Improved High Throughput Free Energy Calculations using Rosetta and Cyrus Bench

Brandon Frenz[1], Frank DiMaio[2], Indigo King[1], Steven Lewis[1], Hahnbeom Park[2], Yuan Liu[2], David Pearlman[1], Yifan Song[1]

[1]Cyrus Biotechnology, 500 Union St, Suite 320, Seattle, WA 98101

[2]Department of Biochemistry, University of Washington, Seattle, Washington 98195

Poster presented at the November 2018 Computational Drug Development for Biologics Summit (currently unpublished)

Abstract

Note: This slide deck contains just a summary and key results; a full publication is in preparation

- A variety of approaches have been developed for approximating the free energy differences arising from protein sequence mutations on stability and affinity.
- Earlier Rosetta ΔΔG method relies upon conformational search in rotamer space
  - Capable of providing useful results and good throughput.
- We describe an improvement to the original approach (Cartesian ΔΔG)
  - Extends sampling to full Cartesian space, and uses improved Rosetta energy function
- Cartesian ΔΔG improves on original Rosetta approach
  - Particularly for mutations which result in a change in the net charge of the system.
- Existing benchmark sets used to calibrate/validate various free energy predictors flawed
  - Massively over-represent of X→Ala mutations
- We propose a mutation-type-normalized data set
  - Better reflects the application scope of these free energy predictors
  - Further demonstrates the improvement of ΔΔG II over previous methods.

New Cartesian ΔΔG improves rank ordering

Mutation Type	Pearson's R		Predictive Index	
	New ΔΔG	Old ΔΔG	New ΔΔG	Old ΔΔG
small to large	0.50	0.53	0.56	0.59
large to small	0.61	0.53	0.71	0.65
positive to negative	0.62	0.53	0.68	0.45
negative to positive	0.26	0.24	0.13	-0.02
negative to hydrophobic	0.53	0.21	0.58	0.24
hydrophobic to negative	0.74	0.80	0.82	0.81
positive to hydrophobic	0.31	0.29	0.33	0.42
hydrophobic to positive	0.53	0.48	0.71	0.60
non-charged polar to positive	0.28	0.39	0.38	0.34
positive to non-charged polar	0.54	0.30	0.75	0.57
non-charged polar to negative	0.48	0.69	0.57	0.70
negative to non-charged polar	0.35	0.18	0.37	0.08
non-charged polar to hydrophobic	0.70	0.75	0.63	0.69
hydrophobic to non-charged polar	0.54	0.29	0.57	0.32
non-charged polar to non-charged polar	0.51	0.58	0.72	0.74
hydrophobic to hydrophobic	0.46	0.54	0.59	0.65
charge to charge	0.39	0.26	0.59	0.18
involves cysteine	0.23	0.09	0.27	0.15
involves proline	0.45	0.08	0.48	0.55
same size	0.40	0.29	0.48	0.31
everything	0.46	0.23	0.59	0.49

Values calculated for 767 data points in renormalized ProTherm database

Predictive Index (PI)‡:
Reflects: How good is method at rank ordering vs. experiment?

PI = +1: Perfect ordering
-1: Perfectly wrong ordering
0: Random (flip a coin)

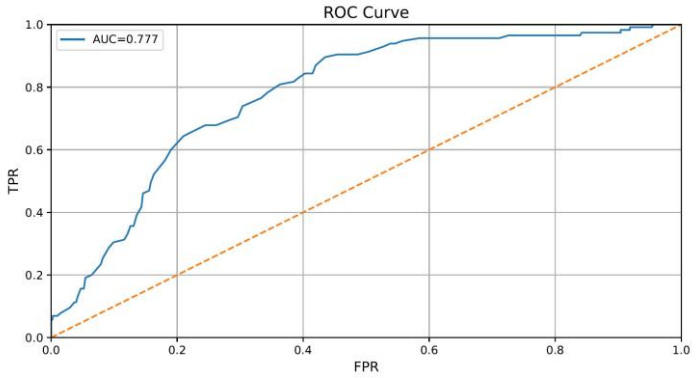
$$PI = \sum_{j>i} \sum_{i} w_{ij} C_{ij} / \sum_{j>i} \sum_{i} w_{ij}$$

$$w_{ij} = \left| E(j) - E(i) \right| \qquad E(i) = \text{expt'l binding energy of cmpd i}$$

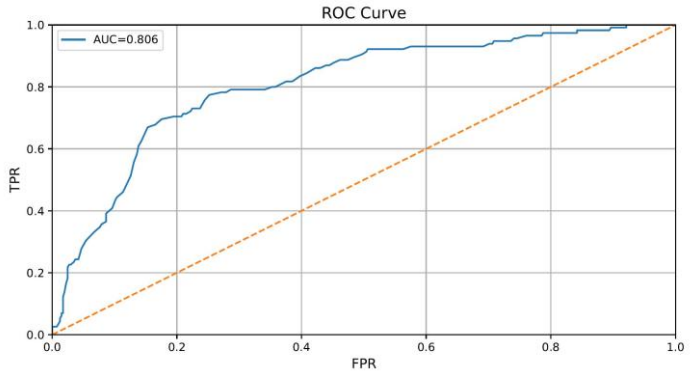
$$C_{ij} = \begin{cases} +1, \text{ if theoretical ranking cmpds (i,j) same as experiment} \\ -1, \text{ if theoretical ranking cmpds (i,j) opposite experiment} \end{cases}$$

‡Pearlman & Charifson (2001) J. Med. Chem. 44 3417

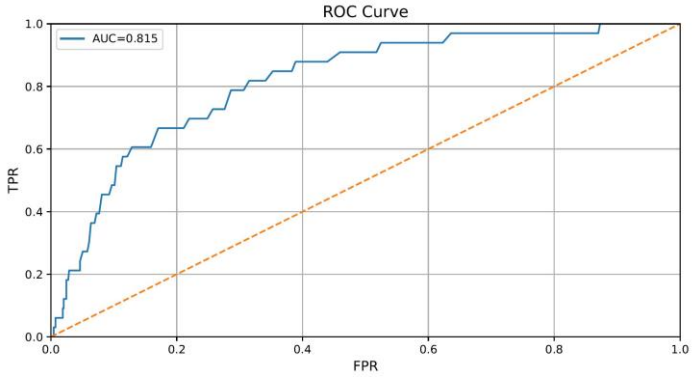
New ΔΔG improves categorization



Old Rosetta ΔΔG
Renormalized ProTherm database
AUC = 0.78



New Rosetta ΔΔG
Renormalized ProTherm database
AUC = 0.81



New Rosetta ΔΔG
"FoldX" database‡
AUC = 0.82

ROC (Receiver Operator Characteristic) plots

- A ROC plot reflects the ability of a predictor to differentiate positive and negative events
  - Plot the True Positive rate on the y axis (Sensitivity) vs. the False Positive rate on the x axis (1-Specificity)
  - A positive event is a mutation with an experimentally measured ΔΔG<0
  - A negative event is a mutation with ΔΔG>0
- The integrated area under the curve (AUC) allows numerical comparison (AUC=0.5 is random)
  - The greater the area under the curve, the stronger the method is for classification
- Plots shown calculated after removing points where |ΔΔG(experimental)| < 0.5 (i.e. ambiguous in light of experimental error)
  - Renormalized ProThem database, 507 total points were used, with 115 < 0 experimentally
  - FoldX database, 700 total points were used, with 33< 0 kcal/mol experimentally

‡R Guerois, JE Nielsen, L Serrano (2002) J. Mol. Biol. 320 (2002) 369–387

Conclusions

- The new Cartesian ΔΔG method performs better overall than the widely-used earlier Rosetta approach
  - Better at binary rank ordering ("is mutation ***i*** better than mutation ***j***?") as determined by Predictive Index
  - Better at classification ("is this a favorable mutation?") as determined by ROC/AUC analysis

- The data sets that have been used to parameterize and benchmark earlier ΔΔG predictors are inherently biased
  - Based on results with our new Cartesian ΔΔG predictor, we expect some of these approaches will look considerably worse if benchmarked against a normalized data set