**Supplementary Material**

# lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests

---

**Table S1: Multi-domain targets and corresponding assessment units (AU) evaluated during the CASP9 experiment.** GDC-all scores for individual AUs were averaged for each target (weighted by AU size) and compared to whole-target lDDT scores to determine the optimal inclusion parameter r.

| Target | AU range |
|---|---|
| T0521 | 1: 1-34, 107-179; 2: 35-104 |
| T0528 | 1: 18-138, 269-351; 2: 139-268, 352-381 |
| T0529 | 1: 7-339; 2: 364-561 |
| T0533 | 1: 7-89, 181-296; 2: 90-180 |
| T0534 | 1: 31-80, 257-384; 2: 81-256 |
| T0537 | 1: 65-350; 2: 351-381 |
| T0542 | 1: 2-302; 2: 303-490, 509-585 |
| T0543 | 1: 56-95; 2: 96-140; 3: 141-457, 468-540; 4: 541-884 |
| T0547 | 1: 1-51, 330-342, 435-553; 2: 52-189, 203-329; 3: 343-421; 4: 554-609 |
| T0548 | 1: 12-46; 2: 47-106 |
| T0550 | 1: 31-177; 2: 178-339 |
| T0553 | 1: 3-65; 2: 66-136 |
| T0571 | 1: 32-196; 2: 197-331 |
| T0575 | 1: 1-63; 2: 64-74, 85-179, 196-216 |
| T0579 | 1: 1-29, 94-124; 2: 30-93 |
| T0582 | 1: 2-122; 2: 123-221 |
| T0586 | 1: 5-84; 2: 85-123 |
| T0589 | 1: 24-65, 96-188, 271-369; 2: 189-270 |
| T0596 | 1: 6-58; 2: 59-188 |
| T0600 | 1: 17-75; 2: 76-122 |
| T0604 | 1: 11-94; 2: 95-291, 497-548; 3: 292-496 |
| T0608 | 1: 29-117; 2: 118-278 |
| T0611 | 1: 3-55; 2: 56-169, 179-213 |
| T0628 | 1: 6-132, 279-295; 2: 133-278 |
| T0629 | 1: 1-49, 209-216; 2: 50-208 |

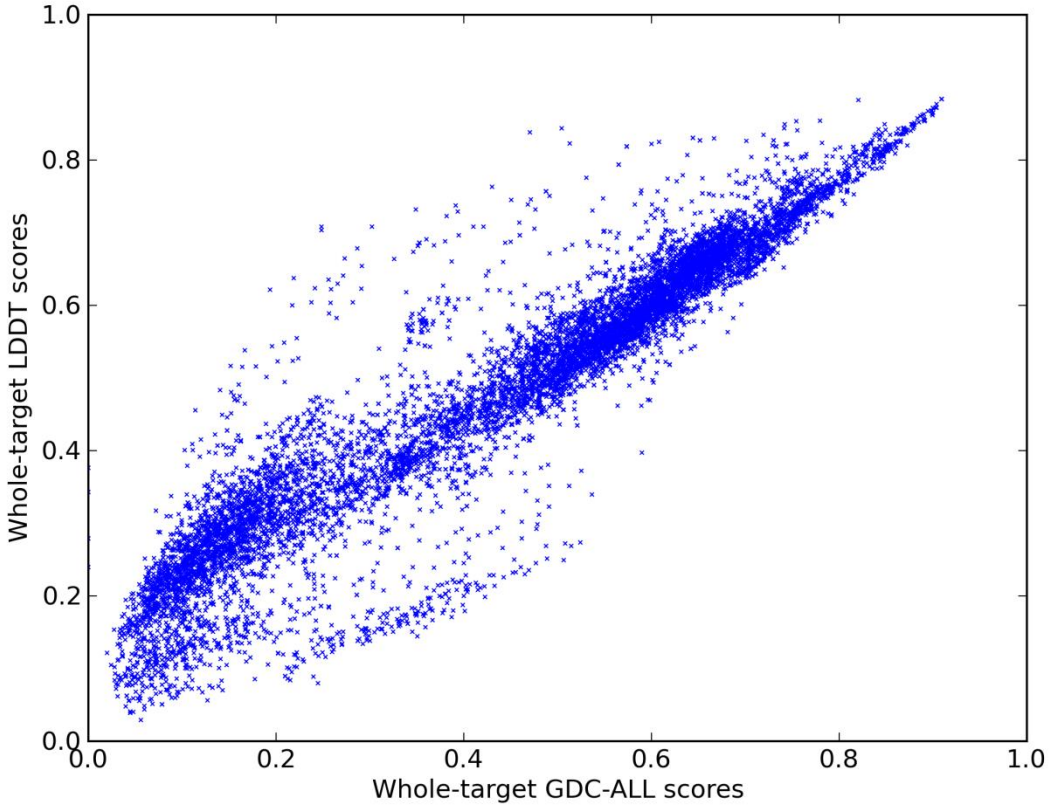



**Fig. S2: Correlation between lDDT scores and GDC-all scores for CASP9 single-domain targets.** Whole-target GDC-all and lDDT scores were computed for CASP9 predictions of single-domain targets. For the lDDT scores, the default value of 15 Å for the inclusion radius was used. The Pearson correlation between the two scores is essentially the same as in the case of multi-domain targets ($R^2$=0.89). Predictions for the following targets were used in the plot: T0515-D1, T0516-D1, T0517-D1, T0518-D1, T0520-D1, T0522-D1, T0523-D1, T0524-D1, T0525-D1, T0526-D1, T0527-D1, T0530-D1, T0531-D1, T0532-D1, T0536-D1, T0538-D1, T0539-D1, T0540-D1, T0541-D1, T0544-D1, T0545-D1, T0551-D1, T0552-D1, T0555-D1, T0557-D1, T0558-D1, T0559-D1, T0560-D1, T0561-D1, T0562-D1, T0563-D1, T0564-D1, T0565-D1, T0566-D1, T0567-D1, T0568-D1, T0569-D1, T0570-D1, T0572-D1, T0573-D1, T0574-D1, T0576-D1, T0578-D1, T0580-D1, T0581-D1, T0584-D1, T0585-D1, T0588-D1, T0590-D1, T0591-D1, T0592-D1, T0593-D1, T0594-D1, T0597-D1, T0598-D1, T0599-D1, T0601-D1, T0602-D1, T0603-D1, T0605-D1, T0606-D1, T0607-D1, T0609-D1, T0610-D1, T0612-D1, T0613-D1, T0614-D1, T0615-D1, T0616-D1, T0617-D1, T0618-D1, T0619-D1, T0620-D1, T0621-D1, T0622-D1, T0623-D1, T0624-D1, T0625-D1, T0626-D1, T0627-D1, T0630-D1, T0632-D1, T0634-D1, T0635-D1, T0636-D1, T0637-D1, T0638-D1, T0639-D1, T0640-D1, T0641-D1, T0643-D1

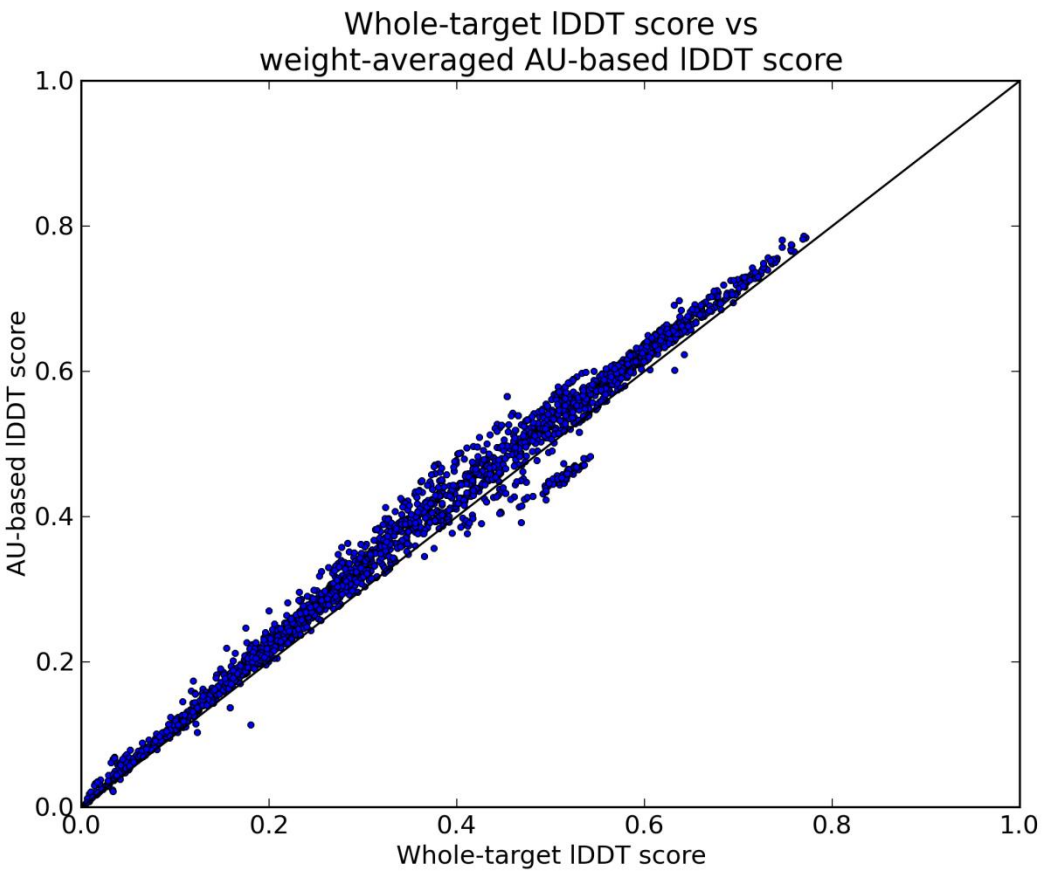



**Fig. S3: Correlation between whole-target lDDT scores and weight-average AU-based lDDT scores for CASP9 multi-domain targets.** The plot shows the correlation between whole-target lDDT scores, and AU-based lDDT scores, weight-averaged following the same procedure applied for Fig.2 of the manuscript (See Materials and Methods), on CASP9 predictions of multi-domain targets (See Table S1). lDDT scores were computed using the default inclusion radius value of 15 Å. The two scores correlate very well (Pearson's correlation $R^2$=0.98) showing a very low sensitivity of the lDDT score, computed using the default parameters, to domain movements.

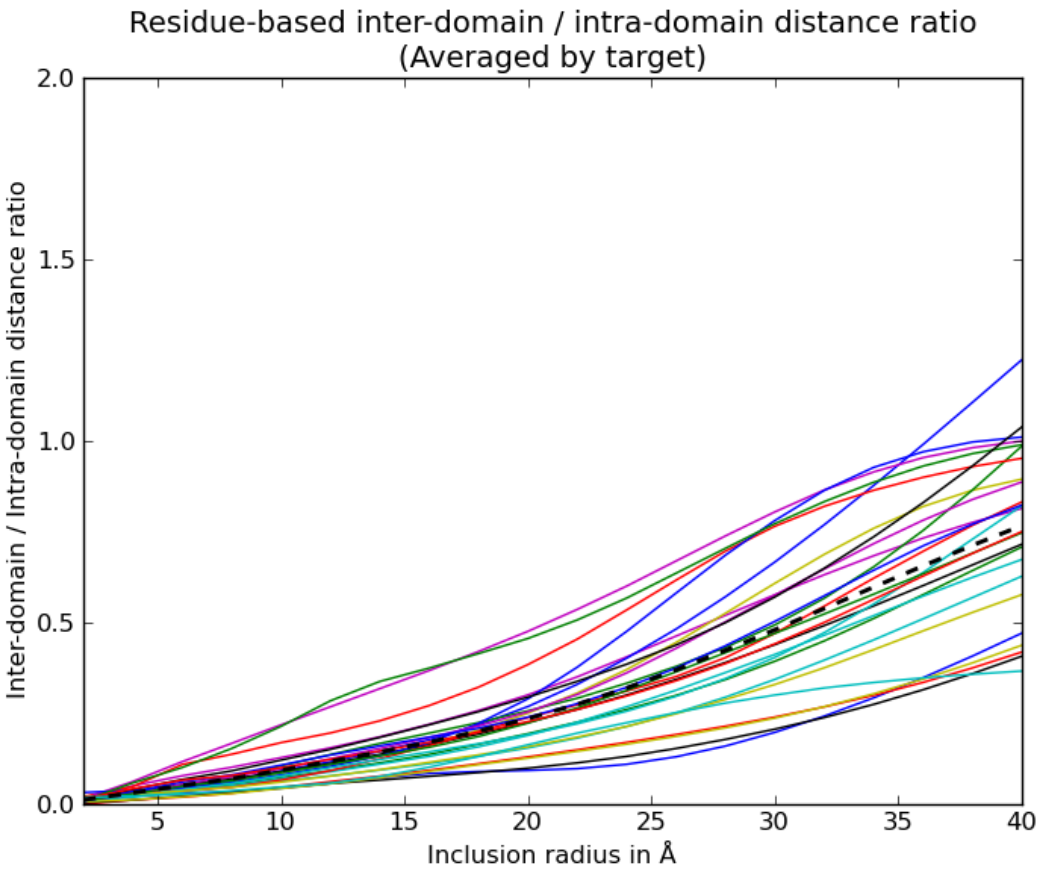


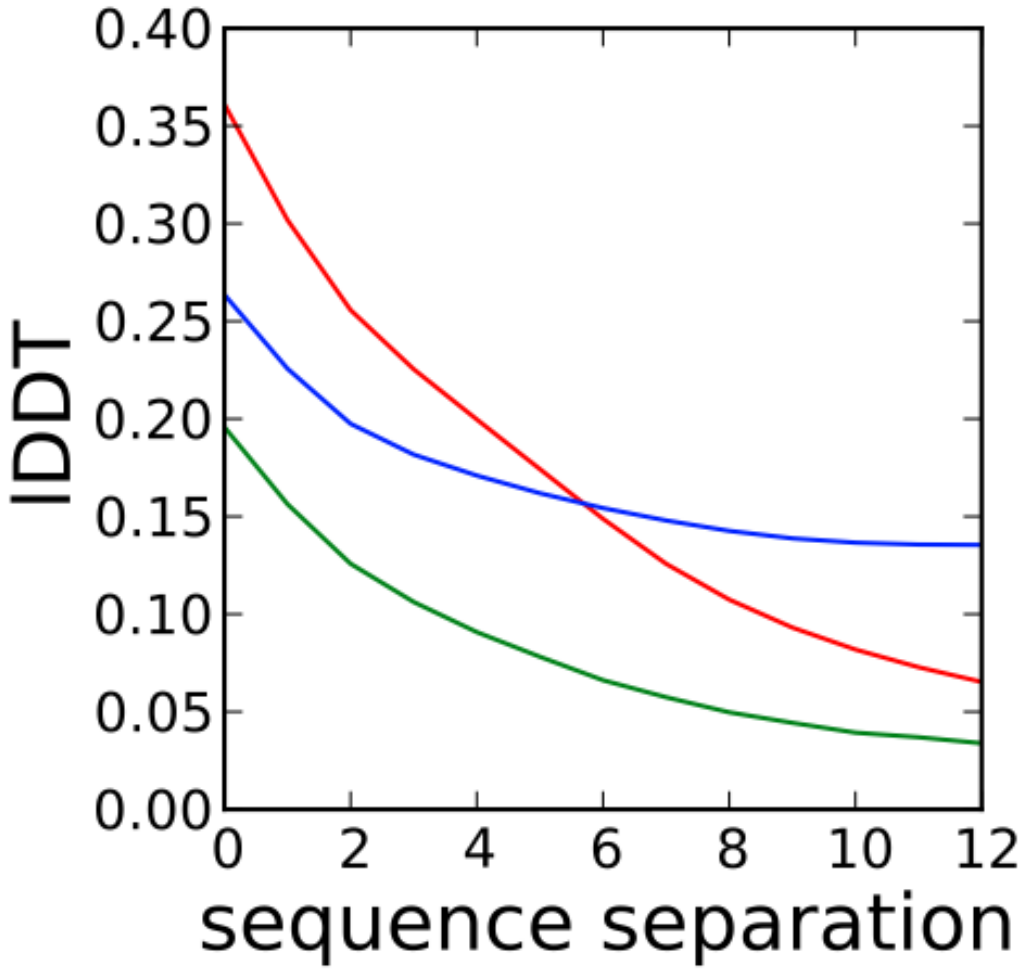

**Fig. S4: Residue-based inter-domain / intra-domain distance ratio.** For each CASP9 multi-domain target (See Table S1), a list of all the distances contributing to lDDT was compiled. For each residue, all distances with at least one atom belonging to the residue were analyzed, and classified either as inter-domain distance (when the end atom belonged to a different domain from the starting one), or as intra-domain distance (when both atoms belonged to the same domain). The inter-domain / intra-domain distance ratio was then calculated for each residue, and averaged over all residues in each target. The calculation was carried out for several values of the inclusion radius, from 2 to 40 Å. The result is shown in the plot, with colored lines representing individual CASP9 multi-domain targets and the thick dashed line representing the variation of the average ratio of all targets. As the inclusion radius increases, each residue connects with a higher number of residues from a different domain, but also to a higher number of residues from the same domain, with a balancing effect. Even for very high inclusion radii, the average ratio is at most slightly above 1 (almost the same number of inter-domain and intra-domain distances), explaining the low sensitivity of the lDDT score to domain movements, except in cases of extreme size difference between domains.

**Evaluation of random lDDT score - Flory-Huggins Polymers**

An ensemble of Flory Huggins (FH) polymers was constructed by creating a polyglycine model of the same length as the reference structure, drawing the backbone torsion angles from the Ramachandran plot. Structures with severe clashes were discarded. For the FH vs FH case, polyglycine models of 150 residues were constructed for both the reference and the models. To assess the behavior of lDDT for random structures, we have constructed different sets of random models and calculated their lDDT score. The following 3 cases were tested:

1. CATH architecture (model) vs CATH architecture (reference)
2. Flory-Huggins polymer (model) vs CATH architecture (reference)
3. Flory-Huggins polymer (model) vs Flory-Huggins polymer (reference)

Figure legend: Average lDDT values for random models at different sequence separations: FH vs FH (red), FH vs CATH architecture (green), CATH architecture vs CATH architecture (blue).




The highest lDDT scores are achieved for the FH vs FH case. Since the Flory Huggins polymers are lacking regular secondary structure elements and are loosely packed, the FH polymers reference structures only have very few local interactions; most of them trivial neighbor contacts which are present both in the model and the reference. At increased sequence separation, trivial neighbor contacts are ignored and the lDDT score starts to rapidly decrease. When comparing CATH architecture vs CATH architecture, the average lDDT score is 0.25, which is considerably lower than

for the FH vs FH case. The lowest lDDT scores are achieved by calculating the lDDT score between FH and CATH architecture. The local packing of "real" protein structures is not well reproduced in Flory-Huggins polymers.

In protein structure prediction typically a number of protein-likeness constraints are imposed on models, and as a result models contain proper secondary structure and are well packed. Thus, they resemble proper folds more than Flory Huggins polymers. Therefore, the CATH architecture vs. CATH architecture comparison was considered to be a better estimate of a "random" protein structure prediction.

**Choice of sigma values for stereo-chemical validation**

Programs such as WHATIF and PROCHECK employ 4 sigma as the threshold for stereo-chemical validation. However, we have found this value to be too strict for current CASP models, since there are orders of magnitude more stereo-chemical violations in CASP models than in experimental structures deposited in the PDB. While it would be desirable to use the same stringent criteria as for experimental structures, we considered doing so in the current CASP experiment as not appropriate as it would penalize a rather large number of predictions. A sigma of 12 appeared as a reasonable balance between penalizing clearly wrong models and still having large fractions of unaffected models in the context of CASP. Over time when the quality of predictions in CASP increases, similar stringent thresholds as for experimental structures should be applied. The following plot illustrates the effect of using different sigma thresholds:

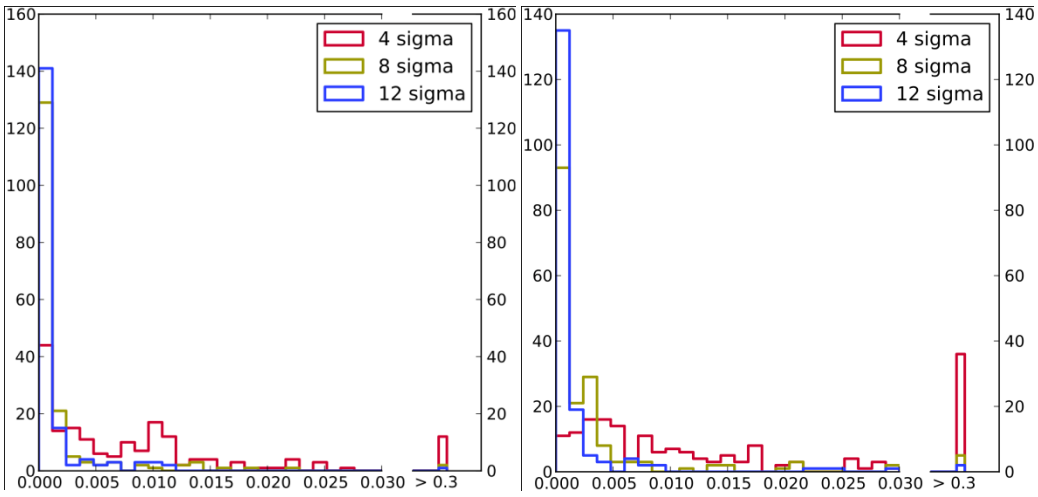



Figure Legend: Histogram of fraction of violations of bonds (left) and angles (right) for all CASP9 structure prediction server methods. Data shown for 4 sigma (blue), 8 sigma (green), and 12 sigma thresholds (red).