Using Rosetta to assess the structural impacts of genetic variants

Igor Petrik, Min-Sun Park, Adam Chamberlin, and R. Bryn Fenwick

The functional impacts of genetic variants are often predicted using sequence based predictors, and for unstructured regions of proteins these predictors are often very efficient and accurate {Pejaver V, *et al.; bioRxiv* 134981; doi: https://doi.org/10.1101/134981.}. However, in structured protein regions these predictors perform significantly worse. The computational prediction of protein stability (ΔΔG) has been shown to be a useful tool in predicting the deleteriousness of genetic variants when properly applied {Pesaran T, *et al.; Int J Breast Cancer*. 2016; 2016:2469523; Kiel C, Serrano L; *Mol Syst Biol* 2014; 10:727}; however, often there are limitations such as handling of backbone flexibility {Guerois R, *et al.*; *J Mol Biol* 2002; 320:369}. Encouraged by Rosetta's versatility and accuracy in protein design, we investigated using the Rosetta score function to determine destabilizing effects of native protein alterations; we found that, while the Rosetta software package is excellent for this purpose, the standard *talaris* energy terms is less well suited. To improve its predictive capacity for variant classification, we have re-weighted the score function and added additional terms. Here we present this score function and its application to the prediction of deleterious changes of the human proteome. From our results we expect that this score function will be useful not just for prediction of native fold disruption, but also in characterization and design of proteins and interfaces.