Eur. Phys. J. D **20**, 601–608 (2002)
DOI: 10.1140/epjd/e2002-00151-x

THE EUROPEAN PHYSICAL JOURNAL D

Essential dynamics and sidechain hydrogen bond cluster studies on eosinophil cationic protein

B.S. Sanjeev and S. Vishveshwara[a]

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

Received 16 January 2002
Published online 13 September 2002 –

Abstract. Eosinophil Cationic Protein (ECP) is a member of RNase A superfamily which carries out the obligatory catalytic role of cleaving RNA. It is involved in a variety of biological functions. Molecular dynamics simulations followed by essential dynamics analysis on this protein are carried out with the goal of gaining insights into the dynamical properties at atomic level. The top essential modes contribute to subspaces and to the transition phase. Further, the sidechain-sidechain/sidechain-mainchain hydrogen bond clusters are analyzed in the top modes, and compared with those of crystal structure. The role of residues identified by these methods is discussed in the context of concerted motion, structure and stability of the protein.

PACS. 82.37.Rs Single molecule manipulation of proteins and other biological molecules

1 Introduction

Biological processes are extremely complex to understand, as reactions in cells do not take place in isolation. Almost any reaction is a dynamical process that is dictated by initial conditions and can be influenced by any of thousands of biomolecules and substances around. However, an insight into biological systems is often obtained by probing biomolecules at atomic and molecular levels and also under *in vitro* conditions. For instance, enzyme functions are often evaluated under laboratory conditions and then compared with *in vivo* results. Consequently a clearer picture related to their biological function and other features emerges.

Several computational techniques have been developed in the last several decades to investigate biopolymers at atomic level so as to understand their structure, dynamics and function using approaches developed in Physics. Computational enzymology [1] is an interesting area of research, which uses quantum mechanical and molecular mechanics techniques. For instance, we have shown earlier how proton transfer during the catalysis of an enzyme (RNase A) can be tremendously influenced by net charge on particular aminoacid residues [2]. Molecular dynamics (MD) simulations are also being extensively used to address a variety of problems. We had carried out simulations on some of the proteins belonging to the family of RNase A [3–11] in order to understand the rigid and flexible parts of the structure and other characteristics of this enzyme in aqueous environment. Molecular modeling is often carried out to model structures of biopolymers such as proteins, as well as to obtain the details of interactions between non-covalently bound species. Using such methods enzyme-substrate interactions, which are difficult to capture experimentally, could be elucidated [3,4]. Monte Carlo simulations are also used to address problems related to molecular biology [12,13].

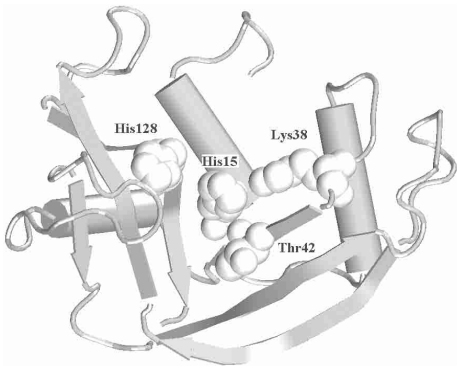

Fig. 1. Crystal structure of human Eosionphil Cationic Protein (PDB code 1DYT, chain A). Active site residues are also shown. This figure (also Figs. 5 and 8) was generated using VMD [14].

In the present article we have investigated the dynamical behavior of a human protein called Eosinophil Cationic Protein (ECP) (Fig. 1 – drawn using VMD [14]). This protein of 133 amino acid residues is found mainly

[a] e-mail: sv@mbu.iisc.ernet.in

in eosinophil granules. It is an enzyme that is believed to have a role in neurological disorders [15,16], exhibits cytotoxicity [17], and is implicated in a few skin diseases [18]. Helminthotoxicity [19], anti-bacterical activity [20], antiviral activity [21] and neurotoxicity [15,16] are some of its other characteristics. Another interesting feature of ECP is that it is the fastest evolving mammalian protein [22].

Biochemically, ECP exhibits ribonuclease activity, which is necessary for its neurotoxicity [15,16]. Structurally, the protein belongs to the family of RNase A, a system on which we had carried out extensive simulations and modeling studies. A remarkable feature of RNase A family of proteins is the diversity of biological functions they carry out despite being structurally similar and carrying an obligatory function of cleaving RNA [23]. We believe that extensive analysis on the structure and dynamics of the RNase A family of proteins will contribute to our understanding of the functioning of these proteins. The present study deals with the simulation of ECP in water, followed by capturing the important modes of dynamics using essential dynamics technique. Further, we have also investigated the clustering pattern of sidechain–sidechain (sc–sc) and sidechain–mainchain (sc–mc) hydrogen bonds using the concepts of graph theory. Although the mainchain–mainchain hydrogen bonds, which form the secondary structures in proteins, are very well characterized over the past several decades, we reckon that a systematic method for sidechain hydrogen bond analysis has been developed for the first time in this article.

2 Methods

The methodology adopted in this work is divided into three parts. In the first part we have performed a one nanosecond simulation on ECP in water at 300 K to obtain the equilibrium properties of the protein. In the second part the trajectory obtained was subjected to essential dynamics analysis to generate important modes in its dynamics. Finally, the analysis of hydrogen bond patterns is carried out using the concepts of graph theory.

2.1 MD simulations

The starting structure was the A-chain of ECP [24] (PDB code 1DYT, 1.75 Å resolution). The MD simulation was performed using AMBER6 suite of programs using parm94 AMBER force-field [25], which is of the form:

$$E_{\text{tot}} = \sum K_r(r - r_{\text{eq}})^2 + \sum K_t(t - t_{\text{eq}})^2 + \sum 0.5V_n[1 + \cos(n\phi - \gamma)] + \sum (A_{ij}/R_{ij}^{12} - B_{ij}/R_{ij}^{6} + q_i q_j/eR_{ij}).$$

All force field parameters like K_r, K_t etc. are listed by Cornell *et al.* [25]. The protein was solvated in TIP3P water [26] such that the farthest atom along each axis had at least 7 Å of water along the same direction. This was necessary to have at least two shells of water for surface atoms. The resulting system had 18 479 atoms comprising 5 437 water molecules. The system was subjected to 5 000 steps of minimization prior to the simulation. During simulation the bond lengths involving hydrogen atoms were constrained. Non-bonded cut-off of 10 Å was used. Periodic boundary conditions were applied. Translational and rotational motions were removed after every 10 ps. A time step of 2 fs was used for integration. Temperature coupling used was 0.5 ps and pressure coupling was 0.5 ps^{-1}. Constant pressure ensemble was used with isotropic position scaling. Coordinates (trajectory) were saved after every 1 ps. Temperature was raised from 0–100 K in the first 10 ps, 100–200 K in the next 10 ps and 200–300 K during 20–30 ps. From 30 ps onwards, the temperature was maintained at 300 K. A total of 1 400 ps simulation was performed consisting of 400 ps of initial equilibration and 1 000 ps of production run. Particle Mesh Ewald (PME) method [27] was used for calculation of electrostatics.

2.2 Essential dynamics

Normal mode analysis is a powerful method of extracting important modes contributing to the motion of molecules. Harmonic or quasi-harmonic mode analysis [28–32] had been carried out earlier to analyse the correlated motions of atoms in proteins. The correlations between atomic positional fluctuations in proteins can be derived from the coordinates obtained from MD simulations. A robust method of picking up the "essential" subspace from protein dynamics has been given by Amadei *et al.* [33], and its utility has been demonstrated in identifying the hinge bending motion in thermolysin [34]. Essential dynamics technique has also been used to address other problems such as disulphide bridge insertion in plastocyanin [35], fluctuations and cross-correlation analysis on BPTI and lysozyme [36], extended sampling [37], domain movements [38], electron transfer partners in metallo proteins [39], internal harmonic motions and diffusion between multiple harmonic wells of proteins [40]. We have adopted this method in our present analysis to pickup the essential subspace from the MD simulation of ECP. A brief account of the method developed by Amadei *et al.* [33] as adopted in the present work is given below.

The 1 000 snapshots obtained from the MD trajectory during the production run were superimposed w.r.t. average MD structure using all C_α atoms. Covariance matrix ($\mathbf{M} = [m_{ij}]$) was then constructed from the positional coordinates of all heavy (*i.e.*, non-hydrogen) atoms as follows:

$$m_{ij} = (1/S)\sum_t (x_i(t) - \langle x_i \rangle)(x_j(t) - \langle x_j \rangle)$$

where

S = total number of configurations,
t = time in ps,

x_i = ith coordinate (i = 1, 2, ..., $3N$) (N being number of atoms),
$\langle x_i \rangle$ = average value of x_i over all configurations.

All 1 094 (non-hydrogen) protein atoms were considered for the construction of covariance matrix. MATLAB6 was used for diagonalization of the covariance matrix and for finding eigenvalues and normalized eigenvectors. Eigenvalues were sorted in descending order of their value. Eigenvectors represent the directions of motions of the atom and eigenvalues represent total mean square fluctuation of the system along the corresponding eigenvectors.

Relative positional fluctuation (RPF) was then calculated as:

$$RPF(n) = \frac{\sum_{i=1,n} \lambda(i)}{\sum_{i=1,3N} \lambda(i)} \qquad (1)$$

where $\lambda(i)$ is ith eigenvalue. $RPF(n)$ gives the amount of motion associated in the subspace spanned by the top n modes. We have analyzed the first (top) ten modes generated by the top eigenvalues and their corresponding vector components. Motion along top eigenvectors is significant in the higher modes as discussed by Amadei *et al.* [33]. Our aim is to analyze the essential top modes and investigate the details of atoms or residues that contribute to such modes. Such an analysis can aid not only in the identification of correlated motions in proteins, but also would be able to pick up the motions important for interaction with ligands. When the ligand-bound structures become available, a comparison of significant essential modes of the native and ligand-bound structures can be made.

2.3 Cross-correlation maps

Cross-correlation maps are used to identify the regions that are moving in or out of phase [36] during simulation. The elements of this matrix are obtained from their position vectors ($\mathbf{r}$) as:

$$b_{ij} = \frac{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)(\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle}{\sqrt{(\langle \mathbf{r}_i^2 \rangle - \langle \mathbf{r}_i \rangle^2)(\langle \mathbf{r}_j^2 \rangle - \langle \mathbf{r}_j \rangle^2)}} \,.$$

We have considered only the C_α displacements in this study. The value of b_{ij} can vary from -1 (completely anti-correlated motion) to $+1$ (completely correlated motion). A cut-off of 0.35 was chosen as the minimum magnitude of b_{ij} and the ones thus selected are used for interpretation.

2.4 Clustering of sidechain hydrogen bonds

The secondary structures such as helices, strands, sheets and turns form the building blocks of protein structures. Proteins are classified into four classes (all-α, all-β, α/β and $\alpha + \beta$) based on their secondary structure composition [41,42]. The secondary structures are formed through hydrogen bonds between the nitrogen and the oxygen atoms of the peptide backbone (mainchain) and have been extensively characterized [41,42]. Although hydrogen bonds play an important role in terms of folding and stability of proteins [43–51], sidechain hydrogen bonds do not necessarily lead to any ordered structural classification and hence a systematic investigation has been difficult.

In the present study, we have employed the graph spectral concepts to cluster the sidechain–sidechain (sc–sc) and sidechain–mainchain (sc–mc) hydrogen bonds. The residues which are connected by these hydrogen bonds are represented in the form of a Laplacian matrix (the construction of which is given below). It has been shown that the eigenvector components corresponding to the second lowest eigenvalue of the Laplacian contains the clustering information [52]. This provides a systematic method to analyze the variations in non-secondary hydrogen bond patterns in the structures obtained from selected modes. The clustering procedure is described below.

Hydrogen atoms are added to the heavy atoms using LEaP program in AMBER6 suite. We then used another program in AMBER6 suite, called Carnal, to pick the atoms that form hydrogen bonds. The criteria used were donor-acceptor distance within 3.2 Å and the angle at hydrogen connecting the acceptor and donor atoms should be between 120–180°.

Atomic coordinates representing a given essential mode are required for hydrogen bond cluster analysis. These coordinates are generated in the following way [33]. The displacement along a single eigenvector (S) is obtained by the inner product of the position vector ($\mathbf{r}$) subtracted by the average position vector ($\langle \mathbf{r} \rangle$) with the eigenvector ($\mathbf{e}$)

$$i.e., \quad S = (\mathbf{r} - \langle \mathbf{r} \rangle) \cdot \mathbf{e}.$$

In this way atomic motions can be projected onto a single eigenvector and hence from such a motion reconstruct the atomic motions due to a single eigenvector. Atomic coordinates due to the motion along a single vector at a particular time can be obtained as:

$$\mathbf{r}'(t) = \langle \mathbf{r} \rangle + S(t)\mathbf{e}.$$

After a few steps of minimization clusters can be obtained exactly the same was as for a crystal structure as discussed above.

We used graph theory to pick the sidechain (*i.e.*, both sidechain–sidechain and sidechain–mainchain) hydrogen bond clusters. The cluster analysis we did was similar in spirit to the analysis that was earlier employed to analyze mainchain and sidechain clusters [53,54] arising from non-covalent interactions. Graphs can be constructed by defining edges and nodes. The aminoacid residues which participate in sc–sc or sc–mc hydrogen bonds are considered as a set of nodes and those which are connected by hydrogen bonds according to the above mentioned criteria are said to have edges. An adjacency matrix $\mathbf{A}$ is constructed such that the elements of the matrix $[a_{ij}]$ have

the following values:

$$a_{ij} = 0.75 + (n_h \times 0.25) \qquad n_h \text{being the number of hydrogen-bonds between residues } i \text{ and } j \text{ and } n_h > 0$$

$$= 0.001 \qquad \text{if } n_h = 0.$$

Such a weighting procedure ensures that all the hydrogen bonds between the residues i and j are connected by a weight of one unit and additional hydrogen bonds between the same residues increases the weight by 0.25 units each.

The degree matrix ($\mathbf{D} = [d_{ij}]$) was constructed from adjacency matrix as:

$$d_{ij} = \sum_{j=1,n} a_{ij} \qquad \text{where } n \text{ is the order of the matrix (number of residues)}$$

$$d_{ij} = 0 \qquad \text{if } i \neq j.$$

Laplacian matrix ($\mathbf{L}$) was then obtained as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}.$$

Now we diagonalize matrix $\mathbf{L}$, sort eigenvalues in descending order and inspect the eigenvector components of the second lowest eigenvalue. All residues belonging to a particular cluster have identical value of eigenvector component and each cluster has unique value associated to it [52]. A cluster is defined as a set of residues where each member is connected to every other member of the same cluster, either directly or indirectly through other residues' sidechain hydrogen bonds, and no connections with residues outside the cluster exist.

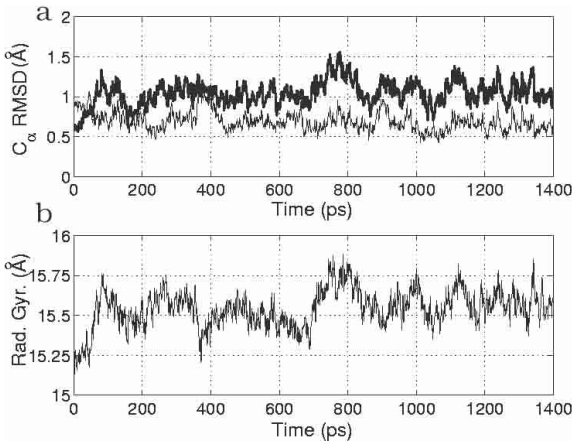

Fig. 2. (a) C_α RMSD w.r.t. minimized crystal structure (thick line) and the average MD structure (thin line) of ECP and, (b) radius of gyration over the trajectory.

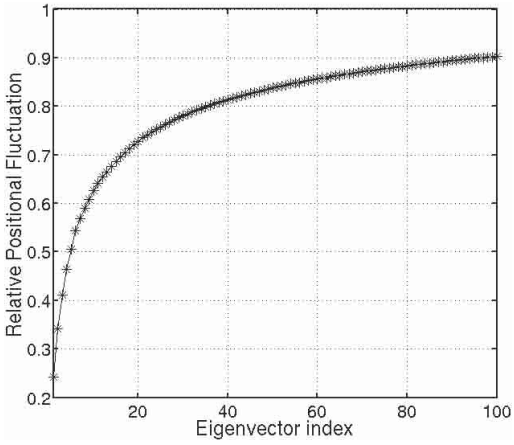

Fig. 3. Relative positional fluctuation as a function of eigenvalue index.

3 Results and discussion

3.1 Molecular dynamics and essential dynamics analysis

Molecular dynamics simulation was carried out on the eosinophil cationic protein in aqueous medium. The first 400 ps was the equilibrium phase followed by 1 000 ps of production run. The root mean square deviations (RMSD) with respect to the minimized crystal structure and the MD average structure are shown in Figure 2a. This confirms that the protein has reached equilibrium by 400 ps. The RMSD fluctuated around 0.7 Å during the production run. The radius of gyration along the trajectory is plotted in Figure 2b. There appears to be a small difference in the average value for the simulation periods before and after 700–800 ps. Interestingly it coincides with a peak in the RMSD w.r.t. minimized crystal structure during the same interval. We will notice the significance of this later in the context of conformational subspaces.

The 1 000 coordinate sets generated from 401 ps to 1 400 ps of simulation was used to extract the essential dynamics features as described in the methods section. The essential dynamics was captured in terms of the eigenvalues and its vector components of the covariance matrix. Relative positional fluctuation (see Eq. (1)) gives the amount of dynamics constituted by the top essential modes. The plot of RPF presented in Figure 3 shows that the top ten modes account for about 62% of the motion in essential space and top three account for more than 40%. The projections along these three modes (Fig. 4a) show a bimodal feature and Figure 4b indicates sampling of two subspaces which are connected by a transition region during the simulation. The subspaces ($\sim$400−650 ps, $\sim$1 200−1 400 ps) span a few hundreds of picoseconds and the time taken for transition is also about a few hundreds of picoseconds. The top two modes dominate the subspaces where as the third mode is mainly involved in the transition. The average structures corresponding to different regions in different modes were generated as explained earlier in the methods section and subjected to 100 steps of minimization to overcome the short contacts. The time intervals 400–650 ps, 1 200–1 400 ps of the first mode and 400–650 ps, 825–955 ps of the second mode were chosen for coordinate generation. The time interval of 725–825 ps of

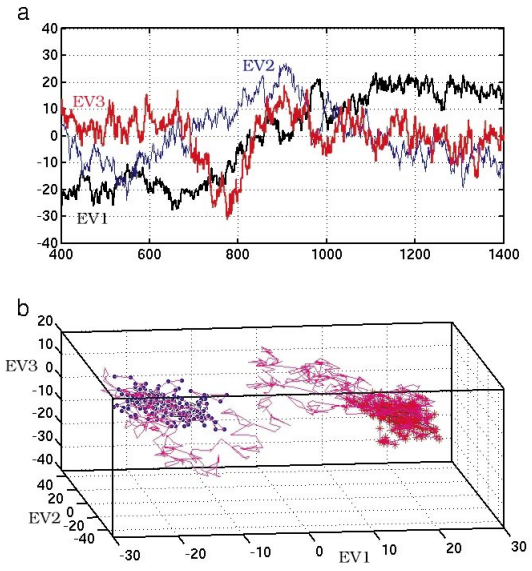

Fig. 4. (a) Projected values (in Å) of top three eigenvectors (EV1, EV2 and EV3) onto MD trajectory. (b) 3-D plot of the three eigenvectors (in Å). The symbols (○) and (*) correspond to A subspace (400–650 ps) and subspace B (1 200–1 400 ps) respectively.

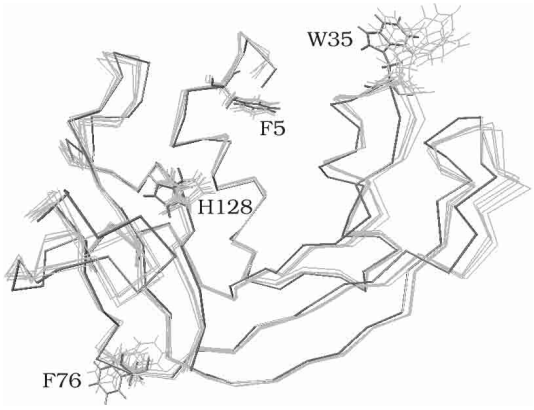

Fig. 5. C_α traces of 5 structures obtained from top three modes (see text) are shows in thin lines. The crystal structure (thick line) is also shown. Conformations of the residues Phe5, Trp35 and Phe76, which contribute significantly to the top modes, and His128, which undergoes conformational change during equilibration, are also shown.

the third mode was selected for coordinate generation corresponding to the transition region. The structures were subjected to 100 steps of minimization. A superposition of these selected five structures along with crystal structure is presented in Figure 5.

Dynamic cross-correlation maps (DCCMs) in the intervals chosen above were computed to see the correlations. Figures 6a and 6b show the DCCMs in the 400–650 ps and 1 200–1 400 ps regions. The correlation patterns are similar in these two time intervals as well as in the in-between time intervals. The expected cross-

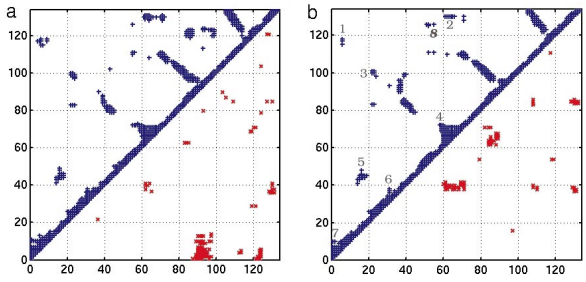

Fig. 6. Atomic displacement cross-correlation maps ($b_{ij} \geq 0.35$) for the two subspaces. The axes correspond to residue numbers. (a) 400–650 ps and (b) 1 200–1 400 ps. The upper triangle and lower triangle correspond to correlation (+) and anticorrelation (×) respectively. In (b) "S–S": strand-strand; and the correlated regions "1–8" are discussed in the text.

correlation of secondary structures, such as the ones due to helices, are seen along the diagonal and those from strands are seen either parallel or perpendicular to the diagonal depending on them being parallel or anti-parallel strands. Some of these regions are marked in Figure 6b. Besides these regular patterns, correlations are also seen in some non-secondary structural regions which are marked by numbers 1–8 in Figure 6b. These correlations arise due to the involvement of sidechain hydrogen bonds, which were detected by the clustering procedure mentioned above and will be discussed in a subsequent section. Thus a systematic side chain hydrogen bond analysis can assist in interpreting the dynamic cross-correlations maps of C_α atoms. The anti-correlations presented in the lower triangle of Figures 6a and 6b show more variability than correlations in different time intervals. However, certain common anti-cross-correlations are seen through out the simulation period.

The magnitude of the vector components of the top modes carry information on the residues contributing significantly to the mode in concern. The top two eigenvectors have similar profiles but differ in amplitudes, while the third does not have any sharp peaks. The residues corresponding to the vector component peaks in the top two modes were examined and many of them were identified as arginines on the surface. Besides, Phe5, Trp35, and Phe76 have high vector components. The follow up of the conformations of these residues during the simulation revealed that Phe5 does not undergo any conformational changes, whereas, Trp35 and Phe76 are involved in conformational changes. Torsion angle trajectories of Trp35 and Phe76 are shown in Figure 7 and the conformations accessed by these residues in different subspaces are shown in Figure 5. Incidentally, the active site residue His128 undergoes a conformational change from its initial (inactive) conformation to the biologically active conformation during the equilibration [55,56]. His128 then remains in the active conformation during the rest of the simulation. Both the conformations are represented in Figure 5.

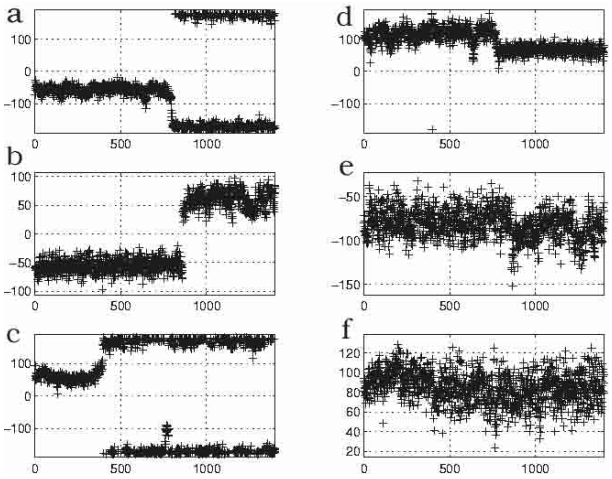

Fig. 7. The sidechain torsion angle trajectories: (a, b, c) correspond to χ_1 and (d, e, f) correspond to χ_{21} for residues Trp35, Phe76 and His128 respectively.

It is interesting to note that the conformational changes correspond to the transition time interval from one subspace to another. The third mode which dominates the transition region (725–825 ps) seems to be involved in global loop motions rather than in any specific residue movement. A comparison of the resulting structure with the crystal structure shows that the fluctuations seem to be related to the opening of the structure at a global level and is also consistent with increase radius of gyration (Fig. 2b) during the transition period. Thus, the vector components of top modes can lead us to detect the local and extended regions of the protein which are important for conformational subspaces and the transitions between them.

A thorough analysis of the essential dynamics including the monitoring of the changes in sidechain torsion angles together with the projections, RMSDs and radius of gyration, has aided in identifying the structural changes that have occurred between 600–900 ps. This transition has been interpreted as a consequence of sidechain conformational changes accompanied by increasing radius of gyration caused by the opening of the structure due to loop movements.

3.2 Sc–sc, sc–mc hydrogen bond clusters

The mainchain–mainchain (mc–mc) hydrogen bonds are well-known indicators of secondary structures in proteins. Here, we have studied the sc–sc and sc–mc hydrogen bonds. Hydrogen bond clusters have been analyzed in the crystal structures as well as in the selected modes. The analysis has yielded hydrogen bond clusters, which reveals that there are no drastic differences in the hydrogen bond patterns between the crystal structure and the selected modes, indicating the intactness of the tertiary structure during simulation. The clusters detected with four or more residues are listed in Table 1.

Table 1. List of major sc–sc/sc–mc hydrogen bond clusters (H-clusters) in the crystal structure. Also included are a few important H-clusters.

Cluster	Residues in sc–sc/sc–mc H-Cluster
C1	Pro3, Thr6, Gln9, Asp115, Arg117, Asp118
C2	Trp10, Gln14, Thr24, Asn31, Lys38, Asn41, Tyr98
C3	Asn87, Asn92, Asn95, Arg97
C4	Arg7, Asp112, Asn113
C5	His64, Asn65, Asp130
C6	Arg34, Trp35, Arg36
C7	Thr42, His82
C8	Gln14, Asn41
C9	Leu18, Arg45

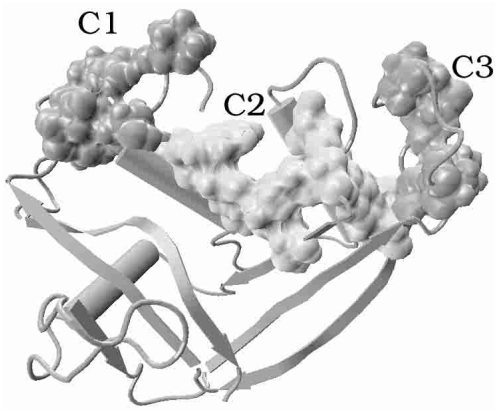

Fig. 8. The three major sc–sc/sc–mc clusters (Tab. 1) seen in crystal structures.

It is interesting to see a strong interaction between the N-terminus, first helix and a long loop involving residues 113–122 (that eventually connects the C-terminal strand) as part of a cluster (C1 in Fig. 8). This loop also has a number of proline residues, imparting rigidity. The network of sc–sc/sc–mc hydrogen bonds coupled with sterically rigid prolines can play an important role in fixing the loops in a structurally ordered fashion. The second cluster (C2) connects the residues from first two helices and the loops on either side of second helix. Tyr98 is also a part of cluster C2. Cluster C3 is formed by residues of a single loop that imparts rigidity to this region of the molecule. Thus, these three major clusters span the molecule from one side to the other, linking several secondary and non-secondary regions. Although no major differences in the clustering patterns in the subspaces are seen when compared to the crystal structure, the clusters C1 and C2, which connect sequentially distant residues, break up into smaller units in some of the subspace structures. Several non-secondary structure correlations in the cross-correlation maps (Fig. 6) were noticed and indeed many of them are from hydrogen bonds. They are either part of big cluster or two residue or three residue cluster. For instance, the correlation of Thr24–Tyr98, and Asn31–Lys38 are due to cluster C2. The residues involved

in the H-bonds Arg1–Phe5, Thr6–Asn118, Leu18–Arg45, His64–Asn65–Asp130 and Gln58–His72 also appear as correlated. The only instance we could not explain a correlation based on sidechain hydrogen bonds was the one between Cys55–Pro126 (marked "8" in Fig. 6b). We speculate that the reason might lie in the disulphide bridge between Cys55 and Cys111, and the hydrogen bond between the strands involving these residues (especially through hydrogen bond between Asp112 and Val125). Hence the hydrogen bond clustering procedure that has been developed can be extremely useful in looking at the tertiary structures of a protein in a systematic way, beyond the identification of secondary structures. It also helps in understanding the correlated global motions in proteins.

4 Conclusions

Molecular dynamics simulation followed by essential dynamics analysis have been performed on a human protein – Eosinophil Cationic Protein (ECP). The top essential modes are analyzed in detail. The essential dynamics analysis of the top modes indicate the access of two conformational subspaces during the simulation, which are dominated by the top two modes and the transition between them being dominated by the third mode. The cross-correlation maps of C_α atoms show interesting correlations and anticorrelations providing the details of the dynamics. From the vector components of the top modes, it was possible to identify Trp35 and Phe76 as the residues having different conformations in the subspaces and the concerted motion of several loops as the important movement during the transition phase.

The clustering of sidechain–sidechain/sidechain–mainchain hydrogen bonds in the crystal structure and in the essential modes has been carried out by the graph theoretic method. This provided a means of systematically analyzing sidechain hydrogen bonding patterns in proteins. The present analysis shows that most of the sidechain interactions in the essential modes are as intact as in the crystal structure. Minor variations, such as breaking up of a large cluster into smaller ones, are seen. The global motion in important modes are controlled by several sc–sc/sc–mc hydrogen bond clusters. The main contributors to the concerted motions identified in the present study bring into focus, perhaps for the first time, the explicit role of sidechain hydrogen bond clusters in protein dynamics in a systematic way. The significance in the context of the residues with large number of sidechain hydrogen bonds in concerted motion, stability and function of biomolecules can perhaps be verified by experiments, and the present study may stimulate further experimental and computational investigations.

We thank the Department of Science and Technology (DST), India, for supporting the project (No. SP/50/D-108/98) and Super Computer Education and Research Center (SERC) at Indian Institute of Science for computational facilities. BSS thanks Council for Scientific and Industrial Research (CSIR), India, for fellowship. It is a pleasure to thank Brinda Mohan and Shilpi Sharma for their help in proof reading the manuscript.

References

1. T.C. Bruice, K. Kahn, Curr. Opin. Chem. Biol. **4**, 540 (2000)
2. S. Vishveshwara, M.S. Madhusudhan, J.V. Maizel Jr, Biophys. Chem. 89 (2001)
3. M.S. Madhusudhan, B.S. Sanjeev, S. Vishveshwara, Proteins: Struct. Func. Gen. **45**, 30 (2001)
4. M.S. Madhusudhan, S. Vishveshwara, Proteins: Struct. Func. Gen. **42**, 125 (2001)
5. M.S. Madhusudhan, S. Vishveshwara, Biopolymers **49**, 131 (1999)
6. M.S. Madhusudhan, S. Vishveshwara, J. Biomol. Struct. Dyn. **16**, 715 (1998)
7. G. Nadig, G.S. Ratnaparkhi, R. Varadarajan, S. Vishveshwara, Protein Sci. **5**, 2104 (1996)
8. K. Seshadri, V.S.R. Rao, S. Vishveshwara, Biophys. J. **69**, 2185 (1995)
9. K. Seshadri, V.S.R. Rao, S. Vishveshwara, J. Biomol. Struct. Dyn. **12**, 581 (1994)
10. K. Seshadri, P.V. Balaji, V.S.R. Rao, S. Vishveshwara, J. Biomol. Struct. Dyn. **11**, 395 (1993)
11. K. Seshadri, V.S.R. Rao, S. Vishveshwara, J. Biomol. Struct. Dyn. **9**, 1253 (1992)
12. U.H. Hansmann, Y. Okamoto, Curr. Opin. Struct. Biol. **9**, 177 (1999)
13. L. Carlacci, A.S. Edison, Proteins **40**, 367 (2000)
14. W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. **14**, 33 (1996)
15. S. Sorrentino *et al.*, J. Biol. Chem. **267**, 14859 (1992)
16. M.R. Synder, G.J. Gleich, in *Ribonulceases: Structures and Functions*, edited by G. D'Alessio, J.F. Riordan (Academic Press, New York, 1997)
17. U. Gullberg *et al.*, Biochem. Biophys. Res. Commun. **139**, 1239 (1986)
18. L. Juhlin, P. Venge, Acta. Derm. Venereol. **71**, 495 (1991)
19. S.J. Ackerman *et al.*, Am. J. Trop. Med. Hyg. **34**, 735 (1985)
20. H.F. Rosenberg, J. Biol. Chem. **270**, 7876 (1995)
21. J.B. Domachowske *et al.*, Nucleic Acids Res. **26**, 3358 (1998)
22. N.A. Singhania *et al.*, J. Mol. Evol. **49**, 721 (1999)
23. J.J. Beintema *et al.*, in *Ribonulceases: Structures and Functions*, edited by G. D'Alessio, J.F. Riordan (Academic Press, New York, 1997)
24. G. Mallorqui-Fernandez *et al.*, J. Mol. Biol. **300**, 1297 (2000)
25. W.D. Cornell *et al.*, J. Am. Chem. Soc. **117**, 765 (1995)
26. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, J. Chem. Phys. **79**, 926 (1983)
27. T. Darden, D. York, L. Perdersen, J. Chem. Phys. **98**, 10089 (1993)
28. T. Ichiye, M. Karplus, Proteins: Struct. Func. Gen. **11**, 205 (1991)
29. T. Horiuchi, N. Go, Proteins **10**, 106 (1991)
30. M.M. Teeter, A.D. Case, J. Phys. Chem. **94**, 8091 (1990)
31. D. Perahia, R.M. Levy, M. Karplus, Biopolymers **29**, 645 (1990)

32. A. Kitao, F. Hirata, N. Go, Chem. Phys. **158**, 447 (1991)
33. A. Amadei, A.B. Linssen, H.J. Berendsen, Proteins: Struct. Func. Gen. **17**, 412 (1993)
34. D.M. van Aalten *et al.*, Proteins: Struct. Func. Gen. **22**, 45 (1995)
35. C. Arcangeli, A.R. Bizzarri, S. Cannistraro, Biophys. Chem. **92**, 183 (2001)
36. P.H. Hunenberger, A.E. Mark, W.F. van Gunsteren, J. Mol. Biol. **252**, 492 (1995)
37. B.L. de Groot *et al.*, Proteins: Struct. Func. Gen. **26**, 314 (1996)
38. J. Lee, S.W. Suh S. Shin, J. Biomol. Struct. Dyn. **18**, 297 (2000)
39. C. Arcangeli, A.R. Bizzarri, S. Cannistraro, Biophys. Chem. **90**, 45 (2001)
40. A. Amadei *et al.*, Proteins: Struct. Func. Gen. **35**, 283 (1999)
41. C. Branden, J. Tooze, *Introduction to Protein Structure* (Garland Publishing, Inc., New York and London, 1991)
42. T. Creighton, *Proteins: Structures and Molecular Properties*, 2nd edn. (W.H. Freeman and Company, New York, 1996)
43. G. Bulaj, D.P. Goldenberg, J. Mol. Biol. **313**, 639 (2001)
44. Y.W. Chen, A.R. Fersht, K. Henrick, J. Mol. Biol. **234**, 1158 (1993)
45. M.P. Byrne, R.L. Manuel, L.G. Lowe, W.E. Stites, Biochemistry **34**, 13949 (1995)
46. C.N. Pace, B.A. Shirley, M. McNutt, K. Gajiwala, FASEB J. **10**, 75 (1996)
47. Y. Yamagata Y *et al.*, Biochemistry **37**, 9355 (1998)
48. K. Takano *et al.*, Biochemistry **38**, 6623 (1999)
49. C.N. Pace, Biochemistry **40**, 310 (2001)
50. V.P. Grantcharova, D.S. Riddle, J.V. Santiago, D. Baker D, Nat. Struct. Biol. **5**, 714 (1998)
51. B.A. Krantz, L.B. Moran, A. Kentsis, T.R. Sosnick, Nat. Struct. Biol. **7**, 62 (2000)
52. K.M. Hall, Manag. Sci. **17**, 219 (1970)
53. S.M. Patra, S. Vishveshwara, Biophys. Chem. **84**, 13 (2000)
54. N. Kannan, S. Vishveshwara, J. Mol. Biol. **292**, 441 (1999)
55. I. Gutman, C. Cvetkovich, Croat. Chem. Acta **49**, 115 (1977)
56. I. Zegers *et al.*, Protein Sci. **3**, 2322 (1994)
57. J. Santoro *et al.*, J. Mol. Biol. **229**, 722 (1993)
58. E. Boix *et al.*, Biochemistry **38**, 16794 (1999)