Thermodynamics and kinetics of ligand–protein binding studied with the weighted histogram analysis method and simulated annealing

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

Agouron Pharmaceuticals, Inc.
3301 North Torrey Pines Court
La Jolla, CA   92037-1022

The thermodynamics of ligand–protein molecular recognition is investigated by the energy landscape approach for two systems: methotrexate(MTX)–dihydrofolate reductase(DHFR) and biotin–streptavidin. The temperature–dependent binding free energy profile is determined using the weighted histogram analysis method. Two different force fields are employed in this study: a simplified model of ligand–protein interactions and the AMBER force field with a soft core smoothing component, used to soften the repulsive part of the potential. The results of multiple docking simulations are rationalized from the shape of the binding free energy profile that characterizes the thermodynamics of the binding process.

1 Introduction

Molecular recognition underlies protein–protein and ligand–protein association and has considerable importance in the field of drug design [1,2]. In many aspects, the molecular recognition and protein folding problems are similar, both requiring accurate energy evaluation and adequate sampling of the associated conformational space. Simplified models that reduce the complexity of both the protein representation and the energy function have been useful in theoretical studies of protein folding, making apparent the salient features of the problem. The energy landscape approach, built upon these simple models, has proven fruitful in unraveling protein folding mechanisms [3]. By combining kinetic and thermodynamic analyses of protein–like heteropolymers, it has been conjectured that thermodynamic parameters such as transition temperatures may dictate the kinetics of protein foldability [4,5].

The assertion that there is a fundamental connection between protein folding and molecular recognition has been recently put forward in the statistical mechanical analysis of binding energy landscapes [6,7]. The complex character of ligand–protein interactions results in a highly frustrated binding energy landscape with many energetically similar but structurally different local minima that present a multitude of binding modes [6,7,8]. Even predicting the structure of a ligand–protein complex with the protein held fixed in its bound conforma-

tion, typically called the docking problem, requires the robust determination of the global energy minimum from an enormous number of conformations.

Considerable progress towards solving the docking problem has been made in recent years with a variety of efficient stochastic algorithms[9]. By designing a simplified molecular recognition energy model, the energy landscape approach has been effective in the analysis of thermodynamic and kinetic requirements for robust computational structure prediction of ligand–protein complexes[6,7]. However, efforts to explain the successes and failures in structure prediction of ligand–protein complexes have been more limited. The energy landscape picture was developed from studies of minimalist molecular recognition models and the funnel–like character of simplified energy landscapes has been invoked to explain the results of ligand–protein docking. While previous studies[3,4] have relied on a cartoon–like picture of folding funnels, the binding energy landscape approach represents the equilibrium thermodynamics and can be used to highlight the connections between kinetic simulations and the energy landscape. It is of significant theoretical and practical interest to provide a more rigorous justification of the funnel hypothesis and determine the relationship between the results of simple models and more realistic force field models.

In this study, we perform equilibrium simulations for two classic ligand–protein systems, MTX–DHFR[10] and biotin–streptavidin[11], and generate their binding free energy profiles with histogram methods. These systems are among the first ligand–protein complexes whose crystal structures were determined at high resolution and are frequently employed to validate computational methods in structure–based drug design. Both a simplified molecular recognition energy model[12] and a more realistic representation of ligand–protein interactions by the AMBER force field[13] are employed in these simulations, and we analyze how the kinetic results of multiple docking simulations relate to the character of the underlying binding free energy landscape.

2 Monte Carlo simulations

The simplified molecular recognition model used in this study includes both intramolecular energy terms for the ligand, given by torsional and non-bonded functions, and simplified intermolecular ligand-protein interaction terms consisting of steric and hydrogen bond contributions[12]. These contributions are calculated from a piecewise linear potential summed over all protein and ligand heavy atoms, together with an angular dependence for the hydrogen bond interaction. The parameters of the pairwise potential depend on the four different atom types: hydrogen-bond donor, hydrogen-bond acceptor, both donor and acceptor, and nonpolar. Primary and secondary amines are defined to be

donors while oxygen and nitrogen atoms with no bound hydrogens are defined
to be acceptors. Crystallographic water molecules and hydroxyl groups are
defined to be both donor and acceptor, carbon and sulfur 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[14].

   We investigate a variant of the AMBER force field that has been adapted
to ligand–protein docking simulations. The short–ranged repulsive interac-
tions present in many molecular force fields such as AMBER, lead to rough
energy surfaces with high energy barriers separating local minima. With this
force field small changes in position can lead to significant energy changes.
For molecular docking simulations, it has been shown that the energy surface
must be smooth for robust structure prediction of ligand–protein complexes[6];
softening the potentials is one way to smooth the force field[15]. In this ap-
proach, the Lennard–Jones interaction and the electrostatic interaction, with
$\epsilon = 2r$, are modified to eliminate the singularity at short–range and soften the
potential, by adding two parameters r_1 and r_2, where

$$V^{soft}_{Lennard-Jones}(r_{ij}) \;=\; \frac{-A_{ij}}{r_{ij}{}^6 + r_1^6} + \frac{B_{ij}}{(r_{ij}{}^6 + r_1^6)^2}$$

$$V^{soft}_{electrostatic}(r_{ij}) \;=\; \frac{q_i q_j}{2(r_{ij}{}^6 + r_2^6)^{1/3}}$$

The AMBER force field with soft core parameters, $r_1 = 2.7$ Å for Lennard–
Jones interactions, and $r_2 = 1.7$ Å for electrostatic interactions, were used
in this study. Here, r_{ij} is the distance between atoms i and j, A_{ij} and B_{ij}
are the parameters of the Lennard–Jones potential, and q_i are the partial
charges. In addition, a desolvation correction term was added to the AMBER
force field to account for the free energy of interactions between the atoms of
the ligand–protein system and the implicitly modeled solvent. This term was
derived by considering the transfer of an atom from an environment where it
is completely surrounded by solvent to an environment in which it has explicit
atomic neighbors[16]. Upon transfer, the neighbors displace solvent, estimated
by a Gaussian weighting function of the effective volumes of the surrounding
atoms. The volume of solvent displaced from an atom i, X_i, is given by the
sum over all neighbor atoms, n, of the van der Waals volume of the neighbor
atom v_n times a Gaussian weighting function of the distance r_{in} between the
atom and the neighbor

$$X_i = \sum_n v_n \exp[-r_{in}^2/\sigma^2]$$

where $\sigma = 3.5$ Å is the first peak in the carbon–water pair correlation function. The desolvation energy of every atom is determined by multiplying the volume of solvent displaced from the atom by the solvent affinity of the atom, where the solvent affinity is assumed to be proportional to the square of its partial atomic charge in the AMBER force field.

Ligand conformations and orientations are searched by a Monte Carlo simulated annealing technique in a parallelepiped that encompasses the binding site obtained from the crystallographic structure of the corresponding complex, with a 2.0 Å cushion added to every side of this box. In each of 200 independent simulated annealing simulations, the temperature was lowered exponentially from 5000 K to 100 K. Each simulation consists of 128,000 sweeps with 40 temperature points in the exponential annealing schedule, 80 cycles per temperature point and 40 sweeps per cycle. A sweep is defined as a single trial move for each degree of freedom of the system.

In addition to the kinetic docking simulations, we performed Monte Carlo equilibrium simulations at T = 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500 and 5000 K. Each consisted of an equilibration stage of 10^6 sweeps and a data collection stage of 10^7 sweeps. The maximum step sizes were updated every cycle of 10^3 sweeps and data was collected at the end of each cycle, resulting in a total of 10^4 data points at each temperature. To facilitate efficient sampling, we employed the dynamically optimized acceptance ratio method [17], whereby the maximum step sizes at each temperature are dynamically chosen to optimize the acceptance ratio, which is the ratio of accepted moves to the total number of trial moves since the previous update.

The multiple histogram method [18] combines data from different temperatures to estimate the density of states, which can then be used to compute equilibrium properties over a continuous range of temperatures. We apply the weighted histogram analysis method to compute ligand–protein binding energy landscapes [19], $F(R,T)$, as a continuous function of reaction coordinate R and temperature T. The potential of mean force $F(R,T)$ at arbitrary temperature relative to a reference position R_c is computed from the probability density $P(R,T)$ as

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

where

$$P(R,T) = \sum_E W(E,R) \exp[-E/k_B T].$$

We define R to be the root mean square difference (rmsd) of the ligand coordinates from the native state, and the native state is chosen to be the reference state, so $R_c = 0.0$. In this work, we use the first bin in the histogram to estimate define $P(R_c, T)$. The density of states $W(E,R)$ is expressed in terms of the histograms $H_i(E,R)$ tabulated during the Monte Carlo simulations[18].

3 Results and Discussion

We investigate the kinetics of ligand–protein docking and the success rate in predicting the crystal structure of the complex by Monte Carlo simulated annealing for the MTX–DHFR and biotin–streptavidin ligand–protein systems. MTX and biotin have seven and five rotatable bonds, respectively. We supplement these kinetic studies with equilibrium studies of the binding energy landscape. These binding free energy profiles do not measure the binding affinity of the ligands but rather characterize the relative thermodynamic stability of various binding domains. They should be regarded as two–dimensional slices with the reference energy $F(R_c, T)$ defined to be zero at each temperature.

3.1 MTX–DHFR

For the MTX system, the standard AMBER force field was investigated along with the soft–core variation, but the standard force field yields only a low success rate in the docking simulations of MTX (Fig. 1), and was not used to investigate the biotin system. The soft–core AMBER force field yields a high success rate in docking simulations, as does the piecewise linear energy function (Figs. 2, 3). The crystal structure is the lowest–energy structure found for the soft–core AMBER force field, and consequently is the predicted structure (Fig. 4).

For the standard AMBER force field, the MTX–DHFR binding energy landscape is determined only within a 2.5 Å range of the crystal structure (Fig. 5). By contrast, the MTX–DHFR binding energy landscapes for the soft–core AMBER and the piecewise linear energy function extend beyond 8 Å of the crystal structure (Figs. 6, 7). The reason for this difference is that only states near to the crystal structure are sampled during the equilibrium simulations with the standard AMBER force field because states that deviate significantly from the crystal structure are too high in energy or are separated by high barriers. The soft–core AMBER and the piecewise linear energy function do
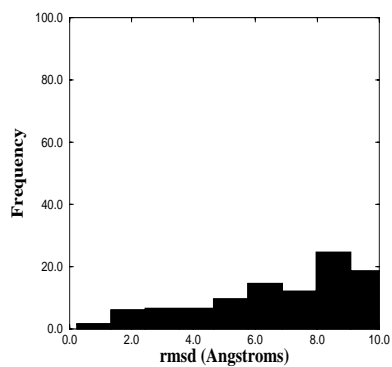
Figure 1: The frequency of predicting the crystal structure of the MTX–DHFR complex with the standard AMBER force field.
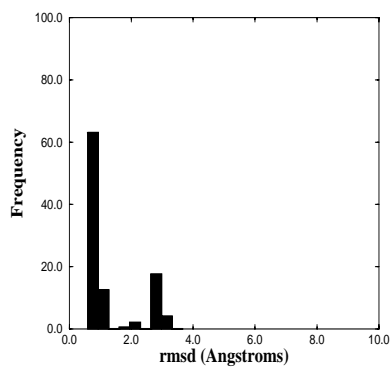

Figure 2: The frequency of predicting the crystal structure of the MTX–DHFR complex with the soft–core AMBER force field.
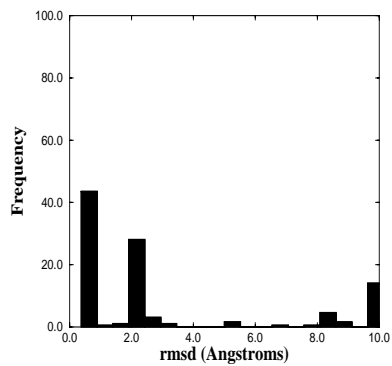

Figure 3: The frequency of predicting the crystal structure of the MTX–DHFR complex with the piecewise linear energy function.
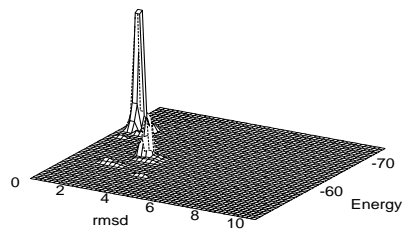

Figure 4: The frequency of predicted structures of the MTX–DHFR complex with the soft–core AMBER force field as a function of energy and rmsd from the crystal.
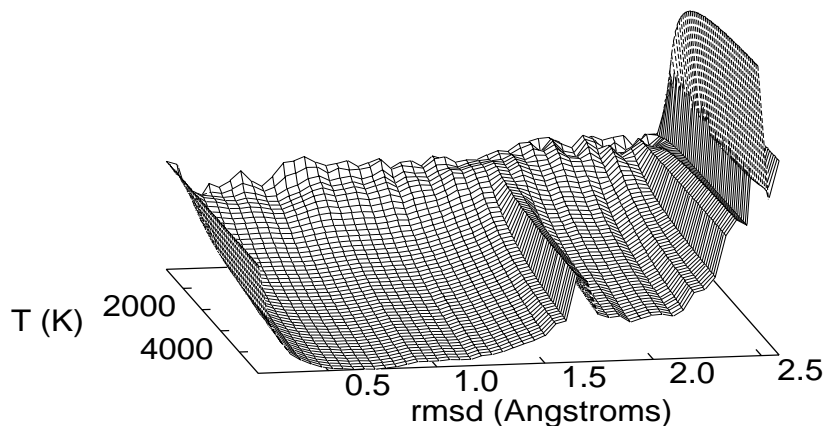
Figure 5: The binding free energy landscape for MTX with the standard AMBER force field. For each two–dimensional temperature slice, the reference energy $F(R_c, T)$ is defined to be zero.

not have singularities at interatomic distances and a much larger fraction of conformational space is accessible, particularly at high temperature.

With the soft–core AMBER force field, a binding mode centered near 2.5 Å is most stable at high temperature, with the crystal favored only at lower temperature (Fig. 6). By contrast, the crystal binding mode is favored throughout the examined temperature range for the piecewise linear force field (Fig. 7). In both cases, the native binding mode belongs to a broad basin; for the soft–core AMBER force field, the funnel extends to approximately 7.0 Å while the funnel for the piecewise linear energy function extends to around 4.0 Å. The binding free energy landscape for the piecewise linear energy function is more rugged than that of the soft–core AMBER force field, with two additional metastable minima centered near 5.0 and 7.0 Å. These additional minima generate additional barriers to the native binding basin, and coupled with the relatively small size of the native binding basin, may be responsible for the erroneous structures predicted by the docking simulations with the piecewise linear energy function, at 5.0 Å and beyond (Fig. 3), that are not found with the soft–core AMBER force field (Fig. 2).

Although the binding free energy landscape generated with the piecewise linear energy function reveals three metastable domains (Fig. 7), the native binding mode dominates the thermodynamic equilibrium even at high temperatures, and alternative local minima are never global minima and therefore
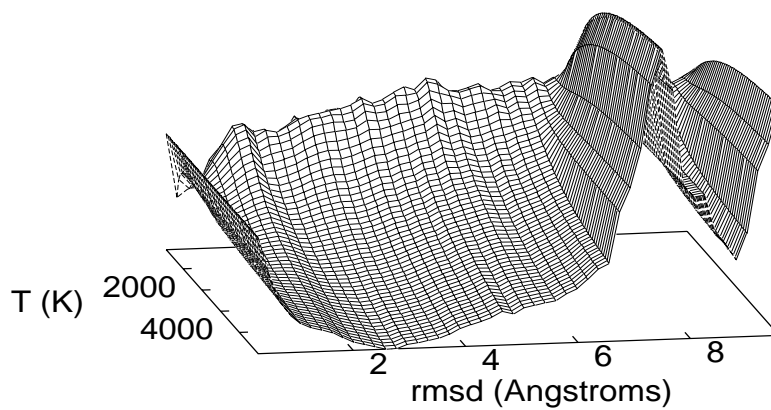
Figure 6: The binding free energy landscape for MTX with the soft−core AMBER force field. For each two−dimensional temperature slice, the reference energy $F(R_c, T)$ is defined to be zero.
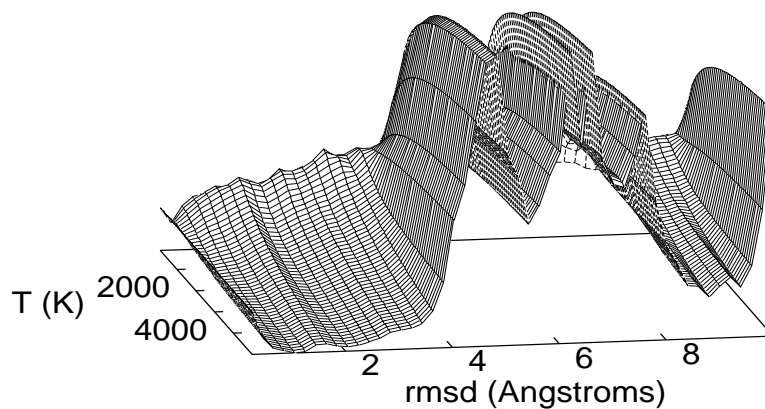

Figure 7: The binding free energy landscape for MTX with the piecewise linear energy function. For each two−dimensional temperature slice, the reference energy $F(R_c, T)$ is defined to be zero.

are not stable. Apparently, the shape of the MTX–DHFR binding energy landscape with the piecewise linear energy function is such that the barriers between high energy states can be overcome at high temperatures, and the probability to remain in these alternate states at low temperature is small. The differences in the shape of the binding free energy profile determined with the piecewise linear energy function and the soft–core AMBER energy function are primarily located in regions that are metastable and therefore may not affect the results of docking simulations too strongly. The presence of extensive funnels for these two force fields, by contrast to the narrow funnel for the standard AMBER force field, rationalizes the high success rate in identifying the crystal structure during the docking simulations.

3.2   *Biotin–streptavidin*

Docking simulations of the biotin–streptavidin complex with the soft–core AMBER energy function predict a number of incorrect structures, with rmsd from the crystal structure between 2.0 and 7.0 Å (Fig. 8). On the other hand, the success rate in determining the crystal structure from multiple docking simulations with the piecewise linear energy function is very high, with 98% of the simulations predicting a structure within 1.0 Å rmsd of the crystal (Fig. 9). The binding free energy profiles generated with these two force fields can rationalize the kinetic docking results. The free energy profile of biotin–streptavidin binding generated with the soft–core AMBER force field is characterized by a broad basin between 1.0 and 7.0 Å, with a series of narrow minima inside this basin (Fig. 10). While the global energy minimum for the biotin–streptavidin complex with the soft–core AMBER force field is thermodynamically stable at low temperature, the alternate minima are stable at high temperature and apparently the system can become trapped in these states during the docking simulation. By contrast, the shape of the binding free energy profile determined with the piecewise linear energy function smoothly connects the state centered about 2.0 Å, which is stable at high temperature, with the state near 1.0 Å, which is stable at low temperature (Fig. 11). The lack of appreciable barriers separating these minima apparently is responsible for the high success rate in the docking simulations.

4   Conclusions

The results suggest that one important factor that governs the success rate in predicting the native ligand-protein complex is the roughness of the binding energy surface. In the absence of broad funnels leading to the native state, as in
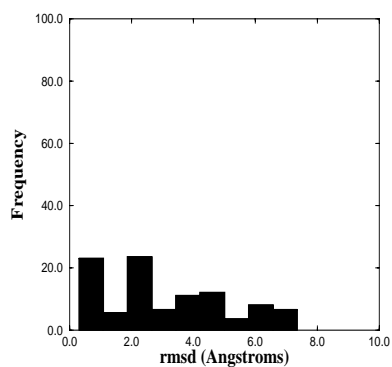
Figure 8: The frequency of predicting the crystal structure of the biotin/streptavidin complex with the soft−core AMBER force field.
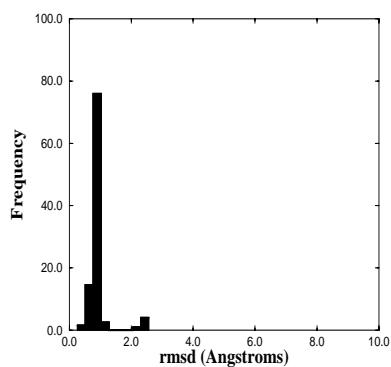

Figure 9: The frequency of predicting the crystal structure of the biotin/streptavidin complex with the piecewise linear energy function.
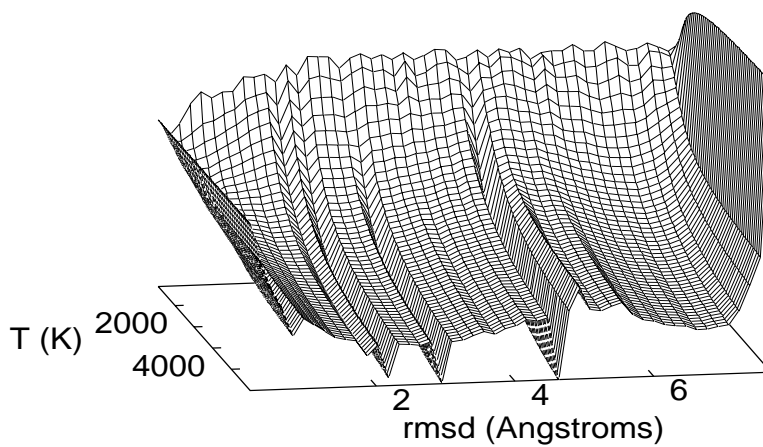

Figure 10: The binding free energy landscape for biotin with the soft−core AMBER force field. For each two−dimensional temperature slice, the reference energy $F(R_c, T)$ is defined to be zero.
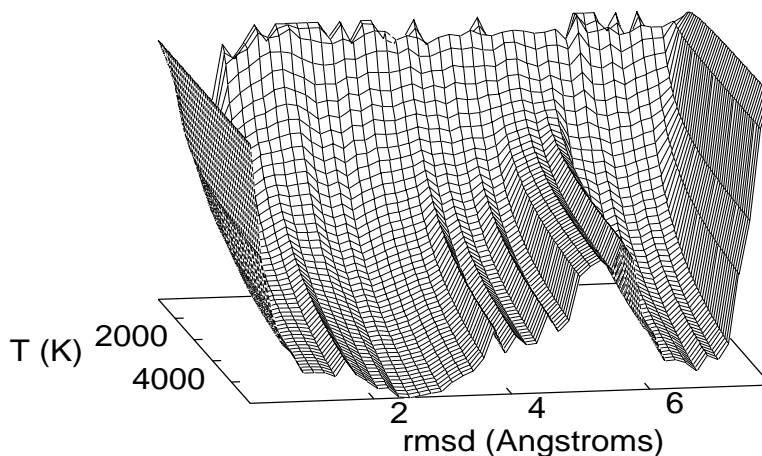
Figure 11: The binding free energy landscape for biotin with the piecewise linear energy function. For each two–dimensional temperature slice, the reference energy $F(R_c, T)$ is defined to be zero.

the case of MTX–DHFR with the standard AMBER force field, the probability of predicting the crystal structure is low. The results on MTX–DHFR with both the soft–core AMBER and the piecewise linear energy function suggest that alternate metastable minima do not necessarily lead to low success rates provided that the global minimum has a significant basin of attraction and the alternative local minima are only marginally stable. The docking of biotin to streptavidin with the soft–core AMBER force field demonstrates a potential complication that arises when alternate minima are stable at high temperature: the system may become trapped in these states during docking simulations.

By analyzing the thermodynamics and kinetics of molecular recognition, we have shown that the results of multiple docking simulations can be rational-ized in the context of the corresponding binding free energy profiles determined from equilibrium simulations. We have shown that for robust ligand-protein docking, the binding energy landscape should be smooth to minimize the like-lihood of trapping in local minima, and have funnels leading to the global free energy minimum. These results also suggest that force fields designed for dock-ing simulations, which may differ from standard force fields, can be optimized by the criterion that they yield wide funnels leading to the crystal structure; this criterion can supplement the previously–used criterion of success rate in docking simulations [6].

References

1. I.D. Kuntz, *Science*, **257**, 1078 (1992).
2. D.W. Miller and K.A. Dill, *Protein Sci.*, **6**, 2166 (1997).
3. J.D. Bryngelson, J.N. Onuchic, N.D. Socci and P.G. Wolynes, *Proteins*, **21**, 167 (1995).
4. L.A. Mirny, V. Abkevich and E.I. Shakhnovich, *Folding & Design*, **1**, 103 (1996).
5. T. Veitshans, D.K. Klimov and D. Thirumalai, *Folding & Design*, **2**, 1 (1997).
6. G.M. Verkhivker, P.A. Rejto, D.G. Gehlhaar and S.T. Freer, *Proteins*, **25**, 342 (1996).
7. P.A. Rejto and G.M. Verkhivker, *Proc. Nat. Acad. Sci. USA*, **93**, 8945 (1996).
8. J. Cherfils and J. Janin, *Curr. Opin. Struct. Biol.*, **3**, 265 (1993).
9. D.R. Westhead, D.E. Clark, and C.W. Murray, *J. Comput. Aided Mol. Des.* **11**, 209 (1997).
10. J.T. Bolin, D.J. Filman, D.A. Matthews, R.C. Hamlin, and J. Kraut, *J. Biol. Chem.* , **22**, 13650 (1982).
11. P.C. Weber, D.H. Ohlendorf, J.J. Wendolowski, and F.R. Salemme, *Science*, **243**, 85 (1989).
12. D.G. Gehlhaar, G.M. Verkhivker, P.A. Rejto, C.J. Sherman, D.B. Fogel, L.J. Fogel, and S.T. Freer, *Chem. Biol.*, **2**, 317 (1995).
13. S.J. Weiner, P.A. Kollman, D.A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.* , **106**, 765, (1984).
14. N.K. Shah, P.A. Rejto and G.M. Verkhivker, *Proteins*, **28**, 421 (1997).
15. T.C. Beutler, A.E. Mark, R.C. van Schaik, P.R. Gerber, and W.F. van Gunsteren, *Chem. Phys. Lett.*, **222**, 529, (1994).
16. P.F.W. Stouten, C. Frömmel, H. Nakamura and C. Sander, *Mol. Simul.*, **10**, 97, (1993).
17. D. Bouzida, S. Kumar and R.H. Swendsen, *Phys. Rev. A*, **45**, 8894 (1992).
18. A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.*, **63**, 1195 (1989).
19. P.A. Rejto, D. Bouzida, and G.M. Verkhivker, *Theo. Chem. Lett.* (in press).