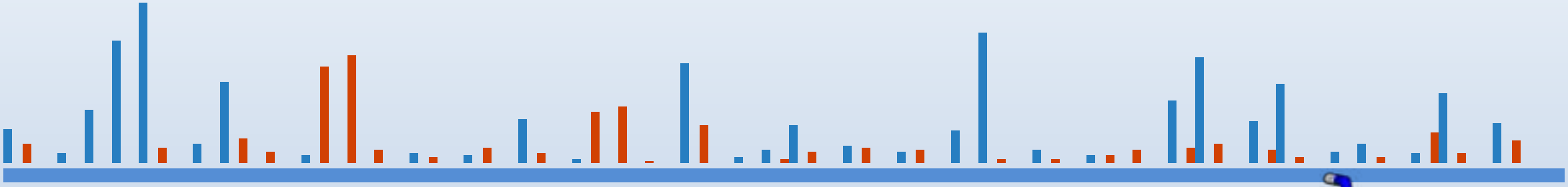
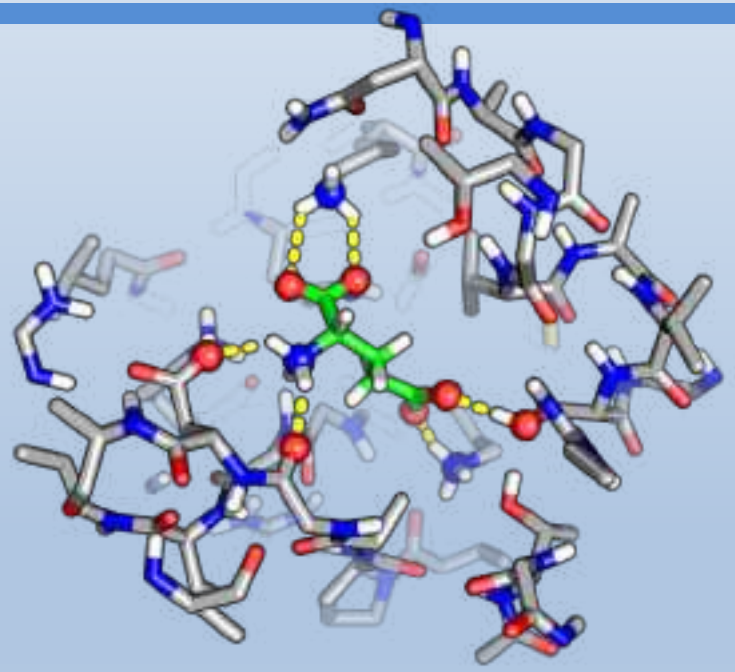


Uncovering Hidden Patterns of Molecular Recognition



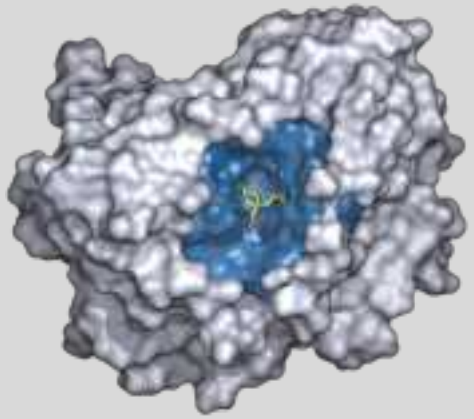
Sebastian Raschka

December 13, 2017

Biochemistry & Molecular Biology and Quantitative Biology

1

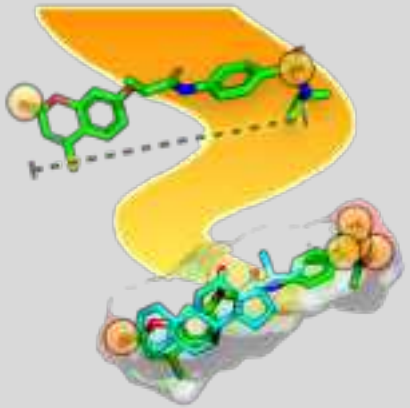
SiteInterlock



Raschka, Bemister-Buffington & Kuhn (2016) *Detecting the native ligand orientation by interfacial rigidity: SiteInterlock.* Proteins Struct Funct Bioinf 84:1888–1901.

2

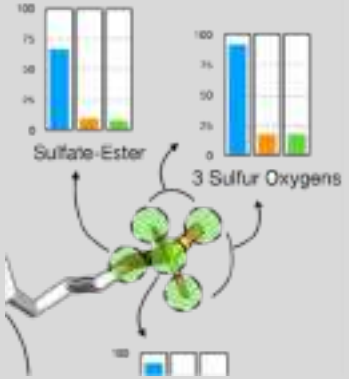
Screenlamp



Raschka, Scott, Liu, Gunturu, Huertas, Li & Kuhn (2017) *Enabling the hypothesis driven prioritization of ligand candidates in big databases: Screenlamp and its application to GPCR inhibitor discovery*. (In revision.)

3

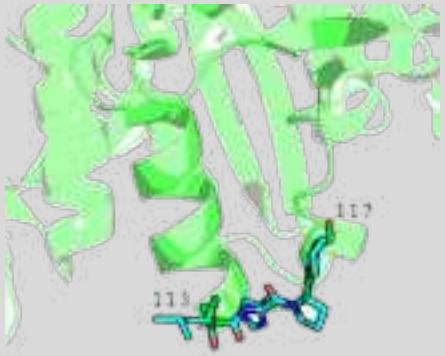
Machine Learning & Chemical Groups



Raschka, Kuhn, Scott, Huertas & Li (2017) *Computational Drug Discovery and Design: Automated inference of chemical group discriminants of biological activity from virtual screening data.* Springer. (In press.)

4

3D Epitope-Based Virtual Screening

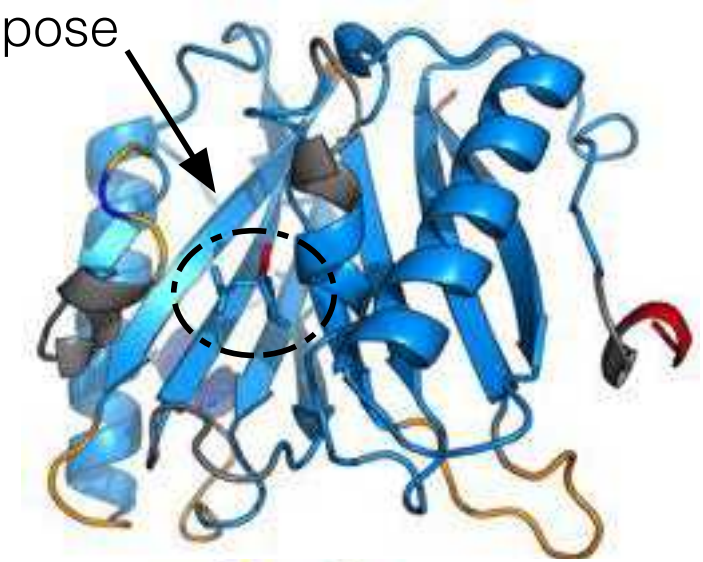


Raschka, Zeng, Basson & Kuhn (2015-present)

1

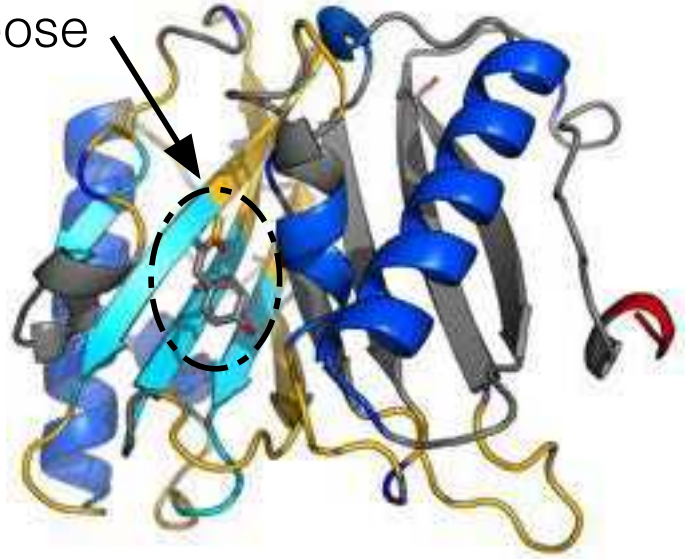
SiteInterlock

Near-native docking pose



Raschka S, Bemister-Buffington J, Kuhn LA (2016)
Detecting the native ligand orientation by interfacial rigidity: SiteInterlock.
Proteins: Structure, Function, and Bioinformatics 84:1888–1901

“Bad” docking pose



- Novel insights: Binding site rigidification is a signature of native protein-ligand complex formation
- Captures the coupling of intermolecular interactions
- Competitive to state-of-the-art scoring functions for pose prediction; robust (no “very bad” predictions); new information (coupling)

🔗 https://psa-lab.github.io/siteinterlock/

screenlamp

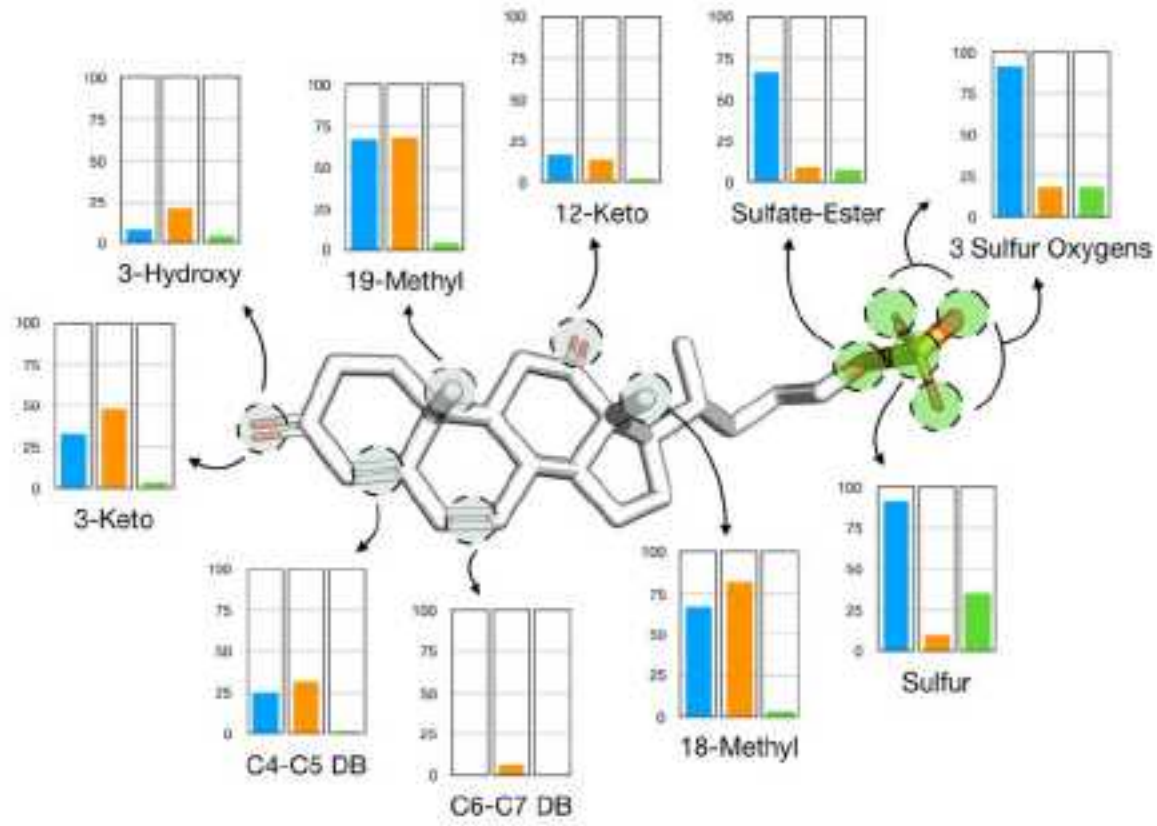
Raschka, Scott, Liu, Gunturu, Huertas, Li & Kuhn (2017)
Enabling the hypothesis driven prioritization of ligand candidates in big databases: Screenlamp and its application to GPCR inhibitor discovery.
(In revision.)

- Discovery of a pheromone antagonist that nullifies the GPCR-mediated signaling response in sea lamprey
- Hypothesis-based virtual screening toolkit for millions of molecules
- Pioneering aquative invasive species control: Antagonists currently tested in streams

https://psa-lab.github.io/screenlamp/

Machine Learning & Chemical Groups

Raschka, Kuhn, Scott, Huertas & Li (2017)
Computational Drug Discovery and Design: Automated inference of chemical group discriminants of biological activity from virtual screening data.
Springer, 2017. (In press.)

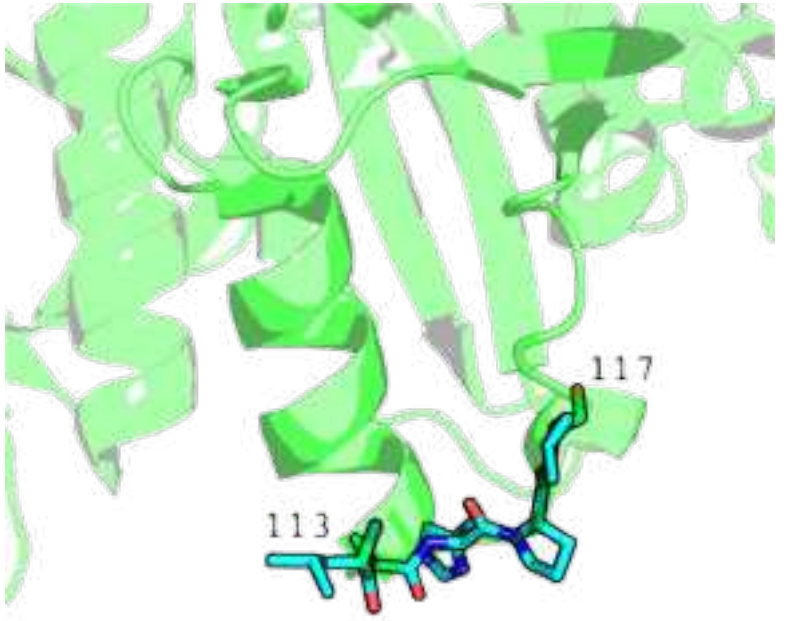


- Identification of chemical groups in pheromone inhibitors that are important for activity
- New knowledge to formulate new screening hypotheses and enable ligand design
- Protocols to determine important chemical groups in other small molecule activity datasets

🔗 https://github.com/psa-lab/predicting-activity-by-machine-learning

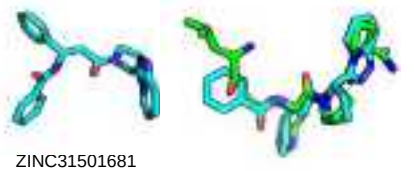
3D Epitope-Based Virtual Screening

- Discovery of small molecules that can block the interaction between two protein kinases involved in cancer metastasis
- Novel protocol for blocking protein-protein interactions using 3D ligand-based virtual screening to mimic a protein epitope (does not require structure of the binding partner)
- Inhibitor candidates from screening >10 million commercially available small molecules currently being tested experimentally (Basson Lab)



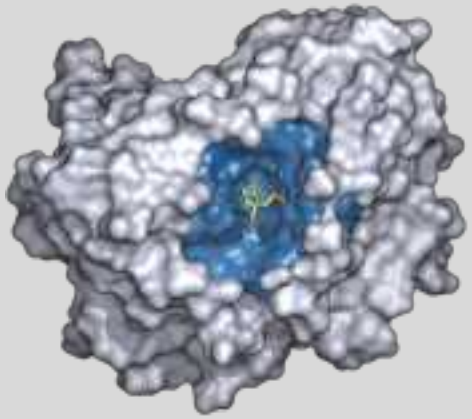

ZINC25757351


ZINC13002691


ZINC31501681

1

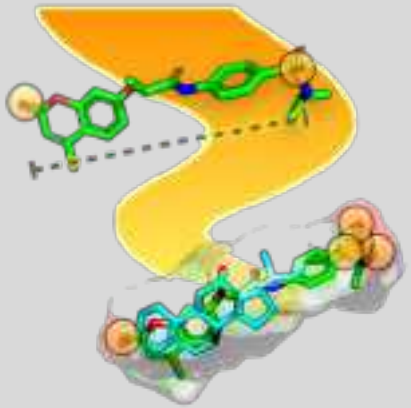
SiteInterlock



Raschka, Bemister-Buffington & Kuhn (2016) *Detecting the native ligand orientation by interfacial rigidity: SiteInterlock.* Proteins Struct Funct Bioinf 84:1888–1901.

2

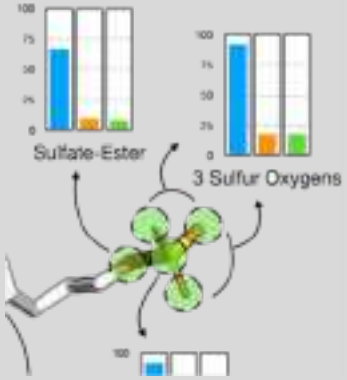
Screenlamp



Raschka, Scott, Liu, Gunturu, Huertas, Li & Kuhn (2017) *Enabling the hypothesis driven prioritization of ligand candidates in big databases: Screenlamp and its application to GPCR inhibitor discovery*. (In revision.)

3

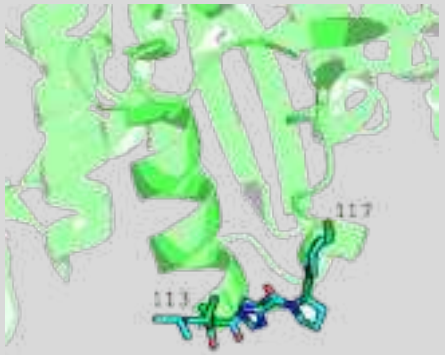
Machine Learning & Chemical Groups



Raschka, Kuhn, Scott, Huertas & Li (2017) *Computational Drug Discovery and Design: Automated inference of chemical group discriminants of biological activity from virtual screening data.* Springer. (In press.)

4

3D Epitope-Based Virtual Screening



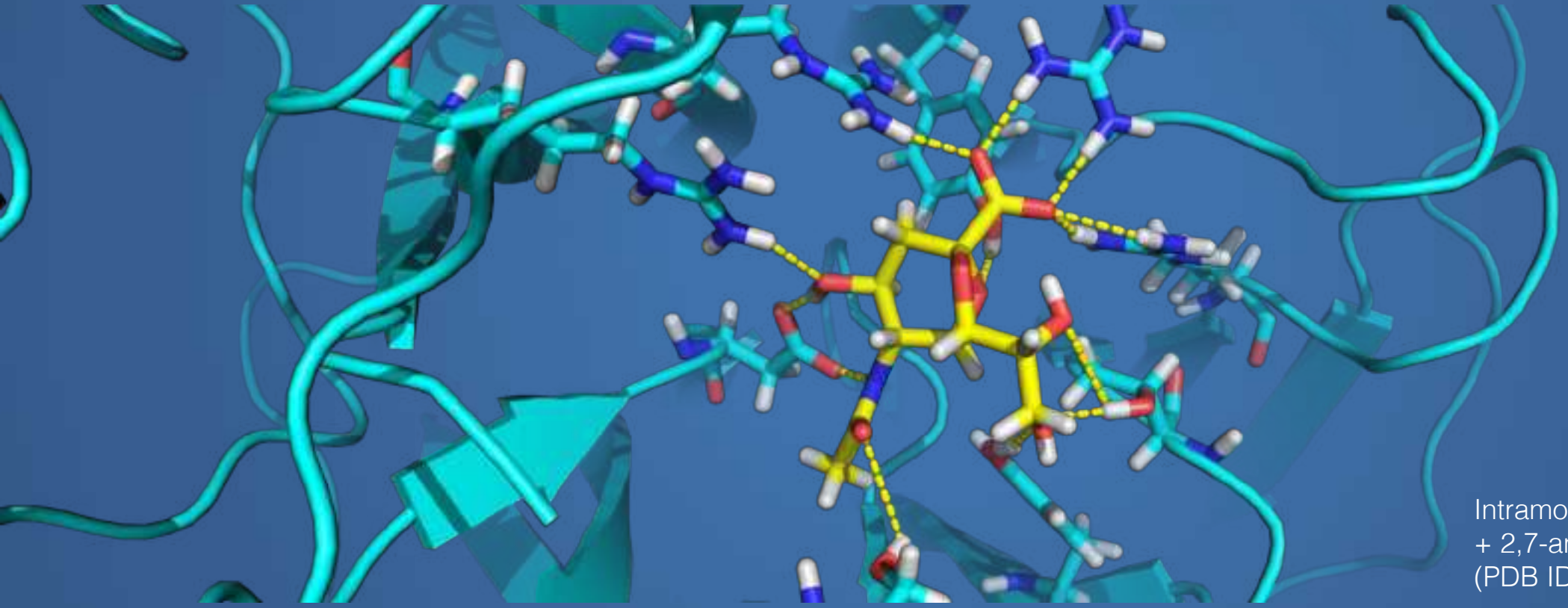
Raschka, Zeng, Basson & Kuhn (2015-present)

5

Protein Recognition Index

Raschka, Wolf, Bemister-Buffington & Kuhn (2017) *Protein-ligand interfaces are polarized: Discovery of a strong trend for intermolecular hydrogen bonds to favor donors on the protein side with implications for predicting and designing ligand complexes.* (Submitted.)

Intermolecular Hydrogen-Bonding Patterns



Intramolecular trans-sialidase
+ 2,7-anhydro-Neu5Ac
(PDB ID: 2sli)

Noted in our previous projects:

1. Protein amine groups frequently H-bond to ligands
2. Hydroxyl groups on small molecules lead to false positives in ligand discovery

Are these general trends?

Methods for analyzing intermolecular hydrogen bond networks

Workflow

Collect dataset of non-homologous proteins in complex with diverse, biological small-molecule ligands

↓

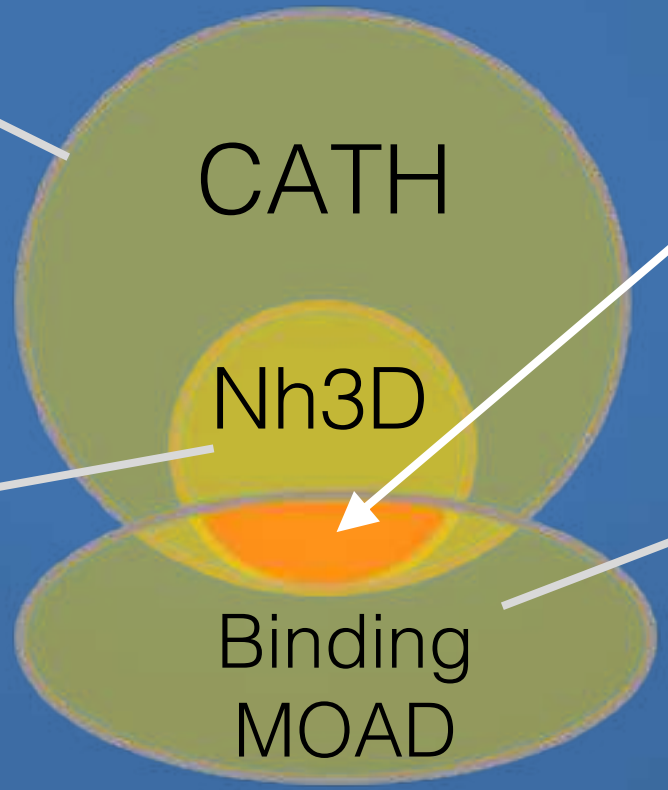
Assign proper protonation states in proteins and ligands (addition and orientation of hydrogen atoms)

↓

Assign and analyze intermolecular hydrogen-bond network

Dataset

CATH database:
Class, Architecture, Topology/fold, Homologous superfamily (http://www.cathdb.info)



136 non-homologous proteins in complex with diverse, biological small-molecule ligands

Non-homologous protein domains based on CATH (Thiruv et al. BMC Structural Biology 2005, 5:12)

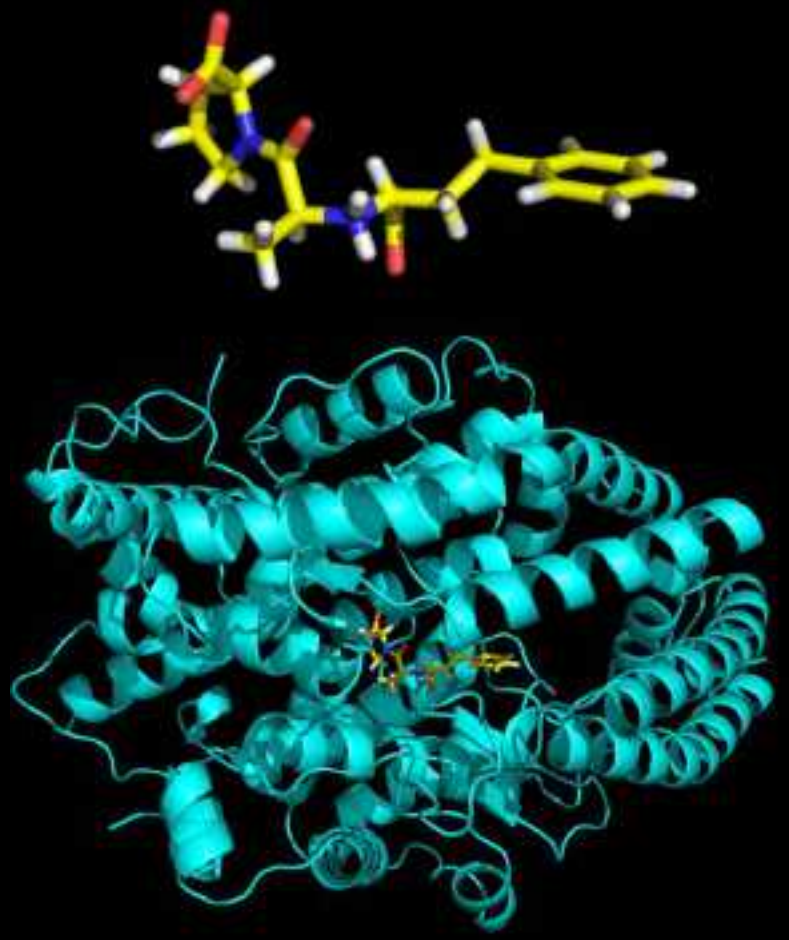
Well-resolved protein structures with biological ligands and experimental binding data (http://bindingmoad.org)

136 Protein-ligand complexes
Non-homologous structures, diverse biological ligands

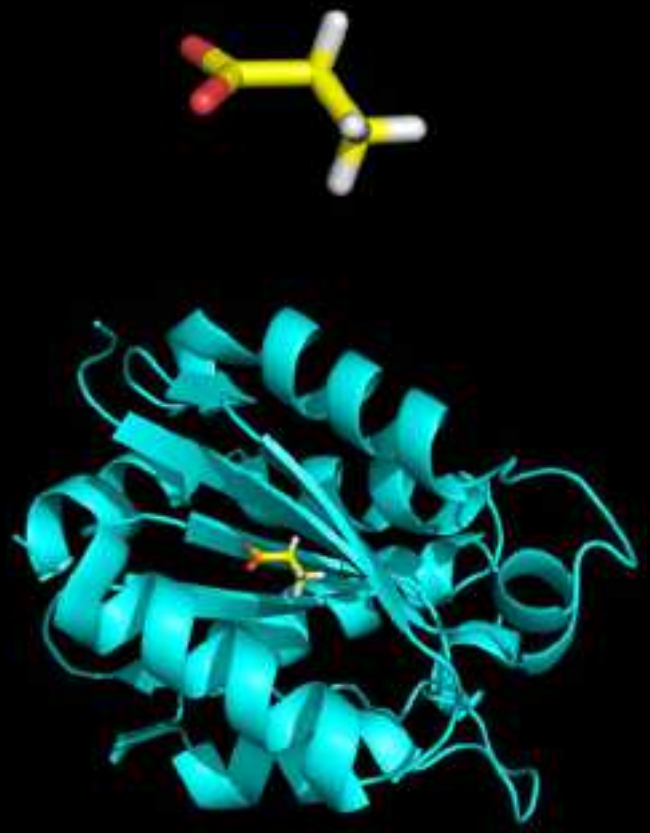
PDB code	Protein description	Ligand code	Ligand category	Lig. chain ID and res. #	Resolution (Å)	R-value work	R-value free
1a9x	Carbamoyl phosphate synthetase	ORN	Peptide-like	A1920	1.8	0.19	-
1af7	Chemotaxis receptor methyltransferase	SAH	Nucleotide-like	A287	2.0	0.20	0.28
1amu	Gramidicin synthetase	PHE	Peptide-like	A566	1.9	0.21	0.25
1awq	Cyclophilin A	Multiple	Peptide-like	B1	1.6	0.34	0.43
1ayl	Phosphoenolpyruvate carboxykinase	OXL	Other	A542	1.8	0.20	0.23
1b4u	Dioxygenase	DHB	Other	D504	2.2	0.16	0.22
1b5e	Deoxycytidylate hydroxymethylase	DCM	Nucleotide-like	B400	1.6	0.19	0.21
1b37	Polyamine oxidase	FAD	Nucleotide-like	A800	1.9	0.20	0.23

. . .

Sulfite oxidase
+ phosphonic acid mono-(2-amino-5,6-dimercapto-4-oxo-3,7,8A,9,10,10A-hexahydro-4H-8-oxa-1,3,9,10-tetraaza-anthracen-7-ylmethyl)ester
(PDB ID: 1sox)



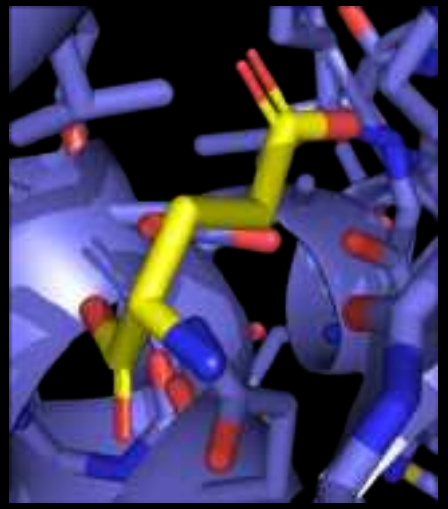
Angiotensin converting enzyme
+ 1-((2s)-2-{[(1s)-1-carboxy-3-phenylpropyl]amino}propanoyl)-L-proline
(PDB ID: 1uze)



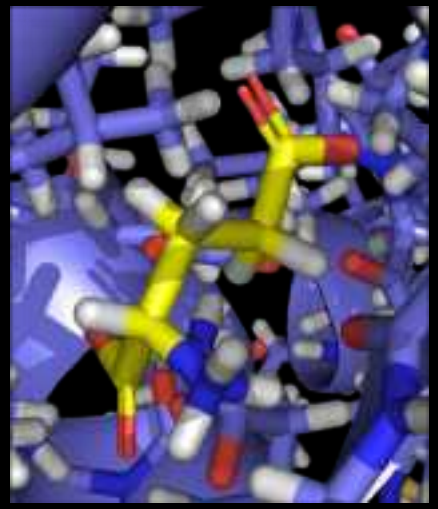
Molybdopterin-bound Cnx1G domain
+ propanoic acid
(PDB ID: 1uuy)

Protonation State Assignment

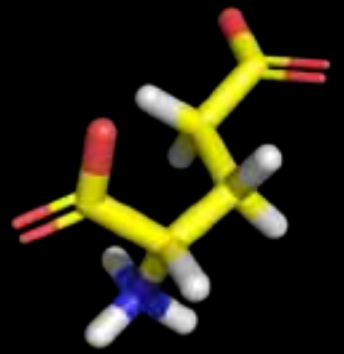
Glutamate dehydrogenase
+ glutamic acid
(PDB ID: 1bgv)



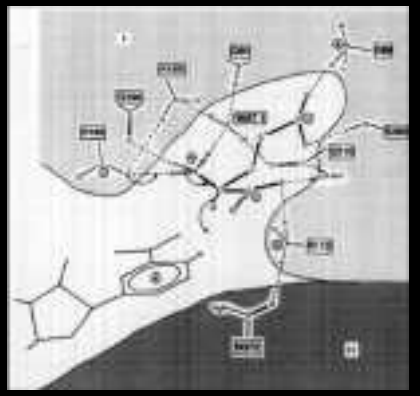
Obtain structure from PDB



Protonate complex with Yasara OptHyd
+ YAMBER force field



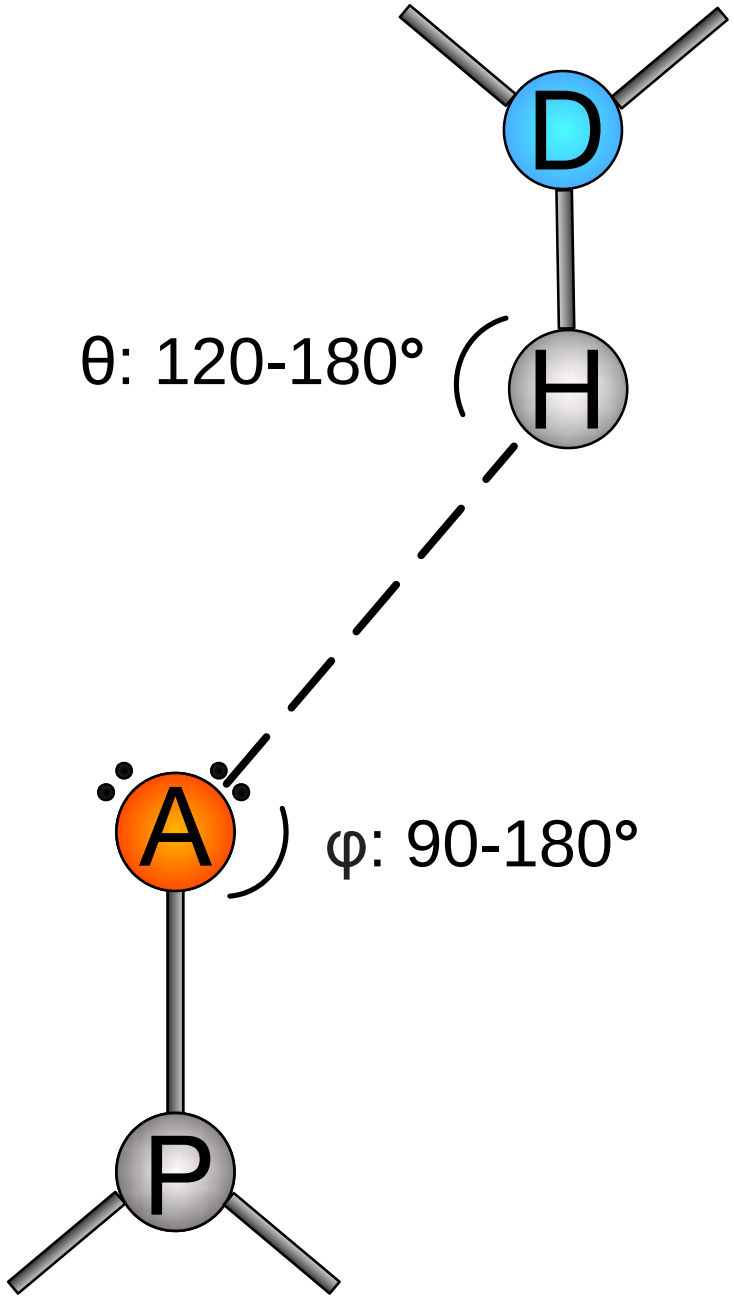
Compare with quantum mechanical computation (OpenEye MolCharge
+ AM1-BCC force field)



Compare with protonation state def. by chemical experts in literature

Hydrogen bond criteria

- Acceptor (**A**)—Donor (**D**) distance: 2.4-3.5 Å
- Acceptor (**A**)—Hydrogen (**H**) distance: 1.5-2.5 Å



Rules based on:

- Ippolito et al 1990. Journal of molecular biology, 215(3), 457-471.
- McDonald, Ian & Janet M Thornton 1994. http://www.biochem.ucl.ac.uk/bsm/atlas

Hbind

Open source, available via GitHub

```
--------------------------------------------------------------------------------
PDB code of protein-ligand complex 1r8s, chain ID: A, ligand res. num.: 401
--------------------------------------------------------------------------------
Hbind (version 1.0)
Protein Structural Analysis & Design Lab, MSU(kuhnlab@msu.edu)

MOL2 file: /home/raschkas/protonated_ligands/1r8s.mol2
PDB file: /home/raschkas/proteins/1r8s.pdb

++++++++++++ SlideScore Summary +++++++++++++++
|
| Protein-Ligand Hydrophobic Contacts :     33
| Protein-Ligand H-bonds              :     16
| Protein-Ligand Salt-bridges         :      4
| Metal-Ligand Bonds                  :      0
|
++++++++++++ Interaction Table ++++++++++++++++
#
#            | Ligand Atom -- Protein  Atom | Bond   D-H-A  Ligand-Protein
#            |  #  type    -- RES   #  type | Dist.  Angle  Interaction
| hbond     1  16  N.am    -- ASP  129  OD1   2.749  173.3  Donor - Acceptor
| hbond     2  18  N.pl3   -- ASP  129  OD2   2.917  165.1  Donor - Acceptor
| hbond     3  22  N.2     -- ASN  126  ND2   3.051  141.5  Acceptor - Donor
| hbond     4  25  O.3     -- LYS  127  NZ    3.221  149.0  Acceptor - Donor
| hbond     5  30  O.2     -- THR   32  N     2.846  150.8  Acceptor - Donor
| hbond     6  30  O.2     -- THR   32  OG1   2.686  178.9  Acceptor - Donor
| hbond     7  31  O.2     -- THR   31  N     2.927  159.3  Acceptor - Donor
| hbond     8  31  O.2     -- THR   31  OG1   2.735  177.4  Acceptor - Donor
| hbond     9  32  O.3     -- LYS  156  NZ    2.757  173.5  Acceptor - Donor
| hbond    10  33  O.3     -- GLY   29  N     3.010  159.5  Acceptor - Donor
| hbond    11  33  O.3     -- LYS   30  N     2.911  160.2  Acceptor - Donor
| hbond    12  33  O.3     -- LYS   30  NZ    2.868  177.9  Acceptor - Donor
| hbond    13  34  O.3     -- GLY   29  N     3.204  123.5  Acceptor - Donor
| hbond    14  39  O.3     -- ALA   27  N     2.850  155.6  Acceptor - Donor
| hbond    15  40  O.2     -- LYS  127  N     3.268  120.7  Acceptor - Donor
| hbond    16  40  O.2     -- ALA  160  N     2.996  131.1  Acceptor - Donor

================================================================================
```

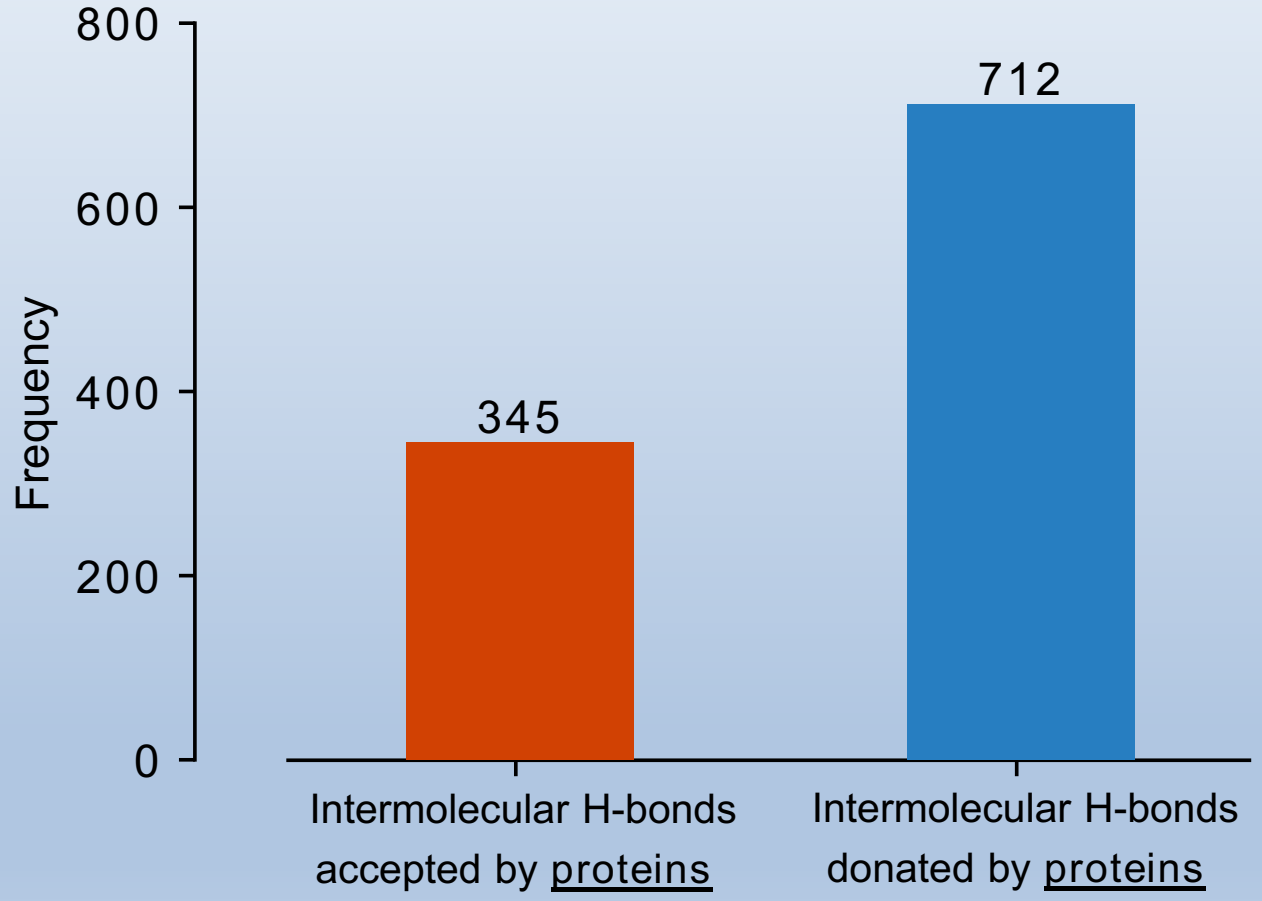
Workflow

Collect dataset of non-homologous proteins in complex with diverse, biological small-molecule ligands

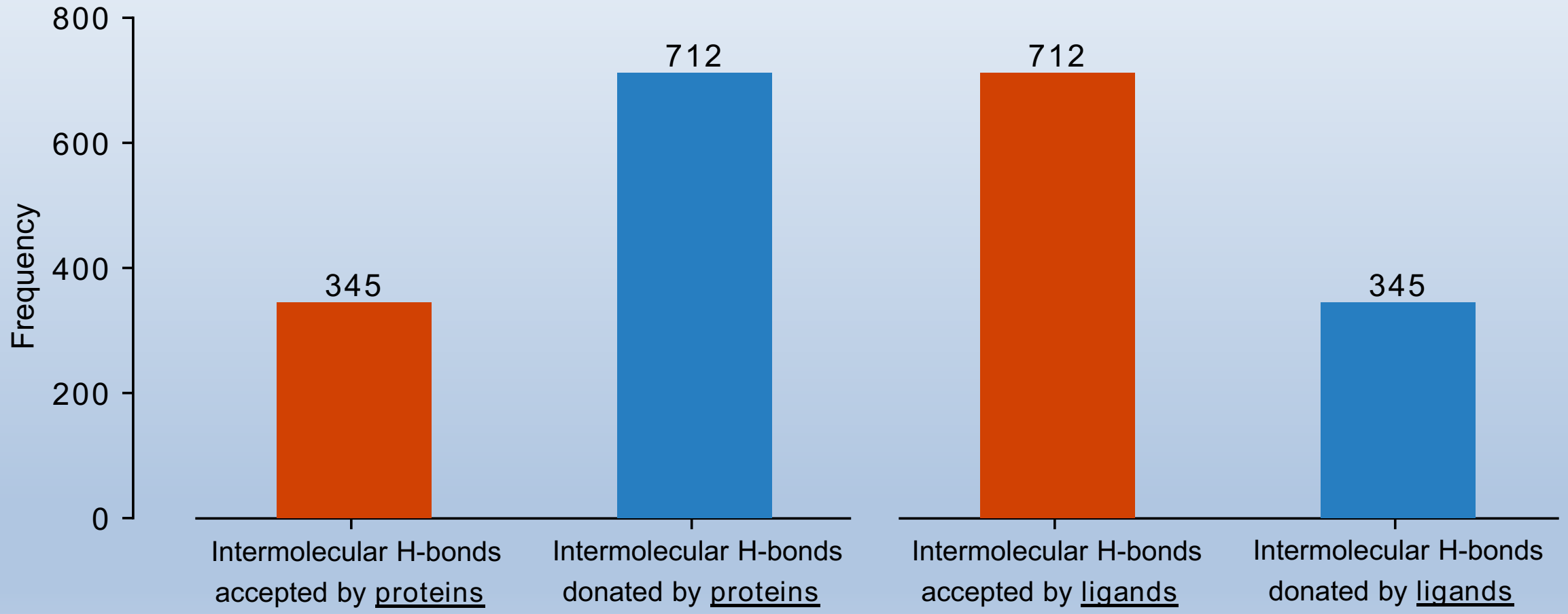
Assign proper protonation states in proteins and ligands (addition and orientation of hydrogen atoms)

Assign and analyze intermolecular hydrogen-bond network

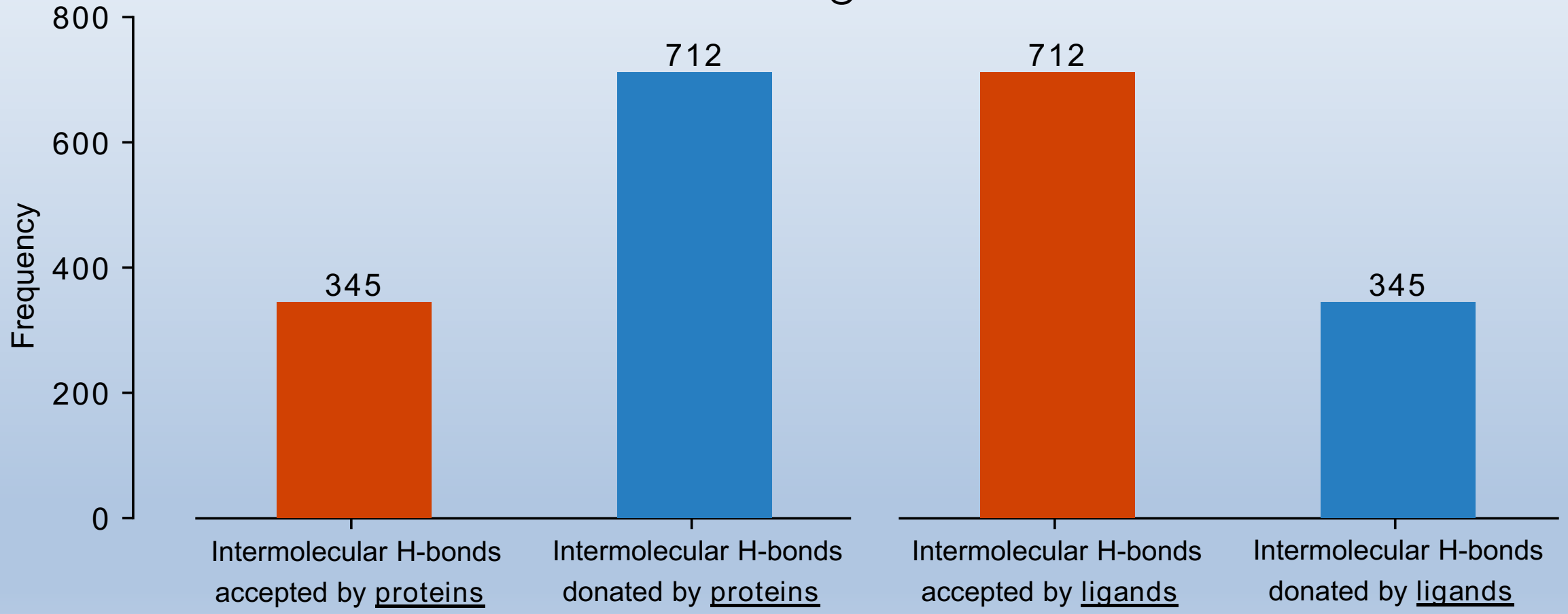
Protein-ligand interfaces are polarized



Proteins donate 2 times as many H-bonds as they accept

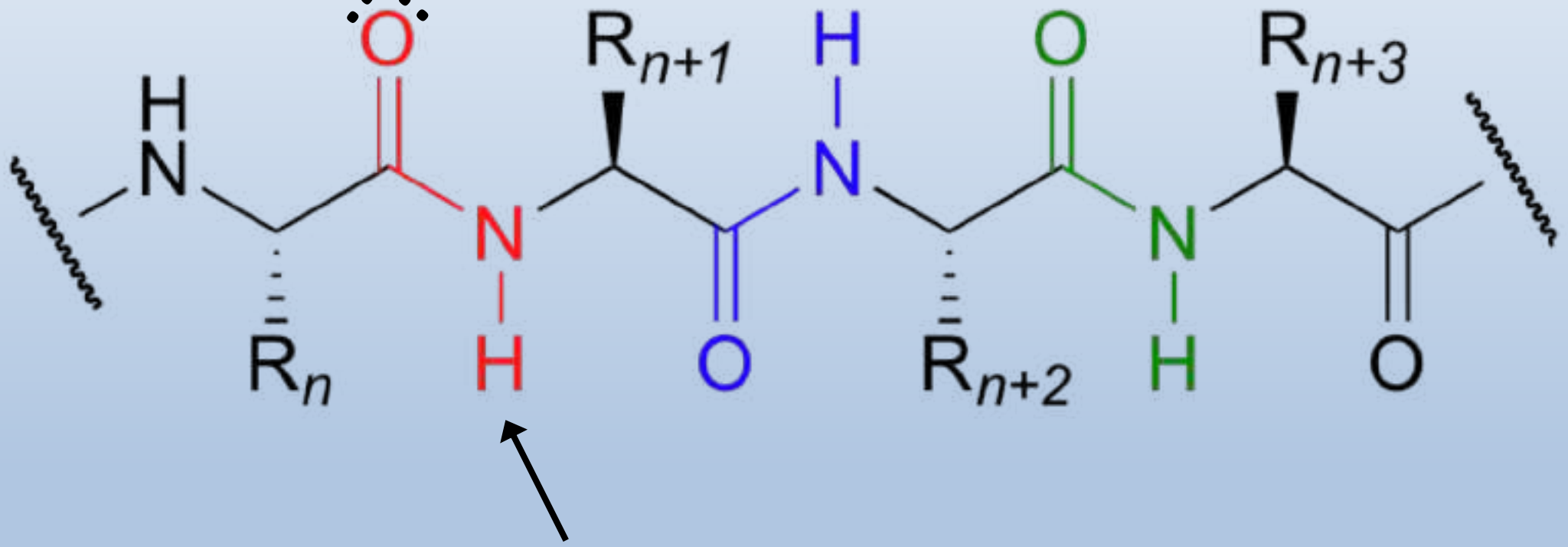


Trend due to high proton to electron lone pair ratio in binding sites?



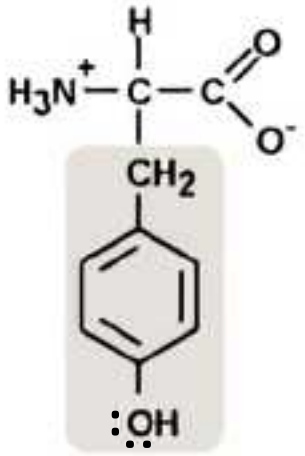
Main Chain

2 hydrogen bond acceptor lone pairs



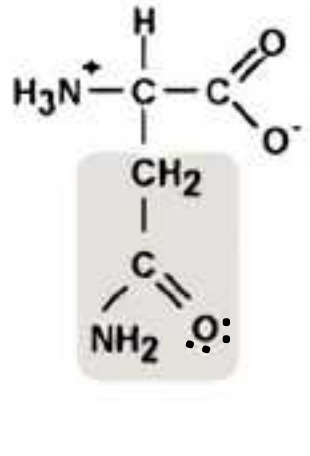
1 hydrogen donor atom

Protons and electron lone pairs on amino acid side chains



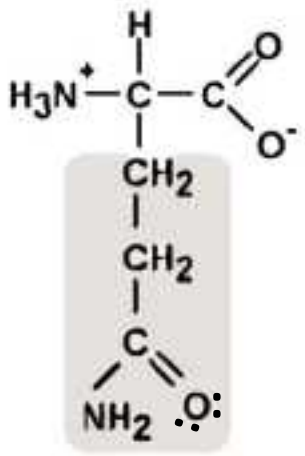
Tyrosine (Tyr)

1 H & 2 LP



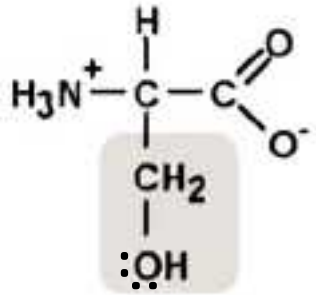
Asparagine (Asn)

2 H & 2 LP



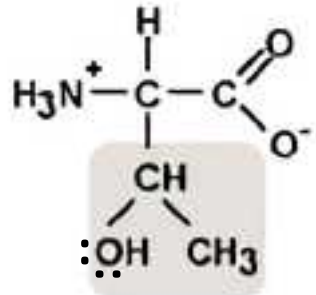
Glutamine (Gln)

2 H & 2 LP



Serine (Ser)

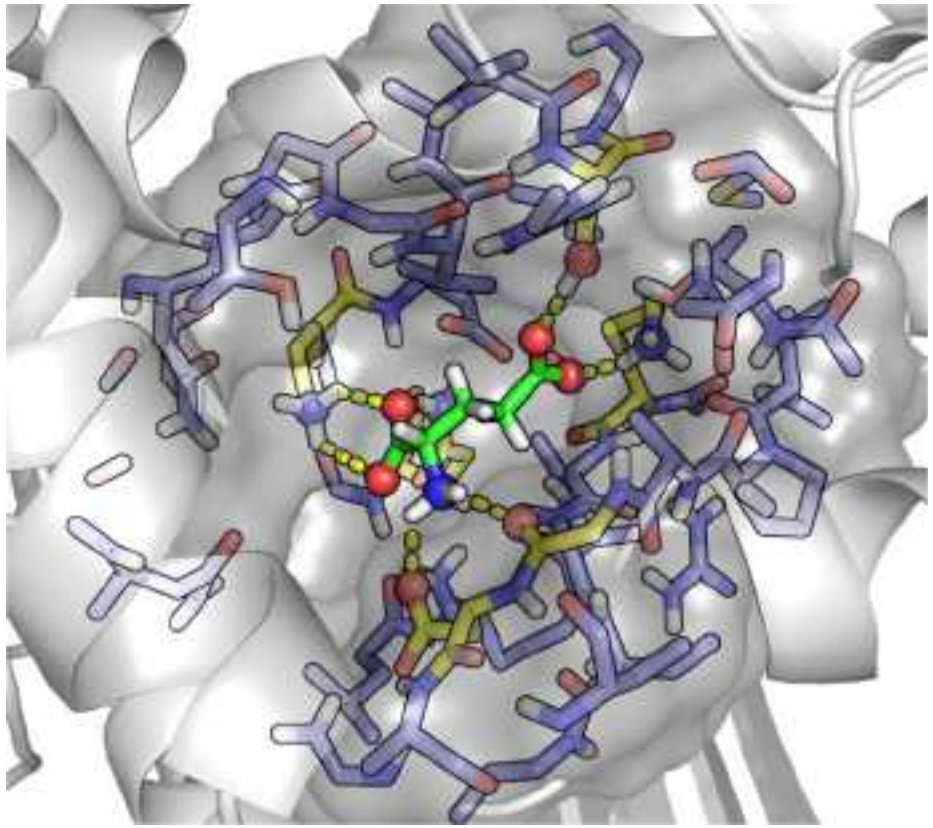
1 H & 2 LP

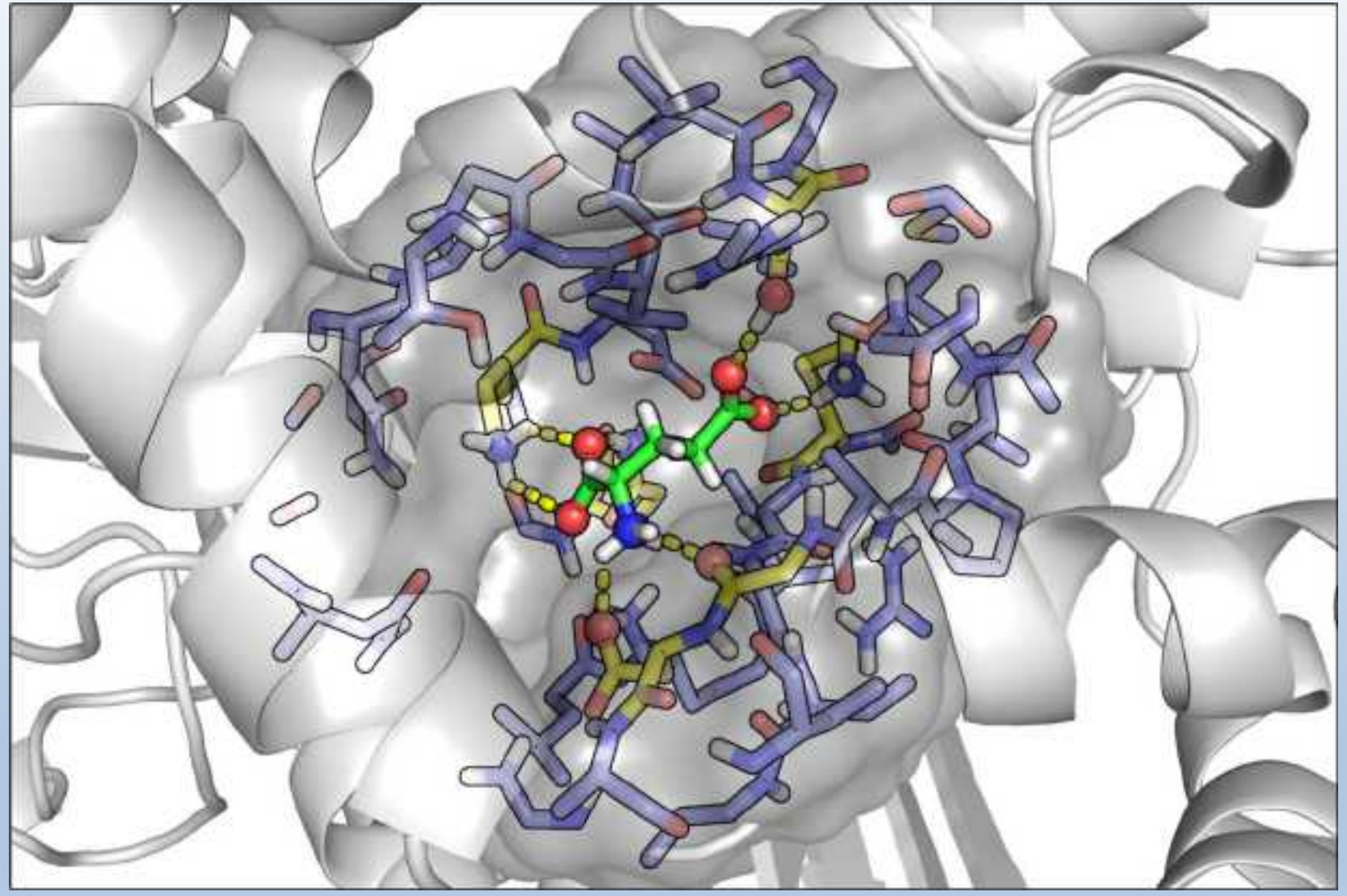


Threonine (Thr)

1 H & 2 LP

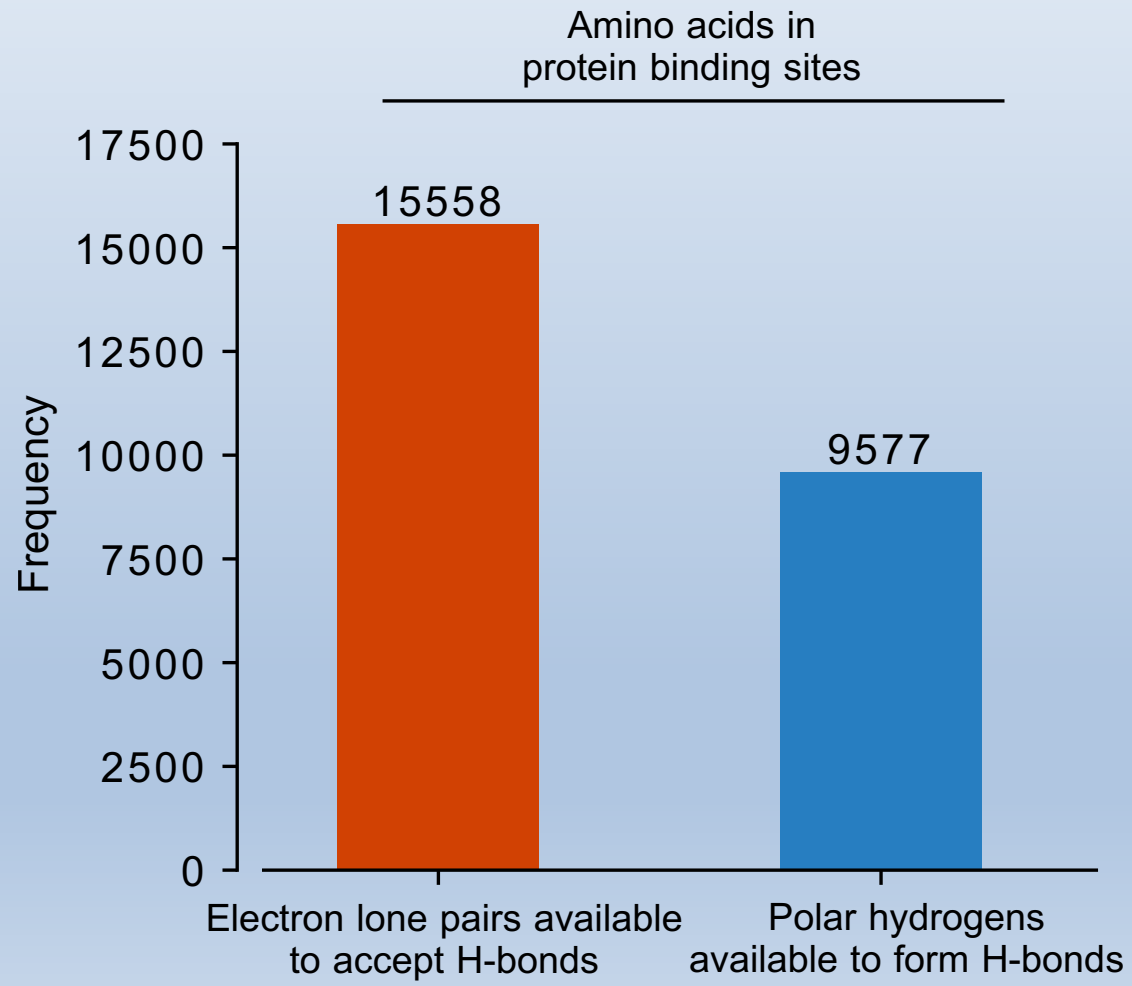
...



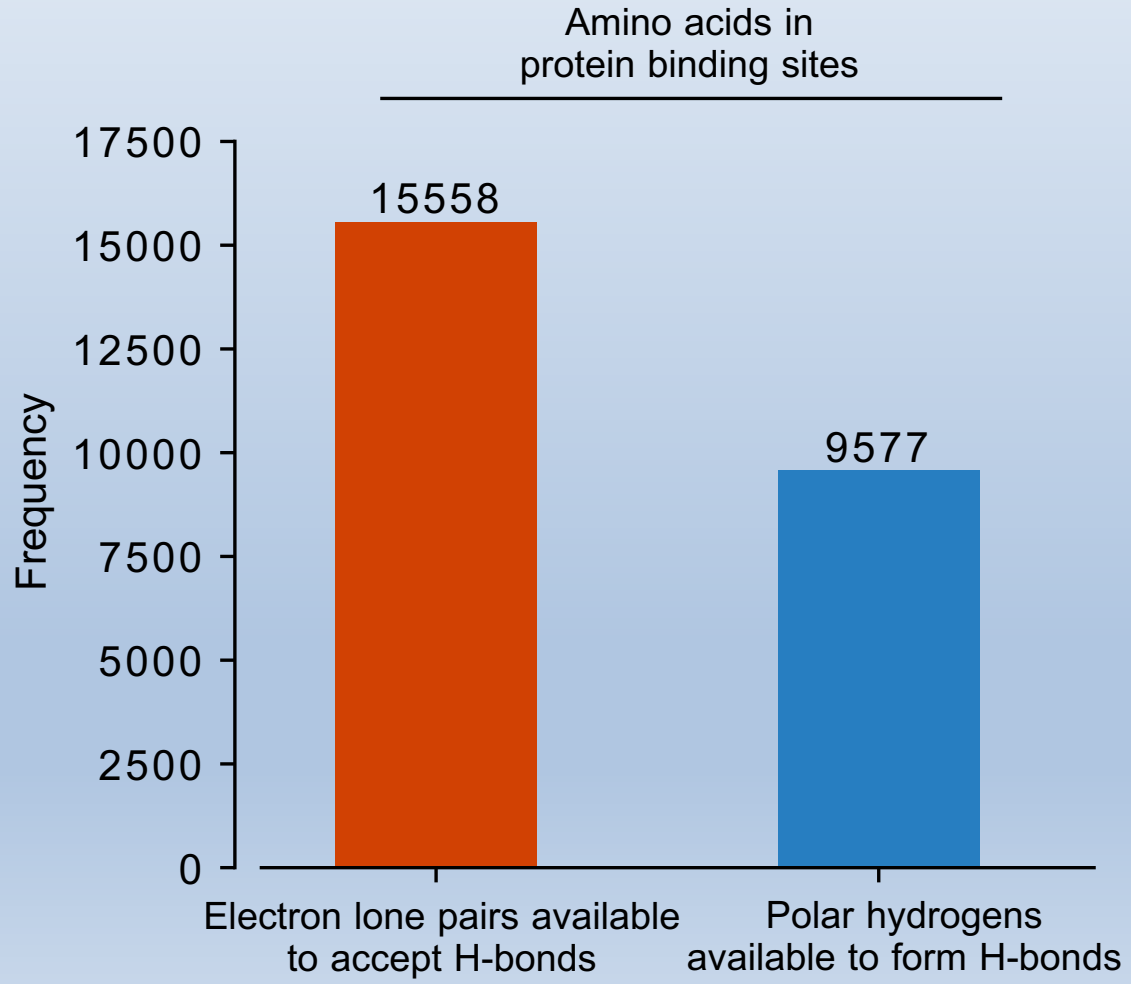


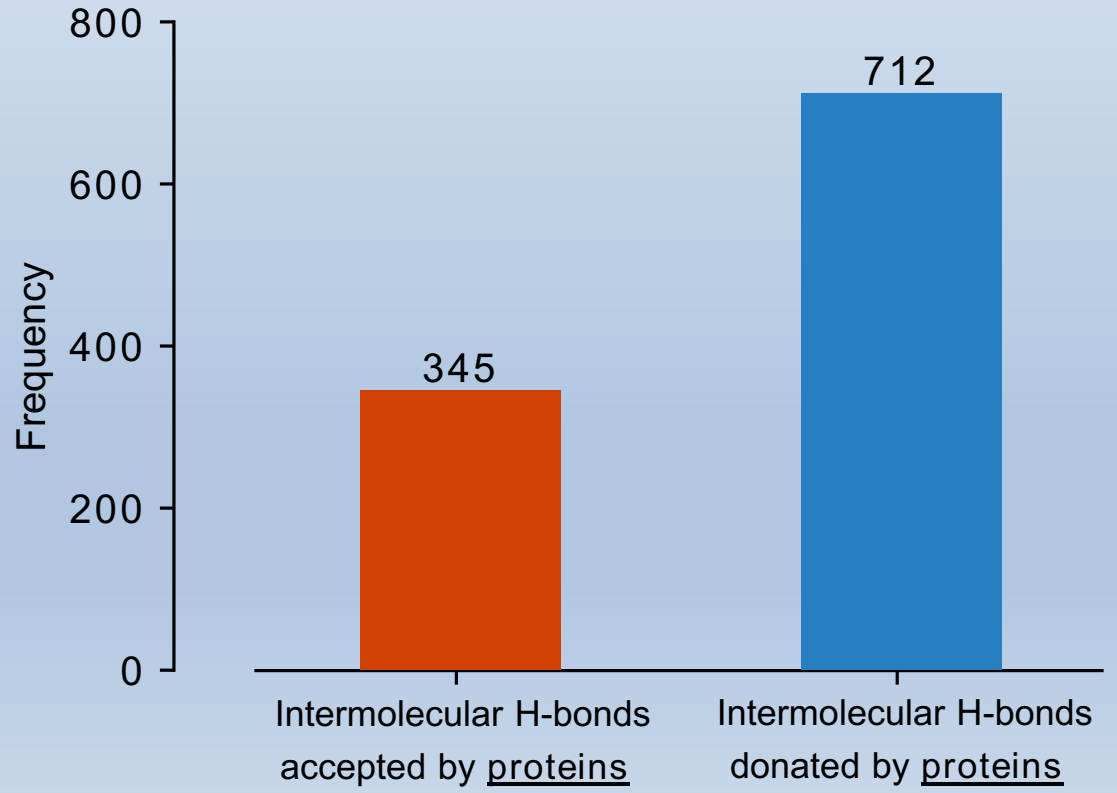
9 Å binding site definition for glutamate hydrogenase interacting with a glutamic acid ligand (PDB ID: 1bgv)

Excess of electron lone pairs does not explain trend that protein atoms favor donating H-bonds

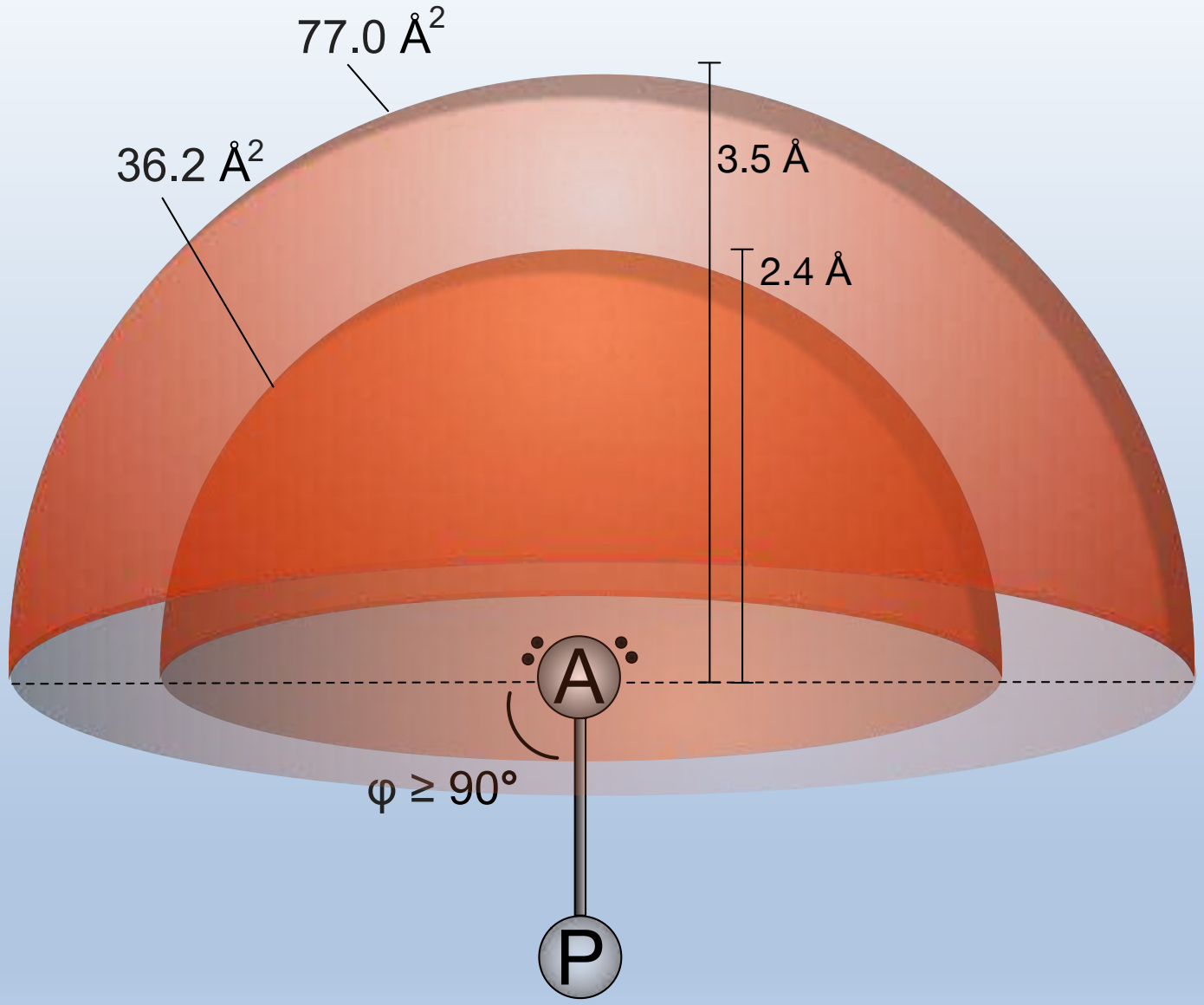


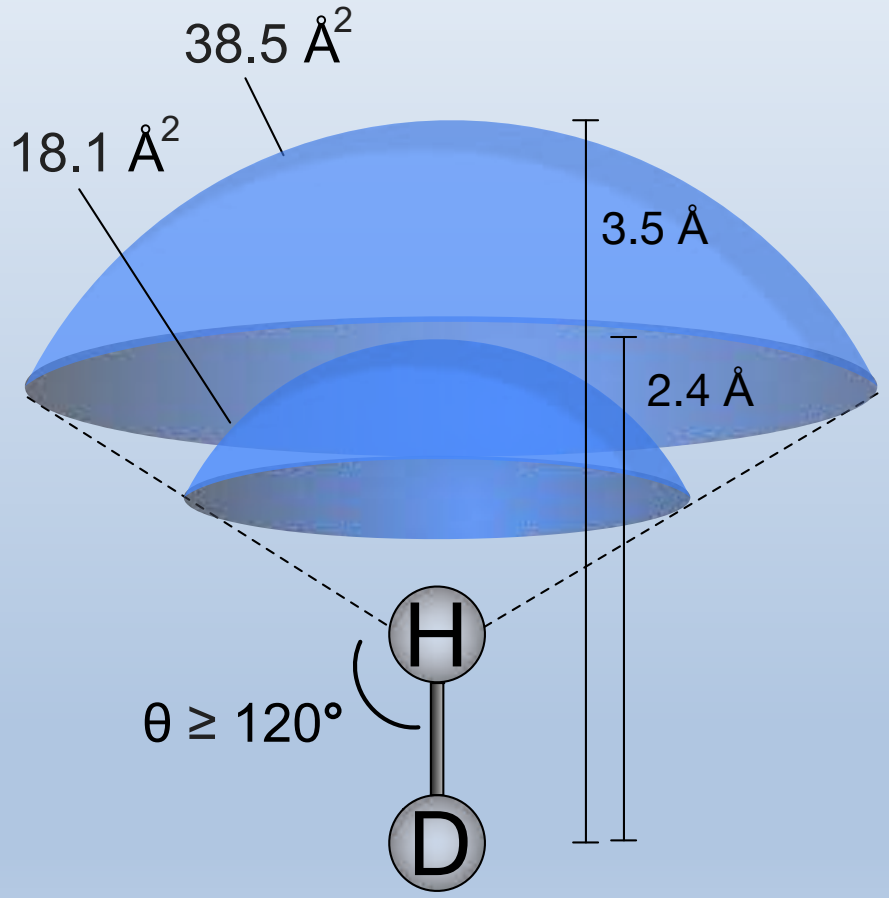
Excess of electron lone pairs does not explain trend that protein atoms favor donating H-bonds





Apparently, there is a strong chemical or evolutionary preference for proteins to act as H-bond donors


77.0 Å2
36.2 Å2
3.5 Å
2.4 Å
A
φ ≥ 90°
P

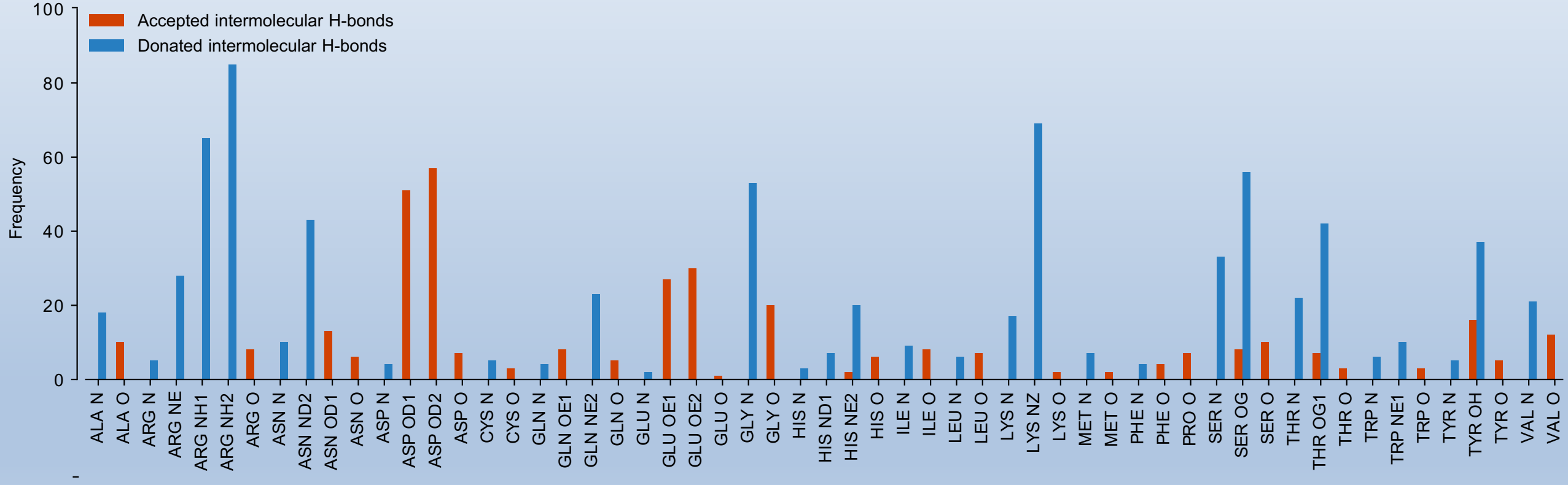

38.5 Å2
18.1 Å2
3.5 Å
2.4 Å
H
θ ≥ 120°
D

Groups that can both donate and accept
(e.g., hydroxyl groups) bring the risk of misrecognition
(promiscuous binding), because many ligands can match in
many different orientations

76% of intermolecular H-bonds are donated by a nitrogen atom

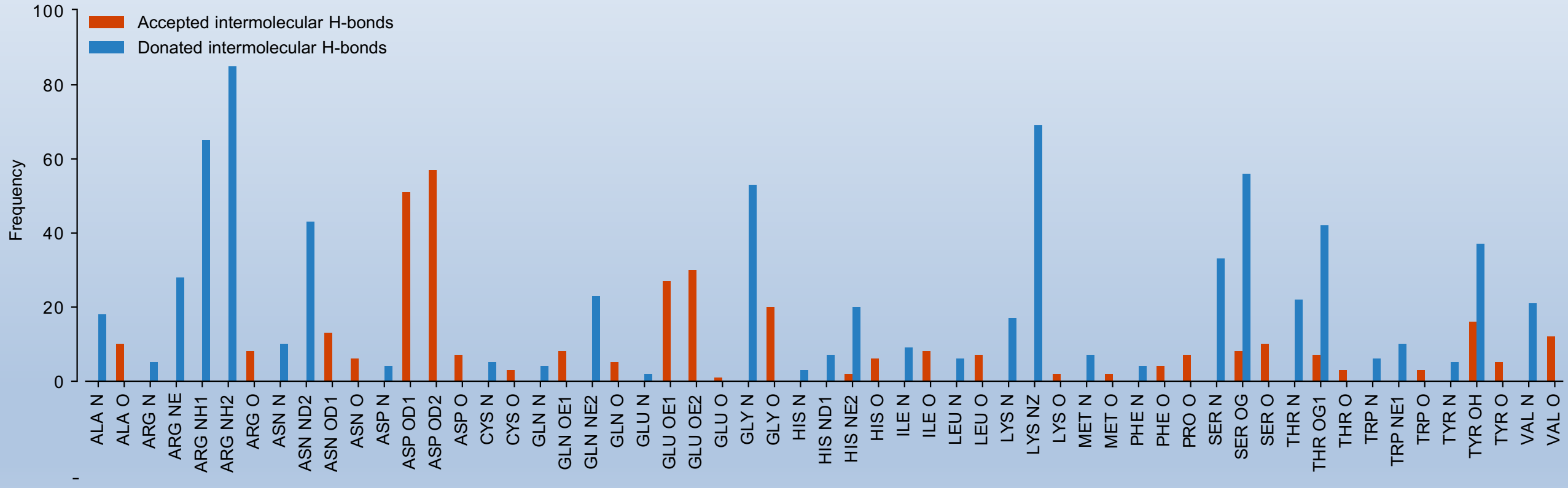
H-bond donor molecule	H-bond type	Frequency	H-bond acceptor molecule
Protein	N-H ··· O	524	Ligand
Protein	N-H ··· N	53	Ligand
Protein	O-H ··· O	127	Ligand
Protein	O-H ··· N	6	Ligand
Ligand	N-H ··· O	219	Protein
Ligand	N-H ··· N	1	Protein
Ligand	O-H ··· O	124	Protein
Ligand	O-H ··· N	1	Protein

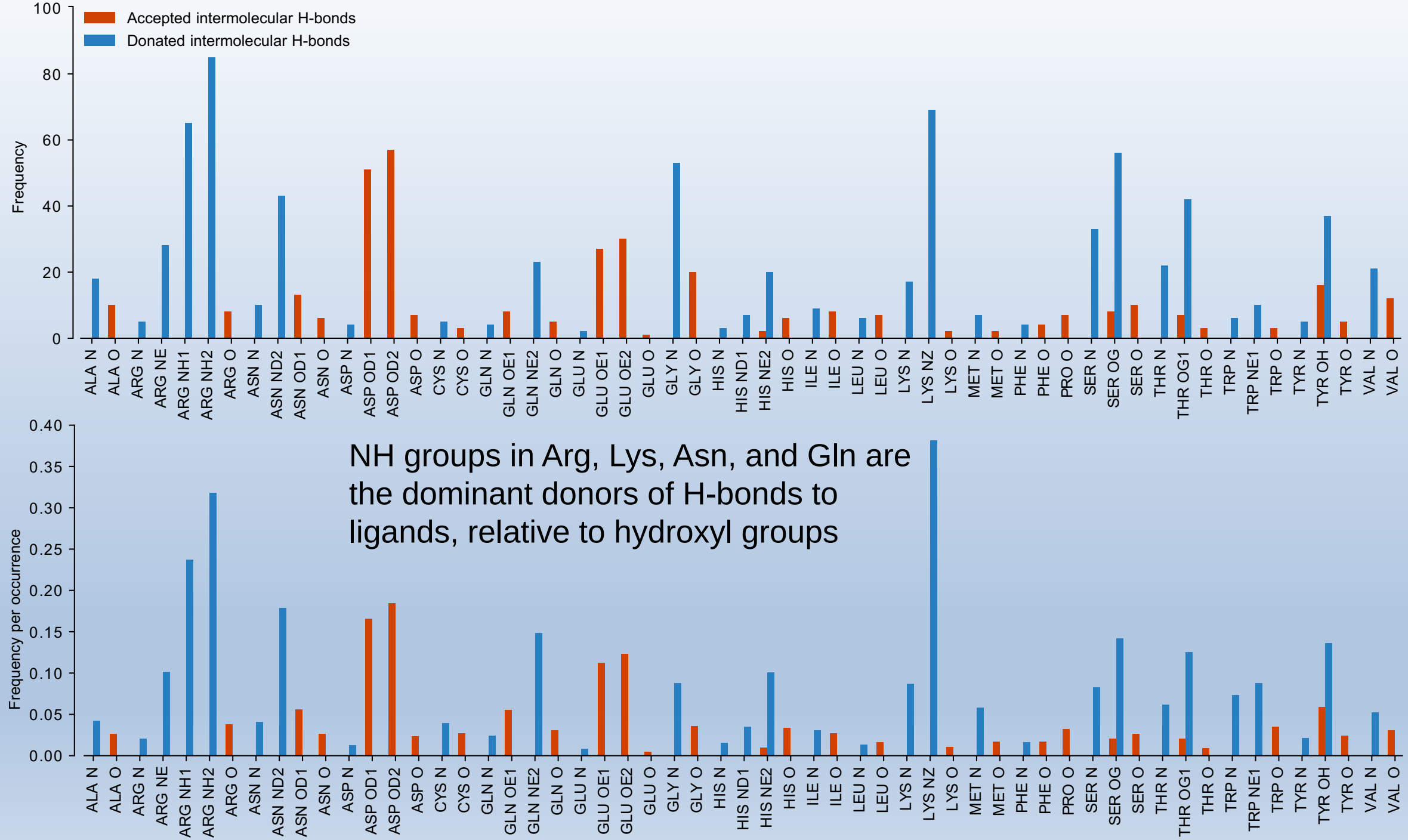
NH groups in Arg and Lys, are the dominant donors of H-bonds to ligands, relative to hydroxyl groups



NH groups in Arg, and Lys, are the dominant donors of H-bonds to ligands, relative to hydroxyl groups

Due to binding site prevalence?



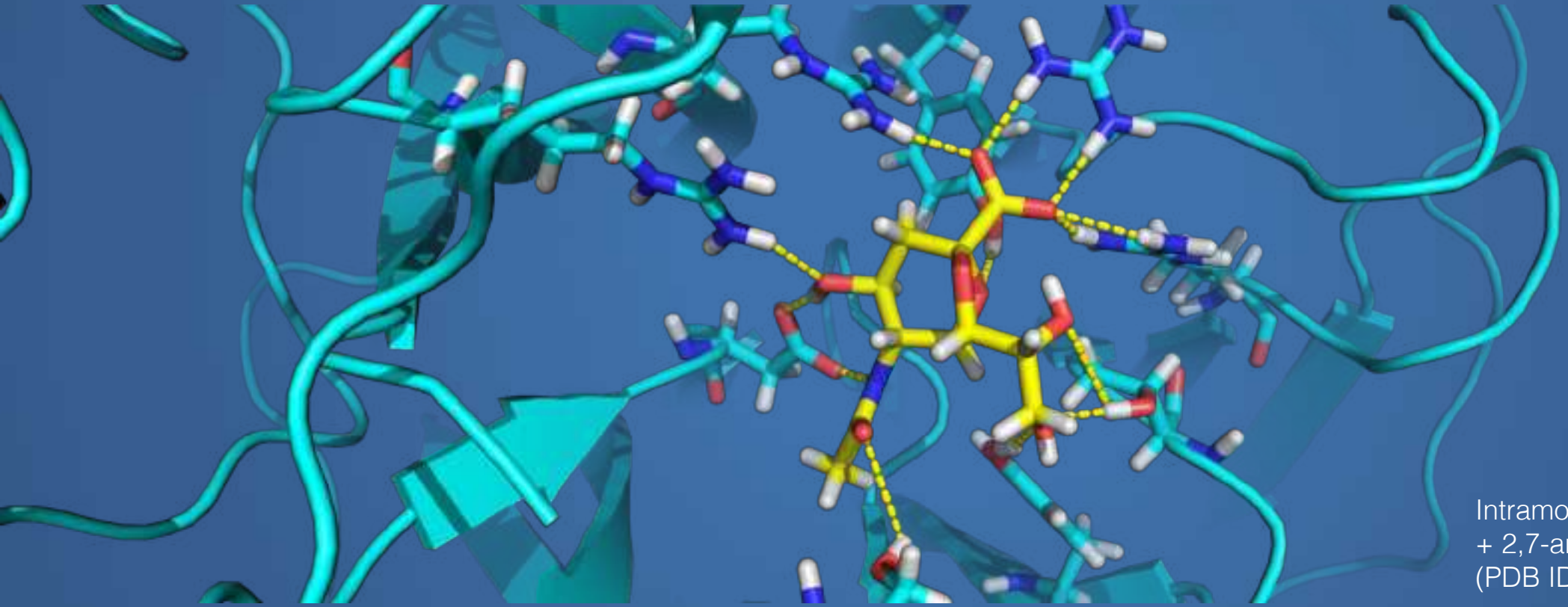


NH groups in Arg, Lys, Asn, and Gln are the dominant donors of H-bonds to ligands, relative to hydroxyl groups

Can the observed H-bonding trends be used to predict protein-ligand interactions?

Protein Recognition Index (PRI)

Computing the Protein Recognition Index

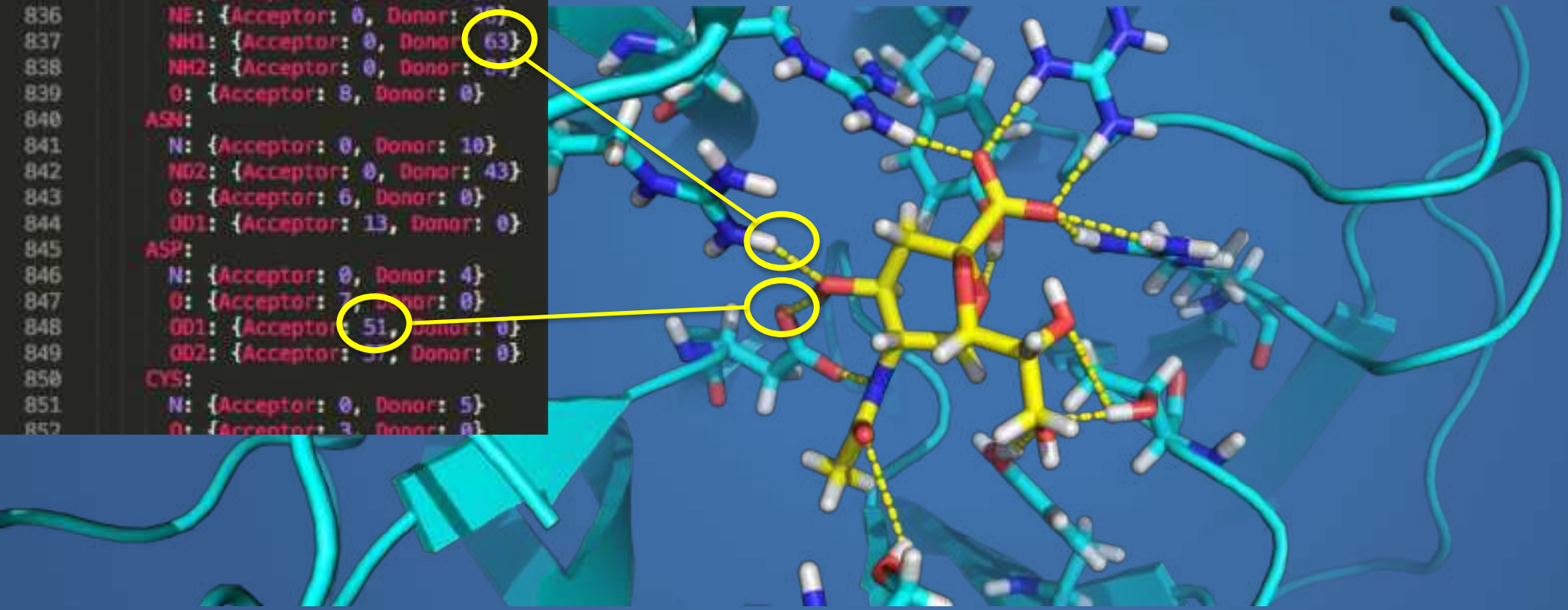


Intramolecular trans-Sialidase
+ 2,7-anhydro-Neu5Ac
(PDB ID: 2sli)

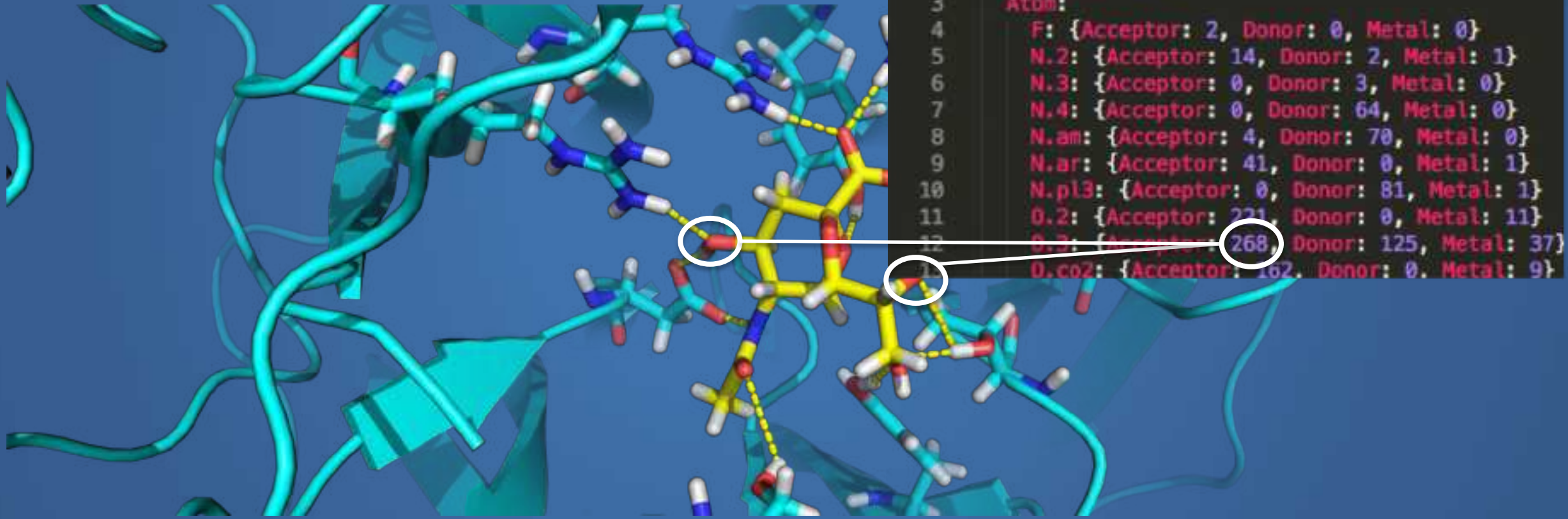
PRI-prot = 63 + 51 + …

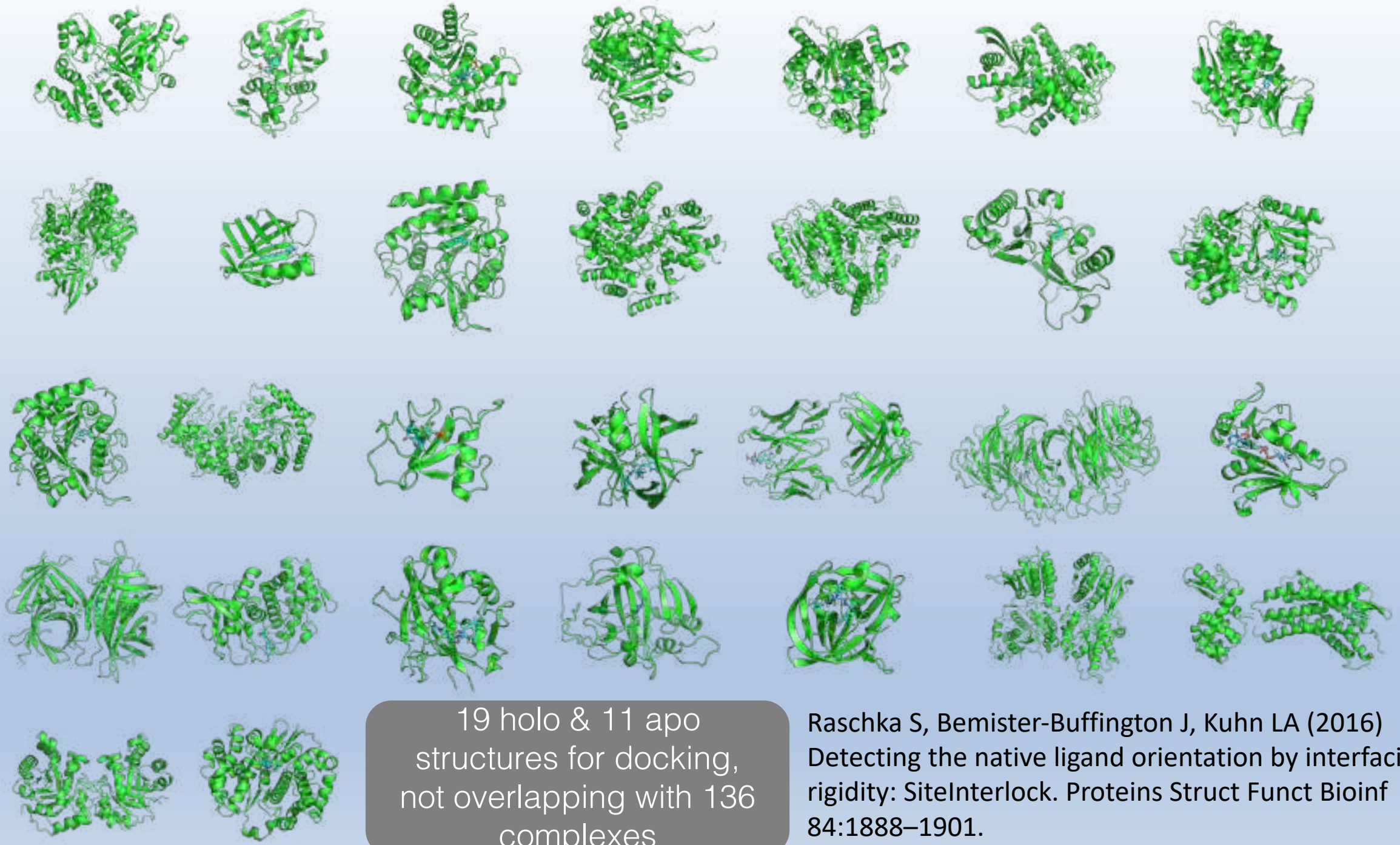
donor_acceptor_cnt.yml

```
Residue_Atom:
  ALA:
    N: {Acceptor: 0, Donor: 18}
    O: {Acceptor: 10, Donor: 0}
  ARG:
    N: {Acceptor: 0, Donor: 5}
    NE: {Acceptor: 0, Donor: 1}
    NH1: {Acceptor: 0, Donor: 63}
    NH2: {Acceptor: 0, Donor: 3}
    O: {Acceptor: 8, Donor: 0}
  ASN:
    N: {Acceptor: 0, Donor: 10}
    ND2: {Acceptor: 0, Donor: 43}
    O: {Acceptor: 6, Donor: 0}
    OD1: {Acceptor: 13, Donor: 0}
  ASP:
    N: {Acceptor: 0, Donor: 4}
    O: {Acceptor: 7, Donor: 0}
    OD1: {Acceptor: 51, Donor: 0}
    OD2: {Acceptor: 5, Donor: 0}
  CYS:
    N: {Acceptor: 0, Donor: 5}
    O: {Acceptor: 3, Donor: 0}
```



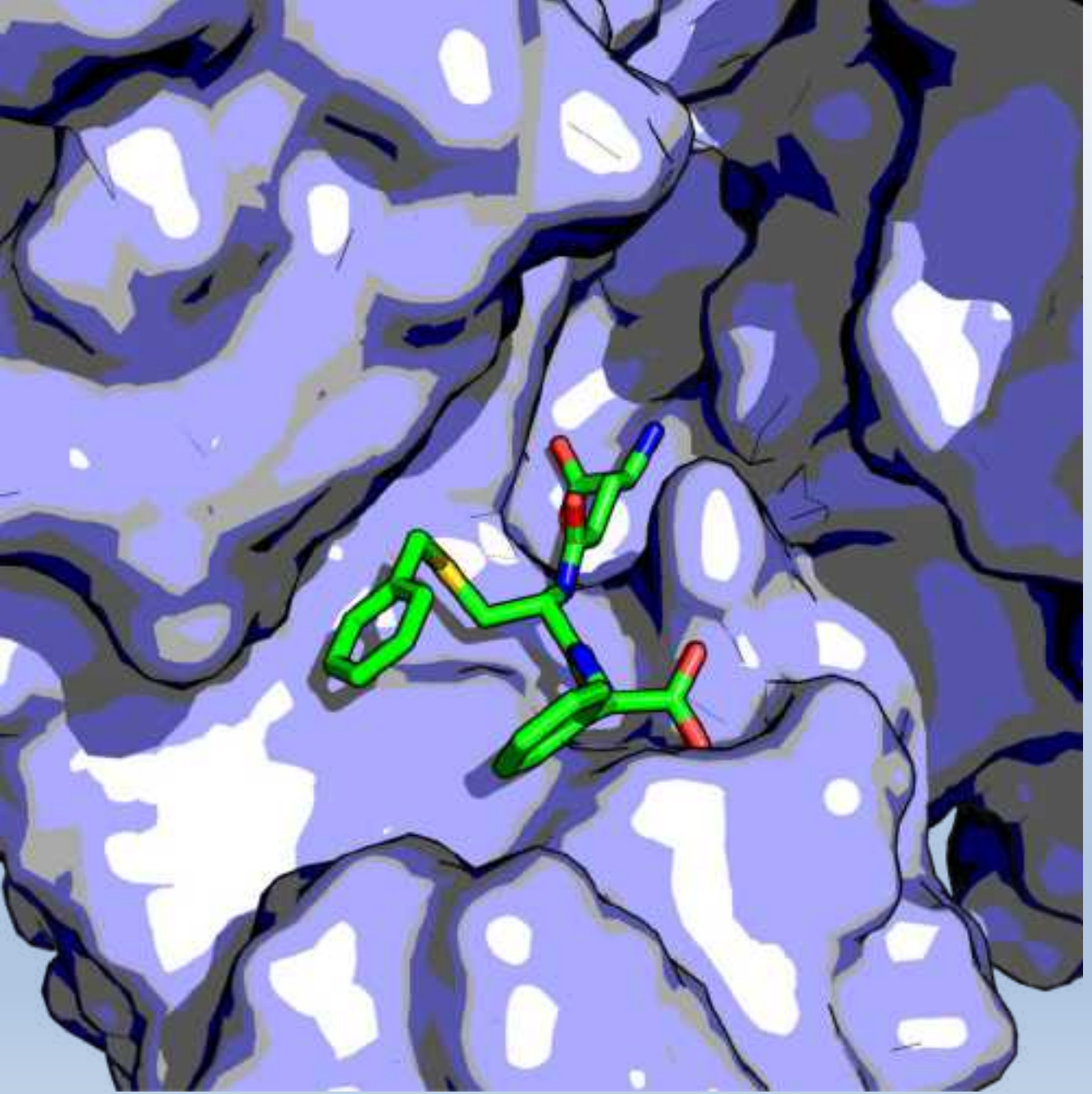
PRI-lig = 268 + 268 + …



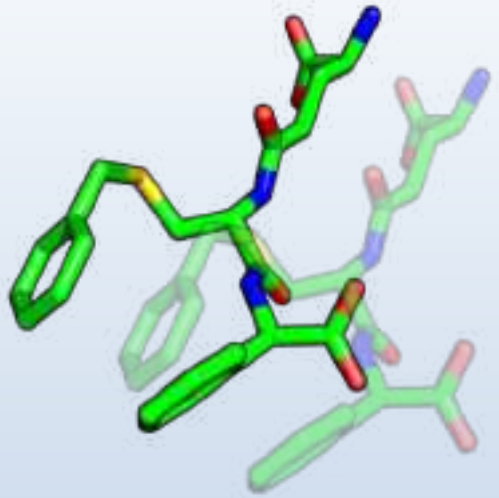


19 holo & 11 apo structures for docking, not overlapping with 136 complexes

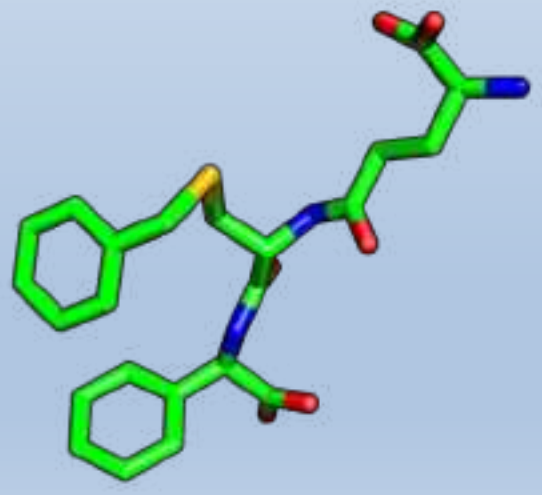
Raschka S, Bemister-Buffington J, Kuhn LA (2016) Detecting the native ligand orientation by interfacial rigidity: SiteInterlock. Proteins Struct Funct Bioinf 84:1888–1901.



Glutathione s-transferase + modified glutathione inhibitor
(PDB ID: 10gs)



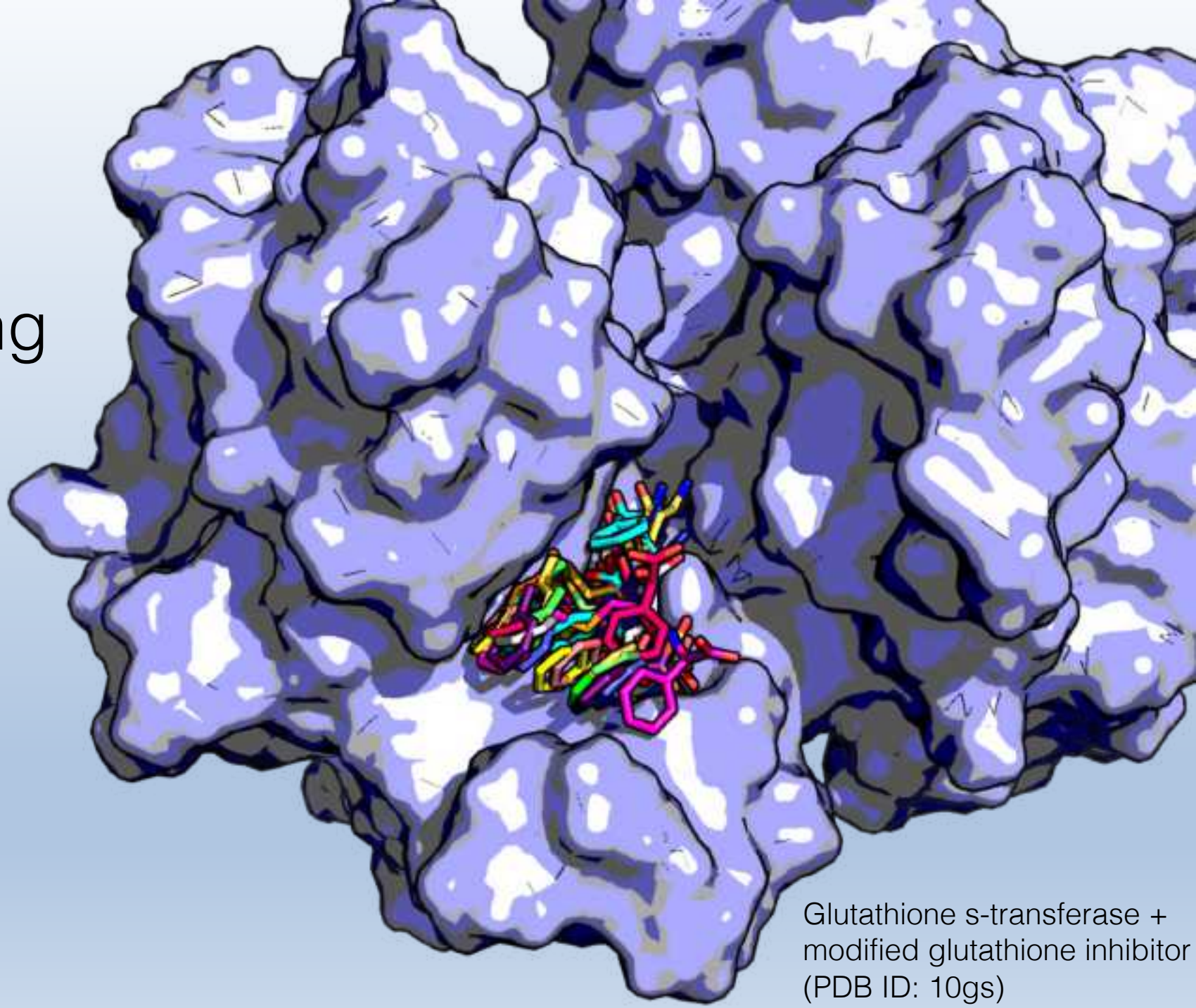
Orientation



Conformation

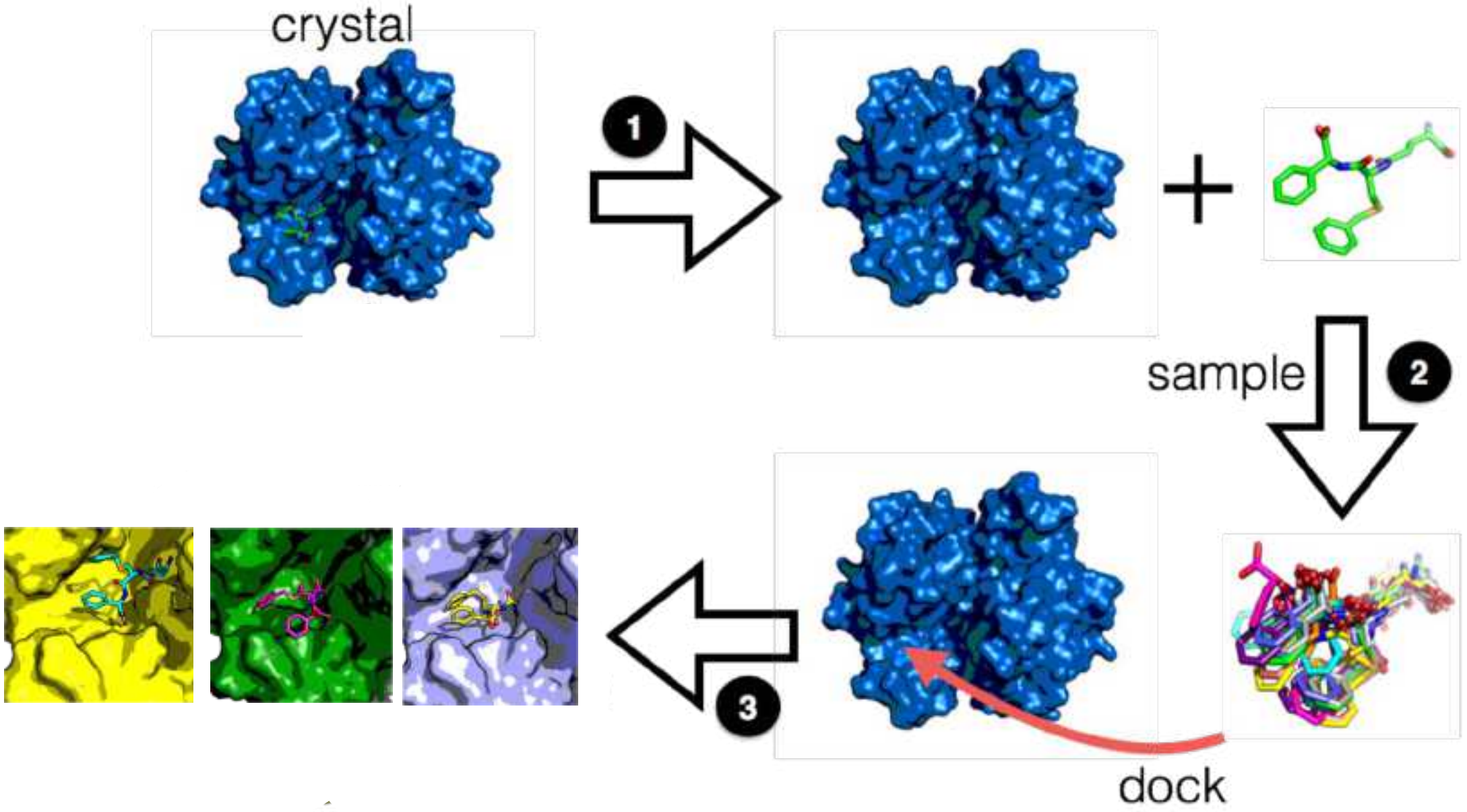
Subset of docking poses sampled for scoring

(sampled protein side-chains not shown)

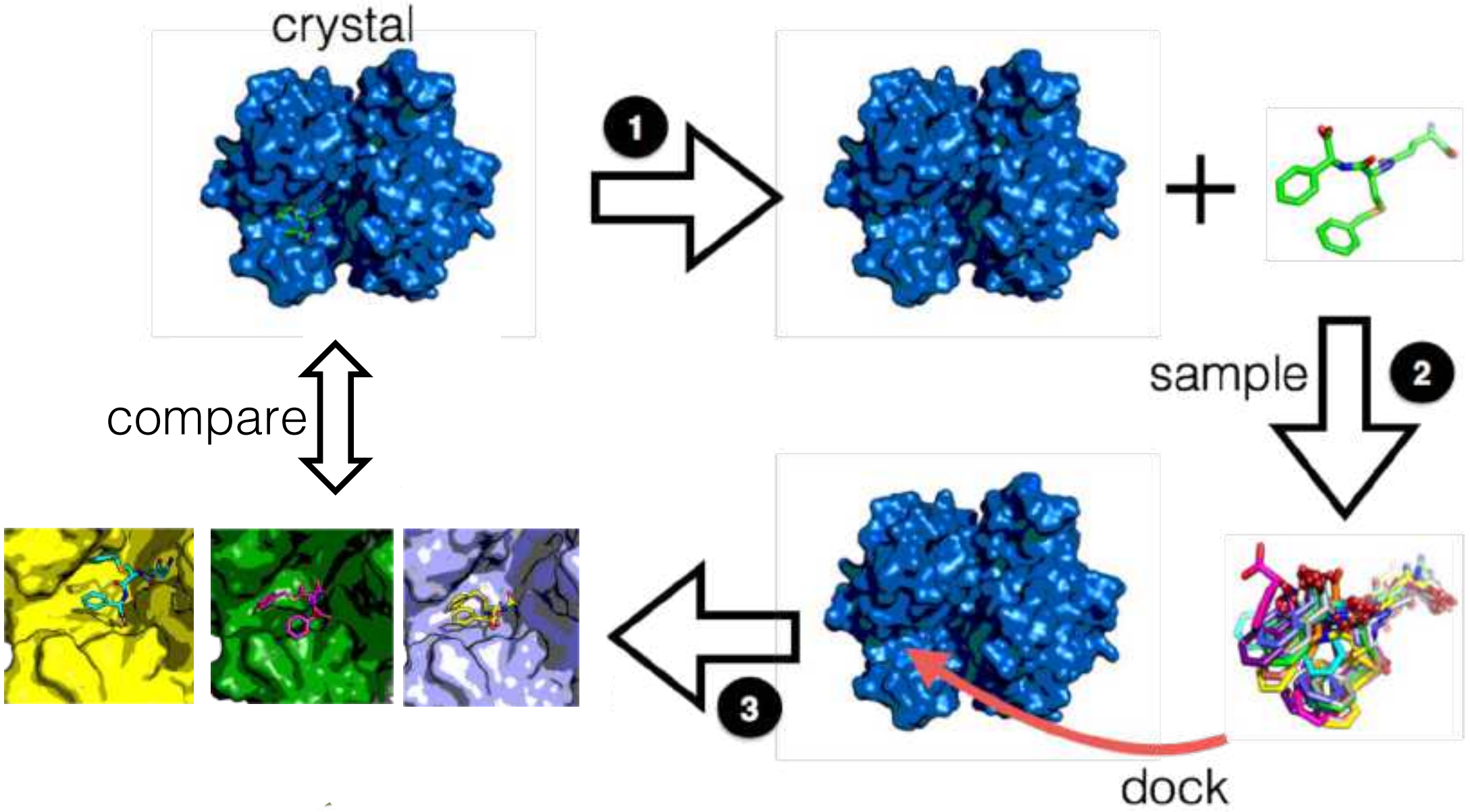


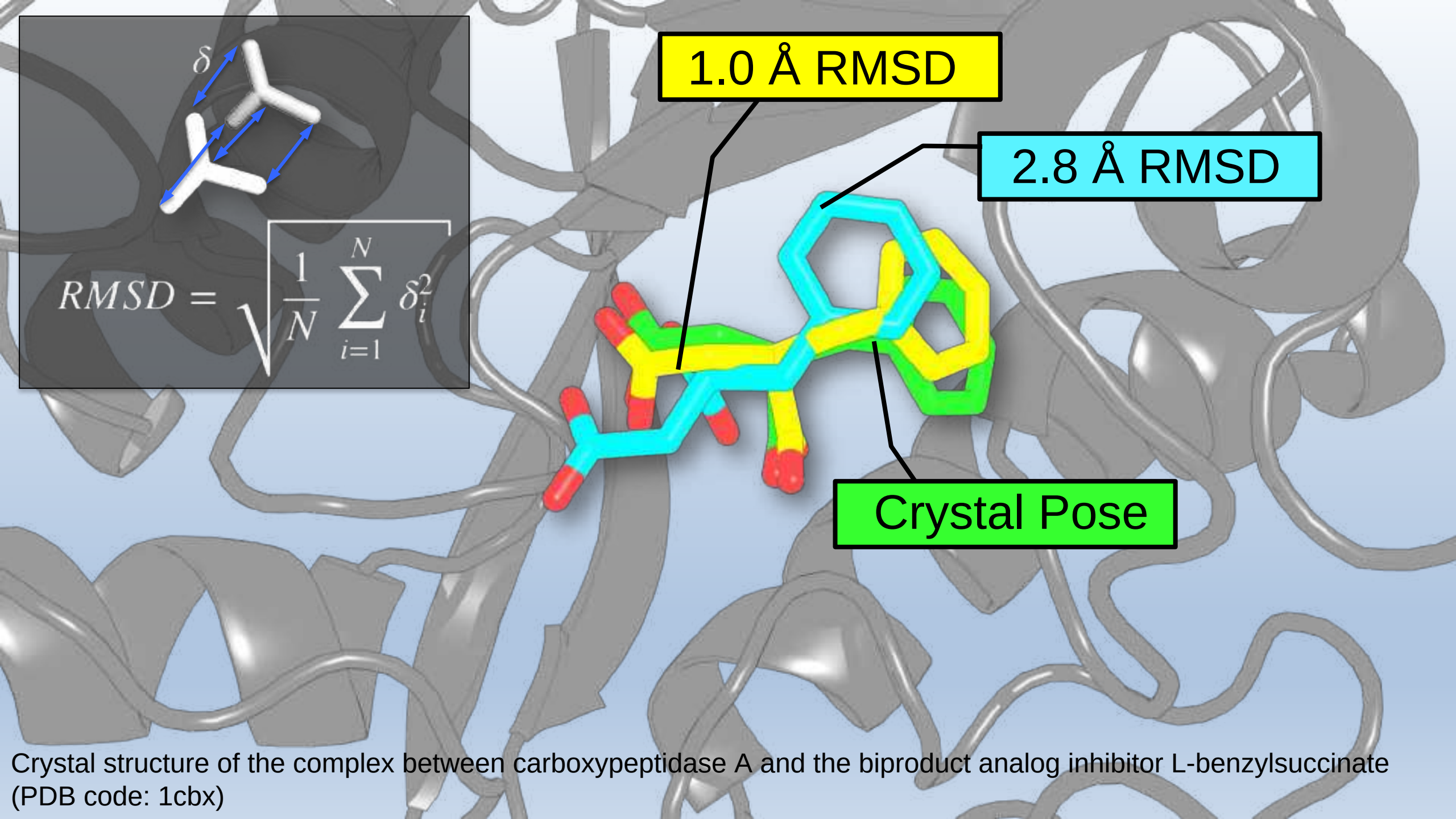
Glutathione s-transferase + modified glutathione inhibitor (PDB ID: 10gs)

Binding pose prediction

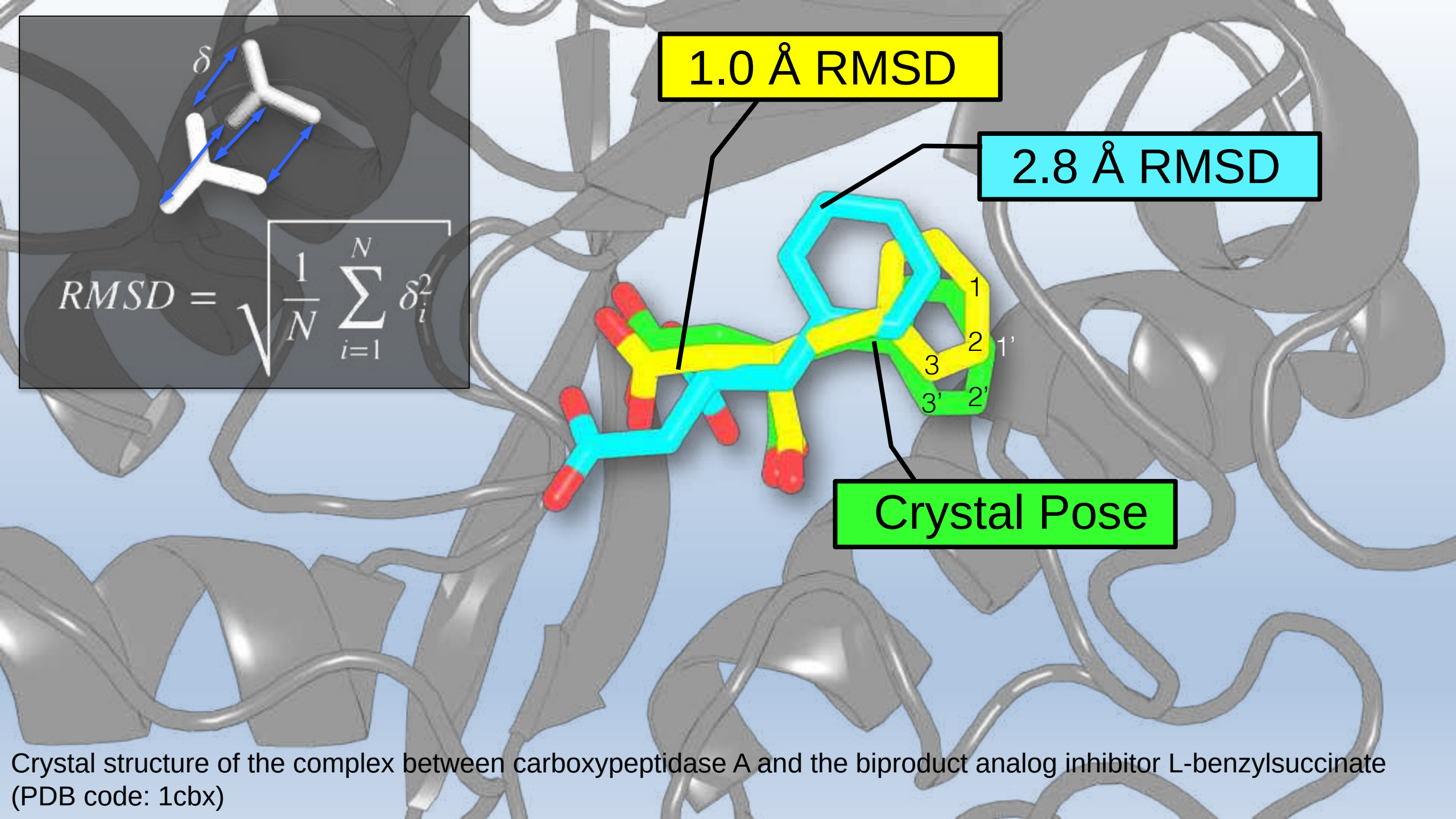


Binding pose prediction





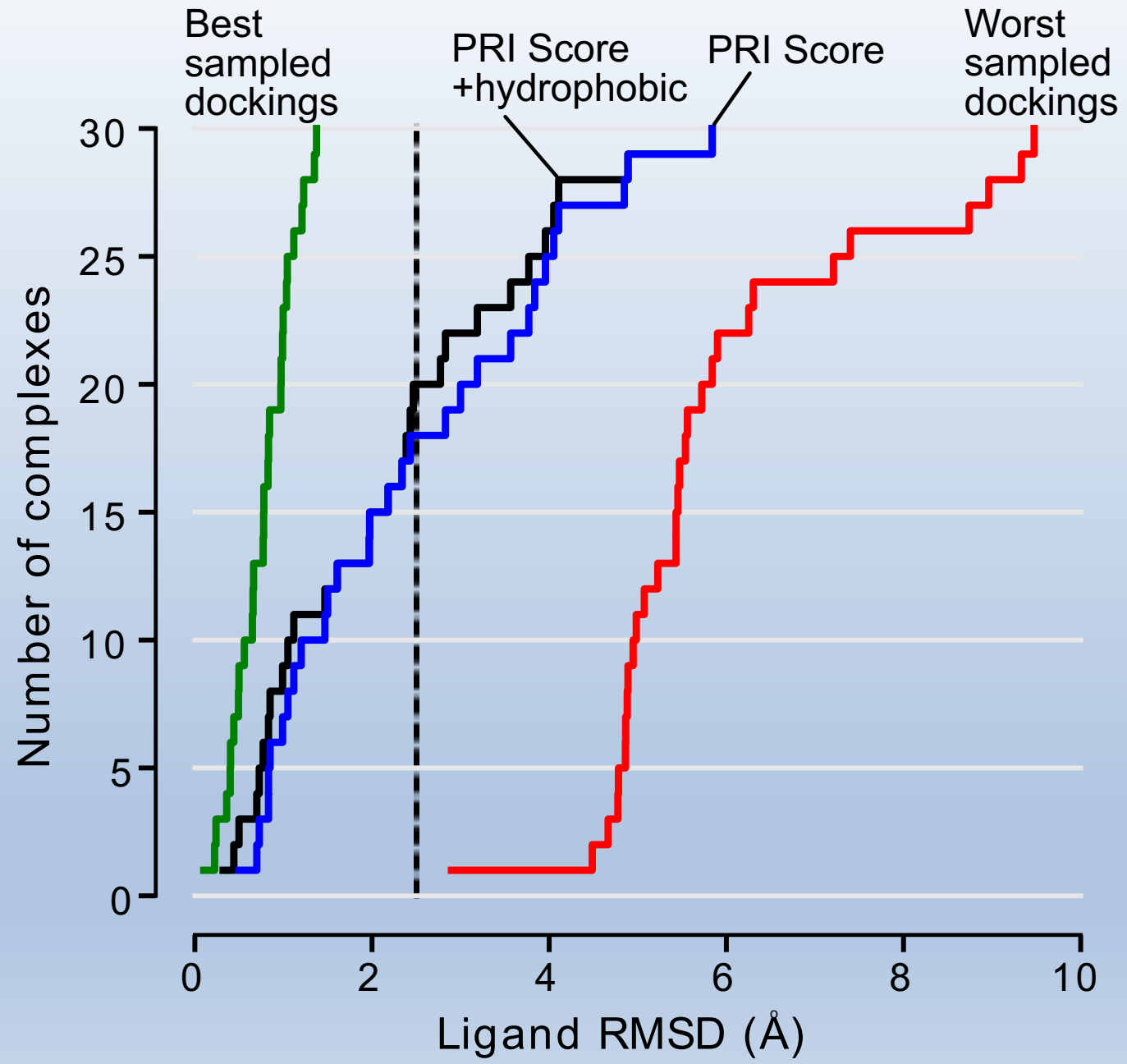
Crystal structure of the complex between carboxypeptidase A and the biproduct analog inhibitor L-benzylsuccinate (PDB code: 1cbx)



Crystal structure of the complex between carboxypeptidase A and the biproduct analog inhibitor L-benzylsuccinate (PDB code: 1cbx)

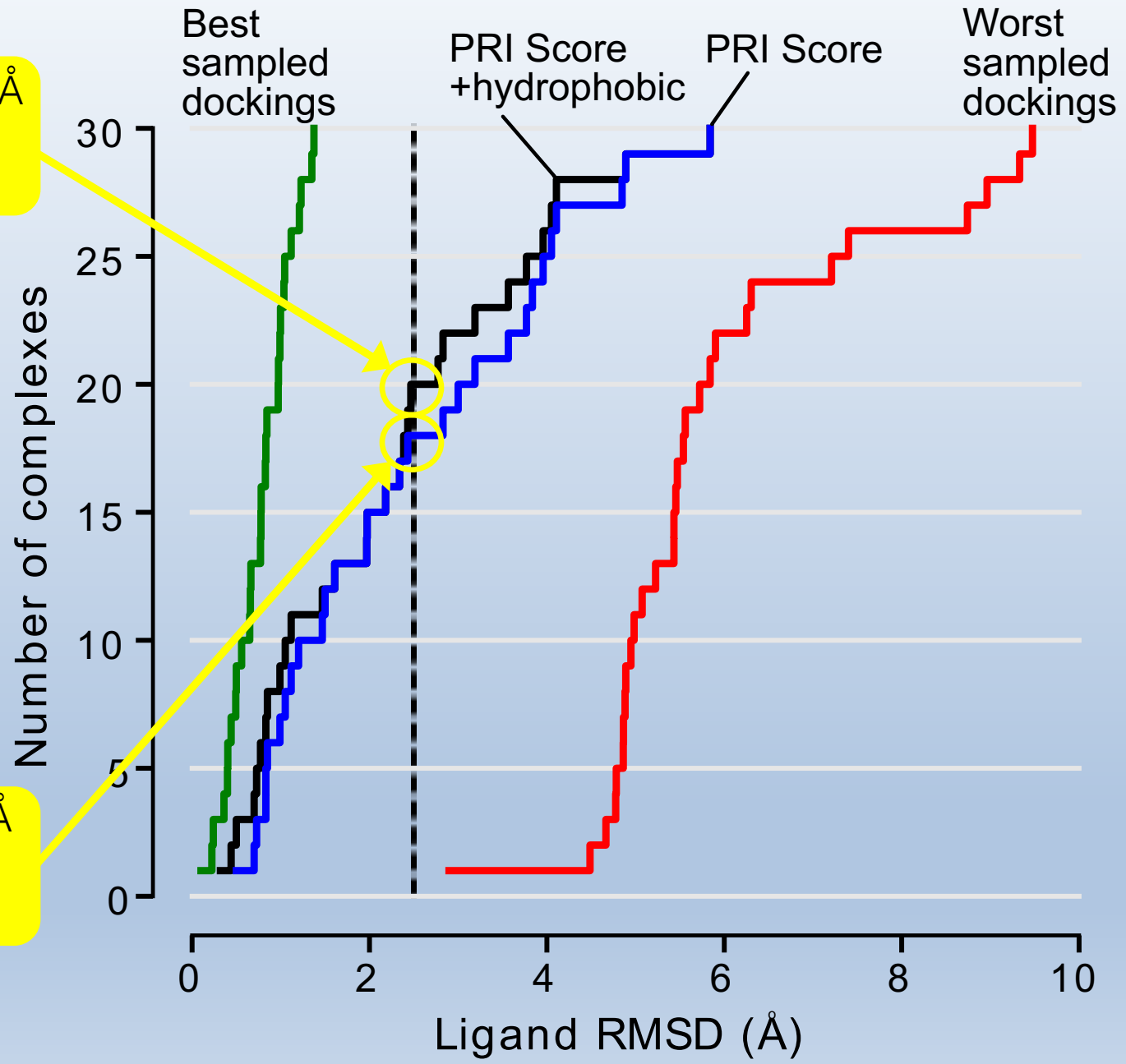
Can the general, observed H-bonding trends be used to predict protein-ligand interactions in individual complexes?

Protein Recognition Index


Best sampled dockings
PRI Score +hydrophobic
PRI Score
Worst sampled dockings
Number of complexes
30
25
20
15
10
5
0
0
2
4
6
8
10
Ligand RMSD (Å)

poses w. less than 2.5 Å predicted in 20 out of 30 cases

poses w. less than 2.5 Å predicted in 18 out of 30 cases



H-bond interaction statistics accumulated across 136 structures capture the essential molecular recognition features that occur within individual structures sufficiently well enough to discriminate native interactions

Comparison to Lipinski's rule of five for orally active drugs

1. No more than 5 hydrogen bond donors
2. No more than 10 hydrogen bond acceptors
3. A molecular mass less than 500 daltons
4. An octanol-water partition coefficient log *P* not greater than 5

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997)
Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.
Adv Drug Deliv Rev 23:3–25.

Violation results in poor absorption or permeability

Comparison to Lipinski's rule of five for orally active drugs

1. No more than 5 hydrogen bond donors
2. No more than 10 hydrogen bond acceptors
3. A molecular mass less than 500 daltons
4. An octanol-water partition coefficient log *P* not greater than 5

* All numbers are multiples of 5 (origin of the name)

Violation results in poor absorption or permeability

Comparison to Lipinski's rule of five for orally active drugs

- Analysis of interactions (rather than physicochemical properties of ligands)
- Twice as many H-bonds being accepted by ligands as donated
- N-H donors are favored over O-H donors
- High preference for certain amino acid side chains (Arg, Lys)
- Protein Recognition Index predictive of how a ligand interacts

Conclusions

Conclusions

- Protein-ligand interfaces are polarized: proteins donate twice as many H-bonds as they accept
- H-bond donors and N-H over O-H groups are preferred, allowing for higher ligand selectivity
- Lys, Arg, Glu, and Asp (charged amino acids) are preferred in intermolecular H-bonds
- A chemical preference key (PRI) provides chemical insights for predicting protein-ligand complexes

Conclusions

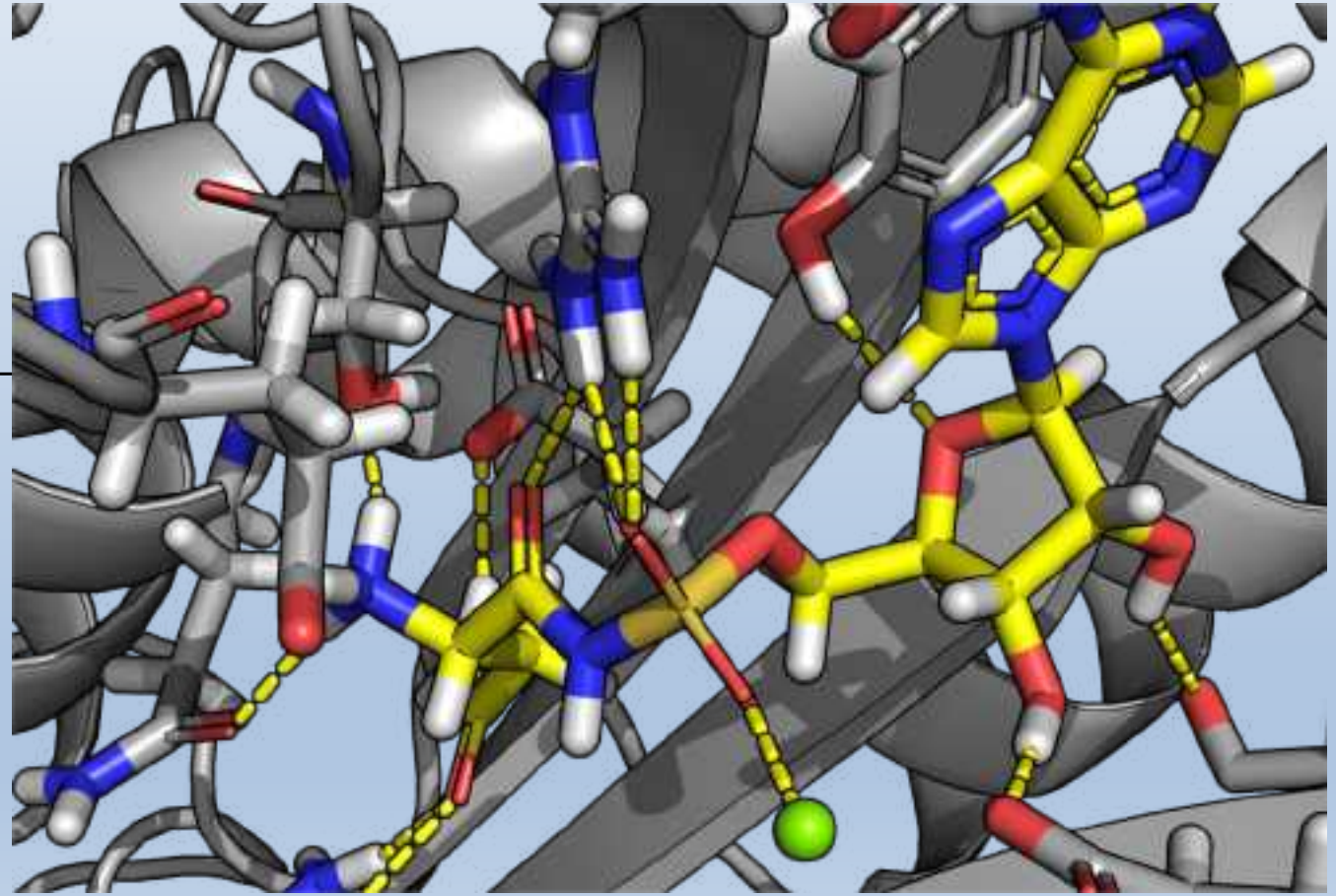
- Protein-ligand interfaces are polarized: proteins donate twice as many H-bonds as they accept
- H-bond donors and N-H over O-H groups are preferred, allowing for higher ligand selectivity
- Lys, Arg, Glu, and Asp (charged amino acids) are preferred in intermolecular H-bonds
- A chemical preference key (PRI) provides chemical insights for predicting protein-ligand complexes

Both *Hbind* and *PRI* software will be made available (open source)

Applications

Protein and Ligand Design

H-bond donor molecule	H-bond type	Frequency	H-bond acceptor molecule
Protein	N-H · · · O	524	Ligand
Protein	N-H · · · N	53	Ligand
Protein	O-H · · · O	127	Ligand
Protein	O-H · · · N	6	Ligand
Ligand	N-H · · · O	219	Protein
Ligand	N-H · · · N	1	Protein
Ligand	O-H · · · O	124	Protein
Ligand	O-H · · · N	1	Protein

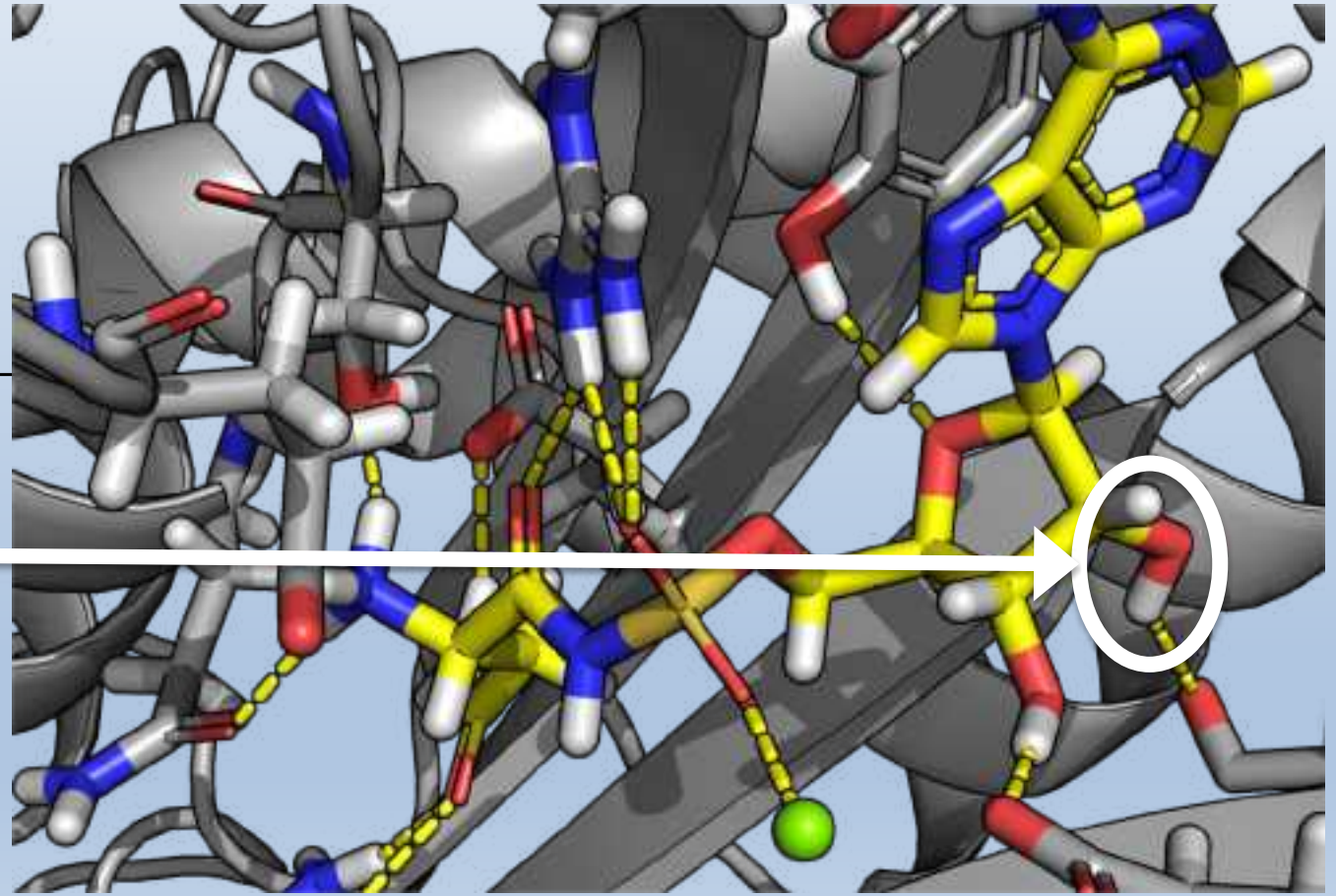


Asparaginyl-tRNA synthetase complexed with the sulfamoyl analog of asparaginyl-adenylate (PDB ID: 2xgt)

Protein and Ligand Design

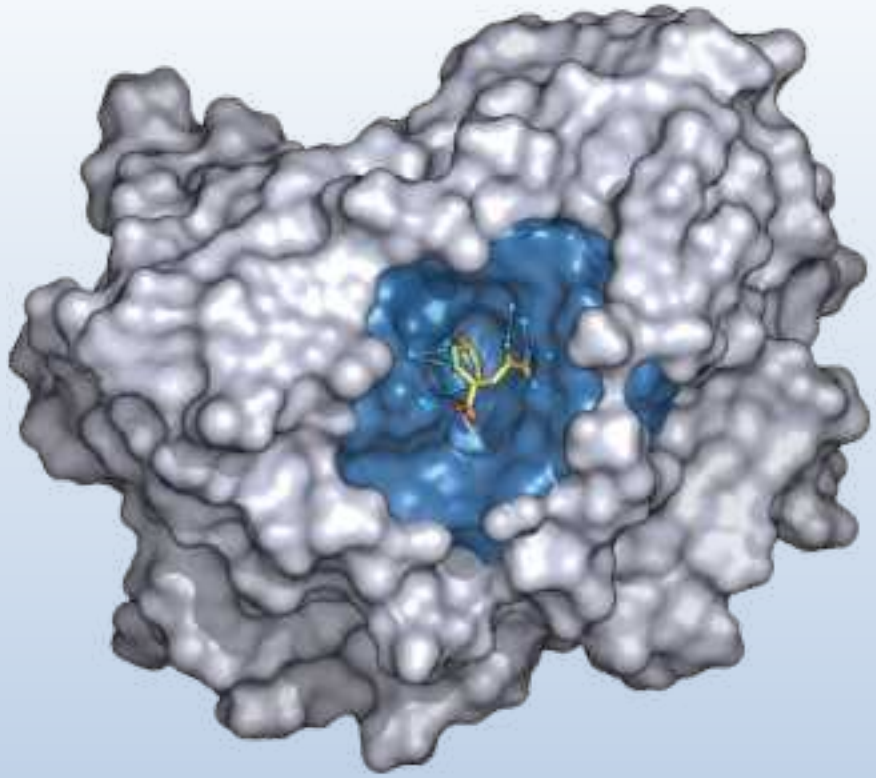
H-bond donor molecule	H-bond type	Frequency	H-bond acceptor molecule
Protein	N-H · · · O	524	Ligand
Protein	N-H · · · N	53	Ligand
Protein	O-H · · · O	127	Ligand
Protein	O-H · · · N	6	Ligand
Ligand	N-H · · · O	219	Protein
Ligand	N-H · · · N	1	Protein
Ligand	O-H · · · O	124	Protein
Ligand	O-H · · · N	1	Protein

OH → NH

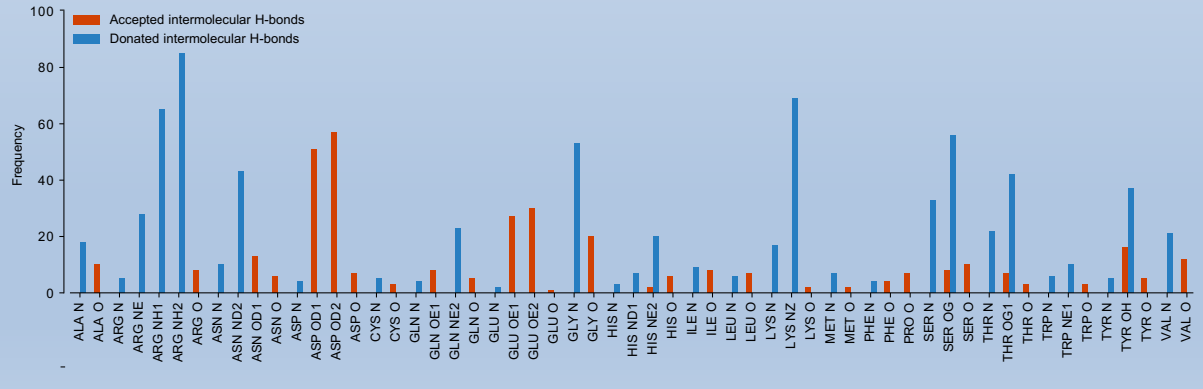


Asparaginyl-tRNA synthetase complexed with the sulfamoyl analog of asparaginyl-adenylate (PDB ID: 2xgt)

Predicting protein-ligand interactions



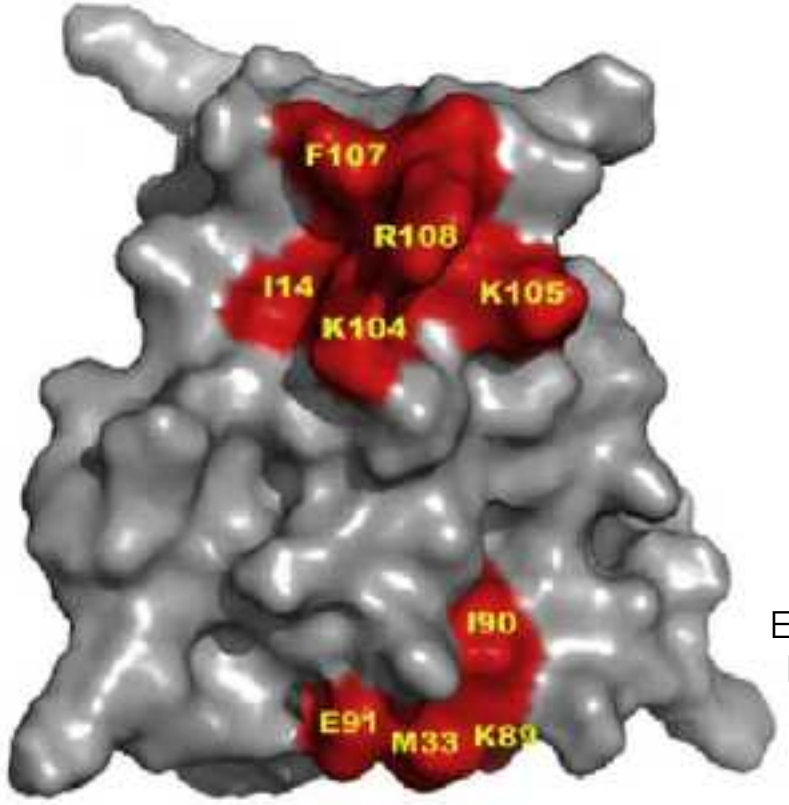
Chemical preference key (PRI)



Coupling of interactions (SiteInterlock)

Raschka, Bemister-Buffington & Kuhn (2016)
Detecting the native ligand orientation by interfacial rigidity: SiteInterlock.
Proteins Struct Funct
Bioinf 84:1888–1901.

Hotspots in Protein-Protein Binding Sites



Experimentally determined hotspot residues in IL-13

Figure adapted from
Agrawal NJ, Helk B, Trout BL (2014).
FEBS Lett 588:326–333. doi:10.1016/j.febslet.2013.11.004

Acknowledgements

Guidance Committee

Dr. David Arnosti
Dr. Titus Brown
Dr. Michael Feig
Dr. Jian Hu
Dr. Cheryl Kerfeld
Dr. Leslie Kuhn (Mentor)

Graduate Programs and BMB Staff

Dr. Michael Garavito
Dr. Jon Kaguni
Dr. John LaPres
Jessica Lawrence
Jeannine Lee
Becky Conat Mansel
Kaillathe (Pappan) Padmanabhan

Kuhn Lab

Joe Bemister-Buffington
Jiaxing Chen
Santosh Gunturu
John Johnston
Dr. Nan Liu
Alex Wolf

Collaborators (Screenlamp)

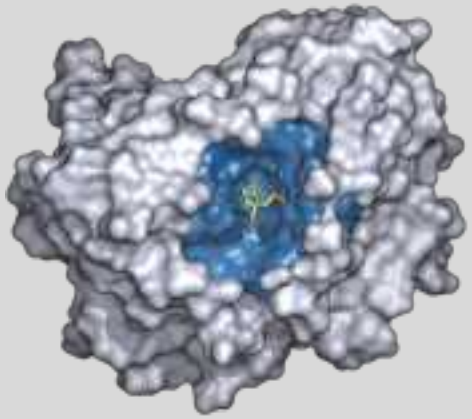
Dr. Mar Huertas
Dr. Weiming Li
Anne Scott

iPRoBe Lab

Dr. Vahid Mirjalili
Dr. Arun Ross

1

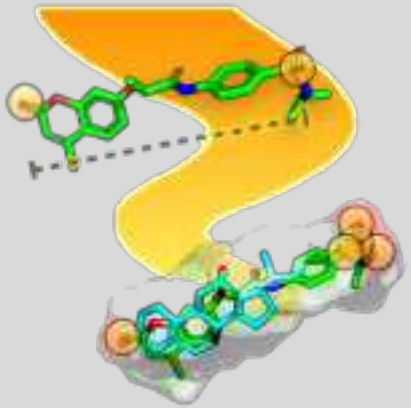
SiteInterlock



Raschka, Bemister-Buffington & Kuhn (2016) *Detecting the native ligand orientation by interfacial rigidity: SiteInterlock.* Proteins Struct Funct Bioinf 84:1888–1901.

2

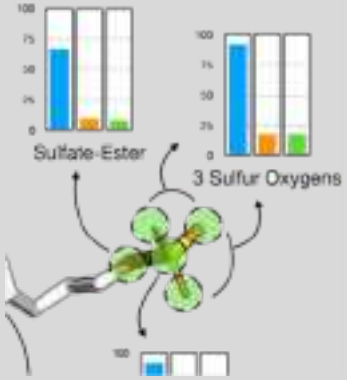
Screenlamp



Raschka, Scott, Liu, Gunturu, Huertas, Li & Kuhn (2017) *Enabling the hypothesis driven prioritization of ligand candidates in big databases: Screenlamp and its application to GPCR inhibitor discovery*. (In revision.)

3

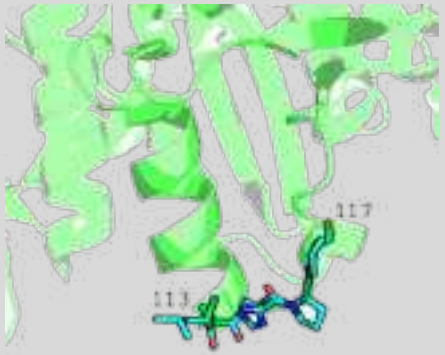
Machine Learning & Chemical Groups



Raschka, Kuhn, Scott, Huertas & Li (2017) *Computational Drug Discovery and Design: Automated inference of chemical group discriminants of biological activity from virtual screening data.* Springer. (In press.)

4

3D Epitope-Based Virtual Screening



Raschka, Zeng, Basson & Kuhn (2015-present)

5

Protein Recognition Index

Raschka, Wolf, Bemister-Buffington & Kuhn (2017) *Protein-ligand interfaces are polarized: Discovery of a strong trend for intermolecular hydrogen bonds to favor donors on the protein side with implications for predicting and designing ligand complexes.* (Submitted.)

Thanks for attending!

Questions?