In silico design of new proteins

Universitat Autònoma de Barcelona

Miguel Correa Marrero, Universitat Autónoma de Barcelona

Introduction

Computational protein design (CPD) is a relatively recent appproach to rational protein design that deals with the inverse folding problem (fig. 1). Its objective is to automate protein design using algorithms that can be guided by different criteria, such as a physical chemistry models that attempt to explain protein folding.

Fig 1. The inverse folding problem. Given a structure, which sequences fold into this structure?

Fundamental concepts

Target structure. The structure for which an adequate sequence is searched for. It is usually fixed (fixed-backbone approach).
Protein design cycle. A popular design strategy (fig. 2).

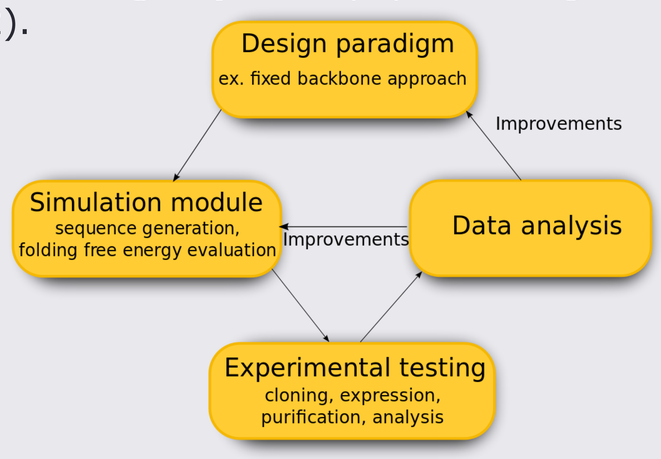

Fig. 2. The protein design cycle is used to empirically improve the design process.

Energy functions. They give an estimation of the folding free energy of possible solutions. Some are based on statistical information, and others on physico-chemical models.
Search algorithms. For a sequence with n residues and r allowed rotamers, there are r^n possible sequences in sequence space. Different strategies to not need to sample all of sequence space exist (fig. 3).

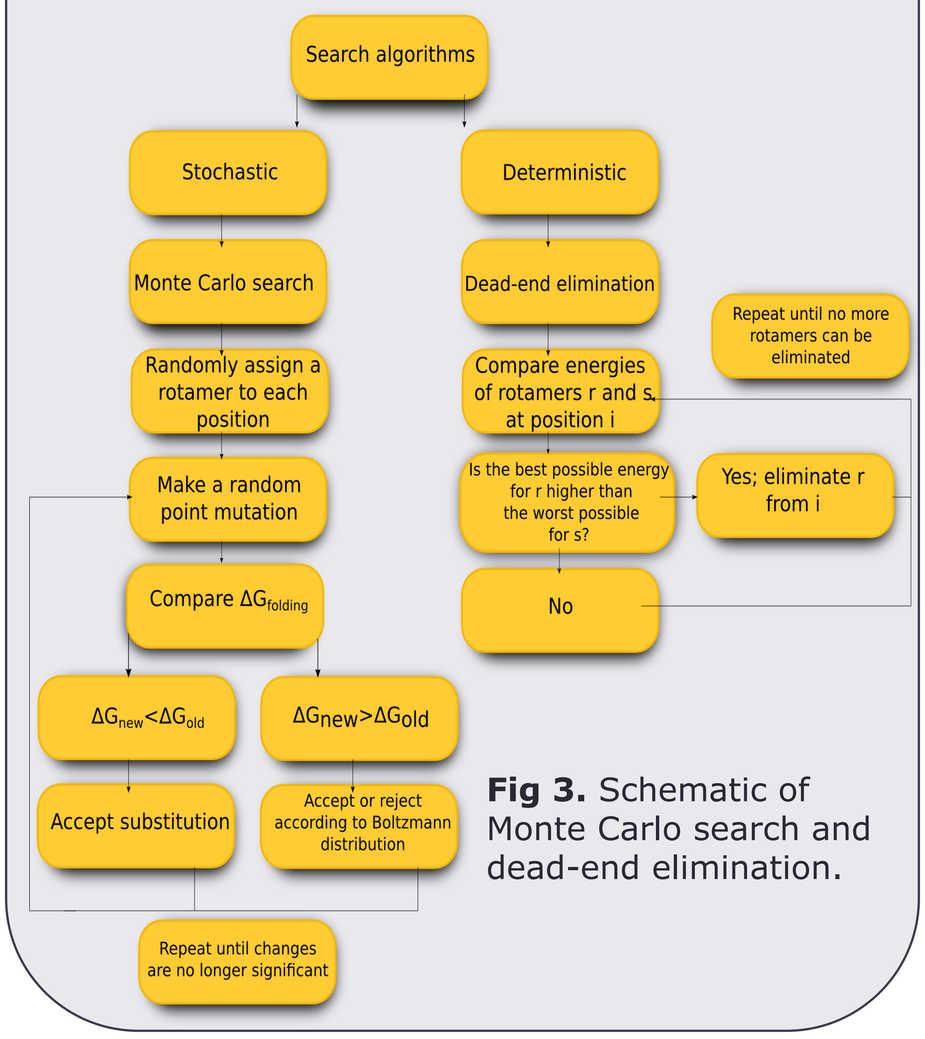

Fig 3. Schematic of Monte Carlo search and dead-end elimination.

Design of Top7

Top7 is a protein designed from scratch with a novel topology [2].

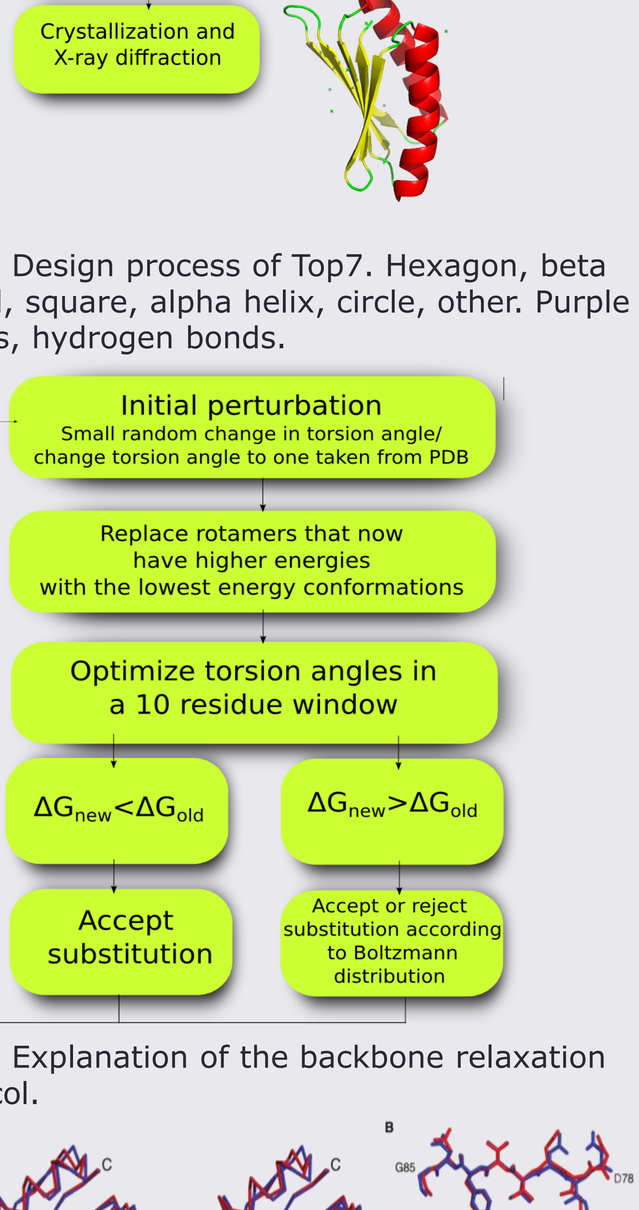

Fig 4. Design process of Top7. Hexagon, beta strand, square, alpha helix, circle, other. Purple arrows, hydrogen bonds.

Fig 5. Explanation of the backbone relaxation protocol.

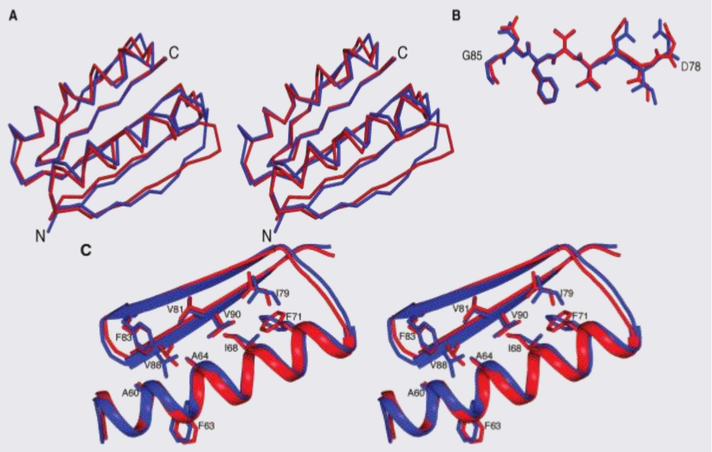

Fig 6. Comparison of theoretical model (blue) to solved structure (red). **A**, backbones overlay. Backbone RMSD: 1,17 angstroms. **B**, C-terminal extremes overlay. All-atom RMSD in this part: 0,79 angstroms. **C**, ribbon model with superposable side chains.

The design of Top7:
- Validates the acuraccy of physical energy functions
- Demonstrates the existence of stable folds not present in nature
- Shows that successful design does not always require taking competing structures into account

Function design examples

A dynamite sensing protein. Looger et al. [3] used *E. coli* proteins as scaffolds to create binding sites for different ligands, so they could be used as fluorescence emitting biosensors.

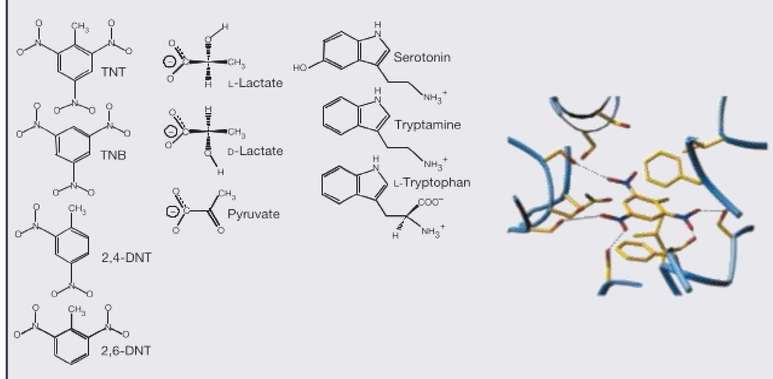

Fig 7. To the left, the different ligands for which binding sites were designed. Note their chemical diversity. To the right, the binding site for TNT.

An endonuclease with altered specificity. Ashworth et al [1]. redesigned an endonuclease to hydrolize a slightly different sequence.

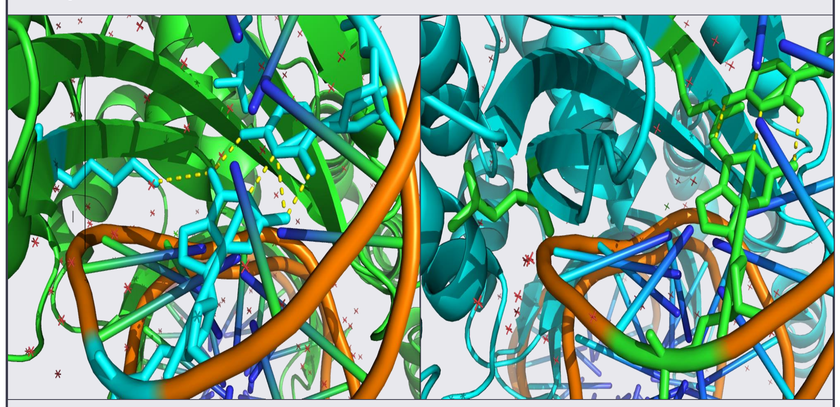

Fig 8. To the left, wild type MsoI bound to wild type DNA. To the right, MsoI redesigned (K28L and T83R) to bind a sequence where a G-C pair has been mutated to C-G. R83 bonds to the introduced G; L28 decreases specificity for WT DNA.

Concluding remarks

Although CPD has seen major developments since its inception, there are still significant issues [4]:
- The traditional fixed-backbone approach causes the rejection of sequences compatible with the target structure. It is necessary to introduce backbone flexibility into designs.
- Negative design strategies to take competing structures into account need to be developed.
- Characterization of designs is slow. High-throughput systems are being developed.
- Computationally designed enzymes are usually inefficient.
- Solvent modelling has much room for improvement

CPD will probably cause a huge impact in biology and materials science, and also the birth of a CPD industry.

References

[1] Ashworth J. et al.(2006), *Nature* **441**(7093), 656--659.
[2] Kuhlman, B et al. (2003), *Science* **302**(5649), 1364--1368.
[3] Looger, L. L (2003), *Nature* **423**(6936), 185--190.
[4] Suárez, M. & Jaramillo, A. (2009), *Journal of the Royal Society Interface* **6**(Suppl 4), S477--S491.