Current Topics

Variations in Clique and Community Patterns in Protein Structures during Allosteric Communication: Investigation of Dynamically Equilibrated Structures of Methionyl tRNA Synthetase Complexes[†]

Amit Ghosh and Saraswathi Vishveshwara*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India 560012

Received April 28, 2008; Revised Manuscript Received August 11, 2008

ABSTRACT: The allosteric concept has played a key role in understanding the biological functions of proteins. The rigidity or plasticity and the conformational population are the two important ideas invoked in explaining the allosteric effect. Although molecular insights have been gained from a large number of structures, a precise assessment of the ligand-induced conformational changes in proteins at different levels, ranging from gross topology to intricate details, remains a challenge. In this study, we have explored the conformational changes in the complexes of methionyl tRNA synthetase (MetRS) through novel network parameters such as cliques and communities, which identify the rigid regions in the protein structure networks (PSNs) constructed from the noncovalent interactions of amino acid side chains. MetRS belongs to the aminoacyl tRNA synthetase (aaRS) family that plays a crucial role in the translation of genetic code. These enzymes are modular with distinct domains from which extensive genetic, kinetic, and structural data are available, highlighting the role of interdomain communication. The network parameters evaluated here on the conformational ensembles of MetRS complexes, generated from molecular dynamics simulations, have enabled us to understand the interdomain communication in detail. Additionally, the characterization of conformational changes in terms of cliques and communities has also become possible, which had eluded conventional analyses. Furthermore, we find that most of the residues participating in cliques and communities are strikingly different from those that take part in long-range communication. The cliques and communities evaluated here for the first time on PSNs have beautifully captured the local geometries in detail within the framework of global topology. Here the allosteric effect is revealed at the residue level via identification of the important residues specific for structural rigidity and functional flexibility in MetRS. This ought to enhance our understanding of the functioning of aaRS in general.

The allosteric effect observed in multimeric enzyme systems (*1*) could potentially exist in any globular protein (*2, 3*). The models such as the "induced fit" [the "MWC" model of Monod and the "KNF" model of Koshland (*4*)] and the "population shift" of Weber (*5*) have been used for years to explain this effect. However, it is argued that the two views are extremes and that generally the actual events involve elements of both models (*6*). Furthermore, it has recently been pointed out that allostery need not involve a change in shape at the protein backbone level and that it can be explained on the basis of a trade-off between entropy and enthalpy (*7*). An insight into allosteric effect can be obtained by a quantitative estimation of the ligand-induced conformational changes at the molecular level, be it noticeable or minor, global or local, main chain topology or side chain orientation. The backbone-based secondary/supersecondary structures, which determine the gross topology of proteins, are optimally packed (*8, 9*) and are unlikely to change in many cases because of minor perturbations such as the binding of a ligand. While the root-mean-square deviations assess the gross conformational changes, large-scale motions such as domain or loop movements or hinge bending can be identified by a variety of techniques, for instance, the Gaussian network model (GNM) (*10*), the elastic network model (ENM) (*11*), and normal mode analysis (NMA) (*12*). However, it is the subtle changes leading to a cumulative effect that are difficult to identify, particularly the changes due to the side chain conformations. The enhanced communication abilities of functionally important residue pairs in proteins have been recently evaluated with the Markovian stochastic model, within the framework of the elastic network model (*13*). The rigidity and flexibility

[†] We acknowledge the computational grants for the Indian Institute of Science from the Department of Biotechnology (DBT) (PartIIB DBT infrastructure program support for basic biological research) and the Department of Science and Technology (support for Mathematical Biology, DSTO773), Government of India.

* To whom correspondence should be addressed: Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India 560012. Phone: +91-80-22932611. Fax: +91-80-23600535. E-mail: sv@mbu.iisc.ernet.in.

10.1021/bi8007559 CCC: $40.75
Published on Web 10/09/2008

of proteins were evaluated earlier in terms of degrees of freedom of vertices using distance and angle constraints on covalent bonds (*14*). Furthermore, sequence-based approaches have been used to measure energetic coupling between positions on a multiple-sequence alignment (MSA) and infer correlations between amino acids in protein families (*15, 16*). In our recent studies (*17, 18*), we have shown that the protein structure network (PSN)[1] perspective is highly illuminating in capturing the global changes, yet not sacrificing the details of the side chain orientations at the local level. For instance, the crystal structures of the complexes of glutaminyl tRNA synthetase were found to differ in their network parameters like hubs, the size of the largest cluster (*19*), and the communication paths between distal residues in methionyl tRNA synthetase (MetRS) were identified from the PSN analysis on the conformational ensembles generated from MD simulations (*18*).

Aminoacyl tRNA synthetases (aaRSs) are excellent models for understanding long-range communication. The aaRSs accurately pick up the cognate amino acid in the active site and cognate tRNA in the anticodon binding site. Both aaRS and tRNA undergo conformational changes in recognizing the cognate amino acid and then transferring the charged amino acid to the acceptor stem of tRNA. However, it has been pointed out (*20*) that the communication takes place through the protein and not through tRNA. A fundamental problem in allosteric enzymes, particularly in aaRS, is understanding the mechanism by which the information of the binding event is transmitted from the tRNA binding site to the catalytic site. Domain−domain interaction plays a crucial role in allosteric communication in aaRS as observed by kinetic and mutational experiments and supported by structural data (*21*−*27*). We have selected MetRS as a representative of aaRS in this study. Here we have demonstrated that the allosteric effect can be investigated at the molecular level by studying the network parameters, cliques (completely connected subgraph), communities (connected cliques) (*28*), and hubs (highly connected nodes), which represent the rigid regions, in the structures of MetRS complexes. These parameters, which are evaluated on the basis of noncovalent interactions in protein structures, identify the relatively high-packing density regions. The variations of clique communities of different liganded systems in MetRS have provided new insights into allosteric functions of aaRS such as the recognition of cognate amino acids and communication between distant sites.

The relevant methodology is described in the next section. In the Results, we present a brief analysis of simulations and a detailed analysis of network parameters, which highlight the ligand-induced changes in rigid and flexible regions in the protein structure. Finally, the insight brought out by the protein structure network on allostery is discussed in the last section.

METHODS

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were performed at 300 K using AMBER9 (*29*) with parm99 (*30*) parameters on five systems of *Escherichia coli* MetRS. These five systems correspond to (A) the unbound MetRS (PDB entry 1QQT) (*31*), (B) MetRS complexed with methionine (PDB entry 1F4L) (*32*), (C) MetRS complexed with ATP (*33*), (D) MetRS complexed with adenylated methionine (MetAMP) (PDB entry 1PFY) (*34*), and (E) the MetRS−$tRNA_f^{Met}$−MetAMP complex, which was modeled (*18*). All the simulations were carried out in aqueous medium for 10 ns, using the TIP3P water model (*35*). The solvation box was 10 Å from the farthest atom along any axis, which resulted in 22145, 21764, 21683, 22259, and 29883 water molecules for systems A−E, respectively. The simulations were performed under NPT conditions. The van der Waals cutoff was set to 10 Å, while the pressure and temperature relaxations were set at 0.5 ps^{-1}. The first 500 ps of the simulations was considered the equilibration phase, and the coordinate sets (snapshots) from the remaining 9.5 ns were used for analysis. A time step of 2 fs was employed with the integration algorithm, and the structures were stored every 1 ps.

Construction of Protein Structure Networks (PSNs). The strength of noncovalent interactions in proteins was evaluated on the basis of the number of interacting atoms between two residues (*36*). Each amino acid in the protein structure was considered as a node in the graph, and an edge was constructed on the basis of the following equation

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} \times 100$$

where I_{ij} is the strength of the interaction between residues i and j, n_{ij} is the number of distinct atom pairs between residues i and j coming within a distance of 4.5 Å, and N_i and N_j are the normalization values for residues i and j, respectively, obtained from a statistically significant data set of proteins based on the maximum interaction the residue is capable of making. We define a cutoff value I_{min}, and when $I_{ij} > I_{min}$, we connect nodes i and j with an edge.

An analysis of a large number of protein structures (*37*) had indicated that the protein structures exhibit an optimal interaction strength ($I_{critical}$), at which point the size of the largest noncovalently connected cluster undergoes a transition. The I_{min} value at which the size of the largest cluster decreases drastically is termed $I_{critical}$. This critical interaction strength ($I_{critical}$) was used to study the network dynamics of lysozyme structure, during MD simulation (*17*). In this study also, we have carried out the analysis of the size of the largest cluster as a function of interaction strength (I_{min}) in the MD-simulated snapshots of MetRS and its complexes. We observe that the critical interaction strength to be around an I_{min} of 3.0% (results not shown). On the basis of this study, any pair of residues (i and j) was considered to be connected, if I_{min} was greater than 3.0%.

k-Clique Community in Protein Structure Networks. The PSNs were constructed as mentioned above for all the snapshots of the MD-simulated structures. The *k*-clique communities were analyzed from these PSNs on the basis of definitions given below.

(i) Definition of the k-Clique Community. A *k*-clique (complete subgraph of *k* nodes) is defined as a set of "*k*" nodes in which all nodes are connected to all other nodes in the set. A community is defined as a union of smaller complete subgraphs that share nodes. In the mathematical literature, a *k*-clique community is defined as the union of

[1] Abbreviations: aaRS, aminoacyl tRNA synthetase; MetAMP, methionyl adenosine monophosphate; PSN, protein structure network.

k-cliques that can be reached from each other through a series of adjacent *k*-cliques. Two *k*-cliques are considered to be connected in a community, if they share *k*-1 nodes (*28*). In this study, we have employed a variant of community definition, in which we have defined a *k*-clique community if two *k*-cliques share *k*-1 or *k*-2 nodes. A community can have two or more of cliques, and the size is given by the number of constituent cliques. The increase in community size can be considered to be proportional to the increase in packing density in proteins.

(ii) k-Clique Community Finding Algorithm. Our community search approach is based on the algorithm proposed by Palla et al. (*28*). We have used CFinder (*38*) to obtain the *k*-clique community from PSNs. A minimum of three nodes is required to make the $k(k = 3)$ clique. In fact, in almost no cases did we obtain cliques greater than 3 (only a few exceptions of $k = 4$ cliques), with the criterion that was used for edge formation ($I_{min} = 3.0\%$). In CFinder, all *k*-cliques are first extracted. Second, the clique−clique overlap matrix is prepared to generate the *k*-clique communities with an overlap of *k*-1 nodes. The communities with *k*-2 node overlaps were extracted by manual inspection of cliques and communities obtained with CFinder.

Hubs. At a given I_{min}, different nodes make different numbers of edges. The residues that make four or more edges are termed hubs.

Communication Paths. The allosteric effect is mediated by communication between distal residues in proteins. In the case of MetRS, the two distal sites, the anticodon binding site and the activation site, are separated by ~70 Å. The identification of noncovalently connected residues along the shortest path between the two end members, from the PSNs of MD snapshots, has been described previously (*18*). Here we compare the residues along the communication path with those found in cliques and communities.

Dynamically Stable Cliques, Communities, and Hubs. The network parameters, cliques, communities, and hubs were evaluated from all the snapshots of the MD simulations in all five systems. They were considered dynamically stable if present in more than 50% of the simulation snapshots. These dynamically stable quantities represent the major conformational populations in the MD ensemble. The two-dimensional graphical representations of dynamically stable cliques and communities were drawn using CFinder (*38*), and the schematic three-dimensional representations were drawn with VMD (*39*).

Euler Angle Analysis. The mutual orientation of the catalytic and anticodon domains was evaluated for the conformations of the simulation. First, the moment of inertia axis of the reference structure (minimized crystal structure) was aligned along Cartesian axes using Orient package available in VMD (*39*). A similar transformation was carried out for all the snapshots along the MD simulation trajectory. Then the catalytic (residues 3−349) and anticodon domains (residues 350−550) of all these MD snapshots were separately superimposed on the respective domains of the reference structure, using the alignment program of VMD. The difference in rotations is measured after a separate optimal alignment of the anticodon domain and the catalytic domain. This difference in rotation is used to calculate the relative rotation of the anticodon domain to the catalytic domain in terms of Euler angles along the $x(\theta)$, $y(\phi)$, and $z(\varphi)$ axes. Euler angles are defined in several ways, and we have adopted the method used in ref *40*. The relative rotations of the domains in systems D and E are presented in panels a and b of Figure 2, respectively.

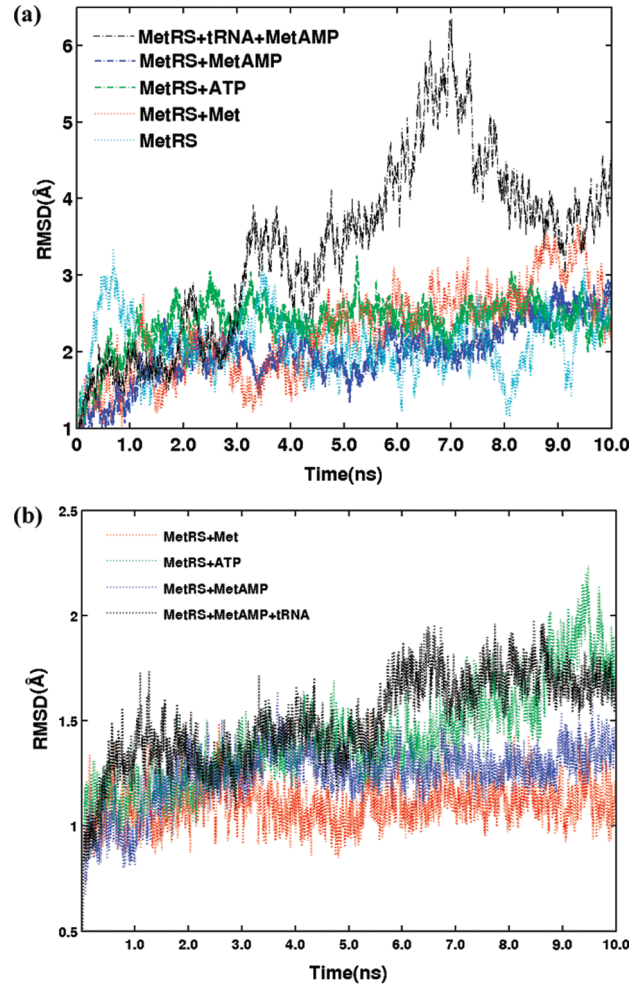

FIGURE 1: Root-mean-square deviation (rmsd) trajectories of five systems of MetRS [(A) MetRS, (B) MetRS with Met, (C) MetRS with ATP, (D) MetRS with MetAMP, and (E) MetRS−MetAMP−tRNA complex] with reference to the minimized crystal structure: (a) superimposed with the whole protein and (b) superimposed with catalytic and anticodon binding domains.

Residue Conservation across Species. Sequence alignment of MetRS of different species was carried out using ClustalW (*41*). The results are shown in the Supporting Information. Conserved residues participating in the top three largest communities and in communication paths are highlighted in different colors.

RESULTS

The trajectories of the conventional parameters such as the root-mean-square deviation (rmsd) and the Euler angle between the anticodon domain and the catalytic domain are presented in the first part of this section. The details of the network parameters such as cliques and communities, hubs, and communication paths obtained from the PSN analysis of MD snapshots are presented in the second part of this section. Furthermore, detailed comparison of the conserved residues participating in cliques, communities, and hubs with those involved in communication paths is also presented.

Simulation Trajectories. MD simulations were carried out on different ligand-bound complexes of MetRS. The five

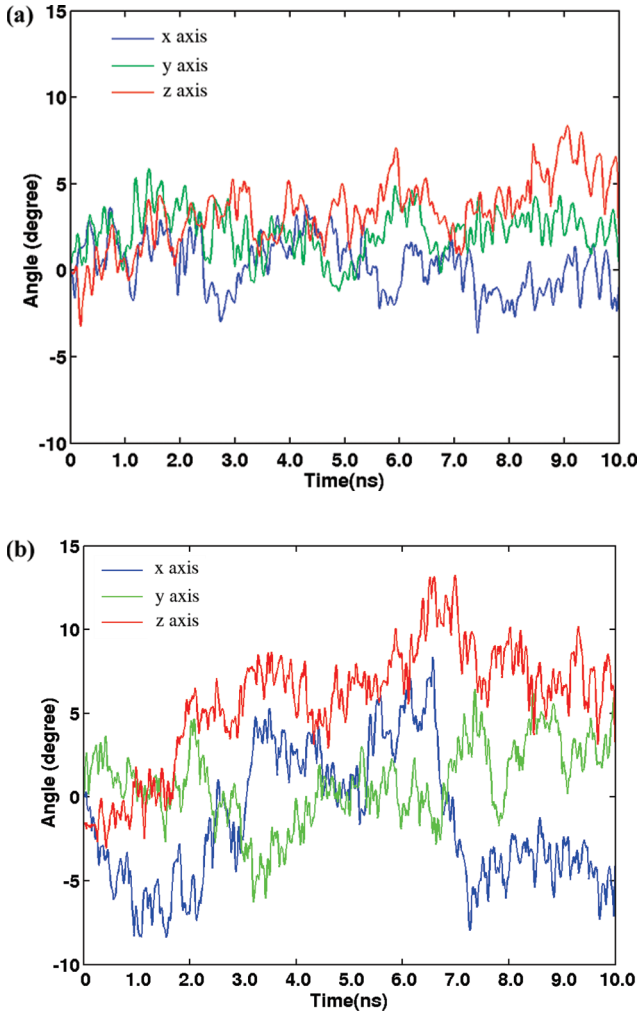

FIGURE 2: From Eulerian angle analysis, the measured rotations of the anticodon domain with respect to the catalytic domain are shown along the MD trajectories in (a) system D and (b) system E.

systems [(A) the unbound MetRS (1QQT), (B) MetRS complexed with Met (1F4L), (C) MetRS complexed with ATP, (D) MetRS complexed with MetAMP (1PFY), and (E) the MetRS−$tRNA_f^{Met}$−MetAMP complex] are analyzed as follows.

(i) Root-Mean-Square Deviation Profiles. The rmsd trajectories of the five systems with respect to their minimized starting structures are shown in Figure 1a. The C_α rmsd values are within 3.5Å with respect to the starting conformation throughout the simulation for systems A−D, and it reaches a maximum value of 6.4 Å around 7 ns in system E. However, this increase is essentially due to the movement of the CP domain which can be seen in Figure 1b. Thus, the conformation of the protein at the C_α backbone level has not undergone any drastic change in the presence of ligand, except for the movement of the CP domain, which results in the opening of the active site pocket. The similarities in MetRS conformations (without the CP domain) in different liganded states can also be seen by comparing the MD average structures. The rmsd values of the pairs of averaged structures presented in Table 1 indicate that the maximum deviation is only 1.25 Å.

(ii) The Anticodon Binding Domain Rotates Relative to the Catalytic Domain. In the absence of any crystal structure of the *E. coli* MetRS−$tRNA_f^{Met}$ complex (system E), it was

Table 1: Pairwise C_α Root-Mean-Square Deviations (angstroms) between Average Structures (averaged over the trajectory) of Different Liganded Systems[a]

	A	B	C	D	E
A	0.00	0.71	1.08	0.9	0.99
B	0.71	0.00	1.1	1.02	1.06
C	1.08	1.11	0.00	1.14	1.21
D	0.9	1.02	1.14	0.00	1.25
E	0.99	1.06	1.21	1.25	0.00

[a] A, MetRS; B, MetRS with Met; C, MetRS with ATP; D, MetRS with MetAMP; and E, MetRS−MetAMP−$tRNA_f^{Met}$ complex.

modeled (*18*) by superimposing the tRNA conformation taken from the complex structure of the *Aquifex aeolicus* MetRS−$tRNA_m^{Met}$ complex (*42*) on the ligand-free conformation of *E. coli* MetRS. However, the anticodon domain and the catalytic domain in the ligand-free *A. aeolicus* MetRS were found to be rotated by ∼15° when compared to the tRNA-bound protein (*42*). Here we monitor the mutual orientation of the two domains to see if such an induced conformational change is also observed in the *E. coli* system. Furthermore, we have evaluated the flexibility in the mutual orientation between the two domains, by Euler angle analysis as described in Methods. The anticodon binding domain of MetRS rotates about *x*, *y*, and *z* axes relative to the catalytic domain (Rossmann fold and connecting polypeptide). The relative rotations evaluated along the trajectory of system D, with respect to the starting conformation, are plotted in Figure 2a. A similar plot for system E is shown in Figure 2b. The values fluctuate in both the systems along the trajectories. However, in general, the maximum deviation and the fluctuations are higher in system E than in system D. Thus, tRNA binding has provided extra flexibility to the MetRS for ligand recognition facilitated by long-range communication.

Network Parameters. The trajectories generated from all the five simulations of MetRS are stored, and the structure networks obtained from each snapshot are analyzed in an effort to understand the dynamical properties of the network. Cliques, communities, and hubs represent the rigid regions of PSNs, and the dynamically stable ones correspond to the major conformation in the MD ensemble. The analysis of these parameters from the simulations of five systems is presented below. The residues in the communication path are more flexible, and the major communication paths from the simulations were described previously (*18*). Here we also investigate if the residues along the communication paths are part of cliques and communities.

(i) Cliques and Communities near the Active Site and the Domain Interface Region of MetRS. All the dynamically stable cliques and communities were evaluated for all five systems (the detailed results are consolidated in Table 1 of the Supporting Information). Below we discuss some of the biologically relevant cases such as those which reorganize in response to the ligand and those that are involved in domain−domain interaction. Specifically, we focus on three regions: (1) the catalytic domain (Rossmann fold), (2) the interface domain (KMSKS domain which connects the catalytic site and the anticodon domain), and (3) the CP-R region (the region where the CP domain connects the two halves of the Rossmann fold). The cliques and communities in these regions detected in all five systems are pictorially represented in Figure 3 in blue (catalytic region), magenta

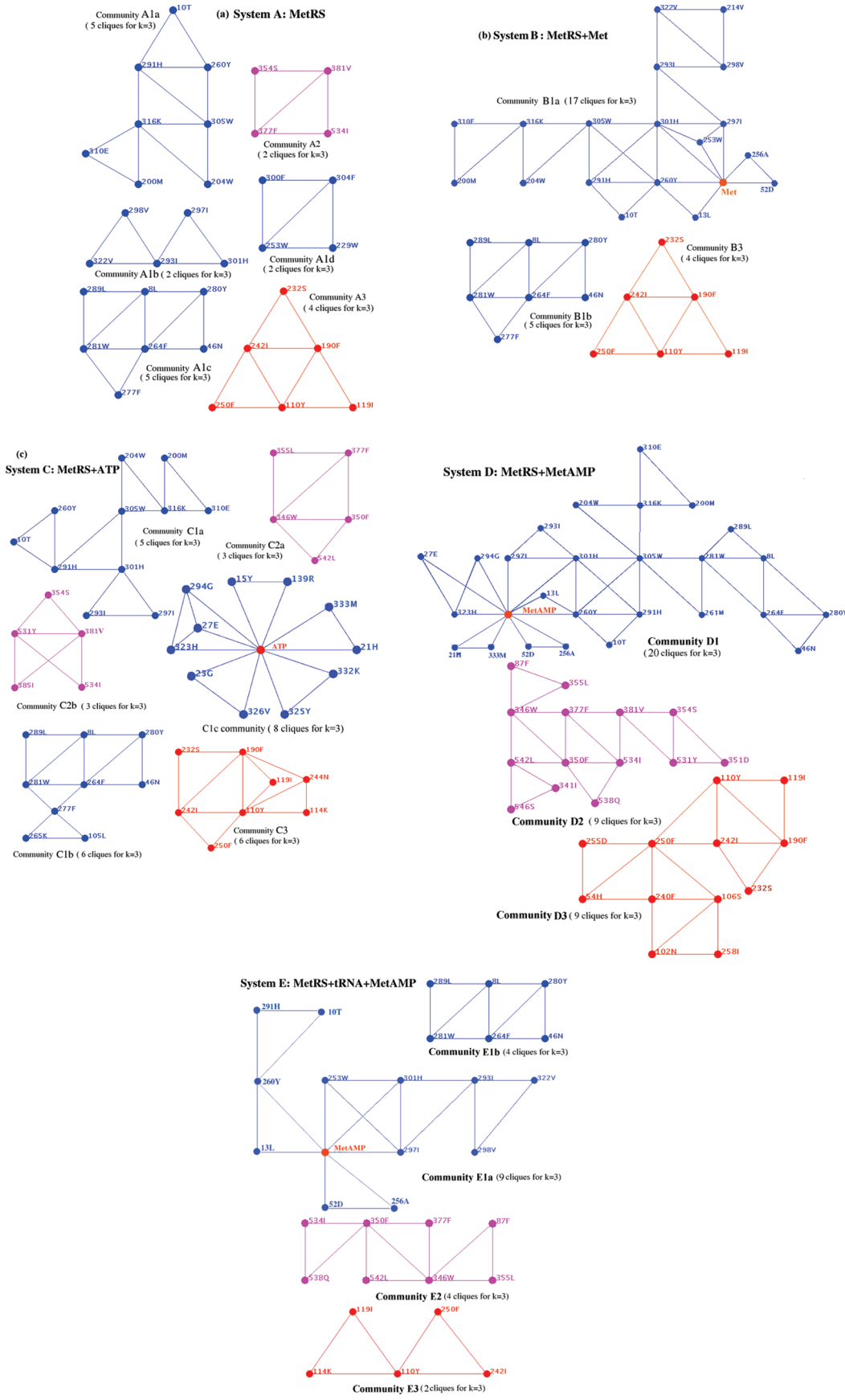

FIGURE 3: Schematic representation of communities. The residue connection details of the three largest communities in system D and their variants in other systems are shown. The communities in the catalytic region, the KMSKS interface domain region, and the CP-R region connecting the CP domain and the Rossmann fold are colored blue, magenta, and red, respectively (details are also given in Table 1 of the Supporting Information). The ligands Met and MetAMP in the different communities are colored red. The communities in systems A−E are represented in panels a−e, respectively. For system E, the community structure for the first part of the simulation (EI) is shown. Figures 3 and 6 were drawn with VMD (*38*).

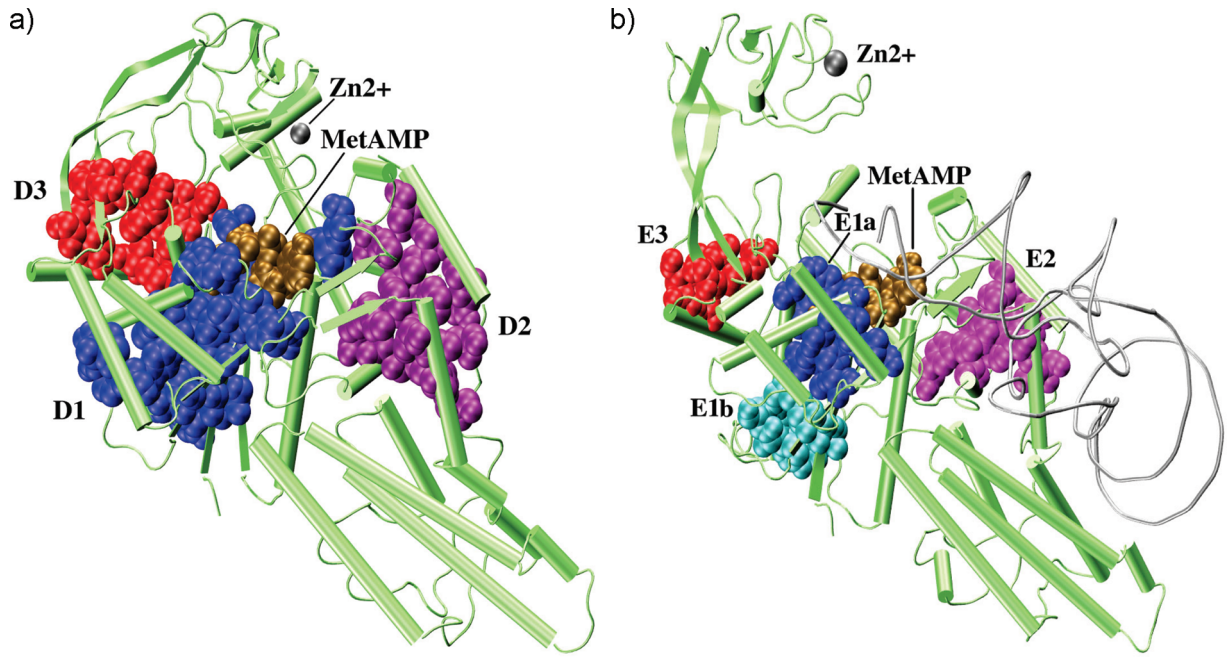

FIGURE 4: Pictorial three-dimensional representation of communities. Figures 4 and 6 were drawn with VMD (*39*). In both these figures, MetRS, MetAMP, and tRNA are shown in green cartoon, ochre van der Waals, and black tube representations, respectively. In this figure, the three largest communities in MetRS [located in the catalytic region (blue), in the KMSKS domain interface region (magenta), and in the CP-R region, which connects the CP domain and the two halves of the Rossmann fold (red)] are shown in van der Waals representation. (a) The sizes of these communities (D1, D2, and D3) are the largest in system D (MetAMP bound to MetRS) (see Figure 3d for residue details). (b) Communities E1, E2, and E3 in system E (MetRS complexed with tRNA and MetAMP) correspond to D1, D2, and D3, respectively, of system D. Community E1 is composed of three separate smaller communities, E1a and E2b (colored blue and cyan, and the residue details are given in Figure 3e). Further, E2 and E3 are smaller than D2 and D3.

(KMSKS, interface region), and red (CP-R region). The ligands in the catalytic region were also considered as nodes in the evaluation of cliques. It is to be noted that the computed cliques consist of three residues ($k = 3$), and the higher-order ones were not obtained [with only a few exceptions in systems B and D near the active site, which have a $k = 4$ clique (clique of residues 10T, 260Y, 301H, 305W, and 291H present in systems B and D and of residues, for MetAMP or Met, 253W, 301H, and 297I present in systems B and E)] for the chosen condition of edge formation.

(1) MetRS (system A). Four $k = 3$ clique communities (A1a, A1b, A1c, and A1d with cliques 5, 2, 5, and 2, respectively) are found near the catalytic region in the ligand-free enzyme as shown in Figure 3a. A small four-residue community, A2, of two cliques (354S, 381V, 377F, and 534I) is present near the KMSKS domain. Relatively strong community A3 with four cliques is present near the CP-R region. It should be noted that several small cliques and communities are present in the catalytic domain, which merge to become larger communities in other ligand-bound systems (which are described in subsequent paragraphs). Such a disjointed community pattern is an indication of flexibility at the active site of the native enzyme, which is required to respond to the approaching ligands.

(2) MetRS with Met (system B). Of four *k*-clique communities present in the catalytic domain of system A, three of them (A1a, A1b, and A1d) have merged with some reorganization to form a single community, B1a, of 17 cliques in system B (community B1b, however, is the same as A1c) (Figure 3b). Thus, one big community is formed near the catalytic region of MetRS, which is responsible for the recognition of the cognate amino acid, methionine. While the rigidity has increased in the catalytic domain with the binding of methionine, it is interesting to see that it has increased the flexibility in the KMSKS domain region which is evident in the absence of community A2. Such a flexibility can facilitate the binding of ATP in this region.

(3) MetRS with ATP (system C). The presence of ATP in the active site has also increased the connectivity of cliques and communities in the catalytic region, when compared with that of the native enzyme. As shown in Figure 3c, communities C1a (consists of five cliques) and C1b (consists of six cliques) represent this reorganization, and a new community, C1c (consists of eight cliques), is formed along with the ligand ATP. Interestingly, we see a more pronounced change in the KMSKS domain region. The community which was present in this region in the free enzyme (system A) and was lost in the complex bound to methionine (system B) has now been substantially strengthened (communities C2a and C2b). As we see later, this is strengthened further in system D, which has MetAMP as the ligand. There is a marginal increase in the size of the community (C3) in the CP-R region.

(4) MetRS with MetAMP (system D). A dramatic transformation in the community configuration is seen in this system (Figure 3d). All the smaller communities which were found in the other systems have merged to form three large communities, D1, D2, and D3, being in the catalytic, KMSKS, and CP-R regions, respectively. This is also pictorially represented in Figure 4a. Thus, the activated methionine (MetAMP) is now surrounded by three large and dense communities. In this context, it may be noted that the binding free energy is minimal for the MetAMP ligand versus several other ligands (*43*). The formation of highly modular

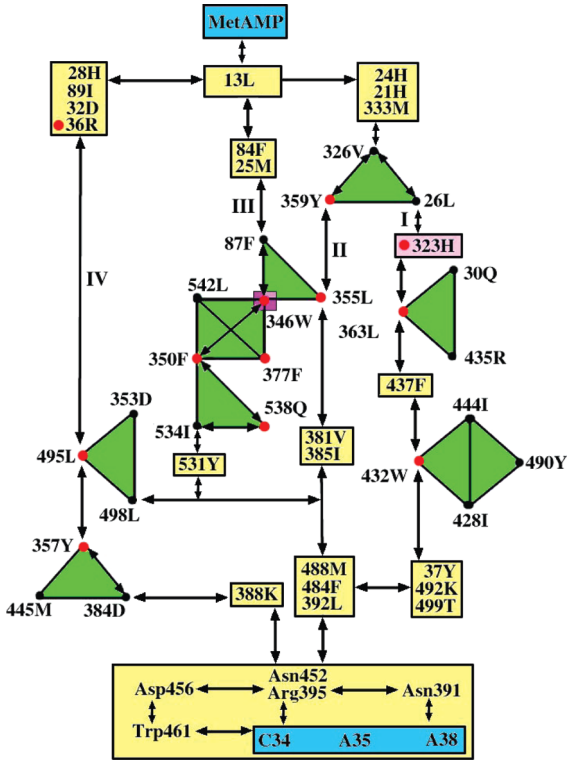

FIGURE 5: Schematic representation of four (I, II, III, and IV) communication paths between the ligand MetAMP and the anticodon bases (both colored blue) in MetRS, as obtained from system E. The residues along the paths that are not part of cliques are shown in yellow rectangular boxes. Green triangles represent cliques or communities from which one or more residues take part in communication. The residue connections along the paths are shown with arrows. Communicating residues, which lie in the interface region, are shown as red circles. The hub residues (at $I_{min} = 5.0\%$) are colored pink. Paths I, II, and III pass through communities and hubs. The most frequently accessed path, path IV, on the other hand does not encounter strong hubs or communities (only three clique residues, 384D, 357Y, and 495L, are part of the path).

structure in MetRS could be the reason for the energetically stable conformation, which may prevent the hydrolysis of MetAMP.

(5) MetRS−MetAMP−tRNA$_f^{Met}$ Complex (system E). On the basis of the rmsd profile (Figure 1b) of system E, the trajectory analysis is broken into two regions (EI, 0.5−5.5 ns; and EII, 5.5−10 ns). The clique−community architecture is also different in these two regions (which can be seen in Table 1 of the Supporting Information). A pictorial representation of the cliques and communities obtained from region EI is shown in Figure 3e. A schematic representation is also shown in Figure 4b. Perhaps the two populations corresponding to regions EI and EII represent the alternate conformations adopted by the system before and after the attachment of MetAMP to the acceptor stem of tRNA. In contrast to system D, the overall number of cliques formed in the EI region of system E is fewer compared to the number in systems B−D, and it has further diminished during the second part of the simulation (EII). The attachment of completely negatively charged tRNA molecule on the surface of MetRS has brought more flexibility to the structure via a reduced size of cliques and communities. This aspect will be further elaborated in relation to the communication paths.

(ii) Hubs. Hubs are highly connected residues (more than three connections), and they also impart rigidity to the structure. We identified the dynamically stable hubs at the interaction strength ($I_{min} = 3.0\%$), a criterion that was chosen for the clique evaluation. This resulted in a large number of hubs (Table 2 of the Supporting Information), and hence, we considered a more stringent criterion of interaction strength ($I_{min} = 5.0\%$). The resultant hubs in all the simulations are listed in Table 2. Whether the hub residue is a part of a clique is also indicated in this table. As one can see in the cliques and communities, the largest number of hubs is present in system D. Furthermore, a large fraction of them are also part of cliques. The smallest number of hubs is seen in system E (EI and EII). An intermediate number between systems D and E is taken in the ligand-free system (A) and in systems B and C. Thus, the hubs also capture the variation in rigidity in response to different ligands.

Table 2: List[a] of Dynamically Stable Hubs[b] ($I_{min} = 5\%$) in Systems A−E (EI and EII)

residue	A	B	C	D	EI	EII
30Q		√ √		√		√
32D	√	√	√	√		
36R			√	√		
54H						
84F		√		√		
91Y		√		√	√	√
93N	√	√ √	√	√ √	√	√
95H		√	√	√ √		
110Y	√ √	√	√ √	√	√	√
129D		√				
189H	√		√ √	√ √		
233R			√ √			
234D	√ √		√	√ √		
240F			√ √	√		
260Y	√ √	√ √		√ √	√	√
264F						√
266N	√	√	√ √	√ √	√ √	√
280Y				√		
281W		√		√ √		
291H	√	√ √	√	√ √	√ √	√ √
292F		√		√		
300F	√ √					
304F	√ √					
305W	√ √	√ √	√ √			
323H						√
346W	√	√	√ √	√ √	√	
357Y	√		√ √			
384D	√		√ √			
495L			√			

[a] All those residues which appear as hubs in at least one of the simulations are listed. A check in the first column of a given system (e.g., A or B) denotes that the residue is a hub in that simulation. A check in the second column indicates that the residue is part of a clique. [b] Fifty percent of the snapshots in which the listed residues appear as hubs.

(iii) Cliques, Communities, Hubs, and the Communication Paths. The residues involved in communication paths between the catalytic site and the anticodon binding site were presented in our earlier studies (*18*). Here we have examined the connection between the residues along the paths and the residues, which take part in cliques, communities, and hubs. The consolidated results as obtained in system E are schematically shown in Figure 5. The residues in the three largest communities in systems D and E are examined for their participation in communication, and the results are listed in Table 3 of the Supporting Information. We see a total lack of communication residues in the communities of the catalytic and CP-R regions. Interestingly, we find that many of the residues involved in communication in the KMSKS

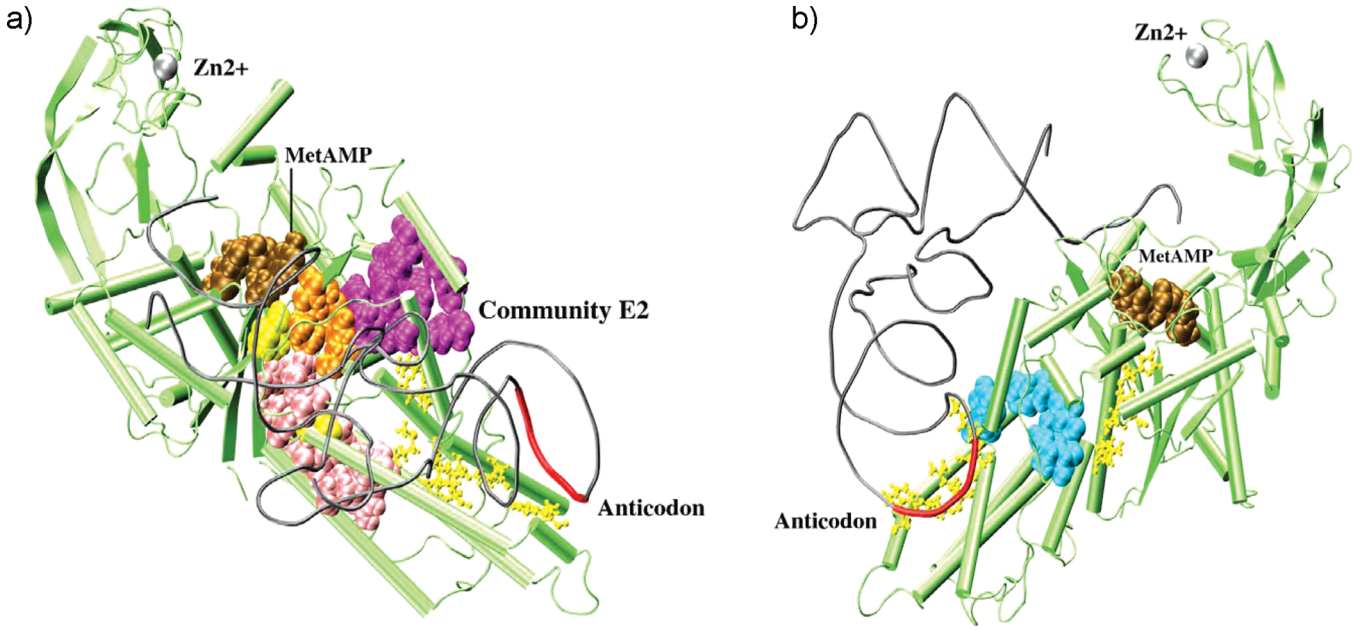

FIGURE 6: Pictorial representation of cliques and communities along the shortest paths of communication shown in Figure 5. Cliques and communities are shown in van der Waals representation, and the connecting residues along the communication paths are shown in yellow ball-and-stick representation. The anticodon bases of tRNA are colored red. Paths I, II, and III are shown in panel a. Here paths II and III pass through one of the top three largest communities (the KMSKS domain region, colored magenta). Path I passes through a small community and a clique (colored pink). Paths I and II also pass through another clique (orange). Path IV is shown in panel b. Here the communication path encounters only two cliques (colored cyan).

domain region are part of cliques or communities [represented by community E2 (Figures 3e and 4b) in system E]. On the other hand, the residues closer to either the anticodon binding region or the catalytic site are not part of cliques which can be seen in Figure 5. Many of the hub residues also come from the interface regions and are part of cliques. Furthermore, the least frequent paths of communication (*18*), paths II and III, pass through a large community, which incorporates hubs like residue 346W. Similarly, another less frequent path, path I, also passes through interface cliques and a hub (323H). The location of mediating residues at the intermodule boundaries, imparting structural rigidity, has also been noticed in other proteins (*44*). Recently, MD simulations have shown the importance of residues near the interface of two domains in the stability of the ligand-bound complex (*45*). The participation of interface residues in cliques, communities, and hubs assures the rigidity of the interdomain regions and explains many of the experimental observations, which have emphasized the intactness of domain−domain interactions (*21*−*23*). In contrast to paths I, II, and III, path IV is distinctly different. It was found to exist only in system E, in which MetRS is bound to both MetAMP and the tRNA, and it was shown to be the most frequently accessed path (*18*). Amazingly, we find that path IV does not encounter any strong hubs and passes through two separate cliques, suggesting more flexibility and adaptability in response to different ligands. A schematic representation of paths I, II, and III and path IV along with the cliques and communities presented in panels a and b of Figure 6, respectively, clearly demonstrates that path IV is made of more flexible residues than the other three paths.

DISCUSSION

The results presented above clearly demonstrate that the network parameters describing the noncovalent interaction of side chains are able to discriminate subtle differences in the conformations in MetRS taking place in response to different ligands. These differences are captured from the simulated snapshots and hence represent the major population in equilibrium. Five simulations were carried out on different liganded states of MetRS to generate the conformational ensembles. The conformational changes in different states were found to be minimal at the backbone topological level. However, the changes at the side chain interaction level have been clearly captured by the network parameters such as cliques, communities, and hubs. Highlights of these analyses are as follows. (1) The number of cliques is dependent on the bound ligand. (2) Some of the cliques are invariant across different complexes. (3) The maximum number of cliques is found in system D in which the MetAMP is surrounded by three large communities (represented as communities D1, D2, and D3 in Figures 3d and 4a). These are present in the catalytic site region, in the KMSKS domain (which connects the catalytic domain and the anticodon domain), and in the region connecting the CP domain and the Rossmann fold. (4) In system E, binding of tRNA to the MetRS−MetAMP complex has shown a major reduction in community size, with a reduced number of cliques (Figures 3e and 4b).

As discussed in the introductory section, there are several aspects related to allosteric mechanism. One component is the identification of conformational changes not only at the gross level but also at the detailed level of side chain interactions. This work has shown that the cliques and communities are able to effectively capture these details. Another important component is the shift in the equilibrium conformation brought about by ligands. It has been pointed out that the allosteric effect is driven by the enthalpy of binding, conformational entropy, or both (*7*). The entropic effect is associated with the changes in the conformational equilibrium, and the evaluation of this contribution requires

a rigorous assessment of conformational changes. This study has effectively captured such a shift in population through the dynamically stable network parameters. This shift which has taken place at the side chain interaction level would not have been able to be identified by conventional parameters such as the backbone rmsd. Additionally, cooperativity is another factor that plays a role in allosteric communication. The changes in cooperativity between residues or subunits have been measured by experimental methods (*46*) and simulations (*47*), and such an effect in MetRS has also been monitored through correlation coefficients (*18*) between residues during simulations.

There are diverse ways in which the allostery can manifest in influencing the binding of the second ligand to proteins. It is important to identify the key residues that take part in the allosteric effect in specific cases to understand the functioning of the chosen protein. As we discussed earlier, some of the residues are involved in maintaining the rigidity of the interdomain region, while some others are needed to communicate the signal from one region to the other. Network representation of protein structures has provided ways of identifying the key residues in allosteric communication (*48*). In this network analysis, the important residues involved in side chain interactions have been identified from the dynamic ensembles. Such an analysis has provided information about important residues in MetRS, which is discussed below.

A significant number of residues [including the ones conserved in sequence across species (Figure 1 of the Supporting Information)] participating in cliques and in communication in MetRS are conserved (40% of the total conservation), and some of them are common to both. While the aromatic residues and isoleucine dominate the cliques, polar and charged residues participate largely in communication. The invariant and the induced cliques and hubs have also been analyzed at the individual residue level. For instance, a set of hub residues (260Y, 291H, 293I, 301H, and 110Y), mainly from the Rossmann fold at the catalytic region, is an invariant component of cliques in all the complexes. The domain−domain communication is induced by the presence of the nucleotide moiety at the catalytic site region in systems C, D, and E, through the formation of a new clique and community, which involves the hub residues (e.g., 346W, 350F, and 377F) from the KMSKS domain region (Figure 3). Yet another set of residues (both hubs and nonhubs) (13L, 21M, 27E, 52D, 256A, 281W, 294G, 305W, 323H, 333M, 280Y, 258I; 240F, 264F, 8L, 316K, 381V; 261M, 531Y, 341I, 546S) responds to ligands in their preference to be part of cliques. Several residues involved in communication are conserved (Figure 1 of the Supporting Information), including a few of them (377F and 531Y) which participate in cliques and also those from functionally important HIGH and KMSKS motifs (21H, 24H, and 333M). While the participation of completely conserved residues is discussed here, it will be interesting to investigate if the other residues in cliques and communication paths have undergone correlated mutations during the evolution (*15*). The biological significance of some of the residues (301H and 305W involved in cliques, residues 21H, 24H, and 333M involved in communication), some of which do not directly interact with ligands, has been experimentally confirmed (*49*−*51*), and this study has unveiled the specific role played by these residues. Further, we also predict that residues like 346W and 350F are crucial for the functioning of MetRS, which may be verified by experiments.

In summary, the rigid regions in ensembles of conformations of different liganded states of MetRS have been evaluated through the network parameters of cliques and communities. The binding of ligands induces changes in its vicinity. Interestingly, the ligands also induce changes in the regions which do not directly interact with them. In the MetAMP-bound system, the active site becomes very rigid with the formation of very large community D1 and thus recognizes the cognate amino acid through an extended network of side chain interactions. This study has elucidated the essence of allostery at the gross topological level while retaining the details at the individual residue level. Investigation of other proteins by this approach should show the diversity in the allosteric mechanism and may prove to be useful in identifying residues important in communication and in maintaining the rigidity required for the function. In addition, the characterization of specific roles of different amino acid residues in the protein structure network may assist in the rational design of new experiments.

ACKNOWLEDGMENT

We are grateful for access to the computational facilities at the Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore.

SUPPORTING INFORMATION AVAILABLE

A consolidated list of dynamically stable $k = 3$ cliques and communities (Table 1), a list of dynamically stable hubs (Table 2), residues in the top three communities in systems D and E (Table 3), and sequence alignment of MetRSs. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: A plausible model. *J. Mol. Biol. 12*, 88–118.
2. Gunasekaran, K., Ma, B., and Nussinov, R. (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins: Struct., Funct., Bioinf. 57*, 433–443.
3. Frauenfelder, H., McMahon, B. H., and Fenimore, P. W. (2003) Myoglobin: The hydrogen atom of biology and a paradigm of complexity. *Proc. Natl. Acad. Sci. U.S.A. 100*, 8615–8617.
4. Koshland, D. E., Nemethy, G., and Filmer, D. (1966) Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry 5*, 365–385.
5. Weber, G. (1972) Ligand binding and internal equilibiums in proteins. *Biochemistry 11*, 864–878.
6. Jardetzky, O. (1996) Protein dynamics and conformational transitions in allosteric proteins. *Prog. Biophys. Mol. Biol. 65*, 171–219.
7. Tsai, C.-J., del Sol, A., and Nussinov, R. (2008) Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *J. Mol. Biol. 378*, 1–11.
8. Przytycka, T., Aurora, R., and Rose, G. D. (1999) A protein taxonomy based on secondary structure. *Nat. Struct. Mol. Biol. 6*, 672–682.
9. Banavar, J. R., and Maritan, A. (2007) Physics of Proteins. *Annu. Rev. Biophys. Biomol. Struct. 36*, 261–280.
10. Bahar, I., Atilgan, A. R., and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des. 2*, 173–181.

11. Yang, L., Song, G., and Jernigan, R. L. (2007) How Well Can We Understand Large-Scale Protein Motions Using Normal Modes of Elastic Network Models? *Biophys. J. 93*, 920–929.
12. Brooks, B., and Karplus, M. (1983) Harmonic Dynamics of Proteins: Normal Modes and Fluctuations in Bovine Pancreatic Trypsin Inhibitor. *Proc. Natl. Acad. Sci. U.S.A. 80*, 6571–6575.
13. Chennubhotla, C., and Bahar, I. (2007) Signal Propagation in Proteins and Relation to Equilibrium Fluctuations. *PLoS Comput. Biol. 3*, e172.
14. Jacobs, D. J., Rader, A. J., Kuhn, L. A., and Thorpe, M. F. (2001) Protein flexibility predictions using graph theory. *Proteins: Struct., Funct., Genet. 44*, 150–165.
15. Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science 286*, 295–299.
16. Dima, R. I., and Thirumalai, D. (2006) Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci. 15*, 258–268.
17. Ghosh, A., Brinda, K. V., and Vishveshwara, S. (2007) Dynamics of lysozyme structure network: Probing the process of unfolding. *Biophys. J. 92*, 2523–2535.
18. Ghosh, A., and Vishveshwara, S. (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U.S.A. 104*, 15711–15716.
19. Sathyapriya, R., and Vishveshwara, S. (2007) Structure networks of *E. coli* glutaminyl-tRNA synthetase: Effects of ligand binding. *Proteins: Struct., Funct., Genet. 68*, 541–550.
20. Gale, A. J., Shi, J. P., and Schimmel, P. (1996) Evidence that Specificity of Microhelix Charging by a Class I tRNA Synthetase Occurs in the Transition State of Catalysis. *Biochemistry 35*, 608–615.
21. Alexander, R. W., and Schimmel, P. (2001) Domain-domain communication in aminoacyl-tRNA synthetases. *Prog. Nucleic Acid Res. Mol. Biol. 69*, 317–349.
22. Alexander, R. W., and Schimmel, P. (1999) Evidence for breaking domain-domain functional communication in a synthetase-tRNA complex. *Biochemistry 38*, 16359–16365.
23. Burbaum, J. J., and Schimmel, P. (1991) Assembly of a class I tRNA synthetase from products of an artificially split gene. *Biochemistry 30*, 319–324.
24. Weygand-Durasevic, I., Rogers, M. J., and Söll, D. (1994) Connecting Anticodon Recognition with the Active Site of *Escherichia coli* Glutaminyl-tRNA Synthetase. *J. Mol. Biol. 240*, 111–118.
25. Sherman, J. M., Thomann, H. U., and Söll, D. (1996) Functional Connectivity Between tRNA Binding Domains in Glutaminyl-tRNA Synthetase. *J. Mol. Biol. 256*, 818–828.
26. Uter, N. T., and Perona, J. J. (2004) Long-range intramolecular signaling in a tRNA synthetase complex revealed by pre-steady-state kinetics. *Proc. Natl. Acad. Sci. U.S.A. 101*, 14396–14401.
27. Zhang, C. M., and Hou, Y. M. (2005) Domain-domain communication for tRNA aminoacylation: The importance of covalent connectivity. *Biochemistry 44*, 7240–7249.
28. Palla, G., Derenyi, I., Farkas, I., and Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature 435*, 814–818.
29. Case, D. A., Darden, T. A., Cheatham, I. T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. (2006) *AMBER 9*, University of California, San Francisco.
30. Cheatham, T. E., III, Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn. 16*, 845–862.
31. Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S., Konno, M., and Blanquet, S. (1999) Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features. *J. Mol. Biol. 294*, 1287–1297.
32. Serre, L., Verdon, G., Choinowski, T., Hervouet, N., Risler, J.-L., and Zelwer, C. (2001) How methionyl-tRNA synthetase creates its amino acid recognition pocket upon L-methionine binding. *J. Mol. Biol. 306*, 863–876.
33. Brunie, S., Zelwer, C., and Risler, J. L. (1990) Crystallographic study at 2.5 Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. *J. Mol. Biol. 216*, 411–424.
34. Crepin, T., Schmitt, E., Mechulam, Y., Sampson, P. B., Vaughan, M. D., Honek, J. F., and Blanquet, S. (2003) Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase. *J. Mol. Biol. 332*, 59–72.
35. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926–935.
36. Kannan, N., and Vishveshwara, S. (1999) Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol. 292*, 441–464.
37. Brinda, K. V., and Vishveshwara, S. (2005) A network representation of protein structures: Implications for protein stability. *Biophys. J. 89*, 4159–4170.
38. Adamcsek, B., Palla, G., Farkas, I. J., Derenyi, I., and Vicsek, T. (2006) CFinder: Locating cliques and overlapping modules in biological networks. *Bioinformatics 22*, 1021–1023.
39. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 33*, 33–38.
40. Budiman, M. E., Knaggs, M. H., Fetrow, J. S., and Alexander, R. W. (2007) Using molecular dynamics to map interaction networks in an aminoacyl-tRNA synthetase. *Proteins: Struct., Funct., Bioinf. 68*, 670–689.
41. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*, 4673–4680.
42. Nakanishi, K., Ogiso, Y., Nakama, T., Fukai, S., and Nureki, O. (2005) Structural basis for anticodon recognition by methionyl-tRNA synthetase. *Nat. Struct. Mol. Biol. 12*, 931–932.
43. Schmitt, E., Meinnel, T., Blanquet, S., and Mechulam, Y. (1994) Methionyl-tRNA Synthetase Needs an Intact and Mobile 332KMSKS336 Motif in Catalysis of Methionyl Adenylate Formation. *J. Mol. Biol. 242*, 566–577.
44. del Sol, A., Arauzo-Bravo, M., Amoros, D., and Nussinov, R. (2007) Modular architecture of protein structures and allosteric communications: Potential implications for signaling proteins and regulatory linkages. *Genome Biol. 8*, R92.
45. Liu, J., and Nussinov, R. (2008) Allosteric effects in the marginally stable von Hippel−Lindau tumor suppressor protein and allostery-based rescue mutant design. *Proc. Natl. Acad. Sci. U.S.A. 105*, 901–906.
46. Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery. *Nat. Struct. Mol. Biol. 13*, 831–838.
47. Kormos, B. L., Baranger, A. M., and Beveridge, D. L. (2006) Do collective atomic fluctuations account for cooperative effects? Molecular dynamics studies of the U1A-RNA complex. *J. Am. Chem. Soc. 128*, 8992–8993.
48. del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol. 2*.
49. Fourmy, D., Mechulam, Y., Brunie, S., Blanquet, S., and Fayat, G. (1991) Identification of residues involved in the binding of methionine by *Escherichia coli* methionyl-tRNA synthetase. *FEBS Lett. 292*, 259–263.
50. Kim, H. Y., Ghosh, G., Schulman, L. H., Brunie, S., and Jakubowski, H. (1993) The Relationship Between Synthetic and Editing Functions of the Active Site of an Aminoacyl-tRNA Synthetase. *Proc. Natl. Acad. Sci. U.S.A. 90*, 11553–11557.
51. Schmitt, E., Panvert, M., Blanquet, S., and Mechulam, Y. (1995) Transition state stabilization by the 'high' motif of class I aminoacyl-tRNA synthetases: The case of *Escherichia coli* methionyl-tRNA synthetase. *Nucleic Acids Res. 23*, 4793–4798.

BI8007559