Computational design of dimeric *de novo* heme-binding helical bundle proteins

Veronica Delsoglio*, Gustav Oberdorfer

Graz University of Technology, Institute of Biochemistry, Austria

INTRODUCTION

The reason why the comprehension of protein biochemistry, biophysics and folding is important, is due to the purpose of creating new protein structures to be able to tackle important challenges in biotechnology. Indeed the amino acid sequence specifies the protein fold and the protein function. The design of protein sequences with altered or novel functions is a tricky aspect of protein engineering. In the last years, the computational methodology has advanced considerably, reaching atomic-level accuracy, and with *de novo* protein design it is now possible to simulate protein sequences and predict which ones exhibit specific properties and desired functions.[1] The *de novo* protein design is performed using the Rosetta software, originally developed at the University of Washington. It includes algorithms for computational modelling and analysis of protein structures.[2] Specifically this work is based on homodimeric bundles of three α-helices – homodimeric 3H5L_2 **A** (H 3H5L_2 **A**), homodimeric 3H5L_2 **B** (H 3H5L_2 **B**), homodimeric 3H5L_2 **C** (H 3H5L_2 **C**). These structures show different amino acid sequences, but are all constituted by two antiparallel untwisted three-helix bundles, with 80-residue helices and an 18-residue repeating unit (5 full circles, or 'layers', until they reach identity) [2]. Moreover, they exhibit four histidine residues, which were designed to coordinate four b-type heme ligands, located on the central plane (Figure 1).

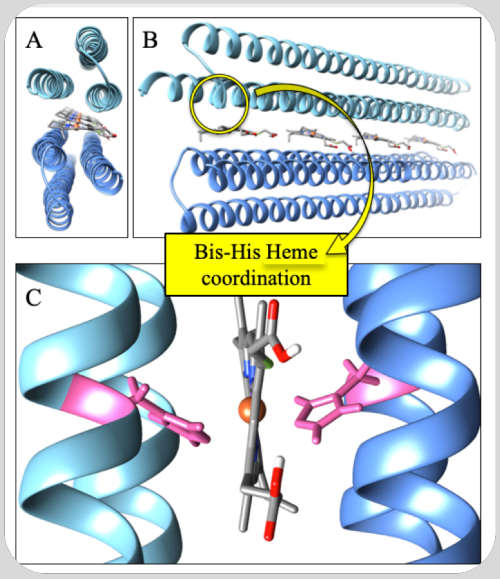
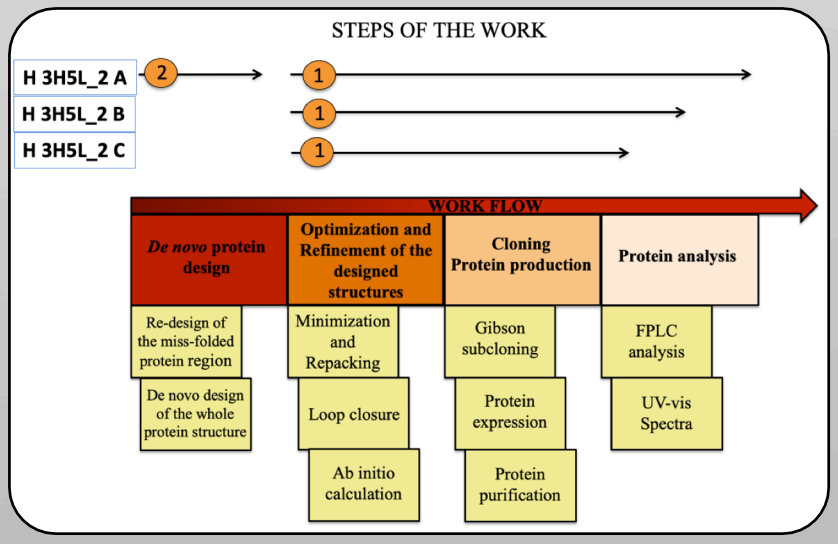


Figure 1. Overall fold of the three homodimeric protein structures H 3H5L_2 **A,** H 3H5L_2 **B,** H 3H5L_2 **C.** **(A)** Side view of the homodimer. **(B)** Front view of the homodimer. **(C)** Heme rings bis-coordination with four different histidine residues located on each first helix of the respective bundles.

METHODS AND RESULTS



Every working cycle is composed of a computational part, where the protein structure is designed and refined using Rosetta software, and a working laboratory one, where the designed constructs are expressed, purified and biochemically and biophysically analyzed.

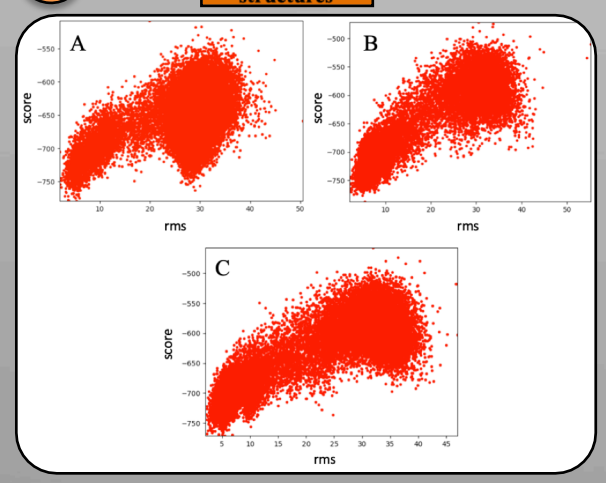


Figure 2. Ab initio calculation results . **(A)** Total-score vs. rmsd plot of H 3H5L_2 **A**. A second population of structures with good total_score values, but different topology than the native one, is present. This means that the optimization work of this structure is not finished yet. The protein sequence regions which fold in a different manner should be redesign. **(B)** Total_score vs. rmsd plot of H 3H5L_2 **B. (C)** Total-score vs. rmsd plot of H 3H5L_2 **C.**

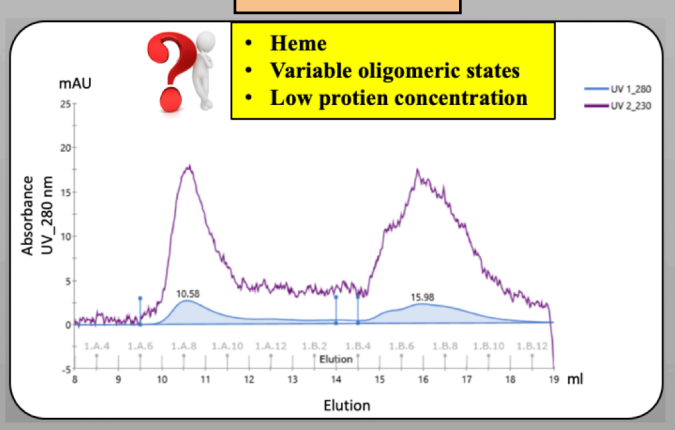


Figure 3. SEC chromatography chromatogram on FPLC of the protein elution. SEC column used: Superdex 200 Increase 10/300 GL. Peak at elution volume 15.98 ml corresponds to a protein molecular weight range 29.000 Da – 44.000 Da (homodimer).

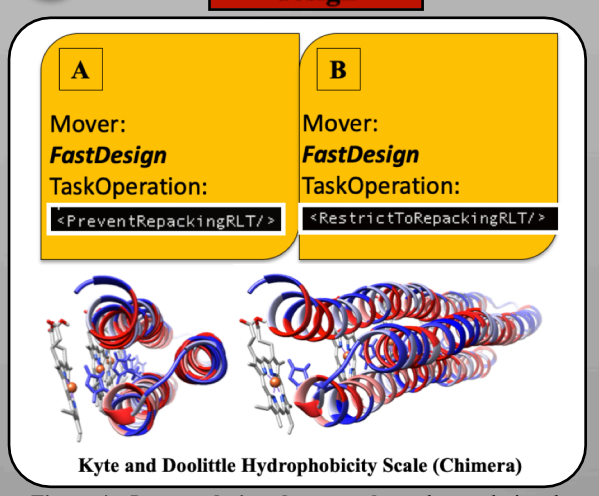


Figure 4. ***De novo*** **designed protocols** used to re-design the protein constructs. The output structures are analyzed using PyRosetta scripts and Chimera tools (etc. kdHydrophobicity).

FUTURE PERSPECTIVES

- Re-design of the three protein constructs writing and testing different scripts;
- Express, produce and analyze the new re-designed proteins.

REFERENCES

[1] P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles, *Science* **346**, 481-484 (2014).

[2] https://www.rosettacommons.org

Contact: veronica.delsoglio@tugraz.at