ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/7763052

Classification of protein complexes based on docking difficulty

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · AUGUST 2005

Impact Factor: 2.63 · DOI: 10.1002/prot.20554 · Source: PubMed

CITATIONS
44

READS
6

1 AUTHOR:

Sandor Vajda
Boston University
189 PUBLICATIONS **7,112** CITATIONS

SEE PROFILE

Available from: Sandor Vajda
Retrieved on: 11 January 2016

PROTEINS: Structure, Function, and Bioinformatics 60:176–180 (2005)

Classification of Protein Complexes Based on Docking Difficulty

Sandor Vajda*
Department of Biomedical Engineering, Boston University, Boston, Massachusetts

ABSTRACT **Based on the results of several groups using different docking methods, the key properties that determine the expected success rate in protein–protein docking calculations are measures of conformational change, interface area, and hydrophobicity. A classification of protein complexes in terms of these measures provides a prediction of docking difficulty. This classification is used to study the targets of the CAPRI docking experiment. Results show that targets with a moderate expected difficulty were indeed predicted well by a number of groups, whereas the use of additional a priori information was necessary to obtain good results for some very difficult targets. The analysis indicates that CAPRI and other relatively large-scale docking studies represent very important steps toward understanding the capabilities and limitations of current protein–protein docking methods. Proteins 2005;60:176–180.** © 2005 Wiley-Liss, Inc.

Key words: CAPRI docking experiment; docking algorithms; hydrophobicity; interface area; conformational change

INTRODUCTION

One of the most important goals of CAPRI is to understand how well current methods can solve real-life (i.e., unbound–unbound) docking problems.[1] However, answering this question is far from trivial. The CAPRI targets are determined by the kindness of crystallographers rather than by any design, and thus can pose docking problems ranging from ones with moderate difficulty to ones that are essentially impossible to solve. The problem is that the CAPRI statistics alone do not necessarily show which target is difficult and which is easy. In fact, the participants are free to use any biochemical or structural information available in the literature, and in some cases this may help to orient the search toward correct solutions even for targets that would be very difficult to predict without such additional information.

We have recently studied the relationship between molecular properties of complexes and the outcome of docking calculations,[2] and introduced a classification of protein complexes based on docking difficulty. Here we apply this correlation to the CAPRI targets. For each target, the predicted difficulty is also compared to the actual "collective success" of the participants [i.e., the ratio of excellent (good, acceptable) solutions and the total number of submitted models for the given target]. As we describe, while the correlation between predicted and actual difficulty is generally good, much better than expected results for some difficult targets indicate the importance of the additional information that was available in the literature.

METHODS

Our correlations are based on the results of 3 research groups[3–6] that independently performed docking calculations for the 52 complexes in the benchmark set of Chen et al.[7] Following the analyses by Janin et al.,[8,9] we have found that the best predictors of success in docking are the conformational change upon binding, the change in the solvent accessible surface area, and the hydrophobicity of the interface. Based on this correlation, 5 classes of protein complexes, Types I through V, have been defined (Table I). The change, ΔASA, in the solvent accessible surface area, was calculated by ΔASA = ASA(complex) − ASA (receptor) − ASA (ligand), where ASA denotes the solvent accessible surface area of the protein indicated in the parenthesis. The hydrophobicity of the interface is given as the free energy, ΔG_{des}, of desolvation upon association, calculated using the atomic contact potential (ACP),[10] an atom-level extension of the Miyazawa and Jernigan potential.[11] A negative value of ΔG_{des} means that the removal of water from the interface is favorable, thereby indicating a largely hydrophobic interface.

According to Table I, docking is generally easy for Type I complexes that have an interface of "standard" size[8,9] that is rather hydrophobic. Most Type I complexes are enzymes and their protein inhibitors. Apart from occasional side-chains in wrong positions, major conformational changes rarely occur in the proteins forming these complexes. Docking is generally also successful for Type II complexes that have a large (but not very large) interface and moderate backbone conformational change upon association. Type III complexes have a "standard" but more polar interface ($\Delta G_{\mathrm{des}} > -4$ kcal/mol). This yields weaker binding and uncertain outcome in docking calculations, which might be affected even by small perturbations in the coordinates of the component proteins. Most antibody–

Grant sponsor: National Institutes of Health; Grant number: GM61867.

*Correspondence to Sandor Vajda, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215. E-mail: vajda@bu.edu

Received 18 January 2005; Accepted 26 January 2005

DOI: 10.1002/prot.20554

© 2005 WILEY-LISS, INC.

TABLE I. Classification of Protein Complexes on the Basis of Docking Difficulty

Type	Definition	Expected difficulty of docking
I	1400 Å^2 < ΔASA < 2000 Å^2 ΔG_{des} < −4 kcal/mol	Easy, unless key side-chains are in wrong conformations
II	2000 Å^2 < ΔASA < 3000 Å^2 C_α RMSD < 2 Å	Moderate difficulty
III	1400 Å^2 < ΔASA < 2000 Å^2 ΔG_{des} > −4 kcal/mol	Unpredictable; can be very difficult, even with know CDRs for the antibody
IV	ΔASA < 1400 Å^2	Very difficult; hits are found, but are generally lost in scoring and ranking
V	ΔASA > 2500 Å^2 C_α RMSD > 2 Å	Rigid-body methods always seem to fail for these complexes

ΔASA, change in the solvent accessible surface area; ΔG_{des}, free energy of desolvation upon association; C_α RMSD, α-carbon root-mean-square deviation between free and bound proteins.

antigen pairs form such Type III complexes, and some of them are far from easy to predict by docking, even when restricting the search to docked conformations in which the interface includes the antibody complementarity determining regions (CDRs). Type IV complexes have a small interface area, and since they are relatively weak, docking their component proteins generally yields poor results. Finally, Type V complexes have a very large interface area, and their component proteins are subject to substantial conformational change upon binding. Since Type II and Type V complexes overlap in terms of the interface area, conformational change is very important here. However, we have found that the conformational change is large if ΔASA > 3400 Å^2, and in such cases the complex is always Type V rather than Type II.

RESULTS AND DISCUSSION

Figure 1(a and b), respectively, shows the properties of the targets in Rounds 1–2 and 3–5 of CAPRI. The ellipses indicate the structures that are most likely solvable with the current docking methods. Table II is a summary of these results, including the level of expected difficulty. From the submissions to CAPRI, we have also calculated a measure to describe the success rate for each target. Based on the criteria defined by Méndez et al.,[12] the evaluators gave three stars (***), two stars (**), and one star (*), respectively, for high, medium, and acceptable quality predictions. Since the maximum number of stars is 3, we defined the percentage success rate for each target as the total number of stars, divided by 3 times the number of submitted models, and multiplied by 100.

According to Figure 1(a), the easiest target in Rounds 1–2 should be Target 6, which is of Type II. As expected, there were high- and medium-accuracy submissions from several groups (Table II). A higher level of difficulty is represented by Targets 1, 2, and 3. All these targets are of Type III, which makes the prediction of the success uncertain. However, Targets 1 and 2 are close to the boundary of the Type I region, which has easy problems (mostly enzyme–inhibitor complexes), whereas Target 3 is close to the boundary of the Type II region, which contains complexes with large interface area and small conformational change. Although these problems are more difficult than those in Target 6, in agreement with the expectations, several groups produced acceptable predictions, and even 2 medium-quality solutions were found for Target 3. Although the success rate was slightly higher for Target 1 than for the other 2, we note that predicting Target 1, it was possible to restrict the search using the known site of phosphorylation.[1] Targets 4 and 5 are of Type III, which includes many antigen–antibody complexes and does not provide enough information to predict the success rate. Target 7 is of Type IV, and thus is expected to be very difficult. Although several groups submitted accurate predictions, these must have been influenced by the analysis of a close homologue, available in the Protein Data Bank (PDB) for this complex.[1] According to our own experience,[3] rigid-body docking methods such as ZDOCK[4] produce some correct docked conformations for this complex, but due to the small and polar interface, the binding is so weak that these near-native conformations cannot be discriminated from false positives with apparently better shape complementarity.

Since Targets 15, 16, and 17 in Rounds 3–5 have been cancelled, we restrict consideration to the remaining targets. Target 11 included modeling one of the component proteins from a close homologue, but otherwise was the same as Target 12, and hence is not shown in Figure 1(b). The targets predicted to be most solvable in Rounds 3–5 are, in the order of increasing difficulty, 13, 12, (11), 18, and 19. Target 8 is of Type III, which makes the prediction of the outcome uncertain. Some of the interface residues were known from the literature,[13] but this target turned out to be of moderate difficulty even without this information, resulting in a few high- and several medium-quality predictions (Table II). Target 9 is definitely of Type V (ΔASA > 3400 Å^2). Indeed, there was a large conformational change, and the single acceptable solution came from the Wolfson group using a method accounting for the hinge motion.[14] Target 10 is also of Type V, and it is a large trimer with substantial conformational change in the subunits. Although this problem is very difficult, it can be solved if symmetry considerations are taken into account. Three acceptable and 1 medium-range prediction was

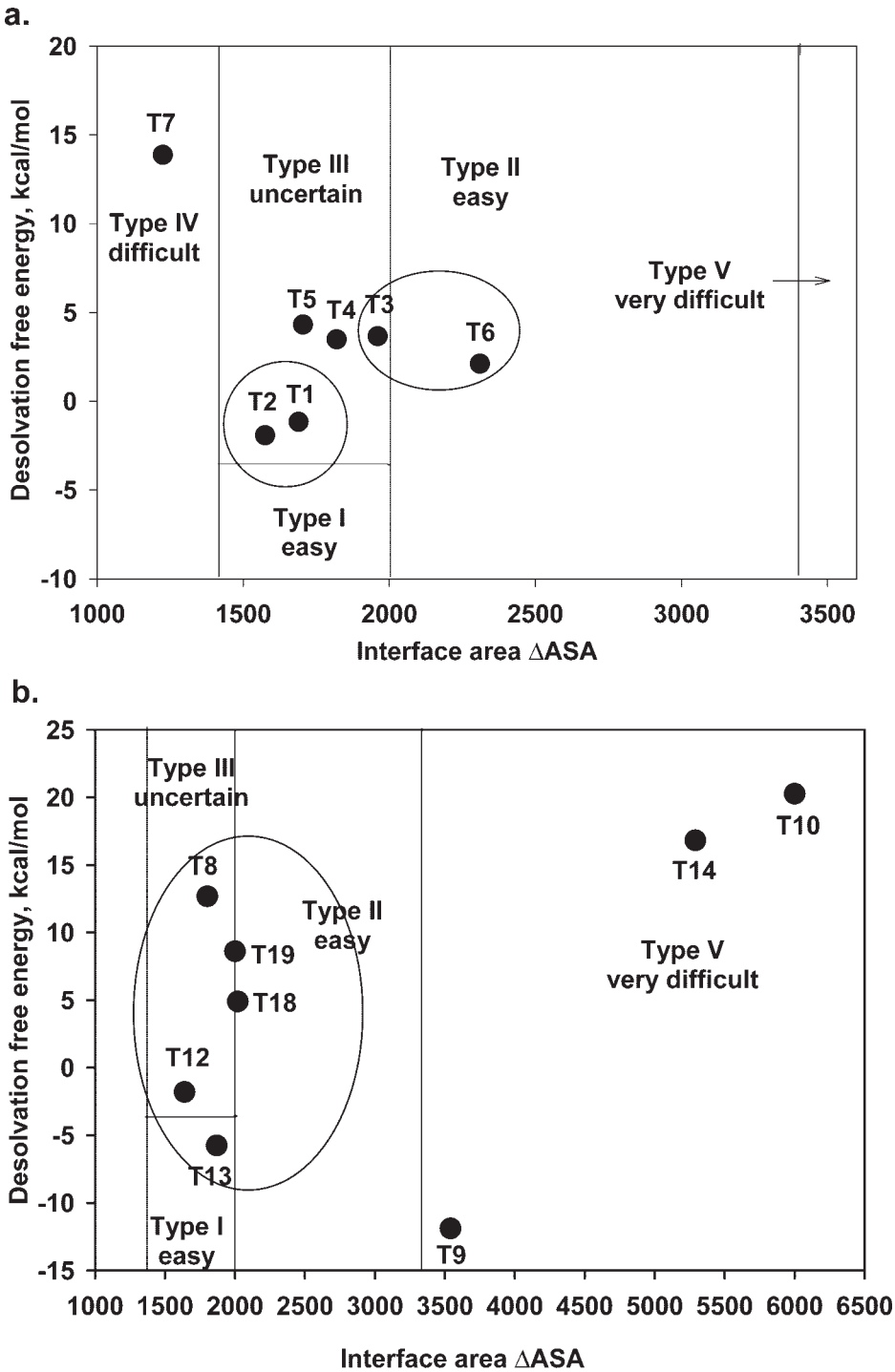

Fig. 1. Classification of CAPRI targets in terms of the change in the solvent accessible surface area (ΔASA) and the desolvation free energy (ΔG_{des}). The ellipses surround the relatively easy targets. (**a**) Rounds 1–2. (**b**) Rounds 3–5.

produced by the Bonvin group using appropriate constraints,[15] but all groups using general-purpose docking methods have failed.

Target 12 is of Type III, but it is close to the Type I region, and thus expected to be of moderate difficulty. In addition, one of the component proteins (dockerin) was given in its bound conformation. Indeed, the success rate was relatively high, and it was only slightly reduced when using the homology model in the solution (Target 11). Target 13 is of Type I [albeit, as shown in Fig. 1(b), is close to the boundary of the region that defines this type of complexes], and thus it is expected to be easy. We recall

TABLE II. Classification and Expected Difficulty of CAPRI Targets

Target	Complex	Type	Expected difficulty	Comment	Success rate, %
1	HPr kinase/HPr	III	Moderate to high	Close to Type I (which is easy)	3.19
2	Rotavirus VP6/FAB	III	Moderate to high	Close to Type I (which is easy)	2.29
3	Flu hemagglutinin/Fab 1ken	III	Moderate to high	Close to Type II (which is of moderate difficulty)	2.19
4	α-amylase/camelid Ab AM-D10	III	Uncertain	Unusual antibody–antigen complex	0.0
5	α-amylase/camelid Ab AM-07	III	Uncertain	Unusual antibody–antigen complex	0.0
6	α-amylase/camelid Ab AM-D9	II	Moderate to low		8.96
7	SpeA superantigen/TCRβ 110×	IV	Very high	Homology model was available	16.91
8	Laminin–nidogen	III	Uncertain	Turned out to be easier than expected	7.45
9	LicT homodimer	V	High	Close to Type II, but large conformational change	0.20
10	TBEV envelope protein trimer	V	Very high	Conformational change, but symmetry constraints help	0.97
11	Cohesin–dockerin complex	III	Moderate to low	Homology model; same as Target 12	9.12
12	Cohesin–dockerin complex	III	Moderate to low	Close to Type I (which is easy)	10.95
13	SAG1–antibody complex	I	Moderate to low	Based on classification, it was expected to be easier	7.18
14	PP1 and myosin phosphatase targeting subunit	V	Very high	Hydrogen bonds in the complex were described in the literature	16.06
18	*Aspergillus niger* xylanase–TAXI	III	Uncertain, moderate	Close to Type II (which is of moderate difficulty)	2.15
19	Ovine prion–Fab complex	III	Uncertain, moderate	Close to type II (which is of moderate difficulty)	4.52

that while Type I complexes are almost exclusively enzyme–inhibitor pairs, Target 13 is an antibody–antigen complex with an interface that is unusually hydrophobic. In fact, Target 13 was somewhat more difficult than expected for a Type I complex, although the antibody was given in the bound conformation, and the docking was restricted by the known CDR residues. Nevertheless, several high- and medium-quality models were submitted. Target 14, the protein phosphatase-1 (PP1) β complex with myosin phosphatase targeting subunit 1, is of Type V, and thus we would not expect any meaningful solutions. However, some critical residues (the RVXF motif) of the myosin phosphatase were known.[16] In addition, the binding mode was known, including the hydrogen bonds, for an analog of PP1 with a peptide that includes the RVXF motif.[17] The additional information was enough to produce a number of high-quality models. Targets 18 and 19 are both on the border between regions for Type II and Type III complexes, suggesting problems that are not very easy but still feasible to solve, and this was actually observed (Table II). It also may have helped that both targets are unbound–bound rather than unbound–unbound complexes.

CONCLUSIONS

On the basis of measures of conformational change, interface area, and hydrophobicity, we defined 5 types of protein–protein complexes to characterize the expected level of difficulty in docking. We emphasize that the correlation between molecular properties and docking difficulty was extracted from docking calculations that, apart from restricting considerations to CDR residues in antibody–antigen complexes, did not use any additional information on the complex.[2] Thus, the classification of complexes in Table I predicts the level of difficulty assuming no a priori information. This classification was used to study the CAPRI targets, resulting in a number of conclusions as follows:

1. The 19 CAPRI targets so far provided a good mixture of problems with different levels of difficulty. There is, however, a major difference between the composition of the protein docking benchmark set of Chen et al.[7] and the CAPRI targets. Almost 40% of the former consists of enzyme–inhibitor complexes that have been classified as Type I and represent relatively easy docking problems, primarily due to the strong hydrophobicity of the interface ($\Delta G_{des} < -4$ kcal/mol). The fraction of enzyme–inhibitor complexes has been reduced in the just released version 2.0 of the Benchmark Set,[18] but it is still around 30%. In contrast, the only Type I complex in CAPRI is Target 13, but it is an antigen–antibody rather than enzyme–inhibitor complex. This seems to

suggest that crystallographers moved beyond the enzyme–inhibitor interactions, heavily studied in the past, and that these "other" complexes generally exhibit a more polar interface.

2. In CAPRI the unbound–unbound targets appear to be intrinsically more difficult than the bound–unbound targets. This difference is not obvious from the docking results[4–6] for the proteins of the benchmark set.[7] However, most bound–unbound complexes in the benchmark are antibody–antigen pairs, which are frequently rather difficult.
3. As shown in Table II, the highest prediction success rates have been achieved for Targets 7 and 14 that, according to our correlations based on the analysis of the docking results for the benchmark set, are expected to be very difficult. In fact, substantial additional information was available in the literature for both targets; hence, the high success rates are somewhat misleading. For all other targets, there is good agreement between predicted and actual success rates. Two more targets, 9 and 10, are also very difficult, yet at least acceptable solutions were produced by 1 group in each case. However, the methods used were able to account for the specificities of the problem (hinge motion for Target 9, symmetry constraints for Target 10). Thus, even very difficult problems can be solved by specialized methods.
4. Since Targets 15, 16, and 17 were withdrawn, CAPRI so far had 16 targets. At least acceptable predictions were submitted for all but Targets 4 and 5. We note that Target 15 was only partly canceled, and the few submissions that were evaluated were rather successful. This is a very good overall success rate. In addition, although Targets 4 and 5 were difficult due to the low affinity of the complexes, most likely there would have been at least acceptable predictions if not the strong belief that antibodies bind via their CDRs, which was not the case here.
5. Targets 7 and 14 demonstrate that information in the literature can have a major impact on the CAPRI results. This type of problems can be fully eliminated if docking methods are implemented as servers that must be able to reproduce the results if the same problem is submitted again, thereby removing the possibility of any unreported intervention. Our server ClusPro was ready to participate only in Rounds 3–5,[19] but we also reran the targets of Rounds 1–2.[3] Using only the structures of component proteins and thermodynamic considerations, the quality of submissions from the server are in excellent agreement with the predicted level of difficulty but are not as good as the results from the best human predictors. Indeed, ClusPro produced high-quality predictions only for the easiest targets, 6 and 12, medium-quality for Targets 8 and 15, and acceptable quality for Targets 1, 13, 17, and 19.[18]

ACKNOWLEDGMENTS

I thank Steve Comeau for calculating the molecular properties of the complexes.

REFERENCES

1. 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.
2. Vajda S, Camacho CJ. Protein–protein docking: is the glass half full or half empty? Trends Biotechnol 2004;22:110–116.
3. 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.
4. Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein-docking algorithm. Proteins 2003;52:80–87.
5. Li L, Cheng R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. Proteins 2003;53:693–707.
6. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
7. Chen R., Mintseris J, Janin J, Weng Z. A protein–protein docking benchmark. Proteins 2003;52:88–91.
8. LoConte L, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. J Mol Biol 1999;285:2177–2198.
9. Chakrabarti P, Janin J. Dissecting protein–protein recognition sites. Proteins 2002;47:334–343.
10. 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.
11. Miyazawa S, Jernigan RL. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
12. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.
13. Stetefeld J, Mayer U, Timpl R, Huber R. Crystal structure of three consecutive laminin-type epidermal growth factor-like (LE) modules of laminin gamma1 chain harboring the nidogen binding site. J Mol Biol 1996;257:644–657.
14. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, Barzilai A, Dror O, Haspel N, Nussinov R, Wolfson HJ. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. Proteins 2003;52:107–112.
15. 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.
16. Toth A, Kiss E, Herberg FW, Gergely P, Hartshorne DJ, Erdodi F. Study of the subunit interactions in myosin phosphatase by surface plasmon resonance. Eur J Biochem 2000;267:1687–1697.
17. Egloff MP, Johnson DF, Moorhead G, Cohen PT, Cohen P, Barford D. Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. EMBO J 1997;16:1876–1887.
18. Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J, Weng Z. Protein–Protein Docking Benchmark 2.0: an update. Proteins 2005;60:214–216.
19. Comeau SR, Vajda S, Camacho CJ. Performance of the first protein docking server *ClusPro* in CAPRI Rounds 3–5. Proteins 2005;60:239–244.