Articles

Origin of Dimeric Structure in the Ribonuclease Superfamily†

Mauro I. Ciglic,‡,§ Philip J. Jackson,|| Sun Ai Raillard,‡ Monika Haugg,‡ Thomas M. Jermann,‡ Jochen G. Opitz,‡ Nathalie Trabesinger-Rüf,‡,§ and Steven A. Benner*,§

Departments of Chemistry and Anatomy and Cell Biology, University of Florida, Gainesville, Florida 32611-7200, and Applied Biosystems Division, Perkin-Elmer, Warrington WA3 7PB, United Kingdom

Received September 5, 1997

ABSTRACT: To enable application of postgenomic evolutionary approaches to understand the divergence of behavior and function in ribonucleases (RNases), the impact of divergent sequence on the divergence of tertiary and quaternary structure is analyzed in bovine pancreatic and seminal ribonucleases, which differ by 23 amino acids. In a crystal, seminal RNase is a homodimer joined by two "antiparallel" intersubunit disulfide bonds between Cys-31 from one subunit and Cys-32′ from the other and having composite active sites arising from the "swap" of residues 1−20 from each subunit. Specialized Edman degradation techniques have completed the structural characterization of the dimer in solution, new cross-linking methods have been developed to assess the swap, and sequence determinants of quaternary structure have been explored by protein engineering using the reconstructed evolutionary history of the protein family as a guide. A single Cys at either position 32 (the first to be introduced during the divergent evolution of the family) or 31 converts monomeric RNase A into a dimer. Even with an additional Phe at position 31, another residue introduced early in the seminal lineage, swap is minimal. A hydrophobic contact formed by Leu-28, however, also introduced early in the seminal lineage, increases the amount of "antiparallel" connectivity of the two subunits and facilitates swapping of residues 1−20. Efficient swapping requires addition of a Pro at position 19, a residue also introduced early in the divergent evolution of the seminal RNase gene. Additional cysteines required for dimer formation are found to slow refolding of the protein through formation of incorrect disulfide bonds, suggesting a paradox in the biosynthesis of the protein. Further studies showed that the dimeric form of seminal RNase known in the crystal is not the only form in vivo, where a substantial amount of heterodimer is known. These data complete the acquisition of the background needed to understand the evolution of new structure, behavior, and function in the seminal RNase family of proteins.

In the postgenomic era, evolutionary histories will be readily available for the protein modules that are independently evolving units of protein sequence found in genomic databases. These histories will be represented by a multiple alignment of the sequences of the proteins in the module (as well as their encoding DNA sequences), an evolutionary tree, and reconstructed ancestral DNA and protein sequences for each branch point in the tree. Together with a detailed model of biomolecular evolution, these histories will be used to connect sequence, structure, chemical reactivity, and biological function (*1*). How these evolutionary histories will be used to guide experimental biochemistry has only begun to be explored, however (*2*).

This is the first in a series of papers that develops evolutionary histories as a guide in experimental biochemistry using the extracellular mammalian ribonuclease (RNase)[1] family as a model (*3*, *4*). The value of RNase as a system for exploring protein evolution was noted in the 1970s by Beintema and co-workers (*5*), who determined several dozen sequences of digestive RNases from a variety of artiodactyls (the mammal order containing, inter alia, pigs, camels, deer, giraffe, sheep, goat, antelope, and ox). In 1984, Nambiar et al. (*6*) brought tools from molecular biology to bear on this

† This work was partially supported by the Swiss National Science Foundation, Sandoz AG, the Swiss Federal Institute of Technology, and the National Aeronautics and Space Administration (NAS 5-97210).

* To whom correspondence should be addressed at the Division of Biochemistry, Department of Chemistry, University of Florida, Gainesville, FL 32611-7200.

‡ Swiss Federal Institute of Technology.

§ University of Florida.

|| Perkin-Elmer.

[1] Abbreviations: BS, bovine seminal; CM, carboxymethyl; DTT, DL-dithiothreitol; DVS, divinyl sulfone; DFDNB, 1,5-difluoro-2,4-dinitrobenzene; EDTA, ethylenediaminetetraacetic acid; GSH, reduced glutathione; GSSG, oxidized glutathione; HPLC, high-pressure liquid chromatography; IPTG, isopropyl β-D-thiogalactopyranoside; MALDI/MS, matrix-assisted laser desorption ionization mass spectroscopy; NEM, *N*-ethylmaleimide; PAGE, polyacrylamide gel electrophoresis; PTH, phenylthiohydantoin derivative; PMSF, phenylmethanesulfonyl fluoride; pUp, uridine 2′,5′/3′,5′-diphosphate; RNase, ribonuclease; SDS, sodium dodecyl sulfate; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane; UpA, uridylyl-(3′→5′)-adenosine.

S0006-2960(97)02203-4 CCC: $15.00
Published on Web 03/05/1998

system, pointing out that the ongoing evolution in this system and the unusual biological activities displayed by some RNase family members made the system ideal for exploring adaptive and neutral evolutionary change in proteins. This point was later expanded and discussed in light of new information by Benner (*7*−*9*).

At the center of this discussion is the discovery that the digestive enzyme in ruminants (*10*) is related by evolutionary ancestry to a large family of nondigestive proteins with unusual catalytic and biological activities (*8*, *9*). Angiogenin, for example, isolated by Vallee and co-workers from tumor-conditioned cell medium while seeking substances important in the vascularization of solid tumors (*11*), is a member of the RNase superfamily of proteins. Eosinophil-derived neurotoxin, which causes neuronal degeneration (*12*); eosinophil cationic protein, which has anti-parasitic activity (*13*); the P-30 protein from amphibian eggs, which recently entered clinical trials as an antitumor agent (*14*); and a sialic acid-binding lectin from frog (*15*) are also all members of the RNase superfamily.

That pancreatic RNases might be part of a superfamily of proteins with such interesting biological activities was anticipated by the unusual biological properties of yet another member of the RNase superfamily, bovine seminal ribonuclease (*7*−*9*, *16*). Seminal RNase, which represents approximately 2% of the total protein in bovine seminal plasma, has antispermatogenic activity (*17*), immunosuppressive activity (*18*−*20*), and cytotoxic activity against many transformed cell lines (*21*, *22*). Each of these activities is largely absent from pancreatic RNase and the protein that was the most probable recent common ancestor of seminal and pancreatic RNase (*23*).

To understand the evolutionary processes that have created this remarkable diversity of behavior in relatively little time (from a geological perspective), we must begin by examining the primary evolutionary event, the divergence of sequence, and what its impact is on dependent structural variables, including tertiary and quaternary structure. In the RNase superfamily, the impact has been remarkable. Twenty-three amino acids (out of 124) separate the bovine seminal and pancreatic RNases. These differences offer the opportunity for different quaternary structures that may have been exploited by divergent evolution. In its best characterized purified form, dimeric seminal RNase is joined covalently by two intersubunit disulfide bonds (*24*, *25*), between Cys-31 of one subunit and Cys-32′ on the other and between Cys-32 of the first subunit and Cys-31′ of the other (the "antiparallel" disulfide connectivity). In pancreatic RNase, which does not form a covalent dimer, these positions contain Lys and Ser residues, respectively.

The divergence of primary and quaternary structure also has an impact on tertiary structure in the dimer. In the crystal structure of the seminal RNase dimer, residues 1−20 (the S-peptide) (*26*, *27*) from one subunit of the seminal RNase dimer are swapped with residues 1−20 from the other (*25*). The S-peptide swap creates composite active sites, with one of the two key catalytic histidine residues (His-12) coming from one polypeptide chain in the dimer and the second (His-119), together with catalytically important Lys-41, coming from the other. Pancreatic RNase A is, under normal circumstances, a monomer that builds its active site from a single polypeptide chain.

Homodimers having active sites formed by elements coming from both subunits are, of course, found throughout biochemistry. Dimers formed by joining two polypeptide chains via a disulfide bond are fewer, but still common (insulin is an example). Divergent quaternary structures are known in homologous proteins; for example, alcohol dehydrogenase from horse liver is a dimer, while its homologue from yeast, with some 70% of the amino acids replaced, forms a tetramer (*28*). Proteins with a swapped secondary structural element are rarer, but not unknown (*29*−*31*).

Some evolutionary context illustrates how remarkable these differences are in the seminal/pancreatic RNase pair. Normally, homologous proteins retain the same overall tertiary fold, even after over 75% of the amino acids have undergone substitution (*32*). The subunit of the swapped seminal RNase dimer and the pancreatic RNase monomer differ in tertiary fold after only 20% divergence. Further, the positioning in the polypeptide chain of active-site residues is normally extremely well conserved; indeed, different positioning of active-site residues can be taken as an argument against homology (*33*).

Further, amino acid substitutions that achieve one goal need not be compatible with other goals (*34*). Introducing cysteines on the surface of a protein appears to be a direct way of allowing a protein to dimerize (*35*−*37*). In a protein that already contains eight cysteines forming four disulfide bonds, however, additional cysteines could form incorrect disulfide bonds that create a kinetic trap that slows the folding process (*38*, *39*).

Bennett and co-workers (*31*) recently suggested that swapping between subunits might be a general mechanism by which dimeric structures might evolve *de novo*. In this model, the contact between two swapped domains resembles the contact in the unswapped monomer; this provides a preevolved interface for forming a dimer contact. The evolution of further contact sites to join the two subunits is assumed to be a second step, made possible once selective advantage conferred by dimeric structure is accessible, at least in part, through the swapped dimer. Although controversial (*40*), this model is supported by the propensity of pancreatic RNase to form noncovalent dimers joined by an S-peptide swap when it is lyophilized from acetic acid (*27*).

Some time ago, we noted that reconstruction of specific intermediates in the divergent evolution of a protein family is useful when analyzing the relationship between structure and function in proteins (*6*, *23*, *41*). In the pancreatic RNase family, this reconstruction can be done in great detail, thanks in large part to sequence data from artiodactyls collected by Beintema and co-workers over the past two decades (*5*). Seminal RNase diverged from the pancreatic lineage after the divergence of camel from the lineage leading to ox but before the divergence of deer (*42*). While the precise structures of these intermediates remain hypothetical, each of the ancestral RNases was monomeric. Thus, the evolutionary reconstruction implies that dimeric structure is a feature of RNase that arose in the seminal lineage after it diverged from the pancreatic lineage.

Molecular biological methods have been applied both in these laboratories (*2*, *35*) and in the laboratories of Raines (*43*) and D'Alessio (*36*, *37*, *44*) to describe the relationship between primary, tertiary, and quaternary structure in seminal RNase. We complete this description here. The evolution-

ary analysis of quaternary structure has focused on the 12 amino acid substitutions that separate the sequence of modern ox seminal RNase from this common ancestor, rather than the 23 amino acid substitutions that separate pancreatic and seminal RNase (for this approach, see for example references *35*, *43*, *45*, *46*). This includes amino acid substitutions at positions 17, 19, 28, 31, 32, 39, 55, 62, 76, 111, 113, and 115, including the two cysteines that join covalently the seminal monomers.

MATERIALS AND METHODS

Bovine pancreatic ribonuclease A (RNase A), IPTG, and calf intestine adenosine deaminase were from Boehringer Mannheim. Bacto Tryptone and Bacto yeast extract were from Difco. Ampicillin (sodium salt), DL-dithiothreitol (DTT), 1,5-difluoro-2,4-dinitrobenzene (DFDNB), and uridylyl-(3′→5′)-adenosine (UpA) were from Sigma. NH_4OAc, NaOAc, and $NaHCO_3$ were from Merck. Modifying and restriction enzymes were from New England Biolabs. Chromatography materials were from Pharmacia, Bio-Rad, and Sigma. Gel staining solutions were from Bio-Rad. All other chemicals were from Fluka and were of the highest purity available.

Assay of RNase Catalysis (UpA Substrate). The catalytic activity of the mutants was measured using uridylyl-(3′→5′)-adenosine (UpA) as a substrate (*47*). Stock solutions of NaOAc buffer (1 M, pH 5.0, sterilized by autoclave) and UpA (580 μM; optical density 13.7 at 265 nm, sterile-filtered and stored at −20 °C) were diluted with sterile water for the kinetic runs; the final concentration of NaOAc was 0.1 M. Adenosine deaminase (5 μL of a 2 mg/mL stock solution per 10 mL of UpA solution) was added directly from the commercial solution. A quartz cuvette (1 mL, path length 0.5 and 1 cm) containing UpA solution (15−200 μM; 745 μL) was preequilibrated (25 °C, 3 min). An aliquot of enzyme solution (5 μL, ca. 10−100 ng of RNase variant) was added, the cuvette was inverted three times, and the decrease of absorbance was measured at 265 or 275 nm. The rate of decrease in absorbance was determined; each assay was repeated 2−3 times, and the results were averaged. K_M and k_{cat} values were obtained from Lineweaver−Burk plots. The coupled assay was internally standardized relative to RNase A in each run.

Purification of Seminal RNase. Seminal RNase was purified by a modification of the procedure of Tamburrini and colleagues (*48*). Bovine seminal plasma (Schweizerische Besamungsstation, Bütschwil) was centrifuged twice (8000*g*, 4 °C, 15 min), and the supernatant was stored at −80 °C. After dilution with NH_4HCO_3 buffer (20 mM, with 1 mM PMSF and 0.02% NaN_3, pH 8, 100 mL) it was loaded onto a CM column (diameter 2.5 cm) which was previously equilibrated with the same buffer. The column was washed successively with the following buffers: (a) NH_4HCO_3 (20 mM, with 1 mM PMSF and 0.02% NaN_3, pH 8, 150 mL), (b) NH_4HCO_3/NaCl (20 mM NH_4HCO_3, 100 mM NaCl, 1 mM PMSF, and 0.02% NaN_3, pH 8, 20 mL), (c) NH_4HCO_3/NaCl (20 mM NH_4HCO_3, 200 mM NaCl, 1 mM PMSF, and 0.02% NaN_3, pH 8, 10 mL), and (d) NH_4HCO_3/NaCl (20 mM NH_4HCO_3, 300 mM NaCl, 1 mM PMSF, and 0.02% NaN_3, pH 8, 10 mL). The bovine seminal RNase was then eluted with elution buffer (20 mM NH_4HCO_3, 400 mM NaCl, 1 mM PMSF, and 0.02% NaN_3, pH 8, 100 mL). Fractions (10 mL) were collected and analyzed by SDS−PAGE. Fractions containing ribonuclease were pooled, frozen with liquid nitrogen, and lyophilized.

Sephadex G-50 (fine grade) was swollen in NH_4HCO_3 buffer (20 mM, with 0.02% NaN_3, pH 8), poured into a 550 mL (total volume) column, and washed with the same buffer (2 L). Lyophilized protein from the previous step was dissolved in NH_4HCO_3 buffer (20 mM, with 1 mM PMSF and 0.02% NaN_3, pH 8, 4 mL) and applied to the G-50 column in two portions (2 mL each). The column was washed with NH_4HCO_3 buffer (20 mM, with 0.02% NaN_3, pH 8, at 4 °C, 40 mL/h), the eluent was monitored by ultraviolet absorbance at 280 nm, and fractions (ca. 5.0 mL) were collected. BS RNase dimer was detected after ca. 150 mL; BS RNase monomer eluted after ca. 240 mL. The different fractions were analyzed by SDS−PAGE (nonreducing) and the pure samples were combined and lyophilized.

Construction of Genes Encoding RNase Variants. The synthetic gene for RNase A in the PiAn7 plasmid used as the starting point for the site-specific mutation is described in detail elsewhere (*6*, *35*, *45*). The gene was designed to have unique restriction sites placed at strategic intervals throughout to facilitate site-specific mutation (*49*), and the construction of the mutated genes in this work took advantages of these sites. The constructions are reported in detail elsewhere (*35*, *46*). We illustrate the general procedure by describing the construction of the **A**(K31C S32C),[2] **A**(Q28L K31C S32C), and **A**(A19P Q28L K31C S32C) variants.

PiAN7 RNase plasmid (*50*) containing a gene for RNase A (*51*) was digested with *Nde*I and *Bam*HI, the fragments were separated on an agarose gel [1% in Tris (40 mM, pH 8)−acetate (20 mM)−EDTA (2 mM) (TAE) buffer], the smaller fragment was discarded, and the major fragment constituting the bulk of the plasmid and the first part of the RNase gene (encoding up to residue 11) was isolated using Gene-Clean. In a separate digestion of PiAN7 RNase with *Sau*96I and *Bam*HI, the fragment of the RNase gene encoding amino acids ca. 35−124 was similarly obtained. Two sets of complementary oligonucleotides bridging the *Nde*I and *Sau*96I sites (corresponding approximately to the region coding for residues 12−38) were prepared by automated solid-phase synthesis (Applied Biosystems) with sequences appropriate to encode the **A**(K31C S32C) and **A**(Q28L K31C S32C) variants. These sequences also introduced a *Pae*R7I site at codon 21 of the gene. Three part ligations joining the bulk of the plasmid (the *Bam*HI−*Nde*I fragment), the synthetic oligonucleotides (*Nde*I−*Sau*96I), and the RNase gene fragment (*Sau*96I−*Bam*HI) yielded plasmids carrying genes encoding the **A**(K31C S32C) and **A**(Q28L K31C S32C) variants of RNase A. The *Escherichia coli* cell line P3 (*46*) was transformed with these plasmids, the DNA was isolated using standard techniques (Qiagen protocol), and the mutant gene was characterized by complete sequencing.

[2] Variants of RNase A and seminal RNase are designated first by a boldface letter, indicating the type of RNase that served as the starting point for the mutation (**A** = RNase A; **S** = bovine seminal RNase), with the amino acid replacements indicated within the parentheses using the one-letter code.

Variant **A**(A19P Q28L K31C S32C) was prepared from the plasmid containing the **A**(Q28L K31C S31C) variant. The plasmid was digested with *Nde*I and *Pae*R7I, the small fragment was discarded, and the large fragment was isolated as previously described. This was then ligated with synthetic duplex encoding RNase from residues 12 to 21 with a Pro at position 19.

Variants **A**(S16G T17N A19P A20S) and **A**(S16G T17D A19P A20S) were constructed by digesting the PiAN7 RNase plasmid with *Pae*R7I and *Nde*I and ligating the large fragment with the synthetic duplex encoding RNase from residues 13 to 21 with the corresponding mutations.

To prepare **A**(K31C), **A**(S32C), **A**(K31F S32C), and **A**(A19P K31C S32C), the synthetic gene for RNase A was cloned in the polylinker site of the pCYTEXP1 vector (Medac) behind a λ promoter using the *Xho*I and *Bam*HI restriction enzymes. This vector also encodes a temperature-sensitive λ repressor (*52*). For the preparation of variants **A**(A19P K31C) and **A**(A19P S32C), the synthetic gene for RNase A was cloned in the polylinker site of the pET-23b vector (Novagen) using *Nde*I and *Hin*dIII restriction enzymes. These variants were constructed using the Muta-Gene Phagemid in vitro mutagenesis kit (version 2) from Bio-Rad, following the procedure recommended by the manufacturer. In all cases, the presence of the desired mutation and the lack of second-site mutations were confirmed by total sequence analysis.

Expression of Genes Encoding RNase Variants. The mutant genes constructed in the PiAN7 vector were transferred to the expression plasmid pUN, described in detail elsewhere (*45*, *53*). pUN RNase A was cleaved with *Nde*I and *Bam*HI and the major fragments were retained. The PiAN7 RNase variants were also digested with these endonucleases and the smaller fragment was isolated. In a two-way ligation, *Bam*HI−*Nde*I fragment from pUN RNase containing most of the plasmid and the insert RNase from *Nde*I to *Bam*HI were ligated, and the product was used to transform the *E. coli* cell line RB 791 (*54*). In the pUN RNase vector, the gene follows a λ promoter, which is controlled by a temperature-sensitive λ repressor.

A lon$^-$ cell line (lon$^-$ hptR$^-$ tet^r) of *E. coli* (*55*, *56*) was transformed with either pUN or pCYTEXP1 vectors containing the variant genes, while a BL21(DE3)pLysS (F$^-$ *ompT* r_B^- m_B^-) cell line was transformed with pET-23b vectors. Cells stored in glycerol at −80 °C were used to inoculate an overnight flask containing LB medium (100 mL; 10 g of Bacto Tryptone, 10 g of NaCl, 5 g of Bacto yeast extract, water to 1 L) with ampicillin (0.1 mg/mL). After growth at 30 °C overnight, aliquots (10 mL) from the overnight flask were used to inoculate 10 Erlenmeyer flasks (each 2 L) containing 400−500 mL of expression medium (16 g of Bacto Tryptone, 10 g of NaCl, 10 g of Bacto yeast extract, water to 1 L) to which solutions of $MgSO_4$ (2 mM), $CaCl_2$ (0.1 mM), and ampicillin (100 μg/mL) had been added. The expression medium was then incubated with shaking (30 °C, 250 rpm) until the cells had reached an optical density of 1.0 (550 nm, ca. 4.5 h). Expression of the mutant proteins in the lon$^-$ strain was then induced by raising the temperature to 42 °C, while expression in BL21 was induces by adding IPTG to a final concentration of 0.4 mM. Aliquots were withdrawn at time intervals, the cells were recovered by centrifugation, and their RNase variant content was analyzed by SDS gel electrophoresis to determine the point of maximum protein production (typically 2.5−3 h). The cells were then recovered by centrifugation (5 min, 4 °C, 7000*g*) and either stored at −80 °C or immediately used. From this step on 0.02% NaN_3 was added to all the buffers.

Recovery of Monomeric RNase Proteins. Cells were suspended in five times the pellet's weight of lysis buffer (50 mM Tris, 100 mM NaCl, 1 mM EDTA, and 0.2 mM PMSF, adjusted to pH 7.8 with concentrated HCl) and lysed via French press (1000 psi, 4 °C). Inclusion bodies containing most of the expressed RNase variants were recovered (together with cell debris) by centrifugation (15−30 min, 4 °C, 7000*g*). After addition of 4 volume equiv of buffer relative to the weight of the pellet denaturing buffer (100 mM NaCl, 5 mM EDTA, 8−10 M urea, deionized with mixed bed ion-exchange resin, 1 mM PMSF, 500 mM 2-mercaptoethanol, and 50 mM Tris, pH 7.8, for proteins expressed in the lon$^-$ strain; 6 M guanidinium hydrochloride, 0.1 M DTT, 10 mM Tris-HCl, pH 8, and 1 mM PMSF for proteins expressed in BL21), the mixture was vigorously vortexed and stirred at room temperature until the inclusion bodies were completely dissolved (30−60 min).

After complete solubilization, proteins expressed in the lon$^-$ strain were incubated in a 30 °C shaker for another 30 min. The solution was centrifuged (30 min, room temperature, 7000*g*) and the supernatant was loaded onto a (carboxymethyl)agarose column (diameter 2.5 cm) previously equilibrated with ammonium acetate (20 mM, pH 6.7). The column was washed with NH_4OAc buffer (20 mM, pH 6.7, 100 mL), and the cationic proteins were eluted with CM elution buffer (0.5 M NaCl and 20 mM NH_4OAc, pH 6.7). Proteins expressed in BL21 were, immediately after denaturation, dialyzed against acetic acid (20 mM) for 18 h, and the precipitate was removed by centrifugation (10 min, 4 °C, 7000*g*).

To refold the monomers, solid oxidized (1 mM) and reduced glutathione (10 mM) were added to the eluate or the supernatant, and the pH was raised to between 7.6 and 7.8. Refolding was achieved by stirring the solution at 4 °C for 12 h. After 12 h no further increase in catalytic activity toward UpA as substrate was observed.

The buffer was then changed by ultrafiltration (YM 3 membrane, Amicon) to NaOAc (pH 5.0, 50 mM), and the solution was concentrated (to <5 mL) and applied to a pUp−agarose (Sigma, 5 mL in a 1 cm diameter) column. The column was washed with buffer (NaOAc, pH 5, 50 mL), and the RNase variants were eluted using a stepwise gradient (0.5, 1, and 3 M sodium chloride in 50 mM sodium acetate, pH 5.0, 20−50 mL each) to yield monomeric proteins as judged by SDS−PAGE and silver staining of the gel. Variants without the mutations K31C and/or S32C were concentrated by ultrafiltration and the buffer was exchanged to NaOAc (100 mM, pH 5).

Recovery of Dimeric RNase Proteins. The buffer of the glutathione-blocked monomeric proteins was exchanged to refolding buffer (100 mM NaCl and 50 mM Tris-HCl, pH 7.8). The variants were incubated at room temperature with DTT (100 mM) for 1 h. To remove glutathione and DTT, the solution was loaded onto a preequilibrated PD10 column (Sephadex) and eluted with refolding buffer following the procedure recommended by the manufacturer. The proteins were refolded at 4 °C for 200−300 h in concentrations

between 65 and 500 μg/mL. This procedure yielded a dimer/monomer ratio of up to 9:1, depending on the variant and its concentration. To separate dimer from monomers and oligomers, an additional affinity chromatography with a stepwise elution profile was performed (see above). Pure dimer was obtained in a buffer (3 M NaCl and 250 mM NaOAc, pH 5.0). Samples of 1.5 mL were pooled and analyzed on SDS−PAGE and silver-stained, and the pure fractions were combined. Typically, 1−4 mg of pure dimer was obtained from 4.5 L of expression medium.

The proteins produced in this way bear an N-terminal methionine residue, as determined by Edman degradation. Control experiments of a variety of RNase mutants with and without this methionine have failed to detect evidence for its impact on the catalytic properties, at least for these mutants (*41*). Concentrations of RNase variants were determined both by SDS−PAGE using known amounts of RNase A as standards and by measuring the absorbance at 280 nm [absorbance of a 1% solution is 7.3; alternatively, at 278 nm, the molar extinction coefficient was assumed to be 9800 (*57*)].

Western Blots. Western blotting was performed at 25 mA and 4 °C for 16 h. The membrane was washed in 0.5× PBS/milk buffer (70 mM NaCl, 5 mM sodium phosphate, pH 7.4, and 2.5% dried milk, 14 mL) for 5 min. Anti-BS RNase antibodies (8.6 mg/mL, 30 μL, raised in rabbit) in 0.5× PBS/milk buffer (1 mL) were added and the solution shaken for 1 h. The membrane was washed twice with 1× PBS buffer (25 mL). After the addition of 0.5× PBS/milk buffer (15 mL) containing goat anti-rabbit antibody− (IgG−) horseradish peroxidase conjugate (25 μL), the membrane was shaken at room temperature for 1 h. The membrane was again washed with 1× PBS buffer (15 mL) and developed in a buffer (2.5 mg/mL 4-chloro-1-naphthol, 60% ethanol, and 0.1% hydrogen peroxide) for 5−60 min until bands were clearly visible.

Assessing the Extent of the S-Peptide Swap. Cross-linking experiments were done using divinyl sulfone (DVS) as a 10% solution in ethanol (*58*, *59*). Enzyme (13 μg, 1 nM/subunit) in NaOAc buffer (100 mM, pH 5.0, 100 μL) and DVS (1 μL of the 10% solution, 1 μM) were incubated at 30 °C. This is approximately a 1000-fold excess of sulfone to each subunit of the protein. Aliquots were withdrawn over a period of 96 h, the remaining catalytic activity (against UpA) was measured, and the reaction was quenched by adding 2-mercaptoethanol (final concentration 200 mM) and incubating for 15−30 min at room temperature. The samples were loaded on a reducing SDS electrophoretic gel and resolved by electrophoresis. The ratio of monomer to cross-linked dimer was estimated by Coomassie blue staining.

DVS induced cleavage of the peptide backbone. The fragments were isolated by HPLC (Vydac RP-C18 column, 218TP54, 4 × 250 mm) using a two-solvent system (solvent A contained 0.1% TFA, solvent B contained 0.086% TFA in 90% acetonitrile, gradient 0−1% solvent B over 25 min, then 1−40% solvent B over 40 min), and analyzed by mass spectrometry and Edman degradation.

An independent cross-linking assay was also done in several proteins using 1,5-difluoro-2,4-dinitrobenzene (DFDNB) (*60*−*62*). Enzyme (1.5 μg) was dissolved in borate buffer (50 mM, pH 8.5, 110 μL). A solution (9 μL) of DFDNB (81.5 μM) in 2% methanol/water was added over a period of 5 h, in the dark at room temperature. The solution was incubated for another 20 h, while aliquots were withdrawn and incubated with 2-mercaptoethanol (final concentration 200 mM) for 30 min at room temperature. The samples were loaded on a reducing SDS electrophoretic gel and resolved by electrophoresis. The ratio of monomer to cross-linked dimer was estimated by silver staining.

Kinetics of Refolding. RNase A variants (0.5−1 mg) in refolding buffer (100 mM NaCl and 50 mM Tris-HCl, pH 7.8) were fully reduced and denatured by incubating them in 200 mM DTT at 37 °C for 2 h and heating them to 80 °C for 15 min. To initiate the folding, the samples were passed through a pre-packed PD10 column (Sephadex), which was preequilibrated in refolding buffer (100 mM NaCl and 50 mM Tris-HCl, pH 7.8), and the denatured proteins were recovered in 3.5 mL. The RNase was reoxidized either in air or by adding reduced and oxidized glutathione (1 mM and 0.2 mM, respectively). Aliquots (36 μL) of the refolding mixture were withdrawn at different intervals and the refolding intermediates were trapped by mixing with NaOAc (4 μL, 1 M, pH 5) and freezing at −80 °C. RNase activity of these aliquots was assayed using UpA as substrate as described above, while the extent of dimer formation was determined from SDS−PAGE.

Free SH content of fully reduced and denatured RNase variants was determined by taking immediately an aliquot of the eluate and treating it with *N*-ethylmaleimide (400 μM) at room temperature for 10 min. The alkylation reaction was quenched by adding aqueous trifluoroacetic acid (2%). MALDI/MS analysis of the alkylated samples revealed a single component with a mass expected for monomeric RNase alkylated with 10 equiv of NEM.

Intersubunit Disulfide Connectivity. Dimeric RNase A variants (100 μg) were purified by reversed-phase HPLC using a Vydac C18 column (218TP54, 4 × 250 mm). A two-solvent system was used: solvent A contained 0.1% TFA, while solvent B contained 0.086% TFA in 90% acetonitrile. The proteins were eluted with a gradient from 0% to 30% solvent B over 30 min and from 30% to 45% solvent B over 30 min with a flow rate of 0.7 mL/min. Protein peaks were detected by absorbance at 210 nm.

The fraction containing the pure dimeric RNase A variant was lyophilized and redissolved in aqueous TFA (70%, 100 μL). The solution was treated with grains of CNBr and incubated in the dark at room temperature for 14 h. After lyophilization, the digested RNase A variants were redissolved in aqueous TFA (0.1%, 20 μL) and the peptide containing residues 14−124 interconnected by disulfide bonds was isolated by HPLC as described. The lyophilized fragment was subjected to five cycles of Edman degradation using a gas-phase sequencer, Applied Biosystems Procise HT.

RESULTS

In the form known by X-ray crystallography, seminal RNase is a dimer joined by two intersubunit disulfide bonds, between Cys-31 from one subunit and Cys-32′ on the second and between Cys-31′ on the second and Cys-32 on the first (antiparallel connectivity) (*25*). Raillard (*35*) and Di Donato et al. (*37*) independently showed that a variant of RNase A carrying cysteines at both positions 31 and 32 also formed

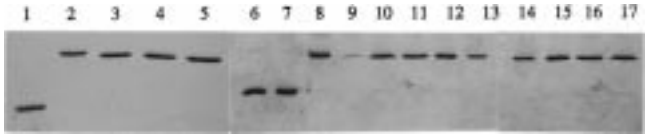

FIGURE 1: Quaternary structure of the RNase A variants as indicated, determined by nonreducing SDS−PAGE. Lane 1, RNase A; lane 2, BS−RNase; lane 3, A(S32C); lane 4, A(K31F S32C); lane 5, A(Q28L K31C); lane 6, RNase A; lane 7, S(C31K C32S); lane 8, BS−RNase; lane 9, A(A19P K31C S32C); lane 10, A(A19P S32C); lane 11, A(A19P K31C); lane 12, A(K31C); lane 13, A(Q28L K31C); lane 14, A(K31C S32C); lane 15, A(Q28L K31C S32C); lane 16, A(A19P Q28L K31C S32C); lane 17, BS−RNase.

a dimer, although the nature of the linkage (parallel or antiparallel) was not investigated. In this work, four variants of RNase A were prepared, **A**(K31C), **A**(S32C), **A**(K31F S32C), and the **A**(K31C S32C) variant reported previously. The Phe at position 31 in the **A**(K31F S32C) variant reflected the fact that reconstructed intermediates throughout much of the evolutionary history of the seminal RNase gene family encode a Phe at this position (*63*−*65*).

Each of these variants retains catalytic activity against UpA as a substrate (Table 1) within a factor of 2 relative to RNase A, indicating that the native fold was achieved. In each variant, refolding under standard reconstitution conditions yields a dimer as the predominant species under oxidizing conditions, judging by a single band with a mass of ca. 27 000 after nonreducing SDS−PAGE (Figure 1). Electrophoresis of the variants in the presence of mercaptoethanol yields single bands migrating at ca. 13 500. These data show that addition of a single Cys to RNase A at either position 31 or position 32 is sufficient for dimer formation.

We next asked whether these dimers swapped the S-peptide to create composite active sites. D'Alessio and coworkers reported an elegant assay for detecting the S-peptide swap in bovine seminal RNase (*66*). In this assay, the dimer is lightly treated with agents that reduce the intersubunit disulfide bonds only. The dimers whose subunit polypeptide chains are entwined disassociate slowly. Raines and coworkers have also used this assay to analyze the biological activities of RNase variants (*43*, *67*), suggesting that dimeric structure with an S-peptide swap is essential for biological activity.

A cross-linking approach exploiting the ability of His-12 and His-119 to react with electrophilic reagents is also conceivable to detect the swap. Divinyl sulfone (DVS) (*58*, *59*) was examined as the cross-linking agent because the sulfone unit mimics the reactive phosphate (*68*) and because the electrophilic CH_2 groups in DVS are at a distance appropriate to react with the two histidines as nucleophiles, judging by the crystal structure of seminal RNase (*25*). If the active site of the dimer is composite, with His-12 coming from one subunit and His-119 coming from the other, the cross-link should join covalently the two subunits even under denaturing conditions, yielding a product of mass 27 000 even after reduction of the intersubunit disulfide bonds. If the active site is not composite, cross-linking would link two histidines from the same subunit, yielding monomers rather than dimers under reducing conditions.

DVS (0.1% solution) gave time-dependent irreversible inactivation of RNase A with a half-life of ca. 12 h. After reduction, inactivated RNase A gave a product that by SDS−PAGE had a mass of ca. 14 000, consistent with the fact that RNase A is not a dimer with swapped S-peptides and composite active sites (Figure 2). MALDI-TOF mass spectroscopy of the product showed that DVS also reacted with the side chains of lysines, presumably to give a cyclic product. A parallel study with difluorodinitrobenzene [DFDNB (*60*)], which cross-links Lys-7 with Lys-41 (and therefore also detects a swap), also gave only monomer. Catalytic activity of the RNase was lost in parallel with formation of the covalently cross-linked dimer.

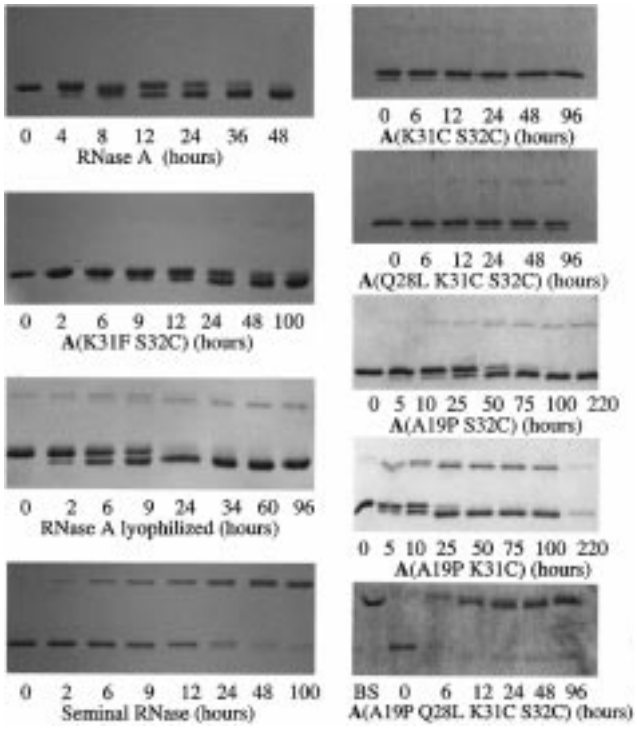

FIGURE 2: SDS−PAGE showing the time-dependent cross-linking of subunits of dimeric RNases with an S-peptide swap using divinyl sulfone. Variants are of RNase A with the substitutions indicated. Time of cross-linking experiment is shown in hours. The upper band (present, for example, in the gel analyzing seminal RNase) arises from swapped dimer (see text); the lower band arises from unswapped dimer or monomer. Both the upper band and lower band lose over time the five carboxyl-terminal amino acids (FDASV), as proven by mass spectrometry (see text). Gels were stained with Coomassie blue. Quantitation is by gel scanning (see text), and the major bands are counted together with their respective cleavage products.

A series of control experiments showed that the appearance of cross-linked band was a reliable quantitative indicator of swapping. Lyophilization of RNase A from acetic acid yields ca. 20% of a dimer held together by swapping of the S-peptide fragments (*27*). Treatment of RNase A samples prepared in this way with DVS also gave ca. 20% covalently joined dimer (Figure 2). Last, purified seminal RNase is a dimer having nearly complete swapping. Cross-linking with DVS gave ca. 85% of the expected dimer (Figure 2), as did cross-linking with difluorodinitrobenzene (data not shown).

These controls established the reliability of DVS cross-linking as a tool for estimating the extent of swap. Interestingly, DVS also caused a time-dependent cleavage of RNase, a reaction that appears to be without precedent. To understand this reaction, the fragments of the cleavage reaction were isolated by HPLC and analyzed. Edman degradation of the large fragment (visible on the gels of

Figure 2) showed that the fragment had the sequence (side-chain-modified K)-E-T-A-A-A, (where the modification on lysine arises from the reaction of the ϵ amino group of lysine with DVS), corresponding to the amino terminus of native RNase. By mass spectrometry, the principal mass of the small peptide was 655.7, corresponding to the peptide sequence FDASV plus a single DVS unit. These results implied that DVS cleaves after His 119 to release the last five amino acids of RNase A, with the amino-terminal amino group of the resulting pentapeptide carrying a DVS modification.

The large fragment arising from the DVS-induced cleavage was also analyzed by mass spectrometry. The mass spectrum of the modified RNase showed a series of peaks separated by 118 Da (the mass of DVS) indicating incorporation of up to 12 DVS units (corresponding to 10 Lys side chains, the amino-terminal amine, plus an additional DVS cross-linking His-12 and His-119). The predominant peak corresponding to the incorporation of seven DVS molecules.

The **A**(K31C) and **A**(S32C) (data not shown) as well as the **A**(K31C S32C) variants, refolded under standard conditions, showed no swapping using the DVS cross-linking assay (Figure 2). Although these proteins form covalent dimers, none form composite active sites under these conditions. Thus, while either Cys-31 or Cys-32 alone is sufficient to support the formation of a covalent dimer, both together are not sufficient to produce an S-peptide swap and a composite active site.

To complete these studies, the corresponding variant was prepared that introduced residues from pancreatic RNase A into a seminal RNase context. The **S**(C31K C32S) variant did not, of course, form a covalent dimer, as expected from the results of Kim and colleagues (*43*). Neither, however, did it form a substantial amount of swapped noncovalent dimer, at least under the standard refolding conditions according to the DVS cross-linking assay (data not shown). These results also show that DVS is a useful tool for examining swapping in RNase variants, giving expected results in control systems and generating coherent data in new systems.

The crystal structure of seminal RNase (*25*) suggests that the orientation of the subunits necessary for S-peptide swap and influencing the parallel/antiparallel orientation of the intersubunit disulfide bridges might be guided by noncovalent interactions between the dimer subunits. In particular, a hydrophobic contact between Leu-28 in one subunit and Leu-28′ in the other might provide this guidance. Evolutionary reconstructions showed that a hydrophobic amino acid was introduced at position 28 early, following the divergence of the seminal RNase lineage from the pancreatic RNase lineage, although the specific hydrophobic amino acid is not conserved (in the Russian antelope saiga, evidently a pseudogene). To explore this hypothesis, the **A**(Q28L K31C S32C) variant was prepared. Again, a catalytically active covalent dimer is formed (Figure 1), confirming results obtained by Raillard (*35*) and Di Donato et al. (*37*). When treated with DVS and reduced, approximately 40% of the product was covalently joined, indicating that the hydrophobic amino acid at position 28 significantly increases the amount of swap (Figure 2), confirming results obtained by Di Donato et al. (*37*). Nevertheless, the predominant form of the dimer retained noncomposite active sites, at least under these refolding conditions.

We then turned to the "hinge region" joining the S-peptide to the rest of the RNase protein, comprising residues 16−22, a rather flexible loop in RNase that adopts different conformations in the swapped and unswapped forms. Parente and D'Alessio (*69*) suggested that the 18−19 peptidyl−proline bond in the cis conformation was important for swapping and perhaps for dimerization. Evolutionary reconstructions suggested that this proline was also introduced early in the evolution of the seminal RNase gene. In this region, however, seminal RNase and pancreatic RNase differ by four amino acids. Two different variants were prepared in the first round of studies, **A**(S16G T17N A19P A20S) and **A**(S16G T17D A19P A20S). Cross-linking with DVS failed to detect swapped dimeric form, at least under the refolding conditions used here (data not shown).

In a second round of mutagenesis, Pro-19 was added to create the **A**(A19P Q28L K31C S32C) variant. This again formed an active covalent dimer (Figure 1), confirming previous results from Raillard (*35*) and Di Donato et al. (*37*). The cross-linking assay suggested that, under standard refolding conditions, ca. 85% of the **A**(A19P Q28L K31C S32C) variant forms a dimer with composite active sites (Figure 2), somewhat higher than that observed with the kinetic measurements reported by Di Donato and colleagues (*37*). This result was therefore confirmed with difluorodinitrobenzene as a cross-linking agent (data not shown).

Evolutionary reconstructions suggest that both Pro-19 and Cys-32 are introduced early in the evolution of the seminal RNase gene, while Cys-31 is introduced very late, perhaps only in the past million years. Therefore, the **A**(A19P S32C), **A**(A19P K31C), and **A**(A19P K31C S32C) variants were examined. Each forms a covalently linked dimer (Figure 1). The three variants gave increasing amounts of S-peptide swap, however, under standard refolding conditions. The first (a possible evolutionary intermediate) gave only a trace of swap (<20%, even at long incubation times). The second (an unlikely evolutionary intermediate) gave more (ca. 30%) (Figure 2). The third (an unlikely evolutionary intermediate) gave the most swap (ca. 40%) under standard refolding conditions (data not shown).

These data do not fully characterize the dimers, however, in particular with respect to their intersubunit disulfide connectivity. In its crystallized state (*24*, *25*), seminal RNase dimer is joined covalently by intersubunit disulfide bonds between Cys-31 of one subunit and Cys-32′ on the other and between Cys-32 of the first subunit and Cys-31′ of the other (the antiparallel disulfide connectivity). Molecular modeling showed that a dimer joined through a 31−31′/32−32′ parallel connectivity is also possible, despite a minor steric interaction of the two loops around residue 91 (Figure 3). Other examples are known of proteins where two alternative intersubunit disulfide connectivities exist under physiological conditions (see, for example, casein) (*70*).

To characterize the connectivity of variants containing Cys residues at both positions 31 and 32 [**A**(K31C S32C), **A**(Q28L K31C S32C), and **A**(A19P Q28L K31C S32C)], each was digested with CNBr. The large fragment containing residues 14−124 interconnected by disulfide bonds was isolated by HPLC, and the purified fragment was subjected to five cycles of Edman degradation. Each fragment contains

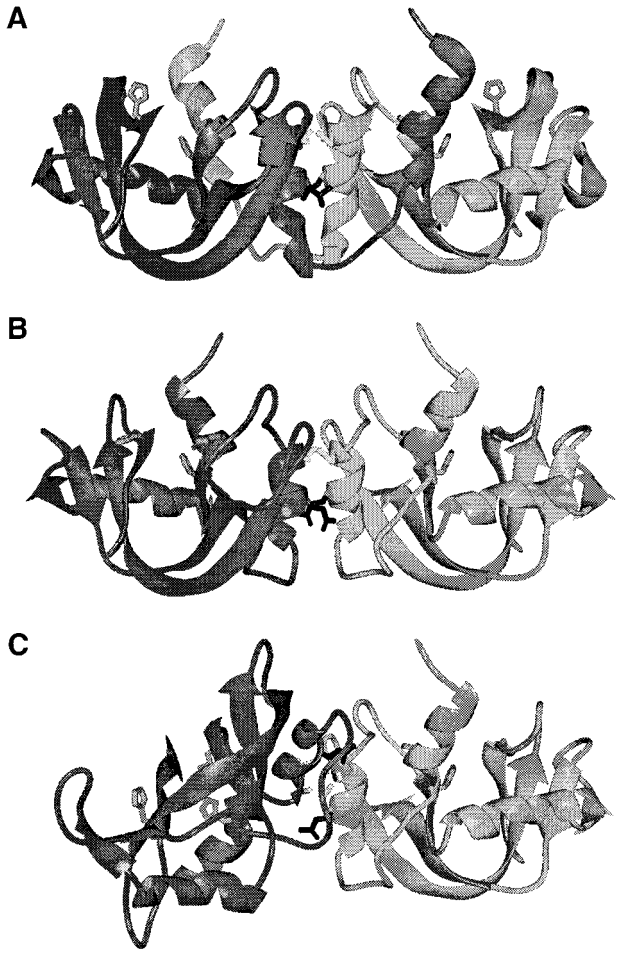

FIGURE 3: Ribbon diagrams representing native seminal RNase with the antiparallel intersubunit disulfide connectivity 31−32′/32−31′ [Brookhaven Protein Data Bank entry 1BSR (*25*)] and with the N-terminal swap (A), models of pancreatic RNase (Brookhaven Protein Data Bank entry 5RSA) monomers superimposed on the seminal RNase structure to retain the antiparallel disulfide connectivity (B), and a model with the left subunit rotated to allow the parallel 31−31′/32−32′ connectivity (C). Residue Leu-28 is highlighted in black. Images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (*97*).

three amino termini, position 14 (Asp, resulting from CNBr cleavage after Met-13), position 80 (Ser, resulting from CNBr cleavage after Met-79), and either position 30 or 31 (Met or Cys, resulting from CNBr cleavage after Met-29 or Met-30, respectively). Positions 80 and 14 release the same products in Edman degradation regardless of the intersubunit disulfide connectivity. From position 31, however, in the parallel connectivity, the first and second cycles release cystine residues. From position 31 in the antiparallel connectivity, the first cycle releases no cystine, while the second cycle releases 2 equiv of cystine.

As this was the first time that any variant of RNase has been completely characterized in this way, and as the experiments were difficult, studies were done to search for factors that might complicate or compromise the analysis. Edman degradation of a Cys−Cys bond where both Cys residues are involved in disulfide bonding might be difficult. The bis(phenylthiohydantoin) (PTH) derivative of cystine can be difficult to quantitate, as it is susceptible to β-elimination to give dehydroalanine, a reaction observed with serine and threonine as well (*71*). Next, cleavage of a Met−Cys bond with CNBr is incomplete when the Cys is involved in a disulfide bond (*72*). Partial cleavage of the Met-30−Cys-31 bond should give rise to a mixture of CCR and MCCR peptides, as the residue preceding Met-30 is also a Met. These products cannot be separated and might confuse the quantitation of the cystine degradation products. Last, disulfide exchange reactions might scramble the connectivity at some point during the analysis. Each of these points is considered below in light of data presented here and examples from the literature.

Disulfide exchange during the analysis was a potential problem. Disulfide exchange was recognized already 40 years ago as a factor complicating the assignment of disulfide links. Ryle and Sanger (*73*) found that, upon treatment with cold concentrated HCl, insulin gave more cystine peptides than could be accounted for by a unique structure, indicating the possibility of disulfide exchange under acidic conditions as well as basic conditions. They investigated this exchange in model systems [cystylbisglycine and bis(2,4-dinitrophenyl)cystine, 10 mL of 12 N HCl and 20 mL of acetic acid, 35 °C, 1 day].

Nevertheless, proteolytic fragmentation and two cycles of manual Edman degradation (TFA, 50 °C, 20 min) was used to assign disulfide connectivity in the ascidian trypsin inhibitor I (*74*) without evidence of rearrangement of the disulfide bonds. Likewise, workers at Amgen (*75*, *76*) exploited automated Edman degradation (Applied Biosystems 470A) to isolate the di-PTH derivative of cystine, reporting a recovery ca. 20% as efficient as that for standard amino acids. The di-PTH-cystine eluted near the PTH derivative of tyrosine. The automated nature of the cycle allows it to be controlled more precisely than manual Edman degradation, permitting quantitation between consecutive cycles. A similar analysis was executed for casein (*70*). This shows that the PTH derivative of cystine can be recovered and quantitated, provided care is taken, even considering its instability (presumably with respect to β-elimination). Bauer and colleagues (*77*) assigned all of the intramolecular disulfide bonds of echistatin, a peptide with four disulfide bonds and 49 amino acids, by incubation for 4 h at 100 °C in 250 mM oxalic acid (first $pK_a = 1.25$). No disulfide exchange was evident, indicating that the acid-catalyzed cleavage of a peptide bond was more rapid than the acid-catalyzed exchange of disulfide bonds under these conditions. The protein includes two disulfide bonds built from adjacent cysteines, as is present in some of the RNase variants described here.

In our experiments, substantial amounts of disulfide exchange would have been detected had it occurred, as it would lead to the release of peptide fragments following CNBr cleavage from the interconnected web. This web was analyzed by HPLC, which isolated the web plus two fragments assigned to residues −1 through 13 and 1 through 13 (the amino-terminal methionine was incompletely cleaved, and this fragment is not joined to the rest of the protein by a disulfide) as the major products. Only minor amounts of unassigned product were observed, amounting to perhaps 10% of the total. This places an upper limit of 10% on disulfide exchange during the analysis.

Difficulties in CNBr cleavage proved to be more significant. In RNase A, it is known that the Met−Met cleavage in RNase A is incomplete when the protein is treated with

CNBr under specific conditions. For example, Link and Stark (*78*) reported that the recovery of free homoserine lactone did not exceed 33% with RNase A and attributed the low yield of cleavage to the assumption that cleavage of the Met-29−Met-30 bond yielded an N-terminal methionine, which itself did not undergo CNBr cleavage. This assumption was supported by several reports that amino-terminal methionines are cleaved inefficiently by CNBr (*79*−*81*). This, in turn, was interpreted as evidence that the positively charged amino-terminal ammonium ion hindered the formation of the sulfonium ion, just four bonds away, postulated to be an intermediate in the cleavage cycle. In a different system, Doyen and Lapresle (*72*) showed that the Met-123−Cys-124 bond in albumin, with the Cys participating in a disulfide bridge, was only partially cleaved by CNBr (75% formic acid, room temperature, 22 h). The fraction not cleaved was recovered as a homoserine−cystine linkage.

The incomplete CNBr cleavage of the Met-29−Met-30−Cys-31 segment in the RNase variants examined here was of special concern, as a mixture of MCCR- and CCR- amino termini would confuse the signal arising from consecutive Edman degradations. In the first cycle, only the parallel CCR- peptide would yield di-PTH-cystine, as before. However, the cystine recovered from the second cycle would contain contributions from the antiparallel CCR- peptide (2 equiv), from the parallel CCR- peptide (1 equiv), and from the parallel MCCR- peptide (1 equiv). In the third Edman cycle, the antiparallel MCCR- peptide would contribute 2 equiv of di-PTH-cystine, while the parallel MCCR- peptide would contribute 1 equiv.

The relative amounts of MCCR- and CCR- peptides arising via CNBr cleavage of the parallel and antiparallel connectivity are difficult to predict. If the rate of reaction of the peptide with CNBr to generate the sulfonium salt was rapid with respect to cleavage, then both the Met−Met and Met−Cys bonds should be cleaved, and the amount of the CCR- peptide should be higher, because the formation of the sulfonium is presumed to commit the Met to further reaction (*82*). If, however, the generation of sulfonium salt was slower than cleavage, the relative amount of MCCR- would be higher. Further, the relative amounts of MCCR- and CCR- peptides would be sensitive to the concentration of CNBr present in the reaction mixture. Finally, the antiparallel connectivity places Arg-33 nearer to Met-30 than the parallel. If Coulombic interactions indeed influence the rate of formation of sulfonium, the fraction of MCCR- formed in the CNBr cleavage reaction might be higher in the antiparallel connectivity than in the parallel connectivity.

These considerations required that independent measurements be made for several of the variables to permit a quantitative analysis of the disulfide connectivity. To estimate the relative amounts of MCCR- and CCR- peptides, the product of the CNBr cleavage was reduced and the Cys residues were reacted with vinylpyridine to give pyridinylethyl derivatives. These are readily quantitated in an Edman degradation. Successive Edman cycles on an automated sequenator permitted the estimation of the relative amounts of MCCR- and CCR- peptides, with estimates of the lag and the repetitive cycle yields. With the **A**(K31C S32C), **A**(Q28L K31C S32C), and **A**(A19P Q28L K31C S32C) variants, the MCCR:CCR ratios were calculated to be 0.6, 0.8, and 0.9, respectively.

As an example, Figure 4 shows the results of the first three Edman cycles of the **A**(Q28L K31C S32C) variant. The di-PTH derivatives of cystine are clearly identified, coeluting with the PTH derivative of Tyr. To correct for overlap between the diPTH-cystine peak and the peak arising from PTH-tyrosine, the Tyr contribution was determined from Edman degradation of the reduced pyridinylethylated variant and was subtracted. The amounts were corrected for a repetitive cycle yield of 80%, and the fraction of antiparallel disulfide connectivity was calculated to be approximately 70%.

In this calculation, four variables, the amounts of parallel and antiparallel forms of the CCR- protein, and the amounts of parallel and antiparallel forms of the MCCR- protein, must be found with four measurements, the amount of cystine in cycles 1, 2, and 3 and the MCCR:CCR ratio. In principle, since no species can have a concentration less than zero, the solution of the four independent equations containing four variables was also bounded by a fifth constraint. With the **A**(A19P Q28L K31C S32C) variant, the solutions found that the amount of parallel MCCR- peptide was close to zero. This is consistent with the notion that the rate of formation of sulfonium salt was fast compared to cleavage in this case, that formation of sulfonium committed the peptide to cleavage (i.e., sulfonium formation is not reversible), and that the antiparallel orientation had more difficulty forming the sulfonium, perhaps because of the proximity of Met-30 to Arg-33′.

The results showed that in the variant **A**(K31C S32C) the parallel connectivity predominated, while in variants **A**(Q28L K31C S32C) and **A**(A19P Q28L K31C S32C) the antiparallel disulfide connectivity predominated (Table 1). There appears to be a general correlation in these three variants between the fraction of antiparallel disulfide connectivity and the extent of S-peptide swap. Variant **A**(A19P Q28L K31C S32C) displays the largest amount of S-peptide swap in this group of variants, while variant **A**(Q28L K31C S32C) displays more S-peptide swap than variant **A**(K31C S32C) (Table 1).

We then asked whether the introduction of additional cysteines into the polypeptide chain creates kinetic barriers to refolding of the monomer. For variant **A**(A19P Q28L K31C S32C) and **A**(K31C S32C), as well as seminal RNase, the rate of appearance of catalytic activity displayed a lag (ca. 1 h) during refolding in air at 4 °C in the absence of glutathione (Figure 5). RNase A displays no lag under these conditions. This lag disappears when the experiments are repeated in the presence of glutathione (1 mM GSH; 0.2 mM GSSG), at concentrations approximating those believed to be present in the endoplasmic reticulum (*83*). This result strongly suggests that the lag arises from the formation of incorrect disulfide bonds, presumably involving the extra Cys residues, during the refolding process. These presumably act as kinetic traps during refolding, diverting the protein into metastable forms with incorrect disulfide connectivity and thereby slowing the folding process.

Finally, to learn more about the state of intermolecular disulfide bonds formed by seminal RNase under physiological conditions, the proteins in crude seminal plasma were resolved by PAGE under both nonreducing denaturing and reducing denaturing conditions. The proteins were visualized by Western blotting using antibody raised against denatured

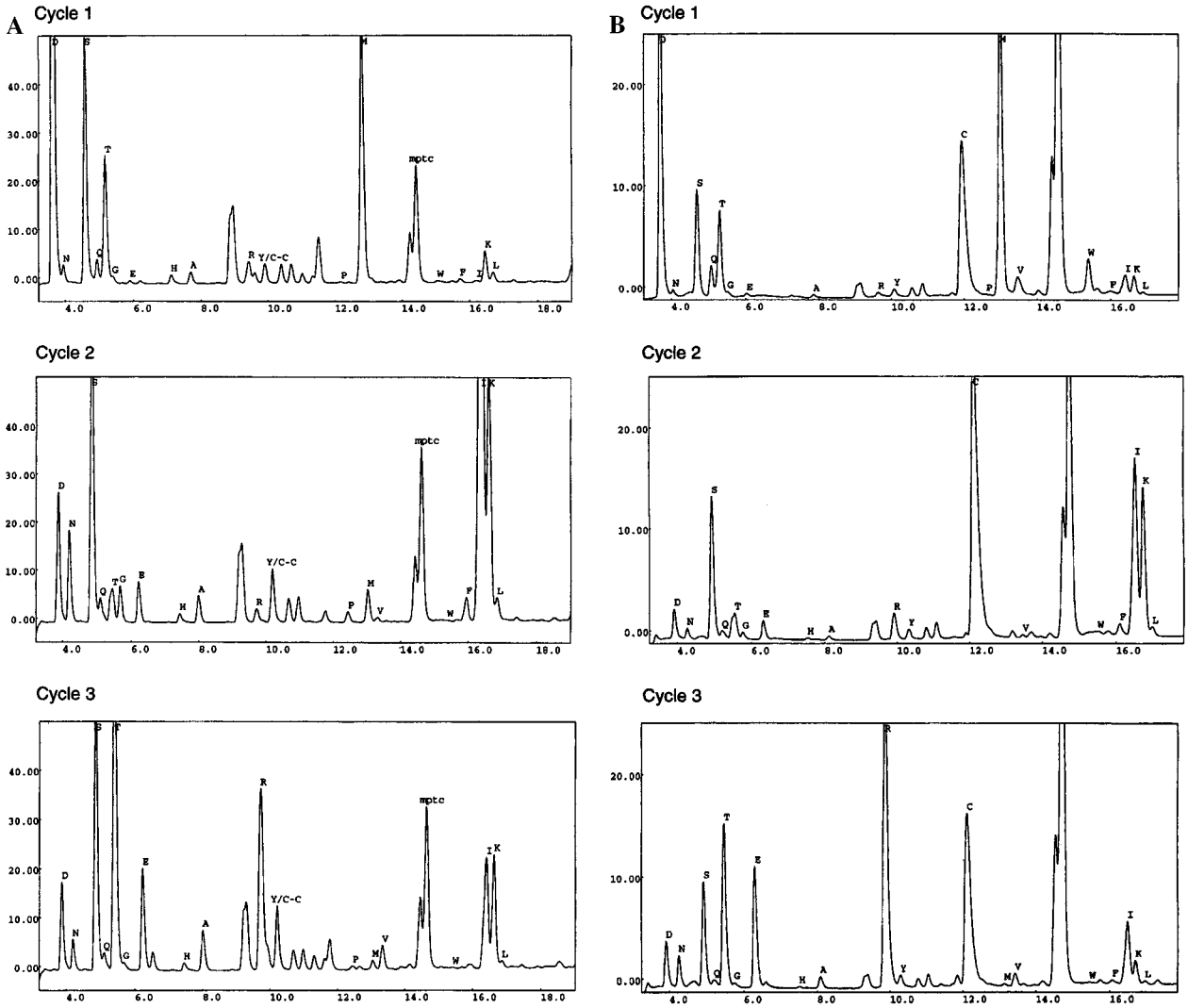

FIGURE 4: (A) HPLC traces showing phenylthiohydantoin derivatives emerging from variant A(Q28L K31C S32C) following cleavage with CNBr and then one, two, and three rounds of Edman degradation. Vertical axis is intensity measured by UV spectroscopy; horizontal axis is retention time (minutes); peaks are indicated by the single-letter code. The PTH-Tyr and PTH-cystine come together. See Results and Discussion for more details. (B) Samples are the same as in panel A but reduced and treated with vinylpyridine. These data provide the background measurements for interpreting the data in panel A. See Results and Discussion for more details.

seminal RNase. To be certain that disulfide exchange did not occur during the analysis, the experiments were repeated under different conditions. Duplicate samples of seminal plasma were obtained with and without *N*-ethylmaleimide, which should react with free thiols, thereby blocking thiol-catalyzed disulfide exchange. Parallel samples with and without *N*-ethylmaleimide were loaded with and without preincubation at 95 °C (5 min); heating should accelerate any residual disulfide exchange that might exist. As can be seen from Figure 6, the pattern of seminal RNase conjugates is unaltered in all four lanes. Were disulfide exchange occurring during the analysis, one would expect different patterns with different analytical conditions. Thus, these results suggest that, under physiological conditions, approximately 50% of seminal RNase is in the form of conjugates with other proteins.

DISCUSSION

The evolution of homodimeric proteins from monomeric ancestors during divergent evolution presents a paradox to molecular evolution. Dimers are typically joined by contacts having surface areas ranging from 700 Å^2 to as much as 5000 Å^2 in area (*84*). Typical dimer contacts resemble the interiors of proteins in terms of hydrophobicity and, to some extent, packing. They generally involve more than one amino acid from each subunit in contact with the other. Taking a fully folded, functional monomer as a starting point, this implies that a hydrophobic contact surface must evolve from a (normally) hydrophilic solvent-exposed surface. This implies an evolution via multiple mutations. If the intermediate stages in this evolution do not confer survival value, then the question arises: If dimer formation requires multiple, simultaneous substitutions, where individual substitutions by themselves serve no selected function (otherwise a gradual evolution of dimeric structure would be admissible by Darwinian theory), how is it possible to evolve dimers (*85*)? This is a specific case of a quite general question in molecular evolution: How can function involving multiple amino acid residues emerge by a process that predominantly involves alteration of single amino acid residues?

Eisenberg and his group (*31*) proposed a solution for evolving dimers that avoids this paradox. Their solution was

Table 1: Catalytic Activity and Dimer Structure of RNase Variants[a]

	k_{cat}/K_M (% of RNase A)	swap (%)	31−32′/32−31′ antiparallel connectivity (%)
RNase A expressed	100	0	no intersubunit disulfides
RNase A dimer lyophilized	nd	20	no intersubunit disulfides
A(K31C)	51	0	antiparallel not possible
A(S32C)	59	0	antiparallel not possible
A(K31F S32C)	91	0	antiparallel not possible
A(Q28L K31F S32C)	97	0	antiparallel not possible
A(A19P K31C)	51	30	antiparallel not possible
A(A19P S32C)	58	20	antiparallel not possible
A(A19P K31C S32C)	nd	40	nd
A(K31C S32C)	140	12	35 ± 20
A(Q28L K31C S32C)	102	40	75 ± 10
A(A19P Q28L K31C S32C)	106	85	75 ± 15
A(Q28L K31C)	79	13	antiparallel not possible
A(S16G T17N A19P A20S)	89	0	no intersubunit disulfides
S(C31K C32S)	57	0	no intersubunit disulfides
bovine seminal RNase (isolated)	59	85	nd

[a] Kinetic parameters were calculated from Lineweaver−Burk analysis of data collected with UpA as substrate and reported relative to RNase A. RNase A was expressed in *E. coli* and contains an N-terminal Met. Seminal RNase was isolated from bull seminal plasma by the procedure of Tamburrini and co-workers (*48*). The k_{cat} values for the dimers were calculated per active site. Swap was estimated by the extent of dimer band cross-linked with divinyl sulfone, scanned, and analyzed using ImageQuant 3.3 (Molecular Dynamics) [Figure 2, where data for **A**(K31C), **A**(S32C), and **A**(A19P K31C S32C) are not shown]. The disulfide bond connectivity of variants having Cys residues at both positions 31 and 32 were determined by CNBr fragmentation followed by Edman degradation, as discussed in Materials and Methods and Results. Antiparallel disulfide connectivity joins 31 from one subunit with Cys-32′ of the other and Cys-32 from one subunit with Cys-31′ of the other. Variants with only a single Cys either at position 31 or at position 32 can form only the parallel intersubunit disulfide connectivity (Cys-31 of one subunit paired with Cys-31′ of the other or Cys-32 of one subunit paired with Cys-32′ of the other). nd = not determined.

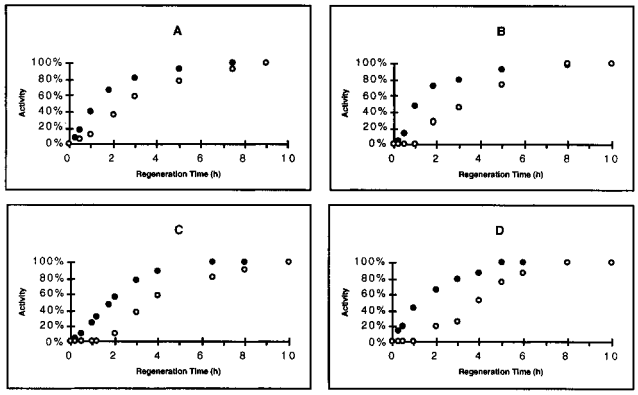

FIGURE 5: Rate of reactivation of fully reduced and denatured RNase variants at pH 7.8, 4 °C. Reoxidation in oxygen (○) and in the presence of a glutathione redox buffer (GSSG, 0.2 mM; GSH, 1 mM) (●). Catalytic activity was measured against 30 μM UpA at 25 °C. (A) RNase A; (B) BS−RNase; (C) A(K31C S32C); (D) A(A19P Q28L K31C S32C).

based on their analysis of a crystal structure of diphtheria toxin, which is sometimes monomeric and sometimes dimeric as it functions. Diphtheria toxin is endocytosed as a monomer. After it is endocytosed, however, it undergoes a conformational change in the acidic environment of the lysosome that allows the protein to leave the lysosome and enter the cytoplasm. This conformational change involves a swap of a portion of the polypeptide chain between two monomeric units to form a noncovalent dimer. In the swap,

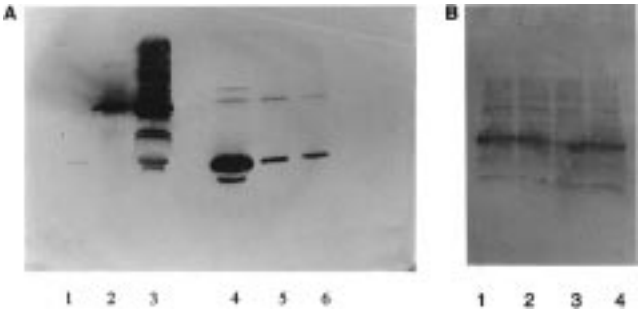

FIGURE 6: (A) Crude bovine seminal plasma resolved by PAGE under both nonreducing denaturing and reducing denaturing conditions, visualized by Western blotting using antibody raised against denatured bovine seminal RNase (from rabbit). Lane 1, 1 μg of monomeric RNase A (Boehringer Mannheim) [the small amounts of immunoreactive bands appearing at high molecular weight are well-known in such samples and presumably arise due to a small amount of misfolded RNase A that is forming an intermolecular disulfide bond]; lane 2, 0.5 μg of purified bovine seminal RNase dimer; lane 3, 0.5 μL of bovine seminal fluid, crude; lane 4, 0.5 μL of bovine seminal fluid, incubated with 100 mM DTT for 5 min at 95 °C; lane 5, 0.5 μg of purified bovine seminal RNase dimer, heated with DTT as in lane 4; lane 6, 1 μg of purified RNase A, heated with DTT as in lane 4. (B) Crude bovine seminal plasma resolved by PAGE under nonreducing denaturing conditions, visualized by Western blotting as in panel A. Lane 1, with heating and with *N*-ethylmaleimide (NEM); lane 2, without heating (5 min, 95 °C) and with *N*-ethylmaleimide (NEM); lane 3, with heating and without *N*-ethylmaleimide (NEM); lane 4, without heating and without *N*-ethylmaleimide (NEM).

a chain from one subunit replaces the corresponding chain in the fold of the second, while the replaced chain from the second fits into the site vacated in the fold of the first subunit.

Bennett and colleagues (*31*) suggested that domain swapping might be the first step in the evolution of dimeric proteins from monomers. The contact between the swapped peptide chain and the second subunit resembles the contact between the chain and its own subunit. Thus, there is no need to evolve a new interface for forming a dimer. Implicit in this model is the notion that swapping is a natural event in any case and can be induced by even a single amino acid substitution. Under the model, once the swap has occurred, a stepwise evolution of the hydrophobic contact site, introducing one hydrophobic amino acid at a time, will stabilize the dimeric structure gradually, presumably improving the survival value of the protein as well by a gradual process entirely acceptable to Darwinian theory.

The Eisenberg "swap early" model for the evolution of quaternary structure postulates that swapping of a key secondary structural element is the first process in the evolution of a dimer. In the evolution of the dimeric structure of seminal RNase, the model is made more plausible by the fact that pancreatic RNases form dimers through this swap, albeit under nonphysiological conditions (*26*, *27*). It is not difficult to imagine how this (presumably) nonfunctional behavior might be exploited during divergent evolution to create a dimeric structure following the Eisenberg model. In this model, the swap might then be enhanced by introduction of a Pro at position 19, stabilized via the introduction of a hydrophobic contact site at position 28, and then fixed through introduction of cysteines at positions 31 and 32 to permit a covalent bond between subunits.

D'Alessio and his group reported that adding *two* cysteines, at positions 31 and 32, in an RNase A background

also yields a protein capable of forming dimers, some swapped and some not (*40*, *44*) (see also ref *35*). This offers an alternative "covalent dimer early" model for the evolution of dimers, one where two cysteine residues were introduced first, with the swap evolving later.

The reconstructed evolutionary history of the seminal RNase family strongly suggests, however, that cysteines were *not* introduced at positions 31 and 32 at the same time (*63*−*65*). Rather, the codon for a single cysteine at position 32 was introduced early in the history of the seminal RNase family, while that for the cysteine at position 31 was introduced only very recently. Thus, neither the D'Alessio nor the Bennett et al. model is consistent with the reconstructed evolutionary history.

Reconstruction of the sequences of ancestral RNases using principles of maximum parsimony and maximum likelihood suggests that the ancestral protein at the point where the seminal and pancreatic lineages diverged contained Cys at neither position 31 nor 32, had no ability to form covalently linked dimers, and showed no propensity to form noncovalent dimers greater than that observed with bovine pancreatic RNase A (*23*). From this it can be inferred that modern dimeric seminal RNase evolved from a monomeric protein. Parsimony analysis of the sequences of various seminal RNase genes in various organisms diverging after the formation of the seminal RNase gene by duplication permits us to identify the order in which residues considered important for dimeric structure (Pro-19, a hydrophobic residue at position 28, Cys-31, and Cys-32) were introduced. Cys-32 appeared in the seminal lineage immediately after it was created and was conserved subsequently. Pro-19 was also introduced immediately and was also conserved. A hydrophobic residue was present at position 28 as early as detectable in the seminal lineage as well but was not conserved. Cys-31 was introduced late, perhaps in the past 1 million years, after the divergence of ox from Cape buffalo.

Guided by the evolutionary history of the seminal RNase family, the experiments reported here examine the mutations likely to be relevant to dimer formation in the same order as they were introduced in the history of the seminal RNase gene family. They show that introduction of a *single* cysteine at position 32 is a direct route to a dimer without requiring multiple simultaneous substitutions. The **A**(A19P S32C) variant, containing two relevant substitutions that emerged early in the seminal RNase gene is also a dimer, but with substantial amounts of swap (Table 1), despite the fact that it is joined by a 32−32′ intersubunit disulfide connectivity. This confirms the role for Pro-19 in encouraging the swap, as suggested by D'Alessio and co-workers (*69*), here in an evolutionary context. As the order in which Pro-19 and Cys-32 were introduced cannot be determined given available data, Cys-32 need not have formed the dimer first, to have the dimeric structure enhanced by the swap later. Pro-19 might have been introduced first. However, all experiments where Pro-19 has been introduced without a cysteine at either position 31 or 32 have failed to detect swap yielding a noncovalent dimer.

Continuing with variants suggested by the reconstructed evolutionary history, a hydrophobic amino acid (Phe), introduced at position 31 early in the seminal RNase lineage, was found not to increase swap. A hydrophobic residue (Leu) at position 28, however, increases the amount of antiparallel disulfide connectivity and the amount of swap in variants that contain cysteines at positions 31 and 32. This is the one residue that appears (from the crystal structure) to form a specific contact site in addition to the swap and the disulfide bonds themselves.

Together these results suggest a model for the origin of dimeric structure different from both the model of Bennett et al. (*31*) and the model of D'Alessio and co-workers (*40*). In this model, the first seminal RNase gene encoded a protein with a single cysteine at position 32 but none at position 31. This was capable of forming a dimer joined by only a single disulfide bond. This dimer has, of course, a "parallel" connectivity. At essentially the same time, Pro-19 and a hydrophobic residue at position 28 were introduced. This implies that early versions of the seminal RNase gene encoded proteins capable of some swap, suggested by the 20% swap of the **A**(A19P S32C) variant. Some 30 million years later, a second Cys was introduced at position 31. This permitted antiparallel connectivity (given a hydrophobic residue at position 28), enabled substantial amounts of swap, and rigidly fixed a specific orientation of the active sites relative to each other. As shown in the following paper (*86*), this orientation leads to higher catalytic activity against duplex nucleic acids, already partially present in the common ancestor of seminal and pancreatic RNase, is increased when residues introduced early in the evolutionary history of seminal RNase are introduced into the pancreatic gene (*86*), and increases substantially in the last episode of evolution of seminal RNase. This model is consistent with the reconstructed evolutionary history of the seminal RNase gene family.

While the reconstructed evolutionary history provides strong evidence that this was the actual sequence of events leading to the seminal RNase dimer, and while the experiments reported here provide information about the derived features of the tertiary and quaternary structure that follow from the change in primary structure (including the nature of the disulfide connectivity and the extent of the swap), they do not answer the "Why?" questions that connect this history with the evolution of behavior and function. Several of the paradoxes of this model must be discussed.

First, addition of a single Cys residue to engineer a new quaternary structure via an intersubunit disulfide bond is a strategy suited only for extracellular proteins. Introduction of extra cysteines raises issues even for extracellular proteins, however. These relate to the rate of monomer folding, the rate of dimer formation, and the specificity of dimer formation in an environment containing other proteins and peptides that present free thiol groups.

For example, RNase A has eight cysteine residues forming four intrasubunit disulfide bonds in the native structure. These permit 104 incorrect disulfide connectivities in the fully oxidized protein and still more when the protein is incompletely oxidized. Each of these is a possible kinetic trap during refolding. Addition of two cysteines increases the number of possible intramolecular disulfide bonding patterns to 945 (in fully oxidized protein), 944 of which are incorrect. While additional cysteines might help create a dimeric protein, at the same time they may create a protein that has difficulties folding.

In fact, introduction of two additional cysteines to RNase A at positions 31 and 32 creates a lag in the refolding in

vitro (Figure 5). This almost certainly arises from the formation of incorrect disulfide linkages within the monomer, as the lag disappears in the presence of total glutathione at 1.2 mM concentrations (GSSG:GSH = 1:5). Introduction of residues presumed to be desirable for one property (dimeric structure) is problematic for another (rapid folding).

Do these results imply that the features of the sequence required to form dimers (additional cysteines) are incompatible with rapid folding in vivo as well? Glutathione is present in the endoplasmic reticulum at low millimolar concentrations (*83*), where the monomer presumably folds, as are enzymes that catalyze disulfide exchange (*87*). Thus, the lag in refolding observed in vitro in these experiments need not encumber seminal RNase in vivo. This observation does not solve the problem, however. It is well-known that seminal RNase does not form a dimer in the presence of glutathione (*69*, *88*). Rather, when seminal RNase is incubated with glutathione, the predominant product is the monomer with Cys-31 and Cys-32 blocked as mixed disulfides with glutathione. This result creates another paradox: Glutathione is necessary for the rapid folding of RNases that *can* form dimers through the existence of extra cysteines, but glutathione blocks the formation of the dimer.

To resolve this paradox, the refolding of RNase variants at concentrations of glutathione (0.12 mM) approximating the lower concentration of glutathione in seminal plasma [0.1 mM, in human (*89*) or 0.20−0.46 mM in various bovids (*84*); possibly 0.026 mM, see ref *90*] was examined. With the **A**(A19P Q28L K31C S32C) variant, dimer and monomer were both formed in a ratio of ca. 1:1 (data not shown). This suggests that at low concentrations of glutathione, seminal RNase monomer with its external cysteines blocked as mixed disulfides with glutathione (generated in the cell, where concentrations of glutathione are higher) might be converted to the native dimer in the seminal plasma itself, presumably via a disulfide exchange mechanisms without reduction of the intrasubunit disulfide bonds.

The statement that seminal RNase fails to form dimer decisively in competition with glutathione at physiological concentrations is, of course, equivalent to the statement that the specific affinity of a seminal RNase monomer for another seminal RNase monomer is not very high. Thus, the contact regions between seminal RNase dimer has none of the complementarity leading to high specific affinity found in classical dimer contact regions.

This could be, of course, simply a manifestation that seminal RNase is an example of "evolution in progress" (*40*), with a more specific dimer contact site yet to emerge. Alternatively, it could indicate that the homodimer known in the crystal structure is not the form of seminal RNase that has physiological relevance. This thought prompted a closer examination of the quaternary structure under physiological conditions in seminal plasma, where other proteins might also present free cysteines on their surface. A Western blot analysis of crude seminal plasma, using an antibody against seminal RNase, showed that the homodimer of seminal RNase represents only part of the RNase in the seminal plasma. The remainder is covalently conjugated with many other proteins (Figure 6), presumably through intermolecular disulfide bonds, together with protein that runs as expected for a glutathione-blocked monomer. Reduction with dithiothreitol disrupts these conjugates to yield monomeric seminal RNase. Identical results (to the limits of detection) were obtained in fresh samples of seminal plasma immediately treated with *N*-ethylmaleimide to block artifacts that may arise by disulfide exchange. This suggests that seminal RNase as it actually exists under physiological conditions is not exclusively the "native" dimer identified by crystallography (*25*).[3]

The fact that heterodimers and heterotrimers between seminal RNase and other proteins in the seminal plasma exist under physiological conditions requires that such structures be discussed when addressing issues relating to the biological function of seminal RNase. Such discussion has been largely absent from the literature on seminal RNase to date. Perhaps relevant to this is the possibility that the adjacent cysteines at positions 31 and 32 might form an intramolecular disulfide bond. As this will severely distort the conformation of the helix, the resulting disulfide bond should be a strong oxidizing agent (although not, of course, as strong as dioxygen). Thus, one would expect RNase containing an intramolecular 31−32 disulfide bond to rapidly form mixed disulfides when encountering a thiol or a dimer when encountering a seminal RNase monomer with a free SH group. Disulfides between adjacent Cys residues in a dipeptide are known both in model systems (*94*, *95*) and in natural enzymes, where they may play functional roles (*96*).

The evolutionary reconstructions used to guide the experiments in this paper are, of course, reconstructions at the level of the gene. The presence of a gene for a seminal RNase in the genome of an ancient organism does not, of course, require the presence of the protein in that organism, much less the presence of the protein in a particular tissue. The gene need not be expressed, or might be expressed in different tissues from those where it is expressed in the modern world. Only if the gene is expressed are the properties of the encoded protein significant; only an expressed gene for seminal RNase be under selective pressure throughout this episode of evolutionary history.

Here, the evidence is uncertain. The seminal RNase gene is an inexpressible, or nonfunctional, pseudogene in many contemporary artiodactyls (*63*, *65*). A parsimony analysis that does not consider sequences in detail infers from this fact that the ancestral genes were also pseudogenes. To assume otherwise would require that several active genes present in ancestral organisms independently lost function in many derived lineages.

A parsimony analysis that considers sequence data draws a different conclusion, however. Analysis of the sequences of a variety of the pseudogenes for seminal RNase from modern organisms fails to identify a single lesion that can be placed by parsimony in all of the ancestral seminal RNase genes. For example, the deletions that disrupt several of the seminal RNase genes (*65*) lie at different points in the

[3] These results are reminiscent of classical work that reports that seminal plasma contains more than one RNase (*91*) as resolved by SE-Sephadex chromatography. We are indebted to D'Alessio for calling this literature to our attention. Subsequently, Reddy and colleagues (*92*) reported a new pyrimidine-specific ribonuclease from bovine seminal plasma that is active on both single- and double-stranded polyribonucleotides and that can distinguish between Mg^{2+}-containing and Mg^{2+}-depleted naturally occurring RNAs. This report was disputed by D'Alessio and colleagues (*93*), and subsequent literature contains no further discussion of this matter. At least some of this controversy might reflect the heterodimers reported here.

protein; they cannot be placed by parsimony in any of the ancestral genes leading to seminal RNase from *Bos taurus*. Nor does parsimony suggest that the evolutionary intermediates were inactive through the loss of His-12, essential for catalytic activity. His-12 is replaced by Leu and Ser in the giraffe and sheep seminal RNase pseudogenes but is present in others. Thus, parsimony analysis reconstructs His at position 12 in all ancestral proteins.

In the absence of additional data, we must entertain the possibility that the seminal RNase gene did have a function through its evolutionary history, a function that was independently lost in giraffe, sheep, and other species of artiodactyls. As discussed elsewhere (Trabesinger-Rüf et al., manuscript in preparation), analysis of the ratio of expressed to silent substitutions during the divergent evolution of the gene suggests that the seminal RNase gene was in fact expressed and under purifying selection throughout much of its evolutionary history. If this is so, then virtually all of the properties of the proteins discussed above have potential physiological relevance. To understand these requires, however, that a better understanding be obtained for the relationship between structure and behavior. This is the subject of the accompanying paper (*86*).

ACKNOWLEDGMENT

We are indebted to Dr. J. Stackhouse for much help and many discussions, to Marianthi Kalamaraki and Sara Mostad for help in constructing two of the mutants, to Dr. Kai Johnsson for helpful comments on how to elucidate the intersubunit disulfide connectivity, and to the Schweizerische Besamungsstation Buetschwil for supplying bovine seminal plasma.

REFERENCES

1. Benner, S. A., Trabesinger-Rüf, N., and Schreiber, D. R. (1998) manuscript in preparation.
2. Allemann, R. K., Presnell, S. R., and Benner, S. A. (1991) *Protein Eng. 4*, 831−835.
3. Richards, F. M., and Wyckoff, H. W. (1971) *Enzymes 4*, 647−806.
4. Blackburn, P., and Moore, S. (1982) *Enzymes 15*, 317−434.
5. Beintema, J. J., Fitch, W. M., and Carsana, A. (1986) *Mol. Biol. Evol. 3*, 262−275.
6. Nambiar, K. P., Stackhouse, J., Stauffer, D. M., Kennedy, W. G., Eldredge, J. K., and Benner, S. A. (1984) *Science 223*, 1299−1301.
7. Benner, S. A. (1988) *FEBS Lett. 233*, 225−228.
8. Benner, S. A., and Allemann, R. K. (1989) *Trends Biochem. Sci. 14*, 396−397.
9. D'Alessio, G., Di Donato, A., Parente, A., and Piccoli, R. (1991) *Trends Biochem. Sci. 16*, 104−106.
10. Barnard, E. A. (1969) *Nature 221*, 340−344.
11. Strydom, D. J., Fett, J. W., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., and Vallee, B. L. (1985) *Biochemistry 24*, 5486−5494.
12. Rosenberg, H. F., Tenen, D. G., and Ackerman, S. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4460−4464.
13. Barker, R. L., Loegering, D. A., Ten, R. M., Hamann, K. J., Pease, L. R., and Gleich, G. J. (1989) *J. Immunol. 143*, 952−955.
14. Ardelt, W., Mikulski, S., and Shogen, K. (1991) *J. Biol. Chem. 266*, 245−251.
15. Okabe, Y., Katayama, N., Iwama, M., Watanabe, H., Ohgi, K., Irie, M., Nitta, K., Kawauchi, H., Tanayanagi, Y., et al. (1991) *J. Biochem. (Tokyo) 109*, 786−790.
16. D'Alessio, G., and Leone, E. (1963) *Biochem. J. 89*, 7P.
17. Dostal, J., and Matousek, J. (1973) *J. Reprod. Fertil. 33*, 263−274.
18. Soucek, J., and Matousek, J. (1981) *Folia Biol. Praha 27*, 334−345.
19. Soucek, J., Hrubá, A., Paluska, E., Chudomel, V., Dostál, J., and Matousek, J. (1983) *Folia Biol. Praha 29*, 250−261.
20. Soucek, J., Chudomel, V., Potmesilova, I., and Novak, J. T. (1986) *Nat. Immun. Cell Growth Regul. 5*, 250−258.
21. Matousek, J. (1973) *Experientia 29*, 858−859.
22. Vescia, S., Tramontano, D., Augusti-Tocco, G., and D'Alessio, G. (1980) *Cancer Res. 40*, 3740−3744.
23. Jermann, T. M., Opitz, J. G., Stackhouse, J., and Benner, S. A. (1995) *Nature 374*, 57−59.
24. Capasso, S., Giordano, F., Mattia, C. A., Mazzarella, L., and Zagari, A. (1983) *Biopolymers 22*, 327−332.
25. Mazzarella, L., Capasso, S., Demasi, D., di Lorenzo, G., Mattia, C. A., and Zagari, A. (1993) *Acta Crystallogr. 49*, 389−402.
26. Richards, F. M., and Vithayathil, P. J. (1959) *J. Biol. Chem. 234*, 1459−1465.
27. Fruchter, R. G., and Crestfield, A. M. (1965) *J. Biol. Chem. 240*, 3868−3874.
28. Karlsson, C., Jornvall, H., and Hoog, J. O. (1991) *Eur. J. Biochem. 198*, 761−765.
29. Schlunegger, M. P., and Grütter, M. G. (1992) *Nature 358*, 430−434.
30. Daopin, S., Piez, K. A., Ogawa, Y., and Davies, D. R. (1992) *Science 257*, 369−373.
31. Bennett, M. J., Choe, S., and Eisenberg, D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3127−3131.
32. Chothia, C., and Lesk, A. M. (1986) *EMBO J. 5*, 823−826.
33. Grütter, M. G., Weaver, L. H., and Matthews, B. W. (1983) *Nature 303*, 828−831.
34. Benner, S. A. (1989) *Chem. Rev. 89*, 789−806.
35. Raillard, S. A. (1993) ETH Dissertation 10261, Swiss Federal Institute of Technology, Zürich, Switzerland.
36. Di Donato, A., Cafaro, V., de Nigris, M., Rizzo, M., and D'Alessio, G. (1993) *Biochem. Biophys. Res. Commun. 194*, 1440−1445.
37. Di Donato, A., Cafaro, V., Romeo, I., and D'Alessio, G. (1995) *Protein Sci. 4*, 1470−1477.
38. Creighton, T. E. (1977) *J. Mol. Biol. 113*, 329−341.
39. Rothwarf, D. M., and Scheraga, H. A. (1993) *Biochemistry 32*, 2671−2679.
40. D'Alessio, G. (1995) *Nat. Struct. Biol. 2*, 11−13.
41. Stackhouse, J., Presnell, S. R., McGeehan, J. M., Nambiar, K. P., and Benner, S. A. (1990) *FEBS Lett. 262*, 104−106.
42. Beintema, J. J. (1987) *Life Chem. Rep. 4*, 333−389.
43. Kim, J. S., Soucek, J., Matousek, J., and Raines, R. T. (1995) *J. Biol. Chem. 270*, 10525−10530.
44. Di Donato, A., Cafaro, V., and D'Alessio, G. (1994) *J. Biol. Chem. 269*, 17394−17396.
45. Trautwein, K., and Benner, S. A. (1991) *FEBS Lett. 281*, 275−277.
46. Trautwein-Fritz, K. (1991) ETH Dissertation 9613, Swiss Federal Institute of Technology, Zürich, Switzerland.
47. Ipata, P. L., and Felicioli, R. A. (1968) *FEBS Lett. 1*, 29−31.
48. Tamburrini, M., Piccoli, R., De Prisco, R. D. D. A., and D'Alessio, G. (1986) *Ital. J. Biochem. 35*, 22−32.
49. Presnell, S. R., and Benner, S. A. (1988) *Nucleic Acids Res. 16*, 1693−1702.
50. Seed, B. (1983) *Nucleic Acids Res. 11*, 2427−2447.
51. Allemann, R. K. (1989) ETH Dissertation 8804, Swiss Federal Institute of Technology, Zürich, Switzerland.
52. Belev, T. N., Singh, M., and McCarthy, J. E. G. (1991) *Plasmid 26*, 147−150.
53. McGeehan, G. M., and Benner, S. A. (1989) *FEBS Lett. 247*, 55−56.
54. Brent, R., and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4204−4208.
55. Goff, S. A., Goldberg, L. P., and Goldberg, A. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6647−6651.
56. Chung, C. H., Waxman, L., and Goldberg, A. L. (1983) *J. Biol. Chem. 258*, 215−221.

57. Libonati, M., Furia, A., and Beintema, J. J. (1976) *Eur. J. Biochem. 69*, 445−451.
58. Koenig, N. H., Muir, M. W., and Friedman, M. (1973) *Text. Res. J. 43*, 682−688.
59. Grant, W. M., and Kinsey, V. E. (1946) *J. Biol. Chem. 165*, 485−493.
60. Lin, S. H., Konishi, Y., Denton, M. E., and Scheraga, H. A. (1984) *Biochemistry 23*, 5504−5512.
61. Kang, I., and Wang, J. H. (1994) *J. Biol. Chem. 269*, 12024−12031.
62. Rahman, M. H., Kang, I., Waterbury, R. G., Narang, U., Bright, F. V., and Wang, J. H. (1996) *Anal. Chem. 68*, 134−138.
63. Trabesinger-Rüf, N., Jermann, T. M., Zankel, T. R., Durrant, B., Frank, G., and Benner, S. A. (1996) *FEBS Lett. 382*, 319−322.
64. Breukelman, H. J., Beintema, J. J., Confalone, E., Costanzo, C., Sasso, M. P., Carsana, A., Palmieri, M., and Furia, A. (1993) *J. Mol. Evol. 37*, 29−35.
65. Confalone, E., Beintema, J. J., Sasso, M. P., Carsana, A., Palmieri, M., Vento, M. T., and Furia, A. (1995) *J. Mol. Evol. 41*, 850−858.
66. Piccoli, R., Tamburrini, M., Piccialli, G., Di Donato, A., Parente, A., and D'Alessio, G. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1870−1874.
67. Kim, J. S., Soucek, J., Matousek, J., and Raines, R. T. (1995) *Biochem. J. 308*, 547−550.
68. Huang, Z., Schneider, K. C., and Benner, S. A. (1991) *J. Org. Chem. 56*, 3869−3882.
69. Parente, A., and D'Alessio, G. (1985) *Eur. J. Biochem. 149*, 381−387.
70. Rasmussen, L. K., Højrup, P., and Petersen, T. E. (1992) *Eur. J. Biochem. 203*, 381−386.
71. Edman, P., and Henschen, A. (1975) in *Protein Sequence Determination* (Needleman, S. B., Ed.) pp 232−279, Springer-Verlag, New York.
72. Doyen, N., and Lapresle, C. (1979) *Biochem. J. 177*, 251−254.
73. Ryle, A. P., and Sanger, F. (1955) *Biochem. J. 60*, 535−540.
74. Kumazaki, T., and Ishii, S. I. (1990) *J. Biochem. (Tokyo) 107*, 414−419.
75. Haniu, M., Acklin, C., Kenney, W. C., and Rohde, M. F. (1994) *Int. J. Pept. Protein Res. 43*, 81−86.
76. Haniu, M., Hsieh, P., Rohde, M. F., and Kenney, W. C. (1994) *Arch. Biochem. Biophys. 310*, 433−439.
77. Bauer, M., Sun, Y., Degenhardt, C., and Kozikowski, B. (1993) *J. Protein Chem. 12*, 759−764.
78. Link, T. P., and Stark, G. R. (1968) *J. Biol. Chem. 243*, 1082−1088.
79. Black, L. W. (1967) Ph.D. Dissertation, Stanford University, Stanford, CA.
80. Chang, J. Y., DeLange, R. J., Shaper, J. H., and Glazer, A. (1976) *J. Biol. Chem. 251*, 695−700.
81. Yeung, C. W. T., Carpenter, F. H., and Busse, W. D. (1977) *Biochemistry 16*, 1635−1641.
82. Schroeder, W. A., Shelton, J. B., and Shelton, J. R. (1969) *Arch. Biochem. Biophys. 130*, 551−555.
83. Hwang, C., Sinskey, A. J., and Lodish, H. F. (1992) *Science 257*, 1496−1502.
84. Janin, J., Miller, S., and Chothia, C. (1988) *J. Mol. Biol. 204*, 155−164.
85. Benner, S. A. (1989) *Adv. Enzyme Regul. 28*, 219−236.
86. Opitz, J. G., Ciglic, M. I., Haugg, M., Trautwein-Fritz, K., Raillard, S. A., Jermann, T. M., and Benner, S. A. (1998) *Biochemistry 37*, 4023−4033.
87. Freedman, R. B. (1992) in *Protein Folding* (Creighton, D., Ed.) pp 455−539, Freeman, New York.
88. Irie, M., and Tsubota, A. (1974) *J. Biochem. (Tokyo) 76*, 1319−1325.
89. Daunter, B., Hill, R., Hennessey, J., and MacKay, E. V. (1981) *Andrologia 13*, 131−141.
90. Slaweta, R., Laskowska-Klita, T., and Sosinska, G. (1985) *Med. Weter. 41*, 685−687.
91. D'Alessio, G., Floridi, A., De Prisco, R., Pignero, A., and Leone, E. (1972) *Eur. J. Biochem. 26*, 153−161.
92. Reddy, E. S. P., Sitaram, N., Bhargava, P. M., and Scheit, K. H. (1979) *J. Mol. Biol. 135*, 525−544.
93. D'Alessio, G., Di Donato, A., Furia, A., Leone, E., Libonati, M., Parente, A., and Suzuki, H. (1981) *J. Mol. Biol. 146*, 269−274.
94. Capasso, S., Mattia, C., and Mazzarella, L. (1977) *Acta Crystallogr. 33*, 2080−2083.
95. Mez, H. C. (1993) *Cryst. Struct. Commun. 3*, 657−660.
96. Blake, C. C. F., Ghosh, M., Harlos, K., Avezoux, A., and Anthony, C. (1994) *Nat. Struct. Biol. 1*, 102−105.
97. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.

BI972203E