Molecular dynamics simulations of modelled angiogenin–mononucleotide complexes

M. S. Madhusudhan and S. Vishveshwara*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Angiogenin, a protein belonging to the RNase superfamily stimulates the formation of blood vessels. For its biological activity, it also has to catalyse the cleavage of single-stranded RNA. Though experimental structures are known for the native protein, ligand-bound complexes of the protein at atomic detail have only been modelled recently. In the present study these modelled mononucleotide-bound angiogenin complexes have been subjected to 1 ns of molecular dynamics simulations. The structures obtained from the simulation are analysed for stability of the complexes and identifying persistent ligand–protein interactions. The results of the simulation are encouragingly similar to experimental observations. The present simulation studies indicate that the modelled complexes are stable and can thus be used as starting structures for further research on binding and catalysis.

1. Introduction

MOLECULAR dynamics (MD) simulation is an indispensable tool in structural biology. Data on macromolecular structure from X-ray crystallography and NMR spectroscopy are refined using MD[1,2]. MD is also used to elucidate structure–dynamics–function relationships. Recent reviews by Doniach and Eastman[3] and Pedersen and Darden[4] excellently elucidate the techniques and application of MD simulations to biological macromolecules. The starting structures for these simulations are usually experimentally determined. In this study, however, we have considered modelled structures of mononucleotide–angiogenin complexes for MD investigations. There are no experimentally determined structures of these complexes. Usually, experimentally determined structures do not include complexes of proteins with their substrates but only of substrate analogues. Modelling substrates onto the protein is thus a vital exercise to decipher protein function and activity[5].

Angiogenin is a 14 kDa protein that stimulates the formation of blood vessels. Though angiogenin is essential for embryonic development and tissue repair, it has also been found to abet the growth of cancer cells. To understand the working and mechanism of a protein, it is imperative that the structures of the native protein and that of its ligand–(substrates, transition state and products) bound complexes be known. Experimentally, structures[6–10] have only been determined for the native enzymes. The modelled complexed structures, which form the crux of this study, will help significantly in understanding angiogenin's activity. Though the exact mechanism of the protein's biological activity is still unclear, it has been structurally well-characterized because of its similarity to one of the best-studied proteins, RNase A. Angiogenin has a 35% sequence similarity to RNase A, besides having a very similar three-dimensional structure. Most of the active and binding site residues are conserved. These facts have been exploited while modelling the angiogenin–ligand complex with the RNase A–ligand complex as a template.

The mononucleotides 3′ CMP (Figure 1 *a*) and 3′ UMP (Figure 1 *b*) have been docked onto bovine angiogenin. The modelled structures of the complexes[11] are presented in Figure 2 *a* and *b*. These mononucleotides are the products of the cleavage of a single-stranded dinucleotide fragment of RNA, a minimal substrate, by angiogenin. This cleavage has been found to be vital for the enzyme's activity[12]. The RNA cleavage activity of angiogenin is, however, 10^5 times weaker than its homologue RNase A.

MD simulations carried out in this study serve as a

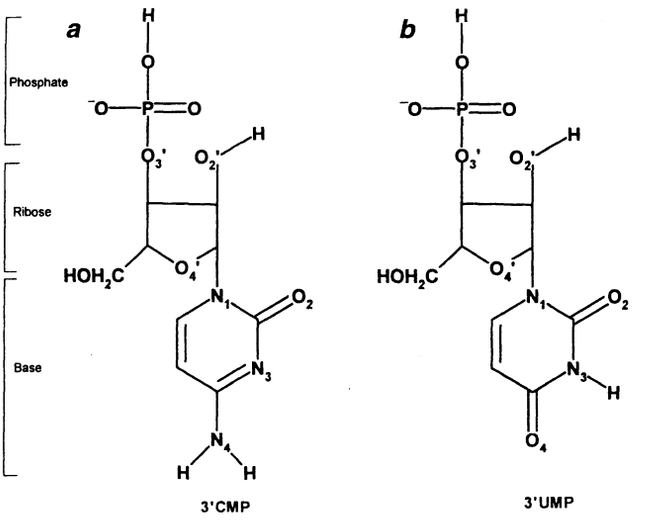

Figure 1. Schematic representation of the ligands 3′ CMP (*a*) and 3′ UMP (*b*). The different regions of the ligand are marked on the left.

*For correspondence. (e-mail: sv@mbu.iisc.ernet.in)

yardstick to estimate the stability (and hence the validity) of the modelled complex (Figure 2 *a* and *b*). Most importantly, in a stable model it will help identify the residues interacting with the ligand. The differences in the dynamics and interactions between the protein and different ligands will help understand the protein's ligand specificity. The interactions will give us vital clues on the protein activity, which may eventually play a significant role in designing drugs against angiogenin. The present study is a step in this direction.

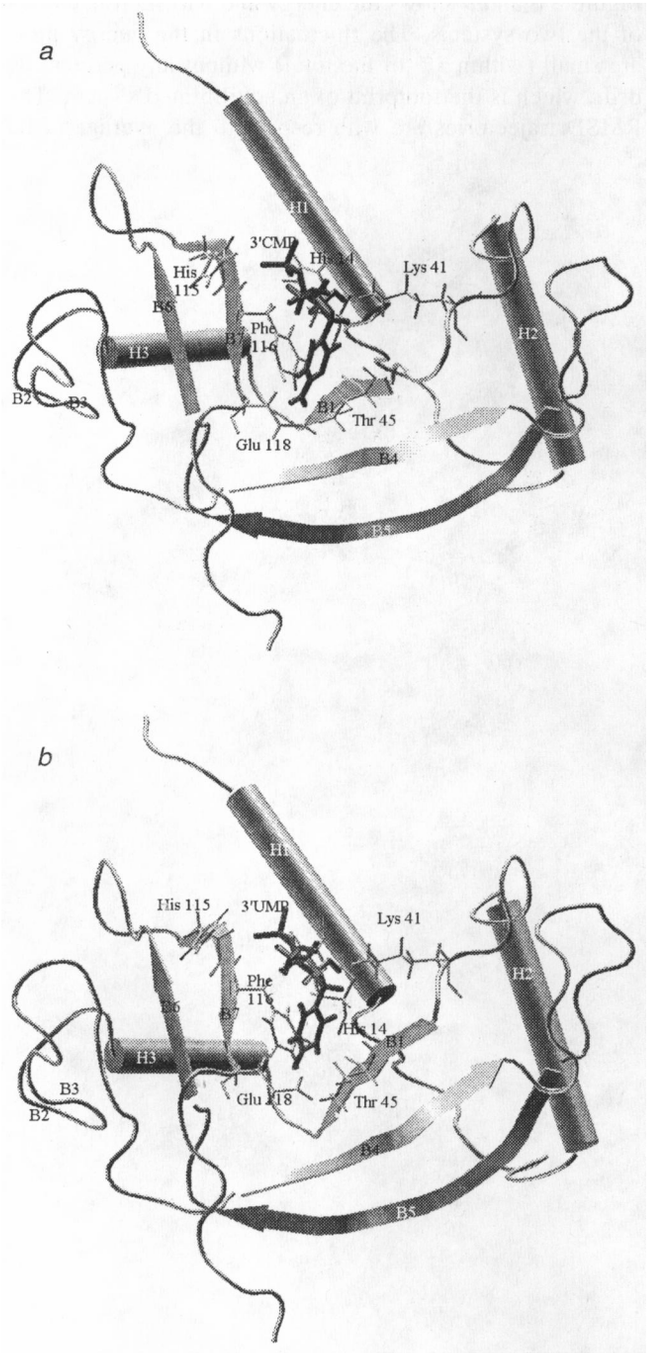

Figure 2. VMD (ref. 24) cartoon representation of the two modelled complexes of bovine angiogenin with 3′ CMP (*a*) and 3′ UMP (*b*). The ligand is shown in thick lines and the interacting residues of the protein in thin lines. Regions of interest are labelled alongside.

2. Methods

MD simulations were carried out using the parallel version of the SANDER module of the AMBER 4.1 suite of programs[13] on a 6 CPU Silicon Graphics power challenge machine. The starting structures of the simulations were the modelled structures of the angiogenin–pyramidine mononucleotide complex.

The systems were simulated using the all-atom force field[14]. The functional form of the force field is given below.

$$E_{\text{total}} = \Sigma K_r (r - r_{eq})^{2+} \Sigma K_\theta (\theta - \theta_{eq})^{2+} \Sigma V_n/2 \\ [1 + \cos(n\phi - \gamma)] + \Sigma(A_{ij}/R_{ij}^{12} - B_{ij}/R_{ij}^{6} + q_i q_j/eR_{ij}). \quad (1)$$

The first two terms in eq. (1) correspond to harmonic bond and angle stretching. The third term represents the torsion potential and the last term accounts for the van der Waals and electrostatic interactions.

MD in the SANDER module utilizes the Verlet algorithm to compute the trajectory of the atoms.

$$R_i^{n+1} = 2R_i^n - R_i^{n-1} + F_i^n h^2/m, \quad (2)$$

$$V_i^n = (R_i^{n+1} - R_i^{n-1})/2h, \quad (3)$$

where R_i^n is the position of the ith particle at the nth instant and V_i^n is its velocity. F is the force on the particle and h is the time-step of integration with m being the mass of the particle.

The system to be simulated was a box of size ~ 58.5 × 52.9 × 42.8 Å^3, enclosing the protein (whose centre of mass coincided with the centre of the box) and around 3200 water molecules. The protein is solvated by immersing it in a box of pre-equilibrated Monte Carlo water molecules. All water molecules within the Van der Waals surface of the protein are removed. Water molecules that surround the protein provide a minimum of two hydration shells. The water molecules and their force field parameters were represented by the TIP3P model[15]. The total number of atoms in the two systems (3′ CMP complex and 3′ UMP complex) was 11772 and 11735, respectively. The side chains of all polar amino acids were considered to be charged. The side chain of Histidine 14, however, was protonated only in the N^ε position[16].

The systems were subject to 1000 cycles of minimization, the first 200 cycles using the steepest descent method and the other 800 by the conjugate gradient algorithm. Minimizing the system rearranges the solvent water

molecules optimally around the protein. The minimized structure was then subject to 1.02 ns of MD simulation where, in the first 0.02 ns the system was coupled to a heat bath (NVT ensemble) with a coupling constant = 0.1 ps (ref. 17). The temperature was gradually increased from 0 to 300 K during this period. For the next 1 ns, the system was decoupled from the heat bath and simulations were carried out in the NVE ensemble. The SHAKE algorithm was used to constrain all bond lengths, iteratively using Lagrange multipliers[18]. An integration step of 2 femtoseconds was used. The dielectric constant was kept at a constant value of 1. The velocities were rescaled every 2 ps if the temperature of the system diverged from 300 by more than 10 degrees. Electrostatic interactions were evaluated using fast Fourier transforms by the Particle Mesh Ewald sum (PME) method, which uses a B-spline interpolation order[19,20]. Periodic boundary conditions were used. A residue based cut-off list (Verlet list) to calculate non-bonded interactions (within 12 Å) was updated every 25 steps. A 1 ns simulation takes about 170 h of CPU time to complete.

3. Analysis

Root mean square deviations (RMSD) of the simulation structures with respect to a reference structure were found using an algorithm of Kearsley[21]. This method uses a 3D least square fit to optimally superimpose structures. The criterion to detect hydrogen bonds was that the donor–acceptor distance be less than 3.5 Å and simultaneously the proton–acceptor distance be no greater than 2.8 Å. All programs for analysis were developed by the group and are coded in FORTRAN 77.

4. Results and discussion

Figure 3 *a* and *b* shows the energy and RMSD trajectories of the two systems. The fluctuations in the energy plots are small (within 3% of the total) without any perceptible drift, which is the footprint of an equilibrated system. The RMSD trajectories are with respect to the averaged MD

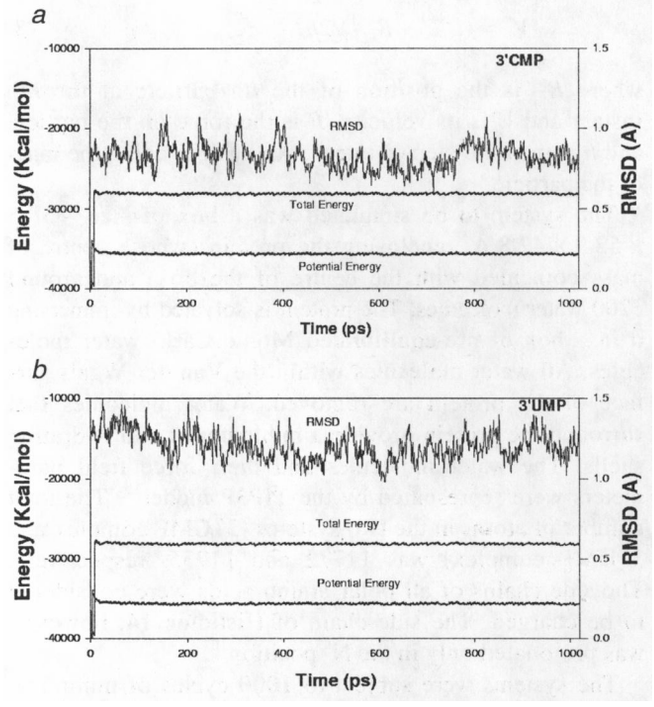

Figure 3. Energy and RMSD trajectories of the two complexes, angiogenin with 3′ CMP (*a*) and 3′ UMP (*b*). The RMSD trajectories are with respect to the ⟨MD⟩.

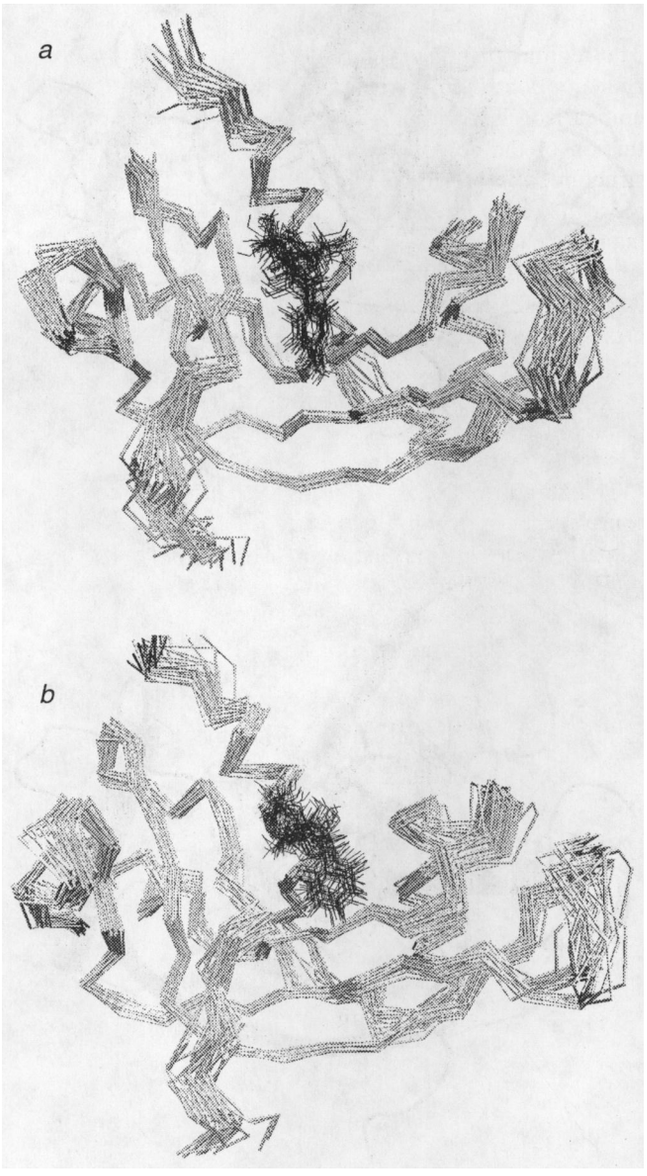

Figure 4. Superimposition of 25 snapshots of the C^{a} traces (grey) of the protein in the two complexes, angiogenin with 3′ CMP (*a*) and 3′ UMP (*b*) along with the ligand (shown in black). Snapshots have been extracted after every 40 ps of MD.

(⟨MD⟩) structure. In both cases the deviation from the ⟨MD⟩ is around 1 Å, indicating that the complexes were stable during the simulation. However, in the 3′ CMP complex, the deviation from the ⟨MD⟩ structure is less than that of the 3′ UMP complex from its ⟨MD⟩ structure.

Figures 4 *a* and *b* are superimposition of 25 snapshots from the MD simulation, taken after every 40 ps of simulation time. The regions of regular secondary structure show lesser flexibility than the loop regions. In general, all the loop regions of the 3′ UMP complex fluctuate more in comparison with the 3′ CMP complex. The effect is visually apparent in the loops connecting the secondary structures H3 and B2, B2 and B3, and B4 and B5 (see Figure 2). These features are also borne out in Figure 5 *a* and *b*, which show the residue-wise RMS fluctuation of the protein residues about their MD averaged position. The regions of non-regular secondary structure stand out as major contributors to the proteins' overall flexibility. A comparison is also made in Figure 5 *a* and *b*, of the fluctuation of the complex structures with that of the native structure[11]. Interestingly, in the case of the 3′ CMP complex, the residues of the complexed protein fluctuate less than that of the native one but the trend is just the opposite in the case of the 3′ UMP complex. This behaviour could be a direct effect of the ligand specificity of the protein. It has been shown experimentally that cytosine is preferred over uracil in the first base position[12]. Protein residues in a tightly bound complex fluctuate little. The residue-wise RMS fluctuation plots qualitatively assert this fact.

The conformational flexibility of the ligands in the complexes is shown in Figure 6 *a* and *b*. The figure shows the trajectories of the phase angle P of the ribose sugar and that of the glycosidic torsion angle χ. The initial ring pucker of the ribose sugar was C3′ endo. But as the simulation progresses the pucker shifts towards C4′ exo. In the case of the 3′ CMP complex there is a further shift to the C2′ exo pucker. Though the 3′ endo pucker is energetically most favoured (in isolation), the persistence of other puckers has also been reported in earlier models[11].

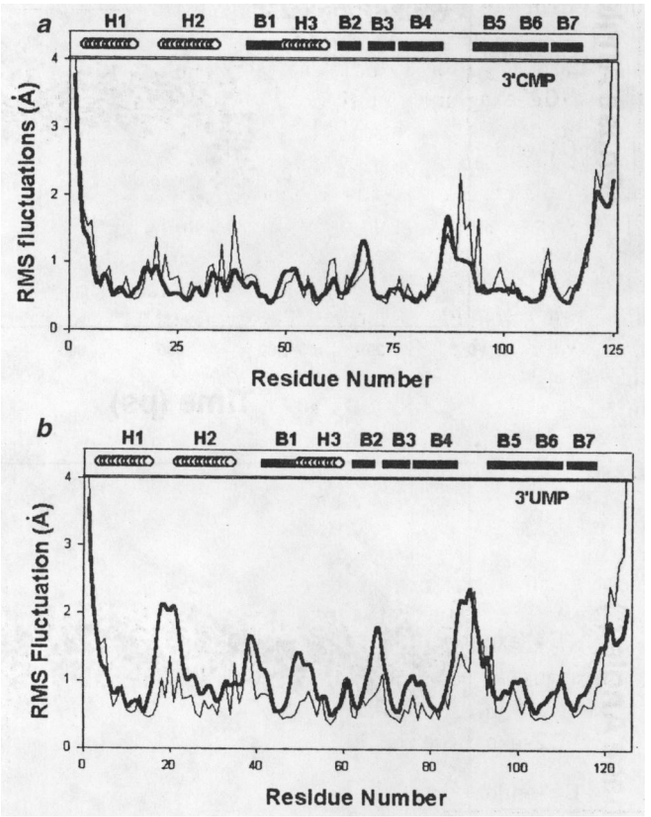

Figure 5. Residue-wise RMS fluctuation of the two complexes {*a*, 3′ CMP and *b*, 3′ UMP} about their ⟨MD⟩ structure (thick lines). The thin lines show the RMS fluctuation of the native structure about its ⟨MD⟩ structure.

Table 1. Hydrogen bonds between angiogenin and the mononucleotides and their average values during the simulation

Protein		Ligand (3′ UMP)	Distance (Å)	Protein		Ligand (3′ CMP)	Distance (Å)
Atom	Residue	Atom		Atom	Residue	Atom	
Nε2	GLN 13	OP3	2.98	Nε2	GLN 13	OP2	3.94
Nε2	HIS 14	OP2	2.94	Nε2	HIS 14	OP2	3.53
Nζ	LYS 41	O2	2.94	Nξ	LYS 41	O2	2.62
N	THR 45	O4	2.76	N	THR 45	N3	3.11
Oγ1	THR 45	O4	2.67	Oγ1	THR 45	N4	2.69
				Oγ1	THR 45	N3	3.37
Nδ1	HIS 115	OP1	2.81	Nδ1	HIP 115	OP1	3.46
N	PHE 116	OP2	2.78	N	PHE 116	OP1	3.43
O	ARG 43	N3	2.93	Oε1	GLU 118	N4	3.32
				Oε2	GLU 118	N4	2.91

The criterion for hydrogen bond detection is given in the methods section. Interactions that are common to the two complexes are shaded.

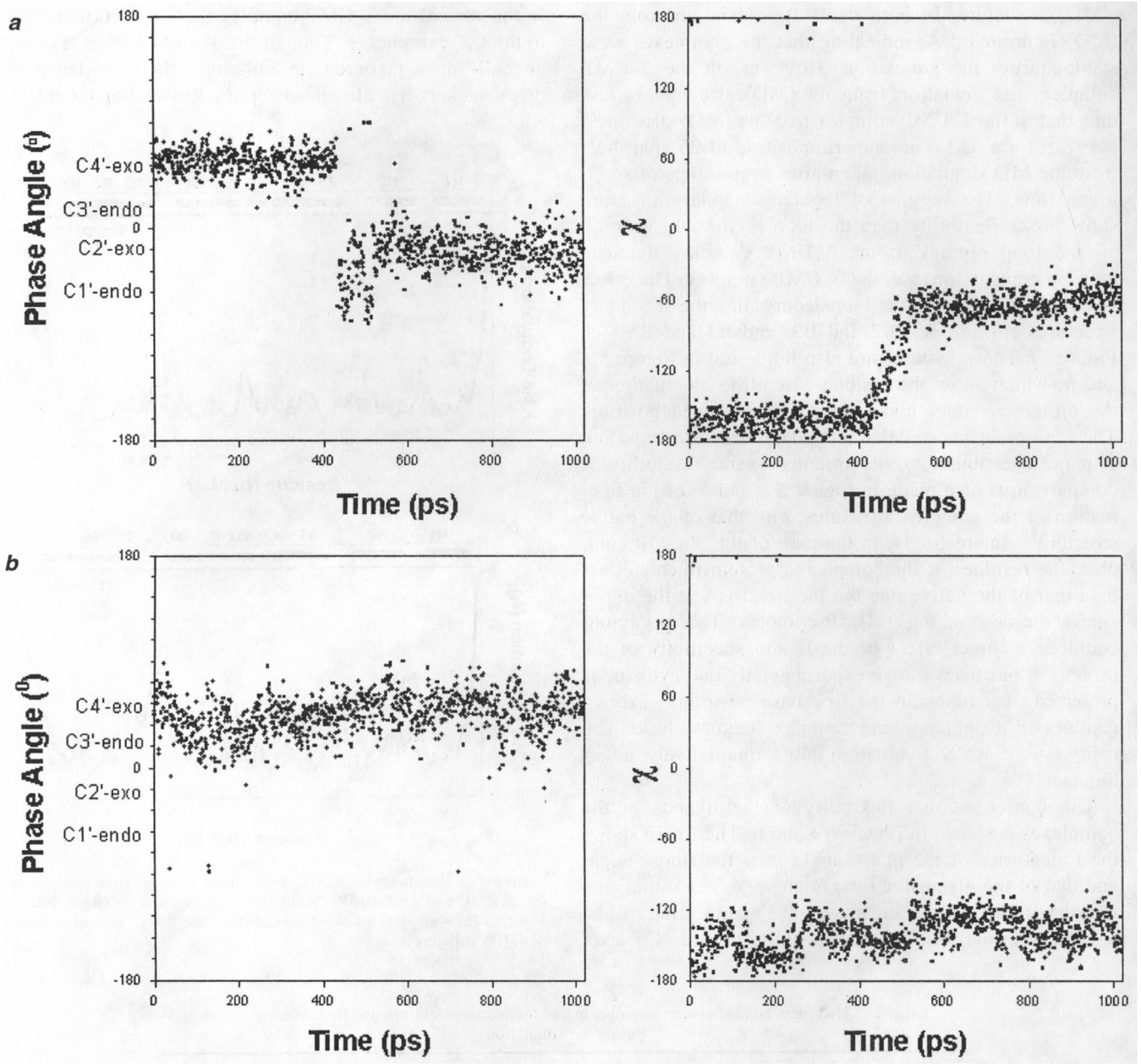

Figure 6. Trajectories of the phase angle of the ribose sugar and the glycosidic torsion angles of the ligands 3′ CMP (*a*) and 3′ UMP (*b*).

In the case of the 3′ CMP ligand, the change in pucker and the change in the glycosidic torsion angle χ are correlated. It is not surprising to see more conformational variation in the sugar and phosphate regions, which are the free end of the ligand. Though changes in the ligand conformation are observed, the position of the pyrimidine relative to the protein is not significantly altered. The details of the protein–ligand interactions, and its changes with ligand conformational flexibility are further discussed below.

The protein–ligand interactions are summarized in Table 1. Interactions that are common to the two complex systems are shaded. Most of the RNase A–ligand interactions are reproduced in our model. The two catalytic Histidines (His 14 and His 115) along with Gln 13 and Phe 116 interact with the phosphate oxygens. Although these residues are constantly hydrogen bonded with the phosphate group, it is not always to the same atom of the 3′ CMP ligand. This is because of the fluctuations in the glycosidic torsion angle χ and the ribose pucker, which result in interactions with different oxygen atoms of the phosphate group at different times. The flexibility of the phosphate group in the mononucleotide complexes leads to a stronger interaction of the phosphate oxygens with the catalytic His 14. Normally, as in RNase A–dinucleotide complexes[22], this histidine interacts with the

O2′ of the ribose. Thr 45, Lys 41 and Glu 118 (only in the 3′ CMP complex) interact with the base. Also, in RNase A–dinucleotide complexes Lys 41 interacts with the O2′ of the ribose. In the present case, this interaction is shifted to O2 of the base, which is in close proximity to the O2′ atom of the ribose. Similarly the interactions of the OG1 of Thr 45 have also been shifted (refer to Figure 1 *a* and *b*) to the N4/O4 atoms of the base instead of the N3 atom. This shift in the ligand position might have an important bearing on the catalytic action of angiogenin when compared to RNase A. As stated earlier, angiogenin is less efficient catalytically than RNase A. The details of the ligand–protein interactions presented here indicate that the differences in interactions of the two complexes have two origins, one being the additional interactions of the Thr 45 with 3′ CMP and the other being the interactions of Glu 118 with the N4 of 3′ CMP. There is no interaction of Glu 118 with any of the base atoms in the case of 3′ UMP. Biochemical studies indicate that specificity may not be linked to interactions with Glu 118 (ref. 23). From the simulations reported in the present study, the protein–ligand interaction energies favour 3′ CMP over 3′ UMP by around 10–15 kcal/mol. Since the mononucleotide ligands only happen to be the products of catalysis of the enzyme, it is not a comprehensive model to give atomistic details of ligand specificity. Simulations on complexes where the ligands are substrates will provide these details. Work in this direction is in progress.

5. Conclusions

We have subjected the first available angiogenin–ligand complex (modelled) to MD simulations to verify its stability and to understand the protein–ligand interactions. The simulations showed that the complexes were stable over a period of a nanosecond, which augured well for the model. The simulated trajectories were analysed for stable protein–ligand interactions. Most of the starting interactions were retained. The interactions varied subtly from those seen in the complexes of RNase A and its ligands. These variations give us a broad hint on angiogenin's reduced RNase activity. The RMSD trajectory of the 3′ CMP complex was consistently less than that of the native structure, whereas the situation was just the opposite in the 3′ UMP complex. The interaction energy for the 3′ CMP complex is stronger than that for the 3′ UMP complex. Interactions of the base of the 3′ CMP with the Oε atoms of the Glu 118 side chain are missing in the case of 3′ UMP. All these features are consistent with the experimental observation that cytosine is preferred over uracil in the first base binding site. The emphasis of this study is more on the stability of ligand-bound complexes than on substrate specificity, because deducing the reasons for substrate specificity from a complex of a ligand that is not a substrate (in this case the ligands are the products) may not yield complete information. In summary, we have the first atomic level detail picture on angiogenin–ligand (product) interaction, which is consistent with biochemical experimental data. This model can serve as a good starting point for further investigations.

1. Brunger, A. T., Adams, P. D. and Rice, L. M., *Structure*, 1997, **5**, 325–336.
2. Brunger, A. T., Kurian, J. and Karplus, M., *Science*, 1987, **235**, 458–460.
3. Doniach, S. and Eastman, P., *Curr. Opin. Struct. Biol.*, 1999, **9**, 157–163.
4. Pedersen, L. and Darden, T., in *Encyclopedia of Computational Chemistry* (ed. Schleyer, P. V.), vol. 2, pp. 1650–1659.
5. Seshadri, K., Rao, V. S. R. and Vishveshwara, S., *Biophys. J.*, 1995, **69**, 2185–2194.
6. Acharya, K. R., Shapiro, R., Riordan, J. F. and Vallee, B. L., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 2949–2953.
7. Acharya, K. R., Shapiro, R., Allen, S. C., Riordan J. F. and Vallee B. L., *Proc. Acad. Natl. Sci. USA*, 1994, **91**, 2915–2919.
8. Lequin, O., Albaret, C., Bontems, F., Spik, G. and Jean-Yves-Lallemand, *Biochemistry*, 1996, **35**, 8870–8880.
9. Lequin, O., Thuring, H., Robin, M. and Jean-Yves-Lallemand, *Eur. J. Biochem.*, 1997, **250**, 712–726.
10. Leonidas, D. D., Shapiro, R., Allen, S. C., Subbarao, G. V., Veluraja, K. and Acharya, K. R., *J. Mol. Biol.*, 1999, **285**, 1209–1233.
11. Madhusudhan, M. S. and Vishveshwara, S., *J. Biol. Struct. Dyn.*, 1998, **16**, 715–722.
12. Riordan, J. F., in *Ribonucleases: Structures and Functions* (eds D'Alessio, G. and James F Riordan), Academic Press, New York, 1997, pp. 445–489.
13. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheetam, T. E. III, Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K. and Kollman, P. A., *AMBER4.1*, University of California, San Francisco, 1995.
14. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A., *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
15. Jorgensen, W. L., Chandrashekar, J., Madura, J. D., Impey, R. W. and Klien, M. L. J., *J. Chem. Phys.*, 1983, **79**, 926–935.
16. Eftink, M. R. and Biltonen, R. L., *Biochemistry*, 1983, **22**, 5123–5134.
17. Berendsen, H. J. C., Postma, J. P., Van Gunstren, W. F., DiNola, A. and Haak., J. R., *J. Chem. Phys.*, 1984, **81**, 3684–3690.
18. Ryckaert, J. P., Ciccotti, G. and Berendsen, H. J. C., *J. Comput. Phys.*, 1977, **23**, 327–341.
19. Darden, T. A., York, D. and Pedersen, L. J., *J. Chem. Phys.*, 1993, **98**, 10089–10092.
20. Cheetam, T. E. III, Miller, J. L., Fox, T., Darden, T. A. and Kollman, P. A., *J. Am. Chem. Soc.*, 1995, **117**, 4193–4194.
21. Kearsley, S. K., *Acta Crystallogr.*, 1989, **A45**, 208–210.
22. Zegers, I., Maes, D., Dao-Thi, M. H., Poortmans, F., Palmer, R. and Wyns, L. *Protein Sci.*, 1994, **3**, 2322–2339.
23. Russo, N., Shapiro, R., Acharya, K. R., Riordan, J. F. and Vallee, B. L., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 2920–2924.
24. Humphery, W., Dalke, A. and Shulten, K., *J. Mol. Graphics*, 1996, **14**, 33–38.

ACKNOWLEDGEMENTS. We thank the Supercomputer Education and Research Centre of the Indian Institute of Science for computational facilities and the Department of Science and Technology for supporting the project.