PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Structural and mechanistic insight into the substrate binding from the conformational dynamics in apo and substrate-bound DapE enzyme

Debodyuti Dutta, and Sabyashachi Mishra*

Department of Chemistry, Indian Institute of Technology Kharagpur, Kharagpur, India

E-mail: mishra@chem.iitkgp.ernet.in

Abstract

Conformational dynamics in large biomolecular systems are often associated with their physiological roles. The dynamics of a dimeric microbial enzyme, DapE, with great potential of an antibiotic target, has been studied employing long molecular dynamics simulations of the enzyme in apo form and its substrate bound complex form. The essential dynamics of apo enzyme and enzyme-substrate complex are extracted from the principal component analysis of the simulations of these two systems where the first two principal components are analyzed in detail. The essential motion of the enzyme in the substrate bound form exhibits a folding motion of its two catalytic domains over the two dimerization domains, which results in the repulsion of water molecules away from the active site of the enzyme-substrate complex. This folding motion also leads to a stabilizing binding free energy of the substrate arising from favorable interaction of the substrate and side chains of the enzyme. The dynamics in the enzyme-substrate complex results in stronger interaction between the catalytic and dimerization domains reflected by increased number of inter-domain hydrogen bonds. The substrate,

*To whom correspondence should be addressed

located in the catalytic domain of DapE, establishes contacts with the side chains of the dimerization domain of DapE by extended chains of hydrogen bonds, which emphasizes the role of the dimerization domain in substrate binding.

Introduction

Bacterial infections together with antibiotic resistance pose a serious health issue which requires discovery of novel targets and new drug molecules.[1–5] To that end, the meso-diaminopimelate (mDAP)/lysine biosynthetic pathway provides promising antibiotic targets.[6–10] The products of this pathway, namely, lysine and mDAP, are inevitable for bacterial growth and multiplication.[11] The most promising part of the importance of the enzymes in the lysine biosynthetic pathway comes from the fact that, while it is very important for bacterial survival, there exists no similar pathways in mammals. Therefore the inhibitors of the enzymes in the mDAP/lysine pathway are expected to provide selective toxicity against bacteria while not affecting its human host.[12] The deletion of the gene encoding for one of the enzymes in this biosynthetic pathway, the dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase (DapE), is lethal to some bacteria.[13,14] The DapE enzyme is responsible for catalyzing the hydrolysis of N-succinyl-L,L-diaminopimelic acid (SDAP) forming succinate and L,L-diaminopimelic acid, which subsequently gets converted to meso-diaminopimelate and lysine, both of which are essential components for bacterial cell wall formation.[6,15]

The DapE enzyme is a dimeric protein consisting of a catalytic domain and a dimeric domain in each monomer.[16] The catalytic domain hosts the active site for catalytic reaction whereas the dimeric domains provide a contact between the two monomers (Figure 1). The catalytic reactions in the wild type and mutant DapE has been studied by several experimental and computational methods outlining the kinetics and energetics of the reaction.[6,7,15,17] However, very little is known about the conformational dynamics of this enzyme system and in particular, the role of the dimerization domain in the catalytic action of the enzyme. Recently, it has been shown that the enzyme

activity is destroyed in the absence of the dimerization domain, which is attributed to the conformational freedom of a loop in the catalytic domain in the absence of the dimerization domain.[18]

Enzyme catalysis involves, besides the actual biochemical reaction, preparation of the apo enzyme to host its substrate, the migration and binding of the substrate to the active site, and release of the products of the catalytic reaction. All these steps of the catalytic action involves conformational dynamics in the enzyme, which holds crucial information to the understanding of enzyme functions.[19] In multi-domain enzymes the large scale domain motions of enzymes are often found essential for their physiological functions and are often evolutionarily optimized.[20–25] The dynamics of large biomolecules involves an interplay of many unspecific and fast stochastic local fluctuations occurring within picosecond timescale together with large-scale rearrangements and movements of domains relative to each other occurring in a much longer timescale.[26–28]

While classical molecular dynamics (MD) simulations provide an excellent platform to determine the multiscale behavior of the complex biological systems, extracting information about the significant large-scale motions, such as those corresponding to domain movements, from a MD trajectory is not always straightforward owing to the high dimensionality of biological systems. Principal component analysis (PCA) offers an elegant route to monitor the concerted motions in proteins that are often found to be biologically significant.[29–31] Within PCA, one decomposes the conformational subspace explored by $3N$ protein degrees of freedom (where N is the number of atoms under consideration) into a small set of modes (p), which forms a low dimensional subspace where concerted, functionally relevant motions are captured, and a large remaining set ($3N - p$), that forms a high dimensional subspace characterized by high frequency independent motions which can be described as small Gaussian type fluctuations.[32,33] PCA has been applied to several protein systems, with results systematically showing that the most important fluctuations of a protein can be accounted for by the first few principal components of motion (principal components) and that a good description of the essential dynamics of the protein can thus be achieved.[29–33]

In the present work, we employ molecular dynamics simulations of the DapE enzyme in apo form and DapE-L,L-SDAP complex to sample the conformational dynamics of these two systems

and analyze the similarities and differences in the essential dynamics of the enzyme when it is bound to substrate as compared to its apo form.

Computational Details

The X-ray crystallographic structure of DapE (at 2.30 Å resolution) retrieved from the protein data bank (pdb ID: 3IC1)[16] was used as the starting point for the computational studies. The DapE enzyme exists in a dimeric form (chain A and chain B) with each monomer consisting of a catalytic and a dimerization domain (Figure 1). The active site in the catalytic domain is composed of two zinc ions bridged by Asp100 and one water molecule, and the side chains of His67, His349, Glu135 and Glu163 directly coordinated to Zn atoms.[16] The enzyme contains a total of 754 residues (377 residues in each chain). The residues 1 to 179 and 293 to 377 in both chains constitute the catalytic domain of the enzyme whereas the residues 180 to 292 form the dimeric domain.

The initial structure of the apo enzyme was taken from the X-ray crystallographic structure of DapE (PDB ID: 3IC1)[16] and the structure obtained from docking L,L-SDAP in the DapE active site[17] served as the initial structure of the enzyme-substrate complex. The apo enzyme and the enzyme-substrate complex were separately immersed in an orthorhombic water box of dimension 165 Å $\times$ 100 Å $\times$ 90 Å, ensuring a distance of 12 Å between the faces of the box and any atom of the enzyme or the substrate. The missing loops in the X-ray crystal structure were modelled using the MMTSB tool set.[34] All Glu, Asp, Lys, and Arg side chains were modelled as charged, while the His side chains were modelled to be neutral. Zn-amino acid interactions are treated as purely nonbonded interactions, which has been proved to be simple and accurate for metalloproteins.[35,36] The time evolution of the distances between Zn and the coordinating ligands show that the force field employed preserves the active site architecture of the enzyme (Figure S1 in the supporting information). The system was made overall charge neutral by adding 22 sodium ions and a salt concentration of 0.15 M was achieved by further adding 108 number of sodium and chlorine ions. The systems contained about 140,000 atoms. To release bad contacts in the system, first the protein

and protein-substrate complex were kept frozen and only the position of the water molecules and that of salt ions were minimized, followed by a full minimization. The system was then slowly heated to 300 K and equilibrated for 5 ns at constant temperature and volume (NVT ensemble), followed by 100 ns molecular dynamics run at constant temperature (300 K) and constant pressure (1 atm) using the Nose-Hoover Langevin thermostat and piston[37] (NPT ensemble). All simulations were done with only hydrogen containing bonds constrained by SHAKE.[38] An integration time step of 2 fs was used and structures were saved every 5 ps for analysis. The long-range electrostatic interactions were treated by the particle mesh Ewald (PME)[39] method with a 12 Å cutoff. The van der Waals interactions were truncated at a cutoff of 12 Å, and a switch function was activated starting at 10 Å. All simulations were performed employing the NAMD program[40] with the CHARMM22 force field[41] and TIP3 potential for water molecules,[42] and Charmm general force field for SDAP generated from Paramchem suite.[43]

The principal component analysis, also known as quasi-harmonic analysis or essential dynamics method, is a well established technique to express the conformational dynamics in high dimensional complex systems in terms of a few principal modes or principal components of motion.[44,45] Although, the time scale of conformational dynamics of biomolecular systems is well beyond nano-second range, in principal component analysis, the eigenvectors corresponding to the larger eigenvalues, which can be used to describe the essential motions, tend to converge to a stable set in the nanosecond timescale, suggesting that MD simulations of nanosecond timescale can provide a reasonable definition of the essential subspace, valid well beyond the nanosecond range.[46–48] The overall translational and rotational motion in the molecular dynamics trajectories of the enzyme and enzyme-substrate were removed by superimposing each frame of the trajectory to the C_α coordinates of the crystal structure. These trajectories were further used to generate covariance matrix by stripping the water molecules of the system. The dimeric enzyme contains 11,608 atoms which results in a 34,824 dimensional covariance matrix for the apo enzyme, whereas the enzyme-substrate complex containing 11,680 atoms results in a covariance matrix of 35,040 dimension. For comparison, the covariance matrices were also generated by considering only the backbone

atoms (consisting of C_α, C, N and O) of the apo-enzyme and enzyme-substrate complex. The 754 residues of the protein contains 3,016 backbone atoms which results in a 9,048 dimensional co-variance matrix for both the apo enzyme and the enzyme-substrate complex. The results from the principal component analysis with the backbone atoms compare well with the all atoms principal component analysis. In the present article, we discuss the principal components obtained from the diagonalization of the covariance matrix of all atoms.

The PCA decomposes the overall protein motion during the molecular dynamics simulation into a set of modes that can be ordered in terms of their degree of contribution to the overall protein motion arranged from the largest to the smallest the eigenvalues of the covariance matrix. The mode having the largest amplitude (also known as principal component 1) is the most significant one to describe the dynamical behavior of the system. The contribution of eigenmodes to the overall dynamics of the system gradually diminishes from the first eigenmode onwards and after a few eigenmodes, the amplitude becomes almost zero denoting local vibrations (Figure S3 in the supporting information). The convergence of the principal component analysis has been rigorously tested by using the cosine content analysis,[49] trajectory overlap analysis,[50] and randomness test of the principal components[29,51] (Figures S4, S5 and Table S1 in the supporting information).

The phenomenon of enzyme-substrate binding affinity can be quantified in terms of binding free energies. In the present work, the interaction energies between SDAP and DapE is calculated by molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method, which is one of the most widely used methods to study ligand-biomolecule interaction energies in noncovalently bound receptor-ligand complexes.[52–55] Within this approach, the free energy of binding ($\Delta G_{binding}$) of a substrate to a protein can be determined from the equation,[54]

$$\Delta G_{binding,solvated} = G_{complex,solvated} - G_{receptor,solvated} - G_{ligand,solvated}, \quad (1)$$

where, the free energy G in the right-hand side of the above equation is estimated as the sum of the

following terms,[54]

$$G_{solvated} = \langle E_{gas} \rangle + \langle \Delta G_{solvation} \rangle - \langle TS_{solute} \rangle. \quad (2)$$

The gas-phase energies E_{gas} are the molecular mechanical energies obtained from the force field with no cut off for non-bonded interactions. The polar solvation free energies $\Delta G_{solvation}$ are calculated using an implicit solvent model, in this case with Poisson-Boltzmann (PB) model. The PB equation was solved using 1000 linear steps of finite difference with dielectric constants of 1 and 80 for the interior and exterior of the molecule, respectively. The nonpolar contribution to the solvation energies are determined with a function of the solvent accessible surface area (SASA) $\Delta G_{nonpol} = \gamma SASA + b$, where γ is the surface tension set to 0.005 kcal mol^{-1}Å^{-2}, SASA was determined with the Molsurf method[56] using a probe radius of 1.4 Å, and the value of parameter b was set to zero. The values of these parameters are well documented in literature.[57,58] The third term in the Equation 2 estimates the entropy contribution, which can be obtained from normal mode analysis but are computationally expensive and tend to have a large margin of error that introduces significant uncertainty in the result.[59] For this reason, it is a common practice to ignore its contribution in binding free energy calculation, under the assumption that the change in solute entropy is small during complex formation.[59] This gives rise to an effective binding free energy which represents the binding free energy without the entropy correction. In the present work, we have employed the quasi-harmonic approach to evaluate the solute entropy change upon binding.[55,60] In addition to the global free energy calculation, decomposing ΔG_{total} in terms of contributions from structural subunits can provide valuable information about the local interactions. The effective binding free energy can be decomposed for each residue by including only those interactions in which one of the residue's atoms are involved, by a scheme called per-residue decomposition.[61,62] For the per-residue decomposition analysis, the entropy contributions are neglected since the entropy is calculated from the global motions spread over the entire complex where every residue contributes to all vibrational modes, and there is no straightforward way to approximately decompose the entropy into per-residue contributions.[55,61,62] The MM-PBSA calculations are performed with Amber software suite.[63]

Results and Discussion

Structural stability of the enzyme

The time evolution of the C_α root-mean-square deviations (RMSD) with respect to the crystal structure shows the overall stability of the DapE and DapE-SDAP complex (Figure 2). The occasional large RMSD values are traced to the rapid fluctuations of the loops in the enzyme. In particular, the loops consisting of residue numbers 71-95, 117-124, 322-328, and 346-353 present in the catalytic domain contribute to the flexibility of the protein. The overall stability of the secondary structure of the protein is further investigated by analyzing the time evolution of the number of residues participating in the stabilizing secondary structures by employing the dictionary of secondary structure prediction (DSSP) algorithm.[64] The number of residues participating in helices and sheets during the molecular dynamics simulations of the apo enzyme and enzyme-substrate complex, compared to their values in the protein crystal structure, show that the secondary structures remain largely intact during the 100 ns molecular dynamics simulation of both the apo enzyme and the enzyme-substrate complex (Figure S2 in the supporting information). To further establish the stability of the protein, we have calculated the number of backbone hydrogen bonds during the molecular dynamics simulation by taking a donor-acceptor cut-off distance of 3.4 Å.[65] The number of backbone hydrogen bonds remain stable during the simulations of the apo enzyme and enzyme-substrate complex (Figure S2 in the supporting information), indicating the stability of the enzyme native state.

Essential dynamics in the apo enzyme and enzyme-substrate complex from principal component analysis of the MD trajectories

The first two principal components, which account for about 70 % of the overall conformational dynamics in both apo enzyme and enzyme-substrate complex, are extremely important to describe the concerted global motions of the system. For the DapE-SDAP complex, the first principal component accounts for 54 % of the dynamics while it is 40 % in case of the apo enzyme. On

the other hand, the second principal component describes 30 % and 18 % of the conformational dynamics in apo enzyme and enzyme-substrate complex, respectively. This suggests that the first two modes are nearly equally important for the apo enzyme, while the first mode is predominant over the other modes in the case of the enzyme-substrate complex.

The root mean square fluctuations of the C_α atoms of the enzyme, averaged over intervals of 10 ns after excluding the first 20 ns of the MD trajectories, are shown in Figure 3 (upper panel), for both the chains in apo enzyme and enzyme-substrate complex. The catalytic domains of both chain A and chain B (residue number 1 to 179 and 293 to 377) exhibit higher fluctuations compared to that of the dimerization domains (residue number 180 to 292) in both the systems, which arises due to the fact that the catalytic domains host several loops while the dimerization domain shows several β sheets (Figure 1a). The DapE enzyme with a catalytic domain and a dimerization domain attains conformational flexibility from the former which hosts several loops while the latter, with stable secondary structures, imparts rigidity. In the middle panel of Figure 3, we have compared the C_α root mean square fluctuations of first two principal components with that of the actual MD trajectory. In order to calculate the atomic fluctuations along a principal component, the actual trajectory is filtered along that eigenmode.[66] The resulting new trajectory describes the dynamics of the system associated with that particular principal component. Both apo enzyme as well as the enzyme-substrate complex show higher root mean squared fluctuations along principal component 1 as compared to that of the principal component 2. For the third principal component onwards, the fluctuations become even smaller and do not show noteworthy contribution to the overall dynamics (data not shown). The enzyme-substrate complex exhibits higher fluctuations along the principal component 1 as compared to that of the apo enzyme, whereas, along the principal component 2, the fluctuation is higher in the apo enzyme as compared to that of the complex (middle and lower panels of Figure 3).

The overlap of six snapshots (red to blue) along the first two principal components for the apo enzyme and the enzyme-substrate complex are shown in Figure 4. The overlapped structures help in visualizing the essential dynamics along the corresponding principal components and provide

information to link the essential dynamics of the complex system in terms of some internal coordinates of the system. The overlapped structures along principal component 1 suggest a global motion involving the two catalytic domains. Thus a possible geometric parameter that quantifies the principal component 1 can be the distance between the centers of mass of the catalytic domain and the dimerization domain in both the chains. This can be quantitatively verified from the high correlation of the distance between the centers of mass of the catalytic domain and dimerization domain in both chains of the apo enzyme and enzyme-substrate complex obtained from MD simulations and the corresponding trajectories filtered along principal component 1 (Figure S6 in the supporting information). The time evolution of the distance between centers of mass of the catalytic domain and dimerization domain along mode 1, shows a decreasing distance between the catalytic and dimerization domains in the DapE-SDAP complex system, while the same is not true for the apo enzyme (Figure S6 in the supporting information). This can be explained by comparing the cross-correlation matrix of C_α atomic fluctuations along principal component 1 for the apo and complex systems (Figure S7 in the supporting information), which shows the motion of the two catalytic domains in the complex exhibit more anticorrelation compared to that in the case of apo enzyme. This indicates that the principal component 1 in the enzyme-substrate complex involves a folding of the catalytic domains onto the dimerization domains, which is further supported from the decreasing value of the radius of gyration of the entire protein along the principal component 1 of the DapE-SDAP complex (Figure 5a). The decreasing radius of gyration in the DapE-SDAP complex along principal component 1 indicates the increase of the compactness of the enzyme-substrate complex, compared to the apo enzyme.

The principal component 2 in both DapE and DapE-SDAP systems corresponds to a rotation of the two catalytic domains about the two dimerization domains in opposite directions (Figure 4c,d). In this case, a possible geometric parameter to describe the principal component 2 is the dihedral angle between the four centers of mass, consisting of the catalytic domain of chain A, the dimerization domain of chain A, the dimerization domain of chain B and the catalytic domain of chain B. The displacement along the principal component 2 in both apo enzyme and enzyme-substrate

complex show excellent agreement with the time evolution of this dihedral angle along the principal component 2, suggesting the accurate description of the principal component 2 by the chosen internal coordinate (Figure S8 in the supporting information).

Enzyme-substrate interactions

The binding propensity of a ligand to a macromolecule can be quantitatively determined by the MM-PBSA approach.[52–54] We have carried out MM-PBSA analysis to quantify the interaction between DapE enzyme and SDAP substrate. Upon SDAP binding, the effective binding free energy (without entropy contribution) is estimated to be -66 kcal/mol, indicating a very strong binding of the substrate to the protein. The contribution of the solute entropy to the substrate binding process is evaluated from quasi harmonic approach and found to be 27 kcal/mol (at 298 K). Hence the net substrate binding free energy of DapE is determined as -39 kcal/mol. To estimate the contribution of different components of the enzyme and substrate towards the net binding free energy, a residue-wise decomposition of the total effective free-energy has been carried out.[61,62] It is found that the substrate SDAP shows a stabilization of -59 kcal/mol, suggesting that when SDAP is transferred from an implicit solvent environment to the interior of the protein, it is significantly stabilized due to the strong electrostatic interactions with the metal ions and certain residues of the protein. The two metal centers, Zn1 and Zn2, exhibit stabilization of free energy by the amount -4.5 and -5.2 kcal/mol, respectively. It is found that the interaction of the substrate with the residues, in particular, Lys175, Arg178, Arg258, Ser290, Thr325, Asp327, Arg329, Asn345, and Lys350 of DapE leads to stabilizing binding energy (Figure 6 and Table S2 in the Supporting Information). The favorable binding energy between the substrate and the active site residues are dominated by electrostatic interaction. Additionally, the glycine residues 322-324 present between α8 and β12 of DapE, which are part of an evolutionarily conserved glycine rich loop (G322-S326) also contribute to the stabilization of the substrate in the enzyme active site. This is in agreement with the earlier finding that the conformational dynamics of this glycine rich loop is associated with the tight binding of the substrate in this active site.[17] Such glycine rich segments are well known to

be evolutionarily conserved and are believed to have important functional roles in many microbial proteins.[67–70]

The consequence of the closing domain motion, seen along the principal component 1 in the enzyme-substrate complex can be further seen in the interaction between the enzyme and substrate. Figure 5b shows the binding free energy of the substrate in the enzyme active site. It is seen, that during the course of the MD simulation of the enzyme-substrate complex, the closing motion of the domain facilitates a stabilizing enzyme-substrate interaction (Figure 5b). To evaluate the effect of conformational dynamics on the enzyme-substrate binding free energy, we have calculated the effective binding free energy of SDAP in DapE enzyme, by considering the MD trajectories filtered along the first two principal components (PC1 and PC2). The average substrate binding free energy along PC1 and PC2 are determined to be -40 and -30 kcal/mol, respectively. The time evolution of the effective binding free energy (Figure 5c) of the substrate along the first two principal components suggests that the substrate binding is stabilized along PC1 during the domain closing motion of the DapE enzyme, in particular after 65 ns of MD simulation (Figure 5c). The strong correlation between the decreasing radius of gyration in the enzyme-substrate complex (Figure 5a), the decreasing total binding free energy of the substrate (Figure 5b), and the binding free energy of the substrate along PC1 (Figure 5c), clearly suggests that the folding motion of the catalytic domains over the dimerization domains in the enzyme-substrate complex (described by the principal component 1) leads to a stronger enzyme-substrate binding. In addition to a strong enzyme-substrate binding, the closing domain motion in the enzyme-substrate complex also leads to repulsion of water molecules away from the enzyme active site, which can be seen from the reduced population of water molecules within 8 Å of the Zn atoms for the enzyme-substrate complex as compared to the apo system (Figure S10 in the Supporting Information). A fewer number of water molecules near the active site indicates the increased hydrophobicity near the active site of the DapE-SDAP complex as a direct consequence of the motion of the catalytic domains towards the dimerization domains.

Inter-domain interactions

The catalytic domains and the dimerization domains of the DapE enzyme exhibit several electrostatic interactions among themselves, which account for the stability of the dimeric protein.[16] Our simulations reveal that the number of hydrogen bonds present between the residues of the catalytic domain and those of dimerization domain are higher in the enzyme-substrate complex, as compared to the apo enzyme (Figure 7a). This fact also emerges from the time-evolution of the interaction energy (electrostatic and van der Waals energy) between the catalytic and dimerization domains of DapE, where the interaction energy in the complex is more stabilized as compared to that of the apo enzyme (Figure 7b). The increased inter-domain interactions impart further stability to the enzyme-substrate complex. Upon comparison of the inter-domain hydrogen bonding in apo and complex form, it is found that the residues Arg179, Pro293, Thr325 and Arg329 of the catalytic domain participate in hydrogen bonding interaction with the residues Ser181, Arg258, and Tyr259 of the dimerization domain in the enzyme-substrate complex (Figure 8). The percentage occupancy of the inter-domain hydrogen bonds in the apo and complex systems are given in Table 1, which shows that the enzyme-substrate complex features more number of stable hydrogen bonds between the two domains, as compared to the apo enzyme. This computational finding is in agreement with the recent kinetic study of the enzyme catalysis by DapE, where the crucial role of the dimerization domain in the enzyme activity is described by inter-domain interactions.[18]

In addition to the direct interactions between the residues of the catalytic domain and dimerization domain of DapE, our simulations also reveal the presence of several water mediated hydrogen bond chains between the catalytic domain and the dimerization domain. Interestingly, these hydrogen bond chains provide a connection between the substrate, bound in the catalytic domain of the enzyme, with the dimerization domain. Figure 9 shows two representative snapshots of such hydrogen bond chains, where the substrate shows a salt-bridge interaction with the Arg178 of the catalytic domain which further participates in a hydrogen bond chain with the side chains of Thr261 via two mediating water molecules (Figure 9a). Similar interactions are also seen, where the substrate exhibits direct electrostatic interaction with Thr325 and Arg329 of the catalytic do-

main, which in turn establish contacts with the side chains of Arg258 and Ser290 of the dimerization domain, respectively. While the former inter-domain contact is water mediated, the latter is direct interaction between the catalytic and dimerization domains (Figure 9b). In summary, it is found that the number of interactions between the catalytic and dimerization domains, both direct as well as bridged through substrate and water molecules, increased during the MD simulation of enzyme-substrate complex, as compared to the apo enzyme.

Inter-chain interaction (dimerization energy)

In addition to the interaction between the dimerization and catalytic domains of a monomer in the dimeric DapE enzyme, we have also evaluated the interaction between the two monomers (chain A and chain B, in Figure 1a). The average value of the effective binding free energy of the dimer formation (the dimerization free energy) is evaluated to be -63 kcal/mol and -60 kcal/mol, for the apo-DapE and DapE-SDAP complex systems, respectively, suggesting a minor effect of the substrate binding on the dimer formation. The per-residue decomposition of the effective binding free-energy in the apo and complex system (Figure S11) reveals that the interaction between the two chains are primarily between the two dimerization domains which is evident from the structure of the enzyme (Figure 1a). A very similar per-residue free energy in the apo and complex system (Figure S11) suggests that the substrate binding has negligible effect on the dimerization free energy. This fact is further supported by the nearly equal number of inter-chain hydrogen bonds in the apo-DapE and DapE-SDAP complex (Figure S12). In both apo DapE as well as DapE-SDAP complex, the interaction between the two chains of the dimeric protein are facilitated by the α-5 helices (Asn203-Tyr218) and β-9 strands (Ser231-Ala239) of both monomers (Figure S13).[16] To account for the contribution of conformational dynamics to the dimer formation, we evaluated the dimerization free energy from the trajectory filtered along principal components 1 and 2. The dimerization free energy along the MD trajectory and the trajectories filtered along principal components 1 and 2 are found to be very similar (Figure S14) which can be explained by the fact that the first two principal components involve significant conformational dynamics of the two catalytic

domains with little change in the relative motion of the two dimerization domains, which on the other hand, contribute almost exclusively to the dimerization free energy.

Conclusions

The microbial enzyme DapE plays a crucial role in the lysine biosynthetic pathway of bacteria which has enormous potential to be the source of several antibiotic targets. We have carried out 100 ns of MD simulations of the apo DapE enzyme and its complex with its natural substrate L,L-SDAP. The analysis of the MD trajectories, in terms of root mean squared deviations, secondary structure analysis, and number of hydrogen bonds, show the plasticity of the native structure of the enzyme both in apo and complex states. While the dimerization domain of the dimeric protein provides rigidity due to strong secondary structure components, the catalytic domain, hosting several loops imparts the flexibility to the protein to undergo conformational dynamics.

The conformational dynamics of the protein involves short timescale local fluctuations coupled with long timescale global motion and domain movements. To decouple these multiscale motions, the high dimensional conformational dynamics of the enzyme is further investigated by the so-called principal component analysis, wherein, the concerted motion of the protein is captured by a few principal components and the local fluctuations by the remaining many principal components. In this work, we have carried out principal component analysis of the MD trajectories of apo enzyme and complex, and characterized the first two principal components in both the systems in terms of the internal coordinates of the enzyme. The principal component 1 in the DapE-SDAP complex depicts the catalytic domains fold over the dimerization domains, which is supported by the C_α-C_α cross-correlation maps and radius of gyration along this mode. The distance between the centers of mass of the catalytic domain and dimerization domain is found to provide an adequate description of the principal component 1. On the other hand, the principal component 2 describes a rotation of the two catalytic domains in opposite direction about an axis formed by the two dimerization domains, in both apo enzyme and enzyme-substrate complex, which is character-

ized by a dihedral angle between the centers of mass of catalytic domain A, dimerization domain A, dimerization domain B, and catalytic domain B.

The folding motion of the catalytic domain over the dimerization domain, described by the principal component 1 in the enzyme-substrate complex, leads to expulsion of water molecules near the active site, as seen from reduced number of water molecules near the active site of the enzyme-substrate complex as compared to the apo enzyme. This motion enhances the tight binding of the substrate in the enzyme active site via the stabilizing interaction energy between the substrate with the DapE residues such as Lys175, Arg178, Thr325, Asp327, Arg329, Asn345, and Lys350. The domain folding motion has also resulted in strengthening the interaction between the dimerization and catalytic domains of DapE, which is reflected from the increased number and lifetime of inter-domain hydrogen bonds as well as the stabilized inter-domain interactions. The substrate, located in the catalytic domain, is seen to participate in indirect interactions with the side chains of the dimerization domain via several chains of hydrogen bonds, such as, SDAP-Arg178-water-T261, SDAP-Arg329-Ser290, SDAP-Thr325-water-Arg258, emphasizing the role of the dimerization domain in the enzyme catalysis.

It is hoped, that the present work will motivate further experimental and computational investigations to explore the links between conformational dynamics and physiological role of this very important class of proteins.

Acknowledgement

SM acknowledges financial support from the DST-India, New Delhi for Ramanujan award and IIT Kharagpur for ISIRD grant. The authors thank Dr. Marc Creus for valuable discussion.

Supporting information

Supporting information available for additional figures and tables.

References

(1) Donadio, S.; Maffioli, S.; Monciardini, P.; Sosio, M.; Jabes, D. *J. Antibiot.* **2010**, *63*, 423–430.

(2) Davies, J.; Davies, D. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 417–433.

(3) Wilke, M. S.; Lovering, A. L.; Strynadka, N. C. J. *Curr. Opin. Microbiol.* **2005**, *08*, 525–533.

(4) Levy, S. B.; Marshall, B. *Nat. Med.* **2004**, *10*, 122–129.

(5) Coates, A.; Hu, Y.; Bax, R.; Page, C. *Nat. Rev. Drug. Discov.* **2002**, *1*, 895–910.

(6) Bienvenue, D.; Gillner, D.; Davis, R.; Bennett, B.; Holz, R. *Biochemistry* **2003**, *42*, 10756–10763.

(7) Uda, N.; Upert, G.; Angelici, G.; Nicolet, S.; Schmidt, T.; Schwede, T.; Creus, M. *Metallomics* **2014**, *6*, 88–95.

(8) Usha, V.; Lloyd, A.; Lovering, A.; Besra, G. *FEMS. Microbiol. Lett.* **2012**, *330*, 10–16.

(9) Gillner, D.; Armoush, N.; Holz, R.; Becker, D. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6350–6352.

(10) Girodeau, J. M.; Agouridas, C.; Masson, M.; LeGoffic, F. *J. Med. Chem.* **1986**, *29*, 1023–1030.

(11) Karita, M.; Etterbeek, M. L.; Forsyth, M. H.; Tummuru, M. R.; Blaser, M. J. *Infect. Immun.* **1997**, *65*, 4158–4164.

(12) Gillner, D.; Becker, D.; Holz, R. *J. Biol. Inorg. Chem.* **2013**, *13*, 155–163.

(13) Pavelka, M.; Jacobs, W. *J. Bacteriol.* **1996**, *178*, 6496–6507.

(14) Velasco, A. M.; Leguina, J. I.; Lazcano, A. *J. Mol. Evol.* **2002**, *55*, 445–459.

(15) Born, T. L.; Zheng, R.; Blanchard, J. S. *Biochemistry* **1998**, *37*, 10478–10487.

(16) Nocek, B.; Gillner, D.; Fan, Y.; Holz, R.; Joachimiak, A. *J. Mol. Biol.* **2010**, *397*, 617–626.

(17) Dutta, D.; Mishra, S. *Phys. Chem. Chem. Phys.* **2014**, *16*, 26348–26358.

(18) Nocek, B.; Starus, A.; Makowska-Grzyska, M.; Gutierrez, B.; Sanchez, S.; Jedrzejczak, R.; Mack, J.; Olsen, K.; Joachimiak, A.; Holz, R. C. *PLOS ONE* **2014**, *9*, e93593.

(19) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290.

(20) Gerstein, M.; Lesk, A. M.; Chothia, C. *Biochemistry* **1994**, *33*, 6739–6749.

(21) Biehl, R.; Hoffmann, B.; Richter, D. *Phys. Rev. Lett.* **2008**, *101*, 138102–138110.

(22) Bhabha, G.; Ekiert, D. C.; Jennewein, M.; Zmasek, C. M.; Tuttle, L. M.; Kroon, G.; Jane Dyson, H. J.; Godzik, A.; Wilson, I. A.; Wright, P. W. *Nat. struct. Mol. Biol.* **2013**, *20*, 1243–1252.

(23) Henzler-Wildman, K. A.; Thai, V.; Kern, D. *Nature* **2007**, *450*, 838–844.

(24) Bennett, W. S.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 4848–4852.

(25) Kay, L. E. *Nat. Struct. Biol.* **1998**, *5*, 513–517.

(26) Inoue, R.; Biehl, R.; Rosenkranz, T.; Fitter, J.; Monkenbusch, M.; Radulescu, A.; Farago, B.; Richter, D. *Biophys. J.* **2010**, *99*, 2309–2317.

(27) Gur, M.; Madura, J. D.; Bahar, I. *Biophys. J.* **2013**, *105*, 1643–1652.

(28) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. *Annu. Rev. Biophys.* **2010**, *39*, 23–42.

(29) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Structure, Function, and Bioinformatics* **1993**, *17*, 412.

(30) Levy, R. M.; Srinivasan, A.; Olson, W.; McCammon, J. *Biopolymers* **1984**, *23*, 1099–1122.

(31) Ichiye, T.; M., K. *Proteins* **1991**, *11*, 205–217.

(32) Lou, H.; Cukier, R. A. *J. Phys. Chem. B* **2006**, *110*, 12796–12808.

(33) Maisuradze, G. G.; Leitner, D. M. *Chem. Phys. Lett.* **2006**, *421*, 5–10.

(34) Feig, M.; Karanicolas, J.; Brooks III, C. L. *J. Mol. Graphics Modell.* **2004**, *22*, 377–395.

(35) Stote, R. H.; Karplus, M. *PROTEINS: Structure, Function, and Genetics* **1995**, *23*, 12–31.

(36) Li, P.; Kenneth M. Merz, J. *J. Chem. Theory Comput.* **2014**, *10*, 289–297.

(37) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comp. Phys.* **1977**, *23*, 327 – 341.

(39) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pederson, L. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(40) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(41) Foster, J. P.; Weinhold, F. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(42) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407–1413.

(43) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2010**, *31*, 671–690.

(44) Kitao, A.; Hirata, F.; Go, N. *Chem. Phys.* **1991**, *158*, 447–472.

(45) Garcia, A. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.

(46) Van Aalten, D. M. F.; Groot, B. L. D.; Findlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. *J. Comput. Chem.* **1997**, *18*, 169–181.

(47) Arcangeli, C.; Bizzarri, A. R.; Cannistraro, S. *Biophys. Chem.* **2001**, *90*, 45–56.

(48) Kubitzki, M. B.; de Groot., B. L. *Biophys. J.* **2007**, *92*, 4262–4270.

(49) Maisuradze, G. G.; Liwo, A.; Scherga, H. A. *J. Mol. Biol.* **2009**, *385(1)*, 312–329.

(50) Skjaerven, L.; Martinez, A.; Reuter, N. *Proteins* **2011**, *79*, 232–243.

(51) Laberge, M.; Yonetani, T. *Biophys. J.* **2008**, *94*, 2737–2751.

(52) Homeyer, N.; Gohlke, H. *Molec. Inf.* **2012**, *31*, 114–122.

(53) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. *J. Biol. Struct. Dyn.* **1998**, *16*, 671–682.

(54) Wang, K.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.

(55) Bill R. Miller, I.; T. Dwight McGee, J.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(56) Connolly, M. L. *J Appl Cryst* **1983**, *16*, 548–556.

(57) Pearlman, D. A. *J Med Chem* **2005**, *48*, 7796–7807.

(58) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J Med Chem* **2006**, *49*, 4805–4808.

(59) Massova, I.; Kollman, P. *Perspect. Drug Discov.* **2000**, *18*, 113–135.

(60) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522–1542.

(61) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891–913.

(62) Metz, A.; Pfleger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K.-H.; Gohlke, H. *J. Chem. Inf. Model.* **2012**, *52*, 120–133.

(63) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(64) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(65) Loof, H.; Nillson, L.; Rigler, R. *J. Am. Chem. Soc.* **1992**, *114*, 4028–4035.

(66) Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.

(67) Wilson, K. A.; Baer, S.; Maerz, A. L.; Alizon, M.; Poumbourios, P. *J. Virol.* **2005**, *79*, 4533–4539.

(68) Wilson, K. A.; Maerz, A. L.; Baer, S.; Drummer, H. E.; Poumbourios, P. *Biochem. Biophys. Res. Commun.* **2007**, *359*, 1037–1043.

(69) Baumann, T.; Kaempfer, U.; Schuerch, S.; Schaller, J.; Largiader, C.; Nentwig, W.; Kuhn-Nentwig, L. *Cell Mol. Life Sci.* **2010**, *67*, 2787–2798.

(70) Harrison, C. F.; Lawson, V. A.; Coleman, B. M.; Kim, Y. S.; Masters, C. L.; Cappai, R.; Barnham, K. J.; Hill, A. F. *J. Biol. Chem.* **2010**, *285*, 20213–20223.

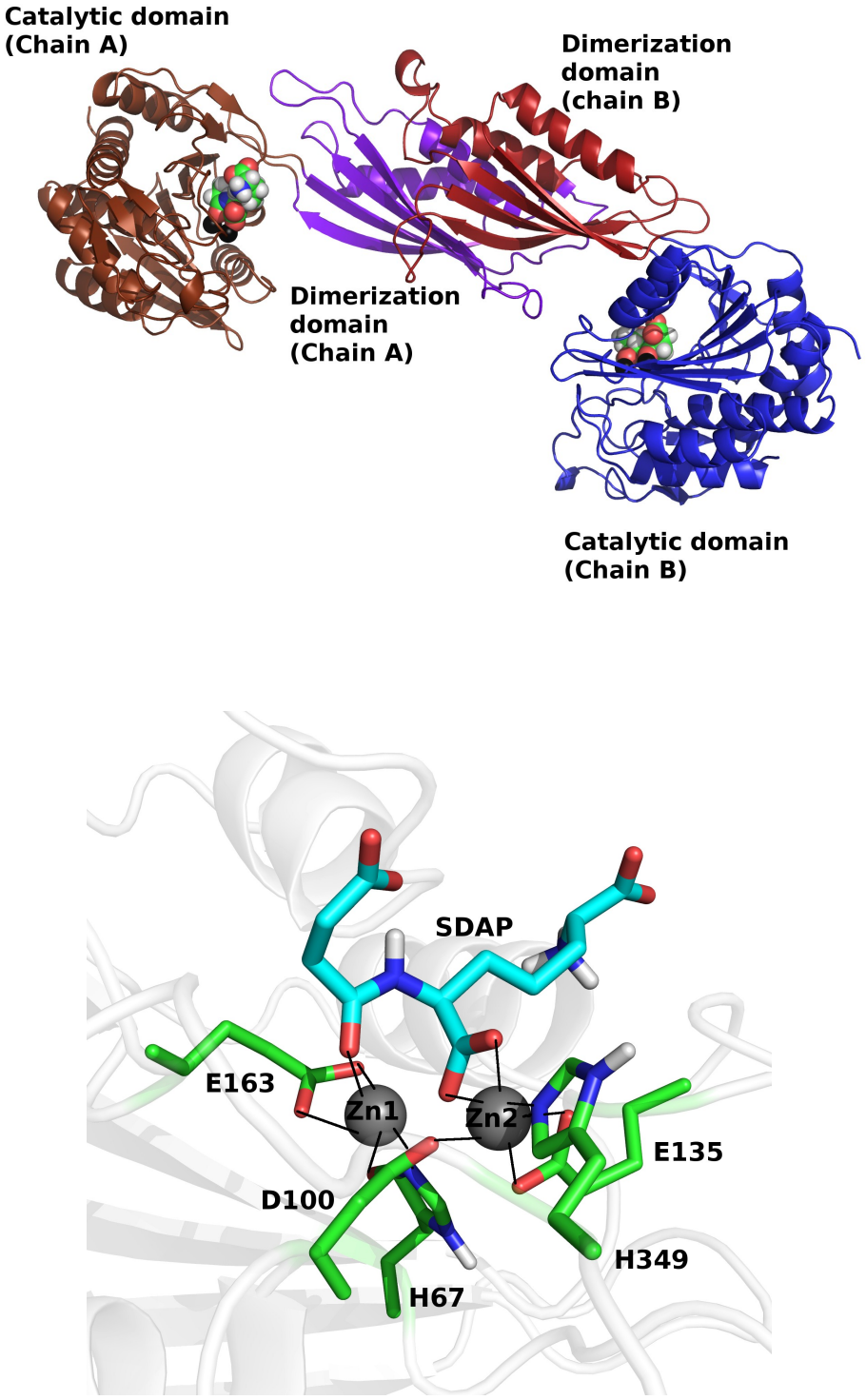


Figure 1: (a) The dimeric structure of DapE showing the catalytic and dimerization domains of chain A and chain B. (b) The active site in the catalytic domain of DapE enzyme. The protein backbones are shown as ribbons. The Zn atoms are shown as spheres and the side chains of the active site residues and the substrate are shown as sticks.

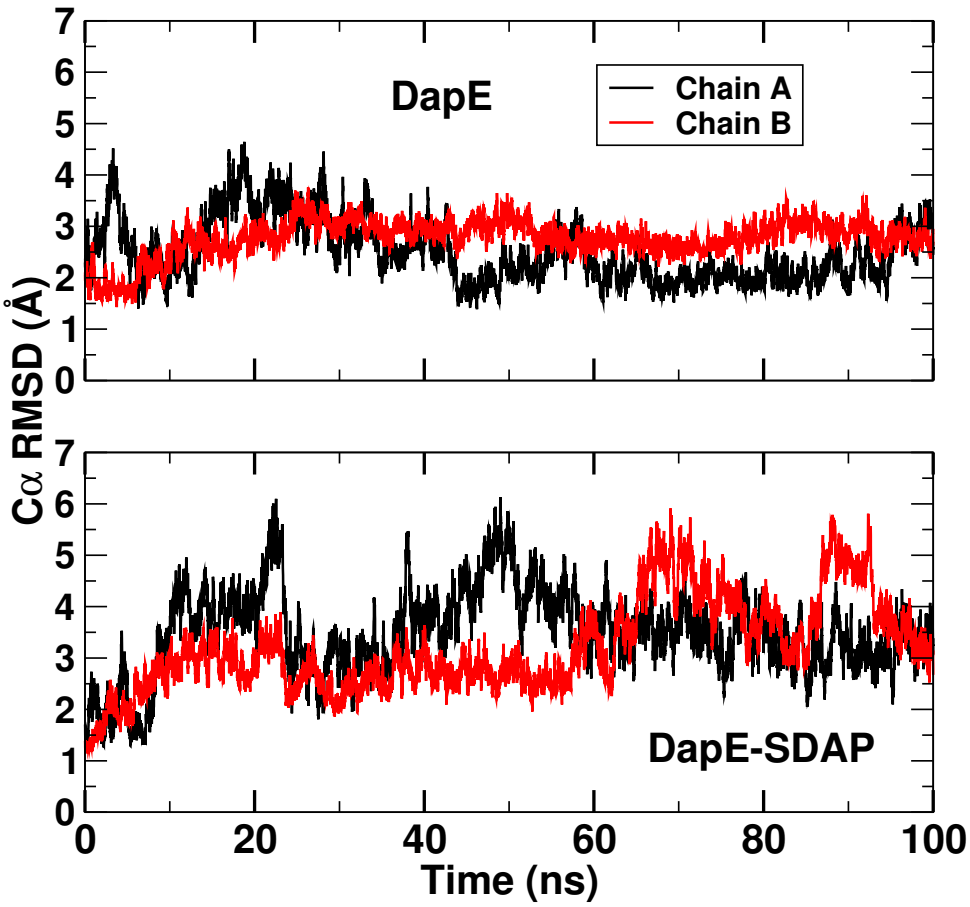


Figure 2: Root mean squared deviation (in Å) of C_α atoms of chain A (black) and chain B (red) in apo DapE (upper panel) and DapE-SDAP complex (lower panel).

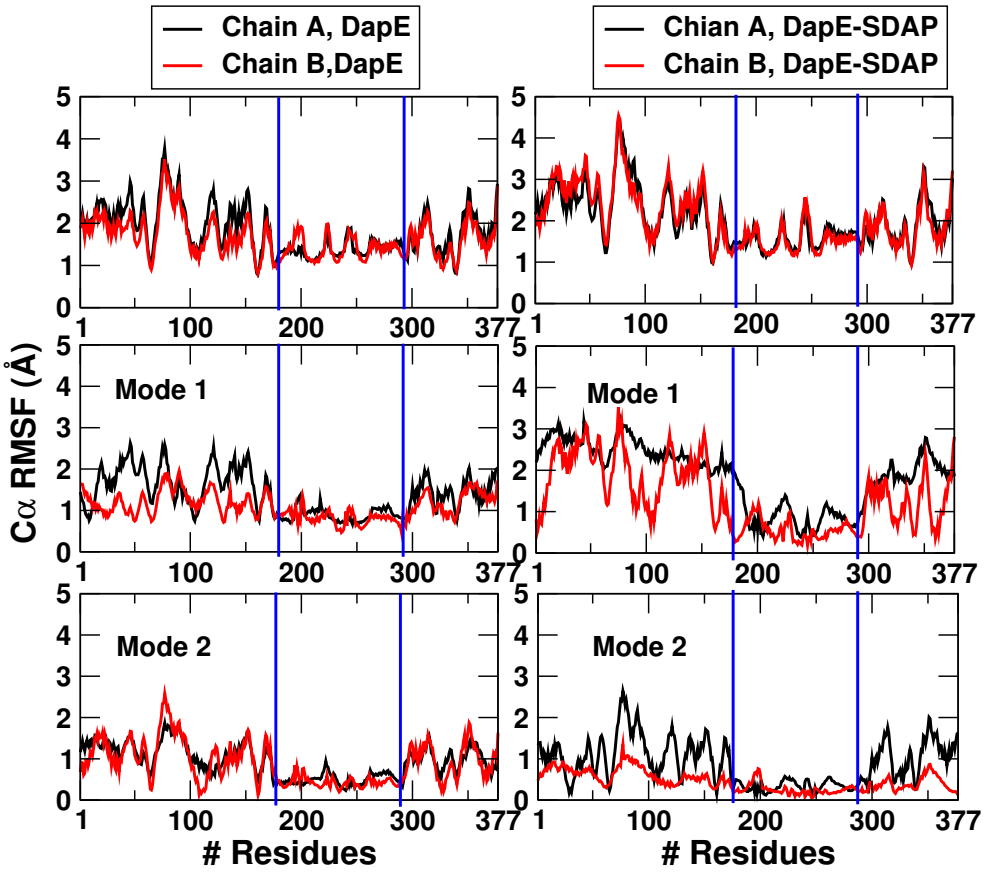


Figure 3: The root mean square fluctuations (in Å) (averaged over intervals of 10 ns after excluding first 20 ns of the MD trajectories) of C_α atoms of chain A and chain B in DapE and DapE-SDAP complex (upper panels) together with the C_α root mean squared fluctuations in principal component 1 (middle panels) and principal component 2 (lower panels). The vertical lines demarcate the dimerization domain (residue 180 to 292) from the catalytic domain.

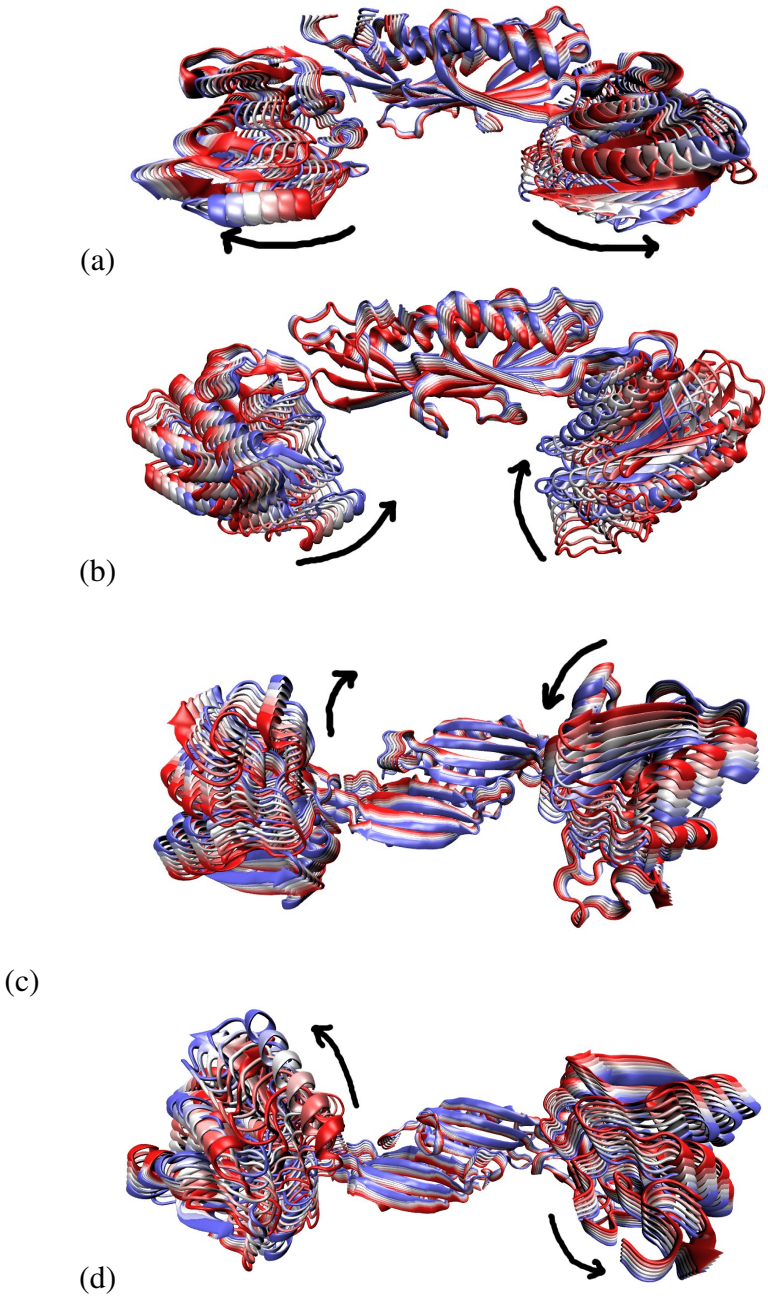


Figure 4: Overlap of six snapshots (red to blue) along principal component 1 of (a) DapE and (b) DapE-SDAP complex and along principal component 2 of (c) DapE and (d) DapE-SDAP complex. The principal component 1 corresponds to the motion of catalytic domains away from the dimerization domains (DapE) or towards the dimerization domains (DapE-SDAP), while principal component 2 corresponds to a rotation of the catalytic domains about the dimerization domains in opposite directions. The direction of motions are indicated by arrowheads.

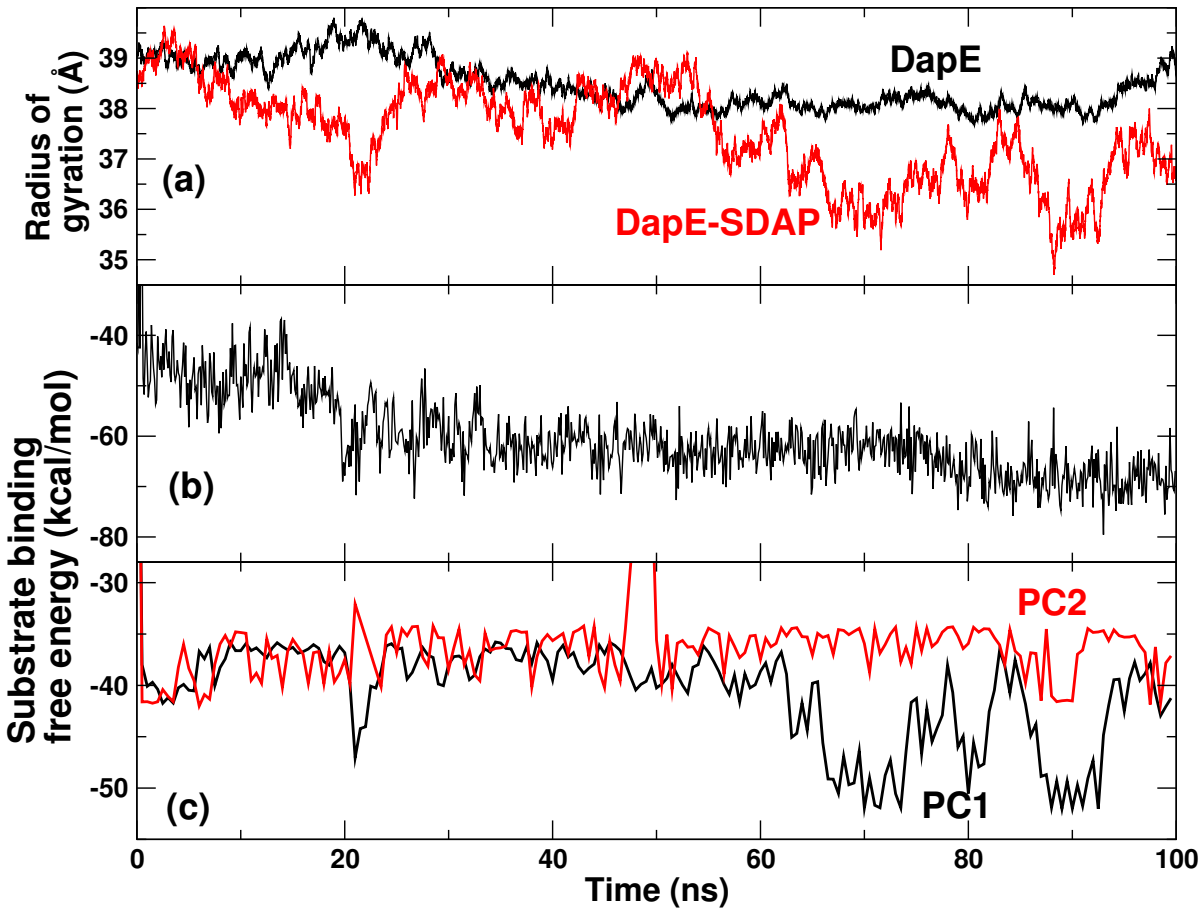


Figure 5: (a) The time series of the radius of gyration along the principal component 1 for DapE and DapE-SDAP complex. (b) The substrate binding free energy during the 100 ns MD simulation. (c) The substrate binding free energy from the trajectories filtered along the principal components 1 and 2.

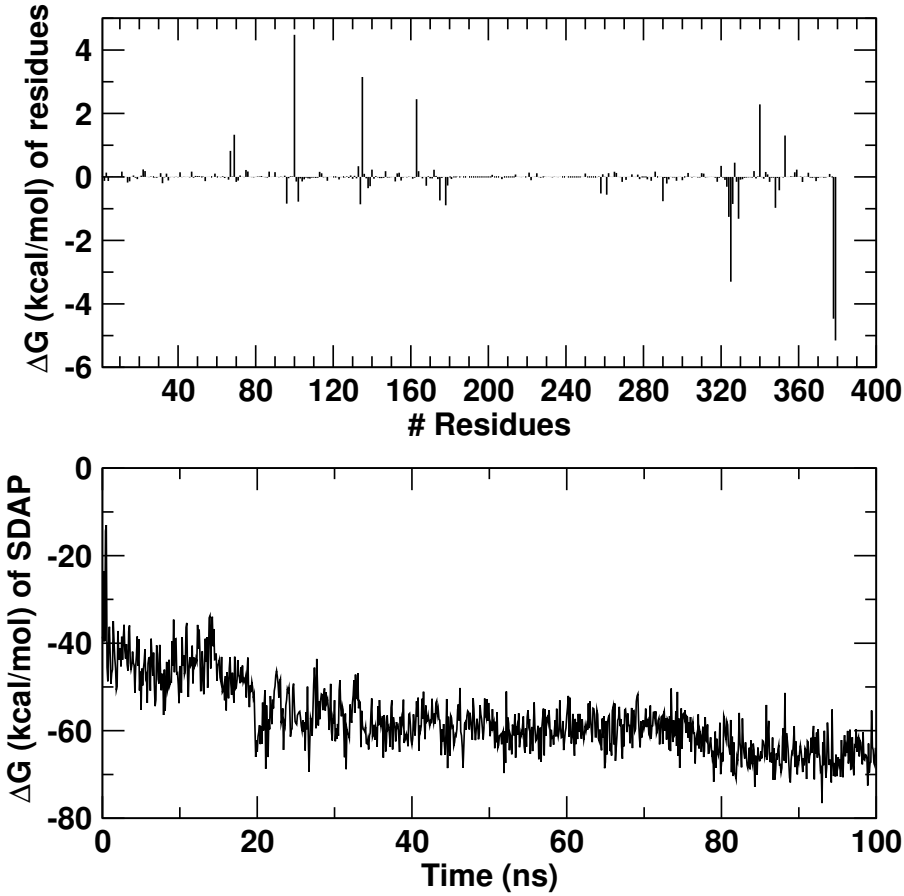


Figure 6: Residue-wise decomposition of the total binding free energy obtained from MM-PBSA analysis of DapE-SDAP MD trajectory (upper panel). The time evolution of the contribution of the substrate to the total binding free energy during the MD simulation of DapE-SDAP complex (lower panel).

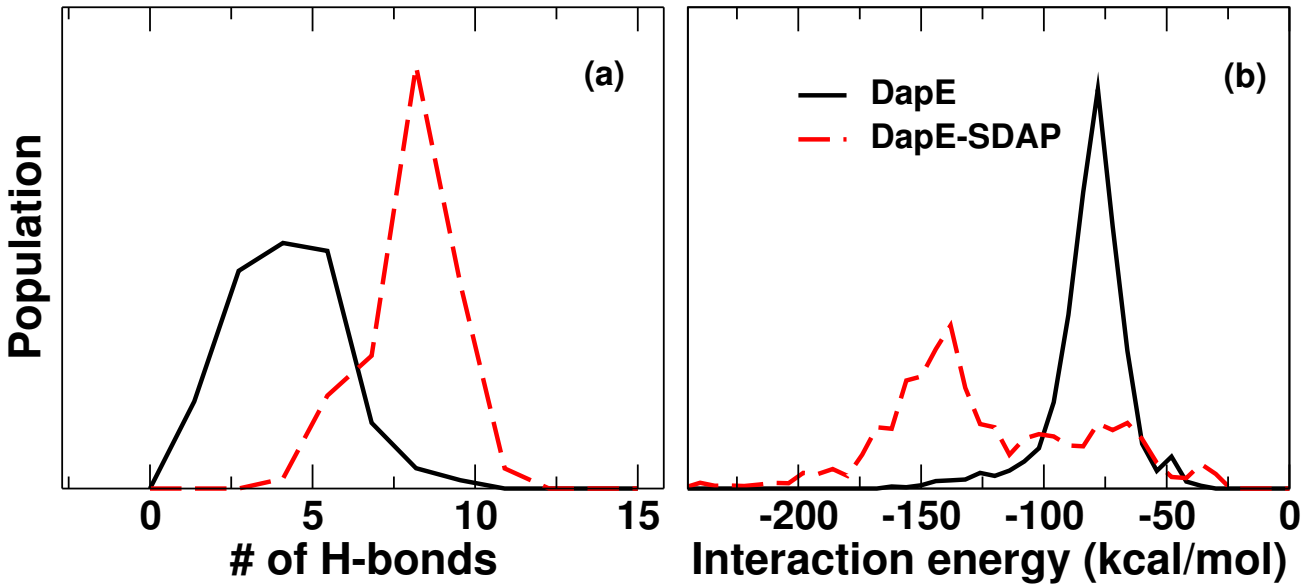


Figure 7: (a) Number of hydrogen bonds between the residues of the catalytic domain and dimerization domain during 100 ns of MD simulation of DapE enzyme and DapE-SDAP complex. (b) Interaction energy (electrostatic and van der Waals) in kcal/mol between the catalytic domain and dimerization domain during 100 ns of MD simulation of DapE enzyme and DapE-SDAP complex (right panel).

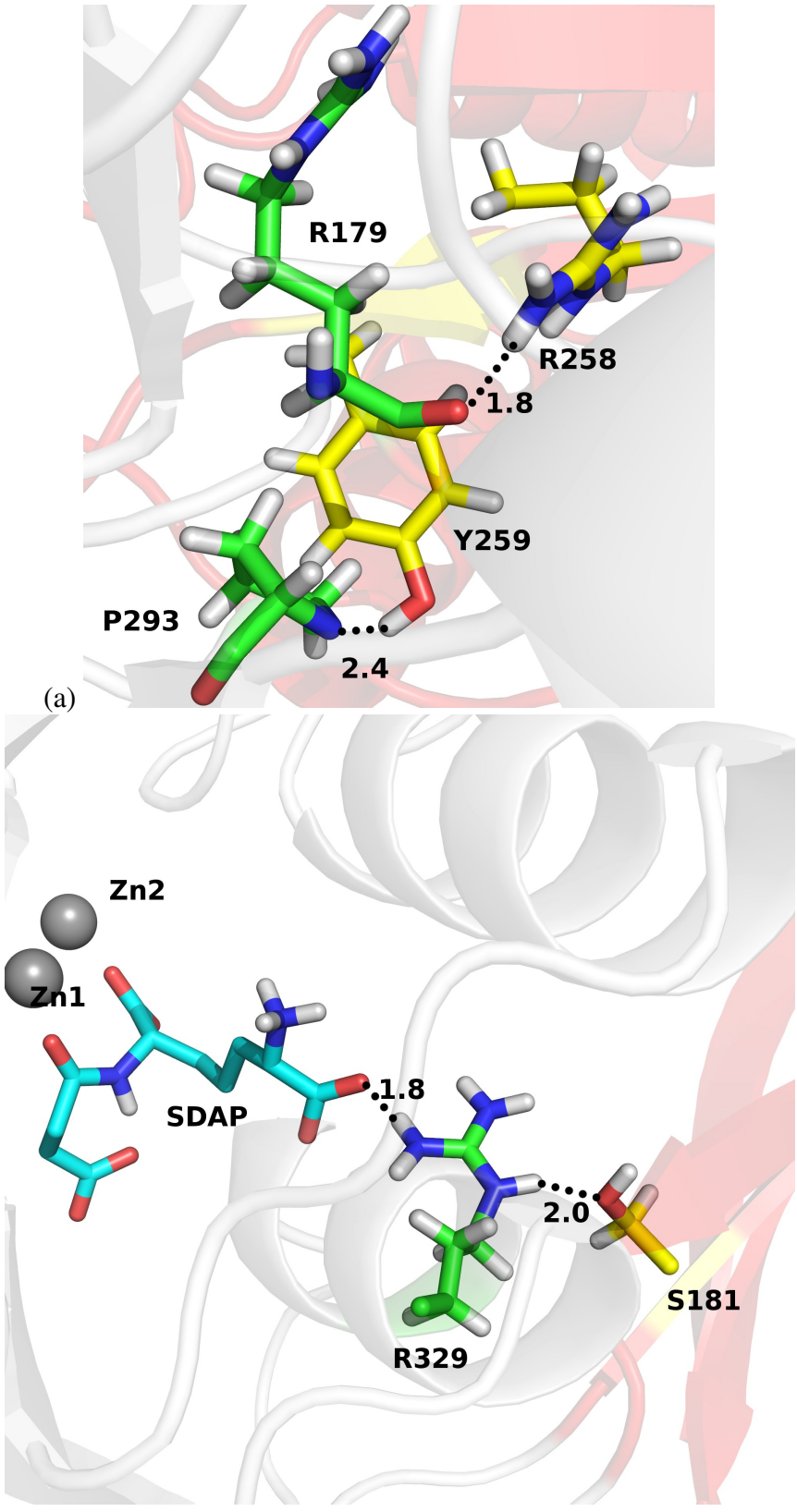
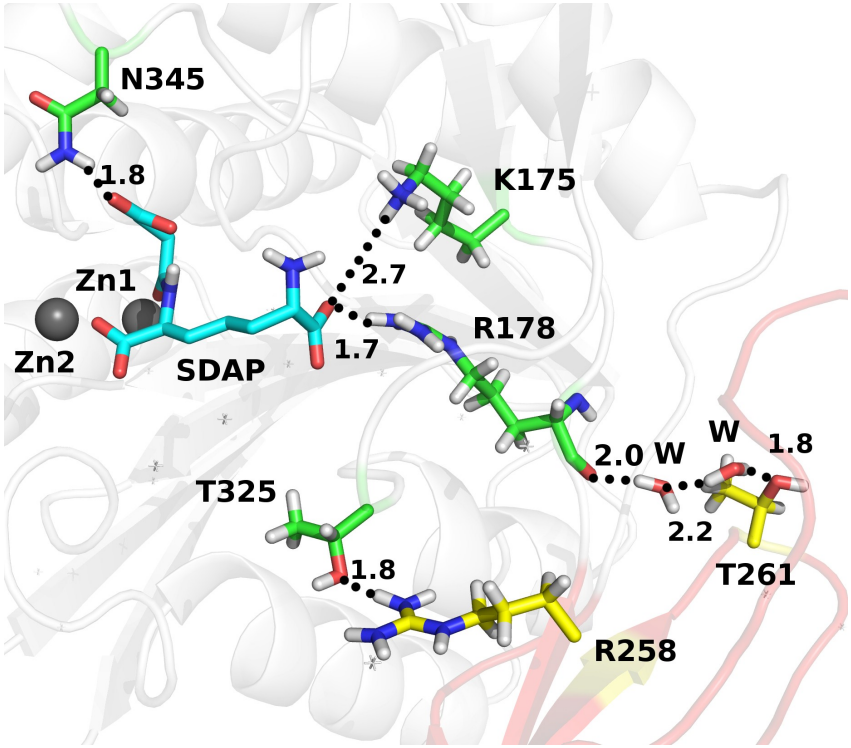
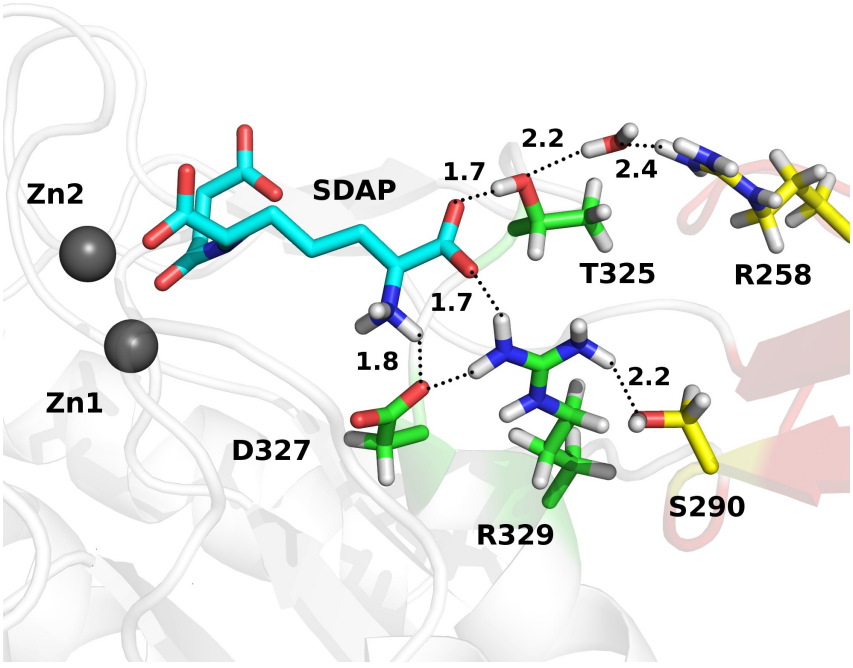


Figure 8: The hydrogen bond interaction between the residues of catalytic domain and dimerization domain, present in the DapE-SDAP complex but absent in apo DapE. The carbon atoms of substrate, catalytic domain, and dimerization domains are shown in cyan, green, and yellow, respectively. The numbers (in Å) indicate the distance between the hydrogen atom and hydrogen bond acceptor atom.



(a)



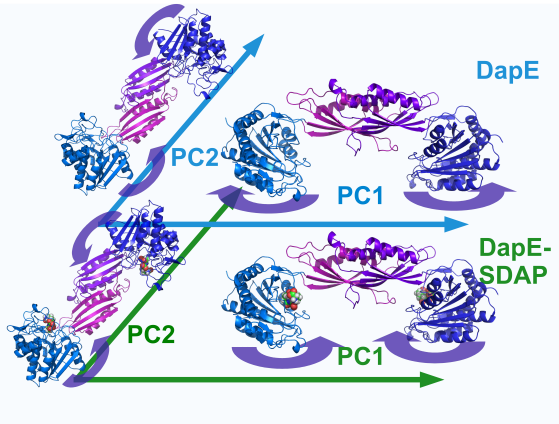
(b)

Figure 9: Hydrogen bond interactions between the substrate SDAP and the side chains of catalytic domain and the dimerization domain of DapE, mediated by water molecules. The carbon atoms of substrate, catalytic domain, and dimerization domains are shown in cyan, green, and yellow, respectively. The numbers (in Å) indicate the distance between the hydrogen atom and the hydrogen bond acceptor atom.

Table 1: The percentage occupancy of hydrogen bonds between the residues of the catalytic and dimerization domains in the apo enzyme and enzyme-substrate complex from 100 ns MD simulations.

Residue		Occupancy (%)	
catalytic domain	dimerization domain	apo	complex
Thr325	Arg258	0	24
Pro293	Tyr259	0	29
Arg329	Ser181	1	35
Arg179	Arg258	2	36
Arg329	Ser290	28	75

Table of Contents Graphics



Conformational dynamics induced by substrate binding in DapE enzyme.