Computer Modeling of Human Angiogenin-Dinucleotide Substrate Interaction

M.S. Madhusudhan and S. Vishveshwara*
Molecular Biophysics unit, Indian Institute of Science, Bangalore, India

ABSTRACT **Structures of substrate bound human angiogenin complexes have been obtained for the first time by computer modeling. The dinucleotides CpA and UpA have been docked onto human angiogenin using a systematic grid search procedure in torsion and Eulerian angle space. The docking was guided throughout by the similarity of angiogenin-substrate interactions with interactions of RNase A and its substrate. The models were subjected to 1 nanosecond of molecular dynamics to access their stability. Structures extracted from MD simulations were refined by simulated annealing. Stable hydrogen bonds that bridged protein and ligand residues during the MD simulations were taken as restraints for simulated annealing. Our analysis on the MD structures and annealed models explains the substrate specificity of human angiogenin and is in agreement with experimental results. This study also predicts the B2 binding site residues of angiogenin, for which no experimental information is available so far. In the case of one of the substrates, CpA, we have also identified the presence of a water molecule that invariantly bridges the B2 base with the protein. We have compared our results to the RNase A-substrate complex and highlight the similarities and differences. Proteins 2001; 42:125–135.** © 2000 Wiley-Liss, Inc.

Key words: human angiogenin-CpA/UpA complex; docking; molecular dynamics; simulated annealing; specificity

INTRODUCTION

Human angiogenin is 33% identical in sequence (see Fig. 1) to bovine pancreatic ribonuclease (RNase A). A sequence comparison of Angiogenin with RNase A shows that key residues are conserved. Angiogenin also has a three-dimensional structure similar to that of RNase A but catalyzes the same reaction, cleavage of single stranded RNA, 10^5 times less efficiently[1] (in terms of specificity constant). Though angiogenin's primary biological role is that of stimulating the growth of blood vessels,[2] the RNase activity nevertheless seems to be an important function of the protein.[3] Structural information on angiogenin has so far been confined to only the native form and its mutants.[4–8]

In order to understand the mechanism of action of angiogenin it is imperative that structures of substrate bound angiogenin complexes become known. We have taken some initial steps in this direction by modeling mono[9,10] and dinucleotide substrates onto angiogenin. In this study we report the docking of dinucleotides CpA and UpA, both of which are substrates, onto human angiogenin.

The docking of substrates onto angiogenin was not straightforward because of two main reasons. The first one being the obstruction of the B1 binding site by the side chain of residue Q117.[11] A similar problem has been dealt with in detail in an earlier study on bovine angiogenin.[9] The other hurdle faced in docking dinucleotide substrates is the dearth of biochemical data on the second base binding region of the protein. The present investigations predict the plausible residues that could interact with the second base. In the docking effort the biochemical evidence[12–14] on the active/binding regions and structures of substrate bound RNase A were used as reference. The substrate bound RNase A complexes included both crystal structures[15] and those modeled by our group in earlier studies.[16–19] In addition to the direct protein substrate interactions we have also predicted the presence of a water-mediated interaction in one case. Our model is the first structure of angiogenin with its substrate.

A nanosecond molecular dynamics simulation is carried out on the modeled complexes, which brings out the stability of the complexes and highlights the persistent interactions. Finally, we have refined our structures by simulated annealing. This method is now frequently used for structure refinement.[20,21] We comment on the substrate specificity of angiogenin and also on the conformational aspects of the docked substrate.

METHODS

Docking Procedure

The torsion angles of the substrate (Fig. 2a,b) that were varied during the docking procedure are shown in Figure 2a. In order to dock the substrate CpA, the coordinates of the pyramidine base of bovine angiogenin -3′CMP complex[9] were transferred onto the crystal structure of human angiogenin (PDB code 2ANG[8]). This transfer was effected after the active and binding site residues (H13, K40, and H114 of angiogenin) of the two proteins were superim-

Grant sponsor: Department of Science and Technology; Grant number: SP/SO/D-107/98.

*Correspondence to: S. Vishveshwara, Molecular Biophysics unit, Indian Institute of Science, Bangalore 560 012, India. E-mail: sv@mbu.iisc.ernet.in

Received 11 February 2000; accepted 1 September 2000

```
human       -QDNSRYTHFLTQHYDAKPQGRDD-RYCESIMRRRGLTSP-CKDINTFIHGNKRSIKAIC
bovine      AQDDYRYIHFLTQHYDAKPKGRND-EYCFNMMKNRRLTRP-CKDRNTFIHGNKNDIKAIC
RNase A     --KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAVC

human       ENKNGNPHRE--NLRISKSSFQVTTCKLHGGSPWPPCQYRATAGFRNVVVACENG--LPV
bovine      EDRNGQPYRG--DLRISKSEFQITICKHKGGSSRPPCRYGATEDSRVIVVGCENG--LPV
RNase A     SQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPV

human       HLDQSIFRRP--
bovine      HFDESFITPRH-
RNase A     HFDASV------
```

Fig. 1. An alignment of the Human and Bovine Angiogenins with RNase A. The conserved active/binding site residues, Q12, H13, K40, T44, H114, and F/L115 (numbering according to Human angiogenin sequence), have been shaded in grey.

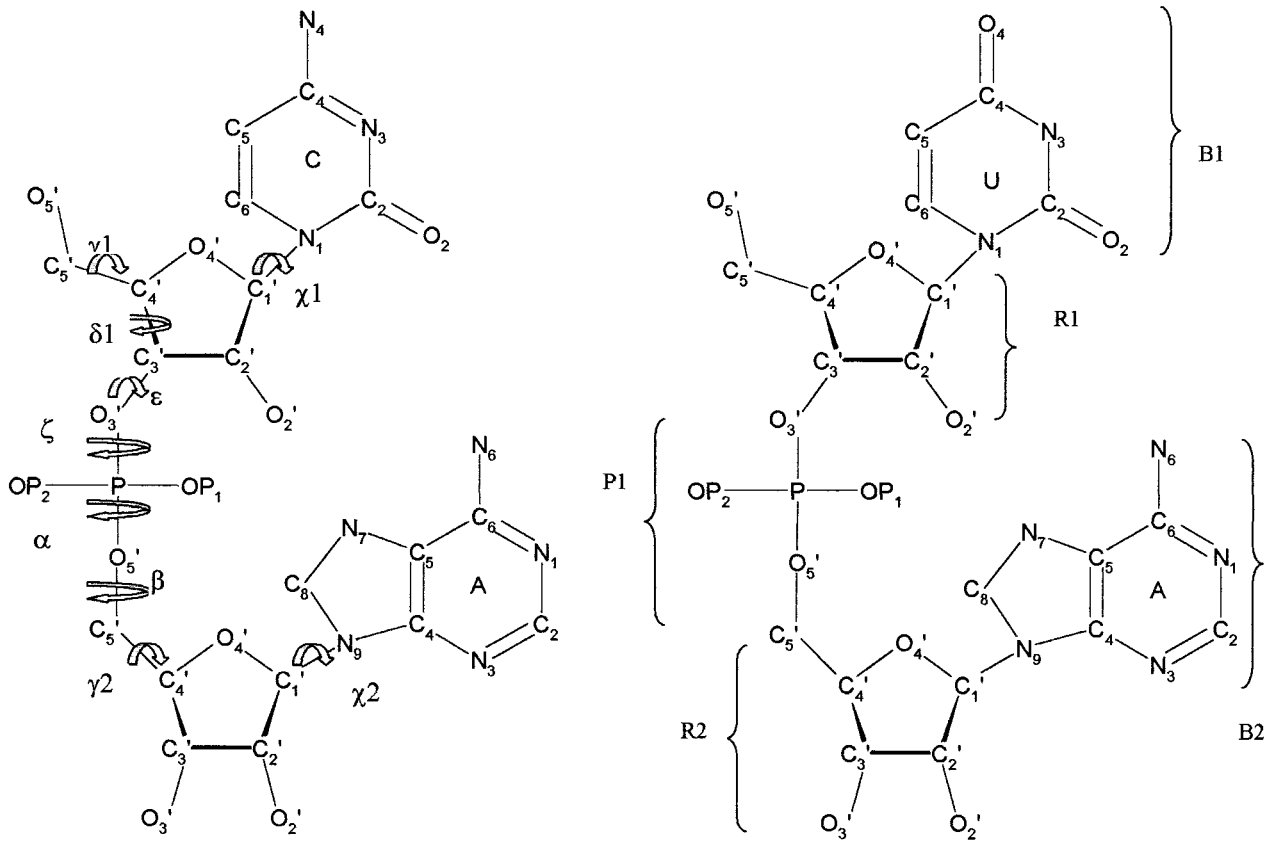

Fig. 2. Schematic representations of the docked substrates (**a**) CpA and (**b**) UpA. The substrate torsion angle α, β, χ, γ, δ, ϵ, and η are marked in a.

posed.[22] Rigid body translation and rotation operations[23] were performed on the base to systematically obtain all steric-free conformations of the base. Of these allowed conformations of the base only those that have interactions with protein residues identified biochemically as important for first-base binding were chosen. Interactions here refer to hydrogen bonds and Van der Waal's interactions. The ribose R1 moiety was added to the chosen conformations of the base. A grid search was done on the glycosidic torsion angle χ1. A coarse grid search takes torsion angles at 30° intervals. A finer search, if necessary, is done on a smaller range of torsion angles for every 6°. All steric-free conformations with most if not all desired interactions are chosen. The phosphate moiety was built onto these templates. Another grid search is done in torsion angle space and the allowed conformations are selected based on the number of favorable interactions. This procedure is repeated again with the uracil base for docking UpA.

TABLE I. The Number of Atoms and Physical Dimensions of the Systems on Which MD Simulations Were Carried Out

System	Total number of atoms (number of waters)	Box size (A^3)
human angiogenin	10912 (2983)	$56.67 \times 48.01 \times 46.12$
human angiogenin + CpA	10140 (2705)	$56.37 \times 47.86 \times 43.87$
human angiogenin + UpA	10064 (2680)	$56.34 \times 48.01 \times 46.12$

The second nucleotide with an adenine base is built upon this template in a procedure similar to that mentioned above. The ribose R2 and base B2 were added serially. With the addition of every moiety, the additional torsion angles are subjected to a grid search to obtain favorable conformations. The pucker of the ribose sugar is also varied during the docking procedure. Such a systematic grid search in torsion angle and Eulerian angle space ensures scanning of all possible productive modes of binding.

In the case of two angiogenin complexes, not only were the substrate torsion angles varied but also some select main-chain and side-chain torsion angles of the protein were systematically varied. These variations were done in conjunction to the rigid body rotation and translation effected on the base of the first nucleotide.

Molecular Dynamics (MD) Simulation Protocol

All simulations reported in this study have been carried out using the AMBER4.1 suite of programs.[24] All force field parameters are listed in AMBER 4.1.[25] Details of the sizes of the simulated systems reported in this study are presented in Table I. The starting structures for these simulations were the modeled structures of the protein-substrate complexes. The starting structure for the native human angiogenin was the 2.0 Å crystal structure[8] taken from PDB (PDB code 2ANG). The starting models were all solvated using a box of TIP3P[26] water molecules using the EDIT module of AMBER. The dimensions of the box (see Table I) were such that there was at least one water molecule 4Å away from every protein atom. When the starting structures were taken from X-ray crystallography, experimentally determined waters were excluded.

All the starting structures were modeled using the AMBER 94 all-atom force field.[25] The amino acids lysine and arginine carried a net charge of +1 while the amino acids glutamic acid and aspartic acid carried a net negative unit charge. All histidines side chains except H13 were doubly protonated. The side chain of H13 was protonated only in the Nϵ position in conforming to the role of general base attributed to this residue during catalysis.[27]

The starting structures were energy minimized. The first 200 steps of minimization followed the steepest descent approach while the next 1800 steps made use of the conjugate gradient algorithm. Equilibration steps followed minimization. During equilibration the system was simulated using the canonical ensemble (NVT) for 4ps at 100K followed by another 4ps of NVT MD at 200K and finally for 12ps at 300K. The temperature during equilibration was held fixed using the Berendsen's coupling process.[28] For the production run the temperature coupling was removed and the systems were simulated in the micro-canonical ensemble (NVE). However, rescaling was resorted to if the temperature deviated from 300K by 10° or more. SHAKE[29] has been used to constrain all covalent bonds at their equilibrium values. The usage of SHAKE also facilitated a time step of 2 femtoseconds. Particle Mesh Ewald sum[30] (PME) was used to calculate electrostatic interactions utilizing a cubic B-spline interpolation order. The size of the grid sides was chosen to be products of 2, 3, and 5, to facilitate fast Fourier transforms. The grid spacing was ~1Å. Periodic boundary conditions were applied. Nonbonded interactions were evaluated using a residue-based list with a cutoff of 12Å, updated every 25 steps. A constant value of 1 was used for the dielectric constant. During the simulation, trajectory snapshots were stored every 1ps. The production runs were all of 1 nanosecond duration.

Analysis

The analysis of all the MD trajectories was carried out using programs coded in FORTRAN. For all the simulations the root-mean-square deviation (RMSD) from its starting and time-averaged structures, residue-wise RMS fluctuation and backbone, and side-chain torsion angles were routinely calculated. Along with these analyses, the acceptor-donor distance of potential hydrogen bonding pairs was also monitored. The criteria of a donor-acceptor distance of less than 3.6Å and the proton-acceptor distance of less than 2.6Å, were used for hydrogen bond detection.

Simulated Annealing

Simulated annealing is a standard technique that takes a system to a global minimum circumventing local minima. The annealing method increases the energy of the system so that it can go to a state that transcends most if not all energetic barriers. The system's energy is then reduced gradually with the introduction of appropriate constraints to navigate its course to the global minimum.[20,21] The temperature of the system, which is a good handle on its energy, is first increased and the system is then slowly cooled.

In the docking studies in which both the substrate and protein were modeled to obtain their complex, it is imperative to ensure that the complexed structures are energetically at their minimum. To this end, simulated annealing was performed on the dinucleotide complexed modeled structures taken from snap shots of MD simulations. Inter-atomic distances between atoms of the substrate and

TABLE II. A Summary of the Protocol Used for Simulated Annealing

Time period of MD simulation	Temperature range (Kelvin)	Time constant for temperature coupling (ps)	Restraints on interatomic distances (Kcal/mol)
0 to 3ps			
0ps–1ps	0°–1200°	0.2	0.1 × 25
1ps–3ps	1200°–1200°	0.2	1.0 × 25
3ps–15ps	1200°–0°		
3ps–11ps		4.0–2.0	1.0 × 25
11ps–13ps		1.0–0.5	1.0 × 25
13ps–15ps		0.05	1.0 × 25

the protein that interacted through a hydrogen bond were used as distance restraints in the simulated annealing run. The restrained hydrogen bonds are shown in Table III.

The system that was annealed was only a part of the whole docked complex and is referred to as a "belly" system. All residues that came within a spherical cutoff of 12Å from the substrate, including water, constituted the belly system for which simulated annealing was performed. The belly systems were subjected to MD at high temperatures initially and the temperature was then gradually reduced. With reduction in temperature the inter-atomic distance restraints were made stronger. For each of the protein-substrate complexes 10 final models were obtained by annealing 10 different complex structures.

Table II summarizes the protocol for annealing. The change in temperature is controlled by the time constant for the heat bath coupled to the solute.[28] The smaller the value of this coupling constant, the tighter is the coupling to the heat bath and hence faster the rate of change of temperature. The value of the coupling constant was 0.2ps for the first 3 picoseconds, where the system was heated to a temperature of 1200K. The system was cooled in the next 12 picoseconds. The time constant for coupling during the first 8 picoseconds of the cooling phase started from 4.0ps and decreased gradually to 2.0ps. This is the slow cooling stage. Similarly the time constant for coupling was reduced from 1.0ps to 0.5ps in a further 2ps. For the last 2 picoseconds of dynamics, the value of the coupling constant is taken as 0.05ps. This very low coupling constant will cool the system rapidly and is akin to an energy minimization.

TABLE III. Important Acceptor-Donor Pairs Identified From the MD Simulations of Angiogenin-Substrate (CpA and UpA) Complexes and Used as Restraints in Simulated Annealing

	Donor-acceptor pair	Presence in the annealed models of CpA complex	Presence in the annealed models of UpA complex
1	NH2 HIS8-O2P Py[a]	✓	✓
2	Nϵ2 HIS 13-O1P Py	✓	✓
3	Nϵ2 HIS 13-O2′ Py	✓	✓
4	Nζ LYS40-O2 Py	✓	✓
5	Nζ LYS40-O2′ Py	✓	✓
6	N THR44-N3 Py	✓	x
7	N THR44-O2 Py	x	✓
8	Oγ1 THR44-N3 Py	x	✓
9	Oγ1 THR44-N4/O4 Py	✓	x
10	Nδ2 ASN68-N1 Pu[b]	✓	✓
11	Nδ1 HIS114-O5′ Py	✓	x
12	Nδ1 HIS114-O1P Py	✓	✓
13	N LEU 115-O1P Py	✓	✓
14	O HIS 8-N GLN 12	✓	✓
15	Oγ1THR44-Oγ1THR 80	✓	✓
16	NH1ARG66-Oϵ1GLU67	✓	x
17	Oϵ2ARG66-Nδ2 ASN68	✓	x
18	WATER-Oϵ2 GLU67	✓	x
19	WATER-N6 Pu	✓	x

[a]Py Pyrimidine (U or C)
[b]Pu Purine (A)

RESULTS AND DISCUSSION

Docking of Dinucleotides CpA and UpA

There are two key questions that need resolution in the docking of substrates (dinucleotides) onto human angiogenin. First, how is the obstruction to the B1 subsite from the side chain of Q117 removed to facilitate the first base to bind? Second, what are the residues that interact with the second base? Biochemical evidences so far only point to the residues that bind to the first nucleotide base.[12,1]

The first step in docking substrates was to superimpose the crystal structure of human angiogenin on the CMP bound complex of bovine angiogenin, which was modeled earlier.[9] The superposition was done using the Nδ and Nϵ atoms of the two catalytic histidines (H13 and H114) and the Nζ atom of K40. The coordinates of the substrate were then transferred onto the human protein. As in the case of the bovine system, the B1 binding site was obstructed by the side chain of Q117. An extensive grid search in side-chain torsion angle space did not reveal an alternate steric-free conformation of Q117 which would open up the B1 binding site to facilitate the binding of the first base of the substrate. The docking strategy was then to alter the main-chain conformation of the C terminus, of which Q117 is a part. The torsion angle ϕ of residue L115 was moved by 18° to effect a change in the Q117 position. This change led to the movement of the C^{α} atom of Q117 by 2.12Å from its original location in the crystal structure. This new conformation of the C terminus facilitated the docking of the first nucleotide of the substrate without steric hindrance. Both

TABLE IVA. Substrate Torsion Angles and Substrate-Protein Interaction Enthalpies in CpA Bound Human Angiogenin Complex

Model	Phase 1	Phase 2	χ1	χ2	α	β	γ1	γ2	δ1	δ2	ϵ	ζ	Interaction energy[a]
RNase A-UpA	18.00		−159	−63	−91	−172	178	173	77	94	−166	81	
Starting model	18.00	−54.00	177.13	−128.10	−61.76	−104.81	−152.00	−156.07	101.32	96.01	−178.91	−170.88	—
1.00	57.89	−47.51	−151.28	−62.51	145.08	−173.06	50.40	63.89	76.07	124.36	−153.11	114.23	−470.05
2.00	56.93	−8.70	−160.31	−67.88	99.25	88.49	53.13	−179.82	81.71	153.70	−108.80	103.08	−469.60
3.00	56.97	−32.08	−146.84	−96.90	58.24	177.76	−44.45	172.61	86.37	139.80	−169.35	131.60	−445.58
4.00	58.93	41.98	−159.66	−99.18	71.39	147.12	−46.07	176.25	89.33	75.05	−152.07	121.15	−468.42
5.00	57.85	64.43	−146.76	−108.08	61.24	177.92	−43.49	169.97	85.06	83.02	−167.74	128.04	−437.24
6.00	54.43	−43.76	−149.79	−60.51	128.99	−166.03	44.06	60.78	74.14	125.10	−154.35	119.28	−466.94
7.00	57.24	−41.02	−150.71	−95.57	82.01	157.78	9.04	−178.30	69.08	131.39	−169.06	135.04	−446.35
8.00	52.34	−20.79	−146.95	−93.74	64.55	168.28	−45.26	178.34	81.12	141.26	−178.54	144.13	−473.28
9.00	54.21	−39.91	−144.62	−88.68	54.90	158.46	32.46	−173.10	68.91	136.32	−163.99	119.82	−454.38
10.00	57.22	47.52	−146.55	−122.97	67.32	169.61	48.70	166.80	82.17	77.62	−178.12	147.26	−473.11

[a]Energy is expressed in terms of kcal/mol

TABLE IVB. Substrate Torsion Angles and Substrate-Protein Interaction Enthalpies in UpA Bound Human Angiogenin Complex

Model	Phase 1	Phase 2	χ1	χ2	α	β	γ1	γ2	δ1	δ2	ϵ	ζ	Interaction energy[a]
1.00	60.23	−21.61	−164.38	−89.31	102.05	92.71	179.75	−179.77	81.94	152.94	−119.55	91.10	−398.38
2.00	44.26	56.23	−169.85	−142.80	66.73	155.03	60.14	179.21	84.77	76.96	−159.84	123.76	−390.23
3.00	73.27	−38.47	170.97	−75.86	103.61	104.37	73.85	−158.37	94.40	134.24	−117.99	98.66	−419.83
4.00	35.54	80.90	−155.30	−116.08	44.78	179.53	55.78	−173.36	87.61	88.36	−164.44	121.42	−436.63
5.00	18.34	−42.82	−148.11	−113.47	48.04	−173.39	53.80	−172.32	82.88	126.74	167.56	164.26	−415.26
6.00	65.45	−24.49	177.80	−90.25	110.22	87.61	−179.58	−161.17	82.96	144.29	−124.11	96.67	−378.19
7.00	47.00	59.68	−152.76	−118.92	68.67	145.89	−158.65	179.56	82.43	74.11	−131.77	109.58	−364.91
8.00	63.28	−31.87	−171.51	−87.33	107.01	77.75	−170.43	−152.80	90.04	139.98	−131.38	99.82	−390.30
9.00	55.96	56.33	174.09	−125.17	111.85	74.41	−164.63	173.93	98.89	88.17	−94.17	103.37	−403.55
10.00	55.42	−80.28	175.00	87.08	77.13	173.49	−168.86	40.78	81.82	90.48	−130.63	104.45	−402.57

[a]Energy is expressed in terms of kcal/mol

pyrimidines Uracil and Cytosine were docked in the first base position. The second base (B2) Adenine was then built using the substrate conformation of the UpA substrate complexed to RNase A.[16]

We identified five polar residues that were in the vicinity of B2 as most probable candidates to interact with it. These were R66, E67, N68, E108, and R5. None of these protein residues, however, seemed to be within proper hydrogen bonding distance of base B2 or ribose R2. An extensive grid search in torsion angle space of the side chains of these residues also did not yield any results. This exercise was carried out on different loop conformations (backbone of residues R66, E67, and N68) by taking structures from different time points of the native human angiogenin MD simulation. This too did not give any good interaction between the residues of the protein and the second base. The interactions were finally obtained by another grid search after changing the pucker of the ribose R2 to C1′-endo (phase angle -54°) from C3′-endo (phase angle 18°). The residue N68 hydrogen bonds to the N1 atom of B2. The residues E67 and R66 in the same loop are still beyond hydrogen bonding distance. R5 interacts with the O3′ atom of the ribose R2. E108, a binding site residue in RNase A, does not interact with the substrate. The torsion angles of the substrate in the unrefined model, referred to henceforth as the starting model, are listed in the first two rows of Table IVa. Both the CpA and UpA substrates were given the same conformation in the starting model, the only difference being that the Oγ1 of T44 was hydrogen bound to the N3 atom of the first base of UpA as opposed to the N4 atom of CpA.

The starting model of the substrate-angiogenin complex had hydrogen bonds that connected the phosphate oxygens of the substrate to the catalytic histidines (H13 and H114), H8 and Q12. The atoms O2 of B1 and O2′ of R1 interacted with the side chain Nζ of Lys40. There is biochemical evidence in the form of site-directed mutagenesis to show the involvement of these residues in the binding process.[4,12] All these interactions are also conserved in the case of substrate bound complexes of RNase A.

In human angiogenin, like in the case of eosinophil-derived neurotoxin,[31] the position 115 is occupied by Leucine, the equivalent residue in bovine angiogenin and RNase A being Phenylalanine, F116 and F120, respectively. The back bone N of F120 in RNase A tethers the phosphate position and allows for the bending of the

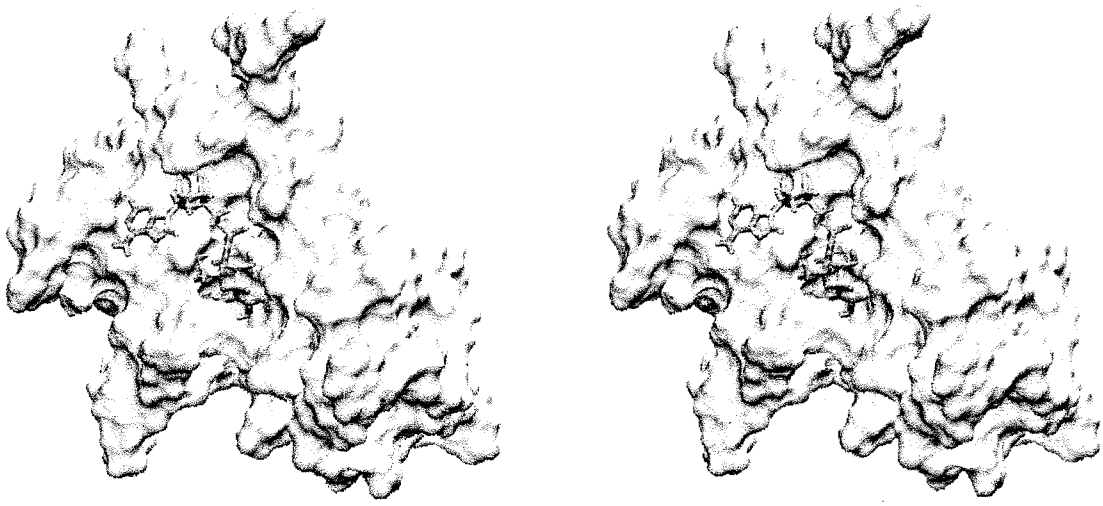

Fig. 3. Stereo diagram of a space filled model of human angiogenin with the docked substrate CpA shown in stick representation.[43]

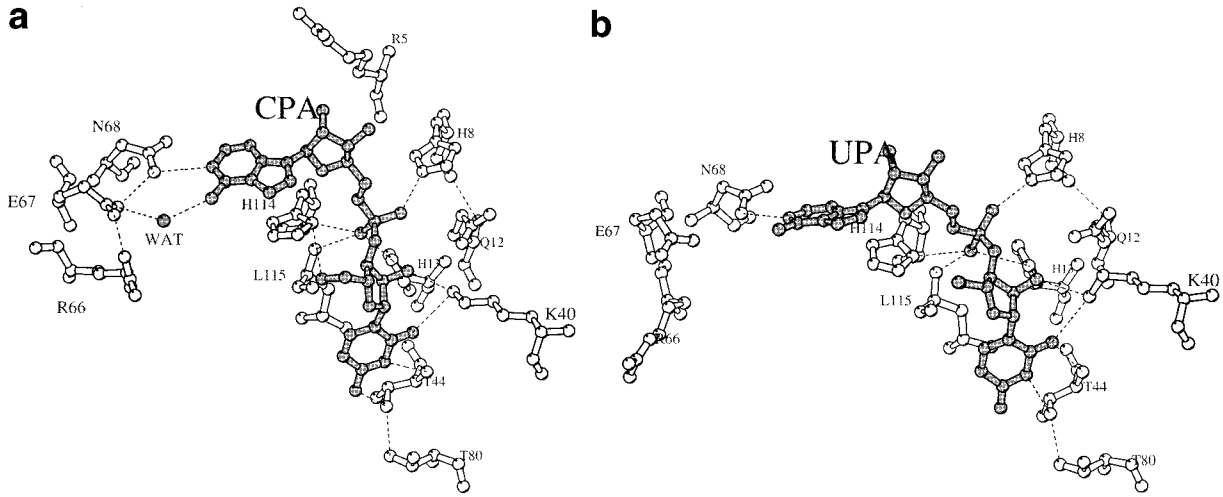

Fig. 4. The active and the binding site residues of human angiogenin. The substrate [(**a**) = CpA and (**b**) = UpA] is in stick representation while the protein residues are in ball and stick representation.

substrate to position B2 to interact with protein residues. Additionally the side-chain ring of F120 stacks with B1 of the substrate. In our starting model of the substrate-angiogenin complex, the hydrogen bond to the backbone of L115 is retained without any stacking interactions. Residues L42 and N43 are also within hydrogen bonding distance of the substrate atoms. These residues were not included among those on whose side chains torsion angle grid searches were done because their role has not been biochemically elucidated.

Figure 3 shows a space-filled model of the protein with the docked substrate in stick representation. It can be noticed that even in the starting model, the substrate fits into pockets on the protein surface. The first base sits deep inside one such pocket. Figures 4a and b show all the residues that interact with the protein. These are the residues whose surface makes the pockets shown in Figure 3.

Molecular Dynamics Simulation of the Docked Model

RMS fluctuations

The native structure and the starting models of the angiogenin-substrate complexes were subjected to 1 nanosecond of MD simulation. Figure 5a shows the RMSD trajectory of native angiogenin (2Ang) and that of its two substrate (CpA and UpA) bound complexes with respect to their respective MD averaged structures. The CpA bound angiogenin structure shows the least RMSD, closely followed by that of the native structure. The UpA bound structure shows a much higher RMSD. This is indicative of the relative stability of the substrate bound structures. This trend is again apparent in the plot of residue-wise

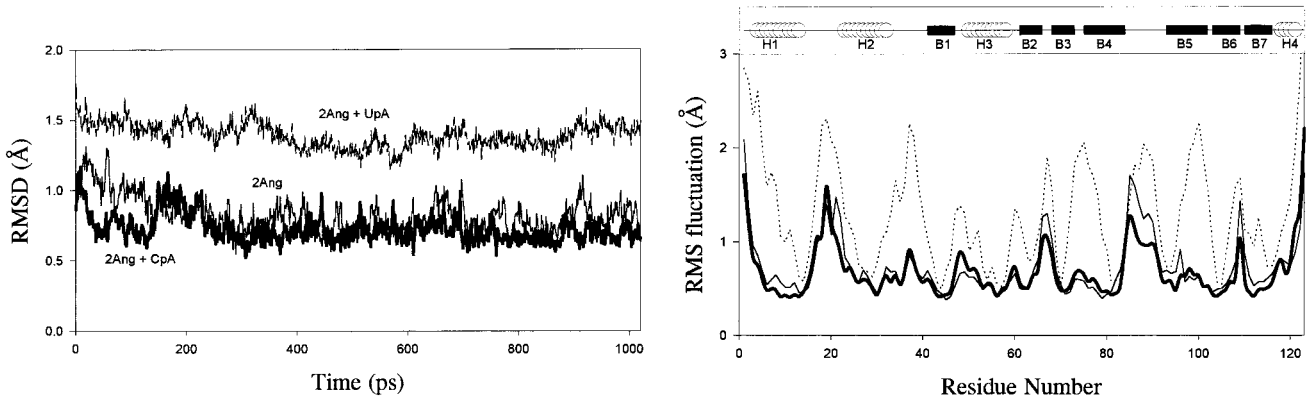

Fig. 5. **A:** RMSD trajectory of the native protein (thin line), CpA complex (thick line), and UpA complex (thin dashed line) from the respective MD simulations. **B:** Residue-wise RMS fluctuation during MD of the native protein (thin line), CpA complex (thick line), and UpA complex (dotted line).

RMS fluctuations shown in Figure 5b. The residues of angiogenin in the CpA bound complex fluctuate lesser about their mean positions than their counterparts in both the native and the UpA bound complex. The residues that fluctuate the most in the native structure are the same that do so in the CpA bound complex, but to a greater extent. These residues are all part of loops that connect secondary structures of the protein. The regions are the loops that connect the first and second helices (residues 14 to 22), strands 2 and 3 (residues 66,67,68), strands 4 and 5 (residue 84 to 92), and strands 6 and 7. Residues 66 to 68 interact with the bound substrate and they are therefore understandably better behaved in the CpA complex than in the native structure. The region encompassing residues 84 to 92 is also better behaved than it is in the native structure even though it is ~20Å away from the bound substrate. Most of the loop regions that connect two secondary structures show considerably high RMSDs in the angiogenin-UpA complex. Some of the loops such as those connecting the secondary structures (H2-B1), (B3-B4), and (B5-B6), which were fairly stable in the native structure exhibited high fluctuations in the UpA complex. The fluctuations of the residues in these loops in the UpA complex are also higher in magnitude than that of their counterparts in the CpA complex. The residue-wise RMS curve for the UpA complex is always above that of the CpA complex or that of the native (Fig. 5b). This indicates that the CpA complex is bound tighter than the UpA complex. This conforms to available experimental evidence that CpA is preferred over UpA as substrate.[4]

Protein-substrate hydrogen bonds

From the analysis of hydrogen bond trajectories between the substrate and the protein, a few key interactions mediated by hydrogen bonds have been picked up. These hydrogen bonds along with some important protein-protein and water-mediated hydrogen bonds have been listed in Table III. From this list, it can be seen that the biochemically[12] crucial interactions such as those of residues H13, K40, and T44 with the first base and ribose are retained. Important interactions of the phosphate with residues H114 and L115 are also retained. Interaction of the second base with residue N68 has emerged. In the case of CpA, the interaction with the second base was stabilized by a water-mediated hydrogen bond. These interactions are listed at the bottom of Table III and a detailed analysis on how these interactions contribute to substrate specificity is presented in the section on simulated annealing. The acceptor-donor distances from these hydrogen bonds were taken as distance restrains in the simulated annealing run.

Besides the hydrogen bonds listed in Table III, an analysis of the MD trajectory of the complex also showed the formation of hydrogen bonds between the substrate and other residues of the protein, I42, N43, and R5. We have mentioned earlier the proximity of the residues I42 and N43 to the substrate. Yet hydrogen bonds between these residues and the substrate are not stable; they form and break frequently in the course of the simulation. It is known that R5 interacts with the second Phosphate (P2) of the single stranded RNA chain.[32] It is therefore not surprising that the long arginine side chain in our model sometimes interacts through a hydrogen bond with O3′ of the ribose R2, the atom that extends as P2 as can be seen from Figure 4a. This interaction is also an indirect indication that the model proposed conforms to known biochemical evidence.

Unusual structure of loop connecting strands 2 and 3

A water molecule was found to be invariantly bound to both the Oϵ atoms of E67 and to the N6 atom of B2-Adenine (Fig. 4a) during the simulation. This interaction bolsters the direct protein-B2 interaction in the case of CpA. Interestingly, this interaction is absent in the case of UpA-angiogenin complex (Fig. 4b) though B2-adenine is the same in both substrates. One of the features that orients this region to forming the water-mediated hydrogen bond in the CpA complex is a network of hydrogen

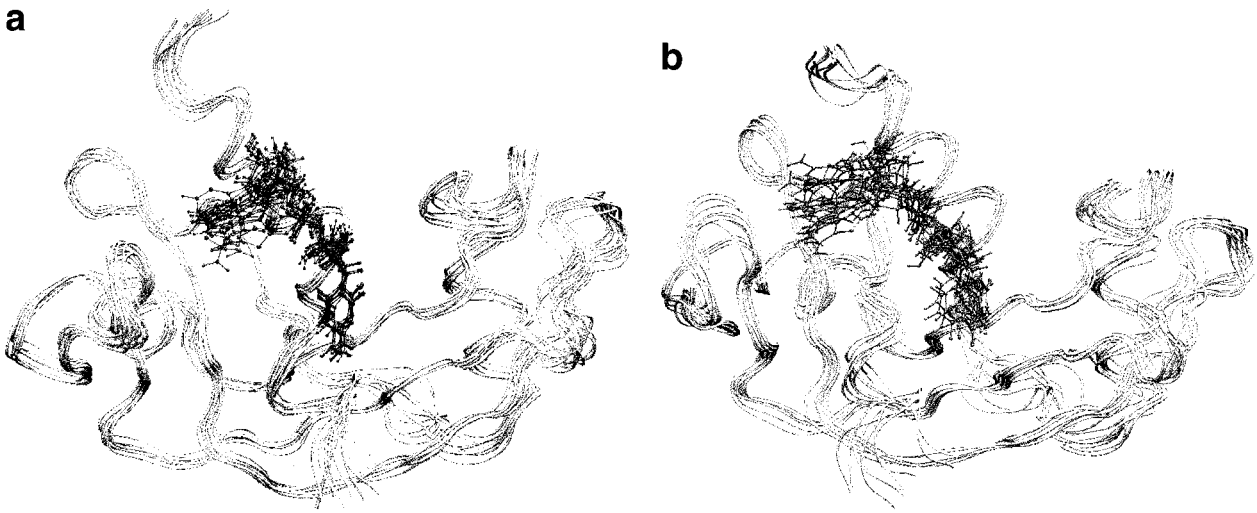

Fig. 6. The C^{α} trace of the 10 annealed structures of (**a**) CpA and (**b**) CpA superimposed on each other. The substrate is shown in ball and stick representation.[43]

bonds formed between the side chains of R66, E67, and N68 as shown in Figure 4a. This network is not present in the UpA-angiogenin complex. This network of side-chain hydrogen bonds in the CpA bound angiogenin structure introduces an unusual conformation, a II′ turn in the protein. The main-chain torsion angles of R66 fall into the unfavorable region of conformational space.[33,34] Peptide units that form a II′ turn take the ϕ–ψ values (60, -120) and (-80,0) in the i+1 and i+2 positions resulting in a hydrogen bond between the carbonyl oxygen of the i^{th} residue and the NH of the $i+3^{ed}$ residue. A distorted type II′ turn can deviate by 30° from these values.[35,36] Residues 65 to 68 in the native protein take up a distorted II′ turn. This turn is retained in the simulation and the annealed models of the CpA bound complex. The unusual II′ turn found in the native structure[7,8] and the CpA bound complex is not unique to this protein; its occurrence has been noticed in 24 proteins of a nonhomologous data set.[37] The ϕ–ψ of the i+1 residue of a II′ turn is normally allowed only for a glycine. However, arginine has occurred in this position in one other protein.[38] It should also be noted that in the case of the UpA-angiogenin complex R66 is in the allowed conformation (β) of the Ramachandran map. Further, the side-chain hydrogen bond network among the residues 66, 67, and 68 is broken in the UpA complex.

Simulated Annealing of the Bound Complexes

The starting structure for the MD simulation was an unrefined model of the substrate-protein complex. From the MD trajectory 10 structures were picked up at random and annealed to get refined models of the protein-substrates complex by constraining the interactions given in Table III. A comparison of the models of the two substrates brings out several interesting features.

Comparing the protein-substrate interactions in the CpA and UpA complexes and origin of substrate specificity

Figures 6a and b show the superimposition of the annealed models of the CpA and UpA complexes, respectively. The UpA complex shows more fluctuation than the CpA complex. This is a reflectance of the MD trajectory of which the starting structures of these annealed models were a part.

The donor-acceptor pairs whose separation was restrained were mainly between the substrate and the protein. There were also a few protein-protein hydrogen bonds that were restrained (see Table III). These connected residues T44 with T80, R66 with E67, and E67 with N68. The hydrogen bond between T44 and T80 is important in keeping Thr44 oriented toward the substrate. The importance of the T80 for enzymatic activity of angiogenin has been investigated biochemically.[39] These hydrogen bonds feature in all the annealed models. The side-chain hydrogen bonds formed between the Residues R66, E67, and N68 have also been used as distance restraints. Even in the annealed models, this network of side-chain hydrogen bonds is seen only in the case of the CpA complex. As mentioned earlier, the existence of this network conforms this loop to a distorted II′ turn in the CpA complex as found in the native enzyme. A strong preference to move away from this strained conformation in the UpA complex is an interesting example of substrate-induced conformational change in the protein structure.

Protein-substrate interactions emerge very clearly from the two sets of annealed structures. Identical atoms of the two dinucleotides interact with residues H13, K40, H114, and L115. Differences are seen in the interactions of the protein with the bases B1 and B2, despite using the same constraints during the process of simulated annealing. The Oγ1 atom of Thr44 interacts with the N4 of the cytosine in

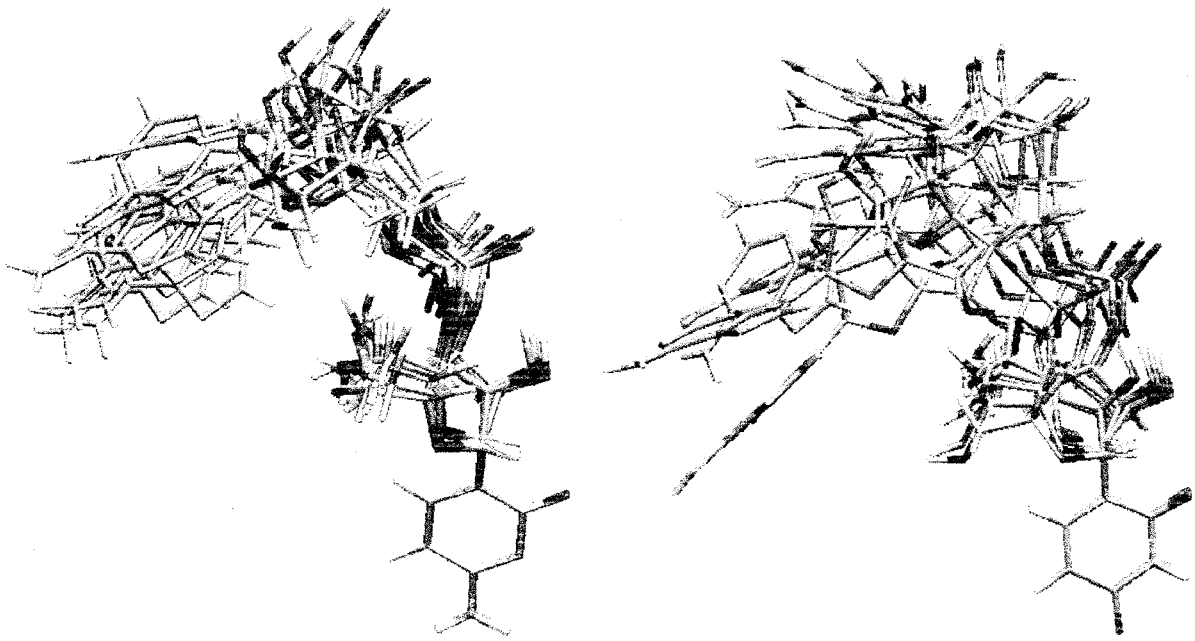

Fig. 7. Superimposition of the substrates (**a**) CpA and (**b**) UpA from the 10 annealed structures. Only the first base atoms were used for the superimposition.

the CpA complex and with that of the N3 atom of Uracil in the UpA complex. The N of Thr44 interacts with the N3 in the case of CpA and with O2 in the UpA bound models. This amounts to a small shift in the positioning of the first base in the two complexes. The absence of the RNase A equivalents of residues S123 and D83 and any water-mediated interactions[40] in this region leaves Thr44 to be the only residue that interacts with the first base. Because both the substrates interact with Thr44 with two hydrogen bonds it is not possible to infer the specificity of the protein from these interactions alone. However, a slight shift in the B1 position seems to have propagated into the B2 site. In the case of the UpA complex there is no water-mediated hydrogen bond between the substrate and E67 nor is there a network of side-chain hydrogen bonds between the residues R66, E67, and N68 that would facilitate the formation of a water-mediated bond. The only common interaction between B2 and the protein is between N68 and the N1 atom of adenine. The absence of an additional water-mediated hydrogen bond in the B2 site of the UpA complex results in less rigid binding of the base B2. Figures 7a and b show the conformational freedom of the substrates in the annealed models. It is clearly discernible that UpA samples a wider range of conformations than CpA, especially in the second base region. This can result in the increase of conformational entropy of the UpA complex. Further, the subtle changes between CpA and UpA binding, particularly with less number of hydrogen bonds in the B2 region, manifests itself in the form of ~50kcal/mol difference in interaction energy, as can be seen from the last columns of Tables IVa and b. The energy here pertains only to the enthalpic contribution to the total energy. We infer that the basis of substrate specificity of angiogenin is due to the subtle shift in substrate position, which cascades into significant differences in conformational entropy and interaction energy.

Comparing the angiogenin-substrate complexes with the RNase-UpA complex

Both angiogenin and RNase A catalyze the same reaction and have most of their active and binding sites conserved, a lot of protein-substrate interactions are also in common. Due to the structural and sequential homology between the two proteins, RNase A and its complexes were used as a standard throughout the docking exercise. The dissimilar interactions indicate the difference between the two systems. In work similar to that reported here, substrates of RNase A were docked onto its active site.[16,17] In angiogenin as in RNase A, the catalytic histidines interact with the phosphate oxygens. The Histidine that acts as a base (H13) also interacts with O2′ of the ribose sugar. The acidic Histidine (His 114) has interactions with the phosphate oxygens. K40 (K41 in RNase A), a residue crucial for the catalytic action, binds to the O2′ atom in both the protein complexes. T45 interacts with the pyrimidine base. However, the other residues that interact with the pyrimidine base in RNase A, S123 and D83, are not conserved in angiogenin and there are no equivalent interactions. Ribonucleotide binding specificity of RNase A is imparted by interaction by all the residues that bind directly (T45, D83) or through water (S123).[40] In the case of angiogenin the specificity is seemingly controlled by the interactions of T44 alone with the B1 base of the substrate. An analysis of this has been done in an earlier section.

Major differences between RNase A and angiogenin's binding to its RNA substrates come from the interaction

that the base B2 has with the proteins. In RNase A this base interacts with N71. Taragona-Fiol et al.[41] have shown that mutating this residue decreases RNase activity of RNase A by nearly three orders of magnitude. The loop to which this N71 belongs is not conserved from RNase A to angiogenin. It has also been shown that in a chimera where this loop in angiogenin is replaced by that from RNase A, the ribonuclease activity increases.[42] In our model of the angiogenin-substrate complex, the role of N 71 is taken up by N68.

CONCLUSIONS

We report here for the first time a modeled structure of human angiogenin bound to its substrates CpA and UpA. The substrates were docked onto the binding sites of the protein after the C-terminus of the protein was modeled to accommodate the first base of the dinucleotide substrates. The sugar pucker of the second ribose of the substrate was altered from C3′-endo to the energetically less favored C1′-endo in order to gain the desired interaction of the second base with the N68 residue of the protein. The modeled starting structures were refined by subjecting them to MD simulations. The stable trajectory of the RMSD plots for both complexes indicate good binding. Hydrogen bonds that stabilized substrate-protein interactions were identified from an analysis of the MD trajectories. These hydrogen bonds were then used as distance restraints in simulated annealing runs to further refine our models.

From an analysis of the annealed models we infer that the substrate specificity of angiogenin towards cytosine over Uracil in the first base position is not based on the interactions of the protein with the first base alone. For both Uracil and Cytosine our models have two interactions with atoms of Thr45. Subtle differences in these interactions lead to a orientation difference in the two substrates finally cascading into an energy difference of about 50kcal/mol in favor of CpA. It was also observed from both the MD and annealing studies that the conformational rigidity of the CpA in the complex is greater than that of UpA. This combined with smaller fluctuations of the protein residues in the CpA complex as compared with the UpA complex makes a more tightly bound complex.

Our models of the CpA complex place the residue R66 in the disallowed region of the Ramachandran map, leading to a distorted II′ turn, which is also a feature of the native structure but is not seen in the UpA complex. This disallowed conformation seems to be maintained because of a stable network of hydrogen bonds that bind the side chains of R66, E67, and N68. In the case of CpA, this network orients the side chain of E67 to have a water-mediated interaction with the N6 atom of the Adenine base. This interaction further stabilizes the CpA complex. No such interaction was found in the simulation of the UpA complex.

ACKNOWLEDGEMENTS

The authors thank K. Seshadri for useful discussions. We thank the Super Computer Education and Research Centre (SERC) of the Indian Institute of Science for computational facilities.

REFERENCES

1. Riordan JF. Structure and function of angiogenin. In: D'Alessio G, Riordan JF, editors. Ribonucleases: structures and functions. New York: Academic Press; 1997. p 445–489.
2. Fett JW, Strydom DJ, Lobb RR, Alderman EM, Bethune JL, Riordan JF, Vallee BL. Isolation and characterization of angiogenin, an angiogenic protein from human carcinoma cells. Biochemistry 1985;24:5480–5427.
3. Hallahan TW, Shapiro R, Vallee BL. Dual site model for the organogenic activity of angiogenin. Proc Natl Acad Sci USA 1991;88:2222–2226.
4. Acharya KR, Shapiro R, Riordan JF, Vallee BL. Crystal structure of bovine angiogenin at 1.5 Å resolution. Proc Acad Natl Sci USA 1995;92:2949–2953.
5. Acharya KR, Shapiro R, Allen SC, Riordan JF, Vallee BL. Crystal structure of human angiogenin reveals the structural basis for its functional divergence from ribonuclease. Proc Acad Natl Sci USA 1994;91:2915–2919.
6. Lequin O, Albaret C, Bontems F, Spik G, Lallemand J-Y. Solution structure of bovine angiogenin by ^{1}H nuclear magnetic resonance spectroscopy. Biochemistry 1996;35:8870–8880.
7. Lequin O, Thuring H, Robin M, Lallemand J-Y. Three-dimensional solution structure of human angiogenin determined by ^{1}H, ^{15}N-NMR spectroscopy. Eur J Biochem 1997;250:712–726.
8. Leonidas DD, Shapiro R, Allen SC, Subbarao GV, Veluraja K, Acharya KR. Refined crystal structures of native human angiogenin and two active site variants: implications for the unique functional properties of an enzyme involved in neovascularisation during tumour growth. J Mol Biol 1999;285:1209–1233.
9. Madhusudhan MS and Vishveshwara S. Modeling angiogenin -3′NMP complex. J Biomol Str Dyn 1998;16:715–722.
10. Madhusudhan MS, Vishveshwara S. Molecular dynamics simulations of modeled angiogenin-mononucleotide complexes. Curr Sci 2000;78:852–857.
11. Russo N, Shapiro R, Acharya KR, Riordan JF, Vallee BL. Role of glutamine-117 in the ribonucleolytic activity of human angiogenin. Proc Natl Acad Sci USA 1994;91:2915–2919.
12. Shapiro R, Weremowicz S, Riordan JF, Vallee BL. Ribonucleolytic activity of angiogenin: essential histidine, lysine and arginine residues. Proc Natl Acad Sci 1987;84:8783–8787.
13. Curran TP, Shapiro R, Riordan JF. Alteration of the enzymatic specificity of human angiogenin by site-directed mutagenesis. Biochemistry 1993;32:2307–2313.
14. Harper JW, Fox EA, Shapiro R, Vallee BL. Mutagenesis of residues flanking Lys-40 enhances the enzymatic activity and reduces the angiogenic potency of angiogenin. Biochemistry 1990; 29:7297–7302.
15. Zegers I, Maes D, Dao-Thi M-H, Poortmans F, Palmer R, Wyns L. The structures of RNase A complexed with 3′-CMP and d(CpA): active site conformation and conserved water molecules. Protein Sci 1994;3:2322–2339.
16. Seshadri K, Rao VSR, Vishveshwara S. Interaction of substrate Uridyl 3′,5′-adenosine with ribonuclease A: a molecular dynamics study. Biophys J 1995;69:2185–2194.
17. Seshadri K. Ph.D. thesis. 1994. Indian Institute of Science.
18. Seshadri K, Rao VSR, Vishveshwara S. Characterization of substrate UpA binding to RNase A— computer modelling and energetics approach. J Biomol Str Dyn 1994;12:581–603.
19. Nadig G, Vishveshwara S. Effects of substrate binding to RNase A: molecular dynamics simulations of UpA bound and native RNase A. Bioploy 1997:42:505–520.
20. Brunger AT, Adams PD, Rice LM. New applications of simulated annealing in X-ray crystallography and solution NMR. Structure 1997;5:325–336.
21. Clore GM, Nilges M, Sukumaran DK, Brunger AT, Karplus M, Gronenborn AM. The three dimensional structure of a-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. EMBO J 1986;5:2729–2735.
22. Kearsley SK. On the orthogonal transformation used for structural comparisons. Acta Crystallogr A 1989;45:208–210.
23. Goldstein H. Classical mechanics. New Delhi: Narosa Publishing House; 1980.

24. Pearlman DA, Case DA, Caldwell JW, et al. AMBER4.1 San Francisco: University of California, 1995.
25. Cornell WD, Cieplak P, Bayly CI, et al. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 1995;117:5179–5197.
26. Jorgensen WL, Chandrashekar J, Madura JD, Impey RW, Klien MLJ. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
27. Eftink MR, Biltonen RL. Energetics of ribonuclease A catalysis. 1. pH, ionic strength, and solvent isotope dependence of the hydrolysis of cytidine cyclic 2′,3′-phosphate. Biochemistry 1983;22:5123–5134.
28. Berendsen HJC, Postma JP, Van Gunstren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
29. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
30. Darden TA, York D, Pedersen LJ. Particle mesh ewald: an N.log(N) method for ewald sums in large systems. J Chem Phys 1993;98:10089–10092.
31. Mosimann SC, Newton DL, Youle RJ, James MN. X-ray crystallographic structure of recombinant eosinophil-derived neurotoxin at 1.83 Å resolution. J Mol Biol 1996;260:540–552.
32. Russo N, Acharya KR, Vallee BL, Shapiro R. A combined kinetic and modeling study of the catalytic center subsites of human angiogenin. Proc Natl Acad Sci USA 1996;93:804–808.
33. Ramachandran GN, Sasishekaran V. Conformations of polypeptides and proteins. Adv Protein Chem 1968;23:283–437.
34. Ramachandran GN, Ramakrishnan C, Sasishekaran V. Stereochemistry of polypeptid chain conformation. J Mol Biol 1963;7:95–99.
35. Wilmot CM, Thornton JM. Analysis and prediction of the different types of β-turn proteins. J Mol Biol 1998;203:21–232.
36. Wilmot CM, Thornton JM. β-turns and their distortions: a proposed new nomenclature. Protein Eng 1990;3:479–493.
37. Narayanan E, Ramakrishnan C. Secondary structure without backbone: an analysis of backbone mimicry by polar side chains in protein structures. Protein Eng 1999;12:447–455.
38. Lah MS, Dixon MM, Pattridge KA, Stallings WC, Fee JA, Ludwig ML. Structure-function in Escherichia coli iron superoxide dismutase: comparisons with the manganese enzyme from Thermus thermophilus. Biochem 1995;34:1646–1660.
39. Shapiro R. Structural features that determine the enzymatic potency and specificity of human angiogenin: threonine-80 and residues 58-70 and 116-123. Biochemistry 1998;37:6847–6856.
40. Gilliland GL, Dill J, Pechik I, Svensson LA, Sjolin L. The active site of bovine pancreatic ribonuclease: an example of solvent modulated specificity. Protein Pept Lett 1994;1:60–65.
41. Tarragona-Fiol A, Eggelte HJ, Harbron S, Sanchez E, Taylorson CJ, Ward JM, Rabin BR. Identification by site-directed mutagenesis of amino acids in the B2 subsite of bovine pancreatic ribonuclease A. Protein Eng 1993;6:901–906.
42. Harper JW, Vallee BL. A covalent angiogenin/ribonuclease hybrid with a fourth disulfide bond generated by regional mutagenesis. Biochem 1989;28:1875–1884.
43. Humphery W, Dalke A, Schulten K. VMD – visual molecular dynamics. J Mol Graph 1996;14:33–38.