Acta Chemica Scandinavica, 1996: 50: 243–248
Printed in Belgium – all rights reserved

ACTA
CHEMICA SCANDINAVICA
ISSN 0904-213X

Developing New Synthetic Catalysts. How Nature Does It†

Steven A. Benner,* Thomas M. Jermann, Jochen G. Opitz, Sun Ai Raillard, Todd R. Zankel, Katrin Trautwein-Fritz, Joseph Stackhouse, Mauro I. Ciglic, Monika Haugg, Nathalie Trabesinger-Rüf and Elmar G. Weinhold

Department of Chemistry, ETH, CH-8092 Zürich, Switzerland

Benner, S. A., Jermann, T. M., Opitz, J. G., Raillard, S. A., Zankel, T. R., Trautwein-Fritz, K., Stackhouse, J., Ciglic, M. I., Haugg, M., Trabesinger-Rüf, N. and Weinhold, E. G., 1996. Developing New Synthetic Catalysts. How Nature Does It. – Acta Chem. Scand. 50: 243–248 © Acta Chemica Scandinavica 1996.

Paleomolecular biochemistry is a new field of science that seeks to understand how life emerged and developed in interaction with its geophysical surroundings. It is an experimental science, involving reconstruction of extinct biomolecules in the laboratory, studying their properties in the laboratory, and inferring details of their behavior and function in the context of geological data. An outline is provided of some tools of this field, together with its application to the study of two specific systems, ribonuclease and alcohol dehydrogenase.

An important goal in contemporary bioorganic chemistry is to invent new molecules that catalyze specific chemical transformations. The invention of such molecules has, of course, occurred often on earth over the past 4 billion years. An understanding of how living systems developed new biological catalysts is useful to understand the catalysts themselves.

It has now been more than a decade since we began to explore in depth the processes by which living systems create new catalysts. This work has yielded investigations of many aspects of living systems, and new ideas in many areas. A few of these are listed below: (a) predicting the structure of proteins from sequence data;[1] (b) organizing the genome sequence database;[2] (c) understanding stereospecificity, kinetics, and other details of enzymatic reactions;[3] (d) redesigning nucleic acids.[4]

This paper concerns a topic that is more general, in fact, which extends over all of these specific research projects. It goes under what we have come to call in Zurich *paleomolecular biochemistry*.

In its broadest sense, paleomolecular biochemistry seeks to unravel the biological, chemical and geological events that have shaped the history of life on earth. However, the interaction is reciprocal. Planet earth and life on earth have evolved together. While geological and geochemical processes create and define the conditions necessary for life, living systems in turn shape the geophysical environment. An experimental approach towards understanding these two systems together not only has scientific value, but is central to our survival as a species. Further, we cannot hope to understand fully life in its various forms, its metabolism, how it regulates this metabolism, and how it expresses itself in multicellular forms, without understanding how it emerged.

How can understanding the history of life on the planet help us understand biological macromolecules? To ask this question requires that we first be able to reconstruct this history at a molecular level. Here, we gain little help from fossils, although we might hope that the first efforts to extract functioning biomolecules from geological sediments might lead to a more extensive means of obtaining information about ancient biomacromolecules. Instead, we draw on an analogy from historical linguistics, where extinct languages are reconstructed from the structure of their descendent languages. Figure 1 illustrates this process, showing how the Proto-Indoeuropean word 'snow' can be reconstructed from the homologous words for

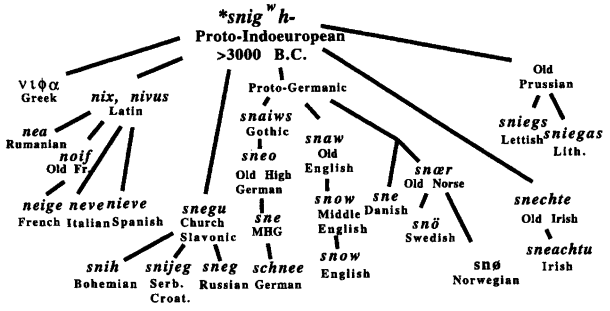

Fig. 1. Reconstructing the extinct word for 'snow'.

† Contribution presented at the Nobel Symposium on *Catalytic Asymmetric Synthesis*, September 3–7, 1995 at Tammsvik, Bro, Sweden.

* To whom correspondence should be addressed.

'snow' in its descendent languages, such as German, Latin, Greek, Slavic and Sanskrit.

This reconstruction is interesting because it tells us (approximately) what the Proto-Indoeuropean would say to refer to snow. This is at one level only an item of trivia. However, the reconstruction says something about Proto-Indoeuropean civilization. In particular, because the Proto-Indoeuropean language had a word that means 'snow', the Proto-Indoeuropean civilization must have been located in a place where it snowed.

Indeed, a rather complete picture can be built describing the Proto-Indoeuropean society from its language. It was less complex than modern society (no gold or silver). But the society was not 'primitive'. It had dogs (*kwón-), horses (*ékwo-), sheep (*H_3éwi-), pigs (*suH-) and cows (*g^wów-), and which ate grain (*yewo-) from vehicles (*wogho-) with wheels (*k^wek^{w}lo-) pulled by animals wearing yokes (*yugo-). Proto-Indoeuropeans fermented honey (*melit-) to make mead (*medhu). They could count to 100 (*kmtóm).

Thirty years ago, Pauling and Zuckerkandl pointed out that the structures of ancient biomolecules (in particular proteins) found in organisms that are now extinct, might be deduced by an analogous evolutionary analysis.[5] The simplest approach is a rule of 'parsimony', where the ancient sequence is reconstructed so that the derived sequences can be obtained with the smallest number of independent evolutionary events.

The key element of an evolutionary analysis is the sequence alignment,[6,7] which attempts to represent the evolutionary relationship between two protein sequences by placing them side by side so that their similarities are as perspicuous as possible (Fig. 2). With an appropriate definition of 'similarity', this maximizes the probability that amino acids paired in the alignment are descendants of single codons in a single ancestral gene.

As when reconstructing ancient languages, it helps if rules of transformation are systematized. With languages, sounds in words tend to evolve with patterns, not randomly. Likewise, amino acid substitutions are not random. By tradition, the non-randomness is described by giving each matched pair of amino acids a score.[8] These pairwise scores are extracted from an empirical scoring matrix, a table that represents the relative probabilities of each of the 210 pairings between the 20 natural amino acids. Gaps in the alignment are penalized, the terms summed, and the resulting score reported. The evolutionary distance between the two sequences is then measured in PAM units, the number of *p*oint *a*ccepted *m*utations

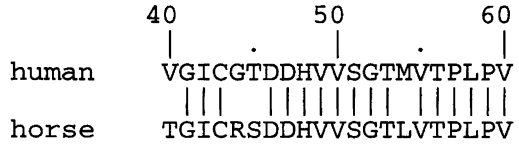

Fig. 2. A segment of an alignment of alcohol dehydrogenase (Adh) sequences.

that the two protein sequences have suffered (per 100 amino acids) since they diverged an unspecified number of aeons ago. The most advanced scoring tools for medium-range sequence divergence, based on an empirical analysis of the entire sequence database, are now available.[9]

The fact that amino acid substitutions are not, in general, random reflects the fact most proteins have a function that contributes to the ability of their host organism to survive, select a mate, and reproduce. To perform this function, proteins adopt a fold, or tertiary structure, a structure that is conserved much more highly than the sequence itself.[10] Function therefore constrains divergent evolution.[11] This implies that intermixed with substitutions that represent neutral drift in the structure,[12,13] an alignment carries information about the protein's conformation. This is the reason why a detailed analysis of patterns of conservation and variation in protein sequences can be used to predict the three-dimensional structure of proteins from sequence data.[1]

Given sequences of two descendents and transformation rules, we can now attempt to reconstruct the most probable sequence for the most recent common ancestor of the two proteins. For example, a parsimony analysis allows us to deduce much (but not all) of the sequence of the alcohol dehydrogenase in the ancestor of man and horse (Fig. 3). Where the two descendents have the same amino acid, the ancestral amino acid is most probably constructed to be the same as in the descendents. Where the two descendents differ, one cannot say what the ancestral amino acid was; the reconstruction is ambiguous.

It is helpful to illustrate this with two positions drawn from the alignment in Fig. 3 (Fig. 4). The reconstruction at position 53 is easy to make. Any assignment other than Thr at position 53 in the ancestor would require that the tree embody at least two point mutations. This is less probable than a tree that embodies no point mutations.

At position 54, the reconstruction is not so clear. Placement of a Met at position 54 in the ancestral sequence yields a tree with one point mutation. So does placement of a Leu in the ancestral sequence. Thus, one reconstruction cannot be preferred over another. The ancestral sequence is ambiguous at this point.

To resolve ambiguities in the reconstructed ancient proteins, one needs more sequence information from ho-

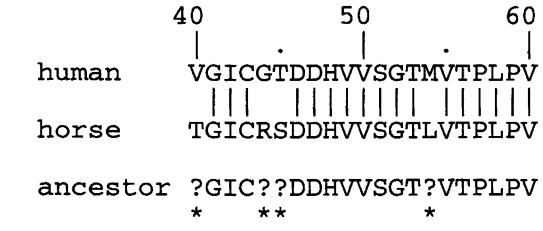

Fig. 3. Reconstruction of the sequence of alcohol dehydrogenase in the most recent common ancestor of horse and man.

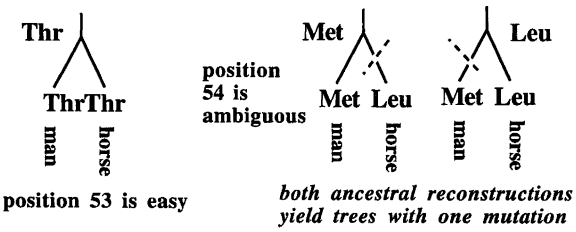

Fig. 4. Reconstruction of residues 53 and 54 in alcohol dehydrogenase in the most recent common ancestor of man and horse.

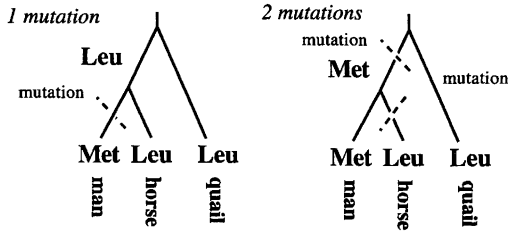

Fig. 6. Reconstruction of residue 54 in alcohol dehydrogenase in the most recent common ancestor of man and horse with information from quail alcohol dehydrogenase.

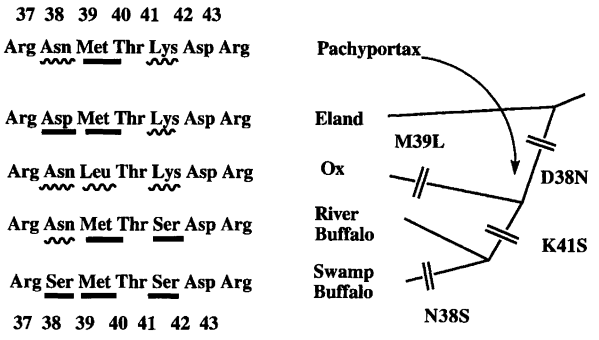

Fig. 7. Reconstruction of a segment of the sequence of ribonuclease (RNase) in the most recent common ancestor of ox, river buffalo and swamp buffalo.

mologous proteins from other sources. For example, addition of the alignable segment of the alcohol dehydrogenase from quail permits a reconstruction of the sequence of the alcohol dehydrogenase in the most recent common ancestor of horse and man that is unambiguous at all positions (Fig. 5). For example, at position 54, the most parsimonious reconstruction of the ancestral Adh is Leu (Fig. 6). The reconstruction on the left (with one mutation), where position 54 is a Leu, is more probable than the reconstruction on the right (with two mutations), where position 54 accommodates a Met.

Given a reconstruction of an ancient sequence, we can begin experimental work. Through the magic of site-directed mutagenesis, we can now make the ancient protein from the extinct organism in the laboratory and study it. As with Proto-Indoeuropean, we are often able to deduce from the protein's properties something about the ancient environment and function as well.

Let us go back in time by exploring ancient ribonucleases (RNases) from artiodactyls, an order of mammals with a cloven hoof that includes the pig and hippopotamus, camel, giraffe, deer and moose, goat, antelope, buffalo and ox. Thanks to extensive work by Beintema *et al.*, RNase is a paradigm of molecular evolution.[14]

Our first step takes us back only a few million years, to the most recent common ancestor of ox, river buffalo, and swamp buffalo (Fig. 7). The organism containing this ancient RNase corresponds approximately to the fossil bovid *Pachyportax*, although the standard caveat, that no fossil remnant is likely to correspond precisely to any particular branch point in an evolutionary tree, must be stated.[15,16]

```
                          54
                          |
human     VGICGTDDHVVSGTMVTPLPV
          |||  |||||||| ||||||
horse     TGICRSDDHVVSGTLVTPLPV

quail     TGICRSDDHVVTGALAMPFPV

ancestor  TGICRSDDHVVSGTLVTPLPV
                          *
```

Fig. 5. With the sequence of the alcohol dehydrogenase from quail, the reconstruction of the sequence of alcohol dehydrogenase in the most recent common ancestor of horse and man becomes unambiguous.

Pachyportax RNase is stable to proteolysis and active against low molecular weight, single stranded RNA substrates (Table 1). These features are expected for an RNase found in the digestive tract, and are similar to those seen with RNase A. Thus, it suggests that the ancient RNase was found in the digestive tract of this extinct artiodactyl.

Going back further in time, RNases from the more ancient artiodactyls *Pachyportax* and *Eotragus* also behave as digestive enzymes. They are stable to trypsin, act on small RNA substrates (UpA), and have low catalytic activity against double stranded RNA (Table 2).

This is not so, however, for RNases for the more ancient artiodactyls *Archaeomeryx* and *Diacodexis*. RNases from *Archaeomeryx* and *Diacodexis* are react more slowly with small substrates and faster with double stranded RNA, and are less stable to proteolysis (Table 3). This is not the behavior of a digestive enzyme. This suggests that

Table 1. Kinetics of RNase from Pachyportax (UpA).

RNase	k_{cat}/s^{-1}	$K_M/\mu M$
Bos taurus	1459+50	211+20
A19S	2055+495	250+20
A19S	1906+301	229+38
L35M	2003+107	208+18
L35M	1602+102	169+18
Pachyportax	1944+136	203+27

Table 2. Kinetics of RNases with UpA as the substrate.

Enzyme	$K_M/\mu M$	k_{cat}/s^{-1}	$(k_{cat}/K_M)/M^{-1}$
Bos taurus	165	1900	11.5×10^6
Pachyportax	190	1944	10.2×10^6
Archaeomeryx	420	468	1.1×10^6
Diacodexis	177	253	2.2×10^6

Table 3. Relative rate of hydrolysis of polyA-polyU by ancient RNases (Bos taurus = 1).

Bos taurus	1.0
Pachyportax	1.2
Achaeomeryx	5.2
Diacodexis	4.6

the digestive function of the RNase arose sometime between *Diacodexis* and *Eotragus*.

Is this possible? We must now return to the geological sciences, and ask not only about the fossil record describing the evolution of artiodactyls, but also the geological environment in which these animals arose. In Fig. 8, we compare the tree describing the evolution of the RNase molecule with the emergence of new biological function, deduced from the fossil record.

The fossil record shows that ruminant artiodactyls, which digest cellulose in a forestomach, arose from non-ruminant artiodactyls 40 million years ago. Interestingly, this corresponds to the time when digestive behavior first appears in this lineage of RNases. It thus appears that ruminants invented a digestive RNase at the same time that rumination itself was invented, implying that the molecular evolution parallels physiological evolution.

This is consistent with a physiological hypothesis advanced 25 years ago by Barnard.[17] Ruminants feed grass to bacteria in the first stomach, and then eat fresh bacteria. Ca. 20% of the nitrogen intake of the ruminant

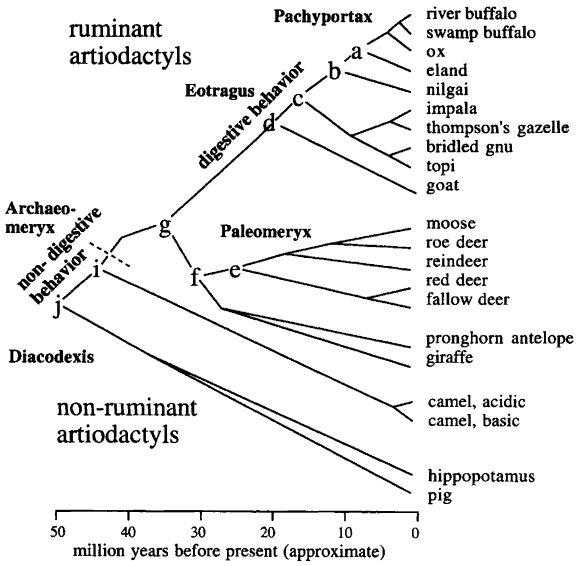

Fig. 8. The evolutionary tree for the RNase A family (adapted from Beintema *et al.*).

comes as RNA. Digestive RNases are needed to break this RNA down. Non-ruminants do not ferment in a forestomach. Rather, fermentation takes place in a cecum following the main stomach. While a digestive RNase is possibly required in cecal fermenters as well, it would most likely evolve under different functional constraints.

Ruminant artiodactyls have been enormously successful in competition with non-ruminant herbivores. Some 170 species of ruminant herbivore have now displaced some 250 species of non-ruminant herbivore (perissodactyls) from the Miocene, reducing these to just three species groups (horse, rhino and tapir). This, incidentally, is a mass extinction event (99% of the dominant mammalian order of non-ruminant herbivores disappeared), far more dramatic than (for example) the extinction of the dinosaurs.

This observation is important. It appears as if innovation in a competing mammalian order caused a massive extinction, not a cosmogenic event (for example, the falling of a meteorite). We suspect that such extinction events have happened frequently. For those of you who are interested in dinosaurs (and who is not?), it is not likely that similar events explain the extinction of dinosaurs. When mass extinctions occur through innovation in a species, the primary surviving lineages are descendents of the species that made the innovation. The other surviving species are generally those that have adapted to fill ecological niches created by the innovator and his/her descendents. With the dinosaurs, most of the mammalian lineages that became dominant in the Tertiary period had diverged before the mass extinction of the dinosaur. It is unlikely that all of these mammalian lineages made the same innovation in parallel.

Returning to our travels back in time, we might ask what the RNase did before becoming a digestive RNase? It is not clear, but an RNase in the brain has been sequenced, which diverged at approximately the same time as ruminant digestion emerged (Fig. 9). The digestive RNase may have evolved from a brain RNase.

This example shows how the paleomolecular biochemist can learn something about the origin of new physiological function through changes in protein sequence. We now must return now to the theme of this conference. Can we use evolutionary analysis to design new catalysts? For example, alcohol dehydrogenases from horse and yeast differ in rates of turnover. The activity of the Adh from horse is 100 times lower with ethanol as a substrate than the Adh from yeast.[18] Horse Adh accepts many substrates, including secondary alcohols and primary alcohols with branching at the 2, 3 and 4 positions. Yeast Adh essentially accepts ethanol alone.[19] The ideal catalyst for a synthetic organic chemist, of course, would be a fast, non-specific alcohol dehydrogenase.

Again, we return to an evolutionary analysis (Fig. 10). A tree suggests that a slow, non-specific alcohol dehydrogenase evolved from a fast, non specific enzyme. Can we use evolutionary reconstructions to understand better how, and then to design a fast, less specific alcohol de-

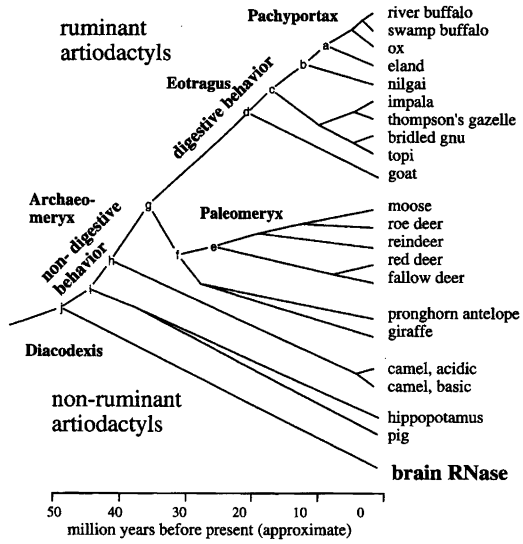

Fig. 9. The evolutionary tree for the RNase A family showing a hypothetical point of divergence of brain RNase.

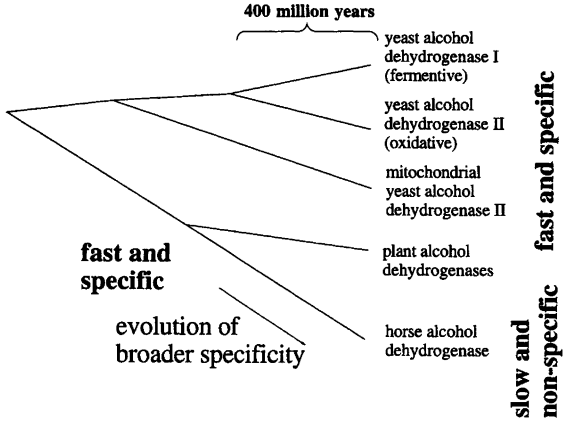

Fig. 10. The evolution of alcohol dehydrogenases.

```
                    54
                    |
human     VGICGTDDHVVSGTMV--TPLPV
          |||  |||||||| |  |||||
horse     TGICRSDDHVVSGTLV--TPLPV

quail     TGICRSDDHVVTGALA--MPFPV

ancestor  TGICRSDDHVVSGTLV--TPLPV

yeast     SGVCHTDLHAWHGDWPLPTKLPL
```

Fig. 11. Part of an alignment of alcohol dehydrogenase sequences.

hydrogenase? We include the sequence of the alcohol dehydrogenase from yeast in the alignment (Fig. 11), and consider again residue 54.

Residue 54 is near an insertion in the protein.[21] In yeast, position 54 holds a large Trp (W). Replacement by Leu, the amino acid found in the alcohol dehydrogenase from horse, together with the insertion, should open up the active site to allow yeast Adh to accept longer primary alcohols, and primary alcohols with 4-position branching. This is in fact the case, both with long-chain primary alcohols (Table 4) and branched-chain primary alcohols (Table 5).[21,22]

Table 4. Native, Trp54Leu and Trp54Met variants of yeast Adh. k_{cat}/K_m relative to ethanol for primary alcoholics.

	Ethanol	Propanol	Butanol	Pentanol	Hexanol
Adh 1[a]	100	22.0	4.7	3.9	8.5
Adh 1[b]	100	11.9	2.9	1.3	2.8
W54M[a]	100	5.2	6.8	9.0	1.4
W54L[b]	100	8.6	14.5	20.4	48.0

[a] Green *et al.*: 83 mM potassium phosphate buffer, pH 7.3, 40 mM KCl, 0.25 mM EDTA, and 10 mM NAD^+ at 30°C.
[b] Weinhold *et al.*: 32 mM sodium pyrophosphate buffer, pH 8.2, 100 mM Na_2SO_4, and 10 mM NAD^+ at 25°C.

We hope that we have presented a case for paleomolecular biochemistry as an experimental field for study that has both scientific interest and technological uses. What we want you to take home with you is the fact that to understand biological catalysts, it is helpful to understand how they emerged.

We should mention that this work was greatly influenced by five scientists, each of whom broke new ground. First, of course, is Linus Pauling and Emil Zuckerkandl. Without the penetrating analysis of Hans Jörnvall and his coworkers, we would not have penetrated so far in the alcohol dehydrogenase field as we have done. Jaap Beintema and his group have, of course, done a remarkable job developing the RNase system as a paradigm in molecular evolution; it was their work that got us started in 1980 in this area. Finally, we must acknowledge Allan Wilson, who with us was the first to prepare in the laboratory an ancient protein whose sequence was reconstructed by parsimony methods.[23]

Table 5. Impact of the Trp54Leu mutation on the V_{max} and K_m values of alcohol dehydrogenase (ratios relative to native Adh).[a]

Alcohol substrate	V_{mut} / V_{wt}	$(K_m)_{mut}$ / $(K_m)_{wt}$	$(V/K)_{mut}$ / $(V/K)_{wt}$
Ethanol	0.33	3.35	0.10
1-Propanol	0.23	3.23	0.07
1-Butanol	0.28	0.55	0.50
1-Pentanol	0.33	0.21	1.58
1-Hexanol	0.81	0.47	1.72
4-Methyl-1-pentanol	2.78	0.12	24.3
Cinnamyl alcohol	0.74	0.33	2.3

[a] V_{mut} = maximal volocity of mutant; V_{wt} = maximal velocity of wild type; $(K_m)_{mut}$ = Michaelis constant of mutant; $(K_m)_{wt}$ = Michaelis constant of wild type.

References

1. Benner, S. A., Gerloff, D. L. and Jenny, T. F. *Science 265* (1994) 1642.
2. Gonnet, G. H., Cohen, M. A. and Benner, S. A. *Science 256* (1992) 1443.
3. Benner, S. A. and Ellington, A. D. *Bioorg. Chem. Front. 1* (1990) 1.
4. Piccirilli, J. A., Krauch, T., Moroney, S. E. and Benner, S. A. *Nature (London) 343* (1990) 33.
5. Pauling, L. and Zuckerkandl, E. *Acta Chem. Scand. 17* (Suppl. 1) (1963) S9.
6. Zuckerkandl, E. *Sci. Am. 212* (5) (1965) 110.
7. Doolittle, R. F. Ed., *Molecular Evolution, Computer Analysis of Protein and Nucleic Acid Sequences, Meth. Enzymol. 183* (1990).
8. Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. In: Dayhoff, M. O., Ed., *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3, pp. 345–352, National Biomedical Research Foundation, Washington DC, 1978.
9. Cohen, M. A., Benner, S. A. and Gonnet, G. H. *Prot. Eng. 7* (1994) 1323.
10. Chothia, C. and Lesk, A. *EMBO J. 5* (1986) 823.
11. Benner, S. A. and Ellington, A. D. CRC *Crit. Rev. Biochem. 23* (1988) 369.
12. King, J. L. and Jukes, T. H. *Science 164* (1969) 788.
13. Kimura, M. In: Kimura, M., Ed., *Molecular Evolution, Protein Polymorphism, and the Neutral Theory*, pp. 3–56, Springer-Verlag, Berlin, 1982.
14. Beintema, J. J., Fitch, W. M. and Carsana, A. *Mol. Biol. Evol. 3* (1986) 262.
15. Stackhouse, J., Presnell, S. R., McGeehan, G. M., Nambiar, K. P. and Benner, S. A. *FEBS Lett. 262* (1990) 104.
16. Jermann, T. H., Opitz, J. G., Stackhouse, J. and Benner, S. A. *Nature (London) 374* (1995) 57.
17. Barnard, E. A. *Nature (London) 221* (1969) 340.
18. Hadorn, M., John, V. A., Meier, F. K. and Dutler, H. *Eur. J. Biochem. 54* (1975) 65.
19. Jones, J. B. and Beck, J. F. In: Jones, J. B., Sih, C. and Perlman, D., Eds., *Applications of Biochemical Systems in Organic Chemistry Part 1*, Wiley, New York 1976, pp. 107–401.
20. Benner, S. A., Cohen, M. A. and Gonnet, G. H. *J. Mol. Biol. 229* (1993) 1065.
21. Weinhold, E. G. and Benner, S. A. *Prot. Eng. 8* (1995). *In press.*
22. Green, D. W., Sun, H.-W. and Plapp, B. V. *J. Biol. Chem. 268* (1993) 7792.
23. Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. and Wilson, A. C. *Nature (London) 345* (1990) 86.

Received September 1, 1995.