Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions

Nir London[1], Barak Raveh[1,2], Eyal Cohen[3], Guy Fathi[3] and Ora Schueler-Furman[1,*]

[1]Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Hadassah Medical School, The Hebrew University, Jerusalem 91120, [2]The Blavatnik School of Computer Science, Tel-Aviv University, Ramat Aviv 69978, Israel and [3]Undergraduate Program of the School of Computer Science and Engineering, The Hebrew University, 91904 Jerusalem, Israel

ABSTRACT

Peptide–protein interactions are among the most prevalent and important interactions in the cell, but a large fraction of those interactions lack detailed structural characterization. The Rosetta FlexPepDock web server (http://flexpepdock.furmanlab.cs.huji.ac.il/) provides an interface to a high-resolution peptide docking (refinement) protocol for the modeling of peptide–protein complexes, implemented within the Rosetta framework. Given a protein receptor structure and an approximate, possibly inaccurate model of the peptide within the receptor binding site, the FlexPepDock server refines the peptide to high resolution, allowing full flexibility to the peptide backbone and to all side chains. This protocol was extensively tested and benchmarked on a wide array of non-redundant peptide–protein complexes, and was proven effective when applied to peptide starting conformations within 5.5 Å backbone root mean square deviation from the native conformation. FlexPepDock has been applied to several systems that are mediated and regulated by peptide–protein interactions. This easy to use and general web server interface allows non-expert users to accurately model their specific peptide–protein interaction of interest.

INTRODUCTION

Protein–protein interactions facilitate most cellular processes. It has lately become apparent that a significant fraction of these interactions are mediated by peptide–protein interactions, which involve the binding of a linear, unfolded peptide stretch onto a globular protein receptor (1–3). Peptide-mediated interactions indeed play key roles in major cellular processes, predominantly in signaling and regulatory networks that require short-lived signals (4), and also in cell localization, protein degradation and immune response (3,4). However, despite their importance and estimated abundance, peptide-protein complexes are underrepresented among solved structures (5,6). Therefore, protocols that can provide accurate structural models of peptide–protein interactions represent an essential tool for the molecular understanding of the cellular network of interactions (7). These models can then be used as ideal starting points for targeted computational and experimental modulation of interactions (8,9). For many real-life peptide docking problems, coarse-grain models can be often obtained from complexes with alternative peptides, unbound structures or homology models where existing structures provide approximate structural information about the receptor and the peptide or the location of the binding site [e.g. peptides that bind to MHC, SH3, WW or PDZ domains (10–13)].

Rosetta FlexPepDock (14) is a high-resolution protocol for the refinement of peptide-protein complex structures that is implemented in the Rosetta modeling suite framework (15). Starting from a coarse model of the interaction, FlexPepDock performs a Monte Carlo-Minimization-based approach to refine all the peptide's degrees of freedom (rigid body orientation, backbone and side chain flexibility) as well as the protein receptor side chains conformations. The Rosetta FlexPepDock web server described here provides a simple interface for the usage of this protocol, and by this aims to increase the accessibility of structural models of peptide–protein interactions to a broad range of scientists.

While a plethora of web servers is available for the docking of a pair of globular proteins [e.g. RosettaDock (16), HADDOCK (17), PatchDock (18), ClusPro (19) and more; see CAPRI (20)], these are not intended for the docking of peptides. In particular, they do not consider the flexibility of the protein backbone during the docking process, and are thus not suitable for the docking of flexible peptides. Web servers are also available for

small-molecule docking [e.g. Autodock (21), DOCK (22), PatchDock (18), ParDock (23), MEDdock (24) and others]. These servers, however, are suitable for molecules with a limited number of rotatable bonds only, and therefore not applicable to peptides, which typically contain many more internal degrees of freedom than small molecules (14,25). Other servers might identify the rough orientation of the peptide (and can serve as a complementary, preliminary step to FlexPepDock), but do not actually model the peptide–protein complex. These include CASTp (26), which aims at detecting pockets on protein surfaces [we previously showed that this feature correlates with peptide binding sites (5)], and PepSite (27), which predicts peptide binding sites and provides a coarse prediction of specific peptide residue locations. Finally, other software that models peptide–protein complexes such as DynaDock (28), or system-specific software for modeling, e.g. PDZ–peptide interactions (29) or MHC–peptide interactions (30), are to our knowledge not accessible to the public in the form of a web server. Consequently, the Rosetta FlexPepDock web server presented here is currently the only server that allows for high-resolution modeling of peptide–protein interactions.

The performance of Rosetta FlexPepDock has been extensively tested against a large set of perturbed peptide–protein complexes and an effective range of sampling was defined (14). Table 1 summarizes the performance of FlexPepDock over a bound docking benchmark that covers a wide range of increasingly divergent starting peptide conformations. More analyses of its performance can be found in Raveh *et al.* (14). For peptides with initial backbone (bb) root mean square deviation (RMSD) of up to $5.5\,\text{Å}$, FlexPepDock is able to create near-native models (peptide bb-RMSD $<2\,\text{Å}$) in 91% of the cases for the bound receptor, and rank them as one of the top five models in 78%. Moreover, the side chains of key residues

in binding motifs are modeled particularly well, typically within $1\,\text{Å}$ of their native conformations (14). In the challenging task of unbound docking, near-native models were sampled in 85% of the cases and ranked correctly in 59% (for starting structures within $5.5\,\text{Å}$ bb-RMSD from the native conformation).

In cases where no information is available about the conformation of the peptide backbone, docking can be started from an extended peptide conformation. In a benchmark in which the peptide was docked starting from an ideal extended backbone conformation ($\pm135°$ for all φ/ψ angles) based on a single anchor residue, near-native solutions could be sampled in 66% of the 71 non-helical complexes (31% $<1\,\text{Å}$ from native), and ranked among the top five solutions in 49% of the cases (24% for $<1\,\text{Å}$ from native).

Rosetta FlexPepDock was tested on peptides of length 5–15 amino acids, and performance shows little to no dependency on the peptide length (see Supplementary Table S1). However, we have also repeatedly applied it successfully to longer peptides.

DESCRIPTION OF WEB SERVER

The main input for the Rosetta FlexPepDock web server is a PDB (31) file of the estimated complex between the receptor (first chain) and the peptide (second chain). The server will dock the peptide starting from this initial conformation. If the native conformation of the peptide lies within the effective range of the protocol (see above), it will most probably produce high-resolution models for this interaction.

Using the default options, the server will perform 100 simulations in full-atom mode and 100 simulations that include a preceding low-resolution centroid-based optimization protocol [see Raveh *et al.* (14) for more details about the protocol]. It will then rank the total of 200 created models by their Rosetta energy score and provide the user with the top 10 predicted models for this interaction, as well as their score and bb-RMSD from the starting conformation. In addition, a plot showing score versus RMSD for each of the created 200 models provides information about overall sampling (Figure 1B).

Table 1. FlexPepDock performance as a function of the starting peptide bb-RMSD

Start RMSD (Å)[b]	Sub-angstrom ($<1\,\text{Å}$)[a]			Near-native ($<2\,\text{Å}$)[a]			Cases (n)
	Rank 1[c] (%)	Top 10[c] (%)	All 200[c] (%)	Rank 1 (%)	Top 10 (%)	All 200 (%)	
0–0.5	61.6	93.0	100.0	93.0	98.8	100.0	86
0.5–1.5	61.6	91.3	97.1	94.2	97.1	99.3	138
1.5–2.5	47.6	77.2	91.8	82.3	93.9	99.7	294
2.5–3.5	36.7	61.5	77.9	67.9	86.9	96.7	390
3.5–4.5	23.4	42.6	54.4	49.8	72.4	85.7	406
4.5–5.5	22.5	41.5	52.2	47.8	65.8	77.8	383
5.5–6.5	17.6	28.7	37.5	30.7	48.3	63.1	352
6.5–7.5	13.6	20.4	27.5	25.5	37.4	48.2	353
7.5–8.5	10.1	17.5	22.4	19.9	28.3	38.8	286
8.5–9.5	5.8	9.7	15.1	14.7	23.3	34.9	258
>9.5	4.7	7.2	10.0	10.3	14.8	21.1	622

[a]We measure the performance by two success criteria—a model is considered successful if the peptide interface bb-RMSD to native is $<1\,\text{Å}$ (sub-angstrom) or $<2\,\text{Å}$ (near-native).
[b]Starting structures were binned according to the starting peptide conformation bb-RMSD. In this case, the bound receptor was used for docking.
[c]Performance when considering just the Top 1 ranking model by energy, Top 10 ranking models, or the entire sample of 200 models.
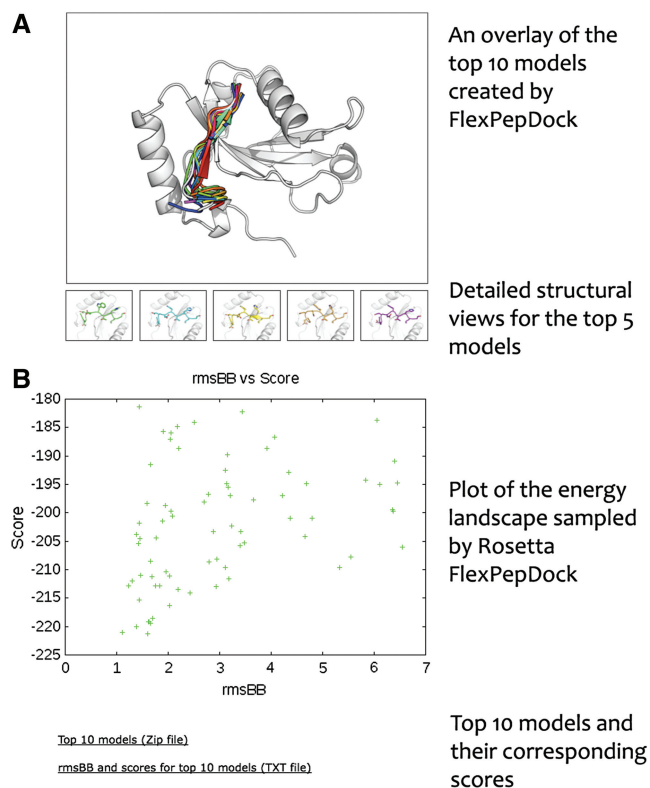
ADVANCED OPTIONS

For more advanced runs, users are able to specify:

- A reference PDB: the user can upload a reference PDB of the peptide–protein interaction. If so, RMSD values of the models will be calculated to the reference peptide conformation found in this file, rather than to the starting conformation. This is useful if for example a structure of a similar interaction is available.
- A constraints file: the user can upload a file that specifies distance constraints between different atoms in the system. This allows the users to incorporate previous experimental knowledge and their intuitions into the simulations. For instance, the distance between a catalytic residue in the receptor and a modified residue

A

An overlay of the top 10 models created by FlexPepDock

Detailed structural views for the top 5 models

B

Plot of the energy landscape sampled by Rosetta FlexPepDock

Top 10 models (Zip file)

rmsBB and scores for top 10 models (TXT file)

Top 10 models and their corresponding scores

Figure 1. Results provided for an example peptide docking run. (**A**) Graphical representation of the top 10 models (superimposed), as well as more detailed figures of the top 5 models (second row). (**B**) Plot of RMSD (*x*-axis) vs score (*y*-axis) of all models created by the simulation run. Bottom panel: The top 10 models (PDB format coordinates), as well as a score file can be downloaded via the provided links. This example is based on a 4.9A bb-RMSD starting conformation and is taken from line 6 in Table 1.

in the peptide, or the distances derived from cross-linking experiments can easily be reinforced with this setup. Another case in which constraints may be useful is if the user wants to fix certain interactions that are present in the starting model.

• The amount of models created with or without the low-resolution preoptimization stage. In cases of high confidence in the initial peptide placement (e.g. when only one point mutation is introduced into an existing structure), the user might want to avoid the larger range sampling of this low-resolution stage and focus the sampling on a closer range. When the initial placement is less confident (e.g. point mutations in the protein indicate a putative binding site, but the exact orientation of the peptide is not known), more sampling with this low-resolution stage might increase the sampling range and allow the identification of the correct conformation.

• Reported scoring terms in the output file: the user can specify which specific Rosetta scoring terms will be reported for the top 10 models, e.g. Lennard-Jones full atom attractive and repulsive terms, the Lazaridis-Karplus solvation term (32), hydrogen bonding score (33) and others.

• Modified amino acids: currently, the server supports the docking of peptides containing modified amino acids (such as phosphorylation, acetylation, etc.) only in high-resolution mode (i.e. without low-resolution preoptimization). When submitting such complexes, the user should consult the FAQ page for exact format of the modified residue. Other non-natural amino that are unrecognized by the server will be ignored.

SUMMARY

We describe here an easy-to-use web server interface to the Rosetta FlexPepDock protocol for the high-resolution modeling of peptide–protein interactions. FlexPepDock has recently been used by us to successfully address several 'real world' modeling tasks (34–37) and we expect that increasing its usability through this web server will open the door for a wide range of new systems and applications.

We have recently extended the FlexPepDock protocol and introduced 'FlexPepDock *ab-initio*', a powerful protocol for simultaneous *de novo* folding and docking of peptides at a known binding site that does not require an initial peptide backbone conformation. FlexPepDock *ab initio* performed well on a benchmark of peptide–protein interactions (38). This protocol is however computationally expensive and therefore not yet available on the web server. It can be downloaded as part of the next Rosetta release.

METHODS

Overview of the protocol

Rosetta FlexPepDock is extensively described in Raveh *et al*. (14). We provide here a short overview of the protocol. The first step in our protocol involves the 'pre-packing' of the input structure, to remove internal clashes: side chain conformations are optimized by determining the best rotamer combination for both the protein and the peptide separately. In order to create a single model, we conduct 10 outer cycles of optimization starting with a reduced repulsive van der Waals term and increased attractive van der Waals term. During refinement, the repulsive and attractive terms are gradually ramped back towards their original values so that in the last cycle the energy function corresponds to the standard Rosetta score. Within each outer cycle, we first optimize the rigid body orientation between the protein and the peptide, and then optimize the peptide backbone for the new orientation, both using Monte Carlo search with energy minimization. Side chain rotamers are recalculated for the interface on-the-fly.

Pre-optimization in low-resolution

We provide an optional fast, low-resolution optimization step prior to the full atom optimization. In this step, side chains are represented as spherical centroids of variable size. Similar to the high-resolution protocol, the rigid body and peptide backbone degrees of freedom are optimized

alternately for several cycles. In this low-resolution representation, sampling range is usually increased.

Rosetta infrastructure

The server is based on Rosetta Release version 3.2 and implements the following command line:

```
FlexPepDocking.release -database minirosetta_database
  -s start.pdb -native ref.pdb -rbMCM -torsionsMCM
  -ex1 -ex2aro -use_input_sc –unboundrot start.pdb
  -scorefile score.sc -ignore_unrecognized_res -nstruct
  200 [-lowres_preoptimize]
```

Queue page

The life cycle of a modeling job submitted by the user goes through the following stages: (i) queued: waiting to be processed by the server; (ii) pre-packing: processing of the input file and repacking of the side chains in each monomer (protein and peptide) to remove internal clashes that are not related to intermolecular interactions; (iii) FlexPepDocking: the actual production run—creation of the requested number of models by high-resolution refinement. This stage consumes the major part of the running time; (iv) processing results: creation of visual representation of results; (v) writing results: creation of the results page; (vi) completed: job has ended successfully or, alternatively, if job failed, the user will receive an error description by e-mail (and a corresponding message in the results page).

Documentation

In addition to an overview page that briefly describes the underlying protocol, and provides information about prior benchmarking results (http://flexpepdock .furmanlab.cs.huji.ac.il/overview.php), users can also read more on the Usage and Frequently Asked Questions (FAQ) page (http://flexpepdock.furmanlab.cs .huji.ac.il/usage.php), which provides details about the input and output of the server, as well as gives answers to the most common anticipated problems. Finally, results of a demo run are also available (http://flexpepdock .furmanlab.cs.huji.ac.il/demo/index.php).

Registration

This web site is free and open to all users and there is no login requirement. However, users can supply an e-mail address (highly recommended), which allows them to receive a notification via e-mail after a simulation finishes, as well as a convenient link to the results page.

System architecture

The server runs on an AMD Sun Cluster of 40 CPUs. Running time of a single simulation takes ~3 min (depending on the peptide and receptor sizes). Each user submitted job is distributed on 6 CPUs and finishes within 1.5–2 h if the queue is empty. Data management is based on an MySQL server (v.5.1.34).

REFERENCES

1. Neduva,V., Linding,R., Su-Angrand,I., Stark,A., de Masi,F., Gibson,T.J., Lewis,J., Serrano,L. and Russell,R.B. (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.*, **3**, e405.
2. Neduva,V. and Russell,R.B. (2006) Peptides mediating interaction networks: new leads at last. *Curr. Opin. Biotechnol.*, **17**, 465–471.
3. Petsalaki,E. and Russell,R.B. (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.*, **19**, 344–350.
4. Pawson,T. and Nash,P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.
5. London,N., Movshovitz-Attias,D. and Schueler-Furman,O. (2010) The structural basis of peptide-protein binding strategies. *Structure*, **18**, 188–199.
6. Vanhee,P., Reumers,J., Stricher,F., Baeten,L., Serrano,L., Schymkowitz,J. and Rousseau,F. (2010) PepX: a structural database of non-redundant protein-peptide complexes. *Nucleic Acids Res.*, **38**, D545–D551.
7. Stein,A., Mosca,R. and Aloy,P. (2011) Three-dimensional modeling of protein interactions and complexes is going 'omics. *Curr. Opin. Struct. Biol.*, **21**, 200–208.
8. Vanhee,P., van der Sloot,A.M., Verschueren,E., Serrano,L., Rousseau,F. and Schymkowitz,J. (2011) Computational design of peptide ligands. *Trends Biotechnol.*, **29**, 231–239.
9. London,N., Raveh,B., Movshovitz-Attias,D. and Schueler-Furman,O. (2010) Can Self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? *Proteins*, **78**, 3140–3149.
10. Madden,D.R., Gorga,J.C., Strominger,J.L. and Wiley,D.C. (1992) The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. *Cell*, **70**, 1035–1048.
11. Cesareni,G., Panni,S., Nardelli,G. and Castagnoli,L. (2002) Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett.*, **513**, 38–44.

12. Songyang,Z., Fanning,A.S., Fu,C., Xu,J., Marfatia,S.M., Chishti,A.H., Crompton,A., Chan,A.C., Anderson,J.M. and Cantley,L.C. (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*, **275**, 73–77.

13. Sudol,M. (1996) Structure and function of the WW domain. *Prog. Biophys. Mol. Biol.*, **65**, 113–132.

14. Raveh,B., London,N. and Schueler-Furman,O. (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins*, **78**, 2029–2040.

15. Das,R. and Baker,D. (2008) Macromolecular modeling with rosetta. *Annu. Rev. Biochem.*, **77**, 363–382.

16. Lyskov,S. and Gray,J.J. (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.*, **36**, W233–W238.

17. de Vries,S.J., van Dijk,M. and Bonvin,A.M. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.*, **5**, 883–897.

18. Schneidman-Duhovny,D., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.

19. Comeau,S.R., Kozakov,D., Brenke,R., Shen,Y., Beglov,D. and Vajda,S. (2007) ClusPro: performance in CAPRI rounds 6-11 and the new server. *Proteins*, **69**, 781–785.

20. Lensink,M.F. and Wodak,S.J. (2010) Docking and scoring protein interactions: CAPRI 2009. *Proteins*, **78**, 3073–3084.

21. Goodsell,D.S., Morris,G.M. and Olson,A.J. (1996) Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.*, **9**, 1–5.

22. Irwin,J.J., Shoichet,B.K., Mysinger,M.M., Huang,N., Colizzi,F., Wassam,P. and Cao,Y. (2009) Automated docking screens: a feasibility study. *J. Med. Chem.*, **52**, 5712–5720.

23. Gupta,A., Gandhimathi,A., Sharma,P. and Jayaram,B. (2007) ParDOCK: an all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein Pept. Lett.*, **14**, 632–646.

24. Chang,D.T., Oyang,Y.J. and Lin,J.H. (2005) MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res.*, **33**, W233–W238.

25. Rubinstein,M. and Niv,M.Y. (2009) Peptidic modulators of protein-protein interactions: progress and challenges in computational design. *Biopolymers*, **91**, 505–513.

26. Dundas,J., Ouyang,Z., Tseng,J., Binkowski,A., Turpaz,Y. and Liang,J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.*, **34**, W116–W118.

27. Petsalaki,E., Stark,A., Garcia-Urdiales,E. and Russell,R.B. (2009) Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput. Biol.*, **5**, e1000335.

28. Antes,I. (2010) DynaDock: a new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins*, **78**, 1084–1104.

29. Niv,M.Y. and Weinstein,H. (2005) A flexible docking procedure for the exploration of peptide binding selectivity to known structures and homology models of PDZ domains. *J. Am. Chem. Soc.*, **127**, 14072–14079.

30. Bordner,A.J. and Abagyan,R. (2006) Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins*, **63**, 512–526.

31. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

32. Lazaridis,T. and Karplus,M. (1999) Effective energy function for proteins in solution. *Proteins*, **35**, 133–152.

33. Kortemme,T., Morozov,A.V. and Baker,D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.*, **326**, 1239–1259.

34. Belitsky,M., Avshalom,H., Erental,A., Yelin,I., Kumar,S., London,N., Sperber,M., Schueler-Furman,O. and Engelberg-Kulka,H. (2011) The Escherichia coli extracellular death factor EDF induces the endoribonucleolytic activities of the toxins MazF and ChpBK. *Mol. Cell*, **41**, 625–635.

35. Crawley,S.W., Gharaei,M.S., Ye,Q., Yang,Y., Raveh,B., London,N., Schueler-Furman,O., Jia,Z. and Cote,G.P. (2011) Autophosphorylation activates Dictyostelium myosin II heavy chain kinase A by providing a ligand for an allosteric binding site in the alpha-kinase domain. *J. Biol. Chem.*, **286**, 2607–2616.

36. Buch,I., Fishelovitch,D., London,N., Raveh,B., Wolfson,H.J. and Nussinov,R. (2010) Allosteric regulation of glycogen synthase kinase 3beta: a theoretical study. *Biochemistry*, **49**, 10890–10901.

37. Chopra,N., Agarwal,S., Verma,S., Bhatnagar,S. and Bhatnagar,R. (2011) Modeling of the structure and interactions of the B. anthracis antitoxin, MoxX: deletion mutant studies highlight its modular structure and repressor function. *J. Comput. Aided Mol. Des.*, **25**, 275–291.

38. Raveh,B., London,N., Zimmerman,L. and Schueler-Furman,O. (2011) Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PLoS ONE*, **6**, e18934.