EVOLUTIONARY GUIDANCE AND THE ENGINEERING OF ENZYMES

Krishnan P. Nambiar, Joseph Stackhouse,
Scott R. Presnell, and Steven A. Benner
Department of Chemistry
Harvard University
12 Oxford Street
Cambridge, Massachusetts 02138 U.S.A.

ABSTRACT. Evolutionary guidance is discussed as a strategy for simplifying the problems associated with the use of recombinant DNA techniques to study structure-function relationships in proteins. Evidence is given that suggests that natural selection "fine tunes" the behavior of proteins. Also progress on the synthesis, cloning, and expression of a gene for pancreatic ribonuclease, designed to permit rapid modification, is reported.

1. INTRODUCTION

Despite efforts in many laboratories to develop enzymes as synthetic reagents, their use in laboratory and industrial organic chemistry remains the exception rather than the rule.

While this may reflect conservatism among organic chemists, it may also simply mean that enzymes are not as useful as traditional organic catalysts in solving synthetic problems. Enzymes are often not easily obtained. Even with recent advances in recombinant DNA technology, expression and purification of proteins are often problematic and expensive. Once in hand, the enzyme may not have the desired substrate or stereospecificity; even rarer are pairs of enzymes with opposite stereospecificities that permit flexibility in synthetic design. Finally, enzymes are often rather unstable, especially in non-native environments.

While enzymes are likely to maintain their position as catalysts for effecting single transformations on biological molecules, will they ever be the catalysts of choice for the broader range of synthetic tasks? Three features of enzymic catalysis make the possibility real:

M. P. Schneider (ed.), Enzymes as Catalysts in Organic Synthesis, 325–340.
© 1986 by D. Reidel Publishing Company.

1. Enzymes are the products of billions of years of biological evolution. Therefore, enzymes may be inherently better (faster, more specific) than any catalyst that an organic graduate student can construct.

2. Enzymes are "combinable". They can remove a single substrate from a pool containing many compounds, and effect as single transformation on it. Thus, they can participate in designed one-pot multistep pathways that entirely rearrange the carbon skeleton of a precursor, all under isoenergetic conditions, just as in a natural biosynthetic pathway. Therefore, enzymes may form the basis of economical, multistep, one pot synthetic routes to complex chemicals.

3. Enzymes as macromolecules can have subtly different structures and subtly different physical and catalytic properties as a result. A series of proteins displaying a continuum of properties, ranging from rate and selectivity to stability and solubility, are in principle possible given the appropriate variations in amino acid sequence.

While these properties of polypeptide catalysts make them distinct from even the most ideal conventional catalysts, they are not obviously advantages. As products of natural selection, the properties that enzymes have evolved are often undesirable to a bio-technologist. For example, instability is a desirable trait of a protein that functions in an inducible pathway. It is almost never a desirable trait in a protein destined for practical application.

Therefore, in order to take full advantage of the special properties of enzymes to make them the catalysts of choice to solve general synthetic problems, sooner or later we must face the problem of how to engineer the properties of the protein by deliberately altering the amino acid sequence of the protein. If "rules" become available that relate the amino acid sequence of a protein to its physical and catalytic behavior, any undesirable property of a protein could, at least in principle, be eliminated.

Science, not technology, presently limits efforts to engineer proteins. Virtually any protein can now be modified using recombinant DNA technology acting on a gene for the protein that has been cloned and expressed. However, while we are confident theoretically that there exists some set of alterations that will confer almost any desired property on any particular protein, in general we have no idea which alterations to make in the protein's structure to have the desired impact on the protein's behavior.

Nor are recombinant DNA techniques alone sufficient to develop an understanding of structure-function relationships in proteins. Naively, one might expect that such understanding would come from changing the structure of a protein and seeing what the effect is. However, even here it is not obvious what changes to make to obtain insights about how the behaviors of proteins (stability, quaternatery structure, substrate specificity) are determined by the amino acid sequence. First, the number of changes possible is astronomically large. For even a small protein like ribonuclease, there are 2356 single variants that are one amino acid different from the native structure, 5,505,972 that are different by two, and nearly 13 billion that are different by three. Introducing mutations via the commonly used strategies is not without effort, requiring several steps, including screening of colonies of bacteria that contain the mutated gene.(1) Even making a variant per week (a generous estimate), it is not likely that a "random walk" through the protein's structure will produce data that are useful for developing a general understanding of protein structure.

It is therefore not surprising that most laboratories exploring structure-function relationships in proteins target residues in the active site that, by inspection, are presumably "important" for catalytic function.(2) These residues are changed, and the altered protein is assayed for activity. If the enzyme is inactive, the residue changes is concluded to be "essential" for catalysis. The "inspect a crystal structure, change an active site residue, and kill the protein" strategy is obviously limited in what it can reveal about changing the structure of protein to improve its physical properties while retaining its catalytic activity. Unfortunately, this information is the kind that the biotechnologist finds most interesting.

The work that I will report today is based on the hypothesis that evolutionary processes have been engineering proteins for billions of years, introducing single amino acid changes into existing proteins to obtain new proteins with desired physical and catalytic properties. If this hypothesis is correct, studying how evolution engineers proteins is a strategy for developing insight that will allow the chemist to do the same. We shall first present evidence supporting this hypothesis, and shall then report our recent progress in synthesizing, cloning, and expressing genes for ribonuclease (RNAse), a protein especially well suited for exploiting "evolutionary guidance" in understanding structure-function relationships in proteins.

2. DOES EVOLUTION ENGINEER THE BEHAVIOR OF PROTEINS?

First, before examining the evolution of proteins to gain insight needed for engineering proteins, we must ask whether there is evidence that the properties of enzymes are indeed engineered by the process of natural selection.

Traits of enzymes are either selected or they are not, depending on whether they influence the survival of the host organism.(3) As natural selection is the only mechanism for obtaining functional behavior in living systems, only selected traits reflect function. Non-selected traits reflect history, including both the traits of ancestral enzymes and any constraints that may prevent those ancestral traits from "drifting."(3) They need not reveal anything fundamental about the workings of biological catalysts.

To use evolutionary history as a guide, we must know whether most of the detailed behavior of proteins reflects function achieved by natural selection, or merely reflects the random accumulation of "neutral" mutations by a process of "drift".(3) For example, if the stability characteristic of pancreatic RNAse is a selected trait, we might expect to see specific amino acids altered in the course of the evolution of the protein to create this stability. If the stability is not a selected trait, its presence in pancreatic RNAse must be an evolutionary accident. The evolution of proteins leading to the stable modern enzyme need not be instructive of how to engineer stability into a protein.

There are many subtle traits of enzymes that bioorganic chemists study where the same question can be asked. For example, dehydrogenases dependent on nicotinamide cofactors transfer only one of the two hydrogens at the 4 position of NADH. About half of the enzymes studied so far transfer the pro-R (or "A") hydrogen, about half transfer the pro-S (or "B") hydrogen.(4) Table 1 shows a list of dehydrogenases listed in the order of the catalog number used by the Enzyme Commission to identify the enzyme.

The impression gained from an inspection of Table I is that the hydrogen transfered by any particular dehydrogenase is "random". Therefore, this trait may not be the product of natural selection, and may not be able to provide fundamental information about protein catalysis. However, in the absence of an alternative testable hypothesis that interprets the same data in terms of function, we cannot be sure whether the appearance of "randomness" reflects actual randomness, or whether it simply reflects our ignorance of the underlying functional significance of the trait.

Table I

E.C.	Reference	Name	Stereochemistry
1.1.1.1	25	Alcohol dehydrogenase (Yeast)	A
1.1.1.3	26	Homoserine dehydrogenase	B
1.1.1.6	27	Glycerol dehydrogenase	A
1.1.1.8	28	Glycerol-3-phosphate dehydrogenase	B
1.1.1.26	29	Glyoxylate reductase	A
1.1.1.27	30	L-Lactate dehydrogenase	A
1.1.1.28	30	D-Lactate dehydrogenase	A
1.1.1.29	31	Glycerate dehydrogenase	A
1.1.1.30	32	3-Hydroxy butyrate dehydrogenase	B
1.1.1.35	33	3-Hydroxy acyl CoA dehydrogenase	B
1.1.1.37	34	Malate dehydrogenase	A
1.1.1.38	34	Malic enzyme	A
1.1.1.40	34	Malic enzyme (NADP)	A
1.1.1.51	35	β-Hydroxysteroid dehydrogenase	B
1.1.1.60	36	Tartronate semialdehyde reductase	A
1.1.1.62	37	Estradiol 17 β-dehydrogenase	B
1.1.1.64	35	Testosterone β-dehydrogenase	B
1.1.1.72	27	Glycerol dehydrogenase (NADP)	A
1.1.1.79	29	Glyoxylate reductase (NADP)	A
1.1.1.81	31	Hydroxypyruvate reductase	A
1.1.1.82	34	Malate dehydrogenase (NADP)	A
1.1.1.91	38	Aryl alcohol dehydrogenase	B
1.1.1.100	39	3-Oxoacyl acyl carrier protein reductase	B
1.1.1.108	40	Carnitine dehydrogenase	B
1.1.1.50	35	3-α-Hydroxysteroid dehydrogenase (P. test.)	B

Stereoselectivity of some dehydrogenases using NADH, listed by Enzyme Commission catalog number. By convention, the "A" and "B" hydrogens are the pro-R and pro-S hydrogens.

Recently, we proposed such a functional theory that suggests that dehydrogenase stereospecificity can be explained as a trait engineered by selective forces to obtain an enzyme with optimal catalytic efficiency. Our model was based on four hypotheses:(5)

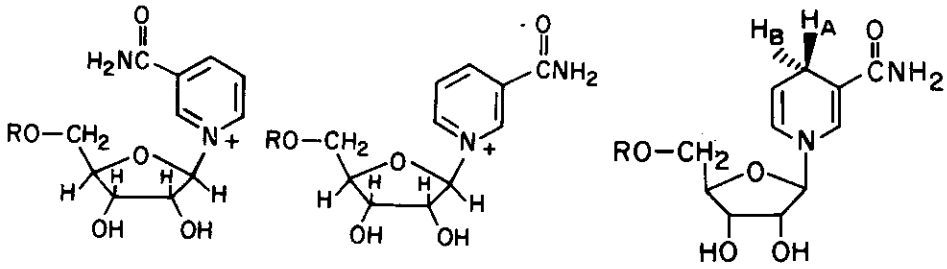

Figure 1: Syn and anti conformations of nicotinamide cofactors, and the two distinct hydrogens (A and B) of NADH.

(a) Dehydrogenases have evolved to reflect stereoelectronic principles. The "A" hydrogen is transfered if the enzyme has evolved to bind the cofactor in the "anti" conformation; the "B" hydrogen is transfered if the enzyme binds the cofactor in the "syn" conformation.(Figure 1).(6)

(b) The "anti" conformation of the cofactor is a weaker reducing agent than the "syn" conformation.

(c) Enzymes are catalytically optimal when they bind cofactor and substrates so that the free energy drop between enzyme-substrate and enzyme-product complexes reflects some of the chemical potential drop of the reaction under physiological conditions.(7)

(d) Dehydrogenases have evolved to be catalytically optimal.

These hypotheses suggest that the stereospecificity of dehydrogenases should correlate with the redox potential of the substrate that the enzyme had evolved to handle. Indeed, upon rearranging the enzymes listed in Table I so as to order them not by catalog number, but rather by the stability of their presumed natural substrate (Table II), a correlation between redox potential and stereospecificity is apparent. Reactive carbonyls are reduced by the pro-R hydrogen, unreactive carbonyls are reduced with the pro-S hydrogen. This correlation is fitted by 125 of the 130 enzymes studied so far.(8)

Table II

enzyme[a]	EC no.	$-\log E_{eq}$[b]	stereochemistry[c]
glyoxylate reductase	1.1.1.26	17.5	*pro-R*
glyoxylate reductase (NADP$^+$)	1.1.1.79	17.5	*pro-R*
tartronate-semialdehyde reductase	1.1.1.60	13.3	*pro-R*
glycerate dehydrogenase[d]	1.1.1.29	13.3	*pro-R*
glycerol 1-dehydrogenase	1.1.1.72	12.8	*pro-R*
hydroxypyruvate reductase (NADP$^+$)[d]	1.1.1.81	12.4	*pro-R*
malate dehydrogenase	1.1.1.37	12.1	*pro-R*
malate dehydrogenase (NADP$^+$)	1.1.1.82	12.1	*pro-R*
malic enzyme	1.1.1.38	12.1	*pro-R*
malic enzyme (NADP$^+$)	1.1.1.40	12.1	*pro-R*
L-lactate dehydrogenase	1.1.1.27	11.6	*pro-R*
D-lactate dehydrogenase	1.1.1.28	11.6	*pro-R*
ethanol dehydrogenase (yeast)	1.1.1.1	11.4	*pro-R*
glycerol 2-dehydrogenase	1.1.1.6	11.3	*pro-R*
glycerol-3-phosphate dehydrogenase	1.1.1.8	11.1	*pro-S*
homoserine dehydrogenase	1.1.1.3	10.9	*pro-S*
carnitine dehydrogenase	1.1.1.108	10.9	*pro-S*
3-hydroxyacyl-CoA dehydrogenase	1.1.1.35	10.5	*pro-S*
3-hydroxybutyrate dehydrogenase	1.1.1.30	8.9	*pro-S*
3β-hydroxysteroid dehydrogenase	1.1.1.50	8.0	*pro-S*
testosterone dehydrogenase	1.1.1.64	7.6	*pro-S*
3-oxoacyl-ACP dehydrogenase	1.1.1.100	7.6	*pro-S*
β-Hydroxysteroid dehydrogenase	1.1.1.51	7.6	*pro-S*
oestradiol 17β-dehydrogenase	1.1.1.62	7.7	*pro-S*

Stereoselectivities of dehydrogenases arranged by the equilibrium constant [NADH][carbonyl][H^+]/[NAD^+][alcohol] for the redox reaction that the enzyme has evolved to catalyze.

The fact that the apparently random data can be organized by a functional model suggests, but does not prove, that evolutionary selection pressures have selected for one stereochemical outcome over the other in different enzymes.

Further evidence that this stereoselectivity reflects selected function rather than random drift can be found by testing predictions that the functional theory makes, especially those that differ from those made by historical models. For example, one generalization regarding dehydrogenase stereoselectivity, known as "Bentley's first rule", is that enzymes from different organisms but acting on the same substrate all have the same stereochemistry.(9)

Historical models treat this as evidence that all enzymes acting on the same substrate are related, with the stereospecificity of the ancestral enzyme being conserved. Functional models such as ours argue that this reflects a selection for a particular stereochemistry for functional reasons related to the redox potential of the natural substrate, the same for all of the enzymes in question. If the functional theory is correct, we would expect that enzymes acting on substrates having redox potential at the break in the correlation would not be under strong selective pressure to favor one stereospecificity over the other. Therefore, we might expect that ethanol dehydrogenases, for example, to come in two stereochemical varieties if the functional model is correct, but not if a historical model is correct.

Therefore, we have examined a number of ethanol dehydrogenases, and find that indeed, some transfer the pro-S hydrogen (the enzyme from Drosophila melanogaster) while others transfer the pro-R hydrogen (the enzyme from yeast). In contrast, all malate dehydrogenases transfer the pro-R hydrogen, while all 3-hydroxyacylcoenzyme A and hydroxybutyrate dehydrogenases transfer the pro-S hydrogen. These data are consistent with our functional model, but not with simple historical models.

Even more subtle details of enzymic catalysis appear to reflect function, and therefore are selected. For example, the free energies of enzyme-substrate and enzyme-product (ES and EP) complexes also appear to reflect selective pressure seeking to evolve enzymes that are catalytically optimal. Based on a simple model (7), enzymes may be catalytically optimized by adjusting the free energies of ES and EP complexes so that they reflect part, but not all, of the chemical potential drop under physiological conditions. This is refered to as a "descending staircase" free energy profile, and is exemplified by the profile displayed by pyruvate kinase.(Figure 2) (7)

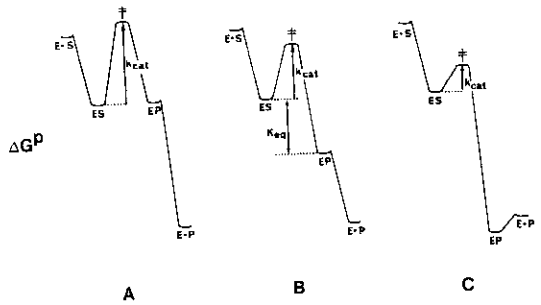

Figure 2: Three possible free energy profiles having different "internal equilibrium constants" between the bound states ES and EP. Profile B is the "descending staircase" profile that gives the fastest rate of catalysis (ref. 7). The ordinate represents physiological free energy, the standard state free energy corrected for physiological concentrations of S and P.

The chemical potential drop under physiological conditions reflects both the intrinsic free energies of the compounds, and their concentrations under physiological conditions. Therefore, the chemical potential drop is different under different conditions, and is reversed when the physiological flux is reversed. Thereofre, the descending staircase model predicts that isozymes whose physiological roles are to catalyze the same reaction in opposite directions should have different internal equilibrium constants. For example, there are two lactate dehydrogenases, one operating under oxidative conditions (from heart), believed to catalyze the conversion of lactate to pyruvate (10), the other under reductive conditions (from muscle), catalyzing the reverse reaction, the conversion of pyruvate to lactate.

The descending staircase model predicts that, for the heart enzyme, the free energy of the Enzyme-pyruvate-NADH complex should be lower than that of the Enzyme-lactate-NAD complex. For the muscle enzyme, the free energies of the two complexes should be reversed. (Figure 2) Estimates of the internal equilibrium constants for both enzymes verify this prediction.(5c,11) The internal equilibrium constant for the heart enzyme appears to be 1.5:1 downhill in the direction of pyruvate;(5c) that for the muscle enzyme appears to be 4:1 downhill in the direction of lactate.(11)

Therefore, enzymes seem to have evolved to adjust the free energies of bound species to reflect both the intrinsic free energies of the compounds that are substrates, but also the nature of the physiological surroundings.

These data do not "prove" functional theories of dehydrogenase stereoselectivity or internal thermodynamics, any more than they "disprove" historical models explaining the same data are incorrect. The statistics associated with stereochemistry requires that many more examples be collected before conclusions can be firmly drawn. Likewise, many more internal equilibrium constants must be measured to provide a large enough number of samples to be confident of our functional interpretation of internal equilibrium constants.

However, the behavior of dehydrogenases appears to conform more closely to functional models than to historical models. It is data such as these that provide the basis of our belief that, in general, enzymes have evolved to be highly refined catalysts optimally suited for effecting specific transformations on specific natural substrates under defined conditions. Thus, within the constraints of Darwinian evolutionary theory, it appears as if natural selection is capable of engineering even subtle details of enzymic behavior without major constraints. This in turn suggests that the process by which evolution engineers proteins is likely to be interesting to those seeking insight into how to engineer proteins in the laboratory.

3. SYNTHESIS, CLONING AND EXPRESSION OF A GENE FOR RNAse

We shall assume that evolutionary processes are constantly searching the "surface" that relates protein structure to protein function, optimizing the protein's structure for optimal survival under the particular conditions it finds itself. Thus, the evolutionary history of a protein that has recently evolved interesting behavior is interesting to study. If we can learn how evolution engineers proteins by changing one amino acid at a time, we may be able to gain insight that will permit the chemist to do the same.

To exploit evolutionary guidance as a rationale for learning about structure-function relationships in proteins, we need to study a protein which has an interesting and reasonably well defined recent evolutionary history, and which can be easily studied structurally and kinetically, and physically. We have chosen ribonuclease (RNAse) as such a system for study, and wish to report our progress in developing the system for structure-function studies. This progress includes synthesis of a gene coding for the protein designed to be easily altered, cloning of that gene into *E. coli*, and expressing the protein as a fusion protein with beta-galactosidase.

RNAse is a small protein with approximately 124 amino acids. It is found widely in the digestive tract of ruminants (artiodactyls), where it has undergone a rapid recent evolutionary divergence as ruminants themselves have diverged to occupy a wide range of environmental niches.(12) Thanks in large part to the extensive work of Beintema and coworkers,(13) the amino acid sequences of many of these modern RNAses are known. From there modern enzymes, a tree can be constructed (Figure 3) which includes hypothetical amino acid sequences of the ancestral RNAses from the (now extinct) ancestors of the modern ruminants. This tree contains a number of evolutionary paths connecting proteins which have altered properties. These include:

1. Primary Structure: Some 70% of the residues in RNAse have been varied by evolutionary processes within the sequences that have been examined from higher organisms.(12)

2. Stability: Certain RNAses are exceptionally stable, others are less stable, all within the same evolutionary series.

3. Quaternary structure: Certain of the related RNAses are monomers, others are dimers covalently linked by disulfide bonds, while still others may transiently form dimers.(14)

4. Anti-tumor activity: Certain RNAses have potent anti-tumor activity, others have no anti-tumor activity at all.(15)

Of course, we would be unable to predict with any particular RNAse what residues to change to alter the proteins stability or quaternary structure simply by inspection of a crystal structure. Further, given the fact that two thirds of the structure is a candidate for alteration, a random walk is unthinkable. Even making good guesses based on fundamental organic principles is unlikely to lead to a dimeric pancreatic RNAse from a monomer it likely to produce countless of inactive failures before the goal is achieved.

However, within the tree in Figure 3, paths connect proteins with such different properties. Proteins that are stable are connected with those that are unstable. Proteins that are monomers are connect by paths involving single amino acid changes with those that are dimers. The paths proceed by single amino acid substitutions, and the intermediate proteins can be made with molecular biological techniques and studied, providing a new rationale for studying the relationships between protein structure and function.

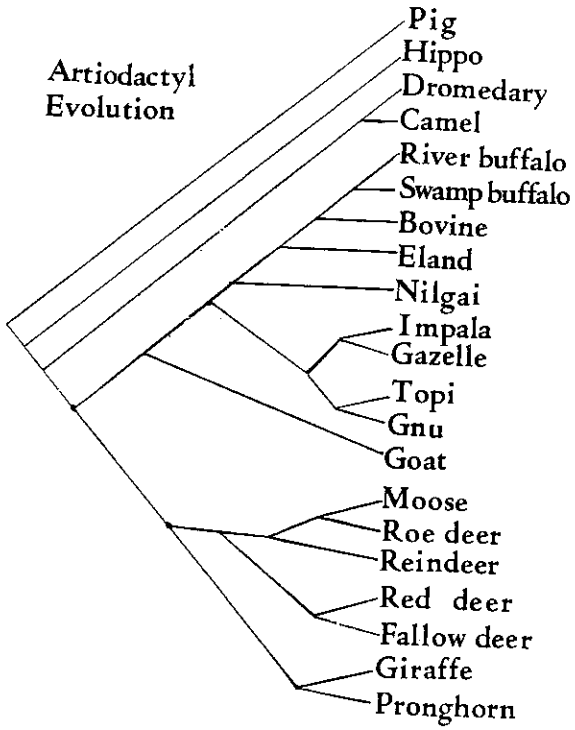

Figure 3. A hypothetical evolutionary tree connecting species whose RNAses have been sequenced. Branch points represent ancestral RNAses with hypothetical sequences that may be deduced from the sequences of the modern enzymes.

In addition to a new rationale for studing structure-function relationships in proteins, we are implementing a new approach to permit rapid alteration of RNAse. This approach relies on a totally synthetic gene.(16) The redundancy of the genetic code is exploited in the design of the synthetic gene to introduce unique restriction sites at regular intervals. These restriction sites facilitate modification of the gene. To introduce a mutation, the gene is cut by digestion with the restriction endonucleases whose sites flank the region to be replaced. The gene is then religated with a segment of synthetic duplex DNA that codes for the desired mutation.

The following criteria were used to construct a sequence best designed to facilitate rapid and controlled alteration:

1. Sites for commercially available restriction endonucleases should be distributed at intervals in the sequence.

2. These restriction sites must be unique; no restriction site should appear more than once in the gene.

3. Repetitions of sequences (or complementary sequences) longer than four base pairs should be eliminated.

4. Provisions should be made for the cloning of the gene into a cloning vector between EcoRl and BamHl sites, and into expression vectors between EcoRl and HindIII sites.

The sequence of the gene for RNAse is in Figure 4.

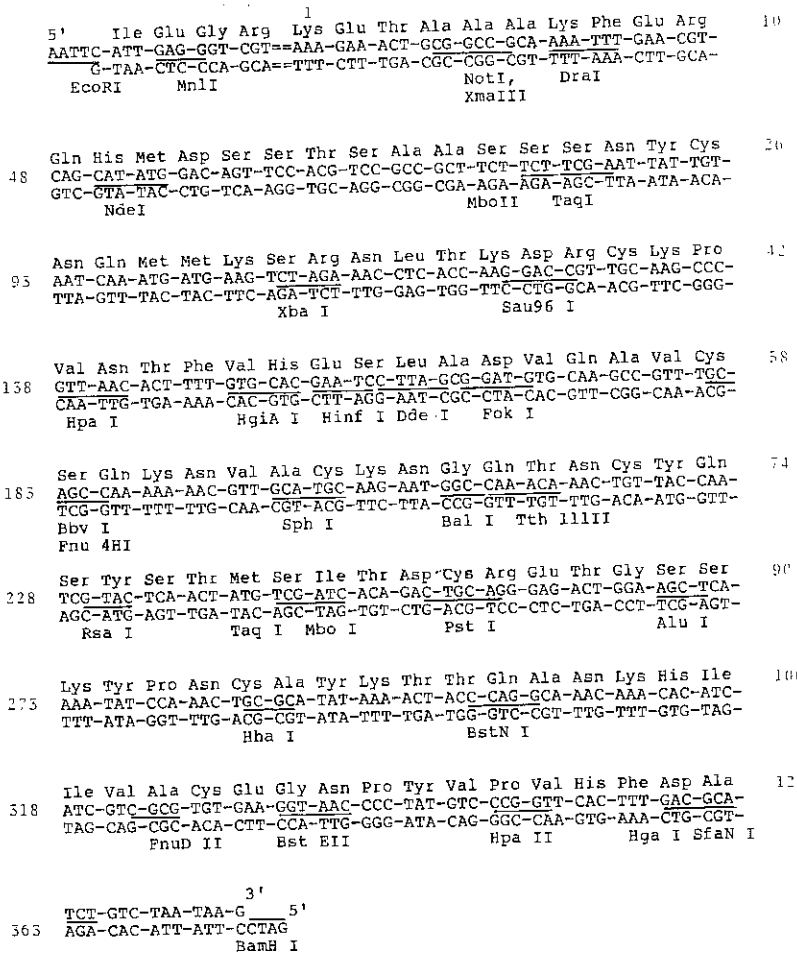

Figure 4: Sequence of one synthetic gene coding for RNAse

Oligonucleotides were synthesized using phosphoramidite reagents to elongate nucleotide chains attached to solid supports.(17) While most of the synthetic work was done by hand, more recent synthesis has been performed on an Applied Biosystems automated DNA synthesizer. Synthetic oligonucleotides were then released from the supports and purified by thin layer chromatography. Purified fragments were sequenced by the Maxam-Gilbert method (18) prior to subsequent ligation. Oligomers were then enzymatically ligated to give the assembled gene.

The synthetic gene was incorporated into a plasmid, and the plasmid was introduced into E. coli via the "calcium shock" method. We adapted a procedure of Hanahan's (19), in which the dimethylsulfoxide was replaced by sulfolane. Transformants were selected for ampicillin resistance, and the plasmids within them were identified by restriction digestion and sequencing.

We have constructed several genes using this general method. The first coded for the ribonuclease S-protein, the proteolytic fragment of RNAse that consists of residues 21-124.(15) Subsequent constructions code for the complete RNAse protein.

As often the case with the production of eukaroytic proteins cloned in E. coli, one challenge is to express the cloned gene. RNAse presents some special problems in this regard, as an active ribonuclease may digest the messenger RNA that codes for its production. While a complete description of the work that we have done cannot be presented here, I would like to show one example of how the protein may be expressed.

One of the constructions prepared (shown in Figure 4) was designed to create a "fusion protein" in which RNAse is covalently fused to beta-galactosidase, a protein found naturally in the expression host, E. coli. The two proteins are joined by a short peptide that is the recognition sequence for Factor Xa, a protease that cleaves after the sequence Ile-Glu-Gly-Arg. Nagai and Thogersen (20) have suggested this protease as generally useful for cleaving fusion proteins.

An SDS electrophoresis gel of the RNAse fusion protein expressed in E. coli shows a band at the appropriate molecular weight. The fusion protein is expressed in substantial quantities, and is found primarily as inclusion bodies within the E. coli cell. The band is identified both by molecular weight, and by "Western blotting" with radioactive antibodies against RNAse.

4. CONCLUSION

We have presented evidence that natural selection has finely tuned the properties of enzymes so as to engineer in those proteins the behaviors that contribute most to the survival of the host organism. Using this as a hypothesis, we have introduced a new strategy for the use of recombinant DNA methods to study structure-function relationships in proteins. Using both synthetic and enzymatic methods, we have synthesized a gene coding for RNAse, a gene that is specially designed to permit rapid alteration. This gene has been cloned and expressed in E. coli.

We are now prepared to retrace the recent evolution of RNAse, using the hypothetical evolutionary tree deduced from the amino acid sequences of the RNAses that are the products of the evolutionary divergence. We hope that by retracing this tree, we will be able to gain insight into the process by which nature has engineered the properties of RNAse. We are hopeful that our next report will contain evidence that this approach has been fruitful in allowing us to mimic the engineering done by Nature in our chemical laboratory.

ACKNOWLEDGEMENT

This work would have been impossible without the support of the National Science Foundation (CHE-8213267 and PCM-8111659), the National Institutes of Health (GM-30110), and many private organizations, including the Searle Foundation, Stuart Pharmaceuticals, Alfred Bader, the Dreyfus Foundation, the Petroleum Research Fund, the Research Corporation, the Clay Science Fund, the Anna Fuller Fund, the Milton Fund, Owens Corning Fiberglass, the Xerox Corporation, the Harvard Society of Fellows, and a gift in memory of Baron Langer von Langensdorf. S. B. is a Fellow of the Alfred Sloan Foundation.

REFERENCES

(1)(a) Gillam, S.; Waterman, K.; Smith, M. Nucl. Acids Res., 1975, 2, 625. (b) Dalbadie-McFarland, G.; Cohen, L. W.; Riggs, A. D.; Morin, C.; Itakura, K.; Richards, J. H. Proc. Nat. Acad. Sci., 1982, 79, 6409-6413.

(2)(a) Wallace, R. B.; Schold, M.; Johnson, M. J.; Dembek, P.; Itakura, K. Nucl. Acid Res., 1981, 9, 3647. (b) Inouye, S.; Soberon, X.; Franceschini, T.; Nakamura, K.; Itakura, K.; Inouye, M. Proc. Nat. Acad. Sci., 1982, 79, 3438. (c) Sigal, I. S.; Harwood, B. G.; Arentzen, R. Proc. Nat. Acad.

Sci., 1982, 79, 7159. (d) Winter, G.; Fersht, A. R.; Wilkinson, A. J.; Zoller, M.; Smith, M. Nature, 1982, 299, 756. (e) Villafranca, J. E.; Howell, E. E.; Voet, D. H.; Strobel, M. S.; Ogden, R. C.; Abelson, J. N.; Kraut, J. Science, 1983, 222 782-788.

(3)(a) Kimura, M. Nature, 1968, 217, 624. (b) Jukes, T. H. Science, 1980, 210, 973. (c) Kimura, M. The Neutral Theory of Molecular Evolution, N.Y., 1983, Cambridge University Press.

(4)(a) You, K. Meth. Enzymol., 1982, 87, 101-126. (b) You, K.; Arnold, L. J.; Allison, W. S.; Kaplan, N. O. Trends in Biochem. Sci., 1978, 3, 365-368.

(5)(a) Benner, S. A. " Experientia, 1982, 38, 633-37. (b) Benner, S. A.; Stackhouse, J. Chemical Approaches to Understanding Enzymic Catalysis, Proceedings of the 26th OHOLO Conference, Zichron Yaacov, Israel, B. S. Green, Y. Ashani, and D. Chipman, editors, 1982, Elsevier, 32-43. (c) Nambiar, K. P.; Stauffer, D. M.; Kolodziej, P. A.; Benner, S. A. J. Am. Chem. Soc., 1983, 105, 5886-5890.

(6) Garavito, R. M.; Rossmann, M. G.; Argos, P.; Eventoff, W. Biochem., 1977, 16, 5065-5071.

(7) Stackhouse, J.; Nambiar, K. P.; Burbaum, J.; Stauffer, D. M.; Benner, S. A. J. Am. Chem. Soc., 1985, 107, 2757-2763.

(8) Benner, S. A.; Nambiar, K. P.; Chambers, G. K. J. Am. Chem. Soc., 1985, in press.

(9) Bentley, R. Molecular Asymmetry in Biology, 1970, 2, 1-89.

(10) Cahn, R. D.; Kaplan, N. O.; Levine, L.; Zwilling, E. Science, 1961, 136, 962-969.

(11) Gutfreund, H. Prog. Biophys. Mol. Biol., 1975, 29, 161-195.

(12) Barnard, E. A. Nature, 1969, 221, 340.

(13) J. J. Beintema and J. A. Lenstra, in Macromolecular Sequences in Systematics and Evolutionary Biology, M. Goodman, Ed., (Plenum, N. Y., 1981).

(14) DiDonato, A.; D'Alessio, G. Biochem. Biophys. Res. Comm., 1973, 55, 919-928.

(15) Vescia, S.; Tramontano D. Augusti-Tocco, G.; D'Alessio, G. Cancer Res., 1980, 40, 3740-3744.

(16) Nambiar, K. P.; Stackhouse, J.; Stauffer, D. M.; Kennedy, W. G.; Eldredge, J. K.; Benner, S. A. Science, 1984, 222, 1299-1301.

(17) Beaucage, S. L.; Caruthers, M. H. Tetrahedron Lett., 1981, 22, 1859-1862.

(18) Maxam A. M.; Gilbert, W. Meth. Enzymol, 1980, 65, 499.

(19) Hanahan, D. J. Mol. Biol., 1983, 166, 557-580.

(20) Nagai, K.; Thogersen, H. C. Nature, 1984, 309, 810-812.