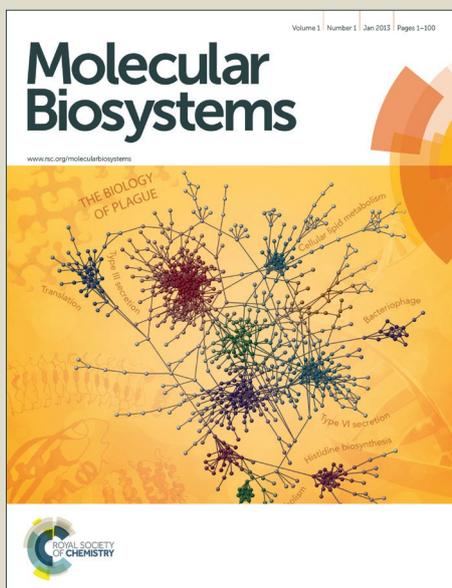


# Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

**Structural and Dynamic Basis of Acid Amido Synthetase GH3.1: An Investigation of Substrate Selectivity and Major Active Site Access Channels**

Xue Wang<sup>a</sup>, Juan Du<sup>a\*</sup>, Xiaojun Yao<sup>b</sup>

<sup>a</sup> Key Lab of Applied Mycology, College of Life Science, Qingdao Agricultural University, Qingdao 266109, China

<sup>b</sup> College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 730000, China

CORRESPONDING AUTHOR FOOTNOTE

\* To whom correspondence should be addressed.

Email: dujuannx@126.com

**Abstract**

Auxin/IAA (Indole-3-acetic acid) plays critical roles in many aspects of plant growth and development. Gretchen Hagen 3.1 (GH3.1) enzyme from grapevine (*Vitis vinifera*) catalyzes the ATP-dependent conjugation of aspartate to IAA suggested a significant way to modulate levels of cellular auxins/IAA. It is reported that VvGH3.1 prefers IAA as substrate than 1-naphthaleneacetic acid (NAA) and benzothiazole-2-oxyacetic acid (BTOA), whereas the detailed interaction mechanism of these substrates remains unclear. In this study, based on the recently reported crystal structure of VvGH3.1 and AtGH3.12, the open form of VvGH3.1 was built. Then combined computational techniques including molecular docking, molecular dynamics (MD) simulations, Molecular Mechanics Generalized Born Surface Area and Poisson–Boltzmann Surface Area (MM-GB/PBSA) methods and active site access channel analysis were utilized to investigate binding mechanism of three substrates, IAA, NAA and BTOA, and dynamic behaviors of GH3.1 induced by substrate binding. The predicted binding free energy is in agreement with the experimental work with an order of IAA < NAA < BTOA. Key residues interacting with three substrates and residues specifically binding to one of them were identified. The strong hydrogen bond interaction formed between IAA and Ser108, Ser339, Gln560 makes it the best substrate over NAA and BTOA. The dynamic behaviors, especially in the dominant access channels of IAA and NAA bound GH3.1 were found to be different from those in BTOA-GH3.1 system with larger bottleneck radius and shorter length. This study provided novel insight to understand the substrate selectivity mechanism of GH3.1 and also for the rational design novel auxin-based herbicides and growth regulators.

## Introduction

Auxins, such as indole-3-acetic acid (IAA), are a class of essential plant hormones which modulate diverse growth and development processes, including elongation and division of cells, tropic responses, vascular development, response to biotic and abiotic stimuli, lateral branching of shoots and roots and fruit development<sup>1-3</sup>. Cellular auxin/IAA levels and their activity can be altered by various ways. The increasing evidence suggests that conjugation reactions play an important role in these regulatory processes<sup>4-6</sup>. Members of Gretchen Hagen 3 (GH3) family can catalyze amino acid conjugating to excess IAA to maintain auxin homeostasis<sup>7</sup>. GH3 enzymes belong to adenylating firefly luciferase (ANL) enzyme superfamily<sup>8</sup>, terms as acyl acid-amido synthetases enzyme found in many plants species. Based the sequence similarities and substrate specificities, GH3 family is classified into 3 groups, group I adenylates jasmonic acid (JA), group II adenylates IAA and salicylic acid (SA) and group III adenylates benzoates, respectively<sup>6,9</sup>. Several GH3 enzymes can join IAA to different amino acids to conduct further reaction. For example, IAA-Asp and IAA-Glu conjugates can be oxidative decomposed. IAA-Ala and IAA-Leu conjugates tend to be converted into hydrolyzed metabolic to release free IAA<sup>10-12</sup>. The catalytic reaction includes two steps: adenylation and transferase reaction<sup>7,9,13-17</sup>, in which ATP binds first followed by IAA and generation of an adenylated IAA intermediate, subsequently reacting with an amino acid to yield the IAA-Amino acids conjugates<sup>16-18</sup>.

Gretchen Hagen 3.1 (GH3.1) enzyme from grapevine (*Vitis vinifera*) catalyzes the ATP-dependent conjugation of aspartate to IAA suggested a significant way to modulate levels of cellular auxins/IAA, thus modulating the initiation of grape berry ripening<sup>2,19</sup>. It is reported that VvGH3.1 prefers IAA as substrate over the synthetic auxin analogues NAA and BTOA<sup>19</sup>, whereas the detailed interaction mechanism between these substrates with the GH3.1 enzyme is still not clear<sup>20</sup>. The crystal structures of two *Arabidopsis* proteins (AtGH3.12/PBS3 and AtGH3.11/JAR1) provide new insights to understand how these enzymes modify plant hormones and

reflect two major conformations of these enzymes during the reaction process, open and close form<sup>21</sup>. Another crystal structure of GH3.1 from grapevine reveals a different hydrophobic site to accommodate the auxins<sup>22</sup>. But this structure was determined with an inhibitor (adenosine-5'-[2-(1*H*-indol-3-yl)ethyl]phosphate [AIEP]) that mimics the adenylated intermediate of the IAA conjugation reaction, thus in a close form. Zubieta *et al.* subsequently reported another crystal structure of AtGH3.12 in the close form in complex with the ATP analogue AMPCPP and salicylate, which indicated the flexibility of these enzymes during the reaction<sup>23</sup>. To sum up, the open conformation is crucial for the initial ATP and hormones binding. After that, the ATP and hormones are positioned in a properly productive binding model. Subsequently, the adenylation reaction is performed and the enzyme adopts a close conformation for the second half-reaction<sup>21-23</sup>. To better understand the substrate binding mechanism, the open form was built based on the structure of VvGH3.1 and AtGH3.12. The GH3.1 structure contains two domains (Fig. 1): a large N-terminal domain includes a barrel and two sheets flanked by helices and a small C-terminal domain consists of a single four-stranded sheet sandwiched by two helices on each side. The active site is located at the interface of the two domains, linked by a flexible hinge loop between  $\beta$ 16 and  $\alpha$ 16, which is supposed to pivot the C-terminal domain during the two-step catalytic reaction.

Davies *et al.* also performed molecular docking to explain the substrate selectivity mechanism of VvGH3.1 based on the close form<sup>22</sup>, which provide limited information for substrate binding and reacting. In view of the two-step reaction mechanism and the flexibility of GH3 enzymes, investigation of the structural and dynamic basis of the open form of VvGH3.1 and binding modes with substrates would provide more interesting information to understand the function of this enzyme. In the present study, the open state conformation of VvGH3.1 was constructed based on the crystal structure of VvGH3.1 and AtGH3.12. The detailed binding interaction mechanism between three substrates: IAA, NAA, BTOA and GH3.1 were investigated by molecular docking, molecular dynamics (MD) simulations and Molecular Mechanics Generalized Born/Poisson–Boltzmann Surface Area (MM-GB/PBSA) methods.

Moreover, active site access channel analysis was performed to identify the possible pathways that are considered to be essential for substrate ingress and egress from the active site to the surface of the enzyme. Our results clearly identified critical residues responsible for substrate binding and the major access channels for substrate access. We expect that our study can be useful to understand the selectivity mechanism of GH3.1 and also for the rational design novel auxin-based herbicides and growth regulators.

## Materials and Methods

### Construction of simulation systems

The crystal structure of an inhibitor (AIEP) bound to the VvGH3.1 (PDB ID: 4B2G) was obtained from the RCSB Protein Data Bank<sup>24, 25</sup>. Since the open state of VvGH3.1 was not available, we used the crystal structure of AMP-CPP (ATP analog) in the AtGH3.12 from the Protein Data Bank (PDB ID: 4EWV) as a reference structure to build the open state of VvGH3.1. The chain A of 4EWV was chosen as the template (protein sequence identity = 49.37%). The SWISS MODEL server<sup>26-29</sup> was used for building of the C-terminal domain of VvGH3.1 (residue 435-586), which was incorporated to the N-terminal domain of 4B2G (residue 16-434). Then the structure was refined to eliminate the bad contact.

### Molecular docking

To investigate the binding interaction between substrates and the open state conformation of the GH3.1 enzyme, molecular docking was carried out using Autodock vina program<sup>30</sup>. The 3D structures of IAA, NAA and BTOA were sketched and geometry optimization was performed by Discovery Studio<sup>31</sup>. AutoDock tools (ADT) was then used to prepare the ligand and protein (deleting all water molecules, adding polar hydrogens and assign atoms charges). The grid center was defined by the position of AIEP in the crystal structure of 4B2G. The searching space size was set to 10 Å. The exhaustiveness of the global search was set to 1000. The maximum energy difference between the best binding mode and the worst one was set to 5 kcal/mol.

The poses were selected according to the position of AIEP in 4B2G. Three complexes, GH3.1-IAA, GH3.1-NAA and GH3.1-BTOA were obtained.

### Molecular Dynamics Simulation

On the basis of the molecular docking, five systems, including the GH3.1-Apo, GH3.1-ATP, GH3.1-ATP-IAA, GH3.1-ATP-NAA and GH3.1-ATP-BTOA were subjected to MD simulations. The force field parameters for the three substrates were generated with the use of Antechamber program from Amber10 package, using General Amber Force Field (GAFF)<sup>32</sup> and restrained electrostatic potential (RESP) partial charges<sup>33</sup>. Geometric optimization and the electrostatic potential calculations were performed at the HF/6-31G\* level of Gaussian09 program<sup>34</sup>. Standard AMBER ff03 force field<sup>35, 36</sup> was assigned to the protein and General Amber Force Field (GAFF) was assigned to the ligands. The force field parameters for Mg<sup>2+</sup> ion and ATP were download from Amber parameter database<sup>37, 38</sup>. The protonation state of ionizable residues was set at the default value for PH 7. Na<sup>+</sup> ions were added to neutralize the overall system. Each system was embedded in a rectangular box of the TIP3P water molecules<sup>39</sup>, keeping 10 Å distance away from any solute atom to the boundary.

Before the MD simulations, molecular mechanics optimizations were employed to relax the system using the sander module in AMBER10 by three steps: first, the water molecules and ions were relaxed by restraining the protein with a harmonic restraints of 5 kcal/mol/Å<sup>2</sup> (5000 cycles of steepest descent and 5000 cycles of conjugate gradient minimizations); second, the side chains of the protein were relaxed by restraining the backbone of the protein (5000 cycles of steepest descent and 5000 cycles of conjugate gradient minimizations); third, the whole system was relaxed without any restraint (5000 cycles of steepest descent and 5000 cycles of conjugate gradient minimizations). After the minimization, each system was gradually heated from 0 to 300 K within 100 ps with the backbone of proteins restrained (5 kcal/mol/Å<sup>2</sup>) in the *NVT* ensemble. Then, the systems were relaxed within 1.05 ns from 5 to 0 kcal/mol·Å<sup>2</sup> in the *NPT* ensemble. Finally, a total of 60 ns were simulated to produce

trajectories. The covalent bonds to hydrogen atoms were constrained using SHAKE algorithm<sup>40</sup> and the Particle Mesh Ewald (PME) method<sup>41</sup> was employed to calculate long-range electrostatic interactions. The real space cutoff was set at 10.0 Å, the same as for van der Waals interactions. The grid-spacing and convergence criteria of PME calculation was set to 1 Å and 0.100E-04 respectively. The time step used for the simulations was set to 2 fs. The atom coordinates were saved every 10 ps for subsequent analysis.

### Binding free energy Calculation

The binding free energies between substrates (IAA, NAA and BTOA) and GH3.1 enzyme were analyzed by the MM-GB/PBSA methods integrated in the Amber10 package. For each system, 1000 snapshots were collected from the last 10 ns of trajectory, equally spaced at 10 ps intervals. The binding free energy between the receptor and ligand was calculated as follows:

$$\Delta G_{\text{binding}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} = \Delta G_{\text{gas}} + \Delta G_{\text{solv}} \quad (1)$$

$$\Delta G_{\text{gas}} = \Delta H_{\text{gas}} - T\Delta S \approx \Delta E_{\text{MM}} - T\Delta S \quad (2)$$

$$\Delta G_{\text{binding}} \approx \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S \quad (3)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{vdW}} + \Delta E_{\text{cle}} \quad (4)$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{GB/PB}} + \Delta G_{\text{SA}} \quad (5)$$

$$\Delta G_{\text{SA}} = \gamma * \Delta A + \beta \quad (6)$$

where  $G_{\text{complex}}$ ,  $G_{\text{receptor}}$ ,  $G_{\text{ligand}}$  are the free energies of complex, receptor, and ligand, respectively. According to MM-PBSA and MM-GBSA theory,  $\Delta G_{\text{binding}}$ , the binding free energy is composed of two parts, the gas phase molecular mechanical (MM) energy ( $\Delta G_{\text{gas}}$ ) and the solvation free energy ( $\Delta G_{\text{solv}}$ ).  $\Delta G_{\text{binding}}$  can be decomposed

into three terms: molecular mechanical energy term ( $\Delta E_{\text{MM}}$ ), solvation energy term ( $\Delta G_{\text{sol}}$ ), and vibrational entropy term ( $T\Delta S$ ).  $\Delta E_{\text{MM}}$  is given as a sum of the changes of  $\Delta E_{\text{int}}$ ,  $\Delta E_{\text{ele}}$ , and  $\Delta E_{\text{vdw}}$ , which are internal energy contribution from bonds, angles and torsions, electrostatic, and van der Waals interaction terms, respectively. By using a single trajectory strategy,  $\Delta E_{\text{int}}$  is canceled between ligand, receptor, and complex, which can significantly reduce the noise in most cases. The change of solvation energy ( $\Delta G_{\text{sol}}$ ), contains the polar part ( $\Delta G_{\text{GB/PB}}$ ) and the nonpolar part of the desolvation energy ( $\Delta G_{\text{SA}}$ ). The polar solvation contribution ( $\Delta G_{\text{GB/PB}}$ ) can be calculated by Poisson–Boltzmann (PB) and Generalized Born (GB) equation. Dielectric constants for solute and solvent were set to 1 and 80, respectively. The nonpolar part of the desolvation energy ( $\Delta G_{\text{SA}}$ ) can be estimated as eqn (6), where  $\Delta A$  represents the change of the solvent accessible surface area (SASA) of the system calculated by the LCPO algorithm<sup>42</sup>, and the fitting coefficients  $\gamma$  and  $\beta$  were set to  $0.005 \text{ kcal/mol} \cdot \text{\AA}^2$  and 0, respectively. The term ( $T\Delta S$ ) of eqn (2) is the change in the conformational entropy upon ligand binding. Here, normal-mode analysis (NMA) was used for the calculation of the conformational entropy, and as the high computational demand, only 100 snapshots extracted from the last 10 ns of the conventional MD trajectories were used.

The calculation error bars are standard errors (SE) that calculated using eqn (7), the STD is standard deviation and N is the number of trajectory snapshots used in the calculation.

$$\text{SE} = \frac{\text{STD}}{\sqrt{N}} \quad (7)$$

### Analysis of Access Tunnels

The program CAVER (version 3.0)<sup>43</sup> was employed to identify the possible tunnels that are considered to be essential for substrate ingress and egress from the active site

to the surface of the protein. The location of the substrate was chosen to be the starting point for tunnel searching. 100 snapshots of each MD simulation trajectory were extracted from the last 10 ns at an interval 100 ps. The probe radius and the clustering threshold were set to 1.0 and 3.5, respectively. Other parameters were set as default throughout the calculations. The tunnels were then visualized using PyMOL<sup>44</sup>.

## Results and Discussion

### Dynamics of GH3.1 Complex with and without ATP and substrates

It is suggested that the C-terminal domain of GH3 enzymes is flexible and its motion makes the enzyme shift from close to open state<sup>21</sup>. In addition, the open form is essential for the first half-reaction—adenylation, while the binding of ATP and substrates simultaneously can lock this conformation<sup>21</sup>. To better understand the substrate binding mechanism, the open form was built based on the structure of VvGH3.1 and AtGH3.12. The root mean squared deviation (RMSD) values of the backbone atoms between the open and close state for the N-terminal domain, the C-terminal domain and the whole protein are 0.06 Å, 26.50 Å and 13.64 Å, respectively, which indicates that the open state can be characterized by C-terminal domain rotation. In order to investigate the structural and dynamics of this enzyme with and without ATP and substrates, we performed MD simulations for five systems for 60 ns (GH3.1-Apo, GH3.1-ATP, GH3.1-ATP-IAA, GH3.1-ATP-NAA and GH3.1-ATP-BTOA). The time dependence of RMSD of the backbone atoms is shown in Fig. 2. It can be seen that all systems seem to reach equilibrium after 25 ns of simulation. The RMSD of the backbone atoms stabilized around 1.85 to 2.20 Å in four systems, except the BTOA binding system with the largest value of 2.70 Å. The heavy atoms of ligand and the GH3.1 protein residues around 5 Å of ligand relative to the initial structure coordinates are also plotted in panel B to E in Fig. 2. We can see that the ATP and its binding pocket reach equilibration after 25 ns in GH3.1-ATP system. IAA, NAA and their binding pocket reach equilibration after 35 ns while 40 ns in BTOA binding system. Compare three substrates binding systems, BTOA induced larger variation of ATP, the binding pocket and the whole protein than IAA

and NAA in respective to the initial structure.

To examine the mobility of protein residues, the root-mean-square fluctuation (RMSF) versus the residue number for the five systems was investigated, as shown in Fig. 3. It can be seen that the fluctuations mostly occur in the loop regions of the protein and other residues are stable with less fluctuation, some of them may be caused by the intrinsic flexibility of this protein, such as loops in  $\alpha 3$ - $\alpha 4$ ,  $\beta 17$ - $\beta 18$  and  $\beta 18$ - $\alpha 17$ .  $\beta 1$ - $\beta 2$  loop (residue 108 to 115) shows reduced flexibility in substrate binding system (mean RMSF in IAA, NAA and BTOA binding system: 1.66 Å, 1.28 Å and 1.45 Å, respectively) compared with those in apo system (mean RMSF: 2.44 Å).  $\alpha 8$ - $\alpha 9$  loop (residue 183 to 189) also exhibits reduced flexibility in substrate binding system (mean RMSF in IAA, NAA and BTOA binding system: 1.31 Å, 1.77 Å and 1.53 Å respectively) compared with those in apo system (mean RMSF: 2.03 Å). It is indicated that substrate binding can stabilize  $\beta 1$ - $\beta 2$  loop and  $\alpha 8$ - $\alpha 9$  loop. While  $\beta 16$ - $\alpha 16$  (residue 439 to 446) and  $\beta 17$ - $\beta 18$  loop (residue 478 to 484) exhibit increased flexibility upon IAA binding (mean RMSF: 1.71 Å and 2.62 Å) compared with apo system (mean RMSF: 1.22 Å and 1.76 Å). The fluctuation of  $\alpha 19$ - $\alpha 20$  loop (residue 549 to 574) in BTOA system (mean RMSF: 2.13 Å) is higher compared with apo system (mean RMSF: 1.24 Å).

### Binding free energy analysis

To get the quantitative estimation of the binding affinity of substrates with GH3.1 enzyme, we performed MM-GB/PBSA calculations of three substrate binding systems by extracting 1000 snapshots from the last 10 ns of the trajectories. The results were summarized in Table 1. Both the MM-GBSA and MM-PBSA values suggested that IAA bound to GH3.1 tighter than NAA and BTOA and ranked these substrates with an order of IAA < NAA < BTOA. The predicted rank is in agreement with the experimental work<sup>19</sup>, in which GH3.1 shows higher affinity with IAA (lower  $K_m$  value) than NAA, while the formation of the BTOA–Asp conjugate is not detected. As a reference, the  $K_m$  value is listed in Table 1, which is sometimes but not always

equal to the  $K_d$ , depending on the enzyme and substrates. In this reaction,  $K_m$  value can reflect the affinity of substrates<sup>18, 19</sup>. MM-GBSA method performs better than MM-PBSA method, while both methods predict that NAA is a non-binder. This result is inconsistent with the previous work showing that NAA is a weak binder. Both methods were able to provide a reasonably ranking of binding affinity of the three substrates, but can't predict the absolute binding free energy. It has been approved that these methods can give good rank with higher efficiency than the rigorous methods, free energy perturbation (FEP) and thermodynamic integration (TI) methods, while they may fail in reproducing the absolute binding free energy<sup>45</sup>. The performance of MM-PBSA and MM-GBSA may be improved by solving the following problems, such as the inaccurate molecular mechanical force field, the solvation free energy error, the entropy error and the inadequate sampling method<sup>46</sup>.

Further analysis suggests that in both methods, the calculated van der Waals contribution to the binding free energies for all of the complexes, are different for each other, ranging from -17.17 to -22.83 kcal/mol. The calculated electrostatic contributions to the binding free energies for the three systems are quite different, with values of -37.17, 18.69 and 12.54 kcal/mol, respectively. Nonpolar solvation terms, which correspond to the burial of SASA upon binding, contribute slightly favorably. Polar solvation terms ( $\Delta G_{GB}$ ) has unfavorable contribution to the IAA binding, while slightly favorable for NAA and BTOA binding. The corresponding entropy contributions for the three systems are different to each other, with the values ranging from 13.58 to 17.92 kcal/mol.

### **Analysis of the binding mode between the substrates and GH3.1**

The initial binding mode of the three substrates obtained from molecular docking was shown in supporting information (Fig. S1). It can be seen from that the three substrates exhibit highly similar binding mode with GH3.1. The hydrogen bond interactions between substrates and Ser339 and Gln560 exist in all three systems. BTOA forms hydrogen bond interaction with Ser108, while IAA and NAA fail to

form that interaction, which caused by the short carboxylate group. Davies *et al.* suggested that those substrates bound GH3.1 in a similar mode with the carboxylate group oriented toward Ser107, Ser108 and Ser339<sup>21</sup>. They concluded that the stable conformation of NAA may collide with the binding of ATP, while BTOA bound quite strongly with residues that are necessary of positioning ATP for catalysis. The bind mode in close form may be difficult to reflect the actual binding site and position of ATP. For example, the C-terminal rotation made the  $\alpha 19/\alpha 20$  loop move away from the binding site, corresponding to  $\alpha 18/\alpha 19$  loop in GH3.12, while Lys550 in  $\alpha 18/\alpha 19$  is confirmed to be important for the activity of AtGH3.12 in the first half-reaction<sup>20</sup>. Consistently, Lys562 is found to interact with the phosphate group of ATP and Gln560 forms hydrogen bond interaction with all the substrates in our docking study. The interaction with Ser108, Ser339 and Gln560 hold the substrates in a properly mode, which facilitates the catalytic reaction.

To better understand the specificity of the substrate recognition under dynamic condition, the intermolecular hydrogen bond networks between the GH3.1 and IAA, NAA and BTOA were also monitored. The hydrogen bonds existing throughout the simulation are listed in Table 2. It can be seen from that the carboxyl of IAA hold a stable hydrogen bond with Ser108 with an occupation of 80.51%, which were demonstrated play a critical role in the catalyze process of adenylation reaction step. Moreover, the carboxyl of IAA forms stable hydrogen bonds with Gln560 and Ser339, existing for more than 60% of the duration of the simulation. The carboxylate group of NAA forms hydrogen bond with Ser108 with an occupation nearly 50%. BTOA forms hydrogen bond with Ser339 and Ser108 at the beginning of MD simulation (the former 20 ns), also forms hydrogen bonds with Gln560 at 1-12 and 41- 44 ns of the MD simulation, but some of these interactions can't be maintained (Fig. S2). It forms hydrogen bond with Ser108 during 25-45 ns with low occupation. The hydrogen bond between BTOA and Ser339 with occupation of 31.66% and 30.41% also indicates weak interaction between BTOA and GH3.1 (Fig. S2). It forms hydrogen bond with Ser108 with lower occupation of 35.33% and 35.26%, suggesting an unstable interacting mode. As seen from Fig. 4, IAA formed hydrogen bond with Ser108,

Ser339 and Gln560 and hydrophobic interaction with Val172, Leu173, Val229, Met335, Ala337, Ser338 and Tyr342, which is found to make a likely contribution to IAA binding Davies' work<sup>21</sup>. NAA formed hydrogen bond with Ser108 and hydrophobic interaction with Ser107, Val172, Val229, Ala337 and Ser338. BTOA lost the hydrogen bond interaction with Ser108, Ser339 and Gln560 during the last 10 ns of trajectories, moving to a hydrophobic pocket surrounded by Val172, Leu173, Val 229, Met335, Ala337 and Ser338. Ser108 and Ser339 were suggested to play key role in the two-step catalytic reaction<sup>21,22</sup>. BTOA lost the hydrogen bond interaction with them, may results in a poor substrate.

### **Analysis of access tunnels in GH3.1-substrates complex structures**

Due to the high flexibility of GH3.1 in the two-step catalytic process, its structural change throughout the simulations may be directly relevant to substrate access and product egress. To explore the possible ingress/egress pathways of ATP and substrates, Caver 3.0 was employed to analyze the 100 snapshots extracted from the last 10 ns of MD simulation trajectory for the five systems. All the pathways with the bottleneck radius equal or larger than 1.0 Å were identified. Then they were clustered into 36 (GH3.1-Apo system), 25 (GH3.1-ATP system), 36 (GH3.1-IAA system), 21 (GH3.1-NAA system) and 26 (GH3.1-BTOA system) pathways.

The pathways were ranked by the priority value, which was calculated by averaging sum of tunnel throughputs for a given pathway cluster over all snapshots. The top five ranked pathway clusters were shown in the Fig. 5, and their characteristics were presented in Table 3. The five tunnels are defined as a, b, c, d, e; a1, d1, c1, g, f; a2, c2, e2, h, i; a3, d3, e3, f3, i3; a4, b4, h4, i4, j, for GH3.1-Apo, GH3.1-ATP, GH3.1-ATP-IAA, GH3.1-ATP-NAA and GH3.1-ATP-BTOA, respectively. For the panel A and B in Fig. 5, path a, c and d are common in GH3.1-Apo and GH3.1-ATP system. But the channels display larger bottleneck radius in the apo system than those in ATP binding system. ATP may access or egress to the active site mostly by path a and c. After ATP binding, path a becomes narrower than those in apo system with a

reduced bottleneck radius of 1.16 Å. Substrates may enter mainly through path f and c1. For the panel C to E in Fig. 5, path a (a2, a3 and a4), located between  $\beta$ 1- $\beta$ 2 loop (P-loop)<sup>21</sup> and  $\alpha$ 19- $\alpha$ 20 loop, is common in three substrate binding systems. It is the most potent channel for substrate ingress and egress to the binding pocket. Comparing these three pathways, a4 shows larger mean bottleneck radius and shorter length than a2 and a3. In addition, path c2 and f3 are major channels for IAA and NAA, which locate between  $\alpha$ 8- $\alpha$ 9 loop and disappear in BTOA binding system. The motions of  $\alpha$ 19 and  $\alpha$ 20 open another channel h4 in BTOA binding system, with larger bottleneck radius than other channels in IAA and NAA system, which suggests that BTOA is easier to escape from the binding pocket than IAA and NAA, thus results in BTOA a poor substrate for VvGH3.1.

### Conclusion

In the present study, the interactions between VvGH3.1 and three substrates, IAA, NAA and BTOA were explored by using an integrated computational protocol, consisting of molecular docking, MD simulations, binding free energy calculations and active site access channel analysis. The predicted binding free energy with an order of IAA < NAA < BTOA is in agreement with the experimental work. Further decomposition of the overall binding free energies into individual energy terms indicates that the electrostatic interaction energy is the dominant force for IAA binding, while van der Waals energy for NAA and BTOA binding. The H-bonds between the carboxyl group of the substrates and the residue Ser108, Ser339, and Gln560 are crucial for IAA binding. Finally, the active site access channel analysis indicates that the channels vary between GH3.1-Apo, GH3.1-ATP and GH3.1-ATP-substrates systems, identifies that the dominant channels for ATP and substrates and reflects the high flexibility of this enzyme. The dominant channels for BTOA-GH3.1 system include a4 and h4 path. The former is common to those in IAA (a2) and NAA (a3) system, which has larger bottleneck radius and shorter length than them. The latter also shows larger bottleneck radius than other channels in IAA and NAA system, which suggests that BTOA is easier to escape than IAA and NAA, thus

results in BTOA a poor substrate for VvGH3.1. This study provides novel insight to understand the substrate selectivity mechanism of GH3.1 and also for the rational design novel auxin-based herbicides and growth regulators.

### Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 31300599, 31301438), the Talents of High Level Scientific Research Foundation (No. 6631318) and the Graduate Research Plan (No. 7601415) of Qingdao Agricultural University and the Shandong Province Higher Educational Science and Technology Program China (Grant No. J14LE13).

### Supporting Information

Fig. S1. Binding modes of the three substrates to the open form of VvGH3.1 predicted by molecular docking. The carbon atoms of substrates are colored in green (IAA), slate (NAA), salmon (BTOA).

Fig. S2. Monitoring for the intermolecular hydrogen bonds between GH3.1 and substrates during the MD simulation, (A) GH3.1-ATP-IAA system; (B) GH3.1-ATP-NAA system; (C) GH3.1-ATP-BTOA system.

### References

1. X. Zhang, G. G. Luo, R. H. Wang, J. Wang and D. Himelrick, *J. Amer. Soc. Hort. Sci.*, 2003, **128**, 316-323.
2. C. Böttcher, R. A. Keyzers, P. K. Boss and C. Davies, *J. Exp. Bot.*, 2010, **61**, 3615-3625.
3. P. J. Davies, *Plant hormones: physiology, biotechnology and molecular biology*, Kluwer Academic Publishers, London, 2004.
4. J. Ludwig-Muller, *J. Exp. Bot.*, 2011, **62**, 1757-1773.
5. Y. Zhao, *Annu. Rev. Plant. Biol.*, 2010, **61**, 49-64.
6. H. Wang, C. E. Tian, J. Duan and K. Wu, *Plant. Growth. Regul.*, 2008, **56**, 225-232.

7. P. E. Staswick, B. Serban, M. Rowe, I. Tiriyaki, M. T. Maldonado, M. C. Maldonado and W. Suza, *Plant Cell*, 2005, **17**, 616-627.
8. A. M. Gulick, *ACS. Chem. Biol.*, 2009, **4**, 811-827.
9. C. S. Westfall, J. Herrmann, Q. Chen, S. Wang and J. M. Jez, *Plant. Signal. Behav.*, 2010, **5**, 1607-1612.
10. S. LeClere, R. Tellez, R. A. Rampey, S. P. Matsuda and B. Bartel, *J. Biol. Chem.*, 2002, **277**, 20446-20452.
11. B. Savic, S. Tomic, V. Magnus, K. Gruden, K. Barle, R. Grenkovic, J. Ludwig-Muller and B. Salopek-Sondi, *Plant. Cell. Physiol.*, 2009, **50**, 1587-1599.
12. S. Tsurumi and S. Wada, *Plant. Cell. Physiol.*, 1986, **27**, 1513-1522.
13. P. E. Staswick, *Plant Cell*, 2002, **14**, 1405-1415.
14. J. Terol, C. Domingo and M. Talón, *Gene*, 2006, **371**, 279-290.
15. R. A. Okrent and M. C. Wildermuth, *Plant. Mol. Biol.*, 2011, **76**, 489-505.
16. Q. Chen, B. Zhang, L. M. Hicks, S. Wang and J. M. Jez, *Anal. Biochem.*, 2009, **390**, 149-154.
17. Q. Chen, C. S. Westfall, L. M. Hicks, S. Wang and J. M. Jez, *J. Biol. Chem.*, 2010, **285**, 29780-29786.
18. R. A. Okrent, M. D. Brooks and M. C. Wildermuth, *J. Biol. Chem.*, 2009, **284**, 9742-9754.
19. C. Böttcher, P. K. Boss and C. Davies, *J. Exp. Bot.*, 2011, **62**, 4267-4280.
20. C. S. Westfall, A. M. Muehler and J. M. Jez, *J. Biol. Chem.*, 2013, **288**, 19304-19311.
21. C. S. Westfall, C. Zubieta, J. Herrmann, U. Kapp, M. H. Nanao and J. M. Jez, *Science*, 2012, **336**, 1708-1711.
22. T. S. Peat, C. Bottcher, J. Newman, D. Lucent, N. Cowieson and C. Davies, *Plant Cell*, 2012, **24**, 4525-4538.
23. A. Round, E. Brown, R. Marcellin, U. Kapp, C. S. Westfall, J. M. Jez and C. Zubieta, *Acta Crystallogr. D: Biol. Crystallogr.*, 2013, **69**, 2072-2080.
24. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I.

- N. Shindyalov and P. E. Bourne, *Nucleic. Acids. Res.*, 2000, **28**, 235-242.
25. H. Berman, K. Henrick and H. Nakamura, *Nat. Struct. Mol. Biol.*, 2003, **10**, 980-980.
26. M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede, *Nucleic. Acids. Res.*, 2014, **42**, W252-258.
27. F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli and T. Schwede, *Nucleic. Acids. Res.*, 2009, **37**, D387-392.
28. K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, **22**, 195-201.
29. N. Guex, M. C. Peitsch and T. Schwede, *Electrophoresis*, 2009, **30**, S162-S173.
30. O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455-461.
31. *Discovery Studio, verison 2.5.5*, San Diego: Accelrys Software Inc. CA, 2010.
32. D. Sitkoff, K. A. Sharp and B. Honig, *J. Phys. Chem.*, 1994, **98**, 1978-1988.
33. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269-10280.
34. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox,

- Gaussian, Inc., Wallingford, CT, USA 2009.
35. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
  36. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct.,Funct., Bioinform.*, 2006, **65**, 712-725.
  37. R. A. Bryce, <http://www.pharmacy.manchester.ac.uk/bryce/amber/>.
  38. K. L. Meagher, L. T. Redman and H. A. Carlson, *J. Comput. Chem.*, 2003, **24**, 1016-1025.
  39. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.
  40. T. G. Coleman, H. C. Mesick and R. L. Darby, *Ann. Biomed. Eng.*, 1977, **5**, 322-328.
  41. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
  42. J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 586-596.
  43. E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor and J. Damborsky, *PLoS Comput. Biol.*, 2012, **8**, e1002708.
  44. Schrödinger, LLC, *The PyMOL Molecular Graphics System, Version 1.3.1*, 2010.
  45. H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719-16729.
  46. J. Wang, T. Hou and X. Xu, *Curr. Comput-Aid. Drug.*, 2006, **2**, 95-103.

Table 1. Binding free energy<sup>a</sup> for three substrates bound to VvGH3.1 by MM-GB/PBSA methods.

Method	MM-GB/SA						MM-PB/SA			$K_m^b$ ( $\mu\text{M}$ )
System	$\Delta E_{\text{ele}}$	$\Delta E_{\text{vdw}}$	$-T\Delta S$	$\Delta G_{\text{SA}}$	$\Delta G_{\text{GB}}$	$\Delta G_{\text{binding}}$	$\Delta G_{\text{SA}}$	$\Delta G_{\text{PB}}$	$\Delta G_{\text{binding}}$	
IAA	-37.17±0.63	-21.06±0.13	17.92±1.02	-2.53±0.01	36.26±0.58	-6.58	-3.50±0.01	30.90±0.54	-12.91	10.4±1.20
NAA	18.69±1.18	-22.83±0.02	13.58±0.90	-2.04±0.00	-5.95±1.14	1.45	-3.74±0.00	-1.90±1.06	3.80	190.3±26.2
BTOA	12.54±0.73	-17.17±0.34	14.57±1.02	-1.60±0.03	-0.32±0.89	8.02	-3.04±0.04	-2.34±0.78	4.56	n/a <sup>c</sup>

<sup>a</sup>All energies are in kcal/mol.

<sup>b</sup> $K_m$  value is listed as a reference, which is sometimes but not always equal to the  $K_d$ , depending on the enzyme and substrates. In this reaction,  $K_m$  value can reflect the affinity of substrates<sup>18, 19</sup>.

<sup>c</sup>Not determined.

Table2. Analysis of hydrogen bond network between substrates and VvGH3.1.

Ligand	Acceptor	Donor	Occupied (%)	Distance (Å)	Angel (°)
IAA	IAA: O2	GLN560: NE2	81.94	2.83	159.99
	IAA: O2	SER108: HG	80.51	2.63	165.10
	IAA: O1	SER339: HG	75.67	2.64	164.69
	IAA: O1	SER339: H	57.57	2.82	150.53
NAA	NAA: O1	SER108: HG	48.70	2.64	165.21
	NAA: O1	SER108: H	35.06	2.81	148.88
	NAA: O2	SER108: HG	33.19	2.66	165.20
BTOA	BTOA: O3	SER108: H	35.33	2.81	151.71
	BTOA: O3	SER108: HG	35.26	2.68	163.86
	BTOA: O2	SER339: HG	31.66	2.68	164.87
	BTOA: O2	SER339: H	30.41	2.81	156.09

Table3. Characteristic of the pathways of five systems.

A. Characteristic of the top 5 ranked pathways of GH3.1-Apo					
Parameter	value				
Pathway	a	b	c	d	e
Occurrence <sup>a</sup>	97%	78%	100%	95%	89%
Mean bottleneck radius <sup>b</sup> [Å]	1.87	1.87	1.66	1.35	1.39
Max bottleneck radius [Å]	2.38	2.35	2.12	1.89	1.96
Mean pathway length <sup>b</sup> [Å]	11.82	9.18	33.04	23.18	22.69
Mean throughput <sup>b,c</sup>	0.78	0.80	0.55	0.54	0.55
Priority <sup>d</sup>	0.76	0.62	0.55	0.49	0.47
B. Characteristic of the top 5 ranked pathways of GH3.1-ATP					
Parameter	value				
pathway	f	c1	g	a1	d1
Occurrence <sup>a</sup>	99%	91%	68%	55%	39%
Mean bottleneck radius <sup>b</sup> [Å]	1.69	1.35	1.16	1.16	1.16
Max bottleneck radius [Å]	2.35	1.77	1.58	1.65	1.51
Mean pathway length <sup>b</sup> [Å]	13.96	20.37	23.64	31.18	26.91
Mean throughput <sup>b,c</sup>	0.69	0.53	0.41	0.35	0.38
Priority <sup>d</sup>	0.69	0.48	0.28	0.19	0.14
C. Characteristic of the top 5 ranked pathways of GH3.1-ATP-IAA					
Parameter	value				
pathway	c2	a2	h	e2	i
Occurrence <sup>a</sup>	99%	96%	81%	79%	60%
Mean bottleneck radius <sup>b</sup> [Å]	1.36	1.18	1.20	1.21	1.20
Max bottleneck radius [Å]	1.65	1.50	1.51	1.55	1.43
Mean pathway length <sup>b</sup> [Å]	18.20	25.34	21.38	22.97	33.94
Mean throughput <sup>b,c</sup>	0.58	0.46	0.49	0.44	0.37
Priority <sup>d</sup>	0.58	0.45	0.39	0.35	0.22
D. Characteristic of the top 5 ranked pathways of GH3.1-ATP-NAA					
Parameter	value				
pathway	f3	a3	d3	e3	i3
Occurrence <sup>a</sup>	99%	95%	76%	89%	72%
Mean bottleneck radius <sup>b</sup> [Å]	1.35	1.38	1.31	1.11	1.18
Max bottleneck radius [Å]	1.73	1.94	1.91	1.37	1.60
Mean pathway length <sup>b</sup> [Å]	20.36	28.03	21.90	26.18	29.37
Mean throughput <sup>b,c</sup>	0.53	0.50	0.57	0.40	0.44
Priority <sup>d</sup>	0.52	0.48	0.43	0.36	0.31
E. Characteristic of the top 5 ranked pathways of GH3.1-ATP-BTOA					
Parameter	value				
pathway	a4	h4	b4	i4	j
Occurrence <sup>a</sup>	100%	93%	97%	96%	80%
Mean bottleneck radius <sup>b</sup> [Å]	1.54	1.70	1.29	1.20	1.19
Max bottleneck radius [Å]	1.91	2.31	2.0	1.54	1.62
Mean pathway length <sup>b</sup> [Å]	20.31	22.57	20.49	35.50	25.65
Mean throughput <sup>b,c</sup>	0.64	0.62	0.59	0.40	0.44

Priority <sup>d</sup>	0.64	0.58	0.57	0.38	0.35
-----------------------	------	------	------	------	------

<sup>a</sup> Occupancy of snapshots in which at least one pathway with bottleneck radius  $\geq 1.0$  Å accounted entire snapshots.

<sup>b</sup> Characteristics averaged over identified pathways.

<sup>c</sup> The mean throughput of a give pathway.  $\text{throughput} = e^{-\text{cost}}$ ,

$\text{cost} = \frac{L}{r^n}$ ,  $L$  and  $r$  are the length and radius of a pathway,  $n$  is a non-negative real number.

<sup>d</sup> Priority is calculated as a sum of throughputs of all pathways in a given cluster, divided by the total number of snapshots that were analyzed.

### Figure Captions

Fig.1 The overall structure of the build open form of VvGH3-1, colored from N-terminal (cyan) to C-terminal (gold) and the structure of three substrates. The binding pocket around ATP and substrate is shown with blue surface.

Fig. 2 RMSD plots of the complex during MD-simulation. (A) GH3.1 Apo system; (B) GH3.1-ATP system; (C) GH3.1-ATP-IAA system; (D) GH3.1-ATP-NAA system; (E) GH3.1-ATP-BTOA system.

Fig. 3 RMSF plots of C $\alpha$  atoms of five systems.

Fig. 4 The represent structures from the last 10 ns of the MD trajectories of the complexes superimposed on the docked ones. (A) GH3.1-ATP-IAA complex; (B) GH3.1-ATP-NAA complex; (C) GH3.1-ATP-BTOA complex. For panels A, B and C, the carbon atoms of substrate are in gray and salmon and the proteins are in gray and cyan for the docked and MD structures, respectively.

Fig. 5 The five top ranked collective pathways identified throughout the molecular dynamics simulations are depicted in one frame as pathways surface. (A) GH3.1 Apo system; (B) GH3.1-ATP system; (C) GH3.1-ATP-IAA system; (D) GH3.1-ATP-NAA system; (E) GH3.1-ATP-BTOA system. Pathways following the ranked order are shown as blue, green, red, cyan, yellow.

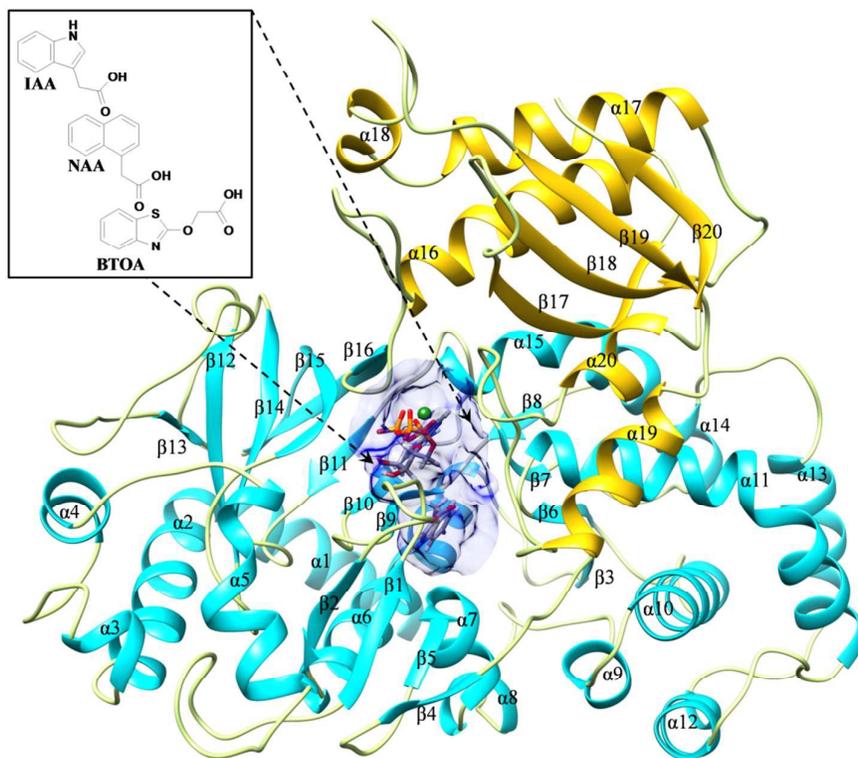


Fig. 1

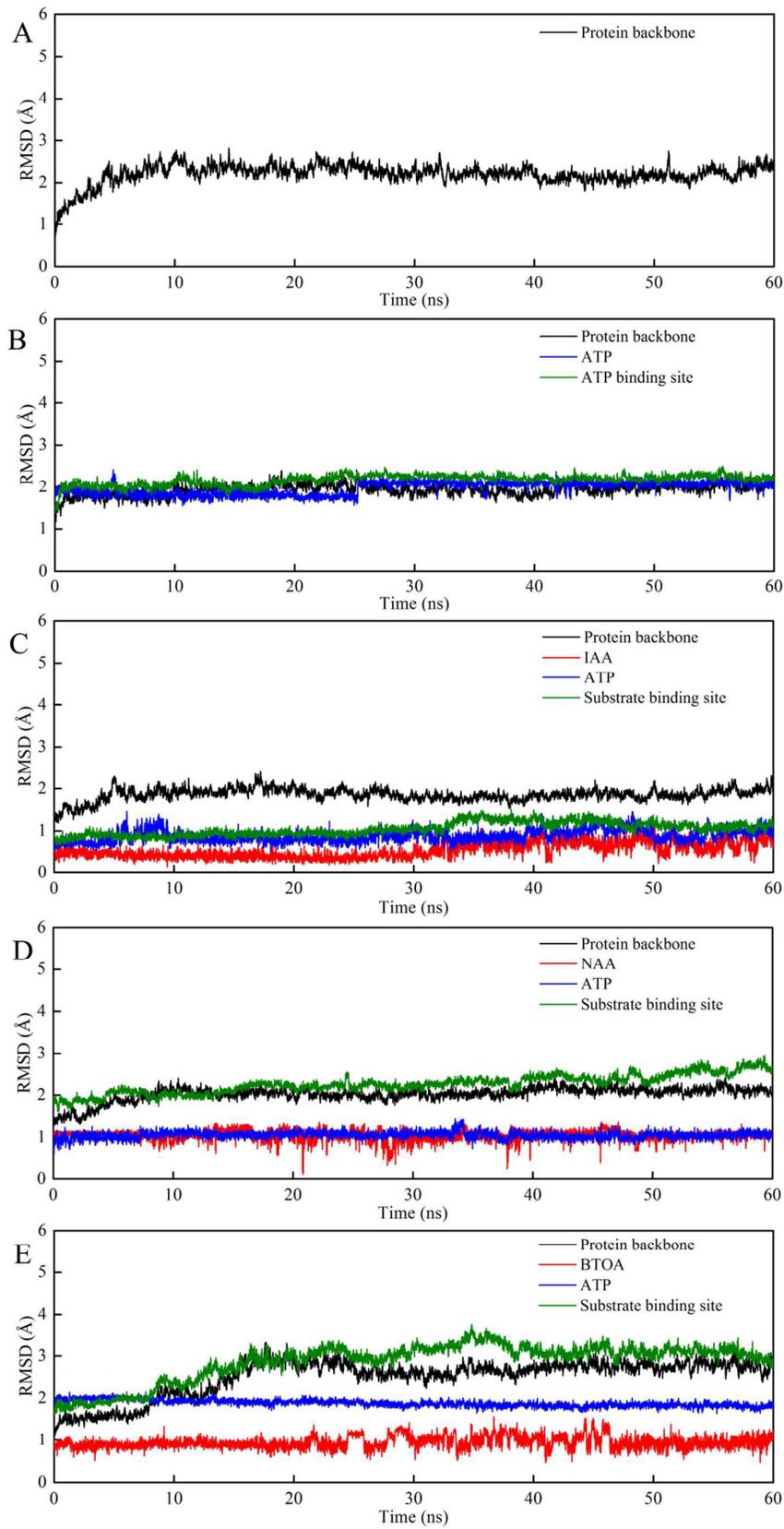


Fig. 2

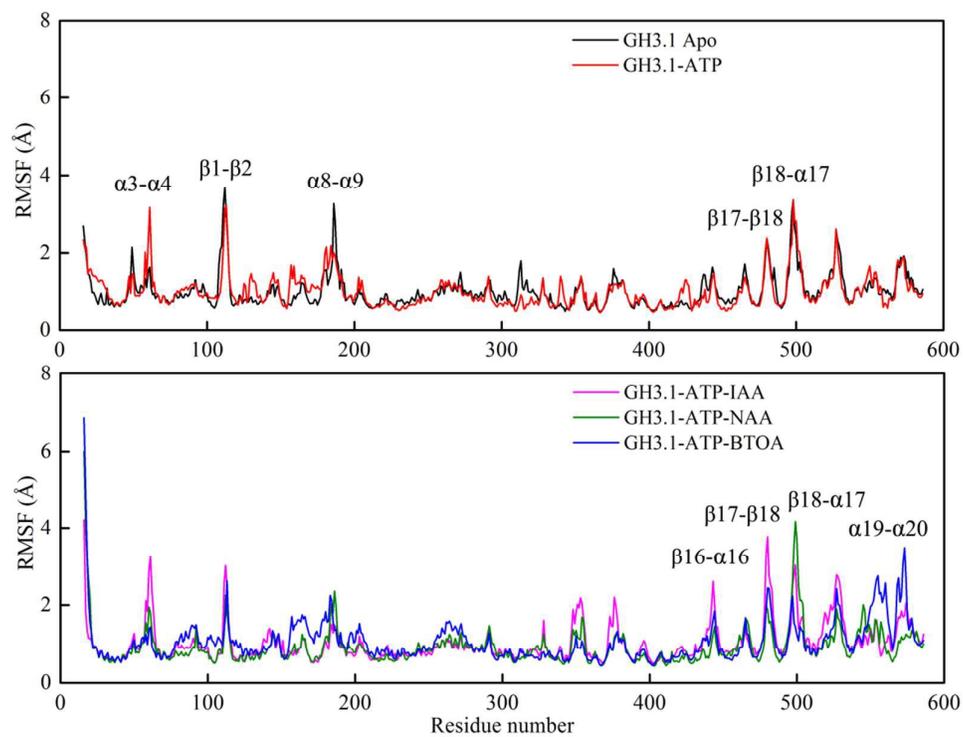


Fig. 3

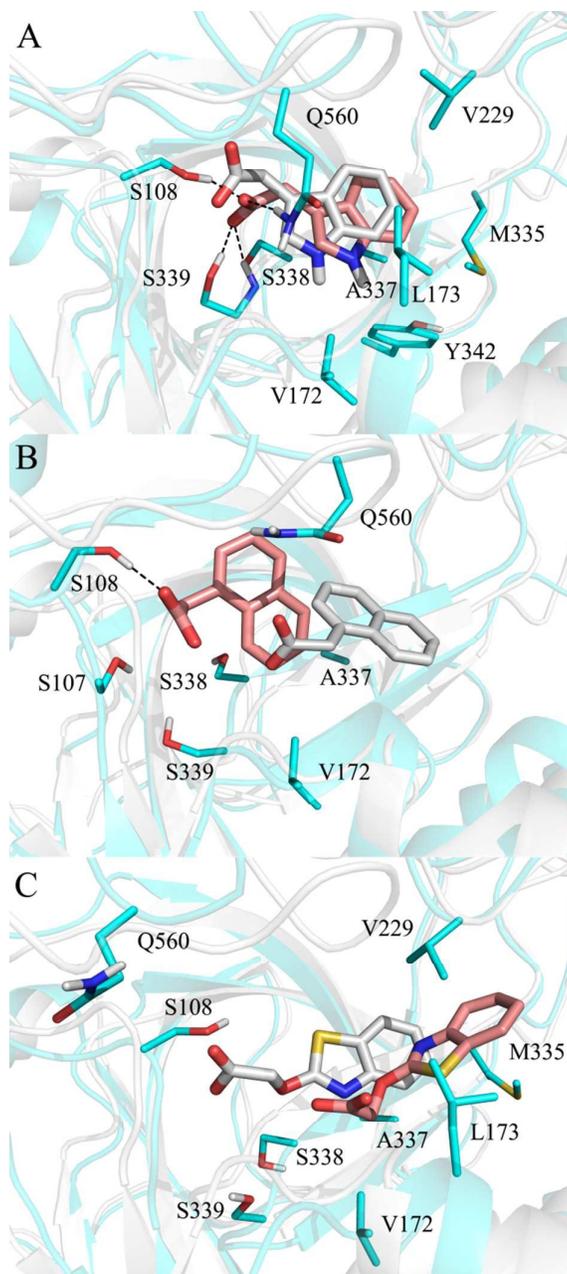


Fig. 4

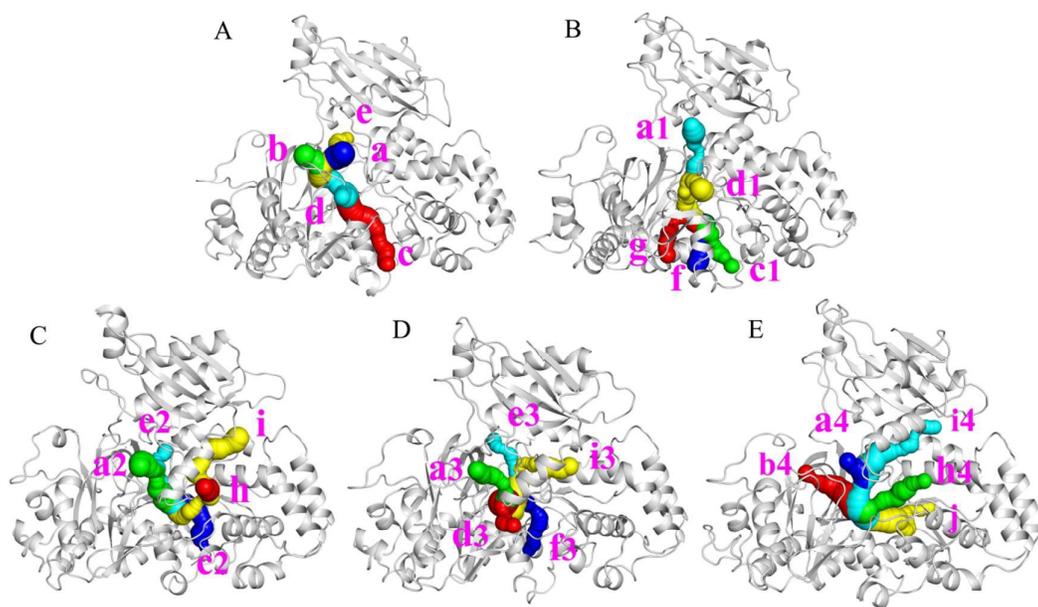


Fig. 5