The resurrection of ribonucleases from mammals: from ecology to medicine

Slim O. Sassi and Steven A. Benner

18.1 Introduction

The family of proteins related to bovine pancreatic ribonuclease A (RNase A) provided the first biomolecular system to be analyzed using experimental paleogenetics. This protein family was chosen in 1979, as it was one of only three families of protein sequences that were sufficiently well represented in the then modest database for which one might consider doing experimental resurrections. Other families that were also considered included cytochrome *c*, which had been developed as a paradigm for molecular evolution by Margoliash (1963, 1964), and hemoglobin, which was studied as a model for biomolecular adaptation (Riggs, 1959; Bonaventura *et al.*, 1974). Cytochromes are substrates for cytochrome oxidases, and it was not considered possible to resurrect both ancestral cytochromes and their ancient oxidases. Hemoglobins are complicated to express, a problem solved only later. This left the RNases.

Fortunately, Jaap Beintema and his colleagues in Groningen had done an excellent job of sequencing (at the level of the protein) ribonucleases from a wide range of ruminants and closely related non-ruminant mammals (Beintema and Gruber, 1967, 1973; Gaastra *et al.*, 1974; Groen *et al.*, 1975; Welling *et al.*, 1975; Emmens *et al.*, 1976; Kuper and Beintema, 1976; Muskiet *et al.*, 1976; Vandenberg *et al.*, 1976; Vandijk *et al.*, 1976; Welling *et al.*, 1976; Gaastra *et al.*, 1978; Beintema *et al.*, 1979, 1984, 1985; Jekel *et al.*, 1979; Lenstra and Beintema, 1979; Beintema and Martena, 1982; Breukelman *et al.*, 2001). Done before the so-called age of the genome,

this work exploited classical Edman degradation of peptide fragments derived by selective cleavage of the protein. Such work required substantial amounts of protein, making convenient the large amount of RNase found in the digestive tracts of oxen and their immediate relatives. Beintema had also inferred the sequences of the ancestral proteins throughout the recent history of the digestive enzymes, using parsimony tools that adapted the ideas that Margaret Dayhoff had laid out (Dayhoff *et al.*, 1978).

18.2 Background

Members of the secreted RNase family of proteins are typically composed of a signal peptide of about 25 amino acids and a mature peptide of about 130 amino acids. Most members of the RNase family have three catalytic residues (one lysine and two histidines, at positions 41, 12, and 119 in RNase A, respectively). These come together in the folded enzyme to form an active site. In addition, RNases generally have six or eight cysteines that form three or four disulfide bonds. Except for these conserved residues, the sequences of RNases have diverged substantially in vertebrates, with sequence identities as low as 20% when comparing oxen and frog homologs (for example). RNase was well known as a digestive enzyme. As expected for enzymes found in the digestive tract, RNases are themselves biochemically robust. For example, the first step in the purification of RNase A involved the treatment of an extract from ox pancreas with 0.25 M sulfuric acid. This procedure precipitates

most other proteins and removes the glycosyl groups from RNase, but otherwise leaves the protein intact.

RNases proved to present several opportunities for biological interpretation and discovery. As digestive enzymes, pancreatic RNases lie at one interface between their host organisms and their changing environments, and are expected to evolve with the environment. Not all mammals, however, have large amounts of pancreatic RNase. In fact, RNase is abundant in the digestive systems primarily in ruminants (which include the oxen, antelopes, and other bovids, together with the sheep, the deer, the giraffe, okapi, and pronghorn) and certain other special groups of herbivores (Barnard, 1969).

In 1969, Barnard proposed that pancreatic RNase was abundant primarily in ruminants because ruminant digestion had a special need for an enzyme that digested RNA (Barnard, 1969). Ruminant digestive physiology is considerably different from human digestive physiology (for example). The ruminant foregut serves as a vat to hold fermenting microorganisms. The ox delivers fodder to these microorganisms, which produce digestive enzymes (including cellulases) that the ox cannot. The microorganisms digest the grass, converting its carbon into a variety of products, including low-molecular-mass fatty acids. The fatty acids then enter the circulation system of the ruminant, providing energy.

The ox then eats the microorganisms for further nourishment. According to the Barnard hypothesis, this digestive physiology creates a need for especially large amounts of intestinal RNase. The fermenting microorganisms are rich in ribosomes and rRNA, tRNA, and mRNA. Fermenting bacteria therefore deliver large amounts of RNA to the gastric region of the bovine stomach and the small intestine. Barnard estimated that between 10 and 20% of the nitrogen in the diet of a typical bovid enters the lower digestive tract in the form of RNA.

Barnard's hypothesis was certainly consistent with the high level of digestive enzymes in ruminants generally. For example, ruminants have large amounts of lysozyme active against bacterial cell walls in their digestive tracts. Was the Barnard hypothesis merely a just-so story, based on correlations that did not require causality or functional necessity? The first experimental paleogenetics program set out to test this.

The available sequences were adequate to support the inference, with little ambiguity, of the sequence of the RNase represented (approximately) by the fossil ruminant *Pachyportax* (Stackhouse *et al.*, 1990). This was also the case for the more ancient *Eotragus*, which lived in the Miocene. The available RNase sequences also permitted the inference, with only modest ambiguity, of sequences for RNases in the first ruminant, approximated in the fossil record by the genus *Archaeomeryx*. With slightly more ambiguity, the contemporary RNase sequences allowed the inference of the sequences of RNase in the first artiodactyl, the order of mammals having cloven hooves that includes the true ruminants as well as the camels, the pigs, and the hippos. This ancestor is approximately represented in the fossil record by the genus *Diacodexis*. A collaboration between Barbara Durrant at the Center for Reproduction of Endangered Species at the San Diego Zoo and the Benner laboratory yielded several additional sequences that assisted in making these inferences (TrabesingerRuef *et al.*, 1996).

Once the ancestral sequences were reconstructed, the Benner group prepared by total synthesis a gene for RNase that was specially designed to support the resurrection of ancient proteins (Nambiar *et al.*, 1984; Stackhouse *et al.*, 1990). From this gene, approximately two dozen candidate ancestral genes for intermediates in the evolution of artiodactyl ribonucleases were synthesized, cloned, and expressed to resurrect the ancestral proteins for laboratory study (Stackhouse *et al.*, 1990; Jermann *et al.*, 1995).

To assess whether reconstructions yielded proteins that were plausible as intermediates in the evolution of the RNase family, the catalytic activities, substrate specificities, and thermal/proteolytic stabilities of the resurrected ancestral RNases were examined. Most of the resurrected proteins, and all of those corresponding to proteins expected in artiodactyls living after *Archaeomeryx*, behaved as expected for digestive enzymes. This was especially apparent from their kinetic

properties (Table 18.1). Modern digestive RNases are catalytically active against small RNA substrates and single-stranded RNA (Blackburn and Moore, 1982). The RNase from *Pachyportax* was also, as were many of the earlier RNases. Thus, if one assumes that these catalytic properties are indicative of a digestive enzyme, these ancestral proteins were digestive enzymes as well.

This was also true quantitatively. Thus, the k_{cat}/K_m values for the putative ancestral RNases with the ribodinucleotide uridylyl $3' \rightarrow 5'$-adenosine (UpA) as a substrate (Ipata and Felicioli, 1968) in many ancient artiodactyls proved not to differ more than 25% from those of contemporary bovine digestive RNase (Table 18.1). With single-stranded poly(U) as substrate, the variance in catalytic activity was even smaller (18%).

Modern digestive RNases, like most digestive enzymes, are stable to thermal denaturation and cleavage by proteases. This suggested another metric for determining whether an ancestral protein acted in the digestive tract. Using a method developed by Lang and Schmid (1986), the sensitivity of the ancestral RNases to proteolysis as a function of temperature was measured (Table 18.2). Again, little change was observed in thermal stability of the ancestral RNases back to the ancestral artiodactyls approximated by

Archaeomeryx in the fossil record. The midpoints in the activity/temperature curves for these ancient proteins varied by only ±1.1°C when compared with RNase A. This can be compared with typical experimental errors of ±0.5°C.

Had all of the ancestral RNases behaved like modern RNases, the resulting evolutionary narrative would have had little interest. The experiments in paleogenetics became interesting because the behavior of RNases resurrected from organisms more ancient than the last common ancestor of the true ruminants (*Archaeomeryx* and earlier) did not behave like digestive enzymes using these metrics.

These more ancient resurrected ancestral RNases displayed a 5-fold increase in catalytic activity against double-stranded RNA (poly(A)-poly(U)). This is not necessarily a digestive substrate. Further, the ancestral RNases showed an increased ability to bind and melt double-stranded DNA. Bovine digestive RNase A has only low catalytic activity against duplex RNA under physiological conditions, and does not bind and melt duplex DNA; these activities are presumably not needed for a digestive enzyme. At the same time, the catalytic activity of the candidate ancestral sequences against single-stranded RNA and short RNA fragments, the kinds of substrate that are

Table 18.1 Kinetic properties of reconstructed ancestral ribonucleases

RNase	Ancestor of	k_{cat}/K_m (UpA $\times 10^6$)	Relative to RNase A		
			k_{cat}/K_m (%)	Poly(U)	Poly(A)-poly(U)
RNase A		5.0	100	100	1.0
a	Ox, buffalo, eland	6.1	122	106	1.4
b	Ox, buffalo, eland, nilgai	5.9	118	112	1.0
c	b, Gazelles	4.5	91	97	0.8
d	Bovids	3.9	78	86	0.9
e	Deer	3.6	73	77	1.0
f	Deer, pronghorn, giraffe	3.3	67	103	1.0
g	Pecora	4.6	94	87	1.0
h1	Pecora and seminal RNase	5.5	111	106	5.2
h2	Pecora and seminal RNase	6.5	130	106	5.2
i1	Ruminata	4.5	90	96	5.0
i2	Ruminata	5.2	104	80	4.3
j1	Artiodactyla	3.7	74	73	4.6
j2	Artiodactyla	3.3	66	51	2.7

RNase names refer to nodes in the evolutionary tree shown in Figure 18.1. All assays were performed at 25°C. UpA, uridylyl $3' \rightarrow 5'$-adenosine.

Table 18.2 Thermal transition temperatures for reconstructed ancient ribonucleases

Enzyme	T_m (°C)	ΔT_m (°C)
RNase A[a]	59.3	0.0
RNase A[b]	59.7	+ 0.4
a	60.6	+ 1.3
b	61.0	+ 1.7
c	60.7	+ 1.4
d	58.4	− 0.9
e	61.1	+ 1.8
f	58.6	− 0.7
g	59.1	− 0.2
h1	58.9	− 0.5
h2	59.3	0.0
i1	58.2	− 1.1
i2	58.7	− 0.6
j1	56.5	− 2.8
j2	57.1	− 2.2

Thermal unfolding/proteolytic digestion temperatures (±0.5°C) were determined by incubating the RNase ancestor in 100 mM sodium acetate (pH 5.0) in the presence of trypsin.

[a] Expressed in *Escherichia coli.*
[b] Boehringer Mannheim.

expected in the digestive tract, was substantially lower (by a factor of five) than in the modern proteins. Proposing that these behaviors can be used as metrics, Jermann *et al*. (1995) concluded that RNases in artiodactyls that were ancestral to *Archaeomeryx* were not digestive enzymes.

A similar inference was drawn from stability studies. The more ancient ancestors displayed a modest but significant decrease in thermal-proteolytic stability using the assay of Lang and Schmid. A less stable enzyme, and a lower activity against single-stranded RNA, for example, might imply simply that the incorrect amino acid sequence was inferred for the ancestral protein. The fact that catalytic activity against double-stranded RNA, and the ability to melt duplex RNA, was higher in the ancestors argued against this possibility.

The issue was probed further by considering the ambiguity in the tree. The connectivity of deep branches in the artiodactyl evolutionary tree is not fully clarified by either the sequence data or the fossil record (Graur, 1993). This created a degree of ambiguity in the ancestral sequences. To manage this ambiguity, Jermann *et al*. (1995) synthesized a variety of alternative candidate ancestral RNase sequences. These effectively covered all of the ambiguity in the tree topology, and the resulting ambiguity in the sequences. The survey showed that the measured behavior and the consequent biological interpretation were robust with respect to the ambiguity.

Site 38 proved to be especially interesting. The variant of h1 (Figure 18.1) that restores Asp at position 38 (as in RNase A) has a catalytic activity against duplex RNA similar to that of RNase A (Jermann *et al*., 1995; Opitz *et al*., 1998). Conversely, the variant of RNase A that introduces Gly alone at position 38 has catalytic activity against duplex RNA essentially that of ancestor h. These results show that substitution at a single position, 38, accounts for essentially all of the increased catalytic activity against duplex RNA in ancestor h.

The reconstructed amino acids at position 38 are unambiguous before and after the *Archaeomeryx* sequence. Thus, it is highly probable that the changes in catalytic activity against duplex RNA in fact occurred in RNases as the ruminant RNases arose. In one interpretation, catalytic activity against duplex RNA was not necessary in the descendent RNases, and therefore was lost. This implies that the replacement of Gly-38 by Asp in the evolution of ancestor g from ancestor h was neutral. Jermann *et al*. (1995) could not, however, rule out an alternative model, that Asp-38 confers positive selective advantage on RNases found in the ruminants.

18.3 Understanding the origin of ruminant digestion

The experimental paleobiochemical data within the pancreatic RNase family suggested a coherent evolutionary narrative consistent with the Barnard hypothesis. RNases with increased stability, decreased catalytic activity against duplex RNA, decreased ability to bind and melt duplex DNA, and increased activity against single-stranded RNA and small RNA substrates emerged near the time when *Archaeomeryx* lived. The properties that increased are essential for digestive function; the properties that decreased are not. *Archaeomeryx*
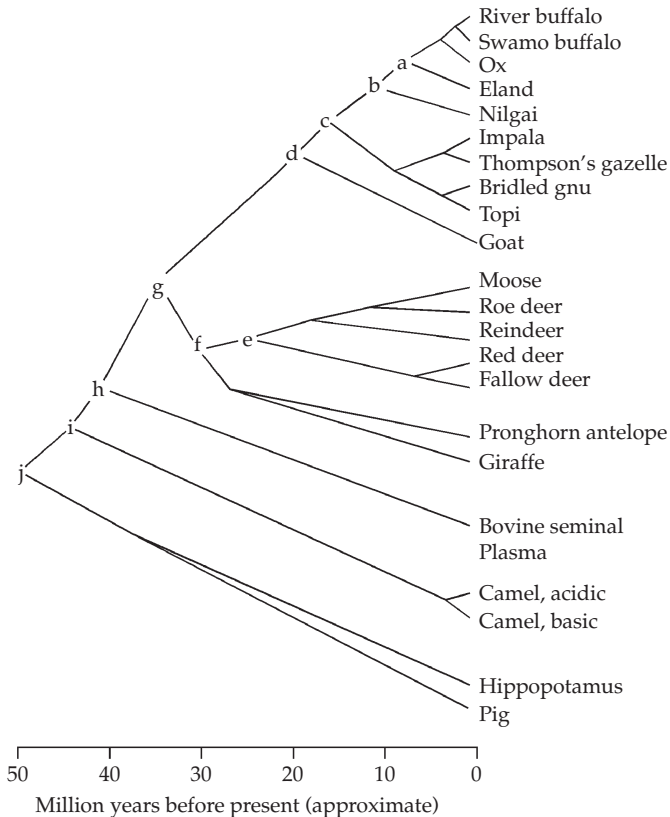
Figure 18.1 The evolutionary tree used in the analysis of ancestral pancreatic RNases. Lower-case letters at the nodes designate putative intermediates in the evolution of the protein family. Upper-case letters (D and G) indicate the residue at position 38 in the contemporary and reconstructed RNases. The time scale is approximate. The tree was adapted from Beintema *et al.* (1988) with a single alteration to join the pig and the hippopotamus together in a separate subfamily that branches together from the main line of descent. In the Beintema–Fitch tree, the pig and the hippopotamus diverge from the main line at separate points. Reprinted with permission from Benner *et al.* (2002) Planetary biology: paleontological, geological, and molecular histories of life. *Science* **296**: 864–868, © 2002 AAAS.

was the first artiodactyl to be a true ruminant. This implies that a digestive RNase emerged when ruminant digestion emerged.

This converts the Barnard hypothesis into a broader and robust narrative. This narrative became still more compelling when the molecular behavior is joined to the historical record as known from the fossil and geological records. These records suggested that the camels, deer, and bovid artiodactyl genera diverged *ca.*40 million years ago, together with ruminant digestion and the digestive RNases to support it, at the time of global climate change that began at the end of the Eocene, extended through the Oligocene, and reached a climax with the ice ages in the Pliocene and Pleistocene.

This climate change eventually involved the lowering of the mean temperature of Earth by approximately 17°C, and the drying of large parts of the Earth's surface (Janis *et al.*, 1998). This, in

turn, was almost certainly causally related to the emergence of grasses as a predominant source of vegetable food in many ecosystems. Tropical rainforests receded, grasslands emerged, and the interactions between herbivores and their foliage changed. Grasses offer poor nutrition compared to many other flora, and ruminant physiology appears to have substantial adaptive value when eating grasses.

This, in turn, may help explain why ruminant artiodactyls were enormously successful in competition with the herbivorous perissodactyls (for example, horses, tapirs, and rhinoceroses) as the global climate change proceeded. Today, nearly 200 species of artiodactyls have displaced the approximately 250 species of perissodactyls that were found in the tropical Eocene. Today, only three species groups of perissodactyl survive. This is the principal reason why resurrection of enzymes from the dawn horse will remain outside

of the reach of contemporary paleomolecular biologists, unless ancestral DNA is extracted from the fossil of the dawn horse directly.

18.4 Ribonuclease homologs involved in unexpected biological activities

The paleobiochemical experiments with pancreatic RNases suggested that RNases having digestive function emerged in artiodactyls from a non-digestive precursor about 40 million years ago. This implies, in turn, that non-digestive cousins of digestive RNases might remain in the genomes of modern mammals, where they might continue to play a non-digestive role there.

This suggestion, generated from the first experiments in paleogenetics, emerged at the same time as researchers were independently discovering non-digestive paralogs of digestive RNase A. These were termed RIBAses (ribonucleases with interesting biological activities) by D'Alessio *et al.* (1991). They include RNase homologs that display immunosuppressive (Soucek *et al.*, 1986), cytostatic (Matousek, 1973), anti-tumor (Ardelt *et al.*, 1991), endothelial-cell-stimulatory (Strydom *et al.*, 1985), and lectin-like activities (Okabe *et al.*, 1991). These proteins all appeared to be extracellular, based on their secretory signal peptides and the presence of disulfide bonds. Their existence suggested to some that perhaps a functional RNA existed outside of cells (Benner, 1988).

These results suggested that the RNase A superfamily was extremely dynamic in vertebrates, with larger than typical amounts of gene duplication, paralog generation, and gene loss. In humans, for example, prior to the completion of the complete genome sequence, eight RNases were already known. These included the poorly named human pancreatic ribonuclease (RNase 1; which does not appear to be a protein specific for the pancreas), the equally poorly named eosinophil-derived neurotoxin (EDN, or RNase 2; which does not appear to have a physiological role as a neurotoxin), the eosinophil-cationic protein (ECP, or RNase 3; aptly named in the sense that the name captured all we knew about the protein), RNase 4, angiogenin (RNase 5), RNase 6 (sometimes known as k6), RNase 7 (Harder and

Schroder, 2002; Zhang *et al.*, 2003), and RNase 8 (Zhang *et al.*, 2002).

The analysis *in silico* of the human genome showed that the human RNase 1–8 genes lie on chromosome 14q11.2 as a cluster of approximately 368 kb. In order from the centromere to the telomere, the genes are angiogenin (RNase 5), RNase 4, RNase 6, RNase 1, ECP (RNase 3), an EDN pseudogene, EDN itself (RNase 2), RNase 7, and RNase 8, separated from each other by 6–90-kb intervals. The genome also helped identification of two new human RNase homologs (RNases 9 and 10) in this cluster, preceding angiogenin. In addition, three new open reading frames sharing a number of common features with other RNases were found. Beintema therefore proposed to name these RNases 11, 12, and 13. RNases 11 and 12 are located between RNase 9 and angiogenin. RNase 13 lies on the centromere side of RNase 7, and has a transcriptional direction opposite to that of RNases 7 and 8. The human genome reveals no other open reading frames with significant similarity to these RNase genes. Therefore, it is likely that all human RNase A superfamily members have been identified.

As in humans, rat RNase genes are located on one chromosome (15p14) in a single cluster. The cluster in the rat genome contains the RNase family in the same syntenic order and transcriptional direction as in human, with only a few exceptions. The RNase 1 family (RNase1h, RNase1g, and RNase1y), the eosinophil-associated RNase family (EAR; R15–17, ECP, R-pseudogene, and Ear3), and the angiogenin family (Ang1 and Ang2) have undergone expansion in the rat (Zhao *et al.*, 1998; Singhania *et al.*, 1999; Dubois *et al.*, 2002). Further, orthologs of human RNases 7 and 8 are not present in the rat genome. This permits us to propose a relatively coherent model for the order of gene creation in the time separating primates and rodents, and a listing of the RNase homologs likely to have been present in the last common ancestor of primates and rodents.

The dynamic behavior of this group of genes is shown by the differences separating the rat and mouse groups. In mouse, two RNase gene clusters are found, on mouse chromosome 14qB–qC1 (bcluster AQ) and chromosome 10qB1 (bcluster

BQ). Cluster A is syntenic to the human and rat clusters and is essentially identical to the rat cluster in gene content and order except for substantial expansions of the EAR and angiogenin gene subfamilies. Cluster B emerged in mouse after the mouse/rat divergence, and contains only genes and pseudogenes that belong to the EAR and angiogenin subfamilies. It also includes a large number of pseudogenes.

This level of diversity presents many why-type questions that might be addressed using molecular paleoscience. To date, three of these have been pursued, one in the Rosenberg laboratory, and two in the Benner laboratory.

18.5 Paleogenetics with eosinophil RNase homologs

In an effort to understand more about the function of these abundant RNase paralogs, Zhang and Rosenberg (2002) examined the EDN and ECP in primates. These proteins arose by gene duplication some 30 million years ago in an African primate ancestral to humans and Old World monkeys. Zhang and Rosenberg first asked the basic question: why do eosinophils have two RNase paralogs? Eosinophils are associated with asthma, infective wheezing, and eczema (Onorato *et al.*, 1996); their role in the non-diseased state remains enigmatic. Some textbooks say that eosinophils function to destroy larger parasites and modulate allergic inflammatory responses. Others suggest that eosinophils defend their host from outside agents, with allergic diseases arising as an undesired side effect.

Earlier work by Zhang, Rosenberg and their associates had suggested that ECP and EDN might contribute to organismic defense in other ways. ECP kills bacteria *in vitro*, whereas EDN inactivates retroviruses (Rosenberg and Domachowske, 2001). *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 the paralogs, a change that might account for their differing behaviors *in vitro* (Zhang *et al.*, 1998). This suggests that, in primate evolution, mutations in EDN and ECP may have adapted them for different,

specialized roles during the episodes of rapid sequence evolution.

To obtain a more densely articulated tree for the protein family, Zhang and Rosenberg (2002) sequenced additional genes from various primates. They used these sequences to better reconstruct ancestral sequences for ancient EDN/ECPs. They estimated the posterior probabilities of these ancestral sequences using Bayesian inference. Then they resurrected these ancient proteins by cloning and expressing their genes. Guiding the experimental work was the hypothesis that the anti-retroviral 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 anti-retroviral and RNA-cleaving activities of EDN. Both the ribonuclease and antiviral activities of the last common ancestor of ECP and EDN, which lived *ca*. 30 million years ago, were low. Both activities increased in the EDN lineage after its emergence by duplication.

Zhang and Rosenberg showed that replacements at sites 64 and 132 in the sequence were required together to increase the ribonucleolytic activity of the protein; neither alone was sufficient. Zhang and Rosenberg then analyzed the three-dimensional crystal structure of EDN to offer possible explanations for the interconnection between sites suffering replacement and the changes in biomolecular behavior that they created. They concluded that in the EDN/ECP family, either of the two replacements at sites 64 and 132 individually had little impact on behavior. Each does, however, provide the context for the other to have an impact on behavior. This provides one example where a neutral (or, perhaps better, behaviorally inconsequential) replacement might have set the stage for a second adaptive replacement.

This observation influences how protein engineering is done in general. Virtually all analyses of divergent evolution treat protein sequences as if they were linear strings of letters (Benner *et al.*, 1998). 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 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 (Benner *et al.*, 1997a), and how to assign function to protein sequences (Benner *et al.*, 1998). The results of Zhang and Rosenberg provide an experimental case where higher-order analysis is necessary to understand a biomolecular phenomenon.

Another interpretive strategy involving resurrected proteins (Benner *et al.*, 1997b) was suggested from the results produced by Zhang and Rosenberg. This strategy identifies physiologically relevant behaviors *in vitro* 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. The properties *in vitro* that are the same at the beginning and end of this episode are not relevant to the change in function. This idea is fully implemented in the example of seminal RNases reviewed next in this chapter.

Whereas 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 (Zhang *et al.*, 1998). The antiviral and ribonucleolytic activities of the proteins before and after the adaptive episode in the EDN lineage are quite different. Benner (2002), interpreting the data of Zhang *et al.* (1998), suggested that these activities are important to the emerging physiological role for EDN. This adds support, perhaps only modest, for the notion that the antiviral activity of EDN became important in Old World primates *ca*. 30 million years ago.

The timing of the emergence of the ECP/EDN pair in Old World primates might also contain information. The duplication occurred near 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. These are the same changes as those that presumably drove the selection of ruminant digestion. 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? If we are able to address these questions we might better understand how to improve our immune defenses against viral infections, an area of biomedical research that is in need of rapid progress.

18.6 Paleogenetics with ribonuclease homologs in bovine seminal fluid

New biomolecular function is believed to arise, at least in recent times, largely through recruitment of existing proteins with established roles to play new roles following gene duplication (Ohno, 1970; Benner and Ellington, 1990). Under one model, one copy of a gene continues to divergently evolve under constraints dictated by the ancestral function. The duplicate, meanwhile, is unencumbered by a functional role, and is free to search protein structure space. It may eventually come to encode new behaviors required for a new physiological function, and thereby confer selective advantage.

This model contains a well-recognized paradox. Because duplicate genes are not under selective pressure, they should also accumulate mutations that render them incapable of encoding a protein useful for any function. Most duplicates therefore should become pseudogenes (Lynch and Conery, 2000) or inexpressible genetic information (junk DNA; Li *et al.*, 1981) in just a few million years (Jukes and Kimura, 1984; Marshall *et al.*, 1994). This limits the evolutionary value of a functionally unconstrained gene duplicate as a tool for exploring protein structure space in the search of new behaviors that might confer selectable physiological function.

One of the non-digestive RNase subfamilies offered an interesting system to use experimental paleogenetics to study how new function arises in proteins. This focused on the seminal RNase paralogs found in ruminants that arose by duplication of the RNase A gene just as it was becoming a digestive protein. In ox, seminal RNase is 23 amino acids different from pancreatic RNase A. As suggested by its name, the paralog is expressed in the seminal plasma, where it constitutes some 2% of total protein (D'Alessio *et al.*, 1972). Seminal RNase has evolved to become a dimer with composite active sites. It binds tightly to anionic glycolipids (Opitz, 1995), including seminolipid, a fusogenic sulfated galactolipid found in bovine spermatozoa (Vos *et al.*, 1994). Further, seminal RNase has immunosuppressive and cytotoxic activities that pancreatic RNase A lacks (Soucek *et al.*, 1986; Benner and Allemann, 1989).

Laboratory reconstructions of ancient RNases (Jermann *et al.*, 1995) suggested that each of these traits was not present in the most recent common ancestor of seminal and pancreatic RNase, but rather arose in the seminal lineage after the divergence of these two protein families. To learn more about how this remarkable example of evolutionary recruitment occurred, RNase genes were collected from peccary (*Tayassu pecari*), Eld's deer (*Cervus eldi*), domestic sheep (*Ovis aries*), oryx (*Oryx leucoryx*), saiga (*Saiga tatarica*), yellow backed duiker (*Cephalophus sylvicultor*), lesser kudu (*Tragelaphus imberbis*), and Cape buffalo (*Syncerus caffer caffer*). These diverged approximately in that order within the mammal order Artiodactyla (Carroll, 1988). The newly sequenced genes complemented the known genes for various pancreatic RNases (Carsana *et al.*, 1988) and seminal RNases from ox (*Bos taurus*; Preuss *et al.*, 1990), giraffe (*Giraffa camelopardalis*; Breukelman *et al.*, 1993), and hog deer.

Seminal RNase genes are distinguished from their pancreatic cousins by several marker substitutions introduced early after the gene duplication, including Pro-19, Cys-32, and Lys-62. By this standard, the genes from saiga, sheep, duiker, kudu, and the buffaloes were all assigned to the seminal RNase family. No evidence for a seminal-like gene could be found in peccary. Thus, these

data are consistent with an analysis of previously published genes that places the gene duplication separating pancreatic and seminal RNases at *ca*. 35 million years before present (Beintema *et al.*, 1988), preceding the divergence of giraffe, sheep, saiga, duiker, kudu, Cape buffalo, and ox, in this order, consistent with mitochondrial sequence data (Allard *et al.*, 1992) and global phylogenetic analyses of Ruminanta (Hassanin and Douzery, 2003; Hernandez Fernandez and Vrba, 2005).

Sequence analysis shows that the seminal RNase genes from giraffe, hog deer, roe deer, and Cape buffalo almost certainly could not produce folded stable protein to serve a physiological function. Deletions or insertions create frame shifts in these genes. Further, the seminal RNase genes from okapi, kudu, and saiga were found to encode substitutions at active-site residues. Thus, these proteins are not likely to have catalytic activity.

To show that these seminal genes were indeed not expressed in semen, seminal plasmas from 15 artiodactyls were examined (ox, forest buffalo (*Syncerus caffer nanus*), Cape buffalo, kudu, sitatunga (*Tragelaphus spekei*), nyala (*Tragelaphus angasi*), eland (*Tragelaphus oryx*), Maxwell's duiker (*Cephalophus monticola maxwelli*), yellow-backed duiker, suni (*Neotragus moschatus*), sable antelope (*Hippotragus niger*), impala (*Aepyceros melampus*), saiga, sheep, and Eld's deer). Catalytically active RNase was not detected in the seminal plasma in significant amounts in any artiodactyl genus diverging before the Cape buffalo, except in *Ovis*. Independent mutagenesis experiments showed that the proteins encoded by these genes, all carrying a Cys at position 32, should form dimers (Trautwein, 1991; Raillard, 1993; Jermann, 1995; Opitz, 1995). By Western blotting, however, only small amounts of a monomeric, presumably pancreatic, RNase were detected in these seminal plasmas. In contrast, the seminal plasmas of forest buffalo, cape buffalo, and ox all contained substantial amounts of Western blot-active RNase (Kleineidam *et al.*, 1999). Only in the seminal plasma of ox, however, is seminal RNase expressed. Even though the gene is intact in water buffalo, no expressed protein could be found in its seminal plasma.

The seminal plasma from the *Ovis* genus (sheep and goat) was a notable exception. Sheep seminal plasma contained significant amounts of RNase protein and the corresponding ribonucleolytic activity. To learn whether RNases in the *Ovis* seminal plasma were derived from a seminal RNase gene, the RNase from goat seminal plasma was isolated, purified, and sequenced by tryptic cleavage and Edman degradation. Both Edman degradation (covering 80% of the sequence) and matrix-assisted laser-desorption ionization (MALDI) mass spectroscopy showed that the sequence of the RNase isolated from goat seminal plasma is identical to the sequence of its pancreatic RNase (Beintema *et al*., 1988; Jermann, 1995). This shows that the RNase in *Ovis* seminal plasma is not expressed from a seminal RNase gene, but rather from the *Ovis* pancreatic gene. To confirm this conclusion, a fragment of the seminal RNase gene from sheep was sequenced, and shown to be different in structure from the pancreatic gene.

These results could be perceived as inconsistent with a model that the seminal RNase gene family gradually developed a new seminal function by stepwise point mutation and continuous selection under functional constraints in the seminal plasma following gene duplication. Rather, the duplicate RNase gene seems initially to have served no function at all. It therefore suffered damage, only to be repaired much later in evolution, after the divergence of kudu, but before the divergence of Cape buffalo, from the lineage leading to ox. Clades containing the saiga, duiker, and sheep are known in the early Miocene (23.8–16.4 million years ago), whereas clades containing the kudu and cape buffalo are known in the late Miocene (11.2–5.3 million years ago). Despite the incompleteness of the fossil record, we might conclude that the damaged gene was repaired extremely rapidly in only a few million years (TrabesingerRuef *et al*., 1996). The paleogenetic study, however, will show support for an alternative scenario.

But what was this new function of bovine seminal RNase? What is the molecular basis of the newly acquired function? To address these questions, we set out to reconstruct and resurrect the ancestral seminal proteins. Figure 18.2 shows the nodes where sequences were reconstructed using a likelihood method. These nodes include the evolutionary period where the new biological function might be arising. Three different evolutionary models, one amino acid-based and two codon-based, were used to make the reconstructions. Two outgroups were also considered, those holding the pancreatic RNases and brain RNases, as the data did not unambiguously force the conclusion that one of these two RNAse subfamilies was the closest outgroup. Next, ambiguity at the level of the phylogeny was considered.

To determine the phylogeny four methods were used: a Bayesian analysis as implemented by the program MrBayes, and three different maximum-likelihood models implemented by the PAUP software package. Both outgroups were considered for the different methods. Using the brain sequences as the outgroup, paleontologically unreasonable topologies resulted, as judged by comparison to species trees based on much larger data-sets. With the pancreatic RNases as the outgroup, however, each tool generated the same set of trees, but with slightly different ranking based on slightly different scores. These trees generally agreed with the accepted species trees based on large sets of data (Hassanin and Douzery, 2003; Hernandez Fernandez and Vrba, 2005). The same was observed if *both* pancreatic and brain RNases were used to construct trees. Therefore, the following tree topologies were considered (see Figure 18.2).

1 Topology 1. Preferred when the work began, this topology also received the highest score from a complete Bayesian analysis. Topology 1 groups the okapi with the deer, and models the saiga and duiker as diverging separately from the lineage leading to oxen after the divergence of deer.
2 Topology 2. Preferred today based on a global analysis of all available sequence and paleontological data. Topology 2 places okapi as an outgroup separate from deer, with the giraffe, and diverging before deer diverged from oxen. It also groups saiga and duiker.
3 Topology 3. This topology places okapi in a clade with deer, and places saiga and duiker together.
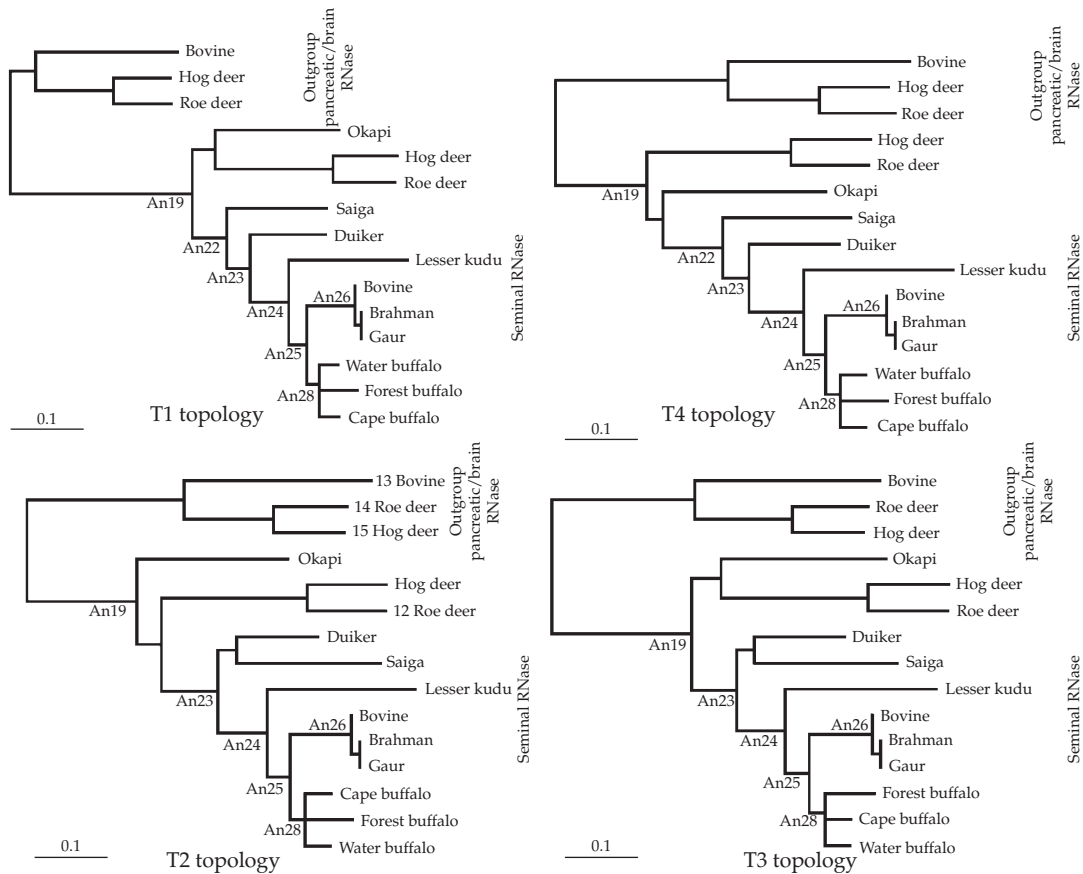
Figure 18.2 Four candidate trees describing the relationship between the artiodactyls providing genes for seminal ribonuclease. Ancestral proteins from the marked nodes (An19–An28) were resurrected in paleogenetic study. See text for discussion.

4 Topology 4. This topology was considered reasonable when this work began, but less so now in light of subsequently emerging data. Here, okapi lies in a clade separate from deer, but diverges from the lineage leading to oxen after deer diverge. Duiker and saiga again were represented as diverging separately from the lineage leading to ox after the divergence of deer.

The closely scoring alternative trees are the consequence of the seminal RNase family having a remarkable amount of homoplasy (parallel and convergent sequence evolution); homoplasy is found at sites 9, 18, 22, 53, 55, 64, 101, and 113. Given this level of homoplasy, no tree can be unambiguously viewed as being correct. Therefore, alternative trees were considered in

proposing candidate ancestral sequences. In an effort to manage ambiguities, all possible sequences were resurrected whenever the reconstructions disagreed.

The distribution of ancestral replacements on the three-dimensional structure of seminal RNase followed a specific pattern. All of the active-site residues remained conserved after the gene duplication. Moreover, the RNA-binding site was also conserved. Most of the replacements were concentrated on the surface of the protein and away from the RNA-binding site. This replacement pattern is consistent with an evolutionary path where the enzymatic function of the protein was conserved; it is not consistent with an inference that the ancestral seminal RNase genes were pseudogenes. Furthermore, the lesions causing the

pseudogene formation in the different lineages are different. These two observations taken together imply that the ancestral seminal RNases were enzymatically active, and that independent inactivation events converted active genes in the different lineages into pseudogenes in many of the modern artiodactyls. Consistent with this model, the resurrected ancestral seminal RNases were all enzymatically active (hydrolyzing a fluorescently labeled RNA substrate; Kelemen *et al.*, 1999).

What then were the properties of seminal RNases that were the targets for natural selection over the past 30 million years? As noted above, there are many behaviors *in vitro* to choose from. For some (such as anti-proliferative activity against cancer cells in culture), it is difficult to rationalize how such behaviors might be important for a protein that exists in seminal plasma. But the site of expression of a protein is changeable over short periods of evolutionary time, meaning that we cannot be certain where seminal RNase has been expressed over its history.

We hypothesized that since seminal RNase is expressed in the seminal fluid and has immunosuppressive activity, it could have evolved to confer a selective reproductive advantage to bulls when the female reproductive tract mounts an immune response against the invading sperm. Indeed, it has been shown in reproductive biology that in many species sperm encounters a defensive immune response and that in many cases seminal plasma is capable of repressing this response (James and Hargreave, 1984; Schroder *et al.*, 1990; Kelly and Critchley, 1997).

To test whether this is true, Benner *et al.* (2007) exploited the strategy to identify the physiologically relevant behaviors *in vitro* for a newly emerging function. As noted above, the strategy examines the behavior of proteins resurrected from points in history before and after the presumed episode of adaptive evolution. The behaviors *in vitro* that are rapidly changing during this episode are inferred to be those relevant to adaptive change. The behaviors *in vitro* that are the same at the beginning and end of this episode are not relevant to the change in function.

Episodes of adaptive evolution are frequently inferred from high normalized non-synonymous/synonymous

(d_N/d_S) ratios (significantly greater than unity), where amino acid replacements conferred new behaviors that conferred enhanced fitness on a protein subject to new functional demands. Thus, they characterize episodes where the derived sequence, at the end of the episode, has (in some sense) a physiological function different from that of the ancestral sequence at its beginning. Low ratios (< 1; although these ratios approach zero in highly conserved proteins) characterize episodes where the ancestral and derived sequences at the beginning and end of the episode have the same physiological function.

The application of this tool in this gene family detected a phase of evolution during the emergence of bovine seminal ribonuclease after the ox diverged from the buffalo. A variety of models within PAML were used to determine d_N/d_S ratios for individual branches in the tree. The Akaike Information Criterion (AIC; Posada and Buckley, 2004) was then used to select the model that best fits the data. Model comparison showed that regardless of the ambiguities in the evolutionary model, the outgroup, or the tree, only the branch leading to the modern seminal RNase in ox, in its three forms (the gaur, Brahman, and ox), underwent adaptive evolution; for no other branch of the tree is this conclusion required. The d_N/d_S ratio was in the range of 1.6–6, depending on the historical model, including the tree topology, choice of outgroup, and choice of codon model.

This strongly suggests that the functional constraints on protein structure, and a correlated change in the physiological function of the protein, occurred in this episode. To identify which *in vitro* behaviors are also changing at this time, the genes encoding ancestral seminal RNase were synthesized by site-directed mutagenesis of a previously prepared RNase synthetic gene. The ancestral RNase candidates were expressed in *Escherichia coli* and purified using newly developed oligonucleotide affinity chromatography.

To address the biomolecular behaviors changing during this adaptive phase, Sassi *et al.* (unpublished data) examined several biochemical and cell-based biomolecular behaviors. The k_{cat}/K_m ratios characterizing the enzyme's ability to catalyze the hydrolysis of a fluorescently labeled

model RNA substrate (carboxyfluoresceinylhexyl-pdAUdAdAp-hexyl-tetramethylrhodamine, IDT) were not significantly different from that of bovine seminal RNase. This implies that this biomolecular behavior is not key to the newly emerging biological function in bovine seminal plasma. All of the candidate ancestral RNases could form dimers under oxidizing conditions, as does seminal RNA, implying that this behavior was not key to the newly emerging function. The rates of folding and other gross physical properties of the ancestors were also not greatly different in the ancestral and modern seminal RNases.

In contrast, the immunosuppressivity of the seminal RNases, measured *in vitro* using a mixed lymphocyte reaction assay exploiting bovine leukocytes isolated from fresh peripheral bovine blood, increased noticeably in the descendents following the branch having adaptive evolution. This suggests that immunosuppression, as measured in this assay *in vitro*, is physiologically relevant for the new function of seminal RNase. This result was paralleled by results in mitogen induction assays which have less physiological relevance. This suggests that the cell-based assays are measuring a property that is important for the new biological function of seminal RNase.

Raines recently suggested that the cell-based activities of RNase might require a swapping of residues 1–20, mentioned above, to form composite active sites (Kim *et al.*, 1995; Lee and Raines, 2005). Accordingly, the extent of the swap was measured using a divinylsulfone crosslinking reagent following the procedure of Ciglic *et al.* (1998). Whereas the extent of swapping may be sensitive to the precise conditions under which the proteins were renatured, the extent of swapping measured *in vitro* also increases during the episode of adaptive evolution. This confirms the hypothesis of Raines, and suggests a structural feature relevant to an adaptive change as well as a biomolecular behavior.

It is important to note that these paleogenetic experiments suggest inferences about the structural changes and behavioral changes that may be important to changing physiological function without recourse to specific studies on the living animals. Further, these inferences are robust with respect to the ambiguities inherent in the reconstruction of historical states from derived sequences.

This study presents an example where the evolutionary history of a gene and the physiological function of the protein were both unknown but the resurrection of the ancestral protein provided evidence for a hypothesis and hints at the evolutionary events shaping this gene's history.

As the number of genomic sequences available increases, the paleogenetic strategy to connect biomolecular behavior *in vitro* to physiological relevance will become easier to apply. Whereas we do not expect it to replace specific studies on the living animals, it should be useful to direct those studies in fruitful directions, and add an interpretive dimension. Paleogenetics is applicable to any biomolecular system where evolutionary reconstructions indicate an episode of adaptive evolution, either through high non-synonymous/synonymous ratios or by high absolute rates of protein sequence divergence deduced using geobiological markers (e.g. fossils) for dating times of divergence.

18.7 Lessons learned

The emergence of systems biology as a paradigm in modern science has brought new attention to a longstanding problem in reductionist biology, which asks whether a particular behavior of an isolated protein, measured *in vitro*, is physiologically relevant in a complex living organism. As Darwinism admits natural selection as the only mechanism for obtaining functional behaviors in biology, this question is equivalent to asking whether the *in vitro* behavior in question, if changed, would lead to a host organism with a diminished ability to survive and reproduce. It has proven remarkably difficult to correlate specific behaviors to fitness although biomolecular behaviors must correlate with fitness in a general sense. A strategy mainly demonstrated in the seminal RNase example uses resurrected ancestral proteins from extinct organisms to help identify biomolecular behaviors *in vitro* that are physiologically relevant to newly emerging biomolecular function.

The ribonuclease family contains the best-developed example of the use of paleomolecular resurrections to understand protein function. It also demonstrates most of the key issues that must be addressed when implementing this paradigm. This includes the management of ambiguities. In all of the cases reviewed here, additional sequences were obtained from additional organisms to increase the articulation of the evolutionary tree, and thereby reduce the ambiguity in the inferred ancestral sequences. When ambiguities remained, multiple candidate ancestral sequences were resurrected to determine that the behavior subject to biological interpretation was robust with respect to the ambiguity.

These examples also show the value of maximum likelihood and empirical Bayes tools in reconstructing ancestral sequences. The simplest parsimony tools, which minimize the number of changes in a tree, are easily deceived by swaps around short branches. Ancestral character states are less likely to be confused by incorrect detailed topology of tree when they are constructed using maximum-likelihood tools than by maximum-parsimony tools.

More important, however, these examples show the potential of molecular paleoscience as a strategy to sort out the complexities of biological function in complex genome systems. Here, the potential of this strategy has only begun to be explored. In the long term, we expect that paleomolecular resurrections will allow us to understand changing biomolecular function in the context of ecological and planetary systems. Margulis and others have referred to this as planetary biology (Margulis and West, 1993; Margulis and Guerrero, 1995; Benner *et al.*, 2002).

Last, these examples show the value of paleomolecular resurrections in converting just-so stories into serious scientific narratives that connect phenomenology inferred by correlation into a comprehensive historical-molecular hypothesis that incorporates experimental data and suggests new experiments. Thus they offer a key example of how paleobiology might enter the mainstream of molecular biology as the number of genome sequences becomes large, and the frustration with their lack of meaning becomes still more widespread.

References

Allard, M.W., Miyamoto, M.M., Jarecki, L., Kraus, F., and Tennant, M.R. (1992) DNA systematics and evolution of the artiodactyl family Bovidae. *Proc. Natl. Acad. Sci. USA* **89**: 3972–3976.

Ardelt, W., Mikulski, S.M., and Shogen, K. (1991) Amino acid sequence of an anti-tumor protein from *Rana pipiens* oocytes and early embryos. Homology to pancreatic ribonucleases. *J. Biol. Chem.* **266**: 245–251.

Barnard, E.A. (1969) Biological function of pancreatic ribonuclease. *Nature* **221**: 340–344.

Beintema, J.J. and Gruber, M. (1967) Amino acid sequence in rat pancreatic ribonuclease. *Biochim. Biophys. Acta* **147**: 612–614.

Beintema, J.J. and Gruber, M. (1973) Rat pancreatic ribonuclease. 2. Amino acid sequence. *Biochim. Biophys. Acta* **310**: 161–173.

Beintema, J.J. and Martena, B. (1982) Primary structure of porcupine (hystrix-cristata) pancreatic ribonuclease: close relationship between african porcupine (an old world hystricomorph) and new world caviomorphs. *Mammalia* **46**: 253–257.

Beintema, J.J., Gaastra, W., and Munniksma, J. (1979) Primary structure of pronghorn pancreatic ribonuclease: close relationship between giraffe and pronghorn. *J. Mol. Evol.* **13**: 305–316.

Beintema, J.J., Wietzes, P., Weickmann, J.L., and Glitz, D. G. (1984) The amino acid sequence of human pancreatic ribonuclease. *Anal. Biochem.* **136**: 48–64.

Beintema, J.J., Broos, J., Meulenberg, J., and Schuller, C. (1985) The amino acid sequence of snapping turtle (chelydra-serpentina) ribonuclease. *Eur. J. Biochem.* **153**: 305–312.

Beintema, J.J., Schuller, C., Irie, M., and Carsana, A. (1988) Molecular evolution of the ribonuclease super-family. *Prog. Biophys. Mol. Biol.* **51**: 165–192.

Benner, S.A. (1988) Extracellular 'communicator RNA'. *FEBS Lett.* **233**: 225–228.

Benner, S.A. (2002) The past as the key to the present: resurrection of ancient proteins from eosinophils. *Proc. Natl. Acad. Sci. USA* **99**: 4760–4761.

Benner, S.A. and Allemann, R.K. (1989) The return of pancreatic ribonucleases. *Trends Biochem. Sci.* **14**: 396–397.

Benner, S.A. and Ellington, A.D. (1990) Evolution and structural theory: the frontier between chemistry and biology. In *Bioorganic Chemistry Frontiers* (Dugas, H., ed.), pp. 1–54. Springer Verlag, Berlin.

Benner, S.A., Cannarozzi, G., Gerloff, D., Turcotte, M., and Chelvanayagam, G. (1997a) Bona fide predictions of protein secondary structure using transparent

analyses of multiple sequence alignments. *Chem. Rev.* **97**: 2725–2843.

Benner, S.A., Haugg, M., Jermann, T.M., Opitz, J.G., Raillard-Yoon, S.-A., Soucek, J. *et al*. (1997b) Evolutionary reconstructions in the ribonuclease family. In *Ribonucleases* (D'Alessio, J.R.a.G., ed.), pp. 214–244. Academic Press, New York.

Benner, S.A., Trabesinger, N., and Schreiber, D. (1998) Post-genomic science: converting primary structure into physiological function. *Adv. Enzyme Regul.* **38**: 155–180.

Benner, S.A., Caraco, M.D., Thomson, J.M., and Gaucher, E.A. (2002) Planetary biology: paleontological, geological, and molecular histories of life. *Science* **296**: 864–868.

Benner, S.A., Sassi, S.O., and Gaucher, E.A. (2007) Molecular paleosciences. Systems biology from the past. In *Advances in Enzymology and Related Areas of Molecular Biology: Protein Evolution* (Toone, E., ed.), vol. 75, pp. 1–132. Wiley, Chichester.

Blackburn, P. and Moore, S. (1982) *Pancreatic Ribonucleases, The Enzymes*, pp. 317–433. Academic Press, New York.

Bonaventura, J., Bonaventura, C., and Sullivan, B. (1974) Urea tolerance as a molecular adaptation of elasmobranch hemoglobins. *Science* **186**: 57–59.

Breukelman, H.J., Beintema, J.J., Confalone, E., Costanzo, C., Sasso, M.P., Carsana, A. *et al*. (1993) Sequences related to the ox pancreatic ribonuclease coding region in the genomic DNA of mammalian species. *J. Mol. Evol.* **37**: 29–35.

Breukelman, H.J., Jekel, P.A., Dubois, J.Y.F., Mulder, P., Warmels, H.W., and Beintema, J.J. (2001) Secretory ribonucleases in the primitive ruminant chevrotain (*Tragulus javanicus*). *Eur. J. Biochem.* **268**: 3890–3897.

Carroll, R.L. (1988) *Vertebrate Paleontology and Evolution*. WH Freeman & Co, New York.

Carsana, A., Confalone, E., Palmieri, M., Libonati, M., and Furia, A. (1988) Structure of the bovine pancreatic ribonuclease gene: the unique intervening sequence in the 5′ untranslated region contains a promoter-like element. *Nucleic Acids Res.* **16**: 5491–5502.

Ciglic, M.I., Jackson, P.J., Raillard, S.A., Haugg, M., Jermann, T.M., Opitz, J.G. *et al*. (1998) Origin of dimeric structure in the ribonuclease superfamily. *Biochemistry* **37**: 4008–4022.

D'Alessio, G., Floridi, A., De Prisco, R., Pignero, A., and Leone, E. (1972) Bull semen ribonucleases. 1. Purification and physico-chemical properties of the major component. *Eur. J. Biochem.* **26**: 153–161.

D'Alessio, G., Di Donato, A., Parente, A., and Piccoli, R. (1991) Seminal RNase: a unique member of the ribonuclease superfamily. *Trends Biochem. Sci.* **16**: 104–106.

Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. (1978) A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.), pp. 345–352. National Biomedical Research Foundation, Washington DC.

Dubois, J.Y.F., Jekel, P.A., Mulder, P., Bussink, A.P., Catzeflis, F.M., Carsana, A., and Beintema, J.J. (2002) Pancreatic type ribonuclease 1 gene duplications in rat species. *J. Mol. Evol.* **55**: 522–533.

Emmens, M., Welling, G.W., and Beintema, J.J. (1976) Amino acid sequence of pike-whale (lesser-rorqual) pancreatic ribonuclease. *Biochem. J.* **157**: 317–323.

Gaastra, W., Groen, G., Welling, G.W., and Beintema, J.J. (1974) Primary structure of giraffe pancreatic ribonuclease. *FEBS Lett.* **41**: 227–232.

Gaastra, W., Welling, G.W., and Beintema, J.J. (1978) Amino acid sequence of kangaroo pancreatic ribonuclease. *Eur. J. Biochem.* **86**: 209–217.

Graur, D. (1993) Towards a molecular resolution of the ordinal phylogeny of the eutherian mammals. *FEBS Lett.* **325**: 152–159.

Groen, G., Welling, G.W., and Beintema, J.J. (1975) Amino acid sequence of gnu pancreatic ribonuclease. *FEBS Lett.* **60**: 300–304.

Harder, J. and Schroder, J.M. (2002) RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *J. Biol. Chem.* **277**: 46779–46784.

Hassanin, A. and Douzery, E.J. (2003) Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Syst. Biol.* **52**: 206–228.

Hernandez Fernandez, M. and Vrba, E.S. (2005) A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biol. Rev. Camb. Philos. Soc.* **80**: 269–302.

Ipata, P.L. and Felicioli, R.A. (1968) A spectrophotometric assay for ribonuclease activity using cytidylyl-(3′, 5′)-adenosine and uridylyl-(3′, 5′)-adenosine as substrates. *FEBS Lett.* **1**: 29–31.

James, K. and Hargreave, T.B. (1984) Immunosuppression by seminal plasma and its possible clinical significance. *Immunol. Today* **5**: 357.

Janis, C.M., Effinger, J.E., Harrison, J.A., Honey, J.G., Kron, D.G., Lander, B. *et al*. (1998) Artiodactyla. In *Evolution of Tertiary Mammals of North America* (Janis, C. M., Scott, K.M., and Jacobs, L., eds), pp. 337–357. Cambridge University Press, Cambridge.

Jekel, P.A., Sips, H.J., Lenstra, J.A., and Beintema, J.J. (1979) Amino acid sequence of hamster pancreatic ribonuclease. *Biochimie* **61**: 827–839.

Jermann, T.M. (1995) *Der Ursprung und die Evolution der Ribonuklease aus dem Pankreas und aus der Samenfluessigkeit*. ETH Dissertation no. 11059, Zurich.

Jermann, T.M., Opitz, J.G., Stackhouse, J., and Benner, S. A. (1995) Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**: 57–59.

Jukes, T.H. and Kimura, M. (1984) Evolutionary constraints and the neutral theory. *J. Mol. Evol.* **21**: 90–92.

Kelemen, B.R., Klink, T.A., Behlke, M.A., Eubanks, S.R., Leland, P.A., and Raines, R.T. (1999) Hypersensitive substrate for ribonucleases. *Nucleic Acids Res.* **27**: 3696–3701.

Kelly, R.W. and Critchley, H.O. (1997) Immunomodulation by human seminal plasma: a benefit for spermatozoon and pathogen? *Hum. Reprod.* **12**: 2200–2207.

Kim, J.S., Soucek, J., Matousek, J., and Raines, R.T. (1995) Mechanism of ribonuclease cytotoxicity. *J. Biol. Chem.* **270**: 31097–31102.

Kleineidam, R.G., Jekel, P.A., Beintema, J.J., and Situmorang, P. (1999) Seminal type ribonuclease genes in ruminants, sequence conservation without protein expression? *Gene* **231**: 147–153.

Kuper, H. and Beintema, J.J. (1976) Amino acid sequence of topi pancreatic ribonuclease. *Biochim. Biophys. Acta* **446**: 337–344.

Lang, K. and Schmid, F.X. (1986) Use of a trypsin pulse method to study the refolding pathway of ribonuclease. *Eur. J. Biochem.* **159**: 275–281.

Lee, J.E. and Raines, R.T. (2005) Cytotoxicity of bovine seminal ribonuclease: monomer versus dimer. *Biochemistry* **44**: 15760–15767.

Lenstra, J.A. and Beintema, J.J. (1979) Amino acid sequence of mouse pancreatic ribonuclease: extremely rapid evolutionary rates of the myomorph rodent ribonucleases. *Eur. J. Biochem.* **98**: 399–408.

Li, W.H., Gojobori, T., and Nei, M. (1981) Pseudogenes as a paradigm of neutral evolution. *Nature* **292**: 237–239.

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.

Margoliash, E. (1963) Primary structure and evolution of cytochrome C. *Proc. Natl. Acad. Sci. USA* **50**: 672–679.

Margoliash, E. (1964) Amino acid sequence of cytochrome C in relation to its function and evolution. *Can. J. Biochem. Physiol.* **42**: 745–753.

Margulis, L. and West, O. (1993) Gaia and the colonization of Mars. *GSA Today* **3**: 277–280, 291.

Margulis, L. and Guerrero, R. (1995) Life as a planetary phenomenon: the colonization of Mars. *Microbiologia* **11**: 173–184.

Marshall, C.R., Raff, E.C., and Raff, R.A. (1994) Dollo's law and the death and resurrection of genes. *Proc. Natl. Acad. Sci. USA* **91**: 12283–12287.

Matousek, J. (1973) The effect of bovine seminal ribonuclease (AS RNase) on cells of Crocker tumour in mice. *Experientia* **29**: 858–859.

Muskiet, F.A.J., Welling, G.W., and Beintema, J.J. (1976) Studies on primary structure of bison pancreatic ribonuclease. *Int. J. Pept. Protein Res.* **8**: 345–348.

Nambiar, K.P., Stackhouse, J., Stauffer, D.M., Kennedy, W.P., Eldredge, J.K., and Benner, S.A. (1984) Total synthesis and cloning of a gene coding for the ribonuclease S protein. *Science* **223**: 1299–1301.

Ohno, S. (1970) *Evolution by Gene Duplication*. Springer Verlag, Berlin.

Okabe, Y., Katayama, N., Iwama, M., Watanabe, H., Ohgi, K., Irie, M. *et al.* (1991) Comparative base specificity, stability, and lectin activity of 2 lectins from eggs of *Rana catesbeiana* and *R. japonica* and liver ribonuclease from *R. catesbeiana*. *J. Biochem. (Tokyo)* **109**: 786–790.

Onorato, J., Scovena, E., Airaghi, S., Morandi, B., Morelli, M., Pizzi, M., and Principi, N. (1996) Role of serum eosinophil cationic protein (s-ECP), neutrophil myeloperoxidase (s-MPO) and mast cell triptase (s-TRY) in children with allergic, infective asthma and atopic dermatitis. *Riv. Ital. Pediatr.* **22**: 900–911.

Opitz, J.G. (1995) *Maximum parsimony: Ein neuer Ansatz zum besseren Verstaendnis von Protein/Nukleinsaeure-Wechselwirkungen*. ETH Dissertation no. 10952, Zurich.

Opitz, J.G., Ciglic, M.I., Haugg, M., Trautwein-Fritz, K., Raillard, S.A., Jermann, T.M., and Benner, S.A. (1998) Origin of the catalytic activity of bovine seminal ribonuclease against double-stranded RNA. *Biochemistry* **37**: 4023–4033.

Posada, D. and Buckley, T.R. (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53**: 793–808.

Preuss, K.D., Wagner, S., Freudenstein, J., and Scheit, K. H. (1990) Cloning of cDNA encoding the complete precursor for bovine seminal ribonuclease. *Nucleic Acids Res.* **18**: 1057.

Raillard, S.A. (1993) *Veraenderung der Struktur und der biologischen Aktivitaet in RNase A mit Hilfe von gezielter Mutagenese*. ETH Dissertation no. 10022, Zurich.

Riggs, A. (1959) Molecular adaptation in haemoglobins. Nature of The Bohr effect. *Nature* **183**: 1037–1038.

Rosenberg, H.F. and Domachowske, J.B. (2001) Eosinophils, eosinophil ribonucleases, and their role in host defense against respiratory virus pathogens. *J. Leukoc. Biol.* **70**: 691–698.

Schroder, W., Mallmann, P., van der Ven, H., Diedrich, K., and Krebs, D. (1990) Cellular sensitization against spermatic and seminal plasma antigens in women after

intrauterine insemination. *Arch. Gynecol. Obstet.* **248**: 67–74.

Singhania, N.A., Dyer, K.D., Zhang, J., Deming, M.S., Bonville, C.A., Domachowske, J.B., and Rosenberg, H. F. (1999) Rapid evolution of the ribonuclease A superfamily: adaptive expansion of independent gene clusters in rats and mice. *J. Mol. Evol.* **49**: 721–728.

Soucek, J., Chudomel, V., Potmesilova, I., and Novak, J.T. (1986) Effect of ribonucleases on cell-mediated lympholysis reaction and on GM-CFC colonies in bone marrow culture. *Nat. Immun. Cell Growth Regul.* **5**: 250–258.

Stackhouse, J., Presnell, S.R., McGeehan, G.M., Nambiar, K.P., and Benner, S.A. (1990) The ribonuclease from an extinct bovid ruminant. *FEBS Lett.* **262**: 104–106.

Strydom, D.J., Fett, J.W., Lobb, R.R., Alderman, E.M., Bethune, J.L., Riordan, J.F., and Vallee, B.L. (1985) Amino acid sequence of human tumor derived angiogenin. *Biochemistry* **24**: 5486–5494.

TrabesingerRuef, N., Jermann, T., Zankel, T., Durrant, B., Frank, G., and Benner, S.A. (1996) Pseudogenes in ribonuclease evolution: a source of new biomacromolecular function? *FEBS Lett.* **382**: 319–322.

Trautwein, K. (1991) *Construction of an Improved Expression System for Bovine Pancreatic Ribonuclease A and Construction and Characterization of RNase A Mutants.* ETH Dissertation no. 9613, Zurich.

Vandenberg, A., Vandenhendetimmer, L., and Beintema, J.J. (1976) Isolation, properties and primary structure of coypu and chinchilla pancreatic ribonuclease. *Biochim. Biophys. Acta* **453**: 400–409.

Vandijk, H., Sloots, B., Vandenberg, A., Gaastra, W., and Beintema, J.J. (1976) Primary structure of muskrat pancreatic ribonuclease. *Int. J. Pept. Protein Res.* **8**: 305–316.

Vos, J.P., Lopes-Cardozo, M., and Gadella, B.M. (1994) Metabolic and functional aspects of sulfogalactolipids. *Biochim. Biophys. Acta* **1211**: 125–149.

Welling, G.W., Groen, G., and Beintema, J.J. (1975) Amino acid sequence of dromedary pancreatic ribonuclease. *Biochem. J.* **147**: 505–511.

Welling, G.W., Mulder, H., and Beintema, J.J. (1976) Allelic polymorphism in arabian camel ribonuclease and amino acid sequence of bactrian camel ribonuclease. *Biochem. Genet.* **14**: 309–317.

Zhang, J.Z. and Rosenberg, H.F. (2002) Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. *Proc. Natl. Acad. Sci. USA* **99**: 5486–5491.

Zhang, J., Rosenberg, H.F., and Nei, M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**: 3708–3713.

Zhang, J.Z., Dyer, K.D., and Rosenberg, H.F. (2002) RNase 8, a novel RNase A superfamily ribonuclease expressed uniquely in placenta. *Nucleic Acids Res.* **30**: 1169–1175.

Zhang, J.Z., Dyer, K.D., and Rosenberg, H.F. (2003) Human RNase 7: a new cationic ribonuclease of the RNase A superfamily. *Nucleic Acids Res.* **31**: 602–607.

Zhao, W., Kote-Jarai, Z., van Santen, Y., Hofsteenge, J., and Beintema, J.J. (1998) Ribonucleases from rat and bovine liver: purification, specificity and structural characterization. *Biochim. Biophys. Acta* **1384**: 55–65.