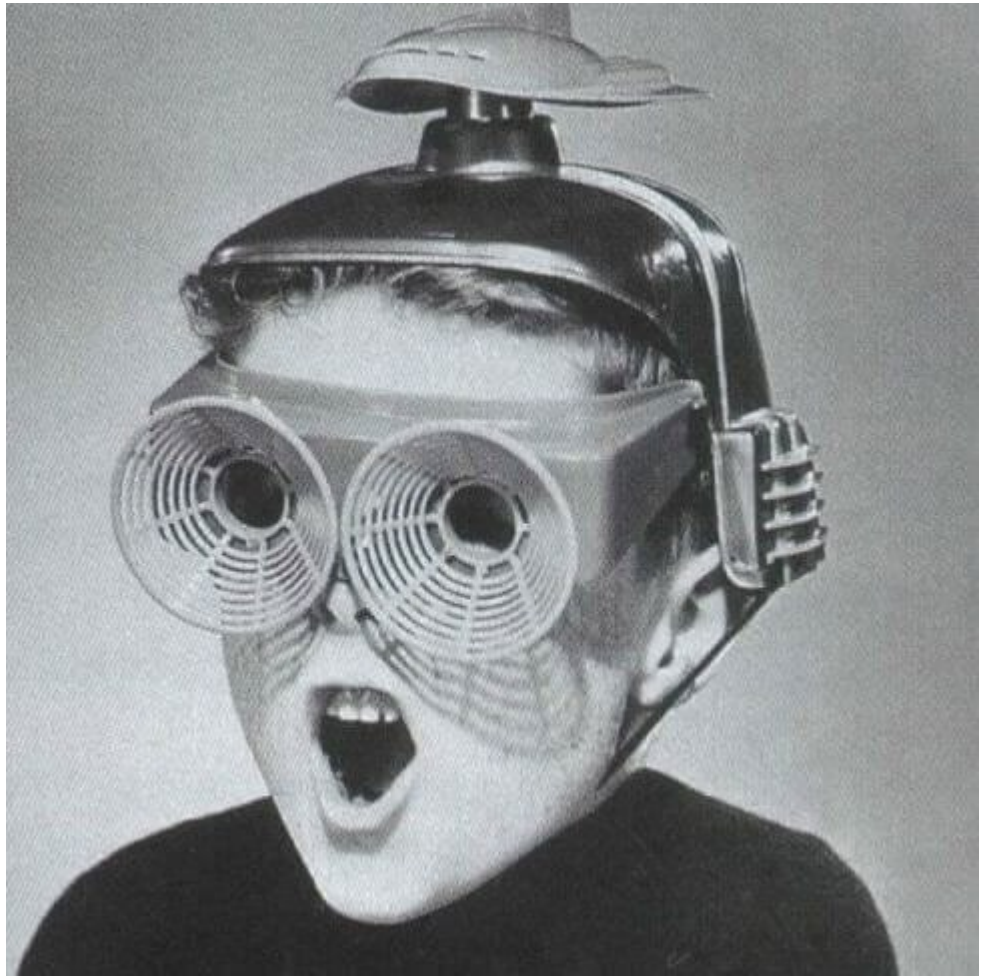
innovative science • intuitive software

Cresset Science – The Future Today

Mark Mackey, CSO

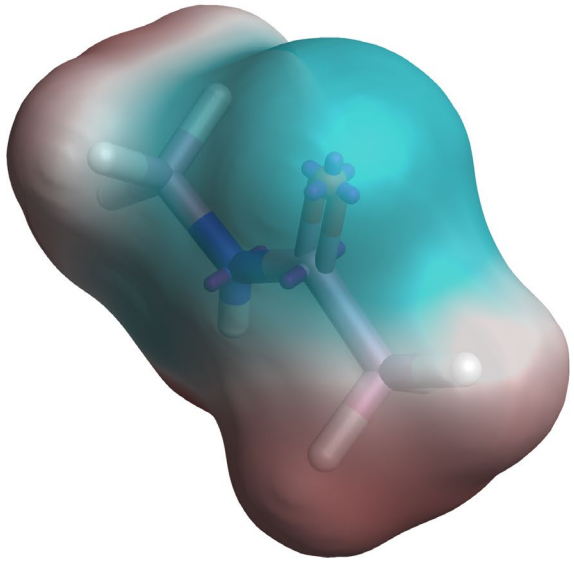
A member of the Cresset science team envisaging the future



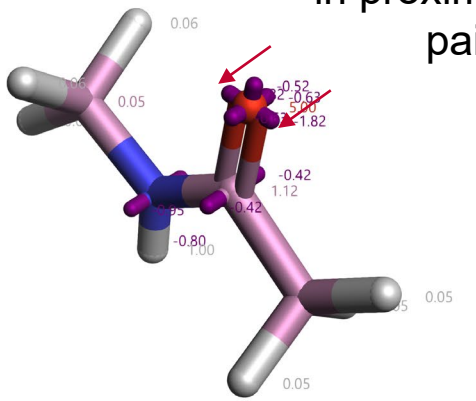
Electrostatic Complementarity™

Anisotropic charge distribution with XED force field

- The polarizable XED force-field is an excellent base for calculating electrostatic properties
  - Description of anisotropic atomic charge distributions at relatively modest computational costs

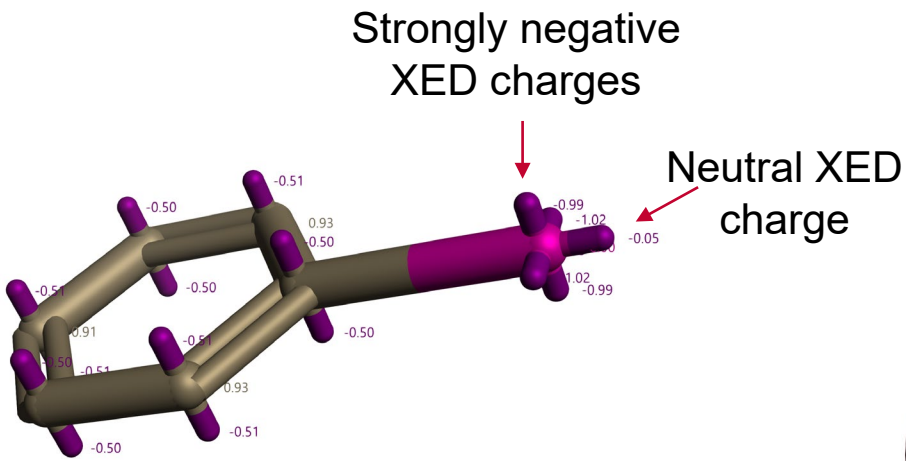


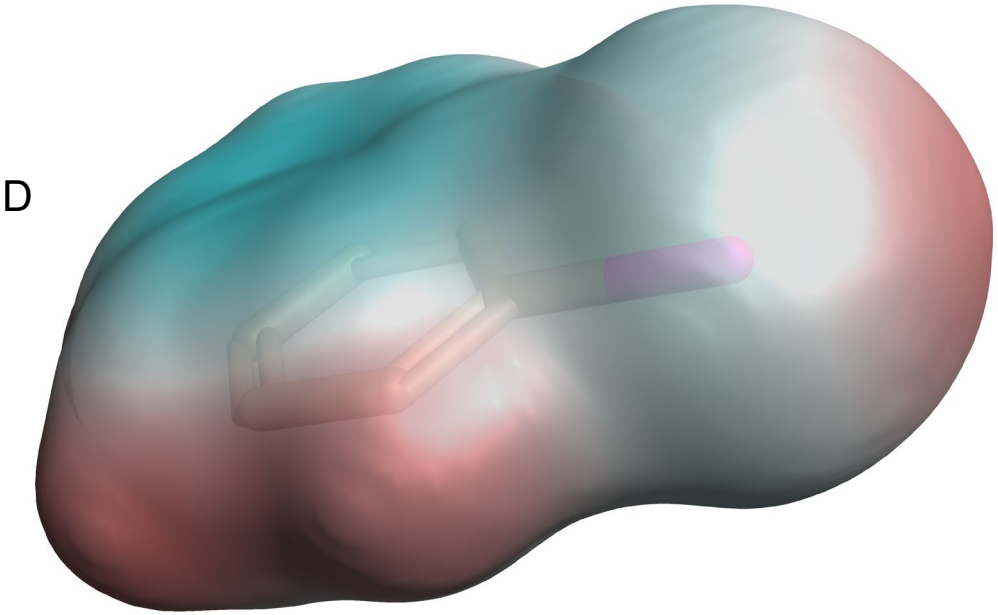
XED charges more negative in proximity to oxygen lone pair orientation



Anisotropic charge distribution with XED force field

- The polarizable XED force-field is an excellent base for calculating electrostatic properties
  - Description of anisotropic atomic charge distributions at relatively modest computational costs

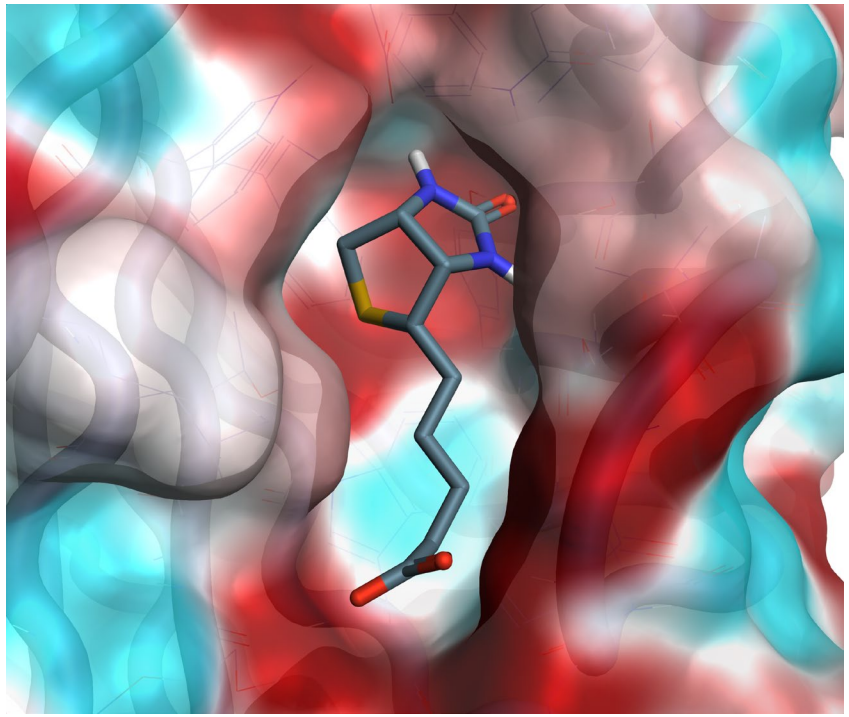




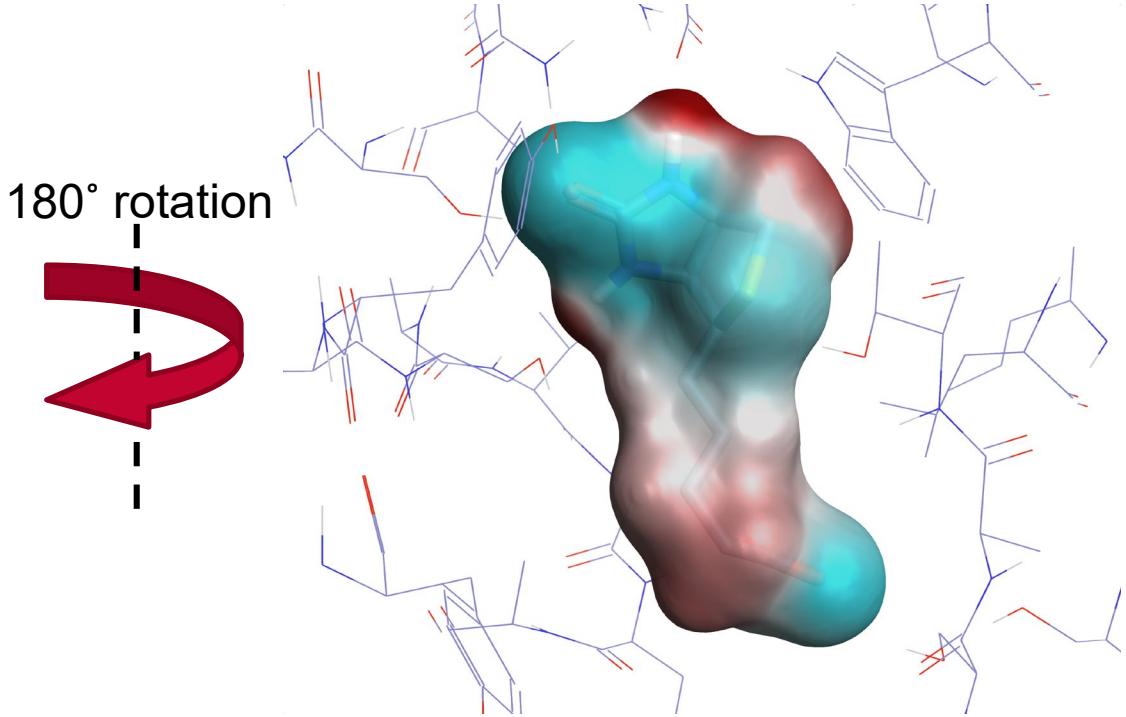
Biotin-Streptavidin example

> Visual inspection of electrostatic potential (Biotin-Strepavidin)

→ red = positive potential and blue = negative potential



XED ESP surface of Streptavidin

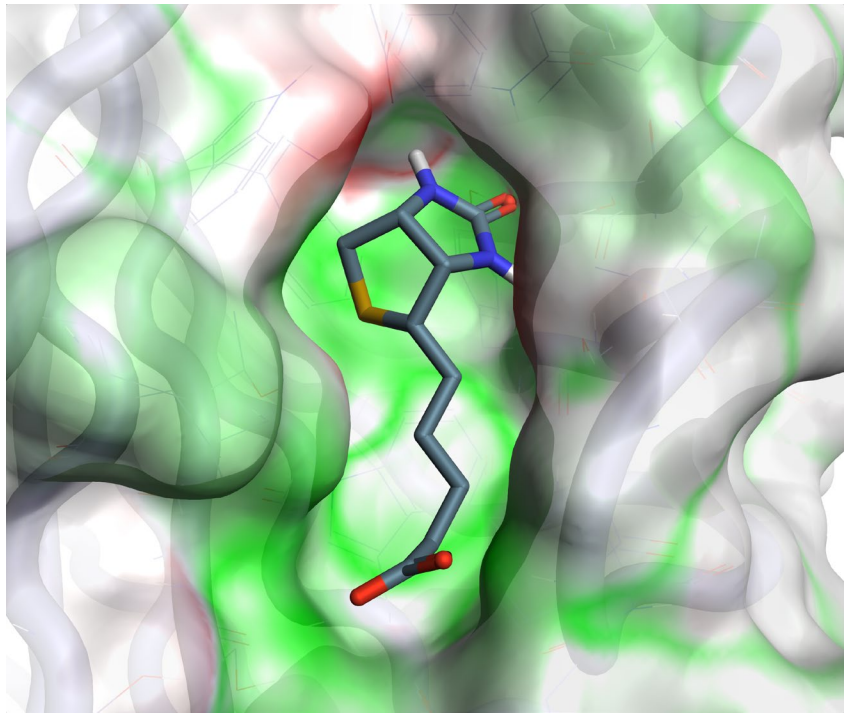


XED ESP surface of Biotin

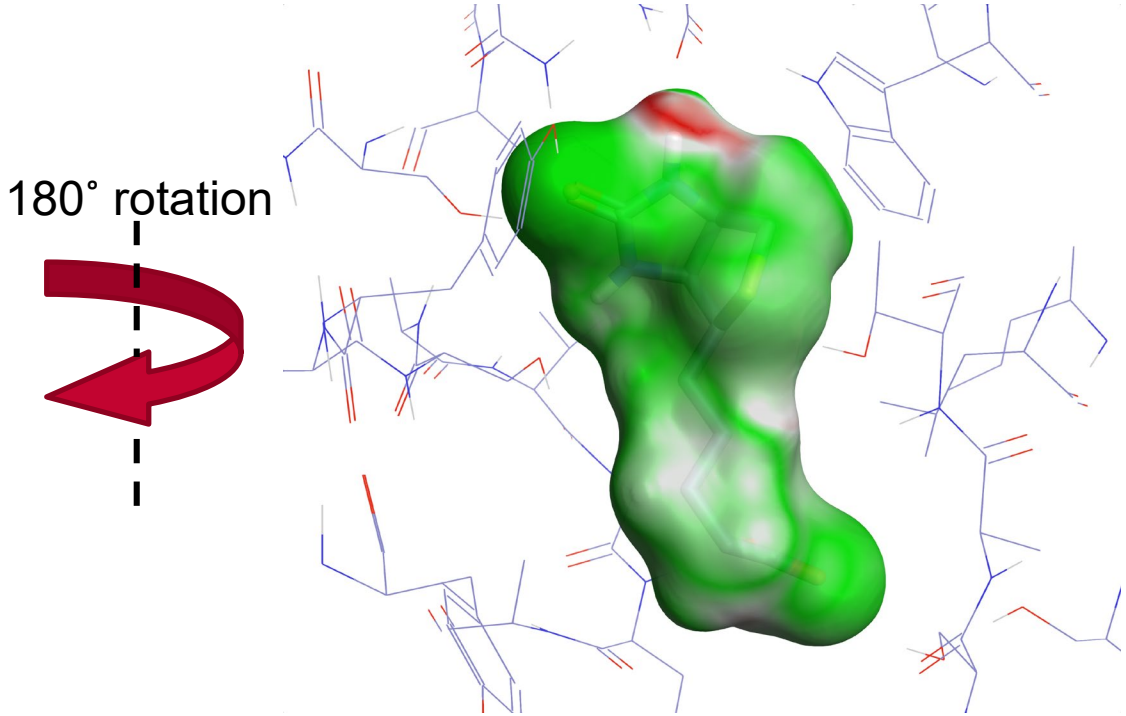
Biotin-Streptavidin example

> Visualization of Electrostatic Complementarity (EC) (Biotin-Strepavidin)

→ green = good complementarity and red = bad complementarity



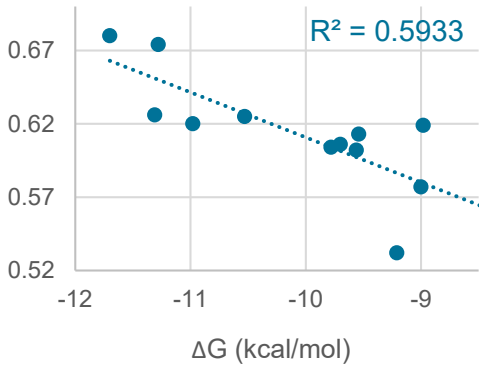
EC surface of Streptavidin



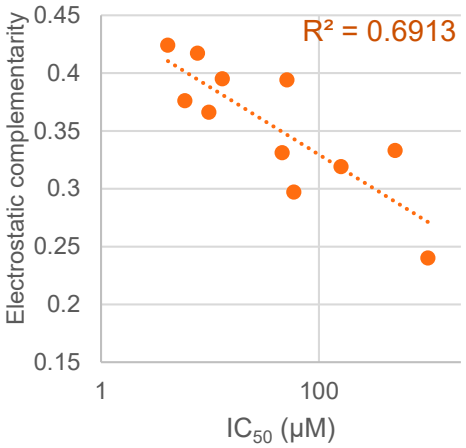
EC surface of Biotin

Application to additional data sets

TYK2



XIAP



Journal of Medicinal Chemistry

Article

Cite This: J. Med. Chem. 2019, 62, 3036−3050

pubs.acs.org/jmc

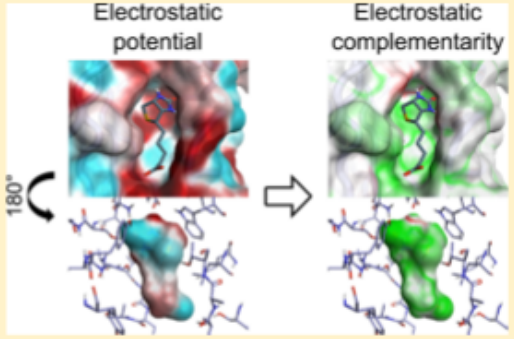
Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein−Ligand Complexes

Matthias R. Bauer* and Mark D. Mackey

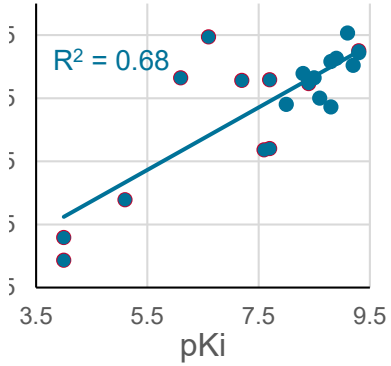
Cresset, New Cambridge House, Bassingbourn Road, Litlington, Cambridgeshire SG8 0SS, U.K.

Supporting Information

ABSTRACT: Electrostatic interactions between small molecules and their respective receptors are essential for molecular recognition and are also key contributors to the binding free energy. Assessing the electrostatic match of protein−ligand complexes therefore provides important insights into why ligands bind and what can be changed to improve binding. Ideally, the ligand and protein electrostatic potentials at the protein−ligand interaction interface should maximize their complementarity while minimizing desolvation penalties. In this work, we present a fast and efficient tool to calculate and visualize the electrostatic complementarity (EC) of protein−ligand complexes. We compiled benchmark sets demonstrating electrostatically driven structure-activity relationships (SAR) from literature data, including kinase, protein−protein interaction, and GPCR targets, and used these to demonstrate that the EC method can visualize, rationalize, and predict electrostatically driven ligand affinity changes and help to predict compound selectivity. The methodology presented here for the analysis of EC is a powerful and versatile tool for drug design.



mGlu5



Complementarity; Complementarity r; Complementarity rho

Comparison to QM

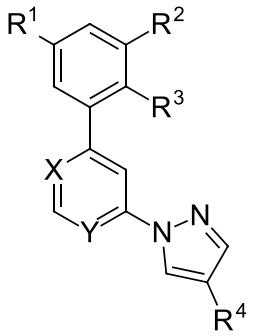
> Is the XED force field giving good enough results?

> Can we compute EC scores at the QM level?

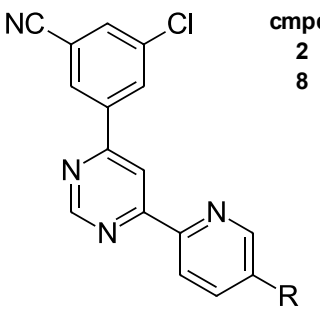
Truncated mGLU5 example (5CGC)

- Truncated binding site mode of 5CGC
- → Corresponds to more or less 6Å binding site definition in Flare™
- → no formal charges
- → analysis of 12 ligands (table 1)

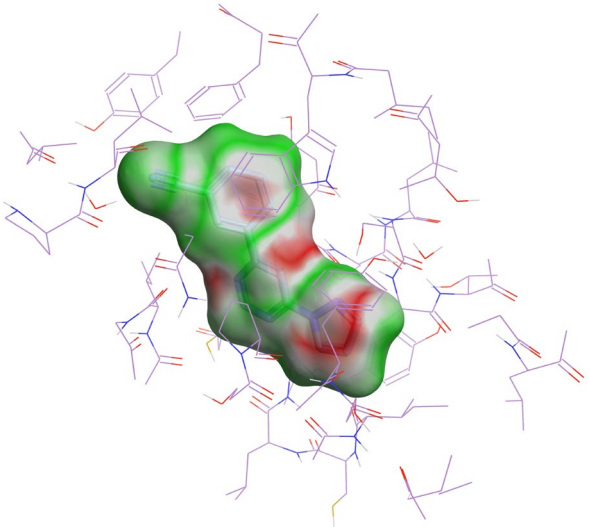
*Christopher et al, **J. Med. Chem.** 2015*



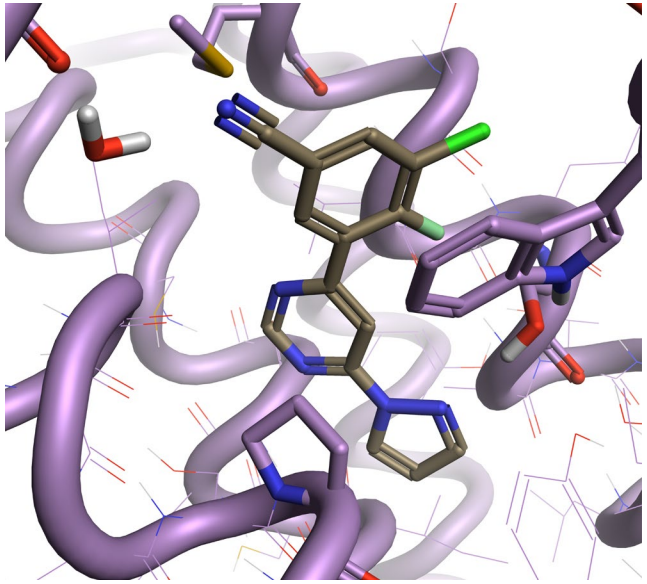
cmpd	X	Y	R¹	R²	R³	R⁴
1	N	N	CN	Cl	F	H
1a	N	N	CN	H	F	H
3	N	N	OMe	F	H	H
4	N	N	F	H	H	H
5	N	N	CN	H	H	H
6	N	N	CN	Cl	H	H
7	N	N	CN	Cl	H	F
9	N	N	CN	Me	H	H
10	N	CH	CN	Me	H	H
11	CH	N	CN	Me	H	H

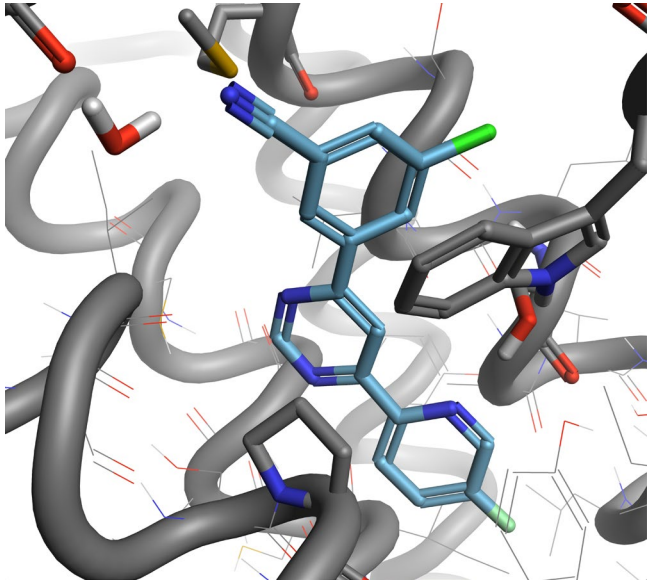


cmpd	R
2	F
8	H



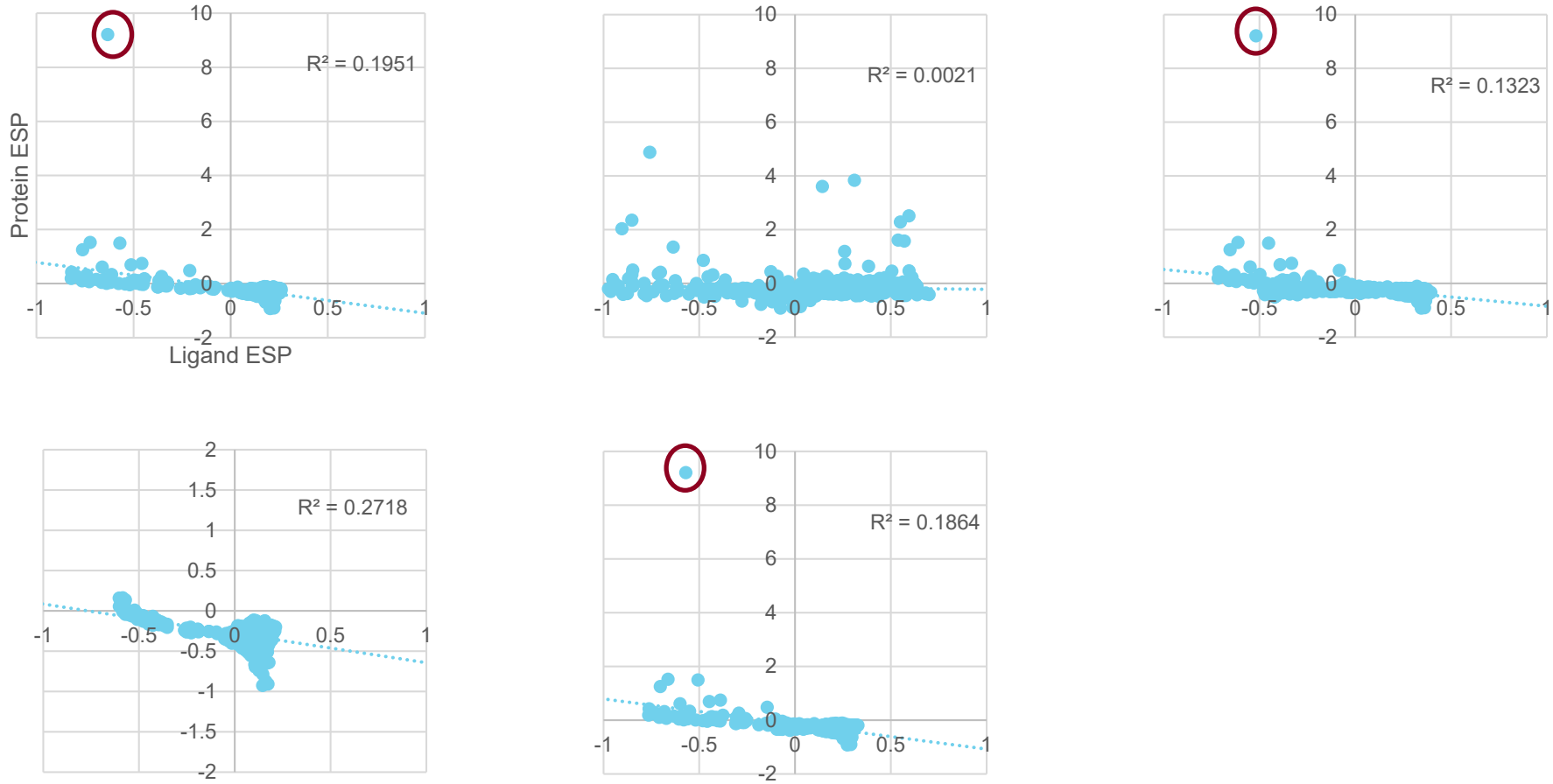
Truncated 5CGC with EC map (**5**)



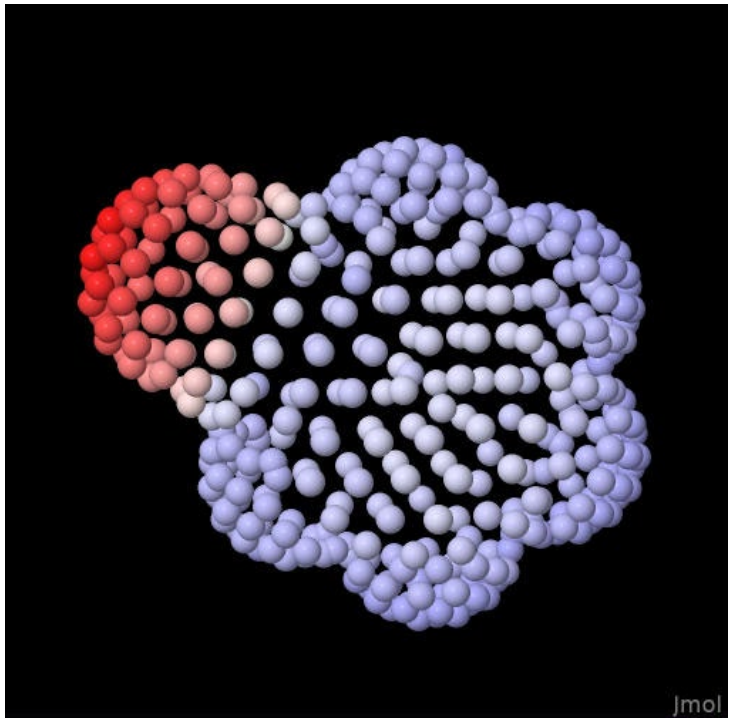


ESP value outliers

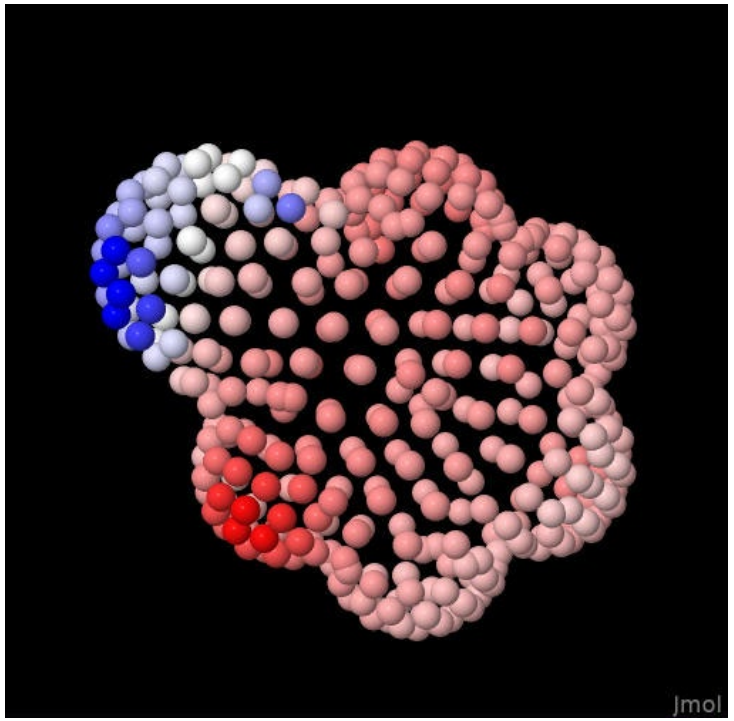
Graphs of ligand electrostatic potential vs protein electrostatic potential over the surface of different ligands



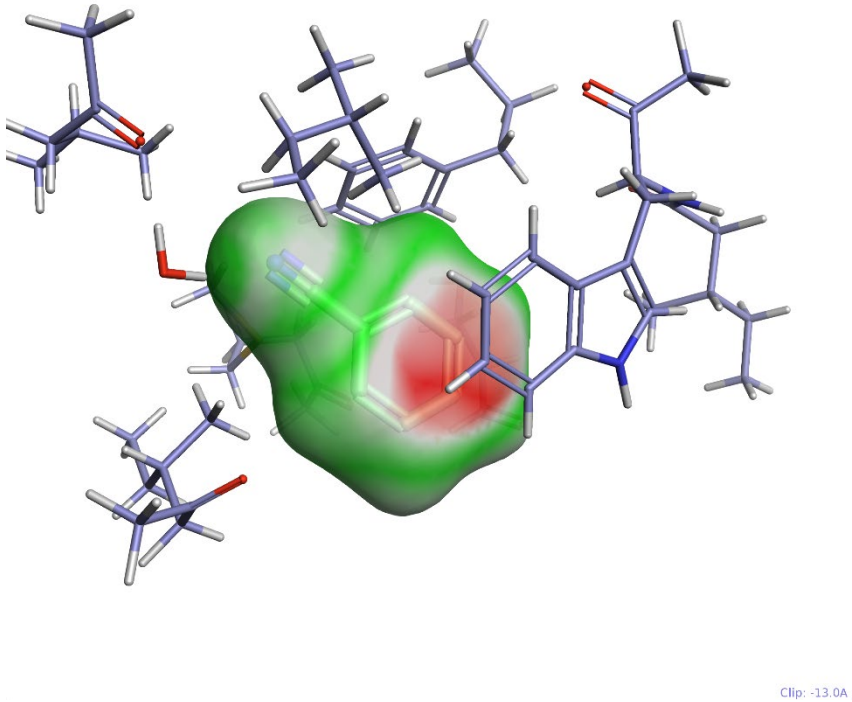
ESP value outliers



Cmpd 8 – Ligand ESP



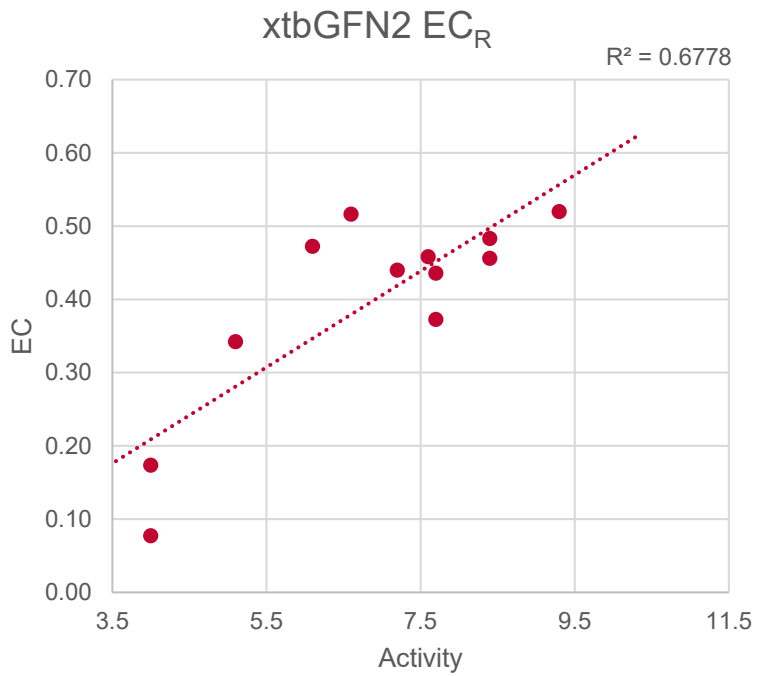
Cmpd 8 – Protein ESP

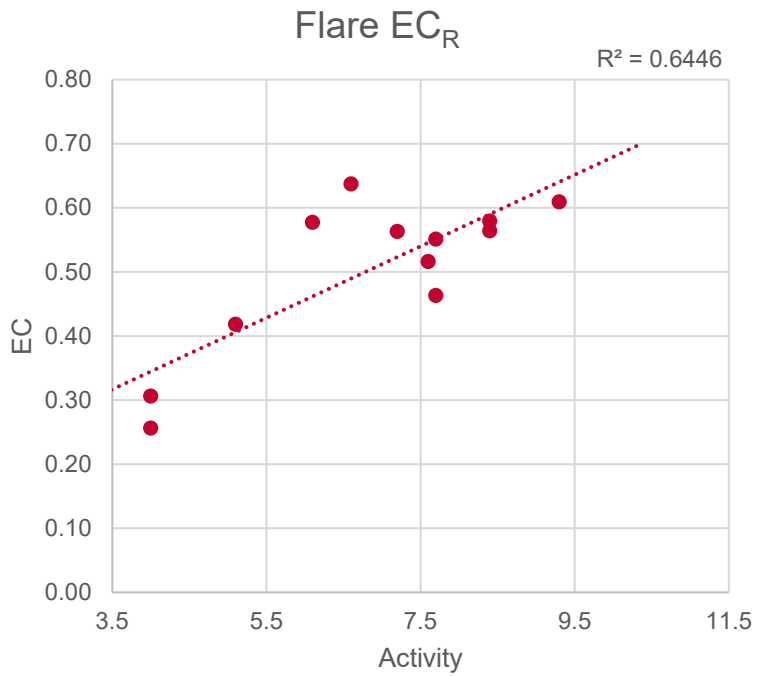


H of water molecule
very close to CN group

It is not just important HOW you calculate the electrostatic potential but also WHERE

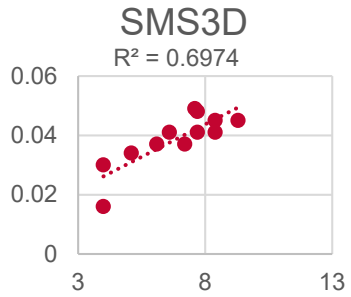
Truncated mGLU5 example (5CGC) - Flare vs XTB EC correlation

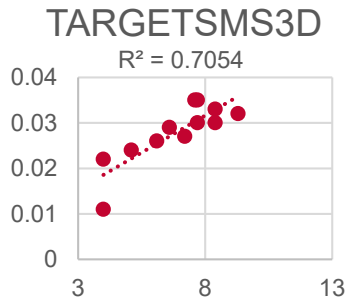


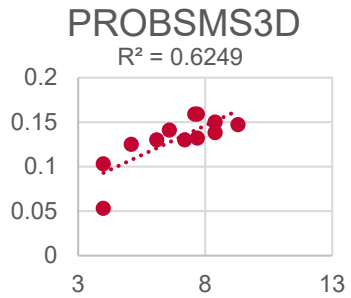


- xtbGFN2 and XED (Flare) are similarly predictive
- Use of truncated 5CGC binding pocket
- Protein ESP outliers for xtbGFN2 (ESP values over 5) were excluded

Truncated mGLU5 example (5CGC) – COSMOsim3D



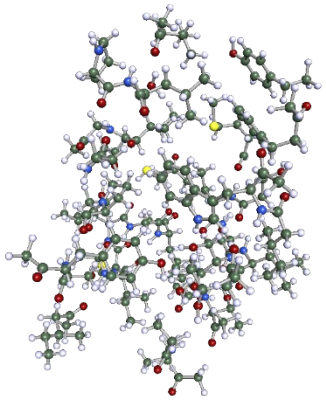


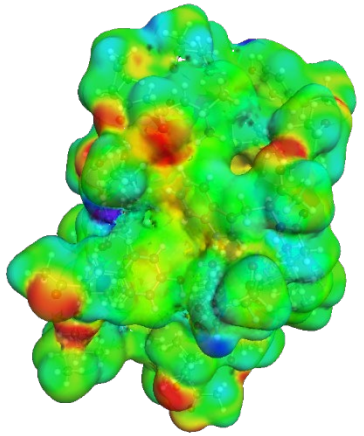


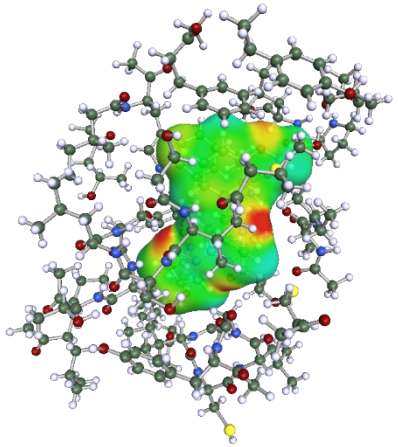
- Calculation of COSMO surfaces with Turbomole (BLYP-D3-SVP level for ligands and HF3c-D3 for receptor)

- Experimental function of COSMOsim3D can calculate similarity between inverse receptor surface and ligand COSMO surfaces

→ good correlation, but takes several hours to compute cosmo surface for truncated receptor (7-8h at HF3c level with TURBOMOLE on a workstation)







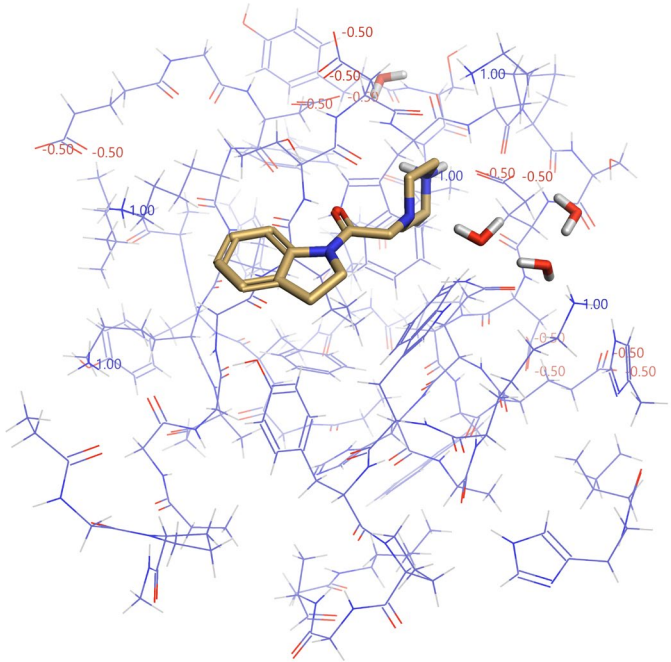
Truncated XIAP (5C7D) example

- PPI target with inhibitors that show electrostatic SAR

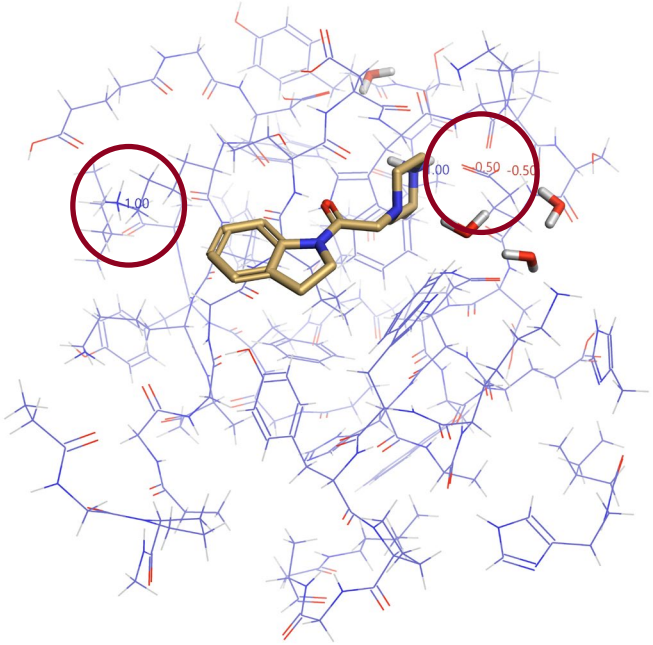
  Chessari et al., ***J. Med. Chem*** *2015*

- Binding site exhibits a large number of formal charges

- Preparation of charged and ‘neutral’ receptor



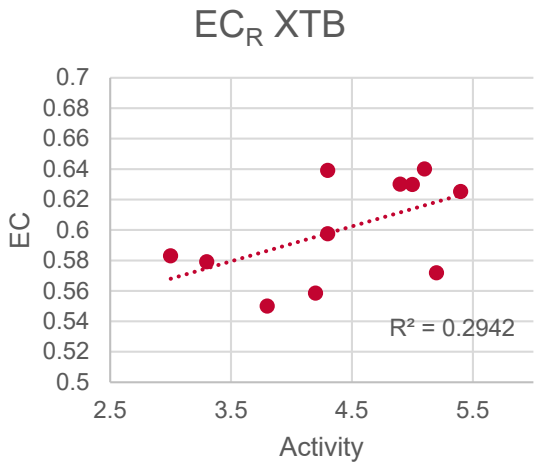
charged

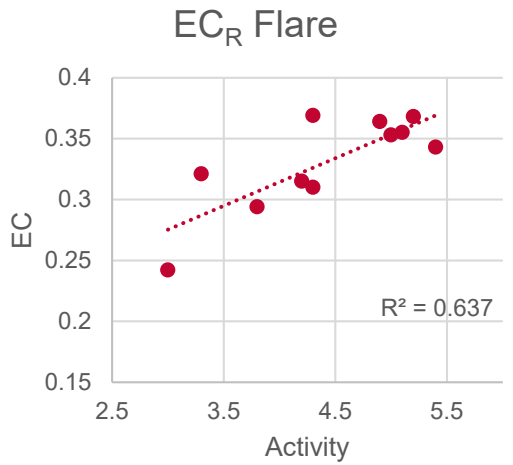


‘neutral’
2 important (close contact to ligand) charges left

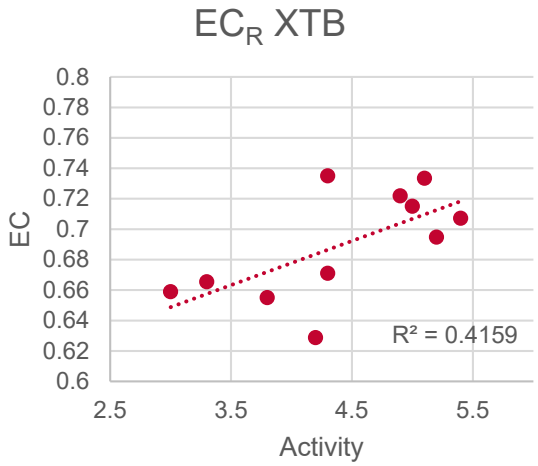
Truncated XIAP example – EC correlation

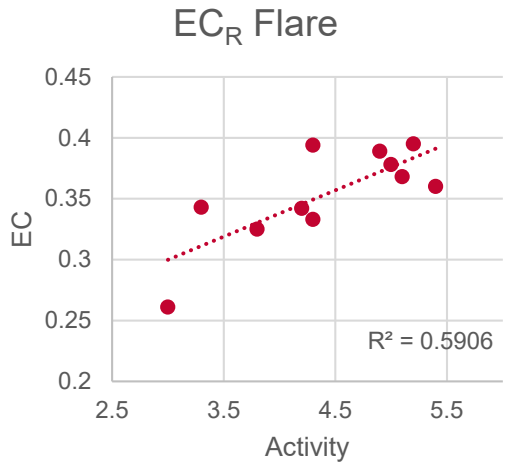
Charged





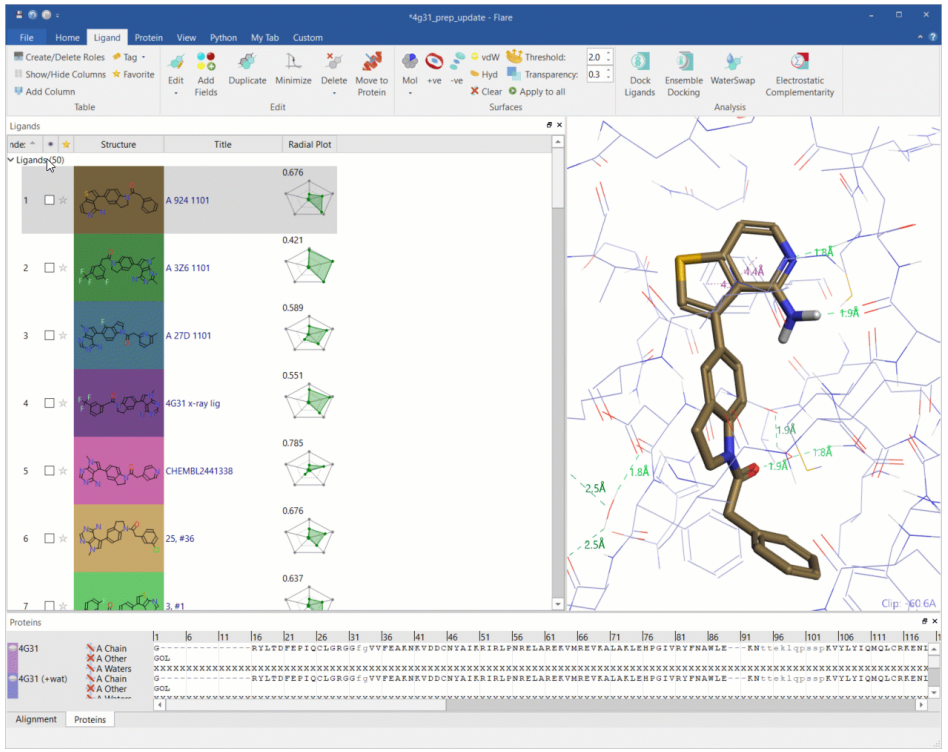
'Neutral'
2 important charges left





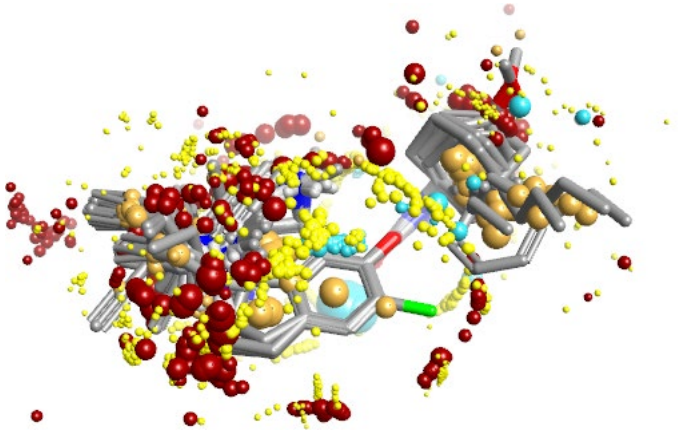
Conclusion and outlook

- Meaningful assessment of electrostatic complementarity at low computational costs
- Possible to rank bioactivities of ligands (provided electrostatics play a main role in affinity changes)
- Caveats: does not calculate free energy of binding ΔG (desolvation, cavity term and space filling, entropic contributions, conformational effects missing)
- Comparison to QM methods shows that XED performs as well or better
- QM methods require a solvation model and have difficulty with charged proteins
- Looking at other improvements:
  - Handling of solvated regions
  - How to handle clipping and EP outliers
  - Ranking docked poses
  - Dynamics/EC

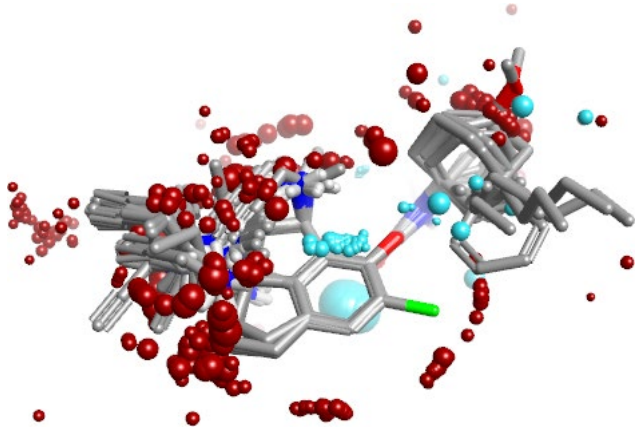


Machine Learning

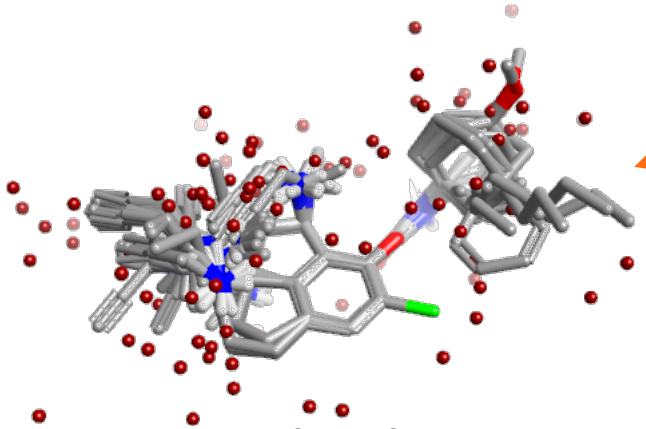
Field QSAR in Forge™



A set of aligned molecules and their field points



Electrostatic field points only



Spaced-out sample of the field point positions, using a sphere exclusion algorithm

	Activity	Electrostatic descriptors				Volume descriptors			
Molecule 1	y_1	e_{11}	e_{12}	e_{13}	e_{14}	v_{11}	v_{12}	v_{13}	v_{14}
Molecule 2	y_2	e_{21}	e_{22}	e_{23}	e_{24}	v_{21}	v_{22}	v_{23}	v_{24}
Molecule 3	y_3	e_{31}	e_{32}	e_{33}	e_{34}	v_{31}	v_{32}	v_{33}	v_{34}
	y	X							

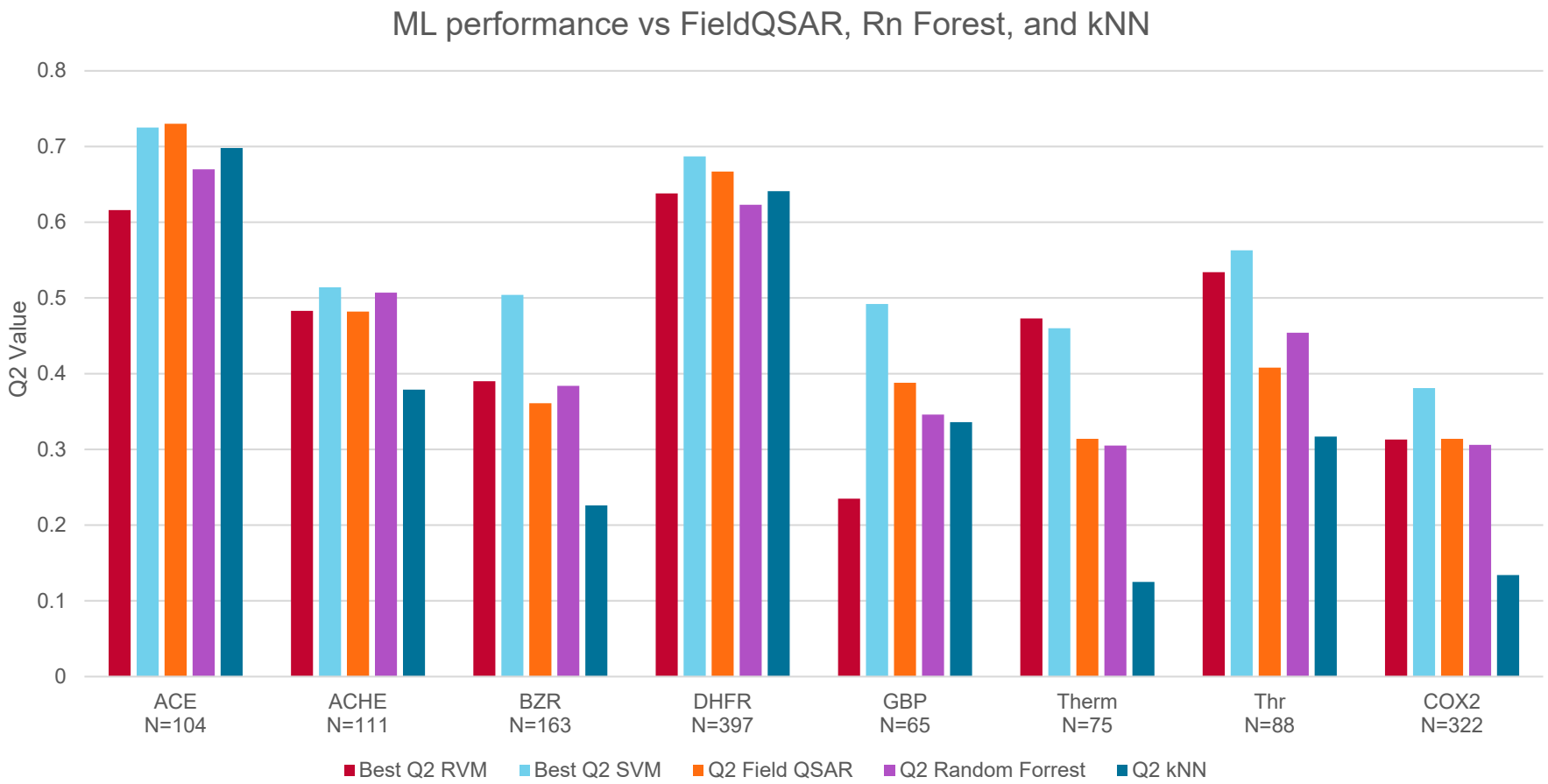
Add more advanced methods than PLS to Forge

New methods:

RVM

SVM

Random Forest



	Best Q^2 RVM	Best Q^2 SVM	Q^2 Field QSAR	Q^2 Random Forrest	Q^2 kNN
mean	0.46	0.54	0.46	0.45	0.36

Conformer Generation

Do we need an improved conformer generator?

Need this

Interpret this as ‘cost to develop/integrate’

This is becoming increasingly important

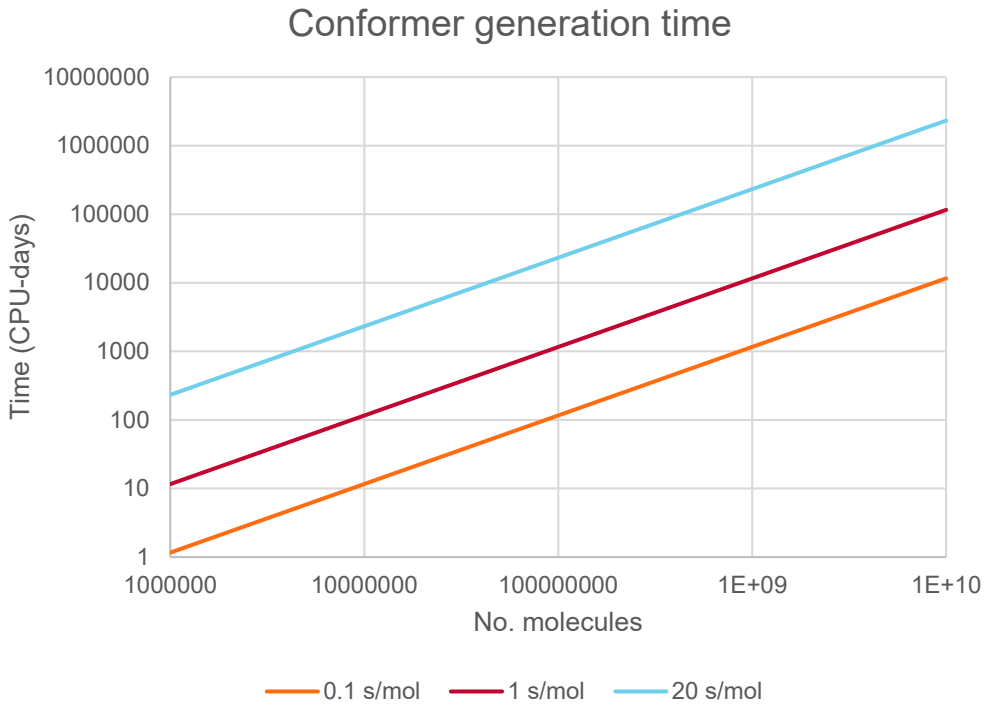
What’s out there? Also, can we integrate it easily?

Simple method producing poor conformations – not interested

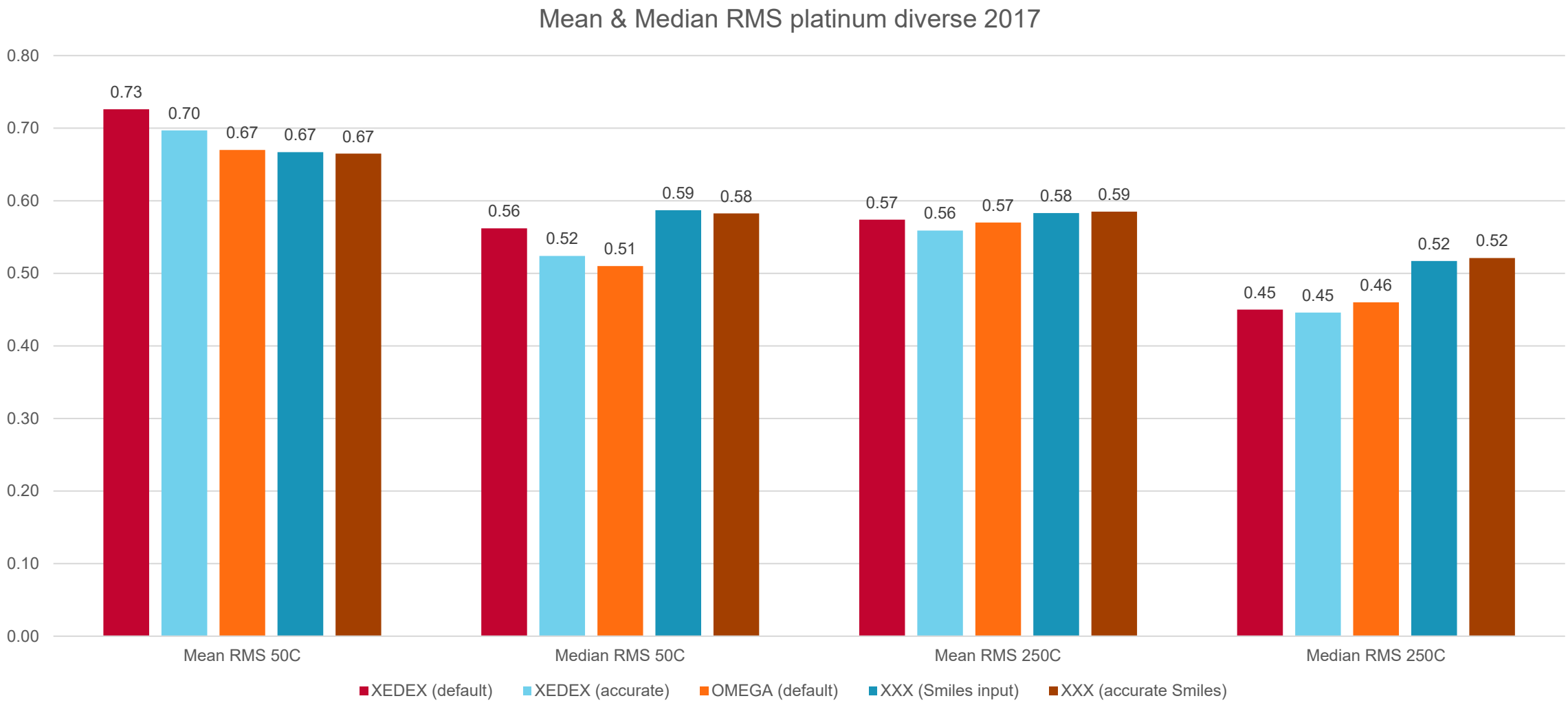
Xedex

Why are we interested?

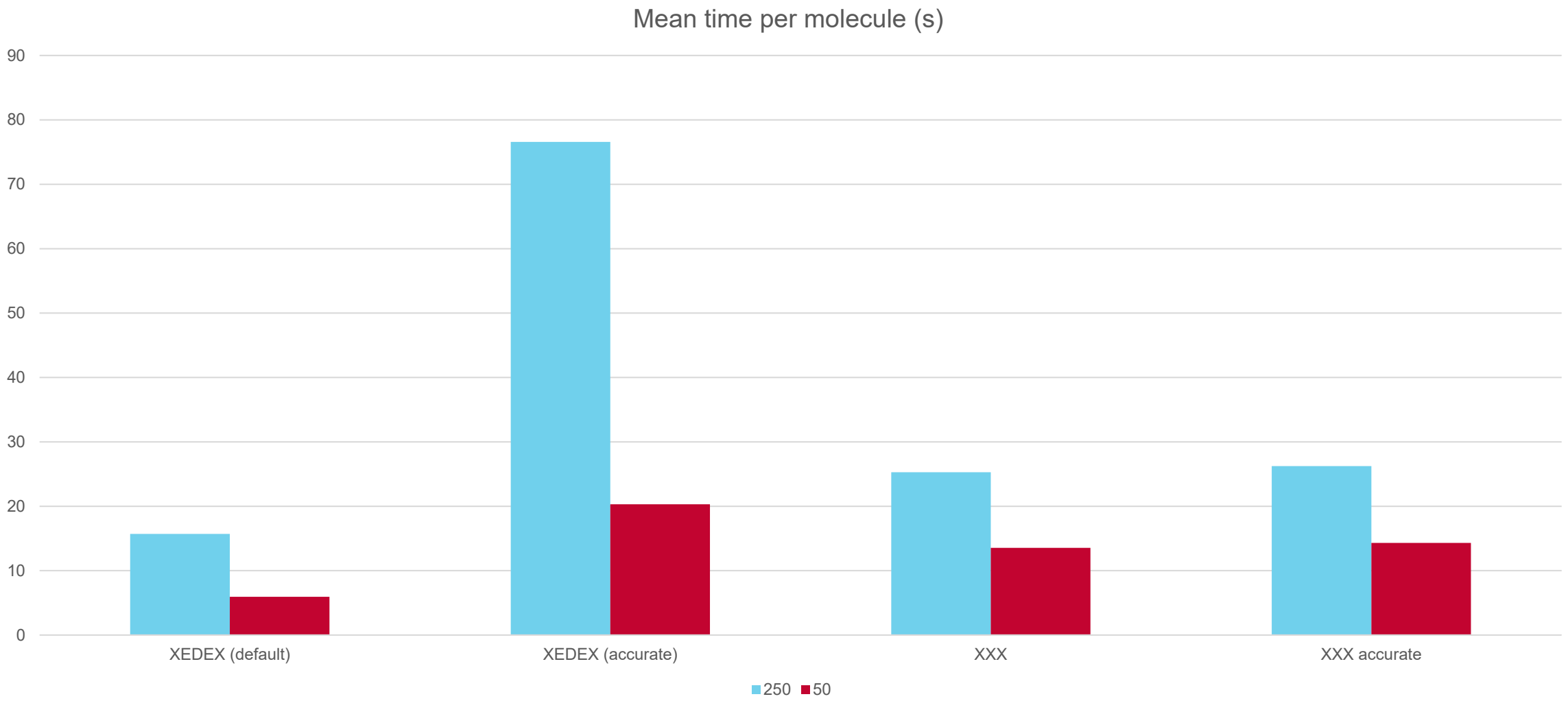
- Better conformations are always nice
- Request from customers: “Can I run Blaze™ on 1Bn molecules?”
  - 1 billion mols @20s/mol = 230 days on 1000 cores
  - Would use ~75TB of disk
- Current Blaze architecture does not scale
  - Re-working Blaze for VLVS
  - Alternative ways to solve the problem
    - Can we use the structure of virtual library spaces to speed up the search?



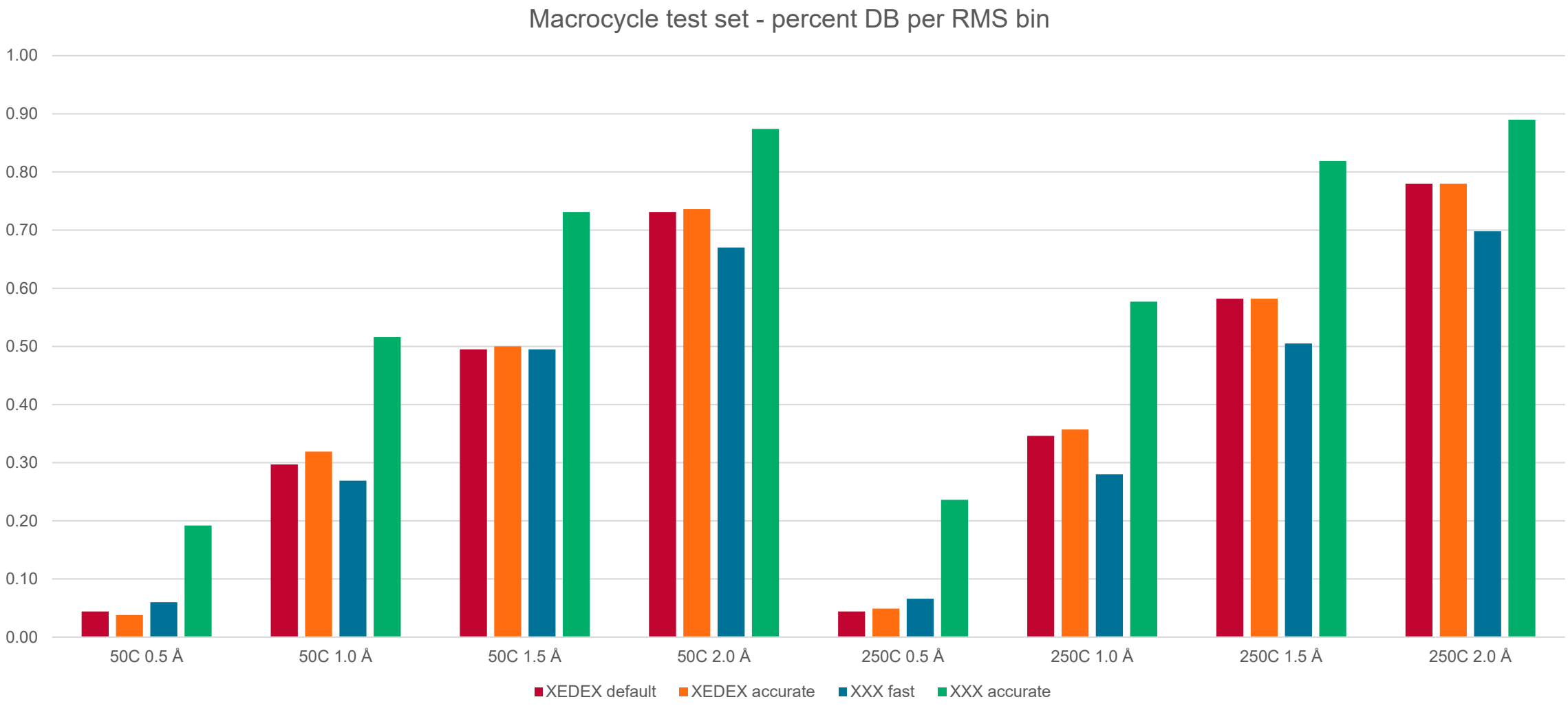
Tried out method XXX from an academic group



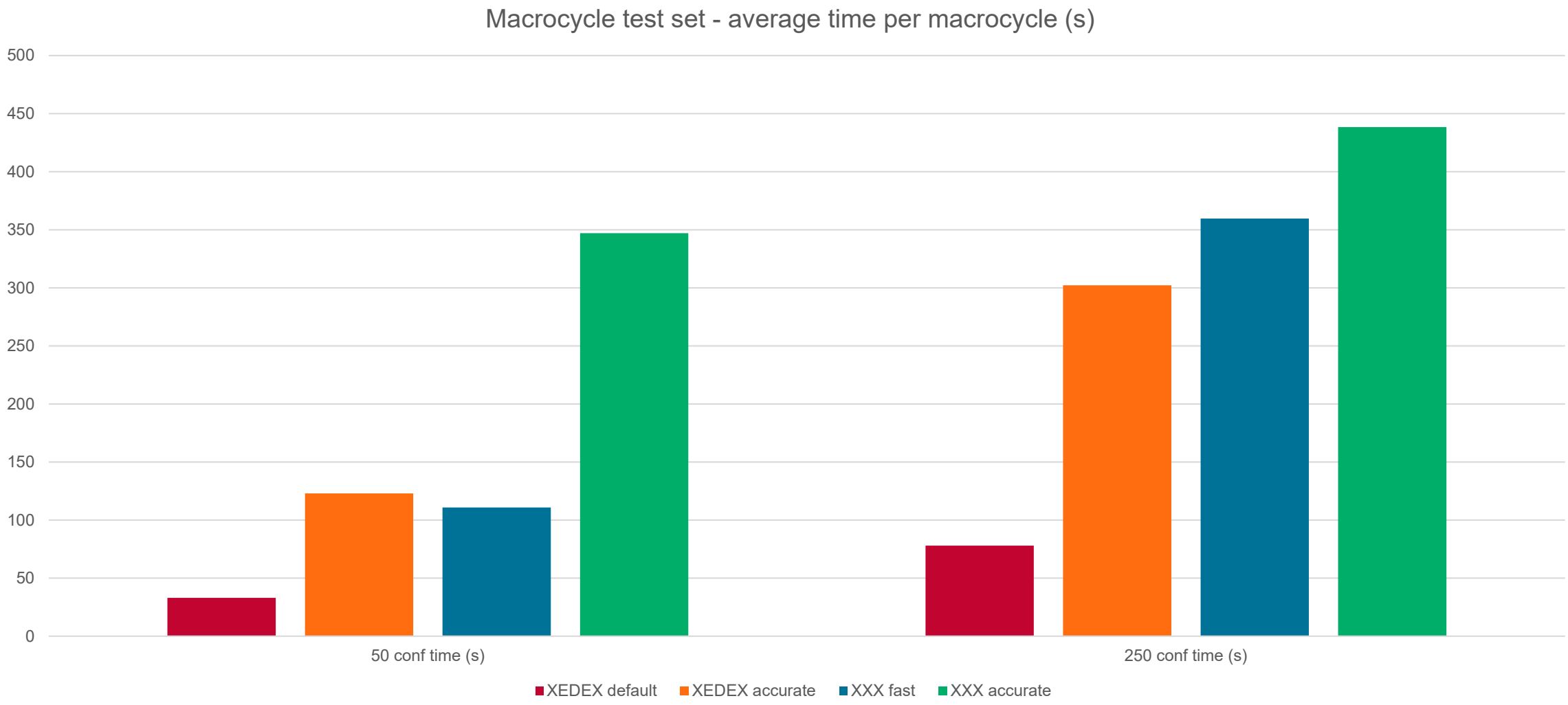
Not significantly faster either



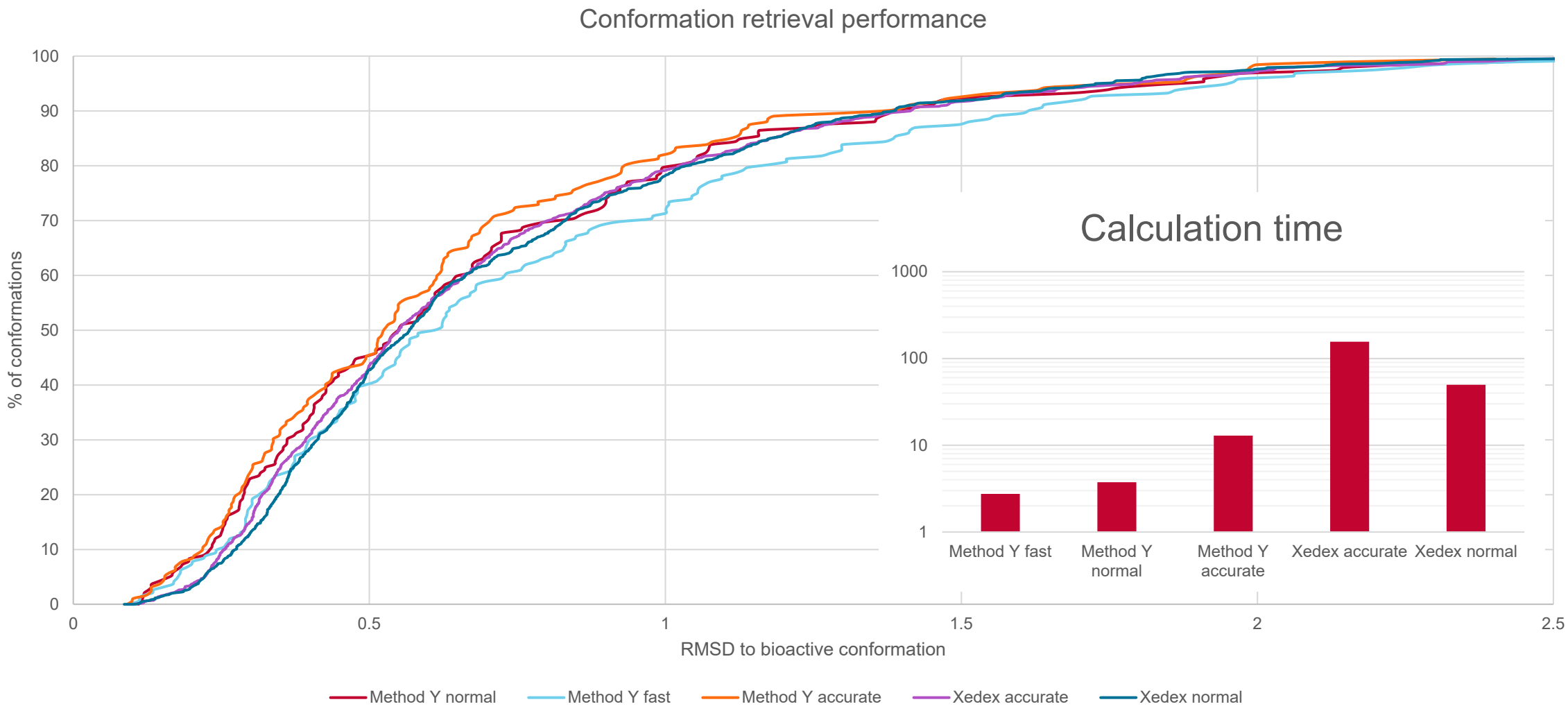
But! Better on macrocycles…



…albeit with a noticeable time cost



Trying another academic method…



FEP

What's Cresset doing about FEP?

- Collaboration with Julien Michel at U. Edinburgh

- Building on top of open-source software
  - AMBER tools
  - OpenMM
  - LOMAP
  - SIRE
  - BioSimSpace
- Launch later in 2019

How does FEP work?

Difference in free energies

Difference in energy between state A and state B

$$\Delta F(\mathbf{A} \rightarrow \mathbf{B}) = F_{\mathbf{B}} - F_{\mathbf{A}} = -k_{\mathbf{B}}T \ln \left\langle \exp\left(-\frac{E_{\mathbf{B}} - E_{\mathbf{A}}}{k_{\mathbf{B}}T}\right)\right\rangle_{\mathbf{A}}$$

Standard Boltzmann stuff

Average over all states of A

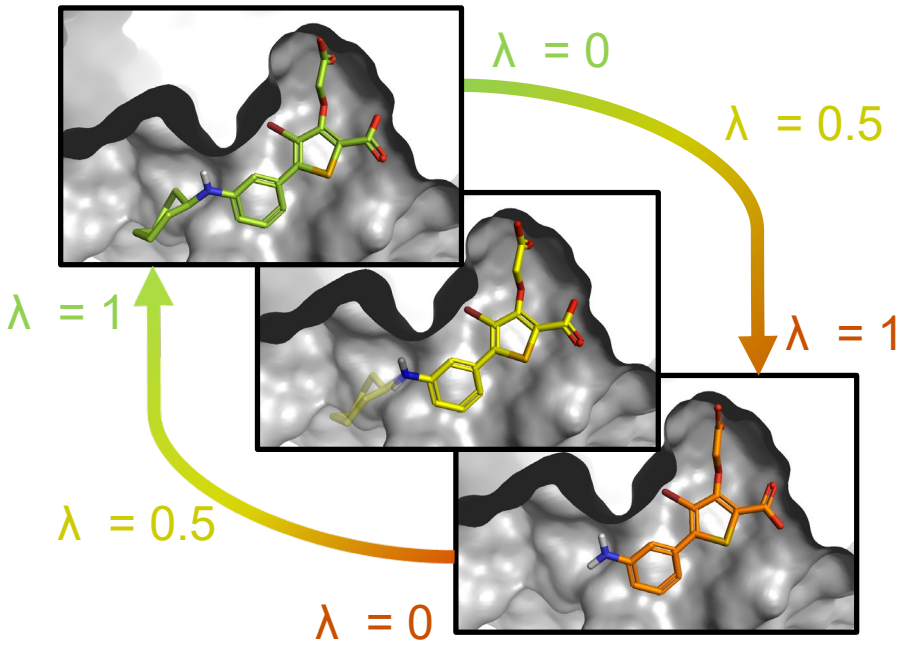
So, do a simulation of state A, calculate the energy at each step of state B as well, and bingo!

Problem: Do we sample all relevant states?

> No. Fix by sampling intermediates!

A → 95% A, 5% B → 90% A, 10% B → → 5% A, 95% B → B

λ =0 λ =0.05 λ =0.1 λ =0.95 λ =1

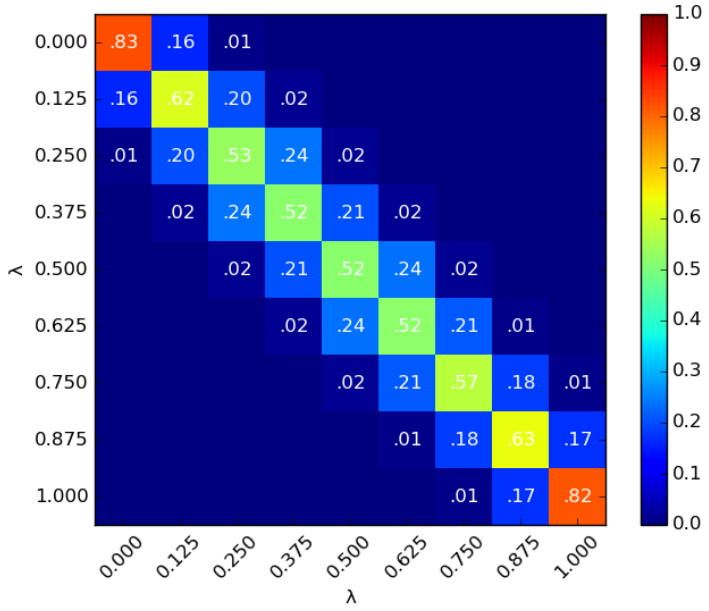


A → 90% A, 10% C → 90% A, 10% B → → C → 90% C, 10% B → → 10% C, 90% B → B

Need to ensure that any adjacent pair of systems are similar enough

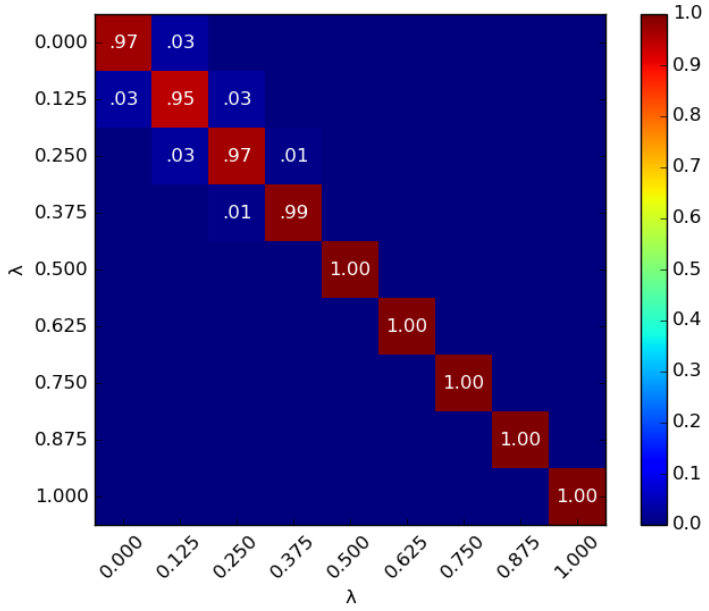
Overlap matrices

Should look like this:



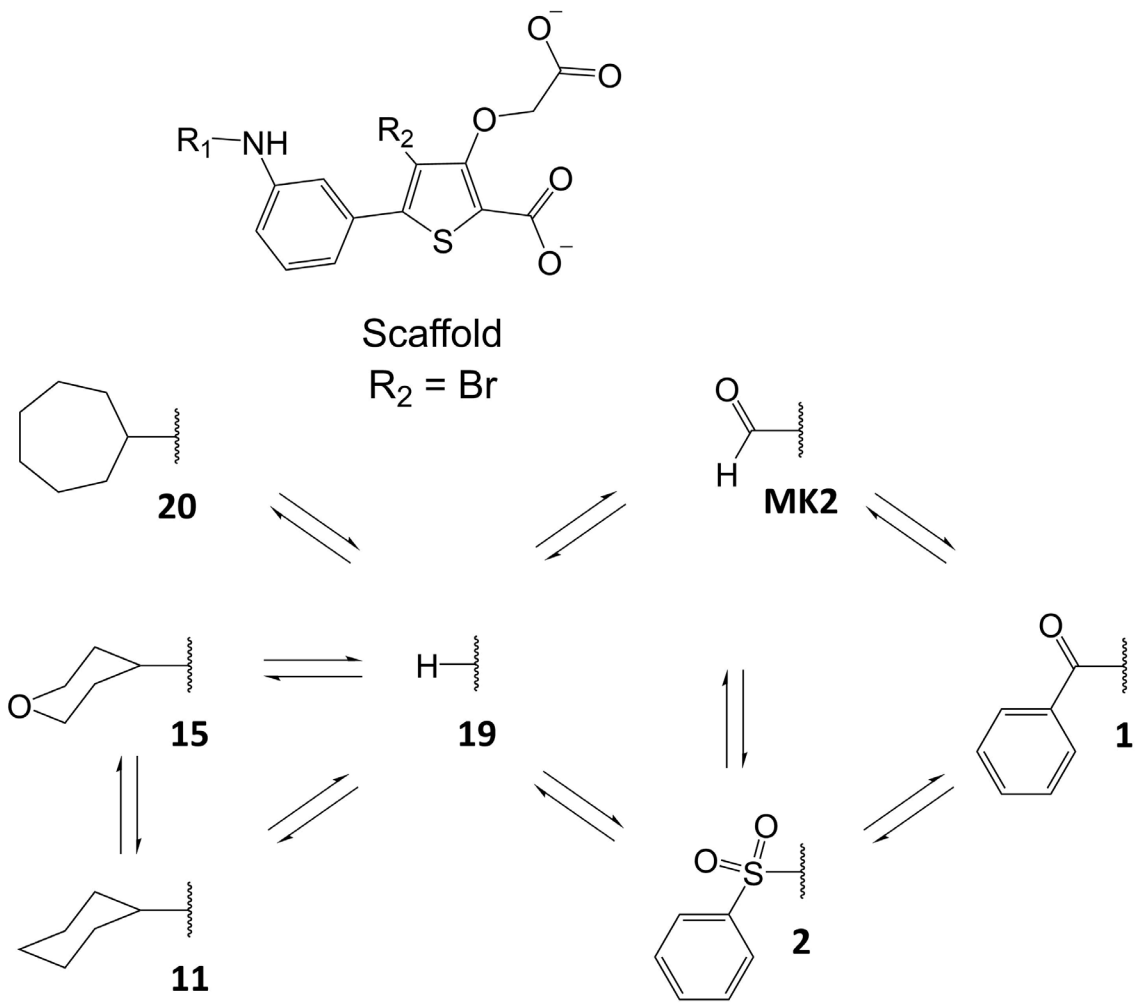
- Visualisation of the phase space overlap between the different states: similarity of microstates
- Some overlap needed so that similar energies can be found between adjacent states

Not like this:



- BUT: unknown how much overlap can be considered (in)sufficient
- Rule of thumb: values in off-diagonal **at least** be 0.02 (preferably higher)

Set up a perturbation map

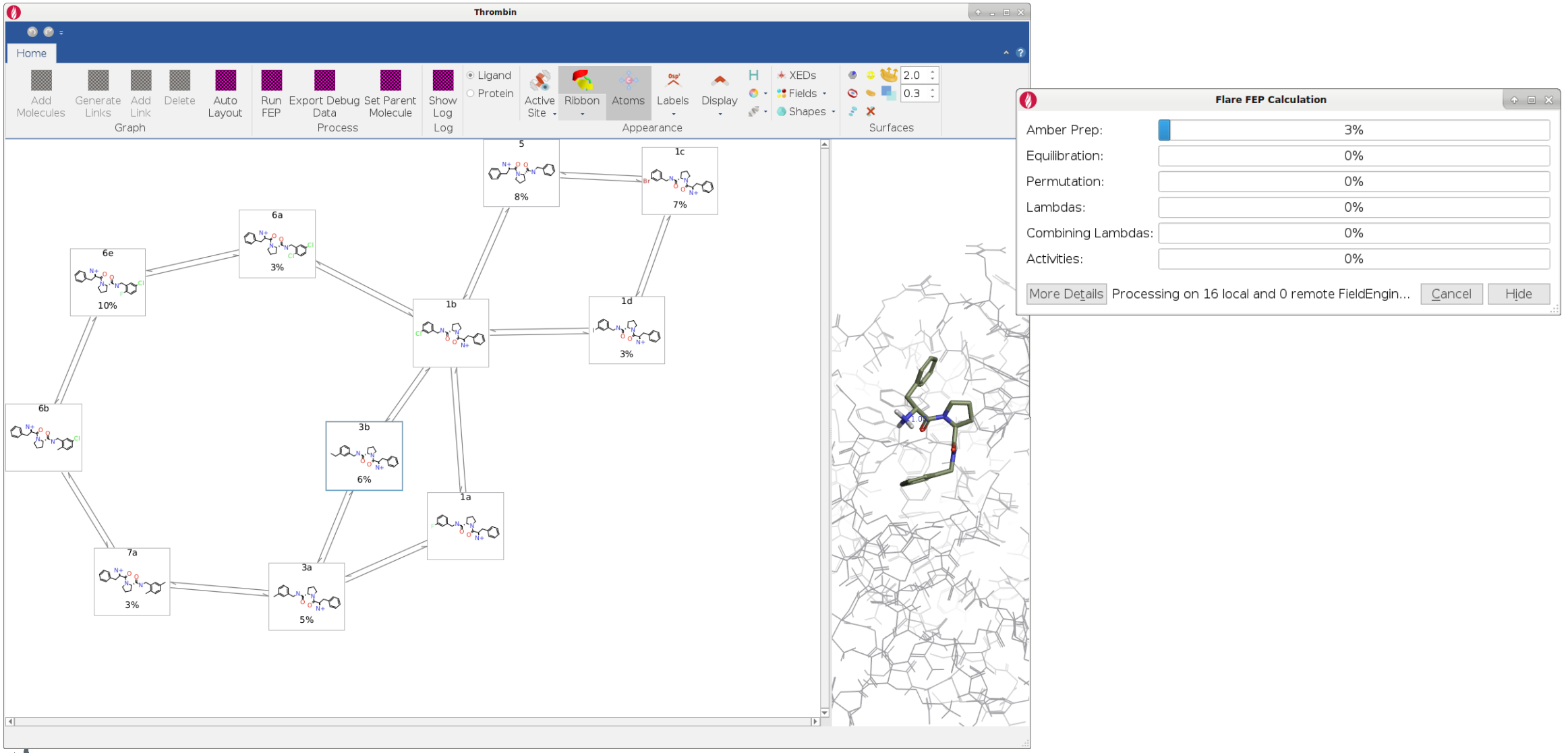


- This is critical for success
  - Sensible peturbations
  - Connected network
- We have an automated method for doing this
- Allow manual modifications
- Decide how many λ windows are needed for each transformation

Preliminary results on standard data sets

Dataset	Cresset/ UoE		Wang et al.		Song et al.	
	R	MUE	R	MUE	R	MUE
Thrombin	**0.88 ± 0.04**	**0.35 ± 0.04**	0.71 ± 0.24	0.76 ± 0.13	0.76	0.46
TYK2	**0.87 ± 0.02**	**0.60 ± 0.04**	0.89 ± 0.07	0.75 ± 0.11	0.57	1.07
PTP1B	**0.83 ± 0.04**	**0.84 ± 0.06**	0.80 ± 0.08	0.89 ± 0.12	0.71	1.06
JNK1	**0.81 ± 0.02**	**0.85 ± 0.04**	0.85 ± 0.07	0.78 ± 0.12	0.47	1.07
MCL1	**0.79 ± 0.02**	**1.30 ± 0.06**	0.77 ± 0.05	1.16 ± 0.10	0.65	1.52
BACE	**0.78 ± 0.03**	**1.08 ± 0.05**	0.78 ± 0.07	0.84 ± 0.08	0.43	1.20
p38	**0.72 ± 0.04**	**1.44 ± 0.05**	0.65 ± 0.09	0.80 ± 0.08	0.38	1.20
CDK2	**0.69 ± 0.09**	**1.02 ± 0.08**	0.48 ± 0.19	0.91 ± 0.12	0.47	0.97

Implementation in Flare



Ongoing research

- Better automated network generation
- Use of overlap matrices to determine optimal λ-window count
- Adaptive λ-window generation
- Parameterisation

Thank you!

Questions welcomed

mark@cresset-group.com

cressetgroup