J Math Chem (2012) 50:421–428
DOI 10.1007/s10910-011-9937-x

ORIGINAL PAPER

Peptide structure prediction using distributed volunteer computing networks

Timo Strunk · Moritz Wolf · Wolfgang Wenzel

Received: 27 September 2011 / Accepted: 20 October 2011 / Published online: 6 November 2011

Abstract Recent investigations to develop novel antimicrobial, antibiotical drugs have focused on the development of artificial protein peptides. As short peptides are naturally involved in many important biological processes in the cell and therefore target many kinds of cells. To functionalize peptides it is vital to design peptides, which can differentially target bacterial and eucariotic cells. Although the length of the peptides investigated in this study was limited to 16 amino acids, the number of possible peptide sequences is still too large to synthesize them in a trial- and error manner, therefore requiring a method for directed, but also high-througput peptide design. By predicting the structure of peptide proteins, this design process can be supported through structure-function analysis and peptide-membrane interaction simulation. In this investigation we could predict peptide structures de-novo, i.e. with the sequence information alone, using a massively parallel simulation scheme. We sample a sizable fraction of the peptide's conformational space using Monte-Carlo simulations in the free-energy forcefield PFF02 on the volunteer computing network POEM@HOME. This forcefield models the protein's native conformation as the global minimum of the free-energy. We could identify peptides of different topologies in a completely automated manner, which allows for the high-throughput screening of large peptide databases for their structural features, which would allow the rapid protopying of peptides needed for novel peptide design.

Keywords Protein structure · Structure prediction · Distributed computing · BOINC · Volunteer computing

T. Strunk (✉) · M. Wolf · W. Wenzel
Institute of Nanotechnology, Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: Timo.Strunk@kit.edu

1 Introduction

Antimicrobial, antibiotical and antifungal peptides [4,10] are hoped to complement current antibiotic drugs. The last novel effective new antibiotic drug was released to market around 1970. As bacteria could adapt to and resist current drugs, effectiveness of those antibiotics decreased significantly.

Although studied peptide sequences are small (10–20 amino acids) compared to many proteins in the human body, the sequence space of peptides is still too large to screen for novel drugs experimentally or in silico by brute-force testing of random sequences. A fast and accurate structure prediction method for peptide structures could aid in the design process with insight into structure-function relationships of peptides and implications of sequence mutations.

Here we present a method for high-throughput peptide structure prediction on our volunteer distributed computing grid POEM@HOME using about 5000 parallel simulations for each peptide. We could verify this technique by predicting the structure of four experimentally known peptide structures of different fold motifs to experimental resolution [2,3,6,7].

2 Methods

2.1 Forcefield

The free energy PFF02 models the native protein conformation as the global free-energy minimum [5,8,9] using potentials for Lennard-Jones, Electrostatic, angle-dependent hydrogen bonding and implicit solvation interactions. To enable the correct sampling of also the β-region of the Ramachandran plot, a semi-empirical torsion potential is included.

2.2 Simulation protocol

In each Monte Carlo step we randomly perturb one angular degree of freedom, i.e. the main-chain and side-chain dihedral angles. The new angles are drawn from Gaussian distributions with a width of 10° around the original angles; they are also selected randomly from a distribution reflecting the naturally occurring distribution of phi- and psi-angles in the Ramachandran plot. To model the Ramachandran plot we used angles randomly drawn from equidistributions of two circles of radii 45 at the centers $(-125, 135)$ for the right-handed helical region and $(-70, -35)$ for the β-sheet region of the Ramachandran plot.

2.3 Poem@Home

All simulations were run on the distributed volunteer computing platform for protein simulation POEM@HOME using the BOINC [1] framework. POEM@HOME is a public server in the spirit of SETI@HOME, where volunteers can download the

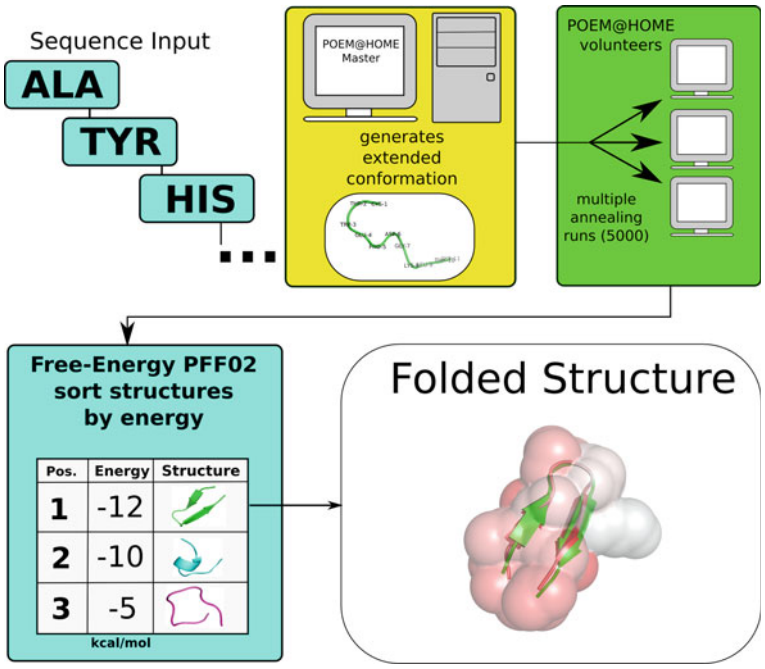

Fig. 1 Prediction Algorithm. From sequence input alone, a multitude of annealing simulations are started on the POEM@HOME volunteer PCs. POEM@HOME keeps track of the energies of these conformations and returns the best energy one as the structure prediction

protein folding client POEM and continously simulate protein structures on their PCs. The combined computing power has grown to 30 TFlop/s on average since its start in August 2008.

Starting from a completely extended conformation, we simulate about 5000 copies of this conformation in 1.5 million step Monte-Carlo runs using geometric temperature scaled simulated annealing in parallel. Once a population converges, i.e. once no new structures are discovered, the lowest energy conformation is then chosen as the predictive model of the experimental structure. This technique was shown to work with 32 medium-size proteins [8]. Figure 1 illustrates the prediction protocol.

2.4 Clustering

The complete population of simulated structures is clustered by RMSD to elucidate the importance of low-energy conformations.

1. The current best energy structure is chosen from the population.
2. All structures in the vicinity of the current best structure (RMSD $< 1.6\,\text{Å}$) are joined to a cluster.
3. The structures are moved out of the population.
4. The algorithm repeats from step 1, until the population is empty.

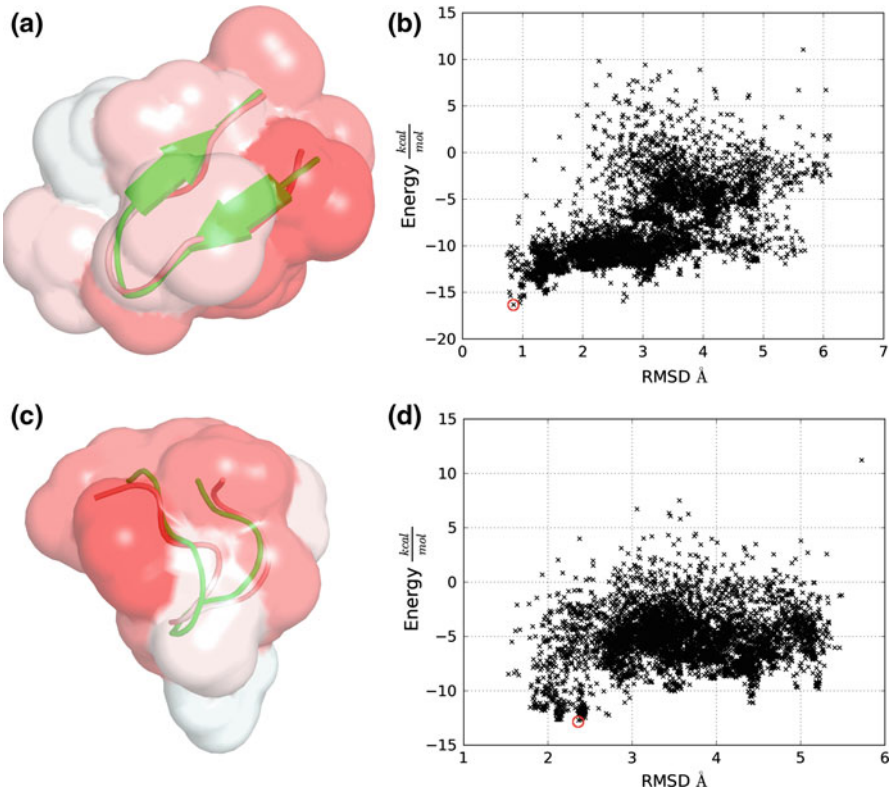

Fig. 2 **a** Result of the folding simulations of the peptide 1N0D (overlay). The predicted β-fold (*green*, in print: *light gray*) corresponds perfectly with the experimental structure (*red–white*, in print: *dark gray*). **b** RMSD-Energy Plot of all simulations of 1N0D. The chosen prediction has an all-atom RMSD of 0.85 Å (*circle*). **c** Comparison of the predicted structure of 1FUV (*green*, in print: *light-gray*) with the experimental structure (*red*, in print: *dark gray*) with a RMSD of 2.4 Å. **d** RMSD versus Energy plot of all simulated conformations for 1FUV. Better conformations could be discovered; they were energetically disfavored. (Both graphs): *Dark red* tones (in print: *dark gray* tones) correspond to hydrophobic amino acids. *Light red* to *white* (in print: *light gray* to *white*) corresponds to hydrophilic amino acids (Color figure online)

This generates clusters around minimum energy conformations and does not chose an energetically not favorable centroid structure as a reprentative conformation of a cluster. The minimum energy conformations are then chosen from the clusters within an energy threshold ($\Delta E < 2.5$ kcal/mol).

3 Results

We verified the prediction framework by folding four peptides of different topology to experimental resolution. Among these are the one-turn helical fold 1EGS [6], the β-like fold 1N0D [7] and the two random coil-like folds 1FUV [2] and 2JQU [3] (four letter codes correspond to RCSB pdb ids).

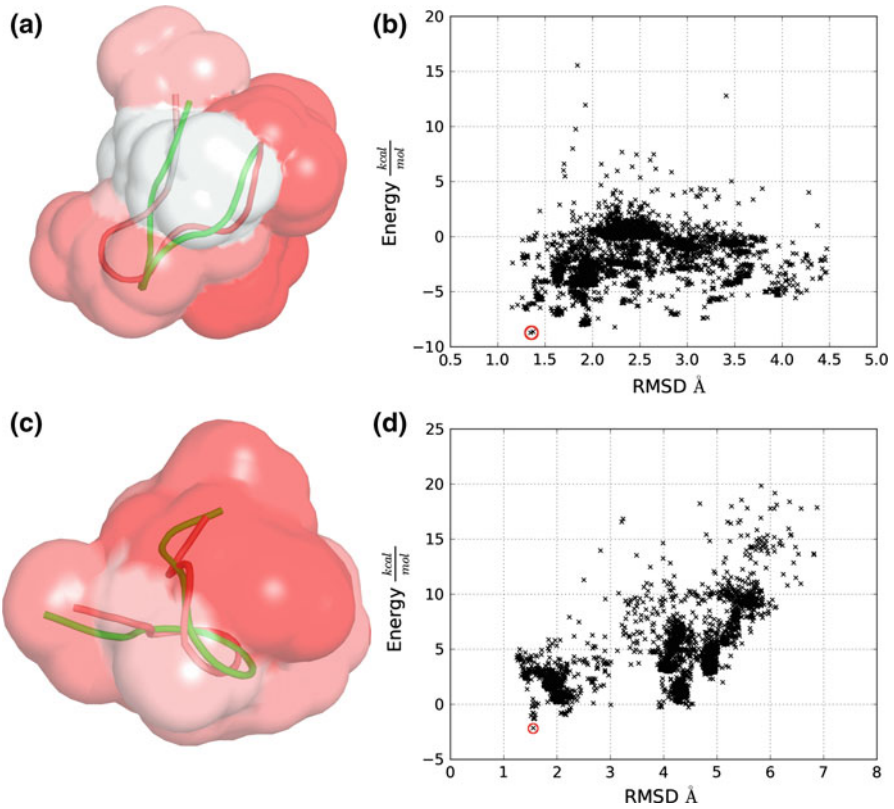

Fig. 3 **a** Result of the folding simulations of the peptide 1EGS. Similiar to the experimental structure (*red*, in print: *dark gray*) the predicted structure shows a shift in the β-like fold. **b** RMSD-Energy Plot of all simulations of 1EGS. The chosen prediction has an all-atom RMSD of 1.4 Å (*circle*). **c** Comparison of the predicted structure (*green*, in print: *light gray*) of 2JQU with the experimental structure (*red*, in print: *dark gray*). The helical-like fold of 2JQU was identified as native. **d** RMSD versus Energy plot of all simulated conformations for 2JQU. Only a few structures of better RMSD than the chosen one of 1.5 Å (*circle*) were identified. (Both graphs): *Dark red* tones (in print: *dark gray*) correspond to hydrophobic amino acids. *Light red* to *white* (in print: *light gray*) corresponds to hydrophilic amino acids (Color figure online)

Figure 2a shows the overlay between the predicted and experimental structure of β-structure 1N0D. The predicted structure agrees perfectly with the experimental structure separated only by a RMSD of 0.85 Å. The RMSD-energy distribution of sampled structures in Fig. 2b shows that the next unfolded structure is separated by an energy-barrier of about 1kcal/mol . It is notable that apart from these two conformations, no distinct other low energy conformations could be found. Furthermore the native structure features an energetically unfavourable hydrophobic patch (dark red spot in Fig. 2a, which could be predicted.

The collapsed coil-fold of 1FUV could be predicted to a RMSD of 2.4 Å (Fig. 2c). Few different low-energy clusters of structures could be identified at 1, 4 and 5 Å RMSD. The best discovered structure has a RMSD of 1.6 Å to the native structure and is separated by a large margin of 5 kcal/mol to the next unfolded structure. The fold

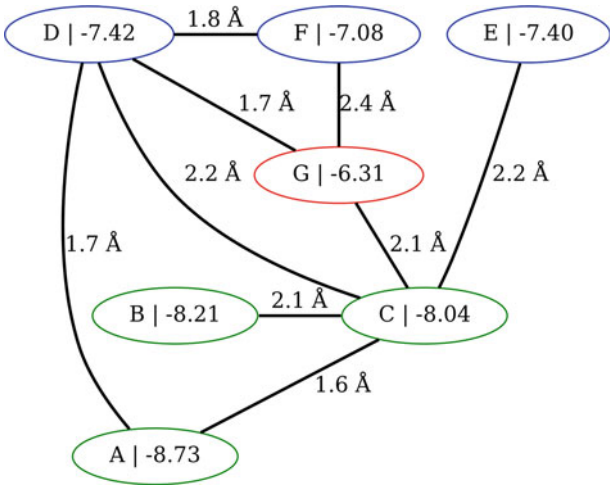

Fig. 4 Connectivity tree of all found distinct conformations of protein 1EGS below an energy of 6.3 kcal/mol. The connections show the relative RMSDs between the structure, while the labels show the protein ID and the respective energy in kcal/mol. IDs correspond to the plots in Fig. 5

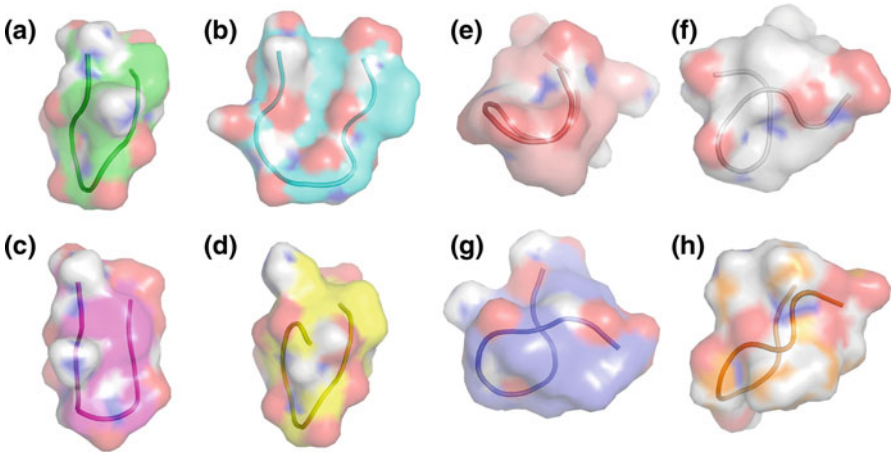

Fig. 5 Low energy conformations of peptide 1EGS. Energies are sorted in increasing order from **a** to **g**. **h** is the experimental native conformation. Conformations **c** and **g** are connected to most of the other conformations as only slight modifications are needed to transform them into either coiled or β-conformations

of 1EGS (Fig. 3a) resembles a distorted β-conformation. The usually stabilizing *Zipper*-mechanism is not pronounced as in 1N0D. 1EGS could be predicted to a RMSD of 1.4 Å, which is separated by about 3 kcal/mol energetically from the next unfolded conformation (Fig. 3c). Additionally the plot shows a reasonable correlation between energy and RMSD for the low energy structures.

This correlation is also apparent and more strongly pronounced in the sampled conformations of 2JQU seen in Fig. 3d. Three kcal/mol separate the native and the next unfolded conformation sampled. There was no pronounced low energy

conformation sampled with a RMSD bigger than 4.5 Å. This can be explained by the fact that the native helical conformation (Fig. 3c) is stabilized mostly by local interactions. Upon unfolding, the protein sacrifices most of its energy attained by hydrogen bonds.

3.1 Low energy conformations of 1EGS

The low energy conformations of 1EGS were analysed using the clustering scheme detailed in Sect. 2.4. Seven distinct low energy conformations were identified within a range between −8.73 and −6.31 kcal/mol. The connectivity of the found structures can be observed in Fig. 4. Two structures were considered adjacent, if their RMSD was less than 2.5 Å. The minimum energy conformation is reachable from two distinct conformations 1.7 and 1.6 Å away. Structure C is the structure with the most connections and acts like a travel hub among the low energy conformations. This is easily understood, when analyzing the conformation of structure C. Figure 5c shows structure C to attain a relaxed β conformation, which is in the vicinity of the native conformation (Fig. 5h), the broken β conformation (Fig. 5b) and the partially helical coiled conformation. (Fig. 5c). The higher energy structures are a warped β-sheet (Fig. 5e) and a coiled helical turn similar to structure (C) in Fig. 5f. Due to the strong shear of the highest analyzed energy conformation (Fig. 5f) , many of the low energy conformations are in the vicinity either by increasing the shear leading to a coiled structure or reducing the shear leading to the correct tertiary β-fold.

4 Conclusion

We have shown that our peptide prediction scheme can reliably predict peptide structure using a physical free-energy based approach for peptides of very different structure, including collapsed folds without apparent secondary structure. We could also parallelize this on our volunteer computing network POEM@HOME to allow these predictions in a short time.

The sampling runs did not only elucidate the structure of the peptides themselves, but also their low energy ensemble of structures, which might be involved during the folding and misfolding of the peptide. Here we could show possible fold-paths of the low energy ensemble, the protein might even shift between in its active conformation. This may enable further analysis to establish structure function relationships of the peptide and thereby elucidate their biological activity.

Acknowledgments We thank all of our POEM@HOME volunteers for their continuous support. Furthermore we thank the Carl-Zeiss Foundation for funding on the project Absolute Quality Control of Protein Structures

References

1. D.P. Anderson, in *5th IEEE/ACM International Workshop on Grid Computing*. Boinc: A System for Public-Resource Computing and Storage (2004), p. 4–10

2. N. Assa-Munt, X. Jia, P. Laakkonen, E. Ruoslahti, Solution structures and integrin binding activities of an RGD peptide with two isomers†. Biochemistry **40**(8), 2373–2378 (2001). doi:10.1021/bi002101f
3. M. Banerjee, E. Meyerowitz, C. Huang, S. Mohanty, Probing the conformation and dynamics of allatostatin neuropeptides: a structural model for functional differences. Peptides **29**(3), 375–385 (2008). doi:10.1016/j.peptides.2007.11.016. http://www.sciencedirect.com/science/article/pii/S0196978107004652
4. K.A. Brogden, Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. **3**(3), 238–250 (2005). doi:10.1038/nrmicro1098
5. S.M. Gopal, W. Wenzel, De novo folding of the DNA-binding ATF-2 zinc finger motif in an all-atom free-energy forcefield. Angew. Chem. Int. Ed. **118**(46), 7890–7892 (2006). http://dx.doi.org/10.1002/ange.200603415
6. S.J. Landry, A. Taher, C. Georgopoulos, S.M. van der Vies, Interplay of structure and disorder in cochaperonin mobile loops. Proc. Nat. Acad. Sci. **93**(21), 11, 622–11,627 (1996). http://www.pnas.org/content/93/21/11622.abstract
7. S.J. Russell, T. Blandl, N.J. Skelton, A.G. Cochran, Stability of cyclic-hairpins: asymmetric contributions from side chains of a hydrogen-bonded cross-strand residue pair. J. Am. Chem. Soc. **125**(2), 388–395 (2003). doi:10.1021/ja028075l. http://pubs.acs.org/doi/abs/10.1021/ja028075l. PMID: 12517150
8. A. Verma, W. Wenzel, Protein structure prediction by all-atom free-energy refinement. BMC Struct. Biol. **7**, 12 (2007)
9. A. Verma, W. Wenzel, A free-energy approach for all-atom protein simulation. Biophys. J. 96(9), 3483–3494 (2009). http://linkinghub.elsevier.com/retrieve/pii/S0006349509003877
10. M.R. Yeaman, N.Y. Yount, Mechanisms of antimicrobial peptide action and resistance. Pharmacol. Rev. **55**(1), 27–55 (2003). http://pharmrev.aspetjournals.org/content/55/1/27.abstract