

Graph-Theoretical Approaches in Drug Discovery at the PDB Scale

Dušanka Janežič and Janez Konc

University of Primorska

Faculty of Mathematics, Natural Sciences and Information Technologies

Koper, Slovenia

Graphs, groups, and more: celebrating Brian Alspach 80th and Dragan Marušič
65th birthdays, UP FAMNIT Koper, May 28 – June 1, 2018

Outline

- Introduction

We have developed new computational tools at the molecular scale for protein-ligand binding in drug discovery, based on graph theoretical approaches, combined with molecular simulations.

- Detecting protein binding sites

ProBiS Tools (algorithm, database, and web servers) for predicting and modeling of medically interesting proteins

- Precision medicine bioinformatics

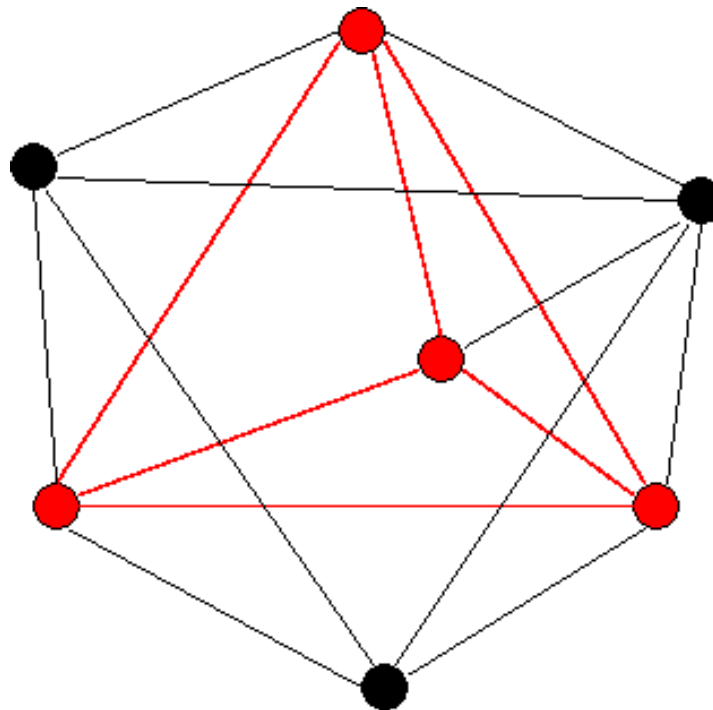
Gene level, Protein level, System level

- Applications: use in drug discovery

Maximum Clique Problem

An algorithm for finding a maximum clique in an undirected graph was developed. Up to 100 times faster than the best comparable algorithm.

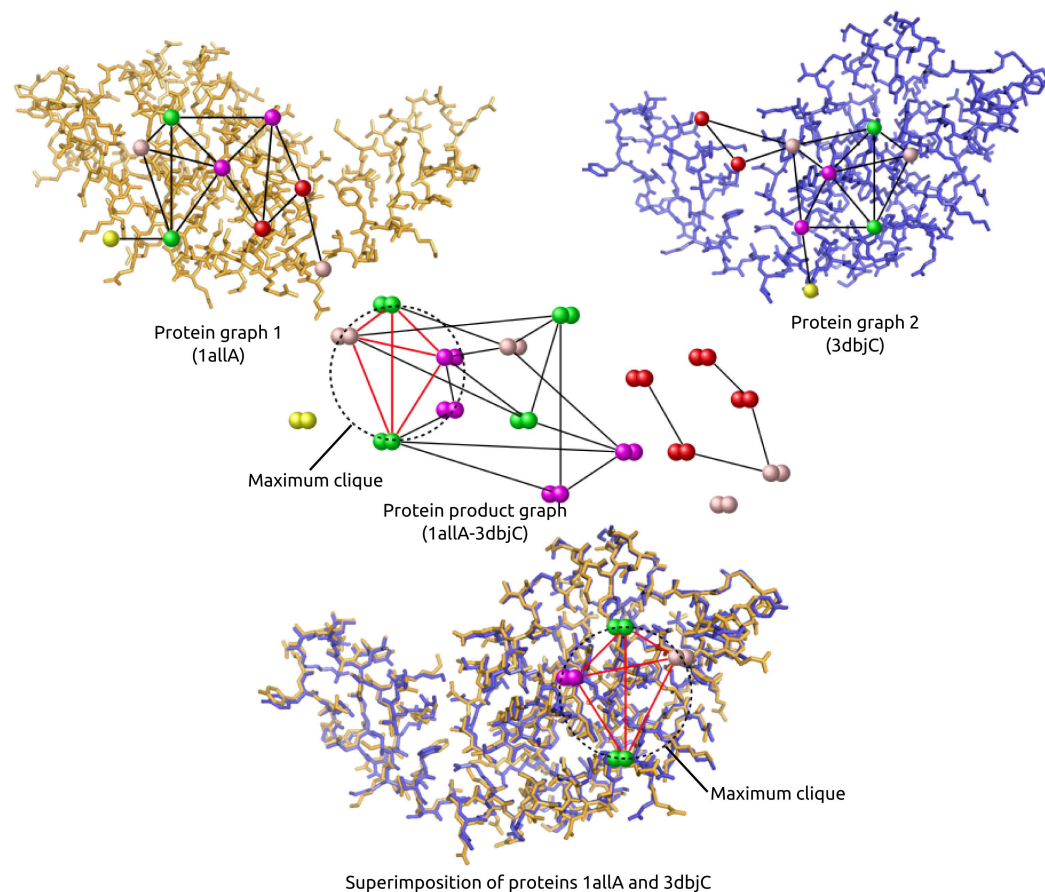
An undirected graph $G=(V,E)$; V is a set of vertices and E is a set of edges.



Maximum clique (red) is the largest fully connected subgraph within a graph.

Janez Konc, Dušanka Janežič, *An Improved Branch and Bound Algorithm for the Maximum Clique Problem*, *MATCH*, 58, 569-590, 2007.

ProBiS Algorithm



Proteins are first converted to protein graphs (top). Vertices are colored according to their physicochemical properties, i.e., acceptor (red), donor (green), pi-pi stacking (pink), and aliphatic (yellow). A protein product graph is constructed from both protein graphs (center). A maximum clique of four vertices connected with red edges indicates the similarity between proteins. The two compared proteins are superimposed (bottom) according to the best alignment of vertices represented by the maximum clique.

J. Konc and D. Janežič, *Bioinformatics*, **26**, 1160 **2010**, & *J. Chem. Inf. Model.*, **53**, 2217 **2013**.

ProBiS Web Server – Input

ProBiS

Search by PDB ID

As of Jul 31, 2015 your protein is compared with 42270 structures

Introduction

- If you are new to ProBiS, you can watch the introductory video available here.

ProBiS-Tools

- ProBiS Input Page
- RESTful Web Services
- Database Access

Download

- ProBiS Algorithm
- Non-redundant PDB (nr-PDB)

Help

- ProBiS-ligands User's Guide
- ProBiS User's Guide
- Examples

Citation

- Konc, J., Miller, B. T., Stulár, T., Lesník, S., Woodcock, H. L., Brooks, B. R., and Janežić, D. ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *J. Chem. Inf. Model.*, 2015, **55**, 2308-2314.

Related Citations

- Konc, J. and Janežić, D. ProBiS-ligands: a web server for prediction of ligands by examination of protein binding sites. *Nucleic Acids Res.*, 2014, **42**, W215-W220.
- Konc, J. and Janežić, D. ProBiS-2012: web server and web services for detection of structurally similar binding sites in proteins. *Nucleic Acids Res.*, 2012, **40**, W214-W221.
- Konc, J., Česník, T., Trykowski, J., Pencina, M., Janežić, D. ProBiS-Database: Precalculated Binding Site Similarities and Local Pairwise Alignments of PDB Structures. *J. Chem. Inf. Model.*, 2012, **52**, 604-612.
- Konc, J. and Janežić, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment.

Enter Query Protein

Pre-calculated local structural similarity profile is ready for a protein with >95% seq. id. [Click here to show ProBiS-Database results for PDB entry 1azeA](#)

PDB ID: Chain ID(s):

Select Input

- ☒ Binding Sites
- ☐ Chain B
- ☐ Custom Selection

Distance (Å)

e.g., A and (12-15,18) or B and (64,6)

Pairwise Align Two Proteins or Binding Sites (optional)

Enter Subject Protein

PDB ID: Chain ID(s):

Your e-mail address (optional):

A link to the results page will be sent to you by e-mail. Alignment of two proteins will take seconds.

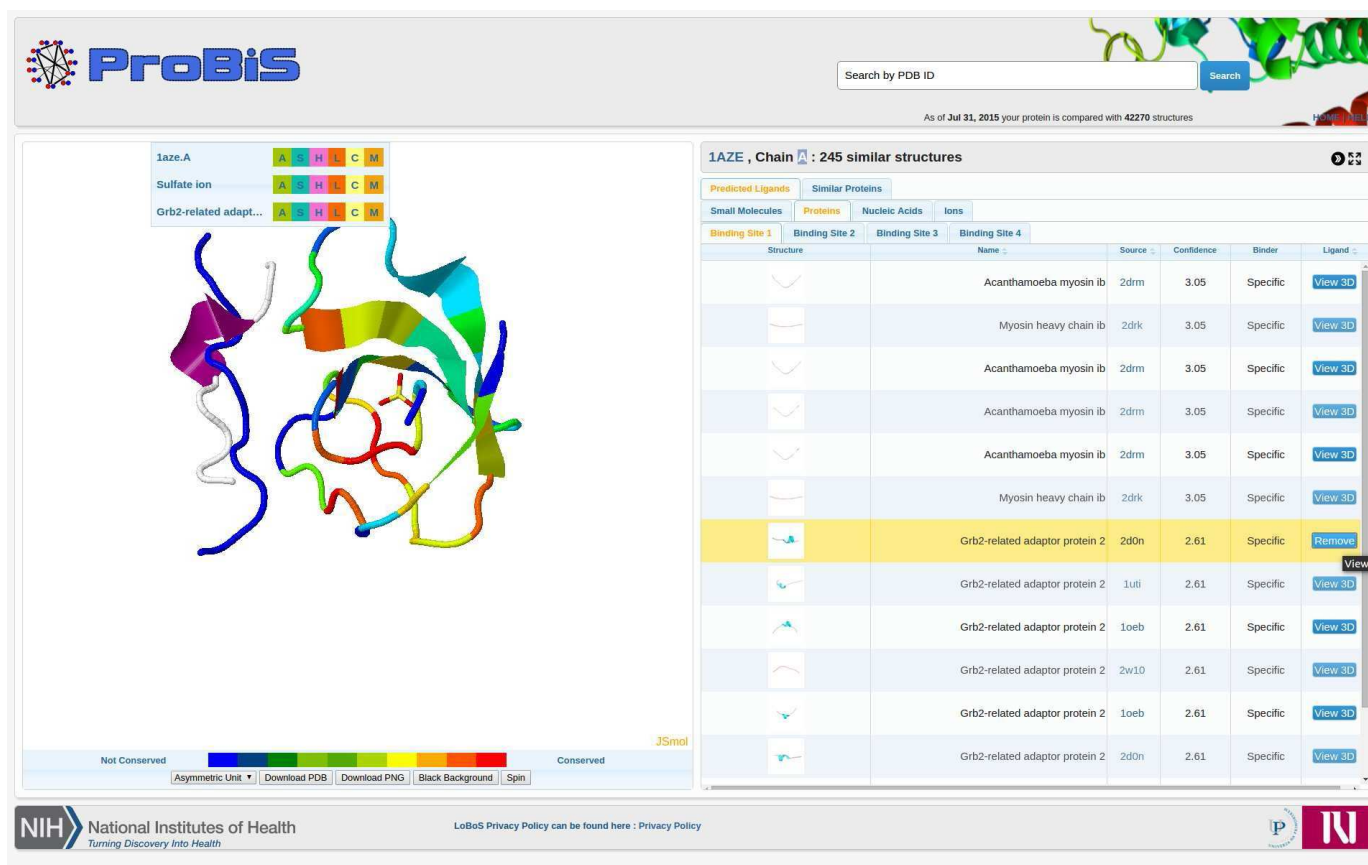
Video Tutorial

Contact

- Scientific: konc[at]cmm[dot]ki[dot]si

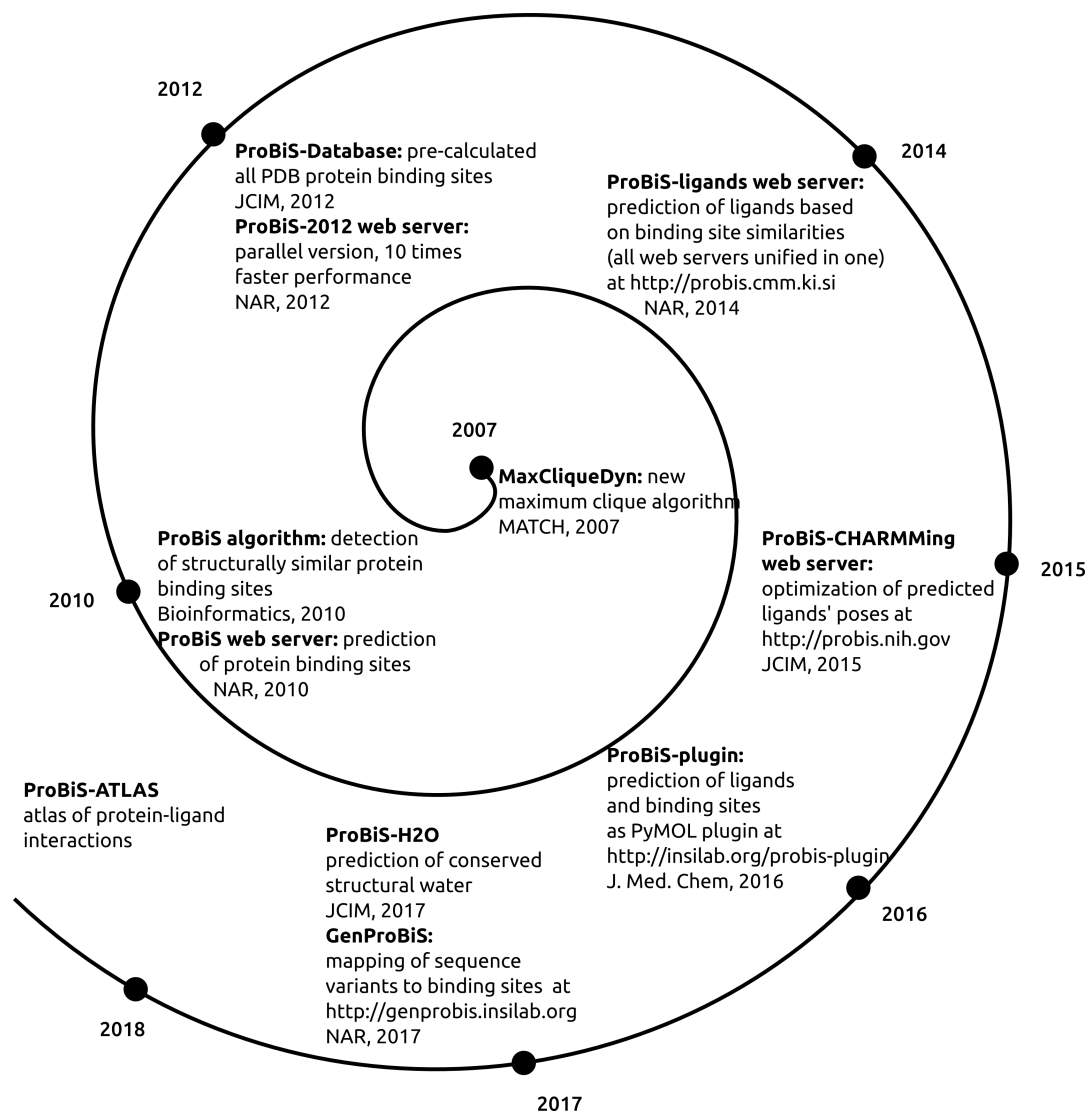
Detection of structurally similar binding sites in PDB and local pairwise alignment of protein structures.

ProBiS Web Server – Output

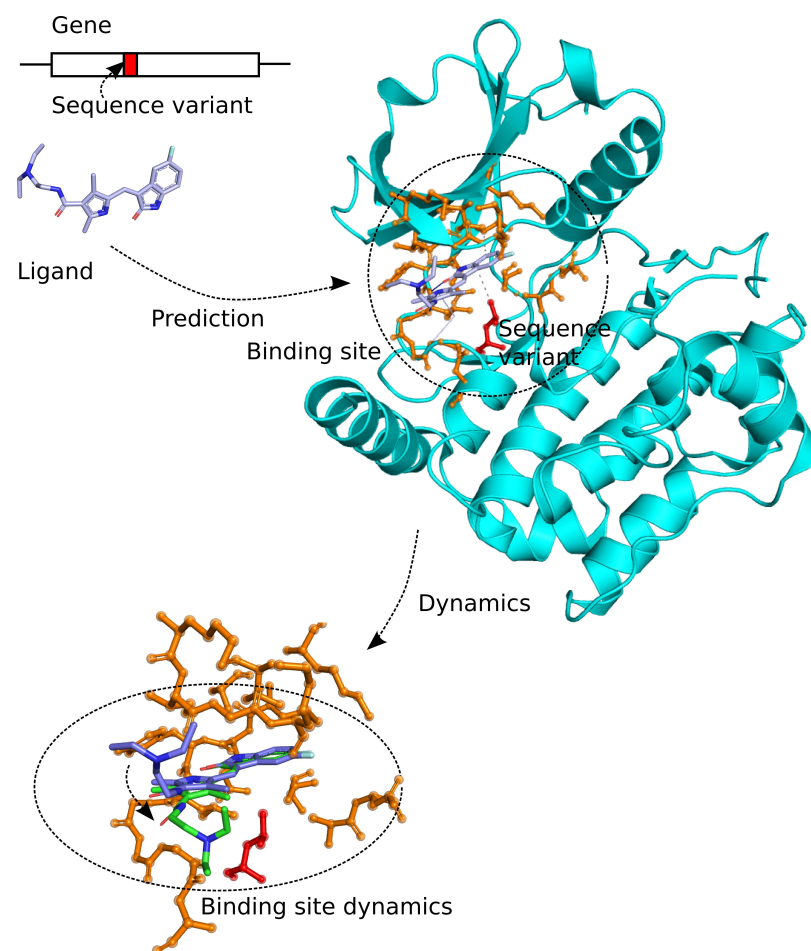


Predicted protein binding site and ligand as visualized in the ProBiS web server.

Evolution of the ProBiS Tools

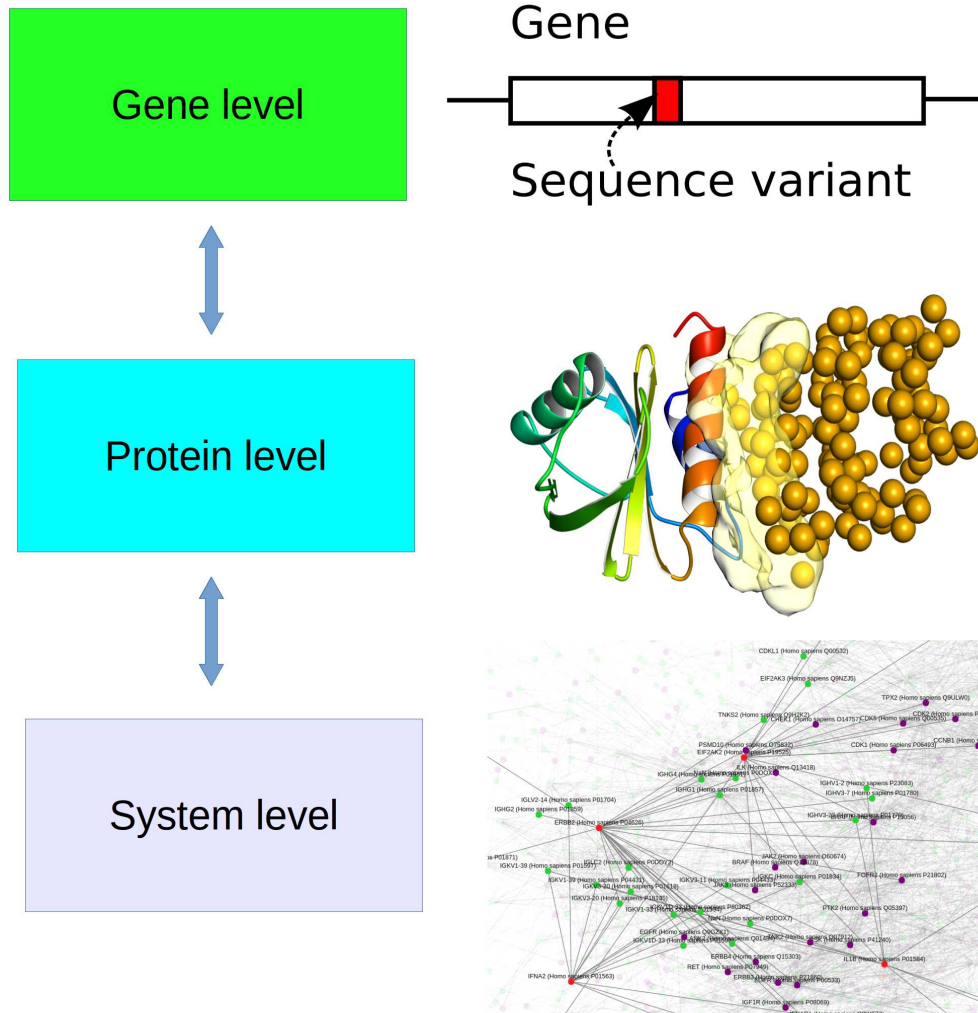


Prediction of the protein binding site, the ligand, the sequence variant, and their binding dynamics - ProBiS Tools



Janez Konc, Dušanka Janežič, ProBiS Tools (algorithm, database, and web servers) for predicting and modeling of pharmaceutically interesting proteins, *Progress in Biophysics & Molecular Biology*, 128, 24-32 **2017**.

ProBiS Tools at the PDB scale for drug discovery



Description and tools:

1. Mapping of sequence variants to protein binding sites (GenProBiS)
2. Prediction of binary protein-ligand interactions (ProBiS, ProBiS-ligands, ProBiS H2O)
3. Molecular simulations to study binding site dynamics (ProBiS-CHARMMing, CHARMM)
4. Protein interaction atlas for prediction of genetic variations involved in drug interactions and disease development (ProBiS-ATLAS)

Protein level

ProBiS: Predicted ligand complex

Internet

CHARMMing: Minimized complex, interaction energy

Molecular simulations to study binding site dynamics:

Combine ProBiS & CHARMMing:

ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites

<http://probis.nih.gov>

ProBiS-CHARMMing predicts minimizes ligands for any protein and can be used to generate holo protein structures from apo proteins (or prepare ligand-receptor complex for molecular dynamics simulation)

Janez Konc, Benjamin T. Miller, Tanja Štular, Samo Lešnik, H. Lee Woodcock, Bernard R. Brooks, Dušanka Janežič, ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *J. Chem. Inf. Model.*, 55, 2308, 2015.

ProBiS-CHARMMing Web Interface - Input: <http://probis.nih.gov>

ProBiS Server x
probis.nih.gov

ProBiS
CHARMMing

Search by PDB ID

As of Nov 29, 2013 your protein is compared with 37643 structures [HELP](#)

Introduction

- If you are new to ProBiS, you can watch the introductory video available here.

ProBiS-Tools

- ProBiS Input Page
- RESTful Web Services
- Database Access

Download

- ProBiS Algorithm
- Non-redundant PDB (nr-PDB)

Help

- ProBiS-ligands User's Guide
- ProBiS User's Guide
- Examples

Citation

Related Citations

- Konc,J. and Janezic,D. ProBiS-ligands: a web server for prediction of ligands by examination of protein binding sites. *Nucleic Acids Res.*, 2014, **42**, W215-W220.

Enter Query Protein

PDB ID: Chain ID(s):

Your e-mail address (optional):

A link to the results page will be sent to you by e-mail. Alignment of two proteins will take seconds.

Video Tutorial

Contact

- Scientific : [konc \[at\] cmm \[dot\] ki \[dot\] gov](mailto:konc[at]cmm[dot]ki[dot]gov)
- Technical : [btmiller \[at\] nhbi \[dot\] nih \[dot\] gov](mailto:btmiller[at]nhbi[dot]nih[dot]gov)

NIH National Institutes of Health
Turning Discovery Into Health

LoBoS Privacy Policy can be found here : [Privacy Policy](#)

ProBiS-CHARMMing Web Interface - Output: <http://probis.nih.gov>

ProBiS
CHARMMing

Search by PDB ID

As of Nov 29, 2013 your protein is compared with 37643 structures

3ks3.A A S H L C M
 2-chloro-4-[[[4,6-di... A S H L C M
 Mini1/2-chloro-4-[... A S H L C M
 Mini2/2-chloro-4-[... A S H L C M

Minimization
 2-chloro-4-[[[4,6-di...
 Inte. Energy:-59.2 kcal/mol
 Download CHARMM scripts
 Close

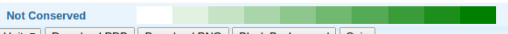
Set Minimization Parameters
 Gradient Tolerance: 0.1
 Steepest Descent Steps: 100
 ABNR Steps: 100
☐ Implicit Solvent

3KS3 , Chain A : 207 similar structures

Predicted Ligands Similar Proteins
 Small Molecules Proteins Nucleic Acids Ions

Binding Site 1 Binding Site 2 Binding Site 3 Binding Site 4 Binding Site 5 Binding Site 6 Binding Site 7 Binding Site 8
 Binding Site 18

Structure	Name	Source	Confidence	Binder	Ligand
	2-chloro-4-[[[4,6-dimethylpyrimidin-2-yl] sulfanyl]benzene	4knm	4.23	Specific	Remove
	2-chloro-4-[(pyrimidin-2-ylsulfanyl) acetyl]benzene	4knn	4.23	Specific	View 3D
	N-(4-chlorobenzyl)-n-methylbenzene-1,4-disulfonamide	3da2	4.23	Specific	View 3D
	N-(4-chlorobenzyl)-n-methylbenzene-1,4-disulfonamide	3da2	4.23	Specific	View 3D
	2,3,5,6-tetrafluoro-4-[(2-hydroxyethyl) sulfonyl]benzoic acid	4hu1	4.23	Specific	View 3D
	Acetate ion	3d0n	4.23	Non-specific	View 3D
	2-chloro-4-[(pyrimidin-2-ylsulfanyl) acetyl]benzene	4knn	4.23	Specific	View 3D
	2,3,5,6-tetrafluoro-4-[(2-hydroxyethyl) sulfonyl]benzoic acid	4hu1	4.23	Specific	View 3D
	2-chloro-4-[[[4,6-dimethylpyrimidin-2-yl] sulfanyl]benzene	4knm	4.23	Specific	View 3D
	5-acetamido-1,3,4-thiadiazole-2-sulfonamide	3czv	4.23	Specific	View 3D
	5-acetamido-1,3,4-thiadiazole-2-sulfonamide	3ml5	3.9	Specific	View 3D
	6-ethoxy-1,3-benzothiazole-2-sulfonamide	3mdz	3.9	Specific	View 3D

Not Conserved  Conserved

Asymmetric Unit Download PDB Download PNG Black Background Spin JSmol

NIH National Institutes of Health
Turning Discovery Into Health

LoBoS Privacy Policy can be found here : [Privacy Policy](#)



Minimization in the ProBiS-CHARMMing web interface.

ProBiS-CHARMMing Web Interface - Output: <http://probis.nih.gov>

probis.nih.gov/browse.php?job_id=1AZE.A#

ProBiS
CHARMMing

Search by PDB ID

As of Jul 31, 2015 your protein is compared with 42270 structures [HOME](#) | [HELP](#)

1AZE, Chain A : 245 similar structures

Download Table

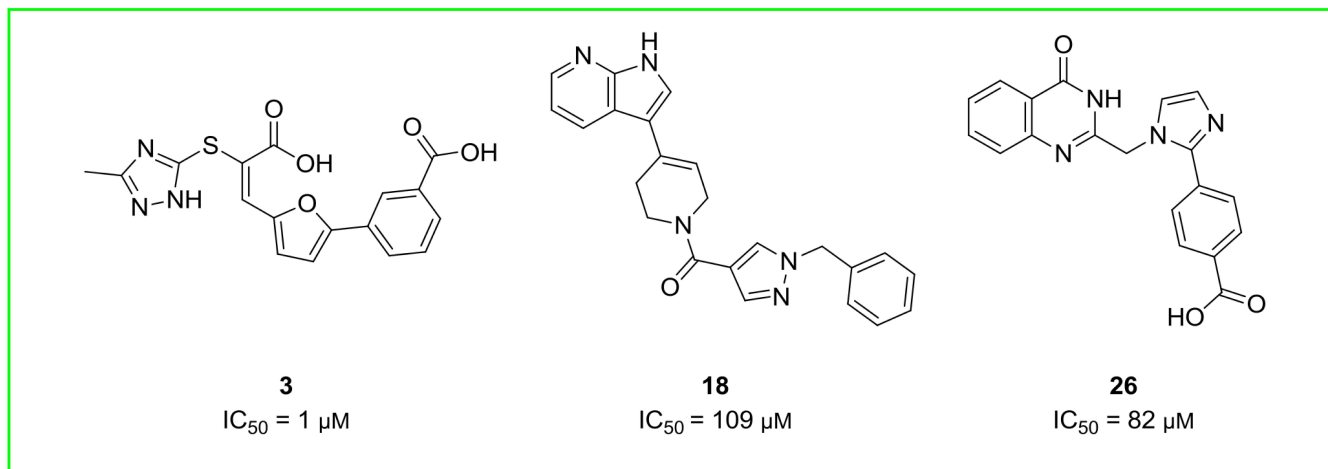
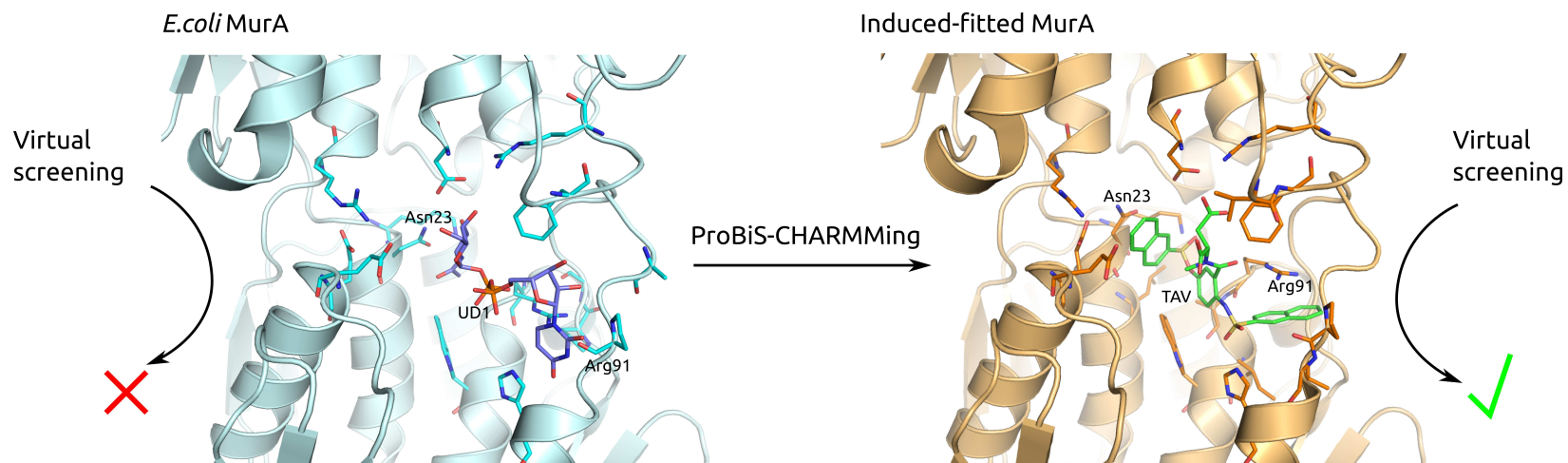
Rank	Alignments	Chain	Name	Pfam	SCOP	UniProt	Z-Score
166	View 3D	1s1n.A	Sh3 domain of tyrosine-protein kinase itk/tsk	PF00018	-	Q03526	2.23
167	View 3D	2r0.A	Mitogen-activated protein kinase kinase kinase	PF14604	-	Q02779	2.23
168	View 3D	2lqw.A	Proto-oncogene tyrosine-protein kinase src	PF00018	-	P00523	2.22
169	View 3D	2egc.A	Cytoplasmic protein nck2	PF00018	b.34.2.1	O43639	2.22
170	View 3D	2b86.A	Bud emergence protein 1	PF00018	-	P29366	2.22
171	View 3D	2jxb.A	Intersectin 2	PF07653	b.34.2.1	Q9NZM3	2.22
172	View 3D	2lpe.A	Phosphatidylinositol 3-kinase regulatory subunit	PF07653	-	P27986	2.20
173	View 3D	1nm7.A	Nephrocystin 1	PF00018	-	O15259	2.20
174	View 3D	2k6d.A	Crk-like protein	PF00017	-	P46109	2.17
175	View 3D	2tpf.A	Sh3 and px domain-containing protein 2a	-	-	-	2.17
176	View 3D	3o5z.A	Cytoplasmic protein nck2	PF00018	-	O43639	2.16
177	View 3D	2rol.A	T-cell surface glycoprotein cd3 epsilon chain, cy	PF00018	-	O43639	2.14
178	View 3D	4le9.A	C-jun-amino-terminal kinase interacting protein	PF14604	-	Q9R237	2.14
179	View 3D	2oj2.A	Peroxisomal membrane protein pas20	PF00018	b.34.2.1	P80667	2.14
180	View 3D	2p21.A	Sh3 domain-containing kinase-binding protein 1	PF14604	-	Q9B897	2.13
181	View 3D	2km.A	C-jun-amino-terminal kinase interacting protein	PF14604	-	Q9R237	2.13
182	View 3D	2d8j.A	Phosphatidylinositol 3-kinase regulatory subunit	PF00018	-	Q8TE68	2.13
183	View 3D	2d0k.A	Epidermal growth factor receptor kinase substr	PF00018	-	Q8TE68	2.13
184	View 3D	2l0a.A	Proto-oncogene tyrosine-protein kinase src	PF00018	-	P00523	2.12
185	View 3D	2djq.A	Hematopoietic cell kinase, sh3 domain	PF00018	-	P08631	2.12
186	View 3D	2d7.A	Rho guanine nucleotide exchange factor 4	PF00169	-	Q9NR80	2.11
187	View 3D	1x6b.A	Cd2-associated protein	PF14604	-	Q9JLQ0	2.11
188	View 3D	1aww.A	Fyn-related kinase	PF00018	-	Q922K9	2.10
189	View 3D	2enm.A	Crk-like protein	PF07653	-	P46109	2.09
190	View 3D	1zuu.A	Signal transducing adapter molecule 1	PF00018	-	Q92783	2.08
191	View 3D	2ega.A	Sh3 domain containing ring finger 2	PF07653	-	Q8BZT2	2.06
192	View 3D	1ycs.B	Kiaa0769 protein	PF14604	-	Q94868	2.06
193	View 3D	4hck.A	Rho guanine exchange factor (gef) 16	PF14604	-	Q5VV41	2.05
194	View 3D	2nrm.A	Bruton's tyrosine kinase	PF00018	b.34.2.1	Q06187	2.05
195	View 3D	2zua.A	Sorting nexin-9	PF00018	-	Q91VH2	2.05
196	View 3D	2d7.A	Bzz1 protein	PF00018	b.34.2.1	P38822	2.03
197	View 3D	2d7.A	Sh3 and px domain-containing protein 2a	PF00018	-	Q5TC21	2.02
198	View 3D	2d7.A	53bp2	PF00018	b.34.2.1 d.21	Q13625	2.01
199	View 3D	2d7.A	Hematopoietic cell kinase	PF00018	b.34.2.1	P08631	1.99
200	View 3D	2d7.A	Activated raf1 kinase	PF00018	b.34.2.1	Q07012	1.99

NIH National Institutes of Health
Turning Discovery Into Health

LoBoS Privacy Policy can be found here: [Privacy Policy](#)

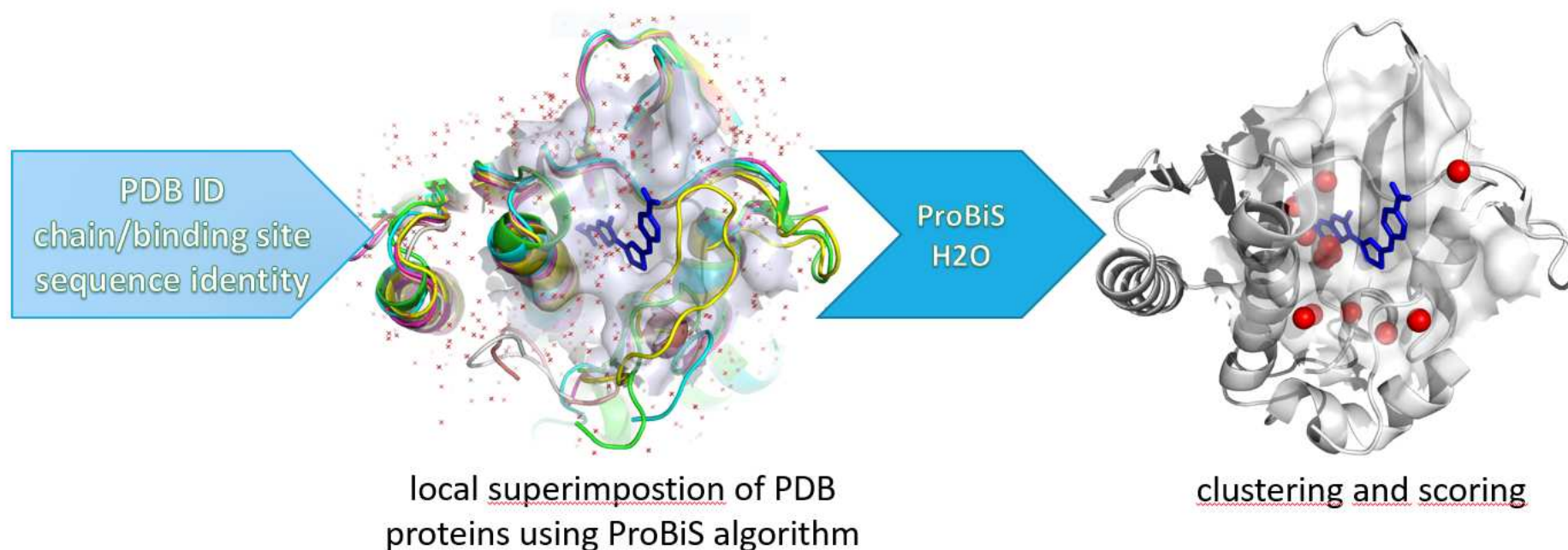
Minimization in the ProBiS-CHARMMing web interface - pairwise alignment with ligand.

Inhibitors of MurA enzyme from *E. coli* - Application in drug discovery



Protein level

ProBiS H2O: Identification of conserved water sites in proteins



J. Chem. Info. Model., 57(12), 3094-3103, 2017.

ProBiS H₂O plugin @ <http://insilab.org/probis-h2o>

Input: PDB ID / Output: identified water clusters

ProBiS H₂O Plugin

define system | cluster analysis | About/Help

define system

PDB ID: **1** 14duh

☒ Analyze binding site (default)

☐ Define water as binding site

☐ Compare whole chain

☐ Debye-Waller correction

☐ Single aligned chain per PDB entry

Sequence identity: 95 **2** Find

Found 19 compl in cluster no.:7547 **(3)** Download

Select BINDING SITE / CHAIN

RLI.301.B

RLI.301.A **4**

Identify

SETUP DB **5** GO

ProBiS H₂O Plugin

define system | cluster analysis | About/Help

Info:

Master water list includes 5683 molecules

Superimposed chains: 24

System volume is: 4435 cubic A

Maximum occupied cluster contains 23 H2O molecules

binding site: RLI.301.B

Calculated clusters, conservation; keep selected display: ☐

1 cluster with 23 H2O molecules, consv. 0.96

2 clusters with 18 H2O molecules, consv. 0.75

3 clusters with 13 H2O molecules, consv. 0.54

4 clusters with 12 H2O molecules, consv. 0.5

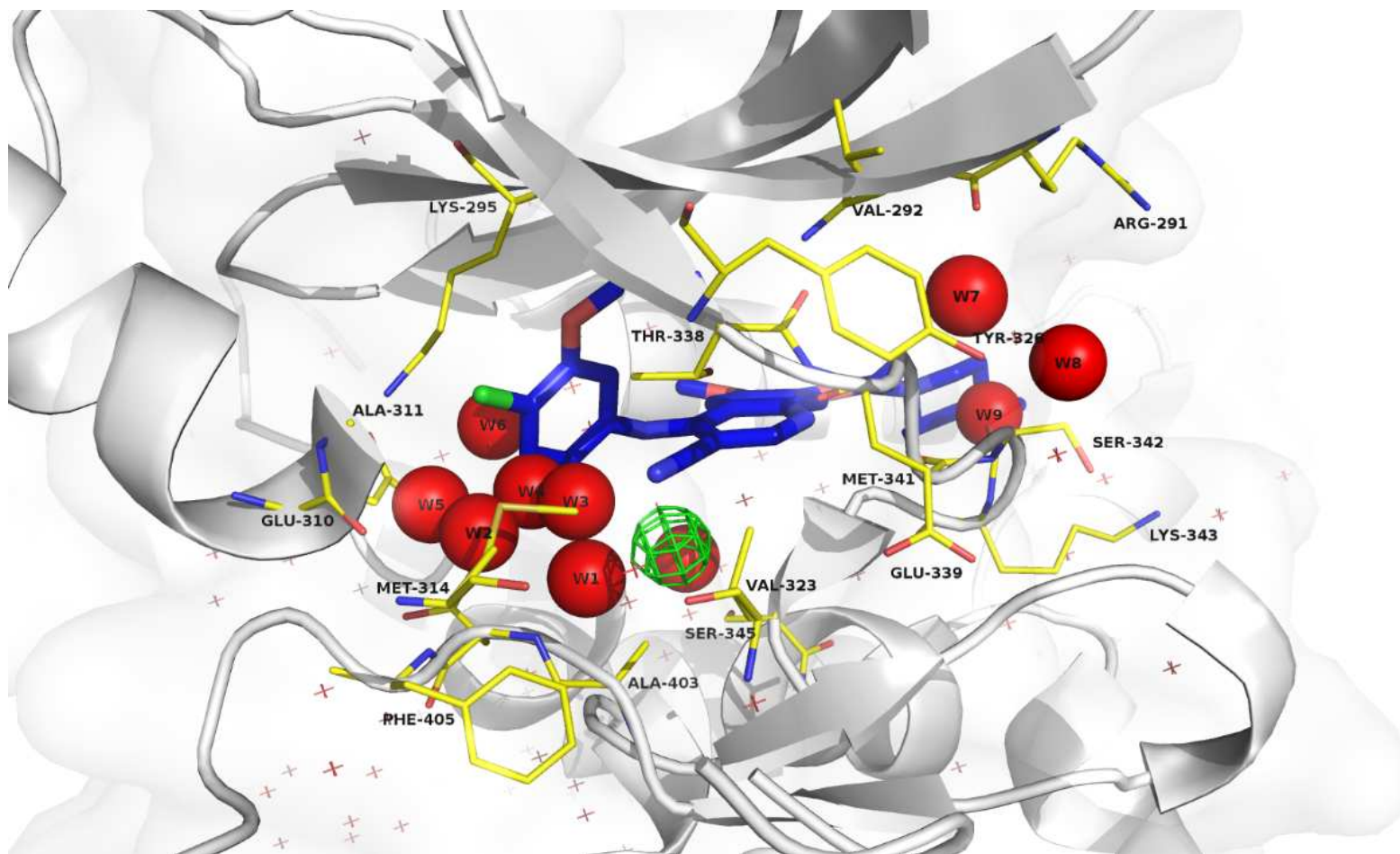
6 clusters with 11 H2O molecules, consv. 0.46

7 clusters with 9 H2O molecules, consv. 0.38

10 clusters with 8 H2O molecules, consv. 0.33

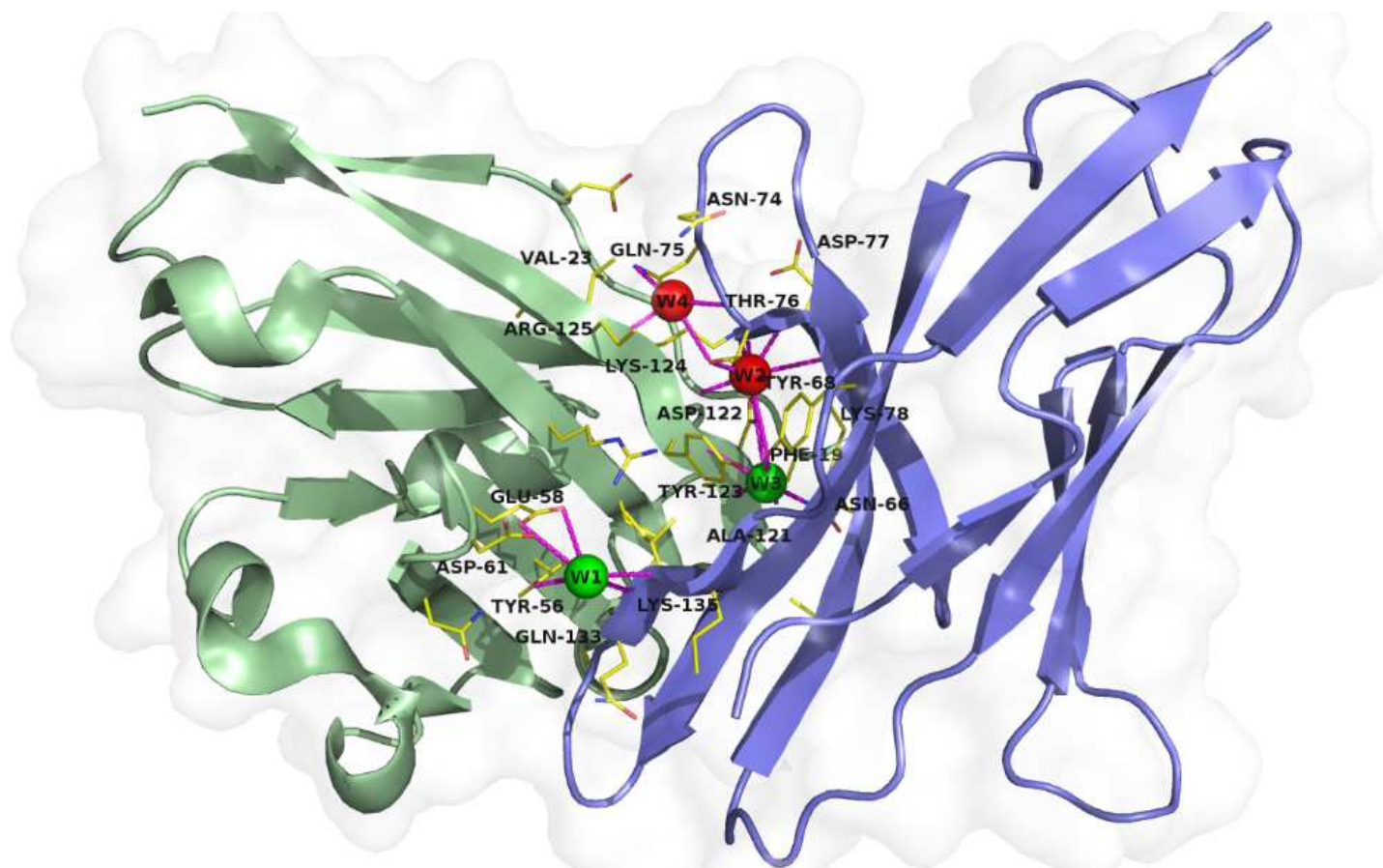
fetch/reset chain box contacts b-site display

ProBiS H2O Validation: Anticancer Drug Selectivity



Bosutinib selectivity towards Src kinase mediated by conserved waters.
Known conserved water clusters W1-W2, newly found W3-W6 and W7-W9.

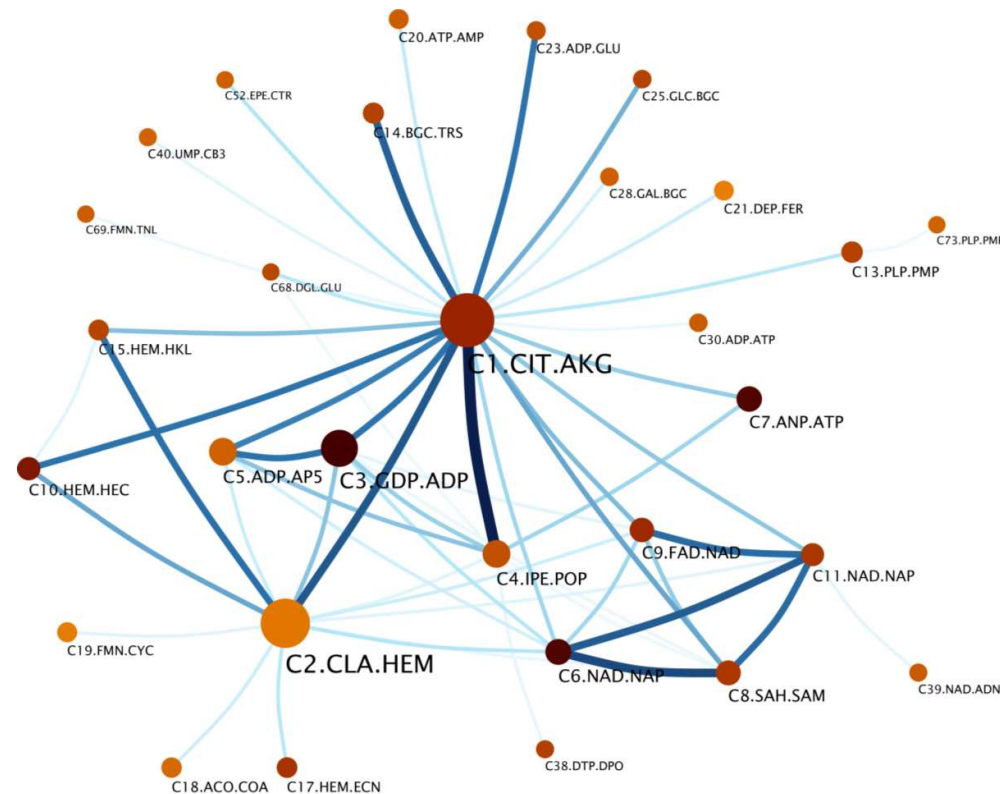
ProBiS H2O Validation: Conserved Waters at Protein-Protein Interface



Interface of hPD-1 protein (light blue) and its ligand hPD-L1 (light green). Known conserved water clusters W1, W3 (green), newly found W2, W4 (magenta).

Protein level

Binding site community network generated by ProBiS



Global organization of a binding site network gives insight into evolution and structure-function relationships of proteins. The binding site included in larger community may be older and have been evolutionary structural scaffolds of more recent ones. *SCIENTIFIC REPORTS*, 7,11652, 2017.

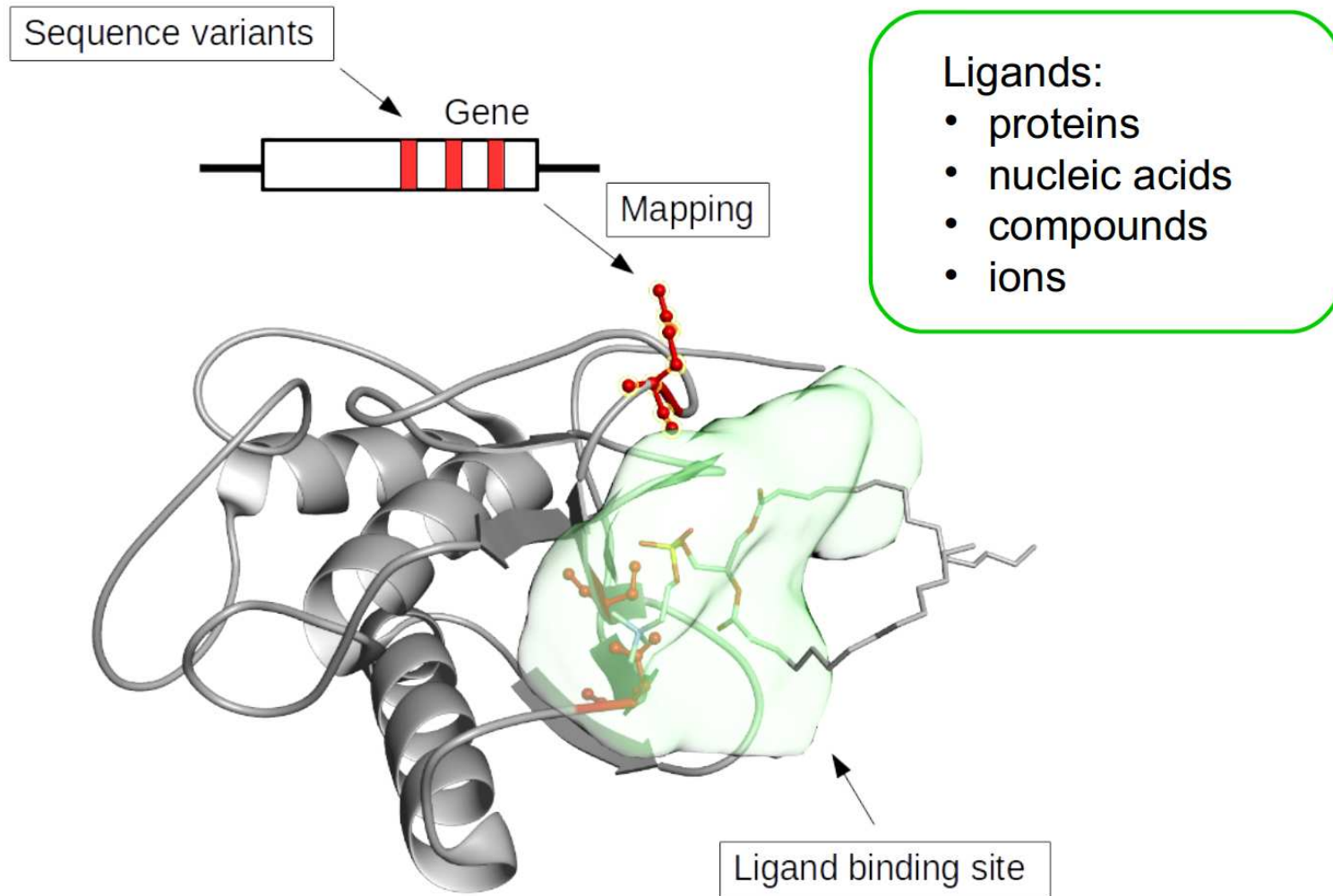
Gene level

GenProBiS: Sequence variants in protein binding sites

<http://genprobis.insilab.org> – *Nucleic Acids Research*, 45, W253-W259, 2017.

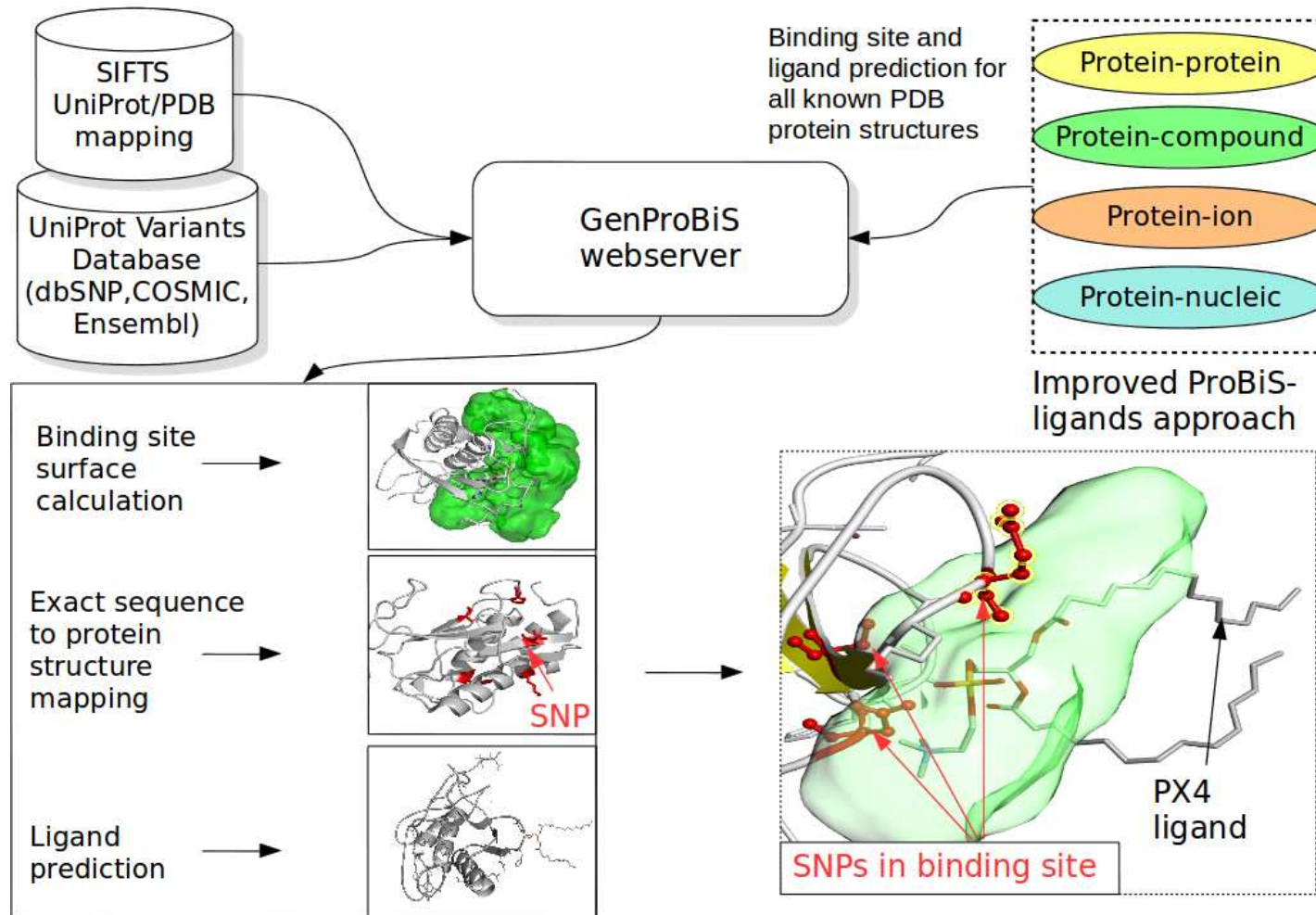
- GenProBiS web server maps sequence variants to protein structures from the PDB, and further to protein-protein, protein-nucleic acid, protein-compound, and protein-metal ion binding sites.
- The concept of a protein-compound binding site is understood in the broadest sense, includes glycosylation and other post-translational modification sites.
- Binding sites and their ligands are predicted with no prior knowledge of binding sites, but based on detected local structural similarities in proteins and transposition of ligands between protein structures irrespective of protein folding.
- GenProBiS allows suggestion of functional effects of mutations on ligand binding and as such represents a key tool in both drug discovery and personalized medicine.

GenProBiS: Sequence variants in protein binding sites



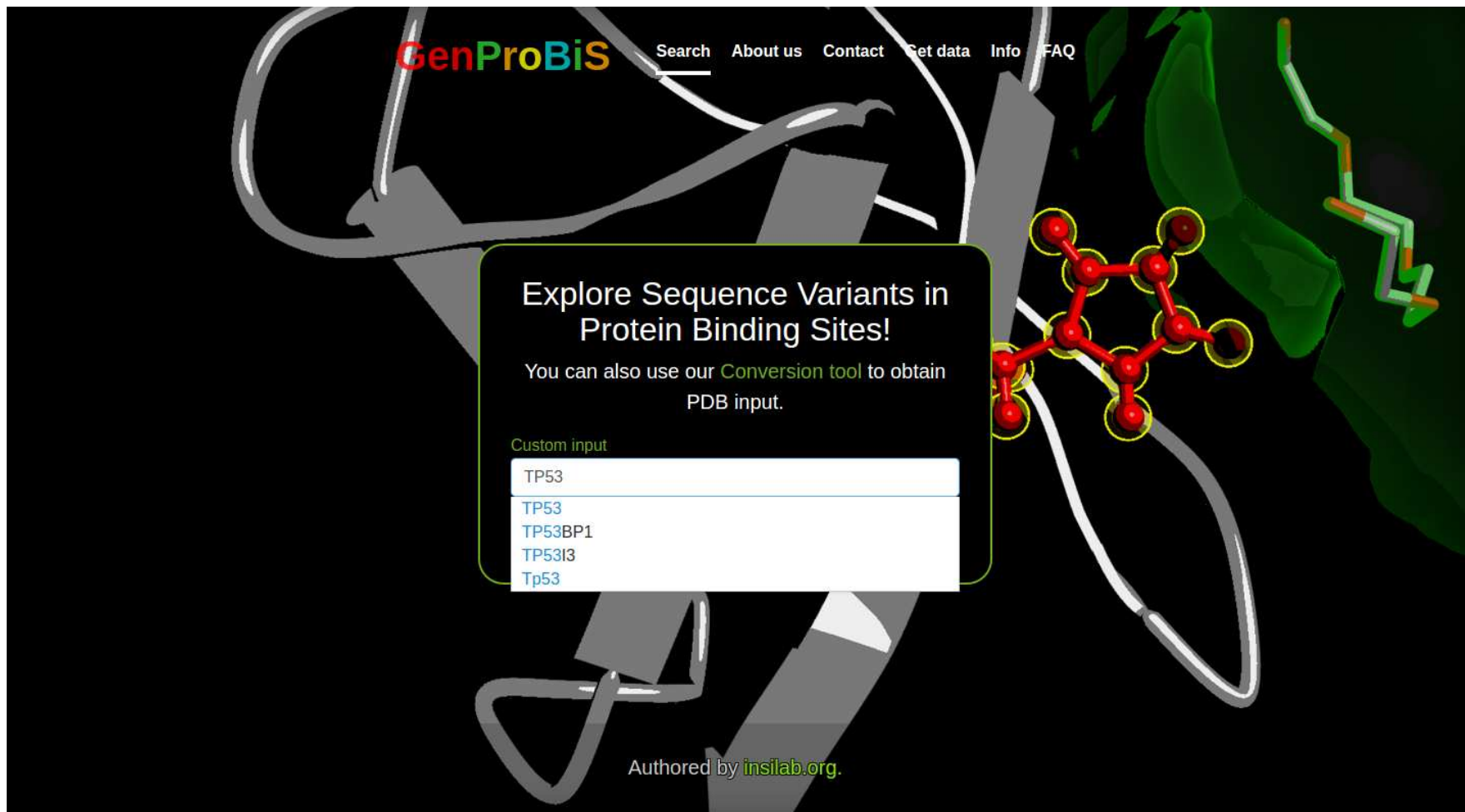
Mapping sequence variants to protein binding site

GenProBiS: Sequence variants in protein binding sites



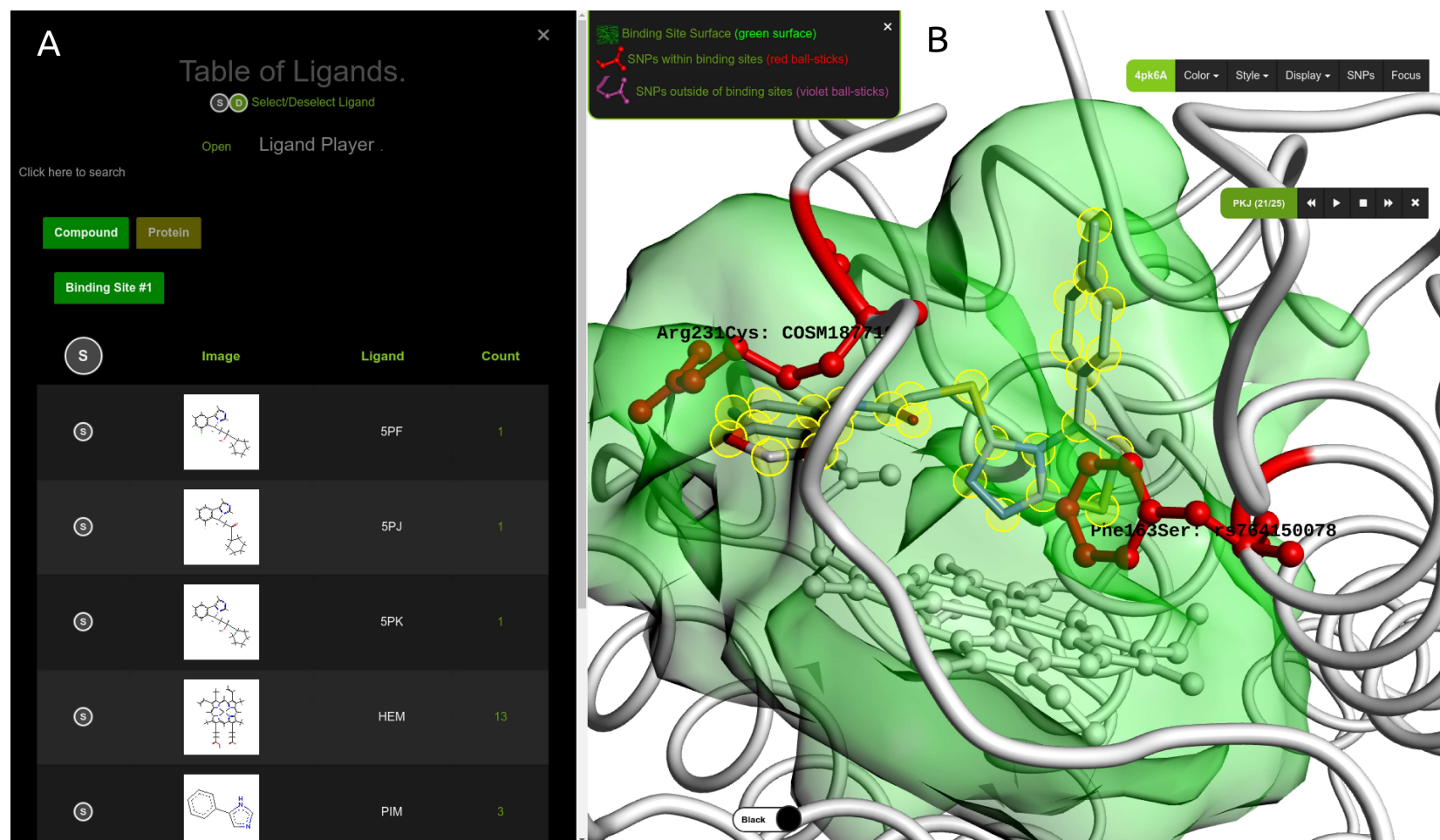
GenProBiS web server approach depicted on example of nsSNPs mapping to a compound (low molecular weight ligand) binding site.

GenProBiS: Sequence variants in protein binding sites



Input: PDB & Chain ID, dbSNPs reference SNP cluster (rs) ID, COSMIC's Mutation ID, Uniprot ID or Gene Symbol

GenProBiS: Sequence variants in protein binding sites



GenProBiS output page for indoleamine 2, 3-dioxygenase as the query protein (PDB and Chain ID: 4pk6A). Interactions of SNP's with inhibitor, example of drug response.

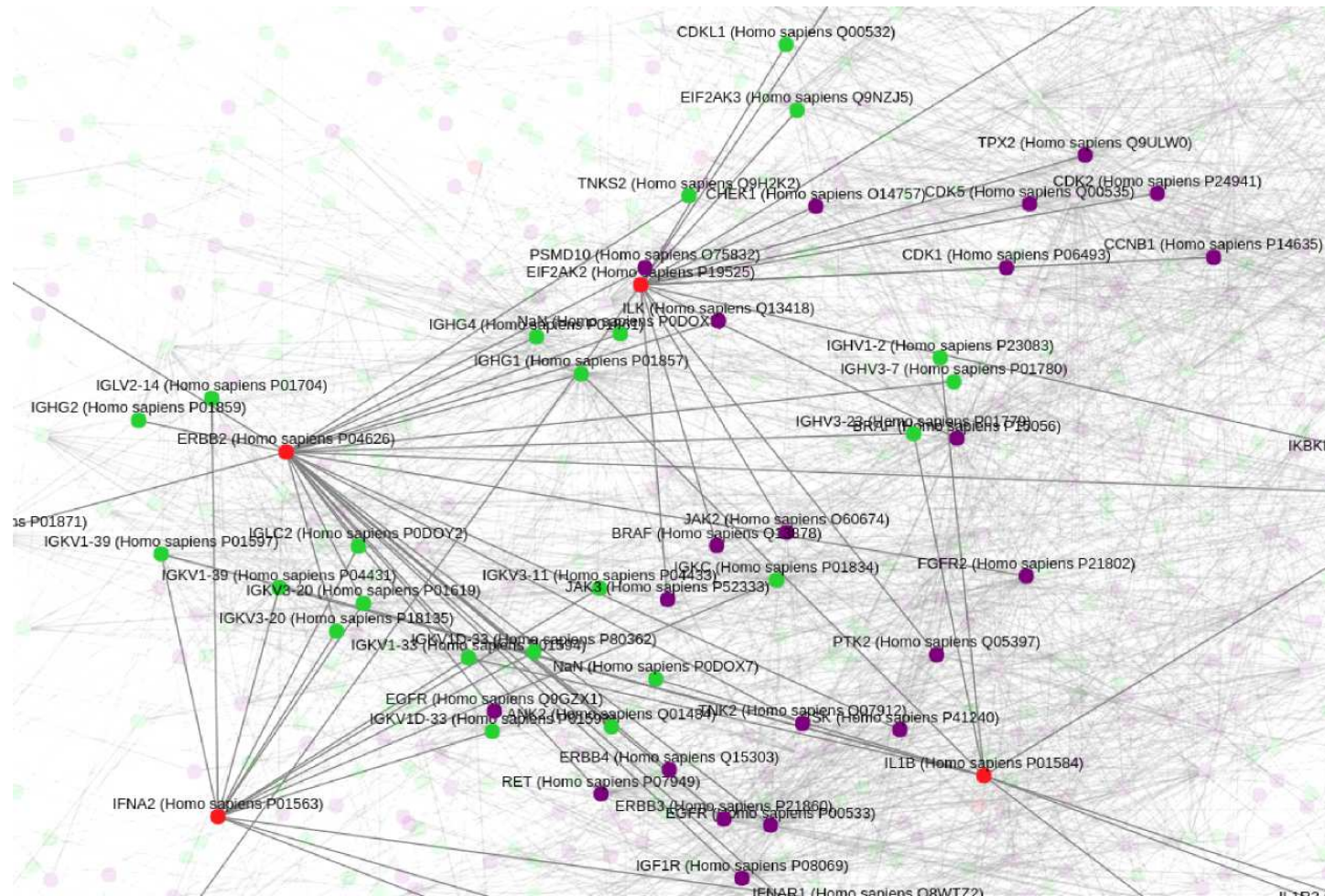
System level

ProBiS-ATLAS: Protein interaction atlas for prediction of genetic variations involved in drug interactions and disease development

- Proteins form a complex network of physical interactions. Understanding these interactions on the scale of the structural interactome is an open problem that, if solved, could provide insights into the development of various diseases and protein function on the system level.
- The global protein-ligand interaction network, in which ligands will be other proteins, compounds, ions, cofactors, nucleic acids or conserved water molecules will be developed.
- Each node in the network will represent a protein or a ligand and an edge will represent a physical protein-ligand interaction.
- Two types of interactions will be considered: those that directly correspond to a determined protein-ligand complex in the PDB, and those that will be predicted to occur based on the determined binding site similarities using our previously developed ProBiS algorithm.

- The network will include many layers of biological annotations, from genomic to proteomic to structural and will consist of more than 100,000 protein nodes, and presumably millions of non-protein nodes representing other ligand types.
- To explore this network in an intuitive manner, we will develop ProBiS-ATLAS, a graphical web-based atlas that will enable the network to be viewed globally, similar to the Earth in Google Maps, or locally by focusing on a layer of interactions around a single protein.
- This atlas will enable finding interactions between protein structures, predicting ligand selectivity and binding, and observing effects of conserved waters and sequence variants on ligand binding.
- This work will provide the scientific community with a free, versatile server readily usable from the web by researchers and clinicians to discover molecules binding to proteins, for example, anti-cancer drugs, and to assess their effects on populations having certain sequence variants.

- ProBiS-ATLAS will be used in precision medicine for prediction of drug response.



Conclusions

- Our newly developed approaches are particularly useful in the context of precision medicine.
- Our tools enable joining several otherwise disconnected areas of research, for example, graph-theoretical approaches, genome sequence studies, protein structures, and MD simulations.

Acknowledgements

SLO Group:

Marko Jukič, Stanislav Gobec, Tanja Kunej, Tanja Štular, Samo Lešnik,
Blaž Škrlj, Nika Eržen

USA Group:

Bernard R. Brooks, Benjamin T. Miller, NIH
H. Lee Woodcock, USF

Funding:

The Slovenian Research Agency (ARRS)

Donation:

We gratefully acknowledge the support of NVIDIA Corporation with the donation of Tesla K40 and two Tesla K80 GPUs used for this research.

University of Primorska Computer Cluster

The cluster is composed of 21 PCs with INTEL Xeon E5-2630v3 processors, 360 cores, connected through 10 GB Ethernet switch, and also includes NVIDIA two Tesla K80 GPUs and Tesla K40 GPU Accelerator 12 GB, a donation from NVIDIA Corporation that we gratefully acknowledge.



Map of Slovenia

