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

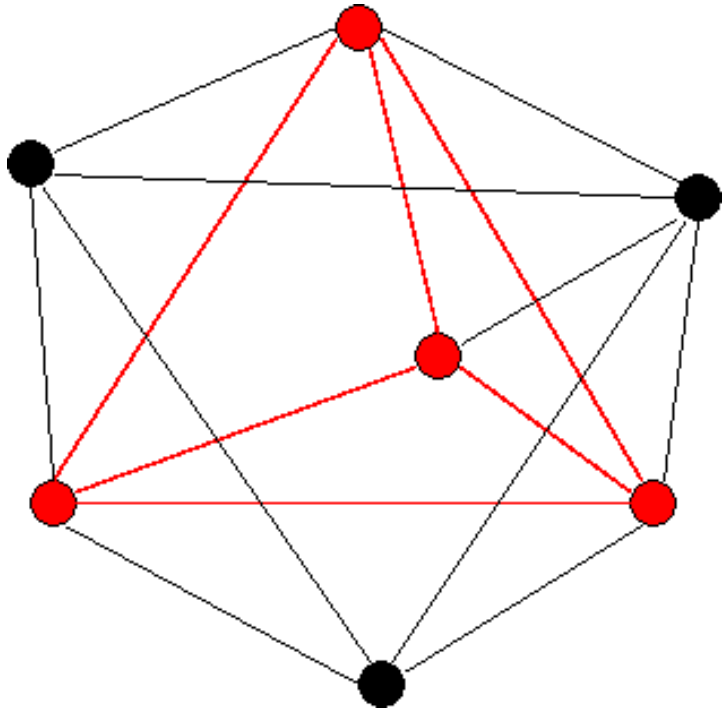
Outline

- 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.

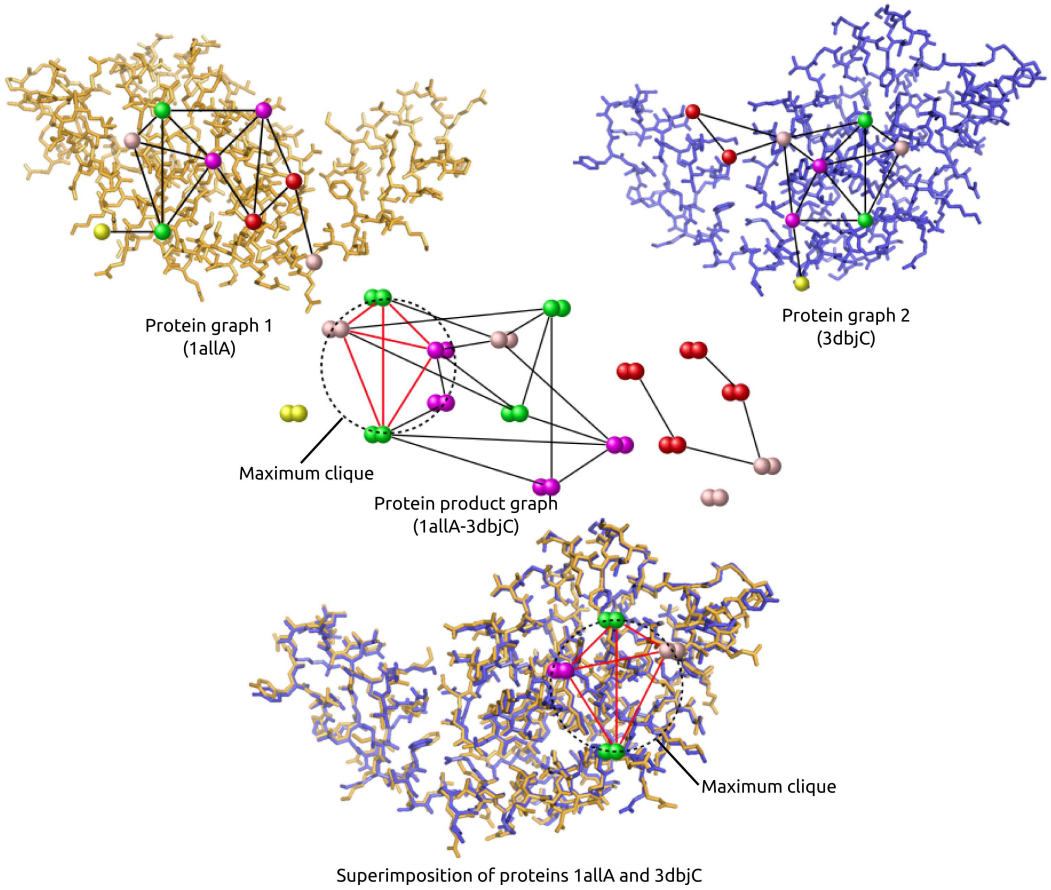
An undirected graph G=(V,E); V is a set of vertices and E is a set of edges.



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

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

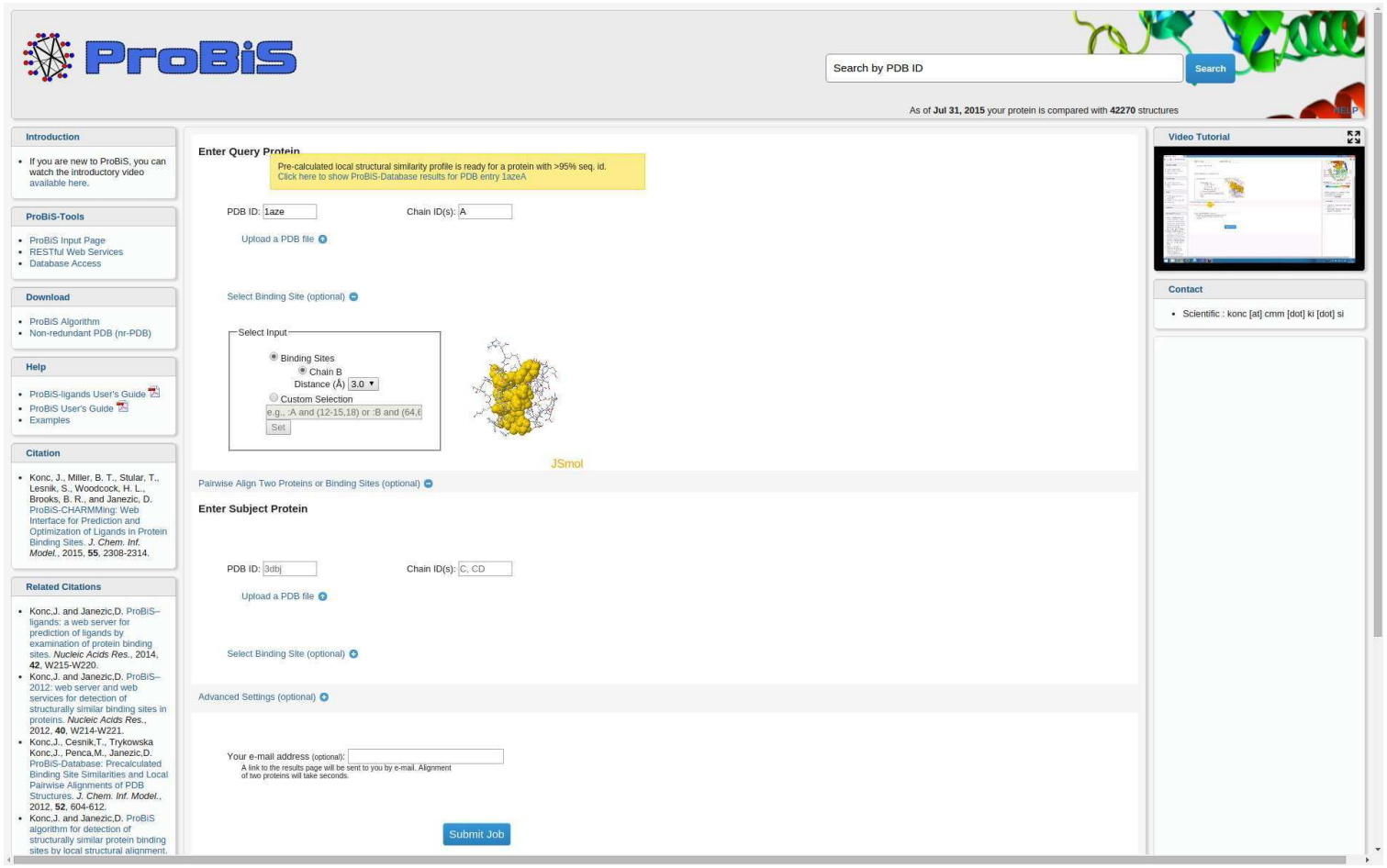
ProBiS Algorithm



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.

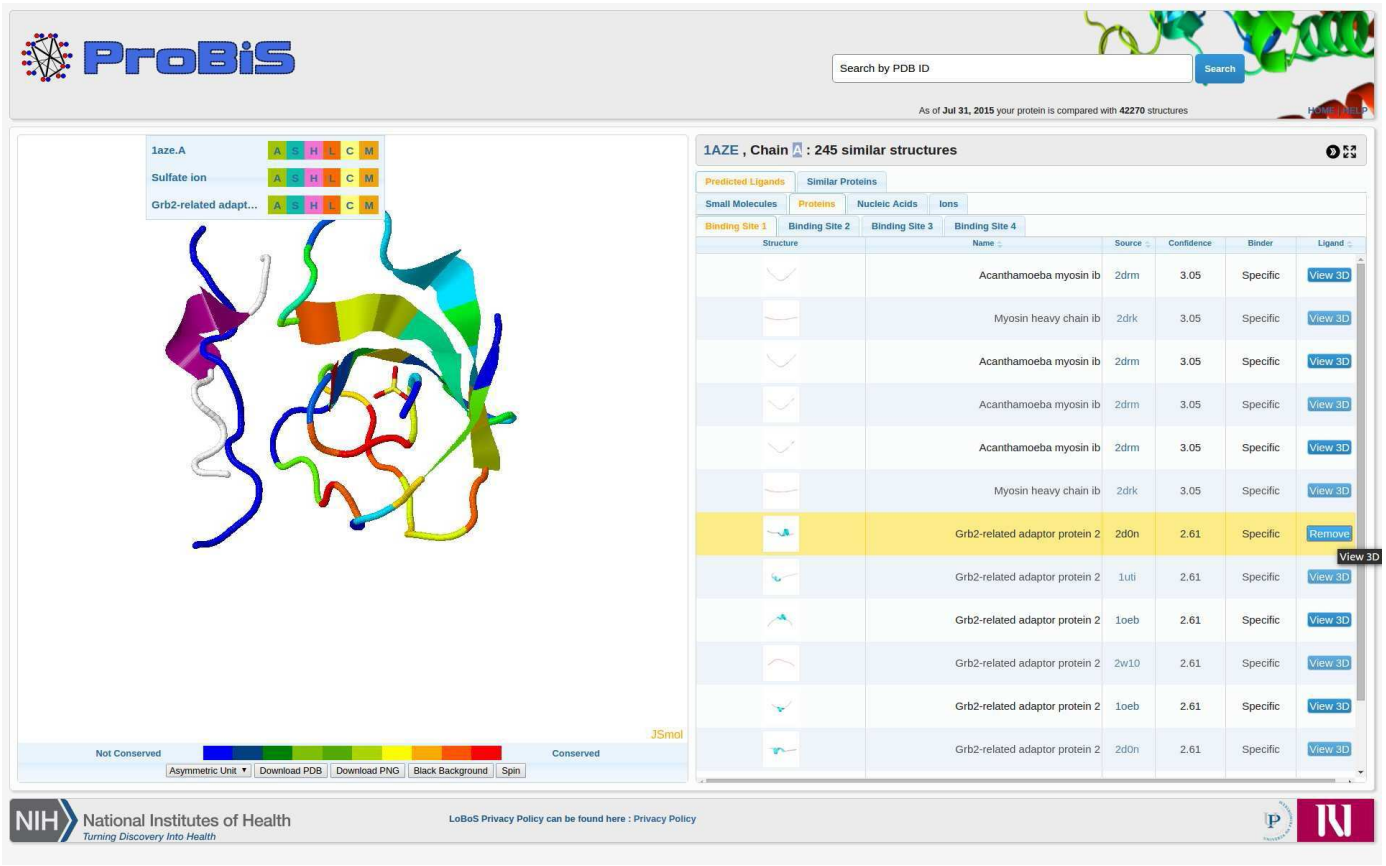
J. Konc and D. Janežič,*Bioinformatics*, 26, 1160 **2010**, & *J. Chem. Inf. Model.*, 53, 2217 **2013**.

ProBiS Web Server – Input



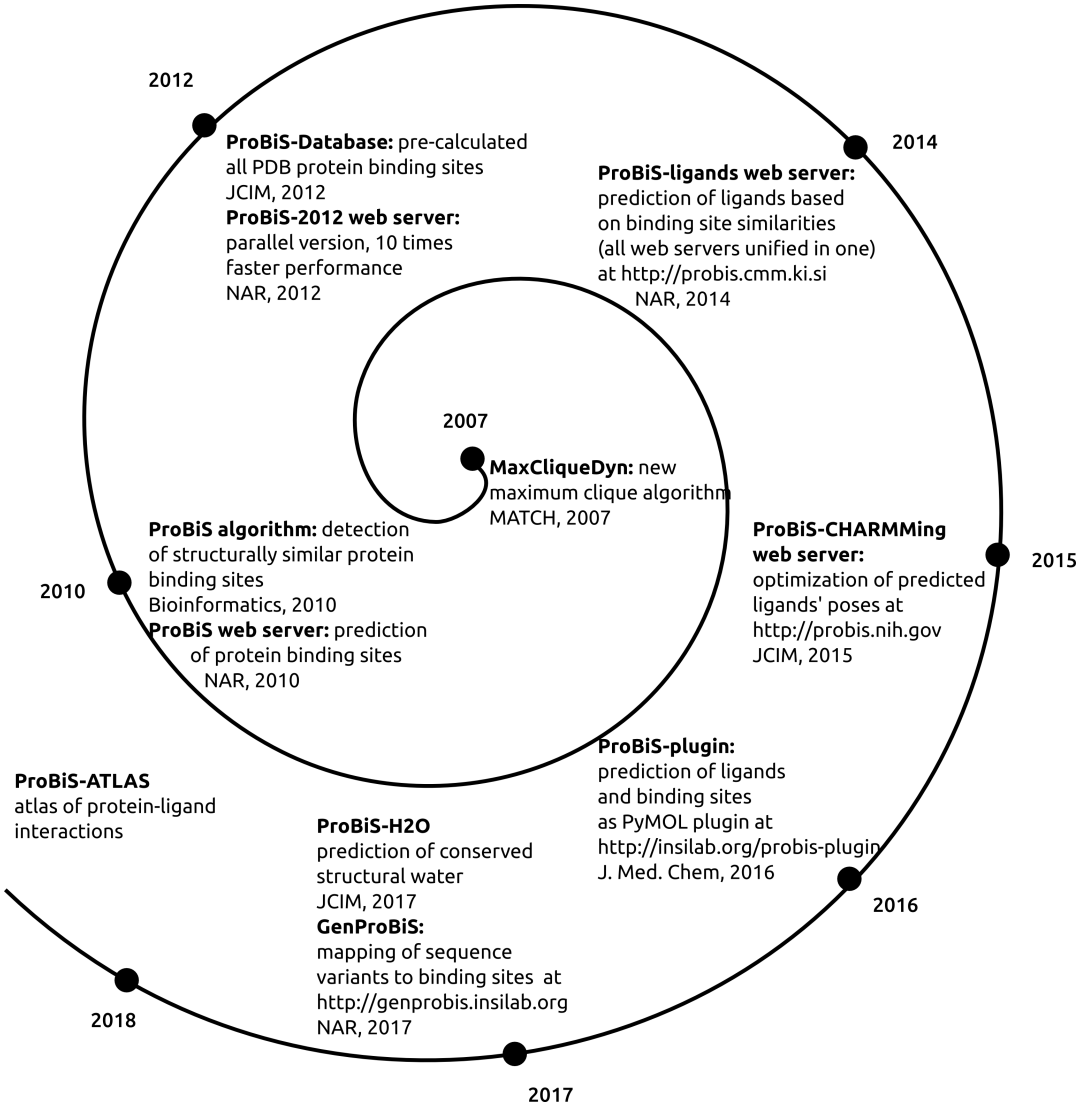
Detection of structurally similar binding sites in PDB and local pairwise alignment of protein structures.

ProBiS Web Server – Output

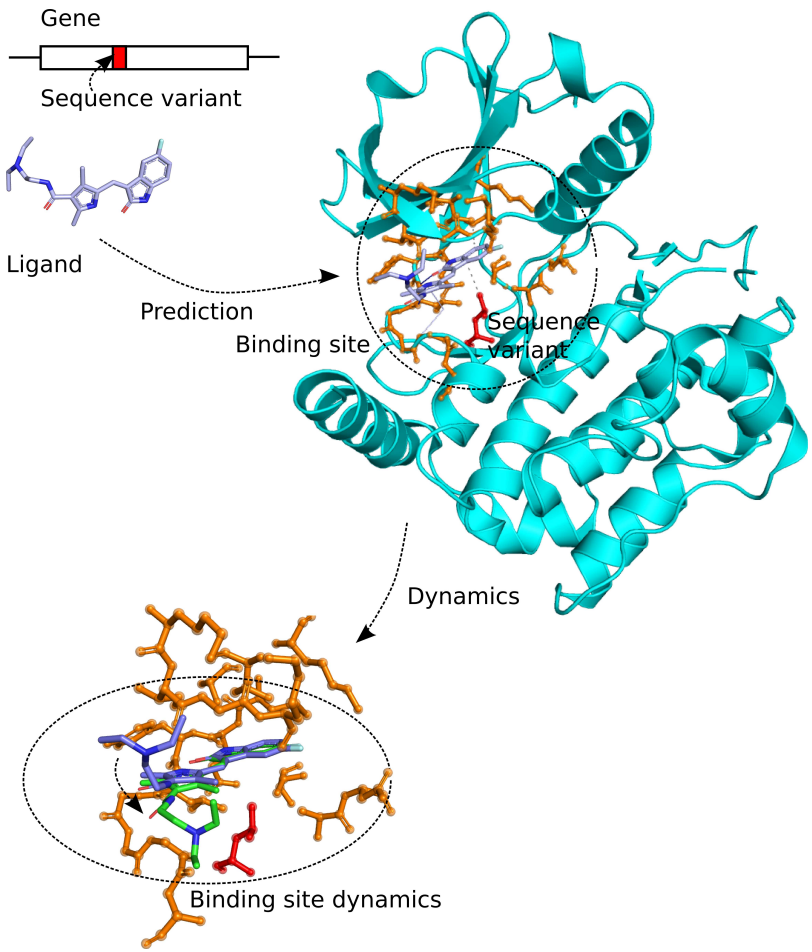


Predicted protein binding site and ligand as visualized in the ProBiS web server.

Evolution of the ProBiS Tools

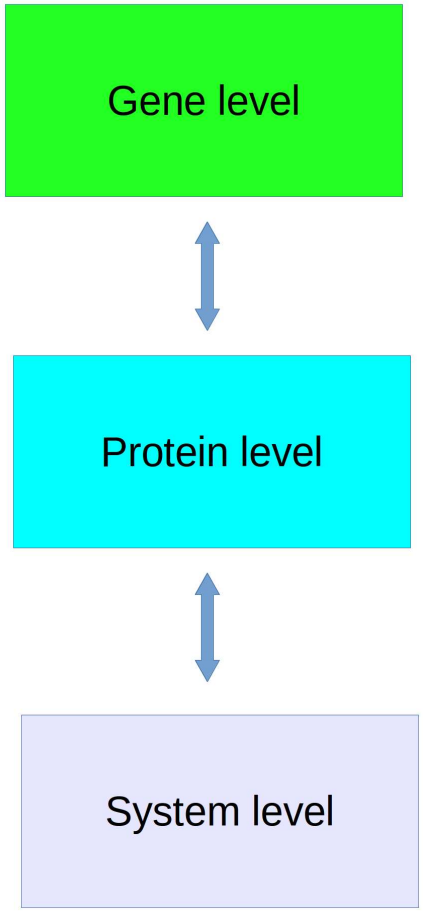


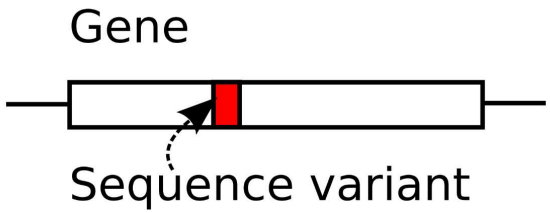
Prediction of the protein binding site, the ligand, the sequence variant, and their binding dynamics - ProBiS Tools

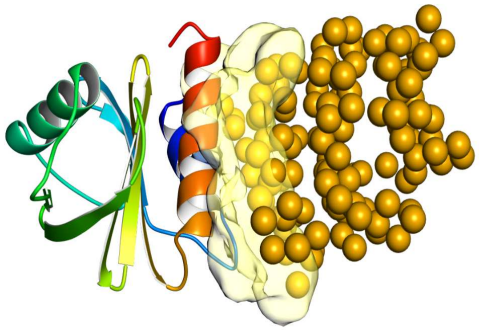


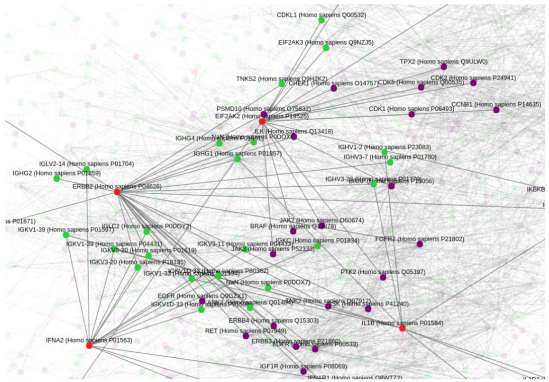
Janez Konc, Dušanka Janežič, ProBiS Tools (algorithm, database, and web servers) for predicting and modeling of pharmaceutically interesting proteins, *Progress in Biophysics & Molecular Biology*, 128, 24-32 **2017**.

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 proten-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:

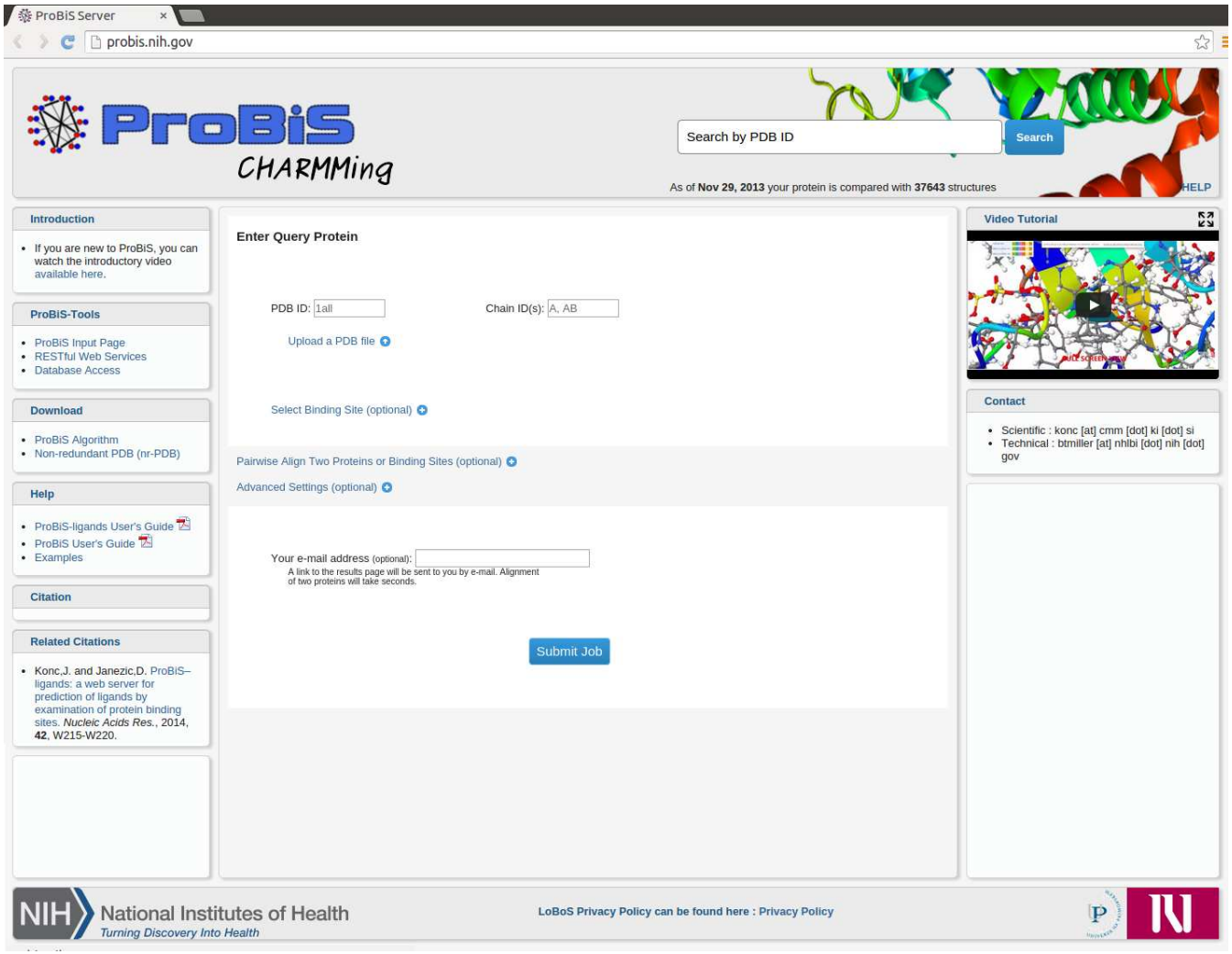
ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites

http://probis.nih.gov

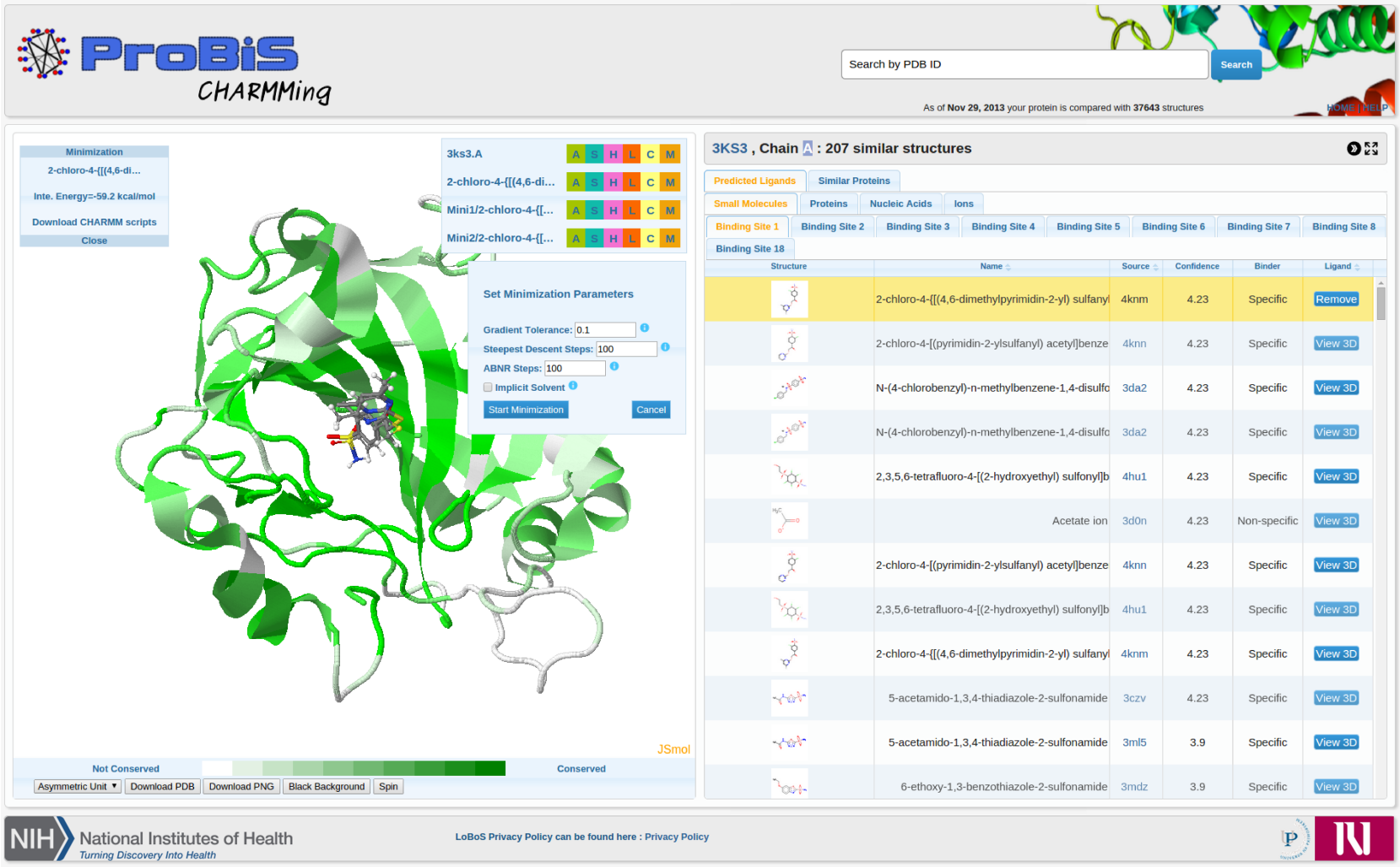
ProBiS-CHARMMing predicts minimizes ligands for any protein and can be used to generate holo protein structures from apo proteins (or prepare ligand-receptor complex for molecular dynamics simulation)

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

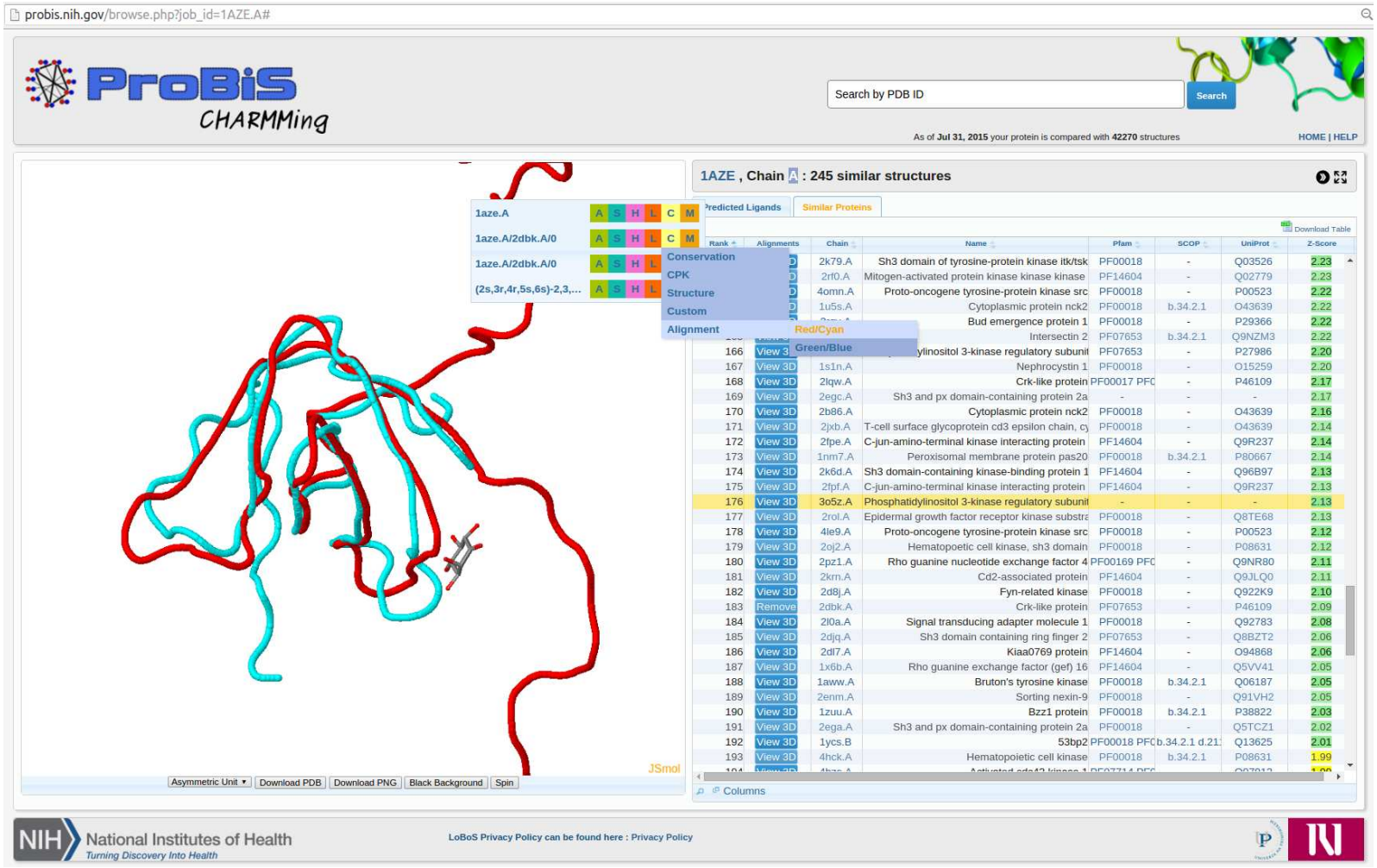


ProBiS-CHARMMing Web Interface - Output: http://probis.nih.gov



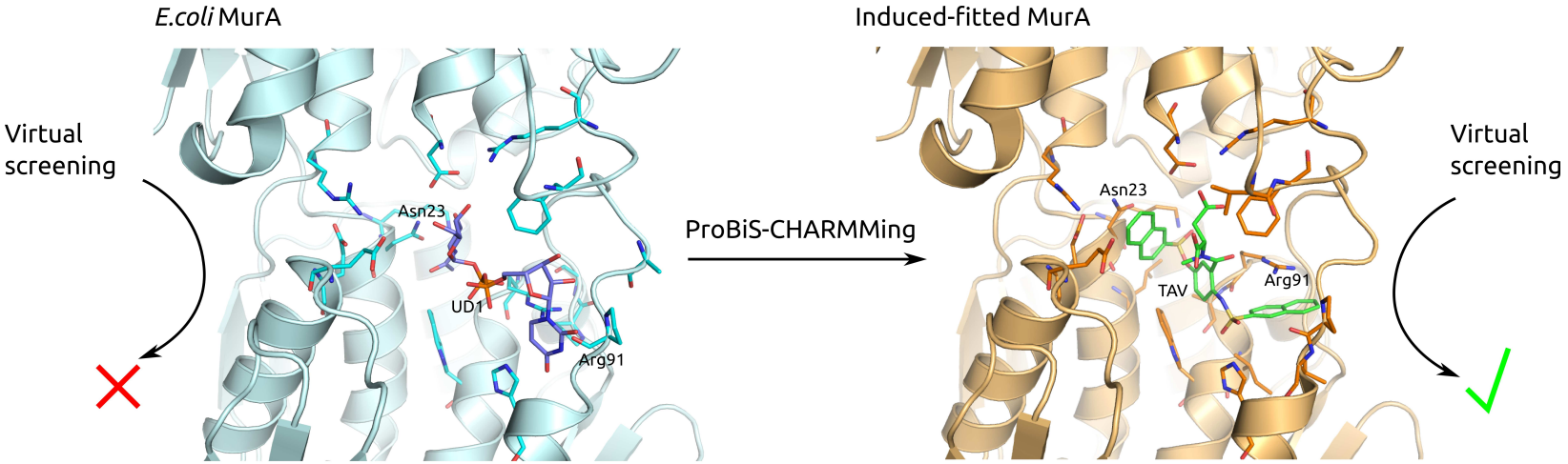
Minimization in the ProBiS-CHARMMing web interface.

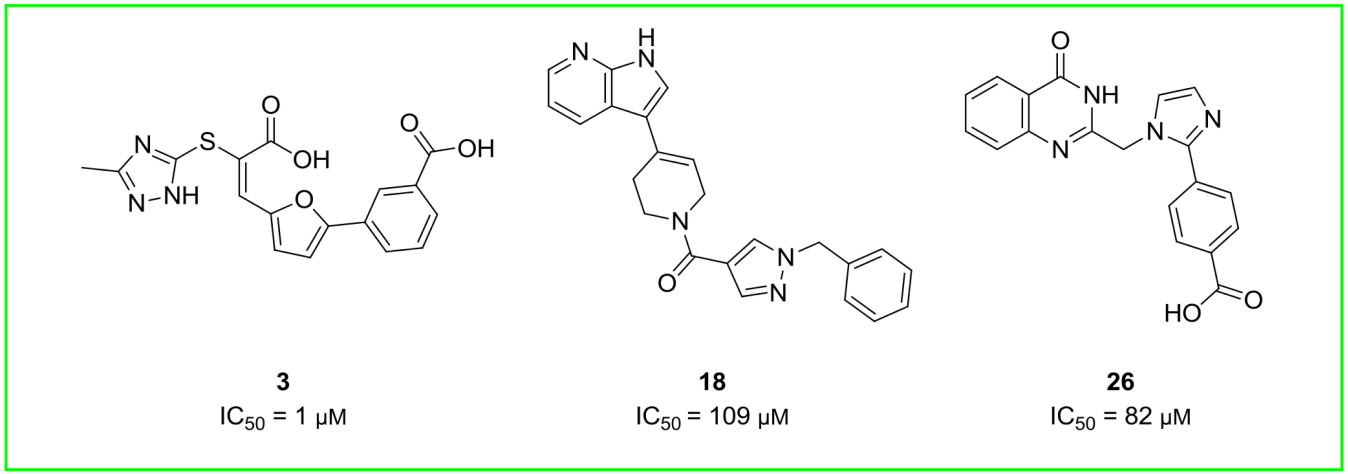
ProBiS-CHARMMing Web Interface - Output: http://probis.nih.gov



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

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

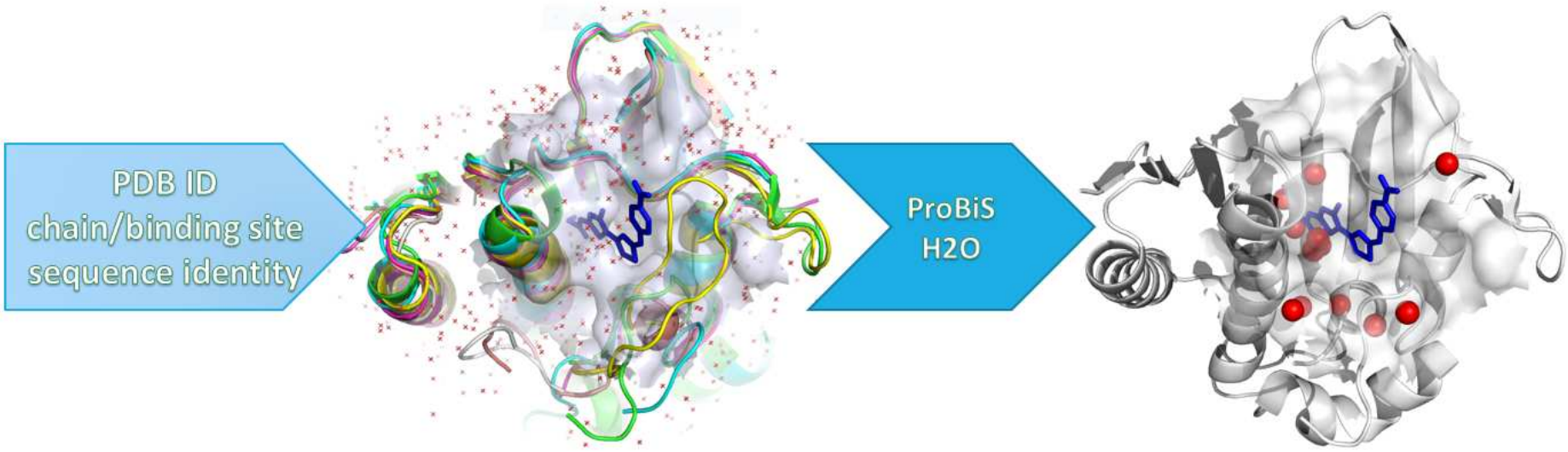




Bioorganic Medicinal Chemistry Letters, 27(4), 944, **2017**.

Protein level

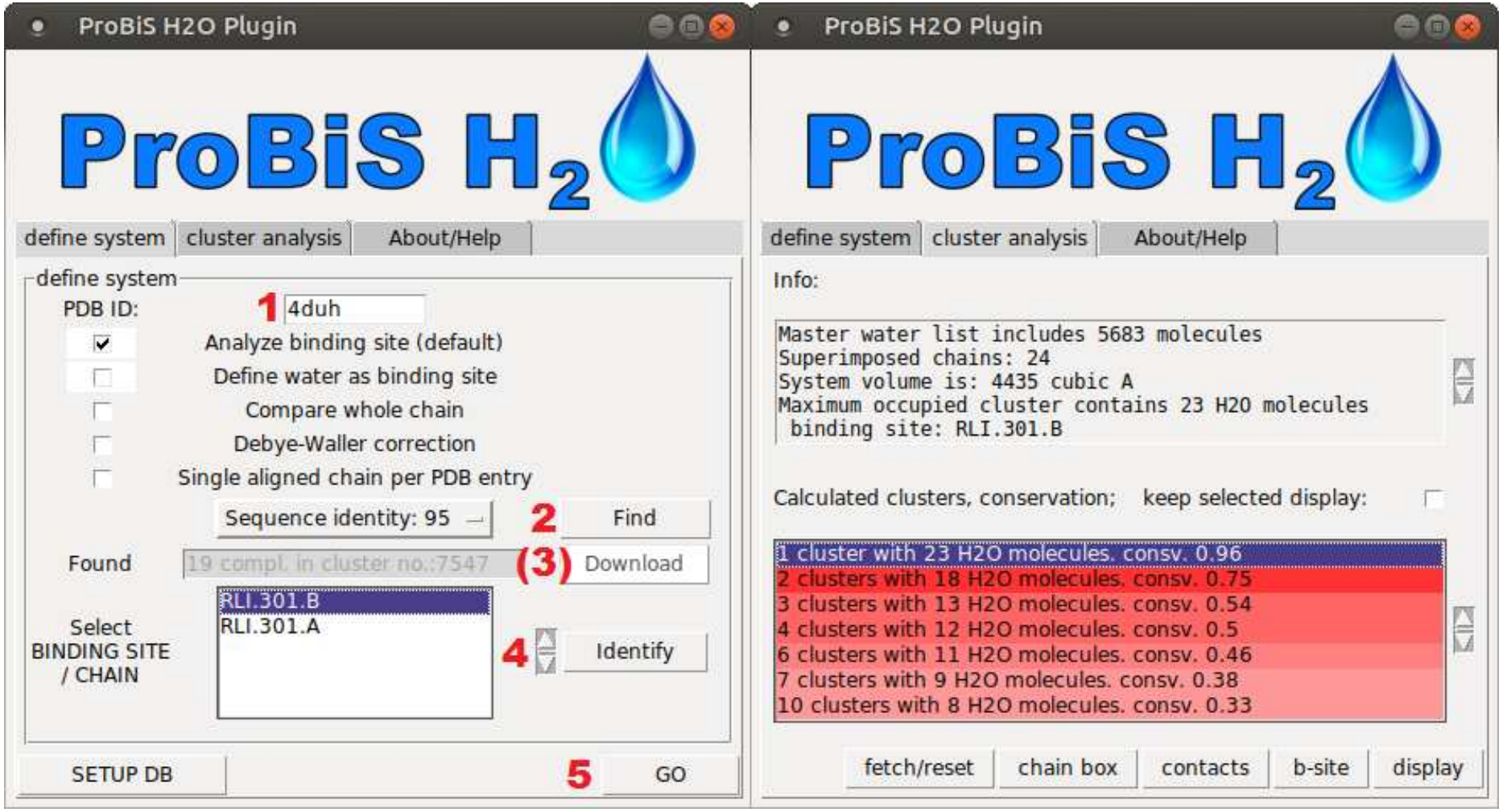
ProBiS H2O: Identification of conserved water sites in proteins



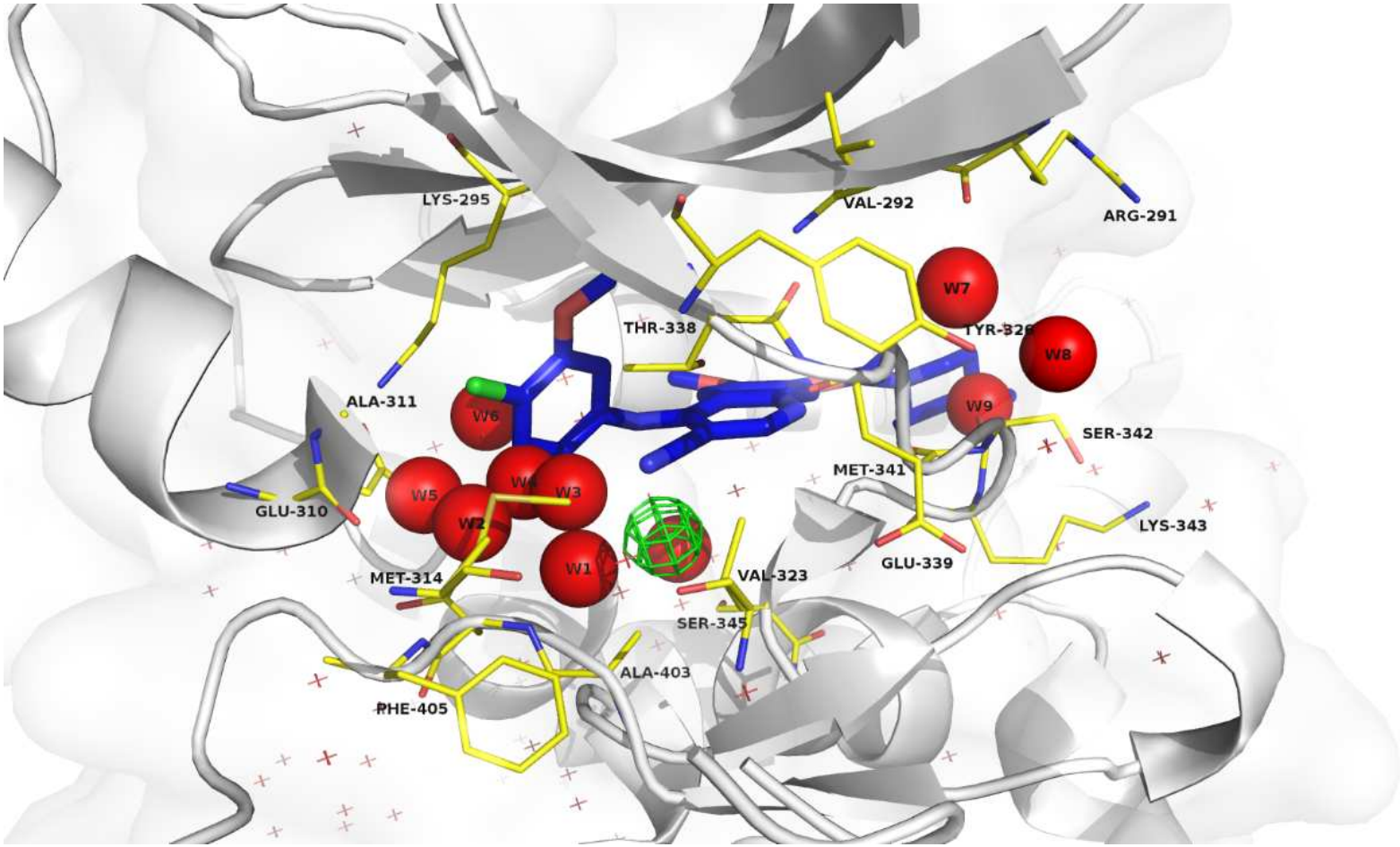
J. Chem. Info. Model., 57(12), 3094-3103, **2017**.

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

Input: PDB ID / Output: identified water clusters



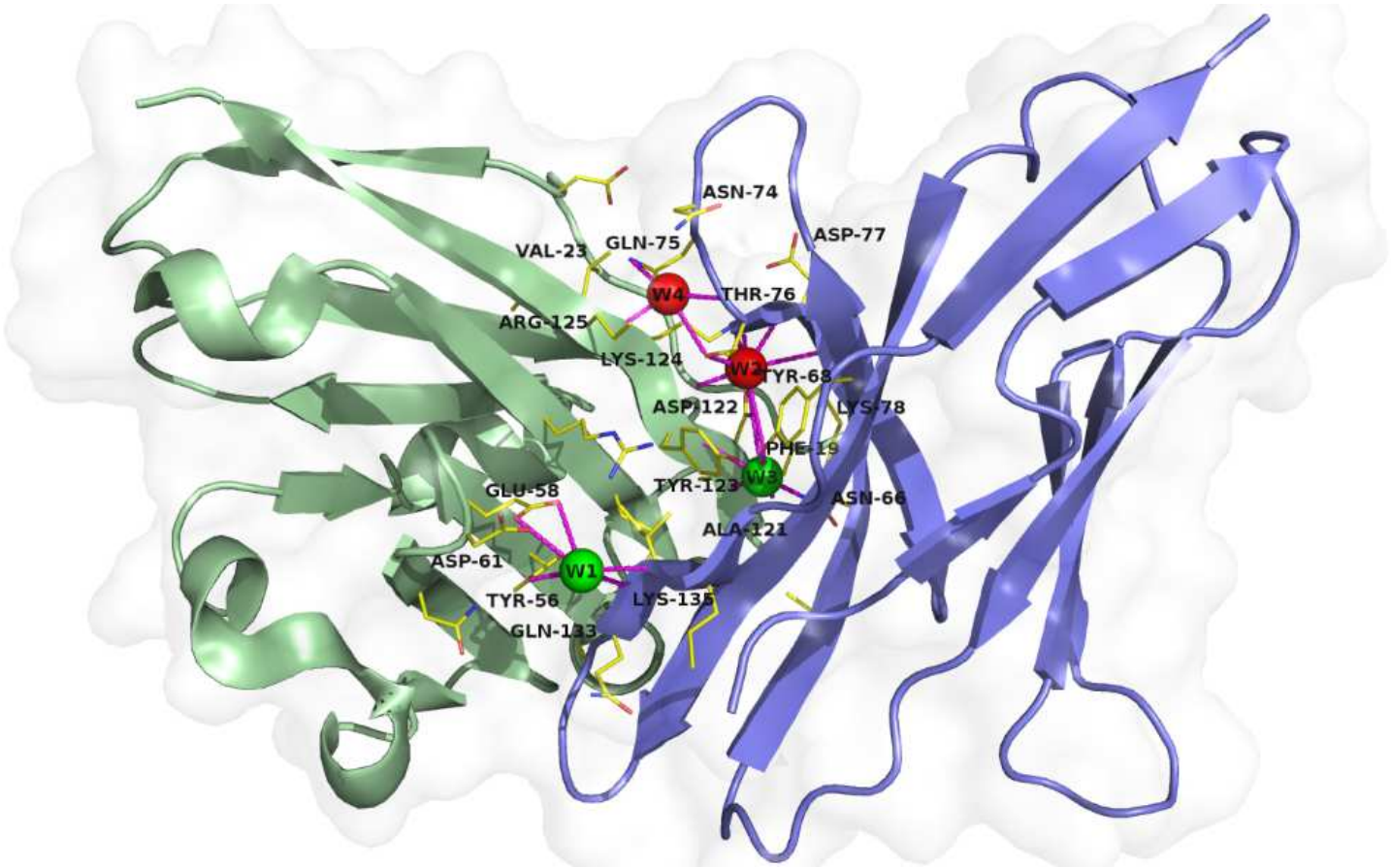
ProBiS H2O Validation: Anticancer Drug Selectivity



Bosutinib selectivity towards Src kinase mediated by conserved waters.

Known conserved water clusters W1-W2, newly found W3-W6 and W7-W9.

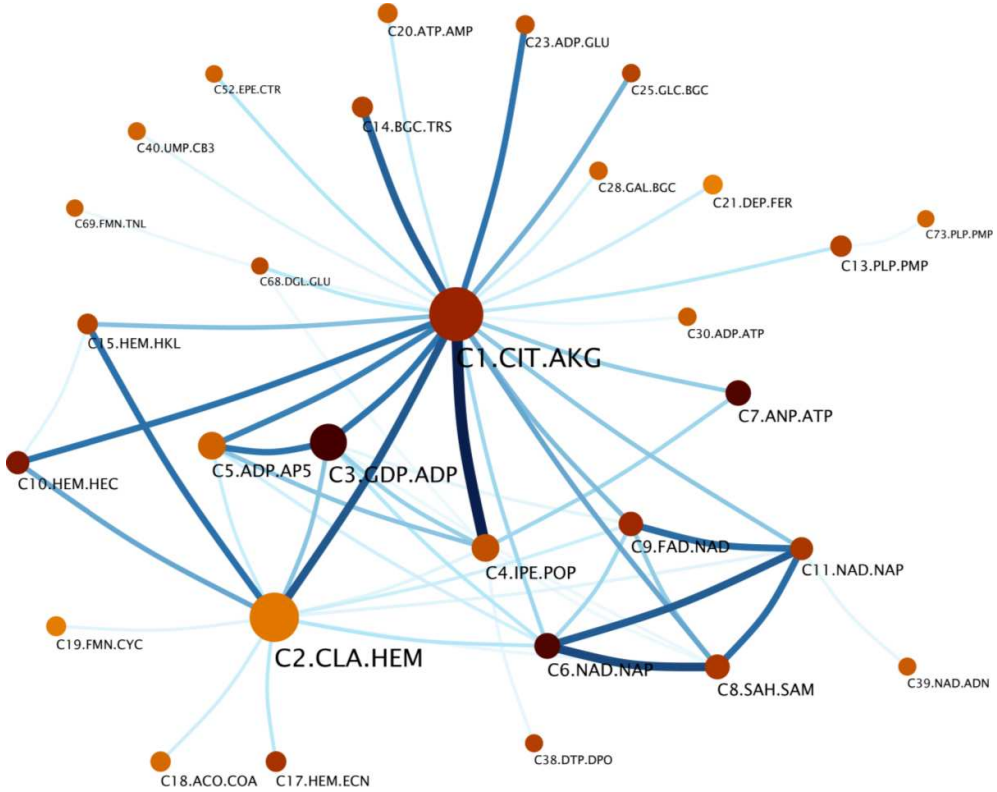
ProBiS H2O Validation: Conserved Waters at Protein-Protein Interface



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 scafolds of more recent ones. *SCIENTIFIC REPORTS*, *7*,*11652*, **2017**.

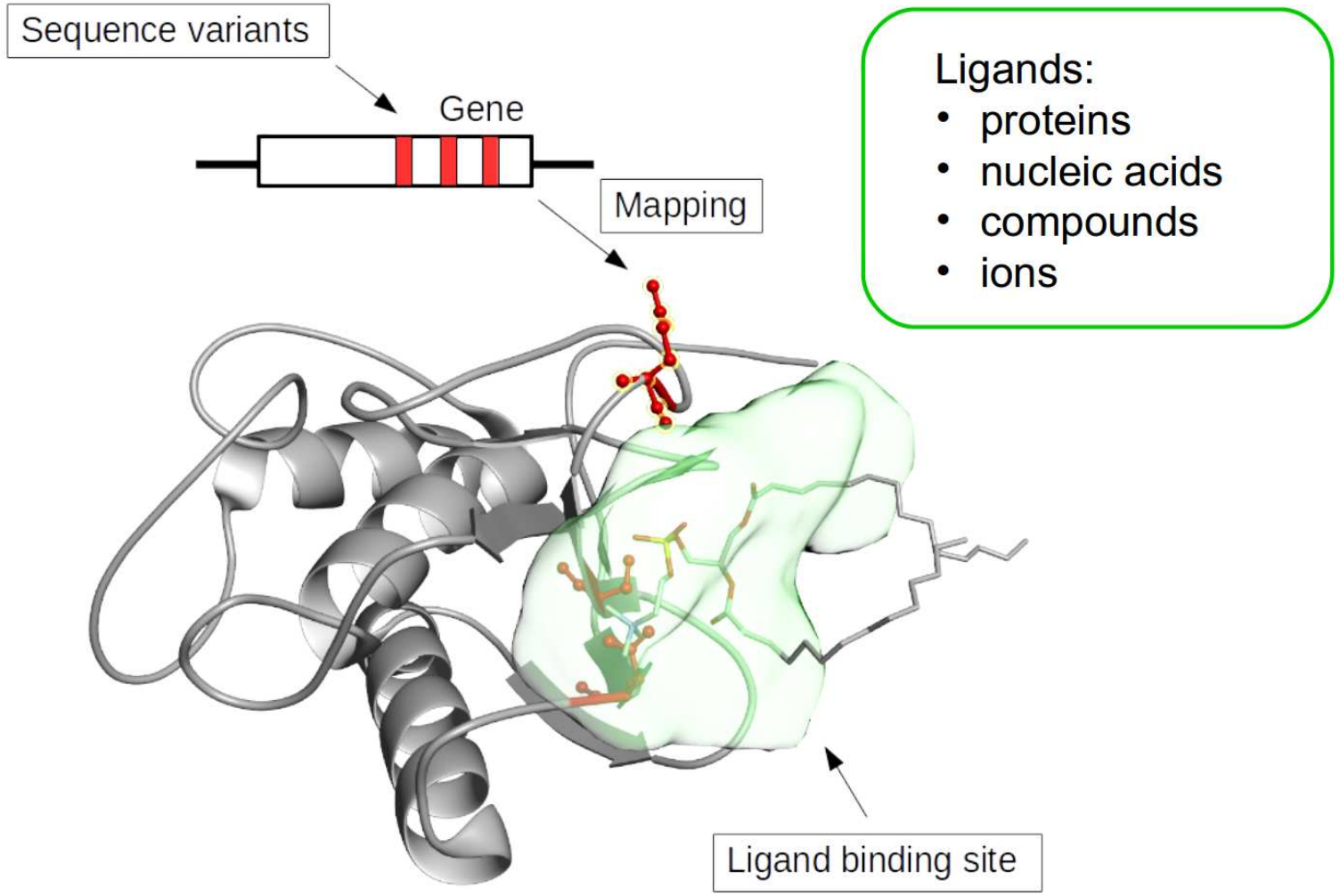
Gene level

GenProBiS: Sequence variants in protein binding sites

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

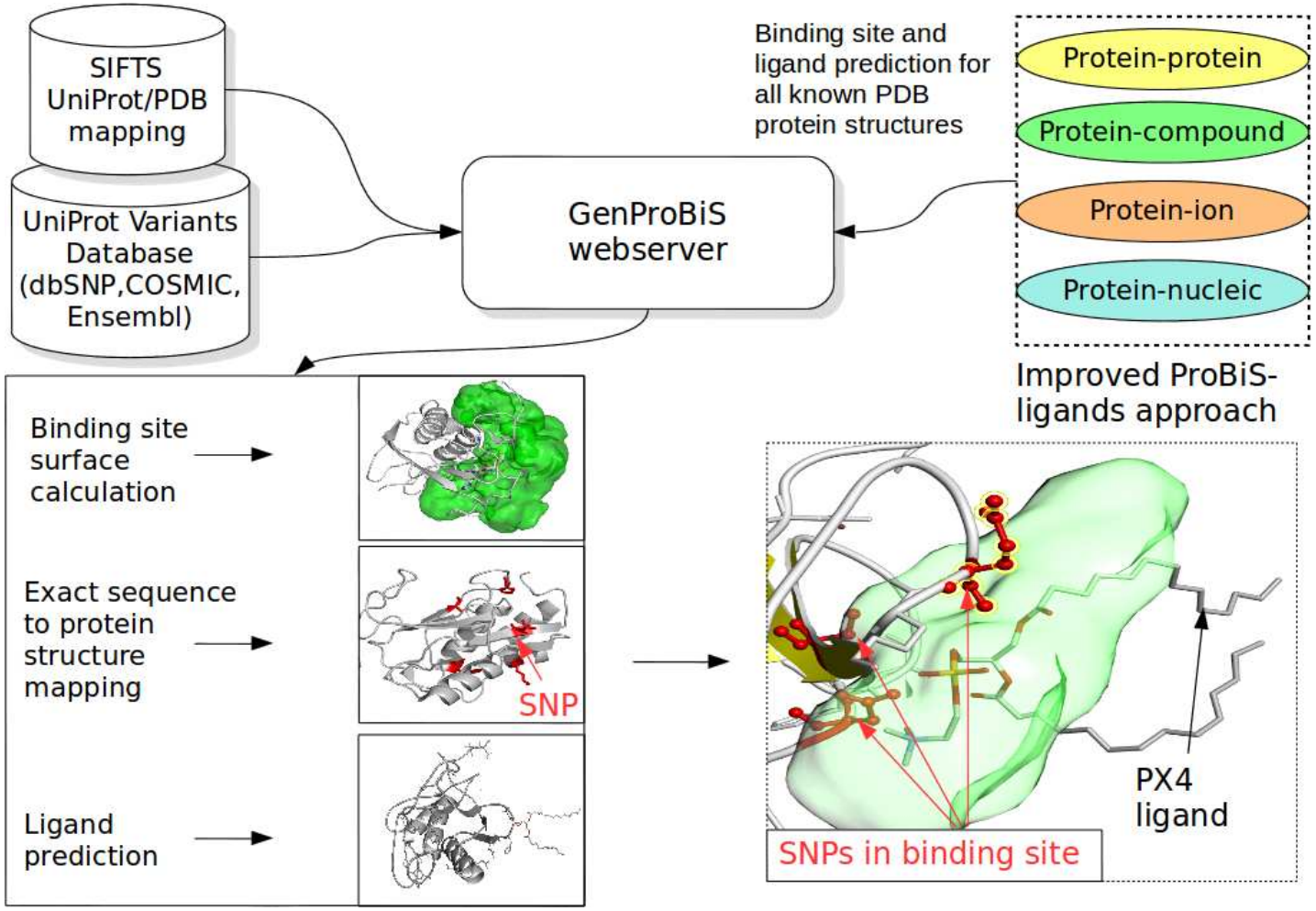
- GenProBiS web server maps sequence variants to protein structures from the PDB, and further to protein-protein, protein-nucleic acid, protein-compound, and protein-metal ion binding sites.
- The concept of a protein-compound binding site is understood in the broadest sense, includes glycosylation and 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



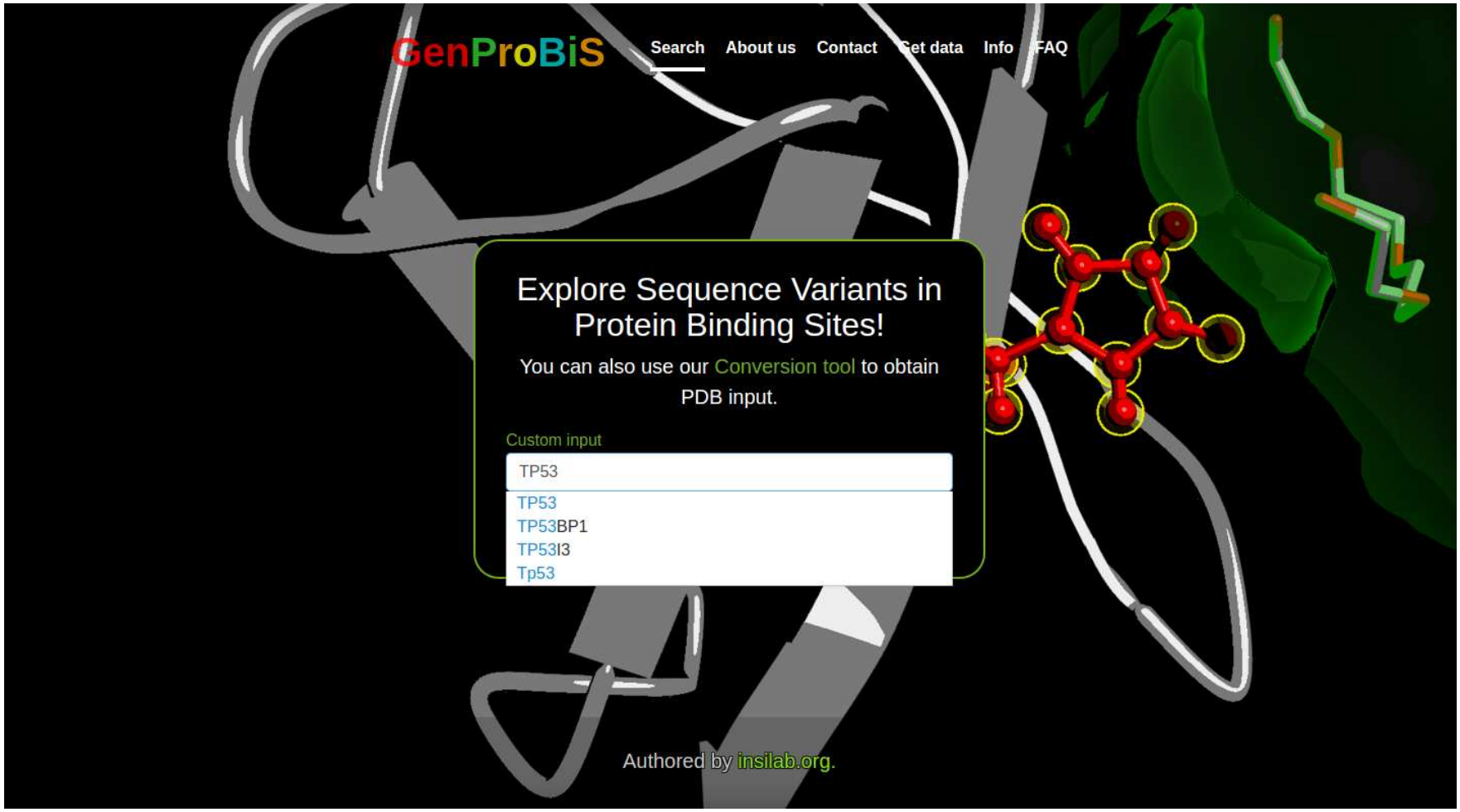
Mapping sequence variants to protein binding site

GenProBiS: Sequence variants in protein binding sites



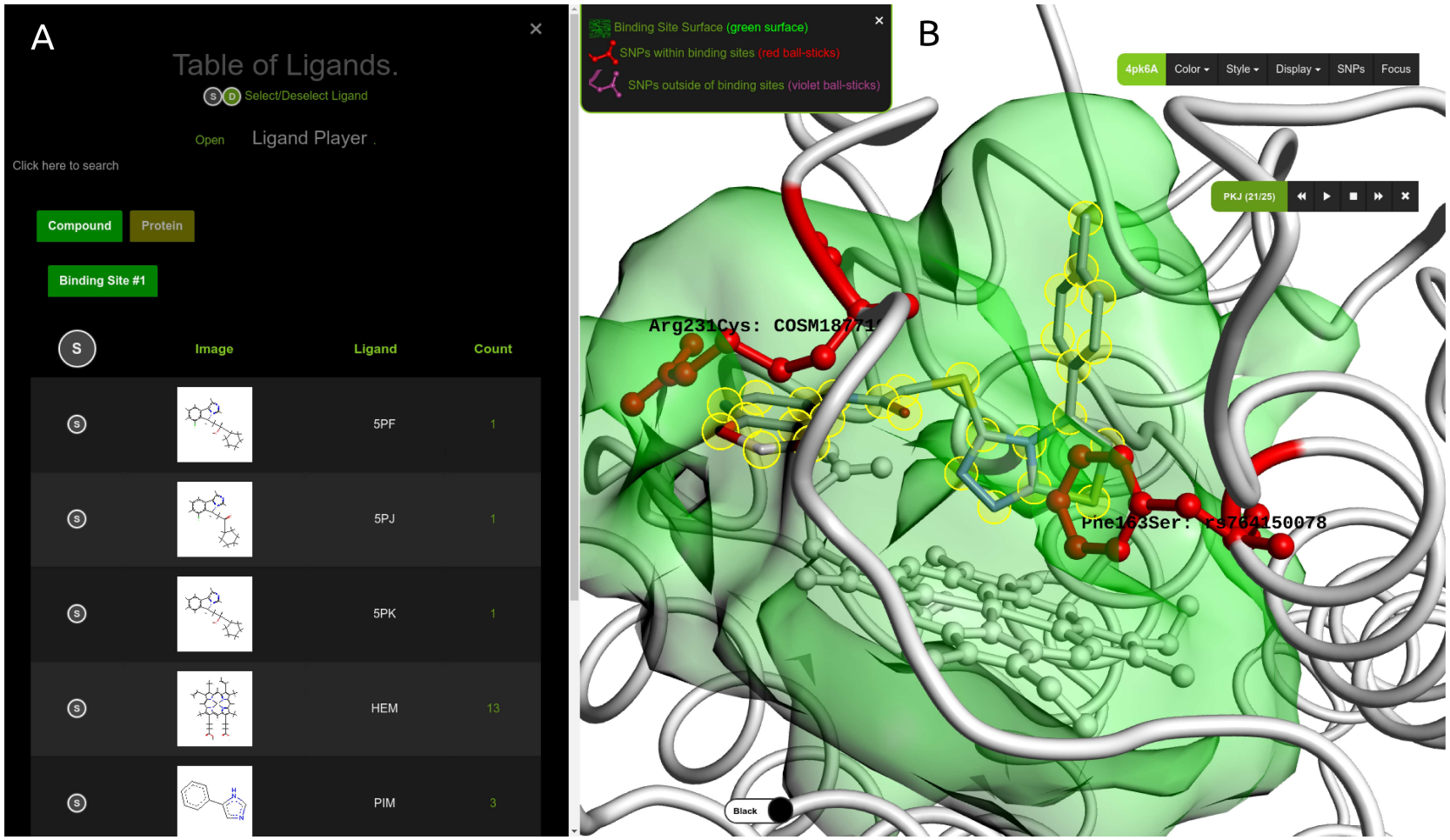
GenProBiS web server approach depicted on example of nsSNPs mapping to a compound (low molecular weight ligand) binding site.

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.

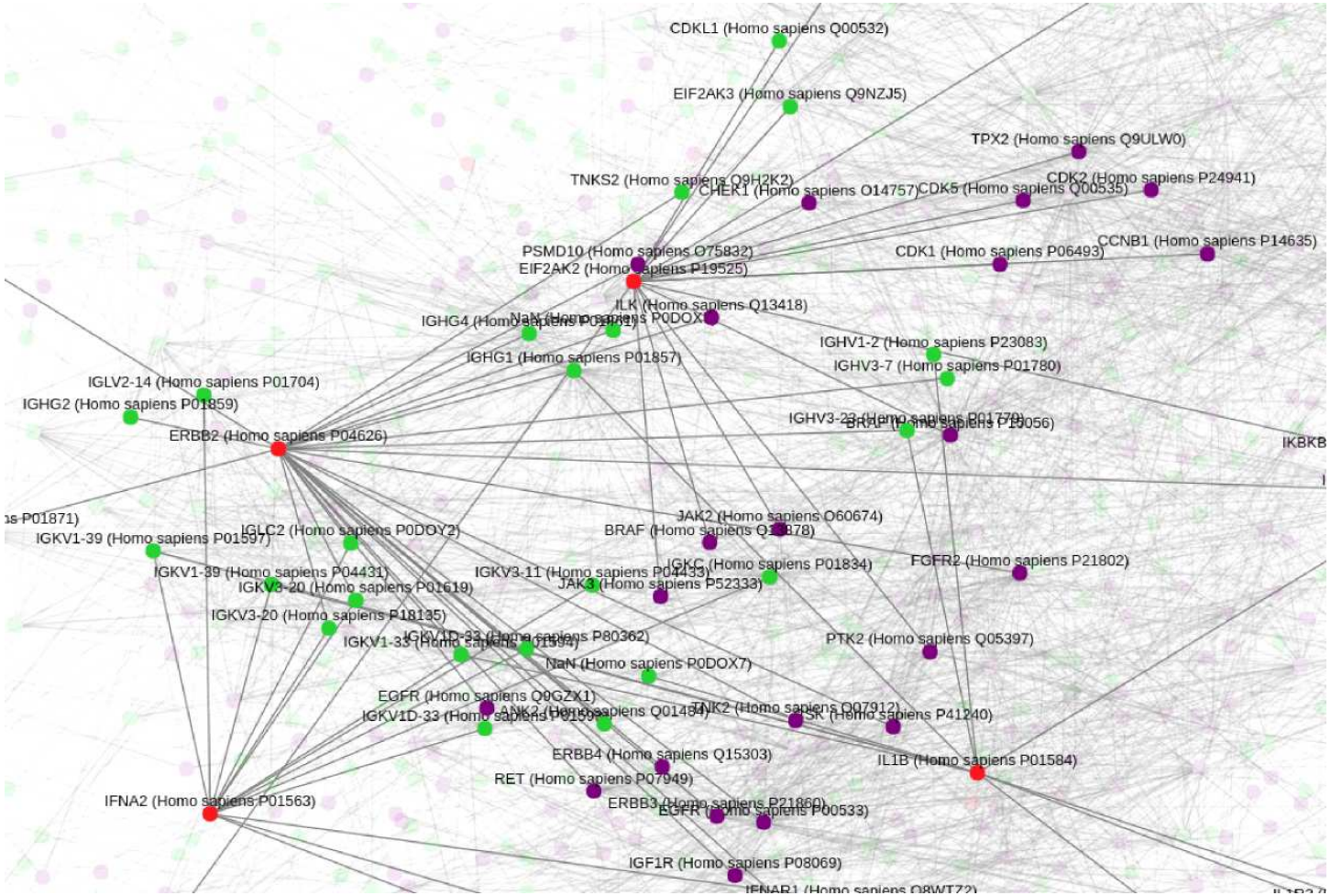
System level

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