -f_{10}$, and $+2f_{10}$ respectively where the appropriate stoichiometric constants $-1, -1, 2$ are obtained from the equation (2.1) for the net reaction. The symbol f_{10} stands here for either the function or its value, depending on the context.

In a general network of reactions, if f_j is the rate expression for unit quantity of the j th enzyme referred to one of its products, then $\lambda_{ij} f_j$ will be the appropriate rate expression of the j th enzyme for the i th metabolite.

The ‘stoichiometric matrix’ $[\lambda_{ij}]$ will consist of small integers, mostly 1 and -1 , and zeros and will in fact contain mostly zeros, since most enzymes remove or produce only a few metabolites. In a system having n metabolites and m enzymes $[\lambda_{ij}]$ will be an $n \times m$ matrix.

The figure below represents a general system existing in a given volume V and with total quantity, E_j^T of the j th enzyme. This can be distributed in any way throughout the volume or, if a transport enzyme, upon the surface.



$$\text{Then: } \frac{\text{(rate of change of total quantity of } S_i \text{ in volume)}}{\text{i}} = \frac{\text{(Sum of contributions due to separate enzyme to the } i\text{th metabolite)}}{\text{}}$$

$$\frac{d(S_i \times V)}{dt} = \sum_j \lambda_{ij} E_j^T f_j(\dots) \quad (2.2)$$

or

$$\dot{S}_i = \sum \lambda_{ij} \left(\frac{E_i^T}{V} \right) \cdot f_j - \left(\frac{1}{V} \dot{V} \right) \cdot \dot{S}_i$$

Writing $\frac{E_j^T}{V} = E_j$ the effective concentration of the j th enzyme in the system and remembering that $(1/V \dot{V}) = 0$ we have the dynamical functions φ_i for the general system as

$$\varphi_i = \dot{S}_i = \sum \lambda_{ij} E_j f_j(\dots) = \sum \lambda_{ij} F_j \quad (2.3)$$

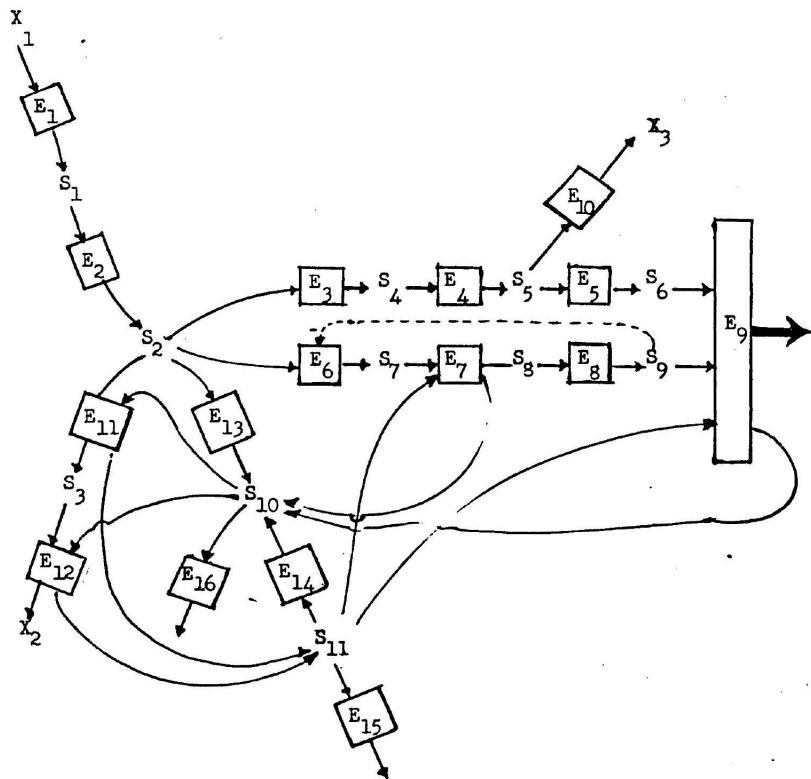


Figure 2.1 METABOLIC STRUCTURE

E_j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
S_i	1	1 -1														
1		1 -1														
2			1 -1		-1						-1	-1				
3				1 -1								1 -1				
4					1 -1											
5						1 -1					-1					
6							1				-1					
7								1 -1								
8									1 -1							
9										1 -1						
10										1	1	-1 -1	1	1	-1	-1
11										-1	-1	1 1	-1	-1	-1	-1

Figure 2.2 STOICHIOMETRIC MATRIX

In (2.3) F_j is the net flux carried by the j th enzyme in unit volume of the system, the flux being defined by reference to one of the substrates of the j th enzyme.

The general equation (2.3) will be made more concrete by means of an example, shown in fig (1), which will illustrate how various metabolic situations can be represented. This example will also be used to discuss certain conditions to which the equations (2.3) must conform.

The metabolic network of fig (1) comprises 16 enzymes, 10 intermediate metabolites and four external substances and the general meaning of the metabolic scheme is as follows. X_1 , which can be regarded as a nutrient, is transported in and converted to S_2 at which point the metabolism divides into several pathways. A catabolic pathway, $S_2 \rightarrow S_3 \rightarrow X_2$, yields a high energy product S_{11} by bimolecular coupling reactions at steps 11 and 12. Two biosynthetic pathways, $S_2 \rightarrow S_4 \rightarrow S_5 \rightarrow S_6$ and $S_2 \rightarrow S_7 \rightarrow S_8 \rightarrow S_9$ converge onto a condensing reaction at step 9, the second of these pathways and the condensing reaction being coupled to the high energy product S_{11}^* .

Clearly S_{10} and S_{11} have a similar role to that of ADP and ATP in a real system and the step E9 would correspond to protein synthesis. Continuous lines represent a net flux of metabolite, whilst the arrow and the number above it indicate the algebraically positive direction and stoichiometric constant respectively. Dotted lines indicate metabolites influencing the rate of a reaction, without themselves being appreciably consumed, the sign above the arrow indicating, when appropriate, whether there is an inhibitory or excitatory effect. Thus in the figure it can be seen that S_9 exerts feedback inhibition on E6 the last enzyme in its own pathway, or again the external substance X_4 activates the last enzyme in the same pathway. All these effects are taken fully into account by multiplying the stoichiometric matrix λ_{ij} , shown below the figure, into the set of rate expressions, f_1, \dots, f_{16} to produce the 10 functions $\varphi_1, \dots, \varphi_{10}$. The dynamical functions can then be used, as explained earlier, to obtain the stationary solution and the sensitivity of any measures to any parameters which are of interest. Thus the problem of describing the system mathematically is just that of writing down the set of rate expressions, f . How exactly this is done will depend on what qualities are considered to be possessed by the separate enzymes. For instance the unimolecular expressions f_1, f_2 may simply be of the usual linear form (1.13b) unless it is suspected that their saturation properties are important. Other expressions clearly have to be more complex just to represent the effects encoded in the metabolic diagram. For example, the simplest possible rate-expression for the bimolecular reaction of E11 would be

$$f_{11}(\dots) = k_x (S_2 \times S_{10} - S_3 \times S_{11}/K_{10})$$

and this, of course, takes no account of saturation effects. Again there are various possibilities for representing E6, which is sensitive to feedback inhibition from S_9 , one of the simplest being

$$f_6(\dots) = k_{1x} \left(S_2 - \frac{S_7}{K_6} \right) \times \frac{1}{(k_2 + S_9)}$$

The situation then is that metabolic ‘structure’ of considerable detail can be included in

the functions $\varphi_1, \dots, \varphi_{10}$ by means of the matrix $[\lambda_{ij}]$ and the set of rate-expressions. In general these rate expressions should be of the simplest form which will express the desired properties of the separate enzymes, However, the general set of equation (2.3) must satisfy a certain condition as will be discussed in the next section.

3. Necessity for linear independence of the functions

This arises because if it happens that one of the φ_i is simply a linear combination of others there will not be sufficient simultaneous equations to determine the stationary solution. Since the rate expressions are clearly all independent functions this condition of linear independence is just that the rows of $[\lambda_{ij}]$ be linearly independent.

An example will show how faulty formulation of a set of equations may generate dependence. In formulating a system involving, say, a set of ADP/ATP dependent reactions, it might be thought sufficient to link the generating system to the absorbing system by a closed loop. In terms of our model fig (1) this is equivalent to omitting the steps E13 (generation ADP), E15 (degrading ATP) E16 (degrading ADP). This might appear to satisfy the linked steps E11 and E12.

The rates of change of the two substrates, S_{10} and S_{11} would then be:

$$\begin{aligned}\dot{S}_{10} &= \varphi_{10} = -F_{11} - F_{12} + F_7 + F_9 + F_{14} \\ \dot{S}_{11} &= \varphi_{11} = +F_{11} + F_{12} - F_7 - F_9 - F_{14}\end{aligned}$$

For the stationary state conditions we find that $\varphi_{10} = 0$, implies that φ_{11} is determined also to be zero and imposes no separate condition. This eliminates one of the n necessary equations for solution. Dynamically the system would have the property $\frac{d(S_{10}+S_{11})}{dt} = 0$ so that the sum of ATP and ADP could not change and would be dependent on the initial conditions (viz, the initial value of the sum). In addition the matrix $D = \left[\frac{\partial \varphi_i}{\partial S_j} \right]$ would have no inverse thus invalidating the iterative method of (1.9). A biochemical system formulated in this way would thus not have the property that the stationary state was dependent only on genetic and environmental parameters, and the whole theoretical approach outlined earlier would be invalid.

It seems reasonable to argue that this difficulty arises out of the unreality of formulating a system in which production of ATP is always coupled to removal of ADP and vice versa. The solution of the difficulty then is to extend the system by adding the reactions E13, E15, E16, which allow S_{10} and S_{11} to participate in non-coupled reactions and thus renders φ_{10} and φ_{11} linearly independent.

It may be objected that the investigator wishes to consider a situation where the sum of ATP and ADP is at a previously prescribed level. This is of course still possible with the extended model. The difference being that instead of the sum being determined by initial conditions it must now be determined by parameter values acting within the system. For

example, suppose the reaction E13 is an irreversible reaction saturated by S_2 and providing a flux/ unit volume determined by a parameter k_1 and that E16 and E15 are linear and irreversible with the same effective rate constants k_2 . Then at S.S.,

$$\varphi_{10} = -F_{11} - F_{12} + F_7 + F_9 + F_{14} + k_1 - (k_2 \times \bar{S}_{10}) = 0$$

and $\varphi_{11} = -F_{11} - F_{12} + F_7 + F_9 + F_{14} + k_1 - (k_2 \times \bar{S}_{11}) = 0$ so that

$$(\bar{S}_{10} + \bar{S}_{11}) = k_1/k_2$$

Therefore by adjusting the parameter k_1 the investigator can prescribe the sum of ATP and ADP.

In general then only systems, having the property that the rows of $[\lambda_{ij}]$ be linearly independent, will be considered, additional reactions being added if the condition is not at first satisfied. This will ensure the applicability of the mathematical methods outlined earlier for calculating the sensitivity coefficients whilst not restricting the investigator in any important way.

4. Independent fluxes in a system

An important ‘measure’ in a system is often the stationary flux/unit vol, \bar{F}_j , at some step in the network, there being m such measures in a system having m enzymes. However, these stationary fluxes are not independent since there are n conditions connecting them, namely

$$\varphi_i = \sum \lambda_{ij} \bar{F}_j = 0 \text{ For } i = 1 \text{ to } n \quad (2.4)$$

The number of independent fluxes will thus be the number of enzymes less the number of conditions connecting them or $(m - n)$.

For example, in a straight chain the number of enzymes will only be one more than the number of pools and hence there is just one independent or ‘pathway’ flux.

For the more complex system of fig (1) there are 16 enzymes and 11 pools so that there are 5 independent fluxes. This means that if 5 selected fluxes are known the remaining 17 can be found using the condition (2.4). This amounts to inverting part of the matrix and may appear algebraically complex. However, the connection between the fluxes can be seen immediately from a metabolic diagram. Thus, if the 5 excretory fluxes $F_0, F_{10}, F_{12}, F_{15}, F_{16}$, in fig(1) are selected as metabolically significant the other fluxes can be written down in

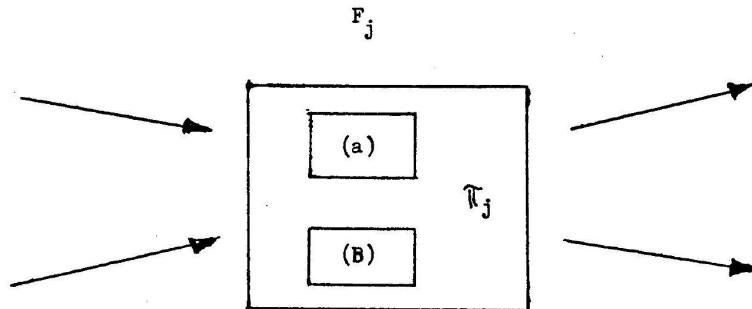
terms of them as for example:

$$\begin{aligned}
 \bar{F}_5 &= \bar{F}_9 \\
 \bar{F}_4 &= \bar{F}_9 + \bar{F}_{10} \\
 \bar{F}_6 &= \bar{F}_9 \\
 \bar{F}_{11} &= \bar{F}_{12} \\
 \bar{F}_{13} &= \bar{F}_{16} + \bar{F}_{15} \\
 \bar{F}_2 &= 2\bar{F}_9 + \bar{F}_{10} + \bar{F}_{12} + \bar{F}_{16} + \bar{F}_{15}
 \end{aligned} \tag{2.5}$$

For most metabolic systems the number of enzymes will only slightly exceed the number of metabolites so that in fact there are relatively few independent fluxes in a network.

5. Introduction of genotypic parameters

Suppose that in the 16 enzyme system of fig. (1) alternative enzymes were available at various steps in the network, then the particular set of enzymes possessed by such a 16 enzyme ‘organism’ would depend on its set of alleles or its ‘genotype’. The introduction of a set of “genotypic parameters” ($\pi_1, \pi_2, \dots, \pi_{16}$) specifying the allelic situation at each step, will serve to describe the genotype. For example in a haploid organism with two alleles at each locus the π integers must each take one of two values, whereas if the organism were diploid there would be three possible values since at each locus there are, physiologically, three alternative states, Figure (2), homozygous either way and heterozygous.



The flux depends on the allelic enzymes E_j^a and/or E_j^b . The π value specifies the allelic state of the locus.

Consideration of the phenotype/genotype, or Ph/Ge relation will clearly be facilitated if these ir parameters can be included in the specification of the functions $\varphi_1 \dots, \varphi_n$. This can conveniently be done by extending the idea of a rate expression, the new formulation depending on the ploidy and the number of alternative alleles at each locus. The case of two alternative alleles per variable locus will be considered for haploid and diploid organisms.

Thus for a haploid organism suppose that the alleles for the j th locus give alternative enzyme concentrations E_j^a and E_j^b and alternative rate expressions $f_j^a(\dots)$ and $f_j^b(\dots)$, it

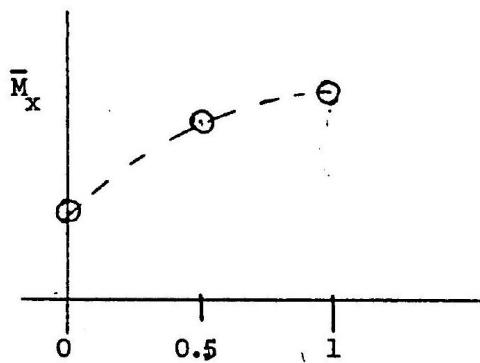


Figure 2.3 The measure \bar{M}_x as a function of the value of π .

being noted that these expressions may be of quite different form as for example one being sensitive to end product inhibition and the other not. A composite expression (cf. Fig. (2)) can be written for the flux/unit volume at the j th step thus.

$$F_j = (1 - \pi_j) E_j^a f_j^a(\dots) + \pi_j E_j^b f_j^b(\dots) \quad (2.6)$$

and clearly $\pi_j = 0$ gives the (a) allele and $\pi_j = 1$ gives the (b) allele so that 0 and 1 are the possible values for π^j

Since the description is formally the same as before, only with a more complex set of rate expressions which involve the if g the same numerical techniques will work. In particular it will be possible to consider how a measure \bar{X} depends on the set of alleles constituting the genotype since the non algebraic functional relationship

$$\bar{M}_X = \bar{M}_X(\pi_1, \pi_2, \dots, \pi_{16}) \quad (2.7)$$

can now be investigated numerically.

For a diploid organism it is only necessary to allow the h_j to take three values, 0, 1/2, 1, where $\pi_j = 1/2$ corresponds to the heterozygous state at the j th locus. This involves the assumption that enzyme quantity is proportional to gene dosage which seems reasonable so long as enzyme quantity is till treated as a genetically specified parameter. The response of a measure to π_j indicated in fig. (2) now gives an indication of the importance and the degree of dominance of the j th locus and the effect on this of varying other π_j , that is the genetic background, can be tried.

The introduction of the or notation makes clearer what is meant by the Ph/Ge relation. It also reveals a difficulty not previously raised. Thus the _Ph/Ge response matrix $[C_{P_y}^{M_X}]$ of (1.9) can now be recognised as $[C_{\pi_y}^{\bar{M}_X}]$ but since π_y can only assume discrete values the meaning and usefulness of the differential sensitivity analysis will have to be considered. The uses of the general relationship (2.7) in the framework of theoretical ! population genetics will be discussed in CH.V.

6. Non-enzymic transformations

In metabolism there will be many situations where non-enzymic transformations are important, e.g. in the case of surface transport by simple diffusion. These can be represented perfectly well using rate expressions the only difference being that instead of enzyme concentration there will be some other factor. Thus 'non-enzymic transport' rate-expressions will involve the quantity area/volume, or A/V . That is the 'concentration of area' rather than the concentration of enzyme. Enzymically mediated transport expressions will not involve A/V since the amount of surface over which the enzyme is distributed is not relevant. The inclusion of nonenzymic transformations will not alter the general form of (2.8) except that a new category of pseudo-enzymes are introduced.

7. Expanding systems

It is often of interest to consider growing systems, particularly since the condition of steady or exponential growth of a culture of such systems is a well defined experimental situation: such a growing system clearly does not possess a stationary state, since certain system parameters are no longer constant with time.(Grainger, 1966, 1968).

However, subject to the conditions that these parameters do not vary too rapidly or over too wide a range and that the system itself is intrinsically stable (see next section) the stationary state approach still proves useful.

Thus an investigator of such a non-stationary system would be interested in the mean values over time, $\langle S_i \rangle_{t_1}$ of system variables. Given the above conditions the $\langle S_i \rangle_t$ will be well approximated by the stationary \bar{S}_i corresponding to average values of varying parameters. In fact fluctuating parameters in a growling system are often quite well behaved In the above sense. For example in a spherical cell which continuously doubles its volume and then divides, perhaps representing a growing bacterium, the parameter (A/V) suffers a sharp increase of $\approx 25\%$ at each division and then gradually falls to reach its former value just prior to the next division. Thus, given reasonable metabolic stability, the system will be close to a stationary solution during most of the time and will suffer its greatest disturbance at each cell division. A similar situation holds for the parameter combinations 'relative gene dosage', whose importance will be seen in the section on control. These double at certain points in the cell cycle when each gene is duplicated and decline steadily during the rest of the cycle due to other genes being replicated. For syncytial growth the stationary state assumption may be a very good approximation, thus (A/V) will vary hardly at all during the tube like growth of hyphae and the 'relative gene dosage' will remain almost constant since in this case a number of asynchronously dividing nuclei inhabit a common metabolic space. Supposing then that it is acceptable to study the stationary state corresponding to average parameter values the effect of growth on this S.S. will still have to be allowed for. This arises because it is no longer permissible to assume $\dot{V} = 0$ when simplifying the

dynamic equations (2.2). Thus instead of the form (2.3) there arises

$$\varphi_i = \dot{S}_i = \sum \lambda_{ij} E_j f_j(\dots) - (1/V \dot{V}) \cdot S_i \quad (2.8)$$

Clearly a stationary solution with growth will require all S_i constant and $\dot{V} \neq 0$ and will only be possible if the term $(1/V \dot{V})$ is constant. This means that such growth will be exponential with growth constant G , where

$$G = \frac{1}{V} \dot{V}$$

Since, for the present, only ‘subsystems’ are being considered G has to be treated as a parameter. The equations (2.8) are then formally equivalent to a metabolic system within a fixed volume and having a “flux to expansion” of $G \times S_i$ away from each S_i . Whether these expansion terms will constitute more than a minor correction will depend on the metabolic pathways contained in the subsystem. In any particular case the G constant will, of course, be known and the problem in setting up the model is to assign levels to the E_j consistent with general data.

For example in the arginine pathway it may be known that arginine in pool form is of the order of 5% of arginine in protein in which case the E_j in arginine pathway would be adjusted until the terms $\lambda_{ij} E_j f_j(\dots)$ are approximately twenty-fold greater than $G S_{\text{arg}}$. This involves some guesswork but can usually be done satisfactorily. The expansion terms can usually be ignored in major pathways such as the glycolytic, which carry a large flux, but may be of dominant importance in pathways carrying a small flux such as co-enzyme synthesis. In addition any mutants tending to block the flux in a pathway and at the same time raise the pool levels will enhance the importance of the expansion terms.

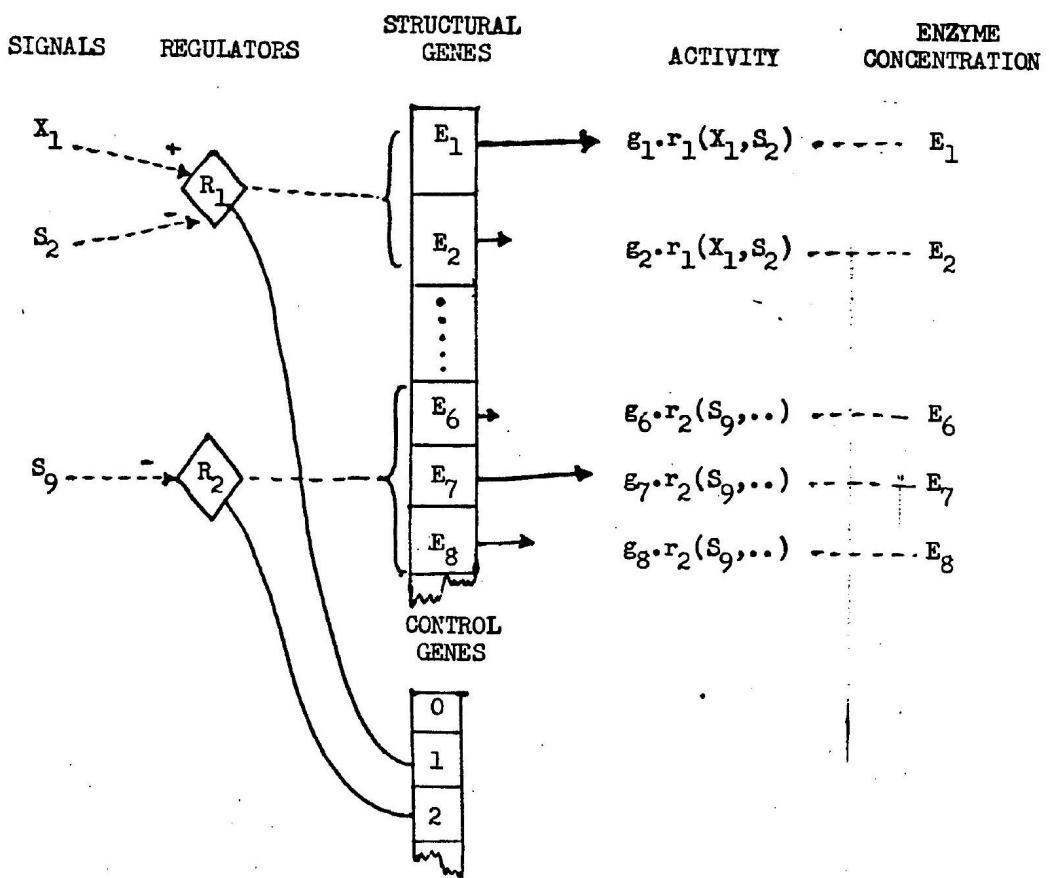
8. Stability of networks of general reactions

The unimolecular systems studies previous were fairly generally γ stable. The more general reactions now under discussion do not, (Higgins, 1968), possess any general, stability property but there is nevertheless reason to suppose that stability may be a common property. This is suspected because in a large system it will often be the case that the dynamics of a part of it, for example a single pathway, will approximate to that of a pseudo-unimolecular system which is known to be stable. However, since the interaction of stable sub-systems can produce large scale instability no general assertion can be made. Further evidence of a tendency to stability is that the dynamical matrix [D] will almost invariably have all its diagonal elements -ve.

9. Control by repression and induction of gene activity

Control of enzyme quantity by repression and induction of gene activity

This phenomenon, which is widespread can be expected to be important in determining the response of a system to its environment. Subject to certain restrictions it can be represented fairly adequately and involves the addition of a 'control level' to the type of system previously discussed. Thus in the example given in fig. (1), which had 16 structural genes and directly specified not only the rate expressions but also the enzyme concentrations there would be added a set of regulator genes. The regulator genes are assumed to produce very small concentrations of control proteins on a more or less constitutive basis. Some of the structural genes then have their activity controlled either separately or in groups by these control proteins which can be thought of as 'transducers' able to receive signals from concentrations of specific intermediate metabolites. This can be represented diagrammatically by extending the genetic level of fig. (1), as indicated below.



Thus the regulatory protein, represented functionally by say a block of type (Diamond R) has a 'regulatory function', $r_1(X_1, S_2)$, representing induction by X_1 and endogenous repression by S_2 and this acts co-ordinately on the enzymes E_1, E_2 of the initial pathway. Or again the three enzymes E_6, E_7, E_8 are coordinately repressed by the end product of

their pathway, S_9 . Enzymes not subject to control can be formally thought of as under, the control of R_0 which produces constant gene activity. The constants g_1, g_2, \dots are determined by the structural genes for enzymes E_1, E_2, \dots and are a measure of their intrinsic messenger producing power, thus if two copies of the same structural gene were present its ‘g’ constant would be doubled. The regulatory functions $r_1(\dots)$, etc. will be represented by suitable algebraic expressions, with values lying between ‘1’ and ‘0’ representing maximum and zero gene activity respectively. Unlike ‘rate-expressions’ there is little theoretical basis for such functions and they are intended only to represent what is known or surmised about the functional aspects of regulation. For example the feedback from S_9 mediated by the function r_2 may be of the form $r_2(S_9) = 1/(1 + S_9)$, the gene having half of maximal activity at unit concentration of S_9 . More complex functions can express the known features of regulation somewhat better and this will be returned to in CH IV.

The relation of enzyme concentration to gene activity involves making a number of rather general assumptions and in addition cannot be discussed without involving a ‘complete’ system. However, for the case of a subsystem which can be regarded as a small part of a larger system then the level of an enzyme in the subsystem will be proportional to the activity of its structural gene. Subject then to this restriction that the subsystem is a small part of the whole of metabolism, it becomes possible to consider control loops acting within the subsystem and also from its ‘environment’, which now includes ‘constant’ internal metabolites, by extending the general equation (2.3) to be

$$\varphi_i = \sum \lambda_{ij} g_j f_j(\dots) \cdot \eta_{jk} r_k(\dots) \quad (2.9)$$

Here summation is over j and k and $[\eta_{jk}]$ is a regulatory matrix showing which regulator genes control which enzymes. Even when such ‘control’ is added the mathematical situation is unchanged except that an additional type of ‘block’ has to be considered in arriving at the $\varphi_i(\dots)$. The formulation of a system now consists in providing all rate expressions and all regulatory functions together with the stoichiometric and regulatory matrices λ_{ij} and η_{jk} .

The introduction of ‘genotypic parameters’, via relation (2.6), previously assumed, in the diploid case, that enzyme concentration was proportional to gene dosage. If genotypic parameters are embedded in a system with a ‘control’ level this becomes the much more reasonable assumption that ‘capacity’ for gene action, or the value of g, constants, is proportional to dosage. If it is further assumed that the alternative alleles at a locus are regulated in parallel, which will often be the case, then (2.9) becomes

$$\varphi_i(\dots) = \sum_{j,k} \lambda_{ij} \left[(1 - \pi_j) g_j^a f_j^a(\dots) + \pi_j g_j^b f_j^b(\dots) \right] \eta_{jr_k}(\dots) \quad (2.10)$$

10. Complete growing systems

On p.58 we considered how to formulate a subsystem within a steadily growing larger system. The effect of growth on pools and fluxes within the subsystem was allowed for by adding an extra flux to expansion' term, $G \times S_i$ away from each pool when formulating the equations for the S.S. The exponential growth constant, $G = (\frac{1}{V}) \dot{V}$, was, in these circumstances, assumed to be given as a fixed parameter. Clearly this is not an adequate formulation if we wish to consider problems such as how the growth constant G itself arises as a system measure and how it is dependant, via the complete system, on the various genetically specified parameters.

The problem of formulating the necessary 'complete' systems is theoretically rather different from our previous considerations insofar as it become necessary to specify how the rate of increase of volume, \dot{V} , is determined.

Account must also be taken in such systems of the fact that the rate of increase of the different enzymes must compete for the currently available overall synthetic capacity of the system. Thus if more of this capacity is allocated to any enzyme less will be available for all the others. such formulations tend to assume specific mechanisms which are at best of dubious generality. However, a simple type of formulation which represents growth and allows for competition whilst making only fairly general assumptions is as follows.

Let the volume of the system at any instant be V then the rate of change of the total quantity of the i th pool can be expressed in the usual way as

$$\frac{d}{dt} (S_i \cdot V) = \sum \lambda_{ij} \cdot V \cdot E_j \cdot f_j(S)$$

Suppose that there is some conceptual enzyme E , which mediates the overall rate of protein synthesis, representing, perhaps, the complete ribosomal machinery. Assume further that this overall rate (which is responsible for the production of all enzymes) is allocated to a particular enzyme, E_j , in proportion to priming activity, g_j , of its structural genes. On this basis we can write down the rate of increase of the total quantity of the j th protein as follows

$$\frac{d (E_j V)}{dt} = \frac{g_j}{\sum g_j} \cdot V \cdot E_p \cdot f_p(S)$$

Where the g_j may now be functions of pool levels if gene activity is supposed subject to control. Finally suppose that a priming activity, g_0 , decides the allocation to non enzymic structural material such as cell wall. Assuming this structural material to be, synonymous with volume we then have an equation for \dot{V} . Thus, where k is a constant, we have

$$\dot{V} = V \times k \times \frac{g_0}{\sum g_j} \cdot E_p \cdot f_p(S)$$

Simplifying these equations in the usual way the S.S. of the growing system is defined by

$$\begin{aligned}\dot{S}_i &= \sum \lambda_{ij} \bar{E}_j \cdot f_j(\bar{S}) - \left(\frac{1}{V} \dot{V} \right) \cdot \bar{S}_i = 0 \quad (a) \\ \dot{E}_j &= \frac{g_j}{\sum g_j} \cdot \bar{E}_p \cdot f_p(S) - \left(\frac{l}{V} \cdot \dot{V} \right) \cdot \bar{E}_j = 0 \quad (b) \\ \left(\frac{1}{V} \dot{V} \right) &= \frac{k g_0}{\sum g_j} \cdot E_p \cdot f_p(\bar{S}) \quad (c)\end{aligned}\tag{2.11}$$

These can be reduced to only n equations, for the pools \bar{S}_i , by using (b) and (c) to eliminate the \bar{E}_j and the term $\left(\frac{l}{V} \dot{V} \right)$ from (a). Thus we have

$$\sum \lambda_{ij} g_j f_j(\bar{S}) - \frac{g_0}{\sum g_j} \cdot g_p \cdot k \cdot f_p(\bar{S}) \cdot S_i = 0 \tag{2.12}$$

Furthermore we can show, from the equation for \dot{E}_p in the set (b), that the growth rate, G, is given by a simple measure on the S.S. system defined by (2.12). Thus

$$G = \left(\frac{1}{V} \dot{V} \right) = \frac{g_p f_p(\bar{S})}{\sum g_j} \tag{2.13}$$

Equations of the type (2.12) are thus suitable for computer studies of the S.S. properties of growing systems with allocation. The question of ‘optimal allocation’ in such systems will be considered at the end of CH. III.

3

Sensitivity Analysis

At the start of CH 1 a general mathematical framework was outlined within which it would be possible to consider the stationary state properties of metabolic systems. Following this, the problem of how any particular system might be formulated in mathematical terms was treated, when it was seen that this necessitated only the provision of a suitable set of rate expressions together with a stoichiometric matrix. Using this formulation it was possible to describe systems ranging from the simple ‘straight chain of linear unimolecular enzymes’ through to complex networks containing bimolecular reactions, enzyme levels modulated by control loops and other non-linear effects.

However, it was pointed out that only for the hypothetical case of a system composed of purely linear reactions was it possible to write down an explicit algebraic solution for the stationary state pools, \bar{S}_i , in terms of the genetic and environmental parameters, $P_1 \dots, P_b$. Even for such linear systems this explicit solution, requiring as it does the inversion of a fixed ‘D’ matrix involved cumbersome algebraic formulae and provided little insight into the way the S.S. is determined by the parameter values. Therefore, since systems of interest tend to possess non-linear features, it appears that even for theoretical ‘models’ we will normally be without any algebraic formulation of the functional relationship connecting the S.S. pools, \bar{S}_i , with the parameter values and represented by the following set of equations:

$$\bar{S}_i = \bar{S}_i (P_1, \dots, P_b), i = 1 \text{ to } n \quad (3.1)$$

Similarly we will be without the corresponding functions for any measure, \bar{M}_x , these functions being represented by the equation below:

$$\bar{M}_x = \bar{M}_x (P_1, \dots, P_b) \quad (3.2)$$

Given that our main purpose is to ‘understand’ how variation in genetic and environmental parameters will affect properties of the S.S. solution it seems that the ‘theorist’ is reduced to a merely numerical investigation, using a computer to evaluate the functional forms of type (3.1), (3.2) for specific values of the parameters.

This places the theorist, using model systems, in a very similar situation to that of the quantitative experimentalist studying real systems, for in each case the basic operation is to modify one or more parameters from some standard position and observe the quantitative effect of this on some system measure(s). The theorist achieves this by using the computer to evaluate this functions (3.2) to any desired accuracy, for a given model and a given set of parameter values, whereas the experimentalist must construct altered genotypes and environments and determine what effect this has on a range of system measures. In both cases, then, data is collected by the procedure of discovering values of the functions (3.2) over some limited region of parameter space. The experimenter and theorist however have different though related aims. The former is usually concerned to establish the structure and parametric values of model metabolic systems which are consistent with his present data within the limits of experimental accuracy. Hopefully this will enable him to predict the probable outcome of future experiments on his ‘real’, system and at the same time provide a basis for understanding its general quantitative behaviour when subject to parametric disturbances. However, in order to achieve this he will have to try out a series of models with varying structures and parametric values until a reasonable fit with the data is obtained and such an activity will require considerable general understanding of how the S.S. properties of a system arise out of its structure and parametric values. This leads to a major aim of the quantitative theorist which is to provide a basis for such ‘understanding’ by studying the response to parametric disturbance of a variety of systems of defined structure.

Despite this impossibility of theory to provide simple expressions for functions of type (3.2), the computer theoretical approach to understanding has distinct advantages over an experimental approach, which attempts to use real systems to achieve the same end. Thus in experimental work the underlying metabolic system, which it is desired to understand, is never known with complete certainty and in addition all measurements and estimates of parameter values are subject to considerable error. Furthermore the effort and time required to construct and measure the necessary genotype/environment combinations usually limits very severely the number which can be employed when compared with those which can be generated and analysed in a computer study. Thus it can be seen that a study of the response to parameter variation in a variety of model systems can rapidly provide the necessary body of reliable data upon which any understanding of the properties of S.S. systems will have to be based.

A final very considerable advantage of a theoretical study is the possibility of using the computer not only to locate the values, \bar{M} of relationship (3.2) but also to discover their sensitivity coefficients with respect to one or more of the parameters P_1, P_2, \dots

This can easily be done for several parameters and measures at a time and, as will be shown in CH IV, involves only a small amount of computation above that already required

to merely locate the S.S. The equivalent experimental determinations are, in terms of time and effort, almost impracticable.

1. Sensitivity coefficients

Such simulation studies are, of course, purely numerical and can deal only with particular cases. This inherent limitation makes it essential, (Lewontin, 1963), to define clearly what indices of response should be computed and, tentatively, what sort of relations may exist between the various indices descriptive of system response. Only in parallel with such theoretical ideas can the simulation techniques of the next chapter be usefully employed.

In this respect a theoretical index for the response of a measure \bar{M}_x to a differential change in a parameter P_y was introduced in the first part of CH.I as $C_{P_y}^{\bar{M}_x}$.

Where

$$C_{P_y}^{\bar{M}_x} = \frac{P_y}{\bar{M}_X} \times \frac{\partial \bar{M}_X}{\partial P_y} = P_y \frac{\partial (\log(\bar{M}_X))}{\partial P_y} \quad (3.3)$$

We now wish to explain in more detail the meaning of such sensitivity coefficients and show how a knowledge of their values can aid understanding in both theoretical and real metabolic systems. We will also prove, in a general way, a number of relations between the different coefficients possessed by a system, some of these relations arising out of the ‘block’ structure of metabolic systems while others are merely formal consequences of differential analysis. Clearly, the understanding provided by a knowledge of the sensitivity coefficients will usually be relevant for only a small region of parameter space around an existing steady state, since the coefficients will themselves vary both absolutely and relative to each other when parameters are subject to appreciable change.

2. Operational meaning of the coefficient

The definition (3.3) gives rise to the approximate relation¹

$$C_P^{\bar{M}} \simeq \frac{\Delta \bar{M}}{\bar{M}} / \frac{\Delta P}{P} \quad (3.5)$$

or

$$C_P^{\bar{M}} \simeq \frac{\text{fractional response in } \bar{M}}{\text{small fractional change in } P \text{ which caused it}} \quad (3.6)$$

Suppose that a system exists in which it is possible to measure \bar{M} without error and also to make arbitrarily small changes in a parameter P . To measure $C_P^{\bar{M}}$ in such a system using the relation (3.6) it is only necessary to create a small disturbance in P and when the system has

¹There appears to be no equation (3.4) in the original document.

settled to a new slightly displaced S.S. measure the fractional response in \bar{M} . Thus if a 10% increase in P produced only a 1% increase in \bar{M} we should have $C_P^{\bar{M}} \simeq 0.1$. Alternatively if the coefficient is known it can be used to estimate the response of \bar{M} to a small finite change in P .

Clearly for the limit of increasingly small disturbances the relation (3.6) defines a measure of response which is independent of the particular units in which \bar{M} and P happen to be measured and also of the actual size of the disturbance in P . Because of this, the sensitivity coefficient is a more meaningful index of response than, say, the differential coefficient $\frac{\partial \bar{M}}{\partial P}$.

Knowledge of the relative values of the sensitivity coefficients in a system can give considerable insight into the relative importance of the different variable parameters in controlling different measures. For example the question of which enzyme in a chain of enzymes is 'rate controlling' for the flux through the chain was considered in CH.I to be equivalent to the question of whether the coefficient of the flux with respect to one of the enzymes was markedly greater than any of the others. Another example arises in relation to a quantitative concept of 'metabolic distance' where we are interested to know whether a disturbance in say, a particular enzyme, E , has a larger effect on pools which are situated close to it on the metabolic chart than on pools which appear to be more distant. In this case the relative values of the coefficient $C_E^{S_1}, C_E^{S_2}, \dots$, would enable us to see whether 'chart' distance, perhaps measured as the number of intervening metabolic steps, is generally similar to quantitative 'metabolic distance'.

The two examples just given show that it would be of little use to know a single coefficient since usually we are interested in comparisons between a group of coefficients, either over different measures or with respect to different parameters. In general, therefore, it is valuable to know a matrix of coefficients $\left[\bar{C}_{P_y}^{\bar{M}_x} \right]$ where x and y run over a suitable set of measures and parameters respectively. It was mentioned previously that such a matrix can be obtained automatically almost as a by-product of this computation necessary to evaluate the set of function (3.2), and it is therefore advantageous in computer studies to obtain this matrix since it provides a great deal of extra information about this response of the system to parameter variation. In real systems, however, considerable effort is necessary to measure each coefficient and it will usually not be possible or worthwhile to build up any sort of matrix by straightforward measurements. However, in many systems the measurement of the response of pools and independent fluxes to only a few parameter modulations together with a knowledge of the ratios of the independent fluxes will enable the relative values of many coefficients of interest to be found without excessive experimental effort. This can be done by using our prior knowledge of the 'structure' of system and employing the methods and relationships set forth later in this chapter.

3. Formal relations of different measures

Although at first sight it appears that there are as many independent measurements on the S.S. of a system as one cares to make; it will in fact often happen that a measure is related in some way to other measures.

For example it was pointed out on p.52 that in a system with m transformational steps and n pools it is always possible to express n of the S.S. fluxes as linear combinations of remaining $(m - n)$. A simple illustration of this is a divided pathway where the flux, F , in the main chain must be equal to the sum of the fluxes (F_1, F_2) in the separate arms or

$$F = F_1 + F_2 \quad (3.7)$$

A general formal result will now be obtained for use whenever such dependance between measures exists. Suppose that a measure \bar{M} depends in some way on a set of other measures $\bar{M}_1, \bar{M}_2, \dots$ and possibly directly on various parameters of type P , this can be expressed in the usual functional notation by

$$\bar{M} = \bar{M}(\bar{M}_1(P_1, \dots), \bar{M}_2(\dots), \dots, P_d, P_e) \quad (3.8)$$

Where each of the measures $\bar{M}_1, \bar{M}_2, \dots$ is itself regarded as a function of the system parameters. By applying the formal rules of partial differentiation to (3.8) we obtain, with respect to P_k

$$C_{P_k}^{\bar{M}} = \frac{P_k}{\bar{M}} \frac{\partial \bar{M}}{\partial P_k} = \frac{P_k}{\bar{M}} \left[\frac{\partial \bar{M}}{\partial P_k} \right] + \sum \frac{M_i}{\bar{M}} \left[\frac{\partial \bar{M}}{\partial \bar{M}_i} \right] \cdot \frac{P_k}{\bar{M}_i} \frac{\partial M_i}{\partial P_k}$$

or

$$C_{P_k}^{\bar{M}} = [C_{P_k}^{\bar{M}}] + \sum C_{P_k}^{\bar{M}_i} [C_{M_i}^{\bar{M}}] \quad (3.9)$$

In this formula the coefficients enclosed within square brackets refer to the sensitivity of function (3.8) considered in isolation from the overall system, whereas the ordinary system coefficients are, unbracketted. Two examples will be given to illustrate what is meant by this:

A) Taking the simple branched pathway case of (3.7) we have

$$[C_{F_1}^F] = \frac{F_1}{F} \frac{\partial (F_1 + F_2)}{\partial F_1} = \frac{F_1}{F_1 + F_2}$$

and similarly $[C_{F_2}^F] = \frac{F_2}{F_1 + F_2}$. Then, since no parameter is directly involved in (3.7) the general relation (3.9) gives:

$$C_P^F = \frac{F_1}{F_1 + F_2} \cdot C_P^{F_1} + \frac{F_2}{F_1 + F_2} \cdot C_P^{F_2} \quad (3.10)$$

This shows that the sensitivity of the main flux to any parameter is just a weighted sum of the sensitivity of the pathway fluxes; the weighting being proportional to the contribution of each branch to the main flux.

B) A simpler form of the relation (3.7) occurs when a measure \bar{M} results from multiplying and dividing together a set of other measures. Consider for example

$$\bar{M} = (\bar{M}_1 \times \bar{M}_2) / (\bar{M}_3 \times \bar{M}_4),$$

In this case $C^M \bar{M}_1 = 1$ and similarly for \bar{M}_2 ,
also $C_{\frac{\bar{M}}{M_3}} = -1$, and similarly for \bar{M}_4 . So that (3.9)

becomes

$$C_P^{\bar{M}} = C_P^{M_1} + C_P^{M_2} - C_P^{M_3} - C_P^{M_4} \quad (3.11)$$

Thus when a measure is formed by multiplication and division of a set of measures its coefficient with respect to any parameter is merely the sum of the separate coefficients of the measures in the numerator minus that of the measures in the denominator.

4. Block coefficients

In the last chapter it was shown that metabolic systems can be regarded as made up of functional blocks. Thus ‘transformational blocks’ which could be enzymic or non enzymic, were represented in a diagram by,



described by a rate expression giving, F , the net flux carried by the block. Similarly there were ‘regulatory blocks’ represented by, \diamond and quantitatively described by a regulatory function, R , giving the fraction by which the maximal activity of a structural locus should be multiplied to take account of regulation.

The concept of the sensitivity coefficient C_P^M was introduced as a means to describing and measuring the nett effect of changes in genetically determined parameters on some systemic property say the measure \bar{M} . Even with the simplest rate expression, a number of parameters are associated with each block and this number increases considerably when bimolecularity, regulation and inhibition has to be introduced. Insofar as all such parametric changes can affect the system only by acting through the one block, it is convenient to consider their common effect by defining a new coefficient, the ‘block coefficient’. It will be shown that, on the other hand, the differences between the effects of parameter within a given block are described by a set of further coefficients descriptive of the behaviour of the isolated block.

The block coefficient is introduced by the following considerations. A fairly general set of

equations defining the S.S. was of the form

$$\sum \lambda_{ij} F_j = 0 \quad (3.12)$$

Where F_j stands for a 'complete' transformational expression involving, in general, both a rate expression f_j and a regulatory function R_k .

Consider a new 'constructed' system related to the original one in that each complete transformational expression, in (3.32), is replaced by a modified form F'_j obtained by multiplying each F in (3.12) by an extra parameter α_j . When the constructed system has all its α_j equal to unity it will, of course, have the same S.S. solution as the original system and will behave identically with regard to other parameter variations. The equations defining the S.S. of the constructed system will be

$$\sum \lambda_{ij} \alpha_j F_j = 0 \quad (3.13)$$

The coefficient of a measure \bar{M} with respect to, say, the 3rd transformational block in the constructed system can now be defined in the usual way as

$$C_{\alpha_3}^{\bar{M}} = \frac{\alpha_3}{\bar{M}} \cdot \frac{\partial \bar{M}}{\partial \alpha_3} \quad (3.14)$$

The block coefficient in the original system is then defined as $C_{\alpha_3}^{\bar{M}}$ evaluated with all α 's equal to unity, that is around the position where the original and the constructed systems behave identically.

It will now be shown how the sensitivity coefficient, $C_P^{\bar{M}}$, is related to $C_{\alpha_3}^{\bar{M}}$ when P is a parameter embedded within the expression F_3 .

The relation can be discovered by considering that it is always possible to make simultaneous small changes $\Delta\alpha_3$ and ΔP such that the S.S. solution, \bar{S}_i , will remain completely unchanged, as also will all measures on this S.S. not directly involving α_3 or P . This gives rise to two relations.

Considering first, the system as a whole it gives rise to the relation

$$C_{\alpha_3}^{\bar{M}} \cdot \frac{\Delta\alpha_3}{\alpha_3} + C_P^{\bar{M}} \cdot \frac{\Delta P}{P} = 0 \quad (a)$$

On the other hand considering the transformational flux expression, F'_3 , in isolation, and noting that the only quantities occurring in it which are subject to change are α_3 and P and that these produce zero net effect we have

$$\left[C_{\alpha_3}^{F'_3} \right] \cdot \frac{\Delta\alpha_3}{\alpha_3} + \left[C_P^{F'_3} \right] \cdot \frac{\Delta P}{P} = 0 \quad (b)$$

in (b) the square brackets are used to denote coefficients of the transformational flux expression, considered in isolation from the rest of the system and with respect to one of its arguments.

Combining (a) and (b) and remembering that, since $F'_3 = \alpha_3 \cdot F_{3'}$

$$\left[C_{\alpha_3}^{F'} \right] = 1 \text{ and } \left[C_P^{F'} \right] \equiv \left[C_P^F \right] \text{ we get}$$

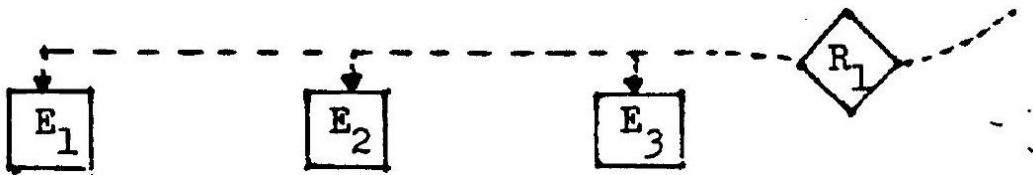
$$C_P^{\bar{M}} = C_{\alpha_3}^{\bar{M}} \cdot \left[C_P^{F_3} \right] \quad (3.15)$$

The relationship (3.15) which applies in both the original and the constructed system, states that the sensitivity of a measure \bar{M} to a parameter P will depend jointly on the sensitivity of \bar{M} to the block in which P acts and on the sensitivity of the flux through this block, when considered in isolation from the rest of the system, to the parameter P itself.

In the more general case where the parameter P occurs in connection with several transformational blocks (such as multiple inhibition by an external substance) the relation (3.15) becomes

$$C_P^{\bar{M}} = \sum C_{\alpha_r}^{\bar{M}} \times \left[C_P^{F_r} \right] \quad (3.16)$$

where the summation effectively extends over all blocks influenced directly by P . Consider for example a system, as indicated below, containing a group of three enzymes having rate expressions f_1, f_2, f_3 and each under co-ordinate control from a block with regulatory function R_1



The separate f flux expressions will then be $f_1 \times R_1, f_2 \times R_1, f_3 \times R_1$, and the sensitivity of any measure \bar{M} to the regulatory block, R_1 , can be considered as the sensitivity to a parameter β , of value unity, multiplying into the regulatory function R_1 . Using the relation (3.16) we get

$$C_{\beta}^{\bar{M}} = C_{\alpha_1}^{\bar{M}_1} \cdot \left[C_{\beta}^{f_1 \times \beta \times R_1} \right] + C_{\alpha_2}^{\bar{M}} \left[C_{\beta}^{f_2 \times \beta \times R_1} \right] + \dots$$

Using the formal relationship (3.11) to simplify the bracketed expressions we get

$$C_{\beta}^{\bar{M}} = C_{\alpha_1}^M + C_{\alpha_2}^M + C_{\alpha_3}^M.$$

This result shows that the sensitivity to the regulatory block is just the sum of the sensitivities to the separate transformational blocks involved in the co-ordinate regulation.

4.1 Coefficient with respect to an enzyme

In a system where enzyme quantities are specified as parameters, that is one in which no gene control loops are operative, all flux expressions will be of the form

$$F_j = E_j \times f_j(--) \quad (3.17)$$

Using, relation (3.15) to connect the block coefficient to the enzyme we have

$$C_{E_j}^M = C_{\alpha_j}^M \cdot [C_E^{F_j}]$$

then using (3.17) and the formal relationship (3.11)

$$C_{E_j}^M = C_{\alpha_j}^M [C_{E_j}^{E^j}] + [C_{E_j}^{f_j}] = C_{\alpha_j}^M (1 + 0)$$

so that

$$C_{E_j}^M \equiv C_{\alpha_j}^M \quad (3.18)$$

Thus for such a system, the block coefficient is synonymous with the coefficient with respect to enzyme quantity previously considered in CH. 1. When, however, the enzyme level is subject to control there is no such identity since E_j is no longer a parameter and $C_{E_j}^M$ has no clear meaning.

4.2 Coefficient with respect to turnover number

Even when enzyme quantity is subject to control, the rate expression will still be of the form $F_j = E_j(\dots)t_j f_j(\dots)$ Where the enzyme level, E_j , is now a function of some sort and t_j is the turnover number. In this case an argument, similar to that above, will show that

$$C_{\alpha_j}^M \equiv C_{t_j}^M \quad (3.19)$$

4.3 Coefficient with respect to gene dosage

When control is operative and on the assumption that the system under consideration is a small part of some larger system the gene dosage parameter g_j introduced in the last chapter will always appear multiplied into the complete flux expression for the jth transformation or

$$F_j = g_j \times t_j \times f_j(-) \times n_{jk} \times R_k(-)$$

Therefore once again

$$C_{\alpha_j}^M \equiv C_{g_j}^M. \quad (3.20)$$

A slight correction to this result will be necessary if it is conceded that the system under consideration is not a small part of a larger system but it will still be approximately true. Another difficulty arises, in relation to the meaning of $C_{g_j}^M$ in (3.2.0), because in most organisms it is not possible to vary g_j continuously. However, this problem will not arise in e.g. Neurospora where g_j can be varied almost continuously

5. Measurement of coefficients in a ‘real’ system

From the foregoing it can be seen that in understanding a real system considerable insight will be obtained if the sensitivity of a selected set of measures can be determined even with respect to each transformational block, rather than for all parameters. These block coefficients cannot usually be measured as $C_E^{\bar{M}}$ since in most systems there is some degree of enzyme control and enzyme quantity is not a true parameter. However, there is no such objection to obtaining the block coefficients via $C_{t_j}^{\bar{M}}$ or $C_{g_j}^{\bar{M}}$ and both of these methods have been employed. Even so several difficulties arise in practice. Thus, since it is not easy, to measure small changes in \bar{M} and P with sufficient accuracy, it will usually be necessary to measure \bar{M} as accurately as possible for a number of more widely spaced ‘P’ values and construct a smooth curve showing the relationship between \bar{M} and P when all other parameters are held constant. From this curve $C_P^{\bar{M}}$ can be estimated around the original P value using the definition (3.6). When the sensitivity with respect to external concentration (of type X) is required no further difficulties arise since such quantities can be varied continuously and measured with reasonable accuracy. However, in order to measure internal ‘block’ coefficients it is necessary to produce a controlled change in a single internal (genetically controlled) parameter and this is not always simple, or indeed possible.

For example, one way of estimating the block coefficient of an enzyme would be to measure $C_t^{\bar{M}}$ using a series of revertants from a mutant in which the activity of the enzyme had been abolished by a mutation at its structural locus. Unfortunately each of these ‘new’ proteins would not only have a different turnover number from the original enzyme but, in general, also different values for all the other kinetic parameters characterising the protein. In such a case it is only possible to plot \bar{M} against turnover number after establishing that other kinetic parameters are effectively uncharged. A method of this general type has been employed by C. Chilcott (Honours Thesis, Edinburgh, 1961) and B. I. Barthelmes (personal communication) using revertants in the arginine pathway of *N. crassa*. Another method, which is intended to avoid the complication of simultaneous changes occurring in several parameters, is to attempt to alter only the quantity of a particular structural gene. In principle this can be done by altering the ‘dose’, that is the parameter g_j , and estimating the block coefficient via $C_{g_j}^{\bar{M}}$.

A way in which a particular g_j can be altered almost continuously without altering anything else is by constructing heterokaryons in fungi between wild type and a strain in which activity at the structural locus in question is abolished. By varying the input of the two types contributing to the heterokaryon the dosage of the structural gene in question can be controlled over a considerable range. A method of this general type has been employed by W. Tateson (Ph.D. Thesis, Edinburgh, 1971), also for steps in the arginine pathway of *Neurospora*.

Since it is clear that the determination of coefficients by these ‘direct’ experiments is extremely laborious, it is important, to establish any further general properties of coefficients, particularly in relation to the expected distribution of control between different blocks and

to any 'diagnostics' which would make it easier to determine the pattern of coefficients experimentally. We will return to this subject after considering the various relations between coefficients and demonstrating how the coefficient values arise from measurable properties of the individual transformation blocks.

6. Summation properties of block coefficients

At the end of CH.I it was shown, relation (1.31), that the sum of the flux-control coefficients in a chain of non linear enzymes was equal to unity. Now that we have suitably extended the definition of coefficient to general networks and networks in which enzyme quantities cannot be treated as a parameter it will be demonstrated that this and other properties still hold, being indeed a very general and simple consequence of the block structure of metabolic systems.

Thus consider the general S.S. equations (3.13) defining the constructed system:

$$\sum \lambda_{ij} \alpha_j F_j = 0$$

I. "The sum of the sensitivity coefficients of any given pool, S , with respect to each of the separate transformation blocks in the system is zero"

$$\sum_{j=1, \dots, m} C_{\alpha_j}^S = 0 \quad (3.21)$$

II. "The sum of the sensitivity coefficients of any given flux, F , in the system with respect to each of the separate transformation blocks in the system is unity"

$$\sum_{j=1, \dots, m} C_{\alpha_j}^F = 1 \quad (3.22)$$

These two results can be obtained by supposing that the same fractional differential change x is made in all the α_j . Then for a given pool the fractional response is found by summing up the responses predicted from the separate sensitivities or:

$$\frac{\Delta S}{S} = \sum C_{\alpha_j}^S \cdot x$$

But $\frac{\Delta S}{S} = 0$, from the co-ordinate property. Therefore $\sum C_{\alpha_j}^S = 0$, as required.

In the case of the flux coefficients we have, similarly.

$$\frac{\Delta F}{F} = \sum C_{\alpha_j}^F \cdot x$$

and in this case the co-ordinate property predicts $\frac{\Delta F}{F} = x$, so that

$$x = \sum C_{\alpha_j}^F \cdot x \text{ or } \sum C_{\alpha_j}^F = 1, \text{ as required}$$

The result (3.22) is particularly useful in relation to the problem of discovering the rate-controlling enzyme for a particular flux in an experimental situation. Thus it predicts that provided that, the structure is such that all coefficients are +ve, the value of the coefficient for a rate controlling enzyme would have to be close to unity. This means that a single experiment on a suspected enzyme can confirm or deny the hypothesis that it is the rate controlling one. Whereas in the absence of an expected value for the coefficient of a rate-controlling enzyme it would be necessary to perform such an experiment for all the enzymes in the system and compare their values, which is hardly practical.. Unfortunately no simple prediction can be made when -ve coefficients are a possibility, as for example in systems having divided pathways.

The result (3.21) for pool coefficients shows that they must be both +ve and -ve but otherwise places no limit on their values. In a later section it will be shown, by algebraic study of a simples system, that there is no general restriction on the values which can be taken by the individual coefficients of either pools or fluxes.

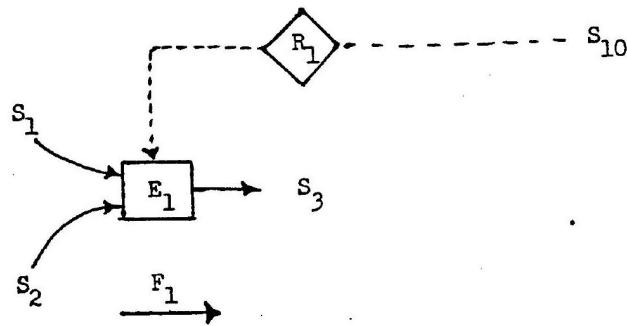
7. Elasticities of complete flux expressions

In the preceding discussion of block coefficients the sensitivity of a complete flux expression, F_j , in isolation from the rest of the system was used, that is for one of its arguments when all others are held at their S.S. levels. This was expressed by enclosing the ordinary sensitivity coefficient notation in square brackets as for example $[C_P^{F_r}]$ in (3.16). Such terms occur frequently in relations between system block coefficients and it is useful to distinguish them from these by a suitable notation. Accordingly we call them ‘elasticities’ writing the above example as $\varepsilon_p^{F_r}$.

$$C_P^{\bar{M}} = \sum C_{\alpha_r}^{\bar{M}} \cdot \varepsilon_p^{F_r} \quad (3.23)$$

The elasticities of a flux expression are, of course, not constant but will depend on the operating conditions of the appropriate enzymes under in vivo S.S. conditions. In principle they could be measured by extracting the appropriate enzymes, reconstructing their exact operating conditions in vitro and determining the sensitivity of the flux to separate modulations in substrates, products, regulators etc. In practice such a procedure is not usually feasible but it will be shown later in this chapter that it is still possible to obtain estimates of all elasticities by considering data from modulation experiments carried out around an existing S.S. in vivo.

Using the idea of elasticity it becomes straightforward to write down the response of a typical flux expression, considered in isolation, to variation in its arguments. Thus consider the subunits indicated below which together define the flux expression F_1



If α_1 is modulated it will produce consequent changes in S_1, S_2, F_1 and throughout the system, but it will always be true that:

$$\frac{\Delta F_1}{F_1} = \varepsilon_{\alpha_1}^{F_1} \cdot \frac{\Delta \alpha_1}{\alpha_1} + \varepsilon_{S_1}^{F_1} \cdot \frac{\Delta S_1}{S_1} + \varepsilon_{S_2}^{F_1} \cdot \frac{\Delta S_2}{S_2} + \dots$$

Here the complete flux expression F_1 involves more than just the rate equation, being of the form:

$$F_1 = \alpha_1 \cdot f_1(S_1, S_2, S_3) \cdot R(S_{10})$$

Consequently the elasticities can be simplified, for example

$$\varepsilon_{\alpha_1}^{F_1} = 1$$

$$\varepsilon_{S_1}^{F_1} = \varepsilon_{S_1}^{f_1}$$

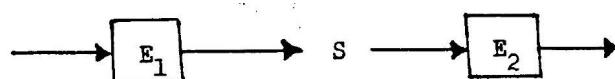
and $\varepsilon_{S_{10}}^{F_1} = \varepsilon_{S_{10}}^{R_1}$. So that

$$\frac{\Delta F_1}{F_1} = \frac{\Delta \alpha_1}{\alpha_1} + \varepsilon_{S_1}^{f_1} \cdot \frac{\Delta S_1}{S_1} + \varepsilon_{S_2}^{f_1} \cdot \frac{\Delta S_2}{S_2} + \varepsilon_{S_3}^{f_2} \cdot \frac{\Delta S_3}{S_2} + \varepsilon_{S_{10}}^R \cdot \frac{\Delta S_{10}}{S_{10}}$$

Expressions of this type will be used in subsequent discussion.

8. Relative Importance of two successive enzymes in a straight chain

The situation often arises, even within a complex system, that two enzymes can be found such that the product of one is the substrate of the next and has no strong interaction with any other protein in the system. Such a situation is indicated below for two hypothetical enzymes, E_1 and E_2 , which may be quite complicated such as being bimolecular and suffering inhibition.



Provided that S only affects E_1 and E_2 it will always be possible to make simultaneous changes in the block parameters α_1, α_2 such that only S is altered. This is achieved by increasing α_1 and then decreasing α_2 so as to exactly cancel the effect on all pools and fluxes other than the common pool S . Suppose that such an operation produces a small fractional change $\frac{\Delta S}{S}$ in S . Considering the flux expression, F_1 , in isolation from the rest of the system and noting that simultaneous changes in α_1 and S produce no change in F_1 , we have, using the elasticity notation

$$\varepsilon_{\alpha_1}^{F_1} \cdot \frac{\Delta \alpha_1}{\alpha_1} + \varepsilon_S^{F_1} \cdot \frac{\Delta S}{S} = 0$$

or

$$\frac{\Delta \alpha_1}{\alpha_1} = -\varepsilon_S^{F_1} \cdot \frac{\Delta S}{S} \quad (a)$$

Similarly for E_2

$$\frac{\Delta \alpha_2}{\alpha_2} = -\varepsilon_S^{F_2} \cdot \frac{\Delta S}{S} \quad (b)$$

Finally, since the simultaneous changes in α_1, α_2 have no effect on any system measure, except S , we have for the whole system,

$$C_{\alpha_1}^M \cdot \frac{\Delta \alpha_1}{\alpha_1} + C_{\alpha_2}^M \cdot \frac{\Delta \alpha_2}{\alpha_2} = 0 \quad (c)$$

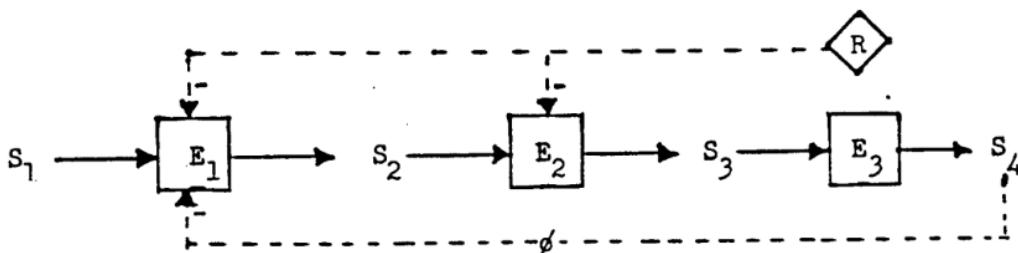
where C_α^M are the ordinary systemic block coefficients of a measure M .

Combining the three equations a, b, c , we have that for two successive enzymes in a chain the ratio of their sensitivity coefficients C_1, C_2 for any measure is given by

$$\frac{C_1}{C_2} = -\frac{\varepsilon_2^{F_2}}{\varepsilon_1^{F_1}} \quad (3.24)$$

This result will be true no matter how complex the system within which the two enzymes are embedded.

For example consider two ‘linear’ enzymes, E_1 and E_2 within some larger system as indicated below and suppose them subject to repression via a manresion function R and the first one to non competitive inhibition with an inhibitory function ϕ .



9.. RELATION BETWEEN THE COEFFICIENTS OF ALL BLOCKS AFFECTED A GIVEN POOL

To find the relative importance of their block coefficients for a.1 measures (other than S_2) we apply (3.24) or

$$\frac{C_1}{C_2} = -\frac{\varepsilon_{S_2}^{F_2}}{\varepsilon_{S_2}^{F_1}} = -\frac{\varepsilon_{S_2}^{(\alpha_2 \cdot f_2 \cdot R)}}{\varepsilon_{S_2}^{(\alpha_1 \cdot f_1 \cdot R \cdot \phi)}}$$

Using the formal relation (3.11) to simplify this we get

$$\frac{C_1}{C_2} = \frac{\varepsilon_{S_2}^{\alpha_2} + \varepsilon_{S_2}^{f_2} + \varepsilon_{S_2}^R}{\varepsilon_{S_2}^{\alpha_1} + \varepsilon_{S_2}^{f_1} + \varepsilon_{S_2}^R + \varepsilon_{S_2}^\phi} = \frac{0 + \varepsilon_{S_2}^{f_2} + 0}{0 + \varepsilon_{S_2}^{f_1} + 0 + 0}$$

or

$$\frac{C_1}{C_2} = \frac{\varepsilon_{S_2}^{f_2}}{\varepsilon_{S_1}^{f_2}}$$

so that the ratio of block coefficients depends on the ratio of the elasticities of the two rate expressions with respect to the common pool. In particular, if the reactions are linear, f_1 will be of the form

$$f_1 = v_1 \left(S_1 - \frac{S_2}{K_1} \right)$$

, and using the definition (3.3) for the coefficient we have

$$\varepsilon_{S_2}^{f_1} = [C_{S_2}^{f_1}] = S_2 \frac{\partial (\log f_1)}{\partial S_2} = \frac{S_2/K_1}{S_1 - S_2/K_1}$$

$$\text{similarly } \varepsilon_{S_2}^{f_2} = \frac{S_2}{S_2 - S_3/K_2}$$

$$\text{so that } \frac{C_1}{C_2} = \frac{(S_1 - S_2/K_1)}{(S_2 - S_3/K_2)} \cdot K_1$$

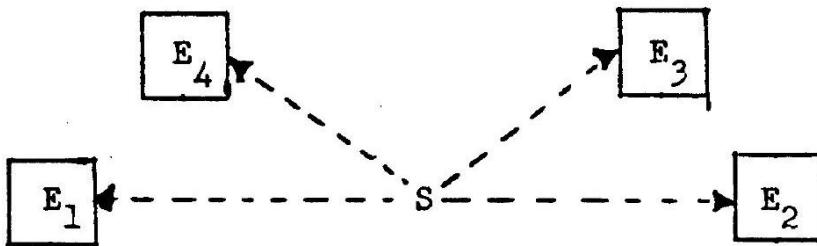
The two terms in numerator and denominator will be recognised as the ‘out-of-equilibrium’ of the two steps. This confirms and extends the result in CH.I concerning the usefulness of the ‘out-of-equilibrium’ term as a diagnostic for rate control in simple cases.

9. Relation between the coefficients of all blocks affected a given pool

The relation (3.24) applying to two successive blocks affected by a common pool can now be generalised to cases where the given pool influences more than two enzymes, as for example if it is a signal for repression or perhaps at the dividing point of a branched pathway.

Since it turns out that the argument does not depend on the flux relationship of the influenced enzymes it will be sufficient to refer to a general pool influence diagram, indicated below for enzymes E_1, E_2, E_3 and E_4 which are all affected by pool S . In this diagram the

dotted lines signify influence, that is non zero elasticities with respect to S , and flux arrows, which may coincide, have been omitted.



As in the previous argument, when S affected only two enzymes, it is always possible to make simultaneous movements in $\alpha_1, \alpha_2, \alpha_3$ and α_4 such that only S is changed, all other pools in the system remaining constant. By considering F_1 in isolation we must have:

$$0 = \frac{\Delta F_1}{F_1} = \varepsilon_{\alpha_1}^{F_1} \cdot \frac{\Delta \alpha_1}{\alpha_1} + \varepsilon_S^{F_1} \cdot \frac{\Delta S}{S}$$

or

$$\frac{\Delta \alpha_1}{\alpha_1} = \varepsilon_S^{F_1} \cdot \frac{\Delta S}{S} \quad (a)$$

and similar relationships for E_2, E_3 and E_4

Also since no measure, other than S , is affected by these simultaneous changes in the α values, we can write for the system as a whole

$$C_{\alpha_1}^M \cdot \frac{\Delta \alpha_1}{\alpha_1} + C_{\alpha_2}^M \cdot \frac{\Delta \alpha_2}{\alpha_2} + \dots = 0 \quad (b)$$

Combining (b) with the four relationships of type (a) the connection between the system coefficients of the four blocks and their elasticities with respect to the chosen pool, S , is found to be:

$$C_{\alpha_1}^M \cdot \varepsilon_S^{F_1} + C_{\alpha_2}^M \cdot \varepsilon_S^{F_2} + C_{\alpha_3}^M \cdot \varepsilon_S^{F_3} + C_{\alpha_4}^M \cdot \varepsilon_S^{F_4} = 0 \quad (3.25)$$

This is the required generalisation of (3.24), when only two pools were involved, and clearly such an equation can be written down for every pool in the system. The relationship for each pool, of course, involves only those blocks having non-zero elasticities with respect to it. Expressed generally the relationship is:

$$\sum C_{\alpha_j}^M \cdot \varepsilon_S^j = 0 \quad (3.26)$$

Where ε_S^j is short for $\varepsilon_S^{F_j}$.

For a system containing n pools there will be a relations of type (3.26) and within each of these most of the ε_S^j will be zero, since most pools influence only a few enzymes. A special

case arises in the above arguments when the pool being modulated is itself taken as the measure. When this is so the result (3.26) will be slightly modified. Thus the relationship (b), above, now becomes

$$C_{\alpha_1}^S \cdot \frac{\Delta\alpha_1}{\alpha_1} + C_{\alpha_2}^S \cdot \frac{\Delta\alpha_2}{\alpha_2} + \dots = \frac{\Delta S}{S}$$

and when combined with (a) and written in the ‘elasticity’ notation yields

$$\sum C_{\alpha_j}^S \cdot \varepsilon_S^j = -1$$

This can be generalised for the case where the measure is any pool S_i and the modulation is performed on any pool S_k , when

$$\sum C_{\alpha_j}^{S_i} \cdot \varepsilon_{S_k}^j = -\delta_k^i \quad (3.27)$$

In relation (3.27), $\delta_k^i = 1$ when $i = k$ and zero when $i \neq k$

Clearly in an n pool m block system it will be possible to write down n relations of type (3.26) or (3.27) connecting the m block coefficients of any given measure. However, since for a general system $m \geq n + 1$, these will not suffice to determine the separate block coefficients and it will be necessary to write down a further $(m - n)$ relations between the block coefficients.

10. System coefficients and a given independent flux

Relations between system coefficients for all blocks contributing to a given independent flux

The $(m - n)$ supplementary relationship referred to above can be written down by considering the different distinct ways in which it is possible to modulate various α_j and change the relative fluxes in the S.S. condition of the network without altering any pools. Once this has been achieved it will be possible to solve the block coefficients in terms of the underlying elasticities.

In general this can be done in exactly $(m - n)$ distinct ways, each one corresponding to that set of modulations which alters a particular one of the $(m - n)$ independent fluxes without affecting any of the other independent fluxes or any of the pools. Thus a local flux can be expressed as a linear combination of the independent fluxes, as in the example previously given on p.53 CH.II.

For the j th dependent flux,

$$F_j = \sum a_{jk} Y_k \quad (3.28)$$

Where Y_k stands for the values of the designated set of independent fluxes at the S.S. and the a_{jk} can be found from the stoichiometric matrix $[\lambda_{ij}]$ Suppose that we wish to find the

set of ‘ α ’ modulation which will affect the independent flux Y_1 and no other independent flux or pool. Clearly if, by differentiation of the stoichiometric relation (3.28), $\frac{\Delta Y_1}{Y_1}$ is the small fractional change in one of the independent fluxes, Y_1 , then the charge in a local flux, F'_j , of the constructed system must be given by

$$\frac{\Delta F'_j}{F'_j} = \left[\frac{Y_1}{F'_j} \cdot \frac{\partial (\sum \alpha_{jk} Y_k)}{\partial Y_1} \right] \cdot \frac{\Delta Y_1}{Y_1} \quad (3.29)$$

a convenient notation for this is

$$\frac{\Delta F'_j}{F'_j} = \varepsilon_{Y_1}^{F'_j} \cdot \frac{\Delta Y_1}{Y_1} \quad (3.30)$$

In relation (3.30) $\varepsilon_{Y_1}^{F'_j}$ stands for what may be called the ‘stoichiometric elasticity’ of F'_j with respect to the independent flux Y_1 . Carrying out the differentiation of (3.28) we have

$$\varepsilon_{Y_1}^{F'_j} = \frac{a_{j1} Y_1}{\sum a_{jk} Y_k}$$

and it appears that the stoichiometric elasticity measures the “proportion of the j th flux which is accounted for by a given independent flux”. If now at each transformational block the α_j is changed proportionally to the stoichiometric elasticity with respect to the independent flux Y_1 , the final result will be that the necessary flux changes are achieved without altering any pools and without altering any independent fluxes other than Y_1 . Since for any pool, S the set of a modulations just described has no effect we have

$$\sum C_{\alpha_j}^S \cdot \frac{\Delta \alpha_j}{\alpha_j} = 0, \text{ where all the } \alpha_j \text{ are equal to } \varepsilon_{Y_1}^{F'_j} \cdot \frac{\Delta Y_1}{Y_1}$$

so that

$$\sum C_{\alpha_j}^S \cdot \varepsilon_{Y_1}^{F'_j} = 0 \quad (3.31)$$

Relations of the type (3.31) provide the necessary $(m - n)$ equations referred to earlier since there is one such relation for each of $(m - n)$ independent fluxes. If the measure involved is the independent flux itself then

$$\sum C_{\alpha_j}^{Y_1} \cdot \frac{\Delta \alpha_j}{\alpha_j} = \frac{\Delta Y_1}{Y_1}, \text{ where } \frac{\Delta \alpha_j}{\alpha_j} = \varepsilon_{Y_1}^{F'_j} \cdot \frac{\Delta Y_1}{Y_1}$$

so that

$$\sum C_{\alpha_j}^{Y_1} \cdot \varepsilon_{Y_1}^{F'_j} = 1 \quad (3.32)$$

or more generally

$$\sum C_{\alpha_j}^{Y_i} \cdot \varepsilon_{Y_k}^{F'_j} = \delta_k^i \quad (3.33)$$

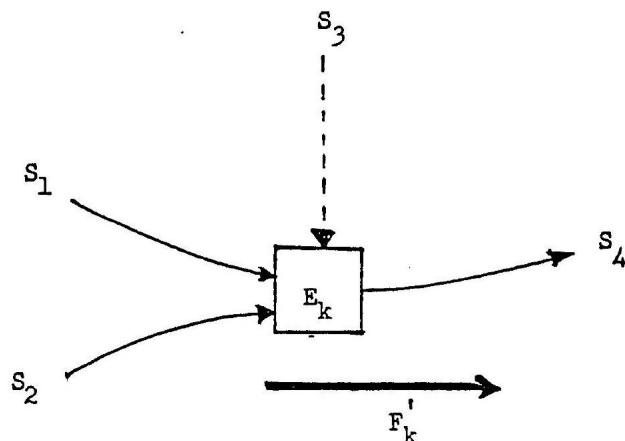
The (m-n) relation of type (3.31) or (3.33) obtained by considering modulation of the different independent fluxes and involving the appropriate ‘stoichiometric elasticities’ are clearly similar to the n relation of type (3.26) or (3.27) obtained by considering modulation of pools and involving flux expression elasticities. Together they are sufficient to determine the block sensitivities for any specified measure. The use of this will be shown in the examples which will follow.

11. Relations between the system coefficients

Relations between the system coefficients of different measures connected with a given block

The relation of (3.26), (3.27) and (3.31), (3.33) all connect the coefficient of a given measure with respect to different blocks. A similar type of argument can be used to establish the relationship between the coefficient of different measures with respect to a single block so that once we have established, say, the flux coefficients $C_{\alpha_j}^F$ it will be possible to calculate from them the various pool measures $C_{\alpha_j}^S$.

Consider for example the k th block and suppose it to be influenced by S_1, S_2, S_3, S_4 as shown below.



Suppose that the α of the j th block is modulated by $\frac{\Delta\alpha_j}{\alpha_j}$, then considering the whole system.

$$\frac{\Delta F'_k}{F'_k} = C_{\alpha_j}^{F'_k} \cdot \frac{\Delta\alpha_j}{\alpha_j}, \frac{\Delta S_1}{S_1} = C_{\alpha_j}^{S_1} \cdot \frac{\Delta\alpha_j}{\alpha_j}, \dots \text{etc.} \quad (3.34)$$

Considering now the kth block in isolation we can write 99.

$$\frac{\Delta F'_k}{F'_k} = \varepsilon_{\alpha_j}^{F'_k} \cdot \frac{\Delta \alpha_j}{\alpha_j} + \varepsilon_{S_1}^{F'_k} \frac{\Delta S_1}{S_1} + \varepsilon_{S_2}^{F'_k} \cdot \frac{\Delta S_2}{S_2} + \dots$$

Replacing the fractional changes by means of relations (3.34) this becomes

$$C_{\alpha_j}^{F'_k} \cdot \frac{\Delta \alpha_j}{\alpha_j} = \varepsilon_{\alpha_j}^{F'_k} \frac{\Delta \alpha_j}{\alpha_j} + \varepsilon_{S_1}^{F'_k} \cdot C_{\alpha_j}^{S_1} \cdot \frac{\Delta \alpha_j}{\alpha_j} + \dots$$

Now $\varepsilon_{\alpha_j}^{F'_k} = \delta_j^k$, $\frac{\Delta \alpha_j}{\alpha_j}$ will cancel out and if we suppose, for example,

that $C_{\alpha_j}^{S_4}$ is unknown and the other coefficients are known we get:

$$C_{\alpha_j}^S = \left(C_{\alpha_j}^{F'_k} - \delta_j^k - C_{\alpha_j}^{S_1} \cdot \varepsilon_1^k - C_{\alpha_j}^{S_2} : \varepsilon_2^k - C_{\alpha_j}^{S_3} \cdot \varepsilon_3^k \right) / \varepsilon_4^k \quad (3.35)$$

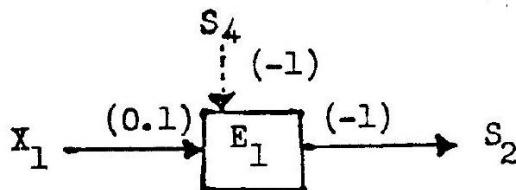
Relations of type (3.35) and those discussed earlier can be used to progressively discover the sensitivity of measures from the known sensitivity of other measures and in many simple networks it is possible to obtain a complete sensitivity matrix in this way using only the known or estimated values of flux elasticities at the existing S.S.

12. Application of theorems relating system coefficients to elasticities

Some examples will now be given to make clear how sensitivity matrices may be computed from elasticities.

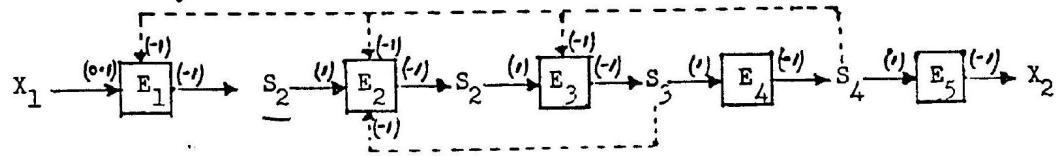
EXAMPLE I: Straight Chain with Multiple Inhibitory Loops

Consider the straight chain system indicated below where values of elasticities around the S.S. working conditions of the transformation blocks are written inside brackets at the appropriate points in the diagram. Thus:



means that the elasticities of enzyme 1 are: 0.1, -1, -1, with respect to substances X₁, S₂ and S₄ respectively. The negative values reflect the 'normal' behaviour that a product inhibits the forward reaction and that S₄ is a true inhibitor. The example includes three inhibitory loops, two of which are intersecting, to demonstrate how this affects the computation. Most of the elasticities have been taken as either +1 or -1, so that the arithmetic

shall remain simple. Although quite feasible they are not intended to represent any real system.



The computation of a sensitivity matrix for the pools and for the pathway flux, $\begin{bmatrix} C_{\alpha_j}^{F, S_i} \end{bmatrix}$ proceeds as follows:

Assume $C_{\alpha_j}^F$ to be scaled to unity, then using relations of type (3.26) arising from the virtual modulation of pools we get 101.

$$\text{from pool } S_1 : \quad C_2 = -C_1 \times \frac{\varepsilon_1^1}{\varepsilon_1^2}, \text{ where } \varepsilon_1^1 \text{ stands for } \varepsilon_1^{F'_1} \text{ and so on}$$

or

$$C_2 = -(1) \times \frac{(-1)}{(1)} = 1$$

$$\text{from pool } S_2 : \quad C_3 = -C_2 \times \frac{\varepsilon_2^2}{\varepsilon_2^3} = 1$$

$$S_3 : \quad C_4 = -\frac{(\varepsilon_3^2 \cdot C_2 + \varepsilon_3^3 \cdot C_3)}{\varepsilon_3^4} = C_2 + C_3 = 2$$

$$S_4 : \quad C_5 = -\frac{(\varepsilon_4^1 C_1 + \varepsilon_4^2 C_2 + \varepsilon_4^3 C_3 + \varepsilon_4^4 \cdot C_4)}{\varepsilon_4^5}$$

$$C_5 = C_1 + C_2 + C_3 + C_4 = 5$$

The actual coefficients must satisfy the summation theorem (3.21) and can be found by normalising the numbers just discovered, which are proportional to the flux coefficients.

This gives

$$[C_1, C_2, C_3, C_4, C_5] = [0.1, 0.1, 0.1, 0.2, 0.5]$$

Having discovered the vector of flux coefficients with respect to the different blocks $C_{\alpha_j}^F$, we may use it to successively generate the vectors for pool coefficients in the order

$$C_{\alpha_j}^{S_4}, C_{\alpha_j}^{S_3}, C_{\alpha_j}^{S_2}, C_{\alpha_j}^{S_1} \text{ using relations of type (3.35)}$$

arising from the consideration of the enzymes E_5, E_4, E_3, E_2 in isolation.

$$\text{Thus for } E_5 \text{ we get: } C_{\alpha_j}^{S_4} = (C_{\alpha_j}^F - \delta_j^5) / \varepsilon_4^5$$

$$\text{which generates the vector } C_{\alpha_j}^{S_4} = [0.1, 0.1, 0.1, 0.2, -0.5]$$

$$\text{For enzyme } E_4 : C_{\alpha_j}^{S_3} = (C_{\alpha_j}^F - \delta_j^4 - C_{\alpha_j}^S \cdot \varepsilon_4^4) / \varepsilon_3^4$$

generating $C_{\alpha_j}^{S_3} = [0.2, 0.2, 0.2, -0.6, 0]$

For enzyme E_3 : $C_{\alpha_j}^{S_2} = \left(C_{\alpha_j}^F - \delta_j^3 - C_{\alpha_j}^{S_3} \cdot \varepsilon_3^3 - C_{\alpha_j}^{S_4} \cdot \varepsilon_4^3 \right) / \varepsilon_2^3$

generating $C_{\alpha_j}^{S_2} = [0.4, 0.4, -0.6, -0.2, 0]$

Finally using the relationship for either E_1 or E_2 , which will yield identical results, we get $C_{\alpha_j}^{S_1} = [0.8, -0.2, -0.2, -0.4, 0]$

The general sequence of computation can be summarised as:

I. <u>FLUX COEFFICIENTS</u>			II. <u>POOL COEFFICIENTS</u>		
<u>Input</u> Flux Coeff. for Block No.	<u>Use Rel.</u> for Pool No.	<u>Output</u> Coeff. for Block No.	<u>Input</u> Coeff. Vector for Pool No.	<u>Use Rel.</u> for Enzyme No.	<u>Output</u> Coeff. Vector for Pool No.
$C_{\alpha_1}^F$ (all)	1	2	-	5	4
2	2	3	4	4	3
2,3	3	4	4,3	3	2
1,2,3	4	5	4,3,2	2	1

Normalise $\rightarrow C_{\alpha_j}^F$

and the result presented as

	F	S ₁	S ₂	S ₃	S ₄
a_1	.1	.8	.4	.2	.1
a_2	.1	-.2	.4	.2	.1
a_3	.1	-.2	-.6	.2	.1
a_4	.2	-.4	-.2	-.6	.2
a_5	.5	0	0	0	-.5

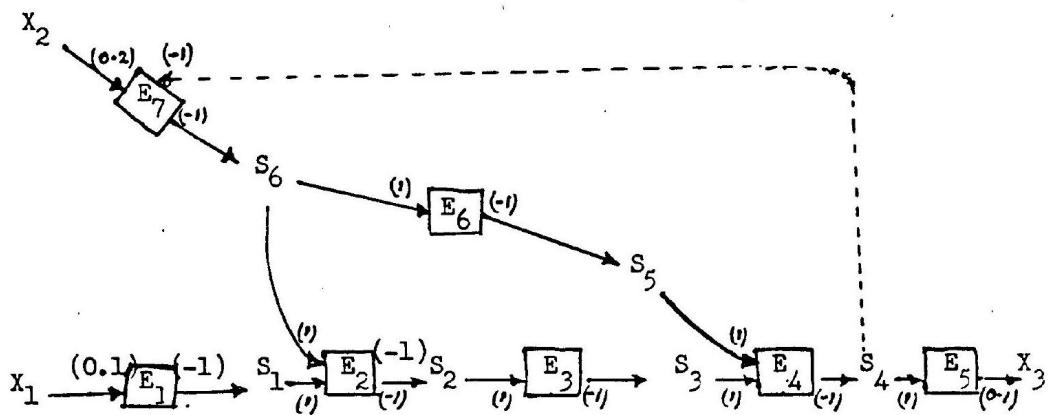
$C_{\alpha_j}^{FIS} =$

Inspection of this matrix shows that various qualitative expectations are fulfilled. Thus a given pool always has positive coefficients for enzymes upstream from it and -ve coefficients for enzymes downstream. Furthermore all flux coefficients are +ve, as would be expected in a straight chain without positive feedback.

Two further examples follow, using the same format. These are intended to illustrate how quite complex networks can be dealt with and also to give some idea of the numerical range and the general pattern which sensitivity coefficient may take.

EXAMPLE II: Bimolecularly Coupled Pathways

12.. APPLICATION OF THEOREMS RELATING SYSTEM COEFFICIENTS TO ELASTICITIES 8



In this case the sequence of computation is:

I. FLUX COEFFICIENTS

Input flux Use relation Output
Coeff. for for pool Coeff. for
Block No. No. Block No.

$(C_i^F \alpha_i)$	1	1	2
	2	2	3
	3	3	4
	4	5	6
6,2	6	7	(NORMALISE)
4,7	4	5	$C_i^F \alpha_i$

II. POOL COEFFICIENTS

Input Coeff. Use rel- Output
Vector for ation for Coeff. Vect.
Pool No. Enzyme No. for Pool No

-	1	1
-	5	4
4	7	6
6	6	5
1,6	2	2
2	3	3

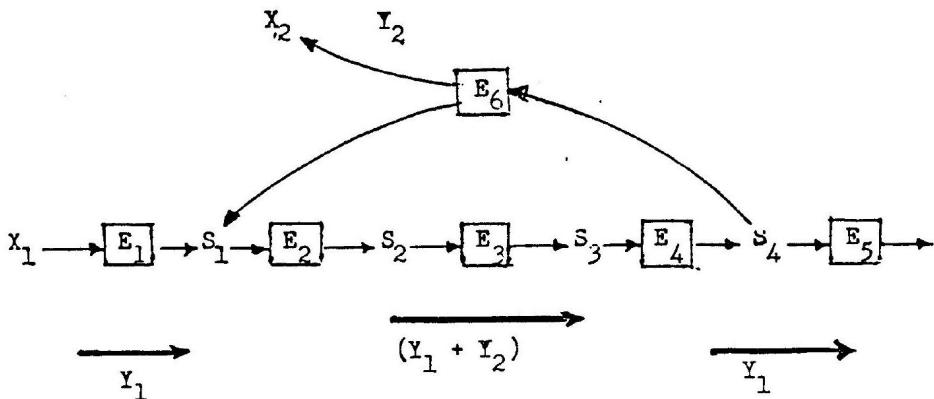
and the resulting sensitivity matrix $\begin{bmatrix} F, S_i \\ C_{\alpha_j} \end{bmatrix}$ is

	F	S_1	S_2	S_3	S_4	S_5	S_6
a_1	.1	.9	.6	.5	.1	-.3	-.2
a_2	.1	-.1	-.4	-.5	.1	-.3	-.2
a_3	.1	-.1	-.4	.5	.1	-.3	-.2
a_4	.1	-.1	-.4	-.5	.1	-.3	-.2
a_5	.3	-.3	-.2	.5	-.7	.1	.4
a_6	.1	-.1	.6	-.5	.1	.7	-.2
a_7	.2	-.2	.2	0	.2	.4	.6

EXAMPLE III: UREA TYPE CYCLE

The two previous examples have considered 'uniflux' systems, that is systems characterised by $m - n = 1$, and therefore possessing only one independent flux. For such systems, once the n relations of type (3.26) arising from pool modulations have been used, only one extra condition relating the coefficients was required and this condition was provided by the fact that $\sum C_{\alpha_j}^F = 1$.

We now wish to consider systems with several independent fluxes where it will be necessary to use relations of type (3.33) which involve stoichiometric elasticities. As an example we will consider the system shown below which contains two independent fluxes Y_1 and I_2 and is similar to the urea cycle if S_1, S_2, S_3, S_4 stand respectively for ORN, CIT, ASA, ARG and x_2 is UREA. In order to define the stoichiometric elasticities it is necessary to know the ratio Y_2/Y_1 , i.e. the ratio of the flux round the cycle to the flux into protein and this is taken in this example to be $1/2$. As before all flux elasticities are taken to be $+1$ or -1 except for ε_1^6 which is taken as zero since this step is regarded as irreversible and ε_4^5 which is taken as $2/7$, again for arithmetic reasons.



Given then the ratio of the independent fluxes and the flux expression elasticities the computation of $[C_{\alpha j}^{Y_1, Y_2, S_i}]$ proceeds as follows:

First of all we note that the stoichiometric elasticities of the local fluxes with respect to the independent fluxes can be written down by considering what fraction of a given flux is accounted for by a particular independent flux. This gives

$$\varepsilon_Y^{F_j} = [1, 2/3, 2/3, 2/3, 1, 0]$$

$$\text{and } \varepsilon_Y^{F_2} = [0, 1/3, 1/3, 1/3, 0, 1]$$

Considering first the sensitivity of the independent flux Y_1 to the separate blocks, let us suppose that $C_1^{Y_1} = x$. Then the pool relationships for $S_1 S_2 S_3$ show that $C_1^Y = C_2^{Y_1} = C_3^{Y_1} = C_4^{Y_1} = x$. Suppose further that $C_5^Y = y$ and $C_6^{Y_1} = 2$.

We can now use the relation for pool S_4 to give

$$(-1) \cdot x + (2/7) \cdot y + (1) \cdot z = 0$$

and the flux modulation relations of type (3.33), where modulation of Y_2 gives

$$1/3 \cdot x + 1/3 \cdot y + 1/3 \cdot x + (1) \cdot z = 0$$

and modulation of Y_1 gives

$$(1) \cdot x + 2/3 \cdot x + 2/3 \cdot x + 2/3 \cdot x + (1) \cdot y = 1$$

These three equations are easily solved to yield

$$x = 0.1, y = 0.7, z = -0.1$$

or

$$C_{\alpha j}^{Y_1} = [0.1, 0.1, 0.1, 0.1, 0.7, -0.1]$$

A similar argument then yields $C_{\alpha_j}^{Y_2} = [0.35, 0.35, 0.35, 0.35, -1.05, 0.65]$ Before proceeding to discover the $C_{\alpha_j}^{S_i}$ using enzyme equations of type (3.35) it is necessary to know the sensitivity to each of the α_j of the flux carried by each enzyme. This has already been established for E_1, E_5, E_6 , but we still require the sensitivities for E_2, E_3, E_4 which all carry the flux $(Y_1 + Y_2)$.

This can be written down using the relation (3.10) proved earlier for a branched pathway.

$$C_{\alpha_j}^{(Y_1+Y_2)} = \frac{Y_1}{Y_1 + Y_2} C_{\alpha_j}^{Y_1} + \frac{Y_2}{Y_1 + Y_2} \cdot C_{\alpha_j}^{Y_1} = 2/3 C_{\alpha_j}^{Y_1} + 1/3 C_{\alpha_i}^{Y_2}$$

or

$$C_{\alpha_j}^{(Y_1+Y_2)} = [0.183, 0.183, 0.183, 0.183, 0.183, 0.116, 0.150]$$

From this point it is straightforward to obtain the $C_{\alpha_j}^{S_1}, C_{\alpha_j}^S, C_{\alpha_j}^3, C_{\alpha_j}^S$ using successively the relation of type (3.35) for E_1, E_2, E_3, E_4 . In this case it is important to remember that the term $C_{\alpha_j}^{F'_k}$ in (3.35) refers in each case to the flux carried by the particular enzyme which is involved.

109.

	$F_{1,5}$	$F_{2,3,4}$	F_6	S_1	S_2	S_3	S_4
a_1	.1	.183	.35	.9	.717	.534	.35
a_2	.1	.183	.35	-.1	.717	.534	.35
a_3	.1	.183	.35	-.1	-.283	.534	.35
a_4	.1	.183	.35	-.1	-.283	-.466	.35
a_5	.7	.116	-1.05	-.7	-.816	-.932	-1.05
a_6	-.1	.150	.65	.1	-.05	-.200	-.35

Suppose now that it is required to know the sensitivity of the internal pools and fluxes to the level of X_1 . To obtain this we can use the relation (3.16) and, since the parameter X_1 only occurs in the block characterised by α_1 , we have

$$C_{X_1}^M = C_{\alpha_1}^{M_1} \cdot C_{X_1}^{F_1} = C_{\alpha_1}^M \cdot \varepsilon_{X_1}^1 \quad (3.36)$$

Noting from the diagram that the 1st block is rather saturated with respect to X_1 , having an elasticity of only 0.1, we can use the first row of the sensitivity matrix, together with (3.36) to obtain

$$C_{X_1}^{F_{1,5}, F_{2,3,4}, F_6, S_i} = 0.01, 0.0183, 0.035, 0.09, 0.0717, 0.0534, 0.0351$$

The above consideration shows that for a given S.S. system a knowledge of the relative values of the fluxes and of the elasticities of the separate flux expression at their working

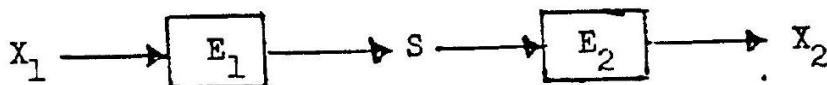
position is sufficient to determine the sensitivity matrix. In the examples given the task of calculating the sensitivity matrix from known flux and stoichiometric elasticities was always straightforward. This was because the relatively simple ‘structure’ of the systems under consideration gave rise in each case to a set of m equations which could be solved by inspection.

13. General limitations on coefficient values

Previously we have seen that in certain uniflux systems we were able to state a general limit for the value of flux coefficients. These are systems in which all the $C_{\alpha_j}^F$ are necessarily +ve, when it follows that $0 < C_{\alpha_j}^F < 1$ for all j . No limitations were present for the pool coefficients. However, as soon as -ve coefficients become possible, either because of branching or perhaps, even in a uniflux system, when certain types of inhibition are operative, it is no longer clear that there are any such limits. Since it is not easy to consider this question in a general way we shall look instead at what limitation may exist in particular systems.

14. Pool sensitivities in a chain of saturable enzymes

Consider the simplest possible case; shown below, in which there is just one pool S . We wish to consider what limits are imposed on C_1^S .



Clearly, by manipulating the relative quantity of E_1 and E_2 , S can be anywhere within the range $X_1 > S > X_2$ (where the a are scaled concentrations taking account of the equilibrium constants). From the point of view of discussing the extreme, values of $C_{\alpha_1}^S$ and $C_{\alpha_2}^S$ we may take S as being variable within this range.

Now using our previous theorem (3.26) and denoting the elasticities by ε_S^1 and ε_S^2 we can write

$$C_{\alpha_1}^F \cdot \varepsilon_S^1 + C_{\alpha_2}^F \cdot \varepsilon_S^2 = 0$$

also $C_{\alpha_1}^F + C_{\alpha_2}^F = 1$

so that $C_{\alpha_1}^F = \frac{\varepsilon_S^2}{\varepsilon_S^2 - \varepsilon_S^1}$

and $C_{\alpha_1}^S = C_{\alpha_1}^F / \varepsilon_S^2 = \frac{1}{\varepsilon_S^2 - \varepsilon_S^1} = -C_{\alpha_2}^S$

Clearly we can go no further without considering the form of the saturable rate expressions f_1, f_2 for the two enzymes. Let us assume that these are the simplest which will take saturation and reversibility fully into account, viz:

$$f_1 = \frac{X_1 - S}{1 + \frac{1}{M_1} + \frac{S}{M'_1}}$$

$$f_2 = \frac{S - X_2}{1 + \frac{S}{M_2} + \frac{X_2}{M'_2}}$$

Where M'_1 and M'_2 are the ‘reverse’ Michaelis constants. then using the formal relation, $C_P^{M_1/M_2} = C_P^{M_1} - C_P^{M_2}$, and finding the coefficients of the numerator and denominator by differentiation we have

$$\varepsilon_S^1 = C_S^{f_1} = - \left[\frac{S}{X_1 - S} + \frac{S/M'_1}{1 + X/M_1 + S/M'_1} \right] = -(a + b)$$

$$\text{and } \varepsilon_S^2 = C_S^{f_2} = \left[\frac{S}{S - X_2} - \frac{S/M_2}{1 + S/M_2 + X_2/M'_2} \right] = (c - d)$$

Where a, b, c, d represent the appropriate algebraic groupings. Clearly a and c represent components of the elasticity which depend mainly on ‘out of equilibriumness’ of the reaction whereas b and d are concerned with the degree of saturation of the reaction. Inspection shows that

$$0 < a < \infty, \quad 1 < c < \infty, \quad 0 < b < 1, \quad 0 < d < 1$$

and that when the reaction steps are close to equilibrium the saturation term will be unimportant. Using this notation we have

$$C_{\alpha_1}^S = \frac{1}{\varepsilon^2 - \varepsilon^1}$$

$$C_{\alpha_1}^S = \frac{1}{a + b + c - d} \quad \text{and it is easy to establish}$$

$$C_{\alpha_1}^S > 0$$

Is there any upper bound for $C_{\alpha_1}^S$? The problem is to consider the maximum value of $C_{\alpha_1}^S$ for any values of M_1, M'_1, M_2, M'_2 and any value of S in the range $X_1 > S > X_2$

We require a ‘minimum’ of $(a + b + c - d)$ and this will be so when

- (i) d is maximal, $= 1$, i.e. E_2 fully saturated
- (ii) b is minimal, $= 0$, i.e. E_1 fully unsaturated

This merely requires sufficiently extreme values of M'_1 and M_2 and can be satisfied regardless of the value of S .

(iii) S lies at a value which will minimise $(a + c)$

$$\text{Now } a + c = \frac{S}{X_1 - S} + \frac{S}{S - X_2}$$

and it can be shown that the minimum occurs at $S = \sqrt{X_1 \cdot X_2}$ So that

$$\begin{aligned}(C_{\alpha_1}^S)_{\max} &= \frac{1}{(a + c)_{\min} + 0 - 1} \\ &= 1/2 \left(\sqrt{\frac{X_1}{X_2}} - 1 \right)\end{aligned}$$

At this maximum position the value of the flux coefficient is

$$C_{\alpha_1}^F = 0.5$$

Thus we have established that in the absence of any knowledge other than a general form for the saturable expression and the values of X_1, X_2

$$0 < C_{\alpha_1}^S < 1/2 \left(\sqrt{\frac{X_1}{X_2}} - 1 \right)$$

Clearly then if $X_1 \gg X_2$ there is no effective limit to the value of $C_{\alpha_1}^S$ and we may expect that under certain circumstances very large values of pool coefficients could occur.

In particular a large pool coefficient appears likely whenever a saturated and almost irreversible enzyme appears at the end of a chain and when the flux control is shared equally between it and the rest of the chain.

15. Pool sensitivities in a chain of linear enzymes

In the case of two linear enzymes an almost identical argument applies, except that the saturation components of the elasticities are absent so that

$$C_{\alpha}^S = \frac{1}{a + c}$$

Once again the minimum occurs when $S = \sqrt{X_1 X_2}$

and gives the result $0 < C_{\alpha}^S < \frac{\sqrt{X_1} - \sqrt{X_2}}{\sqrt{X_1} + \sqrt{X_2}}$, so that $C_{\alpha}^S < 1$

the value of $C_{\alpha_1}^F$ at this maximum is $C_{\alpha_1}^F \simeq \frac{\sqrt{X_1}}{\sqrt{X_1} + \sqrt{X_2}}$

For example if $X_1 = 100, X_2 = 1, (C_{\alpha}^S)_{\max} = \frac{10 - 1}{10 + 1} = 82\%$

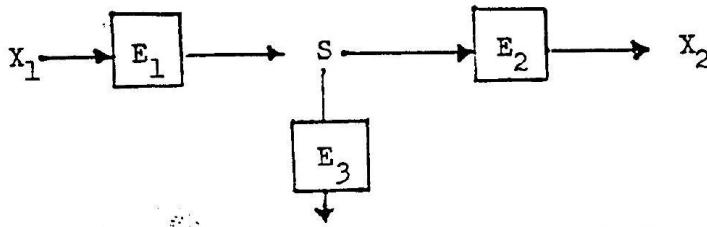
at which position $C_{\alpha_1}^F \simeq \frac{10}{10+1} = 90\%$ This indicates that the

result $C_{\alpha_1}^F = 0.5$ for the saturable case is not general and, more importantly, we have demonstrated that in a chain of linear enzymes the pool coefficients are strictly bounded.

16. Flux coefficients

The results just proved for pool coefficients can be used to show that in systems with branching, that is with more than one independent flux, there is no a priori limit on the values of flux coefficients as there is for a straight chain of enzymes.

Consider the branching system shown below, where E_1 and E_2 may be saturable enzymes.



Let us suppose that the enzyme E_3 is effectively linear and irreversible and that the flux through it, F_3 , is small compared with the fluxes F_1 and F_2 in the main pathway.

Under these circumstances it follows that $C_{\alpha_1}^{F_3} \simeq C_{\alpha_1}^S$ and, from the result just proved for a chain of two saturable enzymes, also

$$(C_{\alpha_1}^S)_{\max} \simeq 1/2 \left(\sqrt{\frac{X_1}{X_2}} - 1 \right)$$

Thus

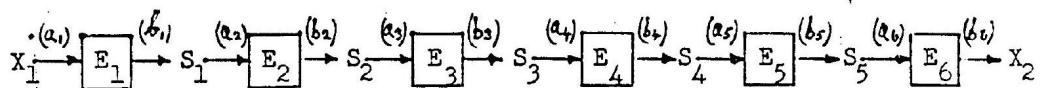
$$(C_{\alpha_1}^{F_3})_{\max} \simeq 1/2 \left(\sqrt{\frac{X_1}{X_2}} - 1 \right)$$

This demonstrates that very large + or -ve values of flux coefficients are possible in a branching system which contains saturable enzymes.

It will be noticed, however, that this result is for the coefficient of a flux with respect to an enzyme not directly in the pathway carrying the flux. In fact it is often possible to identify, within a larger system, major pathways which carry a flux large by comparison with any branches from them. The coefficients of the flux in such major pathways with respect to the enzymes actually in them will under normal circumstances still satisfy the limitation $0 < C_{\alpha}^F < 1$

17. 'Diagnostics' for rate control in non-linear systems

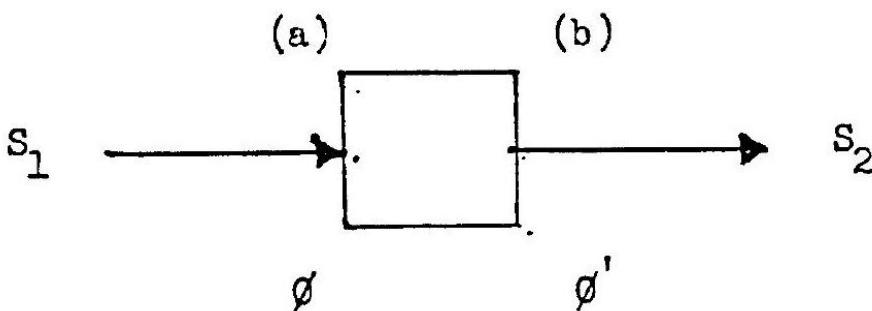
In Ch I it was shown, for a sequence of linear enzymes, that the difference in pool levels across enzymes in a chain was proportional to their respective flux coefficients, $C_{\alpha_i}^F$, provided the pool concentrations were scaled to take account of the product of equilibrium constants along the chain. Thus the pattern of scaled pool levels at the S.S. could be used as a simple diagnostic for rate control. We now wish to consider how this simple diagnostic for rate control in a chain of enzymes will be altered when allowance is made for enzymes having saturable rate expressions.



Such a system is shown above and the substrate and product elasticities 'a' and 'b' respectively, are written next to each enzyme using the appropriate suffixes. By using the pool modulation relations (3.26) and arbitrarily taking the unnormalised value of the 1st flux coefficient, C_1 , as $1/a_1$ we can show that:

$$[C_1, C_2, C_3, \dots] \propto \left[\frac{1}{a_1}, \quad \frac{1}{a_2} \left(-\frac{b_1}{a_1} \right), \quad \frac{1}{a_3} \left(-\frac{b_1}{a_1} \right) \left(-\frac{b_2}{a_2} \right), \dots \dots \dots \right] \quad (3.37)$$

This shows that the elasticities of a given enzyme, say E_3 , affect its own coefficient directly through the factor $\frac{1}{a_3}$ and indirectly by multiplying all terms 'downstream' from it by the factor $\left(-\frac{b_3}{a_3} \right)$. In order to proceed further we must now assume the simplest rate expression which will adequately represent the saturability and reversibility of the typical step shown below.



If S_1 and S_2 are suitably scaled so that the equilibrium constant is eliminated such an expression will be

$$F = \frac{\frac{V}{M} (S_1 - S_2)}{I + \frac{S_1}{M} + \frac{S_2}{M'}} = \frac{\frac{V'}{M'} (S_1 - S_2)}{1 + \frac{S_1}{M} + \frac{S_2}{M'}} \quad (3.38)$$

In this expression V and V' represent the forward and reverse maximum velocities of the enzyme while M and M' represent Michaelis constants for the substrate S_1 and the product S_2 respectively.

Using this rate equation it is now possible to write down the elasticities ‘a’ and ‘b’ with respect to S_1 and S_2 as

$$\begin{aligned} a &= \frac{S_1}{S_1 - S_2} - \frac{\frac{S_1}{M}}{1 + \frac{S_1}{M} + \frac{S_2}{M'}} = \frac{S_1}{S_1 - S_2} - \phi \\ b &= -\frac{S_2}{S_1 - S_2} - \frac{\frac{S_2}{M'}}{1 + \frac{S_1}{M} + \frac{S_2}{M'}} = -\frac{S_2}{S_1 - S_2} - \phi' \end{aligned} \quad (3.39)$$

Where the terms ϕ and ϕ' , arising from the saturation property of the rate expression with regard to S_1 and S_2 respectively, are both positive and can be seen to satisfy the inequality $(\phi + \phi') < 1$. The latter condition means that it is impossible for ϕ and ϕ' to approach their maximum values of unity simultaneously. The elasticities of (3.39) can be written as their ‘linear’ parts multiplied by correction factors which differ from unity when saturation effects are important, thus

$$\begin{aligned} a &= \frac{S_1}{S_1 - S_2} \left(1 - \phi \frac{S_1 - S_2}{S_1} \right) \\ b &= -\frac{S_2}{S_1 - S_2} \left(1 + \phi' \frac{S_1 - S_2}{S_2} \right) \end{aligned} \quad (3.40)$$

Inspection of these correction factors indicates that the effect of saturation depends not only on the terms ϕ and ϕ' but also on the equilibrium situation around the enzyme. For example if the ‘out of equilibrium ration’, $\frac{S_2}{S_1}$, is close to unity the correction will be negligible even though one or other of the saturation terms may be at its maximal value.

The factors $(\frac{1}{a})$ and $(-\frac{b}{a})$ which occur for each enzyme in (3.37) can also be written as a ‘linear’ part multiplied by correction factors β and γ respectively. Thus,

$$\frac{1}{a} = \frac{S_1 - S_2}{S_1} \cdot \frac{1}{\left(1 - \phi \frac{S_1 - S_2}{S_1} \right)} = \frac{S_1 - S_2}{S_1} \cdot \beta \quad (3.41)$$

$$\left(-\frac{b}{a} \right) = \frac{S_2}{S_1} \cdot \frac{\left(1 + \phi' \frac{S_1 - S_2}{S_2} \right)}{\left(1 - \phi \frac{S_1 - S_2}{S_1} \right)} = \frac{S_2}{S_1} \cdot \gamma$$

Using the relation (3.41) with appropriate suffixes we can write (3.37) as

$$[C_1, C_2, C_3, \dots] \cdot \left[\frac{X_1 - S_1}{X_1} \cdot \beta_1, \frac{S_1 - S_2}{X_1} \cdot \beta_2 \cdot \gamma_1, \frac{S_2 - S_3}{X_1} \cdot \beta_3 \cdot \gamma_1 \cdot \gamma_2, \dots \right] \quad (3.42)$$

In the linear case, when $\phi = \phi' = 0$ for all enzymes, all β and γ terms in (3.42) are unity and the successive flux coefficients are just proportional to the scaled pool differences,

confirming the result of CH.I. When however saturation effects are important it will be necessary to have some estimate of the factors β and γ in addition to the ratios of scaled pool levels which were a sufficient diagnostic in the linear case.

One possible way of estimating the correction factor β for an enzyme would be to use the definition, from (3.41), of

$$\beta = \frac{1}{1 - \phi \frac{S_1 - S_2}{S_1}} = \frac{1}{1 - \frac{\frac{S_1}{M}}{1 + \frac{S_1}{M} + \frac{S_2}{M'}} \times \left(\frac{S_1 - S_2}{S_1} \right)} \quad (3.43)$$

However this would involve making a close study of each enzyme, to determine the Michaelis constants M and M' , and also a knowledge of absolute pool levels rather than ratios.

Alternatively we may note from (3.38) that

$$\frac{F}{V} = \left(\frac{S_1}{M} / \left(1 + \frac{S_1}{M} + \frac{S_2}{M'} \right) \right) \times \frac{S_1 - S_2}{S_1} = \phi \cdot \frac{S_1 - S_2}{S_1}$$

so that

$$\beta = \frac{1}{1 - \phi \frac{S_1 - S_2}{S_1}} = \frac{1}{1 - \frac{F}{V}} \quad (3.44)$$

and similarly we can show that

$$\gamma = \frac{1 + \frac{F}{V'}}{1 - \frac{F}{V}} \quad (3.45)$$

Thus estimates for β and γ can be made by measuring the flux F for unit weight of the organism and dividing it by the maximum activity in the forward and reverse directions of the enzyme extracted from unit weight, forming the required factors $\frac{F}{V}$ and $\frac{F}{V'}$.

In principle, then, a knowledge for each enzyme of pool ratios and of in vivo flux against maximum flux can be used to diagnose relative flux control using the following relation, obtained by substituting the expressions (3.44) and (3.45) for β and γ into (3.42). Thus we have

$$\begin{aligned} [C_1, C_2, C_3, \dots] \propto & \left[\frac{X_1 - S_1}{X_1} \frac{1}{\left(1 - \frac{F}{V_1} \right)}, \frac{S_1 - S_2}{X_1} \frac{1}{\left(1 - \frac{F}{V_2} \right)} \frac{\left(1 + \frac{F}{V_1} \right)}{\left(1 - \frac{F}{V_1} \right)}, \right. \\ & \left. \frac{S_2 - S_3}{X_0} \frac{1}{\left(1 - \frac{F}{V_3} \right)} \left(\frac{1 + \frac{F}{V'_1}}{1 - \frac{F}{V_1}} \right), \dots \right] \end{aligned} \quad (3.46)$$

The pattern of pools around an enzyme in fact imposes limits on the values which β and γ can assume. Thus inspection of the expression (3.44) reveals that, for a given value of

S_1 and S_2 , β will be a maximum when the forward saturation, ϕ , approaches unity and a minimum when ϕ approaches zero. So that

$$1 < \beta < \frac{S_1}{S_2} \quad (3.47)$$

A similar but rather more complex argument for (3.45) shows that γ approaches a maximum when $(\phi + \phi')$ approaches unity, irrespective of the separate values of ϕ and ϕ' . So that we also have

$$1 < \gamma < \frac{S_1}{S_2} \quad (3.48)$$

When attempting to estimate flux coefficients from a given set of pool data a useful first step is to take them as proportional to scaled pool differences, using the linear enzyme assumption. If additional evidence indicates that an enzyme may be saturated the effect of allowing for this will be to increase the value of the coefficient of all enzymes downstream from it by the common factor, γ , and also to increase its own coefficient by the factor p , which will be greater than unity and may approach $\frac{S_1}{S_2}$ if the saturation is predominantly forward. The inequalities established for β and γ indicate that it will only be important to allow for saturation effects at steps having a large 'out of equilibrium' ratio, $\frac{S_1}{S_2}$.

For example if all enzymes were known to be saturated from in front, in the sense that they had $\phi = 1$, the application of (3.42) would give

$$[C_1, C_2, C_3, C_4, C_5, C_6] \propto \left[\frac{X_1 - S_1}{S_1}, \frac{S_1 - S_2}{S_2}, \frac{S_2 - S_3}{S_3}, \frac{S_3 - S_4}{S_4}, \frac{S_4 - S_5}{S_5}, \frac{S_5 - X_2}{X_2} \right]$$

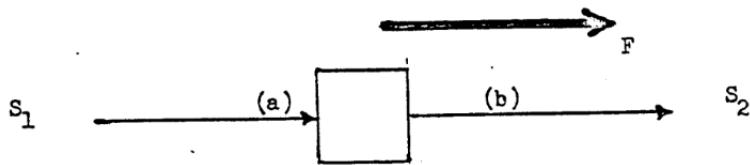
Clearly if $\frac{X_1}{X_2}$ is close to unity, that is the whole chain is close to equilibrium, then a prediction based on the linear assumption will be good. If, however, $\frac{X_1}{X_2}$ is large then enzymes further down the chain may have much greater coefficients than would be expected on a linear hypothesis.

In particular whenever X_1 is close to zero, rendering the last step irreversible, the degree of saturation of the last step becomes of critical importance.

18. Pools attenuate with metabolic distance

Does the effect on pools attenuate with metabolic distance from the site of a modulation?

We are now in a position to consider this question, for our straight chain of non-linear enzymes, taking the modulation to act on the quantity of some enzyme within the chain and 'metabolic distance' to be the number of enzymes between the modulation and the pool in question. Show below is a typical enzyme in the chain, not subject to modulation, and across which such 'attenuation' might occur once again, 'a' and 'b' indicate the appropriate elasticities.



Denoting the coefficients of the pools and of the flux, F , with respect to the remote modulated enzyme as C_1 , C_2 and C_F respectively it is convenient to consider C_1 and C_2 in terms of their deviations Δ_1 and Δ_2 from C_F , the flux coefficient C_F being, of course, the same no matter which enzyme in the chain is considered. Thus

$$C_1 = C_F + \Delta_1 \quad \text{and} \quad C_2 = C_F + \Delta_2$$

We can then discover whether Δ increases or decreases on passing through a particular enzyme by applying an enzyme relationship of type (3.35), which gives

$$C_1 = (C_F - b \cdot C_2) / a$$

or

$$C_F + \Delta_1 = (C_F - b(C_F + \Delta_2)) / a$$

so that

$$\Delta_1 = A\Delta_2 + BC_F \quad (3.49)$$

where

$$A = \left(-\frac{b}{a} \right) \text{ and } B = \frac{(1 - b - a)}{a}$$

Clearly the increase or decrease of Δ on passing through an enzyme will depend on the values assumed by the factors A and B . Assuming the non linear enzyme to be of the same form as in the last section we can use the results obtained there to discover the limits on the numbers A and B in terms of the pools S_1 and S_2 surrounding the given enzyme.

Thus we had, from (3.41), that $A = \left(-\frac{b}{a} \right) = \frac{S_2}{S_1} \cdot \gamma$

and using the inequality (3.48), for γ , this yields

$$\frac{S_2}{S_1} < A < 1 \quad (3.50)$$

Also we have from (3.39) that

$$(1 - a - b) = 1 - \frac{S_1}{S_1 - S_2} + \frac{S_2}{S_1 - S_2} + (\phi + \phi')$$

which simplifies to

$$(1 - a - b) = \phi + \phi'$$

using (3.41) to provide the factor $\left(\frac{1}{a}\right)$ this gives

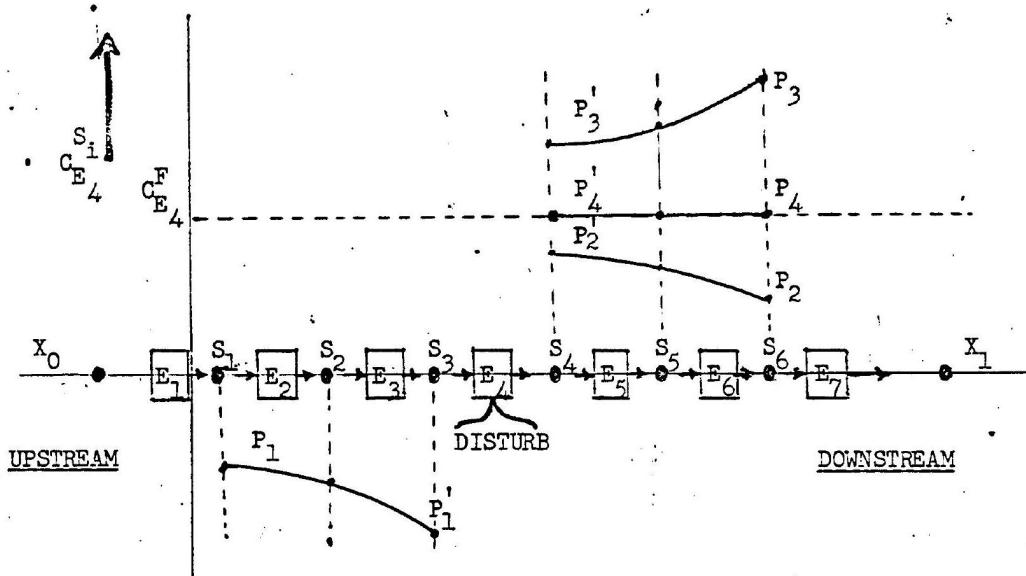
$$B = \frac{1-a-b}{a} = (\phi + \phi') \cdot \frac{(S_1 - S_2)}{S_1} \cdot \beta$$

and finally using the inequality (3.47) for β together with the fact that maximum β coincides with maximum ϕ we have

$$0 < B < \frac{S_1 - S_2}{S_2} \quad (3.51)$$

In both of these inequalities the minimum value occurs when the enzyme is 'linear', the maximum when it is fully saturated.

Referring to the chain of enzymes indicated in the diagram below, where the disturbance acts at E_4 , it is now possible to see under what circumstances attenuation of the disturbance might be expected, by using the relation (3.49) together with the inequalities just established for A and B .



Thus, on the upstream side of the disturbance it is clear that all the ΔS must be -ve and that the effect of moving through any of the enzymes E_3, E_2, E_1 must be to 'shrink' this difference from the dotted line, C_F , by the factor A and then to further add the positive quantity B . C_F . Thus whatever the condition of the enzymes upstream from E_4 they will always attenuate the fractional disturbance in the pools according to a curve such as $P_1 P'_1$ on the diagram.

Downstream such a result does not necessarily apply. What actually happens will depend on the saturation and ‘out of equilibriumness’ of the separate enzymes.

Let us suppose, for example, that enzymes E_5, E_6 are linear. In this case the factor A , for each enzyme, is of the form $\frac{S_2}{S_1}$ and the factor B will be zero. Applying the relation (3.49) to enzymes 6 and 5 therefore gives

$$\Delta_5 = \left(\frac{S_6}{S_5} \right) \cdot \Delta_6 \quad \text{and} \quad \Delta_4 = \left(\frac{S_5}{S_4} \right) \cdot \Delta_5 \quad \text{This shows that}$$

if $C_6^{E_4} < C_{E_4}^F$, that is Δ_6 is -ve, then a curve $P_2 P'_2$ representing attenuation will be the result. If, however $C_{E_4}^{S_6} > C_{E_4}^F$ a curve $P_3 P'_3$, representing amplification of the disturbance will result. A third possibility is that $C_{E_4}^{S_6} = C_{E_4}^F$ in which case the effect is to give a neutral curve $P_4 P'_4$. The sign of $\left(C_{E_4}^{S_6} - C_{E_4}^F \right)$, which determines these possibilities, depends on the condition of the last enzyme and can be considered by applying a relationship of type (3.35) to enzyme E_7 . This gives

$$a_7 \cdot C_{E_4}^{S_6} = C_{E_4}^F$$

and hence

$$C_{E_4}^{S_6} - C_{E_4}^F = C_F \left(\frac{1}{a} - 1 \right) = C_F \cdot Y$$

where Y stands for the factor $\left(\frac{1}{a} - 1 \right)$.

Using relationship (3.41) to replace $\frac{1}{a}$ along with the inequality $1 < \beta < \frac{S_1}{S_2}$ and noting that Y is a maximum when β is a maximum we can show that

$$-\frac{X_1}{S_1} < Y < \frac{(S_1 - 2X_2)}{X_2}$$

This result shows that whenever $S_1 > 2X_1$ it is possible for Y to be either negative or positive so that either an amplifying or an attenuating curve can result. The existence of the attenuating type curve, $P_2 P'_2$, or of the amplifying type, $P_3 P'_3$, depending on the degree of saturation of the last enzyme.

A further point is that if the enzymes E_5 and E_6 are not linear then there need be no steady gradient at all. However, it is never possible for $C_{E_4}^{S_i}$ to descend below $C_{E_4}^F$ if we start with $C_{E_4}^{S_6} > C_{E_4}^F$ and proceed in the direction S_6, S_5, S_4 .

19. In vivo destination of elasticities

In vivo destination of elasticities as a means of discovering the sensitivity matrix

We return now to the idea, proposed earlier in this chapter, that it may be possible to estimable the elasticities of system components around their working position on the basis of

only. a limited number of in vivo parameter modulation S.S. experiments. Having obtained estimates of these elasticities it will then be straightforward to compute the sensitivity matrix $[C_{\alpha_j}^{F, S_i}]$ using the methods already outlined.

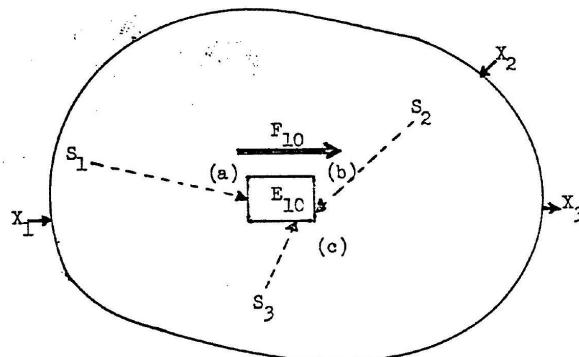
Basically the idea is that by using our information about the 'structure' of the system, that is its' metabolic diagram together with all control information, and measuring only the ratios of observed small fractional responses of type $\frac{\Delta S}{S}$ and $\frac{\Delta F}{F}$ consequent on the in modulation of only a few environmental parameters it will be possible to arrive at a complete sensitivity matrix.

Such an approach to discovering the sensitivity matrix involves no absolute measurements whatsoever, assumes nothing about the form of the rate-expression except what is implied in the structure diagram, and involves only a small number of parameter modulations which can often be external and therefore easily manipulable. Furthermore, since the required rations of small fractional responses do not depend on the details of a parametric modulation but only on its location it becomes relatively straightforward to use revertants, with small but measurable effects, as a source of modulation when the structure of the system necessitates manipulation of genetic parameters.

The direct experimental methods for estimating coefficients outlined earlier, although sharing the above advantage of being in vivo and measuring only fractional responses, contrast unfavourably because they involve complicated experiments to produce exactly known modulations of genetic parameters and because it would be necessary to repeat this laborious procedure for each block in the system if we are to produce a sensitivity matrix.

Similarly the 'diagnostic' approach, although perhaps useful for a quick appraisal from pool levels of possible rate controlling steps, appears even more unfavourable. Thus it involves assumptions about the form of rate-expression which are based on in vitro studies and may be of dubious relevance to the actual in vivo working of the enzyme. In addition it is difficult to determine the necessary factors to produce 'scaled' pool values and the method itself is not easy to extend beyond the straight chain case just presented.

Assume then that the structure of a problem is tentatively known, that the system under study has three external pools and that a particular enzyme in it, E_{10} , has non zero elasticity for the three pools, S_1, S_2, S_3 as shown below.

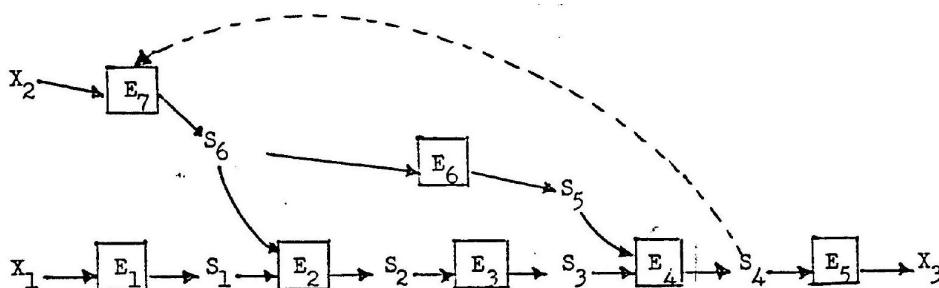


Denoting the elasticities by a, b , and c as shown and the fractional responses of S_1, S_2, S_3 and F_{10} respectively by r, s, t and u and using appropriate suffixes we may consider the enzyme in isolation from the rest of the system. Using the elasticities to write down the response of the flux to variation in the pools we have:

$$\begin{aligned} \text{For response to } X_1, \quad r_1a + s_1b + t_1c &= u_1 \\ " \quad X_2, \quad r_2a + s_2b + t_2c &= u_2 \\ " \quad X_3, \quad r_3a + s_3b + t_3c &= u_3 \end{aligned} \quad (3.52)$$

Provided the three response vectors $(r, s, t, u)_{1,2,3}$ can be considered independent, which will usually be apparent from inspection of the structure diagram, the equations (3.52) will then be sufficient to determine the three elasticities. An example will make the method clear. Suppose that from general evidence the structure diagram is as below:

(Note that because of the stoichiometry there is only one independent flux, F.)



Suppose further that the observed fractional responses of the common flux F and pools S_1, \dots, S_6 to unknown but small modulation in X_1, X_2 and X_3 is given in the table below.

OBSERVED RESPONSE IN:

	F	S_1	S_2	S_3	S_4	S_5	S_6	
MODULATE	X_2	0.2	1.8	1.2	1.0	0.2	-0.6	-0.4
	X_2	0.2	-0.2	0.2	0	0.2	0.4	0.6
	X_3	0.15	0.15	-0.6	0.75	0.15	-0.45	-0.3

Using these data with the method just explained we can apply equations of type (3.52) to find that, for example, the elasticities of E_2 with respect to S_1, S_6 and S_2 are $1, 1, -1$ respectively. For enzymes such as E_3 which have only two elasticities any two out of the three possible equations will be equivalent. Calculating all the elasticities in this way and using them to construct the sensitivity matrix we obtain

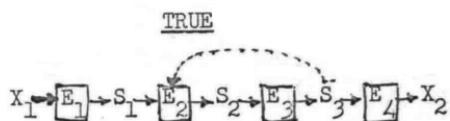
	F	s_1	s_2	s_3	s_4	s_5	s_6	
$\begin{matrix} F, s_i \\ C_{a_j} \end{matrix}$	a_1	.1	.9	.6	.5	.1	-.3	-.2
	a_2	.1	-.1	-.4	-.5	.1	-.3	-.2
	a_3	.1	-.1	-.4	.5	.1	-.3	-.2
	a_4	.1	-.1	-.4	-.5	.1	-.3	-.2
	a_5	.3	-.3	-.2	.5	-.7	.1	.4
	a_6	.1	-.1	.6	-.5	.1	.7	-.2
	a_7	.2	-.2	.2	0	.2	.4	.6

This result will be true even if the different pools have different 'extractibilities' and it involves no knowledge of equilibrium ratios or other absolute kinetic quantities. In addition it has been obtained using data from only, three modulation experiments which themselves involve no manipulation of genetic parameters or even knowledge of the size of the modulations.

When performing the necessary modulation experiments it may happen that an X modulation which produces measurable but small responses around one enzyme will produce large changes around another. In such cases it is quite in order to obtain data for the second enzyme using a lower level of modulation, since the equations (3.52) are entirely 'Local' in nature. For more complex systems having several independent fluxes it will be necessary to use the observed fractional flux responses to calculate also the stoichiometric elasticities by means of equations similar to (3.52). Inspection of the structure diagram will indicate which of these stoichiometric elasticities will be non zero for each of the enzymes. The sensitivity matrix for such a system can then be found using the additional relations (3.33) which involve the stoichiometric elasticities, the arginine cycle example given earlier illustrates this.

We wish now to consider briefly what problems might arise if the given response data were interpreted on the basis of wrong assumptions about the stricture. In particular it is of interest to know whether such errors can be detected by analysis of the given data alone or whether further data becomes necessary.

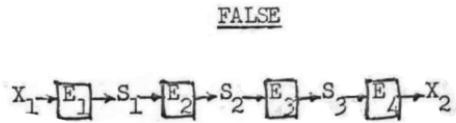
The diagrams below show a simple example of such a situation Indicating, on the left, the 'true' structure, the experimental data on which it is based and the sensitivity matrix which results. On the right of the diagram the situation is shown where the existence of the feedback inhibition is, wrongly, not taken into account and the sensitivity matrix is computed on the basis of only the X_1, X_2 modulation data.



OBSERVED RESPONSE

	F	S ₁	S ₂	S ₃
X ₁	.1	.4	.2	.1
X ₂	.6	-0.6	0	-0.6
E ₃	.1	-0.1	-0.8	0.1

MODULATE



OBSERVED RESPONSE

	F	S ₁	S ₂	S ₃
X ₁	.1	.4	.2	.1
X ₂	.6	-0.6	0	-0.6

MODULATE

Calculated elasticities will be same as for true except for enzyme E₂ where e₃² is assumed to be zero and e₁², e₂² are wrongly found to be -1 and 2.5 respectively, instead of their real values of 1, -1.

$$\begin{bmatrix} F & S_i \\ C_j & a_j \end{bmatrix} = \begin{bmatrix} a_1 & .2 & .8 & .4 & .2 \\ a_2 & .2 & -.2 & .4 & .2 \\ a_3 & .1 & -0.1 & -0.8 & .1 \\ a_4 & .5 & -0.5 & 0 & -0.5 \end{bmatrix}$$

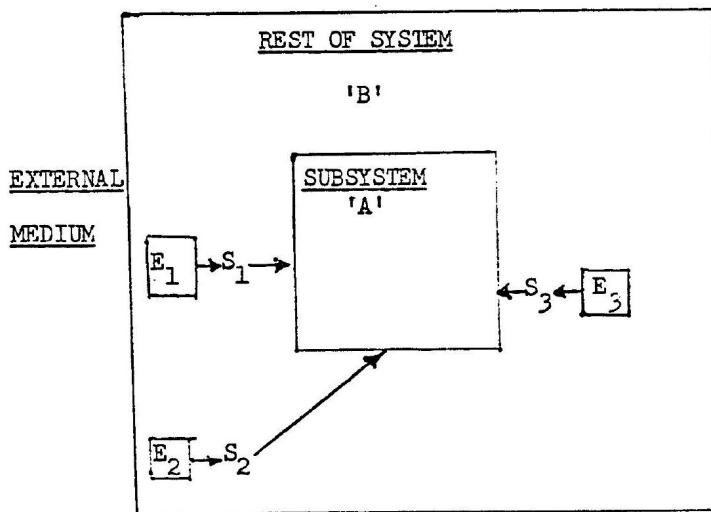
$$\begin{bmatrix} F & S_i \\ C_j & a_j \end{bmatrix} = \begin{bmatrix} a_1 & .2 & .8 & .4 & .2 \\ a_2 & -.2 & 0.2 & -.4 & -.2 \\ a_3 & 0.5 & -0.5 & 0 & 0.5 \\ a_4 & 0.5 & -0.5 & 0 & -0.5 \end{bmatrix}$$

Inspection of the above reveals that the matrix deduced from the limited data and based upon false assumption about structure is nevertheless completely compatible with the given data. By this is meant that the 1st and 4th rows of the matrix are indeed proportional to the observed X₁, X₂ response vectors. However the unusual elasticities found for enzyme E₂, which in this example are the reverse of the normal pattern of substrate activation and product inhibition, should warn the investigator that his assumptions about structure may be false. Only when further investigation has revealed the true structure and suitable extra data has been obtained, by modulating E₃, can the correct matrix be found. Further consideration shows that whenever such 'loops' are in fact present then the discovery of the true matrix will always involve the use of a revertant affecting an enzyme within the loop.

20. Modulation to determine elasticities and coefficients

Use of genetic modulation to determine elasticities and coefficients within a subsystem

Because of the very large number of pools and enzymes in any real biological system the experimenter will normally confine himself to observation on only a small part of it, perhaps to the values of pools and fluxes in a few related pathways. Such a subsystem, A, together with all pools influencing it, S_1 , S_2 and S_3 for example, is indicated below.



Genetic manipulation in B, possibly via the substitution of altered enzymes at E_1 , E_2 or E_3 , can produce several independent response vectors for the pools and fluxes within and around A. How many such modulations will be actually required to estimate the elasticities of all enzymes in A depends on the structure of A. In particular if A contains loops or other complications it may in addition be necessary to use further substitutions actually within the subsystem 'A'. Once these elasticities have been determined we can write as many 'pool' relations of type (3.26) connecting the coefficients with respect to blocks inside A as there are pools inside A.

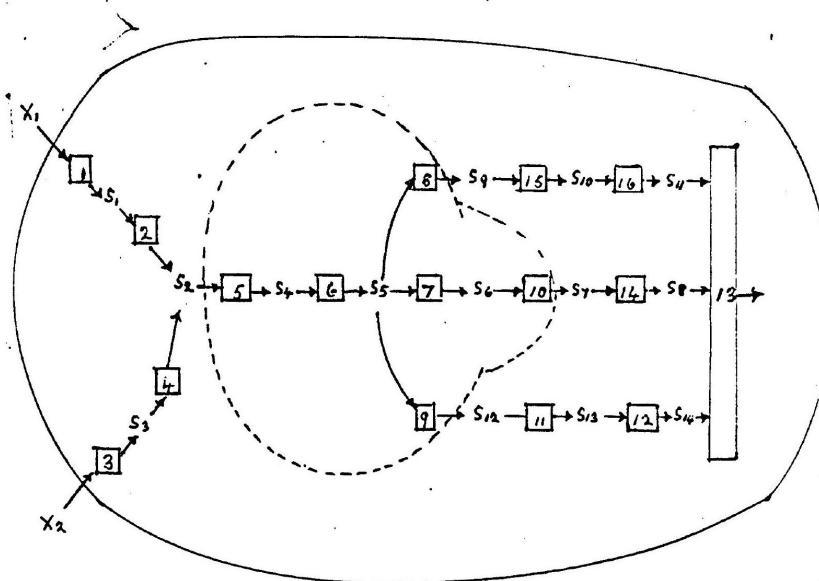
However, further information will still be required since the number of blocks in A must always exceed the number of pools by at least one. This information can be provided by direct measurement, using the methods described earlier, of one or more suitably chosen coefficients within A which can then be used together with the 'pool' relations for A to determine the remaining coefficients. This approach involves the labour associated with the necessary 'directly measured' coefficients but is much less laborious than direct measurement of all coefficients in A.

Alternatively if certain overall assumptions can be made about the structure of B this difficulty can be completely avoided. The difference between these two methods will be made clear in the following example.

21. Determination of $C_{\alpha_J}^{F_1 S_i}$ for a system

a) Making limited use of internal modulation data.

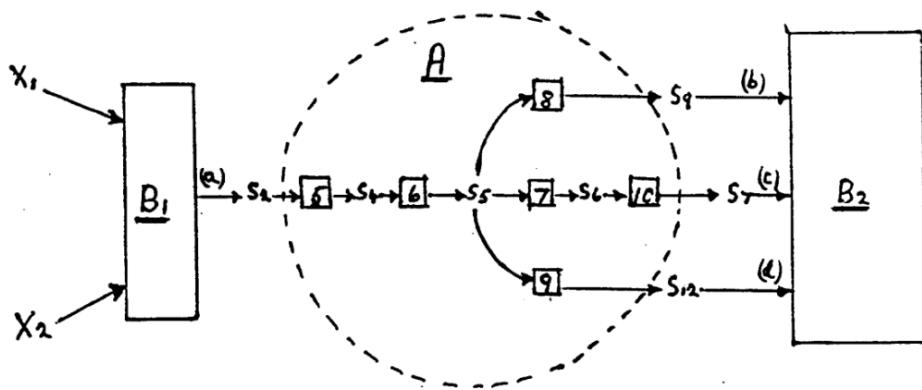
Suppose that the complete metabolic system is shown below and let the subsystem of interest and the rest of the system be 'A' and 'B' respectively:



If no assumptions are made about the structure of B then we may proceed as follows. Since no enzyme inside A has more than two elasticities, substitution of, say, E_2 and E_{14} in B will provide sufficient pool and flux modulations to determine elasticities within A . These elasticities can then be used with pool relation for S_4 , S_5 and S_6 to write down the equations connecting the six block coefficients within A . If now absolute coefficients for E_7 , E_8 and E_9 are determined directly and the others found from them then the process will be complete.

b) Using only internal modulation data.

The alternative process, which avoids the laborious direct determination of coefficients within A , involves knowledge of the structure of B and is as follows. From the point of view of subsystem A the details of B need not be known except that it can be regarded as made up of two lumped parts B_1 and B_2 as shown below.



As before all elasticities within A can be found by substitution in B_1 and B_2 and in addition the effective elasticity 'a' of the subsystem B_1 with respect to S_2 . Finally the same substitution in B_1 together with two more made within A, say at E_8 and E_9 , will together provide sufficient date to determine the elasticities, (b, c and d) of the lumped block B_2 . From these elasticities and the simplified structure the true absolute coefficients within A may be determined. Thus, provided the structural assumptions are correct, the experimental effort is just that of determining pool and flux responses for the subsystem A for four substitutions about which nothing need be known except that their effects be not too large.

This latter approach involves exploiting to the full all structural knowledge of the larger metabolic system which will have been gained by a variety of qualitative methods over many years. It seems to offer the best hope of making a 'quantitative' estimate of the relative importance of different transformational steps, or blocks, since it builds effectively on existing qualitative information and involves only a minimum of quantitative data which need not involve any absolute measurements. The difficulties in the method arise mainly from the difficulty in determining with reasonable accuracy the small fractional changes in pools and fluxes which are needed and this will be particularly severe in the case of flux determinations.

22. Optimal allocation for growth

At the end of CHI.II it was shown that a fairly general model giving the S.S. of a 'complete' growing system was defined by the form

$$\sum \lambda_{ij} g_j f_j(\bar{S}) - \frac{g_0}{\sum g_j} \cdot g_p \cdot k \cdot f_p(\bar{S}) \cdot \bar{S}_i = 0$$

The growth rate, G , being given by the measure:

$$G = \frac{g_p f_p(\bar{S})}{\sum g_j}.$$

We now wish to consider, for a system where the g 's are given as parameters, what configuration of the allocation vector $[g_1, g_2, \dots]$ will maximise t growth rate.

Provided the allocation to structure is small compared with that to active protein $\frac{g_0}{\sum g_j}$ will be small and it is possible to treat the problem in a simple way. Thus under these circumstances the problem becomes identical to one where the E_j are given as parameters and the S.S. is defined by

$$\sum \lambda_{ij} E_j f_j(\bar{S}) = 0$$

The aim being to maximise the growth rate, $G = \frac{E_p f_p(\bar{S})}{\sum E_j} = \frac{F_p}{\sum E_j}$ by adjusting the different enzymes. This can be done by adjusting only the ratio of the enzymes whilst keeping their sum constant, since it is only the ratio of the enzymes which affects G. When the optimal allocation exists a small change in the ratio of any two enzymes such that the sum remains constant will have no effect on the flux to protein F_p . Thus if an amount ΔE is transferred from E_k to E_j around this optimum condition we must have

$$\frac{\Delta F_p}{F_p} = 0 = C_{\alpha_j}^{F_p} \frac{\Delta E}{E_j} - C_{\alpha_k}^{F_p} \cdot \frac{\Delta E}{E_k}$$

So that, when optimal allocation exists, we have for any two enzymes j and k

$$E_j : E_k = C_{\alpha_j}^{F_p} : C_{\alpha_k}^{F_p} \quad (3.53)$$

What this pattern of coefficients and enzyme concentrations actually turns out to be will of course depend on the details of the separate enzymic proteins and the network in which they are embedded. If one of them, for example, has a very low turnover number it will, even at its optimal allocation, require a large amount of protein and have a large coefficient. The following example will make this clear.

23. Straight chain of linear enzymes

Consider the chain of enzymes shown below, where X is the initial substrate and the final reaction is irreversible. The activity, of the separate enzymes is now given as the product, $E_i \cdot t_i$, of enzyme quantity and turnover number



Provided the pools are scaled so that all equilibrium constants are effectively unity the relations (1.18) and (1.19) of CH.I giving the flux and the $C_{\alpha_1}^F$ can be written as follows. Thus

$$F = \frac{X}{\sum \frac{1}{V_i}} = \frac{X}{\sum \frac{1}{E_i \cdot t_i}} \quad (3.54)$$

and

$$C_{\alpha_i}^F = a \cdot \frac{1}{V_i} = a \cdot \frac{1}{E_i \cdot t_i} \quad (3.55)$$

Where a is the same constant for all enzymes.

We can now apply the relation (3.53) in order to discover the optimal allocation of enzymes which will maximise F for a given total enzyme E_T and ensure the maximum growth rate.

Thus (3.53) can be written as

$$E_i = b \cdot C_{\alpha_1}^F \text{ where } b \text{ is the same constant for all enzymes}$$

and combined with the expression (3.55) for the coefficients this gives

$$E_i = b \cdot C_{\alpha_i}^F = a \cdot b \cdot \frac{1}{E_i t_i}$$

or

$$E_i \propto \frac{1}{\sqrt{t_i}}$$

Hence we have that the optimal allocation of enzymes in the chain is such that

$$E_1 : E_2 : E_3 : \dots = \frac{1}{\sqrt{t_1}} : \frac{1}{\sqrt{t_2}} : \frac{1}{\sqrt{t_3}} : \dots \quad (3.56)$$

To find the corresponding maximum growth rate we must use this enzyme ratio in (3.54) to give F_{\max} for a given E_T . Thus from (3.56) we have

$$E_i = E_T \frac{\frac{1}{\sqrt{t_i}}}{\sum \frac{1}{\sqrt{t_i}}}$$

and $\frac{1}{E_i \cdot t_i} = \frac{\sum \frac{1}{\sqrt{t_i}}}{E_T \sqrt{t_i}}$. Putting this into (3.54) we now get

$$F_{\max} = \frac{X \cdot E_T}{\left(\sum \frac{1}{t_i} \right)^2}$$

So that the maximum growth rate is

$$G_{\max} = \frac{F_{\max}}{E_T} = \frac{X}{\left(\sum \frac{1}{\sqrt{t_i}} \right)^2} \quad (3.57)$$

In particular suppose that the chain consisted of three enzymes having turnover numbers of 1, 4 and 9 respectively and that $X = 1$ and $E_T = 9$. If the enzymes are allocated equally the growth rate is given by calculating $\frac{F}{E_T}$ using (3.54) giving

$$G = \frac{1}{9} \cdot \frac{1}{\frac{1}{3} + \frac{1}{3 \times 4} + \frac{1}{3 \times 9}} = 0.25$$

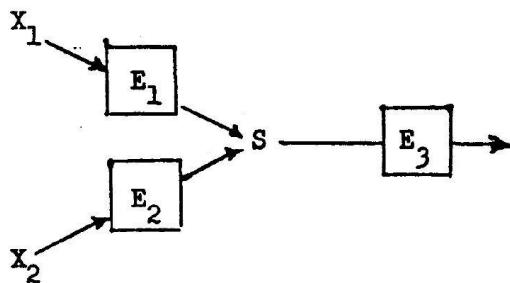
on the other hand G_{\max} for optimal allocation is given by (3.57) as

$$G_{\max} \simeq \frac{1}{\left(1 + \frac{1}{2} + \frac{1}{3}\right)^2} = 0.3$$

In this case optimal allocation of the enzyme has only improved the growth rate by approximately 20% from that obtained with an equal allocation.

24. Optimal solution given alternative food sources

Consider the three linear enzymes shown below.



In this case, if we consider F_3 as the flux to protein, the problem is to find the best allocation between E_1 , E_2 and E_3 . It can be shown by a simple argument that the maximum must be one or other of the solutions in which either E_1 or E_2 is zero. Using the previous example this gives the maximum as being the greater of the two expressions

$$\frac{X_1}{\left(\frac{1}{t_1} + \frac{1}{t_3}\right)^2} \quad \text{and} \quad \frac{X_2}{\left(\frac{1}{t_2} + \frac{1}{t_3}\right)^2}$$

This means that, depending on the ratio of X_1/X_2 , the optimal organism of this type will choose to feed exclusively on one or other of the alternative nutrients but never on both.

Role of regulation

It seems clear that for organisms such as *E.coli* where growth rate is of great selective value, an important role for the regulatory loops found may be to adjust the enzyme allocation so that it approximates to optimal over a variety of environmental situations.

In computer studies it will be of interest to introduce suitable regulatory functions for the g_i , which can respond to both X and S values, and to compare the allocations achieved by such systems with the computed optimal allocations over a range of environmental situations. Such studies could give us some quantitative idea of the value, from the point of view of maintaining a high growth rate, of the various regulatory mechanisms found.

4

Computer Methods For Studying The Steady State Behaviour Of Theoretical

1. Multi-Enzyme systems

In this chapter a brief account will be given of computer methods which have been developed in order to facilitate the study by numerical solution of the otherwise intractable sets of simultaneous non-linear algebraic equations which define the stationary solutions of a multi-enzyme system.

In developing suitable methods two general aspects of such a study were considered particularly important. Firstly that the type of study which we have in mind is always experimental in nature, being in fact an experiment ‘in numero’ on a theoretical system rather than, say, an experiment ‘in vitro’ on a reconstructed system. Because of this it is important that the computer methods used should allow the experimenter great flexibility in varying the values of parameters, in altering the actual structure of the system under consideration and in deciding what ‘measures’ are to be made on the system. Ideally this flexibility will be achieved by having the experimenter ‘on line’ to the computer and able to communicate with it in some sort of ‘biochemical language’ which is simple to learn and use, and will allow him both to specify the metabolic system and to subsequently perform experiments on it. In practice any computer methods developed should be capable of evolving towards this ideal as ‘on-line’ facilities become widely available.

The second aspect of importance is that the demands which such ‘in numero’ experiments place on the available computer resources should not be so large that the user obtains only a poor response from the computing service. For our own purposes it was thought that

the computer methods chosen should be able to perform a significant set of in numero experiments, say 100 S.S. and sensitivity matrix determinations on a medium sized system of 20 non linear enzymes, within a resource allocation which could be available to an average user on at least a daily basis., Keeping these requirements in mind the first decision is whether to employ an analogue or a digital computer.

2. The Analogue computer

For simulation work the analogue computer has two main advantages. Firstly that the analogue components, when suitably interconnected for a particular study, correspond directly to the 'blocks' of the system being simulated and secondly that voltages on the components are directly representative of system variables. Secondly there is the ability to 'interact' directly with the system by, for example, adjusting parameter(s) which correspond to potentiometer settings and then observing the response of system variables, often by direct oscilloscope display.

These advantages largely meet the requirement to be 'online' and to have a simple means of carrying out, experiments on the system. However, the task of specifying the system in a form suitable to the analogue computer may still be formidable.

On the other hand the analogue computer has numerous disadvantages. Thus the size and non-linearity of, systems which can be simulated on any given computer is usually severely limited by the number of analogue components available and, furthermore, to avoid serious inaccuracy, system variables have to be 'amplitude scaled' so that voltages lie within a limited range. Often repeated 'rescaling' of a problem is necessary as an experiment is conducted. In addition there is always difficulty in repeating 'runs' since analogue components are constantly developing minor faults.

For our own requirements the disadvantages just noted considerably outweigh the possible advantages, Thus ruling out the use of analogue computers for all but the smallest problems. For example, even with the quite large analogue computer available to us we would have been limited to the equivalent of about 7 nonlinear unimolecular enzymes with the further limitation that it would not have been easy to produce conveniently even an approximation to a sensitivity matrix.

3. Digital computer systems

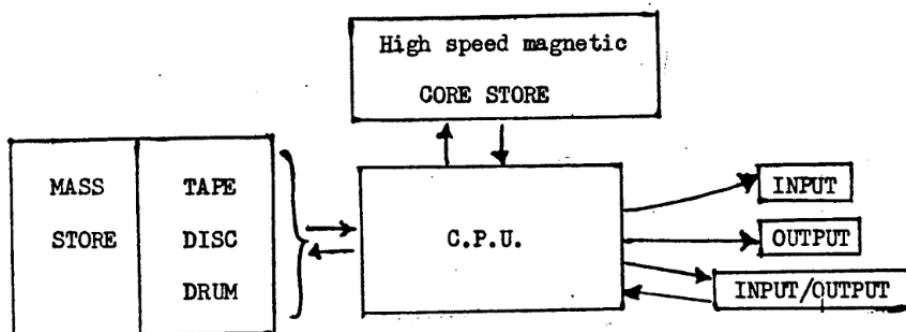
By comparison the digital computer, if it is working at all, is highly accurate and completely reliable, usually being maintained by a service organisation. Furthermore there is, within reason, no limit to the size and non-linearity of the system being simulated except that the larger metabolic systems will require more time. In practice a further important advantage of digital machines is that, for mainly commercial reasons, far more effort is being devoted

to * their development than is to that of analogue machines, so that each year they become more powerful and more widely available.

In recent years a large number of 'simulation languages' have been written in which the general aim is to retain the undoubted advantages of the analogue computer whilst actually using a digital by machine, a useful review of some of these languages is by R.D. Brennan, (1964). In general these languages have been successful in allowing the user to work in terms of the 'block' structure of his problem and in allowing him to easily load the structure and initialise the values of parameters. However, they have been less successful in allowing the user to 'interact' with his system, since the proper development of such interaction must await the general availability of 'conversational' operating systems at computing centres.

In order to discuss the particular methods which have here been used in devising a suitable 'simulation language' for the steady state properties of multi-enzyme systems. it will be necessary to give a brief description of certain concepts and terms commonly employed in digital computer science.

A computer system is indicated in the diagram below in which all details other than those required for a general understanding have been omitted



The C.P.U. or 'Central Processing Unit' is capable of reading binary stored patterns stored in locations within the high speed magnetic core store, and interpreting these patterns as 'instructions' which will control the next action of the C.P.U. When an instruction has been completed the C.P.U. will automatically fetch the next one from a designated location.

A modern machine will be capable of distinguishing and executing about one hundred distinct binary patterns as 'legitimate' instructions each of which causes the C.P.U. to carry out a specific basic operation. These include 'data transfer' instructions which move binary patterns, now considered as data, between registers in the C.P.U. and other registers, either in the core store or associated with various input, output, or mass storage devices. Input devices include teletypes, card and paper punches, graph plotters, line printers and mass storage, includes magnetic tape, magnetic disc, magnetic drum etc. Other 'data manipulation' instructions are concerned with carrying out operations on designated bit patterns, usually held in one or more registers belonging to the C.P.U., these include for example, arithmetic operations such as multiply and divide, logical operations such as 'AND'ing two

bit patterns etc. Usually the C.P.U. will fetch and execute instructions which lie sequentially in core, but certain instructions can determine from which core location the next instruction is to be fetched, so that, for example, it is possible to repeat a sequence of instructions until some condition is satisfied before continuing with further instructions.

A set of such instructions in core is called a 'binary program' and, starting from a given instruction, will cause the C.P.U. to carry out a complete sequence of operations. For example it might read in a set of numbers, previously punched on paper tape and fed to a tape reader, store them as 'data' in core, proceed to find their mean value and standard deviation and output the results on a line printer before returning to the start of the program ready to read in further input from the tape reader.

A modern computing system will provide a number of special 'system programs' which greatly simplify the task of the user and also help to ensure that the computer is being efficiently used.

One type of system program is an 'operating system' or 'O.S.' which has the important task of supervising the activities of all other programs (sequences of instruction in core, or on backing store) which are current. Thus if there are several programs in core at any one time a 'multiprogramming' O.S. will allocate C.P.U. time to each in turn and will ensure that their input and output is connected with the appropriate peripherals. Or again the O.S. may maintain a 'queue' of programs on a disc, so that the C.P.U. never runs out of work, and arrange for programs to be moved from and to backing store so that valuable core space is not wasted on inactive programs. An important goal for advanced O.S. is to allow a user to work from his own input/output station, just as if the computer were dedicated to him alone, this effect being achieved by the O.S. 'time sharing' the C.P.U. and other system resources amongst users active at any moment. This will eventually be important for simulation work (D. Garfinkel et al., 1970) since it will then be possible to 'interact' with the system under study. Unfortunately such 'conversational' O.S. have not proved easy to implement in any efficient way and are only now becoming available to the ordinary user in the U.K.

Another type of system program is a 'compiler' which has the task of translating a program written in a 'high level' source language into an equivalent binary program suitable for running on a particular machine. This relieves the user of the impossibly tedious task of breaking his problem down to the level of a series of binary coded machine instructions and instead allows him, to concentrate on describing his problem in one of the more or less readable high level languages. These include the well known 'ALGOL' (ALGOrithmic Language) and 'FORTRAN' (FORmula TRANslation) languages: compilers are available on most large machines to translate source programs, written in these languages into the equivalent binary code.

With the reasonably advanced O.S. presently available the user can call, by means of 'control statements' which inform the O.S. of his wishes, any stored binary program such as a compiler and feed data to it from any specified input. This allows him for example to call further binary programs: which will receive as data the output of previous programs. If the O.S. provides conversational facilities he will in addition be able to interact with appro-

priately written programs. For example with a conversational 'editor program' or perhaps with his own suitably written simulation program.

4. Design of suitable simulation system

The actual implementation of a simulation system which broadly meets the requirements set forth at the beginning of this Chapter will depend very much on the general facilities provided by the local computing centre. In our own case the situation was that the centre provided a reasonably advanced but non conversational O.S. together with the 'ALGOL' type of language compiler considered capable of producing nearly optimal binary code. With regard to 'computing power' rough estimates indicated that the 'KDF 9' computer, available at the time under this system, would, if carefully programmed, take about 10 minutes to complete the daily simulation work. Neither the limited basic computing power available to us nor the lack of a 'conversational' mode was thought likely to change appreciably for some years.

In these circumstances the basic decision was taken that the 'structural input' for a simulation, that is the data which specifies the rate expression and how they are interconnected to give the functions, $\phi_i(S, P)$, would have to be converted by the available.

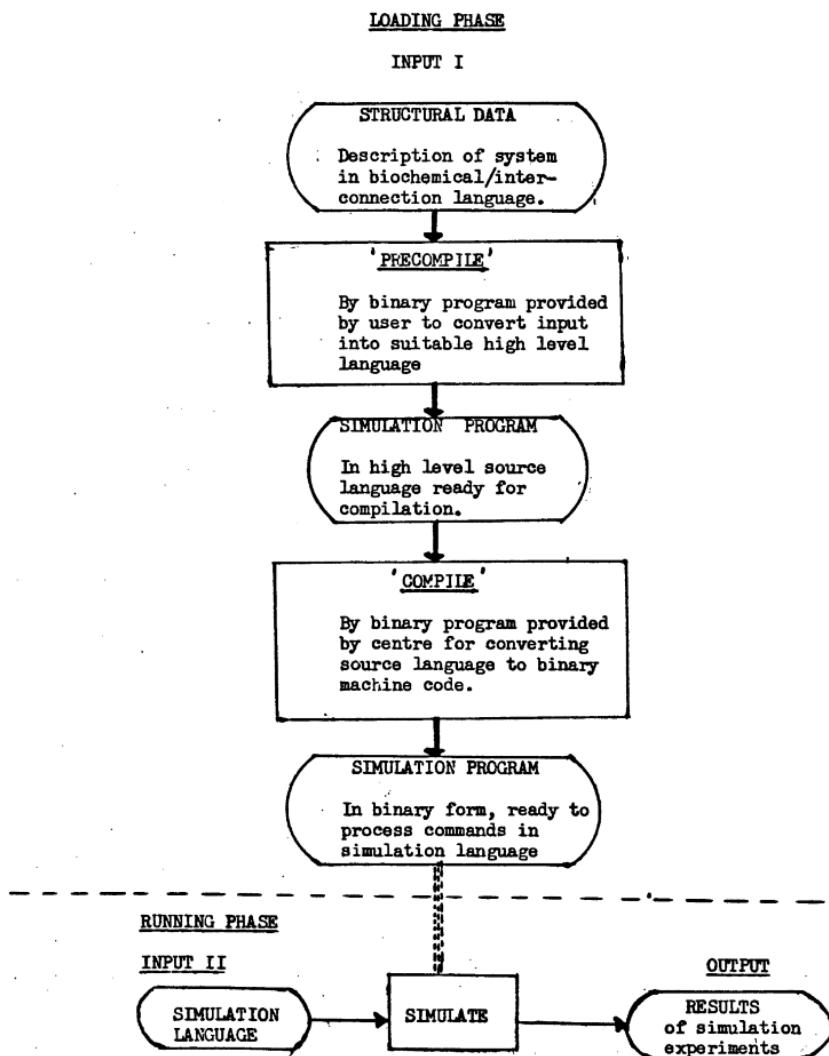
compiler into efficient binary code for evaluating the ϕ_i . This was thought necessary because the 'heart' of a simulation involves repeated numerical evaluation of the ϕ_i and any gain in efficiency at this level will immediately be reflected in the performance of the simulation program. A number of earlier simulation programs have not attempted to produce efficient code in this way and were correspondingly slow, thus limiting their usefulness. For example the earlier programs of Garfinkel were of this so called interpretative type (D. Garfinkel et al., 1970), which in effect meant that each evaluation of the ϕ_1 involved the unnecessary work of rescanning a stored interconnection matrix, a step which in our method would be carried out only once at the compiling phase. G.H. BURGIN, (1966), who converted his originally interpretative 'MIDAS' program so as to include a compilation phase claims an increase in 'speed' of between 5 and 10 fold over the old version.

Since the 'structural input' would not conveniently be in a form suitable for compilation it was decided to feed it as data to a program written by ourselves and designated the 'precompiler'. The purpose of this 'precompiler' was to convert the incoming structural data into a high level language and combine it with existing code to form a 'simulation program' suitable for passing on to the compiler. It was further decided that this simulation program should, after compilation, be itself capable of receiving 'commands' in a simulation language so that the user would not lose any of the advantages of purely interpretative simulation systems which usually provide a flexible means of controlling the actual simulation "running phase".

The use of a 'precompiler' phase to construct and present for compilation a simulation program which can itself be driven by its own simulation language is shown in the figure

opposite, and this general design has several results. Firstly it avoids the 'inefficiency' disadvantages of a purely interpretative system since it ensures that basic and time consuming operations such as evaluation of the ϕ_i , solution of differential equations and inversion of matrices will be in the form of nearly optimal binary code. Secondly it separates simulation into the two distinct phases of 'loading' and 'running' as indicated in the figure. This separation has the advantage that once a structure has been loaded it may be repeatedly investigated by supplying data, in the form of commands in the simulation language, to a stored binary version of the simulation program there being no need to return to the precompiler-compiler level unless it becomes necessary to alter the structure. Thirdly, it will be possible, depending on the sophistication of the precompiler and simulation program, for a user who is unfamiliar with programming to load his enzyme network using a notation which involves only simple algebraic and biochemical concepts and run his experiments by means of a series of simple 'commands' interspersed with the necessary data. This will have the important effect of putting the power of the computing system directly into the hands of the investigator concerned. A further advantage in this respect is that although the main features of the simulation 'command' language are fixed it becomes a simple matter using the precompiler to add new commands as and when they are needed.

figure (3)



Finally the production of the simulation program via a precompiler stage and its command language running technique are well suited for its efficient operation under conversational O.S. as they become available and little reprogramming will be necessary. In this respect the precompiler stage should result in a simulation program of high efficiency and with minimum demands on core storage Both of which are essential attributes if a reasonable response is to be obtained from the conversational O.s. of the near future which will impose a heavy penalty on programs making large demands on time and core. The advantages of a 'fast' simulation program are also very relevant to the 'archive' problem. Thus the experiments to which we have referred consist in moving through 'parameter-space' and

observing the response of all system measures of interest. If the computation of the S.S. were very expensive in computer time it would become necessary to file or 'archive' the results of experiments in case they were needed again for comparison with the results of later experiments as indeed they often would be. Such a strategy is, however, hardly practical on account of the difficulty of devising a filing system which could cope with the continuously expanding exploration of a virtually infinite parameter space, both from the point of view of classification and because of the sheer bulk of information. If, however, the computation of the S.S. is sufficiently fast no such problem need arise since it will be possible to recompute the same S.S. whenever it is required for purposes of comparison.

5. Implementation on the KDF 9

A simulation system of the general design just outlined was implemented on the Edinburgh University KDF 9 computer during 1968, several earlier versions having been discarded, and demonstrated at the FEBS summer school on 'Computing Techniques in Biochemistry' in the same year. General impressions of my own and other, simulation programs are set forth by J.G. Reich, 1969.

In view of the impending removal of the KDF9 and also of proposed modifications to the high level language it was decided to keep the implementation as simple as possible within the general framework of having a 'precompiler' and a 'simulation program'. With this in mind the 'biochemical language' was, for the meantime, taken as an algebraic specification of the rate-expression and of their use in forming the ϕ_i whilst the simulation language was restricted to being entirely numerical, with -ve numbers representing the various commands and +ve numbers the corresponding data. Both of these restrictions could of course be lifted in future versions of the simulation system, their purpose being merely to facilitate the first version. Thus the restricted biochemical language was suitable for direct compilation by the available compiler (Atlas Autocode). Whilst the precompiler, having no need to transform the incoming data, assumed more the form of an EDITOR which would edit the incoming text into the text of a previously filed simulation program. The editing aspects of the precompiler in fact proved so useful to myself and others that a 'public' version, produced in collaboration with Dr. J.G. Burns of Edinburgh University, was made available to other users of the computing centre. A users guide to this is provided in the appendix.

6. Editor (precompiler)

This will not be described in detail since it is marginal to the enzyme simulation problem however its main features will be outlined since any future precompiler would almost certainly include them. The user operated the editor by means of issuing 'commands' which were followed, as appropriate, by the names of text files to be manipulated, line no's to be edited within the files and sections of text to be inserted. The basic operation of the editor

was to take the text from a magnetic tape file, place it in core and edit it in accordance with the commands 'DELETE', 'INSERT' and 'ALTER'. The resulting text could be passed to the compiler, by means of the command 'RUN', for conversion into an equivalent binary program whilst the old version of the text on magnetic tape would automatically be overwritten by the new one unless the command 'NOWRIT' was used. In addition text could be moved from one file to another by means of the command 'BRANCH', inserted direct into a file from an input device by means of the command 'FILE' or stored and retrieved from a 'library' by means of special library commands. The editing system in its final form worked off one magnetic tape and maintained six text files each capable of holding a large program as well as a sizeable library.

In retrospect the editor turned out to be an essential part of the complete simulation system. Thus whilst enabling the non-programmer to 'load' the system using a standard sequence of editing instructions it also allowed the more sophisticated user to carry out the following operations.

- i) Retain 'on file' the text of all 'networks' of interest. So that they could be modified when required, using the editor, and the new simulation program automatically established.
- ii) Extend the simulation program itself by adding new routines and by developing special versions to study particular problems.
- iii) Correct faults which from time to time were discovered in the simulation program.

It may be argued that the functions of the editor amounted to no more than 'shuffling cards' or, as was the case at that time, editing paper tape. In practice, however, the ability to store, edit and list several large programs places the user in an altogether different position and makes feasible the type of activity which is essential to simulation. It turned out also that the programming techniques involved in writing the editor were very similar to those required for a simulation program capable of receiving a simple command language. In each case it is necessary to read in commands, look them up in a dictionary and then, using a 'switch', select an appropriate sequence of routine calls which will carry out the command returning to read in the next command when execution is complete.

7. The simulation language

As mentioned earlier this was kept as simple as possible, being entirely numeric, both to reduce the size and complexity of the simulation program and also to keep the 'commands' as concise as was possible. The integer part of each -ve number in the input stream will select a command and the first two decimal digits the particular 'type' of that command if alternatives exist. A table of 'commands' is shown opposite and an example of their use will be given in the next section. The main features of this table are as follows.

I. Selection of network

If several networks have been previously loaded and the binary version of their corresponding simulation programs stored on magnetic tape it is possible to select among them using the command of type '-0.XX'.

For example the command '-0.02' means that the simulation program for a particular network which was previously compiled and filed in binary form in file number 2 is to receive any subsequent commands. If after some time it wished, say, to investigate the network held in file number 1 it is necessary for operations in file 2 to be brought to a close by means of the command '-0.99' and the new system can be selected by means of the command '-0.01'.

II. Parameter modification

Parameters embedded within the expressions defining the ϕ_i can be altered in two ways, either by multiplying them by a specified factor or by replacing them with a new value.

Multiplication and replacement are brought about respectively by the commands '-3.XX' and '-4.XX' where '.XX' specifies the particular type of parameter being modified as indicated in the table. For example the command '-3.02' indicates that we wish to alter the quantity of an enzyme by multiplying its present quantity by some factor. Commands of this type must of course be accompanied by the appropriate 'data' in the form of a set of +ve numbers preceding the command and provision is made for the user to provide as many pairs of numbers as required each pair specifying the enzyme to be changed and the amount by which it is to be multiplied respectively. For example the line

11 10 4 5.2 - 4.02

causes the operations $e_1 = e_1 \times 10$ and $e_4 = e_4 \times 5.2$ to be carried out, where e_1 and e_4 represent enzyme quantities.

III. Output of the values of parameters and variables and printing of 'messages'

Output of values can be achieved by means of the command '-5.XX' where the 'X' again selects the parameter or variable type which it is required to output. In this case the command always outputs the values of all members of a given parameter type and there is no need for precedent data. For example if there were 10 pools in a given system the command '5.06' would cause the values of S_1, S_2, \dots, S_{10} to be printed out, in that order. For convenience the command '-5.00' causes a complete output, being equivalent to '-5.01', '-5.02' ... '-5.08'. In order to improve the legibility of the output the command '-7' can be followed by any 'message', the effect being cancelled when the end of the line is reached.

PRECEDENT DATA +VE NO	COMMAND CODE -VE NO:	TYPE .XX	FUNCTION
NONE	-0.XX 1 XX 15	FILES 1,...,15	START FILE
NONE	-0.99	NONE	FINISH
NONE	-2	NONE	RECOVER DATA
NONE	-1	NONE	SHUNT DATA
1 XX 7 SUFFIX/MOD/SUFFIX/MOD/.... up to 24 pairs XX = 8 INHIBITION I SWITCH ON, OFF, NEUTRAL 1 3 2 INH NO/SWITCH VALUE	-3.XX 1 XX 8	00 ALL 01 P 02 E 03 T 04 K 05 X	MULTIPLY by 'MOD'
II CHANGES FUNCTION INH NO/SWITCH/CN/SL/LOW/HIGH VALUE	-4.XX 1 XX 8	06 S 07 TIME 08 INH	OVERWRITE by 'MOD'
NONE	-5.XX 0 XX 8		OUTPUT
XX = 0 NONE 0 XX 5 SUFFIX	-6.XX 0 XX 5		SELECT COEFFS
'-7' is followed by a 'message' terminating at the end of the line	-7	NONE	WRITE MESSAGE
1 STORE DESIGNATED 6	-8.XX 1 XX 2 -9.XX 1 XX 2	01 E 02 S	STORE UNSTORE
I ALL FLUXES/ALL POOLS/ II NONE INTERNAL FLUXES POOLS USED	-10	NONE	FLOAT Adjusts e to SS with given fluxes, pool
NONE	-11	NONE	MOVE to S.S.
KEY WORD/STEPS UP/STEPS MOD/DOWN/STEPS MOD (PARTS OF DIGITS TAKEN AS SUFFIXES.)	-12.XX 1 XX 3	01 E 02 P 03 X	SPREAD ON XX

SPECIAL COMMANDS -20 BELONG TO SOME FILES AND ARE DOCUMENTED WITH THEM

IV. Selection of sensitivity coefficients for subsequent output.

Commands of the type '-6.XX' will select which parameters the coefficients are to be found with respect to. Thus whenever a S.S. is subsequently computed only selected rows of the generalised sensitivity matrix $\left[C_{P_j}^{S_i, F_j, M_K} \right]$ will be outputted. In this case, the '.XX' specifies the type of parameter, P , X , etc, and a list of the +ve numbers preceding the command specifies which members of the type vector, X_1, X_2 etc, are required. For example the two command lines,

```
1 5 7 -6.01
4           -6.05
```

will select for subsequent output the rows with respect to P_1, P_5, P_7 and X_4 respectively. The command -6.00 will select all rows of type $C_{\alpha i}$ i.e. the block coefficients, and requires no precedent data.

V. Storage and Retrieval of Values

Often it is convenient, after a sequence of parameter variations for example, to, return to an earlier set of parameter and variable values. There are several commands which are of use in this situation. Thus '-1' will 'shunt' the existing values of all parameters, variables, matrix elements etc. onto magnetic tape and '-2' will 'recover' them. The commands '-8.XX' and '-9.XX' perform a similar function but are more limited, the '.XX' referring to manipulation of either the pool or the enzyme quantity vectors by themselves. For example the command '-8.01' will store the existing enzyme values and the command '-9.01' issued subsequently will then overwrite any new enzyme vector which may have arisen with the original one which was stored.

VI. Computation and output of S.S.

A number of commands are concerned with causing S.S. calculations to be performed and also with outputting the results. Thus '-1', which can be called after any parameter adjustments, will cause variables to move to their S.S. values and will automatically output all pools, fluxes, special measures and any coefficients which have been selected.

If it is required to compute and output a sequence of S.S. solutions corresponding to a steadily increasing value of some parameter this can be achieved concisely by means of the 'SPREAD' command of type '-12.XX'. For example the line

```
'1010203      3      2      3      -12.03'
```

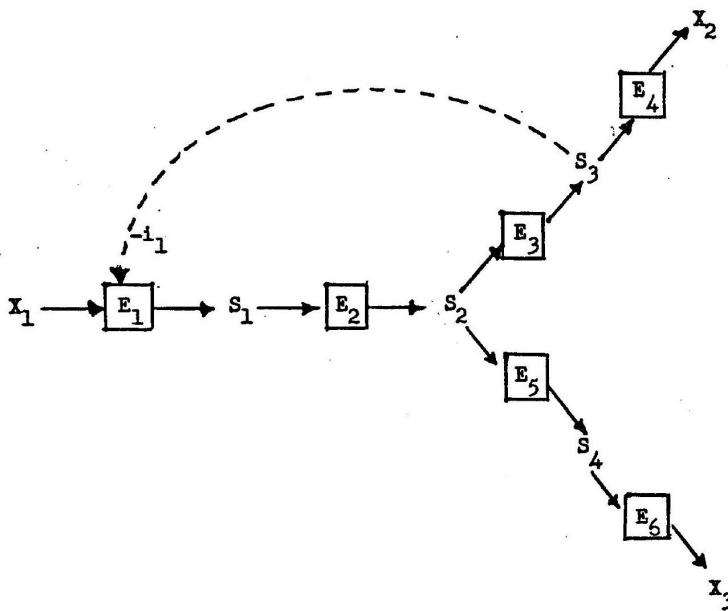
will cause X_3, X_2 and X_1 specified in the leading 'KEYWORD', each to be modulated up and down by three powers of 2, thus computing 9 S.S. around the original position.

VII. Adjustment of enzymes to achieve a given S.S.

A further useful command is '-10' or 'FLOAT' which, when supplied with precedent data for all pools and 'independent fluxes' will adjust or 'float' the values of the enzymes carrying 'dependent' fluxes so that a perfect steady state is achieved corresponding to these specified pools and independent fluxes. Since one may often wish to model a situation where relative fluxes and pools are roughly known this command has proved useful.

8. Loading and running a system

To make clear the steps which are involved in using the simulation system we will consider the six enzyme divided pathway network shown below:



The network involves a metabolic inhibition i_1 , and as an example, is supposed to be a subsystem in an expanding space which has an exponential growth constant P_1

9. Loading

Below are set out the main stages involved in loading the system via the precompiler. The notation used is similar, but not identical, to that actually used on the computer, and corresponds to the notation of CM.I and II. The input is typed onto paper tape or cards or,

if an on-line system is available, perhaps direct into the computer. The precompiler stage can detect certain ‘obvious’ errors in the input and ‘abort’ the loading phase, if these are sufficiently severe.

10. Details of step

STEP DETAILS OF STEP

- A Provide correct control statements to activate the EDITOR/ pre-compiler and find the relevant text for the required simulation program from among the editors files.

- B Type out the six flux expressions giving the flux at each of the six blocks in terms of the surrounding pools etc. Thus

$$F_1 = E_1 t_1 (X_1 - S_1/K_1) \times i (1, S_3)$$

$$F_2 = E_2 t_2 (S_1 - S_2/K_2) / (1 + S_1/P_2 + S_2/P_3)$$

etc

- C Show how the 'Fs' combine to produce the pool rates of change, including the effects of expansion. Thus

$$\phi_1 = F_1 - F_2 - P_1 S_1$$

$$\phi_2 = F_2 - F_3 - P_1 S_2$$

etc

- D Taking the two independent fluxes as the outputs $F_4 F_6$ show how the dependant fluxes are related to them

$$F_3 = F_4 + P_1 S_3$$

$$F_5 = F_6 + P_1 S_4$$

$$F_2 = F_3 + F_5 + P_1 S_2$$

etc

- E Write down expressions defining any special 'measures' which are currently of interest. For example the ratio of fluxes in the two pathways

$$M_1 = \frac{F^4}{F_6}$$

- F Assign values to all parameters and I.C. to pool variables any not assigned will automatically have value unity. Indicate which binary file should hold the compiled version of the simulation program

$$P_1 = 0.001$$

$$P_2 = 10$$

- G Provide a title for the network, which will be automatically printed out whenever it is run. Thus 'SIX ENZYME DIVIDED RATHWAY WITH INHIBITION'

In principle sections *C* and *D* could be omitted since they duplicate information already contained with the rate expressions of *B*.

However they do not represent a large amount of input and they have been included rather than employing a more complex form of precompiler. The representation of inhibition is simplified by means of a standard inhibition routine $i(S, n)$ available within the standard simulation program. Any number of functions can be set up in this way, each being defined by four parameters which control its' slope, maximum inhibition, etc. and which can themselves be modified via the simulation language.

When all sections have been loaded the precompiler incorporates them into the text of the standard simulation program held in its files and feeds the resulting complete program to the compiler. If the program compiles successfully, that is has no obvious errors in it, the resulting binary version will be placed into the designated binary file where it can be repeatedly accessed for use in the 'run' phase by the appropriate -0 .XX command.

11. Running

The system just loaded can now be accessed for experimentation indefinitely unless it is required to alter its structure or perhaps add some extra 'commands'. The following sequence of commands exemplifies one such experiment.

	-0.01	Enter the simulation program contained in 'binary file 1' into the computer.
-7	'EFFECT OF INHIBITION AT DIFFERENT X VALUES'	
	-11	Write suitable message onto output
1 2	-4.08	Find S.S. and output pools, coefficients etc.
	-11	Switch the inhibition function i_1 to 'neutral'
1 1	-4.08	Recalculate S.S., which will be same, and rate coefficient alteration
1 2	-3.05	Switch 'on' inhibition again
	-11	Multiply X_1 by factor of two
1 2	-4.08	Find S.S.
	-11	Again switch inhibition to neutral
.	.	Find S.S.
.	.	.
.	.	and so on
	-0.99	Close simulation run

The above experiment is intended to see what effect the presence of the inhibition i_1 has on the pattern of sensitivity coefficients for increasing levels of X_1

12. The simulation program

Here it seems useful to outline certain aspects of the program these being concerned broadly with the 'numerical' techniques for locating the S.S. and the programming techniques for interpreting the simulation language. The overall program consists in declaring 'arrays' for all the variable types involved, such as E, t and p, the size of these arrays being filled in by the precompiler at 'load' time. Space is also declared for the matrix $\left[\frac{d\phi_i}{dS_j} \right]$ which is required. for the iteration procedure mentioned in CH. I and also for computation connected with the sensitivity coefficients, in addition a large 'switch' is declared for use in selecting between the different commands. Following these declarations are a large number of routines each of which carries out a specified operation on the global 'arrays' or is perhaps concerned with reading in and interpreting the simulation 'language'.

13. Working section

The actual 'working' part of the program is quite short and readable provided the routines which it uses have been given names which clearly indicate their function. It consists essentially of a multiple 'switch' which is selected on the basis of the last command. read in and is represented in the section below.

```

1:      READIN (LIST, COMMAND, TYPE)
        → SWITCH (COMMAND)
        SWITCH (1): Code, mainly routine calls which
                     carries out the operation corres-
                     ponding to command number 1
        → 1
        SWITCH (2): Code for command number 2
        → 1
        SWITCH (3):
        .
        .
        .

```

Let us consider the operation of the above assuming, for example, that the first line of the simulation language is '1 10 4 5 -4.02'

Starting at the 1st line, labelled (1:) the routine 'READIN' fills the 'array' LIST with the positive data 1, 10, 4, 5, sets the integer 'COMMAND' to 4 and the integer 'TYPE' to 2. On the next line the 'SWITCH' instruction directed by the value of 'COMMAND', will cause the program to jump to the line labelled SWITCH (4): which leads to a series of instructions designed to modify the specified quantities multiplicatively using the information contained in 'LIST' and 'TYPE'. When this operation is complete the instruction ' 1' is encountered and the program is ready to return to the routine 'READIN' and obey the next command. The simulation program will thus read and obey any sequence of instructions until the input ceases. It can easily be seen, now, how the precompiler can 'pick up' extra commands, if they are included with a network description at 'load time', and attach them opposite the switch labels 20: and upwards which are left vacant in the standard simulator. Another point is that since each command expects to receive a certain data format it is a simple matter to check the contents of 'LIST' and 'TYPE' to see if they are legitimate. Such error detection facilities were extensively included in the KDF9 implementation and proved to be a valuable feature.

14. Numerical aspects of locating the s.s. and estimating sensitivity coefficients

One general way of getting into the neighbourhood of a steady state solution \bar{S}_j is to integrate the differential equations of motion (1.1), thus simulating the natural movement of the S_j . However, a criterion for when to stop, or indeed certainty as to whether one is in the region of a steady state at all, are not simple matters.

The criterion of closeness to the desired steady state is as follows; after every half hour of simulation (metabolic system time) the values of $(N - n)$ of the F_1 fluxes are accepted and the values of F' which the remaining n must have in order to satisfy equation (1.1) with all $\dot{S}_j = 0$ are calculated. In general these will be different from their existing values and

$$= \sum \frac{F' - F}{F}$$

is a measure of the distance of a constructed exact steady state, with slightly altered enzyme concentrations from the desired exact steady state with the prescribed enzyme concentrations. When the rate equations can be expressed in the form (1a) measures the sum of the fractional enzyme changes needed to move from the constructed to the prescribed steady state. Thus, if 0.001. we can say that we have found a mathematically exact steady state such that no enzyme differs from its prescribed value by more than 1 part in 1000.

Simulation, which is costly, is only employed while 0.20. As soon as 0.20 the matrix \bar{S}_j/E_i is formed, by numerical approximation, for the constructed steady state and is used to estimate the \bar{S}_j obtaining at the correct enzyme values. This is repeated until is reduced to the required accuracy. This process, similar to Newton-Raphson iteration, converges rapidly for all systems so far tried and it is not necessary to recompute \bar{S}_j/E_i very frequently. The sensitivity coefficients C_y^x can all be found from the matrix without appreciable further computation.

The time to recompute the \bar{S}_j/E_i matrix as we move about the parameter space is not great, compared with the time spent on simulation, for systems up to 30 enzymes. Also the necessary matrix inversion is numerically well-behaved provided that we already have a reasonable estimate for \bar{S}_j/E_i , which will be the case if parameter movements are not too violent. Initial estimates of both \bar{S}_j and \bar{S}_j/E_i are more costly, being formed by simulation. They are computed only once and kept available should the current values be lost. Clearly the integration method employed can have a low accuracy and still suffice. At present 4th order Rutta-Merson is used because it is convenient and reliable. For "stiff" systems considerable improvement would result from the employment of a more suitable integration method.

5

DOMINANCE EPISTASIS AND SELECTION

In any concerted attempts to describe the effects of changes in genetic parameters on some systemic measure of the whole system, the phenomena of dominance and epistasis becomes a problem which demands 'explanation'. Not only is it a problem to the 'Mendelian' geneticist insofar as he is concerned with the mechanism of gene action but it has been the concern of the 'quantitative' geneticist insofar as he is concerned with the selection and maximisation of some desirable character. It is immediately obvious that the treatment of biochemical systems, so far attempted, makes it in principle possible to reformulate the question in molecular terms. Given that we can formulate the dependance of some measure on the quantitatively specified parameters of a locus-specific enzyme, we can investigate the changes in the measure when substitutions at this locus occur.

As before, it is worth mentioning how far the relatively abstract treatment may be relevant to the phenomena well known in 'real' organisms. The characters of economic interest, such as growth rate, milk yield etc. cannot, of course, be specified precisely in biochemical terms. It is, however, generally agreed that the rate of synthesis of the various 'end products' is intimately related to the 'flux' through this different pathways leading to them. In these studies this flux was therefore taken as the 'character' since the kinetic formulation in terms of the constituent enzyme activities (and hence the genes) can be given.

Similarly dominance can be measured by observing the effect on this flux of successive substitutions at the locus. Or again epistasis, regarded as the strength of interaction between different loci, can be estimated by observing how the response of the flux to substitution at one locus is influenced by a simultaneous substitution at another. One result which will be

demonstrated is that a considerable degree of dominance and interaction almost invariably arises even when the simplest assumptions are made about the properties of the individual enzymes. With more complex (and more realistic) assumptions, such as saturation, divided pathways and feedbacks, these interactions are even stronger. This appears to arise merely from the fact; true for all organisms, that enzymes are always embedded within a metabolic network of some sort. This seems particularly interesting in the case of dominance since it suggests that the common occurrence of dominance is not in need of a theoretical explanation, (such as the “evolution of dominance”), arising as it does from the general nature of metabolism common to all organisms.

In addition an attempt will be made to understand the implications of this ‘gene-interaction’ for the response to selection on the flux, for a ‘population’ of such ‘organisms’.

Two aspects of such an investigation, however, require special treatment which, so far, it has not been necessary to consider in detail. The first is a method of handling a diploid situation, the second the fact that we shall, in general, be concerned not with differential changes but with discrete (and perhaps ‘large’) steps in the parameter space.

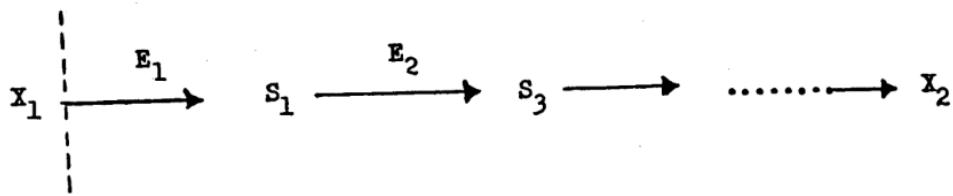
As a first attempt to include diploidy and discrete parameter changes within a theoretical treatment it was realised that an analytic study of a straight chain of simple ‘linear’ enzymes, when thps flux through the chain can be found algebraically, might be of considerable general interest. Such a system, whilst always an oversimplification, allows fully for genes, acting via enzymes which are embedded in a metabolic network, and therefore can be expected to possess the type of irreducible interaction which is common to all organisms. Consequently properties of such a system can be regarded as having a fairly general bearing whenever alleles at many loci contribute, within a given metabolic system, towards a character. Such a treatment will now be given.

1. Dominance and epistasis in a straight chain of enzymes

In order to discuss the response of the flux through the chain of enzymes to the various allelic substitutions it is necessary to define certain quantities. Firstly, there is the quantity of “allelic swing” or, λ , which measures the ratio between the high enzyme activity of the ‘good’ allele and the low activity of the ‘bad’ one at any particular locus. The allelic swing can, of course, have different values at different loci. Secondly, there is the quantity measuring the importance of a particular enzyme step in controlling the flux through the pathway namely the sensitivity coefficient.

Finally there is the “maximum response to selection”, R , measuring the ratio between the flux of the ‘best’ organism (all alleles good) and the worst (all alleles bad).

Consider now the straight chain system indicated below:



The diagram shows the conversion of a precursor at constant concentration X_1 into a product at constant concentration X_2 via a sequence of enzymatically catalysed steps $E_1, E_2 \dots$ etc. The concentrations of the metabolic intermediates being S_1, S_2, \dots etc. At the stationary state, when the intermediates have all stopped changing, the pathway will produce X_2 from X_1 at a steady rate, F , where F is the net-flux in the pathway.

In addition we must specify the enzyme activities corresponding to the two alleles, namely V^H and V^L . Furthermore the allelic state at any one locus is specified by a number.

$$\text{LOCUS 1} \quad \pi_1 = \begin{cases} \text{or} & +1 \\ \text{or} & 0 \\ \text{or} & -1 \end{cases} \quad \boxed{V_1^H} \quad \boxed{V_1^L} \quad \lambda_1$$

$$\text{LOCUS 2} \quad \pi_2 = \begin{cases} \text{or} & +1 \\ \text{or} & 0 \\ \text{or} & -1 \end{cases} \quad \boxed{V_2^H} \quad \boxed{V_2^L} \quad \lambda_2 \text{etc. } \dots$$

The set of numbers π_1, π_2 , etc. can each take one of three values. Thus, $\pi = -1$ corresponds to the homozygote having the low activity enzyme, $\pi = 0$ to the heterozygote and $\pi = 1$ to the high activity homozygote. The set of integers π_2, π_2 , etc. thus specify the genotype of the 'organism', and a finite population of organisms is specified when the π vector of each organism in the population is known.

The enzymes are assumed to convert substrate into product at a rate given by a linear expression of the form $f = V(A - B/K)$, where V is the activity of the enzyme, K , an equilibrium constant, and A and B the concentrations of its substrate and product. This corresponds to an unsaturated 'Michaelis' type enzyme, and, on the assumption that enzyme concentration is proportional to gene dosage, the heterozygous state at the i th locus will also be represented by a linear rate expression having an 'activity' V_i^* , midway between the low and high activities V_i^L and V_i^H . It can be shown, CH.I, that if the effective activities, V_i , at the separate steps are measured in appropriate units the flux F through the pathway is given by

$$F = (X_1 - X_2) / \sum \frac{1}{V_i} \quad (5.1)$$

The effective activity at the i th locus can now be written in terms of V_i^* , its heterozygous value, π_i and which indicates the allelic state at the locus and λ_i the allelic swing at the locus. Thus since λ_i is the ratio between the high and low activities at the i th locus we have

$$\lambda_i = \frac{V_i^H}{V_i^L} \text{ and } V_i^* = \frac{V_i^H + V_i^L}{2}$$

From this it follows that

$$V_i^H = V_i^* \left(1 + \frac{\lambda_i - 1}{\lambda_i + 1} \right)$$

and

$$V_i^L = V_i^* \left(1 - \frac{\lambda_i - 1}{\lambda_i + 1} \right)$$

The activity corresponding to the i th locus, V_i , can now be written, using π_i the genotype variable for the i th locus, as:

$$V_i = V_i^* \left(1 + \pi_i \frac{\lambda_i - 1}{\lambda_i + 1} \right) \quad (5.2)$$

where $\pi = -1$ means the low homozygote, $\pi = 0$ means the heterozygote and $\pi + 1$ means the high homozygote.

Combining this with (5.1) and introducing $\beta_i = \frac{\lambda_i - 1}{\lambda_i + 1}$ we get

$$F = (X_1 - X_2) / \sum \frac{1}{V_i^* (1 + \pi_i \beta_i)} \quad (5.3)$$

When the genotype is known, that is the values of π_i are specified for each locus, the flux through the system can be found using relation (5.3). Thus for the rather simple system under consideration we have now an explicit formula relating the measure F and the logical vector (π_1, π_2, \dots) which specifies the genotype. This genotype/phenotyp mapping function, based on biochemical considerations, can now be used to consider what sort of gene interactions are likely to be significant in this system.

2. Dominance

The response of the flux, measured in suitably scaled units, to successive alterations, made at locus I for example, can be used to consider dominance in the model. Thus by using the expression (5.3) and collecting together all terms not directly affected by the 1st locus we have -

$$\text{writing } a_1 = \sum_{i \neq 1} \frac{1}{V_i^* (1 + \pi_i \beta_i)}$$

$$a_2 = X_1 - X_2$$

$$F = a_2/a_1 + \frac{1}{V_1(1 + \pi_1\beta_1)} \quad (5.4)$$

The rate controlling effect of the 1st enzyme can be found as -

$$C_{V_1}^F = \frac{V_1}{F} \cdot \frac{dF}{dV_1} = \frac{1}{V_1(1 + \pi_1\beta_1)} / \left(a_1 + \frac{1}{V_1(1 + \pi_1\beta_1)} \right) \quad (5.5)$$

Clearly since a_1 is a function of π_2, π_3 , etc. the control effect of V_1 , i.e the value of $C_{V_1}^F$, depends no the genetic background.

When the 1st locus is heterozygous (5.5) takes a particularly simple form. Thus for $\pi_1 = 0$ (i.e. heterozygous) at the let locus we have

$$(C_{V_1}^F)_{\pi_1=0} = \frac{1}{V_1} / \left(a_1 + \frac{1}{V_1} \right) \quad (5.6)$$

In order to consider the degree of dominance present at a locus a suitable measure of dominance is required, as for example that given by Falconer (1959); Thus if F_{-1}, F_0, F_1 represent the flux corresponding to the low homozygote, the heterozygote and the high homozygote at the 1st locus, all on a given background, then the dominance is given as

$$\text{DOMINANCE} = D = \frac{F_0 - 1/2(F_{-1} + F_1)}{1/2(F_{-1} + F_1) - F_{-1}} \quad (5.7)$$

Using this measure of dominance, complete dominance of the high activity allele would correspond to $D = +1$ and over-dominance to $D > 1$, while $D = 0$ indicates a purely additive response to successive substitutions at the 1st locus. The equation (5.4) giving the flux as a function. of the state of the Ist locus can now be used in conjunction with (5.7) to obtain an expression for the degree of dominance at the 1st locus.

This manipulation is made easier if (5.4) is rewritten a follows:

$$\begin{aligned} F &= a_2/a_1 + \frac{1}{V_1(1 + \pi_1\beta_1)} \\ &= a_2 - \frac{a_2}{a_1\beta_1 V_1} \cdot \frac{1}{\pi_1 + 1/\beta_1(1 + 1/a_1 V_1)} \\ F &= a_2 - K \left(\frac{1}{\pi_1 + \gamma} \right) \end{aligned} \quad (5.8)$$

where

$$K = \frac{a_2}{a_1\beta_1 v_1} \quad \text{and} \quad \gamma = \frac{1 + a_1 V_1}{\beta_1}$$

When substituting the appropriate expressions for F_{-1}, F_0, F_1 in (5.7) (i.e. giving values $-1, 0 + 1$ for π_1 in (5.8)) the leading term of unity and the factor K cancel out and we get:

$$D = \frac{F_0 - 1/2(F_{-1} + F_1)}{1/2(F_{-1} + F_1) - F_{-1}} = \frac{1}{\gamma} - \frac{1}{2} \left(\frac{1}{1 + \gamma} + \frac{1}{-1 + \gamma} \right) / \frac{1}{2} \left(\frac{1}{1 + \gamma} + \frac{1}{-1 + \gamma} \right) - \frac{1}{-1 + \gamma}$$

which on simplification yields $D = \frac{1}{\gamma}$

remembering that

$$\beta_1 = \frac{\lambda_1 - 1}{\lambda_1 + 1}$$

this becomes

$$D = \frac{\lambda_1 - 1}{\lambda_1 + 1} \left(\frac{1}{1 + 1/a_1 V_1} \right)$$

which can be further simplified by using (5.6) to replace the factor

$$\frac{1}{1 + 1/a_1 V_1} \text{ resulting in}$$

$$D = \frac{\lambda_1 - 1}{\lambda_1 + 1} \left(1 - {}^*C_{V_1}^F \right) \quad (5.9)$$

or, in general

$$D = \frac{\lambda_{-1}}{\lambda_{+1}} (1 - C)$$

The dominance of the locus thus is seen to depend both on its "allelic swing", λ , and on its heterozygous coefficient 'C' (i.e. its rate controlling influence).

In this formula C stand for the rate controlling influences of the 1st enzyme at its heterozygous position. The influence of the other loci on the dominance of the first locus thus depends on their influence on the rate controlling effect, C ; of the first locus (ce. equation 5.6).

The formula (5.10) is quite general and indicates that the dominance measure at a locus having a small rate controlling effect is almost independent of the background genotype, the small value of C having only a small effect on the factor $(1 - C)$, and depends directly on the λ for the two alleles at the locus in question. Let us consider the 'symmetrical case' where many loci have similar enzyme activities and λ values. In this case the 'maximum response to selection', R , will define the λ values. By 'response' is meant the ratio between the flux of the best organism (all alleles 'good') and the worst organism (all alleles 'bad'), and inspection of relation (3) for the two cases indicates that the separate λ must all be approximately equal to R .

For the many locus symmetrical case we have, therefore; for each locus

$$D \doteq \frac{R - 1}{R + 1} \text{ and } C \text{ very small.}$$

Thus for a population of organisms where the $n_{\text{maximum response}}$ R , is reasonably large, the dominance can be almost complete, even though the effects of inhibition at the individual loci have become vanishingly small.

3. Epistasis

A similar approach can be used to consider the importance of interlocus interactions or epistasis within the same model system. In this case a suitable measure of the interaction between two loci is the increment in flux when both act together over that which would be expected from the sum of their separate effects.

If F_{00} represents the flux with both loci homozygous low, F_{01} with one homozygous high and similarly for F_{10} and F_{11} then the epistatic interaction ' E ' may be defined as -

$$\text{EPISTASIS} = E \frac{(F_{11} - F_{00}) - (F_{10} - F_{00}) - (F_{01} - F_{00})}{(F_{10} - F_{00}) + (F_{01} + F_{00})} \quad (5.10)$$

In terms of the measured increment, F_M , and the increment, F_{AD} which would be expected if the loci acted additively this can be written as

$$E = (F_M - F_{AD}) / F_{AD}, \quad \text{so that: } F_M = F_{AD}(1 + E).$$

The measure E will be zero if the loci act additively whereas if, for example, $E = 1$ it means that joint action of the loci results in a 100% greater effect on F than would be supposed on the basis of their additivity.

It should now be possible, using (5.3), to substitute the various fluxes in (5.10) and obtain an expression measuring interlocus interaction, just as was done for dominance interaction. However, this proves to be rather complex because we now have to consider coefficients and λ ratios for each of two loci and the algebra is correspondingly difficult. It is, therefore, advantageous to consider the case where the two loci have equal control effects and equal λ ratios and interaction might be expected to be greatest. For this situation it is possible to show that -

$$E = C(\lambda - 1)/(\lambda - 2C(\lambda - 1)) \quad \dots (5.11)$$

where C is the sensitivity coefficient of the flux with respect to either enzyme for the homozygous low or F_{00} case and λ is the allelic swing common to the two loci.

This formula shows that when there are many loci with about equal importance the interaction between pairs of loci will tend to zero. This is because C , the rate controlling effect of an individual Locus, becomes small as the number of loci increases. For the case of large λ and $C \ll 1$ equation (5.11) becomes:

$$E \simeq C$$

Consider for example a symmetrical system with 10 enzymes acting in a chain, having a common λ value greater than, say, 5. In such a system the coefficient C of any locus will always be of order 0.1 as also will be the epistatic interaction between any pair of loci. This means that simultaneous substitutions at any two loci will usually produce about 10% greater effect than would be expected from the sum of their separate effects. When there are a larger number of loci of about equal importance interactions between pairs of loci

will tend to become negligible. However, this may not be the case for interactions between more than two loci. A quick way of judging the importance of these high order interactions is to consider a mode? where the loci are split into two equal groups, each having a similar rate controlling effect and a similar λ . In this case the model formally reduces to a two enzyme pathway and the value of C in (5.11) becomes 0.5, since the rate control is shared equally between the two formal enzymes:

$$\text{thus } E = \frac{0.5(\lambda - 1)}{\lambda - 2 \times 0.5(\lambda - 1)}$$

$$\text{or } E = \frac{\lambda - 1}{2}$$

If $\lambda = 10$ we have $E \simeq 5$ which is a very marked gene interaction effect between, say, the first 10 and the last 10 loci in a pathway.

We thus have a situation where high order interactions are quite marked at the, level of physiology of individual ‘model’ organisms and this raises the question of the role of these interactions in the context of ‘population’ genetics. Such questions can be considered most easily for the simple case of a symmetric model, that is for one in which substitutions made at one locus are exactly equivalent to substitutions made at any other.

4. Implications of study for response to selection

Here, for simplicity, the physiologically symmetrical case will be considered, that is, all variable loci are assumed to have the same V^* and λ - Suppose that in the population there are N such variable loci and in addition V_0 non variable loci each of which provides an effective activity having the same value as V^* . In a particular organism let I loci have at least one substitution by a ‘good’ allele and of these let I be the number of heterozygotes. The number of loci homozygous for the ‘bad’ allele will then be $(N - X)$ and the flux F in such an organism will, using equation (5.3), be given by

$$F_i = (X_1 - X_2) / \frac{No}{V^*} + \frac{N - X}{V_2^*/(1 + \lambda)} + \frac{Y_*}{V^*} + \frac{\frac{X}{V^*} - Y}{V^2/(1 + \lambda)} \quad (5.12)$$

The value of F in different organisms will depend only on fixed parameters, having the same value for all the organisms, and on the proportion of loci in any particular organism which are singly or doubly substituted by the ‘good’ allele. Writing $x = \frac{X}{N}$, $y = \frac{Y}{N}$, $n = \frac{No}{N}$ and dropping proportional factor: the relation (5.12) can be written as

$$F \propto 1 / \left(1 + \left(\frac{2n}{1 + \lambda} \right) - x \left(\frac{\lambda - 1}{\lambda} \right) + \frac{y}{\lambda} - \frac{2\pi}{\lambda(\lambda + 1)} \right) \quad (5.13)$$

When λ is reasonably large (about 10) the relationship (5.13) will be well approximated by

$$F \propto 1 / (1 + K - x) \quad (5.14)$$

where

$$K = \frac{2n}{1 + \lambda}$$

This approximate representation (5.24) describes a genotype-phenotype relation which is clearly dominant at all loci, since x is indifferent to a second substitution of a ‘good’ allele. The importance of epistatic interaction in (5.14) can be seen by considering how the response to a substitution depends on the background. Such a substitution represents a small increase, Δx , in x and the response will be approximately given by

$$\Delta F \approx \left(\frac{dF}{dx} \cdot \Delta x \right)$$

or from (5.14) $F \approx \frac{1}{(1 + k - x)^2} \Delta x$

thus for a given Δx and using (5.14) again we have

$$\Delta F \propto F^2 \quad (5.15)$$

This means that for a population of such organism with a maximum response to selection R of, say, 5 fold the effect of a given substitution can vary 25 fold ($= 5^2$) depending on the genetic background. The multiplicative interaction, often used to represent epistatic interaction in theoretical genetics, would give a response to substitution, analogous to (5.15) of

$$\Delta F \propto F \quad (5.16)$$

Clearly the relation (5.14) implies a ‘stronger’ epistatic interaction than does a multiplicative model.

What does this imply for the behaviour of a finite population under selection for F ? If the separate loci are very tightly linked on a single chromosome and the phenotype is defined in the dominant epistatic manner of (5.14) then selection will progressively reduce the number of chromosomes until only two or three chromosomal types remain. This is because with very light linkage the system is equivalent to a large number of alleles at one locus and there are ‘overdominance’ relations between some of them. Selection thus leads to a reasonably complementary pair of chromosomes selected from among those originally present.

An argument will now be given to show that if such a pair of chromosomes predominate in the population a relatively stable situation.

may still exist even when recombination between chromosomes is allowed. Consider for example the relationship (5.14) when $n = 1/2$, that is the ratio of non-variable to variable loci is 0.5, and $\lambda = 9$. Then

$$F \propto 1 / \left(1 + \frac{1}{10} - x \right) \quad (5.17)$$

This relationship between the flux and the proportion of substitutions, x , in a given organism is shown on the graph opposite, where x moves almost continuously in the range zero

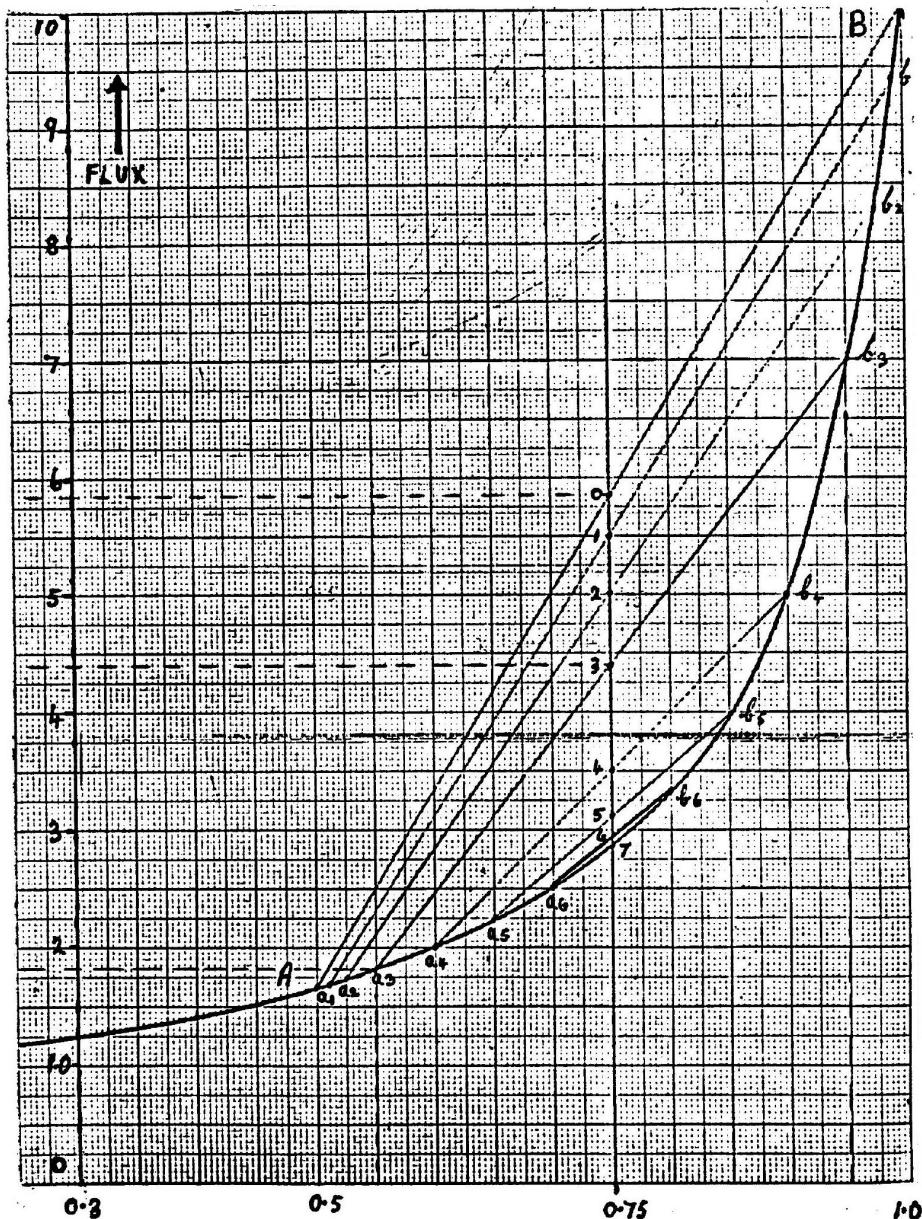
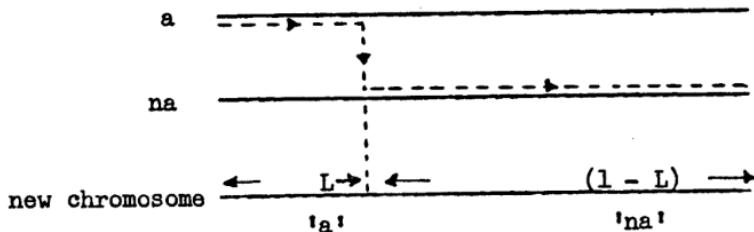


Figure 5.1 The graph shows the plot of function (5.17). As an example consider a recombinant chromosome, formed by a cross-over occurring 1/10 from the end. This corresponds to the points a₃ and b₃ on the graph giving the point “3”. This represents a mean value of 4.3 on the flux scale for this chromosome in the population compared to a value of 5.8 for either of the original chromosomes.

up to unity when there are a large number of loci. Consider the limiting case when the number of loci is large and the chromosome can be regarded as a continuous structure and, suppose, to keep the argument simple, that the two chromosomes have equal probability of good or bad alleles along their length and are complementary in the sense that a good allele in one always occurs opposite a bad allele in the other.

The argument regarding the 'stability' of such a situation under upward selection for F consists in showing that there will be strong selection against recombinant chromosomes and that these could therefore be removed as fast as they are produced. Thus consider the two predominant complementary chromosomes in the diagram below, which will be called 'a' and 'na', 'na' standing for 'not' a and suppose that a recombination occurs at a proportional distance ' L ' from one end. The new chromosome then consists of a portion, L , of 'a' and a portion $(1 - L)$ of 'na' as shown in the diagram below:



This new chromosome will occur in the population equally frequently with 'a' and with 'na' and hardly at all with itself. When combined with 'a' its x value will be $\left(1 - \frac{L}{2}\right)$ and when paired with 'na' its x value will be $\left(\frac{1}{2} + \frac{L}{2}\right)$, these points being symmetrically situated about the value $x = 0.75$. The mean value of the new chromosome in the population will thus be found as the average or mid-point of the corresponding points on the graph, shown as (a_1, b_1) (a_2, b_2) , etc. for values of L_1, L_2 etc. which represent crossover positions increasingly close to the middle of the chromosome. The two predominant chromosomes 'a' and 'na' have identical mean values in the population. This value is equal to the mid-point of the line joining A and B on the graph, that is the point marked '0'.

Thus it can be seen that recombinant chromosomes corresponding to crossover near the middle of the existing chromosomes are at a strong selective disadvantage: Although the effect is less marked for crossovers near the ends of existing chromosomes the new chromosomes so produced will still be selected against.

The general argument just given suggests that the combination of dominance and strong epistatic interaction, represented by the concavity of the graph of F against x , can, when combined with reasonably close linkage, maintain a situation which is very far from linkage equilibrium. However, one important point has not been allowed for and this is that although it is reasonable to consider the chromosomes as continuous and balanced when

considering crossovers within them this will not be so for crossovers at the extreme end.

In this case the new chromosome can actually become ‘better’ than either of the existing ones, by picking up extra ‘good’ alleles at its ends, and be selected in. This ‘end effect’ should then lead to the gradual fixation of ‘good’ loci, although this may happen very slowly since at any time only the outermost unfixed loci in the map are ‘available’ for fixation.

It is, of course, important to see whether these qualitative predictions can be confirmed by direct simulation. A program capable of dealing with a reasonably large number of loci, say thirty, will be required if the ‘end effect’ is to be observed and this has already been developed.

5. The phenotype/genotype relation for more complex systems

In CH II (p53-55) it was shown the genotype parameters, π_1, π_2, \dots , could be embedded within the usual mathematical formulation of a complex metabolic system. The dependence of a phenotypic measure, M_x , on the genotype and the environment then being of the form

$$M_x = M_x (\pi_1, \pi_2, \dots, \dots, \pi_m, X_1, X_2, \dots) \quad \dots \quad (5.18)$$

For the straight chain system just considered this took on the simple form of relation (5.3). In general, however, the relation cannot be expressed by a simple formula and it becomes necessary to employ the computer techniques of CH IV, both to construct and then investigate any particular system. In this way it is hoped that insight may be obtained into the way in which the underlying metabolic system determines general features of the Ph/Ge relation.

Consideration of the analytically obtained expression (5.3) shows clearly that the general relation, (5.18), is unlikely to be approximated by any of the statistical representations commonly used in considering the behaviour of populations subject to selection for the quantitative character, as for example in Fraser (1962, 1967). How important this non applicability of the normally assumed Ph/Ge relationship may be for population studies can be considered by running the usual simulation programmes of theoretical population genetics (Fraser, 1962) in conjunction with metabolic simulation programmes such as my own. However computational limitations soon arise. Thus a Monte-Carlo simulation of a 10 locus problem with an offspring population of 20 animals would take about 3 minutes/generation compared with about 1/10 of a second for a symmetric non epistatic model on the same computer. This situation can be avoided by providing the population programme with a pre-computed table of phenotype values but this method becomes impractical at about 6 loci, (Burns, 1970). In these circumstances it remains important to consider the Ph/Ce relationship at the physiological level in the hope of finding some relatively simple rules for compounding genetic effects operating through a metabolic network. However, even for such physiological studies. of models having alternative alleles at a number of loci,

6.. DOMINANCE, WHEN THE ALLELIC SUBSTITUTION HAS ONLY A SMALL EFFECT ON P

it is not always practical or useful to simply evaluate all the different phenotypes since there will be a very large number of these. In these circumstances the sensitivity coefficient can again provide useful insight, even though the π variations are in fact finite. For example an estimate of the effect of a substitution at the i th locus, SUB_1 , on a given background can be obtained as follows. Thus

$$SUB_i = \frac{\Delta M}{M} \simeq \frac{\partial M}{\partial \pi_i} \cdot \Delta \pi_i$$

Since $\Delta \pi_i$ will be the same no matter what locus is being considered the relative importance of the different variable loci can be seen as

$$[SUB_1, SUB_2, \dots] \propto \left[\frac{\partial M}{\partial \pi_1}, \frac{\partial M}{\partial \pi_2}, \dots \right]$$

The computer can easily produce the vector $\left[\frac{\partial M}{\partial \pi_i} \right]$ on various backgrounds and environments and hence reveal which genes have a major and which a minor effect and to what extent this classification is likely to be altered by changes in environment. In this respect the figure opposite, from a computer simulation of five saturable enzymes in a straight chain, shows how quite small changes in the environment can completely reverse the relative classification of two genes as major and minor.

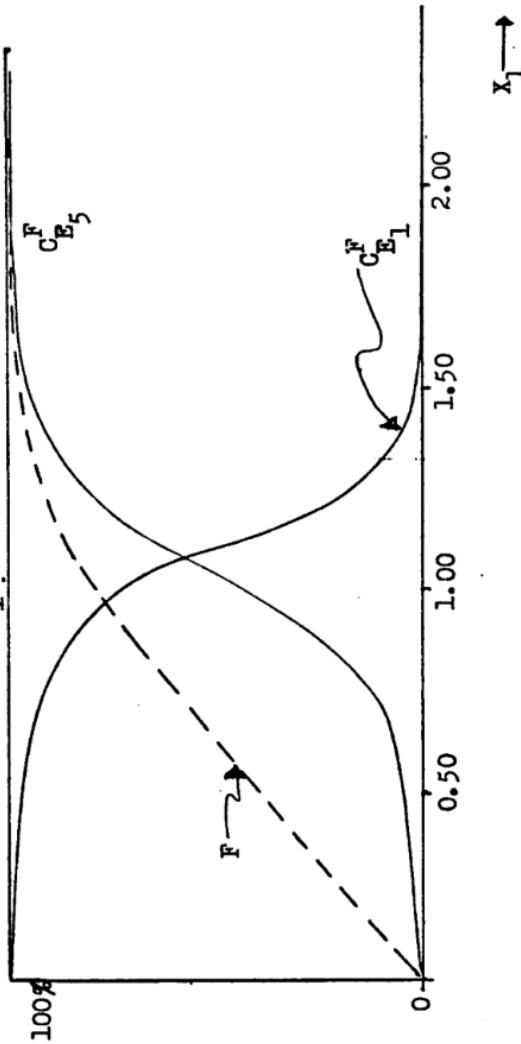
6. Dominance, when the allelic substitution has only a small effect on pools.

In general two alleles at a locus will produce enzymes which are considerably different and complete substitution of one allele for the other can be expected to produce a marked change in at least some pools, although the effect on flux may be much less. Under these circumstances it is difficult to make any general argument about the degree of dominance, that is the position of the stationary solution corresponding to the heterozygous state at the locus.

Consider, however, the restricted case in which the alternative enzymes, whilst having distinct kinetic parameters, are each capable, at the existing steady state, of carrying about the same flux. It then becomes possible to relate the dominance to the two sensitivity coefficients corresponding to the locus being in the alternative homozygous states. The argument is as follows.

SWITCH OF CONTROL IN A FIVE ENZYME PATHWAY

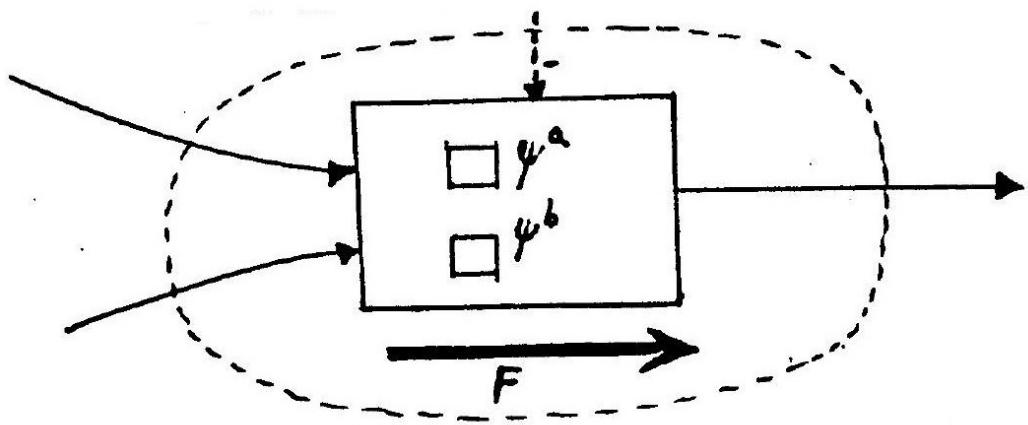
All the enzymes in the model are capable of saturation but the Michaelis constant of E_5 with respect to S_4 is very low compared with all the others, which are set at a large value. When X_1 is small all the pools are small and the chain acts as a chain of linear enzymes in which the rate control is almost entirely on the 1st enzyme E_1 . As X_1 is increased E_5 becomes saturated first and the rate control switches from E_1 on to E_5 . This switch occurs within an X_1 change of only about 50%.



(As a comparison, to the movement of the flux with respect to X_1 is also shown. Although this displays "saturation" it is not the simple Briggs-Haldane type).

Thus suppose the step corresponding to the locus has two alleles 'a' and 'b', as shown below.

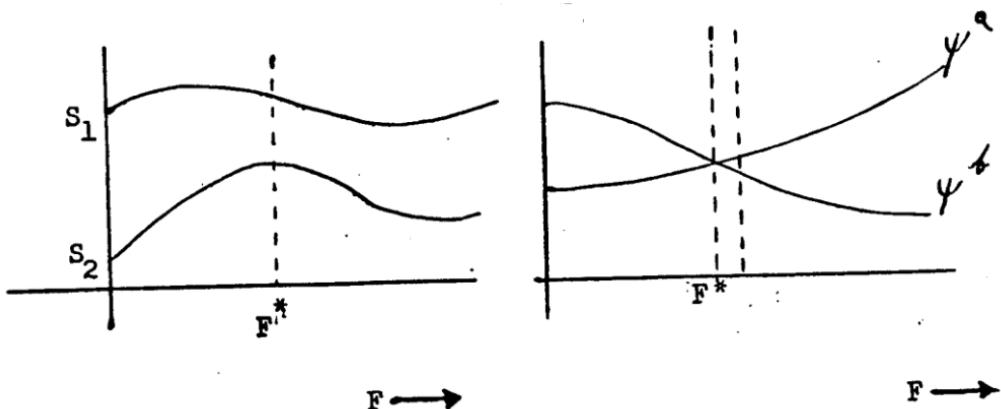
6.. DOMINANCE, WHEN THE ALLELIC SUBSTITUTION HAS ONLY A SMALL EFFECT ON P



Its 'effective' rate expression will be, (p.54)

$$F = (1 - \pi)\psi^a + \pi\psi^b \quad (5.19)$$

Where ψ^a and ψ^b are the complete rate expressions appropriate to the separate alleles in the homozygous state and include enzyme control terms if necessary. Now when the flux, F , carried by the control terms if necessary. Now when the flux, F , carried by the step alters slightly, due for example to an 'a' modulation of the block, the pools S_1, S_2, \dots , will change only to the extent that F itself has been altered and this will also be true of the known functions of the pools, ψ^a and ψ^b . This is shown in the diagrams below where it should be noted that $\psi^a \approx \psi^b$ around the existing value, F^* , of the flux F .



Considering now the composite block, whose rate expression is given by (5.19), in isolation and remembering that the expressions ψ^a and ψ^b are defined by the flux F we can obtain by differentiation the results that

$$C_1 = 1/1 - \left(\frac{\partial \psi^a}{\partial F} \right)_{F=F^*}, C_2 = 1/1 - \left(\frac{\partial \psi^b}{\partial F} \right)_{F=F^*} \quad (i)$$

Where C_1 and C_2 are the coefficients of the flux F for the block when the locus is homozygous for ‘a’ and ‘b’ respectively.

Next consider the relation (5.19) for $\pi = 0, 1/2, 1$, corresponding to the locus being homozygous for ‘a’, heterozygous and homozygous for ‘b’ respectively, and let the respective values of F be F^* , $F^* + \beta\Delta F^*$ and $F^* + \Delta F^*$.

Remembering that the ψ functions are defined by the flux F this gives rise to the three relations

$$F^* = \psi^a(F^*) \quad (\text{ii})$$

$$F^* + \beta F^* = 1/2(\psi^a(F^* + \beta\Delta F^*) + \psi^b(F^* + \beta\Delta F^*)) \quad (\text{iii})$$

$$F^* + \Delta F^* = \psi^b(F^* + \Delta F^*) \quad (\text{iv})$$

If β can be found from these relations it will be possible to find the dominance since the definition given earlier, in relation (5.7), is equivalent to

$$\text{DOM} = 2\beta - 1 \quad (\text{v})$$

Where DOM measures the dominance of the allele ‘b’ over the allele ‘a’. Expanding the ψ^a term in (iii) about F^* , the ψ^b term about $F^* + \Delta F^*$ and using (ii) and (iv) to provide the 1st parts of these expressions we get

$$\beta \cdot \Delta F^* = 1/2 \left(\frac{\partial \psi^a}{\partial F} \cdot \beta \cdot F^* + \frac{\partial \psi^b}{\partial F} \cdot (\beta - 1) \cdot \Delta F^* \right)$$

Which when solved for β gives

$$\beta = \frac{\frac{\partial \psi^b}{\partial F} - 1}{\frac{d\psi^a}{\partial F} + \frac{d\psi^b}{\partial F} - 2} \quad (\text{vi})$$

Replacing the derivatives in (vi) by the coefficients C_1 and C_2 of relation (i) and using the resulting expression in the relation (v) for DOM we can show that

$$\text{DOM} = \frac{C_1 - C_2}{C_1 + C_2} \quad (5.20)$$

The result (5.20), apart from being useful as a numerical estimate in simulation studies, has interesting consequences. Thus in most metabolic situations the coefficients C_1 and C_2 can be expected to be +ve for their own flux so that

$$-1 < \text{DOM} < 1$$

This implies that ‘overdominance’ cannot occur between alleles ‘a’ and ‘b’ having only a small effect on pools.

Again the expression (5.20) can be written as $\frac{C_1/C_2 - 1}{C_1/C_2 + 1}$ and it is clear that even if C_1 and C_2 are themselves small their ratio can in principle take any positive value. Thus almost complete dominance is quite possible and it will be towards the allele possessing the smaller coefficient. This dominance can occur despite vanishingly small pool and flux responses to the substitution.

7. Epilogue

At the outset we stated that our interest was primarily in considering the way in which measures on genetical systems might depend on the underlying genetically specified parameters and on the environment. It is not claimed that we have done more than open up methods for such a consideration and define certain concepts more clearly. However it seems clear in retrospect that this may be a very necessary attempt, aiming, as it does, to understand the quantitative implications of the recent advances in understanding organisms as biochemical systems.

Although considerable work had been carried out by the school of Chance on the dynamics of biochemical systems, which served as a useful guide, our own genetical interests dictated a rather different approach in which stress was placed on the stationary solution and its response to parametric variation.

In this connection the ‘sensitivity coefficient’ of a measure with respect to a differential change in a parameter seems a useful way of describing parametric response and the introduction of a block coefficient enables the relative importance of variation in different functional blocks to be considered. In particular this leads to a clear definition of what is meant by ‘rate control’ in a pathway and the summation property of the flux coefficients (p.85, 3.22) indicates what values of coefficient will indicate a ‘master’ reaction in a pathway. The way in which the block coefficients, which describe the systemic response, are related to the elasticities which describe the behaviour of the separate subunits making up the system, is considered in some detail in CH. III and provides some insight into the way in which the structure of the system and the detailed behaviour of its parts will combine to determine system properties. This treatment makes clear how coefficients of interest within part of a larger system can be determined by an *in vivo* estimation of the relevant elasticities using data obtained from suitable genetically produced mutants. These theoretical/experimental methods for determining coefficients are contrasted with the more laborious ‘direct’ methods which are also available.

The theory relating system coefficients to elasticities also allows the treatment in CH. III of a number of questions. Thus the use of the pool pattern in a chain of enzymes as a diagnostic for control is extended beyond the linear treatment given in CH I and it is shown under what circumstances the saturation effects become important (p.116). Again it becomes possible to consider whether there are any necessary limits to coefficient values (p.110) and whether the effects of a modulation tend to become attenuated with metabolic distance from the site of the modulation (p.124). Finally it is shown that systems having optima allocation for growth will have their enzymes distributed in relation to their sensitivity coefficients (p.141). To some extent all of these theoretical results must remain equivocal in the sense that they are based on particular theoretical assumptions. Accordingly in CH IV emphasis is placed on the idea that computer systems should become available to biochemists and other experimentalists which

Finally in CH.V the implications of our approach for the population genetics of quantitative

characters is considered. Of considerable interest here is the theoretical result, specifically demonstrated for a chain of linear enzymes, that considerable dominance and epistasis will arise quite naturally and will persist even when the individual allelic substitutions have a vanishingly small effect. A qualitative argument is provided to show that such a genotype-phenotype relation could, when combined with suitable linkage between the loci, give rise to very different behaviour under selection from that which would be expected on additive assumptions. For more complex systems it is shown that the sensitivity coefficient aids in the estimation of major and minor loci and also can be related, under certain circumstances, to the degree of dominance between two alleles. Once again, however, theory requires to be supported by flexible means for conducting numerical studies on populations of ‘theoretical organisms’.

If any claim to a further understanding of the organisation of living things could be made, it would perhaps stress less the particular results shown in this thesis, but emphasise the approach.

Acknowledgements

I must express my gratitude to Professor C. H. Waddington who, in the first stages of this investigation, provided support and to Professor D.S. Falconer who generously sustained and encouraged me in the later period of this work. I would also like to thank my friend Dr. H. Kacser for his patient supervision and for the many ideas which he has contributed towards this work.

In addition I wish to acknowledge the many helpful discussions with Dr. I.B. Barthelmess, Dr. C.F. Curtis, Dr. O.J. Gillie, Dr. B. Watson and Mr. R.W. Tateson and many others.

A special acknowledgement is due to the Edinburgh Regional Computer centre who provided computing facilities and 'know how' and also to my brother Dr. J.G. Burns for many discussions on the computing aspects of the work. Finally I would like to thank Miss A. Brown whose assistance helped greatly in finally bringing this work to term.

References

- ACHS, M.J., GARFINKEL, D. Comput. Biomed. Res., 2, 93 (1968).
- BRIGGS, G.E, HALDANE, J.B.S., Biochem. J. 19, 338 (1925).
- BURGIN, G.H., Simulation 6 , No.3, (1966).
- BURNS, J.A. in 'Quantitative Biology of Metabolism ', 75-80, (Locker, A., Ed., Springer Verlag, New York, 1968).
- BURNS, J.A., FEBS Letters, 2, Suppl. "Computing Techniques in Biochemistry", S30-S33, (1969).
- BURNS, J.A., in "Towards a Theoretical Biology, 3: Drafts" (Edinburgh University Press, 1970).
- GHANCE, B., WILLIAMS, F.C. , YANG, C.C., BUSSER, J., YIGGINS, J.J., Rev.Sci.Instr. 22, 683, (1951).
- CHANCE, B., GARFINKEL, D., HIGGINS, J.J., HESS, B., J.Biol.Chem., 235, 2440 (1960).
- CHANCE, B., ESTABROOK, R., GHOSH, A., Proc. Natl.Acad. Sci., U.S., 51, 1244 (1964).
- CHANCE, E.M., SHEPARD, E.P., Comput.Biomed.Res., 2, 321 (1969).
- CLELAND, W.W., Biochim. Biophys.Acta., 67, 104-137 (1963).
- COOPER, G.J., FEBS Letters, $\frac{2}{2}$, Suppl. "Computing Techniques in Biochemistryn, S22-S29 (1969).
- DENBIGH, K.G, 4ICKS, M ., PAGE, F.M., Trans. Faraday Soc., 44, 479. (1948).
- 'ENZYMES, THE', Ed. Boyer, P.D., Lardy, H. , Kyrbäck, K., Academic Prosi, New York, (1959).
- GARFINKEL, D., FRENKEL, R., GARFINKEL, L., Comput. Biomed.Res., 2, 31, (1968)
- GARFINKEL, D., FEBS ietters, 2, Suppl. "Computing Techniques in Biochemistry", S9-S13, (1969).
- GARFINKEL, D., GARFINKEL, L., PRING, M., GREEN, B.S., CHANCE, B., Ann. Rev. Biochem., 39, 473-499, (1970).
- GEAR, C.W., Proc. Intern. Fed. Inform. Proc. Soc., A81, Edinburgh (1968). Goodwin, B., "Temporal Organization in Cells". Academic Press, New York, (1963).
- ** GRAINGER, J.M.R., BASS, L., J. Theoret. Biol. 10, 387, (1966).
- GRIFFITHS, J.S., J. Theoret. Biol., 20, 202-216 (1968a) and (1968b).
- HEARON, J.2., Ann.N.Y.Acad.Sci. 108 (1963).
- HIGGINS, J., Ann.N.Y. Acad.Sci., 108, 305-321, (1963).
- HIGGINS, J., J. Theoret. Biol. 21, (1968).
- HURST, R.O., FEBS Abstr., Intern.Meet., 8th, Prague (1968).

- ISAACSON, F., KELLER, H.B., 'Analysis of Numerical Methods'. John Wiley & Sons N.Y./London (1966).
- KACSER, H. "Some physico-chemical aspects of biological organisation". In "The Strategy of the Genes" (C.H. Waddington, auth.) Allen & Unwin, London (1957).
- KACSER, H., in "Biological Organisation at the Cellular and Supercellular Level". p.25, Academic Press (1963).
- KACSER, H., BURNS, J.A., in "Quantitative Biology of Metabolism" (11-23) (Locker, A., Ed., Springer Verlag, New York) (1968).
- KAUFFMAN, S.A., J. Theoret. Biol., 22,437 , (1969).
- KING, E.L., ALTMAN, C., J. Phys. Chem., 60, 1375 (1956).
- ** GRAINGER, J. M.R., GOFFREY, P.E., WEST, T, T, , J. Theoret. Biol. , 21, 123
- LEWIS, G.N., 'A New Principle of Equilibrium". Proc.Nat.Acad.Sci., i1, 179-83. (1925)
- LEWONTIN, R.C., Genetics, 49, 49-67 (1964).
- PRING, M.J., J. Theoret. Biol., 17, 430 (1967). 210.
- RHOADS, D.G., PRING, M. J., J. Theoret. Biol., 20, 297 (1968).
- ROBERTSON, A., Ann. Rev. Genet., 1 , 295-311 (1967).
- REICH, J.G. FEBS letters, 2 , Suppl. " Computing Techniques in Biochemistry, S3-S8 (1969).
- SEGAL, H. L. in "The Enzymes", Vol.I.p.1, Academic Press, N.Y. (1959).
- WAGNER, P., BERGQUIST, A., Proc.Nat. Acad. Sci. 49, 892 (1963).
- WALTER, C., J. Theoret. Biol., 23, 23-52.
- WATSON, J.D., N Molecular Biology of the Gene, Benjamin, New York, (1965).