

Figure S1: Cumulative average target difficulty over all CASPs. The relative difficulty of each CASP is expressed as a Z-score derived from the average similarity of best templates to those target structures and sequences.(see Methods). The higher the score, the more difficult the target set in a specific CASP. This alternative view of target difficulty confirms that CASP9 human/server targets are harder than those of CASP7 and CASP8, and approximately the same in difficulty as CASP5 and CASP6 targets.

Figure S2: Trend lines for the average GDT_TS score over the best six groups on each target in all CASPs, as a function of target difficulty. The trend lines here are similar to those for the GDT_TS scores using the best model only (Figure 2, main text), showing that best model performance is reasonably representative of the performance of a number of groups.

Figure S3: GDT_TS scores of submitted best models for targets in CASPs 8 and 9, as a function of Z-score ranked target difficulty. Each point represents one target. The three targets ranked easiest and three hardest are omitted for clarity. As in Figure 2, the trend lines show a likely increased accuracy of modeling in the middle range of difficulty in CASP9. This plot differs from Figure 2 in using the Z-score ranking of difficulty, and is included to demonstrate that the apparent mid-range improvement is robust with respect to the choice of difficulty scale.

Figure S4: Average, maximum, and minimum spread among the sets of five models submitted for each target, for those groups included in the analysis of best model selection (Figure 8).