CASP16 & Cryo-EM

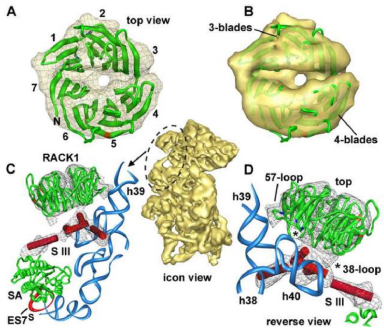
Tom Mulvaney

December, 2024

Cryo-EM and Structure prediction

Cryo-EM and cryo-ET methods have traditionally suffered from low resolutions.

- Large macromolecular assemblies determined by fitting computational structural models to low resolution data.
- Once upon a time, computational models meant homology models
- Today, computational predictions are synonymous with AlphaFold and similar ML approaches.



A ribosome at 8.7Å modelled using homology approaches

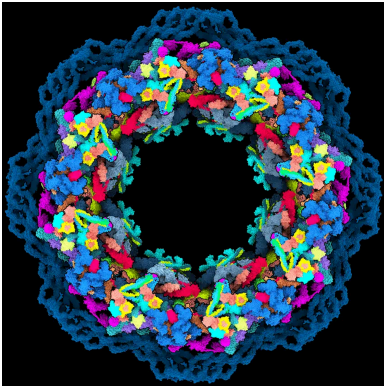
Chandramouli P, Topf M, Ménétret JF, Eswar N, Cannone JJ, Gutell RR, Sali A, Akey CW. Structure of the mammalian 80S ribosome at 8.7 A resolution. Structure. **2008** doi:10.1016/j.str.2008.01.007

CSSB Centre for Structural Systems Biology

Cryo-EM and Structure prediction

Cryo-EM and cryo-ET methods have traditionally suffered from low resolutions.

- Large macromolecular assemblies determined by fitting computational structural models to low resolution data.
- Once upon a time, computational models meant homology models
- Today, computational predictions are synonymous with AlphaFold and similar ML approaches.



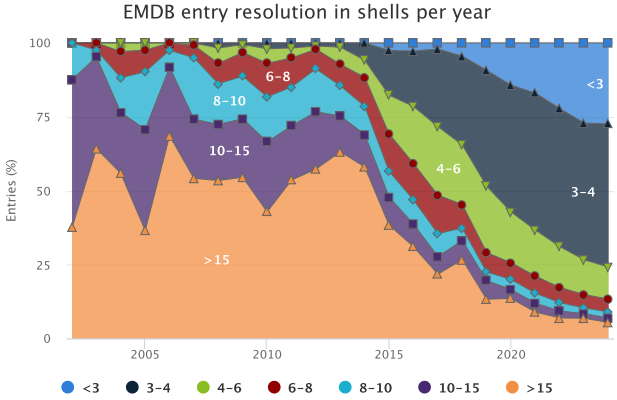
The Nuclear Pore Complex modelled using cryo-ET data AF2 models

Mosalaganti S, Obarska-Kosińska A, Siggel M et al. Science, 2022 doi:10.1126/science.abm9506

CSSB Centre for Structural Systems Biology

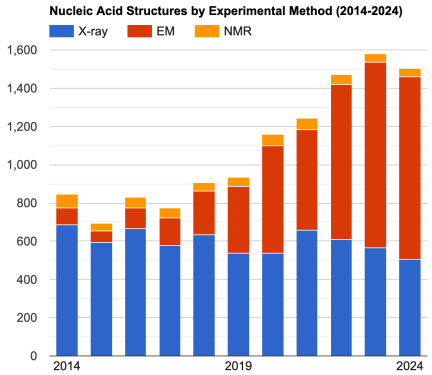
LIV

Cryo-EM and Structure prediction



Resolution statistics from EMDB

The tables have turned



Statistics from the Nucleic Acid Knowledge Base (nakb.org)

The EM community was once a downstream user of computation predictions. The "resolution revolution" has changed things.

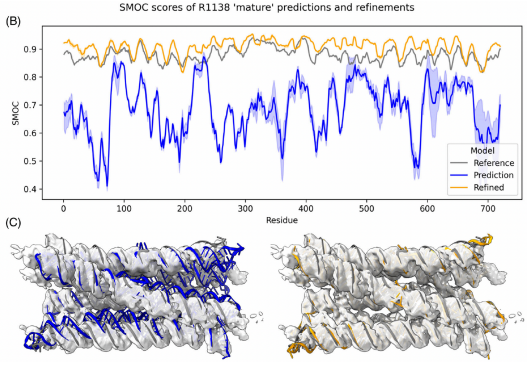
- Now possible to derive models at "near atomic" resolutions
- Cryo-EM models have been targets in recent CASPs.
- Under-represented molecules - e.g. membrane proteins

CSSB Centre for Structural Systems Biology

Assessment of cryo-EM targets

Cryo-EM derived targets are assessed as per the standard assessment criteria. But we also encourage experimentalists to provide experimental data to aid the assessment.

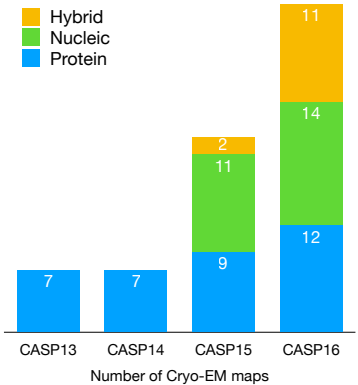
- Model / Map assessment eg. Qscore, SMOC, CC
- "Practical" docking / flexible fitting / refinement of models
- Uncovering modelling errors or ambiguity



Refinement of an RNA prediction from CASP15.
Mulvaney T, Kretsch RC, Elliott L, et al. CASP15 cryo-EM protein and RNA targets: Refinement and analysis using experimental maps. Proteins. 2023;91(12):1935-1951. doi:10.1002/prot.26644

CSSB Centre for Structural Systems Biology

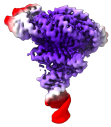
CASP16 Cryo-EM in numbers



This year, experimentalists have kindly provided...

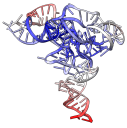
- 37 experimental datasets
- 28 GB data
- Dominated by nucleic acid datasets

Resolution vs RMSD for predictions



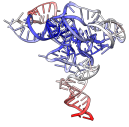
Cryo-EM map of coloured by local resolution

Five models superimposed



Target coloured by local resolution

$$RMSD_i = \sqrt{\sum_{n=0}^{5} \left(\overline{x_i} - x_{n,i}\right)^2}$$

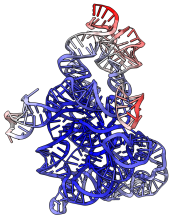


Target coloured by atom RMSD

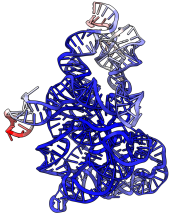
CSSB Centre for Structural Systems Biology

LIV

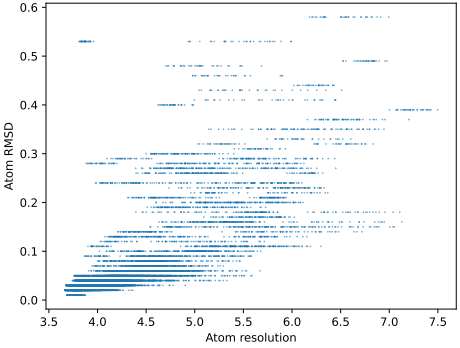
Resolution vs RMSD for predictions - R1241



Target coloured by local resolution



Target coloured by RMSD of predictions from *RNAFOLDX*

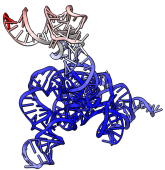


$R^2 = 0.76$

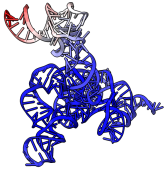
CSSB Centre for Structural Systems Biology

LIV

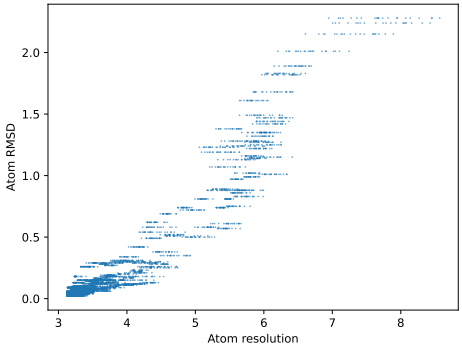
Resolution vs RMSD for predictions - R1289



Map coloured by local resolution



Target colored by RMSD of *Kihara* (294)

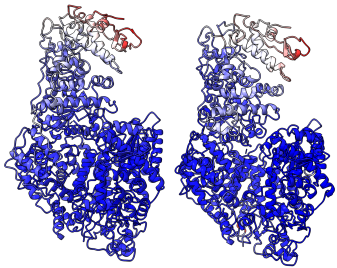


$R^2 = 0.95$

T1220s1 - Resolution vs RMSD

Groups with best model 1 according to GDT_TS

GDT_TS	Group	Mean/Min/Max (Å)	R^2
69.14	HYU_MLLAB	0.1/0.01/0.68	0.69
67.58	CSSB_experimental	0.23/0.01/1.63	0.80
66.70	PEZYFoldings	0.04/0.01/0.28	0.85

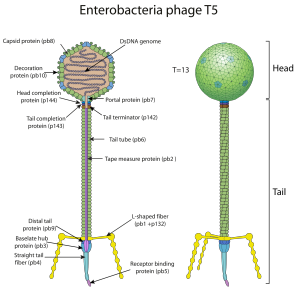


Local resolution (left),
RMSD PEZYFoldings (right)

T1257

Protein complex from the L-tail of the T5 phage.

- Trimer made up of 1263 AA monomers.
- Small error at end terminus.
- Issue highlighted by better fitting predictions.
- Refinement related error.



Schematic of the T5 phage.
Source: ViralZone, SIB Swiss Institute of Bioinformatics, Philippe Le Mercier et al. - VIRION https://viralzone.expasy.org/511 Tequintavirus, CC BY 4.0

T1257

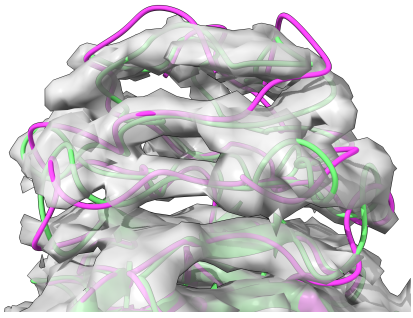
Protein complex from the L-tail of the T5 phage.

- Trimer made up of 1263 AA monomers.
- Small error at end terminus.
- Issue highlighted by better fitting predictions.
- Refinement related error.

T1257

Protein complex from the L-tail of the T5 phage.

- Trimer made up of 1263 AA monomers.
- Small error at end terminus.
- Issue highlighted by better fitting predictions.
- Refinement related error.

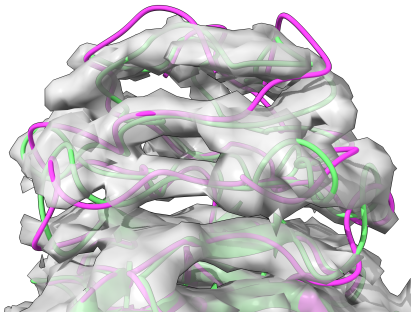


Target (magenta), a prediction (green)

T1257

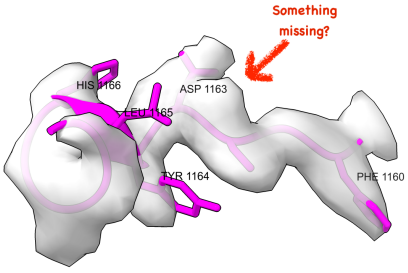
Protein complex from the L-tail of the T5 phage.

- Trimer made up of 1263 AA monomers.
- Small error at end terminus.
- Issue highlighted by better fitting predictions.
- Refinement related error.

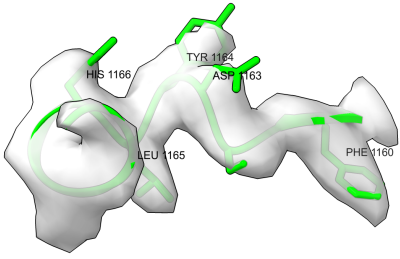


Target (magenta), a prediction (green)

T1220s1 - Fragment 1: 1160-1166

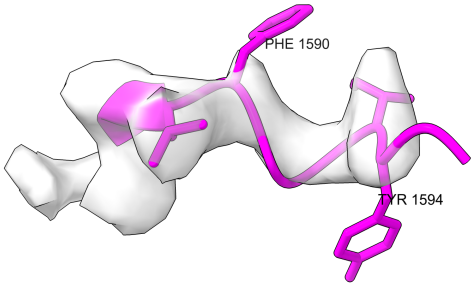


Target

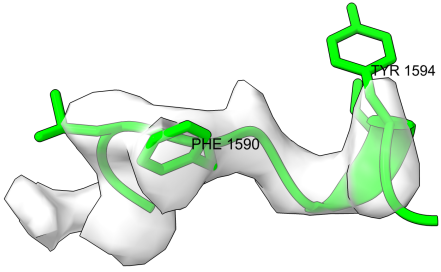


Prediction TS015_1

T1220s1 - Fragment 1: 1588-1596



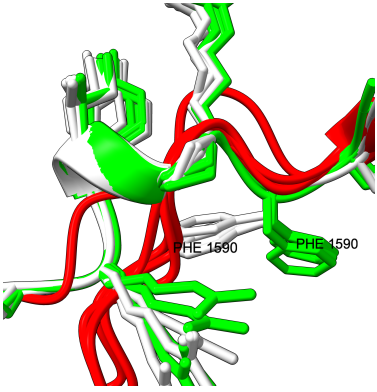
Target



Prediction TS015_1

T1220s1 - Fragment 1: 1588-1596

Rank (GDT)	Model	Group
2	423_5	ShanghaiTech-server
6	023_5*	FTBiot0119
8	198_2	colabfold
9	286_3*	CSSB_experimental
10	286_1	CSSB_experimental
12	293_4*	MRAH
13	079_4*	MRAFold
14	489_2*	Fernandez-Recio



*Correct *PHE:1590* rotamer (green). Rest of Top 10 (red), were inconsistent

Conclusions

- The notion of discrete classes in cryo-EM is an approximation.
  - Variable resolution is one way conformational heterogeneity manifests itself.
  - Conformational variability within predictions from groups reflects this.
- Model building is a difficult task
  - Computational predictions were essential in the blobology era.
  - Good quality predictions still relevant for high resolution data sets.
- We need experimental data sooner rather than later to ensure our targets are accurate. Everyone benefits.

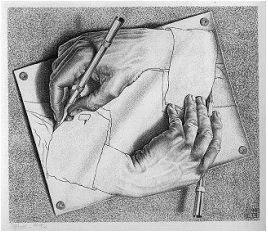
Acknowledgements

CASP Organizers

- John Moult
- Krzysztof Fidelis
- Andriy Krystafovich
- Torsten Schwede
- Maya Topf

CASP Participants

- Target providers
- Cryo-EM data providers
- Predictors



CASP and cryo-EM communities helping each other.
M.C. Escher.