Taylor & Francis
Taylor & Francis Group

Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms

Li-Quan Yang[a,b], Peng Sang[a], Yan Tao[a], Yun-Xin Fu[a,c], Ke-Qin Zhang[a,d], Yue-Hui Xie[e]* and Shu-Qun Liu[a,d,f]*

[a]*Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, P.R. China;* [b]*College of Agriculture and Biological Science, Dali University, Dali 671003, P.R. China;* [c]*Human Genetics Center, School of Public Health, The University of Texas Health Science Center, 1200 Herman Pressler, Room E453, Houston, TX, USA;* [d]*Southwest Biological Diversity Laboratory, Kunming Branch of Chinese Academy of Sciences, Kunming 650223, P.R. China;* [e]*Teaching and Research Section of Computer, Department of Basic Medical, Kunming Medical College, Kunming 650031, P.R. China;* [f]*Key Laboratory for Animal Genetic Diversity and Evolution of High Education in Yunnan Province, Kunming 650091, P.R. China*

Communicated by Ramaswamy H. Sarma

Proteins are dynamic entities in cellular solution with functions governed essentially by their dynamic personalities. We review several dynamics studies on serine protease proteinase K and HIV-1 gp120 envelope glycoprotein to demonstrate the importance of investigating the dynamic behaviors and molecular motions for a complete understanding of their structure–function relationships. Using computer simulations and essential dynamic (ED) analysis approaches, the dynamics data obtained revealed that: (i) proteinase K has highly flexible substrate-binding site, thus supporting the induced-fit or conformational selection mechanism of substrate binding; (ii) Ca^{2+} removal from proteinase K increases the global conformational flexibility, decreases the local flexibility of substrate-binding region, and does not influence the thermal motion of catalytic triad, thus explaining the experimentally determined decreased thermal stability, reduced substrate affinity, and almost unchanged catalytic activity upon Ca^{2+} removal; (iii) substrate binding affects the large concerted motions of proteinase K, and the resulting dynamic pocket can be connected to substrate binding, orientation, and product release; (iv) amino acid mutations 375 S/W and 423 I/P of HIV-1 gp120 have distinct effects on molecular motions of gp120, facilitating 375 S/W mutant to assume the CD4-bound conformation, while 423 I/P mutant to prefer for CD4-unliganded state. The mechanisms underlying protein dynamics and protein–ligand binding, including the concept of the free energy landscape (FEL) of the protein–solvent system, how the ruggedness and variability of FEL determine protein's dynamics, and how the three ligand-binding models, the lock-and-key, induced-fit, and conformational selection are rationalized based on the FEL theory are discussed in depth.

Keywords: protein dynamic personalities; ligand binding; free energy landscape; structure–function relationship; driving force

1. Introduction

Proteins, an essential part of organisms, participate in virtually every process within cells and as such are the materials central to cellular function. The central dogma of structural biology is that a folded protein structure is essential for its biological function (Wright & Dyson, 1999). Understanding proteins' functions at the atomic level requires prediction or experimental determination of their structures. The determination of the protein three-dimensional structures by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy provides a solid basis for structural and functional analyses of

proteins. Thanks to the high-resolution X-ray crystallographic technique, since it has brought out the revolutionary in the field of the structural biology through producing numerous static structures, resulting in a surge in studies of the structure–function relationship of proteins (Henzler-Wildman & Kern, 2007). However, for the high-resolution X-ray crystallography, a general limitation is that the homogeneous crystal is needed. In addition, the limitation of the NMR spectroscopy techniques is that they have generally been confined to structural solution of small soluble proteins. Therefore, only a small number of the proteins will be suitable for experimental structure

---

*Corresponding authors. Email addresses: shuqunliu@gmail.com (S.-Q. Liu); mzhxyh@yahoo.com (Y.-H. Xie)

determination. Consequently, computational solution to the prediction of protein structure becomes the subject of intense activity. The computational structural prediction methodologies are often divided into three main categories: comparative modeling, fold recognition or threading, and *ab initio* folding methods (Fetrow, Giammona, Kolinski, & Skolnick, 2002). These methods involve the development of an energy function capable of identifying the most stable conformation of a protein and a scoring function for evaluating the predicted models.

It is important to keep in mind that proteins should not be regarded simply as static pictures, but rather as dynamic entities with the internal motions and the resulting conformational changes playing an essential role in their functions (Henzler-Wildman & Kern, 2007). In other words, the functional properties of a protein are decided not only by its relatively rigid structure but even more importantly, by its dynamic personalities, as characterized by the thermodynamics (the relative populations/probabilities/lifetimes of the conformational states) and the kinetics (conformational transition that leads to the population redistribution between these conformational states) of the protein under the free energy landscape (FEL) theory (Henzler-Wildman & Kern, 2007; Liu et al., 2012; Ma, Wolfson, & Nussinov, 2001). A full understanding of the structure–function relationship of proteins requires, therefore, analyses of their dynamic behaviors in addition to the static structure. It should also be noted that the dynamics of proteins are characterized not only by the timescale of the fluctuations but also by the amplitude and directionality of the fluctuations, and therefore, the FEL of a protein, which has many atoms, is highly multidimensional (Henzler-Wildman & Kern, 2007). The rugged free energy surface, which is located either on the tube part or at the bottom of the funnel-like FEL and is caused by the noncomplementary change between the entropy and enthalpy of the protein–solvent system, dictates the distribution of relative populations of the protein conformational states/substates and kinetics of conversion between them (Liu et al., 2012). Furthermore, the enormous conformational space of proteins defines a very complex FEL which, in conjunction with the dynamic variability of the landscape, makes it difficult to study both the thermodynamics and kinetics of proteins at atomic resolution in real time using the experimental approaches.

Instead, the molecular dynamics (MD) simulation is an approach that is widely used for investigating the dynamics of proteins, even for modeling the FEL of the protein–solvent system. Using MD techniques, researchers can obtain not only the ultimate details concerning individual particle motions as a function of time, but also the global molecular motions of proteins at both spatial and temporal scale that are difficult to access experimentally (Karplus & McCammon, 2002). Most importantly,

although experiments can determine what is moving and how fast it is, MD simulations include the underlying forces and corresponding energies that can answer why things move (Henzler-Wildman & Kern, 2007). The global large-scale molecular motions can be obtained through applying the combination of MD simulation and essential dynamics (ED) analysis technique (Amadei, Linssen, & Berendsen, 1993; Tao, Rao, & Liu, 2010).

Another computational simulation method that is particularly useful for investigating the large-scale concerted motions is the CONCOORD (de Groot et al., 1997), which generates an ensemble of mutually uncorrelated protein conformers that fulfill a set of predefined upper and lower interatomic distance limits. Such limits are obtained through calculation of the interatomic distance and prediction of the interaction strength of the involved atoms. Although detailed information on the timedependence of atomic motion within protein cannot be reproduced, it has been shown that CONCOORD is capable of more efficiently sampling the conformational space, especially for the large systems, where the conventional MD simulation would require unfeasible large amount of computer time (de Groot et al., 1997; Mello, de Groot, Li, & Jedrzejas, 2002).

In order to demonstrate the relationship between the dynamics of proteins and their functional properties, we review some examples of the research performed in our laboratory. These include the investigations of the molecular motions and dynamic behaviors of serine protease proteinase K and HIV-1 gp120 envelope glycoprotein in relation to their biological functions using computational simulation techniques. We focus mainly on the following aspects: (i) dynamic behavior and large concerted motions of proteinase K (Liu, Meng, Fu, & Zhang, 2010); (ii) changes in molecular motions of proteinase K upon substrate binding (Tao et al., 2010); (iii) influence of Ca^{2+}-binding on the molecular motions of proteinase K (Liu, Meng, Fu, & Zhang, 2011); and (iv) effect of amino acid mutations on the molecular motions of HIV-1 gp120 glycoprotein (Liu, Liu, & Fu, 2007).

Proteases are enzymes that can hydrolyze peptide bonds in other proteins through catalytic reaction. Serine proteases (EC 3.4.21) are present in almost all organisms and their catalytic reactions can bring about diverse functional consequences, such as digestion, immune response, blood coagulation, and reproduction (Hedstrom, 2002). These enzymes exist as two families: the trypsin (EC 3.4.21.4) and the subtilisin (EC 3.4.21.14) families. Serine protease proteinase K (EC 3.4.21.64) from the fungus *Tritirachium album* limber belongs to the subtilisin family, which has attracted intensive research interest from the academic, industrial, and agricultural communities due to: (i) the ready amenability to structural and functional investigation (Wells & Estell, 1988); (ii) the wide application such as protein-degrading

components in washing powders (Shaw, 1987); and (iii) the potential application as bio-control agents against parasite (Liu et al., 2011a; Liu, Meng, Yang, Fu, & Zhang, 2007). Such intensive research efforts revealed that the subtilisin family members have the same catalytic triad Asp–His–Ser as the members of the trypsin family. In this triad, the Ser functions as the primary nucleophile and the His plays a dual role as the proton acceptor and donor at different stages in the reaction. The role of the Asp is believed to bring His into the correct orientation to facilitate the nucleophilic attack by Ser (Dodson & Wlodawer, 1998). Nevertheless, the detailed questions on many dynamic aspects of enzymatic mechanism such as the substrate binding, orientation, catalysis, product release, and how they are regulated remain unanswered. In addition, the mechanism underlying some experimental phenomena exhibiting that Ca^{2+} binding enhances the thermal stability and substrate affinity of proteinase K but does not affect its catalytic activity (Müller, Hinrichs, Wolf, & Saenger, 1994; Martin et al., 1997), needs to be elucidated.

Due to the grim situation of tackling the acquired immunodeficiency syndrome (AIDS) caused by the human immunodeficiency viral infection (HIV), the mechanisms of pathogenesis and the immune evasion of HIV have been extensively and intensively studied for decades. The HIV-1 exterior envelope glycoprotein gp120 undergoes a series of conformational rearrangements while sequentially interacting with the receptor CD4 and the coreceptor CCR5 or CXCR4 on the surface of human T cells to initiate viral entry (Berger, Murphy, & Farber, 1999; Eckert & Kim, 2001; Sattentau, 1998). Such a series of structural rearrangements of gp120, which play a crucial role in the HIV infection process, is governed by gp120's conformational dynamics. Both the crystal structures of the HIV-1 gp120 core in complex with the CD4 and antigen 17b (Kwong et al., 1998, 2000), and of the simian immunodeficiency virus (SIV) gp120 core alone (which is in the free form prior to binding to the CD4 molecule) (Chen et al., 2005) are known. A comparison between these two structures revealed unexpectedly extensive conformational rearrangements, in agreement with thermodynamics evidence suggesting that the binding of CD4 or antibodies to gp120 induces substantial structural rearrangements in gp120 (Myszka et al., 2000). Although these structural and biophysical studies provide invaluable insight into the function of gp120, further comparative analyses of molecular motions and dynamic behavior for gp120 in different conformational states are crucial for a full understanding of the causative mechanisms of conformational preference and transition between the two states.

This study is divided into two parts. Part I presents several case studies on dynamics of proteinase K and HIV-1 gp120. In these studies, the structural trajectories or ensembles of these two proteins were generated using MD or CONCOORD simulation approach, followed by geometrical property analyses to evaluate the structural stability or whether the systems reach the equilibrium during simulation. Furthermore, ED method was used to study the large concerted motions of these two proteins, including the influence of ligand bindings or residue mutations on molecular motions. Part II discusses in-depth, the mechanisms underlying the protein dynamics and protein–ligand–binding models based on the FEL theory. The combination of the dynamics data with the knowledge of mechanisms underlying protein dynamics will facilitate greatly the interpretation of the previously published data and further, the understanding of protein function.

2. Part I: case studies on dynamics and molecular motions of proteinase K and HIV-1 gp120

2.1. The computational and analysis methods used for studying protein dynamics

2.1.1. Molecular dynamics simulation

MD simulations, which are able to provide time-dependent information on dynamic dimension to structural data of biological macromolecules, are important tools for understanding the structure, dynamics, and function of biomolecules. Due to the improvements in computer hardware and simulation algorithms, MD simulations have been more and more widely used for studying dynamic properties of proteins, such as the conformational flexibility, dynamic variation of interaction, molecular motion, and equilibrium fluctuation of protein–solvent system on a certain timescale (Hansson, Oostenbrink, & van Gunsteren, 2002; Karplus & McCammon, 2002). In conjunction with the simulated annealing and replica exchange Monte Carlo techniques, MD simulations are also widely applied to refinement of the structures derived from X-ray crystallographic and NMR data or during comparative protein modeling (Kannan & Zacharias, 2009). Even more importantly, such combined MD approaches enhance substantially the efficiency of sampling for protein conformational space, leading to their applications in *ab initio* folding, investigating the protein folding mechanism, and constructing the FEL of the protein-solvent system. For details, refer to (Fetrow et al., 2002; Henzler-Wildman & Kern, 2007; Liu et al., 2012).

Nowadays, there are many computational software packages that can realize MD simulations, such as the CHARMM (Brooks et al., 1983), AMBER (Weiner & Kollman, 1981), GROMOS (Scott et al., 1999), NAMD (Phillips et al., 2005), and GROMACS (Berendsen, van der Spoel, & van Drunen, 1995; Lindahl, Hess, & van der Spoel, 2001). The GROMACS package is a useful collection of programs for MD simulation and subsequent analysis of the trajectory data. It does not have a force field of its own but is compatible with the force

fields such as GROMOS, OPLS (Jorgensen & Tirado-Rives, 1988), AMBER, and ENCAD (Levitt, Hirshberg, Sharon, & Daggett, 1995). In addition, it can handle polarizable shell models and flexible constraints and is well suited for parallelization on processor clusters (van der Spoel et al., 2005). In our researches, the GRO-MACS package was used to perform MD simulations on the proteinase K structural models with the GROMOS96 43a1 force field. The starting structures used for the MD simulation are the high-resolution crystal structures taken from Protein Database Bank (PDB) (Berman et al., 2000). The detailed setup of GROMACS for each MD simulation was described in the original papers (Liu et al., 2010, 2011b; Tao et al., 2010).

2.1.2.   CONCOORD simulation

CONCOORD simulation was performed to generate respective ensembles of the HIV-1 gp120 structural cores in both the unliganded and CD4-bound conformational states. CONCOORD works in two phases (de Groot et al., 1997). In the first phase, it derives a table of upper and lower distance constraints for various interatomic distances (i.e. covalent, ionic, and hydrogen bond, etc.). In the second phase, CONCOORD generates an ensemble of structures that fulfills the interatomic distance bound for all pairs of atoms. Starting from random coordinates, corrections are applied iteratively to the positions of those pairs of atoms that do not satisfy their specified distance constraints. This process is repeated until all constraints are satisfied. The ensemble generated in this way can be thought of as a set of uncorrelated structural snapshots, with the temporal order of frames scrambled. However, the CONCOORD ensemble can be treated and analyzed in exactly the same way as a structural trajectory produced by a MD simulation. In our studies (Liu, Fu, & Liu, 2007; S. Q. Liu, C. Q. Liu, & Y. X. Fu, 2007; S. Q. Liu, S. X. Liu, & Y. X. Fu, 2007, 2008), CONCOORD simulations generated 500 conformers for each structure models of HIV-1 gp120 core structures.

2.1.3.   Geometrical property analyses

The geometrical properties of the proteins during computational simulations, including the number of hydrogen bonds (NHB), number of native contacts (NNC), number of residues in the secondary structural elements (SSE), radius of gyration (Rg), solvent accessible surface area (SASA), and root mean square deviation (RMSD), were calculated using the programs g_hbond, g_mindist, do_dssp (Kabsch & Sander, 1983), g_gyrate, g_sas, and g_rms within the GROMACS package, respectively.

2.1.4.   Essential dynamics analysis

The ED method (Amadei et al., 1993) is a powerful tool for filtering large concerted motions from an ensemble of

structures derived from crystal structures or computer simulation. This method is based on the diagonalization of the covariance matrix of atomic fluctuations, yielding a set of eigenvectors and eigenvalues. The resulting eigenvectors indicate directions in a 3 N-dimensional (the N is the number of atoms used for constructing the covariance matrix) conformational space and describe the concerted fluctuations of the atoms along those directions. The eigenvalues are the measure of the mean square fluctuations of the system along the corresponding eigenvectors. The central hypothesis of ED is that only a few eigenvectors with large corresponding eigenvalues are important for describing the overall motions of a protein. ED analyses were performed on the MD trajectories or CONCOORD ensembles using the g_covar and g_anaeig programs within GROMACS and, only the C_α atoms were included in such analyses.

The DYNAMITE program (Barrett, Hall, & Noble, 2004) was utilized to generate the "porcupine plot", a way of graphical representation capable of describing vividly the motion of each atom held in an eigenvector. For instance, to visualize eigenvector 1, a cone is drawn for each residue starting from the C_α, and is projected in the direction of the component of the first eigenvector that corresponds to that residue with its length representing the motional amplitude. Given eigenvectors of the ED decomposition, a script was generated by DYNA-MITE to allow the molecular graphics program visual molecular dynamics (VMD) (Humphrey, Dalke, & Schulten, 1996) to plot these cones onto the protein.

2.1.5.   Combined essential dynamics analysis

A useful method for comparing the ED properties of two simulations on similar systems is the combined ED analysis (van Aalten et al., 1995). In this method, ED analysis can be performed on a combined trajectory constructed through concatenating individual trajectories of these systems. Analysis and comparison of the properties of different parts of the projection along the combined eigenvector provide a powerful way for evaluating similarities and differences in essential motions between these different trajectories. There are two main properties to be investigated: (i) the differences in the average values of the projections, which indicate that simulations have different average displacement along eigenvectors, i.e. a static difference in equilibrium fluctuations between individual average structures along that direction; (ii) the differences in the mean square displacement (MSD) of the projections, which can be used to study difference in dynamics/flexibility between equilibrium fluctuations along this direction.

2.1.6.   Essential subspace overlap analysis

If the eigenvectors obtained from the ED analysis are seen as vectors that span a complex 3 N-dimensional

space, then a few "essential" eigenvectors with the largest eigenvalues span a subspace that is called the "essential subspace", and all large concerted motions take place within this subspace. Previous studies (de Groot, Hayward, van Aalten, Amadei, & Berendsen, 1998; Mello et al., 2002; Merlino, Vitagliano, Ceruso, & Mazzarella, 2004; Vreede, van der Horst, Hellingwerf, Crielaard, & van Aalten, 2003) have shown that these large concerted motions usually play a crucial role in the function of proteins and, furthermore, the calculation of essential subspace overlap between different simulated systems can provide a valid way to assess their dynamic or functional similarity. Specifically, the value of the essential subspace overlap can be obtained through calculating the cumulative mean square inner product (CMSIP) between two selected subsets of eigenvectors from each ensemble, with overlap ranging from 0, when the eigenvector subsets are completely dissimilar, to 1 (or 100%) when they are identical. The number of eigenvectors used is typically chosen so as to represent a significant proportion of fluctuations in the simulation.

2.2. The molecular motions and dynamic personalities of proteinase K and HIV-1 gp120

2.2.1. Comparison of structural properties during simulations between substrate-free and substrate-complexed proteinases K

The structures of proteinase K have been resolved in a series of X-ray crystallographic studies. A high-resolution structure of proteinase K at 0.98 Å (PDB code 1IC6) (Kumar, Kaur, Perbandt, Eschenburg, & Singh, 2001) is shown in Figure 1(a). This structure was used

as the starting model for the substrate-free proteinase K MD simulation; the starting model for the substrate-complexed proteinase K MD simulation is the crystal structure with PDB code 3PRK at 2.2 Å resolution (Wolf et al., 1991), where the peptide substrate Ala–Ala–Pro–Ala is located between two segments of the enzyme, Gly100–Tyr104 and Ser132–Gly136, forming a three-stranded antiparallel -pleated sheet via intermolecular hydrogen bonds between the substrate and these two segments (Figure 1(b)). MD simulations on the substrate-free and substrate-complexed proteinases K were run for 40 ns to investigate the influence of substrate binding on the dynamics of proteinase K (Tao et al., 2010). Monitoring the C RMSD revealed that for both the simulations, only the 7–34 ns parts of MD trajectories commonly reached the equilibrium; as such, these 7–34 ns trajectories were selected to be used for the subsequent analyses.

Geometrical properties, i.e. NHB, SASA, NNC, Rg, RMSD, and SSE are calculated (Table 1) over the equilibrium trajectories of the substrate-free and substrate-complexed proteinases K in order to evaluate the structural stability during simulations and the changes in the stability of proteinase K upon substrate binding. The results show that both the free and complexed structures were stable during the MD simulations due to the stable comparable structural properties and, especially, the small values of the standard deviation (SD) from their average values.

In the case of the substrate-free proteinase K, the largest, moderate, and lowest backbone RMSD values were observed in the loops, the entire structure and the SSE, suggesting that the conformational fluctuation
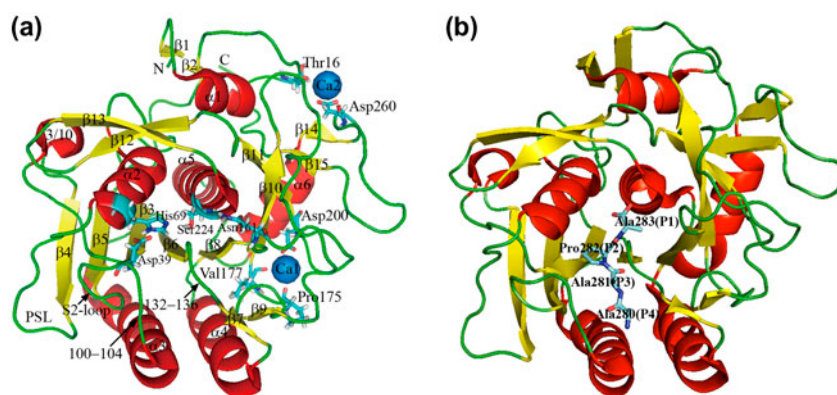

Figure 1. Ribbon structures of (a) substrate-free proteinase K (PDB code: 1IC6) and (b) substrate-complexed proteinase K (PDB code: 3PRK). The structures of the enzyme are color-coded with α helices in red, β sheets in yellow and loops in green. The residues of the catalytic triad (Asp39, His69 and Ser224), oxyanion hole (Asn161), calcium binding sites (only shown in (a); Ca1 site: Pro175, Val177 and Asp200; Ca2 site: Thr16 and Asp260) and peptide substrate (Ala280–Ala281–Pro282–Ala283) are shown in stick models with carbon atoms in cyan, oxygen atoms in red and nitrogen atoms in blue. According to Schechter and Berger nomenclature (Schecter & Berger, 1967), substrate residues Ala280, Ala281, Pro282 and Ala283 were designated as the P4, P3, P2 and P1 residues and their corresponding binding sites as S4, S3, S2 and S1 sites/pockets, respectively. The two calcium cations in (a) are denoted as blue balls. (a) and (b) were cited from (Liu et al., 2011b) and (Tao et al., 2010), respectively.

Table 1. Comparison between average geometrical properties (SD are shown in parentheses) of the substrate-free and substrate-complexed proteinases K during MD simulations. This table is cited from (Tao et al., 2010).

Proteinase K	NHB[a]	SASA[b] ($Å^2$)	NNC[c]	Rg[d] (Å)	RMSD[e] (Å)			SSE[i]		
					All bb[f]	SS bb[g]	Loop bb[h]	α helix	β sheet	Turn
Free	212	15,388.0	134,288	16.7	1.76	1.12	2.23	68	61	28
	(8)	(344.4)	(1244)	(0.74)	(0.10)	(0.09)	(0.13)	(3)	(4)	(5)
Complex	217	15,258.6	135,547	16.8	1.55	1.00	1.94	73	62	28
	(8)	(321.3)	(930)	(0.45)	(0.08)	(0.07)	(0.11)	(2)	(3)	(5)

[a]NHB.
[b]Total SASA.
[c]NNC.
[d]Rg.
[e]RMSD relative to respective starting structures.
[f]Backbone RMSD values of all residues.
[g]Backbone RMSD values of secondary structure elements.
[h]Backbone RMSDs of loop regions.
[i]Number of residues in corresponding secondary structure elements.

originates primarily from the highly flexible loop and/or link regions located between the relatively rigid SSE. Further B-factor analysis (Liu et al., 2010) demonstrated that many of the loops/links involving the largest conformational fluctuations are those residing at or near the substrate-binding site, such as residues 100–104, 132–136, 161–165, and 221–223. These regions participate in the formation of the S1 and S4 substrate pockets (Schecter & Berger, 1967), and therefore, we consider that their high conformational mobility likely facilitates substrate recognition and binding via the induced-fit (Koshland, 1958) or conformational selection mechanism (Berger, Weber-Bornhauser et al., 1999; Foote & Milstein, 1994; Monod, Wyman, & Changeux, 1965; Tsai, Ma, Sham, Kumar, & Nussinov, 2001). In the substrate-complexed proteinase K, the substrate pockets S1 and S4 show reduced mobility relative to the free proteinase K. Nevertheless, their conformational flexibility is still larger than the SSEs, implying a role in facilitating orientation of the bound substrate for catalysis to take place and the subsequent product release (Tao et al., 2010).

Although the differences in geometrical properties between the free and complexed proteinases K are minor, the observed subtle increases or reductions in geometrical properties can reflect the change in stability and compactness of the structural packing upon substrate binding. For instance, the substrate-complexed proteinase K has more NHB, NNC, and SSE than the free form, suggesting the enhanced inter-atomic interactions in protease structure when substrate is present. The reduction in the SASA of proteinase K upon substrate binding suggests that the overall molecular surface is shrunk by the bound substrate. In addition, the lower backbone RMSD values of the substrate-complexed form than those of the free form also suggest that the stability of the enzyme structure is enhanced upon substrate binding. These results, in conjunction with relatively larger SD of the geometrical properties in the free proteinase K than in the complexed form, suggest that the proteinase K in complex with the peptide substrate assumes a more stable and compact conformation than the free proteinase K. It is possible that the highly rigid structural core of proteinase K has evolved as a protective measure against autolysis, while the highly flexible substrate-binding regions play roles in both modulating the thermodynamics and kinetics of the enzyme–substrate interaction so as to efficiently recognize and bind to the substrate, and in allowing interactions with multiple sequence motifs of the substrates, thus broadening the substrate specificity of proteinase K.

2.2.2. The influence of Ca^{2+} on molecular motions of proteinase K

The native proteinase K contains two calcium cations, which were named Ca1 and Ca2 (Figure 1(a)). The Ca1 is tightly bound with high affinity, whereas the Ca2 is weakly bound with lower affinity (Bajorath, Raghunathan, Hinrichs, & Saenger, 1989; Betzel, Pal, & Saenger, 1988; Kumar et al., 2001). It has been reported that the removal of Ca^{2+} decreased the protease's thermal stability and to some extent the substrate affinity (Martin et al., 1997; Müller et al., 1994) but did not affect catalytic activity of the enzyme (Müller et al., 1994). In order to elucidate the mechanism behind these experimental phenomena, we have performed MD simulations on proteinases K with and without the calcium to investigate the effect of calcium on the structural stability and molecular motions of this protease (Liu et al., 2011b).

The comparison between geometrical properties of these two forms of proteinases K during simulations shows slight reductions in the NHB, NNC, and SSE, and slight increases in SASA and Rg upon Ca^{2+} removal. In conjunction with the increased B-factors averaged over

all C_{α} atoms upon Ca^{2+} removal, these results reflect an increase in the overall structural flexibility of the proteinase K when Ca^{2+} is removed. In other words, the Ca^{2+}-free proteinase K assumes a more open and relaxed conformational state than the Ca^{2+}-bound form, thus explaining the experimentally observed lower thermostability of the Ca^{2+}-free proteinase K relative to the Ca^{2+}-bound proteinase K.

A close comparison between B-factors indicates that the Ca^{2+} removal mainly increases the amplitude of mobility in the loops/links and the N- and C-terminal regions, whereas the structural rigidity of the α-helices and central β sheet seems not to be affected significantly (Figure 2) (Liu et al., 2011b). Interestingly, for the segments that participate in the formations of the substrate-binding site, we observed an increase in amplitude of fluctuations in the Ca^{2+}-bound proteinase K compared with the Ca^{2+}-free form. Since many studies have shown that the high substrate-binding affinity requires high structural flexibility at the substrate-binding site (Gudjónsdóttir & Asgeirsson, 2008; Liu et al., 2011a, 2012; Martin et al., 1997; Miller & Agard, 1999; Peters & Bywater, 1999; Peters, Frimurer, Andersen, & Olsen, 1999; Wade, Gabdoulline, Ludemann, & Lounnas, 1998), we consider that the reduced conformational mobility in the substrate binding site caused by Ca^{2+} removal may explain why the Ca^{2+}-free proteinase K has lower substrate-binding affinity than the Ca^{2+}-bound proteinase K. The requirement for high conformational flexibility in the binding site can be explained in terms of the induced-fit and/or conformational selection model. In the induced-fit mechanism, the high flexibility of the substrate-binding site allows conformational adjustment to occur easily to suit the already bound substrate,
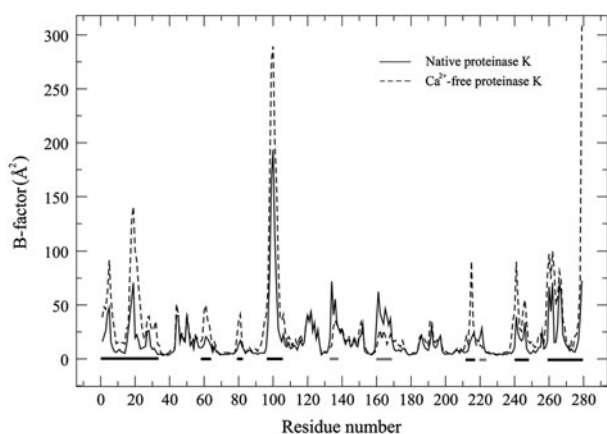
thereby facilitating the formation of the bonds capable of stabilizing the enzyme–substrate complex. In the conformational selection mechanism, the high flexibility of the substrate-binding regions allows the protease to exist in ensembles of different conformational states, thereby being advantageous for the substrate to bind selectively to the enzyme molecule with suitable conformational state. These two binding mechanisms will be discussed in detail in Part II.

Also worth noting is that the differences in B-factors of the catalytic triads between the Ca^{2+}-bound and Ca^{2+}-free proteinases K are only minor, indicating that Ca^{2+} removal has minor influence on the thermal motions of the catalytic triad. Further ED analyses also revealed that the catalytic triad residues do not involve the large concerted fluctuations of proteinase K, either in the Ca^{2+}-bound or in the Ca^{2+}-free form (Liu et al., 2011b); the estimate of the essential subspace overlap between the first 30 eigenvectors of the Ca^{2+}-bound and Ca^{2+}-free proteinases K gave a value of ∼75%, an indication of a high degree of similarity in the conformational freedom between these two forms of proteinase K. These results could be the mechanical reasons for why the removal of Ca^{2+} does not affect dramatically the catalytic activity of proteinase K (Müller et al., 1994).

We also performed a combined ED analysis to investigate the differences in molecular motions between these two forms of proteinase K (Liu et al., 2011b). Figure 3 shows the projections of concatenated trajectories onto the first three combined eigenvectors as well as the distributions of these projections. Only in the case of eigenvector 1 can the significant difference in large concerted motions be found, implying that the equilibrium conformations within the essential subspace of the first eigenvectors of these two forms of proteinase K may reside in different free energy wells that are separated by a large energy barrier. Such different equilibrium conformations could also be used to account for the difference in substrate affinity between the Ca^{2+}-bound and Ca^{2+}-free proteinases K. For example, binding of the calcium cations shifts the conformational equilibrium from the Ca^{2+}-free towards Ca^{2+}-bound state. Because the latter has higher flexibility within the substrate-binding regions, this will provide advantage for conformation changes, thus facilitating substrate recognition and binding via induced-fit or conformational selection process. This will be discussed later in Section 3.2.2., Part II of this paper.

Figure 2. Comparison between C_{α} B-factors calculated from MD simulations of the Ca^{2+}-bound (solid lines) and Ca^{2+}-free (dashed lines) proteinases K. The structural segments demonstrating significant B-factor increase and reduction upon Ca^{2+} removal are indicated with black and gray bars along the horizontal axis, respectively. This figure is cited from (Liu et al., 2011b).

2.2.3. Changes in molecular motions of proteinase K upon substrate binding

In order to investigate the change in large concerted motions of proteinase K caused by substrate binding, combined ED analysis were performed on the concatenated trajectories of the substrate-free and
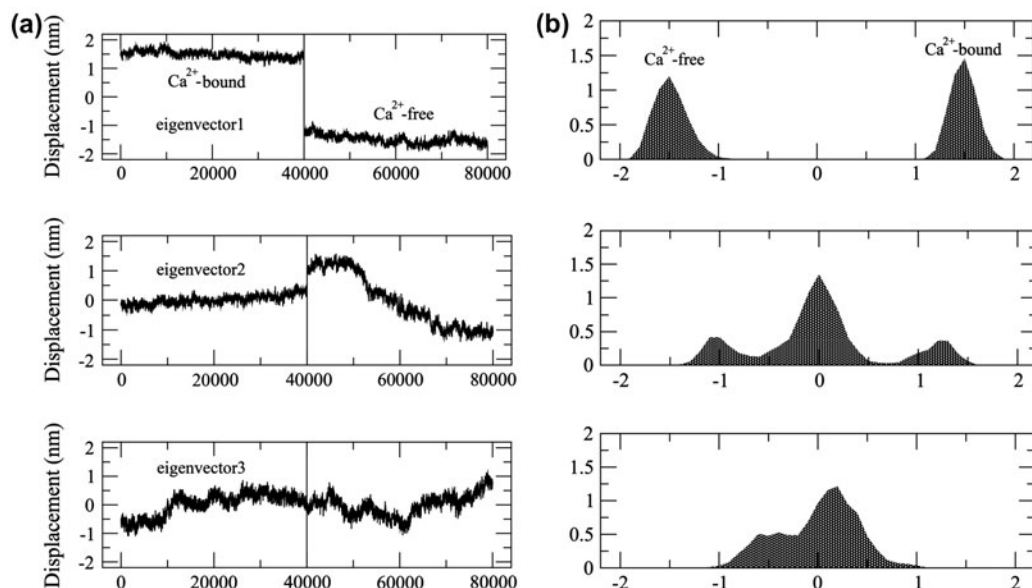
Figure 3. (a) Projections of the concatenated trajectory of the Ca^{2+}-bound and Ca^{2+}-free proteinase K simulations onto the first three "combined" eigenvectors and (b) the distributions of these projections. Distinct distribution can merely be found in the projection along the eigenvector 1. This figure is cited from (Liu et al., 2011b).

substrate-complexed proteinase K simulations (Tao et al., 2010). The result shows that, like what has been observed in the combined ED analysis of the Ca^{2+}-free and Ca^{2+}-bound proteinase K simulations (Liu et al., 2011b), only in the case of the first eigenvector projection can the equilibrium fluctuations be clearly separated. Therefore, such molecular motions occur within different free energy wells over the surface of the FEL, in which the well-separated average/equilibrium structures (conformational states) of the substrate-free and substrate-complexed proteinases K are located.

Figure 4(a) shows the comparison between the extreme structures extracted from the first combined eigenvector projection, from which the static conformational differences between the free and complexed proteinases K can be clearly observed. Although such an image may provide intuitive information about the change in equilibrium fluctuation of proteinase K upon substrate binding, it should be kept in mind that the linear interpolations between the two extremes are not the conformational transition pathway connecting these two states but emphasize only the conformational
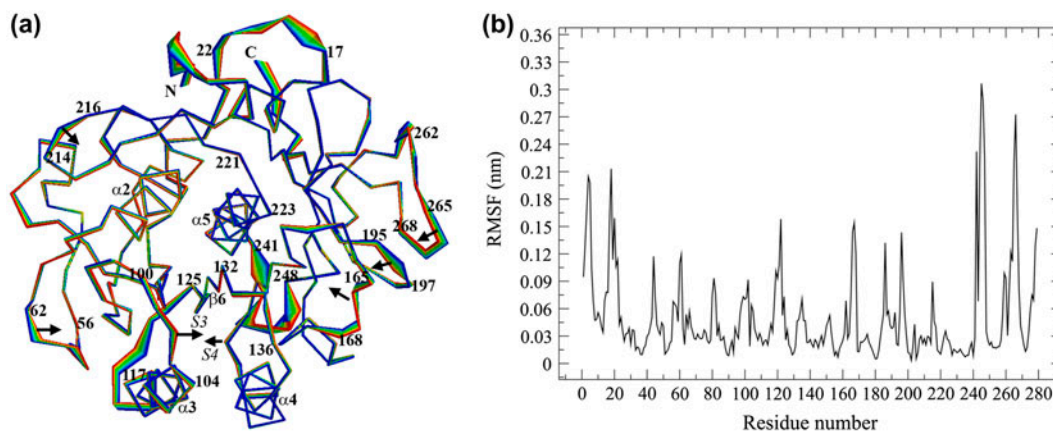

Figure 4. Conformational changes of proteinase K upon peptide substrate binding. (a) Two extreme structures extracted from the projection along the first "combined" eigenvector. The linear interpolations between these two extremes are shown with colour transition from blue (the substrate-free proteinase K) to red (the substrate-complexed proteinase K) to stress the primary difference in conformational state between these two forms of proteinase K, but do not represent the transition pathway. (b) C_{α} RMSF calculated along first "combined" eigenvector projection as a function of residue number. This figure is cited from (Tao et al., 2010).

differences between them. Figure 4(b) shows the C_α root mean square fluctuations (RMSF) of the first combined eigenvector, revealing that, upon substrate binding, the structural regions exhibiting large conformational displacements are located both in some surface-exposed loops/links (such as the residues 121–122, 166–168, 186, 196, 241–248, 258–268) and the N- and C-termini, whereas the majority of the secondary structure components exhibits small shifts. This suggests that the substrate binding has relatively minor influence on dynamics of the internal rigid core but large effect on the external loops of proteinase K. Of particular note is that these loops include not only those located within and near the substrate binding site, but also the ones that are located opposite (such as residues 241–248) and far away from the binding site (such as loop residues 117–125). We consider that the high mobility of the structural regions located opposite or far away from the binding site can modulate the dynamics/structural changes of the substrate-binding site through hinge-bending mechanism or allosteric propagation. The rigid structural segments connecting the high flexibility loops and substrate binding site may serve as hinge points or intermediaries to transmit or mediate the signal of conformational changes. It is reasonable, therefore, to conclude that the conformational changes occurring opposite or far away from the substrate binding site can regulate the process of substrate binding.

We also note that although many surface-exposed loops show large conformational displacements upon substrate binding, no significant structural shift can be observed in the catalytic triad residues Asp39, His69, and Ser224 (Figure 4(a) and (b)), revealing that the architecture of the triad is well retained, whether in the presence or absence of the peptide substrate. The conserved interactions such as hydrogen bonding and electrostatic interactions between the three conserved catalytic triad residues, in conjunction with their positional invariability in the enzyme structure, contribute to the stability of the catalytic triad.

The large concerted motions of proteinase K described by the first few eigenvectors (Figure 5) could be proposed to be related to its functions (Tao et al., 2010). For example, for the substrate-complexed proteinase K, the first ranked motion described by eigenvector 1 enlarges the substrate pockets S1 and S4, thereby providing more space for the P1 and P4 substrate residues to suit the requirements for accommodation and subsequent orientation of these two residues. The second ranked motion described by eigenvector 2 not only narrows the bottoms but also opens the lids of both S1 and S4 pockets, thus leading to lose contacts between the pockets and the substrate and likely facilitating substrate release. The twist motion of the binding groove described by eigenvector 3 may facilitate position adjustments of the P2–P4

substrate residues in their respective binding subsites/ pockets into their bound orientations for catalysis. The large concerted motion described by eigenvector 4 reduces the bottom size of S4 pocket and as thus could be related to P4 substrate residue release. Taken together, the first few modes of the large concerted motions can result in different consequences of conformational changes of the substrate-binding pockets and therefore, could be related to the processes of binding, orientation and release of the substrate.

2.2.4. The effect of amino acid mutations on molecular motions of HIV-1 gp120

The CONCOORD simulation and ED analysis were used to study the influence of amino acid mutations on the molecular motions of HIV-1 gp120 envelope glycoprotein (S. Q. Liu, C. Q. Liu, et al., 2007). The seven gp120 models composed of the structural core and V3 loop, i.e. gp120-CD4 complex, CD4-free gp120, unliganded gp120, CD4-free 375 S/W mutant, unliganded 375 S/W mutant, CD4-free 423 I/P mutant, and unliganded 423 I/P mutant were generated using comparative modeling technique with MODELLER software package (Sali & Blundell, 1993). The structural templates for the CD4-bound and unliganded states of gp120 and for the V3 loop were crystal structures with PDB codes 1G9 M (chain G) (Kwong et al., 2000), 2BF1 (Chen et al., 2005) 1CE4 (Vranken, Budesinsky, Fant, Boulez, & Borremans, 1995), respectively. The starting sequence of gp120 was cleaved from the HIV-1 HXBc2 isolate gp160 precursor (Swiss-Prot accession No. P04578) obtained from the Swiss-Prot protein sequence database (Bairoch et al., 2005). The gp120–CD4 complex model represents the gp120 molecule in complex with the receptor CD4, where the gp120 is in CD4-bound conformational state; the CD4-free model represents that the gp120 is in CD4-bound state but without including the CD4 molecule; the unliganded gp120 model represents gp120 in a state prior to CD4 binding; the CD4-free 375 S/W is a mutant model in the CD4-free state where the Ser at 375 position was replaced by a Trp; the unliganded 375 S/W is a mutant in the unliganded state within which the Ser375 was replaced by a Trp; the CD4-free 423 I/P model represents a mutant in the CD4-free state with Ile423 being replaced by a Pro; the unliganded 423 I/P model represent a mutant in the unliganded state with a Pro substituting for the Ile at 423 position. These seven gp120 models were used as starting structures for the CONCOORD simulations to generate respective structural ensembles.

ED analysis followed by essential subspace overlap calculations were performed on the CONCOORD ensembles to investigate the variations in molecular motions
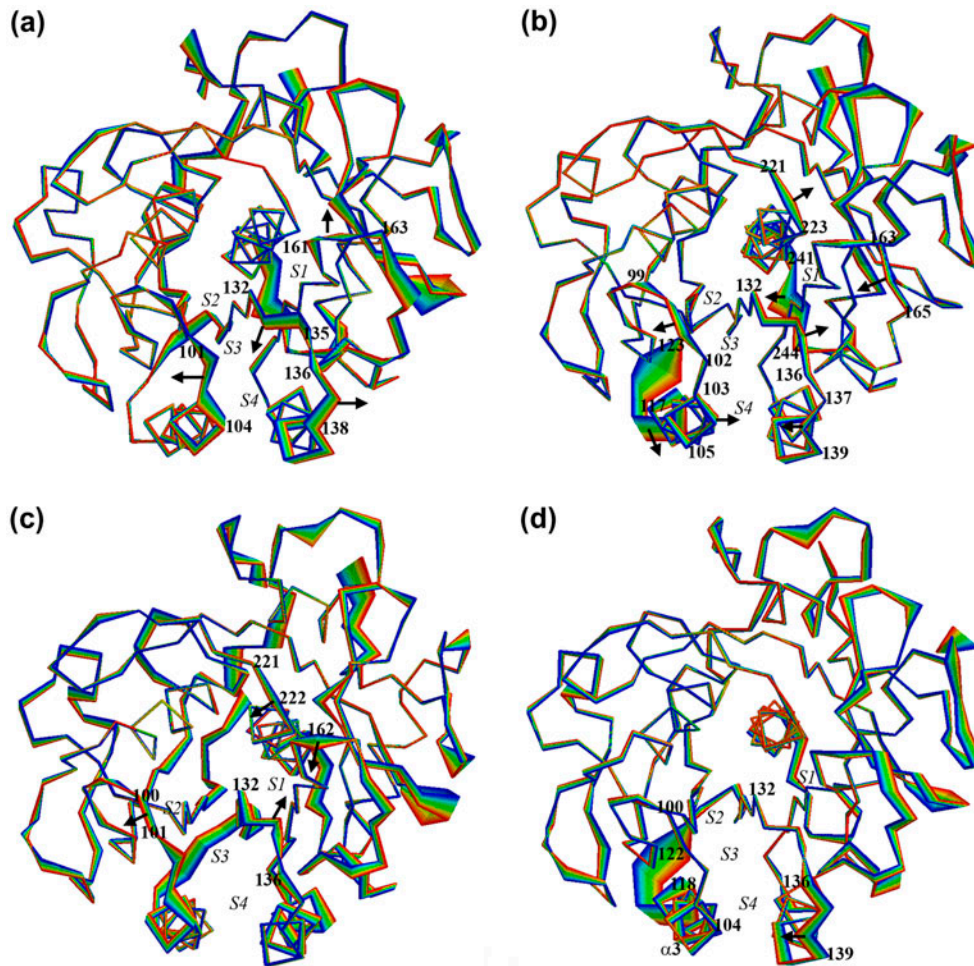
Figure 5.   Large concerted motions of the substrate-complexed proteinase K described by the first four eigenvectors: (a) eigenvector 1, (b) eigenvector 2, (c) eigenvector 3, and (d) eigenvector 4. The linear interpolations between the two extremes are colored from blue to red to highlight the conformational differences between them. The structural segments involving large concerted motions that bring out the dynamic variations of the substrate-binding pockets/subsites are labeled, with the arrows indicating the movement direction. This figure is cited from (Tao et al., 2010).

caused by point mutations, to identify the preference for conformation transition between gp120 mutants and ultimately, to ascertain which conformational states that the 375 S/W and 423 I/P mutants prefer to assume.

The results of the ED analyses reveal that the first 10 eigenvectors, especially the first four eigenvectors, emerge with appreciable freedom and contribute a significant proportion to the total mean square fluctuation, indicating that the first few eigenvectors describe the most significant motions of gp120 molecules. Figure 6 shows the porcupine plots of the first two motional modes of the wild-type and mutated gp120. Although the large concerted motions described by the same ranked eigenvectors in all the simulated models are similar (e.g. the twist of the inner domain with respect to the outer domain), the variations in the twisting directions of certain mobile units (or vortices as shown in Figure 6) caused by point mutants may associate with the conformational transition

from one state to the other or provide capability for certain models to retain their current conformations. The comparative results (for details, see (S. Q. Liu, C. Q. Liu, et al., 2007)) indicate that the 375 S/W mutants have the propensity to assume a conformation close to the CD4-bound state, whereas the mutants 423 I/P tend to adopt a conformation resembling the unliganded state. In addition, our ED analyses also reveal complex combinations of variable motional modes for certain structural components of gp120, that is, the bridging sheet, V1/V2 stem, and V3 loop can move in concert with either the inner domain or outer domain. We consider that such perplexing modes may, on one hand, be related to the dynamic aspects of the mechanism of gp120-receptor association, on the other hand, provide advantage for immune evasion, that is, the failure of immune system to produce potent neutralizing antibodies against the CD4 and coreceptor binding sites of gp120.
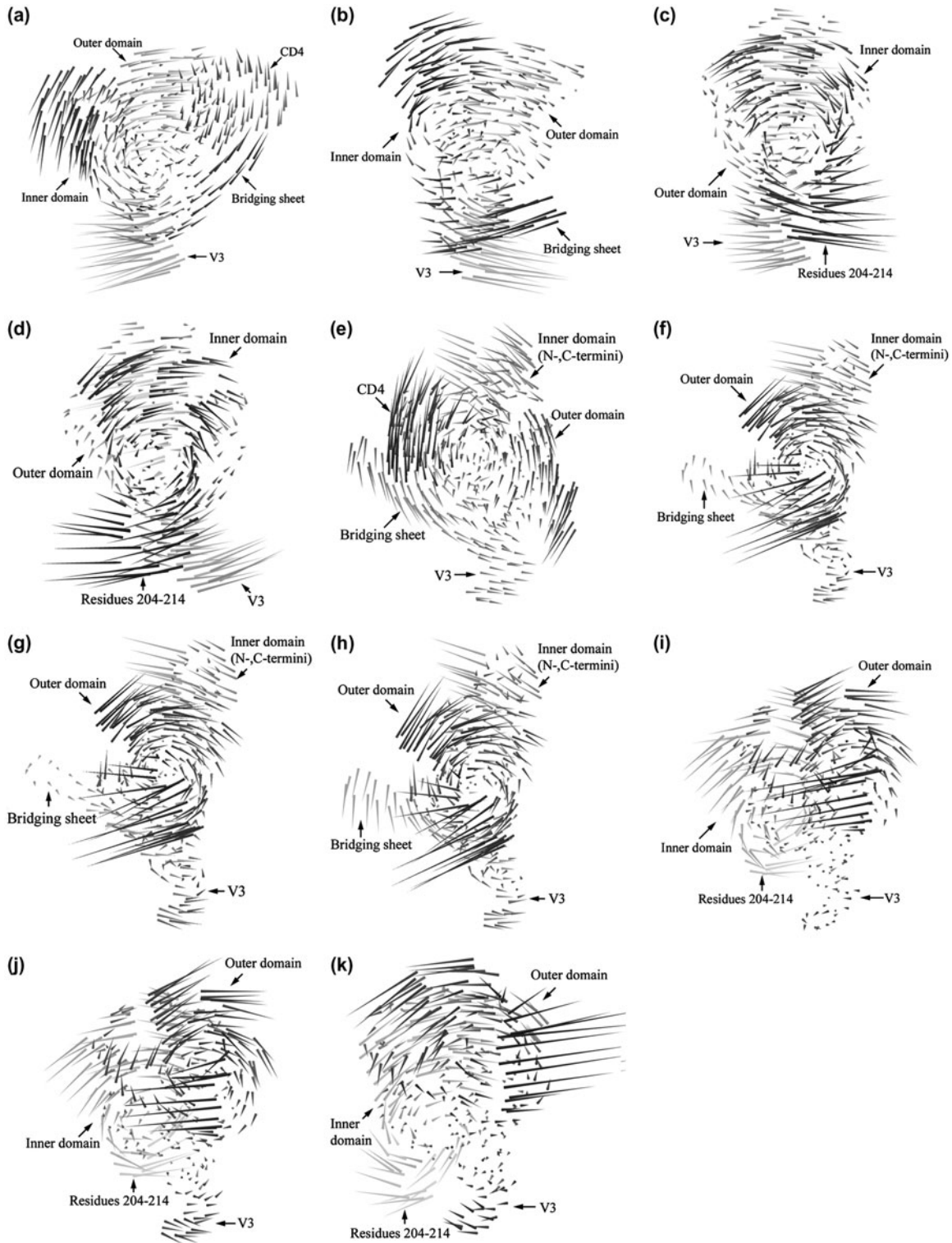
Figure 6.   Porcupine plots of motional modes described by the first two eigenvectors of the wild-type and mutant gp120. (a)–(d) are the most significant large concerted motions along eigenvector 1 for (a) the gp120-CD4 complex, (b) the wild-type CD4-free gp120, (c) the wild-type unliganded gp120, and (d) the unliganded 423 I/P mutant, respectively. The most significant motions of the CD4-free 375 S/W and 423 I/P mutants are identical to that of the wild-type CD4-free gp120 shown in (b). The most significant motion of the unliganded 375 S/W mutant is identical to that of the wild-type unliganded gp120 shown in (c). These views are all from the inner to outer domains. (e)–(k) are the large concerted motions along eigenvector 2 for (e) the gp120-CD4 complex, (f) the wild-type CD4-free gp120, (g) the CD4-free 375 S/W mutant, (h) the CD4-free 423 I/P mutant, (i) the wild-type unliganded gp120, (j) the unliganded 375 S/W mutant, and (k) the unliganded 423 I/P mutant. These views are all from the outer domain to inner domain. These plots were cited from (S. Q. Liu, C. Q. Liu, et al., 2007).

In order to quantitatively evaluate the preference of conformational transition caused by point mutations, we further calculated essential subspace overlaps between ensembles of the wild-type gp120 and mutants (S. Q. Liu, C. Q. Liu, et al., 2007). The values of subspace overlap are shown in Table 2. Here, we mainly focus on the variations in values between the similar essential subspace overlaps, which are defined as the overlaps of all ensembles derived from the same starting conformational state against another ensemble from a different starting state. For example, the values for the similar overlaps of all unliganded ensembles of the wild-type gp120, 375 S/W, and 423 I/P mutants against the CD4-complexed gp120 ensemble are 26.9, 28.6, and 25.7%, respectively. Because the unliganded 375 S/W mutant has the largest value of subspace overlap with the CD4-complexed gp120, we consider that such mutant may have the largest ability to transition toward the CD4-bound conformational state among these unliganded gp120 models. In contrary, the unliganded 423 I/P mutant has the smallest overlap with the CD4-complexed gp120, thus possessing the weakest ability to transition toward the CD4-bound state. Analogously, according to the overlap values of essential subspace listed in Table 2, we infer that the CD4-free 375 S/W mutant possesses the largest capability to maintain the CD4-bound state among all the CD4-free models; the unliganded 423 I/P mutant possesses the largest capability to maintain its unliganded state among all the unliganded models; the CD4-free 423 I/P mutant possesses the smallest capability to maintain its CD4-bound state.

Our theoretical analysis results are in agreement with the physico-chemical studies on these two gp120 mutants (Xiang et al., 2002), demonstrating that both the enthalpic change and the entropic penalty upon 375 S/W mutant binding to CD4 were significantly reduced compared with those observed for the wild-type gp120 and that the 375 S/W mutant was recognized better than

the wild-type gp120 by the CD4 and CD4-induced (CD4i) antibodies. These imply that the 375 S/W mutant samples the conformation space near the CD4-bound state. The other mutant 423 I/P did not bind to the CD4 or CD4i antibodies, but can be recognized efficiently by the CD4-binding site (CD4BS) antibodies, suggesting that this mutant favors the CD4-unliganded state. The structural basis for 375 S/W mutant to prefer for the CD4-bound state is the occupancy of the Trp375 indole group into the Phe43 pocket that resides within the gp120 interior. The favorable interaction between the indole group and the gp120 pocket residues mimics the interaction between the CD4 Phe43 and its binding pocket, thus leading gp120 to sample more CD4-bound conformation states, which in turn lowers the entropic barrier when gp120 binds to CD4 or CD4i antibodies via the conformational selection mechanism. Although our CONCOORD simulations cannot provide the detailed information on the forces underling the indole group–Phe43 pocket interaction, the distance constraints on this interaction may favor the sampling of conformational space close to the CD4-bound state. In the case of the 423 I/P mutant, the steric hindrance of Pro423 makes it difficult for gp120 to form bridging sheet, thus leading to the tendency to maintain the unliganded conformational state. Taken together, our simulation study supports and complements the experimental mutagenesis studies on gp120 molecule and could be extended as a new approach to design or screen mutants that have effects on conformation/function of a protein.

3. Part II: the mechanisms underlying protein dynamics and protein–ligand binding/interaction

3.1. The mechanisms underlying protein dynamics and molecular motions

As mentioned in the "Introduction" section, due to proteins' dynamic nature, a complete understanding of

Table 2. CMSIP values between the first four eigenvectors of one gp120 CONCOORD ensemble and the first ten eigenvectors of another gp120 ensemble. This table is cited from (S. Q. Liu, C. Q. Liu, et al., 2007).

| | Eigenvectors 1–10 | | | | | | |
| | Wild-type | | | 375 S/W | | 423 I/P | |
Eigenvectors 1–4	CD4-free	Unliganded	CD4-complex	CD4-free	Unliganded	CD4-free	Unliganded
CD4-free[a]	1.00	0.180	0.850	0.877	0.212	0.846	0.161
Unliganded[a]	0.259	1.00	0.220	0.231	0.797	0.274	0.929
CD4-complex[a]	0.838	0.269	1.00	0.851	0.286	0.722	0.257
CD4-free 375 S/W[b]	0.866	0.179	0.840	1.00	0.193	NC[c]	NC[c]
Unliganded 375 S/W[b]	0.235	0.787	0.321	0.229	1.00	NC[c]	NC[c]
CD4-free 423 I/P[b]	0.854	0.202	0.709	NC[c]	NC[c]	1.00	0.147
Unliganded 423 I/P[b]	0.258	0.930	0.297	NC[c]	NC[c]	0.240	1.00

[a]Wild-type gp120.
[b]Mutant gp120.
[c]The overlap values were not calculated.

their functions requires the investigation of their dynamics, which is defined as any time-dependent change in atomic coordinates (Henzler-Wildman & Kern, 2007). However, why are proteins inherently dynamic? What are the fundamental physico-chemical principles or mechanisms underlying the protein dynamics? How can these principles dictate protein dynamics and molecular motions? To answer these questions, the concept of the FEL should be introduced.

3.1.1.   The concept of the free energy landscape

The idea of the FEL, despite being now most familiar in the field of protein- folding (Liu et al., 2012), was first proposed to be applied to the folded protein more than 30 years ago by Frauenfelder and colleagues (Austin, Beeson, Eisenstein, Frauenfelder, & Gunsalus, 1975). Dill defined the FEL of a protein–solvent system as a free energy function (Dill, 1999):

$$F(x) = F(x_1, x_2, \dots, x_n) \qquad (1)$$

where, x_1, x_2, …, x_n are variables specifying the protein microscopic states, which can be all the dihedral angles of the amino acids, the eigenvector projections derived from ED analysis (Amadei et al., 1993; Tao et al., 2010), or any degree of freedom (Dill, 1999). $F(x)$ is the free energy of the protein–solvent system, which can be regarded as an approximation of the chemical potential. In thermodynamics the free energy is also called Gibbs free energy (Gibbs, 1873) and is generally expressed as:

$$\Delta G = \Delta H - T\Delta S \qquad (2)$$

where, the enthalpic part (or enthalpic change ΔH) comes from the noncovalent bond formation within the protein interior and the formation and loss of bonds (such as the hydrogen bonds and van der Waals interactions) between the protein and solvent; the entropic part (or entropic change ΔG) comes from all possible solvent configurations and solute (i.e. protein) conformational states. Most importantly, the FEL defines two very important properties of the protein–solvent system: the thermodynamic property (i.e. the relative probabilities of the conformational states), the kinetic property (i.e. the free energy barrier between these conformational states) (Henzler-Wildman & Kern, 2007; Liu et al., 2012). Characterizing these two properties is crucial in resolving protein- folding problem and in understanding protein dynamics.

Under the FEL theory, the process of protein-folding is described by the folding funnel hypothesis (Dill, 1985; Leopold, Montal, & Onuchic, 1992), in which protein-folding is viewed as going down the funnel-like FEL via multiple parallel pathways from a vast majority of individual non-native conformations to the native states located at the bottom of the funnel. This process has been discussed thoroughly elsewhere (Dill & Chan, 1997; Dobson, 2000; Liu et al., 2012) and is not covered by this paper. Here, we mainly focus on the traits at the bottom of the funneled FEL and rationalize how they can dictate the protein dynamics.

3.1.2.   Two important traits of the free energy landscape: the ruggedness and the dynamic variability

First, the bottom of the funnel-like FEL is not smooth but rather, largely rugged or rough due to the existence of the free energy wells and the barriers that separate these wells (Figure 7). Essentially, the free energy wells and barriers are static forms of manifestation of the free energy fluctuation at a certain time point or period, which, actually originate from the noncomplementary changes between the entropy and enthalpy of the protein–solvent system. The change in enthalpy is the consequences of the competitive interactions between amino acid residues within protein and between the protein and solvent; the change in entropy is caused by the nature of protein and solvent to increase their disorder or randomness, that is, a result of the thermal energy of atoms. When the enthalpic change cannot compensate for the entropic change or vice versa, the fluctuation of the free energy occurs (see Equation (2)), leading to the manifestation of a series of hills, valleys, and traps/wells of various heights, depths, and widths over the bottom of the FEL.

Second, the FEL is not static but rather, dynamic. In other words, the shapes of the hills, valleys, and traps/wells composed of the FEL are not invariable; however, any factors capable of perturbing the free energy will alter the profile of the landscape, including changes in the shape, widths, and heights/depths of the hills and valleys/wells and ultimately, leading to redistributions of these "geomorphological features." The factors that are able to perturb free energy are inclusive of physical (e.g. the presence of denaturant, pH, ionic strength, pressure, and temperature) and functional (e.g. the presence of cofactor, small molecular ligand, substrate, and other biomolecules) ones. Both of them change the free energy through bringing out the noncomplementary change between the entropy and enthalpy of the protein–solvent system.

These two crucial traits, the ruggedness and the variability of the bottom region in the funneled FEL, thus determine the dynamic personalities of proteins in the solvent condition. First, the rugged FEL bottom determines that the protein molecules exist in ensembles of different conformational states that coexist in equilibrium at a certain time period. The funneled landscape bottom may contain many free energy wells, within which the
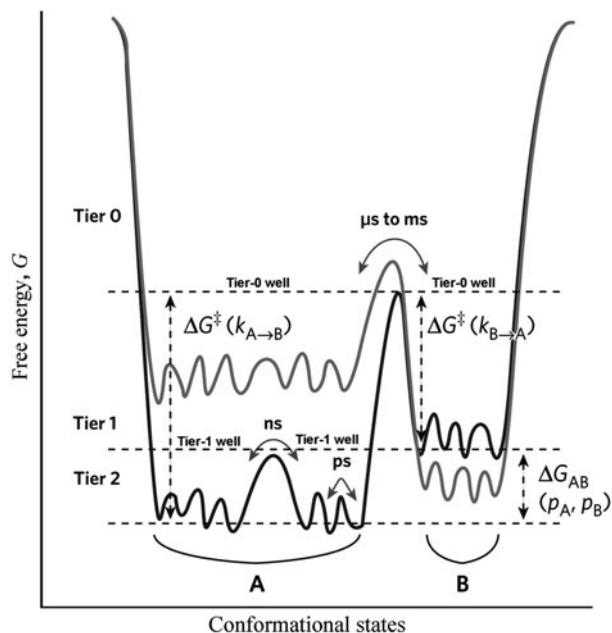
Figure 7. Schematic representation of the FEL to demonstrate the ruggedness, variability and hierarchical organization of the FEL bottom. The tier-0 free energy wells are the largest and deepest wells over the FEL bottom, within which the tier-0 states A and B reside with large populations and high probabilities (p_A, p_B). The difference in distributions between p_A and p_B is determined by their free energy difference ΔG_{AB}. A higher barrier $\Delta G^{\ddagger}$ ($k_{A \to B}$) than $\Delta G^{\ddagger}$ ($k_{B \to A}$) determines that the trasntion rate from state A to B ($k_{A \to B}$) is lower than that from state B to A ($k_{B \to A}$). Although tier-0 states A and B coexist in equilibrium with different population/probability distributions, a change in system will alter the FEL (from the dark line to the gray line) and shift the equilibrium between states and ultimately, the redistributions of the states A and B. The fluctuations around tier-0 states A and B are slow fluctuations that occur on timescales of s to ms (such as the large concerted motions). Each tier-0 well contains many small tier-1 wells, where a large ensemble of closely related substates resides and the fluctuations around these substates occur on fast timescale of ns (such as the loop motions). Even smaller tier-2 wells are contained within each tier-1 well, within which faster fluctuations such as the side chain rotations occur on timescale of ps. Such faster fluctuations originate from the fastest fluctuations occurring on fs timescale such as the bond vibrations. This figure is cited and modified from (Henzler-Wildman & Kern, 2007).

protein molecules in different states are located, with the width and depth of the well determining the relative probability/population/lifetime and the stability of a certain state (i.e. thermodynamic character) (Liu et al., 2012; Yang, Ji, & Liu, 2013). The fluctuations occurring within a relatively large free energy well can be considered as the ones around an average state. Such fluctuations are defined as the equilibrium fluctuations or equilibrium dynamics of the protein and are generally thought to govern the biological function of the protein (Henzler-Wildman & Kern, 2007). Generally, the large

and deep energy wells are able to trap the protein conformational states for a relatively long lifetime, and therefore the equilibrium fluctuations around these individual states can be observed through either experimental techniques (e.g. X-ray diffraction, NMR spectroscopy, fluorescence, infrared spectroscopy, Raman spectroscope and so on) or computational simulations (e.g. MD or CONCOORD simulation). For example, the protein crystal structure determined using the X-ray crystallographic technique is thus a state (tier-0 state in Figure 7) trapped in a large and deep energy well under the crystallization condition, with its B-factors providing information about the spatial distribution around this state (tier-1 and tier-2 states). Second, the variable FEL bottom determines that conformational transition can occur between different conformational states. For example, the changes in solvent conditions or the presence of ligand could alter the width and depth of the free energy wells, or lower the free energy barrier, thus leading to the redistribution of the ensembles of conformational states (the kinetic character; for details, see Figure 7). The fluctuations occurring during the conformational transition are called the nonequilibrium fluctuations and are thought to have minimal effect on the overall rates of biological processes (Henzler-Wildman & Kern, 2007). Owing to the transient lifetimes of the conformational transition intermediates located on the energy barriers, the nonequilibrium fluctuations are hard to detect experimentally.

3.1.3. The multidimensionality of free energy landscape determines the fluctuation amplitudes, directionality, and timescales of protein dynamics

Interestingly, the traits of the ruggedness and variability further determine the high multidimensionality of the FEL for a protein composed of many atoms. In other words, the FEL bottom contains not only some large, deep energy wells, but also many small or even smaller wells within the relatively larger wells (Figure 7). Such nested organization of the energy wells determines, in turn, the multiple hierarchies in not only the timescale but also the amplitude and directionality of protein dynamics. As rationalized by (Henzler-Wildman & Kern, 2007), the fluctuations of proteins occur on different timescales ranging from millisecond (ms) or microsecond (µs) to picosecond (ps) or even femtosecond (fs), with these timescales corresponding to different fluctuation amplitudes ranging from the collective motions (on ms to µs timescale) involving the overall molecule or domain to the local flexibility involving the loop motion (on ns timescale), side-chain rotation (on ps timescale) and bond vibration (on fs timescale) (for details, see Figure 7). Although our MD simulations on proteinase K were only performed for several tens of ns and then are suitable for investigating the loop motions, the large

concerted motions derived from ED analyses were well equilibrated for the first eigenvector and well separated between different forms of the proteases, for example, between the substrate-free and substrate-complexed proteinases K and between the Ca^{2+}-free and Ca^{2+}-bound proteinases K. As thus, such large concerted motions could be regarded as approximations of the equilibrium fluctuations occurring within the large and deep free energy wells and further, be used to investigate the associated biological function and to study the differences in conformational states and in the resulting functional consequence caused by ligand or substrate binding.

Although the amplitude of the equilibrium fluctuations around different average states seems not to provide information about the conformational transition, the motional directions of the overall molecular structure or certain structural components could provide hints on the tendency of the transition. For instance, in our computational simulation study of the HIV-1 gp120 mutants, the observed common or opposite motional directions of the large concerted motions between mutants and reference wild-type gp120 in CD4-complexed or unliganded state suggest that unliganded 375 S/W mutant has a strong preference for transition towards the CD4-bound state, whereas the CD4-free mutant 423 I/P in the CD4-bound state has a strong trend to transition towards the unliganded state. The quantification of conformational subspace overlap between these states supports the conclusions drawn from the qualitative observation of the fluctuation directionality, suggesting that amino acid mutation is able to alter (i.e. increase or lower) the height of free energy barrier between individual energy wells where the different conformational states reside. As mentioned previously, our computational simulations also revealed very complicated and variable motional modes of gp120 molecule, demonstrating that the long variable loops V1/V2 and V3 as well as the bridging sheet are able to move in concert with either the inner domain or outer domain with different amplitude and directions in different conformational states. This suggests that the FEL of the system may be very rugged, with many small energy wells distributing over the bottom of this landscape. This speculation is consistent with a more recent crystallographic study on the HIV-1 gp120 structures, suggesting that the unliganded gp120 may exist in a vast ensemble of distinct conformations, and this provides advantages for HIV-1 virus to escape immune monitoring (Kwon et al., 2012).

From the pure structural point of view, the conformational fluctuations of a protein mainly originate from the highly flexible, surface-exposed loop regions. The high-flexibility nature of the loops comes from the lack of conformational constraints such as the van der Waals and electrostatic interactions and particularly, the periodic hydrogen bonds within themselves or with other structural elements. From the physico-chemical point of view, the high flexibility of the loops comes from their high conformational entropy. Although the competitive interactions of the loops with solvent or other structural components bring out the reduction in enthalpy, such reduction is not sufficient to compensate for the large entropic increase in the loops, thereby leading to the free energy fluctuations and, eventually, the formation of some relatively small free energy wells (the tier-1 wells in Figure 7) within the largest wells (i.e. the tier-0 wells). The loop motions (i.e. tier-1 dynamics) occur faster (on ns timescale) than tier-0 dynamics (i.e. the large concerted motions on µs to ms timescales), allowing protein to assume a large number of closely related substates within each tier-0 well. The multiconformational substates caused by loop motions provide advantages for protein to interaction with multiple structurally dissimilar, but functionally important partners. In the case of proteinase K that we investigated, the loop motions within and/or near the substrate-binding site could, on one hand, allow the protease to recognize and bind to the substrate molecules with multiple sequence motif, on the other hand, to facilitate substrate orientation and catalysis and subsequent product release (Tao et al., 2010).

3.2.   The mechanisms of protein–ligand recognition and binding

Proteins realize their functions mainly through interactions with cofactors, ligands, substrates or other small molecules, and biomolecules. Therefore, elucidating and understanding in depth the mechanism of the protein–ligand recognition and binding is crucial for a thorough understanding of protein functions.

3.2.1.   The three protein–ligand binding models and their relationships

Currently, protein–ligand binding mechanisms are generally described by three models: the "lock-and-key" (Fischer, 1894), the "induced-fit" (Koshland, 1958) and the "conformational selection" (Foote & Milstein, 1994; Monod et al., 1965; Tobi & Bahar, 2005) models (Figure 8). In the lock-and-key model, both the protein (which is likened to the lock) and ligand (which is likened to the key) must be rigid and their surfaces should be accurately complementary so as to guarantee the correct match between the ligand and the binding site of the protein (Figure 8(a)). In the induced-fit model, the ligand binding is able to induce a conformational change at the binding site of the protein, making the final shape of the binding site or, even of the entire protein different from that prior to binding (Figure 8(b)). The prerequisite for induced-fit to occur is that the binding site must have enough conformational flexibility. The conformational selection model is mainly based on the FEL theory. As
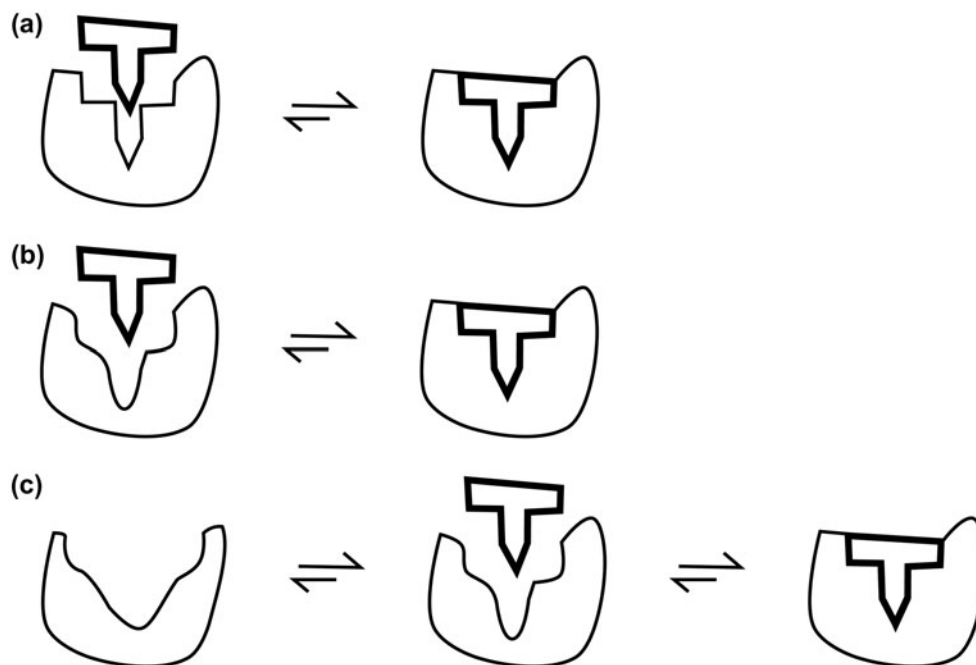
Figure 8.   Schematic representations of the three models that describe the protein–ligand binding mechanisms. (a) Lock-and-key model. (b) Induced-fit model. (c) Conformational selection model. This figure is modified from (Tobi & Bahar, 2005).

discussed earlier, the rugged bottom of the funneled landscape allows the existence of ensembles of different conformational states within individual energy wells. Although these states coexist in equilibrium, the most appropriate conformer can bind selectively to the ligand (Figure 8(c)) and subsequently shift the equilibrium towards complexed state. Similar to the induced-fit, the conformational selection mechanism is also suitable to be applied to flexible proteins, because the high flexibility would lower the free energy barrier and hence, increase the probability of the states appropriate for ligand binding. Generally, the conformational selection involves not only local side chain rearrangements near the binding site, but also the large concerted motions or entire domain movements.

Interestingly, both the lock-and-key and induced-fit models can be regarded as the extreme cases of the conformational selection mechanism under the background of the FEL theory. For example, the lock-and-key binding occurs only for the protein with a single high-rigidity conformational state located within a large and smoothed energy well. Although the conformational selection is suitable for binding of a ligand to a flexible protein that has ensembles of different states within different free energy wells, the binding is selective and occurs only in a single free energy well where the most appropriate conformational state resides. Such process is similar to that described by the lock-and-key model, both being driven by the entropic gain of the protein–ligand–solvent

system (Liu et al., 2012). For the induced-fit, although the conformation of the final bound state differs from that prior to binding, the initial selective collisions and/ or contacts between the protein and ligand decide whether the binding process can proceed forward or the occasional complex will disassociate. Only can the approximately "correct" conformational state provide enough strength and longevity for the interactions within the initial complex to guarantee the subsequent conformational induction to occur (Bosshard, 2001). Therefore, the selectively initial interactions are the prerequisite for the subsequent conformational changes and, as a result, the induced-fit can be considered as an extremity of the conformational selection model.

Taken together, these three models, the lock-and-key, induced-fit and conformational selection, are not independent and mutually exclusive, but rather play a joint role in molecular recognition and binding (Liu et al., 2012; Perica & Chothia, 2010). Because the conformational selection model implies both the lock-and-key and induced-fit processes, it may be more comprehensive and realistic in explaining/describing the binding mechanisms/processes of different ligands and proteins. In this model, the conformational selection (i.e. via the lock-and-key) is important for the initial recognition/contact between the interacting partners and thus determines the ligand specificity, while the subsequent conformational adjustments (i.e. via the induced-fit) play a role in determining the binding affinity.

3.2.2. Interpretation of our MD data with conformational selection and induced-fit mechanisms

In our MD simulations of proteinase K, the combined ED analysis revealed that the large concerted fluctuations described by the first eigenvector were well separated between the Ca^{2+}-free and Ca^{2+}-bound proteinases K and that the substrate-binding regions demonstrated higher flexibility in the Ca^{2+}-bound form than in the Ca^{2+}-free form. This suggests that the Ca^{2+}-bound and Ca^{2+}-free states may be located within different tier-0 free energy wells, of which the Ca^{2+}-bound well is more rugged than the Ca^{2+}-free well, that is, the former has more number of small tier-1 wells than the later. This leads to the speculation that the substrate peptide segment Ala–Ala–Pro–Ala may preferentially choose to bind to the Ca^{2+}-bound state of proteinase K via conformational selection process and, subsequently, the high flexibility of the substrate-binding regions further facilitates substrate binding via induced-fit process, thus explaining the experimentally observed higher substrate-binding affinity of the Ca^{2+}-bound proteinase K compared with the Ca^{2+}-free proteinase K.

In the case of HIV-1 gp120, due to the absence of bridging sheet and fewer number of other structural constraints in the unliganded molecules, the unliganded states may exist in a very rugged tier-0 well with relatively higher free energy. In contrary, the CD4-complexed gp120 molecules are stable because of the formation of the bridging sheet and many other structural constraints. Therefore, such CD4-bound states may exist in another relatively smooth tier-0 well with lower free energy, thus explaining why the isolated unliganded states of gp120 are hard to be "trapped" in the crystal. Interestingly, a more recently published crystallographic study demonstrated that the unliganded HIV-1 gp120 core structures assume the CD-bound conformation (Kwon et al., 2012), supporting our speculation that the CD4-bound states reside within the tier-0 well with lower free energy. As thus, the binding processes of gp120 to the CD4 and CD4i antibodies are largely governed by mechanism of the conformational selection. Ostensibly, the coexistence of unliganded and CD4-bound states provides more opportunity for CD4i antibodies to neutralize the virus, but actually the gp120 molecule exists *in vivo* in the context of a functional viral spike, with its interaction with gp41 providing constraints on the unliganded conformations, thus avoiding as far as possible the neutralization-sensitive CD4-bound state and providing advantages for immune evasion.

3.2.3. The driving forces for protein–ligand binding

Although the process of protein binding to ligand can be explained by the lock-and-key, induced-fit, and more comprehensively and reasonably by the conformational selection model, elucidating the forces that drive protein–ligand binding will facilitate understanding of these three mechanisms and aid in improving binding affinity and specificity by modifying the acceptor or ligand in fields of the medicinal chemistry and drug design.

Like that in the protein-folding process (Ben-Naim, 2012), the driving force of protein–ligand binding is the decrease in the total Gibbs free energy of the protein–ligand–solvent system (Ji & Liu, 2011a, b; Liu et al., 2012; Yang et al., 2013). As described by Equation (2), the free energy change ΔG is dictated by opposing effects involving both the enthalpic (ΔH) and entropic ($T\Delta S$) contributions. In the binding process, the negative ΔH means an exothermic reaction and thus makes favorable contribution to lowering of the free energy; the positive ΔH means an endothermic reaction and thus makes unfavorable contribution to free energy reduction. It is important to keep in mind that enthalpic change of a binding reaction is a global property of the entire system, including not only the contributions from the protein and ligand, but also those from the solvent and protons (Cooper & Johnson, 1994). For example, the breakages of protein–solvent and ligand–solvent hydrogen bonds, electrostatic and van der Waals interactions require energy and thus lead to enthalpic increase (i.e. the positive ΔH); the formation of noncovalent bonds between the protein and ligand and the solvent reorganization near the protein surfaces release energy, thus resulting in enthalpic decrease (i.e. the negative ΔH). The change in binding enthalpy is a combined effect of these individual enthalpic contributions, with specific interactions dominating the binding enthalpy.

The entropy is a measure of how evenly the energy is distributed over the entire protein–ligand–solvent system. Such a trend to distribute energy as evenly as possible will reduce the state of order of the initial system, and therefore, the entropy can be seen as an expression of the disorder or randomness. In the protein–ligand binding process, the positive or negative ΔS makes favorable or unfavorable contribution to lowering of the total Gibbs free energy, respectively. Due to the complexity of the system's components, the total entropic change (ΔS_{tot}) of binding is also a combined contribution from three individual entropic changes: the solvent entropy ΔS_{solv}, conformational entropy ΔS_{conf} and the rotation and translation entropy $\Delta S_{r/t}$, and is expressed as:

$$\Delta S_{tot} = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t} \qquad (3)$$

where, the ΔS_{solv} arises from solvent release upon binding and therefore is positive and contributes favorably to the increase in ΔS_{tot}; the ΔS_{conf} comes from the reduction in rotational degrees of freedom around torsion angles of the protein and ligand upon binding, and therefore is negative and makes a unfavorable contribution to

total entropic increase; the $\Delta S_{r/t}$ represents the loss in translational and rotational freedom degrees of the acceptor and ligand upon complex formation, which also makes a negative contribution to the ΔS_{tot}. The negative entropic changes are unfavorable because they will lead to a trend in increase in the total Gibbs free energy of the system, and thus the ΔS_{conf} and $\Delta S_{r/t}$ are generally called entropic penalty, entropic barrier or entropic cost in the literatures (Amzel, 1997, 2000; Brady & Sharp, 1997). However, because the entropy of hydration of polar and nonpolar groups is large, the positive ΔS_{solv} arising from the displacement of the water networks around the binding interfaces (including both the binding site surface of the protein and the ligand surface) upon binding is generally large and can overcome the negative ΔS_{conf} and $\Delta S_{r/t}$, resulting in an increase in ΔS_{tot} and contributing favorably to lowering of the total Gibbs free energy.

It should be noted that the linear relationship between the changes in enthalpy and entropy, as shown in Equation (2), will lead to a interesting phenomenon called, the entropy–enthalpy compensation (Ji & Liu, 2011a). In other words, the favorable changes in binding enthalpy (i.e. the negative ΔH) are inevitably compensated for by unfavorable changes in binding entropy (i.e. the negative ΔS) and vice versa, thereby resulting in only small change in the binding free energy. In the fields of medicinal chemistry and drug design, such compensation makes it difficult to distinguish between the entropy-driven or enthalpy-driven binding process for a certain ligand and, consequently, makes it difficult to increase the binding affinity through ligand modifications.

3.2.4. The different binding models are driven either by entropic increase or enthalpic decrease or by both.

Distinguishing whether a binding process is driven by entropic increase, enthalpic decrease or the combination of them is necessary because this will facilitate understanding of the binding mechanisms. For purpose of generality, we do not cover here the binding process of a specific ligand to a receptor, but focus mainly on the components of the driving force for the three binding models: the lock-and-key, induced-fit and conformational selection. In the binding process described by the lock-and-key mechanism, due to the perfect surface complementary between the acceptor and ligand, the entropic increase arising from the displacement of water molecules around the binding interfaces (i.e. the large positive ΔS_{solv}) makes a substantial contribution to lowering of the system free energy. In addition, due to the highly structural rigidity of the protein and ligand, such binding process does not involve large backbone conformation change and only some interacting side chains are restricted, leading to a small negative ΔS_{conf}. Generally,

the loss of rotational and translational entropy ($\Delta S_{r/t}$) upon binding is even smaller than ΔS_{conf} (Perozzo, Folkers, & Scapozza, 2004). Thus, these small entropic penalties ($\Delta S_{conf} + \Delta S_{r/t}$) can be easily offset by the enthalpic decrease arising from noncovalent bond formations between the receptor and ligand. Taken together, the binding process of the lock-and-key mechanism is mainly driven by the solvent entropy gain, demonstrating the importance of water molecules in protein–ligand binding. Because the perfect surface complementary between binding partners is the prerequisite for achieving the large entropic gain, the lock-and-key model can be used to explain the specificity of ligand binding.

In the case of the induced-fit model, the initial contact between the interaction interfaces of the protein and ligand exclude limited number of water molecules compared to the lock-and-key model because of the lack of the surface complementarity. This will lead to a small positive ΔS_{solv} that is not sufficient to offset the entropic penalties of ΔS_{conf} and $\Delta S_{r/t.}$ However, the high flexibility of the receptor allows the subsequent large conformational adjustments to suit the incoming ligand and to establish full contacts (i.e. a large number of noncovalent bonds) between the interacting partners. This will result in a large negative enthalpy (ΔH) that is sufficient to overcome the entropic costs of ΔS_{conf} and $\Delta S_{r/t}$, contributing substantially to lowering of the total Gibbs free energy of the protein–ligand–solvent system. Taken together, the induced-fit binding process is mainly driven by the large enthalpic reduction. Because the large number of non-bond interactions resulting from the induced-fit process is able to stabilize the bound ligand within its binding site, we consider that the induced-fit model can be used to explain the ligand-binding affinity.

The conformation selection model, as discussed previously, contains both the "lock-and-key" and "induced-fit" processes. Under this model, the first selective binding step can be viewed as an approximation of the lock-and-key binding process, thus yielding a large positive ΔS_{solv}; the subsequent step of conformational adaption establishes the extensive intermolecular contacts between the binding partners, thus yielding a large negative ΔH. Both the positive ΔS_{solv} and the negative ΔH over-offset the inevitable entropy penalties of ΔS_{conf} and $\Delta S_{r/t}$ and will lead to a significant reduction in the total Gibbs free energy of the system. Taken together, the conformational selection binding process is driven both by the entropic increase and enthalpic decrease and further can be used to explain not only the ligand specificity but also the ligand affinity. In fact, most of the proteins in nature are flexible (especially in the ligand-binding site) and therefore exist as an ensemble of closely related conformational substates. Therefore, the selective binding followed by induced-fit may be a result of evolution, which is advantageous in modulating both

the thermodynamics and kinetics of the protein–ligand interactions and, as thus, allows proteins to recognize and bind to a repertoire of potential partners to realize their biological functions.

4. Conclusions

Because biological function is ultimately rooted in physical motions of biomolecules, a thorough understanding of the function of a protein requires investigating not only its static structures, but even more importantly its dynamic personalities. In this study, several cases of our studies were reviewed to exemplify the importance and necessity of investigating protein dynamics and molecular motions in understanding proteins' biological function. Using computer simulation methods such as MD and CONCOORD, we not only characterized the dynamic properties of proteinase K and HIV-1 gp120, but also related the dynamic behaviors and molecular motion modes to their functional properties. This includes relating the dynamic pockets of the proteinase K to substrate binding, orientation and product release; relating the high flexibility of substrate-binding site to binding mechanism of induced-fit or conformational selection; explaining why Ca^{2+} removal decreases the thermostability and substrate affinity but does not affect dramatically the catalytic activity of proteinase K and explaining why amino acid mutations affect the propensity of gp120 to assume distinct conformational states. Because protein dynamics are rooted in FEL of the protein–solvent system, we further discussed the traits of the FEL, especially focusing on the bottom of the funneled landscape. This includes the ruggedness and variability of the landscape bottom and why it is rugged and variable; the multidimensionality and multihierarchy of the rugged landscape bottom; and how the ruggedness and variability of the landscape dictate the thermodynamic and kinetic characters of proteins. Based on the FEL, the dynamics information obtained from computer simulations or experimental methods can be more reasonably explained and linked to protein functions. The hierarchical organization of free energy wells determines the multitier fluctuations of different protein components on different timescales, thus explaining reasonably the conformational selection mechanism of protein–ligand binding. Finally, we discussed the relationship between the three generally accepted protein–ligand binding models: the lock-and-key, induced-fit, and conformational selection. Because conformational selection model is mainly based on the FEL theory and implies both the binding processes described by the lock-and-key and induced-fit models, we consider that this model may be more realistic and general in explaining the protein–ligand binding mechanisms, either for the rigid or for the flexible proteins.

We believe that the future will see more advanced theoretical and experimental approaches suitable for investigating the dynamics and for reconstructing and characterizing the FEL of proteins–solvent system, which, in turn, will facilitate our understanding of the life process from a more realistically physical level.

References

Amadei, A., Linssen, A. B. M., & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics, 17*, 412–425.

Amzel, L. M. (1997). Loss of translational entropy in binding, folding, and catalysis. *Proteins: Structure, Function, and Genetics, 28*, 144–149.

Amzel, L. M. (2000). Calculation of entropy changes in biological processes: Folding, binding, and oligomerization. *Methods in Enzymology, 323*, 167–177.

Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H., & Gunsalus, I. C. (1975). Dynamics of ligand binding to myoglobin. *Biochemistry, 14*, 5355–5373.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., … Yeh, L. S. (2005). The universal protein resource (UniProt). *Nucleic Acids Research, 33*, 154–159.

Bajorath, J., Raghunathan, S., Hinrichs, W., & Saenger, W. (1989). Long-range structural changes in proteinase K triggered by calcium ion removal. *Nature, 337*, 481–484.

Barrett, C. P., Hall, B. A., & Noble, M. E. (2004). Dynamite: A simple way to gain insight into protein motions. *Acta Crystallographica Section D, 60*, 2280–2287.

Ben-Naim, A. (2012). Levinthal's question revisited, and answered. *Journal of Biomolecular Structure and Dynamics, 30*, 113–124.

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications, 91*, 43–56.

Berger, E. A., Murphy, P. M., & Farber, J. M. (1999a). Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annual Review of Immunology, 17*, 657–700.

Berger, C., Weber-Bornhauser, S., Eggenberger, J., Hanes, J., Pluckthun, A., & Bosshard, H. R. (1999b). Antigen recognition by conformational selection. *FEBS Letters, 450*, 149–153.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research, 28*, 235–242.

Betzel, C., Pal, G. P., & Saenger, W. (1988). Three dimensional structure of proteinase K at 0.15 nm resolution. *European Journal of Biochemistry, 178*, 155–171.

Bosshard, H. R. (2001). Molecular recognition by induced fit: How fit is the concept? *News in Physiological Sciences, 16*, 171–173.

Brady, G. P., & Sharp, K. A. (1997). Entropy in protein folding and in protein–protein interactions. *Current Opinion in Structural Biology, 7*, 215–221.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A Program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry, 4*, 187–217.

Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D. C., & Harrison, S. C. (2005). Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature, 433*, 834–841.

Cooper, A., & Johnson, C. M. (1994). Introduction to microcalorimetry and biomolecular energetics. *Methods in Molecular Biology, 22*, 109–124.

de Groot, B. L., Hayward, S., van Aalten, D. M. F., Amadei, A., & Berendsen, H. J. C. (1998). Domain motions in bacteriophage T4 lysozyme: A comparison between molecular dynamics and crystallographic data. *Proteins: Structure, Function, and Genetics, 31*, 116–127.

de Groot, B. L., van Aalten, D. M. F., Scheek, R. M., Amadei, A., Vriend, G., & Berendsen, H. J. C. (1997). Prediction of protein conformational freedom from distance constraints. *Proteins: Structure, Function, and Genetics, 29*, 240–251.

Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry, 24*, 1501–1509.

Dill, K. A. (1999). Polymer principles and protein folding. *Protein Science, 8*, 1166–1180.

Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology, 4*, 10–19.

Dobson, C. M. (2000). The nature and significance of protein folding. In R. H. Pain (Ed.), *Mechanisms of protein folding* (2nd ed., pp. 1–33). Oxford: Oxford University Press.

Dodson, G., & Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends in Biochemical Sciences, 23*, 347–352.

Eckert, D. M., & Kim, P. S. (2001). Mechanisms of viral membrane fusion and its inhibition. *Annual Review of Biochemistry, 70*, 777–810.

Fetrow, J. S., Giammona, A., Kolinski, A., & Skolnick, J. (2002). The protein folding problem: A biophysical enigma. *Current Pharmaceutical Biotechnology, 3*, 329–347.

Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Berichte der Deutschen Chemischen Gesellschaft, 27*, 2984–2993.

Foote, J., & Milstein, C. (1994). Conformational isomerism and the diversity of antibodies. *Proceedings of the National Academy of Sciences of the United States of America, 91*, 10370–10374.

Gibbs, J. W. (1873). A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. *Transactions of the Connecticut Academy of Arts and Sciences, 2*, 382–404.

Gudjónsdóttir, K., & Asgeirsson, B. (2008). Effects of replacing active site residues in a cold-active alkaline phosphatase with those found in its mesophilic counterpart from Escherichia coli. *FEBS Journal, 275*, 117–127.

Hansson, T., Oostenbrink, C., & van Gunsteren, W. (2002). Molecular dynamics simulations. *Current Opinion in Structural Biology, 12*, 190–196.

Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chemical Reviews, 102*, 4501–4523.

Henzler-Wildman, K. A., & Kern, D. (2007). Dynamic personalities of proteins. *Nature, 450*, 964–972.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics and Modelling, 14*, 33–38.

Ji, X. L., & Liu, S. Q. (2011a). Is stoichiometry-driven protein folding getting out of thermodynamic control? *Journal of Biomolecular Structure and Dynamics, 28*, 621–623.

Ji, X. L., & Liu, S. Q. (2011b). Thinking into mechanism of protein folding and molecular binding. *Journal of Biomolecular Structure and Dynamics, 28*, 995–996.

Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society, 110*, 1657–1666.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers, 22*, 2577–2637.

Kannan, S., & Zacharias, M. (2009). Simulated annealing coupled replica exchange molecular dynamics – an efficient conformational sampling method. *Journal of Structural Biology, 166*, 288–294.

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structure Biology, 9*, 646–652.

Koshland, D. E. J. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 44*, 98–104.

Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S., & Singh, T. P. (2001). Structure of a serine protease proteinase K from *Tritirachium album* limber at 0.98 Å resolution. *Biochemistry, 40*, 3080–3088.

Kwon, Y. D., Finzi, A., Wu, X., Dogo-Isonagie, C., Lee, L. K., Moore, L. R., … Kwong, P. D. (2012). Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternaryinteractions and variable loops. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 5663–5668.

Kwong, P. D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (2000). Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure, 8*, 1329–1339.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV-1 gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature, 393*, 648–659.

Leopold, P. E., Montal, M., & Onuchic, J. N. (1992). Protein folding funnels: A kinetic approach to the sequence–structure relationship. *Proceedings of the National Academy of Sciences of the United States of America, 89*, 8721–8725.

Levitt, M., Hirshberg, M., Sharon, R., & Daggett, V. (1995). Potential-energy function and parameters for simulations of the molecular-dynamics of proteins and nucleic acids in solution. *Computer Physics Communications, 91*, 215–231.

Lindahl, E., Hess, B., & van der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling, 7*, 306–317.

Liu, S. Q., Fu, Y. X., & Liu, C. Q. (2007). Molecular motions and conformational transition between different conformational states of HIV-1 gp120 envelope glycoprotein. *Chinese Science Bulletin, 52*, 3074–3088.

Liu, S. Q., Liang, L. M., Tao, Y., Yang, L. Q., Ji, X. L., Yang, J. K., Fu, Y. X., & Zhang, K. Q. (2011a). Structural and dynamic basis of serine proteases from nematophagous fungi for cuticle degradation. In M. Stoytcheva (Ed.), *Pesticides in the modern world – pests control and pesticides exposure and toxicity assessment* (pp. 333–376). Rijeka: Intech.

Liu, S. Q., Liu, C. Q., & Fu, Y. X. (2007). Molecular motions in HIV-1 gp120 mutants reveal their preferences for different conformations. *Journal of Molecular Graphics and Modelling, 26*, 306–318.

Liu, S. Q., Liu, S. X., & Fu, Y. X. (2007). Dynamic domains and geometrical properties of HIV-1 gp120 during conformational changes induced by CD4-binding. *Journal of Molecular Modeling, 13*, 411–424.

Liu, S. Q., Liu, S. X., & Fu, Y. X. (2008). Molecular motions of human HIV-1 gp120 envelope glycoproteins. *Journal of Molecular Modeling, 14*, 857–870.

Liu, S. Q., Meng, Z. H., Fu, Y. X., & Zhang, K. Q. (2010). Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. *Journal of Molecular Modeling, 16*, 17–28.

Liu, S. Q., Meng, Z. H., Fu, Y. X., & Zhang, K. Q. (2011). The effect of calciums on the molecular motions of proteinase K. *Journal of Molecular Modeling, 17*, 289–300.

Liu, S. Q., Meng, Z. H., Yang, J. K., Fu, Y. X., & Zhang, K. Q. (2007). Characterizing structural features of cuticle-degrading proteases from fungi by molecular modeling. *BMC Structural Biology, 7*, 33.

Liu, S. Q., Xie, Y. H., Ji, X. L., Tao, Y., Tan, D. Y., Zhang, K. Q., & Fu, Y. X. (2012). Protein folding, binding and energy landscape: A synthesis. In P. T. P. Kaumaya (Ed.), *Protein engineering* (pp. 207–252). Rijeka: Intech.

Ma, B., Wolfson, H. J., & Nussinov, R. (2001). Protein functional epitopes: Hot spots, dynamics and combinatorial libraries. *Current Opinion in Structural Biology, 11*, 364–369.

Martin, J. R., Mulder, F. A. A., Karimi-Nejad, Y., van der Zwan, J., Mariani, M., Schipper, D., & Boelens, R. (1997). The solution structure of serine protease PB92 from *bacillus alcalophilus*. *Structure, 5*, 521–532.

Mello, L. V., de Groot, B. L., Li, S., & Jedrzejas, M. J. (2002). Structure and flexibility of streptococcus agalactiae hyaluronate lyase complex with its substrate: Insights into the mechanism of processive degradation of hyaluronan. *Journal of Biological Chemistry, 277*, 36678–36688.

Merlino, A., Vitagliano, L., Ceruso, M. A., & Mazzarella, L. (2004). Dynamic properties of the N-terminalswapped dimer of ribonuclease A. *Biophysical Journal, 86*, 2383–2391.

Miller, D. W., & Agard, D. A. (1999). Enzyme specificity under dynamic control: A normal mode analysisof alpha-lytic protease. *Journal of Molecular Biology, 286*, 267–278.

Monod, J., Wyman, J., & Changeux, J. P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology, 12*, 88–118.

Müller, A., Hinrichs, W., Wolf, W. M., & Saenger, W. (1994). Crystal structure of calcium-free proteinase K at 1.5-Å resolution. *Journal of Biological Chemistry, 269*, 23108–23111.

Myszka, D. G., Sweet, R. W., Hensley, P., Brigham-Burke, M., Kwong, P. D., Hendrickson, W. A., Wyatt, R., Sodroski, J., & Doyle, M. L. (2000). Energetics of the HIV gp120-CD4 binding reaction. *Proceedings of the National Academy of Sciences of the USA, 97*, 9026–9031.

Perica, T., & Chothia, C. (2010). Ubiquitin – molecular mechanisms for recognition of different structures. *Current Opinion in Structural Biology, 20*, 367–376.

Perozzo, R., Folkers, G., & Scapozza, L. (2004). Thermodynamics of protein–ligand interactions: History, presence, and future aspects. *Journal of Receptor and Signal Transduction Research, 24*, 1–52.

Peters, G. H., & Bywater, R. P. (1999). Computational analysis of chain flexibility and fluctuations in Rhizomucor miehei lipase. *Protein Engineering, Design and Selection, 12*, 747–754.

Peters, G. H., Frimurer, T. M., Andersen, J. N., & Olsen, O. H. (1999). Molecular dynamics simulations of protein–tyrosine phosphatase 1B. I. Ligand-induced changes in the protein motions. *Biophysical Journal, 77*, 505–515.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry, 26*, 1781–1802.

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology, 234*, 779–815.

Sattentau, Q. J. (1998). HIV gp120: Double lock strategy foils host defences. *Structure, 6*, 945–949.

Schecter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications, 27*, 157–162.

Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., & van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation package. *Journal of Physical Chemistry A, 103*, 3596–3607.

Shaw, W. V. (1987). Protein engineering. The design, synthesis and characterization of factitious proteins. *Biochemical Journal, 246*, 1–17.

Tao, Y., Rao, Z. H., & Liu, S. Q. (2010). Insight derived from molecular dynamics simulation into substrate-induced changes in protein motions of proteinase K. *Journal of Biomolecular Structure and Dynamics, 28*, 143–157.

Tobi, D., & Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 18908–18913.

Tsai, C. J., Ma, B., Sham, Y. Y., Kumar, S., & Nussinov, R. (2001). Structured disorder and conformational selection. *Proteins: Structure, Function, and Genetics, 44*, 418–427.

van Aalten, D. M. F., Amadei, A., Linssen, A. B. M., Eijsink, V. G. H., Vriend, G., & Berendsen, H. J. C. (1995). The essential dynamics of Thermolysin: Conformation of hinge-bending motion and comparison of simulations in vacuum and water. *Proteins: Structure, Function, and Genetics, 22*, 45–54.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry, 26*, 1701–1718.

Vranken, W. F., Budesinsky, M., Fant, F., Boulez, K., & Borremans, F. A. (1995). The complete consensus V3 loop peptide of the envelope protein gp120 of HIV-1 shows pronounced helical character insolution. *FEBS Letters, 374*, 117–121.

Vreede, J., van der Horst, M. A., Hellingwerf, K. J., Crielaard, W., & van Aalten, D. M. (2003). PAS domains. Common structure and common flexibility. *Journal of Biological Chemistry, 278*, 18434–18439.

Wade, R. C., Gabdoulline, R. R., Ludemann, S. K., & Lounnas, V. (1998). Electrostatic steering and ionic tethering in enzyme–ligand binding: Insights from simulations. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 5942–5949.

Weiner, P. K., & Kollman, P. A. (1981). Amber – assisted model-building with energy refinement – a general program for modeling molecules and their interactions. *Journal of Computational Chemistry, 2*, 287–303.

Wells, J. A., & Estell, D. A. (1988). Subtilisin – an enzyme designed to be engineered. *Trends in Biochemical Sciences, 13*, 291–297.

Wolf, W. M., Bajorath, J., Müller, A., Raghunathan, S., Singh, T. P., Hinrichs, W., & Saenger, W. (1991). Inhibition of proteinase K by methoxysuccinyl–Ala–Ala–Pro–Ala–chloromethyl ketone: An X-ray study at 2.2-Å resolution. *Journal of Biological Chemistry, 266*, 17695–17699.

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure–function paradigm. *Journal of Molecular Biology, 293*, 321–331.

Xiang, S. H., Kwong, P. D., Gupta, R., Rizzuto, C. D., Casper, D. J., Wyatt, R., Wang, L., Hendrickson, W. A., Doyle, M. L., & Sodroski, J. (2002). Mutagenic stabilization and/or disruption of a CD4-bound state reveals distinct conformations of the human immunodeficiency virus type 1 gp120 envelope glycoprotein. *Journal of Virology, 76*, 9888–9899.

Yang, L. Q., Ji, X. L., & Liu, S. Q. (2013). The free energy landscape of protein folding and dynamics: A global view. *Journal of Biomolecular Structure and Dynamics*. doi:10.1080/07391102.2012.748536