Commentary

The past as the key to the present: Resurrection of ancient proteins from eosinophils

Steven A. Benner*

Department of Chemistry, University of Florida, Gainesville, FL 32611-7200

Moving from gene sequences to protein behaviors, and from there to cellular function, organismal biology, and fitness, is an immediate challenge to biology in the postgenomic age. Genome databases are, of course, nothing more than collections of molecular structures for organic "natural products." Unfortunately, the chemical structure of an organic compound does not transparently reveal its behavior or its function in a biological system. This is true even if the molecular structure is a protein sequence.

However, protein sequences do contain information about their historical pasts (1). The similarities between sequences within a family of proteins can be used to construct an evolutionary tree that shows familial relationships. Ancestral sequences can be reconstructed by inference from descendant sequences. Dates can be placed on events in that molecular history. Events in the molecular history can be correlated with events in the geological and paleontological records. From this correlation has emerged a strategy for interpretive proteomics: perhaps if we understand a protein's past, we may be better able to understand its present (2).

A compelling illustration of this strategy is provided by Zhang and Rosenberg (3) in this issue of PNAS. These scientists examined a pair of proteins from the granules of eosinophilic lymphocytes. These proteins are paralogs; they arose by gene duplication some 30 million years ago (Ma) in an African primate that was ancestral to humans and Old World monkeys. The pair of proteins are relatives of digestive ribonuclease in artiodactyls, the mammal order containing ox, giraffe, deer, and antelope (4). This digestive ribonuclease was evidently created ≈40 million years ago, when ruminant digestion first emerged, to degrade the RNA from bacteria growing in the rumen (5).

Zhang and Rosenberg asked: why are these two gene duplicates present in physiology? The names of the two proteins, eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP), are not particularly informative about physiological function. The neurotoxicity of EDN is observed when it is injected into rat brains, hardly a natural physiological happening. The name "ECP" conveys only that the protein is cationic and comes from an eosinophil. Beyond this, past literature does not say much about what these proteins do.

Indeed, the function of eosinophils as cells remains enigmatic (6). We know much about their activities in allergic disease. Eosinophils are associated with asthma, infective wheezing, and eczema, for example (7), but cells do not exist to create diseases. When they function correctly, our own cells must contribute to our evolutionary fitness; helping us survive, select a mate, and reproduce. A current hypothesis suggests that eosinophils do this by defending us from outside agents, with allergic diseases arising as an undesired side effect. Beyond this, we can say very little.

Events in the molecular history can be correlated with events in the geological and paleontological records.

Earlier work by Zhang, Rosenberg, and their associates (6, 8) had already suggested that ECP and EDN might contribute to fitness in new and differentiated ways, some possibly associated with defense. ECP kills bacteria *in vitro*; EDN inactivates retroviruses *in vitro* (6). *In silico* analysis of reconstructed ancestral sequences in primates suggested that the proteins had suffered rapid sequence change near the time of the duplication that generated these two proteins, a change that might account for their differing behaviors *in vitro* (8). Ignoring, for the sake of discussion, some technical issues, this observation suggests that in primate evolution, mutant forms of EDN and ECP conferred more fitness than unmutated forms. This finding indicates that these proteins have roles, and that their roles were changing, adapting, and specializing during the episodes of rapid sequence evolution.

Zhang and Rosenberg have now examined these proteins by using experimental paleomolecular biochemistry, a strategy whereby ancient proteins are resurrected and studied in the laboratory (9, 10). To obtain a more densely articulated tree for the protein family, they sequenced additional genes from various primates. They used these sequence data to better reconstruct ancestral sequences for ancient EDN/ECPs. They estimated the posterior probabilities of these ancestral sequences by using Bayesian inference. Then, they resurrected these ancient proteins by cloning and expressing their genes, and studied their behavior in the laboratory.

Guiding their experimental work was the hypothesis that the antiretroviral activity of EDN might be related to the ability of the protein to cleave RNA. Studies of the ancestral proteins allowed Zhang and Rosenberg to retrace the origins of the antiretroviral and RNA-cleaving activities of EDN. Both the ribonuclease and antiviral activities of the last common ancestor of ECP and EDN, which lived ≈30 Ma, were low. Both activities increased substantially in the EDN lineage after its emergence by duplication. The paleomolecular reconstructions identified individual amino acids that were replaced during these episodes, suggesting hypotheses relating these behaviors to specific amino acids in the sequence. Through careful experimentation, Zhang and Rosenberg showed that two replacements (at sites 64 and 132) in the sequence were together required to increase these activities; neither alone was sufficient. Zhang and Rosenberg then closed the circle back to chemistry, analyzing the three-dimensional crystal structure of EDN to offer possible explanations for the interconnection between sites suffering replacement and the changes in behavior that these replacements created.

Zhang and Rosenberg then drew the principal conclusion from these data.

See companion article on page 5486.

*E-mail: benner@chem.ufl.edu.

www.pnas.org/cgi/doi/10.1073/pnas.092143699

Neutral theory in molecular evolution holds that most amino acid replacements in a protein sequence have little impact on the fitness of the host organism (11). The biochemical correlate of this theory holds that most replacements lack significant (for fitness) impact on behavior. In the EDN/ECP family, either of the two replacements at sites 64 and 132 individually had little impact on behavior, but each provides the context for the other to have a consequence. The notion that a "neutral" (perhaps better, behaviorally inconsequential) replacement might set the stage for a second adaptive replacement may be generally applicable.

These results will have broader impact. For example, virtually all analyses of divergent evolution treat protein sequences as if they were linear strings of letters (12). With this treatment, each site is modeled to suffer replacement independent of all others, future replacement at a site is viewed as being independent of past replacement, and patterns of replacements are treated as being the same at each site. This has long been known to be an approximation, useful primarily for mathematical and statistical analysis (the "spherical cow"). Understanding higher-order features of protein sequence divergence has offered *in silico* approaches to some of the most puzzling conundrums in biological chemistry, including how to predict the folded structure of proteins from sequence data (12), and how to assign function to protein sequences (13). The results of Zhang and Rosenberg provide an experimental case where higher order analysis is necessary to understand a biomolecular phenomenon.

Further, a textbook case illustrating another interpretive proteomics strategy (14) can be constructed from the results produced by Zhang and Rosenberg. This strategy identifies physiologically relevant *in vitro* behaviors for a protein where new biological function has emerged, as indicated by an episode of rapid (and therefore presumably adaptive) sequence evolution. The strategy examines the behavior of proteins resurrected from points in history before and after the episode of adaptive evolution. Those behaviors that are rapidly changing during the episode of adaptive sequence evolution, by hypothesis, confer selective value on the protein in its new function, and therefore are relevant to the change in function, either directly or by close coupling to behaviors that are directly relevant to function. The *in vitro* properties that are the same at the beginning and end of this episode are not relevant to the change in function.

Although the number of amino acids changing is insufficient to make the case statistically compelling, the rate of change in the EDN lineage is strongly suggestive of adaptive evolution (8). The antiviral and ribonucleolytic activities of the proteins before and after the adaptive episode in the EDN lineage are quite different. We conclude, therefore, that these activities are important to the emerging physiological role for EDN. This adds support, though perhaps only modest, for the notion that the antiviral activity of EDN became important in Old World primates $\approx$30 Ma.

We might consider the timing of the emergence of the ECP/EDN pair in Old World primates. Their divergence occurred in the Oligocene, a period of global trauma (15). The Oligocene saw the start of a global climatic deterioration that has continued until the present, with the Ice Ages in the past million years being the culmination (we hope) of this deterioration. Tropical rain forests receded, grasslands emerged, and the interactions between herbivores and their foliage changed. This change may have favored ruminant digestion, which provides another example of recruitment in the ECP/EDN superfamily. If EDN, ECP, and eosinophils are part of a defensive system, it is appropriate to ask what happened during the Oligocene that might have encouraged this type of system to be selected? Why might new defenses against retroviruses be needed at this time? Defenses against bacteria?

As for disease, perhaps recently emerging biomolecular systems do not function optimally, simply because they have not had the time to be optimized. The placental reproduction system, for example, may not yet be perfected by Darwinian processes, and therefore continues to be a site of biochemical experimentation (16, 17). Could this be the case with asthma, allergy, and eczema? Experimental paleobiochemistry with other proteins involved in these diseases, including tryptase and neutrophil myeloperoxidase, might be informative, and may provide insights that lead to improved treatments.

Shortly after the first paleomolecular resurrections were published, Nicholas Wade (18), writing in the *New York Times Magazine*, expressed a hint of displeasure with the enterprise. "The stirring of ancient artiodactyl ribonucleases," he wrote "is a foretaste of biology's demiurgic powers. It may well prove best to keep resurrection an unroutine event."

The contribution of Zhang and Rosenberg makes the opposite view compelling. Molecular resurrection in this case has contributed a piece of a puzzle. More will follow, as a scattering of Federal agencies begin to fund historically based biochemical research; the National Aeronautics and Space Administration Exobiology and Astrobiology programs come to mind. The past is the key to the present. When we understand where we came from, and how we got here, we understand better who we are. This cannot help but have profound and beneficial impact on health, the environment, and the human condition.

1. Pauling, L. & Zuckerkandl, E. (1963) *Acta Chem. Scand.* **17** Suppl. 1, S9–S16.
2. Benner, S. A. (2001) *Nature (London)* **409,** 459.
3. Zhang, J. & Rosenberg, H. F. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 5486–5491.
4. Barker, R. L., Loegering, D. A., Ten, R. M., Hamann, K. J., Pease, L. R. & Gleich, G. J. (1989) *J. Immunol.* **143,** 952–995.
5. Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. (1995) *Nature (London)* **374,** 57–59.
6. Rosenberg, H. F. & Domachowske, J. B. (2001) *J. Leukocyte Biol.* **70,** 691–698.
7. Onorato, J., Scovena, E., Airaghi, S., Morandi, B., Morelli, M., Pizzi, M. & Principi, N. (1996) *Ital. J. Pediatr.* **22,** 900–911.
8. Zhang, J., Nei, M. & Rosenberg, H. F. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3708–3713.
9. Stackhouse, J., Presnell, S. R., McGeehan, G. M., Nambiar, K. P. & Benner, S. A. (1990) *FEBS Lett.* **262,** 104–106.
10. Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. (1990) *Nature (London)* **345,** 86–89.
11. Kimura, M. (1984) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, New York).
12. Benner, S. A., Cannarozzi, G., Chelvanayagam, G. & Turcotte, M. (1997) *Chem. Rev.* **97,** 2725–2843.
13. Benner, S. A., Trabesinger-Ruef, N. & Schreiber, D. R. (1998) *Adv. Enzyme Regul.* **38,** 155–180.
14. Benner, S. A. (1997) *Applications of Protein Structure Predictions*, U.S. Patent Serial No. 08/914375.
15. Prothero, D. R. (1994) *The Eocene-Oligocene Transition. Paradise Lost* (Columbia Univ. Press, New York)
16. Benner, S. A. (1988) *FEBS Lett.* **233,** 225–228.
17. Di Donato, A., Cafaro, V., Romeo, I. & D'Alessio, G. (1995) *Protein Sci.* **4,** 1470–1477.
18. Wade, N. (1995) *New York Times Magazine* (March 26, 1995), pp. 35–36.