

A preliminary three-dimensional structure of angiogenin

(angiogenesis/energy minimization/protein homology)

K. A. PALMER*, H. A. SCHERAGA*, J. F. RIORDAN†, AND B. L. VALLEE†

*Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301; and †Center for Biochemical and Biophysical Sciences and Medicine and Department of Biological Chemistry, Harvard Medical School, and Brigham and Women's Hospital, Boston, MA 02115

Contributed by H. A. Scheraga, November 26, 1985

ABSTRACT A preliminary three-dimensional structure of angiogenin has been computed, based on its homology to bovine pancreatic ribonuclease A. A standard-geometry structure of ribonuclease was first obtained from its x-ray coordinates. The fit of the backbone of angiogenin to that of ribonuclease was then optimized by taking account of amino acid deletions and by minimizing its conformational energy-plus-a-penalty distance function constraining its backbone to that of ribonuclease. Side-chain and backbone dihedral angles were allowed to vary throughout the cycles of energy minimization. In the last stages of minimization, the penalty distance function was removed. A low-energy structure resembling ribonuclease was obtained.

Angiogenin is a basic single-chain protein of 123 amino acid residues that induces *in vivo* angiogenesis, the formation of blood vessels and a vascular system (1-3). The protein has been isolated in pure form from the serum-free conditioned medium of the human colorectal adenocarcinoma cell line HT-29. Both the amino acid sequence (2) and the corresponding DNA sequence (3) have been determined, showing it to have about a 35% homology with several mammalian pancreatic ribonucleases. Since the structure of bovine pancreatic ribonuclease is known (4, 5), the structure of angiogenin can be computed by minimization of its conformational energy on the assumption that it has a three-dimensional backbone structure similar to that of ribonuclease. This approach is used here to compute a preliminary three-dimensional structure of the whole angiogenin molecule. A similar procedure has been used to compute the structure of bovine α -lactalbumin (6) from that of hen egg-white lysozyme (7) and those of several snake venom inhibitors (8) from that of bovine pancreatic trypsin inhibitor (9).

METHODOLOGY AND RESULTS

Optimized Structure of Ribonuclease. First, a model of ribonuclease (RNase) having the standard geometry (bond lengths and bond angles) of the ECEPP (empirical conformational energy program for peptides) algorithm (10, 11) was computed from the x-ray coordinates (4, 5). The objective was to obtain an energy-minimized, standard-geometry structure for ribonuclease that would superimpose well on the known 2.0-Å resolution x-ray structure and be free of any high-energy atomic overlaps. The coordinates from this standard-geometry model could then be used as input to provide a starting conformation for the backbone of angiogenin, as well as for side chains that are conserved between the two molecules. The degrees of freedom in this type of model are dihedral angles, with bond lengths and bond angles fixed at values obtained from high-resolution crystal structures of small molecules. For those dihedral angles that could

be defined by the positions of four heavy atoms, initial values were computed from the neutron- and x-ray-derived Cartesian coordinates (4, 5). For the initial ECEPP model, all peptide bond dihedral angles, ω , were fixed at 180° with the exception of those preceding proline residues, which were left variable. Hydrogen atoms were placed according to ECEPP standard geometry.

Since the bond lengths and bond angles in any protein x-ray structure are derived by fitting to an electron density map, they differ slightly from the values obtained from small-molecule crystals analyzed at high resolution. Thus, when one assigns the calculated x-ray dihedral angles to a model structure having the same sequence but with residues assigned standardized ECEPP bond lengths and bond angles, the cumulative effect of these differences is to distort the overall three-dimensional structure. To circumvent this problem, we divided the first ECEPP version of RNase into five segments of approximately 30 residues each, with the ends of the segments overlapping by 10 residues, for the initial fitting to the x-ray structure.

The ECEPP conformational energy (10, 11) plus an additive penalty function (providing strong distance constraints between the C α atoms) of the starting conformation of each segment was minimized, allowing backbone dihedral angles to vary throughout the chain to enable the computed structure to superimpose optimally on the original x-ray structure. The penalty function was of the form $A_0(r - r_0)^2$, where r and r_0 are the distance between a pair of C α atoms in the ECEPP structure and in the x-ray structure, respectively, and the force constant A_0 was set at 100 kcal/Å² (1 cal = 4.18 J). The minimization algorithm computes analytical first and second derivatives of the function for subsets of up to 250 degrees of freedom. It is coded in assembly language on an FPS-5000 array processor, which is similar to a system described earlier (12). Minimization of this function was continued to convergence for each of the segments. The resulting segments were checked visually for superposition on the x-ray structure, using "A Viewing Program" (AVP), a molecular graphics program developed in this laboratory by E. O. Purisima. The program provides for rigid-body translation and rotation of displayed structures, changes of color, superposition of structures, and variation of side-chain dihedral angles.

The segments were then joined together using the new dihedral angles to form a complete molecule of RNase. The degrees of freedom were divided into four subsets, with the first two consisting of backbone dihedral angles of alternating residues and the second two consisting of side-chain dihedral angles of residues 1-62 and 63-124, respectively. The minimization algorithm cycled through these four subsets, taking either 20 steps or converging in each subset before moving on to the next. The interaction energy of atom pairs separated by a distance greater than 10 Å was not included in the calculation. Distance constraints were included only between

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ECEPP, empirical conformational energy program for peptides.

pairs of C α atoms separated by distances less than 7 Å in the x-ray structure. Minimization was continued until the structure had a total negative ECEPP conformational energy of -297 kcal/mol (not including the contribution from the penalty distance function). This structure was free of atomic overlaps and was observed visually to superimpose closely on the original x-ray structure. The rms deviation between C α atoms in the two structures, calculated after optimized superposition, was 0.35 Å.

Fit of Angiogenin to Ribonuclease. The dihedral angles from this optimized structure of RNase were then used to provide a reasonable starting conformation for the conserved regions of the angiogenin molecule. The first step was to fit the backbone of angiogenin to that of RNase. The procedure was similar to that used to fit the ECEPP RNase structure to the x-ray structure. As shown by the sequence alignment (Fig. 1), angiogenin has six amino acid deletions at four sites relative to bovine RNase, as well as an N-terminal pyroglutamate and four C-terminal residues that extend beyond the end of the bovine RNase sequence. This preliminary structure for angiogenin does not treat the N- and C-terminal extensions. As shown by the RNase structure, both the N and C termini project out into solvent; hence, the addition of a few residues is not expected to cause any significant perturbation in the core of the molecule. We have obtained preliminary results from a full conformational search for low-energy structures for these residues by using a buildup procedure (15). Each of the four deletion sites, at residues 24, 39, 69-70, and 114-115 (bovine RNase numbering), occurs at a loop region of the molecule, as shown in Fig. 2. The hydrophobic core of the molecule, three of the four disulfide bonds, and the residues known to be involved in the catalytic activity of RNase (i.e., histidine-12, lysine-41, and histidine-119) are highly conserved in angiogenin.

The strategy for fitting the backbone and side chains of

angiogenin to those of RNase was as follows. For the initial fitting, glycines were substituted into each of the deletion sites in angiogenin. Backbone dihedral angles were assigned from the ECEPP structure of RNase to their homology-aligned counterparts in angiogenin (Fig. 1). For conserved residues, the side-chain dihedral angles were set at the values obtained from the low-energy RNase structure. In the case of conservative substitutions—e.g., isoleucine to valine—the new residue was assigned side-chain dihedral angles corresponding to the original residue as far along the side chain as the structural analogy was reasonable. For nonconserved residues, the backbone dihedral angles were kept at the same values as those of the original residue. A side-chain conformation was chosen from calculated low-energy conformations (16) for the new residue type having values of ϕ , ψ from the same area of the ϕ , ψ map (17) as the original RNase residue. This procedure reduces the number of high-energy contributions to the total energy resulting from unfavorable interactions within each side chain. The molecule was then divided into five segments corresponding in size and sequence alignment to those used in fitting the standard-geometry segments of RNase to the x-ray coordinates. Minimization of the conformational energies-plus-penalty distance function of these segments was carried out by the same procedure used for the RNase segments. A value of 100 kcal/Å² was chosen for the force constant in the penalty distance function in order to superimpose the backbones of these segments on the backbone of the RNase molecule. The segments were then joined together and minimization of the function for the whole molecule was continued, with the algorithm cycling through four subsets of the degrees of freedom. The subsets were defined in the same manner as those used in minimizing the conformational energy of RNase. The initial value of 100 kcal/Å² for the force constant A_0 was decreased progressively throughout the cycles of

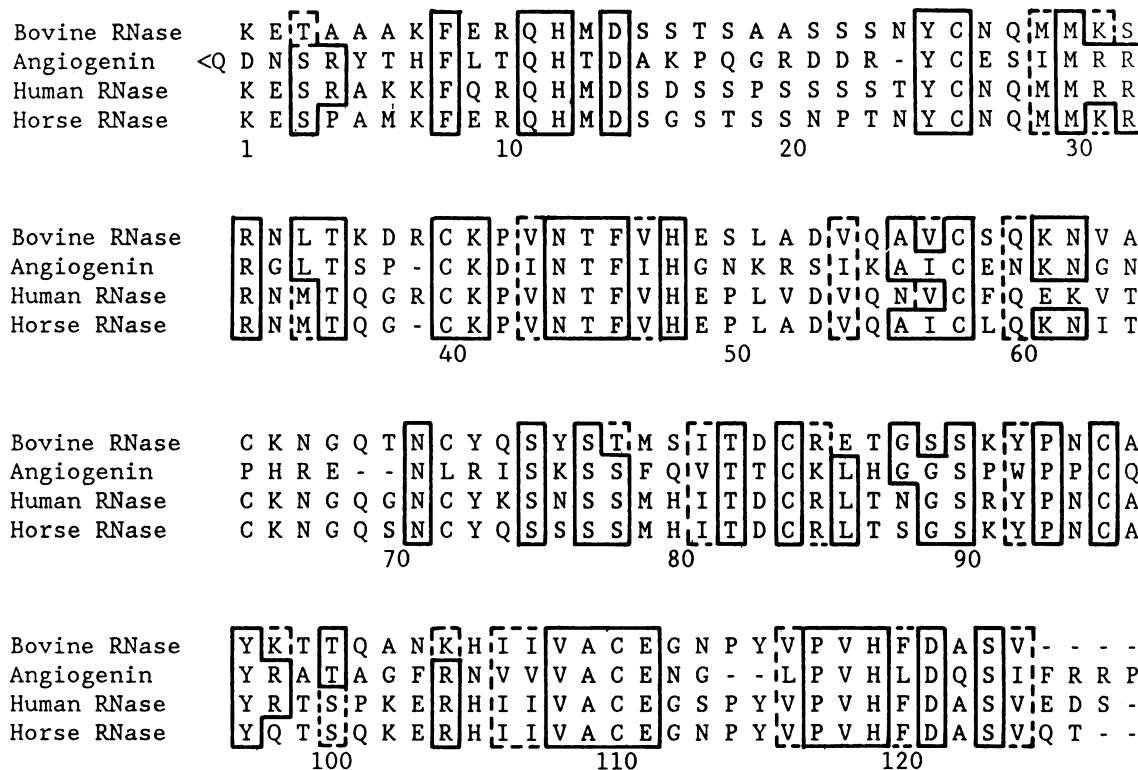


FIG. 1. Amino acid sequences of horse, human, and bovine pancreatic ribonucleases (13, 14) and of angiogenin (2, 3). The residue numbering pertains to bovine ribonuclease. Three of the four disulfide bonds in the RNases are conserved in angiogenin, but the fourth is missing. Conserved residues are enclosed in solid-line boxes. Dashed-line boxes indicate conservative replacements from the RNase sequences. The deletion shown at residue 24 could actually be placed anywhere between residue 15 and residue 24, leading to the same alignment of homologous regions.

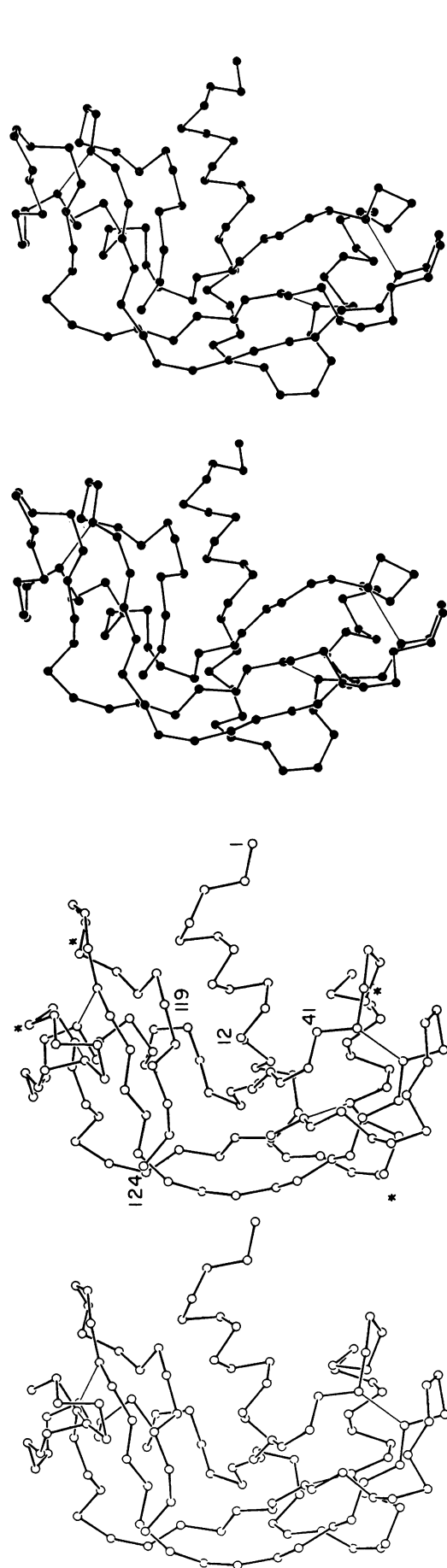


FIG. 2. Stereo ORTEP diagrams of the virtual-bond (C α -C α) structures of RNase (Left) and angiogenin (Right). The N- and C-terminal residues and the active-site residues His-12, Lys-41, and His-119 are indicated for RNase. Deletion sites are indicated by an asterisk.

minimization so as to keep the contribution of the harmonic penalty function the same order of magnitude as that of the conformational energy.

The evolving structure was checked periodically with the graphics system to see how well its backbone superimposed on that of RNase and to identify atomic overlaps. Side chains involved in obvious overlaps were replaced with alternative choices of conformations from the same area of the φ , ψ map until a conformation without overlaps was found. None of the conserved residues had to be replaced in this manner. In a few cases—e.g., tyrosine-6 of angiogenin—a suitable low-energy conformation was not found among the computed low-energy structures. In these cases, the graphics system was used to move the side chain to a position free of overlaps. Distance constraints on the C α atoms were gradually relaxed as the overlap energy from the side-chain subsets became negative. Once the initial high-energy overlaps were relieved, the major positive contributions to the energy came from the loop regions. This is just where one would expect the two structures to differ the most, based on the number of amino acid substitutions in these regions relative to the core.

The six glycines that had been substituted into the deletion areas were next removed. New backbone dihedral angles for the remaining residues in the loops then had to be found so as to satisfy the distance and orientation constraints required to join the ends of the loops to the rest of the molecule without perturbing the overall structure. Solutions for possible values of φ and ψ were generated analytically, and a solution corresponding to an energetically reasonable loop conformation was chosen to close the loop. Where prolines occurred in the loop, the value of φ was set to -75° , and the value of ψ of the preceding residue was adjusted to compensate for this change. The force constant for the penalty distance function in these loop regions was decreased to 5 kcal/Å 2 and eventually removed, since they could no longer reasonably be constrained to fit the RNase backbone in the neighborhood of an amino acid deletion. The *cis* proline at position 93 (bovine RNase numbering) was kept in the *cis* conformation because of its conservation. However, the addition of two other prolines to this loop region (residues 89–94, bovine RNase numbering), as well as several other nonconservative substitutions nearby in the chain, clearly indicates that a number of different arrangements of these residues will have to be energy minimized and their energies compared before a reasonable estimate as to the true conformation of this region can be made. There were several X-to-proline substitutions in loops of angiogenin, and for this initial structure these prolines were assumed to be in the *trans* conformation. Further work, minimizing the energies of alternative starting conformations of angiogenin, will be necessary to explore the possibility that these new prolines may adopt the *cis* conformation.

Minimization was then continued on the 118-residue model of angiogenin with deletions incorporated into the loops. Relief of some overlaps created others during the minimization. These were relieved using the AVP program, by comparing the side chain with its counterpart in RNase and placing it in the same approximate orientation wherever possible. Minimization was continued until the conformational energy of the side-chain subsets was negative and the total overlap energy was of the same order of magnitude as that from the penalty distance function, with distance constraints removed from the loop regions and with the constant A_0 set to 5 kcal/Å 2 . All of the atomic pairwise interactions were checked for overlaps of more than 5 kcal. Those remaining were almost all localized to the surface loop regions. Some overlap energy at this point was doubtless due to satisfaction of the remaining distance constraints. Since the molecule now had a net negative conformational energy, all distance constraints were next removed. This allowed the

structure to relax into the nearest conformational energy well, without an artificial penalty function holding it close to the backbone of RNase. The structure resulting from minimization of this starting conformation without distance constraints had a negative energy, and the backbone of the molecule superimposed well on that of RNase in the core region. This can be seen in Fig. 2, which shows only the connected C α atoms of the two structures for clarity. The conserved regions, including the region corresponding to the active site of RNase, can be compared in Fig. 3, which shows the models of the heavy atoms of both proteins.

DISCUSSION

The structure presented here is the result of minimization of the conformational energy of one starting conformation for the angiogenin molecule and as such should not be interpreted as representing the global energy minimum. No attempt has yet been made to explore alternative conformations for side chains that were free of overlaps throughout this minimization. We expect such a search to reveal, in some cases, alternative side-chain and backbone dihedral angles corresponding to lower-energy conformations, especially in the nonconserved loop regions of the molecule. One could start with the standard-geometry RNase structure and change the nonconserved residues one at a time. Then, a full search through the conformational space available to each residue could be carried out, with energy minimization of the structures obtained after each substitution. The conformation with the lowest computed energy after each amino acid substitution would then serve as the starting conformation for the next cycle of residue substitution and conformational energy minimization. Because of the existence of many local minima in the conformational energy surface, the molecule will usually settle into a minimum near its starting conformation at each stage. Therefore, the energies of a number of starting conformations must be minimized and their final energies be compared before a final structure can be predicted, under the hypothesis that the native structure corresponds to the global minimum of the conformational energy. In the case of homologous proteins, the assumption that sequence homology implies structural similarity makes an adequate sampling of starting conformations possible, since conserved regions can be fixed and not allowed to vary until the later stages of the energy-minimization procedure.

The purpose of this communication is to present a plausible model for angiogenin, particularly of the region corresponding to the active site of RNase, in the hope that the three-dimensional representation of residues in this region will aid in the search for the mechanism of the angiogenic activity of this molecule and in the design of inhibitors of its biological activity. In addition to the catalytic residues, comparison of the sequences and of the structures shows that a number of hydrogen bonds and hydrophobic contacts, as well as three of the four disulfide bonds of RNase, are preserved in angiogenin. These interactions contribute to the stability of both structures and may be important in the binding of postulated substrates for angiogenin.

One of the amino acid changes in the active site region is the substitution of a leucine for phenylalanine-120 in the corresponding site in angiogenin. The effect of this particular substitution on substrate binding, catalytic activity, and the structural stability of the RNase molecule has been characterized previously (18). The replacement of phenylalanine-120 by a leucine did not prevent binding of the substrate and, in fact, had very little effect on the K_m of binding. The catalytic activity of the molecule, however, was diminished by this modification. Phenylalanine-120 was shown to be involved in interactions with the C-terminal region of RNase that are important for its structural integrity. Our preliminary model of angiogenin does not yet allow for comparison of the

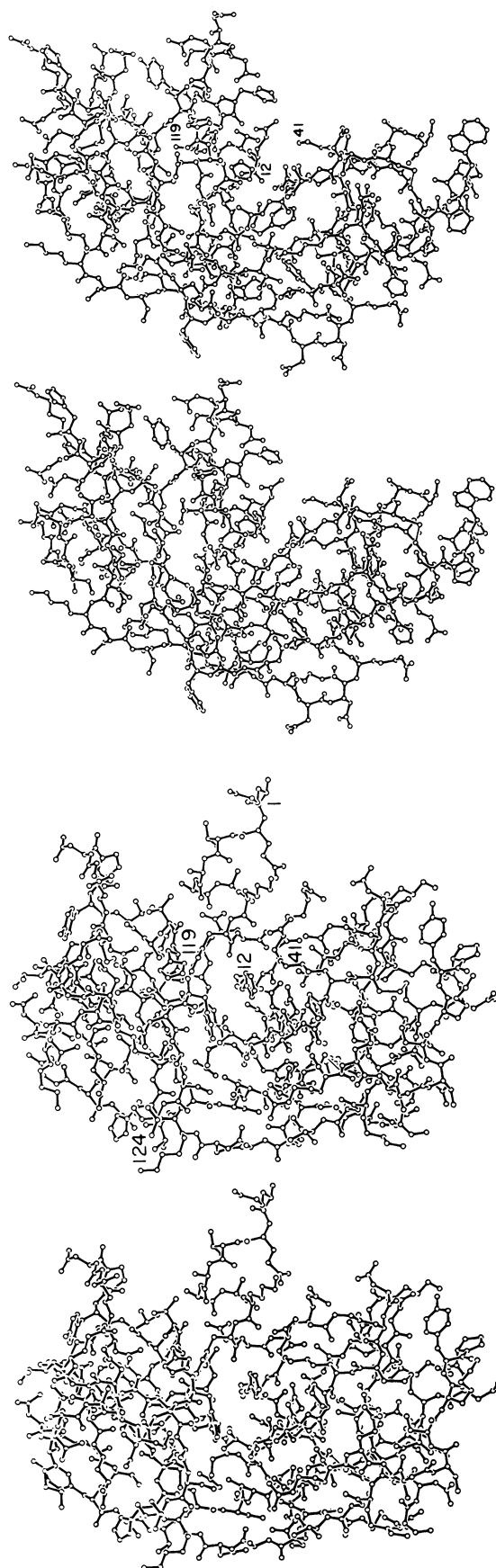


FIG. 3. Stereo ORTEP diagrams of the full (heavy)-atom structures of RNase (*Left*) and angiogenin (*Right*). The active-site residues His-12, Lys-41, and His-119 (bovine RNase numbering) are labeled in each structure.

region of the C-terminal extension, because this extension does not exist in RNase. However, the experimental study (18) does suggest that the leucine-for-phenylalanine substitution in the active site would not be sufficient to prevent binding of an RNase substrate to angiogenin. Nuclear magnetic resonance (19) and x-ray studies (see, e.g., refs. 4 and 5) of inhibitor binding to RNase have identified additional residues that are affected by substrate binding. Among these are threonine-45 and serine-123 in the active site and histidine-48, which is removed from the catalytic residues in RNase but whose spectral changes on binding of substrate analogs may reflect a conformational change in the enzyme (19). On the basis of structural studies, these residues as well as glutamine-11 and aspartate-121 of RNase have been proposed as being important in the binding of substrate (13, 20). All of these residues are conserved in the angiogenin molecule. In comparing our preliminary structure for angiogenin with the x-ray structure of RNase, we see nothing obvious at this point to prevent binding of RNA substrates to angiogenin, with a possible associated catalytic activity. No such activity has been found for this molecule with the substrates tested to date (2, 3). The cloning of the gene for human angiogenin (3) will make it possible to produce sufficient material for further biochemical characterization of the activity of this protein, and, in conjunction with further structural studies, should lead to eventual elucidation of the actual mechanism of its action.

The homology between angiogenin and RNase is less than that seen among the pancreatic ribonucleases, making its structural properties and the specific interactions responsible for these properties a very interesting area for investigation. A number of experimental and theoretical studies have been carried out attempting to identify nucleation sites and possible folding pathways for RNase (see, e.g., refs. 21–23). In the case of homologous proteins, residues critical to proper folding can be deduced from conserved sequences and specific intrachain interactions if the structures are known. The correlation of conserved regions with those shown experimentally to be important in the folding of a polypeptide chain can be taken as supporting evidence for a nucleation site. It is interesting to note that the C-terminal region of angiogenin has a hydrophobic stretch of residues in the same area as that proposed, partly on the basis of hydrophobicity, to be the primary nucleation site for the folding of RNase (21–23).

Using the technique outlined above, we had no problem fitting the backbone of angiogenin to that of RNase in the areas of regular structure—i.e., the three α -helices and the three-stranded β -sheet region. Further exploration of the conformational space for this molecule will be necessary to determine how far the structural analogy between the two molecules holds. However, at this stage, the attainment of a low-energy structure, unconstrained by any penalty distance function, that is close to the structure of RNase in these regions argues for conservation of the overall structural motif of RNase in the angiogenin molecule.

We should like to thank M. J. Dudek, E. O. Purisima, M. Vázquez, and M. Lambert for use of their computer programs and for helpful discussions and S. Rumsey for help with the graphics. This work was supported by research grants from the National Institute of General Medical Sciences (GM-14312) and from the National Science Foundation (DMB84-01811). Support was also received from the National Foundation for Cancer Research. The work on angiogenin leading to this study was supported by funds from the Monsanto Company under agreements with Harvard University.

1. Fett, J. W., Strydom, D. J., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F. & Vallee, B. L. (1985) *Biochemistry* **24**, 5480–5486.
2. Strydom, D. J., Fett, J. W., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F. & Vallee, B. L. (1985) *Biochemistry* **24**, 5486–5494.
3. Kurachi, K., Davie, E. W., Strydom, D. J., Riordan, J. F. & Vallee, B. L. (1985) *Biochemistry* **24**, 5494–5499.
4. Wlodawer, A. (1985) *Protein Data Bank* (Brookhaven National Laboratory, Upton, NY).
5. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–541.
6. Warne, P. K., Momany, F. A., Rumball, S. V., Tuttle, R. W. & Scheraga, H. A. (1974) *Biochemistry* **13**, 768–782.
7. Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967) *Proc. R. Soc. London Ser. B* **167**, 365–377.
8. Swenson, M. K., Burgess, A. W. & Scheraga, H. A. (1978) in *Frontiers in Physicochemical Biology*, ed. Pullman, B. (Academic, New York), pp. 115–142.
9. Deisenhofer, J. & Steigemann, W. (1975) *Acta Crystallogr. Sect. B* **31**, 238–250.
10. Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975) *J. Phys. Chem.* **79**, 2361–2381.
11. Némethy, G., Pottle, M. S. & Scheraga, H. A. (1983) *J. Phys. Chem.* **87**, 1883–1887.
12. Pottle, C., Pottle, M. S., Tuttle, R. W., Kinch, R. J. & Scheraga, H. A. (1980) *J. Comput. Chem.* **1**, 46–58.
13. Blackburn, P. & Moore, S. (1982) *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 15, pp. 317–433.
14. Beintema, J. J., Wietzes, P., Weickmann, J. L. & Glitz, D. G. (1984) *Anal. Biochem.* **136**, 48–64.
15. Vázquez, M. & Scheraga, H. A. (1985) *Biopolymers* **24**, 1437–1447.
16. Vázquez, M., Némethy, G. & Scheraga, H. A. (1983) *Macromolecules* **16**, 1043–1049.
17. Zimmerman, S. S., Pottle, M. S., Némethy, G. & Scheraga, H. A. (1977) *Macromolecules* **10**, 1–9.
18. Lin, M. C., Gutte, B., Caldi, D. G., Moore, S. & Merrifield, R. B. (1972) *J. Biol. Chem.* **247**, 4768–4774.
19. Meadows, D. H., Roberts, G. C. K. & Jardetzky, O. (1969) *J. Mol. Biol.* **45**, 491–511.
20. Stern, M. S. & Doscher, M. S. (1984) *FEBS Lett.* **171**, 253–256.
21. Matheson, R. R., Jr., & Scheraga, H. A. (1978) *Macromolecules* **11**, 819–829.
22. Chavez, L. G., Jr., & Scheraga, H. A. (1977) *Biochemistry* **16**, 1849–1856.
23. Némethy, G. & Scheraga, H. A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6050–6054.