Conformational changes and allosteric communications in human serum albumin due to ligand binding

Navjeet Ahalawat and Rajesh K. Murarka*

Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhauri, Bhopal, Madhya Pradesh 462066, India

Communicated by Ramaswamy H. Sarma

(Received 25 September 2014; accepted 5 December 2014)

It is well recognized that knowledge of structure alone is not sufficient to understand the fundamental mechanism of biomolecular recognition. Information of dynamics is necessary to describe motions involving relevant conformational states of functional importance. We carried out principal component analysis (PCA) of structural ensemble, derived from 84 crystal structures of human serum albumin (HSA) with different ligands and/or different conditions, to identify the functionally important collective motions, and compared with the motions along the low-frequency modes obtained from normal mode analysis of the elastic network model (ENM) of unliganded HSA. Significant overlap is observed in the collective motions derived from PCA and ENM. PCA and ENM analysis revealed that ligand selects the most favored conformation from accessible equilibrium structures of unliganded HSA. Further, we analyzed dynamic network obtained from molecular dynamics simulations of unliganded HSA and fatty acids-bound HSA. Our results show that fatty acids-bound HSA has more robust community network with several routes to communicate among different parts of the protein. Critical nodes (residues) identified from dynamic network analysis are in good agreement with allosteric residues obtained from sequence-based statistical coupling analysis method. This work underscores the importance of intrinsic structural dynamics of proteins in ligand recognition and can be utilized for the development of novel drugs with optimum activity.

Keywords: human serum albumin; elastic network model; principal component analysis; X-ray ensemble; dynamic network analysis; statistical coupling analysis; allostery

Introduction

Understanding of the fundamental mechanisms of molecular recognition is central to understand biology at the molecular level. The conformational transitions that originate in the ligand binding are explained on the basis of two representative models: the induced-fit model proposed by Koshland (1958) and the preexisting equilibrium dynamics or conformational selection model (Boehr, Nussinov, & Wright, 2009; Ma, Kumar, Tsai, & Nussinov, 1999; Monod, Wyman, & Changeux, 1965). In induced-fit model, ligand binding induces a conformational change in the macromolecule to achieve the preferred ligand-bound conformation, whereas according to conformational selection model, ligand selectively binds to a conformational substate available prior to binding with low population in an equilibrium ensemble of substates of the macromolecule, eventually shifting the equilibrium toward the ligand-bound conformation. The Protein Data Bank (PDB) (Berman et al., 2000) is rapidly and continuously growing with structure deposition and more than 89,000 structures are already available. Many proteins have multiple X-ray structures with different ligands and/or under different conditions. Although crystal structures provide significant information about binding interactions but alone they are not sufficient to understand the mechanism or origin of biomolecular recognition. In fact, proteins do not function as static entities; they are rather engaged in functional motions and interactions with other molecules. Understanding of biomolecular interactions helps in identifying the cause of diseases and ultimately to develop therapeutic strategies. Molecular docking is the most commonly used technique in structure-based drug designing. However, incorporation of protein flexibility in docking methods is a major concern because most of the algorithms are based on computationally expensive methods like molecular dynamics (MD) and Monte Carlo simulations. Thus, the prediction of a conformational change upon ligand binding remains a formidable challenge because it is difficult to explore adequate conformational space in a reasonable computation time.

It is widely recognized that the dynamics of proteins manifested in the transition between different conformational substates play key role in their functions.

*Corresponding author. Email: rkm@iiserb.ac.in

© 2015 Taylor & Francis
(Frauenfelder, Sligar, & Wolynes, 1991; Henzler-Wildman & Kern, 2007; Jana, Adkar, Biswas, & Bagchi, 2011; Lange et al., 2008; Smock & Gierasch, 2009; Tokuriki & Tawfik, 2009). Elastic network models (ENMs) based on native contact topology have been highly successful in exploring the functional collective dynamics of biomolecules in recent years (Bahar, Lezon, Bakan, & Shrivastava, 2010; Bahar, Lezon, Yang, & Eyal, 2010; Sanejouand, 2013; Tama & Brooks, 2006; Yang, Song, & Jernigan, 2007; Zheng, Brooks, & Thirumalai, 2006). The collective dynamics predicted by the coarse-grained ENMs (in which the Cα-atoms representing the nodes are connected by harmonic springs) are almost indistinguishable from those that would otherwise be obtained by full-fledged atomic force fields (Ahmed, Villinger, & Gohlke, 2010; Romo & Grossfield, 2011; Rueda, Chacon, & Orozco, 2007; Skjaerven, Martinez, & Reuter, 2011). Further, combination of MD simulation and ENM has been proven as a better sampling approach to find out large protein conformational transitions by determining unique substates (Peng & Head-Gordon, 2011; Peng, Zhang, & Head-Gordon, 2010).

After a drug is absorbed into the bloodstream, it gets distributed to various tissues of the body. These drug molecules can bind to a variety of blood components i.e. red blood cells, leukocytes, and platelets, as well as plasma proteins. Plasma proteins binding can affect the pharmacokinetics of drugs. Human serum albumin (HSA), a most abundant protein in blood plasma, can bind different classes of ligands ranging from metal ions and small molecules to macromolecular compounds such as dendrimers, fullerene derivatives, fatty acids (FAs), hormones, proteins, and drugs at multiple sites (Curry, 2009; Fanali et al., 2012; Ghuman et al., 2005). HSA also helps in increasing solubility in plasma, decreasing toxicity, and/or protection against oxidation of the bound ligand (Curry, 2009; Fanali et al., 2012; Ghuman et al., 2005). This extraordinary ability of HSA has stimulated many efforts to understand what determines the binding of a diverse range of ligands. HSA is a highly flexible 67 kDa monomer containing three homologous helical domains, I (residues 1–195), II (196–383) and III (384–585), each divided into A and B subdomains and the overall structure is stabilized by 17 disulfide bridges. Both ends of HSA protein show high conformational flexibility (Sugio, Kashima, Mochizuki, Noda, & Kobayashi, 1999) and are not available in most of the crystal structures. Three homologous domains of HSA interact with each other according to physiological condition and show a wide variety of conformational changes with allosteric effect (Fanali et al., 2012; Ghuman et al., 2005). Several interaction studies of different drugs/ligands with HSA have been performed using X-ray crystallography (Bhattacharya, Grune, & Curry, 2000; Curry, 2009; Curry, Mandelkow, Brick, & Franks, 1998; Pettitpas, Grune, Bhattacharya, & Curry, 2001; Simard et al., 2005; Zunszain, Ghuman, McDonagh, & Curry, 2008), MD simulations (Fujiiwara & Amisaki, 2008; Sudhamalla, Gokara, Ahalawat, Amoru, & Subramaniam, 2010) and nuclear magnetic resonance (NMR) technique (Fanali, Pariani, Ascenzi, & Fasano, 2009; Krenzel, Chen, & Hamilton, 2013; Simard, Zunszain, Hamilton, & Curry, 2006). However, to the best of our knowledge, a global picture of correlation between mechanism of molecular recognition by HSA and the conformational fluctuations (that is, structural dynamics) in the native state ensemble is yet to be established.

In this article, we address a detailed theoretical study of the molecular mechanism of ligand binding in HSA and aim to identify crucial residues involved in long-range and short-range communications. Toward this goal, we have used essential dynamics-refined ENM (ed-ENM) (Orellana et al., 2010) method for studying the global fluctuations in the native state of HSA. It is well established that handful of experimentally known conformers of a protein determined under different conditions, or with small variations in sequence, capture a representative subset of the true native state ensemble (Best, Linderoff-Larsen, DePristo, & Vendruscolo, 2006). X-ray structural ensemble, derived from 84 crystal structures resolved for HSA with different ligands and/or under different conditions, has been used as the reference for the conformational space accessible to HSA. Principal component analysis (PCA) was employed to derive the functionally important collective motions from X-ray structural ensemble. Furthermore, we utilized dynamical network analysis (Sethi, Eargle, Black, & Luthey-Schulten, 2009; Vanwart, Eargle, Luthey-Schulten, & Amaro, 2012) and community analysis based on mutual information of correlated residue motions to characterize the information flow between communities in the network and identify the crucial residues involved in allosteric communication.

We found that only first few slow modes of ENM are sufficient to capture the major functional subspace explored by PCA modes, thereby suggesting that ligand selects the most favorable structure from accessible conformational space (i.e. conformational selection). Large-scale global conformational changes are observed due to FAs binding, while binding of drugs show relatively small structural deviation in comparison to FAs. This information can be utilized in docking algorithms to improve their predictive power. Major response of FAs binding is found in communities of FA2- and FA3-binding sites known for allosteric effect. We also used statistical coupling analysis (SCA) (Lockless & Ranganathan, 1999; Suel, Lockless, Wall, & Ranganathan, 2003) method based on coevolutionary information of protein sequence to predict the allosteric residues, and compare with the results obtained from community network analysis.
Materials and methods

Principal component analysis

The complete list of PDBs used to generate experimental native state ensemble are available in the Supporting information Table S1. X-ray ensemble is generated by superimposing $C_a$ atoms of crystal structures onto the reference structure (PDB 1AO6) using Kabsch algorithm in an iterative procedure. Each structure is represented by $3N$ Cartesian coordinates, where $N$ is the number of $C_a$ atoms. We build covariance matrix $C = XX^T$ from X-ray ensemble; here, $X = [(q_1 - \bar{q})(q_2 - \bar{q})\ldots(q_M - \bar{q})]_{3N \times M}$ is a matrix of $M$ columns (number of structures), each with $3N$ elements, represent the deviation vector from the average structure of the ensemble. Covariance matrix is diagonalized to get principal directions of structural variations. These principle directions are rank-ordered with decreasing variations and named as principal component 1 (PC1), principal component 2 (PC2), etc.

Elastic network normal mode analysis

ENMs describe the dynamics of a protein by its contact topology, represented as a network of nodes and springs, and only the $C_a$ atoms of the individual residues are considered as nodes and each residue pair located within a cut-off distance is connected by a spring of uniform force constant. ENMs are based on harmonic, near-equilibrium approach, and thus have problems in capturing large anharmonic motions. Many attempts have been made to improve the robustness and generality of ENMs (Atilgan et al., 2001; Hinsen, 1998; Tama, Gadea, Marques, & Sanejouand, 2000; Yang, Song, & Jernigan, 2009; Zheng, 2008). Here, we adopted ed-ENM (Orellana et al., 2010) method based on hybrid potential, which has been calibrated using a large database of MD trajectories to trace better protein flexibility. ed-ENM method captures the flexibility of NMR structural ensembles with remarkable precision. This is a parameter-free ENM method, avoiding any arbitrary cutoff, and uses simple and robust scaling of the local backbone and long-range contacts. This method treats the sequential and spatial (nonsequential) contacts differentially. A fully connected matrix is used for the first $M$ neighbors (sequential contacts), while a continuum distance-dependent function with a calibrated size-dependent cutoff is used for residue pairs that are close in space and not included in the first group. ed-ENM method uses two kinds of inter-residue force constants: (i) very stiff sequence distance-dependent constant, and (ii) Cartesian distance-dependent constant with stiffness scaled down to the sixth power of the exponential term (for more details see Orellana et al., 2010). The overall potential energy of network is given by

$$E = \frac{1}{2} \sum_{i \neq j} K_{ij} \Gamma_{ij} (R_{ij} - R_{ij}^0)^2 = \frac{1}{2} \Delta R^T H \Delta R.$$

Here, $K_{ij}$ is force constant acting between any pair of nodes (residues) $i$ and $j$; $\Gamma_{ij}$ is element of (Kirchhoff matrix) inter-residue connectivity matrix, where $i$ and $j$ represent residue index; $R_{ij}$ and $R_{ij}^0$ are the instantaneous and reference (equilibrium) distances between each pair of $\alpha$-carbon of residues $i$ and $j$; $H$ is the Hessian matrix of $3N \times 3N$, which is second derivative of the potential; $\Delta R = (\Delta x_1, \Delta y_1, \Delta z_1, \ldots, \Delta x_N)^T$ is the position deviation vector for all $C_a$ atoms of the protein; and $N$ is the number of residues ($C_a$-atoms).

Fluctuations’ correlations (cross-correlation) between residue $i$ and residue $j$ are calculated by

$$C(i,j) = \frac{\langle \Delta R_i \Delta R_j \rangle}{\langle \Delta R_i \Delta R_i \rangle^{1/2} \langle \Delta R_j \Delta R_j \rangle^{1/2}}.$$

Here, $\langle \Delta R_i \Delta R_j \rangle = k_BT \text{trace}[H^{-1}]_{ij}$ and $\langle \Delta R_i \Delta R_i \rangle = k_BT \text{trace}[H^{-1}]_{ii}$. The two limits of $C(i,j)$ are 1 and −1 and represent fully correlated and fully anticorrelated motions, respectively.

The overlap $I_j$ describes the similarity between the direction of conformational change along a transition pathway and a given normal mode of protein. This is computed as

$$I_j = \frac{|\mathbf{a}_j \cdot \Delta \mathbf{r}|}{|\mathbf{a}_j| \cdot ||\Delta \mathbf{r}||} = \left|\sum_{i=1}^{3N} a_{ij} \Delta r_i \right| \left|\sum_{i=1}^{3N} a_{ij}^2 \sum_{i=1}^{3N} \Delta r_i^2 \right|^{1/2}$$

where $a_{ij}$ is the eigenvector of mode $j$, $a_{ij}$ is the $i$th coordinate of $j$th mode, $\Delta \mathbf{r}$ represents the deviation vector of crystal structure from the reference structure, and $\Delta r_i$ is the $i$th coordinate of the deviation vector $\Delta \mathbf{r}$.

Collectivity index $k$ (or information entropy) reflects the number of atoms ($C_a$ atoms or residues), which are significantly affected during the conformational change and can be calculated as

$$k = \frac{1}{N} \exp \left\{ \sum_{i=1}^{N} \left( s \Delta r_i^2 \log(s \Delta r_i^2) \right) \right\},$$

where the sum is over the $C_a$ atoms of the protein, and $s$ is a normalization factor that can be calculated by $\sum_{i=1}^{N} (s \Delta r_i^2) = 1$ and $\Delta r_i^2 (= \Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2)$ is square fluctuation of $i$th residue ($C_a$ atom); $k$ is confined to the interval between $N^{-1}$ and 1; $k = 1$ means maximally collective conformational change, and all residues have identical $\Delta r_i^2$.

The B-factor or the crystallographic Debye–Waller factor is related to the expected residue fluctuations and defined as
$B_i = \frac{8\pi^2}{3} \langle (\Delta R_i)^2 \rangle = \frac{8\pi^2}{3} k_B T \text{trace}[H^{-1}]_{ii}$

where $a_k$ and $\lambda_k$ are eigenvector and eigenvalue of $k$th mode, respectively.

**MD simulations and dynamic network analysis**

For dynamic network analysis, we have used the combination of MD simulations, generalized cross-correlations (Lange & Grubmuller, 2006) based on mutual information theory and graph theory. Atomic MD simulations for both states apo-HSA (PDB 1AO6) and HSA-FAs (PDB 1BJ5) were carried out using GROMACS (Van Der Spoel et al., 2005) version 4.5 and AMBER-ff99SB force field. The topologies and force-field parameters (except partial atomic charges) of myristate molecules were generated using ACPYPE (Wang, Wang, Kollman, & Case, 2006) tool, which uses the ANTECHAMBER module with generalized amber force field (Wang, Wolf, Caldwell, Kollman, & Case, 2004). After ab initio optimization at the HF/6-31G* level using Gaussian09 (Gaussian Inc., Pittsburgh, USA), a restrained electrostatic potential fit procedure was used to generate the partial atomic charges of myristate molecules. The initial structures were solvated in a cubic box with 43,712 and 43,580 TIP4P-Ew water (Horn et al., 2004) molecules for apo-HSA and HSA-FAs, respectively. The minimum distance between protein atoms and the box surface was 10 Å. The LINCS (Hess, Bekker, Berendsen, & Fraaije, 1997) algorithm was applied to constrain hydrogen atoms, and an integration time step of 2.0 ps was used. The bonds and the angles of the water molecules were constrained using the SETTLE (Miyamoto & Kollman, 1992) algorithm. To neutralize the systems, 15 Na$^+$ and 14 Na$^+$ were added to apo-HSA and HSA-FAs complex, respectively. Eight replicas of each energy-minimized system were equilibrated starting from different initial atomic velocities. Energy minimization and equilibration steps consist of: (1) a steepest descent energy minimization, (2) a 200 ps position-restrained (all heavy atoms of HSA and myristate molecules) NVT simulation with restraining force constant of 1000 kJ mol$^{-1}$ nm$^{-2}$, (3) a 500 ps position-restrained NPT simulation with restraining force constant of 1000 kJ mol$^{-1}$ nm$^{-2}$, and (4) a 5 ns of NPT equilibration at 300 K and 1 bar pressure using the Berendsen coupling scheme (Berendsen, Postma, Vangunsteren, Dinola, & Haak, 1984) for both temperature and pressure with coupling constants .1 ps and .5 ps, respectively. Each replica was simulated for another 50 ns of production dynamics at 300 K, using separate v-rescale thermostats (Bussi, Donadio, & Parrinello, 2007) for protein, myristate, ions, and the water, with a time constant of $\tau = .1$ ps. The pressure was maintained at 1 bar using the Parrinello–Rahman (Parrinello & Rahman, 1981) algorithm with a coupling constant of 2.0 ps. The electrostatic interactions were evaluated using the particle mesh Ewald (Darden, York, & Pedersen, 1993) method, with a cutoff of 10 Å. The van der waals interactions were evaluated using a cutoff of 10 Å. Generalized cross-correlations of residues in the apo-HSA and HSA-FAs-bound states are calculated based on mutual information between all $C_{\alpha}$ atoms in the protein using the gromacs tool g_correlation developed by Lange and Grubmuller (2006). Dynamic network analysis is carried out using the NetworkView plugin (Eagle & Luthey-Schulten, 2012) in VMD (Humphrey, Dalke, & Schulten, 1996). In dynamical network, each $C_{\alpha}$ atom of residues represents a node and two residues are connected by an edge if in any two of their heavy atoms are within 5 Å for more than 75% of the simulation time. Each edge is weighted by averaged generalized cross-correlations obtained from all (eight) MD trajectories.

**Statistical coupling analysis**

Homologues sequences for HSA (PDB 1AO6) were retrieved from NCBI nr database using psi-blast (Altschul et al., 1997) with cutoff 1.0E-3 e-value. Muscle tool (Edgar, 2004) was used to generate multiple sequence alignment (MSA) considering only less than 96% sequences similarity with more than 80% sequence coverage. Columns having less than 20% gap were considered for all calculations. SCA of MSA consisting of 231 sequences was done using SCA toolbox (SCA 5.0). Further details of the SCA method can be found elsewhere (Lockless & Ranganathan, 1999; Smock et al., 2010; Suel et al., 2003).

**Results and discussion**

**PCA results**

PCA is a widely used technique to reveal the concerted fluctuations with large amplitudes from a set of configurations. PCA results are obtained for HSA using the ensemble of X-ray crystallographic structures. Figure 1 shows the projection of HSA structures onto the subspace spanned by first two principal components (PCs) determined for the X-ray structure ensemble. PC1 and PC2 clearly divide the ensemble of structures into two well-defined clusters. Interestingly, these two clusters differ only on the basis of FAs bound structures and FAs-free structures; cluster-1 (with 45 structures) represents structures without FAs, while cluster-2 (with 39 structures) represents structures with FAs. Minimum and
maximum root mean square deviation (RMSD) within cluster-1 are .36 and 2.48 Å from its average structure, while for cluster-2 they are .22 and 2.53 Å from its average structure, respectively. Minimum and maximum RMSD between these two clusters are 3.79 and 6.73 Å, respectively. When we plotted the structural ensemble on first three PCs, we did not observe any change in cluster classification. This is consistent with the fact that PC1 and PC2 together contribute 92% to the total variation (PC1 and PC2 contribute 89% and 3%, respectively) and, is therefore sufficient to explain the global structural changes. The structural variation represented by highly dominated PC1 is shown in Figure S1.

HSA is a well-known monomeric protein for its allosteric properties. FAs and heme both are known as allosteric modulators of HSA (Ascenzi, Bocedi, Notari, Menegatti, & Fasano, 2005; Fanali et al., 2012; Simard et al., 2006). Although cluster-2 in PC1–PC2 projection (Figure 1) includes the structure that is bound to FAs and heme (PDB 1N5U), but the structure bound to FAs and hemin (PDB 1O9X) has slightly different conformation than typically observed for the FA-bound state; this may be because of the crystal packing effect as reported earlier (Zanssain, Ghuman, Komatsu, Tsuchida, & Curry, 2003). Further it is important to note that the medium/long chain FAs (e.g. myristate) are known to act as allosteric regulators in HSA, but FAs with shorter chain length (e.g. octanoate) are not (Fanali et al., 2012). As structures of HSA complexed with short chain FAs are yet not available, the structural ensemble considered in this study consists of only medium/long chain FAs-bound structures. From the projection of structural ensemble on PC1 and PC2 (Figure 1), it is clearly evident that FA-bound complexes show large-scale global conformational changes, whereas all other ligands (e.g. drugs, proteins, etc.) show relatively small structural deviation in HSA.

ed-ENM analysis
In contrast to PCA, ed-ENM analysis is performed for a single unliganded structure, that is, for apo-HSA (PDB 1AO6). Low-frequency modes (the first six low-frequency modes are zero modes corresponding to the three translational and three rotational degrees of freedom) are assumed to be functional modes and contribute to global molecular motions of protein. To validate the molecular motions predicted by ed-ENM, we have compared the theoretical B-factor obtained from ed-ENM with the experimental B-factor (PDB 1AO6) and found a very
good agreement with a correlation coefficient, $R = .7$ (Figure 2). We also calculated collectivity index, which evaluates how many residues are significantly affected in a particular mode, and found the values .42, .40, and .63 for first three modes, respectively. This suggests that the motions in ed-ENM mode three (ed-ENM3) involve collective displacements of relatively large number of residues compared to the first two modes. Previous experimental studies have shown that subdomains IB-IIA and FA2–FA3-binding sites are allosterically coupled. Binding of heme in subdomain IB allosterically inhibits the binding of warfarin in subdomain IIA, and vice versa (Ascenzi et al., 2005; Ascenzi & Fasano, 2010). Binding of myristate at the FA2-binding site (interfaces of IA, IB and IIA) and at the FA3 site (subdomain IIIA) drives allosteric transitions in HSA (Curry, 2009; Fanali et al., 2012). Our results also show positive cross-correlations emphasizing allosteric links between these ligand-binding sites (Figure 3). It should be noted that anticorrelated motions observed between subdomain IB and IIIB, and mixed (positive and negative correlated) motions between IA and IIIB (Figure 3) helps in producing large conformational changes in HSA.

**Comparison of PCA and ed-ENM results**

To verify the relevance of modes derived from PCA analysis and ed-ENM analysis, we calculated the overlap between the subspaces spanned by the eigenvectors. Figure 4 shows the overlap between the first seven lowest frequency PCA and ed-ENM modes. Such comparisons allow us to identify PCA and ed-ENM modes that are in close correspondence with each other.

---

**Figure 2.** Comparison of experimental B-factor profile with calculated one from ed-ENM. Blue line represents the experimental B-factor (PDB 1AO6) and green line represents calculated B-factor from ed-ENM.

**Figure 3.** Correlation maps for residue fluctuations in HSA calculated over first three ed-ENM modes. The red and blue regions correspond to strongly correlated and strongly anticorrelated pair of residues, respectively. Correlated pairs undergo concerted same directional motions; anticorrelated pairs, on the other hand, undergo concerted opposite directional motions. Rectangular boxes show allosterically linked regions.

**Figure 4.** Correlations between experimentally observed (PCA) and computationally obtained (ed-ENM) modes: Overlaps between top seven modes from PCA and ed-ENM.

---

The counterpart of highly dominated PC1 is found to be ed-ENM1 and ed-ENM3 with overlap values of .50 and .60, respectively (Figure 4). Comparison of structural variations along ed-ENM1, ed-ENM3, and PC1 is shown in Figure 5(A) and (B). Distribution of X-ray structures on the plane of PC1 and ed-ENM1 (Figure 5(C)), and PC1-and ed-ENM3 (Figure 5(D)) show high correlation with an absolute value of correlation coefficients .96 and
respectively. To understand which low-frequency mode makes a significant contribution to the structural changes from unbound to FA-bound state, we calculated the overlap between ed-ENM modes and deviation vector from unliganded HSA (PDB: 1AO6) to FA bound structures. Interestingly, we found that all FA-bound structures show very similar overlap pattern with ed-ENM modes, where largest contributing modes are ed-ENM1 and ed-ENM3 (Figure 6(A), Supporting information Table S2). Overlap between ed-ENM modes and deviation vector from unliganded HSA (PDB: 1AO6) to structures complexed with different ligands (other than FAs) does not show any fixed pattern; however, slow modes do contribute to the binding of these ligands (Figure 6(B), Supporting information Table S2). Further, we calculated the cumulative overlap to check how many ed-ENM modes are required to explore the subspace spanned by PC1 and PC2. Cumulative overlap represents the square root of sum of individual squared overlaps, starting from the lowest nonzero frequency modes. Figure 6(C) shows the cumulative overlap of both PC1 and PC2 with 10 slowest ed-ENM modes. Clearly, first eight ed-ENM modes produced more than 90% overlap with both PC1 and PC2. The present comparison demonstrates that the subspace sampled by the softest ed-ENM modes is able to explore the experimentally accessed conformational space. In other words, the observed conformational changes after ligand binding are likely to be facilitated by low-frequency motions that are intrinsic to the structure. Our results therefore revealed that in the absence of ligands, protein fluctuates into conformations that resemble to the ligand-bound states. This suggests that preexisting equilibrium dynamics could be a preferred mechanism of binding of ligands to highly flexible HSA. To this end, it should be noted that our results are in support of previous findings (Bakan & Bahar, 2009; Yang, Song, Carriquiry, & Jernigan, 2008) where a close correlation between ENM and PCA was observed, thereby confirming the validity of using ENM to understand the functional dynamics of proteins.

**Dynamic network analysis**

From PCA results of X-ray ensemble, it is clear that FAs binding induce large-scale global conformational changes in HSA. MD simulation is a powerful method to provide a detailed atomistic description of these large rearrangements due to FAs binding. Dynamical weighted network
for both HSA and HSA-FAs are examined through community network analysis using the NetworkView (Eargle & Luthey-Schulten, 2012) plugin in VMD (Humphrey et al., 1996). Average values of the generalized cross-correlations of residues obtained from MD simulations are used to weigh the edges of a network. The time-averaged connectivity of the nodes is used to identify subnetworks or communities in the network. Figure 7 shows graph network representation of the optimum structure of communities obtained for HSA and HSA-FAs. The Girvan–Newman algorithm splits the weighted network of HSA into 12 communities and that for HSA-FAs to 13 communities. Flow of information among communities can be traced using coarse-grained network where node represents a community and edge represents intercommunity total betweenness centrality. Comparison of coarse-grained network of both states, HSA and HSA-FAs, shows HSA-FAs has more robust community network than HSA with several routes to communicate among different parts of the protein. Furthermore, inter-community information flow is strong in HSA-FAs than in HSA community network. Community 7 and 3’ in HSA and HSA-FAs network, respectively, are important as they are strongly connected with communities of all domains; they include helix h1 (residue 196–205) which is a major molecular hinge site (predicted by ed-ENM analysis) responsible for large movement of domain I and domain III with respect to domain II. Subdomain IIB, which has very less movements in ed-ENM and PCA modes, has only one independent community 6 and two shared communities 2 and 3 in the community network of HSA; however, there are two independent communities 6’ and 9’ and two shared communities 8’ and 13’ for HSA-FAs. This less coarse-grained structure of subdomain IIB in HSA-FAs shows that this subdomain is now more actively communicating with other
subdomains. Major rearrangements of residues are observed between the communities of domain I and II (Supporting information Figure S2). Interface of subdomain IA and IIA where FA2 binds and interface of domain II and domain III near the FA3-binding site have large rearrangement of residues among communities after FA binding. These two sites (subdomains IA, FA2-binding site and IIIA, FA3-binding site), known for major allosteric effect, are responsible for the perturbation of correlated motions of residues in HSA and can be seen in community network analysis; community network structure depends on correlated motions of residues obtained from dynamics. Intercommunity information flow depends only on few residues that connect them. Mutations in these sites can disrupt complete information flow in distant communities of the network. Therefore, these residues can be considered as critical nodes. Comparisons of these critical residues (Supporting information Table S3) between HSA and HSA-FAs community network shows many of them are common and can be considered as structurally and functionally important for HSA.

**SCA and allosteric hot spots**

A single mutation away from the binding site or functional site can change the whole dynamics of the protein as well as its binding specificity. These single site mutations are not independent; they are always linked with other sites, and if one mutation occurs, it needs to be compensated with other to preserve the functional dynamics. How coevolutionary processes are related to functional diversification within protein families has been studied successively in recent years, and continues to be a topic of central importance. Serum albumin is one of the most rapidly evolving proteins, where ~70–80% of its residues have been replaced in 500 Myr. Thus, it would be interesting to know the biologically relevant correlated residue positions and their role in functional dynamics in HSA. We adopted a sequence-based method to estimate energetic coupling between residues in proteins, named SCA method (Lockless & Ranganathan, 1999). Figure 8(A) shows hierarchical clustering of SCA correlation matrix, which is a global representation of evolutionary coupling between all pairs of residues. Hierarchical clustering clusters residues displaying...
similar and interconnected pattern of significant energetic coupling values. We found that most of the amino acid positions are weakly correlated and evolving independently (Figure 8(A)), only few positions (less than 20%) (Supporting information Table S3) show strong coupling and can be considered as allosteric hotspots. These positions are not clustered in any particular domain of the protein; they are scattered and form a continuously connected network of residues in 3D space, which indicates that different domains of HSA are allosterically linked with cluster of energetically coupled residues. Further, allosteric residues from SCA are in good agreement with residues obtained using dynamic network analysis (Supporting information Table S3). It should be noted that dynamic network analysis and SCA are based on very different principles and assumptions where the former one is purely topology dependent, while SCA relies on information based on MSA.

In several studies, the fluctuation profile of residues along the low- and high-frequency ENM modes has successfully been utilized to identify hinge sites responsible for global conformational changes, and kinetically important residues responsible for protein stability, respectively (Bahar, Atilgan, Demirel, & Erman, 1998; Emekli, Schneidman-Duhovny, Wolfson, Nussinov, & Haliloglu, 2008; Haliloglu, Keskin, Ma, & Nussinov, 2005; Kundu, Sorensen, & Phillips, 2004). In fact, fluctuation minima in slow modes are generally identified as hinge sites, and peaks in fast modes indicate kinetically hot spots responsible for protein stability. To understand the role of critical residues from dynamical network analysis and SCA in intrinsic dynamics, we carried out the fluctuation profile analysis in slow modes and fast modes. Critical residues that are found to be common in both methods (dynamic network analysis and SCA) reside either in minima or near the minima of slow/fast modes (Figure 8(B)). This indicates that these residues are likely to act as hinge sites, driving major conformational changes in HSA. We note that only two residues (Tyr341 and Val469), predicted as critical residues in dynamic network analysis of HSA and HSA-FAs, are at the peak in the fluctuation profile of high-frequency modes (Figure 8(B)) and can be considered as kinetically important residues responsible for protein stability.

Figure 8. (A) A matrix representation of statistical coupling (SCA) values for all amino acid positions of HSA family, and corresponding hierarchical clustering analysis of SCA matrix. (B) Fluctuation profile of residues (yellow star represents common critical residues from dynamical network analysis of HSA and HSA-FAs and arrow represents common residues from dynamical network analysis and SCA). Upper part shows fluctuation profile from first eight slow ed-ENM modes; lower part shows fluctuation profile from last 30 high-frequency ed-ENM modes.
Conclusion

Our network analysis captures information about highly cooperative domain motions important for the protein function (i.e. drug binding) or allosteric communications. From experimental (PCA analysis of X-ray structure ensemble) and theoretical (ed-ENM) analysis, it is clear that low-frequency modes calculated for unbound state account for the conformational changes observed upon ligand binding. This observation reveals that ligand selects the most favored conformation from accessible equilibrium structures of HSA. As small subset of low-frequency modes is able to capture the major functional changes, this information could be used as structural refinement, molecular docking, and the development of drugs that target particular functional motions. Dynamic network analysis of both states HSA and HSA-FAs shows that community network of HSA-FAs is more robust and has more information flow between different domains of the protein. Further, critical residues obtained using community network analysis are common in both states and are in good agreement with allosteric residues obtained from sequence-based SCA method. Predicted allosteric key residues can be utilized to improve the function of HSA. Our study will also facilitate efforts to modify the drugs to control their interactions with HSA, and thereby optimize their distribution across the body. The combination of structural dynamics and sequence-based bioinformatics approach, used in this study, can further be utilized to understand the molecular recognition processes in other biological systems.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2014.996609.

Acknowledgments

We gratefully acknowledge the computing resources of IISER Bhopal. One of the authors (N.A.) thank IISER Bhopal for providing the research scholarship.

References

Ahmed, A., Villinger, S., & Gohilke, H. (2010). Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses. Proteins: Structure, Function, and Bioinformatics, 78, 3341–3352.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25, 3389–3402.

Ascenzi, P., Bocedi, A., Notari, S., Menegatti, E., & Fasano, M. (2005). Heme impairs allosterically drug binding to human serum albumin sudlow’s site I. Biocell, 334, 481–486.

Ascenzi, P., & Fasano, M. (2010). Allosteric in a monomeric protein: The case of human serum albumin. Biophysical Chemistry, 148, 16–22.

Attilan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., & Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophysical Journal, 80, 505–515.

Bahar, I., Attilan, A. R., Demirel, M. C., & Erman, B. (1998). Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. Physical Review Letters, 80, 2733–2736.

Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. Chemical Reviews, 110, 1463–1497.

Bahar, I., Lezon, T. R., Yang, L. W., & Eyal, E. (2010). Global dynamics of proteins: Bridging between structure and function. Annual Review of Biophysics, 39, 23–42.

Bakan, A., & Bahar, I. (2009). The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proceedings of the National Academy of Sciences, 106, 14349–14354.

Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics, 81, 3684–3690.

Berman, H. M., Westbrook, J., Feng, Z., Gililand, G., Bhat, T. N., Weisseg, H., … Bourne, P. E. (2000). The protein data bank. Nucleic Acids Research, 28, 235–242.

Best, R. B., Lindorff-Larsen, K., DePristo, M. A., & Vendruscolo, M. (2006). Relation between native ensembles and experimental structures of proteins. Proceedings of the National Academy of Sciences, 103, 10901–10906.

Bhattacharya, A. A., Grune, T., & Curry, S. (2000). Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. Journal of Molecular Biology, 305, 721–732.

Boehr, D. D., Nussinov, R., & Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. Nature Chemical Biology, 5, 789–796.

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. Journal of Chemical Physics, 126, 014101.

Curry, S. (2009). Lessons from the crystallographic analysis of small molecule binding to human serum albumin. Drug Metabolism and Pharmacokinetics, 24, 342–357.

Curry, S., Mandelkow, H., Brick, P., & Franks, N. (1998). Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. Nature Structural Biology, 5, 827–835.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N⋅log(N) method for Ewald sums in large systems. The Journal of Chemical Physics, 98, 10089–10092.

Eargle, J., & Luthey-Schulten, Z. (2012). NetworkView: 3D display and analysis of protein. RNA interaction networks. Bioinformatics, 28, 3000–3001.

Edgar, R. C. (2004). Muscle: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32, 1792–1797.

Emekli, U., Schneider-Duhovny, D., Wolfson, H. J., Nussinov, R., & Haliloglu, T. (2008). HingeProt: Automated prediction of hinges in protein structures. Proteins-Structure Function and Bioinformatics, 70, 1219–1227.
Fanali, G., di Masi, A., Trezza, V., Marino, M., Fasano, M., & Ascenzi, P. (2012). Human serum albumin: From bench to bedside. Molecular Aspects of Medicine, 33, 209–290.

Fanali, G., Pariani, G., Ascenzi, P., & Fasano, M. (2009). Allosteric and binding properties of Asp1-Glu382 truncated recombinant human serum albumin – An optical and NMR spectroscopic investigation. FEBS Journal, 276, 2241–2250.

Frauenfelder, H., Slijer, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. Science, 254, 1598–1603.

Fujiwara, S., & Amisaki, T. (2008). Identification of high affinity fatty acid binding sites on human serum albumin by MM-PBSA method. Biophysical Journal, 94, 95–103.

Ghurman, J., Zunszain, P. A., Petitpas, I., Bhattacharya, A. A., Otgari, M., & Curry, S. (2005). Structural basis of the drug-binding specificity of human serum albumin. Journal of Molecular Biology, 353, 38–52.

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. (1997). Lincs: A linear constraint solver for molecular simulations. Journal of Computational Chemistry, 18, 1463–1472.

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. Nature, 450, 964–972.

Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. Proteins: Structure, Function, and Genetics, 33, 417–429.

Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., & Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. The Journal of Chemical Physics, 120, 9665–9678.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14, 27–38.

Koshland, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. Proceedings of the National Academy of Sciences, 44, 98–104.

Krenzel, E. S., Chen, Z., & Hamilton J. A. (2013). Correspondence of fatty acid and drug binding sites on human serum albumin: A two-dimensional nuclear magnetic resonance study. Biochemistry, 52, 1559–1567.

Kundu, S., Sorensen, D. C., & Phillips, G. N. (2004). Automatic domain decomposition of proteins by a Gaussian network model. Proteins: Structure, Function and Bioinformatics, 57, 725–733.

Lange, O. F., & Grubmuller, H. (2006). Generalized correlation for biomolecular dynamics. Proteins, 62, 1053–1061.

Lange, O. F., Lakomek, N. A., Fares, C., Schroeder, G. F., Walter, K. F., Becker, S., ... de Groot, B. L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science, 320, 1471–1475.

Lockless, S. W., & Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science, 286, 295–299.

Ma, B., Kumar, S., Tsai, C. J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. Protein Engineering Design and Selection, 12, 713–720.

Miyamoto, S., & Kollman, P. A. (1992). Settle: An analytical version of the shake and rattle algorithm for rigid water models. Journal of Computational Chemistry, 13, 952–962.

Monod, J., Wyman, J., & Changeux, J. P. (1965). On the nature of allostERIC transitions: A plausible model. Journal of Molecular Biology, 12, 88–118.

Orellana, L., Rueda, M., Ferrer-Costa, C., Lopez-Blanco, J. R., Chacon, P., & Orozco, M. (2010). Approaching elastic network models to molecular dynamics flexibility. Journal of Chemical Theory and Computation, 6, 2910–2923.

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics, 52, 7182–7190.

Peng, C., & Head-Gordon, T. (2011). The dynamical mechanism of auto-inhibition of AMP-activated protein kinase. Plos Computational Biology, 7, e1004009.

Peng, C., Zhang, L. Q., & Head-Gordon, T. (2010). Instantaneous normal modes as an unforced reaction coordinate for protein conformational transitions. Biophysical Journal, 98, 2356–2364.

Petitpas, I., Grune, T., Bhattacharya, A. A., & Curry, S. (2001). Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. Journal of Molecular Biology, 314, 955–960.

Renedo, T. D., & Grossfield, A. (2011). Validating and improving elastic network models with molecular dynamics simulations. Proteins: Structure, Function and Bioinformatics, 79, 23–34.

Rueda, M., Chacon, P., & Orozco, M. (2007). Thorough validation of protein normal mode analysis: A comparative study with essential dynamics. Structure, 15, 565–575.

Sanejouand, Y. H. (2013). Elastic network models: Theoretical and empirical foundations. Methods in Molecular Biology, 924, 601–616.

Sethi, A., Eargle, J., Black, A. A., & Luthey-Schulten, Z. (2009). Dynamical networks in tRNA:protein complexes. Proceedings of the National Academy of Sciences, 106, 6620–6625.

Simard, J. R., Zunszain, P. A., Ha, C. E., Yang, J. S., Bhagavan, N. V., Petitpas, I., ... Hamilton, J. A. (2005). Locating high-affinity fatty acid-binding sites on albumin by X-ray crystallography and NMR spectroscopy. Proceedings of the National Academy of Sciences, 102, 17958–17963.

Simard, J. R., Zunszain, P. A., Hamilton, J. A., & Curry, S. (2006). Location of high and low affinity fatty acid binding sites on human serum albumin revealed by NMR drug-competition analysis. Journal of Molecular Biology, 361, 336–351.

Skjaerven, L., Martinez, A., & Reuter, N. (2011). Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. Proteins: Structure, Function and Bioinformatics, 79, 232–243.

Smock, R. G., & Giersch, L. M. (2009). Sending signals dynamically. Science, 324, 198–203.

Smock, R. G., Rivoire, O., Russ, W. P., Swain, J. F., Leibler, S., Ranganathan, R., & Giersch, L. M. (2010). An interdomain sector mediating allostery in Hsp70 molecular chaperones. Molecular Systems Biology, 6, 414.
Sudhamalla, B., Gokara, M., Ahalawat, N., Amooru, D. G., & Subramanyam, R. (2010). Molecular dynamics simulation and binding studies of β-sitosterol with human serum albumin and its biological relevance. *The Journal of Physical Chemistry B, 114*, 9054–9062.

Suel, G. M., Lockless, S. W., Wall, M. A., & Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Structural Biology, 10*, 59–69.

Sugio, S., Kashima, A., Mochizuki, S., Noda, M., & Kobayashi, K. (1999). Crystal structure of human serum albumin at 2.5 A resolution. *Protein Engineering Design and Selection, 12*, 439–446.

Tama, F., & Brooks, C. L. (2006). Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annual Review of Biophysics and Biomolecular Structure, 35*, 115–133.

Tama, F., Gadea, F. X., Marques, O., & Sanejouand, Y. H. (2000). Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure Function and Genetics, 41*, 1–7.

Tokuriki, N., & Tawfik, D. S. (2009). Protein dynamism and evolvability. *Science, 324*, 203–207.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry, 26*, 1701–1718.

Vanwart, A. T., Eargle, J., Luthey-Schulten, Z., & Amaroo, R. E. (2012). Exploring residue component contributions to dynamical network models of allostery. *Journal of Chemical Theory and Computation, 8*, 2949–2961.

Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling, 25*, 247–260.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry, 25*, 1157–1174.

Yang, L., Song, G., Carriquiry, A., & Jernigan, R. L. (2008). Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure, 16*, 321–330.

Yang, L., Song, G., & Jernigan, R. L. (2007). How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophysical Journal, 93*, 920–929.

Yang, L., Song, G., & Jernigan, R. L. (2009). Protein elastic network models and the ranges of cooperativity. *Proceedings of the National Academy of Sciences, 106*, 12347–12352.

Zheng, W. (2008). A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations. *Biophysical Journal, 94*, 3853–3857.

Zheng, W. J., Brooks, B. R., & Thrumalai, D. (2006). Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proceedings of the National Academy of Sciences, 103*, 7664–7669.

Zunszain, P. A., Ghuman, J., Komatsu, T., Tsuchida, E., & Curry, S. (2003). Crystal structural analysis of human serum albumin complexed with hemin and fatty acid. *BMC Structural Biology, 3*, 1–9.

Zunszain, P. A., Ghuman, J., McDonagh, A. F., & Curry, S. (2008). Crystallographic analysis of human serum albumin complexed with 4Z,15E-bilirubin-IXα. *Journal of Molecular Biology, 381*, 394–406.