Table S1: CAMEO lDDT based ranking for the time frame 2016-05-01 - 2016-07-30. All public servers at the time are shown. Performances have been analyzed according to target difficulties.

Server Name	lDDT (all Targets)
Robetta	65.3 ±16.45
RaptorX	63.8±16.57
IntFOLD3-TS**	62.2±17.53
IntFOLD4-TS	62.0 ±16.32
SWISS-MODEL	56.5±22.49
SPARKS-X	56.3±18.25
Princeton_template	55.6±15.68
IntFOLD2-TS**	55.1±17.47
HHpredB*	47.6±17.41
M4T**	45.1±16.74
Phyre2**	44.5 ±23.135
NaiveBLAST**	43.3±25.59
RBO Aleph*	38.6±16.28

* method had technical problems during the CASP12 season, leading to suboptimal or missing data.

** Method is not reflecting the current development and shown in CAMEO for historic comparison.

Table S1 – A: Performance by lDDT for hard targets

Server Name	lDDT (Hard Targets)
Robetta	46.7±13.16
RaptorX	43.6±12.28
IntFOLD3-TS**	41.1±12.54
IntFOLD4-TS	40.7±11.91
Princeton_template	37.9±11.17
SPARKS-X	36.1±11.98
IntFOLD2-TS**	35.4±11.65
HHpredB*	28.5±11.41
SWISS-MODEL	28.5±13.82
RBO Aleph*	25.9±11.90
Phyre2**	21.1±14.38
M4T**	14.2±11.94
NaiveBLAST**	11.6±15.20

Table S1 – C: Performance by lDDT for medium targets

Server Name	lDDT (Medium Targets)
Robetta	69.5±7.57
RaptorX	69.1±8.42
IntFOLD3-TS**	67.9±8.89
IntFOLD4-TS	67.9±7.74
SWISS-MODEL	63.1±10.52
IntFOLD2-TS**	60.8±9.79
Princeton_template	60.7±10.00
SPARKS-X	60.6±10.05
HHpredB*	54.2±9.41
M4T**	51.4±9.79
NaiveBLAST**	50.9±16.46
Phyre2**	49.8±12.18
RBO Aleph*	42.7±13.02

Table S1 – D: Performance by lDDT for easy targets

Server Name	lDDT (Easy Targets)
Robetta	83.9±3.97
SWISS-MODEL	83.8±5.67
RaptorX	81.8±4.96

IntFOLD4-TS	80.5±4.55
IntFOLD3-TS**	80.4±4.26
M4T**	77.9±6.22
SPARKS-X	77.4±5.96
NaiveBLAST**	73.4±6.47
IntFOLD2-TS**	71.3±4.35
Princeton_template	70.4±9.30
Phyre2**	67.8±10.33
HHpredB*	60.4±4.74
RBO Aleph*	47.9±16.85

We have split the targets by assigning difficulties based on the average global lDDT score across public and development servers applying the same criteria that are described in the main text. Analyzing the server performances across the three categories “hard”, “medium” and “easy” illustrates consistent performance for servers Robetta, RaptorX, IntFOLD3-TS and IntFOLD4-TS at all difficulty levels. For “hard” targets (Table S1-B) the largest spread of standard deviations of up to 15 lDDT units is observed. This is expected considering that some targets are modelled *de-novo* and many methods fail to submit predictions for some targets. In contrast for the “easy” targets (Table S1-D) the standard deviations remain smaller than 7 lDDT units for most servers. Methods such as SWISS-MODEL and M4T are clearly focusing on high quality modeling (Table S1-D). SWISS-MODEL is closely following Robetta for the “easy”targets, but exhibiting a clear gap to the top methods for categories “medium” (Table S1 – C) and even more pronounced for “hard”. The NaiveBlast server acts as a base line. For "easy" targets, it is on average 10 lDDT units behind the best servers, but it expectedly fails for harder targets (see methods for more details).s