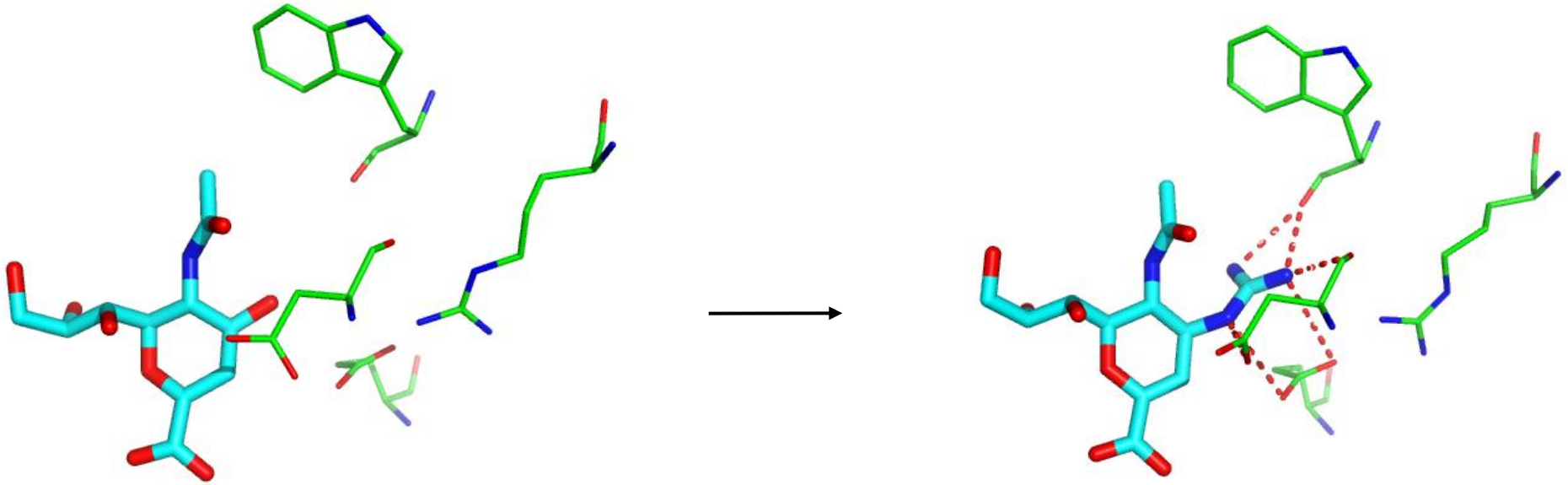
Using Cooperative Networks to Better Understand Tight Binding Ligands

Neil R. Taylor

Desert Scientific Software

Sydney, Australia

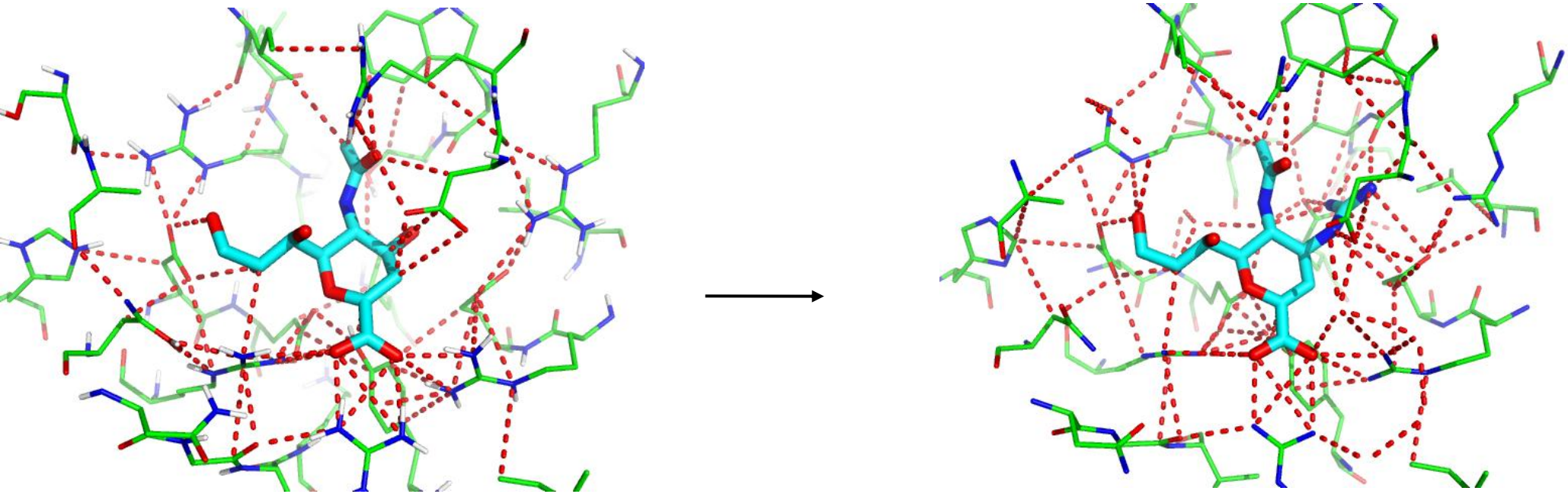
Ligand binding typically understood as the sum of protein-ligand interactions



Additional interactions lead to tighter binding

... not that simple

Beyond the pairwise additive view of protein-ligand interactions

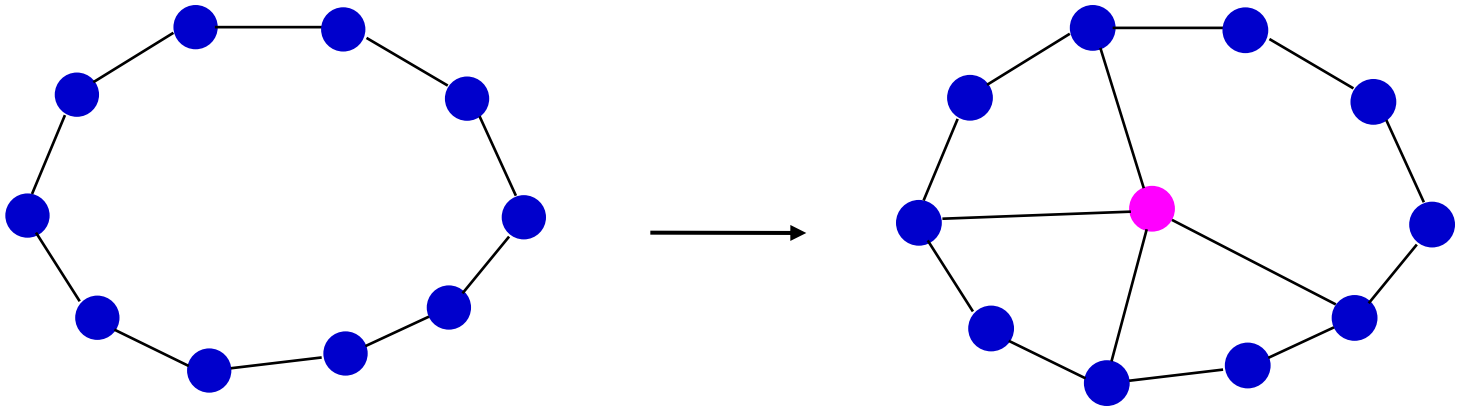


Additional interactions lead to additional *network paths* which can further stabilise the protein-ligand complex

... propose additional network paths lead to tighter binding

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

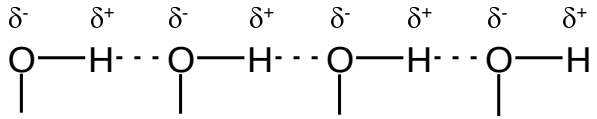
In a small world netwotk popularity is attractive – this is new physics



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

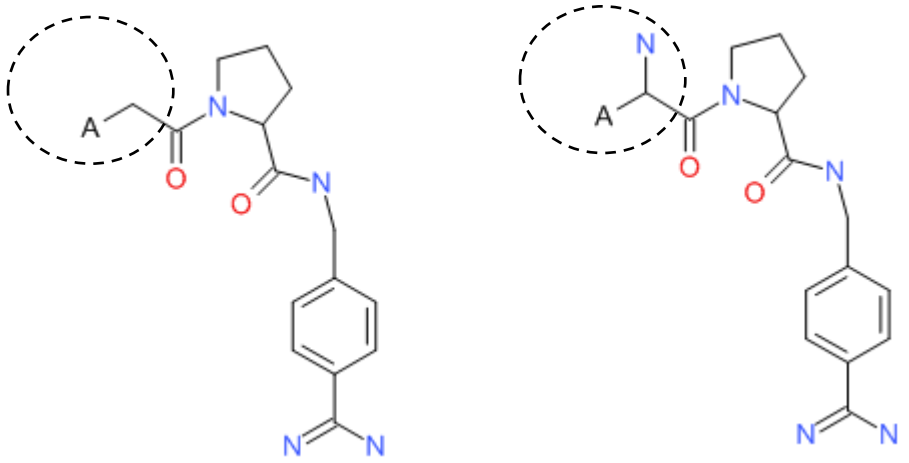
Types of cooperativity



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

Types of cooperativity (cont)

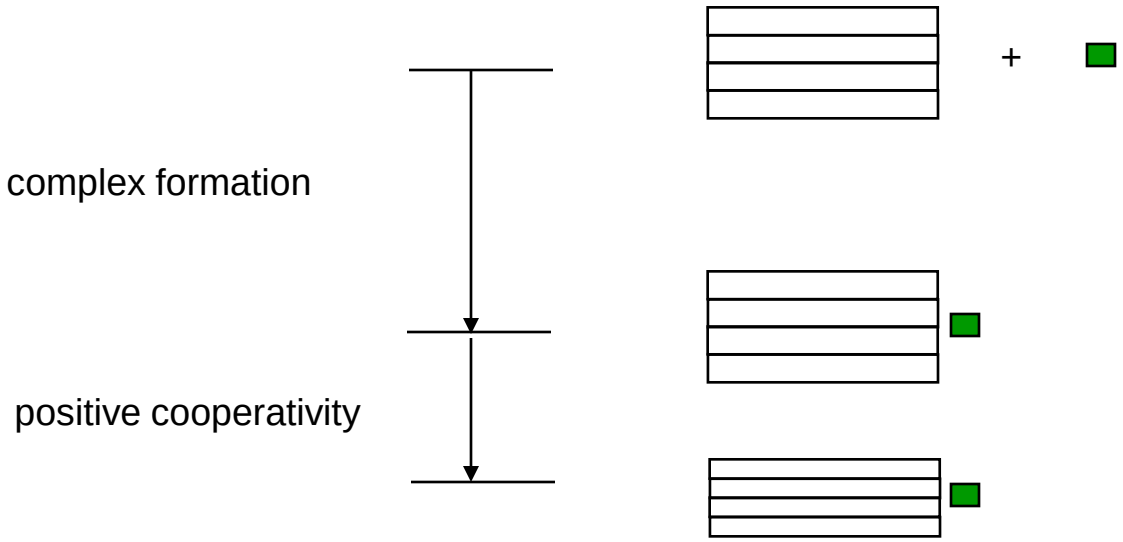
Non-additivity of functional group contributions in a series of thrombin inhibitors



A hydrogen bond reinforces lipophilic interactions in the complex

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

Types of cooperativity (cont)



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 hydogen 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

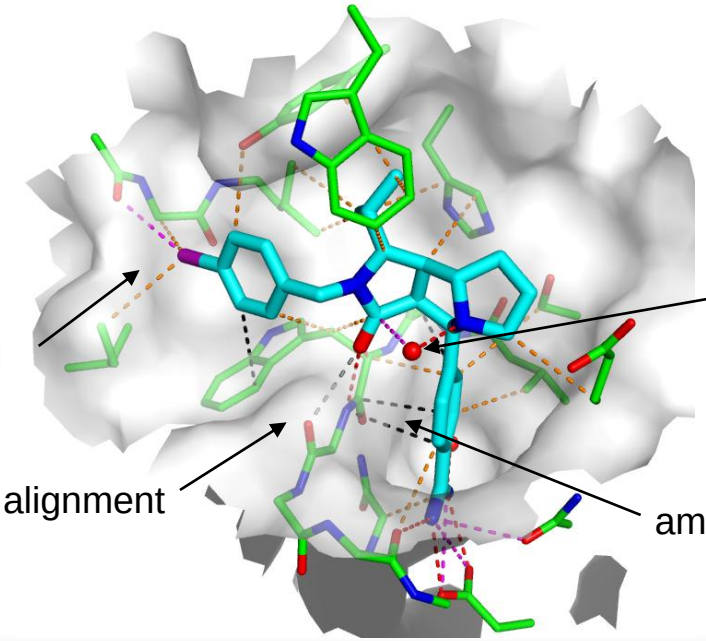
- Identification and classification of different types of favourable and unfavourable close contacts within protein-ligand binding sites
- Combine all covalent and all favourable non-covalent interactions into a single network
- Encode network paths containing ligand atoms into subgraph network descriptors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

Network edges: classification of interactions

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

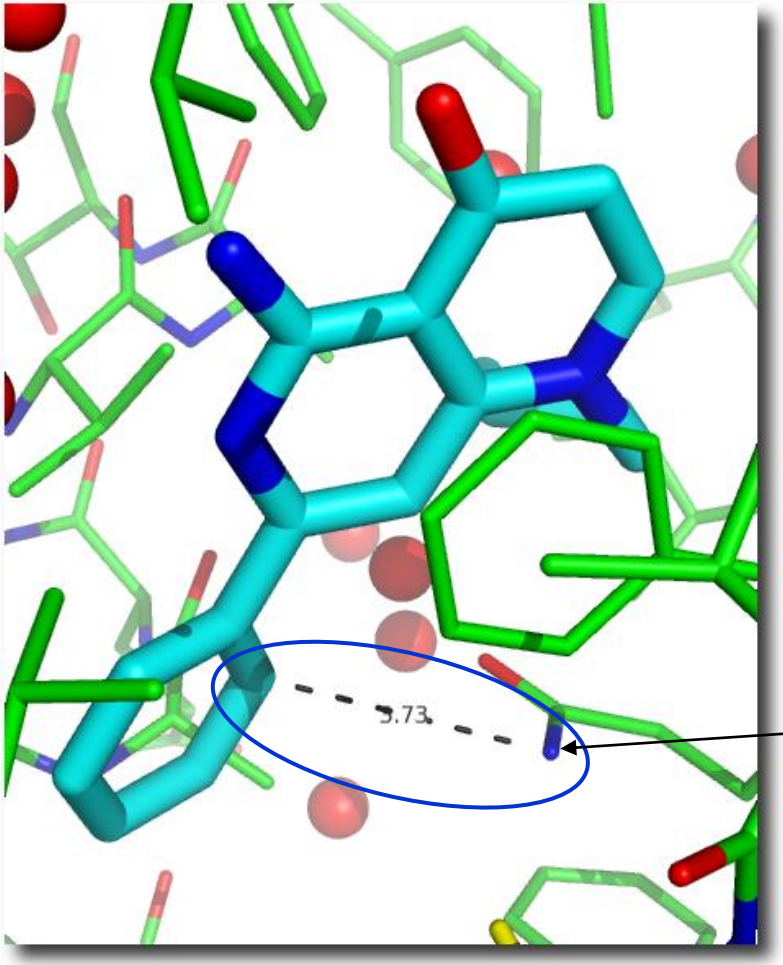
1. hydrogen bond
2. metal
3. ionic
4. cation-dipole
5. cation-pi
6. dipolar
7. σ-hole bond
8. h_donor-pi
9. pi-pi
10. vdW
11. unfavorable of 1, 2, 3, 6
12. polar and non-polar clashes
13. polar-nonpolar contacts with likely desolvation penalties

Ex. 2cf8 (thrombin)



ViewContacts (cont.): 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



Ex. 3hdz (PDE-5)

No hydrogen bond partner for this buried N atom in the binding site → an unfavourable interaction

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

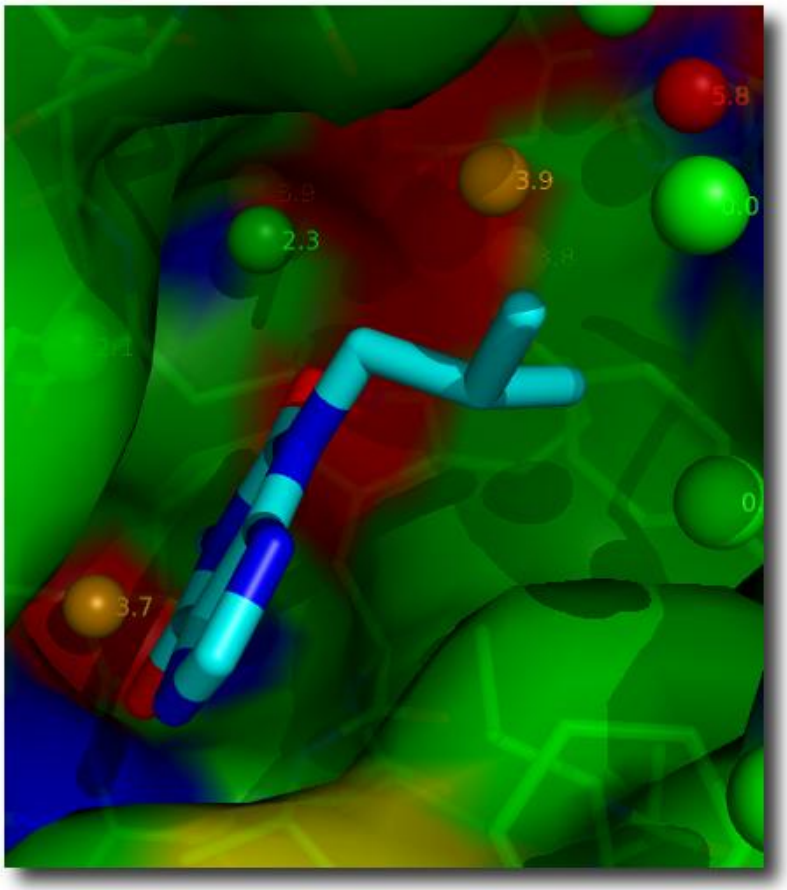
ViewContacts (cont.): handling of water molecules

- score explicit water molecules based on deviation from ideal tetrahedral coordination of protein-bound water molecules

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

- 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)
  - orange: 2.3-3.9 (possible to replace with suitable polar functionalities)
  - red: 3.9-6.0 (unlikely to replace)
- Amadasi et. al., J. Med. Chem., 2008, 51, 1063

Ex. 2r8q (PDE-B1)

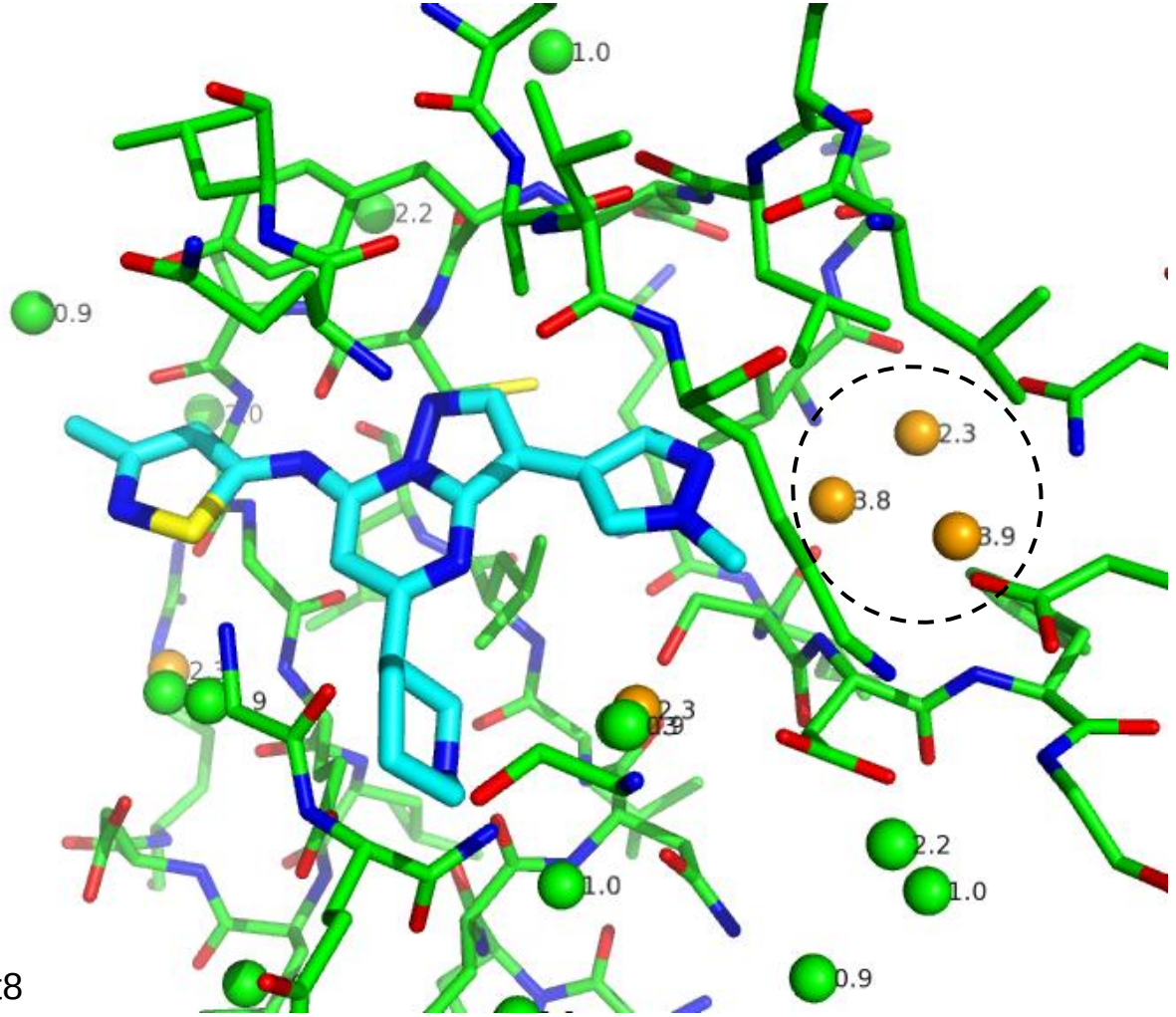


Water molecules with Rank scores ≥ 2.0 are included in networks

Waters in CHK1 Kinase

Three water molecules in CHK1, in a cavity adjacent to ligand, difficult to displace

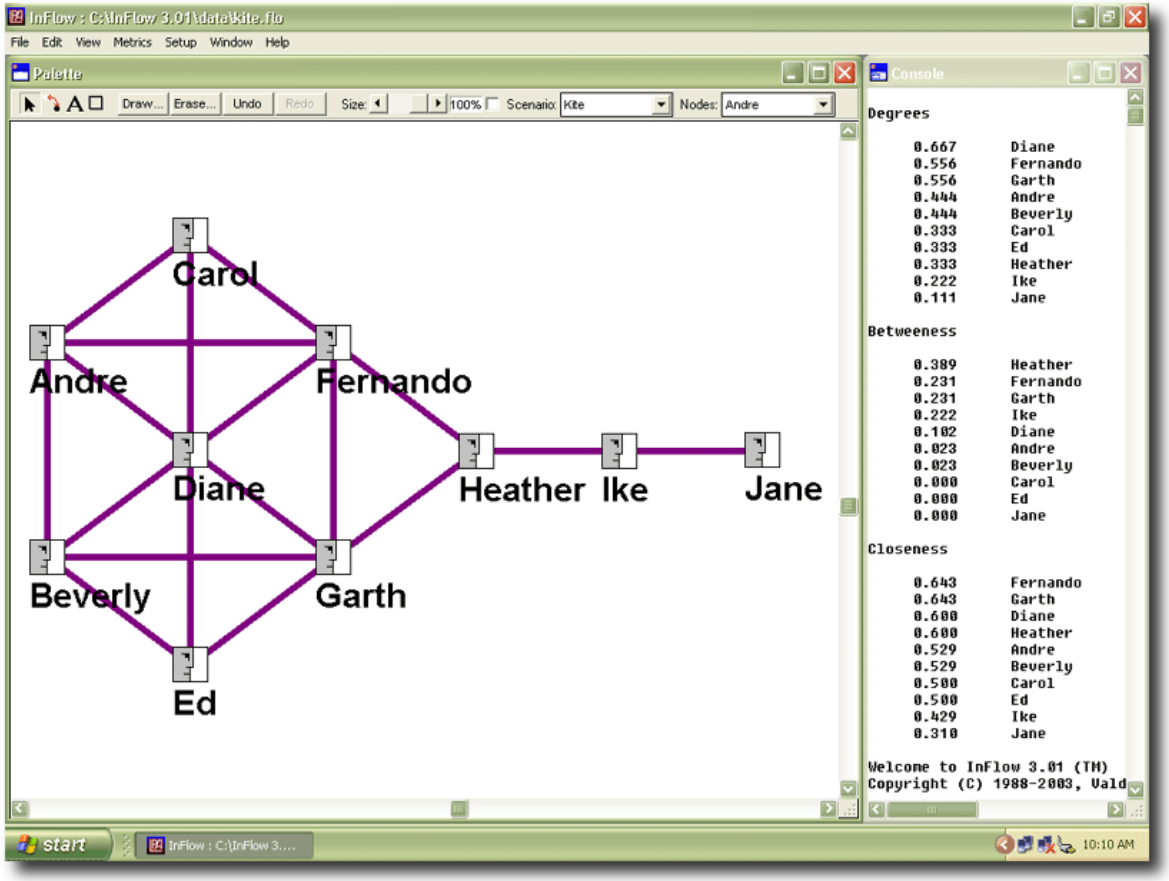
Water Rank Score shows that the ‘front‘ two waters have reasonably high scores



3ot8

Standard small world network (SWN) model

Initially explored using descriptors from Social Network Analysis



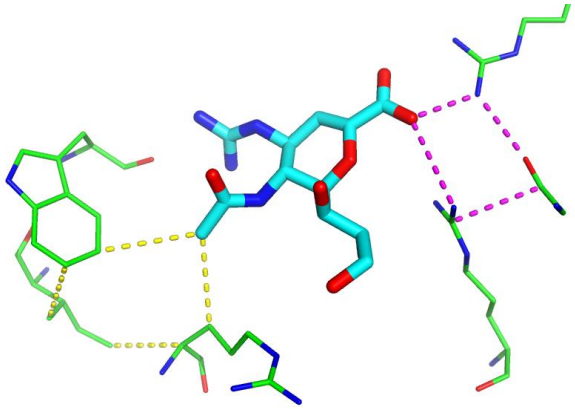
Kite Network, by D. Krackhardt

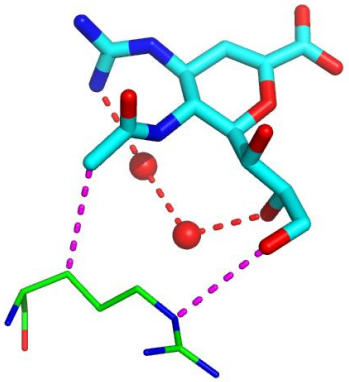
http://www.orgnet.com/sna.html

In our domain, these descriptors are too sensitive to individual contacts, and to geometric constraints associated with maximum number of contacts

Network descriptors: paths involving ligand atoms

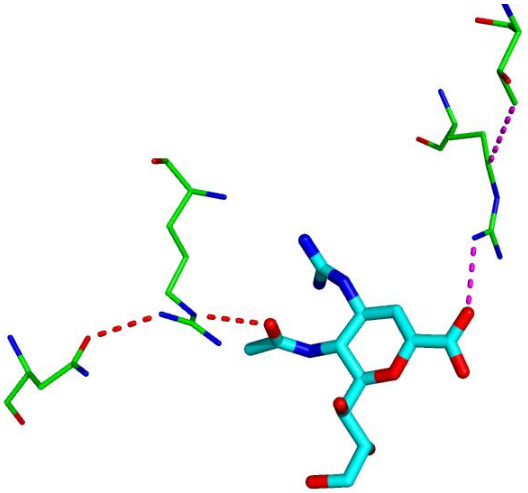
– ligand-protein-ligand (LPL) network elements
  - ligcycles (involving 1 ligand atom) ligloops (involving ≥ 2 ligand atoms)





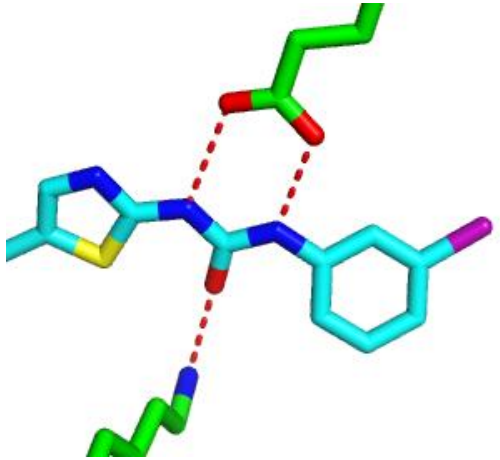
examples from 1nnc

– ligand-protein-protein (LPP) network elements
  - ligpaths (subsets of long ligcycles/ligloops > 8)

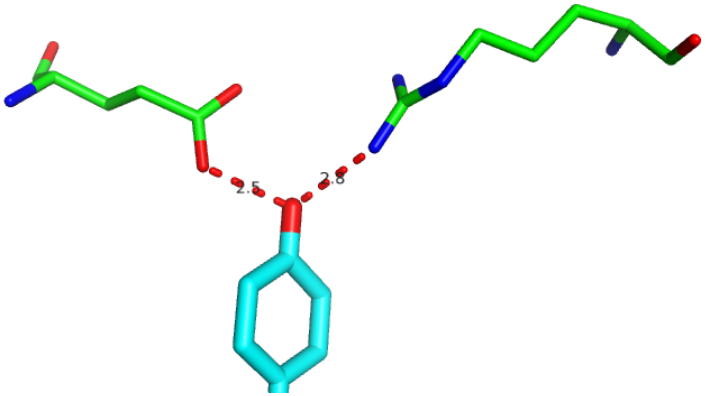


Network descriptors: special treatment of hydrogen bonding

- *privileged pairs* of hydrogen bonds
  - arrangements of hydrogen bonds that can not be achieved in the apo state

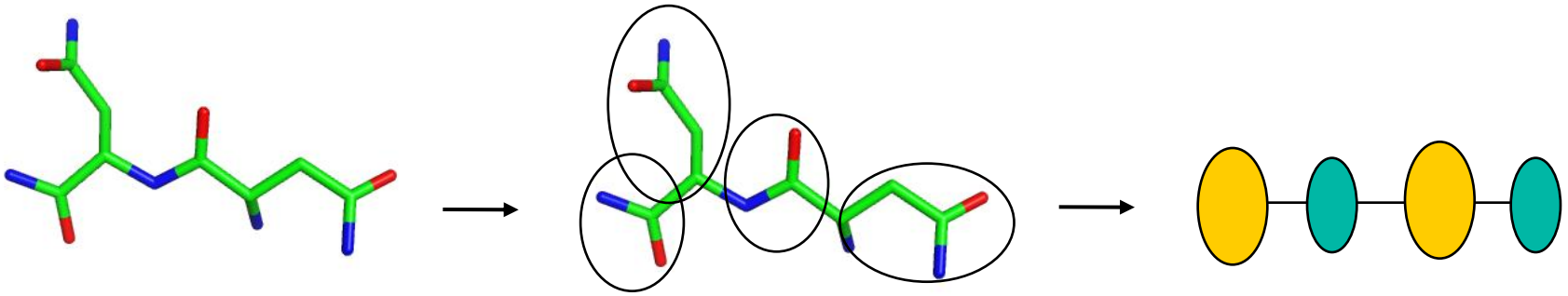


- protein-ligand-protein (PLP)
  - with lower free energy than the sum of the individual bonds due to mutual polarization



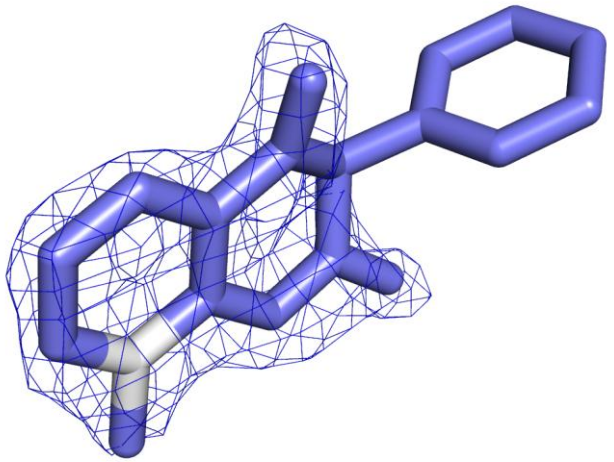
Network descriptors: nodes based on a reduced graph definition of protein structure

- 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

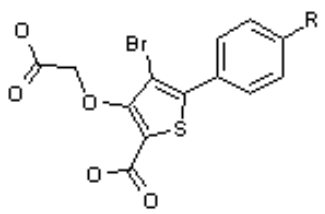
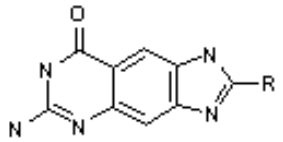
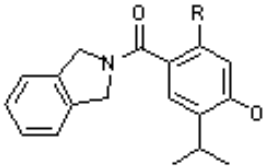


Electon density correlation coefficient is a better measure of model quality than B-factors

Training sets: high quality structures with binding affinity data

I) hard set: 28 compounds:

activity cliff pairs

4	protein tyrosine phosphatase 1B		-OH	2h4g	6.5
			-H	2h4k	5.5
5	trna-guanine transglycosylase		$-NH_2$	2z7k	7.1
			$-CH_3$	3c2y	5.8
6	hsp90		-OH	2xab	9.3
			-H	model	7.2

II) 31 neuraminidase compl

III) 46 PDE10 complexes

IV) 7 subsets with up to 10 structures each:

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

Global optimisation

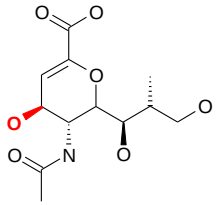
- based on high quality structures and results from docking
- optimisation used genetic algorithm approach
- form of scoring function:

$$S = \sum_{n} f(nt)$$ (without network terms)

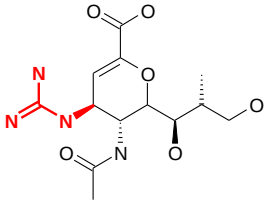
$$S_{Scorpion} = \sum_{n} f(nt) + \sum_{m} g(nt\_nw)$$

- a particular protein-ligand interaction considered networked if [weighted] sum of network elements higher than an interaction-specific threshold

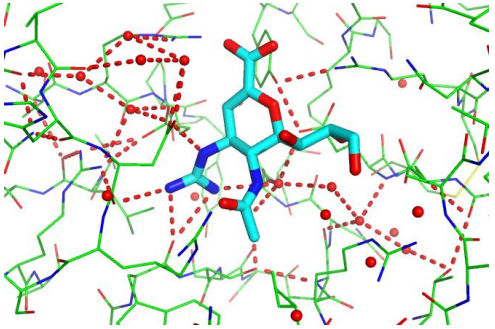
Activity cliffs: predicted vs. experimental energy differences

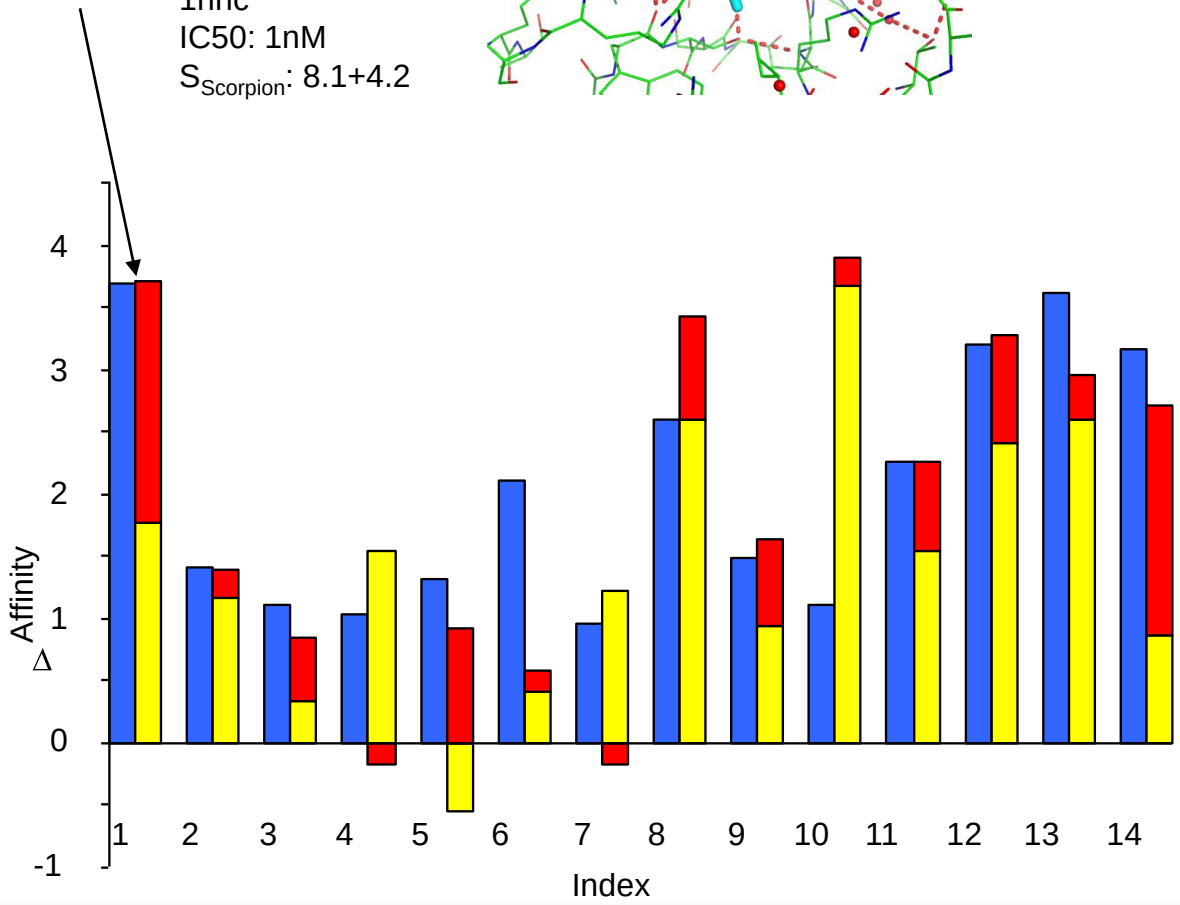


1nnb
IC50: 5uM
$S_{Scorpion}$: 6.4+2.2



1nnc
IC50: 1nM
$S_{Scorpion}$: 8.1+4.2





Scorpion Score

$$S_{Scorpion} = 0.473\,x\,[hbond] + 0.129\,x\,[hbond\_nw]$$
$$+ 0.516\,x\,[vdw] + 0.387\,x\,[vdw\_nw]$$
$$+ 0.188\,x\,[pi - pi] + 0.931\,x\,[pi - pi\_nw]$$
$$+ 0.285\,x\,[cat - dipole] + 0.606\,x\,[cat - pi] + 0.65\,x\,[halogen]$$
$$- 0.387\,x\,[unf\_hbond] - 0.899\,x\,[unf\_desolv] - 1.146\,x\,[unf\_clash] - 1.501\,x\,[unf\_ionic]$$

H-bond donor – nonpolar contact

Results shown from optimisation done in 2010

- scoring function optimisation is on-going, we continue to improve our results

External validation

	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**

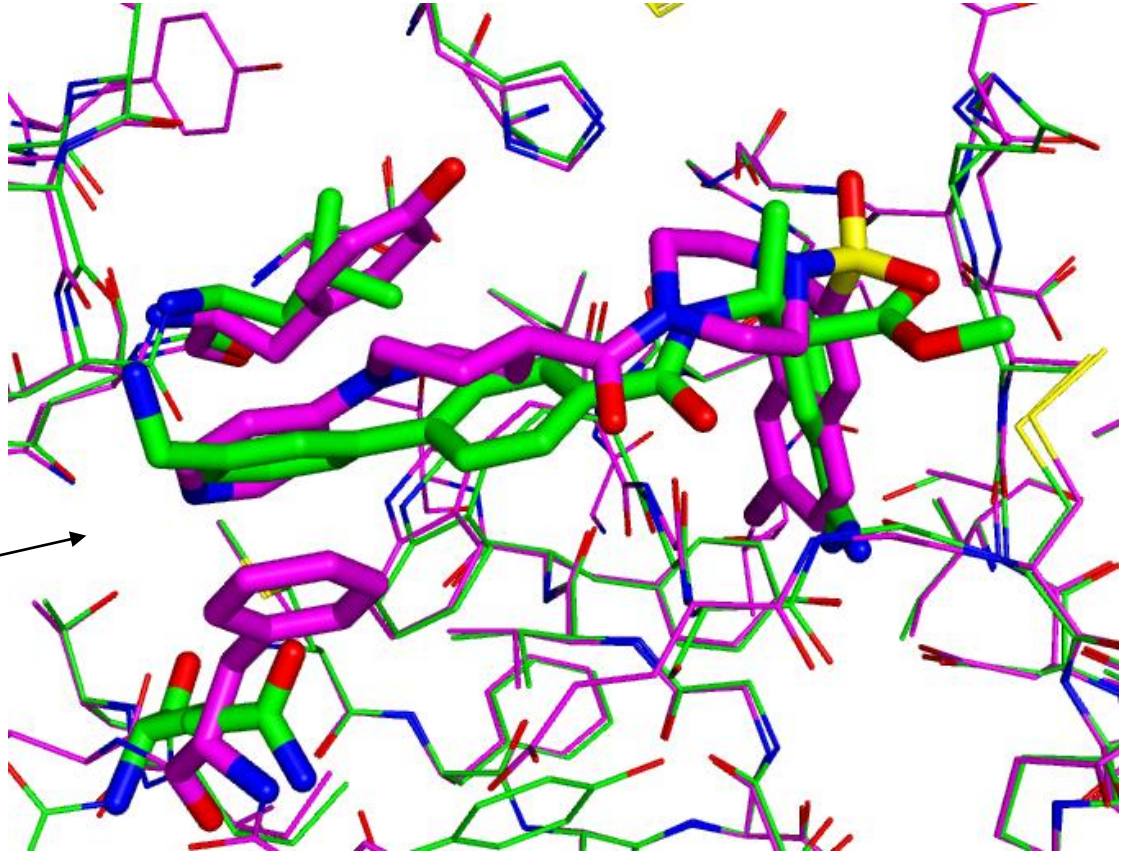
P. Englebienne, N. Moitessier, *J. Chem. Inf. Model.* **2009,** *49,* 1568–1580

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

External validation sets for binding affinity prediction are flawed

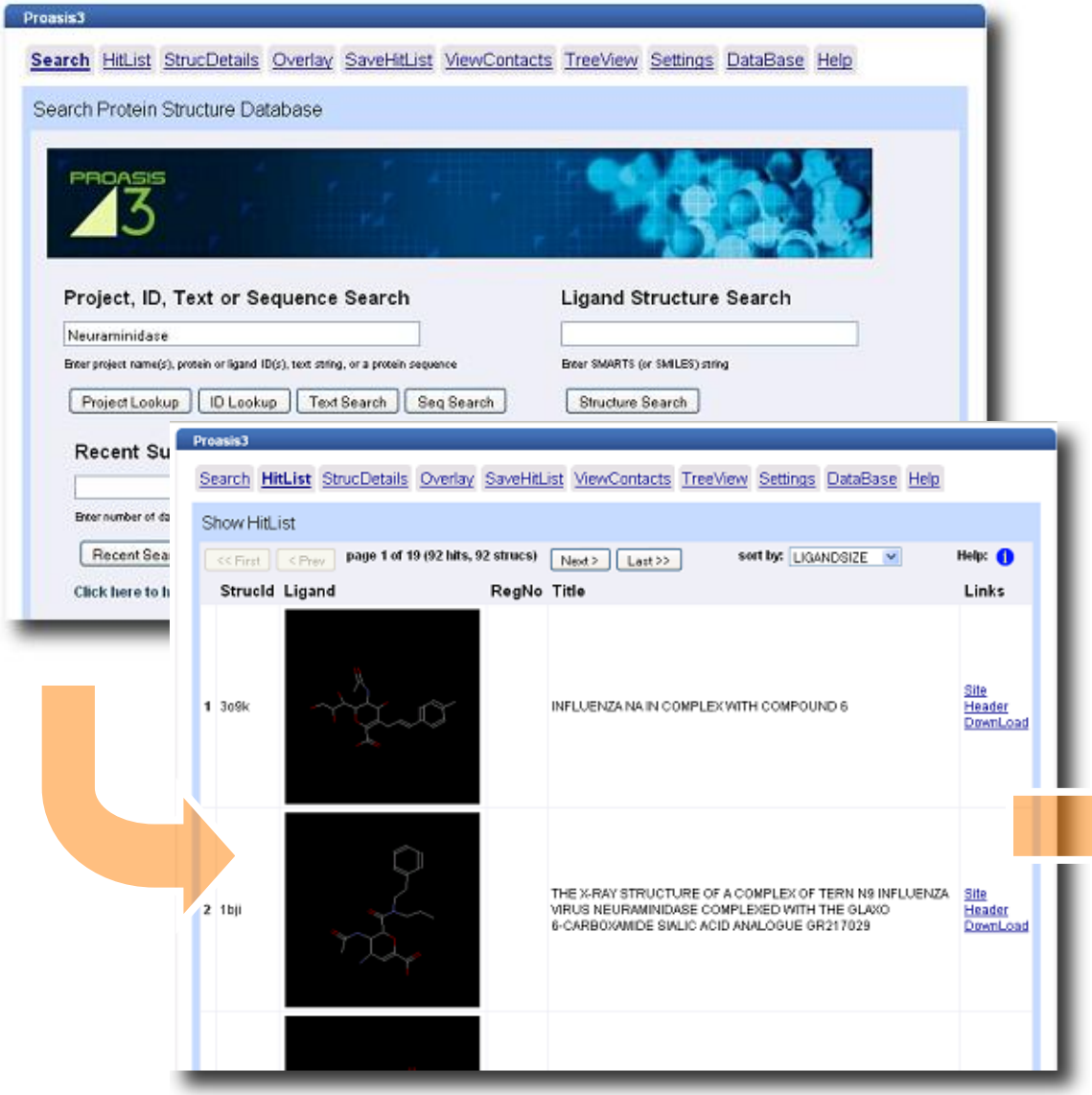
Example: Trypsin subset (13 structures) -> *only one would pass our quality criteria (1f0u)*

1f0u (green) and 1v2k (magenta) both in trypsin set despite large differences in S4 pocket

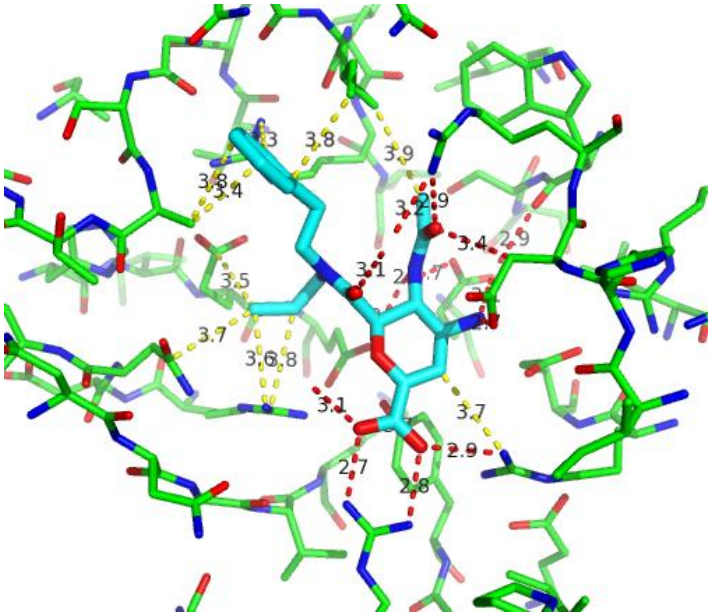


1v2k: FXa S4 pocket
due to mutations

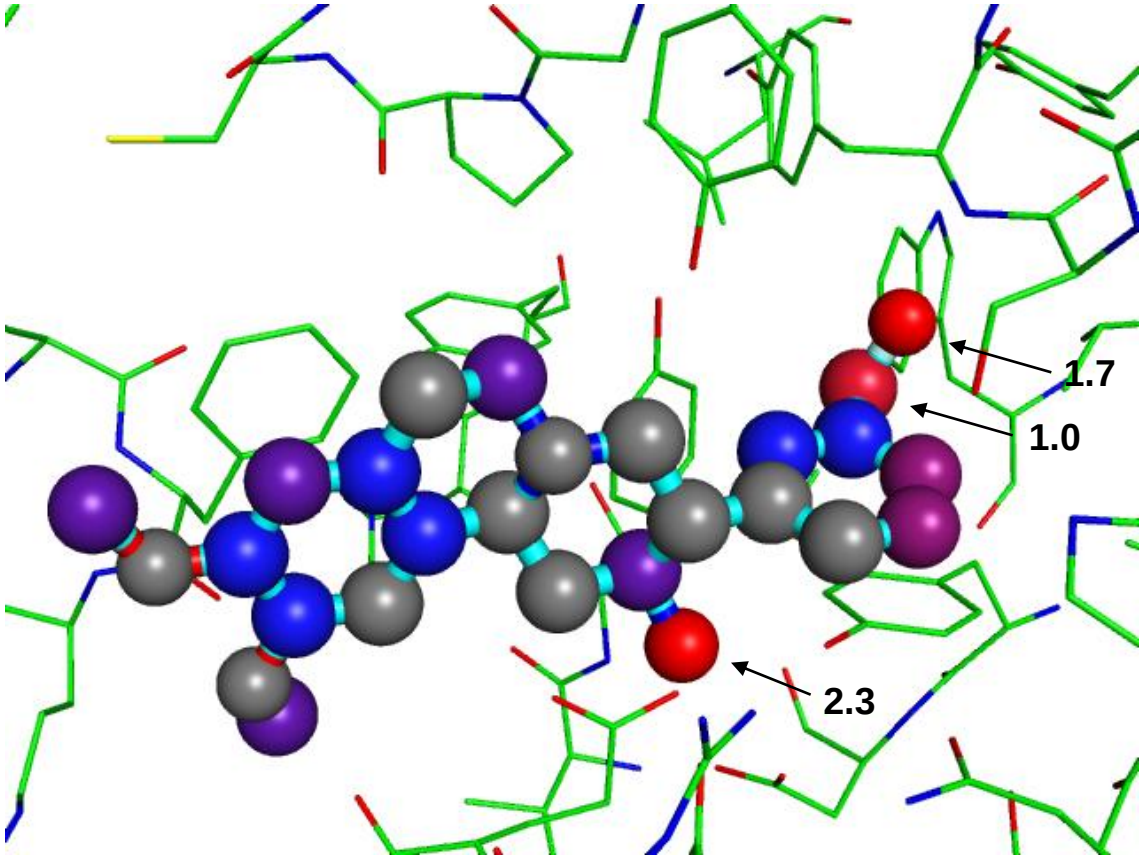
Quick and easy visualisation



… at the click of a button

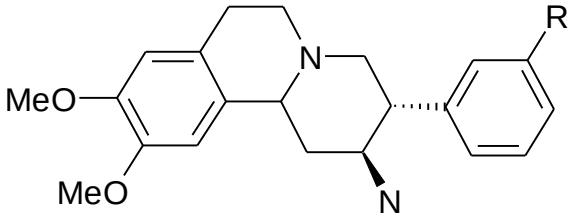


Score contributions mapped onto atoms



3kwj

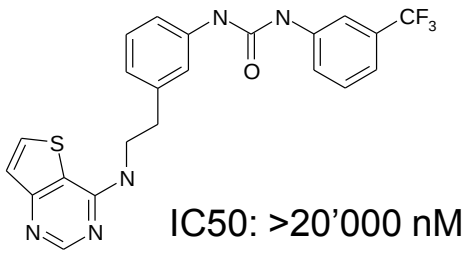
(dpp-4)



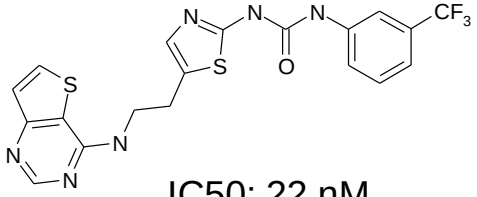
R	K_d (nM)	$S_{Scorpion}$	network
H	200	7.9	0.8
CH_3	4.6	10.8	1.6
CH_2F	0.5	11.3	1.7

color ramp from blue -> red
gray = no score contribution

Aurora A kinase inhibitors

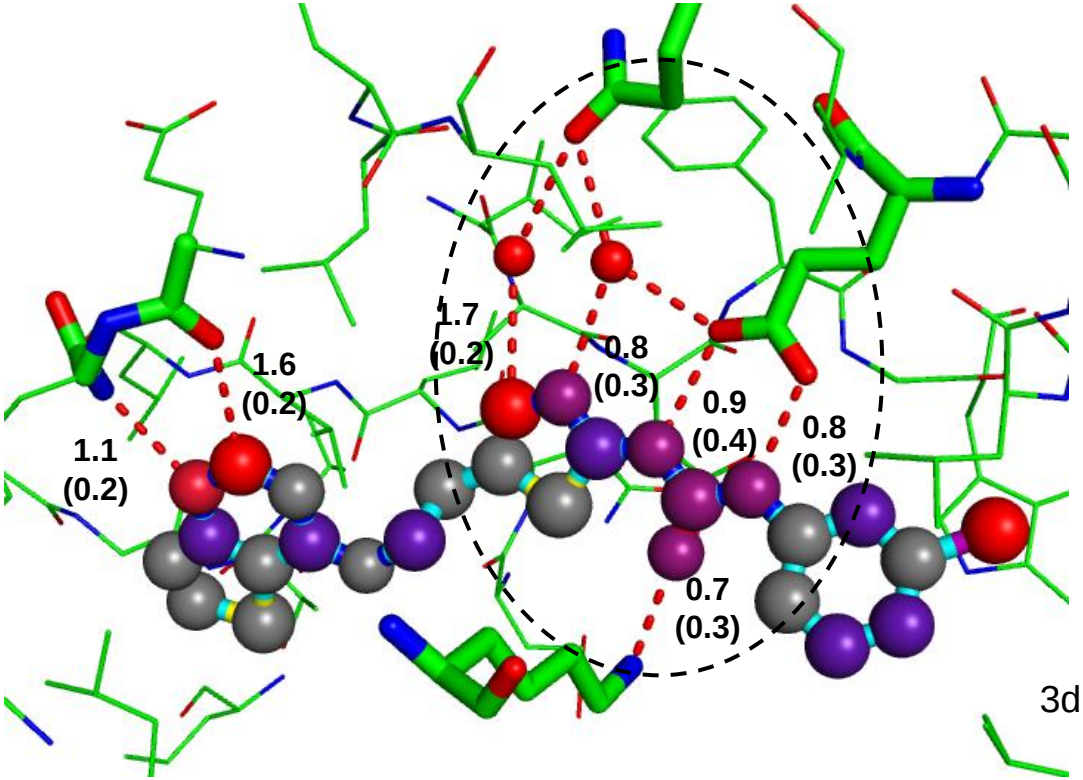


IC50: >20’000 nM



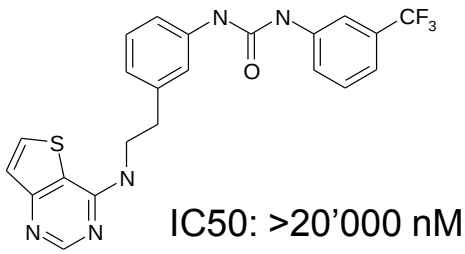
IC50: 22 nM

networked H-bonds with high score
incl. network contribution

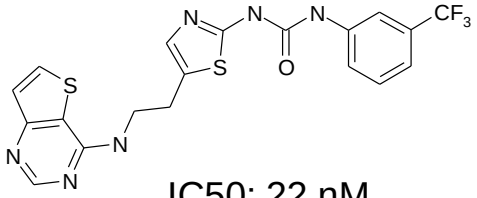


3d15

Aurora A kinase inhibitors (cont.)



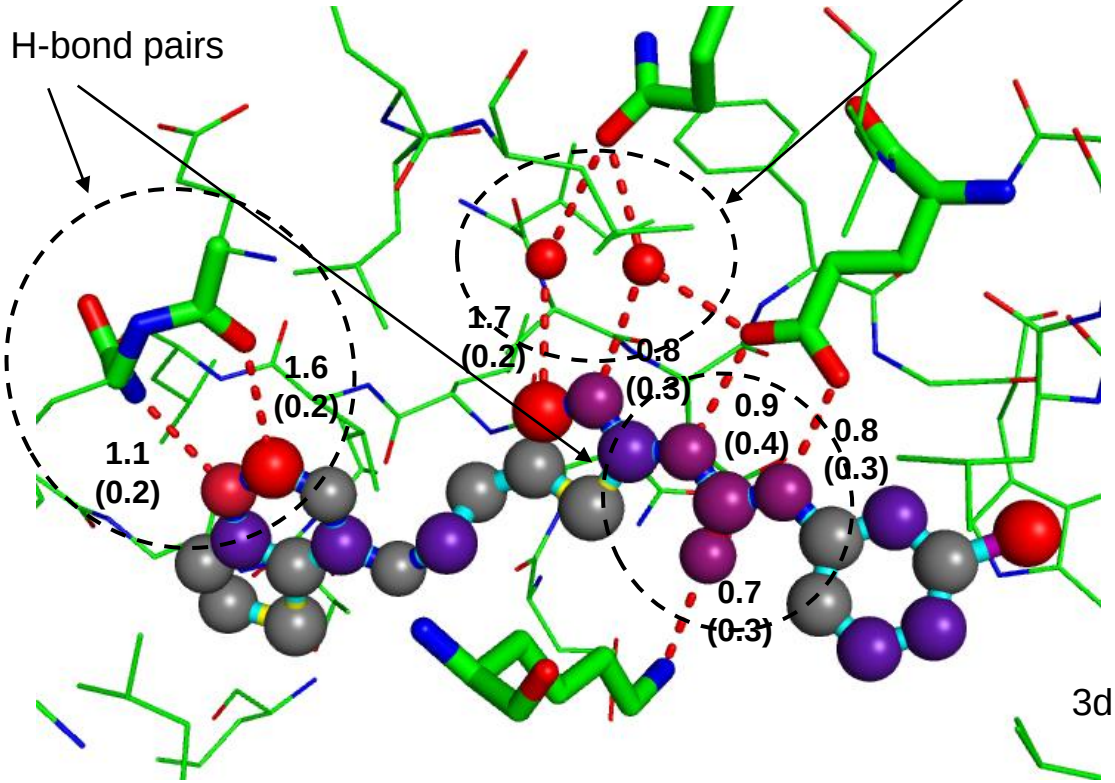
IC50: >20’000 nM



IC50: 22 nM

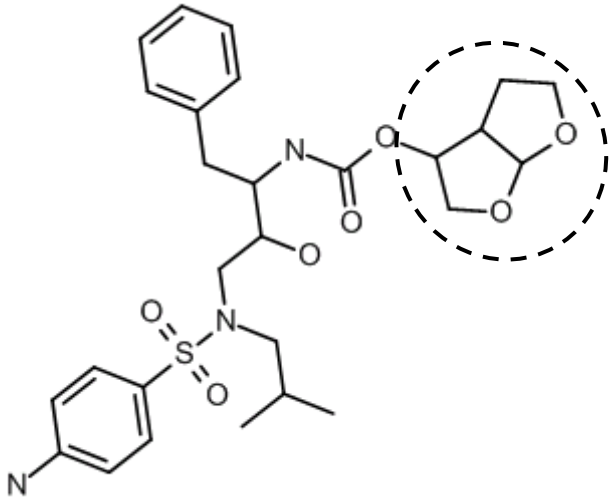
tightly bound waters play important role in networks

privileged H-bond pairs



3d15

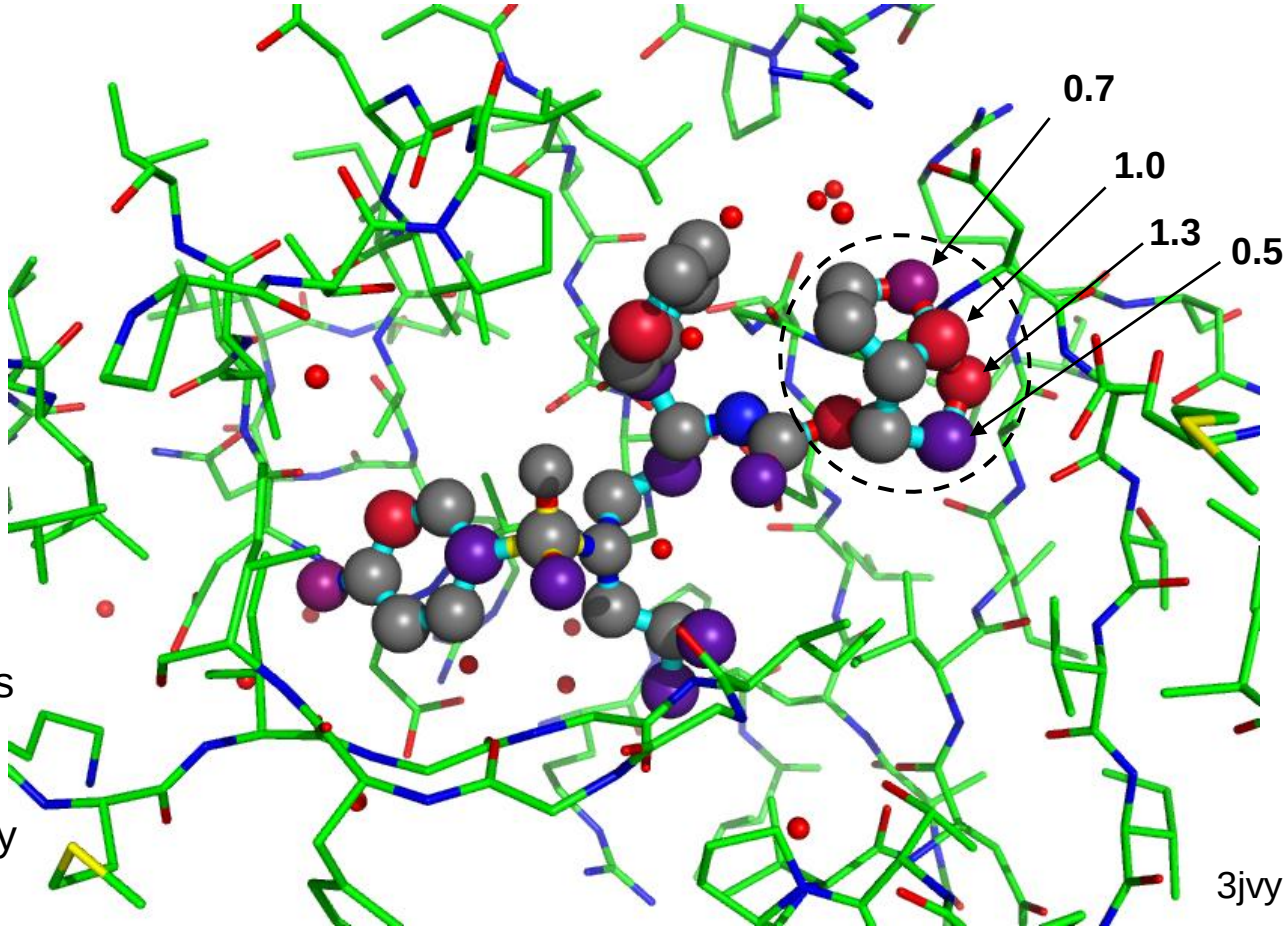
HIV Protease inhibitor - darunavir



Exp: pIC50 ~ 11

Calc: pIC50 ~ 10

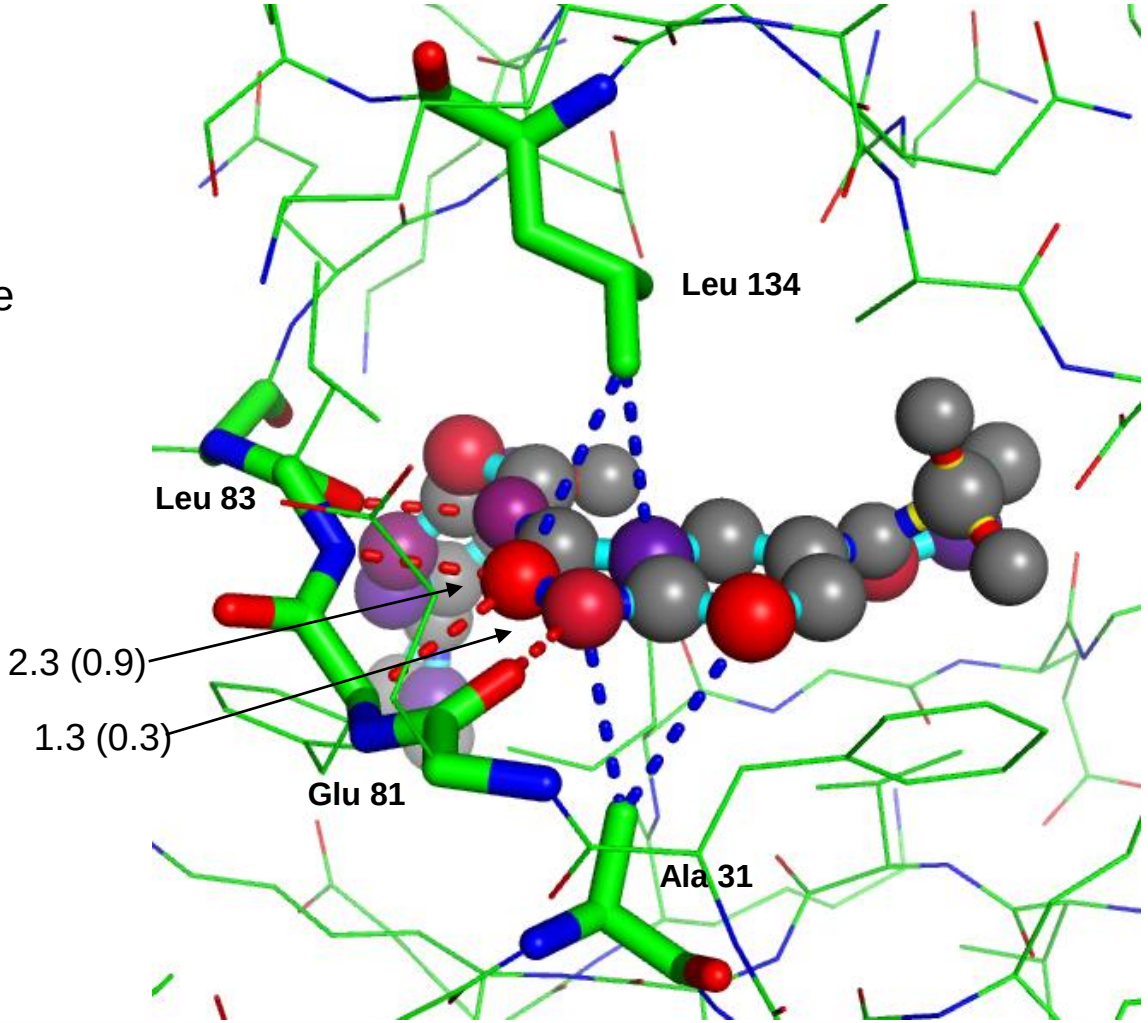
bis-THF displaces two water molecules associated with backbone NH‘s and forms hydrogen bonds and vdW contacts with the protein in optimal way



3jvy

Aminoindazole in CDK2

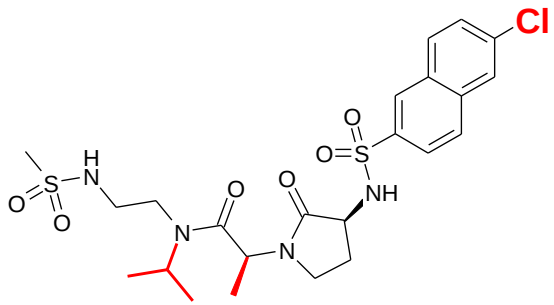
aromatic heterocycle with 1-2 H-bonds and in a hydrophobic environment (e.g. kinase hinge binding)

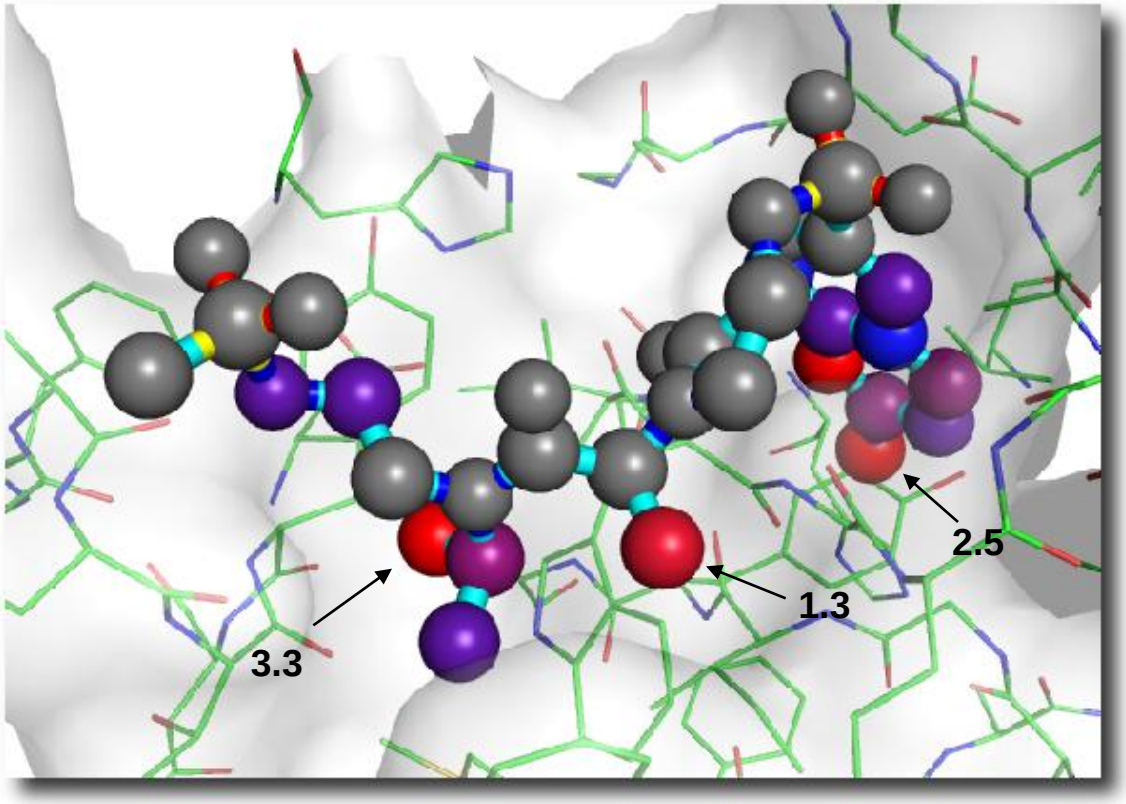


2r64

Examples of highly networked atoms

Atoms in buried pockets with several contacts receive extra network contribution





2j4i (Factor Xa)

other examples:

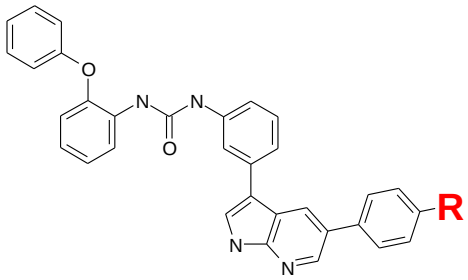
1yvz (hcv polymerase): Cl

1ql7 (trypsin): Cl

2r3r (cdk2): Br

Insulin receptor kinase – pyrrolopyridine complex

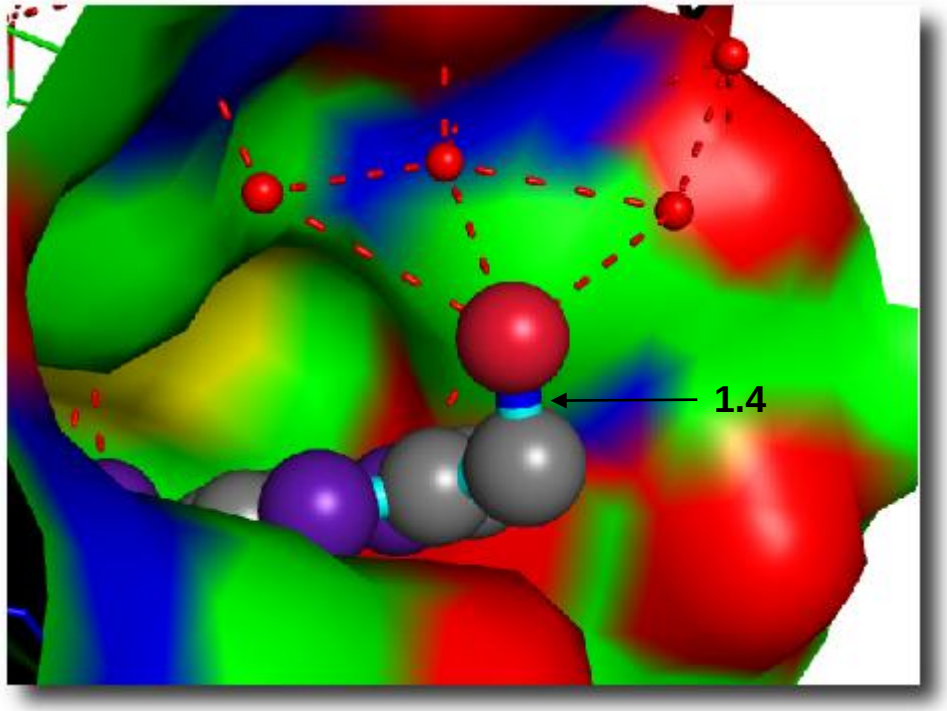
Ligand atoms can have high network scores in spite of being highly solvent-exposed



R:	IC50[nM]
-H	127
$-CH_2NH_2$	14

-aminomethyl group solvent exposed with no direct contact with protein

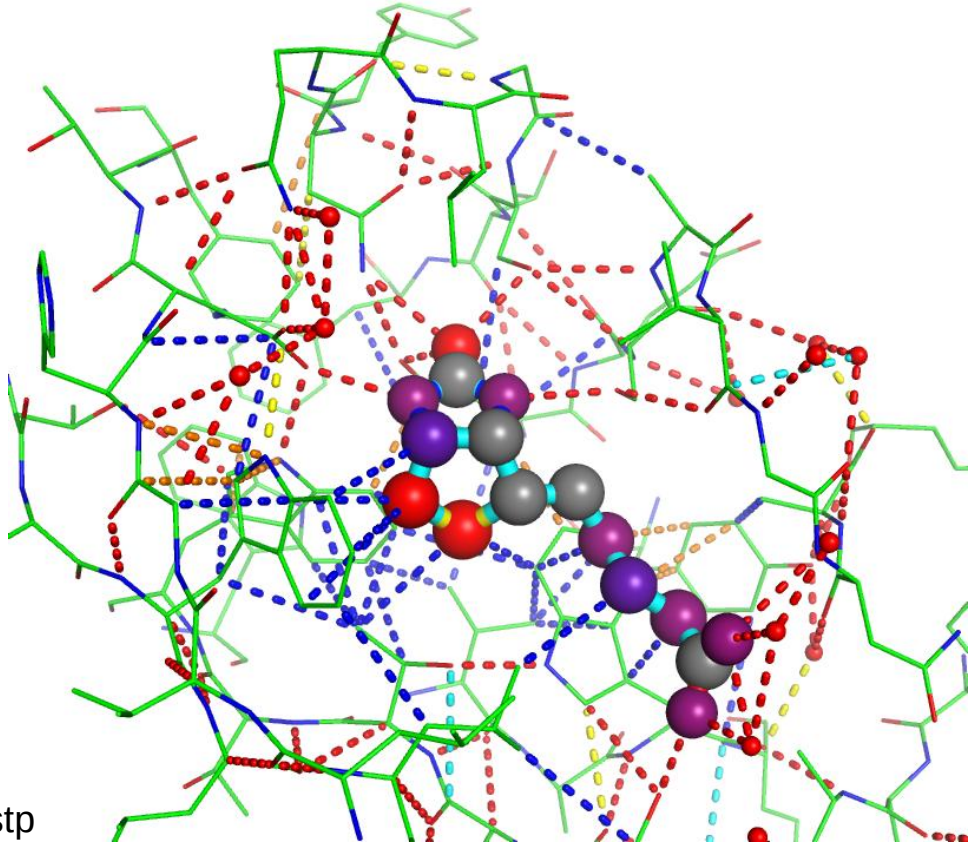
- amino group interacts through protein-bound water molecules with insulin receptor resulting in a high score despite low buriedness



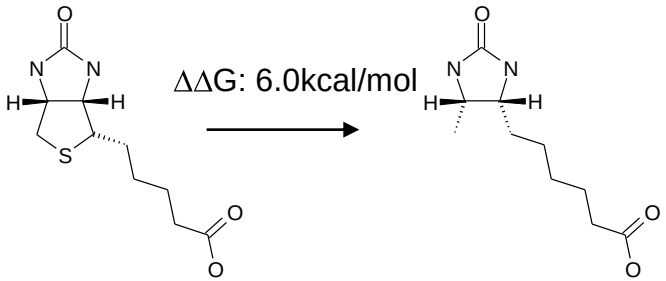
3eta

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)



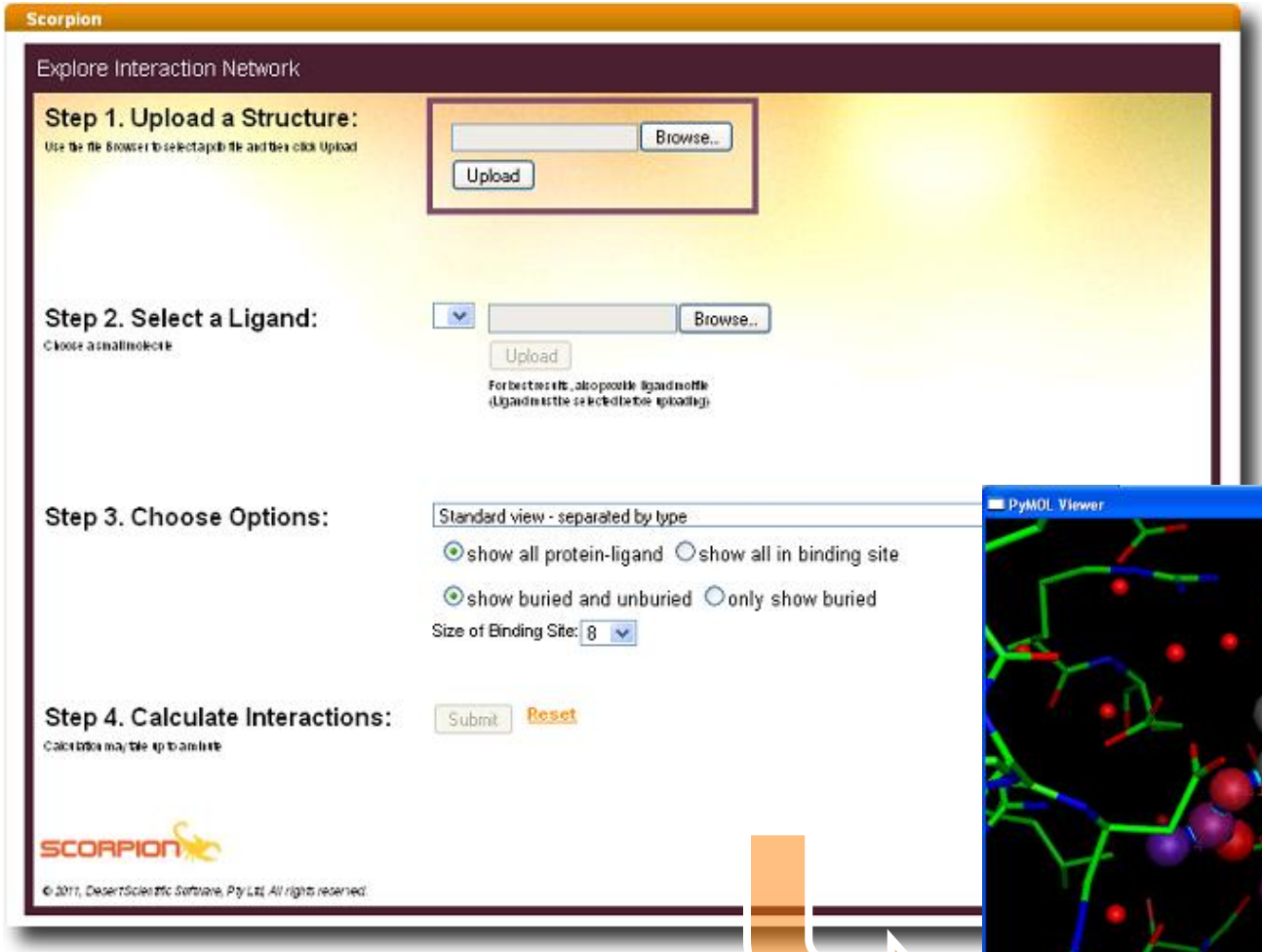
1stp

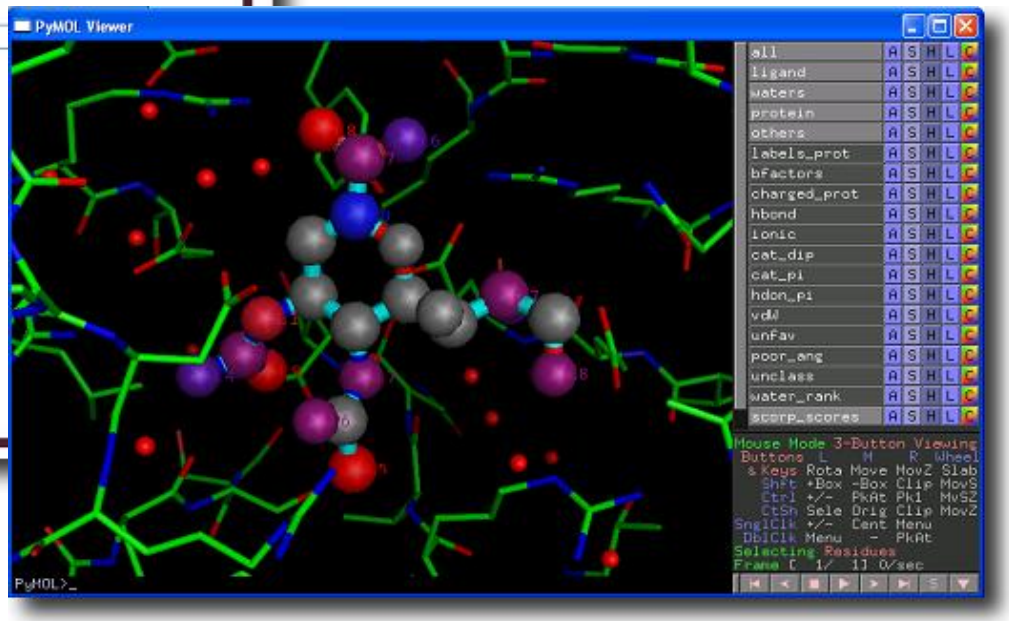


“The streptavidin/biotin system provides a clear example where the binding affinity is the propertry of the whole system”

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

Running Scorpion: using command line tool, Proasis3 system, or ScorpionWeb



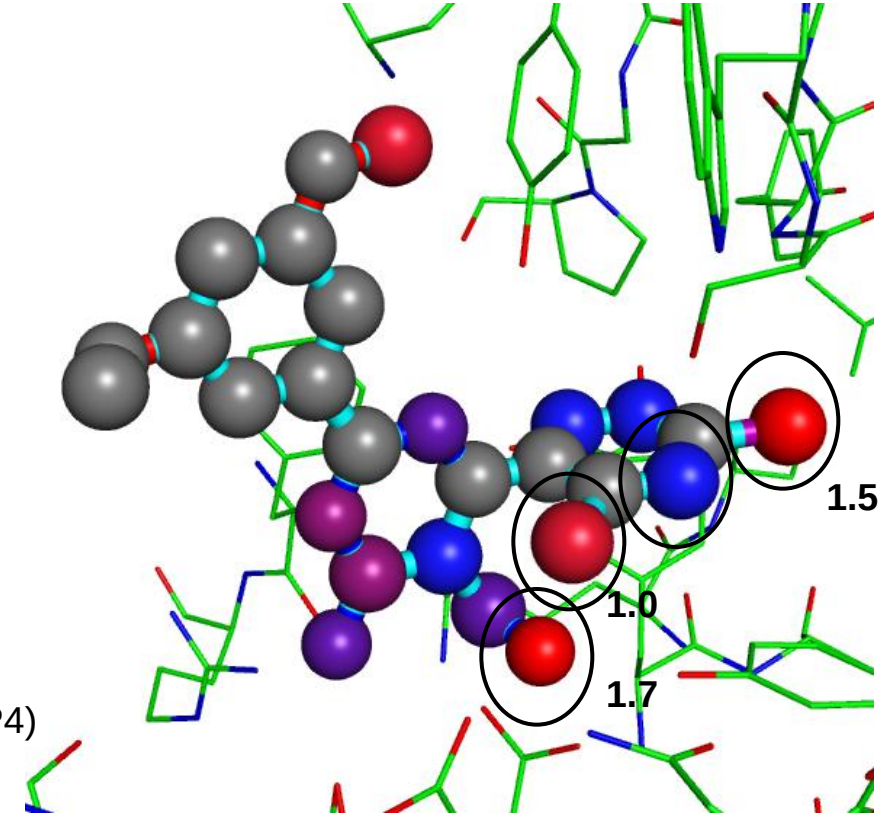


ScorpionWeb is an easy-to-use web GUI that runs highly optimized javascript which is fast to load and fast to execute in all popular browsers

Cooperativity pairs

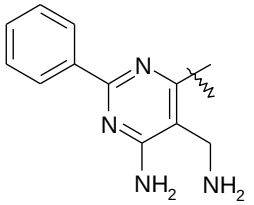
- identification of favorable network motifs

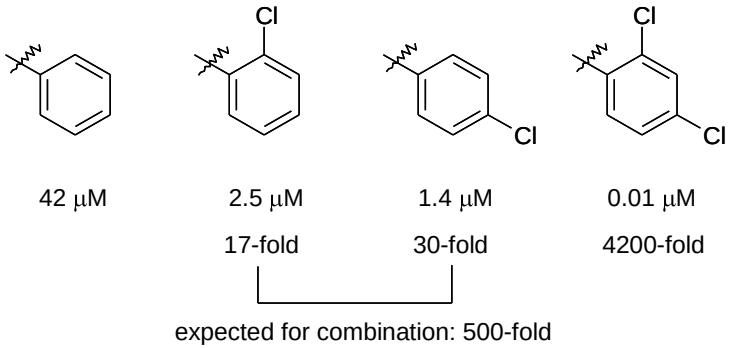
- status: first version completed
  released Mar 2010



1rwq

(DPP4)





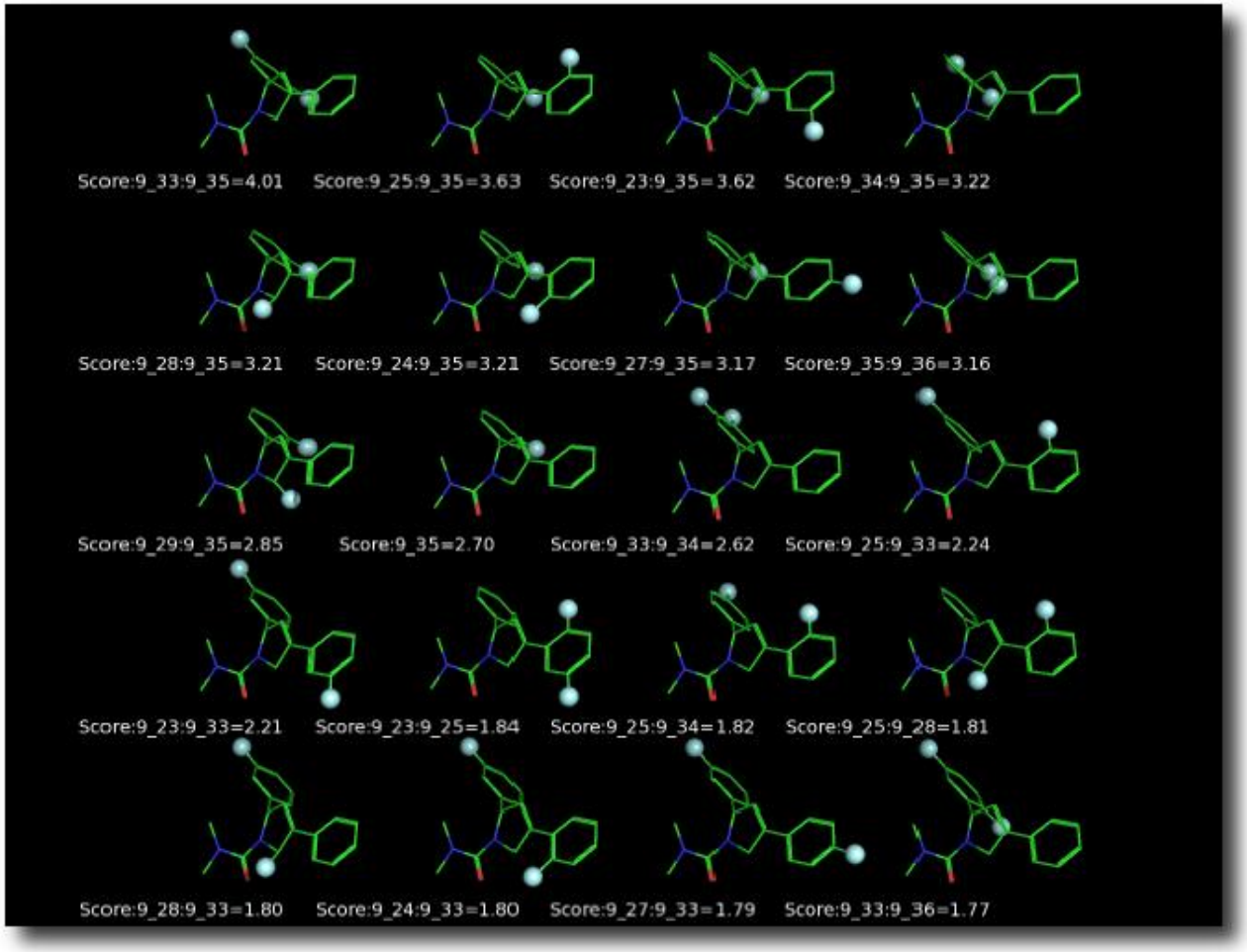
-circled atoms identified as potential cooperativity partners [A,B] - high networkedness of A and B with protein and LPL link from A to B

Ligand design using Viper: atom scan

- command line tool identifying explict 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

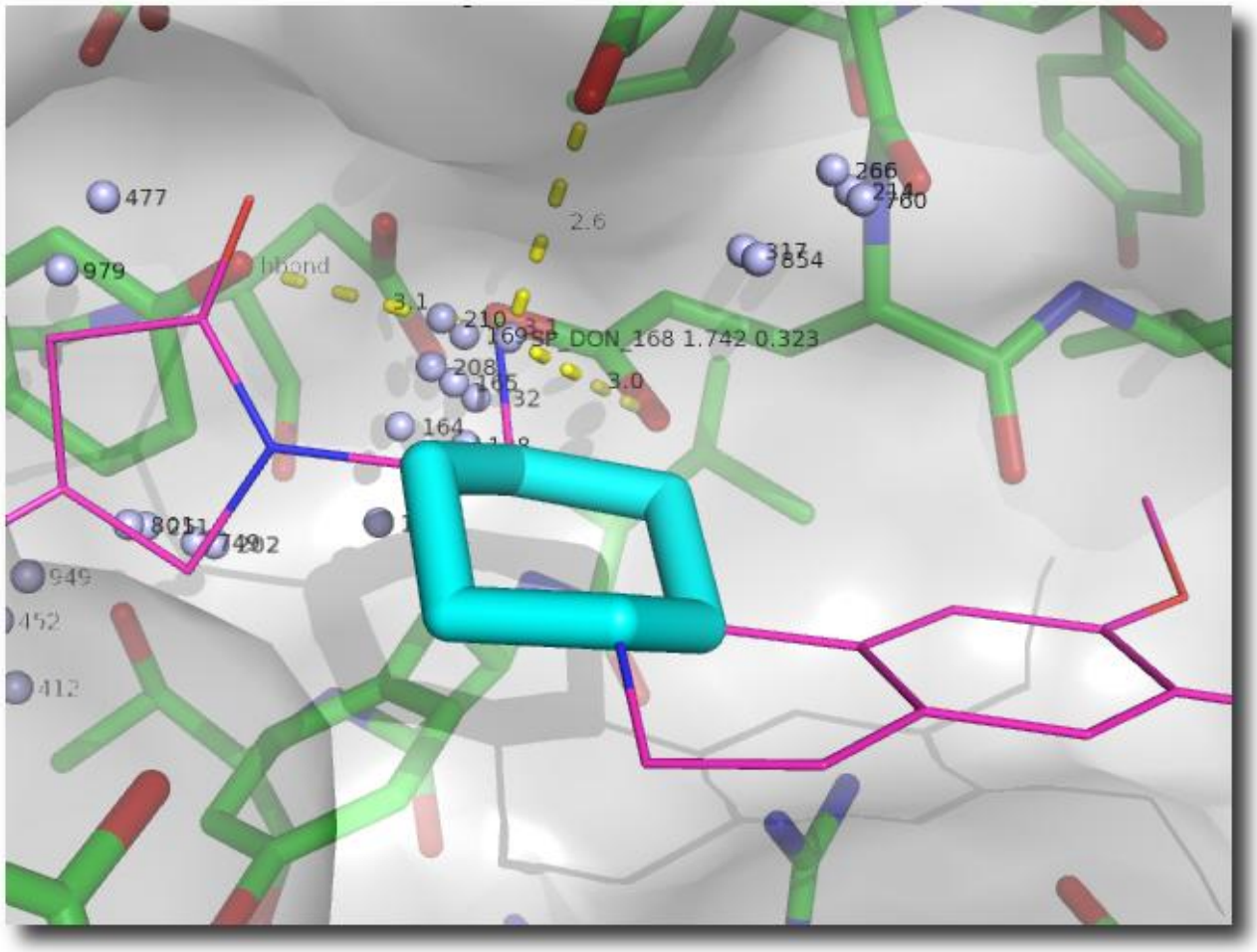
- command line tool highlighting pharmacophores for strong protein-ligand interaction networks

- 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



3kwf

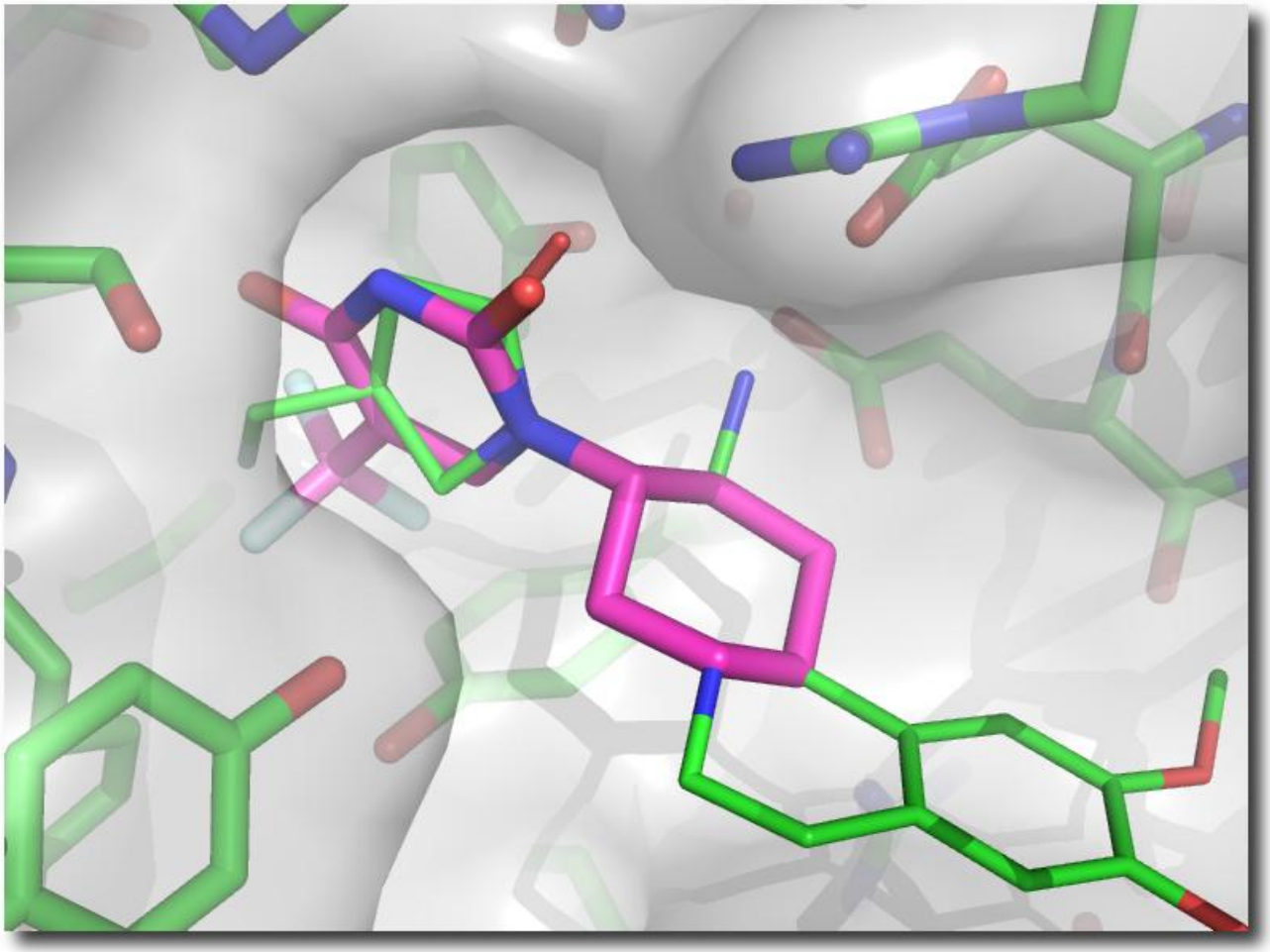
Ligand design using Viper: fragment scan

- 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



3kwf

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