

## SwarmDock: a server for flexible protein–protein docking

Mieczyslaw Torchala<sup>1,†</sup>, Iain H. Moal<sup>2,†</sup>, Raphael A. G. Chaleil<sup>1</sup>, Juan Fernandez-Recio<sup>2</sup> and Paul A. Bates<sup>1,\*</sup>

<sup>1</sup>Biomolecular Modelling Laboratory, Cancer Research UK London Research Institute, London WC2A 3LY, UK and <sup>2</sup>Joint BSC-IRB Research Program in Computational Biology, Life Science Department, Barcelona Supercomputing Center, 08034 Barcelona, Spain

Associate Editor: Anna Tramontano

### ABSTRACT

**Summary:** Protein–protein interactions are central to almost all biological functions, and the atomic details of such interactions can yield insights into the mechanisms that underlie these functions. We present a web server that wraps and extends the SwarmDock flexible protein–protein docking algorithm. After uploading PDB files of the binding partners, the server generates low energy conformations and returns a ranked list of clustered docking poses and their corresponding structures. The user can perform full global docking, or focus on particular residues that are implicated in binding. The server is validated in the CAPRI blind docking experiment, against the most current docking benchmark, and against the ClusPro docking server, the highest performing server currently available.

**Availability:** The server is freely available and can be accessed at: <http://bmm.cancerresearchuk.org/%7ESwarmDock/>.

**Contact:** Paul.Bates@cancer.org.uk

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on November 28, 2012; revised on January 7, 2013; accepted on January 20, 2013

### 1 INTRODUCTION

The great majority of biological functions are mediated not by isolated proteins but by their interactions. Physically, any two proteins can interact. However, interactions of biological significance exist via the pairing of complementary surfaces, and modelling the three-dimensional structure of protein–protein complexes remains a highly active research field.

We recently described the SwarmDock flexible docking algorithm for docking and clustering (Li *et al.*, 2010; Moal and Bates, 2010). This algorithm has been used in the CAPRI blind docking experiment (<http://www.ebi.ac.uk/msd-srv/capri/>), where experimentally determined protein–protein complexes are held in confidence by the organizing committee while the prediction community are invited to predict their structure. Since the last assessment period in May 2010, SwarmDock correctly identified interfaces in seven of the nine targets, a success rate that surpasses all but one of the other participating teams. We have now

automated the docking pipeline, from structure preparation to clustering. We present here the SwarmDock server, which provides the research community with a simple and easy to use interface to the methodology. This automated server has been used in CAPRI without human intervention since round 26, and has correctly identified the structure of all three targets, two of which were not found by any other server. A typical docking run using the server will take up to 36 h, depending on the size of the complex and whether the resources are shared. The server does not, at this time, support non-protein docking.

### 2 IMPLEMENTATION

The SwarmDock algorithm has been described previously (Li *et al.*, 2010; Moal and Bates, 2010). Briefly, a combination of local docking and particle swarm optimization is used to find low energy positions and orientations of the binding partners. Normal modes are used as a component of the optimization vector to model transitions between unbound and bound conformations. The original scoring function has been replaced with DComplex (Liu *et al.*, 2004), and the structures are rescored using the centroid potential described by Tobi (2010) prior to clustering, as these modifications considerably improved the performance of the algorithm (results not shown).

Uploaded PDB structures of the binding partners should conform to three simple requirements: a TER keyword should be placed after each chain, standard residues should be used and structures should preferably not have missing residues. However, if the last two requirements cannot be met, the server will try to replace non-standard residues and model missing residues or residues with missing atoms. After the structures are repaired, they are minimized, docked, minimized again, and then clustered as described previously.

To submit a job, the user can choose between the full-blind docking or local docking with restraints. Starting points are generated uniformly around the receptor. In the latter method, the user may choose the residues belonging to the binding site and consequently the server only uses the starting points in the line of sight of at least one of the chosen receptor residues. Consequently, the server only produces solutions in the region of the receptor site chosen by the user.

Upon completion of the computations, the user receives an email with a link to download the repaired input structures, the docked structures, a ranked list of clusters, details about the residue contacts and the SwarmDock output file.

\*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

**Table 1.** The performance of SwarmDock and ClusPro (high/medium or better/acceptable or better)

Rank	SwarmDock—all 176						SwarmDock—29						ClusPro—29					
	Lowest only			Best			Lowest only			Best			Balanced			Electrostatic		
	High	Medium	Acceptable	High	Medium	Acceptable	High	Medium	Acceptable	High	Medium	Acceptable	High	Medium	Acceptable	High	Medium	Acceptable
Top 1	1	14	18	2	14	19	0	3	3	3	3	0	1	3	0	2	3	3
%	0.6	8.0	10.2	1.1	8.0	10.8	0.0	10.3	10.3	10.3	10.3	0.0	3.4	10.3	0.0	6.9	10.3	10.3
Top 5	3	31	49	7	34	51	1	8	12	12	12	1	3	8	0	5	9	9
%	1.7	17.6	27.8	4.0	19.3	29.0	3.4	27.6	41.4	41.4	41.4	3.4	10.3	27.6	0.0	17.2	31.0	31.0
Top 10	4	38	64	7	42	64	1	8	12	12	12	0	5	10	0	6	10	10
%	2.3	21.6	36.4	4.0	23.9	36.4	3.4	27.6	41.4	41.4	41.4	0.0	17.2	34.5	0.0	20.7	34.5	34.5
Top 50	6	54	100	9	61	101	1	10	17	17	17	0	6	11	0	7	13	13
%	3.4	30.7	56.8	5.1	34.7	57.4	3.4	34.5	58.6	58.6	58.6	0.0	20.7	37.9	0.0	24.1	44.8	44.8
Top 100	6	67	114	9	71	115	1	13	18	18	18	0	6	11	0	7	13	13
%	3.4	38.1	64.8	5.1	40.3	65.3	3.4	44.8	62.1	62.1	62.1	0.0	20.7	37.9	0.0	24.1	44.8	44.8
All	6	71	125	9	74	126	1	13	22	22	22	0	6	11	0	7	13	13
%	3.4	40.3	71.0	5.1	42.0	71.6	3.4	44.8	75.9	75.9	75.9	0.0	20.7	37.9	0.0	24.1	44.8	44.8

Note: For SwarmDock, results are shown for all 176 structures in the Benchmark 4.0, as well as the subset of 29 structures. Results are given when considering only the lowest energy member of each cluster, or all members of each cluster. For ClusPro, results are shown when using the balanced ClusPro scoring function and the electrostatically biased scoring function. The other scoring functions evaluated did not perform as well and are not shown.

### 3 BENCHMARK

Benchmark 4.0 (Hwang *et al.*, 2010) is the most current and up to date benchmark for docking algorithms. Unbound structures were repaired by the SwarmDock server and all the 176 cases (121 rigid body, 30 medium difficulty, 25 difficult) were re-docked, from the unbound receptor/ligand conformations. The server correctly parsed the structures and returned results for all 176 cases. The docking was evaluated by calculating interface RMSD, ligand RMSD and the fraction of native and non-native contacts, and solutions were then classified as incorrect, acceptable, medium or high quality, in accordance with the CAPRI criteria (Lensink *et al.*, 2007). The detailed results are presented in the Supplementary Material. The success rate for blind docking, the number of complexes for which at least one acceptable solution was found, is 71.6%. A more detailed analysis of the results is shown in Table 1.

We also benchmarked SwarmDock against ClusPro, a popular protein–protein docking server (Comeau *et al.*, 2004). ClusPro has consistently outperformed all other servers in the previous and current CAPRI assessment periods, and thus represents the state-of-the-art. We chose a subset of structures to compare the servers: all 19 medium difficulty and difficult complexes from update 4.0 of the Benchmark and 10 rigid-body complexes chosen uniformly. ClusPro returned solutions for all 29 cases. In 7 cases, both servers did not find a solution. In 9 cases, SwarmDock found a correct solution whereas ClusPro did not, and there were no cases of ClusPro finding a solution where SwarmDock could not. Further comparison between the servers is presented in Table 1, which also shows the higher performance of SwarmDock in terms of the quality of docked solutions and their ranking.

### 4 CONCLUSIONS

The SwarmDock server is a tool for predicting the three-dimensional structure of protein–protein complexes. Uploaded structures are automatically repaired, and the server is robust and easy to use. Docking may be performed in two modes: as full blind when nothing is known about interface residues, or in restrained mode, when information about interface residues may be used to restrict the search space. For the full blind mode, tests were performed on the most recent docking benchmark, achieving a success rate of 71.6% when considering all solutions, and 36.4% when considering only the top 10. The full blind mode was also compared with ClusPro, the most successful docking server to date. The SwarmDock algorithm has also been used for generating structures for one of highest performing groups in CAPRI, and the automated server presented here has correctly identified the structure of all three complexes in the CAPRI rounds it has participated in, including two targets for which no other servers returned correct results.

*Funding:* Cancer Research UK and the Spanish Ministry of Science (BIO2010-22324).

*Conflict of Interest:* none declared.

### REFERENCES

Comeau, S.R. *et al.* (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, **20**, 45–50.

- 
- Hwang,H. *et al.* (2010) Protein-protein docking benchmark version 4.0. *Proteins*, **78**, 3111–3114.
- Lensink,M.F. *et al.* (2007) Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins*, **69**, 704–718.
- Li,X.F. *et al.* (2010) Detection and refinement of encounter complexes for protein-protein docking: tacking account of macromolecular crowding. *Proteins*, **78**, 3189–3196.
- Liu,S. *et al.* (2004) A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. *Proteins*, **56**, 93–101.
- Moal,I.H. and Bates,P.A. (2010) SwarmDock and the use of normal modes in protein-protein docking. *Int. J. Mol. Sci.*, **11**, 3623–3648.
- Tobi,D. (2010) Designing coarse grained-and atom based-potentials for protein-protein docking. *BMC Struct. Biol.*, **10**, 40.