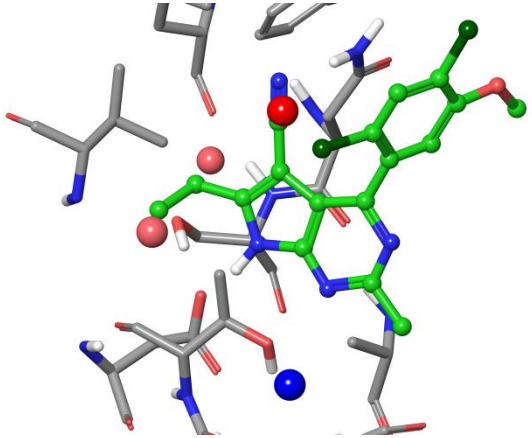
UNIVERSITY OF CAMBRIDGE

Cavendish Laboratory

"Water Networks and Protein Target Ligandability"

David J. Huggins

May 12th 2016

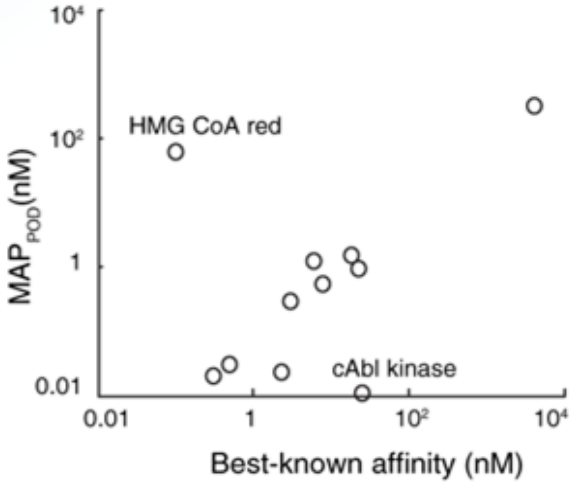


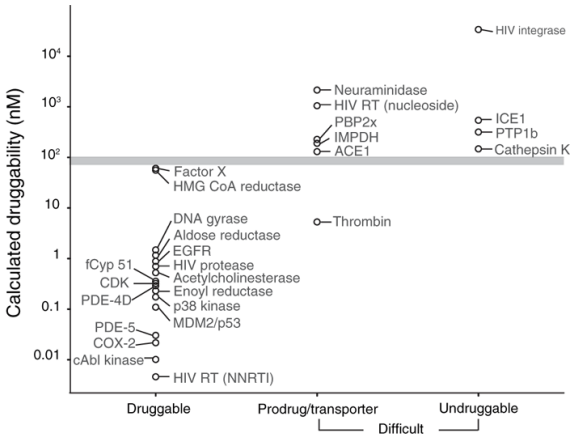
Overview

- Druggability and Ligandability
- Rationale for studying water
  - Protein hydration sites and water networks
  - Water and small-molecule binding
- Theory and methods
  - Inhomogeneous fluid solvation theory (IFST)
  - Combinatorial subgraph search (CSS)
- Validation
  - Water in protein cavities
  - Small-molecule hydration
- Applications
  - Predicting water displaceability (HSP90)
  - Ligandability assessment (Bromodomains)

Druggability and Ligandability

- Druggability can be defined as the relative ease or difficulty of developing a small molecule that can effectively modulate a protein's activity in-vivo.
- In the past, druggability has been quantified in terms of the maximal affinity of small-molecule inhibitors.





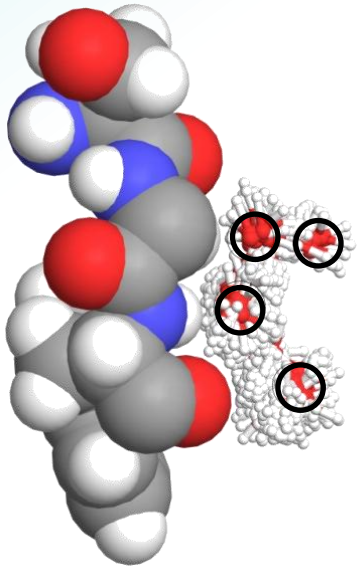
- However, the complex PK/PD issues influence the ability of a small molecule inhibitor to be effective as a drug.
- Ligandability can be defined as the relative ease or difficulty of developing a small molecule that can effectively modulate a protein's activity in-vitro.

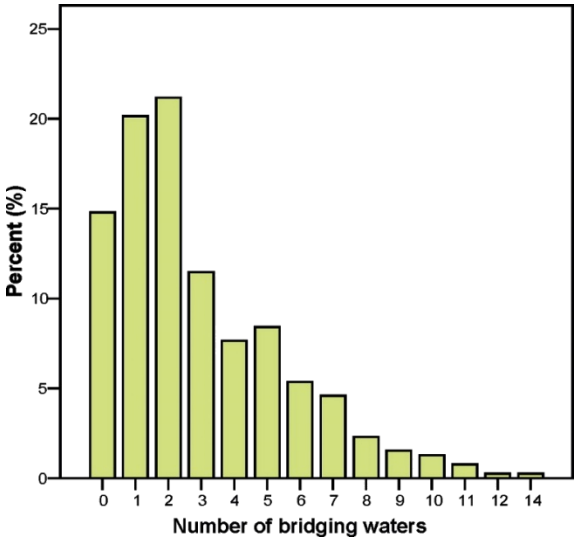
I. D. Kuntz, K. Chen, K. A. Sharp, and P. A. Kollman. "The maximal affinity of ligands." PNAS, 1999 - 96(18)

Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC, Huang ES "Structure-based maximal affinity model predicts small-molecule druggability." Nature Biotechnology, 2007 - 25

Rationale

- Water molecules surround proteins and typically cluster at specific locations at the protein surface, termed hydration sites.



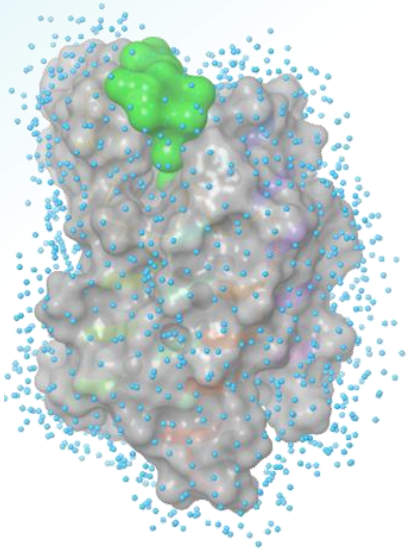


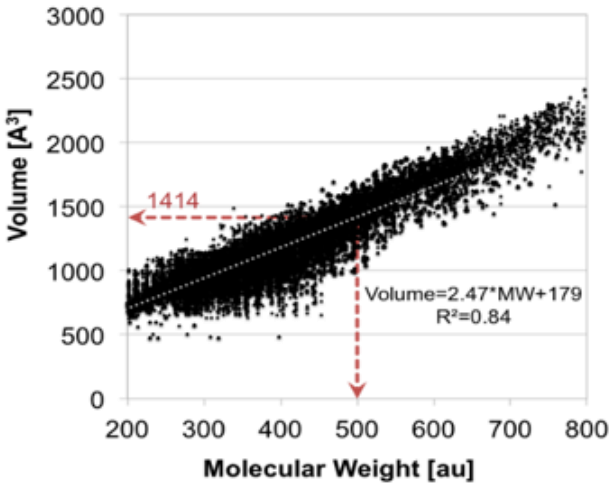
- A 300 residue globular protein typically has around 2000 hydration sites in the first solvation shell.

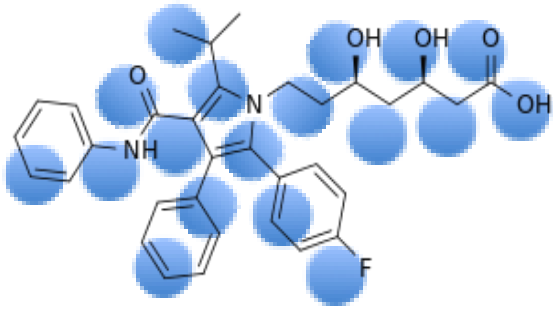
Lu, Y.P., et al. "Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes." Journal of Chemical Information and Modeling, 2007. 47(2): p. 668-675.

Rationale

- Ligand binding is accompanied by unbinding of water molecules from the protein surface.







MW 559 N=18

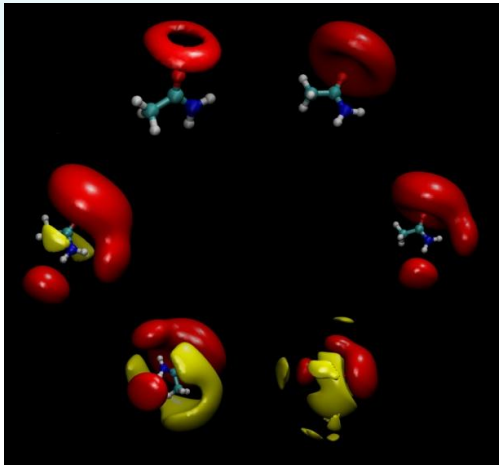
- This leads to a desolvation penalty which disfavours ligand binding.

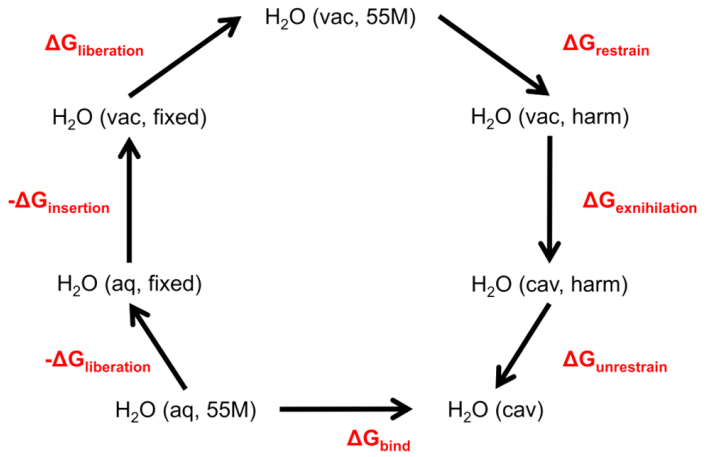
Key questions:

- Can we make accurate estimates of how strongly bound a water molecule is?
- Does ligand binding coincide with displacement of weakly bound water molecules?
- Can we relate the strength of surface water networks with target ligandability?

Methods - Free-Energy Calculations

- We perform all-atom molecular dynamics simulation of proteins in water.





$$\Delta G_{IFST} = \Delta E_{IFST} - T\Delta S_{IFST}$$

$$\Delta E_{IFST} = E_{sw} + E_{ww} - E_{bulk}$$

$$\Delta S_{IFST} = S_{sw} + S_{ww} - S_{bulk}$$

$$\Delta G_{FEP} = \sum_{a=1,b=a+1}^{N} \Delta G_{a\to b}$$

$$\Delta G_{a\to b} = -kT\, ln(\langle exp(-(H_b - H_a)/kT)\rangle_a)$$

- Binding free energies are calculated using inhomogeneous fluid solvation theory (IFST) and free-energy perturbation (FEP).

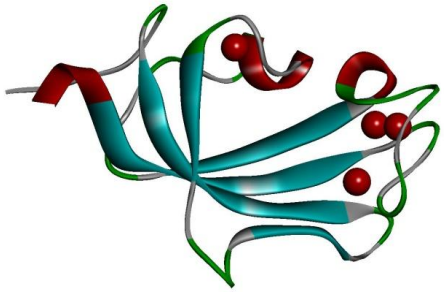
D. Hamelberg and J.A. McCammon "Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method" - J. Am. Chem. Soc., 126 (2004) 7683-7689.

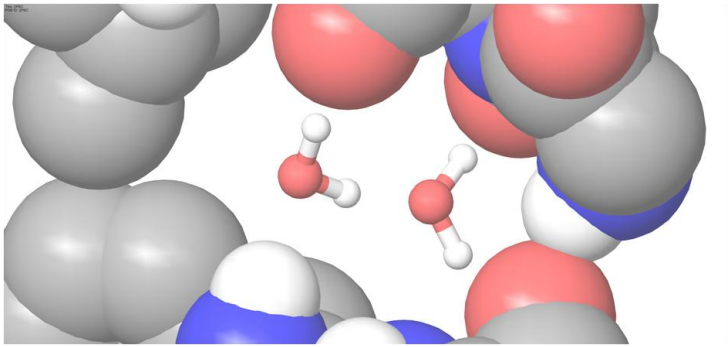
Huggins. "Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations" Biophys. J. - 2015, 108, pp 928-936

Validation

- Can we make accurate estimates of how strongly bound a water molecule is?

Protein	IL-1β	T4 Lysozyme	FKBP-2	CA-II	β-Lactamase
PDBID	2NVH	3DKE	2PBC	3GZ0	2P74
Protein Chain	A	X	A	A	A
Resolution (Å)	1.53	1.25	1.8	1.26	0.88
Single Cavities	2	2	1	5	5
Double Cavities	2	1	1	0	0





- Prediction of binding free energy for water molecules in protein cavities.

Huggins. "Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations" Biophys. J. - 2015, 108, pp 928-936

Validation

- Results for IFST and FEP agree very well for protein cavities.

System	N	ΔG_{FEP} (kcal/mol)	ΔG_{IFST} (kcal/mol)	Signed Difference (kcal/mol)	Unsigned Difference (kcal/mol)
IL-1β	2	-11.77	-11.27	-0.49	0.49
IL-1β	2	-6.18	-5.76	-0.42	0.42
IL-1β	1	-7.09	-7.57	0.49	0.49
IL-1β	1	-6.90	-6.87	-0.03	0.03
T4 Lysozyme	2	-20.41	-19.52	-0.89	0.89
T4 Lysozyme	1	-3.33	-3.00	-0.33	0.33
T4 Lysozyme	1	-8.29	-7.76	-0.53	0.53
FKBP-2	2	-16.70	-16.13	-0.57	0.57
FKBP-2	1	-13.07	-12.50	-0.57	0.57
CA-II	1	-8.21	-7.62	-0.59	0.59
CA-II	1	-4.29	-4.17	-0.12	0.12
β-Lactamase	1	-2.31	-2.55	0.24	0.24
β-Lactamase	1	-16.53	-16.03	-0.51	0.51
			Mean	**-0.31**	**0.45**

- Water molecules are predicted to have substantial binding free energies.

Huggins. "Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations" Biophys. J. - 2015, 108, pp 928-936

Validation

- Dunitz' analysis of experimental data on inorganic crystals suggested that the entropic cost of transferring from the bulk will not typically be greater than 7.0 cal/mol/K.

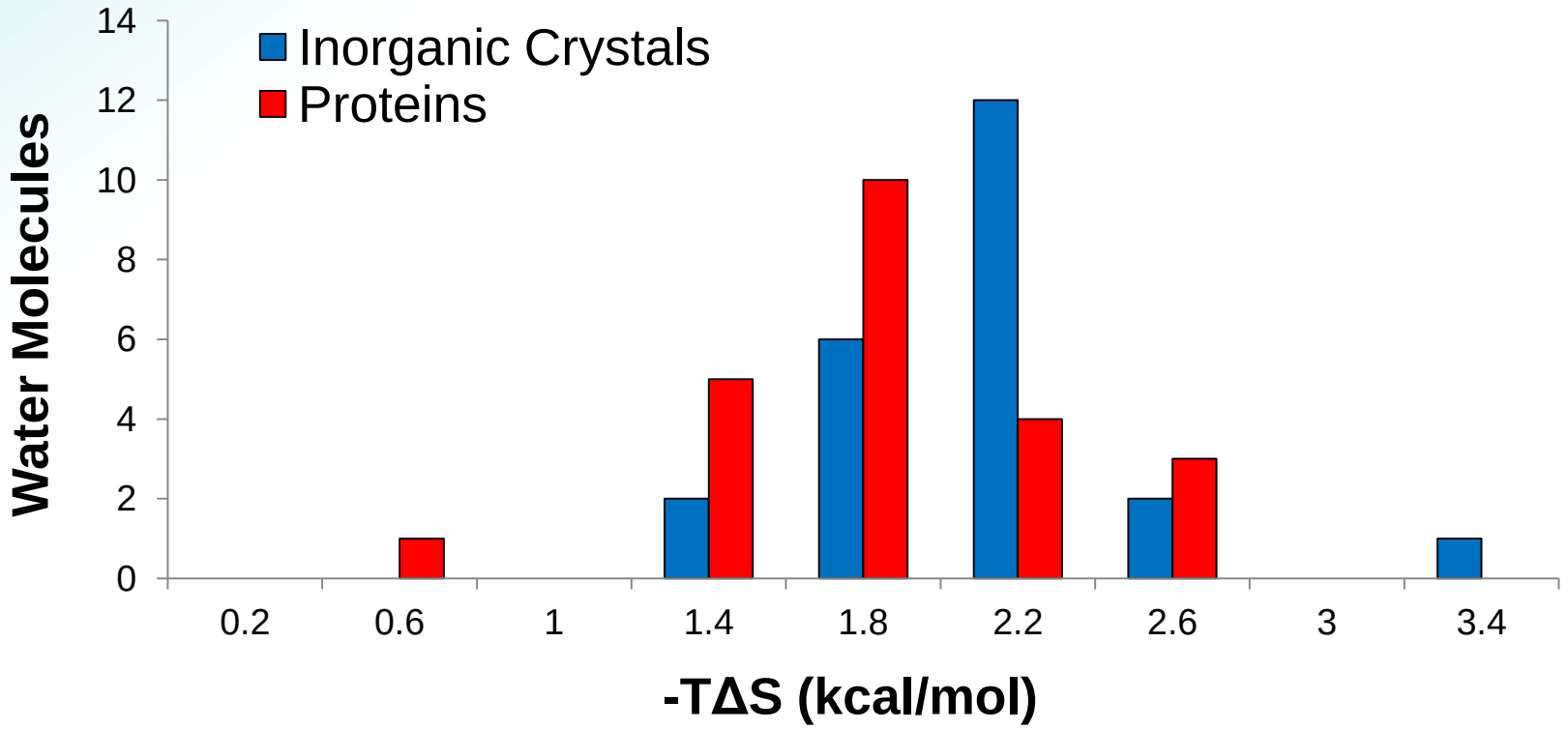
Salt	Waters	ΔS per water (cal/mol/k)	-TΔS per water (kcal/mol)	Salt	Waters	ΔS per water (cal/mol/k)	-TΔS per water (kcal/mol)
ZnSO4	1	-11.6	3.5	NiSO4	6	-7.6	2.3
	6	-7.2	2.1	Al2(SO4)3	6	-7.5	2.2
	7	-7.8	2.3	NH4AI(SO4)2	12	-7.1	2.1
CdCl2	1	-4.2	1.3	MgCI2	1	-5.3	1.6
	5/2	-5.8	1.7		2	-5.9	1.8
CdBr2	4	-6.0	1.8		4	-6.3	1.9
CdSO4	1	-8.4	2.5		6	-5.7	1.7
	8/3	-7.3	2.2	Ca(NO3)2	2	-7.6	2.3
CuSO4	1	-8.0	2.4		3	-7.4	2.2
	3	-7.8	2.3		4	-8.0	2.4
	5	-7.5	2.2	Na2SO4	10	-6.1	1.8
NiCl2	6	-8.4	2.5	KAI(S04)2	12	-7.1	2.1

- This corresponds to a contribution of approximately +2.0 kcal/mol to the free energy.

Dunitz, J. D. "The entropic cost of bound water in crystals and biomolecules" Science – 1994, 264:670

Validation

- The estimates are in good agreement with Dunitz prediction.



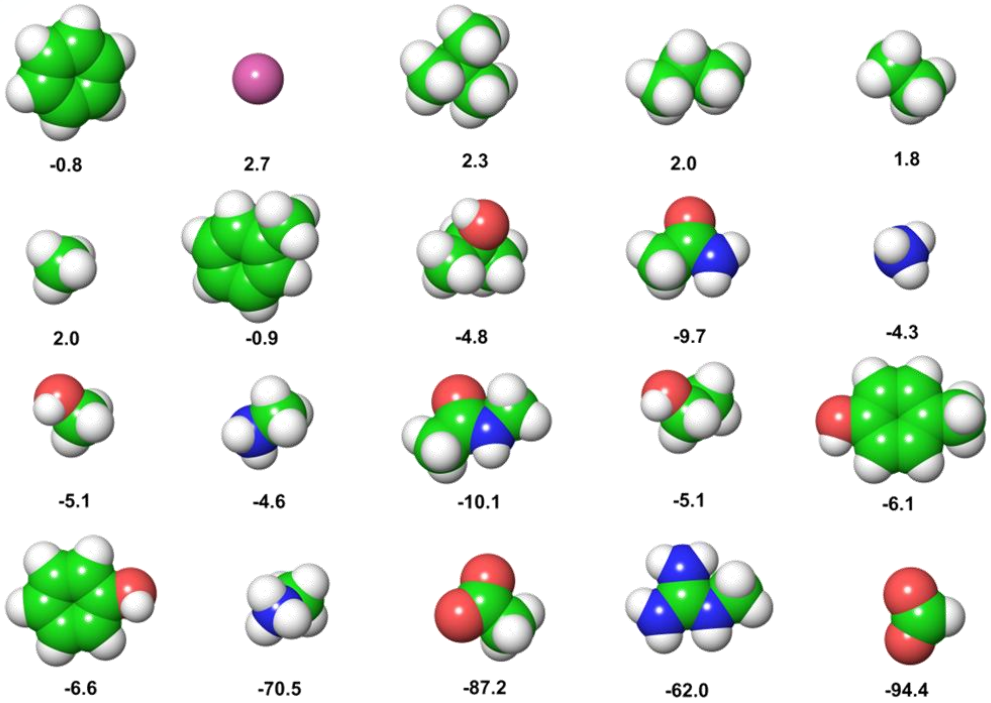
- Water molecules in protein cavities may contribute more than +2.0 kcal/mol.

Dunitz "The entropic cost of bound water in crystals and biomolecules." Science -1994, 264(5159):670.

Huggins. "Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations" Biophys. J. - 2015, 108, pp 928-936

Validation

- $\Delta G_{hydration}$ calculated using FEP and IFST for 20 solutes.
- 7 non-polar solutes, 9 polar solutes and 4 charged solutes.

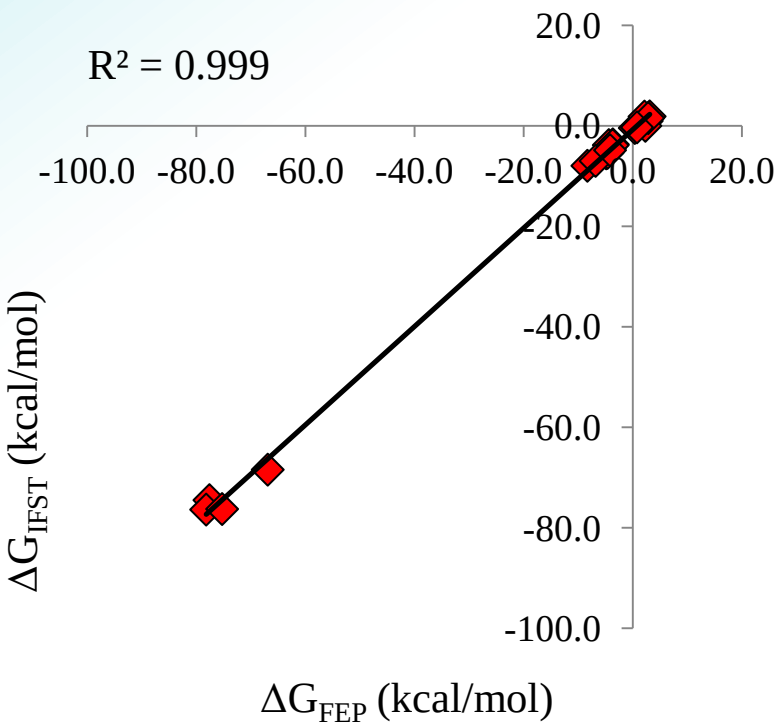


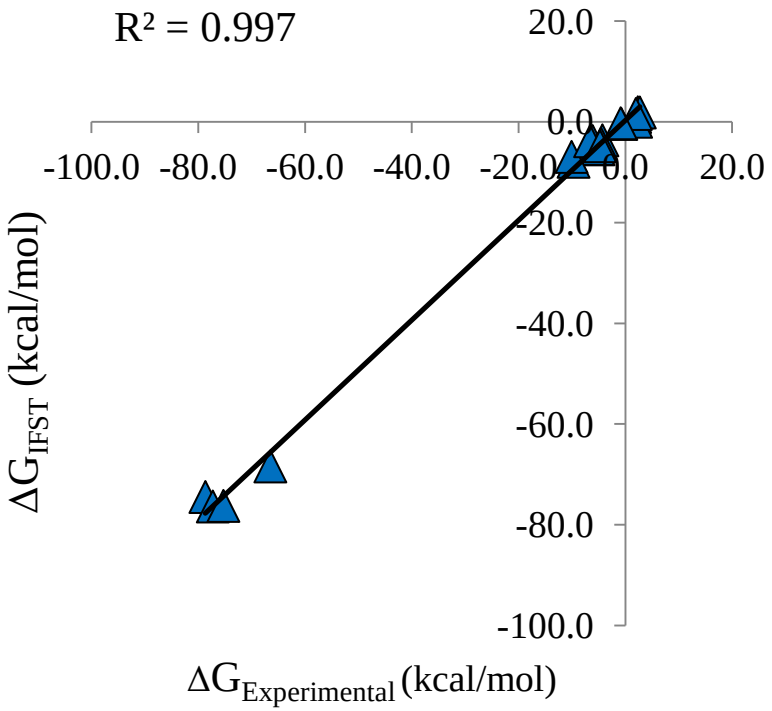
- No heavy atom rotatable torsions.
- Fixed solute geometry for FEP and IFST simulations.

Huggins. "Estimating Translational and Orientational Entropies Using the k-Nearest Neighbors Algorithm" J. Chem. Theory Comput., 2014, 10 (9), pp 3617–3625

Validation

- Excellent agreement with hydration free energies from FEP (red) and experiment (blue).



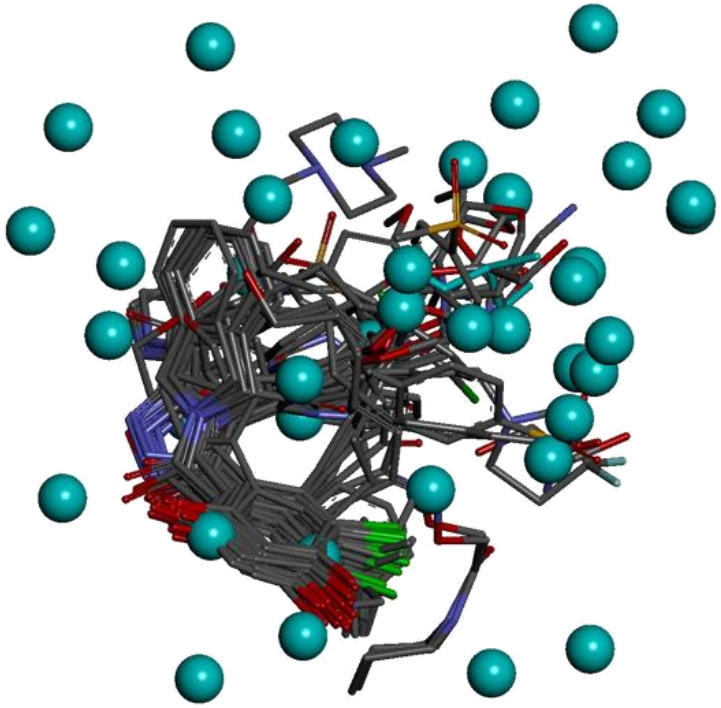


- The slopes are 0.99 and 099
- The MUEs are 0.9 kcal/mol and 1.2 kcal/mol.
- The R^2 values for uncharged species are 0.978 and 0.930

Huggins. "Estimating Translational and Orientational Entropies Using the k-Nearest Neighbors Algorithm" J. Chem. Theory Comput., 2014, 10 (9), pp 3617–3625

HSP90

- Does ligand binding coincide with displacement of weakly bound water molecules?

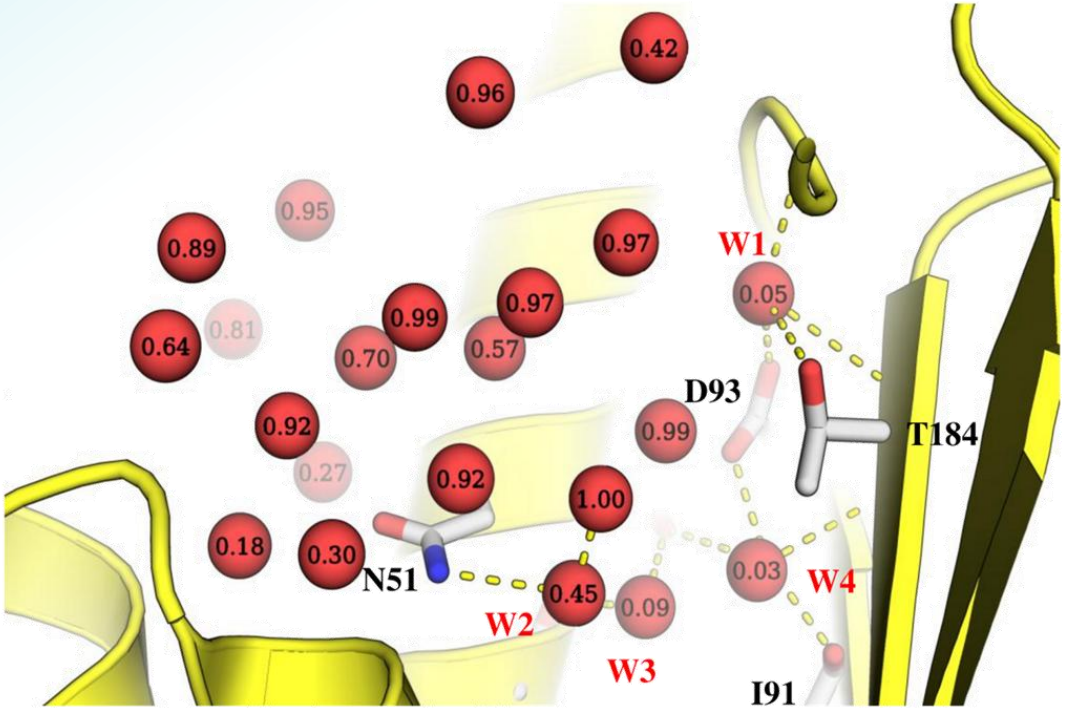


- Dataset consists of 103 overlaid ligand-bound X-ray crystal structures of Hsp90.

Haider and Huggins "Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules" J. Chem. Inf. Model., 2013, 53 (10), pp 2571–2586

HSP90

- We consider the fractional conservation (F) and the displacement fraction (D).



$$F = \frac{N_{hydrated}}{N_{total}}$$

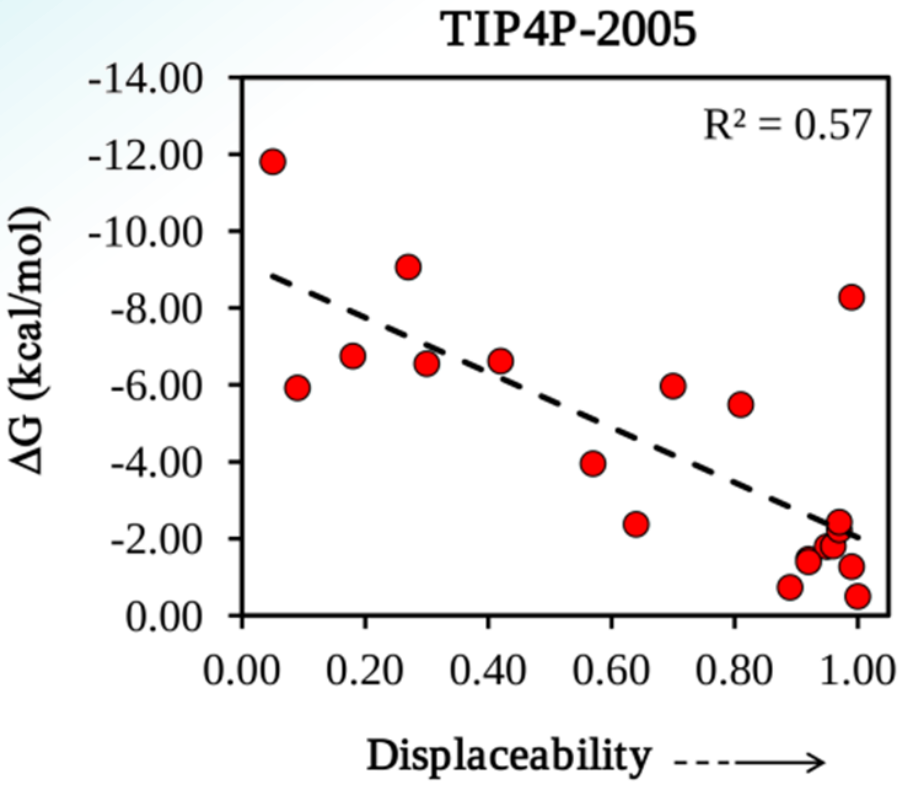
$$D = 1 - F$$

- We calculate the free-energy contribution for each hydration sites using IFST.

Haider and Huggins "Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules" J. Chem. Inf. Model., 2013, 53 (10), pp 2571–2586

HSP90

- The correlation between ΔG_{IFST} and D is reasonable.



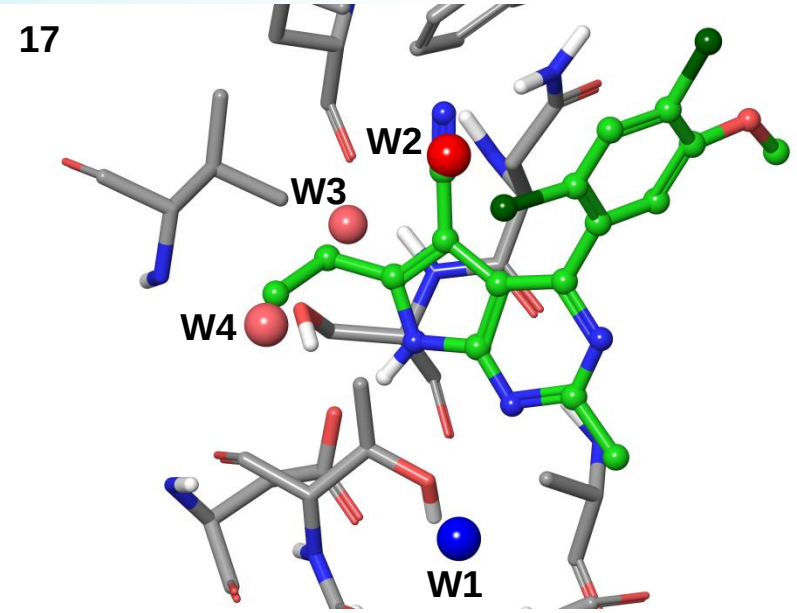
Water	D	ΔG kcal/mol	ΔE kcal/mol	-TΔS kcal/mol	E_{sw} kcal/mol
W301	0.05	-11.81	-13.46	1.65	-23.19
W323	0.27	-9.07	-8.59	-0.48	-14.41
W324	0.57	-3.95	-3.88	-0.07	-7.65
W325	0.30	-6.55	-6.70	0.15	-14.44
W328	0.09	-5.92	-6.42	0.50	-10.36
W336	0.99	-8.28	-8.85	0.57	-14.58
W338	0.92	-1.47	-1.21	-0.25	-3.32
W346	0.18	-6.75	-6.93	0.18	-13.45
W357	1.00	-0.50	-0.22	-0.27	-1.14
W379	0.95	-1.79	-1.55	-0.24	-2.89
W381	0.96	-1.81	-1.46	-0.34	-2.71
W385	0.64	-2.37	-2.37	0.00	-5.29
W405	0.97	-2.23	-2.11	-0.12	-4.82
W412	0.70	-5.97	-5.64	-0.33	-9.67
W435	0.97	-2.43	-1.91	-0.52	-3.86
W476	0.99	-1.27	-0.91	-0.36	-1.83
W529	0.42	-6.62	-6.11	-0.51	-11.11
W536	0.89	-0.73	-0.84	0.11	-1.41
W547	0.92	-1.40	-1.06	-0.34	-1.96
W598	0.81	-5.49	-4.84	-0.65	-8.03
R^2 against D		**0.57**	**0.58**	**0.27**	**0.62**

- The major outlier represents a site where ligands commonly make a salt bridge.

Haider and Huggins "Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules" J. Chem. Inf. Model., 2013, 53 (10), pp 2571–2586

HSP90

- Four of these waters were specifically targeted in a 2011 study.



Compound	Enzyme Ki (μM)	Waters Displaced
7	0.14	-
9	2.0	-
13	0.04	W2
16	0.03	W2, W3, W4
17	0.015	W2, W3, W4

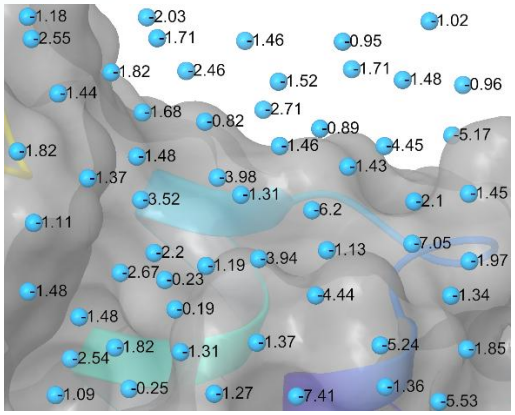
	D (Observed)	D (Predicted)
W1	0.03	0.07
W2	0.45	0.81
W3	0.04	0.49
W4	0.03	0.47

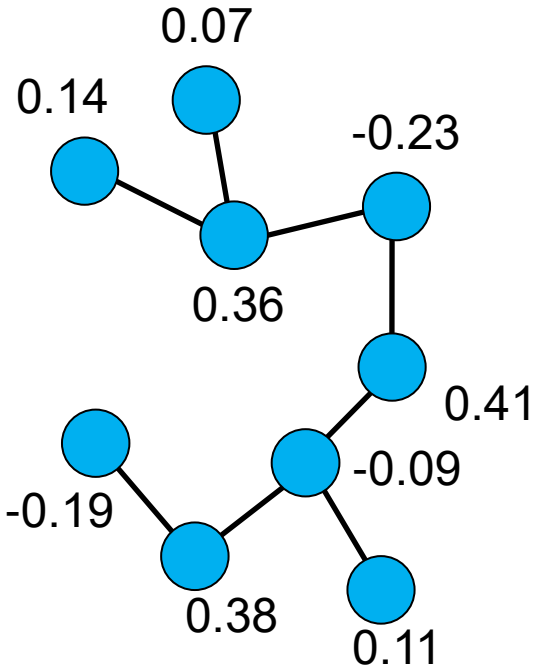
- IFST can be used to identify hydration sites with untapped potential.

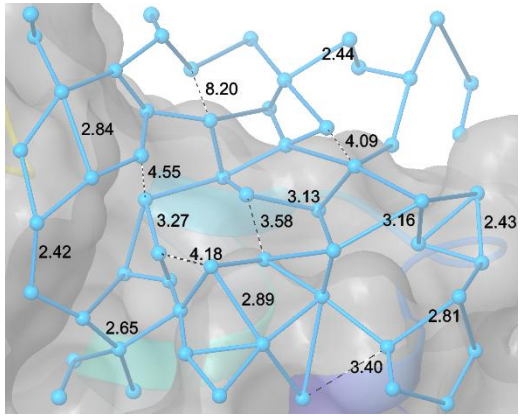
Kung et al. "Design strategies to target crystallographic waters applied to the Hsp90 molecular chaperone" Bioorganic & Medicinal Chemistry Letters, Volume 21, Issue 12, 2011, Pages 3557–3562

Methods - Combinatorial Subgraph Search

- Can we relate the strength of surface water networks with target ligandability?
- We compute the free energy for water molecules in each hydration site and then connect close hydration sites to create a network.





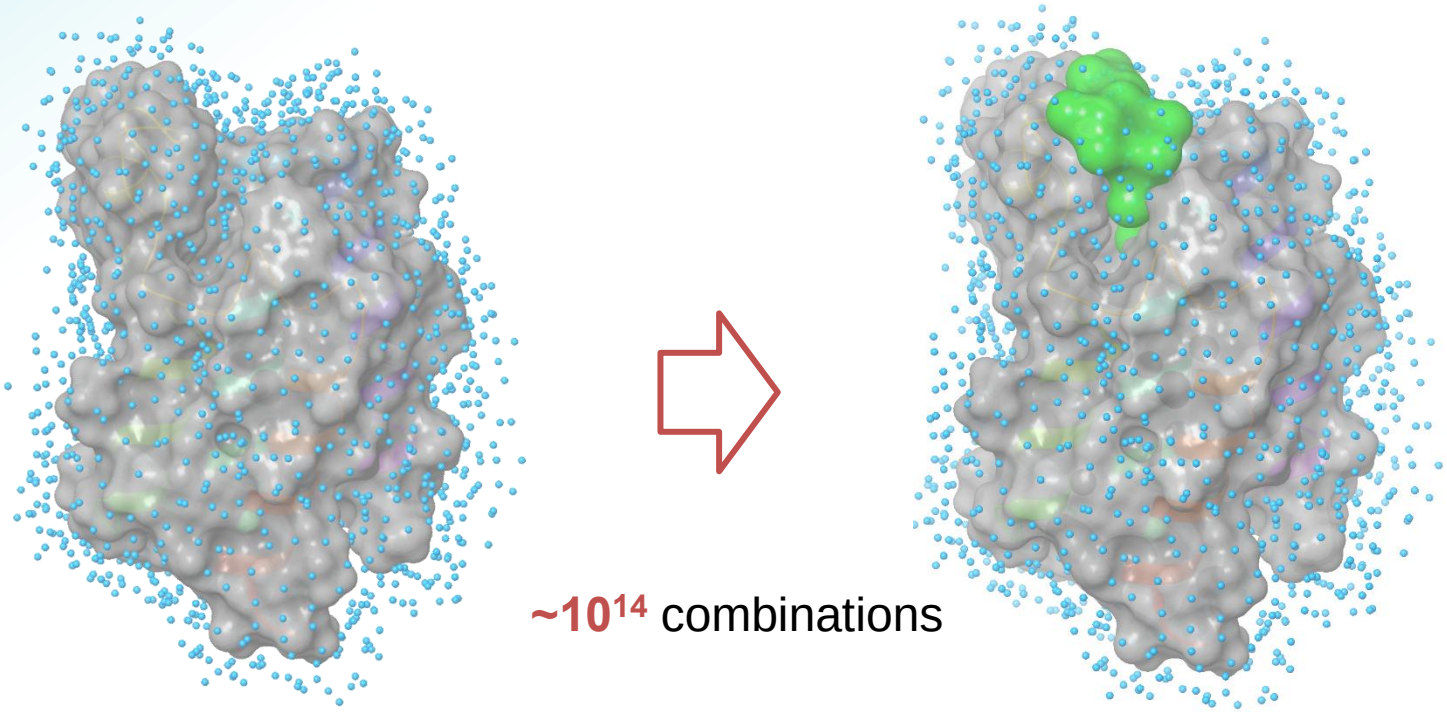


- We enumerate all subgraphs of size 18 and identify the most weakly bound.

Vukovic and Huggins. "Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites" Journal of Physics: Condensed Matter, 2016

Methods - Combinatorial Subgraph Search

- This is the predicted location of the optimal small-molecule binding site.

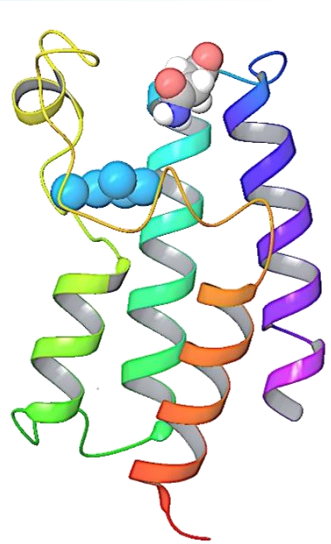


- We compare the scores amongst different protein targets to rank ligandability.

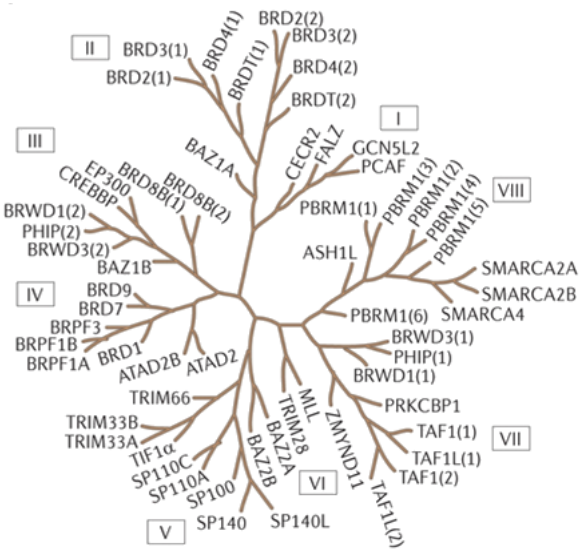
Vukovic and Huggins. "Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites" Journal of Physics: Condensed Matter, 2016

Bromodomains

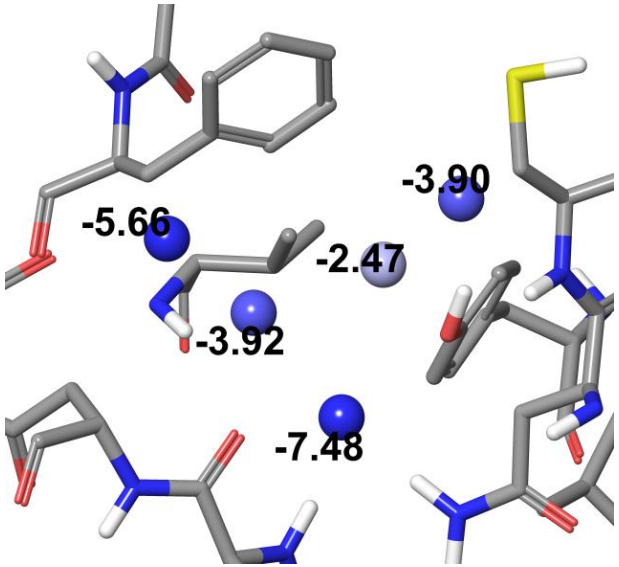
- We have proposed that a concentration of weakly-bound water molecules will be a ligand-binding hotspot.



Bromodomain secondary structure



The bromodomain family tree



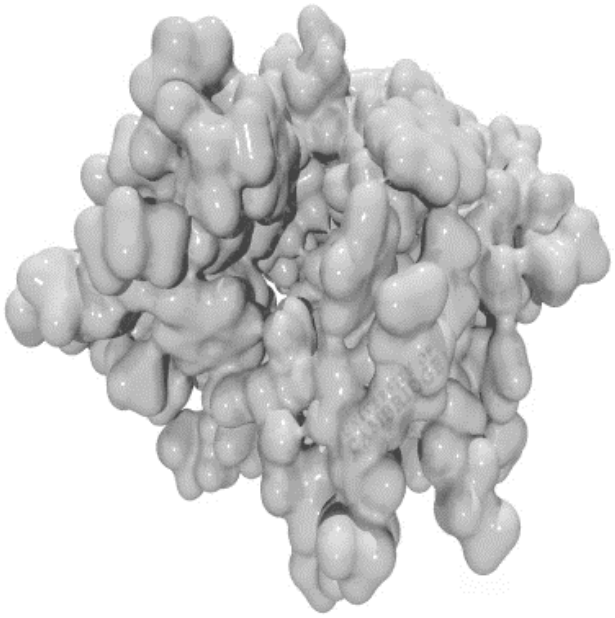
Strongly-bound water molecules in the base of the acetyl-lysine binding pocket of BAZ2B

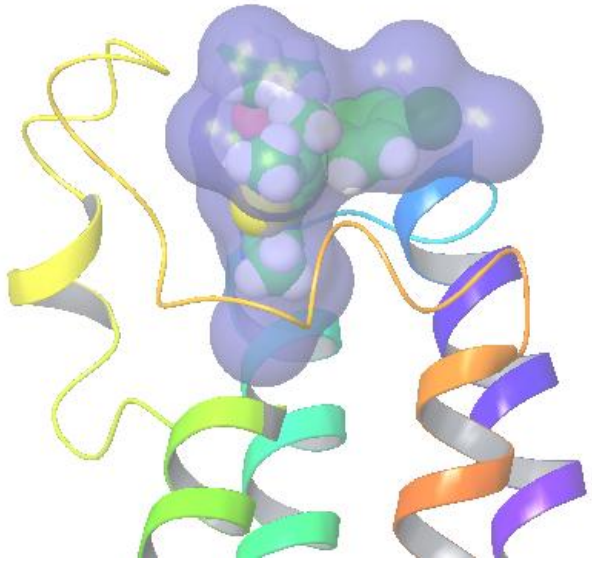
- The aim of the project was to predict the ligandability of the bromodomains family, in terms of the effort required to develop an inhibitor.

Vukovic and Huggins. "Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites" Journal of Physics: Condensed Matter, 2016

Bromodomains

- The software predicts protein ligandability based on the summed binding energy of the displaced water molecules.





Volume overlap for the predicted BRDT hot spot (blue) and the (+)-JQ1 ligand

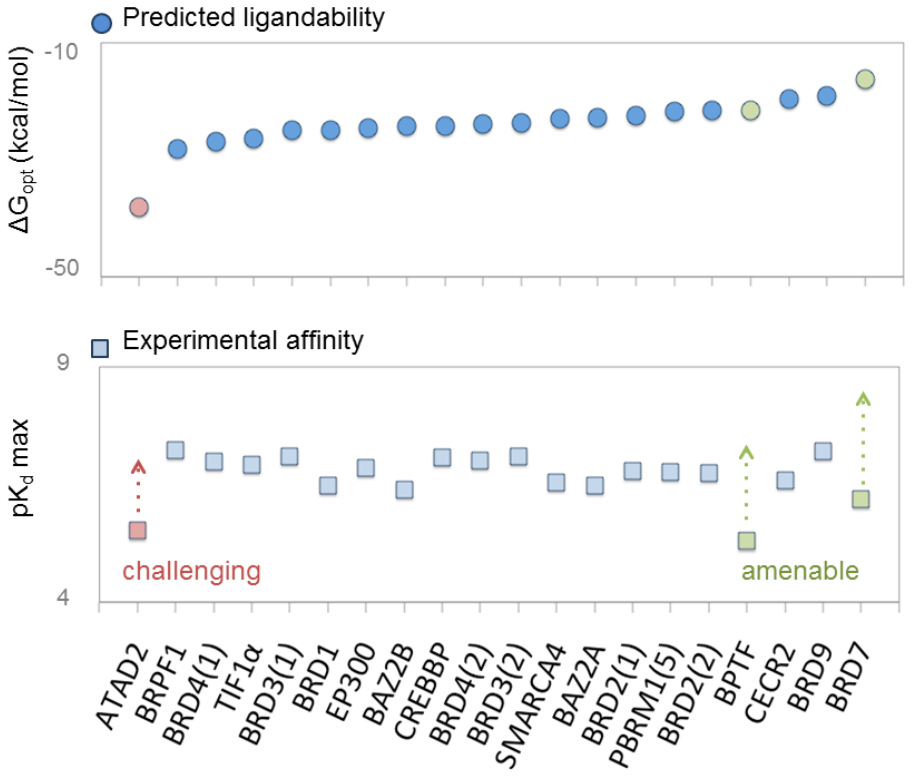
- We can predict the ideal location and shape of an inhibitor.

Vukovic and Huggins. "Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites" Journal of Physics: Condensed Matter, 2016

Bromodomains

- The most ligandable region is identified within the acetyl-lysine binding site for all 165 protein chains tested.

BRD	ΔG_{IFST} (N=18) [kcal/mol]	max pK_D	max K_D (ITC) [nM]	ligand MW
ATAD2	-38.05	5.82	1500	350.40
BRPF1	-28.06	7.89	13	380.45
BRD4(1)	-26.83	7.59	26	423.17
TIF1α	-26.35	7.51	31	643.00
BRD3(1)	-24.91	7.72	19	423.17
BRD1	-24.87	6.97	108	383.42
EP300	-24.53	7.42	38	508.20
BAZ2B	-24.22	6.87	136	371.00
CREBBP	-24.16	7.68	21	508.20
BRD4(2)	-23.93	7.60	25	414.50
BRD3(2)	-23.77	7.72	19	423.17
SMARCA4	-22.94	7.05	89	321.40
BAZ2A	-22.78	6.96	109	357.42
BRD2(1)	-22.44	7.34	46	423.17
PBRM1(5)	-21.66	7.32	48	321.40
BRD2(2)	-21.53	7.28	52	423.17
BPTF	-21.50	5.55	2800	187.00
CECR2	-19.63	7.10	80	495.70
BRD9	-19.07	7.85	14	353.40
BRD7	-16.25	6.62	239	353.40



- We have identified ligandable and difficult bromodomain targets.

Vukovic and Huggins. "Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites" Journal of Physics: Condensed Matter, 2016

Ligandability Metrics

- We would like to predict the amount of effort required to develop a small molecule inhibitor with an affinity of 10 nM (for example).
- To validate predictions, we are currently developing metrics to quantify ligandability from experimental data.

$$L_1 = \frac{[N]_{pK_i>8}}{[N]_{5>pK_i>8}}$$

$$L_2 = \frac{4^{pK_i^{\max}}}{N}$$

$$L_3 = \sqrt{(1-\frac{pK_i^{\max}}{12})^2 + (\frac{N}{1000})^2}$$

Target	max pK_i	N	L_1	L_2	L_3
Carbonic Anhydrase II	10.1	478	0.77	2509	1.99
Factor Xa (fXa)	10.9	466	0.80	7686	2.10
Carbonic Anhydrase I	9.0	416	0.61	630	2.06
Trypsin	8.9	266	0.19	883	2.71
Dipeptidyl Peptidase IV (DPP-IV)	9.4	244	1.18	1838	3.06
Androgen Receptor (AR)	9.7	240	1.26	2878	3.26
Adenosine Receptor A2A	9.7	237	1.05	3006	3.29
Thrombin	10.0	233	0.44	4500	3.49
Carbonic Anhydrase IV	8.8	232	0.42	821	2.81
Carbonic Anhydrase IX	9.0	220	0.34	1192	3.00
Carboxylesterase 1	8.8	219	0.37	938	2.91
Adenosine Receptor A1	9.3	188	0.20	2117	3.41
HIV-1 Protease	10.5	178	1.79	12161	4.62
Cathepsin L	9.0	167	0.71	1482	3.29
Acetylcholinesterase (AChE)	10.6	164	0.64	14387	4.95
Cathepsin S	9.0	157	0.18	1670	3.39
Carbonic Anhydrase VA	8.4	153	0.39	744	2.97
Butyrylcholinesterase (BuChE)	8.7	152	0.16	1136	3.18
Cathepsin K	8.0	151	0.00	410	2.71

- We are using binding-affinity data from the Binding DB to rank well-validated targets.

T Liu, Y Lin, X Wen, RN Jorissen, MK Gilson. "BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities" Nucleic Acids Res. 2007

Summary

- Water Networks
  - Strongly and weakly bound networks
  - Key role in protein-ligand binding
  - Easy to ignore
- IFST
  - Quantitatively accurate
  - Correlation with displaceability
  - Insight from a single simulation
- Ligandability
  - Prediction of target ligandability
  - Identification of optimal binding sites
  - Development of ligandability metrics based on experimental data

We are interested in feedback on these metrics and post-competitive collaboration to understand and predict protein target ligandability.

Acknowledgements

Department of Physics - Cambridge
Professor Mike Payne
Ben Irwin
Dr Sinisa Vukovic (Bromodomains)

City University of New York
Dr Kamran Haider (HSP90)

Alzheimer's Research UK
Dr John Skidmore

Structural Genomics Consortium
Dr Paul Brennan
Dr Oleg Fedorov
Professor Stefan Knapp

Hutchison/MRC Cancer Research Centre
Professor Ashok Venkitaraman

MIT
Professor Bruce Tidor
Dr Michael Altman

Funding

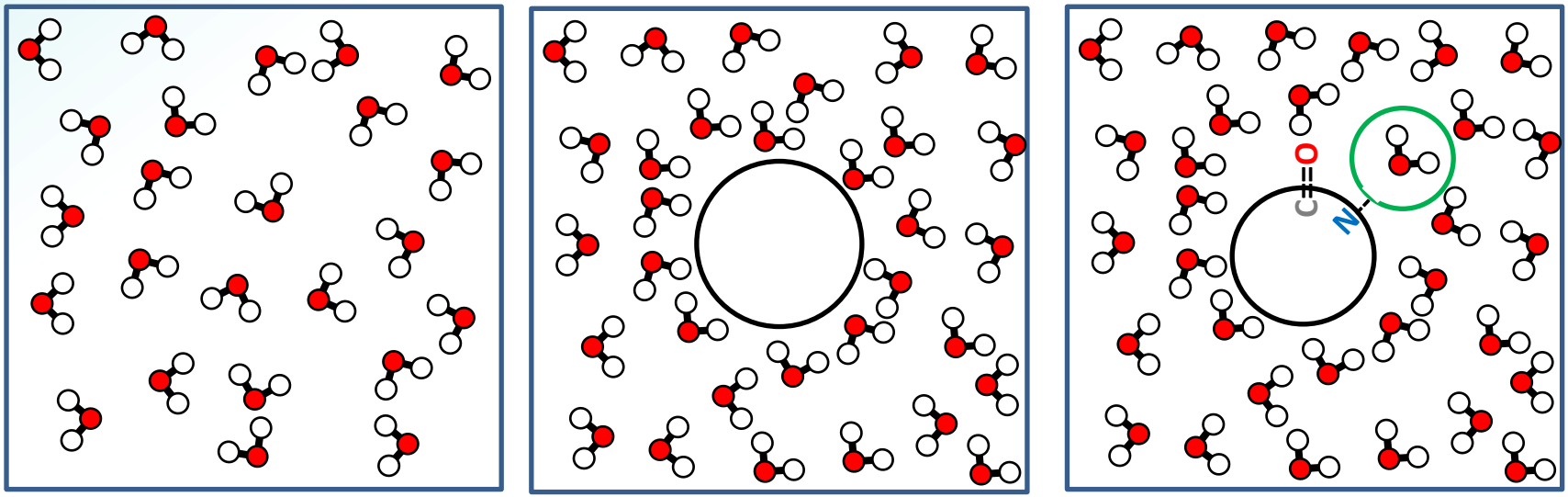
MRC Medical Research Council

EPSRC Engineering and Physical Sciences Research Council

HEC BioSim

IFST

- The presence of a solute (such as a small molecule or protein) perturbs the structure of the water, creating local order.



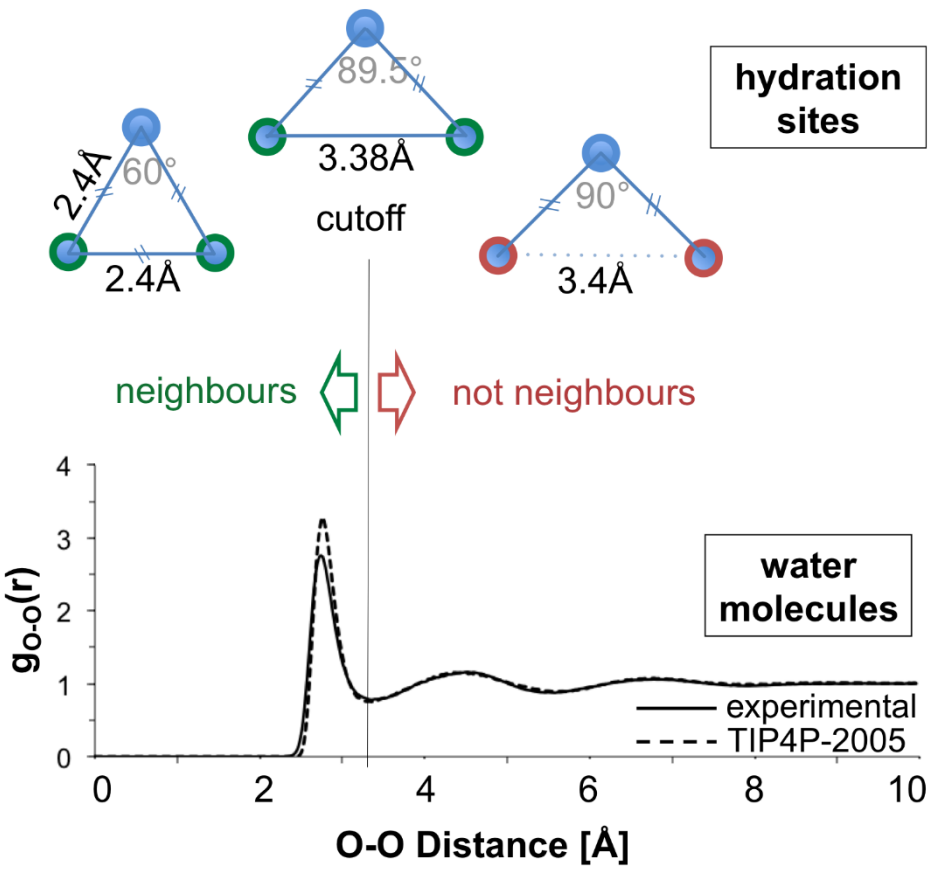
$$S_{sw} = -NR\rho \int g_{w|s} \ln g_{w|s}\, dV$$

$$S_{ww} = {}^1/_2\, NR\rho^2 \int\int g_{ww'|s} \ln \frac{g_{ww'|s}}{g_{w|s} g_{w'|s}}\, dV\, dV'$$

- This order can be captured by computing correlation functions, which can in turn be used to estimate the entropy.

Lazaridis T. "Solvent Reorganization Energy and Entropy in Hydrophobic Hydration" J. Phys. Chem. B (2000), 104, 4964-4979

Methods - Combinatorial Subgraph Search



Vukovic and Huggins. "Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites" Journal of Physics: Condensed Matter, 2016