FIGURES

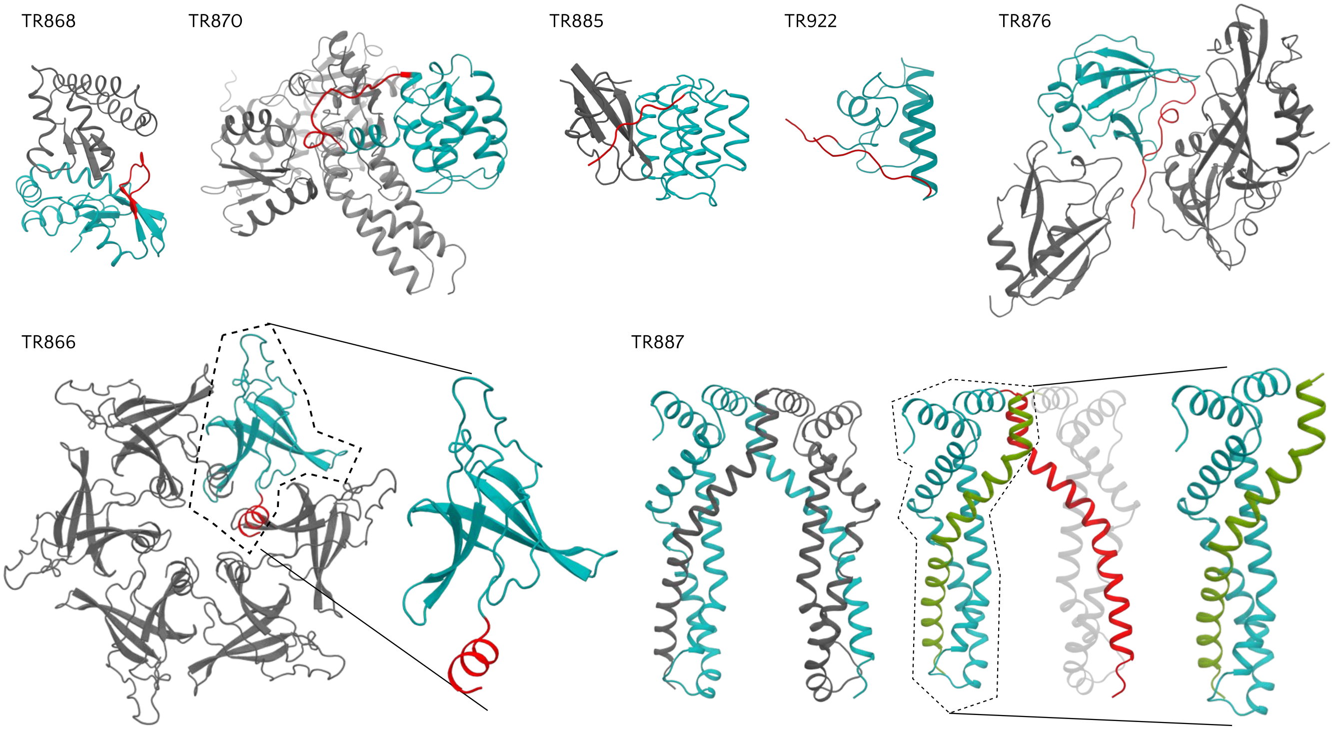

Figure 1. Target structures with inter-domain or inter-chain contacts. The domains to be refined are in cyan, additional domains are in gray. The interacting regions that were removed from the assessment are shown in red. These are: Q65-F75 in TR868, I8-S21 in TR870, F107-A124 in TR876, V142-Y152 in TR866. In the case of TR887, the green region represents the swap segment added to the target structure from the second monomer.

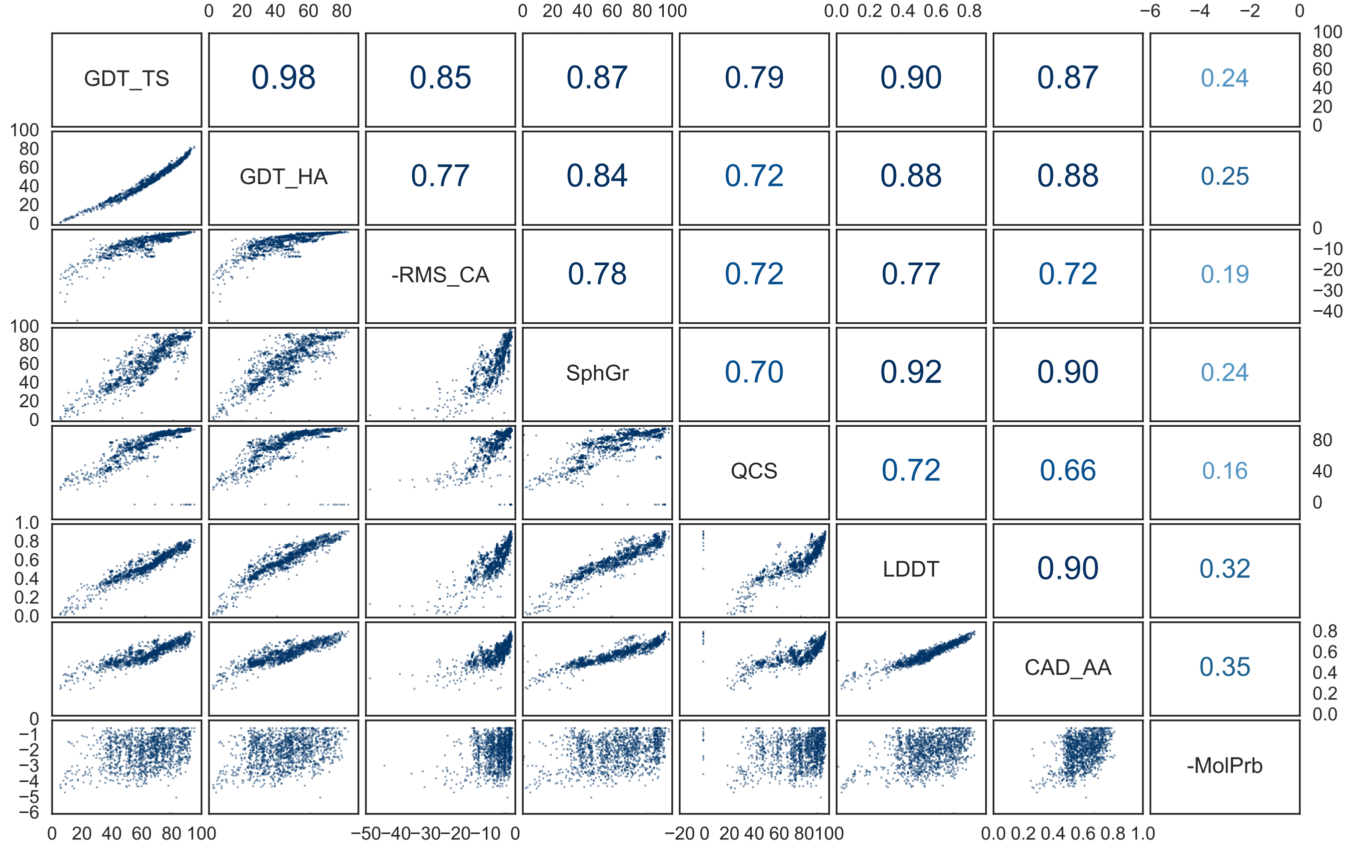

Figure 2. Correlation between eight evaluation metrics for all targets and all submissions. Pair-wise scatter plots are in the left lower triangular part of the table; the correlation coefficients are in the upper one.

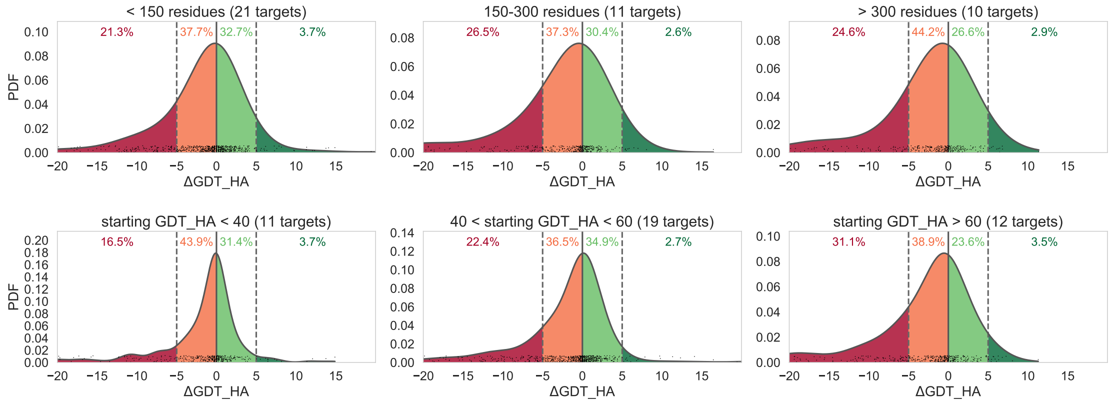

Figure 3. Distributions of ΔGDT_HA differences between the refined and starting models for different target lengths (top row of graphs) and different starting GDT$_{H}A(bottomrow).Dataforfirstsubmittedmodelsarepresented.$

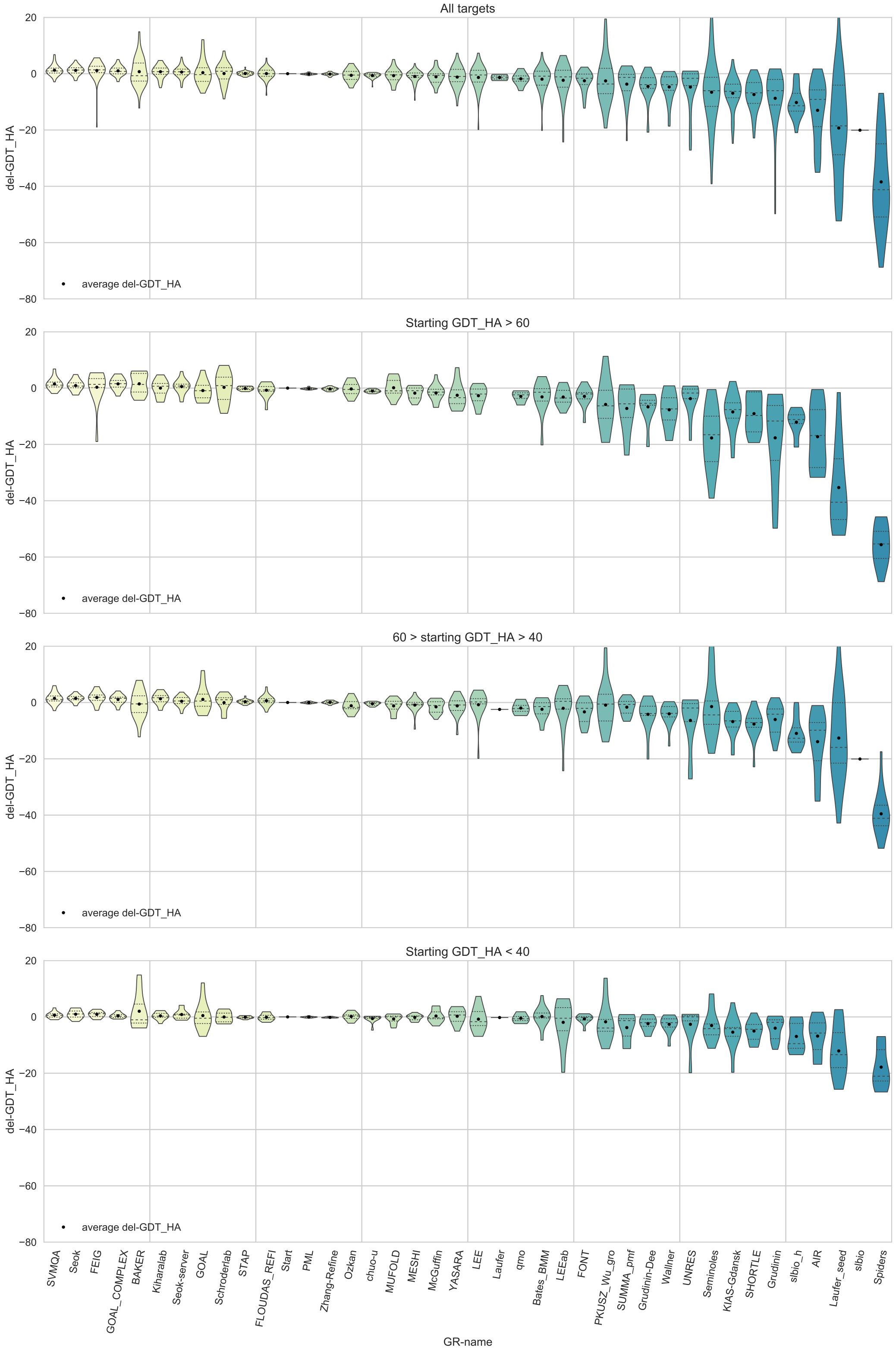

Figure 4. Performance of CASP12 groups as evaluated by the differences in GDT_HA scores between the refined and starting models. The data are shown for all targets (top panel) and for three target subclasses with different GDT_HA scores of starting models (i.e. different difficulties of original targets for tertiary structure prediction). Only models ranked as #1 by the predictors are considered. The quartiles are shown as dotted lines in the violin plots. Groups are sorted according to decreasing ΔGDT_HA mean on all targets (top panel).

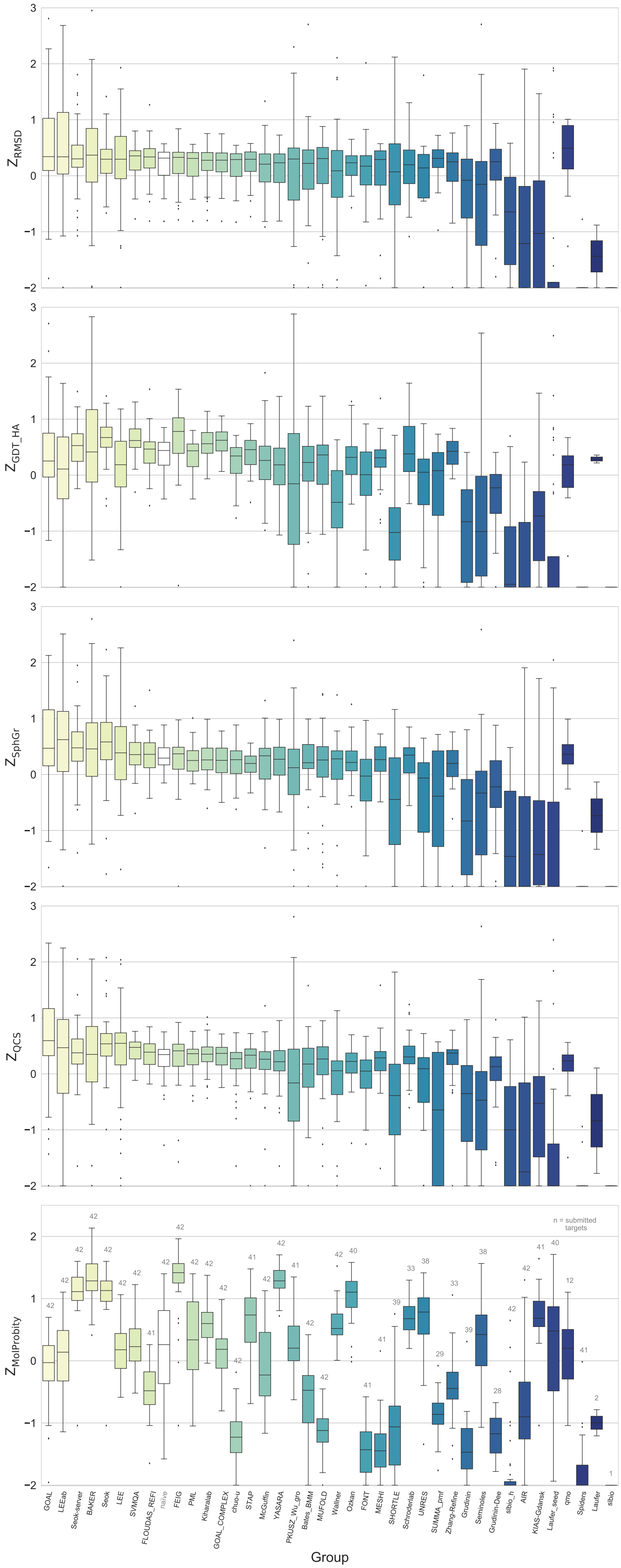

Figure 5. Overall performance by group as measured by RMSD, GDT_HA, SphGr, QCS and MolPrb Z-scores. Each panel shows boxplots of Z-scores for a specific measure. Groups are ordered left to right by the sum of RMSD Z-scores (higher is better) over all targets (higher to lower). Missing predictions are assigned a value of -2 for each target. The number of submitted targets for each group are reported in grey on top of the box plots for MolPrb.

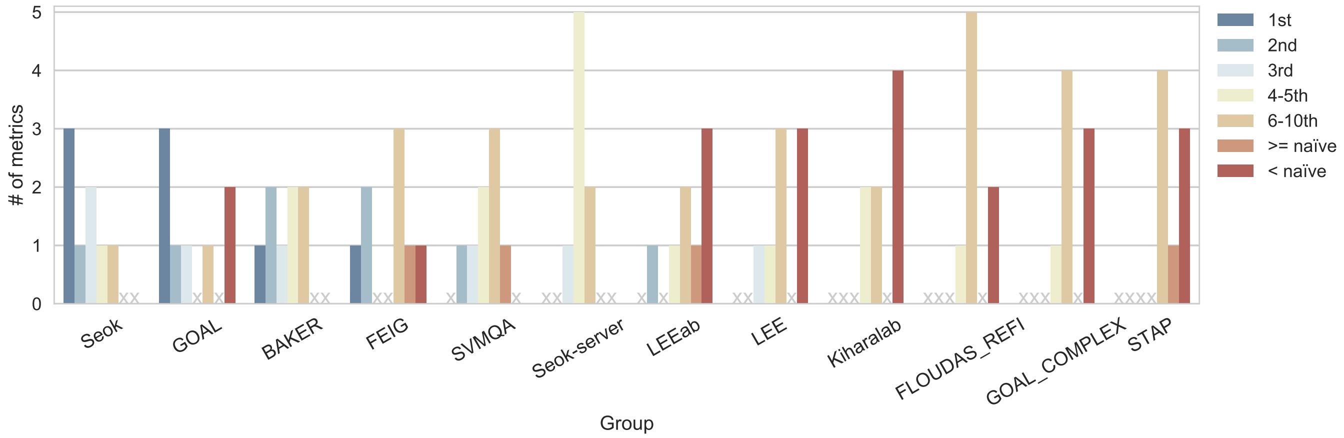

Figure 6. Cumulative group ranking for the eight selected metrics. The plot shows the number of times a group appears with a particular ranking in the best 10 models according to the various metrics considered separately. When a group is not in the best 10, we report whether the score is higher or lower than that of the “naïve” submission. Only groups appearing among the best 10 according to at least 2 metrics are shown.

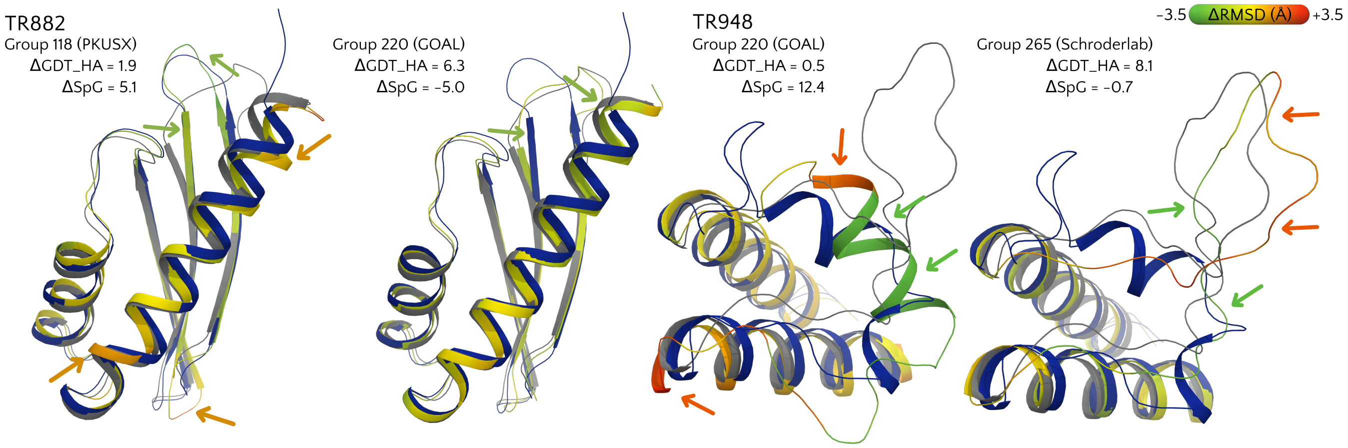

Figure 7. Two representative examples showing remarkable discrepancies between the GDT_HA and SphereGrinder scores. The target structure is shown in blue, the starting model in gray and the prediction with a color scale based on ΔRMSD, ranging from green (improved over starting model) to yellow (no improvement) and red (worse). For clarity, part of the structure have been removed from target TR948.

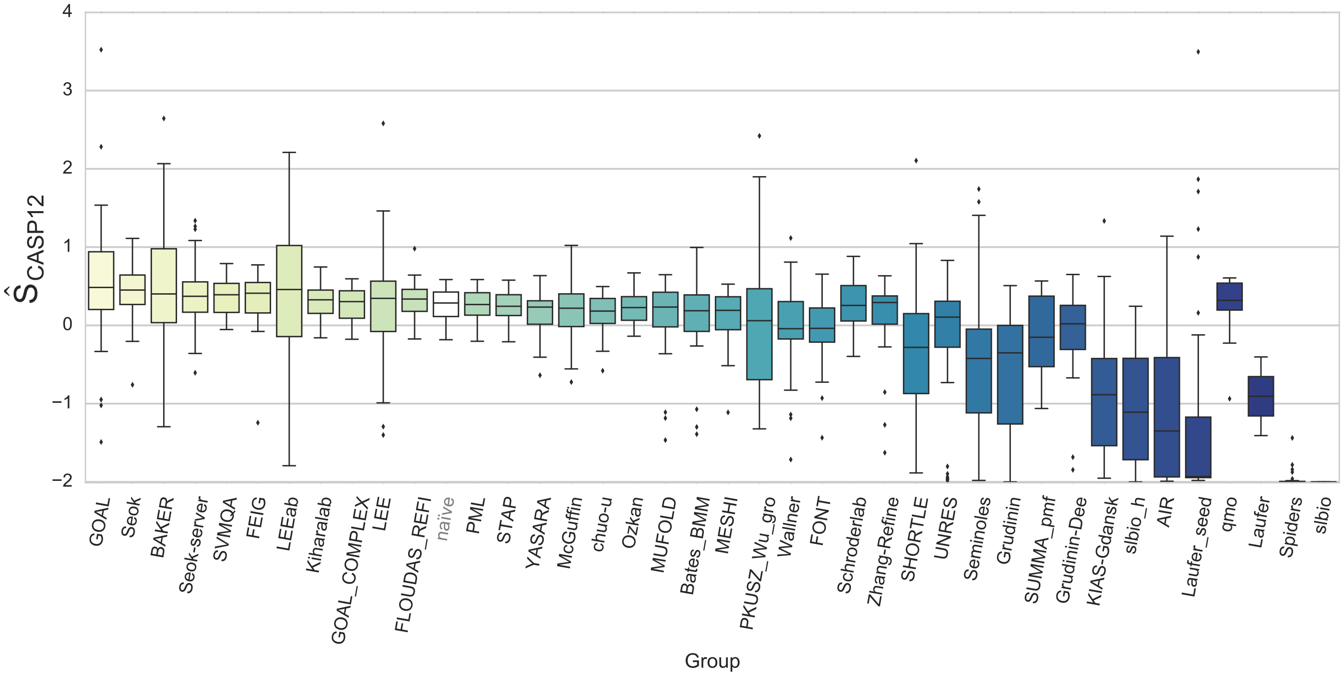

Figure 8. Overall performance by group as measured by the $\widehat{S}$ CASP12 assessors score. Groups are ordered left-to-right by their rank (the sum of $\widehat{S}$ over all targets).

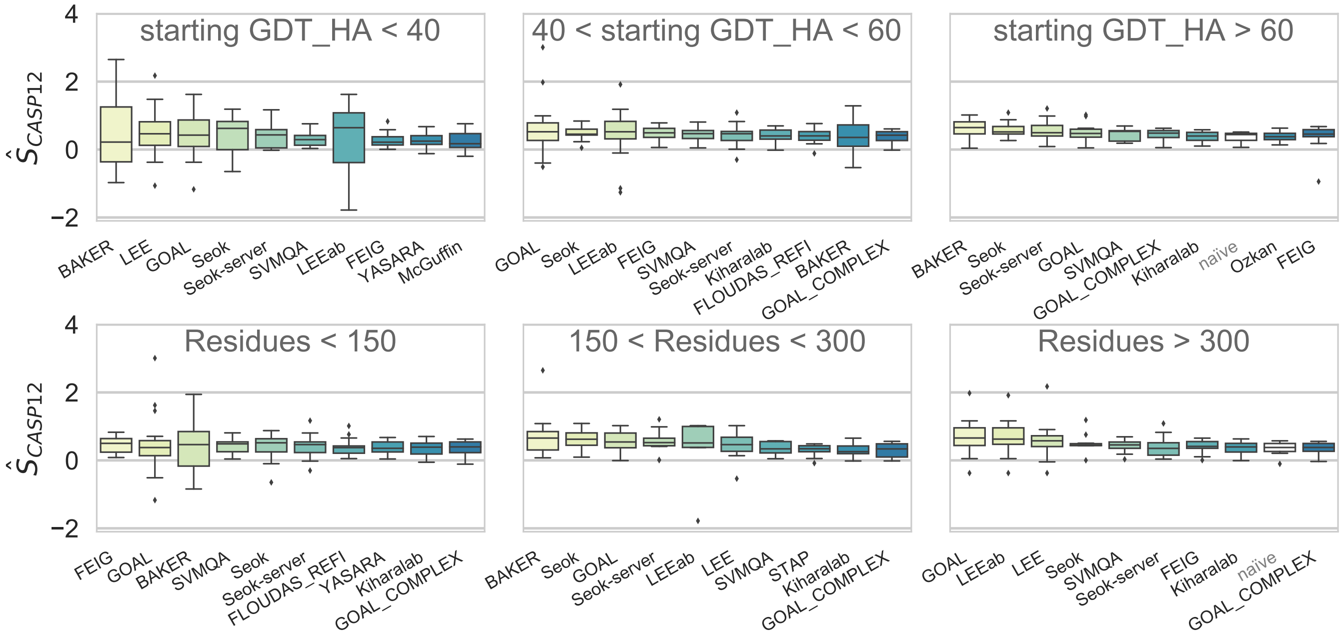

Figure 9. Overall performance by group as measured by the $\widehat{S}$ CASP12 assessors score on the targets grouped into three bins based on the starting model's GDT_HA (top row) and target size (lower row). Groups in each panel are ordered left-to-right by their rank (the sum of $\widehat{S}$ over all targets). Only the first submitted models are considered.

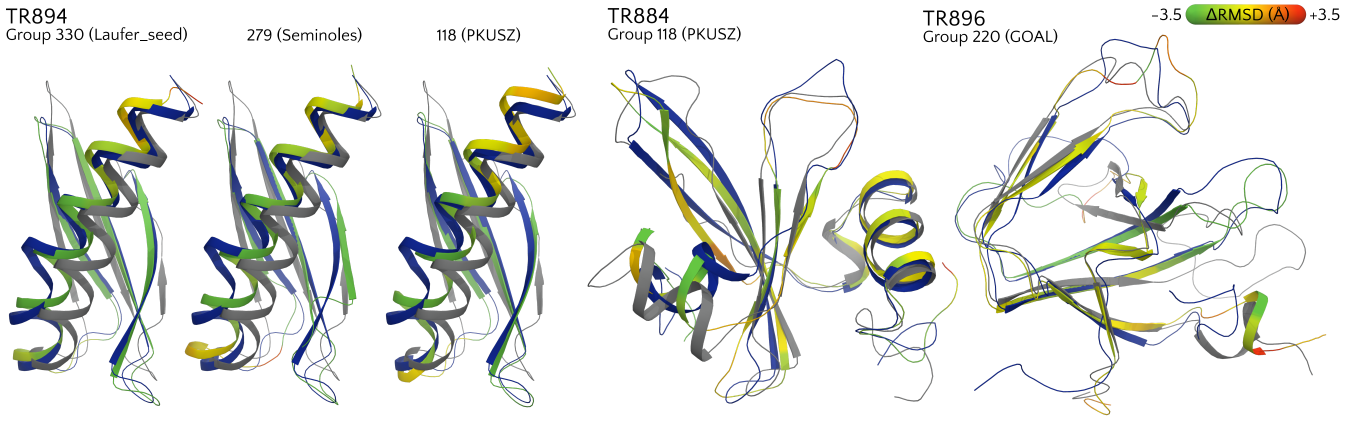

Figure 10. Some examples of notable refinement. The target structure is shown in blue, the starting model in gray and the prediction with a color scale based on residue wise ΔRMSD from green (improved over starting model) to yellow (no improvement) and red (worse).

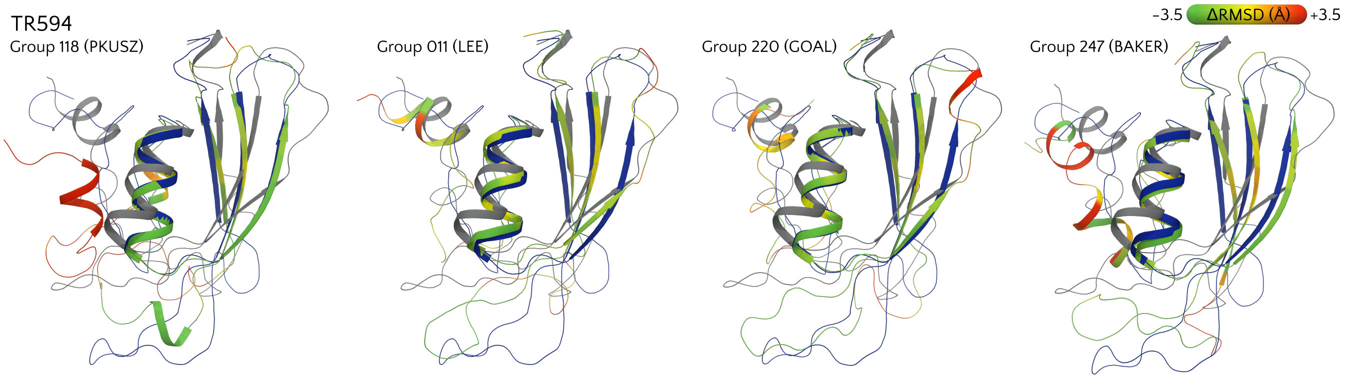

Figure 11. Prediction for target TR594 with a ΔGDT_HA higher than 10. The target structure is shown in blue, the starting model in gray and the prediction with a color scale based on residue wise ΔRMSD from green (improved over starting model) to yellow (no improvement) and red (worse).

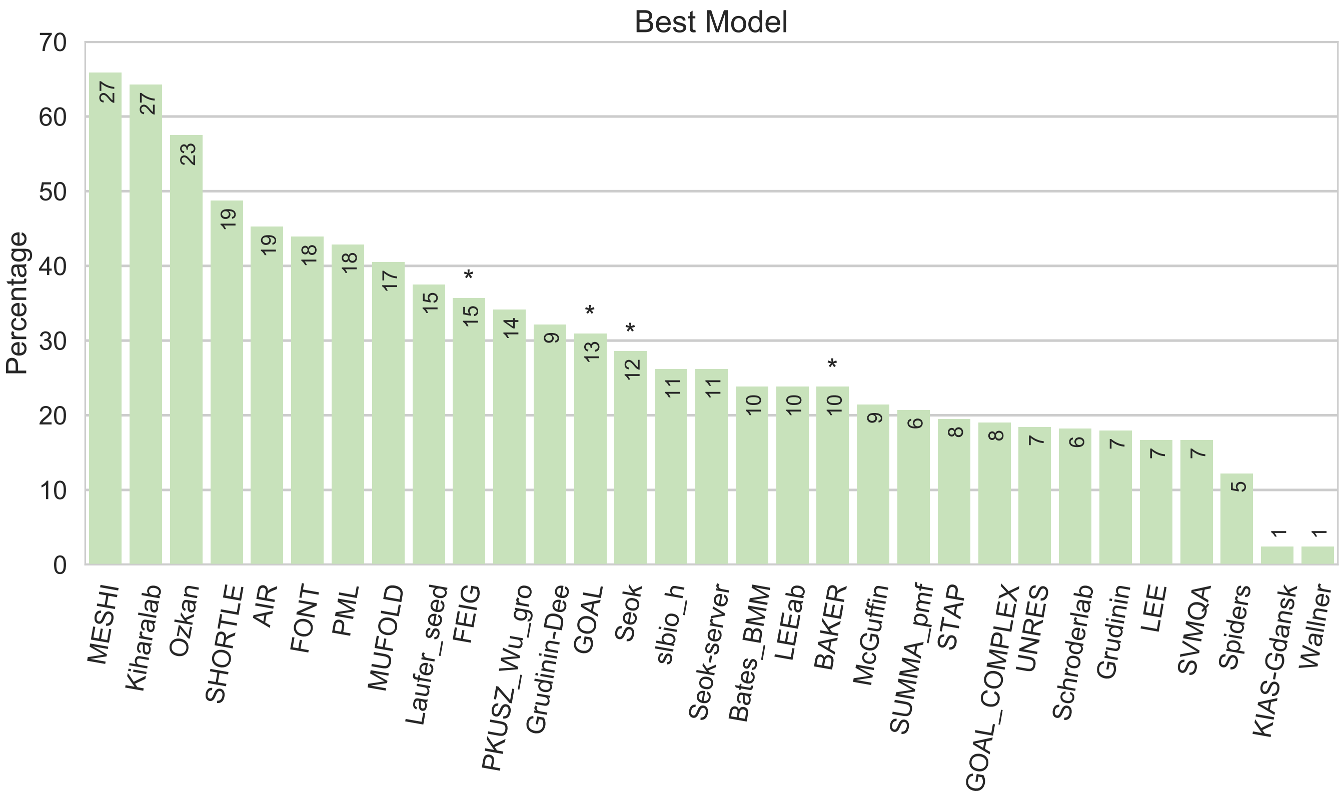

Figure 12. Best model or method selection. The plot reports the percentage of submitted models 1 that correspond to the best of the 5 submitted models. The numbers on top of the bars report the number of model 1s corresponding to the best models (not all groups submitted models for all targets). The asterisks mark the CASP12 top performers.

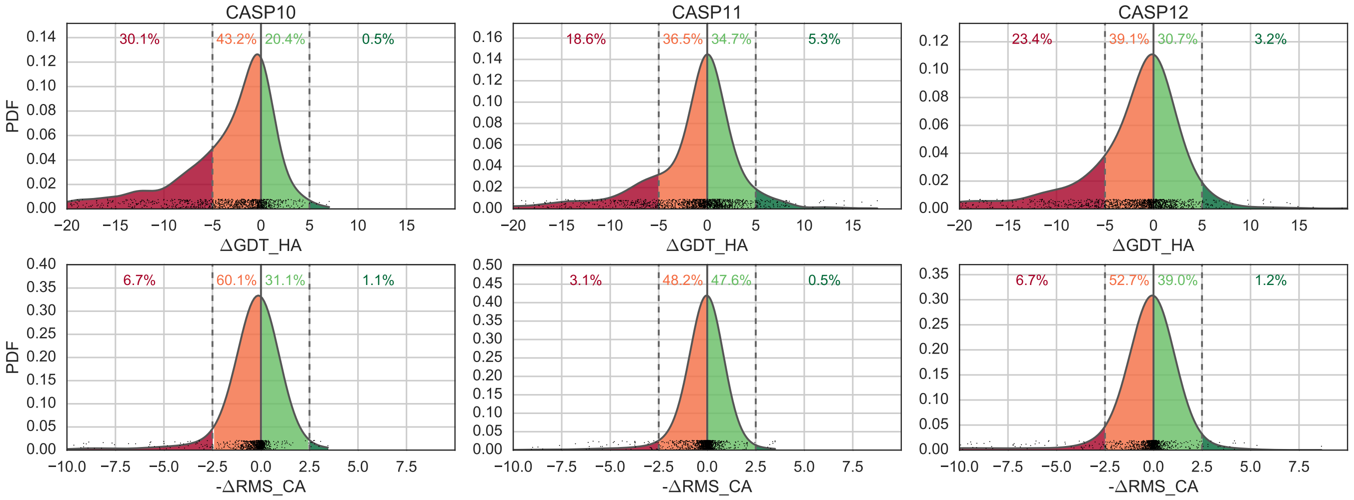

Figure 13. Distributions of GDT_HA and RMSD scores for CASP12 and the two previous CASPs.

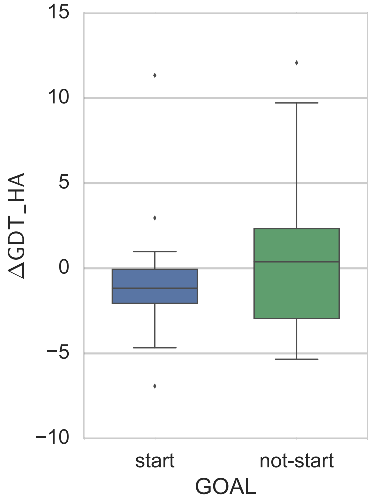

Figure 14. Comparison of the refinement achieved by GOAL in targets for which the starting structure was provided by other groups (”not-start”) or by GOAL itself (”start”).