Chasing Funnels on Protein-Protein Energy Landscapes at Different Resolutions

Anatoly M. Ruvinsky* and Ilya A. Vakser*†
*Center for Bioinformatics and †Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas

ABSTRACT Studies of intermolecular energy landscapes are important for understanding protein association and adequate modeling of protein interactions. Landscape representation at different resolutions can be used for the refinement of docking predictions and detection of macro characteristics, like the binding funnel. A representative set of protein-protein complexes was used to systematically map the intermolecular landscape by grid-based docking. The change of the resolution was achieved by varying the range of the potential, according to the variable resolution GRAMM methodology. A formalism was developed to consistently parameterize the potential and describe essential characteristics of the landscape. The results of gradual landscape smoothing, from high to low resolution, indicate that i), the number of energy basins, the landscape ruggedness, and the slope decrease accordingly; ii), the number of near-native matches, defined as those inside the funnel, increases until the trend breaks down at critical resolution; the rate of the increase and the critical resolution are specific to the type of a complex (enzyme inhibitor, antigen-antibody, and other), reflect known underlying recognition factors, and correlate with earlier determined estimates of the funnel size; iii), the native/nonnative energy gap, a major characteristic of the energy minima hierarchy, remains constant; and iv), the putative funnel (defined as the deepest basin) has the largest average depth-related ruggedness and slope, at all resolutions. The results facilitate better understanding of the binding landscapes and suggest directions for implementation in practical docking protocols.

INTRODUCTION

The adequate description of protein-protein interactions is essential for understanding cell machinery. The intermolecular energy landscape determines structure, kinetics, and thermodynamics of macromolecule complexes. The major characteristics of the landscapes—the folding/binding funnel, the ruggedness of the terrain, etc.—are important for interpreting protein folding and interactions and providing guidelines for modeling (1–13). It has been shown that simple energy functions, including coarse-grained (low-resolution) models, reveal major landscape characteristics. The large-scale, systematic studies of protein-protein complexes confirmed the existence of the intermolecular binding funnel (7,14) and further revealed that the number of distinct energy basins is small and that they are well formed (funnel like) and correlated with known binding modes (15).

Spectrum properties of binding landscapes can be characterized by the z-score or by the modified z-score (16,17). The modified z-score is defined as the ratio of the energy gap between the native state and the average of the energy spectrum and the width of the energy spectrum weighted by entropy per contact in the power $-1/2$. Recent studies of protein-protein association (17) and protein-ligand binding (16,18) showed that the modified z-score has large values for well-formed funneled landscapes and can be used as a descriptor of binding affinity in search of specific inhibitors or potential drugs. Studies of the kinetics of biomolecular binding and kinetic pathways (19–21) showed that the binding time monotonically decreases with the increase of the modified z-score. Cooperative binding/folding was studied by single-molecule dynamics (22,23).

The concept of the binding funnel suggests that a systematic energy bias forms pathways between unbound and bound structures all the way to the bottom of the funnel. These pathways go through a rugged energy surface of the funnel. The frustration of local interactions and its distribution in proteins and their complexes has been studied and reported to correlate with the protein topology and functionality (24–26). Highly frustrated interactions are observed on the protein surface near the binding site (24). The possibilities for experimental estimates of the landscape ruggedness of proteins and RNA have been considered (27). In this study we show that both the ruggedness and the bias (slope) of a basin can be powerful discriminating factors for detecting funnel-type basins.

Accurate predictions of protein-protein complexes are challenging owing to the complexity of the energy landscapes (multiple minima problem (28)), inherent imperfections of the energy functions, and computation time limitations. Low-resolution/minimalist approaches have been shown to simplify the landscape while retaining its large-scale characteristics (basins/funnel) important for docking predictions (7,14), albeit at the cost of the lower precision (29).

An important direction for designing better docking procedures is smoothing the landscape by reducing its ruggedness (28,30), amplifying the funnel/local minima depth ratio (31), etc. Smoothing as a global optimization strategy has

Submitted March 5, 2008, and accepted for publication May 5, 2008.

Address reprint requests to Ilya A. Vakser, Center for Bioinformatics, University of Kansas, 2030 Becker Drive, Lawrence, KS 66047. E-mail: vakser@ku.edu.

Editor: Gregory A. Voth.

0006-3495/08/09/2150/10 $2.00

doi: 10.1529/biophysj.108.132977

been applied to the conformational analysis of molecules (32–34) and the detection of the global intermolecular energy minimum in atom clusters (35–41), water clusters (42), helix dimers (28,43,44), and protein complexes (14,28,45). Four major smoothing strategies have been developed. The first approach is based on a diffusion equation formalism (32,34,37,42,44). The method transforms the original energy hypersurface to obtain a single minimum, which in most cases is related to the global one. The second approach is based on the imaginary time Schroedinger equation (38). The third approach is a coarse-grained mapping of the energy landscape applied in conjunction with a ''basin-hopping'' approach (46,47). The fourth approach utilizes variable range energy functions: the stair potentials (7,14,15,28), the Morse potential (36,39,40), the Lennard-Jones potentials (35), and the gravitational potentials (41). It has been shown (30) that the extension of the potential range decreases the local ruggedness of binding landscapes. According to the principle of minimal frustration (4,9,48), smooth landscapes better approximate the actual energy profile.

The variable-range energy function approach based on the stair potential is implemented in our GRAMM docking program. In the GRAMM algorithm, the extension of the potential range averages the contribution of the neighboring atoms, thus smoothing the energy landscape (28). The approach was used in this study to track changes of the major landscape characteristics at different levels of resolution, from ultralow to atomic. The results show that along the transition from low to high resolution and the corresponding increase of the number of energy minima and the landscape ruggedness, the native/nonnative energy gap remains unchanged, binding funnels are rougher and steeper than the rest of the landscape, and the depth of a basin correlates with its ruggedness and slope, suggesting their use in the funnel detection for docking predictions.

METHODS

The results are obtained on a set of 92 nonobligate protein-protein unbound structures with known cocrystallized structures generated by Dockground (http://dockground.bioinformatics.ku.edu (49,50)). The structures were selected from the Protein Data Bank based on the following criteria: sequence identity between bound and unbound structures is >97%, sequence identity between complexes is <30%, and homomultimers and crystal packing complexes are excluded. The test set contains 26 enzyme-inhibitor complexes, 6 antibody-antigen structures, and 60 other complexes. The average root mean-square deviation (RMSD) between bound and unbound receptors is 2.5 Å and that of the ligands is 1.9 Å.

For the sampling of the landscapes, we used our GRAMM procedure, which has been extensively described in the literature over the years (29,31,51–53). It performs a systematic six-dimensional (6D) rigid body search using correlation by fast Fourier transformation to convolute the translational coordinates—a method extensively utilized in the protein docking community. GRAMM was parameterized according to Eq. 3 and applied to the unbound structures in each complex. In each of the docking runs, we used the grid step equal to half the range of the potential, a 5° interval for rotations, and analyzed the 5000 lowest energy matches per complex.

RESULTS AND DISCUSSION

Energy function

The energy landscapes at different resolutions were described according to GRAMM methodology (29,31,51–53) by atom-atom stair potentials ε with two parameters: repulsion energy U (with attraction energy always equal to -1) and the interaction range R (Fig. 1).

The change of resolution from high to low corresponds to the increase of the potential range 3.4–11 Å (28). Longer potential ranges increase the number of atom-atom interactions and change the energy of these interactions. Since the goal of the study is to explore the binding landscape at different resolutions, it is important to have a consistent route between the resolutions, which would preserve the gross features of the landscape. The surface atoms play an important role in binding, providing a dominant contribution to the formation of the intermolecular energy landscape. Thus we parameterized the potentials based on the condition of invariance of the surface atoms' interaction energies under the following transformation of the potentials $\varepsilon(R_o, U_o) = \varepsilon(R, U)$ and $R > R_o$, where $\varepsilon(R_o, U_o)$ and $\varepsilon(R, U)$ are the energies of a surface atom, U and U_o are the repulsion values,

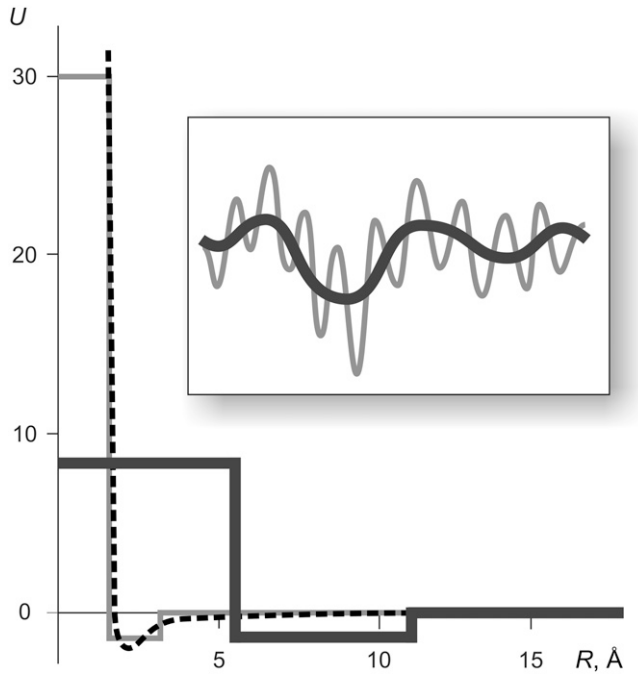

FIGURE 1 The stair potential and the concept of low-resolution smoothing. The sampling (docking) procedure, at high resolution, employs a step function approximation (*thin light gray line*) of the Lennard-Jones potential (*dashed line*). At low resolution, the range of the potential (*thick dark gray line*) is extended. The inset shows a schematic representation of the landscape smoothing. The extended range of the potential removes the ''high-frequency'' fluctuations (local minima) from the landscape (*thin light gray line*) and reveals the ''low-frequency'' characteristics on the underlying, low-resolution landscape (*thick dark gray line*).

and $2R$ and $2R_o$ are the interaction ranges of short (high-resolution) and long (low-resolution) potentials, respectively. The energy of the interface atom at the high resolution can be estimated roughly as

$$\varepsilon(R_o, U_o) = U_o \frac{R_o^3}{3}\left[1 - \frac{3a}{2R_o} + \frac{a^3}{2R_o^3}\right] - \frac{R_o^3}{3}\left[7 - \frac{9a}{2R_o}\right], \quad (1)$$

where a is the shortest distance between protein surfaces (see Appendix and Fig. 2). For the low resolution, we have

$$\varepsilon(R, U) = U \frac{R^3}{3}\left[1 - \frac{3a}{2R} + \frac{a^3}{2R^3}\right] - \frac{R^3}{3}\left[7 - \frac{9a}{2R}\right]. \quad (2)$$

Combining Eqs. 1 and 2 gives

$$U(R) = \frac{3}{R^3}\left[1 - \frac{3a}{2R} + \frac{a^3}{2R^3}\right]^{-1}\left[\varepsilon(R_o, U_o) + \frac{R^3}{3}\left[7 - \frac{9a}{2R}\right]\right]. \quad (3)$$

Using $R_o = 1.7A$ and $U_o = 30$ (28), we obtain the function $U(R)$, which approximates well the empirical parameters used in protein docking earlier (28) (Fig. 3).

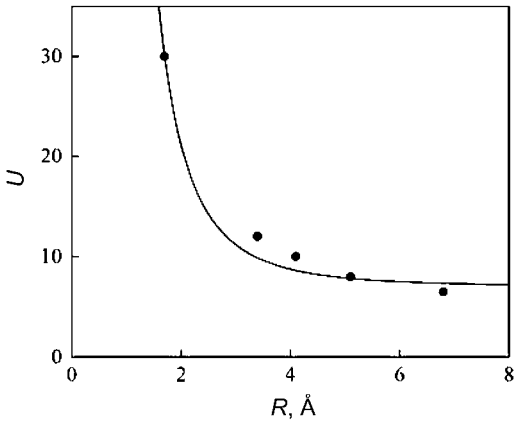

FIGURE 3 Parameters of the stair potential. The repulsion value U is shown as a function of the potential range R. The analytically derived function approximates the empirical values (*dots*) used earlier (28).

Landscape smoothing

A systematic grid-based sampling of energy values by the GRAMM procedure covers the entire landscape. The values below a set energy level form low-energy patches: clusters of points (Fig. 4). Thus the clusters of low-energy matches provide a convenient way to describe the landscape minima.

We partitioned all docked matches into nonoverlapping clusters of matches within a set value of RMSD from the lowest energy match in the cluster. The clustering procedure used the list of all docked matches ordered according to their energies by GRAMM. Then, starting from the lowest energy match (the origin of the first cluster), we computed RMSDs of this match with respect to all matches with higher energies. If the RMSD is lower than the clustering limit, we assign the corresponding higher energy match to the first cluster. The second cluster grows up from the lowest energy match, which is not covered by the first cluster. The RMSDs of this match were computed with respect to all unassigned matches with higher energies. The algorithm iterates until all docked positions are assigned to clusters.

The number of minima (clusters) averaged over the test set versus the potential range is shown in Fig. 5. The data on all clusters, including single match clusters (Fig. 5 *a*) and dense clusters with occupancy higher than average (Fig. 5 *b*), show a steady decrease in the number of minima with the increase of the potential range, illustrating the smoothing of the landscape.

Another view on the landscape-smoothing process can be provided by following the number of near-native matches delivered by systematic grid-based sampling at different smoothing stages. For the purpose of this study, we define a near-native match as one within the binding funnel. If the total number of minima decreases along with the smoothing of the potential, then the number of low-energy matches within the remaining minima, including the funnel, has to increase.

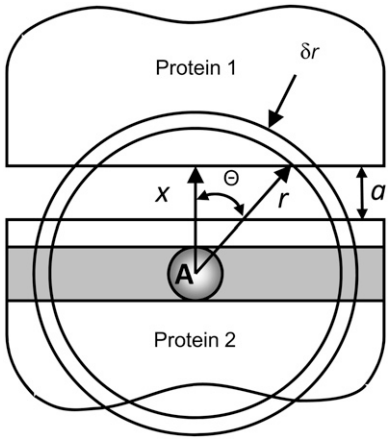

FIGURE 2 The idealized representation of proteins.

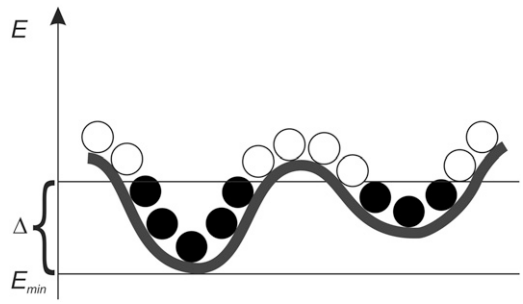

FIGURE 4 Characterization of basins on the energy landscape by clusters of low-energy matches. The grid-based systematic sampling places all matches (*circles*) in uniform distribution on the landscape. The matches with the energy below a set value (*solid circles*) form clusters, thus defining the energy basins.

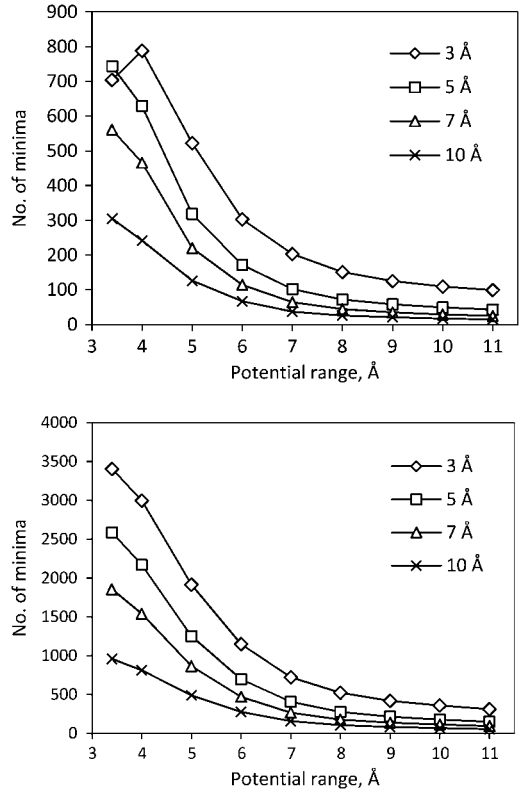

FIGURE 5 The number of minima (clusters) averaged over the test set versus the potential range. The data are shown for clustering radii 3, 5, 7, and 10 Å for all clusters including single match clusters (*upper panel*) and dense clusters with an occupancy higher than average (*lower panel*).

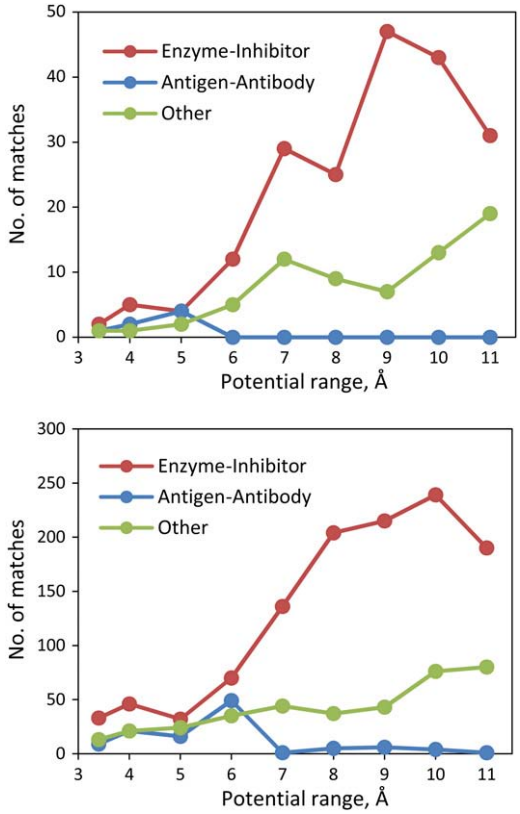

FIGURE 6 The average number of near-native matches in 5000 lowest energy matches for different types of complexes. The near-native matches were defined as those with ligand interface C^{α} RMSD < 5 Å (*upper panel*) and < 10 Å (*lower panel*).

According to the above definition of the near-native match, the criterion for a near-native match was RMSD of C^{α} interface atoms <5 Å and <10 Å from the experimentally determined structure. This corresponds to the earlier estimates of the funnel size (54). The results show that the number of near-native matches increases with the smoothing of the potential (Fig. 6). The enzyme-inhibitor complexes have the largest number of near-native matches. This corresponds to a well-defined funnel, which characterizes this type of complexes (54), largely based on the dominant geometric recognition factor—the active site (for the role of geometric recognition factors, see Nicola and Vakser (55), Nayal and Honig (56), and Binkowski et al. (57)). For the enzyme-inhibitor complexes, the trend of the increasing number of near-native matches breaks down at the longest ranges of the potential. The corresponding high degree of smoothing exceeds the optimum for the moderate size of the funnel in the enzyme-inhibitor complexes (54). The trend, however, continues for the category of ''other'' complexes, which has larger funnels (54). For the antigen-antibody complexes, the trend breaks down early, in complete correspondence with the earlier data on the small size of their funnel (54).

The landscape smoothing is illustrated in the example of one complex in Figs. 7–9, showing different representations of the actual 6D (external/rigid body coordinates) landscape. The convolution of the 6D into a one-dimensional (1D) RMSD plot is shown in Fig. 7. A three-dimensional (3D) cross section as the distribution of minima (low-energy matches) is shown in Fig. 8. The complete energy profile in two-dimensional (2D) cross section is shown in Fig. 9.

Hierarchy of landscape minima

We analyzed the energy gap between the detected lowest point on the landscape and the higher energy states. Such a gap between the native state (presumed to be the lowest energy one) and the nonnative states is characteristic to folding and binding funnel-type landscapes (29,48,58,59). The ratio E^{i}_{ave}/E^{i}_{min} (E^{i}_{ave} and E^{i}_{min} are the average and the minimal energy values detected by the landscape sampling for complex i) was analyzed as a function of the potential range for each of the complexes. These ratios were averaged over all 92 complexes from the test set (see Methods). As $2R$ increased from 3.4 Å to 11 Å, $\sum_{i=1}^{92} E^{i}_{ave}/(92E^{i}_{min})$changed unsubstantially from 0.7 to 0.8. Thus the transformation of the landscape (smoothing) preserves the characteristic energy gap. It

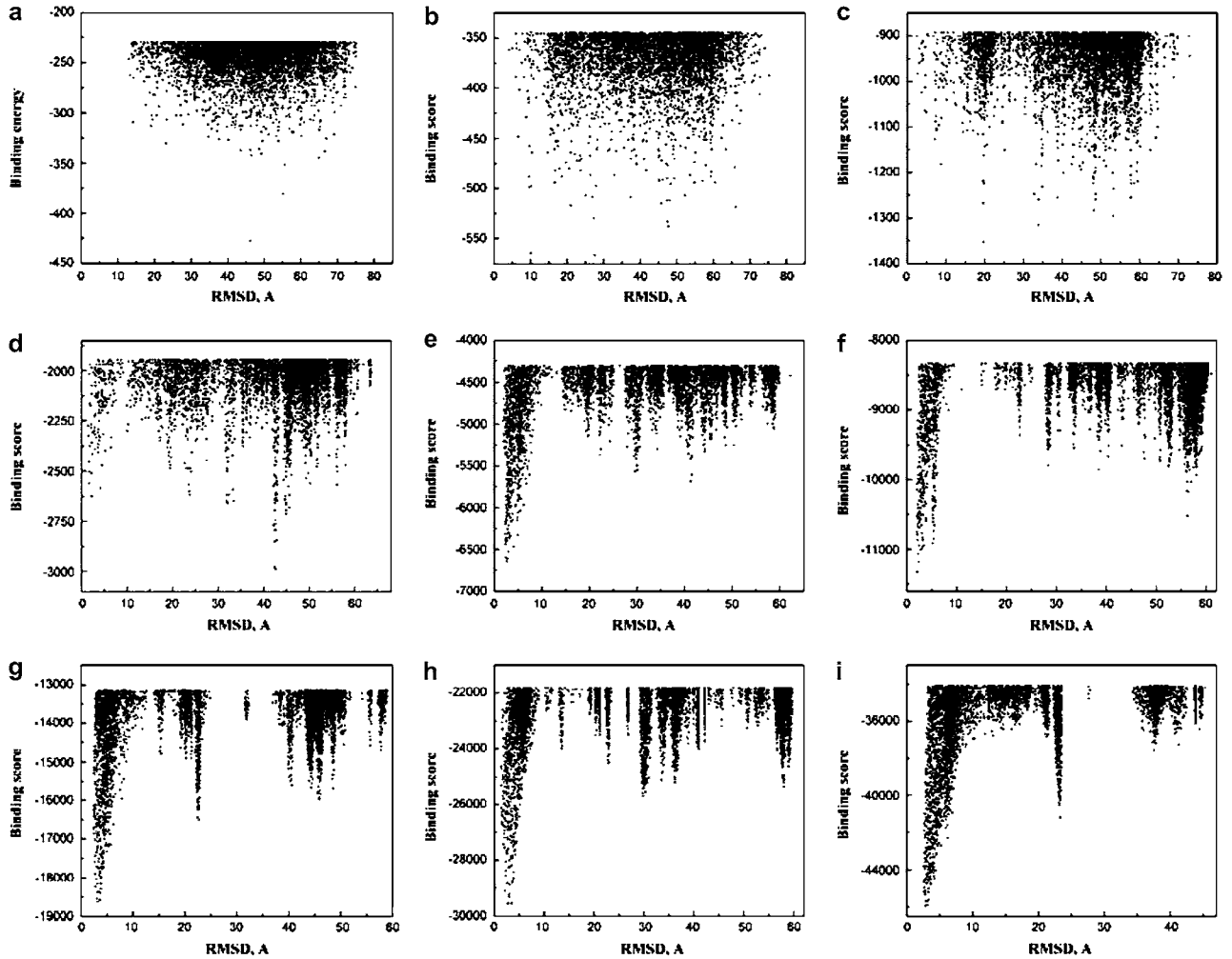

FIGURE 7 Energy landscape at different resolutions convoluted into 1D RMSD plot. Binding scores of 5000 lowest energy matches are shown as a function of the ligand interface C^{α} RMSD from the native position in 1bvn complex. Different resolutions of the landscape correspond to potential ranges of 3.4 Å (*a*), 4 Å (*b*), 5 Å (*c*), 6 Å (*d*), 7 Å (*e*), 8 Å (*f*), 9 Å (*g*), 10 Å (*h*), and 11 Å (*i*). The high-resolution landscape shows multiplicity of scattered minima at different RMSD values. The low-resolution landscape shows a distinct funnel in the small RMSD region.

indicates that the smoothing, although removing the high-frequency landscape fluctuations, preserves the underlying low-frequency characteristics (Fig. 1), which is important for an adequate description of the landscape.

Ruggedness and slope

Important characteristics of the energy landscape are its ruggedness and slope. We calculated average ruggedness as

$$\nu_1(R) = \frac{1}{92}\sum_{i=1}^{92}\frac{1}{N^{i}_{\mathrm{RMSD_I}\le 2}} \times \sum_{\substack{l,m,\\ l\neq m}}^{5000}{}' \left|\frac{E^{i}_{l}(R) - E^{i}_{m}(R)}{E^{i}_{m}(R)}\right| \theta(2 - RMSD^{i}_{I}(l,m)) \tag{4}$$

and average slope as

$$\nu_2(R) = \frac{1}{92}\sum_{i=1}^{92}\frac{1}{N^{i}_{\mathrm{RMSD_I}\le 2}} \times \sum_{\substack{l,m,\\ l\neq m}}^{5000}{}' \frac{\theta(2 - RMSD^{i}_{I}(l,m))}{RMSD^{i}_{I}(l,m)} \left|\frac{E^{i}_{l}(R) - E^{i}_{m}(R)}{E^{i}_{min}(R)}\right|, \tag{5}$$

of the energy landscapes for different potential ranges. Here $E^{i}_{l}(R)$ is the energy of match l in the conformational ensemble of complex i; $RMSD^{i}_{I}(l,m)$ is the interface RMSD between matches l and m; $\theta(x)$ is a θ-function, $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x < 0$; $E^{i}_{min}(R)$ is the minimal energy of the conformational ensemble i; $N^{i}_{\mathrm{RMSD_I}\le 2}$ is the number of pairs of conformations with $RMSD^{i}_{I}(l,m) \le 2A$ in the conformational ensemble i. Primes in sums over m and l in Eqs. 4 and 5

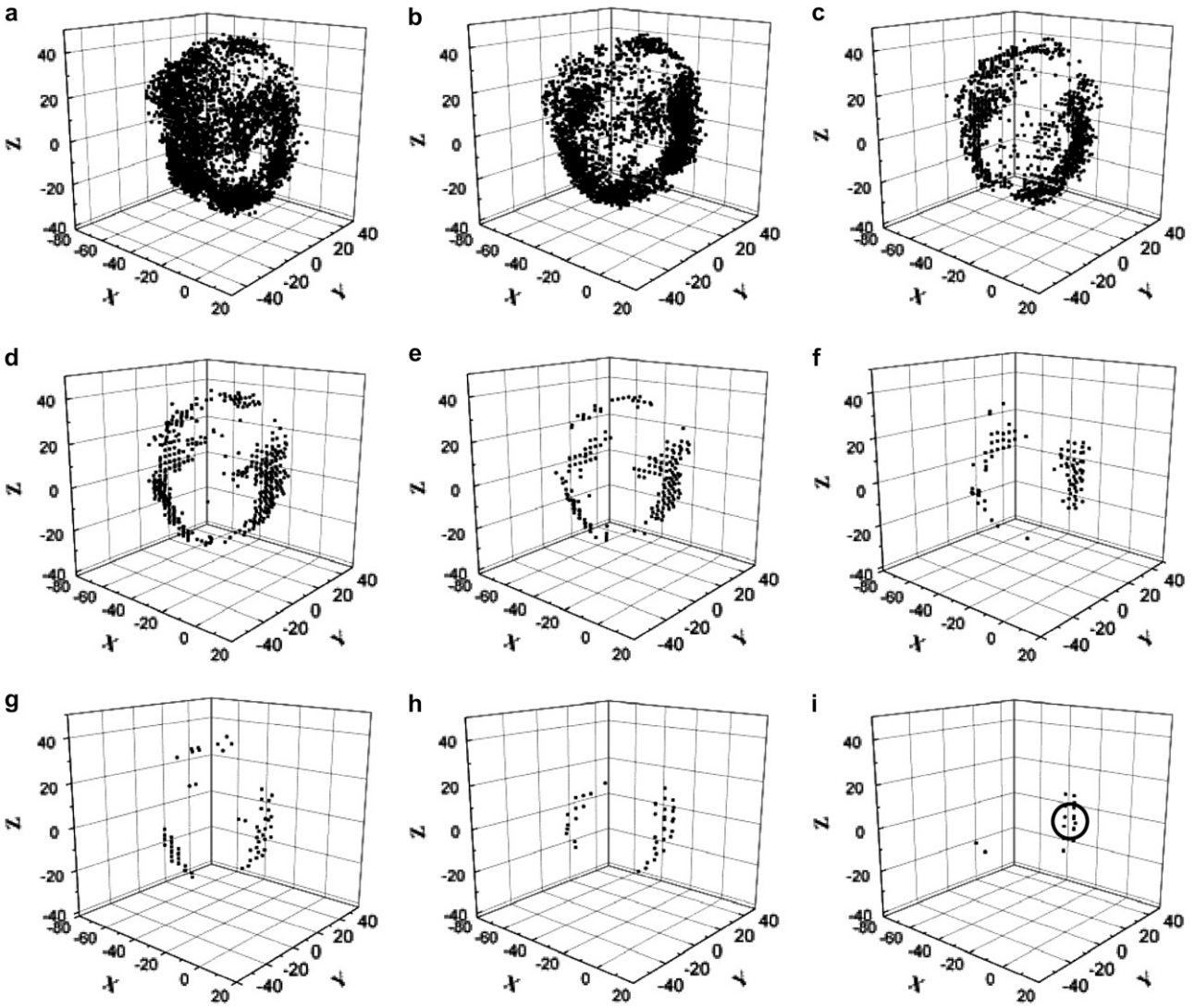

FIGURE 8 3D cross section through the energy landscape at different resolutions shown as distribution of minima (low-energy matches). The data shows positions of the ligand center of mass in 5000 lowest energy matches for 1bvn complex. Different resolutions of the landscape correspond to potential ranges of 3.4 Å (*a*), 4 Å (*b*), 5 Å (*c*), 6 Å (*d*), 7 Å (*e*), 8 Å (*f*), 9 Å (*g*), 10 Å (*h*) and 11 Å (*i*). The circle in panel (*i*) shows the cluster of near-native matches. Matches with different angular orientation of the ligand may have the same position of the ligand center of mass, the effect that significantly increases at lower resolutions (53). Thus the number of matches that appear in the 3D plot significantly decreases along the transition to lower resolution accordingly.

and in Eqs. 7–10 below indicate that a pair of matches l and m are considered one time only. Note that $\nu_1(R)$ is a dimensionless quantity, and $\nu_2(R)$ is measured in Å^{-1}. These values are independent of energy units because of the need of comparison between different potential ranges. To obtain the energy-dependent values, one can easily modify Eqs. 4 and 5 by multiplying the internal sums in Eqs. 4 and 5 by $E^{\text{i}}_{\text{m}}(R)$ and $E^{\text{i}}_{\text{min}}(R)$ accordingly. The definition of $\nu_1(R)$ is similar to the definition of the local ruggedness introduced in Ruvinsky and Vakser (30). The definition of $\nu_2(R)$ is closely related to intermolecular forces measured by chemical force microscopy (60–62). Indeed, the local slope can be calculated as

$$\nu_2 \sim \frac{E_{\text{l}} - E_{\text{m}}}{RMSD_{\text{l}}(l,m)} = -\frac{\Delta E}{\Delta X} = F, \tag{6}$$

where F is the force that acts on the ligand, attached to the cantilever of the force microscope, $RMSD_{\text{l}}(l,m) = \Delta X$, and $\Delta X$is the translational shift of the ligand.

The results show that the averaged ruggedness and the averaged slope decrease with the increasing of the potential range (Fig. 10), as expected in the smoothing of the energy landscape. To compare the ruggedness and the slope of the total landscape with the ruggedness and the slope of the deepest basins (putative binding funnels), we modified Eqs. 4 and 5 as

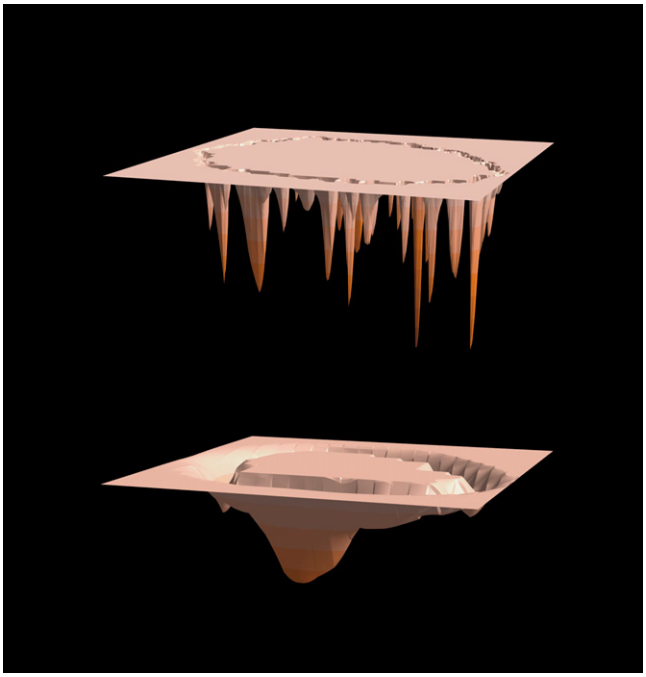

FIGURE 9 2D cross section through the energy landscape at different resolutions for 1bvn complex. As opposed to Figs. 7 and 8, which show only low-energy points, this plot shows all negative (attractive) values, corresponding to the position of the ligand center of mass. The high resolution of 1.7 Å (*upper plot*) corresponds to the unsmoothed, atomic resolution landscape; and the low resolution of 5.5 Å (*lower plot*) corresponds to the smoothed landscape. The large positive energy values inside the perimeter of the minima, which correspond to the overlap between ligand and receptor, are replaced by zero values for clarity.

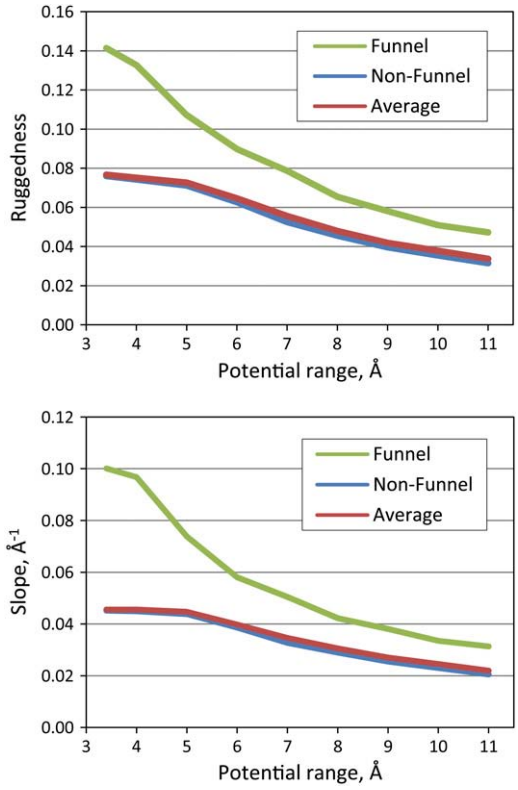

FIGURE 10 The averaged ruggedness and the slope of the energy landscape as a function of the resolution (potential range). See text for details.

$$\nu_1^{\mathrm{f}}(R) = \frac{1}{92}\sum_{\mathrm{i}=1}^{92}\frac{1}{N^{\mathrm{i}}_{\mathrm{RMSD_I}\le 2}}\sum_{\substack{\mathrm{l,m,}\\ \mathrm{l}\neq\mathrm{m}}}^{5000}{}'\left|\frac{E^{\mathrm{i}}_{\mathrm{l}}(R)-E^{\mathrm{i}}_{\mathrm{m}}(R)}{E^{\mathrm{i}}_{\mathrm{m}}(R)}\right|\theta(2-RMSD^{\mathrm{i}}_{\mathrm{I}}(l,m))\theta(7-RMSD^{\mathrm{i}}_{\mathrm{I}}(l))\theta(7-RMSD^{\mathrm{i}}_{\mathrm{I}}(m)) \tag{7}$$

and

$$\nu_2^{\mathrm{f}}(R) = \frac{1}{92}\sum_{\mathrm{i}=1}^{92}\frac{1}{N^{\mathrm{i}}_{\mathrm{RMSD_I}\le 2}}\sum_{\substack{\mathrm{l,m,}\\ \mathrm{l}\neq\mathrm{m}}}^{5000}\frac{\theta(2-RMSD^{\mathrm{i}}_{\mathrm{I}}(l,m))}{RMSD^{\mathrm{i}}_{\mathrm{I}}(l,m)}\left|\frac{E^{\mathrm{i}}_{\mathrm{l}}(R)-E^{\mathrm{i}}_{\mathrm{m}}(R)}{E^{\mathrm{i}}_{\mathrm{min}}(R)}\right|\theta(7-RMSD^{\mathrm{i}}_{\mathrm{I}}(l))\theta(7-RMSD^{\mathrm{i}}_{\mathrm{I}}(m)) \tag{8}$$

for the putative funnels (deepest basins), and

$$\nu_1^{\mathrm{nf}}(R) = \frac{1}{92}\sum_{\mathrm{i}=1}^{92}\frac{1}{N^{\mathrm{i}}_{\mathrm{RMSD_I}\le 2}}\sum_{\substack{\mathrm{l,m,}\\ \mathrm{l}\neq\mathrm{m}}}^{5000}{}'\left|\frac{E^{\mathrm{i}}_{\mathrm{l}}(R)-E^{\mathrm{i}}_{\mathrm{m}}(R)}{E^{\mathrm{i}}_{\mathrm{m}}(R)}\right|\theta(2-RMSD^{\mathrm{i}}_{\mathrm{I}}(l,m))\theta(RMSD^{\mathrm{i}}_{\mathrm{I}}(l)-7)\theta(RMSD^{\mathrm{i}}_{\mathrm{I}}(m)-7) \tag{9}$$

and

$$\nu_2^{\mathrm{nf}}(R) = \frac{1}{92}\sum_{\mathrm{i}=1}^{92}\frac{1}{N^{\mathrm{i}}_{\mathrm{RMSD_I}\le 2}}\sum_{\substack{\mathrm{l,m,}\\ \mathrm{l}\neq\mathrm{m}}}^{5000}{}'\frac{\theta(2-RMSD^{\mathrm{i}}_{\mathrm{I}}(l,m))}{RMSD^{\mathrm{i}}_{\mathrm{I}}(l,m)}\left|\frac{E^{\mathrm{i}}_{\mathrm{l}}(R)-E^{\mathrm{i}}_{\mathrm{m}}(R)}{E^{\mathrm{i}}_{\mathrm{min}}(R)}\right|\theta(RMSD^{\mathrm{i}}_{\mathrm{I}}(l)-7)\theta(RMSD^{\mathrm{i}}_{\mathrm{I}}(m)-7) \tag{10}$$

for the nonfunnel landscape. $RMSD_1(l)$ is the interface RMSD between match l and the match with the minimal energy in conformational ensemble i. The funnel size of 7 Å (54) is used in Eqs. 7–10. The data show that the putative funnels are more rugged and have a steeper slope than the rest of the landscape resolutions. These steeper gradients in the funnel guide the unbound proteins to the bound complex. The steeper gradients mean stronger forces, which suggests detection of the binding mode not only by the lowest energy but also by the strongest force, e.g., by in vitro probing of protein binding sites by force spectroscopy (60–62).

In our earlier studies of protein-protein energy landscapes (15), we defined ruggedness based on the density of matches in the basin (smaller ruggedness corresponds to higher density of matches). We determined that the density of matches in the larger, funnel-like basins is higher than in the nonfunnel basins. Thus, from such a perspective, the funnel is ''smoother'' than nonfunnel basins. Conversely, the current definition of ''ruggedness'' is based on the relative depth of minima. The above results show that such depth/gradient-related ruggedness of the funnels is larger than that of the nonfunnel landscape. The combination of the density-related ruggedness and depth-related ruggedness analyses indicates that a) in comparison with nonfunnel basins, the funnel has fewer energy barriers in its interior and/or these barriers are narrower (lower density-related ruggedness); and b) the funnel gradients are steeper (higher depth-related ruggedness).

Implications for docking

A number of protein-docking approaches implement a multistage/multiscale approach, where the initial global search is performed at lower resolution, followed by the local refinement to a higher resolution (44,63–66). The change of resolution is an essential part of the refinement. A major impediment to the refinement protocols is the uncertainties in the landscape transformation (44). Such uncertainties lead to the loss of the refinement trajectories, unnecessary oversampling, etc. Thus the quantitative description of the landscape change according to the resolution is important for designing refinement procedures for docking.

Another important implication for practical docking directly relates to the ruggedness and slope characteristics. The basin depth-related ruggedness and slope have not been utilized in protein docking. Among various approaches to funnel detection (e.g., knowledge based (67)), docking procedures typically use the energy/score of the top-ranked match and, optionally, the cluster occupancy, in part related to the density-related ruggedness. Thus the existing docking methods do not properly account for the basin shape. As a distinct property of the funnel, the depth-related ruggedness and the slope should complement the energy and the cluster occupancy in the docking funnel detection.

CONCLUSIONS

Studies of intermolecular energy landscapes are important for understanding protein association and adequate modeling of protein interactions. Landscape representation at different resolutions can be used for the refinement of docking predictions and the detection of macro characteristics, like the binding funnel. A representative set of protein-protein complexes was used to systematically map the intermolecular landscape by grid-based docking. The change of the resolution was achieved by varying the range of the potential according to the variable resolution GRAMM methodology. A formalism was developed to consistently parameterize the potential and describe essential characteristics of the landscape.

The results of gradual landscape smoothing, from high to low resolution, lead to the following major conclusions:

1. The number of energy basins, the landscape ruggedness, and the slope decrease accordingly.
2. The number of near-native matches, defined as those inside the funnel, increases, until the trend breaks down at the critical resolution. The rate of the increase and the critical resolution are specific to the type of a complex (enzyme-inhibitor, antigen-antibody, and other), reflect known underlying recognition factors, and correlate with earlier determined estimates of the funnel size.
3. The native/nonnative energy gap, a major characteristic of the energy minima hierarchy, remains constant.
4. The putative funnel (defined as the deepest basin) has the largest average depth-related ruggedness and slope at all resolutions.

The results facilitate better understanding of the binding landscapes and suggest directions for implementation in practical docking protocols.

APPENDIX

Let us consider a pair of proteins (Fig. 2) and an atom A in protein 2 at distance x from the surface of protein 1. The interaction energy between atom A and protein 1 atoms located at the distance r is

$$\varepsilon(x,r) = V(r)r^2\delta r\rho \int_0^{\theta_o} \sin\theta d\theta \int_0^{2\pi} d\phi = 2\pi\rho V(r)(r^2 - rx)\delta r, \tag{11}$$

where δr is the thickness of the spherical layer in protein 1, ρ is protein 1 atom density, and $\theta_o = \arccos(x/r)$ (Fig. 2). For simplicity, we assumed that potential $V(r)$ does not depend on atom type. The total energy of atom A is

$$\varepsilon(x) = \int_x^{2R} \varepsilon(x,r)dr = 2\pi\rho \int_x^{2R} V(r)(r^2 - rx)dr, \tag{12}$$

where $2R$ is the interaction cutoff and a is the closest distance between proteins. Substituting

$$V(r) = \begin{cases} U, & 0 \le r < R \\ -1, & R \le r < 2R \\ 0, & 2R \le r \end{cases} \tag{13}$$

into Eq. 12, we perform integration over r for $x = a$. Allowing clash contacts to override structural differences between unbound and bound structures, we use $a = 1.3$ Å. In Eq. 12 we assumed that the interface, restricted by polar and azimuth angles, is flat (Fig. 2). The energy of a surface atom A is

$$\varepsilon(a) = \int_a^{2R} \varepsilon(a,r)dr = 2\pi\rho \int_a^{2R} V(r)(r^2 - ra)dr = U\frac{R^3}{3}\left[1 - \frac{3a}{2R} + \frac{a^3}{2R^3}\right] - \frac{R^3}{3}\left[7 - \frac{9a}{2R}\right]. \quad (14)$$

I.V. acknowledges undergraduate research of Lei Wang and Amina Sharaf at Stony Brook University, which led to a better understanding of the landscape changes.

The study was supported by National Institutes of Health grant R01 GM074255.

REFERENCES

1. Onuchic, J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Curr. Opin. Struct. Biol.* 14:70–75.
2. Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science.* 267:1619–1620.
3. Levy, Y., S. S. Cho, J. N. Onuchic, and P. G. Wolynes. 2005. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* 346:1121–1145.
4. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA.* 84:7524–7528.
5. Miller, D. W., and K. A. Dill. 1997. Ligand binding to proteins: the binding landscape model. *Protein Sci.* 6:2166–2179.
6. Kumar, S., B. Ma, C.-J. Tsai, N. Sinha, and R. Nussinov. 2000. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* 9:10–19.
7. Tovchigrechko, A., and I. A. Vakser. 2001. How common is the funnel-like energy landscape in protein-protein interactions? *Protein Sci.* 10:1572–1583.
8. Minh, D. D. L., J. M. Bui, C. E. Chang, T. Jain, J. M. J. Swanson, and J. A. McCammon. 2005. The entropic cost of protein-protein association: a case study on acetylcholinesterase binding to fasciculin-2. *Biophys J.* 89:L25–L27.
9. Wolynes, P. G. 2005. Recent successes of the energy landscape theory of protein folding and function. *Q. Rev. Biophys.* 38:405–410.
10. Dill, K. A. 1999. Polymer principles and protein folding. *Protein Sci.* 8:1166–1180.
11. Tsai, C.-J., S. Kumar, B. Ma, and R. Nussinov. 1999. Folding funnels, binding funnels, and protein function. *Protein Sci.* 8:1181–1190.
12. Camacho, C. J., Z. Weng, S. Vajda, and C. DeLisi. 1999. Free energy landscapes of encounter complexes in protein-protein association. *Biophys. J.* 76:1166–1178.
13. Camacho, C. J., and S. Vajda. 2001. Protein docking along smooth association pathways. *Proc. Natl. Acad. Sci. USA.* 98:10636–10641.
14. Vakser, I. A., O. G. Matar, and C. F. Lam. 1999. A systematic study of low-resolution recognition in protein-protein complexes. *Proc. Natl. Acad. Sci. USA.* 96:8477–8482.
15. O'Toole, N., and I. A. Vakser. 2008. Large-scale characteristics of the energy landscape in protein-protein interactions. *Proteins.* 71:144–152.
16. Wang, J., and G. M. Verkhivker. 2003. Energy landscape theory, funnels, specificity and optimal criterion of biomolecular binding. *Phys. Rev. Lett.* 90:188101.
17. Wang, J., L. Xu, and E. K. Wang. 2007. Optimal specificity and function for flexible biomolecular recognition. *Biophys. J.* 92:L109–L111.
18. Wang, J., X. Zheng, Y. Yang, D. Drueckhammer, W. Yang, G. Verkhivker, and E. K. Wang. 2007. Quantifying intrinsic specificity: a potential complement to affinity in drug screening. *Phys. Rev. Lett.* 99:198101.
19. Wang, J., W. M. Huang, H. Y. Lu, and E. K. Wang. 2004. Downhill kinetics of biomolecular interface binding-global connected scenario. *Biophys. J.* 87:2187–2194.
20. Wang, J., K. Zhang, H. Y. Lu, and E. K. Wang. 2006. Quantifying the kinetic paths of flexible biomolecular recognition. *Biophys. J.* 91:866–872.
21. Wang, J., K. Zhang, H. Y. Lu, and E. K. Wang. 2006. Identifying the kinetic paths on biomolecular binding-folding energy landscape. *Phys. Rev. Lett.* 96:168101.
22. Wang, J., Q. Lu, and P. Lu. 2006. Single-molecule dynamics reveals cooperative binding-folding in protein recognition. *PLoS Comput. Biol.* 2:e78.
23. Lu, Q., P. H. Lu, and J. Wang. 2007. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. *Phys. Rev. Lett.* 98:128105.
24. Ferreiro, D. U., J. A. Hegler, E. A. Komives, and P. G. Wolynes. 2007. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA.* 104:19819–19824.
25. Shea, J. E., J. N. Onuchic, and C. L. Brooks. 1999. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA.* 96: 12512–12517.
26. Sutto, L., J. Latzer, J. A. Hegler, D. U. Ferreiro, and P. G. Wolynes. 2007. Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc. Natl. Acad. Sci. USA.* 104:19825–19830.
27. Hyeon, C., and D. Thirumalai. 2003. Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proc. Natl. Acad. Sci. USA.* 100:10249–10253.
28. Vakser, I. A. 1996. Long-distance potentials: an approach to the multiple-minima problem in ligand-receptor interaction. *Protein Eng.* 9:37–41.
29. Vakser, I. A. 1997. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins.* (Suppl. 1): 226–230.
30. Ruvinsky, A. M., and I. A. Vakser. 2008. Interaction cutoff effect on ruggedness of protein-protein energy landscape. *Proteins.* 70:1498–1505.
31. Vakser, I. A., and C. Aflalo. 1994. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins.* 20:320–329.
32. Kostrowicki, J., and H. A. Scheraga. 1992. Application of the diffusion equation method for global optimization to oligopeptides. *J. Phys. Chem.* 96:7442–7449.
33. Head-Gordon, T., and F. H. Stillinger. 1993. Predicting polypeptide and protein structures from amino acid sequence: antlion method applied to melittin. *Biopolymers.* 33:293–303.
34. Hart, R. K., R. V. Pappu, and J. W. Ponder. 2000. Exploring the similarities between potential smoothing and simulated annealing. *J. Comput. Chem.* 21:531–552.
35. Stillinger, F. H., and D. K. Stillinger. 1990. Cluster optimization simplified by interaction modification. *J. Chem. Phys.* 93:6106–6107.
36. Braier, P. A., R. S. Berry, and D. J. Wales. 1990. How the range of pair interactions governs features of multidimensional potentials. *J. Chem. Phys.* 93:8745–8756.
37. Kostrowicki, J., L. Piela, B. Cherayil, and H. A. Scheraga. 1991. Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms. *J. Phys. Chem.* 95:4113–4119.
38. Amara, P., D. Hsu, and J. E. Straub. 1993. Global energy minimum searches using an approximate solution of the imaginary time Schroedinger equation. *J. Phys. Chem.* 97:6715–6721.
39. Doye, J. P. K., and D. J. Wales. 1996. The effect of the range of the potential on the structure and stability of simple liquids: from clusters to bulk, from sodium to C60. *J. Phys. B Atom. Mol. Opt. Phys.* 29: 4859–4894.

40. Miller, M. A., J. P. K. Doye, and D. J. Wales. 1999. Structural relaxation in Morse clusters: energy landscapes. *J. Chem. Phys.* 110: 328–334.

41. Whitfield, T. W., and J. E. Straub. 2002. Gravitational smoothing as a global optimization strategy. *J. Comput. Chem.* 23:1100–1102.

42. Wawak, R. J., M. M. Wimmer, and H. A. Scheraga. 1992. Application of the diffusion equation method of global optimization to water clusters. *J. Phys. Chem.* 96:5138–5145.

43. Vakser, I. A., and S. Jiang. 2002. Strategies for modeling the interactions of the transmembrane helices of G-protein coupled receptors by geometric complementarity using the GRAMM computer algorithm. *Methods Enzymol.* 343:313–328.

44. Pappu, R. V., G. R. Marshall, and J. W. Ponder. 1999. A potential smoothing algorithm accurately predicts transmembrane helix packing. *Nat. Struct. Biol.* 6:50–55.

45. Tovchigrechko, A., C. A. Wells, and I. A. Vakser. 2002. Docking of protein models. *Protein Sci.* 11:1888–1896.

46. Wales, D. J., and H. A. Scheraga. 1999. Global optimization of clusters, crystals, and biomolecules. *Science.* 285:1368–1372.

47. Wales, D. J. 2005. Energy landscapes and properties of biomolecules. *Phys. Biol.* 2:S86–S93.

48. Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.

49. Douguet, D., H. C. Chen, A. Tovchigrechko, and I. A. Vakser. 2006. Dockground resource for studying protein-protein interfaces. *Bioinformatics.* 22:2612–2618.

50. Gao, Y., D. Douguet, A. Tovchigrechko, and I. A. Vakser. 2007. Dockground system of databases for protein recognition studies: unbound structures for docking. *Proteins.* 69:845–851.

51. Katchalski-Katzir, E., I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, and I. A. Vakser. 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA.* 89:2195–2199.

52. Vakser, I. A. 1995. Protein docking for low-resolution structures. *Protein Eng.* 8:371–377.

53. Vakser, I. A. 1996. Low-resolution docking: prediction of complexes for underdetermined structures. *Biopolymers.* 39:455–464.

54. Hunjan, J., A. Tovchigrechko, Y. Gao, and I. A. Vakser. 2008. The size of the intermolecular energy funnel in protein-protein interactions. *Proteins.* 72:344–352.

55. Nicola, G., and I. A. Vakser. 2007. A simple shape characteristic of protein-protein recognition. *Bioinformatics.* 23:789–792.

56. Nayal, M., and B. Honig. 2006. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins.* 63:892–906.

57. Binkowski, T. A., A. Joachimiak, and J. Liang. 2005. Protein surface analysis for function annotation in high-throughput structural genomics pipeline. *Protein Sci.* 14:2972–2981.

58. Guo, Z., D. Thirumalai, and J. Honeycutt. 1992. Folding kinetics of proteins: a model study. *J. Chem. Phys.* 97:525–535.

59. Karplus, M., and E. Shakhnovich. 1992. Protein folding: theoretical studies of thermodynamics and dynamics. *In* Protein Folding. T. E. Creighton, editor. W. H. Freeman & Co., New York. 127–195.

60. Weisel, J. W., H. Shuman, and R. I. Litvinov. 2003. Protein–protein unbinding induced by force: single-molecule studies. *Curr. Opin. Struct. Biol.* 13:227–235.

61. Yang, Y., H. Wang, and D. A. Erie. 2003. Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods.* 29:175–187.

62. Ritort, F. 2006. Single-molecule experiments in biological physics: methods and applications. *J. Phys. Condens. Matter.* 18:R531–R583.

63. Gray, J. J., S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker. 2003. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* 331:281–299.

64. Tovchigrechko, A., and I. A. Vakser. 2005. Development and testing of an automated approach to protein docking. *Proteins.* 60:296–301.

65. Carter, P., V. I. Lesk, S. A. Islam, and M. J. E. Sternberg. 2005. Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. *Proteins* 60:281–288.

66. Li, L., R. Chen, and Z. Weng. 2003. RDOCK: refinement of rigid-body protein docking predictions. *Proteins.* 53:693–707.

67. London, N., and O. Schueler-Furman. 2007. Assessing the energy landscape of CAPRI targets by FunHunt. *Proteins.* 69:809–815.