

# A Comparative Study of All-atom Molecular Dynamics Simulation and Coarse-grained Normal Mode Analysis in Identifying Pre-existing Residue Interaction Networks that Promote Coupled-Domain Dynamics in *Escherichia coli* Methionyl-tRNA Synthetase

Samuel C. Febling, Alexander M. Strom, Brent P. Lehman, Ryan J. Andrews, Sudeep Bhattacharyya\*, and Sanchita Hati\*  
Department of Chemistry, University of Wisconsin-Eau Claire, Wisconsin 54702

Student: andrewrj@uwec.edu\*  
Mentor: batis@uwec.edu

## ABSTRACT

Inter-domain communication plays a key role in the function of modular proteins. Earlier studies have demonstrated that the coupling of domain motions is important in mediating site-to-site communications in modular proteins. In the present study, bioinformatics and molecular simulations were used to trace “pre-existing” residue-residue interaction networks that mediate coupled-domain dynamics in multi-domain *Escherichia coli* methionyl-tRNA synthetase (Ec MetRS). In particular, a comparative study was carried out to evaluate the effectiveness of coarse-grained normal mode analysis and all-atom molecular dynamic simulation in predicting pre-existing pathways of inter-domain communications in this enzyme. Integration of dynamic information of residues with their evolutionary features (conserved and coevolved) demonstrated that multiple residue-residue interaction networks exist in Ec MetRS that promote dynamic coupling between the anticodon binding domain and the connective polypeptide I domain, which are > 50Å apart, through correlated motions. Mutation of residues on these pathways have distinct impact on the dynamics and function of this enzyme. Moreover, the present study revealed that the dynamic information obtained from the coarse-grained normal mode analysis is comparable to the atomistic molecular dynamics simulations in predicting the interaction networks that are essential for promoting coupled-domain dynamics in Ec MetRS.

## KEYWORDS

Domain-domain Communication; Molecular Dynamics; Methionyl-tRNA Synthetase; Normal Mode Analysis; Coupled-domain Dynamics; Course-grained Normal Mode Analysis; Aminoacyl tRNA Synthetases; Statistical Coupling Analysis

## INTRODUCTION

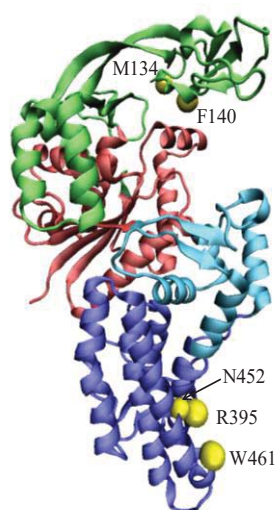
Studies on modular enzymes have suggested that coupling of dynamics between domains is critical for coordinating biological events occurring at distant sites.<sup>1-7</sup> There are two well-known models<sup>8-9</sup> for long-range allosteric communications - the “induced-fit” model (substrate-induced conformational change propagated through a single residue-residue interaction pathway) and the “population-shift” model (a perturbation at a distant site that alters the conformational equilibrium through “pre-existing” multiple pathways of residue-residue interactions).<sup>8,9</sup> Additionally, an updated model suggested that “pre-existing” multiple pathways of long-range site-to-site communications are present in the protein even in the absence of allosteric effectors.<sup>8</sup> After 50 years of debate, there is now growing consensus of opinion that conformational selection followed by conformational adjustment (“population-shift”) is the mechanism of ligand binding and allostery.<sup>7, 10-13</sup> Also, theoretical and experimental studies have shown that the site-to-site communication is propagated by networks of coupled residues and regulated by enthalpic (conformational) and/or entropic (dynamic) changes.<sup>14-17</sup>

For single domain polypeptides, Ranganathan and coworkers have shown that one or more contiguous networks of residues that are evolutionarily conserved or coevolved facilitate communications between distant functional sites.<sup>18-20</sup> However, for multi-domain proteins, the molecular-level picture of the inter-domain communication becomes quite complex as each domain functions more like an independent entity in terms of backbone flexibility and structural stability. As earlier studies have demonstrated that coupling of domain dynamics is prerequisite in coordinating biological events occurring at distant sites, inter-domain communication pathways in modular proteins could be identified by tracing the evolutionarily constrained residues that could mediate coupled domain dynamics.

Molecular dynamic (MD) simulations and normal mode analysis (NMA) have emerged as two important tools for studying protein dynamics. Although MD simulations provide invaluable insight into the atomic-level details of protein dynamics, they are computationally expensive to sample atomic motions that are relevant for biological functions (microsecond to millisecond time-scale motions). The alternative approach to determine proteins’ slow dynamics is NMA<sup>21</sup>. NMA could be used to sample wide-range of protein dynamics, including low-frequency collective motions, as well as high frequency local fluctuations. Interestingly,

Skjaerven et al. have shown that long-time-scale (200 ns) all-atom MD simulations, all-atom NMA, and coarse-grained NMA produce comparable results in depicting protein dynamics.<sup>22</sup> These observations suggest that the computationally inexpensive NMA method could be used to probe the site-to-site communication pathways in large modular proteins.

In this study, an attempt has been made to employ atomistic and coarse-grained simulation methods to trace the pre-existing pathways of residue interaction networks that could facilitate coupled-domain dynamics in the *Escherichia coli* methionyl-tRNA synthetase (Ec MetRS), a member of AARS family. Ec MetRS catalyzes covalent attachment of methionine to the tRNA<sup>Met</sup>. The accurate synthesis of methionyl-tRNA<sup>Met</sup> requires coordination of several events occurring in distant domains (**Figure 1**) of the Ec MetRS.<sup>23,24</sup> Presence of inter-domain communication in Ec MetRS is evident from various experimental studies. For example, the methionylation of tRNA<sup>Met</sup> at the catalytic site is triggered by tRNA binding at the anticodon domain, which is located ~ 50 Å away from the aminoacylation site. A 10<sup>6</sup>-fold decrease in aminoacylation activity ( $k_{cat}/K_M$ ) was observed for tRNAs with truncated anticodon sequence, but identical acceptor stem sequence.<sup>25-27</sup> Moreover, mutations of tRNA anticodon nucleotides or the highly conserved C-terminal residues of MetRS (namely, W461, N452, and R395, which are essential for the anticodon base recognition) have significant (~10<sup>5</sup>-fold) impact on the efficiency of tRNA aminoacylation.<sup>28-31</sup> Although, separate values of these kinetic parameters were not deciphered, the existence of an inter-domain communication, stretching from the anticodon binding domain to the CP domain (**Figure 1**), is believed to be crucial in shaping a functional Ec MetRS molecule.



**Figure 1. Cartoon representation of the 3-dimensional structure of the monomeric form of Ec MetRS (residues 3-548, PDB entry: 1QQT).** The structural domains are colored as follows: red, Rossmann-fold (residues 1-96, 252-323); green, connective polypeptide (CP) domain (residues 97-251); cyan, KMSKS domain (residues 324-384 and 536-547) and blue, the C-terminal  $\alpha$ -helix bundle domain (residues 385-535). The three residues, W461, N452, and R395, in the C-terminal domain and M134 and F140 of the CP domain are shown in yellow beads.

The long-range communication between the anticodon domain and the catalytic domain in Ec MetRS has been previously studied by atomistic MD simulations in the presence of tRNA.<sup>32,33</sup> Ghosh *et al.* have observed that the predicted communication pathways do not involve any residues of the tRNA. This poses an important question – if long-range inter-domain communications propagate through only protein, then the protein sequence, its structure/folding, and the intrinsic dynamics (arising out of the folding) have all necessary information that can promote such communications. In fact, this is evident from our combined bioinformatics and molecular simulation studies on *Thermus thermophilus* leucyl-tRNA synthetase (Tt LeuRS) and Ec prolyl-tRNA synthetase (ProRS), where we observed that the thermally coupled and evolutionarily constrained (coevolved and conserved) residues facilitate coupled-domain dynamics.<sup>34,35</sup> Therefore, in the present work we have employed the recently developed bioinformatics-based Statistical Thermal Coupling Analysis (STCA) method<sup>35</sup> to explore the molecular mechanism of inter-domain communications in Ec MetRS. In addition, we aimed to explore if the dynamical features obtained from the coarse-grained simulations and all-atom molecular dynamics simulations could produce comparable results. Finally, to validate the STCA results, selected residues on the predicted pathways were mutated and the impact of mutations of pathways residues on the collective dynamics of a distant domain were also probed using atomistic simulations.

## METHODS AND PROCEDURES

Based on reported experimental data, W461 (C-terminal domain) and M134/F140 (CP domain) were selected as the two termini between which the interaction networks were mapped.<sup>31</sup> Especially, the M134 region of the CP domain and W461 of the C-

terminal are known to be important for aminoacylation.<sup>24</sup> Visualizations and mutations were performed using VMD software<sup>36</sup>. Statistical coupling analysis (SCA) was carried out using a MATLAB script obtained from Ranganathan lab. Normal mode analysis (NMA) was carried out using the coarse-grained Anisotropic Network Model (ANM)<sup>21,37,38</sup> using the bioinformatics server <http://ignmtest.cccb.pitt.edu/cgi-bin/anm/anm1.cgi>. MD simulations were carried out using NAMD<sup>39</sup> and the all-atom CHARMM22<sup>40</sup> force field. Principal component analysis was carried out using CARMA.<sup>41</sup> The NMA and SCA plots and all data processing were carried out using MATLAB R2006b (The MathWorks Inc., Natick, MA).

#### Statistical Thermal Coupling Analysis

STCA was carried out in four discrete steps as described by Johnsons et al.<sup>35</sup> First, SCA<sup>18,19</sup> was performed to identify conserved and coevolved residues in the MetRS family. Next, the collective motions of various domains were studied by performing coarse-grained NMA,<sup>21,37,38,42,43</sup> as well as all-atom MD simulations. In the third step, the evolutionary dependence of the coupled-domain dynamics was explored by integrating the results of SCA and NMA/MD, which resulted in a subset of residues that are simultaneously coupled through evolution and thermal motion. In the final step, networks of interacting residues between the two distant sites were identified from the pool of dynamically and evolutionarily coupled residues using Dijkstra's algorithm.<sup>44</sup>

#### Statistical Coupling Analysis

SCA is based upon the assumption that the "coupling of two sites in a protein, whether for structural or functional reasons should cause those two sites to coevolve".<sup>18,19</sup> SCA was carried out using the protocol described elsewhere.<sup>18,20</sup> Briefly, this method uses a multiple sequence alignment (MSA) of a protein family as an input file and quantifies the extent of residue conservation, as well as coevolution between two residues by calculating the change in the amino acid distribution at one position with respect to a perturbation at another position. In the present work, the MSA of 478 protein sequences of the MetRS family was constructed using PSI-BLAST.<sup>45</sup> Only MetRS sequences that bear significant sequence identity (> 75 %) with Ec MetRS were included in this study. The conservation constant  $\Delta G_i^{\text{stat}}$  and the coupling constant  $\Delta \Delta G_i^{\text{stat}}$  were obtained using standard procedures described previously.<sup>34,35</sup>

#### Normal mode analysis

Low-frequency (large-amplitude) motions are important for the biological function of a protein. NMA has been shown to be reliable in describing the large-scale conformation changes in biomolecules.<sup>27,46-48</sup> In the present study, coarse-grained NMA was used in which a protein is simplified to a string of beads<sup>49</sup> and each bead represents a  $C_\alpha$  atom.<sup>21,37,38</sup> The NMA was performed using the Anisotropic Network Model (ANM), where fluctuations are anisotropic and the overall potential energy of the protein system is expressed as the sum of harmonic potentials between the interacting  $C_\alpha$  atoms.<sup>49</sup>

$$V_{ANM} = \frac{\gamma}{2} \left[ \sum_{j,i \neq j} \Gamma_{ij} (r_{ij} - r_{ij}^\circ)^2 \right] \quad \text{Equation 1.}$$

In eq. 1,  $\gamma$  represents the uniform spring constant,  $r_{ij}^\circ$  and  $r_{ij}$  are the equilibrium and instantaneous distance between residues  $i$  and  $j$ , respectively, and  $\Gamma_{ij}$  is the  $ij$ -th element of the binary connection matrix of inter-residue ( $C_\alpha - C_\alpha$ ) contacts. Based on an interaction cutoff distance of  $r_c$ ,  $\Gamma_{ij}$  is equal to 1 if  $r_{ij}^\circ < r_c$  and zero otherwise.<sup>37</sup> Previous studies by Eyal *et al.* demonstrated that the 18 Å interaction cutoff resulted in a better correlation between experimental and calculated fluctuations.<sup>38</sup> Therefore, a 18 Å cutoff has been used in the present study. The substrate-unbound enzyme; the crystal structure of monomeric Ec MetRS (residues 3-549, PDB entry: 1QQT) was used.

The correlated or anti-correlated motions between  $C_\alpha$  atoms were analyzed by computing the dynamic cross-correlation matrix  $\mathbf{C}$ , where the  $ij$ -th element  $C_{ij}$  represents the cross-correlation coefficient between fluctuations of residues at sites  $i$  and  $j$ :

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \quad \text{Equation 2.}$$

The atomic ( $C_\alpha$ ) displacements of residues  $i$  and  $j$  are represented by  $\Delta r_i$  and  $\Delta r_j$ , respectively, and the angular brackets represent an ensemble average calculated over structures for combined normal modes.

#### Molecular Dynamics Simulation

MD simulations of Ec MetRS were performed using the X-ray crystal structure of Ec MetRS (PDB entry: 1QQT). All mutants were generated with the 'Mutate Residue' plug-in (version 1.3) of Visual Molecular Dynamics (VMD) version 1.9.1.<sup>36</sup> Simulations were performed in water (TIP3P model)<sup>50</sup> with substrate-free enzymes using the all-atom CHARMM27 force field<sup>40</sup> within the NAMD package.<sup>39</sup> Nonbonded interactions were truncated using a switching function between 10 and 12 Å, and the dielectric

constant was set to unity. The SHAKE algorithm<sup>51</sup> was used to constrain bond lengths and bond angles of water molecules and bonds involving a hydrogen atom. The MD simulations were performed using isothermal–isobaric (NPT) conditions. Periodic boundary conditions and particle mesh Ewald methods<sup>52</sup> were used to account for the long-range electrostatic interactions. In all MD simulations, a time step of 2 fs was used. The pressure of the system was controlled by the implementation of the Berendsen pressure bath coupling<sup>53</sup> as the temperature of the system was slowly increased from 100 to 300 K. During the simulations at 300 K, the pressure was kept constant by applying the Langevin piston method.<sup>54,55</sup> The WT and mutant proteins were solvated with water in a periodic rectangular box with water padding of 12 Å between the walls of the box and the nearest protein atom. The charge neutralization (with sodium ions) of the solvated system was performed with the VMD autoionize extension.<sup>36</sup> The resultant systems were equilibrated by slightly modifying previously described procedures.<sup>56,57</sup> All simulations were carried out with a 20 ps equilibration to minimize the amino acid side chain interactions followed by a 30 ns production MD run. The details of the MD simulation protocol were as described previously.<sup>58</sup> To evaluate the statistical significance of the MD simulation analysis, three replicates were generated for each protein system, as described in the protein simulation studies by Roy and Laughton.<sup>59</sup>

The correlated motions between residue pairs were studied by principal component analysis (PCA) of collective motions<sup>60</sup> as described earlier.<sup>35,58,61</sup> Following the method described in Johnson et al.,<sup>35</sup> the last 25 ns of the MD simulation data were used to generate principal components of atomic (backbone  $C_\alpha$  atoms) fluctuations by Carma.<sup>41</sup> The first three principal components, which represent the low-frequency (high-amplitude) collective motions, were used to perform cluster analysis. This analysis produced a new trajectory of conformations representing the predominant conformational fluctuations and were used for generating dynamic cross-correlation matrix  $\mathbf{C}$ . The  $ij$ -th element,  $C_{ij}$  in matrix  $\mathbf{C}$  represents the cross-correlation coefficient between residue fluctuations at sites  $i$  and  $j$  during the simulation:

$$C_{ij} = \frac{\langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle}{\sigma_{x_i} \sigma_{x_j}} \quad \text{Equation 3.}$$

The atomic ( $C_\alpha$ ) displacements of residues  $i$  and  $j$  are represented by  $x_i$  and  $x_j$ , respectively; the angular brackets represent ensemble averages, and  $\sigma_{x_i}$  and  $\sigma_{x_j}$  represent the standard deviations of these displacements.

#### Integration of Evolutionary and Dynamic Information

To trace the residue-residue interaction networks that play key role in facilitating coupled-dynamics between domains, we constructed a subset of residues that are both, evolutionarily and dynamically coupled.<sup>35</sup> The motional coupling information, obtained from NMA, was integrated with the evolutionary conservation and coevolution dataset from the SCA. The conserved and coevolved residues were treated separately. The conserved and dynamically coupled residues were obtained by selecting only those residues that exhibit significant conservation ( $\Delta G_i^{\text{stat}} \geq 0.5$ ) as well as motional coupling ( $C_{ij} \geq 0.8$ ) with each other. The value of  $C_{ij}$  was set to  $\geq 0.8$  in order to obtain statistically relevant size of the evolutionarily conserved residues. The coevolved and dynamically coupled residues were obtained from the truncated  $\mathbf{C}$  matrix, which was created by including only those columns that are present in the normalized SCA-derived  $\mathbf{G}$  matrix. Next, a new matrix, the coevolutionary dynamic coupling ( $\mathbf{CDC}$ ) matrix, was created by multiplying each  $ij$ -th element of the  $\mathbf{G}$  matrix with the corresponding element of the truncated  $\mathbf{C}$  matrix:

$$CDC_{ij} = \Delta G_{ij}^{\text{stat}} \times C_{ij} \quad \text{Equation 4.}$$

#### Identification of interaction networks using dijkstra's algorithm

From the shortlisted residues, residue-residue interaction networks between W461 (C-terminal domain) and M134/F140 (CP domain) of Ec MetRS were identified using Dijkstra's algorithm as described earlier.<sup>35,44</sup> In this method, each  $C_\alpha$  atom of the protein backbone represents a node. The connectivity between two adjacent nodes [inter-residue ( $C_\alpha - C_\alpha$ ) contacts] is described by a binary connection matrix  $\mathbf{P}$ .<sup>35</sup> The  $C_\alpha - C_\alpha$  distance matrix,  $\mathbf{D}$  was created from the PDB file containing the Cartesian coordinates of all  $C_\alpha$  atoms of the Ec MetRS (PDB entry: 1QQT). Based on a  $C_\alpha - C_\alpha$  cutoff distance  $D_{ij}^\circ$ ,  $P_{ij}$  is equal to 1 if  $D_{ij} < D_{ij}^\circ$  and zero otherwise.<sup>35,44</sup> The interaction networks (pathways) between two functional sites were identified by varying the distance parameter,  $D_{ij}$  and the statistical parameter,  $\Delta G_i^{\text{stat}}$ .

#### Root-mean-square fluctuations

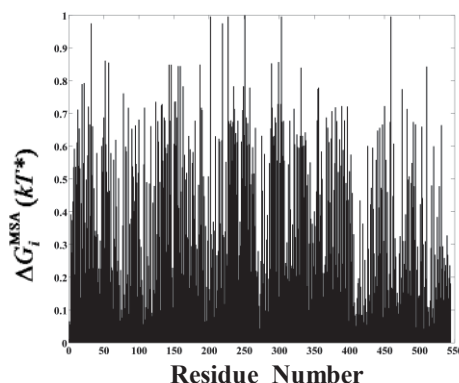
To determine if the motion of CP domain had undergone significant change upon mutation along the interaction networks, the slow dynamics of the CP domain due to a specific mutation was studied by performing PCA for the WT and mutant proteins. Root-Mean-Square fluctuations (RMSF) of  $C_\alpha$  atoms, averaged over three replicate simulations, were obtained for the WT and mutants. In these calculations, the last 25 ns of MD simulation data were used, each comprising an ensemble of 250,000

conformations.<sup>35</sup> Similar to our previous work, an ensemble of 750,000 conformations, obtained by combining three replicate trajectories, was used to perform PCA for each of these protein systems. The first three clusters, which represent the predominant conformational fluctuations, were used in this study.

## RESULTS

### Conserved and coevolved residues

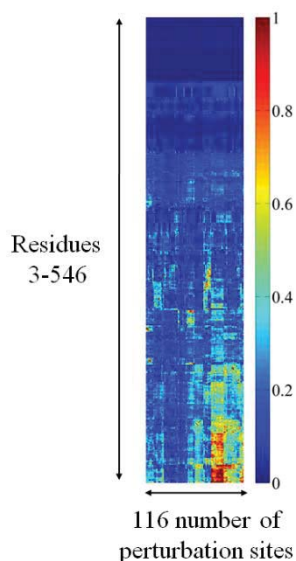
To identify the conserved residues, the  $\Delta G_i^{\text{stat}}$  value of each residue of the Ec MetRS sequence was calculated using SCA.  $\Delta G_i^{\text{stat}}$ , a quantitative measure of the conservation of a residue at the  $i$ th position of the sequence was obtained as a one-dimensional vector normalized to  $1 kT^*$  (Figure 2).



**Figure 2.** A normalized plot of  $\Delta G_i^{\text{stat}}$  vs. position (residue number) to identify conserved residues from SCA. The extent of evolutionary conservation,  $\Delta G_i^{\text{stat}}$ , of a residue at the  $i$ th position of the Ec MetRS sequence was determined by using the following equation:

$$\Delta G_i^{\text{stat}} = kT^* \sqrt{\sum_x [\ln(P_i^x / P_{\text{MSA}}^x)]^2}$$

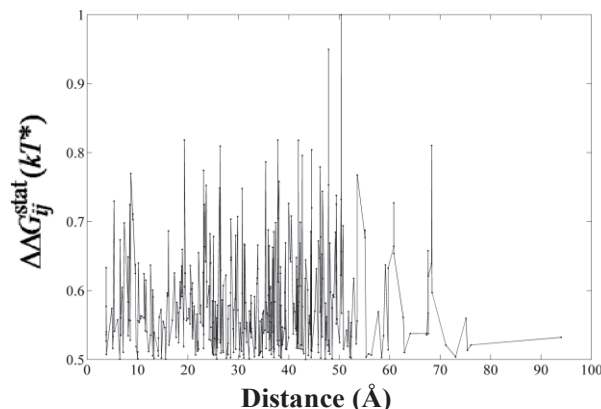
where  $kT^*$  is an arbitrary energy unit,  $P_i^x$  is the binomial probability of observing amino acid  $x$  at site  $i$ ,  $P_{\text{MSA}}^x$  is the probability of observing amino acid  $x$  in the overall MSA, and the summation is over all 20 amino acids.



**Figure 3. Statistical coupling analysis of the MetRS family.** The normalized statistical coupling matrix with one dimensional hierarchical clustering along the perturbation axis is shown. The highly coevolved residues are clustered at the bottom right of the diagram. The color gradient, as indicated in the color bar, is as follows: blue squares represent the lowest ( $0 kT^*$ ) and red squares represent the highest ( $1 kT^*$ ) statistical coupling energies.

In addition, the evolutionary coupling indices ( $\Delta\Delta G_i^{\text{stat}}$ ) of a total 544 residues for 116 perturbation sites were obtained from a  $544 \times 116$  coupling matrix (Figure 3). It is evident that only a small fraction of Ec MetRS residues have high  $\Delta\Delta G_i^{\text{stat}}$  values and those highly coupled residues were found to be located farther apart from each other in the three-dimensional structure of Ec

MetRS (Figure 4). To explore the role of these conserved and coevolved residues in the long-range interactions, the dynamics of the protein segments and their motional coupling was examined.



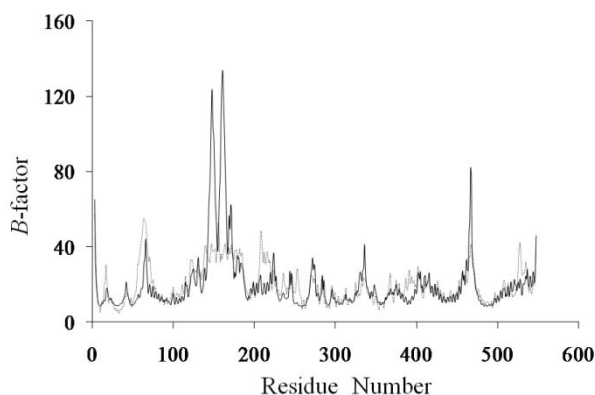
**Figure 4.** The statistical coupling,  $\Delta\Delta G_{ij}^{stat}$  for all residues plotted against their contact ( $C_{\alpha} - C_{\alpha}$ ) distances. The extent of statistical coupling between two sites,  $i$  and  $j$ , of the MetRS protein sequence was estimated from a perturbation analysis, where a sub-alignment of the MSA was created using  $\sim 63\%$  of the total number of sequences and the change in the extent of conservation was calculated. Mathematically, this is represented as a normalized statistical coupling matrix  $\mathbf{G}$ , in which each  $ij$ -th element,  $\Delta\Delta G_{ij}^{stat}$ , measures the perturbation in the conservation of residue  $i$  due to residue  $j$ .  $\Delta\Delta G_{ij}^{stat}$  refers to the evolutionary (statistical) coupling between residues at two functional sites,  $i$  and  $j$  and is expressed as:

$$\Delta\Delta G_{i,j}^{stat} = kT * \sqrt{\sum_x \left[ \ln \left( \frac{P_{i|\delta_j}^x}{P_{MSA|\delta_j}^x} \right) - \ln \left( \frac{P_i^x}{P_{MSA}^x} \right) \right]^2}$$

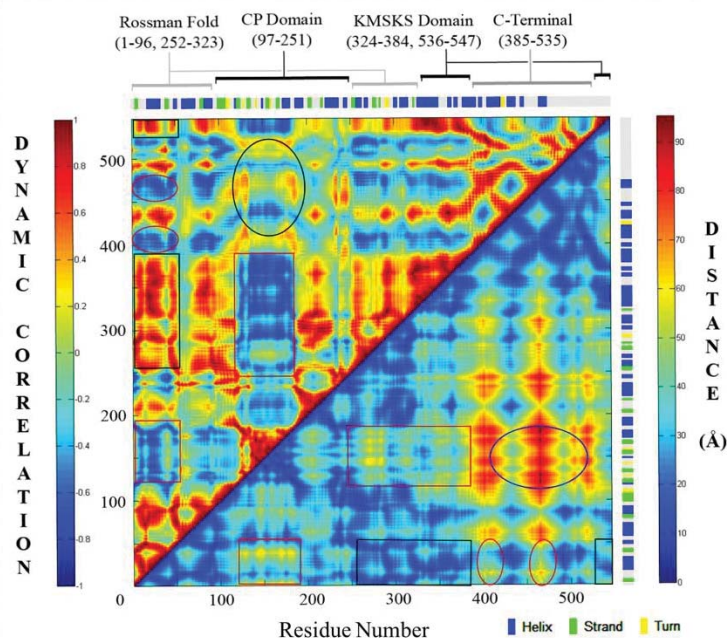
where  $P_{i|\delta_j}^x$  is the probability of  $x$  at site  $i$  being dependent on a perturbation at site  $j$ . A large statistical coupling value,  $\Delta\Delta G^{stat}$ , indicates strong energetic coupling between residues at the two sites  $i$  and  $j$ <sup>1,2</sup>

*Normal mode analysis and flexible protein segments*

The dynamic correlation between residues of the Ec MetRS was extracted from the NMA using a combination of the first 10 lowest-frequency normal modes (modes 1-10), which are usually biologically more relevant. The identical flexible regions in the protein structure are evident from the plot of computed and the experimental (crystallographic)<sup>23</sup>  $B$ -factors of  $C_{\alpha}$  atoms (Figure 5). The study also revealed that the CP domain, especially residues 100-200, is the most mobile. Residues 56-71 in the Rossmann-fold and 454-476 of the anticodon binding domain also exhibit significant mobility, which is consistent to the atomistic simulation study reported earlier.<sup>32,33</sup>

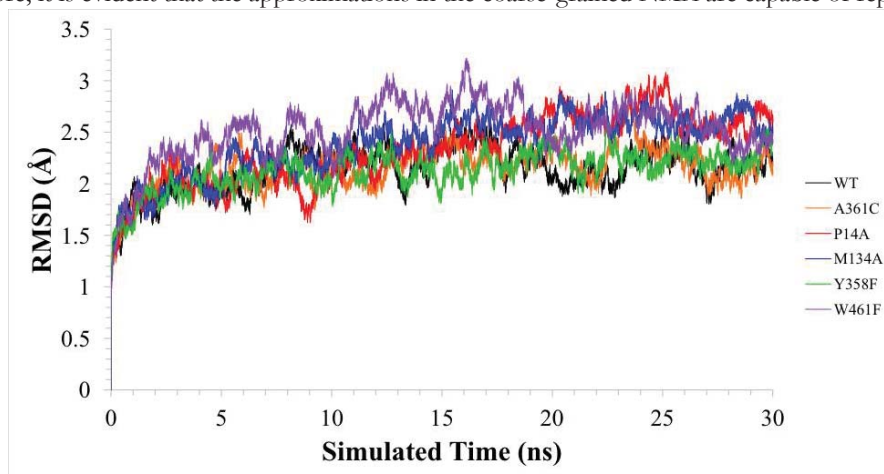


**Figure 5.**  $C_{\alpha}$   $B$ -factor analysis of Ec MetRS. Comparison of the  $C_{\alpha}$   $B$ -factor obtained from the crystal structure (gray, dotted line) (PDB entry: 1QQ1) and the calculated one (black, solid line) using ANM.



**Figure 6.** Correlations in fluctuations between residues (above diagonal) and their contact ( $C_{\alpha} - C_{\alpha}$ ) distances (below diagonal). Dynamic cross-correlations between the  $C_{\alpha}$  atoms of the Ec MetRS (PDB entry: 1QQT) have been calculated based on the lowest 10 normal modes. A value of +1.0 was set for strongly correlated motion and is colored red, whereas -1.0 was used for the strongly anticorrelated motion and is colored blue. The axis values correspond to residue number and the color scale for contact distances is set between 0 – 100 Å; blue for residues in close contact and red for residues located farther away.

Analysis of the dynamic cross-correlation matrix (DCCM) obtained from the combined modes (1-10) revealed that various structural elements within the catalytic domain are engaged in correlated motions (**Figure 6**). The Rossmann-fold domain (residues 1-96 and 245-323) is engaged in strong correlated motion with the KMSKS domain (residues 324-384 and 536-547, **Figure 6**, black rectangles). In contrast, the CP domain (residues 125-200) moves in an anticorrelated pattern with respect to the Rossmann-fold (residues 1-96 and 252-323) and the KMSKS (residues 324–384) domains (**Figure 6**, red rectangles). The observed anti-correlated motion between the CP domain and the main body is significant as this anti-correlated motion provide an adequate space for the 3'-end of tRNA to enter into the synthetic active site for aminoacylation. Similar observations have been made in other class I aminoacyl-tRNA synthetases.<sup>34,62-64</sup> Furthermore, the anticodon binding domain (residues 385-535) and the catalytic site (residues 1-96) exhibit anticorrelated motion (**Figure 6**, red ovals). The CP and the C-terminal domains (separated by a distance > 70 Å, Fig. 4, blue oval) are engaged in partly anticorrelated motion (**Figure 6**, black ovals). These observations are quite consistent with the earlier reported long time-scale (10 ns) MD simulation results,<sup>32,33</sup> as well as the present MD simulation study (*vide infra*). Therefore, it is evident that the approximations in the coarse-grained NMA are capable of reproducing the same

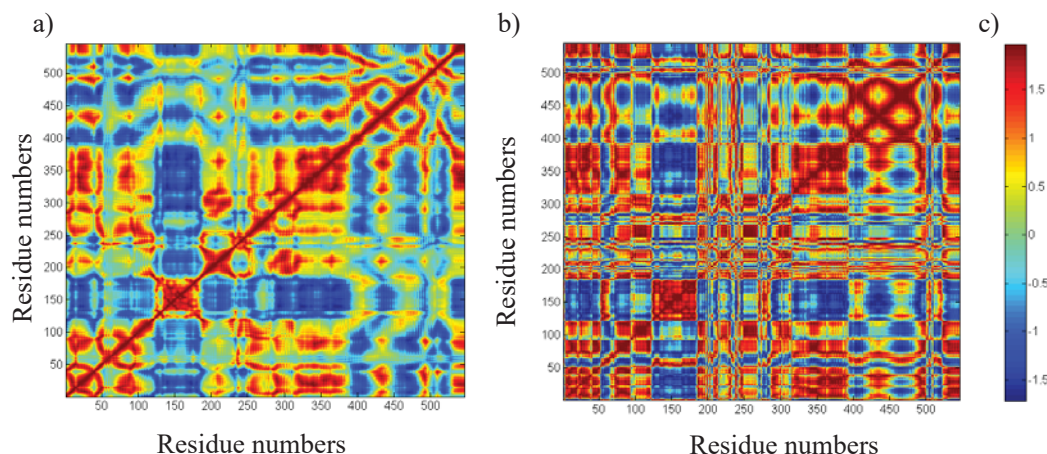


**Figure 7.** RMSD of the  $C_{\alpha}$  atoms from their initial coordinate as a function of time for WT and the five mutants of Ec MetRS. Calculations of RMSDs were performed using 30-ns MD simulation data.

low-frequency dynamics of a large protein like Ec MetRS as obtained from the use of all-atom MD simulation.

#### MD simulations

30 ns MD simulation was carried out for the Ec MetRS to generate the dynamic cross-correlation matrix **C**. The quality of the simulation was first tested by computing the root-mean-square-deviation (RMSD) of the  $C_{\alpha}$  atoms from their initial coordinates. The RMSD obtained from the 30 ns MD simulation trajectory is shown in **Figure 7**. The  $C_{\alpha}$  RMSD values were observed to fluctuate with a mean value of 1.0-1.5 Å during the production period (30 ns) simulations. The dynamic cross-correlation matrix **C** of the WT enzyme (**Figure 8**) was generated using the first three principal components. Analysis of the cross-correlation of fluctuations of residues revealed both inter- and intra-domain dynamic correlations, which are very similar to those obtained from NMA.



**Figure 8.** Dynamic cross-correlation between the  $C_{\alpha}$  atoms of Ec MetRS as obtained from the a) combined first 10 low-frequency modes of NMA; b) 30 ns MD simulation; c) A color scale; a value of +1.0 represents a strongly correlated motion whereas -1.0 represents a strongly anticorrelated motion.

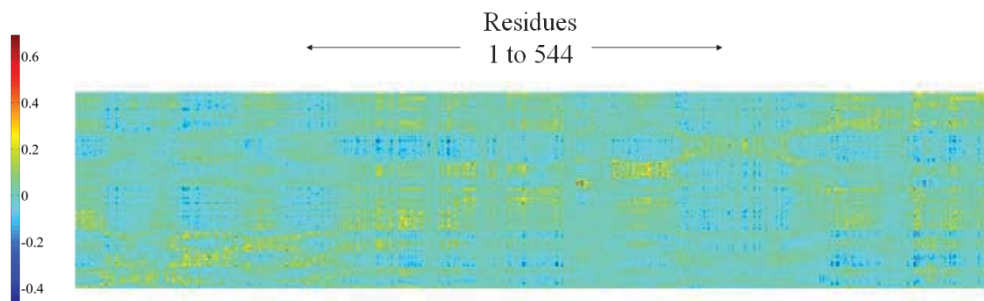
#### Generating residue pools that exhibit statistical and thermal coupling

Results of thermal fluctuations were combined with those of the evolutionary constraints (conservation and coevolution) of residues. Conserved and coevolved residues were treated differently. In order to obtain a reliable set of residues that were dynamically coupled as well as highly conserved in the MetRS family, a controlled experiment was carried out by varying  $\Delta G_i^{\text{stat}}$  cutoff values and monitoring their distributions in various domains. For a specific  $\Delta G_i^{\text{stat}}$  cutoff value, the number of conserved residues varies in each domain of Ec MetRS (**Table 1**). The study revealed that only ~ 15% of C-terminal residues are moderately conserved ( $\Delta G_i^{\text{stat}} \geq 0.5$ ) and exhibit strong motional coupling. In contrast, the Rossmann-fold domain, which is the catalytic domain, contains higher number of conserved residues (~ 40%) that are also dynamically correlated (**Table 1**). In the present study, subsets of conserved and dynamically correlated residues were generated by using a high dynamic coupling constant ( $C_{ij} \geq 0.8$ ) and varying the conservation cutoff ( $\Delta G_i^{\text{stat}}$ ) between 0.50-0.65 (**Table 1**). The cutoffs of these parameters were selected in order to obtain a reasonable size of residue pool consisting of statistically and thermally coupled residues

Domain	$\Delta G_i^{\text{stat}}$			
	$\geq 0.5$	$\geq 0.55$	$\geq 0.60$	$\geq 0.65$
Rossmann-fold (residues 1-96 and 252-323)	72	56	44	32
Connective polypeptide (residues 97-251)	66	57	49	36
KMSKS (residues 324-384 and 536-547)	26	24	19	9
C-terminal (residues 385-535)	29	22	16	12
Total	193	159	128	89

**Table 1.** Number of conserved residues in different domains of Ec MetRS obtained using various cutoff values for  $\Delta G_i^{\text{stat}}$ . The dynamic coupling cutoff ( $C_{ij}$ ) was set to greater than or equal to 0.8.





**Figure 9.** The SCA•NMA plot obtained from the CDC matrix by multiplying individual elements of the SCA matrix with the corresponding elements of the truncated NMA matrix. Values range from +0.67 (coevolved and thermally correlated) to -0.47 (coevolved and thermally anticorrelated).

The coevolving and dynamically correlated residues were extracted from the coevolutionary dynamic coupling matrix CDC (eq. 4), where  $CDC_{ij}$  values range between -0.48 to 0.69 for NMA (Figure 9) and -0.74 to 0.70 for MD simulation (data not shown). In order to obtain a statistically significant set of residues, a systematic study was conducted using variable  $CDC_{ij}$  cutoffs. This study resulted in subsets of residues with various degrees of statistical-thermal coupling and their distributions in different domains are reported in Table 2. In the present study, we have chosen the  $CDC_{ij}$  cutoff value of 0.4, which resulted in a pool of 76 residues.

Domain	$\Delta C_{ij}^{stat}$			
	$\geq 0.3$	$\geq 0.35$	$\geq 0.40$	$\geq 0.45$
Rossmann-fold (residues 1-96 and 252-323)(168 Total)	25%	19%	15%	11%
Connective polypeptide (residues 97-251) (155 Total)	23%	20%	15%	12%
KMSKS (residues 324-384 and 536-547) (73 Total)	33%	29%	20%	9%
C-terminal (residues 385-535) (151 Total)	28%	20%	10%	5%
Total	27%	21%	14%	9%

**Table 2.** Percentage of coevolved residues within different domains of Ec MMetRS obtained using various cutoff values for  $CDC_{ij}$  (eq. 4).

*Interaction networks across domains*

To map interaction networks between C-terminal domain (W461) and CP domain (M134/F140), Dijkstra’s algorithm<sup>44</sup> was used. van der Waal’s radius of an amino acid  $C_{\alpha}$  is  $\sim 3 \text{ \AA}$  and the acceptable distance of non-covalent interaction between two interacting atoms is 2.0–3.0  $\text{ \AA}$ , a strong non-covalent interaction occurs between two  $C_{\alpha}$  atoms of a folded protein when they are within a distance of 8.0-9.0  $\text{ \AA}$ . Therefore, the distance cutoff,  $D_{ij}^{\circ}$  was varied between 8.0-9.0  $\text{ \AA}$  to include only those neighboring  $C_{\alpha}$  atoms that are engaged in strong non-covalent interactions and therefore could propagate site-to-site communications between distant domains.<sup>35</sup>

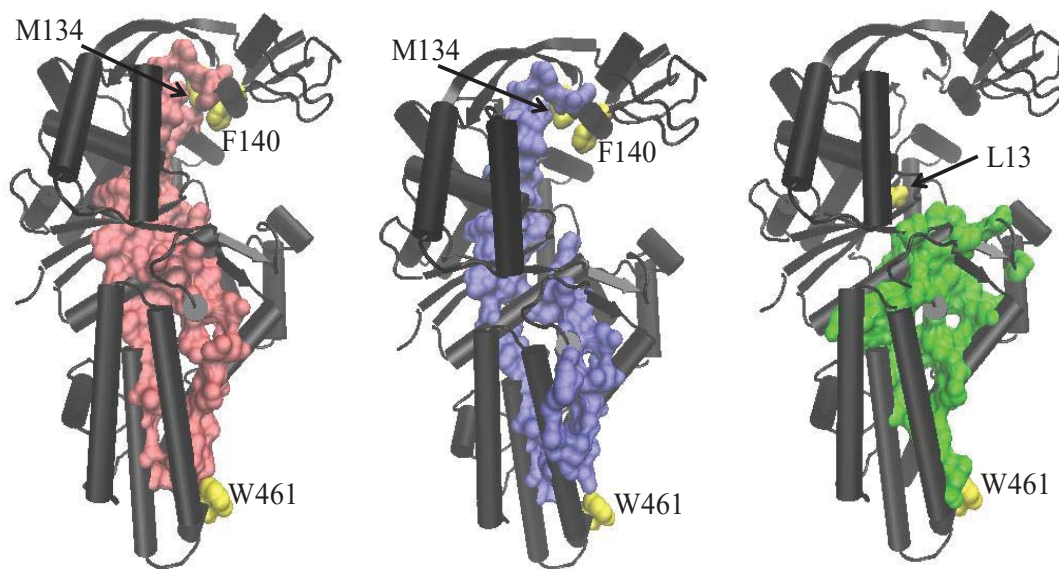
Pathways	Parameters	Residue Network (MD)	Residue Network (NMA)
I	$G_i^{stat} \geq 0.5$ , $D_{ij} \leq 8.0 \text{ \AA}$	W461→P460→C477→ <b>G480</b> →F484→ L487→L491→P493→R36→V35→ <b>F47</b> → <b>C49</b> →A50→D52→H54→F238→P236→M134	W461→P460→C477→ <b>G480</b> →F484→L487 →L491→P493→R36→V35→ <b>F47</b> →C49→ A50→D52→H54→F238→P236→M134
II	$G_i^{stat} \geq 0.5$ , $D_{ij} \leq 8.5 \text{ \AA}$	W461→P460→C477→ <b>G480</b> →F484→ L487→L491→P493→V35→ <b>F47</b> → <b>C49</b> →A50→D52→H54→ <u>Y237</u> →P236→M134	W461→P460→C477→ <b>G480</b> →F484→L487 →L491→P493→ <u>R36</u> →V35→ <b>F47</b> →C49→ A50→D52→H54→ <u>F238</u> →P236→M134
III	$G_i^{stat} \geq 0.5$ , $D_{ij} \leq 9.0 \text{ \AA}$	W461→P460→R395→N391→V386→V381 →Y358→ <b>Y359</b> → <b>L26</b> →H24→P14→D52 →H54→Y237→M134	W461→P460→R395→N391→V386→V381 →Y358→ <b>Y359</b> → <b>L26</b> →H24→P14→D52 →H54→Y237→M134
IV	$G_i^{stat} \geq 0.55$ , $D_{ij} \leq 8.0 \text{ \AA}$	W461→P460→ <b>V455</b> →N452→K388→D384→ <u>Y357</u> →R380→Y358→ <b>Y359</b> →I29→ <b>H28</b> → <b>C11</b> → <u>C49</u> →D51→ <b>A53</b> →H54→F238→P236→ M134	W461→P460→ <u>C477</u> →G480→ <b>A451</b> →D449→ <b>A446</b> →R442→ <b>A361</b> →Y359→I29→ <b>189</b> → <b>Y91</b> → <b>Y94</b> →D51→ <b>A53</b> →H54→F238→P236→ M134
V	$G_i^{stat} \geq 0.55$ , $D_{ij} \leq 8.5 \text{ \AA}$	W461→P460→ <b>V455</b> →N452→K388→D384→ Y357→ <b>R380</b> →Y358→Y356→F87→F84→ <b>H80</b> →D51→ <b>A53</b> →H54→Y237→P236→M134	W461→P460→ <u>C477</u> →R395→N391→K388 →D384→Y357→ <b>R380</b> →Y358→ <b>Y359</b> → <b>L26</b> → H24→P14→D52→H54→Y237→P236→M134
VI	$G_i^{stat} \geq 0.55$ , $D_{ij} \leq 9.0 \text{ \AA}$	W461→P460→ <b>V455</b> →N452→K388→D384 →Y357→ <b>R380</b> →Y358→ <b>Y359</b> → <b>L26</b> →H24 →P14→D52→H54→Y237→M134	W461→P460→R395→N391→K388→D384→ Y357→ <b>R380</b> →Y358→ <b>Y359</b> → <b>L26</b> →H24→P14 →D52→H54→Y237→M134

**Table 3.** Probable pathways of communication between W461 and M134 in the Ec MetRS. Residues in bold are coevolved and the rest are evolutionarily conserved. The residues that are different between MD and NMA predicted pathways are underlined. Values of the two parameters,  $C_{ij}$  and  $CDC_{ij}$ , are set to greater than or equal to 0.8 and 0.4, respectively.

Pathways	Parameters	Residue Network (MD)	Residue Network (NMA)
I	$G_i^{stat} \geq 0.5$ , $D_{ij} \leq 8.0 \text{ \AA}$	W461→P460→C477→ <b>G480</b> →F484→L487 →L491→P493→ <b>D32</b> → <b>A31</b> →T10→Y260 →A256→V252→Y251→R233→D234→ F135→L136→F140	W461→P460→C477→ <b>G480</b> →F484→L487 →L491→P493→R36→V35→V9→T10→Y260 →A256→V252→Y251→R233→D234→F135 →L136→F140
II	$G_i^{stat} \geq 0.5$ , $D_{ij} \leq 8.5 \text{ \AA}$	W461→P460→C477→ <b>G480</b> →F484→L487 →L491→P493→V35→V9→T10→Y260 →A256→Y251→R233→D234→F135→ L136→F140	W461→P460→C477→ <b>G480</b> →F484→L487 →L491→P493→ <u>R36</u> →V35→V9→T10→Y260 →A256→ <u>Y252</u> →Y251→R233→D234→F135 →L136→F140
III	$G_i^{stat} \geq 0.5$ , $D_{ij} \leq 9.0 \text{ \AA}$	W461→P460→R395→N391→V386→ V381→Y358→ <b>Y359</b> → <b>L26</b> →H24→P14→ D52→ <u>G55</u> → <u>T56</u> → <u>P137</u> →F140	W461→P460→R395→N391→V386→V381→ Y358→ <b>Y359</b> → <b>L26</b> →H24→P14→D52→ <u>H54</u> → <u>Y237</u> →M134→L136→F140
IV	$G_i^{stat} \geq 0.55$ , $D_{ij} \leq 8.0 \text{ \AA}$	W461→P460→ <b>V455</b> →N452→K388→ D384→Y357→R380→Y358→ <b>Y359</b> →Q30→ <u>A31</u> →T10→Y260→A256→V252→Y251→ R233→D234→F135→L136→F140	W461→P460→ <u>C477</u> →G480→ <b>A451</b> →D449→ <b>A446</b> →R442→ <b>A361</b> →Y359→I29→ <b>189</b> → <b>Y91</b> → <u>C49</u> → <u>G259</u> → <u>D255</u> →Y251→R233→D234→ F135→L136→F140
V	$G_i^{stat} \geq 0.55$ , $D_{ij} \leq 8.5 \text{ \AA}$	W461→P460→ <b>V455</b> →N452→K388→ D384→Y357→ <b>R380</b> →Y358→ <b>Y359</b> →Q30→ <u>A31</u> →T10→Y260→A256→Y251→R233→ D234→F135→L136→F140	W461→P460→ <u>C477</u> →R395→N391→ K388→D384→Y357→ <b>R380</b> →Y358→ <b>Y359</b> → <b>L26</b> →H24→P14→D52→H54→Y237→P236 →M134→L136→F140
VI	$G_i^{stat} \geq 0.55$ , $D_{ij} \leq 9.0 \text{ \AA}$	W461→P460→ <b>V455</b> →N452→K388→D384 →Y357→ <b>R380</b> →Y358→ <b>Y359</b> → <b>L26</b> → H24→P14→D52→H54→Y237→M134→ L136→F140	W461→P460→R395→N391→K388→D384→ Y357→ <b>R380</b> →Y358→ <b>Y359</b> → <b>L26</b> →H24→P14 →D52→H54→Y237→M134→L136→F140

**Table 4.** Probable pathways of communication between W461 and F140 in the Ec MetRS. Residues in bold are coevolved and the rest are evolutionarily conserved. The residues that are different between MD and NMA predicted pathways are underlined. Values of the two parameters,  $C_{ij}$  and  $CDC_{ij}$ , are set to greater than or equal to 0.8 and 0.4, respectively.

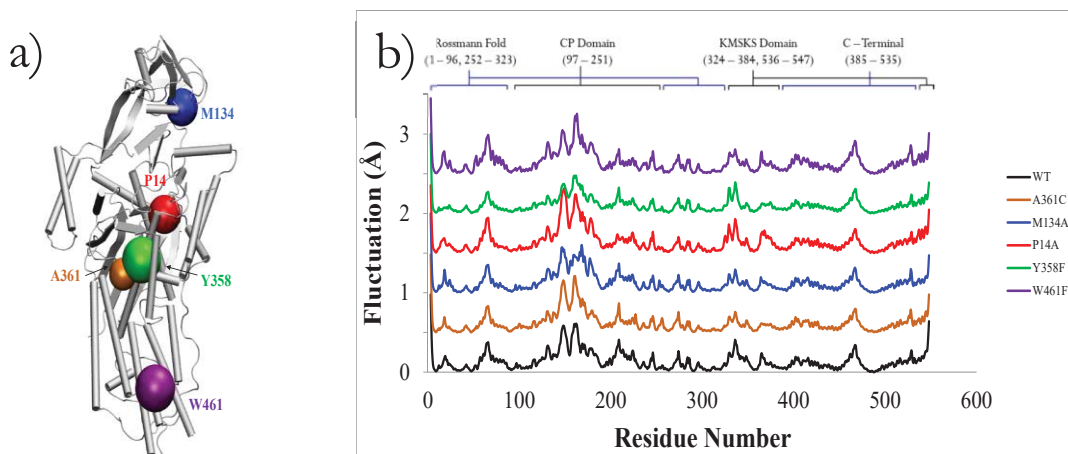
Several interaction networks were identified between W461 and M134/F140 by using various cutoff values for  $\Delta G_i^{stat}$  and  $D_{ij}^o$ . Six contiguous interaction networks between C-terminal domain and CP domain are listed in **Tables 3 and 4**. These pathways (residue interaction networks) between W461 to M134/140 either pass through the  $\alpha$ -helix bundle (residues 385 to 535; pathways I and II) or the helix-loop-strand-helix motif (residues 352 to 385; pathways III - VI) (**Figure 10**). Very similar results were almost identical (**Figure 10 and Tables 3 and 4**). Moreover, analysis of SCA results revealed that conserved residues are dominant over coevolved residues in these predicted pathways (**Table 3 and 4**).



**Figure 10.** Representation of residue-residue interaction networks between the anticodon binding domain (W461) and the CP domain identified in this study. The two terminuses, W461 and M134/F140, are shown in yellow space-filling surface representation. The secondary structure is shown in cartoon representation. From left to right, contiguous pathways as predicted by a) NMA-SCA (12 pathways, Tables 3 and 4), b) MD-SCA (12 pathways, Tables 3 and 4) and c) MD-PSN (4 pathways, reference (33)) are shown in space-filling surface representations.

*Essential dynamics analysis and role of pathway residues*

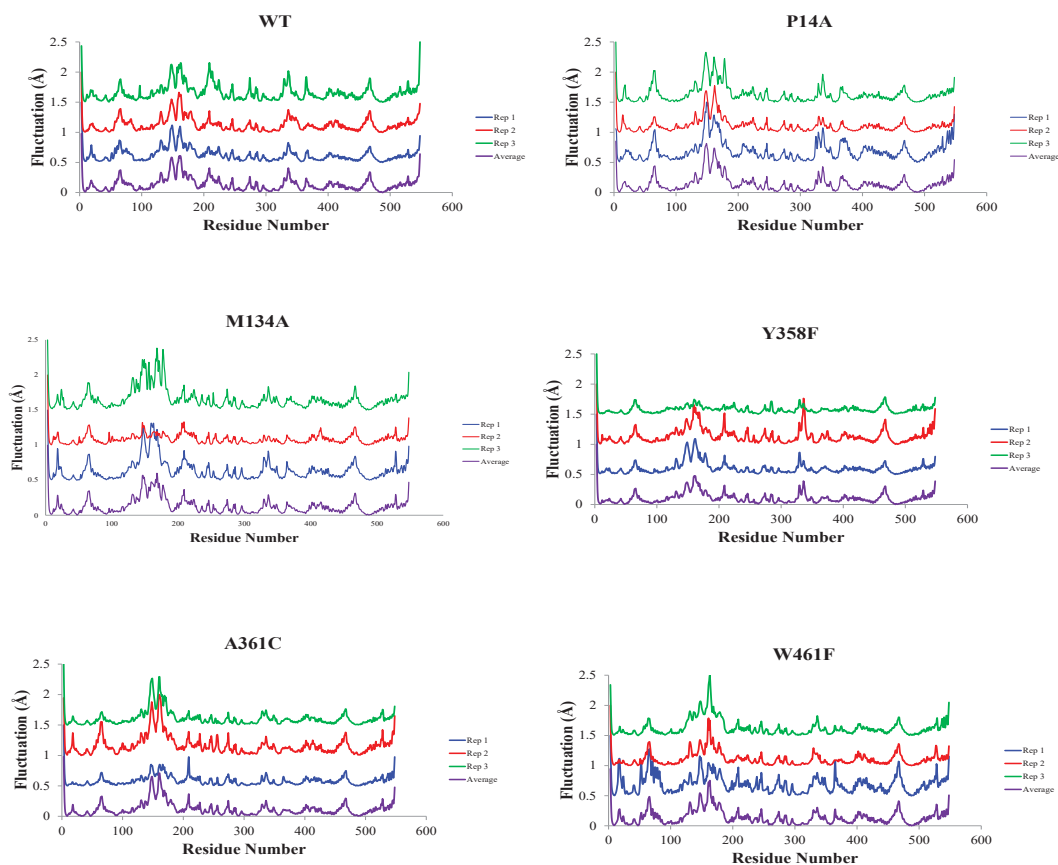
Recent studies have shown that the principal components obtained from long duration simulations (200ns) provide very similar description of the conformational change as that obtained from short duration simulations (10 ns).<sup>22</sup> In this study, we have performed 30 ns simulations for the WT and various mutants to see the impact of mutation of pathway residues on the distant domain dynamics. Essential dynamics analysis of the WT protein was carried out using last 25 ns of 30 ns MD simulation data. The RMSF analysis reveals a flexible CP domain (residues 97-251, **Fig. 11**), similar to what was observed in coarse-grained (NMA) analysis (**Figure 4**). The CP domain, which is adjacent to the catalytic domain, is known to undergo conformational change upon



**Figure 11.** a) The 3D structure of Ec MetRS depicting the 5 sites of mutation, shown in color-coded beads, used for essential dynamics analysis. b) The replica-averaged root-mean-square fluctuations (RMSF) of individual amino acids of WT MetRS and the five variants. The RMSF of C $\alpha$  atoms calculated from the time-averaged structures over the last 25 ns of MD trajectories are shown. The calculated propagated uncertainties are 0.034 Å for WT, 0.049 Å for P14A, 0.046 Å for M134A, 0.037 Å for Y358F, 0.046 Å for A361C, and 0.055 Å for W461F.

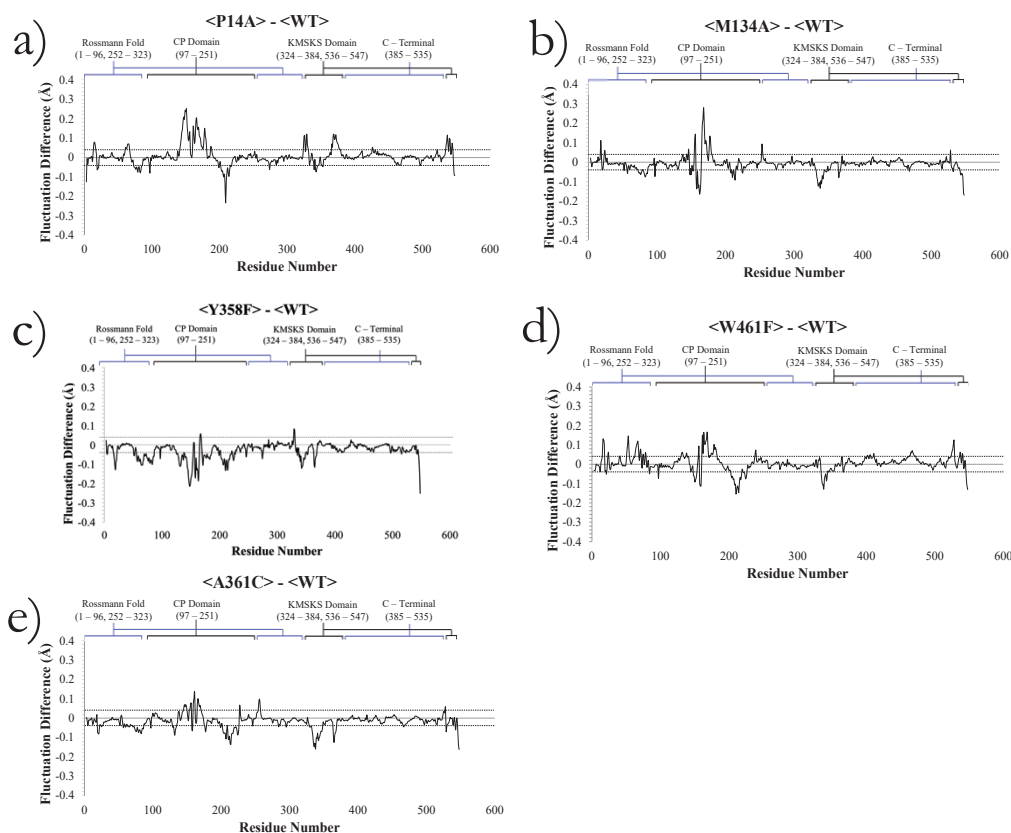
tRNA binding at the anticodon domain and plays an important role in tRNA aminoacylation by guiding the acceptor stem towards the active site.<sup>24,33</sup> Therefore, we hypothesized that a mutation of residues in the predicted pathways would cause a change in the CP domain dynamics. In the subsequent analysis, therefore, we investigated the dynamics of the CP domain (residues 97-251, shown in green in **Figure 1** in response to mutations of a few selected residues in the predicted paths using MD-PCA analysis.

MD simulations were performed in triplicate for the WT and 5 mutants, namely, P14A, M134A, Y358F, A361C, and W461F for a period of 30 ns. These mutational sites are evenly distributed along the communication pathways (Figure 11a). The four mutants - P14A, M134A, Y358F, and W461F, were experimentally found to have significant impact in MetRS function<sup>24,65</sup> (personal communication with Dr. R. W. Alexander). In addition, A361, which is only present in the NMA-derived pathway (Pathway 3) and was experimentally observed to have slight impact on the catalysis, was also chosen for this study.<sup>66</sup> The stability of the dynamics was first evaluated by computing the RMSD of  $C_{\alpha}$  atoms along the simulated time (Figure 7). A sharp change of RMSD was observed in the initial 500-700 ps. Essential dynamics analysis were conducted on the 5-30 ns data, where the RMSDs were within 1.0 Å. The last 25 ns of MD simulation data were used to assess the quality of simulations by computing the RMSF of each amino acid from the time-averaged structure. The RMSF for each replica, as well as the replica-averaged fluctuations of the WT and the five MetRS variants, are reported in Figure 12 and Figure 11b, respectively. The RMSF data demonstrates that the backbone flexibilities are quite reproducible for each of these protein systems, with only a propagated uncertainty of 0.03–0.06 Å for the three replica simulations (Figure 11b). These results indicate that all simulations have reached equilibrated states.



**Figure 12.** Root-mean-square fluctuations of individual amino acids for WT MetRS and five variants. In each stacked plot, the rms fluctuations of  $C_{\alpha}$  atoms calculated from the time-averaged structures over the last 25 ns of MD trajectories are shown. Fluctuations for the three replicas are separated by 0.5 Å and color-coded for the sake of clarity: blue for replica 1, red for replica 2, and green for replica 3. In each case, the bottom plot (purple) represents the replica-averaged rms fluctuations. The calculated propagated uncertainties are 0.034 Å for WT, 0.049 Å for P14A, 0.046 Å for M134A, 0.037 Å for Y358F, 0.046 Å for A361C, and 0.055 Å for W461F.

Next, we probed the impact of these mutations on the collective dynamics of the Ec MetRS. Following the method described in our earlier work,<sup>35</sup> PCA was conducted using the combined 75 ns trajectory (last 25 ns of three replicates) for each protein system. The first three clusters representing the predominant collective dynamics were extracted. The RMSF of  $C_{\alpha}$  atoms were computed from their respective average structures, normalized, and averaged over the three clusters.<sup>35</sup> The impact of mutation of on-pathway residues on the protein flexibility was examined by computing the difference of these cluster-averaged RMSF between the WT and a specific MetRS variant (**Figure 13**). The RMSF analysis demonstrates that the overall flexibility of the protein backbone was altered due to the mutation of on-pathway residues.



**Figure 13.** Changes in the normalized RMSF of  $C_{\alpha}$  atoms observed in the collective dynamics of the five mutants with respect to WT Ec MetRS. The angular bracket indicates that the RMSF are averaged over the first three clusters representing the predominant collective protein motions. The propagated uncertainties for each of these plots are within 0.04 Å and are shown with two parallel dotted lines.

The RMSF analysis revealed that mutations of these “on-pathways” residues have altered the dynamics of the CP domain and other structural elements of this multi-domain protein. For example, alanine substitution of P14 resulted in increased dynamics of residues 324 to 336 and residues 364 to 378 of KMSKS domain (**Figure 13a**); residues of KMSKS domain are important for stabilizing methionyl-adenylate. Also, an increase in the CP domain dynamics was observed. However a sharp decrease in the fluctuation of the residue 209 of the CP domain was noticed in P14A variant.

Mutation of M134 to alanine also demonstrated an increase in CP domain dynamics (**Figure 13b**). The RMSF analysis indicates decreased dynamics of the catalytically important Rossmann-fold domain. However, a small increase in flexibility of W253 of the methionine binding pocket was noticed. Decreased dynamics of the KMSKS domain and C-terminal domain was also observed. In the case of Y358F, an overall decrease in the dynamics of virtually every domain was observed (**Figure 13c**). On the other hand, both increased and decreased fluctuations of various domains were noticed in the case of W461F (**Figure 13d**). Lastly, in the case of A361C mutant, where small impact in catalysis was experimentally observed, the fluctuation of the protein backbone was also altered by the alanine substitution (**Figure 13e**). Taken together, significant alternations in the dynamics of various structural elements were observed for these MetRS variants. These computational analyses demonstrated that mutations along the predicted pathways between the C-terminal and CP domain of MetRS do indeed alter the backbone fluctuations of the distant domains and secondary elements, which might result in reduced catalytic efficiency.

## DISCUSSION

### *Protein dynamism and evolvability*

Dynamism is an intrinsic property of a protein, an inseparable element from its function, and encrypted in its primary structure. Therefore, amino acid residues that are critical for maintaining the protein's intrinsic functional dynamics should either remain conserved or their mutations will be correlated. The present study revealed the existence of multiple interaction networks, involving dynamically and evolutionarily coupled residues, between the CP domain and the C-terminal domain. Scrutiny of these networks of interactions showed that the involved residues are predominantly conserved in this protein family (**Table 3**). Only a small fraction of residues (10-26%, **Tables 3 and 4**) were found to be coevolving and these residues appeared to be important for forming a contiguous network of residue interactions between W461 and M134/F140. Thus, the present findings demonstrated that conserved residues are the key players in regulating distant domain dynamics, while the role of coevolved residues is only complementary. This is consistent with the observations in myosin motor protein<sup>67</sup> and other protein systems<sup>49,35</sup> indicating that integration of evolutionary information with dynamic coupling data is an important criterion for identifying interaction networks between distant functional sites.

### *Cooperative dynamics between the two functional sites*

Previous biochemical and computational studies have established that the cognate anticodon binding at the C-terminal triggers the conformational change of the CP domain, which facilitates the binding of tRNA acceptor stem at the active site.<sup>24,33</sup> Moreover, it has also been reported that the anticodon-triggered conformational change, which is important for efficient tRNA<sup>Met</sup> aminoacylation, propagates through the enzyme, not through the tRNA.<sup>33</sup> Therefore, if the identified residue networks facilitate coupled-domain dynamics, then a mutation along these predicted pathways is expected to have an observable impact on the dynamics of the CP domain. This hypothesis was tested by the combined essential dynamics analysis performed on various mutant protein systems. The four mutants (P14A, M134A, Y358F, and W461F), which are resided on the residue interaction networks between C-terminal and CP domains, were found to have significant impact on catalysis. Also, mutation of A361 to cysteine has impact on the distant domain dynamics. The pair-wise RMSF comparisons of the WT and mutant variants portray the notable effects of discrete mutations along the interaction networks on the distant CP domain dynamics (**Figure 13**). The variation in the CP domain dynamics is somewhat related to the distance between the site of mutation and the CP domain. Taken together, the *in silico* mutational study illustrates the role of predicted interaction networks in maintaining the distant domain dynamics.

### *Existing theoretical and experimental results*

The residue interaction networks (pathways I–VI) identified in the present bioinformatics study also bears a close similarity to previously reported communication pathways (between W461 to L13) obtained from the atomistic simulations and protein structure networks (PSN) analysis.<sup>33,68</sup> However, unlike the previously reported pathways, pathways identified in the present study are contiguous. The involvement of the  $\alpha$ -helix bundle, which encompasses pathways I and II, as well as the helix-loop-strand-helix motif (pathways III–VI) in domain-domain communication of Ec MetRS has also been supported by earlier studies.<sup>66</sup> Therefore, the revelation of similar residue-residue interaction networks, as obtained from the use of two different strategies – i) MD and PSN<sup>33</sup> and ii) MD/NMA and SCA, strongly suggests that the bioinformatics-based STCA method could be used as an alternative, fast yet robust method of predicting long-range communication pathways in multi-domain proteins.

Results obtained in the present study also enable us to explain some of the previously reported experimental mutational results. Previous mutational studies have showed that the mutation of P460, N452, R395, N391 and R233 cause functional defects.<sup>24</sup> These residues actually belong to the predicted interaction networks and appeared to be important for promoting coupled-domain dynamics, which are essential for enzymatic function. Also, it has been reported that the mutation of N452 and N387 to alanine resulted in functionally defective mutants.<sup>69</sup> Our studies show that these two residues are within the range of H-bonding interaction with K388, the residue present in three of the six pathways.

### *Coarse-grained NMA and all-atom MD simulations*

Coarse-grained simulations have emerged as valuable tools for studying conformational changes in large biomolecules. Despite the fact that NMA implies infinitesimal displacements near the local energy minimum, recent studies have demonstrated that micro- to millisecond global motions can be modeled successfully with coarse-grained model.<sup>70</sup> We have also demonstrated that coarse-grained NMA is comparable to the all-atom MD simulation in depicting the intrinsic global dynamics of AARSs including Ec MetRS.<sup>64,71</sup> However, local fluctuations upon substrate binding were poorly captured by the coarse-grained simulations.<sup>71</sup> Although coarse-grained NMA failed to capture time-dependent fluctuations, it has been well-documented that NMA can provide relevant information regarding functional motions and allosteric mechanisms.<sup>22,72-75</sup> Surprisingly, very similar results were obtained from both all-atoms MD simulation and coarse-grained NMA in the present study; four out of six pathways identified by MD and NMA are almost identical (**Tables 3 and 4**). The close similarities between the all-atom and coarse-grained simulations results

suggested that for large biomolecules like AARSs, computationally less intensive coarse-grained NMA could be used to trace the conformational transition (“population-shift”) pathways i.e. residue interaction networks between functional sites.

## CONCLUSIONS

In the present study, pathways of inter-domain communication in Ec MetRS were identified by integrating evolutionary information with that of dynamic coupling. Several pathways (residue interaction networks) were identified through which local perturbation could propagate between two functional sites, 53 Å apart. Residues identified in these pathways are predominantly conserved and are also physically proximate in the structure. These residues are engaged in strong correlated dynamics. Long-duration MD simulation study followed by essential dynamics analysis provided evidence that these residues are important in maintaining protein dynamics and their mutations are capable of altering the dynamics of protein segments over great distances, even across domain interfaces. Therefore, the present STCA method shows promise in identifying and exploring residues that mediate long-range inter-domain communications in large protein systems.

The domain dynamics occur in micro- to millisecond timescale. Therefore, the atomistic simulation of a large protein is quite challenging. Although several hundred nanoseconds of simulations can be performed using the more advance computing systems, it remains quite challenging to simulate functionally important long-timescale collective dynamics of a large protein like Ec MetRS. Therefore, the use of coarse-grained method has substantially reduced the time for obtaining the information of thermal coupling from the collective dynamics of a large protein system like Ec MetRS. However, the coarse-grained simulation does not provide any information about the type of residues involved in coupled motions and statistical coupling analysis complements this by providing the data of residue-residue compatibility. Therefore integration of evolutionary information with dynamic features of residues could enable one to identify long-range interaction networks that are contiguous and critical for maintaining coupled dynamics in modular proteins.

Taken together, the present STCA study has enabled us to identify a set of residues that are potentially involved in maintaining the coupled dynamics among domains in Ec MetRS. The present study also suggested that in order to facilitate long-range communication, multi-domain proteins like Ec MetRS use parallel pathways of residue interactions. This information can be used as a guide to explore more about these interaction networks and the role of intrinsic dynamics in the function of Ec MetRS through spectroscopic, mutational, and theoretical studies. In addition, the STCA method can be extended to other multi-domain proteins to gain molecular-level understanding of the domain-domain communications. Moreover, the present study suggested that the correlated motions derived from NMA can also provide insight into long-range inter-domain communication.

## ACKNOWLEDGEMENT

Authors thank Dr. Ranganathan (University of Texas, Southwestern Medical Center, Dallas) for the MATLAB scripts used in the SCA. We would also like to thank Mr. Alexander Jerome Greene for his initial contributions in developing the STCA script.

## REFERENCES

1. Bu, Z., Biehl, R., Monkenbusch, M., Richter, D., and Callaway, D. J. (2005) Coupled protein domain motion in Taq polymerase revealed by neutron spin-echo spectroscopy, *Proc. Natl. Acad. Sci. U. S. A.* 102, 17646-17651.
2. Yu, H., Ma, L., Yang, Y., and Cui, Q. (2007) Mechanochemical coupling in the myosin motor domain. II. Analysis of critical residues, *PLoS Comput. Biol.* 3, e23.
3. Chennubhotla, C., Yang, Z., and Bahar, I. (2008) Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL, *Mol. Biosyst.* 4, 287-292.
4. Daily, M. D., and Gray, J. J. (2009) Allosteric communication occurs via networks of tertiary and quaternary motions in proteins, *PLoS Comput. Biol.* 5, e1000293.
5. Fidelak, J., Ferrer, S., Oberlin, M., Moras, D., Dejaegere, A., and Stote, R. H. (2010) Dynamic correlation networks in human peroxisome proliferator-activated receptor-gamma nuclear receptor protein, *Eur. Biophys. J.* 39, 1503-1512.
6. Zheng, W., Liao, J. C., Brooks, B. R., and Doniach, S. (2007) Toward the mechanism of dynamical couplings and translocation in hepatitis C virus NS3 helicase using elastic network model, *Proteins* 67, 886-896.
7. Weinkam, P., Pons, J., and Sali, A. (2012) Structure-based model of allostery predicts coupling between distant sites, *Proc. Natl. Acad. Sci. U. S. A.* 109, 4875-4880.
8. del Sol, A., Tsai, C. J., Ma, B., and Nussinov, R. (2009) The origin of allosteric functional modulation: multiple pre-existing pathways, *Structure* 17, 1042-1050.
9. Tsai, C. J., del Sol, A., and Nussinov, R. (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms, *Mol. Biosyst.* 5, 207-216.
10. Changeux, J. P., and Edelstein, S. (2011) Conformational selection or induced fit? 50 years of debate resolved, *F1000 Biol. Rep.* 3, 19.

11. Lee, Y., Mick, J., Furdai, C., and Beamer, L. J. (2012) A coevolutionary residue network at the site of a functionally important conformational change in a phosphohexomutase enzyme family, *PLoS One* 7, e38114.
12. Nussinov, R., Tsai, C. J., and Ma, B. (2013) The underappreciated role of allostery in the cellular network, *Annu. Rev. Biophys.* 42, 169-189.
13. Tsai, C. J., and Nussinov, R. (2014) A unified view of "how allostery works", *PLoS Comput. Biol.* 10, e1003394.
14. Gunasekaran, K., Ma, B., and Nussinov, R. (2004) Is allostery an intrinsic property of all dynamic proteins?, *Proteins* 57, 433-443.
15. 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.
16. Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery, *Nat. Struct. Mol. Biol.* 13, 831-838.
17. Daily, M. D., and Gray, J. J. (2007) Local motions is a benchmark of allosteric proteins, *Proteins* 67, 385-399.
18. Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families, *Science* 286, 295-299.
19. Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins, *Nat. Struct. Biol.* 10, 59-69.
20. Hatley, M. E., Lockless, S. W., Gibson, S. K., Gilman, A. G., and Ranganathan, R. (2003) Allosteric determinants in guanine nucleotide-binding proteins, *Proc. Natl. Acad. Sci. U. S. A.* 100, 14445-14450.
21. Bahar, I., and Rader, A. J. (2005) Coarse-grained normal mode analysis in structural biology, *Curr. Opin. Struct. Biol.* 15, 586-592.
22. Skjaerven, L., Martinez, A., and Reuter, N. (2011) Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit, *Proteins* 79, 232-243.
23. 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.
24. Blanquet, S., Crepin, T., Mechulam, Y., and Schmitt, E. (2005) Methionyl-tRNA Synthetase, *Landes Biosciences/Eurekah.com*, Georgetown, TX
25. Martinis, S. A., and Schimmel, P. (1992) Enzymatic aminoacylation of sequence-specific RNA minihelices and hybrid duplexes with methionine, *Proc Natl Acad Sci U S A* 89, 65-69.
26. Martinis, S. A., and Schimmel, P. (1993) Microhelix aminoacylation by a class I tRNA synthetase. Non-conserved base pairs required for specificity, *J Biol Chem* 268, 6069-6072.
27. Alexander, R. W., Nordin, B. E., and Schimmel, P. (1998) Activation of microhelix charging by localized helix destabilization, *Proc. Natl. Acad. Sci. U. S. A.* 95, 12214-12219.
28. Meinel, T., Mechulam, Y., Blanquet, S., and Fayat, G. (1991) Binding of the anticodon domain of tRNA(fMet) to Escherichia coli methionyl-tRNA synthetase, *J. Mol. Biol.* 220, 205-208.
29. Ghosh, G., Kim, H. Y., Demaret, J. P., Brunie, S., and Schulman, L. H. (1991) Arginine-395 is required for efficient in vivo and in vitro aminoacylation of tRNAs by Escherichia coli methionyl-tRNA synthetase, *Biochemistry* 30, 11767-11774.
30. Ghosh, G., Pelka, H., and Schulman, L. H. (1990) Identification of the tRNA anticodon recognition site of Escherichia coli methionyl-tRNA synthetase, *Biochemistry* 29, 2220-2225.
31. Meinel, T., Mechulam, Y., Le Corre, D., Panvert, M., Blanquet, S., and Fayat, G. (1991) Selection of suppressor methionyl-tRNA synthetases: mapping the tRNA anticodon binding site, *Proc. Natl. Acad. Sci. U. S. A.* 88, 291-295.
32. 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* 68, 670-689.
33. 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.
34. Weimer, K. M., Shane, B. L., Brunetto, M., Bhattacharyya, S., and Hati, S. (2009) Evolutionary basis for the coupled-domain motions in Thermus thermophilus leucyl-tRNA synthetase, *J. Biol. Chem.* 284, 10088-10099.
35. Johnson, J. M., Sanford, B. L., Strom, A. M., Tadayon, S. N., Lehman, B. P., Zirbes, A. M., Bhattacharyya, S., Musier-Forsyth, K., and Hati, S. (2013) Multiple pathways promote dynamical coupling between catalytic domains in Escherichia coli prolyl-tRNA synthetase, *Biochemistry* 52, 4399-4412.
36. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graph.* 14, 33-38, 27-38.
37. Bahar, I., Atilgan, A. R., and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, *Fold Des.* 2, 173-181.
38. Eyal, E., Yang, L. W., and Bahar, I. (2006) Anisotropic network model: systematic evaluation and a new web interface, *Bioinformatics* 22, 2619-2627.
39. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26, 1781-1802.



40. MacKerell, A. D. J., Bashford, D., Bellott, M., Dunbrack, R. L. J., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Gou, J., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E. I., Roux, B., Schelenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) Allatom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102, 3586.
41. Glykos, N. M. (2006) Software news and updates. Carma: a molecular dynamics analysis program, *J. Comput. Chem.* 27, 1765-1768.
42. 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.
43. Tama, F., and Sanejouand, Y. H. (2001) Conformational change of proteins arising from normal mode calculations, *Protein Eng.* 14, 1-6.
44. Dijkstra, E. W. (1959) A note on two problems in connexion with graphs, *Numerische Mathematik* 1, 269-271.
45. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25, 3389-3402.
46. Yang, L. W., and Chng, C. P. (2008) Coarse-grained models reveal functional dynamics--I. Elastic network models--theories, comparisons and perspectives, *Bioinform Biol Insights* 2, 25-45.
47. Skjaerven, L., Hollup, S., and Reuter, N. (2009) Normal mode analysis for protein, *J. Mol. Chem. THEOCHEM* 898, 42-48.
48. Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes, *Structure* 13, 373-380.
49. Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model, *Biophys J* 80, 505-515.
50. 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.
51. Ryckaert, J. P., Ciotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23, 327-341.
52. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An N. Log(N) Method for Ewald Sums in Large Systems, *J. Chem. Phys.* 98, 10089-10092.
53. Berendsen, H. J. C., Postma, J. P., van Gunsteren, M. W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81, 3684-3690.
54. Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method, *J. Chem. Phys.* 103, 4613-4621.
55. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms, *J. Chem. Phys.* 101, 4177-4189.
56. Bhattacharyya, S., Ma, S., Stankovich, M. T., Truhlar, D. G., and Gao, J. (2005) Potential of mean force calculation for the proton and hydride transfer reactions catalyzed by medium-chain acyl-CoA dehydrogenase: effect of mutations on enzyme catalysis, *Biochemistry* 44, 16549-16562.
57. Rauschnot, J. C., Yang, C., Yang, V., and Bhattacharyya, S. (2009) Theoretical determination of the redox potentials of NRH:quinone oxidoreductase 2 using quantum mechanical/molecular mechanical simulations, *J. Phys. Chem. B* 113, 8149-8157.
58. Sanford, B., Cao, B. V., Johnson, J. M., Zimmerman, K., Strom, A. M., Mueller, R. M., Bhattacharyya, S., Musier-Forsyth, K., and Hati, S. (2012) Role of coupled-dynamics in the catalytic activity of prokaryotic-like prolyl-tRNA synthetases *Biochemistry* 51, 2146-2156.
59. Roy, J., and Laughton, C. A. (2010) Long-timescale molecular-dynamics simulations of the major urinary protein provide atomistic interpretations of the unusual thermodynamics of ligand binding, *Biophys. J.* 99, 218-226.
60. van Aalten, D. M., Amadei, A., Linszen, A. B., Eijssink, V. G., Vriend, G., and Berendsen, H. J. (1995) The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water, *Proteins* 22, 45-54.
61. Mueller, R. M., North, M. A., Yang, C., Hati, S., and Bhattacharyya, S. (2011) Interplay of flavin's redox states and protein dynamics: an insight from QM/MM simulations of dihydronicotinamide riboside quinone oxidoreductase 2, *J. Phys. Chem. B* 115, 3632-3641.
62. Silvan, L. F., Wang, J., and Steitz, T. A. (1999) Insights into editing from an ile-tRNA synthetase structure with tRNA<sup>ile</sup> and mupirocin, *Science* 285, 1074-1077.
63. Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S., and Cusack, S. (2005) The crystal structure of leucyl-tRNA synthetase complexed with tRNA<sup>Leu</sup> in the post-transfer-editing conformation, *Nat. Struct. Mol. Biol.* 12, 923-930.
64. Warren, N., Strom, A., Nicolet, B., Albin, K., Albrecht, J., Bausch, B., Dobbe, M., Dudek, M., Firgens, S., Fritsche, C., Gunderson, A., Heimann, J., Her, C., Hurt, J., Konorev, D., Lively, M., Meacham, S., Rodriguez, V., Tadayon, S., Trcka, D.,

- Yang, Y., Bhattacharyya, S., and Hati, S. (2014) Comparison of the Intrinsic Dynamics of Aminoacyl-tRNA Synthetases, *Protein J.* 33, 184-198.
65. Alexander, R. W., and Schimmel, P. (2001) Domain-domain communication in aminoacyl-tRNA synthetases, *Prog. Nucleic Acid Res. Mol. Biol.* 69, 317-349.
66. Banerjee, P., Warf, M. B., and Alexander, R. (2009) Effect of a domain-spanning disulfide on aminoacyl-tRNA synthetase activity, *Biochemistry* 48, 10113-10119.
67. Tang, S., Liao, J. C., Dunn, A. R., Altman, R. B., Spudich, J. A., and Schmidt, J. P. (2007) Predicting allosteric communication in myosin via a pathway of conserved residues, *J Mol Biol* 373, 1361-1373.
68. Ghosh, A., and Vishveshwara, S. (2008) Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes, *Biochemistry* 47, 11398-11407.
69. Alexander, R. W., and Schimmel, P. (1999) Evidence for breaking domain-domain functional communication in a synthetase-tRNA complex, *Biochemistry* 38, 16359-16365.
70. Gur, M., Zomot, E., and Bahar, I. Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions, *J Chem Phys* 139, 121912.
71. Strom, A. M., Fehling, S. C., Bhattacharyya, S., and Hati, S. (2014) Probing the global and local dynamics of aminoacyl-tRNA synthetases using all-atom and coarse-grained simulations, *J. Mol. Model.* 20, 2245.
72. Cui, Q., Li, G., Ma, J., and Karplus, M. (2004) A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase, *J. Mol. Biol.* 340, 345-372.
73. Mouawad, L., and Perahia, D. (1996) Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures, *J Mol Biol* 258, 393-410.
74. Ma, J., and Karplus, M. (1998) The allosteric mechanism of the chaperonin GroEL: a dynamic analysis, *Proc. Natl. Acad. Sci. U. S. A.* 95, 8502-8507.
75. Zheng, W., Brooks, B. R., and Thirumalai, D. (2007) Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations, *Biophys. J.* 93, 2289-2299.

#### ABOUT THE STUDENT AUTHOR

Ryan Andrews recently graduated from the University of Wisconsin – Eau Claire and is currently a PhD student at Iowa State University in the department of Biochemistry, Biophysics and Molecular Biology where he continues to do similar research he found a passion for while an undergraduate.

#### PRESS SUMMARY

Proteins are the machinery of all living cells, and research is conducted every day to further our understanding of how they work. This paper compares two computational methods' ability to study how different parts, or domains, of an individual protein "communicate" with each other, a phenomenon known as inter-domain communication. We found that the less computationally demanding process known as coarse-grained analysis was comparable to the more demanding (though more theoretically accurate) process known as atomistic molecular dynamics in investigating inter-domain communications.