Using the Concept of Transient Complex for Affinity Predictions in CAPRI Rounds 20–27 and Beyond

Sanbo Qin and **Huan-Xiang Zhou**[*]
Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

Abstract

Predictions of protein-protein binders and binding affinities have traditionally focused on features pertaining to the native complexes. In developing a computational method for predicting protein-protein association rate constants, we introduced the concept of transient complex after mapping the interaction energy surface. The transient complex is located at the outer boundary of the bound-state energy well, having near-native separation and relative orientation between the subunits but not yet formed most of the short-range native interactions. We found that the width of the binding funnel and the electrostatic interaction energy of the transient complex are among the features predictive of binders and binding affinities. These ideas were very promising for the five affinity-related targets (T43–45, 55, and 56) of CAPRI rounds 20–27. For T43, we ranked the single crystallographic complex as number 1 and were one of only two groups that clearly identified that complex as a true binder; for T44, we ranked the only design with measurable binding affinity as number 4. For the nine docking targets, continuing on our success in previous CAPRI rounds, we produced 10 medium-quality models for T47 and acceptable models for T48 and T49. We conclude that the interaction energy landscape and the transient complex in particular will complement existing features in leading to better prediction of binding affinities.

Keywords

transient complex; interaction energy landscape; binding affinity; protein docking; protein association

INTRODUCTION

The predictions of binders and binding affinities are important themes in the study of protein-protein interactions. Recent years have seen significant progress in the prediction of binders[1,2] and growing efforts in the prediction of binding affinities.[3–6] These predictions have mostly been based on structural and energetic features pertaining to the native complexes. Potentially the full interaction energy surface of the two subunits encodes information for predicting binders and binding affinities. Here we explore this potential and report our findings for affinity-related targets in CAPRI rounds 20–27.

The inter-protein interaction energy surface is not directly observable but it determines the observable entities such as the structure of the native complex, whether the two subunits have measurable binding affinity (i.e., classification as either binders or non-binders), and if so the magnitude of the binding affinity and the association and dissociation rate constants. The potential value of the full interaction energy surface for predicting such observable

[*]Corresponding author; hzhou4@fsu.edu.

entities has been demonstrated in previous studies. Shen et al.[7] used semi-definite underestimation to search for local minima on the interaction energy surface and to locate the native complex (corresponding to the global minimum). They suggested that this search process mimics the diffusional approach of the subunits in forming the native complex.

The association rate constant is certainly dependent on the full interaction energy surface. In developing a theory for the association rate constant, we mapped the interaction energy surface to introduce the concept of transient complex.[8,9] This is an intermediate on the association pathway, located at the outer boundary of the native-complex energy well (Fig. 1). It separates the "near" region, where the two subunits have significant native interactions but their relative translation and rotation are restricted, and the "far" region, where the two subunits have few specific interactions but have nearly unrestricted translational and rotational freedom. The two subunits reach the transient complex by translational and rotational diffusion. Subsequently they can undergo further conformational rearrangements and tightening of the binding interface to reach the native complex. When this forward step is fast (relative to the breakup of the transient complex by diffusion in the reverse direction), the overall association rate constant is well approximated by the rate constant for reaching the transient complex by diffusion. We have developed an automated method called TransComp (http://pipe.sc.fsu.edu/transcomp/) for predicting the association rate constant in this diffusion-limited regime.[10]

Here we report the use of the interaction energy landscape and the transient complex in particular for the five affinity-related targets (T43–45, 55, and 56) of CAPRI rounds 20–27. We found that the width of the binding funnel and the electrostatic interaction energy of the transient complex are among the features predictive of binders and binding affinities. For the nine docking targets, continuing on our success in previous CAPRI rounds,[11,12] we produced 10 medium-quality models for T47 and acceptable models for T48 and T49.

METHODS

Mapping of interaction energy surface and generation of transient complex

The procedure for mapping the interaction energy surface and identifying the transient complex was published previously.[8–10] Briefly, the mapping involved sampling in the 6-dimensional space of relative translation and relative rotation between the subunits (Fig. 1A). The sampling covered the native-complex basin and the surrounding region. Each subunit was treated as rigid and adopted the native conformation. The 3 translational coordinates were specified by the displacement vector $\mathbf{r}$ between the centers of the binding sites on the two subunits. The 3 rotational coordinates were specified by a unit vector $\mathbf{e}$ attached to subunit B and the rotation angle around this unit vector. In the native complex, this unit vector is perpendicular to the least-squares plane of the interface atoms and points from subunit A to subunit B. Poses in this 6-dimensional space was randomly generated, with the restriction that the magnitude of $\mathbf{r}$ (i.e., r) was below a cutoff r_{cut}.

Of all the randomly generated poses, the clash-free ones were saved. For each of these, the number of contacts, N_{c}, between interaction locus atoms was calculated (Fig. 1B). The interaction locus atoms were selected from interface atoms and came in cognate pairs. In each clash-free pose, contacts formed between cognate atoms were denoted as native and between non-cognate atoms as nonnative. Both were counted in calculating N_{c}. For poses at a given N_{c}, the standard deviation, , of was calculated (Fig. 1C). The dependence of on N_{c} was used to identify the transient complex. An earlier scheme relied on how grew with decreasing N_{c};[9] in the subsequent, automated implementation, the dependence of on N_{c} was fitted to a function for modeling protein denaturation data (Fig. 1C).[10] The midpoint of the transition from the native-complex basin (with high N_{c} and low) to the far region

(with low N_c but high   ) was identified as the transient complex. That is, all the poses with the midpoint N_c (designated as N_c^*) constituted the transient complex.

The width of the binding funnel (see Fig. 1D) can be measured in different ways. We used this parameter for targets T55 and T56, adopting a very simple measure. This was the fraction f_{cf}, of clash-free poses among all the randomly generated poses with $r_{cut} = 6$ Å.

Calculation of electrostatic interaction energy

Except for targets T55 and T56, the electrostatic interaction free energy was calculated by solving the nonlinear Poisson-Boltzmann equation. We used the APBS solver,[13] following a protocol described previously.[10] For the native complex, the electrostatic interaction free energy is

$$\Delta G_{el} = G_{el}(C) - G_{el}(A) - G_{el}(B) \quad \text{(1)}$$

where $G_{el}(C)$, $G_{el}(A)$, and $G_{el}(B)$ are the electrostatic free energies of the complex and their two subunits. The electrostatic interaction free energy of the transient complex, G_{el}^*, was calculated similarly for each pose within the transient-complex ensemble, and then averaged over 100 or 10 representative poses.

For target T55 and T56, because the large amounts of calculations needed to deal with the ~1000 mutants, we used the simple Debye-Hückel potential:

$$\Delta G_{el} = 332 \sum_{i,j} \frac{q_i q_j e^{-\kappa r_{ij}}}{\varepsilon r_{ij}} \quad \text{(2)}$$

where q_i and q_j are the atomic partial charges of subunits A and B, respectively, r_{ij} are the distances between atoms,   = 78.5 is the dielectric constant of water, and   is the Debye-Hückel screening parameter (at ionic strength 0.15 M).

Method for docking targets

Model generation and selection largely followed our previous work.[11,12] For targets where only a homologous template of a subunit was provided, the structure of the subunit was built by using MODELLER 8v2.[14] Except for T47 and T57, all docking poses were generated by ZDOCK 2.3.[15] The docking poses were then selected according to biochemical information if available. The poses were clustered and representatives of clusters were manually inspected. The final selected ten models were subjected to energy minimization by the AMBER program (including 50 steps of steepest descent). Models were generated by homology modeling and HADDOCK[16] for T47 and by Autodock Vina[17] for T57.

RESULTS AND DISCUSSION

CAPRI rounds 20–27 had five affinity-related targets: T43–45, 55, and 56; and nine docking targets: T46-T51, T53, T54, and T58. Below we briefly describe our performance and what we have learned from these exercises.

Native complexes generally have favorable electrostatic interactions

Before going into CAPRI rounds 20–27, we knew that electrostatic interactions between subunits in native complexes are generally favorable (i.e., $G_{el} < 0$). While testing the robustness of our automated TransComp method for predicting association rate constants, we ran this method on the 176 complexes in benchmark 4.0 of Hwang et al.,[18] and

successfully completed these runs for 132 cases (Supplemental Table S1 of Qin et al.[10]). The distribution of the electrostatic interaction free energies (ΔG_{el}^*) in the transient complexes peaks at −0.5 kcal/mol (Fig. 2A). The favorable electrostatic interactions are even more significant in the native complexes (Fig. 2B), with the distribution peaking at −2.5 kcal/mol. However, native complexes sometimes can contain unphysically close contacts (due to low resolution of the structure or poor quality of modeling), which may lead to spurious results for ΔG_{el}. One such case actually occurred among the 132 complexes (the resulting spurious ΔG_{el} was not included in Fig. 2B). Because ΔG_{el}^* and ΔG_{el} show reasonable correlation ($R^2 = 0.48$; Fig. 2C) and the subunits in the transient complex are usually separated by a layer of solvent such that there is less chance for unphysically close contacts, whenever possible we chose to use ΔG_{el}^* for the affinity-related CAPRI targets.

Of course van der Waals interactions should also be favorable in native complexes and may even dominate over electrostatic interactions. However, given their strong dependence on interatomic distances, we reasoned that they might not be particularly useful for discriminating between binders and non-binders among designed complexes. In particular, the effects of subunit conformations, poor contacts, and the neglect of water molecules on van der Waals interaction energies would be very unpredictable. Indeed, while many groups found electrostatic interactions to be useful for the affinity-related targets,[19,20] the same cannot be said about van der Waals interactions.

The finding of generally favorable electrostatic interactions between the subunits is in contrast to the previous conclusion of Sheinerman et al.[21] that "the total effect of electrostatics is generally net destabilizing" for protein-protein complexes. We have noticed that the calculated electrostatic free energy from solving the Poisson-Boltzmann equation is very sensitive to the choice of the boundary between the protein low dielectric and the solvent high dielectric.[22,23] The popular choice of using the molecular surface as the dielectric boundary indeed generally produces unfavorable electrostatic contributions to protein folding and binding. However, the alternative choice of using the van der Waals surface typically reverses the sign of ΔG_{el}, thus predicting a stabilizing effect on binding stability for electrostatic interactions,[24–26] consistent with the results shown in Fig. 2B.

Our systematic assessment of calculated effects of charge mutations on protein folding and binding stability led us to conclude that the use of the van der Waals surface as the dielectric boundary produces better agreement with experimental results.[22–27] Moreover, significant electrostatic enhancements of rate constants observed for many protein complexes can only be produced by the use of the van der Waals surface.[9,28] The Poisson-Boltzmann equation-based electrostatic calculations carried out here followed this protocol.

Target T43

For T43 we were given 20 models generated by Rosetta along with a blinded crystal structure (also based on Rosetta design). We generated transient complexes for these models and calculated the electrostatic interaction free energies of the transient complexes. The results, sorted according to ΔG_{el}^*, are listed in Table I. The top-ranked Model 10 has a significantly more favorable ΔG_{el}^* value (at −7.02 kcal/mol) than the other 20 models (closest ΔG_{el}^* at −4.95 kcal/mol, for Model 9).

Model 10 turned out to be the crystal structure (Protein Data Bank ID 3Q9N).[29] (Only one other group clearly identified Model 10 as a true binder.) The two subunits feature very strong electrostatic complementarity at the interface (Fig. 3A), thus explaining the strong favorable ΔG_{el*}. Our TransComp web server predicted an association rate constant of 4.6×10^8 M^{-1}s^{-1}, with over three orders of magnitude electrostatic rate enhancement. Karanicolas et al.[29] reported a rate constant of $(7–9) \times 10^5$ M^{-1}s^{-1} for a precursor of 3Q9N (before

affinity maturation) using surface plasmon resonance (SPR). However, this experimental technique is limited by mass transport, precluding accurate determination of rate constants higher than about 10^6 $M^{-1}s^{-1}$.[28] The affinity maturation may have further improved the association rate constant, as Karanicolas et al. noted large contribution of charged residue mutations during the affinity maturation. They also noted high salt sensitivity of the binding free energy. These observations all point to a significant role of electrostatic interactions, consistent with our calculations.

Target T44

There were also 21 Rosetta-generated models for T44. We again ranked the models according to G_{el}^* (Table I). This time the G_{el}^* values were all moderate; the top-ranked model had $G_{el}^* = -0.88$ kcal/mol. After the ranking of the models was submitted, the CAPRI participants were informed by the Baker group that Model 2 had measurable binding affinity. We ranked this model as 4th.

Examining of the electrostatic surfaces of the subunits in Model 2 shows that one features a strong positive patch and the other features a strong negative patch (Fig. 3B). However, in the model of the complex generated by Rosetta, these strong complementary patches are not placed within the interface. Rather the interface is positioned at the peripheries of both patches, explaining the moderate G_{el}^* for this model (at -0.56 kcal/mol). It is possible that in the complex actually formed the complementary electrostatic patches on the two subunits are placed in the interface.

Target T45

The goal for this target was to discriminate between Rosetta-designed interfaces from natural complexes. The 87 designed interfaces had favorable computed binding energies but did not show measurable binding affinities; the 120 natural complexes were from benchmark 3.0 of Hwang et al.[30] Again we used G_{el}^* as the role parameter for ranking, with modest success. Our AUC value, 0.69, falls in the lower mid-range among the 28 participating groups (full AUC range is between 0.55 and 0.86).[19]

Our retrospect comparison shows that the designs and natural complexes do show distinct distributions in G_{el}^*, with a significant overall shift toward higher G_{el}^* for the designs. However, the two distributions also have significant overlap over the range of G_{el}^* from -2 to 3 kcal/mol. This suggests that, while electrostatic calculations are useful for characterizing natural complexes (e.g., in the regulation of binding affinity and association rate constant), they have limited ability in discriminating decoy interfaces from natural ones.

Targets T55 and T56

The goal of these two targets was to test the ability of predicting how single mutations affect the binding affinities of two designed protein inhibitors of Influenza hemagglutinin, HB36.4 and HB80.3. The experimental data were provided by the enrichment of mutant sequences in a selection for hemagglutinin binding;[31] the enrichment data were used as the proxy for mutational effects on binding affinity in lieu of direct affinity measurements.

To predict the effect of each mutation, we used an energy function with the combination

$$\Delta G = \Delta G_{el} + \ln f_{cf} \quad (3)$$

where G_{el} is the electrostatic energy calculated according to the Debye-Hückel potential [Eq (2)], and f_{cf} is the fraction of clash-free poses in generating the transient complex. The

latter quantity was expected to capture the influence on the binding affinity by the width of the binding funnel. Our performance on T55 and T56 was reasonably successful. Evaluated according to Kendall's tau-b coefficient, among 22 participating groups, we ranked 7th for T55 and 6th for T56.[20]

We carried out retrospective analysis to assess the relative contributions of the two terms of Eq (3) (Table II). The terms appear to be synergistic in their impact on prediction accuracy; Kendall's tau-b coefficient for the combined energy function is much higher than that for either term. Interestingly, the contribution of f_{cf} seems more significant for T55 whereas that of G_{el} seems more significant for T56.

Docking targets

We submitted 10 medium-quality models for T47, 3 acceptable models for T48 and 1 acceptable model for T49. T47 is the complex of DNase E2 with immunity protein Im2;[32] the CAPRI participants were told that interface water was to be the focus of prediction. We generated docking poses by HADDOCK, using as input 11 homology models of E2 based on the structure of E9 in 2WPT (for the E9-Im2 complex) and 60 models of Im2 in the NMR structure 2NO8 along with the structure of Im2 in 2WPT. In two independent HADDOCK runs, residues K83, F86, and R98 on E2 and D33, R38, E41, and R42 on IM2 were defined as active restraints, based on mutational studies of Li et al.[33] 10 models were selected based on clustering and manual inspection, with all water molecules were stripped. To each of these models, we added 10 interface water molecules conserved between the structures of the E9-Im9 complex (1EMV) and the E9-Im2 complex (2WPT). The final models were subject to 200 steps of energy minimization before submission.

T48 and T49 are the complex of T4moH and T4moC, components of the toluene 4-monooxygenase holoenzyme. T4moH was taken from 3DHH for T48 and was an unpublished unbound structure for T49. The active site in T4moH is buried, with three tunnels leading to the exterior.[34] In identifying a possible binding site for T4moC, we focused on the largest channel of the three. For convenience, we used a residue, S187, near the center of this channel as a representative. We chose models in which the iron-sulfur cluster of T4moC is close to S187. Specifically, the final 10 models all have < 10 Å distances between T4moH S187 and T4moC H47 or H67, which coordinate the iron-sulfur cluster.

T58 serves as an example for the docking targets for which we failed to submit at least an acceptable model. This is the complex of the SalG lysozyme with the inhibitor PliG.[35] Mutation studies of Leysen et al.[36] suggested that mutations on PliG Y47 and R119 significantly reduced inhibitory activity; these authors also noted that these residues are located in an area predicted by our meta-PPISP web server.[37] Our model selection thus placed SalG lysozyme active-site residues E73, D86, and D97 and PliG Y47 and R119 in the interface, assuming that the inhibitor completely blocked the substrate-binding site. It turned out that in the actual complex (4G9S)[35] the inhibitor blocks only half of the substrate-binding site, from a sideway direction. In retrospect our assumption is untenable, since the substrate-binding site is too shallow to hold the inhibitor from a "head-on" direction, so we had to pull the inhibitor away to avoid clash, leading to a physically unreasonable interface.

Additional work with affinity prediction

Partly due to the encouraging results on targets T55 and T56, we attempted to develop a predictor for protein-protein binding affinities, based on linear regression analysis of features defining the full protein-protein interaction surface. For a subset of 33 complexes from an affinity benchmark[3] that have relatively small differences between bound and

unbound structures (I-RMSD < 1 Å), we found a linear combination of parameters, akin to Eq (3), that strongly correlates with the binding free energies ($R^2 = 0.64$). The parameters included the number of nonnative contacts (part of N_c defined above) in the native complex, the curvature of the binding funnel, the largest gap in N_c in the configurational sampling to generate the transient complex, and I-RMSD. We conclude that the interaction energy landscape and the transient complex in particular will complement existing features[3–6] in leading to better prediction of binding affinities.

References

1. Aloy P, Russell RB. Interrogating protein interaction networks through structural biology. Proc Natl Acad Sci U S A. 2002; 99:5896–5901. [PubMed: 11972061]

2. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B. Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature. 2012; 490:556–560. [PubMed: 23023127]

3. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. A structure-based benchmark for protein-protein binding affinity. Protein Sci. 2011; 20:482–491. [PubMed: 21213247]

4. Moal IH, Agius R, Bates PA. Protein-protein binding affinity prediction on a diverse set of structures. Bioinformatics. 2011; 27:3002–3009. [PubMed: 21903632]

5. Malod-Dognin N, Bansal A, Cazals F. Characterizing the morphology of protein binding patches. Proteins. 2012; 80:2652–2665. [PubMed: 22806945]

6. Vreven T, Hwang H, Pierce BG, Weng Z. Prediction of protein-protein binding free energies. Protein Sci. 2012; 21:396–404. [PubMed: 22238219]

7. Shen Y, Paschalidis I, Vakili P, Vajda S. Protein docking by the underestimation of free energy funnels in the space of encounter complexes. PLoS Comput Biol. 2008; 4:e1000191. [PubMed: 18846200]

8. Alsallaq R, Zhou HX. Energy landscape and transition state of protein-protein association. Biophys J. 2007; 92:1486–1502. [PubMed: 17142273]

9. Alsallaq R, Zhou HX. Electrostatic rate enhancement and transient complex of protein-protein association. Proteins. 2008; 71:320–335. [PubMed: 17932929]

10. Qin S, Pang X, Zhou HX. Automated prediction of protein association rate constants. Structure. 2011; 19:1744–1751. [PubMed: 22153497]

11. Qin S, Zhou HX. A holistic approach to protein docking. Proteins. 2007; 69:743–749. [PubMed: 17803232]

12. Qin S, Zhou HX. Selection of near-native poses in CAPRI rounds 13–19. Proteins. 2010; 78:3166–3173. [PubMed: 20589642]

13. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001; 98:10037–10041. [PubMed: 11517324]

14. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using Modeller. Curr Protoc Protein Sci. 2007; 50:2.9.1–2.9.31.

15. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins. 2003; 52:80–87. [PubMed: 12784371]

16. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003; 125:1731–1737. [PubMed: 12580598]

17. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455–461. [PubMed: 19499576]

18. Hwang H, Vreven T, Janin J, Weng Z. Protein-protein docking benchmark version 4. 0. Proteins. 2010; 78:3111–3114. [PubMed: 20806234]

19. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, Mitchell JC, Demerdash ON, Takeda-Shitaka M, Terashi G, Moal IH, Li X, Bates PA, Zacharias M, Park H, Ko JS, Lee H, Seok C, Bourquard T, Bernauer J, Poupon A, Aze J, Soner S, Ovali SK, Ozbek P, Tal NB, Haliloglu T, Hwang H, Vreven T, Pierce BG, Weng Z, Perez-Cano L, Pons C, Fernandez-Recio J, Jiang F, Yang F, Gong X, Cao L, Xu X, Liu B, Wang P, Li C, Wang C, Robert CH, Guharoy M, Liu S, Huang Y, Li L, Guo D, Chen Y, Xiao Y, London N, Itzhaki Z, Schueler-Furman O, Inbar Y, Potapov V, Cohen M, Schreiber G, Tsuchiya Y, Kanamori E, Standley DM, Nakamura H, Kinoshita K, Driggers CM, Hall RG, Morgan JL, Hsu VL, Zhan J, Yang Y, Zhou Y, Kastritis PL, Bonvin AM, Zhang W, Camacho CJ, Kilambi KP, Sircar A, Gray JJ, Ohue M, Uchikoga N, Matsuzaki Y, Ishida T, Akiyama Y, Khashan R, Bush S, Fouches D, Tropsha A, Esquivel-Rodriguez J, Kihara D, Stranges PB, Jacak R, Kuhlman B, Huang SY, Zou X, Wodak SJ, Janin J, Baker D. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol. 2011; 414:289–302. [PubMed: 22001016]

20. Moretti R, Fleishman SJ, Agius R, Torchala M, Bates PA, Kastritis PL, Rodrigues JPGLM, Trellet M, Bonvin AMJJ, Cui M, Rooman M, Gillis D, Dehouck Y, Moal I, Romero-Durana M, Perez-Cano L, Pallara C, Jimenez B, Fernandez-Recio J, Flores S, Pacella M, Kilambi KP, Gray JJ, Popov P, Grudinin S, Esquivel-Rodríguez J, Kihara D, Zhao N, Korkin D, Zhu X, Demerdash ONA, Mitchell JC, Kanamori E, Tsuchiya Y, Nakamura H, Lee H, Park H, Seok C, Sarmiento J, Liang S, Teraguchi S, Standley DM, Shimoyama H, Terashi G, Takeda-Shitaka M, Iwadate M, Umeyama H, Beglov D, Hall DR, Kozakov D, Vajda S, Pierce BG, Hwang H, Vreven T, Weng Z, Huang Y, Li H, Yang X, Ji X, Liu S, Xiao Y, Zacharias M, Qin S, Zhou H-X, Huang S-Y, Zou X, Velankar S, Janin J, Wodak SJ, Baker D. Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. Proteins. 2013 in press.

21. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol. 2000; 10:153–159. [PubMed: 10753808]

22. Vijayakumar M, Zhou HX. Salt bridges stabilize the folded structure of barnase. J Phys Chem B. 2001; 105:7334–7340.

23. Dong F, Zhou HX. Electrostatic contributions to T4 lysozyme stability: solvent-exposed charges versus semi-buried salt bridges. Biophys J. 2002; 83:1341–1347. [PubMed: 12202359]

24. Dong F, Vijayakumar M, Zhou HX. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. Biophys J. 2003; 85:49–60. [PubMed: 12829463]

25. Dong F, Zhou HX. Electrostatic contribution to the binding stability of protein-protein complexes. Proteins. 2006; 65:87–102. [PubMed: 16856180]

26. Qin S, Zhou HX. Do electrostatic interactions destabilize protein-nucleic acid binding? Biopolymers. 2007; 86:112–118. [PubMed: 17326079]

27. Pang X, Zhou HX. Poisson-Boltzmann calculations: van der Waals or molecular surface? Commun Comput Phys. 2013; 13:1–12. [PubMed: 23293674]

28. Schreiber G, Haran G, Zhou HX. Fundamental aspects of protein-protein association kinetics. Chem Rev. 2009; 109:839–860. [PubMed: 19196002]

29. Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S, Montelione GT, Spiegel CP, Liu DR, Baker D. A de novo protein binding pair by computational design and directed evolution. Mol Cell. 2011; 42:250–260. [PubMed: 21458342]

30. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z. Protein-protein docking benchmark version 3. 0. Proteins. 2008; 73:705–709. [PubMed: 18491384]

31. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol. 2012; 30:543–548. [PubMed: 22634563]

32. Wojdyla JA, Fleishman SJ, Baker D, Kleanthous C. Structure of the ultra-high-affinity colicin E2 DNase--Im2 complex. J Mol Biol. 2012; 417:79–94. [PubMed: 22306467]

33. Li W, Keeble AH, Giffard C, James R, Moore GR, Kleanthous C. Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. J Mol Biol. 2004; 337:743–759. [PubMed: 15019791]

34. Bailey LJ, McCoy JG, Phillips GN Jr, Fox BG. Structural consequences of effector protein complex formation in a diiron hydroxylase. Proc Natl Acad Sci U S A. 2008; 105:19194–19198. [PubMed: 19033467]

35. Leysen S, Vanderkelen L, Weeks SD, Michiels CW, Strelkov SV. Structural basis of bacterial defense against g-type lysozyme-based innate immunity. Cell Mol Life Sci. 2012; 70:1113–1122. [PubMed: 23086131]

36. Leysen S, Vanderkelen L, Van Asten K, Vanheuverzwijn S, Theuwis V, Michiels CW, Strelkov SV. Structural characterization of the PliG lysozyme inhibitor family. J Struct Biol. 2012; 180:235–242. [PubMed: 22634186]

37. Qin S, Zhou HX. meta-PPISP: a meta web server for protein-protein interaction site prediction. Bioinformatics. 2007; 23:3386–3387. [PubMed: 17895276]
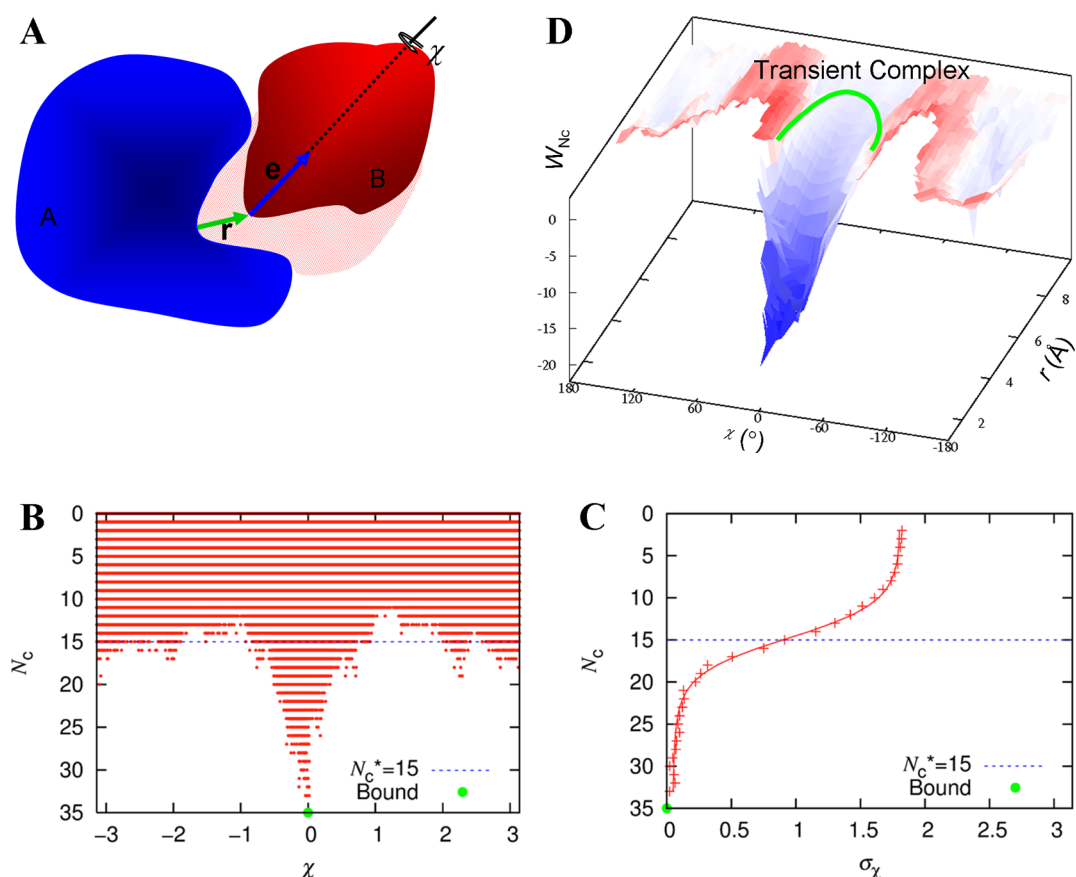
Figure 1.
Identification of the transient complex of two proteins, A and B. (A) The three translational coordinates, as defined by the relative displacement **r** between the centers of the binding sites on the two subunits; and the three rotational coordinates, as defined by the unit vector **e** attached to subunit B and the rotation angle χ around this unit vector. (B) Scatter plot of contact number (N_c) versus χ. Only clash-free poses are present in the scatter plot. (C) Specification of the transient complex, by fitting the standard deviation of χ values of clash-free poses at a given N_c to a function modeling two-state protein denaturation. The set of poses with the N_c value (designated as N_c^*) at the midpoint of the transition constitutes the transient complex. The N_c^* value is highlighted in both (B) and (C) by a blue dashed line. (D) The interaction free energy surface in the r-χ space. The location of the transient complex is sketched. Adapted from Alsallaq and Zhou[8] and Qin et al.[10]
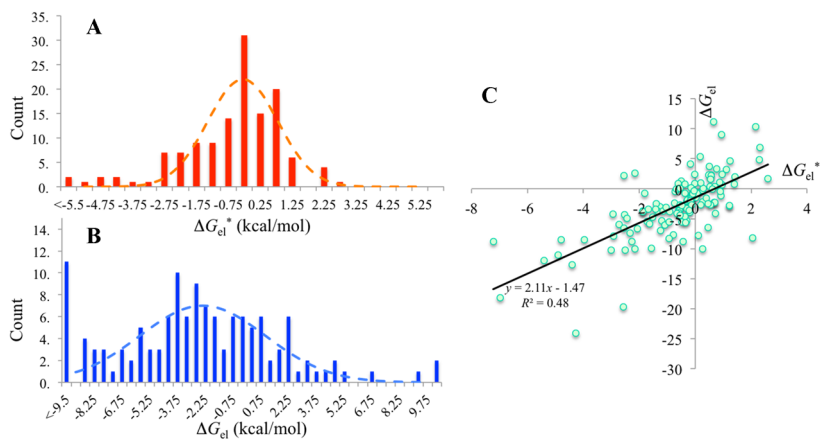
Figure 2.
Generally favorable electrostatic interaction free energies in native complexes and in transient complexes. (A) Histogram of ΔG_{el}^* values for 132 protein-protein complexes in benchmark 4.0 of Hwang et al.[18] (B) Histogram of ΔG_{el} values. (C) Correlation of ΔG_{el} and ΔG_{el}^*. Results were calculated by solving the nonlinear Poisson-Boltzmann equation at a temperature of 298 K (with $\varepsilon = 78.5$) and an ionic strength of 0.15 M.
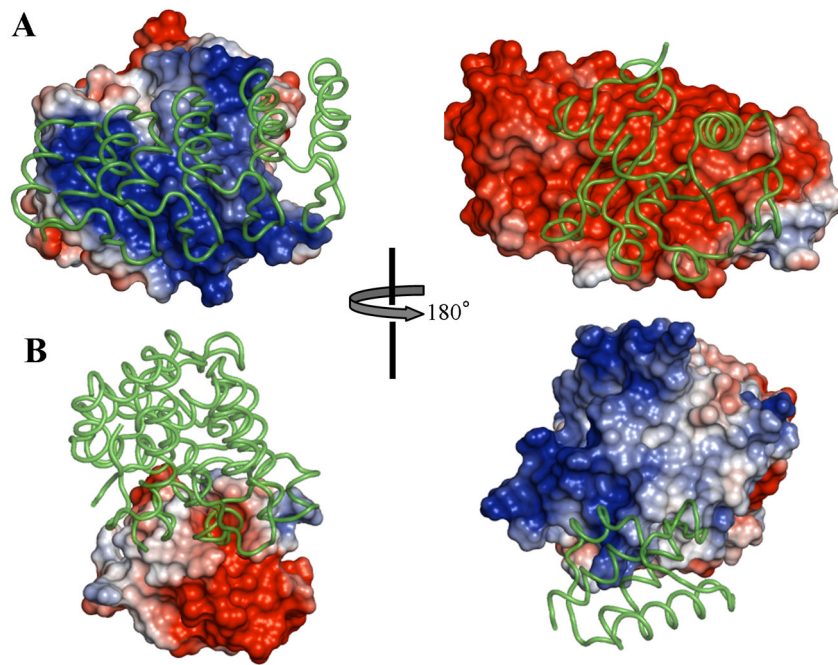
Figure 3.
T43 and T44 models. (A) Crystal structure 3Q9N, corresponding to Model 10 of T43. (B) Rosetta-generated structure for Model 2 of T44. Each model is presented in two views, rotated 180° from each other. In each view, one subunit is represented by the electrostatic surface and the other as ribbon; the representations are then swapped in the other view.
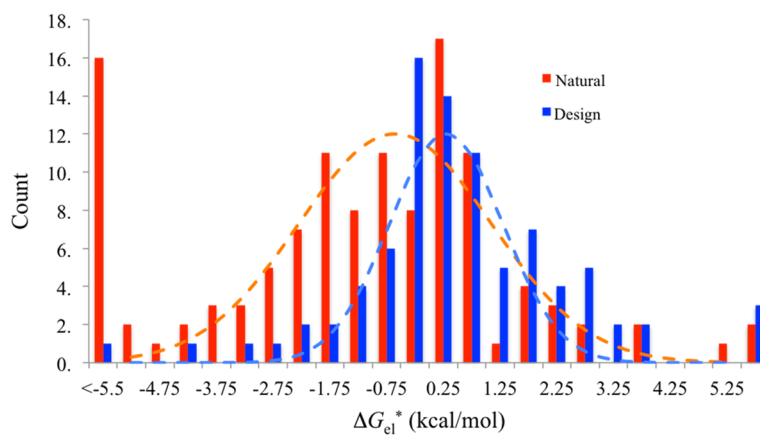
Figure 4.
Distributions of G_{el}^* for 120 natural complexes and 87 designs.

Table I

Ranking of models for T43 and T44 according to $G_{el}{}^*$ (in kcal/mol).

T43		T44	
Model	$G_{el}{}^*$	**Model**	$G_{el}{}^*$
10[a]	−7.02	10	−0.88
9	−4.95	19	−0.73
11	−2.31	17	−0.57
8	−1.23	**2**[b]	−0.56
2	−0.59	16	−0.34
6	−0.48	3	−0.29
20	−0.21	1	−0.19
14	0.14	8	−0.05
7	0.17	15	0.05
3	0.22	11	0.14
16	0.30	20	0.41
15	0.34	14	0.43
19	0.41	5	0.52
12	0.56	6	0.63
21	0.63	18	0.64
17	0.75	4	0.69
18	0.84	9	0.74
4	1.02	21	0.93
5	1.11	12	0.95
13	1.32	7	1.01

[a] This top-ranked model corresponds to crystal structure 3Q9N.

[b] This fourth-ranked model was shown to have measurable binding affinity.

Table II

Retrospective analysis on the contributions of the two terms of our energy function for T55 and T56.

Factor	T55		T56	
	Kendall	**p-value**	**Kendall**	**p-value**
G_{el}	0.0386	0.06	0.0737	9.6E-4
f_{cf}	0.102	1.3E-6	0.0210	0.36
Both	0.145[a]	1.8E-12	0.131[b]	4.21E-9

[a]The corresponding value from the official evaluation is 0.165.

[b]The corresponding value from the official evaluation is 0.147.