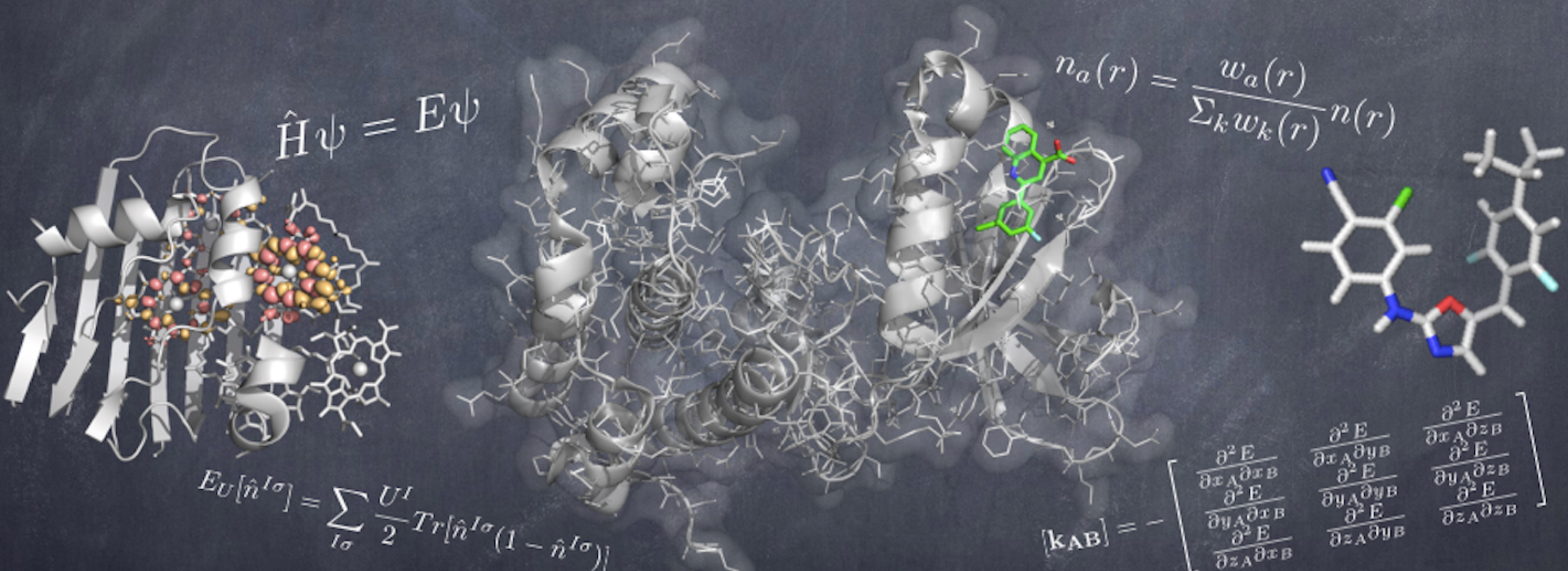


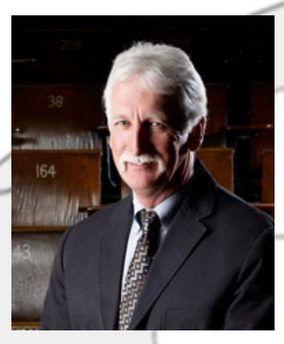
Computational Chemistry: Why, What & How?

Daniel Cole
School of Natural and Environmental Sciences

Newcastle University

MoSMed
Molecular Sciences for Medicine

Biography



Prof. Bill Jorgensen

Yale

2

UK Research and Innovation

Newcastle University

3

UNIVERSITY OF CAMBRIDGE

Prof. Mike Payne FRS

1

NUS National University of Singapore

SEVENTH FRAMEWORK PROGRAMME

MARIE CURIE ACTIONS

Computational Chemistry: Why?

What are the advantages and disadvantages of
using computers in molecular sciences?

Advantages:

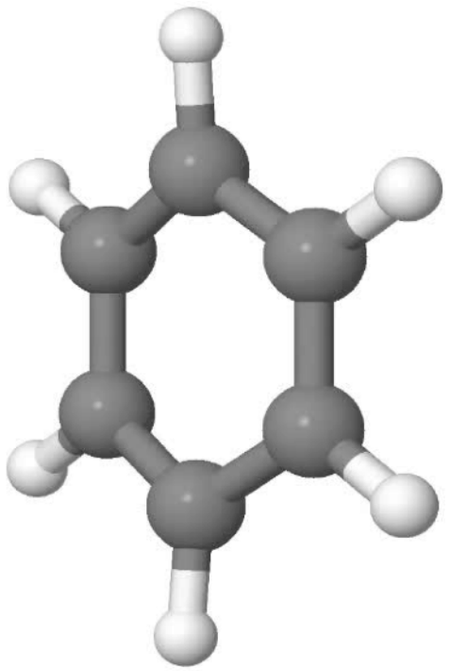
Cost (1 cpu hour = £0.02).
Feasibility (synthetic route to compound may be unknown).
Safety (no COSHH forms!).
May obtain extra information that you cannot ‘see’ experimentally.

Disadvantages:

Limitations in accuracy.
Tendency to use computer as a ‘black box’
(i.e. not understand what it is doing).

Goal of this talk is to help you to understand a) what sort of problem can be accurately addressed and b) how you can get involved.

Computational Chemistry: What?



Atomistic modelling (cf structure-based design)

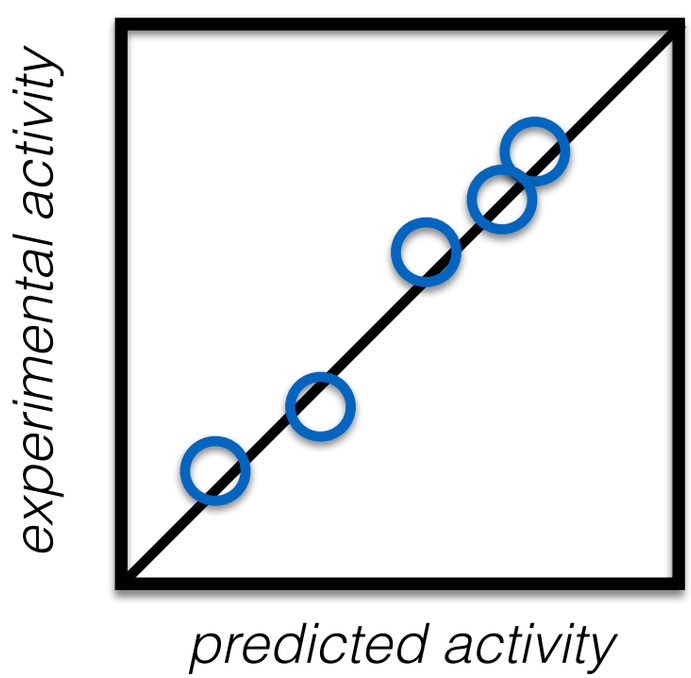
We know the atomic coordinates, aim to use the laws of physics to calculate molecular properties.

Lauren Nelson, Chris Ringrose, Ben Cree

Quantitative structure-activity relationships **(QSAR)**
(cf ligand-based design)

Molecular properties may be too difficult to compute accurately, so we may wish to identify correlations between chemical structures and properties.

Rachael Pirie, Matthew Roberts



Structure-Activity Relationships

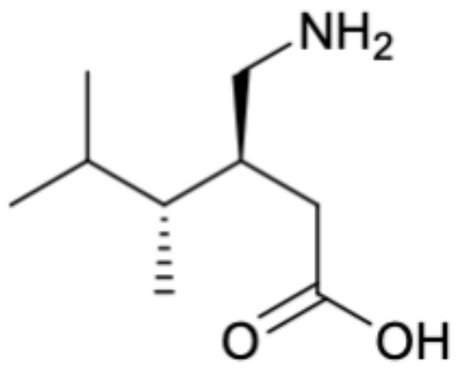
Quantitative structure-activity relationships **(QSAR)**
(cf ligand-based design)

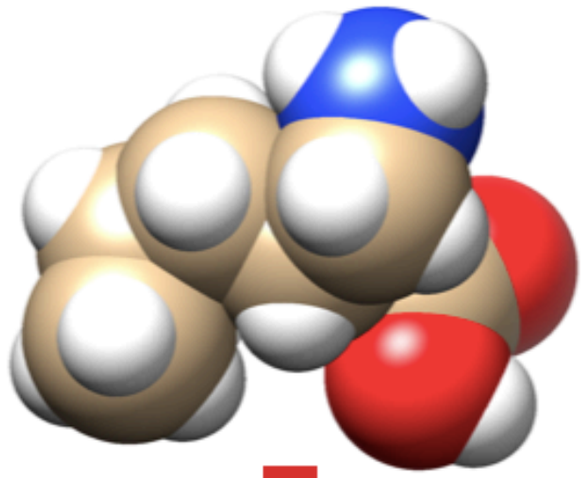
How to compute chemical/ structural similarity is an open research question.

We are looking at techniques from differential geometry to efficiently describe and compare 3D shape.

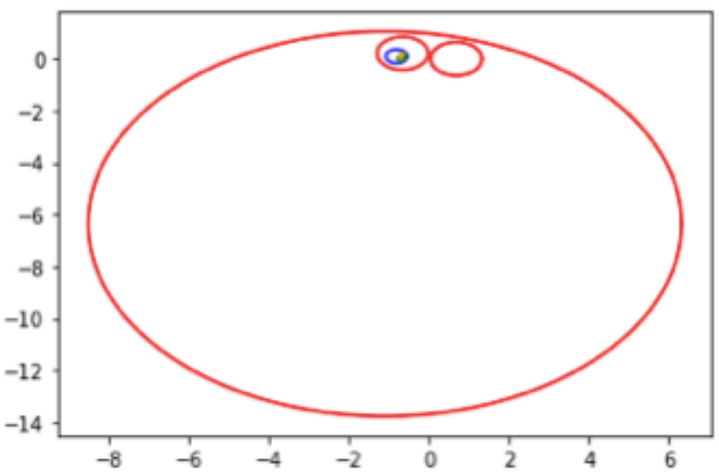
Rachael Pirie, Dr Stuart Hall

Pregabalin



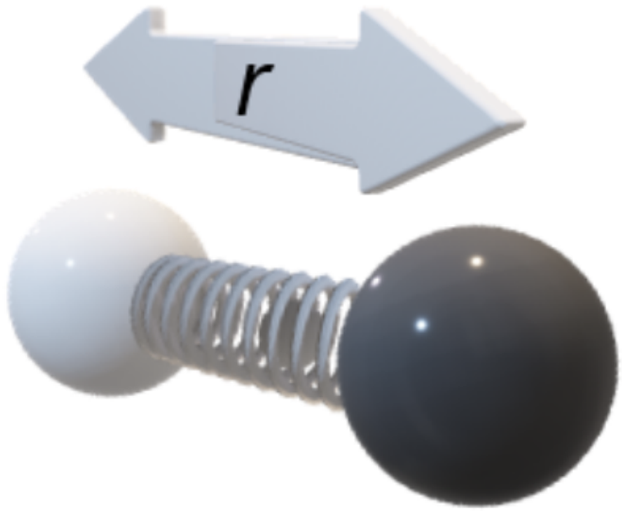


Plot of surface in $\mathbb{CP}^1$: each circle represents an atom.



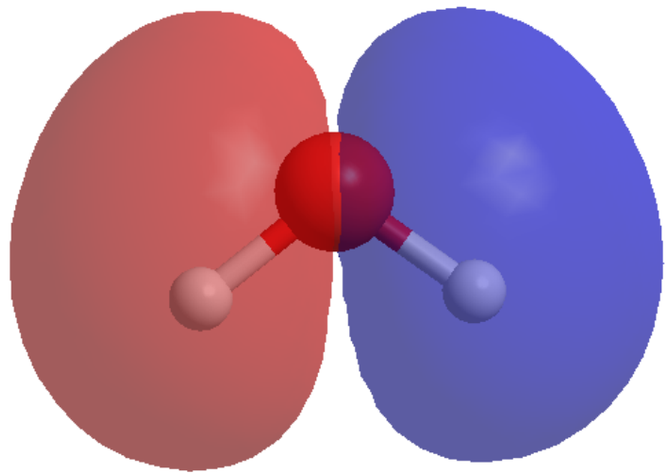
$$\begin{matrix} 7015.18+0j & -4703.64-673.21j & 3086.69+887.98j \\ -4703.64-673.21j & 3246.89+0j & -2188.55-305.06j \\ 3086.69-887.98j & -2188.55+305.06j & 1516.93+0j \end{matrix}$$

Atomistic Modelling



Classical mechanics

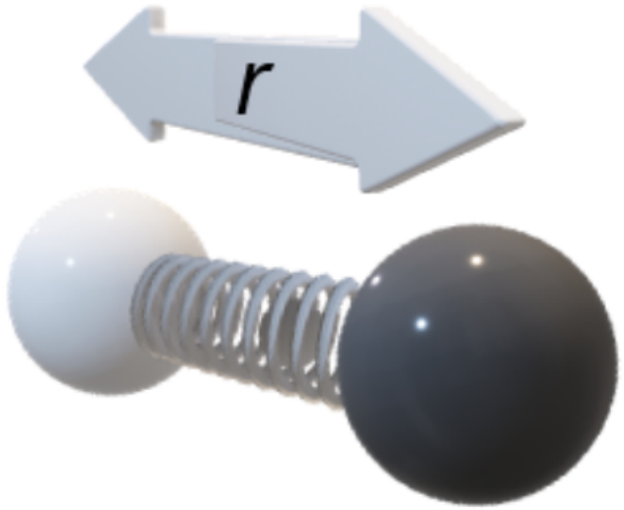
fast to run, large system sizes, not very accurate



Quantum mechanics

very accurate, small system sizes, very expensive to run

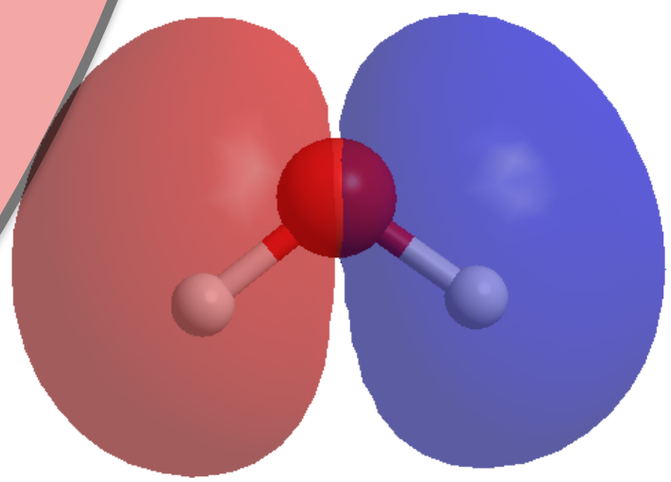
Atomistic Modelling



Classical mechanics

fast to run, large system sizes, not very accurate

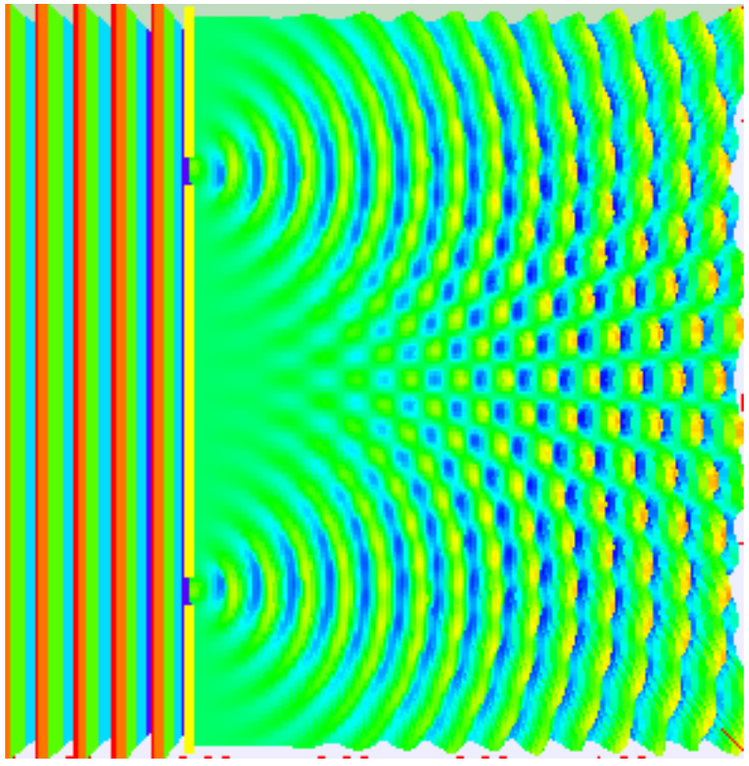
Not much available in this space!

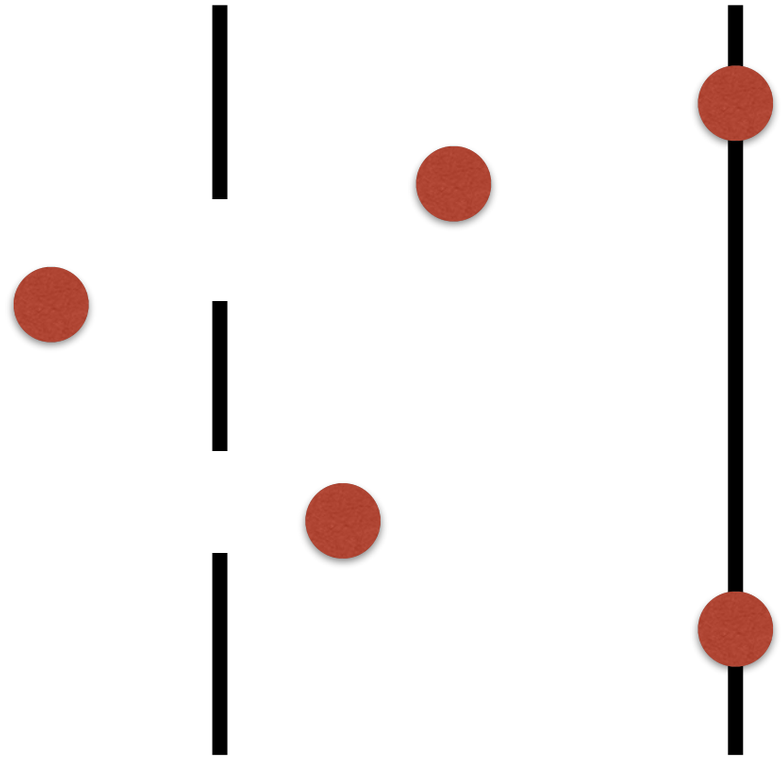


Quantum mechanics

very accurate, small system sizes, very expensive to run

Foundations of Quantum Mechanics



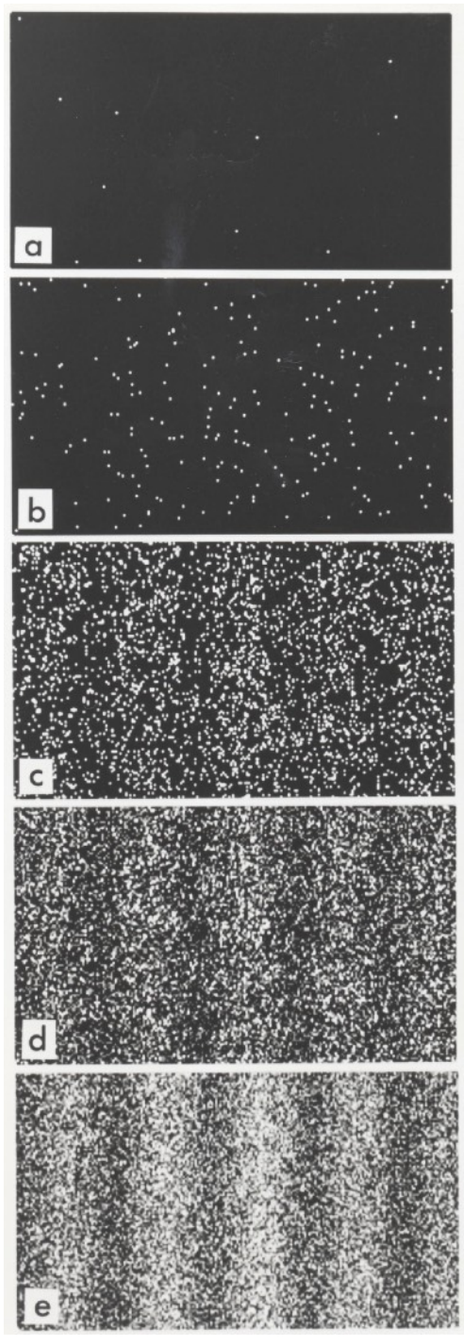


In the famous double-slit experiment, electrons (or other sources) are fired through two closely-spaced parallel gaps and detected on a screen.

In the wave picture, the sources interfere (like water waves) and produce a series of light and dark patches. In the classical picture, only two light patches are expected.

http://weelookang.blogspot.co.uk/2011/10/ejs-open-source-double-slit-diffraction.html

Foundations of Quantum Mechanics



Remarkably, **even when the electrons are sent through one at a time**, we see a wave-like interference pattern on the screen.

Profoundly, this tells us that the electrons are not acting like classical particles (e.g. snooker balls). We cannot know exactly where the electron is, only where it is likely to be found (its **probability density**).

If we are going to model how electrons behave in molecules (the foundations of all of chemistry), we need a theory that is different to the equations of classical mechanics….**quantum mechanics**.

https://commons.wikimedia.org/wiki/File%3ADouble-slit_experiment_results_Tanamura_2.jpg

Foundations of Quantum Mechanics

Erwin Schrödinger
Prize share: 1/2

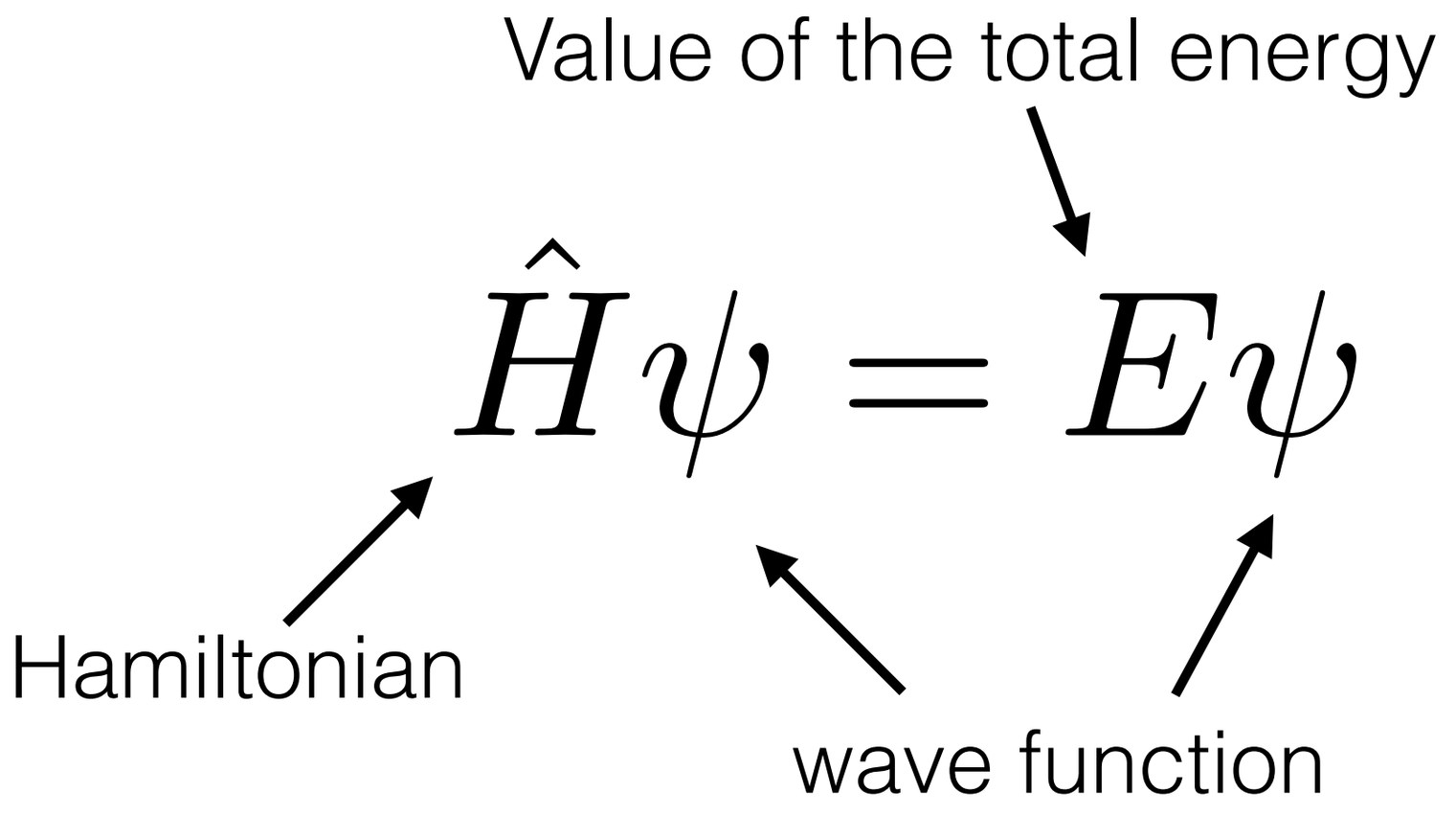
Paul Adrien Maurice Dirac
Prize share: 1/2

The Nobel Prize in Physics 1933 was awarded jointly to Erwin Schrödinger and Paul Dirac “for the discovery of new productive forms of atomic theory”.

Assuming that matter (e.g., electrons) could be regarded as both particles and waves, in 1926 Erwin Schrödinger came up with a wave equation that accurately calculated the energy levels of electrons in atoms.

"The Nobel Prize in Physics 1933". Nobelprize.org. Nobel Media AB 2014. Web. 25 Jan 2018. <http://www.nobelprize.org/nobel_prizes/physics/laureates/1933/>

Schrödinger Equation



Solution of the Schrödinger equation tells us the **wave function** (where the electrons are) and the **total energy** of a configuration of atoms.

Goal is to find E and ψ, such that action of $\hat{H}$ on ψ returns Eψ.

Quantum Mechanics: Case Study

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

8-Substituted O^6-Cyclohexylmethylguanine CDK2 Inhibitors: Using Structure-Based Inhibitor Design to Optimize an Alternative Binding Mode

Benoit Carbain,[⊥,‡] David J. Paterson,[⊥,†] Elizabeth Anscombe,[†] Allyson J. Campbell,[‡] Celine Cano,[‡] Aude Echalier,[†,▽] Jane A. Endicott,*[,†,#] Bernard T. Golding,[‡] Karen Haggerty,[‡] Ian R. Hardcastle,[‡] Philip J. Jewsbury,[§] David R. Newell,[∥] Martin E. M. Noble,[†,#] Celine Roche,[‡] Lan Z. Wang,[∥] and Roger J. Griffin*[,‡]

[†]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

[‡]Newcastle Cancer Centre, Northern Institute for Cancer Research, School of Chemistry, Newcastle University, Bedson Building, Newcastle upon Tyne NE1 7RU, U.K.

[§]AstraZeneca Pharmaceuticals, Alderley Park, Cheshire SK10 4TG, U.K.

[∥]Newcastle Cancer Centre, Northern Institute for Cancer Research, Newcastle University, Paul O'Gorman Building, Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K.

CDK2 Inhibition

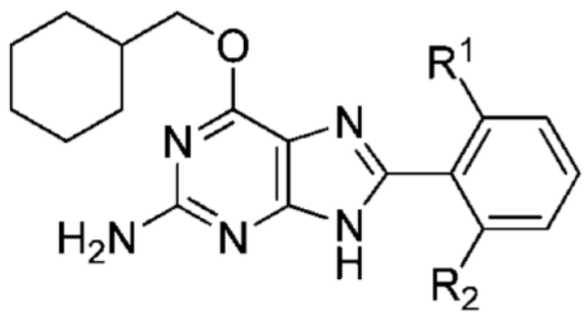
What experimental observation motivated this investigation?

Compelling evidence has recently been found to support a therapeutic role for pharmacological CDK2 inhibition in cancer treatment.

Compound **27** has the potential to form hydrogen bonding interactions with CDK2, as well as *π*-*π* stacking.

However only shows 10% inhibition of CDK2 at 100 μM concentration.

It was hypothesised that a purine-phenyl twist of 40 degrees would optimise packing.



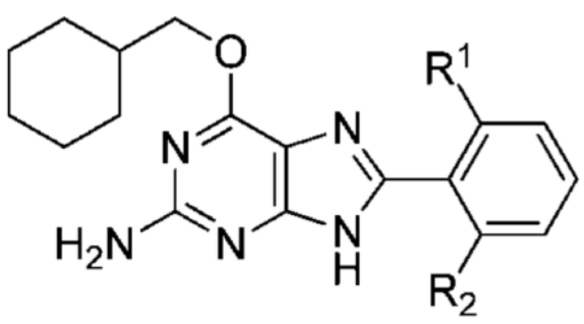
27: $R_1 = R_2 = H$

CDK2 Inhibition

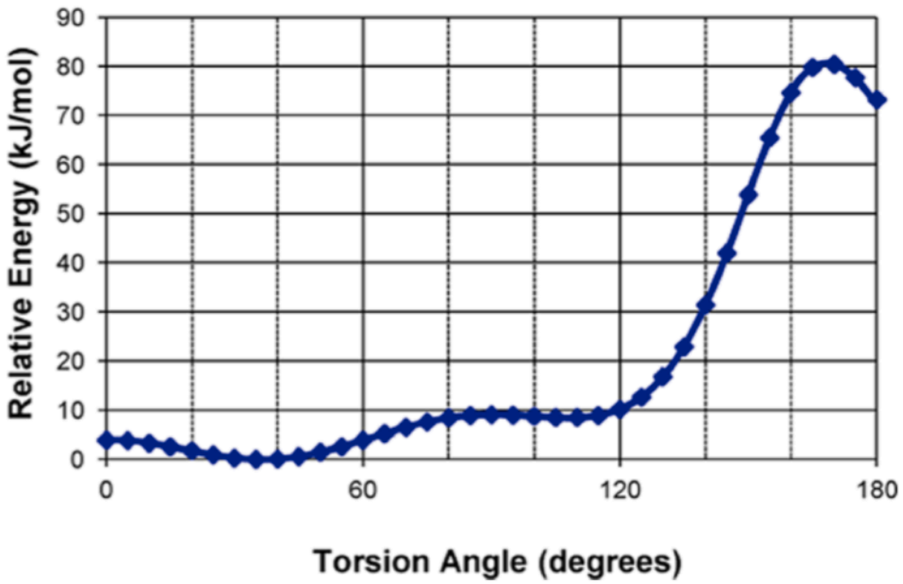
What QM methods are used?

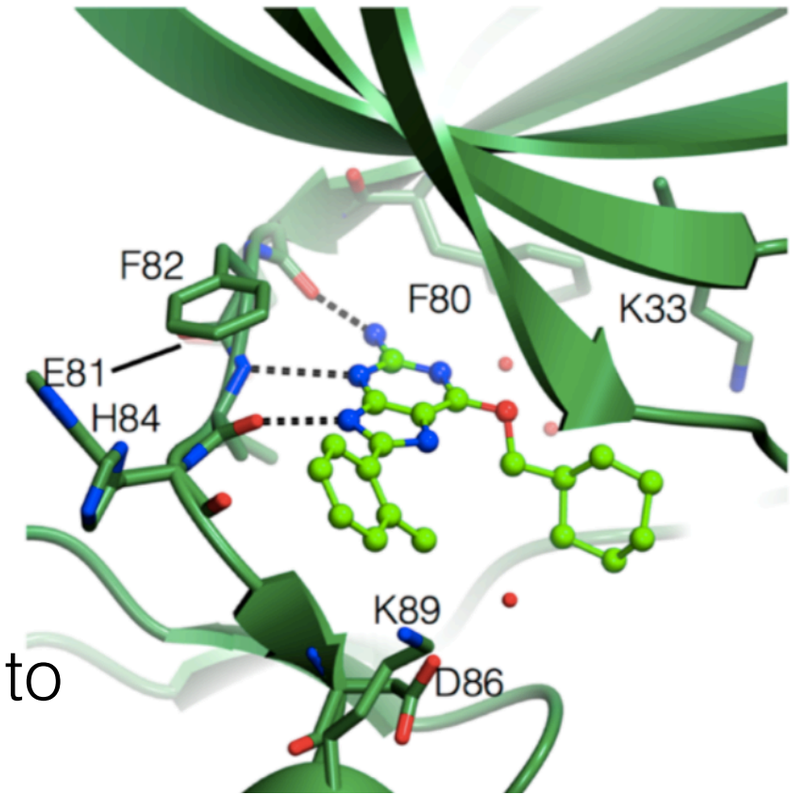
QM dihedral scans with a range of R_1, R_2 substituents are run.

What useful information does QM provide?



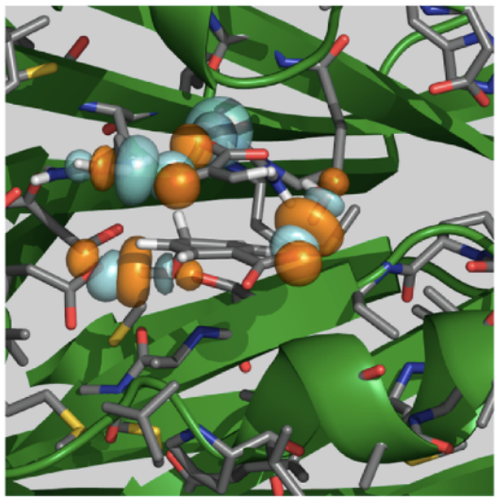
29: R_1 = Me, R_2 = H



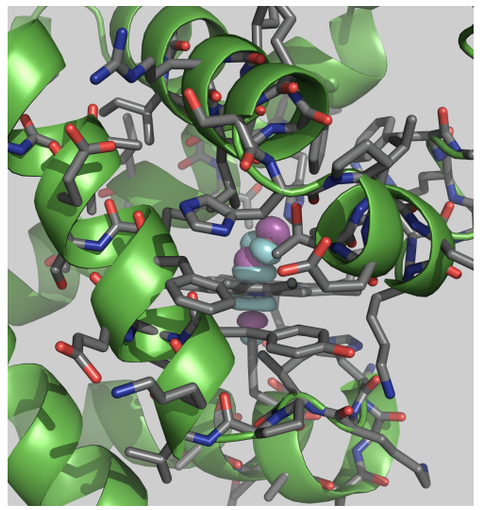


Optimal ortho substitution was synthesised and shown to be relatively potent against CDK2 (IC_{50} = 18 μM).

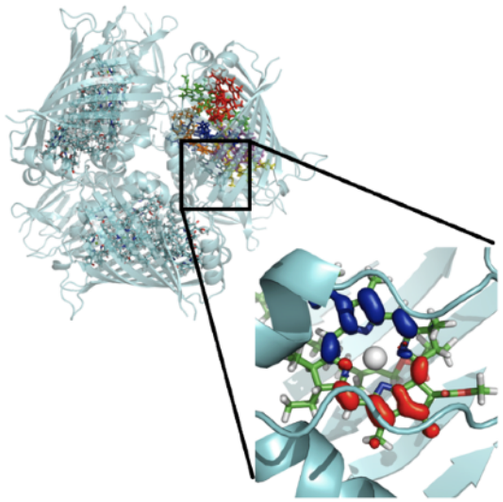
Biological Applications



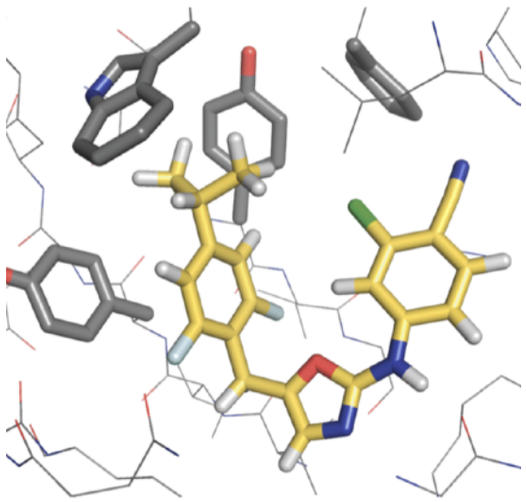
Transition state searching in enzymes



Protein-ligand binding in metalloproteins



Optical spectroscopy in a light-harvesting protein



Classical force field parameterisation for drug discovery

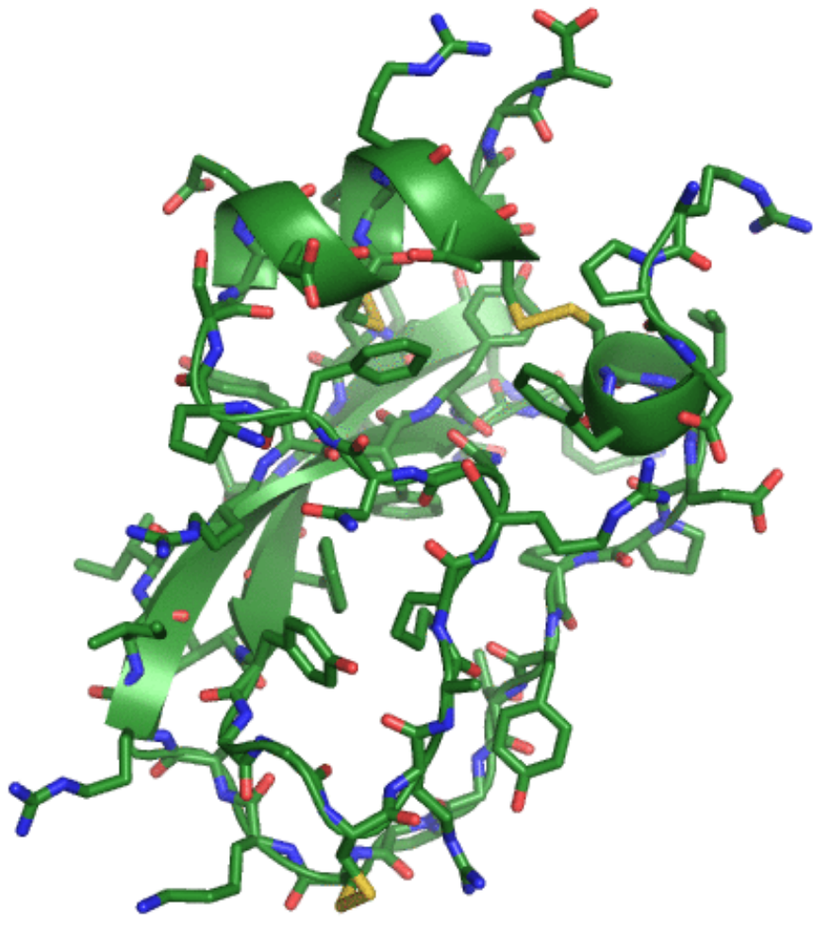
"Applications of Large-Scale DFT in Biology", D. J. Cole & N. D. M. Hine
Journal of Physics: Condensed Matter, 28, 393001 (2016)

Molecular Mechanics

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

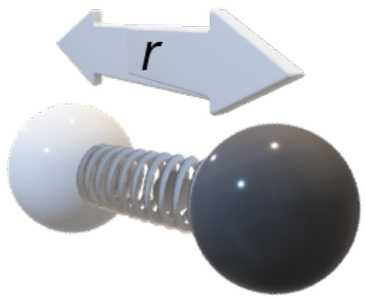
MM can complement experiment in many ways. At the simplest level, we can use molecular dynamics to ‘animate’ the system.

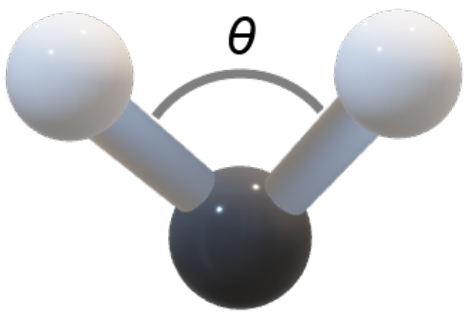
Simulations tend to be limited only by finite sampling and accuracy of the force field.

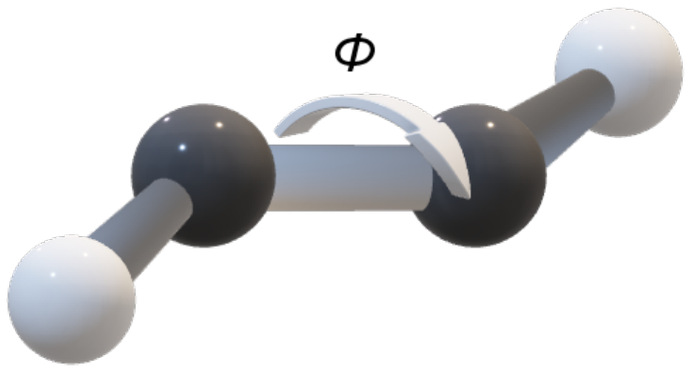


Force Field

$$E_{Total} = \sum_{Bonds} K_r(r - r_0)^2 + \sum_{Angles} K_\theta(\theta - \theta_0)^2 + \sum_{Torsions} \frac{V_n}{2}[1 + cos(n\Phi - \gamma)]$$







Bonded (Intramolecular) Parameters

$$+ \sum_{Non-Bonded} \left[4\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} + \frac{q_i q_j}{r_{ij}} \right]$$

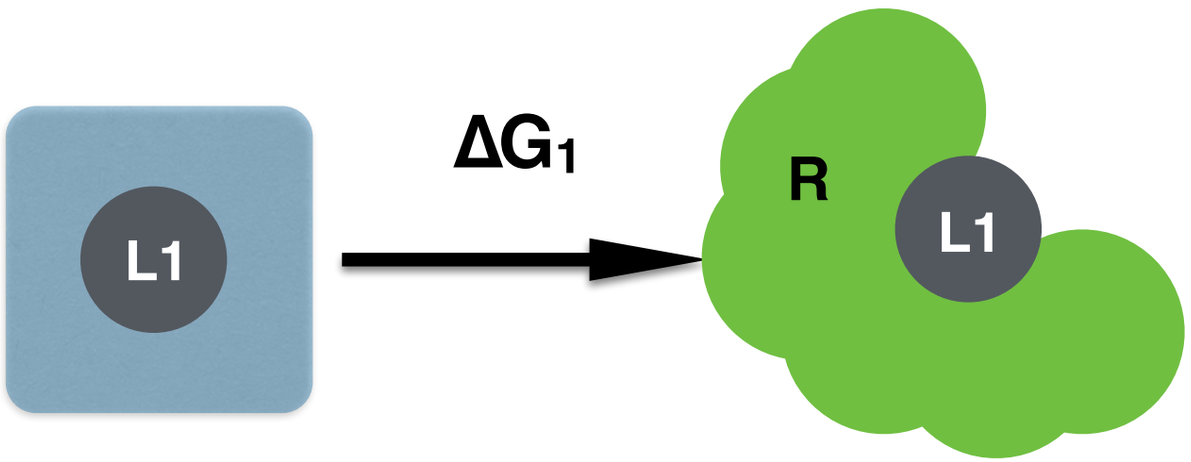
Non-Bonded (Intermolecular) Parameters

An open and collaborative approach to better force fields

https://github.com/qubekit/QUBEKit

open forcefield

FEP for Drug Discovery

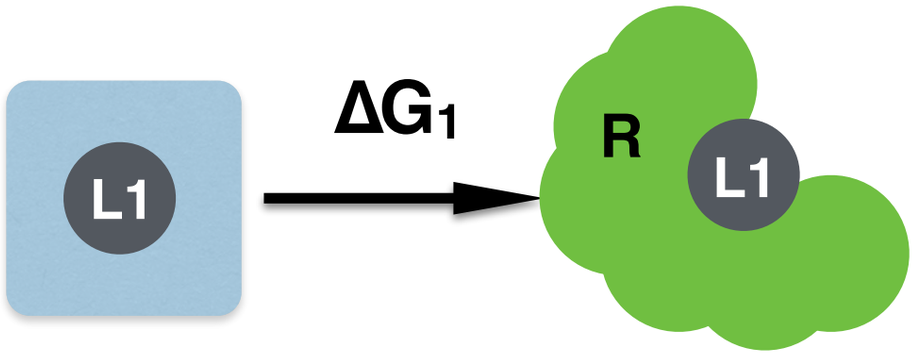


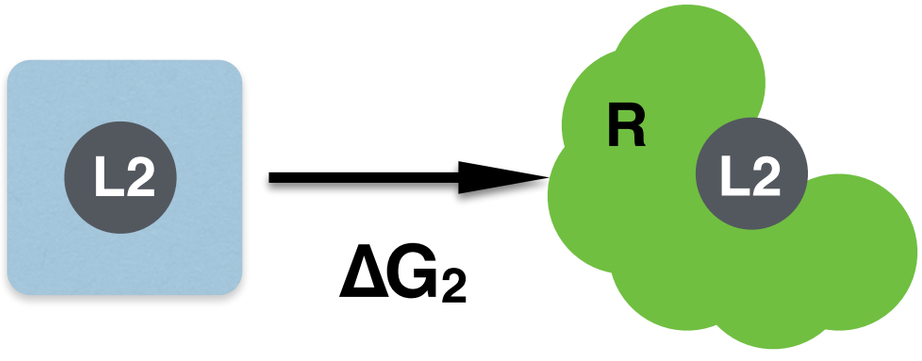
In lead optimisation studies, we are typically interested in optimising the target-ligand binding affinity.

In other words, we need to find the free energy difference between a small molecule (L1) in solution and bound to the protein (R).

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

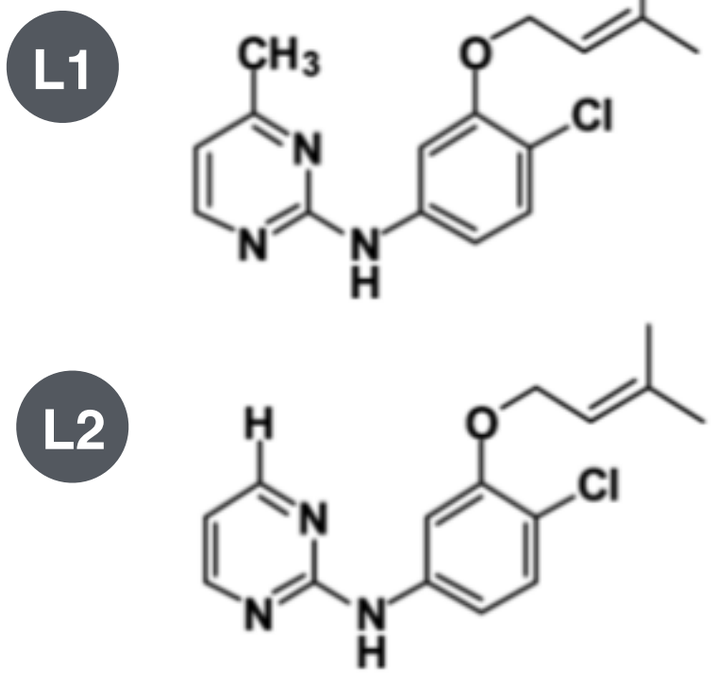
FEP for Drug Discovery





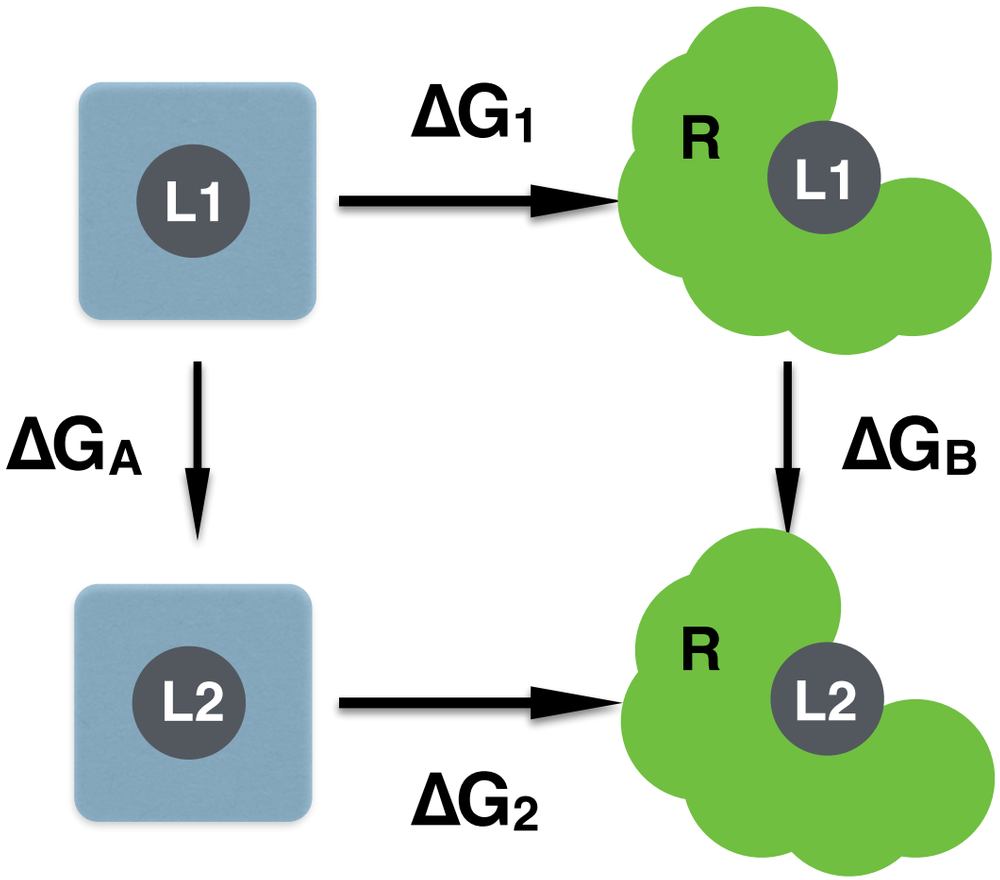
$\Delta\Delta G = \Delta G_2 - \Delta G_1$

For example:



If we have two similar molecules, then often we only need to compute the relative binding free energy ΔΔG.

FEP for Drug Discovery



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

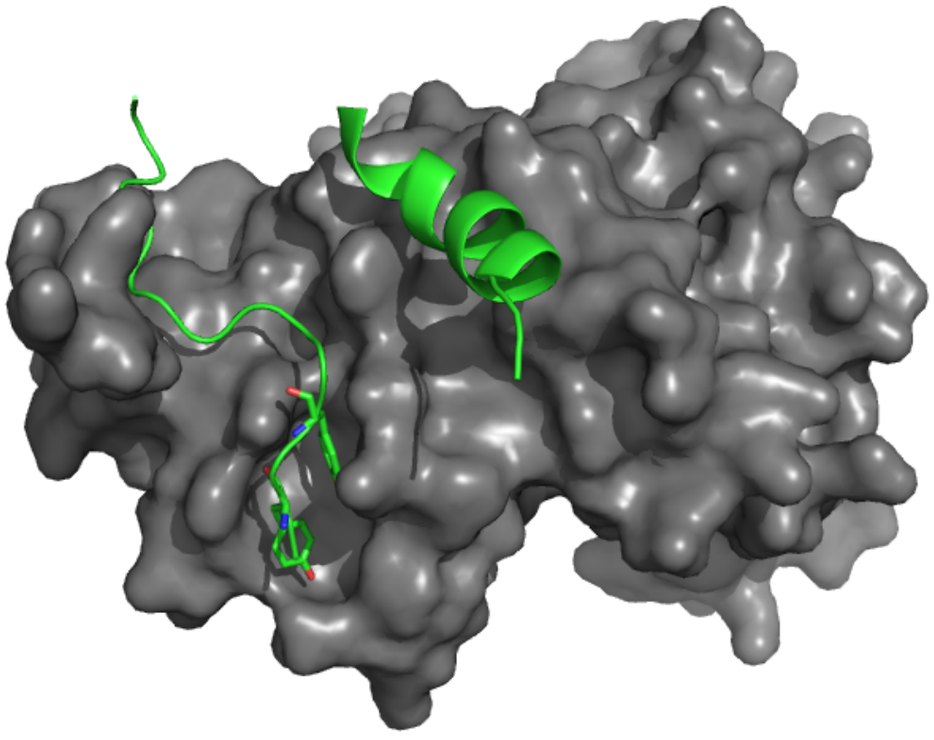
$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A$$

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.

Computer-Aided Drug Design

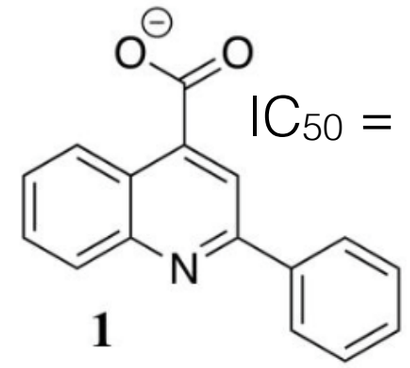


Aurora A kinase plays a central role in cell division. It is oncogenic and over-expressed in various tumour types.

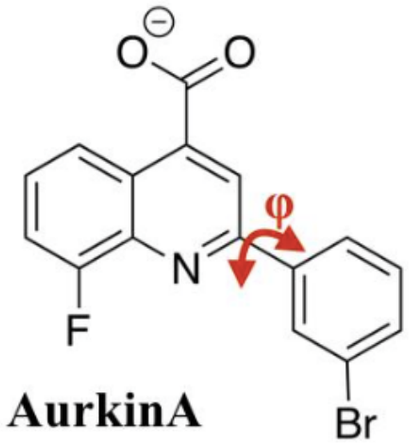
Inhibition of Aurora A leads to cell death in dividing cells and it is a potential drug target in cancer.

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

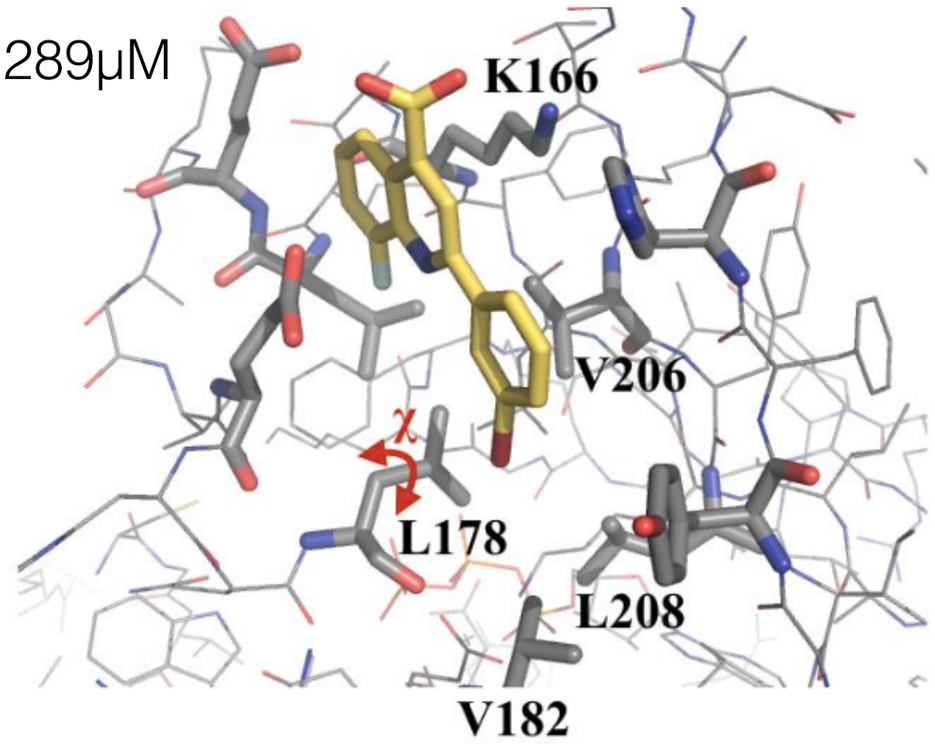
Computer-Aided Drug Design

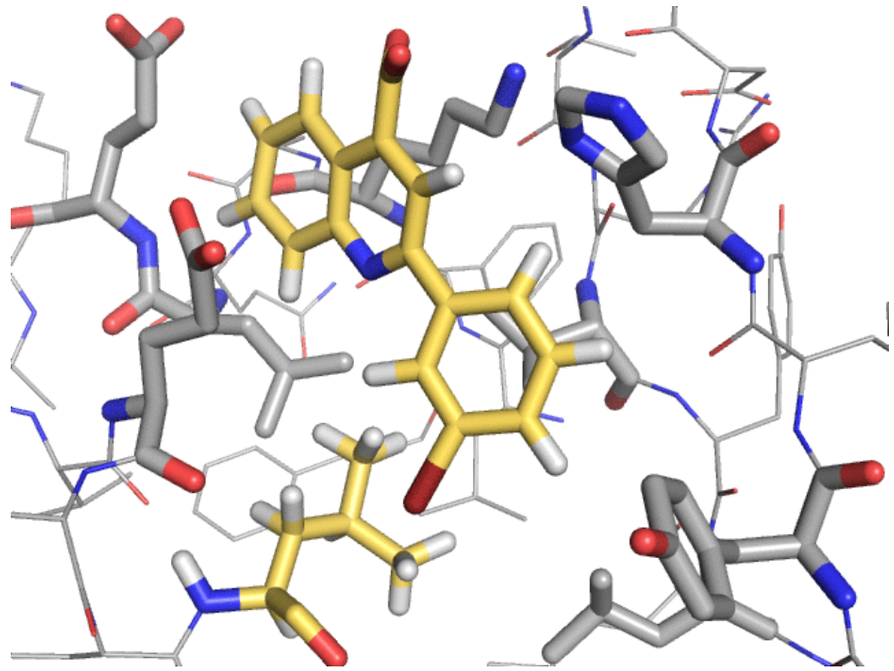


IC_{50} = 289μM



IC_{50} = 12.7μM

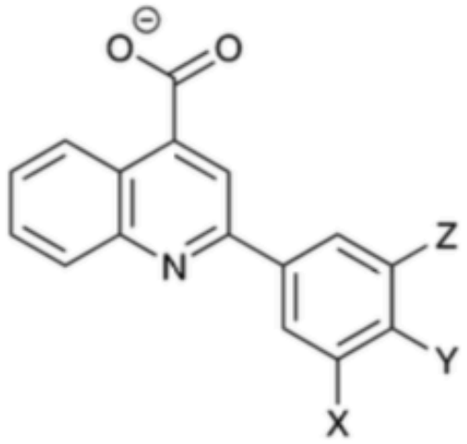




Used FEP to investigate small substitutions around the phenyl ring. Replica exchange with solute tempering (REST) enhanced sampling of ligand and L178.

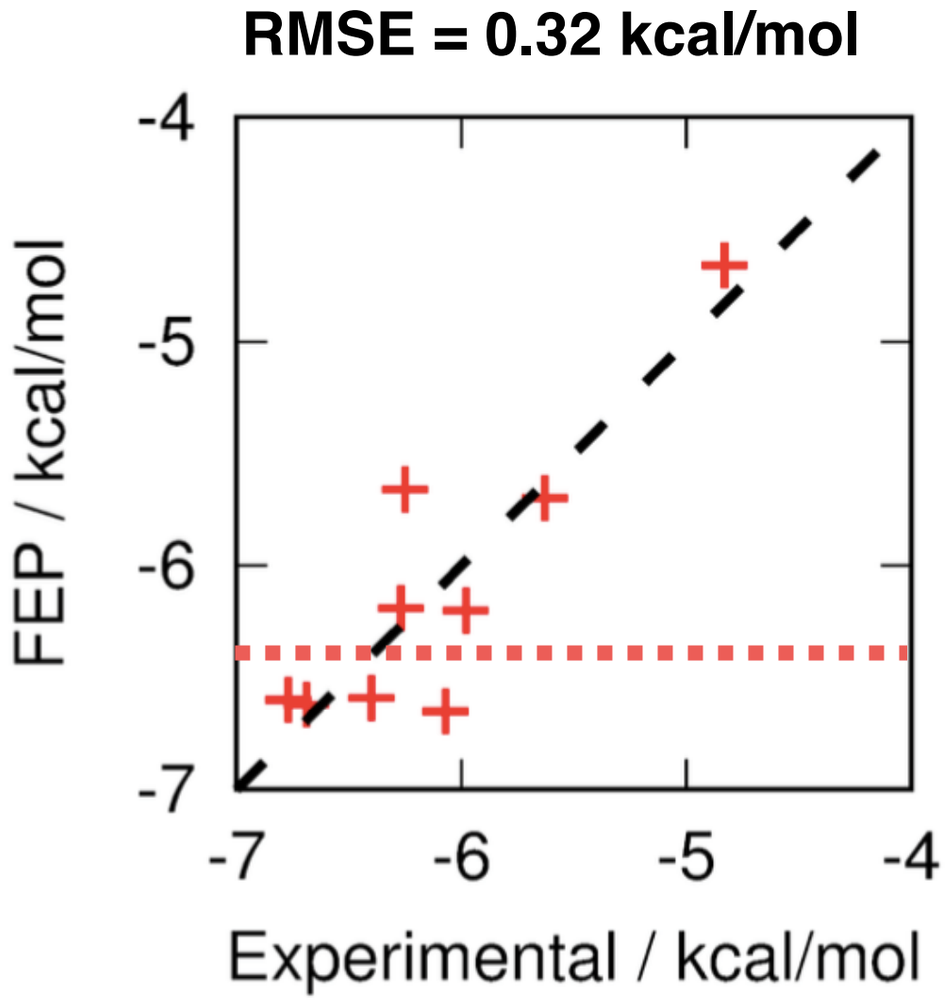
Cole et al., *Chem. Commun.* **2017**, *53*, 9372
Cole, Tirado-Rives & Jorgensen, *JCTC* **2014**, *10*, 565

Computer-Aided Drug Design



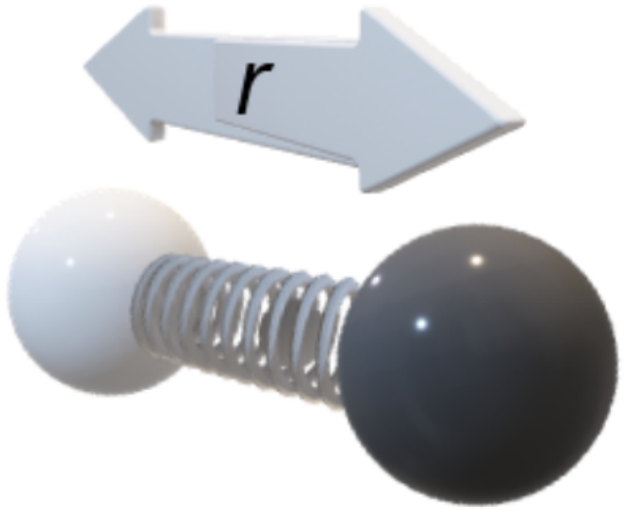
	X	Y	Z	$\Delta\Delta G^a$	$IC_{50}{}^b$	$K_i{}^b$
1	H	H	H	1.05	289	62.5
2	F	H	H	0.00	75.9	16.5
3	F	H	F	−0.94	36.0	7.8
4	Cl	H	H	−0.73	ND	ND
5	Cl	H	F	−0.89	20.5	4.4
6	Br	H	H	−0.49	25.6	5.5
7	CF_3	H	H	0.11	26.5	5.7
8	CH_3	H	H	1.12	ND	ND
9	F	CH_3	H	−0.49	42[c]	8.7[c]
10	Br	CH_3	H	−0.90	11.1[c]	2.3[c]

[a] kcal mol^{-1}. [b] μM. [c] This work (see the ESI).



D. J. Cole et al., *Chem. Commun.* **2017**, *53*, 9372
M. Janecek et al., *Scientific Reports* **2016**, *6*, 28528

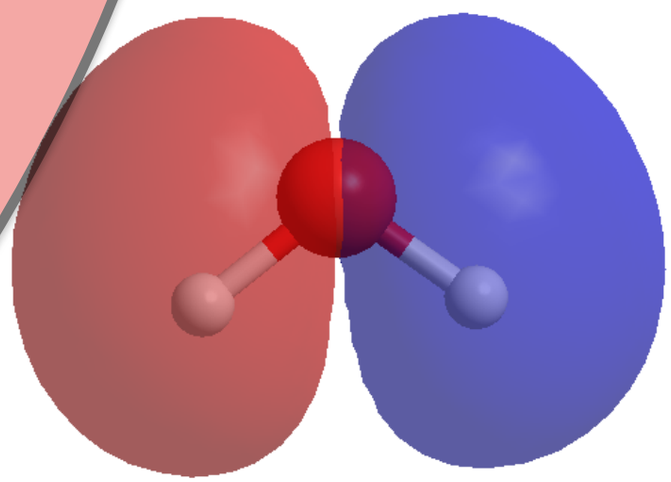
Atomistic Modelling



Classical mechanics

fast to run, large system sizes, not very accurate

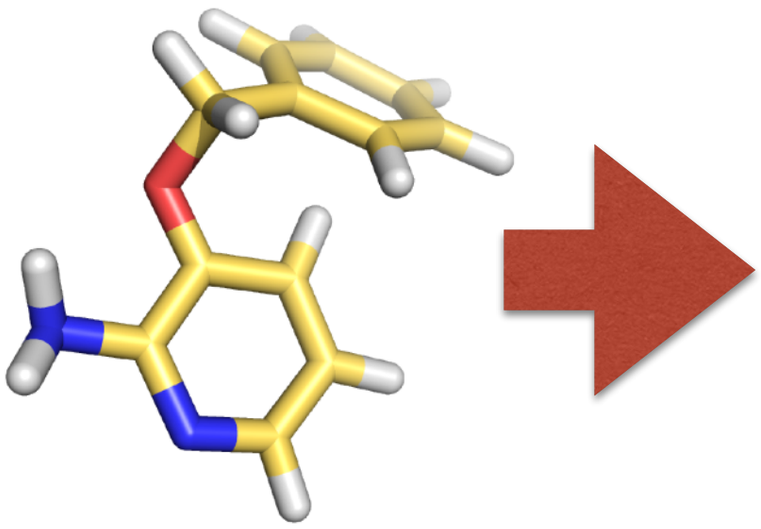
Not much available in this space!



Quantum mechanics

very accurate, small system sizes, very expensive to run

Gaussian Approximation Potential



Total energy and forces

Gaussian approximation potential (GAP) writes the potential energy as a generic function of atomic coordinates. Trained using QM energies and forces.

$$E(\mathscr{A}) = \sum_{\mathscr{B} \in M} x_{\mathscr{B}} K(\mathscr{A}, \mathscr{B})$$

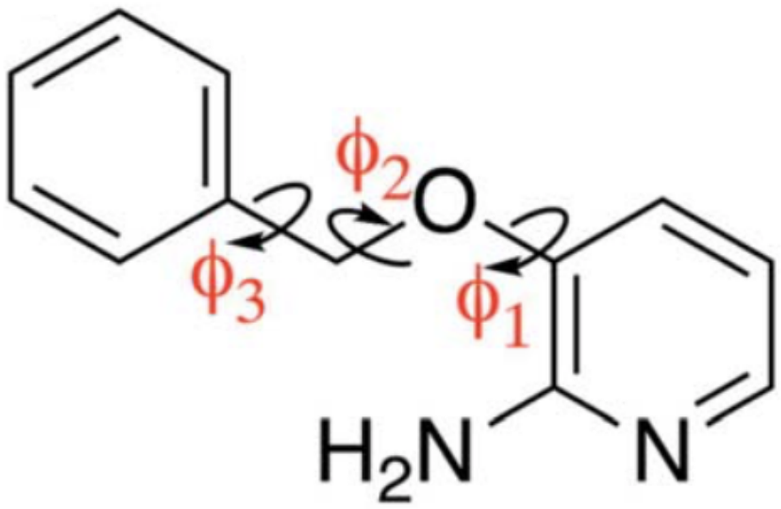
target energy of conformation A (represented by vector of interatomic distances)

unknown coefficients (regularised least squares regression)

squared exponential kernel (similarity function between A and B)

Bartók, Payne, Kondor, Csányi, *Phys. Rev. Lett.* **2010**, *104*, 136403
Bartók, Csányi, *Int. J. Quantum Chem.* **2015**, *115*, 1051-1057

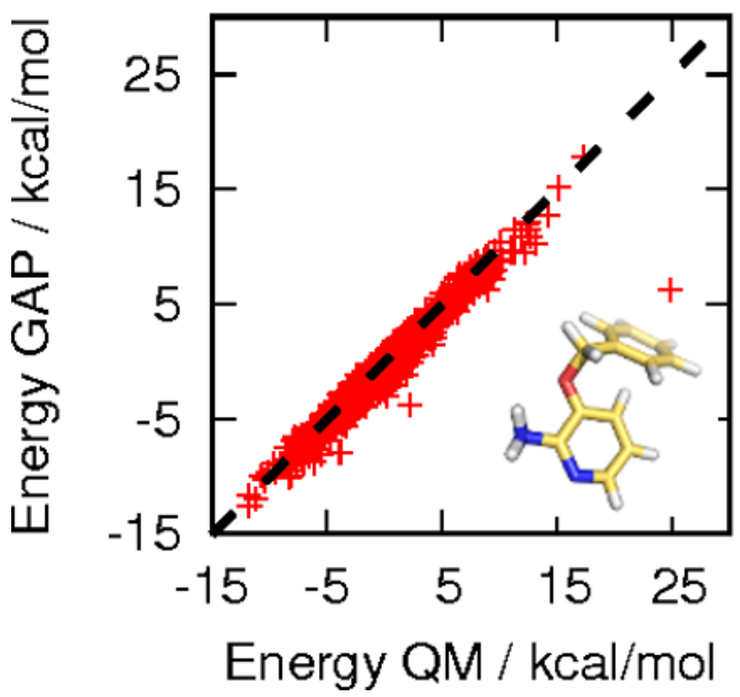
Gaussian Approximation Potential

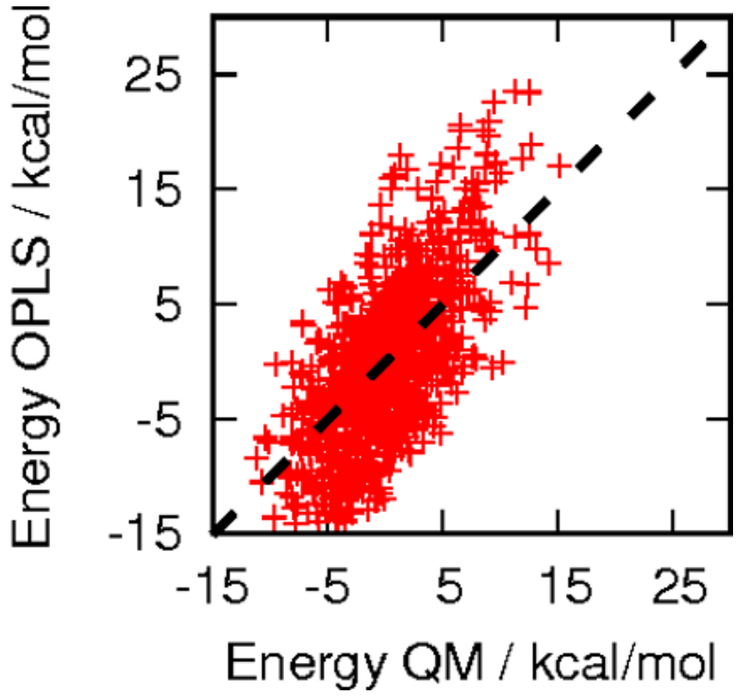


Small molecule GAP trained using QM energies/forces from ~3000 configurations.

Errors on 900 configurations from Monte Carlo simulations:

GAP
RMS error =
1 kcal/mol

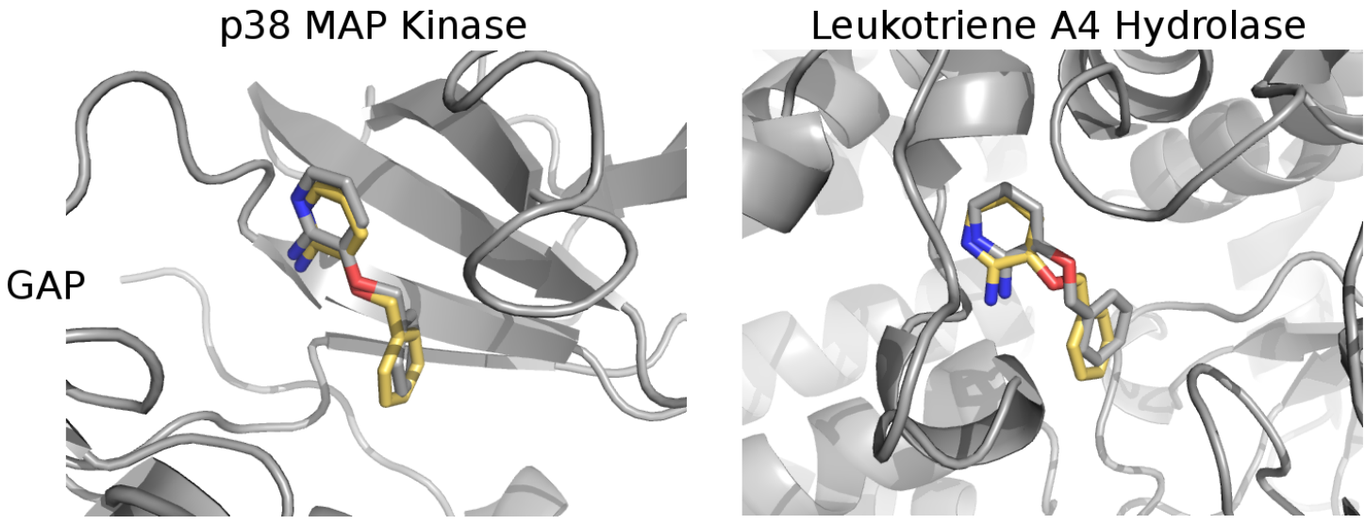




OPLS
RMS error =
4 kcal/mol

DJ Cole, L Mones, G Csányi, *Faraday Discuss.*, **2020**, in press

Interactions with Protein



Implemented in MCPRO molecular modelling software.

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

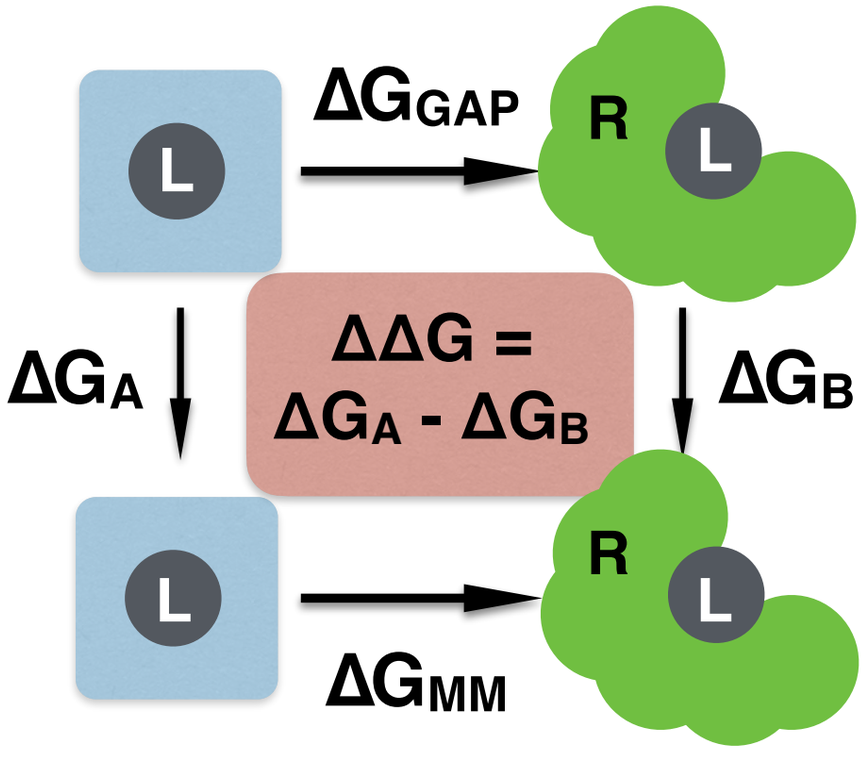
Simulated in 3 environments, bound to two proteins and in water.

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

DJ Cole, L Mones, G Csányi, *Faraday Discuss.*, **2020**, in press

Free energy calculations

It would be preferable to compute the free energy of binding using the GAP, but this is still too expensive at the moment.



Instead we can use FEP to compute the GAP correction to MM free energy of binding (ΔΔG):

	ΔΔG / kcal/mol
Kinase	1.0
Hydrolase	2.0

$$E_L = (1-\lambda)E_{GAP} + \lambda E_{MM}$$

DJ Cole, L Mones, G Csányi, *Faraday Discuss.*, **2020**, in press

Computational Chemistry: How?

Start by learning how to code!

A programming language is an interface between humans and computers.

The ***python*** programming language is very popular at the moment as it has a very intuitive syntax:

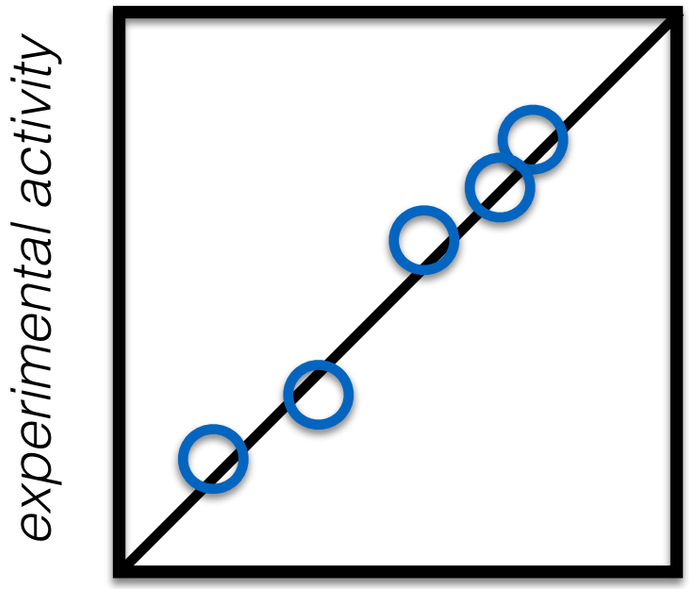
In [46]:
```
raining = True
day = "Saturday"
temperature = 22

if day == "Saturday" and temperature > 15 and not raining:
    print("Go for a walk")
else:
    print("Stay indoors")
```

```
Stay indoors
```

Lots of online courses (e.g. udemy, codecademy), welcome to ask me or Rachael for CHY1610 material.

Computational Chemistry: How?

Quantitative structure-activity relationships **(QSAR)**
(cf ligand-based design)



TeachOpenCADD

A teaching platform for computer-aided drug design (CADD) using open source packages and data

> ⚠️ We are introducing major changes to the TeachOpenCADD repository *very soon*. Before making them available, we are cutting a `1.3.0` version with the current changes on `master`. This will be last release in the `v1.x` series. Subscribe to the repository releases (top-right menu: `Watch` > `Releases`) to stay tuned!

Volkamer Lab

In Silico Toxicology and Structural Bioinformatics
Institute of Physiology
Charité - Universitätsmedizin Berlin
volkamerlab.org

TeachOpenCADD Jupyter Notebooks

DOI 10.1186/s13321-019-0351-x DOI 10.5281/zenodo.2600909 launch binder

TeachOpenCADD KNIME Workflows

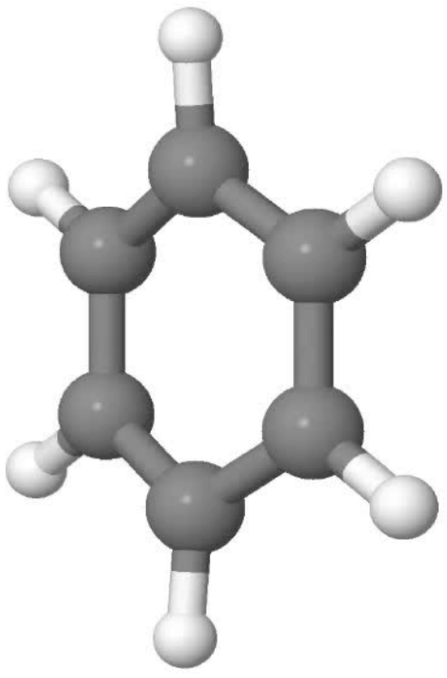
DOI 10.1021/acs.jcim.9b00662 DOI 10.5281/zenodo.3626897 KNIME Hub TeachOpenCADD-KNIME

https://github.com/volkamerlab/teachopencadd

“launch binder”
run tutorials on the cloud
from anywhere

Rachael Pirie, Matthew Roberts

Computational Chemistry: How?



Atomistic modelling (cf structure-based design)

Bit harder to get into without training, though part of Ben's project is to build & deploy tutorials/workflows like the TeachOpenCADD one.

Otherwise come and talk to us!

Lauren Nelson, Chris Ringrose, Ben Cree

Acknowledgements

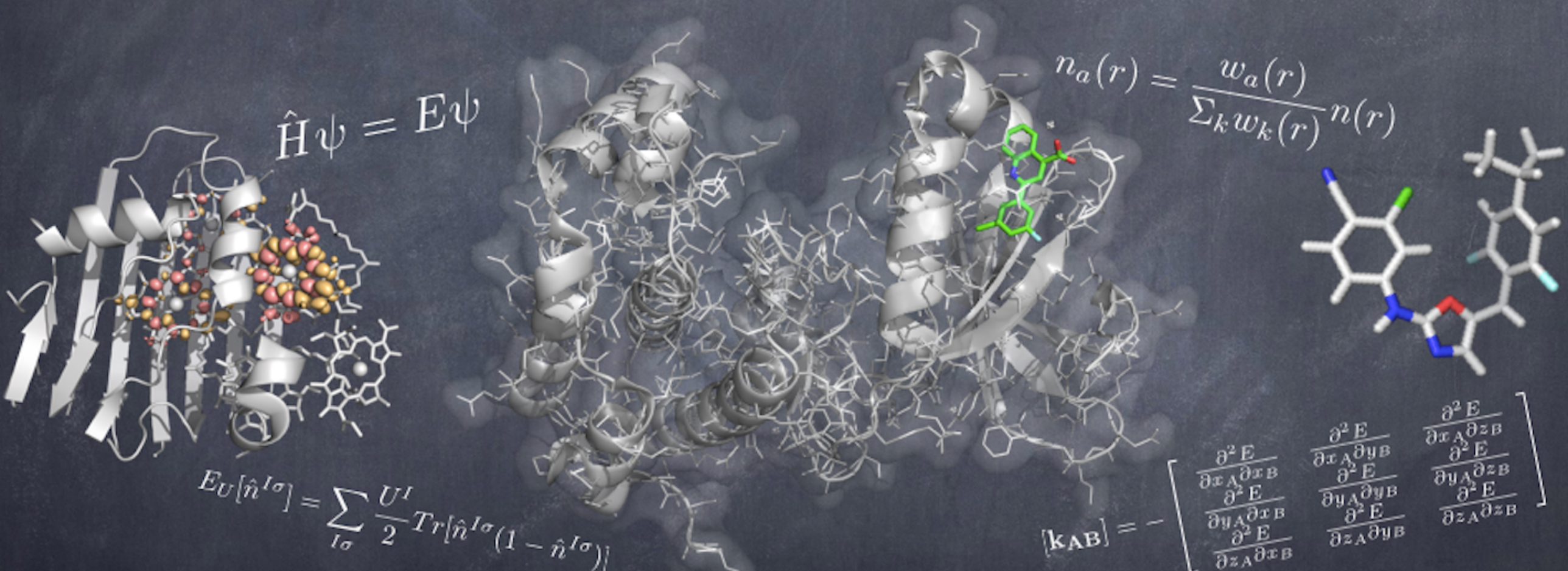
Newcastle University
Josh Horton
Lauren Nelson
Chris Ringrose
Rachael Pirie
Ben Cree
Matthew Roberts
Med Chem & Chem Bio group

Collaborators
William Jorgensen (Yale)
Thomas Manz (New Mexico State)
Gábor Csányi (Cambridge)
Mike Payne (Cambridge)
Julien Michel (Edinburgh)
Open Force Field Initiative

UK Research and Innovation

SEVENTH FRAMEWORK PROGRAMME

MARIE CURIE ACTIONS



https://blogs.ncl.ac.uk/danielcole/
https://github.com/cole-group/
@ColeGroupNCL

Newcastle University

Thank you for your attention