PEP-FOLD3: faster *denovo* structure prediction for linear peptides in solution and in complex

Alexis Lamiable[1], Pierre Thévenet[1], Julien Rey[1], Marek Vavrusa[1], Philippe Derreumaux[2] and Pierre Tufféry[1,*]

[1]Molécules Thérapeutiques in Silico, RPBS, INSERM UMR-S 973, Université Paris Diderot, Sorbonne Paris Cité, 75205 Paris Cedex 13, France and [2]Institut de Biologie Physico Chimique, Laboratoire de Biochimie Théorique, Université Paris Diderot, Sorbonne Paris Cité, CNRS UPR 9080, 75005 Paris, France

ABSTRACT

Structure determination of linear peptides of 5–50 amino acids in aqueous solution and interacting with proteins is a key aspect in structural biology. PEP-FOLD3 is a novel computational framework, that allows both (i) *de novo* free or biased prediction for linear peptides between 5 and 50 amino acids, and (ii) the generation of native-like conformations of peptides interacting with a protein when the interaction site is known in advance. PEP-FOLD3 is fast, and usually returns solutions in a few minutes. Testing PEP-FOLD3 on 56 peptides in aqueous solution led to experimental-like conformations for 80% of the targets. Using a benchmark of 61 peptide–protein targets starting from the unbound form of the protein receptor, PEP-FOLD3 was able to generate peptide poses deviating on average by 3.3Å from the experimental conformation and return a native-like pose in the first 10 clusters for 52% of the targets. PEP-FOLD3 is available at http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3.

INTRODUCTION

Peptides are a category of compounds that meets increasing interest as candidate therapeutics (1), and motivates numerous efforts. Owing to the progress of sequencing techniques, an increasing number of peptide sequences, validated or hypothetical is available. Over 1 800 000 candidate sequences are identified among all prokaryotes (2), and several millions of new sequences are expected from venom analyses (3), but their structures remain largely unknown. The rational design of peptides at protein/protein (4–6) and protein/membrane (7) interfaces is also raising interest (8,9). Overall, computational tools to assist the structural characterization of peptidic sequences and the exploration of peptide–protein interactions are needed (10).

Presently, there are only a few on-line tools specialized for the structural characterization of a peptidic sequence. These include facilities such as CycloPS (11) for the preparation of cyclic peptides, and servers for the *de novo* structure prediction of peptides, including PEPStr and PEP-StrMod (12), PEP-FOLD (13,14) and PEPLook (15). The Quark server (16) can only be used for peptide sizes of more than 20 amino acids. Other tools focus on the characterization of peptide–protein interactions. PEPSite (17) and PEP-SiteFinder (18) have been designed to assist the identification of peptide binding sites on protein surfaces. The RosettaFlexPepDock (19) server will refine the structure of a peptide bound to a receptor, Galaxypepdock (20) performs similarity based docking using existing templates and the CABS-dock server (21) performs a blind docking which can be time consuming. Other tools such as pepATTRACT (22) propose to prepare files to run simulations locally, or to start from 3D coordinates for the HADDOCK server (23). Table 1 summarizes the characteristics of the servers that accept as input a sequence and will return a 3D structure.

Here we present PEP-FOLD3, which is an evolution of the former PEP-FOLD server, which has proven useful to the community (24). It comes with an improved and faster peptide *de novo* structure modeling engine, which makes it possible to open the service to a wide range of peptide sizes, from 5 to 50 amino acids. We also introduce new facilities to bias peptide conformational sampling by the explicit specification of parts of decoy structures, which makes it possible to refine/revisit models. Finally, PEP-FOLD3 also comes with the possibility to generate conformations for protein–peptide complexes, guided by a user-defined patch. Similarly to previous versions, PEP-FOLD3 makes use of a coarse grained representation: the predicted conformation of the complex must be considered as a starting point for further modeling at high resolution. To the best of our knowledge, PEP-FOLD3 is the first on-line computational framework that allows the determination of a peptide conformation either free in solution, or in interaction with a protein.

---

*To whom correspondence should be addressed. Tel: +33 1 5727 8374; Fax: +33 1 5727 8372; Email: pierre.tuffery@univ-paris-diderot.fr

Table 1. Servers proposing the structure prediction of peptides in isolation (free) or in interaction (bound)

	min	max	free	bound
PepStrMod	7	25	✓	
Peplook	2	30	✓	
PEP-FOLD3	5	50	✓	✓
Quark	20	200	✓	
CABS-dock	4	30		✓
GalaxyPepDock		30		✓
RosettaFlexPepDock	5	15		✓

Min, max correspond to the minimal and maximal size of the peptides accepted.
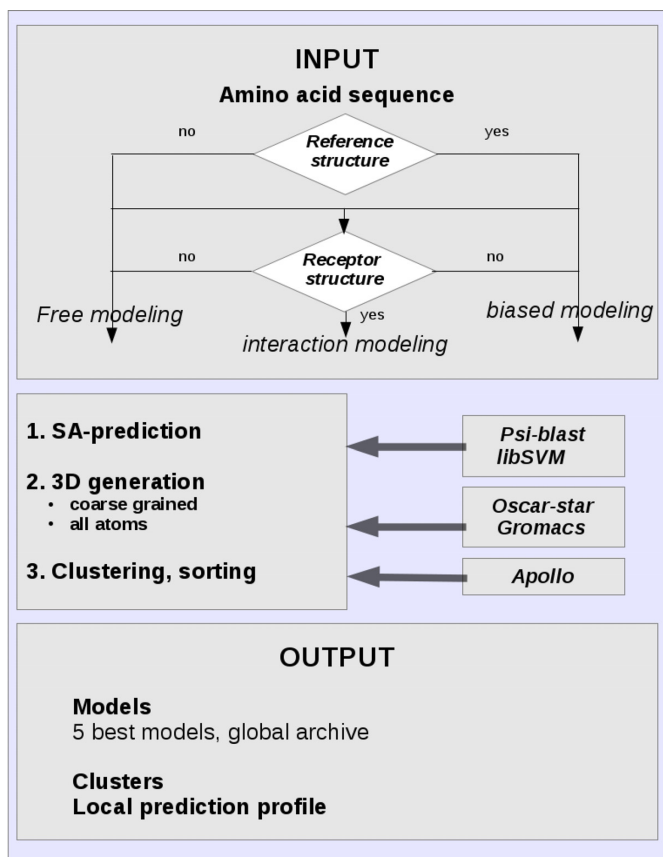
Figure 1. PEP-FOLD3 protocol for linear peptides in solution and in complex.

MATERIALS AND METHODS

PEP-FOLD3 standard protocol

The PEP-FOLD3 protocol is summarized in Figure 1. It relies on three main steps. First, it is important to understand that PEP-FOLD relies on the concept of structural alphabet (SA), a probabilistic framework that can be considered as a generalization of the concept of secondary structure. Here, peptides are described as series of fragments of four amino acids, overlapping by three. Each fragment is associated with geometric descriptors and can be emitted by the 27 states of a Hidden Markov Model as previously described (25). Given the 3D conformation of a peptide of L amino acids, one can calculate the values of the geometric descriptors of each fragment and identify the series of $L -$ 3 states that best describe the conformation, using standard Hidden Markov Models algorithms such as the Viterbi algorithm or the Forward Backward algorithm.

Starting from an amino acid sequence, the first step is to predict the prior probabilities (SA profile) of each fragment of the peptide to be associated with each of the 27 states of the SA. This is achieved by a Support Vector Machine that takes as input the 160 values corresponding to 8 series of 20 values. The 8 series correspond to the 4 amino acids of the fragment extended by two positions each side, and the 20 values to the PSSM inferred by PSI-BLAST (26). These prior probabilities can be biased if requested (see below).

In the second step, and unlike previous versions of PEP-FOLD, we now use the Forward Backtrack algorithm or a Taboo Sampling algorithm we have developed to generate sub-optimal series of states or trajectories - detailed performance analysis will be reported elsewhere. Each series of states then leads to the generation of one conformation using a rigid assembly procedure of prototype fragments. As previously, we grow the peptide one residue at a time. This model generation is performed in the sOPEP coarse grained representation (one bead per side chain)(27). Once the peptide is complete, it is refined using 30 000 Monte-Carlo steps.

In our experience, the number of sub-optimal trajectories required to identify native of near native models can be as low as 200. This makes the 3D generation step of PEP-FOLD $10\times$ faster than that of PEP-FOLD2, making it possible to open the PEP-FOLD3 server for peptides up to 50 amino acids while keeping the processing time reasonable. In addition, the lower limit of the peptides accepted is now 5 amino acids (which corresponds to at least one step of Forward Backtrack). The generation of all atom models from the coarse grained representation is a two step procedure. In a first step, the side chains are added using oscar-star (28). Then a fast minimization procedure using Gromacs 5 (29) is run to ensure the local backbone geometry is correct.

The last step of the protocol corresponds (i) to the identification of clusters and (ii) to the scoring of the conformations. For the modeling of peptides in their unbound conformation, the clustering does not rely on the RMSd any longer, but on the BCscore (30). We use a complete linkage clustering procedure using as distance $d = 1 - BCscore$, and a cut-off value equivalent to a BCscore value of 0.8, a value above which, in our observations, conformations can be considered as native. Clusters are then sorted either using sOPEP (31) or the Apollo scoring (32).

PEP-FOLD3 protocol for peptide–protein complex generation

A new feature of PEP-FOLD3 is the possibility to build peptide models in the vicinity of a protein receptor. Compared to the standard protocol, this requires some adaptations.

First of all, positions must be defined at the vicinity of the receptor from which the model generation will start. In PEP-FOLD3, some knowledge of a candidate binding patch on protein surface is required. The geometric center of the alpha carbon of the residues of the patch is used as the origin from which the starting coordinates are generated using the following procedure: (i) pseudo random directions are drawn using an iterative algorithm that draws directions on the surface of a unit sphere, but making sure that too close positions are avoided by the satisfaction of a minimal distance of 0.55Å. (ii) Scanning along these directions, we use a minimal distance of 4.76Å from the center of mass, which corresponds to a minimal distance between α-carbons of a peptide and a receptor. In order not to position the peptide too far from the receptor, a maximal distance of 7Å is used and a control that the starting points are not too close from each other is made, using a minimal distance between them of 6.Å. (iii) Finally, a last control is performed to ensure the starting points are not located inside the protein, by verifying that for some of the pseudo random directions no atom of the protein is at less than a given cut-off distance value of the starting point. This procedure usually results in only a limited number of starting positions. These positions are then assigned randomly to one amino acid of the peptide to start peptide growth.

A second modification is the possibility to make the peptide position and orientation vary relatively to the receptor during the Monte-Carlo procedure. This is achieved by two mechanisms: (i) we use a pivot residue from which peptide internal conformational changes are propagated. Periodic change in the pivot residue inherently results in changes in the relative orientation of the peptide relatively to the receptor. (ii) The Monte-Carlo includes explicit translation and rotation steps of the peptide as a rigid block. As a result, the RMSd between the initial peptide and final peptide coordinates can be as large as 15Å.

Finally, for complexes, the BCscore and Apollo scores are not relevant since they have been designed for peptides and proteins in isolation, and not for complexes. The clustering is performed using the RMSd between the peptide coordinates without superimposition, given the coordinates of the receptor are fixed during the simulations. We use a cut-off value of 5.5Å which corresponds to the size of the bassin of attraction observed for Rosetta FlexPepDock (19). As well, cluster ranking can only be performed using sOPEP.

IMPLEMENTATION

Input

Standard free modeling parameters. First, PEP-FOLD3 requires the sequence of the peptide to model. Since the PEP-FOLD3 protocol does not currently perform better than the previous one for peptides with disulfide bonds, only linear peptides are considered. The former PEP-FOLD

server is still available for disulfide bonded peptides. Optional parameters are: (i) the specification of a label that will be propagated to the file names of the models. (ii) The number of simulations—it corresponds to the number of sub-optimal trajectories that will be tested. For short peptides, a number of 100 is enough. For larger peptides 200 is recommended. (iii) The criterion to rank the models. It can be either the sOPEP energy value or the Apollo scoring (32), which should be preferred for the larger peptides.

Biased modeling. This section allows the input of a 3D model for the peptide (PDB format), and to specify which part of the model can be considered as acceptable. This is done by specifying a mask which consists in the exact same amino acid sequence as that of the model and that of the peptide, using lower case to identify regions of the model to propagate to the modeling. Note it is not the coordinates that are propagated to the model. Instead, we decode the 3D model using the Viterbi algorithm and bias the prior probabilities by setting to 1 the probabilities of the states corresponding to the residues in lower case. As a consequence, since the SA letters correspond to fragments of four amino acids, the first four amino acids cannot be distinguished and must have the same case (uppercase or lowercase).

Protein receptor. This section allows the specification of a protein receptor in PDB format. PEP-FOLD3 will reject receptor files with missing residues. Also, all ions, ligands, hetero groups will be discarded. Multiple chain receptors are not presently supported. Specifying a receptor, PEP-FOLD3 expects the identification of the residues expected to bind the peptide. Owing to the necessity to specify in a unique way the residues, each must be identified by its chain identifier, its name, its number and its insertion code, using underscore as delimiter, one residue per line. Note that for peptide–protein complexes, the number of simulations is doubled. One series is performed using standard parameters. The second one is performed after biasing the prior probabilities toward extended conformations, an outcome of our preliminary analyses. Finally, PEP-FOLD3 has in theory no hard limit in the size of the peptides bound to a receptor. The upper limit is thus, as for the modeling of peptides in isolation, of 50 amino acids. Note however, that benchmarking has been performed for a size of peptide varying between 5 and 15 amino acids.

Output

The result page of PEP-FOLD3 will first propose the interactive 3D visualization of the best models using the JavaScript viewer pv (http://biasmv.github.io/pv). A second section presents the results of the clustering in a hierarchical manner, and sorted according to either sOPEP or the Apollo score. A link to an archive containing all the models is proposed. Individual access to the PDB files of the representatives of the five best clusters is granted. Finally, a 2D representation of the prior probabilities is depicted using a three color code to facilitate its understanding: red for helical states, green for extended stated, blue otherwise.

Execution times. PEP-FOLD3 is particularly fast. Typical runs take usually on the order of few minutes to up to
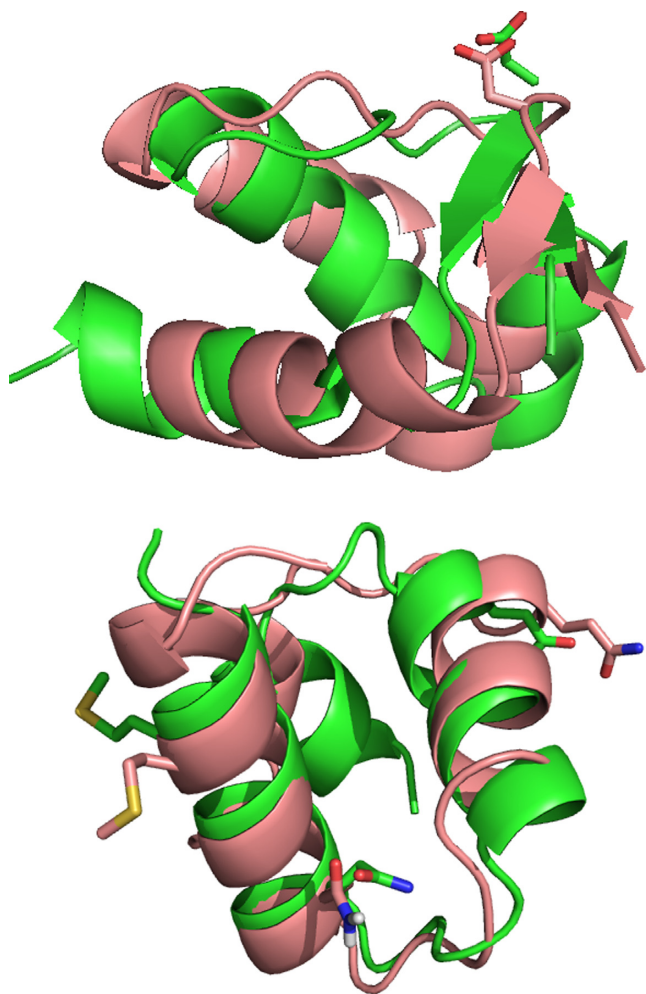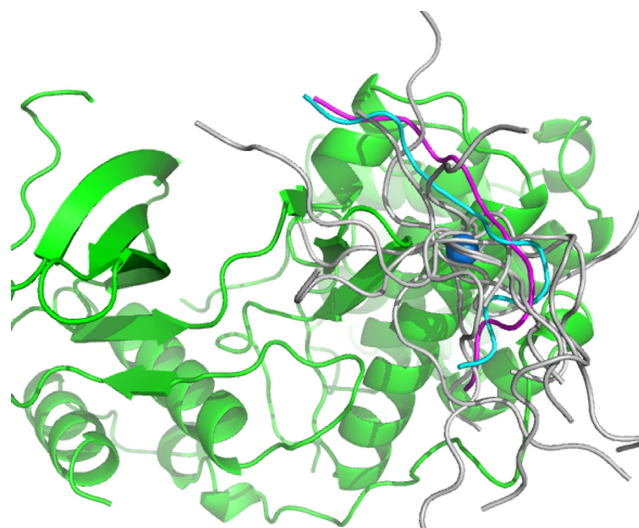
Figure 2. **Top:** example of the N-terminal Subdomain of Translation Initiation Factor IF2 (PDB ID: 1nd9). Green: experimental structure. Wheat: best model generated (BCscore 0.71). **Bottom:** example of the FAF-1 UBA Domain (PDB ID: 3e21). Green: experimental structure. Wheat: best model generated (BCscore 0.83). The details of some side chains are depicted.


Figure 3. Example of Fus3 in interaction with a fragment from Ste7left (PDB ID: 2B9H). Green: the unbound form of the protein. Blue: the sphere corresponds to one starting position. Grey: some initial peptide conformations before Monte-Carlo (RMS deviations up to 20Å). Cyan: the experimental pose. Magenta: one PEP-FOLD3 best pose (RMSd 3Å)

or native conformation in the top five best scored models for 80% of the targets. This performance is slightly better than existing methods such as Rosetta ([34]) which we found able to generate near-native or native models for 88% of the 56 targets. In some cases, users may have partial structural information which can be used to guide the simulation. In our experience, this limits the number of clusters and facilitates the identification of the native structure.

Figure 2 illustrates free modeling for the N-terminal Subdomain of Translation Initiation Factor IF2 (PDB ID: 1nd9), a peptide of 49 amino acids and for the FAF-1 UBA Domain of 40 amino acids (PDB ID: 3e21), illustrating that in few minutes, it is possible to have information about the structure of long peptides, providing a starting point for further investigations.

20 min depending on peptide size. For complexes, due to a larger number of simulations, typical run times are in the order of 20 min depending on peptide and receptor size. These indicative times can also vary depending on server load, but remain in most cases in the order of a few minutes to a few tens of minutes which we hope makes the service user friendly enough.

PERFORMANCE, EXAMPLES

Peptide-free or biased modeling

PEP-FOLD3 performance for linear peptides is identical to that of PEP-FOLD2, which has been benchmarked for a collection of 56 peptides from 25 to 52 amino acids ([33]). We remind that PEP-FOLD was designed for peptides in solution, at neutral pH, not bound to membrane or ligands, and not complexed to or stabilized by metal ions. PEP-FOLD3 is able to generate near-native or native models for 95% of the 56 targets (i.e. all but three), and to return a near-native

Peptide–protein complex

The performance of the PEP-FOLD3 server has been assessed on a collection of 61 peptide–protein complexes of the peptiDB database ([35]) using the unbound structure of the receptor. These 61 complexes correspond to all the peptides of the peptiDB collection for which the unbound conformation of the protein is available, is not a multimer, for which there are no missing residues and the peptide size is of more than four amino acids. Some entries for which the unbound conformation was found to be in interaction with another peptide have also been discarded. A full table of the results is available in Supplementary Table S1. We recall that the aim of PEP-FOLD3 here is not to perform the blind docking of the peptide on the entire protein surface but instead to give the means of a preliminary exploration of the possible modes of interaction between the peptide and a receptor, when the user has some idea of the patch of interaction. The tests have been performed starting from the patch defined as the list of the residues in contact with the
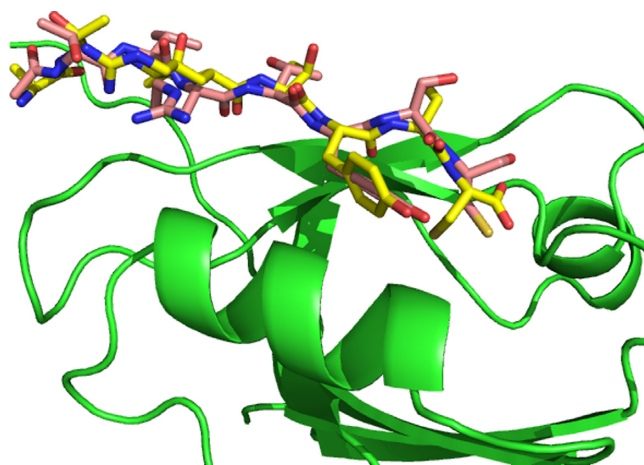
Figure 4. Example of PDZ domain of GRIP1 in complex with liprin C-terminal peptide. Green: the unbound form of the protein. Yellow: the experimental conformation of the peptide. Wheat: the best PEP-FOLD3 pose (RMSd 1.05Å)

peptide in the experimental complex. Note that the exact definition of the patch of interaction is however not crucial: as illustrated in Figure 3, the peptide original poses cover a large part around the starting positions. The initial conformations of the complete peptide before the Monte-Carlo refinement step can deviate by as much as 10Å and in some case by more than 20Å from the peptide in the experimental complex. We find that on average the best poses generated deviate by only 3.48Å from the experimental ones, a value similar to that obtained using the protein bound conformation, with a lowest value of 0.75Å for a peptide of 10 amino acids and a largest value of 6.99Å for a peptide of 15 amino acids. Among the 400 models, PEP-FOLD3 is able to generate a pose at <5.5Å (medium quality) for 90% of the targets. This fraction is of 52% for a distance of 3Å (high quality). Figure 4 shows an example of a pose at high quality. Considering the 10 best clusters, they contain a medium quality pose for 57% of the targets and high quality pose for 32% of the targets. Finally, the representatives of the 10 best clusters contain a medium quality pose for 48% of the cases, and high quality for only 15% of the cases. Compared to other servers such as CABS-dock, and using the common subset, the results by the two servers are very similar. It is worth to note, however, that in the present case the generation of the poses remains guided by the user which is not the case of the CABS-dock server. Using the bound conformation of the proteins does not change much the quality of the results, indicating that our current sampling procedure and the sOPEP force field can be improved for peptide–protein complexes.

CONCLUSIONS

PEP-FOLD3 provides a general framework for the structural characterization of peptides in solution or in interaction with a protein. For peptides in isolation, PEP-FOLD3 returns in a few minutes useful information in the five best models for 80% of the cases. For peptides in interaction with a receptor, PEP-FOLD3 10 best clusters will contain use-

ful information in 50% of the cases, although a more in depth analysis shows medium quality poses or better are generated for over 90% of the targets. Consequently, efforts are now on improving the scoring of peptide–protein complexes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors thank S. de Vries for a careful rereading of the manuscript.

REFERENCES

1. Fosgerau,K. and Hoffmann,T. (2015) Peptide therapeutics: current status and future directions. *Drug Discov. Today*, **20**, 122–128.
2. Rey,J., Deschavanne,P. and Tuffery,P. (2014) Bactpepdb: a database of predicted peptides from a exhaustive survey of complete prokaryote genomes. *Database*, **2014**, bau106.
3. Vetter,I., Davis,J.L., Rash,L.D., Anangi,R., Mobli,M., Alewood,P.F., Lewis,R.J. and King,G.F. (2011) Venomics: a new paradigm for natural products-based drug discovery. *Amino Acids*, **40**, 15–28.
4. MacLaine,N.J. and Hupp,T.R. (2011) How phosphorylation controls p53. *Cell Cycle*, **10**, 916–921.
5. London,N., Raveh,B. and Schueler-Furman,O. (2013) Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr. Opin. Struct. Biol.*, **23**, 894–902.
6. Whitehead,T.A., Baker,D. and Fleishman,S.J. (2013) Computational design of novel protein binders and experimental affinity maturation. *Methods Enzymol.*, **523**, 1–19.
7. Yin,H. and Flynn,A. (2016) Drugging membrane protein interactions. *Annu. Rev. Biomed. Eng.*, **18**, 51–76.
8. Wichapong,K., Alard,J.-E., Ortega-Gomez,A., Weber,C., Hackeng,T.M., Soehnlein,O. and Nicolaes,G.A. (2016) Structure-based design of peptidic inhibitors of the interaction between CC chemokine ligand 5 (CCL5) and human neutrophil peptides 1 (HNP1). *J. Med. Chem.*, doi:10.1021/acs.jmedchem.5b01952.
9. Donsky,E. and Wolfson,H.J. (2011) Pepcrawler: a fast rrt-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics*, **27**, 2836–2842.
10. Vanhee,P., van derSloot,A.M., Verschueren,E., Serrano,L., Rousseau,F. and Schymkowitz,J. (2011) Computational design of peptide ligands. *Trends Biotechnol.*, **29**, 231–239.
11. Duffy,F.J., Verniere,M., Devocelle,M., Bernard,E., Shields,D.C. and Chubb,A.J. (2011) Cyclops: generating virtual libraries of cyclized and constrained peptides including nonnatural amino acids. *J. Chem. Inf. Model.*, **51**, 829–836.
12. Singh,S., Singh,H., Tuknait,A., Chaudhary,K., Singh,B., Kumaran,S. and Raghava,G.P. (2015) Pepstrmod: structure prediction of peptides containing natural, non-natural and modified residues. *Biol. Direct*, **10**, 1–19.
13. Maupetit,J., Derreumaux,P. and Tuffery,P. (2009) Pep-fold: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.*, **37**, W498–W503.

14. Thévenet,P., Shen,Y., Maupetit,J., Guyon,F., Derreumaux,P. and Tufféry,P. (2012) Pep-fold: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.*, **40**, W288–W293.

15. Beaufays,J., Lins,L., Thomas,A. and Brasseur,R. (2012) In silico predictions of 3d structures of linear and cyclic peptides with natural and non-proteinogenic residues. *J. Pept. Sci.*, **18**, 17–24.

16. Xu,D. and Zhang,Y. (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*, **80**, 1715–1735.

17. Trabuco,L.G., Lise,S., Petsalaki,E. and Russell,R.B. (2012) Pepsite: prediction of peptide-binding sites from protein surfaces. *Nucleic Acids Res.*, **40**, W423–W427.

18. Saladin,A., Rey,J., Thévenet,P., Zacharias,M., Moroy,G. and Tufféry,P. (2014) Pep-sitefinder: a tool for the blind identification of peptide binding sites on protein surfaces. *Nucleic Acids Res.*, **42**, W221–W226.

19. London,N., Raveh,B., Cohen,E., Fathi,G. and Schueler-Furman,O. (2011) Rosetta flexpepdock web server–high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.*, **39**(Suppl. 2), W249–W253.

20. Lee,H., Heo,L., Lee,M.S. and Seok,C. (2015) Galaxypepdock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Research*, **43**, W431–W435.

21. Kurcinski,M., Jamroz,M., Blaszczyk,M., Kolinski,A. and Kmiecik,S. (2015) Cabs-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Research*, **43**, W419–W424.

22. Schindler,C.E., deVries,S.J. and Zacharias,M. (2015) Fully blind peptide-protein docking with pepattract. *Structure*, **23**, 1507–1515.

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

24. Sterpone,F., Melchionna,S., Tuffery,P., Pasquali,S., Mousseau,N., Cragnolini,T., Chebaro,Y., St-Pierre,J.-F., Kalimeri,M., Barducci,A. *et al.* (2014) The opep protein model: from single molecules, amyloid formation, crowding and hydrodynamics to dna/rna systems. *Chem. Soc. Rev.*, **43**, 4871–4893.

25. Camproux,A.-C., Gautier,R. and Tuffery,P. (2004) A hidden markov model derived structural alphabet for proteins. *J. Mol. Biol.*, **339**, 591–605.

26. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

27. Maupetit,J., Derreumaux,P. and Tufféry,P. (2010) A fast method for large-scale de novo peptide and miniprotein structure prediction. *J. Comput. Chem.*, **31**, 726–738.

28. Liang,S., Zheng,D., Zhang,C. and Standley,D.M. (2011) Fast and accurate prediction of protein side-chain conformations. *Bioinformatics*, **27**, 2913–2914.

29. Abraham,M.J., Murtola,T., Schulz,R., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19–25.

30. Guyon,F. and Tufféry,P. (2014) Fast protein fragment similarity scoring using a binet-cauchy kernel. *Bioinformatics*, **30**, 784–791.

31. Maupetit,J., Tuffery,P. and Derreumaux,P. (2007) A coarse-grained protein force field for folding and structure prediction. *Proteins*, **69**, 394–408.

32. Wang,Z., Eickholt,J. and Cheng,J. (2011) Apollo: a quality assessment service for single and multiple protein models. *Bioinformatics*, **27**, 1715–1716.

33. Shen,Y., Maupetit,J., Derreumaux,P. and Tufféry,P. (2014) Improved pep-fold approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.*, **10**, 4745–4758.

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

35. London,N., Movshovitz-Attias,D. and Schueler-Furman,O. (2010) The structural basis of peptide-protein binding strategies. *Structure*, **18**, 188–199.