BITS :: Call for Abstracts 2021 - Oral communication

Type	Oral communication
Session	Protein structure and function
Title	Performances evaluation of selected web tools predicting changes in protein stability
All Authors	Anna Marabotti(1), Eugenio Del Prete(2), Bernardina Scafuri(1), Angelo Facchiano(3)

Affiliation

(1) Dept. Chemistry and Biology "A. Zambelli", University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy.
(2) CNR-IAC, National Research Council, Institute for Applied Mathematics "Mauro Picone", Via P. Castellino, 111, 80131 Naples, Italy
(3) CNR-ISA, National Research Council, Institute of Food Science, Via Roma 64, 83100 Avellino, Italy

Motivation

The prediction on how a mutation can affect protein thermodynamic stability is a hard task for computational biology. Despite decades of research and so many predictors developed on purpose, there are still doubts about the reliability of their results. The issues can be ascribed both to the paucity of high quality and reliable data for the creation of a reference database, and to the different approaches developed so far to cope with this task [1]. In this work, we present the assessment we applied to five predictors available online, representing different approaches, and using a reference database of high-quality structures.

Methods

The predictors assessed were: INPS-3D [2], a machine-learning method tailored to face the problem of anti-symmetric property; PoPMuSiC [3], a method using a linear combination of statistical potentials, tailored to correct the bias toward destabilizing mutations; DynaMut [4], one of the most recent web servers developed, based on Normal Mode to take into account the contribution of protein flexibility; MAESTROweb [5], the only web server able to manage both multimeric proteins and compound heterozygous multiple mutations; DUET [6], a consensus predictor combining two other predictors previously developed by the same research group. Starting from VariBench dataset [7], we performed a filtering based on the selection of high-quality reference proteins, in terms of thermodynamic experimental data and quality of the structures associated to them. We created a balanced dataset for number and $\Delta\Delta G$ distribution of destabilizing and stabilizing mutations, in order to evaluate the bias of predictors with respect to abovementioned issue. Finally, we divided the monomeric proteins from the multimeric ones, and assessed separately the predictions made on these two groups, considering that most predictors are not able to handle directly multimeric proteins. To assess the reliability of the predictors, we evaluated if the sign of the $\Delta\Delta G$ predicted by the different tools was in agreement with the sign of the experimental measure associated to the same mutation, and we calculated several statistical parameters to compare the performances of the prediction methods. We computed all the statistics in R language.

Results

Our analysis shows that, although there have been improvements in this field over time, the performances of the assessed predictors are still far from an ideal condition. The most frequent problem detected is a bias towards destabilizing mutations, even in predictors in which this issue is claimed to be solved. Additionally, when the mutation causes a $\Delta\Delta G$ within the interval ±0.5 kcal/mol (generally accepted as the interval error for the measurement of this parameter), the predicted results are generally less reliable than those predicted for mutations causing a $\Delta\Delta G$ outside that interval. Finally, we found that a rough but effective way to increase the reliability of the predictors is the combination of their results into a consensus parameter, based principally on the prediction of the sign of $\Delta\Delta G$.
For these reasons, we suggest to developers to consider in the future the usage of balanced data sets for training their future predictors, and to define the effect of a mutation on the stability of the protein as "uncertain" when its predicted $\Delta\Delta G$ falls within the range ±0.5 kcal/mol. Furthermore, we suggest to users to combine the results of multiple tools, in order to increase the chances of having correct predictions about the effect of mutations on the thermodynamic stability of a protein.

Info

References
[1] Marabotti A, Scafuri B, Facchiano A. Brief Bioinform. 2020; epub ahead of print. doi: 10.1093/bib/bbaa074
[2] Savojardo C, Fariselli P, Martelli PL, Casadio R. Bioinformatics 2016;32: 2542–2544.

[3] Pucci F, Bernaerts KV, Kwasigroch JM, Rooman M. Bioinformatics 2018;34: 3659–3665.
[4] Rodrigues CH, Pires DEV, Ascher DB. Nucleic Acids Res. 2018;46: W350–W355.
[5] Laimer J, Hiebl-Flach J, Lengauer D, Lackner P. Bioinformatics 2016;32: 1414-1416.
[6] Pires DEV, Ascher DB, Blundell TL. Nucleic Acids Res. 2014; 42:W314–W319.
[7] Nair PS, Vihinen M. Hum Mutat. 2018;34: 42-49.

Figure

-

Availability   -

Corresponding Author	
Name, Surname	Anna, Marabotti
Email	amarabotti@unisa.it
Submitted on	07.05.2021