Residue Effect-Guided Design: Engineering of *S. solfataricus* β-glycosidase to Enhance Its Thermostability and Bioproduction of Ginsenoside Compound K

Wenfeng Shen[1,2,3], Paul A Dalby[4], Zheng Guo[5], Weina Li[1,2,3*], Chenhui Zhu[1,2,3], Daidi Fan[1,2,3*]

[1]Engineering Research Center of Western Resource Innovation Medicine Green Manufacturing, Ministry of Education, School of Chemical Engineering, Northwest University, Xi'an, 710069, China.

[2]Shaanxi R&D Center of Biomaterials and Fermentation Engineering, School of Chemical Engineering, Northwest University, Xi'an, 710069, China.

[3]Biotech. & Biomed. Research Institute, Northwest University, Xi'an, 710069, China.

[4]Department of Biochemical Engineering, UCL, London WCIE 6BT, UK.

[5]Department of Biological and Chemical Engineering, Faculty of Science and Technology, Aarhus University, Gustav Wied Vej 10, Aarhus 8000, Denmark.

[*]Corresponding author

E-mail: 20154798@nwu.edu.cn (Weina Li); Daidi Fan: fandaidi@nwu.edu.cn (Daidi Fan)

ABSTRACT.

<i>β</i>-Glycosidase from *S. solfataricus* (SS-BGL), is a highly effective biocatalyst for the synthesis of Compound K (CK) from glycosylated protopanaxadiol ginsenosides. In order to improve the thermal stability of SS-BGL, MD simulations were used to determine the residue-level binding energetics of ginsenoside Rd in the SS-BGL-Rd docked complex, and to identify the top ten critical contributors. Target sites for mutations were determined using dynamic cross-correlation mapping of residues via the Ohm server to identify networks of distal residues that interacted with the key binding residues. Target mutations were determined rationally based on site characteristics. Single-mutants, and then recombination of top hits led to the two most promising variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D with 2.5-fold and 3.3-fold increased half-lives at 95℃, respectively. The enzyme activities relative to WT for ginsenoside conversion were 161% and 116%, respectively.

Keywords: glycosidase, MD simulation, thermal stability, compound K, bioconversion

INTRODUCTION.

SS-BGL is a <i>β</i>-glycosidase from the thermophilic archaea *S. solfataricus*, [1] and has been cloned and successfully expressed previously in *E. coli* [2] and *S. c*erevisiae[3]. Recombinant enzymes have been purified to homogeneity and exhibit the same structural and functional characteristics as natively expressed enzymes. [4-5]

Compound K (CK) is the principal deglycosylated metabolite of ginsenosides, exhibiting intriguing biological activities and holding potential as an anti-cancer, anti-diabetic, and anti-inflammatory agent[6-8]. However, CK has not been identified in ginsenosides naturally. Presently, it is primarily obtained through glycosyl hydrolysis of protopanaxadiol-type ginsenosides. In the current development of methods, biocatalysis shows significant promise because of its unique advantages, and the reconstruction of the biocatalytic system of ginsenosides has become a hot spot for research activity. For the efficient production of CK, a one-pot multi-enzyme-catalyzed strategy based on BG23 and BGA35 (<i>β</i>-galactosidase from *Aspergillus oryzae*) was established.[9] The conversion of ginsenoside Rb1 to CK by snailase (Sna) has also been demonstrated as an effective industrial production method, based on a new immobilization approach to improve the catalyst morphology and enzyme activity.[10]

SS-BGL is implicated as a highly suitable enzyme for the CK production process. In accordance with previous literature,[11] as many as 19 enzymes are known that prepare CK with ginsenoside Rb1 as the substrate, among which six <i>β</i>-glycosidases have been extracted and purified. Of these six enzymes, SS-BGL produced the highest levels of CK.[12] The use of thermophiles means that the temperature of the reaction system can be increased, not only for improved reaction rate, but also to reduce the risk of contamination, and to improve substrate solubility.

The structural modification of SS-BGL has been mainly driven through studies that aimed to clarify the structural characteristics of its thermal stability and heat resistance. A large number of studies have focused on the three-dimensional structure of SS-BGL, and specific structural features such as ion-pair networks and post-translational modifications to elucidate their roles in enzyme stabilization.[13-14] The key catalytic residues E387 and E206 within the active site of the enzyme were also identified through the retention mechanism of <i>α</i>-glycoside hydrolysis and site-specific mutagenesis.[15-17] Most importantly, studies on thermal stability indicate that SS-BGL becomes significantly inactivated at temperatures greater than 85°C.[18-19]

Immobilization is a common method used to enhance enzyme stability. The immobilization of SS-BGL with glutaraldehyde-activated chitosan not only improved the thermostability of the enzyme, but also removed the product-inhibition by glucose.[20] Similarly, the covalent ligation of insoluble chitosan to SS-BGL greatly improved the commercial hydrolysis of lactose.[21]

Approaches such as directed evolution, rational design, and computer-aided design, have proven to be effective strategies for engineering thermophilic enzymes. For example, a BlAsnase variant obtained by site-directed saturation mutagenesis improved its thermal stability by 65.8-fold at 55℃.[22] Similarly, a thermostable cellobiose CtCel6 from *T. thermophilus* was rationally engineered using site-directed mutagenesis based on structure, to give a variant with the $t_{1/2}$ increased by 1.42-fold and 2.40-fold at 80℃ and 90℃, respectively.[23] Rational and semi-rational design strategies have been used to identify mutations at five beneficial sites of <i>β</i>-glycosidase BglY of *T. thermophilus*, which were then recombined in variant HF5 to increase the half-life by 4.7-fold at 93℃.[24] Another approach has combined molecular dynamics simulation with computational prediction of mutations at flexible sites to increase the thermal stability of an antibody fragment[25], transketolase enzyme[26], and recently a thermophilic lipase.[27] This approach was recently extended to target networks of dynamically correlated residues identified from Dynamics Cross Correlation Maps of MD simulations, and increased the melting temperature of a transketolase variant by 3℃[28]. In the present study, initial candidate sites were selected based on the residue-level binding energy contributions in SS-BGL-Rd complexes and then expanded to networked distal residues with correlated dynamics as identified using MD simulations. CD spectroscopy and DSC were used to characterize the stability of the protein structure, and MD simulation was used to rationalize the heat resistance and catalytic mechanisms of SS-BGL at the molecular level.

MATERIALS AND METHODS

Strains, Plasmids, and Chemicals.

E. coli BL21 (DE3) and pET-24a were employed as host strains and expression plasmid vectors, respectively, and *E. coli* DH5α was used for plasmid amplification. p-Nitrophenyl-<i>β</i>-D-glucopyranoside (pNPG) was purchased from Solarbio Biotechnology Co., Ltd. (Beijing, China). Ginsenoside Rb1, CK standards were purchased from Xi'an Giant Biogene Technology Co., Ltd. (Xi'an, China). The site-directed mutation kit was purchased from YEASEN Biotechnology Co., Ltd. (Shanghai, China).

Culture, Expression, and Purification.

Recombinant *E. coli* strains were cultured in Luria-Bertani liquid medium containing 50 μg/mL kanamycin at 37℃ with 220 rpm agitation. When the OD_{600} reached 0.6, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the culture at a concentration of 1.0 mM, inducing enzyme expression. After induction, cells were cultured at 16℃ with shaking at 170 rpm for 16 h.

The cultured cells were harvested and resuspended in 50 mM citric acid/phosphate buffer (pH 5.5) and sonically lysed on ice for 10 min, then centrifugated at 4℃ at 5180×g for 10 min to extract the supernatant.

This crude extract was purified in 50 mM citric acid/phosphate buffer (pH 5.5) at 80℃ for 60 min. The heat-treated protein suspension was then centrifuged at 5180×g for 10 min, with the resulting supernatant filtered with a 0.45 μm sterile membrane to obtain purified protein as the filtrate. Protein concentration was tested using the BCA protein kit (Solarbio, Beijing, China). Protein molecular mass was determined using SDS-PAGE.

MD Simulation

Protein-ligand complex was simulated using AutoDock 4.2. Molecular dynamics (MD) simulations were performed using Amber 22 (San Francisco, CA, USA). The ff19SB force field[29] was employed for calculation of system force field parameters. Within the simulation, solvation was performed using the TIP3P water model and countering ions were added to neutralize the system. Once the system energy was minimized, the system was heated from 0 K to 400 K within 500 ps. System restriction was performed in the canonical ensemble, followed by system pre-balancing at 400 K. Finally, 1000 ns molecular simulations were performed in isothermal isobaric ensembles to maintain periodic boundary conditions. All covalent bonds involving hydrogen were limited by the SHAKE method. Root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) values were calculated. The ligand binding free-energy (ΔG) value was calculated using the MMPBSA.py tool in Amber 22. Residue energy decomposition was obtained by decomposing the solvent-accessible surface area of the protein onto each residue, with the overall binding energy split into the binding energy of each residue. Based on the free energy landscape (FEL) analysis derived from principal component analysis (PCA), Markov state models (MSM) involved in the long timescale molecular simulation were constructed using PyEMMA. [30] Each system was subjected to 5 rounds of simulation repeats to fully sample the dynamic conformation space of the system. The average values and errors of all output data were calculated.

Remote effect pathway exploration was achieved using the Ohm server (https://dokhlab.med.psu.edu/ohm/) which determines pathways of neighboring residues through which allosteric effects can be mediated. As a supplement, SPM (shortest path map) was constructed based on the average distance and dynamic cross-correlation matrix (DCCM) values calculated by MD simulation. The method was implemented in the newly developed *DynaComm.py* Python code,[31] and the structure diagram were drawn using PyMOL (http://www.pymol.org).

Site-Directed Mutagenesis.

The SS-BGL gene (NCBI accession number: WP_ 009992676) was commercially synthesized by Tsingke Biotechnology Co., Ltd. (Beijing, China) and optimized for codons in *E. coli* expression system. The synthesized gene was fused with the vector pET-24a and transferred to *E. coli* DH5α for plasmid amplification. Site-directed mutations were modeled on pET-24a vectors carrying the gene of interest for full sequence amplification. The primers were designed using Oligo 7.0 and synthesized by Sangong Bioengineering Co., Ltd. (Shanghai, China). The primer sequences are shown in Table S1. Sequence verification of positive transformants was completed by Sangong Bioengineering Co., Ltd. (Shanghai, China).

Enzyme Assays.

Enzyme activity was tracked through observing the release of p-nitrophenol from p-nitrophenyl-β-D-glucopyranoside (pNPG), as measured based on the quantity of enzyme required to release 1 μmol of p-nitrophenol per minute. 100 μL of enzyme solution was added to 600 μL of 4 mM pNPG solution and reacted at 85℃ for 10 min, after which 200 μL of 1 mM sodium carbonate solution was added to terminate the reaction. The quantity of p-nitrophenol released was obtained through testing the absorbance at 405 nm. All experiments were conducted in three parallel trials. Based on the results of these trials, the optimal temperature test range of the enzyme was set to 25-95℃, all experiments were carried out in three parallel trials.

Enzyme Stability Assays.

Enzyme thermostability was tested through measuring the residual activity of the enzyme after incubation set as 95℃ for different periods of time, with the initial enzyme activity at 100%.

Bioconversion of Rb1 to CK.

In the ginsenoside conversion process, a 1.0 mg/mL enzyme protein solution was prepared by dissolving it in a 50 mM citric acid/phosphate buffer and then added to a 5 mg/mL Rb1 substrate. The experimental setup included a standard sample group, a negative control group, and enzyme-catalyzed groups. The standard samples, which were methanol-soluble, were directly tested after filtration. The negative control group and enzyme-catalyzed groups underwent a reaction at

85℃ for 1 h, following which an equal volume of methanol was added to stop the reaction. The solution was subsequently sonicated for 30 minutes and filtered using a 0.45 μm sterile membrane. The generation of CK was monitored using high-performance liquid chromatography (HPLC) with the LC-2030C system (Shimadzu), which was equipped with an ultraviolet (UV) detector and a C18 reversed-phase column (4.6 mm × 150 mm, 5 μm, OMNI). The mobile phases consisting of water (A) and acetonitrile (B), underwent low-pressure gradient elution as described in Table S2. The column temperature was set at 35℃, the flow rate was maintained at 1.0 mL/min, and the injection volume was 10 μL, with UV detection at 203 nm. The productivity of the final product CK was calculated using the following equation:

$$CKproductivity=CK(mg)/V(L)\cdot t(h)$$

where V and t represent the reaction system volume and reaction time, respectively.

Further, the transformation products were identified using Ultimate 3000 UHPLC - Q Interactive LC-MS instrument (Thermo Scientific, US), with the HESI as the ion source, the sheath gas flow rate is 40 arb, the auxiliary gas rate is 10 arb, and the spray voltage is 3.2 kV negative ion.

Kinetic Parameter Analysis.

The kinetic parameters K_m and V_{max} at 85℃ were evaluated using the Michaelis-Menten Kinetics method. The activity of the purified enzyme was measured at pNPG, Rb1, Rd, and F2 concentration gradients of 0.01-20 mM. K_m and V_{max} values were then obtained by fitting to the

Michaelis-Menten equation using GraphPad Prism 8.0.2. The k_{cat} value was calculated using the molecular mass value of the enzyme, 60 KDa.

Differential Scanning Calorimetry (DSC).

The protein samples were dissolved in 50 mM citric acid/phosphate buffer (pH 5.5) and measured using a differential scanning calorimeter (Mettler DSC3). Measurements were taken at an initial temperature of 30℃, with temperature increasing at a rate of 5℃/min to a final temperature of 120℃ under nitrogen support.

CD Spectroscopy.

CD spectra for mutants and WT were measured using circular dichroic spectrometry (JASCO J-1500) at 25℃. Protein was dissolved in 5 mM citric acid/phosphate buffer (pH 5.5) at a concentration of 0.2 mg/mL. 190-250 nm’s spectra changes were then continuously recorded using a 1 mm path cell. Finally, data and maps from the recordings were processed and plotted using GraphPad Prism 8.0.2.

Statistical Analysis.

GraphPad Prism 8.0.2 software was used for statistical analysis. Results were presented as the mean ± standard deviation (SD) of the indicated number of replicates. Statistical significance was determined by one-way ANOVA with Dunnett’s multiple comparisons test for more groups. *P* values of <0.05 were considered statistically significant.

RESULTS AND DISCUSSION

Selection of Hotspots.

By breaking down the solvent-accessible surface area of the protein onto each residue, the overall binding energy is split into the binding energy of each residue. It is assumed that the energy value at the residue level is inversely correlated with the contribution of residues to enzyme catalysis; Therefore, residues with lower energy values contribute more to the catalytic process. In this way, the ten residues that contribute the most to the energy decomposition of SS-BGL-Rd complexes were screened out, and were listed in Table 1. Studies have shown that in SS-BGL, a strong ion-pair network composed of a large amount of arginine is the key to its thermophilic properties.[14] The proportion of amino acids in SS-BGL and the location of the top ten contributing residues were shown in Figure S1. The residues E387 and E206, located at the active center of SS-BGL, are the key sites for enzymatic catalysis.[22] The purpose of this study is to improve the thermal stability of proteins, and changes in the surface ion pair network, which play an important role in protein thermal stability, should be avoided during the engineering process. Therefore, the direction of site selection was shifted to residues that can produce synergistic effects with the above key sites. Dynamic residue crosscorrelation is often used to analyze synergistic effects between residues.[32] It has also recently been shown that residue networks can act over long distances at enzyme active sites[33] and are therefore useful targets for mutagenesis[28], so this strategy is also used here to find key target sites. Taking the above 12 key residues as the analysis objective, the residue horizontal dynamic cross-correlation mapping was used to determine other dynamically coupled residues in the SS-BGL-Rd composite structure. Ohm server is used to find and characterize allosteric communication networks in proteins, and unlike previously developed simulation-based methods, it's based solely on protein structure. Figure S2 depicts the analysis of the distal effect pathway by the ohmic server, and Figure 1B shows one of the pathways affecting E206, which is listed in the plot of the sites involved. To provide further support, we supplement this analysis with SPM calculations, as shown in Figure 1C. SPM can map coupled residue networks and provide more details about the nature of predicted interactions.[34] By summarizing the seven effect pathways determined by Ohm and the residual sites involved in SPM analysis, the dynamic cross-correlation domains at the level of six residues were finally identified.

As shown in Figure 1C, 16 sites were selected from within the six dynamic cross-correlation domains, for each of which target mutation determination was carried out in accordance with individual residue characteristics. Protein surface charge design typically involves the transformation of glutamine and asparagine into their corresponding acidic amino acids, a process that affects protein thermal stability and pH.[35-36] N90, Q96, N97, Q103, N113, N128, N302, N377, and N414 were nine residues located on the surface of the protein SS-BGL that were selected for mutagenesis into appropriate charged residues. As judged by RMSF, G221 and G301 were in a flexible loop region and so we introduced proline mutations at these sites based on the rigidifying flexible sites strategy (RFS). Given that the F222 site showed beneficial activity enhancement in previous studies,[37] mutations were introduced at similar residues in accordance with the literature. T386, N388, W425, and V205 are four sites of the catalytically active center that we targeted for mutations according to the nucleophilic center pKa design theory. This argues that positively charged histidines often appear near the active site of glycoside hydrolases and that changes in the pKa value of nucleophilic centers, such as after the introduction of basic amino acids, affect protein thermal stability and activity.[38] See Table S3 for details.

Screening and Combination of Superior Mutants.

Through screening of recombinant enzymes and high-temperature incubation of 16 variants, only Q96E, N97D, N128D, N302D, F222A were found to have even slightly greater residual activity after heating at 95℃ for 30 min (Figure 2A). Figure 2B shows the relative enzyme activity as tested with pNPG as the substrate. Most variants maintained a pNPG hydrolytic activity that was comparable to the WT, with activity abolished only in the N205H variant. As this site is adjacent to the catalytic residue E206, modifications so close to the active site center appears to be more difficult.

Combined variants often show superior properties over single point mutants.[39-40] To further enhance stability, we randomly combined the superior mutations into new variants. As shown in Figure 2C, the thermal stability experiment revealed that except for SS-BGL-N128D/N302D, whose stability was lower than that of WT, the stabilities of combined variants were superior to that of WT by varying degrees. While WT activity decreased to 33% after 120 min at 95℃, the SS-BGL-Q96E/N97D/N128D/N302D and SS-BGL-Q96E/N97D/N128D/F222A variants remained 54% and 48% active after 200 min, respectively.

From the results, the thermal stability experiments at 95℃ showed that the half-lives of the variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D were 2.8 h and 3.6 h, respectively, compared to only 1.1 hour for WT. Furthermore, these compare favourably against the best heat-stable variants of modified cellobiodisaccharide hydrolase from *Trichomonas thermophila*,[41] which had a half-life of 60 min at 90℃, and D-lyxose isomerase from *Hyperthermophilic archaea*,[42] whose activity decreased to 60% after 60 min of heat treatment at 80℃. In comparison to the current commonly used snail enzyme, which is the standard in industrial CK preparation but is unstable, SS-BGL holds great advantages in high-efficiency ginsenoside conversion, and make industrial applications more feasible.

Eight multiple-mutant variants with improved thermal stability were selected for subsequent ginsenoside conversion experiments.

Bioconversion of Rb1 to CK.

The biotransformation of ginsenosides was carried out directly at 85℃ and 95℃, respectively. To produce CK, SS-BGL hydrolyzes the glycosides at site 20 and site 3 of Rb1 successively, with a conversion path of Rb1→Rd→F2→CK, as shown in Figure 3A. The liquid chromatograms of the standard sample group, control group, and sample group are shown in Figure 3B. From the chromatogram, it can be seen that SS-BGL and its variants have Rd accumulation in the entire transformation pathway, while Rb1 and F2 are basically undetectable. The mass spectra of the transformation product CK are shown in Figure S3. Among the eight variants, only SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D showed superior ginsenoside converting activity in comparison to WT at both reaction temperatures. It is worth mentioning that due to the improved thermal stability of the variants, the advantage of variants transformation is highlighted at higher reaction temperatures. The relative enzyme activity at 85℃ was 102% and 109%, respectively, but increased to 161% and 116% at 95℃, as shown in Figure 3C, D. These results indicate that the thermal stability of the SS-BGL-Q96E/N97D/N128D/N302D variant is superior to that of the SS-BGL-Q96E/N97D/N302D variant, with the opposite being true for conversion activity. Such a trade-off is common in enzyme engineering, where increased thermal stability is often accompanied by a decrease in enzyme activity. Although some progress has been made to overcome this limitation, the counteraction mechanism remains obscure. [43] Fortunately, in our study, both thermostability-enhanced variants demonstrated superior conversion rates than WT. Table 2 lists in detail the CK production productivity and enzyme activity data corresponding to variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D at 95℃.

The CK productivity with the best variant at 95℃ reached up to 3727 mg·L^{-1}·h^{-1}. This was a significant improvement over other enymes such as the extracellular enzyme from *Aspergillus tubingensis* (418 mg · L^{-1}·h^{-1}), [44] and <i>β</i>-glucosidase from *Penicillium decumbens* (97 mg·L^{-1}·h^{-1}). [45] The improved conversion rate and strong adaptability to high-temperature production conditions will provide strong support for the industrial application of thermophiles for ginsenoside conversions.

Enzymatic characterization

SDS-PAGE of the expressed and purified WT and variant enzymes (Figure S4) confirm that the molecular mass at around 60 kDa with no significant effects of mutations on the apparent molecular mass.

The superior variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D, displaying increased thermal stability and ginsenoside conversion activity, were tested for activity over the temperature range of 25℃-95℃. As shown in Figure 2E, the combined mutation variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D both exerted maximum activity at 95℃ and above, which is a great improvement over the 85℃ optimal temperature found for WT. This indicates that the increased thermostability of the enzyme variants optimizes their catalytic conditions.

The thermal transition midpoint, T_m, as measured using DSC, represents the temperature at which 50% of protein is denatured. As shown in Table 3, the T_m values of the variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D were respectively 12.6℃ and 14.3℃ higher than for WT. This was consistent with the thermal stability results based on catalytic activity, indicating the superior stability of the variants relative to WT.

Kinetic Parameter Analysis.

For the ginsenoside Rd of the four substrates, the variant exhibited the lowest K_m (by 2x to 4x), but also up to 100x lower k_{cat} values than for pNPG (Table 4). From the perspective of catalytic efficiency, the low catalytic efficiency of variant for ginsenoside Rd is consistent with the rate-limiting step described in the literature.[37] However, it can be seen that the hydrolysis efficiency of the variants SS-BGL-Q96E/N97D/N302D and SS-BGL-Q96E/N97D/N128D/N302D may have improved slightly for ginsenoside Rd compared with WT.

The enzyme kinetic parameters for three types of ginsenoside involved in the hydrolysis pathway showed that catalytic efficiency may have had a small contribution to the increased biotransformation efficiency of the variants, but that this was mostly due to the higher retention of enzyme activity with the elevated reaction temperature.

Secondary Structure Analysis

Through previous studies of thermophiles, protein surface charge, hydrophobic action, hydrogen bonding and disulfide bonding have been recognized as key factors influencing thermostability.[46] To investigate the impact of protein structure on thermostabilisation far-UV CD analysis was performed to estimate the secondary structures of WT and variant SS-BGL-Q96E/N97D/N302D (Figure S5). The CD profiles (Table 5) indicated that the mutations in the variant led to slight increases in structure such as α-helices and decreases in disordered structure. The increased proportion of α-helix and beta-sheet structures in the protein is consistent with forming more intramolecular interactions, such as hydrogen bonding and hydrophobic interactions, which are important factors influencing the stability of protein structures, consistent with studies by others. [47]

Molecular Dynamics (MD) Simulation.

Comparison of structural stability of proteins and docking complexes

In the SS-BGL-catalysed conversion of Rb1 to CK, Rd to F2 is the rate-limiting step. Based on this, the Rd molecule was docked into both the WT and variant SS-BGL-Q96E/N97D/N302D for MD simulations and compared to the undocked proteins. In the absence of Rd, the variant showed higher fluctuations than the WT at around residues 30-55 and 330 -355 indicating that the mutation increased the flexibility of the structure in these two regions (Figure 4E). The docking complexes all exhibited low RMSD values (Figure 4A, B). The definition of protein secondary structure (DSSP) is a standard algorithm for the classification of amino acid residues in protein structures. Figures 4C and 4D show the relationship between the secondary structural transformation of the pure protein throughout the simulation, and the relationship between the secondary structural transformation of the complexes, as shown in Figure S6. Based on DSSP analysis, significant differences in α-helix structure content were identified between variant SS-BGL-Q96E/N97D/N302D and WT, which was consistent with the results of CD spectroscopic analysis. Moreover, as simulation time increased, protein stability gradually decreased, and the protein expansion was manifested by decreases in the secondary structure and increases in random structure, although variant SS-BGL-Q96E/N97D/N302D still showed higher stability through retention of secondary structure than WT.

Analysis of Substrate-Protein Binding

Table 1 shows the results of residue energy decomposition on the complex SS-BGL-Q96E/N97D/N302D-Rd. It was found that most of the key residues in the mutant protein SS-BGL-Q96E/N97D/N302D-Rd were consistent with WT, and the same sites after mutation modification exhibited lowest energy values, compared with the original protein. This suggests that the mutations are more conducive to the functioning of key residues, consistent with the largely unaffected kinetic parameters. Solvent-accessible surface area (SASA) can be used to determine how tightly a protein binds to small molecule ligands and is a key factor in evaluating protein folding and hydrophobicity. As shown in Figure 5A, SASA continued to decline throughout the simulation cycle, with the SASA of the variant SS-BGL-Q96E/N97D/N302D-Rd remaining lower than that of SS-BGL-Rd. Given these results, we see that the mutations limit the interaction of the protein with water molecules of its surrounding microenvironment, improving overall hydrophobicity. As such, the mutant protein more tightly wraps Rd, creating a stronger affinity in SS-BGL-96E/N97D/N302D-Rd. The same conclusion can be drawn from the number of hydrogen bonds in the complex (Figure 5B, C). The number of intermolecular hydrogen bonds in the SS-BGL-Q96E/N97D/N302D-Rd complex increased at the beginning of the simulation, then stabilized at around 11 after 200 ns. In comparison, the SS-BGL-Rd complex system ended the simulation at about 4 intermolecular hydrogen bonds, so much lower than that of the variant complex. These results support the intramolecular interaction forces, overall hydrophobicity, and ligand binding force of proteins as the key factors that improve protein stability and catalytic activity.

Computational analyses of the dominant conformation and molecular interaction

The introduction of mutations in the variants generally led to the redistribution of native protein conformations with the emergence of new conformers and a change in the rate of transition between different conformations, which can also affect protein function.[48] Based on PCA, combined with FEL plots, MSM analysis revealed the transition between the main conformations of the SS-BGL-Rd and SS-BGL-Q96E/N97D/N302D-Rd.

PCA and FEL diagrams comparing the movement and conformational differences of biological macromolecules are plotted in Figures 6A, B, and C. In the FEL plot, the best conformation shows the lowest energy value. In the Q96E/N97D/N302D-Rd system, State 3 is the most stable and the dominant conformation, with the largest proportion of the simulation time, 38.2%. Figure 6D depicts MSM analysis of the transition relationship between the four principal conformations. Rates of transition from conformation 1→3 and 2→3 were significantly faster than those of other conformations. In the SS-BGL-Rd system, S5 is the most stable state, as shown by MSM analysis (Figure S7E). S4→S5, S3→S5 are the main conformational transitions.

The binding free energy analysis of the dominant conformation S5 of WT-Rd and S3 of SS-BGL-Q96E/N97D/N302D-Rd is shown in Table S4, and it is obvious that the bonding stability of SS-BGL-Q96E/N97D/N302D to the small molecule Rd is better than that of WT and Rd, which is consistent with the above results. The interaction study of complexes in the dominant conformations, as shown in Figure 6E, F, results indicate that the main interaction force between protein and Rd is hydrogen bonding. In S5, the three hydrogen bonds between Rd and L337, V216, N205 residues make small molecules malleable. In contrast, in S3, Rd folds inward under the synergistic effect of the five hydrogen bonds formed between it and the W32, W150, H149, F358 residues. This increases the contact area between the small molecule and the active center, making it easier for the reaction to proceed. The analysis of hydrogen bonding forces between substrate and protein in the main conformation showed that mutations led to changes in the binding morphology of substrate Rd and protein, which was the result of the synergistic effect of multiple residues via hydrogen bonding networks, which is also a factor affecting the catalytic efficiency of enzymes.

Abbreviations

SS-BGL, <i>β</i>-glycosidase from the thermophilic archaea *S. solfataricus*; CK, ginsenoside compound K; Rb1, ginsenoside Rb1; Rd, ginsenoside Rd; IPTG, isopropyl-β-thiolactone; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; circular dichroic; DSC, differential scanning calorimetry; pNPG, p-nitrophenyl-β-D-glucopyranoside; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; DSSP, definition of protein secondary structure; SASA, Solvent-accessible surface area; FEL, free energy landscape; PCA, protein principal analysis; MSM, Markov State models; HPLC, high-performance liquid chromatography.

Acknowledgments

This work was supported by the National Key R&D Program of China (2021YFC2101500) and the National Natural Science Foundation of China (22108229). We confirm that we have read and have comply with the statement of ethical standard for manuscripts submitted to *Journal of Agricultural and Food Chemistry*.

Supporting Information

Primers used for cloning and site-directed mutagenesis (Table S1), HPLC detection conditions (Table S2), candidate mutation sites (Table S3), the docking complexes bind free energy by MD simulation analysis (Table S4), identification of the top ten contributing residues in SS-BGL (Figure S1), SS-BGL-RD distal effect path analysis (Figure S2), mass spectrometry of the transformation product CK at 85 °C (Figure S3), SDS-PAGE purification analyses of WT and its nine multiple mutants (Figure S4), protein secondary structure analysis from 190 to 250nm (Figure S5), protein complex DSSP (Figure S6), MSM analysis based on PCA (Figure S7).

Conflicts of interest

The authors declare no competing interests.

References

1. de Rosa, M.; Gambacorta, A.; Bu'lock, J. D., Extremely thermophilic acidophilic bacteria convergent with Sulfolobus acidocaldarius. *Journal of general microbiology* **1975,** *86* (1), 156-64.

2. Moracci, M.; Nucci, R.; Febbraio, F.; Vaccaro, C.; Vespa, N.; La Cara, F.; Rossi, M., Expression and extensive characterization of a beta-glycosidase from the extreme thermoacidophilic archaeon Sulfolobus solfataricus in Escherichia coli: authenticity of the recombinant enzyme. *Enzyme and microbial technology* **1995,** *17* (11), 992-7.

3. Moracci, M.; La Volpe, A.; Pulitzer, J. F.; Rossi, M.; Ciaramella, M., Expression of the thermostable beta-galactosidase gene from the archaebacterium Sulfolobus solfataricus in

Saccharomyces cerevisiae and characterization of a new inducible promoter for heterologous expression. *Journal of bacteriology* **1992,** *174* (3), 873-82.

4. Cubellis, M. V.; Rozzo, C.; Montecucchi, P.; Rossi, M., Isolation and sequencing of a new beta-galactosidase-encoding archaebacterial gene. *Gene* **1990,** *94* (1), 89-94.

5. Pisani, F. M.; Rella, R.; Raia, C. A.; Rozzo, C.; Nucci, R.; Gambacorta, A.; De Rosa, M.; Rossi, M., Thermostable beta-galactosidase from the archaebacterium Sulfolobus solfataricus. *Purification and properties. European journal of biochemistry* **1990,** *187* (2), 321-8.

6. Liu, T.; Zhu, L.; Wang, L., A narrative review of the pharmacology of ginsenoside compound K. *Annals of translational medicine* **2022,** *10* (4), 234.

7. Cho, S. H.; Chung, K. S.; Choi, J. H.; Kim, D. H.; Lee, K. T., Compound K, a metabolite of ginseng saponin, induces apoptosis via caspase-8-dependent pathway in HL-60 human leukemia cells. *BMC cancer* **2009,** *9*, 449.

8. Han, G. C.; Ko, S. K.; Sung, J. H.; Chung, S. H., Compound K enhances insulin secretion with beneficial metabolic effects in db/db mice. *J Agric Food Chem* **2007,** *55* (26), 10641-8.

9. Yang, W.; Zhou, J.; Gu, Q.; Harindintwali, J. D.; Yu, X.; Liu, X., Combinatorial Enzymatic Catalysis for Bioproduction of Ginsenoside Compound K. *J Agric Food Chem* **2023**.

10. Tian, D.; Wu, Z.; Liu, X.; Tu, Z.; Li, R.; Fan, D.; Lan, Y., Synthesis of L-aspartic acid-based bimetallic hybrid nanoflowers to immobilize snailase for the production of rare ginsenoside

compound K. *Journal of materials chemistry. B* **2023,** *11* (11), 2397-2408.

11. Park, S. Y.; Bae, E. A.; Sung, J. H.; Lee, S. K.; Kim, D. H., Purification and characterization of ginsenoside Rb1-metabolizing beta-glucosidase from Fusobacterium K-60, a human intestinal anaerobic bacterium. *Bioscience, biotechnology, and biochemistry* **2001,** *65* (5), 1163-9.

12. Noh, K. H.; Son, J. W.; Kim, H. J.; Oh, D. K., Ginsenoside compound K production from ginseng root extract by a thermostable beta-glycosidase from Sulfolobus solfataricus. *Bioscience, biotechnology, and biochemistry* **2009,** *73* (2), 316-21.

13. D'Auria, S.; Barone, R.; Rossi, M.; Nucci, R.; Barone, G.; Fessas, D.; Bertoli, E.; Tanfani, F., Effects of temperature and SDS on the structure of beta-glycosidase from the thermophilic archaeon Sulfolobus solfataricus. *The Biochemical journal* **1997,** *323* (Pt 3) (Pt 3), 833-40.

14. Cobucci-Ponzano, B.; Moracci, M.; Di Lauro, B.; Ciaramella, M.; D'Avino, R.; Rossi, M., Ionic network at the C-terminus of the beta-glycosidase from the hyperthermophilic archaeon Sulfolobus solfataricus: Functional role in the quaternary structure thermal stabilization. *Proteins* **2002,** *48* (1), 98-106.

15. Moracci, M.; Capalbo, L.; Ciaramella, M.; Rossi, M., Identification of two glutamic acid residues essential for catalysis in the beta-glycosidase from the thermoacidophilic archaeon Sulfolobus solfataricus. *Protein engineering* **1996,** *9* (12), 1191-5.

16. Febbraio, F.; Barone, R.; D'Auria, S.; Rossi, M.; Nucci, R.; Piccialli, G.; De Napoli, L.; Orrù, S.;

Pucci, P., Identification of the active site nucleophile in the thermostable beta-glycosidase from the archaeon Sulfolobus solfataricus expressed in Escherichia coli. *Biochemistry* **1997,** *36* (11), 3068-75.

17. Moracci, M.; Trincone, A.; Perugino, G.; Ciaramella, M.; Rossi, M., Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from Sulfolobus solfataricus: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* **1998,** *37* (49), 17262-70.

18. Moracci, M.; Ciaramella, M.; Rossi, M., Beta-glycosidase from Sulfolobus solfataricus. *Methods in enzymology* **2001,** *330,* 201-15.

19. Ausili, A.; Cobucci-Ponzano, B.; Di Lauro, B.; D'Avino, R.; Perugino, G.; Bertoli, E.; Scirè, A.; Rossi, M.; Tanfani, F.; Moracci, M., A comparative infrared spectroscopic study of glycoside hydrolases from extremophilic archaea revealed different molecular mechanisms of adaptation to high temperatures. *Proteins* **2007,** *67* (4), 991-1001.

20. D'Auria, S.; Pellino, F.; La Cara, F.; Barone, R.; Rossi, M.; Nucci, R., Immobilization on chitosan of a thermophilic beta-glycosidase expressed in Saccharomyces cerevisiae. *Applied biochemistry and biotechnology* **1996,** *61* (1-2), 157-66.

21. Petzelbauer, I.; Kuhn, B.; Splechtna, B.; Kulbe, K. D.; Nidetzky, B., Development of an ultrahigh-temperature process for the enzymatic hydrolysis of lactose. IV. Immobilization of two

thermostable beta-glycosidases and optimization of a packed-bed reactor for lactose conversion. *Biotechnology and bioengineering* **2002,** *77* (6), 619-31.

22. Chi, H.; Wang, Y.; Xia, B.; Zhou, Y.; Lu, Z.; Lu, F.; Zhu, P., Enhanced Thermostability and Molecular Insights for l-Asparaginase from Bacillus licheniformis via Structure- and Computation-Based Rational Design. *Journal of Agricultural and Food Chemistry* **2022,** *70* (45), 14499-14509.

23. Han, C.; Li, W.; Hua, C.; Sun, F.; Bi, P.; Wang, Q., Enhancement of catalytic activity and thermostability of a thermostable cellobiohydrolase from Chaetomium thermophilum by site-directed mutagenesis. *International Journal of Biological Macromolecules* **2018,** *116*, 691-697.

24. Yi, Z.-L.; Zhang, S.-B.; Pei, X.-Q.; Wu, Z.-L., Design of mutants for enhanced thermostability of <i>β</i>-glycosidase BglY from Thermus thermophilus. *Bioresource Technology* **2013,** *129*, 629-633.

25. Zhang, C.; Samad, M.; Yu, H.; Chakroun, N.; Hilton, D.; Dalby, P. A., Computational Design To Reduce Conformational Flexibility and Aggregation Rates of an Antibody Fab Fragment. *Molecular pharmaceutics* **2018**, *15* (8), 3079-3092.

26. Yu, H.; Yan, Y.; Zhang, C.; Dalby, P. A., Two strategies to engineer flexible loops for improved enzyme thermostability. *Scientific reports* **2017**, *7*, 41212.

27. Qu, P.; Li, D.; Lazim, R.; Xu, R.; Xiao, D.; Wang, F.; Li, X.; Zhang, Y., Improved

thermostability of Thermomyces lanuginosus lipase by molecular dynamics simulation and in silico mutation prediction and its application in biodiesel production. *Fuel* **2022,** *327,* 125039.

28. Yu, H.; Dalby, P. A., Exploiting correlated molecular-dynamics networks to counteract enzyme activity-stability trade-off. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115* (52), E12192-e12200.

29. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of chemical theory and computation* **2020,** *16* (1), 528-552.

30. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F., PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of chemical theory and computation* **2015,** *11* (11), 5525-42.

31. Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S., Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catalysis* **2017**, 7 (12), 8524-8532.

32. Sucharski, F.; Gallo, G.; Coelho, C.; Hardy, L.; Würtele, M., Modeling the role of charged residues in thermophilic proteins by rotamer and dynamic cross correlation analysis. *Journal of molecular modeling* **2023,** *29* (5), 132.

33. Yu H.; Dalby PA., Coupled molecular dynamics mediate long and short-range epistasis between mutations that affect stability and aggregation kinetics. P*roc. Natl. Acad. Sci* **2018**,115 (47) E11043-E11052.

34. Crnjar, A.; Griñen, A.; Kamerlin, S. C. L.; Ramírez-Sarmiento, C. A., Conformational Selection of a Tryptophan Side Chain Drives the Generalized Increase in Activity of PET Hydrolases through a Ser/Ile Double Mutation. *ACS Organic & Inorganic Au* **2023**, 3 (2), 109-119.

35. Wu, X.; Zhang, Q.; Zhang, L.; Liu, S.; Chen, G.; Zhang, H.; Wang, L., Insights Into the Role of Exposed Surface Charged Residues in the Alkali-Tolerance of GH11 Xylanase. *Frontiers in microbiology* **2020,** *11*, 872.

36. Wang, X.; Du, J.; Zhao, B.; Wang, H.; Rao, S.; Du, G.; Zhou, J.; Chen, J.; Liu, S., Significantly Improving the Thermostability and Catalytic Efficiency of Streptomyces mobaraenesis Transglutaminase through Combined Rational Design. *J Agric Food Chem* **2021,** *69* (50), 15268-15278.

37. Shin, K. C.; Choi, H. Y.; Seo, M. J.; Oh, D. K., Improved conversion of ginsenoside Rb(1) to compound K by semi-rational design of Sulfolobus solfataricus <i>β</i>-glycosidase. *AMB Express* **2017,** *7* (1), 186.

38. Xie, T.; Zhou, L.; Han, L.; Cui, W.; Liu, Z.; Cheng, Z.; Guo, J.; Zhou, Z., Modulating the pH profile of the pullulanase from Pyrococcus yayanosii CH1 by synergistically engineering the active

center and surface. *Int J Biol Macromol* **2022,** *216*, 132-139.

39. Chen, C.; Guan, B.-H.; Geng, Q.; Zheng, Y.-C.; Chen, Q.; Pan, J.; Xu, J.-H., Enhanced Thermostability of Candida Ketoreductase by Computation-Based Cross-Regional Combinatorial Mutagenesis. *ACS Catalysis* **2023,** *13* (11), 7407-7416.

40. Rao, D.; Huo, R.; Yan, Z.; Guo, Z.; Liu, W.; Lu, M.; Luo, H.; Tao, X.; Yang, W.; Su, L.; Chen, S.; Wang, L.; Wu, J., Multiple approaches of loop region modification for thermostability improvement of 4,6-<i>α</i>-glucanotransferase from Limosilactobacillus fermentum NCC 3057. *International Journal of Biological Macromolecules* **2023,** *233*, 123536.

41. Han, C.; Li, W.; Hua, C.; Sun, F.; Bi, P.; Wang, Q., Enhancement of catalytic activity and thermostability of a thermostable cellobiohydrolase from Chaetomium thermophilum by site-directed mutagenesis. *Int J Biol Macromol* **2018**, *116*, 691-697.

42. De Rose, S. A.; Kuprat, T.; Isupov, M. N.; Reinhardt, A.; Schönheit, P.; Littlechild, J. A., Biochemical and Structural Characterisation of a Novel D-Lyxose Isomerase From the Hyperthermophilic Archaeon Thermofilum sp. *Frontiers in Bioengineering and Biotechnology* **2021,** *9*.

43. Luo, J.; Song, C.; Cui, W.; Han, L.; Zhou, Z., Counteraction of stability-activity trade-off of Nattokinase through flexible region shifting. *Food Chemistry* **2023,** *423*, 136241.

44. Kim, S. A.; Jeong, E. B.; Oh, D. K., Complete Bioconversion of Protopanaxadiol-Type

Ginsenosides to Compound K by Extracellular Enzymes from the Isolated Strain Aspergillus tubingensis. *J Agric Food Chem* **2021**, *69* (1), 315-324.

45. Kim, S. Y.; Lee, H. N.; Hong, S. J.; Kang, H. J.; Cho, J. Y.; Kim, D.; Ameer, K.; Kim, Y. M., Enhanced biotransformation of the minor ginsenosides in red ginseng extract by Penicillium decumbens <i>β</i>-glucosidase. *Enzyme and microbial technology* **2022**, *153*, 109941.

46. Han, H.; Ling, Z.; Khan, A.; Virk, A. K.; Kulshrestha, S.; Li, X., Improvements of thermophilic enzymes: From genetic modifications to applications. *Bioresource Technology* **2019,** *279*, 350-361.

47. Momeni, L.; Shareghi, B.; Saboury, A. A.; Farhadian, S.; Reisi, F., A spectroscopic and thermal stability study on the interaction between putrescine and bovine trypsin. *Int J Biol Macromol* **2017**, *94* (Pt A), 145-153.

48. Óskarsson, K. R.; Nygaard, M.; Ellertsson, B.; Thorbjarnardottir, S. H.; Papaleo, E.; Kristjánsson, M. M., A single mutation Gln142Lys doubles the catalytic activity of VPR, a cold adapted subtilisin-like serine proteinase. *Biochimica et biophysica acta* **2016**, *1864* (10), 1436-43.

Figure captions

Figure 1. Mutation site selection. (A) SS-BGL-Rd complex, constructed from molecular docking. (B) One of the E206 effect pathways. (C) Representation of the shortest path map (SPM) along the evolutionary pathway. (D) Sixteen residue sites selected from residue dynamic cross-correlation. (E) Residue dynamic cross-correlation domain.

Figure 2. Screening and enzymatic activity characterization of WT and variants. (A) Residual enzyme activity incubated at 95℃ for 30 min. (B） Relative enzymatic activity of variants with pNPG as substrate. (C) Thermal stability curve of multiple variants at 95℃. (D) Relative enzymatic activity of multiple variants with pNPG as substrate. (E) Optimal temperature of WT and its variants. Data are presented as mean ± SD (n=5). *$P<0.05$, **$P<0.01$ and ***$P<0.001$ were considered statistically significant vs the WT group, and n.s. indicated no significant difference from the WT group.

Figure 3. Biotransformation of ginsenosides. (A) SS-BGL conversion of Rb1 to CK pathway. (B) HPLC diagrams of the production process of CK. (C) Relative activity of WT and its variant SS-BGL-Q96E/N97D/N302D in converting ginsenosides at 85℃. (D) Relative activity of WT and its variant SS-BGL-Q96E/N97D/N302D in converting ginsenosides at 95℃. Data are presented as mean ± SD (n=5). *$P<0.05$, **$P<0.01$ and ***$P<0.001$ were considered statistically significant vs the WT group, and n.s. indicated no significant difference from the WT group.

Figure 4. RMSD, DSSP, and RMSF analysis of SS-BGL protein and the docked complex. (A) RMSD analysis of WT alone and in complex with Rd. (B) RMSD analysis of SS-BGL-Q96E/N97D/N302D alone and in complex with Rd. (C) DSSP analysis of undocked WT enzyme. (D) DSSP analysis of undocked SS-BGL-Q96E/N97D/N302D. (E) RMSF analysis of WT and SS-BGL-Q96E/N97D/N302D.

Figure 5. SASA and Hydrogen bond analysis of complexes. (A) SASA diagram. (B) SS-BGL-Rd complex hydrogen bond analysis. (C) Hydrogen bond analysis of SS-BGL-Q96E/N97D/N302D-Rd complex.

Figure 6. MSM analysis. (A) FEL diagram of WT pure protein. (B) FEL diagram of SS-BGL-Rd complex. (C) FEL diagram of SS-BGL-Q96E/N97D/N302D-Rd complex. (D) The transformation relationship between the main conformations of the SS-BGL-Q96E/N97D/N302D-Rd complex, analyzed using MSM. Blue represents the α-helix, purple represents the β-lamellae, pink is the flexible ring, and yellow is ginsenoside Rd. (E) Analysis of the interaction forces between Rd and SS-BGL in S5. (F) Analysis of the interaction force between Rd and SS-BGL-Q96E/N97D/N302D in S3, the black circle indicates the intramolecular force of Rd.

Table 1. Ten Critical Protein-ligand Complexation Sites Determined by Residue Energy Decomposition

SS-BGL-Rd		SS-BGL-Q96E/N97D/N302D-Rd	
sites	total energy	sites	total energy
ARG86	-182.2713	ARG86	-188.2713
ARG131	-176.9231	ARG131	-183.0630
ARG185	-178.3453	ARG185	-187.8100
ARG245	-177.6951	ARG440	-182.8310
ARG307	-178.5204	ARG307	-184.0050
ARG324	-175.8521	ARG411	-186.1270
ARG378	-175.1823	ARG378	-183.2030
ARG420	-176.9958	ARG420	-183.5200
ARG140	-177.3881	ARG140	-183.2360
ARG313	-175.3274	ARG234	-183.2360

Table 2. Bioconversion Activity of WT and its Variants at 95℃ .

Variants	CK productivity ($mg\cdot L^{-1}\cdot h^{-1}$)	Enzyme activity (IU)	Enzyme specific activity (IU/mg)
WT	1995	1.07	1.12
SS-BGL-Q96E/N97D/N302D	3727	1.99	1.73
SS-BGL-Q96E/N97D/N128D/N302D	2540	1.37	1.20

Table 3. T_m and $t_{1/2}$ at 95℃ of WT and WT Variants.

Variants	T_m(℃)	△T_m(℃)	$t_{1/2}$ (h) at 95℃
WT	89.6		1.1
SS-BGL-Q96E/N97D/N302D	102.2	12.6	2.8
SS-BGL-Q96E/N97D/N128D/N302D	103.9	14.3	3.6

Table 4. Kinetic Parameters of WT and WT Variants.

Variants	Substrates	K_m(μM)	k_{cat}(min^{-1})	k_{cat}/K_m(min^{-1}·μM^{-1})
WT	pNPG	500±180	21060±660	42
	Rb1	750±480	380±80	0.5
	Rd	140±90	150±20	1.1
	F2	24800±13100	8570±300	0.4
SS-BGL-Q96E/N97D /N302D	pNPG	490±170	21840±660	45
	Rb1	570±420	330±90	0.6
	Rd	235±53	160±70	0.7
	F2	19400±5900	8600±760	0.4
SS-BGL-Q96E/N97D /N128D/N302D	pNPG	440±150	22700±600	52
	Rb1	1610±630	1010±1660	0.6
	Rd	250±220	270±90	1.1
	F2	7700±2900	6200±4000	0.8

Table 5. Secondary Structural Elements of WT and SS-BGL-Q96E/N97D/N302D.

Enzyme	α-helix(%)	β-sheet(%)	turn(%)	unordered (%)
WT	27.2	24.3	22.0	26.6
SS-BGL-Q96E/N97D/N302D	29.3	25.3	20.1	25.5

Figure 1

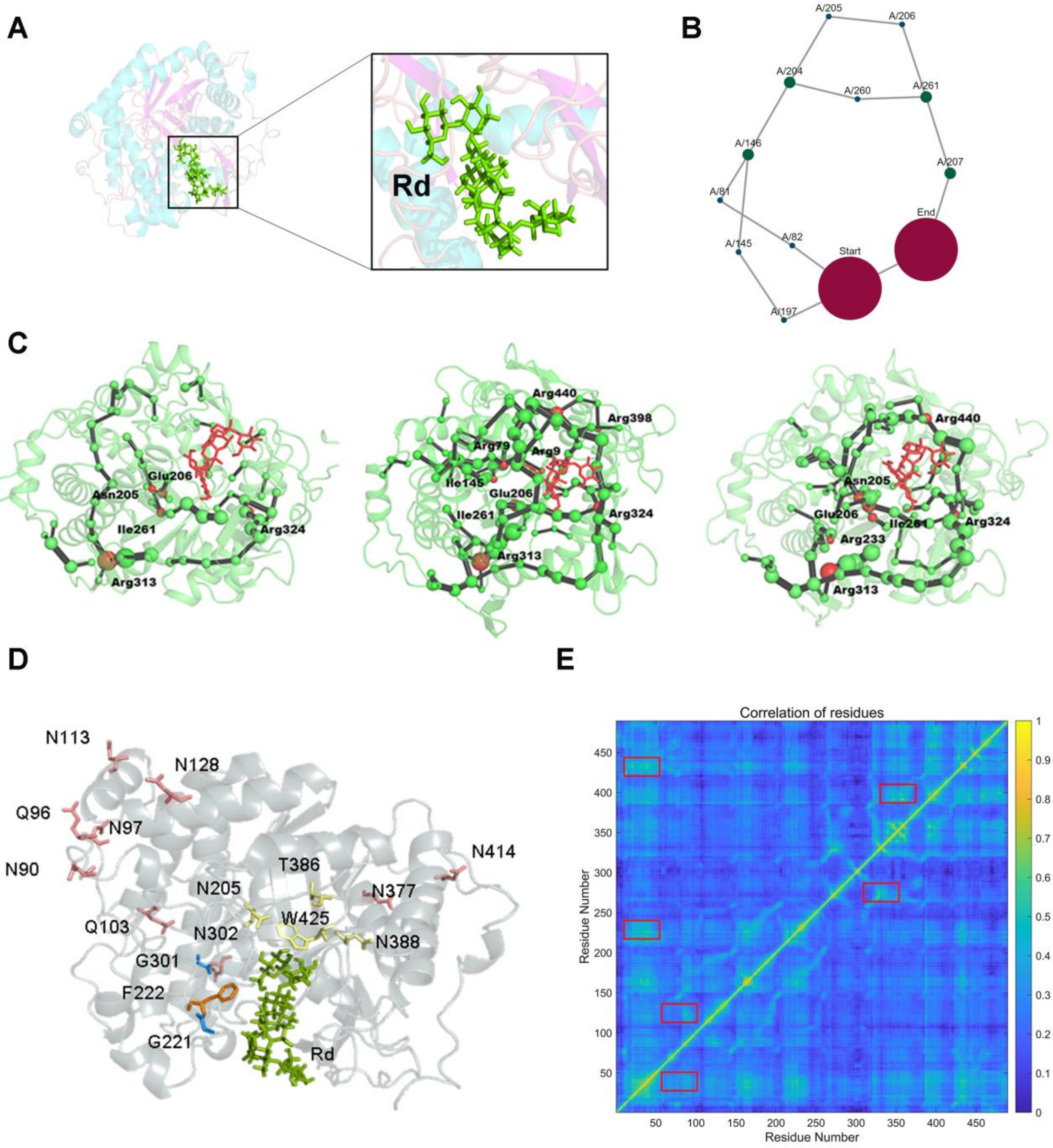

A
Rd
B
A/205
A/206
A/204
A/260
A/261
A/146
A/207
A/81
End
A/145
A/82
Start
A/197
C
Glu206
Asn205
Ile261
Arg324
Arg313
Arg440
Arg398
Arg79
Arg9
Ile145
Glu206
Ile261
Arg324
Arg313
Arg440
Asn205
Glu206
Ile261
Arg324
Arg233
Arg313
D
N113
N128
Q96
N97
N90
N414
T386
N205
N377
W425
Q103
N302
N388
G301
F222
G221
Rd
E
Correlation of residues
Residue Number
Residue Number
450
400
350
300
250
200
150
100
50
50
100
150
200
250
300
350
400
450
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Figure 2

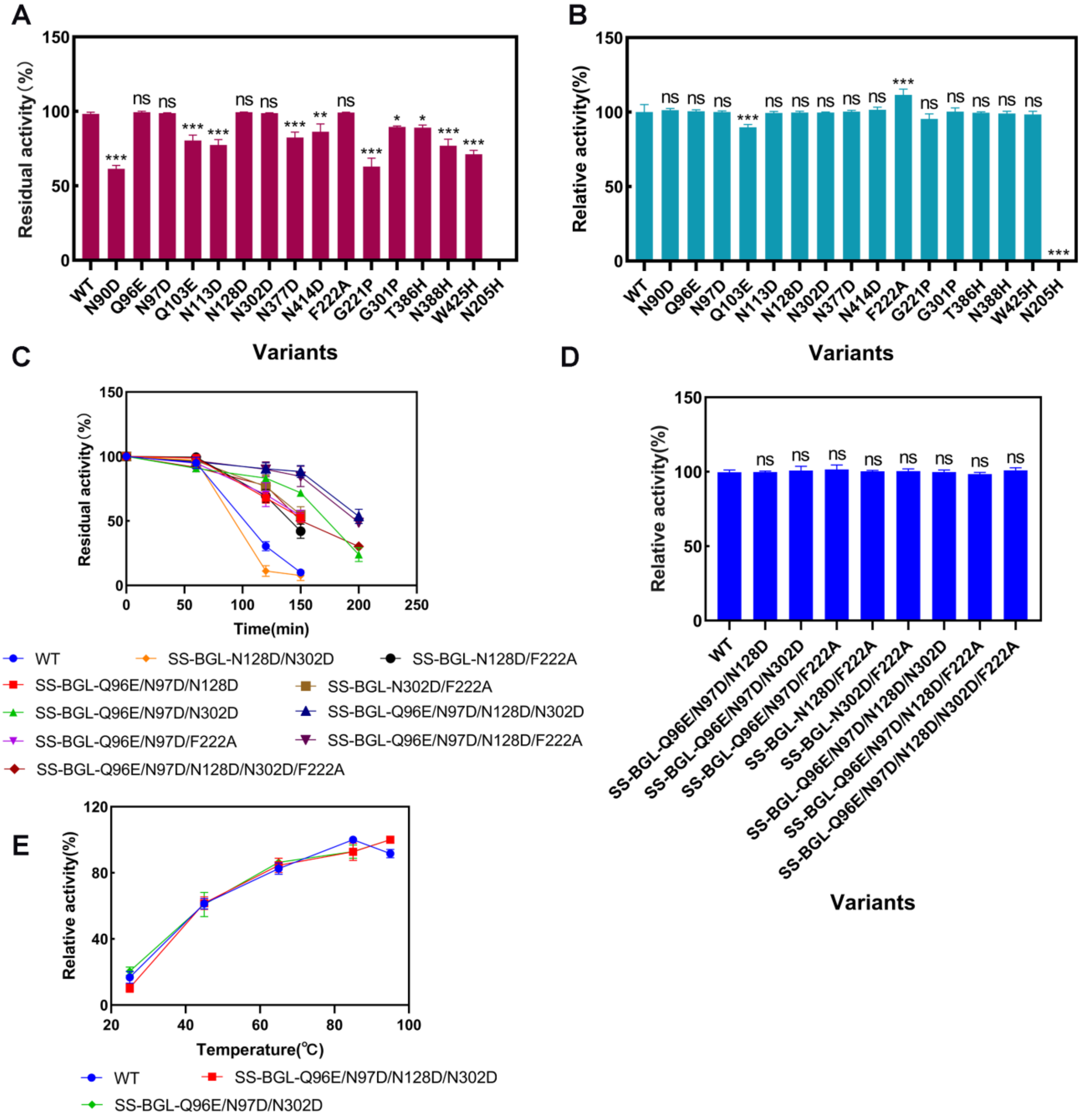

A
150
100
50
0
Residual activity (%)
ns
ns
***
***
***
ns
ns
***
**
ns
***
*
*
***
***
WT
N90D
Q96E
N97D
Q103E
N113D
N128D
N302D
N377D
N414D
F222A
G221P
G301P
T386H
N388H
W425H
N205H
Variants
B
150
100
50
0
Relative activity(%)
ns
ns
ns
***
ns
ns
ns
ns
ns
***
ns
ns
ns
ns
ns
***
WT
N90D
Q96E
N97D
Q103E
N113D
N128D
N302D
N377D
N414D
F222A
G221P
G301P
T386H
N388H
W425H
N205H
Variants
C
150
100
50
0
Residual activity (%)
0
50
100
150
200
250
Time(min)
WT
SS-BGL-N128D/N302D
SS-BGL-N128D/F222A
SS-BGL-Q96E/N97D/N128D
SS-BGL-N302D/F222A
SS-BGL-Q96E/N97D/N302D
SS-BGL-Q96E/N97D/N128D/N302D
SS-BGL-Q96E/N97D/F222A
SS-BGL-Q96E/N97D/N128D/F222A
SS-BGL-Q96E/N97D/N128D/N302D/F222A
D
150
100
50
0
Relative activity(%)
ns
ns
ns
ns
ns
ns
ns
ns
WT
SS-BGL-Q96E/N97D/N128D
SS-BGL-Q96E/N97D/N302D
SS-BGL-Q96E/N97D/F222A
SS-BGL-N128D/F222A
SS-BGL-N302D/F222A
SS-BGL-Q96E/N97D/N128D/N302D
SS-BGL-Q96E/N97D/N128D/F222A
SS-BGL-Q96E/N97D/N128D/N302D/F222A
Variants
E
120
80
40
0
Relative activity(%)
20
40
60
80
100
Temperature(℃)
WT
SS-BGL-Q96E/N97D/N128D/N302D
SS-BGL-Q96E/N97D/N302D

Figure 3

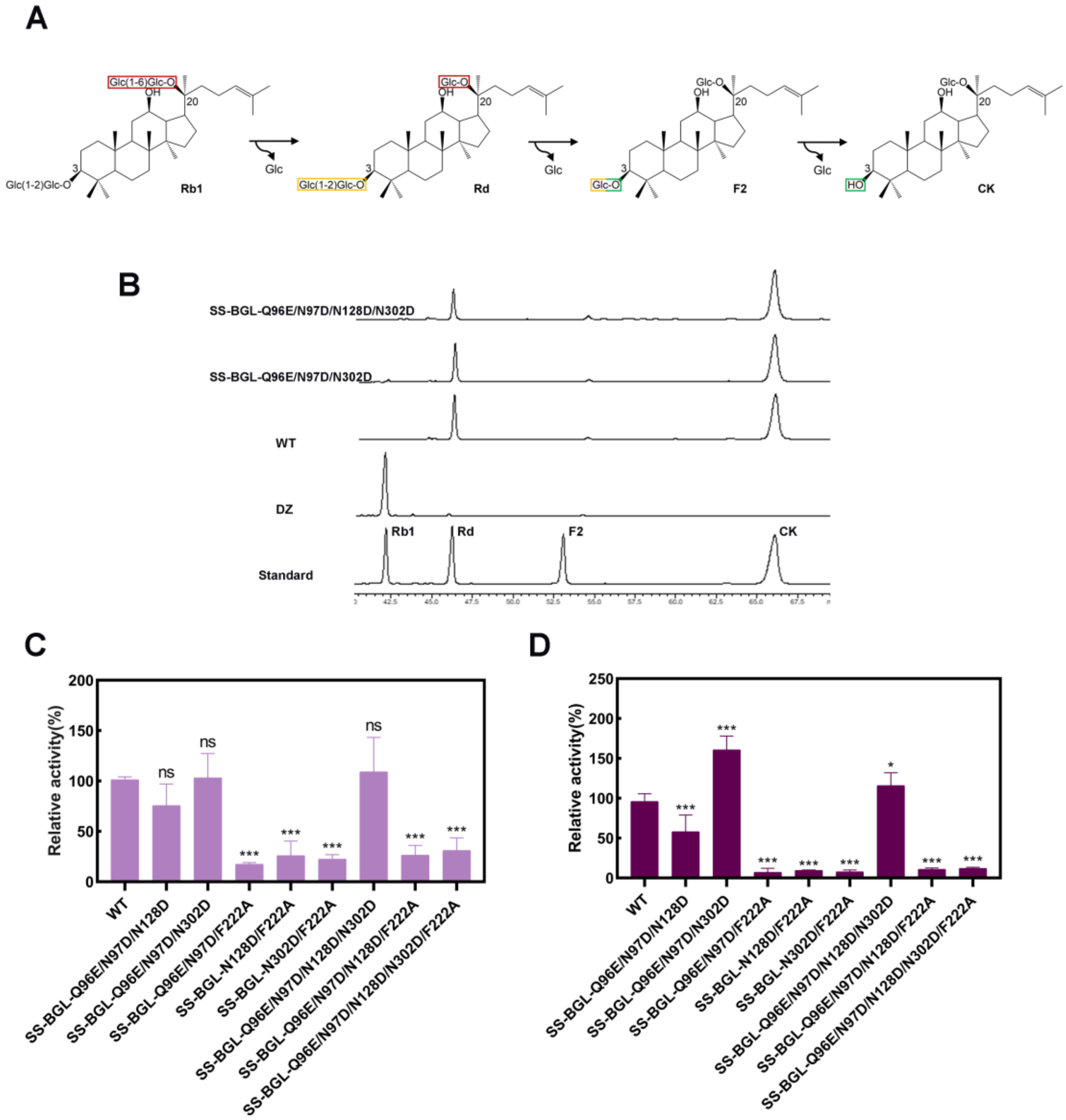

A
Glc(1-6)Glc-O
OH
20
3
Glc(1-2)Glc-O
Rb1
Glc
Glc-O
Glc(1-2)Glc-O
Rd
Glc
Glc-O
Glc-O
F2
Glc
HO
CK
B
SS-BGL-Q96E/N97D/N128D/N302D
SS-BGL-Q96E/N97D/N302D
WT
DZ
Standard
Rb1
Rd
F2
CK
42.5
45.0
47.5
50.0
52.5
55.0
57.5
60.0
62.5
65.0
67.5
C
Relative activity(%)
200
150
100
50
0
ns
ns
ns
***
***
***
***
***
WT
SS-BGL-Q96E/N97D/N128D
SS-BGL-Q96E/N97D/N302D
SS-BGL-Q96E/N97D/F222A
SS-BGL-N128D/F222A
SS-BGL-N302D/F222A
SS-BGL-Q96E/N97D/N128D/N302D
SS-BGL-Q96E/N97D/N128D/F222A
SS-BGL-Q96E/N97D/N128D/N302D/F222A
D
Relative activity(%)
250
200
150
100
50
0
***
***
***
***
***
*
***
***
WT
SS-BGL-Q96E/N97D/N128D
SS-BGL-Q96E/N97D/N302D
SS-BGL-Q96E/N97D/F222A
SS-BGL-N128D/F222A
SS-BGL-N302D/F222A
SS-BGL-Q96E/N97D/N128D/N302D
SS-BGL-Q96E/N97D/N128D/F222A
SS-BGL-Q96E/N97D/N128D/N302D/F222A

Figure 4

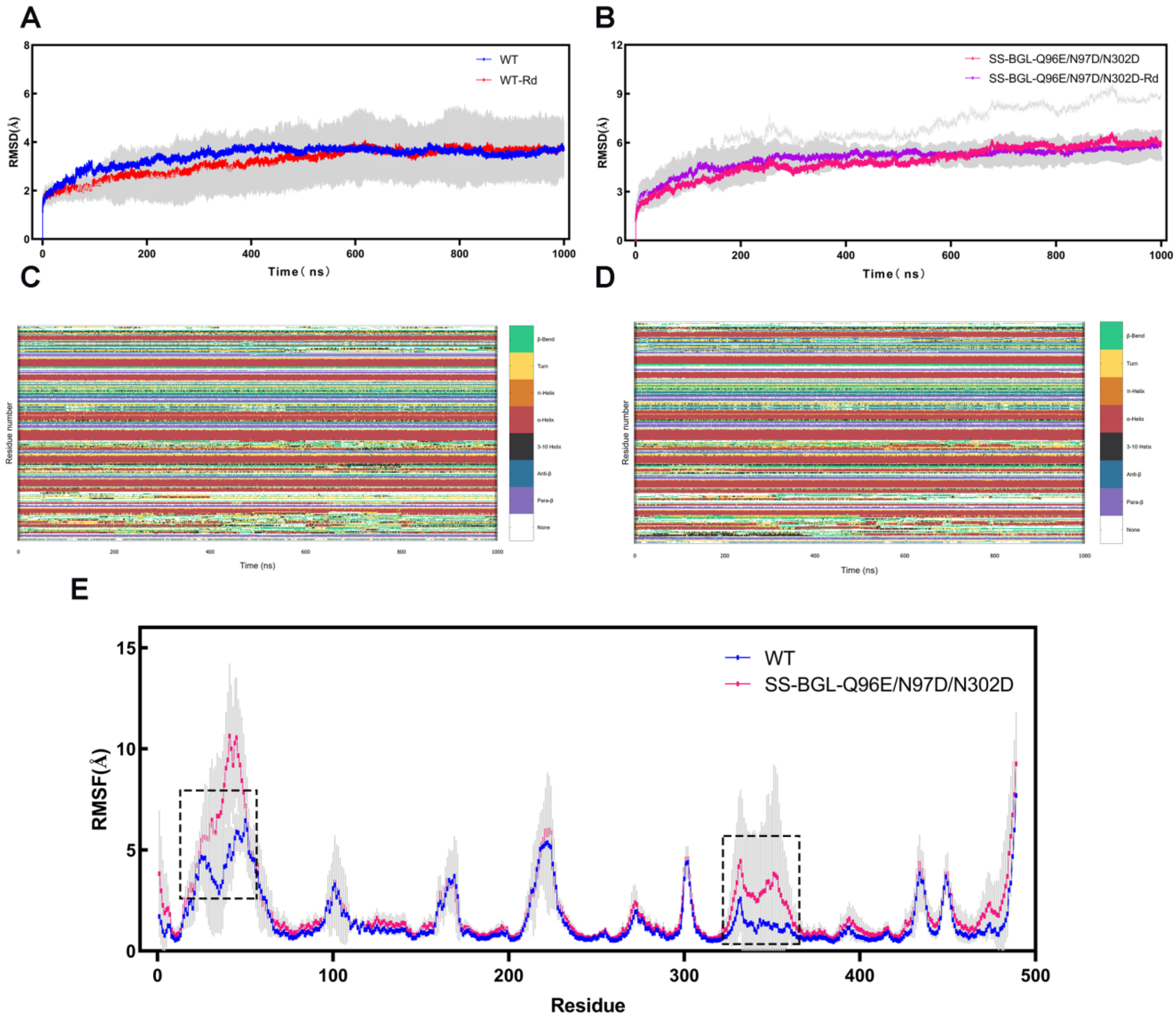

A
RMSD(Å)
0
2
4
6
8
0
200
400
600
800
1000
Time（ns）
WT
WT-Rd
B
RMSD(Å)
0
3
6
9
12
0
200
400
600
800
1000
Time（ns）
SS-BGL-Q96E/N97D/N302D
SS-BGL-Q96E/N97D/N302D-Rd
C
Residue number
0
200
400
600
800
1000
Time (ns)
β-Bend
Turn
π-Helix
α-Helix
3-10 Helix
Anti-β
Para-β
None
D
Residue number
0
200
400
600
800
1000
Time (ns)
β-Bend
Turn
π-Helix
α-Helix
3-10 Helix
Anti-β
Para-β
None
E
RMSF(Å)
0
5
10
15
0
100
200
300
400
500
Residue
WT
SS-BGL-Q96E/N97D/N302D

Figure 5

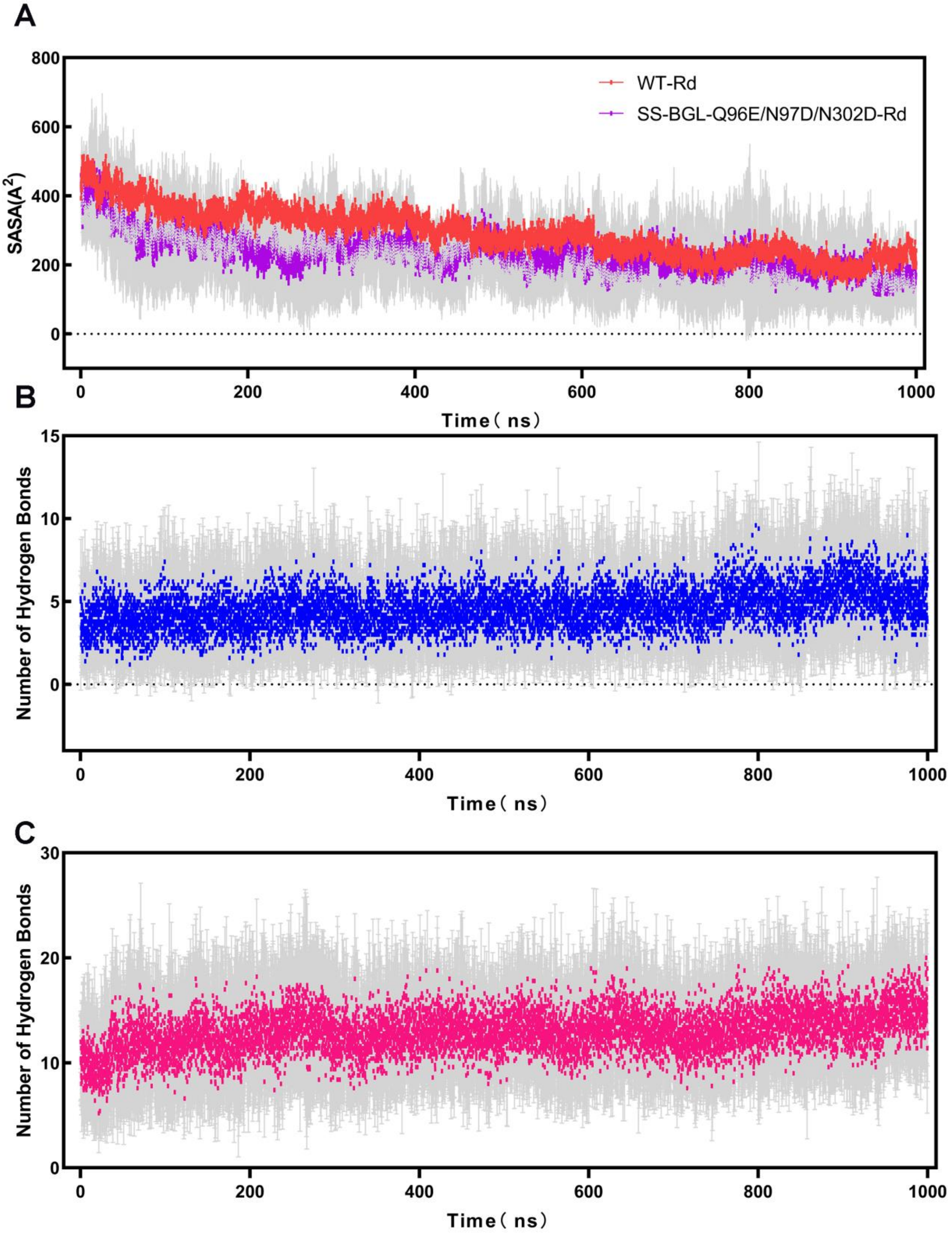
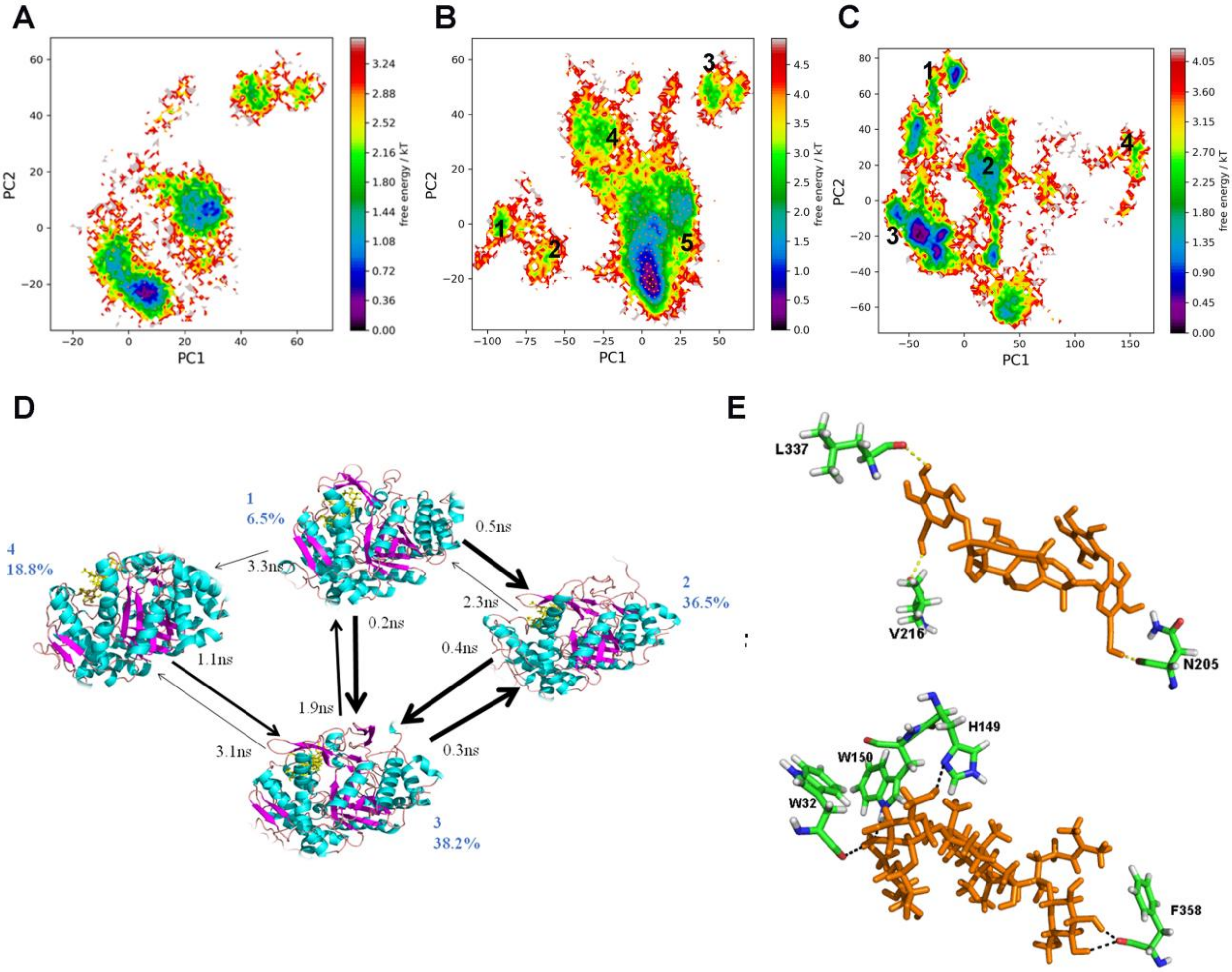
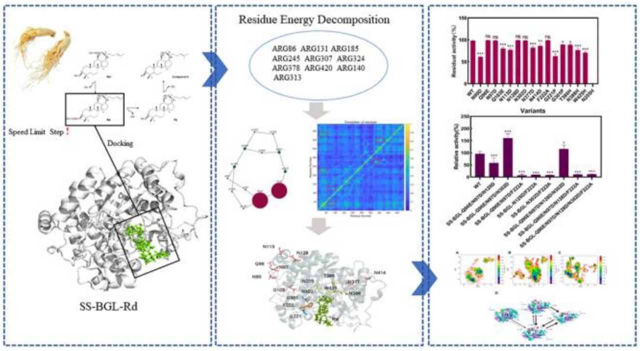

A
800
600
400
200
0
SASA(A^2)
WT-Rd
SS-BGL-Q96E/N97D/N302D-Rd
0
200
400
600
800
1000
Time（ns）
B
15
10
5
0
Number of Hydrogen Bonds
0
200
400
600
800
1000
Time（ns）
C
30
20
10
0
Number of Hydrogen Bonds
0
200
400
600
800
1000
Time（ns）

Figure 6


A
PC1
PC2
free energy / kT
B
1
2
3
4
5
C
1
2
3
4
D
1
6.5%
2
36.5%
3
38.2%
4
18.8%
0.5ns
3.3ns
2.3ns
0.2ns
0.4ns
1.1ns
1.9ns
3.1ns
0.3ns
E
L337
V216
N205
H149
W150
W32
F358

TOC graphic


Speed Limit Step !
Docking
SS-BGL-Rd
Residue Energy Decomposition
ARG86 ARG131 ARG185
ARG245 ARG307 ARG324
ARG378 ARG420 ARG140
ARG313
N113
N128
G96
N97
N90
N225
T386
N377
N414
W425
N388
G103
N302
G301
F222
G221
Rd
Residual activity(%)
Relative activity(%)
Variants
WT