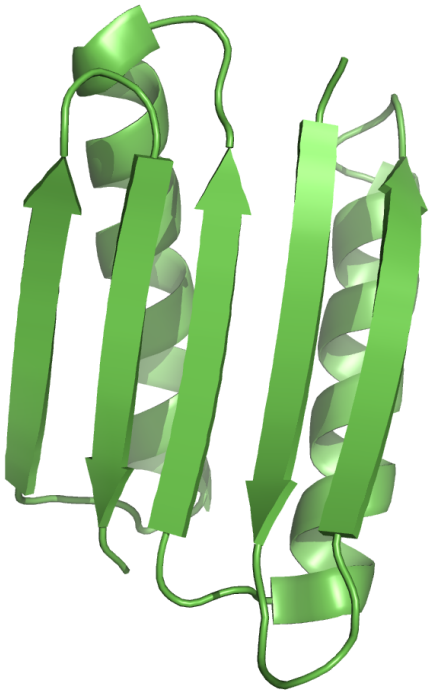
Computational Protein Stability Prediction

Kyle Barlow

PUBS 2015

Why use computation for protein design?

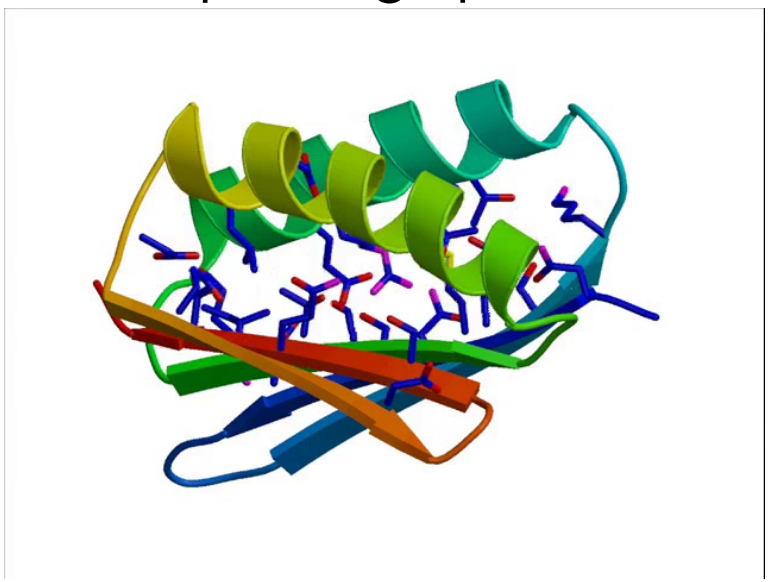
- Computation allows for a range of design problems to be tackled that would otherwise be intractable:
  - Design of new protein structures
  - Novel enzymatic activity
  - **Predict stability of mutants**



Top7 — Kuhlman et al. [2]

How does computational protein design in Rosetta search for low-energy sequences?

Top7 design process



Kuhlman et al. [2]

Conformational sampling

- side chain rotamers
- fixed backbone

Energy function

- physical terms
- atomic packing
- explicit hydrogen bonding
- implicit solvation model

Why is computational design hard?

Sequence search space and structure space are both *very* large. Computational design in Rosetta attempts to reduce this search space through use of efficient sampling and scoring.

Conformational sampling

- Use of predefined side chain rotamers in a rotamer library reduces side chain design search space
- Using a starting scaffold structure (fixed backbone) for design means you can start with a known sequence and matching structure

Energy function (scoring)

- Pairwise score function enables faster score calculations (most scores don't need to be recalculated after a rotamer change)
- Implicit solvation model simplifies search space

You have been doing comprehensive "design" of ubiquitin

- Usually, in protein design, you are interested in finding a low-energy, stable sequence that will fold into your desired structure.
- You have been making mutations to ubiquitin, and computation can provide you with predicted structures (and their predicted stabilities) for your "designs".
- The Rosetta protocol for this: $\Delta\Delta G$

The effect of a mutant on a protein's stability can be expressed as a $\Delta\Delta G$ value

Definitions:

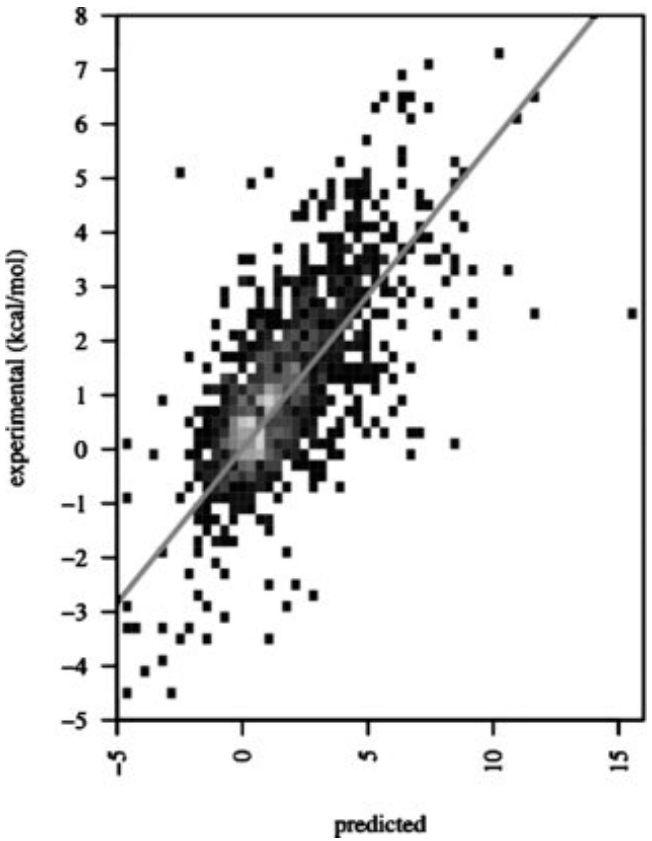
- G — Gibb's free energy
- ΔG – Change in free energy between the unfolded and folded states of a protein

$$\Delta G_{folding} = G_{native\ state} - G_{unfolded\ state} \tag{1}$$

- $\Delta\Delta G$ – Change in free energy of folding caused by a mutation

$$\Delta\Delta G_{I44F} = \Delta G_{folding\ F44} - \Delta G_{folding\ I44} \tag{2}$$

Enter computation: Rosetta $\Delta\Delta G$ Prediction



- Rosetta's predicted $\Delta\Delta G$ value for a single point mutant (starting from a known crystal structure) correlates well with experimental data
- The $\Delta\Delta G$ protocol uses the Rosetta energy function, along with sampling algorithms such as an all side chain "packer" (some backbone minimization is included in this protcol)

Kellogg et al. [1]

Molecular and systems-level constraints acting on proteins

- Maintain protein stability
- Solubility
- Folding kinetics, lifetime
- Ability to switch between conformations (recognition, molecular machines)
- Binding to desired (multiple) partners
- Avoiding other (undesired) partners

Original slide: Tanja Kortemme

Your task: interpret Rosetta's $\Delta\Delta G$ predictions for ubiquitin in the context of your sequencing data

- **Maintain protein stability** — *it should be possible to account for this effect (to some degree) using Rosetta's predicted $\Delta\Delta G$ values*
- Solubility
- Folding kinetics, lifetime
- Ability to switch between conformations (recognition, molecular machines)
- **Binding to desired (multiple) partners** - *Rosetta can also generate binding affinity predictions*
- Avoiding other (undesired) partners

Interface binding energy $\Delta\Delta G$ equation

- The predicted scores of the wildtype or mutant and complex or individual protein chains are combined to create the final $\Delta\Delta G$ score:

$$\begin{aligned}\Delta\Delta G_{bind} &= \Delta G^{MUT}_{bind} - \Delta G^{WT}_{bind} \\ &= (\Delta G^{MUT}_{complex} - \Delta G^{MUT}_{partnerA} - \Delta G^{MUT}_{partnerB}) \\ &\quad - (\Delta G^{WT}_{complex} - \Delta G^{WT}_{partnerA} - \Delta G^{WT}_{partnerB})\end{aligned} \tag{3}$$

Data we will give you

We will provide you with the following Rosetta $\Delta\Delta G$ predictions (all starting from the yeast UBQ sequence):

- Stability - Ubiquitin monomer
- Binding affinity - Ubiquitin in complex with binding partners
- Binding affinity - Ubiqutin dimers

References

[1] Elizabeth H. Kellogg, Andrew Leaver-Fay, and David Baker. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. 79(3):830–838. ISSN 1097-0134. doi: 10.1002/prot.22921. URL http://onlinelibrary.wiley.com/doi/10.1002/prot.22921/abstract.

[2] Brian Kuhlman, Gautam Dantas, Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, and David Baker. Design of a novel globular protein fold with atomic-level accuracy. 302(5649):1364–1368. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1089427. URL http://www.sciencemag.org/content/302/5649/1364.