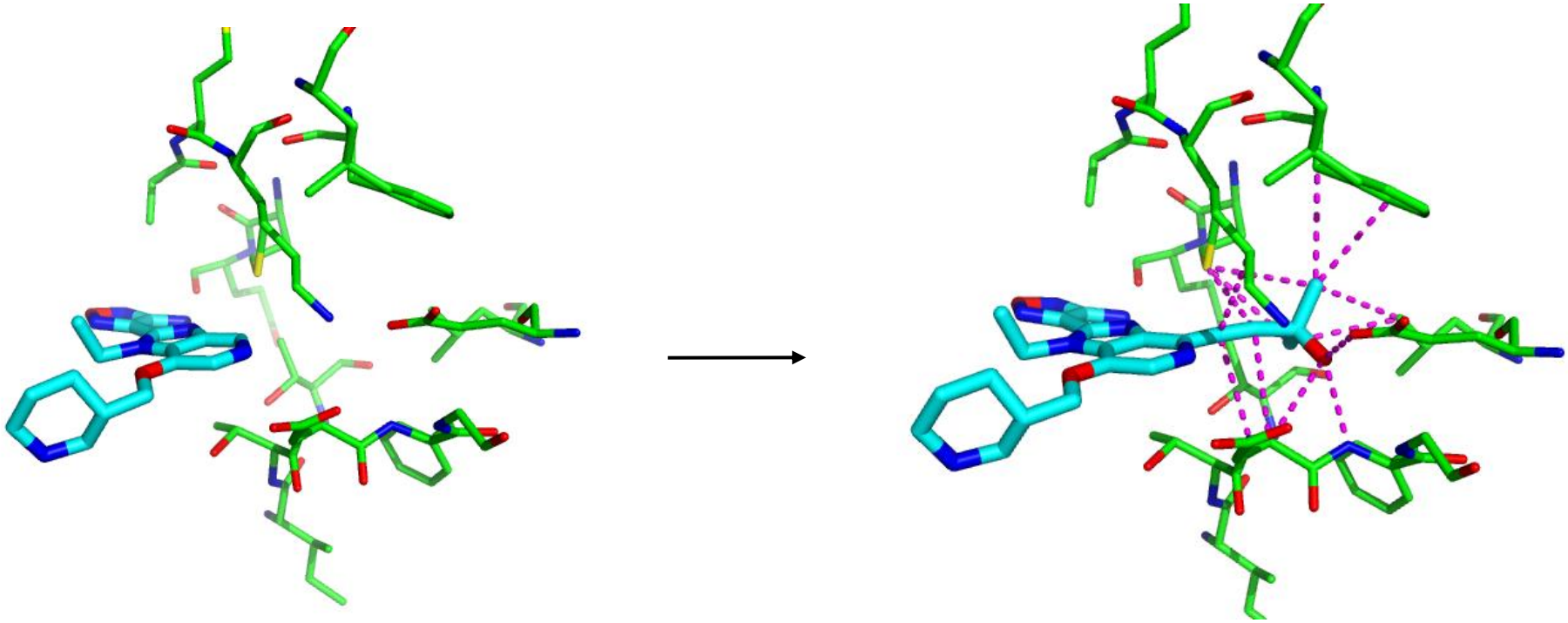
An Exploration of Network Hotspots and Cooperativity in Protein-Ligand Recognition

Neil R. Taylor, Desert Scientific Software, Sydney, Australia

A joint venture between Desert Scientific Software and F. Hoffmann-La Roche

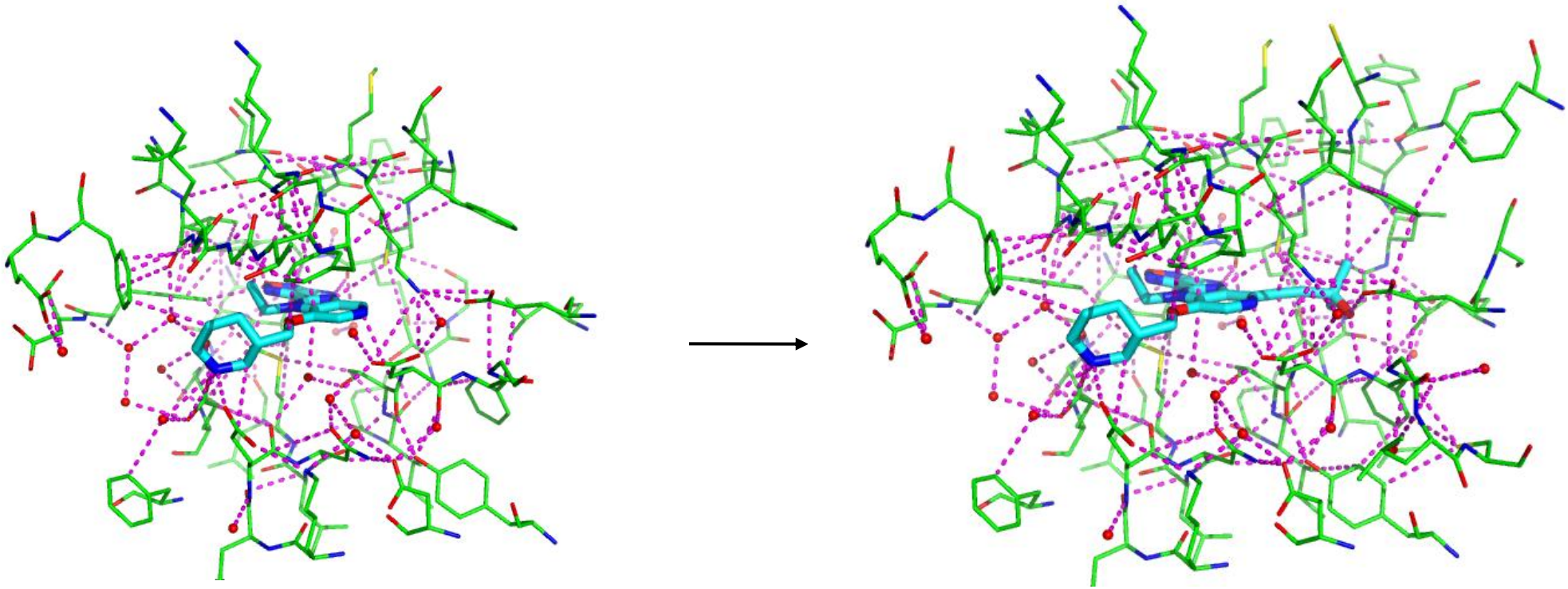
Ligand binding typically understood as the sum of protein-ligand interactions



Additional interactions lead to tighter binding

... not that simple

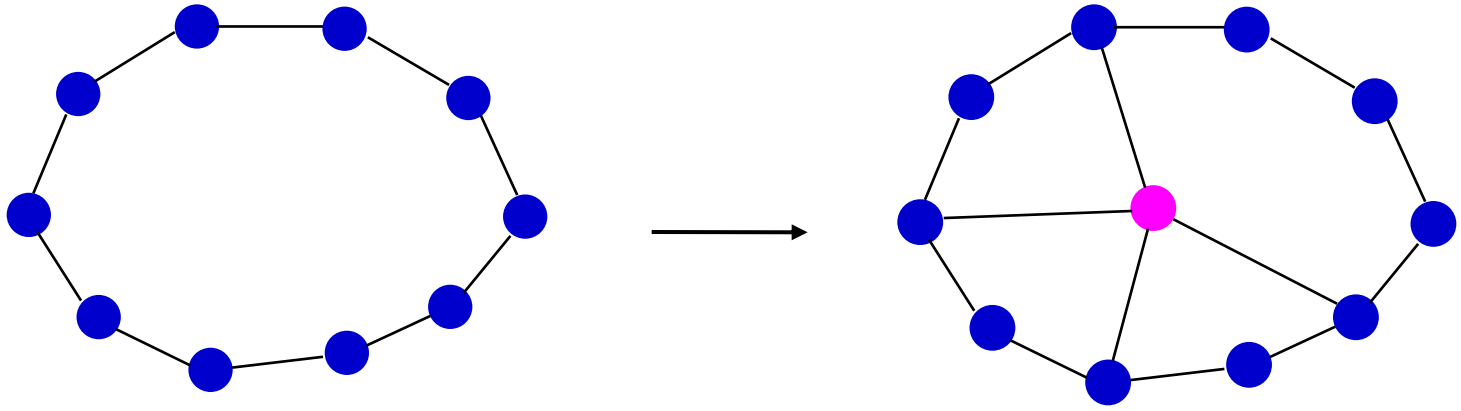
Beyond the pairwise additive view of protein-ligand interactions



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

... propose additional network paths lead to tighter binding

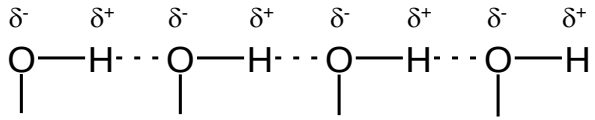
New concept: protein-ligand complex modelled as a small world network (SWN)



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

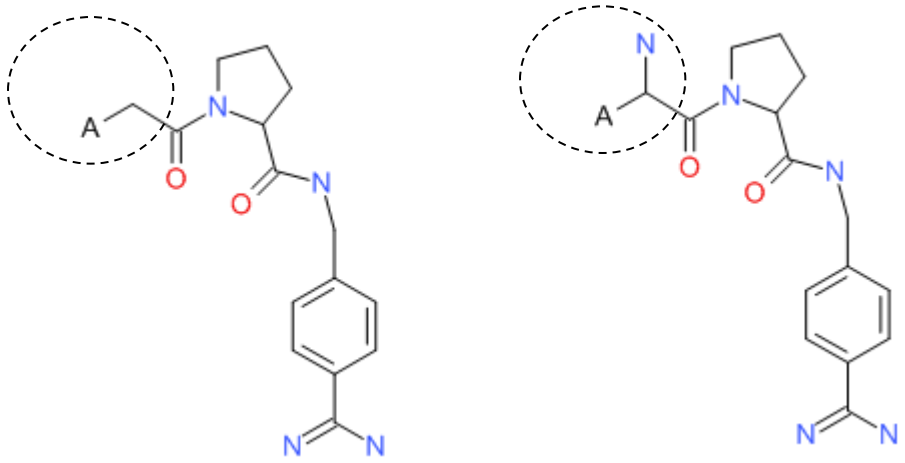
Types of cooperativity



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

Types of cooperativity (cont)

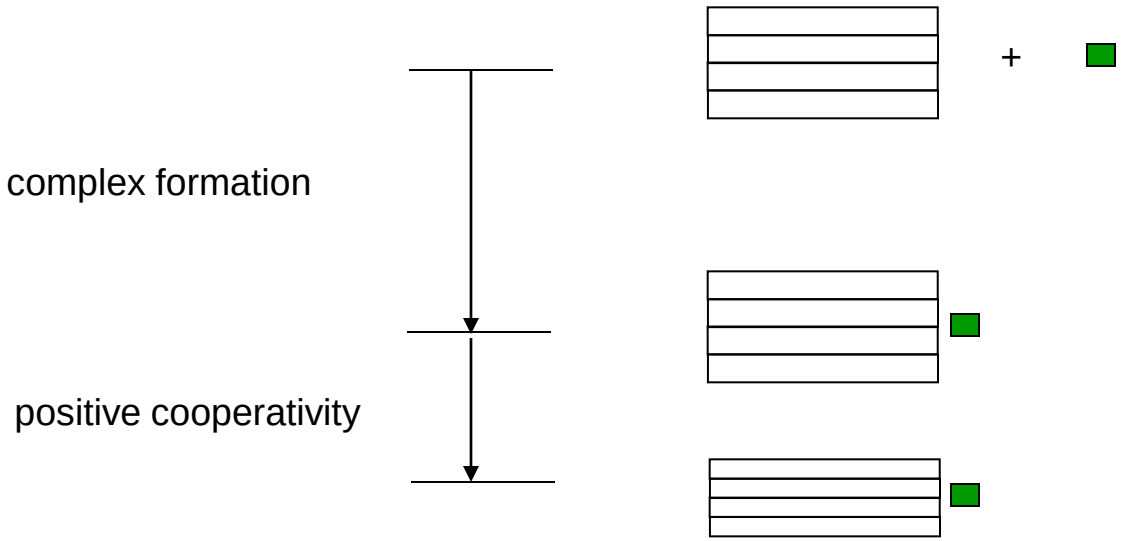
Non-additivity of functional group contributions in a series of thrombin inhibitors



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

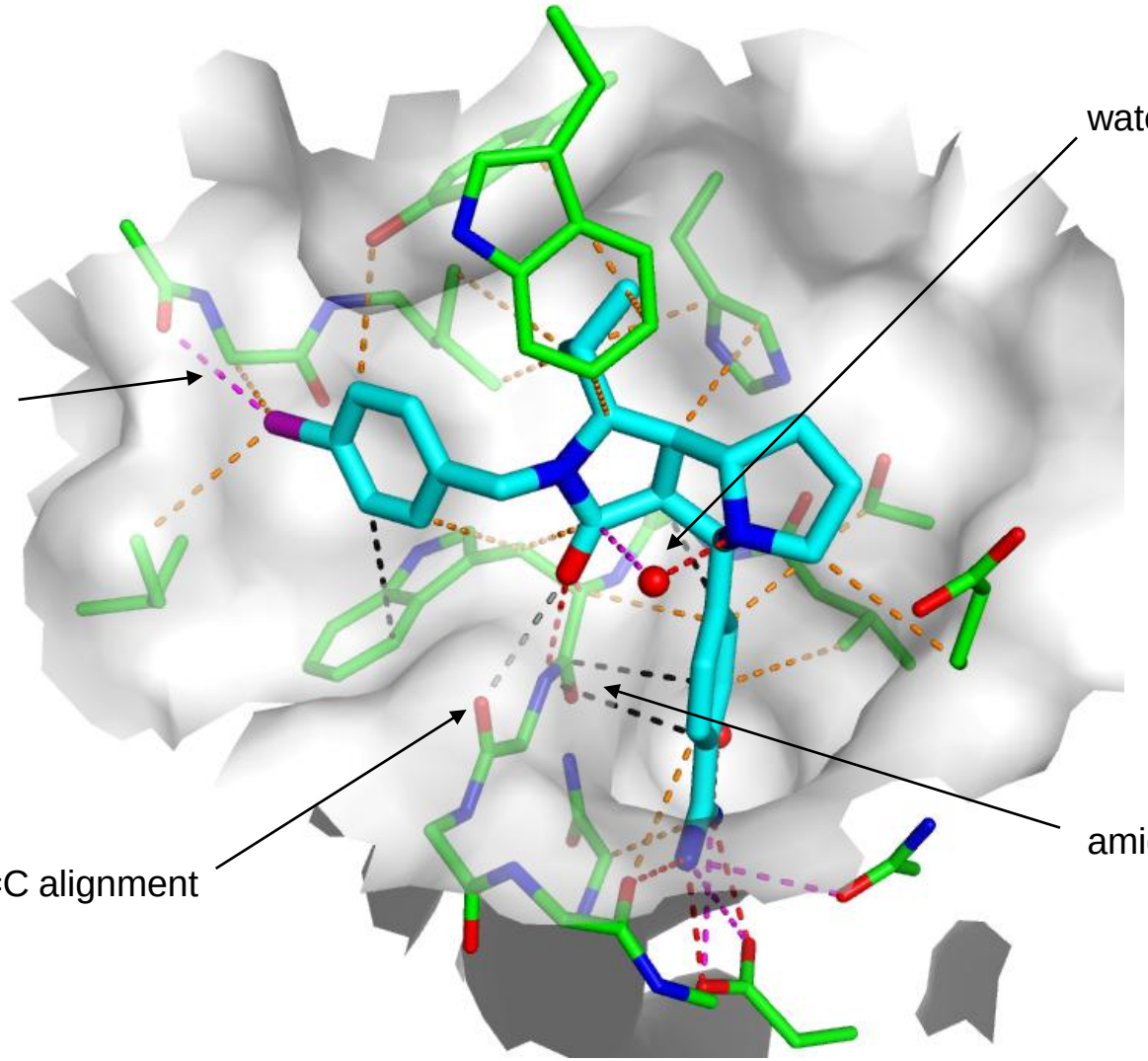
Network edges: indentifcation of favourable and unfavourable interactions using ViewContacts

Implement a broader view of non-covalent interactions

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

ViewContacts: example

Ex. 2cf8 (thrombin)



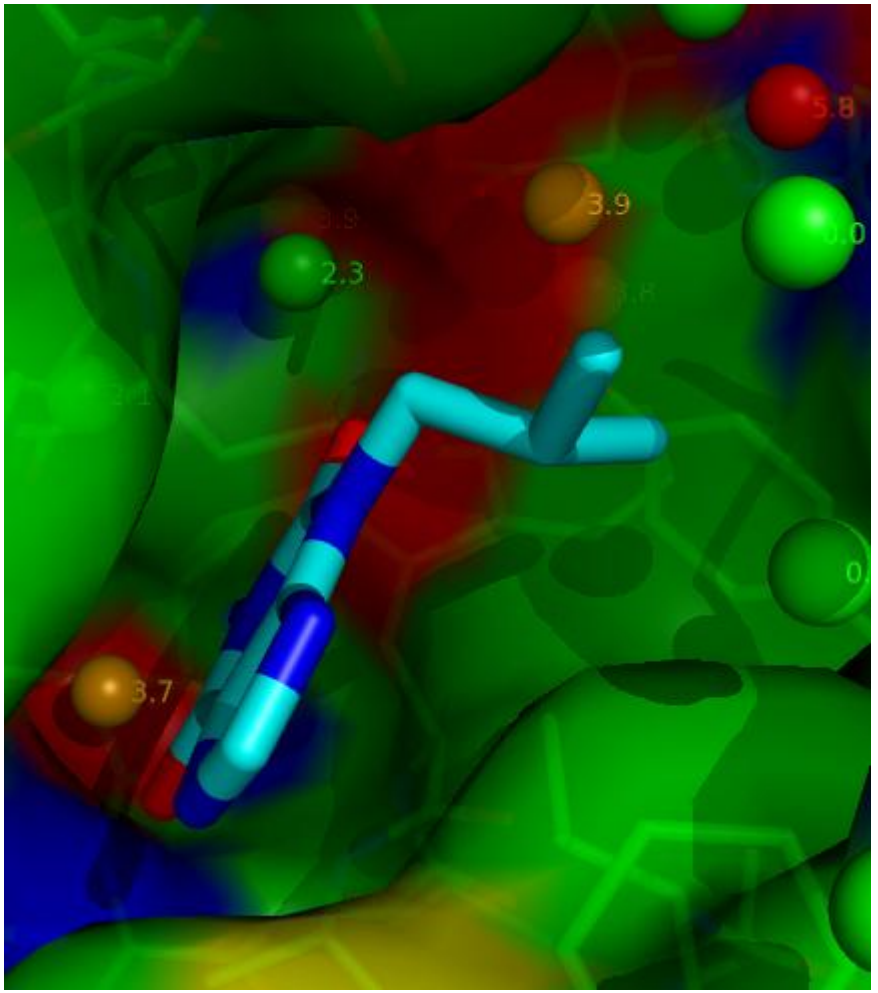
ViewContacts: handling of water molecules

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

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

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

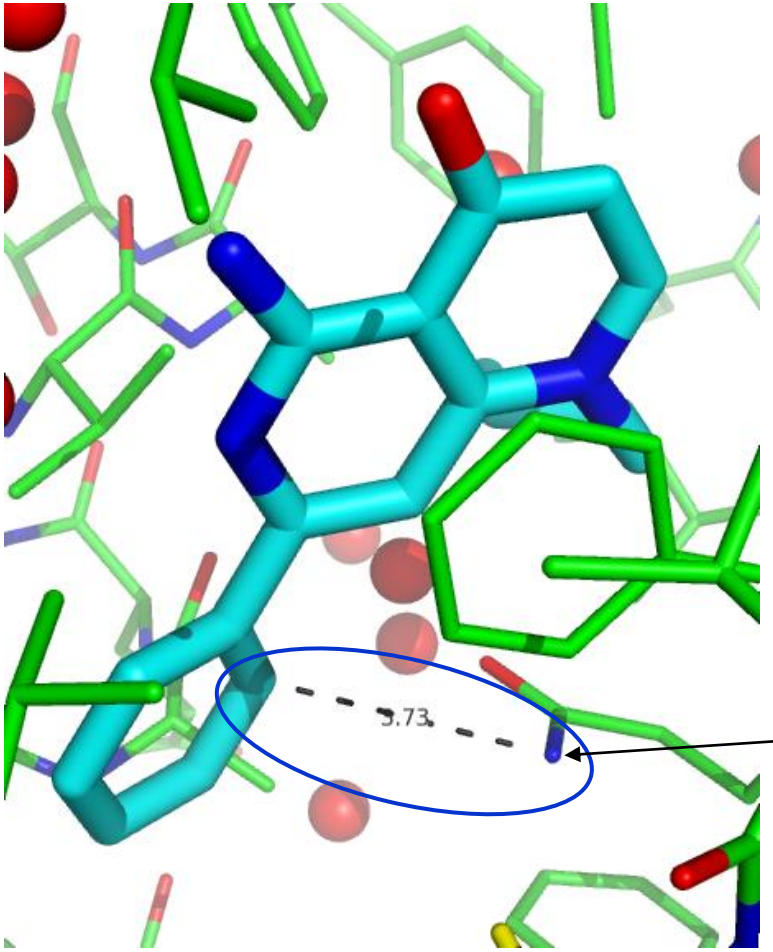
Ex. 2r8q (PDE-B1)



Water molecules with Rank scores ≥ 2.0 are included in networks

ViewContacts: identification of unfavourable interactions

Unfavourable contact if an apolar ligand atom replaced by water molecule fulfills hydrogen bonding requirements



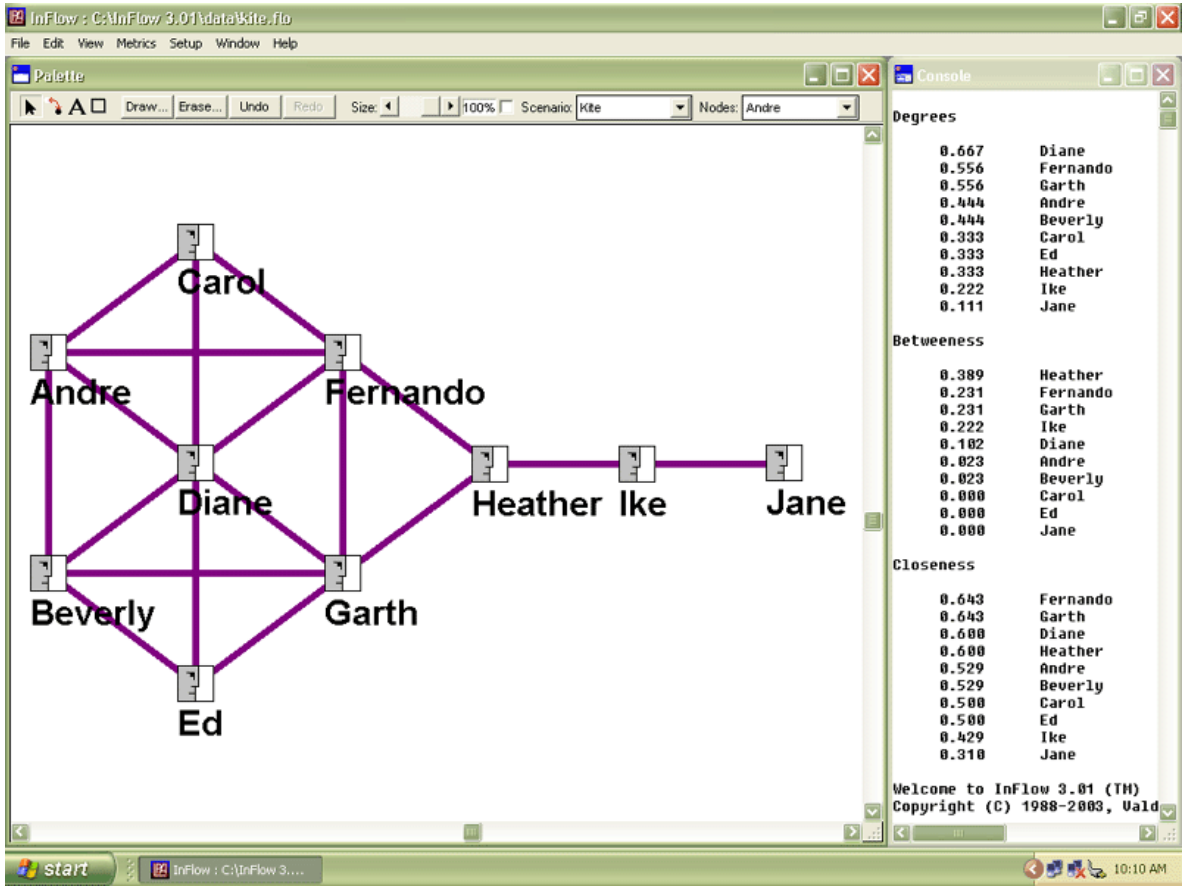
Ex. 3hdz (PDE-5)

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

Standard small world network (SWN) model

Initially explored using descriptors from Social Network Analysis



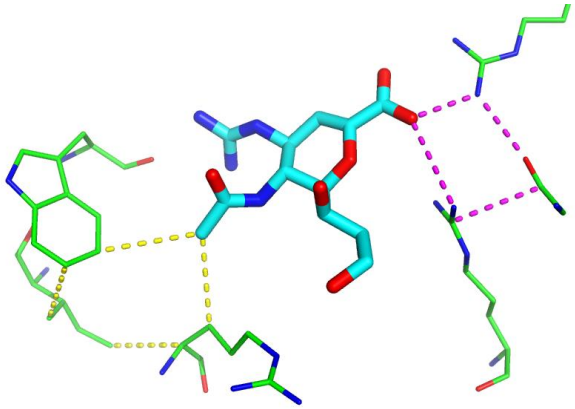
Kite Network, by D. Krackhardt

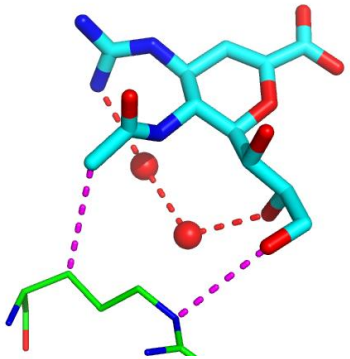
http://www.orgnet.com/sna.html

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

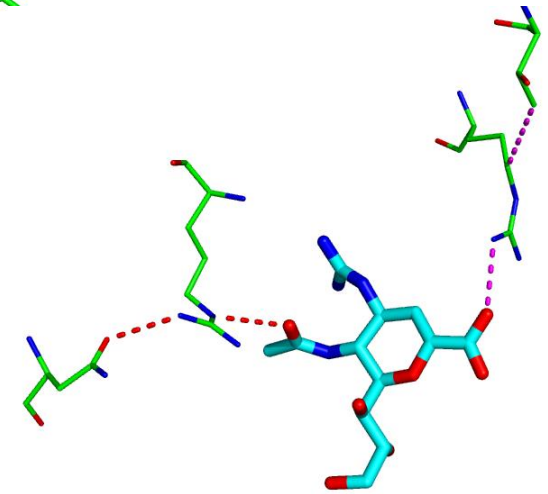
- ligand-protein-ligand (LPL) network elements
  - ligcycles (involving 1 ligand atom) ligloops (involving ≥ 2 ligand atoms)





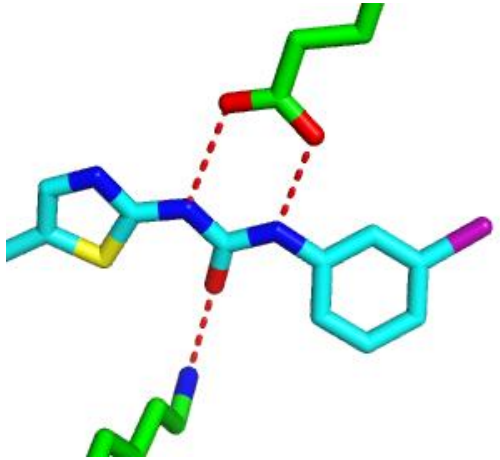
examples from 1nnc

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

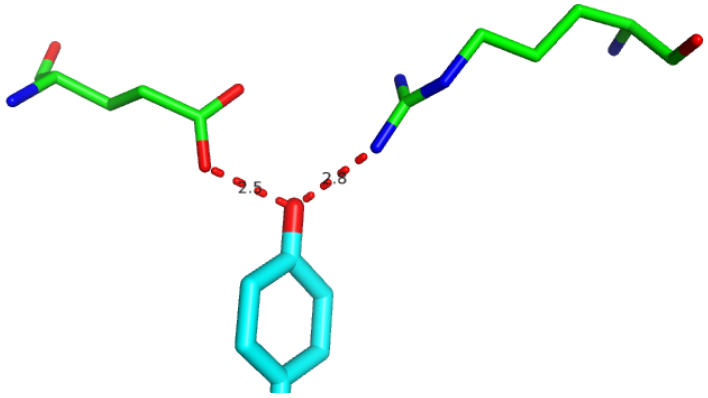


Network descriptors: special treatment of hydrogen bonding

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

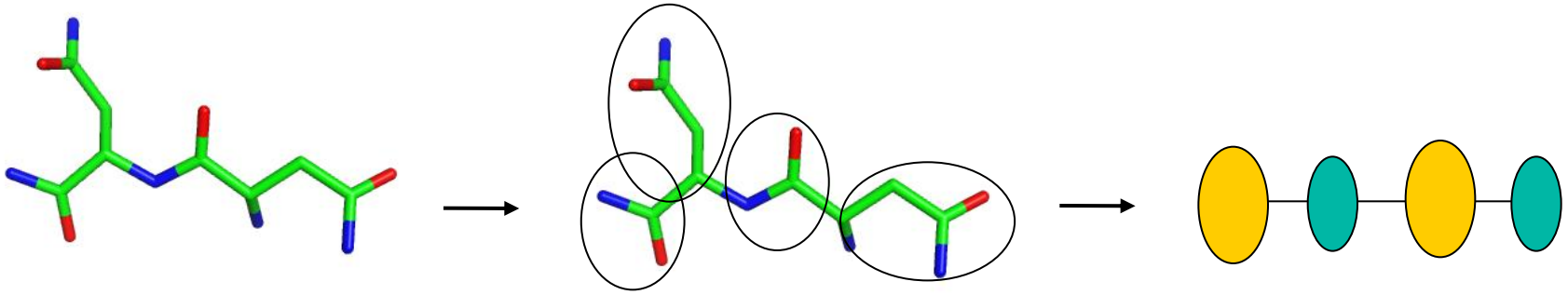


- protein-ligand-protein (PLP)
  - with lower free energy than the sum of the individual bonds due to mutual polarization



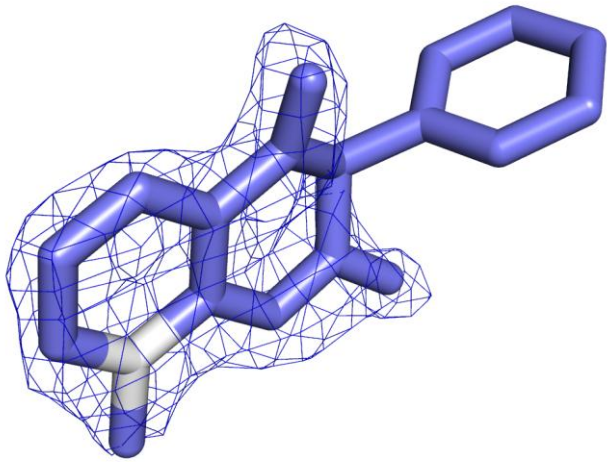
Network descriptors: nodes based on a reduced graph definition of protein structure

Protein structure is treated as a collection of small groups of atoms (functional groups)



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

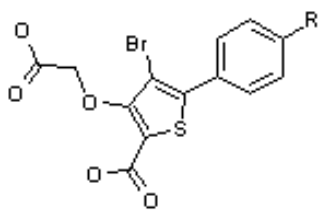
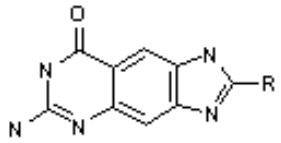
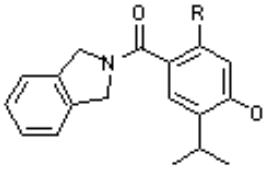


Electon density correlation coefficient is a better measure of model quality than B-factors

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 complexes

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

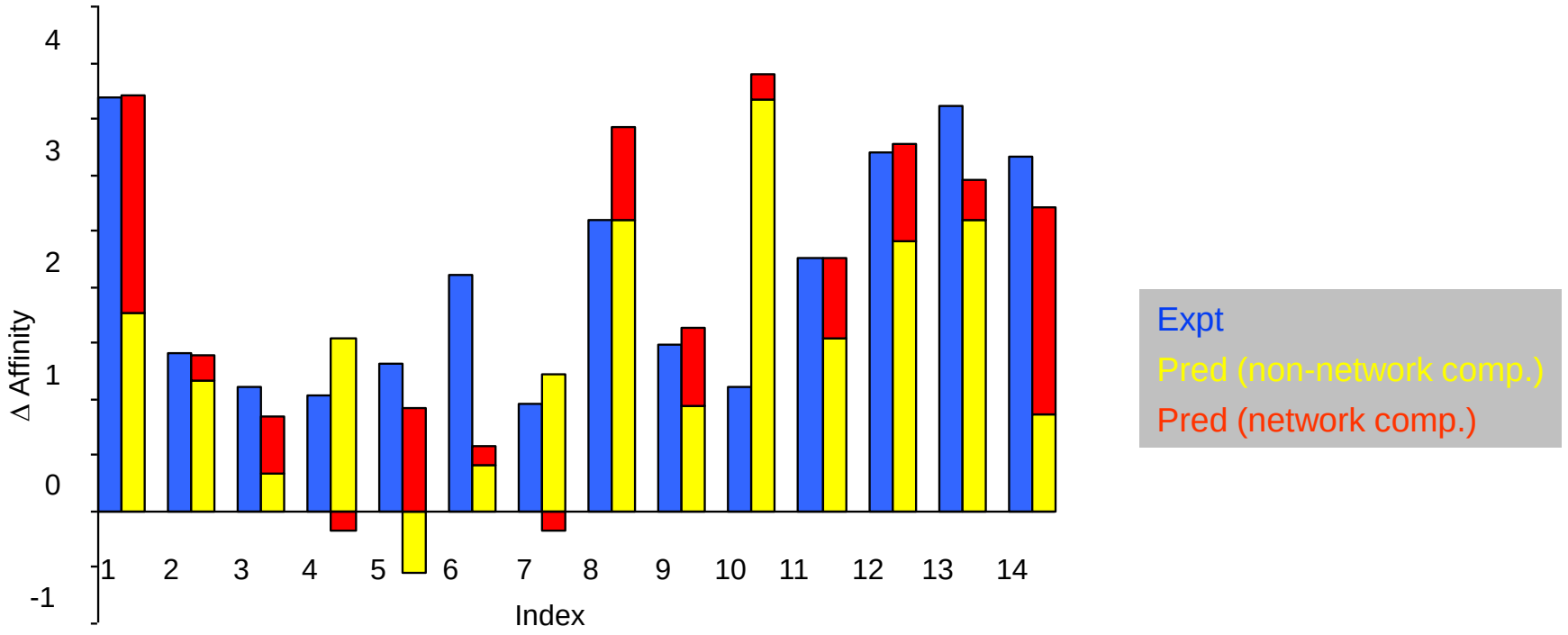
- form of scoring function:

$$S = \sum_{n} f\left(Int\right)$$ (without network terms)

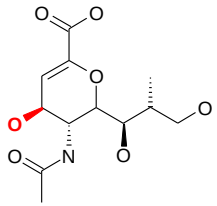
$$S_{Scorpion} = \sum_{n} f\left(Int\right) + \sum_{m} g\left(Int\_nw\right)$$

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

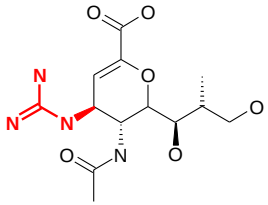
Activity cliffs: predicted vs. experimental energy differences



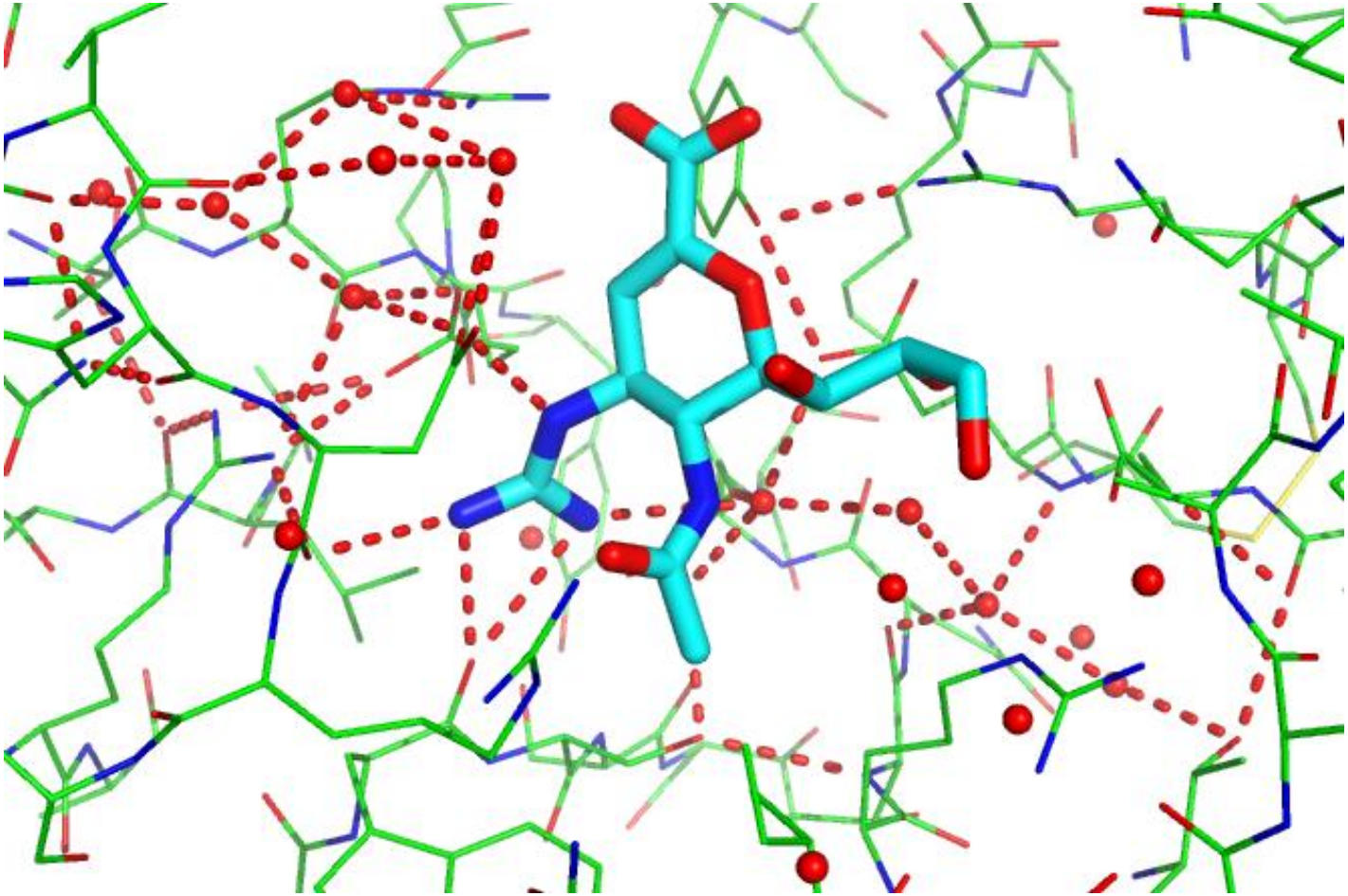
Activity cliffs: Neuraminidase example



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



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



Scorpion Score

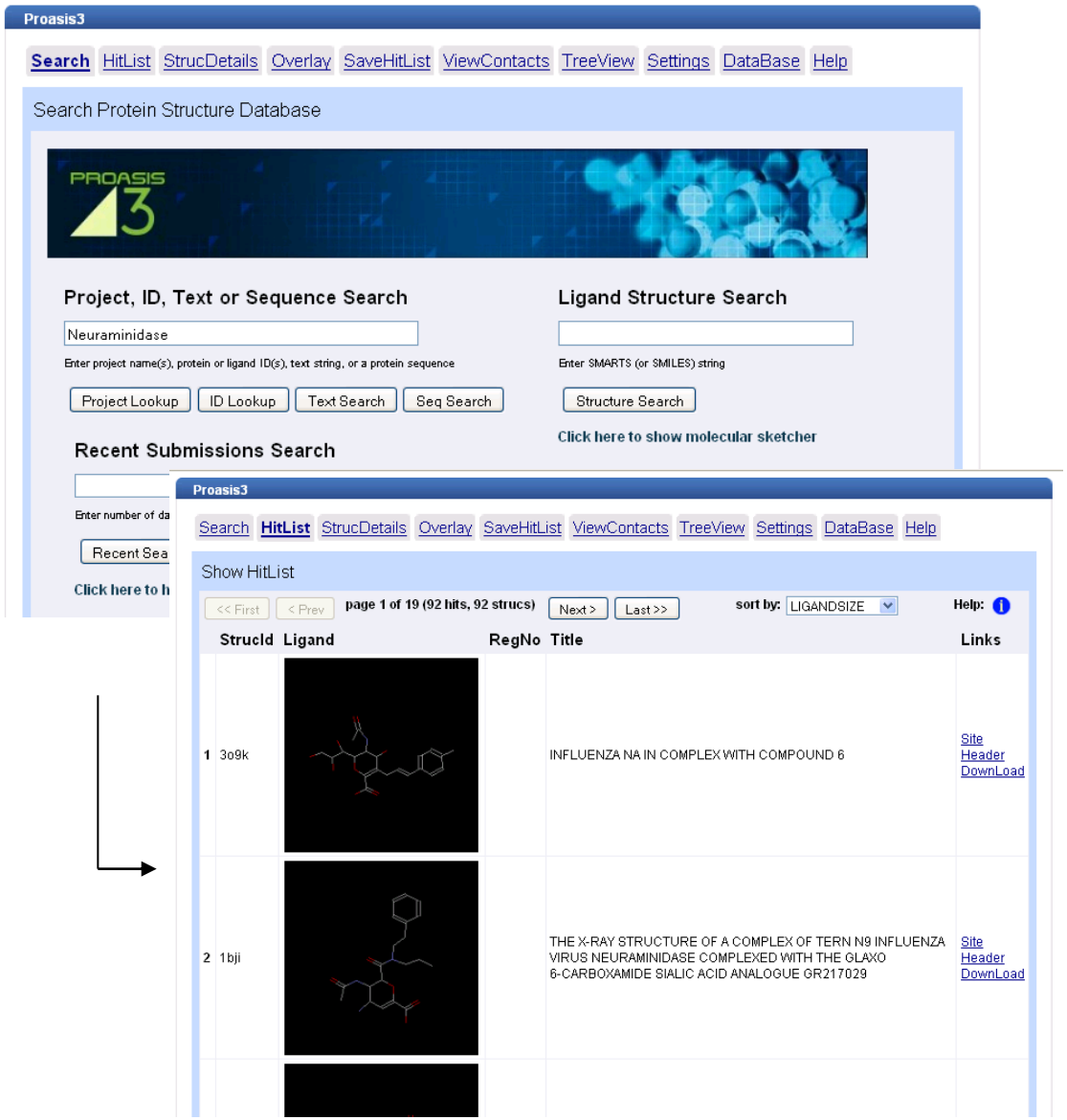
$$S_{Scorpion} = 0.473\,\mathrm{x}\,[\mathrm{hbond}] + 0.129\,\mathrm{x}\,[\mathrm{hbond\_nw}]$$
$$+ 0.516\,\mathrm{x}\,[\mathrm{vdw}] + 0.387\,\mathrm{x}\,[\mathrm{vdw\_nw}]$$
$$+ 0.188\,\mathrm{x}\,[\mathrm{pi\text{-}pi}] + 0.931\,\mathrm{x}\,[\mathrm{pi\text{-}pi\_nw}]$$
$$+ 0.285\,\mathrm{x}\,[\mathrm{cat\text{-}dipole}]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

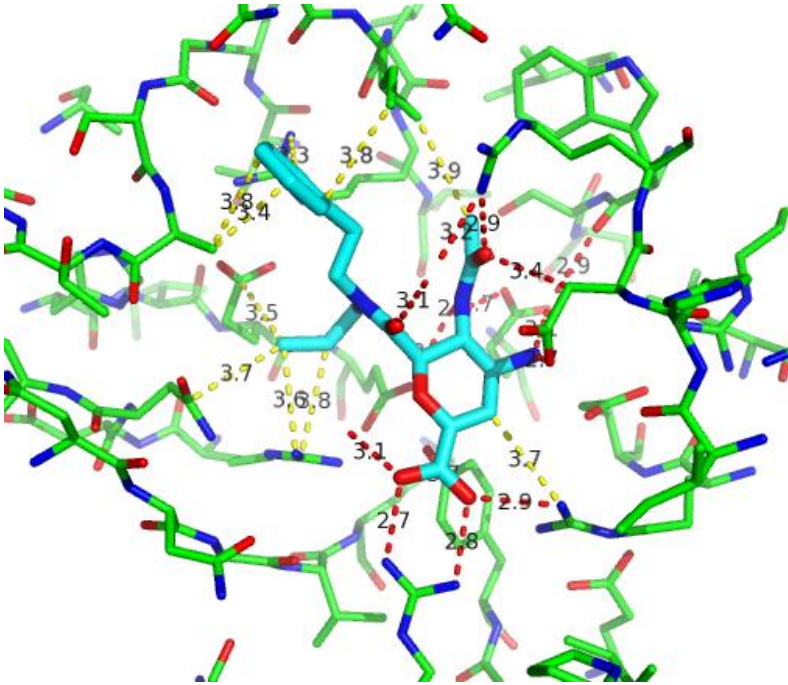
H-bond donor – nonpolar contact

Results shown from optimisation done back in 2010

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

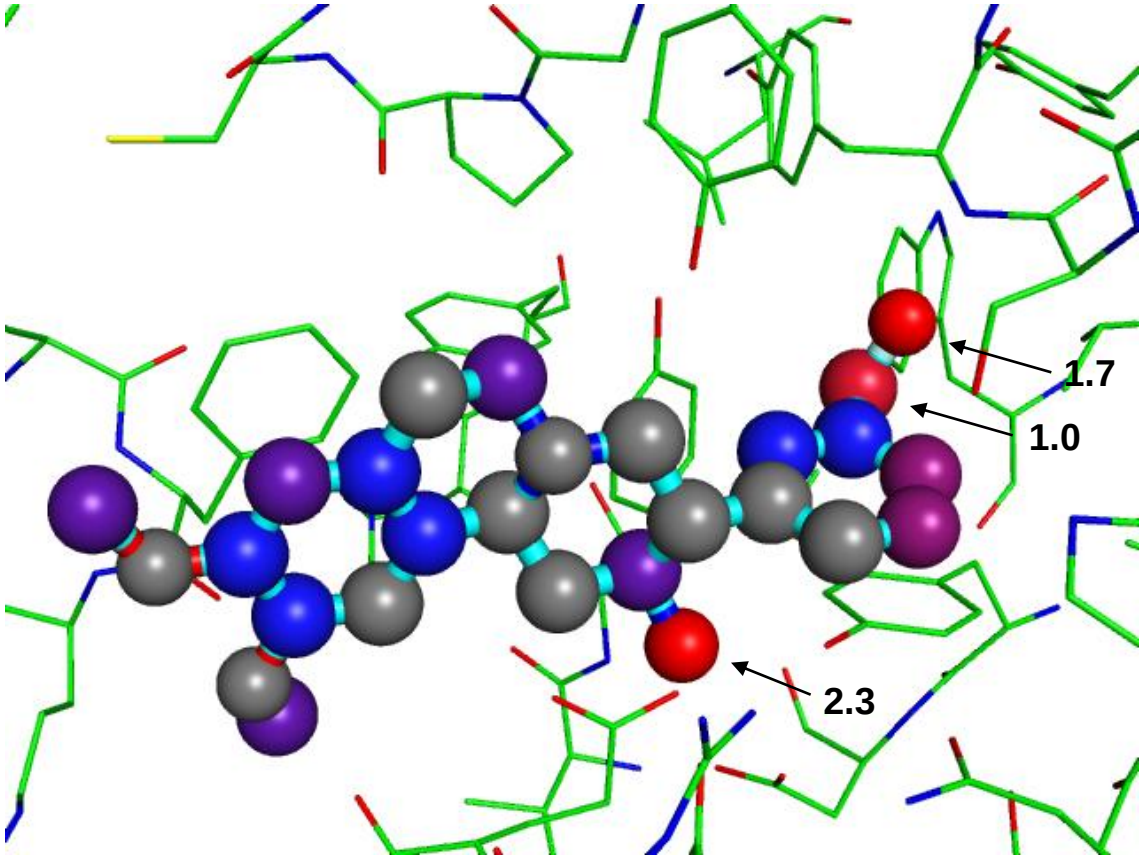
Quick and easy visualisation





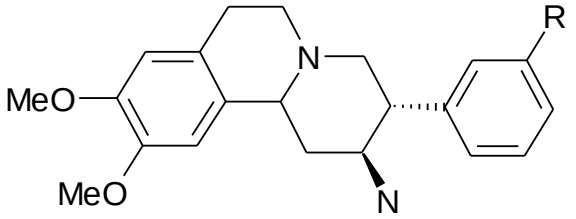
… at the click of a button

Score contributions mapped onto atoms



3kwj

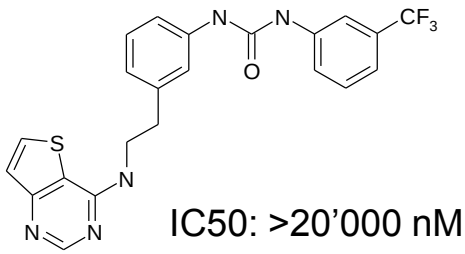
(dpp-4)



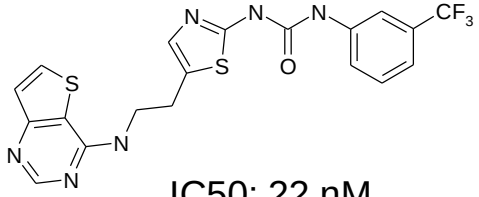
R	K_d (nM)	$S_{Scorpion}$	network
H	200	7.9	0.8
CH_3	4.6	10.8	1.6
CH_2F	0.5	11.3	1.7

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

Aurora A kinase inhibitors

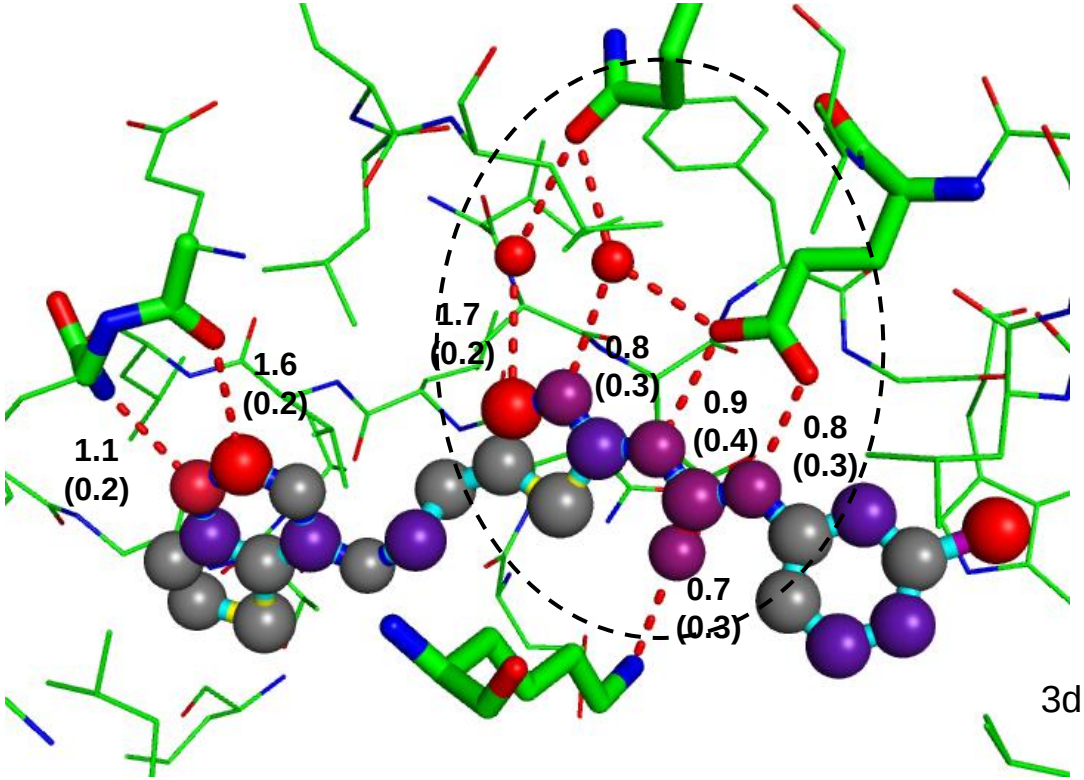


IC50: >20’000 nM



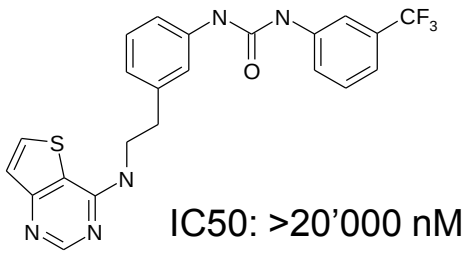
IC50: 22 nM

networked H-bonds with high score
incl. network contribution

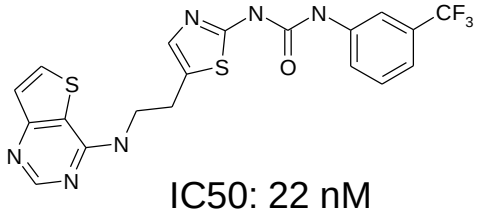


3d15

Aurora A kinase inhibitors (cont.)

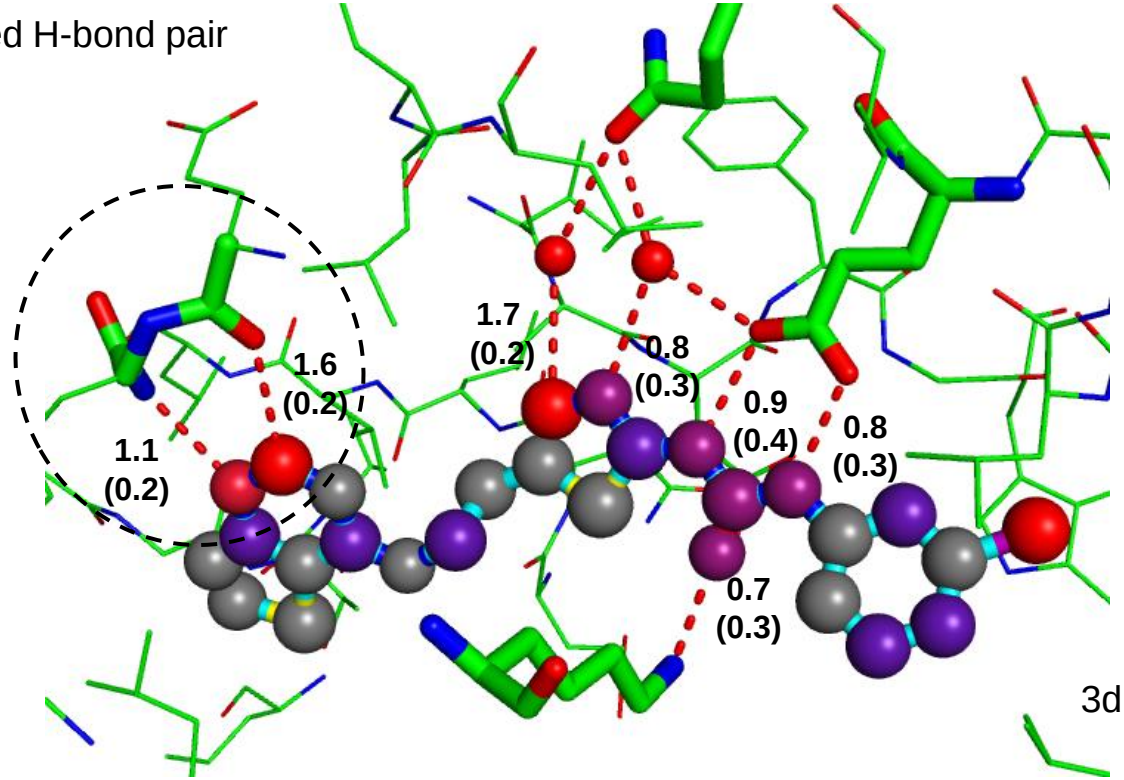


IC50: >20’000 nM



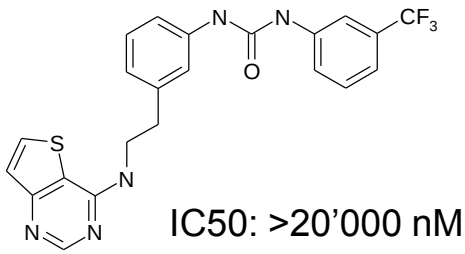
IC50: 22 nM

privileged H-bond pair

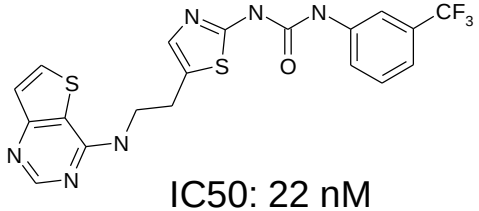


3d15

Aurora A kinase inhibitors (cont.)

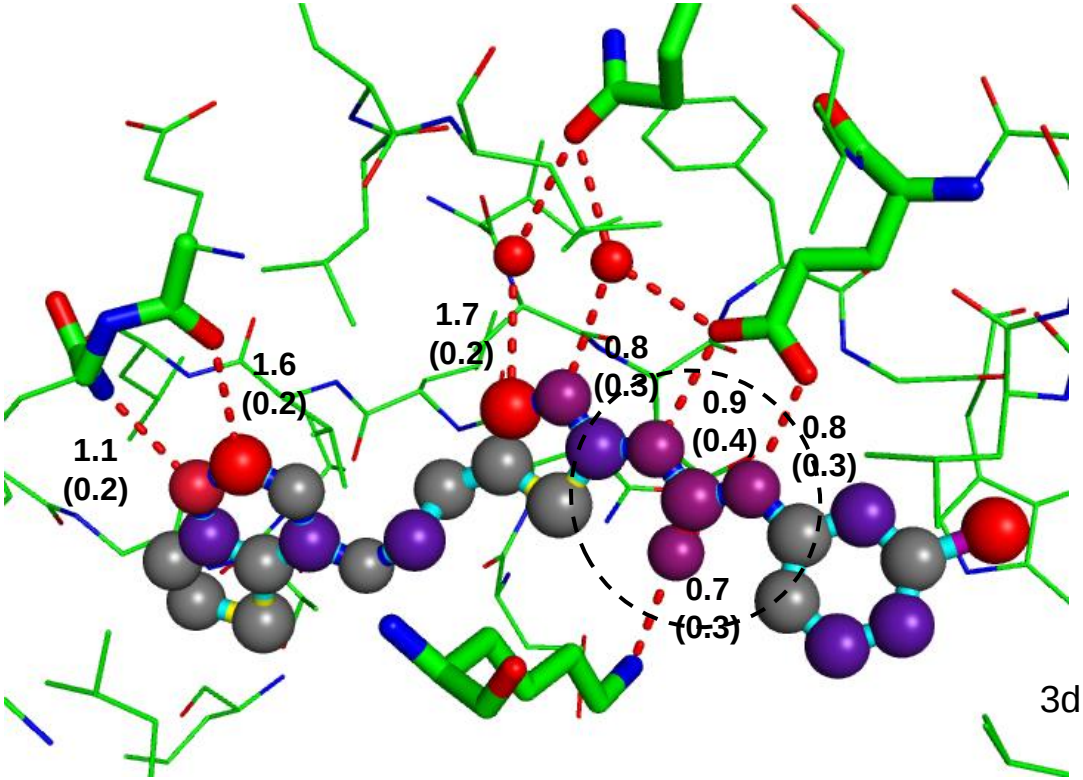


IC50: >20’000 nM



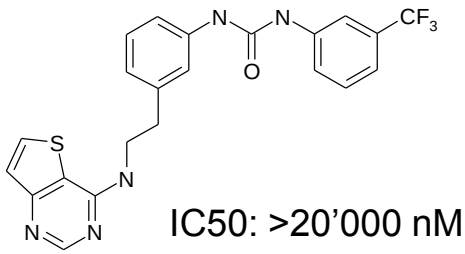
IC50: 22 nM

more privileged H-bond pairs

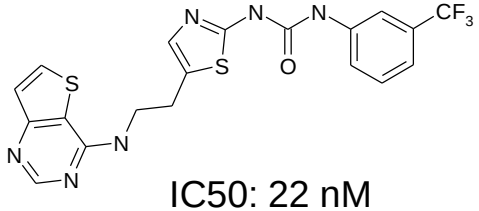


3d15

Aurora A kinase inhibitors (cont.)

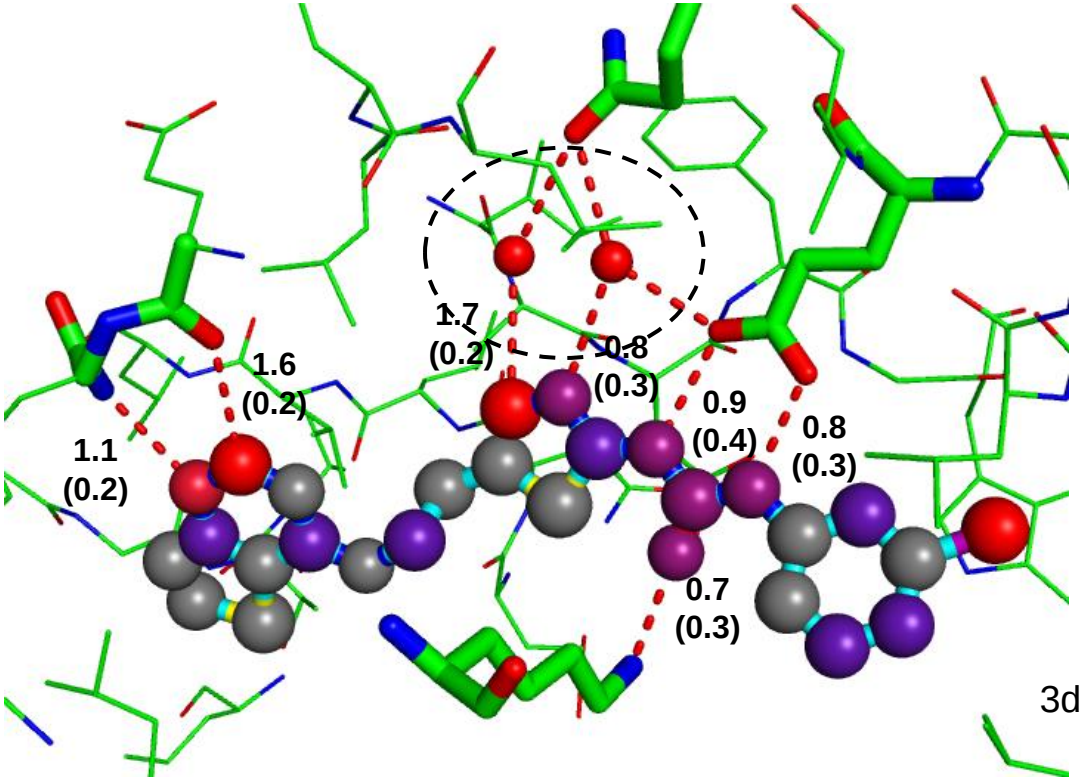


IC50: >20’000 nM



IC50: 22 nM

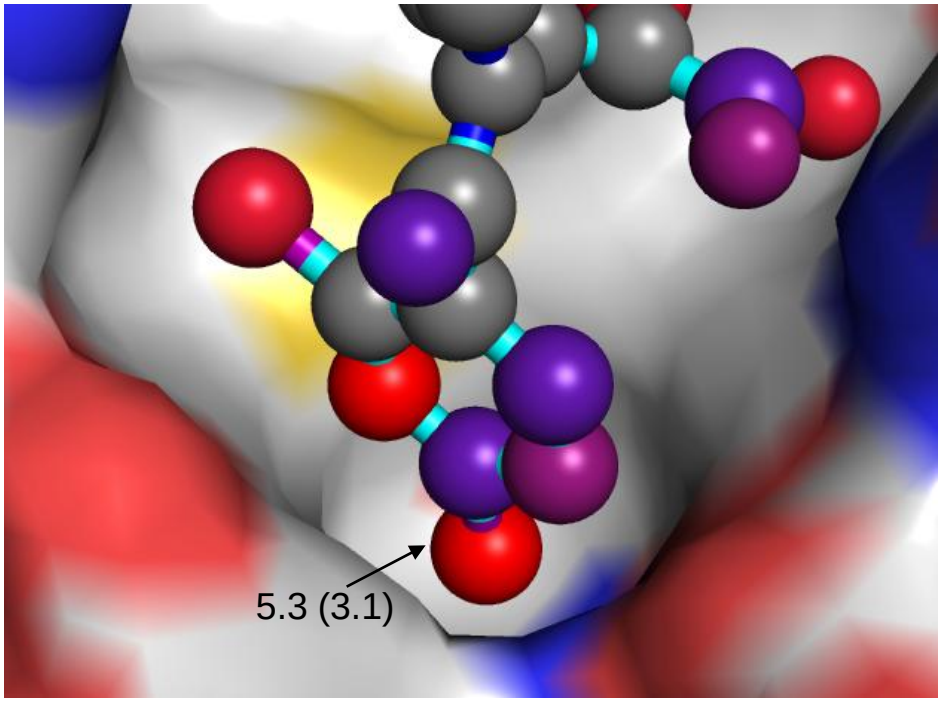
tightly bound waters play
important role in networks



3d15

Examples of highly networked atoms

Atoms in buried pockets with several contacts receive extra network contribution



1yvz

(hcv polymerase)

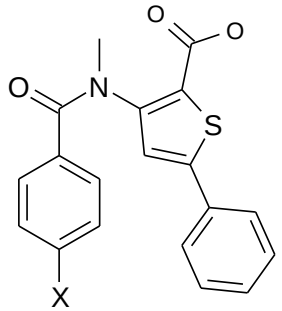
other examples:

1ql7 (trypsin): Cl in S1 pocket

2j4i (FXa): Cl in S1 pocket

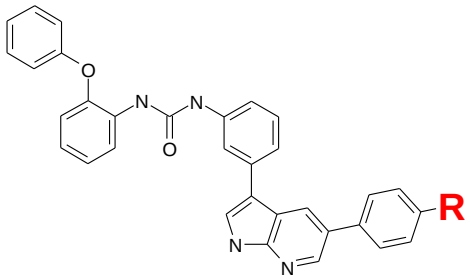
2r3r (cdk2): Br

R:	IC50[uM]
-H	>32
-Cl	1.2



Insulin receptor kinase – pyrrolopyridine complex

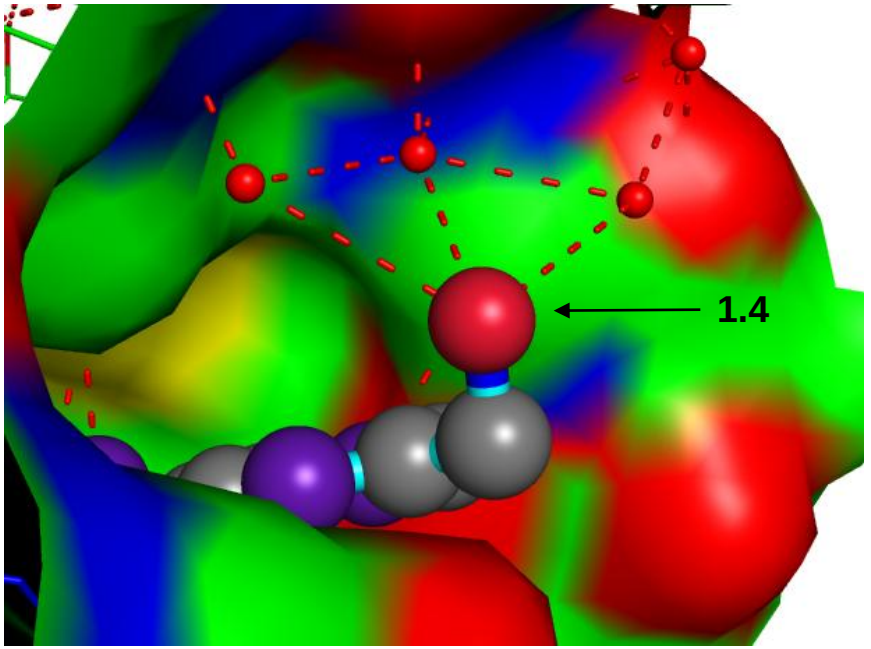
Ligand atoms can have high network scores in spite of being highly solvent-exposed



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

-aminomethyl group solvent exposed with no direct contact with protein

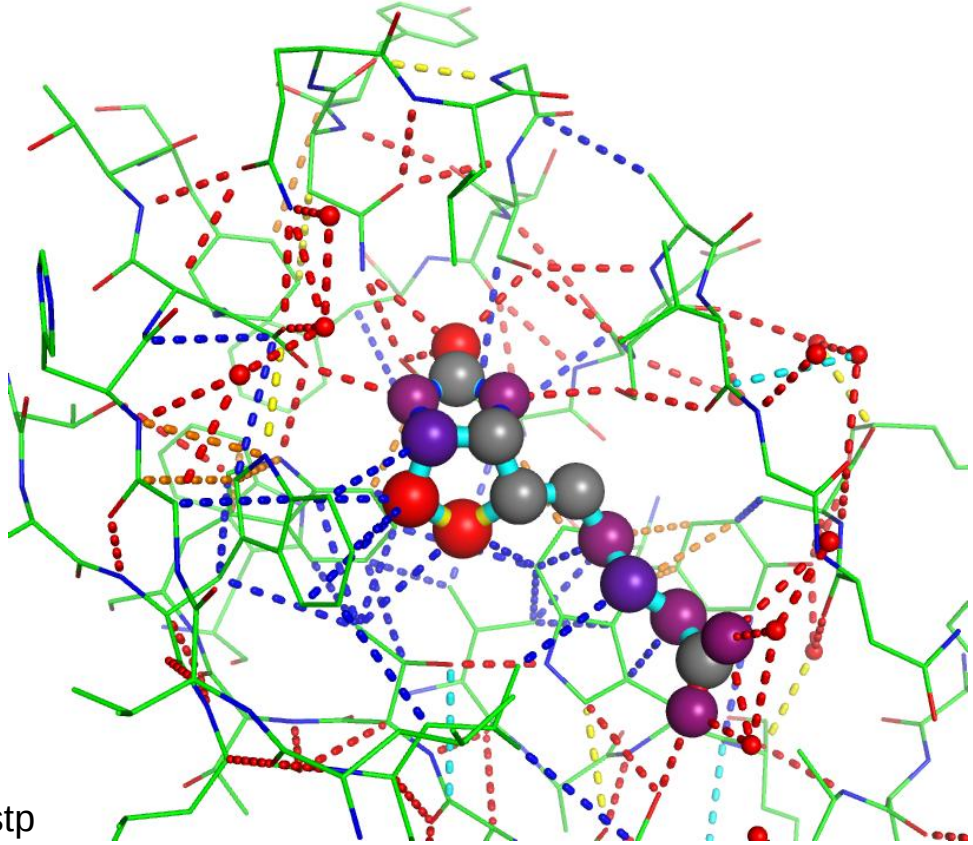
- amino group interacts through protein-bound water molecules with insulin receptor resulting in a high score despite low buriedness



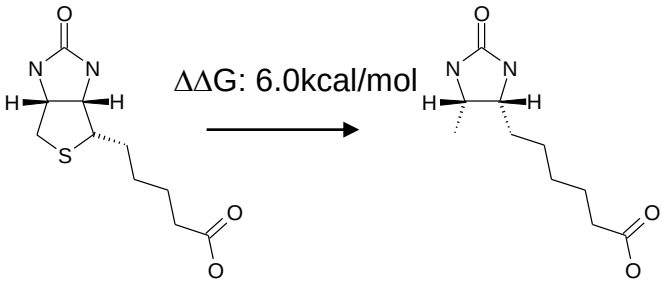
3eta

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

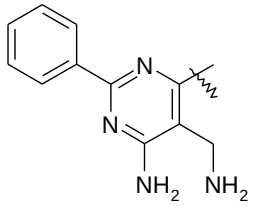


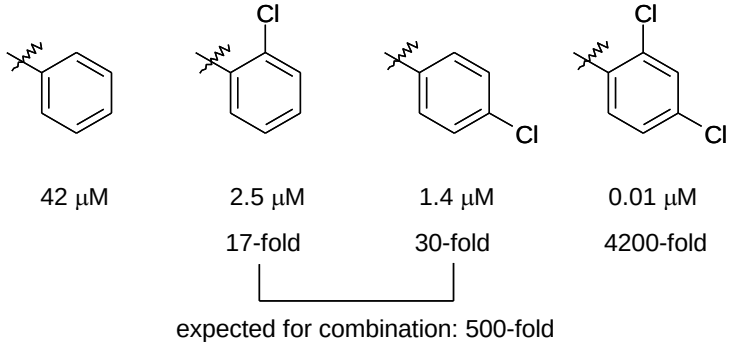
“The streptavidin/biotin system provides a clear example where the binding affinity is the propertry of the whole system”

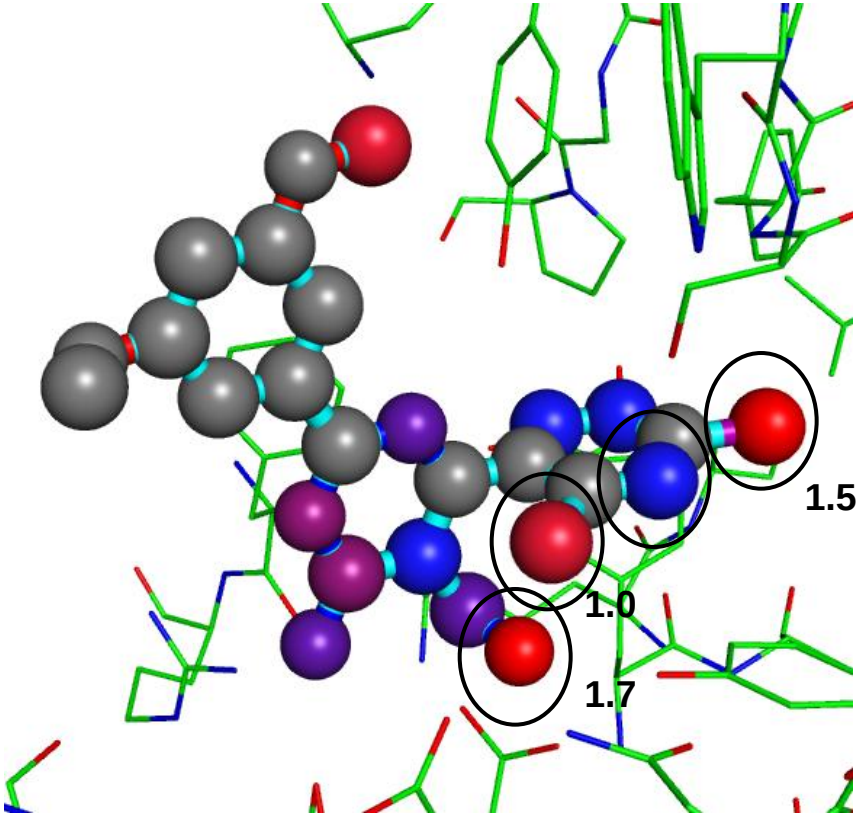
Williams et. al., Angew. Chem. Int. Ed, 2004, 43, 6596

Cooperativity pairs - DPP4

steep and non-additive
SAR in DPP-4:







1rwq

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

Acknowledgements

- Bernd Kuhn*
- Martin Stahl
- Michael Reutlinger for support with generation of the high-quality data sets
- Julian Fuchs for support with genetic algorithm optimisation
- Wolfgang Guba, and other molecular modeling staff from Roche, Basel, for help testing software
- Annabelle Taylor for many fruitful discussions