Hierarchy of Simulation Models in Predicting Structure and Energetics of the Src SH2 Domain Binding to Tyrosyl Phosphopeptides

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., A Pfizer Company, 10777 Science Center Drive, San Diego, California 92121-1111

Received March 13, 2001

Structure and energetics of the Src Src Homology 2 (SH2) domain binding with the recognition phosphopeptide pYEEI and its mutants are studied by a hierarchical computational approach. The proposed structure prediction strategy includes equilibrium sampling of the peptide conformational space by simulated tempering dynamics with the simplified, knowledge-based energy function, followed by structural clustering of the resulting conformations and binding free energy evaluation of a single representative from each cluster, a cluster center. This protocol is robust in rapid screening of low-energy conformations and recovers the crystal structure of the pYEEI peptide. Thermodynamics of the peptide−SH2 domain binding is analyzed by computing the average energy contributions over conformations from the clusters, structurally similar to the predicted peptide bound structure. Using this approach, the binding thermodynamics for a panel of studied peptides is predicted in a better agreement with the experiment than previously suggested models. However, the overall correlation between computed and experimental binding affinity remains rather modest. The results of this study show that small differences in binding free energies between the Ala and Gly mutants of the pYEEI peptide are considerably more difficult to predict than the structure of the bound peptides, indicating that accurate computational prediction of binding affinities still remains a major methodological and technical challenge.

Introduction

The phosphorylation of protein residues is often the first detectable response to an external stimulus, implicating protein tyrosine kinases in mediating a variety of intracellular events, including cell prolifiration, metabolism, and immune response.[1] These proteins selectively bind their targets and initiate a cascade of signaling events through modular domains that control protein−protein interactions.[2] The Src Homology 2 (SH2) domain of the protein tyrosine kinases Src[3,4] plays an important role in signaling cascades by recognizing phosphotyrosine (pTyr) sequences and binding tightly and selectively to substrate proteins. The importance of the Src kinases as a promising target for drug discovery has been recognized since elevated levels of kinase activity have been implicated with breast cancer,[5] colon cancer,[6] and osteoporosis.[7] Structure and energetics of peptide−SH2 domain binding have been extensively studied in recent years to understand structure−function relationships and the molecular recognition mechanism.[8−18] The molecular basis of sequence-specific recognition has been analyzed using three-dimensional (3D) structures of SH2 domains complexed with nonspecific and specific phosphopeptides.[8−11] The crystal structures of the tyrosine kinases Src and Lck SH2 domains bound to the phosphotyrosyl peptide containing the recognition pYEEI motif have suggested a mechanism of sequence-specific binding that resembles a "two-pronged plug" engaging a "two-holed socket".[8,9] The recognition pYEEI peptide binds in an extended conformation with the Ile residue inserted into a hydrophobic pocket, providing an important determinant of specificity and high binding affinity, and the two Glu residues interacting with the protein side chain residues through water molecules.

Calorimetry studies of the Ala and Gly mutants of pYEEI have shown that the differences in the SH2 domain binding affinity between consensus high-affinity peptide and nonconsensus low-affinity peptides are less than 100-fold.[12−18] Mutations in the consensus pYEEI binding sequences cause only modest, less than 10-fold reductions in affinity, whereas a much larger 10 000-fold decline in binding affinity is observed when the pTyr residue is perturbed.[15−18] 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.[15−18] In general, peptide−SH2 domain binding displays the pattern of interactions, where the mutational effects of energetically important residues from binding partners are complementary.[19] Indeed, most of the binding free energy is delivered by hot-spot residues, such as Arg $\beta B5$ in the pTyr binding pocket and Tyr $\beta D5$ in the specificity region. Mutation of a single residue Arg $\beta B5$, in the pTyr binding pocket results in a 200-fold loss in binding affinity, and a nearly 30-fold loss of affinity is observed when Tyr $\beta D5$ is mutated to Ile.[14−17]

The exhaustive thermodynamic mapping of the Src SH2 domain−peptides interface, carried out with the aid of high-sensitivity isothermal calorimetry, has provided the accurate and comprehensive experimental

* To whom correspondence should be addressed. Tel, (858)622-3008; fax, (858)678-8244; e-mail, verk@agouron.com.

10.1021/jm0101141 CCC: $22.00
Published on Web 11/09/2001

estimates of the binding energetics, enthalpy, entropy contributions, and the heat capacity changes.[15−17] The structural and isothermal titration calorimetry binding studies of the specific, high-affinity PQpYEEIPI peptide, derived from the hamster middle T antigen (hmT), and low-affinity TQpYVPMLE and PQpYQPGEN peptides have revealed the major role of the enthalpy contribution in binding of the recognition peptide and the dominant entropy component in binding of the low-affinity peptides.[12] The interactions of the low-affinity peptides with the protein in the region C-terminal to the pTyr are characterized by a relatively small contact surface area (SA), as compared to the recognition peptide.[8,9,12] It has been suggested that the low-affinity TQpYVPMLE and PQpYQPGEN peptides could undergo significant conformational changes upon binding to the SH2 domain, leading to variations in local binding kinetics of the peptide residues.[12] Hence, the interactions at +3 hydrophobic pocket only partially determine high affinity and specificity of the recognition peptide, and the enthalpy contribution is the thermodynamic parameter that distinguishes binding energetics of the recognition high-affinity peptide and the low-affinity peptides. The loss of the hydrophobic contacts by modifying Ile at the pY+3 position with Ala and Gly leads to a less favorable enthalpy, indicative of a loss of favorable van der Waals interactions, that is only partly compensated by a more favorable entropy.[14,15] The heat capacity of phosphopeptides binding to the Src SH2 domain is rather small and quite similar for both high-affinity and low-affinity peptides, confirming a binding mechanism with no appreciable conformational change of the Src SH2 domain upon peptide binding and a profound conservation between the structures of the uncomplexed and complexed forms of the protein.[8,9,12−17]

Recent mass spectrometric and thermodynamic studies have revealed an important role of buried water molecules at the peptide−SH2 interface to the specificity of binding for the recognition peptide.[20] The X-ray crystal structures for the apo SH2 domain and complexes with the specific peptide have shown that the amount of water content in the binding site is retained in complexes with the pYEEI, pYEEA, and pYEEG peptides, with three critical buried water molecules tightly bound in the complexes and interacting with the glutamate residues at +1 and +2 positions. It has been proposed that the three buried water molecules present in complexes with the recognition peptides act as a part of the protein interface and contribute decisively to the promiscuity of the SH2 domain with respect to the specific recognition peptides.[20]

Nuclear magnetic resonance (NMR) relaxation studies of both free and complexed states of the SH2 domains with the recognition and nonspecific peptides have provided a powerful experimental tool to analyze the dynamics of the peptide−protein interfaces and entropy effects in thermodynamics and kinetics of binding.[21−27] These experiments have dissected the role of rigidity and flexibility in the mechanism of sequence-specific recognition. It has been shown that dynamics of the peptide−SH2 domain interface in the region C-terminal to the pTyr plays an important role in modulating binding affinity and regulating rapid exchange of pTyr sequences.[21−27] Flexibility of the hydrophobic residues C-terminal to the pTyr group has been observed in a comparative structural and thermodynamic analysis of binding for the specific EPQpYEEIPIYL peptide and the regulatory ATEPQpYQPGEN peptide with the Fyn member of Src SH2 domains.[28] The NMR shift data have indicated that the interactions of the specific EPQpYEEIPIYL peptide with a number of hydrophobic residues result in a considerable reduction of flexibility for the protein and peptide residues, which is reflected in the favorable negative enthalpy of binding. However, these interactions incur an entropic penalty by restricting side chain vibrational modes, which was suggested to be the cause of the observed unfavorable entropy component $T\Delta S = -2.0$ kcal/mol in binding with the Fyn member of Src SH2 domains.[28] In contrast, the regulatory peptide inserts in a well-defined conformation only in the pY binding pocket, while the C-terminal region fluctuates between a diverse range of low-energy bound conformations. A decreased internal motion of the protein residues has also been observed in NMR studies of the complexed Hck SH2 domain with the recognition EPQpYEEIPIYL peptide using NMR spectroscopy.[25,26]

The enthalpy contribution generally dominates the binding free energy differences between the recognition and the nonspecific phosphopeptides, whereas the entropy contribution plays an important role in the binding thermodynamics of nonspecific peptides. Structural insights of the phosphopeptide−SH2 domain recognition and enhanced understanding of the underlying thermodynamics have been utilized in a recent discovery of potent tetra- and pentapeptide ligands bound to the *pp*60^{c-src} SH2 domain.[29] A combination of isothermal calorimetry, X-ray crystallography, and computational analysis has allowed to rationalize the better potency of the discovered pentapeptide derivatives that have less favorable entropy but more favorable binding enthalpy due to a greater restriction of the ligand flexibility in the complex. The quantitative assessment of entropic effects depends largely on the degree of achieved desolvation, and as a result, obtaining new hydrogen bonds that are often solvent-exposed typically has a net neutral effect on binding affinity. Molecular dynamics simulations and normal mode calculations were able to explain some of the conformational aspects of the peptide binding, but the magnitude of these interactions was often over- or underestimated.[29] Structure-based design of nonpeptide ligands that bind to the *pp*60^{c-src} SH2 domain[30,31] and the crystal structures of the SH2 domain complexes with dipeptide analogues[32] has indicated that the ability to incorporate diverse structural motifs into modest peptides depends critically on the interactions in the hydrophobic pocket at the pY+3 position. Novel bicyclic nonpeptide inhibitors of the Src SH2 domain with a pTyr replacement by a novel moiety have been described.[33,34] The crystal structure of the SH2 domain complex with the inhibitor containing a pTyr replacement by a novel moiety conforms to the general topology of the ligand−protein interface with the bicyclic portion extending deeply into the Ile binding pocket at the pY+3 site.[35]

The binding energetics of the SH2 domain−pYEEI mutants rather poorly correlates with simple structural parameters such as the amount of the buried SA.[17] Moreover, the changes in both enthalpy and entropy

contributions to binding free energies are not in a reasonable agreement with these structural descriptors, suggesting that a detailed assessment of peptide binding energetics based on structural information for the mutant peptides is necessary. The lack of correlation between the experimental and the computed binding free energies has been originally attributed to the failure to account for structural alterations of the interacting components.

The desired synergy of the exhaustive sampling of the conformational space and accurate evaluation of the energetics is difficult to achieve for any given energy model. The rigorous computational approaches for calculation of binding affinities are the free energy perturbation (FEP) and thermodynamic integration methods.[36,37] The decomposition of the total free energies into components has been a subject of much interest and discussion,[38−44] and in principle, the first principles of statistical mechanics do not allow this separation.[39,40] However, complex biophysical phenomena such as molecular recognition are often analyzed on the atomic level with the aid of empirical free energy models, which rely upon a number of energetic terms to describe the binding process, and assume a given structure of the ligand−protein complex.[45−52]

Knowledge-based statistical ligand−protein interaction potentials derived from a database of inhibitor−enzyme complexes[53−55] eliminate any fitting of parameters of the scoring function to the observed binding affinities and avoid complications associated with a delicate balance between large contributions from the van der Waals and electrostatic interactions, hydrogen bonding solvation effects, and conformational entropy, which are difficult to compute accurately.[56]

In the linear interaction energy (LIE) approximation method,[57−68] free energies of binding are obtained from averages of the interaction energies between the ligand and the receptor in the bound state and water in the unbound state, using a total of two molecular dynamics simulations. Electrostatic interaction energies are evaluated using a linear response model, and van der Waals energies are usually scaled by an empirical factor, which is fitted to reproduce the observed binding free energies.

A new methodology, termed MM/PBSA (molecular mechanics Poisson−Boltzmann surface area), has been applied to a wide variety of macromolecules and complexes of macromolecules with ligands.[69−76] The average total free energy of the system is evaluated as the sum of the polar solvation energy, which is computed using a finite-difference Poisson−Boltzmann (PB) approach, the nonpolar solvation term derived from the solvent-accessible SA, and the solute entropy contribution. The molecular mechanical energy of the molecule includes the electrostatic, van der Waals contributions, and internal strain energy. The ensemble of structures for the uncomplexed protein and ligand is generated in the MM/PBSA approach by using the molecular dynamics trajectory of the complex and simply separating the protein and ligand coordinates, followed by an additional minimization of the unbound protein and unbound ligand. This method utilizes molecular dynamics simulations of the system to generate a thermally averaged ensemble of conformations. This methodology has been successfully applied to the analysis of protein−protein interactions,[70] free energy analysis of haptene binding to various forms of the antibody,[71] ligand−protein binding affinity prediction,[72,73] protein folding analysis,[74] and free energy calculations of HIV-1 protease dimer stability.[75]

Correlation between experimental binding free energies and simple theoretical models that describe binding in terms of changes in contributions from the nonpolar and polar components at the ligand−protein interfaces has been found in some complexes that associate as rigid bodies.[77,78] As a result, not only the hydration contributions but also the van der Waals and electrostatic intermolecular interactions may be proportional to the size of the ligand−protein interface.[78−80] However, the presence of packing defects or coupling between local folding and binding may destroy the simple relationship between the binding affinity and the amount of buried solvent-accessible SA. Furthermore, strong interactions between a few residues in the interface can complicate binding affinity prediction.[78]

Empirical solvent-accessible SA-based approaches, successfully applied in folding and binding,[81,82] have been recently used to assess the binding energetics of the recognition high-affinity PQpYEEIPI peptide from the hmT antigen and the effect of mutations at the pY+3 position.[83] None of the employed surface-based approaches that include different treatments of conformational flexibility in the peptide and various models of proximal-ordered water molecules has shown any qualitative agreement with the experimental ranking. Moreover, the adoption of different models had a drastic effect on the computed thermodynamic properties, thereby making the predicted binding free energies highly model-dependent.[83] A model that included the buried water molecules treated as a part of the protein structure and assumed the rigid peptide backbone along with flexible side chains has provided the best agreement between theory and experiment. It has been suggested that the model structures of the peptide mutants may significantly differ from the actual equilibrium ensemble of the peptide conformations interacting with the protein.[83]

Computational analysis of binding affinity for peptide−SH2 domain complexes using rigorous FEP methods is difficult and time-consuming, given the size of the peptides and a diverse range of large perturbations. We propose a multistage protocol, which provides a robust intermediate approach to achieve a synergy of an adequate conformational sampling and an accurate binding energetics, using a hierarchy of energy functions. In this approach, equilibrium sampling of the peptide conformational space is performed by simulated tempering dynamics with the simplified, knowledge-based energy function. This step is followed by structural clustering of the resulting conformations and binding free energy evaluation of a single representative from each cluster, a cluster center, to predict the structure of the peptide bound conformation. We show that this method is adequate for rapid screening of low-energy conformations and recovers the crystal structure of the pYEEI peptide. To analyze the thermodynamics of peptide−SH2 domain binding, binding free energies are computed by averaging the energy contributions over the conformations from the clusters, structurally

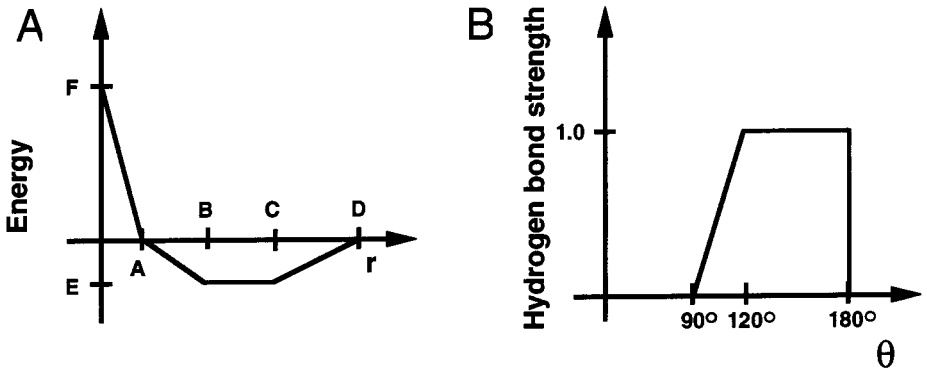

Figure 1. (A) Functional form of the ligand−protein interaction energy. For steric interactions, A = 0.93B, C = 1.25B, D = 1.5B, E = −0.4, F = 15.0, and B = $r_l + r_p$ is the sum of the atomic radii for the ligand and protein atoms. For hydrogen bond interactions, A = 2.3, B = 2.6, C = 3.1, D = 3.4, E = −4.0, and F = 15.0. For sulfur hydrogen bond interactions, A = 2.7, B = 30.0, C = 3.5, D = 3.8, E = −2.0, and F = 15.0. For chelating interactions with the metals, A = 1.5, B = 1.7, C = 2.5, D = 3.0, E = −10.0, and 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, and 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.

similar to the predicted peptide bound structure. Using this approach, binding thermodynamics is studied in this work for a panel of mutant peptides with Ala and Gly sequentially substituted for each residue at +1, +2, and +3 positions C-terminal to pTyr in the recognition pYEEI peptide. The results are compared with binding free energies of the minimized peptide complexes, obtained by introducing mutations in the crystal structure of the pYEEI peptide.

Materials and Methods

Molecular Recognition Energy Model. A hierarchical strategy is pursued with two different energy functions, a simplified energy function[84] in conjunction with Monte Carlo simulations[85−87] and the MM AMBER force field[88] combined with the solvation energy computed using a continuum generalized Born and solvent-accessible SA (GB/SA) solvation model.[89−95] The knowledge-based simplified energetic model includes intramolecular energy terms for the ligand, given by torsional and nonbonded contributions of the DREIDING force field,[96] 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 (Figure 1a).[84] 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, and nitrogen atoms is 1.8 Å; for flourine, it is 1.8 Å; and for sulfur and large nonpolar, it 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, while 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 allow 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 (Figure 1b). 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, while 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; i.e., the penalty is greatest when the angle θ is 180 degrees, fading to zero at 90 degrees and below.

For molecular docking simulations, it has been shown that the energy surface must be smooth for robust structure prediction of ligand−protein complexes;[97,98] 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. It has been shown that the simplified PL energy function produces reliable results in predicting crystal structures of ligand−protein complexes.[84−87,97,98] The PL energy function has no singularities at interatomic distances, effectively explores accessible ligand binding modes, and samples a large fraction of conformational space, particularly at high temperature.

In this study, a hierarchical approach is employed where the PL energy function is used in combination with a powerful searching technique, parallel Monte Carlo simulated tempering dynamics,[99−107] to adequately sample the conformational space and describe the multitude of the inhibitor binding modes. The advantage of the employed simulated tempering dynamics and the simplified energy function is the ability to adequately sample the conformational space and reliably determine the multitude of putative low-energy conformations of the ligand.

We assume that the simplified energy function, which has been found reliable in structure prediction of peptide complexes with SH2 domains, follows the shape of the "true" potential and can detect the density of low-energy states in the regions surrounding favorable binding modes. This assumption implies that the breadth of the local minima basins results from long-range character of hydrophobic interactions and should be recognizable by using the simplified knowledge-based energy function.[84−87,97,98] However, this function is less accurate in detecting the exact location and energetics of the native state because of the inaccuracy in quantifying the exact magnitude of ligand−protein interactions.

In ligand−protein binding, multistage approaches with a hierarchy of energy functions[108−111] have been found more reliable in predicting structures of the ligand−protein complexes. A two-step protocol, including structural clustering of low-energy conformations, generated in equilibrium simulations with the simplified energy function, is followed by the energy minimization with the AMBER force field, supplemented with a solvation correction term. This approach has

led to structure predictions in better agreement with experiment than using either energy function by itself and has allowed us to resolve some common failures in ligand−protein docking.[112]

In this work, a binding free energy approach, which includes a MM AMBER force field and a GB/SA solvation contribution, is used to evaluate the equilibrium samples, generated for each peptide with the PL energy function at $T = 300$ K, and thereby characterize more precisely the energetics of the putative binding modes. The crystal structure of the pYEEI peptide bound conformation with the Src SH2 domain and the minimized bound conformations of the peptide mutants, obtained by introducing mutations in the original crystallographic conformation of pYEEI, have been used as the reference structures in the analysis of equilibrium fluctuations.

The analysis of the equilibrium sampling can be facilitated by conducting structural clustering of the conformational space and binding free energy evaluation of a single representative from each cluster, a cluster center. The peptide−protein complex with the lowest energy, as measured in the AMBER force field with the GB/SA solvation model, determines the predicted peptide bound conformation. In the subsequent thermodynamic analysis, clusters of conformations, structurally similar to the predicted peptide structure, are used to evaluate the average contributions to the binding free energy components of the complex, protein, and peptide, respectively. The ensemble of structures for the uncomplexed protein and peptide are generated by using the corresponding samples of the peptide−protein complex and separating the protein and ligand coordinates, followed by an additional minimization of the unbound protein and unbound ligand. In this procedure, which is conceptually similar to the MM/PBSA approach, more time-consuming PB continuum calculations are replaced with less demanding GB solvation calculation, which generally correlate well with the PB values.[113]

The total entropy contribution is composed of solvation component and changes in conformational entropy in both the receptor and the ligand resulting from association and restricting the respective degrees of freedom in going from the free state to the bound state. The conformational states of the bound SH2 domain are very similar in complexes with different peptides and also virtually identical to the apo form of the protein. As a result, the change in conformational entropy of the SH2 domain is nearly constant for the studied peptide−SH2 domain complexes. The entropy components arising from ligand degrees of freedom are generally difficult to evaluate since they critically depend on flexibility in the bound state, and the magnitude of ligand fluctuations is significantly different from one conformational well to another. The harmonic approximation is used to assess the order of magnitude for the entropy loss during binding. The normal-mode analysis is carried out and vibrational entropy is computed from classical statistical mechanics formula with the AMBER module nmode for the energy minimized structures of the complex, the free protein, and the free peptide without water molecules. A dielectric constant of $4r_{ij}$, where r_{ij} is the distance between atoms i and j of the molecule, is used in the normal mode calculations.

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

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

$$G_{\text{solvation}} = G_{\text{cavity}} + G_{\text{vdw}} + G_{\text{pol}} \quad (2)$$

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

$$G_{\text{SA}} = G_{\text{cavity}} + G_{\text{vdw}} = \sum_i \sigma_i SA_i \quad (3)$$

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

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

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

$$E_{\text{MM}} = E_{\text{es}} + E_{\text{vdw}} + E_{\text{int}} \quad (5)$$

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

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \quad (6)$$

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

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

$$\Delta G_{\text{MM}} = E_{\text{MM}}^{\text{complex}} - E_{\text{MM}}^{\text{boundprotein}} - E_{\text{MM}}^{\text{boundligand}} \quad (8)$$

$$\Delta G_{\text{strain}} = (E_{\text{MM}}^{\text{boundprotein}} - E_{\text{MM}}^{\text{freeprotein}}) + (E_{\text{MM}}^{\text{boundligand}} - E_{\text{MM}}^{\text{freeligand}}) \quad (9)$$

$$\Delta G_{\text{GB/SA}} = G_{\text{solvation}}^{\text{complex}} - G_{\text{solvation}}^{\text{freeprotein}} - G_{\text{solvation}}^{\text{freeligand}} \quad (10)$$

The energy of each ligand−protein complex is subjected to the conjugate gradient minimization as implemented in version 7.0 of the MacroModel molecular modeling software package.[90] All protein residues within a 3 Å radius sphere from the ligand are treated as flexible during minimization. All protein residues within a 2 Å radius from the flexible shell form a first shell of restrained atoms with the force constant 100.0 kJ/mol A^2. A second shell of restrained atoms with the force constant 200 kJ/mol A^2 consists of the residues within a 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/mol A^2. The remaining protein atoms are treated as frozen atoms and do not move during the minimization procedure. The interactions between all type of atoms are counted in the total energy value, including interactions between frozen atoms and restrained atoms and frozen atoms and flexible atoms. A cutoff of 8 Å is set for computing nonbonded van der Waals interactions, and a 20 Å cutoff is used for computing electrostatic interactions. MNDO atomic charges have been derived for each peptide, and the protein atoms have been assigned the AMBER force field charges.

Monte Carlo Simulations of Ligand−Protein Interactions. In simulations of ligand−protein interactions, the protein is held fixed in its bound conformation, while 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 the 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, the bond angles, and the torsional angles of the unrotated bonds were obtained from the crystal structures of the bound ligand−protein complexes. Three critical buried water molecules from the crystal structure of the complex have been included in simulations as a part of the protein structure. Because there are no conformational changes upon 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.[113] The acceptance ratio method is used to update the step sizes every cycle of 1000 sweeps. For all of these simulations, we equilibrated the system for 1000 cycles (or one million sweeps) and collected data during 5000 cycles (or five 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 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

$$\sigma_j^{i+1} = \sigma_j^i \frac{\ln[a\langle P_{\text{ideal}}\rangle + b]}{\ln[a\langle P_j\rangle^i + b]} \tag{11}$$

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

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

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

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

Equilibrium simulations have been carried out 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 between adjacent replicas m and n is accepted or rejected according to Metropolis criterion with the probability

$$p = \min(1, \exp[-\delta]) \tag{14}$$

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 that 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 nonnegligible probability of moving through the entire temperature range and the detailed balance is never violated, which guarantees each replica of the system to be equilibrated in the canonical distribution with its own temperature.[99−107]

Similarity Clustering. The 3D similarity calculations are based on the spatial proximity of atoms in a binding site and the atom type. Four types of atoms are distinguished: hydrogen bond donors, hydrogen bond acceptors, hydrogen bond donors and acceptors, and nonpolar atoms. The atom type compatibility $a(i, j)$ is assigned a value between 0.0 and 1.0, with the compatibility between two atoms of the same type defined to be 1.0, that between a donor and acceptor atom is 0.0, and other combinations of atoms have compatibilities between 0.0 and 1.0.

The spatial proximity between two atoms i and j is evaluated with a Gaussian function $p(i, j) = 10^{(-r_{ij}^2/\sigma^2)}$, where r_{ij} is the distance between atoms i and j and $\sigma = -c^2/\log(p)$, where c and p denote the cutoff distance and proximity threshold, respectively. Both the cutoff distance and the proximity threshold determine the shape of the Gaussian function to evaluate spatial proximity of two atoms, with $c = 3.0$ Å and $p = 0.000\,032$.

A descriptor $d(i, j)$ is calculated from the spatial proximity and the atom type compatibility

$$d(i, j) = p(i, j) \times a(i, j) if r(i, j) \le c \tag{15}$$

$$d(i, j) = 0 if r(i, j) > c \tag{16}$$

An atom descriptor $D_m^n(i)$ for atom i in molecule m is then calculated by summation over all N atoms in molecule n, $D_m^n(i) = \sum_{j=1}^N d_m^n(i, j)$. The intermolecular similarity between molecules m and n is given by the Tanimoto coefficient[115−117]

$$S(m, n) = \left(\sum_{i=1}^M D_m^m(i)\, D_m^n(i) + \sum_{j=1}^N D_n^m(j)\, D_n^n(j)\right) \Big/ \left(\sum_{i=1}^M D_m^m(i)^2 + \sum_{j=1}^N D_n^n(j)^2 - \sum_{i=1}^M D_m^m(i)\, D_m^n(i) - \sum_{j=1}^N D_n^m(j)\, D_n^n(j)\right) \tag{17}$$

Molecules are grouped into clusters by comparing the intermolecular similarity coefficient. The first molecule is assigned to the first cluster. The next molecule is assigned to the cluster in which a cluster member has the highest similarity with the next molecule, if the similarity is above a threshold, chosen to be 0.85. Otherwise, the next molecule is assigned to a new cluster. The first member of the a cluster is called the cluster center. After all molecules are assigned to clusters, the molecules are arranged in a new order, starting with the largest cluster and proceeding to the smallest cluster. The reordered set of molecules is subjected to the same clustering procedure. This procedure is iterated until the information entropy converges to a minimum. The clusters with at least 100 members are analyzed. Because conformations that belong to the same cluster are equivalent with 85% structural similarity, different clusters are compared by analyzing cluster centers.

Results and Discussion

We begin with the structural analysis of the equilibrium simulations conducted for the panel of the Ala and Gly mutants of the pYEEI peptide. The equilibrium simulations of the pYEEI peptide binding (Figure 2a) and the alanine peptide mutants pYAEI, pYEAI, and pYEEA (Figure 2b−d) have shown frequent transitions between the nativelike binding mode, located at root mean square deviation (RMSD) = 2.0 Å from the reference structures and the binding mode at RMSD = 4.0−5.0 Å from the reference states (Figure 2). A considerably broader range of fluctuations with frequent

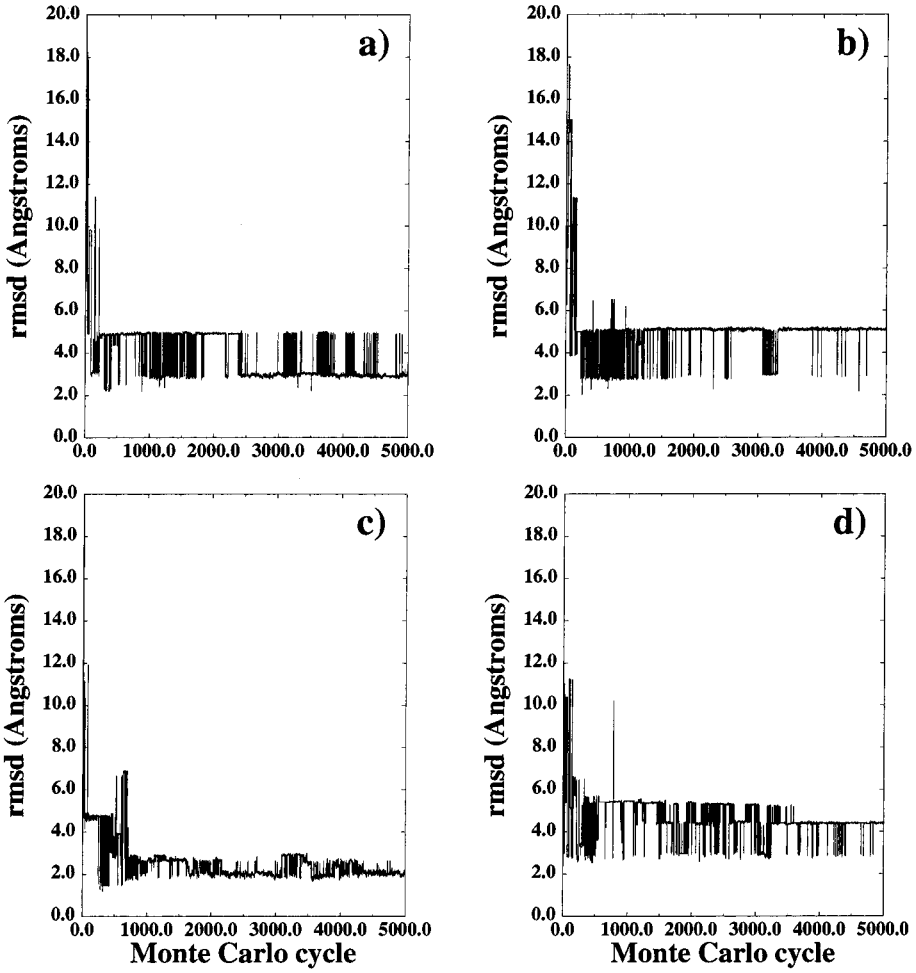

Figure 2. Time-dependent equilibrium history of the Src SH2 domain binding with the pYEEI peptide (a) and its Ala mutants at $T = 300$ K: pYAEI (b), pYEAI (c), and pYEEA (d).

excursions to the alternative low-energy bound conformations, which are situated at RMSD = 8−9 Å from the reference states, has been observed in equilibrium simulations of the Gly mutants pYGEI, pYEGI, and pYEEG (Figure 3b−d). This suggests that the Gly mutants may exhibit a greater degree of conformational flexibility in the bound state than the recognition pYEEI peptide and its Ala mutants, which is consistent with the experimentally observed more favorable binding entropy for the Gly mutants.

A convenient approach to distinguish energy minima that are distant from the reference structure is to generate clusters of structurally similar conformations, since two distinct binding modes will result in two different conformational clusters. Conformations generated in the equilibrium simulations of the pYEEI peptide binding have been clustered, and the cluster centers have been superimposed on the crystal structure of the bound pYEEI (Figure 4). A significant conformational diversity of the bound peptide conformations can be observed, while the topology of the extended bound conformation is still retained. Structural features of the dominant topology for the bound pYEEI peptide that are present in the low-energy bound conformations from various structural clusters constitute an extended conformation of the inhibitors with strong specific interaction in the pTyr pocket and favorable hydrophobic interactions at the pY+3 site (Figure 4).

The pYEEI bound conformations from clusters 3, 5, 9, and 10, which are located at RMSD = 1.5, 2.5, 2.8, and 2.3 Å, respectively, from the crystal structure (Figure 5a) have a more favorable total binding free energy (Figure 5b). A subtle balance between the binding free energy components, the favorable peptide−protein interaction energy, and the solvation energy loss plays a crucial role in determining the lowest energy structure (Figure 5c,d). Nevertheless, there is a relationship between the binding free energies of the determined low-energy peptide conformations from different clusters and their deviation from the crystal structure of the bound pYEEI peptide (Figure 5a,b). The closer the predicted structure conforms to the crystallographic conformation, the lower the energy. This suggests that despite an inherently sensitive nature of the binding free energy functions, consisting of large and compensatory contributions, these models can discriminate the crystal structure of the bound pYEEI peptide as the lowest energy solution. Structural clustering of the equilibrium samples and subsequent binding free energy evaluation of cluster centers requires only a relatively small number of conformational samples to recover the crystal structure of the pYEEI peptide and therefore allows for rapid screening of the low-energy conformations.

The low-energy bound conformations of the pYEEA and pYEEG peptides, superimposed on the crystal

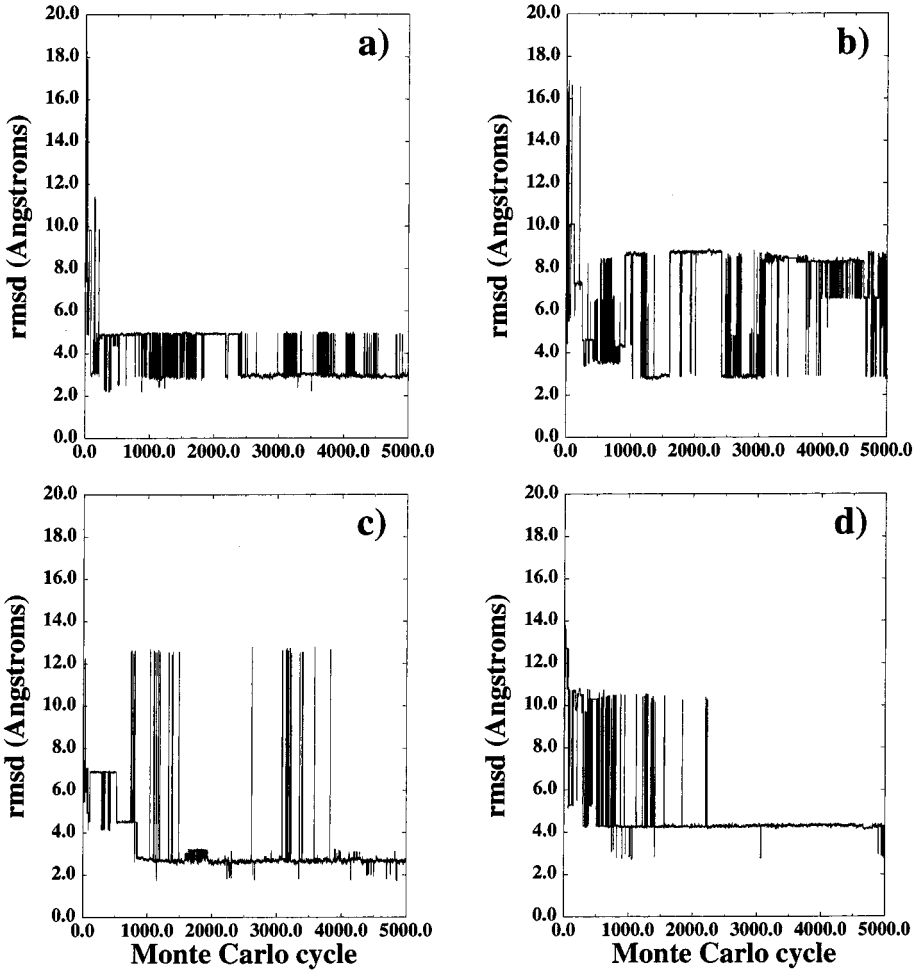

Figure 3. Time-dependent equilibrium history of the Src SH2 domain binding with the pYEEI peptide (a) and its Gly mutants at $T = 300$ K: pYGEI (b), pYEGI (c), and pYEEG (d).

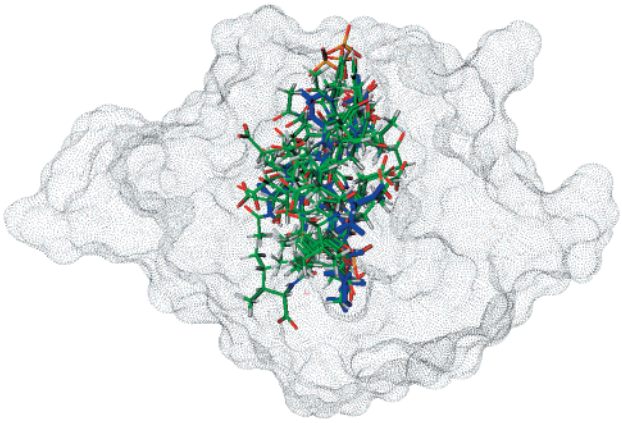

Figure 4. Superposition of the crystallographic conformation of the pYEEI peptide (blue) bound to the Src SH2 domain protein structure with 10 low-energy cluster centers (color-coded by atom type) obtained for pYEEI. Connoly surface of the Src SH2 domain protein in the complex with pYEEI is shown.

structure of the pYEEI bound conformation, demonstrate the increased conformational diversity (Figures 6 and 7). However, the overall topology of the crystal structure of the pYEEI complex is still retained in the predicted conformations. Importantly, the conformations that are the closest to the reference structure have indeed the lowest binding free energy for the Ala mutants (Figure 8) and are predicted as the bound conformations for the respective peptides. However, the sensitive nature of the binding free energy model may lead to alternative low-energy solutions for the pYEAI peptide, corresponding to clusters 6 and 8, which are close energetically to the predicted structure (Figure 8c,d) but are located at RMSD = 4.0 and 6.1 Å, respectively, from the reference structures. For the pYEEA peptide, the low-energy solutions corresponding to clusters 1, 3, and 6 are situated at RMSD = 1.5, 1.6, and 1.7 Å, respectively, from the reference state (Figure 8e,f). However, there is also an energetically similar solution, corresponding to cluster 5, which resides at RMSD = 5.7 Å from the reference structure. Energetically similar but structurally different low-energy conformations have also been observed for the pYGEI and pYEGI peptide mutants (Figure 9a−d), suggesting that the "funnel-like" behavior of the energy function observed for the pYEEI mutant is not general and cannot be readily extended to the energetics of the pYEEI mutants. It is worth mentioning that for all peptides, with the exception of pYEEA and pYEGI, binding free energies are computed by averaging the energy components over conformations from the unique cluster to which the predicted peptide bound structure belongs. For pYEEA, structurally similar, low-energy conformations from clusters 1 and 3 (Figure 8e,f) are also

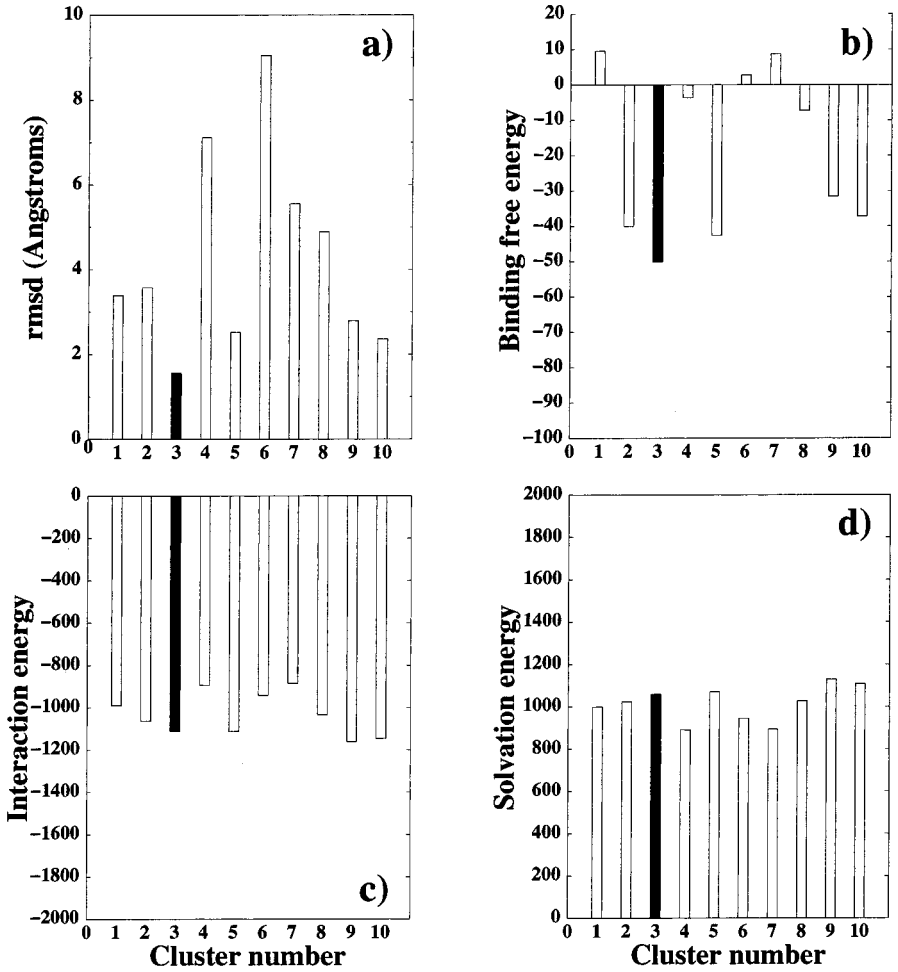

Figure 5. Structural and energetic analysis of the pYEEI peptide: RMSD values of the pYEEI cluster centers from the crystal structure of pYEEI (a), the total binding free energy of the pYEEI cluster centers (b), the interaction energy of the pYEEI cluster centers (c), and the solvation energy of the pYEEI cluster centers (d).

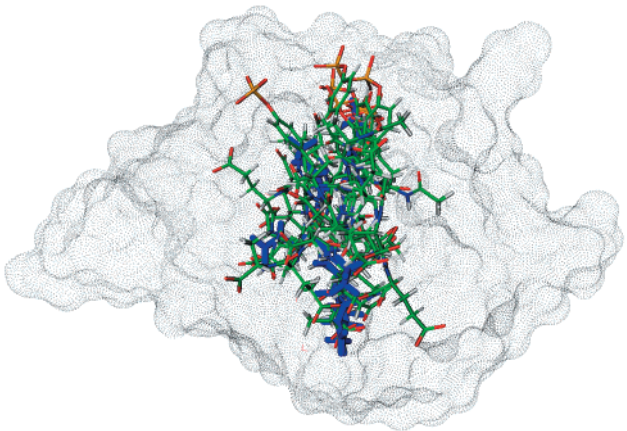

Figure 6. Superposition of the crystallographic conformation of the pYEEI peptide (blue) bound to the Src SH2 domain protein structure with 10 low-energy cluster centers (color-coded by atom type) obtained for the pYEEA mutant peptide. Connoly surface of the Src SH2 domain protein in the complex with pYEEI is shown.

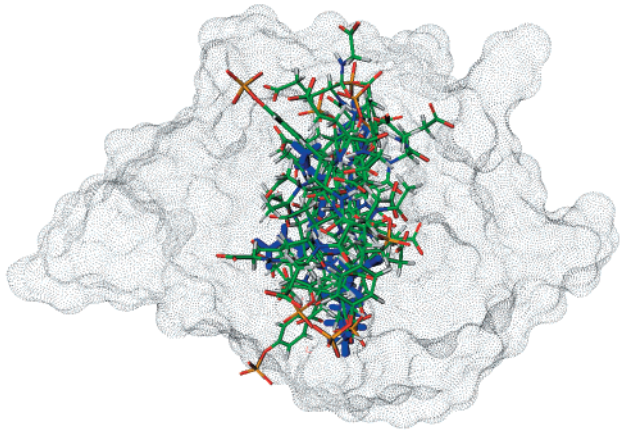

Figure 7. Superposition of the crystallographic conformation of the pYEEI peptide (blue) bound to the Src SH2 domain protein structure with 10 low-energy cluster centers (color-coded by atom type) obtained for the pYEEG mutant peptide. Connoly surface of the Src SH2 domain protein in the complex with pYEEI is shown.

included in calculations of the average energies. In the case of pYEGI, all conformations from cluster 8 (Figure 9c,d) are considered in the binding free energy calculations. The superposition of the pYEEI crystal structure with the predicted bound conformations for Ala and Gly mutants (Figure 10) has indicated that the topology of the extended bound conformation and the energetically favorable interactions formed by the phosphotyrosyl group and hydrophobic residue at the pY+3 position are well-preserved. The increased flexibility is observed for

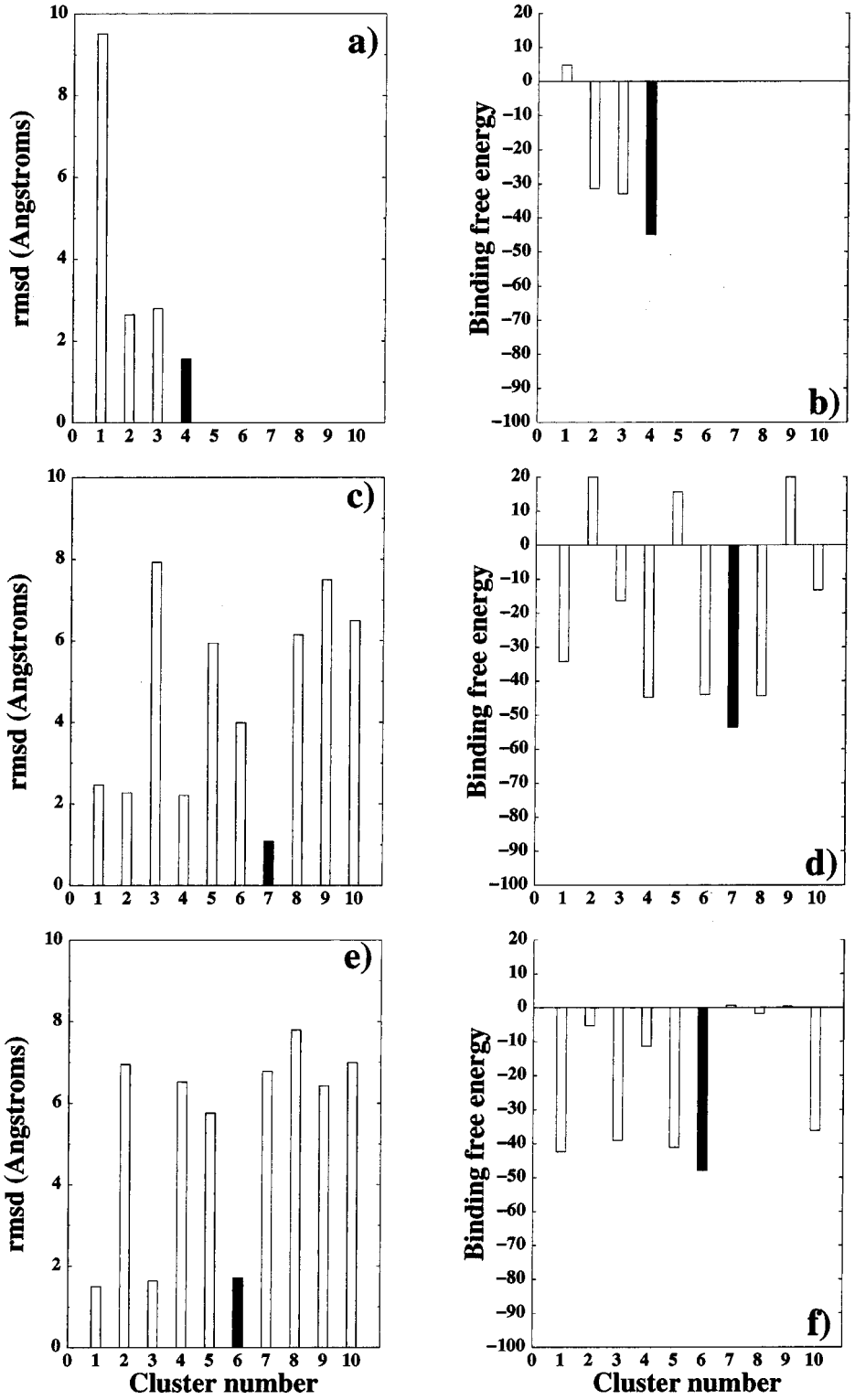

Figure 8. Structural and energetic analysis of the Ala mutants of the pYEEI peptide: RMSD values of the pYAEI cluster centers from the crystal structure of pYEEI (a) and the total binding free energy of the pYAEI cluster centers (b); RMSD values of the pYEAI cluster centers from the crystal structure of pYEEI (c) and the total binding free energy of the pYEAI cluster centers (d); RMSD values of the pYEEA cluster centers from the crystal structure of pYEEI (e) and the total binding free energy of the pYEEA cluster centers (f).

the pYEEA peptide and all Gly mutants, while the topology of the bound peptide is retained for the pYAEI and pYEAI peptides due to important water-mediated specific interactions between the glutamate residues at the +1 and +2 positions and the buried water molecules, treated as a part of the protein structure. The range of structural deviations of the predicted peptide conformations from the reference states (Figure 11) has reflected a general trend for the Gly mutants to further deviate from the initial conformations generated by mutations in the crystallographic conformation of the pYEEI peptide.

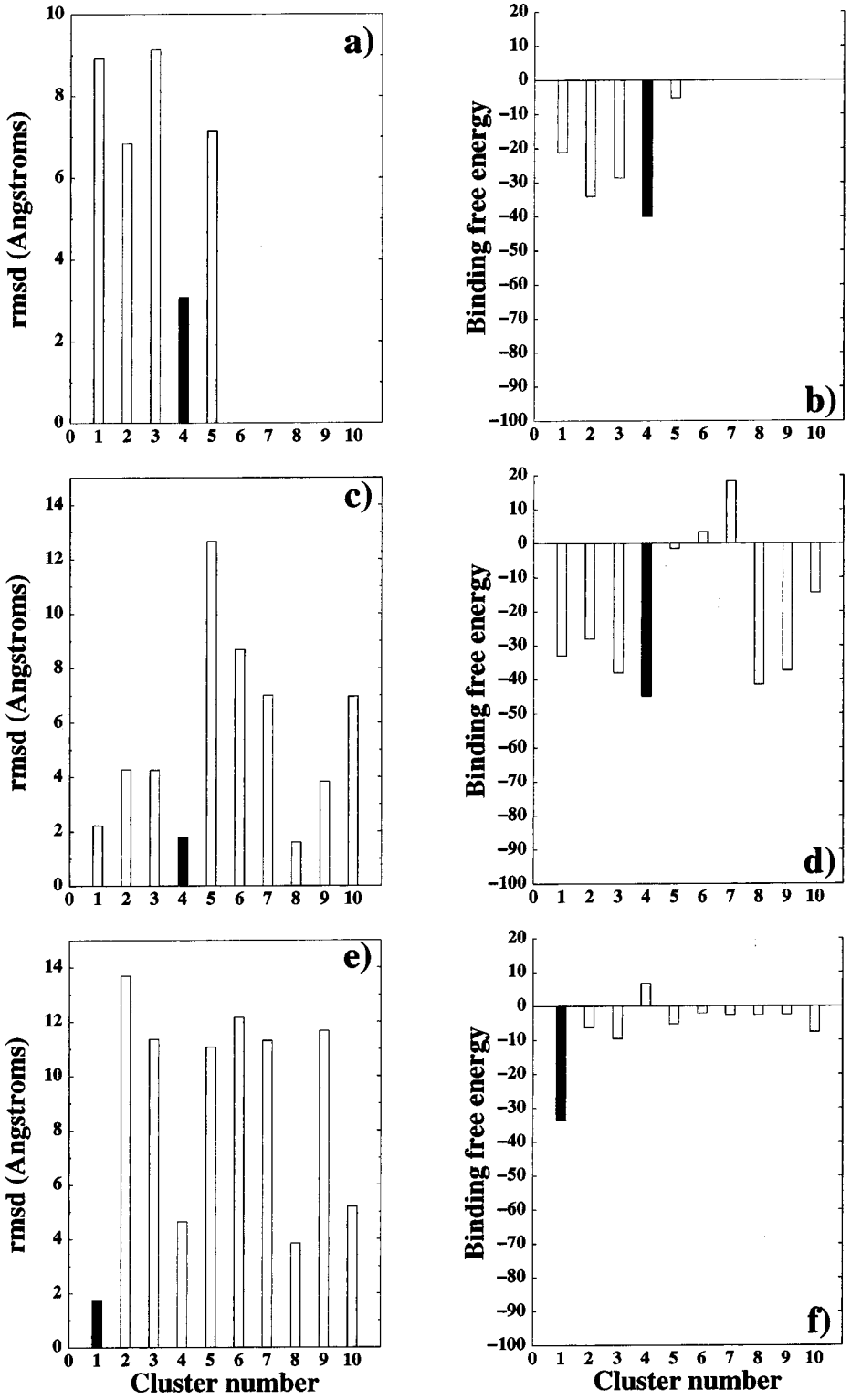

Figure 9. Structural and energetic analysis of the Ala mutants of the pYEEI peptide: RMSD values of the pYGEI cluster centers from the crystal structure of pYEEI (a) and the total binding free energy of the pYGEI cluster centers (b); RMSD values of the pYEGI cluster centers from the crystal structure of pYEEI (c) and the total binding free energy of the pYEGI cluster centers (d); RMSD values of the pYEEG cluster centers from the crystal structure of pYEEI (e) and the total binding free energy of the pYEEG cluster centers (f).

Upon binding, the +3 Ile of the pYEEI peptide makes a number of intimate contacts with a pocket of the Src SH2 domain formed by hydrophobic residues. The loss of the favorable hydrophobic contacts in the pocket by substituting Ala and Gly at the +3 position causes the loss in binding affinity of 1.4 kcal/mol for Ala and 1.8 kcal/mol for Gly.[5−10] The results of binding free energy calculations reveal a consistent decrease in binding affinity of the pYEEI peptide upon mutation to Ala and Gly at +3 position in agreement with the experimental trend (Tables 1−3). This trend, however, could not be reproduced using the minimized reference structures of

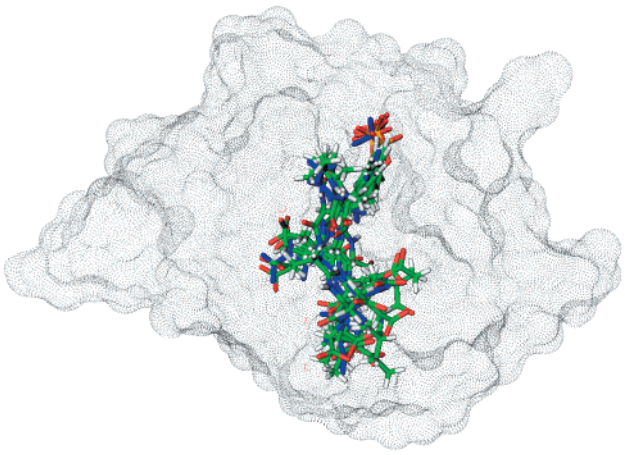

Figure 10. Superposition of the crystallographic conformation of the bound pYEEI peptide (blue) with the lowest energy cluster centers (color-coded by atom type) obtained for the pYAEI, pYEAI, pYEEA, pYGEI, pYEGI, and pYEEG peptides. Connoly surface of the Src SH2 domain protein in the complex with pYEEI is shown.

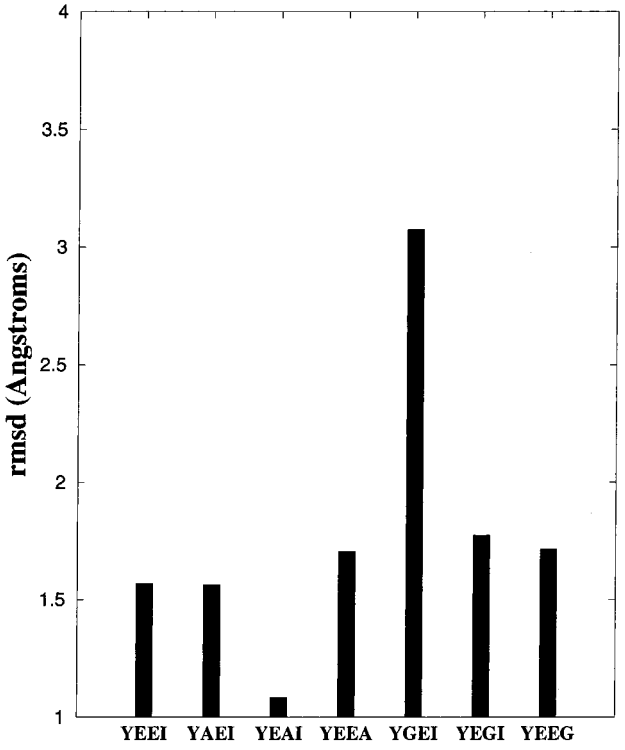

Figure 11. RMSD values for the lowest energy cluster centers obtained for the pYEEI, pYAEI, pYEAI, pYEEA, pYGEI, pYEGI, and pYEEG peptides.

these peptides (Tables 1 and 3). The replacement of Glu at +1 position by Ala and Gly results in a modest reduction of binding for pYAEI and a more dramatic loss of affinity for pYGEI.[5−10] This may result from the loss of the interactions between the peptide side chain at the +1 position and critical Tyr $\beta D5$ protein residue.[10] The computed average binding free energies reproduce the experimental trend of the binding affinity loss upon Ala and Gly mutations at the +1 position (Tables 2 and 3). Moreover, the free energy model suggests that a considerable loss in binding enthalpy upon mutations of Glu at the +1 position results from the loss of favorable hydrophobic interactions of C_γ and especially C_β side chain atoms of Glu with the Tyr $\beta D5$ protein residue. A small favorable contribution of the vibrational entropy to binding of the pYAEI and pYGEI peptides is found (Table 3), suggesting that the interactions of Glu at the +1 position with the Tyr $\beta D5$ protein residue may restrict the flexibility of the peptide. As a result, once this contact is lost, the entropy of the bound peptide tends to compensate for the apparent loss of favorable interactions. This experimentally observed trend could not be reproduced by the binding free energy model when the minimized reference mutant structures were used (Tables 1 and 3).

The computed binding free energies also reproduce a more dramatic loss in affinity for the pYGEI mutant as compared to pYEGI (Table 2). However, this difference in binding affinity is a result of a considerably less favorable binding enthalpy for the pYGEI mutant. The computational analysis attributes this net result to a more significant solvation penalty for the pYGEI mutant, suggesting that there may be a contribution from the enthalpy of solvation that could drive the observed phenomenon. However, some caution is necessary when the net binding free energy differences are rationalized on the basis of individual contributions.

The energetic analysis for the pYEAI and pYEGI peptides reveals a decrease in the peptide−protein interaction energy (Tables 2 and 3) upon substitutions at the +2 position, resulting in part from the loss of favorable water-mediated interaction between the Glu side chain at the +2 position and Arg $\beta D1$. This deficit is compensated by a less detrimental solvation penalty and a more favorable solvation energetics of the complex. A network of interactions connecting the Glu side chain at the +2 position and highly conserved Arg $\beta D1$ is important for integrity of the water structure, which plays a role in ensuring specificity of binding for the recognition peptides.[20] Because the formation of hydrogen bonds contributes to the enthalpy of binding, the experimentally observed changes in enthalpy upon mutations at the +2 position may be a result of loss of water-mediated interactions, abrogated in the presence of Ala and Gly residues.

The computed binding free energies for the minimized Ala and Gly mutants, obtained by modifying the corresponding residues in the crystal structure of the pYEEI peptide, show no correlation with the experimental binding affinity in the absence of the entropy component (Figure 12a). Moreover, these computed free energies reveal a similar lack of agreement with the experimental enthalpy of binding (Figure 12b). However, the addition of the entropy term has only a marginal effect on the resulting correlation (Figure 13a,b). Incorporation of statistically meaningful ensembles of peptide conformations results in a better quantitative agreement with the experimental data than the computed binding free energies of the mutant complexes obtained by introducing mutations in the crystallographic conformation of pYEEI and recent empirical approaches based on the model peptide structures.[59] A correlation between the experimental and the computed binding free energies recovers from $R^2 = 0.41$ to 0.51 when the entropy component is included in calculations (Figures 12c and 13c). The correlation with the experimental

Table 1. Energy Contributions to the Free Energy of Binding for the Minimized Mutant Complexes[a]

	complex		bound protein		bound peptide		unbound minimized peptide		unbound minimized protein	
peptide	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$
YEEI	−2497.09	−1503.97	−1551.16	−1860.10	216.91	−735.91	214.30	−734.93	−1636.28	−1797.35
YAEI	−2482.57	−1449.80	−1550.45	−1858.67	79.68	−528.58	76.23	−526.77	−1636.28	−1797.35
YEAI	−2436.95	−1474.18	−1551.88	−1859.60	89.19	−517.97	87.03	−517.43	−1636.28	−1797.35
YEEA	−2525.96	−1478.87	−1549.44	−1860.14	219.56	−739.93	216.90	−739.00	−1636.28	−1797.35
YGEI	−2482.40	−1447.67	−1550.33	−1858.65	82.59	−529.88	78.93	−527.84	−1636.28	−1797.35
YEGI	−2467.32	−1468.43	−1546.56	−1862.42	72.24	−522.04	69.96	−521.87	−1636.28	−1797.35
YEEG	−2530.27	−1495.08	−1549.68	−1859.94	201.28	−744.71	198.63	−743.76	−1636.28	−1797.35

[a] Mutant complexes are obtained by introducing mutations in the crystallographic conformation of pYEEI. All energies are given in kcal/mol. The experimental binding free energies are taken from ref 14.

Table 2. Average Energy Contributions to the Free Energy of Binding Computed from the Predicted Structures of the Peptides by the Hierarchical Approach[a]

	complex		bound protein		bound peptide		unbound minimized peptide		unbound minimized protein	
peptide	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$	E_{MM}	$G_{GB/SA}$
YEEI	−2534.96	−1474.65	−1561.37	−1849.59	216.21	−740.16	211.20	−737.07	−1636.28	−1797.35
YAEI	−2507.02	−1428.12	−1594.26	−1818.56	92.48	−546.46	90.26	−546.89	−1636.28	−1797.35
YEAI	−2479.70	−1465.60	−1546.79	−1863.84	67.83	−523.45	65.50	−523.54	−1636.28	−1797.35
YEEA	−2550.70	−1472.19	−1555.62	−1852.05	198.34	−736.16	192.60	−733.87	−1636.28	−1797.35
YGEI	−2521.14	−1411.09	−1564.19	−1849.15	66.98	−523.53	63.78	−522.47	−1636.28	−1797.35
YEGI	−2510.02	−1421.51	−1574.48	−1831.66	66.17	−515.67	62.03	−515.05	−1636.28	−1797.35
YEEG	−2583.83	−1436.86	−1547.92	−1854.63	180.00	−731.74	178.04	−731.44	−1636.28	−1797.35

[a] All energies are given in kcal/mol. The experimental binding free energies are taken from ref 14.

Table 3. Energy Contributions to the Binding Free Energy of the Peptides with the Src SH2 Domains[a]

	computational protocol I				computational protocol II				experiment	
peptide	ΔG_{MM}	$\Delta G_{GB/SA}$	$T\Delta S_{vib}$	ΔG total	ΔG_{MM}	$\Delta G_{GB/SA}$	$T\Delta S_{vib}$	ΔG total	ΔH	ΔG
YEEI	−1075.11	1028.31	−6.40(1.5)	−40.40(1.2)	−1109.88	1059.77	−3.76(1.6)	−46.35(1.5)	−7.7(0.2)	−9.2(0.1)
YAEI	−922.52	874.32	−1.43(1.1)	−46.77(0.8)	−961.00	916.12	0.25(1.0)	−45.13(1.1)	−7.7(0.2)	−8.7(0.1)
YEAI	−887.70	840.60	−11.90(1.8)	−35.20(1.1)	−908.92	855.29	−7.10(1.5)	−46.53(1.6)	−6.3(0.1)	−8.2(0.1)
YEEA	−1106.58	1057.48	−11.64(1.6)	−37.46(1.3)	−1107.02	1059.03	−6.72(0.7)	−41.27(1.0)	−5.1(0.4)	−7.8(0.4)
YGEI	−925.05	877.52	−1.25(0.9)	−46.28(0.7)	−948.64	908.73	0.79(0.9)	−40.70(1.1)	−4.8(0.1)	−7.1(0.1)
YEGI	−901.00	850.79	−17.91(1.6)	−32.30(0.9)	−935.77	890.89	−6.64(1.2)	−38.24(1.5)	−6.0(0.1)	−7.8(0.1)
YEEG	−1092.62	1046.03	−8.92(1.2)	−37.67(1.0)	−1125.59	1091.93	−2.26(1.0)	−31.40(1.4)	−3.6(0.1)	−7.4(0.1)

[a] All energies are given in kcal/mol. The experimental binding free energies are taken from ref 14. ΔG_{MM}, $\Delta G_{GB/SA}$, $T\Delta S_{vib}$, and ΔG total contributions to the binding free energy are defined in the Materials and Methods section. In the protocol I, energy contributions to the binding free energy are determined based on the minimized mutant complexes generated by introducing mutations in the crystal structure of the pYEEI peptide. In the protocol II, energy contributions to the binding free energy are determined based on the average contributions computed from the predicted structures of the peptides with the hierarchical approach. The values in parentheses represent the standard error of the mean.

enthalpies improves from $R^2 = 0.47$ to 0.70 in the more detailed model (Figures 12d and 13d). Nevertheless, the average binding free energies based on the predicted bound conformation still exhibit a rather modest correlation with the experimental binding affinity.

The computed changes in vibrational entropy upon binding for the minimized peptide mutants are large and unfavorable, while these values are considerably smaller when determined using the ensembles of structurally similar low-energy conformations (Table 3). The observed magnitude of the vibrational entropy changes becomes much closer to the range of the experimental entropy when the average contributions are computed using the MM/GB/SA model. The changes in vibrational entropy can be directly compared to the contribution due to translational and rotational entropy loss upon binding,[118] because six new internal degrees of freedom in the complex correspond to the lost translational and rotational motions of the individual components. It has been shown in recent experimental[119] and theoretical estimates[120,121] that the loss of translational and rotational entropy appears considerably smaller, approximately $T\Delta S = 3-4$ kcal/mol than the originally suggested range between 12 and 18 kcal/mol.[122,123] The changes in vibrational entropy upon binding with the peptides are generally unfavorable (Table 3); nevertheless, their magnitude is comparable with the recent, more accurate estimates for the loss of rotational and translational entropy. A relatively large error typically associated with computations of the vibrational entropy changes makes the prediction of the entropic changes quite difficult. However, the error bar in calculations of vibrational entropy of studied peptides is quite reasonable given the size of these ligands and results from averaging the entropy contribution over structurally similar and minimized conformations (Table 3). It is worth mentioning that in MM/PBSA studies of ligand−protein binding the errors in vibrational entropy estimates were much larger, up to 5 kcal/mol.[70−73] Only, after averaging over structurally similar conformations,

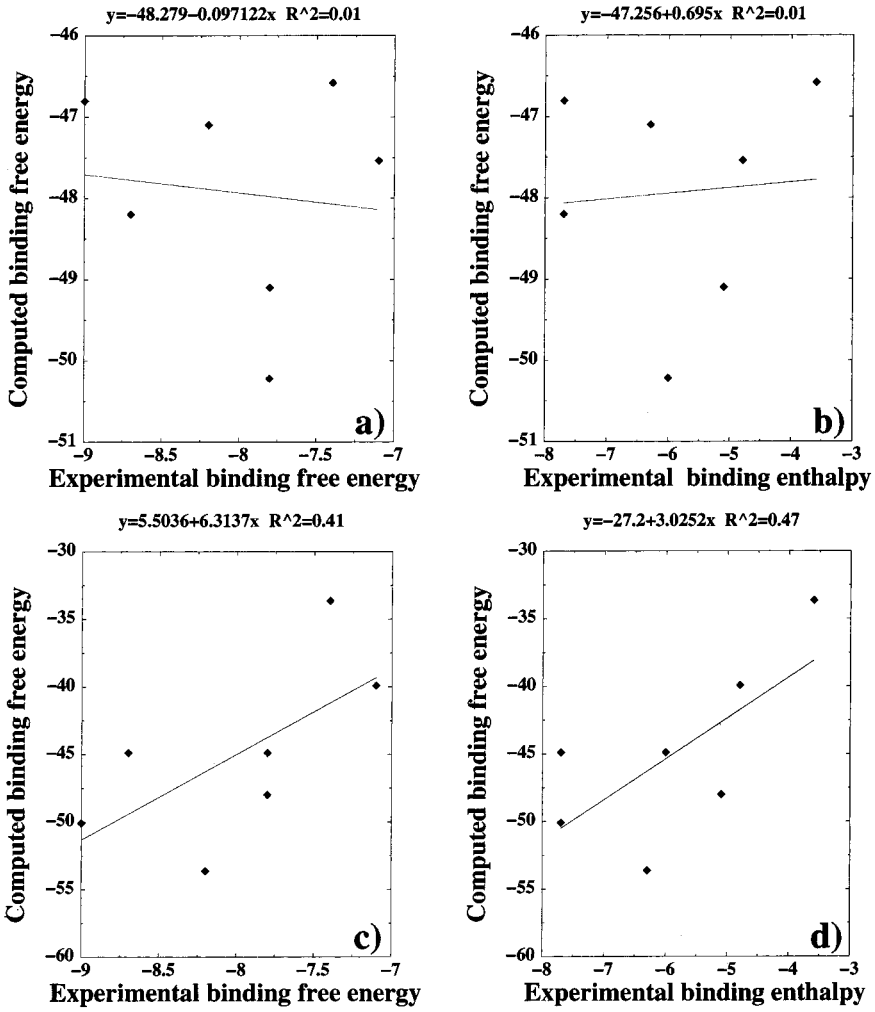

Figure 12. Correlation between the computed binding free energies of the minimized mutant complexes, obtained by introducing mutations in the crystallographic conformation of the pYEEI peptide and the experimental binding free energies (a) and enthalpies (b). Correlation between the computed binding free energies obtained with the hierarchical approach and the experimental binding free energies (c) and enthalpies (d). No entropy contribution is included in the computed binding free energies.

this error was reduced to 0.5−1.5 kcal/mol. The structural analysis of the peptide−SH2 domain complexes suggests that using a representative ensemble of structures for the reference ligand bound to the protein, followed by mutation of the reference ligand into the desired derivative for each snapshot, as typically employed in MM/PBSA applications, may sometimes lead to erroneous results. This assumption is valid only when the equilibrium sampling of the mutant ligand closely resembles the trajectory of the reference ligand. For the pYEEA and pYEEG peptide mutants, the predicted lowest energy conformations considerably deviate from the reference structures. Even when the molecular dynamics trajectories are generated separately for each mutant ligand, the system may get trapped in local minima and never reach the equilibrium at lower temperatures. In the employed hierarchical approach, the conformational space of the peptides can be efficiently explored at the first stage of the protocol. Consequently, the binding free energies are computed by averaging the contributions over low-energy states from the basin, surrounding the most energetically favorable peptide binding mode.

There are a number of limitations of the proposed hierarchical approach in its current implementation, including only a partially flexible receptor and a simplified model for the entropy contribution. A detailed description of the conformational entropy can generally include the following three contributions: the entropy change associated with the transfer of a side chain from the interior of the protein to its surface; the entropy change gained by a surface-exposed side chain when the backbone changes from a unique folded conformation to an unfolded conformation; and the entropy change gained by the backbone upon unfolding from a unique native conformation.[124,125] At the present time, no reliable models for estimates of these contributions are available. Nevertheless, a structure-based thermodynamic analysis approach, which utilized this entropy model, has reproduced the balance of stabilizing contributions and the magnitude of the Gibbs free energy for HIV-1 protease stabilization and binding to various inhibitors in agreement with the experimental measurements.[126] More experimentation is required to assess the limitations of the restrained residues lying outside the active site. In fact, thermodynamic mapping of the Src SH2 domain−pYEEI peptide interface has stressed that the seemingly isolated peptide−protein interactions can be dependent on the composition of the surrounding protein environment.[17]

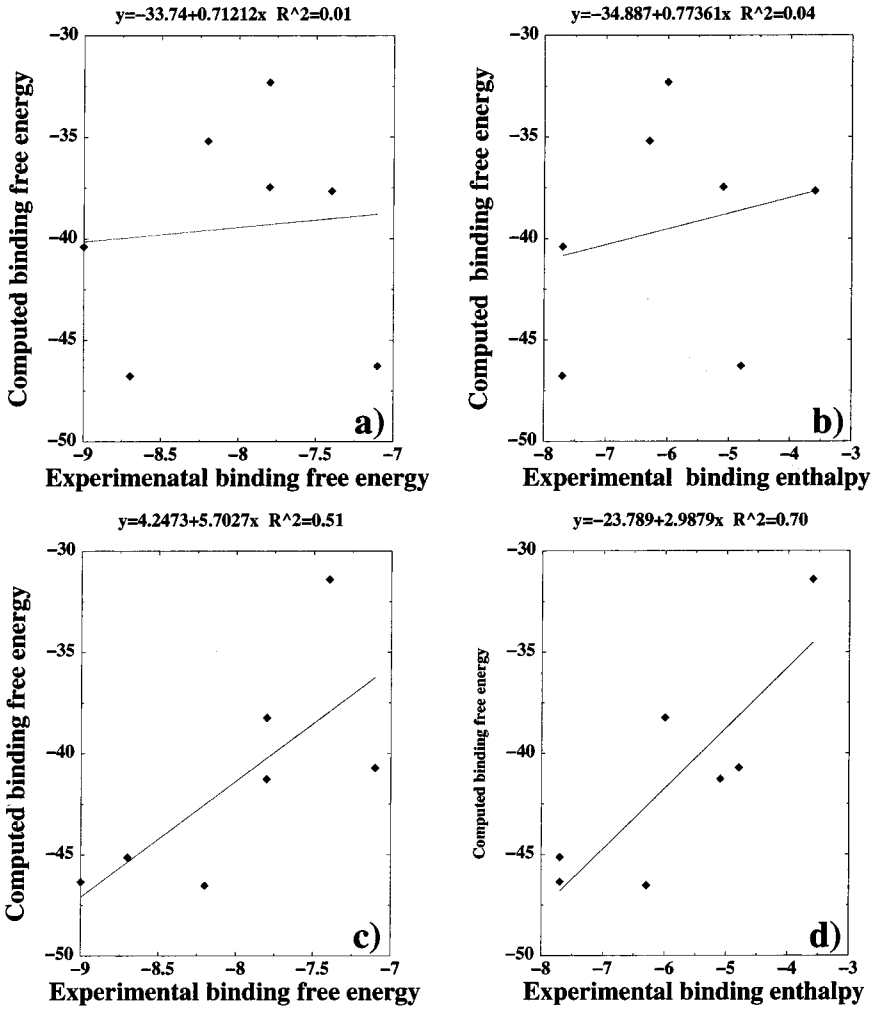

Figure 13. Correlation between the total computed binding free energies of the minimized mutant complexes, obtained by introducing mutations in the crystallographic conformation of the pYEEI peptide and the experimental binding free energies (a) and enthalpies (b). Correlation between the total computed binding free energies obtained with the hierarchical approach and the experimental binding free energies (c) and enthalpies (d). No entropy contribution is included in the computed binding free energies. The entropy contribution is included in the total computed binding free energies.

Conclusions

The structure prediction strategy using equilibrium sampling of the peptide conformational space by simulated tempering dynamics with the simplified, knowledge-based energy function has been shown to be adequate in generating the multitude of available low-energy peptide conformations. The following step of structural clustering and energy evaluation of the resulting conformations provides a reliable procedure for rapid screening of low-energy conformations and can recover the crystal structure of the pYEEI peptide. Structural features of the peptide−SH2 domains binding and the dominant topology of the bound peptides, with specific interaction in the pTyr pocket and hydrophobic interactions provided by the residues C-terminal to the pTyr group, are unambiguously reproduced in computational structure prediction analysis. The range of structural deviations for the predicted peptide conformations reveals a general trend for the Gly mutants to further deviate from the reference conformations generated by mutations in the crystallographic conformation of the pYEEI peptide. While the actual interactions in the specificity region C-terminal to the pTyr may significantly vary from one inhibitor to the other, the structural arrangement of the ligand−SH2 domain interface is conserved. The topology of the native binding mode tolerant to moderate structural perturbations may be an additional mechanism underlying permissiveness of the SH2 domain to the peptide sequence variations in the specificity region.

A synergy of the exhaustive sampling of the conformational space and detailed evaluation of the energetics is achieved using a hierarchy of the energy models. A binding free energy model, which includes the MM AMBER force field and a GB/SA solvation contribution, has been proposed to describe the energetics of the peptide binding. This model presents a modification of the MM/PBSA method, where the total binding free energy is computed by averaging the energy contributions over the conformations, structurally similar to the predicted peptide structure. Although the binding thermodynamics for a panel of studied peptides is predicted with the proposed approach in a better agreement with the experiment than previously suggested models, the overall correlation in binding affinity between theory and experiment is still modest. A considerable uncertainty in evaluating and interpreting the entropic contribution of the binding free energy is also present

in a more complete energetic model. The results of this study indicate that predicting small differences in the binding free energies for the Ala and Gly mutants of the recognition pYEEI peptide is considerably more difficult than to predict the structure of the bound peptides. Hence, despite considerable advances in the application of computational methods to ligand−protein binding, accurate prediction of binding affinities still remains a major methodological and technical challenge.

References

(1) Levitzki, A. Protein Tyrosine Kinase Inhibitors as Novel Therapeutic Agents. *Pharmacol. Ther.* **1999**, *82*, 231−239.

(2) Cohen, G. B.; Ren, R.; Baltimore, D. Modular Binding Domains in Signal Transduction Proteins. *Cell* **1995**, *80*, 237−248.

(3) Shokat, K. M. Tyrosine Kinases: Modular Signaling Enzymes with Tunable Specificities. *Chem. Biol.* **1995**, *2*, 509−514.

(4) Thomas, S. M.; Brugge, J. S. Cellular Functions Regulated by Src Family Kinases. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 513−609.

(5) Ottenhof-Kalff, A. E.; Rijksen, G.; Beurden, A. C. M. V.; Hennipman, M. P.; Michels, A. A.; Staal, G. E. Characterization of Protein Tyrosine Kinases from Human Breast Cancer: Involvement of the C−Src Oncogene Product. *Cancer Res.* **1992**, *49*, 4773.

(6) Weber, T. K.; Steele, G.; Summerhayes, I. C. Different pp60c-src Activity in Well and Poorly Differentiated Human Colon Carcinomas and Cell Lines. *J. Clin. Invest.* **1992**, *90*, 372.

(7) Soriano, P.; Montgomery, C.; Geske, R.; Bradley, A. Targeted Disruption of the c-src Proto-oncogene Leads to Osteoporosis in Mice. *Cell* **1991**, *64*, 693−702.

(8) Waksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B.; Cowburn, D.; Hanafusa, H.; Mayer, B. J.; Overduin, M.; Resh, M. D.; Rios, C. B.; Silverman, L.; Kuriyan, J. Crystal Structure of the Phosphotyrosine Recognition Domain SH2 of v-src Complexed with Tyrosine-Phosphorylated Peptides. *Nature* **1992**, *358*, 646−653.

(9) Waksman, G.; Shoelson, S. E.; 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.

(10) Eck, M.; Shoelson, S. E.; Harrison, S. C. Recognition of a High Affinity Phosphotyrosyl Peptide by the Src Homology 2 Domain of $p56^{lck}$. *Nature* **1993**, *362*, 87−91.

(11) Overduin, M.; Rios, C. B.; Mayer, B. J.; Baltimore, D.; Cowburn, D. Three-Dimensional Solution Structure of the Src Homology 2 Domain of c-Abl. *Cell* **1992**, *70*, 697−704.

(12) Bradshaw, J. M.; Grucza, R. A.; Ladbury, J. E.; 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.

(13) Bradshaw, J. M.; 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.

(14) Bradshaw, J. M.; 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.

(15) Bradshaw, J. M.; Mitaxov, V.; Waksman, G. Investigation of Phosphotyrosine Recognition by the SH2 Domain of the Src Kinase. *J. Mol. Biol.* **1999**, *293*, 971−985.

(16) Grucza, R. A.; Bradshaw, J. M.; 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.

(17) Bradshaw, J. M.; Mitaxov, V.; Waksman, G. Mutational Investigation of the Specificity Determining Region of the Src SH2 Domain. *J. Mol. Biol.* **2000**, *299*, 521−535.

(18) Grucza, R. A.; Bradshaw, J. M.; Mitaxov, V.; Waksman, G. Role of Electrostatic Interactions in SH2 Domain Recognition: Salt-Dependence of Tyrosyl-Phosphorylated 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.

(19) Clackson, T.; Wells, J. A Hot Spot of Binding energy in A Hormone-Receptor Interface. *Science* **1995**, *267*, 383−386.

(20) Chung, E.; Henriques, D. A.; Renzoni, D.; Zvelebil, M.; Bradshaw, J. M.; Waksman, G.; Ladbury, J. E. Mass Spectrometric and Thermodynamic Studies Reveal the Role of Water Molecules in Complexes Formed between SH2 Domains and Tyrosyl Phosphopeptides. *Structure* **1998**, *6*, 1141−1151.

(21) Pascal, S. M.; Yamazaki, T.; Singer, A. U.; Kay, L. E.; Forman-Kay, J. D. 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.

(22) Kay, L. E.; Muhandiram, D. R.; Farrow, N. A.; Aubin, Y.; Forman-Kay, J. D. Correlation between Dynamics and High Affinity Binding in an SH2 Domain Interaction. *Biochemistry* **1996**, *35*, 361−368.

(23) Kay, L. E.; Muhandiram, D. R.; Wolf, G.; Shoelson, S. E.; Forman-Kay, J. D. Correlation between Binding and Dynamics at SH2 domain Interfaces. *Nat. Struct. Biol.* **1998**, *5*, 156−163.

(24) Forman-Kay, Ji. D. The 'dynamics' in the Thermodynamics of Binding. *Nat. Struct. Biol.* **1999**, *6*, 1086−1087.

(25) Zhang, W.; Smithgall, T. E.; Gmeiner, W. H. Self-association and Backbone Dynamics of the HCk SH2 domain in the Free and Phosphopeptide-complexed Forms. *Biochemistry* **1998**, *37*, 7119−7126.

(26) Engen, J. R.; Gmeiner, W. H.; Smithgall, T. E.; Smith, D. L. Hydrogen Exchange shows Peptide Binding Stabilizes Motions in Hck SH2. *Biochemistry* **1999**, *38*, 8926−8935.

(27) Feng, M.-H.; Phillippopoulos, M.; MacKerell, A. D.; 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.

(28) Ladbury, J. E.; Hensmann, M.; Panayotou, G.; Campbell, I. D. Alternative Models of Tyrosyl Phosphopeptide Binding to a Src Family SH2 Domain: Implications for Regulation of Tyrosine Kinase Activity. *Biochemistry* **1996**, *35*, 11062−11069.

(29) Charifson, P. S.; Shewchuk, L. M.; Rocque, W.; Hummel, C. W.; Jordan, S. R.; Mohr, C.; Pacofsky, G. J.; Peel, M. R.; Rodriguez, M.; Sternbach, D. D.; Consler, T. G. Peptide Ligands of pp60-(c-src) SH2 Domains: A Thermodynamic and Structural Study. *Biochemistry* **1997**, *36*, 6283−6293.

(30) Plummer, M. S.; Holland, D. R.; Shahripour, A.; Lunney, E. A.; Fergus, J. H.; Marks, J. S.; McConnell, P.; Mueller, W. T.; Sawyer, T. K. Design, Synthesis, and Cocrystal Structure of a Nonpeptide Src SH2 Domain Ligand. *J. Med. Chem.* **1997**, *40*, 3719−3725.

(31) Lunney, E. A.; Para, K. S.; Rubin, J. R.; Humblet, C.; Fergus, J. H.; Marks, J. S.; Sawyer, T. K. 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.

(32) Pacofsky, G. J.; Lackey, K.; Alligood, K. J.; Berman, J.; Shewchuk, L. M.; Sternbach, D. D.; Rodriguez, M. Potent Dipeptide Inhibitors of the pp60c-src SH2 Domain. *J. Med. Chem.* **1998**, *41*, 1894−1908.

(33) Shakespeare, W. C.; Bohacek, R. S.; Azimioara, M. D.; Macek, K. J.; Luke, G. P.; Dalgarno, D. C.; Hatada, M. H.; Lu, X.; Violette, S. M.; Bartlett, C.; Sawyer, T. K. Structure-based Design of Novel Bicyclic Nonpeptide Inhibitors for the Src SH2 Domain. *J. Med. Chem.* **2000**, *43*, 3815−3819.

(34) Bohacek, R. S.; Dalgarno, D. C.; Hatada, M. H.; Jacobsen, V. A.; Lynch, B. A.; Macek, K. J.; Merry, T.; Metcalf, C. A., III; Narula, S. S.; Sawyer, T. K.; Shakespeare, W. C.; Violette, S. M.; Weigele, M. X-ray Structure of Citrate Bound to Src SH2 Leads to a High-affinity, Bone-targeted Src SH2 Inhibitor. *J. Med. Chem.* **2001**, *44*, 660−663.

(35) Shakespeare, W.; Yang, M.; Bohacek, R.; Cerasoli, F.; Stebbins, K.; Sundaramoorthi, R.; Azimioara, M.; Vu, C.; Pradeepan, S.; Metcalf, C. A., III; Haraldson, C.; Merry, T.; Dalgarno, D.; Narula, S.; Hatada, M.; Lu, X.; van Schravendijk, M. R.; 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 U.S.A.* **2000**, *97*, 9373−9378.

(36) Kollman, P. Free Energy Calculations-Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(37) Lamb, M. L.; Jorgensen, W. L. Computational Approaches to Molecular Recognition. *Curr. Opin. Chem. Biol.* **1997**, *1*, 449−457.

(38) Bash, P.; Singh, C.; Brown, F.; Langridge, R.; Kollman, P. A. Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex. *Science* **1987**, *235*, 574−576.

(39) Mark, A. E.; van Gunsteren, W. Decomposition of the Free Energy of A System in Terms of Specific Interactions. Implications For Theoretical and Experimental Studies. *J. Mol. Biol.* **1994**, *240*, 167−176.

(40) Smith, P. E.; van Gunsteren, W. When are Free Energy Components Meaningful? *J. Phys. Chem.* **1994**, *98*, 13735−13740.

(41) Boresh, S.; Archontis, C.; Karplus, M. Free Energy Simulations−The Meaning of Individual Contributions from a Component Analysis. *Proteins*: *Struct., Funct., Genet.* **1994**, *20*, 25−33.

(42) Boresh, S.; Karplus, M. The Meaning of Component Analysis: Decomposition of the Free Energy in Terms of Specific Interactions. *J. Mol. Biol.* **1995**, *254*, 801−807.
(43) Brady, G. P.; Sharp, K. A. Decomposition of Interaction Free Energies in Proteins and Other Complex Systems. *J. Mol. Biol.* **1995**, *254*, 77−85.
(44) Brady, G. P.; Szabo, A.; Sharp, K. A. On the Decomposition of Free Energies. *J. Mol. Biol.* **1996**, *263*, 123−125.
(45) Horton, N.; Lewis, M. Calculation of the Free Energy of Association For Protein Complexes. *Protein Sci.* **1992**, *1*, 169−181.
(46) Krystek, S.; Stouch, T.; Novotny, J. Affinity and Specificity of Serine Endopeptidase-Protein Inhibitor Interactions. Empirical Free Energy Calculations Based on X-ray Crystallographic Structures. *J. Mol. Biol.* **1993**, *234*, 661−679.
(47) Böhm, H.-J. The Development of A Simple Empirical Scoring Function to Estimate the Binding Constant For a Protein−Ligand Complex of Known Three-Dimensional Structure. *J. Comput.-Aided. Mol. Des.* **1994**, *8*, 243−256.
(48) Böhm, H.-J. Prediction of Binding Constants of Protein Ligands: A Fast Method For the Prioritization of Hits Obtained from De Novo Design or 3D Database Search Programs. *J. Comput.-Aided. Mol. Des.* **1998**, *12*, 309−323.
(49) Verkhivker, G. M.; Appelt, K.; Freer, S. T.; Villafranca, J. E. Empirical Free Energy Calculations of Ligand-Protein Crystallographic Complexes. I. Knowledge-Based Ligand-Protein Interaction Potentials Applied to the Prediction of HIV-1 Protease Binding Affinity. *Protein Eng.* **1995**, *8*, 677−691.
(50) Verkhivker, G. M. Empirical Free Energy Calculations of Ligand-Protein Crystallographic Complexes. II. Knowledge-Based Ligand-Protein Interaction Potentials Applied to the Thermodynamic Analysis of Hydrophobic Mutations. In *Pacific Symposium on Biocomputing-96*; Hunter, L., Klein, T., Eds.; Word Scientific: Singapore, 1996; pp 638−652.
(51) Wallqvist, A.; Jernigan, R. L.; Covell, D. G. A Preference-Based Free Energy Parametrization of Enzyme−Inhibitor Binding. Applications to HIV-1 Protease Inhibitor Design. *Protein Sci.* **1995**, *4*, 1881−1903.
(52) Jain, A. N. Scoring Noncovalent Protein−Ligand Interactions, A Continuous Differentiable Function Tuned to Compute Binding Affinities. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 427−440.
(53) DeWitte, R. S.; Shakhnovich, E. I. SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118*, 11733−11744.
(54) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function For Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(55) Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. Evaluation of PMF Scoring in Docking Weak Ligands to the FK506 Binding Protein. *J. Med. Chem.* **1999**, *42*, 2498−2503.
(56) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis For Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.
(57) Aqvist, J.; Medina, C.; Samuelsson, J. E. New Method For Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.
(58) Hansson, T.; Aqvist, J. Estimation of Binding Free Energies For HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137−1144.
(59) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method For Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667−10673.
(60) Aqvist, J.; Hansson, T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* **1996**, *100*, 9512−9521.
(61) Aqvist, J. Calculation of Absolute Binding Free Energies For Charged Ligands and Effects of Longe-Range Interactions. *J. Comput. Chem.* **1996**, *17*, 1587−1597.
(62) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived from Mannitol Synthesis, Inhibitory Potencies, and Computational Predictions of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885−897.
(63) Hansson, T.; Marelius, J.; Aqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(64) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding Affinities For Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Methodology. *J. Med. Chem.* **1997**, *40*, 1539−1549.
(65) Smith, R. H.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase Using Monte Carlo Simulations in a Linear Response Methodology *J. Med. Chem.* **1998**, *41*, 5272−5286.
(66) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities For HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 145−154.
(67) Wang, J.; Dixon, R.; Kollman, P. A. Ranking Binding Affinities with Avidin: A Molecular Dynamics-Based Interaction Energy Study. *Proteins*: *Struct., Funct., Genet.* **1999**, *34*, 69−81.
(68) Wang, W.; Wang, J.; Kollman, P. A. What Determines the van der Waals Coefficient β in the LIE (Linear Interaction Energy) Method to Estimate Binding Free Energies Using Molecular Dynamics Simulations? *Proteins*: *Struct., Funct., Genet.* **1999**, *34*, 395−402.
(69) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies Of The Stability of DNA, RNA and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(70) Massova, I.; Kollman, P. A. Computational Alanine Scanning to Probe Protein−Protein Interactions: A Novel Approach to Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(71) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. Molecular Dynamics and Free-Energy Calculations Applied to Affinity Maturation in Anti-body 48G7. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14330−14335.
(72) Kuhn, B.; Kollman, P. A. 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.
(73) Kuhn, B.; Kollman, P. A. 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.
(74) Lee, M. R.; Duan, Y.; Kollman, P. A. Use of MM-PB/SA in Estimating the Free Energies of Proteins: Application to Native, Intermediates, And Unfolded Villin Headpiece. *Proteins*: *Struct., Funct., Genet.* **2000**, *39*, 309−316.
(75) Wang, W.; Kollman, P. A. 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.
(76) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics And Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(77) Spolar, R. S.; Record, M. T., Jr. Coupling of Local Folding to Site-Specific Binding of Proteins to DNA. *Science* **1994**, *263*, 777−784.
(78) Janin, J. Elusive Affinities. *Proteins*: *Struct., Funct., Genet.* **1995**, *21*, 30−39.
(79) Weng, Z.; Vajda, S.; DeLisi, C. Prediction of Protein Complexes Using Empirical Free Energy Function. *Protein Sci.* **1996**, *5*, 614−626.
(80) Weng, Z.; Delisi, C.; Vajda, S. Empirical Free Energy Calculation: Comparison to Calorimetric Data. *Protein Sci.* **1997**, *6*, 1976−1984.
(81) Baker, B. M.; Murphy, K. P. Dissecting the Energetics of a Protein−protein Interaction. The Binding of Ovamucoid Third Domain to Elastase. *J. Mol. Biol.* **1997**, *268*, 557−569.
(82) Murphy, K. P.; Freire, E. Thermodynamics of Structural Stability and Cooperative Folding Behavior in Proteins. *Adv. Protein Chem.* **1992**, *43*, 313−326.
(83) Henriques, D. A.; Ladbury, J. E.; Jackson, R. M. Comparison of Binding Energies of Src SH2−Phosphotyrosyl Peptides with Structure-based Prediction Using Surface Area Based Empirical Parametrization. *Protein Sci.* **2000**, *9*, 1975−1985.
(84) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol.* **1995**, *2*, 317−324.
(85) Bouzida, D.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Rose, P. W.; Verkhivker, G. M. 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.
(86) Bouzida, D.; Rejto, P. A.; Verkhivker, G. M. Monte Carlo Simulations of Ligand-Protein Binding Energy Landscapes With the Weighted Histogram Analysis Methodol. *Int. J. Quantum Chem.* **1999**, *73*, 113−121.
(87) Verkhivker, G. M.; Rejto, P. A.; Bouzida, D.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Towards understanding the mechanisms of molecular recognition by computer simulations of ligand-protein interactions *J. Mol. Recognit.* **1999**, *12*, 371−389.

(88) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; 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.
(89) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(90) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel-An Integrated Software System for Modeling Organic and Bioorganic Molecules using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.
(91) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. 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.
(92) Weiser, J.; Weiser, A. A.; Shenkin, P. S.; Still, W. C. Neighbor-List Reduction: Optimization for Computation of Molecular van der Waals and Solvent-Accessible Surface Areas. *J. Comput. Chem.* **1998**, *19*, 797−808.
(93) Weiser, J.; Weiser, A. A.; Shenkin, P. S.; Still, W. C. Erratum: Neighbor-List Reduction: Optimization for Computation of Molecular van der Waals and Solvent-Accessible Surface Areas. *J. Comput. Chem.* **1998**, *19*, 1110.
(94) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217−230.
(95) Weiser, J.; Shenkin, P. S.; Still, W. C. Fast, Approximate Algorithm for Detection of Solvent-Inaccessible Atoms. *J. Comput. Chem.* **1999**, *20*, 586−596.
(96) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. DREIDING: A Generic Force Field for Molecular Simulation. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(97) Verkhivker, G. M.; Rejto, P. A.; Gehlhaar, D. K.; Freer, S. T. 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*: *Struct., Funct., Genet.* **1996**, *25*, 342−353.
(98) Rejto, P. A.; Verkhivker, G. M.; Gehlhaar, D. K.; Freer, S. T. New Trends in Computational Structure Prediction of Ligand-Protein Complexes for Receptor-Based Drug Design. In *Computational Simulation of Biomolecular Systems*; van Gunsteren, W., Weiner, P., Wilkinson, A. J., Eds.; ESCOM: Leiden, 1997; pp 451−465.
(99) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451−458.
(100) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method And Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1607.
(101) Hansmann, U. H. E.; Okamoto, Y. Monte Carlo Simulations in Generalized Ensemble: Multicanonical Algorithm Versus Simulated Tempering. *Phys. Rev. E* **1996**, *54*, 5863−5865.
(102) Hansmann, U. H. E.; Okamoto, Y. Generalized-Ensemble Monte Carlo Method for Systems with Rough Energy Landscape. *Phys. Rev. E* **1997**, *56*, 2228−2233.
(103) Hansmann, U. H. E.; Okamoto, Y. Numerical Comparisons of Three Recently Proposed Algorithms in the Protein Folding Problem. *J. Comput. Chem.* **1997**, *18*, 920−933.
(104) Hansmann, U. H. E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(105) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(106) Verkhivker, G. M.; Rejto, P. A.; Bouzida, D.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Navigating Ligand-Protein Binding Free Energy Landscapes: Universality And Diversity of Protein Folding And Molecular Recognition Mechanisms. *Chem. Phys. Lett.* **2001**, *336*, 495−503.
(107) Verkhivker, G. M.; Rejto, P. A.; Bouzida, D.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Gehlhaar, D. K.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Parallel Simulated Tempering Dynamics of Ligand-Protein Binding with Ensembles of Protein Conformations. *Chem. Phys. Lett.* **2001**, *337*, 181−189.
(108) Totrov, M.; Abagyan, R. Flexible Protein−Ligand Docking by Global Energy Optimization in Internal Coordinates. *Proteins*: *Struct., Funct., Genet.* **1997**, Suppl. 1, 215−220.
(109) Zhou, Y.; Abagyan, R. How and Why Phosphotyrosine-Containing Peptides Bind to the SH2 and PTB Domains. *Folding Des.* **1998**, *3*, 513−522.
(110) Given, J. A.; Gilson, M. K. A Hierarchical Method For Generating Low-Energy Conformers of A Protein−Ligand Complex. *Proteins*: *Struct., Funct., Genet.* **1998**, *33*, 475−495.
(111) Hoffman, D.; Kramer, B.; Washio, T.; Steinmetzer, T.; Rarey, M.; Lengauer, T. Two-Stage Method for Protein−Ligand Docking. *J. Med. Chem.* **1999**, *42*, 4422−4433.
(112) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. Deciphering Common Failures in Molecular Docking of Ligand-Protein Complexes. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 731−751.
(113) Tsui, V.; Case, D. A. Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.
(114) Bouzida, D.; Kumar, S.; Swendsen, R. H. Efficient Monte Carlo Methods for the Computer Simulation of Biological Molecules. *Phys. Rev. A* **1992**, *45*, 8894−8901.
(115) Willet, P.; Winterman, V. A Comparison of Some Measures for the Determination of Intermolecular Structural Similarity. *Quant. Struct.-Act. Relat. Pharmacol. Chem. Biol.* **1986**, *5*, 18−25.
(116) Willet, P.; Winterman, V.; Bawden, D. Implementation of Non-Hierarchic Cluster Analysis Methods in Chemical Information Systems: Selection of Compounds for Biological Testing and Clustering of Substructure Search Output. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 109−118.
(117) Bawden, D. Browsing and Clustering of Chemical Structures. In *Chemical Structures*: *The International Language of Chemistry*; Warr, W. A., Ed.; Springer-Verlag: Berlin, 1988; pp 145−150.
(118) Tidor, B.; Karplus, M. The contribution of Vibrational Entropy to Molecular Association. The dimerization of Insulin. *J. Mol. Biol.* **1994**, *238*, 405−414.
(119) Privalov, P. L.; Tamura, A. Comments on the Comments 'The Entropy Cost of Protein Association'. *Protein Eng.* **1999**, *12*, 187.
(120) Murphy, K. P.; Xie, D.; Thompson, K. S.; Amzel, L. P.; Freire, E. Entropy in Biological Binding Processes: Estimation of Translational Entropy Loss. *Proteins*: *Struct., Funct., Genet.* **1994**, *18*, 63−67.
(121) Amzel, L. M. Loss of Translational Entropy in Binding, Folding, and Catalysis. *Proteins*: *Struct., Funct., Genet.* **1997**, *28*, 144−149.
(122) Finkelstein, A. V.; Janin, J. The Price of Freedom: Entropy of Bimolecular Complex Formation. *Protein Eng.* **1989**, *3*, 1−3.
(123) Karplus, M.; Janin, J. Comment on: 'The Entropy Cost of Protein Association'. *Protein Eng.* **1999**, *12*, 185−186.
(124) D'Aquino, J. A.; Gomez, J.; Hilser, V. J.; Lee, K. H.; Amzel, L. M.; Freire, E. The Magnitude of the Backbone Conformational Entropy Change in Protein Folding. *Proteins*: *Struct., Funct., Genet.* **1996**, *25*, 143−156.
(125) Hilser, V. J.; Gomez, J.; Freire, E. The Enthalpy Change in Protein Folding and Binding: Refinement of Parameters for Structure-Based Calculations. *Proteins*: *Struct., Funct., Genet.* **1996**, *26*, 123−133.
(126) Bardi, J. S.; Luque, I.; Freire, E. Structure-Based Thermodynamic Analysis of HIV-1 Protease Inhibitors. *Biochemistry* **1997**, *36*, 6588−6596.

JM0101141