Bulletin of Mathematical Biology Vol. 46, No. 5/6, pp. 745-764, 1984.
Printed in Great Britain

0092-8240/84$3.00 + 0.00
Pergamon Press Ltd.

STIELTJES INTEGRATION AND DIFFERENTIAL GEOMETRY: A MODEL FOR ENZYME RECOGNITION, DISCRIMINATION, AND CATALYSIS

■ A. H. LOUIE† and R. L. SOMORJAI
Division of Chemistry,
National Research Council,
Ottawa, Ontario,
K1A 0R6, Canada

A model for enzymic catalysis is presented using the mathematical theories of differential geometry and Stieltjes integration. The Stieltjes *integrator* is a complex-valued function of bounded variation which represents the curvature and torsion, hence the conformation, of the backbone of an enzyme molecule. The *integrand* is a complex-valued continuous function which describes the shape of the surface of a substrate molecule. We postulate that enzyme–substrate interactions correspond to evaluations of Stieltjes integrals, and that observables of enzymic catalysis correspond to projections.

Results from the mathematical theory of the Stieltjes integral are discussed together with their biological interpretations. We contrast the difference between structural and functional proteins, and construct analogues of enzyme cofactors, modifications, and regulation. Various techniques of locating the active site on enzymes are also given. We construct a total variation metric, which is particularly useful for detecting similarities among proteins.

An examination on the many different modes of convergence of mathematical functions representing biological molecules leads to a mathematical statement of the fundamental dogma of molecular biology, that 'structure implies function'. Similar arguments also result in the converse statement 'function dictates structure', which is a basic premise of relational biology.

Stepped-helical approximations of the backbone space curves of enzymes provide a concrete computational tool with which to calculate the Stieltjes integrals that model enzymic catalysis, by replacing the integral with a finite series.

The duality between enzymes and substrates (that they are *meters* 'observing' one another) is shown to be a consequence of the mathematical duality of Banach spaces. The Stieltjes integrals of enzyme–substrate interactions are hence shown to be bounded bilinear functionals. The mechanism of enzymic catalysis, the transformation from substrate to product, is also formulated in the Stieltjes integration context via the mathematical theory of adjoints.

The paper closes with suggestions for generalizations, prospects for future studies, and a review of the correspondence between mathematical and biological concepts.

1. Introduction. Recognition and its correlate, discrimination, are important conceptual cornerstones for the understanding of the specificity and control of biological systems and processes. Enzymes provide outstanding

† Present address: 86 Dagmar Avenue, Vanier, Ontario, KIL 5T4, Canada.

examples, with their capability of recognizing specific substrates and their precision in discriminating between molecules on the basis of differences in some structural features.

The selectivity of an enzyme is, to a large degree, related to the three-dimensional conformation of its molecular structure, which must be maintained to retain biological activity. Furthermore, activity is usually associated with a particular site in the catalytic protein—the active site—and any factor which produces a distortion of this site can interfere with the enzyme's function.

Enzymic catalysis proceeds through the intermediate formation of an enzyme–substrate complex, within which rearrangements of the substrate take place to yield products, simultaneously reforming the native enzyme.

An important feature of biological recognition systems is their finite power of resolution. Even the most discriminating enzyme can occasionally be 'tricked' by molecules resembling its substrates: competitive inhibitors vie with natural substrates for the binding to an enzyme.

These are some of the important features of enzymic catalysis. We present an abstract mathematical model of enzyme function which captures these characteristic features. The differential geometry of proteins we discussed in earlier papers (Louie and Somorjai, 1982, 1983) is used as a mathematical tool. The theory itself is derived from the Edelstein-Rosen (1978) model of enzyme–substrate recognition.

2. The Edelstein–Rosen Model. A model for enzyme–substrate recognition is presented in Edelstein and Rosen (1978). The theory rests on the following two postulates:

(1) Substrates can be represented by real-valued continuous functions which vanish outside some closed bounded region in $\mathbf{R}^3$; i.e. substrates are represented by elements $F \in C(K, \mathbb{R})$, the space of all real-valued continuous functions on K, where K is a compact subset of $\mathbf{R}^3$, chosen to be large with respect to molecular dimensions.

(2) Recognition of the substrate corresponds to the evaluation of a linear functional on $C(K, \mathbb{R})$.

—or what is equivalent:

(2′) Associated with a given enzyme is a function α in the dual space of $C(K, \mathbb{R})$, i.e. $\alpha \in NBV(K, \mathbb{R})$, the space of all real-valued, normalized functions on K of bounded variation, and recognition of a substrate F results from the evaluation of a Stieltjes integral of the form $\int F \, d\alpha$.

(The equivalence of (2) and (2′) is due to a theorem in functional analysis known as the Riesz Representation Theorem, which will be discussed in Section 5.)

These postulates are summarized in Fig. 1.

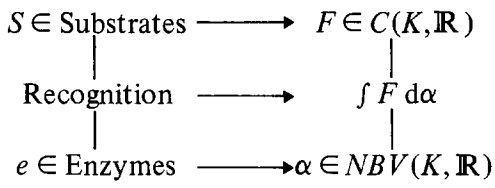

Figure 1.

The representation of molecules by continuous functions and functions of bounded variation on a compact subset K of three-dimensional space is a natural one, in keeping with the usual conception that molecules are three-dimensional objects with 'shapes', occupying definite regions of space. Edelstein and Rosen (1978) suggested that molecules can be viewed 'as continuous three-dimensional distributions (of charge, mass, or other *unspecified* physical parameters)', and it is through these 'complex combinations of physical parameters, some of which are not necessarily observables to our microscopie meters', that the correspondences $S \leftrightarrow F$, $e \leftrightarrow \alpha$ are derived.

While the Edelstein–Rosen theory encompasses some of the features of enzyme–substrate recognition we mentioned in Section 1, it remains unclear *how* the correspondences $S \leftrightarrow F$, $e \leftrightarrow \alpha$ can be mathematically formulated. In the following sections we propose such a formulation, as well as extend the modelling to at least some aspects of discrimination and catalysis *per se*. This concretization depends importantly on being able to represent proteins as bona fide geometric objects.

3. Geometric Representations of Proteins. In our continuing study of the differential geometry of proteins (Louie and Somorjai, 1982, 1983), a protein molecule is regarded as a geometric object. In a first approximation, a protein molecule is represented as a regular parametrized space curve which passes through the C_α-carbons. According to the fundamental theorem of space curves, such a curve is characterized uniquely, except for position in space, i.e. up to a rigid motion, by two continuous functions of the arc-length $s \in I = [a, b] \subset \mathbb{R}$ along the curve, the *curvature* $\kappa \geqslant 0$ and the *torsion* τ. In particular, the collection of space-curve representations of conformations form a subset of $C(I, \mathbb{R}^2)$, the class of continuous mappings from I to $\mathbb{R}^2$.

The pair (κ, τ) of real-valued continuous functions immediately suggests the alternate description $\alpha = \kappa + i\tau$. Thus the collection of protein backbone conformations can be considered as a subset of $C(I) = C(I, \mathbb{C})$, the class of complex-valued continuous functions on I. In Louie and Somorjai (1983)

we showed that the helical approximations of space curves provide a valuable tool to analyze protein patterns. To expand the class of protein backbones to include stepped-helices (connected sequences of helices), we shall relax the restriction on $\kappa + i\tau$ and only require it to be *sectionally* continuous. Furthermore, since protein backbones are 'well-behaved' space curves in which mathematically 'pathological' conditions (e.g. infinite length, unbounded derivative, etc.) do not occur, we can assume without loss of generality that $\alpha = \kappa + i\tau$, representing a protein conformation, is also a function of bounded variation. Note that a sectionally constant $\kappa + i\tau$, which corresponds to a stepped-helical protein backbone, is both sectionally continuous and of bounded variation.

We summarize this section in the following:

Postulate 0. The three-dimensional conformation of a protein backbone is represented by a complex-valued function of a real variable, $\alpha = \kappa + i\tau$: $I \subset \mathbb{R} \to \mathbb{C}$, which is sectionally continuous and has bounded variation.

4. Complex Stieltjes Integration. In order to make the mathematical basis of the ensuing discussion self-contained, we need a digression on the fundamentals of the Stieltjes integral. The exposition will be limited to those results essential for a cogent representation of the mathematical biology of enzymic catalysis. More comprehensive treatises of the Stieltjes integral can be found in Widder (1946), Burrill and Knudsen (1969) and Rudin (1976).

A function $\alpha: I \to \mathbb{C}$, for $I = [a, b] \subset \mathbb{R}$, is of *bounded variation* if there is a constant $M \geqslant 0$ such that for every partition $P = \{a = s_0 < s_1 < \ldots < s_n = b\}$ of I

$$V_P(\alpha) = \sum_{k=1}^{n} |\alpha(s_k) - \alpha(s_{k-1})| \leqslant M. \tag{1}$$

The *total variation* of α is defined by

$$V(\alpha) = \sup\{V_P(\alpha): P \text{ a partition of } I\}. \tag{2}$$

Clearly, $V(\alpha) \leqslant M \leqslant \infty$. α is of bounded variation if and only if $\operatorname{Re}\alpha$ and $\operatorname{Im}\alpha$ are of bounded variation, and

$$V(\alpha) \leqslant V(\operatorname{Re}\alpha) + V(\operatorname{Im}\alpha). \tag{3}$$

An important property of a function of bounded variation is that it possesses limits both from the right and from the left at all points of $[a, b]$. Thus if α is a function of bounded variation on $[a, b]$, the limit $\alpha(s-)$ of α from the left and the limit $\alpha(s+)$ of α from the right exist for every $s \in [a, b]$. (Define $\alpha(a-) = \alpha(a)$ and $\alpha(b+) = \alpha(b)$.) Moreover, a function of bounded variation can have at most countably many discontinuities.

For complex-valued functions F and α on I, the *(Riemann-) Stieltjes integral* of F with respect to α, denoted by

$$\int_I F \, d\alpha, \tag{4}$$

is the limit, if it exists, of sums of the form

$$\sum_{k=1}^{n} F(s_k')[\alpha(s_k) - \alpha(s_{k-1})], \tag{5}$$

where $P = \{a = s_0 < s_1 < \ldots < s_n = b\}$ is a partition of I, each $s_k' \in [s_{k-1}, s_k]$, and the limit is taken over partitions for which

$$\max\{|s_k - s_{k-1}| : 1 \leqslant k \leqslant n\} \to 0. \tag{6}$$

F is the *integrand*, and α is the *integrator.*

If $F \in C(I)$ and α is of bounded variation on I, then $\int_I F \, d\alpha$ exists. In fact

$$\langle \cdot, \cdot \rangle : (F, \alpha) \mapsto \int_I F \, d\alpha \tag{7}$$

is a bounded bilinear functional, with

$$\left| \int_I F \, d\alpha \right| \leqslant \| F \|_\infty V(\alpha), \tag{8}$$

where

$$\| F \|_\infty = \sup\{|F(s)| : s \in I\} \tag{9}$$

is the supremum (L^∞-) norm on $C(I)$ which turns the latter into a Banach space.

5. The Dual Space' of $C(I)$. It follows from above that if α is a complex-valued function of bounded variation on I and

$$\hat{\alpha}(F) = \langle F, \alpha \rangle = \int_I F \, d\alpha \tag{10}$$

for all $F \in C(I)$, then $\hat{\alpha}$ is a bounded linear functional on $C(I)$, i.e. $\hat{\alpha} \in C(I)^*$, and

$$\| \hat{\alpha} \| \leqslant V(\alpha), \tag{11}$$

where $\| \hat{\alpha} \|$ is the operator norm of $\hat{\alpha}$ defined by

$$\| \hat{\alpha} \| = \sup\{|\hat{\alpha}(F)| : \| F \|_\infty \leqslant 1\}. \tag{12}$$

However, the correspondence between bounded linear functionals on $C(I)$ and functions of bounded variation on I is not one-to-one. Indeed, if $a < c < b$ and

$$\gamma(s) = \begin{cases} 0 & \text{if } s \neq c \\ 1 & \text{if } s = c \end{cases} \tag{13}$$

then $\int_I F \, d\alpha = 0$ for all $F \in C(I)$, and $V(\gamma) = 2$. Thus if one is interested only in that linear functional which a function of bounded variation defines on $C(I)$, then γ and 0 are equivalent, or more precisely, $\hat{\gamma} = \hat{0}$. In order to avoid identifying the dual space of $C(I)$ with equivalence classes of functions of bounded variation, a 'normalized' representative from each class is chosen.

Let $NBV(I)$ denote the space of all complex functions on $I = [a, b]$ which are of bounded variation on I, which vanish at a, and which are continuous from the left on (a, b). If α is a function of bounded variation on I, then the function α_N, defined by

$$\alpha_N(s) = \begin{cases} 0 & \text{if } s = a \\ \alpha(s-) - \alpha(a) & \text{if } s \in (a, b) \\ \alpha(b) - \alpha(a) & \text{if } s = b \end{cases} \tag{14}$$

is a normalized function of bounded variation, i.e. $\alpha_N \in NBV(I)$,

$$V(\alpha_N) \leqslant V(\alpha), \tag{15}$$

and

$$\int_I F \, d\alpha_N = \int_I F \, d\alpha \tag{16}$$

for all $F \in C(I)$, i.e. $\hat{\alpha}_N = \hat{\alpha}$.

With respect to pointwise addition and scalar multiplication, $NBV(I)$ is a linear space, and with the norm $\| \cdot \|_V = V(\cdot)$, $NBV(I)$ is a Banach space. Further, we have the

RIESZ REPRESENTATION THEOREM. *The mapping*

$$\alpha \mapsto \hat{\alpha} = \int_I \cdot \, d\alpha \tag{17}$$

is an isometric isomorphism between the Banach spaces $NBV(I)$ and $C(I)^*$.

In particular, the total variation norm of α and the operator norm of $\hat{\alpha}$ coincide:

$$\|\alpha\|_V = \|\hat{\alpha}\|. \tag{18}$$

6. Functional Representations of Substrates. While it is relatively straightforward that the conformation of a protein molecule, hence an enzyme, can be represented by the function $\alpha = \kappa + i\tau \in NBV(I)$ describing the backbone space curve, a geometric model for substrates is less apparent. For those 'string-like' substrates which have well-defined 'backbones', a similar space-curve argument leads to the $F = \kappa + i\tau$ representation. But such a continuous function F has as its domain a real interval $J \subset \mathbb{R}$, which is not necessarily the same as the domain I of the α-representation of its recognizing enzyme. Thus a 'scaling function' $\gamma\colon I \to J$ must be composed to F, resulting in the continuous function $F \circ \gamma \in C(I)$ portraying the 'shape' of the substrate.

With this motivation, we can formulate a geometric model for substrates. When we say it is the 'shape' of a substrate which is identified by an enzyme, we of course mean the 'surface geometry'. Consequently, a natural representation of a general substrate is a surface embedded in three-dimensional space. A regular surface F in $\mathbb{R}^3$ is a two-dimensional object, hence parametrized by two real variables with domain $\operatorname{dom} F \subset \mathbb{R}^2$. Again identifying $\mathbb{R}^2$ with $\mathbb{C}$, we can now investigate the possibility that the range of the function $\alpha \in NBV(I)$ representing an enzyme is contained in $\operatorname{dom} F$. If indeed $\alpha(I) \subset \operatorname{dom} F \subset \mathbb{C}$, then $F \circ \alpha \in C(I)$, and the Stieltjes integral

$$\int_I F \circ \alpha \, d\alpha \tag{19}$$

can be evaluated. (In fact, in complex analysis, the integral (19) is the *line integral* of F along α.) In this case, the substrate can be described alternatively by the continuous function $F \circ \alpha$.

The condition $\alpha(I) \subset \operatorname{dom} F$ has the interesting biological interpretation that since the representation of the enzyme α falls within the domain of the substrate F, the substrate F *recognizes* the enzyme α. This reversal of roles depicts the symmetry of the theory of enzyme–substrate recognition, and this *duality* of enzymes and substrates will be discussed in Section 12.

Thus we see that a general substrate can be portrayed as a continuous function over an appropriate real interval. For simplicity we can choose a 'universal' interval $I = [a, b]$, large enough for the molecular dimensions of all substrates and enzymes. The functional representation of a molecule then has as its support $J \subset I$, and has the value zero on $I \sim J$. Again for

generality, we shall occasionally also consider sectionally continuous functions on I as substrates.

7. Enzyme–Substrate Integration. We now state explicitly the postulates of our model of enzymic catalysis.

Postulate 1. Substrates are represented by complex-valued continuous functions which vanish outside some closed bounded subset of $\mathbb{R}$; i.e. substrates are represented by elements $F \in C(I)$, where I is a compact interval in $\mathbb{R}$ chosen to be large with respect to molecular dimensions.

Postulate 2. Enzymes are represented by complex-valued normalized functions of bounded variation on I, i.e. by elements $\alpha \in NBV(I)$, hence by linear functionals $\hat{\alpha} \in C(I)^*$.

Contrasting our Postulates 1 and 2 with the Edelstein–Rosen postulates in Section 2, we see that our substrates and enzymes are functions from $I \subset \mathbb{R}$ to $\mathbb{C}$, while theirs are from $K \subset \mathbb{R}^3$ to $\mathbb{R}$.

Postulate 3. Enzymic catalysis, the interaction between an enzyme and a substrate, corresponds to a mathematical coupling of F and α, via the evaluation of the Stieltjes integral

$$\langle F, \alpha \rangle = \int_I F \, d\alpha. \tag{20}$$

The Stieltjes integral yields a *complex* number. In accordance with the convention that every observable can be regarded as a mapping from states to *real* numbers (Mackey, 1963; Arnold, 1978; Rosen, 1978), we state

Postulate 4. Observables of enzymic catalysis correspond to projections of the complex number $\langle F, \alpha \rangle$ to the reals.

Thus different observables produce different 'meter-readings' of the same enzyme–substrate process $\langle F, \alpha \rangle$ (Fig. 2). Note, however, that if the enzyme–substrate pair has no interaction, i.e. if $\langle F, \alpha \rangle = 0$, then all observables give the value 0, and in this case we can say that the enzyme α does not interact with F.

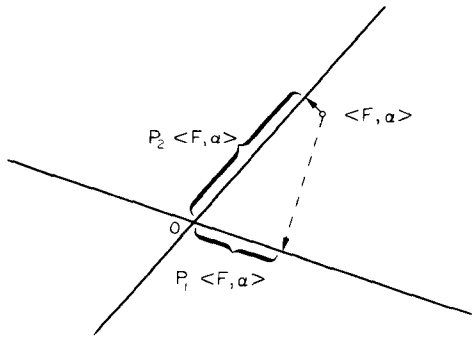

Figure 2.

8. Biochemical Implications. Many mathematical results from the theory of Stieltjes integrals have pleasing biochemical analogues. In this section we discuss several of these correspondences in our modelling relation.

First, let us consider the fibrous (structural) proteins, for example the keratins and collagens. These abundant proteins do not have enzymic activities. Their polypeptide chains are arranged or coiled along a single dimension, often in parallel bundles; i.e. their backbones are regular *helices.* The $\alpha = \kappa + i\tau$ representation of a helix, significantly, is a constant (Louie and Somorjai, 1982, 1983). And when α is a constant,

$$\langle F, \alpha \rangle = \int_I F \, d\alpha = 0 \tag{21}$$

for all $F \in C(I)$ (in fact for *all* complex-valued functions F on I). Thus these fibrous proteins are enzymically inert.

Next, consider this mathematical result: if $\phi \in C(I)$ and $\alpha \in NBV(I)$, then the function α_ϕ defined by

$$\alpha_\phi(s) = \int_a^s \phi \, d\alpha, \tag{22}$$

for $s \in I = [a, b]$, is also a normalized function of bounded variation on I. This leads to the concept of 'modified enzymes'. Given an enzyme $\alpha \in NBV(I)$ and a 'modifier' $\phi \in C(I)$, their interaction results in the modified enzyme $\alpha_\phi \in NBV(I)$, which can in turn operate on other substrates $F \in C(I)$:

$$\langle F, \alpha_\phi \rangle = \int_I F \, d\alpha_\phi = \int_I F\phi \, d\alpha. \tag{23}$$

Clearly $\langle F, \alpha_\phi \rangle$ is not necessarily equal to $\langle F, \alpha \rangle$, hence ϕ does 'modify' the enzyme α. In particular, when $|\langle F, \alpha_\phi \rangle|$ is large compared to $|\langle F, \alpha \rangle|$, i.e. when ϕ increases the activity of α, the above is a model of enzyme cofactors. Consider α as an *apoenzyme.* The modifier ϕ is the *cofactor* (metal ion, coenzyme, or vitamin). The activated form α_ϕ is the corresponding *holoenzyme.* If, on the other hand, $|\langle F, \alpha_\phi \rangle|$ is smaller than $|\langle F, \alpha \rangle|$, then ϕ acts as an *inhibitor* of the enzyme α, and the mathematical theory models enzyme inhibition. Thus depending on the relative values of $\langle F, \alpha_\phi \rangle$ and $\langle F, \alpha \rangle$, various types of modification, regulation, and modulation of enzyme activity can be modelled.

The indefinite Stieltjes integral (22) can also be used to locate the active sites of enzymes. Since interaction of the substrate F and the active site of the enzyme α leads to a slight distortion of the local geometry, a comparison of the two functions α_F and α reveals the location of the active site: candidates

are neighbourhoods of those values of s for which $\alpha_F(s)$ and $\alpha(s)$ are significantly different. An example can be constructed from the results in the Drenth *et al.* (1976) study of the binding of chloromethylketone substrate analogues to crystalline papain. Different chloromethylketones were reacted with the proteolytic enzyme papain (EC 3.4.22.2), and it was found that the conformations of residues cysteine-25 and histidine-159 in the derivative structures (enzyme–substrate complexes) were not the same as those in the native enzyme. These different conformations result in different curvature and torsion values of the corresponding backbone space curves. In other words, for papain α and a chloromethylketone substrate analogue F,

$$|\alpha_F(s) - \alpha(s)| > 0 \tag{24}$$

in neighbourhoods of $s = 25$ and $s = 159$. (The domain interval I can be scaled appropriately so that $s \in I$ represents 'residue number'. For papain, let $I = [1, 212]$.) Furthermore, the operator norm between α_F and α,

$$\|\alpha_F - \alpha\|_V = V(\alpha_F - \alpha), \tag{25}$$

can be used as a measure of the deviation between the enzyme–substrate complex and the native enzyme.

The theory of active-site location can also be formulated as follows. For the substrate F let

$$F_s(\cdot) = F(s - \cdot), \tag{26}$$

i.e. F_s is a lateral translation of F over the real domain. Then for each $s \in I$ we can evaluate the *Stieltjes convolution*

$$\langle F_s, \alpha \rangle = \int_I F_s \, d\alpha. \tag{27}$$

The value of $s \in I$ which gives $\sup |\langle F_s, \alpha \rangle|$ will then correspond to the optimal location of the substrate F relative to the enzyme α.

The operator norm $\|\cdot\|_V$, used in equation (25) to measure the variation between an enzyme and its substrate-bound complex, can be used to model *discrimination.* Given an enzyme α, let F and G be two substrates. Then

$$|\langle F, \alpha \rangle - \langle G, \alpha \rangle| \leqslant \|\alpha\|_V \, \|F - G\|_\infty ; \tag{28}$$

i.e. $\|\alpha\|_V$ is the 'scaling factor' between the distance (difference) of the two substrates and the difference (distance) in the degree of recognition of these two substrates. Since the value of $\|\alpha\|_V$ dictates the relative sizes of $\|F - G\|_\infty$ and $|\langle F, \alpha \rangle - \langle G, \alpha \rangle|$, it is an *indicator* of the discriminating power of the enzyme.

Thus the two correlating biological concepts of *recognition* and *discrimina-*

tion are modelled by the interdependent mathematical concepts of *norm* and *metric* on Banach spaces. First, for a substrate F to be recognized by an enzyme α, i.e. for the 'input' to be identified or associated with one of a set of possible alternatives, the Stieltjes integral $\langle F, \alpha \rangle$ must exist. In other words, *integrability* is a prerequisite for *recognizability*. The *degree* of recognition as an observable is the actual *value* of the integral $\langle F, \alpha \rangle$, or the *norm* $|\langle F, \alpha \rangle|$. Discrimination (the detection of *variations*, or differences, in recognizability between substrates F and G) is the distance, or the *metric*

$$\mathrm{d}(\langle F, \alpha \rangle, \langle G, \alpha \rangle) = |\langle F, \alpha \rangle - \langle G, \alpha \rangle|, \tag{29}$$

which gives the associated real-valued observable. Thus the norm and the metric, hence recognition and discrimination, are related by (29).

Naturally, the operator norm $\|\cdot\|_V$ can also be used as a metric in comparing the functional similarity between different enzymes. For two enzymes α and β, their 'distance' is

$$\mathrm{d}(\alpha, \beta) = \|\alpha - \beta\|_V = V(\alpha - \beta). \tag{30}$$

(Note that since α and β may have different domains I_α and $I_\beta \subset I$, a scaling or translation between I_α and I_β may be necessary.) Functional similarity can also be examined by comparing $\|\alpha_F - \alpha\|_V$ with $\|\beta_F - \beta\|_V$ for arbitrary substrates F. These and other metrics of enzymes, and of proteins in general, provide valuable insights into the structure and function of proteins, which we shall discuss in a subsequent paper.

9. Convergence of Enzymes. In Louie and Somorjai (1983) we show that a pair of continuous functions can be approximated by stepped-mappings, functions which are sectionally constant. Thus every space curve, in particular every protein backbone, can be arbitrarily closely approximated by stepped-helices, since a sectionally constant $(\kappa, \tau) \in S(I)$ corresponds to a linked sequence of helices. ($S(I)$ is the space of stepped-mappings on I.) This approximation is with respect to the supremum (L^∞-) norm on $S(I)^- \supset C(I)$. In particular, for every enzyme α there exists a sequence of stepped-helices, or 'approximating enzymes', $\{\alpha_n\}$, such that as $n \to \infty$

$$\|\alpha_n - \alpha\|_\infty \to 0. \tag{31}$$

This L^∞ (i.e. uniform) convergence of $\{\alpha_n\}$ to α is a 'structural' convergence, since it is a convergence of curvature and torsion (which describe the shapes of protein backbones). The mathematical hierarchy of convergence dictates that uniform convergence implies strong convergence

$$\|\alpha_n - \alpha\|_V \to 0, \tag{32}$$

which in turn implies weak convergence

$$\langle F, \alpha_n \rangle \to \langle F, \alpha \rangle \tag{33}$$

for every (substrate) $F \in C(I)$. Hence structural convergence in fact implies 'functional' convergence, a striking and illuminating mathematical analogue of the fundamental dogma of molecular biology, that 'structure implies function'.

The converse of the dogma, a derivative of Rashevsky's (1960) relational biology, that 'function dictates structure', also has a mathematical analogue (although with a slightly stronger assumption and a slightly weaker conclusion). If $\{\alpha_n\}$ converges to α weakly (in $C(I)^*$, statement (33)), *and* if

$$\|\alpha_n\|_V \to \|\alpha\|_V, \tag{34}$$

then $\{\alpha_n\}$ converges to α strongly (in $NBV(I)$, statement (32)), which then implies pointwise convergence

$$\alpha_n(s) \to \alpha(s) \tag{35}$$

for every $s \in I$. This has the biological interpretation that functionally similar enzymes with close overall shapes (total variations $\|\cdot\|_V$) must be structurally similar as well.

10. Stepped-helical Enzymes. Since a stepped-mapping, i.e. stepped-helix, $\alpha \in S(I)$, has bounded variation, it acts as an enzyme when interacting with a substrate $F \in C(I)$ and produces the complex $\langle F, \alpha \rangle = \int_I F \, d\alpha$. We may assume without loss of generality that α is left-continuous, since changing the value of α at finitely many points (other than the end points) of I will not affect the value of $\langle F, \alpha \rangle$.

Thus a stepped-helical (approximating) enzyme α has the form

$$\alpha(s) = \begin{cases} c_1, & s_0 \leqslant s \leqslant s_1 \\ \sum_{j=1}^{k} c_j, & s_{k-1} < s \leqslant s_k ; k = 2, 3, \ldots, n \end{cases} \tag{36}$$

where $\{a = s_0 < s_1 < \ldots < s_n = b\}$ is a partition of $I = [a, b]$. (If $\alpha \in NBV(I)$ then of course $c_1 = 0$.) Note that c_j is the *difference* in the value of α on the subintervals $(s_j, s_{j+1}]$ and $(s_{j-1}, s_j]$, i.e. the 'jump' at $s = s_j$. With such a stepped-mapping α as integrator, the Stieltjes integral can be written as the finite series

$$\langle F, \alpha \rangle = \int_I F \, d\alpha = \sum_{k=1}^{n} F(s_k) c_k . \tag{37}$$

Thus among all the functional values of F, only those at the partition points

are used. In other words, the enzyme α identifies the *attachment sites* on the substrate F. Along these lines it is also possible to *construct* an inhibitor ϕ for the enzyme α: define $\phi(s_k)$, $k = 1, 2, \ldots, n$, such that

$$\sum_{k=1}^{n} \phi(s_k)c_k = 0 \tag{38}$$

(and let ϕ be continuous on neighbourhoods of s_k); then

$$\langle \phi, \alpha \rangle = \int_I \phi \, d\alpha = 0, \tag{39}$$

hence *inhibition*. Furthermore, $\langle F, \alpha \rangle$ may not exist if F is not continuous on neighbourhoods of the partition points s_k (allowing 'generalized' substrates $F \in C(I)$): an example in which the existence of the Stieltjes integral can be used to reflect the recognizability of the (potential) substrate F by the enzyme α.

The cardinality and density of the partition set $\{a = s_0 < s_1 < \ldots < s_n = b\}$ reveal the size and complexity of the enzyme α, which are related to its selectivity, or resolution power. This is because the larger and denser the number of partition points, the more difficult for $d(\alpha_F, \alpha_G)$ to be small, for the latter requires

$$\left| \sum_{k=1}^{m} F(s_k)c_k - \sum_{k=1}^{m} G(s_k)c_k \right| \tag{40}$$

to be small for $m = 1, 2, \ldots, n$. This cardinality–density argument provides a feasible explanation of the usual conception that the large size of an enzyme is required for its function, in particular for its specificity.

11. Integration by Parts. The assumptions on the integrand and the integrator are reversible. If F is of bounded variation and α is continuous on I, then the Stieltjes integral of F with respect to α also exists, and

$$\int_I F \, d\alpha = F(b)\alpha(b) - F(a)\alpha(a) - \int_I \alpha \, dF, \tag{41}$$

i.e.

$$\langle F, \alpha \rangle + \langle \alpha, F \rangle = F\alpha \Big|_a^b. \tag{42}$$

Equations (41) and (42) are formulae for *integration by parts* for the Stieltjes integral.

In Section 10 we saw that a Stieltjes integral becomes a series if the

integrator is chosen as a stepped-mapping and the integrand is chosen to be continuous at the partition points. The same result can be achieved for a continuous integrator and a stepped integrand, using integration by parts. Indeed, if $\{a = s_0 < s_1 \ldots < s_n = b\}$ is a partition of $I = [a, b]$, if α is continuous at these partition points, and if

$$F(s) = \begin{cases} u_1, & s_0 \leqslant s \leqslant s_1 \\ \sum_{j=1}^{k} u_j, & s_{k-1} < s \leqslant s_k;\ k = 2, 3, \ldots, n \end{cases} \tag{43}$$

then

$$\langle F, \alpha \rangle = \int_I F \, d\alpha = \sum_{k=1}^{n} u_k [\alpha(b) - \alpha(s_k)]. \tag{44}$$

Thus among all the functional values of α, only those at the partition points are used. In other words, the substrate F identifies the *active site(s)* on the enzyme α. Thus, it is also possible to *construct* an enzyme α, by defining its values continuously at $\alpha(s_k)$, to give a prescribed $\langle F, \alpha \rangle$ complex. In particular, these ideas lead to the theoretical possibility of synthesizing *artificial enzymes* which perform specified tasks.

Furthermore, when F is a unit-step function defined by

$$F(s) = \begin{cases} 0, & s \leqslant s_1 \\ 1, & s > s_1 \end{cases} \tag{45}$$

the Stieltjes integral $\langle F, \alpha \rangle$ has the value $[\alpha(b) - \alpha(s_1)]$ (if α is continuous at s_1). Thus one can obtain the functional values of the enzyme α at different s_1 positions via the Stieltjes integral. This gives a method of 'reconstruction' of an 'unknown' enzyme. The method is related to the classical 'moment problem' (where $F(s) = s^k$ for $k = 0, 1, 2, \ldots$), which is discussed in Akhiezer (1965). As a practical application, from a sequence of observed values $\mu_1, \mu_2, \ldots$ due to interactions between a sequence of substrates $F_1, F_2, \ldots$ and an 'unknown' enzyme, one can solve the equations

$$\langle F_k, \alpha \rangle = \mu_k \tag{46}$$

$k = 1, 2, \ldots,$ to obtain the functional representation α of the enzyme.

12. Enzyme–Substrate Duality. The integration by parts formulae (41) and (42) for Stieltjes integrals hold, in fact, without assuming anything about the

nature of the functions F and α except the existence of one of the two integrals. These formulae reflect a certain 'symmetry' inherent in enzyme-substrate interactive systems—we may regard both enzyme and substrate as *meters*, each observing the other. Note, however, that $\langle F, \alpha \rangle$ and $\langle \alpha, F \rangle$ differ in sign, as well as by the constant $F\alpha|_a^b$; the integration by parts formulae therefore do not give the most appropriate description of this enzyme-substrate symmetry. Rather, they point to an inherent *dissymetry* between enzymes and substrates.

A more natural description of this *duality* that exists between enzymes (elements of $C(I)^*$) and substrates (elements of $C(I)$) can be derived from the concept of *second dual space.* The use of the notation $\langle F, \alpha \rangle$ in place of $\hat{\alpha}(F)$, as we have done, is convenient to denote this enzyme-substrate duality: the action of α on F on the one hand and the action of F on α on the other.

Let us first discuss some general concepts. For a normed space X, the dual space X^*, the space of all bounded linear functionals on X, has as its norm the operator norm defined by

$$\|x^*\| = \sup\{|\langle x, x^* \rangle| : \|x\| \leqslant 1\} \tag{47}$$

for every $x^* \in X^*$ (re. equation (12)). The operator norm turns X^* into a Banach space. Dually, for every $x \in X$, the norm of x admits the alternate description

$$\|x\| = \sup\{|\langle x, x^* \rangle| : \|x^*\| \leqslant 1\} \tag{48}$$

Consequently,

$$x^* \mapsto \langle x, x^* \rangle \tag{49}$$

is a bounded linear functional on X^*, of norm $\|x\|$.

The normed dual X^* of a Banach space X, therefore, is itself a Banach space and hence has a normed dual of its own, denoted by X^{**}, called the *second dual space* of X. Statement (48) above shows that every $x \in X$ defines a unique $\phi x \in X^{**}$ by the equation

$$\langle x, x^* \rangle = \langle x^*, \phi x \rangle \tag{50}$$

for every $x^* \in X^*$, and that

$$\|\phi x\| = \|x\| \tag{51}$$

for every $x \in X$. It follows from (50) that $\phi: X \to X^{**}$ is linear, and from (51) that ϕ is an isometry. Furthermore, $\phi(X)$ is closed in X^{**}. Thus ϕ is an isometric isomorphism of X onto a closed subspace of X^{**}. Frequently, X is identified with $\phi(X)$ and hence X is regarded as a subspace of X^{**}.

If $\phi(X) = X^{**}$, then the space X is called *reflexive* (for example, all

L^p-spaces with $1 < p < \infty$ are reflexive). It may also happen that $\phi(X)$ is a proper subspace of X^{**}; the space with which we are concerned, namely $X = C(I)$, is an example.

Thus an enzyme $\alpha \in NBV(I) = C(I)^*$ acts as an operator on a substrate $F \in C(I)$, while dually the substrate $F \in C(I) \subset NBV(I)^* = C(I)^{**}$ acts as an operator on the enzyme α. So the bilinear functional (7) admits the dual description

$$\langle \cdot, \cdot \rangle : \left\{ \begin{array}{l} (F, \hat{\alpha}) \in C(I) \times C(I)^* \\ \\ (\alpha, \phi F) \in NBV(I) \times NBV(I)^* \end{array} \right\} \mapsto \int_I F \, d\alpha. \qquad (52)$$

$$= C(I)^* \times C(I)^{**}$$

Since $C(I)$ is not reflexive, 'substrates', i.e. elements of $C(I)$, do not provide all the operators on 'enzymes', i.e. elements of $NBV(I)$. This suggests the existence of a special class of *enzyme regulators* (in $C(I)^{**} \sim C(I)$). In particular, this class includes the proteolytic enzymes (i.e. the proteases and peptidases).

13. Enzymic Catalysis. Many one-substrate reactions of enzymes involve several complexes, thus the mechanism of enzyme-substrate interaction can be represented as

$$e + S \rightleftharpoons eS \rightleftharpoons e_1 Z_1 \rightleftharpoons e_2 Z_2 \rightleftharpoons \ldots \rightleftharpoons e'P \rightleftharpoons P + e' \qquad (53)$$

where eS is the enzyme-substrate complex, $e_k Z_k$ the sequence of intermediate complexes, and $e'P$ the enzyme-product complex. P is the *product* of the interaction between the enzyme e and substrate S, and e' is the 'modified enzyme', with an induced morpholoty, from which the enzyme e can be recovered (by definition an enzyme is a catalyst). In particular, therefore, $d(e, e')$ must be relatively small.

The kinetic evolution from eS, through the intermediate complexes, to $e'P$ can be formulated mathematically via the concept of *adjoint*. Let X and Y be Banach spaces. For a bounded linear transformation $T: X \to Y$, its *adjoint* $T^*: Y^* \to X^*$ is defined by

$$T^*(y^*) = y^* \circ T \in X^* \qquad (54)$$

for every $y^* \in Y^*$. Then for $x \in X$ and $y^* \in Y^*$

$$\langle x, T^*(y^*) \rangle = \langle Tx, y^* \rangle. \qquad (55)$$

If $y^*, z^* \in Y^*$ and $\|y^* - z^*\| < \epsilon$ for $\epsilon > 0$, then for $x \in X$,

$$|\langle Tx, y^* \rangle - \langle Tx, z^* \rangle| = |\langle Tx, y^* - z^* \rangle| = |\langle x, T^*(y^* - z^*) \rangle|$$

$$\leqslant \|x\| \|T^*\| \|y^* - z^*\| < \|x\| \|T^*\| \epsilon. \qquad (56)$$

It is this continuity property that is useful in the modelling of the evolution of complexes.

Let $X = Y = C(I)$ be collections of substrates, intermediates, and products. Then T: $C(I) \to C(I)$, i.e. $T \in L(C(I))$, maps a substrate F to a potential product TF. The adjoint T^*: $NBV(I) \to NBV(I)$ maps an induced enzyme α', which couples with TF, to the original enzyme α. Thus given a prescribed upper allowable limit $\epsilon > 0$ of variation on the enzyme, an enzyme α and a substrate F, for each substrate-to-(intermediate) product transformation T, one obtains the set of complexes

$$\{\langle TF, \alpha' \rangle : \|\alpha' - \alpha\|_V < \epsilon\}. \tag{57}$$

The selection criterion for the intermeidate (product) can be formulated as

$$\sup\{|\langle TF, \alpha' \rangle| : T \in L(C(I)),\ \|\alpha' - \alpha\|_V < \epsilon\}. \tag{58}$$

These suprema can be taken sequentially, which models the sequential nature of the evolution of complexes from eS to $e'P$.

This mathematical model of enzymic catalysis, an evolution of enzyme-[substrate, intermediate, product] complexes, is depicted in Fig. 3.

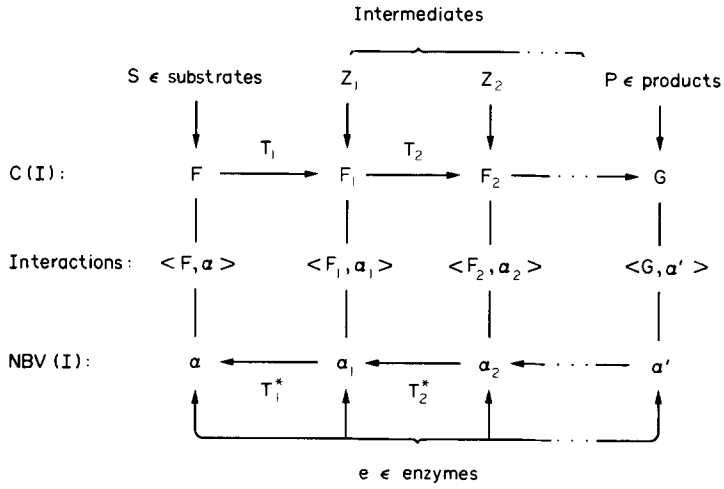

Figure 3.

The same model, with trivial modifications, can also be used to describe multienzyme-multisubstrate chain reactions. Louie *et al.* (1982) present a phenomenological calculus for enzyme-substrate interactive processes. They provide an alternate description of enzymic catalysis using the concept of *response tensor*. The reader is referred to their paper for details. Other phenomenological connections of the present Stieltjes integration model will be discussed in a forthcoming paper.

14. Generalization and Prospects. The proposed theory of enzyme–substrate interactions is based on a combination of mathematical concepts from differential geometry and functional analysis. We saw how the simplest ideas from these topics in mathematics produce a model of the enzyme–substrate interactive process which captures many of the biochemical features of enzymic catalysis. In this final section we discuss further mathematical analogues of other aspects of the biology of enzymes.

In Louie and Somorjai (1982) we presented a dual description of proteins on two differential-geometric hierarchical levels: space curves and surfaces. The characterization of space curves by curvature and torsion leads to the present formulation of the Stieltjes integrator $\alpha = \kappa + i\tau$: $I \subset \mathbb{R} \to \mathbb{C}$. We can similarly use the characterization of surfaces by the two corresponding invariants, *the first and second fundamental forms* I and II, and represent an enzyme by the function $A = \mathrm{I} + i\mathrm{II}$, which has the two-real-dimensional tangent bundle of the surface as domain, and a two-complex-(= four-real-) dimensional space as range. The enzyme–substrate interaction is then described by the Stieltjes integral

$$\int_U F \, \mathrm{d}A \tag{59}$$

over a region $U \subset \mathbb{R}^2$. (The substrate F would still be a function of two real variables, say a regular parametrized surface embedded in $\mathbb{R}^3$, representing the 'shape' of the substrate. Recall Section 6.)

Other functional-analytic concepts which have biological interpretations in enzymic catalysis include:

(a) *annihilators* (elements in dual spaces which vanish on elements in base spaces, and vice versa), which give mathematical analogues of inhibitors, metabolic poisons, etc.;

(b) *bounded operators*, which can be used to describe transformations of enzymes and substrates;

(c) *adjoints*, generalizations of dualities, which further characterize the symmetry and *dissymmetry* of enzymes and substrates; and

(d) *metrics*, which give alternative descriptions of the 'distances' between molecules via the various L^p, operator, etc., definitions, and which can be used to assess the similarity of biological molecules.

15. Review: Correspondence between Mathematics and Biology. Let us close by tabulating the correspondence of the biological and mathematical concepts presented in this paper.

TABLE I

Biology	Mathematics
Protein	Space curve; Pair of functions (κ, τ); Complex-valued function of a real variable, sectionally continuous and of bounded variation, $\alpha = \kappa + i\tau$
Substrate	Complex-valued continuous function of a real variable, $F \in C(I)$
Enzyme	Complex-valued normalized function of bounded variation, $\alpha \in NBV(I)$; Bounded linear functionals on $C(I)$, $\hat{\alpha} \in C(I)^*$
Enzyme–substrate interaction	Stieltjes integral $\langle F, \alpha \rangle = \int_I F \, d\alpha$
Observable of enzymic catalysis	Projections of $\langle F, \alpha \rangle$ to the reals
Structural (fibrous) proteins	Constant α; $\hat{\alpha} = 0$
Cofactor; modifier; inhibitor Apoenzyme Holoenzyme; modified/inhibited enzyme	$\alpha \in C(I)$ $\alpha \in NBV(I)$ $\alpha_\phi(S) = \int_a^S \phi \, d\alpha$.
Induced variation on enzymes; Similarity between enzymes	Total variation metric $\| \cdot \|_V$
Lateral translation/movement of substrate on enzyme	Stieltjes convolution $\langle F_s, \alpha \rangle = \int_I F(s - \cdot) \, d\alpha$
Recognizability Recognition Discrimination	Stieltjes integrability value of $\langle F, \alpha \rangle$ $\lvert \langle F, \alpha \rangle - \langle G, \alpha \rangle \rvert$
Structural similarity	Uniform convergence; Pointwise convergence
Functional similarity	Weak convergence
Structure implies function	Uniform convergence $\Rightarrow$ weak convergence
Function dictates structure	Weak + norm convergence $\Rightarrow$ pointwise convergence
Approximating enzymes	Stepped-helices
Enzyme–substrate dissymmetry	Integration by parts
Enzyme–substrate symmetry	Duality of Banach spaces
Qualitative analysis of unknown enzyme	Moment problem
Mechanism of enzymic catalysis	Transformations and their adjoints

LITERATURE

Akhiezer, N. J. 1965. *The Classical Moment Problem.* Edinburgh: Oliver & Boyd.

Arnold, V. I. 1978. *Mathematical Methods of Classical Mechanics.* New York: Springer-Verlag.

Burrill, C. W. and Knudsen, J. R. 1969. *Real Variables.* New York: Holt, Rinehart, & Winston.

Drenth, J., Kalk, H. H. and Swen, H. M. 1976. "Binding of Chloromethylketone Substrate Analogues to Crystalline Papain." *Biochemistry* **15**, 3731–3738.

Edelstein, L. and Rosen, R. 1978. "Enzyme-substrate Recognition." *J. theor. Biol.* **73**, 181–204.

Louie, A. H. and Somorjai, R. L. 1982. "Differential Geometry of Proteins: a Structural and Dynamical Representation of Patterns." *J. theor. Biol.* **98**, 189–209.

—— and ——. 1983. "Differential geometry of Proteins: Helical Approximations." *J. molec. Biol.* **168**, 143–162.

——, Richardson, I. W. and Swaminathan, S. 1982. "A Phenomenological Calculus for Recognition Processes." *J. theor. Biol.* **94**, 77–93.

Mackey, G. W. 1963. *The Mathematical Foundations of Quantum Mechanics.* New York: Benjamin.

Rashevsky, N. 1960. *Mathematical Biophysics.* New York: Dover.

Rosen, R. 1978. *Fundamentals of Measurement and Representation of Natural Systems.* New York: North-Holland.

Rudin, W. 1976. *Principles of Mathematical Analysis.* New York: McGraw-Hill.

Widder, D. V. 1946. *The Laplace Transform.* Princeton: Princeton University Press.

RECEIVED 8-8-83