Indian Journal of Biochemistry & Biophysics
Vol. 38, February & April 2001, pp. 27-33

A molecular dynamics study based *post facto* free energy analysis of the binding of bovine angiogenin with UMP and CMP ligands

M S Madhusudhan and S Vishveshwara*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India
Achintya Das, Parul Kalra and B Jayaram*
Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

Accepted 3 October 2000

Angiogenin is a protein belonging to the superfamily of RNase A. The RNase activity of this protein is essential for its angiogenic activity. Although members of the RNase A family carry out RNase activity, they differ markedly in their strength and specificity. In this paper, we address the problem of higher specificity of angiogenin towards cytosine against uracil in the first base binding position. We have carried out extensive nano-second level molecular dynamics(MD) computer simulations on the native bovine angiogenin and on the CMP and UMP complexes of this protein in aqueous medium with explicit molecular solvent. The structures thus generated were subjected to a rigorous free energy component analysis to arrive at a plausible molecular thermodynamic explanation for the substrate specificity of angiogenin.

Introduction

Angiogenin is a 14 kd protein and a potent inducer of angiogenesis. The induction of cell proliferation by angiogenin is known to be associated with its binding to an endothelial cellular receptor and transduction of secondary messenger response[1]. Sometime during the process of angiogenesis, angiogenin cleaves single stranded RNA. Recent reviews[2,3] deal with the biological function, structure and specificity of angiogenins. This protein belongs to the superfamily of RNase A and is sequentially and structurally similar to bovine pancreatic RNase A. Although the primary function of angiogenin is to induce neovascularization, its ribonucleolytic activity seems to be vital for its primary function. It has been found that the ribonuclease activity of angiogenin is much weaker than that of RNase A. It cleaves substrates that have a pyrimidine base at the 3′-end and a purine at the second base as in the case of RNase A. However, the most important difference is that a cytosine base is preferred over uracil for the first base position in angiogenin.

The structural details and the concomitant energetics of the ligand bound complexes of the proteins are essential to answer specificity related questions at a quantitative level. So far only the native structures of bovine and human angiogenins[4-8] are available and no ligand bound crystal structure of angiogenin is available. A superposition of the native structure of angiogenin on the ligand bound structure of RNase A, indicates that the active site is blocked by the C-terminal segment, both in the human and bovine angiogenins[4,5]. A continued effort in modeling the ligand bound structure of angiogenin is being made in one of our laboratories[9-11] and recently, we have obtained stable structures of CMP and UMP bound complexes of bovine angiogenin by docking and molecular dynamics procedures. Representative molecular dynamics snapshots of these complexes are given in Figs 1a and 1b.

A detailed analysis of the protein-ligand interactions showed that, crucial interactions such as those between T45 and pyrimidine bases, H14, K41 and H115 with ribose-phosphate group of the ligand are retained during the simulation[11]. Further, the obstructing residue, E118, now interacts with N4 of CMP. The substrate specificity of angiogenin for cytosine over uracil emerged from one nanosecond simulations on CMP and UMP bound complexes in the form of interaction energy, additional interaction of E118 with CMP and reduced RMS fluctuation of the CMP complex when compared to the native enzyme and the UMP complex. The exact reason for preference of cytosine however, could not be deduced from the MD simulations. It is important to understand the reasons for substrate specificity which

*Authors for correspondence
E-mail: sv@mbu.iisc.ernet.in;
bjayaram@chemistry.iitd.ernet.in

can be helpful in drug design strategies. A recent methodology due to Beveridge and Jayaram[12-14] quantifies and analyzes the various components of free energy for non-covalent associations in a rigorous manner and allows one to identify the cause/driving force for binding, in terms of components such as electrostatic, hydrophobic, cavity formation, adaptation and so on. In the present study, one

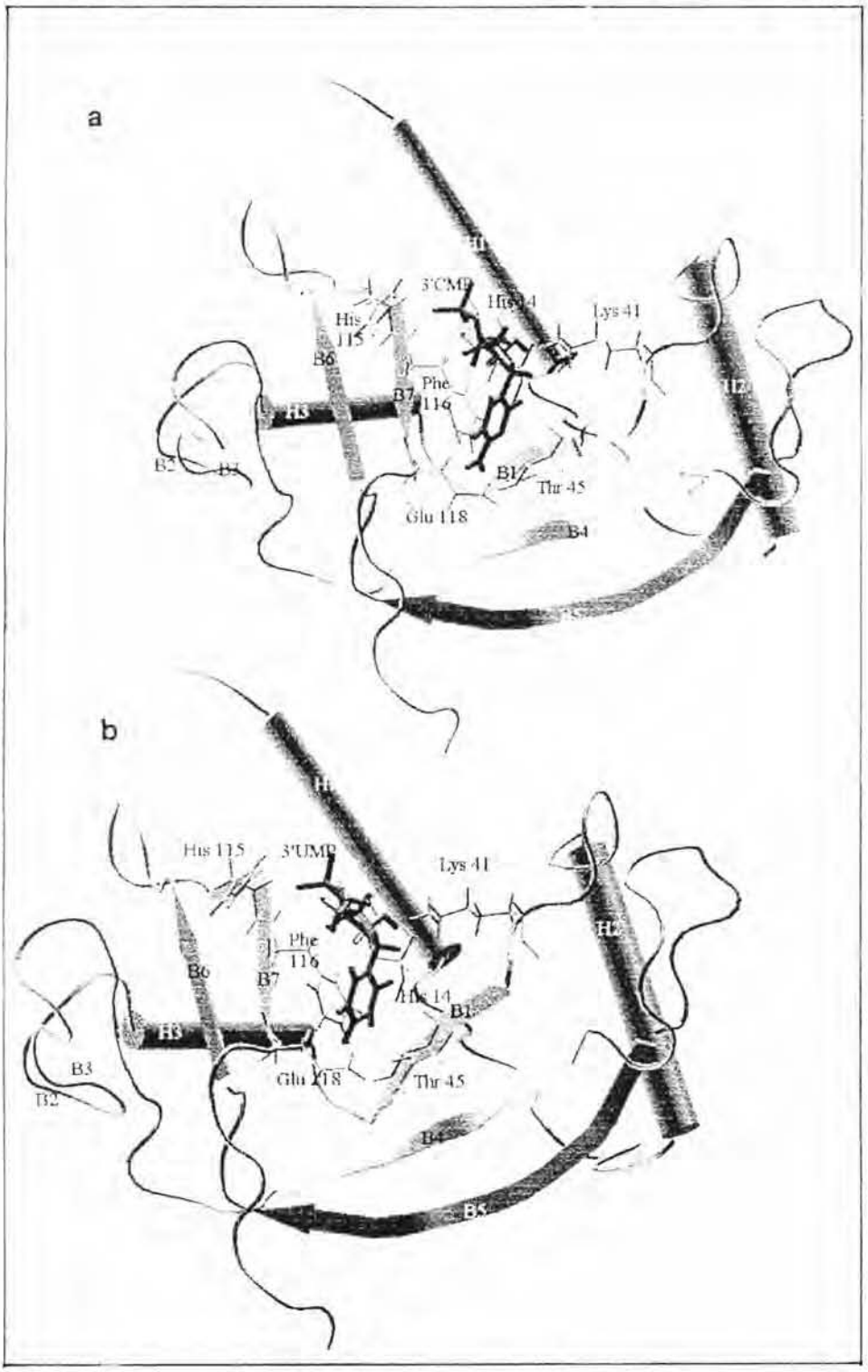

Fig. 1—VMD[18] Diagramatic representation of the two modelled complexes of bovine angiogenin with (a) CMP and (b) UMP. [The ligand is shown in thick lines and the interacting residues of the protein are represented by thin lines. The secondary structures and some of the loops are labeled]

nanosecond MD trajectories of bovine angiogenin[9], UMP and CMP bound complexes of bovine angiogenin[11] were used as input for free energy component analyses. An unambiguous molecular thermodynamic explanation for substrate specificity of angiogenin at the B1 binding site has emerged out of this study.

Methodology

The net standard free energy of binding is treated as a sum of a near comprehensive set of individual contributions. With the assumption of additivity and an arbitrary although rational selection of terms, component analysis is not theoretically rigorous and one can expect at best only a semi-quantitative account, hence expectations must be framed accordingly. However for complex processes such as protein-ligand binding no viable alternative exists at present. Simple identification of the important terms, estimates of their relative magnitudes and determination of whether they make favourable or unfavourable contributions to the free energy of complexation provides potentially useful new knowledge in the context of drug design endeavours.

The thermodynamic cycle for protein-ligand binding in solution used in this study is presented in Fig 2. Here the net binding process is decomposed into six steps. Step I is the process of converting the uncomplexed protein denoted "P" to the form "P*" in which the protein has adapted its structure to that of the ligand bound form. The free energy change for this step is given as :

$$\Delta G_{I}^{0} = \Delta G_{1}^{adpt,P} \quad \ldots (1)$$

Step II is the corresponding structural adaptation of the ligand required to convert the uncomplexed form "I" to the complexed form "I*" in solution. The change in free energy is :

$$\Delta G_{II}^{0} = \Delta G_{2}^{adpt,I} \quad \ldots (2)$$

The next two steps (III and IV) involve desolvation of P* and I* from aqueous medium to vacuum. The free energy change for each of these steps is written as a sum of four components.

$$\Delta G_{m}^{0} = \Delta G_{3}^{el,P} + \Delta G_{4}^{vdw,P} + \Delta G_{5}^{cav,P} + \Delta G_{6}^{DH,P} \quad \ldots (3)$$

$$\Delta G_{IV}^{0} = \Delta G_{7}^{el,I} + \Delta G_{8}^{vdw,I} + \Delta G_{9}^{cav,I} + \Delta G_{10}^{DH,I} \quad \ldots (4)$$

This involves contributions from electrostatic effects of desolvating the macromolecule (el), the van der Waals interactions with the solvent (vdw), elimination of the solvent cavity (cav) in which the molecule is accommodated and the change in added salt effects (DH). The transfer from aqueous medium to vacuum in steps III and IV involves the loss of favourable electrostatics and van der Waals interactions with the solvent and a gain from the cavity term. The latter, of course, is the reverse of the free energy of cavity formation. The free energy of interaction with added salt is also lost on desolvation.

In step V, the protein and the ligand associate as a non-covalently bound complex. The thermodynamics of this step can be described as:

$$\Delta G_{V}^{0} = \Delta H_{11}^{el,C} + \Delta H_{12}^{vdw,C} - T\Delta S_{13}^{tr+rot} - T\Delta S_{14}^{vib+conf} \quad \ldots (5)$$

Complexation involves introducing the electrostatic and van der Waals interactions between the protein and the ligand *in vacuo*. A change in external entropy due to loss of translational and rotational degrees of freedom occurs, which always disfavours complexation. The lost external modes are converted into low frequency internal vibrational and configurational

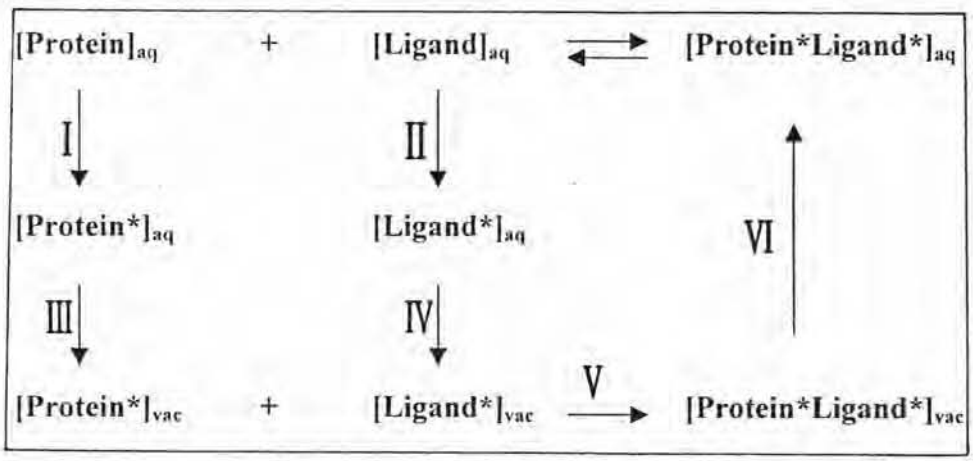

Fig. 2—The thermodynamic cycle used for a component-wise estimation and analysis of the binding free energies.

degrees of freedom in the complex and are reflected along with motional changes occurring as a consequence of the burial of amino acid side chains on complexation in the corresponding change in vibrational and configurational entropy.

In step VI, the complex is transferred from vacuum back to aqueous solution and the free energy change is due to solvation of the complex.

$$\Delta G^{0}_{VI} = \Delta G^{el,C}_{15} + \Delta G^{vdw,C}_{16} + \Delta G^{cav,C}_{17} + \Delta G^{DH,C}_{18} \quad \ldots (6)$$

Here again an electrostatic component, a van der Waals component, a cavity formation component and added salt effects are involved. While the cavitation term is unfavourable, all the other terms are favourable to solvation in this step.

In summary, the binding process in solution as considered here consists of six well-defined thermodynamic steps each of which can be decomposed into physically meaningful thermodynamic components. The total number of individual contributions to the free energy of binding in this model is 18.

The electrostatic contribution to solvation is calculated via the generalized Born model using the effective radii parameters derived by Jayaram *et al.*[12] consistent with the AMBER parm94 set. The nonelectrostatic contribution to solvation involves molecular surface area calculations, performed using the ACCESS program based on the algorithm of Lee and Richards. Further details on the evaluation of each component, the full theory and methodology for obtaining thermodynamic indices of macromolecular complexation are available[13-16]. The new methodological feature in this work, in relation to our earlier studies on protein-ligand interactions[15,16], is the incorporation of molecular dynamics structures and ensemble averages in the energy analysis.

A *post facto* free energy analysis of binding based on MD trajectories is implemented as follows. About 100 structures are culled from each of the MD trajectories i.e. on the native protein, and on the complexes with UMP and CMP developed separately with explicit waters. The various components of the binding free energy (totaling to 18 for each system) are then calculated, stripping off explicit waters, with a generalized Born dielectric continuum treatment of the solvent consistent with AMBER internal energetics. The structural adaptation expense of the ligands UMP and CMP, i.e. ΔG^{0}_{II}, is considered to be negligible–a reasonable assumption for fairly rigid small ligands, as independent MD trajectories on ligand-alone are not available. Also, salt effects are not included in this analysis to be consistent with MD set up. Results on each of the components enumerated above as contributing to binding for each system are listed in Table 1 and an analysis of binding in terms of compounded subsets signifying conventionally implicated physico-chemical forces are presented below.

Results and Discussion

The net free energy of binding is a result of various compensatory effects, mainly between (i) the internal and solvation electrostatics, (ii) the direct van der Waals interactions between the protein and the ligand and loss in van der Waals interactions with the solvent due to desolvation upon complexation and (iii) the cavitation effects and the entropic losses of the protein and the ligand molecules upon binding. Fig 3 illustrates the primary contributions to free energy of binding presented as compounded subsets. The electrostatics is a sum of terms 3, 7, 11 and 15 from equations (1-6). Similarly, the van der Waals represents a combination of terms 4, 8, 12 and 16. The hydrophobic part includes terms 5, 9 and 17. The entropic effects are a sum of components 13 and 14. The adaptation contribution is due to component 1 in equation (1). A final state analysis (considering the energies of the complexed state only) reveals that both UMP and CMP exhibit comparable binding energetics to bovine angiogenin. While electrostatics and entropy effects are similar, CMP shows more favourable van der Waals, indicating marginally better steric complementarity, and UMP shows more favorable hydrophobic effects. The last pair of histograms in Fig. 3 dramatically illustrates the differences between UMP and CMP when both initial and final states are considered. This adds an extra term i.e. ΔG^{0}_{I}, to the final state analysis arising due to structural adaptation. As mentioned before, ΔG^{0}_{II} is set to zero. On inclusion of the adaptation energy of the protein, UMP and CMP yield binding energies of +2.0 and –23.7 kcal/mol respectively. This difference is due to the fact that the structural adaptation in UMP is estimated to be about +15.1 kcal/mol while in the case of CMP it is –11.6. Further analysis of the MD structures reveals that the protein assumes a more compact conformation in the presence of CMP with smaller surface area relative to the native form whereas the presence of UMP stretches the protein. The differences in accessible surface areas between the bound form of the protein and the native protein

Table 1—Calculated contributions to the standard free energy of binding (in kcal/mol) for bovine angiogenin – UMP/CMP complexes

Term	Component	UMP	CMP
	Step I: Structural adaptation of protein		
$\Delta G_1^{adpt,P}$	Free energy change for the process P → P*	15.1	–11.6
	Step II: Structural adaptation of ligand		
$\Delta G_2^{adpt,I}$	Free energy change for the process I → I*	—	—
	Step III: Desolvation of protein		
$\Delta G_3^{el,P*}$	Electrostatic component of P* desolvation	2862.7	2780.3
$\Delta G_4^{vdW,P}$	VdW component of P* desolvation	332.4	313.2
$\Delta G_5^{CAV,P}$	Cavity component of P* desolvation	–392.6	–369.8
$\Delta G_6^{DH,P}$	Loss of added salt interactions	—	—
	Step IV: Desolvation of ligand		
$\Delta G_7^{el,I}$	Electrostatic component of I* desolvation	78.8	88.1
$\Delta G_8^{vdW,I}$	VdW component of I* desolvation	19.3	19.4
$\Delta G_9^{CAV,I}$	Cavity component of I* desolvation	–22.8	–22.9
$\Delta G_{10}^{DH,I}$	Loss of added salt interactions	—	—
	Step V: Complex formation in vacuo		
$\Delta H_{11}^{el,C}$	Electrostatic interactions between P* & I*	–337.5	–343.4
$\Delta H_{12}^{vdW,C}$	VdW interactions between P* & I*	–24.9	–24.5
$-T\Delta S_{13}^{tr\&rot}$	Rotational, translational entropy change	24.3	24.3
$-T\Delta S_{14}^{vib\&conf}$	Vibrational, configurational entropy change	5.3	5.5
	Step VI: Solvation of complex		
$\Delta G_{15}^{el,C}$	Electrostatic component of complex solvation	–2617.5	–2539.0
$\Delta G_{16}^{vdW,C}$	VdW component of complex solvation	–328.4	–313.0
$\Delta G_{17}^{CAV,C}$	Cavity component of complex solvation	387.8	369.7
ΔG_{18}^{DH}	Added salt interactions with complex	—	—
$\Delta G^0_{NET\ (FINAL)}$	Net free energy of binding	–13.1	–12.1
$\Delta G^0_{NET\ (FINAL-INITIAL)}$		2.0	–23.7

averaged over the MD trajectories are 173.6 Å^2 for UMP and –311.1 Å^2 for CMP. The overall binding of course is accompanied by a loss in surface area as seen from the histograms for the hydrophobic component. In a nut-shell, the protein adapts itself structurally, dynamically and thermodynamically with marked ease to bind to CMP.

At the biological level, the free energy component analysis carried out here clearly supports the experimental observation of angiogenin preferring cytosine over uracil at the first base binding site. Although the residue E118 that obstructs the first base binding site in the native enzyme interacts favourably with N4 of cytosine, it has not emerged as a major reason for substrate specificity (through electrostatic component) from our analysis. This is also in agreement with the experimental observation that mutation of E118 does not significantly alter the substrate specificity[17]. The marked structural adaptability of angiogenin to CMP emerging from our analysis indicates a cumulative effect for substrate specificity. Thus, it is gratifying to note that MD and rigorous free energy component analysis combination is able to explain the specificity of enzyme-ligand interaction. The results are also suggestive of an inverse relationship of RMS fluctuations of the protein during MD simulation with the structural adaptability of the protein. These results generate confidence in pursuing such an approach for the quantitative evaluation of free energy components in protein-ligand interactions.

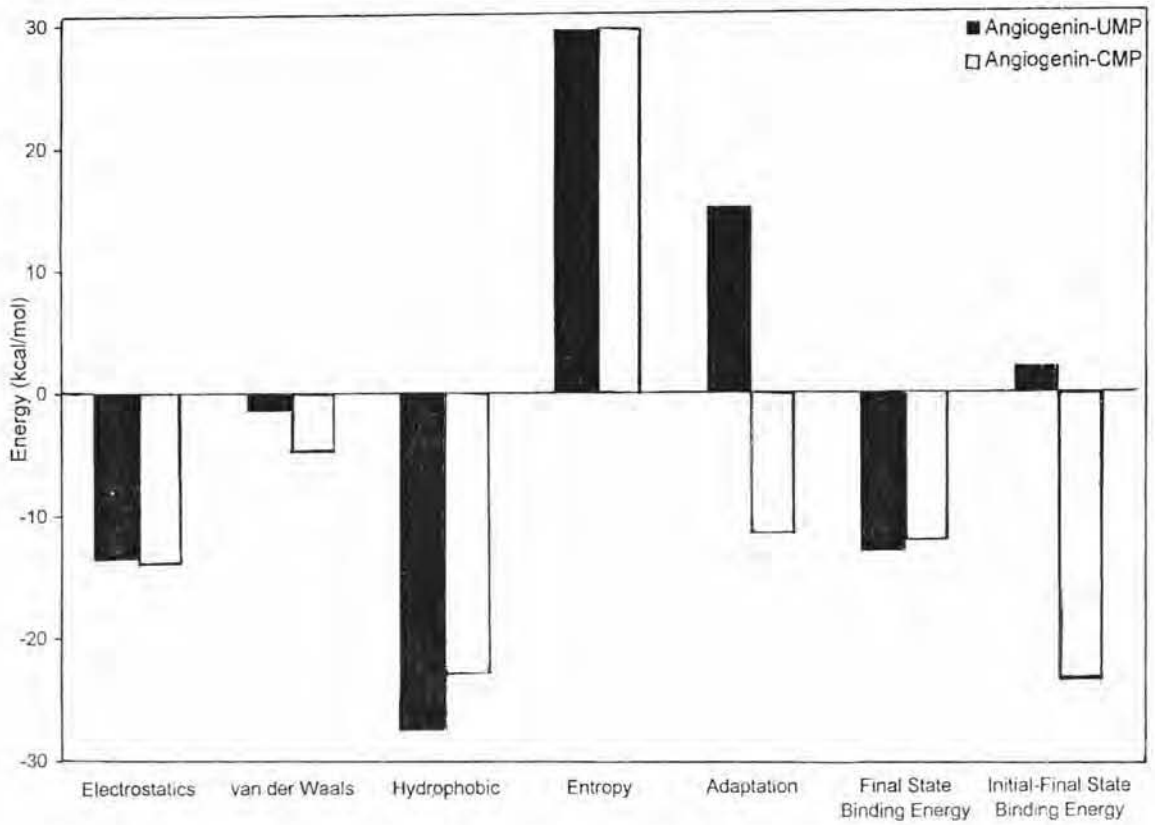

Fig. 3—A histogram-view of the calculated contributions to the binding free energy of the bovine angiogenin-UMP/CMP complexes. [The binding energy without the adaptation term represents a final state analysis while the binding energy with adaptation considers both the initial and the final states of the protein]

Concerning the relative thermodynamic stability of the native protein vis-à-vis that in the complex, particularly in the case of CMP, a question arises as to whether the adaptation expense can ever be negative. Does this suggest that the presence of CMP lets the protein explore thermodynamically more stable substates? At this stage, we have not been able to resolve the physical implication of this result. Results with UMP are of course intuitive. Even if we assumed that errors in the estimates of adaptation expense are of the order of 12 kcal, implying essentially zero adaptation penalty for CMP and a positive value for the complex with UMP, the conclusions of this study on the net stabilization of CMP complex relative to that of UMP and the origins of this preference remain valid.

Conclusions

It was difficult to rationalize the reduced activity of angiogenin on uracil base at the B1 site compared to cytosine, especially since there was only a small difference in the structure of the two bases and all the known protein-ligand interactions were very much similar in both the systems. We have carried out a free energy component analysis on the MD snapshots of native, UMP and CMP bound complexes of bovine angiogenin. Our present analysis predicts that CMP is a better ligand than UMP, a result that is consistent with experiments. It is interesting to note that many of the terms such as electrostatics, van der Waals interactions, hydrophobic interactions and entropies are very much similar for the binding of CMP and UMP to angiogenin. A major difference comes in the adaptation term. In a physical sense, it is the energy required for the enzyme to adapt itself to interact with/accommodate the ligand. This better adaptation of protein to CMP is also reflected in a compact protein conformation and reduced RMS fluctuations in the presence of CMP. Thus, a quantitative reasoning for the preference of cytosine over uracil in the first base binding region [B1 site] of angiogenin has emerged from the analysis of free energy components and it is suggested that such an approach can be pursued in understanding the protein-ligand interactions in general.

Acknowledgement

Financial support from the Council of Scientific and Industrial Research (CSIR) to BJ and that from Department of Science and Technology (DST) to SV are gratefully acknowledged. The Supercomputer Education and Research Centre of IISc is acknowledged for computational facilities.

References

1 Vallee B L, Riordan J F, Higachi N, Feit J W, Crossley G, Bukhler R, Budzik G, Bredham K, Bethune J L, Alderman E M (1985) *Experimentia* 41(1), pp. 1-15
2 Riordan J F (1997) in *Ribonucleases: Structures and Functions* (G D'Alessio & James F Riordan, eds), pp. 445-489, Academic Press, New York
3 Strydom D J, (1998) *Cell Mol Life Science* 54, 811-24
4 Acharya K R, Shapiro R, Riordan J F & Vallee B L (1995) *Proc Nat Acad Sci USA* 92, 2949-2953
5 Acharya K R, Shapiro R, Allen S C, Riordan J F & Vallee B L (1994) *Proc Nat Acad Sci USA* 91, 2915-2919
6 Lequin O, Albaret C, Bontems F, Spik G & Jean-Yves-Lallemand (1996) *Biochemistry* 35, 8870-8880
7 Lequin O, Thuring H, Robin M & Jean-Yves-Lallemand (1997) *Eur J Biochem* 250, 712-726
8 Leonidas D D, Shapiro R, Allen S C, Subbarao G V, Veluraja K & Acharya K R (1999) *J Mol Biol* 285, 1209-1233
9 Madhusudhan M S & Vishveshwara S (1999) *Biopolymers* 49, 131-144
10 Madhusudhan M S & Vishveshwara S (1998) *J Bio Str Dyn* 16(3), 715-722
11 Madhusudhan M S & Vishveshwara S (2000) *Current Science* 78, 852-857
12 Jayaram B, Sprous D & Beveridge D L (1998) *J Phys Chem B* 102, 9571-9576
13 Jayaram B, Sprous D, Young M A & Beveridge D L (1998) *J Am Chem Soc* 120, 10629-10633
14 Jayaram B, Mc Connell K J, Dixit S B & Beveridge D L (1999) *J Comp Phys* 151, 333-357
15 Kalra P, Das A, Dixit S B & Jayaram B (2000) *Indian J Chem* 39A, 262-273
16 Kalra P, Das A & Jayaram B (2001) *J Appl Biochem Biotech, (in press)*
17 Ru sso N, Shapiro R, Acharya K R, Riordan J F & Vallee BL (1994) *Proc Nat Acad Sci USA* 91, 2920-2924
18 Humphery W, Dalke A & Schulten K (1996) *J Mol Graphics* 14, 33-38