INTERPRETING THE BEHAVIOR OF ENZYMES PURPOSE OR PEDIGREE?

Authors: **Steven Benner**
Organische Chemie
Eidgenössische Technische Hochschule
Zürich, Switzerland

Andrew D. Ellington
Department of Biochemistry and Molecular Biology
Harvard University
Cambridge, Massachusetts

Referee: Gerald D. Fasman
Department of Biochemistry and Molecular Biology
Brandeis University
Cambridge, Massachusetts

I. INTRODUCTION

In the last 2 decades, biochemists have compiled a remarkable collection of data describing the structure and behavior of enzymes. Virtually no detail of the structure and behavior of proteins has been considered too small to escape experimental attention.

Protein structure has been probed in extraordinary depth. Data bases now contain over 4000 sequences of proteins. Crystal structures of over 500 individual proteins are available at moderate resolution or better.[1] Structures range in size from small oligopeptides to complete viruses. For many enzymes, three-dimensional structures are available where substrate or substrate analogs are bound in the active site. In a few cases, low-temperature crystallography has trapped reaction intermediates in the active site, and structures of these intermediates are available as well.[2]

Only slightly less appreciated is the wealth of data describing the physical and catalytic behavior of enzymes. Macroscopic kinetic parameters, such as K_m and V_{max}, have been measured for an uncounted number of enzymes.[3] Kinetic mechanisms, which provide a description of the order of the events in a catalytic reaction, have been reported for hundreds of these, along with selected microscopic rate constants describing the rates of individual reaction steps in the active site. In a few cases, biochemists can boast of having a nearly complete "free energy profile" for an enzymatic reaction, a quantitative description of all kinetic events that take place during the catalytic process.[4]

At the level of chemical mechanism, an enormously sophisticated body of literature has also developed. The stereochemical course of nearly 1000 enzymatic reactions has been described in detail.[5] Probes of transition-state structure have suggested conclusions regarding the number of transition states and intermediates in the chemical transformation occurring in the active site.[6] The state of the art is sufficient to support arguments about the nature of the transition state for these reactions. In several cases, biochemists are close to accounting, at least semi-quantitatively, for the entire amount of rate acceleration produced by enzymes.

The physical properties of biological macromolecules have also not been neglected. Proteins have been the targets of virtually every type of modern spectroscopy. Physical processes occurring in proteins ranging in scale from a microsecond to a second have been described. The microscopic motions of proteins with time constants on the order of nanoseconds have been measured.[7] Sophisticated calculations have been used to examine and predict protein structure and behavior,[8] and the effect of small changes in structure on that behavior.

Patterns of structural diversity and similarity in proteins, especially differences in the primary structures of proteins, have been analyzed intensively over the last 20 years.[9,10] However, less well appreciated are the examples of behavioral diversity displayed by proteins. Following are some examples that may be regarded as provocative for the purpose of this review (see Figure 1).

A. Mechanism

Enzymatic transfer of a phosphoryl group from ATP to ADP proceeds via an enzyme-phosphoryl intermediate. However, transfer of a phosphoryl group from ATP to *AMP* is direct; there is no intermediate.[11] Histidine decarboxylases from mammals and *Escherichia coli* use pyridoxal phosphate as a cofactor.[12,13] However, histidine decarboxylases from *Lactobacillus* and *Micrococcus* use an enzyme-bound pyruvyl residue.[14]

B. Stereochemistry

Alcohol dehydrogenase from yeast transfers the pro-R hydrogen at the 1-position of ethanol to NAD^+. However, alcohol dehydrogenase from *Drosophila* transfers the pro-S hydrogen at the 1-position of ethanol to NAD^+.[15]

Acetolactate decarboxylase catalyzes the decarboxylation of acetolactate with net inversion of configuration, malic enzyme effects the decarboxylation of malate with net retention of configuration, and acetoacetate decarboxylase effects decarboxylation with net racemization of configuration.[16] Yet all three enzymes catalyze the same fundamental reaction, the decarboxylation of a beta-ketoacid.

C. Kinetics and Thermodynamics

The internal equilibrium constant of lactate dehydrogenase from rabbit muscle appears to favor the enzyme-lactate-NAD^+ ternary complex. The internal equilibrium constant of lactate dehydrogenase from mammalian heart appears to favor the enzyme-pyruvate-NADH ternary complex.[17,18]

D. Macromolecular Structure

The amino acid sequence of ribonuclease has diverged at 70% of the positions in the time since rat, whale, cow, and kangaroo diverged.[19] In the same time, the amino acid sequence of cytochromes has hardly diverged at all.[20]

Within a population of *Drosophila*, there is hardly any variation of the structure of the alcohol dehydrogenase.[21] However, in the same population, the esterase is highly polymorphic.[22]

E. Substrate Structure

Nicotinamide adenine dinucleotide, flavin adenine dinucleotide, S-adenosyl methionine, coenzyme A, and thiamin pyrophosphate all have a ribonucleotide portion, while biotin, pyridoxal, and lipoic acid do not.[23] Cofactors involved in the synthesis and degradation of fatty acids come in a range of structures, all related, some with ribonucleotide portions and some without.[24]

These data are tabulated in many reviews, which are available for those seeking examples of the peculiarities of living systems or, perhaps better, the ingenuity of biological chemists in elucidating these peculiarities. However, as the focus has been primarily on developing methods to elucidate such behaviors, there has been remarkably little focus on the biological significance of the data per se. This review seeks to fill this gap.

Nothing in biology can be understood except in the light of evolution. This is true even for most chemical details of enzymatic behavior. Macromolecular structure and behavior are products of two contrasting evolutionary processes, natural selection and neutral drift.[25,26] Macromolecular traits are selected if they influence the survival of a host organism; selected

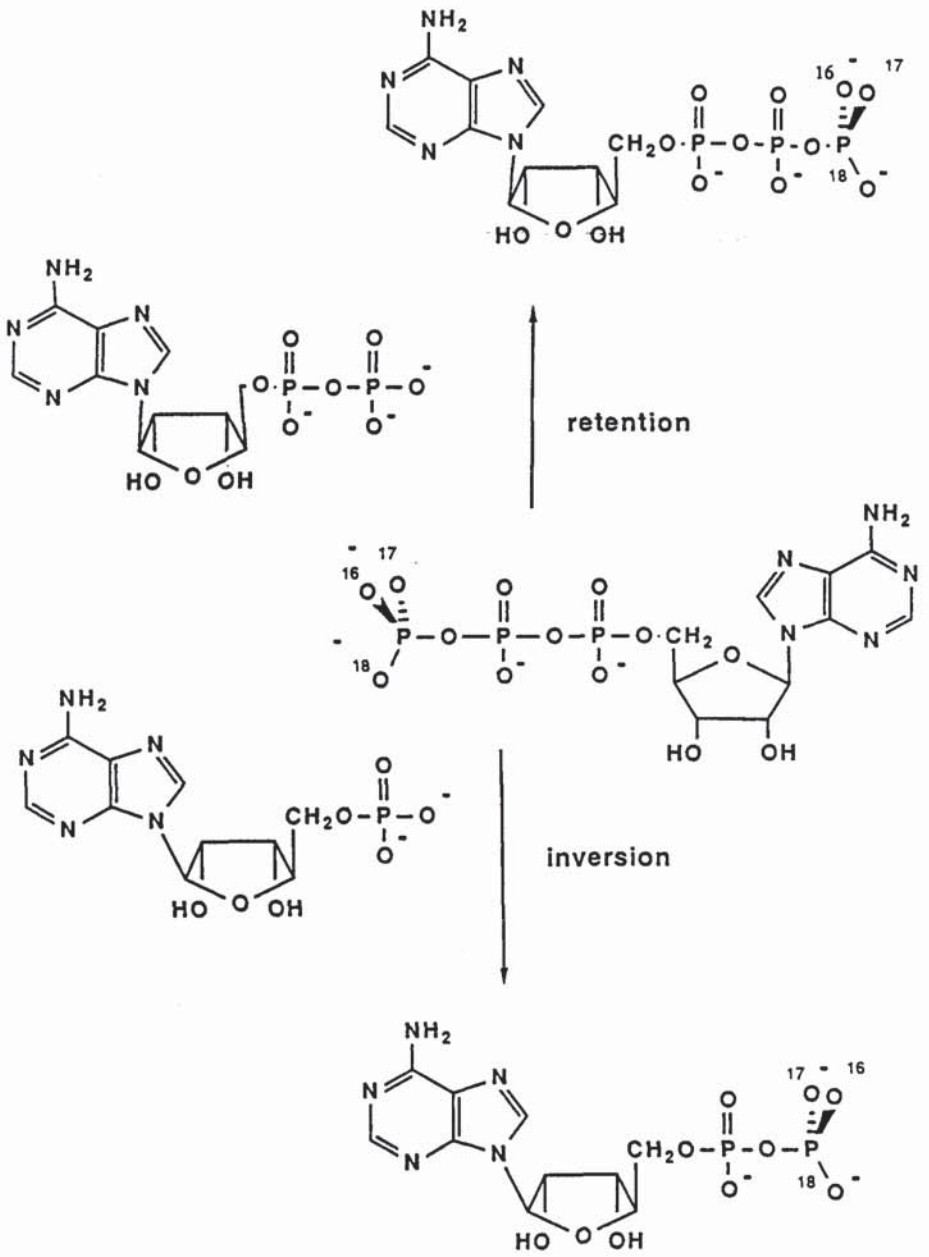


FIGURE 1A. Some representative examples of diversity in behavior in biological macromolecules are depicted in Figures 1A to F. The mechanism for enzymatic transfer of a phosphoryl group from ATP to ADP is different from the mechanism for the transfer from ATP to AMP. The former proceeds via an enzyme-phosphoryl intermediate; the latter transfer is direct.[11] Yet the two reactions are highly analogous, both chemically (both are S_n^2 displacements at phosphorus, with phosphorus as the attacking nucleophile) and structurally (the substrates are all nucleotides).

traits reflect biological function. Traits that do not influence survival of the entire host organism are not selected. Variation in these traits is termed "neutral" and will drift as the structure of the protein diverges.

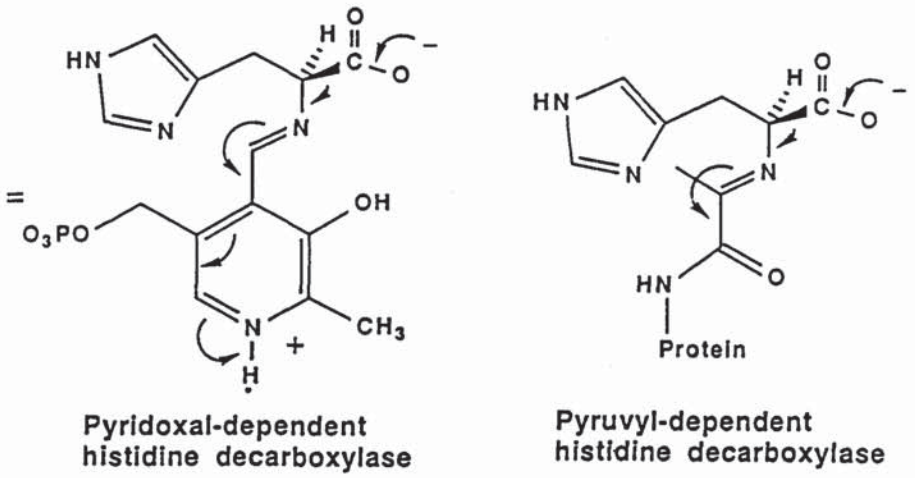


FIGURE 1B. The mechanism for the decarboxylation of histidine in mammals and *E. coli* is different from the mechanism in *Lactobacillus* and *Micrococcus*. Enzymes in the first two organisms use pyridoxal phosphate as a cofactor.[12,13] In the second two organisms, an enzyme-bound pyruvyl residue is used.[14] Otherwise, The reactions appear identical.

The distinction is complicated by the possibility that neutral behavior in proteins is coupled structurally to selected behaviors. For example, it is possible that the stereospecificity of dehydrogenases alluded to previously serves no intrinsic function, but cannot be reversed without damaging structures that are responsible for selected behavior (e.g., catalytic activity). Thus, it is conceivable that nonselected behaviors will be functionally constrained from drifting.

Thus, at a molecular level, it is very difficult to distinguish selected and nonselected traits. If the behaviors of two homologous enzymes from different sources are the same, the behavior can be with equal plausibility *a priori* (1) functional (in which case it was conserved for functional reasons) or (2) nonfunctional (in which case it was conserved because it was structurally coupled to another, functional trait). If the behaviors of two homologous enzymes are different, it seems equally plausible *a priori* that the behaviors are (1) functional (in which case the different behaviors represent adaptation to different environments of the homologous enzymes) or (2) nonfunctional (in which case the difference reflects behavioral drift).

Nevertheless, the distinction remains important, as the interpretation of macromolecular behaviors depends on whether or not they reflect selection. Behaviors that are the products of natural selection can be interpreted in terms that reflect biological function; behaviors that are not the products of natural selection can only be interpreted in terms of events in the historical past. In both selecting macromolecular behaviors to study and in interpreting the results of those studies, we must know whether the similarities and differences in macromolecular behavior reflect functional adaptation in response to natural selection or nonfunctional, perhaps random, historical accidents.

This article reviews the current literature that forms the basis of an effort to answer these questions. We first consider methods for formulating hypotheses that explain the structure and behavior of biological macromolecules in terms of two constrasting types of models, "functional" models and "historical" models. Then, we focus on recent experimental results that suggest conclusions about the interaction between natural selection, macromolecular behavior, and functional constraints on behavioral drift.

F. Why Bother?

As we shall see, it is difficult to determine whether historical or functional models better

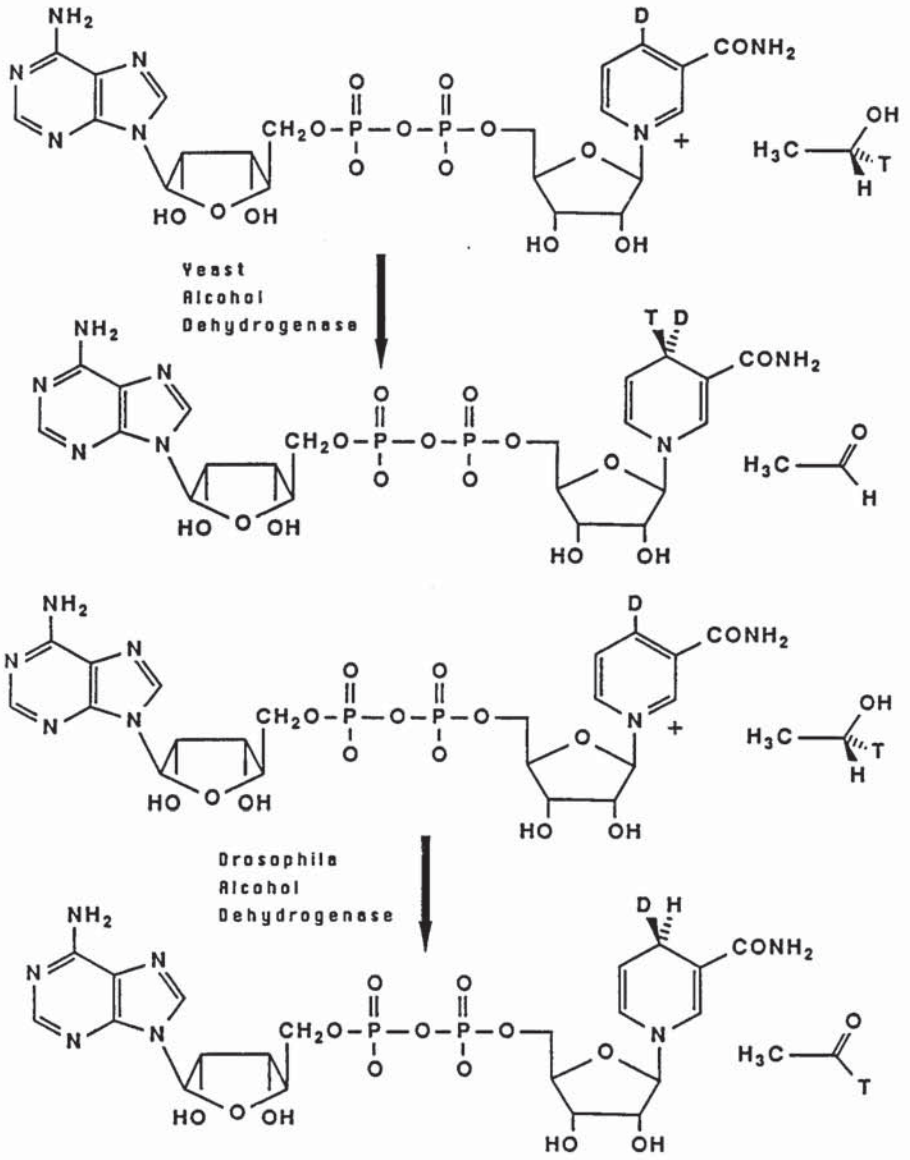


FIGURE 1C. Alcohol dehydrogenase from yeast transfers the pro-R hydrogen at the 1-position of ethanol to NAD^+. However, alcohol dehydrogenase from *Drosophila* transfers the pro-S hydrogen at the 1-position of ethanol to NAD^+.[15] Otherwise, the reactions appear identical.

explain particular behaviors of particular enzymes. Arguments are often inconclusive and require evaluation of possibly incomplete data. Therefore, one might well wonder whether it is worth addressing the question at all.

Aside from the intrinsic intellectual interest in protein behavior and evolution, there are three special reasons why this question is important to examine.

First, understanding protein evolution should be valuable to guide the biological chemist seeking to engineer the physical and catalytic behavior of proteins. The tools for "engineering"

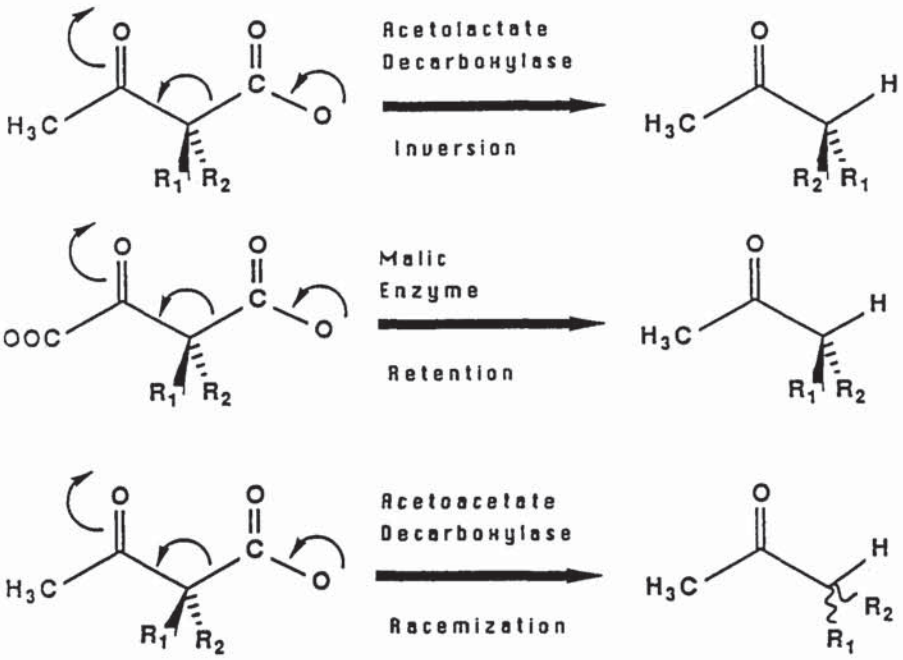


FIGURE 1D. Acetolactate decarboxylase catalyzes the decarboxylation of acetolactate with a net inversion of configuration; malic enzyme effects the decarboxylation of malate with a net retention of configuration, and acetoacetate decarboxylase effects decarboxylation with a net racemization of configuration.[16] Yet all three reactions are fundamentally the same, i.e., the decarboxylation of a beta-ketoacid.

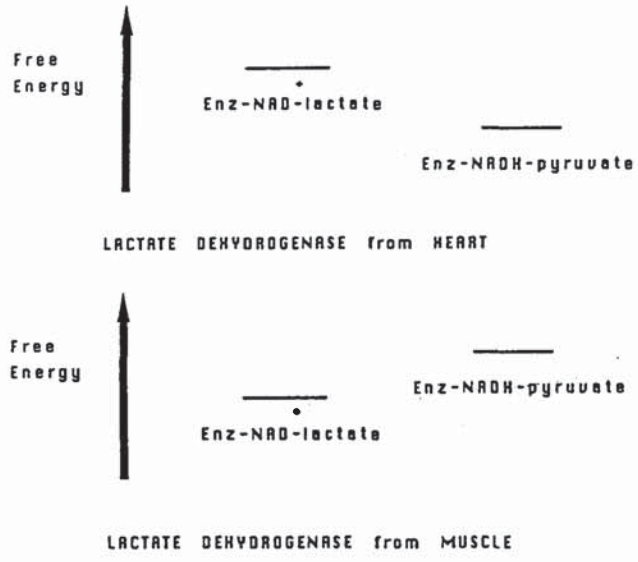


FIGURE 1E. The internal equilibrium constant of lactate dehydrogenase from rabbit muscle appears to favor the enzyme-lactate-NAD^+ ternary complex. The internal equilibrium constant of lactate dehydrogenase from mammalian heart appears to favor the enzyme-pyruvate-NADH ternary complex.[17,18] Otherwise, the enzymatic reactions appear the same.

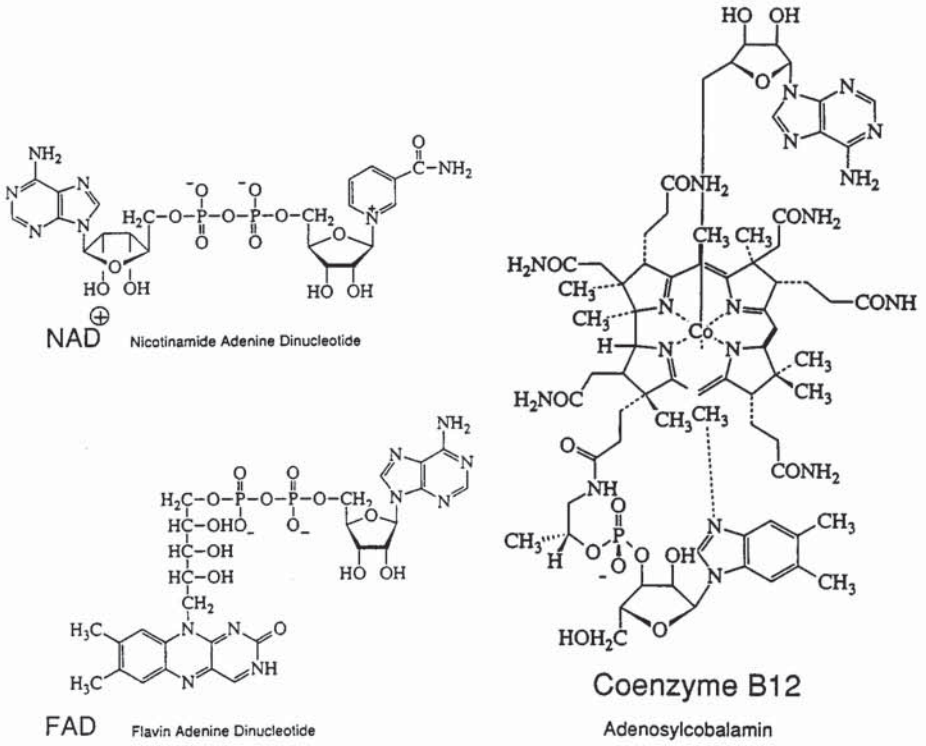


FIGURE 1F. Nicotinamide adenine dinucleotide, flavin adenine dinucleotide, S-adenosyl methionine, coenzyme A, and thiamine pyrophosphate all have a ribonucleotide portion. Biotin, pyridoxal, and lipoic acid do not.[23]

the behavior of proteins are now widely available, as molecular biology enables the deliberate manipulation (at least in principle) of the structure of genes and proteins. However, the problem is enormously complex, given the number of possible alterations that can be made in a protein's sequence. Here, "evolutionary guidance" might be of help. Although enzyme engineering is new in the laboratory, it is not new on the planet. Understanding how evolution developed proteins with altered physical and catalytic abilities should help guide biochemists as they attempt to do the same.

Second, understanding evolution in protein catalysts may shed light on underlying principles in chemistry. Macromolecular behaviors that are the products of natural selection have been refined by billions of years of biological evolution. This time, corresponding approximately to 2×10^8 graduate student Ph.D theses, represents a level of experimentation that is unattainable in the laboratory. Traits of biological macromolecules that have been optimized by natural selection may well be informative about chemical principles.

In contrast, traits that are explained historically reflect the conservation of random accident. Understanding how these accidents occur and are conserved will not shed light on underlying chemical principles. However, which accidents are conserved during divergent evolution may suggest information about constraints on macromolecular evolution and may provide information on intrinsic properties of biological macromolecules. Furthermore, studies of historical models for protein evolution may allow studies of macromolecular behavior to join with macromolecular structure as tools for elucidating the pattern of organism evolution.

Finally, bioorganic chemists frequently argue that detailed information on the behavior of macromolecules is a necessary prerequisite for the "rational" treatment of human disease. This

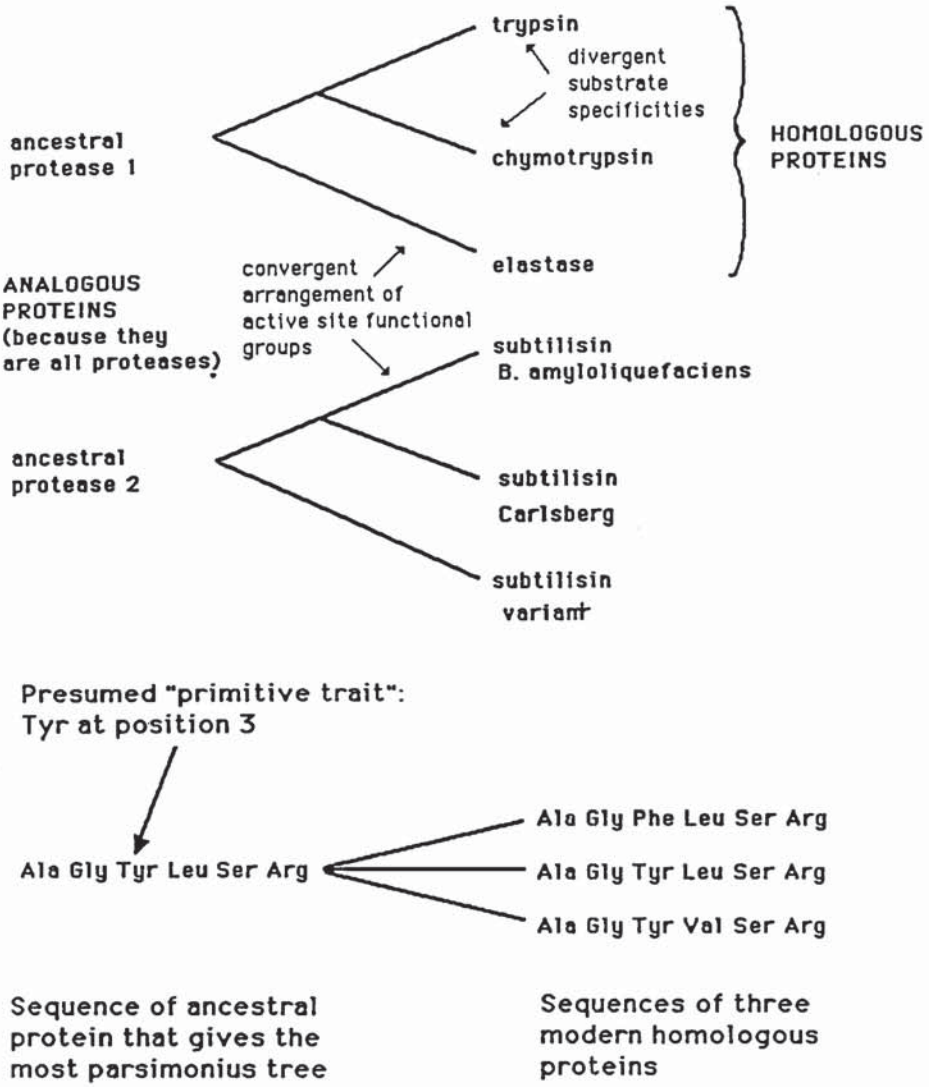


FIGURE 2. The nature of the evolutionary trees describing the pedigree of groups of proteins determines whether any two proteins are homologous or nonhomologous.

may indeed be true. However, exactly how bioorganic data are applied to "rational" medicine is largely dependent on whether the behavior is viewed as adaptive or nonadaptive.

II. DEFINITIONS

This review explores the explanations for the behavior of biological macromolecules of two general classes, "historical" and "functional". Much of the language that we use comes from evolutionary biology, a field unfamiliar to many bioorganic chemists. Therefore, some definitions are in order (Figure 2).[9,10]

Two proteins are "homologous" if they share a common ancestor. This descriptor makes no statement about similarities or differences in any other property of the two proteins.

Two proteins are "analogous" if they are similar in some specified way, e.g., in function, behavior, or mechanism. This descriptor makes no statement about the relatedness of the two proteins.

Behaviors that are the same in two nonhomologous macromolecules are said to have arisen by "convergent evolution". Behaviors that are different in two homologous macromolecules are said to have "diverged". Behaviors that are the same in two homologous macromolecules have been "conserved". None of these statements carry any implication as to whether the convergence, divergence, or conservation was adaptive.

Furthermore, it is frequently difficult to tell whether two macromolecules are homologous. Therefore, there is often a corresponding uncertainty whether identical traits arose by convergent evolution or are conserved vestiges of a common ancestor. Conversely, there is often uncertainty as to whether behaviors that are different reflect divergent evolution or independent origin.

In constructing "evolutionary trees" that assign a "pedigree" to a homologous set of enzymes, one often examines molecular "traits" and constructs a "parsimonious" tree. A parsimonious tree is the tree that connects a set of enzymes so that the minimum number of evolutionary events must be presumed to create the existing diversity within the set. Trees are often constructed considering a single trait (e.g., the amino acid sequences of a set of homologous proteins). Parsimonious trees that consider different traits may yield trees with different "connectivities", and the connectivity of different species based on a tree of a single protein need not reflect the actual evolutionary relationship of the species themselves.[26]

A "primitive trait" is a behavior that was present in an ancestral organism. If a primitive trait is identified, it can be useful in determining homology. A "derived" trait is a trait that is not primitive.

"Polymorphism" refers to the existence in a population of several structural variants of a protein. Usually, all of the variants display biological activity. Polymorphism is distinguished from "heterozygosity", which refers to different structural genes on each chromosome of a pair in a single polyploid organism, and "isozymes", which implies the existence of multiple genetic loci for a protein, each performing a separate function within the same organism. Again, polymorphism, heterozygosity, and isozymes may or may not reflect functional adaptation.

A functional explanation interprets the behavior of a biological macromolecule by assigning a purpose or function to that behavior. As Darwinian selection at the level of the organism is presumed to be the only mechanism for obtaining functional behavior in living systems, a functional explanation presumes that, were the behavior different, the organism that is the host of the macromolecule would be less able to survive and reproduce.

A historical explanation interprets the behavior of a biological macromolecule in terms of the macromolecule's pedigree, its immediate past, and accidents of history incorporated into the macromolecular structure as the macromolecule evolved. Historical models require assumptions about the behavior of ancestral macromolecules, the interrelatedness of various macromolecules, and the degree to which the behavior in question might have diverged or been conserved.

III. METHODOLOGY

A. Intuition, Preconceptions, and Prejudices

We must begin with a general statement that adaptation at the molecular level is a "virgin field".[27] As Lewontin has noted,[28] "It has proved remarkably difficult to get compelling evidence for changes in enzymes brought about by selection, not to speak of adaptive changes." Nevertheless, it is difficult to find a biological chemist who does not hold strong opinions on at least some aspect of molecular evolution. One of the themes of this review is that very often what appears obvious is in fact often at the heart of a valid, and often hotly contested, scientific dispute. Therefore, they cannot be simply presumed to be true.

For example, bioorganic chemists often presume that a pair of enzymes that are analogous in some behaviors are analogous generally. In this view, the malate dehydrogenase from yeast is "the same enzyme" as the malate dehydrogenase from *E. coli*, a view that suggests that the behaviors of the enzyme from *E. coli* are generally the same as those of the enzyme from yeast. Such a presumption is encouraged by the classification of enzymes by reaction type and substrate specificity by the Enzyme Commission, the officiating body for nomenclature in the field.[3]

Of course, this presumption is not without basis. For example, as Dixon and Webb noted many years ago:[29] "It is a remarkable fact that in general the catalytic properties, specificity, affinities, etc. of a given enzyme vary little with the source. Although there may be slight physical differences in a given enzyme when it is produced by different cells they are usually unimportant, and the enzyme remains essentially the same enzyme."

"Given enzyme" here refers to a set of proteins that catalyze analogous reactions. However, "slight" and "unimportant" are terms that reflect no more than a presumption on the part of a biochemist, not the realities of natural selection. In fact, we rarely know what variations in macromolecular behavior are important to a living organism struggling to survive and reproduce. Therefore, two macromolecules should be considered "the same enzyme" only when they have the same amino acid sequence. The fact that a behavior of one protein is known does not mean that it is known for a functionally analogous protein with a different sequence. It is our task to decide which behavioral differences observed in macromolecules with different structures are in fact important or unimportant, and which variations in behavior are functional.

Two other (contradictory) assumptions are frequently made: (1) a similarity in behavior in two macromolecules (analogy) is necessarily evidence for homology and, conversely, (2) analogous behaviors in two macromolecules are necessarily evidence that the behaviors arose as a convergent, adaptative response. Neither is appropriate as a general statement. Assigning homology, or deciding which traits are conserved, is a difficult task that requires the distinction between primitive and derived macromolecular traits. The fact that two malate dehydrogenases both act on S-malate is neither a strong argument for relatedness, nor a strong argument for functional adaptation since we do not know *a priori* whether the trait (acting on S-malate) is primitive.

B. Model Construction

Another theme of this review is that distinguishing between adaptive and nonadaptive behavior in enzymes requires a careful, logically rigorous process of model building and testing. The goal is to construct hypotheses, in the form of historical or functional models, concerning the behavior of biological macromolecules. These models are then tested by experiment. Much of the bioorganic discussion of the evolution of proteins is flawed by an absence of models that are logically consistent, clearly expressed, and consistent with known fact.

Whether a model is good or bad depends both on how consistent it is with known fact and how readily it suggests experiments that might test it. In the latter regard, functional and historical models each have their own characteristic problems, and the value of each depends in large part on how it avoids those problems.

IV. FUNCTIONAL MODELS

A. Natural Selection and Neutral Drift

Natural selection is presumably the only mechanism for obtaining functional behavior in living systems. A functional interpretation of a specific behavior therefore must assume that variation in that behavior influences the survival of the host organism.

In formulating functional theories, the biochemist must recognize that any statement that a subtle enzymatic behavior is adaptive is instantly controversial. The reader is directed to the

literature[30] for examples of the often heated arguments generated in the biological community by these assumptions.

There is good reason for this controversy. Much of the evolution of macromolecules today is viewed in the context of the "Neutral Theory" of molecular evolution. First outlined by Kimura[31] and Jukes and Holmquist,[32] the "Neutral Theory" argues that most of the diversity in molecular structure among homologous proteins reflects random "drift", not functional adaptation. In this model, neutral mutations, those that do not influence the survival of an organism, occur randomly in a population, producing polymorphism. With a frequency depending on the population size, these mutations are randomly fixed. Through the random fixation of neutral mutations, the structure of a protein is expected to drift through geological time.

Regardless of the validity of the details of the Neutral Theory, one point is certain: macromolecular structure will drift if it is not constrained by a selectable function. As macromolecular behavior depends on macromolecular structure, behaviors too will drift if they are not functionally constrained.

There is little doubt that many gross behaviors of biological macromolecules are functionally constrained from drifting. Examples are known of mutations that destroy the catalytic activity of a protein that confer measurable selective disadvantage on a host organism. Thus, many of these behaviors undoubtedly do not drift.

However, many of the behaviors that biochemists presently study are not obviously linked to any selectable function. For example, there is no reason *a priori* to believe that it makes any difference to the survival of an organism whether its decarboxylases produce retention or inversion. It is equally not obvious that were the dynamic motions of lysozyme on a nanosecond time domain different, a cow would be less able to survive and reproduce. In the context of neutral theory, it is natural to view variation in these behaviors at this level as also neutral. It is these views that must be scrutinized.

At the heart of the relationship between enzymatic behavior and natural selection is the question: what enzymatic behavioral variation is sufficiently subtle that it does not influence the survival of the host organism? A major problem in constructing a functional model is to incorporate statements that address (1) how the drift of behaviors of macromolecules is constrained by function, and (2) exactly why such a subtle behavior of a single macromolecule is expected to influence the ability of the organism to survive and reproduce.

The prominent dispute between Neutral Theorists and Neo-Darwinists in molecular evolution for the last 2 decades calls for a high level of skepticism. This extends to the application of chemical principles that are well documented in reactions of small molecules to reactions in the active sites of enzymes. Enzymatic reactions are presumed to be governed by orbital symmetry,[33,34] stereoelectronic theory,[35,36] or "Baldwin's Rules".[37] Possible reaction mechanisms are often excluded on the grounds that these organic principles are violated.[33,35,37] Such exclusions may in fact be reasonable. However, they do presume that natural selection operates to select for enzymes that obey these rules, and this presumption is controversial.

The greater the domain of neutral theory, the greater the domain of "unreasonable" functional models. The boundary dividing those behaviors where functional models are appropriate from those where they are not is the boundary between selected and nonselected traits. Until that boundary is found, the bioorganic chemist must recognize the possibility that the traits that he studies, including stereochemical, kinetic, structural, and dynamic behaviors, may simply be irrelevant to biological function.

Thus, a functional model must assign the evolutionarily relevant function of the macromolecule being examined and must include statements explaining why the survival of a host organism depends on the enzyme behaving in one way as opposed to another.

B. Assigning Natural Function

A functional theory can only make predictions about, and thus can only be tested by,

behaviors of enzymes whose function is known. Unfortunately, there is no way to determine with certainty the evolutionarily relevant function of a protein. However, there are standard tests that can be applied to support an argument for or against a particular role for a particular enzyme. These tests can be divided into three groups:

1. Is the proposed physiological role clearly defined and plausible?
2. Are the properties of the enzyme *in vitro* consistent with the proposed physiological role?
3. Is the proposed physiological role consistent with *in vivo* data concerning the protein, especially with phenotypes of organisms with and without the enzyme?

A proposed physiological role for an enzyme can be little more than a hypothesis that it converts a specific substrate to a specific product. At the very least, the presumed physiological substrate should be present in the natural environment of the enzyme, at least under some circumstances. Indeed, if the substrate is not present physiologically, a strong argument can be made against the suggested role.

For example, a D-malic dehydrogenase has been reported in the eggs of sea urchins.[38] The enzyme is active only with the acetylpyridine analog of NAD^+ (Figure 3), which itself cannot serve as a cofactor. As it is unlikely that sea urchin eggs contain acetylpyridine analogs of NAD^+, the assignment of function to this enzyme is doubtful. Likewise, the suggestion that the enzyme dehydrocortisone reductase acts on dehydrocortisone is doubtful since dehydrocortisone (Figure 3) is not found in the organism where the enzyme is found.[39]

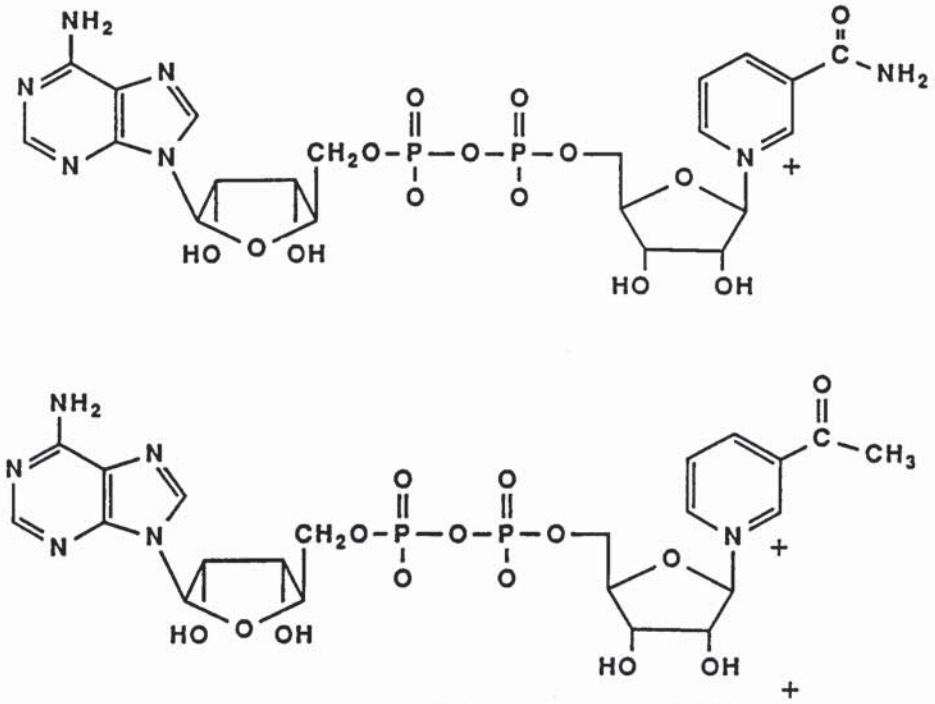
Such arguments are stronger if the enzyme occupies a clear role in a well-established metabolic pathway, especially if the pathway can be documented in the organism that is the source of the enzyme.

The *in vitro* behavior of the enzyme is also commonly used to support a proposal for a physiological role. Especially important in this regard are the *in vitro* kinetic properties of the enzyme, including the magnitude of various rate parameters, substrate specificity, and patterns of inhibition.

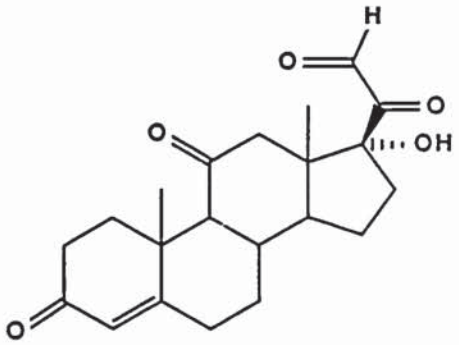
For example, if a purified preparation of the protein shows catalytic activity against only a single substrate and not against a range of structurally similar substrates, the narrowness of substrate specificity may be used as an argument to assign a biological function. The argument is not entirely reliable. For example, it was used to argue that lactaldehyde reductase acted physiologically on lactaldehyde.[40] One might conclude on other grounds that this argument is incorrect.[41] However, it is clear that catalytic activity against multiple substrates diminishes the plausibility of the assignment of functions. Examples from the recent literature have cast doubt on the physiological plausibility of functions assigned to 15-hydroxyprostaglandin dehydrogenase,[42] dihydrodiol dehydrogenase,[43] cortisone-5-beta-reductase,[44] and protein disulfide isomerase.[45]

Kinetic parameters are also frequently used to argue for or against biological function. For example, the Michaelis constant (K_M) for an alleged substrate is expected to be within an order of magnitude of the physiological concentration of the substrate. If the K_M of the assigned natural substrate is far from its physiological concentration, this fact may form the basis of a challenge to the assignment.

The magnitude of maximal velocities measured *in vitro* can also be used to assess the plausibility of an assignment of a natural substrate to an enzyme. The specific activity of enzymes catalyzing different types of reactions are available in the literature and can be compared to the kinetic properties of the enzyme in question. For example, typical dehydrogenases have specific activities between 1 and 10,000 mmol/min/g of enzyme for the reduction of carbonyls ranging in stability from acetoacetate (stable) to glyoxylate (unstable).[3] One might argue that enzymes having specific activities out of this range are not catalyzing the physiologically relevant reaction. For example, the 3-beta-hydroxysteroid dehydrogenase from



Acetyl pyridine analog of NAD



21-Dehydrocortisone

FIGURE 3. The acetylpyridine analog of NAD^+ and dehydrocortisone are two molecules probably not present in organisms physiologically which nevertheless are the substrates of certain enzymes.

Pseudomonas has a specific activity of 32, while the corresponding activity from rat liver is 0.016. This comparison argues that the natural substrate for the 3-beta-hydroxysteroid dehydrogenase from liver is not in fact a 3-beta-hydroxysteroid. This argument is reasonable on other grounds as well. The substrate specificity of the enzyme from liver is quite broad; it catalyzes redox reactions on dihydrobenene diol, chloral hydrate, and quinones with an efficiency similar to its catalytic activity with steroids.[15,46]

Of course, in comparing kinetic parameters, analogous reactions proceeding in analogous directions must be examined. It makes little sense to compare specific activities for two enzymes acting in opposite directions, especially if the reactions being compared have quite different equilibrium constants. A Haldane equation relates the Michaelis constants and maximal rate constants to the external equilibrium constant for enzymatic reactions. Strongly exergonic reactions are expected to have higher V_{max} values than weakly exergonic reactions to reflect (at least some) of the increase in equilibrium constant.

For example, the alcohol dehydrogenase from *Drosophila* catalyzes the oxidation of secondary alcohols at a maximal rate five times larger than it catalyzes the oxidation of primary alcohols. This fact is occasionally interpreted as evidence that alcohol dehydrogenase from *Drosophila* acts naturally on secondary alcohols. However, the equilibrium constant for the ketone-secondary alcohol couple favors the oxidation reaction by perhaps a factor of 100 in comparison with the aldehyde-primary alcohol couple. Some of this difference in equilibrium constant is expected to appear in the forward V_{max}. Therefore, the higher V_{max} for secondary alcohols is not *a priori* a strong argument that secondary alcohols are the natural substrate of the enzyme.

Furthermore, kinetic parameters are best measured under conditions resembling those found physiologically. For example, isozyme B of carbonic anhydrase is inhibited by physiological concentrations of chloride ion. Furthermore, the abundance and activity of the C isozyme suggests that isozyme B is not essential for the metabolism of carbon dioxide, and carbonic anhydrases catalyze a wide range of ester hydrolyses. These facts suggest that the physiological role of isozyme B might not be to hydrate CO_2.[47]

Some of the strongest tests of a hypothesis regarding physiological function are those that connect the protein directly with a biological phenotype *in vivo*. For example, the activity of an enzyme can be interrupted *in vivo* by mutation, and the phenotype of the mutant organism can be studied. In microorganisms, this test is especially valuable. For example, mutants of *E. coli* lacking homoserine dehydrogenase require threonine and methionine for growth, almost certainly establishing aspartate semialdehyde as the natural substrate.[48] The hydroxysteroid dehydrogenases from strains of *Pseudomonas* are necessary to permit the organism to grow on steroids as the sole carbon source.[49]

Furthermore, evidence can be obtained for selection of the enzyme in a controlled environment, for a correlation of the level of the enzyme present in the organism with the environment the organism naturally inhabits, or for variation of enzyme levels in a rational fashion in different environments. For example, the phenylalanine dehydrogenase from *Bacillus badius* is induced in the presence of phenylalanine, implying that it plays a role in the degradation of phenylalanine rather than in the biosynthesis of phenylalanine.[50]

Many enzymes listed in standard compilations are assigned a function using only *in vitro* tests. It cannot be emphasized too strongly that firm conclusions generally cannot be based on these tests alone. Several examples drawn from the recent literature provide illustrations that make this point.

Proteins that bind nucleic acids and destabilize the double helix are well studied.[51,52] For the proteins from prokaryotic sources, mutants are available and support the hypothesis that these proteins are important in DNA repair, replication, and recombination. However, an analogous protein was isolated from rat liver and was proposed to perform an analogous function in rats.[53] Supporting the idea was the fact that antisera to the protein from rat cross-reacted with a protein in *Drosophila* salivary gland.[54] The protein was presumed to have a physiological role relevant to its ability to bind to double stranded DNA and destabilize the helix and therefore to be important for gene regulation, differentiation, and perhaps disease processes.

This assignment was called into question by the discovery [55] that the amino acid sequence of "DNA-binding protein" was nearly identical (about 93%) to the lactate dehydrogenase from pig. The DNA-binding protein was then shown to catalyze the oxidation of lactate.

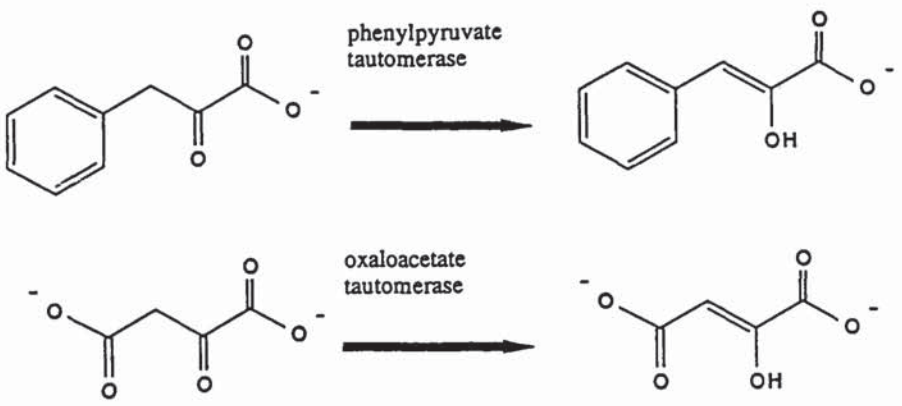


FIGURE 4. Reactions catalyzed by phenylpyruvate tautomerase and oxaloacetate tautomerase. Neither reaction seems to be important for metabolism.

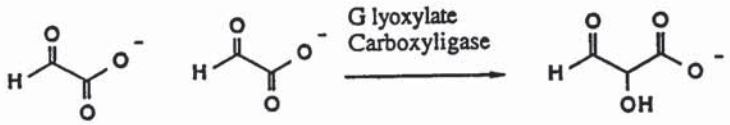
The physiological significance of the DNA binding of lactate dehydrogenase is not clear. Analogous binding is seen with glyceraldehyde-3-phosphate dehydrogenase.[56] However, the fact that a protein has a well-known biological activity (lactate dehydrogenase) suggests that the protein's ability to bind DNA may not be relevant to its biological function. However, at this point, it is impossible to distinguish between the possibility that the DNA binding displayed by these dehydrogenases is fortuitious (and not functional) and the possibility that the DNA binding has some as-yet undiscovered physiological role.

Similar cases where *in vivo* data have proven inadequate to resolve ambiguities regarding biological function can be found among catalytically active proteins. For example, an "oxaloacetate decarboxylase" from cod appears to be physiologically important as a pyruvate kinase.[57]

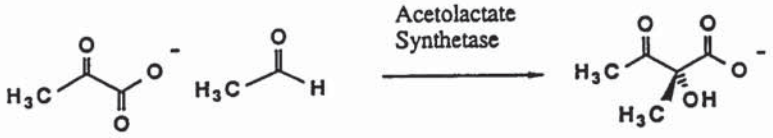
For other enzymes, the physiological role is simply ambiguous. For example, there is no obvious biological function for oxaloacetate keto-enol tautomerase (Figure 4).[58] The reaction catalyzed by this enzyme appears to be superfluous. The enzyme is believed to be distinct from common enzymes catalyzing reactions on oxaloacetate. Details of the enzyme catalytic reaction have been well studied, but the significance of the information gathered depends on functional assumptions. Similarly ambiguous is the function of phenylpyruvate tautomerase. Again, the reaction appears superfluous, but it has been suggested that the reaction is important in creating the enol form of p-hydroxyphenyl pyruvate as the starting point in thyroxine biosynthesis.[59]

There are several examples where bioorganic data create functional ambiguity where it is not expected. Phosphorylase A contains pyridoxal, with no obvious functional role.[60] Acetolactate synthetases from plants contain flavin,[61] as does glyoxylate carboligase from *E. coli*[62,63] and chorismate synthase.[64,65] The reactions catalyzed by each enzyme are shown in Figure 5. While, by several tests, these proteins function biologically to synthesize glucose-1-phosphate, acetolactate, hydroxymalonic semialdehyde, and chorismate, respectively, the presence of unusual cofactors raises the possibility that they have additional functions (a functional model) or that the unusual cofactors are vestiges of ancestral function (historical model).

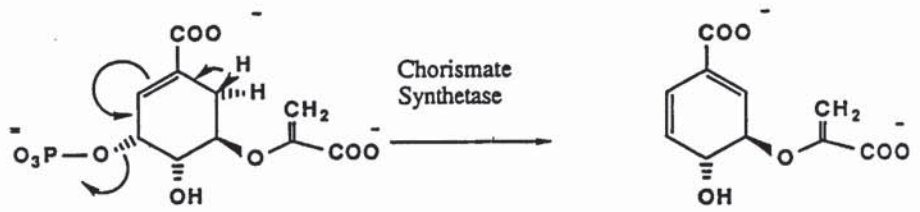
Because proteins with one fundamental activity can be mistaken for proteins with a completely different function, it is not surprising that physiological roles can be mistaken when a choice must be made from among a set of chemically similar reaction types. A classical example of this ambiguity is the alcohol dehydrogenase from horse liver. A wide range of substrates have been suggested. For example, Dutler and colleagues have found that the enzyme can oxidatively decarboxylate porphyrin derivatives and they suggested a physiological impact (if not an evolutionarily selected function) of this behavior.[66] Other examples from alcohol dehydrogenases are known. Human liver "high K_M" aldehyde dehydrogenase may be better



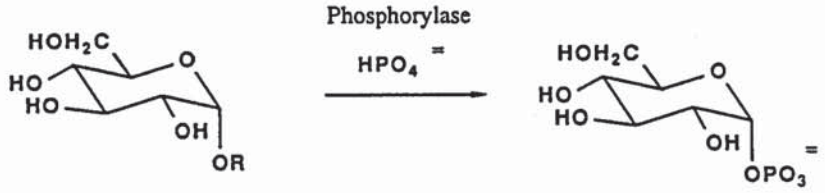
Expected Cofactor: Thiamine Pyrophosphate
"Superfluous" Cofactor: Flavin Adenine Dinucleotide



Expected Cofactor: Thiamine Pyrophosphate
"Superfluous" Cofactor: Flavin Adenine Dinucleotide



"Superfluous" Cofactor: Reduced Flavin Mononucleotide



"Superfluous" Cofactor: Pyridoxal Phosphate

FIGURE 5. Reactions catalyzed by four enzymes that appear to bind "superfluous" cofactors, i.e., cofactors not obviously important for the chemical reaction.

classified as a glutamate-gamma-semialdehyde dehydrogenase.[67] Lactaldehyde reductase from pig kidney may be indistinguishable from an enzyme that reduces glucuronate.[41]

These examples illustrate the difficulties of assigning biological function to an enzyme using only *in vitro* tests. However, before a functional theory can be constructed, an assignment must be made. Fortunately, biological tests can help identify function even if *in vitro* tests 1 and 2 fail to provide an unambiguous assignment. While there are many examples where such tests have been used in complicated cases, two deserve special mention.

One case, discussed in detail by White and co-workers,[68] concerns the assignment of physiological function to isozymes of glycerol-3-phosphate dehydrogenase from chicken liver

and chicken muscle. Kinetic data suggest that the muscle enzyme operates primarily to oxidize NADH formed during anaerobic glycolysis and that the enzyme from liver has a primary role in the biosynthesis of triglycerides.

Two facts formed the basis for a challenge to the latter assignment. First, the enzyme from liver uses NAD^+ rather than $NADP^+$. Since NAD^+ is rarely involved in a reductive synthetic pathway, it was considered unlikely that the liver enzyme acted in a reductive synthetic pathway. Second, the effect of starvation on the level of the enzyme was measured and found to change in a fashion characteristic of an enzyme involved in glycerol catabolism rather than triglyceride biosynthesis.

Biological tests have been extensively applied to assess the physiological function of the alcohol dehydrogenase from *Drosophila* (DADH). With DADH, *in vitro* data do not permit a conclusion to be drawn about the natural substrate. The enzyme acts on ethanol and acetaldehyde. However, its specific activity when acting on ethanol is lower (250 vs. 10) than that of the alcohol dehydrogenase from yeast, an enzyme presumed to act on ethanol naturally. Nevertheless, the specific activity of DADH acting on ethanol is still respectable when compared with that for other dehydrogenases from multicellular organisms.

Furthermore, DADH has a substrate specificity that is somewhat broader than the yeast alcohol dehydrogenase. It is catalytically active against a range of long-chain primary and secondary alcohols. However, the substrate specificity of DADH is not as broad as that for liver alcohol dehydrogenase, an enzyme presumed not to act on ethanol as a natural substrate. Furthermore, much of the *in vitro* data that are adduced to show that the enzyme has a broad substrate specificity were collected at pH values that are unphysiologically high.[69,70]

These *in vitro* data allow no firm conclusions regarding the physiological role for DADH. However, the "physiological role" concerns those properties of DADH that are important in the survival of the fly in its natural environment. Fortunately, a large amount of biological evidence is available to address this issue directly. These data suggest that the primary, and probably the only, selected role of DADH is the oxidation of ethanol, both as a source of metabolic energy and as a mechanism for detoxifying environmental ethanol.

Different species of *Drosophila* have different levels of DADH which correlate with the level of ethanol in the environment where they feed (which can be as high as 15%).[71] The different species characterized by different DADH activities lay eggs in different media. The level of ethanol in the media in which the eggs are laid corresponds to the tolerance of the individual strain of flies to ethanol, as measured experimentally,[72] implicating ethanol oxidation as at least one physiological function of DADH.

Further studies concern strains of flies genetically adapted in the laboratory. When placed under stress in the form of high environmental ethanol, flies are selected that constitutively produce increased levels of DADH,[73] a demonstration of a mechanism for selecting a DADH that oxidizes ethanol.

Data from mutant flies in which the DADH gene is deleted also suggest that the selected function of DADH is to oxidize ethanol. In the presence of toxic levels of ethanol, flies lacking DADH survive less well than flies having DADH. Indeed, mutants deficient in ADH are killed by rather low concentrations of ethanol.[74] However, when fed ethanol as food, wild-type flies survive substantially better than mutant flies deficient in ADH.

In contrast, data from mutant flies all but rule out the possibility that the physiological role of DADH is to detoxify or metabolize long-chain primary and secondary alcohols. In the presence of toxic levels of secondary alcohols, wild-type flies and mutant flies lacking DADH survive equally well.[75] Similarly, tolerance to long-chain primary alcohols is not substantially different in flies with and without DADH.[76] Likewise, there is little or no difference between survival ability of wild and ADH-null mutant flies fed on long-chain primary alcohols as food.[77]

One final study is worth mentioning. Kreitman sequenced 11 cloned genes for DADH from a population of *Drosophila* and found 43 polymorphisms (structural variations) in the gene

coding for DADH.[21] Remarkably, *only one* of the variants actually had an impact on the amino acid sequence of the protein expressed from the gene. The remaining 42 variations were silent. The implication of this experiment is that DADH is under strong selective pressure to maintain its amino acid sequence.

Together, these studies support two conclusions. First, experiments demonstrate a mechanism for selecting a DADH suited for the oxidization of ethanol. Thus, DADH almost certainly has an important role in helping flies survive in their natural environment by its ability to oxidize ethanol, the most abundant alcohol. Second, experiments failed to find a mechanism for selecting for a DADH suited to oxidize long-chain primary and secondary alcohols. Thus, oxidation of long-chain primary and secondary alcohols is most likely not a physiological role of DADH.

As convincing as they are, these data still do not constitute "proof". However, this illustration shows the value of biological data in addressing a question where *in vitro* data do not permit conclusions to be drawn about physiological function.

On occasion, bioorganic chemists remain resistant to arguments based on biological data.[78,79] For example, Schneider-Bernloehr and co-workers recently wrote that *Drosophila* ADH "has not evolved for optimal catalysis of ethanol oxidation", even though they apparently were aware of the biological data presented above.[79] They argued that DADH had a metabolic role as a secondary alcohol dehydrogenase, a conclusion based on *in vitro* kinetic data that showed that the kcat/K_M ratio for secondary alcohols "is up to 50 times higher for *Drosophila* ADH and about 400 times lower for yeast ADH compared to kcat/K_M for ethanol" (Tests 1 and 2).

They made two additional arguments. First, they presented an argument premised on the assumption that:

> Enzymes can be divided into two groups. One shows high substrate specificity and low heterozygosity, which may be compatible with high evolutionary pressure. The other group is 'substrate-nonspecific', and is characterized by a high heterozygosity. Alcohol dehydrogenase from different species of *Drosophila* have been classified as belonging to this latter group.
>
> In conclusion, DADH shows the typical behavior of a multiple-substrate-specific enzyme: high heterozygosity, which may be compatible with low functional constraints, broad substrate specificity, and low turnover numbers.

Second, they recognized that the biological arguments made above "indicate that this enzyme is responsible for detoxification of ethanol." However, they dismissed this role as a selectable one, writing that "it is clear that it [DADH] is not essential for energy production since flies without this ADH can also survive."

In the spirit of a critical review, we may use the discussion presented here to analyze the plausibility of these arguments. The first argument is not factually accurate. As the work of Kreitman shows *(vide supra)*,[21] DADH, in fact, has surprisingly low "heterozygosity" (in this case meaning polymorphism). The work of Kreitman is almost certainly incompatible with any suggestion that the enzyme has "low functional constraints".

Furthermore, the assumption that enzymes can be divided into two groups where low substrate specificity is correlated with high heterozygosity (and vice versa) is itself rather controversial.[29,30] It cannot be assumed as a premise. How selection pressures determine substrate specificity and how heterozygosity is connected with selection pressure are two unresolved questions in molecular evolution. DADH itself contradicts the assumption that Schneider-Bernloehr and co-workers state as obviously true; it is an enzyme apparently under extreme functional constraints, but which does not have high substrate specificity.

The second argument, that the natural function of DADH cannot be the oxidation of ethanol but rather is the oxidation of secondary alcohols because DADH is not essential for energy production, is both logically and biologically flawed. The biological experiments searched for, and failed to find, a selectable phenotype associated with the ability of DADH to oxidize secondary alcohols. Thus, Schneider-Bernloehr et al. are arguing that DADH has a particular

metabolic function (oxidation of secondary alcohols) when experimental evidence strongly suggests that it makes no difference to the survival of flies challenged with secondary alcohols whether or not they have DADH.

The authors are correct in their statement that mutants lacking DADH do indeed survive in the absence of ethanol; clearly, DADH is not essential for energy production. However, these reviewers know of no statement in the literature that DADH is essential for energy production under these conditions. Rather, the argument is that, in the presence of ethanol, DADH catalyzes one step in the utilization of ethanol for energy and in the detoxification of ethanol. In order for Schneider-Bernloehr's argument to be relevant to the issue of natural function, one must assume that only enzymes required for "energy production" are under selective pressure (almost certainly false), or that there is no ethanol in the natural environment of *Drosophila* (clearly false).

It is most remarkable, given the recognized fallibility of *in vitro* methods for determining natural function, that these authors (and others)[78] rely on them to the exclusion of excellent relevant biological data. This reliance should be regarded as a weakness. In the hope that it will encourage others to abandon this approach, we would like to abandon it here ourselves in the one case where one of us attempted to apply it and made an incorrect judgement as a result *(vide infra)*.[40]

Unfortunately, enzymes from multicellular organisms are only infrequently studied as thoroughly as DADH. However, in microorganisms, mutants are frequently available to determine whether the absence of an enzyme of interest has a detectable phenotype. Therefore, microorganisms often are the best source of enzymes useful for testing functional theories.

V. HISTORICAL MODELS

Historical models make presumptions about relatedness, origin, and conservation in biological macromolecules. Constructing and testing these models requires methods for determining homology in macromolecules. Three questions might be of concern. The first simply asks whether two enzymes are homologous or not. The second concerns the detailed connectivity of the tree that describes the pedigree of several homologous macromolecules. The third concerns the correlation of points on the tree where specific enzymes diverged with absolute geological time.

Bioorganic chemists seeking to construct historical models are most interested in asking whether or not two enzymes are homologous. In some cases, this question is easy to answer. If the amino acid sequences of two proteins can be aligned with few gaps, and if the amino acids are the same at greater than 30% of the positions in the alignment, the two proteins are almost certainly homologous. Homology between parts of the sequence also can be established by such alignments. Examples of such proteins are the alcohol dehydrogenases from yeast and horse liver, the proteases trypsin, chymotrypsin, and elastase,[80] and the isozymes of lactate dehydrogenase.[81]

However, many historical models rely on comparisons of macromolecules where the level of sequence identity is inadequate to establish homology with statistical significance, or where the sequence similarities are borderline. Doolittle and co-workers have presented a penetrating review on the statistics associated with the analysis of sequence similarities[82,83] and have discussed several cases where alignments claimed to suggest homology in fact contained sequence similarities no greater than those statistically expected for the comparison of two proteins with random sequences. Analogous statistical arguments form the basis for the analysis of aligned DNA sequences.[84]

Of course, failure of an alignment to reveal a statistically significant level of sequence similarity does not necessarily mean that two proteins are not homologous. However, the sequence data are not adequate to support an assumption of homology needed as the basis for a historical model explaining the behavior of macromolecules.

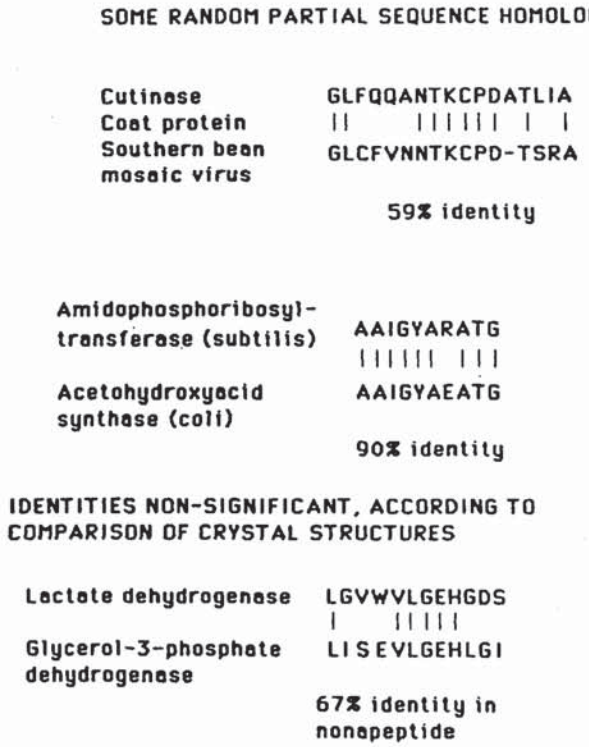


FIGURE 6. Computer matching of sequences readily identifies short stretches of similar sequences in proteins that are not obviously homologous. The significance of these similarities is not clear. However, in one case (bottom pair), an alignment based on the sequence similarities around an "essential" histidine does not correspond to an alignment based on tertiary structural similarities.

Even when sequence similarities are statistically significant, it is still difficult to construct trees with a defined connectivity, to relate these trees with trees constructed independently from fossil and physiological data, and to assign an absolute time scale to the branch points of the tree. Such construction would be aided were the rate of change in protein structure linear as a "clock-like" function of geological time.[85] Regrettably, it appears not to be, even within a given class of enzymes.

For example, the divergence of superoxide dismutase has apparently not been clock-like, judging by the sequences of the enzymes from *Drosophila*, man, horse, cow, and yeast.[86] Among mammals, the protein's structure has diverged very rapidly (30.9 substitutions per 100 amino acids per 100 million years). This divergence rate is one of the fastest observed for any protein. When the enzymes from *Drosophila* and mammals are compared, the divergence rate is much slower (10.6). However, when the enzymes from yeast and animals are compared, the divergence rate is slower still (5.8) and similar to the divergence rate in cytochromes (2.2) and other slowly diverging proteins.

Furthermore, the sequences of superoxide dismutases show that the enzymes from *Drosophila* and cow both have a deletion of two amino acids at a position within the protein, a deletion that is not present in the enzymes from yeast and horse. This seems to suggest that the enzymes from *Drosophila* and cow are more closely related than are the enzymes from cow and horse, a suggestion that contradicts much biological information. While the data can be

rationalized using an *ad hoc* assumption (the deletion occurred independently at the same place in two separate lineages), the contrast between the most simple interpretation of the data and the close relationship between cow and horse (as assessed by paleontology) is disquieting. As many of the microscopic behaviors measured by bioorganic chemists, including kinetics and dynamic behavior on the nanosecond time scale, might presumably be influenced by this deletion, the connectivity of this tree is important

Behavior that is not clock-like is also observed in the divergence of DNA sequences. For example, comparisons between nuclear and mitochondrial DNA sequences reveal extreme variation in the rate at which two sequences diverge.[87] Thus, it is virtually impossible from sequence data alone to draw conclusions about the relative times elapsed since divergence of sequence.

The possibility that genetic material can be transferred between unrelated species further complicates the construction of trees. For example, Martin and Fridovich suggested that the superoxide dismutase gene of ponyfish, *Leiognathus slendens*, may have been transferred to its bacterial luminescent symbiont, *Photobacter leiognathi*, in recent evolutionary times.[88] It is difficult to evaluate at present the extent to which interspecies transfer of genetic information has occurred. To the extent that it has occurred, the evolutionary distance between two organisms will not be a good judge of the evolutionary distance between two homologous enzymes from these organisms.

Fortunately, for most historical models, the only important issue is whether two proteins are homologous. The connectivity of a tree is often only important if the model builder wishes to suggest which pairs of homologous enzymes are more likely to share behavioral traits.

When comparisons of sequence data reveal only borderline homology, a battery of statistical tests are available. For example, improvements in statistics can be obtained by comparing more than two related sequences.[89,90]

If a comparison of sequence data fails to detect statistically significant homology, other criteria are needed to assess homology. For example, it appears that the tertiary structures of proteins drift more slowly than their primary structures.[20,91-93] Thus, if two sequences have drifted so that no statistically valid sequence similarities can be detected, crystallographic evidence may still indicate that two proteins are related.

Chothia and Lesk have very recently suggested an analogous correlation between divergence in tertiary structure and divergence in primary structure.[94] The root mean square (rms) deviation in the positions of the backbone atoms in two proteins with identical sequences is approximately 0.4 Å. This is a measure of the variation expected to arise from different crystal packing forces and the error in the crystal structure. When 25% of the amino acid residues have diverged, the rms deviation increases to 0.7 Å. With 50% sequence divergence, the rms deviation is approximately 1 Å. With 75% sequence divergence, the rms deviation is 1.6 Å.[94]

Many investigators have attempted to exploit the slow divergence of tertiary structure as a tool for establishing homology. For example, Rossmann noted that the tertiary structure of the dinucleotide-binding domain in several dehydrogenases is quite similar and suggested that dehydrogenases sharing this domain are homologous.[92] In many cases, the homology is difficult or impossible to detect by sequence alignments. By this argument, lactate dehydrogenase, malate dehydrogenase, alcohol dehydrogenase, and glyceraldehyde-3-phosphate dehydrogenase are all related. Similarly, flavodoxin and adenylate kinase may be related by this criterion.[92,93]

As a case in point, the dinucleotide-binding domains of lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase show no detectable sequence homology.[95] However, their tertiary structures are "topologically equivalent" at 57% of their residues. The probability of this degree of similarity occurring at random is almost certainly low, but difficult to assess quantitatively.

Matthews and co-workers have used similar reasoning to argue that the lysozymes from

goose (GEWL), chicken (HEWL), and bacteriophage (T4L) all have a common ancestor.[96] By sequence comparisons, these proteins are not obviously related. However, Matthews and co-workers argued that the folding patterns of the protein are sufficiently similar to conclude that these three proteins are related. This argument is noteworthy, as several active site residues, which often are assigned catalytic roles, are not conserved among the three proteins. For example, Glu 73 of GEWL corresponds with Glu 35 of HEWL and Glu 11 of T4L. However, there is no obvious counterpart to Asp 52 in GEWL and Asp 20 T4L. If these proteins are indeed homologous, mechanism apparently can diverge faster than tertiary structure, and can do so while substrate specificity remains the same.

There is, of course, considerable subjectivity involved in aligning and comparing the folding patterns of enzymes having no clear sequence homologies. Furthermore, arguments that very similar tertiary structures can arise independently by convergent evolution are difficult to exclude.[1,97] There is some attractiveness to the idea that only a few folded forms of polypeptides are favored. If so, similarity in crystal structure cannot be taken as a strong argument for homology.

Richardson and co-workers[98] applied similar arguments in their analysis of the structures of immunoglobulin and a subunit from a Cu/Zn superoxide dismutase. These two macromolecules have no apparent functional similarity and display no sequence similarity. Based on these facts, it is simplest to assume that the two proteins are not homologous. However, the folding pattern of the two proteins is analogous, suggesting that either these enzymes are distantly related or the same folded form arose by convergent evolution favoring a folded structure that had intrinsic stability. Richardson et al. calculated that the "probability" that the two proteins would have the same folded form by chance is 1/2853.[98] Of course, this calculation has large elements of subjectivity.

In such an extreme case, it is difficult to draw any conclusion regarding homology. The probablility of these two folded forms being the same by chance may be small. On the other hand, since many different comparisons are possible given the substantial number of protein structures available, the probability that at least some nonhomologous proteins will have similar folded forms may, in fact, not be low. Furthermore, it is not clear how one should distinguish between the two possibilities. With two macromolecules unrelated by sequence and nonanalogous by function, yet having the same folding pattern, an independent demonstration of homology, or lack thereof, is needed to decide whether the similar folding pattern arose by convergent evolution or was conserved following an ancient divergence of the two proteins. These reviewers are unaware of the existence of any method for making such an independent demonstration.

Crystallographic data can be used to detect sequence homologies below the statistical level. For example, once the sequences of the lysozymes discussed earlier were aligned using the presumably homologous folding patterns, Matthews and co-workers could identify residues that were conserved, even though the level of conservation alone would not have been enough to align the sequences.[96]

Other methods exist for confirming the significance of sequence similarities that are statistically borderline. An example is provided by comparisons of two classes of dehydrogenases.[99] The first class includes alcohol dehydrogenases from a variety of yeast strains, and alcohol and sorbitol dehydrogenases from a variety of mammals. The second includes alcohol dehydrogenase from *Drosophila*, and ribitol dehydrogenase from *Klebsiella*. The first class contains enzymes that are clearly interrelated, as does the second class. However, the sequence homology between the two classes in the dinucleotide-binding domain is borderline; it approaches 30% in only one region. Are the dinucleotide-binding domains of the four enzymes descendents of a common ancestor?

By sequence analysis alone, Jornvall concluded that they are.[99] However, the sequence similarity was not sufficient to be generally convincing. However, comparing sequence

divergence within each class, one can identify residues that have drifted within each class and residues that have been conserved within each class. The conservation of specific residues can often be rationalized using the structural data from crystallography. For example, four of the conserved residues are glycines involved in the formation of the dinucleotide-binding fold. These are conserved within both the first group and the second group of dehydrogenases, and they are also conserved between the two groups. This strengthens the argument that both classes of dehydrogenases are related, but of course does not rule out the possibility that dinucleotides are uniquely bound by a protein structure with four glycines in the conserved positions.

A. Conservation Principles

A plausible pedigree is necessary, but not sufficient, for historical models to be predictive. If two enzymes share a nonfunctional behavior, and if this similarity is explained by assuming that both enzymes descended from a common ancestor where this trait was primitive, the model must also explain why this individual trait has not drifted in the time since the two proteins diverged. Thus, historical models must embody a "conservation principle" which comments on the relationship between structure, behavior, and natural selection relevant to the conservation of the particular behavior being examined.

Without hypotheses about the conservation of nonfunctional traits, historical models are not predictive, as the conservation principle is the basis for the expectation that a primitive trait will be found in modern proteins. Yet many historical models in the bioorganic literature are not accompanied by statements regarding conservation of behavior, and in many cases the conservation principles implied by these historical models are inconsistent with what is known about the divergence of behavior in proteins.

Behaviors should drift if (1) they are not functionally constrained and (2) if there has been enough time since the two macromolecules diverged. Therefore, a historical model must explain the conservation of a nonfunctional trait either by assuming that there is a function linked with the trait that is constrained from drifting (a "functional constraint"), or by assuming that the two proteins only recently diverged.

To illustrate these two assumptions, let us consider two decarboxylases that both catalyze decarboxylation with retention of configuration, and a historical model that argues that decarboxylation with retention is a primitive trait conferring no selective advantage over decarboxylation with inversion. The historical model is incomplete unless it explains why this nonfunctional stereospecificity has not drifted since the two enzymes diverged.

Two arguments are possible. First, we can argue that there has not been enough time since the two proteins diverged for stereospecificity to drift. Such an argument is strong if the sequences of the two enzymes are quite similar. For example, if the sequences are 50% identical, and if we assume that for stereospecificity to diverge the sequence of two proteins must be less than 30% identical, conservation is expected. The enzymes have not had the time to accumulate enough mutations for stereospecificity to have been reversed.

Alternatively, we might argue that, although enzymes producing retention and inversion equally contribute to the survival of the host organism, stereospecificity is indirectly linked to catalytic efficiency. Enzymes having structures intermediate between those that produce retention and those that produce inversion are catalytically inefficient and this inefficiency damages the survival of the host organism. Thus, drift from an enzyme producing retention to one producing inversion is indirectly *functionally constrained.*

The first argument requires assumptions regarding relationships between structure and behavior. To decide whether divergence in stereospecificity is expected, given the level of structural divergence, one must make assumptions about the level of structural divergence that is needed to create a protein with divergent stereospecificity. In the second argument, the historical model incorporates a conservation principle based on a functional constraint. Such a historical model is itself partly a functional model, and issues of functional role relevant to functional models are relevant here as well *(vide infra)*

The second argument is usually considered so obvious as to not require statement. However, as we shall see below, an assumption that certain traits are linked with others is highly controversial, as is the assumption that retentive and invertive enzymes are separated by proteins that are catalytically inactive. Ultimately, assumptions about what traits are conserved as two sequences of amino acids diverge must be based on an understanding of how enzymatic behavior is determined by amino acid sequence. This understanding is lacking in modern protein chemistry, although it is currently the object of intensive study.

These issues are discussed in detail in the next section. For now, we should only note that one of the most common flaws of historical models is that they lack a conservation principle. Historical models with this flaw do not make interesting predictions. They do not define how a nonfunctional trait will be conserved as structure diverges and therefore do not predict the behavior in newly studied divergent enzymes.

B. The Dialectic

This discussion covering the construction of functional and historical models should not be regarded as philosophically obscure.

The bioorganic chemist must draw *some* conclusion about whether the data he examines do or do not reflect a selected behavior. This is the only way he can understand its significance. This creates a dilemma. On one hand, constructing a functional model requires a detailed consideration of biological function, a function that is not always apparent for the system being studied. Furthermore, it requires recognizing the strong trend in modern population biology that treats many subtle details of protein structure and behavior as "neutral", the nonfunctional products of historical accidents.

On the other hand, historical explanations are not predictive (and therefore neither testable or falsifiable) without conservation principles that connect an accident in the historical past with the behaviors of modern proteins. Such conservation principles must ultimately be based on a deeper understanding of relationships between structure and function in proteins. This understanding is presently unavailable.

Furthermore, should the chemist wish to assume a conservation principle based on a functional constraint of a linked enzymatic behavior, he must construct a functional argument as well and must solve the problems associated with constructing functional models.

Finally, to test a historical model, the pedigree of the systems under study must be assessed. If the bioorganic chemist is unfortunate enough to be studying a protein for which sequence data or crystal data are not available, this means much work on many systems. The best resolution of these problems requires model building that is logically formal. Several of these models, covering the kinetic power, stereospecificity, and substrate specificity have been covered elsewhere.[100-102] In the following, we review the state of present knowledge concerning the interaction between natural selection and behavior and attempt to divide in a general way those traits that are selected from those that are drifting.

VI. THE SURFACE

To construct historical and functional models, assumptions must be made about the relationship between the behavior of a protein, which can be measured, and the structure of the protein, which contains the best information regarding the history of the protein. In this section, we discuss the data which form the basis of our current knowledge of relationships between natural selection, macromolecular behavior, and macromolecular structure.

Macromolecular behavior is a function of macromolecular structure. In a protein n amino acid residues in length, a surface in n + 1 dimensions describes this behavior, where the n independent variables are the amino acid substitutions at each of n positions.[103] For a protein with n amino acids, the surface consists of 20^n points corresponding to this number of proteins with

3 dimensional surface relating behavior and structure in a dipeptide (n=2)

FIGURE 7. Surfaces describe the relationship between a property of an enzyme (k_{cat}, K_M, stability, survival value) and the amino acid sequence of an enzyme. Here, for clarity, we show the surface for a dipeptide (n = 2) where there are only ten amino acid substitutions at each of the two positions.

Two types of surfaces might be considered separately. The first type are "behavioral surfaces", those that relate structure to behavior in a protein. The second are "survival surfaces", those that relate structure and "survival value" in a protein.

Survival value is the degree to which a macromolecule assists a host organism as it struggles to survive and reproduce. In principle, it is defined by an experiment in which two organisms, otherwise identical but differing in the structure of the macromolecule of interest, compete for survival in a defined environment. The relative amounts of the two "isomacromolecules" in the population in equilibrium defines their relative "survival value".[104] In practice, survival value is only occasionally measured quantitatively *(vide infra)*.

In contrast, behavioral surfaces describe properties of a macromolecule that can be directly measured in the laboratory. Since molecular biological techniques now enable bioorganic chemists to synthesize variants of proteins with specific amino acid replacements, these surfaces can be directly and systematically explored.[105] The only serious limitation is that only a small portion of a surface can be examined during the length of a normal research project.

The evolution of protein behavior corresponds to motion across these surfaces.[106] Alterations of structure that have no impact on survival value (i.e., that reflect motion on a flat region of a survival surface) correspond to neutral "drift". Motion of a macromolecule "downhill" on a survival surface is impelled by natural selection; motion "uphill" is constrained by natural selection. Like a ball rolling on a suface in a gravitational field, the structure of the macromolecule will move along the survival surface to an optimum, drawn here as a low point on the surface.

Of several possible models for describing the rate of evolution across surfaces of these types, one simple model is analogous to rate processes in chemistry. Survival value can be defined so that the equilibrium ratio of two proteins in a population is an exponential function of the "survival free energy difference" between the two proteins. The rate of appearance of protein B from protein A is simply the product of the number of genes in a population coding for A times an intrinsic mutation rate (Figure 9). The intrinsic mutation rate is presumed to be constant for different genes, although it may vary from organism to organism.

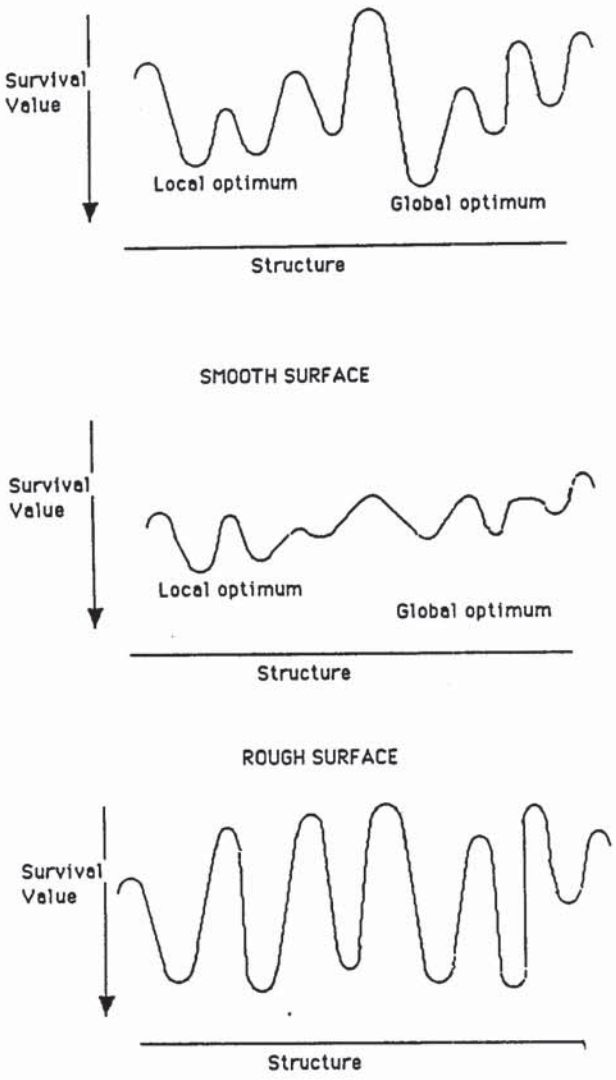
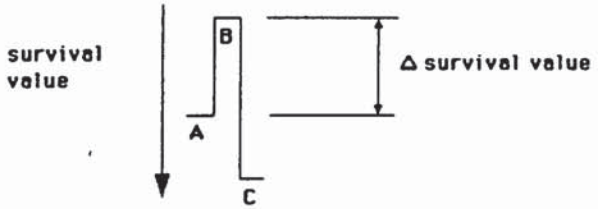


FIGURE 8. For the sake of clarity, complex surfaces are represented in this article as continuous, and in only two dimensions. "Smooth" and "rough" are qualitative descriptions of the surface without more precise definition. The surface is drawn so that the enzyme with the best survival value is the lowest point on the surface. Thus, under natural selection, enzymes will gravitate to this lowest point.

Thus, evolving from one local minimum to another when the environment is static is a function of the "concentration" of species at the "transition state", defined as the structural intermediate between the two minima with the least favorable survival value. As in transition state theory describing chemical kinetics, this model assumes that proteins up to and including the highest barrier are present in equilibrium amounts in a population. As with chemical kinetics, this assumption is not necessarily universally valid.

The number of total points on the surface is large (20^n). For example, the surface describing



$$\frac{\text{concentration of B}}{\text{concentration of A}} = \exp(-\Delta \text{ survival value})$$

$$\text{rate of formation of C} = (\text{microscopic mutation rate}) \times (\text{concentration of B})$$

$$= (\text{microscopic mutation rate})\ (\text{concentration of A})\ \exp(-\Delta \text{ survival value})$$

FIGURE 9. By analogy with transition state theory in organic chemistry, we might define relative survival value by the ratio of isoenzymes in a population at equilibrium. A naive expression for the rate of appearance of a variant can be written from this definition.

a protein with 58 amino acids (e.g., bovine pancreatic trypsin inhibitor). has 3×10^{75} points. However, because of the high dimensionality of the surface, no two points on the surface are far apart. Any point on this surface can be reached from the most distant point in only n – 1 (here, 57) steps. Furthermore, the number of different direct pathways between two distant points (representing two proteins with no sequence identities) is equal to n!, a very large number when n is large.

The result of such high dimensionality is that there are many paths connecting points on the surface. For example, for a protein the size of bovine pancreatic trypsin inhibitor, the number of direct paths between the two most distant points on the surface is 2×10^{78}. Here, the number of direct paths between two distant points is 1000 times greater than the total number of points on the surface. With many paths, the probability declines that the evolution of one structure to another must proceed via a path with a protein having unacceptable properties as an intermediate. Thus, surfaces with many dimensions are less likely to have isolated optima than sufaces with few dimensions. However, even with many dimensions, it is still possible for optima to be isolated.

The two types of surfaces are functions of the environment. When the environment is perturbed, surfaces describing selective value can undergo upheaval, substantially changing the relative survival value of different structures. If the upheaval changes what was formally a local optimum into a nonoptimum, then the macromolecule with the previously optimal structure will evolve downhill in search of a new optimum. Again, if the population was at "equilibrium" before the upheaval, the initial rate at which the protein evolves will depend on the relative amounts of various structures in the population, as defined by the initial surface, moving along the surface that results after the upheaval.

In contrast, most bioorganic behaviors of macromolecules do not change drastically with changes in external conditions. Thus, behavioral surfaces are likely to undergo more subtle variation after a modest environmental perturbation.

To explain macromolecular behavior in terms of function, one would like to convolute the two surfaces and relate survival value directly to macromolecular behavior. This convolution is complicated. The survival value of a macromolecule depends, at least in principle, on many of its behaviors. Since different behaviors do not necessarily contribute equally to survival value, these behaviors must be weighted. Furthermore, different properties in the same protein

may not be independent of each other; achieving an optimal level of one behavior may require sacrificing it in another.

One can hope that the relationship between behavior and survival value is a simple one. For example, it may be that the higher the k_{cat}/K_M of the enzyme, the higher its survival value. While such relationships often have intuitive appeal,[4] the appeal is often only superficial.

There are behaviors of proteins where "more" is almost certainly not "better." For example, a protein with enormous thermal stability probably does *not* have greater survival value than a protein with intermediate stability. High thermal stability, if it implies that the protein cannot be degraded *in vivo*, is almost certainly disadvantageous in a protein that must be turned over in response to changes in metabolism. In contrast, great instability may be undesirable because the protein has to be resynthesized too often. Thus, a surface relating survival value and stability in a protein might be expected to have an optimum, where the stability of a protein is intermediate between "very stable" and "very unstable".

Further difficulties arise when different behaviors of proteins are not independent of each other. High turnover numbers for an enzymatic reaction, by presuming a precisely optimized fit between a substrate and catalytic groups on the protein, may not be compatible with broad substrate specificity. High turnover numbers may require high dynamic motion in proteins, implying that they are likely to be unstable with respect to thermal denaturation. High turnover numbers may require that the protein have a large size and hence a higher biosynthetic cost.

If such relationships between behaviors exist, a natural protein may be evolutionarily optimal overall, and yet be sub-optimal in each individual behavior. Thus, even if natural selection can optimize the behavior of proteins, the degree to which any individual behavior is optimized will depend on the relative values, costs, and tradeoffs between each behavior. Of course, these "tradeoffs" represent costs and values that are not easy to specify quantitatively.

Problems of this sort are frequently addressed in other branches of science, most notably in economic theory and engineering. These fields offer intellectual tools for discussing optimizations that involve qualitative evaluations of relative costs, constraints, and tradeoffs. However, economic theory and engineering are areas unfamiliar to bioorganic chemists. If they consider the problem at all, most bioorganic chemists view molecular adaptation in terms of an optimization of a single unconstrained variable.

The existence of these different surfaces presents a dilemma. Behavioral surfaces are appealing because they concern behaviors that can be measured directly and quantitatively. Discussion is uncomplicated by considerations about other properties of the macromolecule or the impact of environment. However, behavioral surfaces are uninformative about the biological relevance of the property being discussed.

Survival surfaces are appealing because they are directly biologically relevant. However, they are uninformative about behavior and their discussion always requires an element of uncertainty since the dependent parameter (survival value) is difficult to assess experimentally.

Convoluting the surfaces in an attempt to construct a function that directly relates behavior and selective value is appealing because such a relationship directly concerns functional interpretations of bioorganic data. However, the convolution requires understanding the tradeoffs between different behaviors, including their relative costs and relative contributions to survival under different environments. While a qualitative understanding of such tradeoffs is not inaccessible to experimental analysis, bioorganic chemists are so uncomfortable with the complexities of such an analysis that they often doubt *a priori* that such a complex analysis might be informative.

As complicated as this dilemma is, it cannot simply be ignored. These surfaces are important to interpreting the data collected from studies on biological macromolecules. Historical models for behaviors are likely to be more important if the survival surfaces in general have many structures with essentially equal survival value, where these structures are isolated on the surface (a "rough" surface). Rough surfaces lend themselves to conservation principles since surface

regions having equal functional value are isolated by regions that have poor survival value, implying that proteins displaying one behavior cannot drift to those displaying another. The rougher the surface, the more likely the behavior of a modern protein reflects the behavior of its ancestors.

Functional models are likely to be more important if structures with similar survival values are not isolated on the surface, permitting the protein's structure to evolve toward a global optimum determined by underlying chemical principles. Functional models are best suited to surfaces that are smooth in some dimensions and rough in other dimensions. On such surfaces, variability in behavior is possible by making only a small number of changes in structure. Thus, the protein can adapt to changing environments. However, smooth dimensions exist where a protein can reach large regions of the surface without having to evolve through proteins with unacceptable behaviors.

Finally, if these surfaces are generally flat, models explaining the behavior of proteins in terms of drift are likely to be most generally applicable. If so, the behaviors studied by bioorganic chemists will be truly random, explainable neither functionally nor in terms of ancestral events, but rather reflecting recent accident. Behaviors are likely to be different in closely homologous proteins and likely to be still different in such proteins after continued evolutionary time.

Fortunately, recent literature is adequately developed to suggest some general features of these surfaces. We present the conclusions that might be inferred from the literature reviewed below.

For behavioral surfaces, available data suggest that, in general, individual behaviors of proteins are very sensitive to changes in amino acid substitution at a few positions and largely insensitive to substitution at most other positions. This means that, in general, behavioral surfaces are "rough" in only a few dimensions at any one point, and "smooth" in most other dimensions, where the surface is scaled based on bioorganic intuitions and the sensitivity of experimental method. This is a description of the optimal surface for adaptation since it allows the most opportunity for behaviors to vary independently of each other and for evolution from one local optimum to another without the need to cross insurmountable barriers.

However, for survival surfaces relating structure and selective value, evidence suggests that natural selection can discriminate between two proteins having very few amino acid alterations. Furthermore, evidence suggests that survival surfaces are rather variable given small changes in environmental conditions. However, the scale that relates behavioral differences to selectability appears to be different in different organisms.

Finally, available data suggest a general distinction between behavioral details that are selected and those that are not. The line between selected and nonselected traits appears to cut through the center of the behaviors that currently interest bioorganic chemists. Where the line is drawn appears to be different in different organisms.

Thus, behavioral surfaces appear to allow proteins to evolve to globally optimal states, while survival surfaces appear to generally constrain the properties of a protein to a narrow region. The first statement favors functional models; the second is less clear. Three possibilities make the second statement also consistent with broad applicability of functional models.

First, the high dimensionality of the surface means that any two points are connected by an astronomically large number of paths. It is possible that, between any two points on a survival surface, at least some of these paths do not travel over parts of the surface that are extremely disadvantageous under selective pressure.

Second, it is possible that, in any given environment, locally optimal points are indeed isolated in all dimensions. However, one expects the surface to undergo upheaval with small environmental changes. These upheavals can, at least in principle, make the entire surface accessible since what formally was a local optimum is no longer a local optimum.

Third, it may be that regions on the survival surface are rough in many dimensions, while other regions have the mixture of rough and smooth dimensions that is optimal for adaptation.

If this is the case, we expect that organisms with proteins in the second region would have been more adaptable than organisms with proteins in the first region. Proteins in their early stage of evolution would have been selected for adaptability and therefore modern proteins would have structures in regions of survival surfaces where adaptability was optimal.

These considerations raise a final question: is the rate with which a protein finds an optimum on a survival surface fast or slow compared with the rate at which the surface itself undergoes upheaval as the environment changes? If motion across the surface is fast compared with the rate at which the surface changes, and if widely distant points on the surface are generally accesible, the protein is nearly always optimally adapted. If the rate of motion across the surface is slow compared with the rate at which the surface is changing, the protein is rarely optimally adapted.

A. Sources of Information

Several classes of experiments provide data that directly address the nature of these surfaces.

1. Experiments on the properties of proteins where point mutations have been introduced, either randomly or by directed mutagenesis, provide data regarding behavioral surfaces since they directly relate structural changes and behavioral changes. Information concerning survival surfaces can come from such mutants if the impact of the mutant on the growth and survival of a host organisms is examined. These experiments provide a direct statement of the structural change needed to produce a specific growth advantage/disadvantage under specified (but possibly arbitrary) conditions.[104]
2. Experiments in which proteins are forced to "evolve" under laboratory conditions provide information on the number of amino acid changes that are needed to produce a measurable change in selectable behavior and directly measure specific adaptive responses under specific conditions.
3. Experiments on the properties of homologous proteins isolated from various biological sources (i.e., correlation of behavior with structure) provide an upper limit on the structural variation required to achieve a given behavioral variation. Furthermore, these data provide direct information on the ability of evolutionary processes, drift or adaptation, to produce behavioral diversity. However, whether the diversity reflects drift or adaptation is generally not obvious.

1. Mutants

Over a century of biological experimentation in microorganisms, plants, fungi, and animals has produced countless mutants that have been mapped to specific proteins.[107] These mutant proteins are informative about the general nature of the surface-relating structure, survival value, and behavior. However, a sampling bias somewhat limits their usefulness. Mutants are generally isolated and scored only if they have a detectable influence on phenotype. Therefore, mutants that are behaviorally neutral are underrepresented in classical studies. Furthermore, many protocols for identifying mutant proteins with different behaviors are highly qualitative, and would not identify many mutants that retain activity at a lower level. Thus, a casual review of the literature would give the impression that point mutations are likely to be more deleterious than they actually are.

Nevertheless, the literature contains many examples where a small number of changes in a protein or nucleotide substitutions in a gene produce major changes in kinetic, physical, solubility, and other properties. Mutants arising spontaneously are now supplemented by a growing number of mutations made deliberately using recombinant DNA techniques. A wide range of behaviors are now known to be alterable by a single amino acid change. Kinetic behavior is especially sensitive. A point mutation (from Asp to Ser) causes the catalytic activity of phosphofructokinase to decrease by over four orders of magnitude.[108] Similar effects on the rate of electron transfer is seen upon point mutation of cytochrome c.[109] A single point mutation

(Arg to Met) renders mitochondrial ATPase inactive.[110] Replacement of Glu 49 in tryptophan synthase by all 19 amino acids leads to catalytically inactive protein.[111] A single amino acid change increases the rate of the chemical reaction in tyrosyl aminoacyl t-RNA synthetase.[112]

Point mutations can similarly affect binding by large amounts. Examples include single-stranded DNA-binding protein,[113] aspartate aminotransferase,[114] and lactose repressor.[115] Here, examples are more evenly divided among those that increase affinity and those that decrease affinity. Likewise, examples are now known where regulatory behavior is altered by point mutations, including those in proteins with allosteric regulations[116,117] and where regulation involves disassociation of subunits.[118] Indeed, Fersht and Lau have recently uncovered a point mutation that changes the effect of allosteric interactions from repression to activation.[119]

Examples are also known where a single (or small number) of point mutations influence a range of biological, physical, or catalytic properties. For example, single amino acid substitutions in ribosomal proteins raise and lower the fidelity of translation.[120] A single base change is sufficient to destroy the promoter activity in certain genes.[121] Six mutations in alcohol dehydrogenase from horse liver significantly alter the substrate specificity of the enzyme.[122] The stereospecificity of yeast alcohol dehydrogenase appears to be substantially destroyed by a point mutation in the active site.[123]

There are many examples where these behavioral differences can be clearly correlated with selective disadvantage. For example, a single amino acid substitution in 5-enolpyruvylshiimate-3-phosphate synthase alters the ability of the enzyme to be inhibited by the herbicide glyphosate.[124] A point mutation in the 16S rRNA molecule from *Nicotiana tabacum* renders the plant resistant to streptomycin.[124] A single amino acid change alters antigenicity and replication in an influenza virus.[126] A single amino acid substitution in the outer membrane protein of *E. coli* causes conjugation deficiency.[127] Point mutations in bacteria (*E. coli*) or fungi (*Neurospora*, yeast) that create auxotrophic phenotypes that are selectively disadvantageous in specific environments are a major tool of bacteriologists and molecular biologists.[128]

Sickle cell anemia is a classical example where a point mutation in a single protein has an impact on survival. Perutz reviewed variants of hemoglobin and myoglobin in view of their selective advantage.[28] In addition to mutant hemoglobins that are associated with disease, many variants are known that have no obvious impact on survival.

Furthermore, evolutionary divergence in structure appears to correlate with behavior in a fashion which suggests that such divergence is adaptive. For example, a single amino acid change in human interferon alpha 2 renders it biologically active on mouse cells.[129]

As the body of data expands, a list of enzymatic behaviors that can be changed by substituting a few amino acids grows. Hence, these behaviors are almost certainly likely to drift if they are not directly functionally important. Thus, historical models explaining their distribution are not likely to be valid; conservation principles needed for historical models to explain such behaviors in proteins are *a priori* unreasonable. Conversely, if these behaviors are conserved in two homologous proteins, they are likely to be conserved for functional reasons. Behaviors included in this category are kinetic behavior, substrate specificity, stability, regulatory behavior, stereospecificity, and biological activity.

The selection bias mentioned above can be avoided by examining essentially all of a collection of mutants created randomly. For example, Miller and co-workers collected nearly 100 nonsense mutations (those introducing a stop signal into the gene) in the gene coding for the lac repressor protein.[130] Since all nonsense mutants have selectable phenotypes (except perhaps for mutants very near the 3′ end of the gene), these mutants are biased only by the biases inherent in the mutation event. The mutants were extensively mapped and sequenced. The mutants were then placed into suppressor strains of *E. coli*. These suppressor strains introduced serine, glutamine, tyrosine, leucine, or lysine (depending on the strain) into the lac repressor at the position of the nonsense mutation. This process created a library of repressors with structures extensively varied along nearly 100 of the coordinates in the structural dimensions. The altered repressors were then examined for their ability to bind inducer and DNA.

The key result from these experiments was that introduction of any of these five amino acids at most positions in the protein did not influence the ability of the repressor to bind inducer. More strikingly, those mutations that did diminish the biological activity of the lac repressor to bind to inducer appeared at nearly regularly spaced intervals throughout the protein. These were called "hot spots" by Miller, positions in the polypeptide chain where almost any amino acid replacement led to loss of the ability to bind DNA or inducer.[130]

These results suggest that variations in repressor structure are generally behaviorally neutral, judged by a qualitative assay for an active lac repressor. The occurrence and distribution of "hot spots" was explained in terms of a two-domain structure for the repressor and a hypothesis that beta turns, which appear at approximately regular intervals in proteins, were both very critical for protein folding and very sensitive to minor alteration in sequence. Variations at these positions were presumed to prevent the protein from assuming its native folded state.

Furthermore, Miller found that the inducer- and DNA-binding properties of the protein behaved more or less independently. Mutations that altered the ability of the repressor to bind DNA did not in general destroy the repressor's ability to bind inducer. Finally, two mutants were found that bound DNA more tightly than wild-type repressor.

If we may generalize from repressor structure to proteins and behaviors, these experiments imply that for any behavior there exists a set of mutations which markedly influence the behavior. However, this set is only a small fraction of the total number of point mutations possible. Most point mutations do not lead to large behavioral differences. The behavioral surface is "rough" in only 5 to 10% of its dimensions. In the remaining 90 to 95% of the dimensions, it is smooth.

Miller's experiments do not address the nature of survival surfaces since the survival properties of *E. coli* with different repressors were not examined. However, the fact that two mutants were found that bound DNA more tightly than wild-type repressor suggests either that the tightest binding repressor is not evolutionarily the best repressor, or that tightness of binding in the wild-type repressor is sacrificed to obtain optimality in other behaviors. While the first option seems more reasonable, the second cannot be ruled out.

Of course, the behavior of point mutations tells us only about the shape of the surface around the wild-type structure. While the combination of rough and smooth dimensions suggests that the protein is behaviorally optimized in a "local" sense, there remains the question of whether the behavior surface has "mountains" in all dimensions separating local behavioral optima from "global" optima elsewhere on the surface.

Shortle and co-workers have published data suggesting that regions are not isolated.[131,132] They semi-randomly introduced a large number of missense mutations into the gene for staphylococcal nuclease and isolated the mutant genes by screening. This provided a collection of nucleases where a single point mutation had largely destroyed the catalytic activity or binding of the protein. These mutant genes were then further examined to find mutations at second sites that would restore the catalytic function of the protein. Several of these "second site suppressors" were found. The unexpected result was that several of these repressors were global, i.e., the introduction of a single mutation at one site would restore activity in many of the 77 missense mutants where activity had been lost.

The experiments of Shortle and co-workers focus on mutations that have a major impact on the catalytic behavior of the protein. These are the mutants that are easily detected by the screen. Thus, the mutants involve motion on the behavior surface along dimensions that are rough, where behavior changes substantially with small changes in structure. If a nuclease with low catalytic activity is presumed to place the organism at a selective disadvantage, these are movements on the surface that would not occur in an environment where the organism is under strong selective pressure.

However, the second site suppressors are mutations that suppress the detrimental effect of the mutation on catalytic activity and (we presume for the sake of discussion) presumably suppress

the deleterious effect of these mutations on survival value. Thus, the suppressor mutations allow motion on the surface that would otherwise be constrained by selective pressures. In other words, the suppressor mutations are paths on the surface that provide a path in one dimension around an insurmountable mountain in another dimension.

Again, Shortle's assay was crude. Mutants were examined only for gross catalytic competence and microscopic kinetic data were not collected. Furthermore, no data were collected regarding the survival of organisms with mutant nucleases. Therefore, no conclusion can be drawn about survival surfaces. However, as the nuclease presumably is involved in obtaining nucleic acid subunits from nucleic acid polymers in solution, some adaptive disadvantage is likely to be associated with a defective nuclease in some environments. However, we cannot say whether undetected catalytic differences between the wild-type nuclease and in nucleases with suppressed mutations would, given the substitution of one by the other in an organism, be "neutral".

If we generalize from Shortle's results, the implication is that paths around barriers on behavioral surfaces can generally be avoided by crossing over into another dimension on the multidimensional surface. This is a property of the high-dimensional surface that was anticipated on theoretical grounds above. If there are paths around general barriers, it is unlikely that regions of a surface are isolated from other regions.

Have such second site suppressor mutations been used in evolution to circumvent barriers? We do not know. However, in this regard, it is interesting to note that one of the second site suppressor residues uncovered by Shortle is present in a homologous nuclease from the Foggi strain of *Staphylococcus*.[133] If, as Shortle suggests, the second site suppressor generally stabilizes conformation, thereby permitting the introduction of mutants that destabilize the folded form, the second site suppressor may actually have played a role in the evolution of this specific protein in these particular organisms.

For point mutations that are damaging to the thermal stability of a protein, it is quite likely that there are second site suppressors that can correct the problem. After all, if stability is the result of an addition of many small stabilizing interactions (as many seem to believe), it is reasonable that the loss of a stabilizing interaction at one position can generally be compensated by introducing a stabilizing interaction elsewhere.

This need not be so with mutations that destroy catalytic activity, especially when the mutation destroys a group that is commonly known as an "essential" residue. For example, it is hard to see how a change elsewhere in chymotrypsin would compensate for changing serine to alanine in the active site.

However, in at least one case, exactly such a second site suppressor has been found. In their work with site-directed mutagenesis of triose phosphate isomerase,[134,135] Knowles and coworkers replaced a glutamate residue at the active site with an aspartate. The glutamate has been implicated as a base directly involved in the catalytic reaction. Not surprisingly, the k_{cat} of the mutant enzyme was smaller by a factor of 10^3.

However, a subsequent mutation restored a substantial fraction of the catalytic activity. Of course, one might argue that the alteration of glutamate to an aspartate is an extremely conservative change. Had the glutamate been altered to an alanine, one would imagine that a second site suppressor would have been much more difficult to find.

Examples of second site reversion are known in nonenzymic proteins dealing with more complicated behaviors. For example, introducing a charged residue into a signal peptide responsible for the secretion of a fused protein alters the overall hydrophobicity of the signal and diminishes secretion.[136] The change can be compensated for by increasing the overall hydrophobicity of the signal peptide. Mutant bacteria lacking certain methylases and demethylases do not display chemotaxis. However, strains have been recently reported that lack these enzymes, but nevertheless are capable of chemotaxis [137] as a result of a second site deletion.

In most cases, studies of this sort are not informative about survival surfaces. However, for

triose phosphate isomerase, the second site revertants were selected because they conferred evolutionary advantage to the host organism. This suggests a general statement that decreasing the catalytic activity by a factor of 100 in an enzyme that is being used in a microorganism almost certainly puts the microorganism at a selective disadvantage. Furthermore, these data again suggest a limitation on what nonfunctional behaviors can be presumed to be conserved in proteins. Since catalytic activity can be changed by changing a single amino acid residue, it is difficult to assume that similar catalytic properties of two homologous proteins are due to a nonfunctional constraint on drift. Again, a class of conservation principles that might be incorporated into a historical model are unreasonable *(vide supra)*.

These conclusions can be further supported by mutations introduced into beta-lactamase in *E. coli*. A mutant where the serine residue at the active site was converted to a cysteine shows a reduced, but still significant, catalytic activity.[138] The protein with reduced catalytic activity is less valuable for the survival of the bacteria that contain it. Richards and co-workers have found that introducing substitutions at the active site also leads to instability by heat and urea to denaturation, which is selectively disadvantageous in the presence of antibiotic.[139]

Two other "designed" variants are worth mentioning. Juillerat and Taniuchi[140] studied the effect of replacement of Leu32 in cytochrome c. The residue is invariant in the divergent evolution of this protein. Consistent with this fact, substitution with Ile had a significant impact on the properties of the protein. In contrast, Weissmann and co-workers examined structural variants of alpha-interferon. A number of alterations had no detectable impact on the antiviral activity of the protein.[141]

These results suggest that a behavioral surface is a mixture of rough and smooth dimensions. Such a surface is exactly the one where behavioral adaptation is expected to be the easiest. Were the surface smooth in all dimensions, many amino acid replacements would have to be made in order to get a significant change in behavior. Behavioral adaptation, even though unconstrained, would be slow. In contrast, were the surface rough in all dimensions, proteins would be isolated in local optima, unable to seek out global optima because they could only be reached via proteins with unacceptable properties.

One final point should be remembered. In this section, we have mentioned or discussed examples where point mutations influence catalytic activity, V_{max}, K_M, substrate specificity, binding strength, inhibition properties, antigenicity, level of expression, and stability in a protein. If a single point mutation can change these properties, the properties are expected to drift if they are not functionally constrained. Historical models that must assume these traits are nonfunctionally conserved are therefore implausible.

2. In Vitro Evolution

Mutants that destroy catalytic activity in an enzyme are uninspiring, even if some of the activity is recovered by mutation at a second site. More interesting are those mutations that create catalytic power, as these presumably are also the ones that create, rather than destroy, "survival value". How catalytic power is created is especially interesting if the creation of catalytic power in a protein requires the protein structure to traverse a substantial distance across the "surfaces" defined previously.

In several examples, an existing enzyme has evolved by point mutation to create an enzyme with a different substrate specificity as a selective response to an environmental challenge. These examples provide some information about the interaction between selective pressure and enzymatic catalysis. Many of these examples are based on the fact that when microorganisms are placed under selective pressure in the laboratory, they frequently undergo genetic alteration that allows them to adapt to new environments.[142-148]

The experiment places a microorganism in an environment where its fitness would be enhanced by a particular catalyst, but where the organism lacks a protein that has the appropriate catalytic properties. Often under these circumstances, the first adaptive genetic change is the evolution of a "constitutive overproducer", an organism that synthesizes a large amount of an

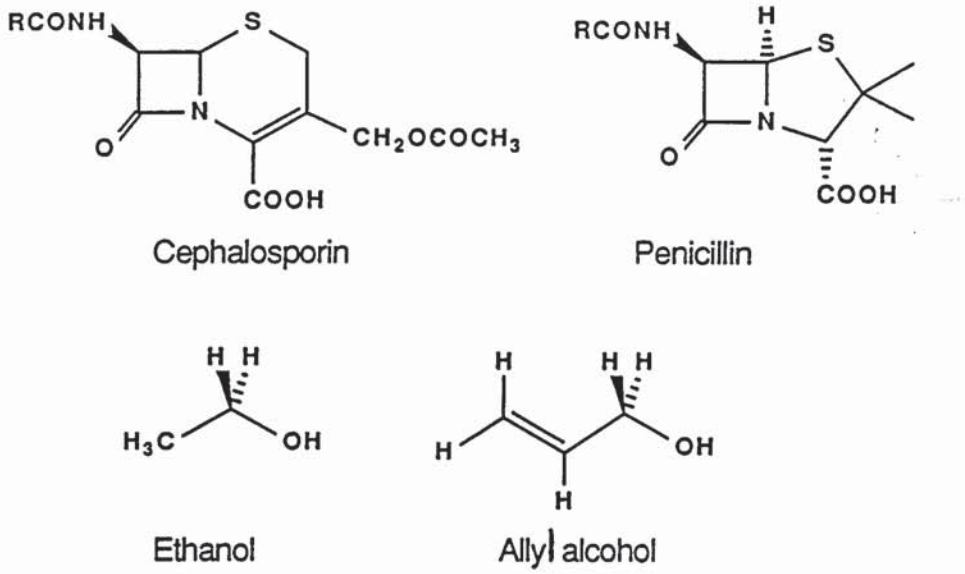


FIGURE 10. In *in vitro* evolution experiments, the substrate specificity of beta-lactamases has evolved from one favoring the hydrolysis of penicillin to one favoring the hydrolysis of cephalosporin. Likewise, alcohol dehydrogenases have been selected that act poorly on allyl alcohol, again in response to selective pressures.

enzyme that catalyzes the desired reaction rather poorly.[146] The frequency with which this adaptation is observed suggests that the probability of a mutation favorably affecting regulation of expression is higher than the probability of a mutation that favorably alters the microscopic catalytic properties of an enzyme. This suggestion is itself interesting.

However, in several cases, further evolution yields proteins with improved characteristics as a result of amino acid substitution in an existing protein. The fact that new proteins "evolve" in a short time (compared to geological time) suggests that the substrate specificity of enzymes is rather adaptable and that the creation of new behaviors is relatively simple. At the very least, this suggests that the behavioral surface describing substrate specificity is relatively smooth, at least in the dimensions along which the new function is created. We discuss some specific examples later.

When fed butyramide as a primary source of carbon, microorganisms have been found to evolve enzymes capable of hydrolyzing butyramide as a first step in their degradation.[146]

Beta-lactamases act only poorly on cephalosporin (Figure 10). However, when challenged with cephalosporin, enzymes with a single amino acid substitution have evolved that are better able to hydrolyze cephalosporin and less able to hydrolyze penicillin itself.[147]

Allyl alcohol is a substrate for alcohol dehydrogenase from yeast (Figure 10). However, the product, acrolein, is toxic. Thus, when yeast is grown in the presence of allyl alcohol, there is an advantage for an enzyme that has an altered substrate specificity that disfavors the oxidation of allyl alcohol.[148] This adaptation in substrate specificity has been observed in the laboratory when growing yeast in the presence of allyl alcohol.

These data show that substrate specificity can evolve significantly by creation of a small number of point mutations, given normal selective pressures. A certain degree of behavioral variation, in this case substrate specificity, has demonstrable impact on fitness, at least in a microorganism. Also, since substrate specificity can be changed by changing a single amino acid residue, it is difficult to assume that, if the substrate specificities of two divergent proteins are similar, similarity is due to a nonfunctional constraint on drift.

Occasionally, "laboratory selection" experiments provide information about the relative probability of events that permit an adaptive response. For example, Brooker and Wilson selected for mutants in lactose transport proteins in *E. coli* that would transport maltose.[149] A total of 18 mutants were isolated and sequenced. Surprisingly, all of the mutants with the desired phenotype were mutants at either position 177 or 236 in the transport protein. These mutants included proteins where Ala-177 had become Val or Thr, and where Tyr-236 had become Phe, Asn, Ser, or His. Variants at position 177 retained the ability to transport galactosides, while variants at 236 were defective in the transport of galactosides.

These results provide some insight into the first steps in the adaptation of a protein. At least in this case, there seem to be only two dimensions along which a lactose carrier can evolve to transport maltose by a single change. Along one of those dimensions, lactose- and maltose-transport properties are interdependent; along another, they are independent.

Of course, these experiments did not clearly explore the possibility of double changes that had the same phenotypic effect. Also, the microscopic kinetic properties of the altered transport proteins were not measured. However, the existence of a pair of behaviors (ability to transport maltose and ability to transport lactose) whose variation is correlated in changes along one dimension, and not correlated in changes along another dimension, is quite interesting and suggests that we cannot in general rule out *a priori* either that any two behaviors are independent of each other or that any two behaviors are interdependent.

An especially detailed investigaton into the adaptation of proteins under artificial selection pressure was done by Hall, Sinnott, and co-workers.[150-154] The gene for beta-galactosidase was deleted from *E. coli* and the organism grown with lactose as a carbon source. Under this selection pressure, the bacterium evolved a new enzyme for hydrolyzing lactose from another (as yet undefined) protein. The "evolved beta-galactosidase" gene emerged by two point mutations, and the detailed kinetics of the intermediate proteins have been studied. The evolution was clearly adaptive and a linear relationship was found between growth rate (with lactose as a carbon source) and the k_{cat}/K_m for hydrolysis of lactose of the evolving gene.[150-154]

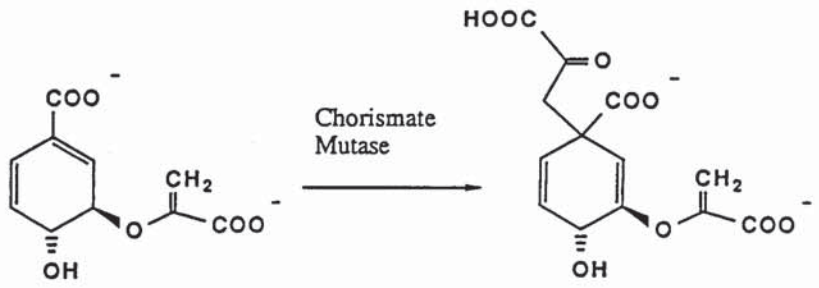
The kinetic data on the variant proteins showed an evolution of a free energy profile that was uneven, but overall improved. Consistent with the preceding discussion, these data again show that specificity and microscopic rate constants can undergo substantial variation by changing only a small number of amino acids. Again, one expects the values of microscopic rate constants and substrate specificity to drift unless they are functionally constrained.

All of the experiments described above involve the creation of a new catalytic activity from a protein that already displayed a related catalytic activity. These results might be viewed as unimpressive since, although they show that the specificity of binding can drift/adapt, they do not show the creation of a new type of catalytic activity. The hydrolysis of cephalosporin is akin to the hydrolysis of penicillin; the transport of maltose is similar to the transport of lactose. The complaint is similar to the creationist argument against the existence of macroevolution: of course, evolution within a species may occur, but there is never the creation of a "new kind".

The creation of a new catalytic activity does indeed appear to be a more formidable problem than the mere redirection of an existing catalytic activity toward a new substrate. One might expect such a "creative adaptation" to be infrequent or impossible in laboratory selection experiments. However, there are a few documented examples where it seems to have occurred.

The first committed step in the biosynthesis of aromatic amino acids is catalyzed by the enzyme 3-deoxy-arabinoheptulosonate-7-phosphate synthase (DAHP synthase). Consistent with this function, mutants in this enzyme in bacteria are auxotrophs for aromatic amino acids, and the enzyme is allosterically regulated by prephenate, the end product at a branch point in aromatic amino acid biosynthesis. This regulation is a result of an allosteric site on DAHP synthase that binds prephenate (but not chorismate, the immediate precursor of prephenate).[155]

In wild-type *B. subtilis*, prephenate is synthesized from chorismate by chorismate mutase, an enzyme that (formally) catalyzes a Claisen rearrangement on chorismate (Figure 11). However,



Chorismate Prephenate

A catalytic site for rearranging chorismate arose from an allosteric site for prephenate

FIGURE 11. A new "active site" which catalyzes the conversion of chorismate to prephenate appears to have evolved from an allosteric site in the last 40 years.

in a laboratory mutant that has been widely used in microbial studies over the last 40 years, the chorismate mutase gene is deleted. Remarkably, despite this deletion, this laboratory strain is not an auxotroph for aromatic amino acids. Another protein has taken over the job of converting chorismate to prephenate.

Llewellyn et al. analyzed this laboratory strain and came to the following remarkable conclusion.[155] In the time since the deletion occurred (perhaps 40 years), it appears as if the allosteric site on DAHP synthetase that binds prephenate has undergone mutation to bind chorismate. Furthermore, the site binds the transition state for the chorismate mutase reaction approximately 6.5 kcal/mol more tightly than chorismate itself. Thus, what was an allosteric site for binding prephenate has picked up point mutations necessary to convert it to an active site for the chorismate synthetase reaction. DAHP synthase is now a bifunctional enzyme catalyzing both the first and last steps in the biosynthesis of prephenate. Furthermore, in this new bifunctional enzyme, allosteric regulation by prephenate is not present.

It is not known how many amino acid substitutions were required to convert a binding site for prephenate into a chorismate mutase. The fact that it has evolved so rapidly suggests that either (1) there are few changes or (2) motion of a protein across a surface is quite rapid. However, as a potential example of a new catalyst arising from an old binding site, the result is remarkable. It supports the hypothesis, discussed later, that enzymes catalyzing consecutive steps in a reaction have evolved sequentially from one another. Furthermore, it represents an example of a process that is the reverse of an event assumed to be frequent: the evolution of an allosteric site from a catalytic site.

These experiments also provide insight into the limitations of laboratory evolution. The new chorismate mutase, having only 40 years to evolve, is not quite as catalytically active as chorismate mutases from other sources that have had millions of years to evolve. The evolved chorismate mutase produces a rate enhancement of 4×10^4, substantially less than the rate acceleration of approximately 2×10^6 found in the chorismate mutase from *Aerobacter aerogenes*.[156,157] There are several possibilities as to why the new chorismate mutase does not meet the same high standards.

First, it may be that the recently evolved chorismate mutase is trapped in a local catalytic optimum. A faster mutase may be accessible only via intermediate enzymes that have no mutase activity. Forty years may be insufficient time to accumulate the many mutations needed to evolve a faster mutase. A faster mutase may, in this bifunctional enzyme, be possible only at the expense of the DAHP activity. Each of these possibilities is a constraint on the evolution of a faster mutase.

Alternatively, it may be that the limiting factor in the growth of these mutants is not the chorismate mutase reaction and that no selection pressure is being brought to bear on the protein to force it to evolve to a better catalyst. This implies that, should the normal chorismate mutase lose catalytic activity by a factor of 100, its host organism would not be placed at a selective disadvantage, at least under the conditions where this laboratory strain is normally grown. As discussed earlier, this is unlikely, and the "artificially evolved" chorismate mutase, having done the easiest part of its evolution, may well be trapped in a local optimum, at least for the time being.

Have similar processes led to the creation of new catalytic activities in natural evolution? We do not know for sure, but there is strong evidence that it has. Several examples suggest that proteins with similar function are homologous. For example, a cytochrome P-450 reductase from pig seems homologous, at least in part, with glutathione reductase, ferredoxin reductase, and flavodoxin.[158] Carbamoyl-phosphate synthetase may have arisen by the fusion of two genes, one for a glutaminase and one for a synthetase.[159] Enzymes catalyzing consecutive steps in metabolic pathways, including the biosynthesis of methionine,[160] may have arisen from common ancestors. Ornston and colleagues have recently suggested that enzymes in the beta-ketoadipate pathway in *Pseudomonas putida* may be homologous.[161]

The conclusions from these studies is that protein stability, substrate specificity, microscopic kinetic properties, and, at least to a certain extent, reaction type can diverge easily if there is a selective pressure. This fact has an important implication. If the behaviors are *not* functionally constrained, these traits are expected to drift rapidly. Thus, it is reasonable to assume that two homologous enzymes with the same stability, substrate specificity, kinetic behavior, or reaction type have this similarity because of a functional constraint on drift. Historical models that require such traits to be conserved for nonfunctional reasons are not likely to be correct.

Furthermore, in constructing historical models, it is risky to treat these traits as primitive and to argue that enzymes sharing these traits are homologous. Finally, as is noted later, these traits in fact are not observed to drift rapidly in natural evolution. The conclusion, therefore, is most probably that these traits are functionally constrained from drifting.

3. Homologous Proteins

Data from mutant proteins show that most enzymatic behaviors, including kinetic parameters, substrate specificity, stability, stereospecificity, and even reaction type, can be changed by making small changes in primary structure. *In vitro* evolution experiments argue that such evolution does in fact occur under appropriate selective pressures. We now turn to the structures and behaviors of homologous proteins from nature, where it appears that the behavioral adaptability of proteins has been exploited by evolution.

Comparing homologous enzymes can provide a quantitative estimate of how much the structure of a protein must be changed to achieve a particular behavioral difference (the converse of the more common observation that two proteins with quite different structures can have the same behavior). Table 1 suggests a "scale" of behavioral variability. Kinetic behavior appears to have drifted/adapted (we still have not yet developed the arguments needed to decide which) faster than quaternary structure, quaternary structure faster than substrate specificity, substrate specificity faster than stereospecificity, stereospecificity faster than catalytic mechanism, and catalytic mechanism faster than gross tertiary structure.

Such a scale only approximately suggests the relative ease of divergence of enzymatic behaviors. However, even an approximate scale is valuable as we try to distinguish between functional and historical models designed to account for the kinetic, stereochemical, or physical behavior of enzymes.

A few examples illustrate how such a table can be constructed. Pancreatic and seminal RNAs have identical amino acids in 81% of the positions.[19] Yet their quaternary strutures are different (one is a dimer; the other is a monomer),[162] their substrate specificities are different (one acts on

Table 1
SPECIFIC EXAMPLES RELATING BEHAVIORAL AND STRUCTURAL DIVERGENCE

Variable by point mutation
Kinetic properties: k_{cat}, K_M, internal equilibrium constants
Regulatory properties, allosteric inhibition, and activation
Thermal stability, substrate specificity, solubility, biological activity

Variable with 10% sequence divergence
Substrate specificity
Ethanol vs. sterols (liver alcohol dehydrogenase) (2%)

Variable with 20% sequence divergence
Quaternary structure
Seminal (dimer) and pacreatic (monomer) ribonuclease (19%)

Variable with 30% sequence divergence
Number of introns in gene
Lysozyme

Variable with 40% sequence divergence
Substrate specificity[235]
2-Oxoglutarate dehydrogenase vs. pyruvate dehydrogenase 45%

Variable with 50% sequence divergence
Substrate specificity[236-238]
Cycloisomerase 1 vs. clc B muconate cycloisomerase
Tryptophan hydroxylase vs. phenylalanine hydroxylase (49%)
Tyrosine hydroxylase vs. tryptophan hydroxylase (55%)
Phenylalanine hydroxylase vs. tyrosine hydroxylase (55%)
Acetylcholinesterase vs. butyrylcholinesterase (45%)
Mechanistic differences[239]
Superoxide dismutases (Mn vs. Fe) 49%
Reaction type[240]
Phosphoglycerate mutase vs. diphosphoglycerate mutase (49%)

Variable with 60% sequence divergence
Intradomain disulfide bonds
Mammalian RNase[4] vs. turtle RNase and angiogenin[3] (60%)
Reaction type[241]
Tryptophan synthetase vs. threonine synthetase (58%)

Variable with 70% sequence divergence
Reaction types[242]
Eukaryotic repressor qa-1S *(Neurospora crassa)* vs. Shikimate dehydrogenase *(S. cerevisiae)* 26%

Variable with 80% sequence divergence
Reaction type
Fumarase vs. aspartase (76%)

Variable with >80% sequence divergence
Mechanistic differences and stereospecificity[243]
Alcohol dehydrogenases (Zn^{2+} vs. no Zn^{2+})
"Essential" active site residues diverge
Lysozymes (homology by X-ray structure)
Reaction type[244,245]
Malate synthase (cucumber)/uricase (soybean nodule)/glycolate oxidase (spinach) (marginal alignment)
Fructose-2,6-bisphosphatase vs. phosphoglycerate mutase (active site peptides)
Different binding types[246]
Corticosteroid binding globin vs. serine protease inhibitors

single-stranded nucleic acid; the other prefers double-stranded nucleic acid),[163,164] and their biological activities are different (seminal RNase has potent antitumor activity; pancreatic RNase has essentially no antitumor activity).[165,166]

Further divergence in behavior is seen as the homologous proteins drift still farther apart. "Angiogenins", proteins thought to have a role in the vascularization of solid tumors, have sequences that are only 40% identical to bovine pancreatic RNase.[167] If, as suggested, angiogenins are site-specific RNAses,[168] a sophisticated variation in substrate specificity has evolved during this divergence. Furthermore, the number of disulfide bonds has drifted as 60% of the amino acid sequence has drifted. Angiogenins have at most three disulfide bonds; mammalian pancreatic RNAses have four. Similar divergence is seen in turtle RNase, which has at most three disulfide bonds.[169]

A full range of divergence can be seen among alcohol dehydrogenases. The three isozymes from yeast (95% identical) of alcohol dehydrogenases have different substrate specificities and stabilities.[170] The enzymes from yeast and horse liver (30% identity) have grossly different substrate specificities, kinetic properties, and quaternary structures.[171] Glucose dehydrogenases and ribitol dehydrogenases (25% identity)[172] catalyze somewhat different types of redox reactions. Glucose dehydrogenase catalyzes the oxidation of a hemiacetal, while ribitol dehydrogenase catalyzes the oxidation of a simple alcohol. The alcohol dehydrogenases from *Drosophila* and yeast (with 25 to 30% sequence similarities in only one domain) have different catalytic mechanisms, different substrate stereospecificities, and different cofactor specificities.[16] The enzyme from yeast uses a metal ion; the enzyme from *Drosophila* does not. Stereoselectivity is opposite, both at nicotinamide and at ethanol.[16]

Table 1 provides an upper limit for the amount of sequence divergence that is needed to produce a specific divergence in behavior. For example, trypsinogen and chymotrypsinogen have sequences that are 39% identical. They have different substrate specificities. Thus, fewer than 61% of the residues must be changed to change substrate specificity. Exactly how many differences are necessary to account for the divergent substrate specificity is not known. It has been argued[173] that a single change in sequence could produce the same effect. This argument can be directly tested using site-specific mutagenesis.[174,175]

If the divergence in sequence is too great to ascertain homology with confidence, alternative methods must be used (e.g., crystal structures, *vida supra*) to establish homology. If one assumes that proteins with similar tertiary structures are homologous, families of homologous proteins differ in virtually every property measured. This suggests that all catalytic behaviors in an enzyme can diverge more rapidly than gross tertiary structure.

Such tables can be useful to form general expectations about how rapidly traits should drift. For example, Vennesland has suggested that the pattern of protein behavior can be used to construct evolutionary trees.[176] The behavior she considered most appropriate was stereoselectivity. This suggestion is based on the presumption that stereoselectivity is more highly conserved than sequence and that it therefore is a better indicator of homology than sequence.

The suggestion is reasonable. However, there are now a small (but growing) number of examples where stereoselectivity has apparently diverged in proteins where homology is still apparent from a comparison of sequence data.[16] These examples suggest that stereoselectivity is not sufficiently well conserved to serve as an indicator of distant homology and may be less satisfactory as an indicator of homology than alignments of primary sequence.

The extreme conservation of tertiary structure is most remarkable. For example, some suggestions that mitochondrial and nuclear genes have different genetic codes arise from a comparison of the genes for proteins, such as cytochrome c oxidase which both organelles synthesize. It is remarkable that the mitochondrial and nuclear proteins have sufficiently similar sequences to permit this comparison. This may indicate that there was a transfer of information between the nucleus and the mitochondria after the two genetic codes diverged. Alternatively, it may indicate that codon use has drifted faster than cytochrome oxidase structure, at least at some point in their evolution.[177]

Table 1 suggests that there is no effective barrier that prevents any enzymatic behavior from diverging. Naively, this argues against models that explain conserved enzymatic behaviors historically. However, this argument is not strong unless we know the absolute rate at which a sequence, unconstrained by function, drifts. If such drift is slow, behaviors might be the same in homologous proteins not for functional reasons, but rather because there simply has not been enough time for the structure to have diverged.

Several estimates can be made from the rate of structural drift that is not functionally constrained. For example, introns have no apparent function in many eukaryotic genes and drift in their structure is believed to be unconstrained by function.[178] The choice of codons in higher organisms is likewise generally unconstrained from drifting. Finally, the rate of drift of the structure of pseudogenes, nonfunctional remnants of gene duplication that do not code for a functioning protein,[179,180] is also unconstrained by function.

The rate of structural divergence of pseudogenes, introns, and third positions of codons is on the order of 10^{-9} to 10^{-8} changes per site per year.[180,181] For example, the rate of drift of globin pseudogenes is 3 to 4×10^{-9} changes per site per year.[182,183]

Thus, about 100 million years is required for 50% divergence in the structure of homologous genes provided there are no functional constraints on drift, while 500 to 1000 million years is required for sequence identity to be lost entirely. The first number is on the order of time between major geological epochs. The latter number is on the same order as the time that multicellular life has been present.

Thus, if divergence in behavior requires less than a 50% divergence in structure, there would have been sufficient geological time since the divergence of the various mammalian orders for the behavior to have diverged if it was not functionally constrained. Likewise, if not functionally constrained, behavioral divergence corresponding to an 80% divergence in structure will have occurred in the time that has elapsed since the divergence of the different phyla of multicellular life. Conversely, if behavioral divergences of these levels are not seen, conservation of such behaviors is likely to reflect function.

B. Survival Surfaces

This statement is significant. However, it is only a small part of the information needed to understand the interaction between natural selection and behavior in proteins. It remains to be asked whether differences in behavior in homologous proteins reflect drift or adaptation. Furthermore, no statement has yet been made about whether behavioral "neutrality" in the eyes of a bioorganic chemist (e.g., a divergence in K_M of "only" 5%) corresponds to functional "neutrality" in the eyes of natural selection. It is frequently worth reminding ourselves in this context that a difference of only 1% in the generation time of a microorganism is sufficient to permit it to dominate rapidly a resource-limited culture.

If small (judging from the bioorganic chemist's point of view) changes in behavior have a significant impact on survival value, behavioral drift might well be highly constrained. Furthermore, if small behavioral differences in macromolecules are selectable, this implies that macromolecules with subtly different behaviors will be optimal in subtly different environments. Therefore, different behaviors in homologous proteins might well reflect adaptation of the proteins to different environments. Finally, the implication that movement across a surface is highly constrained, that only a small fraction of the surface may have been examined in geological time, in turn implies that the optimum chosen may be quite different from a global one.

Alternatively, if large changes in behavior from the bioorganic chemist's point of view have little impact on survival, much of the behavioral divergence that is observed should reflect drift. Since this view implies that there are relatively few functional constraints on drift, it is less likely in this case that functional constraints have prevented the protein from achieving optimal behavior in those traits that are important for survival.

How good is selection at distinguishing between proteins whose behavioral differences are quantitatively small? What variations in performance at the molecular level are felt by natural selection? These questions address the shapes of suvival surfaces, those relating structure of a macromolecule as the independent variable with "survival value". As noted previously, this surface is a more complicated one.

It is clear that the divergence in structure of some proteins is under little functional constraint. This conclusion can be securely drawn by comparing the rate of divergence in protein structure with the rate of divergence of pseudogenes, introns, and other structures that are clearly not functionally constrained from drifting *(vide supra).*

For example, drift in albumins is generally assumed not to be strongly functionally constrained. Genetically transmitted variations in the structure of albumin in humans are known, but these variations are associated with very few clinical symptoms.[184] The divergence of albumins is on the order of 5 to 7×10^{-9} changes per site per year, comparable to the rate of drift of pseudogenes. Wide variation in the structure of albumin appears to have little impact on the fitness of the host organism. The survival surface in this protein appears to be rather flat.

Similarly rapid divergence is seen in certain globin genes[185,186] and in the structure of the C peptide region of the preproinsulin gene. The latter, which is removed and discarded before the hormone becomes biologically active, diverges at a rate of 7×10^{-9} mutations per site per year.[187] The rate of change of fibrinopeptides is 6×10^{-9} per site per year.[188] The rate of amino acid substitutions of the alpha-fetoprotein gene is 1.5×10^{-9} per site per year.[189] In each case, the rapid rate of drift is consistent with few functional constraints on the structure of the protein.

If structural divergence in proteins is faster than the rate displayed by pseudogenes, such divergence is regarded as adaptation to a rapidly changing environment. The scarcity of examples of proteins that more rapidly evolve has often been used as evidence in favor of the neutral theory of evolution. However, the rapidly diverging growth factors[190] may be a case of adaptation in a protein, as might be the rapidly diverging serine protease inhibitors and ovomucoids.[190,192]

Slower rates of divergence can be interpreted as evidence for functional constraints on divergence. Divergence at a rate comparable to that observed in pseudogenes is normally considered to be evidence that the structure is not under selective pressure. For example, the structure of cytochromes diverge at a rate of 0.2×10^{-9} per base per year. This slow divergence may be explained functionally. One might argue that the cytochromes evolved to selective "perfection" long ago and that recent behavioral variation is slow because the protein cannot improve on that perfection. However, the slowness of drift alternatively results from the existence of many constraints on adaptation, constraints that isolate the cytochromes that we have from other cytochromes that, were they only accessible, could have far better survival value. For example, cytochrome c is a substrate of other proteins and therefore cannot undergo structural drift without the drift of other proteins as well. Here, the presence of "multiple users" constrains drift. In this model, cytochrome c's might well be trapped in a local optimum determined by historical accidents early in their evolution.

It is clear that, in some cases, natural selection tolerates essentially no variation in the structure of an enzyme. For example,[21] 11 genes for alcohol dehydrogenase were independently cloned and sequenced from 5 geographically distinct populations of *Drosophila melanogaster*. Structural variation at the level of the DNA was abundant: 43 positions were found to be polymorphic — 14 polymorphisms in the 3 coding regions (765 bp), 11 in the adult intron (6654 bp), 7 in 2 other introns (135 bp), 3 in the 5′ and 3′ nontranslated sequences (332 bp), and 8 in the 5′ and 3′ flanking regions (863 bp). Of these 43, only one polymorphism in the gene influenced the structure of the protein. This variation (Thr vs. Lys at position 192) has been long known and appears to be an adaptive variant for *Drosophila* living in different environments.

The absence of structural variation in the protein in this population is remarkable. Given the level of gene polymorphism, one would expect about 35 nonsilent mutations to be found at the

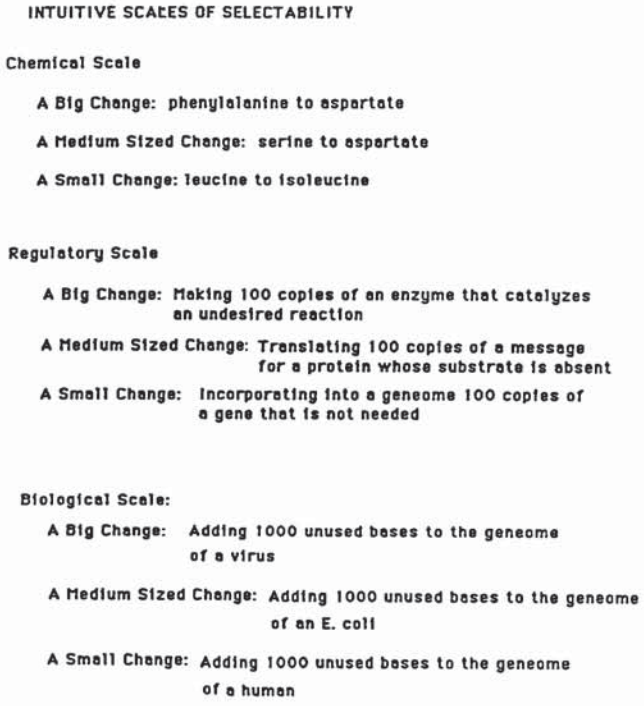


FIGURE 12. Three intuitive scales of selectability can be defined, one "chemical", one "regulatory", and one "biological".

level of protein. The fact that these were not found suggests that they were selectively disadvantageous to the fly. This result is especially remarkable because 28% of the amino acid residues in the protein are either Val (22 residues), Ile (23 residues), or Leu (26 residues). Yet no "conservative" interconversions of these three appear to be tolerated by natural selection. Apparently, if one adds or subtracts a single methylene group in this protein with a molecular weight of 27,000, the fly dies.

Site-directed mutagenesis on alcohol dehydrogenase from yeast[247] suggests that at least some of the point mutations eliminated by natural selection would, *in vitro*, have kinetic and physical behaviors essentially indistinguishable from wild-type enzyme. This suggests an enormously stringent functional constraint on behavioral drift in this protein.

Thus, some proteins (such as alcohol dehydrogenase from *Drosophila*) are almost certainly under enormous selective pressure; essentially no divergence in structure (or behavior) is tolerated. In contrast, some proteins (such as albumens) appear to be under very little selective pressure; considerable variation in structure is tolerated. Complicating the analysis is the possibility that divergence can be adaptive (as in growth factors or protease inhibitors).

1. Scales of Selectability

The literature supports three general statements regarding the influence of natural selection on macromolecular behavior (Figure 12). The first is simply that structural changes that produce a bigger impact on behavior *(in vitro)* are more likely to be selectable than those that are smaller (the "chemical" scale). The second (a "regulatory" scale) suggests that the impact of a structural perturbation in macromolecular structure depends on the position of the perturbation in the hierarchy of information flow in an organism. Finally, a "biological" scale suggests that the

impact of a given behavioral diversity is larger on the survival of small organisms than big organisms.

The first scale relating the impact of a macromolecular structural variation on survival to the magnitude of its impact on the *in vitro* chemical behavior of the macromolecule is intuitive. Altering a lysine to an arginine in a protein is likely to have less impact than altering a lysine to an isoleucine. Altering residues lying on the surface of a protein is likely to have less impact than altering one inside the protein. A change of 5% in the k_{cat}/K_M of an enzyme is expected to have less selective impact than a change of 50%.

These intuitions are supported by fact. Surface residues drift faster than residues interior to the protein. "Conservative" substitutions are observed more frequently than "nonconservative" substitutions, even after correcting for the degeneracy of the genetic code.[193] In hemoglobins, many of the the changes that chemical intuition suggests are "neutral" seem to be neutral *in vivo*.[28]

A second scale relates impact on survival to the position in the hierarchy of information transfer where the structural perturbation occurs. Thus, including unused DNA in a genome incurs a cost associated with the replication of the DNA at each generation.[194] This cost is expected to be small, corresponding to two ATP equivalents per excess base per generation, plus the cost required to synthesize the additional bases, plus perhaps an "overhead" cost associated with maintenance and repair.

More expensive is the undesired deregulation of the gene leading to the synthesis of unused protein. In a "typical" prokaryote, a gene can make 10^3 copies of a protein per minute, requiring 4 ATP units for each amino acid (including the cost of the message) in each protein expressed, the ATPs needed to synthesize the amino acids, and additional overhead costs. Thus, expressing unused protein in *E. coli* is roughly 10^6 times more expensive per gene than carrying excess DNA.

Most expensive is the undesired catalytic activity of an expressed protein, provided that the protein catalyzes a reaction that is not physiologically at equilibrium. As enzymes dissipate free energy by allowing a system coming to chemical equilibrium, the cost of an enzyme doing undesired reactions is the cost of undoing those reactions. Hence, a protease can cut 10^3 peptide bonds per minute, requiring the expenditure of roughly 10^4 ATP molecules to repair them. Thus, the catalytic activity of an undesirably active protein costs roughly 10^{12} times more than carrying the DNA silently in the chromosome.

Such a scale is qualitatively entirely consistent with what is known about genetic regulation and evolution in microorganisms. However, it is important to note that measuring quantitatively the "cost" associated with the waste of energy in synthesizing an unused protein remains an unsolved challenge.[104,195-198]

The biological scale argues that the impact on survival of a specific perturbation in macromolecular behavior will depend on the size and complexity of the host organism and will be smallest in multicellular organisms, intermediate in single-cell eukaryotes, and largest in prokaryotes and viruses. This scale is also not unreasonable. A variation in behavior of a specific magnitude in a bacterium represents a larger fraction of the "total behavior" than it does in a mammal. Furthermore much of the survival of a multicellular animal presumably depends on physiology, a factor influencing the survival of bacteria far less.

The value of such scales is that they allow *a fortiriori* arguments about the selectability of behaviors for which we have no specific data for other enzymes in other organisms where we have solid evidence that a behavior is selectable. For example, any structural variation observed in the alcohol dehydrogenase in *Drosophila* probably reflects functional adaptation. The biological scale suggests that similar structural variation in the isozymes of alcohol dehydrogenase in yeast is similarly functional. This appears to be the case.[18]

Experimental data are generally consistent with these scales. Supporting the "regulatory scale" is the fact that all organisms carefully regulate the expression of proteins and the catalytic

activity of expressed proteins. However, multicellular organisms appear to carry excess DNA, while microorganisms such as *E. coli*, if burdened with excess DNA (as on plasmids), generally lose the extra DNA in a process commonly referred to as plasmid "curing". Finally, viruses are especially conservative with DNA; the overlapping genes found in several viruses are frequently interpreted as a result of a functional constraint preventing viruses from surviving in competition with others if they carry too much DNA.

To illustrate the value of this approach, let us examine a commonly discussed problem in macromolecular structure: is the choice of triplet codons for amino acid functional or nonfunctional? The literature suggests both possibilities. Arguments favoring functional choice of codons often are based on a detection of a bias in codon usage in a given organism. A functional rationale for the biases observed is that the level of expression is controlled by codon choice; this model assumes that the levels of tRNA for different codons are different. Arguments against function are based on an absence of bias in codon usage and the existence of drift or polymorphism in codon use. According to the "biological" scale, codon usage should matter more in viruses than in bacteria, more in bacteria than in unicellular eukaryotes, and more in unicellular eukaryotes than in multicellular eukaryotes.

A series of elegant studies by Weissmann and co-workers provides an interesting argument that codon usage can be selected in viruses.[199] These workers examined the fingerprint patterns of RNA from Q-beta phage by digestion with RNAse T1, a procedure that detects structural variation in about 10% of the gene. These experiments detected "polymorphism" in the structure of the phage gene in a wild-type population. This means that a population of viruses is in fact a mixture of viruses with slightly different DNA sequences. Each "variant" phage differs from the most abundant variety of phage by a few bases in the genome.

Weissmann and co-workers further discovered that when a single phage was cloned and regrown, the polymorphism reappeared. The polymorphism arises from spontaneous mutations and suggests that the population of phage is "searching" the surface in the vicinity of the "wild-type" structure. Finally, and most interestingly, essentially the same distribution of variants were obtained in clones descended from different variants.

These results suggest that some structural variation in the virus only slightly perturbs the ability of the virus to grow. Slightly disadvantageous structures are each represented as a somewhat smaller fraction of the total population, while slightly advantageous structures are represented by a somewhat larger fraction of the total population. If one structure is isolated and then regrown, it rapidly mutates to reestablish the distribution of structural types that was present in the original population.

One of the variants present in the population was found to have a substitution of U for C that converted a UUC triplet (Phe) to the synonomous UUU triplet. Competition experiments showed that this variant was at a selective disadvantage compared to wild-type phage. Weissmann and colleagues did not establish a mechanism for this selective disadvantage. Two possibilities were considered, one that the change in codon influenced survival of the phage at the level of translation of the gene, the other that the base change influenced a structural behavior of the nucleic acid important to packing. However, the result of this experiment suggests rather strongly that codon selection is not completely neutral, at least in viruses.

However, the experiments also suggested conclusions about what structural perturbations could be viewed as neutral in this virus. In the population, the average number of nucleotide variations from the norm was about 1 to 2 per RNA molecule. Thus, only about 14% of the phages in the population had the "wild-type" genetic structure. Furthermore, a purely statistical analysis suggested that the population of 10^{11} viruses could contain representatives of all selectively neutral single, double, and triple mutations and about 0.05% of all quadruple mutants. While the number of representative examples of each type was not measured for obvious reasons, one can conclude from these data that the optimum on the survival surface in this virus is surrounded by walls that are fairly steep.

Based on these studies, Weissmann concluded that "probably none of the Q-beta variants which we have observed could be classified as neutral in the sense of Kimura and King and Jukes."[199] The study makes a strong statement about survival surfaces at the bottom of the "biological scale". Virtually any change in structure has an impact on the survival of a virus. However, for structures clustered around the native structure, the selective disadvantage is not so great as to prevent a heterogeneity in the population that could form the basis for future evolution, should the environment change. Finally, should the environment change, the phage is expected to evolve extremely rapidly to a new genetic structure.

Other studies, while less complete, suggested similar conclusions. The complete sequence of bacteriophage T7 has been examined for bias in codon use; the conclusion reached was that the codons selected reflect the abundance of different types of host tRNA, tRNA-mRNA interactions, and absence of short palindromes.[200] The last perhaps reflects functional adaptation to avoid digestion by restriction endonucleases.

The authors also observed a bias in this virus in favor of codons with the general structure purine-nucleotide-pyrimidine. Surprisingly, they considered a historical model rather than a functional model to explain this observation. The model was based on a proposal by Eigen and co-workers [201] which assumes that the original genetic code evolved from a primitive code composed only of (purine-base-pyrimidine) triplets. In view of the data discussed earlier and the possibility that selection is functional, any model of codon usage that reflects the structure of presumed primitive code (established 1 billion years ago) seems most unlikely.

In bacteria and yeast, analysis of codon usage suggests that codon selection is biased. Codon usage has been correlated with tRNA abundance in *E. coli* and in yeast, which presumably is the mechanism by which codon biases are selected and prevented from drifting.[202-205] Again, the variation in codon usage is presumed to be functional, perhaps controlling the level of gene expression. Furthermore, Grosjean et al. have suggested that codons with intermediary GC contents are optimal for translation.[206]

Evidence is far less convincing that there is a bias in codon usage in higher organisms. The rate of drift in codon use is only slightly slower than the rate of drift in the structure of pseudogenes (*vide supra*) or introns. It has been argued that this difference reflects some selective pressure on codon use.[179]

There have been many suggestions that the selection of codons in higher organisms is biased.[207,208] Hamada et al.[208] have suggested that codon usage is selected for structures that form the Z conformation of DNA and that Z structures might have functional roles in gene regulation, recombination, or expression. Other repeated structures are known in eukaryotic DNA.[209,210] Others have argued that codons in human globin genes are functionally selected to increase mutational stability.[211]

Finally, there are suggestions that the choice of codon depends on the context. For example, Lipman and Wilbur have suggested that the choice of the third base in a codon is influenced by the 3′ neighboring codon.[212] Blaisdell has suggested that in eukaryotes the optimum codon is a function of bases up to two codon sites away.[207]

In general, these suggestions are based on biases that are statistically marginal. Indeed, in many multicellular organisms, no bias at all can be detected. For example, in rat, mouse, and bovoids, codon selection is essentially random; in humans some bias was reported.[213] Again, the rate at which the third codon "drifts" is rather fast.[179] For example, in the gene coding for preproinsulin, the rate of drift in codon usage is 4.6×10^{-9} per site per year, inconsistent with a strong functional constraint. However, this rate of drift is some three times smaller than the value for mitochondrial genes.[214]

Finally, Kreitman's work with *Drosophila*[21] discussed previously argues against strong selection pressure on codon use in *Drosophila* and, given the "biological scale" discussed earlier, in more complex organisms as well. Within a population, Kreitman found considerable variation in the choice of codons for the gene for alcohol dehydrogenase. This contrasted

strongly with amino acid selection. The codon polymorphism is most easily explained in terms of selective neutrality or near-neutrality for codon usage, at least in this protein in this organism. Nevertheless, there has been some argument that codon selection in *Drosophila* is more biased in genes coding for proteins that are highly expressed than in those that are expressed only at low levels.

These data are clearly consistent with the notion of a "biological" scale. In viruses, codon choice is almost certainly selected, and the extent of selection has been measured. In bacteria, codons also appear determined by natural selection. However, in *Drosophila*, data on codon polymorphism suggest strongly that codon selection drifts as if incompletely constrained by function. Finally, in mammals, the codon selection is nearly random, suggesting that there is little selection pressure favoring one codon over another.

Chemical intuition also suggests that certain classes of protein function will be more sensitive to structural variation than others. For example, binding appears (based on a variety of model studies) to be less sensitive to structural variation than to catalysis. This is consistent with the apparent ability of natural selection to tolerate structural variation in binding proteins, when compared with catalytic proteins. The classical example in human biology is hemoglobin, developed by studies of Perutz and co-workers.[215,216] Over 200 variants of hemoglobin are now known. In some cases, variants are associated with disease, most notably sickle cell anemia. Here, homozygosity in a protein with a single-point mutation has a measurable selective disadvantage (although the heterozygote may have selected advantage).[217] However, most hemoglobin variants are not associated with a disease and do not obviously influence the survival of the host.

2. Polymorphism

The absence of polymorphism at the level of the protein in the presence of silent polymorphism at the level of the gene has been interpreted (above) as an indication of functional constraints on structural drift. The presence of polymorphisms in some proteins from yeast,[218] *E. coli*,[219,220] *D. melanogaster*,[221] and man [222] is more difficult to interpret. *A priori*, they are equally likely to reflect neutral drift (it makes no difference to natural selection what amino acid is found at certain positions) or functional adaptation to slightly different environments (it is critically important what amino acid is found at certain positions, so much so that the amino acid must be altered if the environment changes slightly).

To illustrate the difficulties in interpreting polymorphism, consider the one polymorphism at the level of the protein that does exist in the alcohol dehydrogenase from *Drosophila*. Which variant is found appears to correlate with latitude and altitude.[223] This suggests (but does not prove) that different forms of the dehydrogenase are preferred depending on mean temperature or other environmental variables. In this light, the sequence differences between the DADH in *D. simulans* (differing at 2.8% of the positions) and in *D. mauritiana* (differing at 3.1% of the positions) [224] seem more likely to reflect adaptation than drift. However, this appearance has not gone unquestioned.

In the absence of detailed studies on the distribution of polymorphism in a population, it is impossible to know whether the polymorphism is adaptive or reflects neutral variation. Thus, occasional discussions that assume polymorphism is neutral or, conversely, that polymorphism is adaptive need reevaluation.[225,226]

As mentioned previously, functional constraints may be different for different types of proteins. For example, there may be fewer constraints on the drift of a binding protein (such as hemoglobin) than there are on the drift of a "binding and catalyzing protein" (such as alcohol dehydrogenase). Indeed, even within alcohol dehydrogenases, enzymes acting via one mechanism (e.g., requiring a metal ion) may be more or less constrained than those acting via an alternative mechanism.

A scale that ranks protein types in order of functional constraint would be most desirable.

Proteins with fewer functional constraints might be expected to drift more rapidly than those with more functional constraints. Constructing such a scale would presumably consider stereoelectronic, binding, and other constraints on the reaction itself and may well be a goal for future work. For example, one possible view, so far not defended in the literature, is that the functional constraints on enzymes that catalyze stereoelectronically demanding reactions are higher than on those that do not.

3. Correlating Specific Behaviors with Survival Value

Correlating variation in specific behaviors with variation in survival value is often a complicated task. In extreme cases, adaptive behavior is well known and generally undisputed.[25] Proteins from thermophiles are more stable and this stability is directly related to and functionally necessary for adaptation to higher temperatures. However, in most behaviors studied by bioorganic chemists, the correlation between behavior and environment is far less apparent.

Only a few studies exist that attempt to understand the degree to which behavioral variation is under selective influence. An especially elegant study concerns the two isozymes of the lactate dehydrogenases from the killifish *Fundulus heteroclitus*..[227] The structure of lactate dehydrogenase B (also known as "heart type") varies in different individual fish drawn from waters along the Atlantic coast. The allozymes have different temperature optima, and the relative frequencies of the allozymes correlate with the latitude where the fish were collected. The allozyme with the higher temperature optimum was found more abundantly in the southern waters; the allozyme with the lower temperature optimum was found more frequently in the northern waters.

Thus, it appears that a rather subtle kinetic behavior of this protein was under the control of natural selection. Similar studies on lactate dehydrogenase (LDH) from cow and fish showed that the temperature optima *in vitro* correlated with the temperature in the environment where the enzyme was adapted.[228] Stability of a protein to hydrostatic pressure appears to correlate with the depth of water in which a fish lives.[229]

Correlation between structure, a parameter of the environment, and the relevant bioorganic property (in this case, the temperature dependence of the kinetic parameters) presents an elegant picture of the interaction between environment and subtle macromolecular behavior in a single species. Similar examples are now beginning to appear for several other behaviors of enzymes,[25,230-233] suggesting that natural selection is capable of highly refining the kinetic behavior of proteins.

4. Site-Directed Mutagenesis Studies

Deliberately altered genes can, at least in principle, be returned to microorganisms to directly determine how behavioral perturbation in a single protein influences the survival of the host. Recent work with alcohol dehydrogenases from yeast[247] suggests that minor behavioral perturbations (from a bioorganic chemist's point of view) can be strongly selected.

However, there is an alternative approach that can be followed, provided that specific behaviors of large numbers of random mutant proteins have been measured. To illustrate, let us assume that the k_{cat}/K_M value of an enzyme is maximized (i.e., the enzyme with the highest value provides the highest survival value), while the disassociation constant for a substrate (K_s) is optimized to an intermediary value. The k_{cat}/K_M and K_s values of the mutant proteins will show different distributions around these values for the native protein. Specifically, most of the mutants will have lower values for k_{cat}/K_M than the wild type. In contrast, the K_s values will be distributed uniformly above and below the K_s value of the wild-type enzyme.

Of course, the numerical values for nonselected behaviors are also expected to be distributed uniformly around the value for the native enzyme. Thus, a uniform distribution will not (necessarily) indicate whether an experimentally measurable parameter is a target of natural

selection or not. However, nonuniform distributions will indicate that both selection and selection pressure favor an optimization of a parameter.

In the case of tyrosyl aminoacyl-tRNA synthetase, Fersht and co-workers have accumulated enough quantitative data [234] for such an approach to be applied. In this enzyme, it does appear that k_{cat} is maximized, while various K_s values are not. This does not exclude the possibility that these K_s values are *optimized*, only that their optimal value is neither a maximum nor a minimum.

In this regard, it is interesting to note that Fersht (and other groups working on other enzymes) report, from time to time, mutants that have larger values for k_{cat} than the native enzyme. These are potentially interesting to study. It may be that larger k_{cat} values are obtained only upon the sacrifice of other selectable properties. Alternatively, it may indicate limits on the degree to which kinetic behavior in enzymes is optimized. This topic is discussed at length elsewhere.[101]

VII. CONCLUSIONS

We can conclude that behavioral surfaces relating behavior to structure are "rough" in only a few dimensions around individual points and are "smooth" in most dimensions. Even in those cases where a structural alteration has a large behavioral impact, it appears as if the impact can generally be suppressed by mutations at second sites. Thus, regions of the behavioral surface are not isolated from other regions by proteins with unacceptable behavior. Nevertheless, by making alterations along the "rough" dimensions, kinetics, stability, substrate specificity, stereospecificity, and biological activity can all be changed by changing just a few amino acids in the protein's sequence. This means that conservation principles necessary for historical models to explain behavioral similarities in divergent proteins are generally not plausible.

Furthermore, different behaviors appear to be able to vary independently of others. This suggests that few nonfunctional behaviors will be constrained from drifting because they are structurally coupled to other, functional behaviors.

In contrast, survival surfaces relating structure to "survival value" appear to be rough in many dimensions, at least for many catalytic proteins. Natural selection appears to be able to select for structural differences in such enzymes if it has even the slightest impact on behavior. In others, selection pressures appear to be rather weak. The extent to which selection pressure is felt depends on the chemical magnitude of the perturbation, the chemical function of the biological macromolecule, the total impact on energy consumption in the organism, and the complexity of the organism. Finally, subtle variations in the environment can influence the survival surface enough to change the structure of the evolutionarily optimal protein. This includes temperature, pH, and pressure and applies to microorganisms and eukaryotes alike.

Upon first inspection, the conclusions regarding the two surfaces are somewhat contradictory. Behavioral surfaces appear to be well suited to favor functional models since the combination of rough and smooth dimensions provides for rapid variability in behavior while permitting access to all regions of the surface. However, the roughness of survival surfaces appears to strongly constrain drift, favoring historical models that imply that the behavior of proteins in the modern world has only been locally optimized and therefore primarily reflects historical accidents in the evolution of the protein.

The resolution of this apparent paradox lies in the fact that small changes in the environment cause "upheavals" in survival surfaces. Upheavals change the relative survival values of different proteins with different structures, effectively removing obstacles to divergence in survival surfaces. Thus, rapid adaptation to an optimal protein with entirely satisfactory properties (even if it does not have a structure that is globally optimal) is expected. The evolution of the chorismate mutase discussed earlier may be an example of this.

Furthermore, we expect that modern proteins will already have searched for regions of the surface where such adaptation is possible. Proteins in regions of the surface where adaptation is not possible (e.g., because the survival surface is too rough) would be selected against because

they would not be readily adaptable following perturbations in the environment. Interestingly, the relatively small number of folded forms of proteins may reflect the selection of adaptable protein structures rather than of stable protein structures.

It remains unresolved as to whether the behavioral optimum displayed by proteins is truly global. The data for the evolved chorismate mutase and the evolved beta-galactosidase suggest that rapid adaptation produces catalysts that are 1 to 2 orders of magnitude poorer than those produced by long-term adaptation. Whether long-term adaptation is able to produce a truly global optimal catalyst cannot be said at this point. These topics are discussed at length elsewhere.[101-103]

REFERENCES

1. **Richardson, J. S.,** The anatomy and taxonomy of protein structure, *J. Adv. Prot. Chem.*, 34, 167, 1981.
2. **Douzou, P. and Petsko, G. A.,** Proteins at work: "stop-action" pictures at subzero temperatures, *Adv. Protein Chem.*, 86, 245, 1984.
3. **Barman, T. E.,** *Enzyme Handbook*, Springer -Verlag, Berlin, 1969.
4. **Albery, W. J. and Knowles, J. R.,** Evolution of enzyme function and the deveopment of catalytic efficiency, *Biochemistry*, 15, 5631, 1976.
5. **Hanson, K. R. and Rose, I. A.,** Interpretations of enzyme reaction stereospecificity, *Acc. Chem. Res.*, 8, 1, 1975.
6. **Hermes, J. D., Tipton, P. A., Fisher, M. A., O'Leary, M. H., Morrison, J. F., and Cleland, W. W.,** Mechanisms of enzymatic and acid-catalyzed decarboxylations of prephenate, *Biochemistry*, 23, 6263, 1984.
7. **Frauenfelder, H. and Wolynes, P. G.,** Rate theories and puzzles of hemeprotein kinetics, *Science*, 229, 337, 1985.
8. **Karplus, M. and McCammon, A. J. A.,** Dynamics of protein: elements and function, *Annu. Rev. Biochem.*, 52, 263, 1983.
9. **Goodman, M.,** *Macromolecular Sequences in Systematic and Evolutionary Biology*, Plenum Press, New York, 1982.
10. **Nei, M. and Koehn, R. K., Eds.,** *Evolution of Genes and Proteins*, Sinauer, Sunderland, MA, 1983.
11. **Frey, P. A.,** Stereochemistry of enzymatic reactions of phosphates, *Tetrahedron*, 38, 1541, 1982.
12. **Lawson, A. and Quinn, A. G.,** Studies on amino acid decarboxylases in *Escherichia coli*, *Biochem. J.*, 105, 483, 1967.
13. **Tran, V. T. and Snyder, S. H.,** Histidine decarboxylase. Purification from fetal rat liver, immunologic properties, and histochemical localization in brain and stomach, *J. Biol. Chem.*, 256, 680, 1981.
14. **Recsei, P. A. and Snell, E. E.,** Histidine decarboxylase of lactobacillus 30a.VI. Mechanism of action and kinetic properties, *Biochemistry*, 9, 1492, 1970.
15. **Benner, S. A., Nambiar, K. P., and Chambers, G. K.,** A stereochemical inperative in dehydrogenases: new data and criteria for evaluating function-based theories in bioorganic chemistry, *J. Am. Chem. Soc.*, 107, 5513, 1985.
16. **Rozzell, J. D., Jr. and Benner, S. A.,** Stereochemical imperative in enzymic decarboxylations. Stereochemical course of the decarboxylation catalyzed by acetoacetate decarboxylase, *J. Am. Chem. Soc.*, 106, 4937, 1984.
17. **Gutfreund, H.,** Kinetic analysis of the properties and reactions of enzymes, *Prog. Biophys. Mol. Biol.*, 29, 161, 1975.
18. **Ellington, A. D.,** Evolutionary Optimization of Isozymes, Ph.D. thesis, Harvard University, Cambridge, MA, 1988.
19. **Beintema, J. J., Fitch, W. M., and Carsana, A.,** Molecular evolution of pancreatic-type ribonucleases, *Mol. Biol. Evol.*, 3, 262, 1986.
20. **Dickerson, R. E.,** Structural conservatism in proteins over three billion years: cytochrome with a touch of collagen, in *Biomolecular Structure, Conformation, Function, and Evolution*, Pergamon Press, Elmsford, NY, 1981, 227.
21. **Kreitman, M.,** Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*, *Nature*, 304, 412, 1983.
22. **Ramshaw, J. A. M., Coyne, J. A., and Lewontin, R. C.,** The sensitivity of gel electrophoresis as a detector of genetic variation, *Genetics*, 93, 1019, 1979.
23. **Visser, C. M. and Kellogg, R. M.,** Biotin. Its place in evolution, *J. Mol. Evol.*, 11, 171, 1978.

24. **Oppenheimer, N. J., Singh, M., Sweeley, C. C., Sung, S.-J., and Srere, P. A.,** The configuration and location of the ribosidic linkage in the prosthetic group of citrate lyase *(Klebsiella aerogenes)*, *J. Biol. Chem.*, 254, 1000, 1979.
25. **Hochachka, P. W. and Somero, G. N.,** *Biochemical Adaptation*, Princeton University Press, Princeton, NJ, 1984.
26. **Wu, T. T., Fitch, W. M., and Margoliash, E.,** The information content of protein amino acid sequences, *Annu. Rev. Biochem.*, 43, 539, 1974.
27. **Perutz, M. F.,** Species adaptation in a protein molécule, *Adv. Prot. Chem.*, 36, 213, 1984.
28. **Lewontin, R. C.,** Adaptation, *Sci. Am.*, 239, 156, 1979.
29. **Dixon, M. and Webb, E. C.,** *Enzymes*, Longmans, London, 1958, 563.
30. **Kimura, M.,** The neutral theory as a basis for understanding the mechanism of evolution and variation at the molecular level, in *Molecular Evolution, Protein Polymorphism, and the Neutral Theory*, Springer-Verlag, Berlin, 1982.
31. **Kimura, M.,** Evolutionary rate at the molecular level, *Nature*, 217, 624, 1968.
32. **Jukes, T. H. and Holmquist, R.,** Evolutionary clock: nonconstancy of rate in different species, *Science*, 177, 530, 1972.
33. **Rose, I. A.,** Enzymology of proton abstraction and transfer reactions, *Enzymes*, 2, 281, 1980.
34. **Grimshaw, C. E., Sogo, S. G., Copley, S. D., and Knowles, J. R.,** Synthesis of stereoselectively labeled [9-^{2}H,^{3}H]chorismate and the stereochemical course of 5-enolpyruvoylshikimate-3-phosphate synthtase, *J. Am. Chem. Soc.*, 106, 2699, 1984.
35. **Schwab, J., Klassen, J. B., and Habib, A.,** Stereochemical course of the hydration reaction catalyxed by beta-hydroxydecanoylthioester dehydrase, *JCS Chem. Commun.*, p.357, 1986.
36. **Zydowsky, T. M., Courtney, L. F., Kobayashi, F. K., Shimizy, H., Yuen, L. D., Matthews, R. G., Benkovic, S. J., and Floss, H. G.,** Stereochemical analysis of the methyl transfer catalyzed by cobalamin-dependent methioninesynthase from *Escherichia coli* B, *J. Am. Chem. Soc.*, 108, 3152, 1986.
37. **Ramalingam, K., Lee, K.-M., Woodard, R. W., Bleecker, A. B., and Kende, H.,** Stereochemical course of the reaction catalyzed by the pyridoxal phosphate-dependent enzyme 1-aminocyclopropane-1-carboxylate synthase, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 7820, 1985.
38. **Wiggert, B. and Villee, C. A.,** Malic dehydrogenases in sea urchin eggs, *Science*, 138, 508, 1962.
39. **Monder, C. and White, A.,** The 21-hydroxysteroid dehydrogenases of liver. A nicotinamide adenine dinucleotide phosphate dehydrogenase and two nicotinamide adenine dinucleotide dehydrogenases, *J. Biol. Chem.*, 71, 1965.
40. **Nambiar, K. P., Stauffer, D. M., Kolodziej, P. A., and Benner, S. A.,** A mechanistic basis for the stereoselectivity of enzymatic transfer of hydrogen from nicotinamide cofactors, *J. Am. Chem. Soc.*, 105, 5886, 1983.
41. **You, K.-S.,** Stereospecificity for nicotinamide nucleotides in enzymatic and chemical hydride transfer reactions, *CRC Crit. Rev. Biochem.*, 17, 313, 1985.
42. **Chung, H., Harvey, R. G., Armstrong, R. N., and Jarabak, J.,** Polycyclic aromatic hydrocarbon quinones and glutathione thioethers as substrates and inhibitors of the human placental NADP-linked 15-hydroxyprostaglandin dehydrogenase, *J. Biol. Chem.*, 262, 12448, 1987.
43. **Matsuura, K., Hara, A., Nakayama, T., Nakagawa, M., and Sawada, H.,** Purification and properties of two multiple forms of dihydrodiol dehydrogenase from guinea-pig testis, *Biochim. Biophys. Acta*, 912, 270, 1987.
44. **Furuebisu, M., Deguchi, S., and Okuda, K.,** Identification of cortisone 5 beta-reductase as delta-3-ketosteroid 5 beta-reductase, *Biochim. Biophys. Acta*, 912, 110, 1987.
45. **Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllyla, R., Huhtala, M.-L., Koivu, J., and Kivirikko, K. I.,** Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene, *EMBO J.*, 6, 643, 1987.
46. **Boutin, J. A.,** Camphoroquinone reduction: another reaction catalyzed by rat liver cytosol 3 alpha-hydroxysteroid dehydrogenase, *Biochim. Biophys. Acta*, 870, 463, 1986.
47. **Chapman, S. K. and Maren, T. H.,** A search for the function of human carbonic anhydrase B, *Biochim. Biophys. Acta*, 527, 272, 1978.
48. **Bachmann, B. I. and Low, K. B.,** Linkage map of *Escherichia coli* K-12, Edition 6, *Microbiol. Rev.*, 44, 1, 1980.
49. **Boyer, J., Baron, D. N., and Talalay, P.,** Purification and properties of a 3 alpha-hydroxysteroid dehydrogenase from *Pseudomonas testosteroni*, *Biochemistry*, 4, 1825, 1965.
50. **Asano, Y., Nakazawa, A., Endo, K., Hibino, Y., Ohmori, M., Numao, N., and Kondo, K.,** Phenylalanine dehydrogenase of *Bacillus badius*, *Eur. J. Biochem.*, 168, 153, 1987.
51. **Kowalczykowski, S. C., Bear, D. G., and von Hippel, P. H.,** Single-stranded DNA binding proteins, in *The Enzymes*, Vol. 14, 3rd ed., Boyer, P. D., Ed., Academic Press, Orlando, FL, 1981, 373.
52. **Matthews, C. K., Kutter, E. M., Mosig, G., and Berget, P. B.,** in *Bacteriophage T4*, Williams, K. R. and Konigsberg, W. H., Eds., American Society for Microbiology, Washington, D.C., 1983, 8.

53. **Karpel, R. L.**, *Gene Amplification and Analysis*, Vol. 2, Chirikjian, J. G. and Papas, T. S., Eds., Elsevier/North-Holland, NY, 509, 2.
54. **Patel, G. L. and Thompson, P. E.,** Immunoreactive helix-destabilizing protein localized in transcriptionally active regions of *Drosophila* polytene chromosomes, *Proc. Natl. Acad. Sci. U.S.A.*, 77, 6749, 1980.
55. **Williams, K. R., Redigari, S., and Patel, G. L.,** Identification of a nucleic acid helix destabilizing protein from rat liver as lactate dehydrogenase-5, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 5260, 1985.
56. **Grosse, F., Nasheuer, H.-P., Scholtissek, S., and Schomburg, U.,** Lactate dehydrogenase and glyceraldehyde-phosphate dehydrogenase are single-stranded DNA-binding proteins that affect the DNA-polymerase-alpha-primase complex, *Eur. J. Biochem.*, 160, 459, 1986.
57. **Creighton, D. J. and Rose, I. A.,** Studies on the mechanism and stereochemical properties of the oxalacetate decarboxylase activity of pyruvate kinase, *J. Biol. Chem.*, 251, 61, 1976.
58. **Johnson, J. D., Creighton, D. J., and Lambert, M. R.,** Stereochemistry and function of oxalacetate keto-enol tautomerase, *J. Biol. Chem.*, 261, 4535, 1986.
59. **Blasi, F., Fragomele, R., and Covelli, I.,** Enzymic pathway for thyroxine synthesis through p-hydroxy-3,5-diiodophenylpyruvic acid, *Endocrinology*, 85, 542, 1969.
60. **Graves, J. D. and Wong, J. H.,** Alpha glucan phosphorylases, *Enzymes*, 7, 435, 1972.
61. **Stormer, F. C. and Umbarger, H. E.,** The requirement for flavine adenine dinucleotide in the formation of acetolactate by *Salmonella typhimurium* extracts, *Biochem. Biophys. Res. Commun.*, 17, 587, 1964.
62. **Gupta, N. K. and Vennesland, B.,** Glyoxylate carboligase of *Escherichia coli:* some properties of the enzyme, *Arch. Biochem. Biophys.*, 113, 255, 1966.
63. **Gupta, N. K. and Vennesland, B.,** Glyoxylate carboligase of *Escherichia coli*: a flavoprotein, *J. Biol. Chem.*, 239, 3787, 1984.
64. **Welch, G. R., Cole, K. W., and Gaertner, F. H.,** Chorismate synthase of *Neurospora crassa:* a flavoprotein, *Arch. Biochem. Biophys.*, 165, 505, 1974.
65. **Welch, G. R., Cole, K. W., and Gaertner, F. H.,** Chorismate synthase of *Neurospora crassa:* a flavoprotein, *Arch. Biochem. Biophys.*, 165, 505, 1974.
66. **Dutler, H. and Ambar, A.,** Ligand sphere transitions: a new concept in zinc-enzyme catalysis, in *The Coordination Chemistry of Metalloenzymes*, Bertini, I., Drago, R.S., and Luchinat, C., Eds., Reidel, Dordrecht, The Netherlands, 1983.
67. **Forte-McRobbie, C. M. and Pietruszko, R.,** Purification and characterization of human liver "high Km" aldehyde dehydrogenase and its identification as glutamic gamma-semialdehyde dehydrogenase, *J. Biol. Chem.*, 261, 2154, 1986.
68. **Harding, H. W., Jr., Pyeritz, E. A., Copeland, E. S., and White, H.B., III,** Role of glycerol 3-phosphate dehydrogenase in glyceride metabolism.Effect of diet on enzyme activities in chicken liver, *Biochem. J.*, 146, 223, 1975.
69. **Winberg, J. O., Thatcher, D. R., and McKinley-McKee, J. S.,** Alcohol dehydrogenase from the fruitfly *Drosophila melanogaster*. Substrate specificity of the alleloenzyme AdhS and AdhUF, *Biochim. Biophys. Acta*, 704, 7, 1983.
70. **Heinstra, P. W. H., Eisses, K. Th., Schoonen, W. G. E. J., Aben, W., de Winter, A. J., van der Horst, D. J., Marrewijk, W. J. A., Beenakkers, A. M., Scharloo, W., and Thorig, G. E. W.,** A dual function of alcohol dehydrogenase in *Drosophilia, Genetica*, 60, 129, 1983.
71. **McDonald, J. F. and Avise, J. C.,** Evidence for the adaptive significance of enzyme activity levels; interspecific variation in alpha-GPDH and ADH in *Drosophila, Biochem. Genet.*, 14, 347, 1976.
72. **Hougouto, N., Lietaert, M. C., Libion-Mannaert, M., Feytmans, E., and Elens, A.,** Oviposition-site preference and ADH activity in *Drosophila melanogaster, Genetica*, 58, 121, 1982.
73. **McDonald, J. F., Chambers, G. K., David, J., and Ayala, F. J.,** Adaptive response due to changes in gene regulation: a study with *Drosophila, Proc. Natl. Acad. Sci. U.S.A.*, 74, 4562, 1977.
74. **Grell, E. H., Jacobson, K. B., and Murphy, J. H.,** Alterations of genetic material for analysis of alcohol dehydrogenase isozymes of *Drosophila melanogaster, Ann. N.Y. Acad. Sci.*, 151, 441, 1968.
75. **David, J. R., van Herrewege, J., de Scheemaker-Louis, M., and Pla, E.,** *Drosophila* alcohol dehydrogenase; detoxification of isopropanol and acetone, substances not used in energy metabolism, *Heredity*, 47, 263, 1981.
76. **Vigue, C. and Sofer, W.,** Chemical selection of mutants that affect ADH activity in *Drosophila*. III. Effects of ethanol, *Biochem. Genet.*, 14, 127, 1976.
77. **David, J. and Bocquet, C.,** Compared toxicities of different alcohols for two *Drosophila* sibling species: *D. melanogaster* and *D. simulans, Comp. Biochem. Physiol.*, 54C, 71, 1976.
78. **You, K.-S.,** Benner's model: a successful prediction, *Chem. Eng. News*, 64, 3, 1986.
79. **Schneider-Bernloehr, H., Adolph, H.-W., and Zeppezauer, M.,** Coenzyme stereospecificity of alcohol/polyol dehydrogenases: conservation of protein types vs. functional constraints, *J. Am. Chem. Soc.*, 108, 5573, 1986.
80. **Stroud, R. M.,** Family of protein-cutting proteins, *Sci. Am.*, 231, 74, 1974.
81. **Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H.-J., Meyer, H., Keil, W., and Kiltz, H.-H.,** Structural adaptations of lactate dehydrogenase isozymes, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 2677, 1977.

82. **Doolittle, R. F.,** Similar amino acid sequences: chance or common ancestry?, *Science*, 214, 149, 1981.
83. **Feng, D. F., Johnson, M. S., and Doolittle, R. F.,** Aligning amino acid sequences: comparison of commonly used methods, *J. Mol. Evol.*, 21, 112, 1985.
84. **McLachlan, A. D. and Boswell, D. R.,** Confidence limits for homology in protein or gene sequences. The c-myc oncogene and adenovirus E1 a protein, *J. Mol. Biol.*, 185, 39, 1985.
85. **Wilson, A.C., Carlson, S. S., and White, T. J.,** Biochemical evolution, *Annu. Rev. Biochem.*, 46, 573, 1977.
86. **Lee, Y. M., Friedman, D. J., and Ayala, F. J.,** Superoxide dismutase; an evolutionary puzzle, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 824, 1985.
87. **Vawater, L. and Brown, W. M.,** Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock, *Science*, 234, 194, 1986.
88. **Martin, J. P., Jr. and Fridovich, I.,** Evidence for a natural gene transfer from the ponyfish to its bioluminescent bacterial symbiont *Photobacter leiognathi*, *J. Biol. Chem.*, 256, 6080, 1981.
89. **Gentry, G. A.,** Locating a nucleotide-binding site in the thymidine kinase of vaccinia virus and of herpes simplex virus by scoring triply aligned protein sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 6815, 1985.
90. **Murata, M., Richardson, J. S., and Sussman, J. L.,** Simultaneous comparison of three protein sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 3073, 1985.
91. **Schulz, G. E.,** Gene duplication in glutathione reductase, *J. Mol. Biol.*, 138, 335, 1980.
92. **Rossmann, M. G., Moras, D., and Olsen, K. W.,** Chemical and biological evolution of a nucleotide-binding protein, *Nature*, 250, 194, 1974.
93. **Schulz, G. E. and Schirmer, R. H.,** Topological comparison of adenyl kinase with other proteins, *Nature*, 250, 142, 1974.
94. **Chothia, C. and Lesk, A. M.,** The relation between the divergence of sequence and structure in proteins, *EMBO J.*, 5, 823, 1986.
95. **Rossmann, M. G. and Argos, P.,** The taxonomy of protein structure, *J. Mol. Biol.*, 109, 99, 1977.
96. **Weaver, L. H., Gruetter, M. G., Remington, S. J., Gray, T. M., Isaacs, N. W., and Matthews, B. W.,** Comparison of goose-type, chicken-type, and phage-type lysozymes illustrates the changes that occur in both amino acid sequence and three-dimensional structure during evolution, *J. Mol. Evol.*, 21, 97, 1985.
97. **Braenden, C.-I.,** Relation between structure and function of alpha/beta-proteins, *Q. Rev. Biophys.*, 13, 317, 1980.
98. **Richardson, J. S., Richardson, D. C., and Thomas, K. A.,** Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit, *J. Mol. Biol.*, 102, 221, 1976.
99. **Jornvall, H., Persson, M., and Jeffrey, J.,** Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type, *Proc. Natl. Acad. Sci. U.S.A.*, 78, 4226, 1981.
100. **Benner, S. A. and Piccirilli, J. A.,** Functional and historical interpretations of enzymatic stereospecificity, *Top. Stereochem.*, in press.
101. **Benner, S. A.,** Evolutionary adaptation of the kinetic behavior of enzymes, *Chem. Rev.*, submitted.
102. **Benner, S. A.,** Reconstructing the evolution of proteins, in *Redesigning Life*, Springer-Verlag, Heidelberg, 1988, 155.
103. **Wright, S.,** *Evolution and the Genetics of Populations. IV. Variability within and among Natural Populations*, University of Chicago Press, Chicago, IL, 1978.
104. **Dykhuizen, D. E. and Hartl, D. L.,** Selection in chemostats, *Microbiol. Rev.*, 47, 150, 1983.
105. **Ackers, G. K. and Smith, F. R.,** Effects of site-specific amino acid modification on protein interactions and biological function, *Annu. Rev. Biochem.*, 54, 597, 1985.
106. **Eigen, M. and Schuster, P.,** The hypercycle, a principle of natural self-organization. C. The realistic hypercycle, *Naturwissenschaften*, 65, 341, 1978.
107. **Bachmann, B. J. and Low, K. B.,** Linkage map of *Escherichia coli* K-12, edition 6, *Microbiol. Rev.*, 44, 1, 1980.
108. **Hellinga, H. W. and Evans, P. R.,** Mutations in the active site of *Escherichia coli* phosphofructokinase, *Nature*, 327, 437, 1987.
109. **Liang, N., Pielak, G. J., Mauk, A. G., Smith, M., and Hoffman, B. M.,** Yeast cytochrome c with phenylalanine or tyrosine at position 87 transfers electrons to (zinc cytochrome c peroxidase)$^+$ at a rate ten thousand times that of the serine-87 or glycine-87 variants, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 1249, 1987.
110. **Willson, T. A. and Nagley, P.,** Amino acid substitutions in subunit 9 of the mitochondrial ATPase complex of *Saccharomyces cerevisiae*, *Eur. J. Biochem.*, 167, 291, 1987.
111. **Yutani, K., Ogasabara, K., Tsujita, T., Kanemoto, K., Matsumoto, M., Tanaka, S., Miyashita, T., Matsushiro, A., Sugino, Y., and Miles, E. W.,** Tryptophan synthase alpha subunit glutamic acid 49 is essential for activity, *J. Biol. Chem.*, 262, 13429, 1987.
112. **Ho, C. K. and Fersht, A. R.,** Internal thermodynamics of position 51 mutants and natural variants of tyrosyl-tRNA synthetase, *Biochemistry*, 25, 1891, 1986.

113. **Khamis, M. I., Casas-Finet, J. R., and Maki, A. H.,** Investigation of the role of individual tryptophan residues in the binding of *Escherichia coli* single-stranded DNA binding protein to single-stranded polynucleotides, *J. Biol. Chem.*, 262, 10938, 1987.
114. **Toney, M. D. and Kirsch, J. F.,** Tyrosine 70 increases the coenzyme affinity of aspartate aminotransferase, *J. Biol. Chem.*, 262, 12403, 1987.
115. **Betz, J. L.,** Affinities of tight-binding lactose repressors for wild-type and pseudo-operators, *J. Mol. Biol.*, 195, 495, 1987.
116. **Kawasaki, H., Bauerle, R., Zon, G., Ahmed, S. A., and Miles, E. W.,** Site-specific mutagenesis of the alpha subunit of tryptophan synthase from *Salmonella typhimurium*, *J. Biol. Chem.*, 262, 10678, 1987.
117. **Cherfils, J., Vachette, P., Tauc, P., and Janin, J.,** The pAR5 mutation and the allosteric mechanism of *Escherichia coli* aspartate carbamoyltransferase, *EMBO J.*, 6, 2843, 1987.
118. **Clarke, A. R., Wigley, D. B., Barstow, D. A., Chia, W. N., Atkinson, T., and Holbrook, J. J.,** A single amino acid substitution deregulates a bacterial lactate dehydrogenase and stabilizes its tetrameric structure, *Biochim. Biophys. Acta*, 913, 72, 1987.
119. **Lau, F. K. T. and Fersht, A. R.,** Conversion of allosteric inhibition to activation in phosphofructokinase by protein engineering, *Nature*, 326, 1987.
120. **Dequard-Chablat, M.,** Different alterations of the ribosomal protein S7 lead to opposite effects on translational fidelity in the fungus *Podospora anserina*, *J. Biol. Chem.*, 261, 4117, 1986.
121. **Biswas, T. K. and Getz, G. S.,** A critical base in the yeast mitochondrial nonanucleotide promoter, *J. Biol. Chem.*, 261, 3928, 1986.
122. **Pietruszko, R.,** Alcohol dehydrogenase from horse liver, steroid-active SS isozyme, *Methods Enzymol.*, 89, 428, 1982.
123. **Glasfeld, A.,** Evolution of Dehydrogenase Stereospecificity, Ph.D. thesis, Harvard University, Cambridge, MA, 1983.
124. **Stalker, D. M., Hiatt, W. R., and Comai, L.,** A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase convers resistance to the herbicide glyphosate, *J. Biol. Chem.*, 260, 4724, 1985.
125. **Etzold, T., Fritz, C. C., Schell, J., and Schreier, P. H.,** A point mutation in the chloroplast 16 S rRNA gene of a streptomycin resistant *Nicotiana tabacum*, *FEBS Lett.*, 219, 343, 1987.
126. **Both, G. W., Shi, C. H., and Kilbourne E. D.,** Hemagglutinin of swine influenza virus: a single amino acid change pleiotropically affects viral antigenicity and replication, *Proc. Natl. Acad. Sci. U.S.A.*, 80, 6996, 1983.
127. **Ried, G. and Henning, U.,** A unique amino acid substitution in the outer membrane protein OmpA causes conjugation deficiency in *Escherichia coli* K-12, *FEBS Lett.*, 223, 387, 1987.
128. **Strathern, J. N., Jones, E. W., and Broach, J. R.,** *The Molecular Biology of the Yeast Saccaromyces: Metabolism and Gene Expression*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, Appendix 1.
129. **Weber, H., Valenzuela, D., Lujber, G., Gubler, M., and Weissmann, C.,** Single amino acid changes that render human IFN-alpha2 biologically active on mouse cells, *EMBO J.*, 6, 591, 1987.
130. **Miller, J. H.,** Genetic studies of the lac repressor. XI. On aspects of lac repressor structure suggested by genetic experiments, *J. Mol. Biol.*, 131, 249, 1979.
131. **Shortle, D. and Lin, B.,** Genetic analysis of staphylococcal nuclease: identification of three intragenic 'global' suppressors of nuclease-minus mutations, *Genetics*, 110, 539, 1985.
132. **Serpersu, E. H., Shortle, D., and Mildvan, A. S.,** Kinetic and magnetic resonance studies of effects of genetic substitution of a Ca(2+)-liganding amino acid in staphylococcal nuclease, *Biochemistry*, 25, 68, 1986.
133. **Cone, J. L., Cusumano, C. L., Taniuchi, H., and Anfinsen, C. B.,** Staphylococcal nuclease (Foggi strain): the amino acid sequence, *J. Biol. Chem.*, 246, 3103, 1971,
134. **Knowles, J. R.,** *Evolution of Catalytic Function*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988, 52.
135. **Strauss, D., Raines, R., Kawashima, E., Knowles, J. R., and Gilbert, W.,** Active site of triosephosphate isomerase: in vitro mutagenesis and characterization of an altered enzyme, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 2272, 1985.
136. **Ryan, J. P., Duncan, M. C., Bankaitis, V. A., and Bassford, P. J., Jr.,** Intragenic reversion mutations that improve export of maltose-binding protein in *Escherichia coli* malE signal sequence mutants, *J. Biol. Chem.*, 261, 3389, 1986.
137. **Stock, J., Borczuk, A., Chiou, F., and Burchenal, J. E. B.,** Compensatory mutations in receptor function: a reevaluation of the role of methylation in bacterial chemotaxis, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 8364, 1985.
138. **Sigal, I. S., Harwood, B. G., and Arentzen, R.,** Thiol-beta-lactamase: replacement of the active-site serine of RTEM beta-lactamase by a cysteine residue, *Proc. Natl. Acad. Sci. U.S.A.*, 79, 7157, 1982.
139. **Dalbadie-McFarland, G., Neitzel, J. J., and Richards, J. H.,** Active-site mutants of beta-lactamase: use of an inactive double mutant to study requirements for catalysis, *Biochemistry*, 25, 332, 1986.
140. **Juillerat, M. A. and Taniuchi, H.,** A study on the role of evolutionarily invariant leucine 32 of cytochrome c, *J. Biol. Chem.*, 261, 2697, 1986.

141. **Valenzuela, D., Weber, H., and Weissmann, C.,** Is sequence conservation in interferons due to selection for functional proteins?, *Nature*, 313, 698, 1985.
142. **Zimmermann, T., Kulla, H. G., and Leisinger, T.,** Evolution of new bacterial enzyme activities during adaptation to azo dyes, *Experientia*, 39, 1429, 1983.
143. **Campbell, H. H., Lengyel, J. A., and Langridge, J.,** Evolution of a second gene for beta-galactosidase in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 1841, 1973.
144. **Boronat, A. and Aguilar, J.,** Experimental evolution of propanediol oxidoreductase in *Escherichia coli*. Comparative analysis of the wild-type and mutant enzymes, *Biochim. Biophys. Acta*, 672, 98, 1981.
145. **Burleigh, B. D., Jr., Rigby, P. W. J., and Hartley, B. S.,** A comparison of wild-type and mutant ribitol dehydrogenases from *Klebsiella aerogenes*, *Biochem. J.*, 143, 341, 1974.
146. **Betz, J. L., Brown, P. R., Smyth, M. J., and Clarke, P. H.,** Evolution in action, *Nature*, 247, 261, 1974.
147. **Hall, A. and Knowles, J. R.,** Directed selective pressure on a beta lactamase to analyze molecular changes involved in development of enzyme function, *Nature*, 264, 803, 1976.
148. **Wills, C.,** Production of yeast alcohol dehydrogenase isoenzymes by selection, *Nature*, 261, 26, 1976.
149. **Brooker, R. J. and Wilson, T. H.,** Isolation and nucleotide sequencing of lactose carrier mutants that transport maltose, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 3959, 1985.
150. **Burton, J. and Sinnott, M. L.,** Catalytic consequences of experimental evolution. I, *J. Chem. Soc. Perkin Trans. 2*, p. 359, 1983.
151. **Li, B. F L., Osborne, S., and Sinnott, M. L.,** Catalytic consequences of experimental evolution. II. Rate-limiting degalactosylation in the hydrolysis of aryl beta-D-galactopyranosides by the experimental evolvants ebg(a) and egb(b), *J. Chem. Soc. Perkin Trans. 2*, p.365, 1983.
152. **Murray, M., Osborne, S., and Sinnott, M. L.,** Catalytic consequences of experimental evolution. III, *J. Chem. Soc. Perkin Trans. 2*, p.1595, 1983.
153. **Hall, B. G., Murray, M., Osborne, S., and Sinnott, M. L.,** Catalytic consequences of experimental evolution. III. Construction of reaction profiles for hydrolysis of lactose by ebg(o), egb(a), and egb(b) enzymes via measurement of the enzyme-catalysed exchange of p-[1-(18)0] by c(13) NMR spectroscopy, *J. Chem. Soc. Perkin Trans. 2*, p.1595, 1983.
154. **Hall, B. G., Yokoyama, S., and Calhoun, D. H.,** Role of cryptic genes in microbial evolution, *Mol. Biol. Evol.*, 1, 109, 1984.
155. **Llewellyn, D. J., Daday, A., and Smith, G. D.,** Evidence for an artificially evolved bifunctional 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase-chorismate mutase in *Bacillus subtilis*, *J. Biol. Chem.*, 255, 2077, 1980.
156. **Andrews, P. R., Smith, G. D., and Young, I. G.,** Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate, *Biochemistry*, 12, 3492, 1973.
157. **Gorisch, H.,** On the mechanism of the chorismate mutase reaction, *Biochemistry*, 17, 3700, 1978.
158. **Haniu, M., Izanagi, T., Miller, P., Lee, T. D., and Shively, J. E.,** Complete amino acid sequence of NADPH-cytochrome P-450 reductase from porcine hepatic microsomes, *Biochemistry*, 25, 7906, 1986.
159. **Nyunoya, H., Broglie, K.E., and Lusty, C.J.,** The gene coding for carbamoyl-phosphate synthetase I was formed by fusion of an ancestral glutaminase gene and a synthetase gene, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 2244, 1985.
160. **Belfaiza, J., Parsot, C., Martel, A., Bouthier de la Tour, C., Margarita, D., Cohen, G. N., and Saint-Girons, I.,** Evolution in biosynthetic pathways: two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 867, 1986.
161. **Yeh, W.-K., Fletcher, P., and Ornston, L. N.,** Repetitions in the NH2-terminal amino acid sequence of beta-ketoadipate enol-lactone hydrolase from *Pseudomonas putida* (revised), *J. Biol. Chem.*, 255, 6342, 1980.
162. **Sante Capasso, F. G., Mattia, C. A., Mazzarella, L., and Zagari, A.,** Refinement of the structure of bovine seminal ribonuclease, *Biopolymers*, 22, 327, 1983.
163. **Piccoli, R. and D'Alessio, G.,** Relationships between nonhyperbolic kinetics and dimeric structure in ribonucleases, *J. Biol. Chem.*, 259, 693, 1984.
164. **Taniguchi, T. and Libonati, M.,** Action of ribonuclease BS-1 on a DNA-RNA hybrid, *Biochem. Biophys. Res. Commun.*, 58, 280, 1974.
165. **Vescia, S. and Tramontano, D.,** Antitumoral action of bovine seminal ribonuclease, *Mol. Cell. Biochem.*, 36, 125, 1981.
166. **Levy, C. C. and Karpetsky, T. P.,** Human ribonucleases, in *Enzymes as Drugs*, Holcenberg, J. C. and Roberts, J., Eds., Wiley-Interscience, New York, 1981.
167. **Palmer, K. A., Scheraga, H. A., Riordan, J. F., and Vallee, B. L.,** A preliminary three-dimensional structure of angiogenin, *Proc Natl. Acad. Sci. U.S.A.*, 83, 1965, 1986.
168. **Shapiro, R., Riordan, J. F., and Vallee, B. L.,** Characteristic ribonucleolytic activity of human angiogenin, *Biochemistry*, 25, 3527, 1986.
169. **Beintema, J. J., Broos, J., Meulenberg, J., and Schueller, C.,** The amino acid sequence of snapping turtle *(Chelydra serpentina)* ribonuclease, *Eur. J. Biochem.*, 153, 305, 1985.

170. **Young, E. T. and Pilgrim, D.,** Isolation and DNA sequence of ADH3, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in *Saccharomyces cerivisiae*, *Mol. Cell. Biol.*, 5, 3024, 1985.
171. **Joernvall, H.,** Differences between alcohol dehydrogenases, *Eur. J. Biochem.*, 72, 443, 1977.
172. **Joernvall, H., von Bahr-Lindstroem, H., Jany, K.-D., Ulmer, W., and Froeschle, M.,** Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases, *FEBS Lett.*, 165, 190, 1984.
173. **Petsko, G.,** personal communication.
174. **Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R., and Rutter, W. J.,** Redesigning trypsin: alteration of substrate specificity, *Science*, 228, 291, 1985.
175. **Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., and Wells, J. A.,** Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering, *Science*, 233, 619, 1986.
176. **Vennesland, B.,** Stereospecificity in biology, *Top. Current Chem.*, 48, 41, 1974.
177. **Hall, B. D.,** Mitochondria springs surprises, *Nature*, 282, 129, 1979.
178. **Ng, R., Domdey, H., Larson, G., Rossi, J. J., and Abelson, J.,** A test for intron function in the yeast actin gene, *Nature*, 314, 183, 1985.
179. **Jacq, C., Miller, J. R., and Brownlee, G. G.,** A pseudogene structure in 5s DNA of *Xenopus laevis*, *Cell*, 12, 109, 1977.
180. **Vanin, E. F.,** Processed pseudogenes.Characteristics and evolution, *Biochim. Biophys. Acta*, 782, 231, 1984.
181. **Miyata, T. and Hayashida, H.,** Extraordinarily high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons, *Proc. Natl. Acad. Sci. U.S.A.*, 78, 5739, 1981.
182. **Li, W.-H., Wu C.-I., and Luo, C.-C.,** Nonrandomness of point mutations in nucleotide substitutions in pseudogenes and its evolutionary implications, *J. Mol. Evol.*, 21, 58, 1984.
183. **Li, W.-H., Gojobori, T., and Nei, M.,** Pseudogenes as a paradigm of neutral evolution, *Nature*, 292, 237, 1981.
184. **Gitlin, D. and Gitlin, J. D.,** *The Plasma Proteins*, Vol. 2, Putman, F. W., Ed., Academic Press, New York, 1975.
185. **Li, W.-H. and Gojobori, T.,** Rapid evolution of goat and sheep globin genes following gene duplication, *Mol. Biol. Evol.*, 1, 94, 1983.
186. **Kafatos, F. C., Efstradiatis, A., Forget, B. G., and Weissman, S. M.,** Molecular evolution of human and rabbit beta-globin mRNAs, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 5618, 1977.
187. **Perler, F., Efstradiatis, A., Lomedico, P., Gilbert, W., Kolander, R., and Dodgson, J.,** The evolution of genes: the chicken preproinsulin gene, *Cell*, 20, 556, 1980.
188. **Crabtree, G. R., Comeau, C. M., Fowlkes, D. M., Fornace, A. J., Jr., Malley, J. D., and Kant, J. A.,** Evolution and structure of fibrinogen genes, *J. Mol. Biol.*, 185, 1, 1985.
189. **Minghetti, P. P., Law, S. W., and Dugaiczyk, A.,** The rate of molecular evolution of alpha-fetoprotein approaches that of pseudogenes, *Mol. Biol. Evol.*, 2, 347, 1985.
190. **Yang, Y.-C., Ciarletta, A. B., Temple, P. A., Chung, M. P., Kovacic, S., Witek-Giannotti, J. S., Leary, A. C., Kriz, R., Donahue, R. E., Wong, G. G., and Clark, S. C.,** Human IL-3 (multi-CSF): identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3, *Cell*, 47, 3, 1986.
191. **Hill, R. E. and Hastie, A.,** Accelerated evolution in the reactive centre regions of serine protease inhibitors, *Nature*, 326, 96, 1987.
192. **Brown, A. L.,** Positively Darwinian molecules, *Nature*, 326, 12, 1987.
193. **Dayoff, M. O.,** Atlas of protein sequence and structure, *Natl. Biomed. Res. Found.*, 4, 1, 1978.
194. **Lewin, R.,** "Computer genome" is full of junk DNA, *Science*, 232, 578, 1986.
195. **Diamond, J. M.,** Why do disused proteins become genetically lost or repressed?, *Nature*, 321, 565, 1986.
196. **Dykhuizen, D. E.,** Selection for tryptophan auxotrophs of *Escherichia coli* in glucose-limited chemostats as a test of the energy conservation hypothesis of evolution, *Evolution*, 32, 125, 1978.
197. **Koch, A. L.,** The protein burden of lac operon products, *J. Mol. Evol.*, 19, 455, 1983.
198. **Li, W.-H.,** Retention of cryptic genes in microbial population, *Mol. Biol. Evol.*, 1, 213, 1984.
199. **Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C.,** Nucleotide sequence heterogeneity of an RNA phage population, *Cell*, 13, 735, 1978.
200. **Sharp, P. M., Rogers, M. S., and McConnell, D. J.,** Selection pressures on codon usage in the complete genome of bacteriophage T7, *J. Mol. Evol.*, 21, 150, 1985.
201. **Eigen, M., Gardiner, W., Schuster, P., and Winkler-Oswatitsch, R.,** The origin of genetic information, *Sci. Am.*, 244, 78, 1981.
202. **Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R.,** Codon catalog usage is a genome strategy modulated for gene expressivity, *Nucleic Acids Res.*, 9, r43, 1981.
203. **Gouy, M. and Gautier, C.,** Codon usage in bacteria: correlation with gene expressivity, *Nucleic Acids Res.*, 10, 7056, 1982.
204. **Konigsberg, W. and Godson, G.N.,** Evidence for use of codons in the dnaG gene and other regulatory genes of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.*, 80, 687, 1983.

205. **Bennetzen, J. and Hall, B.,** Codon selection in yeast, *J. Biol. Chem.*, 257, 3026, 1982.
206. **Grosjean, H. and Fiers, W.,** Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes, *Gene*, 18, 199, 1982.
207. **Blaisdell, B. E.,** Choice of base at silent codon site 3 is not selectively neutral in eucaryotic structural genes: it maintains excess short runs of weak and strong hydrogen bonding bases, *J. Mol. Evol.*, 19, 226, 1983.
208. **Hamada, H., Petrino, M. G., and Kakunaga, T.,** A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes, *Proc. Natl. Acad. Sci. U.S.A.*, 79, 6465, 1982.
209. **Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T., and Schmid, C. W.,** Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences, *Nature*, 284, 372, 1980.
210. **Duncan, C., Biro, P. A., Choudary, P. V., Elder, J. T., Wang, R. R. C., Forget, B. G., deRiel, J. K., and Weissman, S. M.,** RNA polymerase III transcriptional units are interspersed among human non-α-globin genes, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 5095, 1979.
211. **Modiano, G., Battistuzzi, G., and Motulsky, A. G.,** Nonrandom patterns of codon usage and of nucleotide substitutions in human alpha- and beta-globin genes: an evolutionary strategy reducing the rate of mutations with drastic effects?, *Proc. Natl. Acad. Sci. U.S.A.*, 78, 1110, 1981.
212. **Lipman, D. J. and Wilbur, W. J.,** Contextual constraints on synonymous codon choice, *J. Mol. Biol.*, 163, 363, 1983.
213. **Lanave, C., Preparata, G., and Saccone, C.,** Mammalian genes as molecular clocks?, *J. Mol. Evol.*, 21, 346, 1985.
214. **Lanave, C., Preparata, G., Saccone, C., and Serio, G.,** A new method for calculating evolutionary substitution rates, *J. Mol. Evol.*, 20, 86, 1984.
215. **Perutz, M. F. and Lehmann, H.,** Molecular pathology of human haemoglobin, *Nature*, 219, 902, 1968.
216. **Weatherall, D. J. and Clegg, J. B.,** Molecular genetics of human hemoglobin, *Annu. Rev. Genet.*, 10, 157, 1976.
217. **Allison, A. C.,** Aspects of polymorphism in man, *Cold Spring Harbor Symp. Quant. Biol.*, 20, 239, 1955.
218. **Wills, C. and Joernvall, H.,** The two major isozymes of yeast alcohol dehydrogenase, *Eur. J. Biochem.*, 99, 323, 1979.
219. **Milkman, R.,** Electrophoretic variation in *Escherichia coli* from natural sources, *Science*, 182, 1024, 1973.
220. **Sawyer, S. A., Dykhuizen, D. E., and Hartl, D. L.,** Confidence interval for the number of selectively neutral amino acid polymorphisms, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 6225, 1987.
221. **Hubby, J. L. and Lewontin, R. C.,** *Drosophila, Genetics*, 54, 595, 1966.
222. **Harris, C.,** Genetics of man. Enzyme polymorphisms in man, *Proc. R. Soc. London B*, 164, 298, 1966.
223. **Oakeshott, J. G., Gibson, J. B., Anderson, P. R., Knibb, W. R., Anderson, D. G., and Chambers, G. K.,** Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents, *Evolution*, 36, 86, 1982.
224. **Cohn, V. H., Thompson, M. A., and Moore, G. P.,** Nucleotide sequence comparison of the Adh gene in three drosophilids, *J. Mol. Evol.*, 20, 31, 1984.
225. **Johnson, G. B.,** Enzyme polymorphism and metabolism, *Science*, 184, 28, 1974.
226. **Zours, E.,** Electrophoretic variation in allelozymes related to function or structure?, *Nature*, 254, 446, 1975.
227. **Place, A. R. and Powers, D. A.,** Genetic variation and relative catalytic efficiencies: lactate dehydrogenase B allozymes of *Fundulus heteroclitus*, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 2354, 1979.
228. **Borgman V. and Moon, T. W.,** A comparison of lactate dehydrogenases from an ectothermic and an endothermic animal, *Can. J. Biochem.*, 53, 998, 1975.
229. **Hennessey, J. P., Jr. and Siebenaller, J. F.,** Inactivation of NAD-dependent dehydrogenases from shallow- and deep-living fishes by hydrostatic pressure and proteolysis, *Biochim. Biophys. Acta*, 913, 285, 1987.
230. **Koehn, R. K.,** Esterase heterogeneity: dynamics of a polymorphism, *Science*, 163, 943, 1969.
231. **Merritt, R. B.,** Geographic distribution and enzymatic properties of lactate dehydrogenase allozymes in the fathead minnow, *Pimephales promelas*, *Am. Nat.*, 106, 173, 1972.
232. **Day, T. H., Hiller, P. C., and Clarke, B.,** Properties of genetically polymorphic isozymes of alcohol dehydrogenase in *Drosophila melanogaster*, *Biochem. Genet.*, 11, 141, 1974.
233. **Day, T. H. and Needham, L.,** Properties of alcohol dehydrogenase isozymes in a strain of *Drosophila melanogaster* homozygous for the Adh-slow allele, *Biochem. Genet.*, 11, 167, 1974.
234. **Fersht, A. R., Leatherbarrow, R. J., and Wells, T. C.,** Structure-activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships, *Biochemistry*, 26, 6030, 1987.
235. **Bradford, A. P., Aitken, A., Beg, F., Cook, K. G., and Yeaman, S. J.,** Amino acid sequence surrounding the lipoic acid cofactor of bovine kidney 2-oxoglutarate dehydrogenase complex, *FEBS Lett.*, 222, 211, 1987.
236. **Frantz, B. and Chakrabarty, A. M.,** Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 4460, 1987.
237. **Grenett, H. E., Ledley, F. D., Reed, L. L., and Woo, S. L. C.,** Full-length cDNA for rabbit tryptophan hydroxylase: functional domains and evolution of aromatic amino acid hydroxylases, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 5530, 1987.

238. **Sikorav, J.-L., Krejci, E., and Massoulie, J.,** cDNA sequences of *Torpedo marmorata* acetylcholinesterase: primary structure of the precursor of a catalytic subunit; existence of multiple 5′-untranslated regions, *EMBO J.*, 6, 1865, 1987.
239. **Schinina, M. E., Maffey, L., Barra, D., Bossa, F., Puget, K., and Michelson, A. M.,** The primary structure of iron superoxide dismutase from *Escherichia coli*, *FEBS Lett.*, 221, 87, 1987.
240. **Shanske, S., Sakoda, S., Hermodson, M. A., DiMauro, S., and Schon, E. A.,** Isolation of a cDNA encoding the muscle-specific subunit of human phosphoglycerate mutase, *J. Biol. Chem.*, 262, 14612, 1987.
241. **Parson, C.,** A common origin for enzymes involved in the terminal step of the threonine and tryptophan biosynthetic pathways, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 5207, 1987.
242. **Anton, I. A., Duncan, K., and Coggins, J. R.,** A eukaryotic repressor protein, the qa-1S gene product of *Neurospora crassa*, is homologous to part of the arom multifunctional enzyme, *J. Mol. Biol.*, 197, 367, 1987.
243. **Jornvall, H., Hoog, J.-O., Bahr-Lindstrom, von H., and Vallee, B. L.,** Mammalian alcohol dehydrogenases of separate classes: intermediates between different enzymes and intracless isozymes, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 2580, 1987.
244. **Volokita, M. and Somerville, C. R.,** The primary structure of spinach glycolate oxidase deduced from the DNA sequence of a cDNA clone, *J. Biol. Chem.*, 262, 15825, 1987.
245. **Pilkis, S. J., Lively, M. O., and El-Maghrabi, M. R.,** Active site sequence of hepatic fructose-2,6-bisphosphatase, *J. Biol. Chem.*, 262, 1987.
246. **Hammond, G. L., Smith, C. L., Goping, I. S., Underhill, D. A., Harley, M. J., Reventos, J., Musto, N. A., Gunsalus, G. L., and Bardin, C. W.,** Primary structure of human corticosteroid binding globulin, deduced from hepatic and pulmonary cDNAs, exhibits homology with serine protease inhibitors, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 5153, 1987.
247. **Glasfeld, A., Ellington, A. D., and Benner, S. A.,** unpublished.