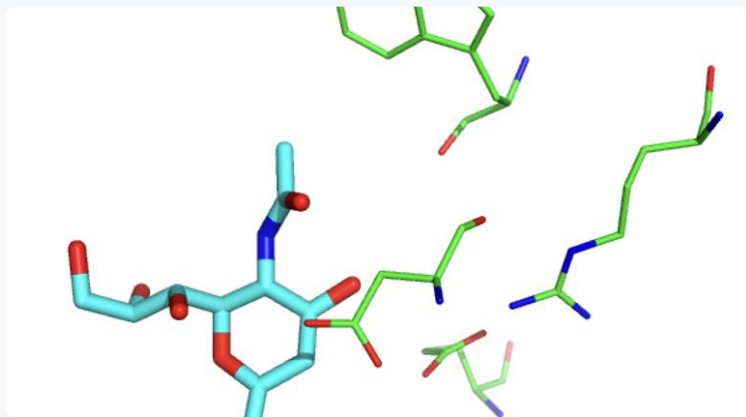




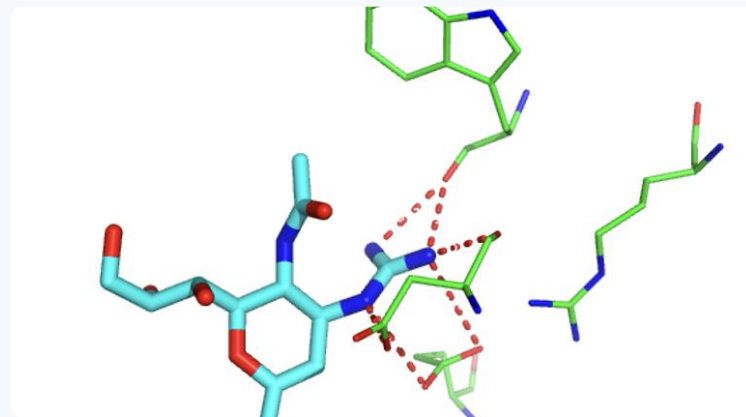
Using cooperative networks to better understand tight binding ligands

Neil R. Taylor
Desert Scientific Software
Sydney, Australia

Beyond the pairwise additive view of protein-ligand interactions

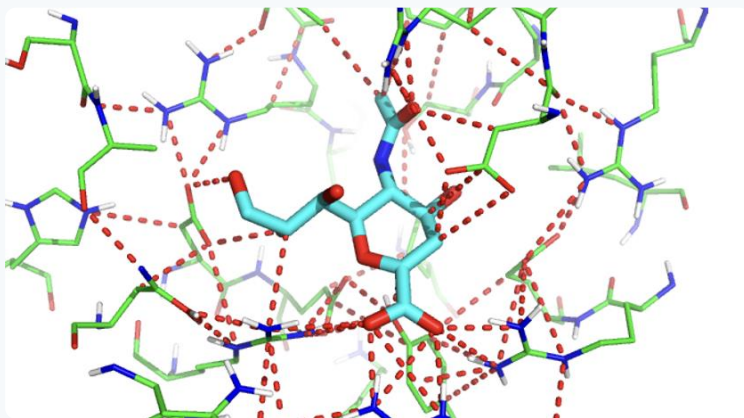


Additional interactions lead to tighter binding

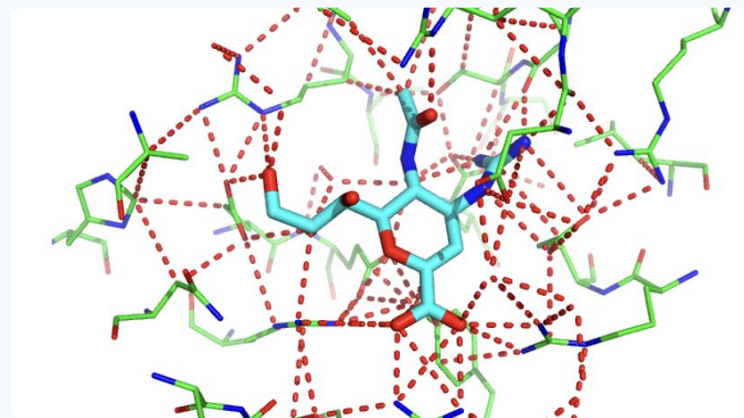


... not that simple

Ligand binding typically understood as the sum of protein-ligand interactions

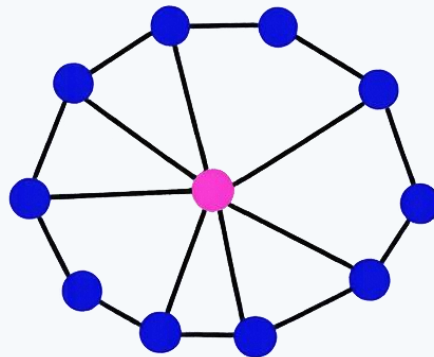
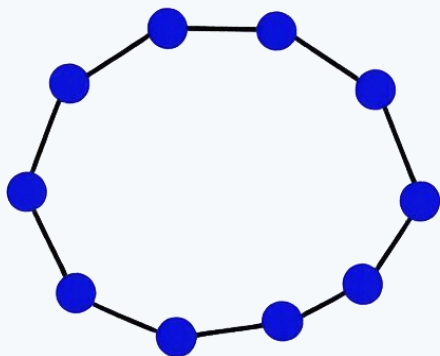


Additional interactions lead to additional network paths which can further stabilise the protein-ligand complex



... propose additional network paths lead to tighter binding

New concept: protein-ligand complex modelled as a small world network (SWN)

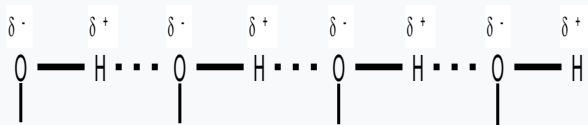


In a small world network popularity is attractive – this is new physics...

Addition of an extra node and just a few extra edges can reduce shortest path lengths between many pairs of nodes. We use network approach to capture cooperativity in protein-ligand complexes.

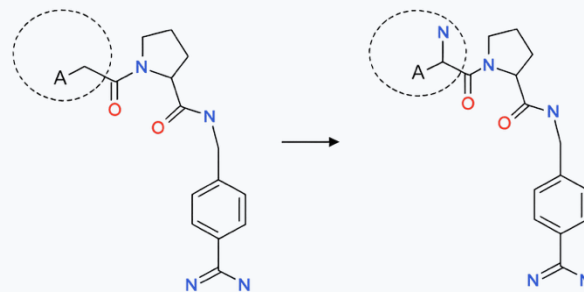
Types of cooperativity

Correlated H-bonds have lower free energy than sum of individual hydrogen bonds due to mutual polarization



Non-additivity of functional group contributions in a series of thrombin inhibitors. A hydrogen bond reinforces lipophilic interactions in the complex.

Baum et. al., J. Mol. Biol., 2010, 397, 1042

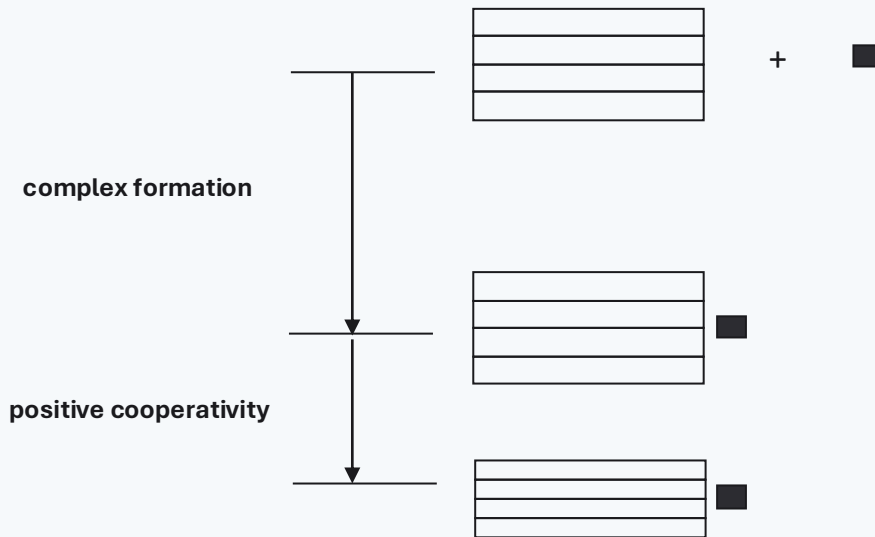


Types of cooperativity

The binding of biotin to streptavidin is 1000 times stronger than sum of the parts.

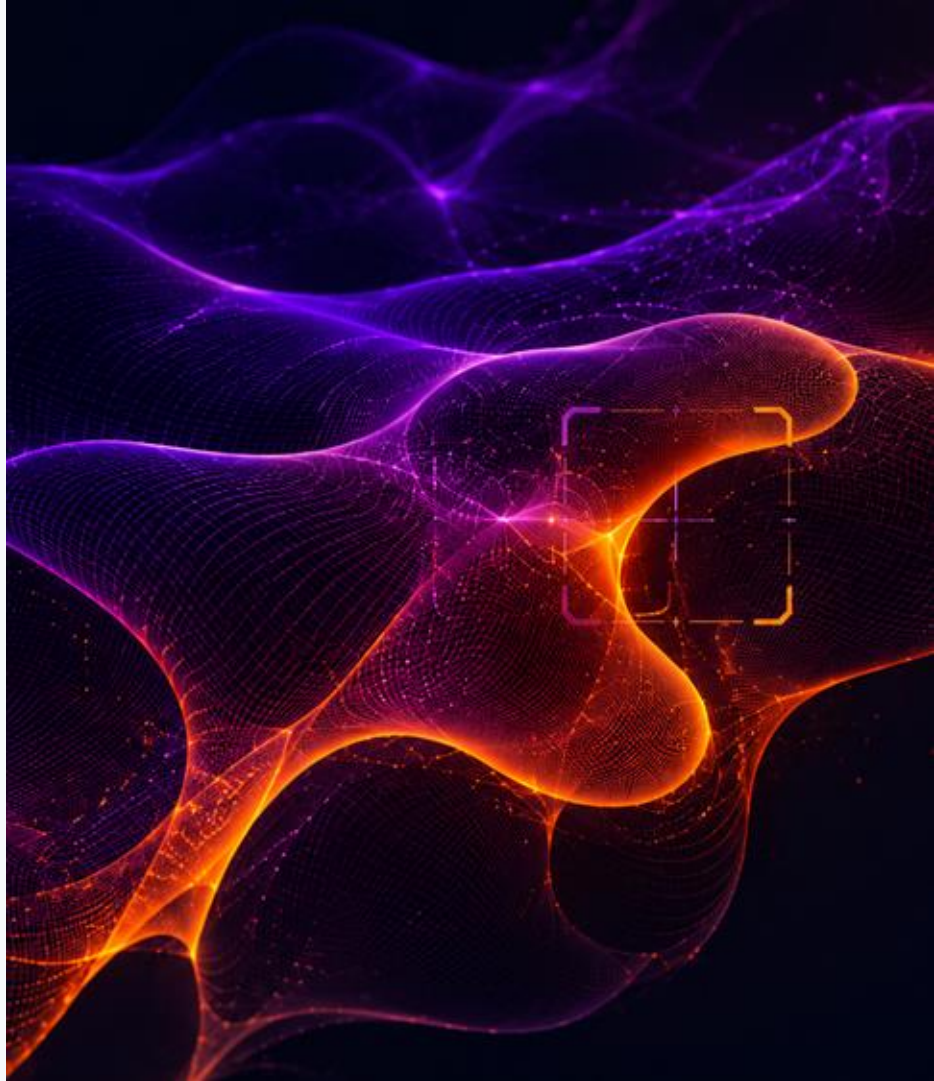
“Very large ligand binding energies ... derived by decreasing the lengths of numerous hydrogen bonds of a protein (upon binding a small molecule) by as little as about 1%.”

Williams et. al., Angew. Chem. Int. Ed, 2004, 43, 6596



Overview of approach: Scorpion

- Identification and classification of different types of favourable and unfavourable close contacts within protein-ligand binding sites.
- Combine all covalent and all favourable non-covalent interactions into a single network.
- Encode network paths containing ligand atoms into subgraph network descriptors.
- Define a reduced graph representation of protein structure.
- Parametrise using genetic algorithm based on high quality data sets.

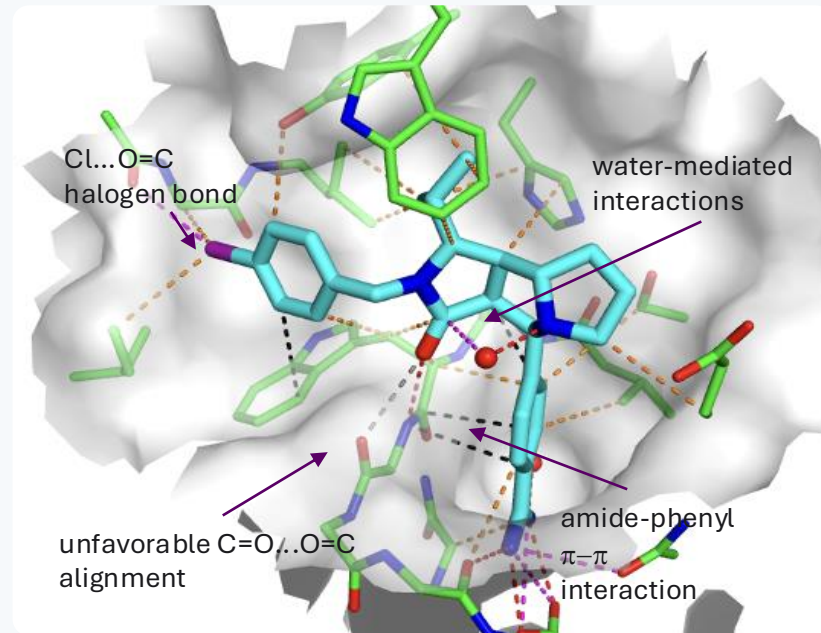


Network edges: classification of interactions

Raptor does automatic assignment of SMARTS-based atom types and then detection of different types of interactions (distance & angle constraints, line-of-sight test)

Classification of interaction types:

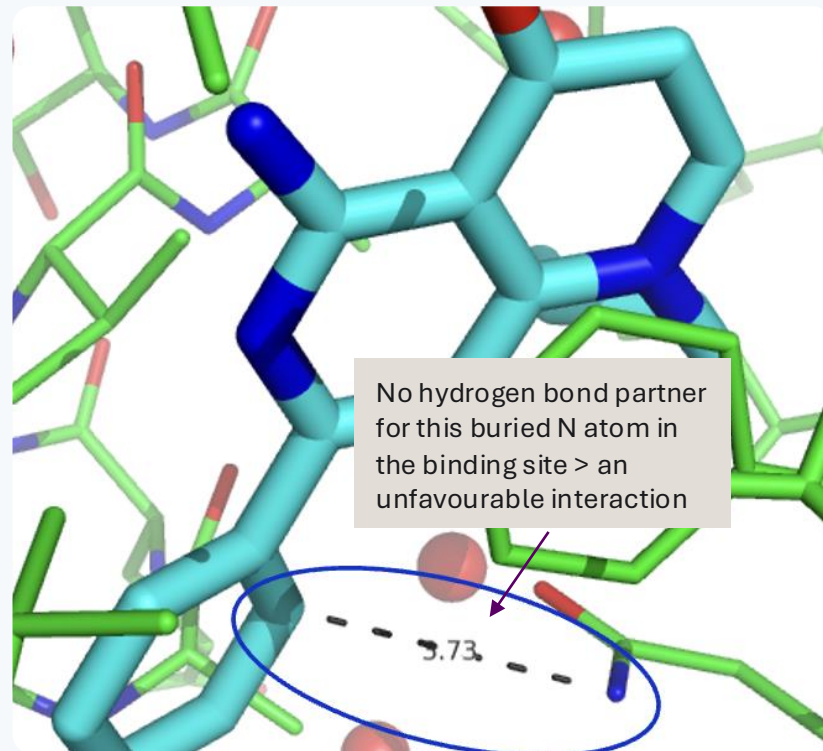
- | | |
|------------------------|---|
| 1. Hydrogen bond | 8. H donor-pi |
| 2. Metal | 9. Pi-pi |
| 3. Ionic | 10. vdW |
| 4. Cation-dipole | 11. Unfavourable of 1, 2, 3, 6 |
| 5. Cation-pi | 12. Polar and nonpolar clashes |
| 6. Dipolar | 13. polar-nonpolar contacts with likely desolvation penalties |
| 7. σ -hole bond | |



Raptor: Identification of unfavourable interactions

- Unfavourable contact if apolar ligand atom replaced by water molecule fulfills hydrogen bonding requirements.
- Distances and angles are checked for all polar/apolar close-contact pairs.
- Solvent exposure of each atom taken into account when testing for unfavorable contacts.

Allows for the detection of desolvation penalties that negatively affect target binding.



EX. 3HDZ (PDE-5)

Handling of water molecules

Scoring of water replacement potential based on deviation from ideal tetrahedral coordination of protein-bound water molecules.

$$Rank = \sum_n \left\{ (2.80 A / r_n) + \left[\sum_m \cos(\Theta_{Td} - \Theta_{ymn}) \right] / 6 \right\}$$

Rank scores:

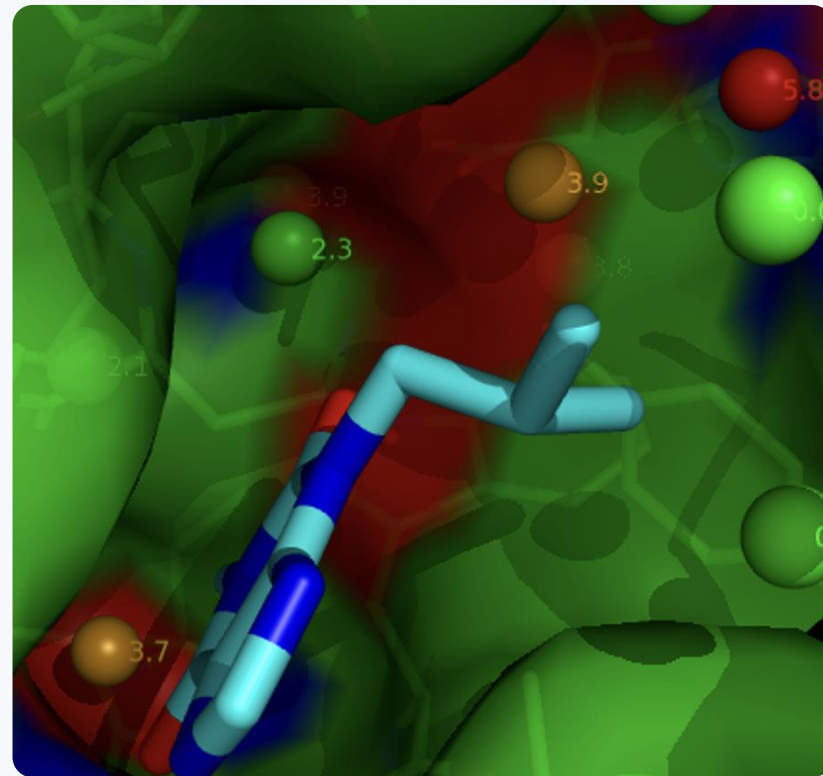
- 1 ideal H-bond: 1
- 2 ideal H-bonds: 2.3
- 3 ideal H-bonds: 4.0
- Tetrahedral coordination: 6.0

Maximum of 4 protein atoms is counted (≤ 2 donors and ≤ 2 acceptors). Any angle less than 60° is rejected.

Implemented into graphic analysis with color-coding of Rank score:

- **Green:** 0 - 2.3 (easy to replace)
- **Amber:** 2.3 - 4.0 (possible to replace with suitable polar functionalities)
- **Red:** 4.0 - 6.0 (unlikely to replace)

Amadasi et. al., J. Med. Chem., 2008, 51, 1063

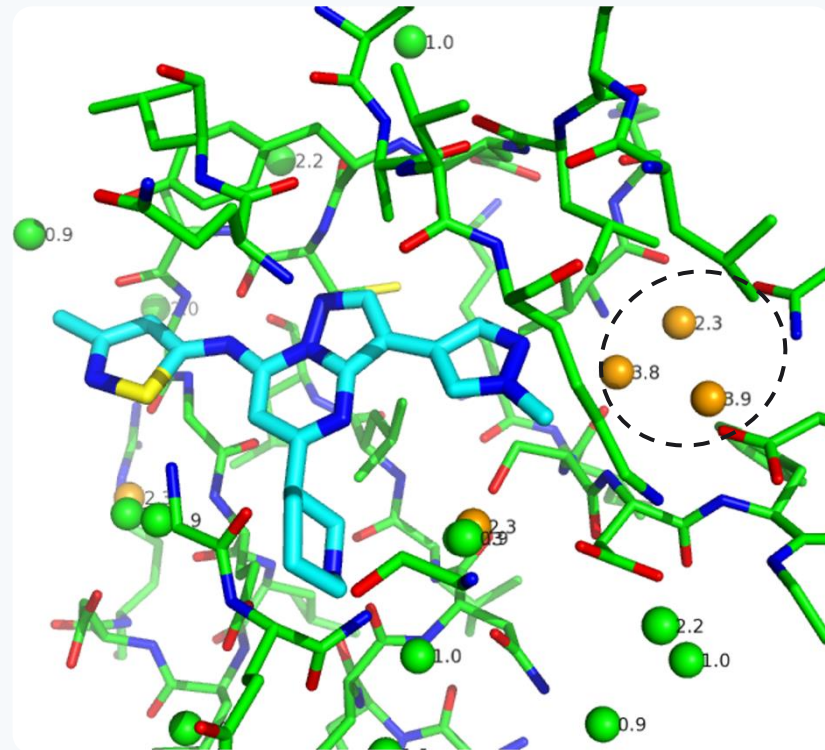


EX. 2R8Q (PDE-B1)

Water molecules with Rank scores ≥ 2.0 are included in networks.

Waters in CHK1 kinase

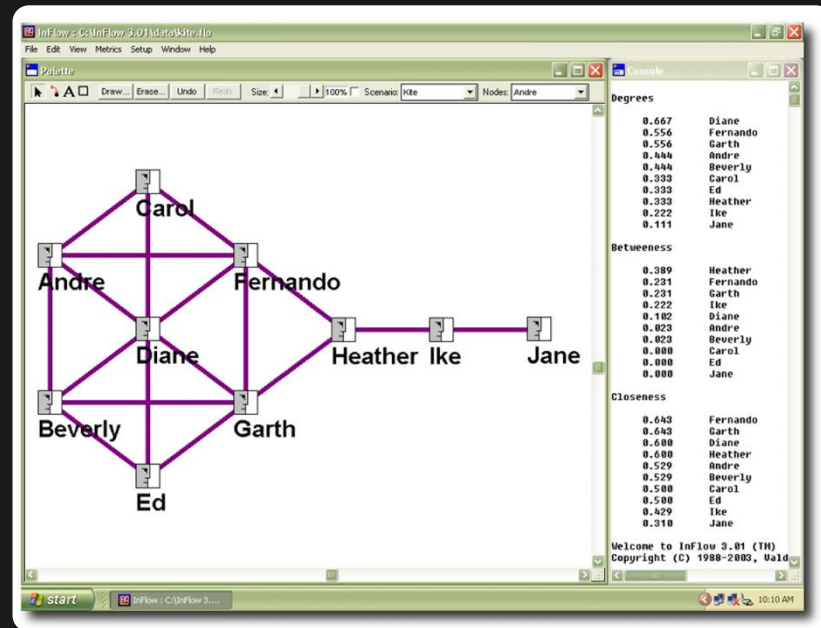
Three water molecules in CHK1, in a cavity adjacent to ligand, difficult to displace. WaterRank Score shows that the 'front' two waters have reasonably high scores.



3OT8

Standard small world network (SWN) model

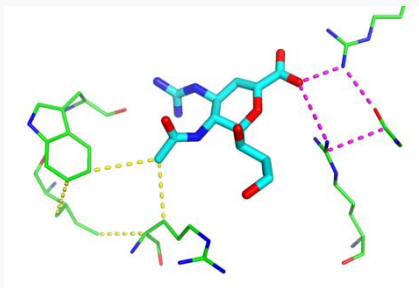
Initially explored using descriptors from Social Network Analysis. In our domain, these descriptors are too sensitive to individual contacts, and to geometric constraints associated with maximum number of contacts



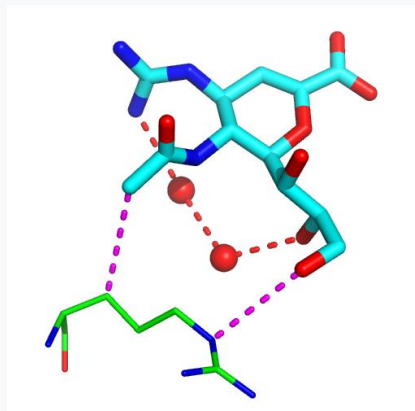
Kite Network, by D. Krackhardt
<http://www.orgnet.com/sna.html>

Network descriptors: paths involving ligand atoms

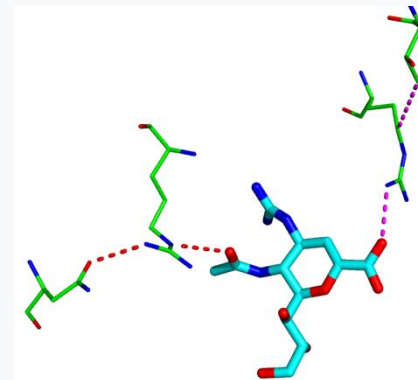
Ligand-protein-ligand (LPL) network elements: Ligcycles (involving 1 ligand atom) .



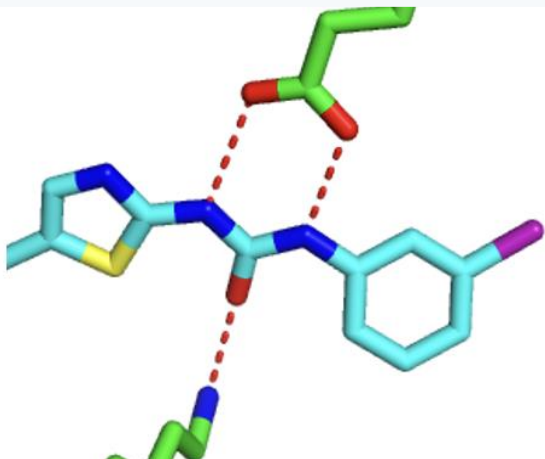
Ligloops (involving ≥ 2 ligand atoms) examples from 1nnc.



Ligand-protein-protein (LPP) network elements: Ligpaths (subsets of long ligcycles/ligloops > 8)

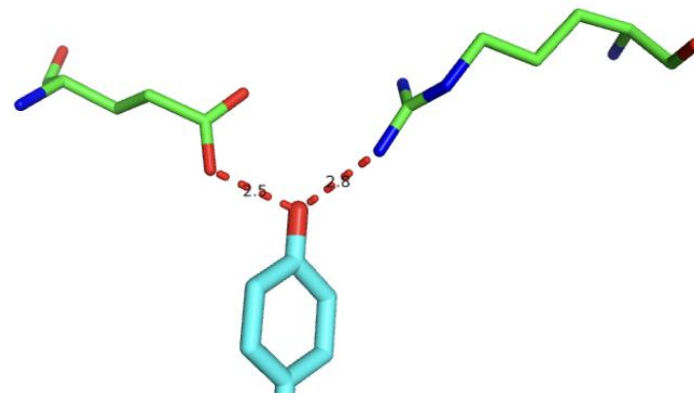


Network descriptors: special treatment of hydrogen bonding



Privileged pairs of hydrogen bonds

Arrangements of hydrogen bonds that can not be achieved in the apo state .



Protein-ligand-protein (PLP)

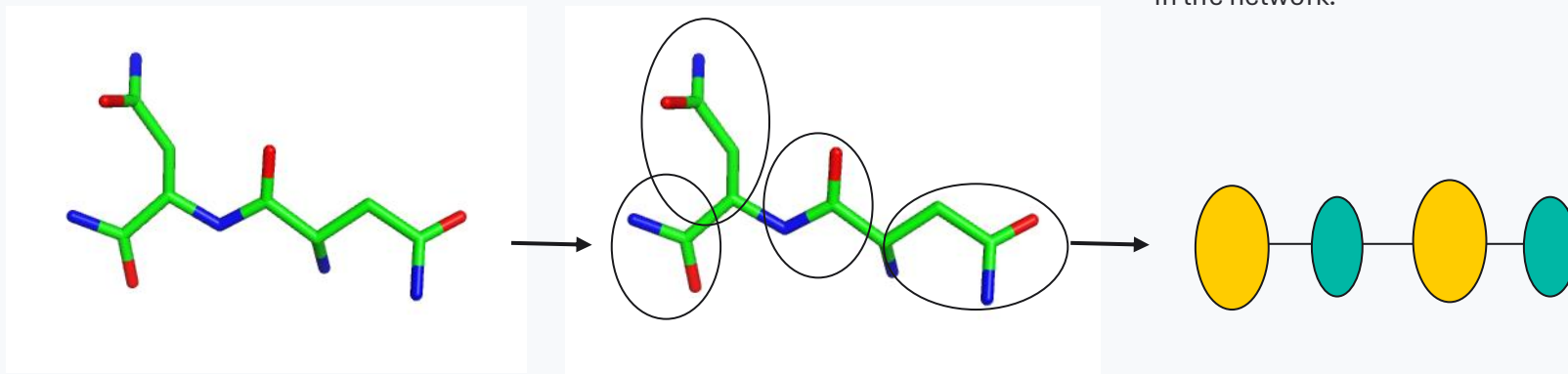
With lower free energy than the sum of the individual bonds due to mutual polarization.

Network descriptors: nodes based on a reduced graph definition of protein structure

For our network study the all-atom based approach was too fine-grained and the residue approach too coarse.

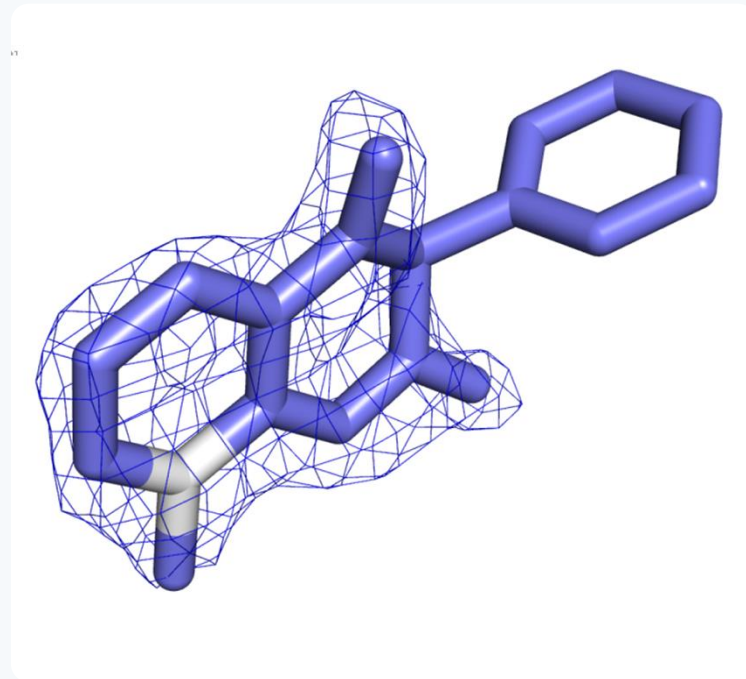
A protein structure is treated as a collection of small groups of atoms (functional groups).

The functional groups of sidechains - rings, acid groups, etc - and backbone amide bonds, treated as single nodes in the network.



Stringent quality criteria for training sets

- X-ray structure with crystallographic resolution ≤ 2.5 Å.
- Successful match of ligand topology (best Proasis ligand quality).
- Noncovalent binding between ligand and protein.
- No symmetry contacts
- No alternative conformations.
- No clashes.
- No missing atoms.
- No broken residues.
- Minimum occupancy = 1.0.
- Minimum real space correlation coefficient ≥ 0.7 .
- Ligand strain energy ≤ 8 kcal/mol
- Drug/lead-like ligands.
- Binding data available (K_i , K_d , IC_{50}) and measured with same assay.



Electron density correlation coefficient is a better measure of model quality than B-factors.

Training sets: high quality structures with binding affinity data

I. Hard set: 28 compounds:

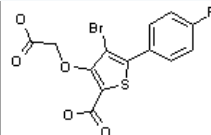
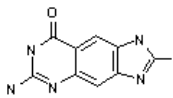
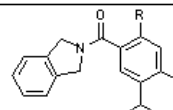
- Activity cliff pairs.

II. 31 neuraminidase complexes.

III. 46 PDE10 complexes.

IV. 7 subsets with up to 10 structures each:

- IRAK4, BTK, HCV polymerase, HIV protease, DPP-4, PKACA, LCK.

4	protein tyrosine phosphatase 1B		-OH	2h4g	6.5
			-H	2h4k	5.5
5	tma-guanine transglycosylase		-NH ₂	2z7k	7.1
			-CH ₃	3c2y	5.8
6	hsp90		-OH	2xab	9.3
			-H	model	7.2

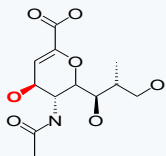
Global optimisation

- Based on high quality structures and results from docking.
- Optimisation used genetic algorithm approach.
- Form of scoring function:
- A particular protein-ligand interaction considered networked if [weighted] sum of network elements higher than an interaction-specific threshold.

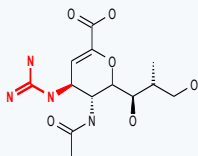
$$S = \sum_n f(Int) \quad (\text{without network terms})$$

$$S_{Scorpion} = \sum_n f(Int) + \sum_m g(Int_nw)$$

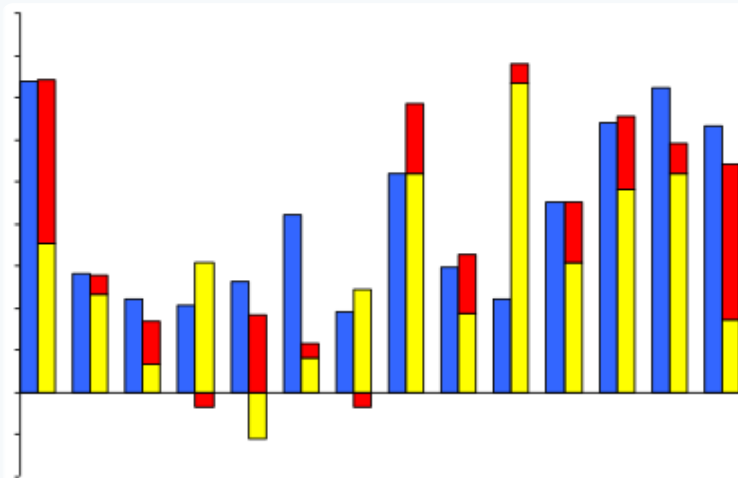
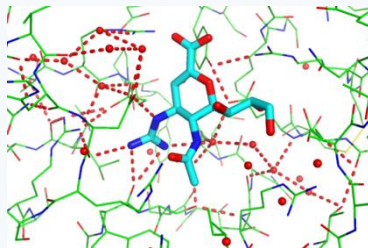
Activity cliffs: predicted vs. experimental energy differences



1nnb
IC₅₀: 5uM
S_{Scorpion}: 6.4+2.2



1nnc
IC₅₀: 1nM
S_{Scorpion}: 8.1+4.2



Expt

Pred (non-network comp.)

Pred (network comp.)

Scorpion Score

Results shown from optimisation done in 2010 - scoring function optimisation is on-going, we continue to improve our results.

$$\begin{aligned} S_{\text{Scorpion}} = & 0.473 \times [\text{hbond}] + 0.129 \times [\text{hbond_nw}] \\ & + 0.516 \times [\text{vdw}] + 0.387 \times [\text{vdw_nw}] \\ & + 0.188 \times [\text{pi - pi}] + 0.931 \times [\text{pi - pi_nw}] \\ & + 0.285 \times [\text{cat - dipole}] + 0.606 \times [\text{cat - pi}] + 0.65 \times [\text{halogen}] \\ & - 0.387 \times [\text{unf_hbond}] - 0.899 \times [\text{unf_desolv}] - 1.146 \times [\text{unf_clash}] - 1.501 \times [\text{unf_ionic}] \end{aligned}$$

H-bond donor – nonpolar contact

External validation

We do not attach too much weight to these results – public affinity data sets are flawed and our focus is to better understand tight binding.

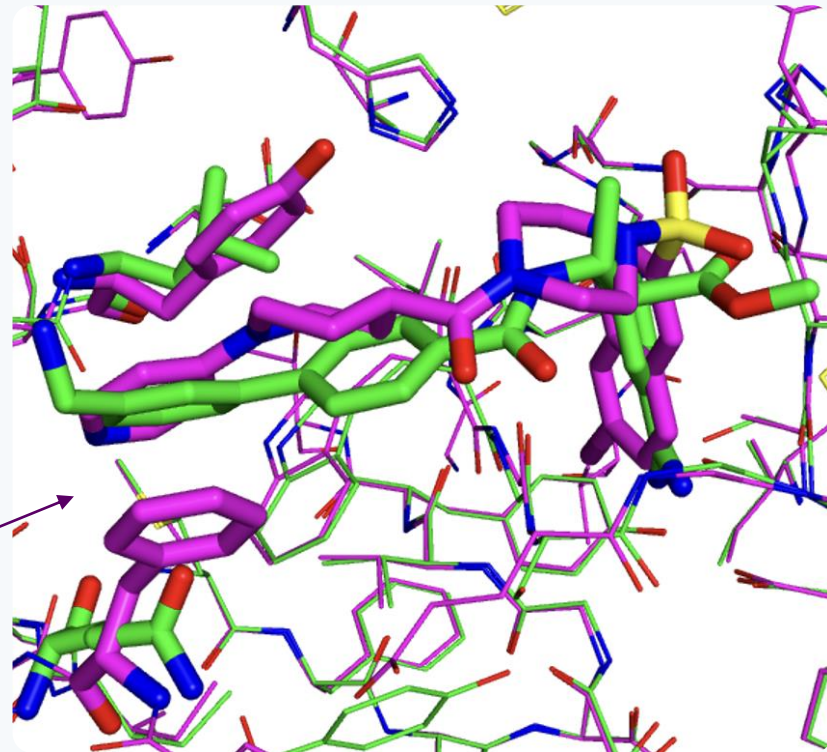
P. Englebienne, N. Moitessier, J. Chem. Inf. Model. 2009, 49, 1568–1580

HIV protease (11)	Thrombin (22)	Trypsin (13)	Thr/Try/FXa (42)	Average	
RankScore	0.55	0.68	0.36	0.61	0.55
XScore	0.73	0.31	0.44	0.41	0.47
DrugScoreCSD	0.55	0.42	0.46	0.44	0.47
PLP1 (Cerius2)	0.59	0.48	0.31	0.49	0.47
Scorpion Affinity	0.59	0.49	0.23	0.51	0.46
DockScore (Sybyl)	0.62	0.37	0.39	0.46	0.46
DrugScorePDB	0.48	0.56	0.18	0.56	0.44
PLP2 (Cerius2)	0.59	0.39	0.31	0.45	0.43
GoldScore (Sybyl)	0.44	0.55	0.18	0.50	0.42
LigScore2 (Cerius2)	0.51	0.45	0.25	0.42	0.41
Hammerhead (Cerius2)	0.48	0.30	0.49	0.31	0.40
PMF (Cerius2)	0.55	0.25	0.36	0.35	0.38
ChemScore (Sybyl)	0.62	0.06	0.56	0.27	0.38
GlideScore	0.51	0.31	0.31	0.33	0.37
LigScore1 (Cerius2)	0.44	0.43	0.18	0.37	0.35
MW	0.26	0.49	0.13	0.46	0.33
eHiTS SF	0.37	0.46	0.08	0.39	0.32
Surflex SF	0.48	0.18	0.31	0.30	0.32
FlexXScore (Sybyl)	0.66	-0.02	0.39	0.08	0.27
PMF (Sybyl)	-0.18	0.29	-0.03	0.30	0.10

External validation sets for binding affinity prediction are flawed

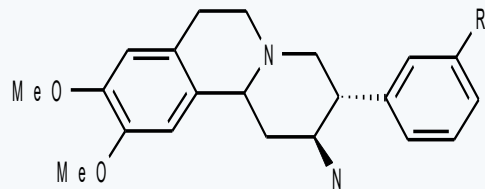
- Example: Trypsin subset (13 structures) -> only one would pass our quality criteria (1f0u) .
- 1f0u (green) and 1v2k (magenta) both in trypsin set despite large differences in S4 pocket.

1v2k: FXa S4 pocket due to mutations

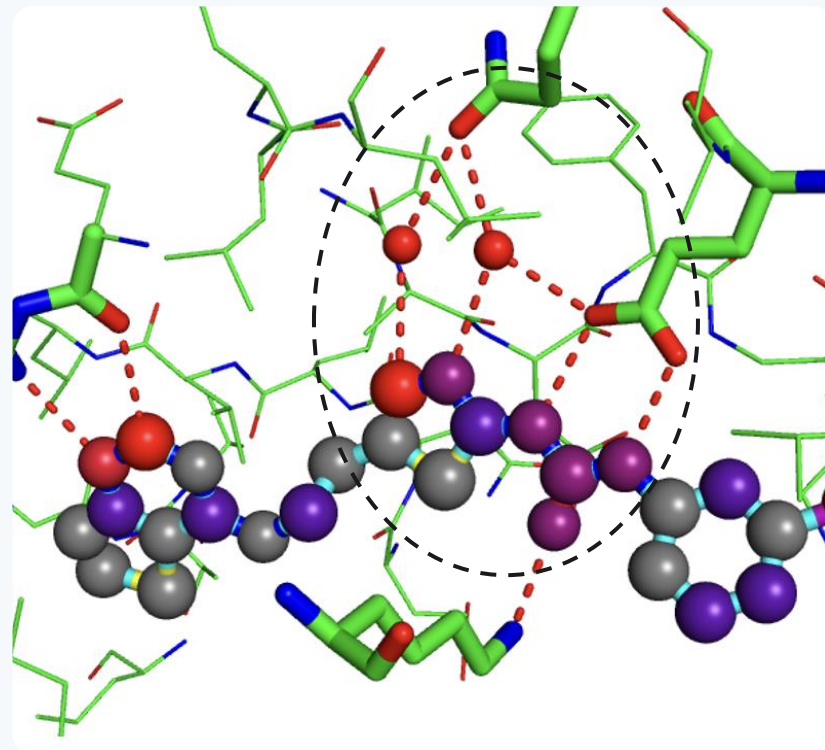


Score contributions mapped onto atoms

- Color ramp from blue -> red
- Fray = no score contribution



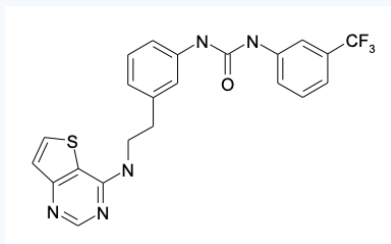
R	K_d (nM)	$S_{Scorpion}$	network
H	200	7.9	0.8
CH ₃	4.6	10.8	1.6
CH ₂ F	0.5	11.3	1.7



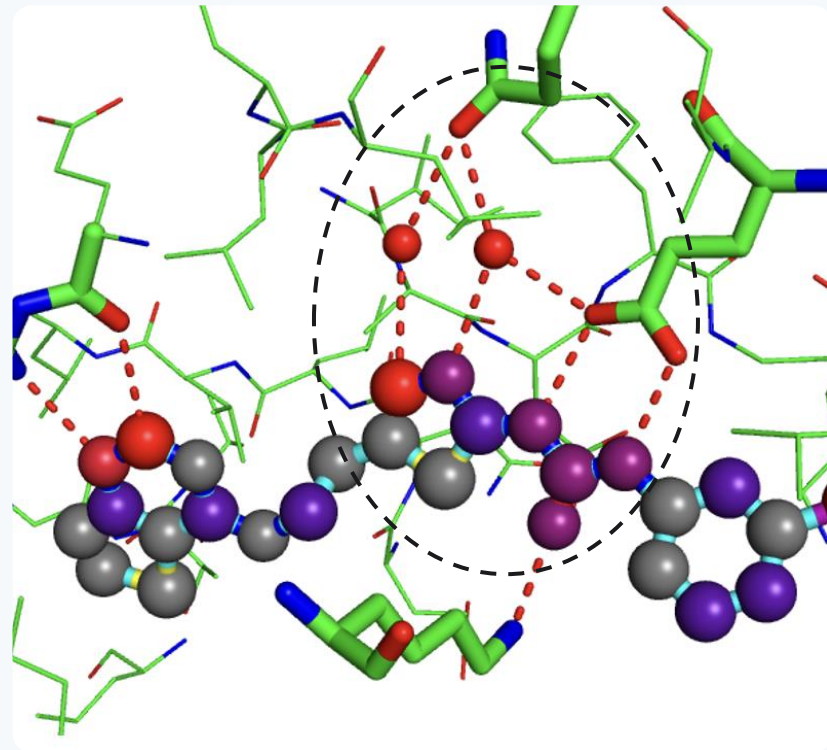
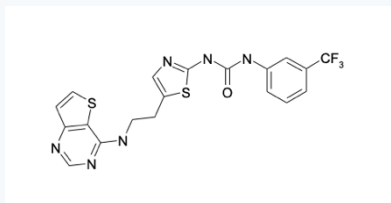
3KWJ (DPP-4)

Aurora A kinase inhibitors

IC₅₀: >20'000 nM



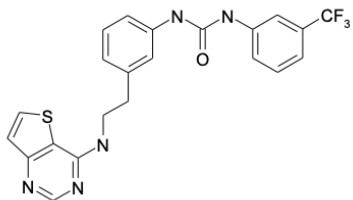
IC₅₀: 22 nM



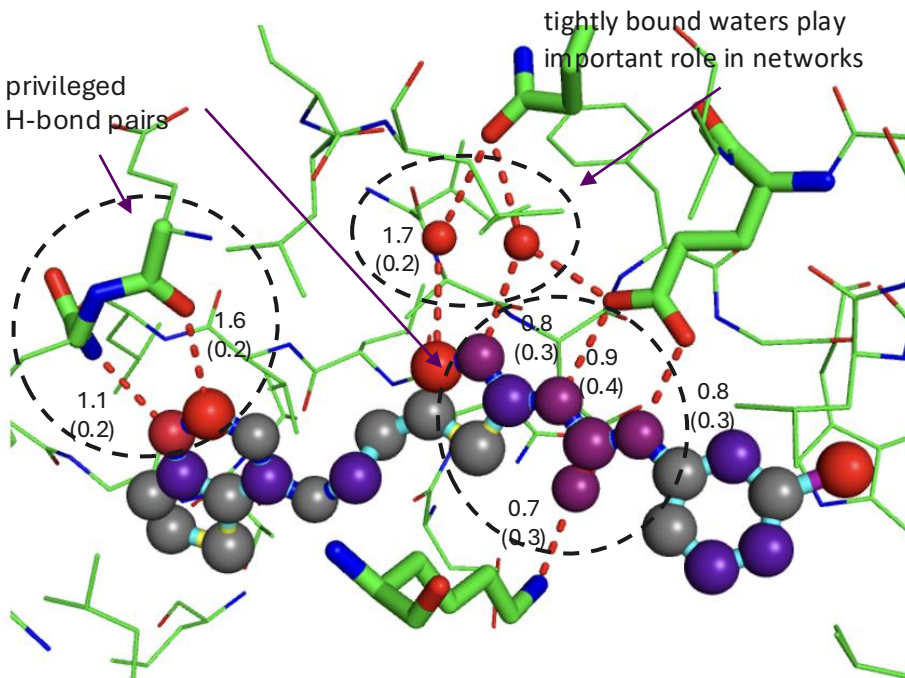
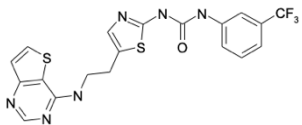
Networked H-bonds with high score incl. network contribution

Aurora A kinase inhibitors (cont.)

IC₅₀: >20'000 nM



IC₅₀: 22 nM

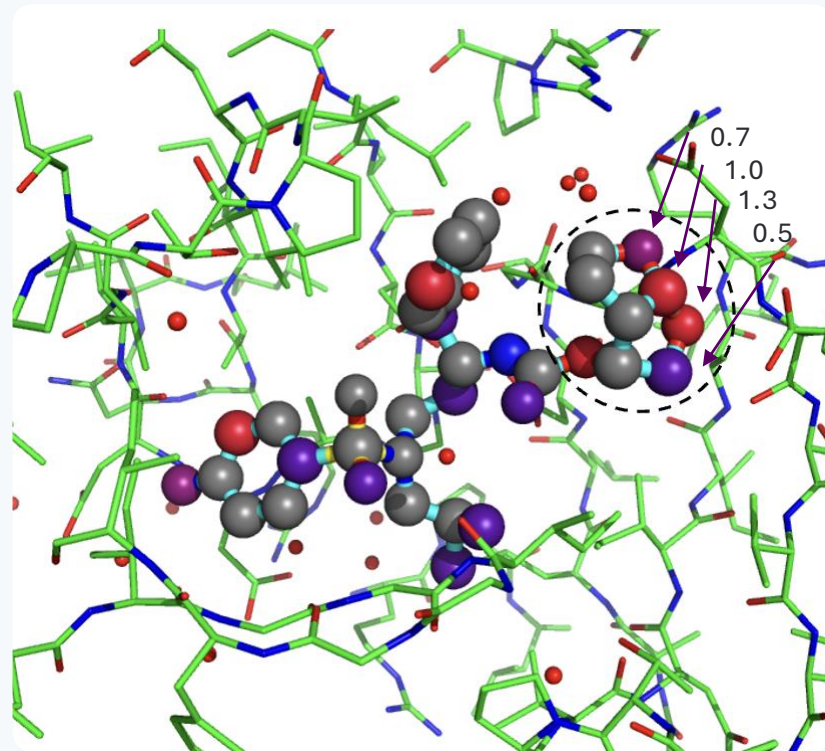
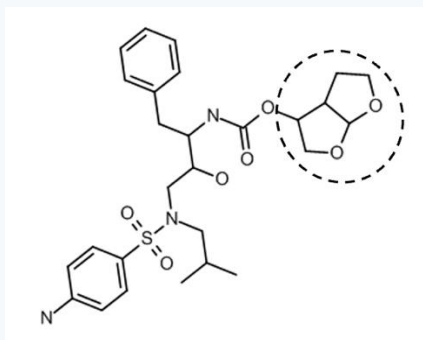


HIV protease inhibitor - darunavir

Bis-THF displaces two water molecules associated with backbone NH's and forms hydrogen bonds and vdW contacts with the protein in optimal way.

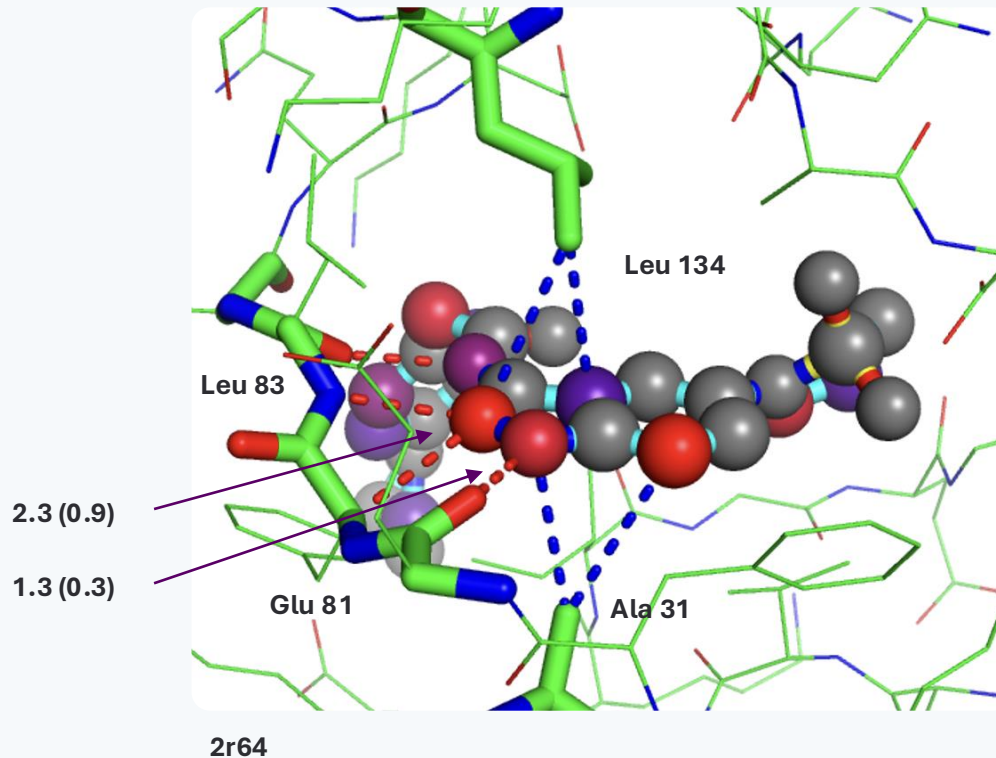
Exp: pIC50 ~ 11

Calc: pIC50 ~ 10



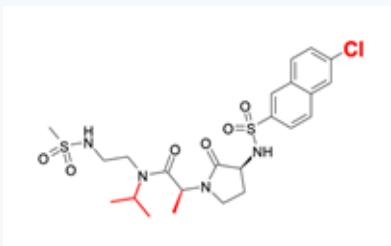
Aminoindazole in CDK2

Aromatic heterocycle with 1-2 H-bonds and in a hydrophobic environment (e.g. kinase hinge binding)



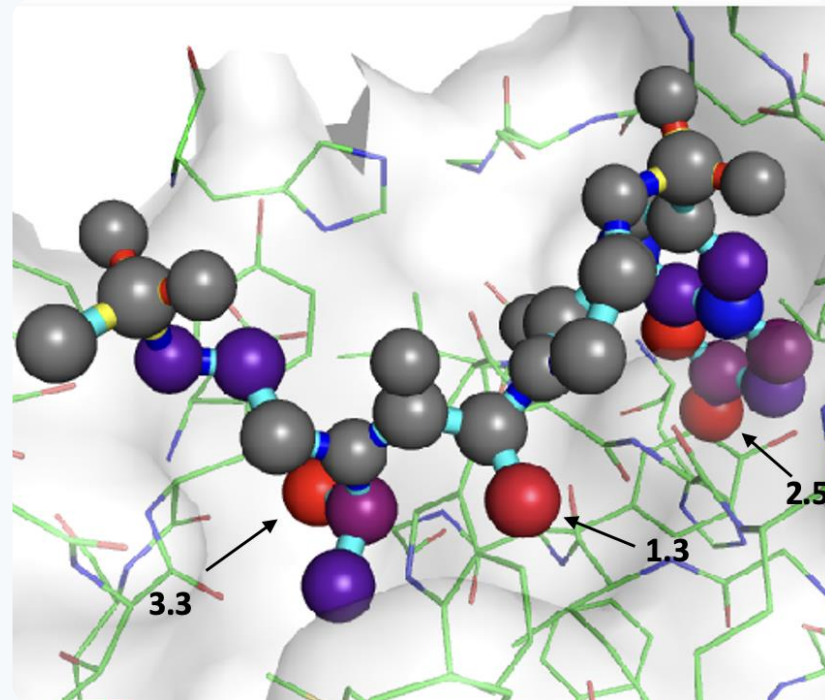
Examples of highly networked atoms

Atoms in buried pockets with several contacts receive extra network contribution.



Other examples:

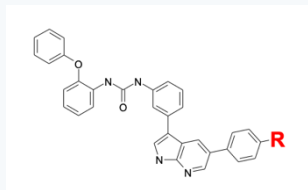
- 1yvz (hcv polymerase): Cl
- 1ql7 (trypsin): Cl
- 2r3r (cdk2): Br



2J4I (FACTOR XA)

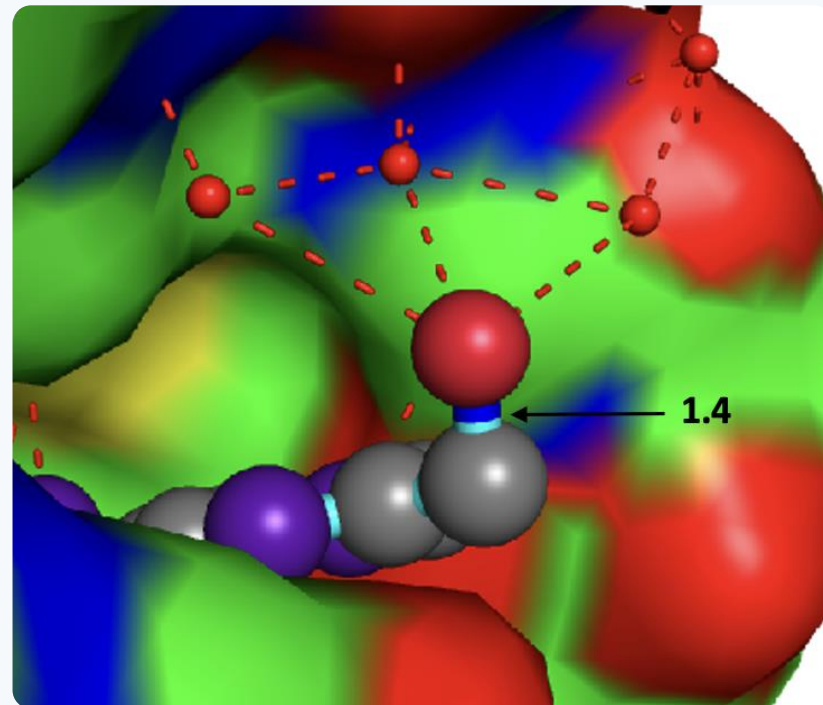
Insulin receptor kinase – pyrrolopyridine complex

Ligand atoms can have high network scores in spite of being highly solvent-exposed.



R:	IC50[nM]
-H	127
-CH ₂ NH ₂	14

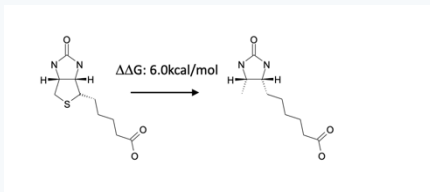
- Aminomethyl group solvent exposed with no direct contact with protein.
- Amino group interacts through protein-bound water molecules with insulin receptor resulting in a high score despite low buriedness.



3ETA

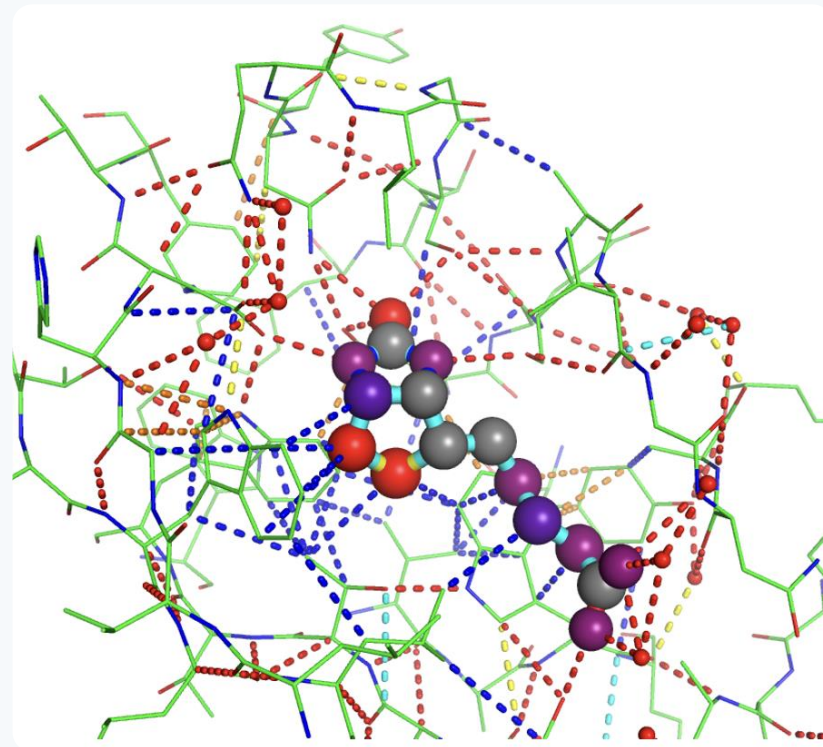
Streptavidin - biotin

- Femtomolar binding affinity, not explainable with standard methods.
- Experimental evidence for tighter packing in complex - reduced H/D exchange.
- High Scorpion scores for S (4.9), adjacent C (2.1) and carbonyl O (1.7) atoms, unusually high network contribution for S atom (3.4).

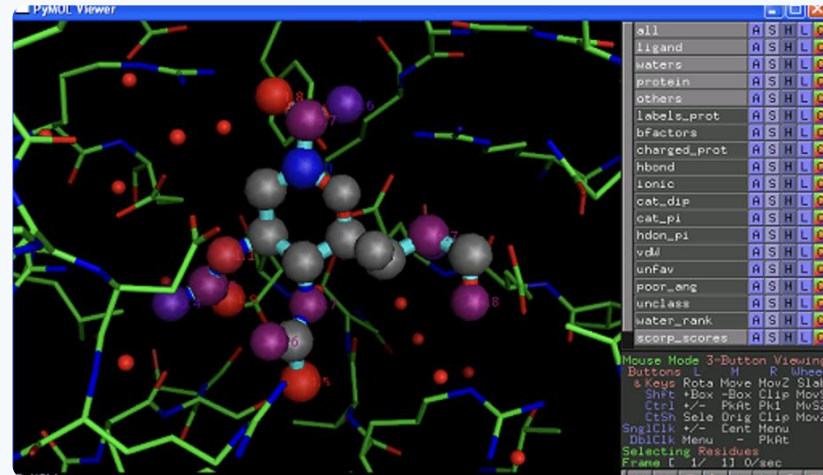
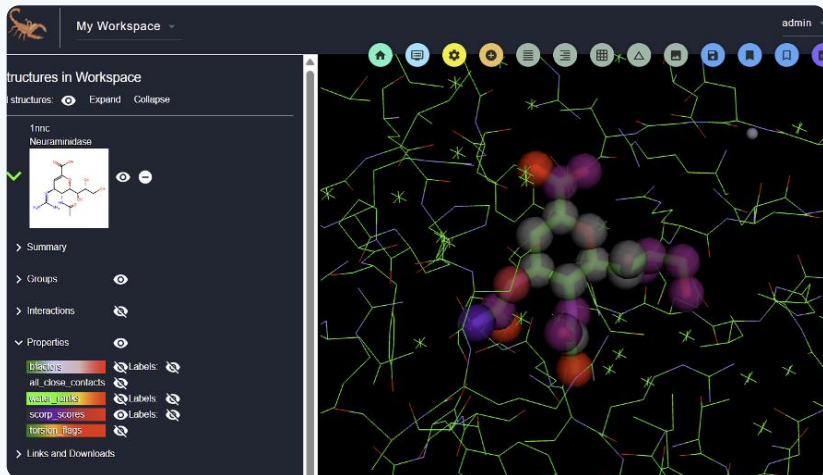


“The streptavidin/biotin system provides a clear example where the binding affinity is the property of the whole system.”

Williams et. al., Angew. Chem. Int. Ed, 2004, 43, 6596

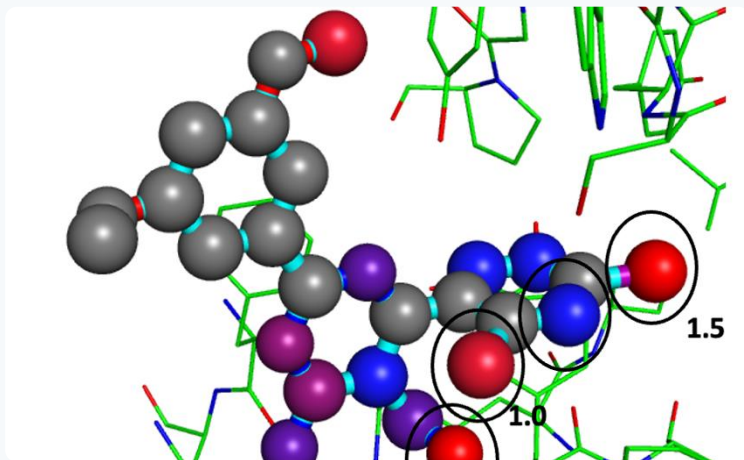


Running Scorpion: using command line tool, Proasis system, or API resources



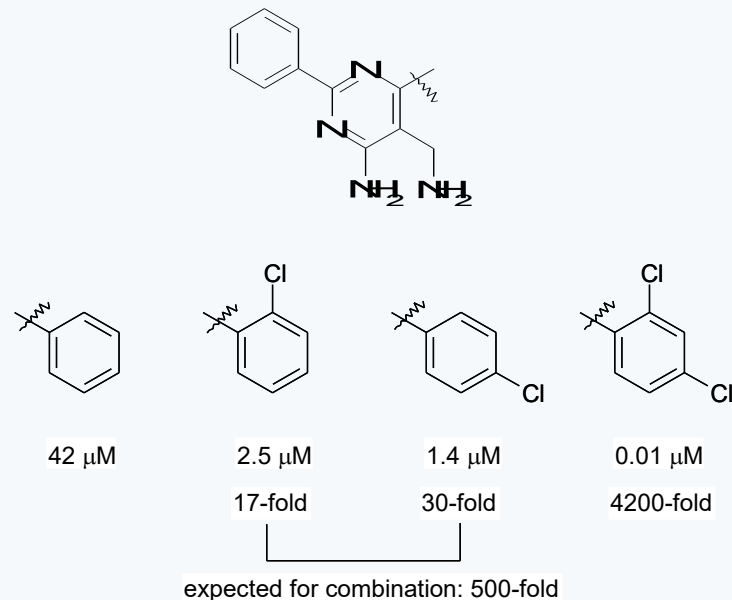
Proasis also enables export directly into PyMol

Cooperativity pairs often the key to tight binding



1RWQ (DPP4)

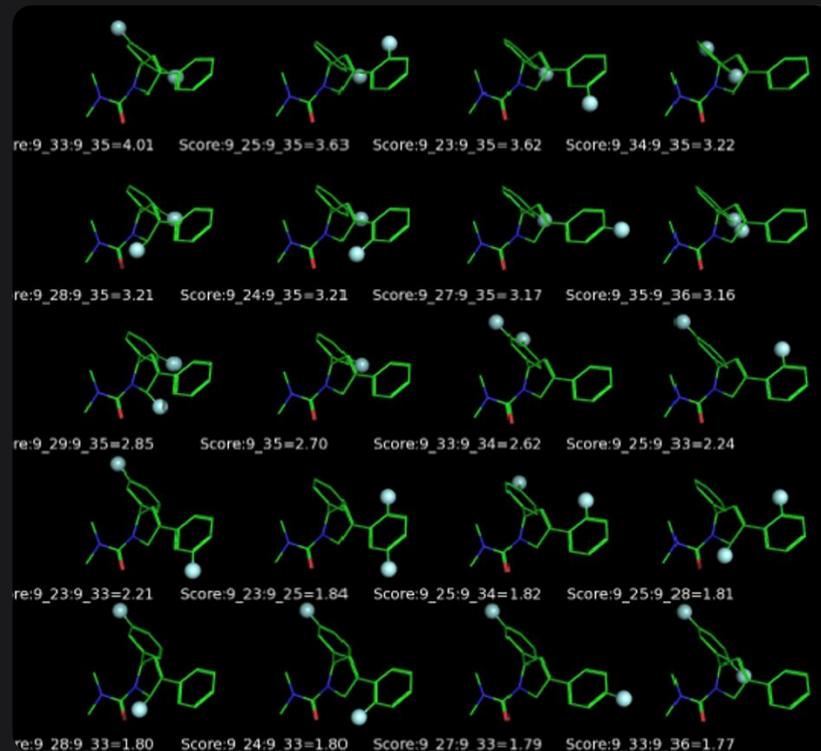
Favorable network motifs involving two different ligand substituents can provide the tipping point in lead optimisation.



Circled atoms identified as potential cooperativity partners [A,B] - high networkedness of A and B with protein and LPL link from A to B.

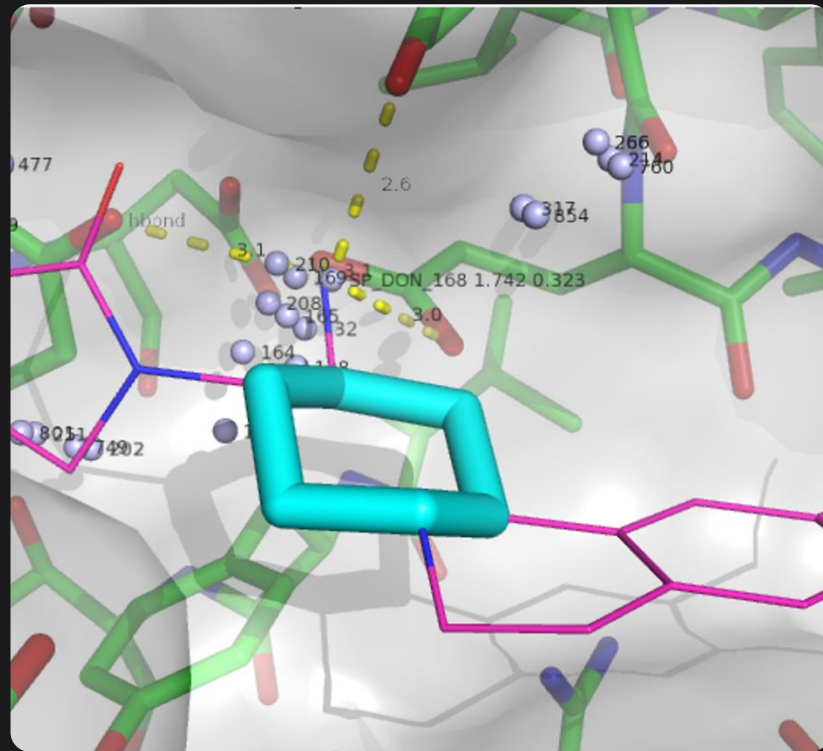
Ligand design using Viper: atom scan

- Command line tool identifying explicit atom substituents that lead to high network interaction scores.
- Enables single, double, and triple substitution, for any C-H, by a halide or any other element.
- Ranks according to Scorpion.



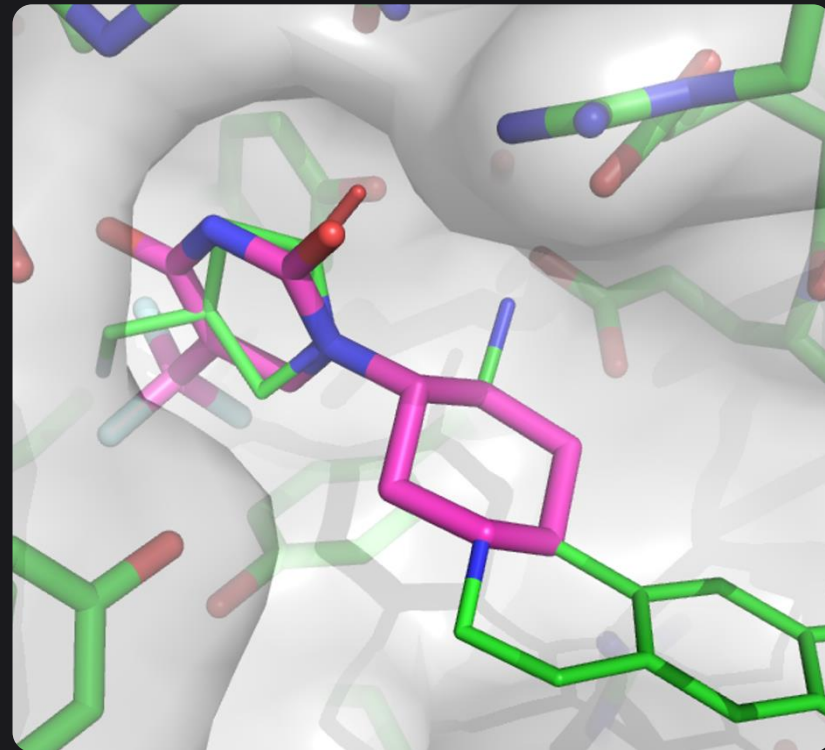
Ligand design using Viper: Hotspot search

- Command line tool highlighting pharmacophores for strong protein-ligand interaction networks.
- Grid-like sampling of binding site.
- Uses combinations of ViewContacts atom types.
- Hotspots, including Scorpion score, network score, and contacts involved, all easily viewed in PyMol.
- Also identifies favourable water binding sites.



Ligand design using Viper: fragment scan

- Command line tool that extends a ligand with small optimised substituents.
- Starts with 3D fragment library, such as BRICS, and guided by Scorpion hotspots.
- Sample fragments, taking account of chemistry, torsion strain, clashes, binding affinity.
- Generates suggestions for ligand substituents that provide high network interaction scores.

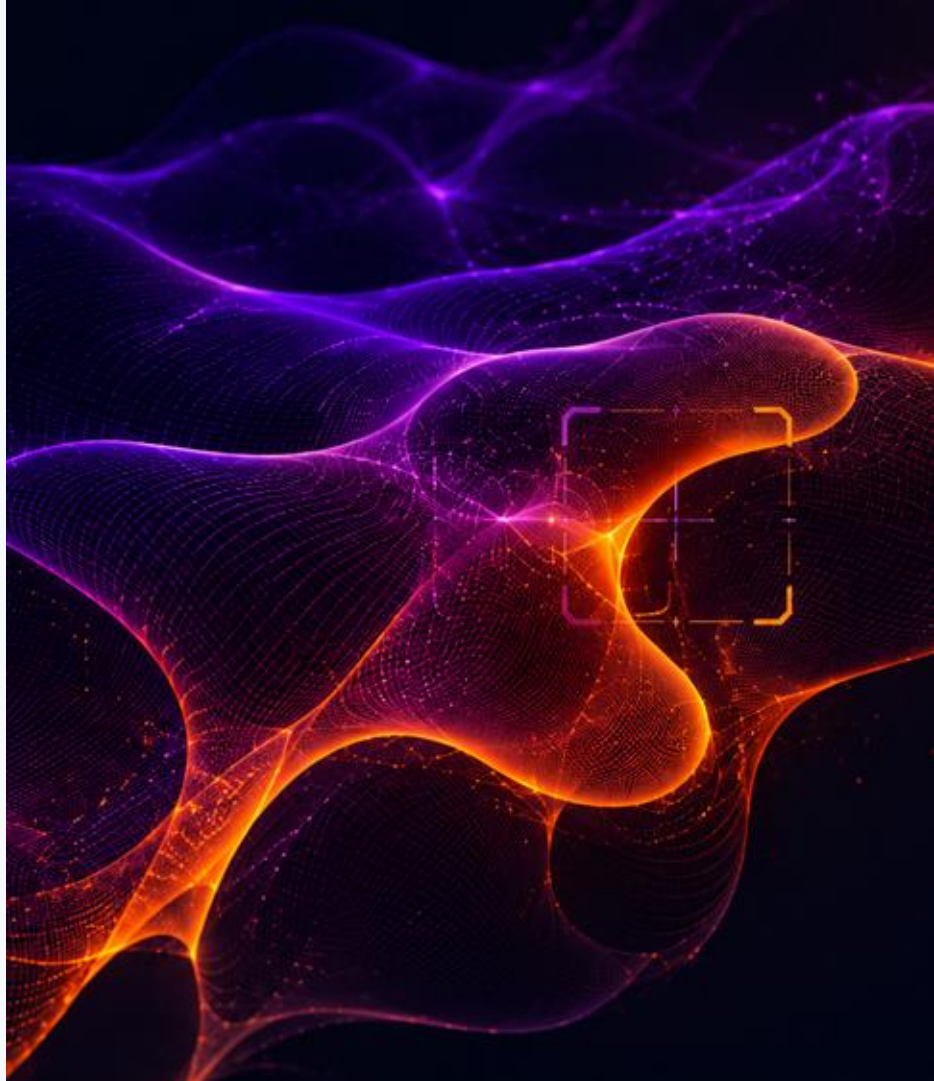


Acknowledgements

Scorpion project is a joint venture between
Desert Scientific Software and
F. Hoffmann-La Roche:

- Bernd Kuhn.
- Martin Stahl.
- Wolfgang Guba, and other molecular modeling staff from Basel.
- Michael Reutlinger and Julian Fuchs.

"Rationalizing Tight Ligand Binding Through Cooperative Networks" J. Chem. Inf. Model., 2011, 51 (12), pp 3180–3198 [Rationalizing Tight Ligand Binding through Cooperative Interaction Networks](#) .





Get in touch

CONTACT US

info@desertsci.com
www.desertsci.com

From overseas Tel: 1300 442 500
Australia Tel: 02 7262 6793

OUR LOCATION

Desert Scientific Software
Suite 8, 7 Sefton Road
Thornleigh NSW 2120