

# Structure prediction of protein segments with low target/template sequence identity based on a reduced protein model

Andrzej Szymoszek and Martin Zacharias, AG Theoretische Biophysik, Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, D-07745 Jena, Germany

## Introduction

Modelling of the 3D-structure of a protein based on sequence similarity to a known protein is limited by the degree of the sequence identity with respect to the template. Frequently, the quality of the target-template sequence alignment is non-uniform along the sequence: parts can be modelled with high confidence, whereas other parts differ strongly from the template. Segments of the target sequence that have no equivalent region in the template structure (insertions or loops) are the most difficult regions to model [1]. They are often larger than small loop regions. Since at atomic resolution the accurate loop prediction is limited to short loops of up to 9 residues [1, 2], one needs to evaluate a large number of possible conformations.

For preselection of possible protein segment 3D topologies, we suggest an application of a reduced protein model containing two pseudo atoms per amino acid residue. Our model allows a very rapid generation of protein segment conformations, compatible with boundaries imposed by those parts of the protein chain, that can be accurately modelled based on the template structure. For the pseudo-atom interactions, a knowledge-based potential has been used. The idea is schematically presented in Fig. 1 and outlined in the following. Preliminary tests of the approach on two small test proteins with known structure will be discussed.

## Methods

Fig. 2 presents a comparison of the reduced structure representation to the 3D atomic resolved structure of the CRO repressor protein. Of two pseudo atoms per each amino acid, one represents a backbone C-alpha atom, and the other one the complete side chain. The equilibrium distance between the two pseudo atoms is calculated with respect to the center of mass of the side chain. Harmonic bond, bond-angle, proper and improper (due to chirality) dihedral angle potentials are assumed. They are based on bond and angle distributions in known protein structures. Non-bonded interactions are described by soft Lennard-Jones energy functions, parameterized according to the concept of inter-residue contact energies, extracted from folded protein structures [3]. Examples of energy functions are presented in Fig.3.

Parts of the structure can be restrained to template positions. Other parts are free to move, or they can be constrained to the predicted secondary or loop structure. Molecular dynamics and constraint energy minimization are used to force the mobile part of the structure into a preset structural topology. Generated segment topologies can be evaluated using non-bonded interactions of the final structures in combination with secondary structure propensities for the segment. Generation of an alternative topology for the present test proteins takes only seconds on a workstation computer.

## Results and Discussion

As test examples, two small proteins of known 3D structures were chosen: CRO repressor protein (ref. P1, [4]) and amino terminal domain of phage 434 repressor (ref. P2, [5]). Their backbone structures are presented in Fig. 4. In all cases, a part of the modelled structure was restrained to the template (known experimental structure). Distance constraints were imposed and the structures energy minimized in order to generate alternative conformations for the mobile protein segment. The following tests were performed:

I Prediction of the position of an alpha-helix conformation in P1 (marked pink, Fig.4)

II Comparison of various topologies in the helical region of P2 (marked pink, Fig.4)

In I, it was assumed that the helix of seven residues can be placed in its experimentally derived site, or shifted 1-3 residues in both directions. Results are presented in Fig. 5. The position as found in the experimental structure was identified as the energetically most favorable topology.

In II, original helix constraints were replaced by various constraints, including beta-sheet, to force part of the structure into alternative possible conformations. The procedure led to structures presented in Fig. 6 and as found for test I selected the experimental segment topology as lowest energy conformation.

## Perspectives

Initial tests on proteins with known structure indicate that the approach might be helpful to preselect likely conformations for a protein segment with low or no sequence identity to a template if the rest of the protein can be modelled accurately. Since the research is still in its early phase, more different tests, including various structural aspects of investigated proteins, should be performed, so that the model can be further validated.

### References:

- [1] M.A.Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F.Melo, A. Sali. *Annu. Rev. Biophys. Biomol. Struct.* 2000, 29:291-325.
- [2] H.W.T. Vlijmen, M. Karplus. *J.Mol.Biol.* 1997, 267:975-1001.
- [3] S. Miyazawa, R.L. Jernigan. *PROTEINS: Struct., Funct. and Gen.* 1999, 34: 49-68.
- [4] D.H. Ohlendorf, D.E. Tronrud, B.W. Matthews. *J. Mol. Biol.* 1998, 280:129-136.
- [5] A. Mondragon, S. Subbiah, S.C. Almo, M. Drottler, S.C. Harrison. *J. Mol. Biol.* 1989, 205:189-200.

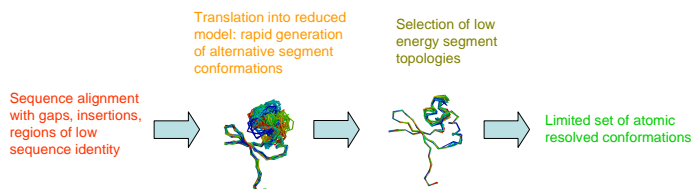


Fig. 1 Schematic presentation of the reduced-model approach to comparative modelling

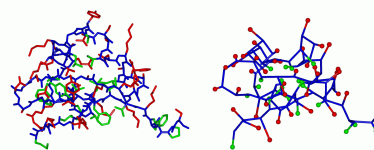


Fig. 2 Reduced model representation of the CRO repressor protein (right panel), compared to atomic resolved structure (left panel). Protein backbone is plotted in blue, side chains of hydrophobic residues are in green, side chains of hydrophilic residues are in red.

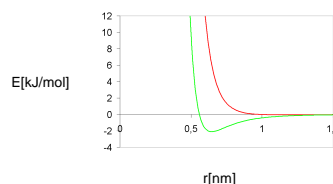


Fig. 3 Potential energy of the LEU-LEU (green) and LYS-LYS (red) pseudo atom interactions. All other pair-potentials are similar to one of the two types presented.

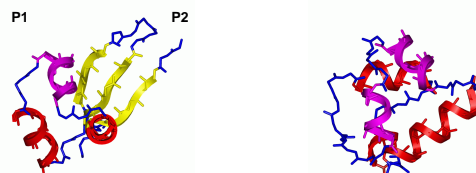


Fig.4 Experimental backbone structures of test proteins P1 and P2. Helical regions of mobile segments are plotted in pink.

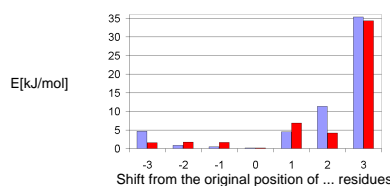


Fig.5 Total (blue) and non-bonded (red) energy of P1 subjected to constraints that shift the second helix (residues 17-24 in experimentally determined structure) by up to +/-3 residues.

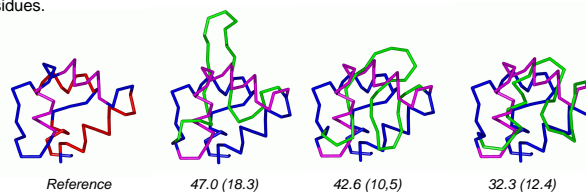


Fig. 6 Structures corresponding to various alternative topologies generated for residues 14-39 in reference structure and green in newly generated conformations of P2. Differences in total energy in kJ/mol with respect to the reference structure are listed (non-bonded energy in brackets).