

Allosteric Mechanism of Human Mitochondrial Phenylalanyl-tRNA Synthetase: An Atomistic MD Simulation and a Mutual Information-Based Network Study

Qi Shao, Zhongjie Han, Jingmin Cheng, Qiankun Wang, Weikang Gong, and Chunhua Li*



Cite This: <https://doi.org/10.1021/acs.jpcb.1c03228>



Read Online

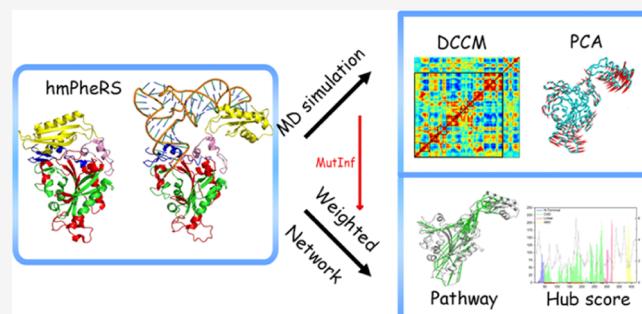
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Aminoacyl-tRNA synthetases (aaRSs), a family of ubiquitous and essential enzymes, can bind target tRNAs and catalyze the aminoacylation reaction in genetic code translation. In this work, we explore the dynamic properties and allosteric communication of human mitochondrial phenylalanyl-tRNA synthetase (hmPheRS) in free and bound states to understand the mechanisms of its tRNA^{Phe} recognition and allostery using molecular dynamics simulations combined with the torsional mutual information-based network model. Our results reveal that hmPheRS's residue mobility and inter-residue motional coupling are significantly enhanced by tRNA^{Phe} binding, and there occurs a strong allosteric communication which is critical for the aminoacylation reaction, suggesting the vital role of tRNA^{Phe} binding in the enzyme's function. The identified signaling pathways mainly make the connections between the anticodon binding domain (ABD) and catalytic domain (CAD), as well as within the CAD composed of many functional fragments and active sites, revealing the co-regulation role of them to act coordinately and achieve hmPheRS's aminoacylation function. Besides, several key residues along the communication pathways are identified to be involved in mediating the coordinated coupling between anticodon recognition at the ABD and activation process at the CAD, showing their pivotal role in the allosteric network, which are well consistent with the experimental observation. This study sheds light on the allosteric communication mechanism in hmPheRS and can provide important information for the structure-based drug design targeting aaRSs.



INTRODUCTION

The aminoacyl-tRNA synthetases (aaRSs), a family of ubiquitous and essential enzymes for protein biosynthesis, catalyze the aminoacylation of specific amino acids on their cognate tRNAs, enabling the genetic code translation.¹ Mutations in aaRSs are associated with various diseases such as cancer, neuronal pathologies, and autoimmune disorders.² Human mitochondrial phenylalanyl-tRNA synthetase (hmPheRS), due to owning a minimum set of structural domains capable of aminoacylation, attracts wide attention from researchers.^{3,4} Experiments found that certain mutations in the hmPheRS anticodon binding domain (ABD) can significantly affect its catalytic efficiency at the active sites although the two regions are separated by up to 80 Å.^{3,4} Currently, the molecular mechanism of the long-range allosteric communication between the two regions and the reason why these residue mutations can affect the enzyme function are not completely clear. Thus, exploring the allosteric transition and accurately identifying the key residues involved in the allosteric modulation are important for the understanding of the hmPheRS working mechanism and the structure-based drug design.

HmPheRS with a single chain is the smallest known member of class II aaRS. The structures of hmPheRS in tRNA-bound and -free states have been experimentally solved, which are composed of four major structural blocks: N-terminal (residues 1–47), catalytic domain (CAD, residues 48–289), linker (residues 290–322), and ABD (residues 323–415).⁷ The ABD undergoes a 160° hinge-type rotation upon tRNA^{Phe} binding, and furthermore, the enzyme's whole conformation turns from the inactive "close" to active "open" state (Figure 1). In the complex, the ABD typically establishes contacts with the tRNA anticodon loop with its anticodon recognition sites mainly including Asp364, Phe366, Ser375, and Arg414, while the CAD binds tRNA 3'-CCA end sequence and phenylalanine, which synthesizes phenylalanyl adenylate and transfers phenylalanine to the tRNA. HmPheRS ensures accurate tRNA

Received: April 10, 2021

Revised: June 9, 2021

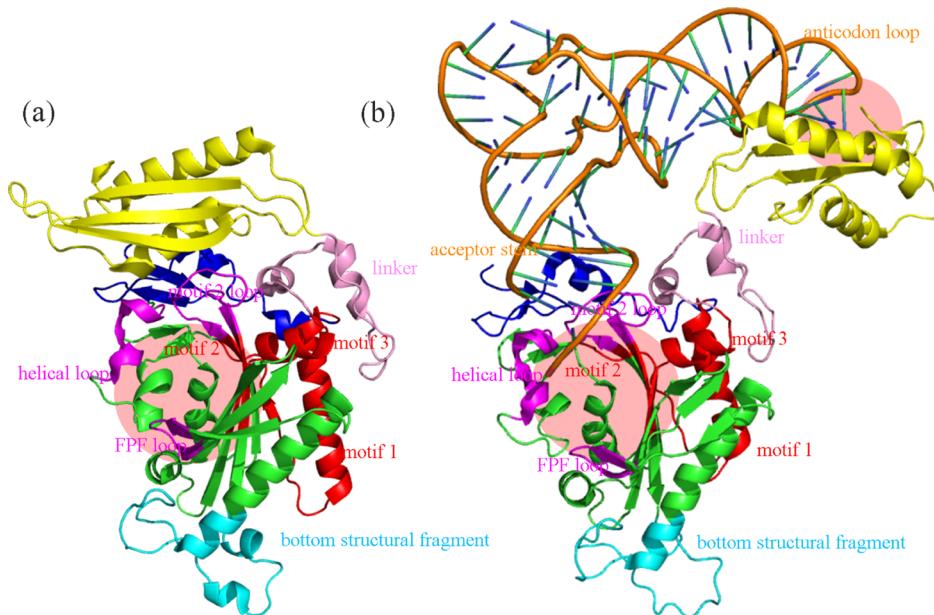


Figure 1. Crystal structures of hmPheRS in tRNA^{Phe}-free (a) and tRNA^{Phe}-bound states (b). The N-terminal, CAD, linker, and ABD are colored blue, green, pink, and yellow, respectively. In the CAD, motifs 1–3, characteristic loops (helical loop, motif 2 loop, and FPF loop) and the bottom structural fragment are colored red, magenta, and cyan respectively. The active sites and anticodon recognition sites are marked with a transparent ball. tRNA^{Phe} is shown in orange.

recognition and aminoacylation through coordinating the movements of the ABD and CAD.⁸ The CAD, due to its catalytic role, has many functionally important segments. Among the three specific motifs, motif 1 (residues 48–82) was classically thought to be involved in subunit association, but its function in the monomer remains unknown;⁴ motif 2 (residues 139–160) and motif 3 (residues 276–288) are involved in the active site formation, and the former (mainly its loop) plus a helical loop and the N-terminal is responsible for binding tRNA 3'-CCA end sequence like a clamp⁸ (Figure 1b). Additionally, three functional loops, a helical loop (residues 99–108), motif 2 loop (residues 141–155, a part of motif 2), and FPF loop (residues 230–238), are all involved in closing up the active sites, and the latter is also responsible for binding phenylalanine.⁸ For the motif 2 loop, its another role is to help the transition of hmPheRS between the close and open conformations.⁴ Besides, a long insertion sequence between motifs 2 and 3 creates a structural fragment (residues 165–202, called the bottom structural fragment here), whose function remains unclear completely.⁷

Allostery plays a fundamental role in innumerable biological processes in all organisms.⁹ The investigation of allosteric mechanisms is a current research hotspot^{10–13} and expedites the discovery of allosteric drugs.^{14,15} However, it is difficult for experimental methods to explore allosteric communication in proteins. Theoretically, many computational methods have been developed to investigate the issue. One of the most widely used methods has been the statistical coupling analysis (SCA) pioneered by Ranganathan and co-workers^{16,17} who utilized statistical methods on the sequential evolution data to identify allosteric pathways. Despite meaningful results achieved, it needs large-scale alignment of homologous sequences to obtain statistically significant evolutionary coupling, which is a major limitation for the general application of the SCA-based approaches. Other techniques involve the elastic network models combined with structural perturbation

methods,¹⁸ which provide valuable information on which residues contribute largely to the communication. These methods are topology-based and coarse-grained and often lose some atomic details. The atomistic-level methods mainly start with obtaining the movement-related information from molecular dynamics (MD) simulation trajectories. Bowerman detected thrombin's allosteric networks using MD simulations where the pairwise residue motional correlations are combined with the contact information to identify allosteric pathways.¹⁹ Kumawat et al. elucidated the molecular basis for the allosteric modulation in the PDZ3 domain by constructing a pairwise residue energetic perturbation network with noncovalent energies extracted from MD trajectories.²⁰ These methods, although obtain the atomistic movement information, introduce high-frequency noises into the correlation calculation due to the use of cartesian coordinates, leading to spurious correlations between residues to some extent. Aiming at the issue, McClendon et al. developed the MutInf method²¹ where the residue mutual information is extracted in internal coordinates (residue torsion angles) rather than in cartesian coordinates, which can identify the significant correlations from multiple short simulations and eliminate high-frequency noises to some extent. Utilizing MutInf, McClendon et al. explored the motional correlations between the residues located at two small-molecule binding sites in human interleukin-2, successfully identifying not only local correlations but also the long-range ones between the key functional sites.²¹ The research indicates that MutInf is an effective method in quantifying the function-related motional correlations and in identifying the key residues that mediate the cross-talk between functional sites.

In this work, we explore the effect of tRNA^{Phe} binding on hmPheRS dynamics using atomistic MD simulations combined with principal component analysis (PCA). The allosteric communication is investigated based on the mutual information obtained using MutInf from MD trajectories.

Finally, we identify the signaling pathways and key mediating residues between the hmPheRS anticodon binding region and the active sites and within the CAD.

MATERIALS AND METHODS

Biomolecular Systems and Molecular Dynamics Simulations.

The crystal structures of the hmPheRS-tRNA^{Phe} complex and hmPheRS monomer were obtained from Protein Data Bank (PDB ID: 3TUP and 3CMQ respectively), on which MD simulations were performed. In the solvation stage, for the former (the latter), about 55,000 (40,000) TIP3P water molecules were added in a cubic periodic box of $123 \times 123 \times 123 \text{ \AA}^3$ ($111 \times 111 \times 111 \text{ \AA}^3$), and the total number of atoms was about 174,000 (130,000). Sodium and chloride ions were added to the systems to get a final ion concentration of 0.15 mol/L. On the systems, two independent MD simulations were performed by using NAMD 2.12²² with CHARMM36 all-atom force field.^{23–25} The SHAKE algorithm was used to constrain the covalent bonds involving hydrogen atoms.²⁶ Van der Waals interactions were calculated using a switching function with twin-range cutoffs of 10 and 12 Å. Electrostatic interactions were computed at every timestep by the particle mesh Ewald method with 1 Å grid size.²⁷ The system pressure was maintained at 1 atm by Langevin piston control with a decay period of 50 fs and a damping time of 25 fs, while the system temperature was kept at 303.15 K using a Langevin thermostat with a damping coefficient of 1 ps⁻¹. Each system was energetically minimized with 20,000 steps. All backbone atoms of solutes were restrained with a harmonic constraint of 0.1 kcal mol⁻¹ Å⁻² and the systems were slowly heated up from 0 to 303.15 K over a period of 1.0 ns. Finally, the non-constrained MD simulations were performed at constant pressure (1 atm) and constant temperature (303.15 K) for 200 ns. The trajectory snapshots were saved every 2.0 ps, and 100,000 conformations were collected for further analyses.

Movement Cross-Correlations of Residues. The movement cross-correlation coefficient C_{ij} between two atoms i and j during the simulation is defined as follows

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle (\Delta r_i \cdot \Delta r_i) \langle (\Delta r_j \cdot \Delta r_j) \rangle}} \quad (1)$$

where vectors Δr_i and Δr_j are the instantaneous displacements of the i th and j th atoms from their mean positions, respectively, and $\langle \rangle$ represents trajectory average. Positively correlated residues move in the same direction, that is, $C_{ij} > 0$, while negatively correlated residues move in the opposite direction, that is, $C_{ij} < 0$. C_{ij} values are calculated for Ca atoms in hmPheRS and P atoms in tRNA^{Phe}.

Principal Component Analysis. The slow and functional motions of biomolecules can be extracted by applying the dimensional reduction methods, and PCA is one of the widely used approaches.²⁸ Based on the movement cross-correlations (unnormalized), the covariance matrix is constructed, and then, it is diagonalized to obtain the orthogonal eigenvectors and corresponding eigenvalues. The eigenvectors are also called the principal components (PCs), which indicate the directions of the concerted motions. The first few PCs usually represent the slow and global functional motions of biomolecules.²⁹ The corresponding eigenvalues describe the magnitude of the motions along the directions. Here, the Bio3d package³⁰ is used to perform the PCA, and Ca atoms in hmPheRS are analyzed by the PCA method.

Calculation of the Torsional Mutual Information. The MutInf method²¹ is used to calculate the torsional mutual information of pairwise residues. An internal coordinate system of bond lengths, bond angles, and torsion angles is utilized to build the protein configurational space. The torsional mutual information between two torsional degrees of freedom is calculated as the difference between their individual entropies and joint entropy

$$\text{MutInf}(x, y) = S(x) + S(y) - S(x, y) \quad (2)$$

For entropy and mutual information-based approaches, the small sample size is notoriously challenging. A corrected histogram entropy estimate³¹ over adaptive partitions³² is adopted as follows

$$\begin{aligned} \text{MutInf} = & \sum_{i=1}^r \frac{n_i}{N} \left(\ln N - \Psi_{n_i} - \frac{(-1)^{n_i}}{n_i + 1} \right) \\ & + \sum_{j=1}^s \frac{n_j}{N} \left(\ln N - \Psi_{n_j} - \frac{(-1)^{n_j}}{n_j + 1} \right) - \\ & \sum_{ij=1}^{rs} \frac{n_{ij}}{N} \left(\ln N - \Psi_{n_{ij}} - \frac{(-1)^{n_{ij}}}{n_{ij} + 1} \right) \end{aligned} \quad (3)$$

where r and s are the numbers of marginal bins; n_i , n_j , and n_{ij} are the histogram counts; N is the number of data points; and Ψ is the digamma function. When measuring the mutual information between residues a and b , the mutual information of all pairs of the residues' torsion angles is summed up as shown in the formula

$$\text{MutInf}_{ab} = \sum_{k=\varphi, \phi, \chi} \sum_{l=\varphi, \phi, \chi} \text{MutInf}(k, l) \quad (4)$$

where $\text{MutInf}(k, l)$ is the mutual information between torsion angles k and l , belonging to residues a and b , respectively. Here, the main-chain (φ, ψ) and all the side-chain torsion angles ($\chi_1, \chi_2, \chi_3, \dots$) are taken into account as they are generally related to the description of the motions of biological importance.²¹

Construction of the Residue Complex Network. A protein structure is represented as an undirected graph with individual residues as nodes and inter-residue contacts as edges. A contact forms if $\text{C}\alpha$ atoms of a residue pair are within 8.0 Å.^{33,34} The edge weight is calculated as $\text{MutInf}_{\max} - \text{MutInf}_{ab}$, where MutInf_{\max} is the maximum mutual information of all residue pairs and MutInf_{ab} is the mutual information between the considered terminal residues a and b of the edge. The stronger the correlation between residues, the shorter the distance between them in the network.

Calculation of Allosteric Pathways. The shortest pathway between two residues is defined as the connecting route which minimizes the sum of the edge weights of the connecting route, and its length is just the minimized sum. To take into account the strong couplings between the residues far apart in space, the shortest pathways whose terminal residues' mutual information is higher than the average value and their distances are more than 10 Å are selected and their lengths are calculated with the Floyd-Warshall algorithm.³⁵ All the selected shortest pathways are ranked in increasing order of their terminal residues' mutual information values, and the top 10% of them are chosen as allosteric pathways. Then, these

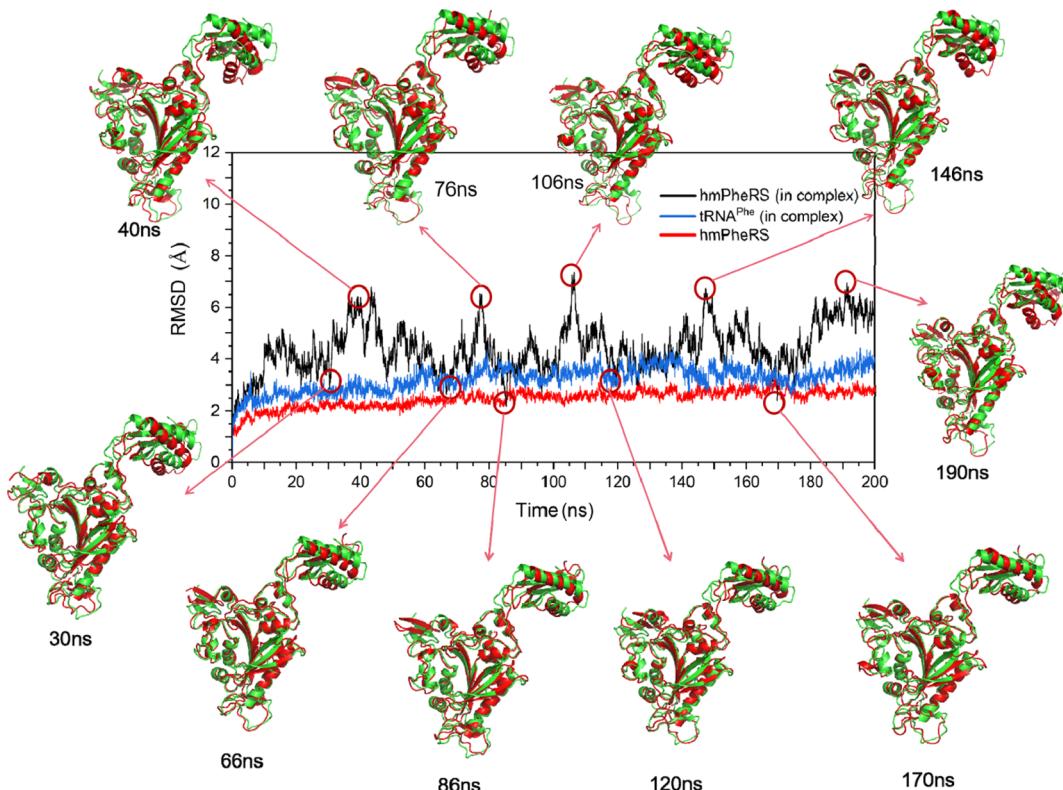


Figure 2. Time evolutions of RMSDs of hmPheRS in two independent MD simulations for its tRNA^{Phe}-free and -bound states, with the corresponding result of tRNA^{Phe} in the bound state also shown. Here, Ca atoms in protein and P atoms in tRNA are considered in RMSD calculation. Several snapshots (green) at some time points corresponding to the peaks and bottoms of the RMSD curve in 200 ns were extracted and superimposed on the original conformation (red).

allosteric pathways are clustered together according to their spatial proximity. To evaluate the spatial proximity between two allosteric pathways, their overlap is calculated as the proportion of the pairs of nodes (belonging to two different pathways) within a cutoff distance (8.0 Å) in all node pairs. Taking the overlap value as a similarity metric, these pathways are clustered using the hierarchical clustering algorithm. To determine the optimal number of clusters, two parameters are defined: intraoverlap (the average overlap within one cluster) and interoverlap (the average overlap between clusters). Then, the cluster separation efficiency (SE) is calculated using the following equation

$$SE = \sum_{i=1}^n \frac{\text{intraoverlap}(i)}{n} - \sum_{i,j(j>i)} \frac{\text{interoverlap}(i, j)}{n(n-1)/2} \quad (5)$$

where i and j represent two different clusters and n is the number of clusters. The optimal number of clusters is determined by maximizing the value of SE.

RESULTS AND DISCUSSION

Analyses of RMSDs of the Systems and RMSFs of Residues from MD Trajectories. To detect the effect of tRNA^{Phe} binding on the structural stability of hmPheRS, the conformational deviations of hmPheRS in tRNA^{Phe}-bound and -free systems from its corresponding starting structures over the course of simulations were monitored using the root mean square deviations (rmsd) of Ca atoms as a function of time, with the results shown in Figure 2. As shown in Figure 2, both hmPheRSs in the two systems reach their individual equilibrium states after a 20 ns relaxation. At equilibrium

stages, hmPheRS in the tRNA^{Phe}-free system remains relatively stable with a rmsd of 2.49 ± 0.34 Å, while that in the tRNA^{Phe}-bound system has a large structural fluctuation with a rmsd of 4.42 ± 0.95 Å although tRNA^{Phe} in the system is quite stable with a RMSD of 3.21 ± 0.46 Å (also shown in Figure 2). In order to detect the reason for the high rmsd fluctuation for the bound hmPheRS, we extracted the snapshots at some time points corresponding to the peaks and bottoms of the RMSD curve in 200 ns, and superimposed them on the original one, as shown in Figure 2. From Figure 2, the regions with large rmsds are mainly in the N-terminal region, the bottom structural fragment, and especially in the ABD which has an open–close movement toward the CAD (detailed analyses given in Figure Sf). Thus, we think that the enhanced mobility of the ABD, N-terminal region, and bottom structural fragment caused by the binding of tRNA^{Phe} mainly contributes to the larger structural fluctuation in the bound state than in the free state, which can be seen in the following analyses of residue root mean square fluctuations (RMSFs).

To see which parts of hmPheRS have an evident flexibility change when binding tRNA^{Phe}, we calculated RMSFs of hmPheRS in tRNA^{Phe}-bound and -free states respectively, as shown in Figure 3. Although a large conformational difference (“open” from “close”) for hmPheRS in the two states is observed, its overall RMSF contours do not have a significant difference. Additionally, note that almost all the regions of hmPheRS in the bound state display much higher flexibilities than the corresponding ones of the protein in the tRNA^{Phe} free state, which can explain the protein’s larger structural fluctuation in its bound state than in its free state. Compared with the close structure, the losses of inter-residue contacts in

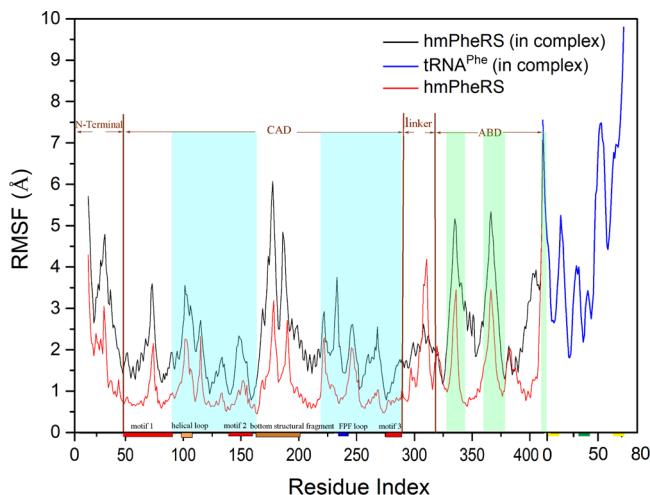


Figure 3. Comparison of residue RMSFs of hmPheRS in two independent MD simulations for its tRNA^{Phe}-free and -bound states. Here, Ca atoms for residues and P atoms for nucleotides are considered in RMSF calculation. The regions of the N-terminal, CAD, linker, and ABD are labeled. Motifs 1–3 are labeled by red rectangles along the axis of the residue index. The protein helical loop, bottom structural fragment, FPF loop, tRNA acceptor stem, and anticodon loop are labeled by orange, brown, blue, yellow, and green rectangles, respectively. The active site region in the CAD and anticodon recognition sites in the ABD are marked with transparent blue and green rectangles, respectively.

the open one contribute partially to the increased flexibility. Researchers have found that the structural flexibility of functional domains is essential for the aminoacylation activity of hmPheRS, and only the “open” structure is competent for aminoacylation.³⁶ Additionally, it is also critical for the active sites to possess a marked degree of natural plasticity to exert their control over the catalytic efficiency.⁷ Besides, the bottom structural fragment also becomes more flexible, which mainly attributes to the turning of its several helices in the free state into loops in the bound state (Figure 1). However, it should be pointed out that different from other regions, the linker becomes less flexible in the bound state, which we think maybe is related to its “hinge” role in the allosteric modulation in the bound state. With regard to tRNA^{Phe}, from Figure 3, we note that its anticodon loop has a relatively low flexibility due to its interactions with the ABD, while its receptor stem has a high flexibility, which has been proven to be conducive to the transfer of amino acid in the activation process.⁸ In summary,

tRNA^{Phe} binding changes hmPheRS’s dynamical properties, making its structure much more flexible.

Analyses of Movement Coupling in hmPheRS in tRNA^{Phe}-Free and -Bound States. The movement cross-correlations can be used to measure the strength of the collective motions between residues, helping to understand function-related movement couplings. To detect the effect of tRNA^{Phe} binding on residue collective motions, we calculated (eq 1) and compared the movement cross-correlations for hmPheRS in tRNA^{Phe}-free and -bound states, as shown in Figure 4a,b. Compared with hmPheRS in the free state (Figure 4a), hmPheRS in the bound state (Figure 4b) has stronger and more extensive residue movement correlations regardless of the positive or the negative ones, implying that tRNA^{Phe} binding enhances the intra- and inter-domain couplings. The strengthened and extended positive correlations mainly occur within the ABD, linker, active sites, motif 1, bottom structural fragment, and N terminal, which are beneficial for the allosteric signal transmission within the regions. Meanwhile, the enhanced negative correlations mainly occur between the ABD and some regions in the CAD including the bottom structural fragment, motif 1, motif 2 loop, and motif 3, which are helpful for the long-range allosteric coupling between the anticodon binding at the ABD and catalytic activation at the CAD. With regard to the correlations between hmPheRS and tRNA^{Phe}, from Figure 4b, the tRNA^{Phe} anticodon loop is positively correlated with the ABD due to its direct interactions, while the tRNA^{Phe} acceptor stem is negatively correlated with the ABD, which is consistent with the negative correlations between the CAD and ABD due to the acceptor stem’s direct interactions with the CAD. In summary, tRNA^{Phe} binding causes a significant enhancement of the motional correlations in the whole structure of hmPheRS, modulating intra-domain and inter-domain motional couplings and is beneficial for the allosteric communication from the ABD to CAD.

Motional Modes of hmPheRS in tRNA^{Phe}-Free and -Bound States. To better understand the complicated motions related to the function exertion of hmPheRS, the PCA was performed on the equilibrium ensembles from MD simulations. The results are shown in Figure 5a–c for the enzyme in the free state and in Figure 5d–f for it in the bound state. Figure 5a,d displays that the eigenvalues at the very beginning are relative to larger global concerted motions (low-frequency and usually function-related motion modes) but later decrease quickly and reach more localized fluctuations.

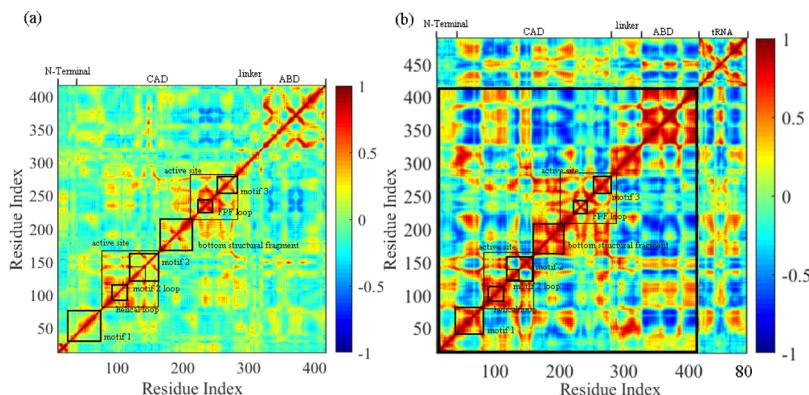


Figure 4. Movement cross-correlations of residues for hmPheRS in the tRNA^{Phe}-free state (a) and tRNA^{Phe}-bound state (b).

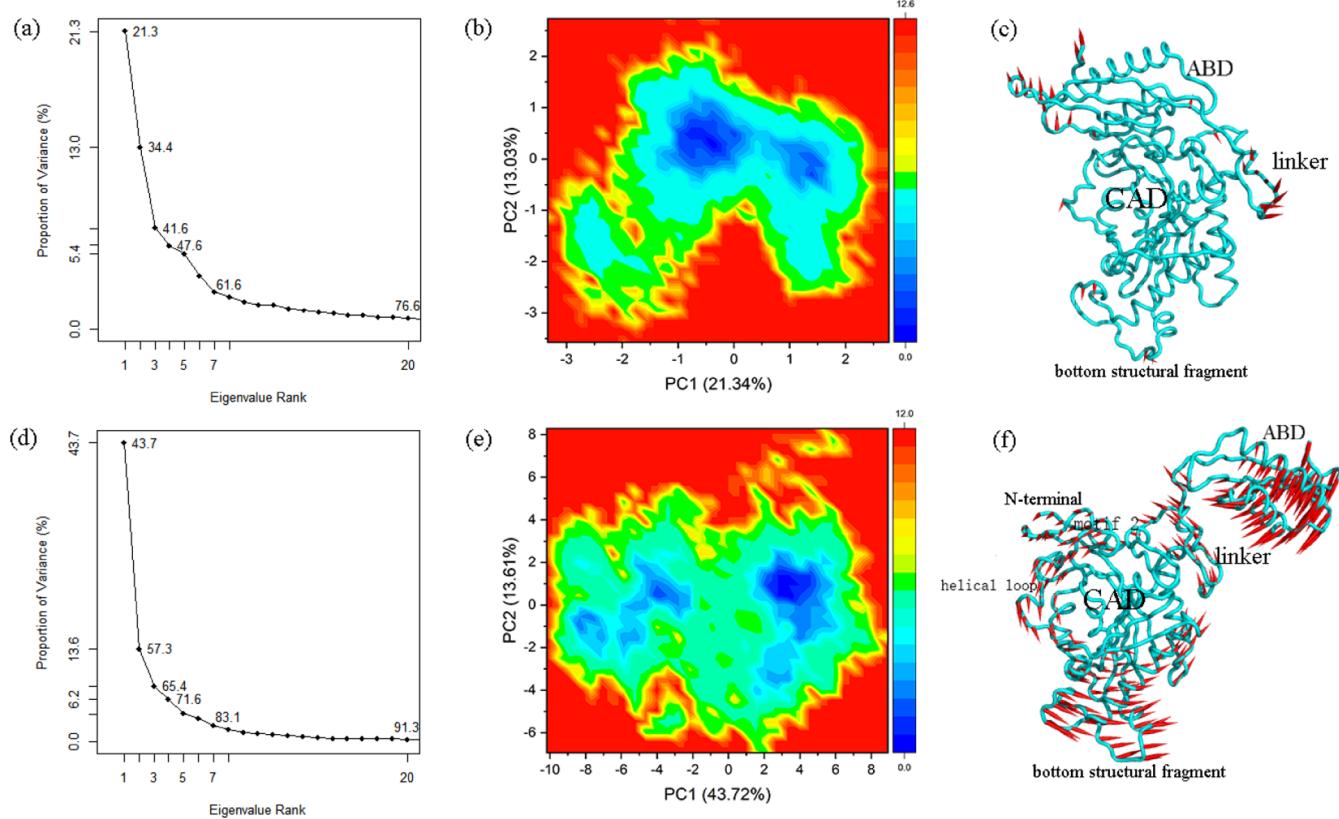


Figure 5. Results from the PCA performed on the two independent trajectories of MD simulations for hmPheRS in the tRNA^{Phe}-free state (a–c) and tRNA^{Phe}-bound state (d–f). Ca atoms are considered in PCA analysis. (a,d) Eigenvalue rankings based on the percentage of the total mean square displacement (or variance) of atoms from the average structure captured by the corresponding eigenvectors. (b,e) Free energy landscapes along the first two PCs. Blue indicates low free energy and red indicates high free energy. (c,f) First two slowest motion modes (depicted with a cone model) mapped on the average structures. The cone's length is proportional to the motion magnitude, and the cone's orientation indicates the motion direction.

Compared with 21.3 and 34.4% of total variance captured by the top 1 and 2 PCs, respectively, for the enzyme in the free state, the corresponding values are 43.7 and 57.3% for it in the bound state, which suggests that the large-amplitude cooperative motions are enhanced by tRNA^{Phe} binding, consistent with the analyses of movement cross-correlations.

To explore the main conformational behaviors of the two systems, the analyses of their free energy landscapes along the first two PCs were performed and the results show a big difference as plotted in Figure 5b,e. Compared with the tRNA^{Phe}-free system, the conformations of the tRNA^{Phe}-bound system span a much larger range and present more separate local minimal basins in the PC1–PC2 plane, hinting that the latter undergoes much larger conformational changes along the two directions and indicating again that the binding of tRNA^{Phe} leads to the enhancement of the motions in the tRNA^{Phe}-bound system. For clarity, the major structural changes along the motion mode PC1 are illustrated in each of the two systems, as shown in Figure 5c,f. Usually, the first few motion modes represent the low-frequency large-amplitude motions related to protein functions. For the free hmPheRS, the whole system is of low mobility with weak intrinsic motions observed for the ABD and linker (Figure 5c). For the bound one, the whole system has a remarkable collective motion, especially for the N-terminal region, surrounding regions of active sites, bottom structural fragment, and ABD (Figure 5f). Experiments showed that tRNA binding induces the conformational changes throughout the protein-

tRNA interface and active sites.⁷ The N-terminal and motif 2, located at the interface, have consensus motions, which is helpful for their common role in specifically binding tRNA^{Phe} 3'-CCA end sequence. The surrounding regions of active sites, such as the helical loop, motif 2 loop, and FPF loop, also have a concerted motion, which is conducive to the binding of the tRNA 3'-CCA end and phenylalanine and their closing up the active sites, facilitating the aminoacylation reaction. We note that the center of the active sites is of a relatively low mobility, which is in accordance with the precise and tight positioning of catalytic sites—a requirement for mechanochemical activity of enzymes.³⁷ Additionally, an evident negatively correlated motion is observed between the ABD and bottom structural fragment, consistent with the analyses of movement cross-correlations of residues. Previous research studies suggested that the bottom structural fragment can characterize protein hmPheRS,⁷ but the detailed function is not clear until now. We think that the fragment, like a tail, extends the CAD and can strengthen the correlated motion between the ABD and CAD, facilitating their allosteric communication involved in the aminoacylation reaction. In a word, the above results imply that the tRNA^{Phe} binding induces a strong long-range coupled motion between the ABD and CAD of hmPheRS and thereof triggers the allosteric communication between them. The following detailed analyses of the allosteric communication also validate this point.

Allosteric Communication between hmPheRS Residues. To explore the effect of tRNA^{Phe} binding on the

allosteric communication within the hmPheRS system, we calculated the mutual information of all residue pairs based on eqs 2–4 from two MD simulation trajectories. Then, the average mutual information of all pairwise residues within a given range of Ca distances 0–2, 2–4, 4–6 Å, and so forth were calculated. Figure 6 shows the change in average mutual

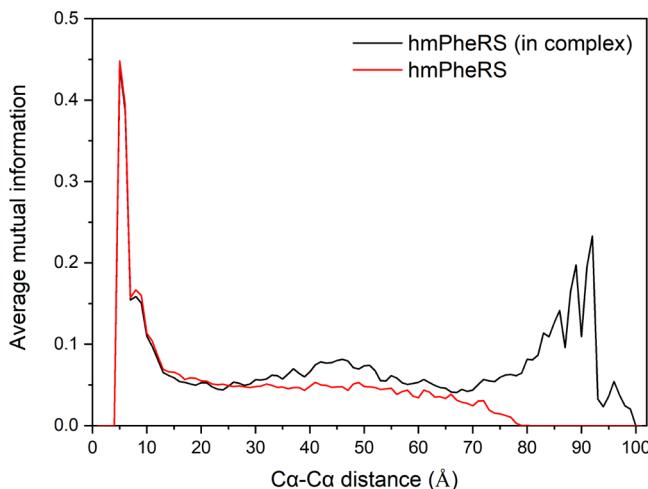


Figure 6. Changes in the average mutual information as a function of the pairwise residue $C\alpha$ distance for hmPheRS in tRNA^{Phe}-free and tRNA^{Phe}-bound states. The average mutual information means the average value of the mutual information of all pairwise residues within a given range of $C\alpha$ distances.

information as a function of $C\alpha$ distance between residues. The higher the average mutual information of residue pairs, the stronger their movement correlations.

From Figure 6, it can be seen that for the pairwise residues within 5 Å, their motion correlations are very strong due to their direct atom contacts. Beyond 5 Å, the average mutual information has an evident decrease, which is the case for the two systems. Then, for the tRNA^{Phe}-free state, the value continues to drop slowly to zero at about 80 Å distance. Interestingly, for the tRNA^{Phe}-bound state, the value, after a fast drop, has a slow ascent and then reaches the second peak of the curve at an inter-residue distance of about 45 Å,

suggesting that there exists a relatively strong correlated motion between the residues with this distance apart. Through examination, the correlated motions mainly occur within the CAD, which may help to modulate the active sites at the CAD. Then, after a small drop, the curve again shows a rise at about 80 Å which approximately corresponds to the inter-residue distances between the anticodon recognition sites at the ABD and the active sites at the CAD,⁴ indicating well the existence of long-range allosteric communication between them to keep hmPheRS in an aminoacylation active state. Afterward, the curve reaches the next high peak which is mainly caused by the highly correlated motions between the ABD and bottom structural fragment, consistent with the analyses of the movement cross-correlations between them, which hints again the important role of bottom structural fragment in the long-range allosteric coupling.

Allosteric Communication Pathways. To investigate the allosteric signal communication between residues far apart in space, we selected the shortest pathways with the mutual information of the terminal residue pairs greater than the average value and the residue pair distance larger than 10 Å and calculated their pathway lengths (see Materials and Methods). Then, among them, the 500 shortest pathways with the smallest pathway lengths were taken as allosteric pathways. Afterward, these allosteric pathways were clustered, and the optimal number of clusters was determined with eq 5. Figure S1 in Supporting Information gives the change in cluster SE as a function of the number of clusters. From Figure S1, the optimal number of clusters is 12, which maximizes the value of SE.

The first two pathway clusters with the most allosteric pathways (Table S1) were selected for further analyses, as shown in Figure 7. From Figure 7, as expected, both pathway clusters go through the linker and connect the anticodon recognition site and active site regions. Especially, the active site region, linker, and ABD represent a high density of allosteric pathways. Meanwhile, there are quite a few pathways that connect the N-terminal, important motifs, and bottom structural fragment, which are pivotal to the enzyme's function exertion.

From Figure 7a, the pathway cluster is distributed in the ABD, linker, and some regions of the CAD. In the ABD, the

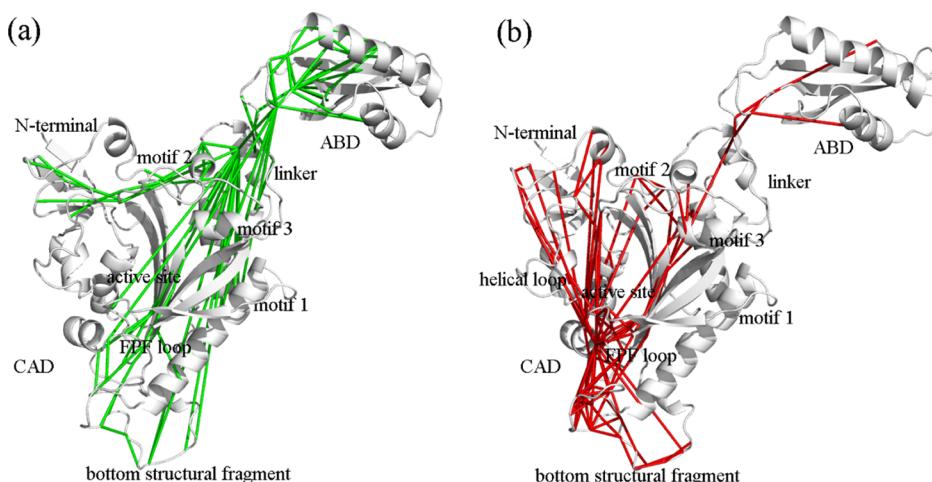


Figure 7. Identified two major allosteric pathway clusters. The motifs 1–3, characteristic loops (helical loop, motif 2 loop, and FPF loop), and bottom structural fragment are labeled.

pathways mainly start from residues Arg350, Asp364, Ser375, and Arg414, of which the latter three participate in the specific recognition of G34, the first nucleotide of the tRNA^{Phe} anticodon GAA.⁷ Then, they go through the linker to the CAD, and within the CAD, they group into several main branches. One branch goes through the motif 2 loop to the N-terminal region. The motif 2 loop is critical for correctly positioning and stabilizing the tRNA 3'-CCA end sequence in the active sites and also important for the transition of hmPheRS between the close and open conformations.⁴ The N-terminal contributes to the tRNA^{Phe} recognition and complex stability via electrostatic interactions with tRNA^{Phe}.⁴ We speculate that this branch plays an important role in regulating motif 2 and N-terminal to bind tRNA to ensure the subsequent aminoacylation reaction; one branch reaches the FPF loop of the active sites, which we think occurs mainly to enable the FPF loop to specifically bind phenylalanine.⁸ One branch reaches motif 1. Considering its enhanced mobility (Figure 3) and negative coupling with the ABD by tRNA binding (Figure 4b) and its conformational differences in tRNA^{Phe}-free and -bound states (Figure 1), we believe it is likely that motif 1 plays a critical role in the allosteric transition of hmPheRS; the other branch reaches the bottom structural fragment, which we think may contribute to the strengthened motion coupling between the ABD and CAD, promoting the aminoacylation reaction.

From Figure 7b, this pathway cluster is mainly distributed in the CAD. There is a small number of pathways transmitting from ABD residues Arg414 (involved in anticodon recognition) and Arg350, which goes through the linker and motif 3 and then clusters together with other pathways at the active sites. Most of the allosteric pathways in the CAD start from the N-terminal and motif 2 loop, go through the helical loop, center of the active sites or motif 3, and then cluster at the FPF loop and the bottom of the active sites. Afterward, a part of them reaches the bottom structural fragment. For the pathways, we speculate that the activated N-terminal and motif 2 by using the former pathway cluster probably regulate motif 3 (a part of active sites), the helical loop, and FPF loop (responsible for positioning the tRNA^{Phe} 3'-CCA end and phenylalanine in active sites) to exert their functions coordinately, ensuring the smooth running of the aminoacylation reaction. Thus, the aminoacylation activity is the result of the co-regulation from multiple allosteric signaling pathways.

In summary, these two clusters of allosteric pathways demonstrate how the allosteric signals transmit from the ABD to CAD active sites and modulate multiple important regions of the enzyme hmPheRS to act together and finally achieve the aminoacylation function.

Hubs in Allosteric Communication Pathways. In our work, the structure of the enzyme hmPheRS in the bound state is represented as a weighted residue network with individual residues as nodes and inter-residue contacts ($C\alpha$ atoms within 8.0 Å) as edges whose weights are determined from the mutual information between the terminal residues of the edges. Thus, on one hand, the network reflects the inter-residue interactions, and on the other hand, it reflects the inter-residue dynamical couplings within hmPheRS and can be regarded as a signal-communication network. To explore the key residues and regions in allosteric communication, we assigned each node in the residue network a hub score (Figure 8) which is defined as the sum of the numbers of the allosteric pathways

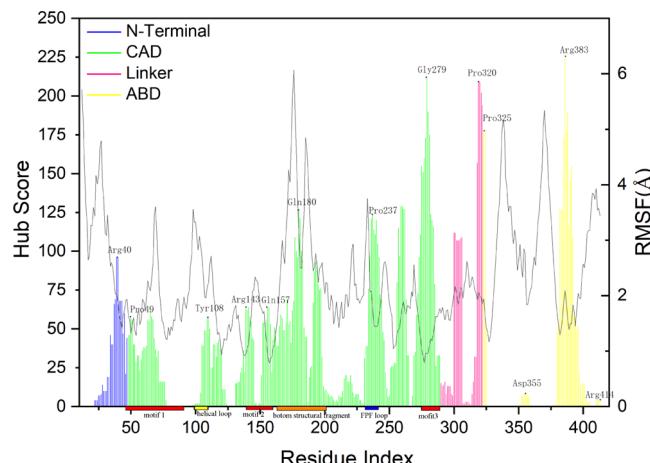


Figure 8. Hub score (colored line) of each residue of hmPheRS in the bound state with RMSFs (black line) of $C\alpha$ atoms also displayed. The N-terminal, CAD, linker, and ABD are represented by blue, green, pink, and yellow lines, respectively. Motifs 1–3 are labeled by red rectangles along the axis of residue index. The helical loop, bottom structural fragment, and FPF loop are labeled by yellow, orange, and blue rectangles, respectively.

mediated by the node and its neighbors (± 3 residues along the sequence). Thus, the residues with a high hub score play key roles in allosteric communication and in the local structural stability of the enzyme, both of which are intimately associated with the enzyme's catalytic activity and efficiency. In the following, combining the hubs' mediating roles, we analyze their functions in terms of the dynamics of the regions where the hubs occur.

From Figure 8, each domain of the enzyme has the residues with relatively high hub scores, which mediate multiple allosteric pathways and play important roles in allosteric communication. For the N-terminal region, the peak of the hub score is located at residue Arg40 which mediates some allosteric pathways bridging the N-terminal and active sites. Upon binding tRNA^{Phe}, the N-terminal interacts with the tRNA acceptor stem and stabilizes the complex structure, which facilitates the aminoacylation reaction. The residue Arg40 contributes significantly to the complex stability via electrostatic interactions with tRNA^{Phe}.⁴ Meanwhile, this residue has a low RMSF in this region, which is consistent with its role in this place.

In the CAD, the residues with a relatively high hub score are located in the specific motifs and regions, including Pro49 in Motif 1, Arg143 and Gln157 in Motif 2, Gly279 in motif 3, Tyr108 in the helical loop, Pro237 in the FPF loop, and Gln180 in the bottom structural fragment. Motif 1 mediates many allosteric pathways from the ABD. Experiments showed the mutation Pro49Ala in motif 1 changes the movement of the ABD toward CAD, thereof causing an impaired aminoacylation activity,⁵ which supports our deduction that motif 1 plays an important role in the enzyme's conformational transition. Motif 2 specifically recognizes and positions the tRNA acceptor stem and further coordinates the activities of the CAD and ABD.⁵ In the crystal structure, besides the interactions with tRNA^{Phe} acceptor stem, the motif 2 residues Arg143 and Gln157 interact with the Phe-AMP phosphate group, which facilitates the smooth transfer of phenylalanine to tRNA^{Phe}.⁷ Motif 3 mediates multiple allosteric pathways bridging the linker and active sites. Its Gly279 is located in

the shape-generating fragment of the CAD,⁵ which is crucial to the formation and stability of the CAD conformation. Thus, the residue's roles in allosteric communication and the stabilization of the CAD are consistent with its high hub score and low mobility. For the helical loop, multiple pathways go through it from the N-terminal to active sites. Research studies have found that the mutation Tyr108Cys in the helical loop causes a decrease in catalytic efficiency of the tRNA charging with phenylalanine,³⁸ hinting its modulating role in catalytic activity.

For the ABD, the residues Pro325, Asp355, Arg383, and Arg414 have a relatively large hub score. Residue Asp355 mediates the allosteric pathways bridging the anticodon recognition sites and linker region. The experiments showed that the mutation Asp355Val causes a decrease in the binding affinity of the active sites for phenylalanine, impairing the aminoacylation activity.³⁸ Arg383 is located in the shape-generating fragment of the ABD which contributes largely to the stabilization of the ABD conformation.⁵ The mutation Arg383Cys partially disrupts a salt bridge between Arg383 and Glu358, leading to an approximately two-fold reduction in the catalytic efficiency.⁵ Arg414 specifically recognizes G34, the first nucleotide of tRNA^{Phe} anticodon GAA. The mutation Arg414Ala displays a considerable reduction in the aminoacylation activity.⁶ In summary, the above experimental data validate the pivotal roles of the identified key mediating residues Asp355, Arg383, and Arg414 in the long-range allosteric communication between the ABD and active sites in the CAD.

It should be pointed out that for the other identified key residues including Gln180 in the bottom structural fragment, Pro237 in the FPF loop, Pro320 in the linker region, and Pro325 in the ABD, we did not find the experimental data to validate their importance. The bottom structural fragment gathers some pathways going through the active sites and linker region. Combining its strong couplings with the ABD, we think that the residue Gln180 may be important in responding to the ABD motion and increasing the motion correlations between the ABD and CAD. The FPF loop is relatively stable, consistent with its function in the recognition of phenylalanine.⁸ Its residue Pro237 may provide hydrophobic interactions with phenylalanine for easy binding. The linker as a unique and necessary region bridges the ABD and CAD, playing a critical role in the enzyme's allostery and function. We believe that its residue Pro320 and the adjacent one Pro325 with a considerably high hub score play an important role in transmitting the anticodon recognition signal to the CAD active sites. Hope our findings will be helpful for guiding future experiments to explore the enzyme's functional sites.

In conclusion, the residues identified with relatively high hub scores are either structurally vital or dynamically critical for the enzyme's aminoacylation activity. They mediate multiple allosteric pathways connecting functionally important motifs and regions, which thereof represent the coordinated motions to ensure the smooth running of the enzyme's aminoacylation function.

CONCLUSIONS

In this work, we explore the dynamic properties and allosteric communication of human mitochondrial phenylalanyl-tRNA^{Phe} synthetase (hmPheRS) in tRNA^{Phe}-free and -bound states with full atomistic MD simulations combined with the torsional

mutual information-based network model. Through MD simulations, we find that the tRNA^{Phe} binding significantly affects the dynamics of hmPheRS. The RMSF analysis shows that the molecular flexibilities are enhanced considerably by tRNA^{Phe} binding. The movement cross-correlation analysis and PCA reveal that the intra-domain and inter-domain motional couplings are also enhanced and extended substantially, and there occur evident motional couplings between the ABD and CAD. Additionally, the torsional mutual information analysis from the MD trajectories demonstrates that there exists long-range allosteric communication between the ABD anticodon recognition sites and CAD active sites only in the tRNA^{Phe}-bound state. Furthermore, we identify the allosteric pathways transmitting the signals from the anticodon binding at the ABD to the aminoacylation reaction at the CAD, revealing the co-regulation of multiple allosteric pathways in the important regions to enable them to act coordinately and achieve the hmPheRS's aminoacylation function. Finally, we identify the key hub residues which mediate multiple allosteric pathways involved in the enzyme's allosteric regulation, in good agreement with the experimental observation. In addition, we propose the possible functions for motif 1 and the bottom structural fragment based on the dynamics data, which provides a complement to the deficiency of their function annotation. Importantly, we find several important hub residues not yet validated by experiments, which can provide a guide for researchers to explore the enzyme's functional sites. This work sheds light on the allosteric mechanism in hmPheRS and is helpful for the understanding of the biological function exertion of the class-II aaRS.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.1c03228>.

Change in the cluster SE as a function of the number of clusters and number of allosteric pathways in each cluster when 12 clusters are selected ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

Chunhua Li – Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China;
Email: chunhuali@bjut.edu.cn

Authors

Qi Shao – Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Zhongjie Han – Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China;
 orcid.org/0000-0002-6088-2469

Jingmin Cheng – Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Qiankun Wang – Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China

Weikang Gong – Faculty of Environmental and Life Sciences, Beijing University of Technology, Beijing 100124, China;
 orcid.org/0000-0001-8797-784X

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcb.1c03228>

Author Contributions

Q.S. and Z.H. contributed equally to this work. Q.S., Z.H., and C.L. designed the research. Z.H. performed the MD calculations. Z.H. and Q.W. performed data analyses. Q.S., Q.W., and W.G. developed the program. Z.H., C.L., and J.C. wrote the article.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (31971180, 11474013).

REFERENCES

- (1) Ibba, M.; Söll, D. Aminoacyl-tRNA Synthesis. *Annu. Rev. Biochem.* **2000**, *69*, 617–650.
- (2) Park, S. G.; Schimmel, P.; Kim, S. Aminoacyl tRNA Synthetases and their Connections to Disease. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11043–11049.
- (3) Sissler, M. Mitochondrial aminoacyl-tRNA Synthetase. In *The Aminoacyl-tRNA Synthetases*; Ibba, M., Francklyn, C., Cusack, S., Eds.; Landes Biosciences: Georgetown, TX, 2005; pp 271–284.
- (4) Klipcan, L.; Moor, N.; Finarov, I.; Kessler, N.; Sukhanova, M.; Safro, M. G. Crystal Structure of Human Mitochondrial PheRS Complexed with tRNAPhe in the Active “Open” State. *J. Mol. Biol.* **2012**, *415*, 527–537.
- (5) Kartvelishvili, E.; Tworowski, D.; Vernon, H.; Moor, N.; Wang, J.; Wong, L.-J.; Chrzanowska-Lightowers, Z.; Safro, M. Kinetic and Structural Changes in HsmtPheRS, Induced by Pathogenic Mutations in Human FARS2. *Protein Sci.* **2017**, *26*, 1505–1516.
- (6) Peretz, M.; Tworowski, D.; Kartvelishvili, E.; Livingston, J.; Chrzanowska-Lightowers, Z.; Safro, M. Breaking a Single Hydrogen Bond in the Mitochondrial tRNA(Phe)-PheRS Complex Leads to Phenotypic Pleiotropy of Human Disease. *FEBS J.* **2020**, *287*, 3814–3826.
- (7) Klipcan, L.; Levin, I.; Kessler, N.; Moor, N.; Finarov, I.; Safro, M. The tRNA-induced Conformational Activation of Human Mitochondrial phenylalanyl-tRNA Synthetase. *Structure* **2008**, *16*, 1095–1104.
- (8) Moor, N.; Kotik-Kogan, O.; Tworowski, D.; Sukhanova, M.; Safro, M. The Crystal Structure of the Ternary Complex of Phenylalanyl-tRNA Synthetase with tRNAPhe and a Phenylalanyl-Adenylate Analogue Reveals a Conformational Switch of the CCA End†. *Biochemistry* **2006**, *45*, 10572–10583.
- (9) Lu, S.; Shen, Q.; Zhang, J. Allosteric Methods and their Applications: Facilitating the Discovery of Allosteric Drugs and the Investigation of Allosteric Mechanisms. *Acc. Chem. Res.* **2019**, *52*, 492–500.
- (10) Lu, S.; Zhang, J. Small Molecule Allosteric Modulators of G-Protein-Coupled Receptors: Drug-Target Interactions. *J. Med. Chem.* **2019**, *62*, 24–45.
- (11) Lu, S.; Ni, D.; Wang, C.; He, X.; Zhang, J. Deactivation Pathway of Ras GTPase Underlies Conformational Substates as Targets for Drug Design. *ACS Catal.* **2019**, *9*, 7188.
- (12) Lu, S.; Chen, Y.; Wei, J.; Zhao, M.; Duan, N. C.; Xh, A.; Jian, Z. Mechanism of Allosteric Activation of SIRT6 Revealed by the Action of Rationally Designed Activators. *Acta Pharm. Sin. B* **2021**, *11*, 1355.
- (13) Ni, D.; Jiacheng, W.; Xinheng, H.; Urrehman, A.; Xinyi, L.; Yuran, Q.; Jun, P.; Shaoyong, L.; Jian, Z. Discovery of Cryptic Allosteric Sites Using Reversed Allosteric Communication by a Combined Computational and Experimental Strategy. *Chem. Sci.* **2021**, *12*, 464–476.
- (14) Lu, S.; He, X.; Ni, D.; Zhang, J. Allosteric Modulator Discovery: From Serendipity to Structure-Based Design. *J. Med. Chem.* **2019**, *62*, 6405–6421.
- (15) Ni, D.; Li, Y.; Qiu, Y.; Pu, J.; Lu, S.; Zhang, J. Combining Allosteric and Orthosteric Drugs to Overcome Drug Resistance. *Trends Pharmacol. Sci.* **2020**, *41*, 336–348.
- (16) Shulman, A. I.; Larson, C.; Mangelsdorf, D. J.; Ranganathan, R. Structural Determinants of Allosteric Ligand Activation in RXR Heterodimers. *Cell* **2004**, *116*, 417–429.
- (17) Stiel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily Conserved Networks of Residues Mediate Allosteric Communication in Proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69.
- (18) Chennubhotla, C.; Bahar, I. Signal Propagation in Proteins and Relation to Equilibrium Fluctuations. *PLoS Comput. Biol.* **2007**, *3*, 1716–1726.
- (19) Bowerman, S.; Wereszczynski, J. Detecting Allosteric Networks Using Molecular Dynamics Simulation. *Methods Enzymol.* **2016**, *578*, 429.
- (20) Kumawat, A.; Chakrabarty, S. Hidden Electrostatic Basis of Dynamic Allostery in a PDZ Domain. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E5825–E5834.
- (21) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theor. Comput.* **2009**, *5*, 2486–2502.
- (22) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (23) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
- (24) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O’Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation On Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
- (25) MacKerell, A. D.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (26) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (27) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (28) Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys. Rev. Lett.* **1998**, *80*, 2733–2736.
- (29) Chang, S.; He, H.-Q.; Shen, L.; Wan, H. Understanding Peptide Competitive Inhibition of Botulinum Neurotoxin a Binding to SV2 Protein Via Molecular Dynamics Simulations. *Biopolymers* **2015**, *103*, 597–608.
- (30) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695–2696.
- (31) Grassberger, P. Finite Sample Corrections to Entropy and Dimension Estimates. *Phys. Lett. A* **1988**, *128*, 369–373.
- (32) Steuer, R.; Kurths, J.; Daub, C. O.; Weise, J.; Selbig, J. The Mutual Information: Detecting and Evaluating Dependencies Between Variables. *Bioinformatics* **2002**, *18*, S231–S240.
- (33) Bagler, G.; Sinha, S. Network Properties of Protein Structures. *Phys. A* **2012**, *346*, 27–33.
- (34) Brinda, K. V.; Surolia, A.; Vishveshwara, S. Insights Into the Quaternary Association of Proteins through Structure Graphs: A Case Study of Lectins. *Biochem. J.* **2005**, *391*, 1–15.
- (35) Floyd, R. W. Algorithm 97: Shortest Path. *Commun. ACM* **1969**, *5*, 345.
- (36) Yadavalli, S. S.; Klipcan, L.; Zozulya, A.; Banerjee, R.; Svergun, D.; Safro, M.; Ibba, M. Large-Scale Movement of Functional Domains Facilitates Aminocylation by Human Mitochondrial phenylalanyl-tRNA Synthetase. *FEBS Lett.* **2009**, *583*, 3204–3208.

- (37) Yang, L.-W.; Bahar, I. Coupling Between Catalytic Site and Collective Dynamics: A Requirement for Mechanochemical Activity of Enzymes. *Structure* **2005**, *13*, 893–904.
- (38) Elo, J. M.; et al. Mitochondrial phenylalanyl-tRNA Synthetase Mutations Underlie Fatal Infantile Alpers Encephalopathy. *Hum. Mol. Genet.* **2012**, *21*, 4521–4529.