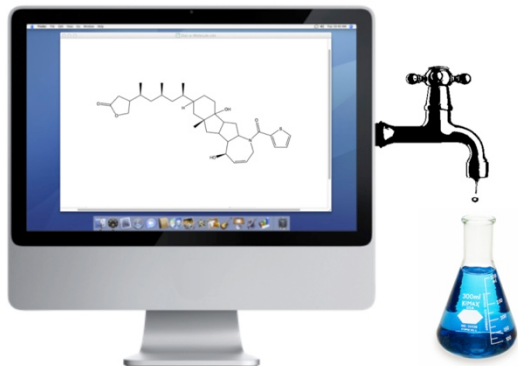




Dial-a-Molecule

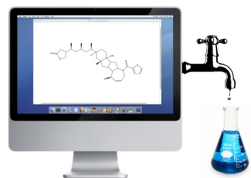
An EPSRC Grand Challenge Network

Big Data and the Dial-a-Molecule Grand Challenge

Richard Whitby, University of Southampton

From Big Data to Chemical Information

RSC, 22nd April 2015



Dial-a-Molecule

An EPSRC Grand Challenge Network

Grand Challenge

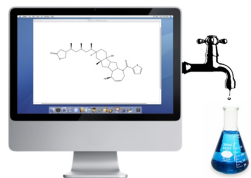
Vision: In 20-40 years, scientists will be able to deliver any desired molecule within a timeframe useful to the end-user, using safe, economically viable and sustainable processes.



Delivery of novel molecules should be as quick and efficient as it currently is for stock chemicals.

How can we make novel molecules in DAYS not YEARS?

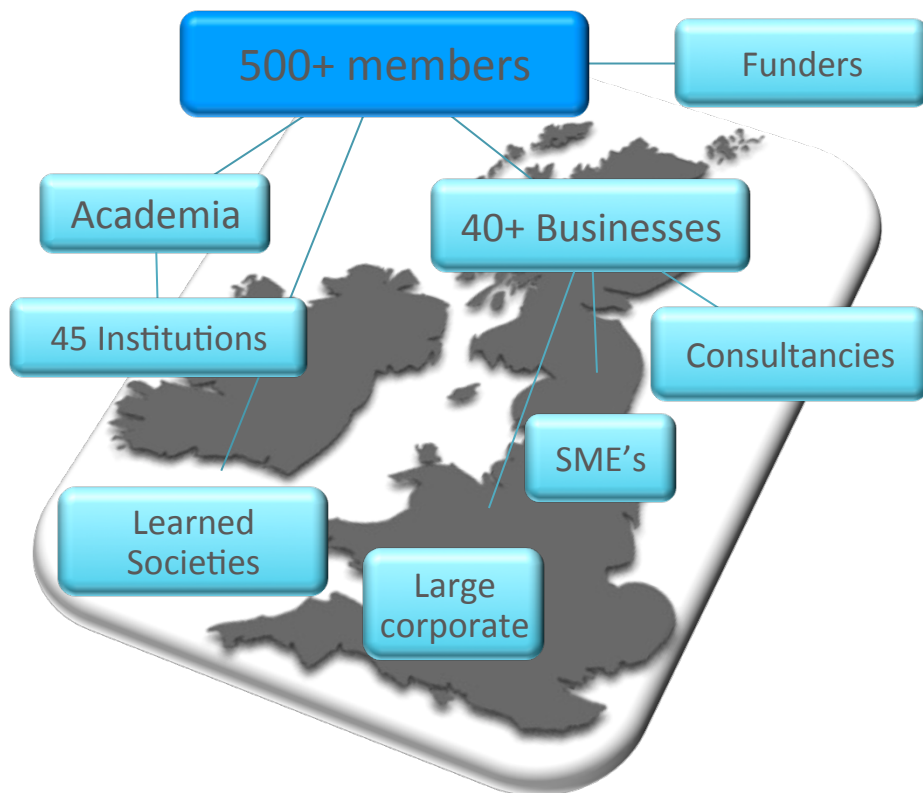
Selected as one of three Grand Challenges for Chemistry and Chemical Engineering in a competitive exercise run by EPSRC/RSC/ICChemE/CIKTN between June 2008 and Sept 2009.



Dial-a-Molecule

An EPSRC Grand Challenge Network

The Network



STEERING GROUP

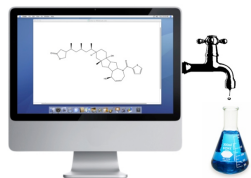
Prof. Richard Whitby^a
Prof. Steve Marsden^b
Prof. David Harrowven^a
Prof. Joe Sweeney^c
Dr. Harris Makatsoris^d
Prof. Asterios Gavriilidis^e
Dr. David Hollinshead^f
Dr. David Fox^g
Dr. Mimi Hii^h

Dr. Andrew Russellⁱ
Dr. Robin Attrill^j
Prof. Nick Turner^k
Dr. Matt Tozer^l
Dr. John Clough^m
Dr. Gill Smithⁿ
Prof. John Leonard^o
Dr. Natasha Richardson^p
Dr. Simon Rushworth^q

Network
Co-ordinator

Dr Kelly Kilpin
University of Southampton

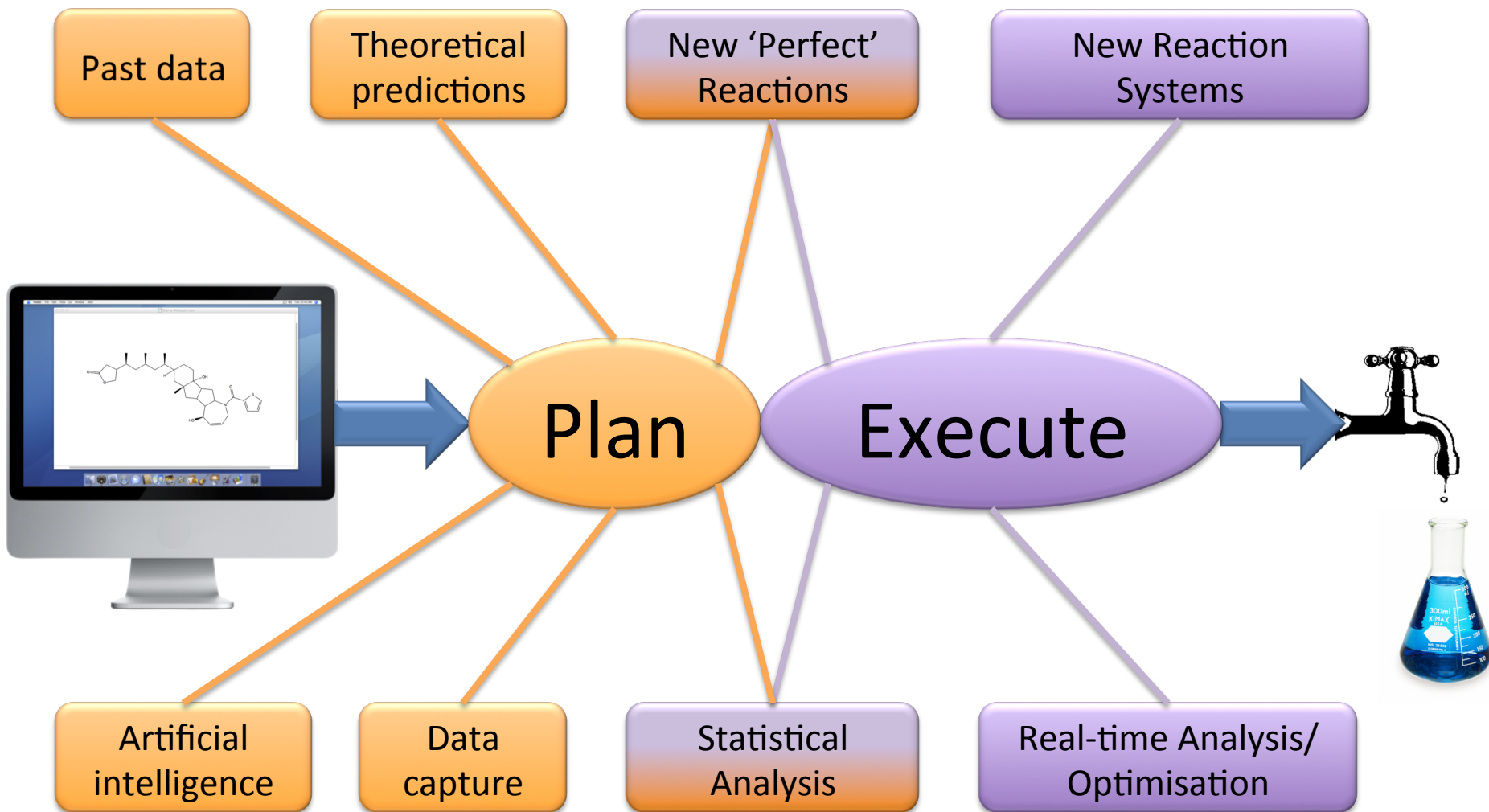
^aUniversity of Southampton, ^bUniversity of Leeds, ^cUniversity of Huddersfield, ^dBrunel University, ^eUCL, ^fSTB Associates, ^gRSC, ^hImperial College London, ⁱUniversity of Reading, ^jGSK, ^kManchester University, ^lPeakdale, ^mSyngenta, ⁿGillian Smith Associates, ^oAstraZeneca, ^pEPSRC, ^qCIKTN

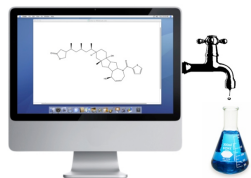


Dial-a-Molecule

An EPSRC Grand Challenge Network

How to solve it





Dial-a-Molecule

An EPSRC Grand Challenge Network

Roadmap: 2011.

(www.dial-a-molecule.org).

Key Barriers:

Making Synthesis Predictable

Smart Synthesis by Design

Sustainable Synthesis for a Sustainable Future



**Transforming synthesis,
enabling science**

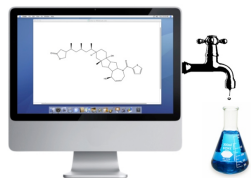


**Roadmap for
synthesis in the 21st Century**

EPSRC

Engineering and Physical Sciences
Research Council

www.Dial-a-Molecule.org



Dial-a-Molecule

An EPSRC Grand Challenge Network

Optimum Reaction and Route Design

Why does synthesis take so long?

Key problem is route selection.

Many reactions do not work!

- Forces route changes.

Others work, but require optimisation to give acceptable yields

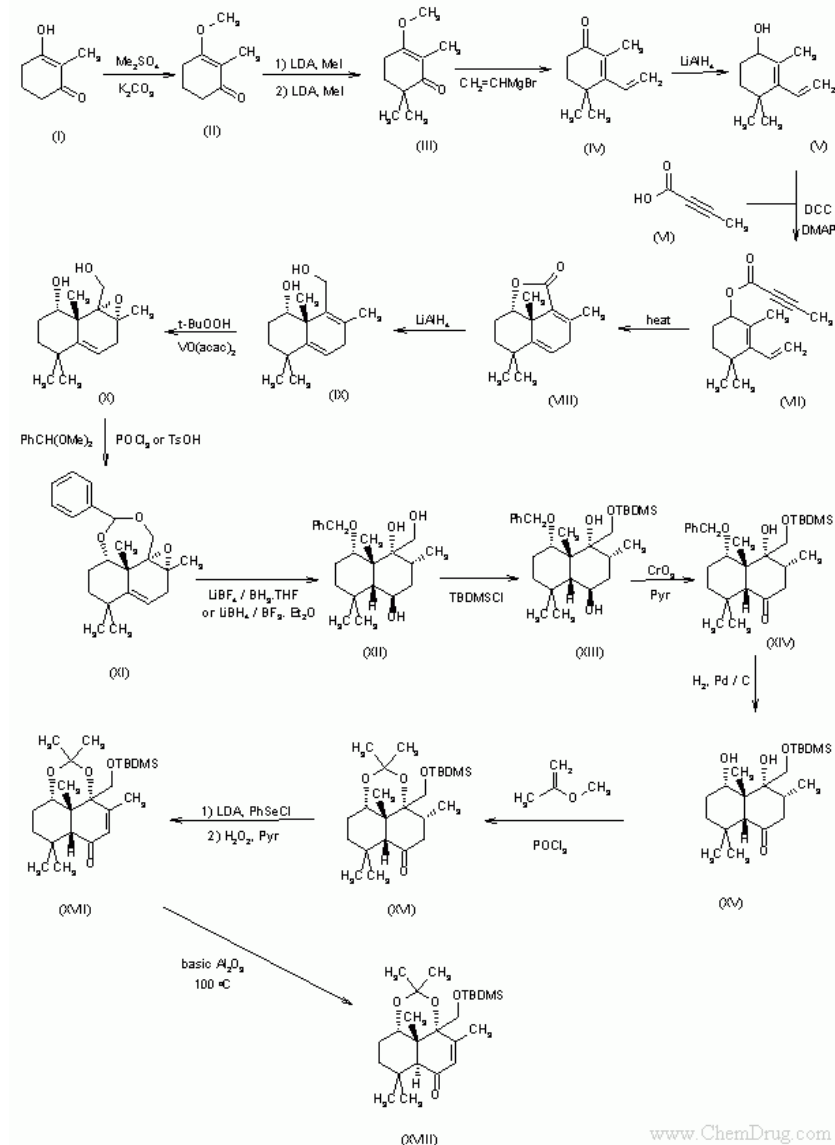
For moderate complexity targets:

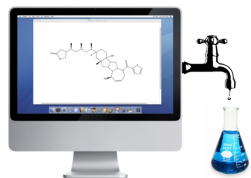
Number of reactions tried

No. in final synthetic route

Can be >100

Synthesis of a Forskolin intermediate



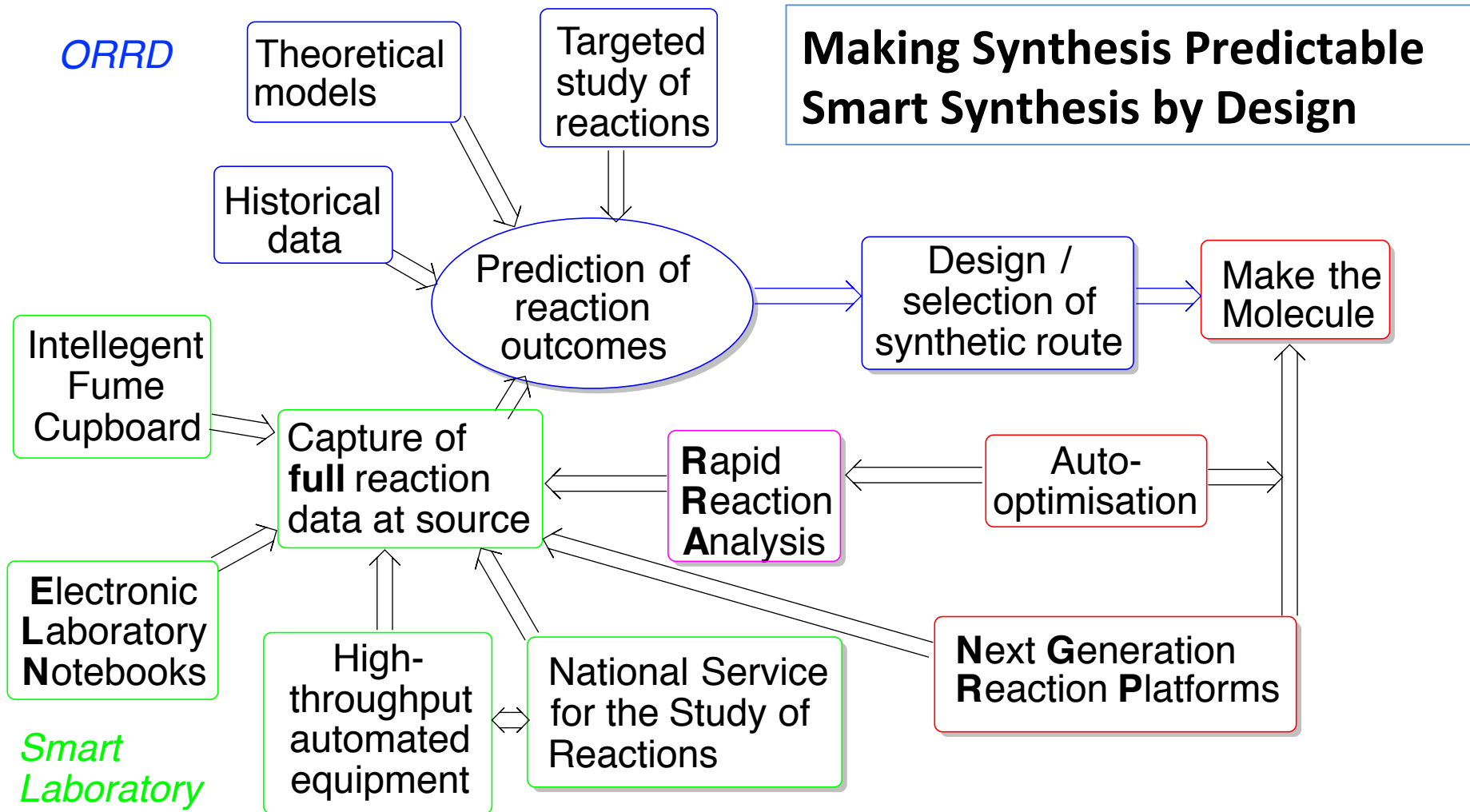


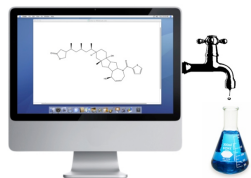
Dial-a-Molecule

An EPSRC Grand Challenge Network

Dial-a-Molecule Grand Challenge – key challenges									
Lab of the Future/Synthetic Route Selection				A Step Change in Molecular Synthesis		Catalytic Paradigms for Efficient Synthesis			
Optimum Reaction and Route Design	The Smart Laboratory	Next Generation Reaction Platforms	Rapid Reaction Analysis	Stepwise perfection (1000 Click Reactions)	Holistic approach to molecular synthesis	New reactivity: target-driven catalysis	Intervention-free synthesis	Engineering control through understanding	
Need for high quality reaction data	Electronic Laboratory Notebooks	Reactor Platforms	Equipment development into the lab.	Establish criteria for perfection	Tandem & telescoped reactions: generalising the concept	Efficient transformations across chemical space	Phase-separated catalysts	Rapid (self-)optimisation of reactions	
New ways to analyse reaction data to predict unknown outcomes	Automated and high throughput equipment for synthesis	Intelligent Feedback Control	Software development: automation of expert tasks.	Establish an inventory of reactions required & with potential for automation	Rational design & implementation of catalysts for multicomponent reactions	Complexity: building reactions	Mutually compatible catalysts	Full elucidation of catalytic mechanisms	
Planning of synthetic routes subject to constraints	The intelligent fume cupboard	Microfluidics and Lab-on-a-chip	Equipping academia – overcoming the cost barrier.	Reagentless "zero-waste" transformations (energy driven & catalytic)	Determining the reactions needed and priorities in targeting new chemical space	Sustainability: feedbacks	Switchable catalysts	Theoretical chemistry through understanding to prediction	
Theoretical prediction of reaction outcomes									
Active study and optimisation of reactions									

Making Synthesis Predictable Smart Synthesis by Design





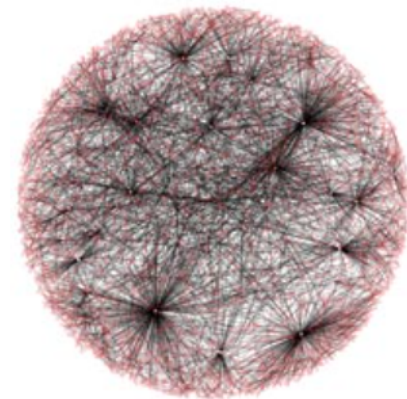
Dial-a-Molecule

An EPSRC Grand Challenge Network

Predicting reaction outcomes

How big is reaction space?

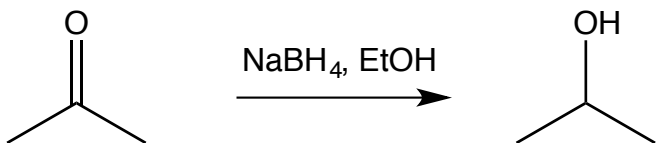
“The size of the chemical space that is of interest to drug developers is estimated to lie between 10^{18} and 10^{200} compounds”
 10^{60} stable organic compounds with MW<500 is figure often given.



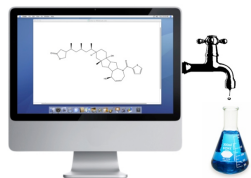
Reaction space is connections between molecules. $10^?$ x number of molecules?

Each reaction can be carried out in many ways, with many different conditions ($\times 10^?$)

Functional group approximation. Assumes that particular arrangement of atoms will always react the same way independent of the rest of the molecule



Partially true for some reactions,
Much less so for others (e.g. skeletal rearrangements)



Dial-a-Molecule

An EPSRC Grand Challenge Network

Predicting reaction outcomes

What do we know?

Reaxys, CASReact, SPRESI, PCD... Perhaps 30×10^6 unique reactions

All of chemistry so far – perhaps 10^8 reported reactions.

Each with some information (% yield, reagent, solvent, temperature, time, work-up...)

Potential to include calculated data

How well can we do?

Computer Aided Synthesis design programs

LHASA (hand coded rules)

ARCHEM (rules from database)

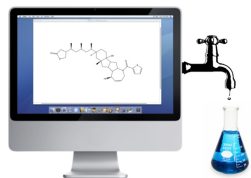
ICSynth (rules from database)

Chematica (hand coded rules)

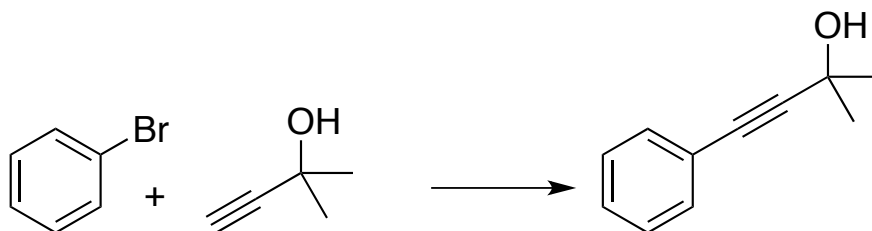
And many other attempts...

Sparse
Incomplete
Inconsistent
Error strewn

Challenge for Big Data – how to do better?



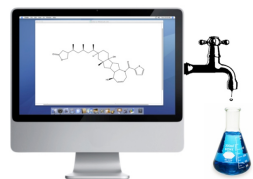
Using full experimental data



Database: $(\text{PPh}_3)_2\text{PdCl}_2$, CuI, NEt_3 , 80°C , 12h

Publication: 2-methylbut-3-yn-2-ol (3.21 g, 38.16 mmol) and 15 mL NEt_3 were added to a mixture of bromobenzene (5.00 g, 31.8 mmol), CuI (181 mg, 954 μmol), PPh_3 (996 mg, 3.80 mmol), and $\text{PdCl}_2(\text{PPh}_3)_2$ (1.34 g, 1.90 mmol) under nitrogen. The reaction mixture was heated to 80°C for 12 h. The reaction was quenched with water (100 mL), extracted with dichloromethane (DCM) (3 x 100 mL). The combined organic phase was dried over MgSO_4 , filtered off, and concentrated under reduced pressure. The product was obtained as a yellow oil by column chromatography (silica, hexane/ethyl acetate (4:1), $R_f=0.33$) (yield: 4.84 g, 95percent).

Reality: Many, many variables, most not even recorded.
Eg S88 batch process description may occupy 80 pages!



Dial-a-Molecule

An EPSRC Grand Challenge Network

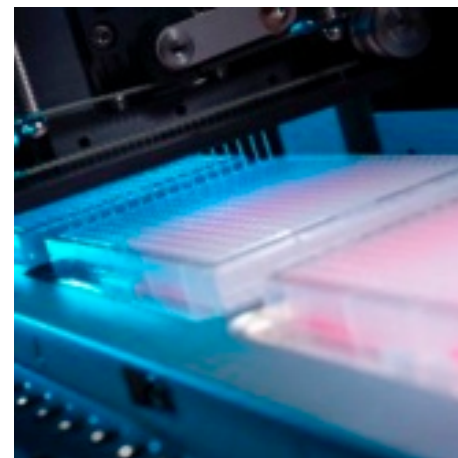
Predicting reaction outcomes

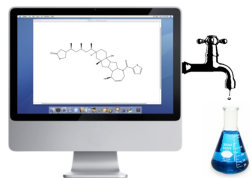
Using full experimental data

Much more data is coming!

Electronic Laboratory Notebooks,
Automated experimentation,
High throughput parallel experimentation.
In situ spectroscopy

How to make effective use of the flood of data produced?





Dial-a-Molecule

An EPSRC Grand Challenge Network

Big Data and the Dial-a-Molecule Grand Challenge

More effective use of existing data to predict reaction outcomes

Making use of the flood of detailed data starting to be generated.

Determine 'best' synthetic routes.

Acknowledgements.

EPSRC for funding

Dial-a-Molecule coordinators: Bogdan Ibanescue, Susanne Coles, Kelly Kilpin

Dial-a-Molecule Steering group.