W498–W503 *Nucleic Acids Research, 2009, Vol. 37, Web Server issue* *Published online 11 May 2009*
doi:10.1093/nar/gkp323

PEP-FOLD: an online resource for *de novo* peptide structure prediction

Julien Maupetit[1,2], Philippe Derreumaux[3] and Pierre Tuffery[1,2,*]

[1]MTi, INSERM UMR-S 973, - Paris 7, 35 rue H. Brion, F75205, Paris, [2]RPBS, Université Paris Diderot - Paris 7, 35 rue H. Brion, F75205, Paris and [3]Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F75005, Paris, France

Received January 30, 2009; Revised April 5, 2009; Accepted April 20, 2009

ABSTRACT

Rational peptide design and large-scale prediction of peptide structure from sequence remain a challenge for chemical biologists. We present PEP-FOLD, an online service, aimed at *de novo* modelling of 3D conformations for peptides between 9 and 25 amino acids in aqueous solution. Using a hidden Markov model-derived structural alphabet (SA) of 27 four-residue letters, PEP-FOLD first predicts the SA letter profiles from the amino acid sequence and then assembles the predicted fragments by a greedy procedure driven by a modified version of the OPEP coarse-grained force field. Starting from an amino acid sequence, PEP-FOLD performs series of 50 simulations and returns the most representative conformations identified in terms of energy and population. Using a benchmark of 25 peptides with 9–23 amino acids, and considering the reproducibility of the runs, we find that, on average, PEP-FOLD locates lowest energy conformations differing by 2.6 Å Cα root mean square deviation from the full NMR structures. PEP-FOLD can be accessed at http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD

INTRODUCTION

While computational biologists and chemists have developed convincing *ab initio* or *de novo* approaches for protein structure determination [see CASP experiments (1)], accurate and fast peptide 3D conformation prediction from bioinformatics-based or physics-based methods is still an open challenge. Yet, peptides play many biological functions ranging from hormones, neurotransmitters to antibiotics, among others (2,3).

The obstacles for elucidating the relationship between peptide sequences, structures and functions arise from several factors. First, in contrast to proteins, short peptides do not systematically adopt a stable well-defined tertiary structures (4). Then, the experimental flow of peptide structure determination based on Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography remains very low, and structures can also be determined by using non-physiological conditions to increase their solubility such as trifluoroethanol (TFE)/water mixtures. In addition, many peptides are marginally stable, and early peptide structure determinations may be questioned. For instance, the 20-residue betanova peptide was initially proposed as adopting a three-stranded β-structure in aqueous solution (5), but its content of β-sheet was revisited and reduced by long molecular dynamics (MD) simulations and new experiments to 10% (6). Finally, on the theoretical front, in spite of constant effort, large-scale *in silico* prediction using all-atom models is still not feasible. The application of the Rosetta approach to peptides remains to be evaluated (7), and all-atom molecular dynamics or Monte Carlo simulations in explicit or implicit solvent models with replica exchanges are hampered by the computer time and resources (8,9).

To accelerate conformational search and reduce the number of local minima, many studies have resorted to simplified representations and energy models, but the main problem is to preserve the physics of the systems (10). In this context, Ichikawa *et al.* (11) proposed a growing chain algorithm using a number of discrete dihedral ϕ/ψ choices and a sum of hydrophobic and hydrogen bond interactions. Raghava *et al.* (12) designed PepStr based on secondary structure and β-turn prediction with a short MD-based energy refinement. For their part, Thomas *et al.* (13) developed Peplook based on a Boltzmann-stochastic algorithm coupled to 64 ϕ/ψ backbone combinations, but their method failed to predict all topologies. More recently, Nicosia and Stracquadanio (14) proposed a generalized pattern search algorithm (GPS) using an all-atom energy model and secondary structure prediction (for systems with more 15 amino acids). Using a benchmark of 42 peptides with 9–20 amino acids in aqueous and non-aqueous solutions, GPS approached the experimental structures at 3.2 Å Cα root mean square deviation

*To whom correspondence should be addressed. Tel: +331 5727 8374; Fax: +331 5727 8372; Email: pierre.tuffery@univ-paris-diderot.fr

(cRMSD) (14). Currently, the best server available to biologists for peptide structure prediction is the PepStr server reaching an averaged cRMSD accuracy of 4.0 Å on a set of 42 peptides (12).

Here we present, the PEP-FOLD server, which builds on a new *de novo* approach to predict 3D peptide structures from sequence information. PEP-FOLD is based on the concept of structural alphabet (SA) and uses a Hidden Markov Model (HMM)-derived SA of 27 letters to describe proteins as series of overlapping fragments of four amino acids (15). PEP-FOLD uses a two-step procedure: prediction of a limited set of SA letters at each position from sequence, and then assembly of the prototype fragments associated with each SA letter using a revised version of our greedy algorithm (16,17) and a generic protein coarse-grained force field (18).

PEP-FOLD is very fast and can be used with high confidence for any natural peptide sequence of 9–25 amino acids in aqueous solution, and, with more caution, in non-aqueous solutions. All details about the algorithm, force field and the results can be found elsewhere (Maupetit,J. *et al.*, submitted for publication).

CONCEPTS AND METHODS

Figure 1 presents a general overview of PEP-FOLD.

SA profile prediction from amino acid sequence

PEP-FOLD uses a HMM-derived SA of 27 letters, describing proteins as series of overlapping fragments of four amino acids (15). The conformations of consecutive fragments are not independent and follow a first-order Markovian process. An SA letter at a position cannot therefore be followed by all 27 SA letters at the adjacent position. A protein of L amino acids is represented by a SA sequence of $L-3$ SA letters. To learn SA letter profiles from the amino acid sequence, PEP-FOLD uses the following two-step procedure.

In the first step, the SA letter profiles are predicted at each independent position, using an amino acid profile obtained by running PSI-BLAST (19) against the uniref data-bank (20). As support vector machine (SVM) input, we use the amino acid profile of each four-residue fragment enlarged by two residues on both sides, and therefore a 20×8 dimensional vector to predict one SA letter. The SVM output is a profile of $L-3 \times 27$ probabilities that gives the predicted probability that each SA letter fits each fragment of the protein. We emphasize that this prediction step is free of any bias. The SVM training was performed by using a collection of structures excluding our peptide set.

In the second step, these probabilities are used as input data of a *forward-backward* algorithm along with the HMM-SA model. We have observed that the prediction of only the eight best letters at each position contains the native conformation information. Thus, the output of the SA prediction is a profile of $L-3 \times 8$ SA letters.

Note that the PEP-FOLD server also allows the user to constrain the predicted SA profile by PSIPRED secondary structure prediction (21). This is done, for the amino acid positions predicted by PSIPRED as helices (or β-strands) with a confidence index >5, by restraining their possible SA letters to helical (or β-strand) letters. Although we found that the PSIPRED option has a marginal impact in most cases, it can sometimes help identify the native conformation.

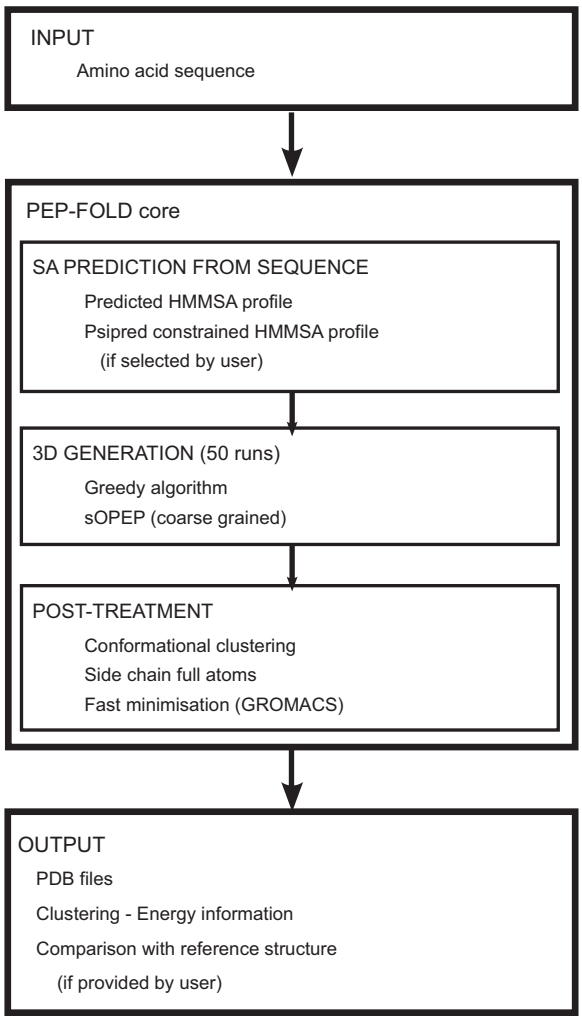

Figure 1. PEP-FOLD flowchart.

3D assembly of prototype conformations

To build 3D structures from the predicted SA letter profiles, we perform a rigid assembly of the predicted fragments by using an enhanced version of our greedy algorithm (16,17). Whereas the early algorithm uses forward (from N- to C-terminal) and backward (from C- to N-terminal) incremental operators to grow the chain, our revisited version uses a zip operator to start the building process at any position of the structure, alternatively adding S residues at each side of the growing structure. After extensive tests on 9–25 residue systems, we found that an S value of 1 is appropriate, the starting position can be selected randomly, and very high accuracy can be achieved by using a total of 74 prototypes to describe the 27 SA letters. At greedy completion, the best model obtained is refined by a Monte Carlo simulation of

300 000 steps, using a random selection of one prototype at each step and $T = 300$ K in the Metropolis acceptance criterion.

In contrast to our previous studies where the assembly was guided by an energy function dependent on the knowledge of the native long-range Cα–Cα contacts (16,17), PEP-FOLD assembly is driven by the generic protein coarse-grained force field, OPEP 3.1 version (18), slightly adapted for greedy algorithm. As many force fields, OPEP includes local interactions for stereochemistry (e.g. bond lengths, bond angles, etc.), van der Waals interactions between all particles and terms allowing the formation of hydrogen bonds. OPEP uses the positions of the backbone N, H, Cα, C and O atoms and one bead for each side chain, and is designed for basin hopin (10) and molecular dynamics (22) simulations in Cartesian coordinate space. Since the greedy algorithm works in a discrete space and uses well-defined fragments of four-residues, we derived sOPEP. The main difference between OPEP and sOPEP comes from the formulation of the side chain—side chain interactions that result in smoother interactions at short distances, reducing therefore the number of steric clashes during rigid assembly.

3D models post-treatment

Each polypeptide is subject to one run, i.e. 50 greedy simulations differing in the starting position (see previous section) and randomness in the retained candidates at each position. The conformational diversity of the 50 models is characterized by cluster analysis using a complete linkage procedure and a threshold of 2 Å cRMSD for peptides less than 20 amino acids and 3 Å above. For each cluster, we return the lowest energy centroid conformation.

The Greedy-OPEP assembling procedure generates coarse-grained structure models. These models are post-treated in two ways. First, the side chain beads are discarded and all-atom side-chains are positioned using a fast backbone-dependent procedure previously used in the SABBAC server (23). Secondly, to provide high quality structures, a fast minimization is performed with GROMACS (24).

INPUT/OUTPUT

PEP-FOLD accepts, as input, sequences of lengths between 9 and 25 amino acids. Most of the frequent sequence formats including the FASTA or RAW formats are accepted. Only standard amino acids are presently accepted, and no hetero groups—including ions—are considered. Disulfide bonds are not managed. It is possible to constrain the SA local structure prediction using PSIPRED, although this is not active by default. Finally, for the sake of convenience, the user can specify a reference structure for comparison with the *de novo* models.

On program termination, the server gives the results of the cluster analysis. Although the number of clusters, Ncs, can vary from 1 (all conformations similar) to 50, Ncs is usually less than 5. The output first returns a 2D plot with the population and energy of each cluster and the localization of the lowest energy (Figure 2A), and then a picture

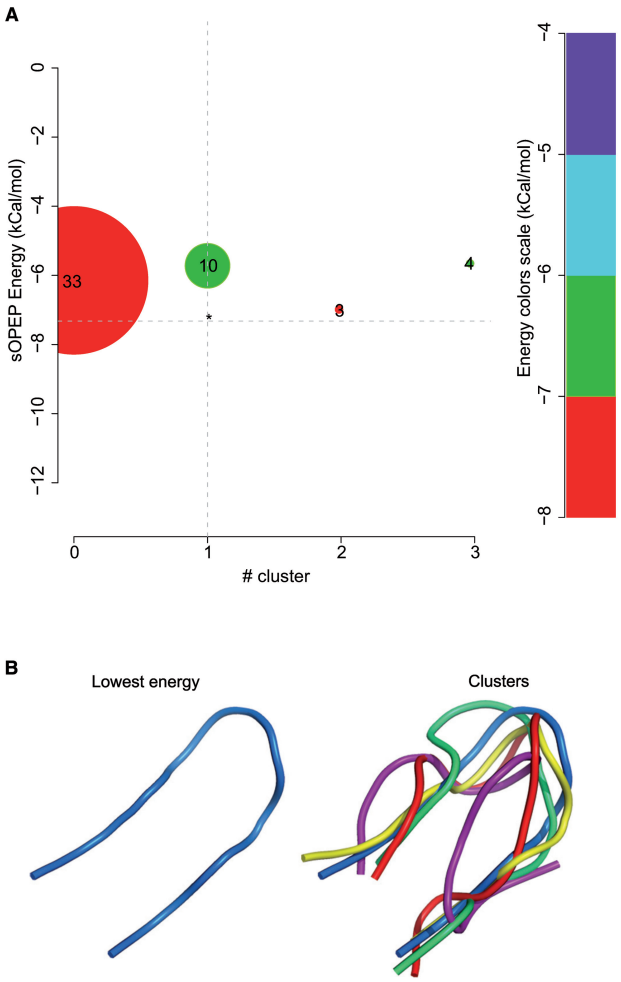

Figure 2. PEP-FOLD sample output for the peptide 1UAO. A 2D energy plot. For each cluster (*x*-axis), we show the energy of its centroid (*y*-axis). The size of the circle is proportional to the cluster population which is also given. The dashed lines help locate the LEC. (**B**) Left: LEC. Right: the representatives of the best clusters are superimposed on the LEC to illustrate conformational variability. Pictures are obtained using the PyMol software.

of the centroid of each cluster (Figure 2B). Subsequently, the output gives for each of the 50 conformations its cluster assignment and energy. If a reference structure was specified, the cRMSD, the GDT-TS (25) and TM scores (26) of each model are also reported. Finally, an archive containing all cluster centroid conformations and the lowest energy conformation (LEC) (PDB format) is available for download. The LEC is also available separately.

Since PEP-FOLD is based on a chain growth algorithm, the CPU times vary linearly with peptide length. Typical execution times, using 10 Intel Xeon 3GHz processing units in parallel, vary for one run from 10 min for a 10-residue peptide to 45 min for a 25-residue peptide.

PERFORMANCES

In a previous report, the PEP-FOLD performances were assessed using a benchmark of 25 peptides with 9–25

Table 1. PEP-FOLD results on 25 peptides with 9–25 amino acids in aqueous solution

PDB	*L*	Class	LEC energy	Max ΔE	Native cRMSD	LEC cRMSD
1a13	14	–	−26.5	0.0	1.8	0.0
1b03	18	B	−23.6	1.4	2.0–3.1	2.1
1dep	15	A	−30.7	0.0	1.7	0.0
1du1	20	A	−45.1	0.0	5.1	0.0
1e0q	17	B	−26.7	0.2	2.7–5.7	2.2
1egs	9	–	−3.9	0.0	1.5	0.0
1gjf	14	A	−24.7	0.0	2.4–2.5	0.1
1in3	12	A	−25.8	0.0	2.3	0.0
1k43	14	B	−20.9	0.4	1.5–1.7	0.8
1l2y	20	A	−28.5	0.1	2.1–2.3	0.2
1l3q	12	–	−6.1	0.7	3.3–5.3	2.5
1lcx	13	–	−28.0	0.3	2.8	0.0
1le1	12	B	−18.2	0.0	1.0	0.0
1le3	16	B	−24.2	3.9	1.2–2.7	1.1
1niz	14	B	−19.1	0.7	2.1–2.7	0.7
1nkf	16	A	−18.9	0.0	4.3–4.5	0.3
1pef	18	A	−58.2	0.0	0.9	0.0
1pei	22	A	−50.3	0.0	1.6	0.0
1pgbF	16	B	−21.6	2.7	2.1–2.4	2.4
1rpv	17	A	−28.9	0.0	0.6	0.1
1uao	10	B	−7.3	0.1	2.0	0.1
1wbr	17	–	−33.9	0.0	3.5	0.0
1wz4	23	A	−31.2	0.6	5.6–6.4	2.8
2bta	15	–	−23.5	0.0	4.5	0.0
2evq	12	B	−16.9	0.0	0.9	0.0

PDB: Protein Data Bank identifier. *L*: peptide length. Class: structural class (A for α, B for β and - for none), according to STRIDE (29) program. Each target is subject to five independent runs, each of 50 greedy simulations. For each peptide, we report, the LEC (kcal/mole) from run 1 and the maximal energy difference (Max ΔE) between the five LECs from runs 1–5. We also give the minimal and maximal cRMSD (in angstroms) of the five LECs with respect to the NMR structure (native cRMSD). Finally, we report the average cRMSD between the five LECs (LEC cRMSD). Note that 1pgbF corresponds to the fragment 41–56 of 1pgb, and the NMR structure used for cRMSD calculation is model 1, except for 1a13, 1b03, 1e0q, 1egs, 1lcx, 1pei, 1rpv, 1wbr and 1wz4 where the reference models are 10, 4, 8, 3, 25, 10, 12, 23 and 2, respectively.

amino acids in aqueous solution. Since PEP-FOLD is stochastic in character, we do not expect the same ensemble of higher energy conformations from one run to another, each consisting of 50 greedy simulations. The Ncs, their sizes and their centres vary, as should be expected.

Table 1 summarizes the performances without any PSIPRED constraints, using five independent runs. It reports for each peptide the LEC obtained from run 1, the maximal energy difference between the five LECs, the extremal cRMSD between the five LECs and the NMR structure, and the average cRMSD between the five LECs. Averaged on 25 peptides and five runs, the PEP-FOLD LEC reproduces the NMR structure at 2.6 Å cRMSD. Considering a strict criterion to accept LEC identity: the variation in energy <1 kcal/mole and variation in cRMSD <1.2 Å, we find that for 20 peptides, the LEC remains the same between the 5 runs, and the average cRMSD between the LECs is of 0.6 Å. For 1le3, 1b03 1pgbF, the energy differences between the LECs do not change the predicted topology. The maximal energy difference of 1.4 kcal/mole with a cRMSD difference of 1.1 Å leads, however, for 1b03, to a β-hairpin with out-of-register strands. In contrast, for 1l3q and 1e0q, the NMR topology is not identified systematically: for 1e0q, a LEC of −27 kcal/mole displays the NMR β-hairpin topology, but PEP-FOLD frequently returns a LEC with helix-turn-strand character.

In summary, using our strict criteria for reproducibility, the probability to locate the LEC by using one run of 50 greedy simulations for a peptide with 9–25 amino acids is 86%. The probability reaches 96% in terms of convergence towards the same topology.

We present in Figure 3 the results on the previously studied 10-residue peptide chignolin (PDB code: 1UAO) (27), and a new system, the 23-residue magainin antibiotic peptide (PDB code: 2MAG) (3).

For the chignolin, the 18 NMR models are very similar and display a β-hairpin. PEP-FOLD generates four distinct conformations. The lowest energy model displays a fully native structure (cRMSD of 2.0 Å) with the experimental β-strands spanning residues 2–3/8–9 correctly predicted. The β-carbon RMSD is 1.0 Å from the NMR structure, indicating correct orientation of the side chains relative to the strands as seen in Figure 3.

The magainin antibiotic peptide is soluble, but unstructured in aqueous solution. Its NMR structure in dipalmitoylphosphatidylcholin (DPC), sodium dodecyl sulfate (SDS) and TFE/water environments is characterized by a long α-helix that is kinked at residues Phe12-Gly13 (3). The conformation of the first cluster identified by PEP-FOLD is very similar to the NMR structure in non-aqueous solution (0.7 Å cRMSD, see the green and magenta structures in Figure 3) and correctly predicts the helix kink. The second cluster of lowest energy energy, only 2 kcal/mole more stable than the first cluster, displays two α-helices spanning 1–12 and 17–22 stabilized by an hydrophobic patch between the three phenylalanines (Phe5, Phe12 and Phe16). Compared to the

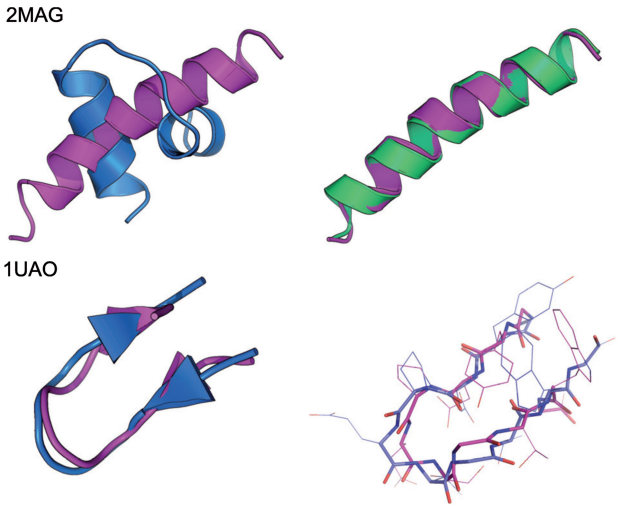

Figure 3. PEP-FOLD server-generated models. The lowest energy model (blue) is superposed on the experimental structure of 1UAO in aqueous solution and 2MAG in TFE/water (magenta) using iSuperpose online facility (http://mobyle.rpbs.univ-paris-diderot.fr/). For 2MAG, we also show the lowest RMSD structure (green) with respect to reference state. For 1UAO, we also show the all-atom predicted and experimental structures.

magainin NMR structure in aqueous solution, it is clear that PEP-FOLD can sometimes generate over-structured conformations. We note, however, that the transition of magainin from random coil to α-helical occurs at very low concentration of TFE (28), indicating that our predicted conformations exist in aqueous solution, albeit with a lower probability.

DISCUSSION AND FUTURE WORK

Compared to the PepStr and GPS methods, PEP-FOLD is a more accurate approach leading to a cRMSD of 2.6 Å between the predicted, LECs and the full NMR structures using a benchmark of 25 peptides. It is, however, important to note that PEP-FOLD can sometimes fail to recognize the experimental structure in aqueous solution. This is illustrated here on the 23-residue magainin antibiotic peptide, where PEP-FOLD predicts the conformations observed in sodium dodecylsulfate micelles and trifluoroethanol/water solutions rather than the unstructured conformations as observed in aqueous solution. It is possible that in these few cases, the conformational entropy, which is neglected in the present treatment, modifies entirely the free energy profile. In this context, short replica-exchange molecular dynamics simulations of 30 ns with the OPEP force field starting from one PEP-FOLD solution might be valuable (9).

In its present form, PEP-FOLD is parametrized for aqueous solution and does not offer the possibility to tune the force field parameters according to the experimental conditions. These include: pH variation in aqueous solution and non-aqueous environments such as TFE/water and SDS micelles. Although, we found that PEP-FOLD generated very reasonable structures for a set of 27 peptides in non-aqueous solutions (2.7 Å cRMSD) (results not shown), we are currently optimizing sOPEP parameters for these environments. Another direction we are exploring is the possibility to treat linear and cyclic peptides combining both D- and L-amino acids, and peptides with disulfide bridges. Overall, we believe that PEP-FOLD, with its present accuracy and speed, should be a useful server for chemical biologists.

FUNDING

Funding for open access charge: INSERM UMR-S 973.

Conflict of interest statement. None declared.

REFERENCES

1. Moult,J., Fidelis,K., Kryshtafovych,A., Rost,B., Hubbard,T. and Tramontano,A. (2007) Critical assessment of methods of protein structure prediction-round vii. *Proteins*, **69 (Suppl. 8)**, 3–9.
2. Liu,F., Baggerman,G., Schoofs,L. and Wets,G. (2008) The construction of a bioactive peptide database in metazoa. *J. Proteome Res.*, **7**, 4119–4131.
3. Gesell,J., Zasloff,M. and Opella,S.J. (1997) Two-dimensional 1h NMR experiments show that the 23-residue magainin antibiotic peptide is an alpha-helix in dodecylphosphocholine micelles sodium dodecylsulfate micelles, and trifluoroethanol/water solution. *J. Biomol. NMR*, **9**, 127–135.
4. Gellman,S.H. and Woolfson,D.N. (2002) Mini-proteins trp the light fantastic. *Nat. Struct. Biol.*, **9**, 408–410.
5. Kortemme,T., Ramírez-Alvarado,M. and Serrano,L. (1998) Design of a 20-amino acid, three-stranded beta-sheet protein. *Science*, **281**, 253–256.
6. Colombo,G., Roccatano,D. and Mark,A.E. (2002) Folding and stability of the three-stranded beta-sheet peptide betanova: insights from molecular dynamics simulations. *Proteins*, **46**, 380–392.
7. Bradley,P., Chivian,D., Meiler,J., Misura,K.M.S., Rohl,C.A., Schief,W.R., Wedemeyer,W.J., Schueler-Furman,O., Murphy,P., Schonbrun,J. *et al.* (2003) Rosetta predictions in casp5: successes, failures, and prospects for complete automation. *Proteins*, **53 (Suppl 6)**, 457–468.
8. Jang,S., Shin,S. and Pak,Y. (2002) Molecular dynamics study of peptides in implicit water: ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif. *J. Am. Chem. Soc.*, **124**, 4976–4977.
9. Chebaro,Y., Dong,X., Laghaei,R., Derreumaux,P. and Mousseau,N. (2009) Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J. Phys. Chem. B.*, **113**, 267–274.
10. Derreumaux,P. (1997) A diffusion process-controlled Monte Carlo method for finding the global energy minimum of a polypeptide chain. i. formulation and test on a hexadecapeptide. *J. Chem. Phys.*, **106**, 5260–5270.
11. Ishikawa,K., Yue,K. and Dill,K.A. (1999) Predicting the structures of 18 peptides using geocore. *Protein Sci.*, **8**, 716–721.
12. Kaur,H., Garg,A. and Raghava,G.P.S. (2007) Pepstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept. Lett.*, **14**, 626–631.
13. Thomas,A., Deshayes,S., Decaffmeyer,M., Van Eyck,M.H., Charloteaux,B. and Brasseur,R. (2006) Prediction of peptide structure: how far are we? *Proteins*, **65**, 889–897.
14. Nicosia,G. and Stracquadanio,G. (2008) Generalized pattern search algorithm for peptide structure prediction. *Biophys. J.*, **95**, 4988–4999.
15. Camproux,A.C., Gautier,R. and Tufféry,P. (2004) A hidden Markov model derived structural alphabet for proteins. *J. Mol. Biol.*, **339**, 591–605.
16. Tuffery,P., Guyon,F. and Derreumaux,P. (2005) Improved greedy algorithm for protein structure reconstruction. *J. Comput. Chem.*, **26**, 506–513.

17. Tuffery,P. and Derreumaux,P. (2005) Dependency between consecutive local conformations helps assemble protein structures from secondary structures using go potential and greedy algorithm. *Proteins*, **61**, 732–740.
18. Maupetit,J., Tuffery,P. and Derreumaux,P. (2007) A coarse-grained protein force field for folding and structure prediction. *Proteins*, **69**, 394–408.
19. 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.
20. Suzek,B.E., Huang,H., McGarvey,P., Mazumder,R. and Wu,C.H. (2007) Uniref: comprehensive and non-redundant uniprot reference clusters. *Bioinformatics*, **23**, 1282–1288.
21. Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
22. Derreumaux,P. and Mousseau,N. (2007) Coarse-grained protein molecular dynamics simulations. *J. Chem. Phys.*, **126**, 025101.
23. Maupetit,J., Gautier,R. and Tufféry,P. (2006) Sabbac: online structural alphabet-based protein backbone reconstruction from alpha-carbon trace. *Nucleic Acids Res.*, **34**, W147–W151.
24. Van Der Spoel,D., Lindahl,E., Hess,B., Groenhof,G., Mark,A.E. and Berendsen,H.J.C. (2005) Gromacs: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–1718.
25. Zemla,A. (2003) Lga: a method for finding 3d similarities in protein structures. *Nucleic Acids Res.*, **31**, 3370–3374.
26. Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.
27. Honda,S., Yamasaki,K., Sawada,Y. and Morii,H. (2004) 10 residue folded peptide designed by segment statistics. *Structure*, **12**, 1507–1518.
28. Marion,D., Zasloff,M. and Bax,A. (1988) A two-dimensional NMR study of the antimicrobial peptide magainin 2. *FEBS Lett.*, **227**, 21–26.
29. Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.