

# Protein Folding, Misfolding and Diseases: The I-Mutant Suite

Calabrese R, Capriotti E\*, Fariselli P, Martelli PL and Casadio R

Laboratory of Biocomputing, CIRB/CIG/Department of Biology, University of Bologna, via Irnerio 42,  
40126 Bologna, Italy.

\* now at: Structural Genomics Unit, Department of Bioinformatics,  
Centro de Investigacion Principe Felipe, Autopista del Saler 16,  
46013 Valencia, Spain

## Motivation

Motivation A basic question of protein structural studies is to which extent mutations affect the stability. This question may be addressed starting from sequence and/or from structure. In proteomics and genomics studies prediction of protein stability free energy change ( $\Delta G$ ) upon single point mutation may also help the annotation process. The experimental  $\Delta G$  values are affected by uncertainty as measured by standard deviations. Most of the  $\Delta G$  values are nearly zero (about 32% of the  $\Delta G$  data set ranges from -0.5 to 0.5 kcal/mole) and both the value and sign of  $\Delta G$  may be either positive or negative for the same mutation blurring the relationship among mutations and expected  $\Delta G$  value.

## Methods

Methods In order to overcome this problem we describe a new predictor that discriminates between 3 mutation classes: destabilizing mutations ( $\Delta G < -1.0$  kcal/mol), stabilizing mutations ( $\Delta G > 1.0$  kcal/mole) and neutral mutations ( $-1.0 \leq \Delta G \leq 1.0$  kcal/mole). We recently developed the I-Mutant Suite that incorporates different machine learning algorithms to predict the effects of non synonymous Single Nucleotide Polymorphism (nsSNPs) in coding regions. The effect of nsSNPs can be studied considering two main aspects: folding stability and protein functionality loss (affecting human health). I-Mutant Suite provides predictions for both aspects, integrating and improving previously developed predictors.

## Results

Results In particular the new version of I-Mutant- $\Delta G$  predicts both the sign and the value of the free energy change ( $\Delta G$ ) upon single point protein mutations, scoring in the classification task with a 80% accuracy and with a correlation coefficient of 0.70 to experimentally detected  $\Delta G$  values, when structure is adopted as input. Furthermore, to better grading the stability predictions, I-Mutant3.0 also provides the capability of discriminating between 3 classes, namely destabilizing, neutral and stabilizing ( $\Delta G > 0.5$  Kcal/mol), scoring with an accuracy of 64% and 68% when protein sequence and structural information are considered, respectively. The I-Mutant Suite also includes I-Mutant-Disease, a new SVM-based predictor able to discriminate between neutral and disease-related polymorphisms. The predictor is endowed with 76% overall accuracy and with a correlation coefficient of 0.52. I-Mutant Suite is the first web server that integrates in a unique framework the predictions of the folding free energy changes and disease-related effects upon single point protein mutation. Our method improves the quality of the prediction of the free energy change due to single point protein mutations by adopting a hypothesis of thermodynamic reversibility of the existing experimental data. By this we both recast the thermodynamic symmetry of the problem and balance the distribution of the available experimental measurements of free energy changes. This eliminates possible overestimations of the previously described methods trained on an unbalanced data set comprising a number of destabilizing mutations higher than stabilizing ones.

## Supplementary informations

- Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics*. 2008 Mar 26;9 Suppl 2:S6.
- Vassura M, Margara L, Di Lena P, Medri F, Fariselli P, Casadio R. FT-COMAR: fault tolerant three-dimensional structure reconstruction from protein contact maps. *Bioinformatics*. 2008 Apr 1. [Epub ahead of print]
- Capriotti E, Arbiza L, Casadio R, Dopazo J, Dopazo H, Marti-Renom MA. Use of estimated evolution-

ary strength at the codon level improves the prediction of disease-related protein mutations in humans. *Hum Mutat.* 2008 Jan;29<sup>(1)</sup>:198-204.

- Capriotti, E., Calabrese, R., and Casadio, R. (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with Support Vector Machines and evolutionary information. *Bioinformatics* 22(22):2729-2734
- Capriotti, E., et al. (2005a) Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, 21, Suppl. 2, ii54–ii58
- Capriotti E, Fariselli P, Casadio R. (2005b). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* Jul 1;33(Web Server issue):W306-10.
- Capriotti E, Fariselli P, Casadio R. (2004) A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics.* Aug 4;20 Suppl 1:i63-i68

**Contact :** remo@biocomp.unibo.it