

SCIENTIFIC REPORTS



OPEN

QM/MM free energy Simulations of an efficient Gluten Hydrolase (Kuma030) Implicate for a Reactant-State Based Protein-Design Strategy for General Acid/Base Catalysis

Xia Wang , Ruirui Li, Wenchao Cui, Qiang Li & Jianzhuang Yao 

It is a grand attraction for contemporary biochemists to computationally design enzymes for novel chemical transformation or improved catalytic efficiency. Rosetta by Baker *et al.* is no doubt the leading software in the protein design society. Generally, optimization of the transition state (TS) is part of the Rosetta's protocol to enhance the catalytic efficiency of target enzymes, since TS stabilization is the determining factor for catalytic efficiency based on the TS theory (TST). However, it is confusing that optimization of the reactant state (RS) also results in significant improvement of catalytic efficiency in some cases, such as design of gluten hydrolase (Kuma030). Therefore, it is interesting to uncover underlying reason why a better binding in the RS leading to an increased k_{cat} . In this study, the combined quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy (PMF) simulations, pK_a calculation, and the statistical analysis such as the ANOVA test were carried out to shed light on the interesting but elusive question. By integration of our computational results and general acid/base theory, we answered the question why optimization of RS stabilization leads to a better TS stabilization in the general acid/base catalysis. In addition, a new and simplified protein-design strategy is proposed for the general acid/base catalysis. The idea, that application of traditional well-defined enzyme mechanism to protein design strategy, would be a great help for methodology development of protein design.

Enzymes are fundamental catalysts that are involved in most life processes^{1,2}. enzymes are no doubt highly efficient catalysts, with significantly enhanced reaction rates compared to the corresponding uncatalyzed reactions in solution³⁻⁵. Therefore, it is of fundamental and practical importance to understand the origin of the detailed catalytic mechanisms of the enzymes. In order to explain the origin of the catalytic effect of enzymes, the uncatalyzed reaction in solution is generally used as a reference state. Thus, the catalytic effect (k_{cat}/k_o) is evaluated by the rate constant of k_{cat} for enzymatic reaction divided by the rate constant (k_o) for the corresponding uncatalyzed reaction in water. The enhancement factor of the catalytic effect sometimes can achieve more than $10^{17.5}$. The origin of the catalytic effect has been studied extensively and ideas have been proposed to explain the origin of the catalytic effect of the enzymes⁶. One general and traditional explanation proposed by Linus Pauling is the transition state theory (TST) in which the catalytic efficiency of enzyme is due to the better binding between enzyme and substrate in the transition state (TS) rather than in the reactant state (RS)⁷. In other words, enzyme and substrate binds tighter in TS rather than RS. Another basic idea related to the catalytic power of enzymes is that the functional groups of the catalytic residues are perfectly oriented⁸. Despite the rightness of this classical idea, it is still not clear how the stronger binding and perfect orientation achieved by enzymes in the transition state. To fill the gap, catalytic opinions are proposed continually. The important modern understanding of enzymatic catalytic

School of Biological Science and Technology, University of Jinan, Jinan, 250022, China. Correspondence and requests for materials should be addressed to J.Y. (email: bio_yaojz@ujn.edu.cn)

efficiency is summarized here by terms, including electrostatic preorganization, near-attack conformer, proton tunneling, reactant destabilization, dynamical effects, acid/base catalysis, covalent catalysis, and so on.

Compared to the fruitful insights by understanding the enzyme mechanism, protein (enzyme) design seems still at its infancy, in the sense that it remains challenging to screen mutations by both experimental work and make quantitative predictions on the catalytic proficiency ($k_{\text{cat}}/K_{\text{M}}$) of an engineered enzyme with an entirely generic computational approach⁹. In recent decades, protein design is advanced in two directions, including *de novo* enzymes design (creation of new enzymes) and re-design of naturally-evolved enzymes to gain a significantly improved catalytic proficiency^{10,11}. As is well-known, design methodology provided by Rosetta is the leading computational approach in the enzyme design society¹². A couple of successful design cases had been shown by previous studies using Rosetta^{10,11,13,14}. The design protocol available in Rosetta can be briefly summarized as below¹². First, modeling the active site with minimal residues according to the chosen catalytic mechanism; second, searching the minimal active site in a scaffold protein database; third, enhancement of the transition state stabilization by re-design of the surrounding residues in the active site; fourth, proving and ranking the designed enzymes by experimental work. Before the optimization of the TS complex by mutation of the active site residues, the detailed catalytic mechanism of the enzymes must be understood clearly and the rate-limiting transition-state complex needs to be modeled firstly by the QM method. The whole process is still too complicated for non-computational biochemists, so a simplified procedure is expected and would be a great help for protein design society. The well-defined catalytic mechanisms may be a potential theory database for development of the easy-to-use protein design approach. Thereby, it is interesting to ask whether the theories of the catalytic mechanism of enzymes (e.g., general acid/base catalysis) discussed above can help to simplify the procedure of protein design and thus advance the progress of next level methodology development in enzyme design.

Generally speaking, all discovered naturally-evolved enzymes play their catalytic roles strictly following the well-known rules (as discussed above), such as enzymes in the Sedolisin family¹⁵. As described by our previous computational studies^{16,17}, Sedolisins adopt (but not limited to) the acid/base and the covalent catalysis to perform their peptidase role. For example, the wild type kumamolisin-As (KumaWT) shares the similar catalytic mechanism with other Sedolisins¹⁸, but has some specific properties^{19,20}. First, the optimal pH for enzymatic activity is pH2–4; Second, the high activity and stability at the 37 °C (the physiologic temperature); Third, KumaWT as a peptidase exhibits a substrate specificity for Pro at the P2 position; Fourth, a good amount of soluble protein is obtained by expression and purification of the recombinant KumaWT using a standard protocol in *E. coli*. These properties make KumaWT as an ideal enzyme scaffold to develop an oral enzyme therapeutic for celiac sprue caused by gluten in everyday food consuming²¹. The PQ motif found in the immunogenic peptides of gluten is proved to be the immunological antigens leading to a series of immune response (e.g., inflammatory reactions) in human the digestive tract²². However, the substrate specificity of KumaWT favors positive charged residues (e.g., His, Arg, and Lys) in the S1 binding pocket²⁰. To change the substrate specificity of KumaWT toward PQ-containing peptides, Gordon *et al.* computationally designed 261 mutants by Rosetta, and named the most active design as Kuma010²¹. Mature Kuma010 owns six mutations: S73K, N102D, D104T, G130S, D169G, D179H. Experimental study proved that Kuma010 shows more than 100 fold better peptidases reactivity than KumaWT. Using Rosetta, Kuma010 was further re-designed by Wolf *et al.*²³, in order to gain greater catalytic activity and broader substrate specificity. Compared to Kuma010, the new design named Kuma030 represents a 30-fold better catalytic efficiency. Seven mutations (K73E, E80T, S165Q, G169S, D210Q, A260Q, I274T) were introduced to Kuma030. The Kuma010 and Kuma030 design approach basically adopted the general design protocol of Rosetta with a confusing assumption that the better enzyme-substrate binding in the RS leads to a better stabilization in TS, and thus increases the k_{cat} ^{21,23}. In other words, Kuma030 was not designed by optimization of the TS complex, but the RS complex. According to the TST, an improved k_{cat} value must root in a better binding in TS than RS. The change of the binding free energy in RS can be roughly used to evaluate the Michealis constant (K_{M}), not the turnover number (k_{cat}). So, a gap exists between TST and Kuma030-design protocol adopted by Rosetta. Considering the assumption of Rosetta works in the design of gluten hydrolases, this elusive question motivated us to uncover the underlying reason why a better binding in the RS leading to an increased k_{cat} . To clarify this confusing and fill the gap, the detailed catalytic mechanism of KumaWT and Kuma030 must be firstly understood by calculating the free energy profile along the RC. On the basis of completely understanding the fundamental enzymatic reaction mechanism and the origin of the improved catalytic efficiency of the mutant, next generation of protein design strategy is expected to be proposed.

With these questions in mind, we performed free energy (PMF) simulations based on the combined quantum mechanical/molecular mechanical (QM/MM)²⁴ molecular dynamics (MD), pK_{a} calculation, and the statistical analysis such as the ANOVA test. This computational study firstly uncovers the fundamental catalytic mechanism of the KumaWT towards the peptide (PFPQPQQPF), and then investigates the origin of the improved catalytic efficiency of Kuma030. Finally, a simple, fast, and reliable approach for enzyme design on general acid/base catalysis is proposed.

Results

The catalytic mechanism of the substrate (peptide PFPQPQQPF) hydrolysis by Kuma010 had been studied by our previous QM/MM study²⁵, as shown in Fig. 1. Consistent with other Sedolisins and autocatalytic process of KumaWT, the reaction pathway consists of acylation and deacylation processes^{16–18}. Consistent with the previous studies, the potential energy profile also shows the acylation process of Kuma030 with PFPQPQQPF is the rate-limiting step. Considering the catalytic turnover number (k_{cat}) is related to the activation free energy barrier, only rate-limiting steps (acylation processes) were simulated by free energy simulations for both enzymes (KumaWT and Kuma030) toward peptide (PFPQPQQPF).

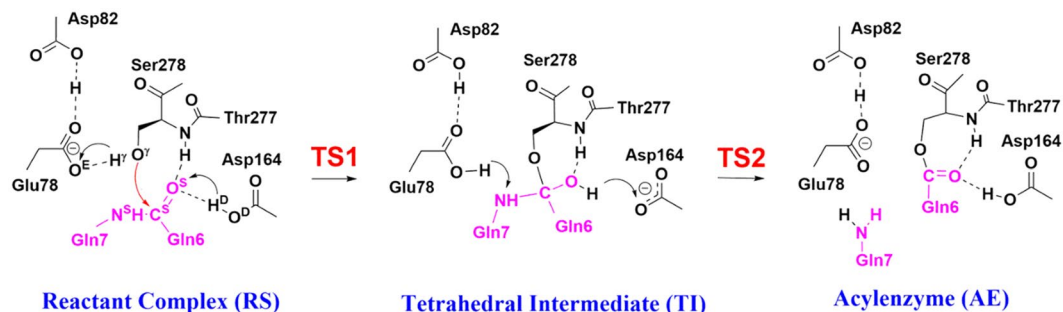


Figure 1. The proposed acylation catalytic mechanism of KumaWT, and Kuma030 complexed with a substrate peptide (PFQPQQPF) and the role of important active-site residues. The reaction coordinates (RC) applied in this study are denoted as $RC1 = r(O^{\gamma} \dots H^{\gamma}) - r(O^{\epsilon} \dots H^{\gamma}) - r(C^{\delta} \dots O^{\gamma})$, $RC2 = r(O^{\delta} \dots H^{\delta}) - r(O \dots H^{\delta})$, and $RC3 = r(N^{\delta} \dots C^{\delta}) + r(O^{\epsilon} \dots H^{\gamma}) - r(N^{\delta} \dots H^{\gamma})$. The RCs are all in Å.

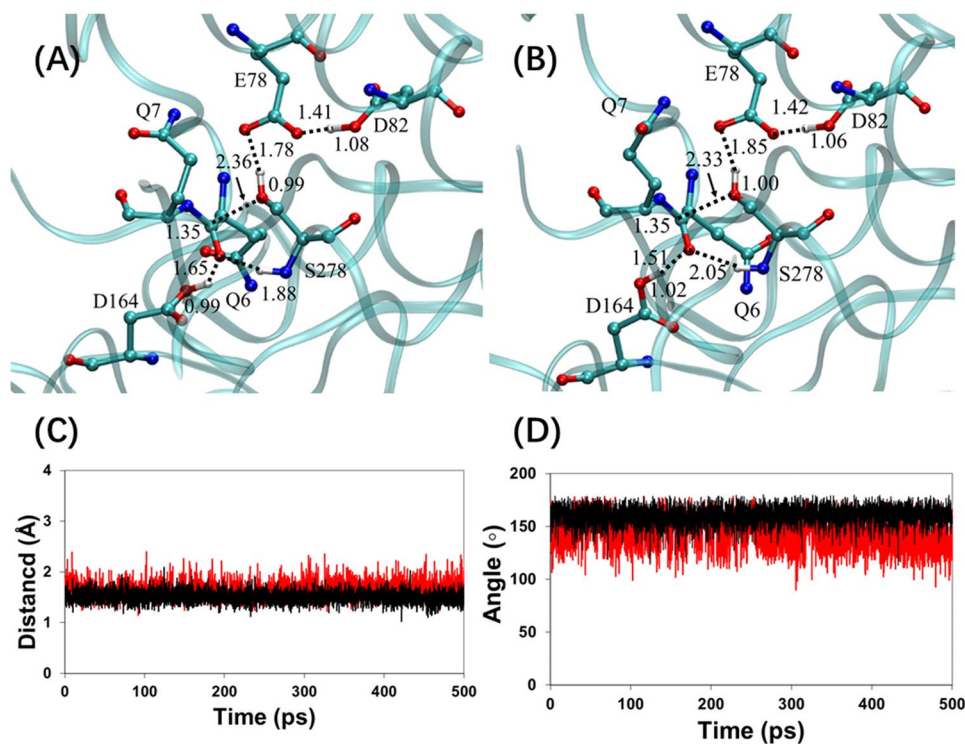


Figure 2. Reactant state (RS) complexes and distributions of distance and angle obtained by the QM/MM(DFTB3/CHARMM36) MD simulation. (A) Average structure of kumamolisin-As RS complex generated by 500 ps QM/MM MD simulation. (B) Average structure of Kuma030 RS complex generated by 500 ps QM/MM MD simulation. (C) Distance distribution of the hydrogen bond $r(O^{\delta} \dots H^{\delta})$ during 500 ps MD simulation. (D) Distribution of the hydrogen bond angle $\alpha(O^{\delta} \dots H^{\delta} - O^{\delta})$ during 500 ps MD simulation.

Enzymes-PFQPQQPF binding mode. The QM/MM(DFTB3/CHARMM36) MD simulations, in which the atoms included in QM region were described by the third-order self-consistent charge density functional tight-binding (DFTB3, 3ob-2-1)²⁶ method and the MM atoms were described by the classical CHARMM36 force field²⁷, were performed to reveal the binding properties in the RS complexed formed by KumaWT and Kuma030 with the substrate (PFQPQQPF). As shown in Fig. 2, average structures of active sites show a good interaction between substrates and the active sites of KumaWT and Kuma030. The important residues of KumaWT and Kuma030 consist of the catalytic triad (E78, D82, and S278) and residues forming the oxyanion hole. The oxyanion hole is constructed by the side chain of Asp164 and the amide groups of Ser278. The hydrogen bond networks were formed within the catalytic triad in both RS complexes. The hydrogen bond network is formed among the catalytic triad^{16–18}. KumaWT and Kuma030 share the same patterns on the interactions between enzymes and substrates in the active sites, but the strength of the interactions is slightly different with each other. On one hand, the Ser278 O^{γ} atoms of KumaWT and Kuma030 interact well with the P1-Gln6 C^{δ} of substrate (PFQPQQPF). The corresponding interaction distances are 2.36 and 2.33 Å, respectively. Those distances make the O^{γ} atoms of Ser278 in good position for nucleophilic attack toward the C^{δ} atoms of the substrates. On the other hand, the

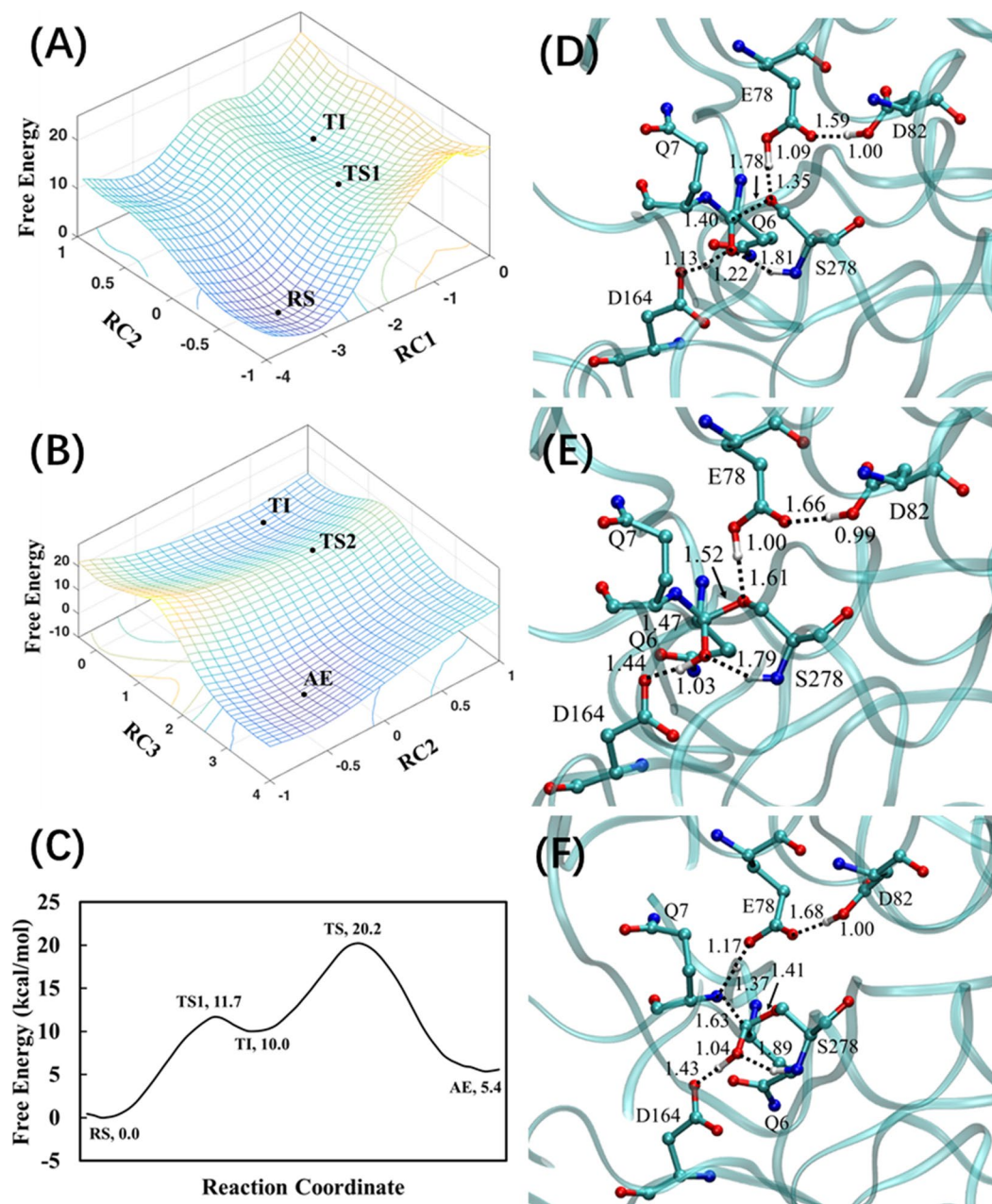


Figure 3. Free energy maps and average structures of KumaWT with the substrate peptide (PFPQQQPF) for acylation process. (A) Use of RC1 and RC2 to simulate the formation of the TI in the acylation process; (B) RC2 and RC3 to simulate the decomposition of the TI. (C) Minimum free energy profile plotted on the basis of Fig. 3A,B. Average structures of TS1 (D), TI (E), and AE (F). The distances are all in Å. The energies are in kcal/mol. All transition states are approximate TS.

carbonyl O^s atom of the P1-Gln6 backbone forms two hydrogen bonds in the oxyanion hole, which are donated by the protonated side chain of Asp164 and the NH group of Ser278 backbone. The Asp164 in Kuma010 and corresponding Asp170 in Sedolisins are suggested to function as general acid/base catalysts in the acylation processes^{15,28}. Interestingly, the average hydrogen-bond distances are 1.65 and 1.51 Å for RS complexes of KumaWT and Kuma030, respectively. The stronger hydrogen bond of Kuma030 may partly account for the better binding designed by Rosetta in the RS compared to KumaWT.

Acylation reaction pathway in KumaWT and substrate complex. Using the KumaWT-PFPQQQPF structure as the RS (Michaelis-Menten) complex (depicted in Fig. 2A), the hybrid QM/MM(DFTB3/CHARMM36) PMF (free energy) simulations were performed to simulate the acylation process of KumaWT toward the peptide (PFPQQQPF). With two-sets RCs, the two-dimensional (2D) free energy (PMF) profiles were determined and plotted in Fig. 3A,B. Figure 3C shows the one dimensional minimum free energy profile

generated based on Fig. 3A,B. All free energy profiles demonstrate that the RS complex went through two TSs and one tetrahedral intermediate (TI). Acyl-enzyme (AE) was finally formed at the end of the acylation process of KumaWT. The whole acylation process follows the classical catalytic mechanism of the Sedolisins family^{15,25}.

As shown in Fig. 3D–F, the acylation process includes TS1, TI, TS2, and AE stages. The acylation process consists of formation of a C^S–O^γ bond between the P1-Gln6 backbone of substrate and the Ser278 side chain of KumaWT (related to TS1) and breaking of the C^S–N^S peptide bond of the substrate (related to TS2). Simultaneously, two protons are transferred associated with these two TSs. In the TS1, the H^γ atom located at Ser278 sidechain is transferred to Glu78 O^E atom, while the H^D atom located at Asp164 is transferred to the P1-Gln6 backbone O^S atom. In the rate limiting TS2, the H^γ atom of Ser278 currently located at the O^E atom of the Glu78 side chain is transferred to the backbone N^S atom of the P2-Gln7, while the H^D atom currently located at the P1-Gln6 backbone O^S atom is transferred back to carboxylate group of Asp164 sidechain.

As shown in Fig. 3, the chosen reaction coordinates are similar with previous Kuma010 QM/MM study²⁵. For instances, RC1 = $r(\text{O}^\gamma \dots \text{H}^\gamma) - r(\text{O}^E \dots \text{H}^\gamma) - r(\text{C}^S \dots \text{O}^\gamma)$, RC2 = $r(\text{O}^D \dots \text{H}^D) - r(\text{O}^S \dots \text{H}^D)$, and RC3 = $r(\text{N} \dots \text{C}) + r(\text{O}^E \dots \text{H}^\gamma) - r(\text{N} \dots \text{H}^\gamma)$ and RC2.

Acylation reaction pathway in Kuma030 and substrate complex. Using RS complex as the starting point, the same QM/MM(DFTB3/CHARMM36) PMF simulations were carried out to simulate the acylation process of Kuma030 with the substrate peptide (PFPPQPPF), as shown in Fig. 4A,B. The minimum free energy profile (depicted in Fig. 4C) demonstrates that Kuma030 shares the similar acylation mechanism with KumaWT, including a TI sandwiched by two TSs. Therefore, the two sets of RCs are chosen identically with KumaWT simulations.

Figures 4D–F show the two-steps chemical transformation from the reactant state (RS) to the tetrahedral intermediate (TI) through transition state (TS1), and from TI to acyl-enzyme (AE) through the transition state (TS2). In the TS1 stage of Kuma030, the H^D atom of Asp164 has been transferred to the P1-Gln6 O^S atom with a distance $r(\text{O}^S \dots \text{H}^D)$ of 1.00 Å (see Fig. 4D), while the corresponding proton transfer process in KumaWT is still at the middle stage with the $r(\text{O}^S \dots \text{H}^D) = 1.22$ Å (see Fig. 3D). The similar trend was also found in the H^γ atom proton transfer process. For example, the distances of $r(\text{O}^E \dots \text{H}^\gamma)$ are 1.35 and 1.26 Å, respectively. In the TI as shown in Figs 3E and 4E, all of the H^γ atom of KumaWT and Kuma030 have been transferred to the Glu78 O^E atom. Similarly, all of the H^D atom of KumaWT and Kuma030 have been transferred to the O^S atom of the P1-Gln6 with the distances of 1.03 and 0.99 Å, respectively. In the TS2 as shown in Figs 3F and 4F, a similar proton transfer pattern was found between KumaWT and Kuma030. For instance, the transfers of the H^γ atom to the N^S are all in their middle stage with distances $r(\text{N} \dots \text{H}^\gamma)$ of 1.37 and 1.40 Å, respectively for KumaWT and Kuma030, while the transfers of the H^D atom back to Asp164 are all in its beginning stage with the corresponding distances $r(\text{N} \dots \text{H}^D)$ of 1.43 and 1.50 Å.

Calculated free energy profiles and experimental kinetic data. According to the minimum free energy profiles depicted in Figs 3C and 4C, both acylation processes of KumaWT and Kuma030 are two steps reactions involving a RS, two TSs, a TI, and an AE. As discussed above, the acylation is the rate limiting step for the KumaWT and Kuma030 toward the substrate peptide (PFPPQPPF), so the activation free energy barriers can be read from the minimum free energy profiles. The free energy barriers (corresponding to TS1 and TS2) of Kuma030 are 10.7 and 18.1 kcal/mol, respectively. So, the rate-limiting step is associated with the second TS in the acylation process of Kuma030, which is consistent with the wild type Sedolisins and Kuma010 catalyzed reaction. The k_{cat} value is not available for KumaWT with the substrate peptide (PFPPQPPF), while the k_{cat} value for Kuma030 can be calculated based on the experimental Michaelis-Menten curves²³. As shown in the Figure S1 of the Kuma030 experimental study²³, the k_{cat} value for Kuma030 is roughly 75.7 S^{-1} . According to the conventional TST²⁹, the experimental derived activation free energy barrier (under 37°C) is 15.5 kcal/mol. Our previous benchmark calculation for Kuma010 with the substrate peptide (PFPPQPPF) showed that DFTB3/MM overestimates the potential energy barrier by around 3.1 kcal/mol in the rate-limiting TS, compared to B3LYP/MM method³⁰. After an energy correction of DFTB3/MM systematical error, the calculated activation free energy barrier of Kuma030 with the substrate peptide (PFPPQPPF) is 15.1 kcal/mol, which is in a good agreement with the experimental result (15.5 kcal/mol).

Discussion

Although numerous chemical transformations had been found and characterized in the naturally evolved enzymes, it is still super attractive for computational biochemists to design proteins with novel or improved catalytic functions. The major challenge for modern protein design is to develop efficient methods for ranking of designed enzymes before experimental proof¹².

Computational ranking of the k_{cat} is possible, because the enzymatic reaction rates (k_{cat}) and the corresponding activation free energy barrier can be accurately converted based on the TST³¹. To computationally determine the value of k_{cat} for the chemical steps catalyzed by native or engineered enzymes, the corresponding activation free energy barrier can be obtained based on the calculations of the free energy profiles along the reaction coordinate (RC). The combined QM/MM approaches²⁴, are the choice to compute the free energy difference along the RC by potential of mean force (PMF) simulations and thus determine the activation free energy barriers for enzyme-catalyzed reactions. QM/MM methods based on high level QM basis set are still too computationally expensive to perform thousands of QM/MM simulations simultaneously, which is required by mutates screening of enzyme design protocol of Rosetta. Alternatively, semiempirical QM approaches, such as the self-consistent-charge density-functional-tight-binding (SCC-DFTB), have been developed^{32,33} to expedite the QM/MM MD simulations. However, supercomputers with hundreds of the processors are still needed to calculate thousands of free energy profiles for the mutants designed by Rosetta. Although possible, it is still a tremendous

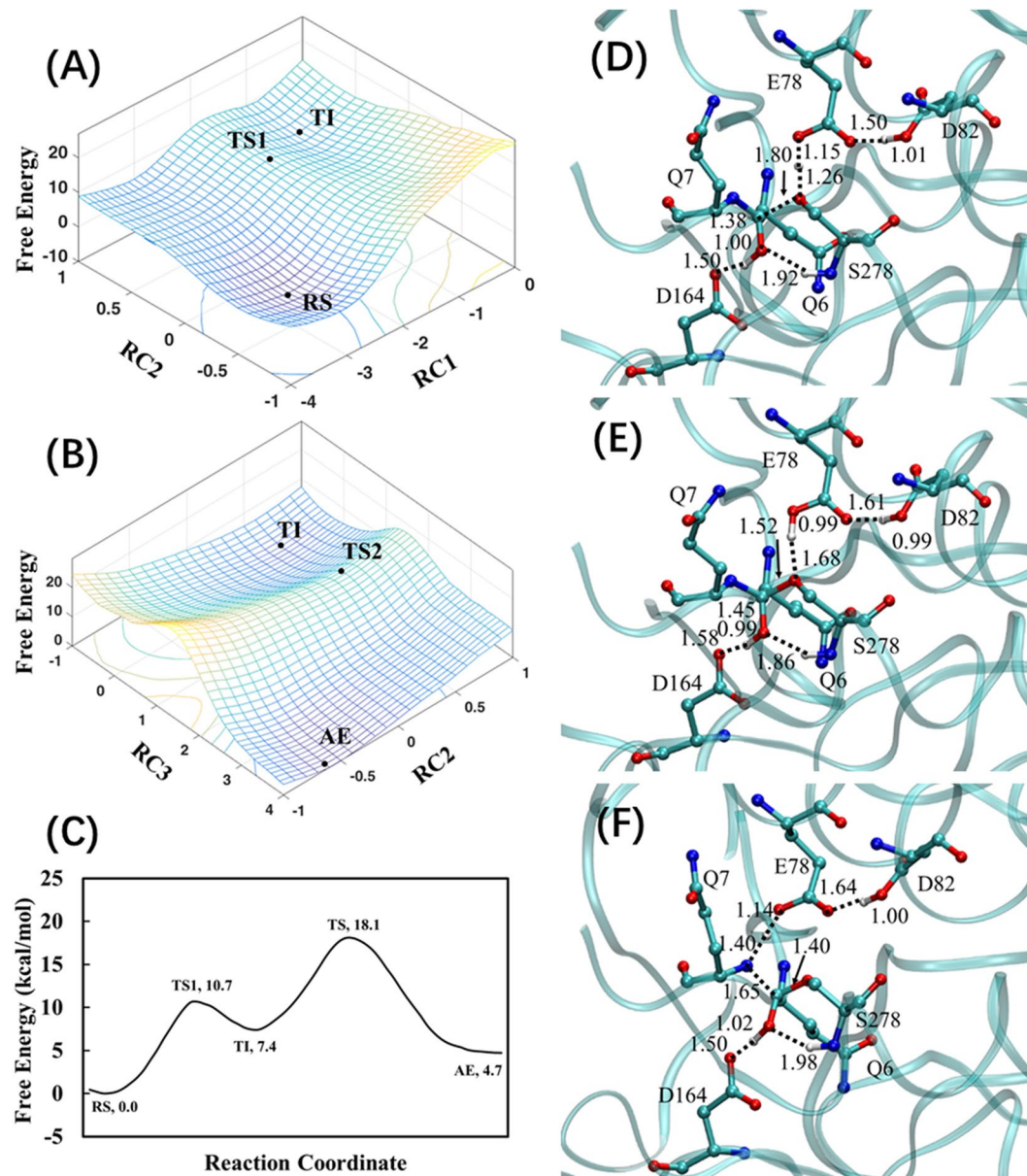


Figure 4. Free energy maps and average structures of Kuma030 with the substrate peptide (PPFQPQQPF) for acylation process. (A) Use of RC1 and RC2 to simulate the formation of the TI in the acylation process; (B) RC2 and RC3 to simulate the decomposition of the TI. (C) Minimum free energy profile plotted on the basis of Fig. 4A,B. Average structures of TS1 (D), TI (E), and AE (F). The distances are all in Å. The energies are in kcal/mol. All transition states are approximate TS.

burden for experimental labs and even computational labs. To this end, it is an urgent but challenging work to develop simple, fast, and reliable method(s) to rank the k_{cat} values of designed enzymes before the experimental work. In this study, we tried to build up efficient ranking method by examination the well-known catalytic mechanism of enzymes, such as general acid/base catalysis.

General acid/base catalysis is the fundamentally important catalytic mechanism widely adopted by naturally-evolved enzymes, such as serine proteinases^{34,35}. Acid/base catalysis improves the catalytic activity by proton-transfer process. In the rate limiting step of enzyme-catalyzed reactions, a proton can be transferred either to substrates (reactant) from active site acid residues or from substrates to active site base residues, which results in reducing the activation free energy barriers. Interestingly, the activation free energy barrier of the proton transfer reaction can be lowered by reducing the pK_a difference between proton donor and acceptor^{36,37}. Under the matching pK_a condition, the coupling of proton transfer can be maximized, leading to the shortening of the length of the hydrogen bond³⁷⁻³⁹. The length of the hydrogen bond can be analyzed by the strength and thus angle of the hydrogen bond. Thus, k_{cat} changes of mutants can be roughly measured by pK_a difference between proton donor and acceptor for proton-transfer reaction involved in the rate-limiting step. Because length and strength

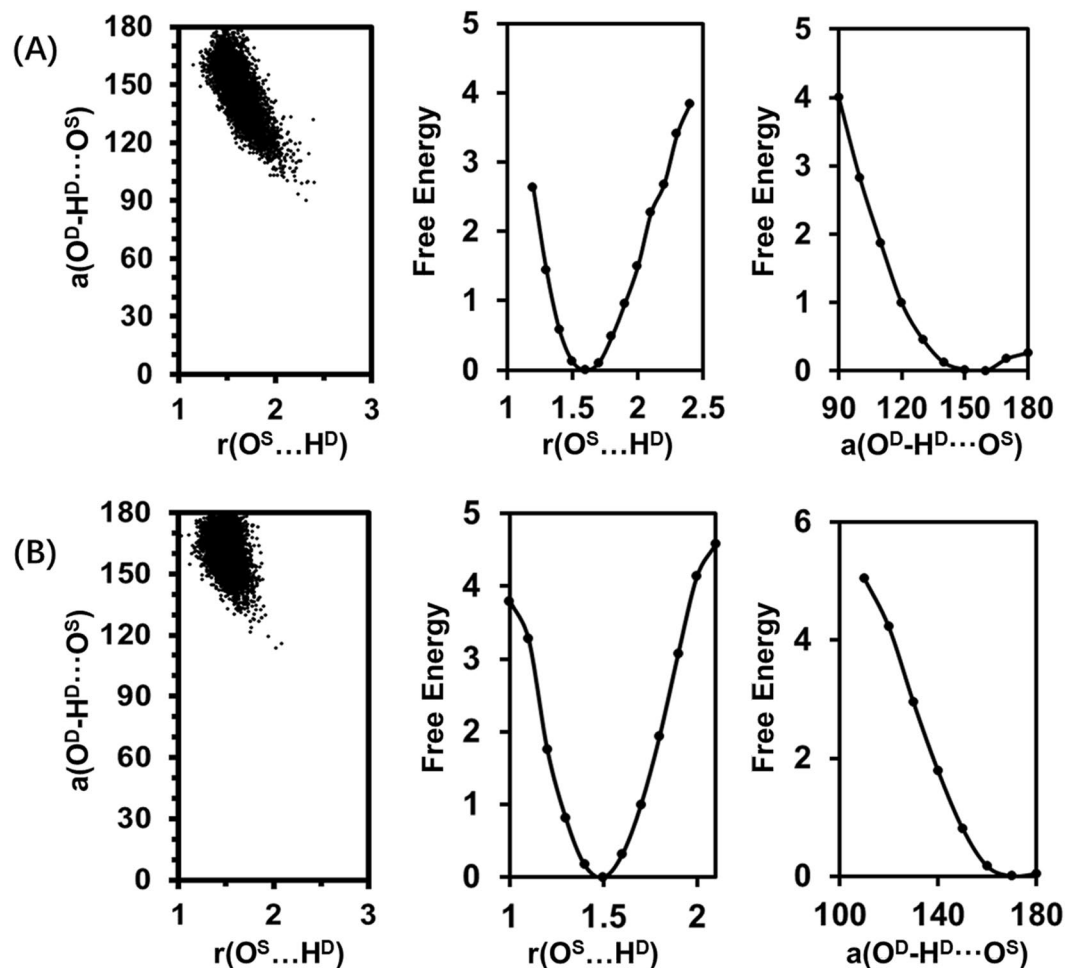


Figure 5. The distributions of $r(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}})$ and $a(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}}\text{-O}^{\text{D}})$ and the corresponding free-energy changes from the 500 ps QM/MM MD simulations on the reactant complexes of kumamolisin-As (A) and Kuma030 (B). (Left) The two-dimensional plot of $r(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}})$ and $a(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}}\text{-O}^{\text{D}})$ distributions. (Center) The free-energy change as a function of $r(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}})$ obtained from the distributions. (Right) The free-energy change as a function of $a(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}}\text{-O}^{\text{D}})$ obtained from the distributions. The distances are given in Å. The angles are in degree ($^{\circ}$). The energies are in kcal/mol.

of the hydrogen bond is coupled with the $\text{p}K_{\text{a}}$ difference^{37–39}, it is reasonable to propose that the k_{cat} changes may also be ranked by comparison of $\text{p}K_{\text{a}}$ difference, length and angle of hydrogen bonds in the RS.

In this study, we use KumaWT and Kuma030 with higher k_{cat} designed by Rosetta as an example to test our hypothesis. First, the relationship between $\text{p}K_{\text{a}}$ difference and corresponding free energy barrier was studied. In the first step of acylation process, the TS1 of Kuma030 shows a lower free energy barrier (10.7 kcal/mol) compared to KumaWT (11.7 kcal/mol). The proton transfer from the Asp164 to the P1-Gln6 is involved in the TS1. To measure the $\text{p}K_{\text{a}}$ difference in the RS, we assume that the $\text{p}K_{\text{a}}$ change of the O^{S} atom of the P1-Gln6 can be negligible upon the mutation of KumaWT to Kuma030 and the $\text{p}K_{\text{a}}$ change of the O^{S} atom of the P1-Gln6 along the RC due to the developing charge increase is the same between KumaWT and Kuma030. So, the change of $\text{p}K_{\text{a}}$ difference from KumaWT to Kuma030 can be determined by the $\text{p}K_{\text{a}}$ change of the O^{D} atom of Asp164. The optimized RS complexes from the last snapshots of the MD simulation for both KumaWT and Kuma030 were submitted to H++ using default setting (pH 4.0)⁴⁰. The predicted $\text{p}K_{\text{a}}$ of KumaWT is 6.30, while the predicted $\text{p}K_{\text{a}}$ of Kuma030 is 5.86, which implicate for a reducing $\text{p}K_{\text{a}}$ difference from KumaWT to Kuma030, thus resulting in a decreased TS1 free energy barrier of Kuma030 compared to KumaWT. The result is consistent with our PMF calculation, suggesting that a simple $\text{p}K_{\text{a}}$ calculation is good enough for the k_{cat} ranking purpose in the general acid/base catalysis.

Second, to test our hypothesis that the free energy barriers can be ranked by comparison of the hydrogen bond lengths in the RSs, the relationship between k_{cat} and bond lengths and angles is investigated by analyses of the distributions of the hydrogen bond length $r(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}})$ and angle $a(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}}\text{-O}^{\text{D}})$ as well as the corresponding free energy changes, as shown in Fig. 5. Compared to KumaWT, the Kuma030 shows a large population for the structures of hydrogen bond with relatively short $r(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}})$ and large $a(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}}\text{-O}^{\text{D}})$, which is consistent with the average distances and angles depicted in Fig. 2. In the wise of free energy changes along with $r(\text{O}^{\text{S}}\cdots\text{H}^{\text{D}})$

data	Source of Variation	SS	df	MS	F	P-value
DIS	Between Groups	53.13601861	1	53.13602	2827.083	< 2e-16***
	Within Groups	187.9159414	9998	0.018795		
ANG	Between Groups	664067.812	1	664067.812	4707.510	< 2e-16***
	Within Groups	1410373.927	9998	141.066		

Table 1. One-way ANOVA test result. DIS: $r(O^S \dots H^D)$; ANG: $a(O^S \dots H^D \dots O^D)$.

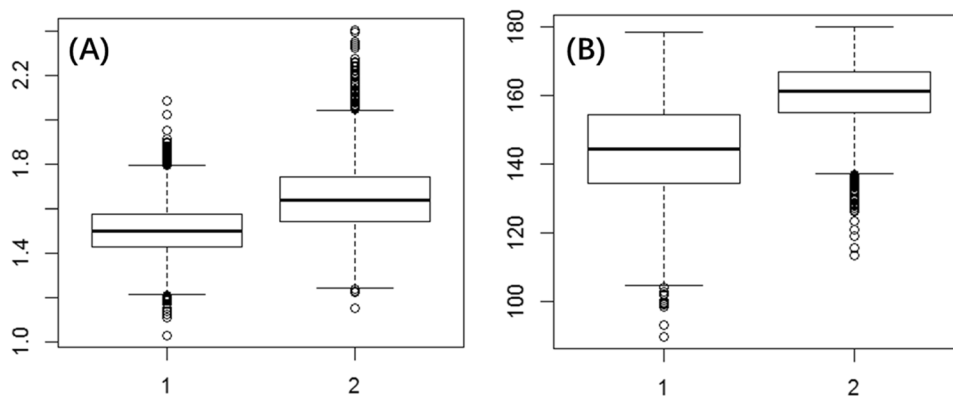


Figure 6. The one-way ANOVA test between RS complexes of kumamolisin-As and Kuma030. (A) distance $r(O^S \dots H^D)$ comparison; (B) $a(O^S \dots H^D \dots O^D)$ comparison. Sample 1 is Kuma030 and Sample 2 is kumamolisin-As. 5000 windows were obtained based on 500 ps QM/MM MD simulation.

and $a(O^S \dots H^D \dots O^D)$, Kuma030 shows relatively low free energies at short $r(O^S \dots H^D)$ and large $a(O^S \dots H^D \dots O^D)$, and implicates for an easy proton transfer process in term of the free energy barrier, which agrees well with the free profiles of the first proton transfer reactions between Asp164 and P1-Gln6 obtained by PMF calculations (see Figs 3C and 4C). Indeed, this is also the case for the second proton transfer processes between Asp164 and P1-Gln6 with a reverse result. In the TIs depicted in Figs 3E and 4E, KumaWT shows a better $r(O^D \dots H^D)$ with a distance of 1.44 Å compared to the Kuma030 corresponding bond length with a distance 1.58 Å, suggesting that the free energy barrier of KumaWT for the second proton transfer process would be lower than Kuma030. Indeed, the free energy barriers between TI and AE are 10.1 and 10.7 kcal/mol, respectively. Our PMF results again support our hypothesis that short length of hydrogen bond leads to low free energy barrier of proton transfer process. Therefore, k_{cat} ranking task for general acid/base may be simply be done by comparing the corresponding hydrogen bond length without knowing the activation free energy barrier by performing expensive QM/MM simulations. Furthermore, the hydrogen bond lengths are 1.66 and 1.50 Å for the last snapshots minimized RS complex of KumaWT and Kuma030, respectively. Therefore, it is reasonable to propose that optimized hydrogen bond lengths can be used for the ranking purpose as well as the average distances.

Finally, although bond length analysis seems as an easy way for the k_{cat} ranking task, we need to confirm that the difference of average bond lengths is statistically meaningful before the experimental work. In the RSs of KumaWT and Kuma030, the average bond lengths of $r(O^S \dots H^D)$ are 1.65 and 1.51 Å, respectively. Since the bond lengths of $r(O^S \dots H^D)$ are averaged from 5000 snapshots during the 500 ps MD simulation, we originally have 5000 data for each $r(O^S \dots H^D)$ from KumaWT and Kuma030. The available data from MD simulations were further analyzed by the one-way ANOVA test implanted in R package⁴¹. ANOVA shows that distance and angle are significantly different between KumaWT and Kuma030, as shown in Table 1 and Fig. 6. Here, we propose for the first time the k_{cat} values of the general acid/base catalysis of designed enzymes may be ranked by comparison of the corresponding hydrogen bond length with the ANOVA test.

Computational protein design has been emerging as a leading and important research area in the biophysics/biochemistry. In this study, with the help of well-defined idea of the general acid/base catalysis, we firstly demonstrate that the ranking task for designed enzymes can be fulfilled not only by QM/MM studies, but also can be easily performed by comparison of the difference of pK_a or the bond lengths for the corresponding proton acceptor or donor in the general acid/base catalysis. For the proteins with the general acid/base catalysis as the rate-limiting step, the simplified protein-design protocol can be summarized as below. First, modeling the reactant state of the enzyme-substrate complex; second, mutants screening and equilibration and/or optimization of the RS complex by MD simulation and/or energy minimization; third, ranking the enhancement of the TS stabilization (k_{cat}) by comparison of the hydrogen bond lengths involved in rate-limiting proton transfer step; fourth and optionally, ranking the enhancement of the binding free energy (K_m) between substrate and protein; finally, proving the design experimentally. As a simple, fast, and reliable approach, it is expected to be a great help for protein design not only by Rosetta but also by the classical MD simulation and energy minimization.

Methods

Model construction. The binding model of Kuma030 and substrate (PFPQPQQPF) were built similarly with the Kuma030-design protocol described by Wolf *et al.*²³. Briefly, the Kuma030-PFPQPQQPF complex were manually mutated based on our previously-simulated reactant state complex of Kuma010-substrate (PFPQPQQPF)²⁵. Thus, six mutations (K73E/E80T/S165Q/G169S/D210Q/A260Q/) were introduced to Kuma010 in order to build Kuma030. The model of wild type kumamolisin As-substrate complex were from our previously study²⁵.

Hydrogen atoms of enzyme-substrate complexes were added by the HBUILD module⁴² of CHARMM⁴³. The protonation states of acidic and basic residues were determined under pH 4.0 condition depending on surrounding environment. All protonation states are confirmed by H⁺⁴⁰. Using the O^γ atom of Ser278 as the enzyme-substrate complex center, the solvation of the system was performed with a 22 Å radius water droplet. Solvent water molecules close to crystal atoms (within 2.8 Å) were removed. The TIP3P water model^{27,44–46} was applied. The QM treated atoms include the side chains of E78, D82, D164, and S287, the carbonyl of substrate P5, the backbone of substrate Q6, and the C^α and amide group of substrate Q7. The MM treated atoms are the other atoms of the systems excluding the QM region. The QM and MM boundaries were treated by the divided frontier charge (DIV) link-atom scheme^{43,47,48}. The SCC-DFTB module implemented in the CHARMM⁴⁹ was employed for the QM-region atoms and the all-hydrogen CHARMM36 force field^{27,44} was employed for the MM atoms. The cut-off of the non-bonded interaction was 13 Å.

For the stochastic boundary⁵⁰ MD simulation, the reference center is the side chain O^γ atom of Ser278. The radius (r) of the reaction region was of 20 Å, while the radius of buffer region was within $20 \text{ \AA} \leq r \leq 22 \text{ \AA}$. Specifically, the reaction region was simulated using Newtonian equations-of-motion, while the buffer region was treated by solving Langevin equations-of-motion with a 298.15 K temperature bath of Langevin thermostat⁵¹. All atoms neither included in the reaction region nor buffer region were fixed. All hydrogen-involved covalent bonds were treated by the SHAKE algorithm⁵². The initial structures of the enzyme-substrate complexes were minimized firstly by the steepest descent (SD) method and then by the adopted-basis Newton-Raphson (ABNR) methods. A 1 femtosecond (fs) time step was set up for the following MD simulations. Starting from 50 K, the systems were heated to 298.15 K in 100 picosecond (ps) gradually and 500 picosecond MD simulations (production runs) were carried out for reactant state complexes (RS) of all enzyme-substrate complexes.

The hydrogen bond length $r(\text{O}^{\text{S}} \dots \text{H}^{\text{D}})$ and angle $a(\text{O}^{\text{S}} \dots \text{H}^{\text{D}} - \text{O}^{\text{D}})$ and the free energy change. The methodology was followed the previous protocol^{53–55}. Briefly, based on the QM/MM MD simulations of KumaWT and Kuma030, the distributions of $r(\text{O}^{\text{S}} \dots \text{H}^{\text{D}})$ and $a(\text{O}^{\text{S}} \dots \text{H}^{\text{D}} - \text{O}^{\text{D}})$ were monitored during the 500 ps MD simulations for the RS complexes and applied to calculate the free energies required to generate the reactive structure. A one-way ANOVA on means was conducted in R package⁴¹ using the function *aov.out* to see if the important hydrogen bond related distances and angles are significantly different between KumaWT and Kuma030.

The free energy (Potential mean force, PMF) simulation. The free energy (PMF) simulations were performed on the basis of the QM/MM(DFTB3/CHARMM36) MD simulations. The umbrella sampling method⁵⁶ and the Weighted Histogram Analysis Method⁵⁷ were employed to calculate the PMF profile as a function of the RCs. As the first step of free energy profiles determination for the acylation processes, the potential energy maps were determined by adiabatic-mapping calculations starting from the last snapshot of the 500 ps QM/MM MD simulations on the reaction systems. To obtain the 2D free energy maps, around 2000 windows were generated for each enzyme complex. For one window, 100 ps QM/MM MD simulation was carried out with the first 50 ps for equilibration. One snapshot was saved per 0.5 ps, so one hundred snapshots were obtained for one window. The force constant applied for the harmonic biasing potential was $150 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for all of the PMF calculations.

References

- Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. (Macmillan, 1999).
- Warshel, A. *Computer modeling of chemical reactions in enzymes and solutions*. (Wiley New York, 1991).
- Bruice, T. C. Computational Approaches: Reaction Trajectories, Structures, and Atomic Motions. *Enzyme Reactions and Proficiency*. *Chem. Rev.* **106**, 3119–3139 (2006).
- Lad, C., Williams, N. H. & Wolfenden, R. The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl. Acad. Sci.* **100**, 5607–5610 (2003).
- Wolfenden, R. & Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc. Chem. Res.* **34**, 938–945 (2001).
- Knowles, J. R. Enzyme catalysis: not different, just better. *Nature* **350**, 121–124 (1991).
- Pauling, L. Molecular Architecture and Biological Reactions. *Chemical & Engineering News Archive* **24**, 1375–1377 (1946).
- Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
- Frushicheva, M. P. *et al.* Computer aided enzyme design and catalytic concepts. *Curr. Opin. Chem. Biol.* **21**, 56–62 (2014).
- Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387–1391 (2008).
- Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
- Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **6**, e19230 (2011).
- Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
- Siegel, J. B. *et al.* Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* **329**, 309–313 (2010).
- Oda, K. New Families of Carboxyl Peptidases: Serine-Carboxyl Peptidases and Glutamic Peptidases. *J. Biochem.* **151**, 13–25 (2012).
- Xu, Q., Yao, J., Wlodawer, A. & Guo, H. Clarification of the Mechanism of Acylation Reaction and Origin of Substrate Specificity of the Serine-Carboxyl Peptidase Sedolisin through QM/MM Free Energy Simulations. *J. Phys. Chem. B* **115**, 2470–2476 (2011).
- Yao, J., Xu, Q. & Guo, H. QM/MM and Free-Energy Simulations of Deacylation Reaction Catalysed by Sedolisin, a Serine-Carboxyl Peptidase. *Mol. Simul.* **39**, 206–213 (2013).
- Yao, J., Wlodawer, A. & Guo, H. Understanding the Autocatalytic Process of Pro-Kumamolisin Activation from Molecular Dynamics and Quantum Mechanical/Molecular Mechanical (QM/MM) Free-Energy Simulations. *Chem. - Eur. J.* **19**, 10849–10852 (2013).

19. Comellas-Bigler, M. *et al.* 1.2 A Crystal Structure of the Serine Carboxyl Proteinase Pro-Kumamolisin; Structure of an Intact Pro-Subtilase. *Structure* **12**, 1313–1323 (2004).
20. Wlodawer, A. *et al.* Crystallographic and Biochemical Investigations of Kumamolisin-As, a Serine-Carboxyl Peptidase with Collagenase Activity. *J. Biol. Chem.* **279**, 21500–21510 (2004).
21. Gordon, S. R. *et al.* Computational Design of an α -Gliadin Peptidase. *J. Am. Chem. Soc.* **134**, 20513–20520 (2012).
22. Shan, L. *et al.* Structural Basis for Gluten Intolerance in Celiac Sprue. *Science* **297**, 2275–2279 (2002).
23. Wolf, C. *et al.* Engineering of Kuma030: A Gliadin Peptidase That Rapidly Degrades Immunogenic Gliadin Peptides in Gastric Conditions. *J. Am. Chem. Soc.* **137**, 13106–13113 (2015).
24. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).
25. Yao, J., Luo, H. & Wang, X. Understanding the Catalytic Mechanism and the Substrate Specificity of an Engineered Gluten Hydrolase by QM/MM Molecular Dynamics and Free Energy Simulations. *J. Chem. Inf. Model.* **57**, 1179–1186 (2017).
26. Gaus, M., Cui, Q. & Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **7**, 931–948 (2011).
27. Huang, J. & MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
28. Reichard, U. *et al.* Sedolisins, a New Class of Secreted Proteases from *Aspergillus fumigatus* with Endoprotease or Tripeptidyl-Peptidase Activity at Acidic pHs. *Appl. Environ. Microbiol.* **72**, 1739–1748 (2006).
29. Garrett, B. C. & Truhlar, D. G. Transition State Theory in Encyclopedia of Computational Chemistry. John Wiley & Sons: Chichester, UK (1998).
30. Yao, J., Wang, X., Luo, H. & Gu, P. Understanding the Catalytic Mechanism and the Nature of the Transition State of an Attractive Drug-Target Enzyme (Shikimate Kinase) by Quantum Mechanical/Molecular Mechanical (QM/MM) Studies. *Chemistry – A European Journal* **23**, 16380–16387 (2017).
31. Garcia-Viloca, M., Gao, J., Karplus, M. & Truhlar, D. G. How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations. *Science* **303**, 186–195 (2004).
32. Thiel, W. Semiempirical quantum-chemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 145–157 (2014).
33. Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B* **58**, 7260 (1998).
34. Eigen, M. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. *Angewandte Chemie International Edition in English* **3**, 1–19 (1964).
35. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **102**, 4501–4524 (2002).
36. Schutz, C. N. & Warshel, A. The low barrier hydrogen bond (LBHB) proposal revisited: The case of the Asp ... His pair in serine proteases. *Proteins: Struct., Funct., Bioinf.* **55**, 711–723 (2004).
37. Gerlt, J. A. & Gassman, P. G. Understanding enzyme-catalyzed proton abstraction from carbon acids: details of stepwise mechanisms for beta-elimination reactions. *J. Am. Chem. Soc.* **114**, 5928–5934 (1992).
38. Guillén Schlippe, Y. V. & Hedstrom, L. A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch. Biochem. Biophys.* **433**, 266–278 (2005).
39. Ishikita, H. & Saito, K. Proton transfer reactions and hydrogen-bond networks in protein environments. *Journal of The Royal Society Interface* **11** (2014).
40. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **40**, W537–W541 (2012).
41. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2017).
42. Brünger, A. T. & Karplus, M. Polar Hydrogen Positions in Proteins: Empirical Energy Placement and Neutron Diffraction Comparison. *Proteins: Struct., Funct., Bioinf.* **4**, 148–156 (1988).
43. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
44. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
45. Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids .10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers - Application to Liquid Water. *J. Am. Chem. Soc.* **103**, 335–340 (1981).
46. Neria, E., Fischer, S. & Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **105**, 1902–1921 (1996).
47. Field, M. J., Bash, P. A. & Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **11**, 700–733 (1990).
48. König, P. H., Hoffmann, M., Frauenheim, T. & Cui, Q. A Critical Evaluation of Different QM/MM Frontier Treatments with SCC-DFTB as the QM Method. *J. Phys. Chem. B* **109**, 9082–9095 (2005).
49. Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T. & Karplus, M. A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method. *J. Phys. Chem. B* **105**, 569–585 (2000).
50. Brooks, C. L., Brunger, A. & Karplus, M. Active-Site Dynamics in Protein Molecules - A Stochastic Boundary Molecular-Dynamics Approach. *Biopolymers* **24**, 843–865 (1985).
51. Adelman, S. A. & Doll, J. D. Generalized Langevin Equation Approach For Atom-Solid-Surface Scattering - General Formulation For Classical Scattering Off Harmonic Solids. *J. Chem. Phys.* **64**, 2375–2388 (1976).
52. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
53. Guo, H.-B. & Guo, H. Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *Proc. Natl. Acad. Sci.* **104**, 8797–8802 (2007).
54. Yao, J., Chu, Y., An, R. & Guo, H. Understanding Product Specificity of Protein Lysine Methyltransferases from QM/MM Molecular Dynamics and Free Energy Simulations: The Effects of Mutation on SET7/9 beyond the Tyr/Phe Switch. *J. Chem. Inf. Model.* **52**, 449–456 (2012).
55. Iii, C. L. B., Karplus, M., Pettitt, B. M. & Austin, R. H. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. *Phys. Today* **43**, 120 (1990).
56. Torrie, G. M. & Valleau, J. P. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **28**, 578–581 (1974).
57. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **13**, 1011–1021 (1992).

Acknowledgements

The National Natural Science Foundation of China (grant 21703079), the Shandong Provincial Natural Science Foundation, China (grants ZR2017BB056 and ZR2017QD013) are thanked for supporting this work.

Author Contributions

X.W., R.L., and W.C. performed the computational modeling/simulations and analyzed the computational data. Q.L. and J.Y. drafted the manuscript and prepared the figures and draft the main manuscript. J.Y. designed the study and finalized the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-25471-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018