The last mile of the protein folding problem – a pilgrim's staff and skid-proof boots

Elmar Krieger, Sander B. Nabuurs, Chris.A.E.M. Spronk and Gert Vriend

CMBI, Center for Molecular and Biomolecular Informatics, Radboud University Nijmegen, the Netherlands, www.cmbi.ru.nl
Elmar.Krieger@cmbi.ru.nl, **www.YASARA.org**

Today's energy functions are not able yet to turn a homology model into an experimental-like structure, for example during a molecular dynamics simulation. Here we describe a set of approaches that e.g. allowed to reduce the Cα-RMSD of the T0231 core from 0.96 to 0.79 Å using the programs YASARA and WHAT IF:

1) Self-parameterizing force fields that make optimal use of empirical energy functions.
2) Thousands of parallel simulations combined with predictions for the outcome of these simulations and the ColonyMorphScore to pick out the pearls.
3) Recalculation of NMR template structures using the original experimental restraints if available.

INTRODUCTION

Improving near-native models is currently a major bottle-neck in homology modeling or experimental structure determination at low resolution. Increasingly accurate energy functions are required to walk along the '***last mile of the protein folding problem***'.

RESULTS

Self-parameterizing force fields

Here we provide a pilgrim's staff: self-parameterizing force fields[1], that were obtained from the AMBER force field[2] by simulating complete protein crystals (Fig.1) and iteratively adjusting the force field parameters to minimize the damage done to known structures[3] (Fig.2).

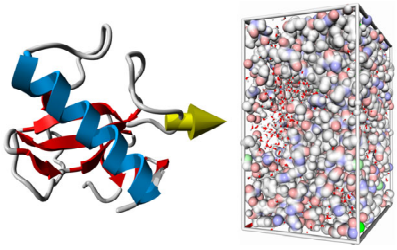

Fig.1: Reconstructed unit cells of high-resolution X-ray structures provide the gold standard.

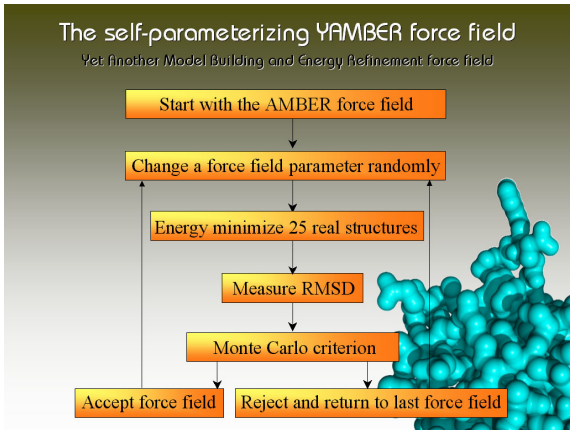

Fig.2: Making optimal use of empirical energy functions by minimizing the damage done to 25 crystal unit cells during an energy minimization.

The resulting YAMBER (www.yasara.org/yamber) and YASARA force fields are then used to run accurate simulations of homology models in aqueous solution (Fig.3).

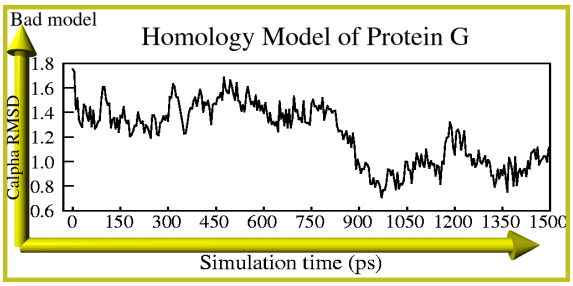

Fig.3: Molecular dynamics simulation of a Protein G homology model and RMSD from the known X-ray structure.

Parallel simulations of models

Additional skid-proof boots are needed to avoid a common pitfall: even with an ideal force field, homology models cannot be expected to always approach the native conformation directly. That is why we run 100 simulations in parallel[4], first in theory using CONCOORD[5] and then in practice with YASARA's new force fields[1,3].

Picking pearls: ColonyMorphScore

To identify those trajectories where the simulation proceeded in the right direction, a sophisticated scoring function based on WHAT IF checks[6] and YASARA energies is used (Fig. 4). False positives with artificially high scores are avoided by considering the number of structural neighbors in the model ensemble[7], similar to an approach described for loops[8].

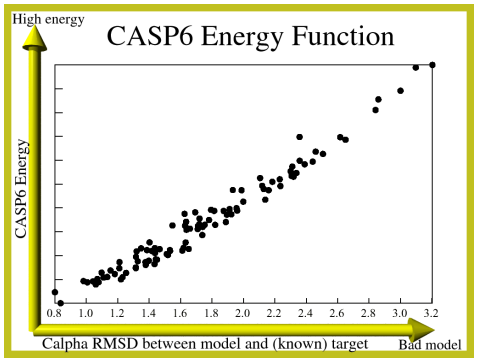

Fig.4: Example of the ColonyMorphScore

T0231: High resolution refinement

Even models very close to the native structure can be improved: The closest template for Target 231 was an NMR structure with 80% sequence identity and 0.96 Å Cα RMSD (excl. the long flexible surface loop (residues 74 to 84) involved in crystal contacts on the right side of Fig. 5). During the refinement, this RMSD could be reduced to 0.79 Å (Fig. 5).

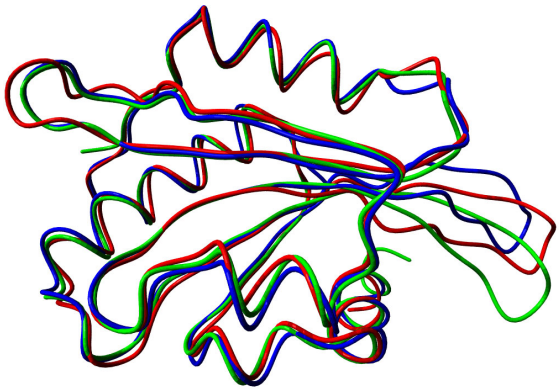

Fig.5: Superposition of template (NMR structure 1V6F, red), model 1 (blue) and true target (1VKK, green).

Improving NMR templates: DRESS

For targets T0221 and T0276, the experimental NMR restraints of the closest templates were available, which allowed us to redetermine the template structures using a standardized NMR refinement procedure in explicit solvent (Fig.6): **DRESS**, the **D**atabase of **RE**fined **S**olution **S**tructures[9]. The true target coordinates are not available yet to judge the impact.

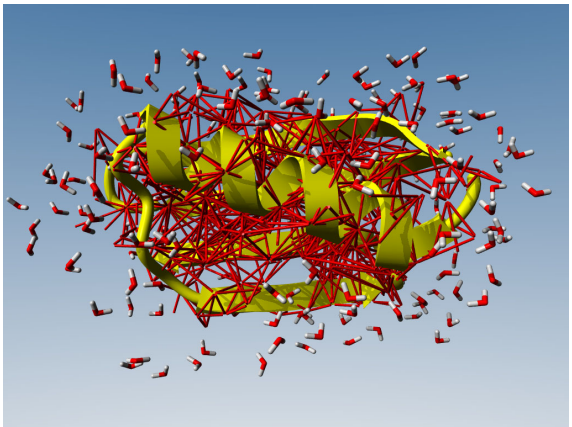

Fig.4: DRESS, database of refined NMR structures.

REFERENCES

1. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. Proteins 2002;47:393-402.
2. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? JComp-Chem 2000;21:1049-1074.
3. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. Making optimal use of empirical energy functions: force field parametrization in crystal space. Proteins 2004;in press.
4. Krieger E, Vriend G. Models@Home: distributed computing in bioinformatics using a screensaver based approach. Bioinformatics 2002;18:315-318.
5. de Groot BL, van Aalten DM, Scheek RM, Amadei A, Vriend G, Berendsen HJ. Prediction of protein conformational freedom from distance constraints. Proteins 1997;29:240-251.
6. Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272-272.
7. Krieger E, Ivankov DN, Finkelstein A, Vriend G. WHAT IF YASARA folds a protein? ; 2002. p A-169.
8. Xiang Z, Honig B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. ProcNatlAcadSciUSA 2002;99:7432-7437.
9. Nabuurs SB, Nederveen AJ, Vranken W, Doreleijers JF, Bonvin AM, Vuister GW, Vriend G, Spronk CA. DRESS: a database of refined solution NMR structures. Proteins 2004;55:483-486.