PDB: 6vn1

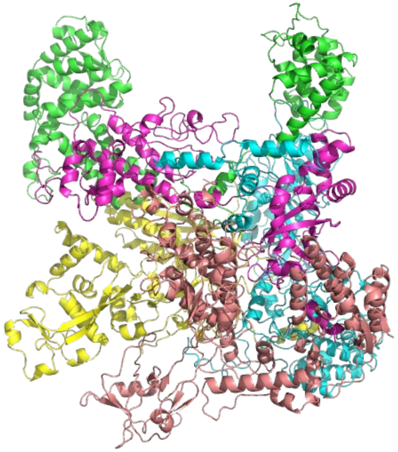
Cryo-EM Targets and their evaluation in CASP14

Andriy Kryshtafovych (UC Davis)

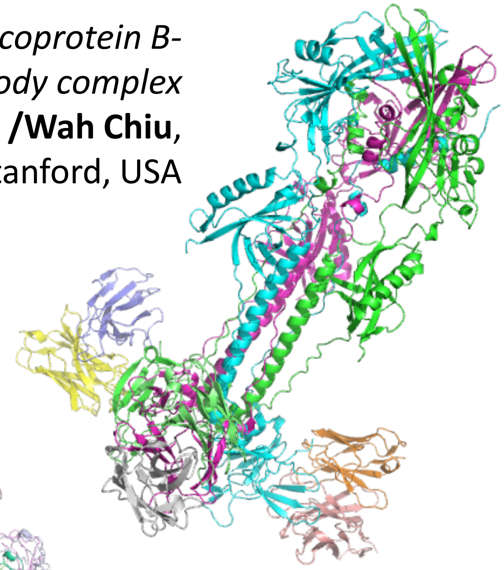
Tristan Cragnolini, Maya Topf (Birkbeck, U London)

Cryo-EM Targets in CASP14

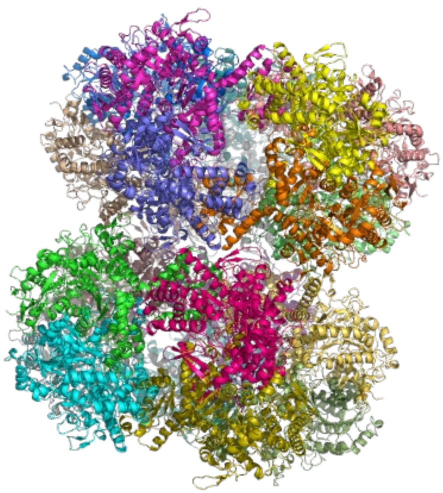
H1097 (3.8 Å) :: *AR9 RNA polymerase*
Petr Leiman,
UTexasMB, USA



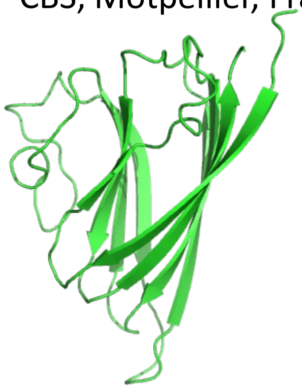
H1036 (2.8 Å) :: *glycoprotein B-neutralizing antibody complex*
Stefan Oliver /Wah Chiu,
Stanford, USA

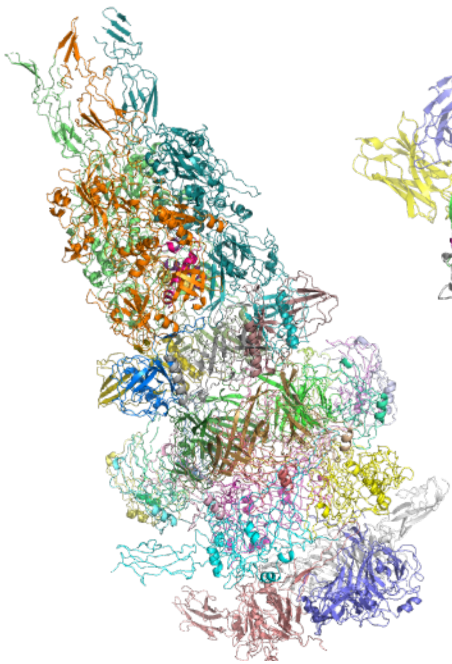


H1081 (2.1 Å) :: *decarboxylase*
Ambroise Desfosses,
IBS, Grenoble, France

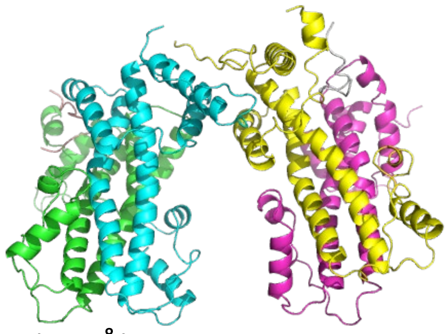


T1026 (3.2 Å) :: *faba bean necrotic stunt virus*
Stefano Trapani,
CBS, Motpellier, France



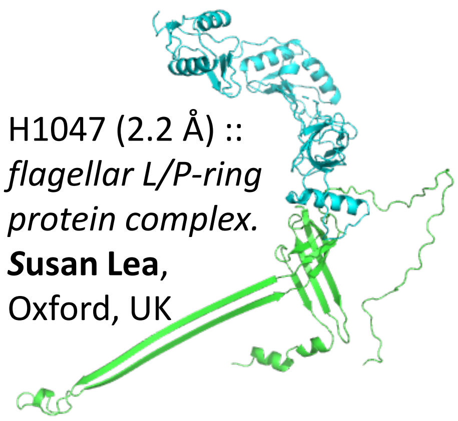


H1060 (3.2 Å) ::
T5 bacteriophage tail complex
Romain Linares,
IBS, Grenoble, France



T1099 (3.7 Å) ::
capsid of duck hepatitis B virus
Bettina Boettcher,
University of Würzburg, Germany

H1047 (2.2 Å) ::
flagellar L/P-ring protein complex.
Susan Lea,
Oxford, UK



Cryo-EM Targets
(7 complexes = 23 evaluation units)

Single-domain:

1. T1026
2. T1036s1
3. T1092-D1
4. T1092-D2
5. T1093-D1
6. T1093-D2
7. T1093-D3
8. T1094-D1
9. T1094-D2
10. T1095
11. T1096-D1
12. T1096-D2
13. T1099

Multi-domain:

1. T1092
2. T1093
3. T1094
4. T1096

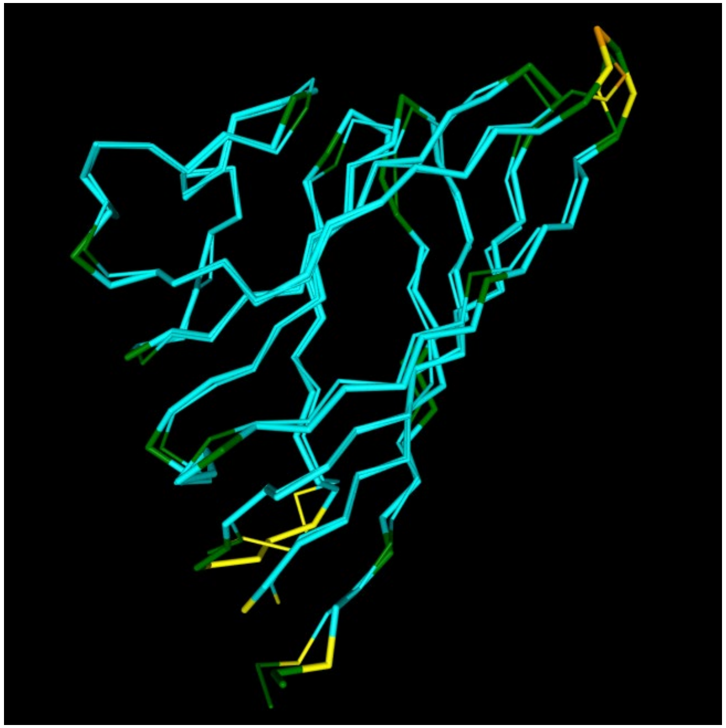
Complexes:

1. H1036 (A3B3C3 complex)
2. H1047 [T1047s1, T1047s2]
3. H1060 [T1061]
4. H1081 (A20)
5. H1097 (A1B1C1D1E1 complex)
6. T1099ov0 (A4 substructure of a 240-mer)

CASP Models on Cryo-EM targets

Evaluation vs Reference structures

T1026 (3.2 Å)
faba bean necrotic stunt virus
Stefano Trapani



T1026
TBM-hard
Best model: AlphaFold2 (TS427_1)
GDT_TS=**94**

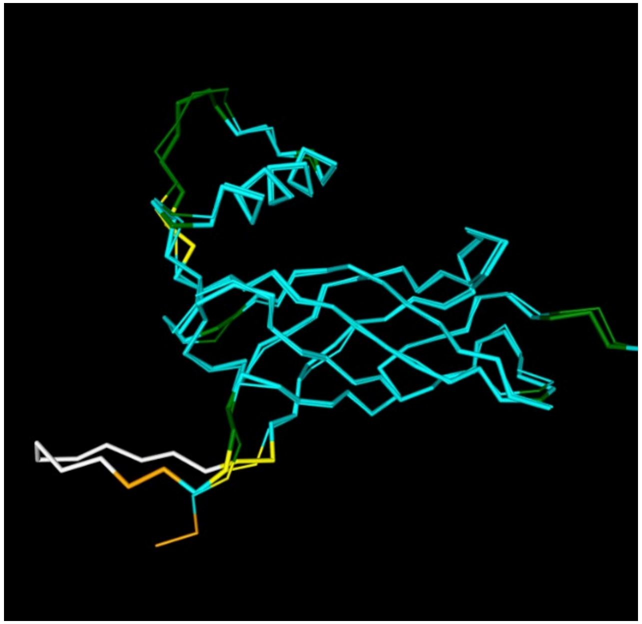
CASP Models on Cryo-EM targets

Evaluation vs Reference structures

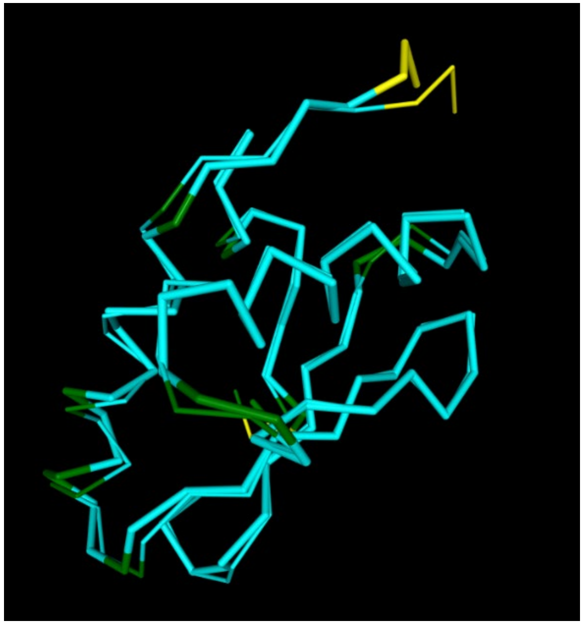
T1047s2 (2.2 Å)

flagellar L/P-ring protein complex

Susan Lea



T1047s2-D1
FM
Best model: AlphaFold2
GDT_TS(TS427_1)=**96**

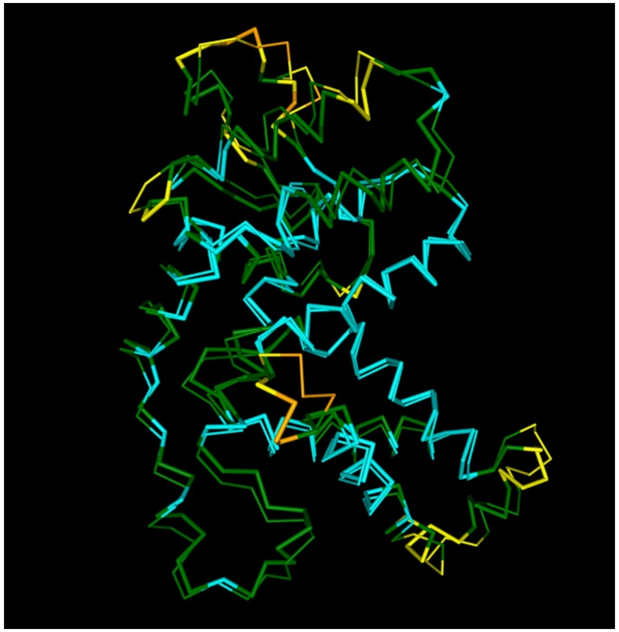


T1047s2-D2
FM/TBM
Best model: AlphaFold2
GDT_TS(TS427_2)=**95**

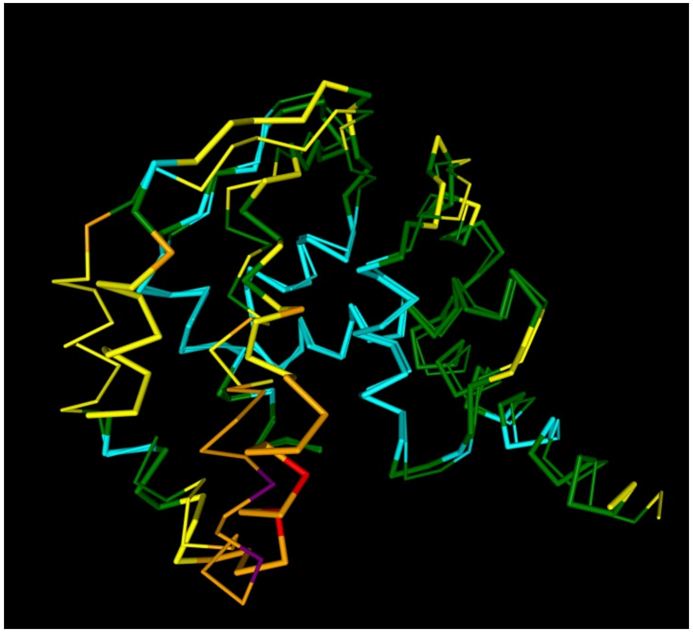
CASP Models on Cryo-EM targets

Evaluation vs Reference structures

T1096 (3.8 Å)
g226 from AR9 RNAP
Petr Leiman



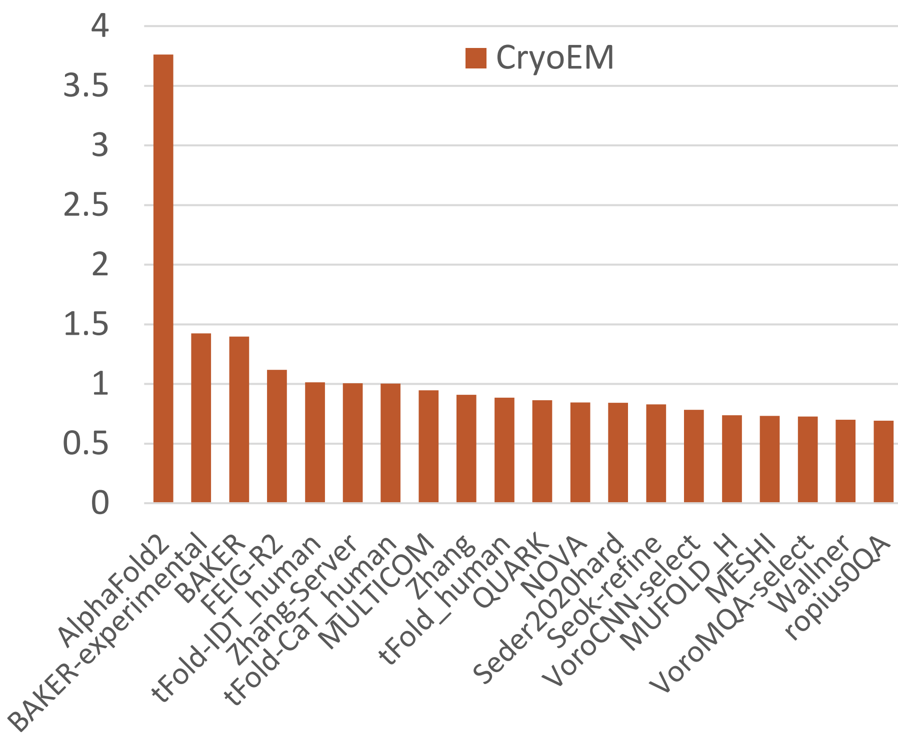
T1096-D1
FM
Best model: AlphaFold2
GDT_TS (TS427_4-D1) =**85**

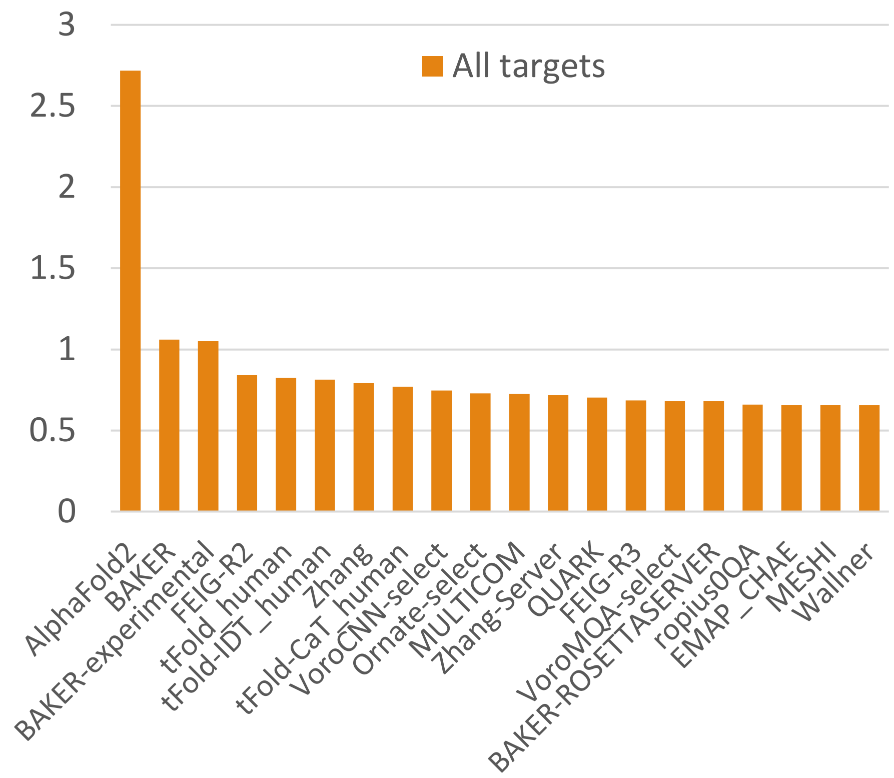


T1096-D2
FM
Best model: AlphaFold2
GDT_TS (TS427_2-D2)=**79**

CASP Models on Cryo-EM targets
Evaluation vs Reference Structures

Tertiary structure (3GDT_TS + LDDT + CADaa + SG)





Cryo-EM Targets (evaluation units)

Single-domain:

1. T1026 (59)
2. T1036s1 (77)
3. T1092-D1 (5)
4. T1092-D2 (112)
5. T1093-D1 (22)
6. T1093-D2 (5)
7. T1093-D3 (5)
8. T1094-D1 (5)
9. T1094-D2 (83)
10. T1095 (2)
11. T1096-D1 (12)
12. T1096-D2 (27)
13. T1099 (5)

Multi-domain:

1. T1092 (5)
2. T1093 (0)
3. T1094 (5)
4. T1096 (0)

Complexes:

1. H1036 (22)
2. H1081 (0)
3. H1097 (0)
4. T1099ov0 (0)
5. H1047 [T1047s1, T1047s2]
6. H1060 [T1061]

For evaluation versus maps we need 'high quality models' defined here as those scoring GDT_TS>70 (monomers) or LDDT>70 (multimers)

Outcomes of the 2019 EMDataResource model challenge: validation of cryo-EM models at near-atomic resolution

Metric Class	**Package Metric Definition**
Correlation Coefficient, all voxels	Phenix CCbox full grid map vs model-map density correlation coefficient[18] TEMPy CCC full grid map vs model-map density correlation coefficient[23]
Correlation Coefficient, selected voxels	Phenix CCmask map vs model-map density, only modelled regions[18] Phenix CCpeaks map vs model-map density, only high-density map and model regions[18] TEMPy CCC_OV map vs model-map density, overlapping map and model regions[25] TEMPy SMOC Segment Manders' Overlap, map vs model-map density, only modelled regions[25]
Correlation Coefficient, other density function	TEMPy LAP map vs model-map Laplacian filtered density (partial 2nd derivative)[22] TEMPy MI map vs model-map Mutual Information entropy-based function[22] TEMPy MI_OV map vs model-map Mutual Information, only modelled regions[25]
Correlation Coefficient, atom positions	Chimera/MAPQ Qscore map density at each modeled atom vs reference Gaussian density function[14]
Fourier Shell Correlation	Phenix FSC05 Resolution (distance) of Map-Model FSC curve read at point FSC=0.5[18] CCPEM/Refmac FSCavg FSC curve area integrated to map resolution limit[19,59]
Atom Inclusion	EMDB/VisualAnalysis AI_all Atom Inclusion, percentage of all atoms inside depositor-provided density threshold[20] TEMPy ENV Atom Inclusion in envelope corresponding to sample MW; penalizes unmodeled regions[22]
Side Chain Density	Phenix EMRinger evaluates backbone positioning by sampling map density around Cɣ-atom ring-paths for non-branched residues[21]

Lawson C, Kryshtafovych A, et al. 2020 *Nature Methods (accepted)*

CASP14 webpage – Global analysis

Assessors Home | General Discussions | Domain Definitions and Classifications | Summary on Experimental Sequences | Summary on Target Structures | Available Structures | Target List | TS Results | EMA Results | RR Results | Tarballs & Plain Files

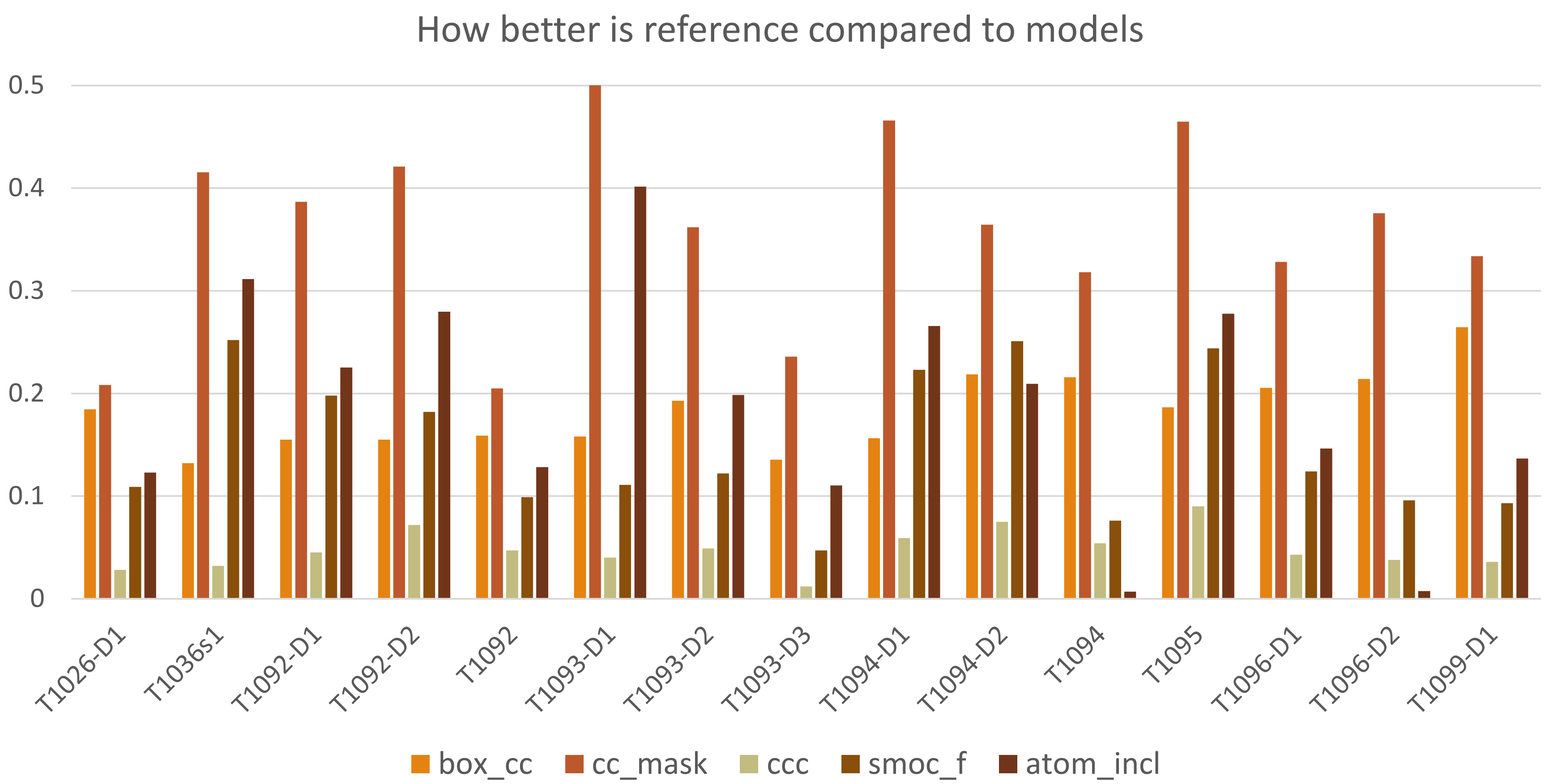
Global Scores | Per Residue Analysis

Table | Atom Inclusion Plot

Target: T1026-D1

Text

		TemPy			Phenix			EMRinger	Atom inclusion	
#	Model	CCC	MI	SMOC(d)	CC(mask)	CC(volume)	CC(peak)	EMRinger score	ALL	BB
1.	T1026-D1.pdb	0.589	0.209	0.783	0.807	0.799	0.629	3.329	0.771	0.892
2.	T1026TS427_1-D1	0.561	0.189	0.674	0.599	0.666	0.468	2.715	0.648	0.815
3.	T1026TS427_3-D1	0.561	0.190	0.666	0.586	0.662	0.459	2.026	0.645	0.805
4.	T1026TS427_4-D1	0.559	0.190	0.667	0.580	0.661	0.458	2.622	0.636	0.796
5.	T1026TS427_2-D1	0.558	0.189	0.661	0.582	0.656	0.455	2.773	0.639	0.791



How better is reference compared to models
0.5
0.4
0.3
0.2
0.1
0
T1026-D1
T1036s1
T1092-D1
T1092-D2
T1092
T1093-D1
T1093-D2
T1093-D3
T1094-D1
T1094-D2
T1094
T1095
T1096-D1
T1096-D2
T1099-D1
box_cc
cc_mask
ccc
smoc_f
atom_incl

CASP14 webpage – Local analysis
T1096