ORIGINAL PAPER

Evaluation of active designs of cephalosporin C acylase by molecular dynamics simulation and molecular docking

Qing Li · Xiaoqiang Huang · Yushan Zhu

Abstract Optimization to identify the global minimum energy conformation sequence in in silico enzyme design is computationally non-deterministic polynomial-time (NP)-hard, with the search time growing exponentially as the number of design sites increases. This drawback forces the modeling of protein-ligand systems to adopt discrete amino acid rotamers and ligand conformers, as well as continuum solvent treatment of the environment; however, such compromises produce large numbers of false positives in sequence selection. In this report, cephalosporin acylase, which catalyzes the hydrolytic reaction of cephalosporin C to 7-aminocephalosporanic acid, was used to investigate the dynamic features of active-site-transition-state complex structures using molecular dynamics (MD) simulations to potentially eliminate false positives. The molecular docking between cephalosporin C and wild type acylase N176 and its eight mutants showed that the rate-limiting step in the hydrolytic reaction of cephalosporin C is the acylation process. MD simulations of the active-site-transition-state complex structures of the acylation processes for N176 and its eight mutants showed that the geometrical constraints between catalytic residues and small molecule transition states are always well maintained during the 20 ns simulation for mutants with higher activities, and more hydrogen bonds between binding residues and functional groups of the ligand side chain in the active pocket are formed for mutants with higher activities. The conformations of the ligand transition states were changed greatly after the simulation. This indicates that the hydrogen bond network between the ligand and protein could be improved to enhance the activity of cephalosporin C acylase in subsequent design.

Introduction

Green process development is becoming an increasingly central theme in the chemical manufacturing industry because of the associated negative impact this industry has on the environment. Here, enzymes play a critical role, and are being applied as industrial catalysts to develop environmentally benign processes that can replace high pollutant processes— the latter usually catalyzed by chemical catalysts in organic solvents [1]. However, the availability of suitable enzymes directly amenable to non-natural substrates is limited. One of the overwhelming advantages of enzymes compared with chemical catalysts is their exceptional selectivity, such as chemo-, region- and stereoselectivity. Unfortunately, an accompanying disadvantage that comes with these merits is the extreme difficulty of engineering subtle changes in the active site of an enzyme to accommodate different substrates [2]. To overcome such drawbacks, directed evolution as an empirical approach has been used with significant success during the last two decades [3]. An alternative to this technology is a structure-based computational enzyme design approach. This method can lead to the cost-effective, rapid design of a new enzyme with the aid of high-performance computing and advanced molecular modeling. Pauling proposed that enzymes work by stabilizing the transition state of the target reaction relative to its ground state [4]. Therefore the enzyme design concept can be taken as a special molecular recognition problem as we can identify just the transition state of the rate-

Q. Li · X. Huang · Y. Zhu (✉)
Department of Chemical Engineering, Tsinghua University,
Beijing 100084, People's .Republic of China
e-mail: yszhu@tsinghua.edu.cn

limiting step of the reaction and search for an amino acid sequence that folds spontaneously into the target structure, and that binds the transition state tightly and specifically. However, de novo design of the protein fold is not a routine approach. The three most striking in silico enzyme designs, i.e., artificial enzymes for Kemp elimination [5], the retro-aldol reaction [6] and the stereo-selective Diels-Alder reaction [7], were all created based on the selection of a suitable scaffold from a scaffold library, which could hold the theoretical active site described by the catalytic geometrical constraints between the catalytic residues and the transition state of the target reaction.

The success of automatic design of artificial enzymes has inspired a body of research on why they work and how they can be improved, because the highest catalytic efficiency of the de novo designed enzymes is still well below those of natural enzymes. Here, we take the Kemp eliminase design as an example. Alexandrova et al. [8] used mixed quantum and molecular mechanics (QM/MM) tools to study four active designs for Kemp elimination [5], i.e., KE07/1thf, KE10/1a53, KE15/1thf and KE16/1thf. This study failed to obtain a correlation between computed activation barriers and the activation barriers derived from experiments, and the trend of their calculated barriers was found to be opposite to the values derived from experiments. Frushicheva et al. [9] used an empirical valence bond method to correctly reproduce the overall catalytic effect of the designed Kemp eliminases. Based on their calculation of the activation barrier they could suggest mutations that reduced overall catalysis, but they encountered difficulties in improving the activities of Kemp eliminases. This observation clearly illustrates that it is extraordinarily challenging to create an accurate preorganized active site environment that discriminates between the transition and ground states effectively. Kiss et al. [10] concluded that full enzyme QM/MM calculations are inadequate to differentiate between active and inactive designs, but the implementation of molecular dynamics (MD) simulations may be able to eliminate effectively the false-positives. Privett et al. [11] combined the computational protein design method with MD simulations to develop an iterative approach, which led to the development of the most catalytically efficient computationally designed enzyme, i.e., HG3/1gor, for Kemp elimination. Starting from HG3/1gor, Blomberg et al. [12] evolved an artificial enzyme, i.e., HG3.17, that accelerated the Kemp elimination 6×10^8 fold, and its catalytic efficiency k_{cat}/K_m reached 2.3×10^5, which is comparable with that of highly optimized natural enzymes such as triosephosphate isomerase. The crystal structure of the HG3.17-inhibitor (6-nitrobenzotriazole) complex showed that the evolved Kemp eliminase resembles the idealized active site targeted by design; however, a computational approach alone did not find such a design.

In all available computational enzyme design programs, searching of sequence space remains a bottleneck for current combinatorial optimization algorithms [13]. Therefore, discrete amino acid side-chain rotamers and ligand placement schemes are always used, which restricts the side-chain functional groups that can adopt the high-precision conformations necessary for high catalytic efficiency. In addition, the protein scaffold is always assumed to be rigid in order to mitigate computational complexity. An additional source of inaccuracy for enzyme design comes from the simplified electrostatic interaction models, which struggle to handle correctly the long-range effects of electrostatics. All these shortcomings can be accounted for by MD simulations where explicit water molecule models are applied. In this research, the wild-type and eight mutants of a cephalosporin C acylase, i.e., N176, which catalyzes the hydrolytic reaction of cephalosporin C to 7-aminocephalosporanic acid (7-ACA), were used to investigate whether or not the dynamic features identified by the MD simulations could evaluate active designs with different activities, in order to facilitate the discovery of further mutations that lead to higher catalytic efficiency.

Materials and methods

Calculation of the active-site-transition-state complex

The hydrolytic reaction of cephalosporin C (CPC) shown in Fig. 1 is catalyzed by cephalosporin acylase (CA), which converts CPC into 7-amino cephalosporanic acid (7-ACA) and D-α-aminoadipate under very low activity as the primary substrate of the cephalosporin acylase from *Pseudomonas* sp. N176 is glutaryl-7-aminocephalosporanic acid (GL-7-ACA). The structure of the enzyme active-site-transition-state complex was obtained by the PROtein Design Algorithmic (PRODA) [14–16] package, where the three-dimensional (3D) scaffold structure of cephalosporin acylase was taken (PDB ID: 4HSR) without further minimization. The structure of CPC was modeled based on the crystal structure of GL-7-ACA (PDB ID: 1JVZ) [17], and its transition state structure was generated according to the placing rules shown in Supporting Information Table S1 based on the tetrahedral intermediate assumption. The catalytic geometrical relationships between catalytic residues and transition state small molecules are presented in Fig. 2, where the four catalytic residues are SerB1, HisB23, HisB70 and AsnB242. CA belongs to the N-terminal nucleophile amidohydrolase superfamily and shares the same catalytic mechanism as penicillin G acylase [18], in which the hydroxyl group of the N-terminal serine of the B-chain, i.e., SerB1, is the nucleophilic group. The design scheme for scaffold 4HSR is presented in Supporting Information Table S2. The total number of rotamers for all design sites was 97,942, where the number

Fig. 1 Reaction scheme catalyzed by N176. *CPC* Cephalosporin C, *CA* cephalosporin acylase, *7-ACA* 7-aminocephalosporanic acid
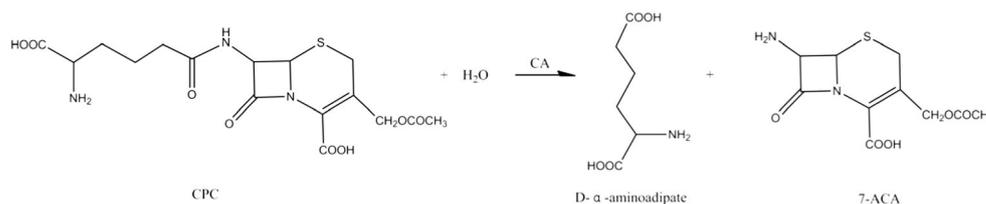
of small molecule rotamer library was 2,110, and the computational complexity reached 6.27×10^{80} for optimal rotamer selection. The active-site-transition-state complexes were calculated by PRODA for the wild-type and eight mutants, i.e., M1(HisB57Asn), M2(HisB57Phe), M3(HisB57Ser), M4(HisB57Thr), M5(HisB70Ser), M6(HisB70Tyr), M7(HisB57Ser, HisB70Ser) and M8(HisB57Ser, HisB70Ser, MetA165Ser). The experimental kinetic parameters, i.e., V_{max} and K_M, for wild-type and all eight mutants are given by Pollegioni et al. [19] and Golden et al. [20]. The eight mutants were selected because their mutation sites are in direct interaction range with the substrate and the activities of these mutants are different from that of wild type N176, the specific positions of the mutation sites are shown in supporting information Fig. S5.

Molecular dynamics simulation of active-site-transition-state complex

The dynamic behaviors of the active-site-transition-state complex for N176 and its eight mutants were investigated by MD simulations, which were implemented by running GROMACS 4.5.5 [21] using the GROMOS 96 force field on a computer-cluster with 208 cores. The topologies of N176

and its eight mutants were generated by GROMACS automatically. While the topologies of the small molecule, i.e., CPC, were prepared using the online PRODRG sever [22], and the atom types and bond parameters of CPC were built to be consistent with the definitions of the GROMOS 96 force field. However, the settings of the charge groups and group charges of CPC generated by PRODRG are different from those of the same functional groups present in the GROMOS96 43A1 parameter library. To overcome this discrepancy, the charge groups and group charges of CPC were modified to be consistent with those in the GROMOS 96 parameter library, which are shown in Fig. S1 [23]. The calculated protein-TS complex structure was immersed in a dodecahedral box whose size was determined by setting the distance between the solute and box to be 10 Å, and the box was filled by the addition of up to 23,000 explicit water molecules represented by a simple point charge model. Subsequently, the whole system was neutralized by explicit counter ions.

The MD simulation process was initiated by a systematic energy minimization (EM) for all atoms in the system. A solvent equilibrium process was then run, which was composed of seven rounds of position-restrained (PR) simulations prior to the system being subjected to MD simulations under constant pressure and temperature periodic-boundary conditions. Specifically, the EM was conducted using the steepest descent minimization algorithm to reduce excessively large forces between atoms to adjust the randomly generated starting configurations of water molecules and counter ions. The PR simulations were run in a six-step (NVT1–NVT6) heating process at 50, 100, 150, 200, 250 and 300 K, under constant volume periodic boundary conditions, and each NVT step was conducted for 50 ps with a time-step of 1 fs. This was followed by an equilibrium process (named NPT step) at constant temperature 300 K and constant pressure 1 bar for 2 ns with a time-step of 2 fs, to avoid drastic rearrangements of the protein-TS complex structure caused by not fully equilibrated solvent [10], where heavy atoms of the protein-TS complex were harmonically restrained to their starting positions. The temperature was controlled using a modified Berendsen thermostat with velocity rescaling (v-rescale) [24], and the pressure was maintained by Parrinello-Rahman coupling [25] in the NPT step. In both PR and MD simulations, the LINCS algorithm [26] was used to impose constraints on bonds and angles of the protein-TS complex, and
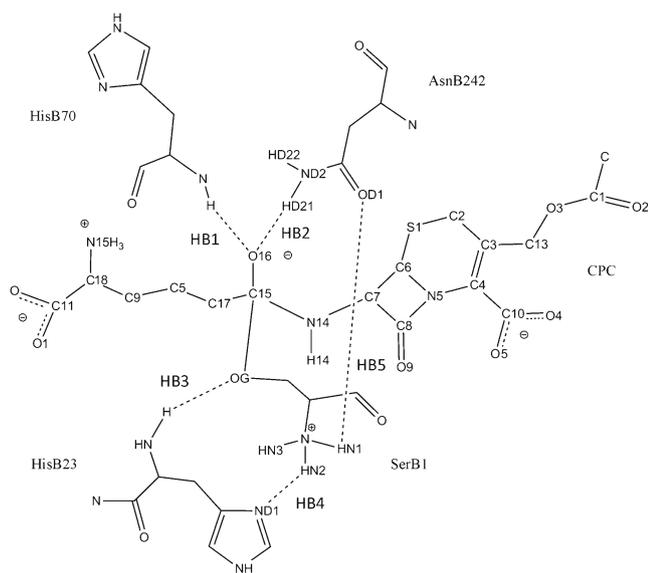

Fig. 2 Catalytic geometrical constraints for reaction catalyzed by CA. *Dashed lines* Five catalytical hydrogen bonds (HB1–HB5)

the Particle-Mesh-Ewald [27, 28] method was used to model long-range electrostatic effects, while the differential equations of motion were integrated by a leap-frog algorithm [29]. A 20 ns MD simulation was run for each system. In the MD simulations, geometries and velocities were saved every 1,000 steps (1 ps, step size = 1 fs), which generated a total of 20,000 frames for each run. Post-simulation data extraction and analysis were performed using the GROMACS analyzing tools, and DiscoveryStudio2.1 was used to observe geometries visually.

Molecular docking between enzyme and substrate

The molecular docking process of substrate CPC into the active site of CA was implemented using the LigandFit protocol in DiscoveryStudio2.1, which is a fast and accurate docking tool based on shape comparison between ligand and the active site. All candidate site regions for docking were found by a cavity detection algorithm in LigandFit and the site region in the vicinity of the catalytic residues, i.e., SerB1, HisB23, HisB70 and AsnB242, and the key binding residues, i.e., ArgB24, TyrB32, HisB57 and HisB178, were selected as the initial site for docking. This site was further edited by contracting and expanding using the tools of LigandFit and a final manual modification. In the docking process the Dreiding force field was used to calculate the energy grid where the Gasteiger charging method was employed for determining charges on the protein and the ligand. Diverse CPC conformations were randomly generated using a Monte Carlo algorithm based on the flexible torsion angles of CPC. After initial alignment of the ligand in the active site the rigid body pose was minimized by running up to 1,000 SD iterations and BFGS iterations until the energy of the conformation reached a local minimum. Several rounds of docking were always run for each mutant of CA and each round generated 20 poses. Before the poses were saved clustering analysis was performed with a 1.0-Å RMS threshold. After docking the pose was further minimized using Smart Minimizer as the optimization tool with 0.001 gradient tolerance and CHARMm as the force field. Finally, the produced poses were filtered using three geometric criteria: (1) the distance between the OG atom of SerB1 and the C15 atom of CPC should be no more than 3.1 Å, which is the sum of van der Waals radius of atom O (1.4 Å) and that of atom C (1.7 Å) [30]. (2) The side chain of CPC should lie in the interior, which means that the distances between the side chain of CPC and the key binding residues, i.e., ArgB24, TyrB32, HisB57 and HisB178, should be less than that between the cepham ring and those residues. (3) The distance between the O16 atom of CPC and the N atom of HisB70 should be less than 4.2 Å, which is the upper limit of the distance between donor and acceptor for a weak hydrogen bond [31]. The filtered poses were evaluated using consensus scoring based on seven scoring functions, i.e., LigScore1,

LigScore2, PLP1, PLP2, Jain, PMF, and DOCK SCORE, and the favorable pose was identified as the one with the top consensus score among all filtered poses. The binding energy between CPC and the enzyme active site was scored based on the output pose using LigScore2 [32], which is an empirical scoring function that estimates the free energy of binding. The LigScore2 function contains three contributions that describe the van der Waals interaction, the electrostatic attraction between ligand and protein, and the desolvation penalty attributed to the binding of polar ligand atoms to the protein and vice versa. The scored binding energy was then converted into the dissociation constant (K_s) of the enzyme–substrate complex.

Results and discussion

Molecular docking results between enzyme N176 and substrate CPC

The enzyme–substrate complex structures between CPC and N176, as well as its eight mutants were obtained by LigandFit, and the calculated K_s values between the enzyme and substrate based on the binding energy, which were scored by LigScore2, and the experimentally measured kinetic parameter K_m are shown in Fig. 3. The specific data are presented in the Supporting Information Table S3. Figure 3 shows that most of the calculated K_s values are consistent with the kinetic
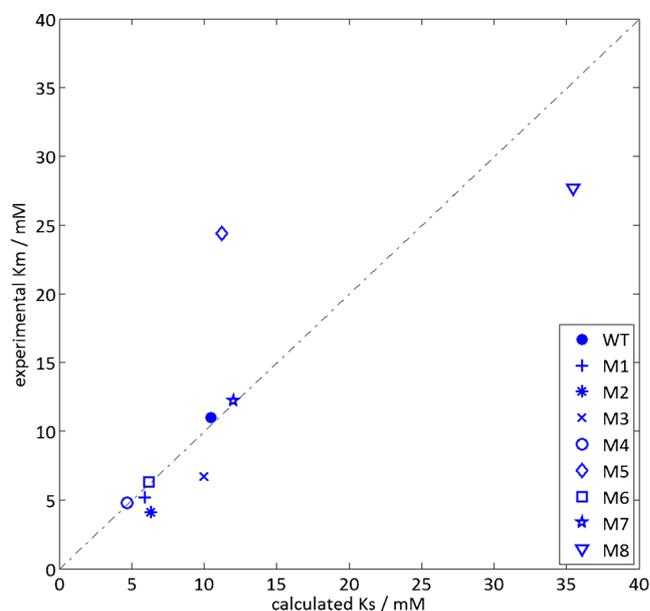

Fig. 3 Plot of experimental K_m against calculated K_s for wild type N176 and its eight mutants with CPC as the substrate. WT(wild type N176), M1(HisB57Asn), M2(HisB57Phe), M3(HisB57Ser), M4(HisB57Thr), M5(HisB70Ser), M6(HisB70Tyr), M7(HisB57Ser, HisB70Ser), and M8(HisB57Ser, HisB70Ser, MetA165Ser)

parameter K_m. According to the kinetic mechanism of the serine protease reaction [33], the acylation process is always the rate-limiting step for amide hydrolysis when compared with the deacylation process, this implies $k_3 \gg k_2$, as that shown in Fig. 12 of Hedstrom [33]. The CA from N176 and its mutants are poor catalysts of the CPC hydrolytic reaction. This indicates that the kinetics of CPC hydrolysis is reaction-controlled instead of diffusion-controlled, which implies $k_{-1} \gg k_2$. Therefore, the Michaelis-Menten parameter K_m can be approximated using the dissociation constant K_s between CPC and N176 or its mutants, respectively, and this approximation further implies that the kinetics of N176 and its mutants abide by the catalytic mechanism of serine proteases and that the acylation process is the rate-limiting step for the CPC hydrolytic reaction. This conclusion showed that decreasing activation energy of the acylation process significantly promotes the catalytic efficiency of CPC acylase. It should be noted that the calculated K_s of M5 deviated from the experimental K_m greatly, this was led by the defects of the scoring function LigScore for the binding energy between the protein and ligand. The binding pose between CPC and active site of M5 is shown in Fig. S2, and this figure implies that no hydrogen bond is formed between atom ND1 of HisB57 and atom O1 of CPC since the donor-hydrogen-acceptor angle is acute. But the distance between donor and acceptor is 2.82 Å; therefore, this hydrogen bond was mistakenly included in the scoring function due to the surface descriptor calculation method used by LigandFit. Moreover, the positive amino group of CPC is oriented towards the positive guanidine group of ArgB24 and the distance between N15 atom of CPC and the NH1 atom of ArgB24 is only 4.39 Å. Such an arrangement would produce strong long-range electrostatic repulsions physically. But, the long-range electrostatic interaction cannot be considered in the scoring function of LigandFit due to the surface descriptor calculation method. As a result, these two miscalculations led to a lower calculated K_s value for M5.

Structure validation of the MD simulations

The specific MD protocols introduced in Materials and methods , such as the choice of force field, settings of the force field parameters and the water molecule models, were tested by the structural differences before and after simulations. The MD simulations were carried out on the crystal structures of N176 (PDB ID: 4HSR) in the absence of the ligand. The overlay of the N176 structure after a 20 ns MD simulation and that of the crystal is presented in Fig. 4, and the geometries of the active site residues were maintained throughout the 20 ns MD simulation. The backbone geometries show a root-mean-standard deviation (RMSD) from those of the crystal structure of 1.16 Å, whereas the side chain RMSD was 2.03 Å. It should be noted that SerB1 is the N-terminal residue of the B-Chain and AsnB242 lies on a loop,

and that these two residues deviate largely from their crystal positions shown in Fig. 4. The RMSD of the backbone geometries of the active site regardless of these two flexible residues reduces greatly, and is ≈ 0.57 Å; however, the RMSD of the side chain geometries does not change very much. Therefore, we can say that the MD simulations can recover the crystal structure of N176 and will behave equally well on the active-site-TS complex structure calculated by PRODA. Moreover, prolonged, i.e., 40 ns, MD simulations were run for the active-site-TS complex structures of wild-type N176 and its double mutant M7; the RMSD figures of main chains and side chains are shown in supporting information Fig. S3. These figures imply that the simulation is stable at 20 ns.

MD simulation results for hydrogen bonds within the catalytic sites

Transition-state theory states that enzymes accelerate reactions because they reduce the activation free energies of the corresponding reactions by stabilizing the transition state of the reacting system. The transition state or the tetrahedral intermediate of CPC at the active site of CA is stabilized by a covalent bond formed between the OG atom of SerB1 and the C15 atom of CPC, as well as five hydrogen bonds. The backbone N atom of HisB70 and the side chain atom ND2 of AsnB242 comprise the oxyanion hole of the partial positive charge that activates the carbonyl of the amide and stabilizes the negatively charged O16 atom of CPC; two hydrogen bonds hold the O16 atom of CPC between these two nitrogen atoms. The third hydrogen bond formed between the N atom
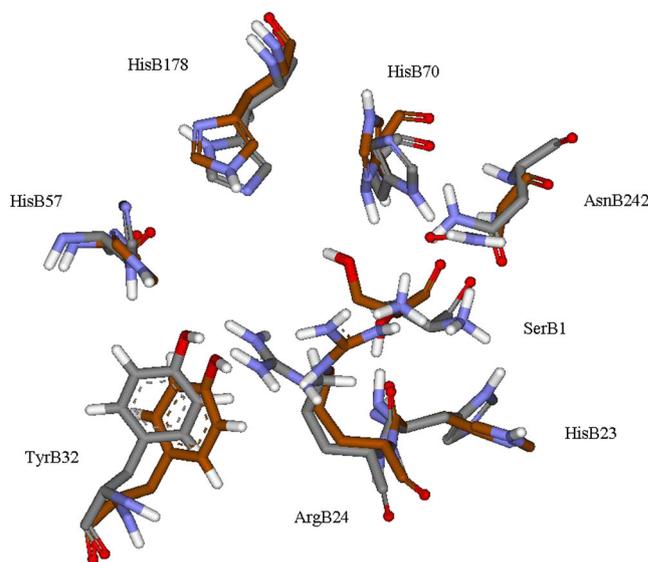

Fig. 4 Overlay of representative molecular dynamics (MD) geometry and crystal structure for wild type N176. *Grey* Carbon atoms in crystal structure, *brown* carbon atoms in MD geometry. Atoms: *Red* Oxygen, *cyan* nitrogen, *white* hydrogen

of HisB23 and the OG atom of SerB1 is maintained during the reaction path, which helps in accurately positioning the latter atom for nucleophilic attack. The other two hydrogen bonds are formed between the N atom of SerB1 and the ND1 atom of HisB23 and the OD1 atom of AsnB242. These two hydrogen bonds facilitate the terminal amino group to abstract the proton from the hydroxyl group of SerB1 and enhance its ability for nucleophilic attack. The category criteria for calibrating the hydrogen bonds quantitatively were adopted from [31], and the specific parameters are presented in Fig. 5. Table 1 shows that at least three out of the five hydrogen bonds at catalytic sites are maintained through the 20 ns MD simulation, although some hydrogen bonds lie in the "weak" categories. These results are consistent with the conclusion given by Kiss et al. [10] that MD simulations are robust enough to distinguish between active and inactive designs, because all eight CA mutants and the wild-type show appreciable activities towards CPC hydrolysis. Although we cannot discriminate mutants with higher activities from those with lower activities just from the number of hydrogen bonds maintained through 20 ns MD simulation, mutants with higher activities always have more hydrogen bonds, which lie in the categories of "strong" and "moderate". Figure 5 shows the angle versus distance scatter plots of five hydrogen bonds at catalytic sites for the wild-type CA and M7(HisB57Ser, HisB70Ser) mutant, and the latter shows the highest catalytic efficiency among all mutants, which is four times higher than that of the wild-type. Figure 5a–e states that only a hydrogen bond between ND2 atom of AsnB242 and O16 atom of CPC lies in the moderate category of bond strength, and its densely populated clusters of data points peak at 1.64 Å and 169.8°. Meanwhile, the other four hydrogen bonds either lie mainly in the weak categories of hydrogen bond strength, i.e., for hydrogen bonds between the N atom of HisB70 and the O16 atom of CPC, the N atom of HisB23 and the OG atom of SerB1, the OD1 atom of AsnB242 and the N atom of SerB1, respectively, or they have been lost completely during the simulation time, i.e., for the hydrogen bond between the ND1 atom of HisB23 and the N atom of SerB1. For the mutant M7(HisB57Ser, HisB70Ser) shown in Fig. 5f–j, all hydrogen bonds lie in the moderate categories of hydrogen bonding strength and their densely populated data points peak at favorable distances and nearly linear angles, except for the one between the ND1 atom of HisB23 and the N atom of SerB1, which was partly lost during the simulation time. The two mutations, i.e., HisB57Ser and HisB70Ser, of M7 are both substitutions of large polar amino acids with smaller polar amino acids; therefore the active pocket of N176 is expanded to accommodate the larger substrate, i.e., CPC compared with its natural substrate GL-7-ACA. The well maintained hydrogen bonds formed between the transition state of CPC and the active site of M7 stabilize the tetrahedral intermediate and reduce the activation free energies of the CPC hydrolytic

reaction. In fact, two single mutation mutants, i.e., M3(HisB57Ser) and M5(HisB70Ser), both have promoted catalytic efficiencies compared with that of wild-type N176. According to the data of V_{max} and K_m for wild-type, M3, M5 and M7 presented by Pollegioni et al. [19], we found that the single mutation HisB70Ser increased V_{max} from 0.7 to 2.2, but also increased K_m from 11 to 24.4. These changes imply that the single mutation HisB70Ser facilitated transition state stabilization, but jeopardized the binding between the CPC substrate and the active site. In contrast, the single mutation HisB57Ser promoted its catalytic efficiency by improving binding between the CPC substrate and the active site, while not destroying the binding between the CPC transition state and the active site. It should be noted that the double mutation, i.e., HisB57Ser and HisB70Ser, is synergistic, because the kinetic parameter V_{max} was promoted more than the V_{max} values determined by both single mutations and the K_m was just increased to less than the average of the K_m of both single mutations.

MD simulation results for hydrogen bonds at the binding sites

The side chain of CPC is polar and buried inside the active pocket of N176, therefore an effective hydrogen bonding network should be formed between CPC and the polar or charged residues at binding sites in order to compensate for the desolvation free energy penalty for burying the polar groups. Such requirements have been confirmed by our former statistical analysis for a series of complex structures of enzymes or receptors with their ligands [34, 35]. The polar binding residues in the vicinity of functional groups of the CPC side chain are ArgB24, TyrB32, HisB57 and HisB178, and up to ten hydrogen bonds could be formed between the functional groups of these polar residues and the carboxyl and amino groups of CPC. Therefore, the average number of hydrogen bonds formed between them during the MD simulation time is a reliable measure with which to judge the binding between the CPC transition state and the active site of CA, and the hydrogen bonding numbers for wild type N176 and its eight mutants with CPC are presented in Table 2. Moreover, the relationship between the average number of hydrogen bonds of each mutant and its catalytic efficiency is illustrated in Fig. 6, which is plotted as a quadrantal

Fig 5a–j Angle versus distance scatter plots for hydrogen bonds at ▶ catalytic sites. **a**–**e** Hydrogen bonds in wild type N176, **f**–**j** hydrogen bonds in mutant M7(HisB57Ser, HisB70Ser). Data points are from 20 ns MD simulation, and the angle and distance distributions are projected onto the axes. The hydrogen bond categories and the explicit description of each hydrogen bond are outlined in the *inset* of each figure. The *green points* show the distance and angle of hydrogen bonds designed by PRODA
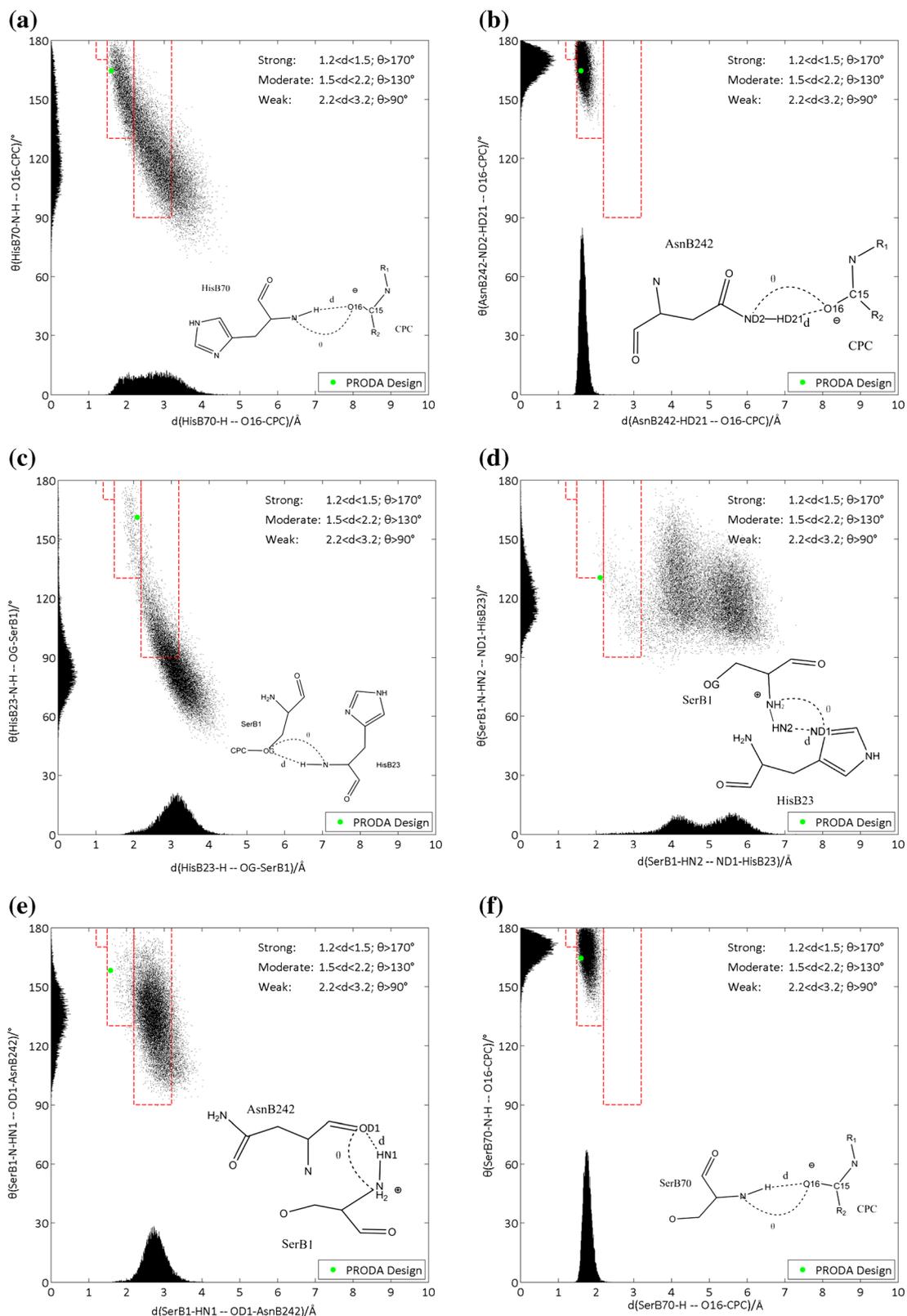
(a) Strong: 1.2<d<1.5; θ>170°
Moderate: 1.5<d<2.2; θ>130°
Weak: 2.2<d<3.2; θ>90°

(b) Strong: 1.2<d<1.5; θ>170°
Moderate: 1.5<d<2.2; θ>130°
Weak: 2.2<d<3.2; θ>90°

(c) Strong: 1.2<d<1.5; θ>170°
Moderate: 1.5<d<2.2; θ>130°
Weak: 2.2<d<3.2; θ>90°

(d) Strong: 1.2<d<1.5; θ>170°
Moderate: 1.5<d<2.2; θ>130°
Weak: 2.2<d<3.2; θ>90°

(e) Strong: 1.2<d<1.5; θ>170°
Moderate: 1.5<d<2.2; θ>130°
Weak: 2.2<d<3.2; θ>90°

(f) Strong: 1.2<d<1.5; θ>170°
Moderate: 1.5<d<2.2; θ>130°
Weak: 2.2<d<3.2; θ>90°

diagram with the wild-type as the origin. Figure 6 shows that mutants M7, M8 and M5, which lie in the first quadrant and have higher catalytic efficiencies than that

of the wild type N176, formed more hydrogen bonds than those of other mutants Moreover, mutants M6 and M4 lying in the third quadrant show less catalytic efficiencies
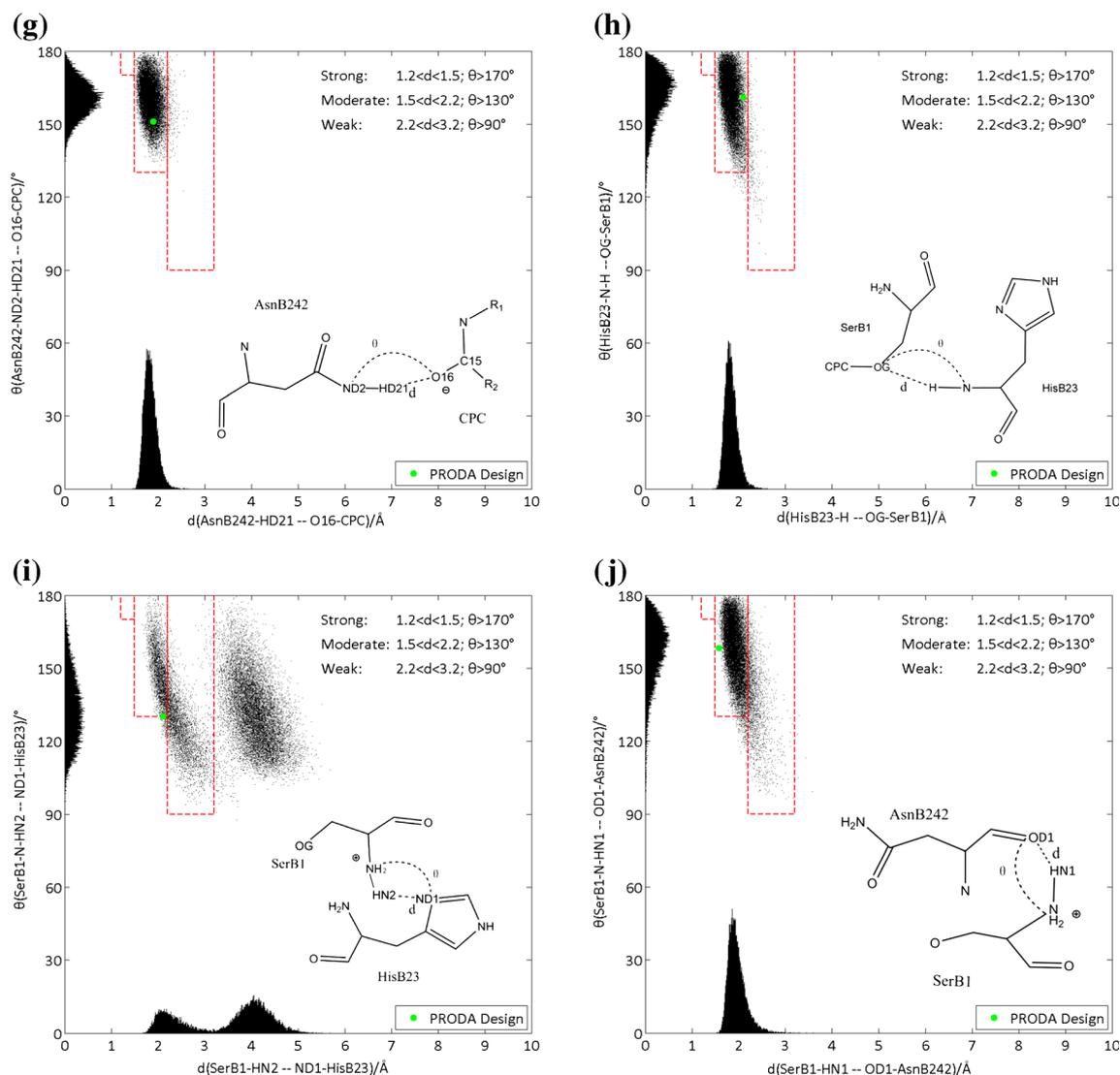
(g)

(h)

(i)

(j)

Fig. 5 (continued)

and fewer hydrogen bonds when compared with that of the wild type protein. A representative snapshot of the MD frames for wild type N176 is shown in Fig. 7a, where only three hydrogen bonds were formed between the polar functional groups of CPC and the neighboring binding residues. Three hydrogen bonds are formed between the imidazole groups of HisB57 and HisB178 with the carboxyl group of the CPC side chain, and no hydrogen bonds were observed for the guanidine group of ArgB24, the hydroxyl group of TyrB32 and the amino group of the CPC side chain. These polar groups, which are buried inside the active pocket of N176, contribute to a significant desolvation free energy penalty for the active-site-CPC-TS complex and result in higher activation energy for the target reaction. Such a poor hydrogen bonding network comes from the larger volume of CPC than that of GL-7-ACA, which is the natural substrate of N176.

Therefore, mutations that save space to accommodate CPC and bridge enough hydrogen bonds with the carboxyl and amino groups of CPC will promote the catalytic efficiency of CA towards CPC. Such mutations are embodied in mutant M7, where the binding residue HisB57 was mutated to Ser, and the hydrogen bonding network between polar residues of M7 and the CPC side chain is reflected in the representative geometry of the MD frames shown in Fig. 7b. Five hydrogen bonds were formed between the carboxyl group of the CPC side chain and the polar groups of the binding residues, i.e., the imidazole group of HisB178, the hydroxyl groups of TyrB32 and SerB57 and the guanidine group of ArgB24. Only the amino group of the CPC side chain does not form any hydrogen bonds, which presents opportunities for further improvement of the catalytic efficiency of N176-based mutants towards CPC hydrolysis.

Table 1 Peak values of hydrogen bond distances and angles in 20 ns MD simulations for wild type N176 and its eight mutants

Mutants[a]	HB1[b]		HB2		HB3		HB4		HB5		V_{max}/K_m
	d[c]	θ[d]	d	θ	d	θ	d	θ	d	θ	
WT	3.06	113.9	1.64	169.8	3.17	82.7	5.52	119.8	2.73	136.1	0.06
M1	1.7	172.6	1.66	168.2	2.04	159.5	2.51	116.4	1.94	140.6	0.05
M2	1.75	170.4	1.68	167.6	2.75	114.5	4.4	146	1.93	142.9	0.01
M3	1.77	168	1.73	165	2.4	145.8	2.87	116.6	2.58	104.5	0.1
M4	1.79	166.6	1.69	162.8	2.77	122.6	4.16	147.8	2	140.9	0.04
M5	2.51	68.4	2.77	120.2	3.07	88.9	1.95	158.3	2.56	109.9	0.09
M6	4.42	138.5	1.66	167.6	2.79	95.4	4.25	137.6	2.9	116.6	0.04
M7	1.76	172.1	1.76	163.2	1.78	166.9	4.05	128	1.87	161.2	0.24
M8	1.82	166.6	1.74	162.9	1.84	162	2.16	111.1	1.88	151.1	0.15

[a] WT for wild type N176, and M1(HisB57Asn), M2(HisB57Phe), M3(HisB57Ser), M4(HisB57Thr), M5(HisB70Ser), M6(HisB70Tyr), M7(HisB57Ser, HisB70Ser), and M8(HisB57Ser, HisB70Ser, MetA165Ser)

[b] The descriptions of hydrogen bonds from HB1 to HB5 are shown in Fig. 2

[c] Distance of the hydrogen bond (Ångstroms)

[d] Angle of the hydrogen bond (degrees)

MD simulation results for structural integrity of the active site

The RMSD values of the backbone and side chain geometries of the active sites after the 20 ns MD simulations and those in the PRODA design for wild type N176 and its eight mutants are shown in Table 2, where the backbone and side chain geometries always show RMSD values of ≈ 1.0 and 2.0 Å, respectively. Moreover, the main-chain and side-chain RMSD figures for wild type N176 and its eight mutants in 20 ns MD simulations are

also provided in the supporting information (Fig. S4) to show the trend of the simulation process. Only mutant M6 deviates largely from the wild-type and other mutants with backbone and side chain geometries that give RMSD values of 1.70 and 2.41 Å, respectively. These values state that the designed structures of the active sites were always maintained during the simulation time. This implies that PRODA has found the global minimum, or the near
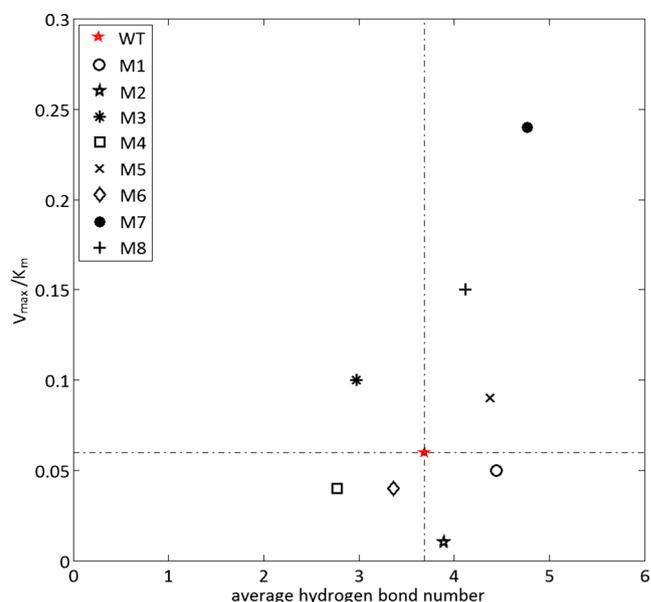
Table 2 MD simulation results for wild type N176 and its eight mutants

Mutants[a]	HB number[b]	RMSD-BB/Å	RMSD-SC/Å	V_{max}/K_m[c]
WT	3.7	1.00	2.11	0.06
M1	4.4	0.89	1.73	0.05
M2	3.9	1.15	1.87	0.01
M3	3.0	1.03	2.10	0.1
M4	2.8	1.16	1.91	0.04
M5	4.4	1.16	1.89	0.09
M6	3.4	1.70	2.41	0.04
M7	4.8	1.00	1.77	0.24
M8	4.1	0.83	1.39	0.15

[a] WT for wild type N176, and M1(HisB57Asn), M2(HisB57Phe), M3(HisB57Ser), M4(HisB57Thr), M5(HisB70Ser), M6(HisB70Tyr), M7(HisB57Ser, HisB70Ser), and M8(HisB57Ser, HisB70Ser, MetA165Ser)

[b] Average number of the hydrogen bonds at the binding sites generated from 20,000 frames for each 20 ns MD simulation

[c] The units of V_{max} and K_m are U/mg and mM, respectively.

Fig. 6 The quadrant plot of catalytic efficiency V_{max}/K_m against the average hydrogen bond number for wild type N176 and its eight mutants, where WT is given as the origin point marked with a *red star*, and the four quadrants are divided by *dashed lines*
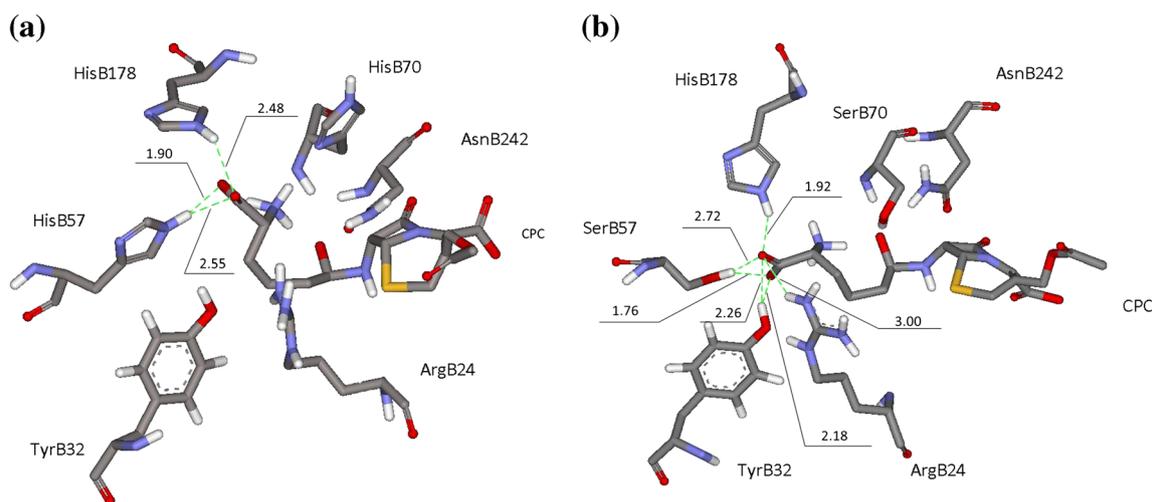
(a)

(b)

Fig. 7a,b Hydrogen bonds between functional groups of the CPC side chain and the binding residues of CA in representative MD geometry. **a** Wild type N176, **b** M7(HisB57Ser, HisB70Ser). *Green dashed lines*

Hydrogen bonds (distances in Å). Atoms: *red* oxygen atoms, *cyan* nitrogen, *grey* carbon, *white* hydrogen

global minimum, on the potential free energy landscape for each sequence. However, it is difficult to distinguish mutants with high activities from those with low activities only by examining RMSD values. The structural super-imposition of the designed geometries by PRODA and those after the 20 ns MD simulation for M7 is shown in Fig. 8. Although the binding residues in the active pocket did not move away from their geometries calculated by the PRODA design, the conformation of the CPC side chain has changed greatly. In fact, the carboxyl group of the CPC side chain is orientated in the opposite direction of the designed conformation. This orientation gives rise to a hydrogen bond between the amino group of the CPC side chain and its O16 atom, as well the carboxyl group of the CPC side chain forming a stronger hydrogen bond network. In fact, such a conformation change has been observed in the 20 ns MD simulation for the wild type N176 and all other mutants besides M7. The reason is that the electrostatic environment of the N176 active pocket has evolved to accommodate the side chain of GL-7-ACA, which has just one negatively charged group, i.e., the carboxyl group, therefore the binding residues of N176 are always positively charged, such as ArgB24, or suitable for forming hydrogen bonds with the carboxyl group of GL-7-ACA, such as TyrB32, HisB57 and HisB178. In contrast, such an electrostatic environment does not accommodate the side chain of CPC as well, because the CPC side chain not only has the negatively charged group, i.e., the carboxyl group, but also the positively charged group, i.e., the amino group. Although the mutation HisB57Ser has created enough space to accommodate CPC, the electrostatic dispersion arising from ArgB24 and HisB178 has pushed the amino group of CPC to move close to the O16 atom of CPC and the

OG atom of SerB1 to form hydrogen bonds, which are shown in Fig. 8. It should be noted that the O16 atom is negatively charged in the transition state. Therefore, further mutations that improve the electrostatic environment of the CPC side chain might promote the catalytic efficiency of CA towards CPC hydrolysis. The electrostatic attraction between the amino group and the O16 atom of CPC belongs to a long-range coulombic interaction that was not well characterized in the free energy model of PRODA. This implies that the MD simulation could be used to complement the PRODA design to reflect the dynamic behaviors of the computationally designed enzymes.
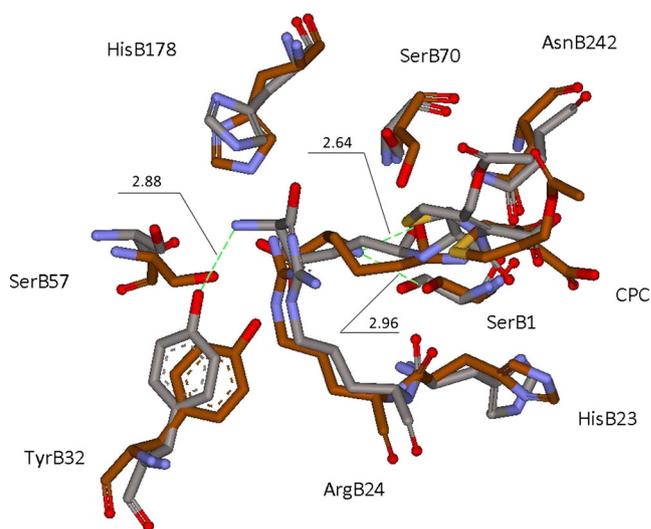

Fig. 8 Overlay of MD geometry in last frame and the structure designed by PRODA for M7(HisB57Ser, HisB70Ser). *Grey* Carbon atoms in designed structure, *brown* carbon atoms in MD geometry. *Green dashed lines* hydrogen bonds (distances in Å). Atoms: *red* oxygen, *cyan* nitrogen, *grey* carbon, *white* hydrogen

Conclusions

Active designs of cephalosporin C acylase were analyzed using MD simulations starting from the active-site-transition-state complex structures of its catalyzed reaction. The enzyme-substrate complex structures between substrate cephalosporin C and the active site of the enzymes were obtained using a molecular docking program, i.e., LigandFit, and the docking results showed that the kinetic parameter K_m can be approximated by the dissociation constant between the substrate cephalosporin C and acylase. This implies that the rate-limiting step of the hydrolytic reaction of cephalosporin C to 7-aminocephalosporanic acid is the acylation process. The active-site-transition-state structures of the acylation process were calculated using our computational enzyme design program PRODA, i.e., the PROtein Design Algorithmic package, and their dynamic trajectories were obtained using the MD simulation program GROMACS with the GROMOS 96 force field. The dynamics simulations results of the wild type acylase N176 and its eight mutants with the cephalosporin C transition state showed that the key catalytic geometrical constraints between the catalytic residues and transition state for all designs were maintained well during the 20 ns simulation, but hydrogen bonding networks at the binding sites were different for designs with different catalytic efficiencies. More hydrogen bonds were formed between the binding residues and the functional groups of cephalosporin C in the active pocket for mutants with higher activities. For all mutants, the starting calculated conformations of small molecule transition states were changed greatly following completion of the dynamics simulation. These observations imply that novel designs of acylase with higher catalytic efficiencies towards cephalosporin C hydrolysis could be potentially discovered by inserting mutations that improve the hydrogen bonding networks in the active pocket to stabilize the transition states, thereby reducing the activation barrier.

References

1. Wenda S, Illner S, Mell A, Kragl U (2011) Industrial biotechnology—the future of green chemistry? Green Chem 13:3007–3047
2. Arnold FH (2001) Combinatorial and computational challenges for biocatalyst design. Nature 409:253–257
3. Bornscheuer U, Huisman G, Kazlauskas R, Lutz S, Moore J, Robins K (2012) Engineering the third wave of biocatalysis. Nature 485:185–194
4. Pauling L (1946) Molecular architecture and biological reactions. Chem Eng News 24:1375–1377
5. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O (2008) Kemp elimination catalysts by computational enzyme design. Nature 453:190–195
6. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF (2008) De novo computational design of retro-aldol enzymes. Science 319:1387–1391
7. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, Clair JLS, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 329:309–313
8. Alexandrova AN, Röthlisbergeer D, Baker D, Jorgensen WL (2008) Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. J Am Chem Soc 130:15907–15915
9. Frushicheva MP, Cao J, Chu ZT, Warshel A (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. Proc Natl Acad Sci USA 107:16869–16874
10. Kiss G, Rothlisberger D, Baker D, Houk KN (2010) Evaluation and ranking of enzyme designs. Protein Sci 19:1760–1773
11. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL (2012) Iterative approach to computational enzyme design. Proc Natl Acad Sci USA 109:3790–3795
12. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grütter MG, Privett HK, Mayo SL, Hilvert D (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature 503:418–421. doi:10.1038/nature12623
13. Lassila JK, Privett HK, Allen BD, Mayo SL (2006) Combinatorial methods for small-molecule placement in computational enzyme design. Proc Natl Acad Sci USA 103:16710–16715
14. Zhu Y (2007) Mixed-integer linear programming algorithm for a computational protein design problem. Ind Eng Chem Res 46:839–845
15. Lei Y, Luo W, Zhu Y (2011) A matching algorithm for catalytic residue site selection in computational enzyme design. Protein Sci 20:1566–1575
16. Huang X, Han K, Zhu Y (2013) Systematic optimization model and algorithm for binding sequence selection in computational enzyme design. Protein Sci 22:929–941
17. Kim Y, Hol WG (2001) Structure of cephalosporin acylase in complex with glutaryl-7-aminocephalosporanic acid and glutarate: insight into the basis of its substrate specificity. Chem Biol 8:1253–1264
18. Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC (1995) Penicillin acylase has a single-amino-acid catalytic centre. Nature 373:264–268
19. Pollegioni L, Lorenzi S, Rosini E, Marcone GL, Molla G, Verga R, Cabri W, Pilone MS (2005) Evolution of an acylase active on cephalosporin C. Protein Sci 14:3064–3076
20. Golden E, Paterson R, Tie WJ, Anandan A, Flematti G, Molla G, Rosini E, Pollegioni L, Vrielink A (2013) Structure of a class III engineered cephalosporin acylase: comparisons with class I acylase and implications for differences in substrate specificity and catalytic activity. Biochem J 451:217–226
21. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854

22. Schuttelkopf AW, Van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr Sect D: Biol Crystallogr 60:1355–1363

23. Lemkul JA, Allen WJ, Bevan DR (2010) Practical considerations for building GROMOS-compatible small-molecule topologies. J Chem Inf Model 50:2221–2235

24. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:14101–014107

25. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52:7182–7190

26. Hess B, Bekker H, Berendsen HJ, Fraaije JG (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472

27. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092

28. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593

29. Apol E, Apostolov R, Berendsen H, Van Buuren A, Bjelkmar P, Van Drunen R, Feenstra A, Groenhof G, Kasson P, Larsson P (2010) GROMACS user manual version 4.5. 4. GROMACS user manual version 4.5. 4. Royal Institute of Technology and Uppsala University, Stockholm and Uppsala

30. Pauling L (1939) The nature of the chemical bond. Cornell University Press, Ithaca

31. Steiner T (2002) The hydrogen bond in the solid state. Angew Chem Int Ed 41:48–76

32. Krammer A, Kirchhoff PD, Jiang X, Venkatachalam C, Waldman M (2005) LigScore: a novel scoring function for predicting binding affinities. J Mol Graph Model 23:395–407

33. Hedstrom L (2002) Serine protease mechanism and specificity. Chem Rev 102:4501–4524

34. Huang X, Yang J, Zhu Y (2013) A solvated ligand rotamer approach and its application in computational protein design. J Mol Model 19:1355–1367

35. Luo W, Pei J, Zhu Y (2010) A fast protein-ligand docking algorithm based on hydrogen bond matching and surface shape complementarity. J Mol Model 16:903–913