Monte Carlo simulations of the peptide recognition at the consensus binding site of the constant fragment of human immunoglobulin G: The energy landscape analysis of a hot spot at…

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2002

Impact Factor: 2.63 · DOI: 10.1002/prot.10164 · Source: PubMed

6 AUTHORS, INCLUDING:

Peter W Rose
University of California, San Diego
44 PUBLICATIONS **1,522** CITATIONS

PROTEINS: Structure, Function, and Genetics 48:539–557 (2002)

Monte Carlo Simulations of the Peptide Recognition at the Consensus Binding Site of the Constant Fragment of Human Immunoglobulin G: the Energy Landscape Analysis of a Hot Spot at the Intermolecular Interface

Gennady M. Verkhivker,* Djamal Bouzida, Daniel K. Gehlhaar, Paul A. Rejto, Stephan T. Freer, and Peter W. Rose
Agouron Pharmaceuticals, Inc., A Pfizer Company, San Diego, California

ABSTRACT **Monte Carlo simulations of molecular recognition at the consensus binding site of the constant fragment (Fc) of human immunoglobulin G (Ig) protein have been performed to analyze structural and thermodynamic aspects of binding for the 13-residue cyclic peptide DCAWHLGELVWCT. The energy landscape analysis of a hot spot at the intermolecular interface using alanine scanning and equilibrium-simulated tempering dynamics with the simplified, knowledge-based energy function has enabled the role of the protein hot spot residues in providing the thermodynamic stability of the native structure to be determined. We have found that hydrophobic interactions between the peptide and the Met-252, Ile-253, His-433, and His-435 protein residues are critical to guarantee the thermodynamic stability of the crystallographic binding mode of the complex. Binding free energy calculations, using a molecular mechanics force field and a solvation energy model, combined with alanine scanning have been conducted to determine the energetic contribution of the protein hot spot residues in binding affinity. The conserved Asn-434, Ser-254, and Tyr-436 protein residues contribute significantly to the binding affinity of the peptide–protein complex, serving as an energetic hot spot at the intermolecular interface. The results suggest that evolutionary conserved hot spot protein residues at the intermolecular interface may be partitioned in fulfilling thermodynamic stability of the native binding mode and contributing to the binding affinity of the complex. Proteins 2002;48:539–557.**

Key words: molecular recognition; consensus binding site; energy landscapes; Monte Carlo simulations; hot spot; alanine scanning

INTRODUCTION

Molecular recognition between proteins and flexible target molecules, including other proteins, nucleic acids, and small molecules, is often accompanied by a considerable flexibility of the protein binding sites and structural rearrangements on binding between the associated partners.[1–4] Accessibility of alternative conformational states is important for protein function, including assembly, molecular recognition, regulation of biological activity, and enzymatic catalysis.[4–9] It has been recognized that proteins are not adequately described by a single conformational state but are better represented by a manifold of low-energy protein conformations, conformational substates, on a rugged energy landscape.[4–9] The current view of the protein energy landscape picture implies that the conformational substates, which represent local minima of the protein, are organized in hierarchical tiers that are separated by barriers that can be crossed by thermal activation.[6,7] Within this hierarchy, alternative conformational states are defined by significant differences in protein conformation and large energy barriers, whereas modest coordinate changes, with concomitantly smaller energy barriers, characterize alternative protein conformational substates.[8] The distribution of protein conformational substates is heterogeneous and may depend on the topology of the protein structure where the proteins with similar architectures can exhibit similar large-scale dynamic behavior.[10,11] The analysis of point mutations in proteins has revealed that topology of the native structure may determine which regions of the protein structure respond to sequence changes, often regardless of the sequential or spatial location of the mutation. A similar effect has been observed in protein superfamilies in which different sequences with similar topologies display similar large-scale dynamic fluctuations.[10,11]

Protein flexibility and dynamics of the intermolecular interfaces can have a profound effect in determining binding thermodynamics, kinetics, and, consequently, in modulating binding affinity and specificity of molecular recognition.[12–16] NMR relaxation studies have provided a quantitative view of motions at specific positions of intermolecular interfaces and site-specific entropic contributions.[12,13] The mobility of protein polar residues with favorable enthalpy contributions to binding may become locally restricted at the intermolecular interface, leading to a considerable entropic cost.[14] The increase in motions

*Correspondence to: Gennady M. Verkhivker, Department of Computational Chemistry, Pfizer Global R&D La Jolla, 10777 Science Center Dr., San Diego, CA 92121-1111.
E-mail: gennady.verkhivker@pfizer.com

Received 11 October 2001; Accepted 15 March 2002

of hydrophobic residues can rescue this energetic loss by increasing both backbone and side-chain flexibility and compensating the entropic component of the binding free energy.[15] Local flexibility within a protein, which often changes on binding, involves a significant change of conformational entropy that is typically expressed by equilibrium fluctuations between conformational substates around the native structure.

The presence of mobile regions undergoing local folding and unfolding transitions on ligand binding may result in local cooperativity effects that are not limited to the binding site residues but can propagate to a small subset of remote protein regions.[17–20] It has been found that protein binding sites can have a dual character and are characterized by the presence of the regions with both high and low structural stability.[19,20] Structural disorder of the HIV-1 protease and a dual character of the binding site with regions of high and low structural stability can have important biological implications, in particular, conferring resistant mutations in the HIV-1 protease.[21,22]

Despite a typical large size of intermolecular interfaces, binding affinity and specificity may be determined by a relatively small fraction of the interface, forming energetically important hot spots at protein-binding sites.[23] Alanine-scanning mutagenesis of interfacial residues, combined with structural and thermodynamic studies, have been valuable experimental tools in both detecting and analyzing the energetic hot spots at the macromolecular interfaces.[23,24] A systematic analysis of a wide range of protein–protein interfaces[25,26] has shown a diversity of interaction patterns and a lack of general rules for hydrophobicity, polarity, or shape, which can be used to unambiguously predict hot spots at the intermolecular interfaces. Analysis of the anatomy of hot spots compiled from 22 protein interfaces has shown that free energy of binding is not equally distributed across protein–protein interfaces,[27] which destroys a simple relationship between binding affinity and the amount of the buried solvent-accessible surface area. A recent analysis of conserved residues in 11 clustered interface families comprising a total of 97 crystal structures has shown that the composition of hot spots is typically enriched by certain residues, such as Trp, Tyr, Arg, His, Gln, Asn, and Pro, and can be surrounded by a shell of less important residues, shielding hot spot residues from solvent and, thereby, providing a requirement for their favorable interactions.[27–29] This pattern could have a more simple explanation,[24] indicating that peripheral residues are more easily replaced by water than the hot spot residues in the center of the intermolecular interface.[30]

Structural plasticity in protein–protein interfaces[31–33] can rescue large functional changes by a limited number of mutations at the intermolecular interface and provides a mechanism for mutations to be accommodated during coevolution of high-affinity binding partners. Furthermore, protein-binding sites can be not only structurally flexible but also functionally adaptive, with a diverse range of protein systems capable of binding with high affinity to ligands different from their natural binding partners in composition, size, and shape.[34,35]

The relationship between conformational flexibility and convergent evolution was shown in a recent study of the recognition in the hinge region of the constant fragment (Fc) of human immunoglobulin G (Ig) protein, which represents the consensus binding site for natural proteins and synthetic peptides.[34] Multiple binding partners, including domain B1 of protein A, domain C2 of protein G, rheumatoid factor, neonatal Fc-receptor, and a combinatorially selected peptide, interact with Ig at the consensus binding site. It is important that a combinatorially selected high-affinity synthetic peptide mimics specific interactions of the natural proteins while using the interacting groups from a completely different structural scaffold.[34] The crystal structures of the Fc fragment of Ig in complexes with different binding partners have shown that the consensus region of Fc undergoes local conformational changes to complement the distinct surfaces of each binding partner. These fluctuations are primarily reflected in movements of two critical Met residues at the binding site, which help to adapt pockets for protein G and for the peptide, while presenting a much flatter surface to protein A and rheumatoid factor.[34] The consensus binding region on the surface of the Fc protein is characterized by a high degree of solvent accessibility and a predominantly nonpolar character of the interface, suggesting that the hydrophobic effect is the driving force behind binding. It is of interest that the low hydrogen bonding ability in the binding region (19% of the surface is capable of hydrogen bonding compared with 37% on average) has indicated that this site places fewer geometric constraints on binding partners, because fewer complementary polar interactions are necessary for binding.[34] These studies have revealed that evolution can find convergent solutions to stable intermolecular interfaces by using structural plasticity of the binding site and stabilizing contributions of the hot spot residues to accommodate to different binding partners.

It has been suggested that specific recognition and binding to multiple ligands at the same binding site may result from the dynamic distribution of the protein conformational states at equilibrium near the native structure and may be associated with the ruggedness of the underlying binding energy landscape.[4,9,36–38] The less stable and less selective complexes may have a more rugged bottom of the binding energy funnel with low barriers between conformers of the complex. Highly specific peptide–protein complexes are likely to be relatively rigid, with a steep funnel of conformations leading to the native structure.[4,9,36–38] Molecular recognition involves a large number of conformational states available to a system, which necessitates the use of a statistical characterization and the energy landscape analysis, originally introduced in protein folding[39–45] and further developed in ligand–protein binding.[46–55] The concept of the energy funnel on the landscape[39–43] implies the presence of a deep and broad native basin surrounding the native binding mode, with the stabilizing native interactions stronger on aver-

age than non-native interactions of the alternative binding modes. This results in a gradual energy decrease as the native state is approached. Equilibrium binding energy landscapes generated for a number of ligand–protein complexes have revealed a funnel-like character of the binding energy landscape near the native structures.[51–55] It was shown that robust peptide–protein docking on complex energy surfaces may require the underlying binding energy landscape to have a funnel leading to the global free energy minimum.[53–55] Furthermore, a minimally frustrated energy surface with a unique and stable native complex can be an important prerequisite for receptor-specific binding.[52] The funnels have been detected near the binding sites of protein– protein complexes, providing a plausible explanation for the observed rapid association rates.[56] A comprehensive docking of protein–protein complexes has revealed the existence of low-resolution recognition and the presence of binding funnels near the native structure in protein–protein interactions on a broad scale, suggesting that this may be a general feature in protein–protein association.[57,58]

In this work, we use the energy landscape analysis to study structural and energetic aspects of molecular recognition between a 13-residue cyclic peptide DCAWHLGELVWCT and the Fc fragment of the Ig protein.[34] To represent alternate protein conformational substates, we apply a simulation approach to model ligand binding with the ensembles of multiple protein conformations.[59,60] This method can account for protein flexibility by considering a finite number of protein states that have significant differences in both side-chain and main-chain conformations. The protein-bound conformations found in complexes with various binding partners mimic a range of protein equilibrium fluctuations near the native structure. Equilibrium simulations of the peptide–protein binding are performed with protein mutants, obtained by alanine scanning of the critical residues in the binding site. We set out to determine the role of individual hot spot protein residues to the funnel-like character of the energy landscape near the binding site and their contribution to the thermodynamic stability of the native binding mode of the complex.

The energy landscape analysis with the simplified energy models can faithfully describe a multitude of the available binding modes for the complex and reproduce relative thermodynamic stability of the native complex for intermediate complexes and alternative binding modes.[51–56] However, the assessment of binding affinity for the complexes requires not only an adequate sampling of the conformational space, but more importantly, accurate estimation of the binding energetics. Computational analysis of binding affinity for the peptide–Ig complexes using rigorous free energy perturbations methods is difficult and time-consuming, given the size of the peptide and a diverse range of large perturbations. A multistage strategy[61,62] with a hierarchy of different energy functions is pursued in this work to achieve a synergy of robust conformational sampling and accurate estimation of binding energetics. The simplified energy function[62,63] is used in conjunction with Monte Carlo simulations[63–65] to sample

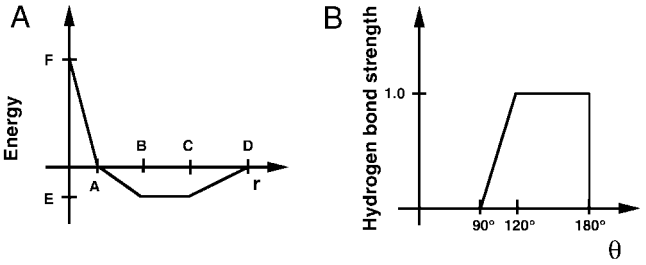

Fig. 1. **A:** The functional form of the ligand–protein interaction energy. For steric interactions, A = 0.93B, C = 1.25B, D = 1.5B, E = −0.4, F = 15.0, and B = $r_{ligand} + r_{protein}$ is the sum of the atomic radii for the ligand and protein atoms. For hydrogen bond interactions, A = 2.3, B = 2.6, C = 3.1, D = 3.4, E = −4.0, F = 15.0. For sulfur hydrogen bond interactions, A = 2.7, B = 3.0, C = 3.5, D = 3.8, E = −2.0, F = 15.0. For chelating interactions with the metals A = 1.5, B = 1.7, C = 2.5, D = 3.0, E = −10.0, F = 15.0. For repulsive interactions, A = 3.2, E = 0.1, F = 15.0. The repulsive potential is then linearly scaled from E = 0.1 to zero between 3.2 and 5.0 Å. The units of A, B, C, and D are Å; for E and F the units are kcal/mol. **B:** The hydrogen bond interaction energy and the repulsive term are multiplied by the hydrogen bond strength term, which is a function of the angle θ determined by the relative orientation of the protein and ligand atoms.

the conformational space and adequately describe the multitude of the low-energy states available to the system. The resulting conformational states are evaluated with detailed binding free energy model, which includes the molecular mechanics AMBER force field[66] and the solvation energy term based on continuum generalized Born and solvent accessible surface area (GB/SA) solvation model.[67–73] This procedure is conceptually similar to the MM/PBSA (molecular mechanics Poisson-Boltzmann surface area) approach[74–81] and replaces time-consuming PB continuum calculations with less demanding GB solvation calculations, correlating well with the PB results.[82] We apply this approach to evaluate binding affinity of the peptide with various protein mutants and analyze role of the protein hot spot residues in contributing to the binding affinity with the peptide.

MATERIALS AND METHODS

Molecular Recognition Energy Model

The knowledge-based simplified energetic model includes intramolecular energy terms for the ligand, given by torsional and nonbonded contributions of the DREIDING force field,[83] and intermolecular ligand–protein steric and hydrogen bond interaction terms calculated from a simplified piecewise linear (PL) potential summed over all protein and ligand heavy atoms [Fig. 1(a)]. The parameters of the pairwise potential depend on the following different atom types: hydrogen bond donor, hydrogen bond acceptor, both donor and acceptor, carbon-sized nonpolar, sulfur-sized nonpolar, flourine-sized nonpolar, and large nonpolar. The atomic radius is 1.4 Å for fluorine and 1.8 Å for carbon, oxygen, and nitrogen atoms. The atomic radius of 2.2 Å is assigned to sulfur and phosphorus, chlorine, and bromine atoms, modeled as large nonpolar atom type. Electronegative atoms with an attached hydrogen are defined as donors, whereas oxygen and nitrogen atoms with no bound hydrogens are defined

as acceptors. Sulfur is modeled as being capable of making weak hydrogen bonds, which allows for sulfur donor closer contacts that are seen in some of the crystal structures. Crystallographic water molecules and hydroxyl groups are defined in this model to be both donor and acceptor, and carbon atoms are defined to be nonpolar. An empirical desolvation correction is applied to the attractive portion of the interactions between nonpolar and polar atoms. This correction is defined as the ratio between the attractive well depth for nonpolar–polar contacts and the one for nonpolar–nonpolar contacts and can range between 0 and 1. The parameter is set to 1.0 in this work, thereby imposing a desolvation penalty by disadvantaging the burial of polar groups with the nonpolar atoms.

A hydrogen bond interaction term is assigned to interactions between donor and acceptors, a repulsive interaction contribution is computed for donor–donor and acceptor–acceptor contacts, and a steric intermolecular term is assigned for other contacts. The steric and hydrogen bondlike potentials have the same functional form, with an additional three-body contribution to the hydrogen bond term and the repulsive term for donor–donor and acceptor–acceptor contacts. Both the hydrogen bond interaction energy and the repulsive interaction contribution between donor–donor and acceptor–acceptor close contacts are modulated by an approximate angular dependence [Fig. 1(b)]. These terms are multiplied by the hydrogen bond strength term, which is a function of the angle θ determined by the relative orientation of the protein and ligand atoms [Fig. 1(b)]. The scaling for the repulsive interactions is equivalent to the dependence used for the hydrogen bond interaction term, but in this cases it implies a maximum penalty when the angle θ is 180°, fading to zero at 90° and below. θ is the angle between two vectors, one of which points from the protein atom to the ligand atom. For protein atoms with a single heavy atom neighbor, the second vector connects the protein atom with its heavy atom neighbor, whereas for protein atoms with two heavy atom neighbors, it is the bisector of the vectors connecting the protein atom with its two neighbors.

For molecular docking simulations, it has been shown that the energy surface must be smooth for robust structure prediction of ligand–protein complexes[53–55]; softening the potentials is a way to smooth the force field and enhance sampling of the conformational space while retaining adequate description of the binding energy landscape. The PL energy function has no singularities at interatomic distances, effectively explores accessible ligand binding modes, and samples a large fraction of conformational space. This function produces reliable results in predicting crystal structures of ligand–protein complexes.[53–55,61,62]

A hierarchical approach is used where the PL energy function is used in combination with a powerful searching technique, parallel Monte Carlo simulated tempering dynamics[59,60,84–91] to adequately sample the conformational space and describe the multitude of the ligand-binding modes. The advantage of the simulated tempering dynamics and the simplified energy function is the ability to adequately sample the conformational space.

Although the PL energy function is proven to be more adequate for sampling nonpolar and hydrogen bonds patterns, this simplified energy model does not include a direct electrostatic component and, therefore, may be less accurate in detecting the exact energetics of the binding modes, especially when extensive networks of electrostatic interactions are present in the crystal structure. This function is less accurate in detecting the exact location and energetics of the native state because of the inaccuracy in quantifying the magnitude of peptide–protein interactions.

A binding free energy model, which includes a molecular mechanics AMBER force field and a GB/SA solvation contribution, is used to evaluate the equilibrium samples, generated for the peptide with the PL energy function at T = 300 K and thereby characterize more precisely the binding energetics.

The ensemble of structures for the uncomplexed protein and peptide are generated by using the equilibrium trajectory of the complex and separating the protein and peptide coordinates, followed by an additional minimization of the unbound protein and unbound peptide.

The average total free energy of the molecule G is evaluated as follows:

$$G_{\text{molecule}} = G_{\text{solvation}} + E_{\text{MM}} - TS_{\text{solute}} \tag{1}$$

$$G_{\text{solvation}} = G_{\text{cavity}} + G_{\text{vdw}} + G_{\text{pol}} \tag{2}$$

In the GB/SA model, the G_{cavity} and G_{vdw} contributions are combined together via evaluating solvent-accessible surface areas:

$$G_{\text{SA}} = G_{\text{cavity}} + G_{\text{vdw}} = \sum_i \sigma_i SA_i \tag{3}$$

where G_{SA} is the nonpolar solvation term derived from the solvent-accessible surface area (SA).

$$G_{\text{pol}} = -166.0\left(1 - \frac{1}{\varepsilon}\right)\sum_i \sum_j \frac{q_i q_j}{(r_{ij}^2 + \alpha_{ij}^2 \exp(-D_{ij}))^{0.5}} \tag{4}$$

where G_{pol} is the polar solvation energy, which is computed by using the GB/SA solvation model. S_{solute} is the vibrational entropy of the molecule. E_{MM} is the molecular mechanical energy of the molecule summing up the electrostatic E_{es} interactions, van der Waals contributions E_{vdw}, and the internal strain energy E_{int}:

$$E_{\text{MM}} = E_{\text{es}} + E_{\text{vdw}} + E_{\text{int}} \tag{5}$$

Using these equations, the binding free energy of the peptide–protein complex is computed as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{peptide}} \tag{6}$$

From this equation, one can determine contributions of the peptide–protein interaction energy ΔG_{MM}, strain energy ΔG_{strain}, and solvation energy $\Delta G_{GB/SA}$ to the total binding free energy.

$$\Delta G_{\text{bind}} = \Delta G_{\text{interaction}} + \Delta G_{\text{strain}} + \Delta G_{\text{solvation}} \quad (7)$$

$$\begin{aligned}\Delta G_{\text{interaction}} &= E_{\text{MM}}^{\text{complex}} - E_{\text{MM}}^{\text{boundprotein}} - E_{\text{MM}}^{\text{boundpeptide}} \\ &= E_{\text{es}}^{\text{complex}} - E_{\text{es}}^{\text{boundprotein}} - E_{\text{es}}^{\text{boundpeptide}} \\ &\quad + E_{\text{vdw}}^{\text{complex}} - E_{\text{vdw}}^{\text{boundprotein}} - E_{\text{vdw}}^{\text{boundpeptide}} \\ &\quad + E_{\text{int}}^{\text{complex}} - E_{\text{int}}^{\text{boundprotein}} - E_{\text{int}}^{\text{boundpeptide}}\end{aligned} \quad (8)$$

$$\begin{aligned}\Delta G_{\text{strain}} &= (E_{\text{MM}}^{\text{boundprotein}} - E_{\text{MM}}^{\text{freeprotein}}) \\ &\quad + (E_{\text{MM}}^{\text{boundpeptide}} - E_{\text{MM}}^{\text{freepeptide}})\end{aligned} \quad (9)$$

$$\Delta G_{\text{solvation}} = G_{\text{solvation}}^{\text{complex}} - G_{\text{solvation}}^{\text{freeprotein}} - G_{\text{solvation}}^{\text{freepeptide}} \quad (10)$$

The energy of each peptide–protein complex is subjected to the conjugate gradient minimization as implemented in the version 7.0 of the MacroModel molecular modeling software package.[68] All protein residues within 3 Å radius sphere from the peptide are treated as flexible during minimization. All protein residues within 2 radius from the flexible shell form a first shell of restrained atoms with the force constant 100.0 $kJ/molA^2$. A second shell of restrained atoms with the force constant 200 $kJ/molA^2$ consists of the residues within 2 Å radius from the first shell and, finally, the third shell of restrained atoms is generated by the residues that reside within 2 Å from the second shell, and they are restrained with the force constant 300 $kJ/molA^2$. The remaining protein atoms are treated as frozen atoms and do not move during the minimization procedure. The interactions between frozen atoms and restrained atoms, and frozen atoms and flexible atoms are included in the total energy value. A residue-based cutoff of 8 Å is set for computing nonbonded van der Waals interactions, and 20 Å residue-based cutoff is used for computing electrostatic interactions. The peptide and protein atoms have been assigned the AMBER force-field charges. Because of significant variances in computing solute entropy using the MM/GBSA approach, this term was not included in the total binding free energy value.

Monte Carlo Simulations of Peptide–Protein Interactions

In simulations with ensembles of multiple protein conformations, each peptide replica of the peptide–protein system is associated with a protein conformation from a given ensemble. The protein conformations are linearly assigned to each temperature level, which implies a consecutive assignment of protein conformations starting from the highest temperature level and allows each protein conformation from the ensemble at least once be assigned to a certain temperature level. We have conducted equilibrium simulations with the ensembles of protein conformations using parallel simulated tempering dynamics with 50 replicas of the peptide–protein system attributed, respectively, to 50 different temperature levels that are uniformly distributed in the range between 5300K and 300K. Local Monte Carlo moves are performed independently for each replica at the corresponding temperature level, but after a simulation cycle is completed for all replicas, configuration exchanges for every pair of adjacent replicas are introduced. The mth and nth replicas, described by a common Hamiltonian $H(X)$, are associated with the inverse temperatures β_m and β_n and the corresponding conformations X_m and X_n. The exchange of conformations between adjacent replicas m and n is accepted or rejected according to Metropolis criterion with the probability $p = \min(1, \exp[-\delta])$, where $\delta = [\beta_n - \beta_m][H(X_m) - H(X_n)]$. Starting with the highest temperature, every pair of adjacent temperature configurations is tested for swapping until the final lowest value of temperature is reached. This process of swapping configurations is repeated 50 times after each simulation cycle for all replicas whereby the exchange of conformations presents an improved global update which increases thermalization of the system and overcomes slow dynamics at low temperatures on rough energy landscapes, thereby permitting regions with a small density of states to be sampled accurately. During simulation, each replica has a non-negligible probability of moving through the entire temperature range, and the detailed balance is never violated, which guarantee each replica of the system to be equilibrated in the canonical distribution with its own temperature.[84–91] Hence, we generate the canonical distribution of the ligand–protein system and the equilibrium distribution of protein conformations at each temperature. At equilibrium, the fraction of time that the peptide–protein system spends at a protein conformation $\lambda = i$ to time spent at a protein conformation $\lambda = j$ is determined by the Boltzmann distribution

$$\frac{P(\lambda_i = 1, \lambda_{m \neq i} = 0)}{P(\lambda_j = 1, \lambda_{l \neq j} = 0)} \quad (11)$$

and provides a measure for ordering protein conformations according to their interaction free energies with the ligand. The protein conformations that deliver the lowest interaction energy for the ligand during equilibrium simulation would dominate the distribution with the highest probability.

In simulations of peptide–protein interactions, the protein is held fixed in its bound conformation, whereas rigid body degrees of freedom and rotatable angles of the peptide are treated as independent variables. Peptide conformations and orientations are sampled in a parallelepiped that encompasses the binding site obtained from the crystallographic structure of the corresponding complex with a 10.0 Å cushion added to every side of this box to accurately reproduce both the unbound and bound peptide conformations. Bonds allowed to rotate include those linking sp^3 hybridized atoms to either sp^3 or sp^2 hybridized atoms and single bonds linking two sp^2 hybridized atoms. The initial peptide bond lengths, bond angles, and the torsional angles of the unrotated bonds were obtained from the crystal structures of the bound peptide–protein complexes.

Monte Carlo simulations dynamically optimize the step sizes at each temperature by taking into account the inhomogeneity of the molecular system.[92] The acceptance ratio method is used to update the step sizes every cycle of

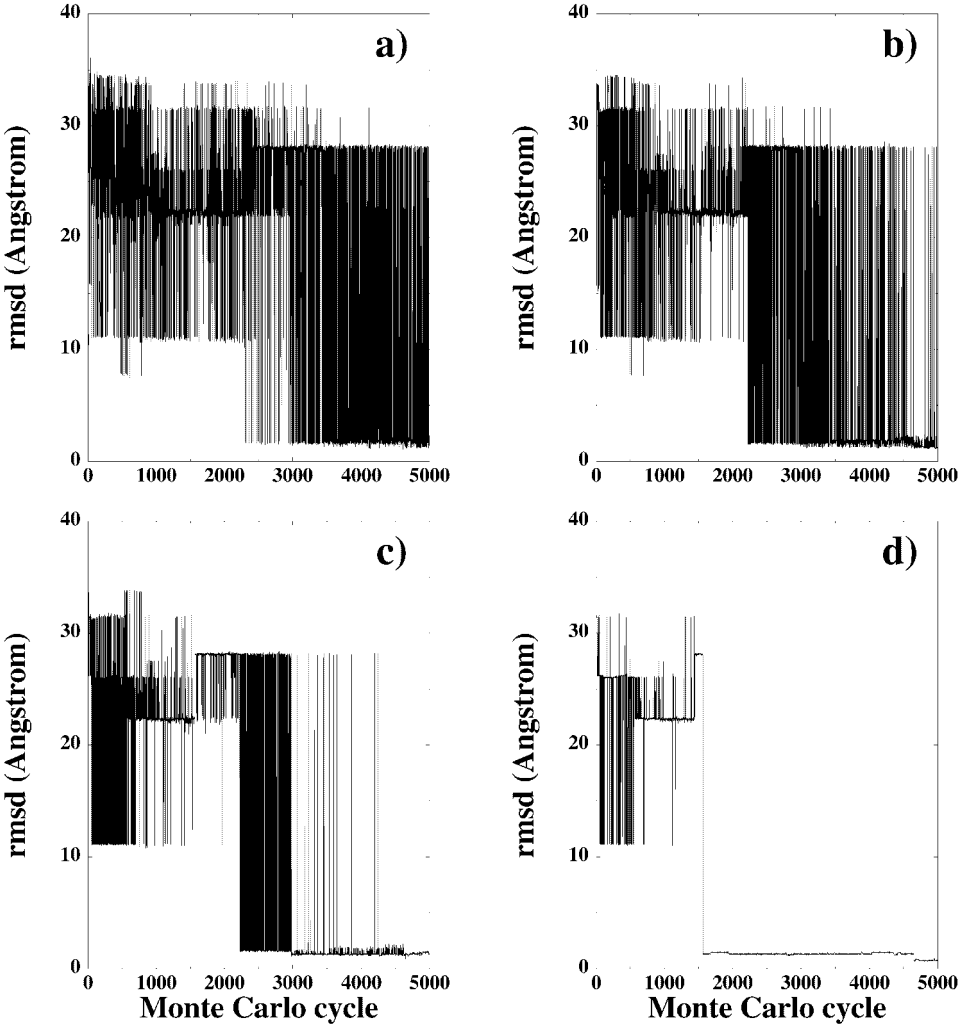

Fig. 2. Time-dependent equilibrium history of the DCAWHLGELVWCT peptide binding with Ig: T = 600K (**a**), T = 500K (**b**), T = 400K (**c**), and T = 300K (**d**).

1000 sweeps. For all these simulations, we equilibrated the system for 1000 cycles (or 1 million sweeps) and collected data during 5000 cycles (or 5 million sweeps) resulting in 5000 samples at each temperature. A sweep is defined as a single trial move for each degree of freedom of the system.

A key parameter is the acceptance ratio, which is the ratio of accepted conformations to the total number of trial conformations. At a given cycle of the simulation, each degree of freedom can change randomly throughout some prespecified range determined by the acceptance ratio obtained during the previous cycle. This range varies from 1 degree of freedom (df) to another because of the complex nature of the energy landscape. At the end of each cycle, the maximum step size is updated and used during the next cycle.

Simulations are arranged in cycles, and after a given cycle i, where the average acceptance ratio for each df j is P_j^i, the step sizes $\sigma^i{}_j$ for each df are updated for cycle $i + 1$ according to the formula

$$\sigma_j^{i+1} = \sigma_j^i \frac{\ln[a\langle P_{\text{ideal}}\rangle + b]}{\ln[a\langle P_j\rangle^i + b]} \tag{12}$$

where $\langle P_{ideal}\rangle$ is the desired acceptance ratio, chosen to be 0.5. The parameters a and b are used to ensure that the step sizes remain well behaved when the acceptance ratio approaches 0 or 1. They are assigned so that the ratio σ^{i+1}/σ^i is scaled up by a constant value s for $\langle P_j\rangle^{\text{i}} = 0$, and down by the same constant for $\langle P_j\rangle^{\text{i}} = 1$. Solving the equations

$$s^{-1} = \frac{\ln[a\langle P_{\text{ideal}}\rangle + b]}{\ln[b]} \tag{13}$$

$$s = \frac{\ln[a\langle P_{\text{ideal}}\rangle + b]}{\ln[a + b]} \tag{14}$$

with $s = 3$ yields $a = 0.673$ and $b = 0.065$.

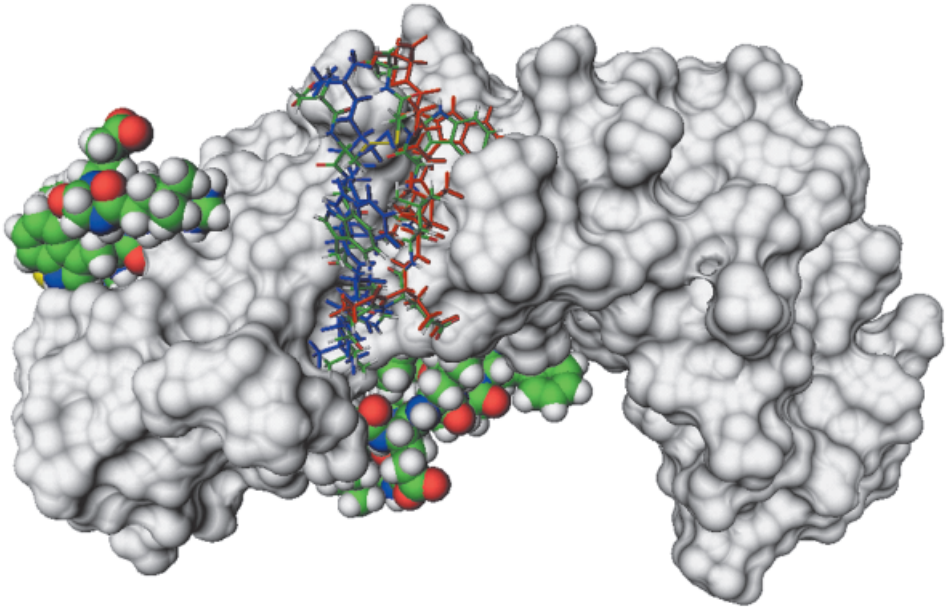

Fig. 3. Superposition of the crystallographic conformation of the bound DCAWHLGELVWCT peptide (GELVWCT portion of the peptide in red, DCAWHL portion of the peptide in blue) with the predicted lowest energy bound conformation (color coded by atom type). Two major alternative binding modes of the 13-residue peptide are shown in CPK representation and color coded by atom type. Connolly surface of the protein in the complex with the peptide is shown in gray.

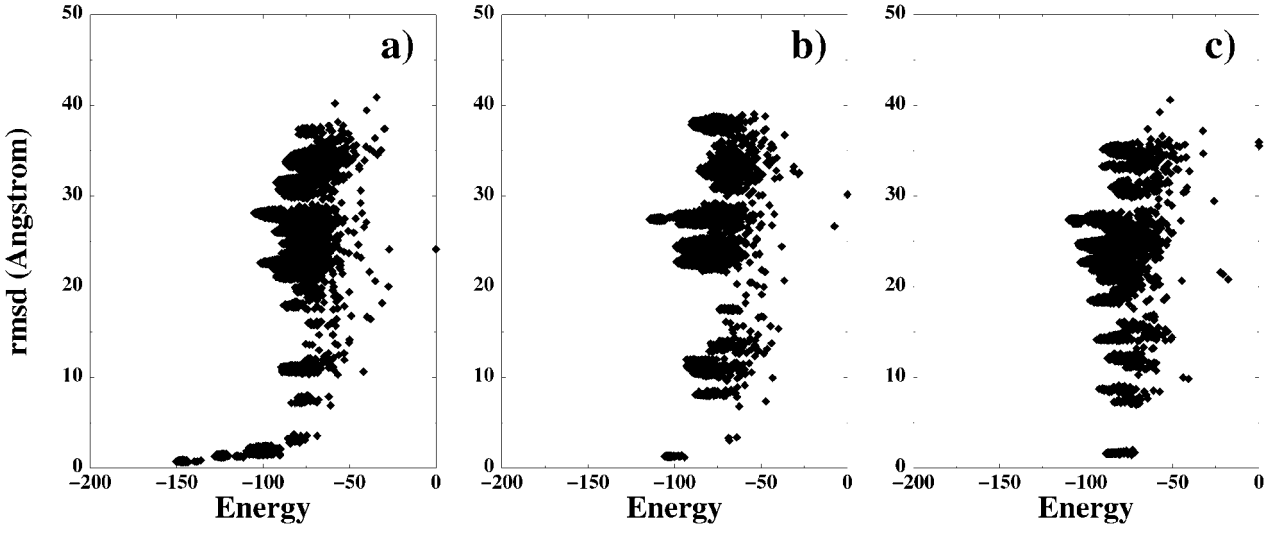

Fig. 4. The scatter graphs between the RMSDs from the crystal structure of the peptide and energies generated from equilibrium binding simulations at a range of temperatures between T = 300K and T = 1000K with the native protein conformation (**a**) and protein bound conformations from the complexes with the protein G (**b**) and protein A (**c**).

RESULTS AND DISCUSSION

Understanding structural and energetic aspects of molecular recognition implies determination of the ensemble of many similar conformations that describe the thermodynamically stable native basin of the global energy minimum rather than a single structure. The equilibrium fluctuations of the DCAWHLGELVWCT peptide, shown at the temperatures between T = 600K [Fig. 2(a)] and T = 300K [Fig. 2(d)], indicate that at higher temperatures the low-energy peptide-bound conformations fluctuate between the native-like structures at the consensus binding site and alternative binding modes of the bound peptide. As temperature lowers, the native-like conformations begin to dominate the thermodynamic equilibrium (Fig. 2). The lowest energy structure of the peptide determined from the equilibrium binding simulations with the Fc fragment of the Ig protein deviates only by root mean square deviation (RMSD) = 1.1 Å from the native struc-

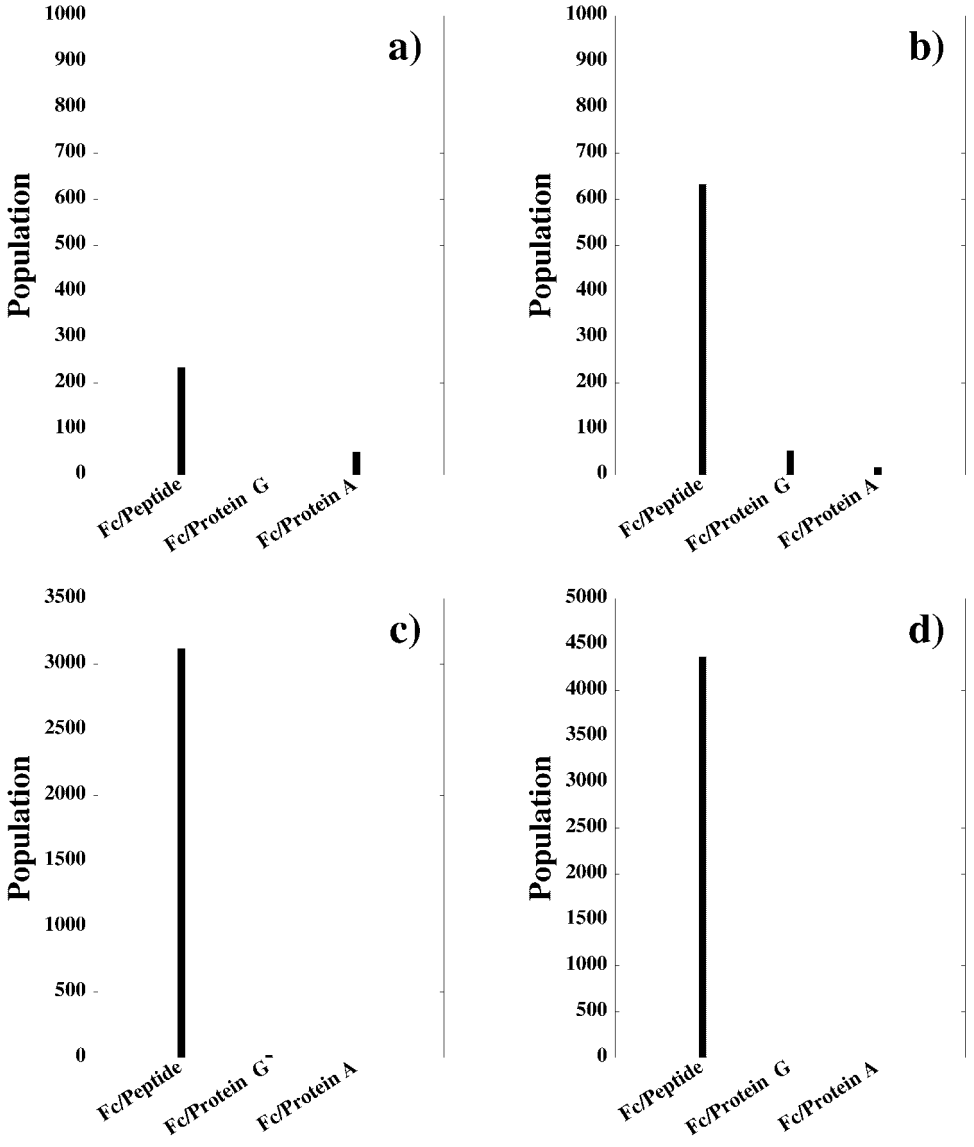

Fig. 5. The equilibrium distribution of protein conformations at T = 600K (**a**), T = 500K (**b**), T = 400K (**c**), T = 300K (**d**) determined from binding simulations of the peptide with the ensemble of three protein conformations of the Fc fragment of Ig obtained from the crystal structures in the complexes with the peptide, protein G, and protein A.

ture (Fig. 3). It is of interest that misdocked low-energy peptide conformations populated at higher temperatures interact with another highly accessible and predominantly nonpolar site on the protein surface (Fig. 3), indicating that other properties, such as shape and structural adaptability, contribute primarily to making the actual binding site a consensus locus for binding.[34]

The essence of the funnel concept is a competition between tendency toward the native structure and trapping in alternative binding modes due to ruggedness at the-bottom of the landscape. The scatter graphs between the RMSDs from the crystal structure of the peptide and energies generated from equilibrium binding simulations at a range of temperatures between 300 and 1000K reveal energy gradients, or binding funnels near the binding site, which are characterized by a steady decrease in energy of the system as the degree of similarity between the native and docked peptide structures increases [Fig. 4(a)]. However, the energy overlap between the native structure of the complex and the alternative binding modes persists throughout a wide temperature range. Only at lower temperatures, as the system begins to sample low-energy states, we observe a considerable stability gap between the crystal structure of the complex and alternative low-energy peptide conformations. This gap provides the thermodynamic stability of the native state at lower temperatures and locates unambiguously the crystal structure of the complex as the lowest energy conformation in equilibrium simulations.

To understand the role of the native protein conformation in specificity of recognition, equilibrium simulations

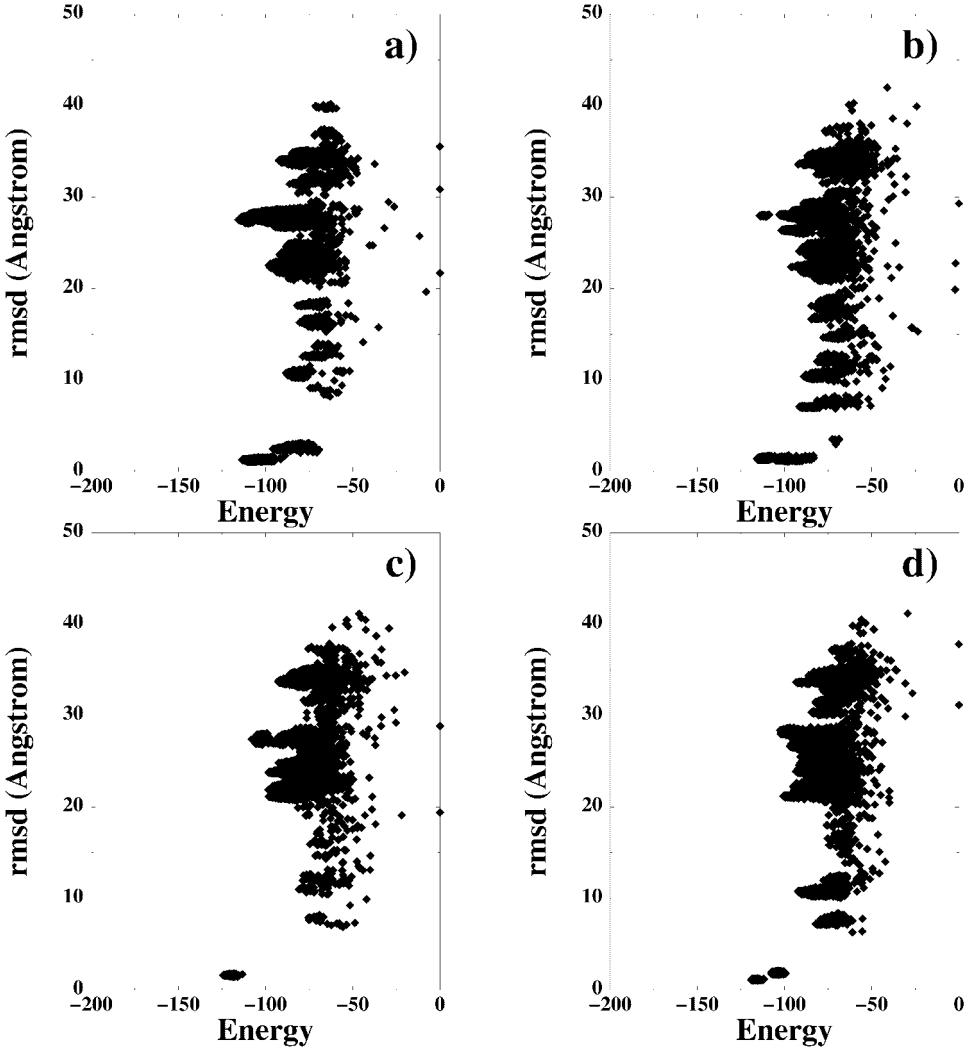

Fig. 6. The scatter graphs between the RMSDs from the crystal structure of the peptide and energies generated from equilibrium binding simulations at a range of temperatures between T = 300K and T = 1000K with the native protein conformation where Met-252 is mutated to Ala (**a**), Ile-253 is mutated to Ala (**b**), His-433 is mutated to Ala (**c**), and His-435 is mutated to Ala (**d**).

of the peptide–protein binding have been also performed with the protein conformations of the Fc fragment of Ig obtained from the crystal structures in the complexes with the protein G and protein A, binding partners that interact with the consensus binding site of Ig.[34] The corresponding scatter plots between energies and the deviations from the crystal structure of the bound peptide in the wild-type complex [Fig. 4(b,c)] display a lack of the energy separation between the crystallographic conformation of the peptide and alternative binding modes. Simulations with the ensemble of multiple protein conformations of the Fc fragment of Ig, which mimic the effects of protein flexibility in peptide–protein binding, result in the distribution of protein conformational states, which is dominated by the native protein structure (Fig. 5). Hence, the binding energy landscape has a funnel-like shape near the binding site and a distinct energy gradient toward the native structure of the complex, which are not present when the peptide interacts with alternative protein conformations of Ig.

Conserved interactions that are present in all binding interfaces with the Fc fragment of Ig share a set of contacts with the side-chains of key protein residues: His-433, His-435, Ser-254, Ile-253, Met-252, Met-428, Arg-255, Asn-434, and Tyr-436. The conserved peptide–protein interactions in the crystal structure of the native complex include hydrophobic packing onto Met-252 and His-435 residues, hydrogen bonding to the main-chain of Ile-253, and salt bridges with His-433, whereas specific interactions with the peptide are also formed via hydrogen bonding to Asn-434, Ser-254, and salt bridges with Arg-255.[34] We set out to determine the contribution of individual residues to the funnel-like character of the energy landscape near the binding site of the native complex.

Equilibrium binding simulations of the peptide–protein binding have been performed with the protein mutants

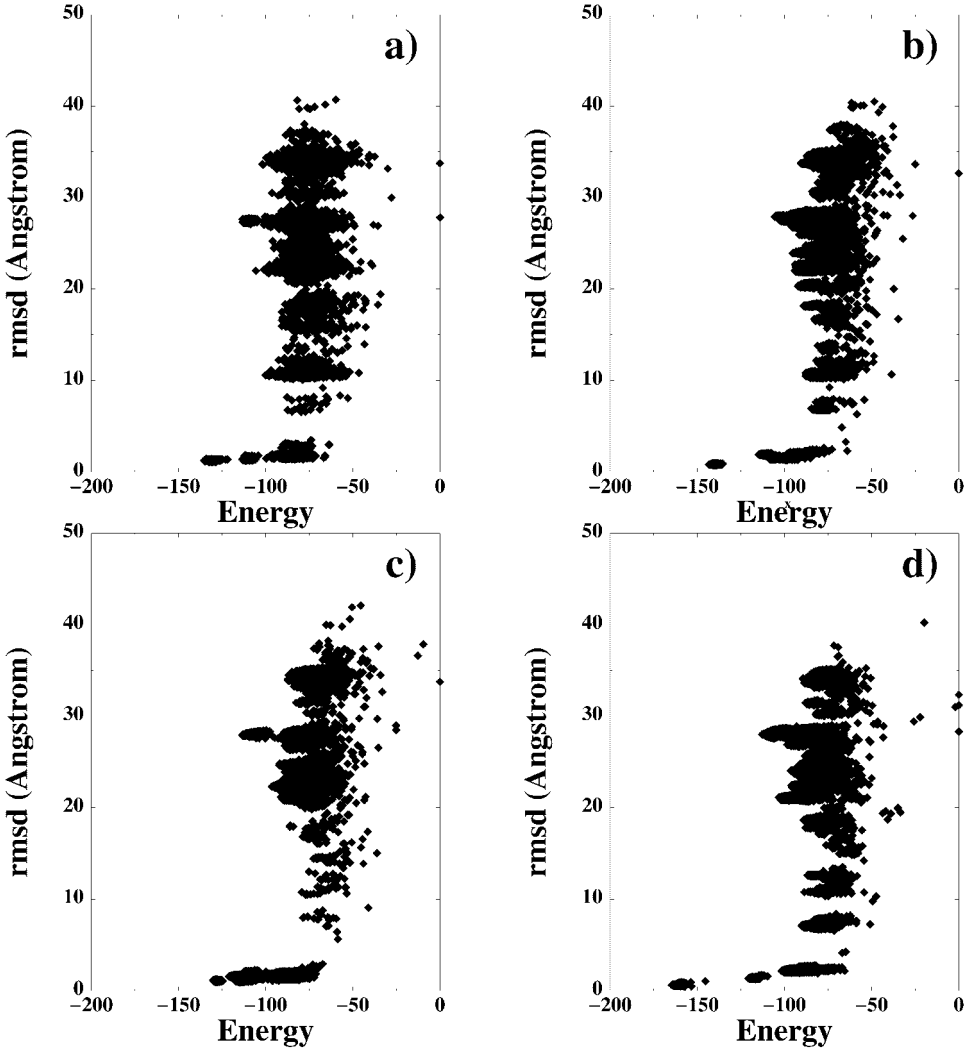

Fig. 7. The scatter graphs between the RMSDs from the crystal structure of the peptide and energies generated from equilibrium binding simulations at a range of temperatures between T = 300K and T = 1000K with the native protein conformation where Ser-254 is mutated to Ala (**a**), Arg-255 is mutated to Ala (**b**), Asn-434 is mutated to Ala (**c**), and Met-428 is mutated to Ala (**d**).

where the hot spot residues His-433, His-435, Ser 254, Ile-253, Met-252, Met-428, Arg-255, Asn-434, and Tyr-436 were mutated to Ala one at a time. The results have revealed that deletions of hydrophobic side-chains for Met-252 [Fig. 6(a)] and Ile-253 [Fig. 6(b)] and the mutations in His-433 [Fig. 6(c)] and His-435 residues [Fig. 6(d)] have a detrimental effect on the funnel-like shape of the binding energy landscape near the binding site and eliminate the energy gradient toward the native structure of the complex. The stability gap, present in the wild-type complex, is reduced for these mutants, and a significant overlap in the energies of native-like conformations and alternative binding modes is observed even at lower temperatures. In contrast, alanine scanning of Ser-254 [Fig. 7(a)], Arg-255 [Fig. 7(b)], Asn-434 [Fig. 7(c)], and Met-428 residues [Fig. 7(d)] has no effect on the funnel-like character near the binding site, which is retained when Ala substitutions of these residues are introduced in the wild-type protein. These results suggest that hydrophobic contacts and specific interactions between the peptide and the Met-252, Ile-253, His-433, and His-435 residues at the binding site may ultimately determine the funnel-like shape at the bottom of the binding energy landscape and thermodynamic stability of the native binding mode. It is important that these residues provide structural plasticity of the protein-binding site in interacting with distinct binding partners.[34] Positional and side-chain adjustments of hydrophobic residues Met-252 and Ile-253 provide a considerable extent of adaptability in the consensus binding site, and His-433 assumes different rotamer conformations depending on the binding partner (Fig. 8).

We also analyze the effect of mutations in the binding site residues on equilibrium population of the protein conformational substates on binding with the peptide. Equilibrium binding simulations of the peptide have been conducted with the ensemble of three protein conforma-

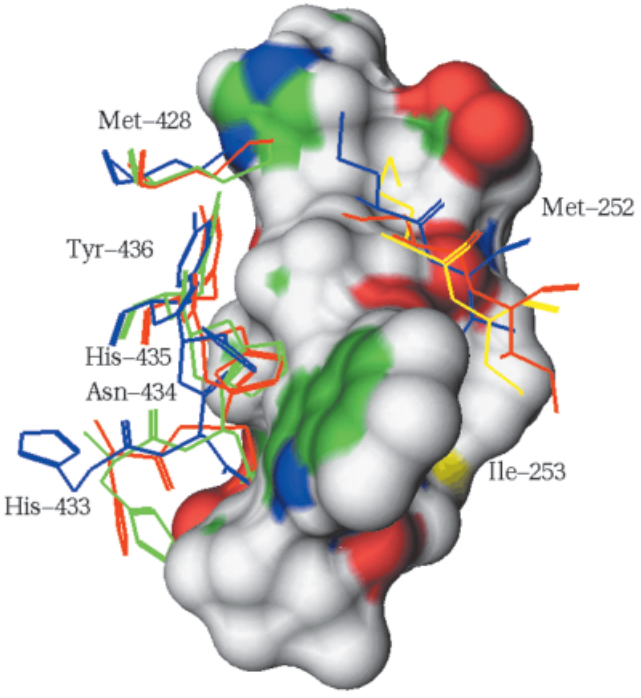

Fig. 8. Conformational variations of the key protein residues forming a hot spot at the intermolecular interface from the native complex (side-chains are shown in yellow), from the complexes with the protein G (side-chains in blue), and protein A (side-chains are shown in red). Connolly surface of the DCAWHLGELVWCT peptide color coded by atom type.

tions, obtained from the crystal structures in the complexes with the peptide, protein G, and protein A. In each of these simulations, alanine mutations of key protein-binding site residues are introduced one mutation at a time, and in each protein conformation from the ensemble. We find that alanine substitutions of the Met-252, Ile-253, His-433, and His-435 residues result in the equilibrium distribution of protein conformations with a considerable population of alternative conformational substates (Fig. 9). Therefore, mutations in the Met-252, Ile-253, His-433, and His-435 residues can shift the equilibrium distribution of the protein conformational substates from the native protein conformation to the alternative conformations, that may affect specificity of binding. In contrast, Ala mutations of Ser-254, Arg-255, Asn-434, and Met-428 residues have no effect on the equilibrium distribution of protein conformations, which continues to be solely dominated by the native structure (Fig. 10). Hence, the thermodynamic stability of the native structure is relatively insensitive to modifications of the protein residues Ser-254, Arg-255, Asn-434, and Met-428.

These results provide a quantitative assessment of the driving force behind selection of a particular highly populated protein conformation on binding to the peptide. It appears that specificity of recognition between the peptide and the native protein conformation may be primarily determined by the unfrustrated interactions with the Met-252, Ile-253, His-433, and His-435 hot spot protein residues, which are also responsible for the funnel-like shape of the energy landscape near the binding site.

Equilibrium simulations of the peptide–protein binding also have been conducted with the ensemble of 10 alanine chimeras of the native protein conformation, obtained from the crystal structure of the peptide–protein complex. This ensemble is generated by mutating Arg-255, Glu-380, Ser-254, Met-252, Met-428, Tyr-436, Ile-253, His-435, Asn-434, and His-433 residues to Ala by using the native protein structure of the peptide–protein complex. The distribution of protein conformations is presented at a range of temperatures: T = 1000K [Fig. 11(a)], T = 800K [Fig. 11(b)], T = 500 [Fig. 11(c)], and T = 300K [Fig. 11(d)]. Computationally engineered chimeras obtained by Ala mutations of the Ile-253, His-433, and His-435 residues do not contribute to the equilibrium distribution of protein fluctuations near the native binding mode of the complex. In contrast, the chimeras of the native protein with alanine substitutions in the Ser-254 and Asn-434 residues dominate the equilibrium distribution of protein conformations near the native state at lower temperatures.

Therefore, hydrogen bonding between the peptide and the Ser-254 and Asn-434 residues may be less important in providing thermodynamic stability of the native binding mode, and the hydrophobic interactions between the peptide and the Met-252, Ile-253, His-433, and His-435 residues are the driving force for providing the energy gradients near the binding site toward the native structure.

Equilibrium simulations of the peptide–protein binding have been conducted with the simplified energy function determined primarily by short-range interactions. Although long-range electrostatic forces and “electrostatic

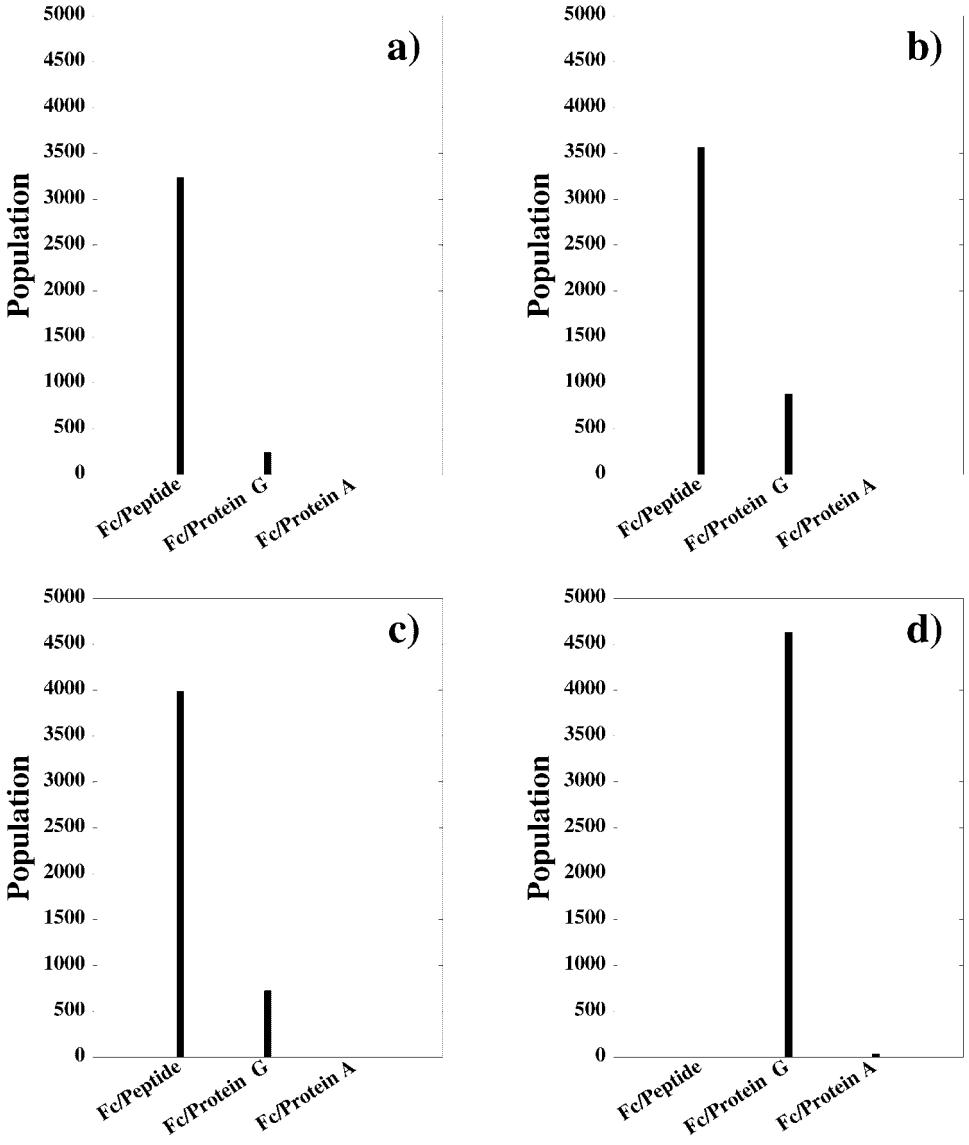

Fig. 9. The equilibrium distribution of protein conformations T = 300K determined from binding simulations of the peptide with the ensemble of three protein conformations of the Fc fragment of Ig obtained from the crystal structures in the complexes with the peptide, protein G, and protein A. In every protein conformation from the ensemble Met-252 is mutated to Ala (**a**), Ile-253 is mutated to Ala (**b**), His-433 is mutated to Ala (**c**), and His-435 is mutated to Ala (**d**).

steering" are important in enhancing the rates of association by orienting a pair of binding partners on a productive association pathway,[93,94] the final formation of the native complex is ultimately determined by a balance of van der Waals interactions and other short-range contributions, often resulting in energetically similar but structurally different binding modes separated by high-energy barriers on a frustrated binding energy landscape. Multistage binding mechanisms are mimicked in various docking approaches where a broad binding region on the protein surface is identified first by mapping electrostatic and desolvation components of the free energy, whereas a more demanding task of determining the native structure of the complex depends critically on the extent of frustration of the binding energy landscape.[95–97]

In our study, we observe the equilibrium coexistence of structurally dissimilar binding modes of the peptide at a broad temperature range, where the energy gradients at the bottom of the energy landscape near the binding site emerge only at lower temperatures. This could serve to guide the peptide and the protein into their native binding mode and accelerate the final stage of the peptide–protein association, thereby further enhancing the rate of association and providing specificity of recognition in accordance with recent findings by Zhang and coworkers.[56]

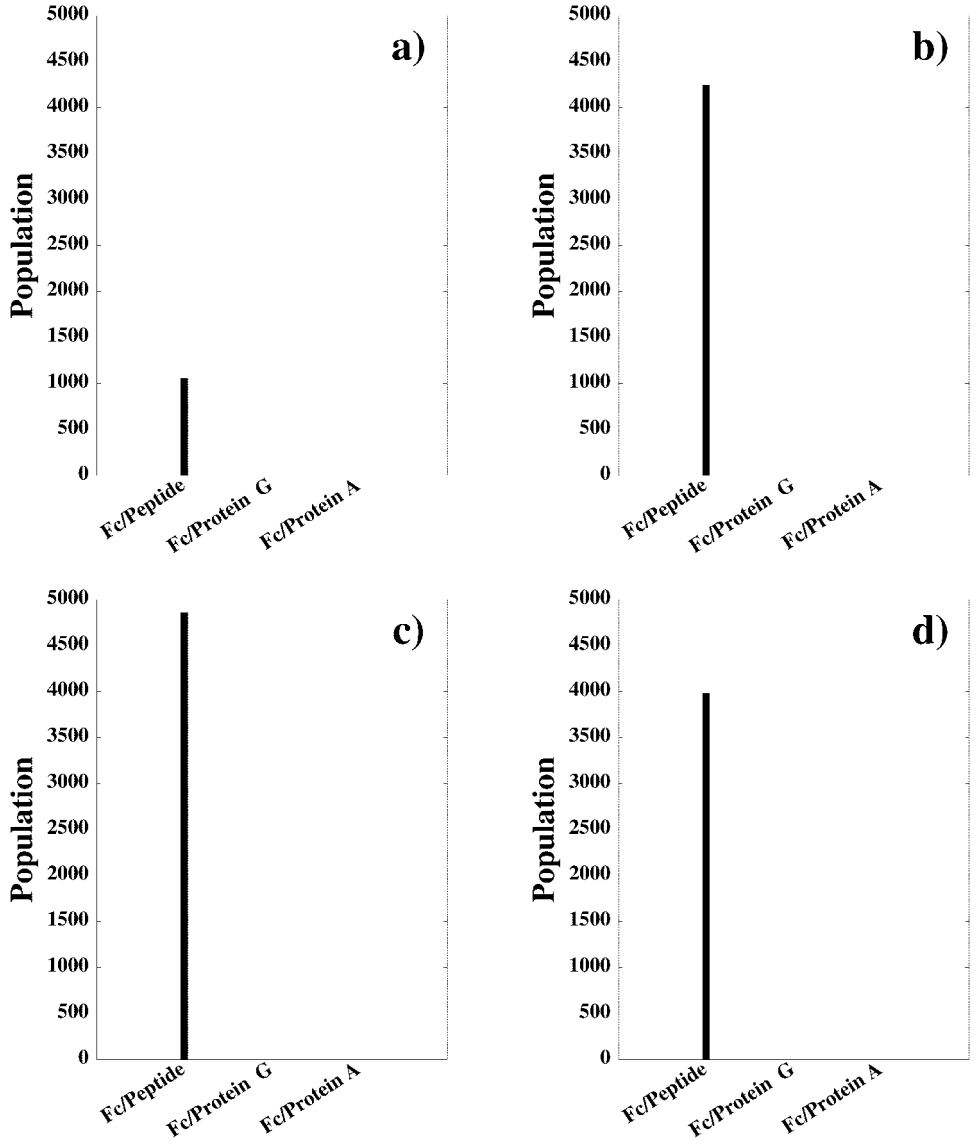

Fig. 10. The equilibrium distribution of protein conformations T = 300K determined from binding simulations of the peptide with the ensemble of three protein conformations of the Fc fragment of Ig obtained from the crystal structures in the complexes with the peptide, protein G, and protein A. In every protein conformation from the ensemble Ser-254 is mutated to Ala (**a**), Arg-255 is mutated to Ala (**b**), Asn-434 is mutated to Ala (**c**), and Met-428 is mutated to Ala (**d**).

Inferring binding mechanisms and association rates directly from the results of equilibrium simulations requires the knowledge of the transition coordinate, which is defined in kinetics as the coordinate along which the system progresses most slowly and not necessarily associated with any particular path in the phase space.[45,98] The structural properties of the transition state ensemble may be determined by sampling the thermal distribution of states in the dominant free energy barrier only for highly unfrustrated energy landscapes.[99] With an increased frustration, the free energy profile may become more complicated, and the transition state ensemble can no longer be adequately described as the free energy barrier along a simple reaction coordinate. An estimation of the kinetic rate of the association process based on equilibrium simulations can be in principle obtained by considering binding as a diffusion process and by applying a Fokker-Plank equation to relate the diffusion on a multidimensional surface to the shape of the binding energy landscape.[99] The approach to directly infer kinetics and determine the transition state ensemble from equilibrium sampling can be problematic in both protein folding[45,99] and ligand–protein binding.[100] Simple order parameters measuring the degree of similarity with the native structure may not be rigorous in approximating the reaction coordinate even for a relatively unfrustrated binding energy landscape of

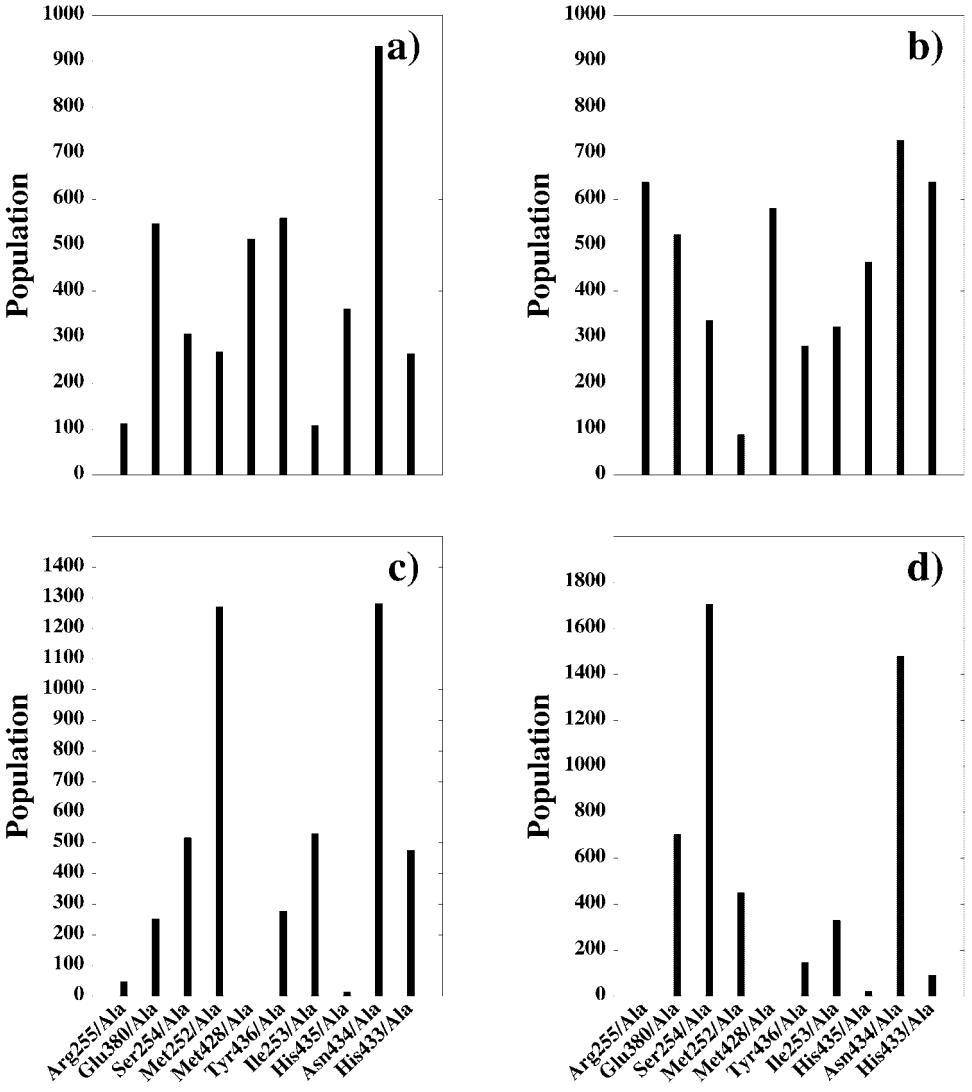

Fig. 11. The equilibrium distribution of protein conformations at T = 1000K (**a**), T = 800K (**b**), T = 500 (**c**), and T = 300K (**d**) determined from binding simulations of the peptide with the ensemble of 10 chimeras of the native protein bound conformation taken from the original crystal structure of the peptide–protein complex. Each of 10 chimeras was generated by introducing the corresponding Ala mutation in the native protein conformation for each of the following residues: Arg-255, Glu-380, Ser-254, Met-252, Met-428, Tyr-436, Ile-253, His-435, Asn-434, and His-433.

the methotrexate (MTX)–dihydrofolate reductase (DHFR) system.[100] A rigorous but computationally very demanding approach is to determine the transition state as the ensemble of conformations that have equal probability of reaching the unbound states and the native structure. This method can in principle determine the transition state ensemble without explicitly knowing the reaction coordinate and regardless of the complexity of the free energy landscape. This analysis extends beyond the scope of the present study and will be discussed elsewhere.

A common GELVWCT sequence emerged as a consensus motif among the recognition peptides capable of binding to the consensus site of the Fc fragment of Ig.[34] To understand this experimental observation, we have dissected the 13-residue peptide into the consensus motif GELVWCT and the remaining 6-residue DCAWHL peptide and conducted equilibrium binding simulations with the Fc fragment of Ig for each of these smaller peptide motifs. The consensus motif GELVWCT rapidly achieves the thermodynamically stable native conformation, which is in structural harmony with its conformation when embedded in the complete peptide structure (Fig. 12). In contrast, binding simulations with the DCAWHL motif have revealed a number of distinct low-energy conformations, even when the DCAWHL peptide conformation was kept regid in its crystallographic conformation. It is important that only a small fraction of the low-energy bound conformations for the DCAWHL motifs resembles the conforma-

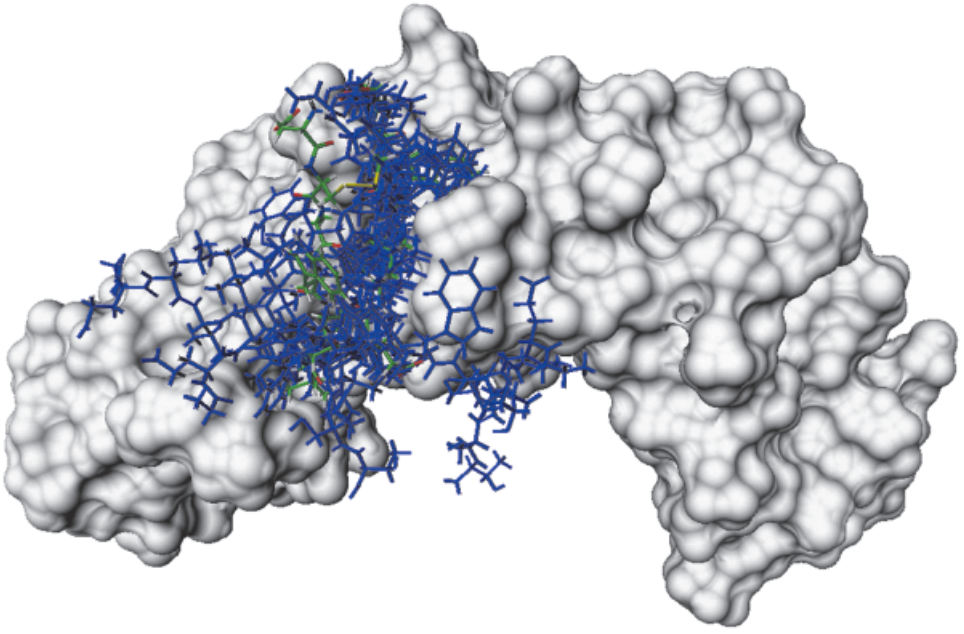

Fig. 12. Superposition of the crystallographic conformation of the bound DCAWHLGELVWCT peptide (color coded by atom type) with the pool of determined low-energy bound conformations of the GELVWCT recognition motif of the peptide (shown in blue). Connolly surface of the protein in the complex with the peptide is shown in gray. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

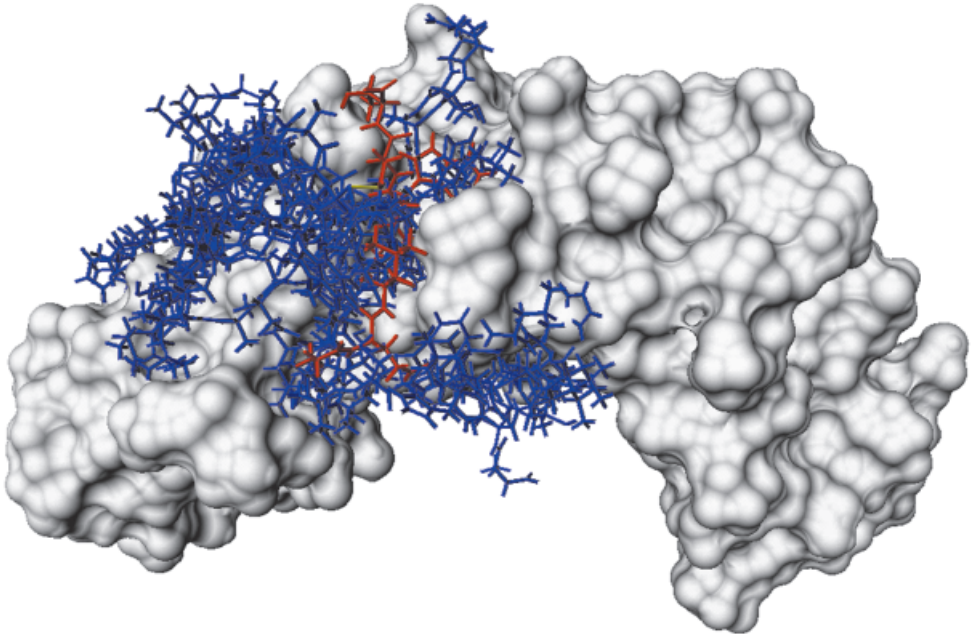

Fig. 13. Superposition of the crystallographic conformation of the bound DCAWHLGELVWCT peptide (color coded by atom type) with the pool of determined low-energy bound conformations of the DCAWHL portion of the peptide (shown in blue). Connolly surface of the protein in the complex with the peptide is shown in gray. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tion of this segment in the native complex of the complete 13-residue peptide (Fig. 13). The scatter graphs between energies and the RMSDs from the crystallographic conformations of the GELVWCT and DCAWHL motifs, when embedded in the structure of the bound 13-residue peptide, have been collected from equilibrium binding simulations at a range of temperatures between 300 and 1000K (Fig. 14). Although there is binding funnels near the

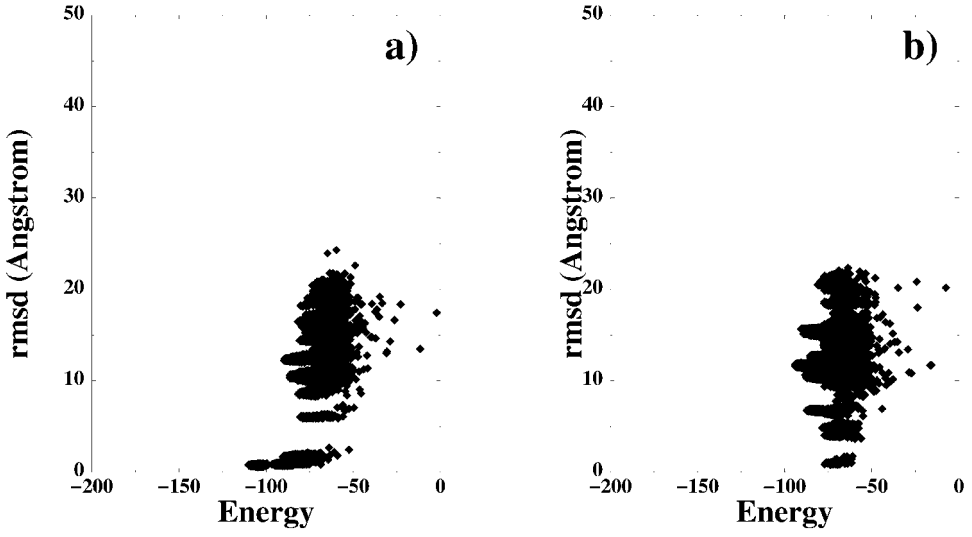

Fig. 14. The scatter graphs between the RMSDs from the crystal structure of the peptide and energies generated from equilibrium binding simulations with the ensemble of three protein conformations at a range of temperatures between T = 300K and T = 1000K for the GELVWCT recognition motif of the peptide (**a**) and the DCAWHL portion of the peptide (**b**).

binding site for the recognition GELVWCT motif [Fig. 14(a)] there is a considerable overlap between the low-energy solutions for the DCAWHL motif [Fig. 14(b)], and the lowest energy conformation for this motif is located away from the binding site [Figs. 13 and 14(b)].

These results suggest that there may be a thermodynamic advantage for the recognition peptides to share the consensus motif, which provides a stable structural platform for the remainder of the peptide in recognizing the binding site. Structural stability of the recognition GELVWCT sequence is consistent with the principle of minimal frustration[39–43] (i.e., the unfrustrated interactions between the recognition motif and the protein at the binding site result in a high propensity to form a stable low-energy structure in isolation, which is consistent with the one found in the native peptide–protein complex.

We have also determined the contribution of key protein binding site residues of the Fc fragment of Ig in the binding affinity with the specific 13-residue peptide. Binding free energy calculations have been conducted for the native complex and peptide complexes with 10 different protein chimeras generated by mutating Arg-255, Glu-380, Ser-254, Met-252, Met-428, Tyr-436, Ile-253, His-435, Asn-434, and His-433 residues, respectively, to Ala using the native protein structure of the peptide–protein complex. The binding free energies are computed by using the MM/GBSA approach by averaging the energy contributions over the conformational samples generated in equilibrium simulations with the simplified energy function at T = 300K.

We find that the most dramatic loss of binding affinity occurs when the Asn-434, His-433, His-435, and Tyr-436 residues are replaced by alanine (Fig. 15). These results agree with the mutagenesis experiments, which have shown that alanine substitution of Asn-434, His-435, or Tyr-436 disrupts binding of the selected peptide.[34] Hence,

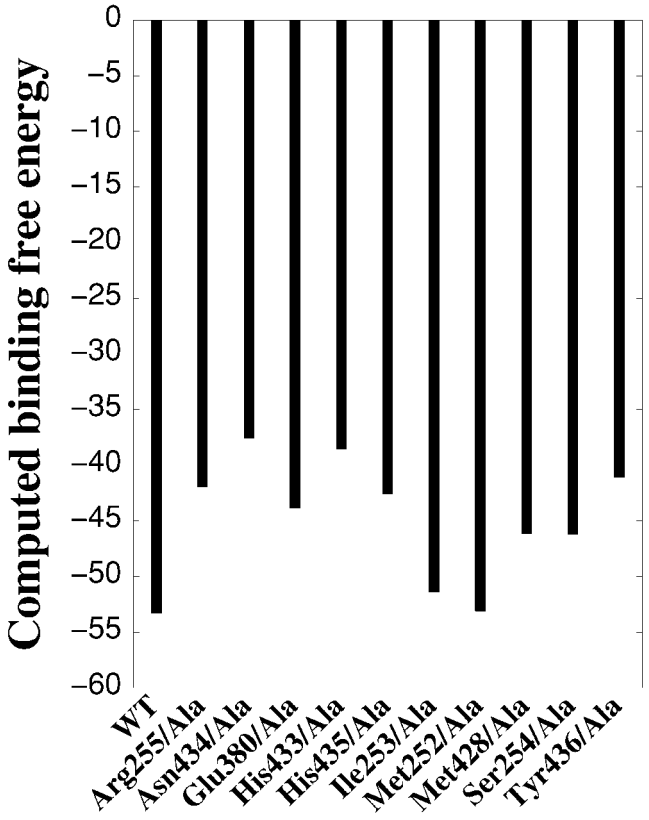

Fig. 15. Computed binding free energies (kcal/mol) for the complexes between a panel of alanine mutants of the Fc fragment of Ig and the DCAWHLGELVWCT peptide.

the His-433 and His-435 residues may critically contribute to both binding affinity of the peptide–protein complex and to the thermodynamic stability of the native binding mode.

In contrast, the conserved protein residues Asn-434, Ser-254, Tyr-436, which contribute significantly to the binding affinity with the peptide, have been found to be less critical in providing thermodynamic stability of the native complex.

CONCLUSIONS

Monte Carlo simulations of molecular recognition at the consensus binding site of the constant fragment (Fc) of human immunoglobulin G (Ig) protein have been performed to analyze structural and thermodynamic aspects of binding for the 13-residue cyclic peptide DCAWHLGELVWCT. The energy landscape analysis of a hot spot at the intermolecular interface using alanine scanning and equilibrium simulated tempering dynamics with the simplified, knowledge-based energy function has enabled the role of the protein hot spot residues in providing the thermodynamic stability of the native structure to be determined.

To represent alternate protein conformational substates, we apply a simulation approach to model ligand binding with the ensembles of multiple protein conformations This method can account for protein flexibility by considering a finite number of protein states that have significant differences in both side-chain and main-chain conformations.

We have found that specificity of recognition between the peptide and the native protein conformation may be primarily determined by the unfrustrated interactions with the Met-252, Ile-253, His-433, and His-435 hot spot protein residues, which are also responsible for the funnel-like shape of the energy landscape near the binding site. We have also discovered that binding of the consensus peptide motif GELVWCT with the Fc fragment of Ig is characteristic of a minimally frustrated energy landscape, where in isolation this motif maintains structural stability of the dominant bound conformation observed in the crystal structure of the full peptide. The interactions of the consensus GELVWCT motif with the hot spot protein residues are primarily responsible for providing energy gradients on the energy landscape near the binding site.

Alanine scanning and binding free energy calculations using a molecular mechanics force field combined with a solvation energy term have been performed to determine the energetic contribution of the protein hot spot residues in binding affinity. The conserved Asn-434, Ser-254, and Tyr-436 protein residues contribute significantly to the binding affinity of the peptide–protein complex, serving as an energetic hot spot at the intermolecular interface. Although the consensus binding site is mostly hydrophobic, the conserved polar residues contribute to specific recognition with the peptide and largely determine binding affinity of the peptide–protein complex serving as the energetic hot spots at the intermolecular interface.

These results suggest that evolutionary conserved hot spot protein residues at the intermolecular interface may be partitioned in fulfilling the binding affinity of the peptide–protein complex and providing the thermodynamic stability of the native binding mode.

REFERENCES

1. McCammon JA. Theory of biomolecular recognition. Curr Opin Struct Biol 1998;8:245–249.
2. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. Mol Pharmacol 2000;57:213–218.
3. Van Regenmortel MH. Molecular recognition in the post-reductionist era. J Mol Recognit 1999;12:1–2.
4. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R. Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci 2000;9:10–19.
5. Frauenfelder H. Complexity in proteins. Nat Struct Biol 1995;2: 821–823.
6. Leeson DT, Wiersma DA. Looking into the energy landscape of myoglobin. Nat Struct Biol 1995;2:848–851.
7. Frauenfelder H, Leeson DT. The energy landscape in non-biological and biological molecules. Nat Struct Biol 1998;5:757–759.
8. Rejto PA, Freer ST. Protein conformational substates from X-ray crystallography. Prog Biophys Mol Biol 1996;66:167–196.
9. Tsai CJ, Ma B, Nussinov R. Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci USA 1999;96:9970–9972.
10. Sinha N, Nussinov R. Point mutations and sequence variability in proteins: redistributions of preexisting populations. Proc Natl Acad Sci USA 2001;98:3139–3144.
11. Keskin O, Jernigan RL, Bahar I. Proteins with similar architecture exhibit similar large-scale dynamic behavior. Biophys J 2000;78:2093–2106.
12. Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD. Correlation between binding and dynamics and dynamics at SH2 domain interfaces. Nat Struct Biol 1998;5:156–163.
13. Forman-Kay JD. The "dynamics" in the thermodynamics of binding. Nat Struct Biol 1999;6:1086–1087.
14. Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol 2000;7:72–77.
15. Zidek L, Novotny MV, Stone MJ. Increased protein backbone conformational entropy upon hydrophobic ligand binding. Nat Struct Biol 1999;6:1118–1121.
16. Cavanagh J, Akke M. May the driving force be with you—whatever it is. Nat Struct Biol 2000;7:11–13.
17. Todd MJ, Freire E. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. Proteins 1999;36:147–156.
18. Freire E. The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. Proc Natl Acad Sci USA 1999;96:10118–10122.
19. Luque I, Freire E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. Proteins 2000; Suppl 4:63–67.
20. Freire E. Can allosteric regulation be predicted from structure? Proc Natl Acad Sci USA 2000;97:11680–11682.
21. Velazquez-Campoy A, Todd MJ, Freire E. HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. Biochemistry 2000;39:2201–2207.
22. D'Aquino JA, Freire E, Amzel LM. Binding of small organic molecules to macromolecular targets: evaluation of conformational entropy changes. Proteins 2000;Suppl 4:93–107.
23. Clackson T, Wells J. A hot spot of binding energy in a hormone–receptor interface. Science 1995;267:383–386.
24. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 2002;12:14–20.
25. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
26. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
27. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
28. Hu Z, Ma B, Wolfson H, Nussinov R. Conservation of polar residues as hot spots at protein interfaces. Proteins 2000;39:331–342.
29. Ma B, Wolfson HJ, Nussinov R. Protein functional epitopes: hot spots, dynamics and combinatorial libraries. Curr Opin Struct Biol 2001;11:364–369.
30. Janin J. Wet and dry interfaces: the role of solvent in protein–

protein and protein–DNA recognition. Structure 1999;7:R277–R279.
31. Atwell S, Ultsch M, De Vos AM, Wells JA. Structural plasticity in a remodeled protein-protein interface. Science 1997;278:1125–1128.
32. Sundberg EJ, Mariuzza RA. Luxury accommodations: the expanding role of structural plasticity in protein-protein interactions. Structure Fold Des 2000;8:R137–142.
33. Demchenko AP. Recognition between flexible protein molecules: induced and assisted folding. J Mol Recognit 2001;14:42–61.
34. DeLano WL, Ultsch MH, de Vos AM, Wells JA. Convergent solutions to binding at a protein-protein interface. Science 2000; 287:1279–1283.
35. Tondi D, Slomczynska U, Costi MP, Watterson DM, Ghelli S, Shoichet BK. Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. Chem Biol 1999;6:319–331.
36. Tsai C-J, Xu D, Nussinov R. Protein folding via binding and vice versa. Curr Biol 1998;3:R71–R80.
37. Tsai C-J, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels and protein function. Protein Sci 1999;8:1181–1190.
38. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding ata single protein site: a matter of pre-existing populations. Protein Sci 2002;11:184–197.
39. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding, a synthesis. Proteins 1995;21:167–195.
40. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principle of protein folding—a perspective from simple exact models. Protein Sci 1995;4:561–602.
41. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
42. Shakhnovich EI. Theoretical studies of protein—folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
43. Onuchic JN, Lutney-Schilten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
44. Shea J-E, Brooks CL III. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 2001;52:499–535.
45. Mirny L, Shakhnovich E. Protein folding theory: from lattice models to all-atom models. Annu Rev Biomol Struct 2001;30:361–396.
46. Kuntz ID, Meng EC, Shoichet BK. Structure-based molecular design. Acc Chem Res 1994;27:117–123.
47. Kollman P. Free energy calculations-applications to chemical and biological phenomena. Chem Rev 1993;93:2395–2417.
48. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand-receptor complexes. J Med Chem 1995;38: 4953–4967.
49. Rosenfeld R, Vajda S, DeLisi C. Flexible docking and design. Annu Rev Biophys Biomol Struct 1995;24:677–700.
50. Janin J. Quantifying biological specificity: the statistical mechanics of molecular recognition. Proteins 1996;25:438–445.
51. Rejto PA, Verkhivker GM. Unraveling principles of lead discovery: from unfrustrated energy landscapes to novel molecular anchors. Proc Natl Acad Sci USA 1996;93:8945–8950.
52. Verkhivker GM, Rejto PA. A mean field model of ligand–protein interactions: implications for the structural assessment of human immunodeficiency virus type 1 protease complexes and receptor-specific binding. Proc Natl Acad Sci USA 1996;93:60–64.
53. Verkhivker GM, Rejto PA, Gehlhaar DK, Freer ST. Exploring energy landscapes of molecular recognition by a genetic algorithm, analysis of the requirements of robust docking of HIV-1 protease and FKBP-12 complexes. Proteins 1996;25:342–353.
54. Rejto PA, Verkhivker GM, Gehlhaar DK, Freer ST. New trends in computational structure prediction of ligand-protein complexes for receptor-based drug design. In: van Gunsteren W, Weiner P, Wilkinson AJ, editors. Computational simulation of biomolecular systems. Leiden: ESCOM; 1997. p 451–465.
55. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Schaffer L, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Hierarchy of simulation models in predicting molecular recognition mechanisms from the binding energy landscapes: structural analysis of the peptide complexes with SH2 domains. Proteins 2001;45:456–470.
56. Zhang C, Chen J, DeLisi C. Protein–protein recognition: exploring the energy funnels near the binding sites. Proteins 1999;34: 255–267.
57. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein–protein complexes. Proc Natl Acad Sci USA 19990;96:8477–8482.
58. Tovchigrechko A, Vakser IA. How common is the funnel-like energy landscape in protein-protein interactions? Protein Sci 2001;10:1572–1583.
59. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Schaffer L, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Monte Carlo simulations of HIV-1 protease binding dynamics and thermodynamics with ensembles of protein conformations: incorporating protein flexibility in deciphering mechanisms of molecular recognition. In: Eriksson LA, editor. Theoretical biochemistry—processes and properties of biological systems. Theoretical and computational chemistry. Amsterdam: Elsevier; 2001. p 289–340.
60. Verkhivker GM, Rejto PA, Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Marrone T, Rose PW. Parallel simulated tempering dynamics of ligand–protein binding with ensembles of protein conformations. Chem Phys Lett 2001;337:181–189.
61. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Deciphering common failures in molecular docking of ligand-protein complexes. J Comput Aided Mol Des 2000;14:731–751.
62. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Schaffer L, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Hierarchy of simulation models in predicting structure and energetics of the Src SH2 domain binding to the tyrosyl phosphopeptides. J Med Chem 2002;45:72–89.
63. Bouzida D, Rejto PA, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Rose PW, Verkhivker GM. Computer simulations of ligand–protein binding with ensembles of protein conformations: a Monte Carlo study of HIV-1 protease binding energy landscapes. Int J Quantum Chem 1999;72:73–84.
64. Bouzida D, Rejto PA, Verkhivker GM. Monte Carlo simulations of ligand–protein binding energy landscapes with the weighted histogram analysis method. Int J Quantum Chem 1999;73:113–121.
65. Verkhivker GM, Rejto PA, Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Marrone T, Rose PW. Towards understanding the mechanisms of molecular recognition by computer simulations of ligand-protein interactions. J Mol Recognit 1999;12:371–389.
66. Weiner SJ, Kollman PA, Case DA, Singh UC, Chio C, Alagona G, Profeta S, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765–784.
67. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
68. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J Comput Chem 1990;11: 440–467.
69. Qiu D, Shenkin, PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate born radii. J Phys Chem A 1997;101: 3005–3014.
70. Weiser J, Weiser AA, Shenkin PS, Still WC. Neighbor-list reduction: optimization for computation of molecular van der Waals and solvent-accessible surface areas. J Comput Chem 1998;19:797–808.
71. Weiser J, Weiser AA, Shenkin PS, Still WC. Erratum: Neighbor-list reduction: optimization for computation of molecular van der Waals and solvent-accessible surface areas. J Comput Chem 1998;19:1110.
72. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comput Chem 1999;20:217–230.
73. Weiser J, Shenkin PS, Still WC. Fast, approximate algorithm for

detection of solvent-inaccessible atoms. J Comput Chem 1999;20: 586–596.
74. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA and phosphoramidate–DNA helices. J Am Chem Soc 1998;120:9401–9409.
75. Massova I, Kollman PA. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. J Am Chem Soc 1999;121:8133–8143.
76. Chong LT, Duan Y, Wang L, Massova I, Kollman PA. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. Proc Natl Acad Sci USA 1999;96:14330–14335.
77. Kuhn B, Kollman PA. A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. J Am Chem Soc 2000;122:3909–3916.
78. Kuhn B, Kollman PA. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J Med Chem 2000;43:3786–3791.
79. Lee MR, Duan Y, Kollman PA. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. Proteins 2000;39:309–316.
80. Wang W, Kollman PA. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. J Mol Biol 2000;303:567–582.
81. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
82. Tsui V, Case DA. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. J Am Chem Soc 2000;122:2489–2498.
83. Mayo SL, Olafson BD, Goddard WA III. DREIDING: a generic force field for molecular simulation. J Phys Chem 1990;94:8897–8909.
84. Marinari E, Parisi G. Simulated tempering: a new Monte Carlo scheme. Europhys Lett 1992;19:451–458.
85. Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc (Jap) 1996;65: 1604–1607.
86. Hansmann UHE, Okamoto Y. Monte Carlo simulations in generalized ensemble: multicanonical algorithm versus simulated tempering. Phys Rev E 1996;54:5863–5865.
87. Hansmann UHE, Okamoto Y. Generalized-ensemble Monte Carlo method for systems with rough energy landscape. Phys Rev E 1997;56:2228–2233.
88. Hansmann UHE, Okamoto Y. Numerical comparisons of three recently proposed algorithms in the protein folding problem. J Comput Chem 1997;18:920–933.
89. Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. Chem Phys Lett 1997;281: 140–150.
90. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–151.
91. Verkhivker GM, Rejto PA, Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Marrone T, Rose PW. Navigating ligand–protein binding free energy landscapes: universality and diversity of protein folding and molecular recognition mechanisms. Chem Phys Lett 2001;336:495–503.
92. Bouzida D, Kumar S, Swendsen RH. Efficient Monte Carlo methods for the computer simulation of biological molecules. Phys Rev A 1992;45:8894–8901.
93. Selzer T, Albeck S, Schreiber G. Rational design of faster associating and tighter binding protein complexes. Nat Struct Biol 2000;7:537–541.
94. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein–protein association. Biophys J 1999;76:1166–1178.
95. Camacho CJ, Vajda S. Protein docking along smooth association pathways. Proc Natl Acad Sci USA 2001;98:10636–10641.
96. Camacho CJ, Vajda S. Protein–protein association kinetics and protein docking. Curr Opin Struct Biol 2002;12:36–40.
97. Fernandez-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. Protein Sci 2002;11:280–291.
98. Du R, Pande V, Grosberg A, Tanaka T, Shakhnovich E. On the transition coordinate for protein folding. J Chem Phys 1998;108: 334–350.
99. Nymeyer H, Socci ND, Onuchic JN. Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration. Proc Natl Acad Sci USA 2000;97:634–639.
100. Verkhivker GM, Rejto PA, Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Marrone T, Rose PW. Navigating ligand–protein binding free energy landscapes: universality and diversity of protein folding and molecular recognition mechanisms. Chem Phys Lett 2001;336:495–503.