Scoring a diverse set of high-quality docked conformations: A metascore based on electrostatic and desolvation interactions

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · JUNE 2006

Impact Factor: 2.63 · DOI: 10.1002/prot.20932 · Source: PubMed

3 AUTHORS, INCLUDING:

Christoph Champ
University of Washington Seattle
5 PUBLICATIONS **110** CITATIONS

SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 63:868–877 (2006)

Scoring a Diverse Set of High-Quality Docked Conformations: A Metascore Based on Electrostatic and Desolvation Interactions

Carlos J. Camacho,[1]* Hui Ma,[2] and P. Christoph Champ[1]
[1]*Department of Computational Biology, University of Pittsburgh, Pittsburgh, Pennsylvania*
[2]*Molecular Biophysics Program, University of Pittsburgh, Pittsburgh, Pennsylvania*

ABSTRACT **Predicting protein–protein interactions involves sampling and scoring docked conformations. Barring some large structural rearrangement, rapidly sampling the space of docked conformations is now a real possibility, and the limiting step for the successful prediction of protein interactions is the scoring function used to reduce the space of conformations from billions to a few, and eventually one high affinity complex. An atomic level free-energy scoring function that estimates in units of kcal/mol both electrostatic and desolvation interactions (plus van der Waals if appropriate) of protein–protein docked conformations is used to rerank the blind predictions (860 in total) submitted for six targets to the community-wide Critical Assessment of PRediction of Interactions (CAPRI; http://capri.ebi.ac.uk). We found that native-like models often have varying intermolecular contacts and atom clashes, making unlikely that one can construct a universal function that would rank all these models as native-like. Nevertheless, our scoring function is able to consistently identify the native-like complexes as those with the lowest free energy for the individual models of 16 (out of 17) human predictors for five of the targets, while at the same time the modelers failed to do so in more than half of the cases. The scoring of high-quality models developed by a wide variety of methods and force fields confirms that electrostatic and desolvation forces are the dominant interactions determining the bound structure. The CAPRI experiment has shown that modelers can predict valuable models of protein–protein complexes, and improvements in scoring functions should soon solve the docking problem for complexes whose backbones do not change much upon binding. A scoring server and programs are available at http://structure.pitt.edu. Proteins 2006; 63:868–877.** © 2006 Wiley-Liss, Inc.

Key words: docking; protein interactions; scoring; complex structure; binding mechanism; recognition; desolvation; free energy

INTRODUCTION

Barring some large structural rearrangement, it is now possible to rapidly sample the space of protein docked conformations.[1] In fact, the limiting step for the successful prediction of protein interactions is the scoring function used to select the native-like models. Ideally, one would like to have a scoring function that would take all models for a given target and appropriately differentiate between native-like models and nonnative ones. However, because docking methods, force fields, and degrees of refinement vary widely,[2] the goal of having a single scoring function for every model regardless of its source might not be easily accomplished.

The challenge of scoring dates back to Wodak and Janin seminal paper on the computer analysis of protein–protein interactions,[3] where the scoring function used to dock two proteins was the surface buried upon forming the complex. This function was motivated by earlier studies of Chothia[4] and Janin,[5] who pointed out the importance of buried hydrophobic surface area in proteins. Ever since, surface complementarity—a smooth estimate of the van der Waals interaction—has been an essential component of most docking methods and scoring functions.[2] On its own, this scoring function has proven to be effective in significantly reducing the space of possible docked conformations from billions to several thousands.[1,6]

Early in the 1980s it was shown that electrostatic played a key role in the recognition of some receptor–ligand complexes[7,8] fueling the rapid development of computational estimates of the intermolecular electrostatic potential.[9,10] These estimates range from the direct Coulombic interaction energy[11] to sophisticated free-energy calculations that solved the Poisson-Boltzman equation.[12] These approaches, however, have the problem that lacked an ideal way for accounting for nonpolar interactions. The above notwithstanding, the successful electrostatic description of the binding interaction in some complexes[7,8] casted a shadow over the relative importance of nonpolar interactions.

Solvation effects on polar and nonpolar groups were first incorporated in binding free energy estimates using atomic

Grant sponsor: the National Science Foundation; Grant number: MCB-0444291.

*Correspondence to: Carlos Camacho, Department of Computational Biology, University of Pittsburgh, 3501 Fifth Avenue, Pittsburgh, PA 15213. E-mail: ccamacho@pitt.edu

Received 6 September 2005; Revised 21 November 2005; Accepted 13 December 2005

Published online 27 February 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20932

TABLE I. Table of Targets Analyzed in This Article

ID	Receptor	PDB code	Structure	Ligand	PDB	Structure	Complex PDB
8	Nidogen-G3	1NPE	Bound	Laminin	1KLO	Unbound	1NPE
12	Cohesin	1ANU	Unbound	Dockerin	1OHZ	Bound	1OHZ
13	Fab	1V74	Bound	SAG1	1KZQ	Unbound	1V74
14	Ser/Thr phosphotase-1	1FJM	Template	MYPT1	1S70	Bound	1S70
18	TAXI xylanase inh.	1T6G	Bound	Xylanase (A. niger)	1UKR	Unbound	1T6G
19	VRQ14 Fab	1TPX	Bound	Ovine prion	1DWY	Template	1TPX

solvation parameters.[13] These estimates, based on transfer free energies of molecules between organic solvents and water, gave a more complete accounting of the thermodynamic interactions involved in protein–protein binding. However, it was not until the late 1990s that solvation effects alone (including nonpolar interactions) were found to be dominant interactions capable of providing the binding specificity for some receptor–ligand complexes.[14]

Motivated by the above, we developed a free energy based scoring function $\Delta G_{\mathrm{RUGGED}}$ of docked conformations that accounted for the three aforementioned interactions[15]

$$\Delta G_{\mathrm{RUGGED}} = \Delta G_{\mathrm{SMOOTH}} + \Delta E_{\mathrm{vdW}}, \quad (1)$$

where

$$\Delta G_{\mathrm{SMOOTH}} = \Delta E_{\mathrm{Coul}} + \Delta G_{\mathrm{des}}. \quad (2)$$

The smooth (or soft) component of the scoring function $\Delta G^{16}_{\mathrm{SMOOTH}}$ accounts for the electrostatic and solvation terms, or chemical affinity, responsible for the broad free-energy minima found around the binding site.[14] ΔE_{Coul} corresponds to the direct Coulombic interaction plus screening modeled by a distance dependent dielectric, and ΔG_{des} is a term accounting for the desolvation of polar and nonpolar groups plus side chain entropy loss.[17] The van der Waals energy (vdW), ΔE_{vdW}, is the standard 6-12 Lennard-Jones potential, and one of the main sources of the ruggedness of the binding free-energy landscape (see, e.g., ref. 18). It is important to stress that $\Delta G_{\mathrm{RUGGED}}$ is by no means the full binding free energy, because it does not account for, say, internal energies, the loss of rotational and translational entropy, nor for solute–solvent vdW interactions.

Scoring functions of docked conformations are tested in the ongoing Critical Assessment of PRediction of Interaction[19] (CAPRI; http://capri.ebi.ac.uk) experiment in which researchers are given the three-dimensional coordinates of the unbound structures before the cocrystallized complexes are published. The researchers are then given a few weeks to dock the two structures together, and can use any information necessary, including biological information and literature searches. This community-wide experiment has had two evaluations meetings. The scoring function $\Delta G_{\mathrm{RUGGED}}$ has been used in the predictions of the Camacho group since the beginning of CAPRI in 2001; $\Delta G_{\mathrm{SMOOTH}}$ is used as a free-energy filter in the implementation of the *ClusPro* server[20] (http://structure.pitt.edu) , the first protein–protein docking server in its class. In rounds 1 and 2 of CAPRI, Camacho and Gatchell[21] produced some of the best model structures, appropriately distinguishing between near-native and false positive structures for three targets. In rounds 3 to 5, the automated server *ClusPro* (the only server participating in CAPRI) predicted good models for five targets,[22] while our manual predictions resulted in good predictions for six targets (missing the three targets that had a significant structural rearrangement upon binding). For all these predictions, $\Delta G_{\mathrm{RUGGED}}$ identified near-native models as those with the lowest free energy.[23]

In this article, we analyze the full set of models submitted for CAPRI-II (rounds 3 to 5) for the six targets that did not undergo a large structural rearrangement upon binding. We show that the scoring function $\Delta G_{\mathrm{SMOOTH}}$ is enough to discriminate near-native predictions from docked conformations far from the binding site for five of the targets, and for all but one of the manual predictions submitted to CAPRI. False positives were found for two groups, Eisenstein[24] and the automated server *ClusPro*.[22] Furthermore, Weng[25] and Camacho's[23] models are refined in part by vdW minimization. For these authors, $\Delta G_{\mathrm{RUGGED}}$ results in better discrimination of native-like complexes than $\Delta G_{\mathrm{SMOOTH}}$ alone. The most striking result is that the scoring function predicts the nearest-native models for each of the sampling and refinement techniques used by 16 groups in CAPRI. $\Delta G_{\mathrm{SMOOTH}}$ provides a fast and robust measure of the relative binding affinity of refined docked conformations.

METHODS AND METERIALS

The proteins studied in this article are listed in Table I; they correspond to targets 8, 12, 13, 14, 18, and 19 of the CAPRI experiment. For each of these targets, we downloaded the models submitted by all participants. Then, we used CHARMm[26] to add polar hydrogens in the complexes, and performed three rounds of 20 constrained backbone ABNR minimization steps (dielectric $4r$ and cutoff of nonbonded interactions of 15 Å). The vdW energy was obtained from the CHARMm output, and the electrostatic and desolvation free energy from the output of the program *FastContact*[16] on the minimized complex structures. The CHARMm minimization was necessary to obtain a meaningful vdW and hydrogens to compute the electrostatic energy. The structural changes due to the minimization were very small (less than 0.1 Å).

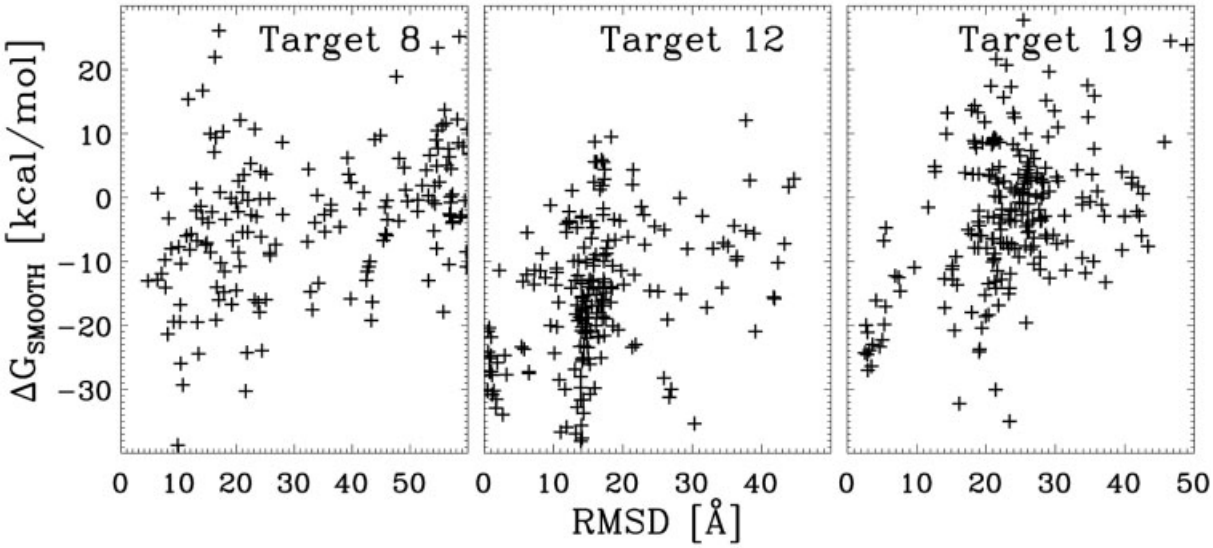

Fig. 1. Free energy score ΔG_{SMOOTH} as a function of RMSD for all the models submitted to three blind targets of CAPRI.

The program *FastContact*[16] is available to download and as a server at http://structure.pitt.edu/, the server computes the free energy score of two structures in PDB form. The interactions, discussed in Equations (1) and 2, are also easily reproducible by using the CHARMm19 partial charges and the atomic contact potential parameters available to download at http://structure.pitt.edu/software/fastcontact/.

RESULTS

In what follows, we show the analysis of the rescoring of the high-quality models submitted to CAPRI using ΔG_{SMOOTH} and ΔG_{RUGGED}. Figure 1 shows ΔG_{SMOOTH} for targets 8, 12, and 19. The chemical affinity of the receptor–ligand docked conformations is plotted as a function of the root-mean-square-deviation (RMSD) of the ligand with respect to the structure of the cocrystal after the receptors have been superimposed—an RMSD of 0 Å implies that the model is identical as the cocrystallized structure. Although native-like models have low free energy scores, Figure 1 does not distinguish target 12 native models; the second best score for target 8 is a false positive, and target 19 has several false positives with high RMSD. Because ΔE_{vdW} is heavily dependent on whether or not (also how and how much) modelers minimized the van der Waals energy for their predictions, ΔG_{RUGGED} is very noisy and provides no discrimination of native-like models.

A different approach is to score predictions for each modeler separately, because presumably the same method is used to generate all their predictions. The assumption here is that a consistent sampling and refinement procedure should better normalize the comparison between different models. In particular, in ref. 23 we showed evidence that ΔG_{RUGGED} is indeed able to differentiate the near-native models submitted by the Camacho group to CAPRI.

Figure 2 shows ΔG_{SMOOTH} as a function of RMSD for all the 11 groups that according to the assessors of CAPRI[2] had a near-native prediction for target 8. The plots also indicate the top-ranked submission of each group. We find that ΔG_{SMOOTH} discriminates a native-like prediction for eight groups. For four of these groups, ΔG_{SMOOTH} improves the ranking of the best model with respect to the ranking submitted by the own modelers. Weng group [Fig. 2(B)] refines their predictions using CHARMm,[25] accordingly we use ΔG_{RUGGED} to identify the native-like structure as the one with the lowest free energy score. ΔG_{RUGGED} for the Eisenstein group [Fig. 2(B)] has one false positive. A recurrent feature of Eisenstein predictions is the poor electrostatic, and therefore a ΔG_{SMOOTH} score of their native-like models (see Discussion). Finally, the server *ClusPro* rigid-body predictions are selected and ranked based on a measure of the broadness of the different free energy minima, without performing any refinement. Therefore, it is not surprising that an atomic level scoring function fails to properly rank the native-like models.

The ranking for target 12 is shown in Figure 3. It has been suggested that the cohesion/dockerin complex has two binding modes.[27,28] Most models were closer to a twofold symmetry structure of the crystal,[28,23] around 14 Å RMSD from the cocrystal (see Fig. 1; ref. 23 has a snapshot of this second binding mode). Figure 3(A) shows that for eight groups ΔG_{SMOOTH} ranks the very best model as the one with the lowest chemical affinity. It is important to stress here that five of these groups ranked the wrong model as their top prediction. In Figure 3(B), Camacho and Weng native models are correctly selected using ΔG_{RUGGED}. Valencia's group does not optimize the vdW energy; however, the large overlaps of some of their models led to a misleading ΔG_{SMOOTH} score. Ranking using ΔG_{RUGGED} eliminates the worst overlaps, leaving the best native-like model with the best score.

Figure 3(C) shows that four groups (Umeyama,[29] Wolfson,[30] Krippahl and Bates[31]) submitted almost exclusively models close to the twofold symmetry complex (around 14 Å RMSD), and the best chemical affinity corresponded to

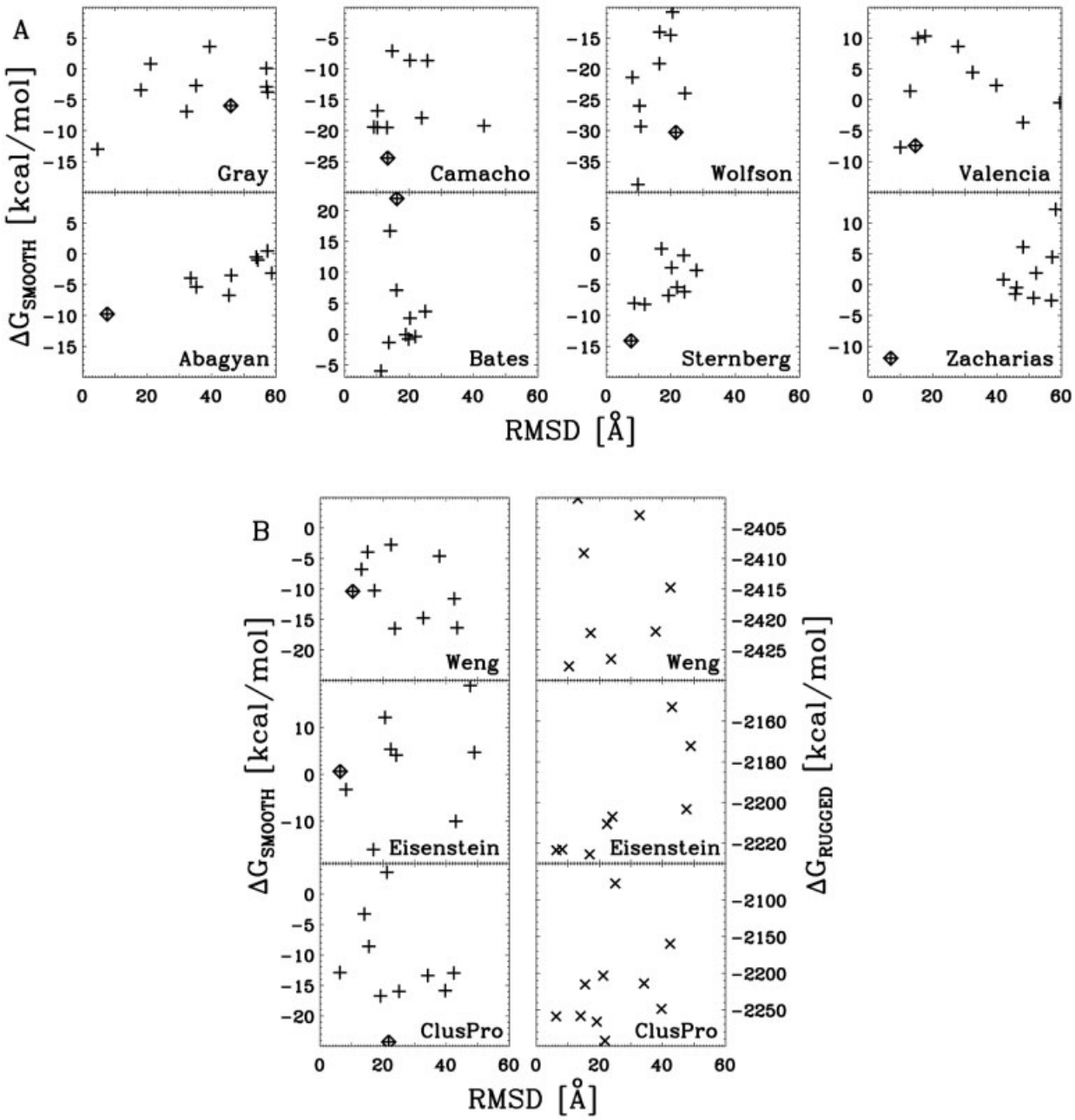

Fig. 2. Free energy score as a function of RMSD for the models of participants to CAPRI that had a native-like model among their predictions to Target 8. Diamond symbol indicates the model that was ranked No. 1 by the submitter. (**A**) Results for submitters that were correctly ranked using $\Delta G_{\mathrm{SMOOTH}}$; (**B**) Others.

these models. Nevertheless, some of these models were assessed as acceptable,[2] but too far from the bound state to have a low enough free energy score. It is very likely that these native-like models are near the transition state between these two binding modes. Therefore, the free-energy funnel is not a monotonically increasing function with RMSD. Finally, as for target 8, Eisenstein and *ClusPro* predictions failed to get properly ranked (see Discussion).

The rankings for target 13 are shown in Figure 4. Here, $\Delta G_{\mathrm{SMOOTH}}$ improves the ranking for all 10 groups. Camacho's models have one false positive with relatively high vdW; $\Delta G_{\mathrm{RUGGED}}$ eliminates this conformation, and scores the best native model with the lowest free energy.[23] We note that models with RMSD's larger than 60 Å, as for Baker[32] and ClusPro, correspond to false positives involving the membrane bound surface of the antigen SAG1. In reality, these regions are not accessible to the antibody, but it is interesting to note that they could generate low free energies. Poupon's predictions include one false positive (Model ranked fifth). This model has the Fab binding to the framework region of the antigen. This region is

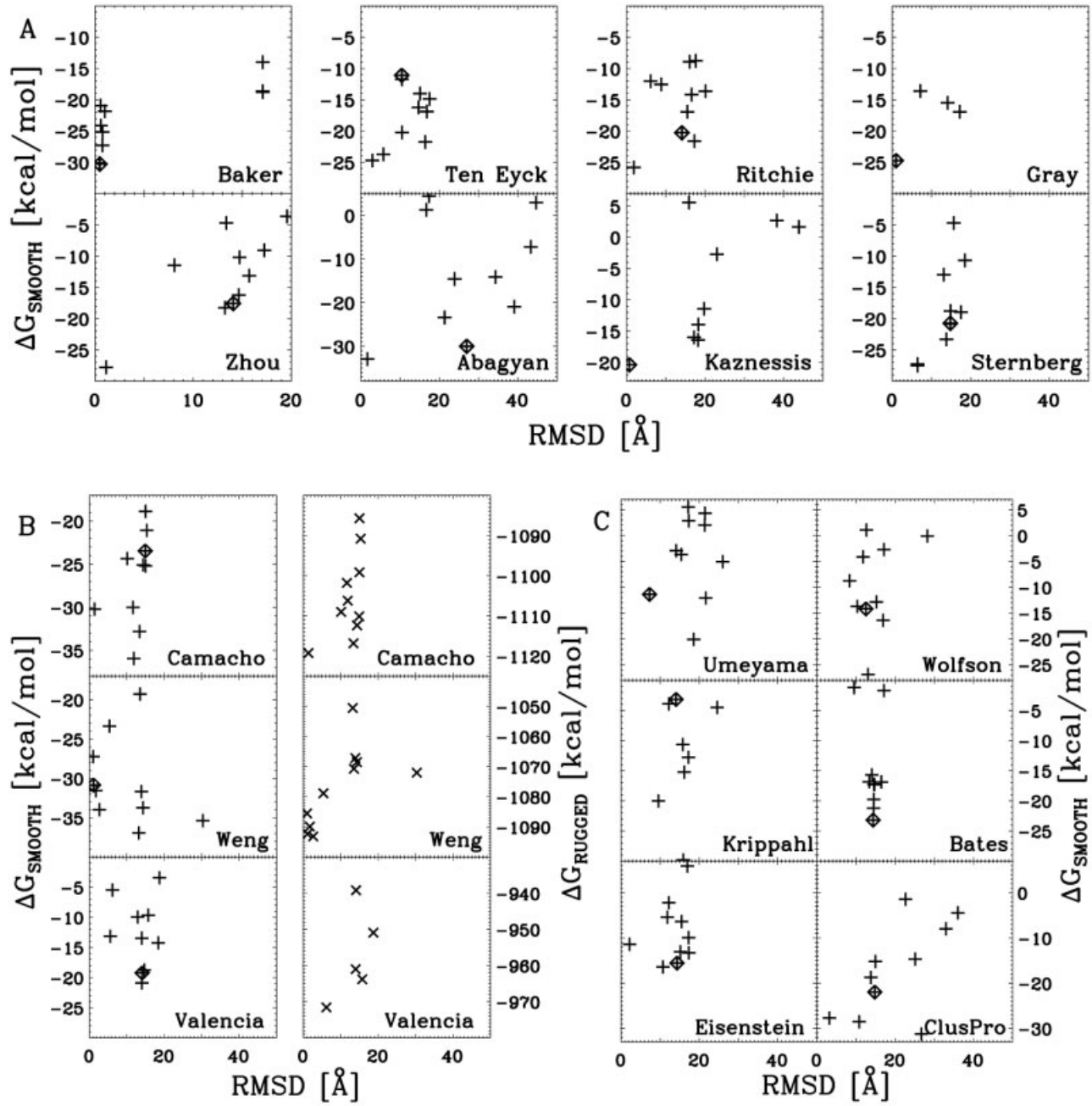

Fig. 3. Free energy score as a function of RMSD for the models of participants to CAPRI that had a native-like model among their predictions to Target 12. Diamond symbol indicates the model that was ranked No. 1 by the submitter. (**A**) Results for submitters that were correctly ranked using $\Delta G_{\mathrm{SMOOTH}}$; (**B**) Results for submitters that were correctly ranked using $\Delta G_{\mathrm{RUGGED}}$; (**C**) Others.

blocked by the N-terminal of SAG1 in the bound structure but not in the unbound.

For target 14 (Fig. 5), the chemical affinity $\Delta G_{\mathrm{SMOOTH}}$ predicts a native-like model for 12 different groups. The discrimination of near-native structures is quite striking for almost all groups. Notice that for, say, Baker[32] and Ritchie[33] in Figure 5(A), $\Delta G_{\mathrm{SMOOTH}}$ differentiates very small RMSD differences even though the relative scale for their models is 70 kcal/mol apart. It would be hard to conceive such a level of discernment with a nonatomic level scoring function. For Wolfson's models, the lowest $\Delta G_{\mathrm{SMOOTH}}$ score is found for their submission numbered 1, 7.3 Å RMSD from the crystal. This model has more than 200 clashes, and an unrealistic vdW energy of 622 kcal/mol, eliminating this complex or scoring using $\Delta G_{\mathrm{RUGGED}}$ results on model 3 (4.1 Å RMSD) as the one with lowest free-energy score. As already mentioned, Camacho's scoring improves using $\Delta G_{\mathrm{RUGGED}}$.[23] Zhou,[34] Weng, and Eisenstein models in Figure 5(C) failed to rerank properly; only for Zhou

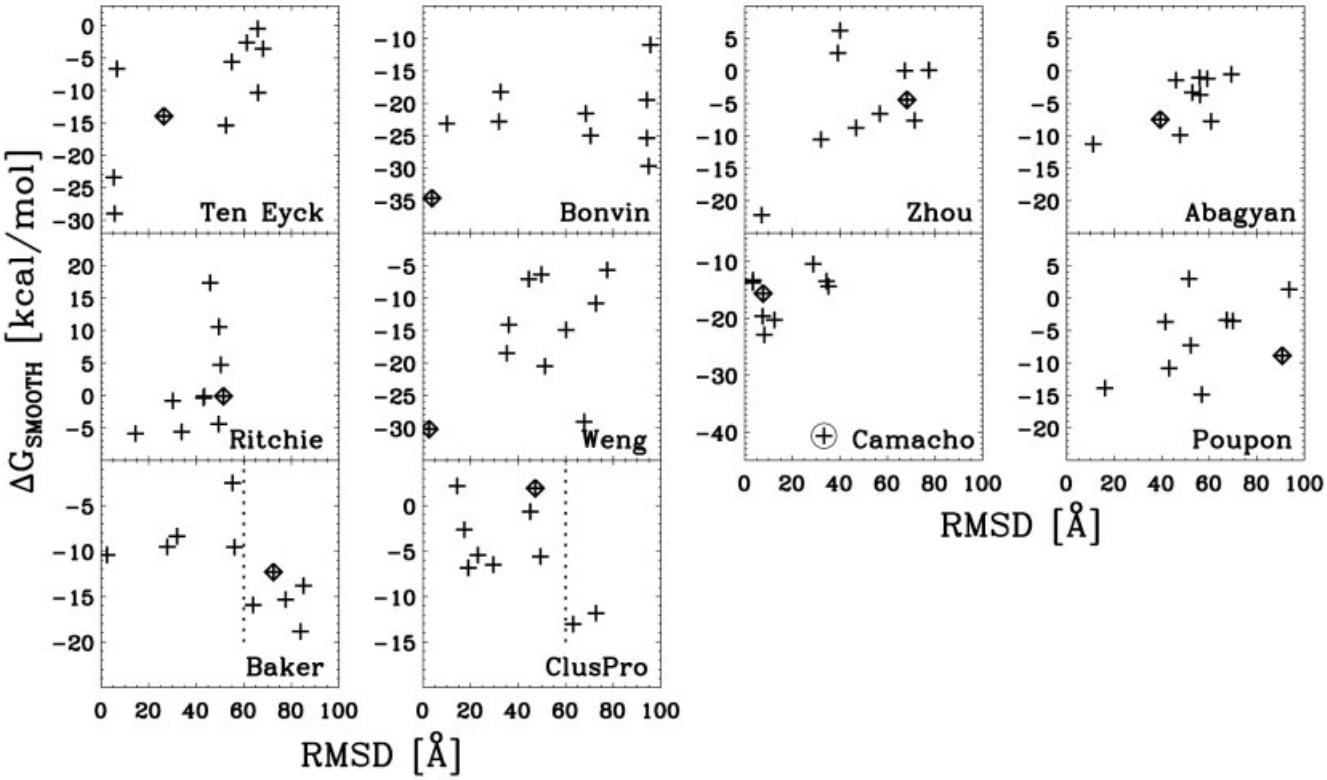

Fig. 4. Free energy score as a function of RMSD for the models of participants to CAPRI that had a native-like model among their predictions to Target 13. Diamond symbol indicates the model that was ranked No. 1 by the submitter. Enclosed in a circle we indicate a model with poor vdW that is properly ranked using $\Delta G_{\mathrm{RUGGED}}$. Models with RMSD larger than 60 Å (to the right of the dashed lines) interact with the membrane bound domain of SAG1.

submissions $\Delta G_{\mathrm{RUGGED}}$ eliminates all false positives (not shown).

Target 18 was the most challenging target among the set of receptor/ligand complexes analyzed in this article. Only seven groups had a meaningful prediction, and only Weng's manual ranking correctly identify their native-like model number 1.[25] As shown in Figure 6, Vakser's models are reranked correctly using $\Delta G_{\mathrm{SMOOTH}}$, and Bates has one false positive; Camacho and Wolfson predictions are correctly reranked using $\Delta G_{\mathrm{RUGGED}}$, while Weng's have one false positive. Bates,[31] Zhou,[34] and Abagyan[35] native-like structures are only reranked correctly after a full atom vdW minimization (i.e., without constraining the backbone). Overall the chemical affinity for this complex is relatively small (around −10 kcal/mol), while the vdW was not optimal (see Discussion).

Results for target 19 are plotted in Figure 7. With the exception of the server *ClusPro* and Sternberg's[36] (acceptable model is 15 Å away from bound), $\Delta G_{\mathrm{SMOOTH}}$ ranks well the models of eight groups that submitted near-native models. We observed two false positives: Zacharias[37] group built a complex by significantly rearranging the N-terminal domain of the prion protein away from the crystal, creating extra contacts that distorted the comparison between his different models; and, Bate's false positive has a large number of overlaps and unrealistic vdW energy of 688 kcal/mol (the average vdW is −1500 kcal/mol for this complex).

DISCUSSION

For models that do not have significant rearrangements upon binding, the scoring function $\Delta G_{\mathrm{SMOOTH}}$ in Equation (2) provides a relative estimate of the affinity of docked conformations for the models submitted to the Critical Assessment of Protein Interactions (CAPRI) experiment.[19] However, the discrimination of native-like structures is heavily dependent on the method used for refinement. Namely, the overall scale of the scoring function depends on the amount of overlap left on the predicted models, the extent of side chain optimization, and hydrogen bond pairing. This is quite apparent in Figure 1, where the results for target 12 seem to suggest that the correct binding site is 14 Å RMSD away from the cocrystal, while a more detailed analysis consistently identified the native-like models for each submitter [Fig. 3(A)]. We also note that the free-energy scale obtained for native-like complexes can vary by 20 kcal/mol or more (see, e.g., Ritchie and Bates scores in Figure 5 that differ by as much as 100 kcal/mol). The above notwithstanding, the score $\Delta G_{\mathrm{SMOOTH}}$ identifies native-like complexes as those with the lowest free energy for the models of 16 (out of 17) human predictors for five of the targets, while at the same time the modelers failed to do so in more than half of the cases.

Including van der Waals interactions in the scoring function ($\Delta G_{\mathrm{RUGGED}}$) is certainly helpful for those groups that optimize this component of the free energy, mainly

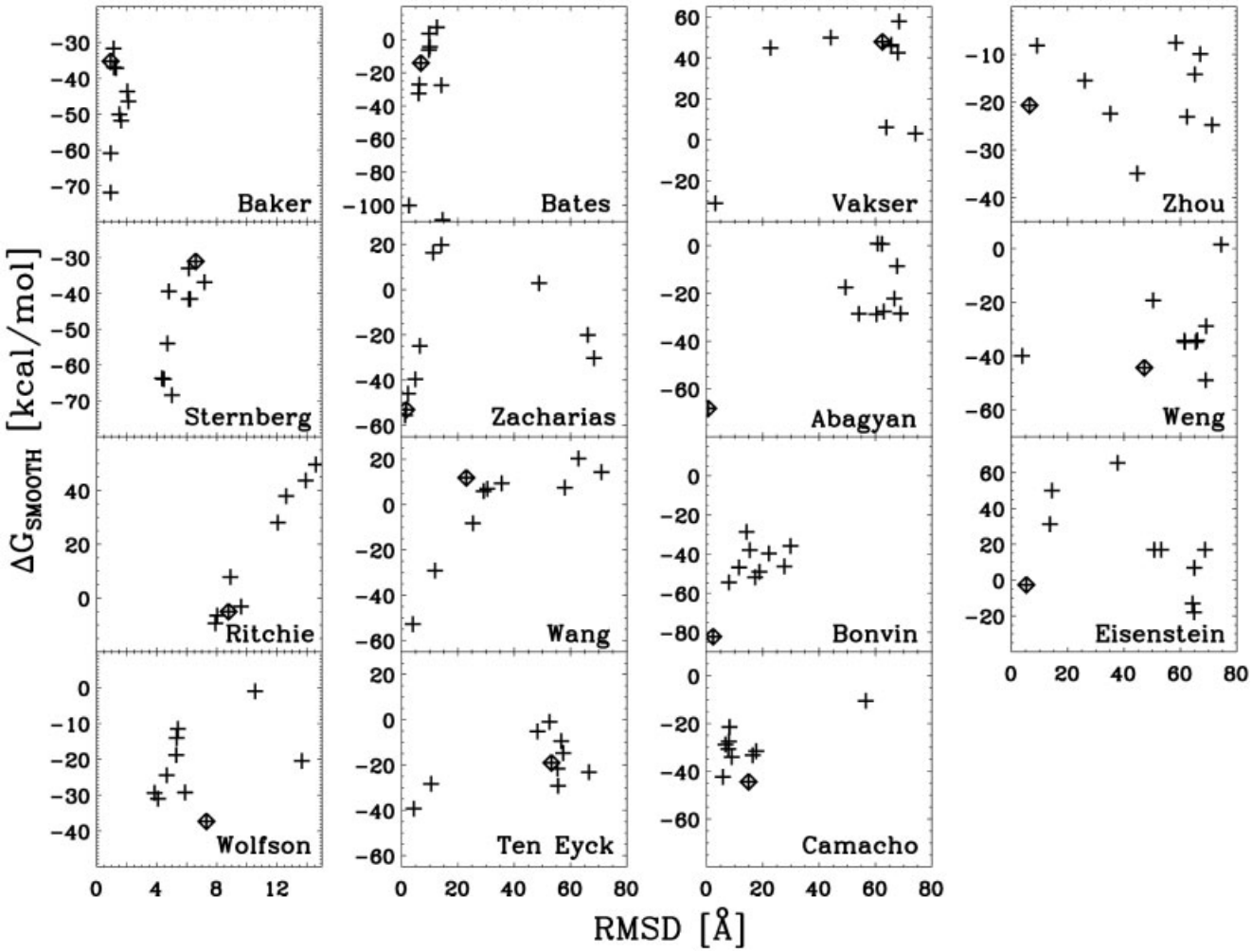

Fig. 5. Free energy score as a function of RMSD for the models of participants to CAPRI that had a native-like model among their predictions to Target 14. Diamond symbol indicates the model that was ranked No. 1 by the submitter.

Camacho's[23] and Weng's[25] group that used similar scoring functions. However, the success of $\Delta G_{\mathrm{SMOOTH}}$ scoring the rest of the models suggests that one might be able to refine docked conformations without an explicit vdW energy term. For instance, Abagyan,[35] Baker,[32] and Gray[38] used a truncated vdW term in their energy score, and $\Delta G_{\mathrm{SMOOTH}}$ was able to consistently score their models well.

It is not surprising that $\Delta G_{\mathrm{SMOOTH}}$ is an effective scoring function for docked conformations, because electrostatics and desolvation are the dominant interactions responsible for recognition and docking. For the most part, the electrostatic component of $\Delta G_{\mathrm{SMOOTH}}$ reranks these high-quality models better than the desolvation component. For target 13, it is the desolvation (not electrostatics) component of $\Delta G_{\mathrm{SMOOTH}}$ the one that appropriately identifies the native-like models, including those of Baker's and *ClusPro*'s that failed to be properly ranked by $\Delta G_{\mathrm{SMOOTH}}$. This evidence (not shown) confirms that a general scoring function needs to account for both terms of the free energy as suggested by the biophysics of molecular recognition.[14]

Our results yielded few false positives that are interesting to discuss in some detail because they shed light into the shortcoming of scoring functions. The electrostatic energy is often problematic, particularly in conjunction with high repulsive desolvation. For instance, in Figure 1 target 19, the three false positives (RMSD >10 Å) have a desolvation energy $\Delta G_{\mathrm{des}} > 15$ kcal/mol. It is worth mentioning that several of these low electrostatic/high desolvation complexes were found by the Bonvin group,[39] which only accounted for solvation effects at the very end of their docking algorithm. In this regime, our free-energy estimate probably fails because of the breakdown of the additivity assumption entailed by $\Delta G_{\mathrm{SMOOTH}}$. These type of complexes pose a serious limitation to the performance of our scoring function on a wide set of docked conformations that have not been properly screened.

We also note that clustering alone is able to identify near-native docked conformations within the top 10 models (server *ClusPro*),[20,22] but without refinement of the clusters the chemical affinity estimate $\Delta G_{\mathrm{SMOOTH}}$ is not

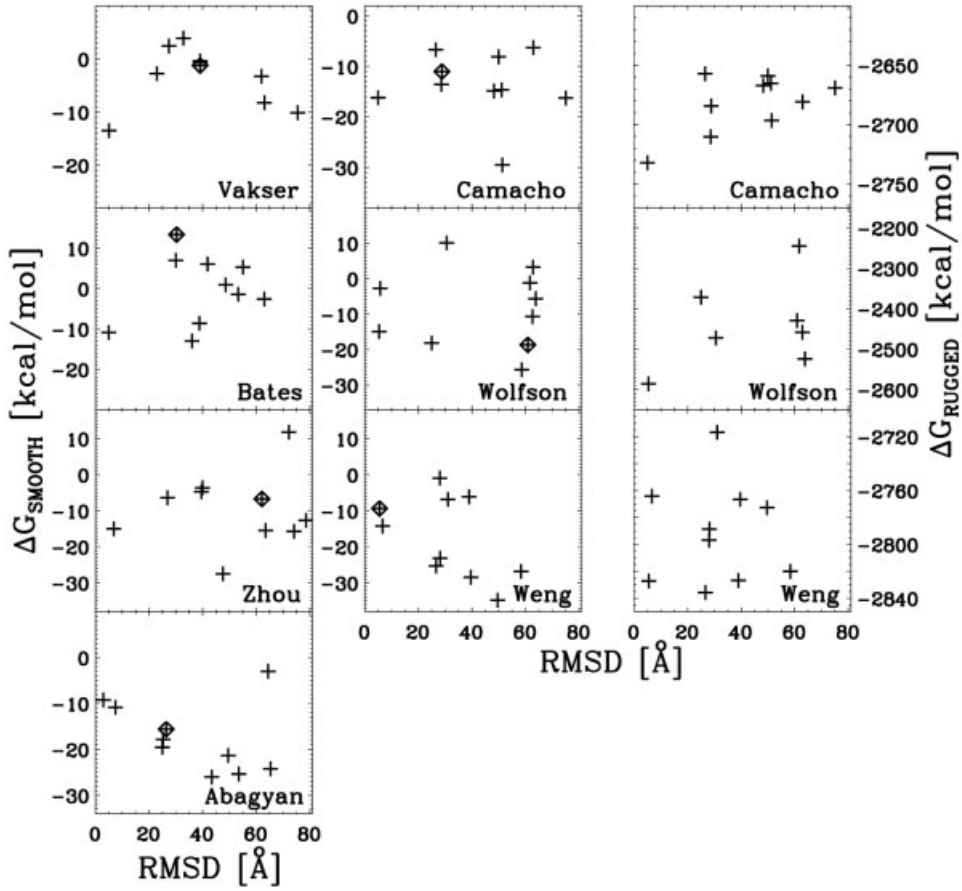

Fig. 6. Free energy score as a function of RMSD for the models of participants to CAPRI that had a native-like model among their predictions to Target 18. Diamond symbol indicates the model that was ranked No. 1 by the submitter. For three groups, we show both ΔG_{SMOOTH} and ΔG_{RUGGED}.

enough to provide an appropriate ranking among the cluster centers.

For target 13, the Baker group (and *ClusPro*) found models with very low scores that docked the Fab region to the membrane bound region of the SAG1 antigen. Experience tells us that these false positives are quite common. Indeed, often the membrane bound region of a protein or antibody has a nonspecific hydrophobic patch that yields good desolvation free energies with other proteins. In vivo, these regions are not visible to the proteins. Thus, it is not clear whether the false positives are a product of a faulty scoring function or of the fact that these contacts are not supposed to happen. Poupon's group has one false positive that involves the framework region of the antigen. Interestingly, this false positive again raises the issue of accessibility because the contact area is actually blocked by the N-terminal of SAG1 in the bound structure, but the unbound structure is missing these residues.

Structural rearrangement can also cause the appearance of false positives. A good example is Zacharias[37] model No. 2 submission for target 19, where they rearranged the N-terminal of the Prion to improve its chemical affinity (This model had the lowest ΔG_{SMOOTH} of all submissions; see Fig. 7). This type of rearrangements cannot be accounted for by a scoring function that does not include internal energy, which is high for this model. This case is a good example of the limitations of current scoring functions to relatively rigid backbones, including the traditional internal energy term on a scoring function increases the ruggedness of the landscape to a whole new level.

Finally, structural overlaps, as listed by the CAPRI assessors, varied significantly among modelers. In few occasions, it was more than 100, while the typical number of contacts was more like 10–20. These overlaps often yielded very high, even positive, vdW energies, for example, Valencia's[40] models for target 8, Wolfson's[30] for target 14, and Bates's[31] for target 19. At the same time, they resulted in good electrostatic and desolvation energies because an artificially higher number of attractive contacts were generated. This problem is likely to arise for any type of contact potential with an empirical cutoff distance. Based on a free-energy measure, models with large overlaps are very difficult to properly assess.

Models submitted for target 18 did not rerank properly, except for Vakser's[41] models. Several aspects of this target converged to make this a difficult target. One problem was a small structural rearrangement on the binding loop of the TAXI inhibitor given to the CAPRI participants. The latter despite the fact that the inhibitor was announced to be in its bound structure. In all likelihood, the modelers

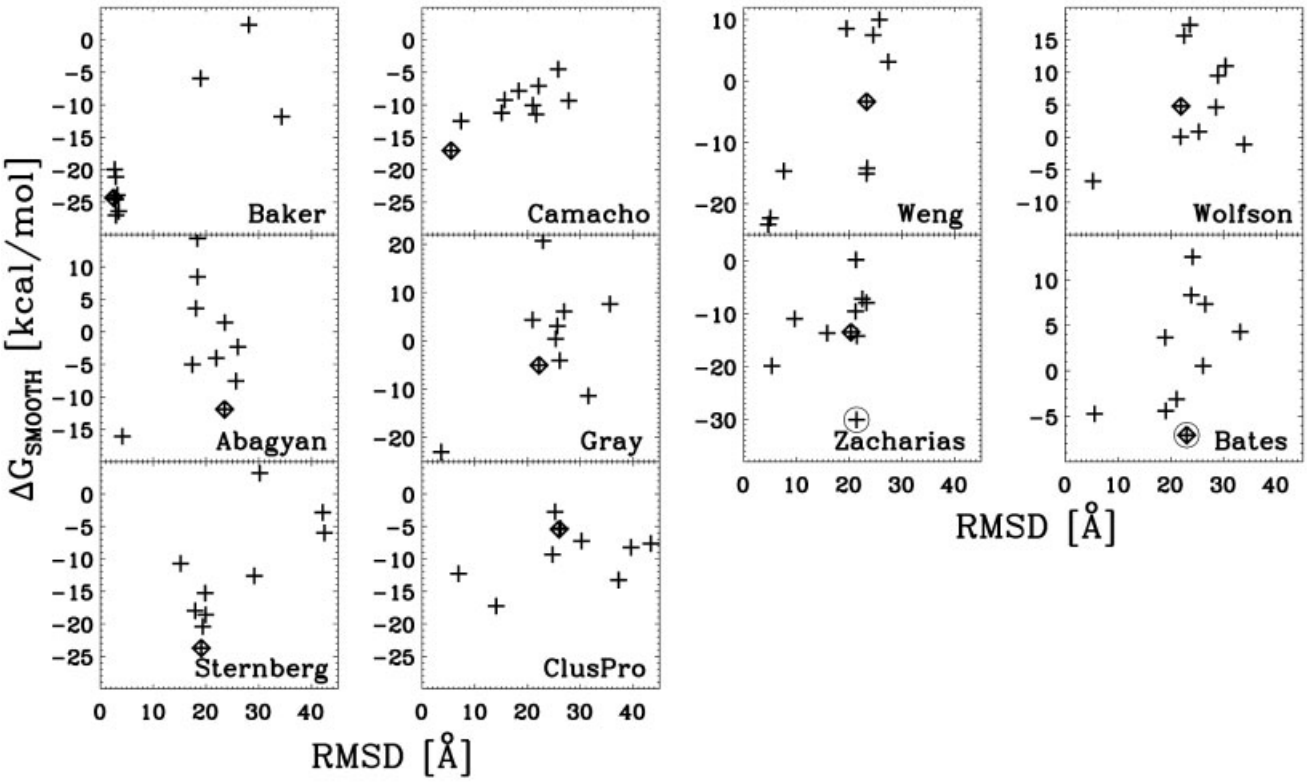

Fig. 7. Free energy score as a function of RMSD for the models of participants to CAPRI that had a native-like model among their predictions to Target 18. Diamond symbol indicates the model that was ranked No. 1 by the submitter. Enclosed in an open circle, we highlight a model by Zacharias that have significant backbone rearrangement, and a model by Bates that have large overlaps (see text).

expectation that TAXI was in its bound structure had a negative influence in the refinement of the models that kept poor contacts that could have been relaxed or eliminated. This is consistent with the fact that a full atom structural minimization improves the scoring of Bates, Zhou, and Abagyan models. Indeed, two groups leaders indicated that they missed this target because the constraints they imposed on the Taxi inhibitor. Another source of error was the protonation state of His^{374} of Taxi-I.[42] Most researchers optimized the docked conformation with a neutral Histidine, which biased their sampling and optimization procedures.

An intriguing observation of our analysis was the fact that $\Delta G_{\mathrm{SMOOTH}}$ was not able to rerank the models submitted by the Eisenstein group, not even their high quality models (targets 8 and 12). The problem was traced back to the poor electrostatic energy estimates found for these models, mostly due to missing hydrogen bonds. We note, however, that hydrogen bonds are missing from the original models. Indeed, these bonds are built and refined by CHARMm as part of the preprocessing of all models (see Methods). At the same time, the Eisenstein[24] method includes a manual refinement of these bonds at the very last stage of their docking technique. In all likelihood, there is some incompatibility in the way that the hydrogen bonds are satisfied in the models with respect to those placed by CHARMm that have stringent constraint in terms of bond angles and lengths.

FINAL REMARKS

The diversity of scoring functions, force fields, sampling, and refinement techniques used by CAPRI participants produced models of docked conformations with very different properties. We found that native-like models (low RMSD with respect to the cocrystal) often have varying degrees of atom clashes, and poor intermolecular contacts or hydrogen bonding properties, making the problem of constructing a normalized universal scoring function to properly rank these complexes very difficult.

Based on estimates of the electrostatic and desolvation free energy, the function of $\Delta G_{\mathrm{SMOOTH}}$ was proven successful in scoring the native-like models of individual groups as those with the lowest free energy. Dealing with backbone structural rearrangement remains an open problem.

ACKNOWLEDGMENTS

C.J.C. is grateful to Dr. Jeff Gray for encouraging this project. Hui Ma worked as a rotation student in the automation of the analysis of the CAPRI models. Christoph Champ built the server.

REFERENCES

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

2. Mendez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins 2005;60:150–169.
3. Wodak SJ, Janin J. Computer analysis of protein–protein interaction. J Mol Biol 1978;124:323–342.
4. Chothia C. The nature of the accessible and buried surfaces in proteins. J Mol Biol 1976;105:1–14.
5. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
6. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein–protein complexes. Proc Natl Acad Sci USA 1999;96:8477–8482.
7. Getzoff ED, Tainer JA, Weiner PK, Kollman PA, Richardson JS, Richardson DC. Electrostatic recognition between superoxide and copper, zinc superoxide dismutase. Nature 1983;306:287–290.
8. Mauk MR, Mauk AG, Weber PC, Matthew JB. Electrostatic analysis of the interaction of cytochrome *c* with native dimethyl ester heme substituted cytochrome *b*5. Biochemistry 1986;25:7085–7091.
9. Weiner PK, Langridge R, Blaney JM, Schaeffer R, Kollman PA. Electrostatic potential molecular surfaces. Proc Natl Acad Sci USA 1982;79:709–713.
10. Gilson MK, Honig B. Energetics of charge–charge interactions in proteins. Proteins 1988;3:32–52.
11. Pickersgill RW. A rapid method of calculating charge–charge interaction energies in proteins. Protein Eng 1988;2:247–248.
12. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149.
13. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
14. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein–protein association. Biophys J 1999;76,1166–1178.
15. Camacho CJ, Gatchell DW, Kimura SR, Vajda S. Scoring docked conformations generated by rigid-body protein–protein docking. Proteins 2000;40:525–537.
16. Camacho CJ, Zhang C. *FastContact*: rapid estimate of contact and binding free energies. Bioinformatics 2005;21:2534–2536.
17. Zhang C, Vasmatzis G, Cornette JL, DeLisi, C. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 1997;267:707–726.
18. Camacho CJ, Vajda S. Protein docking along smooth association pathways. Proc Natl Acad Sci USA 2001;98:10636–10641.
19. Janin J, Henrick K, Moult J, Ten Eyck L, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a critical assessment of predicted interactions. Proteins 2003;52:2–9.
20. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. *ClusPro*: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004;20:45–50.
21. Camacho CJ, Gatchell DW. Successful discrimination of protein interactions. Proteins 2003;52:92–97.
22. Comeau SR, Vajda S, Camacho CJ. Performance of the first protein docking server *ClusPro* in CAPRI rounds 3–5. Proteins 2005 60:239–244.
23. Camacho CJ (2005) Modeling side chains using molecular dynamics improve recognition of binding region in CAPRI targets. Proteins 60:245–251.
24. Ben-Zeev E, Kowalsman N, Ben-Shimon A, Segal D, Atarot T, Noivirt O, Shay T, Eisenstein M. Docking to single-domain and multiple-domain proteins: old and new challenges. Proteins 2005; 60:195–201.
25. Wiehe K, Pierce B, Mintseris J, Tong WW, Anderson R, Chen R, Weng Z. ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. Proteins 2005;60:207–213.
26. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMm: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
27. Schaeffer F, Matuschek M, Guglielmi G, Miras I, Alzari PM, Beguin P. Duplicated dockerin subdomains of Clostridium thermocellum endoglucanase CelD bind to a cohesin domain of the scaffolding protein CipA with distinct thermodynamic parameters and a negative cooperativity. Biochemistry 2002;39:5013–5021.
28. Carvalho AL, Dias FM, Prates JA, Nagy T, Gilbert HJ, Davies GJ, Ferreira LM, Romao MJ, Fontes CM. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. Proc Natl Acad Sci USA 2003;100:13809–13814.
29. Terashi G, Takeda-Shitaka M, Takaya D, Komatsu K, Umeyama H. Searching for protein–protein interaction sites and docking by the methods of molecular dynamics, grid scoring, and the pairwise interaction potential of amino acid residues. Proteins 2005;60:289–295.
30. Inbar Y, Schneidman-Duhovny D, Halperin I, Oron A, Nussinov R, Wolfson HJ. Approaching the CAPRI challenge with an efficient geometry-based docking. Proteins 2005;60:217–223.
31. Smith GR, Fitzjohn PW, Page CS, Bates PA. Incorporation of flexibility into rigid-body docking: applications in rounds 3–5 of CAPRI. Proteins 2005;60:263–268.
32. Schueler-Furman O, Wang C, Baker D. Progress in protein–protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. Proteins 2005;60:187–194.
33. Mustard D, Ritchie DW. Docking essential dynamics eigenstructures. Proteins 2005;60:269–274.
34. Zhang C, Liu S, Zhou Y. Docking prediction using biological information, ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function. Proteins 2005;60:314–318.
35. Fernández-Recio J, Abagyan R, Totrov M. Improving CAPRI predictions: optimized desolvation for rigid-body docking. Proteins 2005;60:308–313.
36. Carter P, Lesk VI, Islam SA, Sternberg MJE. Protein–protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. Proteins 2005;60:281–288.
37. Zacharias M. ATTRACT: protein–protein docking in CAPRI using a reduced protein model. Proteins 2005;60:252–256.
38. Daily MD, Masica D, Sivasubramanian A, Somarouthu S, Gray JJ. CAPRI rounds 3–5 reveal promising successes and future challenges for RosettaDock. Proteins 2005;60:181–186.
39. van Dijk ADJ, de Vries SJ, Dominguez C, Chen H, Zhou HX, Bonvin AMJJ. Data-driven docking: HADDOCK's adventures in CAPRI. Proteins 2005;60:232–238.
40. Tress M, de Juan D, Graña O, Gómez MJ, Gómez-Puertas P, González JM, López G, Valencia A. Scoring docking models with evolutionary information. Proteins 2005;60:275–280.
41. Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. Proteins 2005;60:296–301.
42. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns A. Structural basis for inhibition of *Aspergillus niger* xylanase by triticum aestivum xylanase inhibitor-I. J Biol Chem 2004;279:36022–36028.