Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools

Chia-En A. Chang[*], **Yu-Ming M. Huang**, **Leonard J. Mueller**, and **Wanli You**

Department of Chemistry, University of California, Riverside, CA 92521, USA

Yu-Ming M. Huang: yu-ming.huang@email.ucr.edu; Leonard J. Mueller: leonard.mueller@ucr.edu; Wanli You: wyou001@ucr.edu

Abstract

This review discusses the use of molecular modeling tools, together with existing experimental findings, to provide a complete atomic-level description of enzyme dynamics and function. We focus on functionally relevant conformational dynamics of enzymes and the protonation states of substrates. The conformational fluctuations of enzymes usually play a crucial role in substrate recognition and catalysis. Protein dynamics can be altered by a tiny change in a molecular system such as different protonation states of various intermediates or by a significant perturbation such as a ligand association. Here we review recent advances in applying atomistic molecular dynamics (MD) simulations to investigate allosteric and network regulation of tryptophan synthase (TRPS) and protonation states of its intermediates and catalysis. In addition, we review studies using quantum mechanics/molecular mechanics (QM/MM) methods to investigate the protonation states of catalytic residues of β-Ketoacyl ACP synthase I (KasA). We also discuss modeling of large-scale protein motions for HIV-1 protease with coarse-grained Brownian dynamics (BD) simulations.

Keywords

force field; calculation; energy; substrate binding

1. Introduction

Conformational changes of enzymes are often related to regulating and creating an optimal environment for efficient chemical catalysis. The dynamics and conformational changes in enzymes may range from the fluctuation of side chains to large-scale protein motions. The former helps adjusting the environment for chemical reactions and the latter enables the protein to create a tertiary structure for substrate binding. The internal motions of proteins may serve as a "gate" in some systems, such as HIV-1 protease, which controls ligand–protein association [1]. In many enzymes or enzyme complexes, protein motions involve allosteric communication to coordinate the function and reactions, which may be intrinsic to

[*]Correspondence: chiaenc@ucr.edu; Tel.: +1-951-827-7263.
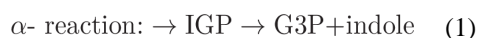
Conflicts of Interest: The authors declare no conflicts of interest.

many enzymes [2,3]. Residues in the active site directly involved in chemical catalysis may even change allosteric networks in different states during catalysis; example states are a ligand-free resting state and a substrate-bound working state.

In addition to conformational rearrangements of protein side-chains and substrates, protonation states of substrates and key catalytically important residues play critical roles in chemical reactions. For β-Ketoacyl ACP synthase I (KasA), the enzyme that involved in fatty acid synthesis, changes in protonation states on key catalytic residues at different steps are essential in enzyme activity [4,5]. The protonation states may significantly affect protein motions as well. With the cofactor pyridoxal-5′-phosphate (PLP), the bioactive form of vitamin B6, the protonation states of different sites on the coenzyme, such as the phenolic oxygen and PLP ring nitrogen in particular, are thought to be critical for establishing the efficient reaction pathway [6–9]. Moreover, changing a single proton location may interrupt the overall protein–substrate stability and restrain enzyme catalysis.

In this review, we discuss three well-known model systems, tryptophan synthase (TRPS) in Figure 1, β-Ketoacyl ACP synthase I (KasA) in Figure 2, and HIV-1 protease in Figure 3. The systems have been studied extensively for understanding large-scale protein motions, allosteric regulations and/or protonation states of key residues and ligands.

Bacterial TRPS is a bifunctional tetrameric $\alpha\beta\beta\alpha$ enzyme complex that catalyzes the last two reactions in the biosynthesis of L-tryptophan (L-Trp). In its α-subunit, the enzyme catalyzes the α-reaction, which involves cleavage of indole-3-glycerol phosphate (IGP) to indole and glyceraldehydes-3-phosphate (G3P) [10]. Indole then diffuses in the channel within the enzyme complex and is condensed with serine at the β active site (β-reaction) (Figure 4). This pathway involves at least nine different intermediates—E(Ain), E(GD$_1$), E(Aex$_1$), E(Q$_1$), E(A-A), E(Q$_2$), E(Q$_3$), E(Aex$_2$), and E(GD$_2$)—formed from PLP and the reacting substrates (Figure 4) [11,12]. The $\alpha\beta$-reaction involves the following combination:

$$\alpha\text{- reaction:} \rightarrow \text{IGP} \rightarrow \text{G3P+indole} \quad (1)$$

$$\beta\text{- reaction:} \rightarrow \text{Indole+L- Ser} \rightarrow \text{L- Trp+H}_2\text{O} \quad (2)$$

TRPS complexes with E(Q$_3$) bound to the β-site; the complex readily reacts and is difficult to crystallize, so substrate analogues equivalent to the intermediates formed along the β-reaction pathway were widely studied. For example, the indole analogues indoline and 2-aminophenol (2AP) can rapidly react with E(A-A) to produce the long-lived indoline quinonoid E(Q)$_{indoline}$, and 2AP quinonoid E(Q)$_{2AP}$, respectively [13,14]. TRPS is a good model system for investigating PLP-dependent enzymatic mechanisms and for studying substrate channeling and allosteric networks that regulate protein functions.

β-Ketoacyl ACP synthase I (KasA) is a key enzyme in the survival of *mycobacterium tuberculosis* (*M. tuberculosis*), the causative pathogen of tuberculosis [15]. It has been proven that KasA is important in the biosynthesis of mycolic acid, which is one of the

building blocks of the cell wall in *M. tuberculosis*. Therefore, KasA is a promising drug target in tuberculosis treatment. The catalytic cycle of enzyme mechanism consists of three steps, acylation, decarboxylation and condensation (Figure 5). First, Cys171 of the active site attacks and binds with acetyl chain delivered by acyl carrier protein (ACP). ACP is subsequently eliminated, and His311 is deprotonated in the acylation step. Decarboxylation of another ACP in the active site generates an enolate, which is stabilized by His311 and His345. The enolate tautomerizes to a carbanion that attacks the carbonyl carbon of acylated Cys171, which finally forms an elongated acyl chain. However, the enzyme mechanism has been debated and discussed [5], due to the multiple possible protonation states of its catalytic residues, Cys171, His311 and His 345.

HIV-1 protease is a dimeric aspartic protease that cleaves the premature polypeptide of HIV-1 and plays an essential role in viral replication [16–18]. Inhibitors targeting HIV-1 protease have been developed and represent among the major antiretroviral therapies for AIDS. Their development is considered a milestone success of structure-based drug design [19]. The motion in the flaps of HIV-1 protease plays a critical role in substrate binding and catalysis. Because of the intrinsic flexibility of HIV-1 protease, the free enzyme contains various flap conformations, usually termed semi-open, open, and closed [20,21]. Most apo structures show a semi-open flap conformation; however, the flaps must open for binding to a natural substrate. The protonation states of the two catalytically important aspartates, Asp-25 and Asp-25′, have been investigated to understand the catalytic mechanism and inhibitor recognition [22]. Because hydrogen atoms of Asp-25 and Asp-25′ can contribute to an important hydrogen bond, the protonation states of the residues need to be considered in structure-based drug design.

Recently developed molecular modeling tools have advanced our knowledge of the atomistic details of enzyme dynamics and catalysis. Through combining experimental techniques such as NMR, single molecule spectroscopy and computational tools, one can not only unveil the protonation states of catalytically significant intermediates of enzymes, which serves as a key component to reveal the catalytic mechanism, but also examine large-amplitude conformational movements of proteins in greater details. In this review, we discuss the applications of molecular dynamics (MD) simulations to investigate atomistic information of TRPS, MD and multi-scale quantum mechanics/molecular mechanics (QM/MM) methods to investigate protonation states of catalytic key residues of TRPS and KasA, and Brownian dynamics (BD) simulations to model large-scale conformational changes in HIV-1 protease for substrate recognition.

2. Computational Tools

2.1. Atomistic Molecular Dynamics Simulations

MD simulations, first developed over 30 years ago [23,24], have advanced from a method to simulate movements of several hundreds of atoms to a widely used way to study the structure and dynamics of macromolecules such as proteins or nucleic acids. Simulation of systems with ~50,000–100,000 atoms are now routine. The simulated system can be represented with different levels of detail, among which atomistic representations can best

reproduce the actual systems; coarse-grained representations are useful with large systems or long simulations [25].

In MD simulations, an initial model of the system is prepared from nuclear magnetic resonance (NMR), crystallographic, or built by homology-modeling when experimental data are not available. Different solvation models, including both explicit and implicit representations, can be used for the simulation [26–34]. Once the simulation system is built, forces acting on each atom can be obtained by deriving equations with parameters such as equilibrium bond length or angle, partial atomic charge, and van der Waals atomic radii (called "force-field" [35,36]) (Figure 6). Forces arising from interactions occur from both bonded and non-bonded atoms. Chemical bonds, atomic angles and improper angles are modeled by harmonic motions, and dihedral angles are modeled by using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. Non-bonded forces include van der Waals interactions, modeled using the Lennard–Jones potential, and electrostatic interactions, modeled using Coulomb's law.

Forces acting on individual atoms are then used for the calculation of accelerations and velocities with classical Newton's law of motion. Therefore atom positions are updated after each time step. Here comes one principle challenge that limits the performance of atomistic MD simulation. To avoid possible collision, the simulation time is advanced, often by only 1 or 2 fs. The time step needs to be shorter than the timescale of bond stretching, the fastest motion in the molecule movements. Tens to hundreds of nanoseconds are currently standard simulation lengths for systems with ~100,000 atoms. However, the achievement of adequate sampling of conformational states may require at least microsecond-long simulations, which would cost much more computational time. While for coarse-grained simulation, which involves a more simplified representation of the system, we can use much larger time steps to greatly extend the length of simulations, in which case the long time length simulation is obtained at the cost of the accuracy including neglecting detailed atomic movements. The good news is, with the rapid development of computational technology and algorithms, we have greatly improved the performance of atomistic MD simulations.

During the past decade, molecular dynamics simulations performing on computer clusters or supercomputers using hundreds of processors in parallel has become very common. With the Message Passing Interface (MPI), by using multiple processors for one calculation task simultaneously, we can largely reduce computation time [41]. The most popular simulation software packages (AMBER [37], CHARMM [38], GROMACS [39], NAMD [40] or TINKER [43]) have long been compatible with MPI. In recent years, the use of graphical processing unit (GPU) cards to accelerate calculations is a major breakthrough in computational simulation field. With the use of GPU cards, which include many arithmetic units working in parallel, MD simulations can be accelerated by tens of times, so a single PC with such a card has the power similar to that of a cluster of workstations with multiple processors. Many major MD codes have already been rewritten to incorporate GPUs [42,44]. Moreover, with the Anton machine that specifically designed for MD simulations [45,46], it is possible to study protein dynamics on a millisecond timescale [47].

Another strategy to improve of efficiency of MD is through enhanced MD simulations. Various enhanced MD simulation methods have been developed over the past decade, among which coarse grain is one of the early approaches [48–50]. Other approaches like hyperdynamics [51], accelerated MD [52], RaMD-db [53] and Gaussian accelerated MD (GaMD) [54] accelerate the simulation by raising the potential energy well to lower the energy barrier. Another group of enhanced MD methods improve sampling by employing additional forces to the region of interest, including steered MD [55,56], target MD [57] and self-guided Langevin dynamics (SGLD) [58,59]. There are also other important MD based simulation methods that can enhance sampling. LowModeMD [60] uses low frequency modes from normal mode analysis to guide MD simulation. Transition interface sampling [61] performs intensive sampling at the transition interface, and obtains the kinetics information from probability. Replica exchange molecular dynamics (REMD) [62,63] makes configurations at high temperatures available to the simulations at low temperatures and vice versa, leading to a robust ensemble sampling both low and high energy configurations. Both computational technologies and enhanced MD simulations have made great progress enhancing computational calculations, it is anticipated that future development will continue improving the efficiency and lead to a new chapter for MD simulation.

2.2. Quantum Mechanics/Molecular Mechanics (QM/MM) Method

Though MD is useful in simulating both local atomistic movements and large-scale protein motions, it lacks the ability to model chemical reactions that involve changes in electronic structures, such as bond breaking, bond forming and charge transfer. Quantum Mechanics (QM) would be the perfect method to describe the changes of electronic structures; however, the large computational demand for electronic structure calculation limits the use of QM methods to only small systems with several hundred atoms. To study large biomolecules, Warshel and Levitt first came up with a hybrid quantum mechanics/molecular mechanics (QM/MM) approach in 1976 [24], which is further developed and evaluated by Karplus by coupling semi-empirical QM methods to MM force field [64]. These three together won the 2013 Nobel Prize in Chemistry for "the development of multiscale models for complex chemical systems".

The QM/MM methods combine the accuracy of QM and speed of MM; thus allowing for the study of chemical reactions in enzymes. The general idea of QM/MM methods is to divide a system into an inner QM region that covers substrates and residues in the enzyme active site, and an outer MM region that includes the rest part of the protein and solvent. If water molecules also play critical roles in catalysis, explicit water molecules in the active site are also included in the inner region. The inner region is treated by a highly accurate QM method, and the outer region is treated by an inexpensive MM method. Therefore, the total potential energy for the system is a sum of MM energy terms, QM energy terms and QM/MM coupling terms:

$$V_{QM/MM} = V_{QM}(QM) + V_{MM}(MM) + V_{QM\text{-}MM}(QM+MM)$$

The QM/MM method is now widely used to model biomolecular systems [65–67], as well as other systems such as inorganic/organometallic [68,69], solid-state [70,71] and explicit

solvent systems [72,73]. Moreover, combined with X-ray crystallography or NMR, QM/MM methods are also useful in the refinement of protein structures [74–77].

2.3. Combining Calculation, X-ray Crystallography and Solid-State NMR for Determining Protonation States

Using X-ray crystallography to determine high resolution protein structures is the first step to provide atomistic details for understanding enzyme function and mechanism. Even at high resolution, protonation states are difficult to be determined by solely X-ray crystallography. Therefore, we use the chemical shift in NMR spectroscopy as an extremely sensitive probe of the chemical environment to determine charge states and electrostatic fields at enzyme active sites. With NMR spectroscopy, isotropic chemical shifts can be used in two specific ways: (1) for directly reporting the chemical state of a probe atom to examine the protonation states or hybridization state at that site; or (2) analyzed for reporting the chemical and structural environment surrounding a probe nucleus. Notably, although solution NMR may not be applicable to proteins with more than several hundreds of residues, solid-state NMR can be used with large protein systems such as TRPS. Solid-state NMR spectra can be obtained on microcrystalline protein samples, prepared basically under the same conditions used to determine X-ray crystal structures [78]. Computational chemistry plays a key role in achieving a high level of structural details for determining protonation states. Typical X-ray and NMR experiments mostly show a protein structure in its static state, and molecular modeling tools are used to illustrate a dynamic view for protein motions and hydrogen bond networks. Although NMR chemical shifts are directly available from experiments, a proton may exchange between multiple ionizable sites. To more accurately predict the protonation states in the equilibrium, a proton can be placed in multiple sites, and *ab initio* calculations are then carried out for first-principle predictions of NMR chemical shifts. This calculation allows for further judging and ranking potential protonation site(s) based on their agreement with experiments [79].

2.4. Coarse-Grained Brownian Dynamics Simulations

Coarse-grained (CG) models for proteins have become popular in recent decades because of the need for modeling large-scale and/or long-time-scale protein motions such as protein folding or large-amplitude conformational fluctuations [80,81]. The CG model and BD simulations are also widely used approaches to investigate ligand–protein association processes. Although the association step is not discussed in this review, forming a substrate-enzyme complex is the very first step for any chemical reaction to happen in the active site of an enzyme. Typically a CG model uses one to six interacting centers (beads) to present each amino acid. A CG model such as a MARTINI CG force field can be a generic force field and the parameter set can be applied to most proteins [82]. Other CG parameters may be protein-dependent [83]. For example, Tozzini and co-authors developed a CG model for HIV-1 protease [84,85]. All the HIV-1 protease X-ray structures available in the Protein Data Bank were included in a statistical set, and the force-field parameters were derived by analyzing the set with the Boltzmann inversion procedure. The potential energy function is a sum of five kinds of interactions:

$$U=U_{bond}+U_{angle}+U_{dihe}+U_{vdw}+U_{elec}$$

The representation of the one-bead model and the force-field functional forms are shown in Figure 7. The CG model has been implemented in the University of Houston Brownian Dynamics (UHBD) simulation package and the Reduced Molecular Dynamics (RedMD) simulation package [86,87].

Because microseconds or longer simulation lengths may be required to sample and obtain reasonable statistics for large-scale protein motions, BD simulations are used to further accelerate the simulations. Using the BD algorithm, we solved the Langevin equation of internal motion in the overdamped limit. This Ermak–McCammon equation is then used to gives a Brownian trajectory [89]. During the simulation, the time step dt may be set to 10–100 fs, which is significantly larger than with atomistic MD simulations, where time step is usually set to 1–2 fs. Hydrodynamic interactions are usually included in BD simulations, and this term is essential to simulate molecular systems under flow. However, calculating hydrodynamic interactions is time-consuming, and this term may be ignored if the hydrodynamic effects are negligible in modeling conformational fluctuations of a protein.

3. Examples of Modeling Enzymes and Substrates

3.1. TRPS: A Model System for Allosteric and Network Regulation in Enzyme Catalysis

Allosteric regulation is important in coordinating function and chemical reactions in many enzymes and protein complexes. A good model system that exhibits allosteric communication, synergistic regulation, and substrate channeling to enhance protein function is TRPS. As shown in Figure 8, TRPS shows synchronization of the α- and β-catalytic activities and conformational switching in its free and substrate bound states [90]. Atomistic MD simulations have been used to study the structure and dynamics of key residues for the α/β-dimer of TRPS and the motions of α-L6 and the COMM domain in TRPS with or without ligands. The MD simulations started from X-ray structures, the reference state. Because the flexible α-L6 loop is missing in all X-ray structures in the apo form, no open α-L6 loop structure is available experimentally. Therefore, the conformations of the open α-L6 were constructed by running MD simulations for a closed α-L6 loop conformation to open the loop [91,92]. With the absence of a ligand, the flexible α-L6 loop can sample open or partially closed conformations, but the loop tends to shift to fully closed conformations when its substrate is placed in the binding site. The fully closed conformations are induced by favorable ligand–protein interactions, which was not seen when the substrate is missing.

Post-analysis of MD simulations also reveals the most common pattern of protein motion and the dynamic correlations. For example, principal component analysis shows that α-L6 and α-L2 loops and the COMM domain have a strong tendency to move in concert (Figure 1). This synergistic movement is important because α-L6 and α-L2 facilitate the formation of the fully closed conformation to catalyze the α-reaction. In the meantime, the COMM domain receives the signal from allosteric communication to induce the fully closed conformation of the β-site and thereby enhance the overall rate of the α/β-reaction. The

correlated motion that supports the experimental suggestions is shown in Figure 8, whereby a bound α-substrate induced a fully closed conformation of α-L6 and α-L2 loop and then further accelerated the COMM domain closure. A detailed cross-correlation matrix can be plotted with programs such as T-Analyst or Bio3D [93,94], which illustrates residue interactions during MD simulations. Correlations between pairs of residues are presented in cyan (correlated motion) or pink (anti-correlated motion) in Figure 9; correlated motions between α-L6 and α-L2 and the COMM domain are highlighted in squares. The COMM domain and α-L2 loop tend to move to the same direction (region I in Figure 9) and the motion of α-L6 is anti-correlated with that of the COMM domain (regions II and III).

In addition to large amplitude loops and domain motions to open or close the active sites, some residues can form networks of correlated motions for chemical catalysis. NMR chemical shift covariance analyses (CHESCA) and MD simulations are powerful tools to detect such amino acid networks when substrates or products are absent (resting state) and present (working state) in the active site [95–97]. We investigated TRPS in its resting and working states to identify the correlation networks from a cluster of residues. We found that a cluster of residues can form a strongly correlated network in its working state, but the same correlation does not exist in its resting state (Figure 10). Mechanistic studies suggested that Glu49 of α-subunit is a key residue for chemical reaction. NMR studies also indicated that Glu49 and a cluster of other residues formed a network motion in the working state for the α-subunit of TRPS from *Escherichia coli*. Figure 10A shows a cluster of residues that form a correlation network: Glu49, Leu48, Thr39, Val197, Gly98, Met101, Phe107, Ile151, Ile166, Ala167, Ser168, Gly170, Ala198, Ala205, Aal254, Val257 and Thr266. Post-analysis for a 100-ns MD simulation for an isolated α-subunit of TRPS suggested that the correlated motions changes significantly between the working and resting state, which are consistent with the results from NMR experiments (Figure 10B). In particular, the catalytic important Glu49 shows considerable correlated motions with residues identified in the above cluster when the substrate is in the bound form (working state) but not in the resting state. Notably, X-ray structures show that Glu49 can stay at different locations, so proper positioning of Glu49 may play a role in regulating enzyme activity. This observation is consistent with results from NMR and MD that Glu49 performs highly correlated motions with other residues. These studies of TRPS demonstrated that examining the networks for catalysis and synergistic allosteric regulation is critical for fully understanding the functions of an enzyme or enzyme complex. The information may be useful for engineering enzymes and for other applications.

3.2. TRPS: How Protonation States Affect Protein Dynamics and Catalysis

Determining the protonation states of ionizable groups in β-reaction intermediates is critical for understanding the mechanisms of chemical reactions in TRPS catalysis since the relocation of a single proton is enough to promote or initialize a chemical reaction in the enzyme active site. On each PLP-bound intermediate, the protonation states play a crucial role in enhancing or weakening the attractions between the enzyme and substrate at six different locations: the phosphoryl group (PG), pyridoxyl nitrogen (PN), pyridoxyl oxygen (PO), Schiff Base (SB) nitrogen, and both carboxyl oxygens (COs) of the L-Ser substrate (Figure 11A). Combined studies of solid-state NMR and MD simulations have shown that

the proton tends to stay at the SB nitrogen in the E(Ain) intermediate, whereas the proton transfers to PO in the E(A-A), E(Q)$_{indoline}$, and E(Q)$_{2AP}$ state. Moreover, the locations of PG, PN and both COs are all deprotonated during most of the lifetime [98,99]. The preference of protonation states in the four β-intermediates is shown in Figure 11B.

The catalysis of TRPS involves a series of proton transfers. Protonation or deprotonation of atoms at the catalytic site or the substrate directly affects structural stability, conformational changes and enzyme activity. For example, in the E(A-A) system, the proton at the PO helps stabilize the substrate in the binding site, so that the PG can form H-bonds with several nearby residues; the PN forms interactions with the β-hydroxyl group of the Ser377 side chain; and the CO atoms interact with the hydroxyl group of the Thr110 side chain and the backbone nitrogens of Gly111 and His115, which keeps the TRPS-E(A-A) complex in a closed conformation of the β-subunit, with no water molecules near the PN, PO, SB nitrogen, and COs. However, the simulations of the proton at the COs, SB nitrogen, PO and PG indicate that the weaker interactions between the COs and nearby residues result in the fluctuation of other residues around the substrate. As well, the H-bonds formed between the PO and Gln114 cause the opening of the β-site, so the solvent moves from bulk solution to the substrate binding site. Thus, changing the location with a single proton not only affects the stability of the local binding site but also alters the overall protein dynamics and molecular motions.

The proton at the PG and PN position has been widely investigated in PLP substrates. Until today, all evidence shows that in most systems of PLP catalysis, the PG is deprotonated. In addition, the PG binding sites are highly conserved—the oxygen atoms of the PG form stable H-bonds with a series of backbone Gly residues, which become major contacts between the PLP enzyme and substrate [100]. The protonated PG would result in a missing H-bond between the carboxylate group of L-Ser and Thr110/His115 and further affect the movement of the Gln114 loop. Thus, this would result in the opening of β-subunit from the original closed conformation. Similar effects have been identified with the change in protonation states at PN. Although the protonated PN has been extensively accepted because of a zwitterionic structure in coenzyme for stabilizing resonance carbanionic intermediates [101–103], recent studies suggest that some PLP enzymes, such as alanine racemase, O-acetylserine sulfhydrylase and TRPS, allow the PN in an unprotonated form [104–106]. Even without the proton at PN, a strong H-bond between the β-hydroxyl group of Ser377 and the PN can still form the carbanionic intermediates. In addition, charges at PN also vary along with the changes of the PN protonation state. The study of charge distribution at the PN shows that stronger negative charge, such as −0.9e, helps to stabilize the carbanionic intermediates as well as the protein conformations near the β-active site.

3.3. KaxA: Using QM/MM Methods to Determine Protonation States of Key Residues in Chemical Reactions

The protonation states of catalytic residues, Cys171, His311 and His 345, of KasA are important in enzyme activity and understanding the catalytic mechanisms. Lee and Engels have been using QM/MM methods to investigate the reaction mechanisms [107–109]. The three catalytic residues can have total eighteen possible protonation states. From multiple X-

ray crystal structures, it can be inferred that His345 is Nε protonated since Nδ forms a hydrogen bond with backbone amide of Ile347 as a hydrogen acceptor, which narrows down the number of possible protonation states to six. Among the six, Cys171/His311(ε) and Cys171/His311(+) can be excluded because they are unable to attack the substrate. Thus, researchers end up with four possible protonation states at the resting state of KasA (Figure 12). Since Lys340/Glu354 pair is at the active site and its protonation states, either neutral or zwitterionic, may influence the result of QM/MM calculation, their protonation states are also investigated.

QM/MM calculation was first conducted for the most probable protonation state, the zwitterionic state, Cys171(−)/His311(+). The QM part comprised Cys171, His311, Lys340, Glu354 and water molecules Wat1 and Wat2. The simulation was 100 ps with DFT (BLYP/6-31G**) for the QM region. The positions of protons along distances α and β can be measured (Figure 13). The distribution of distance α shows that neutral state Cys171/His311(δ) is preferred (~90%), while a population of ~10% indicates that Cys171(−)/His311(+) is less stable. Lys340/Glu354 pair (distance β) shows similar population for neutral and zwitterionic states. Free energy perturbation (FEP) was then conducted to compare the three protonation states, Cys171(−)/His311(δ), Cys171(−)/His311(ε) and Cys171(−)/His311(+), and pKa values were further derived from the computed free energy differences. Cys171(−)/His311(δ) can be ruled out due to its high acidity of the Nδ, while an equilibrium between Cys171(−)/His311(ε) and Cys171(−)/His311(+) can be predicted. Therefore, at the end of catalytic cycle, Cys171(−)/His311(ε) can change back to its initial neutral protonation state via zwitterionic state. By applying the QM/MM methods and FEP, a complete catalytic cycle has been obtained.

3.4. Investigating Large-Scale Conformational Changes in Enzymes: HIV-1 Protease

The dynamics of HIV-1 protease flaps are essential for substrate binding and catalysis. Atomistic MD and coarse-grained BD have been used to study the flap motions, flap rearrangements and substrate/inhibitor binding processes. HIV-1 protease is a C2-symmetric homodimer with the active site consisting of the residues Asp25/Asp25′, Thr26/Thr26′ and Gly27/Gly27′. The active site is gated by the two flaps that are extended beta hairpin loops and act as clamps to bind to a substrate or inhibitor. NMR studies suggested that the free HIV-1 protease has a substantial conformational change in the flap region, and the large-scale motion occurs on a micro- to millisecond (μs–ms) time scale [21]. Because HIV-1 protease is a well-studied system, hundreds of structures have been solved experimentally. Superimposing the experimental structures showed that most parts of the protein have similar conformations except for the flap regions. Three major conformations of HIV-1 protease have been widely discussed—fully-open, semi-open and closed conformation. Atomistic MD simulations suggested that the closed form is rare when a ligand is absent [110–112]. The semi-open conformations are the predominant form, and only <15% of apo HIV-1 protease appears with a fully open conformation (Figure 14 top). Both MD and BD simulations revealed that the flap elbows act as "hinges" to control the movement of the flexible regions, whereas the beta hairpin loops move relatively rigidly. Multiple 20-μs BD simulations with a coarse-grained model suggested that the average flap fully open and closed times are ~70 and 430 ns, respectively [85]. The orientation of flaps is called

handedness. In addition to the large-scale conformational changes of the flaps, the handedness also varies along with the alternation of flaps motions from open to close. In general, semi-open and closed flap conformations are with semi-open and closed handedness, respectively (Figure 14 bottom).

In the apo protein, semi-open conformations, the predominant form of the flaps, still do not open widely enough for substrate association. Although the flaps can open spontaneously, the opening may be induced when a ligand is approaching to the enzyme, which results in an induced flap motion that may accelerate ligand binding [88,113,114]. After substrate/inhibitor binds to HIV-1 protease, the conformations of the flaps will switch from semi-open form with semi-open handedness to closed form with closed handedness (Figure 14 middle). Simulations complement existing observations from experiments and also provide more complete pictures of large-scale enzyme motions and residue rearrangements for ligand binding to help understand binding/catalysis mechanisms and for inhibitor design.

4. Outlook

Understanding enzyme dynamics and functions is important to building a complete picture of the catalytic mechanism and further help for enzyme design. Mechanistic studies of enzymes usually begin with investigating X-ray structures, which provide atomistic details for the first examination of chemical reactions and enzyme functions. With recent advances in the synergistic combination of NMR, MD simulations and *ab initio* QM calculations, we can now specify all atomistic information, including the protonation states, of an enzyme system. MD simulations and NMR further reveal dynamic features important to the enzyme functions. Although less directly observable, the allosteric regulation and correlated motions of a residue network are also involved in controlling substrate binding and chemical catalysis. Various computer modeling techniques, such as coarse-grained BD simulations discussed in this review, have been applied to complement atomistic MD simulations to sample large-amplitude and/or long-time-scale protein motions. The field of enzymology has evolved, and describing enzyme functions and catalysis is no longer based solely on simple organic chemical interactions. In addition, we need to consider protein motions involving communication between various residues and multiple regions of an enzyme complex. With more powerful computation tools and resources, molecular simulations will more significantly contribute to our understanding of enzyme systems and assist in enzyme engineering in the future.

Acknowledgments

References

1. McCammon JA. Gated diffusion-controlled reactions. BMC Biophys. 2011; 4:4. [PubMed: 21595999]

2. Csermely P, Sandhu KS, Hazai E, Hoksza Z, Kiss HJ, Miozzo F, Veres DV, Piazza F, Nussinov R. Disordered proteins and network disorder in network descriptions of protein structure, dynamics

and function: Hypotheses and a comprehensive review. Curr Protein Pept Sci. 2012; 13:19–33. [PubMed: 22044146]

3. Tsai CJ, Del Sol A, Nussinov R. Protein allostery, signal transmission and dynamics: A classification scheme of allosteric mechanisms. Mol Biosyst. 2009; 5:207–216. [PubMed: 19225609]

4. Kremer L, Dover LG, Carrere S, Nampoothiri KM, Lesjean S, Brown AK, Brennan PJ, Minnikin DE, Locht C, Besra GS. Mycolic acid biosynthesis and enzymic characterization of the beta-ketoacyl-acp synthase a-condensing enzyme from mycobacterium tuberculosis. Biochem J. 2002; 364:423–430. [PubMed: 12023885]

5. White SW, Zheng J, Zhang YM, Rock. The structural biology of type ii fatty acid biosynthesis. Annu Rev Biochem. 2005; 74:791–831. [PubMed: 15952903]

6. Toney MD. Controlling reaction specificity in pyridoxal phosphate enzymes. Biochim Biophys Acta. 2011; 1814:1407–1418. [PubMed: 21664990]

7. Crugeiras J, Rios A, Riveiros E, Richard JP. Substituent effects on electrophilic catalysis by the carbonyl group: Anatomy of the rate acceleration for plp-catalyzed deprotonation of glycine. J Am Chem Soc. 2011; 133:3173–3183. [PubMed: 21323335]

8. Major DT, Gao J. A combined quantum mechanical and molecular mechanical study of the reaction mechanism and alpha-amino acidity in alanine racemase. J Am Chem Soc. 2006; 128:16345–16357. [PubMed: 17165790]

9. Major DT, Nam K, Gao J. Transition state stabilization and alpha-amino carbon acidity in alanine racemase. J Am Chem Soc. 2006; 128:8114–8115. [PubMed: 16787057]

10. Sachpatzidis A, Dealwis C, Lubetsky JB, Liang PH, Anderson KS, Lolis E. Crystallographic studies of phosphonate-based alpha-reaction transition-state analogues complexed to tryptophan synthase. Biochemistry. 1999; 38:12665–12674. [PubMed: 10504236]

11. Drewe WF, Dunn MF. Detection and identification of intermediates in the reaction of l-serine with escherichia-coli tryptophan synthase via rapid-scanning ultraviolet visible spectroscopy. Biochemistry. 1985; 24:3977–3987. [PubMed: 3931672]

12. Drewe WF, Dunn MF. Characterization of the reaction of l-serine and indole with escherichia-coli tryptophan synthase via rapid-scanning ultraviolet visible spectroscopy. Biochemistry. 1986; 25:2494–2501. [PubMed: 3087420]

13. Ngo H, Kimmich N, Harris R, Niks D, Blumenstein L, Kulik V, Barends TR, Schlichting I, Dunn MF. Allosteric regulation of substrate channeling in tryptophan synthase: Modulation of the l-serine reaction in stage i of the ss-reaction by alpha-site ligands. Biochemistry. 2007; 46:7740–7753. [PubMed: 17559232]

14. Lai J, Niks D, Wang Y, Domratcheva T, Barends TR, Schwarz F, Olsen RA, Elliott DW, Fatmi MQ, Chang CEA, et al. X-ray and nmr crystallography in an enzyme active site: The indoline quinonoid intermediate in tryptophan synthase. J Am Chem Soc. 2011; 133:4–7. [PubMed: 21142052]

15. Bhatt A, Kremer L, Dai AZ, Sacchettini JC, Jacobs WR Jr. Conditional depletion of kasa, a key enzyme of mycolic acid biosynthesis, leads to mycobacterial cell lysis. J Bacteriol. 2005; 187:7596–7606. [PubMed: 16267284]

16. Debouck C, Gorniak JG, Strickler JE, Meek TD, Metcalf BW, Rosenberg M. Human immunodeficiency virus protease expressed in escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. Proc Natl Acad Sci USA. 1987; 84:8903–8906. [PubMed: 3321060]

17. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci USA. 1988; 85:4686–4690. [PubMed: 3290901]

18. Krausslich HG, Ingraham RH, Skoog MT, Wimmer E, Pallai PV, Carter CA. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. Proc Natl Acad Sci USA. 1989; 86:807–811. [PubMed: 2644644]

19. Wlodawer A, Vondrasek J. Inhibitors of hiv-1 protease: A major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct. 1998; 27:249–284. [PubMed: 9646869]

20. Ishima R, Freedberg DI, Wang YX, Louis JM, Torchia DA. Flap opening and dimer-interface flexibility in the free and inhibitor-bound hiv protease, and their implications for function. Structure (London, England 1993). 1999; 7:1047–1055.

21. Freedberg DI, Ishima R, Jacob J, Wang YX, Kustanovich I, Louis JM, Torchia DA. Rapid structural fluctuations of the free hiv protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. Protein Sci Publ Protein Soc. 2002; 11:221–232.

22. Adachi M, Ohhara T, Kurihara K, Tamada T, Honjo E, Okazaki N, Arai S, Shoyama Y, Kimura K, Matsumura H, et al. Structure of hiv-1 protease in complex with potent inhibitor kni-272 determined by high-resolution X-ray and neutron crystallography. Proc Natl Acad Sci USA. 2009; 106:4641–4646. [PubMed: 19273847]

23. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977; 267:585–590. [PubMed: 301613]

24. Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol. 1976; 103:227–249. [PubMed: 985660]

25. Orozco M, Orellana L, Hospital A, Naganathan AN, Emperador A, Carrillo O, Gelpi JL. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. Adv Protein Chem Struct Biol. 2011; 85:183–215. [PubMed: 21920324]

26. Anandakrishnan R, Drozdetski A, Walker RC, Onufriev AV. Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations. Biophys J. 2015; 108:1153–1164. [PubMed: 25762327]

27. Bashford D, Case DA. Generalized born models of macromolecular solvation effects. Annu Rev Phys Chem. 2000; 51:129–152. [PubMed: 11031278]

28. Haberthur U, Caflisch A. Facts: Fast analytical continuum treatment of solvation. J Comput Chem. 2008; 29:701–715. [PubMed: 17918282]

29. Kleinjung J, Fraternali F. Design and application of implicit solvent models in biomolecular simulations. Curr Opin Struct Biol. 2014; 25:126–134. [PubMed: 24841242]

30. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins. 1999; 35:133–152. [PubMed: 10223287]

31. Luchko T, Gusarov S, Roe DR, Simmerling C, Case DA, Tuszynski J, Kovalenko A. Three-dimensional molecular theory of solvation coupled with molecular dynamics in amber. J Chem Theory Comput. 2010; 6:607–624. [PubMed: 20440377]

32. Orozco M, Luque FJ. Theoretical methods for the description of the solvent effect in biomolecular systems. Chem Rev. 2000; 100:4187–4226. [PubMed: 11749344]

33. Roux B, Simonson T. Implicit solvent models. Biophys Chem. 1999; 78:1–20. [PubMed: 17030302]

34. Vorobjev YN. Advances in implicit models of water solvent to compute conformational free energy and molecular dynamics of proteins at constant ph. Adv Protein Chem Struct Biol. 2011; 85:281–322. [PubMed: 21920327]

35. Rueda M, Ferrer-Costa C, Meyer T, Perez A, Camps J, Hospital A, Gelpi JL, Orozco M. A consensus view of protein dynamics. Proc Natl Acad Sci USA. 2007; 104:796–801. [PubMed: 17215349]

36. Perez A, Lankas F, Luque FJ, Orozco M. Towards a molecular dynamics consensus view of b-DNA flexibility. Nucleic Acids Res. 2008; 36:2379–2394. [PubMed: 18299282]

37. Case D, Babin V, Berryman J, Betz R, Cai Q, Cerutti D, Cheatham T III, Darden T, Duke R, Gohlke H. AMBER. 2014; 14

38. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. Charmm: The biomolecular simulation program. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]

39. Hess B, Kutzner C, van der Spoel D, Lindahl E. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008; 4:435–447. [PubMed: 26620784]

40. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with namd. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]

41. Larsson P, Hess B, Lindahl E. Algorithm improvements for molecular dynamics simulations. Wiley Interdiscip Rev Comput Mol Sci. 2011; 1:93–108.

42. Harvey MJ, Giupponi G, Fabritiis GD. Acemd: Accelerating biomolecular dynamics in the microsecond time scale. J Chem Theory Comput. 2009; 5:1632–1639. [PubMed: 26609855]

43. Ponder JW, Richards FM. An efficient newton-like method for molecular mechanics energy minimization of large molecules. J Comput Chem. 1987; 8:1016–1024.

44. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang LP, Shukla D, et al. Openmm 4: A reusable, extensible, hardware independent library for high performance molecular simulation. J Chem Theory Comput. 2013; 9:461–469. [PubMed: 23316124]

45. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, et al. Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM. 2008; 51:91–97.

46. Shaw, DE.; Grossman, JP.; Bank, JA.; Batson, B.; Butts, JA.; Chao, JC.; Deneroff, MM.; Dror, RO.; Even, A.; Fenton, CH., et al. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. Proceedings of the SC14: International Conference for High Performance Computing, Networking, Storage and Analysis; New Orleans, LA, USA. 16–21 November 2014; p. 41-53.

47. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010; 330:341–346. [PubMed: 20947758]

48. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. J Phys Chem B. 2007; 111:7812–7824. [PubMed: 17569554]

49. Nagarajan A, Andersen JP, Woolf TB. Coarse-grained simulations of transitions in the e2-to-e1 conformations for ca atpase (serca) show entropy-enthalpy compensation. J Mol Biol. 2012; 422:575–593. [PubMed: 22684148]

50. Tozzini V. Coarse-grained models for proteins. Curr Opin Struct Biol. 2005; 15:144–150. [PubMed: 15837171]

51. Voter AF. Hyperdynamics: Accelerated molecular dynamics of infrequent events. Phys Rev Lett. 1997; 78:3908–3911.

52. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys. 2004; 120:11919–11929. [PubMed: 15268227]

53. Doshi U, Hamelberg D. Achieving rigorous accelerated conformational sampling in explicit solvent. J Phys Chem Lett. 2014; 5:1217–1224. [PubMed: 26274474]

54. Miao Y, Feher VA, McCammon JA. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. J Chem Theory Comput. 2015; 11:3584–3595. [PubMed: 26300708]

55. Leech J, Prins JF, Hermans J. Smd: Visual steering of molecular dynamics for protein design. IEEE Comput Sci Eng. 1996; 3:38–45.

56. Nicolini P, Frezzato D, Gellini C, Bizzarri M, Chelli R. Toward quantitative estimates of binding affinities for protein-ligand systems involving large inhibitor compounds: A steered molecular dynamics simulation route. J Comput Chem. 2013; 34:1561–1576. [PubMed: 23620471]

57. Schlitter J, Engels M, Kruger P. Targeted molecular-dynamics—A new approach for searching pathways of conformational transitions. J Mol Graph. 1994; 12:84–89. [PubMed: 7918256]

58. Wu XW, Brooks BR. Self-guided langevin dynamics simulation method. Chem Phys Lett. 2003; 381:512–518.

59. Wu X, Brooks BR, Vanden-Eijnden E. Self-guided langevin dynamics via generalized langevin equation. J Comput Chem. 2016; 37:595–601. [PubMed: 26183423]

60. Labute P. Lowmodemd-implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. J Chem Inform Model. 2010; 50:792–800.

61. Swenson DWH, Bolhuis PG. A replica exchange transition interface sampling method with multiple interface sets for investigating networks of rare events. J Chem Phys. 2014; 141doi: 10.1063/1.4890037

62. Voter AF. Parallel replica method for dynamics of infrequent events. Phys Rev B. 1998; 57:13985–13988.

63. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999; 314:141–151.

64. Field MJ, Bash PA, Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem. 1990; 11:700–733.

65. Hu H, Yang W. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. Annu Rev Phys Chem. 2008; 59:573–601. [PubMed: 18393679]

66. Senn HM, Thiel W. QM/MM methods for biomolecular systems. Angew Chem (Int Ed Engl). 2009; 48:1198–1229. [PubMed: 19173328]

67. Van der Kamp MW, Mulholland AJ. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry. 2013; 52:2708–2728. [PubMed: 23557014]

68. Chatwin SL, Davidson MG, Doherty C, Donald SM, Jazzar RFR, Macgregor SA, McIntyre GJ, Mahon MF, Whittlesey MK. H–X bond activation via hydrogen transfer to hydride in ruthenium N-heterocyclic carbene complexes: Density functional and synthetic studies. Organometallics. 2006; 25:99–110.

69. Genest A, Woiterski A, Krüger S, Shor AM, Rösch N. The imomm (integrated molecular orbitals/molecular mechanics) approach for ligand-stabilized metal clusters. Comparison to full density functional calculations for the model thiolate cluster $Cu_{13}(SCH_2CH_3)_8$. J Chem Theory Comput. 2006; 2:47–58. [PubMed: 26626378]

70. Shiota Y, Suzuki K, Yoshizawa K. QM/MM study on the catalytic mechanism of benzene hydroxylation over fe–zsm-5. Organometallics. 2006; 25:3118–3123.

71. To J, Sherwood P, Sokol AA, Bush IJ, Catlow CRA, van Dam HJJ, French SA, Guest MF. AM/MM modelling of the ts-1 catalyst using hpcx. J Mater Chem. 2006; 16:1919–1926.

72. Gunaydin H, Acevedo O, Jorgensen WL, Houk KN. Computation of accurate activation barriers for methyl-transfer reactions of sulfonium and ammonium salts in aqueous solution. J Chem Theory Comput. 2007; 3:1028–1035. [PubMed: 26627421]

73. Alexandrova AN, Jorgensen WL. Why urea eliminates ammonia rather than hydrolyzes in aqueous solution. J Phys Chem B. 2007; 111:720–730. [PubMed: 17249815]

74. Hsiao YW, Sanchez-Garcia E, Doerr M, Thiel W. Quantum refinement of protein structures: Implementation and application to the red fluorescent protein dsred.M1. J Phys Chem B. 2010; 114:15413–15423. [PubMed: 20977248]

75. Ryde U. Combined quantum and molecular mechanics calculations on metalloproteins. Curr Opin Chem Biol. 2003; 7:136–142. [PubMed: 12547438]

76. Marti S, Moliner V, Tunon I. Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems. J Chem Theory Comput. 2005; 1:1008–1016. [PubMed: 26641916]

77. Hsiao YW, Drakenberg T, Ryde U. Nmr structure determination of proteins supplemented by quantum chemical calculations: Detailed structure of the ca2+ sites in the egf34 fragment of protein s. J Biomol NMR. 2005; 31:97–114. [PubMed: 15772750]

78. Mueller LJ, Dunn MF. Nmr crystallography of enzyme active sites: Probing chemically detailed, three-dimensional structure in tryptophan synthase. Acc Chem Res. 2013; 46:2008–2017. [PubMed: 23537227]

79. Lai J, Niks D, Wang Y, Domratcheva T, Barends TR, Schwarz F, Olsen RA, Elliott DW, Fatmi MQ, Chang CEA, et al. X-ray and nmr crystallography in an enzyme active site: The indoline quinonoid intermediate in tryptophan synthase. J Am Chem Soc. 2011; 133:4–7. [PubMed: 21142052]

80. Clementi C. Coarse-grained models of protein folding: Toy models or predictive tools? Curr Opin Struct Biol. 2008; 18:10–15. [PubMed: 18160277]

81. Gorfe AA, Chang CEA, Ivanov I, McCammon JA. Dynamics of the acetylcholineasterase tetramer. Biophys J. 2008; 94:1144–1154. [PubMed: 17921202]

82. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The martini force field: Coarse grained model for biomolecular simulations. J Phys Chem B. 2007; 111:7812–7824. [PubMed: 17569554]

83. Leonarski F, Trovato F, Tozzini V, Les A, Trylska J. Evolutionary algorithm in the optimization of a coarse-grained force field. J Chem Theory Comput. 2013; 9:4874–4889. [PubMed: 26583407]

84. Tozzini V, McCammon JA. A coarse grained model for the dynamics of flap opening in hiv-1 protease. Chem Phys Lett. 2005; 413:123–128.

85. Chang CE, Shen T, Trylska J, Tozzini V, McCammon JA. Gated binding of ligands to hiv-1 protease: Brownian dynamics simulations in a coarse-grained model. Biophys J. 2006; 90:3880–3885. [PubMed: 16533835]

86. Gorecki A, Szypowski M, Dlugosz M, Trylska J. Redmd-reduced molecular dynamics package. J Comput Chem. 2009; 30:2364–2373. [PubMed: 19247989]

87. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution: Simulations with the university of houston brownian dynamics program. Comput Phys Commun. 1991; 62:187–197.

88. Chang CEA, Trylska J, Tozzini V, Andrew McCammon J. Binding pathways of ligands to hiv-1 protease: Coarse-grained and atomistic simulations. Chem Biol Drug Des. 2007; 69:5–13. [PubMed: 17313452]

89. Ermak DL, McCammon JA. Brownian dynamics with hydrodynamic interactions. J Chem Phys. 1978; 69:1352–1360.

90. Dunn MF. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Arch Biochem Biophys. 2012; 519:154–166. [PubMed: 22310642]

91. Fatmi MQ, Ai R, Chang CEA. Synergistic regulation and ligand-induced conformational changes of tryptophan synthase. Biochemistry. 2009; 48:9921–9931. [PubMed: 19764814]

92. Fatmi MQ, Chang CEA. The role of oligomerization and cooperative regulation in protein function: The case of tryptophan synthase. PLoS Comput Biol. 2010; 6:e1000994. [PubMed: 21085641]

93. Ai R, Qaiser Fatmi M, Chang CEA. T-analyst: A program for efficient analysis of protein conformational changes by torsion angles. J Comput Aided Mol Des. 2010; 24:819–827. [PubMed: 20689979]

94. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: An r package for the comparative analysis of protein structures. Bioinformatics (Oxford, England). 2006; 22:2695–2696.

95. Axe JM, Yezdimer EM, O'Rourke KF, Kerstetter NE, You W, Chang CEA, Boehr DD. Amino acid networks in a (beta/alpha)(8) barrel enzyme change during catalytic turnover. J Am Chem Soc. 2014; 136:6818–6821. [PubMed: 24766576]

96. Axe JM, Boehr DD. Long-range interactions in the alpha subunit of tryptophan synthase help to coordinate ligand binding, catalysis, and substrate channeling. J Mol Biol. 2013; 425:1527–1545. [PubMed: 23376097]

97. Axe JM, O'Rourke KF, Kerstetter NE, Yezdimer EM, Chan YM, Chasin A, Boehr DD. Severing of a hydrogen bond disrupts amino acid networks in the catalytically active state of the alpha subunit of tryptophan synthase. Protein Sci A Publ Protein Soc. 2015; 24:484–494.

98. Caulkins BG, Bastin B, Yang C, Neubauer TJ, Young RP, Hilario E, Huang YMM, Chang CEA, Fan L, Dunn MF, et al. Protonation states of the tryptophan synthase internal aldimine active site from solid-state nmr spectroscopy: Direct observation of the protonated schiff base linkage to pyridoxal-5′-phosphate. J Am Chem Soc. 2014; 136:12824–12827. [PubMed: 25148001]

99. Huang YMM, You W, Caulkins BG, Dunn MF, Mueller LJ, Chang CEA. Protonation states and catalysis: Molecular dynamics studies of intermediates in tryptophan synthase. Protein Sci A Publ Protein Soc. 2016; 25:166–183.

100. Schnackerz KD, Andi B, Cook PF. P-31 nmr spectroscopy senses the microenvironment of the 5′-phosphate group of enzyme-bound pyridoxal 5′-phosphate. Biochim Biophys Acta Proteins Proteom. 2011; 1814:1447–1458.

101. Eliot AC, Kirsch JF. Pyridoxal phosphate enzymes: Mechanistic, structural, and evolutionary considerations. Annu Rev Biochem. 2004; 73:383–415. [PubMed: 15189147]

102. Fogle EJ, Liu WS, Woon ST, Keller JW, Toney MD. Role of q52 in catalysis of decarboxylation and transamination in dialkylglycine decarboxylase. Biochemistry. 2005; 44:16392–16404. [PubMed: 16342932]

103. Sharif S, Fogle E, Toney MD, Denisov GS, Shenderovich IG, Buntkowsky G, Tolstoy PM, Huot MC, Limbach HH. Nmr localization of protons in critical enzyme hydrogen bonds. J Am Chem Soc. 2007; 129:9558–9559. [PubMed: 17628065]

104. Griswold WR, Toney MD. Role of the pyridine nitrogen in pyridoxal 5′-phosphate catalysis: Activity of three classes of plp enzymes reconstituted with deazapyridoxal 5′-phosphate. J Am Chem Soc. 2011; 133:14823–14830. [PubMed: 21827189]

105. Shaw JP, Petsko GA, Ringe D. Determination of the structure of alanine racemase from bacillus stearothermophilus at 1. 9-angstrom resolution. Biochemistry. 1997; 36:1329–1342. [PubMed: 9063881]

106. Tai CH, Cook PF. Pyridoxal 5′-phosphate dependent alpha,beta-elimination reactions: Mechanism of o-acetylserine sulfhydrylase. Acc Chem Res. 2001; 34:49–59. [PubMed: 11170356]

107. Lee W, Luckner SR, Kisker C, Tonge PJ, Engels B. Elucidation of the protonation states of the catalytic residues in mtkasa: Implications for inhibitor design. Biochemistry. 2011; 50:5743–5756. [PubMed: 21615093]

108. Lee W, Engels B. Clarification on the decarboxylation mechanism in kasa based on the protonation state of key residues in the acyl-enzyme state. J Phys Chem B. 2013; 117:8095–8104. [PubMed: 23768199]

109. Lee W, Engels B. The protonation state of catalytic residues in the resting state of kasa revisited: Detailed mechanism for the activation of kasa by its own substrate. Biochemistry. 2014; 53:919–931. [PubMed: 24479625]

110. Karthik S, Senapati S. Dynamic flaps in hiv-1 protease adopt unique ordering at different stages in the catalytic cycle. Proteins. 2011; 79:1830–1840. [PubMed: 21465560]

111. Deng NJ, Zheng W, Gallicchio E, Levy RM. Insights into the dynamics of hiv-1 protease: A kinetic network model constructed from atomistic simulations. J Am Chem Soc. 2011; 133:9387–9394. [PubMed: 21561098]

112. Hornak V, Okur A, Rizzo RC, Simmerling C. Hiv-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. Proc Natl Acad Sci USA. 2006; 103:915–920. [PubMed: 16418268]

113. Huang YM, Kang M, Chang CEA. Switches of hydrogen bonds during ligand-protein association processes determine binding kinetics. J Mol Recognit JMR. 2014; 27:537–548. [PubMed: 25042708]

114. Kang M, Roberts C, Cheng Y, Chang CEA. Gating and intermolecular interactions in ligand-protein association: Coarse-grained modeling of hiv-1 protease. J Chem Theory Comput. 2011; 7:3438–3446. [PubMed: 26598172]
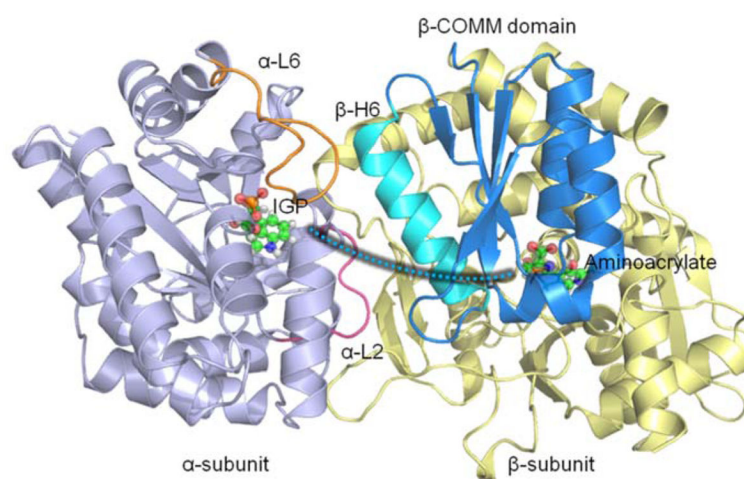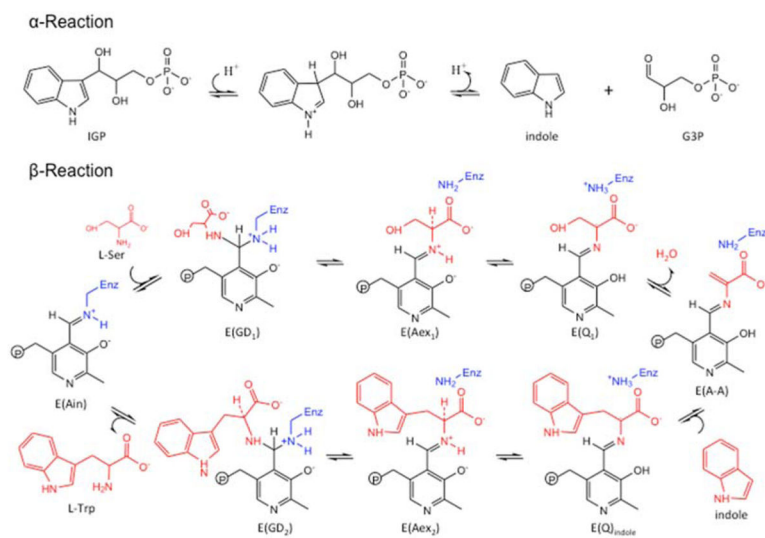
Figure 1.
Overall structure and chemical reactions of tryptophan synthase (TRPS). TRPS is composed of an α- (purple) and β-subunit (yellow). The two ligands binding to each subunit are shown in bead representation. The open, partially closed, and fully closed conformations of the α-subunit are controlled by α-L2 (pink), α-L6 (orange), and β-Helix-6 of the COMM domain (cyan). The tunnel used to channel indole from the α-site to the β-site is marked as a cyan dashed shaded line.
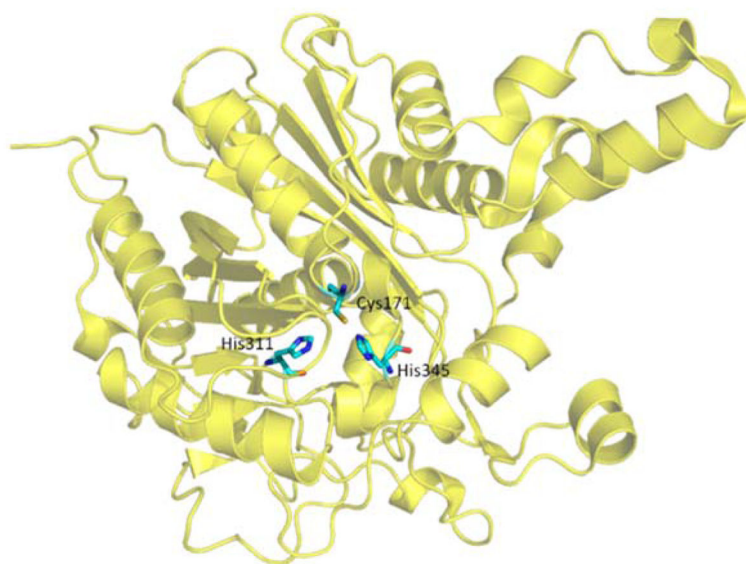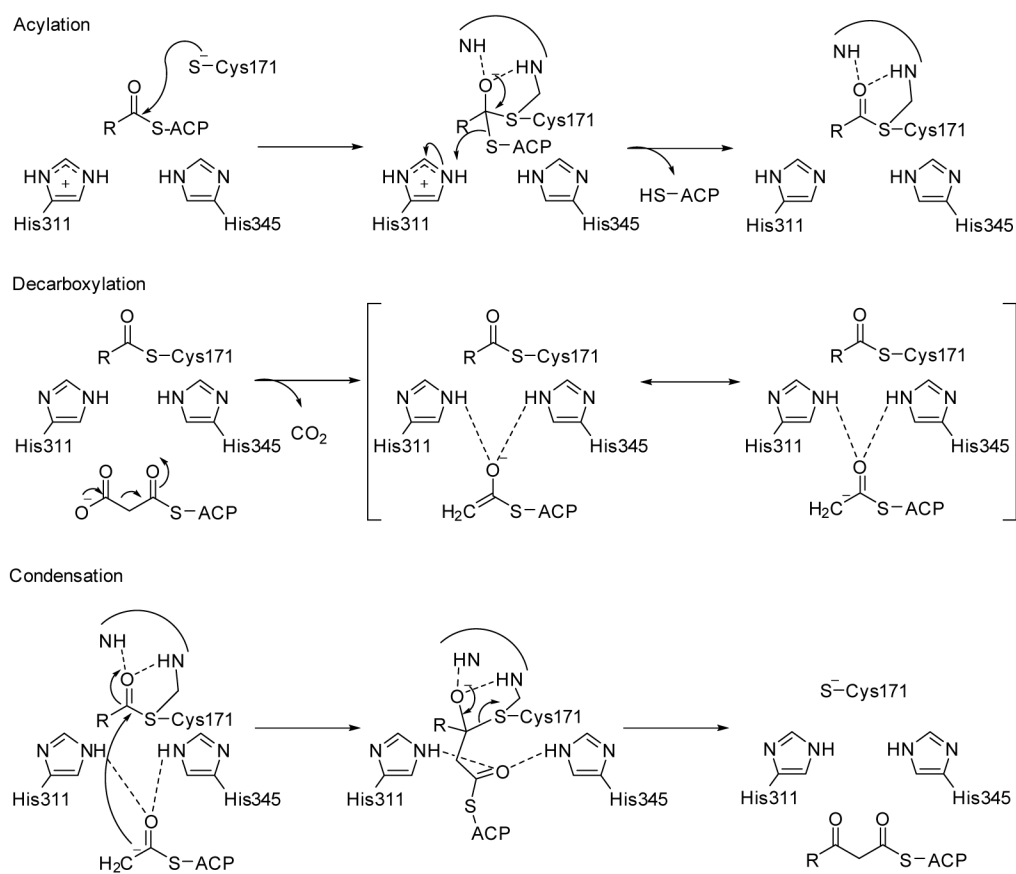
Figure 2.
Overall structure of KasA. Three residues in the active site, Cys171, His311 and His345, important for catalysis are highlighted in bond representation.
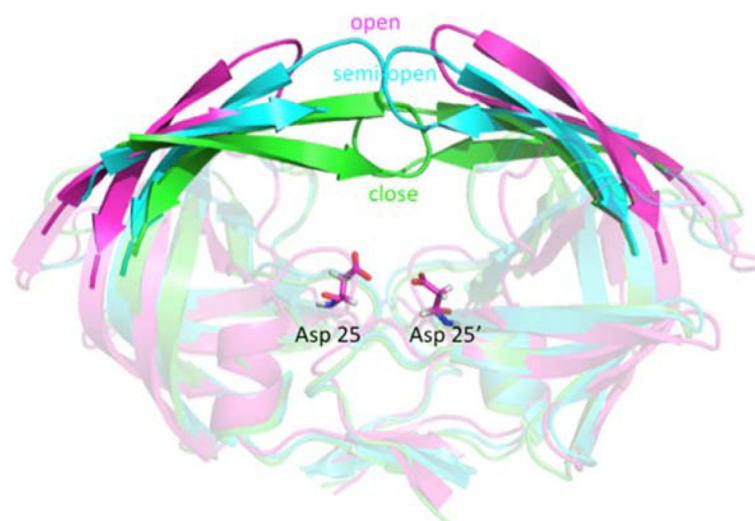
Figure 3.
Structure of HIV-1 protease. Pink, cyan, and green indicate open, semi-open and closed flap conformation, respectively. The residues Asp-25 and Asp-25′, highlighted in bond representation, are associated with chemical catalysis and their protonation states have been investigated.

Figure 4.
Overall chemical reactions in TRPS.

Figure 5.
Catalytic mechanism of KasA.

$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 \qquad \text{\textit{bond}}$$

$$+ \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \qquad \text{\textit{angle}}$$

$$+ \sum_{dihedrals} k_\phi (1 + \cos[n\phi - \gamma]) \qquad \text{\textit{dihedral}}$$

$$+ \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 \qquad \text{\textit{improper}}$$

$$+ \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] \qquad \text{\textit{van der Waals}}$$

$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \text{\textit{electrostatic}}$$

Figure 6.
Classical force fields used for MD simulations: (**Right**) potential energy terms in a force field; and (**Left**) energy function used to derive atomic forces for molecular movement. r is the bond length; θ is the atomic angle; φ is the dihedral angle; ω is the improper dihedral angle; r_{ij} is the distance in between atom i and j; k_r, k_θ, k_φ, and k_ω are force constants; r_{eq}, θ_{eq} and ω_{eq} are equilibrium positions; the dihedral term is a periodic term characterized by a force constant (k_ω), multiplicity (n), and phase shift (γ); ε_{ij} is related to the Lennard–Jones well depth; r_m is the distance at which the potential reaches its minimum; q_i and q_j are the charges on the respective atoms; and ε_0 is the dielectric constant.
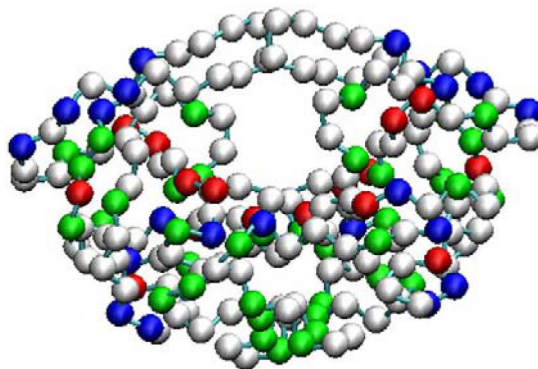
Bonded interactions energy function

$$E_{bond} = \frac{1}{2}k_b(b-b_0)^2, E_{morseb} = E_0\left[\left(1-e^{-k(r_{ij}-r_0)}\right)^2 - 1\right]$$

$$E_{angle} = \frac{1}{2}k_a(\theta-\theta_0)^2 + \frac{1}{3}k_a'(\theta-\theta_0)^3 + \frac{1}{4}k_a^*(\theta-\theta_0)^4$$

$$E_{dihedral} = \frac{1}{2}k_d(\phi-\phi_0)^2$$

Non-bonded interactions energy function

$$E_{intra} = 0.20708\left[\left(1-e^{-0.70711(r_{ij}-9.5)}\right)^2 - 1\right]$$

$$E_{inter} = 0.5\times\left[\left(\frac{r_i+r_j}{r_{ij}}\right)^8 - 1.5\left(\frac{r_i+r_j}{r_{ij}}\right)^6\right]$$

$$E_{elec} = \frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}}, \varepsilon_0 = 4r_{ij}$$

Figure 7.
Representation of HIV-1 protease in the coarse-grained model and energy function of the force field. (**Top**) Each bead represents a residue, and positively charged, negatively charged and polar residues are in blue, red and green, respectively. (**Bottom**) The potential energy function of the force field for BD simulations, where r is the bond length; θ is the atomic angle; φ is the dihedral angle; r_{ij} is the distance in between atom i and j; k_b, k_a, k_d, and E_0 are force constants; and b_0, θ_0 and φ_0 are equilibrium positions. r_i and r_j are the bead radius, and q_i and q_j are the charges on the respective bead. ε_0 is the dielectric constant. E_{morseb} is the Morse potential, a term for computing the bond energy between two consecutive beads. E_{intra} and E_{inter} are non-bonded interactions for $r_{ij} < 8$ Å and > 8 Å, respective. Detailed parameters are in ref [85,88].
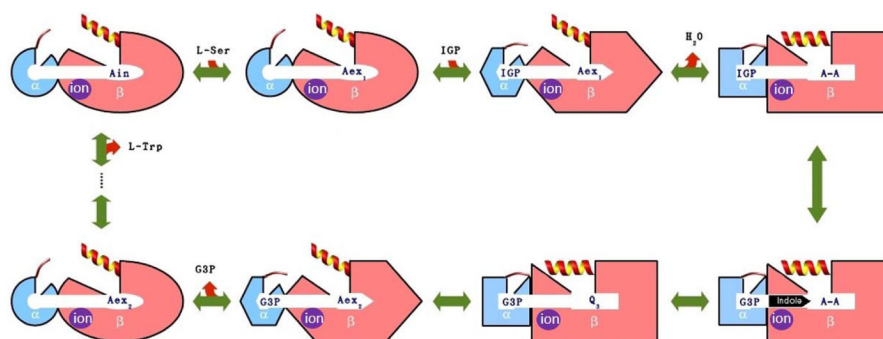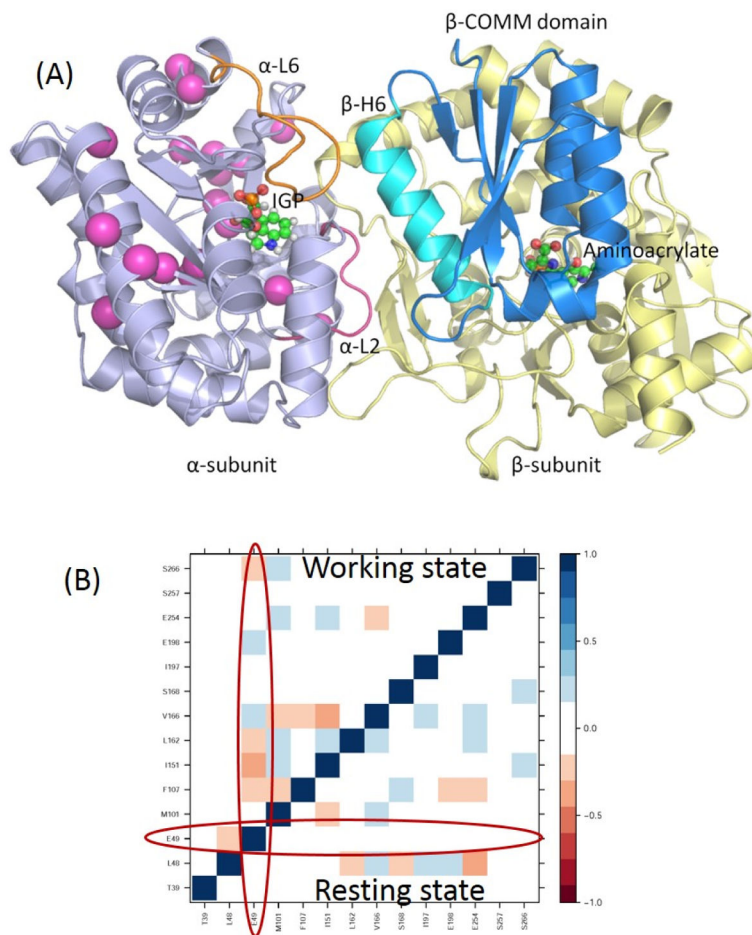
Figure 8.
Cartoon representation of synergistic regulation in TRPS. The α- and β-subunit are represented by blue and pink, respectively. Different shapes represent different protein conformations, and α-L6 (thin line), and β-Helix-6 (helix) control open/closed of the active sites during different steps of chemical reactions.
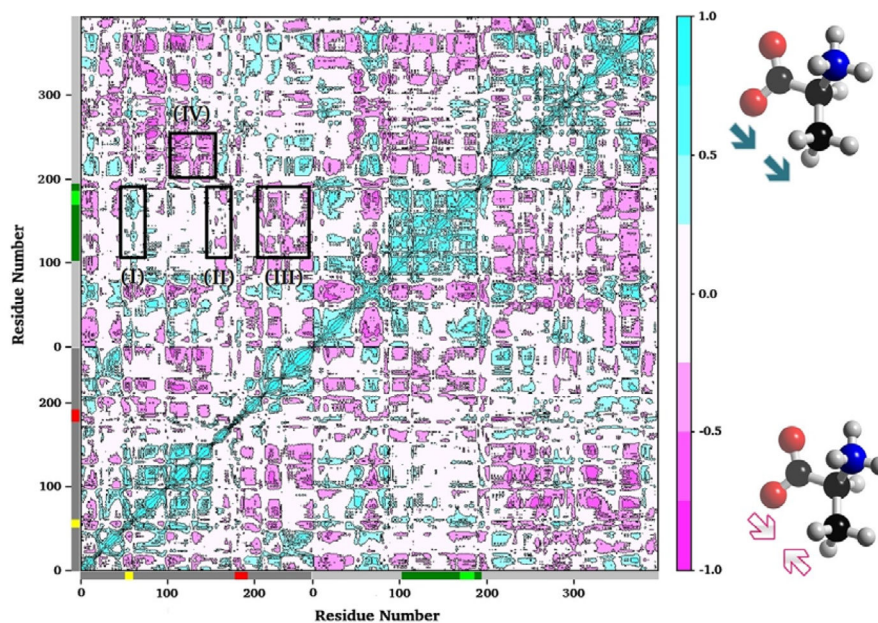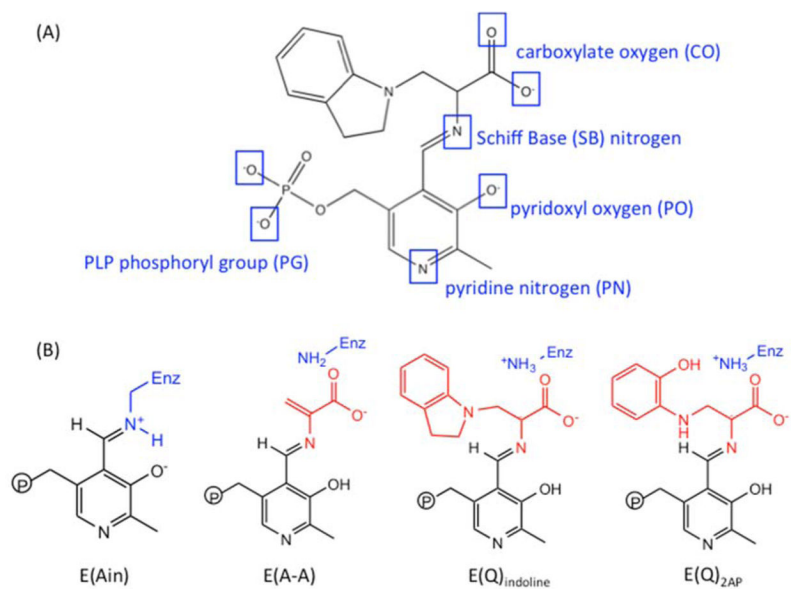
Figure 9.
Plot of the residue–residue cross-correlation matrix. The plot was generated from a MD simulation of an apo TRPS by using the Bio3D program. Only Cα atom of each residue was used in the analysis. Negative values (pink) indicate anti-correlated motions that Cα atoms move along opposite directions. Positive values (cyan) represent correlated motions that Cα atoms move along the same direction. Square I shows the correlations between α-L2 and the COMM domain. Squares II and III show that residues of the two edges of α-L6 have anti-correlation motions with the COMM domain. Another example for correlated motions between the α/β interfaces is shown in square IV.

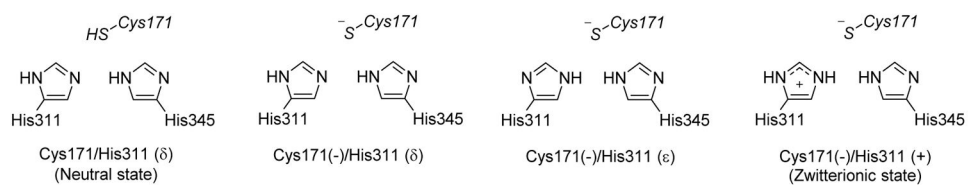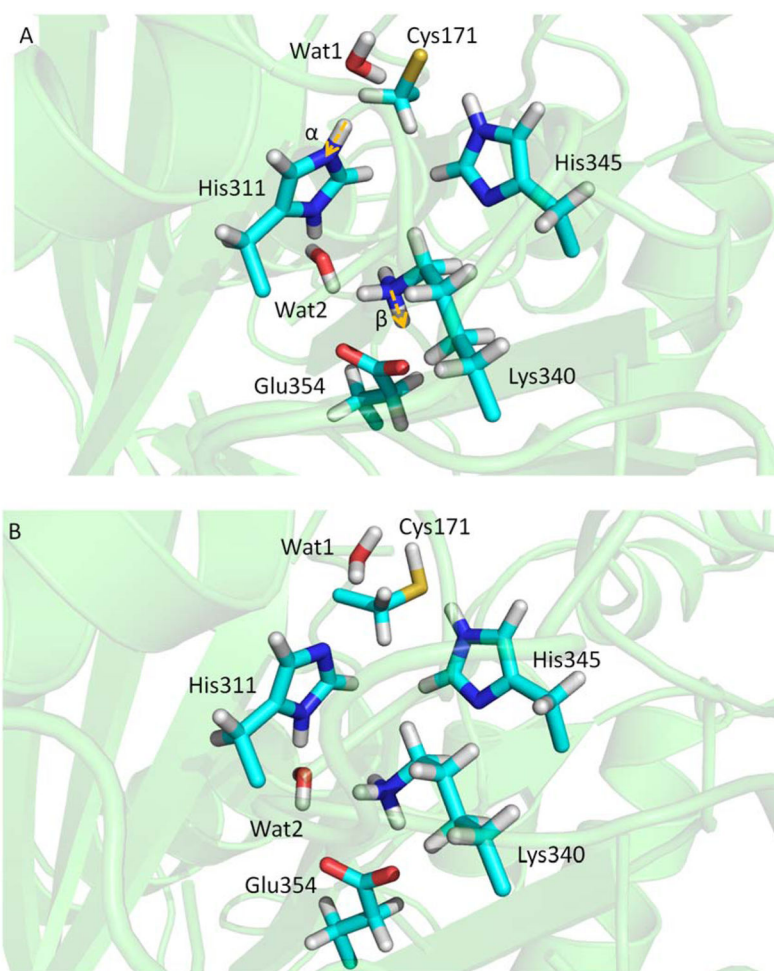**Figure 10.**
A cluster of residues of the TRPS α-subunit and their cross-correlation map: (**A**) residues identified by CHESCA that show correlated motions in the working state are presented as pink spheres; and (**B**) a cross-correlation map generated from a 100-ns MD simulation for the isolated α-subunit of TRPS from *Salmonella typhimurium.* The residues are shown in (**A**). The program T-Analyst was used to compute the side chain dihedral angles and their correlations with other residues. If a residue has more than one side chain dihedral angle, only the one closest to the protein backbone is used in the analysis. The lower triangle is for the ligand-free (resting) state, and the upper triangle is from the substrate-bound (working) state. The circles indicate the cross-correlations of Glu49 to other residues in the resting and working state, with no significant correlations with the residues in the resting state.
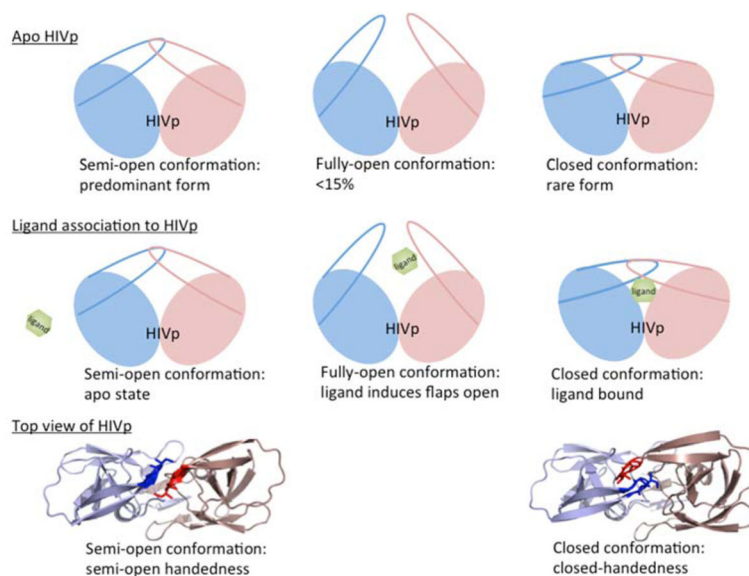
Figure 11.
(**A**) Potential sites of protonation on an indoline quinonoid substrate; and (**B**) protonation states of four PLP substrates: E(Ain), E(A-A), E(Q)$_{indoline}$ and E(Q)$_{2AP}$.

Figure 12.
Four possible protonation states of catalytic residues of KasA.

Figure 13.
(**A**) Conformations of key residues of KasA in the resting state with protonation Cys171(−)/His311(+); and (**B**) conformations of key residues of KasA in the resting state with protonation Cys171/His311(δ).

Figure 14.
Cartoon representation of conformational rearrangements of HIV-1 protease. (**Top**) Three major HIV-1 protease conformations have been widely discussed. Semi-open conformations are the dominant form in the free state; however, fully-open and closed conformations are rare. (**middle**) A ligand binding to HIV-1 protease. The protein conformations change from semi-open, fully-open to closed forms. (**bottom**) The top view of HIV-1 protease shows that semi-open and closed conformations are with semi-open and closed handedness, respectively.