J. Indian Inst. Sci., July–August 1997, 77, 327–337.

Protein Dynamics: Molecular simulations of RNase A and related proteins

G. NADIG, R.VARADARAJAN# AND S.VISHVESHWARA*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.
E-mail: # varadar@mbu.iisc.ernet.in * sv@mbu.iisc.ernet.in

1. Introduction

Proteins have unique structures. At equilibrium, the atoms in these molecules fluctuate around a mean structure and the extent of fluctuation is determined by the force acting on each atom of the molecule. The motion of atoms about the mean position can elucidate a number of features of proteins such as: which of the noncovalent interactions are rigid, which ones are flexible, what kind of conformational change the protein undergoes when it binds to its ligands, how does protein dynamics change when mutations occur at specific positions in the protein and so on. Elucidation of such properties ultimately gives an insight into the basic problems of protein folding and the mechanism of action of proteins. In our laboratory, we have carried out molecular dynamics simulations on the protein Ribonuclease A (RNase A), its complexes, fragments and mutants in order to address problems such as protein–ligand interactions[1,2] and protein dynamics in fragment complemented systems[3] using the computational facilities at the SERC, IISc. The results of some of these investigations are presented below.

2. Studies on RNase S – A fragment-complemented system

Fragment complementation systems are systems with two or more fragments of a protein that can be reconstituted to give a complex with structure and activity similar to that of the native uncleaved protein. Ribonuclease S is one of the most well studied and oldest fragment complementation systems[4]. RNase S is a noncovalent complex of two fragments of RNase A, the smaller 20-residue fragment called the S peptide and the larger 104-residue fragment called the S protein. RNase S has been used as a model system to study the thermodynamics of protein folding[5]. RNase A and RNase S have quite different heat capacities of folding despite having identical folded structures. Differences in the dynamic properties of structurally similar molecules have been suggested as a possible explanation for these differences in the heat capacity. Several independent experiments indicate that RNase S shows more fluctuations about its time-averaged structure than RNase A. There are two different probes that have been used for monitoring the dynamics of RNase A and RNase S, namely, hydrogen exchange and trypsin digestion. In a hydrogen exchange experiment the rates at which the amide protons exchange with protons of the solvent (for example, tritiated water) are measured, while in the case of tryptic cleavage the rates at which two proteins are digested by the enzyme trypsin are measured. It has been shown

a) b)

FIG. 1. Superposition of the Cα traces of the structures extracted at intervals of 5 picoseconds from the <M.D>structure of A) RNase A and B) RNase S.

that a major contributing factor to the above-mentioned rates is the dynamics displayed by the two proteins. It has been shown that RNase S is readily cleaved by trypsin while RNase A is indefinitely stable. The tritium and deuterium exchange studies of RNase S indicate that the rates of exchange of amide protons from RNase S are much higher than those from RNase A [Richards and Wycoff, 1971]. A detailed comparison of the various crystal structures of RNase A and RNase S to investigate the possible role of structural differences in the experimentally observed differences in dynamics, and the susceptibility to acid- and temperature-induced denaturation between the two proteins did not suggest any reasons for the above-mentioned differences. We, therefore, have attempted to characterize the dynamics of RNase A and RNase S at 300 K and 330 K by carrying out 100 ps of molecular dynamics simulations (MD) of the protein immersed in a box of roughly 2300 water molecules. The greatest advantage of using MD simulations is the availability of information at atomic resolution as a function of time.

2.1. *Results and discussion*

Although many biologically relevant motions in proteins occur on time scales of milliseconds to seconds, MD simulations of protein molecules in water are feasible for time scales of a few hundred picoseconds. The simulation protocol adopted by us is outlined in the methods section. The primary objective of our simulations was to characterize fluctuations about the time-averaged structures of RNase S and RNase A. These fluctuations are studied by computing the root mean square deviations of the coordinates of all the atoms over the simulation period. Shown in Figures 1a and 1b are superpositions of the Cα traces of structures extracted at intervals of 5 ps from the RNase A and RNase S simulations on their molecular dynamics average structure, respectively. The RMSD of the backbone and side-chain of RNase A and RNase S are comparable, with RNase S surprisingly having a marginally smaller value (Figures 2 & 3). The RMSD values indicate that the loop regions of the two proteins show the largest fluctuations, while the regions forming

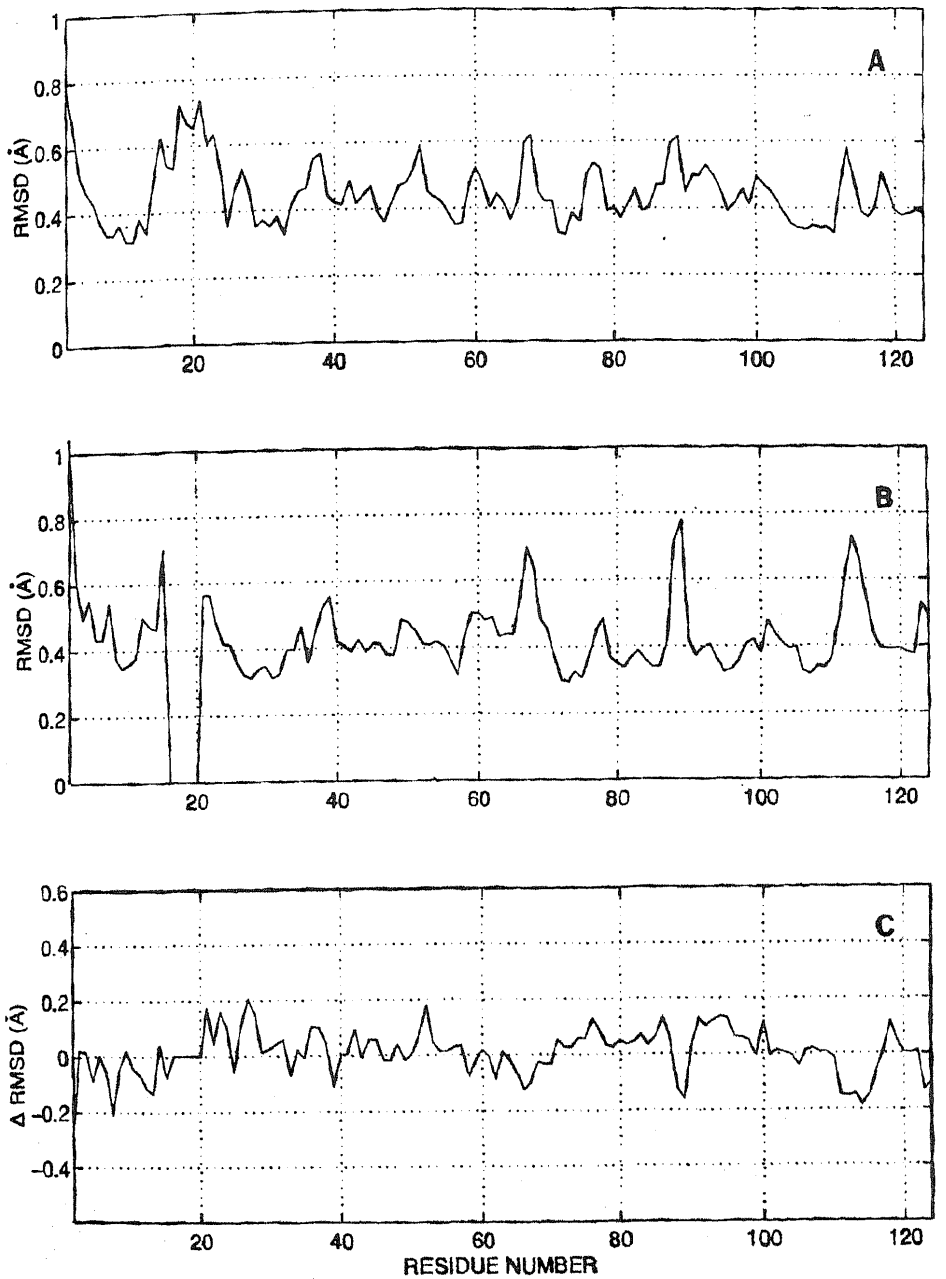

FIG. 2. Average RMSD from the <MD> structures of the main chain :A) RNase A , B) RNase S, C) difference in RMSD between RNase A and RNase S on superposing the residues 21-124 only. The residues 16-20 are absent in RNase S.

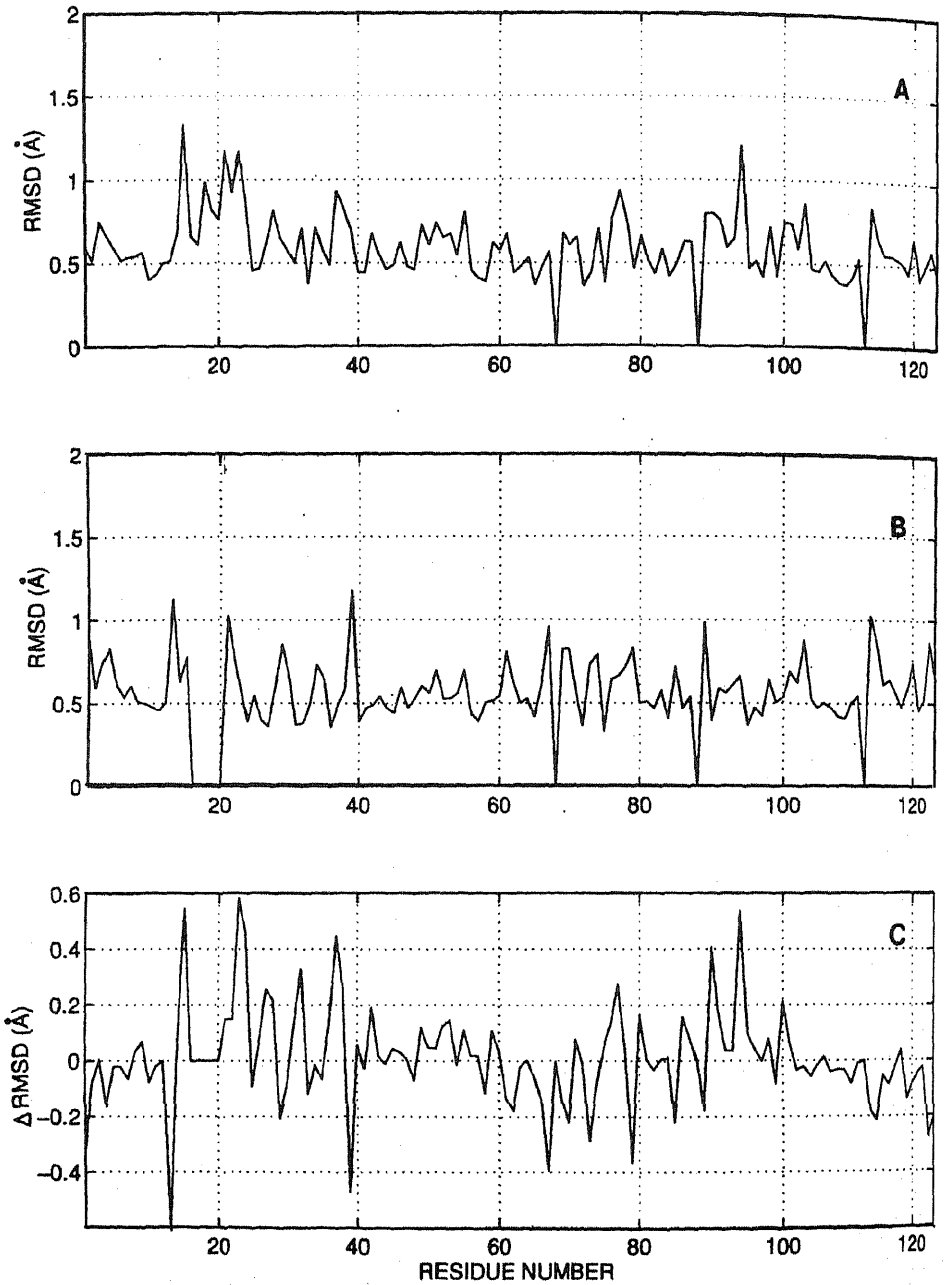

FIG. 3. Average RMSD from the <MD> structures of the side chain :A) RNase A, B) RNase S, C) difference in RMSD between RNase A and RNase S on superposing the residues 21-124 only. The residues 16-20 are absent in RNase S.

secondary structure have much lesser fluctuations. The S peptide region of RNase S shows larger fluctuations in RNase S than in RNase A: This could easily be explained by the fact that the S peptide is bound to S protein solely through noncovalent interactions, while in RNase A there is, in addition, a covalent linkage which could constrain the motion of the S peptide. The increased mobility of the S peptide linkage does not affect the motion of residues of the S protein which have noncovalent interactions with it.

2.2. *Correlation between visits of water and hydrogen exchange rates*

Tritium and deuterium exchange experiments indicate faster amide exchange rates in RNase S than in RNase A [Richards & Wyckoff, 1978]. We have calculated the number of visits that water molecules make to the hydration shells of each atom in the protein in RNase S and RNase A during the simulations and attempted to correlate this with experimentally observed exchange rates. The number of visits in the case of RNase A and RNase S are similar. This number depends on the residue type as well as on the shape of the protein surface. The number of visits is lowest for nonpolar residues and largest for polar residues, which is consistent with the fact that water is more ordered around hydrophobic residues than around the polar or charged residues. The exchange rates of amide protons of RNase A have been studied by NMR spectroscopy[7]. All the main-chain amide protons of RNase A have been divided into three classes, protons that exchange with a rate constant lesser than 1.5×10^{-4} min^{-1}, protons that exchange with a rate constant greater than 1.5×10^{-2} min^{-1}, protons that have an exchange rate constant that lies between these two limits. We find that there is a definite correlation between the number of visits to a given residue observed in the simulation and the exchange class of that residue determined by NMR. The set of residues that have 5 visits have only protons that exchange the slowest; as we move to the set of residues that have 35 visits, the fraction of the residues that have the highest exchange rates increases (figure 4). The time scales of these simulations are very different from those of the experiment, yet there is a very good correlation between the results of MD simulations and experimental exchange studies for RNase A. This implies that genuine differences, if any, between the dynamics of RNase A and RNase S should have manifested in these simulations. It is well known that the potential energy surface of a protein in the native state consists of a large number of local minima[8]. MD simulations on the time scale of 100 ps sample only a fraction of the conformational space available to the protein. The obvious way to enhance the conformational sampling is to carry out simulations for a very long period. Currently, the largest simulations of proteins of this size have been on a time scale of the order of 1 nanosecond. We have, instead of choosing to carry out such long simulations, performed an additional 100 ps of simulation of the two proteins at a higher temperature, 330 K. High-temperature simulations have been used previously to study the unfolding of globular proteins. The fluctuations of the atoms about their respective [MD] structure show very similar differences. Thus, the simulations at 300 K and 330 K indicate similar dynamic behaviour about their time-averaged structures. An exhaustive analysis of the simulation data, which includes the study of fluctuations in the solvent accessible surface area, the depth of the protein atoms from the surface and the fluctuations in the dihedral angles of the proteins revealed only minor differences in the dynamics of the two proteins. These differences were confined to

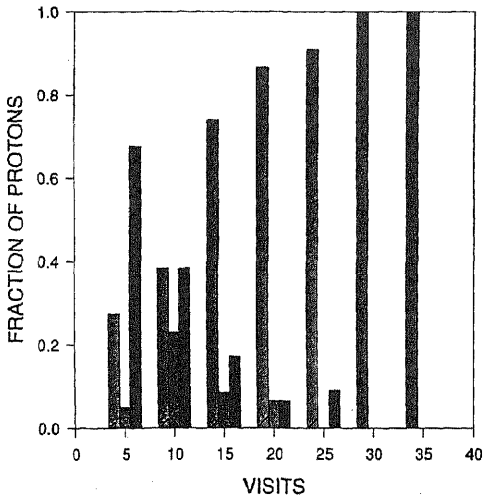

FIG. 4. Correlation between the number of visits to a given residue and its amide proton exchange class. The residues have been classified into 7 types according to the number of visits ie; from 5 visits to 35 visits. For each class of residues the fraction of protons that are in the three exchange classes are drawn on the Y axis as a histogram (exchange rates less than 1.5×10^{-4} min^{-1} ( ), exchange rates greater than 1.5×10^{-2} min^{-1} ( ) , exchange rates between 1.5×10^{-4} min^{-1} and 1.5×10^{-2} min^{-1} ( )).

the loop regions and the regions around the termini. Although, there were a few residues which exhibited different behaviour in the dynamics, it was not clear as to whether this was because of the differences in the starting model for the two simulations. These simulations strongly suggest that both RNase A and RNase S in solution exhibit similar fluctuations about their time-averaged structures. This is in apparent contradiction to previous proteolysis and hydrogen exchange data which led to the re-examination of experimental results.

2.3. *Experimental studies*

Since it was observed from the MD simulations that RNase A and RNase S exhibited similar dynamic behaviour, we re-examined the experimental data that had suggested that RNase S should exhibit greater dynamic flexibility than RNase S. A fundamental difference between RNase A and RNase S is that RNase S is a bimolecular system while RNase A is a unimolecular system. Thus, the fraction of the folded state should be independent of the total protein concentration in the case of RNase A, while in the case of RNase S the fraction of the folded state is proportional to the total protein concentration. Since the dissociation constant of the S peptide from the S protein was very small, it was assumed that the small amount of S peptide present in most experiments could be neglected. A thorough analysis revealed that this is not true. It was shown using a simple kinetic model that the observed apparent differences could be accounted for by the presence of the small

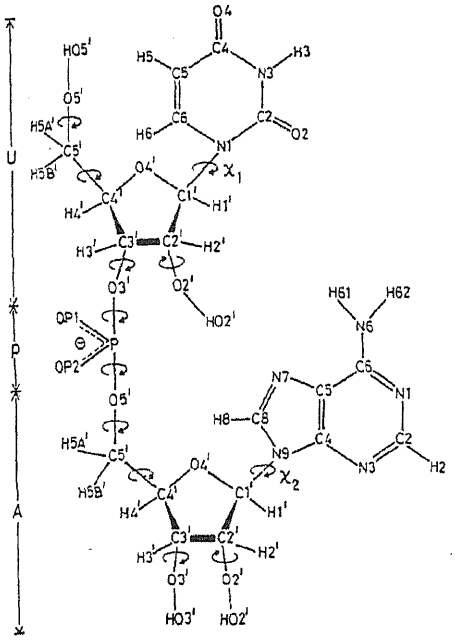

FIG. 5. Schematic digram of the chemical structure of the ligand, UpA.

FIG. 6. The structure of the RNase A-UpA complex that was used as the starting model for the simulations. The ligand, UpA is drawn in ball and stick type. This picture has been obtained using MOLSCRIPT (Kraulis, 1991).

amounts of free S peptide and S protein, which are in equilibrium in solution[3]. Thus, MD simulations along with the analysis of experimental data demonstrate that RNase A and RNase S have similar dynamic behaviour. These results have provided us with a satisfactory answer to the long-standing question about the origins of the differences in RNase A and RNase S dynamics.

2.4. *Conclusions*

The MD simulations and analysis of a number of parameters of the protein RNase A and its fragment complementation system RNase S showed that the dynamical behaviour around its equilibrium structure is very much similar, contrary to the belief that the dynamics of the two systems are strikingly different. This result prompted us to re-examine the experimental data and explain satisfactorily the apparent differences as being due to RNase S being a bimolecular system.

3. Studies of RNase A – ligand interactions

Understanding the mechanism of action and substrate specificity of enzymes at a molecular level is an important area of research in the field of enzyme catalysis. RNase A has been an enzyme of intense investigation in the past two decades. A wealth of information is available from X ray crystallographic studies on enzyme -ligand complexes. However, computational studies further contribute to our understanding for the reasons such as 1)

the stable enzyme -ligand complexes which are amenable for experimental study are generally inhibitors and analogs of substrate or transition state. The complex with real substrate can be studied by computer simulations. 2) Quite often, the ligand binding sites are poorly characterized at the molecular level by experiments and computer modelling can give lucid details and 3) time averaged static pictures are obtained from experiments, where as the dynamics which is an essential feature of enzyme activity is best understood by molecular dynamics studies. Towards these goals, systematically docking of ligands such as pyrimidine 3' phosphate and Uridyl 3'5' adenosine (UpA) into the active site of RNase A [Seshadri *et al.*, 1994] were carried out in our laboratory. The docked RNase A-UpA complex is further investigated by molecular dynamics simulations and a comparison with native RNase A is made [Seshadri *et al.*, 1995, Nadig & Vishveshwara, 1997]. A summary of these investigations are presented below.

3.1. *Results and Discussions*

The structure of the substrate, UpA is shown in figure 5. The rotations around the various bonds (defined by the dihedral angles) necessary to describe the three dimensional conformation of the molecule are also indicated in the figure. Various UpA conformations were generated systematically by varying the dihedral angles. The conformations of UpA thus generated were docked into the active site of the enzyme RNase A and tested for contact free conformations. The contact free conformations of such complexes were further subjected to energy minimization. The minimized structures satisfying the experimental observations were considered for detailed analysis. The computer generated RNase A-UpA complex shown in figure 6 was selected as the starting point for the molecular dynamics studies. The molecular dynamics simulations were conducted on RNase A and RNase A-UpA complex in aqueous medium as described in the methods section. The dynamics of enzyme-ligand interactions were monitored in detail by following the trajectories of parameters such as those related to UpA conformations, protein conformation and the protein -ligand hydrogen bonds. The analysis of the trajectories indicate that the active site residues of RNase A and the substrate, UpA form a network of hydrogen bonds. The nucleotide bases, Uridine and Adenine are held in their respective positions by stable hydrogen bonds with the protein residues T45, S123, Q67, N71 and E111. The hydrogen bond between the 2' hydroxyl group and the H12 Nε2 seems to be consistent with the fact that the role of H12 is to abstract a proton from the O2' atom of the substrate in the reveresible trans-phosphorylation step in the catalysis. The phosphate oxygen, OP1, makes a very stable hydrogen bond to the amide nitrogen of F120. The other oxygen atom, OP2, makes hydrogen bonds to the side chains of the catalytically important residues H119 and to K7.

A comparison of the properties between free RNase A and RNase A-UpA complex brings out the differences in the enzyme behaviour upon ligand binding. Such differences are monitored by comparing the root mean square deviation (RMSD), dihedral angles and the principal component analysis. The difference RMSD is presented in figure 7. On the average, the atomic positional fluctuations about the <M.D> structure in both the simulations are similar. Yet, various regions of the proteins display different dynamic behaviour. These differences are due to the presence of the susbstrate, UpA in the RNase A-UpA

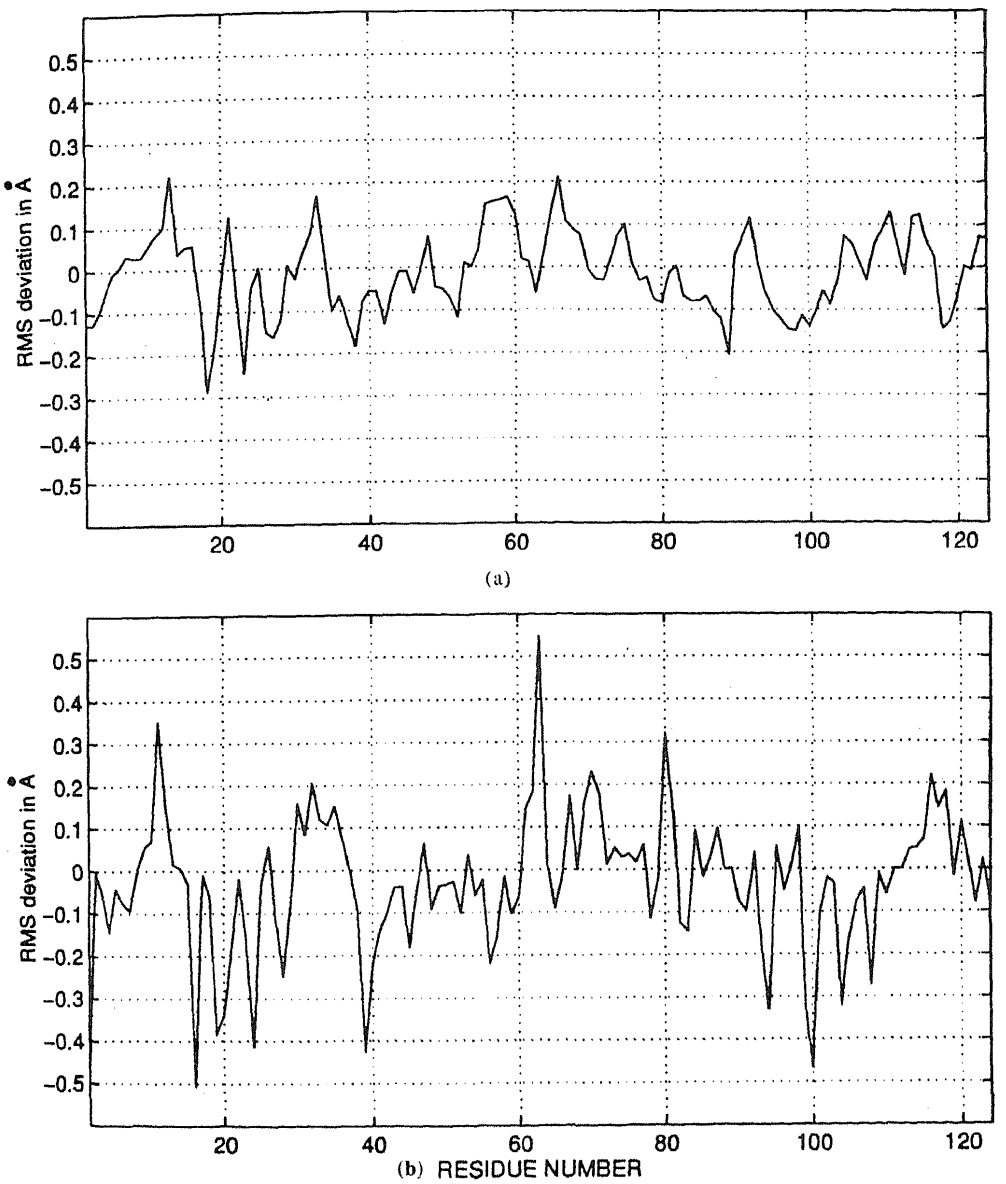

FIG. 7. The differences in RMSD about the <MD> structures of the free RNase A and RNase A-UpA complex of a) main chain and b) the side chains.

complex. The RMSD clearly indicate a reduction in the magnitude of the atomic positional fluctuations of the regions that form the three helices (residue 3-13, 24-34 and 50-60) in the RNase A-UpA complex. The active site residues do not show any discernable differences between the two simulations.

3.2. *Conclusions*

The enzyme substrate interaction was modelled on a computer by docking UpA into the active site of RNase A. Molecular dynamics studies on the RNase A-UpA complex elucidates the catalytically important interactions between RNase A and UpA. Further, a comparison of the dynamics of native and UpA bound RNase A elucidates the conformational changes in the protein upon UpA binding.

3.3. *Methods*

A pre-requisite for performing any Molecular dynamics simulation is the availability of a force field that accurately simulates the energetics of the biomolecules. The force field that we have used is the parm91.dat of AMBER v4.0[Pearlman et.al., 1991] and has been emperically derived from a huge body of spectroscopic, crystallographic and thermodynamic data of model peptides. The emperical potential energy function used is given by

$$E = \Sigma k_{bi}(b_i-b_0)^2 + \Sigma k_{\theta i}(\theta_i-\theta_0)^2 + \Sigma k_{\varphi i}[1-\cos n(\varphi_i-\varphi_0)] + \Sigma(A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6} + q_iq_j/r_{ij})$$

The first two terms in the above equation allow for the harmonic stretching of bond lengths and angles around the equilibrium positions, the third term describes the changes in energy due to rotations of the torsion angles and the last term represents the Lennard-Jones potential energy between any two non bonded atomic pair. The simulations of RNase A, RNase S and the RNase A -UpA complex were carried out using the suite of computer programmes, AMBER v4.0 [Pearlman et.al, 1991]. An identical simulation protocol was followed in carrying out these simulations. The protein molecules were placed at the center of a box of pre-equilibrated TIP3P monte carlo water molecules. Those water molecules making bad van der Waals contact with many protein atoms and also the those water molecules which are farther than 5.0 Å along the X, Y and Z axes from any protein were removed from the starting model. Thus our starting model for the solvated protein consisted of about 8300 atoms. The simulations are usually started by minimizing the potential energy of the system to a gradient of 0.1 kcal/Mol/ Å using a combination of steepest descent and conjugate gradient minimization procedures. The sole purpose of this procedure is to releive bad non-bonded contacts in the starting model. The M.D simulation runs are started by assigning velocities drawn from a maxwellian distribution to all atoms in the system in a random fashion. The Newton's equations of motion are then integrated using the verlet algorithm with an integration time step of 1 femtosecond. The system is brought upto the desired temperature of 300 K in three stages by coupling it to a heat bath, a procedure called equilibration. In the first stage, the temperature is raised to 200 K and the simulations are carried out for a further 2 ps, the last stage comprises of a 6 ps simulation at 300 K. In the case of the RNase A -UpA complex we equilibrated the structure for an additional period of 15 ps. Now the coupling to the heat bath is removed and a production run of 90 ps duration is then carried out. The co-ordinates obtained from the simulation are stored at intervals of 0.1 ps. During the entire simulation period, all the bond lengths were held at their equilibrium values using the SHAKE algorithm [Ryckert *et.al.*, 1977] . The non-bonded interactions were computed using a residue based list with a cut-off of 12 Å . The list was updated every 10 time steps. A dielectric constant of $\varepsilon = 1$ was used. A typical 5 ps simulation requires 12 hours of CPU time on the IBM Rs 6000 580

machines. All computations were carried out at the Supercomputer Education and Research Center, Indian Institute of Science.

References

1. ELBER, R. AND KARPLUS, M., Science, 1987, **235**, 318–321.
2. KIM, E. E., VARADARAJAN, R, WYCKOFF, H. W. AND RICHARDS, F. M. *Biochemistry*, 1992, **31**, 12304–1231.
3. KRAULIS, P. J. *J. Appl. Crystallogr.*, 1991, **24**, 946–950.
4. NADIG, G. AND VISHVESHWARA, S. *Biopolymers*, 1997 (in press).
5. NADIG, G., RATNAPARKHI, G. S., VARADARAJAN, R AND VISHVESHWARA, S. *Protein Science*, 1996, **5**, 2104–2114.
6. PEARLMAN, D. A., CASE, D. A., CALDWELL, J. C., SEIBEL, G. L., SINGH, U. C., WEINER, P AND KOLLMAN, P. A **AMBER v4.0**. University of California, San Francisco, 1991.
7. RICHARDS, F. M AND VITHAYATHIL, P. J. *J. Biol. Chem.*, 1959, **234**, 1459–1465.
8. RICHARDS, F. M. AND WYCKOFF, H. A. *The Enzymes*, 1971, **4**, 647.
9. RYCKAERT, J. P., CICCOTTI, G. AND BERENDSEN, H. J. C. *J. Computational Phys.*, 1977, **23**, 327–341
10. SANTORO, J., GONZALEZ, G., BRUIX, M., NEIRA, J. L., NIETO, J. L., HERRANZ, J AND RICO, M. *J. Mol. Biol*, 1993, **229**, 722–734
11. SESHADRI, K., RAO, V. S. R AND VISHVESHWARA, S., *Biophysical J.*, 1995, **69**, 2185–2194.
12. SESHADRI, K., RAO, V. S. R AND VISHVESHWARA, S., *J. Biomol. Struct. Dyn*, 1994, **11**, :395–415
13. VARADARAJAN, R., CONNELLY, P. R., STURTEVANT, J. M AND RICHARDS, F. M., *Biochemistry*, 1992, **31**, 1421–1426