Using 3DLigandSite and CombFunc

Mark Wass

m.n.wass@kent.ac.uk





Predicting protein function using Gene Ontology

3DLigandSite

3DLigandSite



Homepage - submission

Paste your sequence here:





Results page

Submission details

Submission Details						
Email:	mark@wass	s.com				
Unique Job identifier:	8ce9f8caffc	285eb JOB	D			
Description:	eg4					
Date:	Tue Sep 20	11:36:45 BST 20	11			
Submission Type:	sequence	Submissi	on type	– seque	ence/struc	tu
Query Seq:	GLAACEGEY GKVTLEIAI IFRQLVSAV YCAPEVLMG VSGLLQPVP LSDVAQAQE	SQKYSTMSPLGSGA LSRVEHANIIKVLD GYLRLKDIIHRDIK NPYRGPELEMWSLG ERRTTLEKLVTDPW LCGGE	FGFVWTAVDKE IFENQGFFQLV DENIVIAEDFT /TLYTLVFEEN /TQPVNLADYT	KNKEVVVKFI MEKHSGLDLF IKLIDFGSAA PFCELEETVE WEEVFRVNKP	KKEKVLEDCWIEDE AFIDRHPRLDEPLA YLERGKLFYTFCGT AAIHPPYLVSKELM ESGVLSAASLEMGN	KL SY IE ISL IRS

3DLigandSite



3DLigandSite



London

Structural model

Structural Mo	odel	JOB ID –same a
Phyre2 job:	8ce9f8c	affc285eb 3DLig job id
Phyre2 template:	3dlsA_	
Phyre2 confidence score:	100.0	Model confidence 0 (low) -100 (high)
confidence data from search of Mammoth	ral library with	
Average InE:	29.965	
Maximum LnE:	34.64	Similarity of structural bits
Min LnE:	28.45	(higher value =
Min InE value used = 7 Predictions using low LnE values e.g. < ~12-15 should I treated with caution	be	structures more similar)

3DLigandSite



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



Imperial College London

Interpreting predictions – what ligands?

Heterog	jens p	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



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

Structural View of Prediction



Download model and pymol Script

Display Modification

Whole protein

colour by:
 prediction
 Jensen Shannon Divergence

spacefill: Off O20% O100%

wireframe: Off Owireframe Owireframe 50 Owireframe 100

✓ cartoon

Predicted residues

spacefill: ⊙off ○20% ○100%

wireframe: Off Oon Owireframe 50 Owireframe 100

✓ cartoon

label

Heterogens

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

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

wireframe: Off Ostandard Owireframe 50 Owireframe 100

ew

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

Display Modification

Whole protein

colour by: • prediction • Jensen Shannon Divergence

wireframe: • off • wireframe • wireframe 50 • wireframe 100

Cartoon

Predicted residues

spacefill: ● off ○ 20% ○ 100%

wireframe: Off Oon Owireframe 50 Owireframe 100

✓ cartoon

label

Heterogens

View

Display of Metalic heterogens spacefill: Off O20% •100%

```
Display of Non Metalic heterogens spacefill: • off 20% 100%
```

wireframe: Off Ostandard Owireframe 50 Owireframe 100

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



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

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.

Interactive Example

Structural View of Prediction



Download model and pymol Script

Display Modification

Whole protein

colour by:
 prediction
 Jensen Shannon Divergence

spacefill: Off O20% O100%

wireframe: Off Owireframe Owireframe 50 Owireframe 100

✓ cartoon

Predicted residues

spacefill: ⊙off ○20% ○100%

wireframe: Off Oon Owireframe 50 Owireframe 100

✓ cartoon

label

Heterogens

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

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

wireframe: O off O standard O wireframe 50 O wireframe 100

/iew

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



ConFunc/CombFunc



CombFunc - submission

Paste your sequence here:	(example input formats)	
	Paste sequence	//
Uniprot id:	Uniprot Id is optional	optional
email address:		optional
Job description:		optional
	Reset CombFunc search	

a few seconds to respond after hitting submit, please do not press more than once



All predictions are:

- Gene Ontology Terms (e.g. GO:0003924, GTPase activity)
- In either Molecular Function or Biological Process ontology
- Associated with a confidence score (0-1 1 highest confidence)
- Presented in multiple formats for users to investigate

The overall results are displayed first

Followed by results from each individual method

CombFunc Overall results

Molecular Function Predictions



Equivalent for Biological Process Terms

Biological Process Predictions











Predictions also displayed in a list view



- GO:0008601 protein phosphatase type 2A regulator activityGO
- GO:0004722 protein serine/threonine phosphatase activityGO
- GO:0005488 bindingGO

Click to show parent term

Molecular Function Predictions

Expand All Collapse All

GO:0008601 - protein phosphatase type 2A regulator activityGO
 GO:0019888 - protein phosphatase regulator activityGO
 GO:0019208 - phosphatase regulator activityGO
 GO:0030234 - enzyme regulator activityGO
 GO:0003674 - molecular functionGO

GO:0004722 - protein serine/threonine phosphatase activity GO

□ GO:0004721 - phosphoprotein phosphatase activityGO

GO:0016791 - phosphatase activityGO

GO:0042578 - phosphoric ester hydrolase activityGO

GO:0016788 - hydrolase activity, acting on ester bondsGO

GO:0016787 - hydrolase activityGO

GO:0003824 - catalytic activityGO

GO:0003674 - molecular_functionGO

CombFunc Results of Individual methods

ConFunc

The terms predicted by ConFunc are disaplyed below. subalignment that was generated for that term.

ConFunc Molecular Function Results Show/hide

ConFunc Biological Process Results Show/hide

Go Term	Function	Z Score	Z Score Ratio	Number Sequences
GO:0042578GO	phosphoric ester hydrolase activity	25.163	1	5
<u>GO:0016791</u> GO	phosphatase activity	25.163	1	5
GO:0016788GO	hydrolase activity, acting on ester bonds	25.163	1	5
<u>GO:0016787</u> GO	hydrolase activity	25.163	1	5
GO:0004722GO	protein serine/threonine phosphatase activity	y 25.114	0.99804	4
<u>GO:0004721</u> GO	phosphoprotein phosphatase activity	25.114	0.99804	4
GO:0019888GO	protein phosphatase regulator activity	23.702	0.94193	14
GO:0019208GO	phosphatase regulator activity	23.702	0.94193	14
<u>GO:0008601</u> GO	protein phosphatase type 2A regulator activi	ty 23.702	0.94193	14
<u>GO:0030234</u> GO	enzyme regulator activity	21.059	0.8369	17
<u>GO:0003824</u> GO	catalytic activity	20.541	0.81631	7
<u>GO:0003674</u> GO	molecular_function	15.72	0.62472	44
		1	1	
	Higher the better m	Zscore ax(Zsco	/ #sec ore) aligni	ιs in ment

BLAST

The top 3 sequence hits with either to 6 sequences in total are shown - annotations

BLAST data Show/hide

Shows top 3 GO annotated BLAST hits

BLAST data Show/hide



Interpro

The Interpro hits to the query sequence are shown below Ontology functions that are commonly associated with the will also have some of these functions. Links to further in clicking the link on the interpo accession. The domain him where on the sequence the domains are identified

Interpro Hits - Tabular Data Show/hide

Interpro Hits - Graphical View Show/hide

Interpro Hits - Tabular Data Show/hide

Database	Accession	start	finish	e-value	Interpro	Desc
FPrintScan	PR00600	267	294	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	105	133	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	295	322	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
HMMPIR	PIRSF037309	4	452	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	182	209	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
HMMPanther	PTHR11871	1	453	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	323	348	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	76	104	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
PatternScan	PS01025	180	194	0.00e+00	IPR018067	Protein phosphatase 2A, regulatory subunit PR55, conserved site
FPrintScan	PR00600	419	448	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	238	266	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
PatternScan	PS01024	89	103	0.00e+00	IPR018067	Protein phosphatase 2A, regulatory subunit PR55, conserved site
FPrintScan	PR00600	41	61	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	349	375	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
FPrintScan	PR00600	210	237	0.00e+00	IPR000009	Protein phosphatase 2A, regulatory subunit PR55
superfamily	SSF50978	34	444	3.40e-34	IPR011046	WD40 repeat-like-containing domain
Gene3D	G3DSA:2.130.10.10	417	443	6.90e-28	IPR015943	WD40/YVTN repeat-like-containing domain
Gene3D	G3DSA:2.130.10.10 G3DSA:2.130.10.10	30 170	383	0.90e-28 6.90e-28	IPR015943	WD40/YVTN repeat-like-containing domain
HMMPfam	PF00400	100	129	5.00e-01	IPR019781	WD40 repeat, subgroup
HMMSmart	SM00320	171	210	1.20e+00	IPR001680	WD40 repeat
HMMSmart	SM00320	89	129	4.70e+00	IPR001680	WD40 repeat

GO Terms	Go function
GO:0008601	protein phosphatase type 24 regulator activity
GO:0007165	signal transduction
GO:0000159	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
GO:0000159	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
GO:0000159	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
GO:0000159	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
<u>GO:0000159</u>	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
<u>GO:0000159</u>	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
<u>GO:0000159</u>	protein phosphatase type 2A complex
GO:0008601 GO:0007165	protein phosphatase type ZA regulator activity
GO:000159	protein phosphatase type 24 complex
00.000100	protein phosphatase type 24 complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
GO:0000159	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
GO:0000159	protein phosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
<u>GO:0000159</u>	protein phosphatase type 2A complex
<u>GO:0008601</u>	protein phosphatase type 2A regulator activity
GO:0007165	signal transduction
<u>GO:0000159</u>	protein prosphatase type 2A complex
GO:0008601	protein phosphatase type 2A regulator activity
GO:000159	signal transduction
GO:0000109	protein priosphalase type 2A complex
60.0005515	protein binding

GO:0005515	protein binding	
<u>GO:0005515</u>	protein binding	

protein binding

protein binding

GO:0005515

GO:0005515

Interpro Hits - Tabular Data Show/hide





Search of structural library

The Phyre2 fold library has been searched using HHsear that are homologous to the query sequence. The table sh of the structures are displayed in the last column.

Structural Search Results Show/hide



Interactome Analysis

MINT and IntAct have been searched for int of the interactions partners have been ident

Molecular Function Data Show/hide

Biological Process Data Show/hide

GO Term	Description		Count direct	% direct	Count indirect		% indirect
GO:0005488	binding		22	95.652	750		91.575
GO:0005515	protein binding		22	95.652	717		87.546
GO:0003824	catalytic activity		8	34.783	283		34.554
GO:0016787	hydrolase activity		6	26.087	123		15.018
GO:0051082	unfolded protein binding		6	26.087	37		4.518
GO:0016788	hydrolase activity, acting on es	ter bor	5	21.739	62		7.570
GO:0016740	transferase activity		3	13.043	148		18.071
GO:0030234	enzyme regulator activity		3	13.043	148		18.071
GO:0016772	transferase activity, transferring	g phos	3	13.043	136		16.606
GO:0016301	kinase activity		3	13.043	134		16.361
	#direct with this GO annot	% of direc intera	all atly acting	#indir with t annot	rect his GO	% d ind inte	of all lirectly eracting
		prote	ins			pro	teins



GO Term	Description		Count direct	% direct	Count indirect		% indirect
GO:0005488	binding		22	95.652	750		91.575
GO:0005515	protein binding		22	95.652	717		87.546
GO:0003824	catalytic activity		8	34.783	283		34.554
GO:0016787	hydrolase activity		6	26.087	123		15.018
GO:0051082	unfolded protein binding		6	26.087	37		4.518
GO:0016788	hydrolase activity, acting on es	ster bor	5	21.739	62		7.570
GO:0016740	transferase activity		3	13.043	148		18.071
GO:0030234	enzyme regulator activity		3	13.043	148		18.071
GO:0016772	transferase activity, transferring	g phos	3	13.043	136		16.606
GO:0016301	kinase activity		3	13.043	134		16.361
	#direct with this GO annot	% of a direc intera	all tly acting	#indir with t annot	rect his GO	% c ind inte	of all irectly eracting
		proteins				pro	teins

Gene Expression

- CoxpressDB has been searched for co-expression data for the query pr coexpressed proteins have been identified and are summarised in the ta
- Molecular Function Coexpression Data Show/hide
- Gene Co-expression data for GO Biological Process Show/hide
- Gene Co-expression data Show/hide

Go Term	#CoExp Av MR min MR max MR	GO Term Description
---------	----------------------------	---------------------

GO:0009987	6	26.9	11.4	41.0
GO:0044237	5	26.8	11.4	41.0
GO:0008152	5	26.8	11.4	41.0
GO:0044238	5	31.7	11.4	41.0
<u>GO:0043170</u>	4	29.4	11.4	41.0
<u>GO:0044260</u>	4	29.4	11.4	41.0
GO:0043412	4	32.6	11.4	41.0
GO:0019538	3	29.8	11.4	41.0
GO:0006464	3	29.8	11.4	41.0

cellular process
cellular metabolic process
metabolic process
primary metabolic process
macromolecule metabolic process
cellular macromolecule metabolic process
macromolecule modification
protein metabolic process
protein modification process

#genes coexpressed with function

Mutual rank

Av – average over co-exp genes

- Min smallest over co-expressed genes
- Max largest over co-expressed genes

ConFunc/CombFunc







Predicting protein function using Gene Ontology