



optibrium

Automatic Generation of Compound Ideas to Guide Optimization

Innovative Lead Optimization and Candidate Selection by *in Silico* Synthesis and ADMET Prediction. December 3rd 2015

Matthew Segall, CEO, Optibrium Ltd.

Overview

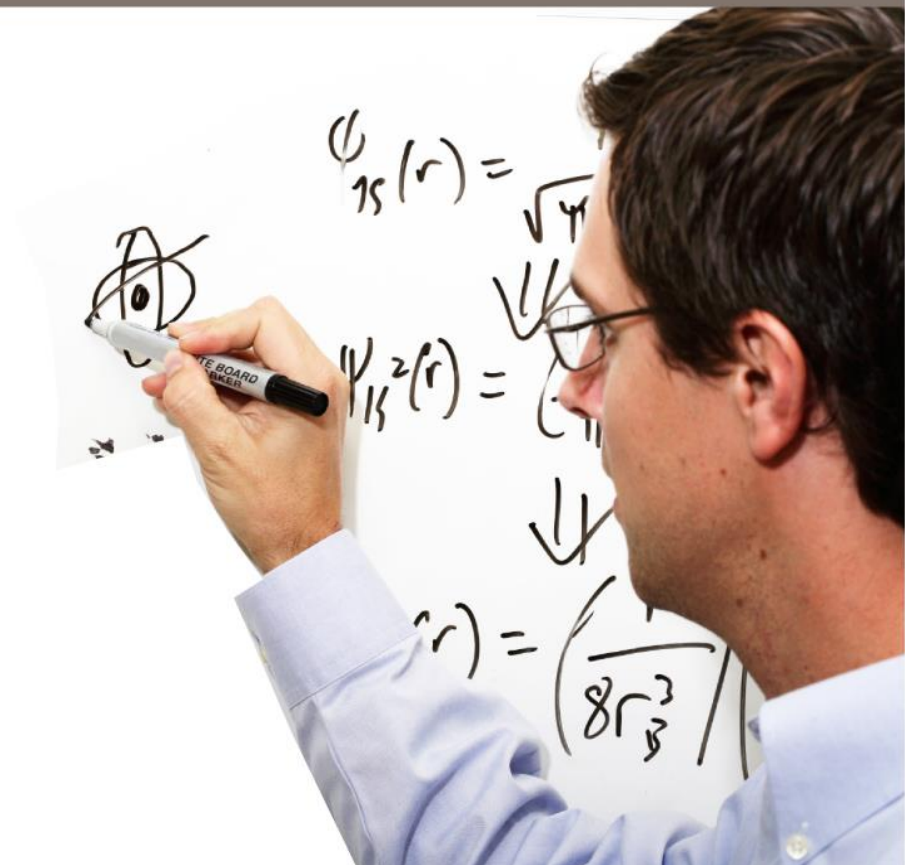
- Automatic compound idea generation
 - *De novo* design
- Matched Series Analysis
 - SAR Transfer
 - Matsy™
- Medicinal chemistry transformation rules
- Conclusions

Automatic Generation of Compound Ideas

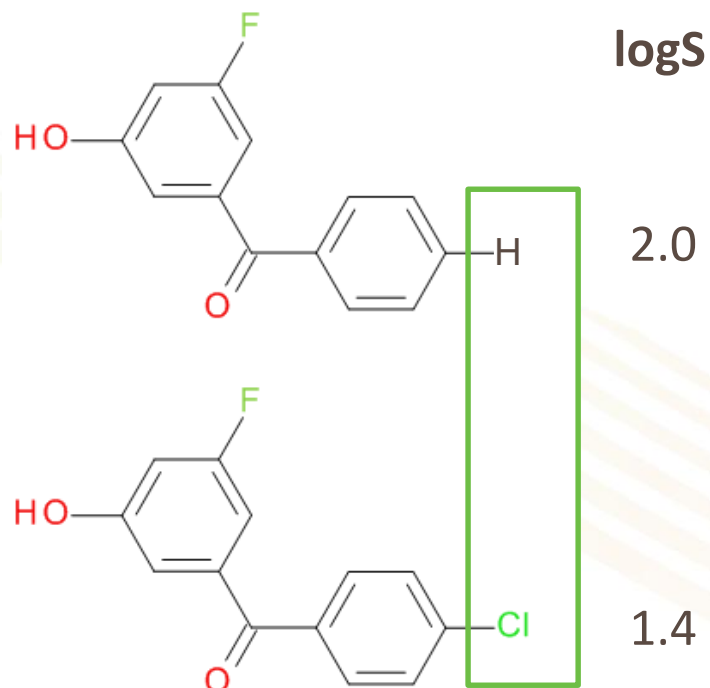
Objectives

- Traditionally have a scarcity of data
 - Time consuming and expensive to generate data
 - Easy to think of more ideas than can be synthesised and tested
 - Manually analyse all of the available data to select compounds
- With predictive models and MPO, it is easy to evaluate very large numbers of ideas
 - More ideas than one scientist can design and draw
- Generate new ideas to stimulate exploration of chemistry
 - *In silico* analysis helps to prioritise ideas for detailed consideration
- But... compound ideas must be **relevant and accessible**
 - First generation of *de novo* design methods tended to generate poor quality, non-synthesizable structures

Matched Series Analysis



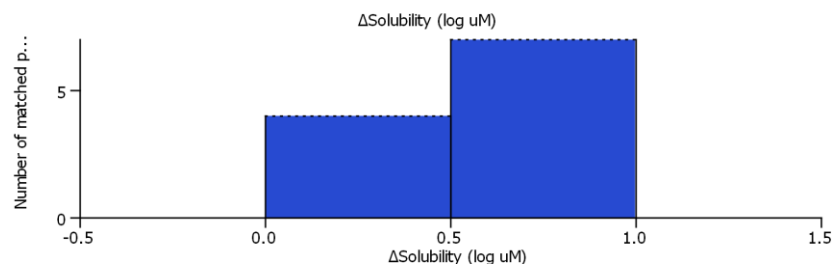
Matched Molecular Pairs



Matched Pair

Two compounds that differ only by a small replacement at a single position

From	To	Count	Δ Solubility (log uM)
		15	0.5927
		11	0.5642
		6	0.5443
		15	0.54

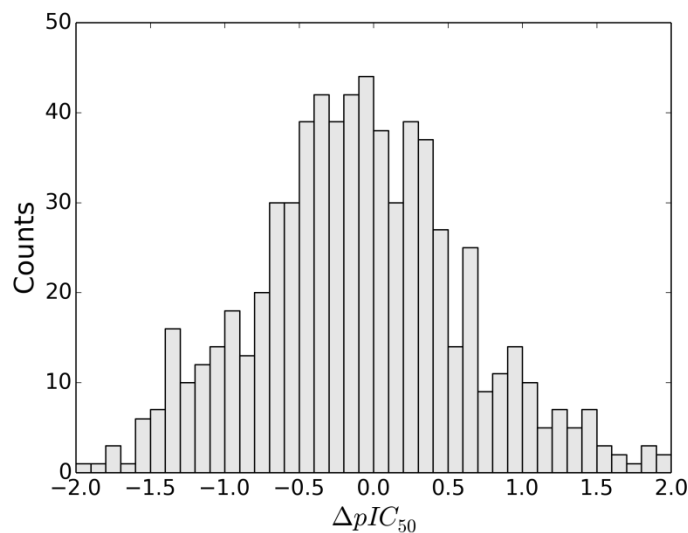


Limitations of Matched Molecular Pairs

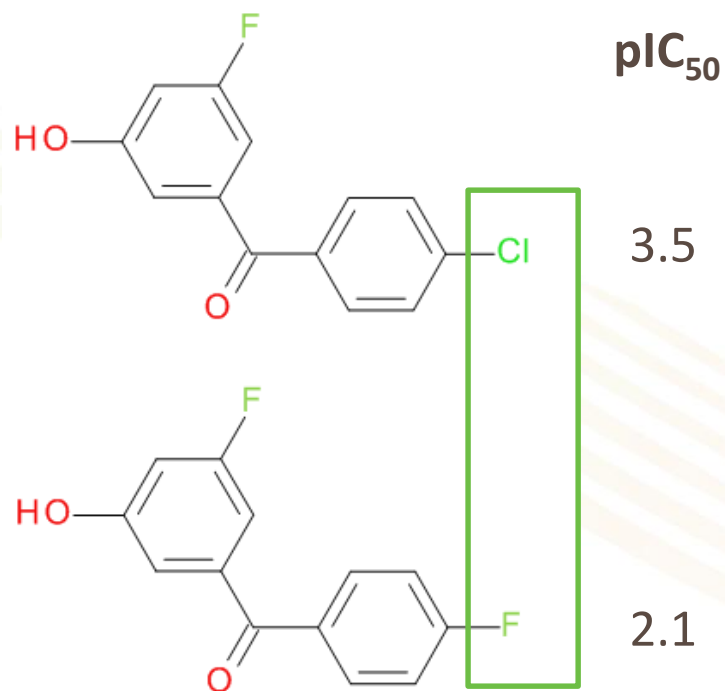
- If we look at all examples of a matched pair across diverse chemistry and target classes, they provide little information on target activity

Distribution of change in pIC_{50}

replacing CCCC* with CC*



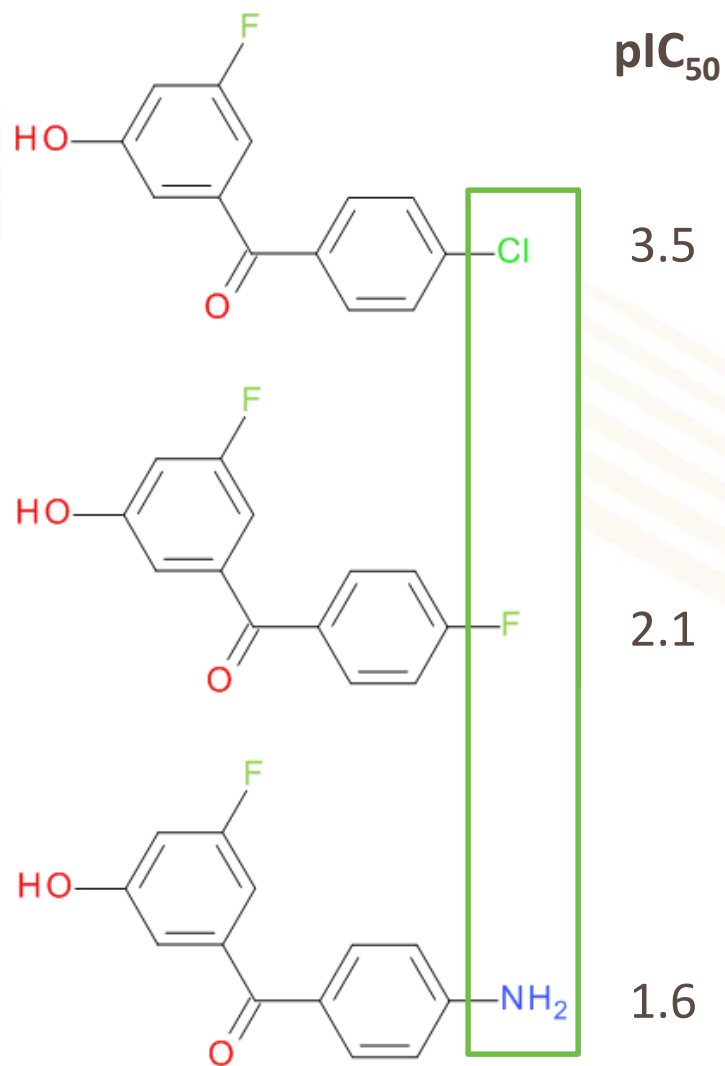
Matched Series Principles



Matched Pair
(Matched Series of length 2)

[CL > F]

Matched Series Principles



Matched Series of length 3

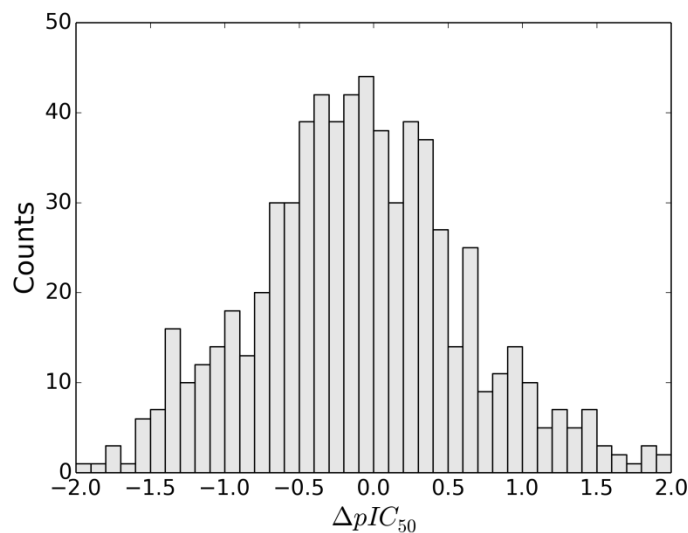
[CL > F > NH₂]

Matched Series Principles

- Longer series that match an observed order of activity improve our ability to make relevant predictions

Distribution of change in pIC_{50}

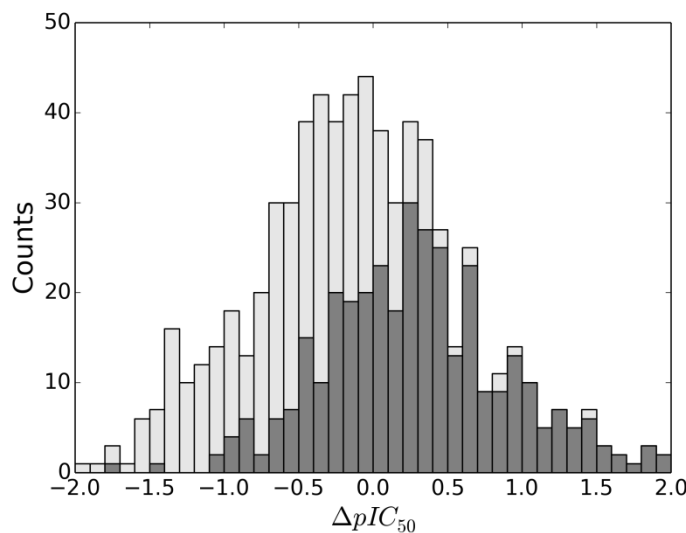
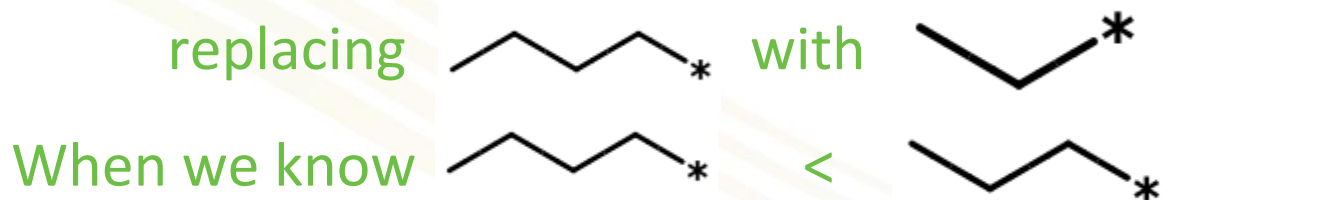
replacing CCCC* with CC*



Matched Series Principles

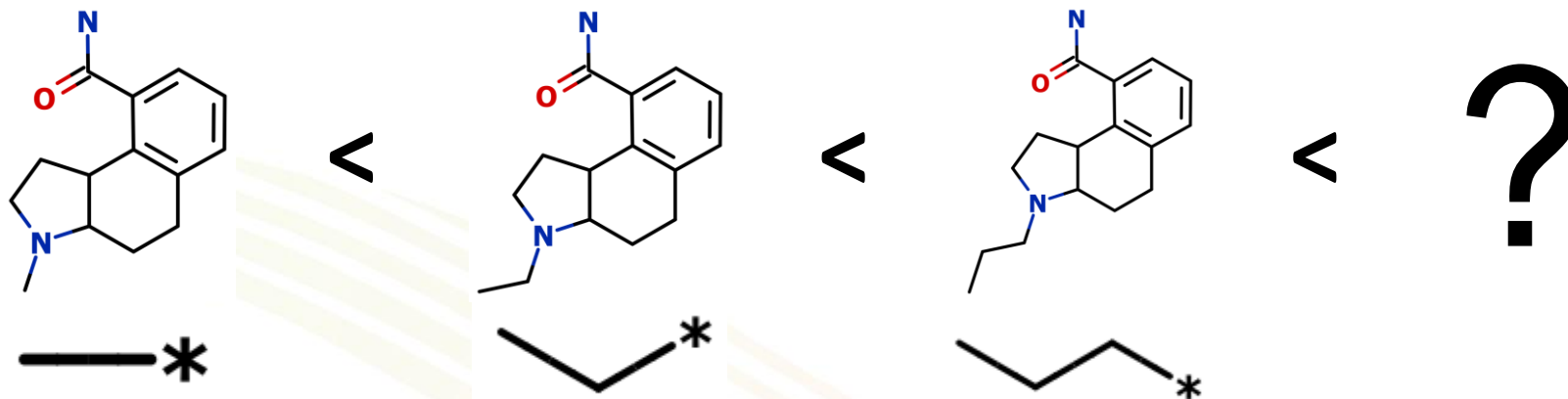
- Longer series that match an observed order of activity improve our ability to make relevant predictions

Distribution of change in pIC_{50}



Matched Series Principles

Look up existing examples from a database

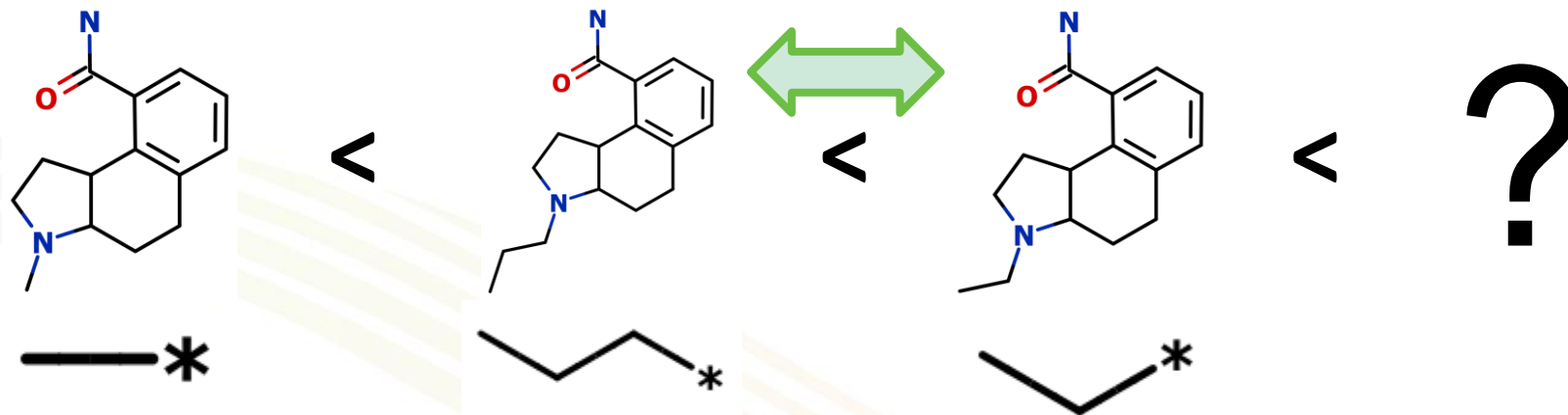


Matched series suggest
larger hydrophobes will
be more active

	Structure	RGroup	% that improve	Observations
1			90.5	21
2			72.9	59
3			68.8	32
4			61.5	26
5			59.1	44
6			59	39

Matched Series Principles

Look up existing examples from a database




















Smaller branched hydrophobes suggested by this series

	Structure	RGroup	% that improve	Observations
1			39.1	23
2			35	20
3			34.6	26
4			32.7	107
5			28.1	32
6			27.3	22

Matched Series Principles

















- Matsy™

- Shorter matched series (≥ 3)
- High number of observations (≥ 20)
- Key metric: % of observed series that improve activity

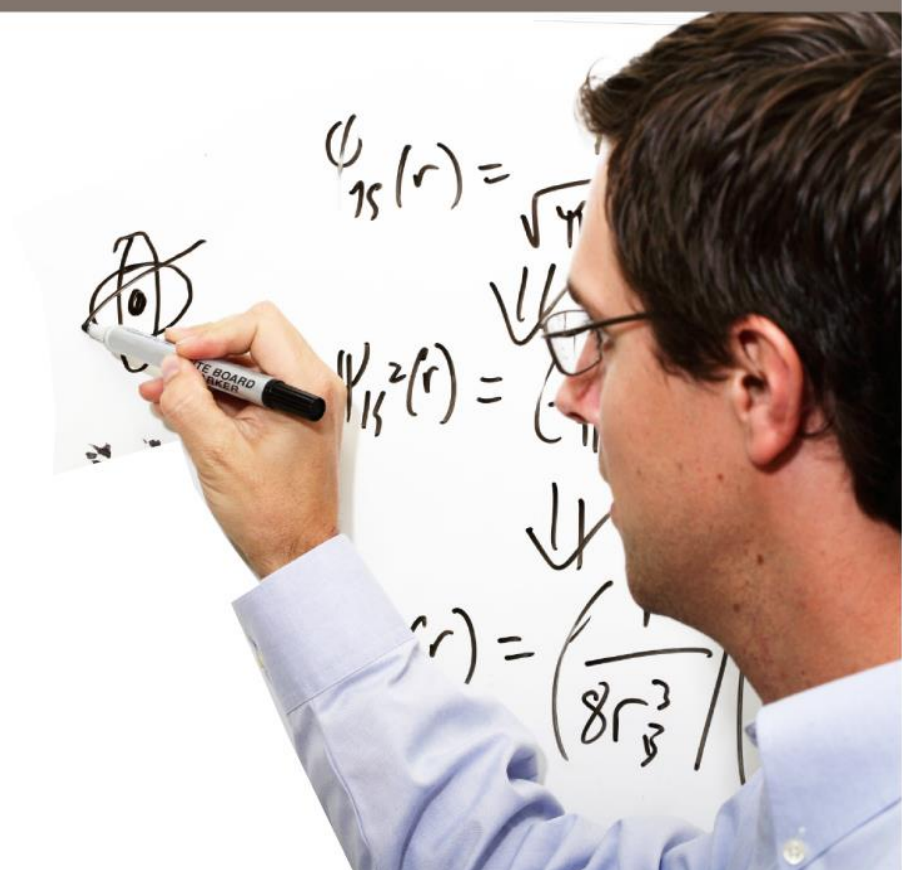
Scaffold	Target			H	
			8.9	8.79	8.69
	MAP kinase p38 ...	7.85	7.8	7.72	6.52
	Epoxide hydratase	6.85	6.35	6.29	5.72
	Cyclin-depende...	6.23	6.22	6.21	5.91
	Serine/threonine...	8.22	7.72	7.68	7.29
	Cholecystokin...	7.8	7.52	7.26	7.19
	Unchecked	6.8	6.68	6.44	6.3
	Tyrosine: protein...	7.89	7.07	7.05	6.57
	Epidermal growt...	7.15	6.88	6.6	6.07
	Unchecked	6.74	6.22	6.19	6.01
	Unchecked	5.27	5.04	5.01	4.99
	FK506 binding pr...	5.92	5.51	5.1	5.09
	PI4-kinase beta s...	6.06	5.23	5.09	4.89
	Cyclooxygenase-1	4.29	4.22	4.2	4.1

- SAR Transfer

- Longer matched series (≥ 8)
- Few observations
- Key metric: Correlation (R^2) between observed series

Scaffold	Target	Correlation												
			9.1	8.96	8.09	7.7	7.49	7.45	7.26	6.82	6.56	6.22	5.63	
	Serotonin 7 (5-H...	0.8676	7.72	7.07	6.72	7.04	6.35	6.11	6.03	5.96	6.11	5.82	5.66	6.03
	Serotonin 7 (5-H...	0.7939	7.92	6.92	6.89	7.13	6.57		6.33	6.46	6.38	6.7	5.92	6.21
	Serotonin 1a (5-...	0.8095	7.24	7.13	6.78	6.69		6.48	5.57	5.07	6.09		5.72	

Medicinal Chemistry Transformation Rules



Generating Compound Ideas

Applying Med. Chem. 'Transformation Rules'

- Compounds generated must 'make sense' from a medicinal chemistry perspective
- Apply 'transformation rules', derived from medicinal chemistry experience, to initial compound(s)*
 - Library of >200 transformations, generate ~180 new compounds per input
 - >94% of structures generated acceptable to med. chemists
 - Not only functional group replacement but also framework

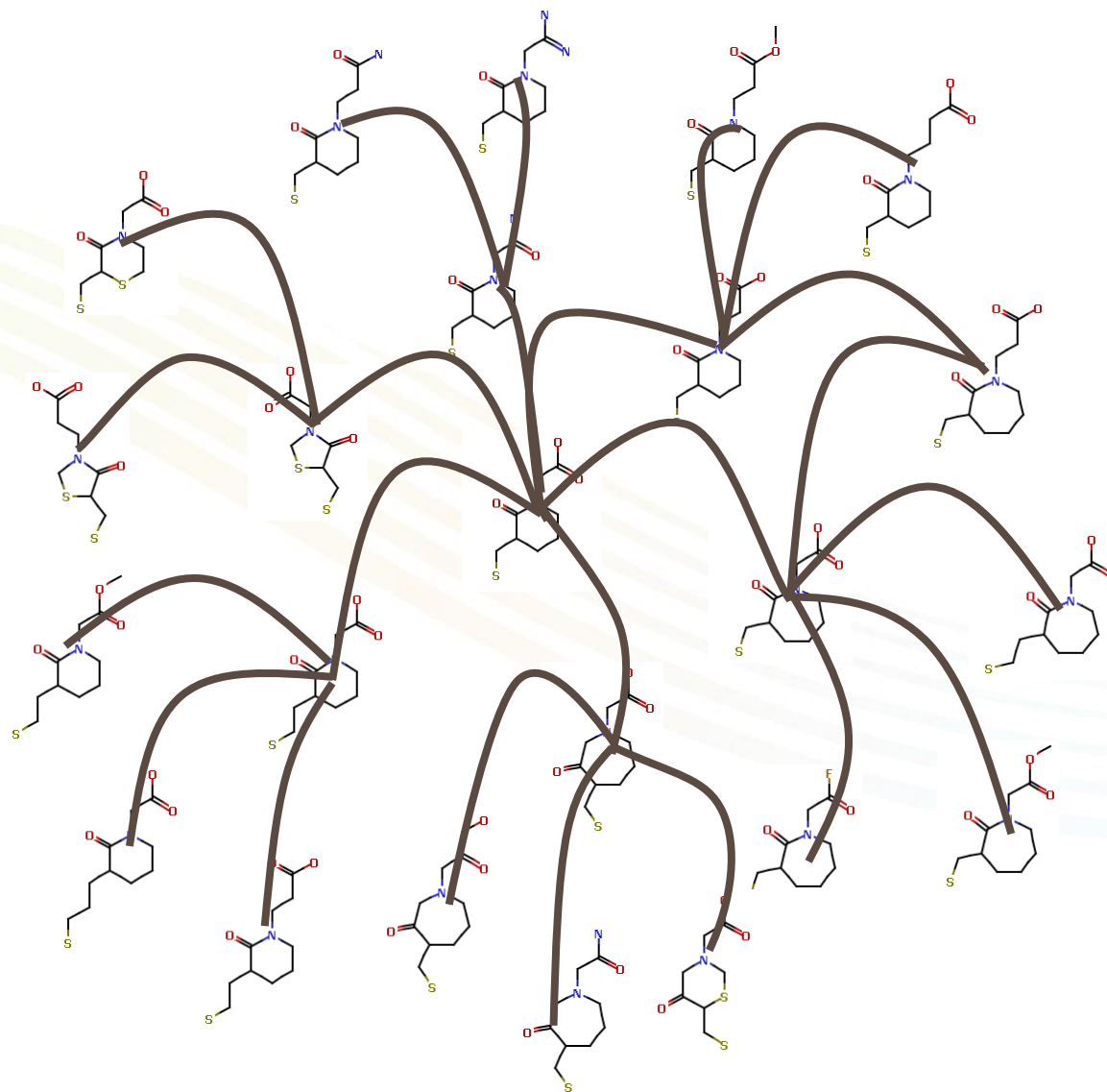


Ring identification and addition:
e.g. benzene to indole

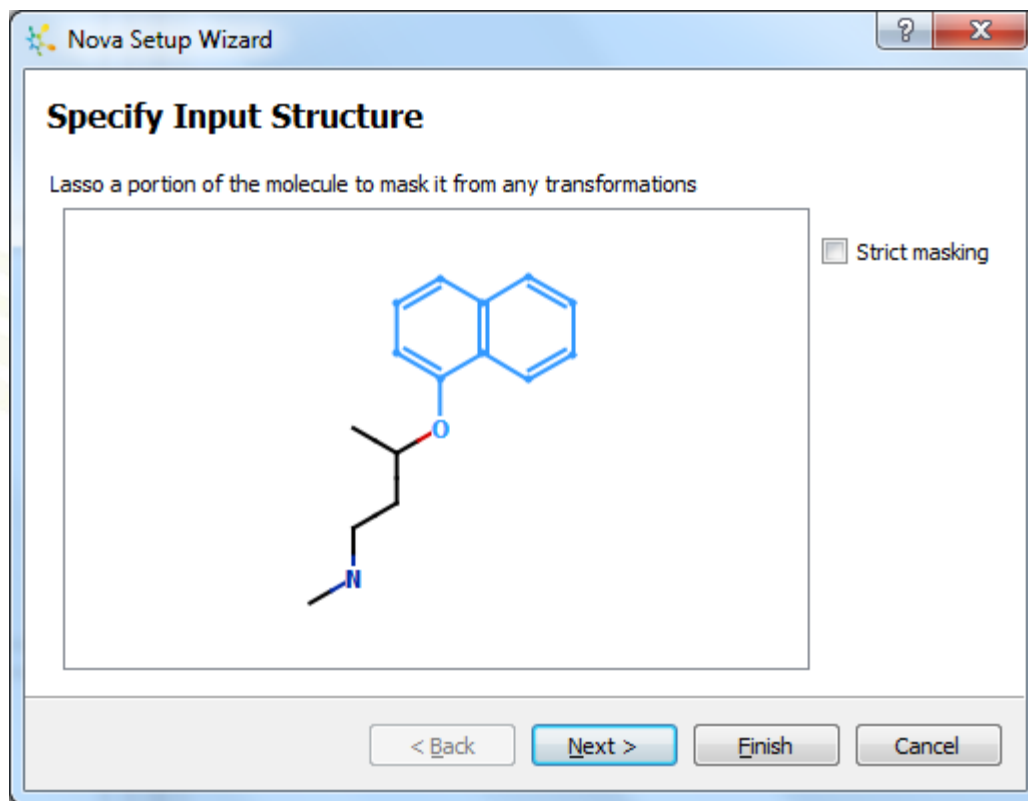
*Stewart *et al.* Bioorg. Med. Chem. (2006) **14** p. 7011

*Segall *et al.* J. Chem. Inf. Model. (2011) **51** pp. 2967-2976

Exponential Growth!

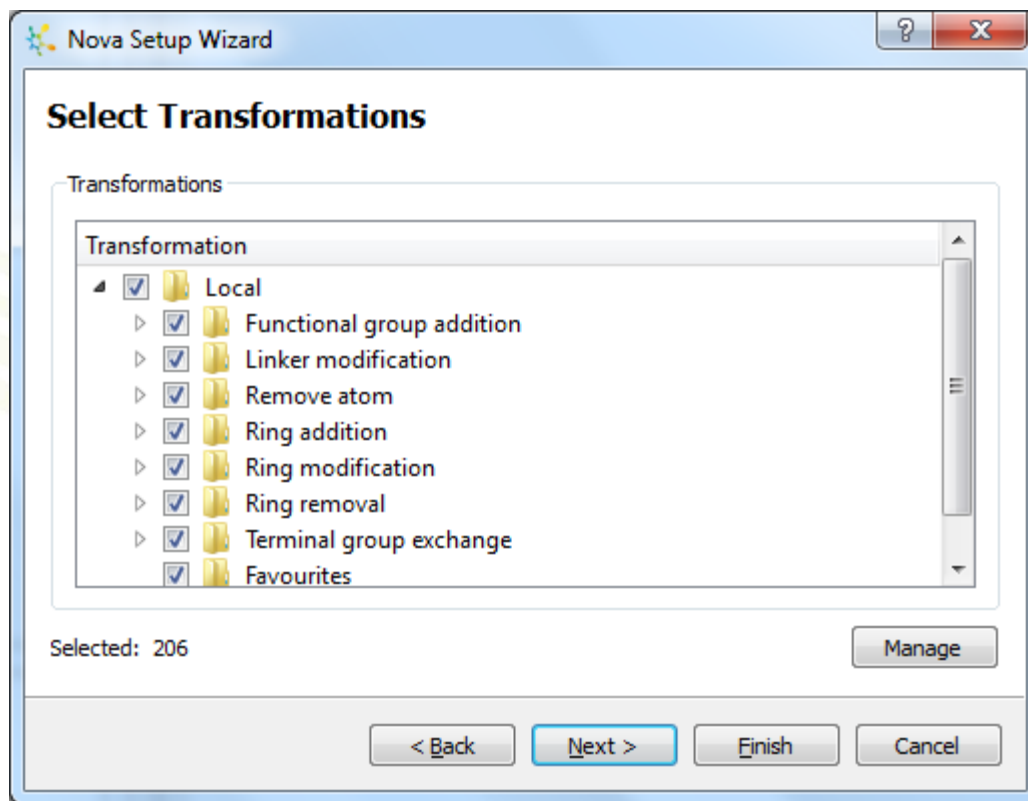


Controlling the Process



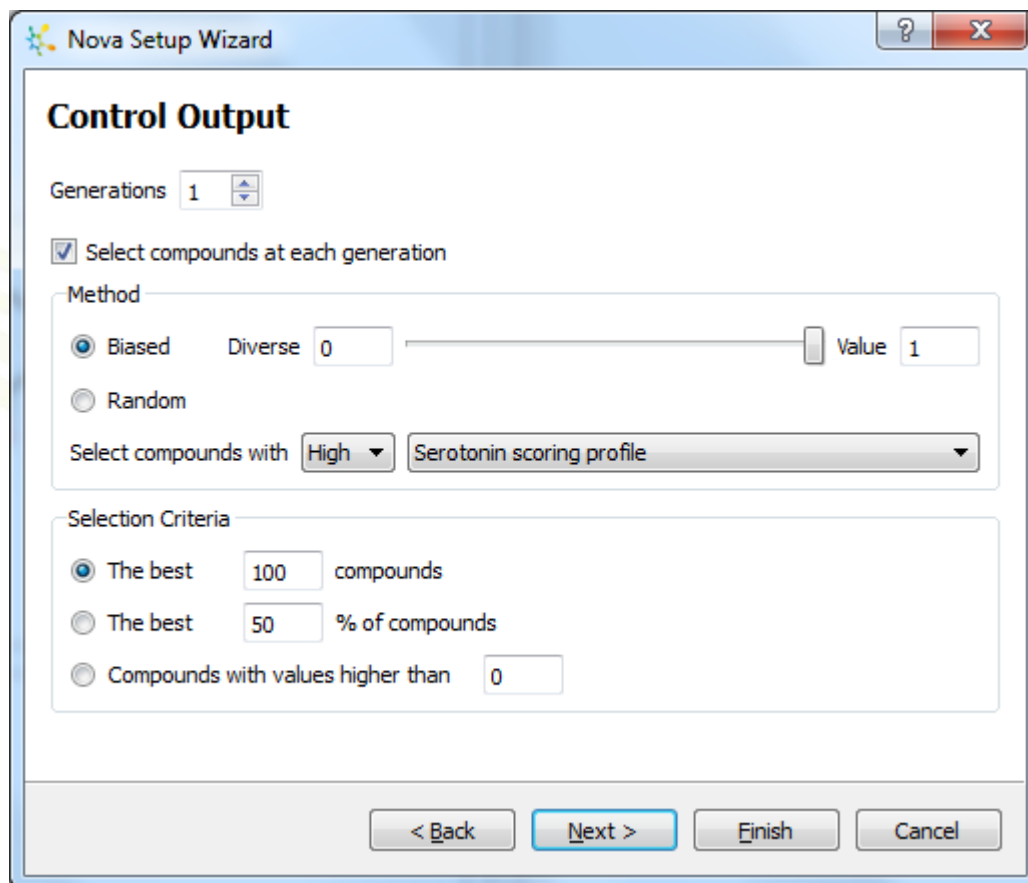
- Specify initial structure
- A region can be selected to be fixed (no changes allowed)

Controlling the Process



- Select transformations to apply
- Transformations can be managed for specific objectives

Controlling the Process



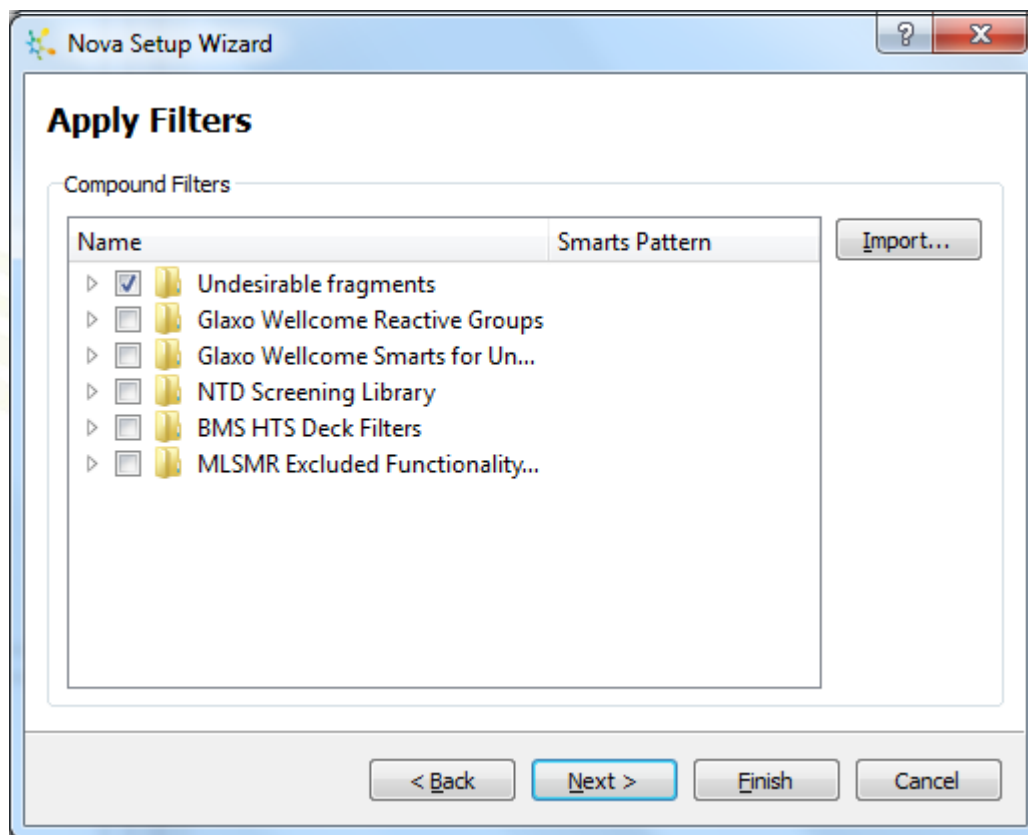
The screenshot shows the 'Control Output' window of the 'Nova Setup Wizard'. The window has a title bar with a question mark icon and a close button. The main content area is titled 'Control Output' and contains several settings:

- Generations:** A numeric input field set to '1'.
- Select compounds at each generation:** A checked checkbox.
- Method:** A section containing two radio buttons: 'Biased' (selected) and 'Random'. The 'Biased' option has a 'Diverse' slider set to '0' and a 'Value' input field set to '1'.
- Select compounds with:** A dropdown menu set to 'High' and a dropdown menu set to 'Serotonin scoring profile'.
- Selection Criteria:** A section containing three radio buttons: 'The best' (selected), 'The best', and 'Compounds with values higher than'. The first 'The best' option has a numeric input field set to '100' and the text 'compounds'. The second 'The best' option has a numeric input field set to '50' and the text '% of compounds'. The third option has a numeric input field set to '0'.

At the bottom of the window are four buttons: '< Back', 'Next >', 'Finish', and 'Cancel'.

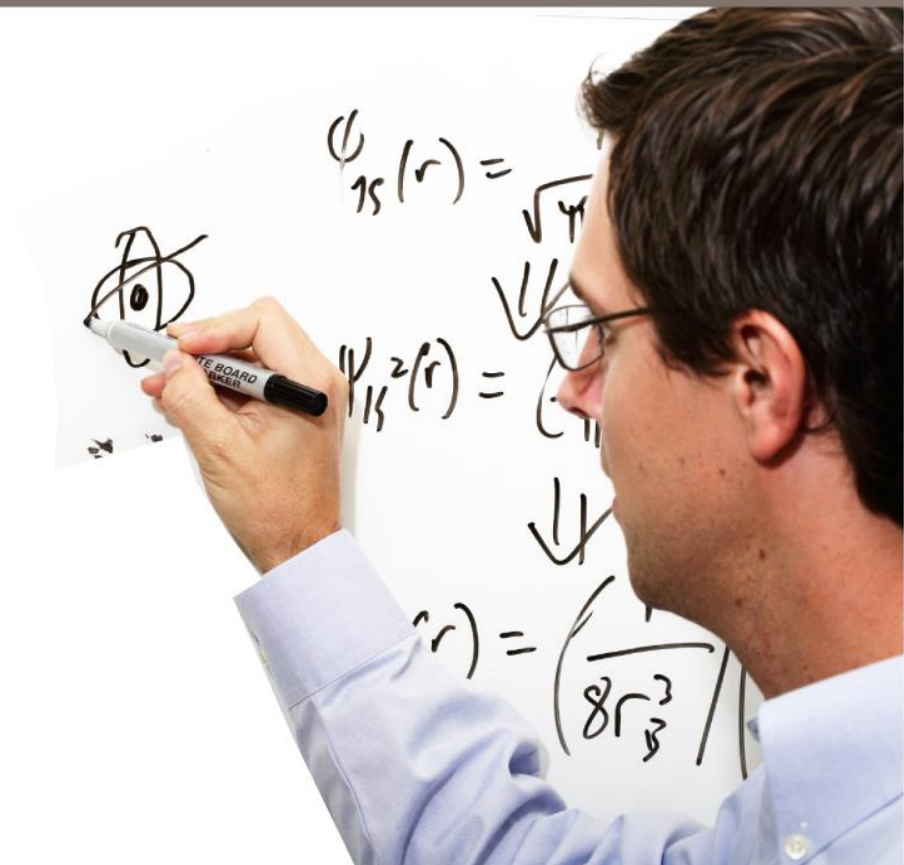
- Apply multiple generations of transformations
- Bias selection in favour of property, score or diversity

Controlling the Process



- Apply substructure filters to results

Extending the Database of Transformations

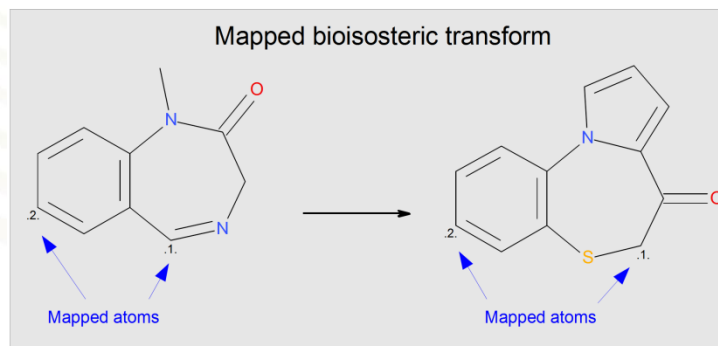
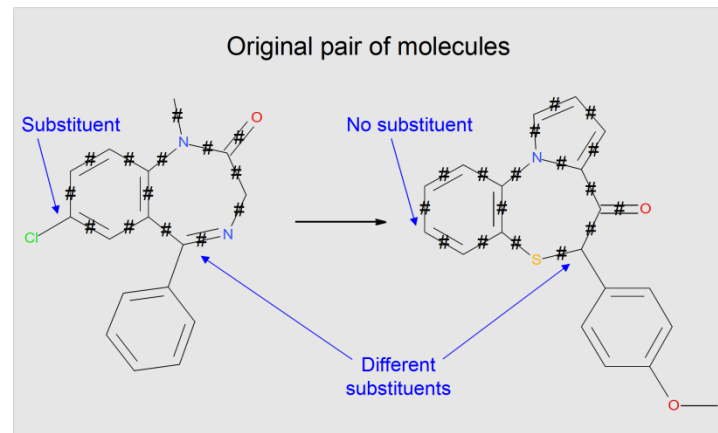


- Database of >32,000 pairs of compounds representing optimisation steps in chemistry projects*
 - Manually curated from the literature by Dr István Ujváry
- Library of transformations defined from pairs of molecules
- Addresses additional questions, e.g.
 - What modifications have been successfully applied to similar compounds?
 - Identify potential lead-hopping strategies
 - Search for patent protection/busting strategies

* Ujváry and Hayward, in Methods and Principles in Med. Chem. (Vol. 54), N. Brown (ed)

Generating Transformations

- Substructure replacement manually defined
- Handle different substitutions
 - Matched using heuristics
- Avoid promiscuity
 - Constraints on application
- 25,351 transformations generated (~80% success)



SMIRKS generated

```
([C;$ (C1@c2@c (@N (!@C) @C (=; !@O) @C@N=1) @c@c@c@c2) :1] 1c2c  
(N ([CH3]) C(=O) [CH2]N=1) [cH] [cH] [c:2] [cH] 2)>>(n21-c3c  
(S [C;$ (C1 (@S@c3@c (-; @n2@c (@C1=; !@O) @c@c@c@c2) @c@c@c@c3)  
!@ [H]) :1] (C (c1 [cH] [cH] [cH] 2)=O) [H]) [cH] [c:2] [cH] [cH] 3)
```

Transformation Library

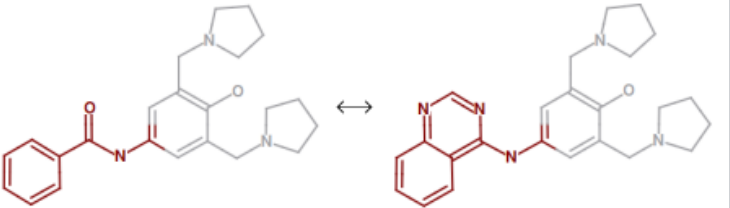
Transformation Manager

Search:

Transformation

- ★ Acylsulfamide to sulfonylurea
- ★ Formamide to thioamide
- ★ Styrylformamide to benzylidenemalo...
- ★ (Z)-alkene to amide
- ★ Amide to cyanoamidine
- ★ Amide to nitrile
- ★ Amide to vinyllogous amino acid deri...
- ★ Amide to sulfonomethyl
- ★ Amine to formamidine
- ★ Amine to aminothiazoline
- ★ Amine to aminotriazole
- ★ Amine to aminopyrimidinol
- ★ Amine to alkylamine
- ★ Amine to imidazoline-2-thione
- ★ Amine to imidazoline-1
- ★ Amide to imidazoline
- ★ Amine to dimethyltriazene
- ★ Amide to amidine
- ★ Amide to imidazolidone
- ★ Amide to carbamate
- ★ Amine to alkylsilane
- ★ Quaternary amine to sulfone
- ★ Quaternary amine to pyridinium
- ★ Benzamide to aminoquinazoline
- ★ Amine to nitroguanidine
- ★ Amine to cyanoguanidine
- ★ Benzamide to naphthalimide
- ★ Amide to oxime
- ★ Amide to N-acylpyrrole
- ★ Lactam to cyanoguanidine
- ★ Amide to sulphinamide
- ★ Amine to ethylenediamine
- ★ Amine to imidazoline-2
- ★ Amide to triazole

Benzamide to aminoquinazoline



SMIRKS: [[c;\$(!@N!@C(!@c1@c@c@c@c1)=;!@O):1][NH]C(c1[ch][ch][ch][ch][ch]1)=O]>>(c21c([NH])[c;\$(!@N!@c1@c2@c(@n@c@n1)@c@c@c@c2):1)n[ch]nc1[ch][ch][ch]

Bioster No: 282 ID code: AMI055

Names & Key Phrases: Antiarrhythmic

References: Glowka M L et al, J Med Chem, 34() p. 2678, 1991
Stout D M et al, J Med Chem, 28() p. 295, 1985
Wustrow D et al, J Med Chem, 41() p. 760, 1998
Cummins J G et al, Bioorg Med Chem Lett. 14() p. 5389, 2004

Import...
Export...
New Group
Delete

OK Cancel

Conclusions

- Matched Series Analysis
 - Improves on matched molecular pair analysis to generate more relevant suggestions to improve target activity
- Medicinal Chemistry Transformation Rules
 - Based on medicinal chemistry experience to generate relevant and accessible compounds ideas
- Combine with predictive models and MPO to improve *balance* of properties
- Please see:
 - M.D. Segall Expert Opin. Drug. Discov. (2014) 9(7), pp. 803-817
 - O'Boyle *et al.* Drug Discovery World, Fall 2015, pp. 55-59