

Quantum Mechanics based Potentials for Computer-Aided Drug Design

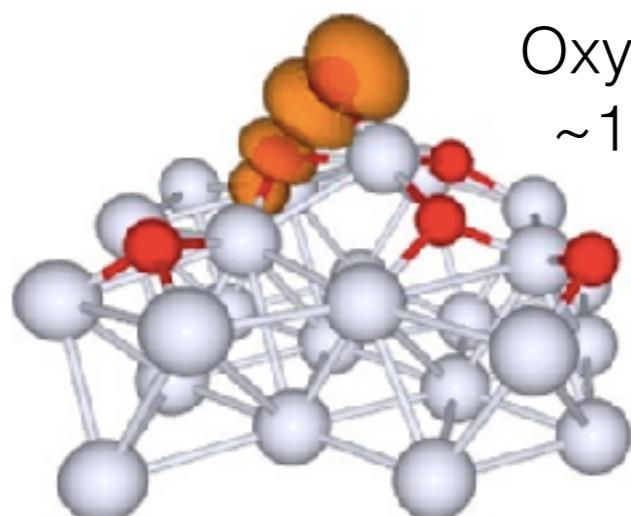
Daniel Cole

School of Natural and Environmental Sciences



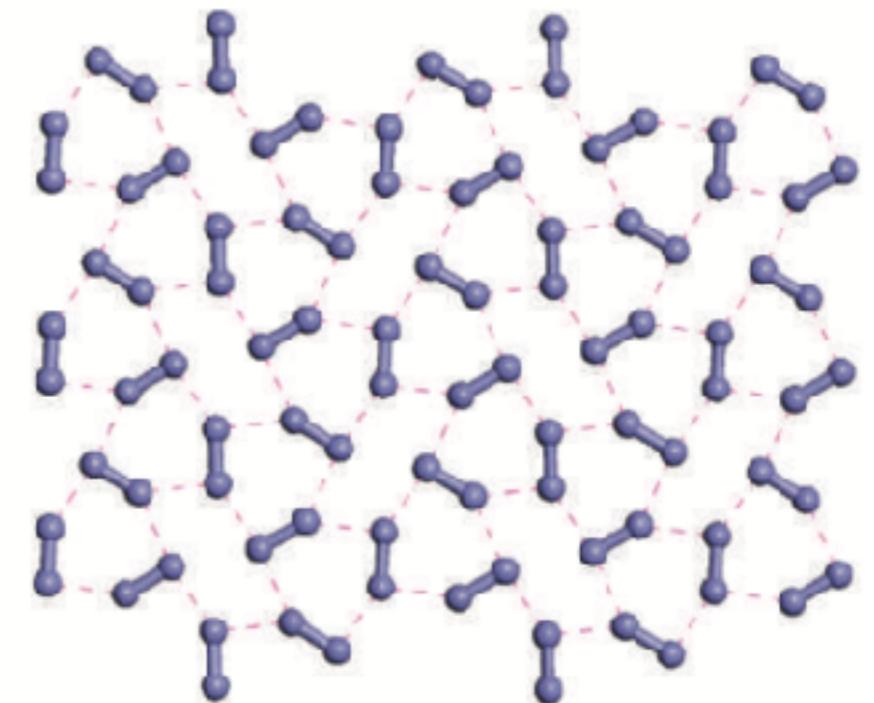
First Principles Quantum Mechanics

Density functional theory (DFT) has been very successful in e.g. the electronics industry for predicting properties of materials and functional mechanisms.



Oxygen absorption on silicon
~100 atoms, 10^{-12} seconds

*L. Colombi Ciacchi and M. Payne,
Phys. Rev. Lett. 95, 196101 (2005)*

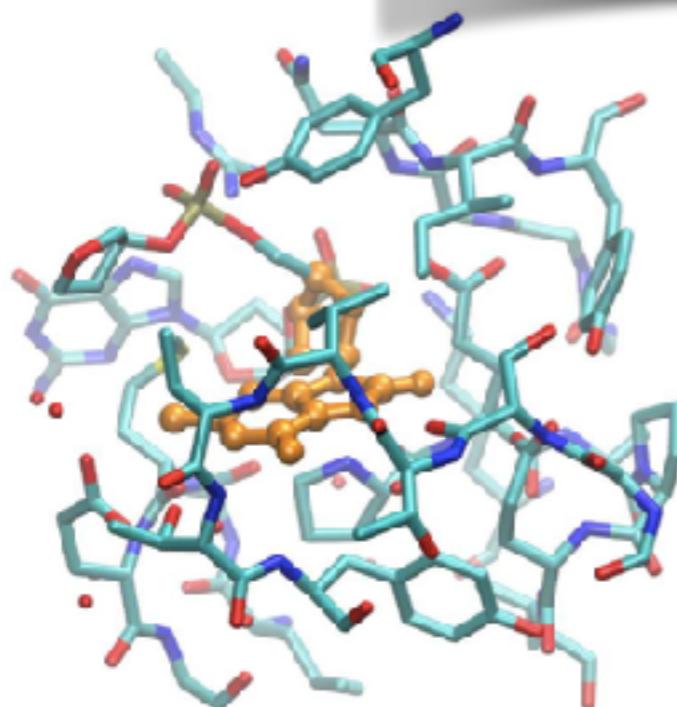
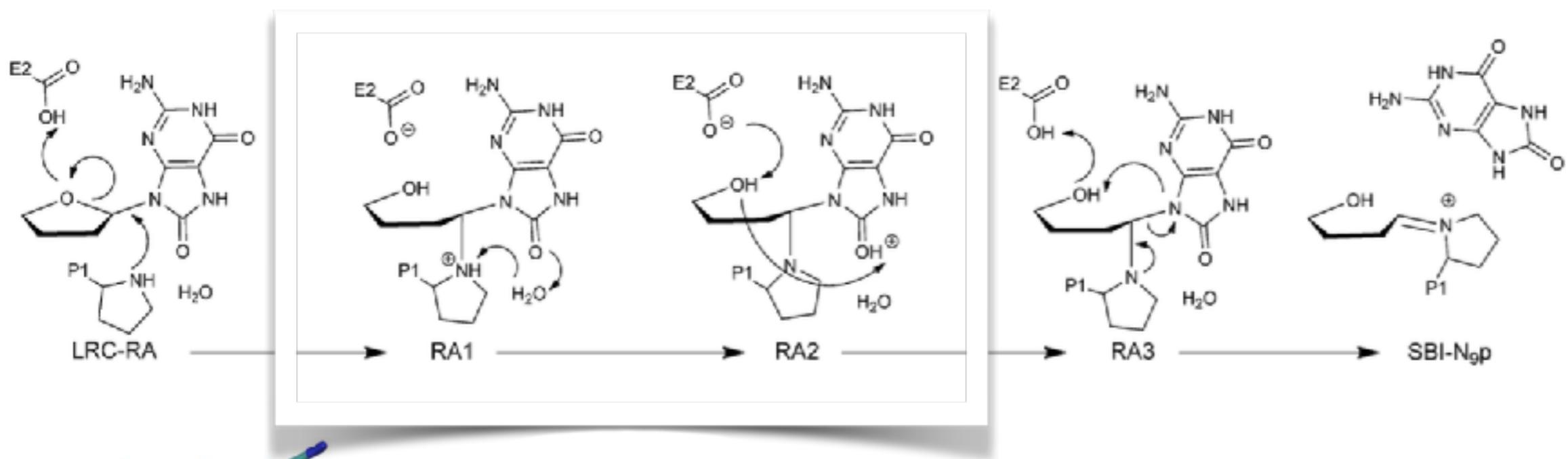


Proposed structure of solid hydrogen at 300 GPa
*C. J. Pickard & R. J. Needs
Nat. Phys. 3, 473 (2007)*

Can the same methods be used in biology for the prediction of structure/function relationships from first principles?

The Length Scale Problem

Deprotonation of proline in the second stage of repair of oxidised guanine by the bacterial glycosylase, MutM



QM atoms	Barrier (kcal/mol)
143	28
278	6
493	14
606	14

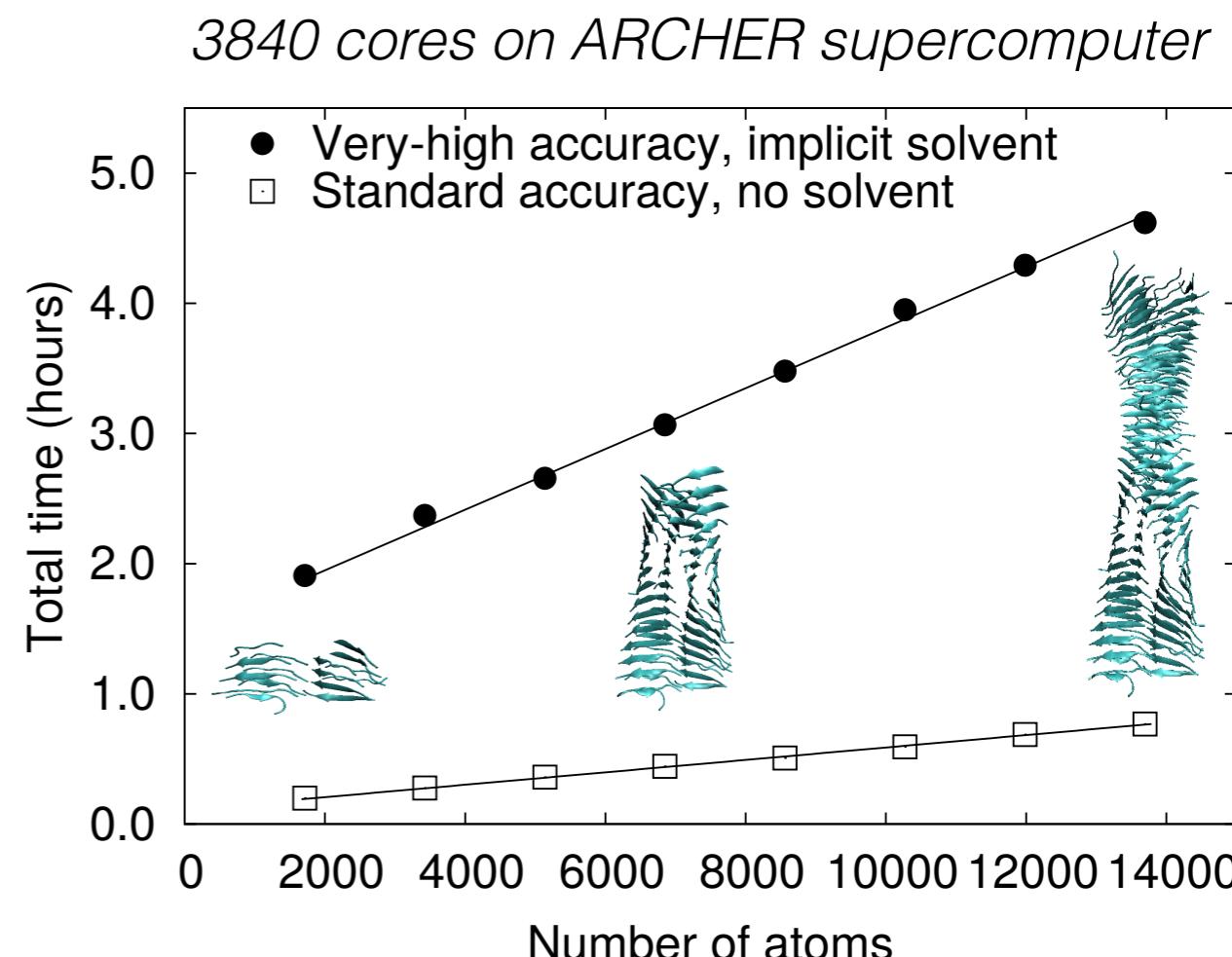
Linear-Scaling DFT

The ONETEP linear-scaling DFT approach combines near-complete basis set **accuracy** with a computational cost that scales **linearly** with system size.

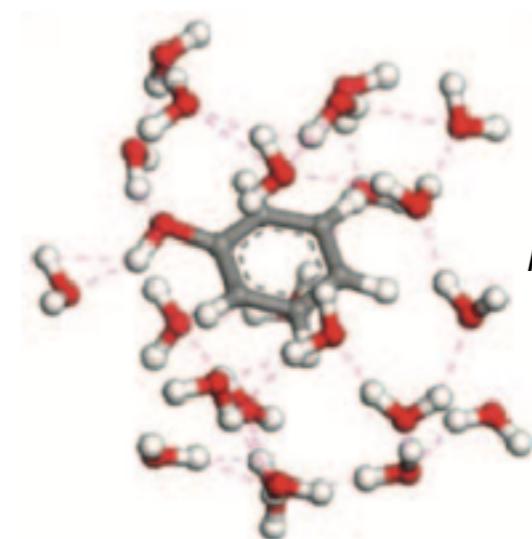
DFT simulations can now be employed in much larger systems, *including entire proteins*.

www.onetep.org/

C. K. Skylaris, P. D. Haynes, A. Mostofi, M. C. Payne
J. Chem. Phys. **2005**, 122, 084119

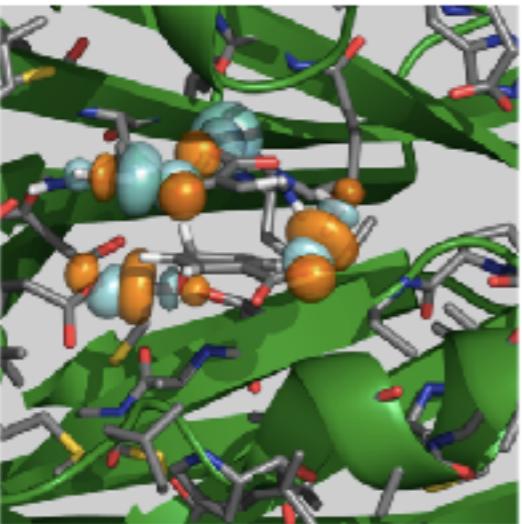


	No. functions	ΔE / kcal/mol
ONETEP	166	-7.04
Gaussian (cc-PVQZ)	3780	-7.22

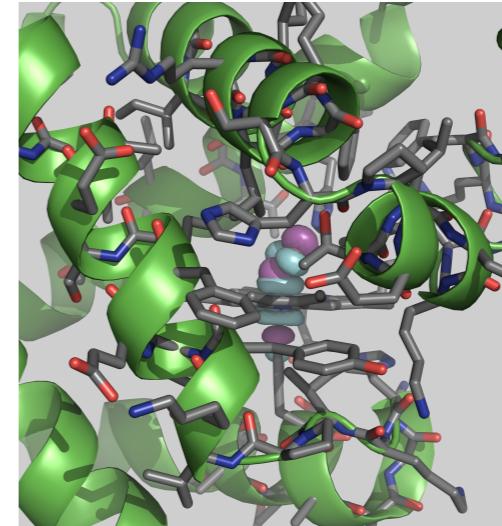


Fox et al., *J. Chem. Phys.* 135, 224107 (2011)

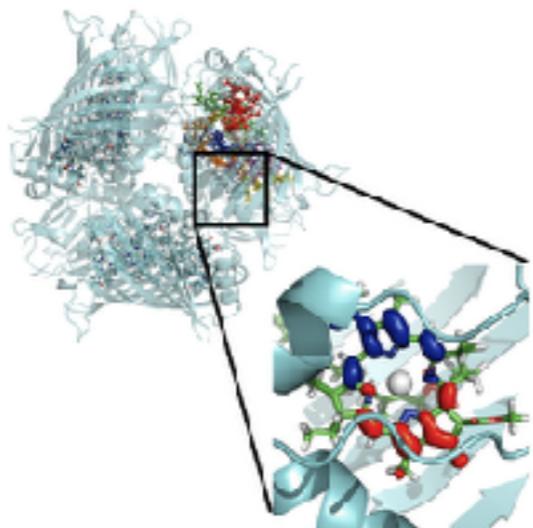
Biological Applications



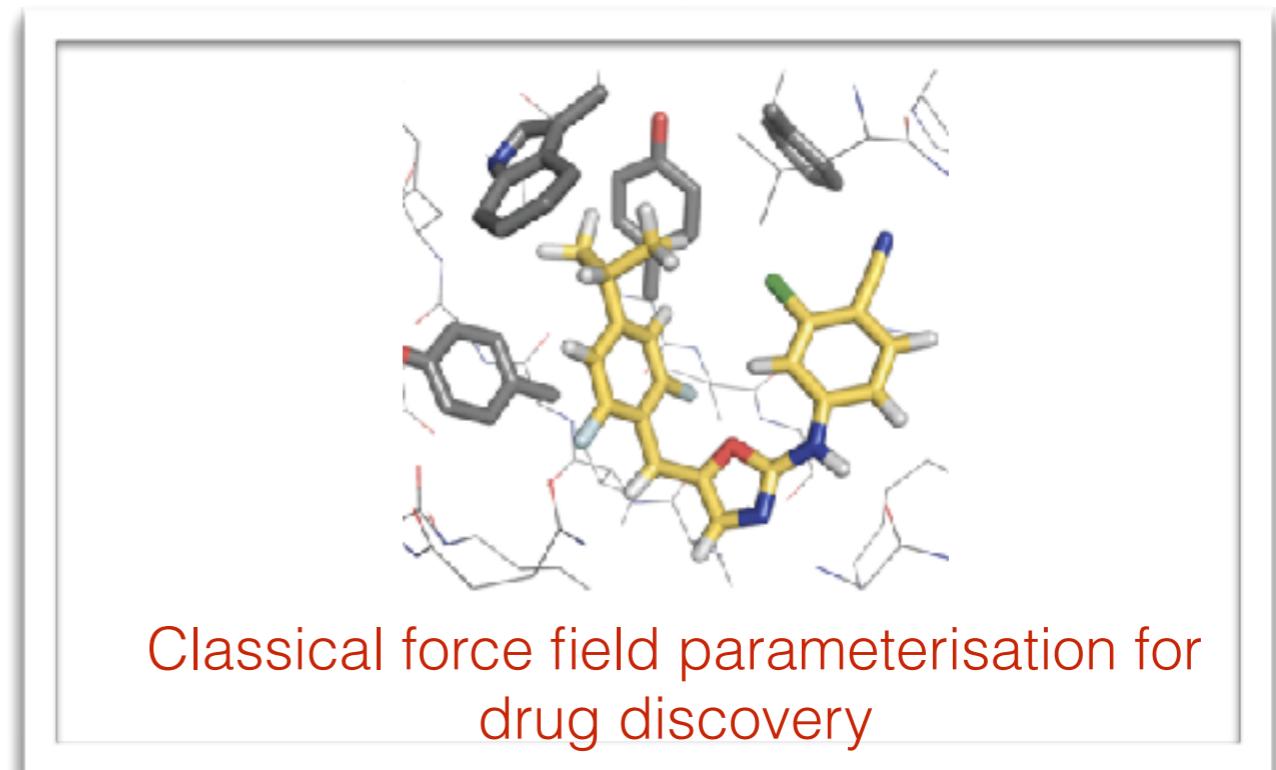
Transition state searching in enzymes



Protein-ligand binding in metalloproteins



Optical spectroscopy in a light-harvesting protein



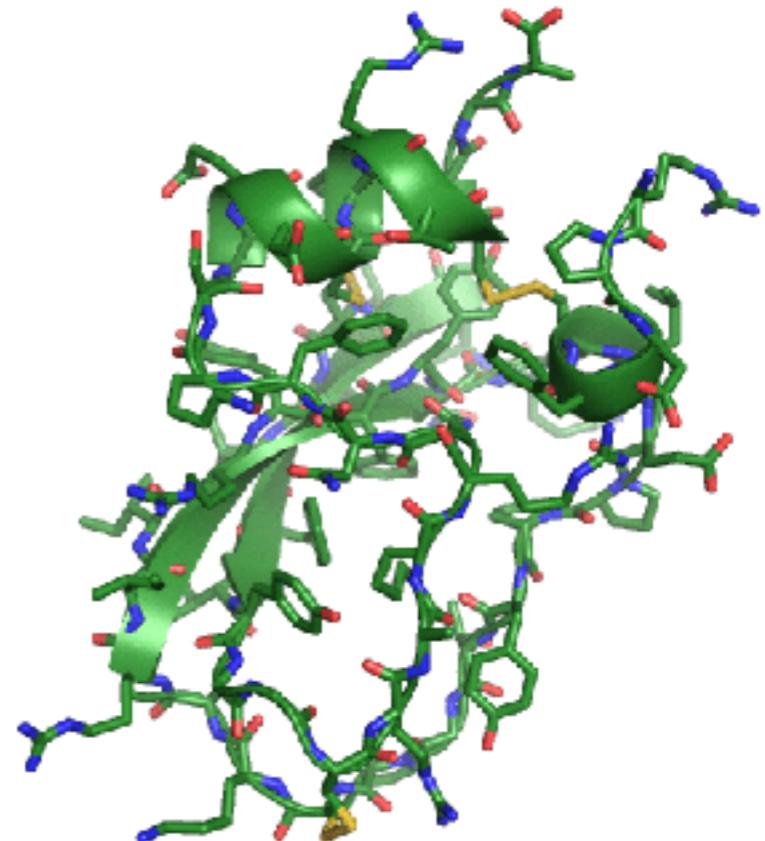
Introduction

Molecular mechanics (MM) force fields are widely used in computer-aided drug design, protein folding, protonation states, protein-surface interactions, allosteric mechanisms, photochemistry, QM/MM....

$$E^{MM} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right)$$

MM can complement experiment in many ways. At the simplest level, we can use molecular dynamics to ‘animate’ the system.

Simulations tend to be limited only by finite sampling and **accuracy of the force field**.



FEP for Drug Discovery

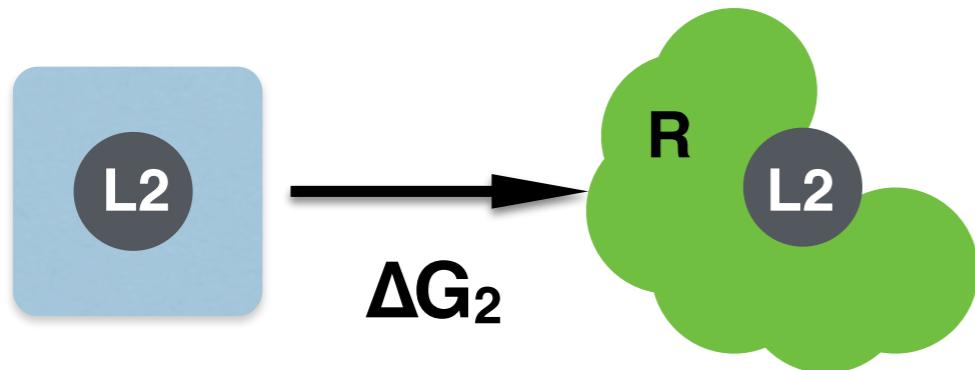
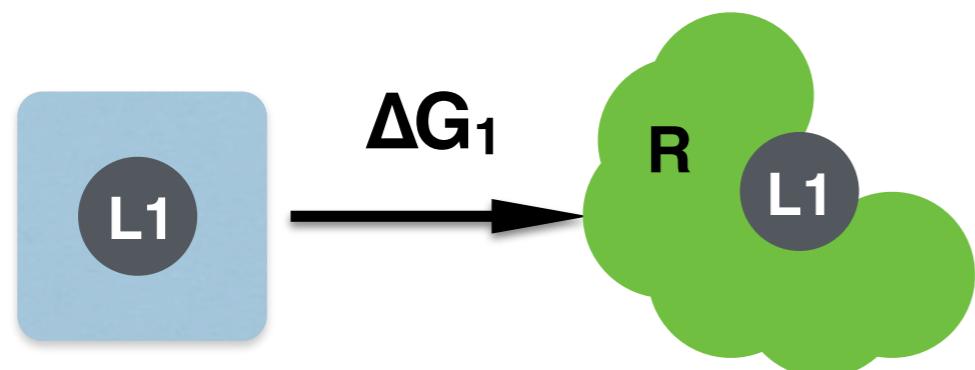


In lead optimisation studies, we are typically interested in optimising the target-ligand binding affinity.

In other words, we need to find the free energy difference between a small molecule (L_1) in solution and bound to the protein (R).

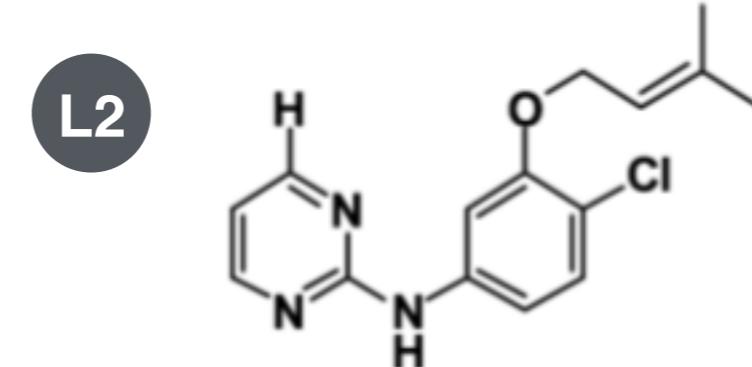
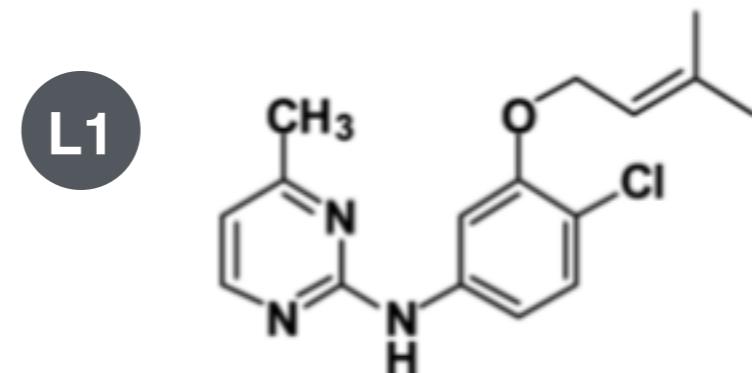
Free energy perturbation (FEP) theory provides a rigorous means to compute the binding free energy.

FEP for Drug Discovery



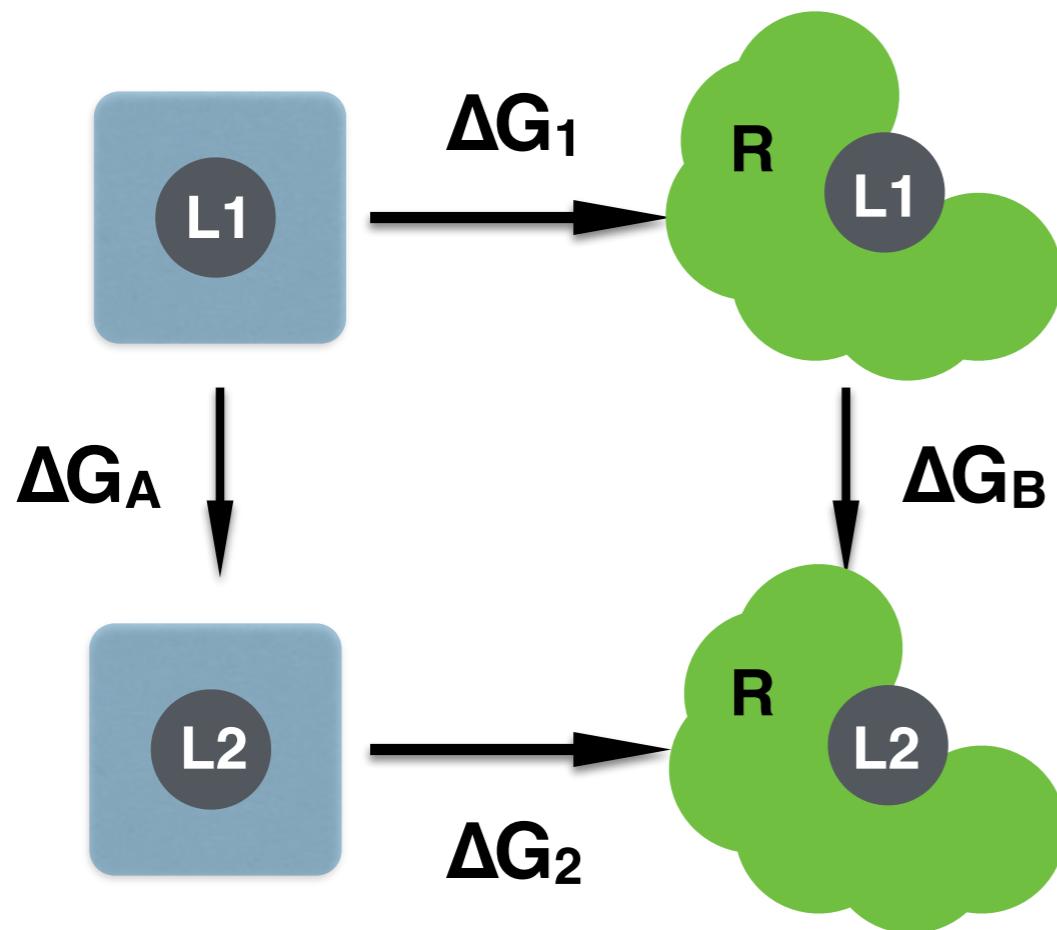
$$\Delta\Delta G = \Delta G_2 - \Delta G_1$$

For example:



If we have two similar molecules, then often we only need to compute the relative binding free energy $\Delta\Delta G$.

FEP for Drug Discovery



The total free energy change around a closed loop is zero:

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A$$

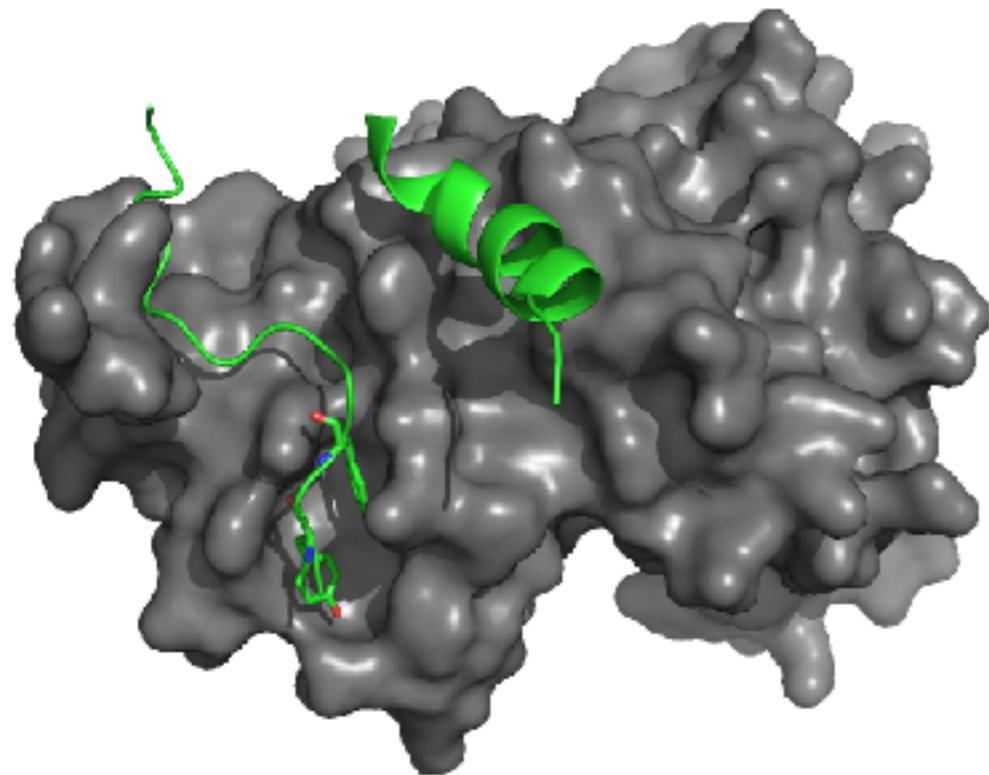
Free energy changes computed using Zwanzig equation:

$$\Delta G_A = -kT \ln \left\langle \exp \left[\frac{-(U_{L2} - U_{L1})}{kT} \right] \right\rangle_{L1}$$

We can use FEP to transform molecule L1 into molecule L2 in the protein and in water. Conformational sampling performed using force field:

$$E^{MM} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right)$$

Computer-Aided Drug Design

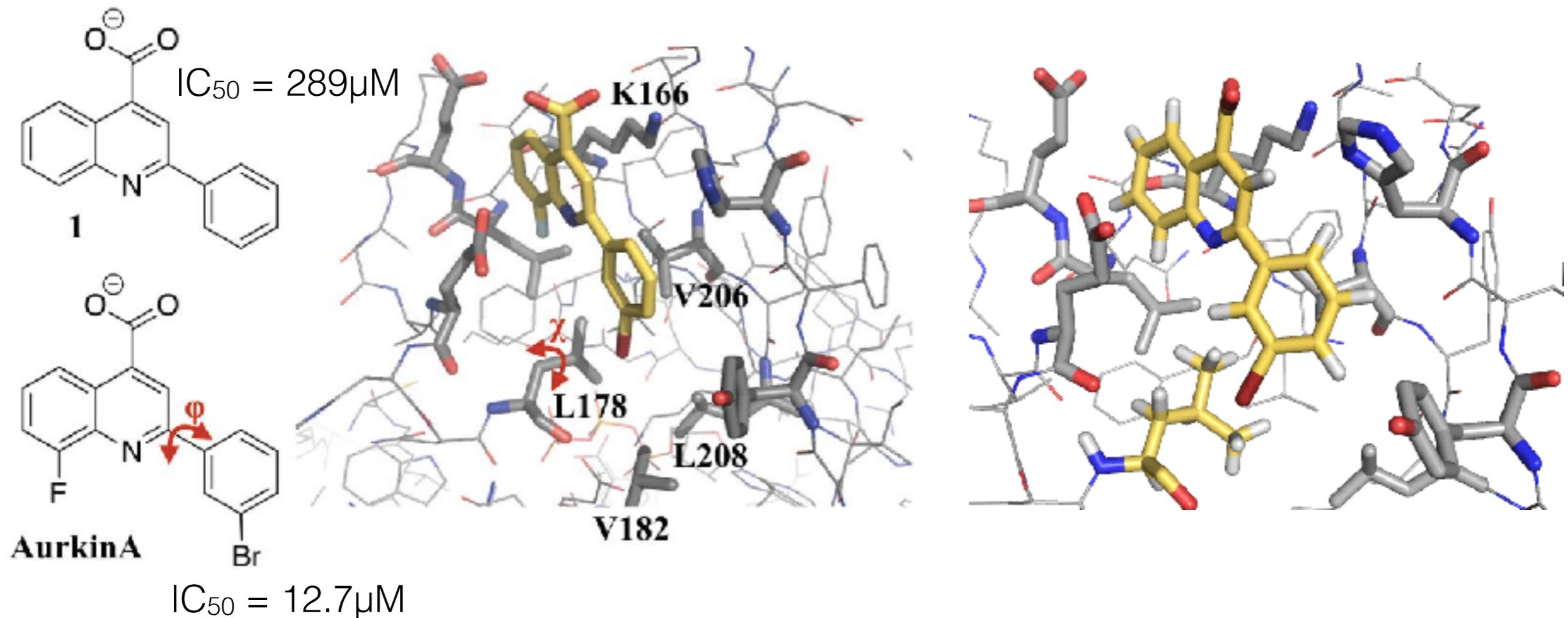


Aurora A kinase plays a central role in cell division. It is oncogenic and over-expressed in various tumour types.

Inhibition of Aurora A leads to cell death in dividing cells and it is a potential drug target in cancer.

It has recently been shown that an allosteric surface site can be targeted. Interruption of the protein-protein interaction with TPX2 reduces kinase activity.

Computer-Aided Drug Design



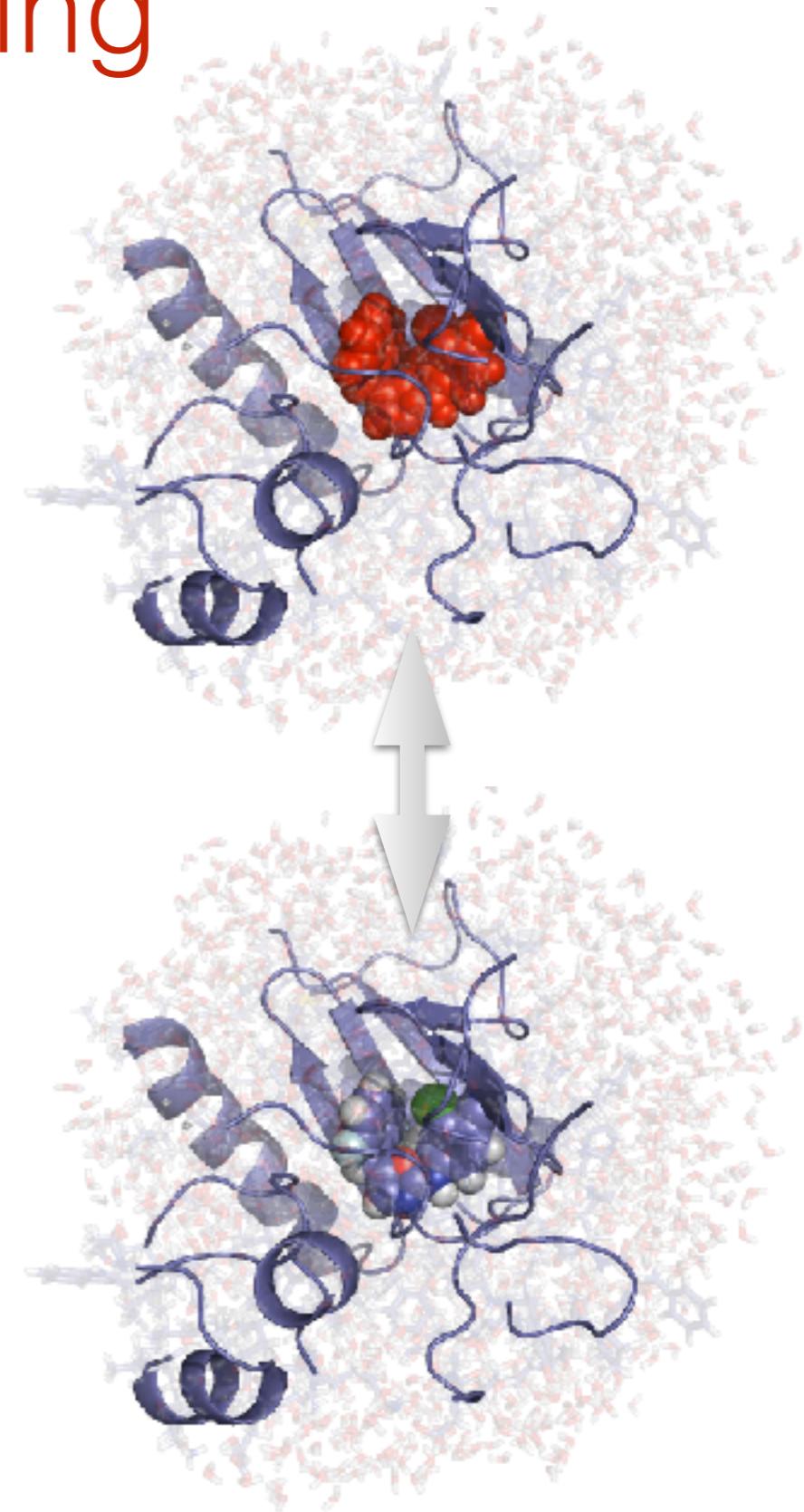
Used FEP to investigate small substitutions around the phenyl ring.
Replica exchange with solute tempering (REST) enhanced sampling of ligand and L178.

Enhanced Sampling

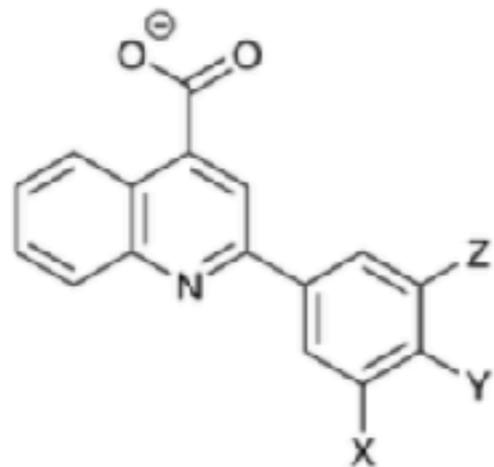
In replica exchange with solute tempering (REST) method, the potential energy is scaled with temperature → molecule/fragments of interest appears to get hotter, while the surroundings remain ‘cold’.

High temperature replicas explore more conformational space.

Replica exchange propagates conformational changes to room temperature ensemble.

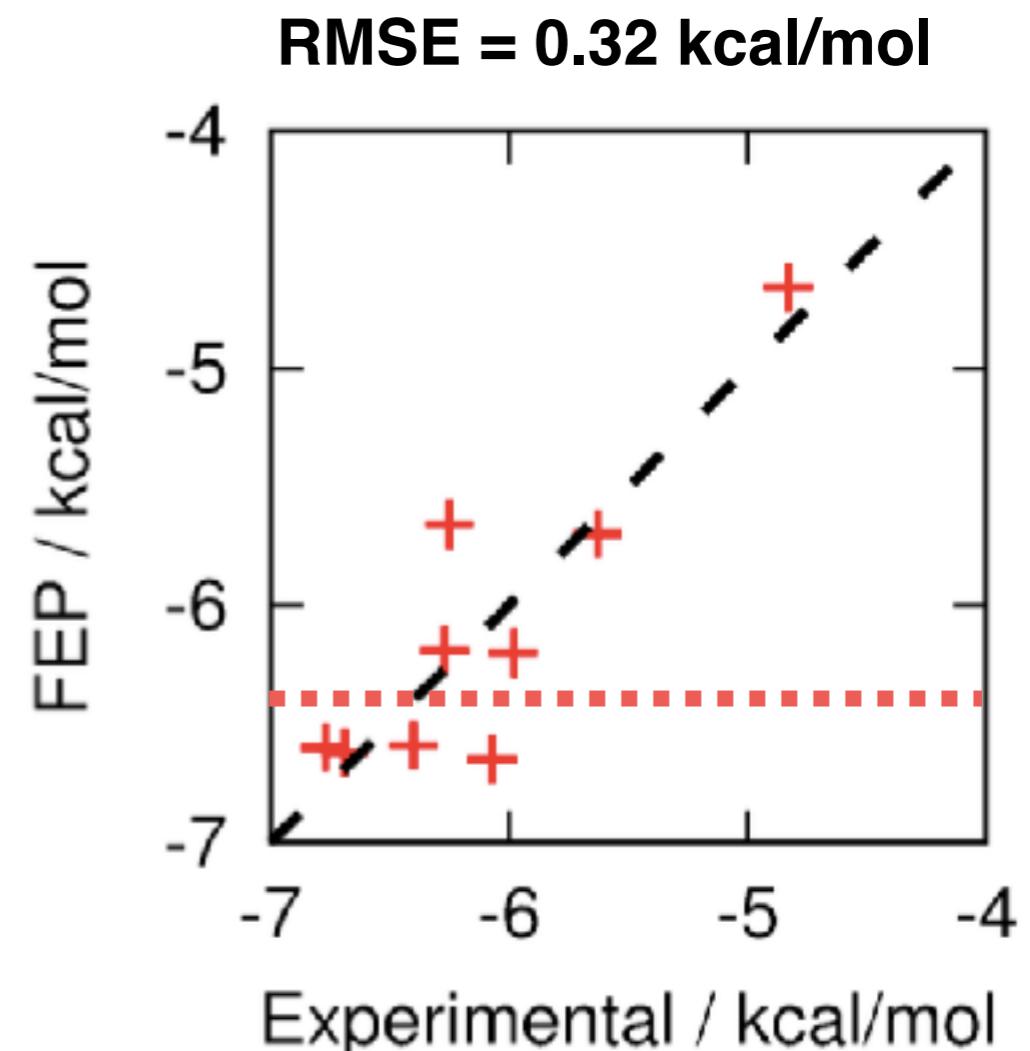


Computer-Aided Drug Design



	X	Y	Z	$\Delta\Delta G^a$	IC_{50}^b	K_i^b
1	H	H	H	1.05	289	62.5
2	F	H	H	0.00	75.9	16.5
3	F	H	F	-0.94	36.0	7.8
4	Cl	H	H	-0.73	ND	ND
5	Cl	H	F	-0.89	20.5	4.4
6	Br	H	H	-0.49	25.6	5.5
7	CF_3	H	H	0.11	26.5	5.7
8	CH_3	H	H	1.12	ND	ND
9	F	CH_3	H	-0.49	42 ^c	8.7 ^c
10	Br	CH_3	H	-0.90	11.1 ^c	2.3 ^c

^a kcal mol⁻¹. ^b μM . ^c This work (see the ESI).



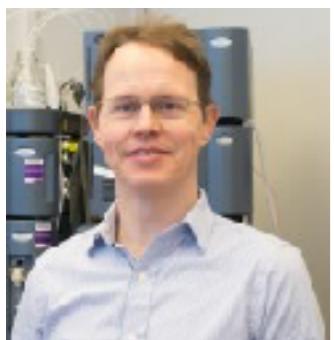
Medicinal Chemistry Research at Newcastle



**Mike
Waring**



**Celine
Cano**



**Ian
Hardcastle**



**Agnieszka
Bronowska**

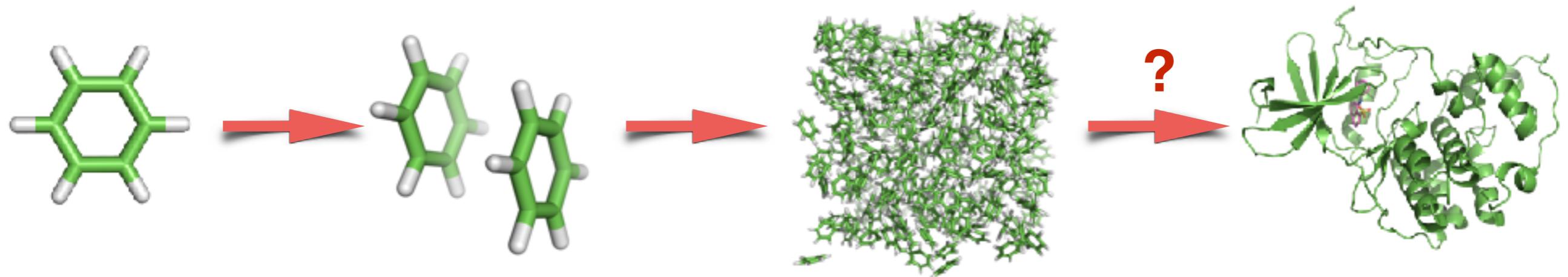
Close ties between Chemistry,
Northern Institute for Cancer
Research, and Medical School.



Strategic alliance with Astex
Pharmaceuticals (Cambridge)
and Cancer Research UK.

Disadvantages of Force Field Design

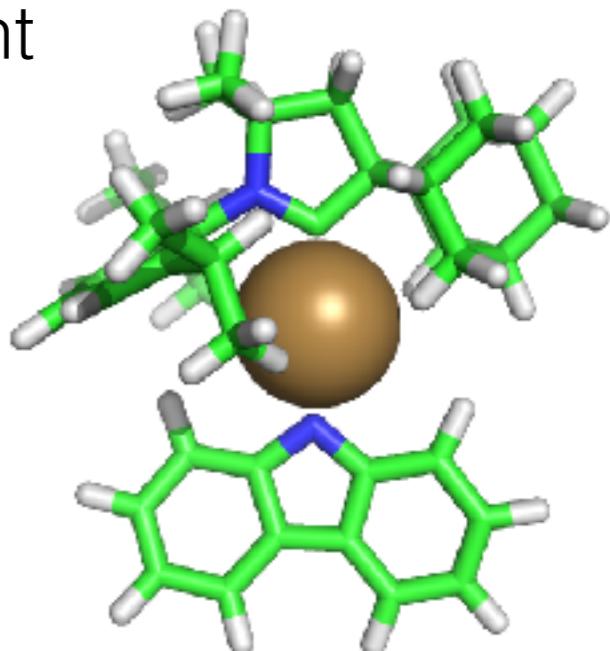
Transferability: Force field parameters are obtained by fitting to properties of small molecules:



Labour-intensive: Typically requires large groups, significant time to produce a model useful for practical problems.

Functional form: Is all of the required physics contained within the simple functional form of MM force fields?

Ease of use: Missing parameters are problematic for inexperienced force field users (metals, excited states?).

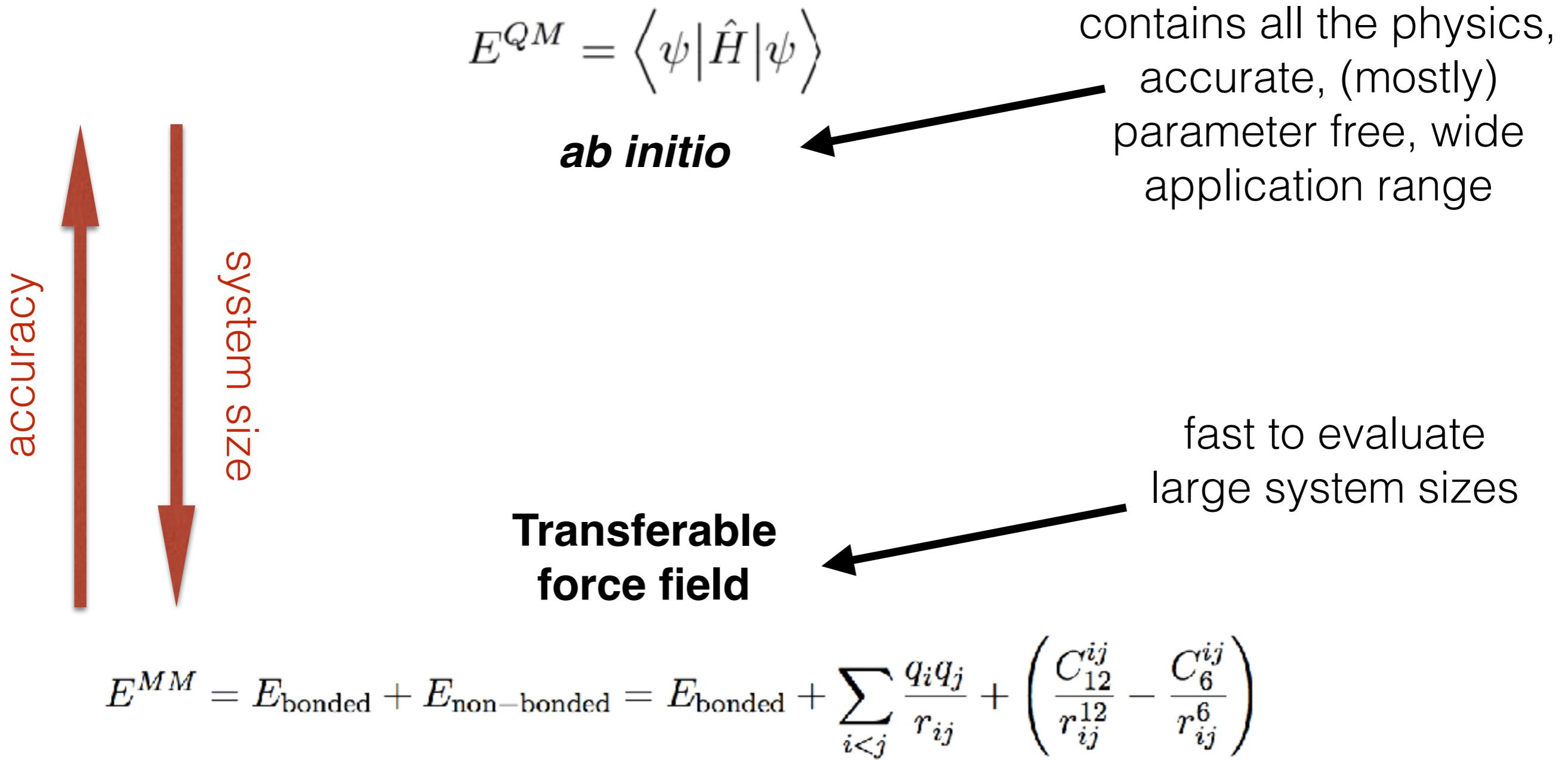


Disadvantages of Force Field Design

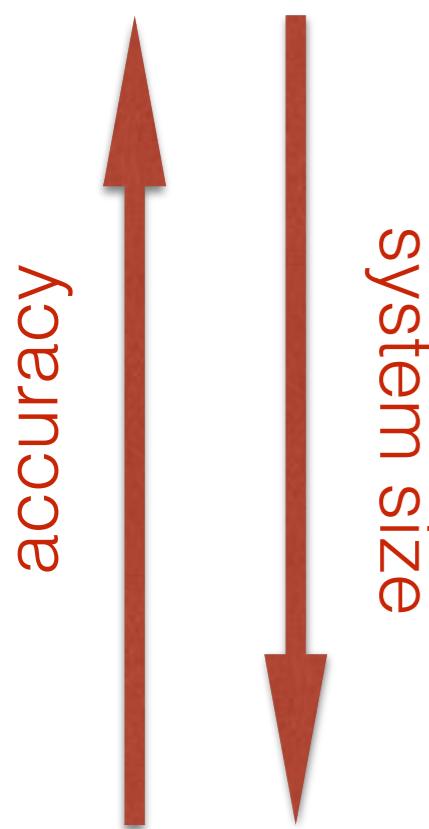
“Classical force field parameterization
is a bottomless pit of despair”

- Twitter user

QUBE Force Field



QUBE force field



$$E^{QM} = \langle \psi | \hat{H} | \psi \rangle$$

ab initio

**Quantum mechanical
BEspoke (QUBE) force field**

**Transferable
force field**

system-specific
potentially more accurate
(virtually) parameter free
fast to evaluate
large system sizes
automated
wide application range

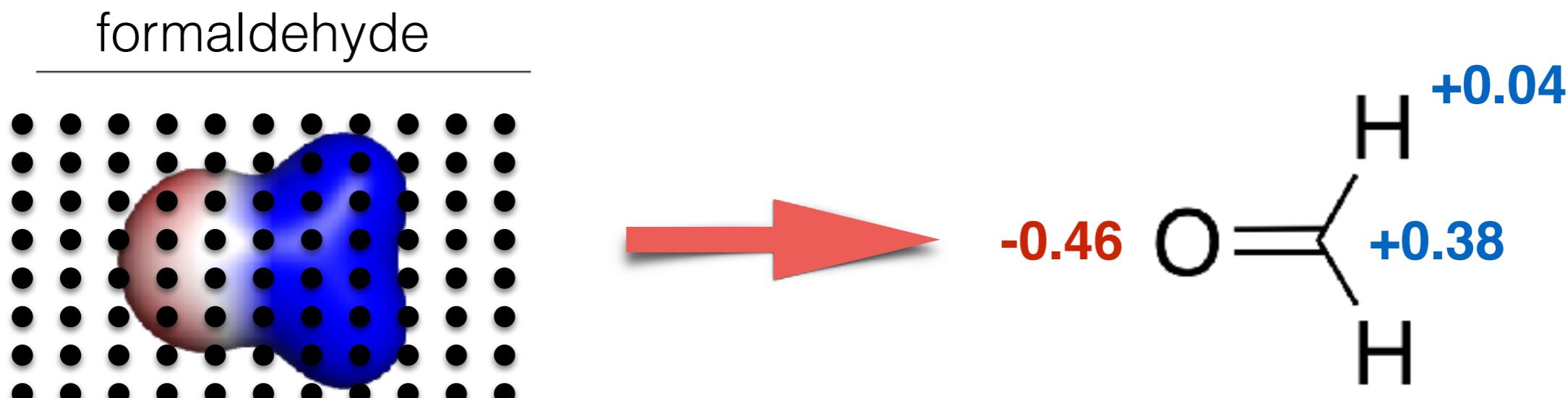
I will now go through each part of the force field in turn...

$$E^{MM} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right)$$

ESP Charges

In force fields we attempt to represent the full QM electrostatic potential using a simple charge placed on each atom.

ESP charges are usually used in standard force fields to best reproduce the electrostatic potential of a small molecule. However, this procedure is infeasible for proteins.

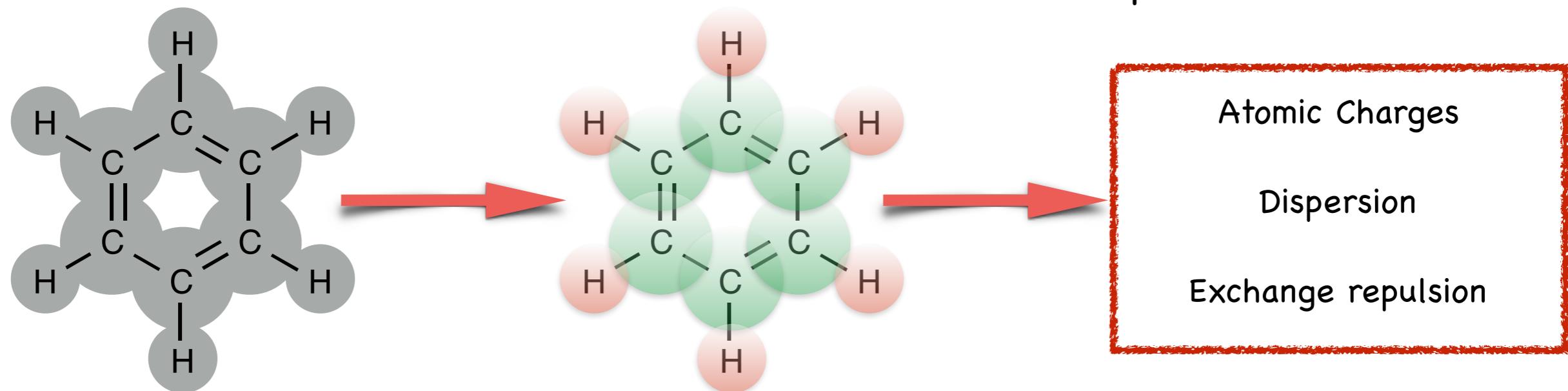


Atoms-in-Molecule

DFT Calculation computes
total electron density

Electrons partitioned amongst
the atoms in the system

Atomistic force field
parameters computed directly
from partitioned electron density

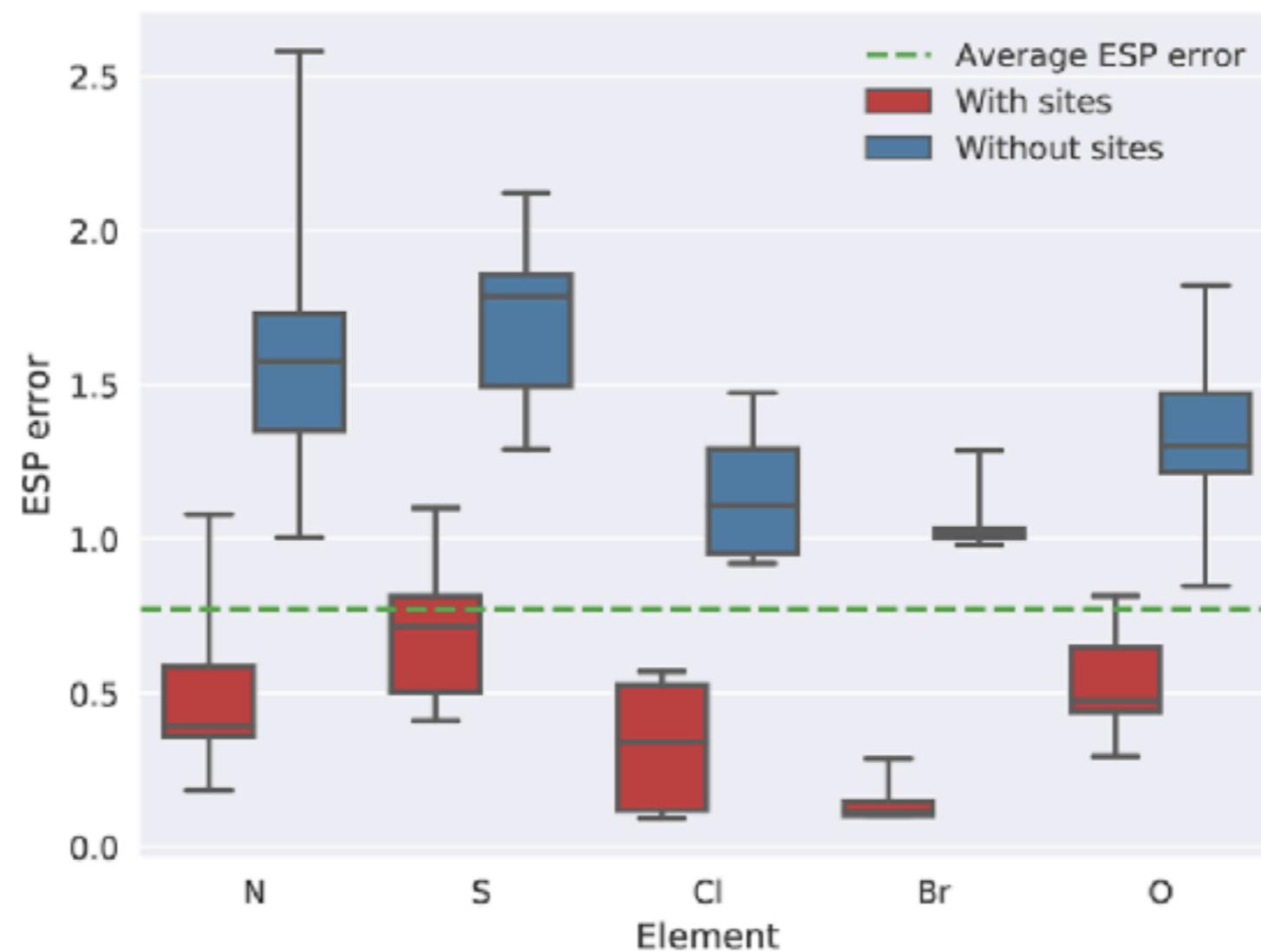


Density derived electrostatic and chemical (DDEC) electron density partitioning
(good reproduction of the ESP and not too conformation dependent).

Charges are computed in implicit solvent to account for induction effects.

Off-Centre Charges

Anisotropic electron density cannot be accurately modelled using atom-centred point charges:

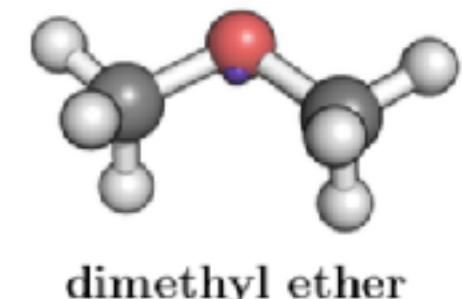
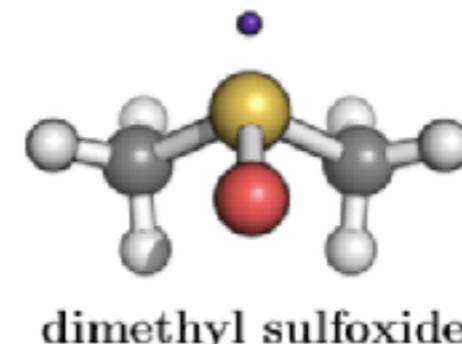
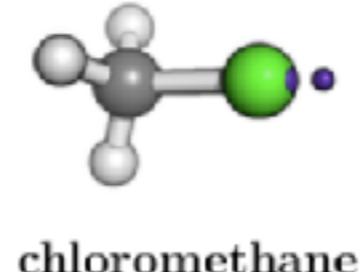
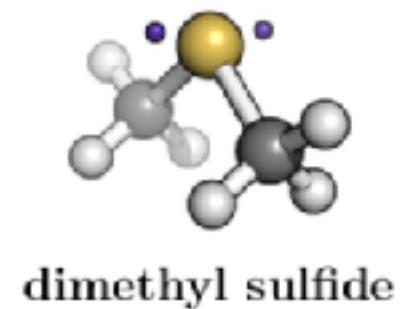
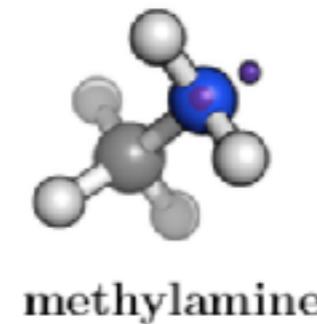
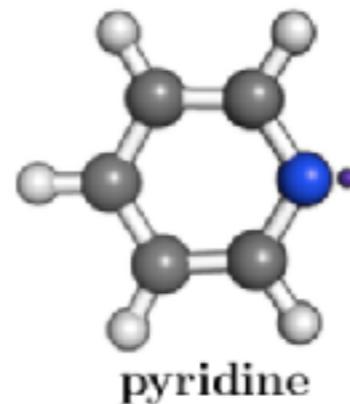


Off-Centre Charges

We derive the positions and charges of off-centre virtual sites, by:

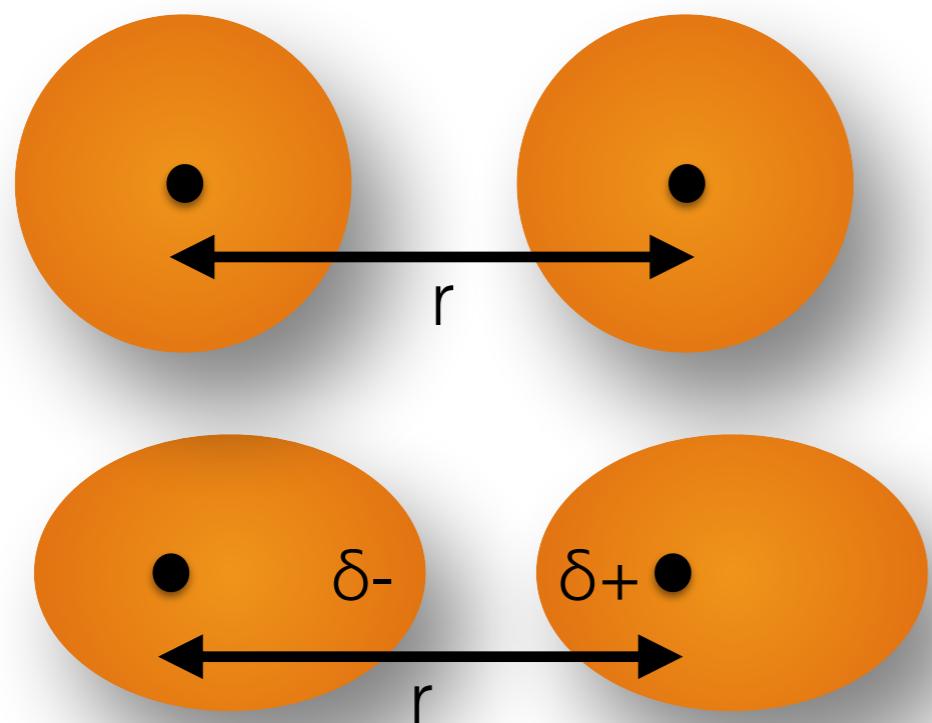
- 1) minimising the difference between the MM electrostatic potential (ESP) and the ESP of the partitioned electron density;

- 2) maintaining the symmetry of the atom's bonding environment.



van der Waals Interactions

C_6 coefficients derived using Tkatchenko-Scheffler approach:

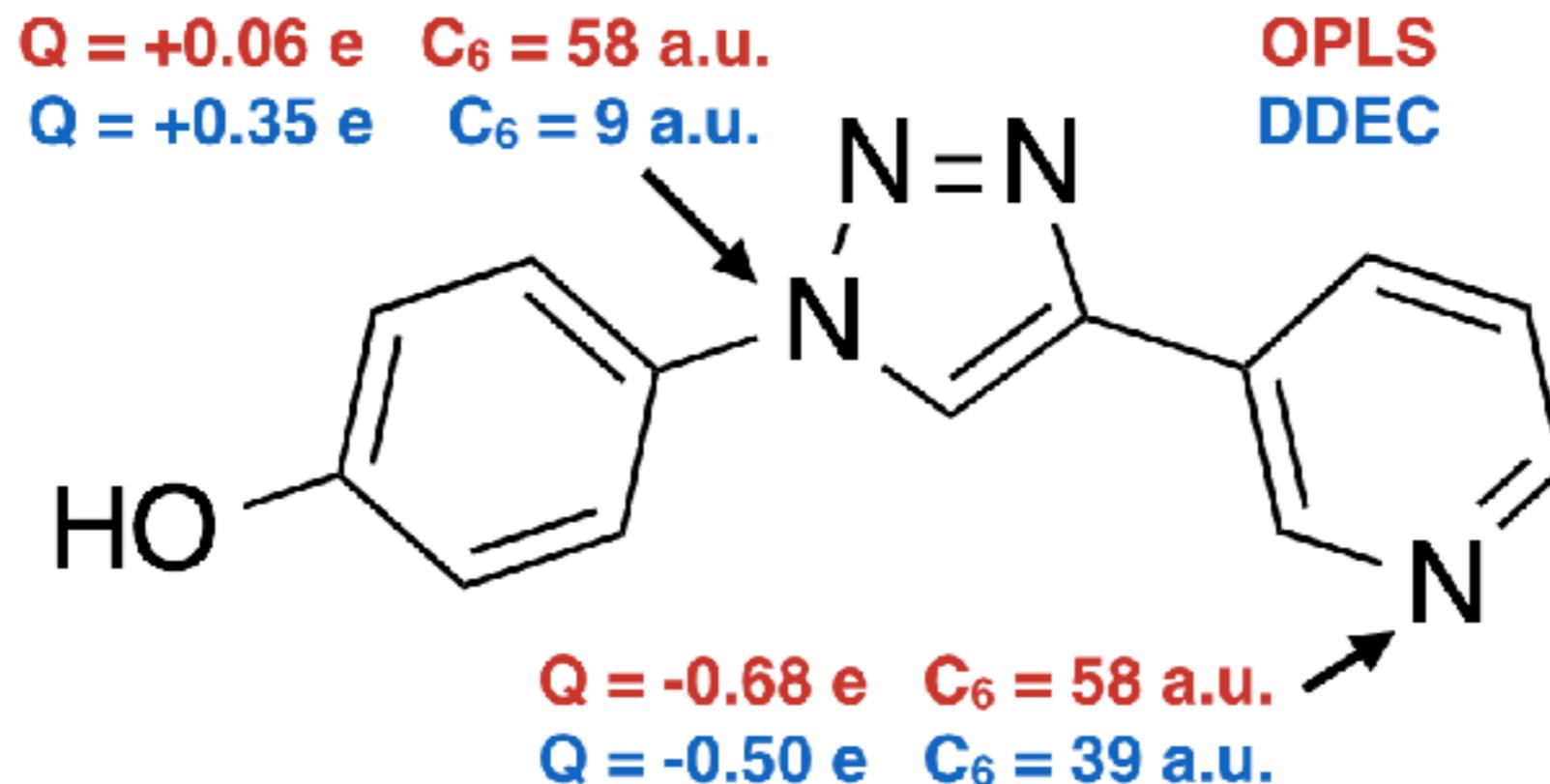


$$V_A = \int r^3 n_A(\mathbf{r}) d^3\mathbf{r}$$
$$C_6 = \left(\frac{V_A}{V_A^0} \right)^2 C_6^0$$

$$E^{MM} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right)$$

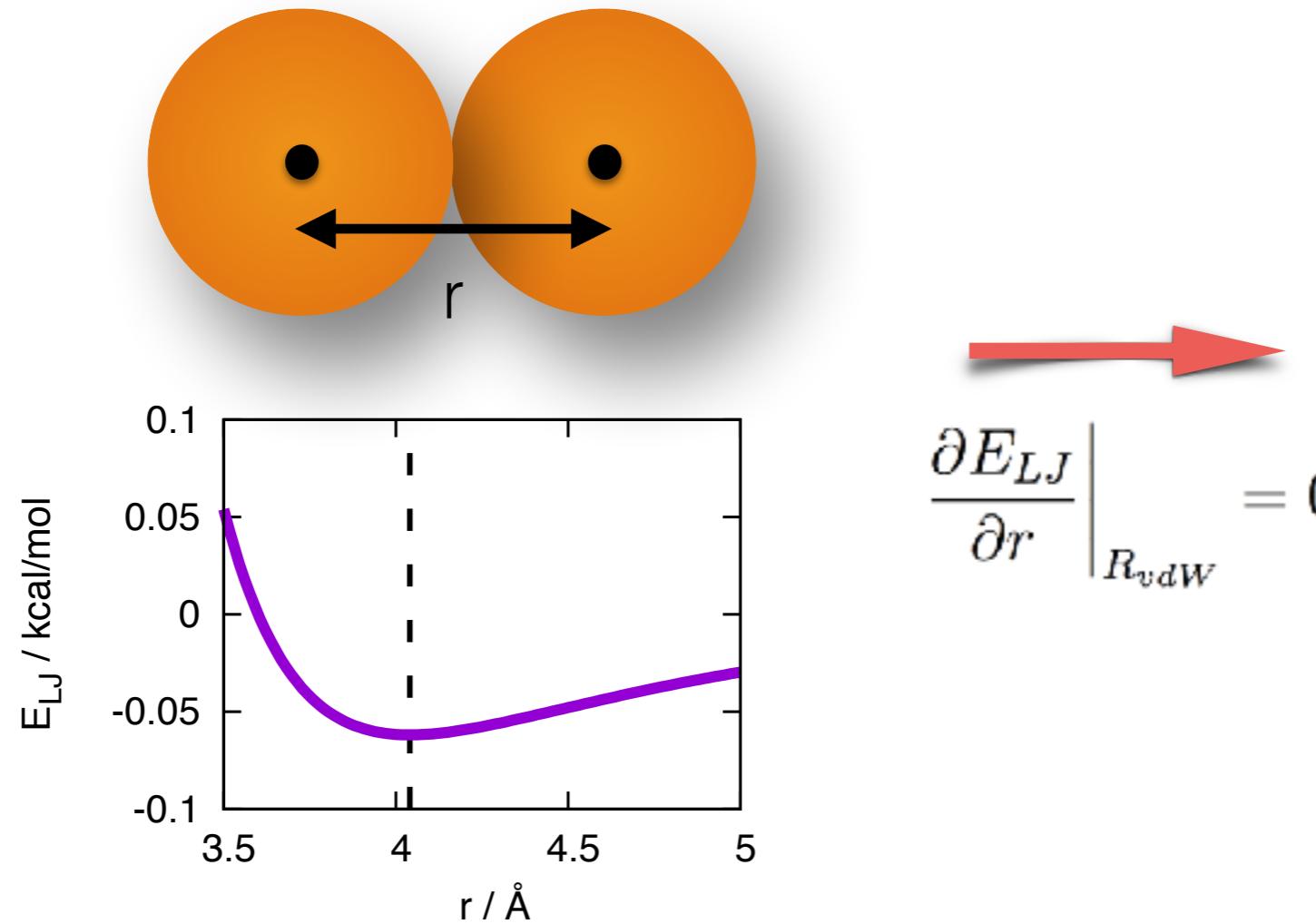
van der Waals Interactions

C₆ coefficients for a small drug-like molecule:



Exchange-Repulsion

C_{12} coefficients derived from vdW radius of the atom:



$$\frac{\partial E_{LJ}}{\partial r} \Big|_{R_{vdW}} = 0$$

Exchange-Repulsion

$$C_{12} = \frac{1}{2} C_6 (2R_{vdW})^6$$

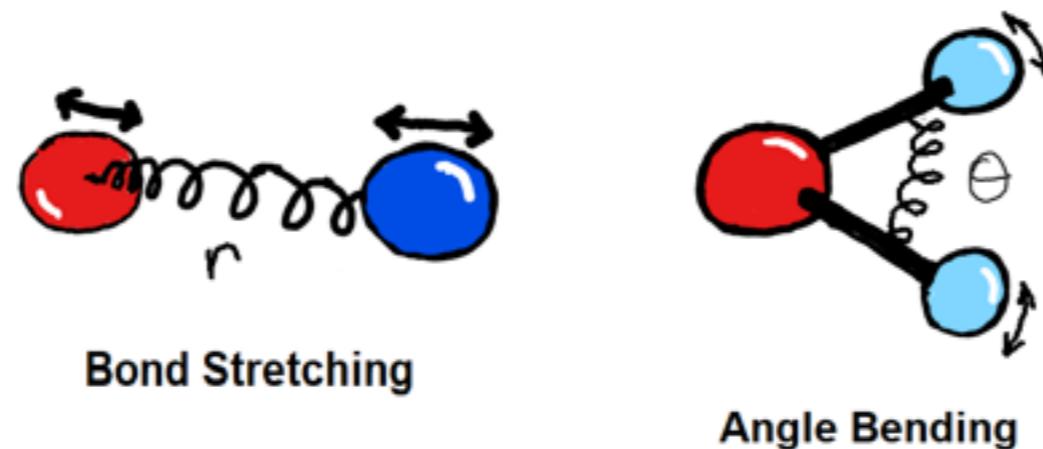
$$R_{vdW} = \left(\frac{V_A}{V_A^0} \right)^{1/3} R_{vdW}^0$$

The only fitting parameters used in the non-bonded part of the force field.
One per element, trivial to fit against liquid densities.

Bonded parameters

What about the bonded parameters?

$$E_{\text{MM}} = \sum_{\text{bonds}} k_r(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} k_\phi(1+\cos(n\phi+\phi_0)) + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right)$$



Bond and angle harmonic force constants are typically fit to reproduce QM or experimental vibrational frequencies.

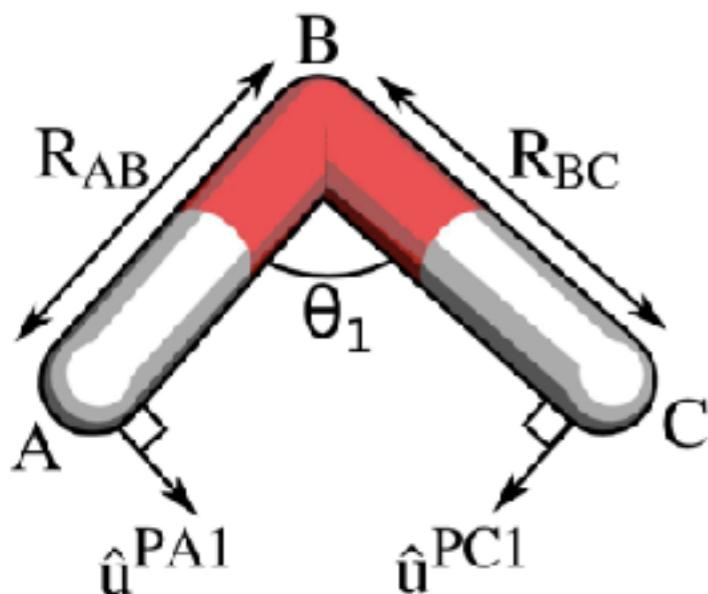
Bonded parameters

What about the bonded parameters?

$$E_{MM} = \sum_{\text{bonds}} k_r(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} k_\phi(1+\cos(n\phi+\phi_0)) + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right)$$

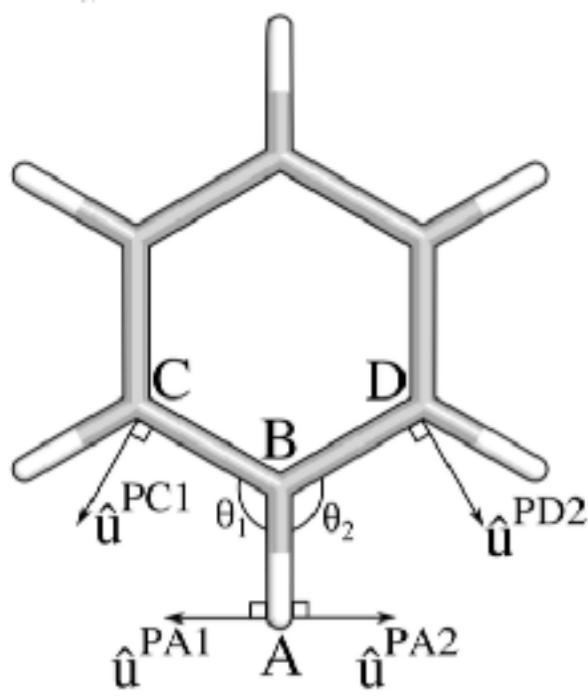
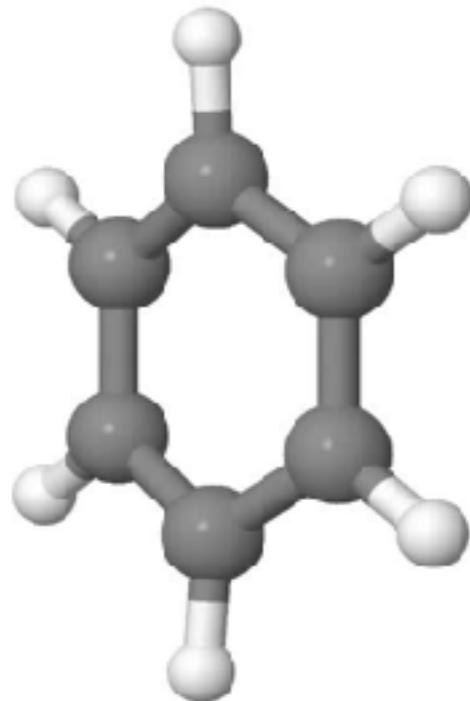
Seminario method

- 1) Compute $3N \times 3N$ QM Hessian matrix \mathbf{k} ;
- 2) Extract 3×3 partial matrix involving atoms A and B;
- 3) Project Hessian eigenvectors onto directions parallel and perpendicular to AB;
- 4) Compute k_r, k_θ



$$[\mathbf{k}_{AB}] = - \begin{vmatrix} \frac{\partial^2 E}{\partial x_A \partial x_B} & \frac{\partial^2 E}{\partial x_A \partial y_B} & \frac{\partial^2 E}{\partial x_A \partial z_B} \\ \frac{\partial^2 E}{\partial y_A \partial x_B} & \frac{\partial^2 E}{\partial y_A \partial y_B} & \frac{\partial^2 E}{\partial y_A \partial z_B} \\ \frac{\partial^2 E}{\partial z_A \partial x_B} & \frac{\partial^2 E}{\partial z_A \partial y_B} & \frac{\partial^2 E}{\partial z_A \partial z_B} \end{vmatrix}$$

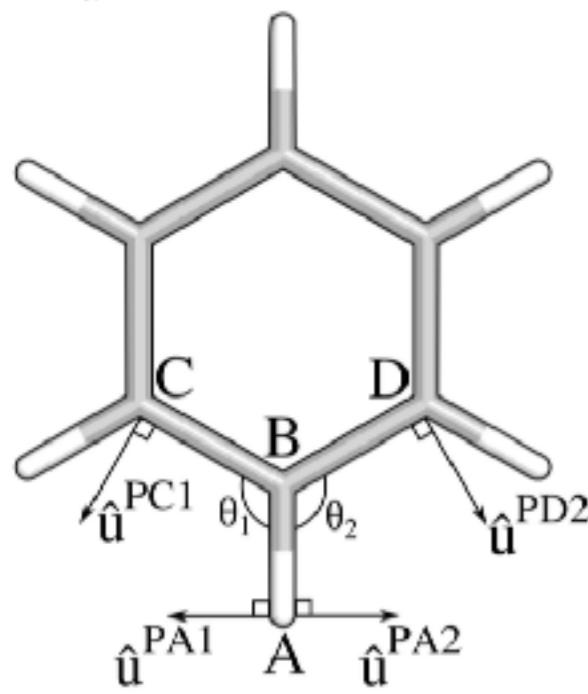
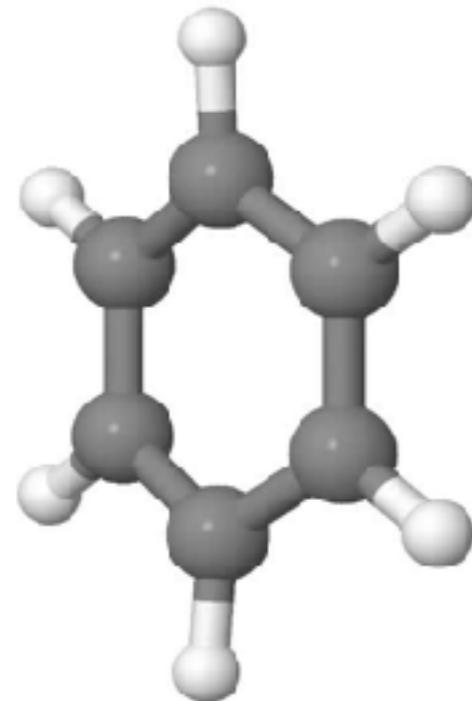
Bonded parameters



Average error in MM vibrational frequencies of
70 molecules relative to QM (cm^{-1}):

	OPLS	Seminario
Small molecules	60	120
Heterocycles	83	132
Dipeptides	47	104

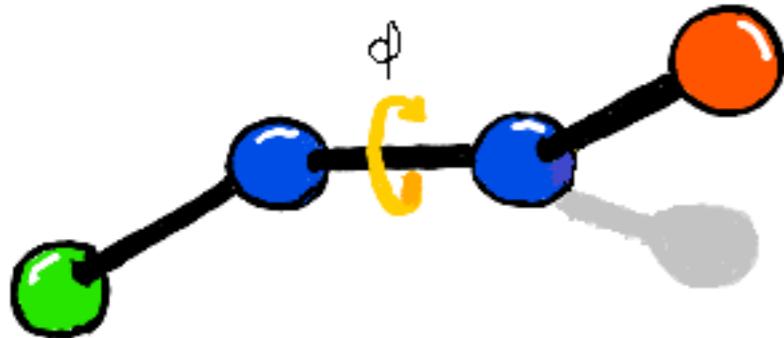
Bonded parameters



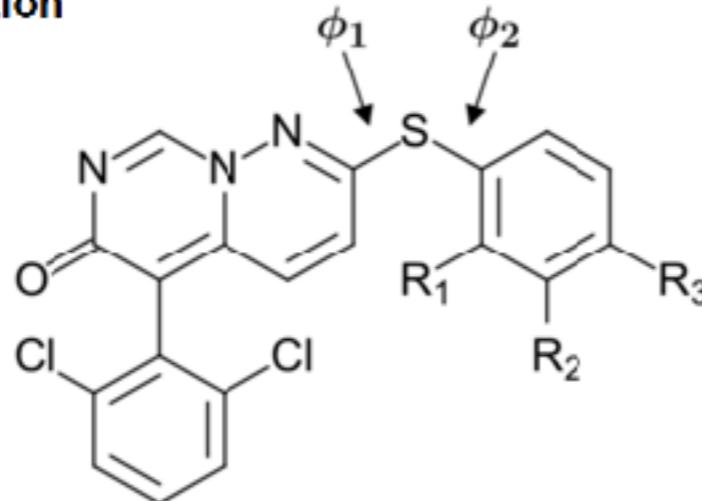
Average error in MM vibrational frequencies of
70 molecules relative to QM (cm^{-1}):

	OPLS	Seminario	Modified Seminario
Small molecules	60	120	52
Heterocycles	83	132	53
Dipeptides	47	104	40

Torsions

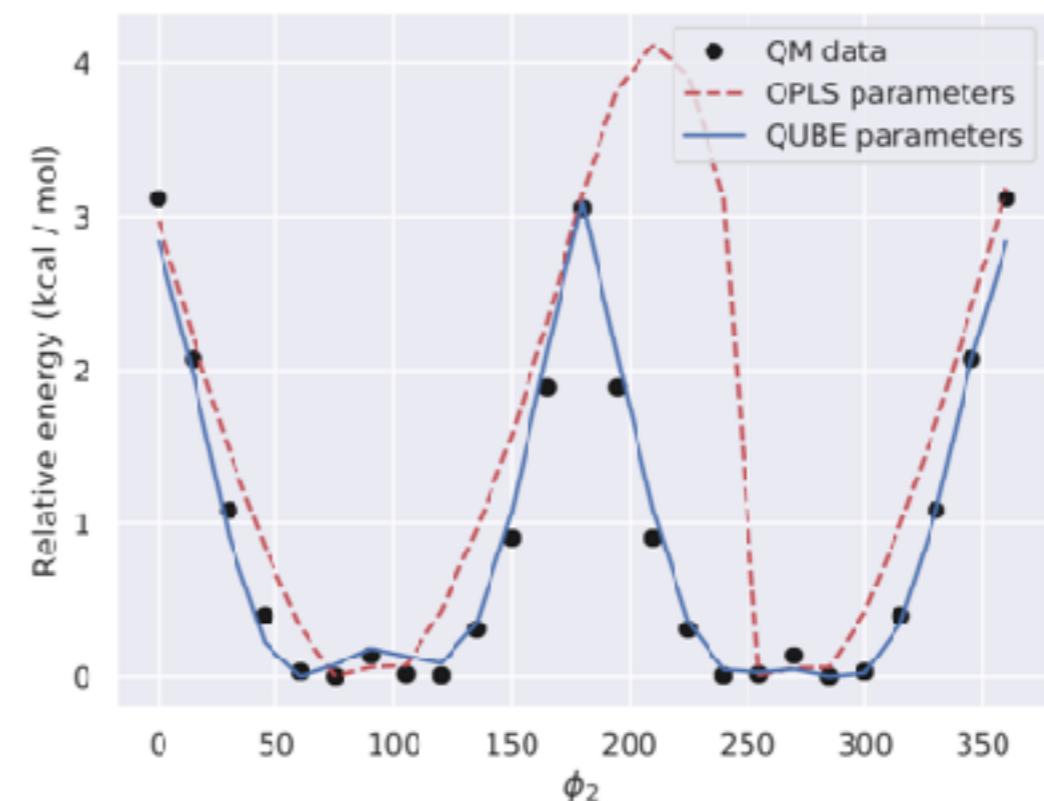
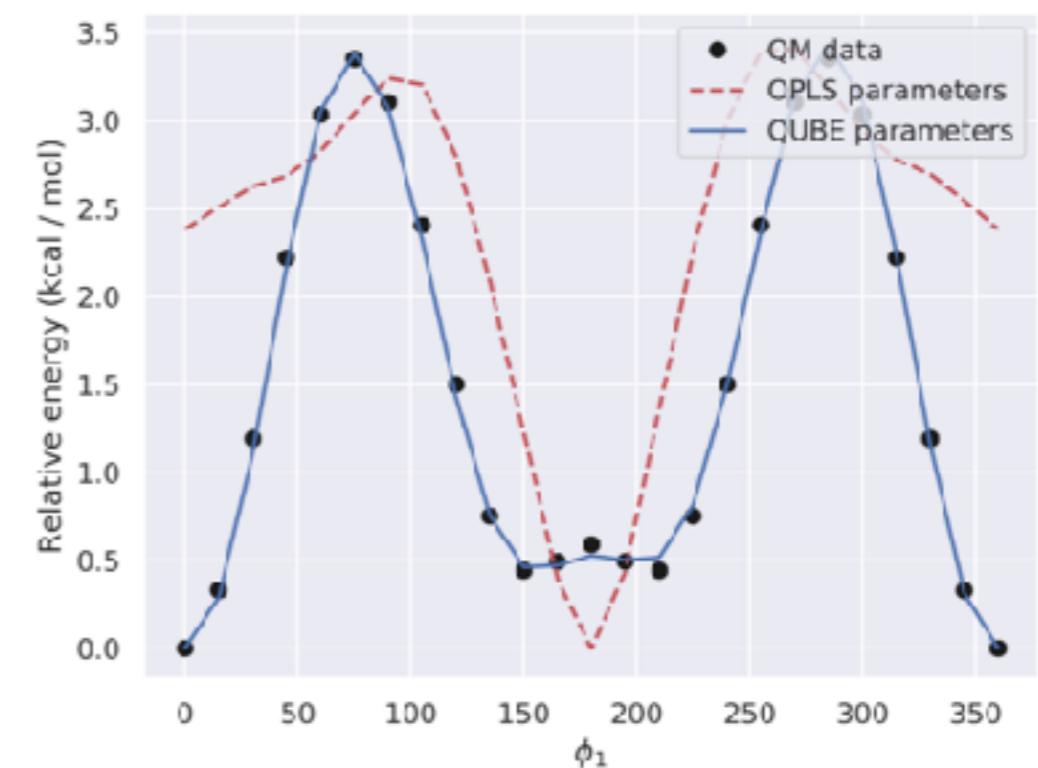


Dihedral Angle Rotation

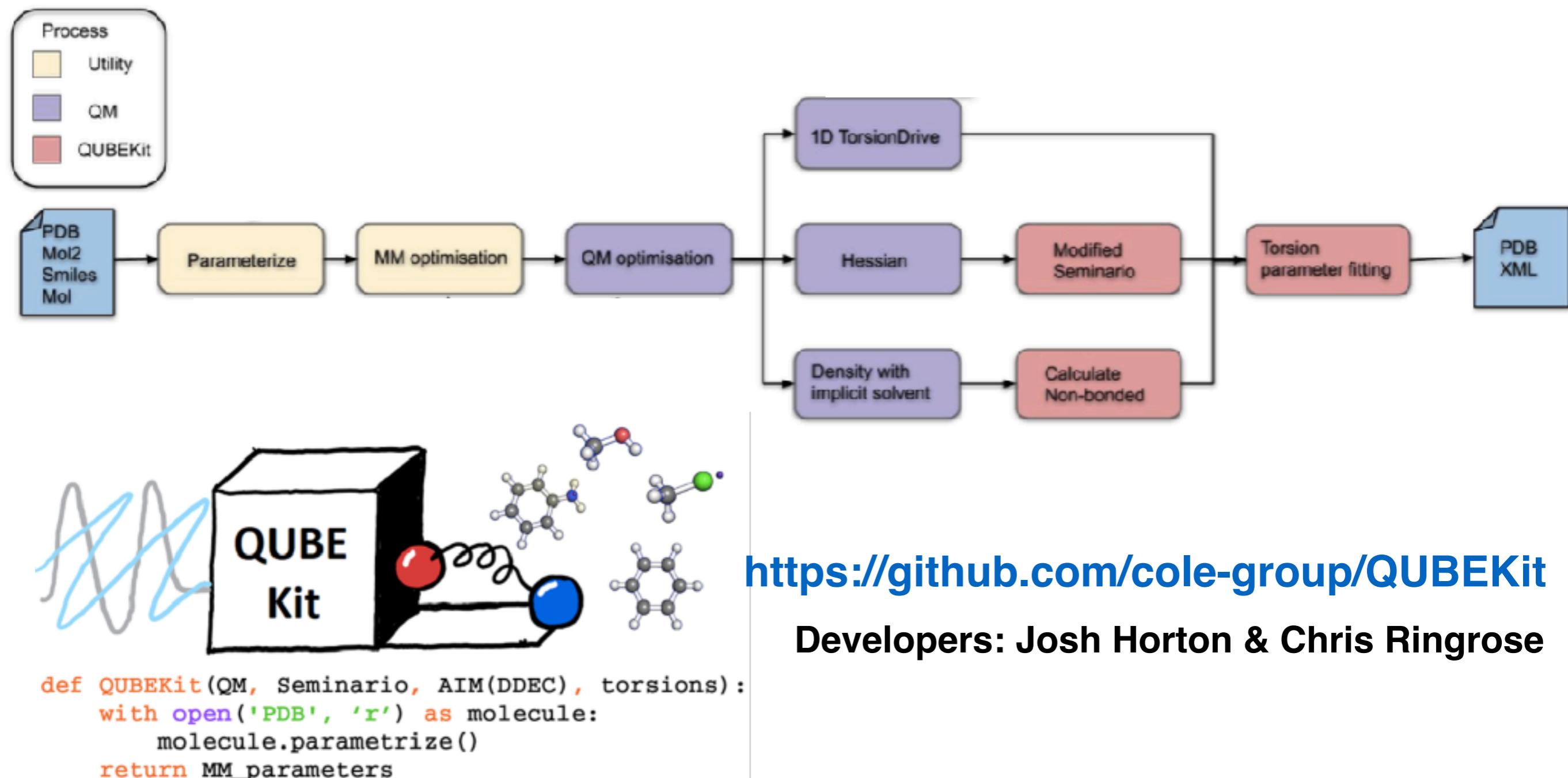


Torsion parameter derivation via fitting to QM 1D potential energy surfaces.

Originally via BOSS code, but now facilitated by an interface with torsiondrive (Lee-Ping Wang).



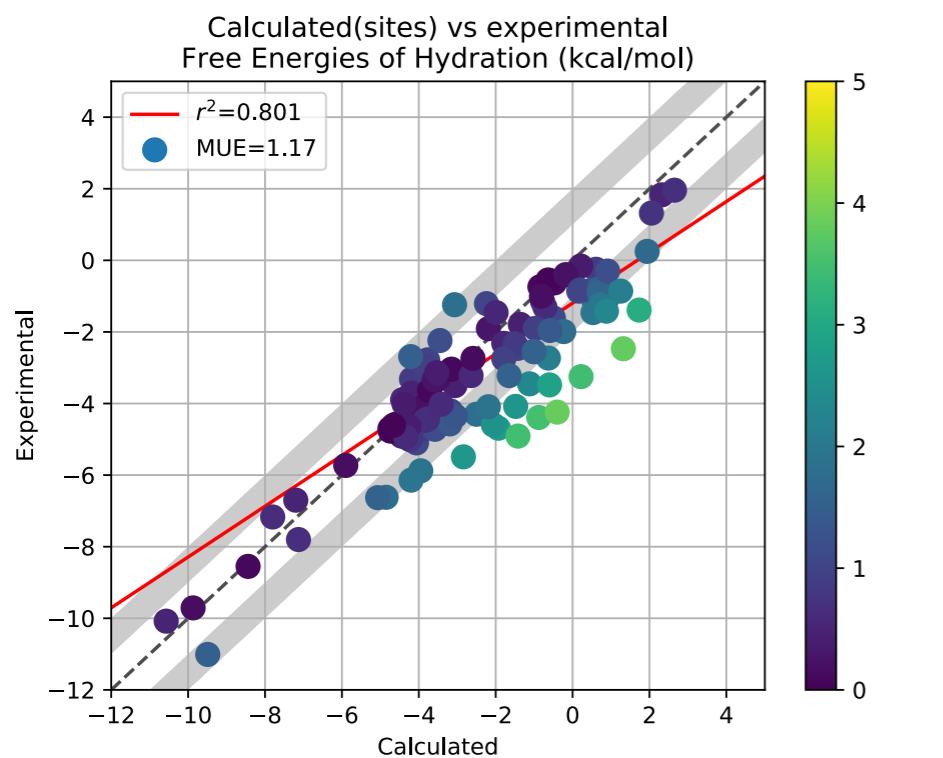
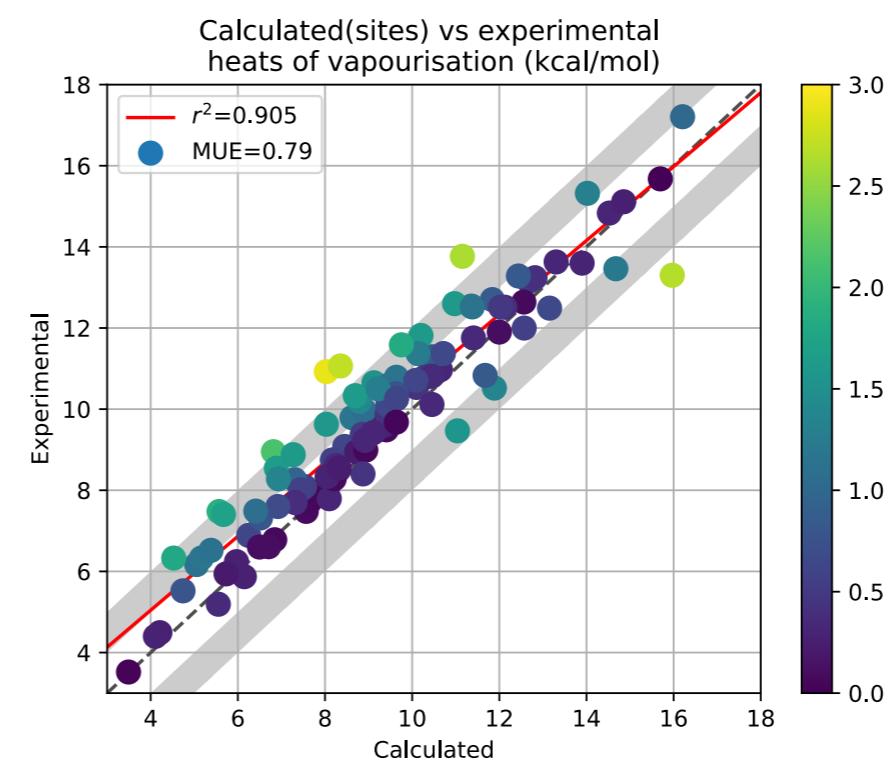
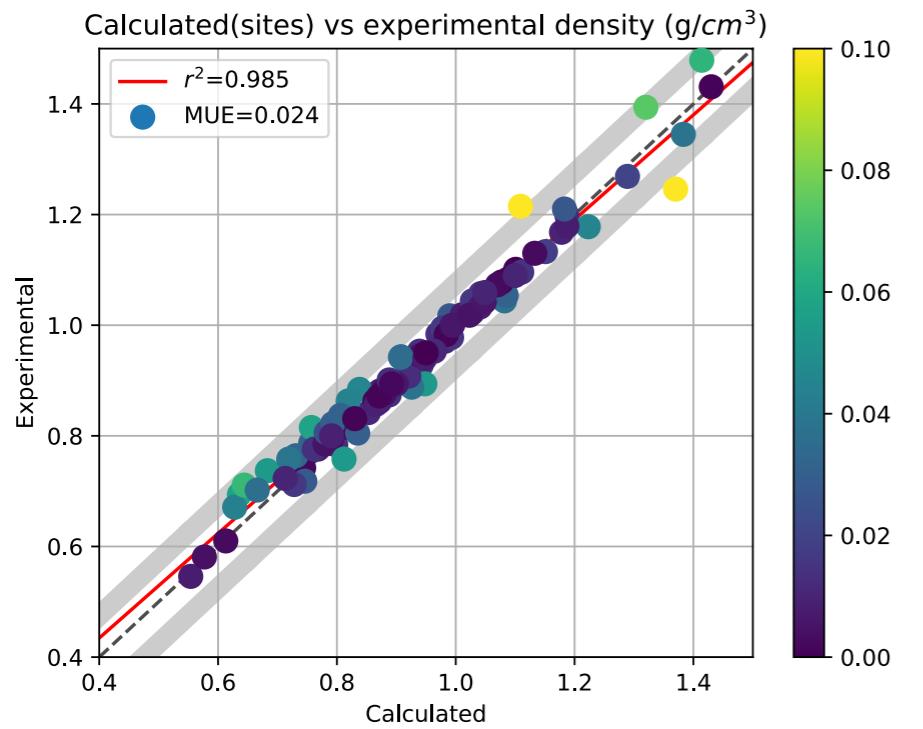
Automation



<https://github.com/cole-group/QUBEKit>

Developers: Josh Horton & Chris Ringrose

Liquid Data



Full force field implemented in OpenMM,
and used to compute liquid properties.
Competitive with standard force fields that
have been fit to these data.

Horton JT, Allen AEA, Dodd L, Cole DJ.
J Chem Inf Model 2019, **59**(4), 1366-1381.

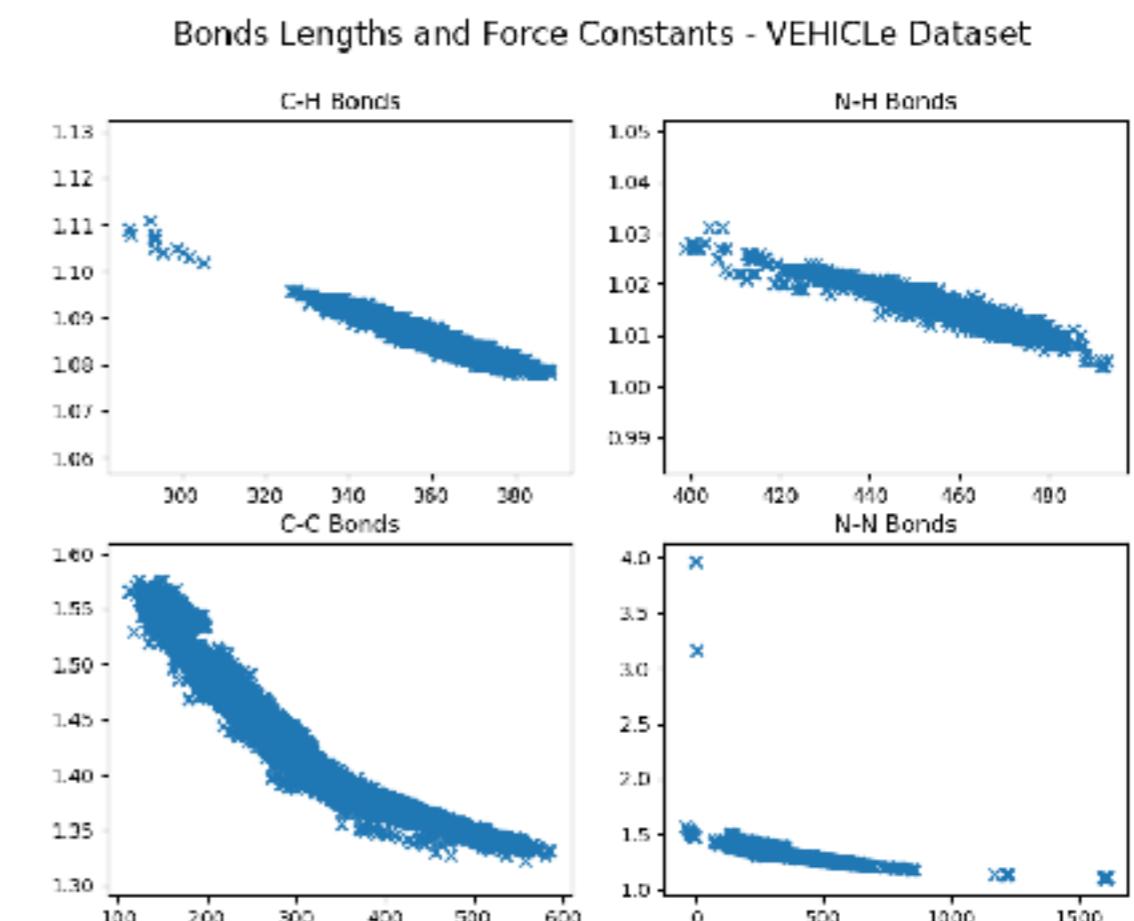
QCArchive

QUBEKit interfaces with a lot of open source code (including QM, MM, cheminformatics, ...). This is made significantly easier by an interface with MolSSI's **QCEngine**.

The **QCArchive** project (sponsored by the Open Force Field Initiative, qcarchive.molssi.org) organises and computes a database of millions of QM results, and user friendly front ends to explore the data.

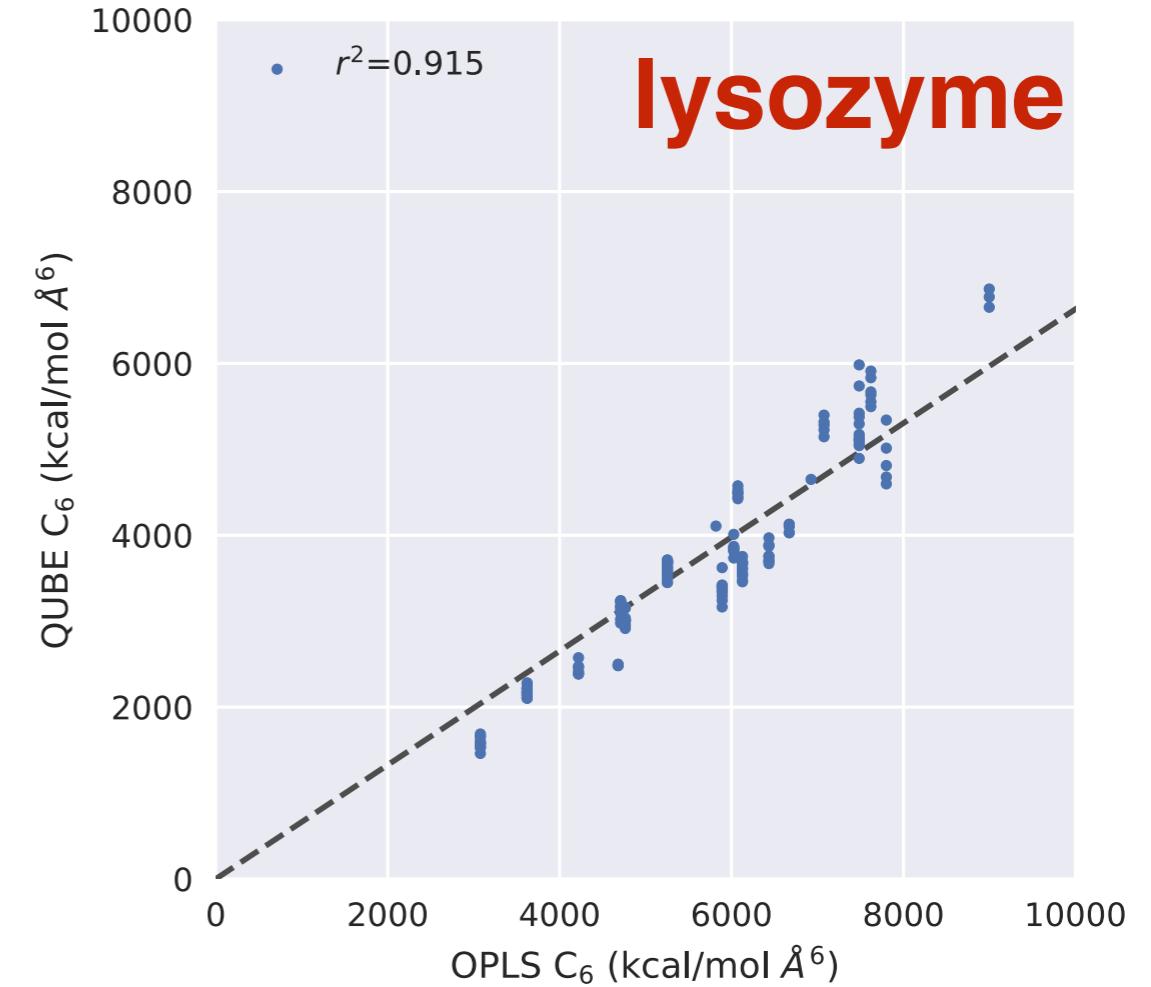
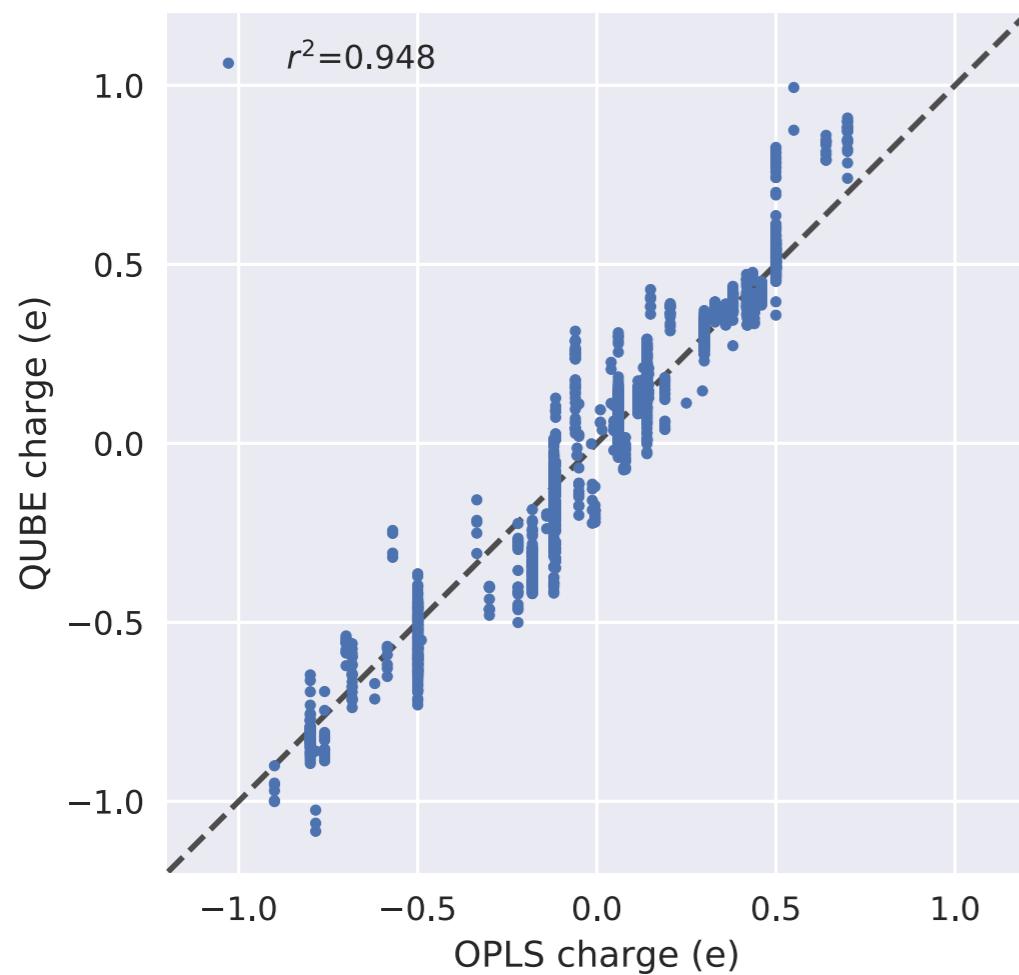
Example use case (work in progress):

Analyse Hessian matrices and optimised geometries of ~20K small molecules stored on QCArchive using QUBEKit.

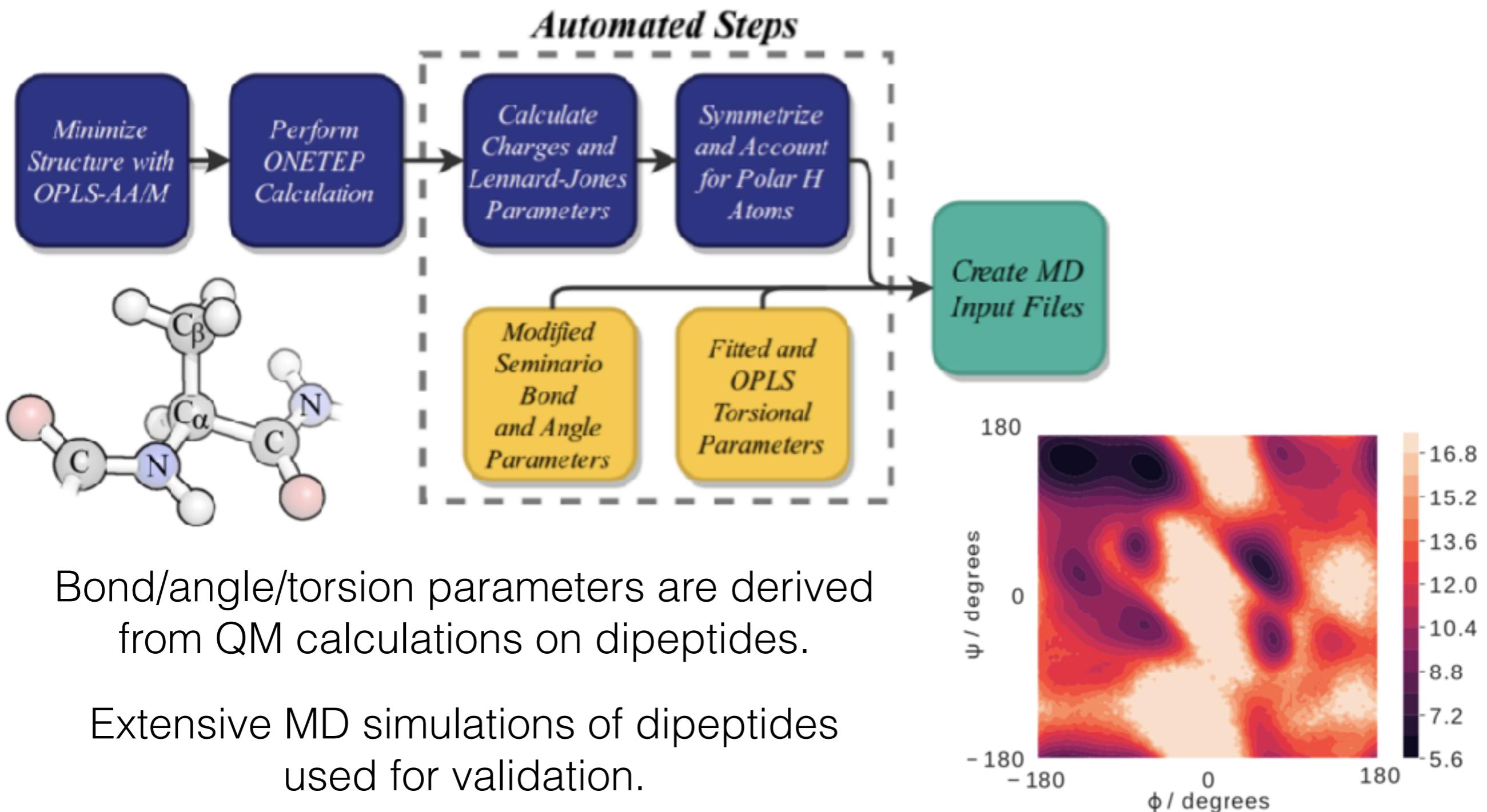


QUBE Protein Force Field

We have implemented all of the non-bonded parameterisation methods in the ONETEP linear-scaling DFT code (www.onetep.org).

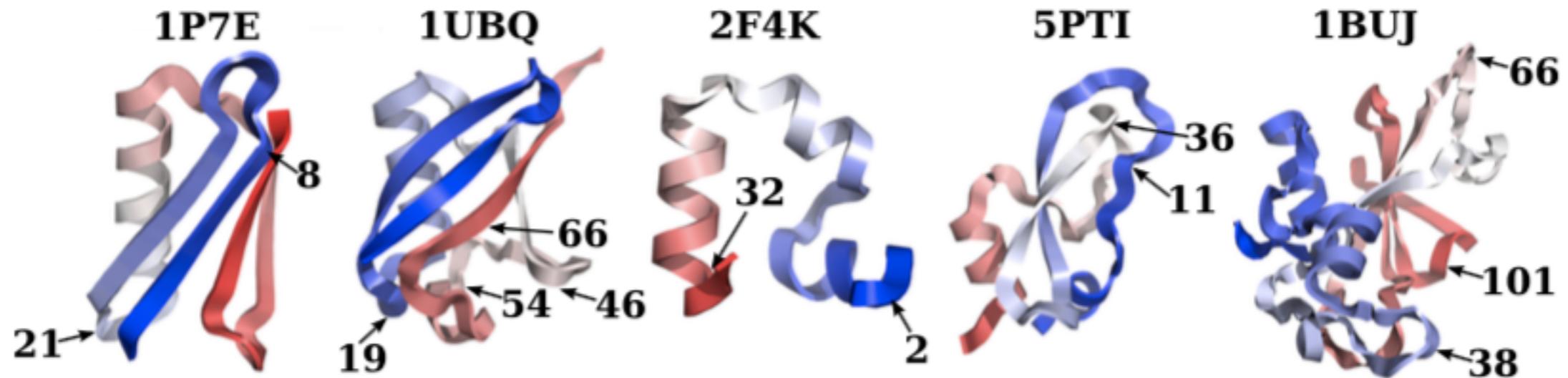
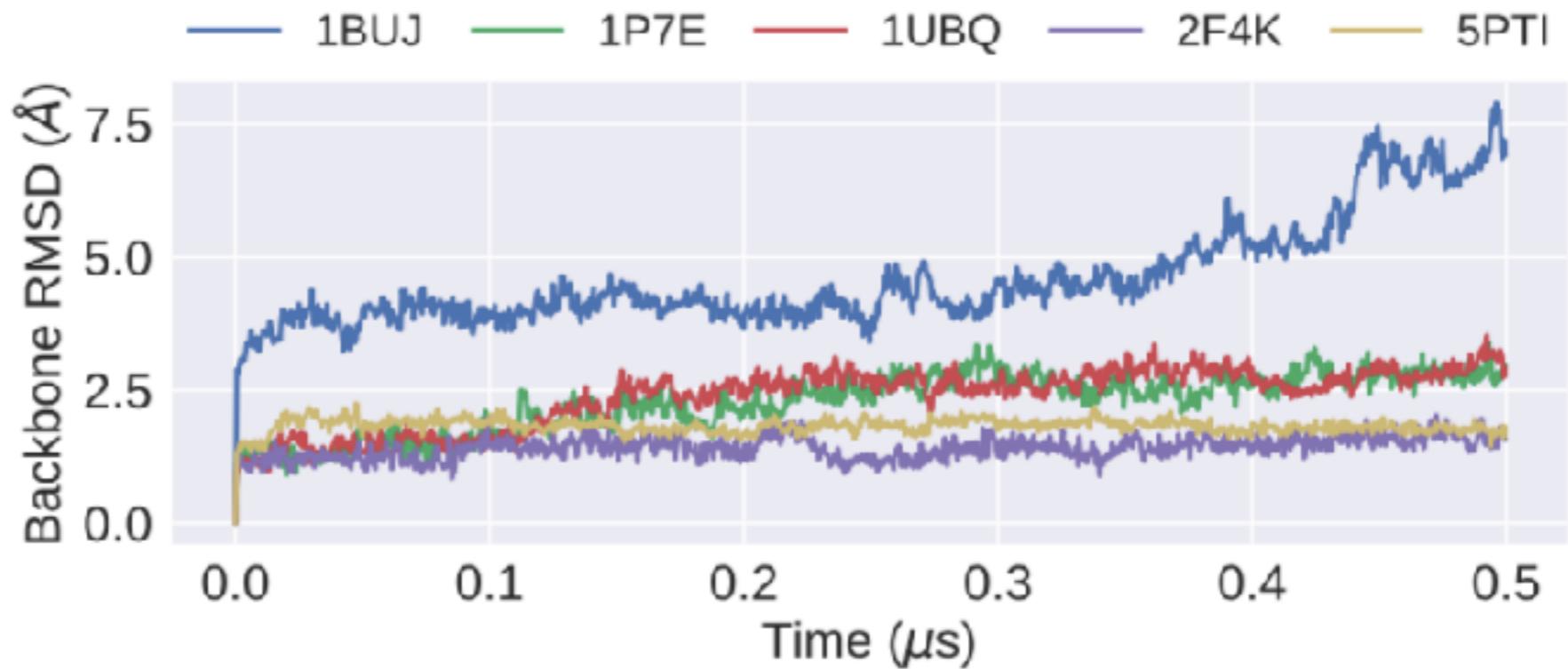


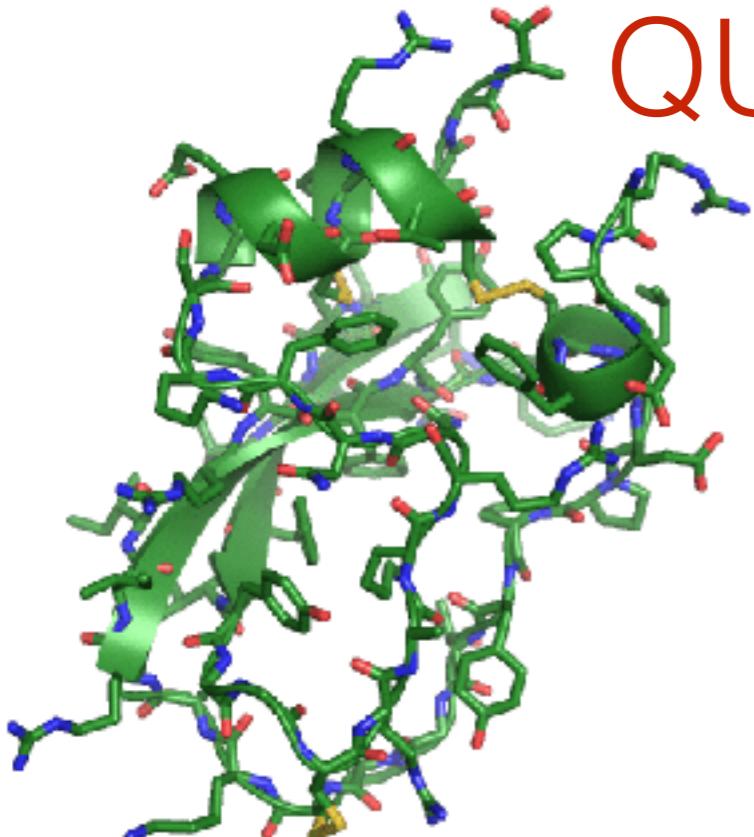
QUBE Protein Force Field



QUBE Protein Force Field

RMSD relative to
crystal structure
($3 \times 0.5\mu\text{s}$ MD):





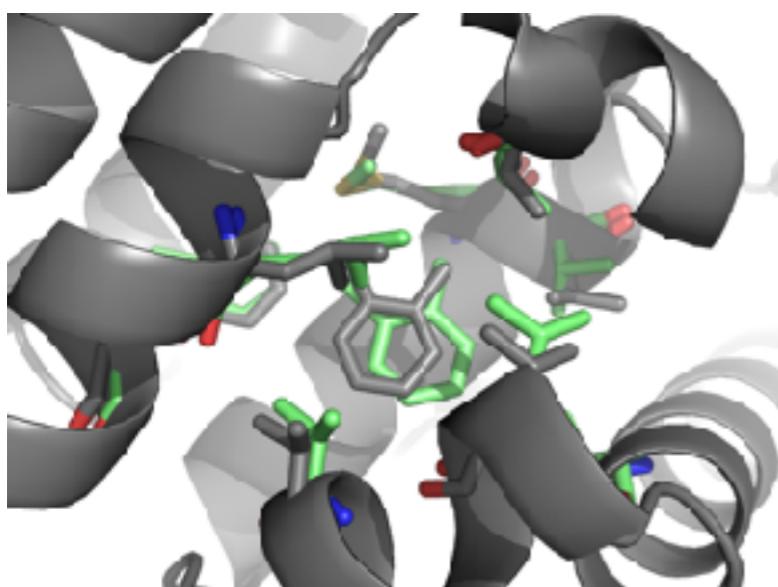
QUBE Protein Force Field

MD simulations can be benchmarked against experimental NMR J-couplings which describe conformational fluctuations.

	OPLS-AA	OPLS-AA/M	QUBE
Dipeptides	1.0	0.4	0.4
Ala ₅	2.3	1.2	0.9
1UBQ	1.8	1.1	1.5
1P7E	1.5	0.9	1.1

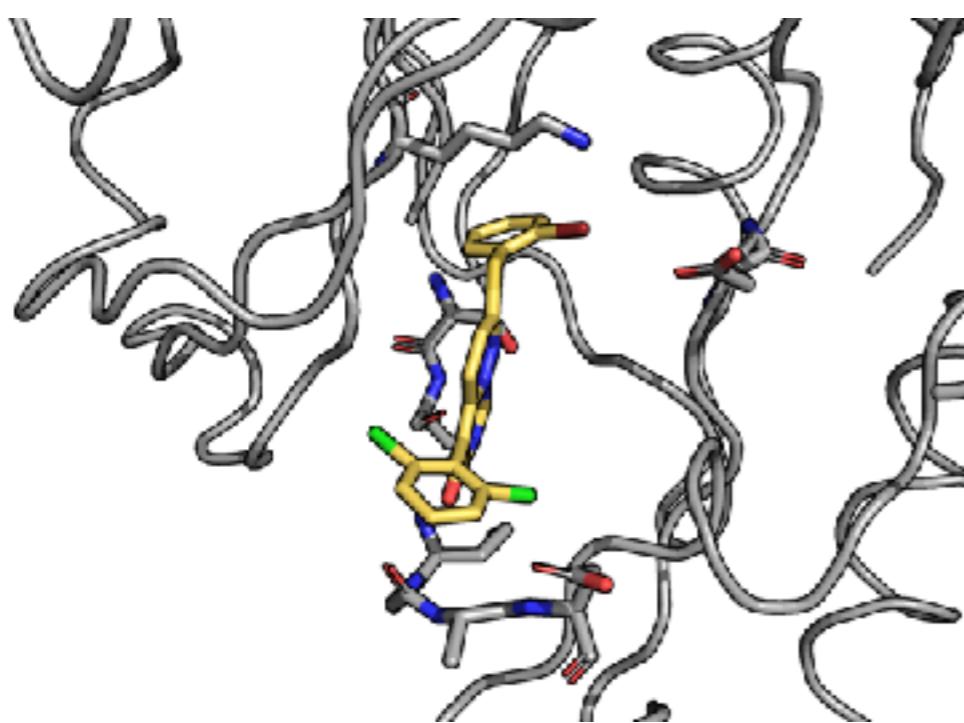
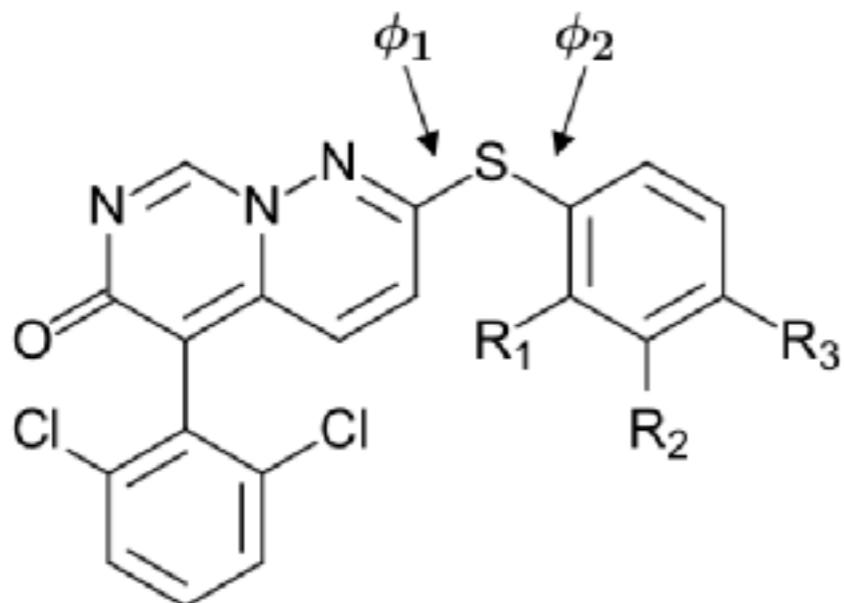
Free Energy Perturbation

Absolute binding free energies of small molecules to the L99A mutant of T4 lysozyme (kcal/mol):



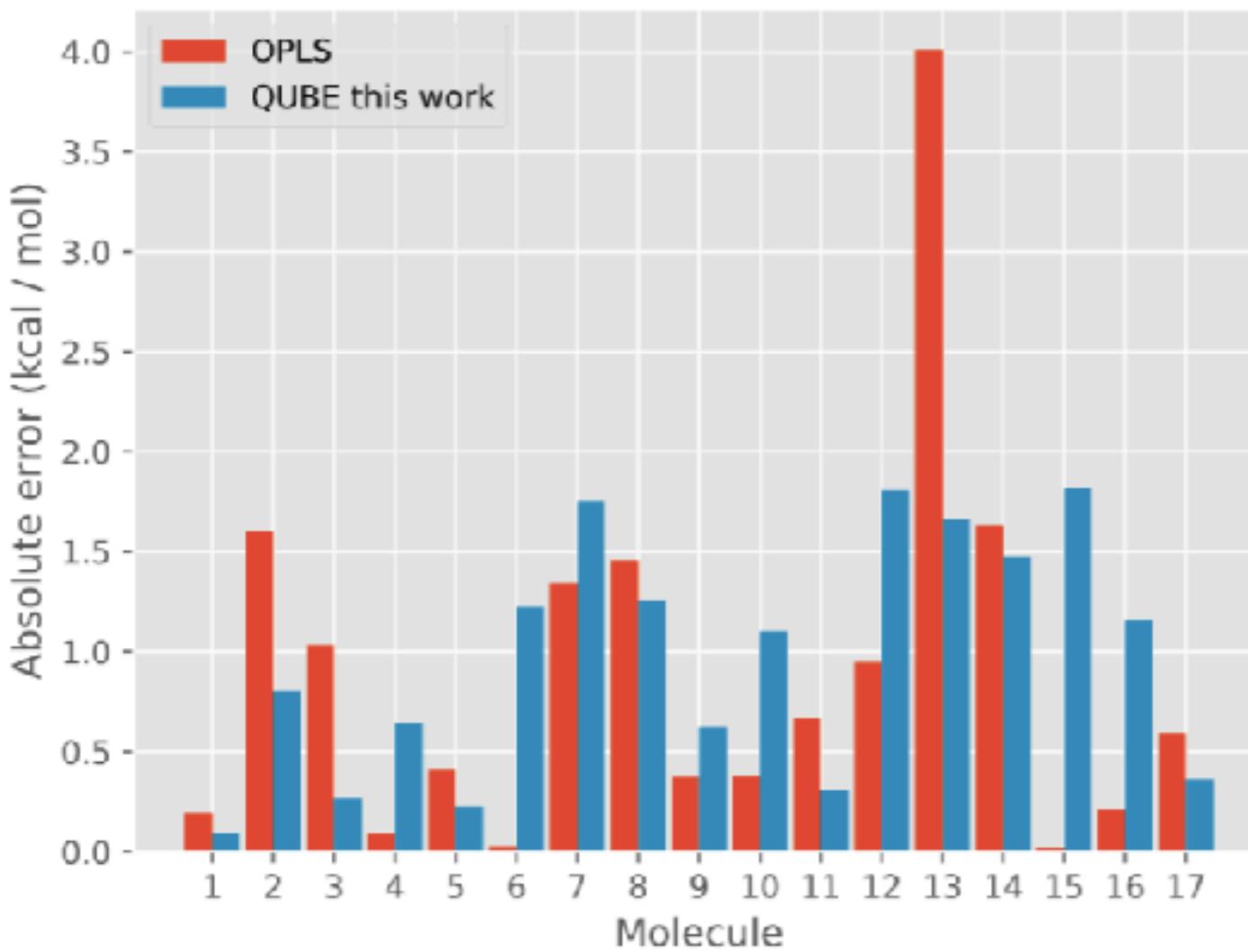
	OPLS	QUBE	Experiment
benzene	-7.7	-6.0	-5.2
p-xylene	-5.0	-4.4	-4.7
o-xylene	-2.9	-5.0	-4.6
benzofuran	-7.2	-7.0	-5.5
indole	-4.4	-3.8	-4.9
indene	-5.9	-4.0	-5.1
MUE	1.3	0.9	—

p38 Kinase FEP

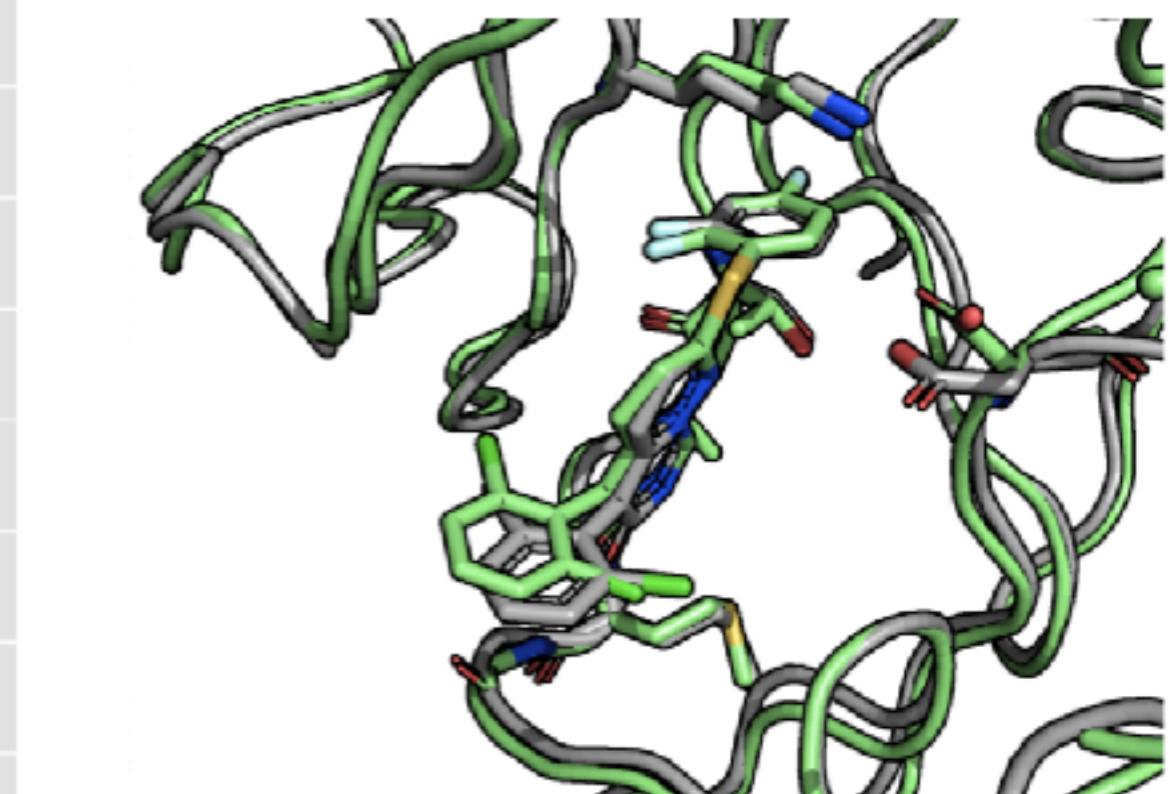


Compound	R1	R2	R3	pIC ₅₀
1	H	H	H	6.602
2	H	H	F	7.000
3	H	H	CH ₃	5.854
4	H	Cl	Cl	6.097
5	H	CH ₃	H	5.854
6	H	CH ₃	CH ₃	5.721
7	H	F	H	6.347
8	CH ₃	H	H	6.699
9	H	Cl	F	6.301
10	H	Cl	H	6.553
11	CH ₃	H	Cl	6.745
12	Br	H	H	6.602
13	CH ₃	H	CH ₃	6.577
14	OH	H	H	6.444
15	NH ₂	H	F	6.658
16	Cl	H	F	7.444
17	F	F	F	8.046
18	F	H	H	N/A

p38 Kinase FEP

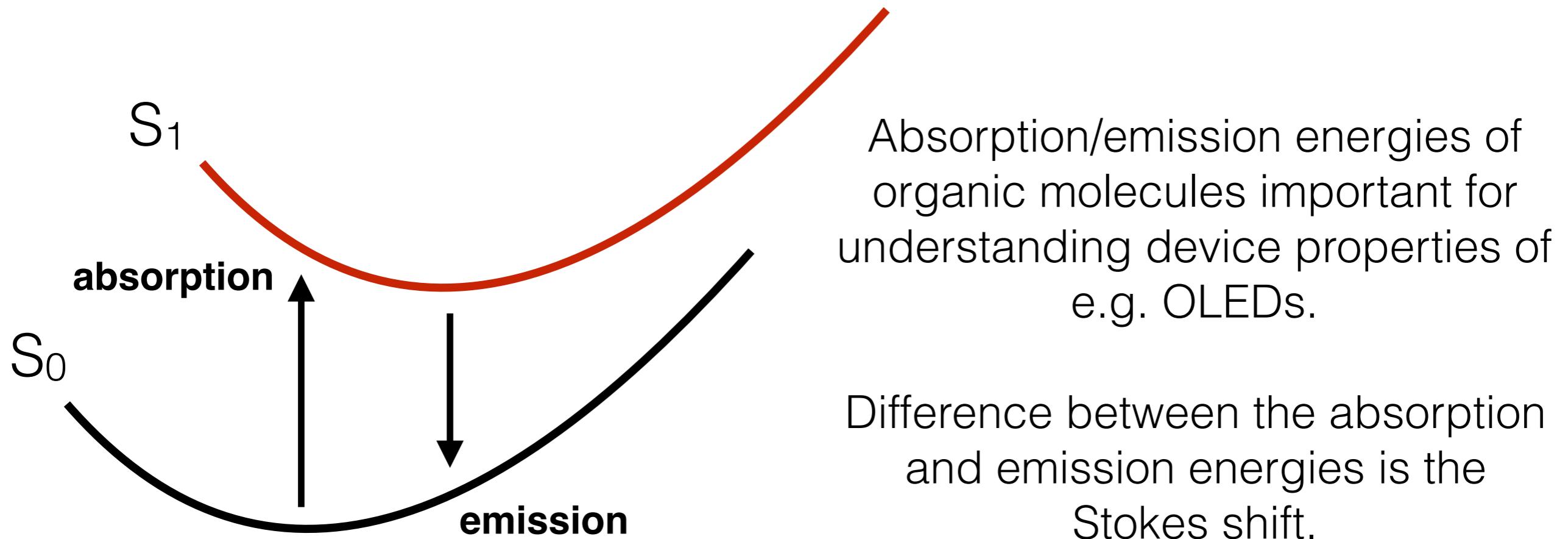


MUE(OPLS) = **0.88** kcal/mol
MUE(QUBE) = **0.98** kcal/mol



Good structural/energetic agreement for 1st generation MM force field.

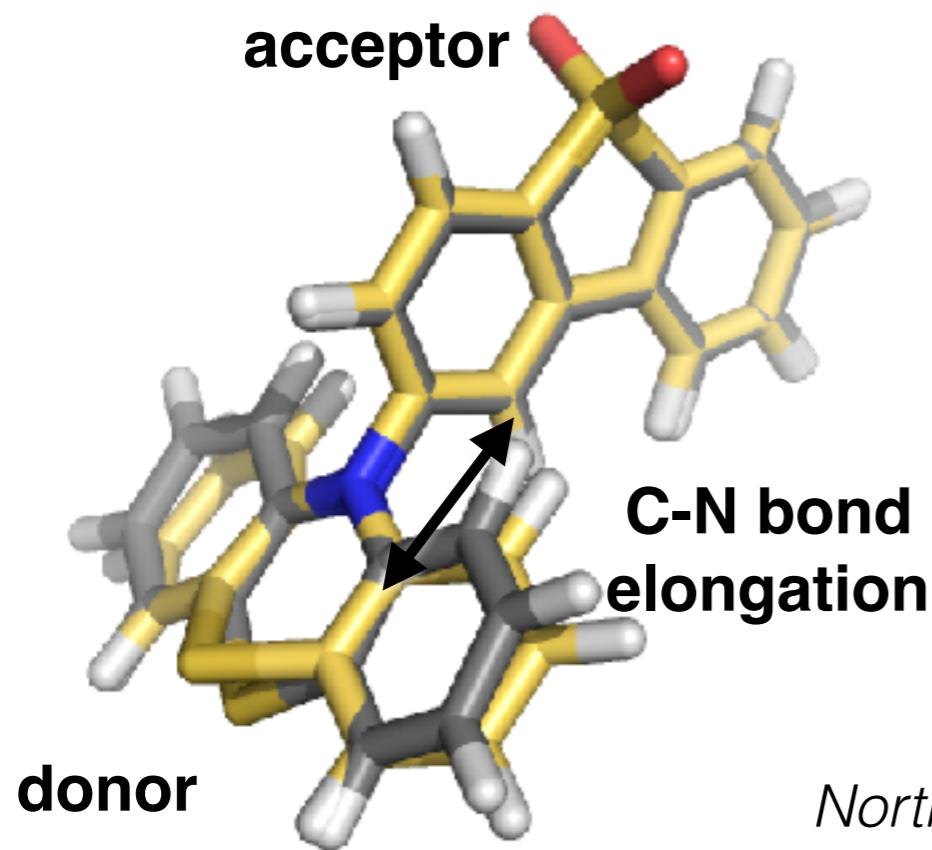
Photochemistry



Structural rearrangement of the emitter in the excited state and electrostatic interactions with its environment can influence the Stokes shift.

MM force fields are often used to model the dynamics of these organic molecules.

Photochemistry



Studied prototypical donor-acceptor molecule (PTZ-DBTO2).

S₀ ground state structure shown in yellow.

S₁ lowest singlet charge transfer state structure in grey.

Northey, Stacey, Penfold, J. Mater. Chem. C, 2017, 5, 11001-11009

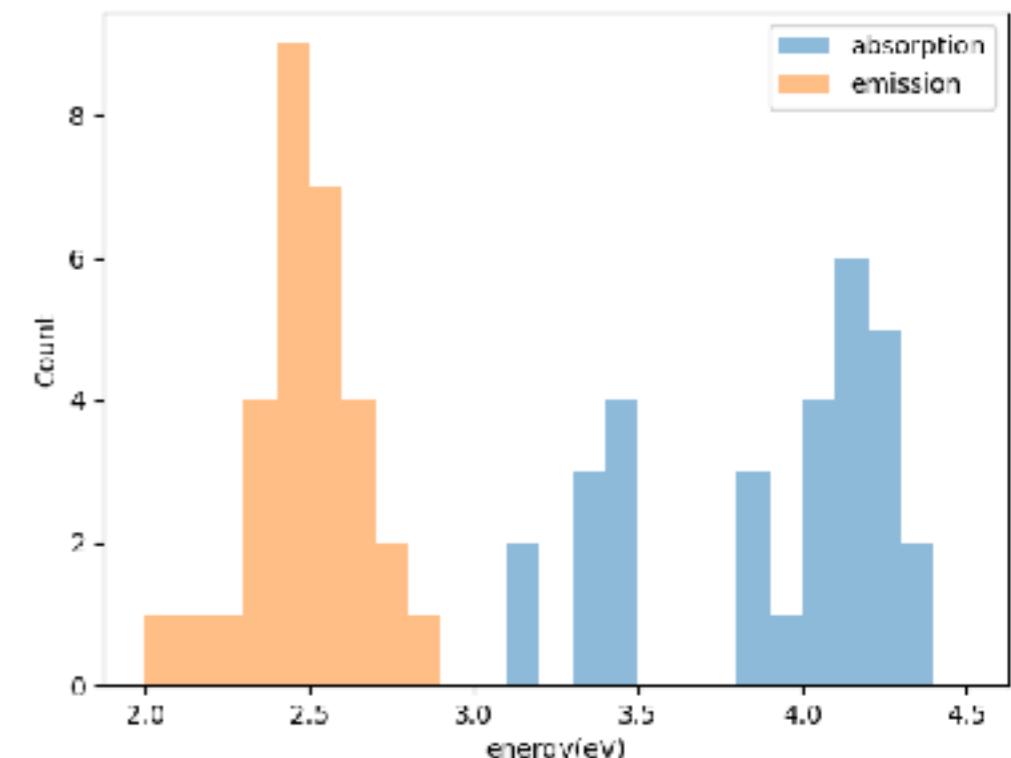
Separate QUBE force fields generated for S₀ and S₁ states.
Standard force fields use same bonded parameters, independent of electronic structure.

Photochemistry

Run MD of PTZ-DBTO2 in two different hosts
(liquid and solid state).

Extract snapshots and compute average absorption/emission energies using standard QM/MM approach.

QUBE appears to capture both structural and electrostatic contributions to Stokes shift.



Host	Standard	QUBE	Experiment
Toluene	absorption	3.4	3.7
	emission	2.8	2.1
PYD2	absorption	3.3	—
	emission	2.5	—

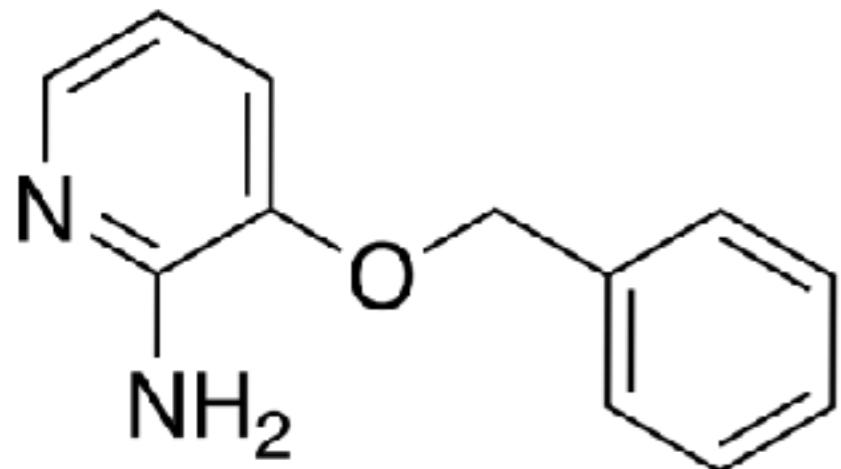
Machine Learning

$$E_{MM} = \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi (1 + \cos(n\phi + \phi_0)) + \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \left(\begin{array}{cc} C_{12}^{ij} & C_6^{ij} \\ r_{ij}^{12} & r_{ij}^6 \end{array} \right)$$

Gaussian approximation potential (GAP) applies Bayesian inference to learn the potential energy surface of a molecule using quantum mechanical training data (energies and forces).

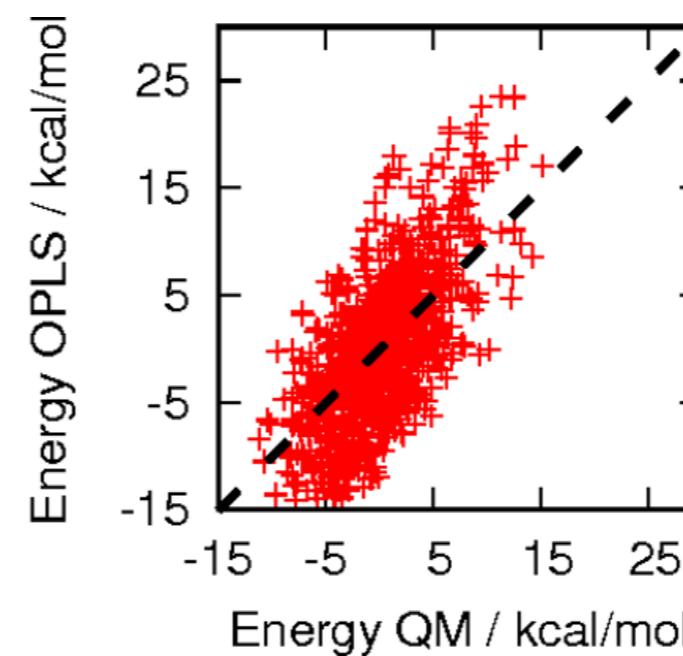
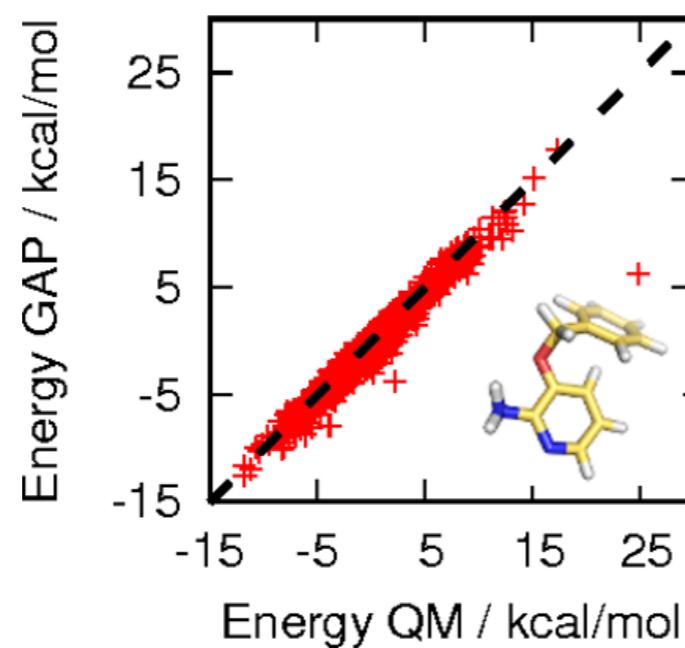


Machine Learning



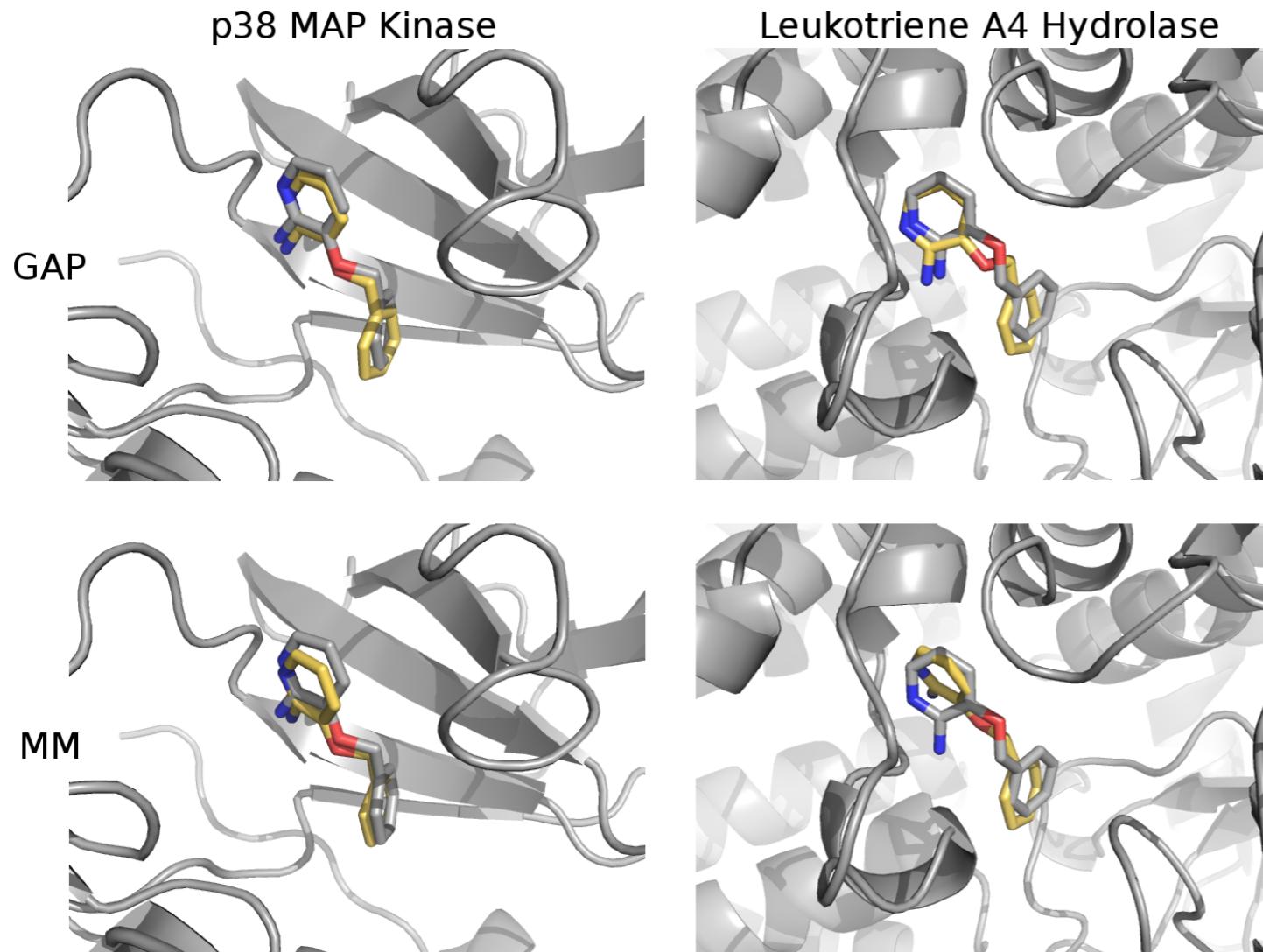
Small molecule GAP trained using ~3000 configurations.
QM energies/forces computed using MP2/6-311G(2d,p).

GAP
RMS error =
1 kcal/mol



OPLS
RMS error =
4 kcal/mol

Machine Learning



Implemented in MCPRO molecular modelling software.

Intermolecular interactions computed using OPLS force field (similar to QM/MM).

Simulated in 3 environments, bound to two proteins and in water.

Good agreement with crystal structure for kinase, but propose alternative structure for hydrolase that is consistent with x-ray data.

Summary

Have described steps towards our goal of the automated parameterisation of molecular mechanics force fields directly from QM data.

The **non-bonded** components of biomolecular force fields are obtained from QM atoms-in-molecule analysis. The **bonded** components come from the QM Hessian matrix and automated torsional scans (or machine learning).

Ongoing validation against protein dynamics and protein-ligand binding free energies.

All QUBE force field parameterisation is performed via the open source **QUBEKit** software.

<https://github.com/cole-group/QUBEKit>

Future Work

Continuing development of **QUBEKit** to improve ease-of-use and widen applicability.

Interfaces with **Open Force Field Initiative** infrastructure will give access to large quantum chemistry databases for force field derivation and workflows for force field validation.

Continue to investigate **machine learning** approaches for learning both force field parameters or entire potential energy surfaces from large quantum chemistry datasets.

There are very few empirical parameters, so the force field can easily be **re-designed** with a different functional form. E.g. we have been investigating the derivation of screened C₆, C₈ and polarisability coefficients from QM.

Potential to move into areas of **unmet need** for force fields, e.g. metals, excited state potential energy surfaces, transition states.

Acknowledgements

Newcastle University

Josh Horton

Lauren Nelson

Chris Ringrose

Vadiraj Kurdekar

Yale University

Jonah Vilseck

Leela Doddha

Michael Robertson

Israel Cabeza de Vaca

Julian Tirado-Rives

William Jorgensen



Engineering and Physical Sciences
Research Council



University of Cambridge

Alice Allen

Lupeng Yang

Edward Linscott

Louis Lee

Mike Payne

Collaborators

David Huggins (Cornell)

Ashok Venkitaraman (Cambridge)

Chris-Kriton Skylaris (Southampton)

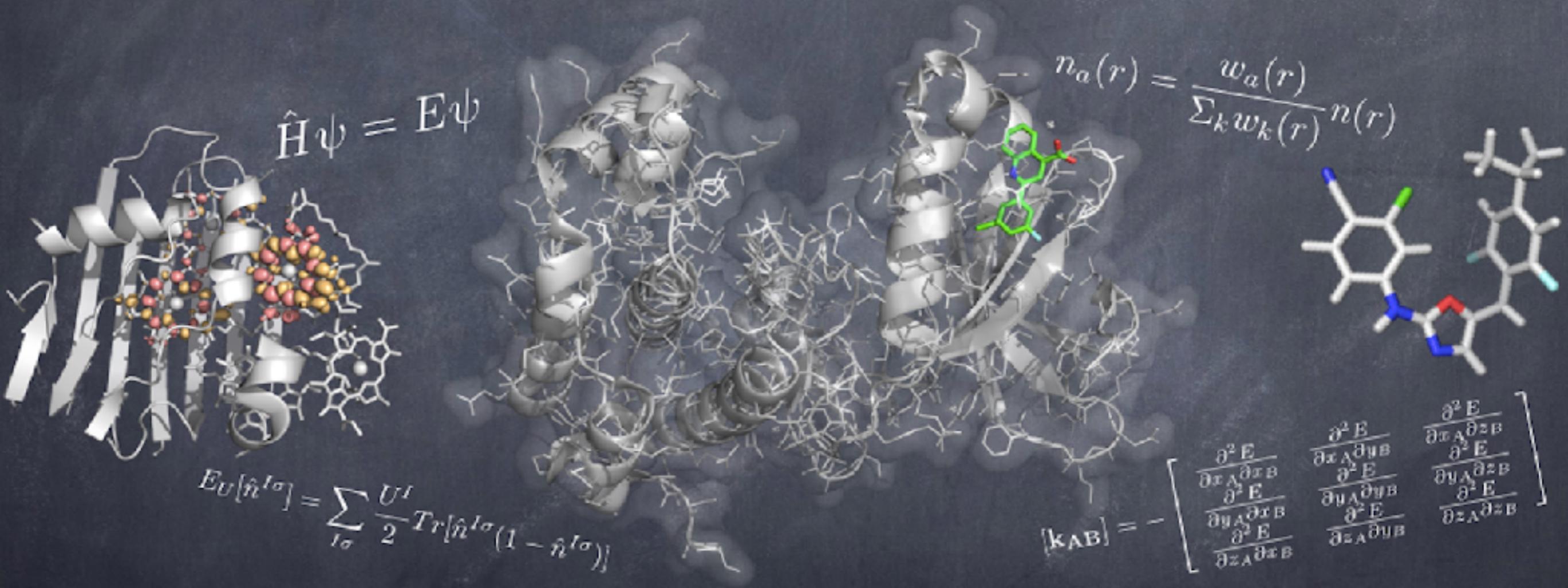
Nick Hine (Warwick)

Thomas Manz (New Mexico State)

Tom Penfold (Newcastle)

Gábor Csányi (Cambridge)

Lee-Ping Wang, David Mobley, John Chodera,
Daniel Smith (Open Force Field Initiative)



<https://blogs.ncl.ac.uk/danielcole/>

<https://github.com/cole-group/>

 @ColeGroupNCL



Thank you for your attention