# Self-consistency test reveals systematic bias in programs for prediction destabilization upon mutation

Dmitry N. Ivankov

*Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), 88 Dr. Aiguader, 08003 Barcelona, Spain,* ivankov13@gmail.com

Computational prediction of the effect of amino acid substitutions on protein stability is utilized by researchers in many fields. Specifically, researchers may be interested in exploring the effect of combination of different mutations, such as when considering the maintenance of protein stability in the course of accumulation of substitutions in evolution. Such programs, while relatively inaccurate, are not known to provide systematically biased results. We explored the suitability of using two of the most popular algorithms, FoldX (1, 2) and I-Mutant (3, 4), for prediction of substitutions on structure.

We devised a self-consistency test that queries the reciprocity of the prediction of amino acid substitutions. Unbiased algorithms should, on average, predict that the effect of an amino acid substitution is equal but opposite to the reverse substitution. To test this prediction, we selected among the available crystal structures those that differ from each other by only a few substitutions. We found that both of the tested algorithms have an inherent bias, whereby for many instances the effect of the forward and the reverse substitution was predicted to be substantially different in magnitude. The systematic bias for single mutants was $0.63 \pm 0.04$ kcal/mol for FoldX and $0.59 \pm 0.43$ kcal/mol for I-Mutant.

We hypothesize that FoldX displays this bias because the underlying algorithm does not change the backbone of the structure while I-Mutant is influenced by the content of the training set, which likely includes many more damaging mutations than benign one. Authors seeking to use these and, possibly, other algorithms for prediction of the effect of amino acid substitutions should be aware of the inherent bias described here.

The author thanks Dinara Usmanova, Natalya Bogatyreva, Joan Ariño Bernad, Anastasiya Gorshkova, Aleksandra Eremina, German Kanevskiy, Lyubov Lonishin, Alexander Meister,

Alisa Yakupova, and Fyodor Kondrashov for making the research. The work was started at the School of Molecular and Theoretical Biology supported by the Dynasty Foundation. The study was in part funded by HHMI International Early Career Scientist Program (55007424), the MINECO (BFU2015-68723-P), Spanish Ministry of Economy and Competitiveness Centro de Excelencia Severo Ochoa 2013-2017 grant (SEV-2012-0208), Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat's AGAUR program (2014 SGR 0974), and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013, ERC grant agreement 335980_EinME).

1. R.Guerois, J.E.Nielsen, L.Serrano (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations, *J Mol Biol*, **320:**369–387.

2. J.Schymkowitz, J.Borg, F.Stricher, R.Nys, F.Rousseau, L.Serrano (2005) The FoldX web server: an online force field, *Nucleic Acids Res*, **33(Web Server issue):**W382-388.

3. E.Capriotti, P.Fariselli, R.Calabrese, R.Casadio (2005) Predicting protein stability changes from sequences using support vector machines, *Bioinformatics*, **21 Suppl 2:**ii54-58.

4. E.Capriotti, P.Fariselli, R.Casadio (2005) I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res*, **33(Suppl. 2):**306–10.