

ProQM-resample: Improved Model Quality Assessment for Membrane Proteins by Limited Conformational Sampling

Björn Wallner^{1,2*}

¹Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden. bjornw@ifm.liu.se

²Swedish e-Science Research Center

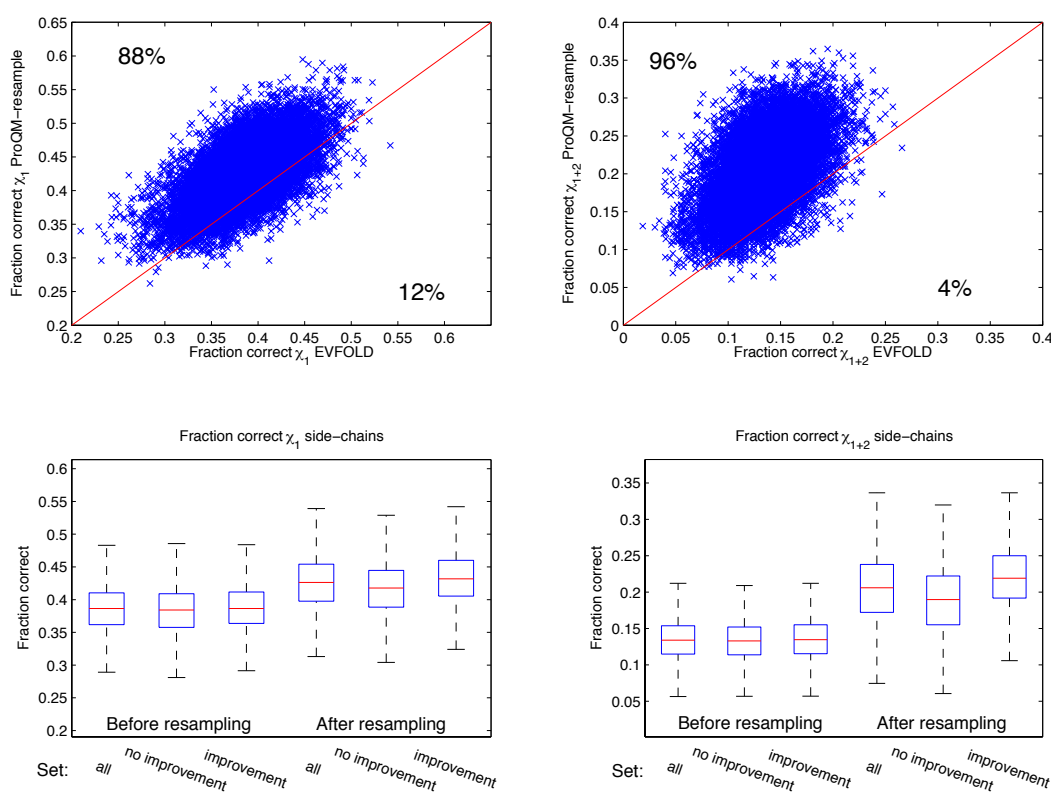


Fig. S1. Side-chain quality before and after resampling, a rotamer is considered correct if it is within 40 degrees of the correct rotamer. A) Fraction correct χ_1 before vs. after resampling; B) Fraction correct χ_{1+2} before vs. after resampling; C) Box-and-whisker plots showing the distribution of fraction correct χ_1 for the complete set (all), the targets showing no improvement in model selection (no improvement), and targets showing improved selection (improvement) before and after resampling, respectively. The median is showed in red, the box represent the first and third quartiles and the whiskers extends to 1.5 of the interquartile range covering 99.3% of the data assuming normal distribution; D) Same as in C) but for χ_{1+2} . The targets included in the *no improvement* and *improvement* sets are defined in Fig S2.

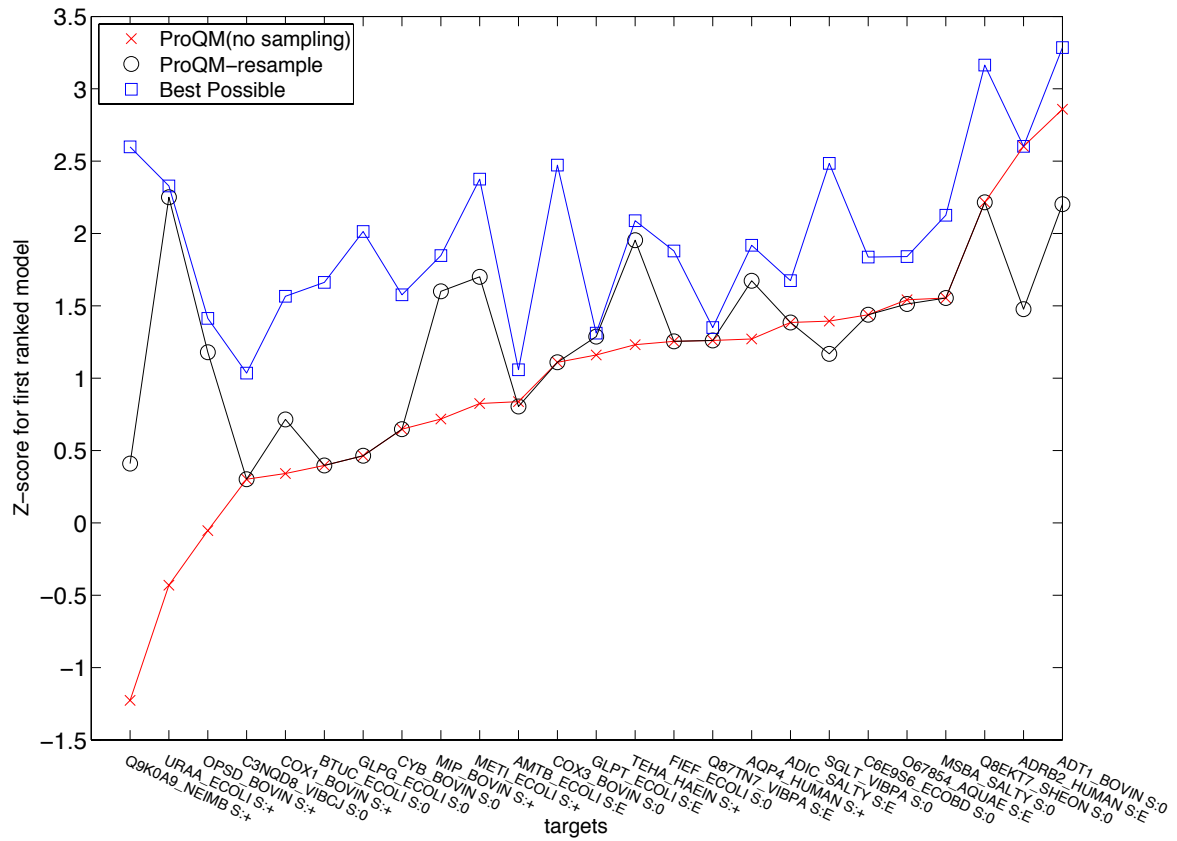


Fig. S2. Target selection per target as measured by the Z-score for the first ranked model for ProQM, ProQM-resampling and the best possible selection, respectively. To analyze the reason for improvement targets were group into three sets, the *excluded* set for which no improvement is possible since the initial choice is already close to optimal, the *improvement* set, and *no improvement* set. Targets with $Z_{\text{best}} - Z_{\text{ProQM}} < 0.3$ were put in the *excluded* set, marked by S:E after the target name on the x-axis, targets with $Z_{\text{ProQM-resampling}} - Z_{\text{ProQM}} = \Delta Z > 0.3$, marked by S:+, are included in the *improvement* set, and targets with and $\text{abs}(\Delta Z) < 0.3$, marked by S:0, are included in the *no improvement* set.

Fig. S3. Scatter plots of predicted scores for ProQM-resample, ProQM, Pcons and EVFOLD-MQAP vs TMscore for each of the 25 targets. The title of each scatter plot has correlation coefficient, R, the TMscore and the Z score for the top 1 model, and the TMscore and Z score for the best possible model.

