Design of a Novel Globular Protein Fold with Atomic-Level Accuracy

Brian Kuhlman, Gautam Dantas, Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, David Baker

Presented by Kate Stafford

4 May 05

Protein Structure Prediction Methods

- *Ab initio* structure prediction: uses quantum mechanics to model atom by atom
  - Relatively inaccurate, very slow
  - The only option for totally novel structures
- Homology modeling: relies on sequence alignment with a template of known structure
  - More accurate, uses less computer time
  - Very dependent on quality of alignment, which is dependent on sequence identity
  - Impossible if no similar sequences have solved structures
- Tertiary structure is very difficult, but good algorithms for secondary structure prediction exist

Protein Design: Why?

- Designing a protein *de novo* is a good test of current modeling methods, especially energy functions and solvent models
- New protein designs could lead to the development of novel catalytic functionality

Top7: A Totally Artificial Protein

- Top7 is a 93-residue α/β protein whose topology and sequence are not found in the PDB or in SCOP
  - This means Top7's fold is in a region of conformational space not explored by (currently known) biological structures
- Top7's crystal structure has a backbone RMSD of 1.7Å compared to the designed model
  - For comparison: <3 is excellent for a homology-based model of a small- to medium-sized protein with high sequence identity to the template

Figure removed due to copyright reasons. Please see:
Kuhlman B., G. Dantas, G. C.Ireton, G. Varani, B. L.Stoddard, and D. Baker. "Design of a novel globular protein fold with atomic-level accuracy." *Science* 302, no. 5649 (Nov 21, 2003): 1364-8.

Blue = model

Red = crystal structure

Top7 Design Protocol I: Starting Structures

- The target topology was selected first, specifically because it was not known in nature
- For a given topology, geometric constraints must be identified (e.g., hydrogen bonding in sheets and helices)
- A set of 3D models satisfying the constraints was generated by combining 3- and 9-residue PDB fragments of the appropriate predicted secondary structure

→ This gave 172 backbone structures differing from each other by 2-3Å (not very much)

Figure removed due to copyright reasons. Please see: Kuhlman B., G. Dantas, G. C.Ireton, G. Varani, B. L.Stoddard, and D. Baker. "Design of a novel globular protein fold with atomic-level accuracy." *Science* 302, no. 5649 (Nov 21, 2003): 1364-8.

Target topology with final Top7 sequence. Arrows represent hydrogen bonds. Pink hexagons are sheets, blue squares are helices, green circles are loops.

Top7 Design Protocol II: Starting Sequences

- A starting sequence was designed for each structure by searching an amino acid rotamer library
- All amino acids except cysteine were allowed at 71 positions
- The remaining 22 positions are on the sheet surfaces and were restricted to polar amino acids
- Unsurprisingly, these starting models had a high free energy

Top7 Design Protocol III: Optimization

- Backbone optimization: use Monte Carlo minimization to alter structure so that it better accommodates existing sequence
  1. Perturb 1-5 randomly selected torsion angles
  2. Optimize sterics by cycling through rotamers
  3. Optimize torsion angles within 10 residues
- Sequence optimization: rotamers of new amino acids explored for low-energy side-chain packing
- 15 alternating cycles of each gives a final energy-minimized structure

Top7 Characteristics

- Final backbone model is only 1.1Å different from starting model
- Only 31% of residues are retained from initial to final design
- The synthesized protein is thermally stable to 98C
- Many side chains are superimposable (for comparison: 50% of native residue-residue contacts are reproduced in a decent model)

Figure removed due to copyright reasons. Please see:
Kuhlman B., G. Dantas, G. C.Ireton, G. Varani, B. L.Stoddard, and D. Baker. "Design of a novel globular protein fold with atomic-level accuracy." *Science* 302, no. 5649 (Nov 21, 2003): 1364-8.

Blue = model

Red = crystal structure

Top7 Structure

Figure removed due to copyright reasons. Please see:

Kuhlman B., G. Dantas, G. C.Ireton, G. Varani, B. L.Stoddard, and D. Baker. "Design of a novel globular protein fold with atomic-level accuracy." *Science* 302, no. 5649 (Nov 21, 2003): 1364-8.