J Mol Model (2011) 17:1241–1249
DOI 10.1007/s00894-010-0819-0

ORIGINAL PAPER

Experimental and computational studies indicate the mutation of Glu12 to increase the thermostability of oligomeric protease from *Pyrococcus horikoshii*

Dongling Zhan · Weiwei Han · Yan Feng

Received: 16 April 2010 / Accepted: 21 July 2010 / Published online: 15 August 2010

Abstract The intracellular protease from *Pyrococcus horikoshii* (PhpI) is a member of the DJ-1/ThiJ/PfpI superfamily, which is suggested to be involved in cellular protection against environmental stresses. In this study, flexible docking approach was employed to dock the ligand into the active site of PhpI. By analyzing the results, active site architecture and certain key residues responsible for substrate specificity were identified on the enzyme. Our docking result indicates that Glu12 plays an important role in substrate binding. The kinetic experiment conducted by Zhan shows that the E12T mutant is more stable than that of the wild-type. We also predict that Glu15, Lys43, and Tyr46 may be important in the catalytic efficiency and thermostability of enzyme. The new structural and mechanistic insights obtained from computational study should be valuable for detailed structures and mechanisms of the member of the DJ-1 superfamily.

Keywords Intracellular protease *Pyrococcus horikoshii* (PhpI) · Molecular docking · Molecular dynamics · Thermostabilities

Weiwei Han and Dongling Zhan contributed equally to this work.

D. Zhan · W. Han (✉) · Y. Feng (✉)
Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, Jilin University,
130023 Changchun, China
e-mail: weiweihan@jlu.edu.cn
e-mail: yfeng@mail.jlu.edu.cn

D. Zhan
College of Food Science and Engineering,
Jilin Agricultural University,
130118 Changchun, China

Introduction

Oligomeric protease from *Pyrococcus horikoshii* (PhpI) is a member of the DJ-1/ThiJ/PfpI superfamily, which includes proteins with diverse functions, such as chaperone (*E. coli* Hsp31) [1–4], protease (PhpI) from *Pyrococcus horikoshii* (PDB code 1G2I) [5], *Escherichia coli* YhbO [6], the Parkinson disease protein DJ-1 [7–12], the protein DR1199 [13] and so on. Furthermore, all members characterized of the DJ-1 superfamily are oligomers, though the oligomer polymerization order and interface vary and appear to be related to the particular set of insertions into the core fold [14, 15]. Although the function of many members of the DJ-1 superfamily remains unknown, those that have been characterized are mostly involved in the cell response to stress [8–17]. For example, *Escherichia coli* Hsp31 is involved in thermal stress protection and acid resistance [1–4], human DJ-1 in cellular protection against oxidative stress [7–12, 15–17]. In 2003, this superfamily has attracted a great deal of interest because one of its members, the human protein DJ-1, has been identified independently as both an oncogene when overexpressed and a neuroprotective protein whose absence or inactivation can cause certain types of early-onset Parkinsonism [7].

PhpI is an intracellular cysteine peptidase characterized by its stability and proposed to represent the predominant proteolytic activity of this thermophilic archaebacterium, and is classified in the C56 family of peptidase (http://merops.sanger.ac.uk). It forms a hexameric structure and the active sites are formed at the interfaces between three pairs of monomers. The shared active site between subunits A and C of PhpI performs proteolytic cleavage through a Cys–His–Glu catalytic triad: Cys100 and His101 reside on the A subunit, whereas Glu74 is provided by the neighboring C subunit, and Cys500 and His 501 is on the C subunit,

whereas Glu74 is in the A subunit (Fig. 1b). Cysteine acts as nucleophile, and His–Glu functions as catalytic base, which assists the rate-determining nucleophilic attack of sulfur (seen from Fig. 1a). PhpI is the only member that shows proteolytic activity with hexamers ring in DJ-1 superfamily. The unique structure of PhpI may influence its substrate specificity.

In the present investigation, the docking method was used when Glu12 plays an important role in substrate binding and validated by experiment. The new structural and mechanistic insights obtained from computational study should be valuable for detailed structures and mechanisms of the members of the DJ-1 superfamily.

Methods

Flexible molecular docking

The initial structures of substrates were built by Builder module. The Affinity module of InsightII [18] was employed to dock the ligand into the active site of PhpI. Affinity is applied with a procedure combining Monte Carlo (MC) and simulated annealing (SA) to search for the optimal orientation of ligand binding [19]. This method employs a purely repulsive bonded quartic potential for modeling van der Waals (vdW) interactions, and Coulombic interactions are zero.

MD simulation

An all atom model of PhpI was generated with xleap module in Amber9 [20] on the basis of the initial model. To release conflicting contacts among residues, energy minimization was performed. The protein was then solvated with water in a truncated octahedron periodic box. The TIP3P [21] water model was used. Finally, 3.5 ns MD simulation was conducted at 1 atm and 358 K with the NPT ensemble. During the simulation, SHAKE algorithm [22] was applied to constrain the covalent bonds to the hydrogen atoms. The ff03 all atom Amber force field was used for the protein.

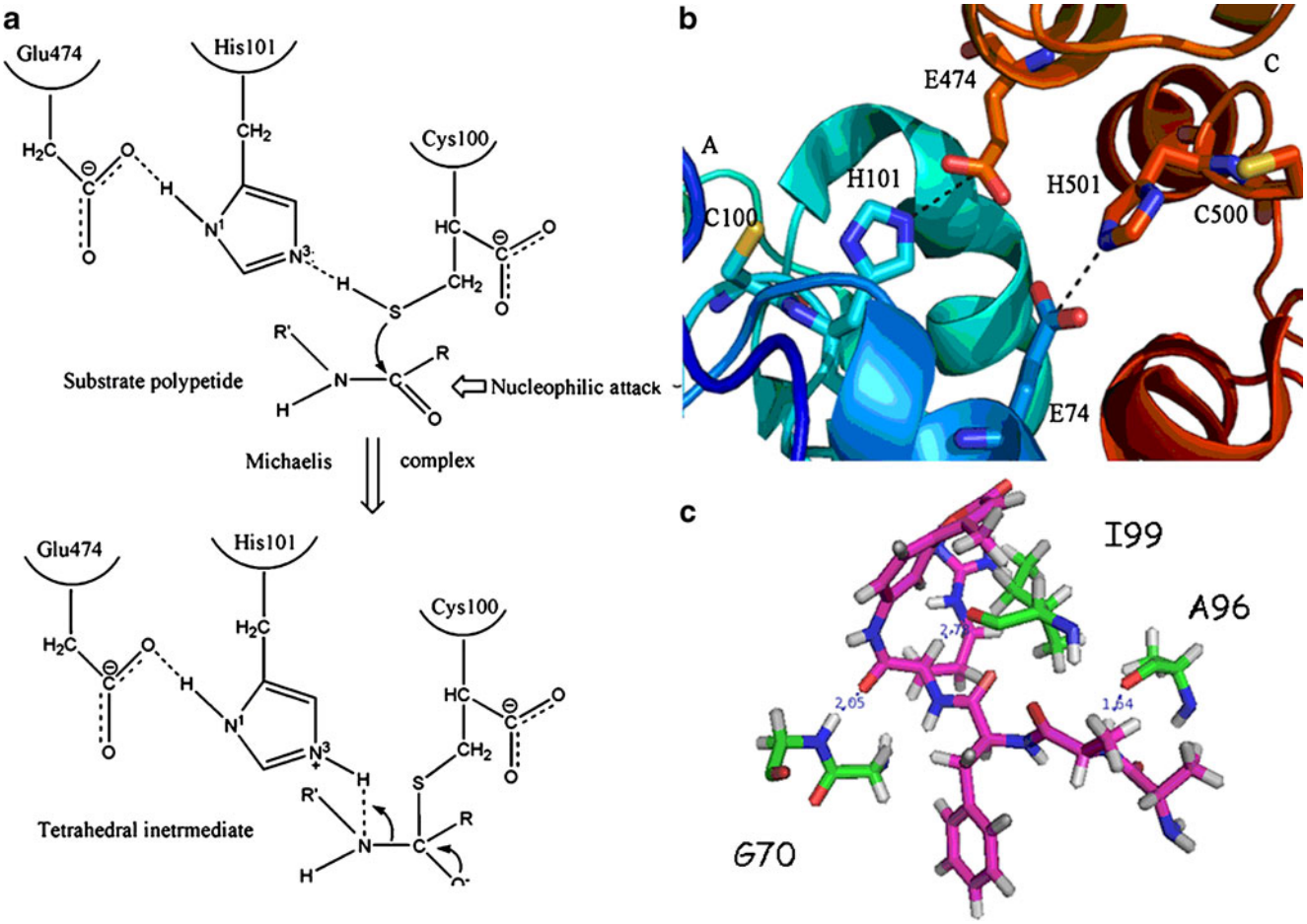

Fig. 1 **a** Proposed mechanism of PhpI (see text for details). The nucleophilicity of Cys100 is enhanced by the catalytic base His101-Glu474. **b** The catalytic triad Cys100, His101, Glu474 and its symmetry-related counterparts Cys500, His501, Glu74. **c** The hydrogen bonds between PhpI and AAFR-AMC

Model quality assessment

The overall quality of the mutant model of PhpI was assessed with respect to its geometry and energy. Profiles-3D and Prosa2003 [23, 24] were used to evaluate the quality of consistency between the native fold and the sequence and to examine the energy of residue–residue interactions. The energy is transformed to a z-score by:

$$Z_{s,c} = \frac{E_{S,C} - \hat{E}_{S,C}}{\sigma_S}, \quad (1)$$

where $E_{S,C}$ and $\hat{E}_{S,C}$ are, respectively, the residue–residue interaction energy and average residue–residue interaction energy of the sequence S in conformation space C, and $\sigma_{S,C}$ is the associated standard deviation.

Expression and purification of the wild type PhpI and the mutant E12T

The expression vector pET15b containing the PhpI protease gene was constructed previously; in this study, the plasmid pET15b was used as the template to obtain mutant E12T. According to the instruction manual of the QuikChange Site-Directed Mutagenesis, the product of the reaction was transformed into *E.coli* DH5α, and then plated on LB agar containing 100 μg/ml ampicillin. The mutated plasmid, pET15b was transformed into *E.coli* BLP and plated on 2YT agar containing 100 μg/ml ampicillin. The colonies which expressed the target proteins were acquired from the plates, respectively. The mutant and WT were then cultivated in 1 L of 2YT culture medium containing 100 μg/ml ampicillin and grown to mid-exponential phase (OD_{600}=0.6-0.8) at 310 K. The cells were then induced for 12 h with 1 mM IPTG (isopropyl-b-D-thiogalactopyranoside) at 299 K. The cells were deposited and resuspended in 50 mM Tris buffer at pH 7.5. After sonicating with middle pulse extension on ice, the supernatant was heat-treated at 358 K for 30 min and obtained by centrifugation (12000 g at 277 K for 20 min), and the purified target proteins of WT and E12T were dialyzed and analyzed by SDS–PAGE.

Enzyme assays

Protease activity was detected by cleavage of 7-amido-4-methylcoumarin of AAFR-AMC. Each assay tube contained 200 μl of substrate stock (100 mM in dimethyl sulfoxide), 50 μl of diluted enzyme sample, and 250 μl of 50 mM sodium phosphate buffer, pH 7.5. The assay tubes were incubated at 358 K in the thermocycler for 60 min, chilled on ice, and centrifuged. and fluorescence was measured at 360±40 nm excitation and 460±40 nm emission. Fluorescence units measured were converted to picomoles of AMC released by using a standard curve prepared with known dilutions of AMC in 5% dimethyl sulfoxide in 50 mM sodium phosphate, pH 7.5.

Stability assays were carried out by incubating enzyme samples in centrifuge tubes sealed with liquid paraffin. Residual activity was measured by adding 50 μl of sample at each time point to a centrifuge tube, together with 200 μl of substrate, diluted in 50 mM sodium phosphate buffer, pH 7.5, to 500 μl. The tubes were incubated at 358 K for 60 min prior to recording fluorescence. In all assays, the fluorescence of a blank containing 300 μl of sodium phosphate buffer, pH 7.5, and 200 μl of substrate.

Results and discussion

Flexible docking

In this investigation, the molecular docking is used to indentify important residues in the active site of PhpI. AAFR-AMC is docked to the enzyme, and a flexible docking study is carried out to reveal the interaction between the different substrates and enzyme.

The PhpI-AAFR-AMC complex is generated with the InsigtII/Affinity module. AAFR-AMC is stabilized by hydrogen bond and hydrophobic interactions in the center of the active site. Seen from Fig. 1c, Gly70, Ile99 and Ala96 make a hydrogen bond with PhpI, respectively. Figure 1a shows proposed mechanism of PhpI. The nucleophilicity of Cys100 is enhanced by the catalytic base His101-Glu474. Gly70 and His101 function as oxynion hole, and the presence of two residues previously identified as conserved residues in DJ-1 superfamily sequences. In PhpI-AAFR-AMC complex, the side chain NH of the Gly70 forms a hydrogen bond to the oxygen of AAFR-AMC in a perfect position (2.05 Å). However, the distance between hydroxyl oxygen atom in AAFR-AMC and the NH group of backbone of His101 is 5.78 Å, and His101 can not form the hydrogen—bonding network to constitute a catalytic mechanism. The reason maybe lies in that His101 is near the catalytic residue Cys100, and the big substrate (AAFR-AMC) can not be close to it.

In order to find the key residues responsible for the binding of AAFR-AMC, the interaction energies of the substrate were calculated with eath of the residues in the active site. The interaction energies including the total, vdW, and electrostatic energies of key residues with the substrate were listed in Table 1. The PhpI- AAFR-AMC complex had a total interaction energy of -89.35 kcal·mol^{-1}, in which the electrostatic energy and vdW energies were -72.80 kcal·mol^{-1} and -16.55 kcal·mol^{-1}, respectively. From Table 1, it indicates that His101, Ala96, Glu474, Tyr120, Glu12, Gly70 and Cys100 are the important anchoring residues for

Table 1 The total energy E_{tota}, van-der-Waal energy E_{vdw} and electrostatic E_{ele} between AAFR-AMC and individual residues (E_{total}<-1.00kcal·mol^{-1} listed in energy rank order)

Residues	E_{vdw}(kcal ·mol^{-1})	E_{ele}(kcal· mol^{-1})	E_{total}(kcal ·mol^{-1})
His101	−8.35	−5.42	−13.77
Ala96	−7.01	−2.10	−9.11
Glu474	−1.86	−6.60	−8.46
Tyr120	−8.74	2.21	−6.53
Glu12	−3.41	−0.53	−3.95
Gly70	−3.15	−0.75	−3.90
Cys100	−0.15	−2.02	−2.17
Ile99	−1.22	0.09	−1.13

binding for they have strong van der Waals and electrostatic interactions with AAFR-AMC. Cys100, His101 and Glu474 function as catalytic triad. Gly70 is oxyion hole. Our theoretical results are in good agreement with the results by Du et al. [5]. Obviously, Tyr120 has strong van der Waals with AAFR-AMC, which forms P-π stacking interaction with the aromatic group of AAFR-AMC. Seen from Fig. 1, the sequence alignment and structure superposition show that PhpI shares characteristics with the members of the DJ-1 family belonging to four different kingdoms of life: 47% sequence identity with respect to the bacterial YhbO [6], 35% sequence identity with respect to the protein DR1199

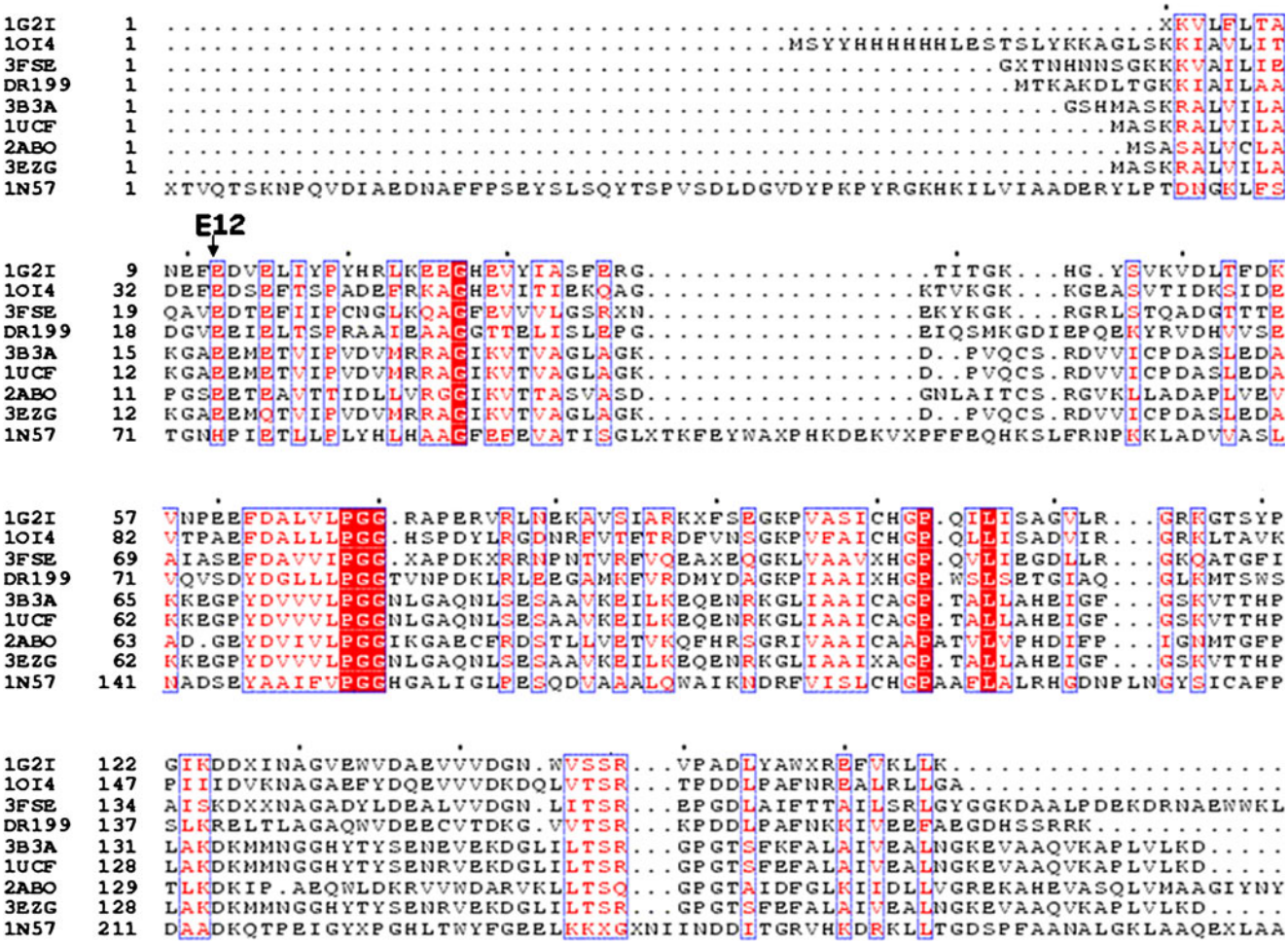

Fig. 2 Sequence alignment of PhpI homologs. Abbreviations are 1OI4, Yhbo from *Escherichia Coli* (47%); 3FSE, two-domain protein containing DJ-1/ThiJ/PfpI-like and ferritin-like domains (41%); DR199, the stress response protein DR1199 from *Deinococcus radiodurans* (35%); 3B3A, protein DJ-1 (25%); 1UCF, DJ-1, a protein related to male fertility and Parkinson's disease (24%); 2ABO, the YajL (ThiJ) protein from *Escherichia coli* (25%); 3EZG, protein DJ-1(23%); 1 N57, *Escherichia coli* Hsp31(19%)

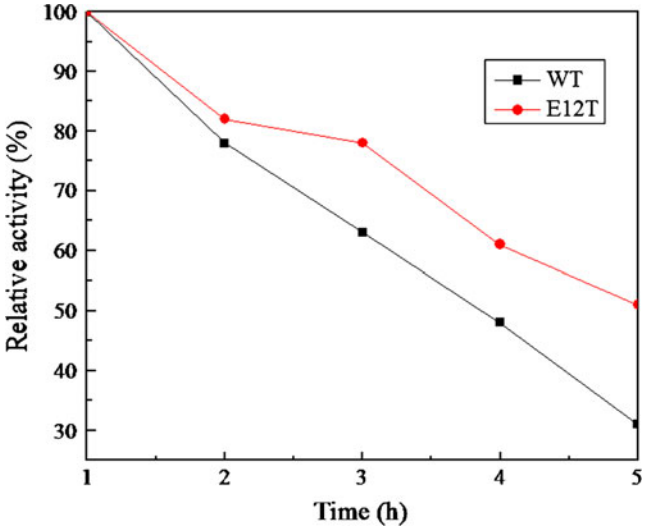

Fig. 3 Thermostability of WT and E12T

[13], 24% sequence identity with respect to human DJ-1 [8–12, 15–17] and 19% sequence identity with respect to the Hsp31 [1–4]. From the sequence alignment, we can see that Glu12 is highly conservative in DJ-1superfamily (Fig. 2) and this indicates Glu12 plays an important role in substrate binding. So we select Glu12 for further experiment.

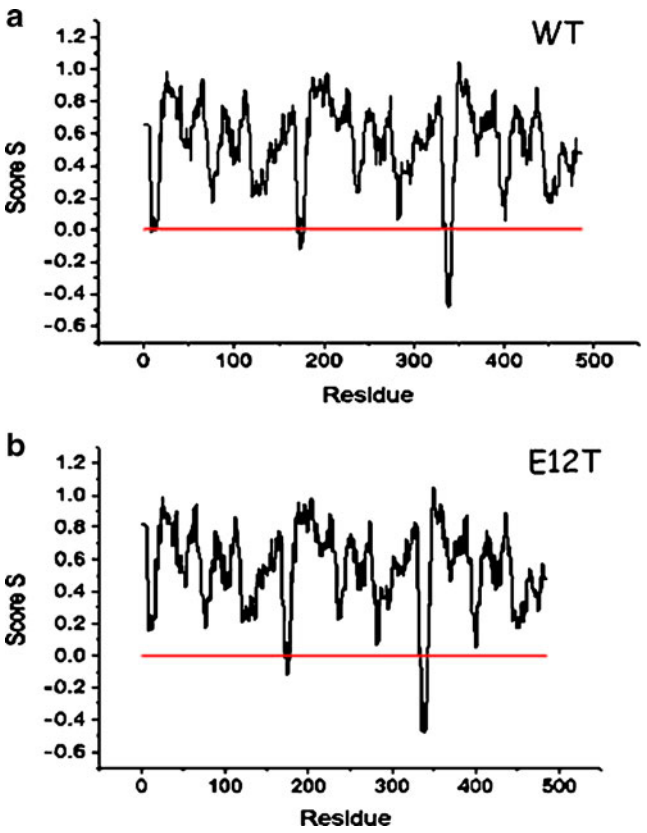

Fig. 4 The Profiles-3D score of PhpI and E12T mutant

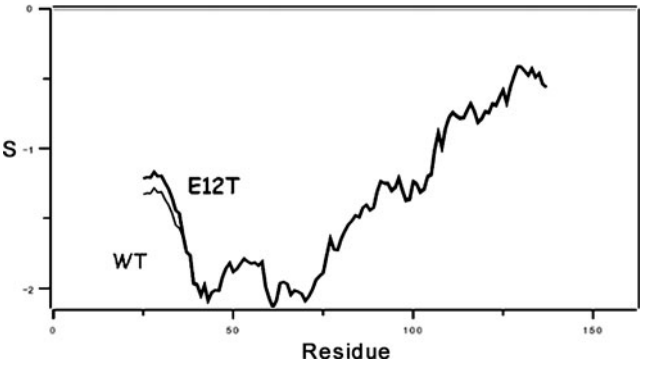

Fig. 5 Pair energy graph of PhpI (a chain) and E12T mutant (a chain). It shows that E12T mutant fold is better than that of PhpI's

Experiment results

According to the plasmid of WT, E12T mutant was successfully constructed. Through sonicating and heat-treating, the WT and E12T were purified into target proteins. And the protease activity of E12T was 3.8 fold higher than WT (data not shown). From the curve of thermostability (Fig. 3), we can see that the mutant was more stable than WT at 358 K.

Key residue Glu12 responsible for substrate binding

In order to explore the key residue Glu12 responsible for substrate binding, MD simulation is used. The 3D structure

Table 2 Profile-3D score for the wild type enzyme and the E12T mutant

Residue WT	Score	ResidueE12T mutant	Score
Lys2	0.649	Lys2	0.817
Val3	0.649	Val3	0.817
Leu4	0.649	Leu4	0.817
Phe5	0.649	Phe5	0.817
Leu6	0.649	Leu6	0.817
Thr7	0.649	Thr7	0.817
Ala8	0.354	Ala8	0.522
Asn9	0.310	Asn9	0.478
Glu10	−0.012	Glu10	0.156
Phe11	0.022	Phe11	0.190
Glu12	0.026	Thr12	0.194
Asp13	0.085	Asp13	0.253
Val14	−0.005	Val14	0.163
Glu15	−0.006	Glu15	0.162
Leu16	0.048	Leu16	0.216
Ile17	0.054	Ile17	0.222
Tyr18	0.297	Tyr18	0.297
Pro19	0.601	Pro19	0.601
Tyr20	0.536	Tyr20	0.536

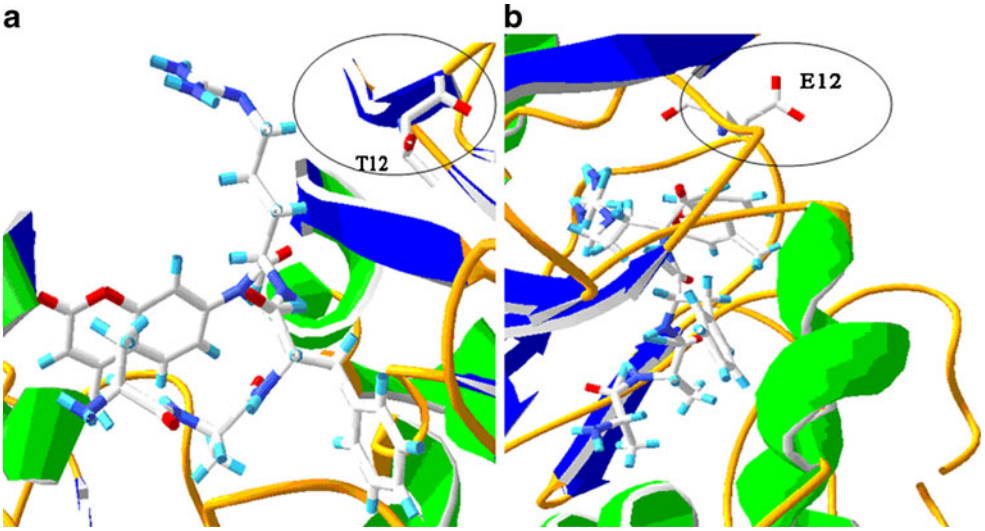

Fig. 6 a AAFR-AMC in the **a** E12T **b** WT

of the wild-type and E12T mutant are checked by Profiles-3D. The overall self-compatibility score for the E12T PhpI mutant is 245, which is higher than the wild-type's (240) (Fig. 4). Moreover, the analysis with ProSa2003 program was used to perform on the wild-type and the mutant shows that the mutant is more reasonable than the wild-type fold (Seen from Fig. 5). From Table 2, we can see that residues 2-17 in the E12T PhpI mutant become better than that of the wild-type enzymes'. By checking with two different criteria mentioned above, we believe that the E12T PhpI mutant is more reliable

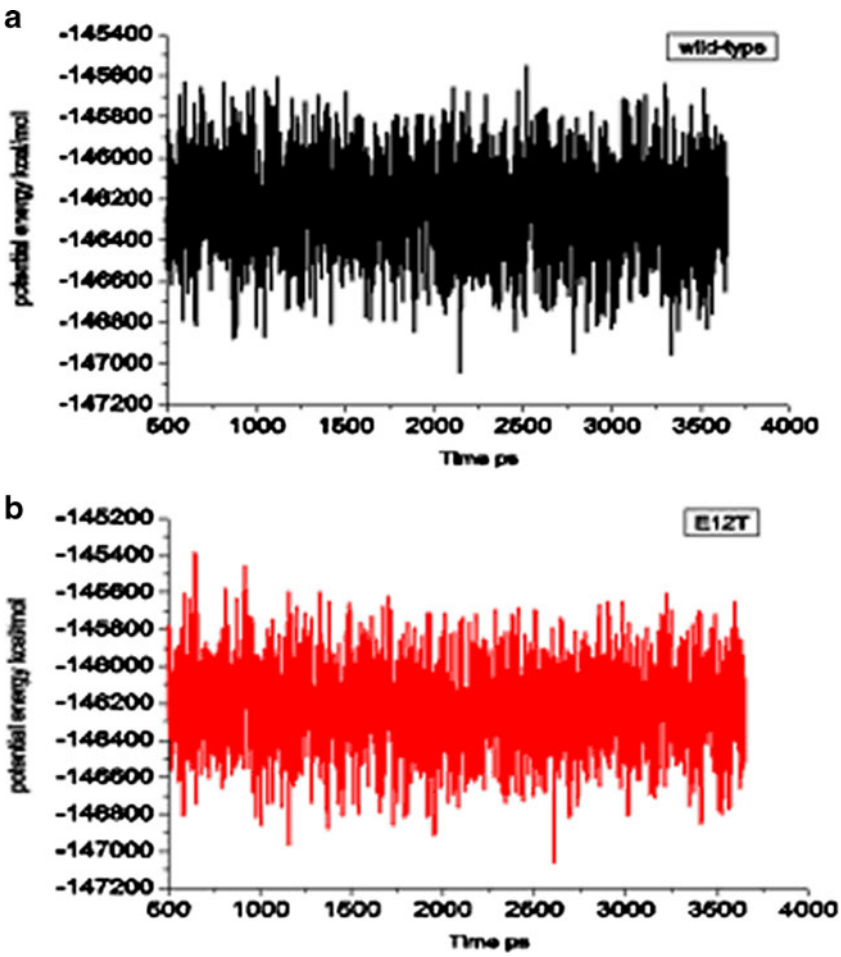

Fig. 7 **a** Potential energy (black) is obtained from the 500-4000 ps molecular dynamics trajectory for the PhpI. **b** Potential energy (red) is obtained from the 500-4000 ps molecular dynamics trajectory for the E12T mutant

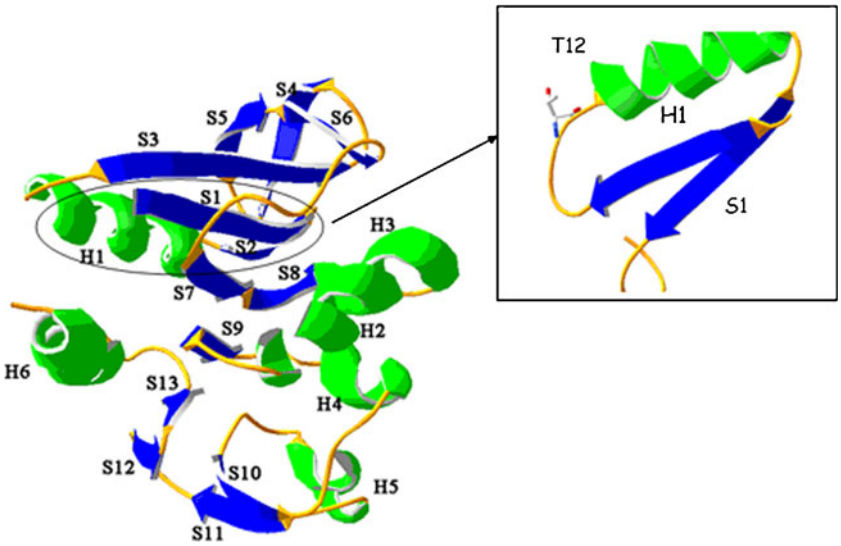

Fig. 8 Secondary structure analysis is based on the crystal structure of. PhpI and E12T mutant. "H" and "S" refer to helix and sheet, respectively

and stable than that of the wild-types'. The kinetic experiment by Zhan shows that the E12T mutant is more stable than that of the wild-type (Fig. 3). Our theoretical result as above is in harmony with kinetic experiment. We also docked AAFR-AMC to the wild-type, as well as E12T mutant (Fig. 6). From Fig. 6, we can see that in the wild-type, Glu12 locates in a sheet, whereas in the mutant, Thr12 locates in a loop.

In order to explain the conformational change, MD simulation is used. In this work, the docked complexes have been used as the initial structure in further simulations.

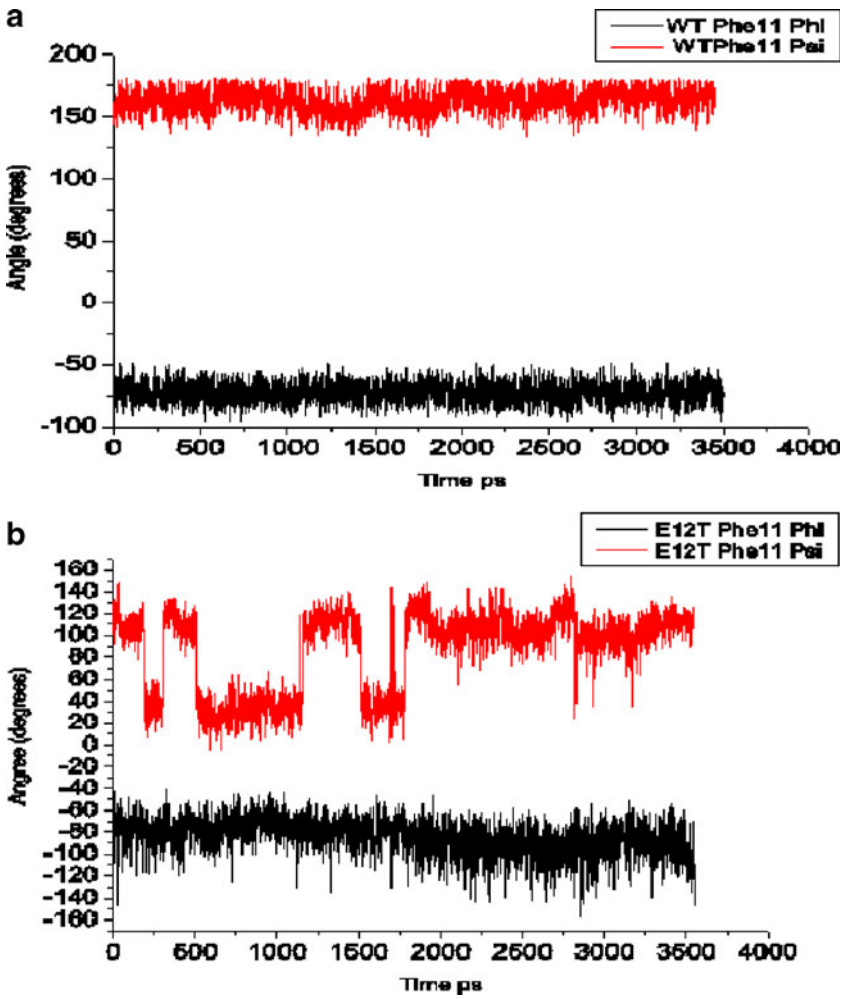

Fig. 9 Phi and psi dihedral angles of Phe11 in (**a**) PhpI, (**b**) E12T mutant during MD

The stability of the two complexes was refined by performing 3.5 ns MD simulations. Figure 7 displays potential energy of energy of the simulated system of the WT (Fig. 7a)-AAFR-AMC and E12T-AAFR-AMC mutant (Fig. 7b) during the 0.5-3.5 ns of molecular dynamics. As seen from Fig. 7, the potential energy remains constant after 500 ps simulated time of molecular dynamics of two complexes, which indicates that two complexes are structurally stable after 500 ps MD time. Of these, we can study the stable system during MD time. Figure 8 shows the comparison of the WT with the E12T mutant shows similar distributions of secondary structures. In the secondary structure of the wild-type protein, there are 13-β-sheets. Whereas the E12T PhpI mutant includes 13-β-sheets. Among these, sheet 2 (S2), for the wild-type, includes Glu10, Phe11, and Glu12. However, for E12T PhpI mutant, Glu10, Phe11, and Thr12 are not positioned in the sheet (Seen from Fig. 8). As a result, Phe11 in the mutant may be flexible. In order to investigate its flexibility, phi and psi dihedral angles are plotted against MD time, shown in Fig. 9a, in which the variations of phi and psi dihedral angles are placed within 60 for Phe11 in the wild-type protein. However, for the mutant, phi and psi dihedral angles position placed with 140 (Fig. 9b) shows variation. As a result, the difference between the same residues (Phe11) in the wild-type protein included in S2 and Phe11 in the mutant which are not included in a sheet can make a difference in stability of substrate binding, and this can result in different binding affinity for AAFR-AMC in wild-type and the mutant. This result is consistent with the point that proteins are conformationally dynamic and exhibit functional promiscuity [23, 24].

Distance deviations from the starting structure cannot necessarily reflect the mobility of structural elements. Some more comprehensive information on flexibility is achieved by comparing root mean square fluctuations (RMSF). Averaged RMSF may dump local difference in the mobility when comparisons are made between different conformations. To identify flexible regions in the molecule and to facilitate comparisons, we plot Ca-RMSF on per residue basis (Fig. 10). It is obvious that residue 1-50 (S1, S2, H1, S3, S4, S5, and S6) in E12T mutant-AAFR-AMC complex show distinctly large RMSF in all simulations, but there are variations among the trajectories. We can conclude that substrate binding does alter the structural flexibility of PhpI in E12T mutant. In fact, the fluctuations of S2 in E12T mutant increase substrate binding. Hence, high flexibility in this region appears to be an intrinsic property of the E12T mutant protein. The RMSF analysis was calculated. The results also reflect that the residues of the S1, S3, S4, S5 and H1 region are highly flexible and can visit relatively larger conformational space. We note that these regions contribute to a lot of critical residues for affinity and specificity.

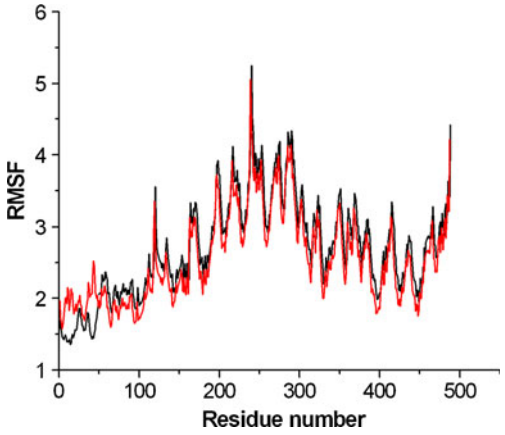

Fig. 10 Root mean square fluctuation (RMSF) of the Cα atoms around the average MD structure: blackline, protein in the wild-type protein-substrate complex; red line, protein in the E12T protein-substrate complex

Seen from the crystal structure, Glu15 forms a hydrogen bond with Glu12, and there is a hydrogen bond between Phe11 and Lys43. Tyr46 makes a hydrogen bond with residue Asp13 (Fig. 11). So their flexible nature may point to helping dynamics local interactions. Site-directed mutagenesis shows that the substitutions of Glu12 with Thr caused about 3.8 fold increases for AAFR-AMC. S2 becomes loop, which is very flexible in substrate binding, and this is helpful for catalytic power. Hydrogen bond may play an important role in catalytic efficiency. It can be concluded that Glu15, Lys43 and Tyr46 may also take part in the catalytic efficiency and thermostability of enzyme. This theoretical result may be helpful for the future site-directed mutagenesis.

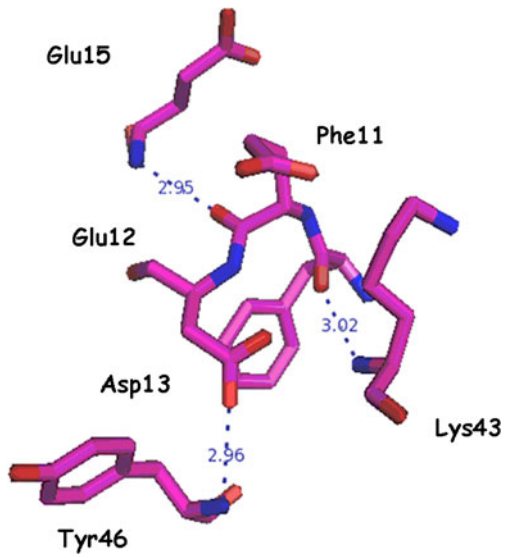

Fig. 11 Hydrogen bonds among residue Phe11, Glu12, Asp13, Glu15, Lys43, and Tyr46

Acknowledgments This work was supported by the National High Technology Program of China (863 Program, 2006A020203) and the basic research fund of Jilin University (No. 200810019). We thank Dr. Dave Case for his kindness in offering us the Amber9 program (amber.scripps.edu) as a freeware.

References

1. Malki A, Caldas T, Abdallah J, Kern R, Eckey V, Kim SJ, Cha SS, Mori H, Richarme G (2005) J Biol Chem 280:14420–14426
2. Mujacic M, Bader MW, Baneyx F (2004) Mol Micro 51:849–859
3. Quigley PM, Korotkov K, Baneyx F, Hol Wim GJ (2004) Protein Sci 13:269–272
4. Quigley PM, Korotkov K, Baneyx F, Hol Wim GJ (2003) Proc Natl Acad Sci USA 100:3137–3142
5. Du XL, Choi IG, Kim R, Wang W, Jancarik J, Yokota H, Kim SH (2000) Proc Natl Acad Sci USA 97:14079–14084
6. Abdallah J, Kern R, Malki A, Eckey, Richarme G (2006) Protein Expression Purif 47:455–460
7. Bonifati V, Rizzu P, Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MCJ, Squitieri F, Ibanez P, Joosse M, Dongen JW, Vanacore N, Swieten JC, Brice A, Meco G, Duijn CM, Oostra BA, Heutink P (2003) Science 299:256–259
8. Canet-Avilés RM, Wilson MA, Miller DW, Ahmad R, McLendon C, Bandyopadhyay S, Baptista MJ, Ringe D, Petsko GA, Cookson MR (2004) Proc Natl Acad Sci USA 101:9103–9108
9. Wilson MA, Collins JL, Hod Y, Ringe D, Petsko GA (2004) Proc Natl Acad Sci USA 100:9256–9261
10. Lee SJ, Kim SJ, Kim IK, Koo J, Jeong CS, Kim GH, Park C, Kang SO, Suh PG, Lee HS, Cha SS (2003) J Biol Chem 278:44552–44559
11. Lakshminarasimhan M, Maldonado MT, Zhou W, Fink AL, Wilson MA (2008) Biochemistry 47:1381–1392
12. Kazuya H, Nobuo NS, Masataka H, Takeshi N, Takahiro T, Hiroyoshi A, Fuyuhiko I (2003) J Biol Chem 278:31380–31384
13. Fioravanti E, Dura MA, Lascoux D, Micossi E, Franzetti B, McSweeney S (2008) Biochemistry 47:11581–11589
14. Bandyopadhyay S, Cookson MR (2004) BMC Evol Biol 4:1–9
15. Wilson MA, Amour CV, Collins JL, Ringe D, Petsko GA (2004) Proc Natl Acad Sci USA 101:1531–1536
16. Wei Y, Ringe D, Wilson MA, Ondrechen MJ (2007) PLoS Com Biol 3:0120–0126
17. da Costa C (2007) Curr Mol Med 7:650–657
18. Insight II, version 2000. San Diego, Accelrys Inc 2000
19. Affinity User Guide, Accelrys Inc, San Diego, USA 1999
20. Amber 9 users' manual (2007) University of California, San Francisco
21. Jorgensen MJ, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
22. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341
23. Bershein S, Tawfik DS (2008) Chem Biol 12151-158
24. Tkuriki N, Tawfik DS (2009) Science 324:203–207