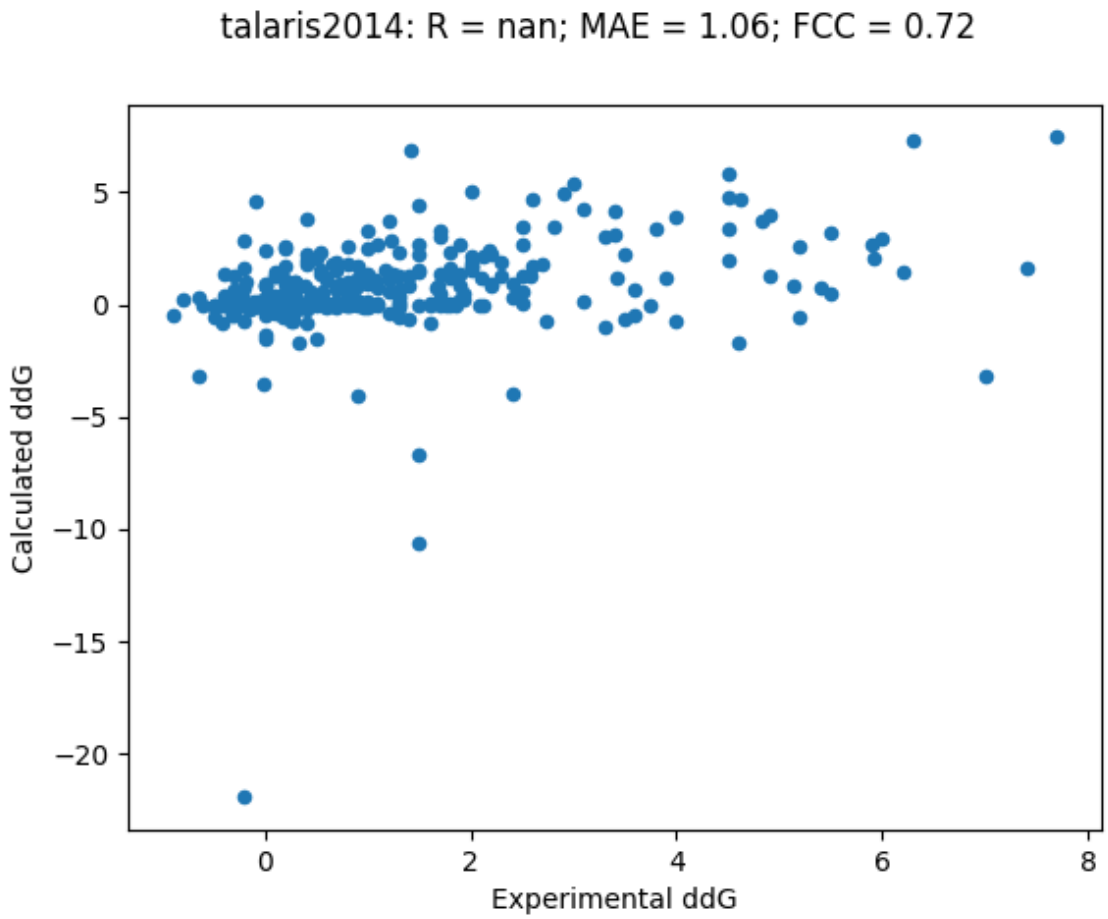
Scientific test: ddg_ala_scan

FAILURES

None

RESULTS



AUTHOR AND DATE

Set up by Ajasja Ljubetic (ajasja.ljubetic@gmail.com), Baker lab, Sept. 2019

This is a port of the Kortemme lab benchamrk (https://github.com/Kortemme-Lab/ddg).

Original benchmark done by Shane Connor, Kyle Barlows, Andrew Leaver-Fay, Tanja Kortemme & David Baker.

Metrics used in evaluation

R: correlation coefficient (>.33)

MAE: Mean Absolute Error (<1.1)

FCC: Fraction Correctly Classifed (>.72)

PURPOSE OF THE TEST

The benchmark test the correlation between predicted and experimental ddG upon a mutation from a native residue to Ala.

The alanine scanning protocol avoids any perturbation of the backbone or side chains, other than the residue being mutated, which is placed into a low-energy rotamer using the Rosetta â€œpackerâ€. This minimal perturbation relies on the fact that the overall protein structure is unlikely to change much after a single point mutation to alanine, making the input crystal structure a good approximation for the mutant structure.

As the alanine scanning protocol does not perturb the protein backbone or side chains (other than the mutant residue), this protocol is not suitable for use on mutations outside of the interface. A mutation outside of the interface will result in a negligible change in total score without the use of a more intensive sampling protocol.

As in Î”Î”G, the metrics used to measure success in this benchmark are: i) the linear correlation (Pearson coefficient) between experimental and predicted values; ii) the mean absolute error (MAE) of same; and iii) the FCC (fraction correctly classifed). FCC is stability classification accuracy, which measures whether a mutation was correctly predicted to be stabilizing, destabilizing, or neutral.

BENCHMARK DATASET

The benchmark is comprised of experimental data for 381 mutations. The datasets are taken from the following publications (PubMed IDs are specified):

7504735, 9571026, 10970748, 9050852, 1281426, 2479414, 9480775, 8494892, 7739054, 9425068, 8784199, 7654692,

10678837, 10880432, 8332602, 10452608, 2402498, 9500785,

11123892, 9878445, 9579662, 8703938, 8263942, 9609690,

10338006

The input files were creating using the instructions found here: https://github.com/Kortemme-Lab/ddg/tree/master/protocols/alanine-scanning

No minimisation is done on the input structures.

This benchmark includes:

a previously published set of alanine mutations in 19 different protein-protein interfaces with known crystal structures (see Kortemme & Baker, 2002);

scripts to run a new RosettaScripts protocol which has been designed to emulate the protocol described in Kortemme & Baker (2002);

an analysis script that output the metrics used for analysis. The script also outputs a scatterplot plotting experimental Î”Î”G values against predicted values (in whichever scoring unit is used by the protocol);

PROTOCOL

The protocol uses Rosetta scripts to change a residue from it's native type to an Ala. The ddG of this mutation is calculated.

There is no minimization/relaxation step of the entire structure.

PERFORMANCE METRICS

The performance metrics are the the correlation coefficient (R) between the experimental and calculated ddG values. The cutoffs were chosen so that the benchmark passed in 74cba0b67deb889d2ae4e4cf519fba9ae9210 commit of Rosetta.

KEY RESULTS

The correlation is compared to experimental data. The correlation should be better than 0.33 for all the energy score terms.

DEFINITIONS AND COMMENTS

LIMITATIONS

Some experimental values are listed as >4. These are currently ignored.

There is no minimization/relax after the mutation is performed. Adding a minimization and repacking might improve the correlation.

Also currently only talaris2014 is tested.

Technically resfiles are not needed (the XML script could be rewritten to take the position IDs directly.)

REFERENCES

Kortemme, T, Baker, D. A simple physical model for binding energy hot spots in proteinâ €“protein complexes. Proc Natl Acad Sci U S A. 2002 Oct 29;99(22):14116-21. Epub 2002 Oct 15. doi: 10.1073/pnas.202485799.

Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. Sci STKE. 2004 Feb 3;2004(219):pl2. doi: 10.1126/stke.2192004pl2.