Research Article

Journal of PeptideScience

Received: 30 March 2011 Revised: 14 June 2011 Accepted: 1 July 2011 Published online in Wiley Online Library: 27 October 2011

(wileyonlinelibrary.com) DOI 10.1002/psc.1410

# In silico predictions of 3D structures of linear and cyclic peptides with natural and non-proteinogenic residues

**Jérôme Beaufays, Laurence Lins,* Annick Thomas and Robert Brasseur**

**We extended the use of Peplook, an in silico procedure for the prediction of three-dimensional (3D) models of linear peptides to the prediction of 3D models of cyclic peptides and thanks to the *ab initio* calculation procedure, to the calculation of peptides with non-proteinogenic amino acids. Indeed, such peptides cannot be predicted by homology or threading. We compare the calculated models with NMR and X-ray models and for the cyclic peptides, with models predicted by other in silico procedures (Pep-Fold and I-Tasser). For cyclic peptides, on a set of 38 peptides, average root mean square deviation of backbone atoms (BB-RMSD) was 3.8 and 4.1 Å for Peplook and Pep-Fold, respectively. The best results are obtained with I-Tasser (2.5 Å) although evaluations were biased by the fact that the resolved Protein Data Bank models could be used as template by the server. Peplook and Pep-Fold give similar results, better for short (up to 20 residues) than for longer peptides. For peptides with non-proteinogenic residues, performances of Peplook are sound with an average BB-RMSD of 3.6 Å for 'non-natural peptides' and 3.4 Å for peptides combining non-proteinogenic residues and cyclic structure. These results open interesting possibilities for the design of peptidic drugs. Copyright © 2011 European Peptide Society and John Wiley & Sons, Ltd.**

*Supporting information may be found in the online version of this article*

**Keywords:** cyclic peptides; non-natural peptides; structure prediction; Boltzmann-Stochastic; modeling

## Introduction

Determination of protein and peptide structures is an important field to understand their functions. Physicochemical approaches such as X-ray crystallography or NMR spectroscopy are used to determine three-dimensional (3D) structures. However, these techniques show some limits because of long time processing and/or important costs. In the meantime, computational methods have been developed to calculate 3D models from sequences. They can be classified into three main categories: homology [1–3], threading [4–6], and *ab initio* [7–9].

In the first category, the structure is calculated based on a resolved structure from the Protein Data Bank (PDB) whose sequence is homologous to the target sequence. Threading is based on the limited number of protein folds with respect to the huge possibilities of sequences. This method tries to identify which resolved proteins will share its fold with the target sequence. The *ab initio* method does not require a template to build models. It is only based on physicochemical principles and aims to predict folds of lower free energy.

For the prediction of peptide structures, few servers exist such as Pep-Fold [10,11], Pepstr [12], Protinfo [13,14], Hmmstr/Rosetta [15,16], I-Tasser [17,18]. . .

Recently, our laboratory developed Peplook [19], an *ab initio* method. Peplook is an iterative Boltzmann-Stochastic algorithm to predict 3D models from sequences up to 30 residues. The program uses Φ/Ψ angles derived from the structural alphabet of Etchebest [20]. It generates series of random peptide conformations and selects the lowest energy model.

In this paper, we present two new applications of Peplook. In the first application, we show that distance restraints can be used to enable the prediction of cyclic peptide structures. Calculated Peplook models were compared with experimental data and with models obtained by two other structure prediction servers, Pep-Fold [10,11] and I-Tasser [17,18].

Pep-Fold uses a hidden Markov model-derived structural alphabet of 27 motifs composed of 4 residues. It first determines structural alphabet letters of the sequence and then builds model by assembling the fragments using a greedy algorithm driven by a coarse-grained force field. Pep-Fold can be used for peptides of 9–25 amino acids. I-Tasser combines threading and *ab initio* methods. The sequence is first threaded through a PDB structure library. Fragments are then assembled to build a global structure and unaligned regions are generated by an *ab initio* approach.

In the second application, Peplook is used to predict 3D models of peptides containing non-proteinogenic amino acids such as the posttranslationally modified (phosphorylated, sulfated, hydroxylated and carboxylated) amino acids, D-amino acids, natural amino acids of non ribosomal peptides and non-natural synthetic amino acids. To our knowledge, this is the first description of an *ab initio* method able to predict structure of 'non-natural peptides'.

* *Correspondence to: Laurence Lins, Centre de Biophysique Moléculaire Numérique, ULg, Gembloux Agro Bio Tech, Passage des Déportés, 2, 5030 Gembloux, Belgium. E-mail: l.lins@ulg.ac.be*

*Jérôme Beaufays and Laurence Lins contribute equally to the paper.*

*Centre de Biophysique Moléculaire Numérique, ULg, Gembloux Agro Bio Tech, Passage des Déportés, 2, 5030 Gembloux, Belgium*

**Abbreviations used:** *BB-RMSD, Backbone Root Mean Square Deviation; NMR, Nuclear magnetic resonance; PDB, Protein Data Bank*

## Materials and Methods

### Peplook

Peplook is an iterative Boltzmann-Stochastic algorithm described elsewhere [19]. Briefly, from a sequence, Peplook generates 100–500 times $10^4$ structures using Φ/Ψ couples randomly selected among 64 couples derived from the structural alphabet for protein structures proposed by Etchebest et al. [20]. In this paper, they described the alphabet as being composed of 16 'protein blocks' of five residues in length. Consequently, each 'protein block' can be defined as a succession of four Φ/Ψ couples. The 64 couples of angles are indicated in the Table S1. In the next steps, the probability of each combination of Φ/Ψ couples of angles is increased or decreased if they generate energetically favorable or unfavorable structures, respectively. Step after step, the probability of Φ/Ψ value for each residue varies. When the mean probability of peptide Φ/Ψ couples of angles remains constant, the process is stopped, and 99 structures of lower energy are selected.

For the cyclic peptides, energy gaps are added to force minimal distances: 2.2 Å between the SG of the cysteines implicated in the disulfide bonds and 1.3 Å between the N and the C of amino acids implicated in amide bonds.

For peptides with non-proteinogenic residues, initially, structures of these residues were constructed using HyperChem 5.0 (Hypercube, Gainesville, FL, USA), starting from the backbone of alanine and optimized by the Polak-Ribiere algorithm using the AMBER-95 force field with a gradient $\delta$ inferior to 0.1 kcal/(Å mol), as previously described [21]. Non-proteinogenic amino acids were then added to the set of the 20 natural residues used in Peplook, and calculation of 3D models was run as previously described for peptides consisting of proteinogenic amino acids.

### I-Tasser

In the initial step, I-Tasser [17,18] uses Lomets [22], a meta-threading approach to identify templates for the query sequence in a non-redundant PDB structure library. Then, fragments excised from the consensus threading templates are assembled by modified replica-exchange Monte-Carlo simulations into 3D models. Models are clustered using Spicker [23]. The cluster centroids are

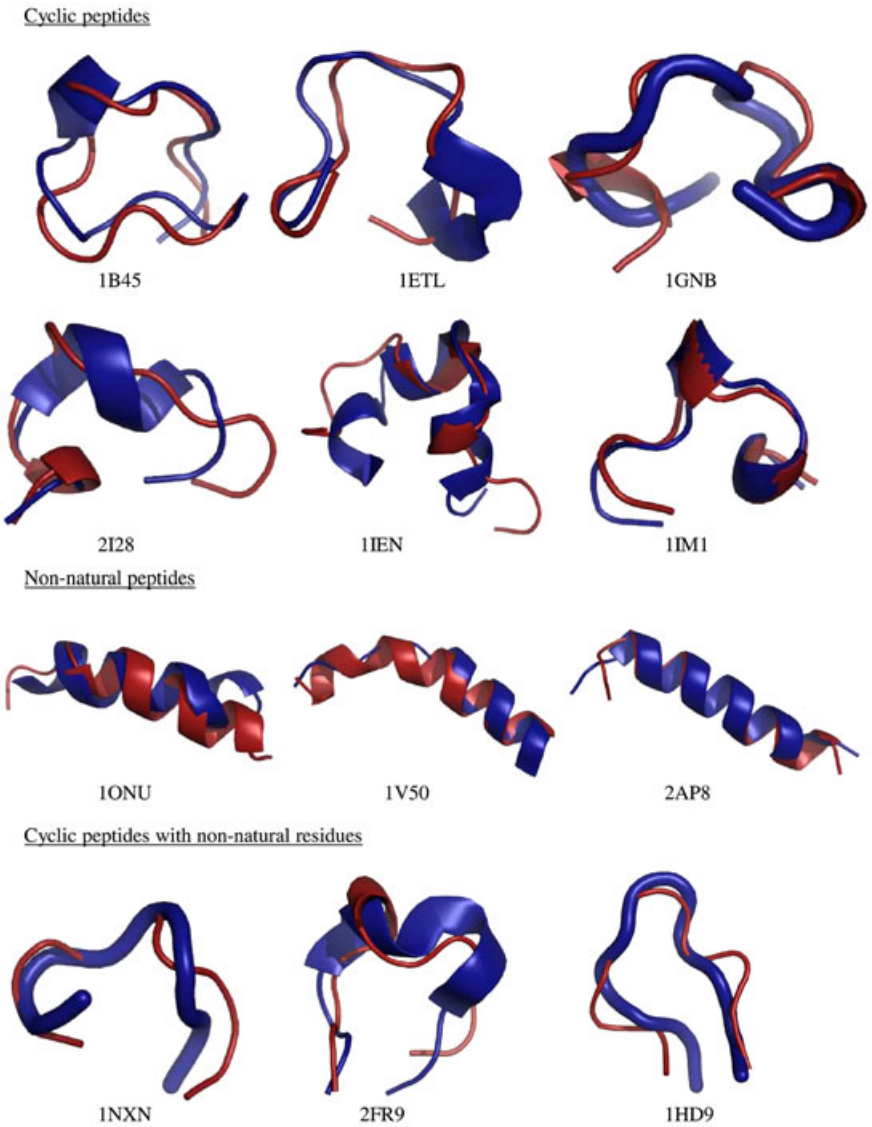


**Figure 1.** Examples of Peplook models. NMR/X-ray structures are in blue and Peplook models in red. Pictures were generated using Pymol software.



  

**Table 1.** List of cyclic peptides used in this study

| | | Peplook | | | | I-Tasser | | | | Pep-Fold | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BB-RMSD | | Cβ-RMSD | | BB-RMSD | | Cβ-RMSD | | BB-RMSD | | Cβ-RMSD | |
| PDB | Length | Prime | Best | Prime | Best | First | Best | First | Best | First | Best | First | Best |
| 2P7R | 5 | 1.0 | 1.0 | 0.2 | 0.2 | | | | | | | | |
| 1QVL | 6 | 1.5 | 1.5 | 2.7 | 2.7 | | | | | | | | |
| 1FOZ | 7 | 1.6 | 1.6 | 2.0 | 2.0 | | | | | | | | |
| 1N0C | 10 | 3.2 | 3.1 | 4.2 | 4.1 | 2.1 | 2.1 | 3.1 | 3.1 | 2.1 | 2.1 | 2.6 | 2.6 |
| 1N0A | 11 | 2.8 | 2.8 | 3.7 | 3.7 | 0.9 | 0.9 | 1.2 | 1.2 | 0.6 | 0.6 | 1.0 | 1.0 |
| 1IM1 | 12 | *2** | *2** | *2.6** | *2.6** | *1** | *1** | *1.2** | *1.2** | *2.3** | *2.3** | *2.0** | *2.0** |
| 1ETL | 12 | 2.0 | 2.0 | 3.2 | 3.2 | 3.3 | 3.3 | 4.0 | 4.0 | 2.5 | 2.5 | 3.6 | 3.6 |
| 1IXU | 12 | 3.5 | 3.5 | 5.1 | 5.1 | 3.3 | 3.0 | 4.6 | 3.7 | 4.1 | 3.8 | 5.4 | 5.4 |
| 1IM7 | 13 | *2.3** | *1.7** | *3.4** | *2.7** | *2.3** | *2.0** | *3.8** | *3.8** | *2.5** | *2.1** | *4.1** | *3.7** |
| 1XGB | 13 | *2.4** | *2.4** | *4.2** | *4.2** | *3** | *2.7** | *4.7** | *4.6** | *3.1** | *3.1** | *4.0** | *4.0** |
| 2I28 | 13 | 2.1 | 2.1 | 3.4 | 3.4 | 4.3 | 3.5 | 4.2 | 4.2 | 3.0 | 2.5 | 3.0 | 3.3 |
| 1GNB | 13 | 2.4 | 2.4 | 3.1 | 3.1 | 5.0 | 1.8 | 4.9 | 2.5 | 5.1 | 5.1 | 6.1 | 6.1 |
| 1HJE | 13 | 3.7 | 3.7 | 4.3 | 4.3 | 3.1 | 2.5 | 3.6 | 3.2 | 2.9 | 2.9 | 4.0 | 4.0 |
| 1JBL | 14 | 3.2 | 3.2 | 4.3 | 4.3 | 2.8 | 1.7 | 3.4 | 2.1 | 2.9 | 2.3 | 3.9 | 3.4 |
| 1B45 | 14 | 2.2 | 2.2 | 2.8 | 2.8 | 2.2 | 1.8 | 2.9 | 2.9 | 5.1 | 2.4 | 4.3 | 3.1 |
| 1R8T | 15 | 2.7 | 2.7 | 3.9 | 3.9 | 3.0 | 2.9 | 4.0 | 4.0 | 3.6 | 3.6 | 4.9 | 4.9 |
| 1MII | 16 | 3.2 | 3.2 | 4.7 | 4.7 | 1.9 | 1.9 | 1.8 | 1.8 | 3.5 | 3.4 | 3.7 | 4.8 |
| 2EFZ | 16 | 3.4 | 3.4 | 4.5 | 4.5 | 2.6 | 2.6 | 3.7 | 3.7 | 3.6 | 3.6 | 5.2 | 5.2 |
| 1KWD | 16 | 3.0 | 3.0 | 4.1 | 4.1 | 2.3 | 2.3 | 3.0 | 3.0 | 3.5 | 3.2 | 4.9 | 4.9 |
| 1NIM | 17 | 3.7 | 2.9 | 4.2 | 3.6 | 4.6 | 4.0 | 5.7 | 5.2 | 4.7 | 3.6 | 5.9 | 5.2 |
| 1IEN | 19 | 3.3 | 3.3 | 4.4 | 4.4 | 2.2 | 2.2 | 2.7 | 2.7 | 3.3 | 2.7 | 4.2 | 3.4 |
| 1X7K | 19 | 3.6 | 3.6 | 4.7 | 4.7 | 2.6 | 2.3 | 4.6 | 4.4 | 5.9 | 4.9 | 7.0 | 5.8 |
| 1V6R | 21 | 5.8 | 4.7 | 8.8 | 5.2 | 7.0 | 6.8 | 7.8 | 7.8 | 7.1 | 5.7 | 8.0 | 6.9 |
| 1TER | 21 | 3.7 | 3.7 | 4.9 | 4.9 | 1.1 | 1.1 | 1.6 | 1.6 | 6.6 | 4.7 | 7.5 | 6.1 |
| 1KCN | 21 | 4.7 | 4.7 | 6.0 | 6.0 | 1.9 | 1.9 | 2.4 | 2.4 | 5.9 | 3.7 | 7.4 | 5.0 |
| 2AJW | 22 | *2.4** | *2.4** | *4.0** | *4.0** | *1.7** | *1.3** | *1.6** | *1.3** | *2.3** | *1.1** | *1.2** | *1.1** |
| 1RPC | 22 | 6.5 | 6.5 | 8.0 | 6.8 | 3.5 | 3.5 | 4.2 | 4.2 | 6.1 | 5.3 | 7.1 | 6.1 |
| 1HP9 | 22 | 5.4 | 4.3 | 6.3 | 5.4 | 1.8 | 1.2 | 1.7 | 1.6 | 3.3 | 2.7 | 4.3 | 3.1 |
| 1OIG | 24 | 6.4 | 6.4 | 7.4 | 7.4 | 1.8 | 1.8 | 2.4 | 2.4 | 6.0 | 5.6 | 7.3 | 6.9 |
| 1ORX | 24 | 6.8 | 6.8 | 7.2 | 7.2 | 1.1 | 1.1 | 1.1 | 1.1 | 7.3 | 4.0 | 6.7 | 5.0 |
| 2OQ9 | 24 | *3.8** | *3.8** | *5.4** | *5.4** | *2.8** | *2.6** | *4.0** | *3.8** | *3.4** | *3.3** | *4.8** | *4.7** |
| 1SP7 | 24 | 4.6 | 4.6 | 5.3 | 5.3 | 2.4 | 2.0 | 3.1 | 3.0 | 6.9 | 4.5 | 7.9 | 6.1 |
| 1WQC | 26 | 4.8 | 4.7 | 5.6 | 5.6 | 2.3 | 2.0 | 2.7 | 2.6 | — | — | — | — |
| 2NX7 | 28 | 7.4 | 6.2 | 8.4 | 6.9 | 0.9 | 0.9 | 0.7 | 0.7 | — | — | — | — |
| 1WM8 | 28 | 5.7 | 5.7 | 6.1 | 6.1 | 3.0 | 3.0 | 3.7 | 3.7 | — | — | — | — |
| 1V5A | 28 | 5.4 | 5.1 | 6.2 | 6.2 | 1.2 | 1.2 | 1.4 | 1.4 | — | — | — | — |
| 2IT7 | 28 | 6.7 | 5.5 | 7.0 | 5.9 | 0.7 | 0.7 | 0.6 | 0.6 | — | — | — | — |
| 1MMC | 30 | *5** | *4.4** | *5.9** | *5.5** | *1.9** | *1.9** | *2.6** | *2.6** | — | — | — | — |

  

**Table 1.** *(Continued)*

| PDB | Length | Peplook | | | | I-Tasser | | | | Pep-Fold | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BB-RMSD | | Cβ-RMSD | | BB-RMSD | | Cβ-RMSD | | BB-RMSD | | Cβ-RMSD | |
| | | Prime | Best | Prime | Best | First | Best | First | Best | First | Best | First | Best |
| Mean | All | 3.8 | 3.6 | 4.8 | 4.5 | 2.5 | 2.2 | 3.1 | 2.9 | 4.1 | 3.4 | 4.9 | 4.4 |
| | 10-20AA | 2.9 | 2.8 | 3.9 | 3.9 | 2.8 | 2.4 | 3.6 | 3.2 | 3.4 | 3.0 | 4.2 | 4.0 |
| | 21-30AA | 5.3 | 4.9 | 6.4 | 5.9 | 2.2 | 2.1 | 2.6 | 2.5 | 5.5 | 4.1 | 6.2 | 5.1 |

RMSD_P, RMSD_Prime; RMSD_B, RMSD_Best; RMSD_F, RMSD_First; BB-RMSDs in italic followed by a star correspond to BB-RMSDs calculated on the rigid core of the peptide.

subjected to a second iteration process in order to remove clashes and to refine global topology. The final all-atom models are generated by Remo through optimization of hydrogen bonding networks [24]. To evaluate predicted models, a C-score is defined based on the quality of the threading alignments and the convergence of parameters of the structure assembly simulations. Web address: http://zhanglab.ccmb.med.umich.edu/I-TASSER/

### Pep-Fold

Pep-Fold uses hidden Markov model-derived structural alphabet of 27 motifs to describe the conformations of four consecutive residues [10,11]. It first determines structural alphabet letters of the sequence and then builds 3D models by assembling the fragments using a greedy algorithm driven by a coarse-grained force field. For each target sequence, 50 greedy simulations are performed and the 50 models are clustered. The all-atom models are then generated followed by a fast minimization performed with Gromacs [25]. As distance restraint cannot be applied with Pep-Fold, the procedure used for cyclic peptide structure determination was the same as for linear peptides. Web address: http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD/

## Results

### Structure Prediction of Cyclic Peptides

We tested the performance of Peplook to predict structure of cyclic peptides. A set of 38 sequences of PDB models was used for this study, with lengths varying from 5 to 30 residues. Most of the 3D models were solved by NMR, only two were obtained by X-ray crystallography. Cyclization of these peptides occurs either by a disulfide bond or by an amide bond between the amino and carboxyl groups of side chains or backbones.

Peplook, Pep-Fold, and I-Tasser were used to generate the models. To cyclize the peptides, distance restraints, corresponding to the lengths of disulfide bridge and amide bond, were applied in Peplook and I-Tasser. Figure 1 shows some examples of predicted models fitted on the corresponding NMR or X-ray models. To evaluate the accuracy of predictions, we calculated the root mean square deviation of backbone atoms (BB-RMSD) between predicted and experimental models. Table 1 reports the BB-RMSD of the Prime (Peplook) and First model (I-Tasser) (corresponding to the lowest energy conformation in each procedure) and the BB-RMSD of the best model (corresponding to the model with the lowest BB-RMSD in each procedure). We have to note that some PDB models are partially flexible, notably at the N- and C- ends, as usually observed for NMR models. For them, RMSD from the average structure was actually calculated on the structurally monomorphic part ('rigid core') of the peptide. Consequently, the BB-RMSDs calculated in our study are also based on the same part of the peptide. These peptides are indicated in italic followed by an asterisk in Table 1.

For the Peplook Primes, the overall average BB-RMSD of the 38 models is 3.8 Å. Interestingly, the score is very close to that of the best models (3.6 Å). This is comparable with the score of Pep-Fold models (4.1 Å for the first and 3.4 Å for the best). BB-RMSD for models predicted by I-Tasser are lower (2.5 and 2.2 Å for the first and the best models, respectively). It should be noted that the use of Pep-Fold is restricted to peptides of 9–25 amino acids and that of I-Tasser, to peptides of at least 10 residues, so that only 29 of the 38 models were calculated with Pep-Fold and 35 with I-Tasser.

Journal of PeptideScience

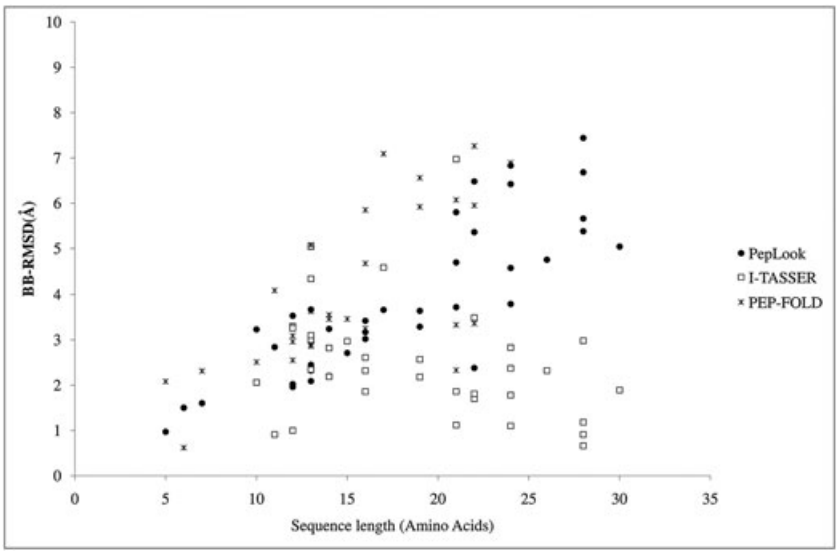


**Figure 2.** RMSD-sequence length relationship for Peplook, I-Tasser, and Pep-Fold models.

We also analyzed the performance of Peplook as a function of peptide length. Figure 2 shows that BB-RMSD values increase for peptides of more than 20 amino acids. Models of peptides with 5–20 residues have a BB-RMSD lower than 4 Å (Figure 3), with individual values between 1 Å (1IM1) and 3.7 Å (1HJE). The average BB-RMSD is 2.9 Å for the Primes and 2.8 Å for the best models for the 10–20 amino acids peptides. These results are similar to those obtained with structures predicted with I-Tasser (2.8 Å BB-RMSD for the first and 2.4 Å BB-RMSD) and with Pep-Fold (3.4 Å BB-RMSD for the first and 3.0 Å BB-RMSD).

For peptides of 21–30 amino acids, models predicted with Peplook have BB-RMSD values between 2.4 Å (2AJW) and 7.4 Å (2NX7) (Figure 3). The mean BB-RMSDs for the 16 longer peptides are 5.3 Å (first models) and 4.9 Å (best models). These values are comparable with the Pep-Fold predictions (5.5 and 4.1 Å for the first and for the best models, respectively). In contrast, averaged BB-RMSD of the I-Tasser models is close to 2 Å irrespective of the size of peptides (2.2 Å for the first models and 2.1 Å for the best models for longer peptides).

In a general manner, the BB-RMSD of cyclic peptides is lower than linear ones (data not shown). To evaluate more accurately the models, C$\beta$-RMSD was therefore calculated (Table 1). The results with C$\beta$-RMSD are similar to those obtained with BB-RMSD, whereas the C$\beta$-RMSD values being slightly higher. The overall average C$\beta$-RMSD of Peplook models is comparable with the score of structures predicted with Pep-Fold. The value for the 38 Peplook models is 4.8 Å (Prime) and 4.5 Å (best) and 4.9 Å (first) and 4.4 Å (best) for the Pep-Fold models. C$\beta$-RMSD of I-Tasser predicted structures is also lower (3.1 Å for the first and 2.9 Å for the best).

### Structure Prediction of 'Non-Natural Peptides'

Thirteen 'non-natural peptides' were chosen with length varying from 9 to 27 amino acids. What we name here 'non-natural peptides' are peptides with natural, but non-proteinogenic amino acids and with non-natural amino acids. Some examples of predicted models fitted to NMR structures are shown in Figure 1.

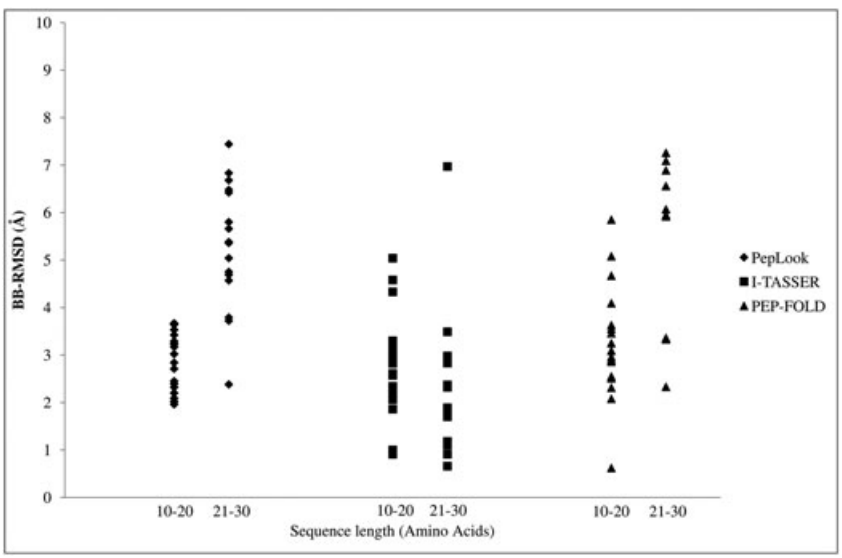


**Figure 3.** RMSD distribution for 10–20 and 21–30 amino acid peptides models.



  

**Table 2.** List of 'non-natural peptides' used in this study

| PDB | Length | RMSD_P | RMSD_B | Sequence | NPAA |
|---|---|---|---|---|---|
| 2RLL | 9 | 2.8 | 2.6 | SPI **TYS** DIN **TYS** Y | TYS: *O*-sulfo-L-tyrosine |
| 2JQC | 15 | 2.4 | 2.1 | DWEYHAHPK **HYP** NSFWT | HYP: hydroxyproline |
| 1ONU | 17 | 2.0 | 1.7 | GE **CGU** **CGU** LQ **CGU** NQ **CGU** LIR **CGU** KSN | CGU: gamma-carboxyglutamic acid |
| 1V50 | 17 | 2.6 | 2.1 | KISSPTE **TPO** ERCIESLIA | TPO: phosphothreonine |
| 2CEZ | 19 | 4.4 | 3.4 | CRKAGVGQ **PSE** WKENSPLNVS | PSE: phosphoserine |
| 2CEF | 19 | 4.3 | 3.9 | CRKAGVGQ **PSE** WKEN **PSE** PLNVS | SEP: phosphoserine |
| 1VQX | 19 | 4.5 | 4.1 | DDEA **SEP** **TPO** **TPO** V **SEP** K **TPO** E **TPO** **SEP** QVAPA | SEP: phosphoserine, TPO: phosphothreonine |
| 2AP8 | 20 | 1.4 | 1.3 | I **DIL** GPVLGLVGSALGGLLKKI | DIL: D-isoleucine |
| 1GEA | 21 | 4.2 | 3.4 | HSDGIFTDSYSRYRKQMAVK **LYN** | LYN: 2,6-diamino-hexanoic acid amide |
| 1ONT | 21 | 3.3 | 2.9 | GE **CGU** **CGU** YQKML **CGU** NLR **CGU** AEVKKNA | CGU: gamma-carboxyglutamic acid |
| 1T8J | 23 | 5.3 | 4.6 | YRV **DPR** SYDFSRSDELAKLLRQHAG | DPR: D-proline |
| 1HCW | 24 | 4.8 | 4.7 | YTVPS **PYA** TFSRSDELAKLLRLHAG | PYA: 3-(1,10-phenanthrol-2-yl)-L-alanine |
| 2G57 | 27 | 5.0 | 4.0 | KAAVSHWQQQSYLD **SEP** GIH **SEP** GATTTAP | SEP: phosphoserine |
| Mean | 9-30 AA | 3.6 | 3.1 | | |
| | 9-20AA | 3.0 | 2.6 | | |
| | 21-27AA | 4.5 | 3.9 | | |

NPAA, nature of the non-proteogenic amino acid.

The averaged BB-RMSD of the Primes and the best models with respect to the NMR models are 3.6 and 3.1 Å, respectively (Table 2). These RMSD values vary from 1.4 to 5.3 Å for the primes and from 1.3 to 4.7 Å for the best models. As observed for the cyclic peptides, the BB-RMSD increases with the peptide length but the increase is less important: the averaged RMSDs are 3.0 Å (Prime) and 2.6 Å (Best) for the 9–20 amino acids peptides and 4.5 Å (Prime) and 3.9 Å (Best) for the 21–27 amino acids peptides (Table 2).

Because no available software is dedicated to the prediction of 'non-natural peptides', we had no other prediction for comparison. Nevertheless, we compared predictions for natural and 'non-natural peptides' to assess the potentiality of Peplook (Table 3). We used three 'non-natural peptides' (1 V50, 2CEF, and 2AP8), which had a related natural peptide experimentally solved. For these three examples, the BB-RMSDs with respect to the experimental models are similar for the natural and 'non-natural peptides'.

### Structure Prediction of Cyclic Peptides with Non-Proteinogenic Residues

We also tested Peplook with peptides that have both non-proteinogenic amino acids and a cyclic structure. We selected 14 PDB sequences for this study. Two of these peptides were cyclized through an amide bond (1JAR and 1T9E) and the others by a disulfide bond. Figure 1 shows fitted Peplook models on the NMR structures. The overall averaged BB-RMSD is 3.4 Å for the Prime models and is 3.2 Å for the best models (Table 4). Although our analysis carries on a limited number of peptides, we observed that BB-RMSDs are lower for the shorter peptides.

For four peptides (1JAR, 2FR9, 1T9E, 1ZWU), 3D models of the homologous natural peptides were solved, and we compared their conformations with those of Peplook models for the natural and the 'non-natural peptides' (Table 5). In addition, for two of them (1JAR, 2FR9), 3D structures of two 'non-natural peptides' were determined because the non-proteinogenic amino acids is located at two different positions. BB-RMSDs of 'non-natural peptides' are similar to those of the natural peptides, as already observed for the non-cyclic peptides. Furthermore, the same values are observed when 'non-natural models' are available. As an example, 2FRB and 2FR9 with a *p*-(benzoyl)-phenylalanine instead of Asn4 (2FRB) or Ser12 (2FR9) were compared with 1NOT, the related natural peptide. The BB-RMSDs are 2.4, 2.4, and 2.7 Å for the Primes of 1NOT, 2FRB, and 2FR9, respectively, and 2.4, 2.4, and 2.3 Å for the best models, respectively.

## Discussion

Peplook is an iterative Boltzmann-Stochastic algorithm to predict 3D models of peptide sequences up to 30 amino acids. In this paper, we show that Peplook can also be used for cyclic and 'non-natural peptides'. BB-RMSD values were used to evaluate the

**Table 3.** Comparison of PepLook performance between natural and non-natural homologous peptides

| PDB | RMSD_P | RMSD_B | Sequence | NPAA |
|---|---|---|---|---|
| 1V4Z | 2.3 | 1.6 | KISSPTETERCIESLIA | |
| 1V50 | 2.6 | 2.1 | KISSPTE **TPO** ERCIESLIA | TPO: phosphothreonine |
| 2CEH | 4.7 | 3.0 | CRKAGVGQSWKENSPLNVS | |
| 2CEZ | 4.4 | 3.4 | CRKAGVGQ **PSE** WKENSPLNVS | PSE: phosphoserine |
| 2CEF | 4.3 | 3.9 | CRKAGVGQ **PSE** WKEN **PSE** PLNVS | PSE: phosphoserine |
| 2AP7 | 1.5 | 1.2 | IIGPVLGLVGSALGGLLKKI | |
| 2AP8 | 1.4 | 1.3 | I **DIL** GPVLGLVGSALGGLLKKI | DIL: D-isoleucine |

NPAA: see Table 2.

Journal of PeptideScience

**Table 4.** List of non-natural cyclic peptides used in this study

| PDB | length | RMSD_P | RMSD_B | Sequence | NPAA |
|---|---|---|---|---|---|
| 1YL8 | 8 | 2.7 | 1.4 | **DPN** CY **DTR** KTCT | DPN: D-phenylalanine, DTR: D-tryptophan |
| 1NXN | 9 | 1.9 | 1.9 | GDCP **DTR** KPWC | DTR: D-tryptophan |
| 1HD9 | 11 | 2.4 | 2.4 | **NOL** CTASIPPQCY | NOL: norleucine |
| 1XY6 | 12 | 2.5 | 2.5 | YCKFE **DTR** **IAM** TFKSC | DTR: D-tryptophan, IAM: 4-[(isopropylamino)methyl]phenylalanine |
| 1JAR | 13 | 2.6* | 2.6* | IWGDSGKLI **DAB** TTA | DAB: 2,4-diaminobutyric acid |
| 1J9V | 13 | 2.4* | 2.4* | IWG **DAB** SGKLIDTTA | DAB: 2,4-diaminobutyric acid |
| 2FRB | 13 | 2.4 | 2.4 | ECC **PBF** PACGRHYSC | PBF: *p*-(benzoyl)-phenylalanine |
| 2FR9 | 13 | 2.7 | 2.3 | ECCNPACGRHY **PBF** C | PBF: *p*-(benzoyl)-phenylalanine |
| 1XBH | 13 | 3.2 | 3.2 | CIYYKDGEALKY **DCY** | DCY: D-cysteine |
| 1T9E | 14 | 4.8 | 4.8 | GR **ABA** TKSIPPI **ABA** FPD | ABA: alpha-aminobutyric acid |
| 1MTQ | 19 | 3.8 | 3.8 | IRD **CGU** CCSNPACRVNN **HYP** HVC | CGU: gamma-carboxyglutamic acid, HYP: hydroxyproline |
| 1TCG | 22 | 3.7 | 3.5 | RDCCT **HYP** **HYP** KKCKDRQCK **HYP** QRCCA | HYP: hydroxyproline |
| 2JUY | 28 | 6.7 | 6.0 | FFCPFGCALVDCGPNRPCRDTGF **SME** SCDC | SME: methionine sulfoxide |
| 1ZWU | 30 | 6.4 | 5.1 | VGECVRGRCPSGMCCSQ **NAL** GYCGKGPKYCGR | NAL: beta-(2-naphthyl)-alanine |
| Mean | | 3.4 | 3.2 | | |

NPAA: see Table 2.

calculated models with respect to their experimental NMR or X-ray counterparts. In addition, for cyclic peptides, Peplook models were compared with those obtained with two other prediction servers: I-Tasser and Pep-Fold. For the ensemble of peptides, the BB-RMSDs of Peplook models are similar to the values obtained with Pep-Fold models but are higher than those of I-Tasser models. However, all target sequences that we tested in this study come from the PDB. As I-Tasser uses a PDB library to thread the sequence, in many cases, it used the template model to build the model, an artifact that could not be prevented.

Results show that Peplook is very reliable for peptides of 5–20 residues. For the 19 structures of that series, BB-RMSDs never exceed 4 Å (values are between 2 and 3.7 Å) with an average of 2.9 Å. However, for longer peptides, the results are less convincing. Indeed, BB-RMSD values are higher, especially for the 26–30 residues peptides with an average BB-RMSD of 5.3 Å. Although Peplook can predict structure of 'linear' peptide up to 30 residues, for cyclic peptides, its optimal performances are restricted to sequences up to 20–25 amino acids. In addition, some of the tested peptides have a proline in a *cis* conformation (e.g. 1IXU, 1ORX, 1JBL, 2OQ9). Because Peplook is able to predict structure of peptides containing *cis*-proline, we compared the results obtained with both proline conformations, and we noted that BB-RMSDs are similar.

We also presented another original option of Peplook: its capacity to predict 3D structure of peptides with non-proteinogenic residues. We showed that performances of Peplook are not affected by the presence of these residues. Indeed, results were similar for peptides with or free of non-proteinogenic amino acids. It should be noted that Peplook uses non-proteinogenic residues that differ from the natural ones only by their side chains, meaning that backbones of the non-proteinogenic residues are the same as those of the 20 linear natural amino acids. BB-RMSD values of peptides with non-proteinogenic amino acids are between 1.4 and 5.3 Å with an average of 3.6 Å. As observed for cyclic peptides, the BB-RMSDs increase with peptide lengths, but the difference is smaller than for the cyclic structures (3.0 Å for 10–20 amino acid peptides and 4.5 Å for 21–27 amino acid peptides) supporting the conclusion that Peplook is an interesting procedure to predict 3D structures of 'non-natural peptides' up to 30 amino acids.

In the same manner, we showed that Peplook can be used to predict structure of peptides that combine a cyclic structure and non-proteinogenic amino acids. Results are similar to those

**Table 5.** Comparison of Peplook performance between natural and non-natural homologous cyclic peptides

| PDB | RMSD_P | RMSD_B | Sequence | NPAA |
|---|---|---|---|---|
| 1NOT | 2.4 | 2.4 | ECCNPACGRHYSC | |
| 2FRB | 2.4 | 2.4 | ECC **PBF** PACGRHYSC | PBF: *p*-(benzoyl)-phenylalanine |
| 2FR9 | 2.7 | 2.3 | ECCNPACGRHY **PBF** C | PBF: *p*-(benzoyl)-phenylalanine |
| 1IM7 | 2.3 | 2.3 | IWGCSGKLICTTA | |
| 1J9V | 2.4 | 2.4 | IWG **DAB** SGKLIDTTA | DAB: 2,4-diaminobutyric acid |
| 1JAR | 2.6 | 2.6 | IWGDSGKLI **DAB** TTA | DAB: 2,4-diaminobutyric acid |
| 1JBL | 3.2 | 3.2 | GRCTKSIPPICFPD | |
| 1T9E | 4.8 | 4.8 | GR **ABA** TKSIPPI **ABA** FPD | ABA: alpha-aminobutyric acid |
| 1MMC | 6.0 | 5.3 | VGECVRGRCPSGMCCSQFGYCGKGPKYCGR | |
| 1ZWU | 6.4 | 5.1 | VGECVRGRCPSGMCCSQ **NAL** GYCGKGPKYCGR | NAL: beta-(2-naphthyl)-alanine |

NPAA: see Table 2.



  

obtained for the cyclic and the 'non-natural peptide' predictions with an averaged BB-RMSD of 3.4 Å.

In conclusion, we have shown that Peplook is suitable to predict 3D conformation of 1–25 residues long cyclic peptides and containing non-proteinogenic residues. For cyclic peptides, Peplook is as potent as other in silico methods, such as Pep-Fold or I-Tasser, when the peptide length is less than 25 residues. For longer peptides, up to 30 amino acids, I-Tasser performs better than the two other methods, but it introduces a bias because of the potential use of the query structure as template. The improvement of Peplook to reach better prediction for longer peptides, natural, cyclic, or non-natural, is currently under investigation.

In a general way, Peplook can be used to predict the conformation as well as the lability of a peptide (linear, cyclic, and/or non-natural) and to evaluate the influence of mutations on the structure as already stated for natural peptides [19].

It is worth noting that Peplook is, to our knowledge, the only method able to predict the structure of 'non-natural peptide'. This is notably because of its *ab initio* search procedure for structure; indeed, template models of 3D structures with non-proteinogenic residues are not available for homology and threading protocols.

Peplook could be an interesting tool in the field of peptide therapy. A rapid expansion in the use of peptides as drugs has been observed in the last few years. However, development of peptides as therapeutic drugs is limited by their weak metabolic stability and low bioavailability. A solution to overcome these problems could be to use modified peptides including non-proteinogenic residues and/or cyclic peptides.

#### Acknowledgements

This work was supported by the Ministère de la region Wallonne contract n°516268 (PEPSEIN) and by the Interuniversity Attraction Pole contract (IAP. P6/19 PROFUSA). R. Brasseur and L. Lins are, respectively, Research Director and Research Associate at the National Funds for Scientific Research (FNRS) of Belgium. Annick Thomas is Research Director at the 'Institut National de la Santé et de le Recherche Médicale' (INSERM, France).

## References

1 Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993; **234**: 779–815.
2 Sanchez R, Sali A. Advances in comparative protein-structure modelling. *Curr. Opin. Struct. Biol.* 1997; **7**: 206–214.
3 Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 2000; **29**: 291–325.
4 Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 1991; **253**: 164–170.
5 Jones DT, Taylor WR, Thornton JM. A new approach to protein fold recognition. *Nature* 1992; **358**: 86–89.
6 Luthy R, McLachlan AD, Eisenberg D. Secondary structure-based profiles: use of structure-conserving scoring tables in searching protein sequence databases for structural similarities. *Proteins* 1991; **10**: 229–239.
7 Bradley P, Misura KM, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science* 2005; **309**: 1868–1871.
8 Monge A, Friesner RA, Honig B. An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure. *Proc. Natl Acad. Sci. U S A* 1994; **91**: 5027–5029.
9 Eyrich VA, Standley DM, Felts AK, Friesner RA. Protein tertiary structure prediction using a branch and bound algorithm. *Proteins* 1999; **35**: 41–57.
10 Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.* 2009; **37**: W498-503.
11 Maupetit J, Derreumaux P, Tuffery P. A fast method for large-scale de novo peptide and miniprotein structure prediction. *J. Comput. Chem.* 2009; **31**: 726–738.
12 Kaur H, Garg A, Raghava GP. PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept. Lett.* 2007; **14**: 626–631.
13 Hung LH, Samudrala R. PROTINFO: secondary and tertiary protein structure prediction. *Nucleic Acids Res.* 2003; **31**: 3296–3299.
14 Hung LH, Ngan SC, Liu T, Samudrala R. PROTINFO: new algorithms for enhanced protein structure predictions. *Nucleic Acids Res.* 2005; **33**: W77-80.
15 Bystroff C, Thorsson V, Baker D. HMMSTR: a hidden Markov model for local sequence-structure correlations in proteins. *J. Mol. Biol.* 2000; **301**: 173–190.
16 Bystroff C, Shao Y. Fully automated *ab initio* protein structure prediction using I-SITES, HMMSTR and ROSETTA. *Bioinformatics* 2002; **18** Suppl. 1: S54-61.
17 Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. *Proc. Natl Acad. Sci. U S A* 2004; **101**: 7594–7599.
18 Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008; **9**: 40.
19 Thomas A, Deshayes S, Decaffmeyer M, Van Eyck MH, Charloteaux B, Brasseur R. Prediction of peptide structure: how far are we? *Proteins* 2006; **65**: 889–897.
20 Etchebest C, Benros C, Hazout S, de Brevern AG. A structural alphabet for local protein structures: improved prediction methods. *Proteins* 2005; **59**: 810–827.
21 Lins L, Charloteaux B, Heinen C, Thomas A, Brasseur R. "De novo" design of peptides with specific lipid-binding properties. *Biophys. J.* 2006; **90**: 470–479.
22 Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 2007; **35**: 3375–3382.
23 Zhang Y, Skolnick J. SPICKER: a clustering approach to identify near-native protein folds. *J. Comput. Chem.* 2004; **25**: 865–871.
24 Li Y, Zhang Y. REMO: a new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins* 2009; **76**: 665–676.
25 Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 2005; **26**: 1701–1718.

  