

SUPPLEMENTARY INFORMATION

Abinitio structure prediction protocol

A sample command line as employed for *de novo* structure prediction using the *AbinitioRelax* protocol of Rosetta3.7:

```
ROSETTA_DIR/main/source/bin/AbinitioRelax.linuxgccrelease @  
general_flags.txt
```

Parameter input file “general_flags.txt”:

```
-nstruct 1000  
  
-in  
-file  
-fasta ./${TARGET}.fasta  
-frag3 ./t000_.200.3mers.gz  
-frag9 ./t000_.200.9mers.gz  
-native ./${TARGET}-refe.pdb  
-gdtmm true  
  
-out  
-sf fsc_files/${ID}.fsc  
-file  
-silent out_files/${ID}.out  
-silent_struct_type binary  
  
-constraints  
-cst_file ${TARGET}_${GROUP}${LIST}.cst  
-viol  
-viol_level 10
```

Sample restraint file “T0859_111_2L.cst” (truncated):

```
AtomPair CB 67 CB 81 BOUNDED 2.0 8.0 0.50 HMC  
AtomPair CA 68 CB 80 BOUNDED 2.0 8.0 0.50 HMC  
AtomPair CB 66 CB 82 BOUNDED 2.0 8.0 0.50 HMC  
...
```

Figures

Figure S1 Correlation between the alignment depth and precision for the Top5 groups according to the order in Table 2.

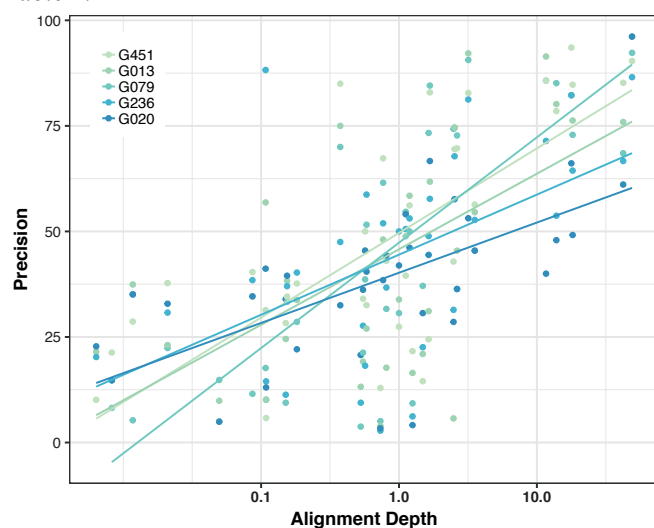


Figure S2: Number of medium and long range contacts in relation to the sequence length for all CASP12 domains. The majority of targets have their number of contacts between one and two and a half times the sequence length, with some outliers having as few as 0.5 and as many as three times the sequence length.

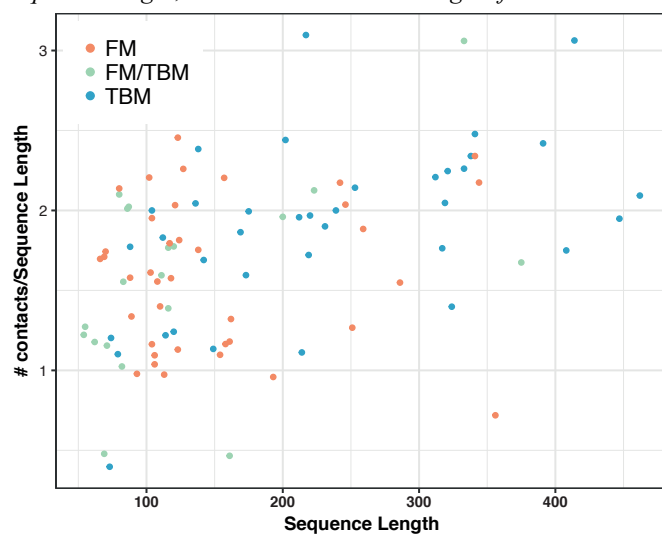


Figure S3: The GDT_TS of contact-guided Rosetta models built by us for different contact prediction groups as a function of the precision of underlying contact prediction on three representative targets - T0866, T0904 and T0941. The tertiary structure predictions were built separately for six lengths of contact lists ($0.2L$ - $3L$) used to guide the modeling. Points in the graph represent the median GDT_TS score within the top 100 structures built for each contact prediction group. The median GDT_TS of the Top 100 models without contacts is indicated by the dashed vertical line. The median GDT_TS of the Top 100 models without contacts is indicated by the dashed vertical line. The median GDT_TS correlates with the precision. Hardly any improvement in respect to the run without constraint is observed for T0941 (right). For the other two targets an improvement in median GDT_TS depends on a combination of list size, precision, and target.

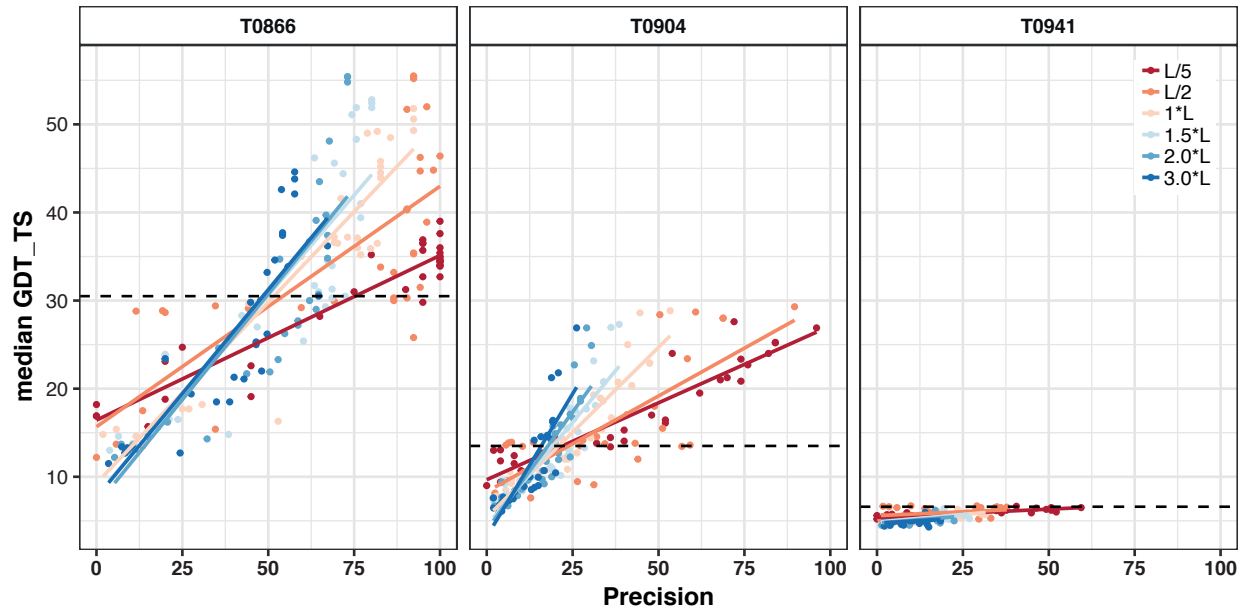


Figure S4: The GDT_TS of contact-guided Rosetta models (see caption Figure S3 for details). GDT_TS within the Top5 models by score based on the different numbers of contacts used ($L/5$ - $3*L$) for all 14 single-domain targets correlates in most cases with the precision. The best GDT_TS for the run without restraints is indicated by the dashed line.

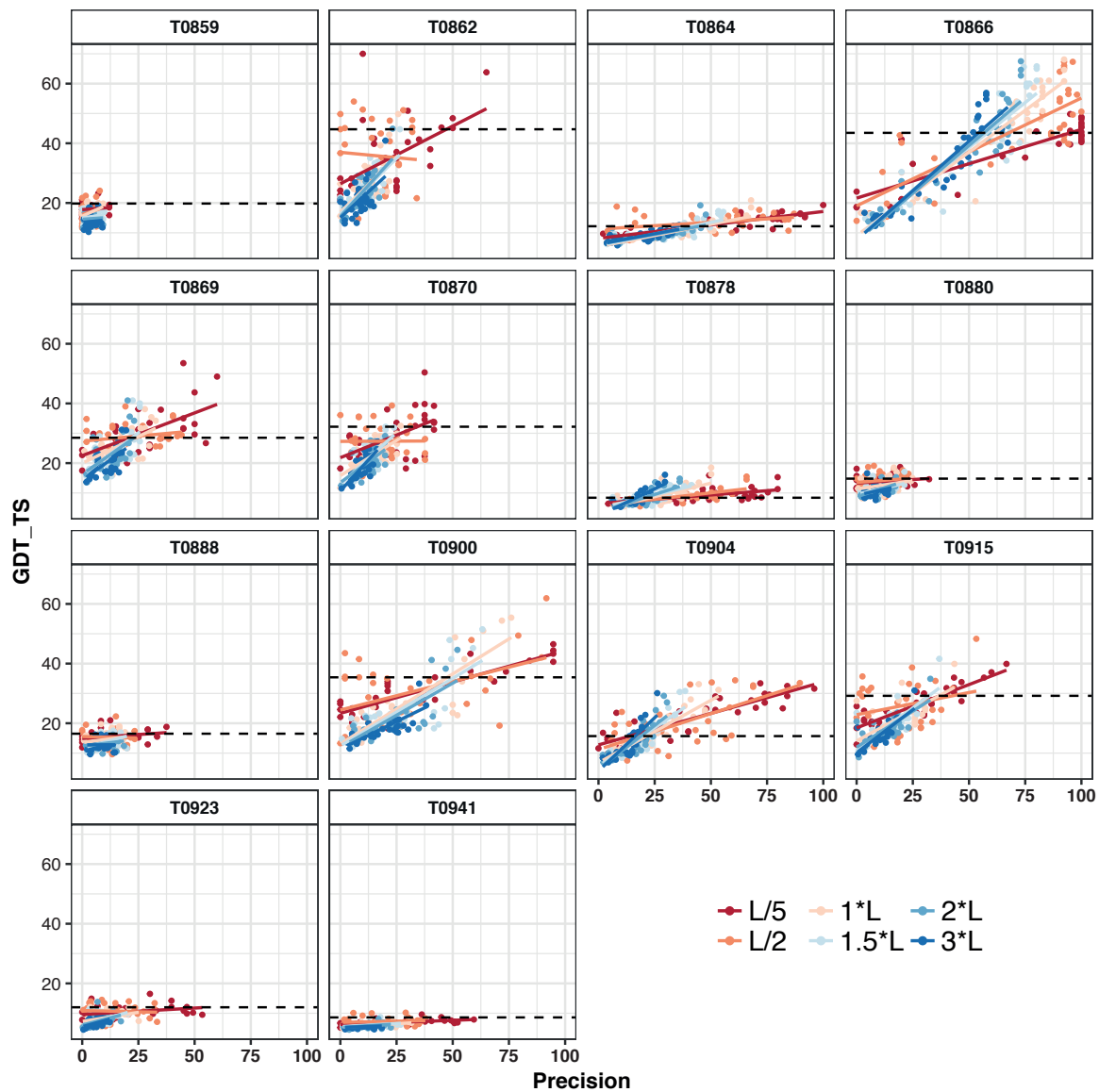


Figure S5: Distribution of delta GDT_TS values compared to the modelling without restraints for the best GDT_TS within the Top 5 structures by score over all targets reaching GDT_TS values above 0.3 (L/2 medium and long range contacts.) The data are provided for the best 5 groups according to Table 2 and a consensus set comprised of the best 1000 models by score from all models of the five groups. In this set the Rosetta score (Consensus) is capable of picking out the best models in terms of GDT_TS in most cases.

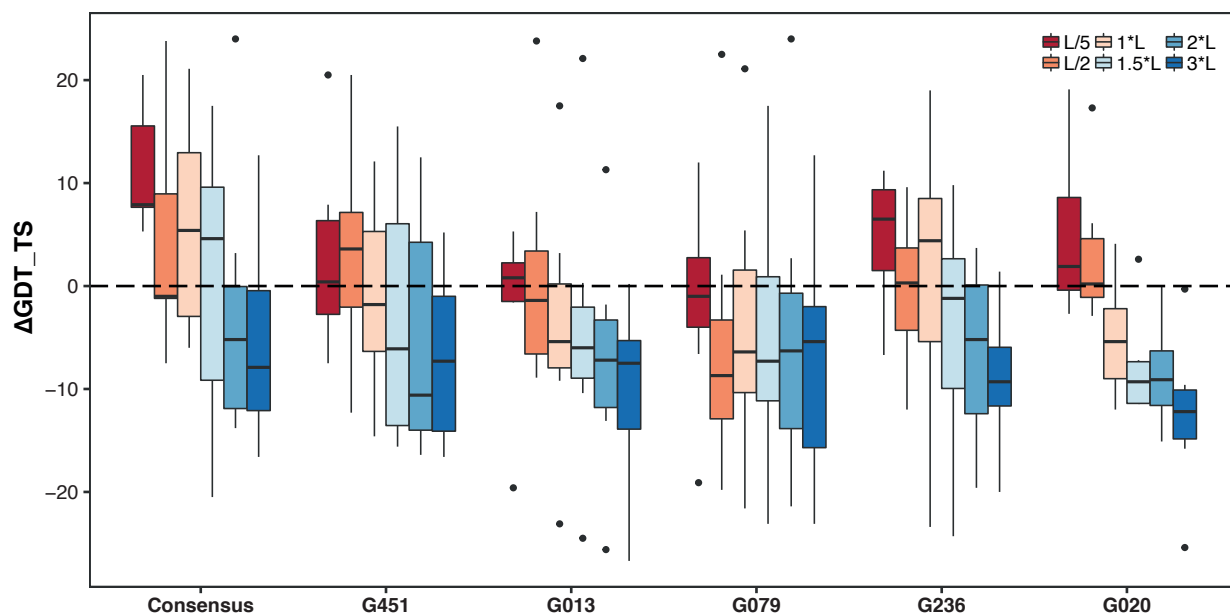


Figure S6: The GDT_TS of contact-guided Rosetta models (see caption Figure S3 for details). The best GDT_TS within the Top5 models by score based on the different numbers of contacts used ($L/5$ - $3*L$) for all 14 targets is anti-correlated to the Average restraint violation in Å. The best GDT_TS for the run without restraints is indicated by the dashed line. The average restraint violation plotted is the mean of the distance above 8Å for all false positive predicted contacts.

