Journal of Computer-Aided Molecular Design, **14:** 731–751, 2000.
KLUWER/ESCOM

Deciphering common failures in molecular docking of ligand-protein complexes

Gennady M. Verkhivker, Djamal Bouzida, Daniel K. Gehlhaar, Paul A. Rejto, Sandra Arthurs, Anthony B. Colson, Stephan T. Freer, Veda Larson, Brock A. Luty, Tami Marrone & Peter W. Rose

Agouron Pharmaceuticals, Inc., A Warner-Lambert Company, 10777 Science Center Drive, San Diego, CA 92121-1111, U.S.A.

Received 5 January 2000; Accepted 28 April 2000

Key words: binding free energy profile, clustering, ligand-protein docking, molecular recognition, Monte Carlo simulations, structural similarity

Summary

Common failures in predicting crystal structures of ligand-protein complexes are investigated for three ligand-protein systems by a combined thermodynamic and kinetic analysis of the binding energy landscapes. Misdocked predictions in ligand-protein docking are classified as 'soft' and 'hard' failures. While a soft failure arises when the search algorithm is unable to find the global energy minimum corresponding to the crystal structure, a hard failure results from a flaw of the energy function to qualify the crystal structure as the predicted lowest energy conformation in docking simulations. We find that neither the determination of a single structure with the lowest energy nor finding the most common binding mode is sufficient to predict crystal structures of the complexes, which belong to the category of hard failures. In a proposed hierarchical approach, structural similarity clustering of the conformations, generated from equilibrium simulations with the simplified energy function, is followed by energy refinement with the AMBER force field. This protocol, that involves a hierarchy of energy functions, resolves some common failures in ligand-protein docking and detects crystallographic binding modes that were not found during docking simulations.

Introduction

Computational structure prediction of ligand-protein complexes using docking simulations has greatly advanced an understanding of the molecular recognition phenomenon and has become an important tool in drug discovery by facilitating structure-based ligand design [1–11]. Docking simulations require the energy of the ligand-protein complex crystal structure to be the global minimum on the binding energy landscape, that represents a thermodynamic condition on the employed in simulations energy function. The requirement to determine consistently and rapidly the global free energy minimum addresses a kinetic aspect of the docking problem. Synergy of energetic models based on either surface complementarity [12–19] or atom–atom representations of the intermolecular interactions [20–25], combined with stochastic optimization techniques [26–33], has led to powerful structure prediction strategies, which have expanded the scope of the docking problem. Recent advances in computational structure prediction of ligand-protein complexes incorporate hierarchical approaches by employing Monte Carlo minimization simulations in flexible binding sites [34, 35] and molecular dynamics docking simulations with flexible receptors [36], utilizing rotamer libraries of side-chains combined with the dead-end elimination (DEE) algorithm to optimize protein side-chains [37–39], and including explicit protein flexibility [40–43]. Current applications of

*To whom correspondence should be addressed. E-mail: verk@agouron.com

flexible ligand docking range from simulations with ensembles of multiple ligands [44] and multiple protein conformations [45–47] to analysis of the binding energy landscapes [48], lead discovery [49, 50], database mining [51], and structure-based combinatorial ligand design [52].

Docking simulations usually determine a single structure of the complex with the lowest energy and postulate that the lowest energy conformation corresponds to the native structure. The number of low-energy structures is usually very large and a computationally demanding task of finding the lowest energy structure does not imply its thermodynamic stability. Nevertheless, the structure prediction problem implies determination of the ensemble of many similar conformations, which describe the thermodynamically stable native basin of the global energy minimum, rather than a single structure. The conjecture, that there are more low energy conformations surrounding the native state than non-native minima, was used to recognize near-native protein structures in ensembles of misfolded decoys [53]. It was suggested that uniform sampling of the conformational space with the low resolution energy function followed by identification of the largest cluster of structurally related low-energy conformations may be more efficient in finding the region of the conformational space that contains the native structure than the energy-based criteria [53]. In ligand-protein binding, we have introduced a structural consensus approach that is focused on the size of the low-energy clusters predicted in docking simulations and searched for a common topology of the binding mode, rather than a single lowest energy structure [54].

Misdocked predictions in computer simulations of ligand-protein docking can be categorized as soft and hard failures. A soft failure is due to a flaw in the search algorithm, which is unable to locate the global energy minimum corresponding to the crystal structure, and is defined as the case when the energy of the crystal structure, after minimization with the chosen force field, is lower than the energy of the lowest energy conformation predicted in docking simulations. A hard failure results from inaccuracy in reproducing differences in the relative energies of alternate binding modes and arises when the global energy minimum corresponds to a misdocked structure with the energy lower than the energy of the minimized crystal structure. Comparing the results of docking experiments performed on a large validation set of Protein Data Bank (PDB) ligand-protein complexes using the GOLD program [31] and the AGDOCK approach [21, 22], we have detected a relatively small number of examples, where both methods have been unable to predict the crystal structure of the complex because of a hard failure in docking simulations. A simplified energy function, used in validation experiments with the AGDOCK approach, was originally developed to satisfy both thermodynamic and kinetic requirements in docking simulations by reducing frustration of the underlying binding energy landscape [21, 22, 55]. Robust structure prediction of flexible ligands given a fixed conformation of the native protein was achieved with this energy function by generating binding energy landscapes with co-existing correlated, funnel-like and uncorrelated, rugged features. The simplified energy function does not have singularities at interatomic distances and can effectively explore accessible ligand binding modes by sampling a large fraction of conformational space, particularly at high temperature. While adequate for non-polar and hydrogen bond patterns, this simplified energy model does not include a direct electrostatic component and therefore can be less reliable when extensive networks of electrostatic interactions are present in the crystal structures. By contrast, the GOLD method employs a template of protein hydrogen bond donors and acceptors to sample intermolecular hydrogen bonds networks and ligand conformations, but lacks a desolvation component and is more susceptible to failures in predicting hydrophobic interactions.

In this work, we propose a hierarchical approach to investigate three ligand-protein complexes that belong to the category of common hard failures, found in docking simulations with both GOLD and AGDOCK approaches. We employ a protocol based on structural similarity clustering of the conformations, generated from equilibrium simulations with the simplified energy function, that is followed by energy refinement with the AMBER force field. This strategy, that involves a hierarchy of energy functions, resolves some common failures in ligand-protein docking and detects crystallographic binding modes that were not found during docking simulations.

Materials and methods

Molecular recognition energy models

We have pursued a ‘plug-and-play’ strategy with two different energy functions, a simplified energy function and a molecular mechanics force field in conjunction with two sampling techniques, evolutionary

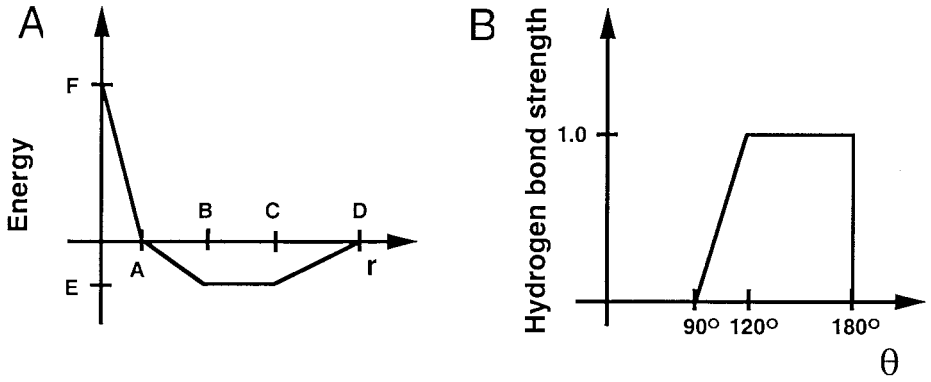

Figure 1. (A) The functional form of the ligand-protein interaction energy. For steric interactions, A = 0.93B, C = 1.25B, D = 1.5B, E = −0.4, F = 15.0, and B = $r_l + r_p$ is the sum of the atomic radii for the ligand and protein atoms. For hydrogen bond interactions, A = 2.3, B = 2.6, C = 3.1, D = 3.4, E = −4.0, F = 15.0. For sulfur hydrogen bond interactions, A = 2.7, B = 30.0, C = 3.5, D = 3.8, E = −2.0, F = 15.0. For chelating interactions with the metals A = 1.5, B = 1.7, C = 2.5, D = 3.0, E = −10.0, F = 15.0. For repulsive interactions, A = 3.2, E = 0.1, F = 15.0, and B, C, and D are not relevant. The units of A, B, C, and D are Å; for E and F the units are kcal/mol. (B) The hydrogen bond interaction energy is multiplied by the hydrogen bond strength term, which is a function of the angle θ determined by the relative orientation of the protein and ligand atoms.

programming [21] and Monte Carlo simulations [46, 48, 56–58]. The knowledge-based simplified energetic model includes intramolecular energy terms for the ligand, given by torsional and nonbonded functions [59], and intermolecular ligand-protein steric and hydrogen bond interaction terms calculated from a piecewise linear potential (PL) summed over all protein and ligand heavy atoms (Figure 1a). 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, fluorine-sized nonpolar and sulfur-sized nonpolar. We have added an additional iodine-sized nonpolar atom type in simulations with the ligand-protein complex that contains iodines. The atomic radius for carbon, oxygen, nitrogen atoms is 1.8 Å, for fluorine it is 1.8 Å, for sulfur 2.2 Å and 2.35 Å for iodine. A multiplicative desolvation penalty of 1.0 is applied to the attractive portion of the interaction between non-polar 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 allows for sulfur-donor closer contacts that are seen in some of the crystal structures. Crystallographic water molecules and hydroxyl groups are defined to be both donor and acceptor, and carbon atoms are defined to be nonpolar. The steric and hydrogen bond-like 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. The parameters were refined to yield the experimental crystallographic structure of a set of ligand-protein complexes as the global energy minimum [21, 22]. No assumptions regarding either favorable ligand conformations or any specific ligand-protein interactions were made, and all buried crystallographic water molecules are included in the simulations as part of the protein structure. The standard AMBER force field [60, 61] is used in conjunction with a solvation term [46, 48, 62], which is added to the interaction potential to account for the free energy of interactions between the explicitly modeled atoms of the ligand-protein system and the implicitly modeled solvent. The term was derived by

considering the transfer of atom, from an environment where it is completely surrounded by solvent, to an environment in which it has explicit atomic neighbors [62].

Computer simulations of ligand-protein binding

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 5.0 Å cushion added to every side of this box. Bonds allowed to rotate include those linking sp^3 hybridized atoms to either sp^3 or sp^2 hybridized atoms and single bonds linking two sp^2 hybridized atoms. The initial ligand bond lengths, bond angles, and the torsional angles of the unrotated bonds were obtained from the crystal structures of the bound ligand-protein complexes. Crystallographically determined ligand structures may have somewhat distorted valence angles, nitrogen geometry, and in some cases, distorted ring geometry which could affect the conformational search. The structures of the studied ligands were minimized and MNDO atomic charges were calculated using the MOPAC program [63, 64]. Crystallographic buried water molecules are included in the simulations as part of the protein structure.

Evolutionary algorithm, a stochastic optimization technique based on the ideas of natural selection, was used in ligand-protein docking simulations. During the search, a population of candidate ligand conformers competes for survival against a fixed number of opponents randomly selected from the remainder of the population. A win is assigned to the competitor with the lowest energy and the number of competitions that a member wins determines the survival probability to the next generation [21]. All surviving members produce offspring, subject to a constant population size. For each docking simulation, the evolutionary search was performed for a total of 120 generations with a population size of 1200 members. The minimized best member of the final generation defines the predicted structure for the ligand-protein complex.

We have carried out equilibrium ligand-protein simulations with a fixed bound protein conformation using parallel simulated tempering dynamics [65–70] with 50 replicas of the ligand-protein system attributed respectively to 50 different temperature levels that are uniformly distributed in the range between 5300 K 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 mth and nth replicas, described by a common Hamiltonian $H(X)$, are associated with the inverse temperatures β_m and β_n, and the corresponding conformations X_m and X_n. The exchange of conformations between adjacent replicas m and n is accepted or rejected according to a Metropolis criterion with the probability $p = \min(1, \exp[-\delta])$ where $\delta = [\beta_n - \beta_m][H(X_m) - H(X_n)]$ [66]. 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 enhances conformational sampling at low temperatures on rough energy landscapes and permits regions with a small density of states to be sampled more accurately. During the course of the simulation, each replica has a non-negligible 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 [65–70].

During Monte Carlo moves, performed independently for each replica of the ligand-protein system, we employ the dynamically optimized acceptance ratio method [71]. The maximum step sizes at each temperature are dynamically chosen in this procedure to optimize the acceptance ratio, which is the ratio of accepted conformations to the total number of trial conformations. We update the maximum step sizes every cycle of 1000 sweeps, and store both the energy and the coordinates of the system at the end of each cycle. For all these simulations, we equilibrated the system for 1000 cycles (or one million sweeps), and collected data during 10 000 cycles (or ten million sweeps) resulting in 10 000 samples at each temperature. A sweep is defined as a single trial move for each degree of freedom of the system.

The weighted histogram analysis method

The multiple histogram method [72–74] optimally combines simulation data obtained at many discrete

temperatures to provide an improved estimate of the density of states, which can then be used over a range of continuous temperatures. A generalization of the multiple histogram method, the weighted histogram analysis method (WHAM), estimates the density of states from data collected using umbrella sampling [72–74].

In this work, we apply the weighted histogram analysis method to compute ligand-protein binding energy landscapes, $F(R, T)$, as a continuous function of temperature and reaction coordinate. They are determined by first tabulating two-dimensional histograms $H_i(E, R)$ from the various constant-temperature equilibrium simulations i, and then solving the self-consistent multiple histogram equations [72] to yield the density of states

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

where

$$\exp[-F_j/k_B T_j] = \sum_E W(E) \exp[-E/k_B T_j], \quad (2)$$

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

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

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

$$W(E) = \sum_R W(E, R) = \frac{\sum_{i=1}^{M} g_i{}^{-1} H_i(E)}{\sum_{j=1}^{M} g_j{}^{-1} n_j \exp[-(E - F_j)/k_B T_j]}, \quad (3)$$

where

$$H_i(E) = \sum_R H_i(E, R), \quad (4)$$

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

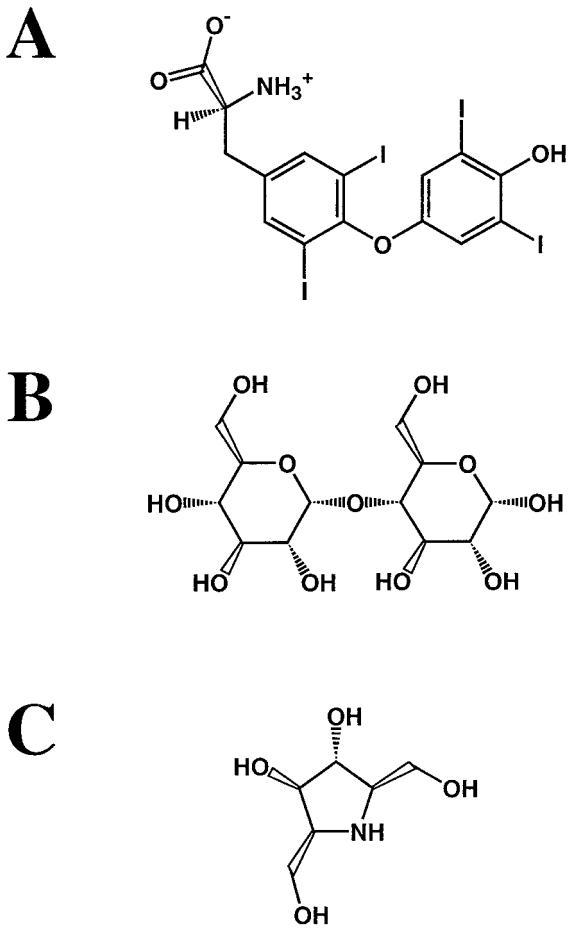

Figure 2. The structures of (A) thyroxine ligand (3,5,3′,5′-tetraiodo-L-thyronine), (B) maltose, (C) 2,5-dideoxy-2,5-imino-D-glucitol.

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

$$F(R, T) = -k_B T \ln[P(R, T)/P(R_c, T)], \quad (5)$$

where

$$P(R, T) = \sum_E P_T(E, R) \quad (6)$$

$$P_T(E, R) = W(E, R) \exp[-E/k_B T]. \quad (7)$$

We define R to be the root mean square deviation (RMSD) of the ligand coordinates from the native

state, calculated between ligand heavy atoms, and the native state is chosen to be the reference state, so $R_c = 0.0$. Minimization of the RMSD between corresponding heavy atoms in two structures is performed using a technique that is based on quaternions [75]. In order to ensure that the minimal RMSD is computed for structures with internal symmetry, all possible mappings between corresponding atoms are determined [76] and the lowest RMSD from this set of mappings is reported.

Similarity clustering

The 3D-similarity calculations are based on the spatial proximity of atoms in a binding site and the atom type. We distinguish four types of atoms: 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 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 $p(i, j) = 10^{(-r_{i,j}^2/\sigma^2)}$, where r_{ij} is the distance between atoms i and j, and $\sigma = -c^2/\log_{10}(p)$, 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 $p = 0.000032$.

We calculate a descriptor $d(i, j)$ from the spatial proximity and the atom type compatibility:

$$d(i, j) = p(i, j) * a(i, j) \quad \text{if} \quad r(i, j) \leq c,$$
$$d(i, j) = 0 \quad \text{if} \quad r(i, j) > c.$$

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 [77]:

$$S(m, n) = \frac{\sum_{i=1}^{M} D_m^m(i) D_m^n(i) + \sum_{j=1}^{N} D_n^m(j) D_n^n(j)}{\sum_{i=1}^{M} D_m^m(i)^2 + \sum_{j=1}^{N} D_n^n(j)^2 - \sum_{i=1}^{M} D_m^m(i) D_m^n(i) - \sum_{j=1}^{N} D_n^m(j) D_n^n(j)}$$

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 ofthe a cluster is called the cluster leader. After all molecules are assigned to clusters, the molecules are arranged in 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 [78] converges to a minimum. We analyze clusters with at least 100 members. Since conformations which belong to the same cluster are equivalent with 85% structural similarity, different clusters are compared by analyzing cluster leaders.

Results

We investigate the dynamics and thermodynamics of molecular recognition for three ligand-protein systems that represent hard failures in ligand-protein docking (Figure 2).

Transthyretin-thyroxine complex

The first studied complex is the transthyretin protein with a 1:1 mixture of Val and Met residues at position 30 bound with the thyroxine ligand (3,5,3′,5′-tetraiodo-L-thyronine) (pdb entry 1eta) [79, 80]. Docking simulations with the PL energy function result in two distinct binding modes: a native-like binding domain at RMSD = 2–3 Å from the crystal structure and an alternative binding mode at RMSD = 8–9 Å from the native state (Figure 3A). The lowest energy solution determined during docking simulations belongs to the alternative binding domain, located at RMSD = 8.97 Å from the native state (Figure 3B). In docking simulations with AMBER, we have observed a broad distribution of low energy states at RMSD = 5–8 Å from the crystal structure (Figure 3C) and the lowest energy conformation is located at RMSD = 6.74 Å from the native state (Figure 3D, Table 1). Hence, both energy functions predict an incorrect structure for this complex. The binding energy landscape constructed with the PL energy function has a broad funnel of conformations leading to a shallow minimum near RMSD = 8.0 Å from the native structure (Figure 4A). However, the lowest energy

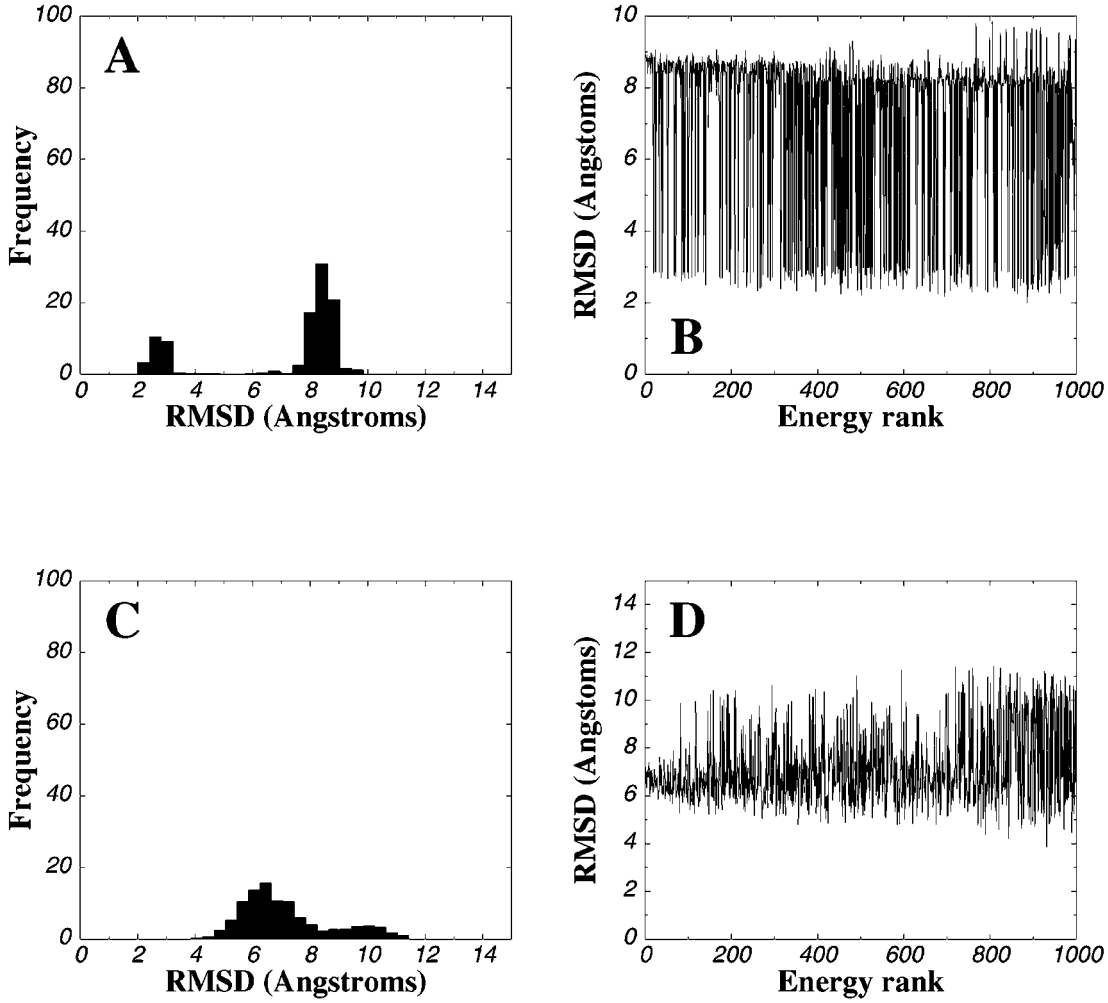

Figure 3. (A) The frequency of predicting binding modes of the 1eta ligand-protein complex with the piecewise linear energy function and (B) the RMSD of the docked conformations from the crystal structure ranked by energy. (C) The frequency of predicting binding modes of the 1eta ligand-protein complex with the AMBER force field and (D) the docked conformations as a function of RMSD from the crystal structure ranked by energy.

structure determined from equilibrium simulations belongs to the native-like binding mode with RMSD = 2.83 Å from the native structure. The failure of ligand-protein docking for this system is seen to result in part from incomplete conformational sampling because the structure predicted from equilibrium simulations has a lower energy than the one found in docking simulations. Thermodynamic analysis helps to explain this hard failure in structure prediction by complementing the results of kinetic docking simulations. We find that the hard failure in ligand-protein docking for this system may result in the existence of a conformational funnel leading to the misdocked binding mode which is, nevertheless, only a meta-stable local minimum at low temperatures. The binding free energy profile constructed with the AMBER energy function is flat between 4.0 and 10.0 Å RMSD from the crystal structure, with several marginally stable binding modes at low temperatures (Figure 4B).

The true binding free energy landscape is multi-dimensional, whereas the binding free energy profile that we analyze is only a one-dimensional projection of this surface onto a single coordinate. One of the disadvantages of using RMSD as a reaction coordinate for the ligand-protein binding process is that it is difficult to distinguish energy minima that are distant from the crystal structure. For example, it is not possible to unambiguously conclude whether a single low-energy

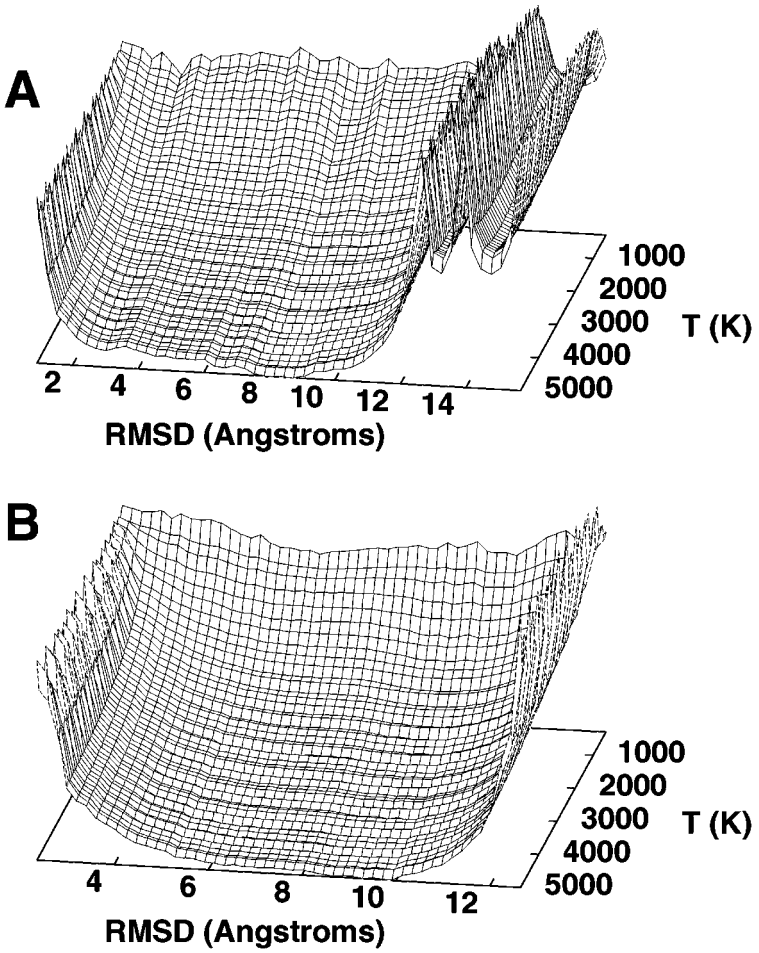

Figure 4. The binding free energy landscape of the 1eta complex with (A) the piecewise linear energy function and (B) the AMBER force field. For each two-dimensional temperature slice, the reference energy $F(R = 0, T)$ is defined to be zero.

binding is located at RMSD = 8.0 Å from the crystal, or there are different binding modes at this same distance. A convenient method to resolve this complication is to generate clusters of structurally similar conformations, since two distinct binding modes will result in two different conformational clusters. The largest cluster of conformations generated with the PL energy function from equilibrium simulations in the temperature range between 300 K and 5000 K is centered near 3.0 Å RMSD from the crystal structure, and has the native-like topology (Figure 5A). It is also apparent that there are two distinct clusters of states located approximately at 8.0 Å and 9.0 Å RMSD from the crystal structure. After minimization with the AMBER energy function the lowest energy structure resides at RMSD = 2.0 Å from the native state (Figure 5B, Table 1). There are no stable near-native structures with the AMBER energy function and the largest cluster of conformations generated on the AMBER energy landscape is centered near RMSD = 6–7 Å from the crystal structure. There is not a single cluster in the native binding domain (Figure 5C), a clear case when the PL energy function is more robust than AMBER in describing the low-energy binding modes of the ligand-protein complex.

Examination of the misdocked conformation reveals a subtle interplay between van der Waals and electrostatic interactions. In the crystal structure of the 1eta complex, N-8 of the ligand forms a hydrogen bond with the Glu-154 residue of the protein. In the alternative binding mode, the ligand flips and forms two hydrogen bonds with Leu-110 and Ser-117 (Figure 6). The mobility of Ser-117 has been implicated in reducing the binding site cavity and tighter binding with the

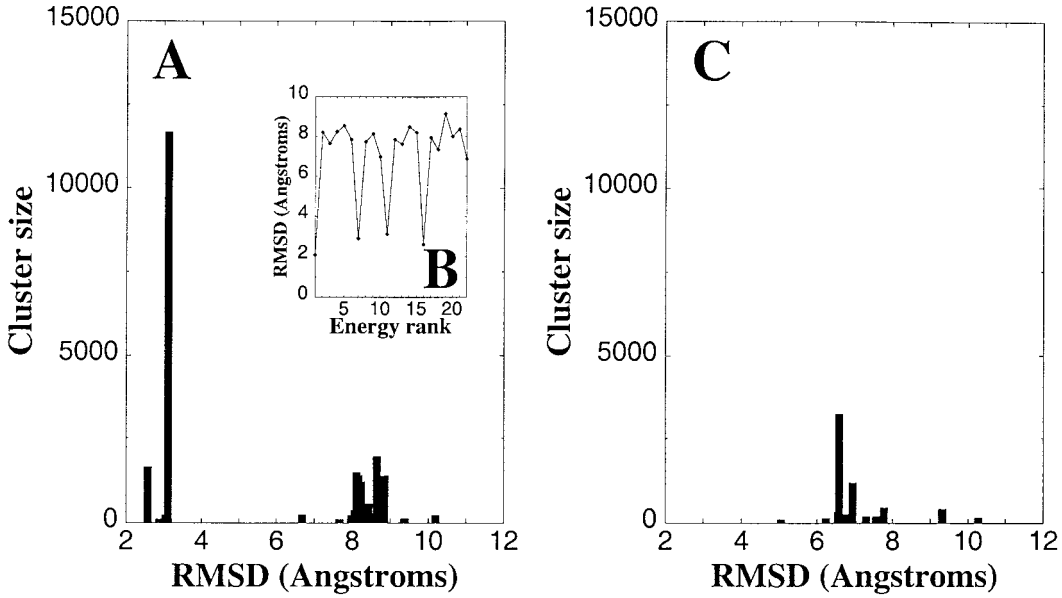

Figure 5. (A) The size of structurally similar clusters for the 1eta ligand-protein complex with the piecewise linear energy function as a function of RMSD from the crystal, and (B) the RMSD of the conformations minimized with the AMBER energy function, ranked by energy. (C) The size of structurally similar clusters for the 1eta ligand-protein complex with the AMBER energy function as a function of RMSD from the crystal.

Thr-109 mutant protein. The analysis of the thyroxine-binding sites in the Thr-109 substituted protein with the Met-30 residue has suggested that there is a variation in binding affinity for thyroxine between three different mutants of the transthyretin protein which may arise from differences in the size of the binding pocket [96, 97]. It was crystallographically determined that the shift in the position of the Ser-117 residue in the Thr-109 mutant complex results in a smaller binding cavity and possibly is the major factor contributing to the increase in the thyroxine ligand affinity with this protein variant. It is possible that subtle differences in stability of the alternative binding modes, found in the native wild-type complex with Ala-109 (Figure 6), are regulated by mutations at this position.

Cyclodextrin glycosyltransferase-maltose complex

The second hard failure examined is a complex of cyclodextrin glycosyltransferase with maltose bound molecule in the domain C (pdb entry 1cdg) [81]. Docking simulations using either the PL or the AMBER energy function fail to predict the crystallographic binding mode of the complex as the lowest energy structure. The native binding mode was located in only a small fraction of docking simulations (Figures 7A, 7C). The lowest energy structure determined from docking simulations with the PL energy function is located at RMSD = 6.1 Å from the crystal structure (Figure 7B). The spectrum of low-energy docking solutions consists primarily of the conformations which belong to the misdocked binding mode and the conformations with the native-like binding mode are congregated at the tail of the spectrum (Figure 7B). The lowest energy structure predicted with the AMBER energy function belongs to the same misdocked binding mode and is located at RMSD = 6.28 Å from the crystal structure (Figure 7D, Table 1).

The binding free energy profile constructed with the PL energy function has at least four different binding domains, one of which is within 2.0 Å RMSD from the crystal structure (Figure 8A). Nevertheless, the misdocked binding mode at RMSD = 6 Å from the crystal structure dominates thermodynamic equilibrium at the entire temperature range and is more favorable than the crystallographic conformation. At high temperature, when the system can efficiently explore the conformational space, a broad basin in the region between 4 Å and 6 Å RMSD from the crystal structure is more stable and contributes to the thermodynamic equilibrium (Figure 8A). The transition to the well-defined misdocked binding mode occurs only at lower temperatures. The region on the binding energy landscape directed to the native structure

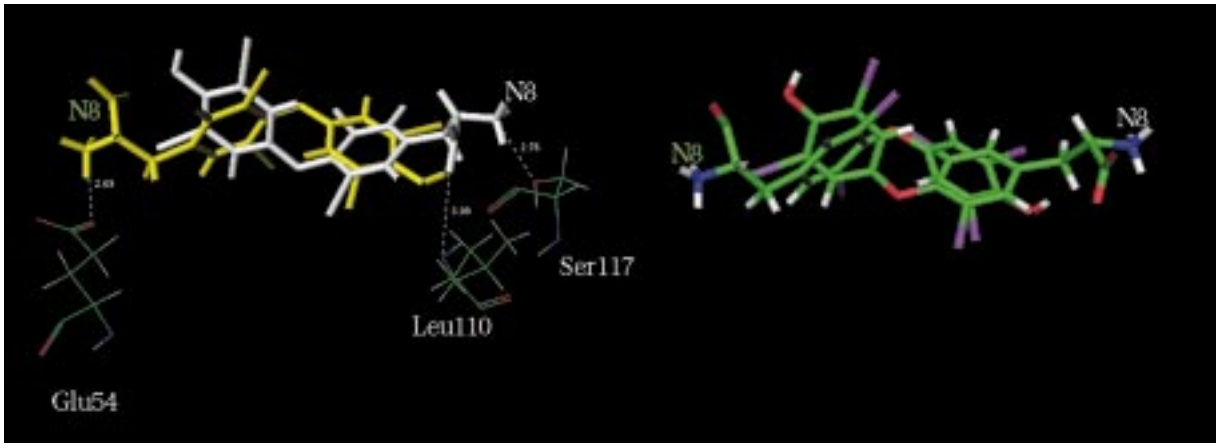

Figure 6. Left panel: Superposition of the crystal structure of the 1eta ligand-protein complex (yellow) with the lowest energy conformation (gray) obtained from docking simulations, which defines its predicted structure. Right panel: Superimposed inhibitor structures color coded by atom type.

Table 1. The RMSD values (Å) from the crystal structure for the lowest energy conformations obtained in docking, equilibrium simulations and with the structural clustering/minimization procedure

PDB entry	RMSD of the lowest energy docked conformation (PL)	RMSD of the lowest energy docked conformation (AMBER)	RMSD of the lowest energy equilibrium conformation (PL)	RMSD of the lowest energy minimized conformation after clustering
1eta	8.97	6.74	2.83	2.07
1cdg	6.15	6.28	6.19	0.78
1did	4.53	3.55	4.15	4.10

becomes narrow as temperature decreases, and the native binding mode is represented by an isolated region of conformations in the close vicinity of the crystal structure. Moreover, a significant energy barrier, that emerges at lower temperature, separates this region from the rest of the conformational space (Figure 8A). While there is a narrow, though well-defined funnel of conformations that leads to the native structure, the alternative binding modes are thermodynamically stable at all temperatures and this promotes the consistent acquisition of the misdocked binding mode in docking simulations. Structural clustering of the conformations, generated from equilibrium simulations with the PL energy function, produced the largest size clusters located at RMSD = 5 Å and RMSD = 6 Å from the crystal structure (Figure 9A). Only a relatively small cluster of native conformations was detected at RMSD = 1.0 Å from the native state, reflecting a narrow region on the binding energy landscape in the proximity of the crystal structure. However, energy minimization with the AMBER energy function of the cluster leaders yields the lowest energy structure located at RMSD = 0.78 Å from the crystal structure (Figure 9B, Table 1). Hence, using a two-step protocol of first identifying clusters of structurally similar conformations generated from equilibrium sampling with the PL energy function, followed by minimization of the corresponding cluster leaders with the AMBER force field, resolves the hard failure in ligand-protein docking and results in structure prediction consistent with the crystallographic binding mode.

The binding energy landscape generated with the AMBER energy function is rather flat with a number of shallow meta-stable local minima separated by small barriers. At higher temperatures there is a broad funnel of conformations in the region between RMSD

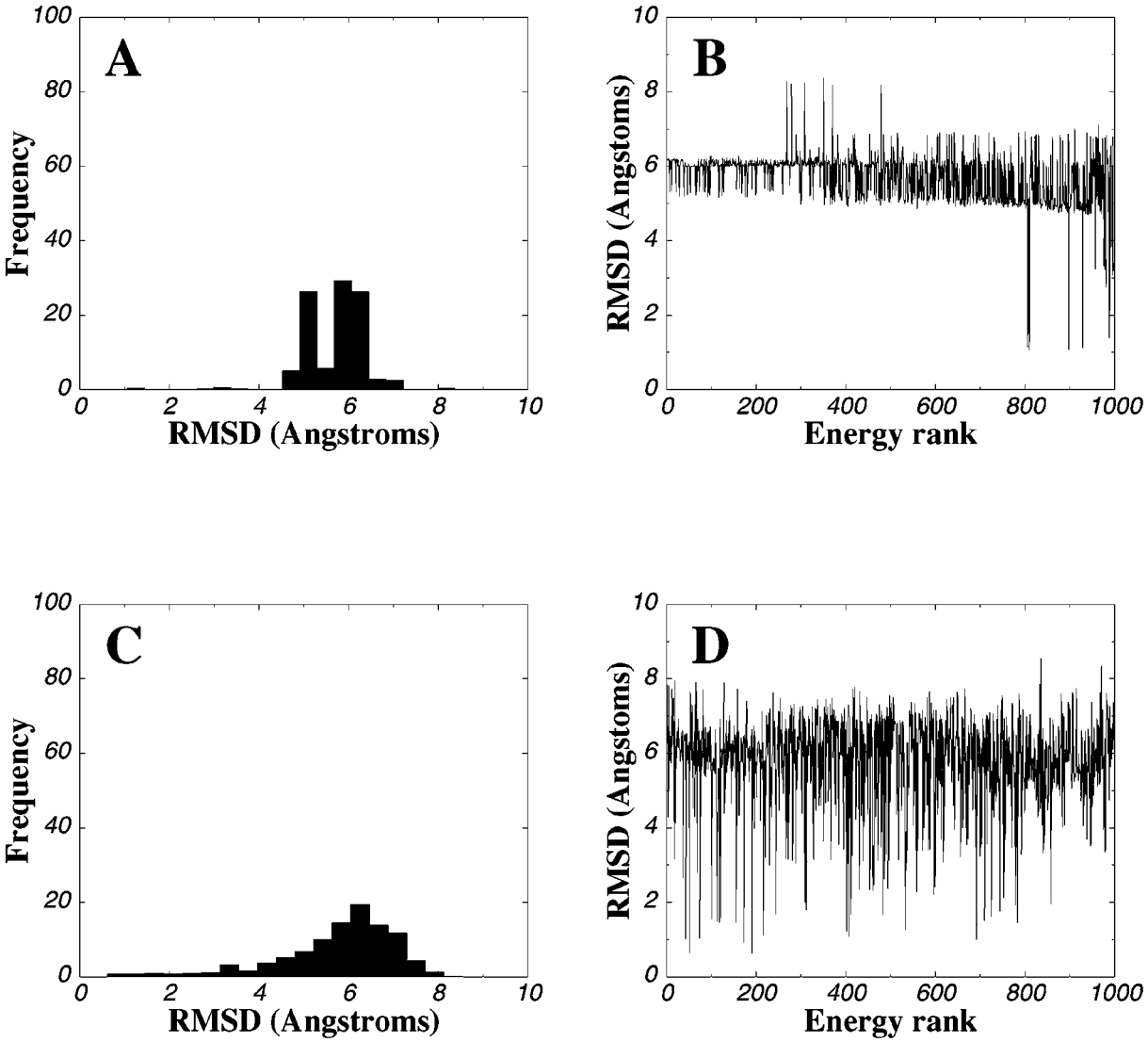

Figure 7. (A, C) The frequency of predicting binding modes of the 1cdg ligand-protein complex; and (B, D) the RMSD of the docked conformations from the crystal structure ranked by energy. See the caption of Figure 3 for more details.

= 4 Å and RMSD = 8 Å from the native structure which contributes significantly to the thermodynamic equilibrium (Figure 8B). However, the relative stability of the binding modes changes as the temperature lowers and the binding modes at RMSD = 3 Å and RMSD = 7 Å from the crystal structure start to dominate thermodynamic equilibrium at lower temperatures. Interestingly, most of the low docking solutions determined with the AMBER energy function belong to the region of the conformational space that is more stable at higher temperatures, and only a very small fraction of the predicted conformations populate regions that become more stable at lower temperatures. There are no clusters of structurally similar conformations sampled with AMBER that are centered closer than 4.0 Å RMSD from the crystal structure (Figure 9C). The hard failure in docking with the AMBER energy function for the 1cdg ligand-protein system results from a number of misdocked frustrated binding modes on a flat binding energy landscape with no significant barriers.

The bound conformation of the ligand in the crystal structure interacts by its aromatic portion of the reducing sugar with the aromatic side-chain of Trp-413. The bound maltose molecule makes five direct hydrogen bonds to the protein. The carboxyl oxygen

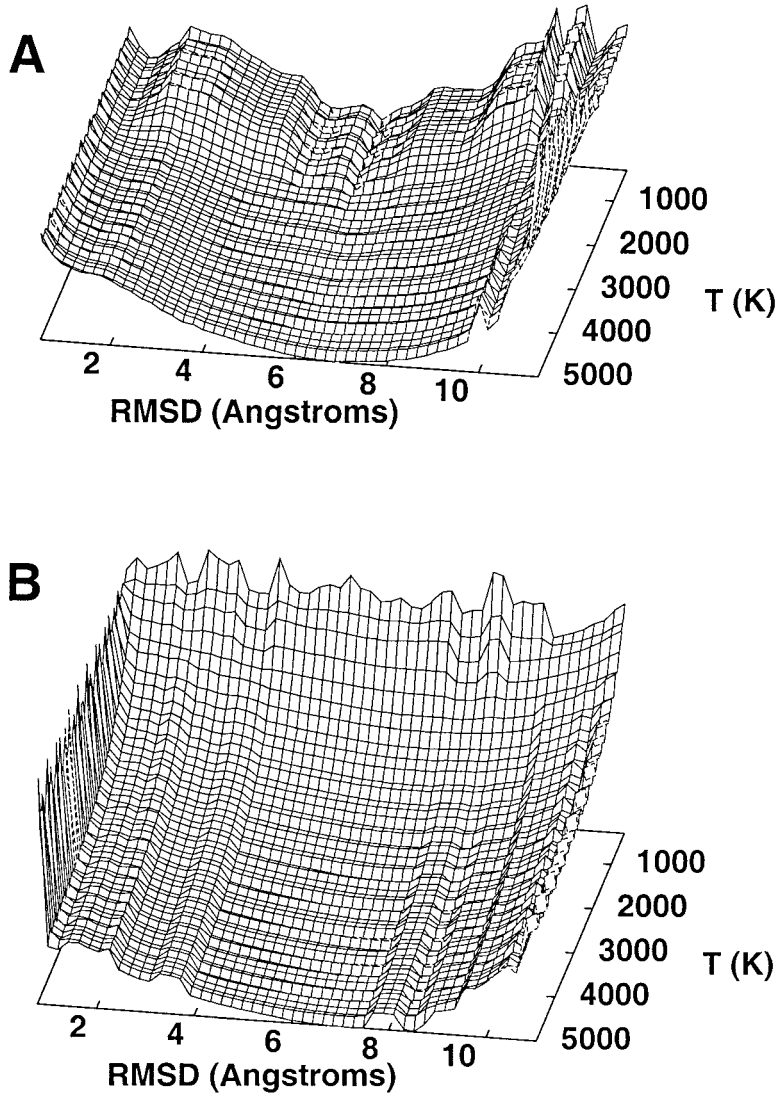

Figure 8. The binding free energy landscape of the 1cdg complex. See the caption of Figure 4 for more details.

and NH-group of Ile-414 form hydrogen bonds with O-2′ and O-3′ atoms of the ligand, respectively. In addition, the carboxyl oxygen of Glu-411 interacts with O-2 and the carboxyl oxygen of Gly-446 forms a hydrogen bond with O-6 of the ligand (Figure 10). In the crystal structure there is also an additional hydrogen bond with the carbonyl oxygen of the symmetry related residue Asp-540 that is not shown. The most populated binding mode, the largest cluster of structurally similar conformations generated with the PL energy function interacts with the same key protein residues, including Glu-411 and Ile-414, as the crystal structure. This binding mode has different networks of hydrogen bonds and, moreover, forms more hydrogen bonds than in the crystal structure. The O-3 and O-3′ atoms of the maltose bound molecule form two hydrogen bonds with OE2 of Glu-411 and O-2′ interacts with the carboxyl oxygen of Glu-411. Other hydrogen bonds are formed with the carboxyl oxygen of Ile-414, nitrogens of the Trp-413 and Arg-412 side-chains (Figure 10).

D-xylose isomerase-D-glucitol complex

The third example investigated is a complex of D-xylose isomerase, an enzyme dependent on a divalent ion for catalytic activity [82], with 2,5-dideoxy-2,5-imino-D-glucitol (pdb entry 1did). Two metal ions are

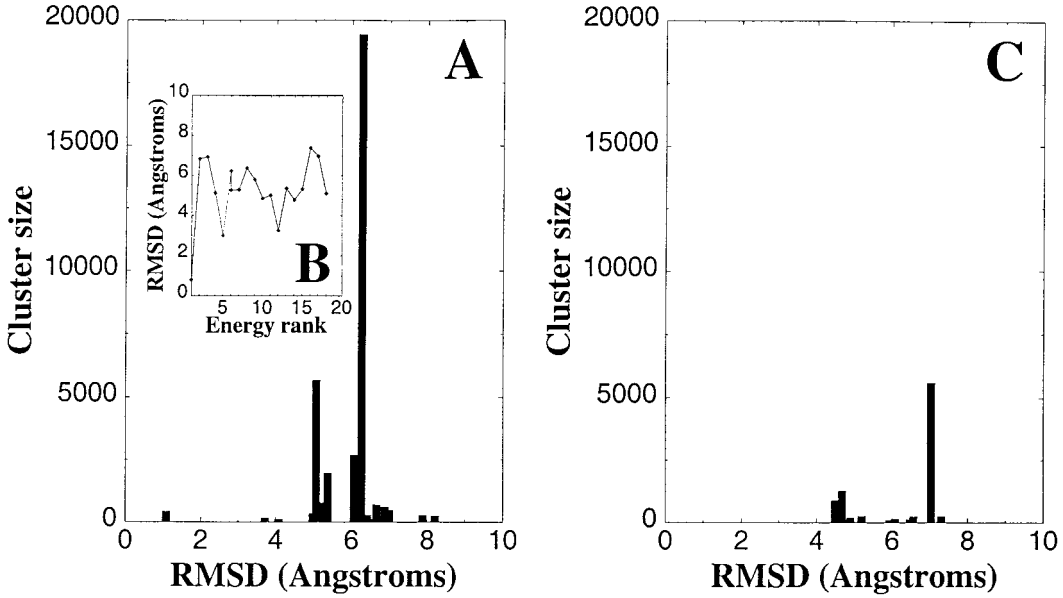

Figure 9. (A, C) The size of structurally similar clusters for the 1cdg ligand-protein complex; and (B) the RMSD of the AMBER minimized conformations ranked by energy. See the caption of Figure 5 for more details.

required per monomer and the high affinity manganese (II) site has been identified in the crystallographic study. The metal ion is complexed by four carboxylate side-chains of the conserved residues Glu-180, Glu-216, Asp-244 and Asp-292, and it was suggested that failure of the GOLD program to predict the native binding mode may have resulted from underestimation of the coordination energy with the manganese (II) cation [31].

Docking simulations with the PL energy function predicted structures near RMSD = 3.5 Å and RMSD = 4.5 Å from the crystal structure (Figure 11A), with the lowest energy structure registered at RMSD = 4.53 Å from the native state (Figure 11B). Although a conformation 0.87 Å RMSD from the crystal structure was located during simulations with the PL energy function, its energy was higher than the lowest energy structure. Docking simulations with the AMBER force field produced two distinct peaks, located at 3.5 Å and 6 Å RMSD from the crystal structure (Figure 11C), with the lowest energy structure located at RMSD = 3.55 Å from the native state (Figure 11D). The lowest energy structure obtained from equilibrium simulations with the PL energy function is located at RMSD = 4.1 Å (Figure 12A) from the crystal structure and the predicted conformation with the AMBER force field resides at RMSD = 5.55 Å from the native state (Figure 12B, Table 1). The binding energy profile generated with the PL energy function has a dominant binding domain at RMSD = 3.5 Å (Figure 12A) that contributes significantly to the thermodynamic equilibrium, but there are several nearby local minima. The energy landscape constructed with AMBER is considerably more rugged and the misdocked binding mode at RMSD = 6 Å from the crystal structure dominates thermodynamic equilibrium at all temperatures (Figure 12B). While local minima on the PL energy landscape represent shallow basins, higher energy barriers are present on the AMBER energy landscape. However, with both energy functions, the native binding domain is only a marginal local minimum at lower temperatures. The broad basins of local minima determined with the PL energy function are reflected in the large clusters of structurally similar conformations located near 3.0 Å, 3.7 Å, 4.1 Å and 5.5 Å RMSD from the crystal structure (Figure 13A). There is also a small cluster of conformations in the native region at RMSD = 0.87 Å from the crystal structure (Figure 13A). The minimized conformation from the native cluster, which is only fourth lowest, is located at RMSD = 2.2 Å from the native state and moves away from its original close proximity to the crystal structure (Figures 13A, 13B, Table 1).

The number of structurally different clusters is significantly larger with the AMBER energy function, indicating that the energy landscape is rugged.

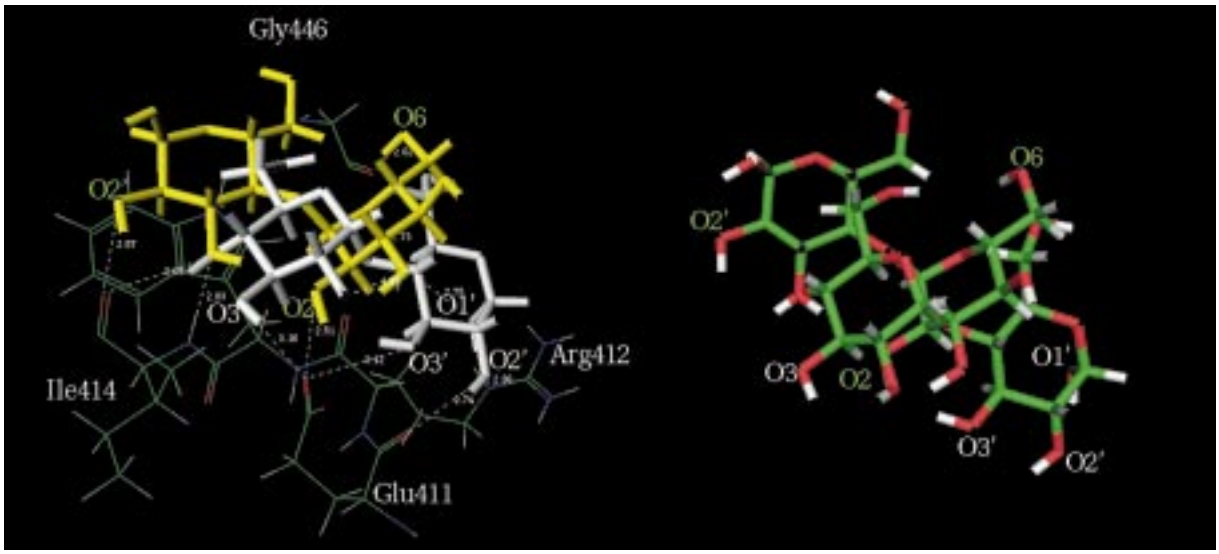

Figure 10. Left panel: Superposition of the crystal structure of the 1cdg ligand-protein complex (yellow) with the lowest energy conformation (gray) obtained from docking simulations. Right panel: Superimposed inhibitor structures color coded by atom type.

The largest clusters are centered near 5.5 Å from the crystal structure and the native-like binding domain is not recognized as a significant cluster of structurally similar conformations (Figure 13C). Unlike the previous two systems, the failure in the 1did ligand-protein complex is truly 'hard'; it is not possible to identify the correct binding mode even with the two-step minimization procedure. The hard failure in this system may result from inaccuracy in reproducing the exact magnitude of the electrostatic and van der Waals interactions, most noticeably interactions with the manganese cation. The structure of the 1did complex could not be successfully predicted using either the PL or the AMBER energy functions, alone or in combination. Even though structural similarity clustering of conformations, generated with the PL energy function, succeeded in characterizing the multitude of the binding modes for the 1did complex, including the crystal structure, the accuracy of reproducing intermolecular interactions is not sufficient to correctly rank the minimized conformations.

In the crystal structure, there is a network of six hydrogen bonds where O-1 of the ligand forms two hydrogen bonds with side-chains OD1 and OD2 of Asp-244, O-3 and O-6 form hydrogen bonds with oxygens of Glu-180, N-5 interacts with Asp-292 and O-4 mediates a hydrogen bond with a water molecule in the active site (Figure 14). In the low-energy conformation that represents the largest cluster leader obtained from sampling with the PL energy function, O-1 of the ligand makes favorable hydrogen bond interactions with Glu-180 and Asp-244; O-3 of the ligand interacts with Asp-292 and with both oxygens of Glu-180 (Figure 14). The bound conformation of the ligand in the crystal structure complements interactions with the protein residues in completing the coordination shell for the manganese cation. The interactions with the cation formed in the alternative binding mode are not as optimal as in the crystal structure. Superposition of the crystal structure and the predicted misdocked binding mode shows that the O-6 ligand atom in the crystal structure and O-3 in the alternative binding mode occupy the same position and form similar hydrogen bonds suggesting different binding modes could share the same type of interactions with key protein residues, critical for functional activity of the complex.

Discussion

An understanding of the common failures and successes in ligand-protein docking helps to establish useful connections between topography of the energy landscapes [83–93] and the results of docking simulations and between the thermodynamic and kinetic requirements of the docking problem. We have found that the stability of the native binding modes for the hard failures determined with different energy models is marginal, and minor structural changes will lead to changes in the relative stability of the native binding mode. The interactions that stabilize the crystal

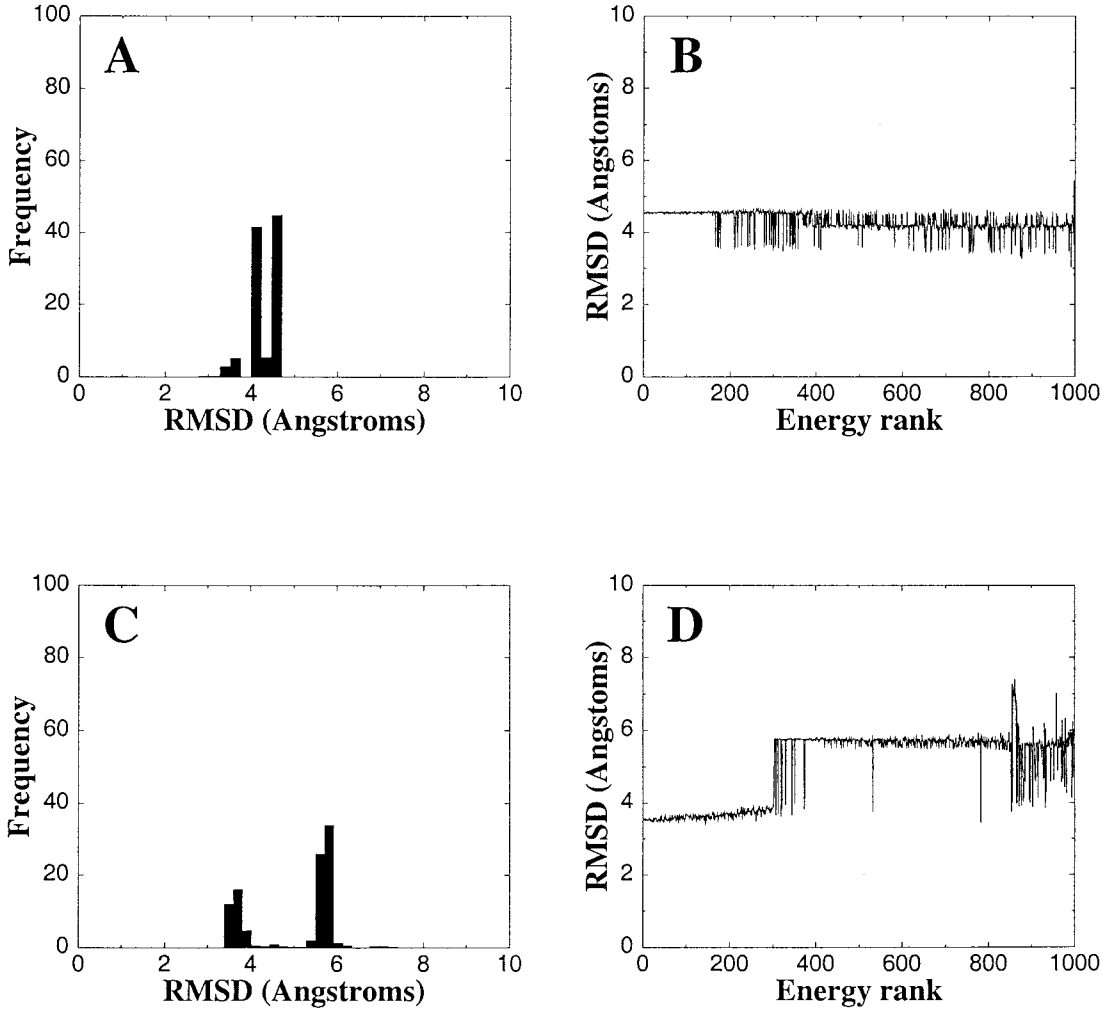

Figure 11. (A, C) The frequency of predicting binding modes of the 1did ligand-protein complex; and (B, D) the RMSD of the docked conformations from the crystal structure ranked by energy. See the caption of Figure 3 for more details.

structure for these complexes may not stabilize near-native conformations, which leads to a diversity of binding modes and a rugged energy landscape. This marginal stability of the binding modes for hard failure complexes coupled with narrow regions in the conformational space that correspond to the native-like conformations could make these systems highly sensitive to protein conformational fluctuations. As a result, reliable structure prediction for studied ligand-protein complexes becomes even more problematic with a flexible protein model. Subsequently, a very sensitive and optimized energy function is required to rank correctly the relative stabilization energy of the crystallographic binding mode of the hard failure complex. This explains why common failures in molecular docking have been detected using different energy functions and searching methods.

On the other hand, the recent analysis of the binding energy landscape for the methotrexate-dihydrofolate reductase (MTX–DHFR) system has provided a plausible rationale of a common success in molecular docking for this ligand-protein complex [57, 58]. We have shown that a pronounced thermodynamic stability of the native structure and robust topology of the native binding mode for the MTX–DHFR complex can explain a common success in molecular docking simulations obtained with the different energy functions [58]. The interactions that favor the native crystallographic binding mode are significantly stronger on average than interactions sta-

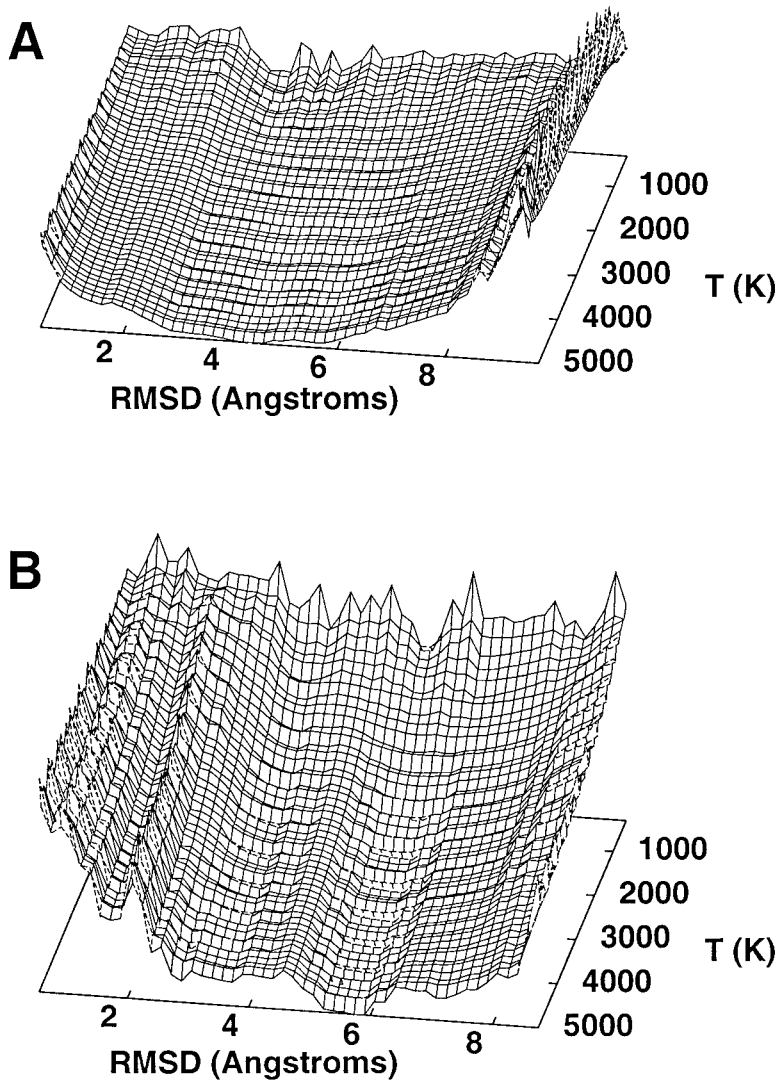

Figure 12. The binding free energy landscape for the 1did complex. See the caption of Figure 4 for more details.

bilizing alternative binding modes, which results in a gradual energy decrease as the native interactions are progressively formed and a broad basin of low-energy native-like conformations in the vicinity of the crystal structure. In particular, the simplified PL energy function has been adequate in structural and thermodynamic analysis of MTX–DHFR and biotin-streptavidin complexes [48]. Moreover, this energy model reproduces subtle differences in relative stability of the binding modes for the MTX–DHFR system and favors the crystal structure over an additional binding mode, corresponding to a conformation of the pteridine ring where an amino group of the pteridine ring is flipped relative to the native conformation [58].

Analogous to a typical folded protein, ligand-protein complexes generally have a well-defined native structure, but on a microscopic level a ligand-protein system may exhibit structural disorder that is revealed on different length and time scales: by rotation of a local protein side-chain, by conformational change of the ligand in the active site or by a collective conformational change associated with a movement of the protein backbone, side-chains and a change of the ligand binding mode. On the basis of structural and thermodynamic data for ligand-protein binding, it has been shown that local folding events and disorder-order transitions could couple to the binding process for HIV-1 protease [94], avidin [95], streptavidin

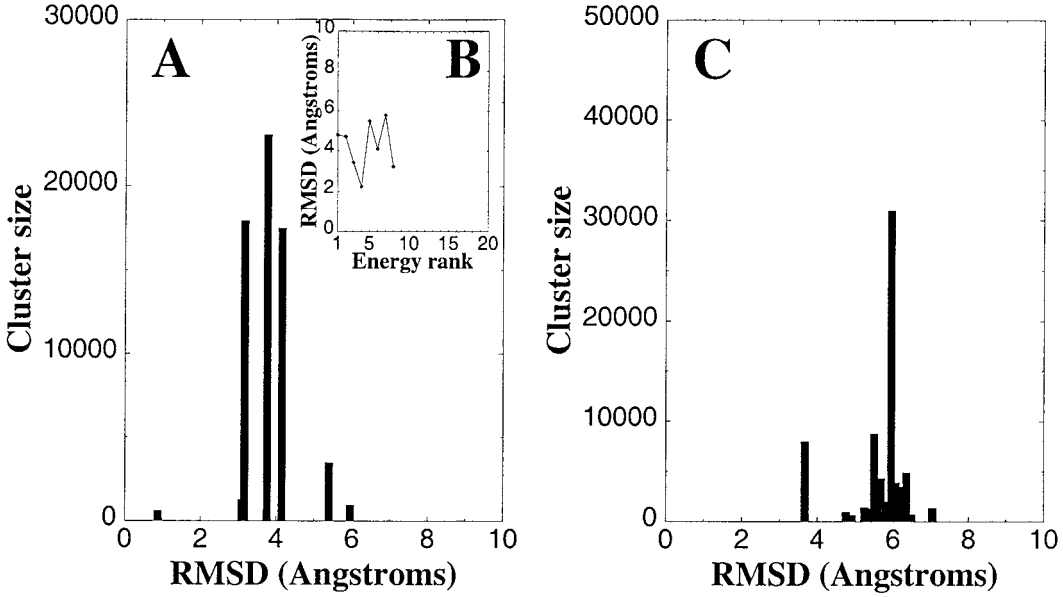

Figure 13. (A, C) The size of structurally similar clusters for the 1did ligand-protein complex; and (B) the RMSD of the AMBER minimized conformations ranked by energy. See the caption of Figure 5 for more details.

[96] and trp repressor [97]. Accurate prediction of low-energy binding modes, associated with significant conformational changes and protein loop motions, can be limited in this case because the search algorithms may fail to adequately characterize the multitude of local minima on the complex energy landscape. A more realistic representation of ligand-protein interactions may be necessary in ligand-protein systems where molecular recognition is primarily determined by specific hydrogen bonds and salt bridges. These energy landscapes are characterized by a multitude of ligand binding modes that can be accompanied by energetically compensating yet structurally different protein side-chain arrangements of the active site residues. A more sensitive and thermodynamically accurate energetic model would be necessary to distinguish between correct and incorrect solutions and locate the global energy minimum on the binding energy landscape.

Nevertheless, hierarchical-based docking strategies can be useful in approaching the more challenging problem of predicting ligand-protein complexes, whereby folding transitions occur upon ligand binding to create an ordered ligand-protein interface. A hierarchical computational approach was developed for predicting structures of ligand-protein complexes and analyzing binding energy landscapes, which combines the Monte Carlo simulated annealing technique to determine the ligand bound conformation with the DEE algorithm for side-chain optimization of the protein active site residues [39]. In this method, each of the docked ligand conformations is used to generate the template for a subsequent step of protein side-chain optimization with the DEE procedure. Local minimizations and energy evaluations of the generated solutions are performed at the final stage of this protocol. In a hierarchical 'double-energy' docking approach, that operates in internal coordinate space, a set of random and biased probability moves are applied to the ligand positions, receptor side-chains and local deformations of flexible loops [41, 42]. In this method, each move is followed by local minimization and ranking according to the energy function that includes an entropy component, surface tension and electrostatic solvation contribution. Importantly, the robustness of this method has improved as the energy landscape was made less frustrated in the final conformational stack of best low-energy conformations for a lysozyme-antibody complex [41].

In the current study, we combine the thermodynamic and kinetic analysis of the binding energy landscapes with structural similarity clustering, which enables an efficient search of the conformational space, but avoids the generation of a large number of conformations for which more detailed energy functions would have to be applied to identify the native binding mode. This hierarchical strategy can be more suit-

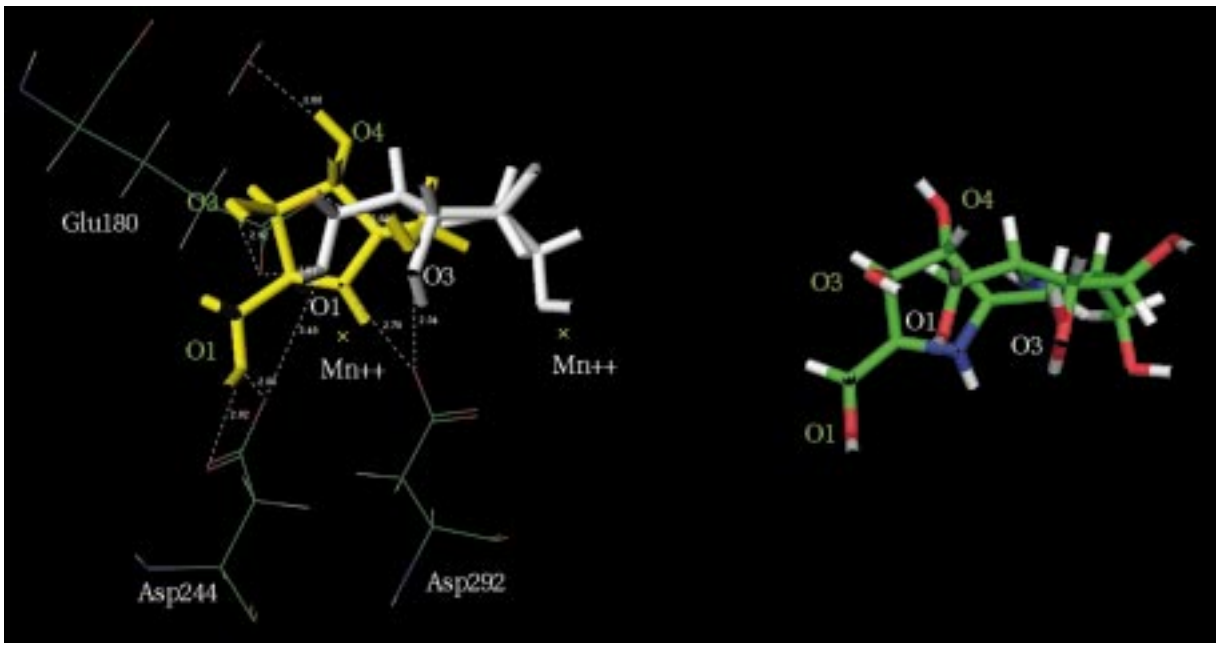

Figure 14. Left panel: Superposition of the crystal structure of the 1did ligand-protein complex (yellow) with the lowest energy conformation (gray) obtained from docking simulations. Right panel: Superimposed inhibitor structures color coded by atom type.

able in the analysis of ligand-protein complexes that exhibit multiple conformational substates and ligands with many rotatable bonds. Ironically, for small rigid ligands, that contain only few or no rotatable bonds, multiple low-energy binding modes may be difficult to sort out based on the structural similarity clustering procedure. For these ligand-protein systems, incorporating energy minimization into a Monte Carlo procedure can be powerful in discriminating between native and misdocked conformations, by locating the native conformation directly during the conformational search [35].

Protein structures determined in different environments, at high pressure, under various pH and solvent conditions, in different crystal forms as well as bound to inhibitors provide information about accessibility of alternative conformational states that is important in molecular recognition [98], but typically only a single rigid protein conformation is used in docking simulations. A recently introduced molecular docking technique employs a set of related crystal structures as 'snapshots' of a dominant protein conformation perturbed by different ligands, crystallization conditions and simple mutations [45]. Different crystal forms of the same ligand-protein complex provide conformational changes due to various packing environments. The analysis of the effect of multiple protein conformational substates in response to ligand binding has led to some practical recipes to effectively account for the types of protein flexibility in docking simulations with ensembles of protein structures [45–47]. In general, common failures in molecular docking to converge upon the crystal structure of the ligand-protein complex may result not only from an inaccurate search algorithm or energy function, but also from the fact that computer simulations can be performed under different 'free energy' conditions as the experiment. A failure to reproduce the crystal structure in docking simulations can have these two broad motivations, and both of which may be relevant for a given ligand-protein system. In particular, in the 1cdg complex three maltose binding sites have been observed on the surface of the enzyme, two in domain E and one in domain C [81]. It has been suggested that the third maltose binding site in domain C can mediate contacts between symmetry related molecules. In this crystallographically determined binding site, that is used in docking simulations, one additional direct hydrogen bond is made between maltose and a symmetry-related protein residue, and therefore crystal contacts can contribute, but apparently play a secondary role in stabilization of the complex. It is important to realize, that there can be additional fundamental reasons for a failure in predicting the crystal structure of a

ligand-protein complex, beyond the scope of the soft and hard failures in computer simulations of molecular docking, studied in this work. Consequently, optimization of the energetic models in docking simulations is desirable to conduct for ligand-protein complexes with the crystal structures corresponding to thermodynamically stable physiological states. Importantly, binding energy landscapes of the ligand-protein complexes, that are primarily determined by the dominant and thermodynamically stable native binding mode, appeared to be robust to protein structural perturbations and relatively insensitive to the accuracy of the energetic model describing ligand-protein interactions [46, 48, 56–58]. For these ligand-protein complexes the thermodynamically stable native binding modes can be unambiguously recovered in simulations with either a single protein conformation or an ensemble of protein conformations [46].

Conclusions

We have determined that for 1eta and 1cdg ligand-protein systems a two-step protocol of first identifying clusters of structurally similar conformations generated in equilibrium simulations with the simplified PL energy function, followed by energy minimization with the AMBER energy function, led to predictions in better agreement with experiment than using either energy function by itself. Structural similarity clustering of low-energy conformations allows to detect isolated regions of the conformational space that correspond to the crystallographic binding modes of the studied ligand-protein complexes. The binding energy landscape is less sensitive to the details of specific interactions and a more adequate sampling of the available binding modes has been achieved with the PL energy function.

The energy landscapes generated with AMBER are either more rugged, with a number of misdocked meta-stable binding modes, or are characterized by shallow and flat binding regions corresponding to the marginally stable local minima. A limited sampling of low-energy states in simulations with AMBER indicates that it is unlikely to achieve adequate sampling of the low-energy binding basins in equilibrium simulations with a molecular mechanics force field. It is assumed in our study that the simplified energy function tracks the low-energy regions of the 'true' potential. This assumption implies that the breadth of the local minima basins results from the long-range character of hydrophobic interactions and should be recognizable by using the simplified knowledge-based energy function. As a result, clusters of the low-energy samples, generated with the simplified PL energy function, can be mapped onto the corresponding local minima basins of the AMBER energy landscape by a direct minimization procedure. We have applied a hierarchy of energy functions to resolve thermodynamic and kinetic requirements of the docking problem for the hard failure systems in a multi-stage hierarchical procedure. This approach resolves the kinetic aspect of the docking problem by utilizing the robustness of the PL energy function in describing the overall topography of the binding energy landscape. Parallel simulated tempering dynamics on the simplified binding energy landscape allows to effectively characterize the multitude of the available low-energy binding modes and avoids kinetic traps, typically produced by more detailed force fields, that are sensitive to the precise geometry of the binding modes and are characterized by high energy barriers. The standard AMBER force field is less amenable to searching, but, in principle, it should describe more adequately the energetics of ligand-protein interactions, which is critical in recognizing the native binding mode for hard failure systems in molecular docking as the global free energy minimum. Relative energetic stability of the binding modes is determined by the AMBER force field, which ranks the minimized conformations by quantifying the exact magnitude of ligand-protein interactions.

The results of this work suggest that for some of the hard failure complexes, the native binding mode with the lowest stabilization energy may correspond to a narrow and isolated region on the binding energy landscape. Consequently, neither the determination of a single structure with the lowest energy, nor finding the largest cluster of structurally similar conformations and the common inhibitor binding mode is sufficient to predict crystal structures of the complexes which belong to the category of hard failures. It appears that a hierarchical procedure that incorporates energy refinement of all clusters may resolve a common hard failure in molecular docking. In the analysis of other cases, that represent true hard failures in ligand-protein structure prediction, with no *a priori* knowledge of the correct binding mode, flexible protein models combined with optimized energy functions may provide a next step in developing improved docking strategies.

References

1. Kuntz, I.D., Science, 257 (1992) 1078.
2. Shoichet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D. and Perry, K.M., Science, 259 (1993) 1445.
3. Shoichet, B.K., Leach, A.R. and Kuntz, I.D., Proteins Struct. Funct. Genet., 34 (1999) 4.
4. Kuntz, I.D., Meng, E.C. and Shoichet, B.K., Acc. Chem. Res., 27 (1994) 117.
5. Rosenfeld, R., Vajda, S. and DeLisi, C., Annu. Rev. Biophys. Biomol. Struct., 24 (1995) 677.
6. Straatsma, T.P. and McCammon, J.A., Annu. Rev. Phys. Chem., 43 (1992) 407.
7. Cherfils, J. and Janin, J., Curr. Opin. Struct. Biol., 3 (1993) 265.
8. Kollman, P., Chem. Rev., 93 (1993) 2395.
9. Ajay and Murcko, M.A., J. Med. Chem., 38 (1995) 4953.
10. Jones, G. and Willett, P., Curr. Opin. Biotechnol., 6 (1995) 652.
11. Gschwend, D.A., Good, A.C. and Kuntz, I.D., J. Mol. Recognit., 9 (1996) 175.
12. Shoichet, B.K. and Kuntz, I.D., J. Mol. Biol., 221 (1991) 327.
13. Walls, P.H. and Sternberg, M.J.E., J. Mol. Biol., 228 (1992) 277.
14. Vakser, I.A. and Aflalo, C., Proteins Struct. Funct. Genet., 20 (1994) 320.
15. Jackson, R.M. and Sternberg, M.J.E., J. Mol. Biol., 250 (1995) 258.
16. Fisher, D., Lin, S.L., Wolfson, H.J. and Nussinov, R., J. Mol. Biol., 248 (1995) 459.
17. Norel, R., Lin, S.L., Wolfson, H.J. and Nussinov, R., J. Mol. Biol., 252 (1995) 263.
18. Gabb, H.A., Jackson, R.M. and Sternberg, M.J., J. Mol. Biol., 272 (1997) 106.
19. Jackson, R.M., Gabb, H.A. and Sternberg, M.J.E., J. Mol. Biol., 276 (1998) 265.
20. Friedman, A.R., Roberts, V.A. and Tainer, J.A., Proteins Struct. Funct. Genet., 20 (1994) 15.
21. Gehlhaar, D.K., Verkhivker, G.M., Rejto, P.A., Sherman, C.J., Fogel, D.B., Fogel, L.J. and Freer, S.T., Chem. Biol., 2 (1995) 317.
22. Verkhivker, G.M., Rejto, P.A., Gehlhaar, D.K. and Freer, S.T., Proteins Struct. Funct. Genet., 25 (1996) 342.
23. Rarey, M., Kramer, B., Lengauer, T. and Klebe, G., J. Mol. Biol., 261 (1996) 470.
24. Rarey, M., Kramer, B. and Lengauer, T., J. Comput.-Aided Mol. Design, 11 (1997) 369.
25. Welch, W., Ruppert, J. and Jain, A.N., Chem. Biol., 3 (1996) 449.
26. Caflish, A., Niederer, P. and Anliker, M., Proteins Struct. Funct. Genet., 13 (1992) 223.
27. Hart, T.N. and Read, R.J., Proteins Struct. Funct. Genet., 13 (1992) 206.
28. Di Nola, A., Roccatano, D. and Berendsen, H.J.C., Proteins Struct. Funct. Genet., 19 (1994) 174.
29. Clark, K.P. and Ajay, J. Comput. Chem., 16 (1995) 1210.
30. Oshiro, C.M., Kuntz, I.D. and Dixon, J.S., J. Comput.-Aided Mol. Design, 9 (1995) 113.
31. Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor R., J. Mol. Biol., 267 (1997) 727.
32. Westhead, D.R., Clark, D.E. and Murray, C.W., J. Comput.-Aided Mol. Design, 11 (1997) 209.
33. Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R. and Eldridge, M.D., Proteins Struct. Funct. Genet., 33 (1998) 367.
34. Apostolakis, J., Pluckthun, A. and Caflish, A., J. Comput. Chem., 19 (1998) 21.
35. Trosset, J.-Y. and Scheraga, H.A., J. Comput. Chem., 20 (1999) 244.
36. Mangoni, M., Roccatano, D. and Di Nola, A., Proteins Struct. Funct. Genet., 35 (1999) 153.
37. Leach, A.R., J. Mol. Biol., 235 (1994) 345.
38. Desmet, J., Wilson, I.A., Joniau, M., De Mayer, M. and Lasters, I., Faseb J., 11 (1997) 164.
39. Schaffer, L. and Verkhivker, G.M., Proteins Struct. Funct. Genet., 33 (1998) 295.
40. Wasserman, Z.R. and Hodge, C.N., Proteins Struct. Funct. Genet., 24 (1996) 227.
41. Totrov, M. and Abagyan, R., Nat. Struct. Biol., 1 (1994) 259.
42. Totrov, M. and Abagyan, R., Proteins Struct. Funct. Genet., Supplement 1 (1997) 215.
43. Sandak, B., Wolfson, H.J. and Nussinov, R., Proteins Struct. Funct. Genet., 32 (1998) 159.
44. Lorber, D.M. and Shoichet, B.K., Protein Sci., 7 (1998) 938.
45. Knegtel, R.M., Kuntz, I.D. and Oshiro, C.M., J. Mol. Biol., 266 (1997) 424.
46. Bouzida, D., Rejto, P.A., Arthurs, S., Colson, A.B., Freer, S.T., Gehlhaar, D.K., Larson, V., Luty, B.A., Rose, P.W. and Verkhivker, G.M., Int. J. Quantum Chem., 72 (1999) 73.
47. Carlson, H.A. and McCammon, J.A., Mol. Pharmacol., 57 (2000) 213.
48. Bouzida, D., Arthurs, S., Colson, A.B., Freer, S.T., Gehlhaar, D.K., Larson, V., Luty, B.A., Rejto, P.A., Rose, P.W. and Verkhivker, G.M., In Altman, R.B., Dunker, A.K., Hunter, L., Klein, T. and Lauderdale, K. (Eds.), Pacific Symposium on Biocomputing-99, World Scientific, Singapore, 1999, pp. 426–437.
49. Rejto, P.A. and Verkhivker, G.M., Proc. Natl. Acad. Sci. USA, 93 (1996) 8945.
50. Oshiro, C.M. and Kuntz, I.D., Proteins Struct. Funct. Genet., 30 (1998) 321.
51. Ewing, T.J.A. and Kuntz, I.D., J. Comput. Chem., 18 (1997) 1175.
52. Sun, Y., Ewing, T.J.A., Skillman, A.G. and Kuntz, I.D., J. Comput.-Aided Mol. Design, 12 (1998) 597.
53. Shortle, D., Simons, K.T. and Baker, D., Proc. Natl. Acad. Sci. USA, 95 (1998) 11158.
54. Shah, N., Rejto, P.A. and Verkhivker, G.M., Proteins Struct. Funct. Genet., 28 (1997) 421.
55. Rejto, P.A., Verkhivker, G.M., Gehlhaar, D.K. and Freer, S.T., in van Gunsteren, W., Weiner, P. and Wilkinson, A.J. (Eds.), Computational Simulation of Biomolecular Systems. ESCOM, Leiden, 1997, pp. 451–465.
56. Bouzida, D., Rejto, P.A. and Verkhivker, G.M., Int. J. Quantum Chem., 73 (1999) 113.
57. Rejto, P.A., Bouzida, D. and Verkhivker, G.M., Theor. Chem. Acc., 101 (1999) 138.
58. 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. and Rose, P.W., J. Mol. Recognit., 12 (1999) 371.
59. Mayo, S.L., Olafson, B.D. and Goddard, III, W.A., J. Phys. Chem., 94 (1990) 8897.
60. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Chio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
61. Jorgensen, W. L. and Tirado-Rives, J., J. Am. Chem. Soc., 110 (1988) 1657.
62. Stouten, P.F.W., Frömmel, C., Nakamura, H. and Sander, C., Mol. Simul., 10 (1993) 97.

63. Dewar, M.J.S. and Thiel, W., J. Am. Chem. Soc., 99 (1977) 4899.
64. Besler, B.H., Merz, Jr., K.M. and Kollman, P.A., J. Comput. Chem., 11 (1990) 431.
65. Marinari, E. and Parisi, G., Europhys. Lett., 19 (1992) 451.
66. Hukushima, K. and Nemoto, K., J. Phys. Soc., 65 (1996) 1604.
67. Hansmann, U.H.E. and Okamoto, Y., Phys. Rev., E54 (1996) 5863.
68. Hansmann, U.H.E. and Okamoto, Y., Phys. Rev., E56 (1997) 2228.
69. Hansmann, U.H.E. and Okamoto, Y., J. Comput. Chem., 18 (1997) 920.
70. Hansmann, U.H.E., Chem. Phys. Lett., 281 (1997) 140.
71. Bouzida, D., Kumar, S. and Swendsen, R.H., Phys. Rev., A45 (1992) 8894.
72. Ferrenberg, A.M. and Swendsen, R.H., Phys. Rev. Lett., 63 (1989) 1195.
73. Boczko, E.M. and Brooks III, C.L., J. Phys. Chem., 97 (1993) 4509.
74. Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A. and Rosenberg, J.M., J. Comput. Chem., 13 (1992) 1011.
75. Kearsley, S.K., Acta Crystallogr., A45 (1989) 208.
76. Morgan, H.L., J. Chem. Doc. 5 (1965) 107.
77. Bawden, D., In Warr, W.A. (Ed.), Chemical Structures: The International Language of Chemistry, Springer-Verlag, Berlin, 1988, pp. 145–150.
78. Shannon, C.E., The Bell System Techn. J., 27 (1948) 379.
79. Hamilton, J.A., Steinrauf, L.K., Braden, B.C., Liepnieks, J., Benson, M.D., Holmgren, G., Sandgren, O. and Steen, L., J. Biol. Chem., 268 (1993) 2416.
80. Steinrauf, L.K., Hamilton, J.A., Braden, B.C., Murrell, J.R. and Benson, M.D., J. Biol. Chem., 268 (1993) 2425.
81. Lawson, C.L., van Montfort, R., Strokopytov, B., Rozeboom, H.J., Kalk, K.H., de Vries, G.E., Penninga, D., Dijkhuizen, L. and Dijkstra, B.W., J. Mol. Biol., 236 (1994) 590.
82. Henrick, K., Collyer, C.A. and Blow, D., J. Mol. Biol., 208 (1989) 129.
83. Dill, K.A., Protein Sci., 8 (1999) 1166.
84. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. and Wolynes, P.G., Proteins Struct. Funct. Genet., 21 (1995) 167.
85. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D. and Chan, H.S., Protein Sci., 4 (1995) 561.
86. Dill, K.A. and Chan, H.S., Nature Struct. Biol., 4 (1997) 10.
87. Shakhnovich, E.I., Curr. Opin. Struct. Biol., 7 (1997) 29.
88. Janin, J., Proteins Struct. Funct. Genet., 25 (1996) 438.
89. Verkhivker, G.M. and Rejto, P.A., Proc. Natl. Acad. Sci. USA, 93 (1996) 60.
90. Verkhivker, G.M. and Rejto, P.A., Proteins Struct. Funct. Genet., 28 (1997) 313.
91. Tsai, C.-J., Xu, D. and Nussinov, R., Curr. Biol., 3 (1998) R71.
92. Tsai, C.-J., Kumar, S., Ma, B. and Nussinov, R., Protein Sci., 8 (1999) 1181.
93. Zhang, C., Chen, J. and DeLisi, C., Proteins Struct. Funct. Genet., 34 (1999) 255.
94. Wlodawer, A. and Erickson, J.W., Annu. Rev. Biochem., 62 (1993) 543.
95. Livhan, O., Bayer, E.A., Wilcheck, M. and Sussman, J.L., Proc. Natl. Acad. Sci. USA, 90 (1993) 5076.
96. Weber, P.C., Wendolski, J.J., Pantoliano. M.W. and Salemme, F.R., J. Am. Chem. Soc., 114 (1992) 3197.
97. Zhao, D., Arrowsmith, C.H., Jia, X. and Jardetzky, O., J. Mol. Biol., 229 (1993) 735.
98. Rejto, P.A. and Freer, S.T., Prog. Biophys. Mol. Biol., 66 (1996) 167.