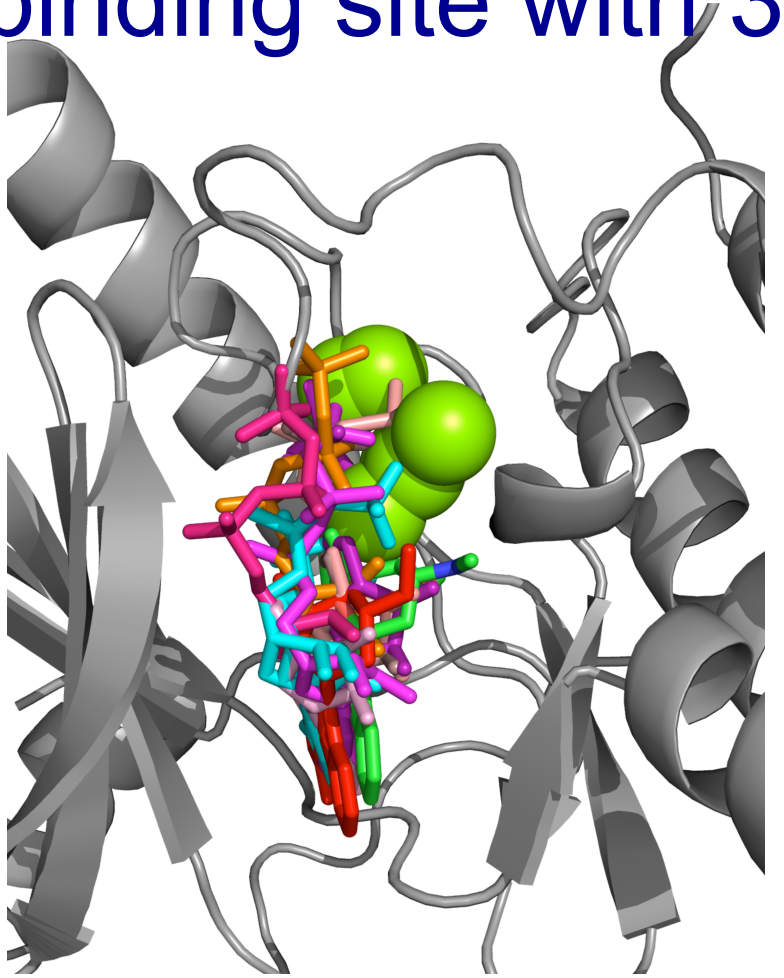
Modelling binding site with 3DLigandSite



Mark Wass
m.n.wass@kent.ac.uk

CASP

MEEYKVVVCGSGPVALGCF Target sequence
(2 per day)

Human
predictors

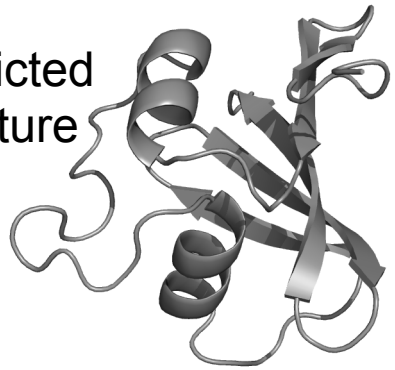
Server
predictors

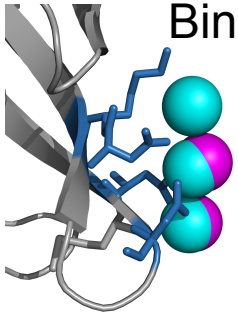
3 days

3 weeks

Predicted
Binding site

Predicted
structure



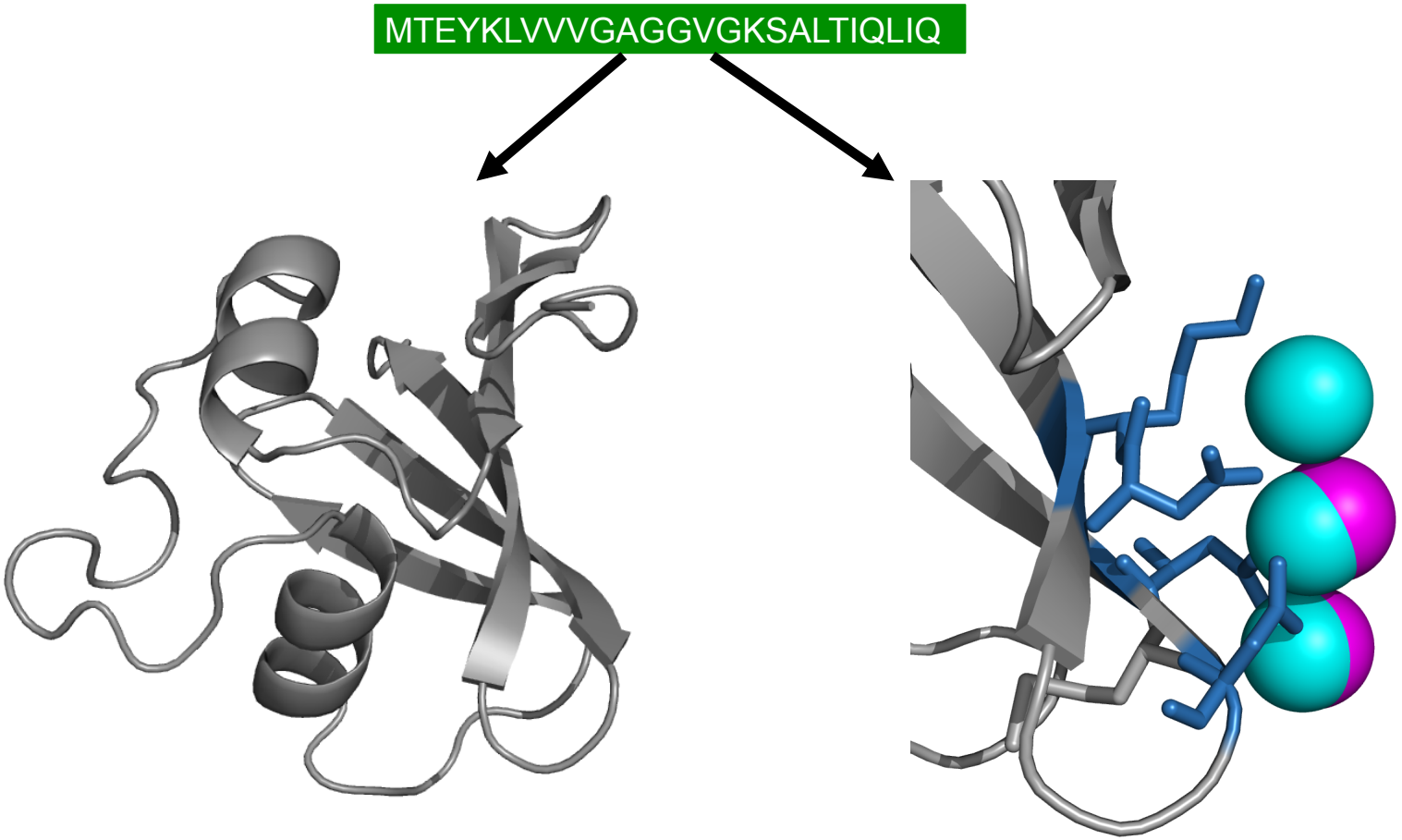


Assessment

Results
Performance
Compared to
other groups

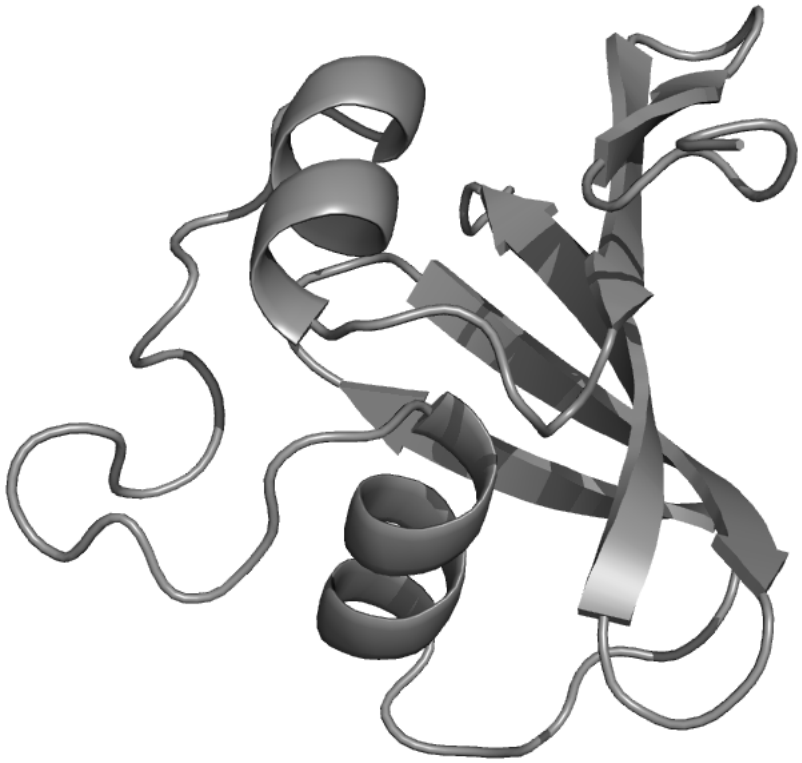
3DLigandSite

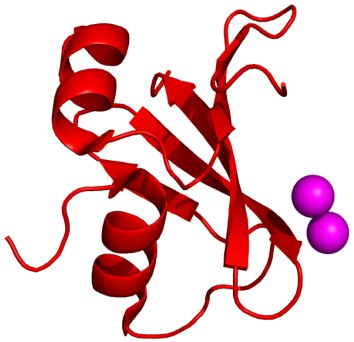
Developed as a result of participation in CASP



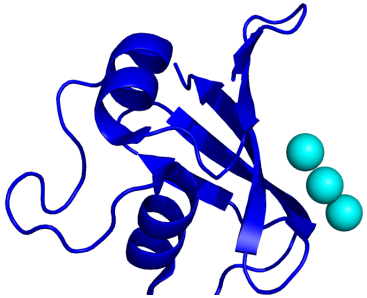
3DLigandSite

Homologous structures

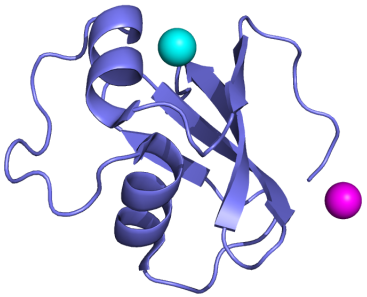




2oai



2pls



2p4p

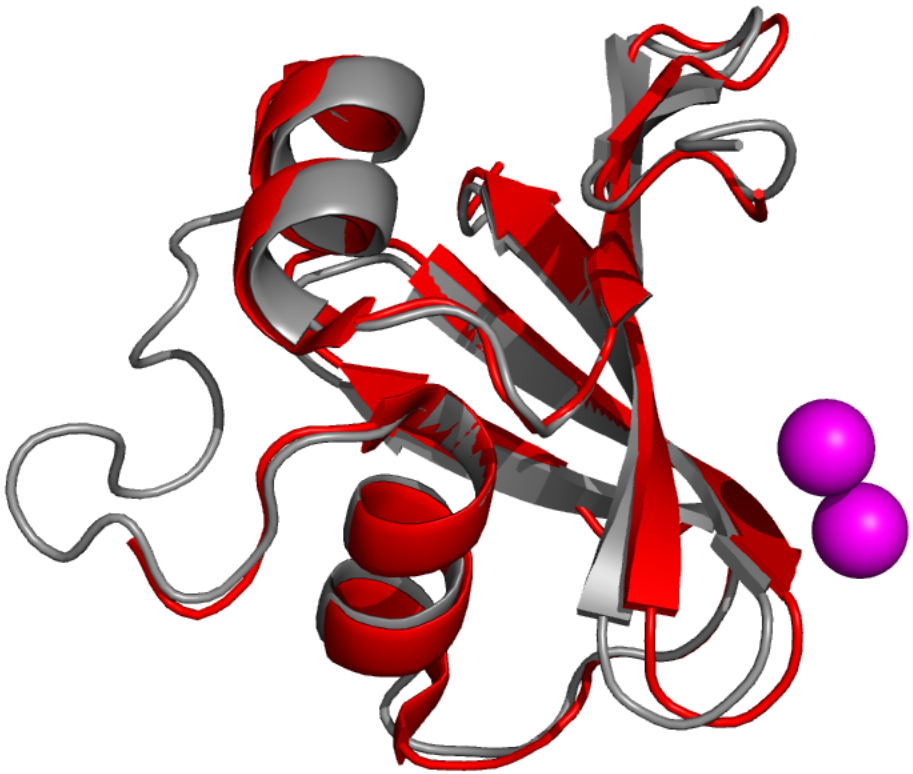
Magnesium

Calcium

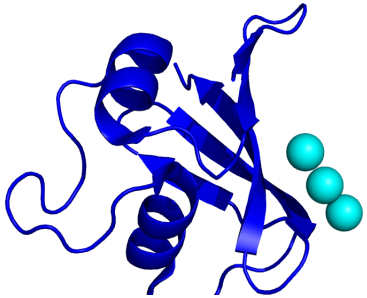
Wass & Sternberg *Proteins* 2009

3DLigandSite

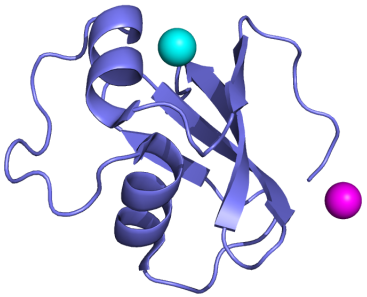
Homologous structures



2oai



2pls



2p4p

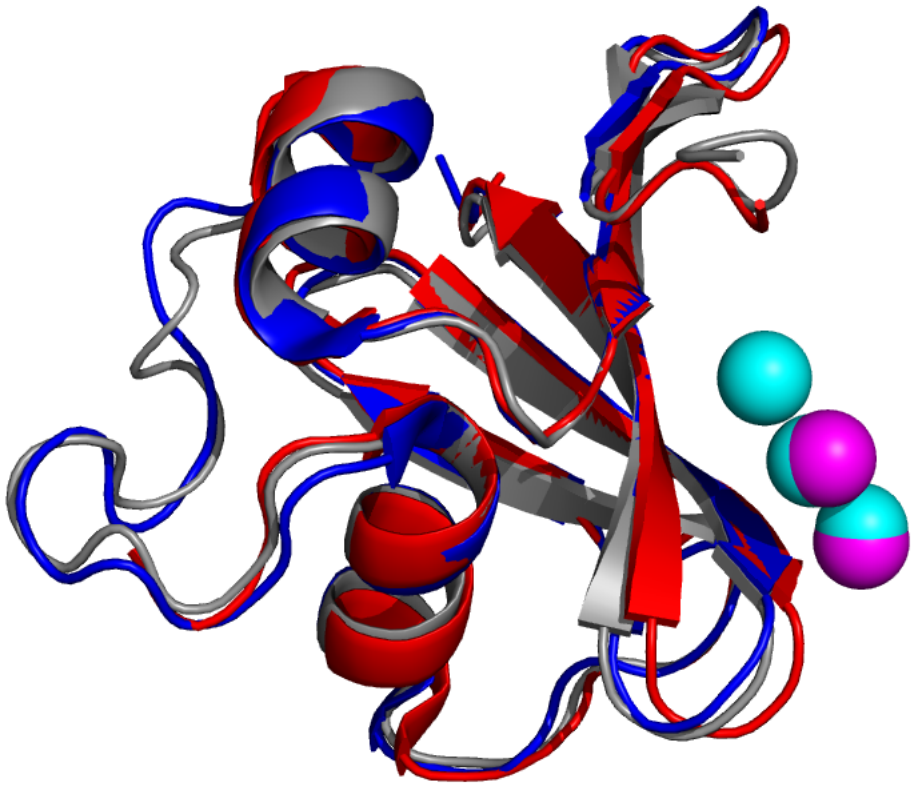
Magnesium

Calcium

Wass & Sternberg *Proteins* 2009

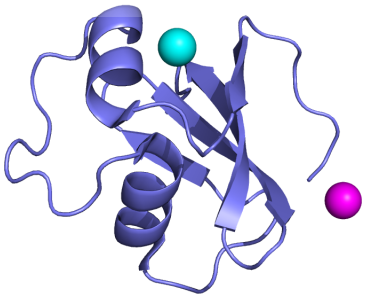
3DLigandSite

Homologous structures



2oai

2pls



2p4p

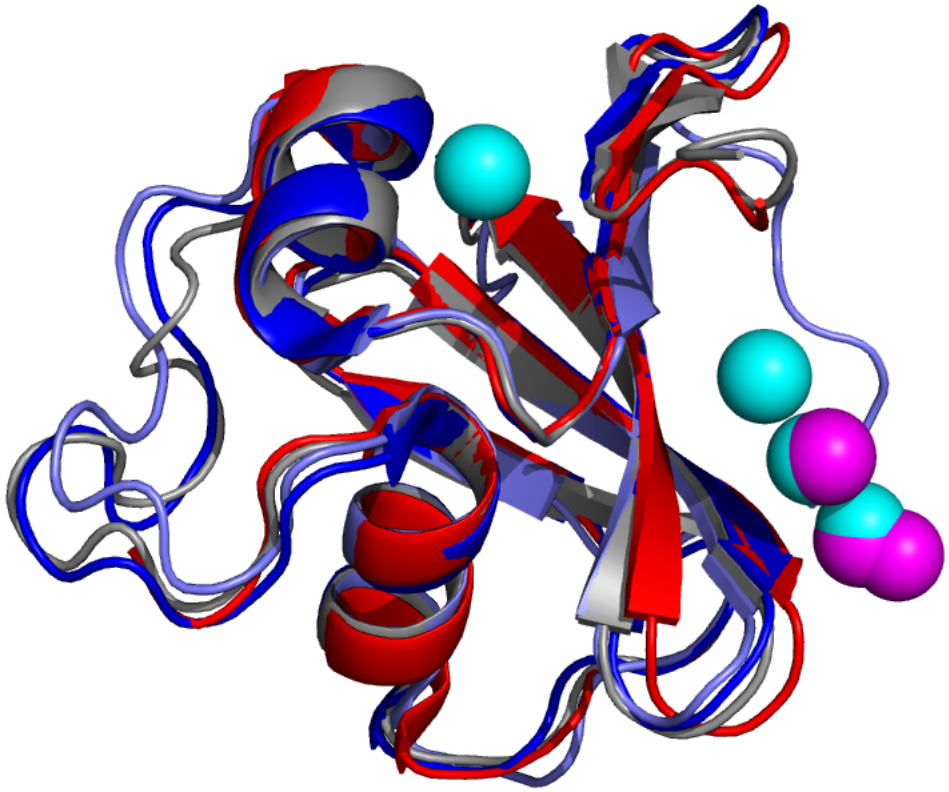
Magnesium

Calcium

Wass & Sternberg *Proteins* 2009

3DLigandSite

Homologous structures

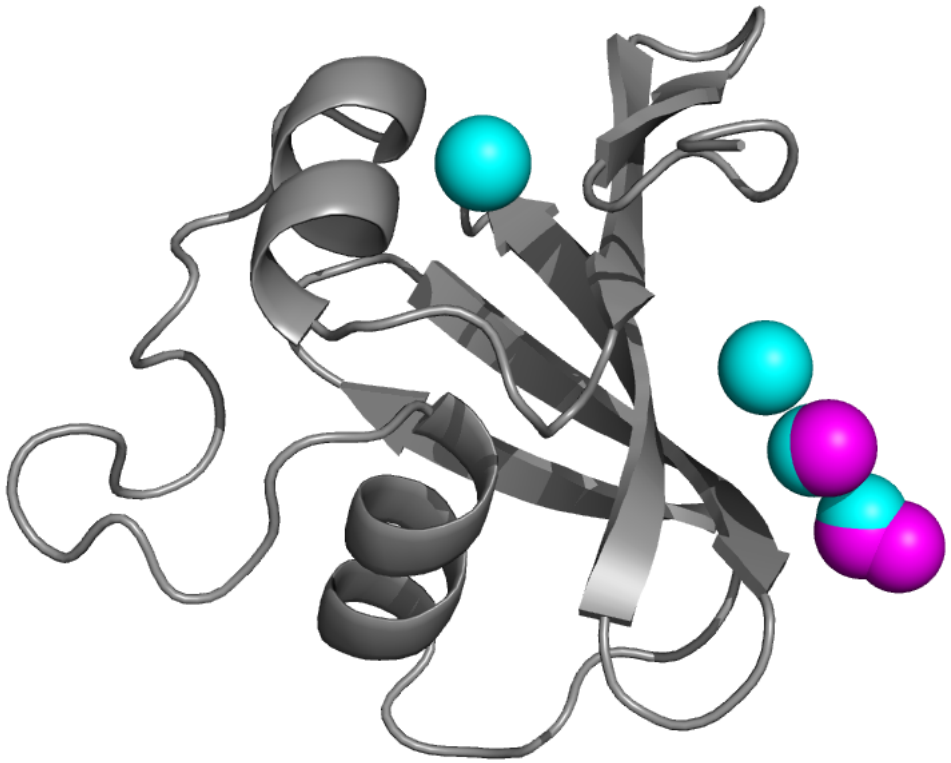


Magnesium Calcium

Wass & Sternberg *Proteins* 2009

3DLigandSite

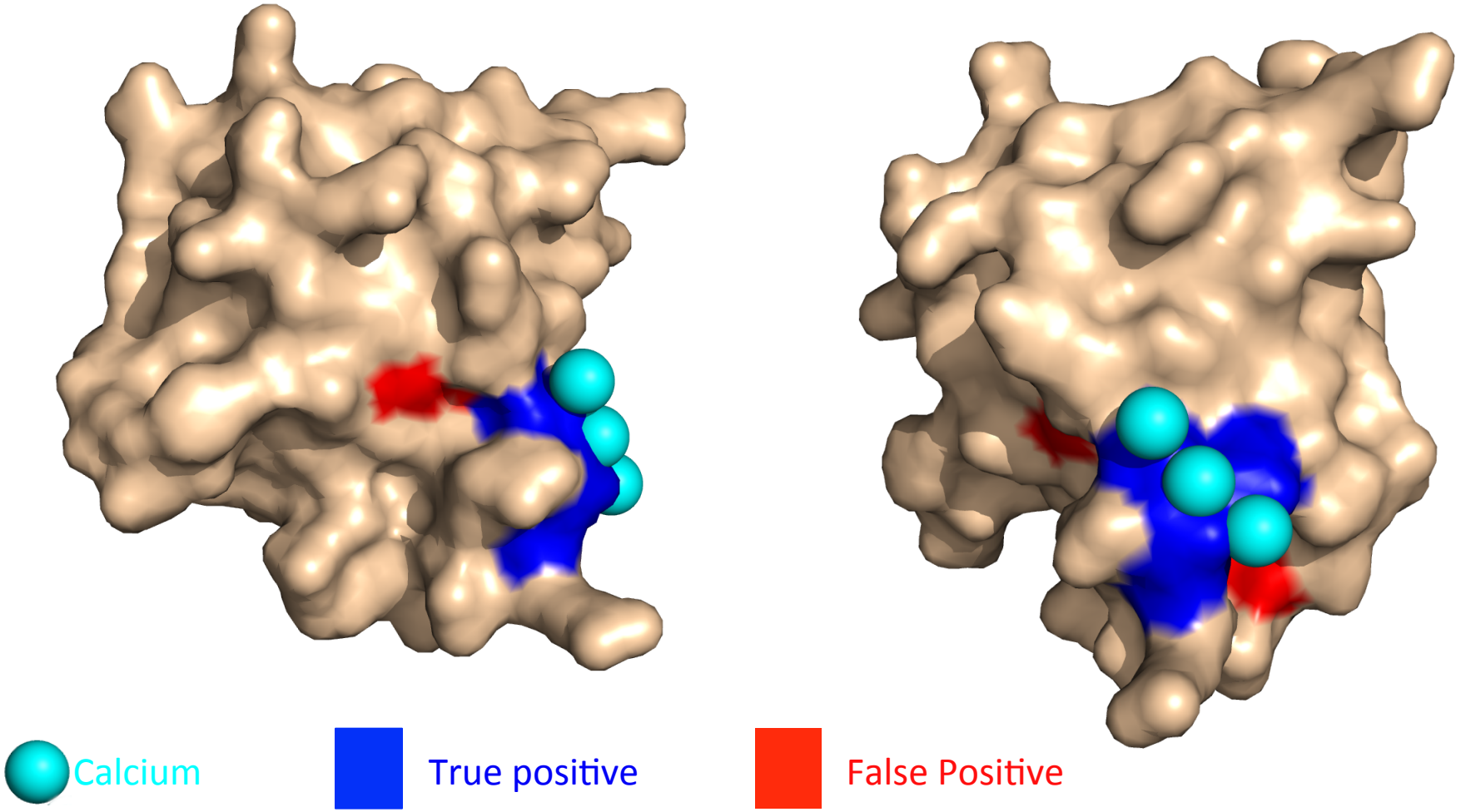
Homologous structures



Calcium Magnesium

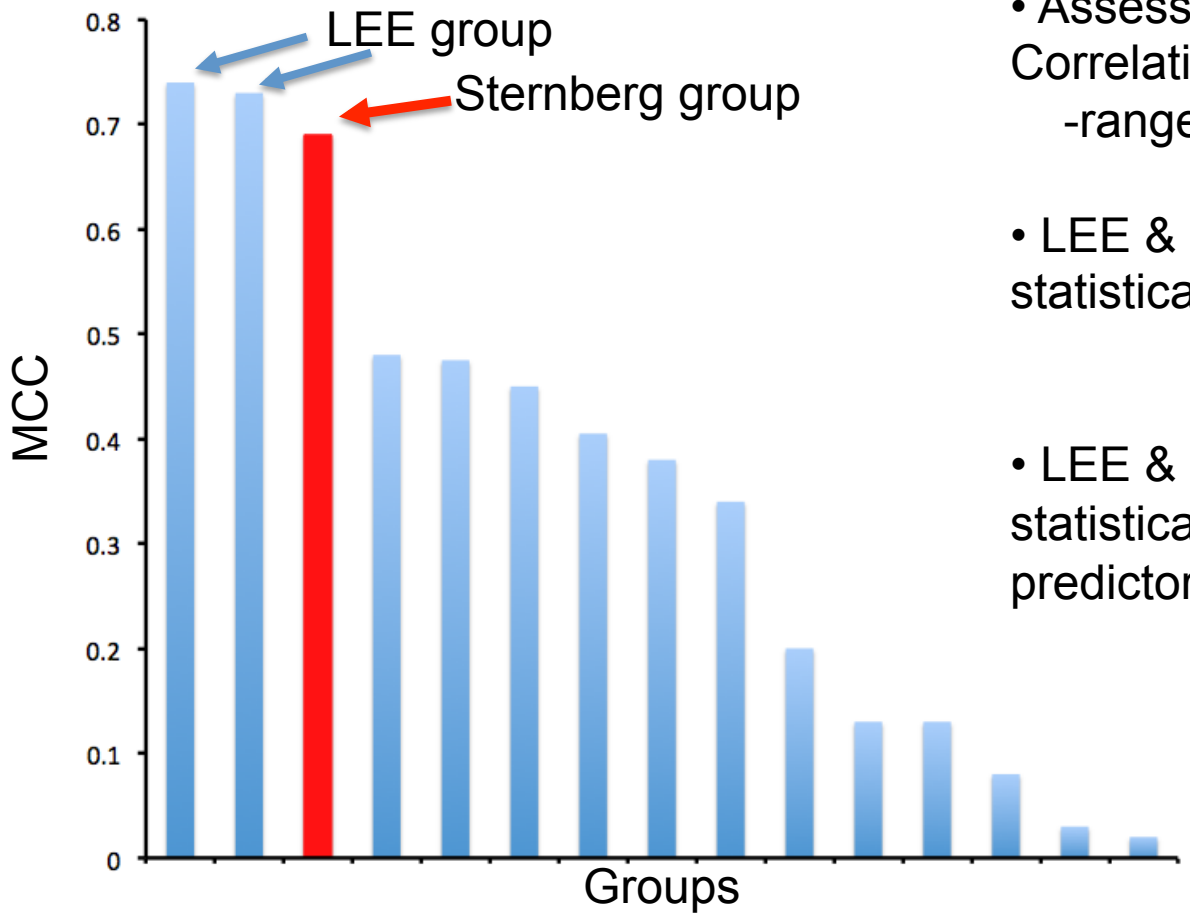
Wass & Sternberg *Proteins* 2009

3DLigandSite



Wass & Sternberg *Proteins* 2009

Performance at CASP8



- Assessed using Matthews' Correlation coefficient –
  -range -1 – +1

- LEE & Sternberg not statistically different.

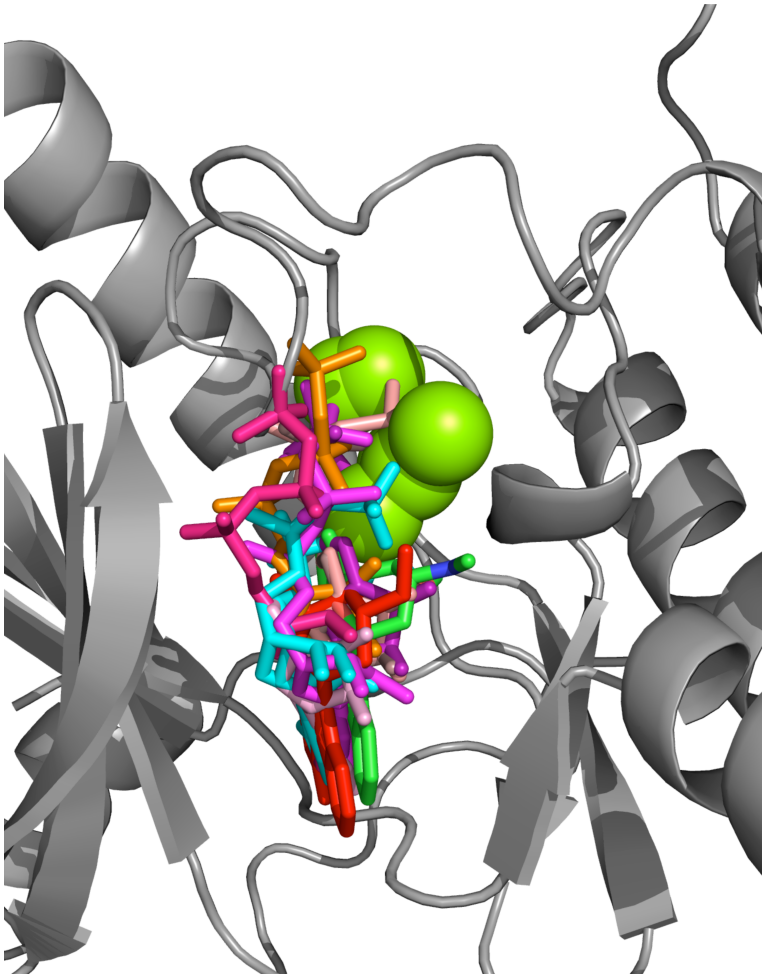
- LEE & Sternberg are statistically different to all other predictors ($p < 0.01$).

Adapted from Lopez et al., 2009

3DLigandSite

Automating our CASP8 approach

Structural View of Prediction



Jmol

Display Modification

Whole protein

colour by: ⦿ prediction ○ Jensen Shannon Divergence

spacefill: ⦿ off ○ 20% ○ 100%

wireframe: ⦿ off ○ wireframe ○ wireframe 50 ○ wireframe 100

☑ cartoon

Predicted residues

spacefill: ⦿ off ○ 20% ○ 100%

wireframe: ⦿ off ○ on ○ wireframe 50 ○ wireframe 100

☑ cartoon
☐ label

Heterogens

Display of Metalic heterogens
spacefill: ○ off ○ 20% ⦿ 100%

Display of Non Metalic heterogens
spacefill: ⦿ off ○ 20% ○ 100%

wireframe: ○ off ○ standard ○ wireframe 50 ⦿ wireframe 100

View

Reset to original orientation

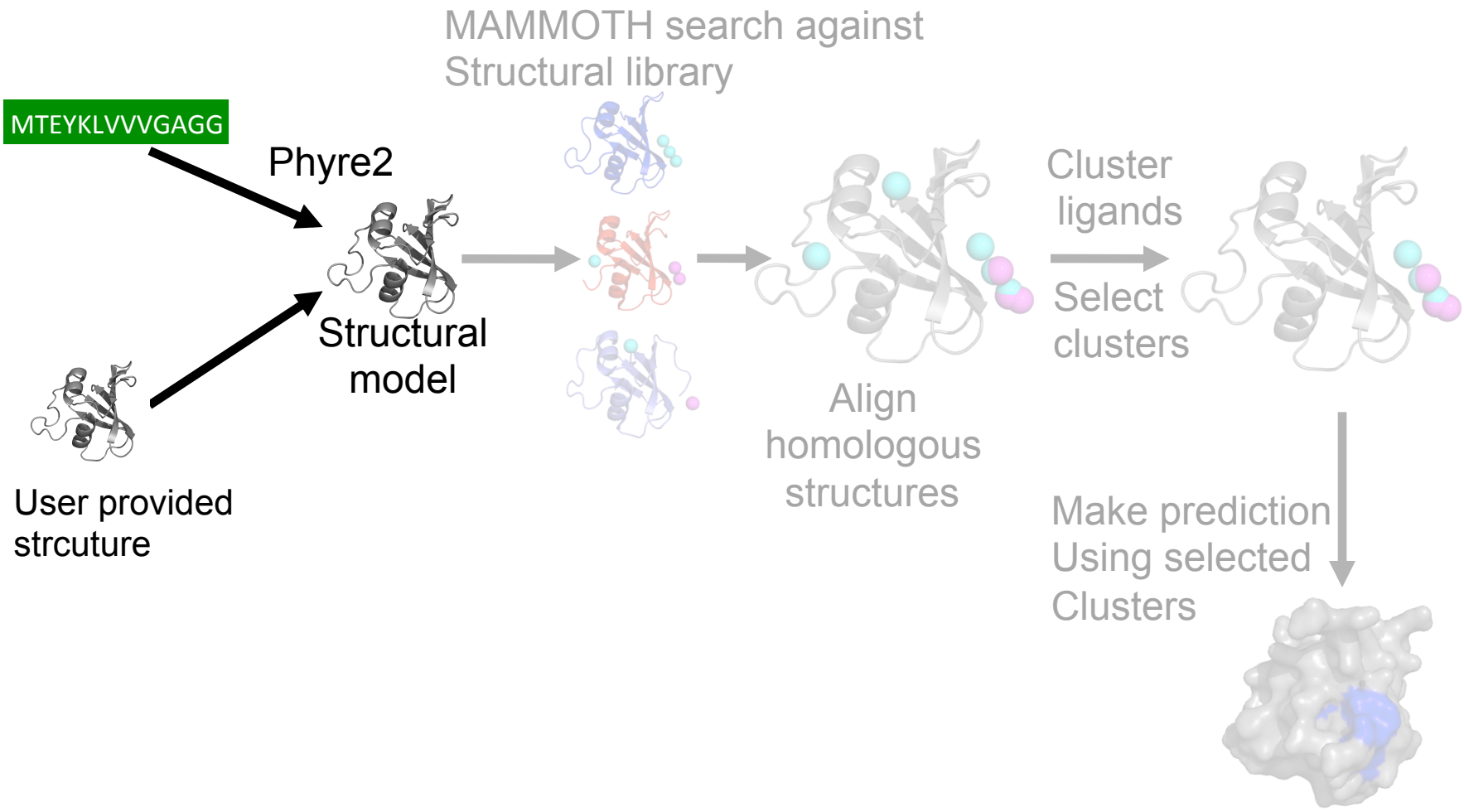
☐ spin

background black

Prediction colour legend:	Other residues	Predicted Binding Site			
Conservation Score Colour legend:	0-0.15	0.16-0.30	0.31-0.40	0.41-0.50	
	0.51-0.60	0.61-0.70	0.71-0.80	0.81-1.00	

Wass et al., NAR 2010

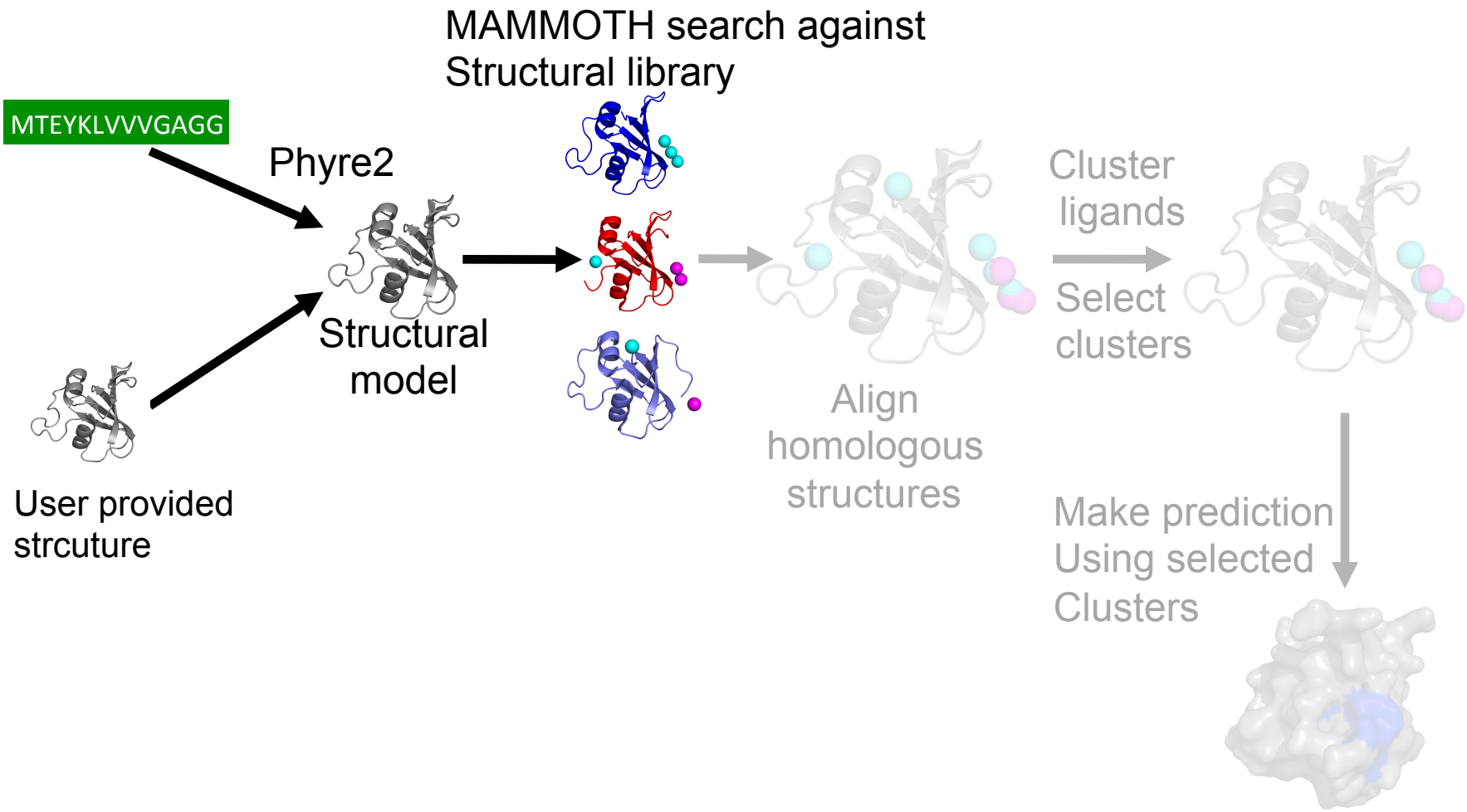
3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

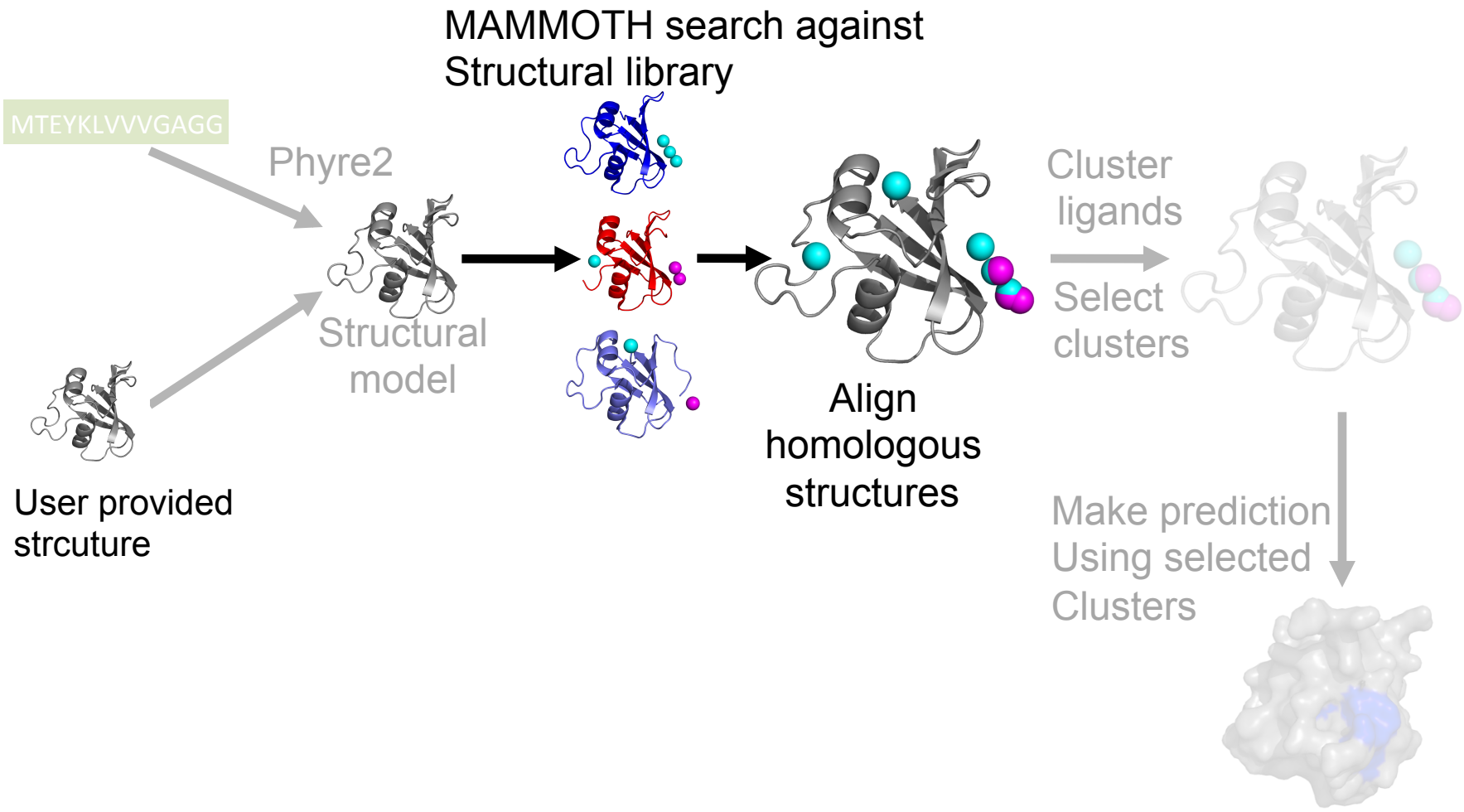
3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

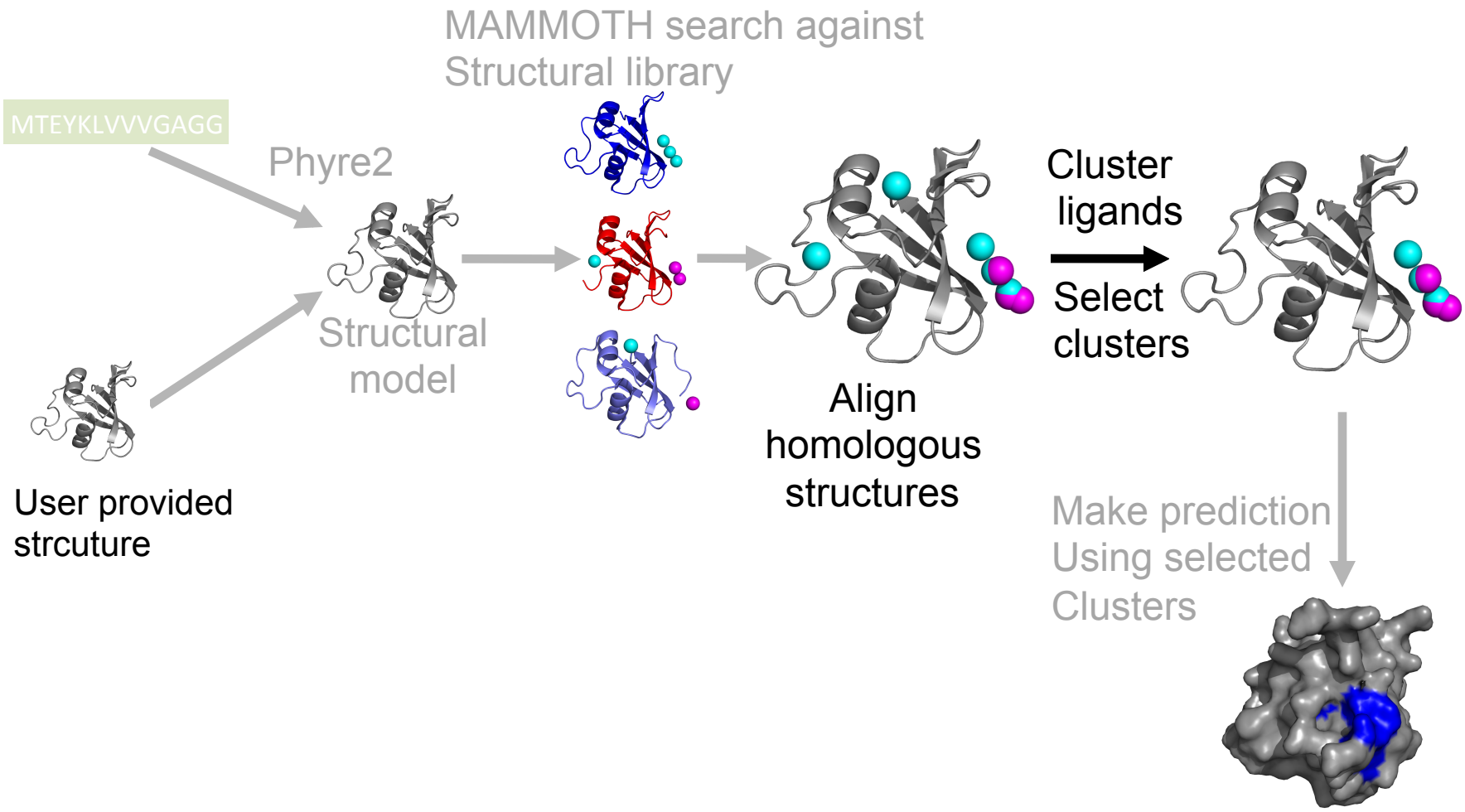
3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

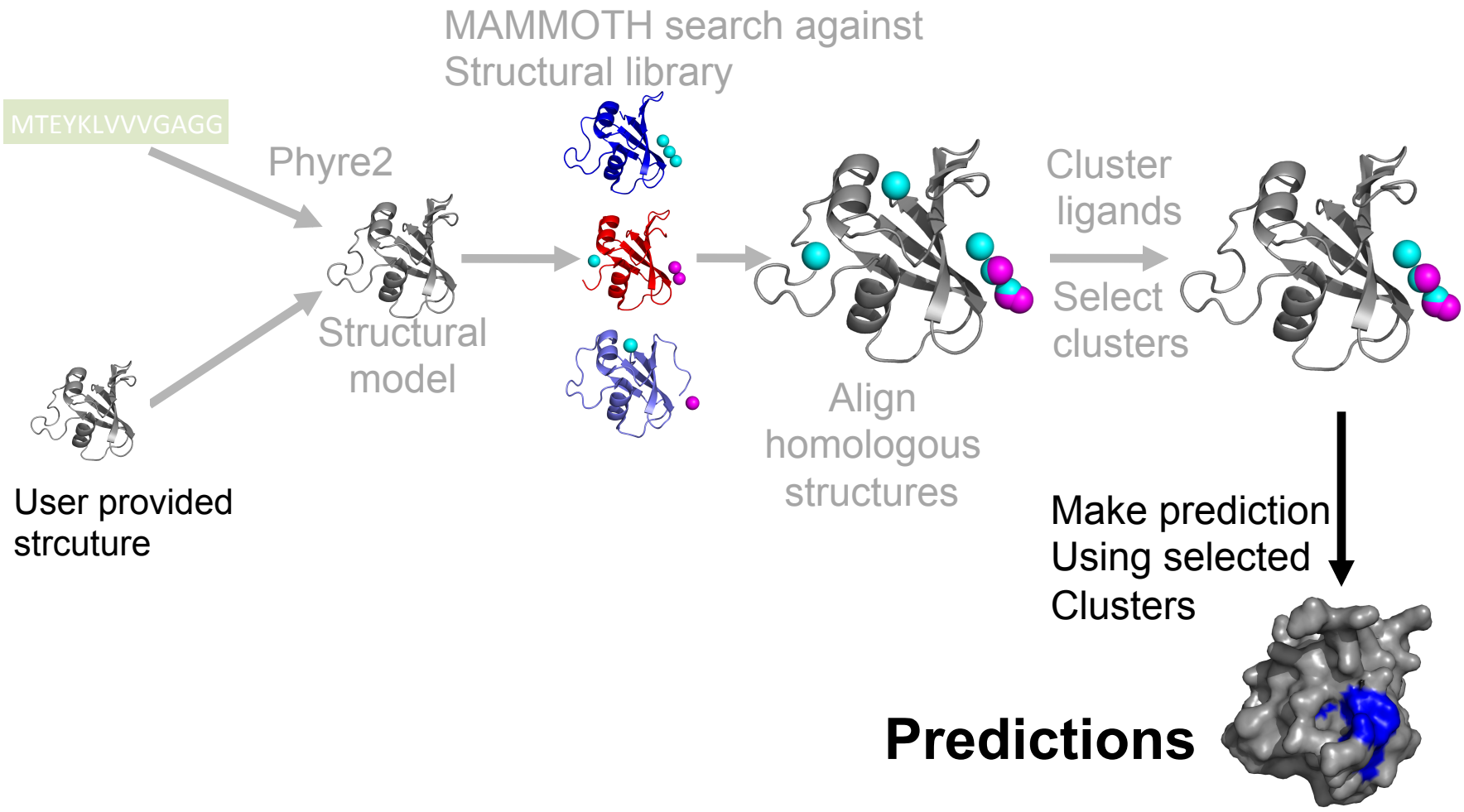
3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

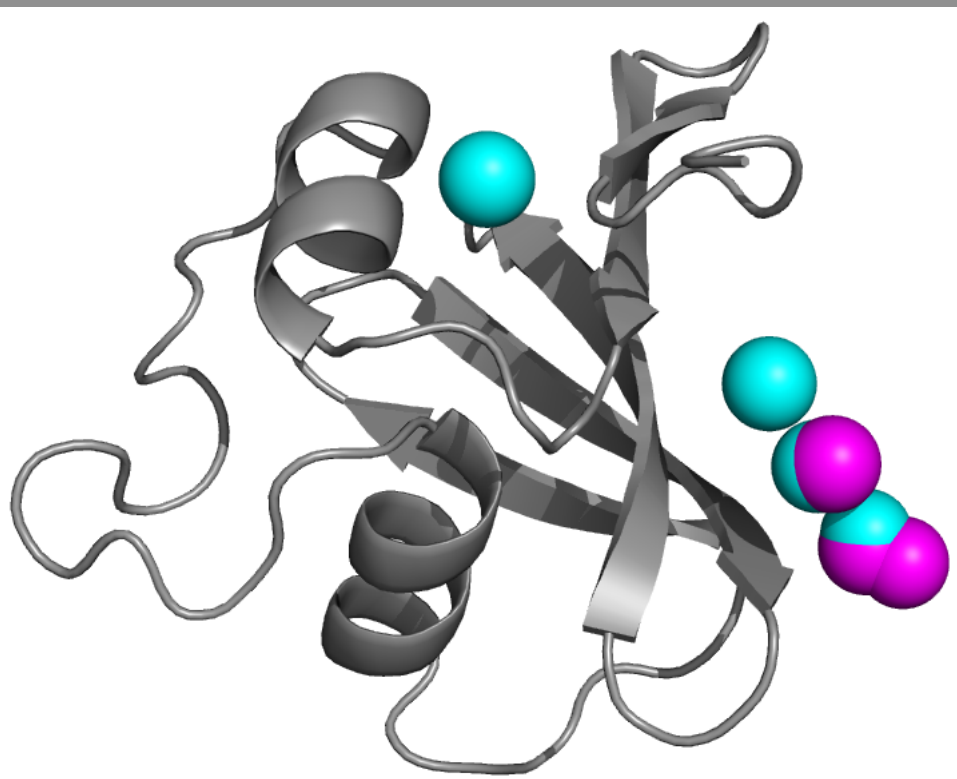
3DLigandSite

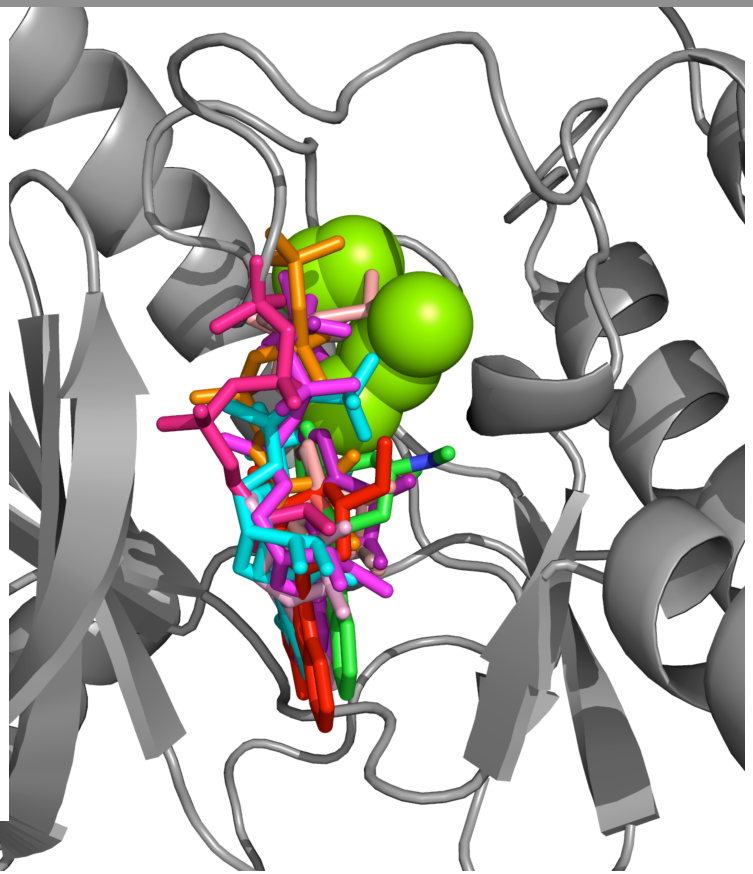


Wass *et al*., NAR 2010

Imperial College London

Predicting contacting residues





Multiple molecules in cluster but where is the actual binding site?

Threshold for prediction = Contact 25% ligands

3DLigandSite Benchmarking

FINDSITE set (617)

Measure	3DLigandSite
MCC	0.68
Recall	70%
Precision	70%

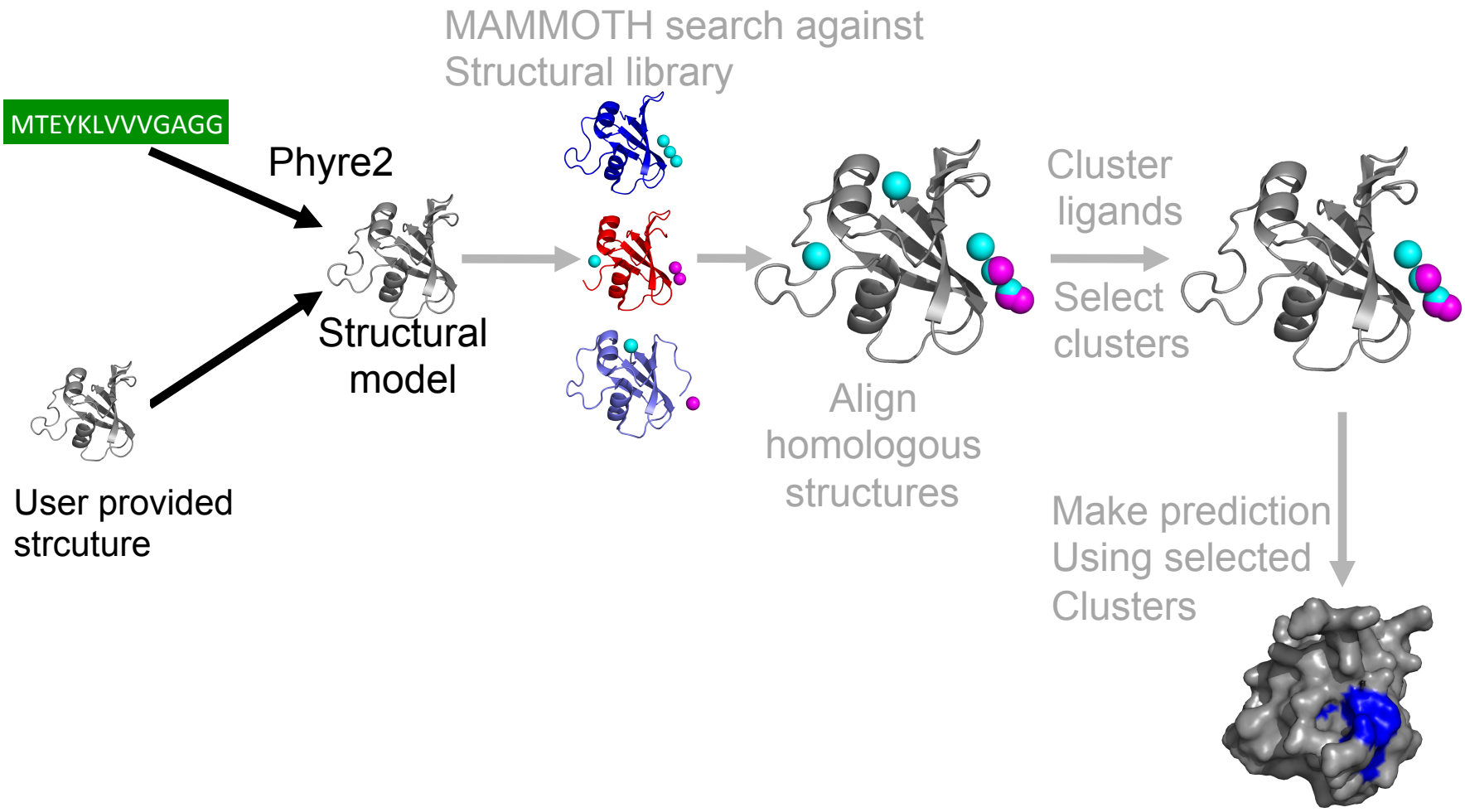
CASP8 targets (28)

Measure	3DLigandSite	Human CASP8
MCC	0.64	0.63
Recall	71%	83%
Precision	60%	56%

MCC – Matthews Correlation Coefficient
Recall– percentage of binding sites that are predicted (TP/(TP+FN))
Precision– percentage of predicted residues that are correct (TP/(TP+FP))

Using 3DLigandSite

3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

Homepage - submission

Paste your sequence here:

Paste sequence and Run

email address: optional

Job description: optional

Reset 3dligandSite search

Or submit your own structure

click here to submit your own protein structure (i.e. a pdb file)

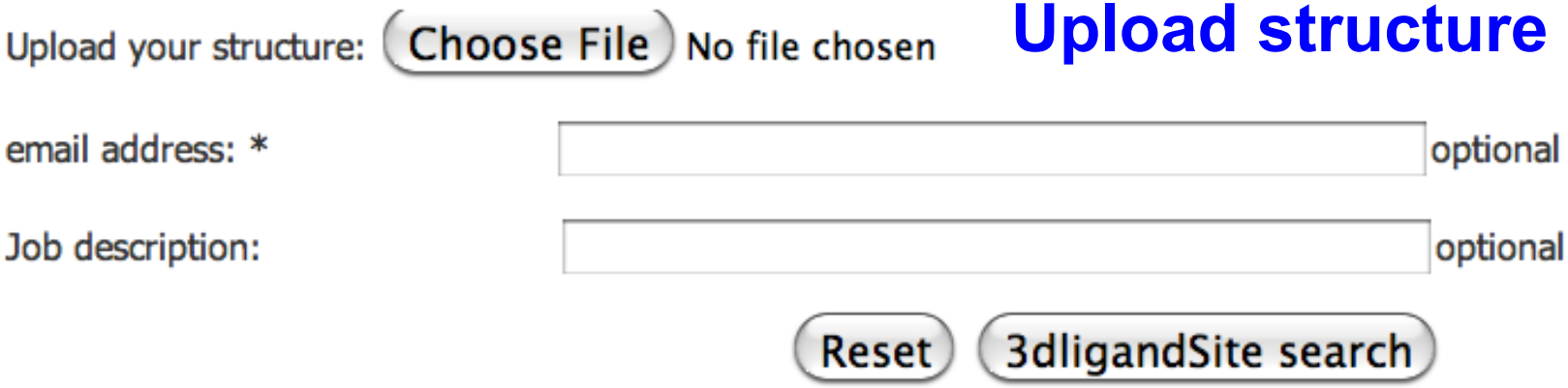
Retrieve results

3DLigandSite Results

Enter job id to view 3DLigandSite results: Fetch

About News Help FAQ Example Contact Disclaimer

Homepage - submission



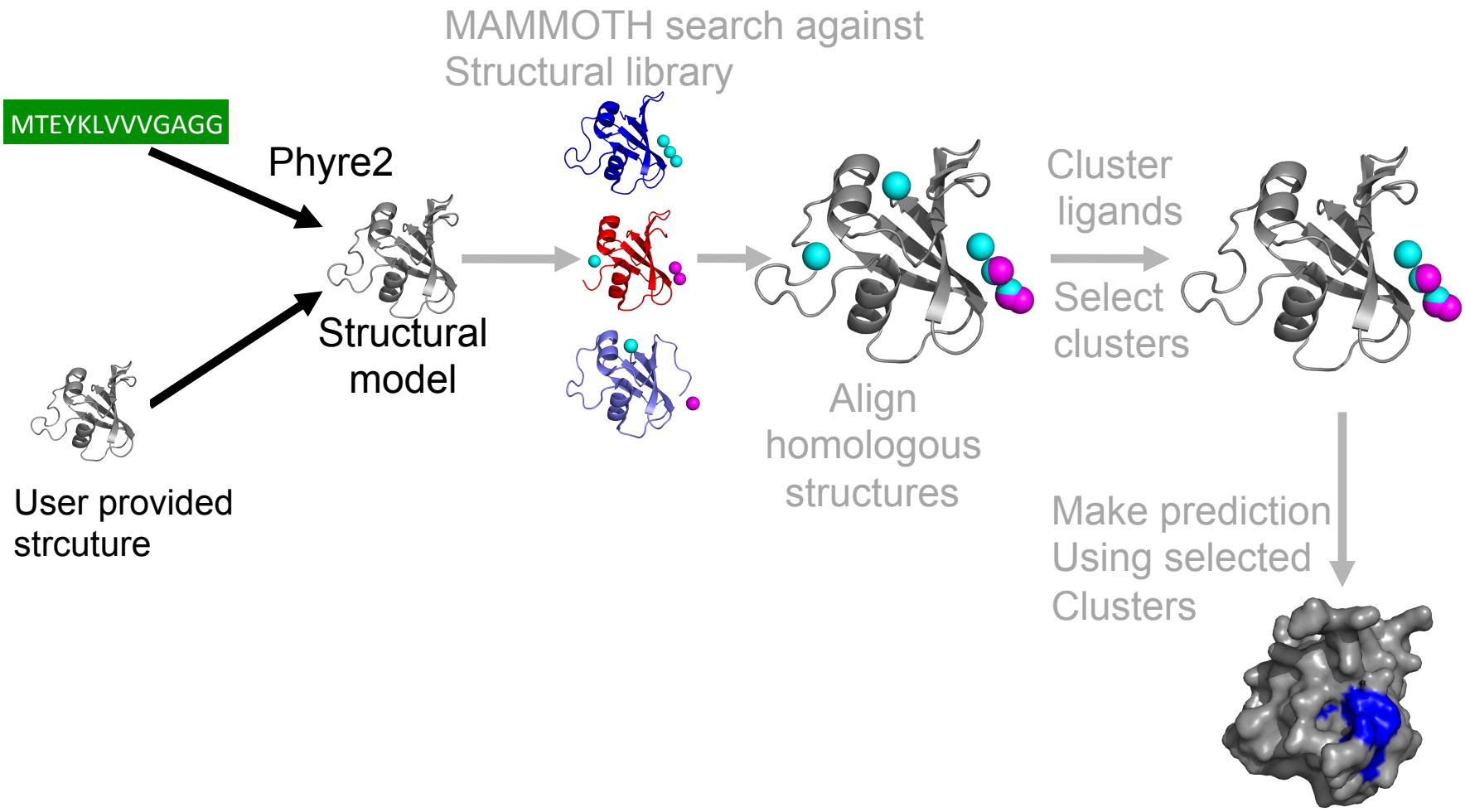
Results page

Submission details

Submission Details

Email:	mark@wass.com
Unique Job identifier:	8ce9f8caffc285eb **JOB ID**
Description:	eg4
Date:	Tue Sep 20 11:36:45 BST 2011
Submission Type:	sequence **Submission type – sequence/structure**
Query Seq:	GLAACEGEYSQKYSTMSPLGSGAFGFVWTAVDKEKNKEVVVKFIKKEKVLEDCWIEDPKL GKVTLEIAILSRVEHANIIKVLDIFENQGFFQLVMEKHSGLDLFAFIDRHPRLDEPLASY IFRQLVSAVGYLRLKDIIHRDIKDENIVIAEDFTIKLIDFGSAAYLERGKLFYTFCGTIE YCAPEVLMGNPYRGPELEMWSLGVTLYTLVFEENPFCELEETVEAAIHPPYLVSKELMSL VSGLLQPVPERRTTLEKLVTDPWVTQPVNLADYTWEEVFRVNKPESGVLSAASLEMGNRS LSDVAQAQELCGGE

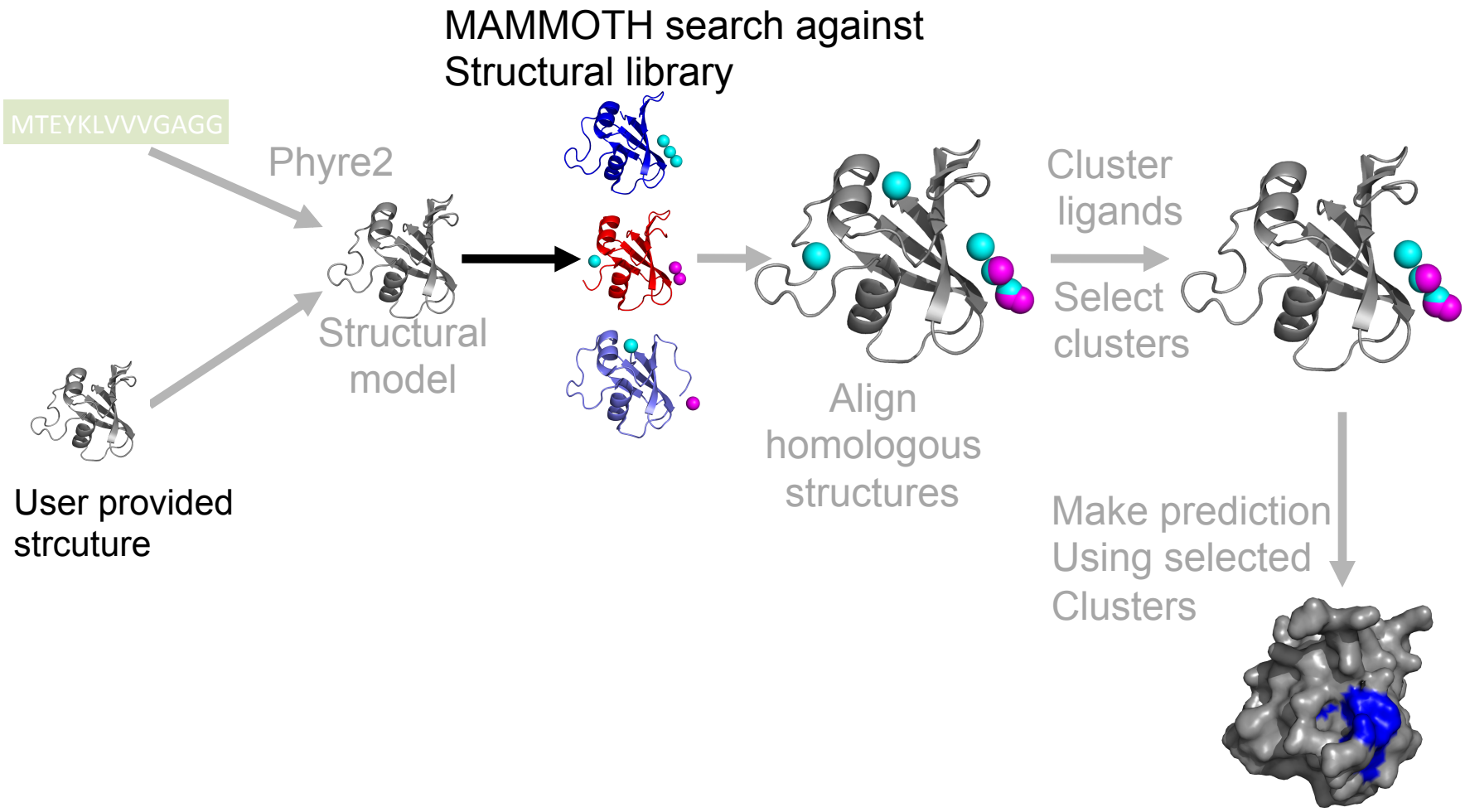
3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

Structural model

Structural Model	
Phyre2 job:	8ce9f8caffc285eb
Phyre2 template:	3dlsA_
Phyre2 confidence score:	100.0

Structural Search confidence data from search of structural library with Mammoth	
Average lnE:	29.965
Maximum LnE:	34.64
Min LnE:	28.45

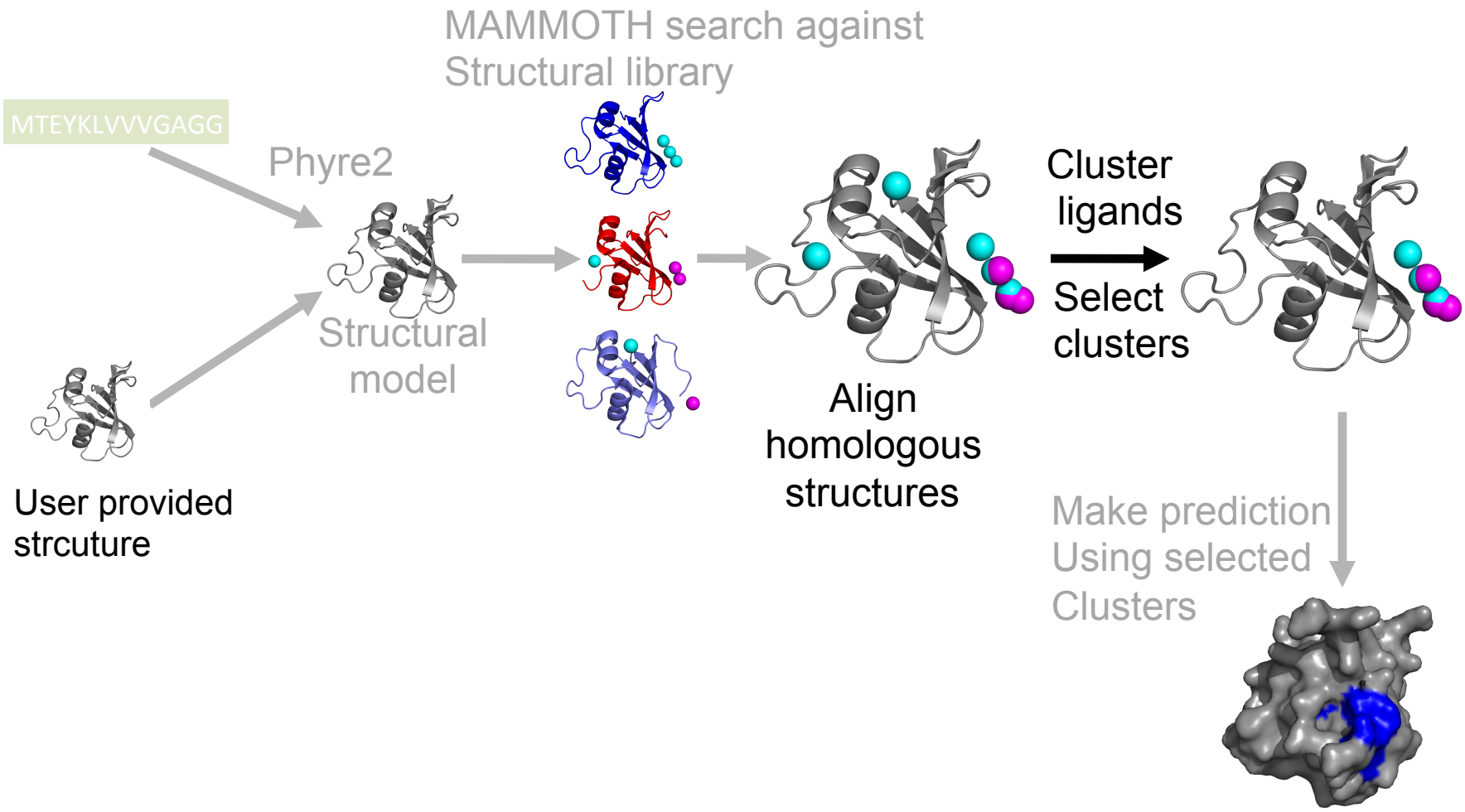
JOB ID –same as 3DLig job id

Model confidence
0 (low) -100 (high)

Similarity of structural hits (higher value = structures more similar)

Min lnE value used = 7
Predictions using low LnE values e.g. < ~12-15 should be treated with caution

3DLigandSite



Wass *et al*., NAR 2010

Imperial College
London

Ligand Clusters

Ligand Clusters Identified

Note prediction based on first cluster

Click on other clusters to view the potential sites associated with them

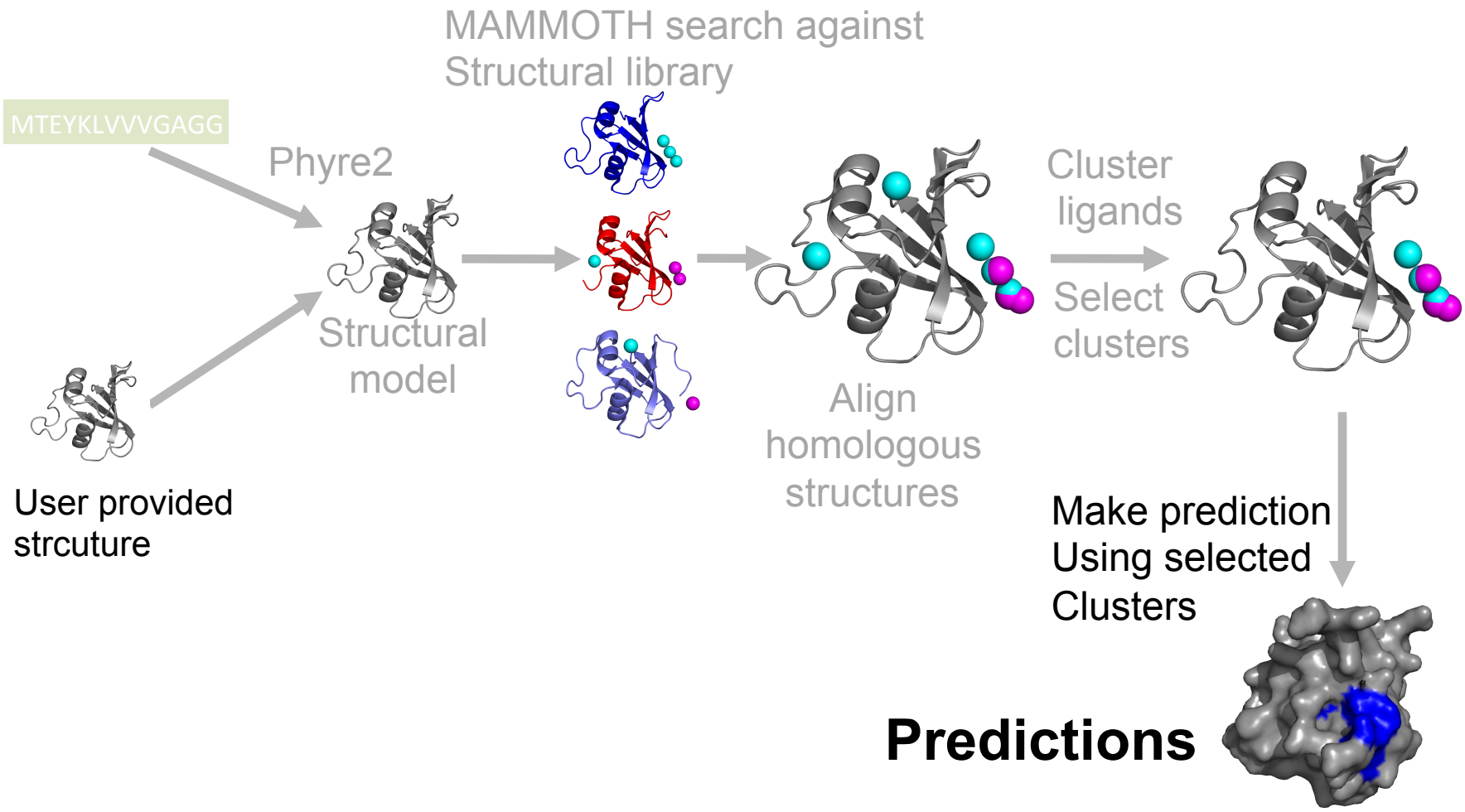
			MAMMOTH Scores		
Cluster	Ligands	Structures	Av	min	max
1	33	22	30.0	28.4	34.6
2	2	2	28.5	28.4	28.6
3	2	2	28.5	28.4	28.6
4	1	1	28.7	28.7	28.7

Clusters ranked by number of ligands.

Mammoth scores for cluster displayed to indicate how similar the structures are that contributed the ligands in the cluster.

Top cluster displayed as main prediction. Click on rows to view predictions for the other clusters.

3DLigandSite



Wass *et al*., NAR 2010

Imperial College London

Interpreting predictions – what ligands?

Heterogens present in Predicted Binding Site

Heterogen	Count	source structures
STU	9	1yhs_A,3ckx_A,1u59_A,1qpd_A,1qpj_A, 2dq7_X,3cd3_A,3bkb_A,3cbl_A
ADP	6	3dls_F,3d5w_A,1ol7_A,1mq4_A,1ol5_A, 2g2i_B
MG	16	2ou7_A,3dls_F,1xr1_A,3f2a_A,1ol7_A, 2v7a_B,1mq4_A,1ol5_A,3cly_A
AMP	1	1yxu_C
ATP	1	1ql6_A

Lists the ligands that are present in the cluster and the structures that they are from

Interpreting predictions

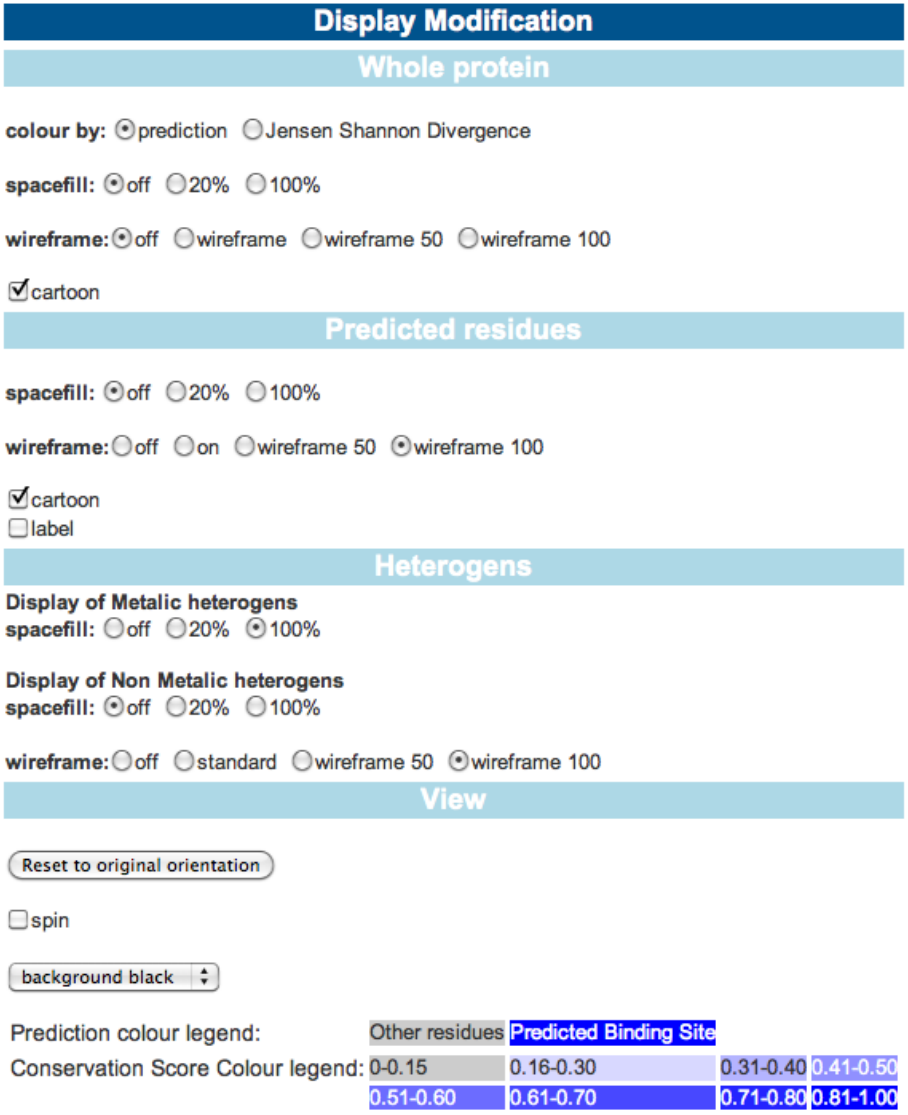
Residue	Amino acid	contact	av distance	JS divergence
19	LEU	17	0.03	0.62
20	GLY	17	0.05	0.76
21	SER	17	0.15	0.34
27	VAL	17	0.29	0.79
40	VAL	17	0.01	0.72
42	LYS	22	0.15	0.84
79	ILE	16	0.33	0.54
95	MET	15	0.19	0.60
96	GLU	17	0.31	0.82
97	LYS	13	0.34	0.65
98	HIS	15	0.34	0.53
100	GLY	10	0.02	0.70
101	LEU	17	0.00	0.00
102	ASP	17	0.10	0.76
103	LEU	23	0.09	0.60
145	GLU	18	0.07	0.83
146	ASN	22	0.09	0.87
148	VAL	9	0.40	0.65
158	ILE	17	0.08	0.49
159	ASP	31	0.02	0.85

Predicted residue table

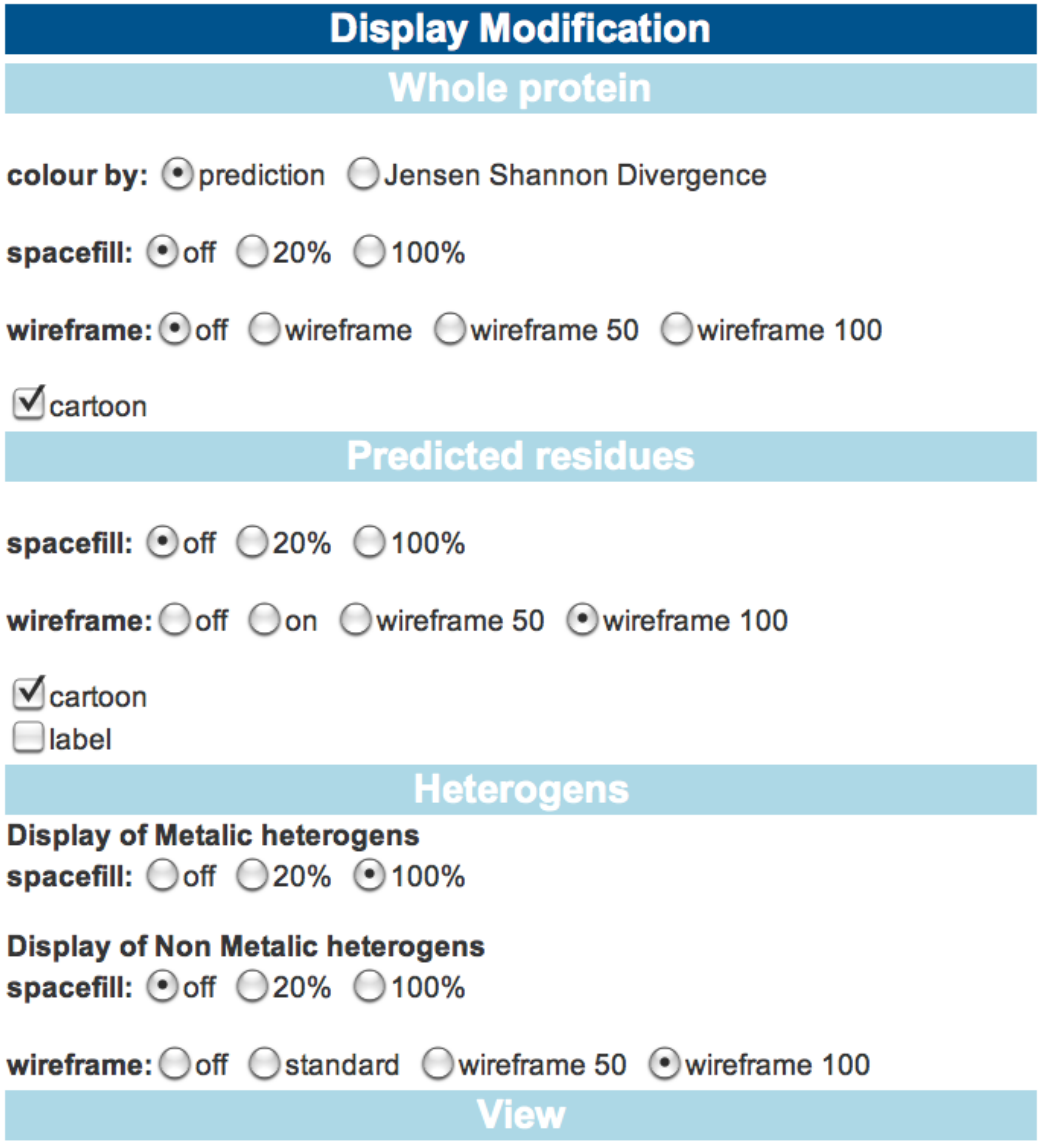
- **Residues in cluster that are < 0.5A +vdw of 25% of cluster predicted**
- **Number of ligand contact**
- **Av distance between residue and these ligands**
- **JS Divergence – conservation score (range 0 – 1).**

- **These values can be used to refine the prediction – e.g.**
  - **residues that contact few of the ligands**
  - **are further from the ligands**
  - **Have low conservation scores**

Interpreting predictions



Interpreting predictions



Control:
Colouring of protein – by prediction or conservation

Display of protein:
Spacefill/wireframe/cartoon

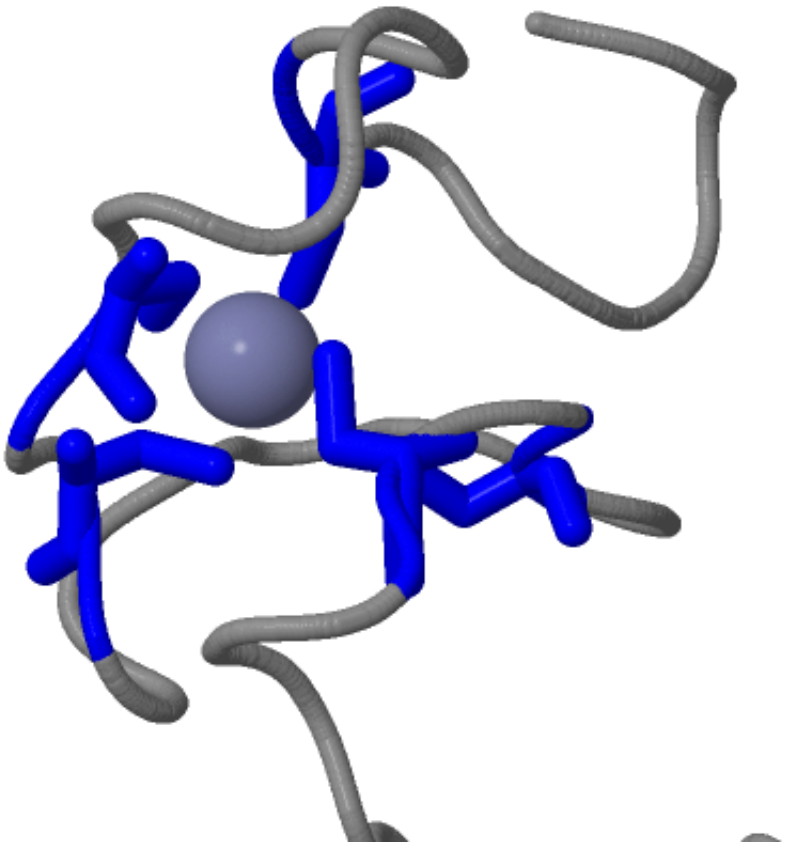
Label predicted residues so they can be identified in the graphical view.

Separate controls for display of predicted residues

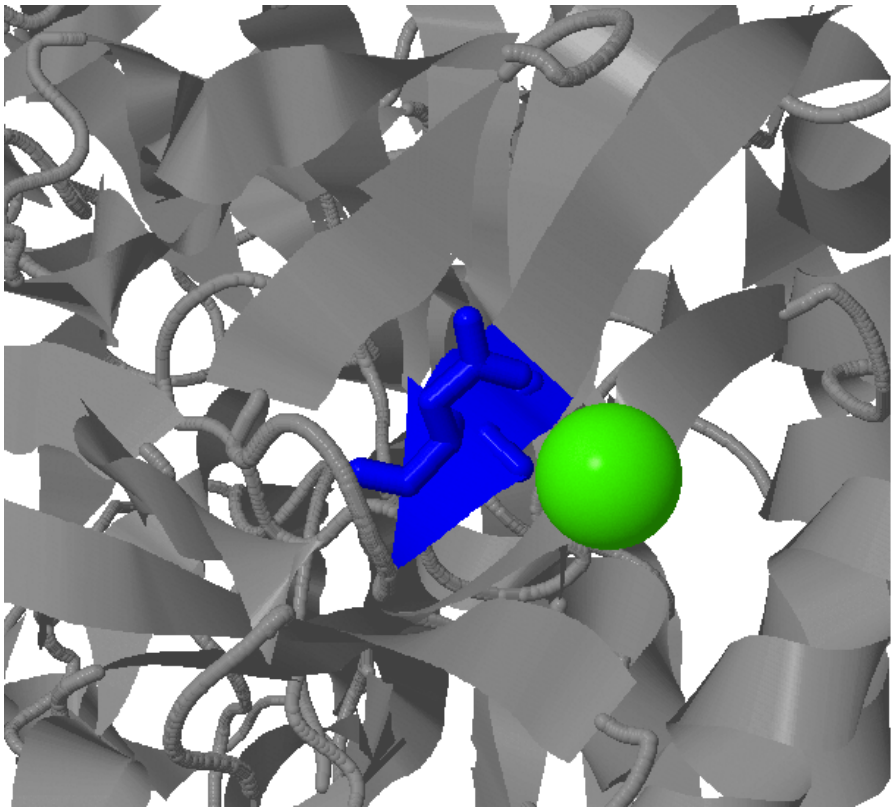
Modify display of ligands:
Spacefill/wireframe

Overall:
Make protein rotate
Change background colour

Interpreting predictions - Metals

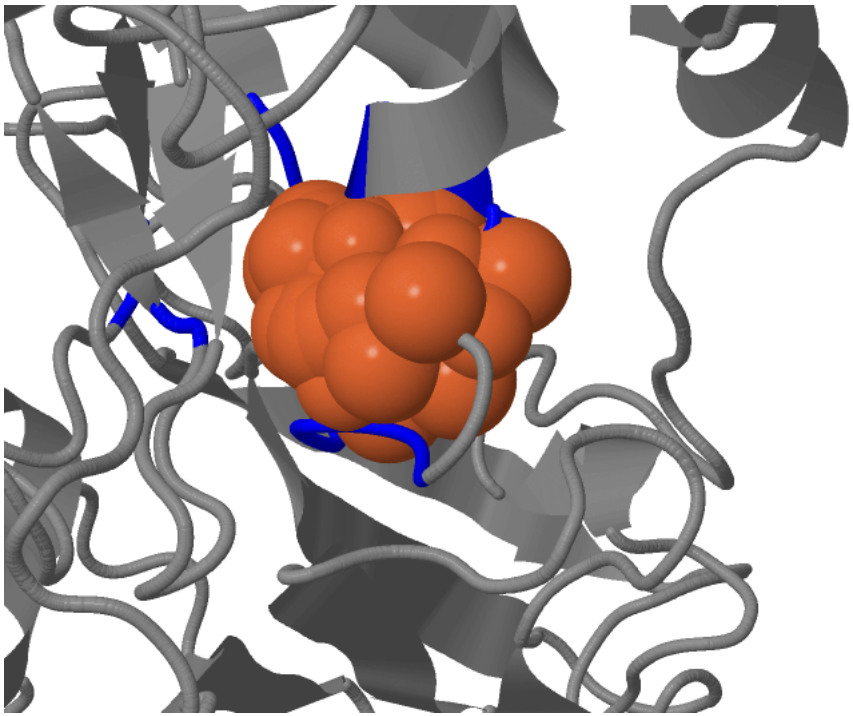


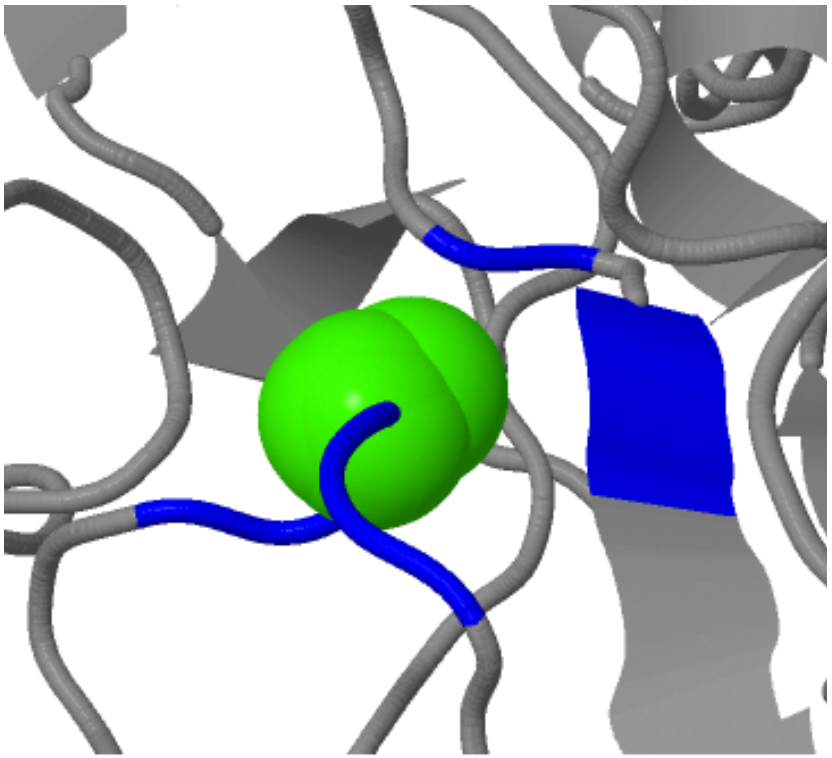
Metals found bound like this – with 3-6 residues
Often the residues aren't sequential



Binding sites with a single residue contacting the ligand are likely to be wrong

Interpreting predictions - Metals

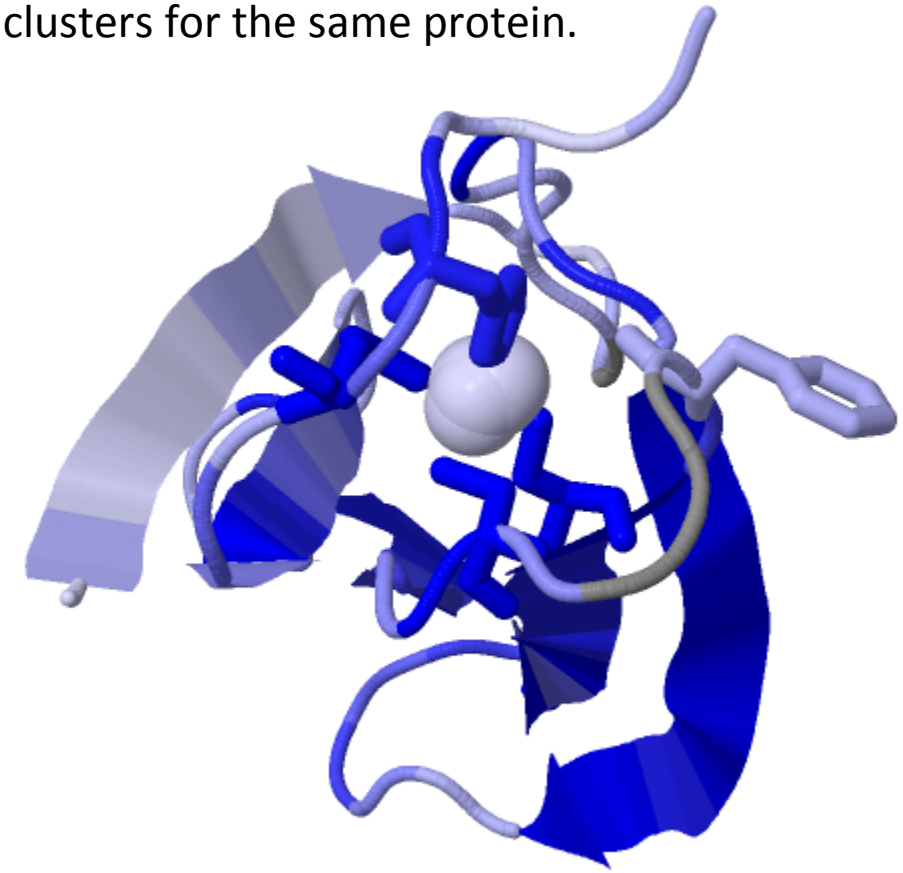




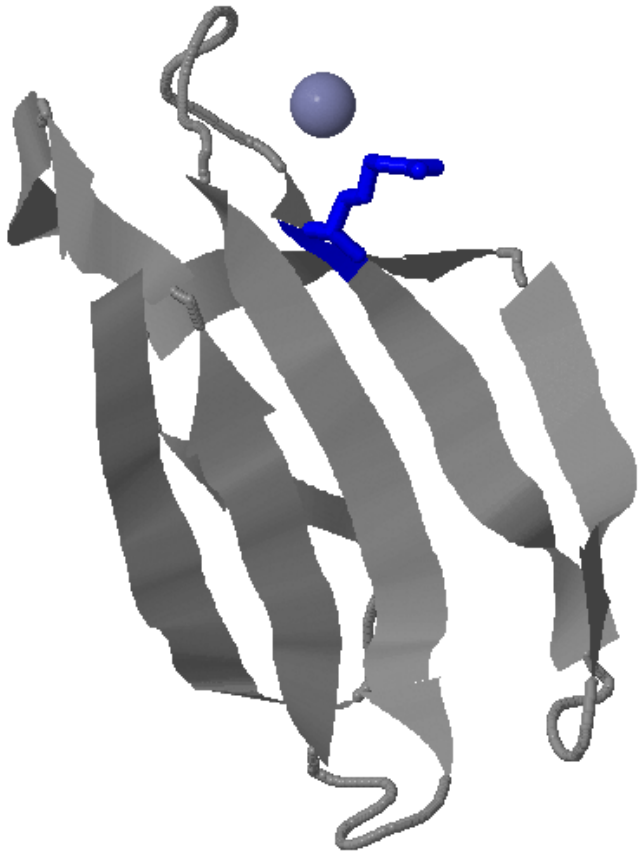
Sometimes the cluster of residues might overlap with the protein structure as in the examples above. This is more likely where the cluster is close to a loop. The prediction may be good but it might also be slightly affected by the overlap of the cluster and the structure

Interpreting predictions - Metals

2 clusters for the same protein.

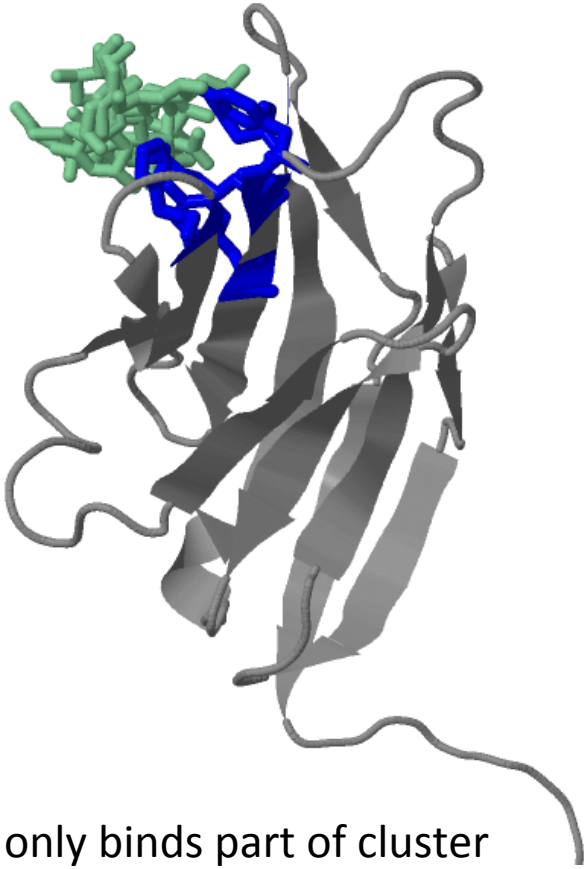


Multiple ligands in cluster
Multiple residues contacting ligand
Looks like it could be a ligand binding site
Divergence colouring help suggest residue that might not be part of the binding site.

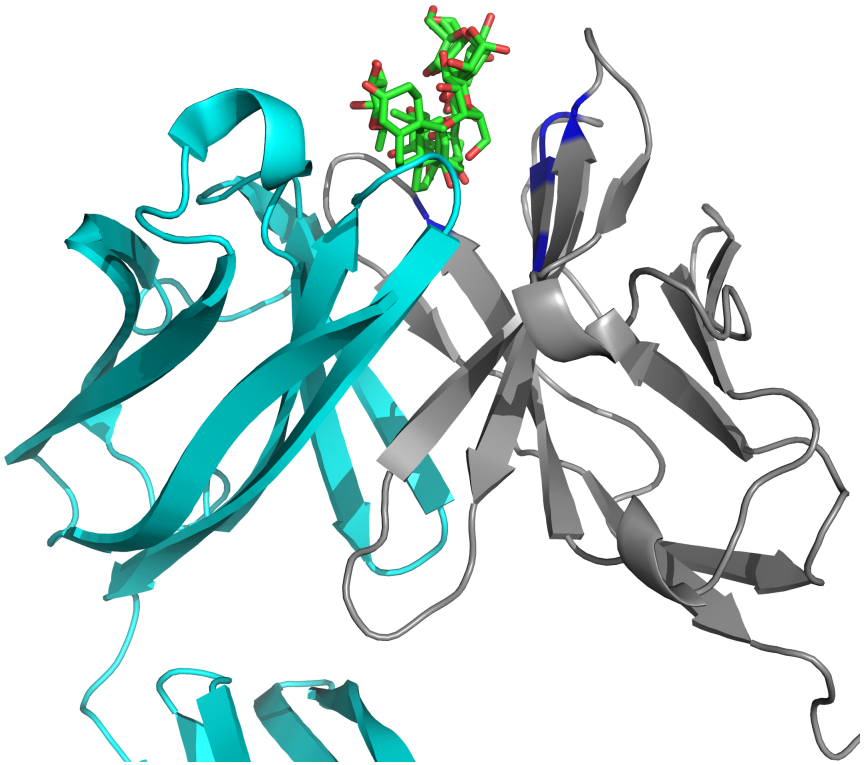


Single ligand in cluster
Single residue binds the ligand
Unlikely to be a ligand binding site

Interpreting predictions - Oligomers



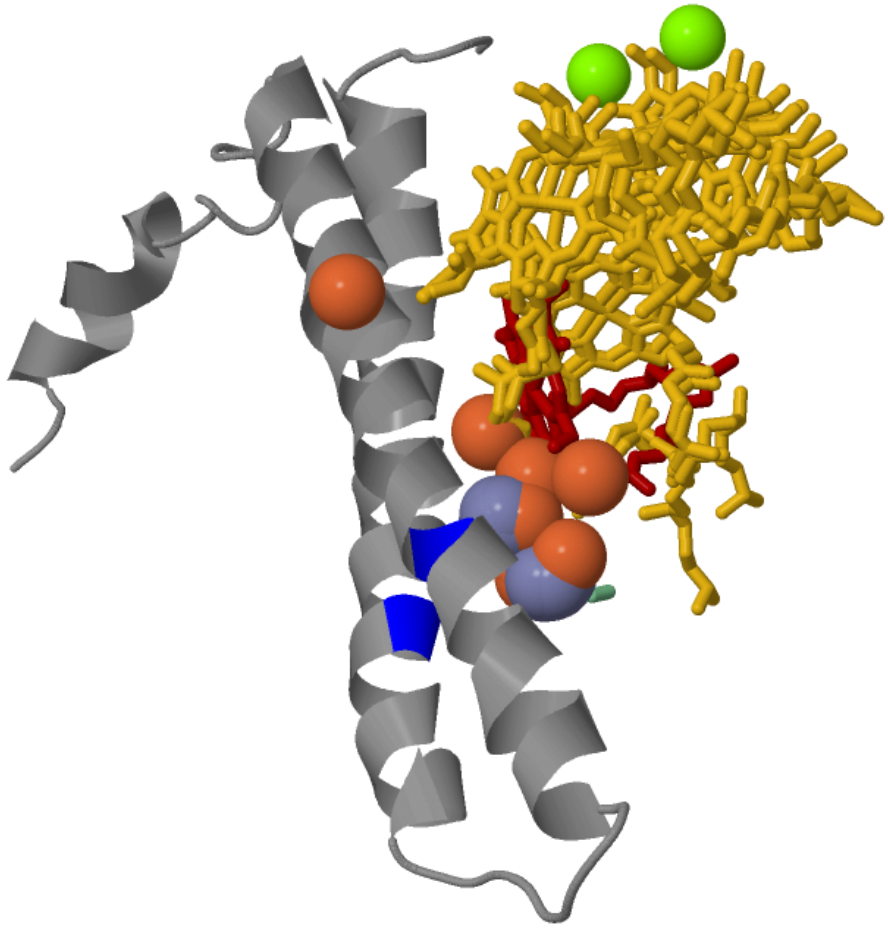
Site only binds part of cluster



Prediction viewed with other chain of dimer from one of the templates

When predictions only seem to contact part of the ligand in some example this is because the ligand is bound between chains in an oligomer. Therefore part of the binding site might be missed. Different clusters predicted for the binding site may predict different residues that when combined contain the full binding site

Interpreting predictions – large Clusters



Large cluster of many different ligands.

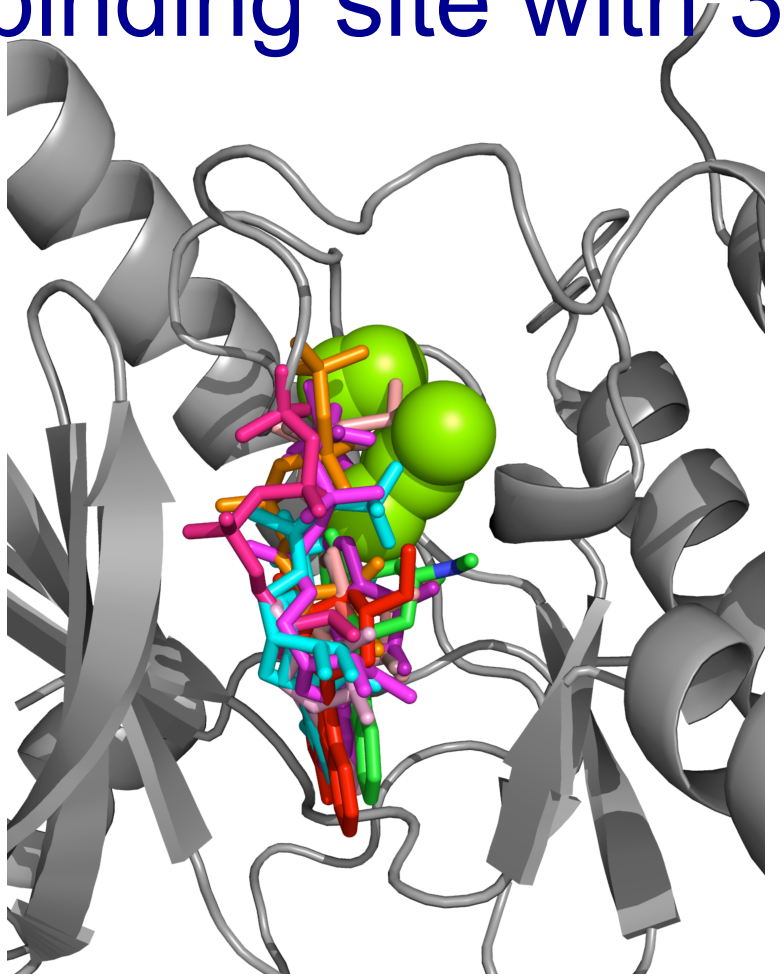
This is unlikely to be a binding site

Interpreting predictions

Suggestions for interpreting results:

- Consider the similarity between the structure and the hits
- The number of ligands in a cluster may be indicative of how likely it is for the region to be a binding site
- Use of the JS Divergence score may help refine predictions
- Metal binding site predictions can have high levels of false positive.
    - Especially if there are many clusters and the clusters only contain a single metal ion
    - Metal ions a generally contact multiple residues
    - Checking the conservation score may be helpful here to remove false predictions
- Clusters can occasionally become very large – with many ligands covering a large are of the protein. Such a large site is likely to be incorrect, although part of it may be ligand binding.

Modelling binding site with 3DLigandSite



Mark Wass
m.n.wass@kent.ac.uk