Natural Selection, Protein Engineering, and the Last Riboorganism: Rational Model Building in Biochemistry

S.A. BENNER, R.K. ALLEMANN, A.D. ELLINGTON, L. GE, A. GLASFELD, G.F. LEANZ, T. KRAUCH, L.J. MACPHERSON, S. MORONEY, J.A. PICCIRILLI, AND E. WEINHOLD
Laboratory for Organic Chemistry, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland

The behavior of biological macromolecules can be interpreted both functionally and historically (Benner et al. 1985). Functional interpretations presume natural selection and require a distinction between macromolecular behaviors that are the products of selection and those that reflect neutral drift. Historical interpretations require distinctions between "primitive traits" present in a common ancestor and "derived" traits that arose more recently.

These distinctions are extremely difficult to make. Therefore, appropriate research strategy involves construction of formal models that can be set in opposition to each other and experimentally tested. The building and testing of historical and functional models are described in this paper. We hope to illustrate how rigorous model building can (1) help distinguish between selected and nonselected behaviors in proteins, (2) permit the engineering of the catalytic properties of enzymes, and (3) define the role of RNA in early catalysis.

To show how model building can help distinguish between selected and nonselected behaviors in proteins, let us consider a single trait, stereospecificity, of a single class of enzymes, alcohol dehydrogenases dependent on nicotinamide cofactors (NAD^+ and $NADP^+$). NADH bears two hydrogens (Fig. 1) at the 4-position, labeled R and S. The hydrogens are different, and individual dehydrogenases catalyze the transfer of only one of them. This is a trait that displays diversity; about half of the dehydrogenases examined transfer H_R and half transfer H_S.

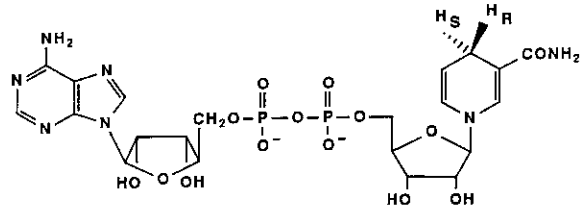


Figure 1. Two hydrogens at the 4-position of reduced nicotinamide cofactors are different and are designated R and S. Dehydrogenases selectively transfer only one of these hydrogens (Table 1).

When dehydrogenases are organized by their Enzyme Commission catalog number, stereospecificity appears to be random (see Table 1), consistent with the hypothesis that the trait is not selected. Either stereospecificity is drifting or it is a randomly fixed historical accident that has been conserved for unspecified reasons. Data exist that rule out the possibility that stereospecificity is drifting (You 1985). For example, all lactate dehydrogenases, including enzymes from bacteria, plants, and animals, transfer H_R, inconsistent with the notion that this trait is drifting. Rather, a historical model must make two assumptions: (1) All lactate dehydrogenases are descendants of a single

Table 1. Stereospecificity of Dehydrogenases Organized by Enzyme Commission Catalog Number

E.C. no.	Enzyme	Stereospecificity
1.1.1.1	ethanol dehydrogenase	R
1.1.1.3	homoserine dehydrogenase	S
1.1.1.6	glycerol dehydrogenase	R
1.1.1.8	glycerol-3-phosphate dehydrogenase	S
1.1.1.26	glyoxylate reductase	R
1.1.1.27	L-lactate dehydrogenase	R
1.1.1.28	D-lactate dehydrogenase	R
1.1.1.30	3-hydroxybutyrate dehydrogenase	S
1.1.1.35	3-hydroxyacyl-CoA dehydrogenase	S
1.1.1.37	malate dehydrogenase	R
1.1.1.38	malic enzyme	R
1.1.1.50	3-hydroxysteroid dehydrogenase	S
1.1.1.60	tartronate semialdehyde reductase	R
1.1.1.62	estradiol dehydrogenase	S
1.1.1.100	3-oxoacyl ACP Reductase	S
1.1.1.108	carnitine dehydrogenase	S

ancestral lactate dehydrogenase that transferred H_R (for no functional reason) and (2) stereospecificity is rigorously conserved during the divergence of plants, animals, bacteria, and mycoplasma (e.g., *Acholeplasma laidlawaii*).

Alternatively, a functional model might argue that the uniform stereospecificity displayed by lactate dehydrogenases reflects either convergent evolution of stereospecificity or a functional constraint on the drift of stereospecificity during the divergent evolution of lactate dehydrogenases. In both cases, the transfer of H_R is assumed to confer a selective advantage on the host organism.

In the two decades preceding 1980, the historical explanation came to be accepted as the only explanation for the origin of this particular trait (You 1985). This was largely due to the absence of an alternative, mechanistically reasonable functional model. One such model was recently proposed (Benner et al. 1985) on the basis of three hypotheses: (1) Which hydrogen is transferred is controlled by the conformation of bound NADH and is determined in part by stereoelectronic considerations, (2) different conformers of NADH have different redox potentials, and (3) functionally optimal enzymes match the redox potential of bound NADH to the redox potential of the natural substrate of the dehydrogenase.

A model based on these assumptions makes a simple prediction: Enzymes that have evolved to reduce unstable carbonyl substrates should have evolved to transfer H_R, whereas enzymes that have evolved to reduce stable carbonyl substrates should have evolved to transfer H_S. As shown in Table 2, this prediction is, in general, confirmed by experiment.

The ability of a functional model to create order in the distribution of an apparently random macromolecular behavior itself strongly supports the model, at least as a working hypothesis. However, the correlation in Table 2 is also consistent with a modified historical model that assumes that (1) there existed *two* ancestral dehydrogenases, one transferring (randomly) H_R and the other transferring (randomly) H_S; (2) the ancestral enzyme transferring H_R had a preference for unstable carbonyl substrates, whereas the enzyme transferring H_S must have had a preference for stable carbonyl substrates; and (3) both stereospecificity *and* substrate type were rigorously conserved during divergent evolution.

Implicit in the functional model and in the correlation shown in Table 2 is the statement that stereospecificity is not strongly selected in those enzymes acting on substrates with intermediate stabilities. A simple interpretation of the functional model suggests that if stereochemical diversity is to be observed in analogous enzymes from different organisms, it will be observed in enzymes acting on substrates such as ethanol or glycerol-3-phosphate, where the equilibrium constant of the overall reaction is approximately 10^{-11} M. The historical model predicts that such enzymes will have the same stereospecificity. This then might provide some grounds for distinguishing between the two models.

We have examined several enzymes that catalyze the reduction of substrates with the general structure R-CH_2CHO. The equilibrium constants for the redox reactions of these substrates are at the "break" in the correlation in Table 2. Thus, one naively expects that enzymes from different organisms acting on these substrates might have different stereospecificities. This is in fact the case. Ethanol dehydrogenases (e.g., from *Drosophila* and yeast) and hydroxymethyl glutaryl–coenzyme A (HMG-CoA) reductases (e.g., from *Acholeplasma* and mammal) both display stereochemical heterogeneity and act on substrates with redox potentials such that the functional model predicts that either particular stereospecificity would not be strongly selected.

Of course, these data do not by themselves rule

Table 2. Stereospecificity of Dehydrogenases Organized by Redox Potential of Presumed Natural Substrate

log K_{eq}	Enzyme	Stereospecificity
−17.5	glyoxylate reductase	R
−13.5	tartronate semialdehyde reductase	R
−12.8	glycerate dehydrogenase	R
−12.1	malate dehydrogenase	R
−12.1	malic enzyme	R
−11.6	L-lactate dehydrogenase	R
−11.6	D-lactate dehydrogenase	R
−11.2	ethanol dehydrogenase	R
−11.1	glycerol-3-phosphate dehydrogenase	S
−10.9	homoserine dehydrogenase	S
−10.9	carnitine dehydrogenase	S
−10.5	3-hydroxyacyl-CoA dehydrogenase	S
−8.9	3-hydroxybutyrate dehydrogenase	S
−7.7	estradiol 17-dehydrogenase	S
−7.6	3-oxoacyl ACP Reductase	S
−7.6	3-hydroxysteroid dehydrogenase	S

Of the 130 dehydrogenases studied to date, about 120 fit the correlation, perhaps 10 do not. For discussion, see Benner et al. (1985) and Oppenheimer (1984).

out all historical models. Perhaps all enzymes from organisms such as *Drosophila* are stereochemically "unusual," or perhaps enzymes acting on substrates with redox potentials far from the break also display stereochemical diversity. We have looked for evidence to support either argument, but have found none. For example, the functional model predicts that malate dehydrogenases from all organisms should transfer H_R. The enzymes from mammals, birds, fish, insects (e.g., *Drosophila*), fungi, eubacteria (e.g., *Clostridium acidiurici*), plants, and archaebacteria (e.g., *Halobacter halobium*) in fact all transfer H_R.

To explain the facts, the historical model must now assume two ancestral dehydrogenases for both ethanol and HMG-CoA as substrates (but not for malate and lactate as substrates), or postulate that stereospecificity can drift in ethanol dehydrogenases or HMG-CoA reductases (but not lactate dehydrogenases or malate dehydrogenases). These modifications, although possible, are ad hoc. Furthermore, in the absence of explanations as to why some enzymes have multiple ancestors or why stereospecificity can diverge in some enzymes and not in others, historical models with such ad hoc modifications are not predictive.

An independent argument suggests that stereospecificity may not be absolutely conserved within a group of homologous enzymes. We recently showed that the ethanol dehydrogenase from *Drosophila* transfers H_S, in contrast to the ethanol dehydrogenase from yeast, which transfers H_R (Benner et al. 1985). Earlier, Jornvall et al. (1981) noted sequence similarities in the dinucleotide-binding domain of these two proteins and suggested that these two enzymes might therefore be homologous. If their suggestion is correct, stereospecificity with respect to cofactor is not the same in at least some homologous dehydrogenases.

Historical models consistent with these data are too ad hoc to permit direct experimental test. Reasonable criteria for separating dehydrogenases into separate lineages, including sequence analysis, metal ion requirements, and substrate specificity, all fail to predict stereospecificity in any but the most trivial examples. In contrast, the functional model, although controversial (Oppenheimer 1984), remains a subject for intensive experimental investigation and has proved to be remarkably successful in anticipating experimental and theoretical results on enzymatic and model systems related to dehydrogenases. Several of these are listed in Table 3.

On the basis of similar models, well-studied macromolecular behaviors can be divided into two groups (Table 4). In one, working functional models exist that are predictively satisfactory, whereas in the other group, historical models are proving to be more satisfactory.

Even a limited distinction between functional and nonfunctional behavior in proteins can be valuable. For example, it can assist those attempting to engineer the behavior of proteins via molecular biological tools that allow the biochemist to alter the structure of catalytic proteins. Although "engineered enzymes" are widely recognized as desirable entities, there is at present no good theory to suggest which residues to change to achieve a desired perturbation in behavior. Therefore, residues generally are changed in the active sites of proteins and are chosen by inspection of a crystal structure. Not surprisingly, many of these changes have enormous (and often undesirable) impact on the behavior of the enzyme. Loss of a factor of 10^5 in catalytic activity is not uncommon (Cronin et al. 1987).

Alterations in structure at a distance from the active site offer the prospect of engineering more delicate changes; this strategy is more consistent with that followed by Nature as she engineers the behavior of proteins. However, the biochemist has difficulty with this strategy because of the number of mutations that are possible in residues remote from the active site.

We have argued for some time (Nambiar et al. 1986) that if we can distinguish in a general way between the results of selection and drift in macromolecular structure, an alternative rationale can be implemented for engineering of the behaviors of proteins by deliberate alteration of amino acid residues remote from the active site. In a strategy termed "evolutionary guidance," comparison of the structure and behavior of homologous enzymes, tempered by available understanding of selection and drift in proteins, provides guidance for changing the structure of a particular protein to yield new proteins with properties that are not simply the average of the properties of modern homologous proteins.

Consider, for example, a simply stated problem in

Table 3. Recent Results Consistent with the Functional Model

Computational verification of stereoselectivity based on conformation in reduced nicotinamide cofactors (Wu and Houk 1987)
Correlation between stereoselectivity and redox potential in nicotinamide model systems (Ohno et al. 1986)
Experimental and theoretical confirmation of arguments regarding adjustment of internal equilibrium constants in certain enzymes (Ellington and Benner 1987)
Convergent stereospecificity of nonhomologous dihydrofolate reductases (Matthews et al. 1986)
Stereospecificity of $NAD^+/NADP^+$ transhydrogenases (Kaplan 1967)

Table 4. Candidates for Selectable Macromolecular Traits

Stereoselection between diastereomeric transition states
NADH-dependent redox reactions
phosphoryl transfer reactions
Internal equilibrium constants
Kinetic parameters ±10%
Stability/instability
Substrate specificity against compounds present physiologically
Candidates for neutral macromolecular traits
Stereoselection between enantiomeric transition states
decarboxylation of beta-keto acids
pyridoxal-dependent decarboxylation of amino acids
Nonequilibrium dynamic motion of proteins (Stackhouse et al. 1985)
Substrate specificity against compounds not present physiologically

protein engineering: How can we make yeast alcohol dehydrogenase, an enzyme that reduces acetaldehyde to ethanol, reduce acetaldehyde faster? The task is especially difficult for two reasons. ADH from yeast is already an exceptionally fast enzyme. Furthermore, a crystal structure for the enzyme from yeast is not available. The closest structure is for the alcohol dehydrogenase from horse liver, an enzyme sharing only 35% sequence identity with the enzyme from yeast. However, a simple rationale exists for designing a faster enzyme. The rate-limiting step in the reduction reaction is not the hydride transfer, but the release of NAD^+. Therefore, a mutant form of the enzyme, where the free energies of all bound species are raised together (Fig. 2), might well be faster. Such a "uniform binding change" (Albery and Knowles 1976) can be effected by weakening the interaction between the enzyme and the adenine portion of the cofactor somewhat, but not by too much.

A comparison of the sequences of homologous dehydrogenases (Fig. 3) in a region near the adenine-binding site identified residue 211 as a position where variation in amino acid substitution might alter the binding of cofactor uniformly. The residue was changed from arginine to threonine. Representative kinetic constants of the resulting mutant enzyme are shown in Table 5. The results are notable in several respects. First, the increased dissociation constants of NAD^+ and NADH in the mutant dehydrogenase were as predicted. Second, the residue that was changed was remote from the active site. It would not have been targeted by a rationale based on simple inspection of a crystal structure. Third, the mutant enzyme had kinetic properties that were not the average properties of the enzymes found in nature. Indeed, mutant R211T appears to be a faster "acetaldehyde reductase" than any natural enzyme studied so far. "Evolutionary guidance" as a strategy for deliberately engineering the behavior of proteins must still be explored. However, it has passed an important test.

How can this approach be used to address questions regarding RNA catalysis and the origin of life? Here again, the interesting questions revolve around an analysis of contrasting functional and historical models, proposals that distinguish between primitive and derived traits, and distinctions between selection and drift as mechanisms for evolving observed behaviors. Here again, there is a virtue to rigorous model building.

A simple model for the involvement of RNA molecules as early catalysts is based on three assumptions: (1) The first life form consisted of an RNA-directed RNA polymerase that itself was an RNA

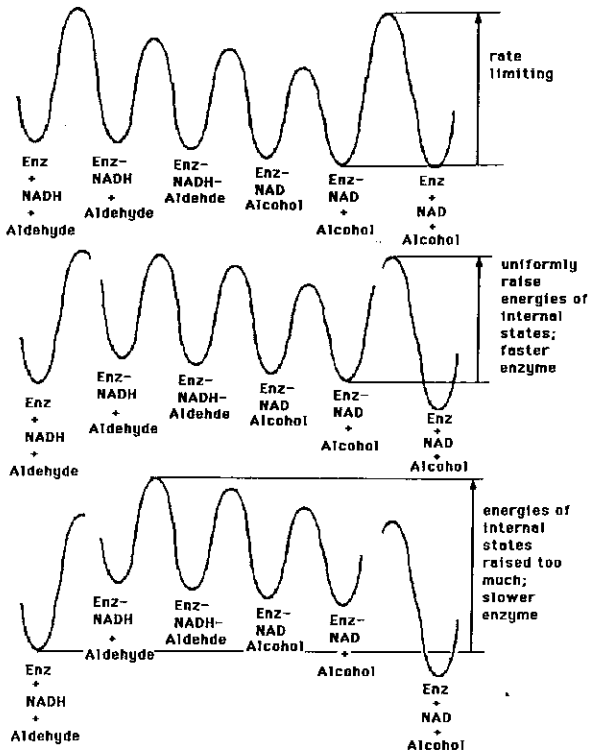
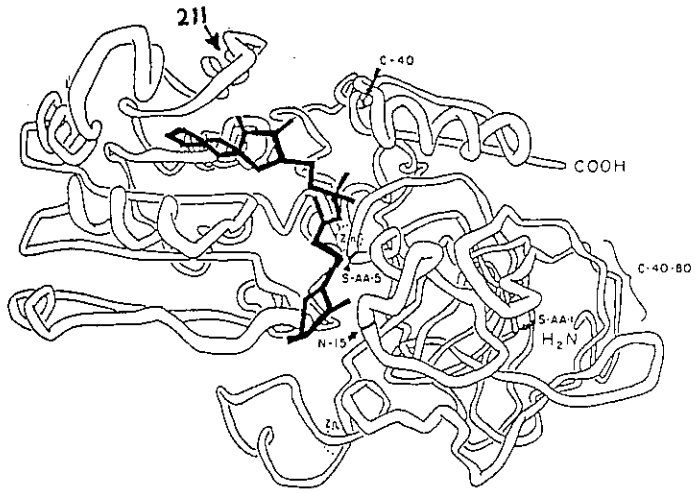


Figure 2. (*Top*) Abstract "free-energy profile" for the reaction catalyzed by yeast alcohol dehydrogenase. As evidenced by the small primary deuterium isotope effect ($V_H/V_D=1.4$), the rate-limiting step is not the transfer of hydrogen, but rather the release of NAD^+ cofactor. Therefore, moving the free energies of all internal states up by a modest amount (*middle*) is expected to increase the rate of the reaction. However, a uniform loosening of the binding of all internal states can easily be overdone (*bottom*), leading to a slower enzyme with another rate-limiting step. Thus, mutations at the active site are likely to be insufficiently delicate to produce the change desired to the extent desired. We have used "evolutionary guidance," based on an analysis of different sequences and behaviors of a set of homologous proteins, to identify residues remote from the active site that, when altered, will produce mutants with the desired kinetic properties.



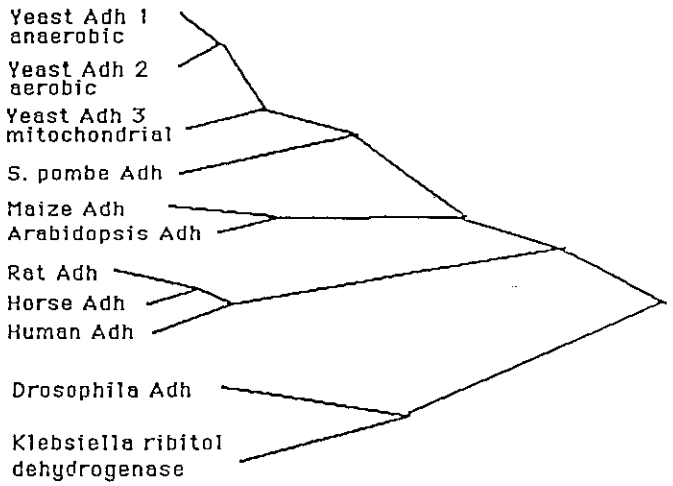


Figure 3. Analysis of the sequences and behaviors of a set of homologous proteins connected by the tree shown above identifies a region above the adenine-binding site where mutations might influence uniformly the binding of NAD^+ and NADH. Residues here, remote from the active site, would not be the targets of site-directed mutagenesis experiments had simple inspection of a crystal structure been the sole basis for the mutagenesis rationale. However, alterations here produce the desired change in binding and a faster enzyme (Table 7).

molecule; (2) the first synthesis of proteins was done by RNA catalysts; and (3) at some point, proteins took over most of the catalytic tasks previously performed by RNA catalysts.

Although a rudimentary form of this model was proposed by Woese, Crick, and Orgel in the late 1960s (Woese 1967; Crick 1968; Orgel 1968), the development of the model in the 1970s should not be overlooked. For example, in 1976, Usher and his co-workers discovered what we believe to be the first clear example of catalytic RNA (Usher and McHale 1976). Furthermore, the significance of the structure of many "ribo-cofactors," mentioned by Orgel (1968), was explored by White in 1976. Finally, in what might be regarded as the apogee of theoretical development of this model to date, Visser and Kellogg (1978) were able to correlate the model, the structure of ribo-cofactors, and their chemical reactivities.

Table 5. Kinetic Properties of Native and Mutant Alcohol Dehydrogenases

Protein	NAD^+ (K_D)	NADH (K_D)	V_H/V_D	Red (V_{max}; relative)
ADH 1	1.2 mM	0.075 mM	1.4	100
R211T	1.6 mM	0.185 mM	1.8	141

The model, especially as developed in the 1970s, solves four specific problems presented by the biochemistry of modern organisms: (1) the "chicken or egg" problem, originating in the simultaneous need for DNA to make proteins and proteins to make DNA, (2) the intermediacy of mRNA between DNA and proteins, (3) the presence of rRNA as the prinicipal component of ribosomes, and (4) the fact that many cofactors contain RNA-like moieties that do not participate in the chemical reaction and that can be removed with no ill effect on catalysis. The last point may be the most important (Fig. 4). Ribo-cofactors are distributed in nature in a pattern that suggests that these structures are primitive. The RNA parts of the cofactors almost certainly have no intrinsic function. Analogous molecules with similar reactivities that perform identical biological functions are known that lack any RNA component (Fig. 4). Without the historical model out-

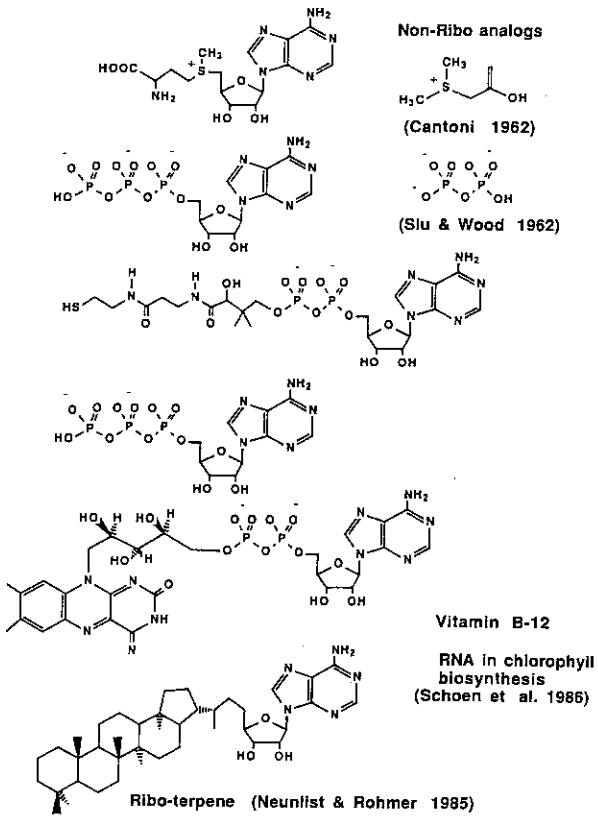


Figure 4. Many cofactors contain fragments of RNA that serve no apparent role in catalysis. Although arguments occasionally are made that these fragments serve some function (Fierke and Jencks 1986), their presence cannot reflect intrinsic function, as simpler variants of several of these cofactors (right-hand column) are chemically similar in other enzymatic systems. The wide distribution in nature of this apparently nonfunctional trait strongly suggests that ribo-cofactors are primitive structures. A model that presumes a *complex* ribometabolism based nearly exclusively on RNA as a catalyst offers a simple explanation for these structures. Models that presume a *primitive* ribometabolism preceding the advent of translation are logically inadequate to provide an analogous explanation. Finally, cofactors that arose during a period when proteins and RNA molecules both served as catalysts (a period that apparently extends from the time of the breakthrough organism to the present day) are expected to contain chemically nonessential fragments that include protein-like moieties. Acyl-carrier protein may well be an example of such a cofactor.

lined above, the repeated occurrence of nonfunctional RNA in cofactors whose structures are almost certainly primitive represents a problem. With the model, these enigmatic structures are explained as vestiges of an ancient metabolism based on RNA catalysts.

In more recent discussions of the "ribo-world" (Table 6) (Sharp 1985; Darnell and Doolittle 1986; Gilbert 1986; Westheimer 1986), the model has lost some of its logical force for three reasons.

1. The revised model has focused almost exclusively on RNA-catalyzed intron splicing and RNA processing, beginning with an assumption that these traits, and the related genetic structure of various species, are primitive vestiges of an earlier world.
2. The revised model views the geological time during which RNA served as catalyst as short.
3. The revised model focuses on the development of translation machinery as the first step in the evolution of a primitive riboorganism.

Each of these assumptions is problematical. Introns and their self-catalyzed splicing are not "obviously primitive traits" either chemically or by their distribution in nature. Transesterification reactions of phosphodiesters are readily catalyzed (see below); introns, self-splicing, and protein-catalyzed splicing display peculiar distribution in the modern world, and the diversity of splicing mechanisms is perplexing a priori. Modern examples of RNA catalysis can presently be explained equally well as vestiges of an ancient state or as a trait derived recently as an inexpensive way to remove introns that serve roles that are either regulatory (as perhaps is the case in T4 phage) (Ehrenman et al. 1986) or unknown. In the context of the historical model outlined above, RNA-catalyzed splicing may be viewed as confirming the historical model based on other grounds. However, it is a poor foundation for a historical model and does not make a convincing case if forced to stand alone. Furthermore, a model based on the notion that translation evolved before complex metabolism is not logically equipped to explain the data it is intended to explain. If the model is intended to explain the structure of NAD^+ as a vestige of a riboorganism, it must assume that the *breakthrough riboorganism*, the first organism to synthesize proteins by translation, had ribozymes catalyzing redox reactions using NAD^+. In fact, the model must assume that the breakthrough organism had several ribodehydrogenases; otherwise, it cannot easily explain why the ribocofactor structure was conserved across the breakthrough. This is also true for *S*-adenosylmethionine (implying that the breakthrough organism had ribo-

Table 6. Some Representative Comments Concerning the Plausibility of a Complex Metabolism in Riboorganisms

Lehninger (1972): "It is quite plausible that a primitive nucleotide-based life could have existed in the absence of proteins. But it is quite clear that not much evolutionary progress could have been made without proteins."
Visser (1984): "The catalytic capabilities of RNA molecules suggest that a short but decisive evolution towards primitive metabolism was possible even before the origin of the genetic code."
Orgel (1986): "The transition to a more modern biochemistry required the invention of genetic coding."

transmethylases), flavin adenine dinucleotide (ribo-oxidases), adenosine triphosphate (ribophosphate metabolism), riboterpenoids (ribozyme-based isoprenoid chemistry), CoA (ribozyme-catalyzed Claisen condensations), and other ribo-cofactors (Fig. 4).

This is simply not a description of an organism with a primitive metabolism. Rather, for the model to explain the vestigal RNA structures that are distributed in catalytic roles throughout the biological world, it must assume that the breakthrough organism had a fully developed metabolism.

One cannot dilute this picture without damaging the explanatory coherence of the model. For example, many current models attempt to introduce protein participation in the RNA world as early as possible to assist RNA as a catalyst. This attempt is apparently based on the assumption that RNA molecules "need" proteins, without which they are ineffective catalysts. Although this view can be questioned on chemical grounds (see below), explanations based on this modified model are necessarily weakened. To the extent that models assume the involvement of proteins in the world where ribo-cofactors evolved (which must also be the world where metabolism dependent on ribo-cofactors evolved), the model cannot explain why the nonfunctional fragments of cofactors are RNA and not (for example) proteinaceous. In the context of a logically explanatory model, acyl carrier protein is a cofactor that evolved in a ribonucleic acid–protein world; nicotinamide adenine dinucleotide is not. In the context of the model, the modern world is the world where proteins doubled with RNA as catalysts, not the ancient world.

To provide a logically coherent solution to the four problems mentioned above, the model must postulate a breakthrough riboorganism with complex metabolism. With an eye toward the chemistry of modern metabolic processes, one might inspect modern metabolic pathways to generate a model for the metabolism of this breakthrough organism. Although the details of the model vary depending on the rules one uses to construct the model, one plausible metabolism is shown in Table 7.

Table 8 illustrates how modern bioorganic data can be used to address a single question arising in the model: Was the breakthrough organism photosynthetic? Here again, the value of the discussion is not that it resolves this issue, but that it is a rich source of suggestions for future experimental work. The question is especially intriguing for another reason. The participation of RNA in the biosynthesis of chlorophyll in the modern world (Schoen et al. 1986) is consistent with the possibility that the breakthrough organism was photosynthetic (Table 8). The origin of photosynthesis can be approximately dated by the appearance of oxidized sediments in datable geological strata (Strother and Barghoorn 1980). Fossils of organisms are known preceding this time. Thus, if the metabolism outlined in Table 7 is correct, these fossils are fossils of riboorganisms.

The assumption that complex metabolism arose before ribosome-based translation has advantages in addition to its making the model explanatory. Translation is among the most complex biological processes known. It requires many chemical transformations, including the charging of tRNA molecules with appropriate amino acids, the translation step itself, and the modification of RNA bases. In chemical terms, it is more complex than many pathways. It is unreasonable on chemical grounds to argue that translation emerged before complex metabolism. Furthermore, many of the most perplexing problems discussed regarding the origin of translations are perplexing because it is assumed that translation arose in a metabolically primitive organism. These problems are lessened (but not solved) by models that presume a metabolically complex breakthrough organism.

Could RNA catalyze the chemical reactions that the historical model must presume are part of the breakthrough metabolism? Many have doubted this possibility (Table 6), although recent dramatic discoveries by Cech (Bass and Cech 1984), Altman (Guerrier-Takada and Altman 1983), Szostak (1986), and their co-workers have removed some of these doubts. However, accounting for catalysis by RNA molecules is not difficult in many cases, and catalysis by RNA in general is not a priori more problematic than catalysis by proteins. For example, in accounting for the transesterification reactions catalyzed by self-splicing RNA molecules, hydrophobic interactions have been largely overlooked (Fig. 5). Binding a phosphate in a hydrophobic region of RNA will raise the pK_a of the phos-

Table 7. One Model for the Breakthrough Organism

Reactions Part of the Breakthrough Organism's Metabolism
Oxidation/reduction reactions, aldol condensations, Claisen condensations, transmethylations
Lived in an aerobic environment, were photosynthetic
Degraded fatty acids, synthesized terpenes
Used DNA to store information
Energy metabolism based on ATP
Modified RNA bases
Reactions Not Part of the Breakthrough Organism's Metabolism
Biotin-dependent carboxylations
Fatty acid synthesis
Pyridoxal chemistry, transaminations

Table 8. Was the Breakthrough Organism Photosynthetic?

YES: The first step in the "C_5" pathway for the biosynthesis of chlorophyll involves RNA intimately in the catalytic step (Schoen et al. 1986).

NO: The C_5 pathway is not distributed in modern organisms as one would expect for a primitive trait. Plants have it. But both mammals and bacteria on opposite sides of the tree have an alternative pathway.

YES: But this distribution merely suggests that the enzymes catalyzing the first step of the alternative pathway in bacteria and mammals are not homologous. Bio-organic data suggest that this may be true. The enzyme from *Rhodopseudomonas spheroides* has a lysine in the active site, whereas the corresponding enzyme from rat may have an active-site cysteine (Nandi 1978).

AMBIGUITIES:
1. Arguments for an active-site cysteine in the mammalian enzyme are not strong.
2. At least some bacterial enzymes appear homologous by sequence to mammalian enzymes. However, a cysteine is highly conserved, not a lysine.
3. Some organisms have both pathways.

The area is rich with suggestions for new experiments.

phodiester from 2 to well above 10. Favorable binding interactions between the substrate and the RNA catalyst can compensate for the unfavorable energy of interaction between the catalyst and the phosphate group. The "unhappy" phosphate monoanion will pick up a proton from solvent; the protonated phosphate is 10^6 times more reactive toward transesterification than the diester monoanion. Proper orientation of the attacking nucleophile is worth an additional two to three orders of magnitude in rate enhancement. Although this mechanism postulates the involvement of a proton from solvent in the reaction, it does not require a general acid or general base attached to the catalyst itself. The proposed mechanism illustrates that hydrophobicity in an active site, if used judiciously, can, in principle, account for the majority of the catalysis observed in self-splicing.

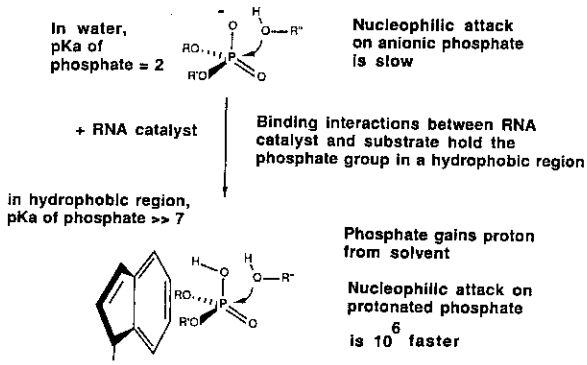


Figure 5. Mechanism for RNA-catalyzed transesterifications. Binding a phosphodiester (via favorable interactions between the RNA molecule and other portions of the diester) between two bases in an RNA catalyst brings the diester into a region with an effective dielectric of approximately 12. This will raise the pK_a of the phosphate from 2 to well above 10. The bound phosphate monoanion will therefore gain a proton from solvent. The protonated phosphate is 10^6 times more reactive toward transesterification than the diester monoanion. Proper orientation of the attacking nucleophile should provide an additional two to three orders of magnitude in rate enhancement. Thus, hydrophobic regions and "proximity" together can account for most of the rate acceleration observed in RNA self-splicing and processing reactions.

For more complicated reactions implicit in the ribometabolism listed in Table 7, catalysis by RNA need not be viewed as more problematic than catalysis of similar reactions by proteins. Although it appears to be a widespread belief that catalysis by proteins is easier to account for than catalysis by RNA because proteins have more functional groups than RNA, this view of the catalytic power of RNA molecules, especially in a metabolically sophisticated breakthrough organism, is not supported by chemical considerations.

Table 9 lists functional groups known to occur in RNA and protein molecules. In terms of available

Table 9. Functional Groups in RNA and Protein Molecules

On Building Blocks	
Hydrogen bonding	Hydrogen bonding
Hydrophobic groups	Hydrophobic groups
Phosphates	
Sugars	
	Aliphatic amines
	Carboxylates
	Sulfur
	Hydroxyl groups
	Imidazole
Posttranscriptional (Translational) Modification	
Aliphatic amines	
Carboxylates	
Sulfur	
Hydroxyl groups	
	Phosphates
	Sugars
Ketones	Ketones
Selenium	Selenium
Available on Cofactors	
Nicotinamide	Nicotinamide
R-SH	
Flavin	Flavin
Sulfonium ions	Sulfonium ions
Acyl anions	Acyl anions
Metals	Metals

Data from Adams et al. (1976).

functional groups, RNA and protein catalysts are not very different. Both biopolymers have certain functional groups built into the building blocks. Both lack other functional groups, but obtain these by posttranscriptional (translational) modification and from cofactors. This fact presents an interesting puzzle. The fact that most modern catalysts are *not* RNA molecules suggests, in the context of this model, that proteins are superior to RNA molecules as catalysts. What is the chemical basis for this superiority? There are two interesting possibilities. There may be some special features of a polypeptide backbone (e.g., lack of charge) that make proteins better suited as catalysts. Alternatively, a higher fraction of the functional groups found in RNA molecules are obtained by posttranscriptional modification. Posttranscriptional modification requires information, which is undoubtedly costly. This may represent a disadvantage for an organism that uses RNA as its principal catalyst.

The latter problem reflects a shortage of building blocks for RNA molecules, which, in turn, reflects the fact that RNA molecules exploit only two of eight possible hydrogen-bond-pairing schemes theoretically available for a "pyrimidine-purine-like" base pair (Fig. 6). The two that are exploited (uracil and cytosine as the "pyrimidine"; see Fig. 6) are the only two where the "pyrimidine-like" base is joined to the ribose by a carbon-heteroatom bond. To construct base pairs using other hydrogen-bonding schemes, pyrimidine analogs must be joined to a ribose via a *carbon-carbon* bond. These constructions are difficult to envision as products of prebiotic chemistry. This may be why these other hydrogen-bonding schemes are not used in RNA.

As organic chemists, we are not constrained by prebiotic chemistry. We have synthesized a new base pair, where a new hydrogen-bonding pattern is exploited using pyrimidine analogs joined to ribose via a carbon-carbon bond. An outline of the synthetic scheme is shown in Figure 6. Thus, we have expanded the number of "letters" in the RNA "alphabet," offering the potential of increasing the number of functional groups that can be introduced into RNA directly, diminishing the number that must be introduced posttranscriptionally, and improving the intrinsic power of RNA as a catalyst.

This advance comes 2 billion years too late to help those organisms that tried to use only RNA catalysis to compete with the breakthrough organism. However, pseudouridine, a naturally occurring nucleoside formed posttranscriptionally with a base joined to a ribose by a carbon-carbon bond, may reflect an effort by a breakthrough organism to do by chance billions of years ago what we have done in the last year using organic synthesis.

An important final comment about the origin of life also follows from the historical model. The traits of a primitive organism can be guessed only by comparison of the traits of many of its descendants. We cannot infer the structure of a primitive organism by examining a single descendant simply because we cannot distinguish traits in the descendant that are primitive from those that are derived.

A metabolically complex breakthrough organism living in an ecologically diverse world does not resemble the first organism. Furthermore, evidence suggests that only a single breakthrough organism existed (Lake 1985). This implies that competing riboorganisms (those that did not participate in the breakthrough) became extinct (Fig. 7). Therefore, even if we were to deduce a complete and accurate model of the breakthrough organism, we could not extrapolate past the evolutionary bottleneck back to the properties of more primitive organisms. Our view to primeval times is necessarily obscured. Examining the modern world for clues to the origin of life is futile.

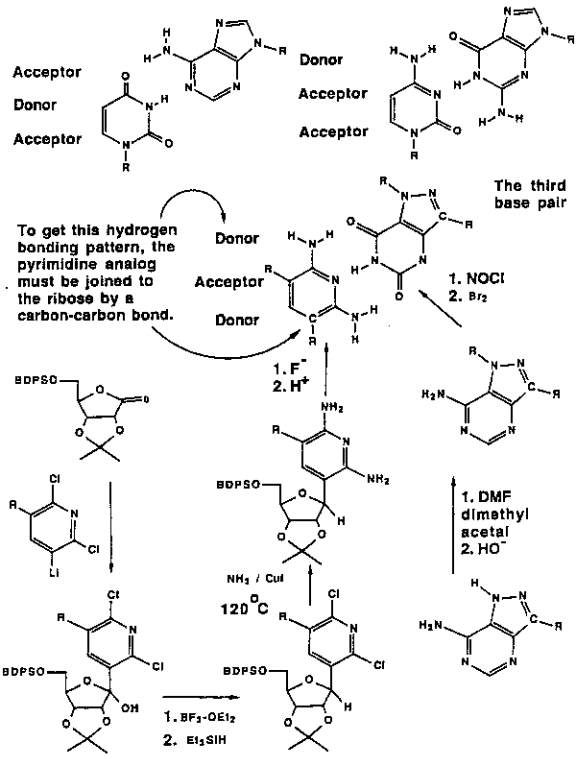


Figure 6. Only two of the eight possible "standard" Watson-Crick hydrogen-bonding schemes available to RNA are used in modern biochemistry. Those involving U and C as the pyrimidine-like unit are the only ones where the pyrimidine base is attached to the sugar by a carbon-nitrogen bond. To achieve other base-pairing schemes (*center*), the pyrimidine-like unit must be attached to the sugar by a carbon-carbon bond. Such molecules are almost certainly not found in prebiotic soups, and their absence in modern biochemistry may reflect a constraint imposed upon the evolution of nucleic acids by prebiotic chemistry. However, the presence of pseudouridine in modern RNA may reflect an ancestral effort by riboorganisms to relax this constraint. Organic chemists are not constrained by prebiotic chemistry. We have synthesized the molecules necessary to construct the "third base pair" (*bottom*). This is the first step toward the development of RNA molecules built from eight or more building blocks, capable of bearing a more diverse set of functional groups, and (presumably) inherently better catalysts.

Ellington, A. and S.A. Benner. 1987. Free energy differences between enzyme bound states. *J. Theor. Biol.* (in press).

Fierke, C.A. and W.P. Jencks. 1986. Two functional domains of coenzyme A activate catalysis by coenzyme A transferase. *J. Biol. Chem.* **261:** 7603.

Gilbert, W. 1986. The RNA world. *Nature* **319:** 818.

Guerrier-Takada, C. and S. Altman. 1983. Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science* **223:** 285.

Jornvall, H., M. Persson, and J. Jeffrey. 1981. 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.* **78:** 4226.

Kaplan, N.O. 1967. Beef heart TPNH-DPN pyridine nucleotide transhydrogenases. *Methods Enzymol.* **10:** 317.

Lake, J.A. 1985. Evolving ribosome structure: Domains in archaebacteria, eocytes, and eukaryotes. *Annu. Rev. Biochem.* **54:** 507.

Lehninger, A. 1972. *Biochemistry*. Worth, New York.

Matthews, D.A., S.L. Smith, D.P. Baccanari, J.J. Burchall, S.M. Oatley, and J. Kraut. 1986. Crystal structure of a novel trimethoprim-resistant dihydrofolate reductase specified in *Escherichia coli* by R-plasmid R67. *Biochemistry* **25:** 4194.

Nambiar, K.P., J. Stackhouse, S.R. Presnell, and S.A. Benner. 1986. Evolutionary guidance and the engineering of enzymes. In *Enzymes as catalysts in organic synthesis* (ed. M. Schneider), p. 325. D. Reidel, New York.

Nandi, D.L. 1978. Delta-aminolevulinic acid synthase of *Rhodopseudomonas spheroides*. *Arch. Biochem. Biophys.* **188:** 266.

Neunlist, S. and M. Rohmer. 1985. A novel hopanoid, 30-(5′-adenosyl)hopane, from the purple non-sulphur bacterium *Rhodopseudomonas acidophila*, with possible DNA interactions. *Biochem. J.* **228:** 769.

Ohno, A., M. Ohara, and S. Oka. 1986. $NAD(P)^+$-NAD(P)H models. 61. An interconversion between central and axial chiralities as an evidence for a functional model of chemical evolution of an enzyme. *J. Am. Chem. Soc.* **108:** 6438.

Oppenheimer, N.J. 1984. Stereoselectivity of enzymic transfer of hydrogen from nicotinamide coenzymes: A stereochemical imperative? *J. Am. Chem. Soc.* **106:** 3032.

Orgel, L.E. 1968. Evolution of the genetic apparatus. *J. Mol. Biol.* **38:** 381.

———. 1986. RNA catalysis and the origins of life. *J. Theor. Biol.* **123:** 127.

Schoen, A., G. Krupp, S. Gough, S. Berry-Lowe, C.G. Kannangara, and D. Soell. 1986. The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. *Nature* **322:** 281.

Sharp, P.A. 1985. On the origin of RNA splicing and introns. *Cell* **42:** 397.

Siu, P.M.L. and H.G. Wood. 1962. Phosphoenolpyruvic carboxytransphosphorylase, a CO_2 fixation enzyme from propionic acid bacteria. *J. Biol. Chem.* **237:** 3044.

Stackhouse, J., K.P. Nambiar, J.J. Burbaum, D.M. Stauffer, and S.A. Benner. 1985. Dynamic transduction of energy and internal equilibria in enzymes: A reexamination of pyruvate kinase. *J. Am. Chem. Soc.* **107:** 2757.

Strother, P.K. and E.S.H Barghoorn. 1980. *Origins of life and evolution* (ed. H.O. Halvorsen and K.E. van Holde), p. 14. A.R. Liss, New York.

Szostak, J.W. 1986. Enzymatic activity of the conserved core of a group I self-splicing intron. *Nature* **322:** 83.

Usher, D.A. and A.H. McHale. 1976. Hydrolytic stability of helical RNA: A selective advantage for the natural 3′, 5′-bond. *Proc. Natl. Acad. Sci.* **73:** 1149.

Visser, C.M. 1984. Evolution of biocatalysis. *Origins Life* **14:** 291.

Visser, C.M. and R.M. Kellogg. 1978. Biotin. Its place in evolution. *J. Mol. Evol.* **11:** 171.

Westheimer, F.H. 1986. Polyribonucleic acids as enzymes. *Nature* **319:** 534.

White, H.B. III. 1976. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7:** 101.

Woese, C.R. 1967. *The origins of the genetic code*. Harper and Row, New York.

Wu, Y.-D. and K.N. Houk. 1987. Theoretical transition structures for hydride transfer to methyleniminium ion from methylamine and dihydropyridine. On the nonlinearity of hydride transfers. *J. Am. Chem. Soc.* **109:** 2226.

You, K.-S. 1985. Stereospecificity for nicotinamide nucleotides in enzymatic and chemical hydride transfer reactions. *CRC Crit. Rev. Biochem.* **17:** 313.