

# FastContact: a free energy scoring tool for protein–protein complex structures

P. Christoph Champ and Carlos J. Camacho\*

Department of Computational Biology, University of Pittsburgh, Pittsburgh, PA, USA

Received January 31, 2007; Revised March 23, 2007; Accepted April 17, 2007

## ABSTRACT

**'FastContact' is a server that estimates the direct electrostatic and desolvation interaction free energy between two proteins in units of kcal/mol. Users submit two proteins in PDB format, and the output is emailed back to the user in three files: one output file, and the two processed proteins. Besides the electrostatic and desolvation free energy, the server reports residue contact free energies that rapidly highlight the hotspots of the interaction and evaluates the van der Waals interaction using CHARMM. Response time is ~1 min. The server has been successfully tested and validated, scoring refined complex structures and blind sets of docking decoys, as well as proven useful predicting protein interactions. 'FastContact' offers unique capabilities from biophysical insights to scoring and identifying important contacts.**

## INTRODUCTION

The most intuitive decomposition of the binding free energy involves four terms (1–3): van der Waals (vdW) interactions, electrostatic, hydrophobicity and configurational entropy. The relative contribution of the changes between the bound and free states of these four terms is not the same. For stability (1), the main contributions appear to be electrostatic and desolvation interactions. For refined docked conformations, vdW interactions are expected to balance between the bound and unbound state, as they seemingly do in protein folding (1). This is good news, since it is not yet possible to readily estimate solute–solvent interactions. It should be noted, however, that solute–solute vdW has been shown to be an important consideration for complex refinement (4). Configurational entropy loss upon binding, including rotational and translational degrees of freedom, is always important, rough estimates based on crystal complexes varying between 5 and 15 kcal/mol (2,5–8). For the most part, this entropy depends on the flexibility of the unbound or free state with respect to the bound,

with smaller corrections depending on the docking geometry. Since there is no robust estimate of entropy for a given protein, empirical free energy estimates, like 'FastContact', are always subject to an entropic correction. Hence, the server is most useful for discrimination between protein–protein docked complexes, and, more generally, for identifying energetically important contacts at the interface.

'FastContact', originally published in (9,10) rapidly estimates the electrostatic and desolvation component of the free energy based on a classic distance dependent dielectric  $4r$  (11) and an empirical contact potential for the desolvation contribution (7) developed using a database of crystal (no complexes) structures from the PDB. Because of the pairwise nature of the empirical interactions, 'FastContact' is also able to report the contribution of individual residues and pairs of residues to the free energy. The latter should prove useful for site-directed mutagenesis studies since rankings of these interactions consistently identify the hot spots in the interface.

## BRIEF DESCRIPTION OF THE ALGORITHM

The code behind the server was written in Fortran 77 and the server itself was written in PHP. 'FastContact' performs a fast computational estimate of the binding free energy between two proteins based on atomic pairwise interactions:

- (i) Electrostatic energy: the standard intermolecular Coulomb electrostatic potential with a distant-dependent dielectric constant equal to  $4r$ , enforcing a minimum atom-to-atom distance separation equal to the sum of their corresponding vdW radii to avoid artificial overlaps.
- (ii) Desolvation free energy: knowledge-based contact potential that accounts for hydrophobic interactions, self-energy change upon desolvation of charged and polar atom groups and side-chain entropy loss.
- (iii) vdW energy: the standard 6–12 Lennard–Jones potential is evaluated using the program CHARMM (12) as part of the optimization of polar Hydrogens and overlaps.

\*To whom correspondence should be addressed. Tel: +4126487776; Email: ccamacho@pitt.edu

The first two values (i–ii) can be used to calculate the overall free energy of the protein–protein interactions, assuming solute and/or solvent vdW cancellation between the bound and free proteins, and a correction factor for the configurational entropy loss. The application uses the definition of the atomic composition of each amino acid consistent with CHARMM19 parameters.

## USING THE FASTCONTACT SERVER

### Required user-input information

Figure 1 shows a snapshot of the input page. The user uploads two Protein Data Bank (PDB) format files (13), one ‘receptor’ and one ‘ligand’, along with their email address. The web server currently makes no distinction between chains; it simply reads in each line in the PDB file starting with an ‘ATOM’ field. The maximum number of residues is limited to 1500. The email address is where the output/results will be sent (as a file attachment). Hydrogen bonds and missing atoms are built and optimized on the uploaded structures using the molecular software CHARMM.

### Optional parameters

*Range of desolvation interaction.* The default range is 6 Å, such that the potential smoothly goes to zero between 5 and 7 Å. This range is suggested for refined models, without overlaps and relatively snugly fit interfaces, e.g. (4,10). The user has the option of changing the range to 9 Å, approaching zero between 8 and 10 Å. This modality is suggested for encounter complexes, e.g. (14,15) for a rigid body docking validation.

*Minimization.* The default setting for Hydrogen bond optimization and removal of minimal overlaps prescribes a short  $3 \times 20$  ABNR minimization steps with fixed backbone using the program CHARMM and the PARAM19 residue topology file (RTF).

However, the user is free to change this setting to a full atom minimization. This setting will work for single chains only and no gaps.

*Patch end terminals with  $-NH_3^+$  and  $-COOH$ .* By default, the end terminal residues will be patched by CHARMM. In case the end terminals are missing from the structure, the user has the option of turning the patching feature off.

### Output format and explanation

The results from a ‘FastContact’ server run are returned to the user via email as a file attachment (with a normal response time of  $\sim 1$  min). The attached file is a gzipped archive (.tar.gz) containing three results files: (i) the main results file (‘output.txt’); and, the processed (including H-bonds) and renumbered (ii) receptor PDB file (‘protein1’) and (iii) the ligand PDB file (‘protein2’). All of the files are prefixed with the user name (email prefix) and timestamp of the server run for easy reference.

The main source of errors in the output file relates with the format of the input PDB files. For instance, columns usage must strictly follow the PDB standards, and ATOM keyword must describe only protein amino acids. The server cannot minimize the backbone of sequences with gaps, and missing heavy atoms are sometimes not able to be reconstructed by the server. If the server detects an error, it will report a message with possible problems and suggestions.

The main results file (‘output.txt’) returns two components of a free energy function, electrostatic energy and desolvation free energy, and evaluates the solute vdW energy using CHARMM. The latter is sometimes useful to compare between different models (16), but here it is given only as a reference since it is not used in the analysis of contacts. Often vdW energies larger than about  $-500$  kcal/mol suggest structural overlaps. Although ‘FastContact’ smoothes the potentials to tolerate some limited overlaps, these are, in general, detrimental to the quality of the computational estimates. Figure 2 shows the summary energy output and part of the contact analysis, for the barnase–barstar complex 1BRS. We should caution that, when submitting co-crystallized receptor and ligand structures, the automated minimization implemented in the server leads to an over optimization of the electrostatic contacts of  $\sim 10$ – $20\%$  (5). The reason is because the direct electrostatic term used in the server does not have an angular dependence for Hydrogen bonds. Hence, these interactions tend to ‘double-dip’. This effect is compensated when scoring unbound models that always have some built in frustration due to the less optimal backbone and side chain conformations.

## DISCUSSION

### Critical assessment of protein interactions (CAPRI)

The method implemented in ‘FastContact’ has been successfully applied in the CAPRI experiment both as a free energy filtering procedure of the ‘ClusPro’ server (14) that predicts protein complexes and in protein–protein

Figure 1. Snapshot of input page.

```

----- SUMMARY ENERGIES -----
Electrostatic (4r) Energy: -40.4498808 kcal/mol
Desolvation Free Energy: 11.2128888 kcal/mol
van der Waals (CHARMM18): -1046.66 kcal/mol
-----
If vdW > -600 kcal/mol, structures might overlap
Affinity = Elec + Desol + Config. entropy*
*For a given pair of proteins,
Config. entropy is between 6-to-16 kcal/mol
-----
Top 20 Min & Max ligand residues contributing
to the binding free energy
-10.228 39 ASP
-8.808 74 GLU
-2.897 36 ASP
-2.312 44 TRP
-2.147 78 GLU
-1.418 34 LEU
-1.322 48 GLU
...
-----
...
0.420 30 TYR
0.684 32 GLU
0.843 1 LYS
0.788 31 GLY
0.817 33 ASN
1.787 38 TRP
-----
...
-----
Top 20 Min & Max receptor-ligand residue
free energy contacts
-10.802 81 ARG 38 ASP
-8.440 67 ARG 74 GLU
-8.263 86 ARG 39 ASP
-2.860 67 ARG 36 ASP
-2.790 26 LYS 39 ASP
-1.383 80 LYS 35 ASP
-1.383 67 ARG 78 GLU
-1.289 26 LYS 78 GLU
...
-----
...
0.887 62 ASP 78 GLU
0.880 26 LYS 38 TRP
0.874 80 LYS 33 ASN
0.884 100 HIS 38 ASP
1.217 101 TYR 38 ASP
1.217 67 ARG 38 TRP
1.268 71 GLU 36 ASP
2.340 73 ASP 38 ASP
3.844 71 GLU 38 ASP
-----

```

**Figure 2.** Output page. It includes a summary of electrostatics, desolvation and vdW energies, followed by a list of the 20 most attractive and 20 most repulsive residues and contacts for the electrostatic, desolvation, and the sum of these two components that correlates with the binding free energy. For modeling, the repulsive information is sometimes useful as an indication of a wrong structural motif.

refinement (4) (using a 9 and 6 Å desolvation range, respectively). ‘FastContact’ has been instrumental in the success of our group in blind predictions (17,18). In rounds 1 and 2 of CAPRI, Camacho and Gatchell (19) produced some of the best model structures, appropriately distinguishing between near-native and false positive structures for three targets. In rounds 3–5, the automated server ‘ClusPro’ (the only server participating in CAPRI) predicted good models for 5 targets (15), while our manual predictions resulted in good predictions for 6 targets (20) (missing the 3 targets that had a significant structural rearrangement upon binding).

The robustness of our method was further supported by the analysis of the full set of models submitted for CAPRI (rounds 3–5) for the 6 targets that did not undergo a large structural rearrangement upon binding (18). For these targets, we showed that ‘FastContact’ was able to discriminate near-native predictions from docked conformations far from the binding site for 5 of the targets (10), and for all but one of the manual predictions submitted to CAPRI. For instance, Figure 3 shows the re-scoring of models submitted for targets 8 and 12 by 13 different groups around the world. In all cases, ‘FastContact’ correctly identified the near-native conformation, even when the modeler failed to do so.

By splitting the free energy between electrostatics and desolvation, ‘Fastcontact’ also provides immediate insights into the nature of the binding interactions. Namely, negative desolvation is associated with a hydrophobic pocket at the binding site, whereas positive desolvation characterizes mostly polar interfaces. This is important since sometimes electrostatic or desolvation alone could lead to better discrimination than the combination of the two (21,22). The latter is, of course, due to the intrinsic limitations of empirical free energies. In particular, reliable estimates for solvent and entropic interactions are not yet available.

## OTHER SERVERS

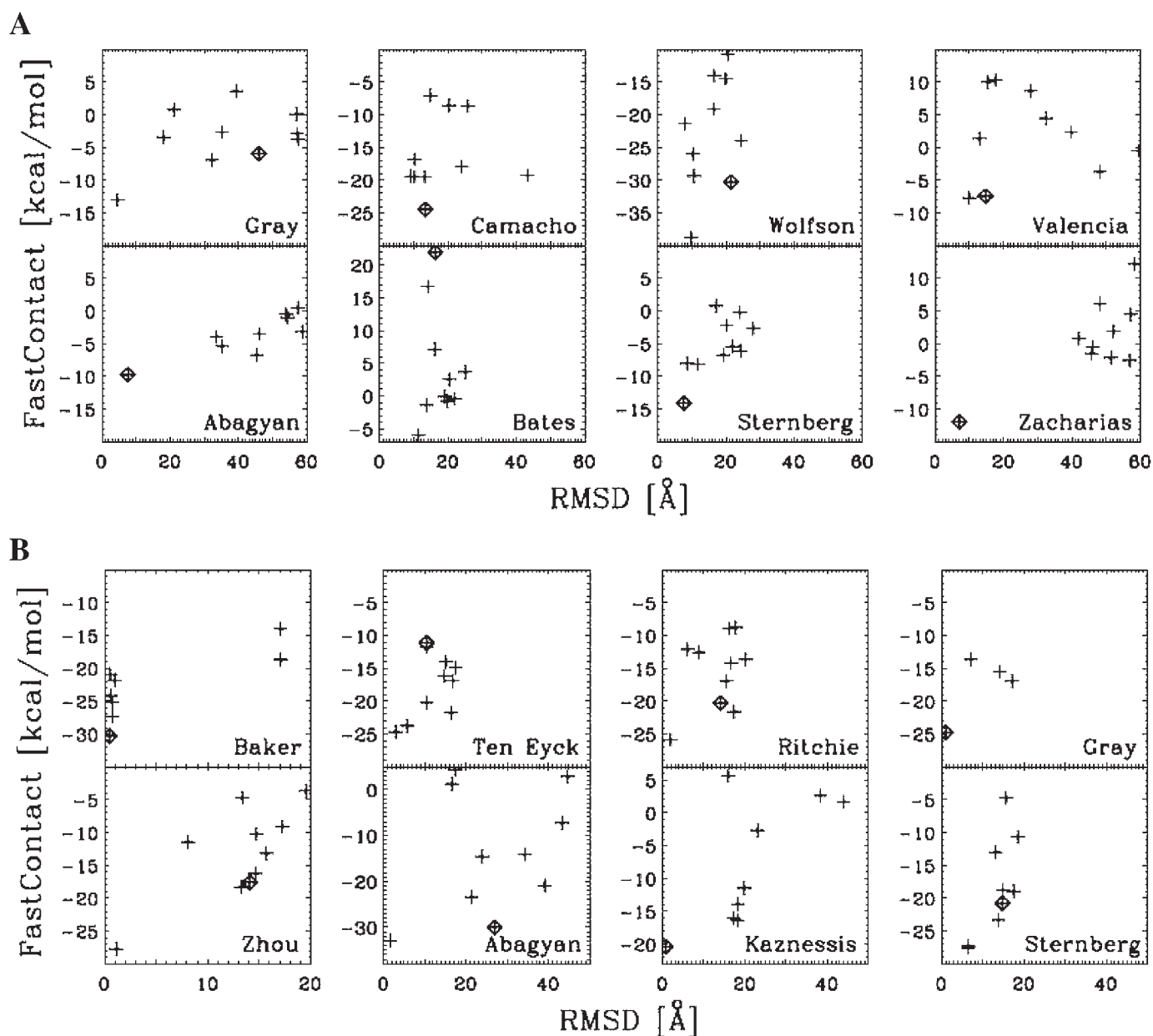
We are aware of only one server that estimate binding free energies of complex structures: <http://sparks.informatics.iupui.edu/czhang/complex.html> by the Zhou Lab (8). The server returns a single binding free energy estimate in kcal/mol that was shown to correlate with experimental values ( $\pm 1.8$  kcal/mol) for some 69 crystal structures.

## AVAILABILITY

The web server is available freely and without registration at: <http://structure.pitt.edu/servers/fastcontact/>

## ACKNOWLEDGEMENTS

The ‘FastContact’ Server has been thoroughly tested by over 500 runs from users all over the world. We are grateful to the many people around the world who tested our server and provided constructive feedback. This material is based upon work supported by the National Science Foundation under Grant No. MCB-0444291.



**Figure 3.** Examples of 'FastContact' scoring for a subset of high quality docked models from eight groups for targets 8 and 12 of CAPRI rounds 3-5, from <http://capri.ebi.ac.uk/>. For each of these targets, we run the models in our server and re-rank the models accordingly. In all cases, the server was able to correctly rank a low RMSD model as the one with the lowest free energy score. For comparison, we also marked with a diamond symbol the model ranked number 1 by the modeler (23-33). (A) target 8; (B) target 12. For a complete analysis of all models please see reference (10).

Funding to pay the Open Access publication charges for this article was provided by NSF.

*Conflict of interest statement.* None declared.

## REFERENCES

- Bueno, M., Camacho, C.J. and Sancho, J. (2007) SIMPLE estimate of the free energy change due to aliphatic mutations: superior predictions based on first principles. *Proteins*, (in press).
- Novotny, J., Bruccoleri, R.E. and Saul, F.A. (1989) On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry*, **28**, 4735-4749.
- Vajda, S., Weng, Z. and DeLisi, C. (1995) Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments. *Protein Eng.*, **8**, 1081-1092.
- Camacho, C.J. and Vajda, S. (2001) Protein docking along smooth association pathways. *Proc. Natl Acad. Sci. USA*, **98**, 10636-10641.
- Kimura, S.R., Brower, R.C., Vajda, S. and Camacho, C.J. (2001) Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys. J.*, **80**, 635-642.
- Vajda, S., Weng, Z., Rosenfeld, R. and DeLisi, C. (1994) Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry*, **33**, 13977-13988.
- Zhang, C., Vasmatzis, G., Cornette, J.L. and DeLisi, C. (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.*, **267**, 707-726.



8. Liu,S., Zhang,C., Zhou,H. and Zhou,Y. (2004) A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins*, **56**, 93–101.
9. Camacho,C.J. and Zhang,C. (2005) FastContact: rapid estimate of contact and binding free energies. *Bioinformatics*, **21**, 2534–2536.
10. Camacho,C.J., Ma,H. and Champ,P.C. (2006) Scoring a diverse set of high-quality docked conformations: a metascoring based on electrostatic and desolvation interactions. *Proteins*, **63**, 868–877.
11. Pickersgill,R.W. (1988) A rapid method of calculating charge-charge interaction energies in proteins. *Protein Eng.*, **2**, 247–248.
12. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J. and Karplus,M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.
13. Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
14. Comeau,S.R., Gatchell,D.W., Vajda,S. and Camacho,C.J. (2004) ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.*, **32**, W96–W99.
15. Comeau,S.R., Vajda,S. and Camacho,C.J. (2005) Performance of the first protein docking server ClusPro in CAPRI rounds 3-5. *Proteins*, **60**, 239–244.
16. Camacho,C.J., Gatchell,D.W., Kimura,S.R. and Vajda,S. (2000) Scoring docked conformations generated by rigid-body protein-protein docking. *Proteins*, **40**, 525–537.
17. Mendez,R., Leplae,R., De,M.L. and Wodak,S.J. (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins*, **52**, 51–67.
18. Mendez,R., Leplae,R., Lensink,M.F. and Wodak,S.J. (2005) Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, **60**, 150–169.
19. Camacho,C.J. and Gatchell,D.W. (2003) Successful discrimination of protein interactions. *Proteins*, **52**, 92–97.
20. Camacho,C.J. (2005) Modeling side-chains using molecular dynamics improve recognition of binding region in CAPRI targets. *Proteins*, **60**, 245–251.
21. Camacho,C.J., Weng,Z., Vajda,S. and DeLisi,C. (1999) Free energy landscapes of encounter complexes in protein-protein association. *Biophys. J.*, **76**, 1166–1178.
22. Comeau,S.R., Gatchell,D.W., Vajda,S. and Camacho,C.J. (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, **20**, 45–50.
23. Tress,M., de,J.D., Grana,O., Gomez,M.J., Gomez-Puertas,P., Gonzalez,J.M., Lopez,G. and Valencia,A. (2005) Scoring docking models with evolutionary information. *Proteins*, **60**, 275–280.
24. Mustard,D. and Ritchie,D.W. (2005) Docking essential dynamics eigenstructures. *Proteins*, **60**, 269–274.
25. Carter,P., Lesk,V.I., Islam,S.A. and Sternberg,M.J. (2005) Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. *Proteins*, **60**, 281–288.
26. Zacharias,M. (2005) ATTRACT: protein-protein docking in CAPRI using a reduced protein model. *Proteins*, **60**, 252–256.
27. Fernandez-Recio,J., Abagyan,R. and Totrov,M. (2005) Improving CAPRI predictions: optimized desolvation for rigid-body docking. *Proteins*, **60**, 308–313.
28. Daily,M.D., Masica,D., Sivasubramanian,A., Somarouthu,S. and Gray,J.J. (2005) CAPRI rounds 3-5 reveal promising successes and future challenges for RosettaDock. *Proteins*, **60**, 181–186.
29. Smith,G.R., Fitzjohn,P.W., Page,C.S. and Bates,P.A. (2005) Incorporation of flexibility into rigid-body docking: applications in rounds 3-5 of CAPRI. *Proteins*, **60**, 263–268.
30. Zhang,C., Liu,S. and Zhou,Y. (2005) Docking prediction using biological information, ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function. *Proteins*, **60**, 314–318.
31. Inbar,Y., Schneidman-Duhovny,D., Halperin,I., Oron,A., Nussinov,R. and Wolfson,H.J. (2005) Approaching the CAPRI challenge with an efficient geometry-based docking. *Proteins*, **60**, 217–223.
32. Schueler-Furman,O., Wang,C. and Baker,D. (2005) Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins*, **60**, 187–194.
33. Law,D., Hotchkko,M. and Ten,E.L. (2005) Progress in computation and amide hydrogen exchange for prediction of protein-protein complexes. *Proteins*, **60**, 302–307.