Hierarchy of simulation models in predicting molecular recognition mechanisms from the binding energy landscapes: Structural analysis of the peptide complexes with SH2 domains

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · DECEMBER 2001

Impact Factor: 2.63 · DOI: 10.1002/prot.10019

PROTEINS: Structure, Function, and Genetics 45:456–470 (2001)

Hierarchy of Simulation Models in Predicting Molecular Recognition Mechanisms From the Binding Energy Landscapes: Structural Analysis of the Peptide Complexes With SH2 Domains

Gennady M. Verkhivker,* Djamal Bouzida, Daniel K. Gehlhaar, Paul A. Rejto, Lana Schaffer, Sandra Arthurs, Anthony B. Colson, Stephan T. Freer, Veda Larson, Brock A. Luty, Tami Marrone, and Peter W. Rose
Agouron Pharmaceuticals, Inc., San Diego, California

ABSTRACT **Computer simulations using the simplified energy function and simulated tempering dynamics have accurately determined the native structure of the pYVPML, SVLpYTAVQPNE, and SPGEpYVNIEF peptides in the complexes with SH2 domains. Structural and equilibrium aspects of the peptide binding with SH2 domains have been studied by generating temperature-dependent binding free energy landscapes. Once some native peptide–SH2 domain contacts are constrained, the underlying binding free energy profile has the funnel-like shape that leads to a rapid and consistent acquisition of the native structure. The dominant native topology of the peptide–SH2 domain complexes represents an extended peptide conformation with strong specific interactions in the phosphotyrosine pocket and hydrophobic interactions of the peptide residues C-terminal to the pTyr group. The topological features of the peptide–protein interface are primarily determined by the thermodynamically stable phosphotyrosyl group. A diversity of structurally different binding orientations has been observed for the amino-terminal residues to the phosphotyrosine. The dominant native topology for the peptide residues carboxy-terminal to the phosphotyrosine is tolerant to flexibility in this region of the peptide–SH2 domain interface observed in equilibrium simulations. The energy landscape analysis has revealed a broad, entropically favorable topology of the native binding mode for the bound peptides, which is robust to structural perturbations. This could provide an additional positive mechanism underlying tolerance of the SH2 domains to hydrophobic conservative substitutions in the peptide specificity region. Proteins 2001;45:456–470.**

Key words: molecular recognition; Monte Carlo simulations; molecular docking; binding free energy profiles; weighted histogram analysis; SH2 domains; peptide optimization; protein folding

INTRODUCTION

Structure and energetics of peptide–SH2 domain binding have been studied extensively to understand the molecular basis of sequence-specific recognition.[1–18] The crystal structures of two closely related SH2 domains of the tyrosine kinases Src and Lck, bound to the phosphotyrosyl peptide that contains the recognition pYEEI motif, have suggested a mechanism of sequence-specific binding that resembles a "two-pronged plug" engaging a "two-holed socket."[1,2] The peptide is bound in an extended conformation with the phosphotyrosyl pY group located in a deep pocket and stabilized by a network of hydrogen bonds and electrostatic interactions. The original mechanism of the Src SH2 domain recognition has proposed that the occupation of the +3 binding pocket by the Ile residue is the critical determinant of specificity and high binding affinity.[5,6] However, thermodynamic evaluations of peptide–SH2 domain binding have shown that the differences in the SH2 domain binding affinity between consensus, high-affinity peptide, and nonconsensus low-affinity peptides are <100-fold.[6–11] Mutations in the consensus pYEEI binding sequences cause only modest, <10-fold reductions in affinity, whereas a much larger 10,000-fold decline in binding affinity is observed when the pTyr residue is perturbed.[8–11] It has been demonstrated that more than half of the binding free energy of the pYEEI peptide is derived from the phosphate in the pTyr. In contrast, mutations of the residues at +1, +2, or +3 positions C-terminal to the pTyr cause only minor changes in binding affinity.[8–11]

The structural and thermodynamic studies of binding for the specific, high-affinity PQpYEEIPI peptide and low-affinity TQpYVPMLE and PQpYQPGEN peptides have revealed the dominant role of the enthalpy contribution for the recognition peptide and the major entropy component in binding of the low-affinity peptides.[6] The crystal structures for these peptides in complexes with the SH2 domain have shown that the interactions of the low-affinity peptides with the protein in the region C-terminal to the phosphotyrosine are characterized by a relatively small contact surface area, compared with the recognition peptide and typical protein–protein interactions.[1,2,6] It has

*Correspondence to: Gennady M. Verkhivker, Agouron Pharmaceuticals, Inc., A Pfizer Company, 10777 Science Center Drive, San Diego, CA 92121-1111. E-mail: verk@agouron.com.

Received 1 February 2001; Accepted 7 August 2001

been suggested that the low-affinity TQpYVPMLE and PQpYQPGEN peptides could undergo significant conformational changes on binding to the SH2 domain, leading to variations in local binding kinetics of the peptide residues.[6] Flexibility of the hydrophobic residues C-terminal to the pTyr group has been also observed in a comparative structural and thermodynamic analysis of binding for the specific EPOpYEEIPIYL peptide and the regulatory ATEPQpYQPGEN peptide with the Fyn member of Src SH2 domains.[19] The regulatory peptide inserts in a well-defined conformation only in the pY binding pocket, whereas the C-terminal region fluctuates between a diverse range of low-energy bound conformations. The favorable entropy of binding and a multistep kinetic mechanism of complex formation have been detected for the regulatory peptide, whereas the kinetics of the specific peptide binding has revealed a simple exchange between the unbound and the complexed states.[19]

Nuclear magnetic resonance (NMR) relaxation studies of both free and complexed states of the SH2 domains with different peptides have analyzed dynamics of the peptide–protein interfaces and entropy effects in thermodynamics and kinetics of binding. These experiments have dissected the role of rigidity and flexibility in the mechanism of sequence-specific recognition.[20–26] It has been shown that dynamics of the peptide–SH2 domain interface in the region C-terminal to the phosphotyrosine plays an important role in modulating binding affinity and regulating rapid exchange of phosphotyrosine sequences.[20–26] SH2 domains can exhibit high affinity to the tyrosine-phophorylated peptides as a result of high on-rate of the peptide–protein association. It is important that peptide–SH2 domain complexes may also demonstrate rapid off-rates of dissociation, so that binding of proteins to phosphotyrosine sequences can rapidly exchange.[27] Binding affinity is greatly reduced for unphosphorylated peptides, allowing for rapid regulation by phosphorylation and dephosphorylation in the phosphotyrosine site.

However, nonlinear binding kinetics of the tandem-SH2 domain of the Syk kinase complexed with a dually phosphorylated peptide has been detected, where a single conformer of the SH2 domain is dominant only at low temperatures, and one or more additional conformers become significantly populated as temperature increases.[10,28] Furthermore, binding kinetics experiments of the peptide–SH2 domain complex formation, monitored by NMR spectroscopy, have also shown that the lifetime of the peptide–SH2 domain interactions may vary by almost two order of magnitude for different protein residues due to different dynamics even after reaching the native bound conformation.[20–31] Although modifications in the peptide composition do not typically alter the dominant topology of the extended peptide conformation, changes in the nature of the peptide can be reflected in altered local binding kinetics at different regions of the peptide–SH2 domain interface.[20–31]

By probing dynamics of some peptide–SH2 domain complexes, NMR relaxation experiments have shown that restriction of motion at the interfacial region may play a role in determining binding energy and molecular basis for peptide specific binding.[22,23] In particular, a high-affinity binding of the C-terminal SH2 domain of the phospholipase (PLCC SH2) to the 12-residue pY-containing peptide displays restriction of motion in the pY binding region, whereas the hydrophobic binding site interacting with the C-terminal residues of the recognition peptides demonstrates a significant ps–ns motional disorder.[21,26] The protein flexibility provides a framework for understanding a relaxed specificity from +2 up to +6 positions despite extensive peptide–protein contacts in this region. However, despite multiplicity of the residues that are permissible at these positions, the PLCC SH2 domain has shown a certain preference for hydrophobic side-chains at +1, +2, and +3 positions.

In contrast, NMR relaxation studies have determined that the hydrophobic binding surface is rather rigid in the amino-terminal Syp SH2 domain–peptide complexes.[22] A low mobility of the Syp SH2 domain and restriction of ps–ns dynamics contributes to an increased stabilization of the hydrophobic interactions with the residues C-terminal to the pTyr group, compensating for the observed decrease in the strength of the electrostatic interactions in the phosphotyrosine pocket.[22] The extensive hydrophobic interactions in the crystal structures of the Syp SH2 domain complexes, coupled with the observed rigidity of the protein residues at the peptide–protein interface, could not explain a considerable promiscuity to hydrophobic substitutions at the positions immediately C-terminal to the pTyr group.[22,32] The crystal structures of the Syp SH2 domain in complexes with the SVLpYTAVQPNE and SPGEpYVNIEF peptides[32] have revealed a similar peptide binding mode. The three positions immediately following the phosphotyrosine are virtually identical in the Src SH2 domain complex with the YEEI peptide compared with either of the Syp SH2 domain complexes. A high degree of structural congruence in the positions of the carboxy-terminal residues coexists with a diversity of binding orientations at the positions preceding the phosphotyrosine, suggesting that the amino-terminal residues may not play a critical role for specificity.[32] There is also a profound conservation between the structures of the uncomplexed and complexed forms of the Syp SH2 domains; therefore, the Syp SH2 domains do not undergo any appreciable conformational changes on peptide binding.[32]

Peptide–SH2 domain recognition even with the rigid protein structure involves a large number of conformational states available to flexible peptides, which necessitates the use of a statistical characterization and the energy landscape analysis, originally introduced in protein folding[33–39] and further developed in ligand–protein binding.[40–51] We previously showed that for robust peptide–protein docking on the complex energy surfaces, the underlying binding energy landscape must have a funnel shape leading to the global free energy minimum.[47] The funnels have been detected near the binding sites of protein–protein complexes, providing a plausible explanation for the observed rapid association rates.[51] Because it is more convenient and computationally feasible to use the folded protein as a reference state in ligand–protein binding, the question may arise as to whether the binding

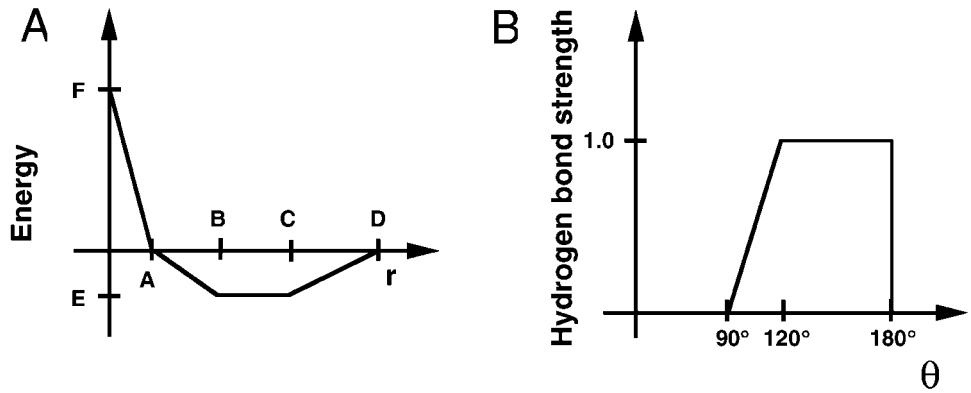

Fig. 1. **A:** The functional form of the ligand-protein interaction energy. For steric interactions, A = 0.93B, C = 1.25B, D = 1.5B, E = −0.4, F = 15.0, and B = $r_l + r_p$ is the sum of the atomic radii for the ligand Tand protein atoms. For hydrogen bond interactions, A = 2.3, B = 2.6, C = 3.1, D = 3.4, E = −4.0, F = 15.0. For sulfur hydrogen bond interactions, A = 2.7, B = 30.0, 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, and B, C, and D are not relevant. 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 is 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.

process can be adequately portrayed by the binding free energy landscapes. Even in the case of a rigid protein bound conformation, the large number of rotational degrees of freedom for a flexible peptide and the complexity of intermolecular interactions would require the determination of the global free energy minimum on a highly complex, multidimensional landscape.

In the present study, we investigate structural and equilibrium aspects of peptide binding with SH2 domains and analyze dynamics of the peptide–SH2 domain interactions. On the basis of the experimental evidence, revealing the rigidity of the Syp SH2 domain interface in both free and the complexed forms, we suggest that simulating dynamics of the peptide equilibrium fluctuations even with the rigid protein structure could be an adequate computational model in the analysis of the peptide–SH2 domain interactions. Predicting structures for the peptide–SH2 domain complexes presents a challenging computational problem as apparent from previous docking studies of the SVLpYTAVQPNE and SPGEpYVNIEF peptides that failed in identifying the native or even near-native conformations.[52] Computational simulations are performed for the pYVPML peptide complex with the v-Src tyrosine kinase SH2 domain, the SVLpYTAVQPNE and SPGEpYVNIEF peptides with the Syp SH2 domain. The effect of the native topological constraints on accuracy of structure prediction for peptide–SH2 domain complexes is studied. We also establish a relationship between the shape of the resulting binding energy landscapes, tolerant to structural perturbations, and the function of specificity for the peptide–SH2 domain complexes.

MATERIALS AND METHODS

Molecular Recognition Energy Model

Computer simulations are conducted by using the simplified energy function[53] in combination with parallel Monte Carlo-simulated tempering approach.[54–59] Simplified, knowledge-based energy function, used in this study, has been adequate in structure prediction of numerous ligand–protein complexes by generating binding energy landscapes with coexisting correlated, funnel-like and uncorrelated, rugged features.[47,53,60–66] The energy function that was used does not have singularities at interatomic distances, effectively explores conformational space, and identifies putative ligand binding modes. More detailed molecular mechanics energy function can produce kinetic traps because of high-energy barriers and rough-energy landscapes susceptible to the precise geometry of the binding modes.[65,66] The advantage of simulated tempering approach is the ability not only to generate an accurate canonical distribution of the ligand–protein system at a wide temperature range but also to search for the global energy minimum.

The knowledge-based simplified energetic model includes intramolecular energy terms for the ligand, given by torsional and nonbonded contributions of the DREIDING force field,[67] and intermolecular ligand–protein steric and hydrogen bond interaction terms calculated from a piecewise linear (PL) potential summed over all protein and ligand heavy atoms [Fig. 1(A)]. The parameters of the PL energy function have been derived to yield the experimental crystallographic structure of a database of ligand–protein complexes, primarily HIV-1 protease and FKBP-12 protein complexes as the global energy minimum.[47,53] 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, flourine-sized nonpolar, sulfur-sized nonpolar and large nonpolar. The atomic radius for carbon, oxygen, nitrogen atoms is 1.8 Å, for flourine is 1.8 Å, and for sulfur and large nonpolar is 2.2 Å. A multiplicative desolvation penalty of 1.0 is applied to the attractive portion of the interaction between nonpolar and polar atoms. Primary and secondary amines are defined to be donors, whereas oxygen and nitrogen atoms with no bound hydrogens are defined to be acceptors. Sulfur is modeled as

being capable of making long-range, weak hydrogen bonds which allows for sulfur donor closer contacts that are seen in some of the crystal structures. Chlorine and phosphorus are modeled as large nonpolar atom types. 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. The steric and hydrogen bondlike potentials have the same functional form, with an additional three-body contribution to the hydrogen bond term. The hydrogen bond interaction energy is 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)]. θ is defined to be 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. The long-range component of the repulsive term used for donor–donor, acceptor–acceptor, and donor–metal close contacts is scaled according to the relative positioning of the two atoms. The scaling is equivalent to that used for hydrogen bonding, that is, the penalty is greatest when the angle θ is 180°, fading to zero at 90° and below.

Monte Carlo Simulations of Ligand–Protein Interactions

In the energy landscape approach, temperature-dependent free energy profiles are generated as a function of the appropriately chosen order parameter from Monte Carlo simulations with the aid of the optimized data analysis and the weighted histogram analysis technique.[68–73] These energy landscapes represent free energies averaged over the degrees of freedom that are not explicitly included in the structure of the landscape, such as solvent degrees of freedom.

In simulations of ligand–protein interactions, the protein is held fixed in its bound conformation, whereas rigid body degrees of freedom and rotatable angles of the ligand are treated as independent variables. Ligand 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 ligand bond lengths, bond angles, and the torsional angles of the unrotated bonds were obtained from the crystal structures of the bound ligand–protein complexes. Crystallographic buried water molecules have been excluded from the simulations to avoid any bias toward the native structure in the course of simulations. Because there are no conformational changes on peptide binding and the structure of the complexed and uncomplexed forms of the SH2 domains are similar, computational structure prediction of the peptide bound conformation is carried out with the protein bound conformation taken from the crystal structure of the corresponding peptide–protein complex.

Monte Carlo simulations allow to dynamically optimize the step sizes at each temperature by taking into account the inhomogeneity of the molecular system.[68] The maximum step sizes are updated by using the acceptance ratio method every cycle of 1000 sweeps and stored both the energy and the coordinates of the system at the end of each cycle. For all these simulations, we equilibrated the system for 1000 cycles (or 1 million sweeps) and collected data during 10,000 cycles (or 10 million sweeps), resulting in 10,000 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 one degree of freedom 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 degree of freedom j is $\langle P_j \rangle^i$, the step sizes σ_j^i for each degree of freedom are updated for cycle $i + 1$ according to the formula

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

where $\langle P_{\text{ideal}} \rangle$ 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 $\langle P_j \rangle^i = 0$ and down by the same constant for $\langle P_j \rangle^i = 1$. Solving the equations

$$s^{-1} = \frac{\ln[a\langle P_{\text{ideal}}\rangle + b]}{\ln[b]} \tag{2}$$

$$s = \frac{\ln[a\langle P_{\text{ideal}}\rangle + b]}{\ln[a + b]} \tag{3}$$

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

Equilibrium simulations have been carried out by using parallel simulated tempering dynamics with 50 replicas of the ligand–protein system attributed respectively to 50 different temperature levels that are uniformly distributed in the range between 5300 and 300 K. Independent 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 m-th and n-th 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

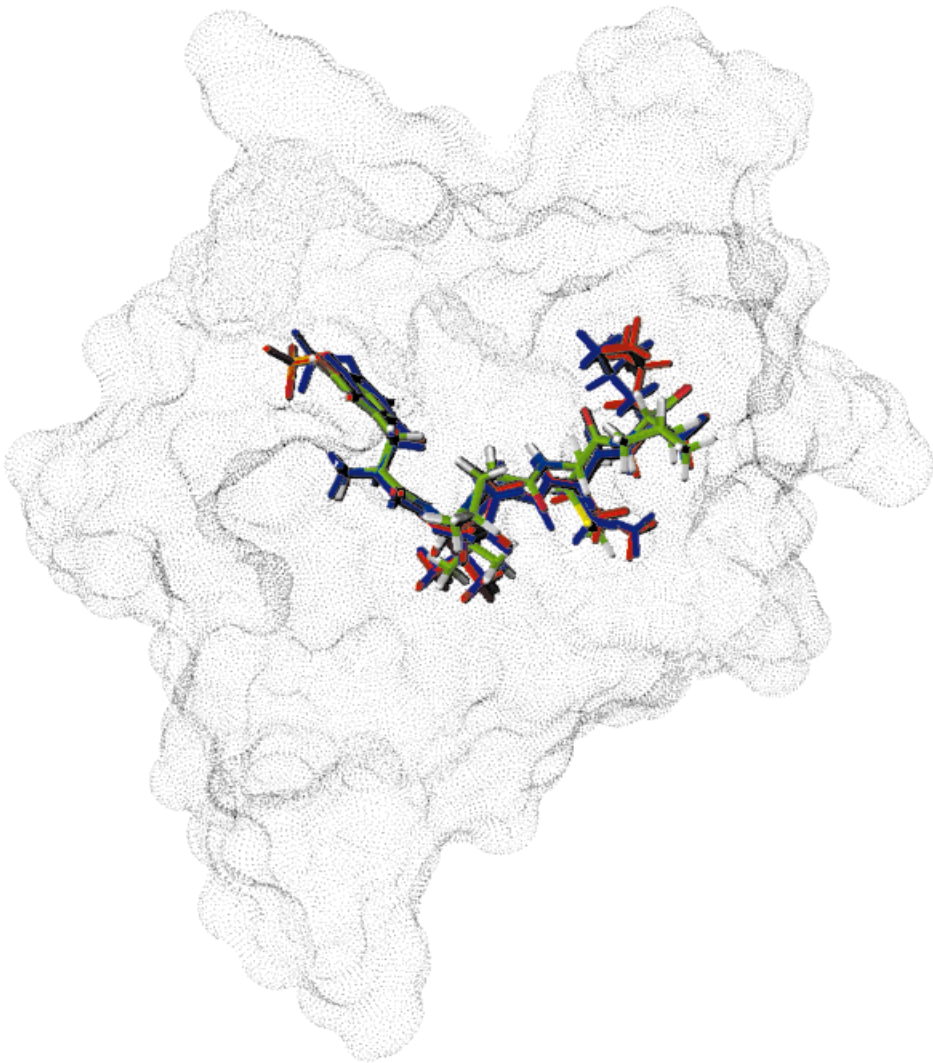

Fig. 2. Superposition of the crystal structure complex of the v-Src tyrosine kinase SH2 domain bound with the pYVPML peptide (color coded by atom type) with the lowest energy structure determined from equilibrium simulations of the flexible peptide (blue) and the peptide with the constrained phosphorus atoms (red). Connoly surface of the v-Src tyrosine kinase SH2 domain protein in the complex with pYVPML is shown.

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.[54–59]

Monte Carlo Data Analysis With the Weighted Histogram Method

The energy landscape approach evaluates equilibrium thermodynamic properties of the system from Monte Carlo simulations of the system at a broad temperature range with the aid of the optimized data analysis and the weighted histogram analysis technique.[69–73] The multiple histogram method[70,71] optimally combines simulation data obtained at many discrete temperatures to provide an improved estimate of the density of states, which can then be used over a range of continuous temperatures. A generalization of the multiple histogram method, the weighted histogram analysis method (WHAM), estimates the density of states from data collected by using umbrella sampling.[72,73]

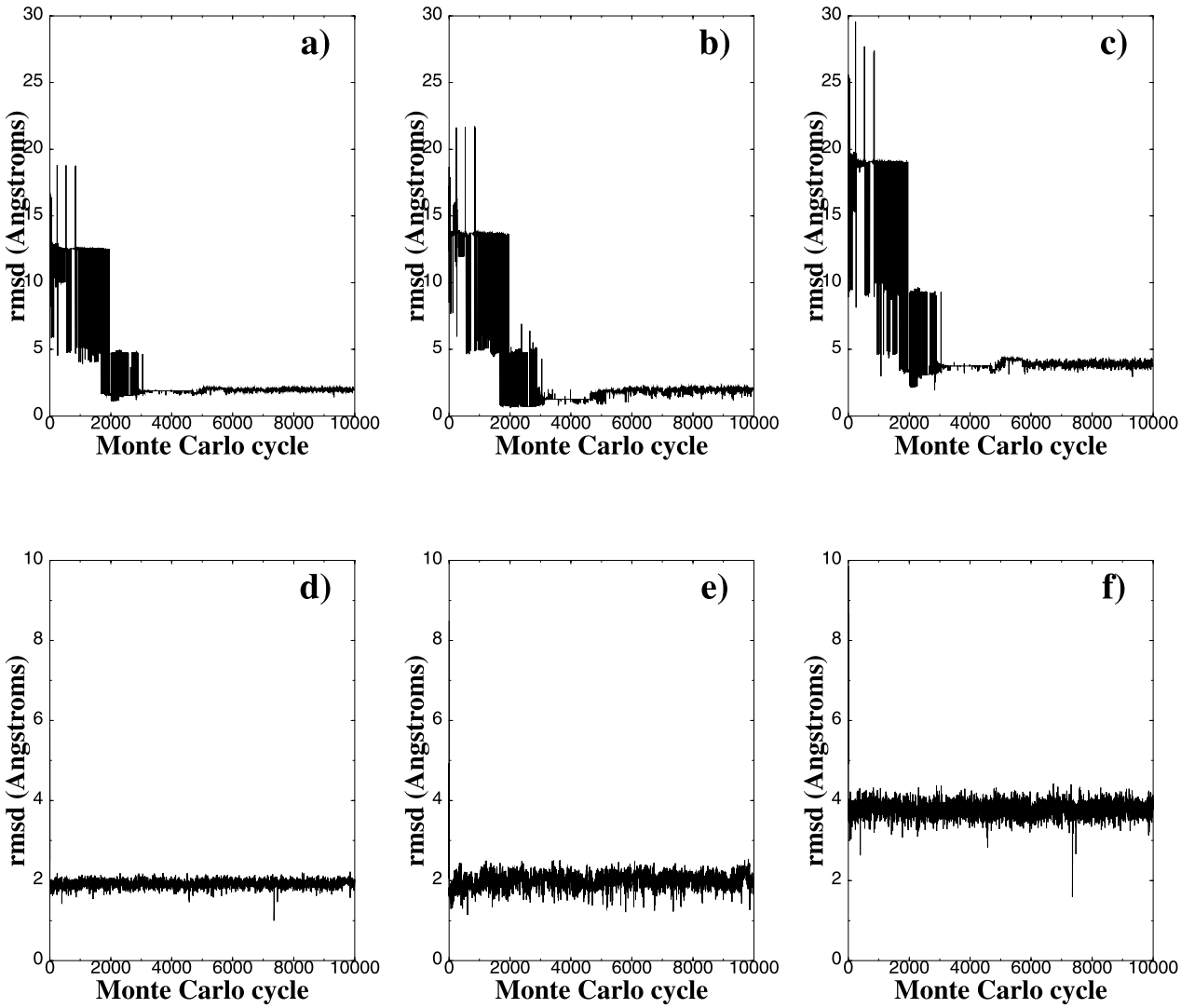

Fig. 3. Time-dependent equilibrium history of the flexible pYVPML peptide binding at T = 300 K: complete peptide (**a**), Met residue (**b**), Leu reside (c). Time-dependent equilibrium history of the pYVPML peptide binding with the constrained phosphorus atom at T = 300 K: complete peptide (**d**), Met residue (**e**), and Leu reside (**f**).

In this work, the weighted histogram analysis method is applied to compute ligand–protein binding energy landscapes, $F(R, T)$, as a continuous function of temperature and reaction coordinate. They are determined by first tabulating two-dimensional histograms $H_i(E, R)$ from the various constant-temperature equilibrium simulations i, and then solving the self-consistent multiple histogram equations[70] to yield the density of states

$$W(E, R) = \frac{\sum_{i=1}^{M} g_i^{-1} H_i(E, R)}{\sum_{j=1}^{M} g_j^{-1} n_j \exp[-(E - F_j)/k_B T_j]} \quad (4)$$

where

$$\exp[-F_j/k_B T_j] = \sum_E W(E) \exp[-E/k_B T_j] \quad (5)$$

and

$$W(E) = \sum_R W(E, R)$$

g_j depends on the correlation time τ_j as $g_j = 1 + 2\tau_j$ and n_j is the number of samples at the temperature T_j.

Although these equations are expressions for the density of states as a function of both energy and reaction coordinate, the free energies are identical to those obtained from the standard one-dimensional multiple histogram equation.

$$W(E) = \sum_R W(E, R) = \frac{\sum_{i=1}^{M} g_i^{-1} H_i(E)}{\sum_{j=1}^{M} g_j^{-1} n_j \exp[-(E - F_j)/k_B T_j]} \quad (6)$$

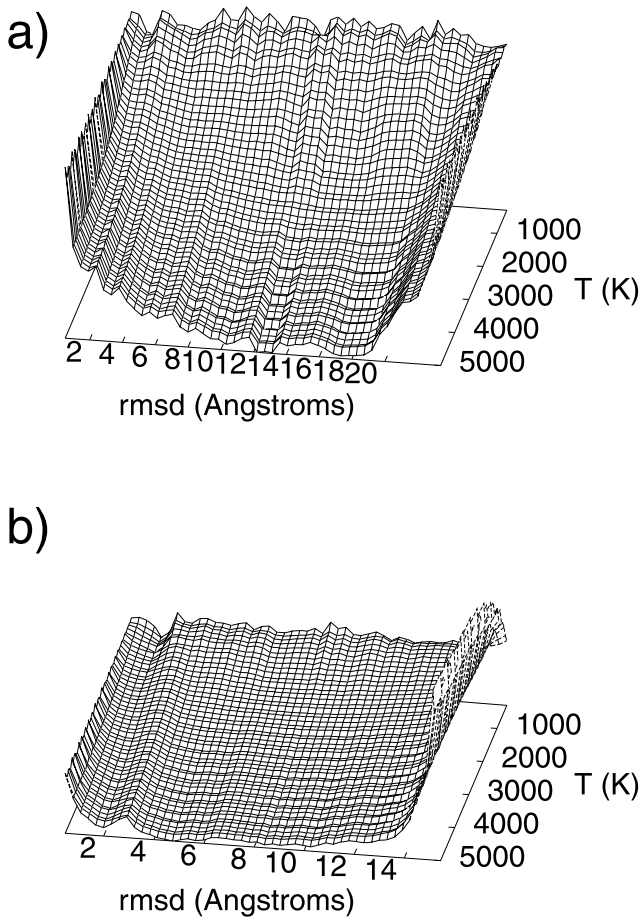

Fig. 4. The binding free energy landscape of the flexible pYVPML peptide (**a**) and the pYVPML peptide with the constrained phosphorus atom (**b**) For each two-dimensional temperature slice, the reference energy $F(R = 0, T)$ is defined to be zero.

where

$$H_i(E) = \sum_R H_i(E, R) \tag{7}$$

and $H_i(E)$ is the standard one-dimensional histogram as a function of energy. These equations are precisely the self-consistent equations for the free energies in the one-dimensional multiple histogram equations. Hence, the one-dimensional equations can be used to determine the free energies F_j and then to compute the multidimensional density of states $W(E, R)$. In this way, calculating the multidimensional density of states as a function of E and R requires no additional computational effort beyond tabulating the simulation data as a function of reaction coordinate as well as energy; the only difficulty is that more sampling is required to ensure adequate statistics.

The potential of mean force $F(R, T)$ at arbitrary temperature relative to a reference position R_c can be computed from the probability density $P(R, T)$ as

$$F(R, T) = -k_B T \ln[P(R, T)/P(R_c, T)] \tag{8}$$

where

$$P(R, T) = \sum_E P_T(E, R) \tag{9}$$

$$P_T(E, R) = W(E, R)\exp[-E/k_B T] \tag{10}$$

Temperature-dependent free energy profiles are generated as a function of the appropriately chosen order parameter. A suitable order parameter must differentiate between ensembles of the native-like structures, non-native conformations and unbound conformations, providing a small dispersion in the energies of the states with a similar value of the order parameter. We define the order parameter R to be the root-mean-square deviation (RMSD) of the ligand coordinates from the native state. The chosen order parameter measures the structural overlap of a given ligand conformation with the ligand conformation in the native structure of the complex. All structures within $R_{\text{native}} = 1.5$ Å from the native conformation constitute the native state. The native state is chosen to be the reference state, so $R_c = 0.0$.

RESULTS AND DISCUSSION

We begin with the computational structure prediction performed for the pYVPML peptide complex with the v-Src tyrosine kinase SH2 domain. The lowest energy structure determined from the equilibrium simulations of the pYVPML peptide binding belongs to the native binding domain with RMSD = 2.0 Å from the native structure (Fig. 2). The phosphotyrosine and Met (pY + 3) peptide residues are deeply buried in the protein surface, whereas the peripheral Leu residue is exposed to solvent. This residue has higher temperature B-factors in the crystal structure, reflecting the inherent flexibility in this region of the peptide–receptor interface (Fig. 2). Equilibrium simulations of the flexible pYVPML peptide converge to the thermodynamically stable native conformation after approximately 2000 simulations cycles [Fig. 3(a)]. Although the range of equilibrium thermal fluctuations for the pYVPML peptide and its Met residue is similar and is confined in the native state [Fig. 3(a,b)], a solvent exposed Leu residue exhibits much larger fluctuations up to RMSD = 4 Å from its crystal coordinates [Fig. 3(c)]. Simulations with the constrained phosphorus atom in its native location in the phosphotyrosine pocket result in a rapid acquisition of the native peptide structure [Fig. 3(d)]. Although Val, Pro, and Met rapidly attain their native structures and fluctuate in the close proximity of the native conformations throughout equilibrium [Fig. 3(e)], Leu continues to display much larger fluctuations from the native state [Fig. 3(f)].

The binding energy landscape of a flexible peptide is characterized by a funnel of conformations leading to a broad, entropically favorable native basin via low-energy barriers [Fig. 4(a)]. Although the dominant native topology of the bound peptide is stabilized at higher temperatures, the exact crystallographic conformation of the bound pYVPML peptide becomes thermodynamically favorable only at lower temperatures, competing with a broader local minimum at RMSD = 4.0 Å from the crystal structure. The latter local minimum corresponds to the native binding topology with the fluctuating Leu residue. A

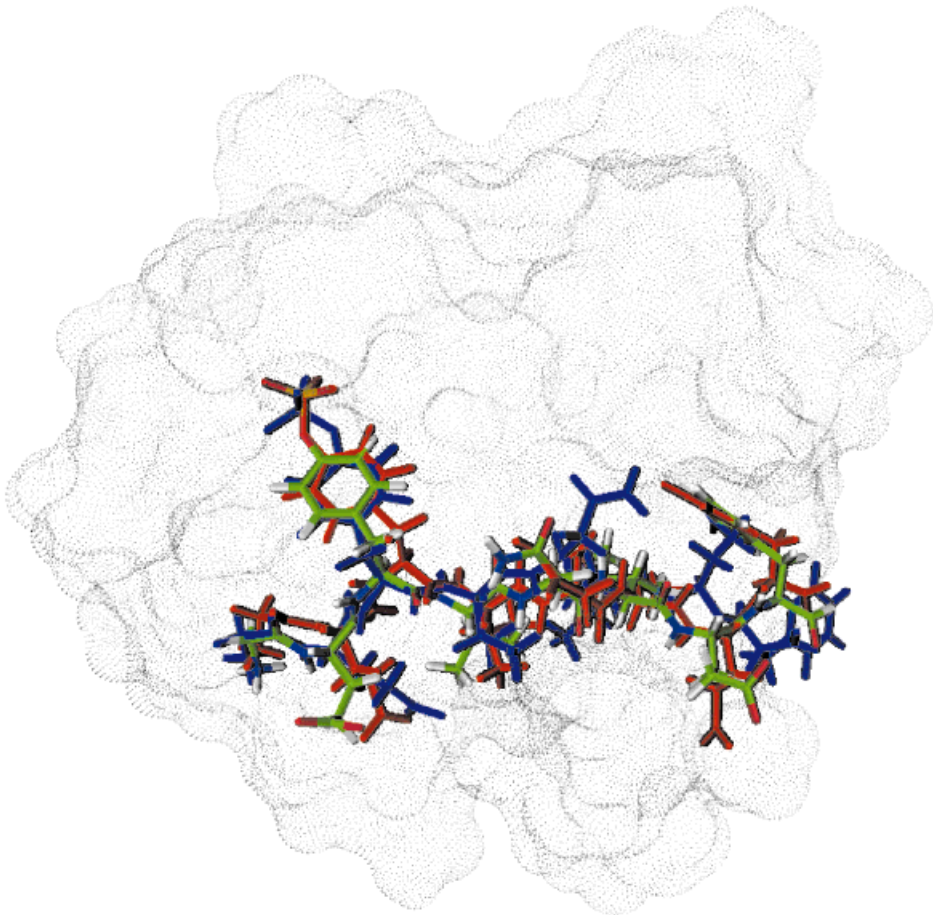

Fig. 5. Superposition of the crystal structure complex of the Syp SH2 domain bound with the GEpYVNIEF portion of the SPGEpYVNIEF peptide (color coded by atom type) with the lowest energy structure determined from equilibrium simulations of the flexible peptide (blue) and the peptide with the constrained phosphorus atoms (red). Connoly surface of the Syp SH2 domain bound with SPGEpYVNIEF is shown.

considerable flexibility of the Leu residue in the course of equilibrium simulations presents only a minor source of frustration because of the low-energy barriers separating the alternative conformations. A well-defined energy barrier between these minima emerges only at lower temperature. The pYVPML peptide undergoes considerable conformational changes on binding during equilibrium simulations even at moderate temperatures, in agreement with the experimental thermodynamic data.[6] A slower convergence to the crystallographic conformation at lower temperatures may lead to variations in local binding kinetics in the flexible region of the peptide as temperature decreases.[6] The results suggest that the lifetime of the native peptide–SH2 domain interactions, observed in the crystal structure of the pYVPML peptide complex, may differ considerably in distinct regions of the peptide–SH2 domain interface. The equilibrium fluctuations of the flexible pYVPML peptide at the bottom of the dominant binding funnel do not change the topology of the native structure but may lead to different dynamics of the peptide–SH2 domain interactions even after reaching the native structure. The binding free energy profile generated in simulations with the constrained phosphorus atom is characterized by a unique and well-defined native binding domain that is thermodynamically favorable almost at the entire temperature range [Fig. 4(b)]. The formation of the favorable interactions in the phosphotyrosine pocket leads to a funnel-like shape of the resulting binding energy hypersurface and rapid descent to the thermodynamically stable native structure without overcoming any major free energy barrier.

Computer simulations of the SPGEpYVNIEF peptide binding with the Syp SH2 domain have been carried out by using a flexible GEpYVNIEF portion of the peptide because no crystallographic coordinates were present for Ser and Pro in the crystal structure.[32] Because the structures of the complexed and uncomplexed forms of the Syp SH2 domains are essentially identical,[32] the bound form of the protein was originally used in simulations. The lowest energy structure obtained from the equilibrium simulations reproduces a general topology of the native peptide complexes with the Syp SH2 domains but differs by RMSD = 3.5 Å from the crystallographic conformation. Nevertheless, a number of salient features of the native topology have been accurately determined in the predicted structure. The orientation of the C_α—C_β bond of the peptide side-chain at +1 position is parallel to the protein surface, so that the hydrophobic Val side-chain is only

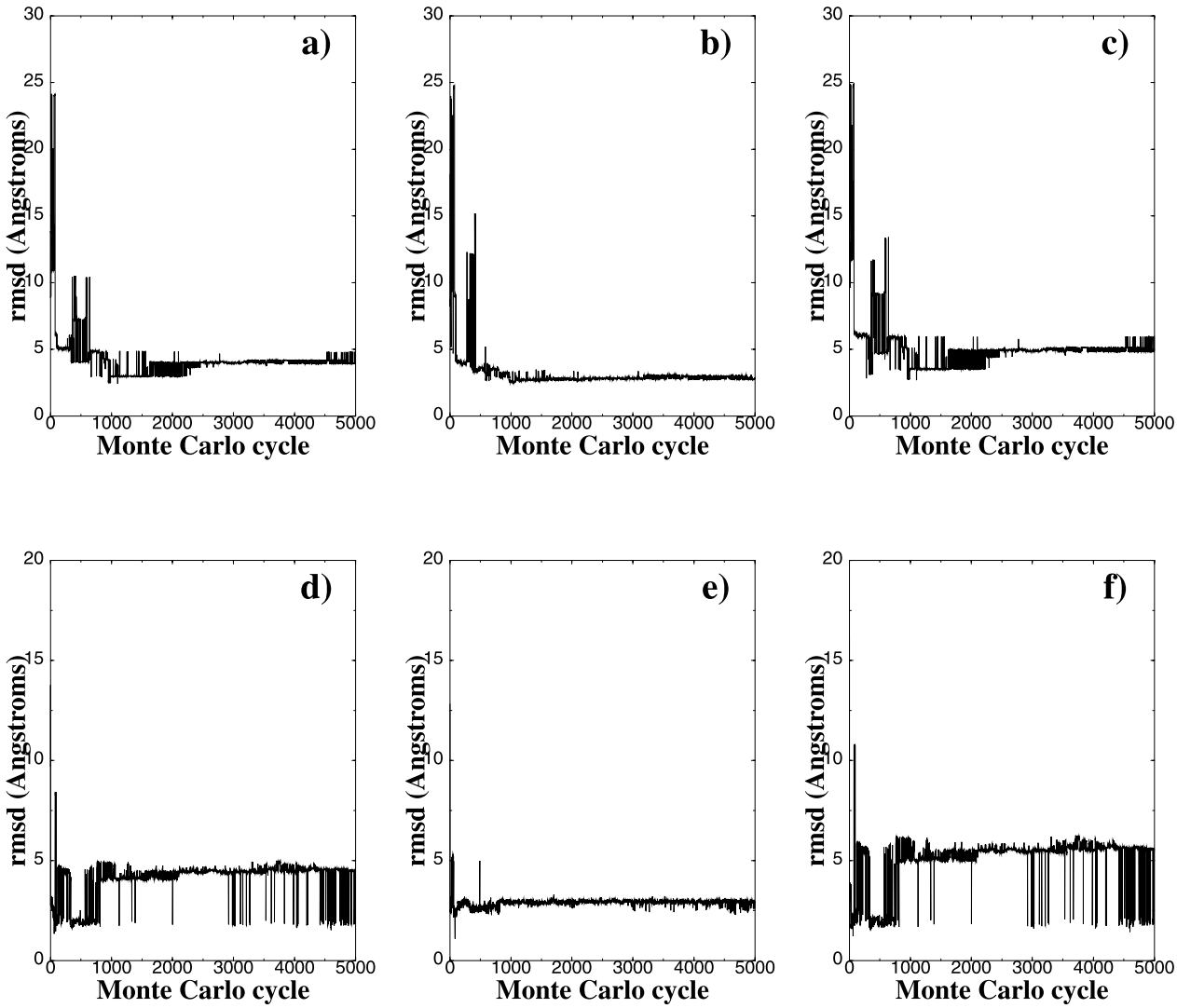

Fig. 6. Time-dependent sampling history of the flexible GEpYVNIEF peptide binding at T = 300 K: complete peptide (**a**), GE motif (**b**), and VNIEF motif (**c**). Time-dependent sampling history of the flexible GEpYVNIEF peptide binding with the constrained phosphorus atom at T = 300 K: complete peptide (**d**), GE motif (**e**), and VNIEF motif (**f**).

partially buried in a pocket. In the computationally predicted structure, the side-chain of Ile at position +3 is buried into the surface of the SH2 domain, making extensive interactions with the protein, in agreement with the crystallographic conformation (Fig. 5). Equilibrium simulations of the flexible GEpYVNIEF peptide converge to the native-like topology of the bound conformation after approximately 1000 Monte Carlo cycles [Fig. 6(a)]. Thermal fluctuations of the bound peptide conformation at RMSD = 4 Å from the crystal structure is due to a combined effect of both the GE motif, preceding the pTyr recognition residue [Fig. 6(b)], and the VNIEF portion of the peptide [Fig. 6(c)]. Constraining a single phosphorus atom leads to a rapid collapse to the dominant topology of the bound peptide after only three simulation cycles [Fig. 6(d)]. However, frequent transitions have been observed between the near-native states, located at RMSD = 4 Å from the crystal structure and the native conformations with RMSD = 2 Å from the crystal structure [Fig. 6(d,f)]. The binding free-energy profile of the flexible and partially constrained GEpYVNIEF peptide [Fig. 7(a,b)] are fairly shallow. Binding domains at RMSD = 2 Å and RMSD = 4 Å from the native structure dominate thermodynamic equilibrium at a broad temperature range [Fig. 7(b)]. Only at low temperatures the native binding mode with RMSD = 2 Å from the crystal structure becomes more favorable.

The coexistence of two favorable binding modes within the same dominant native topology at a broad temperature range is due to considerable equilibrium fluctuations of the hydrophobic residues C-terminal to the pTyr group. This suggests that peptide dynamics may play a role in moderating the extensive static interactions observed in the crystal structure of the SPGEpYVNIEF complex, which would otherwise prohibit accommodation of the hydrophobic substitutions in the peptide specificity region. The variations in structural stability and flexibility during equilibrium for the peptide–SH2 domain complexes are associated with the ruggedness at the bottom of the

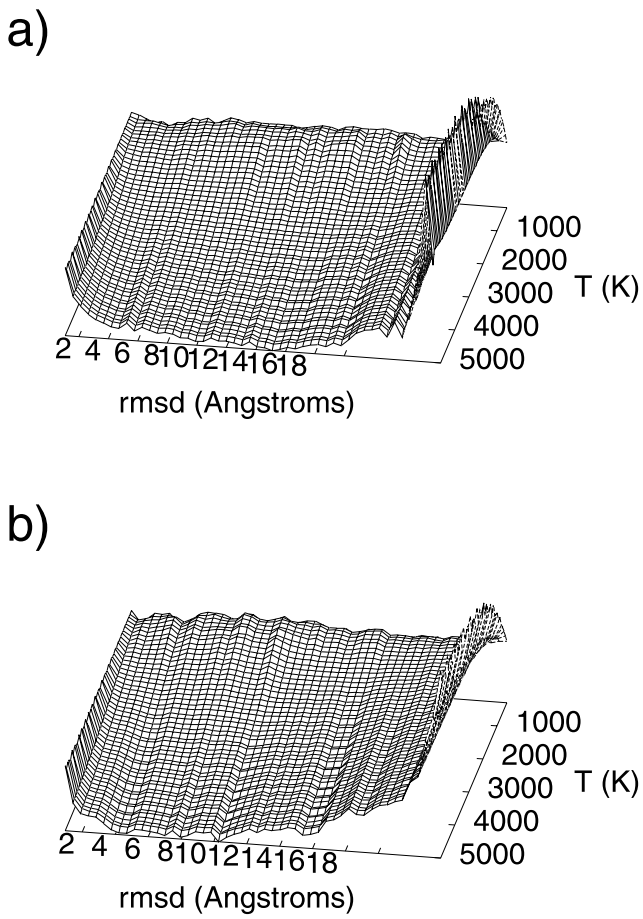

Fig. 7. The binding free energy landscape of the flexible GEpYVNIEF peptide complex (**a**) and the GEpYVNIEF peptide with the constrained phosphorus atom (**b**). For each two-dimensional temperature slice, the reference energy $F(R = 0, T)$ is defined to be zero.

underlying energy landscape. The differences in the shape of the binding energy landscape can be related to various functions such as specificity or permissiveness in peptide–protein recognition.[45,46,49,50] Highly specific peptide–protein complexes are likely to be relatively rigid, with a steep funnel of conformations leading to the native structure. 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. The binding energy landscape for the GEpYVNIEF–Syp SH2 domain complex has a broad, entropically favorable topology with low-energy barriers between coexisting local minima. The shape of the binding funnel for this peptide–SH2 domain complex underscores the robustness of the native topology to structural perturbations. This could provide an additional positive mechanism underlying tolerance of the SH2 domains to hydrophobic mutations in the peptide specificity region.

Computer simulations of the SVLpYTAVQP variant of the original SVLpYTAVQPNE peptide interacting with the Syp SH2 domain have been performed because the coordinates of two peripheral residues were not present in the crystal structure.[32] The lowest energy structure determined from equilibrium simulations of the flexible SVLpYTAVQP peptide has the native topology of the crystal structure, although it is located at RMSD = 4 Å from the crystal structure (Fig. 8). The binding mode at three crucial positions immediately following the phosphotyrosine is virtually identical to the crystal structure, with the TAVQP portion of the peptide deviating only by RMSD = 2.5 Å from the native conformation (Fig. 8). The observed difference between the predicted and crystal structures can be attributed to flexibility of the amino-termimal Ser, Val, and Leu residues, which may represent a more frustrated region of the peptide–SH2 domain interface and, therefore, could be less important for specific recognition of the peptide (Fig. 8).

By constraining a C_α position for each residue of the SVLpYTAVQP peptide, the complexity of the energy landscape can be somewhat reduced, and its effect on the accuracy of structure prediction for the peptide complexes can be analyzed. Equilibrium simulations with the constrained phosphorus atom have revealed a rapid collapse to the native topology of the bound conformation. The lowest energy conformation is located at RMSD = 4.5 Å from the crystal structure [Fig. 9(a)], but the TAVQP motif of the peptide fluctuates at the native-like conformations at RMSD = 2.0 Å from the crystal structure [Fig. 9(c)]. When C_α positions of the flexible, amino-terminal Ser and Val residues of the peptide are fixed in the correct orientations, the remainder of the peptide instantly achieves the crystallographic conformation and maintains the thermodynamically stable state with RMSD = 1.5 Å from the crystal structure [Fig. 9(d,e)]. However, when the C_α position of the Leu residue is constrained, the thermodynamically stable conformations fluctuate within RMSD = 3 Å from the native structure [Fig. 9(f)]. When the C_α atoms of Thr, Val, or Gln carboxy-terminal peptide residues are fixed in the native locations, the lowest energy structure with the native-like topology of the bound conformation resides at RMSD = 4 Å from the crystal structure [Fig. 9(g–i)]. Although the native topology is formed early and is steadily maintained throughout equilibrium simulations, flexible amino-terminal residues continue to fluctuate in structurally alternative conformations, suggesting that these peptide–protein interactions are more frustrated than the interactions with the recognition residues.

The binding free-energy profiles generated from simulations of a flexible SVLpYTAVQP peptide [Fig. 10(a)] are characterized by a number of competing binding modes at low temperatures. However, the binding free energy profiles obtained from simulations with the fixed phosphorous atom [Fig. 10(b)] and the constrained C_α atom of the Ser residue [Fig. 10(c)] reveal a funnel-like shape of the binding energy landscape leading to the native structure at high temperatures with no energy barriers. We find that no matter which peptide residue of SVLpYTAVQPNE is constrained in the native C_α position, the native topology of the bound peptide is still formed rapidly and consistently during equilibrium simulations, despite a large number of degrees of freedom (37 flexible rotatable bonds) that are still available to the system. The convergence to the native structure independently of the nature of the constrained peptide–protein contacts even at high tempera-

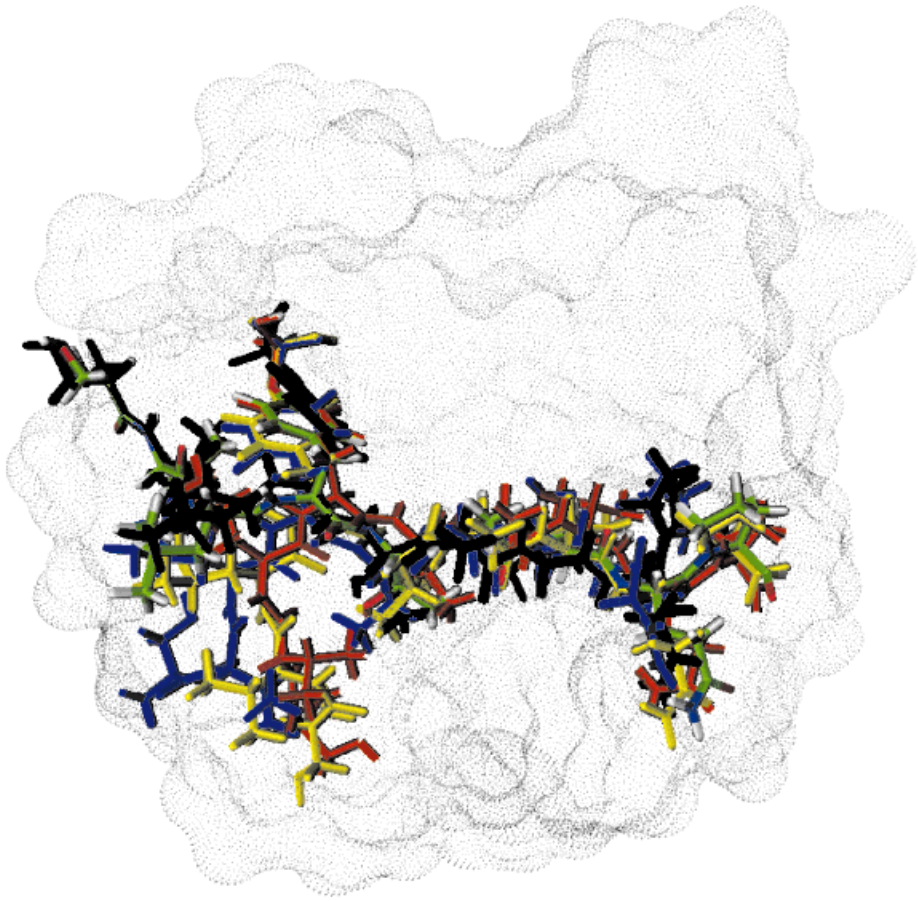

Fig. 8. Superposition of the crystal structure complex of the Syp SH2 domain bound with the SVLpYTAVQP variant of the SVLpYTAVQPNE peptide (color coded by atom type) with lowest energy structure determined from equilibrium simulations of the flexible peptide (blue), of the peptide with the constrained phosphorus atoms (red), of the peptide with the fixed C_α atom of Val residue in the carboxy-terminal (gold), of the peptide with the fixed C_α atom of Ser residue (black). Connoly surface of the Syp SH2 domain bound with SVLpYTAVQPNE is shown.

tures is due to a funnel-like shape of the underlying energy landscape with the dominant free energy minimum, robust to structural perturbations.

It may be energetically advantageous for the peptide–SH2 domain complex to form the native contacts in the most thermodynamically stable region early in the binding reaction, because a moderate loss of conformational entropy, compensated by strong, favorable native interactions in the pTyr pocket, may lower the free energy barrier. The importance of the native topology has been originally recognized in protein folding, where local native interactions and simple topological motifs may facilitate a rapid transition to the native structure.[74–78] However, rigorous kinetic simulations and determination of the structure of the transition state ensemble are necessary to provide a detailed picture of the mechanism of the peptide–SH2 domain association. This analysis extends beyond the scope of the present study and will be discussed elsewhere.

CONCLUSIONS

Structural and equilibrium aspects of the SH2 domains binding to a set of phosphopeptides have been studied with the energy landscape approach. Computer simulations using the simplified, knowledge–based energy function and simulated tempering dynamics have accurately determined the native structure of the SH2 domain complexes with large and highly flexible pYVPML, SVLpYTAVQPNE, and SPGEpYVNIEF peptides. The binding energy landscape approach is used to determine not only the lowest energy structures of the peptide complexes but also characterize structural stability and dynamics at the peptide–SH2 domain interface. The topological features of the peptide–SH2 domain interface are primarily determined by the thermodynamically stable phosphotyrosyl group. A diversity of structurally different binding orientations has been observed for the amino-terminal residues to the phosphotyrosine, in agreement with experimental data. The dominant native topology for the peptide residues carboxy-terminal to the phosphotyrosine is tolerant to flexibility in this region of the peptide–SH2 domain interface observed in equilibrium simulations. The results suggest that a broad, entropically favorable topology of the native binding mode, robust to moderate structural perturbations may be an additional mechanism underlying permissiveness of the SH2 domain to the peptide sequence variations in the specificity region.

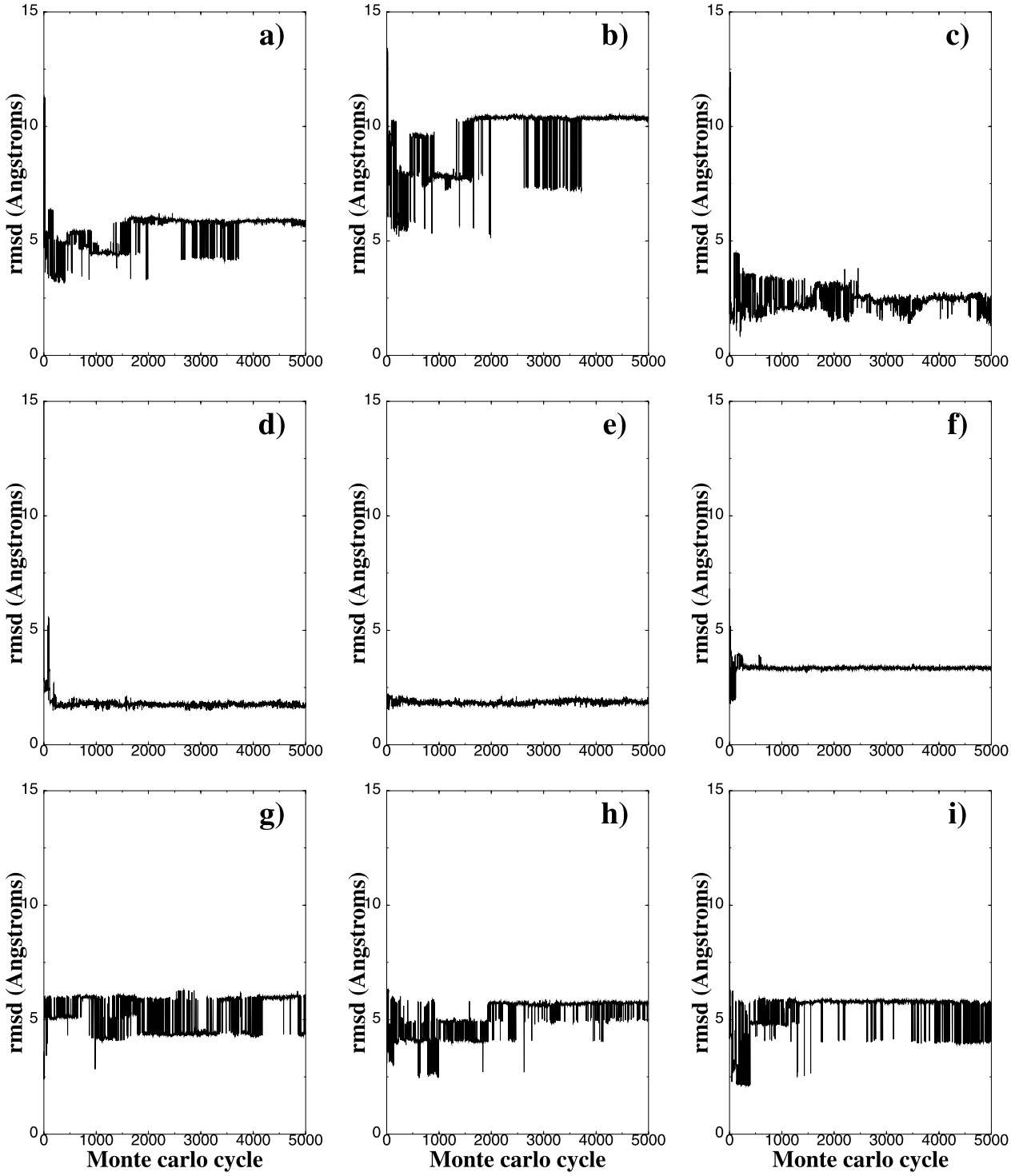

Fig. 9. Time-dependent sampling history of the SVLpYTAVQP peptide binding with the constrained phosphorus atom at $T = 300$ K: complete peptide (**a**), SVL motif (**b**), and TAVQP motif (**c**). Time-dependent sampling history of the SVLpYTAVQP peptide at $T = 300$ K with the fixed C_α atom of Ser residue (**d**), Val residue of amino-terminal (**e**), Leu residue (**f**), Thr residue (**g**), Val residue of carboxy-terminal (**h**), and Gln residue (**i**).

The effect of the native topology on the robustness and accuracy of structure prediction for the peptide–SH2 domain complexes has been studied by simulating the process of the native structure assembly given that some native peptide–SH2 domain contacts are constrained. The underlying binding free energy profile has the funnel-like shape that leads to a progressive acquisition of the native structure, once some native peptide–SH2 domain contacts

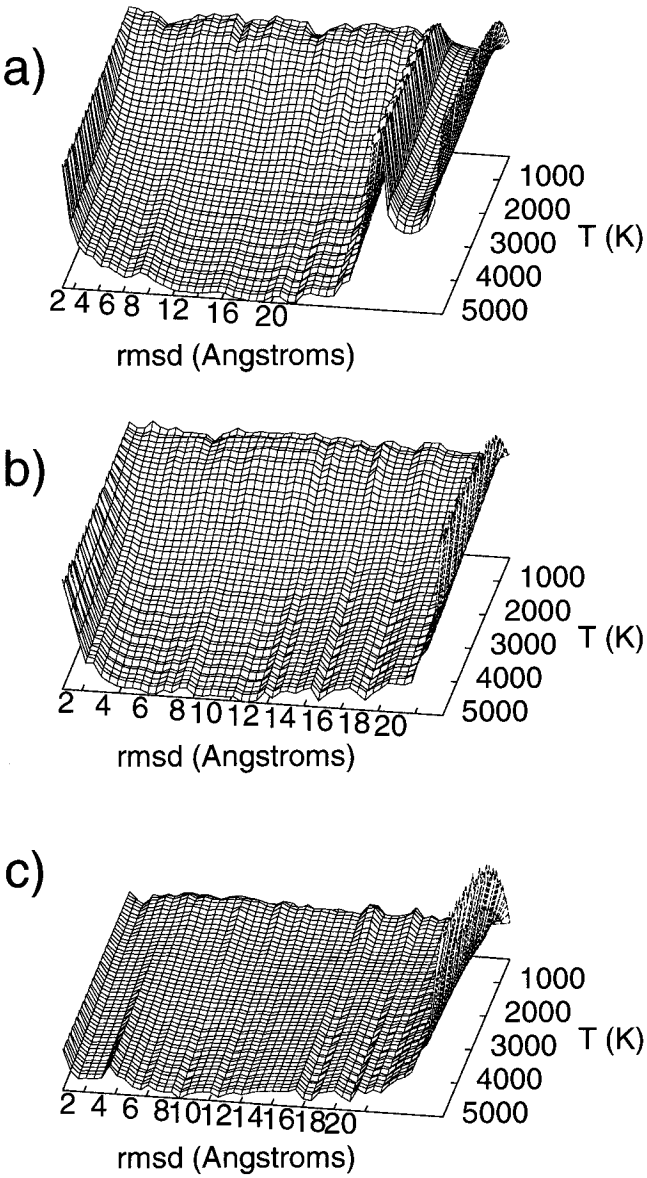

Fig. 10. The binding free energy landscape of the flexible SVLpYTAVQP peptide complex (**a**) the SVLpYTAVQP peptide with the constrained phosphorus atom (**b**), and the SVLpYTAVQP peptide with the fixed C_α atom of Ser residue (**c**). For each two-dimensional temperature slice, the reference energy $F(R = 0, T)$ is defined to be zero.

are fixed in their native positions. Hence, incorporation of a limited knowledge about the peptide–SH2 domain interface can improve the robustness of structure prediction. The results of our study indicate that there may a relationship between the shape of the binding energy landscapes and the function of specificity for the peptide–SH2 domain complexes. The results of this study suggest that a useful physical insight into molecular basis of sequence-specific recognition for the peptide–SH2 domain complexes can be obtained by using the simplified model of peptide–protein interactions and the energy landscape analysis. Nevertheless, rigorous free energy calculations of peptide mutations in the specificity region are necessary to provide atomic details and quantitative assessment of the enthalpy and entropy components in the thermodynamics of the peptide–SH2 domain binding.

REFERENCES

1. Waksman G, Kominos D, Robertson SC, Pant N, Baltimore D, Birge RB, Cowburn D, Hanafusa H, Mayer BJ, Overduin M, Resh MD, Rios CB, Silverman L, Kuriyan J. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosinephosphorylated peptides. Nature 1992;358:646–653.
2. Waksman G, Shoelson SE, Pant N, Cowburn D, Kuriyan J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 Domain: crystal structures of the complexed and peptide-free forms. Cell 1993;72:779–790.
3. Eck M, Shoelson SE, Harrison SC. Recognition of a high affinity phosphotyrosyl peptide by the Src homology 2 domain of $p56^{lck}$. Nature 1993;362:87–91.
4. Overduin M, Rios CB, Mayer BJ, Baltimore D, Cowburn D. Three-dimensional solution structure of the Src homology 2 Domain of c-Abl. Cell 1992;70:697–704.
5. Gilmer T, Rodriguez M, Jordan S, Crosby R, Alligood K, Green M, Kimery M, Wagner C, Kinder D, Charifson P, Hassell AM, Willard D, Luther M, Rusnak D, Sternbach DD, Mehrotra M, Peel M, Shampine L, Davis R, Robbins J, Paterl IR, Kassel D, Burkhart W, Moyer M, Bradshaw T, Berman J. Peptide inhibitors of src SH3-SH2-phosphoprotein interactions. J Biol Chem 1994;269: 31711–31719.
6. Bradshaw JM, Grucza RA, Ladbury JE, Waksman G. Probing the "two-pronged plug two-holed socket" model for the mechanism of binding of the Src SH2 domain to phosphotyrosyl peptides: a thermodynamic study. Biochemistry 1998;37:9083–9090.
7. Bradshaw JM, Waksman G. Calorimetric investigation of proton linkage by monitoring both the enthalpy and association constant of binding: application to the interaction of the Src SH2 domain with a high-affinity tyrosyl phosphopeptide. Biochemistry 1998;37: 15400–15407.
8. Bradshaw JM, Waksman G. Calorimetric examination of high-affinity Src SH2 domain-tyrosyl phosphopeptide binding: dissection of the phosphopeptide sequence specificity and coupling energetics. Biochemistry 1999;38:5147–5154.
9. Bradshaw JM, Mitaxov V, Waksman G. Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. J Mol Biol 1999;293:971–985.
10. Grucza RA, Bradshaw JM, Futterer K, Waksman G. SH2 domains: from structure to energetics, a dual approach to the study of structure-function relationships. Med Res Rev 1999;19:273–293.
11. Bradshaw JM, Mitaxov V, Waksman G. Mutational investigation of the specificity determining region of the Src SH2 domain. J Mol Biol 2000;299:521–535.
12. Charifson PS, Shewchuk LM, Rocque W, Hummel CW, Jordan SR, Mohr C, Pacofsky GJ, Peel MR, Rodriguez M, Sternbach DD, Consler TG. Peptide ligands of pp60(c-src) SH2 domains: a thermodynamic and structural study. Biochemistry 1997;36:6283–6293.
13. Plummer MS, Holland DR, Shahripour A, Lunney EA, Fergus JH, Marks JS, McConnell P, Mueller WT, Sawyer TK. Design, synthesis, and cocrystal structure of a nonpeptide Src SH2 domain ligand. J Med Chem 1997;40:3719–3725.
14. Lunney EA, Para KS, Rubin JR, Humblet C, Fergus JH, Marks JS, Sawyer TK. Structure-based design of a novel series of nonpeptide ligands that bind to the pp60-src SH2 domain. J Am Chem Soc 1997;119:12471–12476.
15. Pacofsky GJ, Lackey K, Alligood KJ, Berman J, Charifson PS, Crosby RM, Dorsey GF Jr, Feldman PL, Gilmer TM, Hummel CW, Jordan SR, Mohr C, Shewchuk LM, Sternbach DD, Rodriguez M. Potent dipeptide inhibitors of the pp60c-src SH2 domain. J Med Chem 1998;41:1894–1908.
16. Shakespeare W, Yang M, Bohacek R, Cerasoli F, Stebbins K, Sundaramoorthi R, Azimioara M, Vu C, Pradeepan S, Metcalf C III, Haraldson C, Merry T, Dalgarno D, Narula S, Hatada M, Lu X, van Schravendijk MR, Adams S, Violette S, Smith J, Guan W, Bartlett C, Herson J, Iuliucci J, Weigele M, Sawyer T. Structure-based design of an osteoclast-selective, nonpeptide src homology 2 inhibitor with in vivo antiresorptive activity. Proc Natl Acad Sci USA 2000;97:9373–9378.
17. Shakespeare WC, Bohacek RS, Azimioara MD, Macek KJ, Luke GP, Dalgarno DC, Hatada MH, Lu X, Violette SM, Bartlett C,

Sawyer TK. Structure-based design of novel bicyclic nonpeptide inhibitors for the src SH2 domain. J Med Chem 2000;43:3815–3819.
18. Bohacek RS, Dalgarno DC, Hatada M, Jacobsen VA, Lynch BA, Macek KJ, Merry T, Metcalf CA III, Narula SS, Sawyer TK, Shakespeare WC, Violette SM, Weigele M. J Med Chem 2001;44:660–663.
19. Ladbury JE, Hensmann M, Panayotou G, Campbell ID. Alternative models of tyrosyl phosphopeptide binding to a Src family SH2 domain: implications for regulation of tyrosine kinase activity. Biochemistry 1996;35:11062–11069.
20. Pascal SM, Yamazaki T, Singer AU, Kay LE, Forman-Kay JD. Structural and dynamic characterization of the phosphotyrosine binding region of an SH2 domain-phosphopeptide complex by NMR relaxation, proton exchange, and chemical shift approaches. Biochemistry 1995;34:11353–11362.
21. Kay LE, Muhandiram DR, Farrow NA, Aubin Y, Forman-Kay JD. Correlation between dynamics and high affinity Binding in an SH2 domain interaction. Biochemistry 1996;35:361–368.
22. Kay LE, Muhandiram DR, Wolf G, Shoelson SE, Forman-Kay JD. Correlation between binding and dynamics at SH2 domain interfaces. Nat Struct Biol 1998;5:156–163.
23. Forman-Kay JD. The 'dynamics' in the thermodynamics of binding. Nat Struct Biol 1999;6:1086–1087.
24. Zhang W, Smithgall TE, Gmeiner WH. Self-association and backbone dynamics of the HCk SH2 domain in the free and phosphopeptide-complexed forms. Biochemistry 1998;37:7119–7126.
25. Engen JR, Gmeiner WH, Smithgall TE, Smith DL. Hydrogen exchange shows peptide binding stabilizes motions in Hck SH2. Biochemistry 1999;38:8926–8935.
26. Feng M-H, Phillippopoulos M, MacKerell AD, Lim C. Structural characterization of the phosphotyrosine binding region of a high-affinity SH2 domain-phosphopeptide complex by molecular dynamics simulation and chemical shift calculations. J Am Chem Soc 1996;118:11265–11277.
27. Felder S, Zhou M, Hu P, Urena J, Ullrich A, Chaudhuri M, White M, Shoelson SE, Schlessinger J. SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. Mol Cell Biol 1993;13:1449–1455.
28. Grucza RA, Futterer K, Chan AC, Waksman G. Thermodynamic study of the binding of the tandem-SH2 domain of the Syk kinase to a dually phosphorylated ITAM peptide: evidence for two conformers. Biochemistry 1999;38:5024–5033.
29. Grucza RA, Bradshaw JM, Mitaxov V, Waksman G. Role of electrostatic interactions in SH2 domain recognition: salt-dependence of tyrosylphosphorylated peptide binding to the tandem SH2 domain of the Syk kinase and the single SH2 domain of the Src kinase. Biochemistry 2000;39:10072–10081.
30. Weber T, Schaffhausen B, Liu Y, Gunther UL. NMR structure of the N-SH2 of the p85 subunit of phosphoinositide 3-kinase complexed to a doubly phosphorylated peptide reveals a second phosphotyrosine binding site. Biochemistry 2000;39:15860–15869.
31. Kristensen SM, Siegal G, Sankar A, Driscoll PC. Backbone dynamics of the C-terminal SH2 domain of the p85alpha subunit of phosphoinositide 3-kinase: effect of phosphotyrosine-peptide binding and characterization of slow conformational exchange processes. J Mol Biol 2000;299:771–788.
32. Lee C-H, Kominos D, Jacques S, Margolis B, Schlessinger J, Shoelson SE, Kuriyan J. Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphotase. Structure 1994;2:423–438.
33. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding, a synthesis. Proteins 1995;21:167–195.
34. 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.
35. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
36. Shakhnovich EI. Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
37. Lazaridis T, Karplus M. "New view" of protein folding reconciles with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
38. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
39. Pande VS, Grosberg AY, Rokshar D, Tanaka T. Pathways for protein folding: is a new view needed? Curr Opin Struct Biol 1998;8:68–79.
40. Kuntz ID, Meng EC, Shoichet BK. Structure-based molecular design. Acc Chem Res 1994;27:117–123.
41. Kollman P. Free energy calculations-applications to chemical and biological phenomena. Chem Rev 1993;93:2395–2417.
42. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand-receptor complexes. J Med Chem 1995;38:4953–4967.
43. Rosenfeld R, Vajda S, DeLisi C. Flexible docking and design. Annu Rev Biophys Biomol Struct 1995;24:677–700.
44. Janin J. Quantifying biological specificity: the statistical mechanics of molecular recognition. Proteins 1996;25:438–445.
45. 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.
46. 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.
47. Verkhivker GM, Rejto PA, Gehlhaar DK, Freer ST. Exploring energy landscapes of molecular recognition by a genetic algorithm, analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes. Proteins 1996;25:342–353.
48. 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.
49. Tsai C-J, Xu D, Nussinov R. Protein folding via binding and vice versa. Curr Biol 1998;3:R71–R80.
50. Tsai C-J, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels and protein function. Protein Sci 1999;8:1181–1190.
51. Zhang C, Chen J, DeLisi C. Protein-protein recognition: exploring the energy funnels near the binding sites. Proteins 1999;34:255–267.
52. Zhou Y, Abagyan R. How and why phosphotyrosine-containing peptides bind to the SH2 and PTB domains. Fold Des 1998;3:513–522.
53. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. Molecular recognition of the inhibitor AG–1343 by HIV–1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 1995;2:317–324.
54. Marinari E, Parisi G. Simulated tempering: a new Monte Carlos scheme. Europhys Lett 1992;19:451–458.
55. Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc (Jap) 1996;65:1604–1607.
56. Hansmann UHE, Okamoto Y. Monte Carlo simulations in generalized ensemble: multicanonical algorithm versus simulated tempering. Phys Rev E 1996;54:5863–5865.
57. Hansmann UHE, Okamoto Y. Generalized-ensemble Monte Carlo method for systems with rough energy landscape. Phys Rev E 1997;56:2228–2233.
58. Hansmann UHE, Okamoto Y. Numerical comparisons of three recently proposed algorithms in the protein folding problem. J Comput Chem 1997;18:920–933.
59. Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. Chem Phys Lett 1997;281:140–150.
60. Shah N, Rejto PA, Verkhivker GM. Structural consensus in ligand–protein docking identifies recognition peptide motifs that bind streptavidin. Proteins 1997;28:421–433.
61. Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Rejto PA, Rose PW, Verkhivker GM. Thermodynamics and kinetics of ligand–protein binding studied with the weighted histogram analysis method and simulated annealing. In: Altman RB, Dunker AK, Hunter L, Klein T, Lauderdale K, editors. Pacific symposium on biocomputing-99. Singapore: Word Scientific; 1999, p 426–437.
62. 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.
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. Rejto PA, Bouzida D, Verkhivker GM. Examining ligand–protein interactions with binding energy landscapes. Theor Chem Acct 1999;101:138–142.
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 Recog 1999;12:371–389.
66. 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 Design 2000;14:731–751.
67. Mayo SL, Olafson BD, Goddard WA III. DREIDING: a generic force field for molecular simulation. J Phys Chem 1990;94:8897–8909.
68. Bouzida D, Kumar S, Swendsen RH. Efficient Monte Carlo methods for the computer simulation of biological molecules. Phys Rev A 1992;45:8894–8901.
69. Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. Phys Rev Lett 1988;61:2635–2638.
70. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195–1198.
71. Boczko EM, Brooks CL III. Constant-temperature free energy surfaces for physical and chemical processes. J Phys Chem 1993;97:4509–4513.
72. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. J Comp Chem 1992;13:1011–1021.
73. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional free energy calculations using the weighted histogram analysis method. J Comp Chem 1995;16:1339–1350.
74. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
75. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305–11310.
76. Galzitskaya OV, Finkelstein AV. A theoretical approach for folding/unfolding nuclei in three-dimensional protein structures. Proc Natl Acad Sci USA 1999;96:11299–11304.
77. Munoz V, Eaton WA. A simple model for calculating kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 1999;96:11311–11316.
78. Alm E, Baker D. Matching theory and experiment in protein folding. Curr Opin Struct Biol 1999;9:189–196.