Capturing and Applying Knowledge to Guide Compound Optimisation

Streamlining Drug Discovery – 31st May 2018

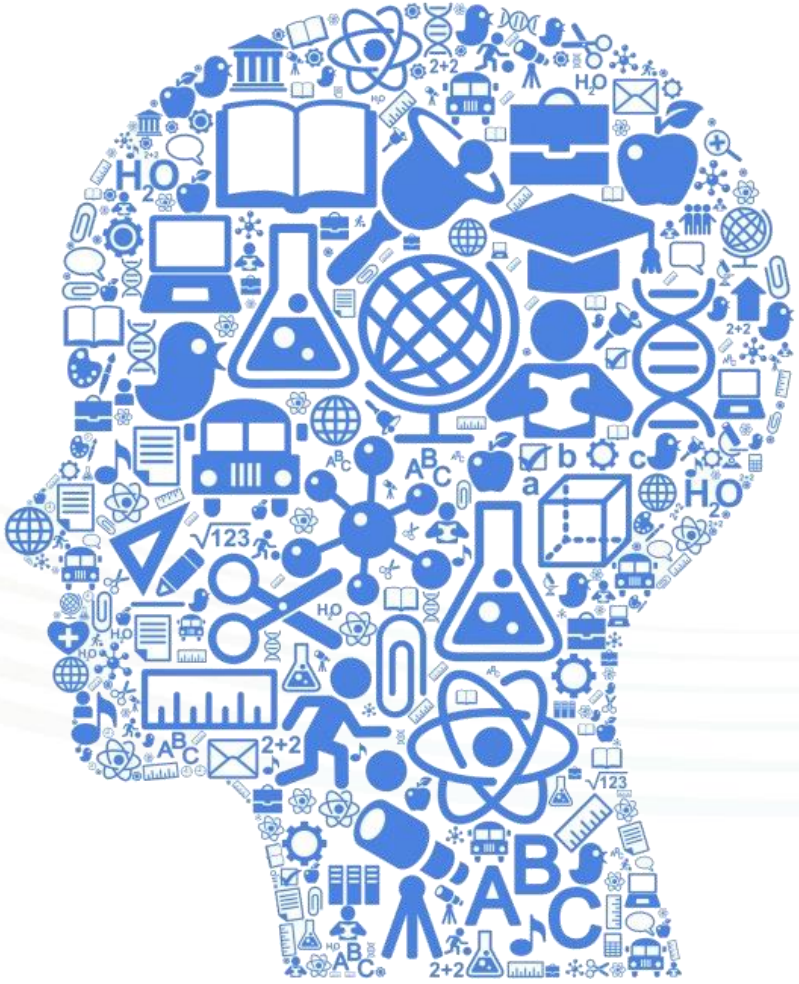
Edmund Champness, Matthew Segall, Peter Hunt and Tamsin Mansley

Overview

- Why capture and apply knowledge computationally?
- Capturing and applying different types of knowledge
  - Chemistry
  - Structure activity relationships
  - Project objectives
- Example application
  - Optimisation of high quality, selective Anagliptin analogues
- Conclusions

Why Capture Knowledge Computationally?

- Different viewpoints
  - Drug discovery is a multi-disciplinary field
  - Knowledge transfer
- Limited experience
  - Even within our own field, we each have limited experience
- Limited memory
  - We can only easily remember a fraction of what we have experienced
- Computers have virtually unlimited capacity and perfect recall



Why Apply Knowledge Computationally?

- Being able to search large databases of information/knowledge is useful, but limited
- Computers can apply stored knowledge much faster than a person
- Even scientists have hidden biases*
  - We choose which knowledge to 'believe'
- Computers can explore ideas more quickly and rigorously than a person

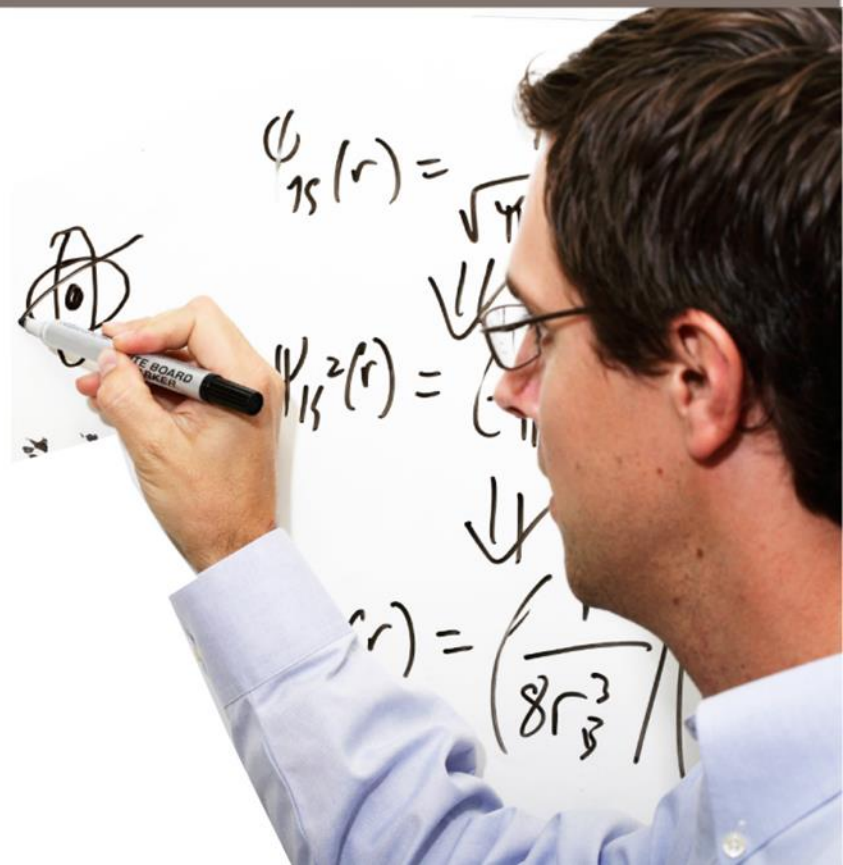
Knowledge is like paint… It does no good until it is applied

- Doe Zantamata

*Chadwick and Segall, Drug Discov. Today (2010) **15**(13/14), pp. 561-569

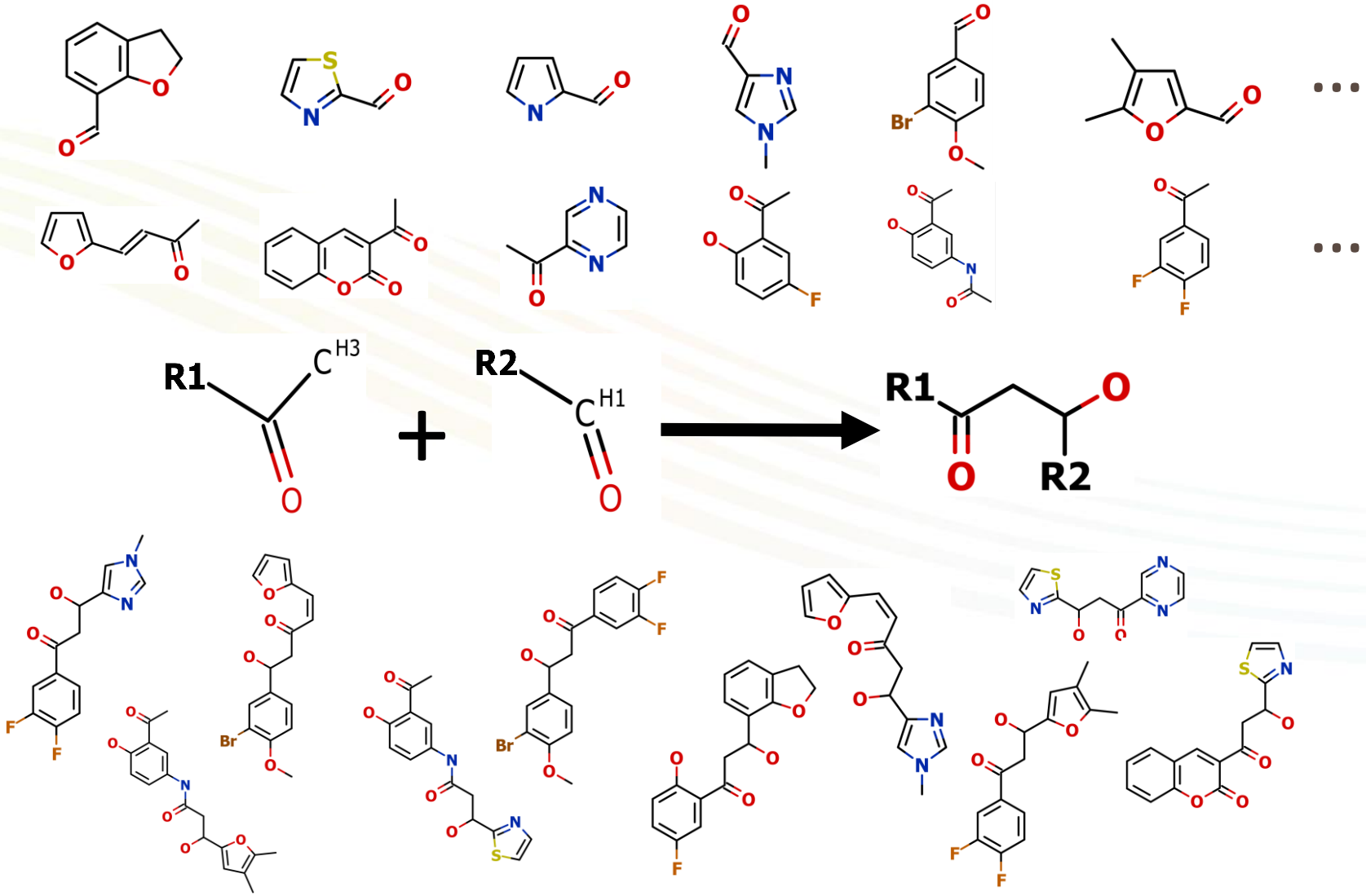
Capturing and Applying Knowledge
Chemistry – What might we make?

optibrium



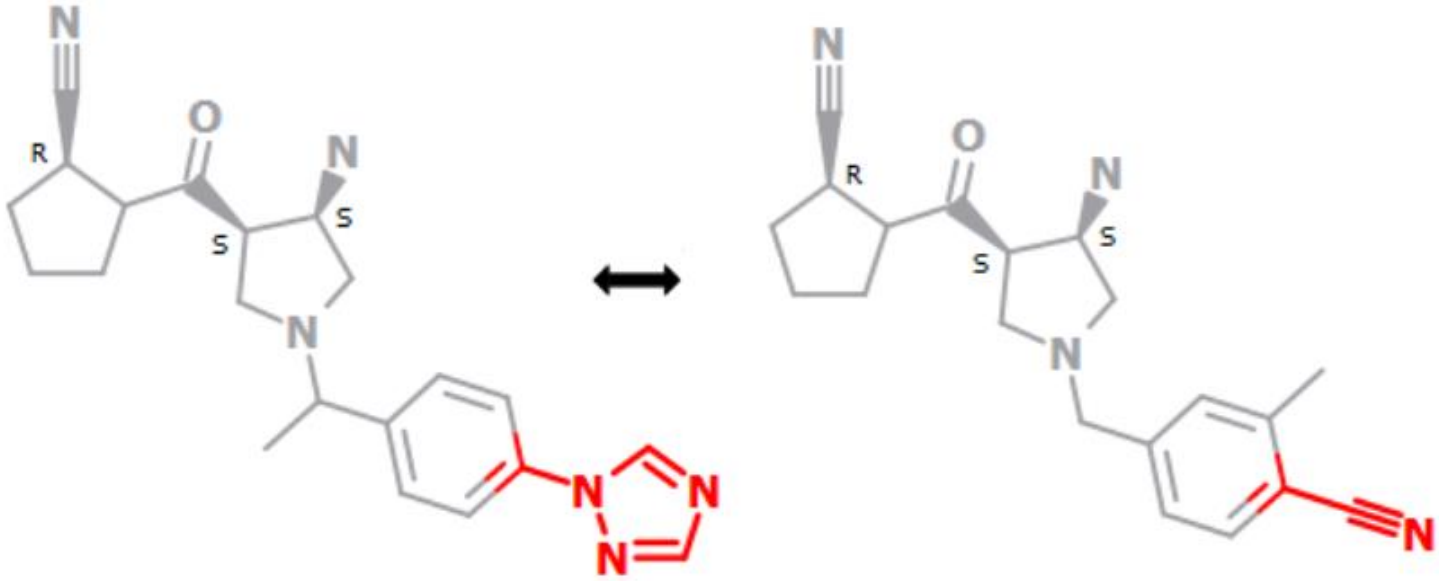
Chemistry – What could we make?

Reactions and Reagents



Chemistry – What might we make?
Transformations

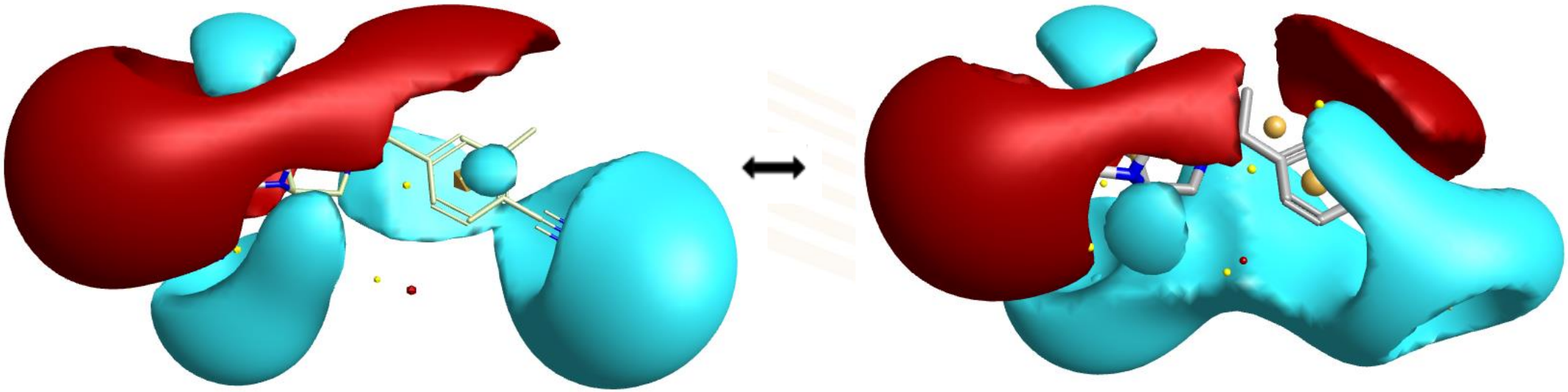
E.g. Bioisosteric replacement



SMIRKS: ([c;$(c!@n1@c@n@c@n1):1]n1[cH]n[cH]n1)>>([c;$(c!@C#;!@N):1]C#N)

Chemistry – What might we make?
Transformations

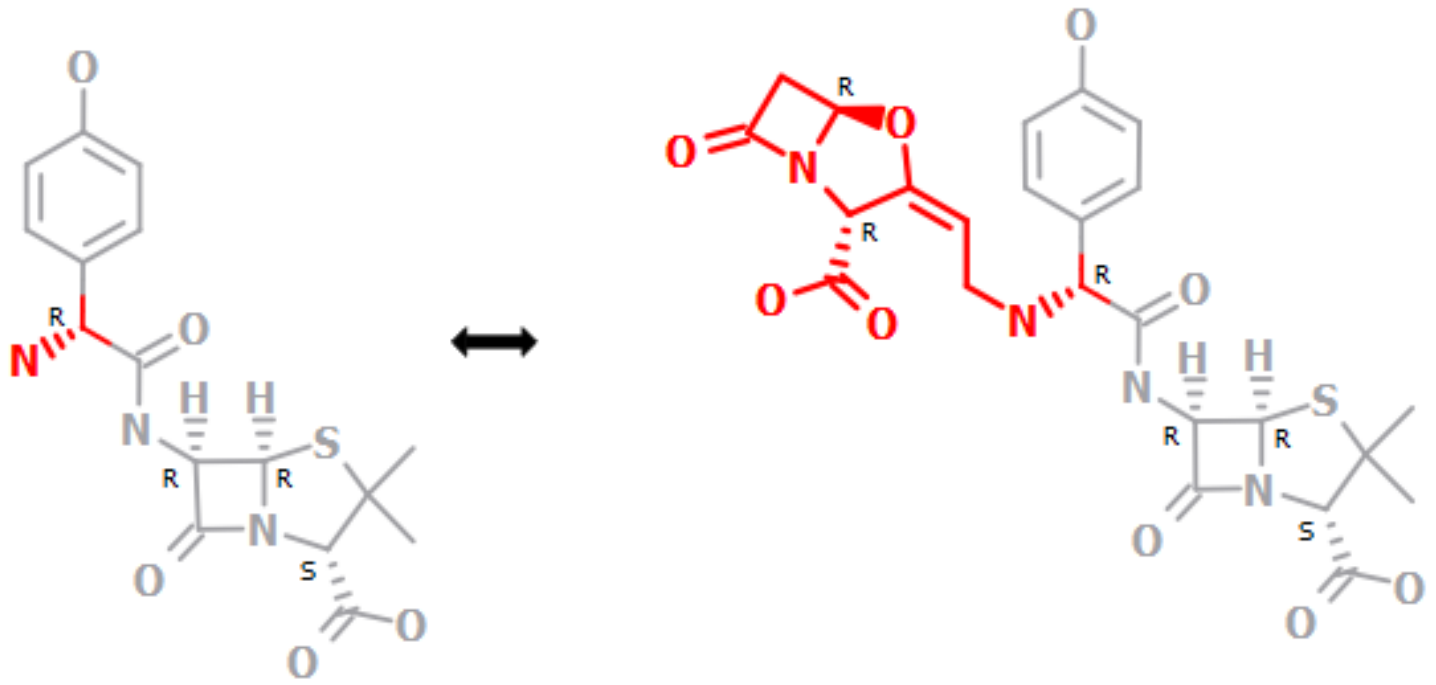
E.g. Bioisosteric replacement



SMIRKS: ([c;$(c!@n1@c@n@c@n1):1]n1[cH]n[cH]n1)>>([c;$(c!@C#;!@N):1]C#N)

Chemistry – What might we make?
Transformations

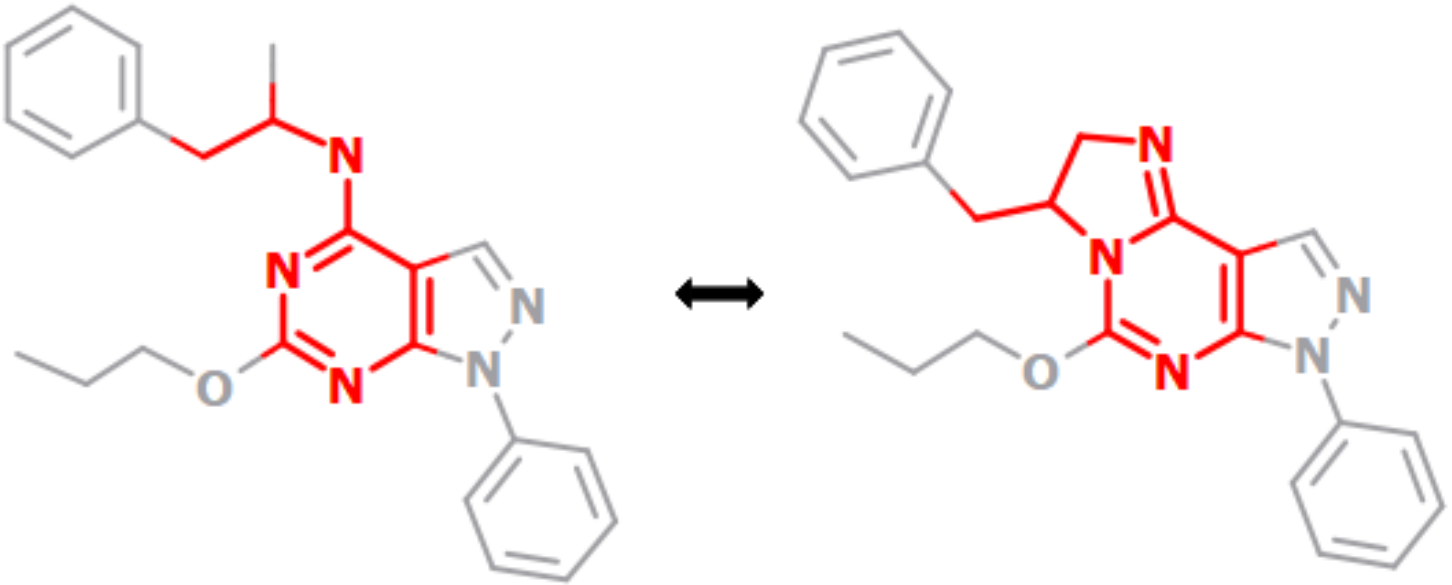
E.g. Prodrug strategy



SMIRKS: ([C;$(C!@N):1][NH2])>>([C;$(C!@N!@C!@C=;!@C1@C(@N2(@C(=;!@O)@C@C2@O1))!@C(!@O)=;!@O):1][NH][CH2][CH]=C1[C@@H](N2C(=O)[CH2][C@H]2O1)C([OH])=O)

Chemistry – What might we make?
Transformations

E.g. Ring opening/closing

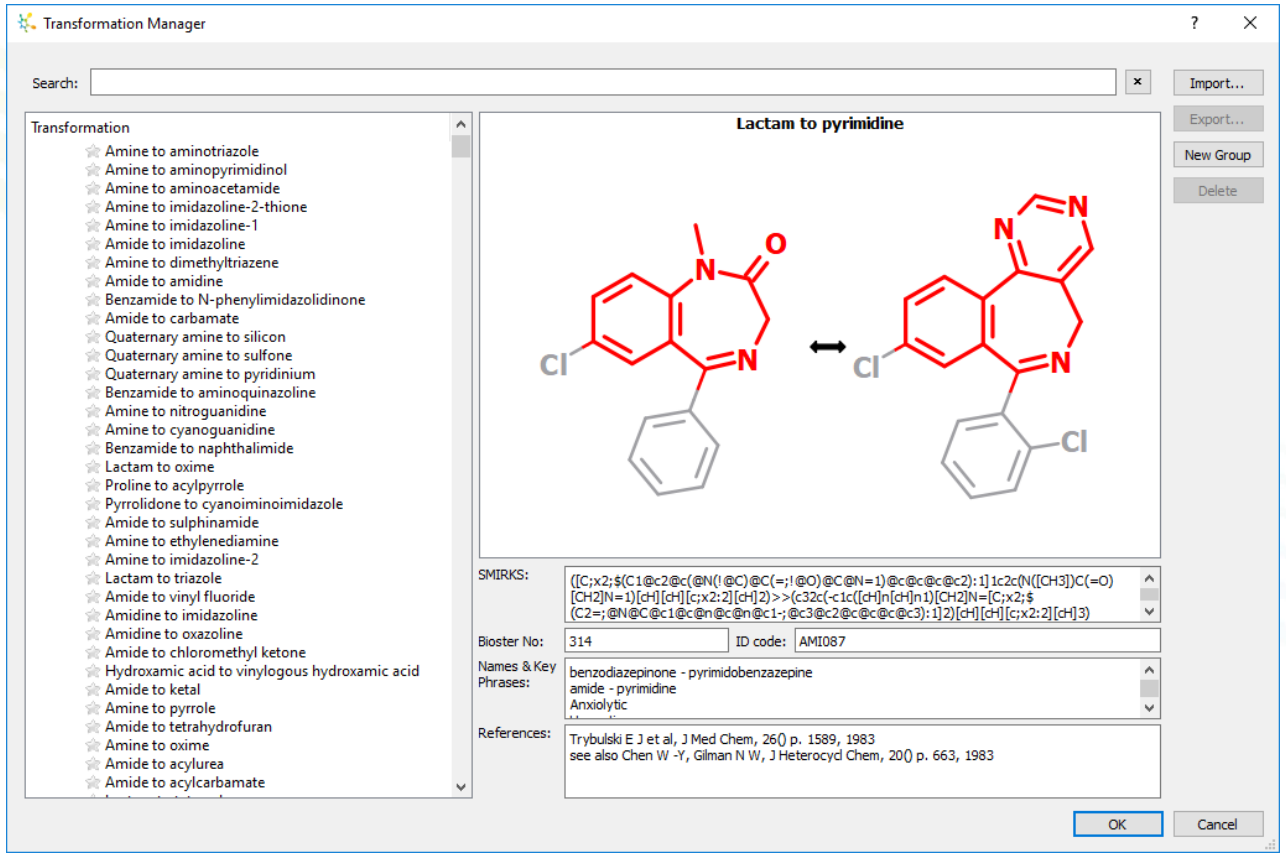


SMIRKS: ([c;$(c1@c@c(@n@c@n1)!@N!@C!@C):2]1[c:1]c(n[c;x2:4]n1)[NH][C:5][C:3])>>(N21C([c:1]([c:2](N=[C;x2;$(C1=;@N@c(@c(@C2@N1@C(@C@N=2)!@C)!@[H])!@[H]):4]1)[H])[H])=N[C;x2:5][CH]2[C:3])

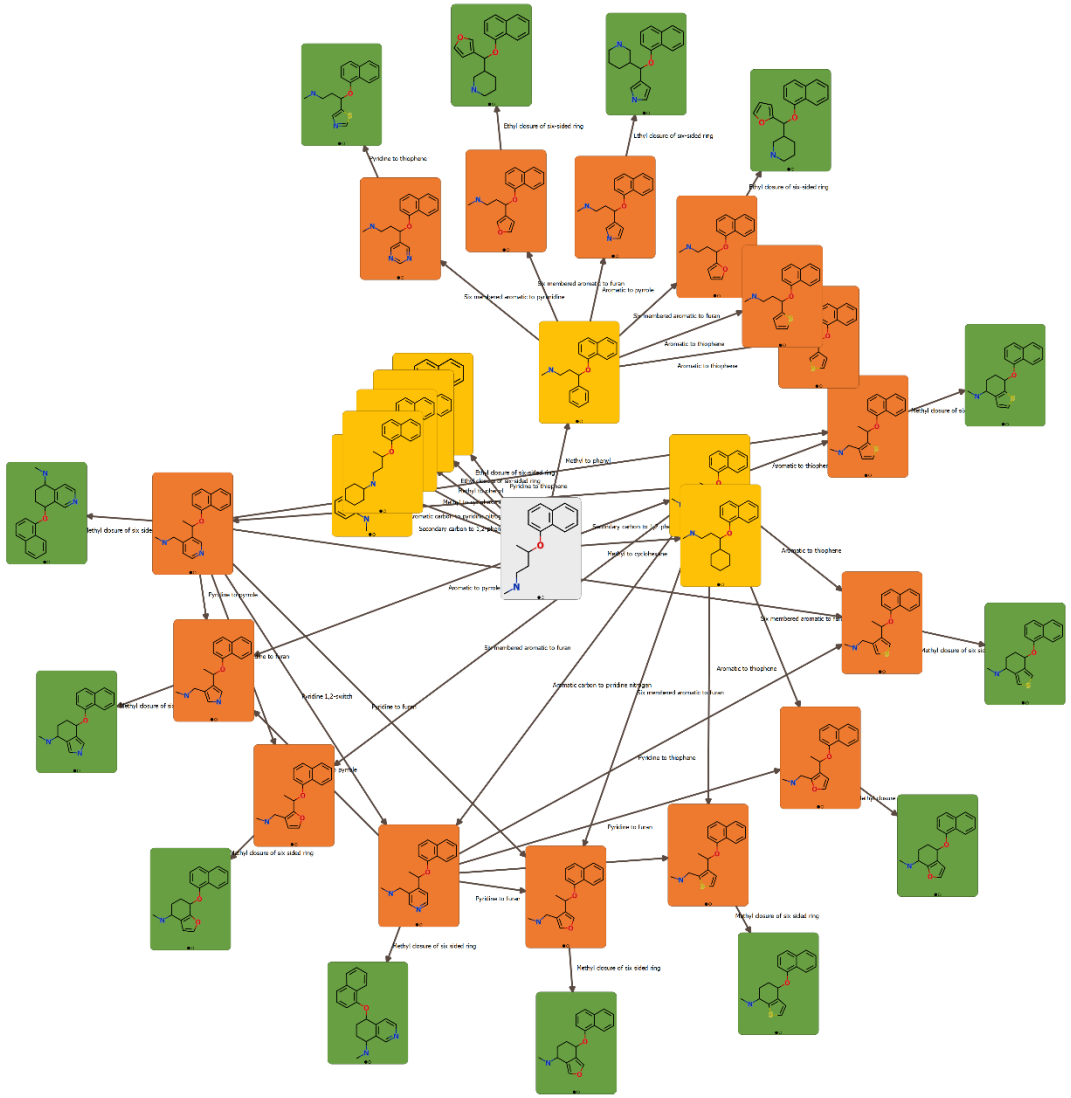
Chemistry – What might we make?

BIOSTER™

Database of ~30,000 precedented transformations from the chemistry literature



Chemistry – What could we make?
Applying Transformations

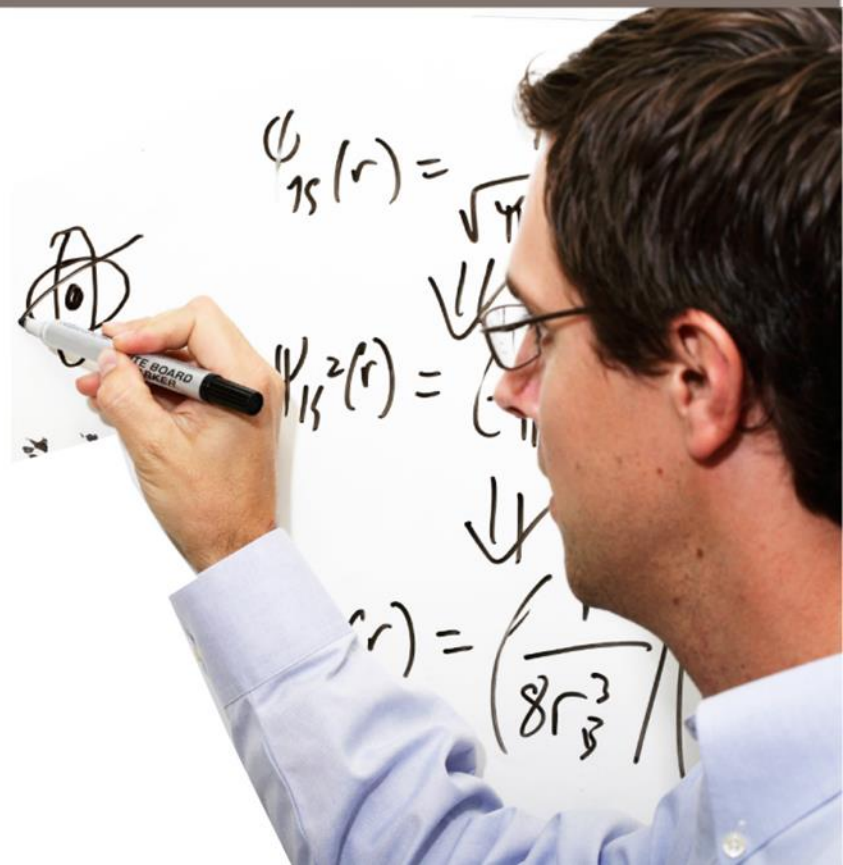


Segall *et al.* J. Chem. Inf. Model. (2011) **51**(11) pp. 2967-2976

Capturing and Applying Knowledge

SAR – How are molecular properties likely to change?

optibrium



SAR – How are Molecular Properties Likely to Change?
Quantitative Structure-Activity Relationships

$$y = f(x_1, x_2, x_3, \ldots) \pm \varepsilon$$

Statistical uncertainty

- Data
  - Quality data is essential
  - Public data need very careful curation* (and may not be good enough)
- Descriptors, e.g.
  - Whole molecule properties, e.g. logP, MW, PSA…
  - Structural descriptors, SMARTS, fingerprints…
- Machine learning method, e.g.
  - Artificial neural networks, support vector machines, random forests, Gaussian processes…

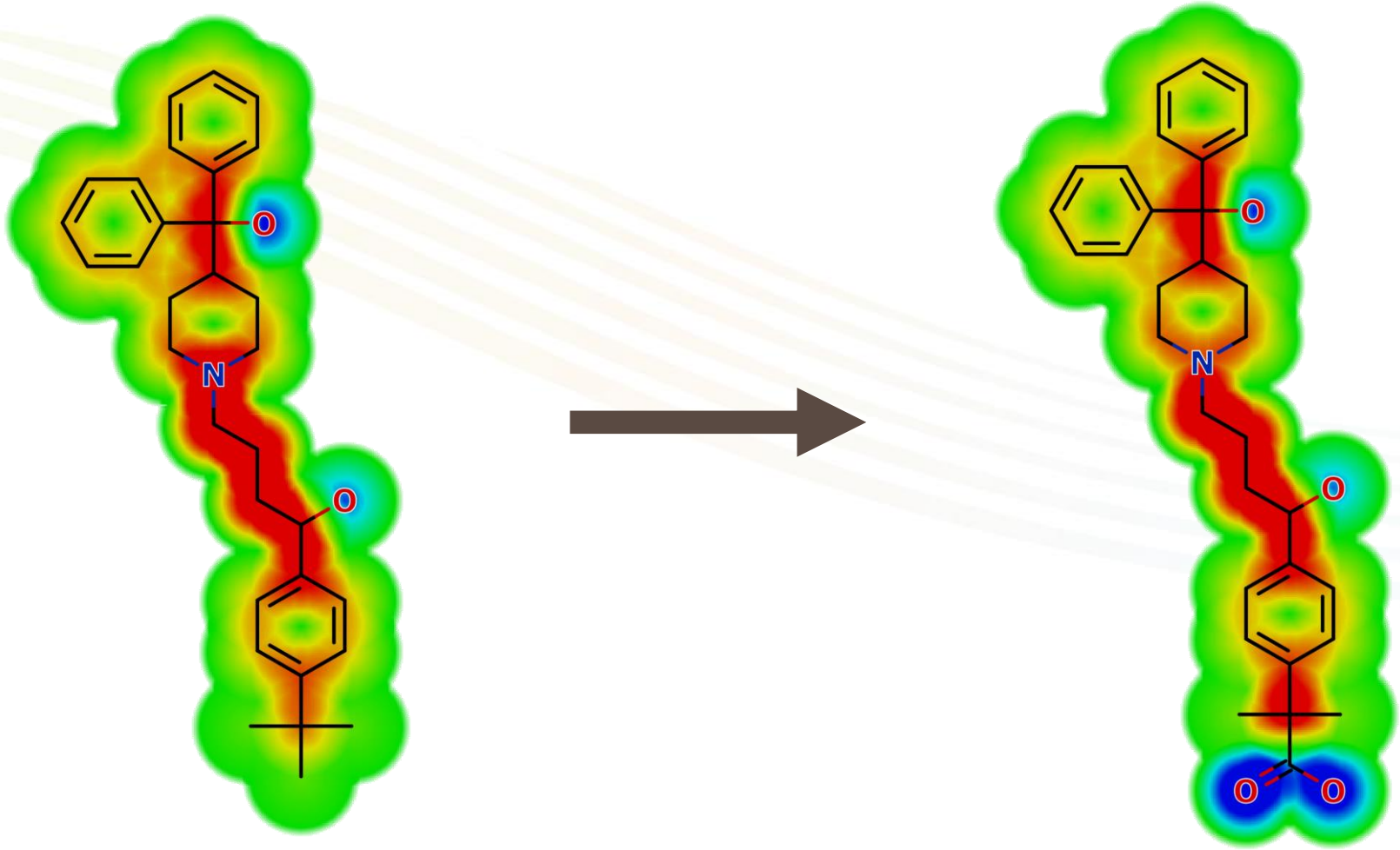
*Waldman et al. JCAMD (2015) DOI: 10.1007/s10822-015-9865-0

SAR – How are Molecular Properties Likely to Change?

Quantitative Structure-Activity Relationships

Terfenadine: hERG pIC_{50}=6.7

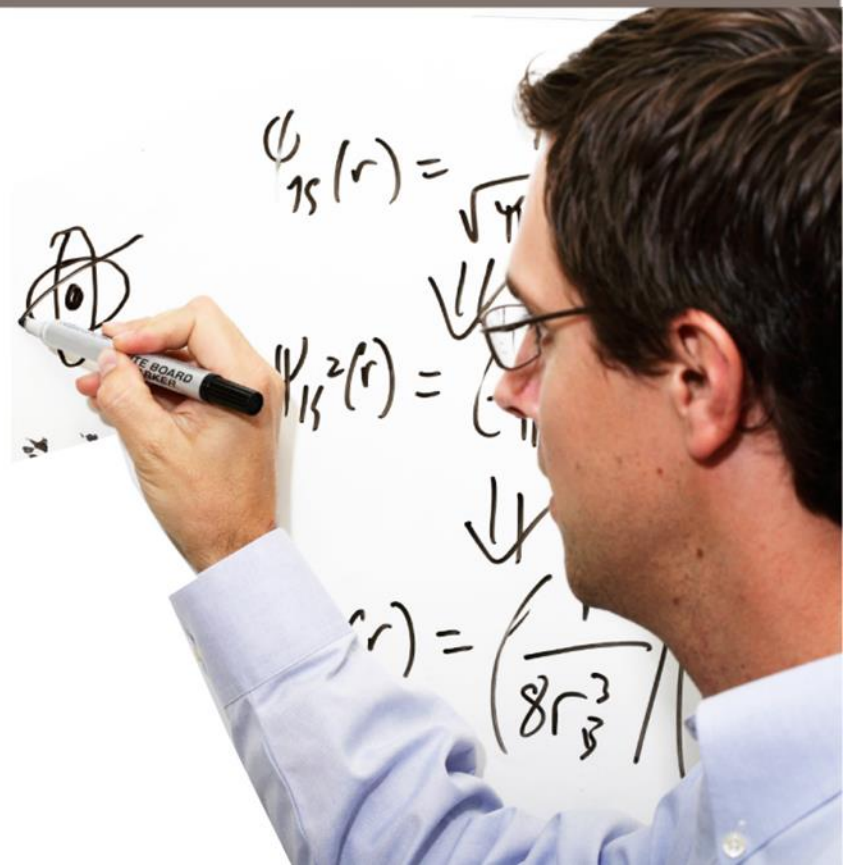
Fexofenadine: hERG pIC_{50}=5.2



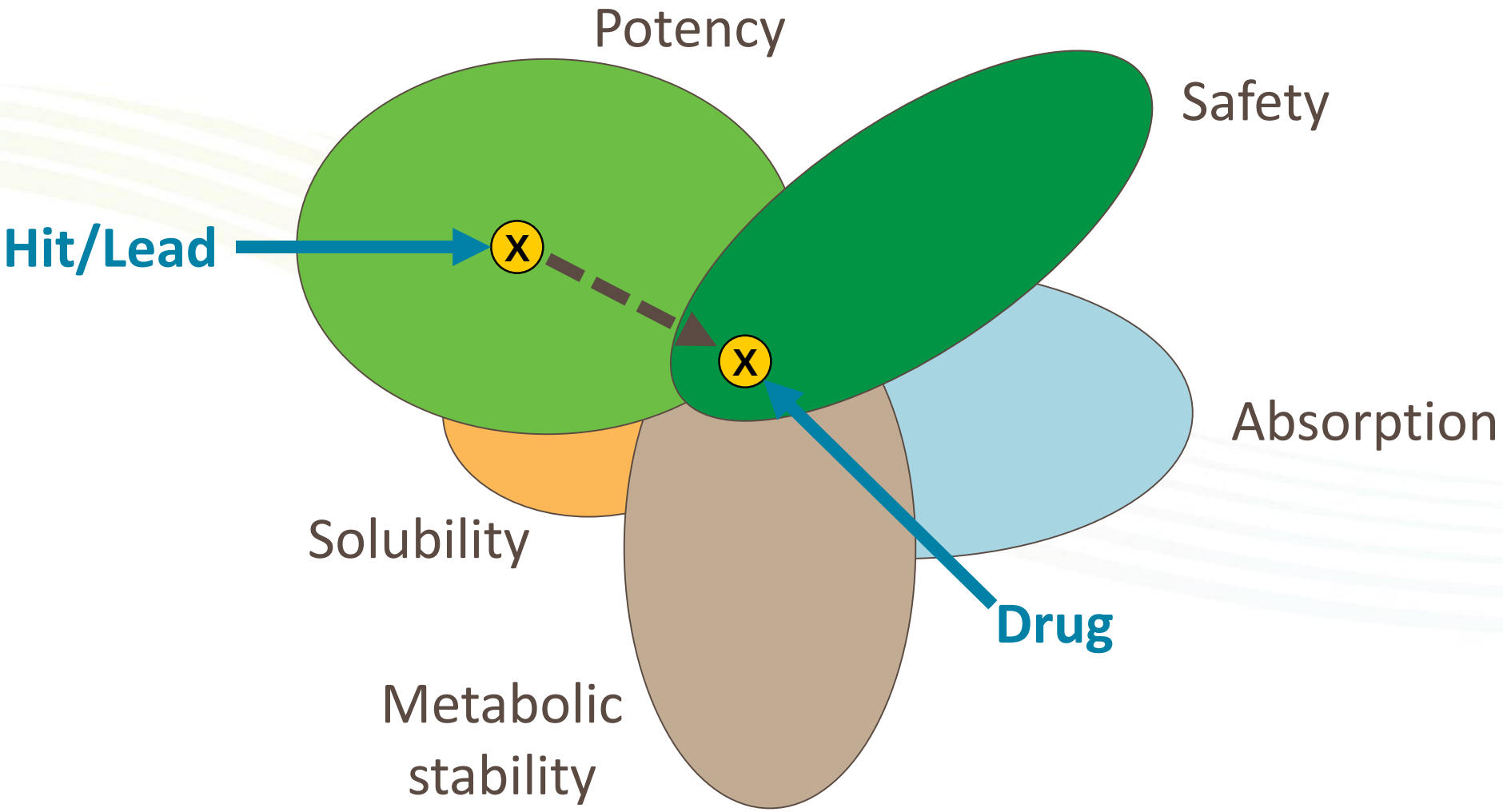
M.D. Segall *et al.* (2009) Chemistry & Biodiversity **6** pp. 2144-2151

Capturing and Applying Knowledge
Project objectives

optibrium



Project Objectives
Multi-parameter optimisation



Project Objectives

Multi-parameter optimisation – Probabilistic Scoring

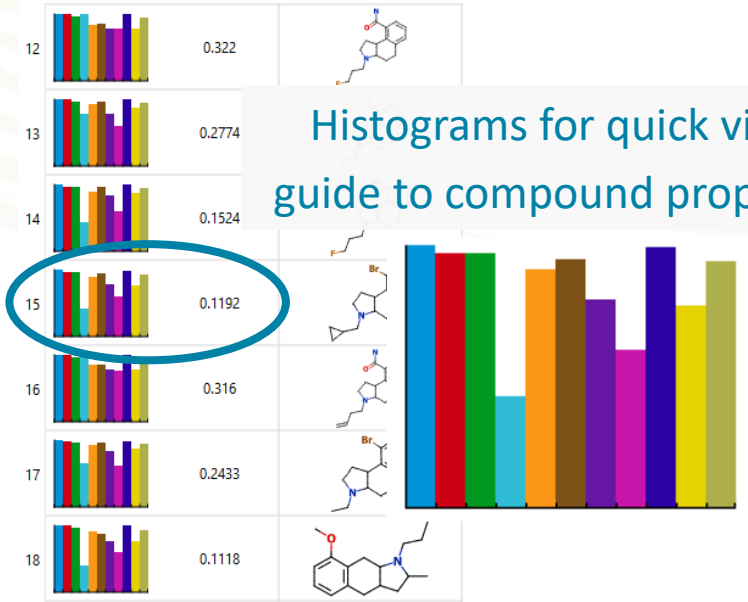
- Evaluates all available data against project criteria
- Accounts for the uncertainties in all compound-related data
- Objective assessment of compounds' chances of success

User-defined scoring profile

Profile: 5HT1A Project Profile

Property	Desired Value	Importance
5HT1a affinity (pKi)	> 7	
logS	> 1	
HIA category	+	
logP	0 -> 3.5	
BBB log([brain]:[blood])	-0.2 -> 1	
BBB category	+	
P-gp category	no	
hERG pIC50	⩽ 5	
2C9 pKi	⩽ 6	
2D6 affinity category	low medium	
PPB90 category	low	

Compounds ranked by likelihood of success



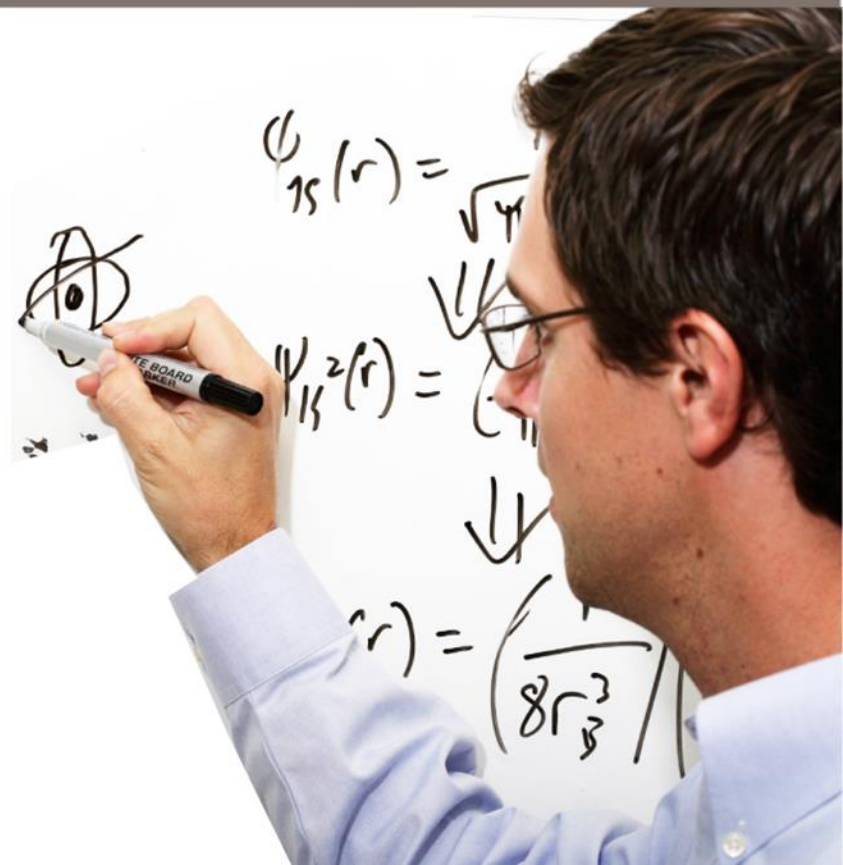
Histograms for quick visual guide to compound properties

M.D. Segall (2012) Curr. Pharm. Des. **18**(9) pp. 1292-1310

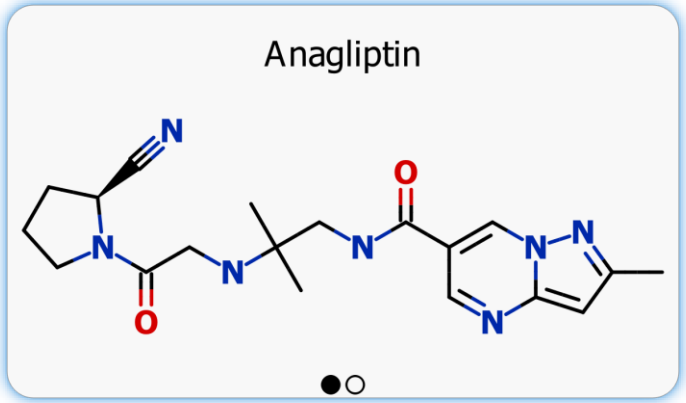
Example Application

Optimisation of high quality, selective Anagliptin analogues

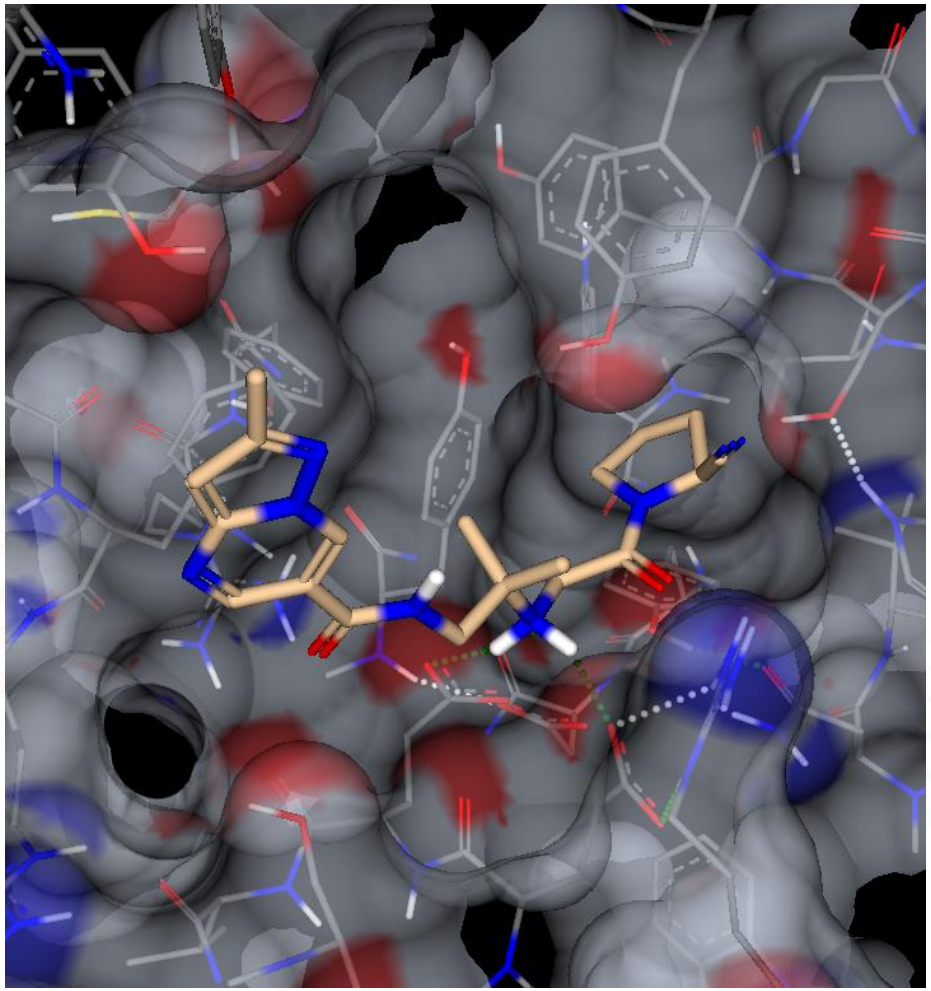
optibrium



Anagliptin



- Dipeptidyl Peptidase-4 (DPP-4) inhibitor
  - Treatment for type-2 diabetes mellitus
- Withdrawn in most markets due to animal toxicity*
  - Concerns raised about selectivity over DPP-2, DPP-8 and DPP-9
  - Since disproved…
- Explore strategies to improve DPP-4 activity and selectivity over other DPP isoforms



PDB 3WQH

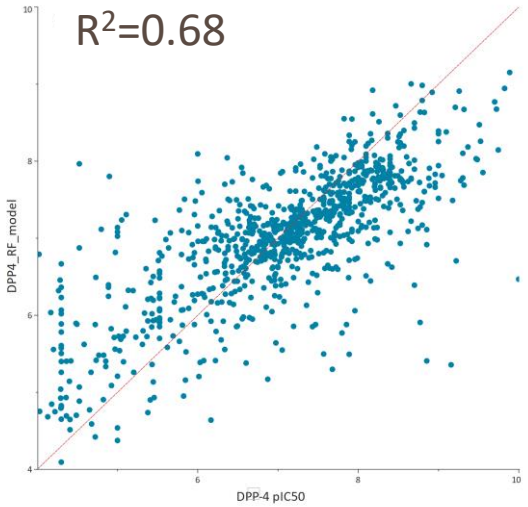
*R.N. Kushwaha *et al.* (2014) Curr. Med. Chem. **21**(35) pp. 4013-4045

Capturing SAR for DPP Activities

- QSAR models built with random forests method in StarDrop's Auto-Modeller*
  - 2D SMARTS descriptors and whole molecule properties, e.g. logP, MW, TPSA…
- Data sets of pIC_{50} data from ChEMBL†
- Validation on independent test sets:

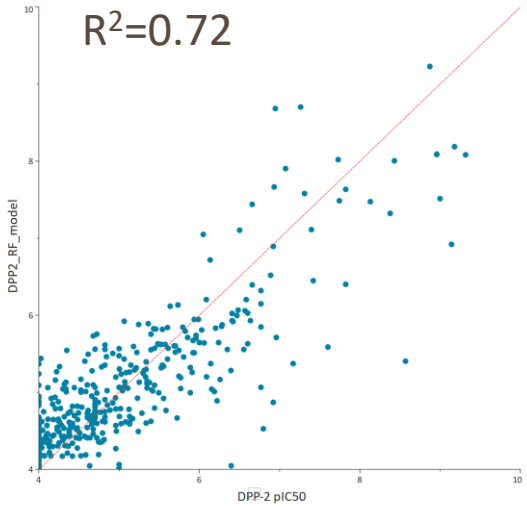
DPP-4

R^2=0.68



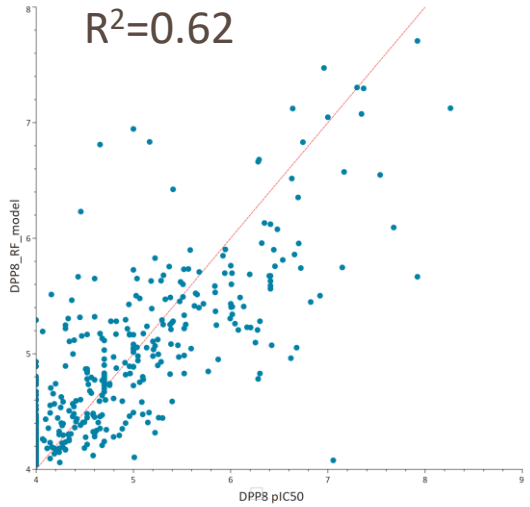
DPP-2

R^2=0.72



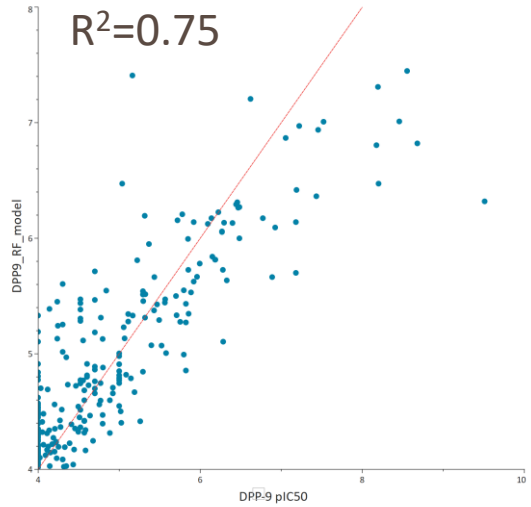
DPP-8

R^2=0.62



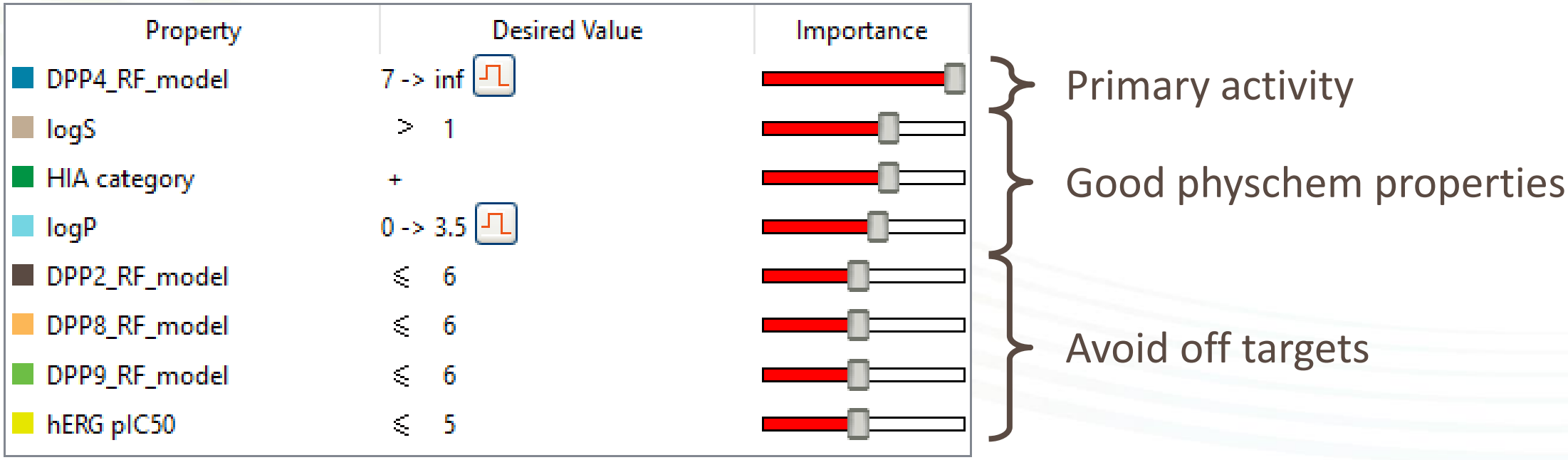
DPP-9

R^2=0.75

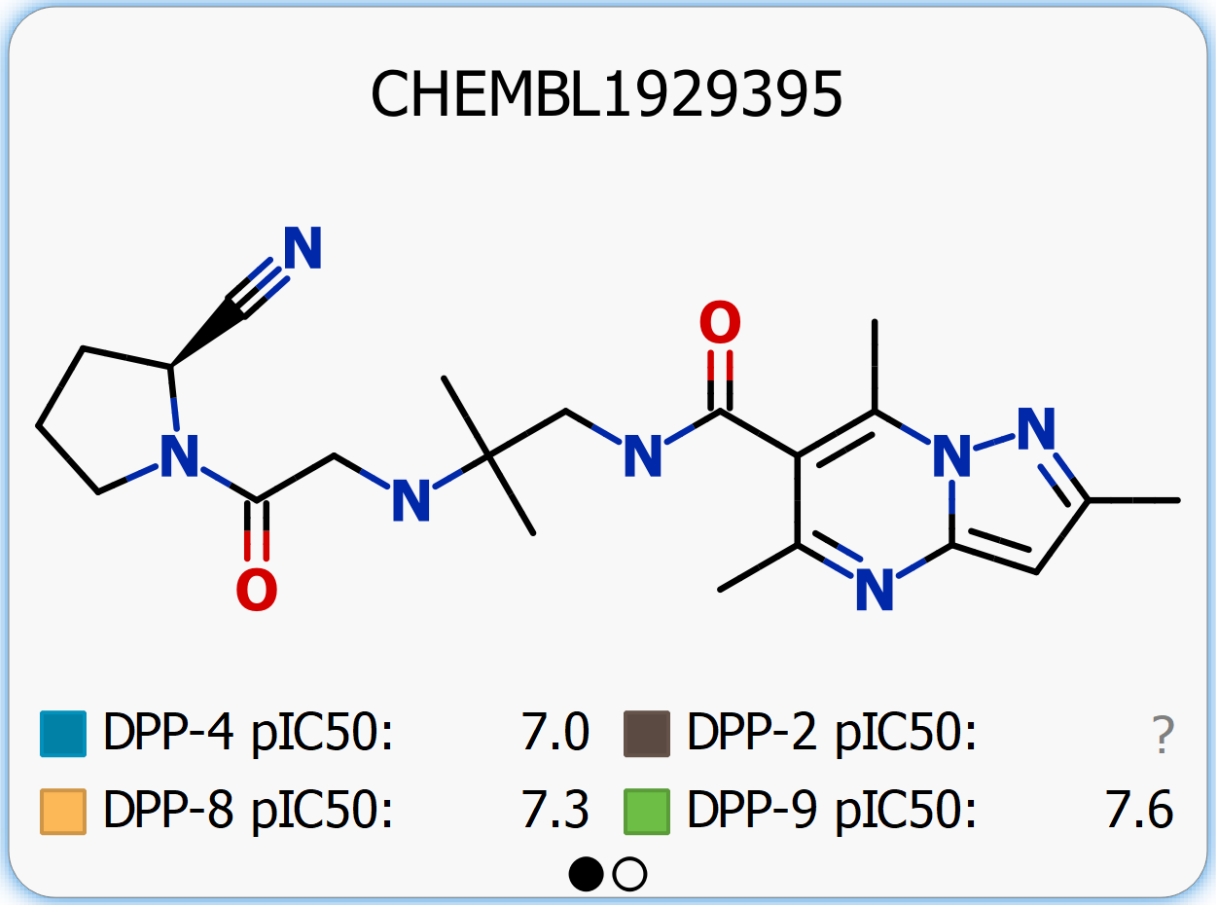


*O. Obrezanova *et al.* (2008) JCAMD **22**(6-7) pp. 431-440 †Bento *et al.* (2014) Nucleic Acids Res. pp. 1083-1090

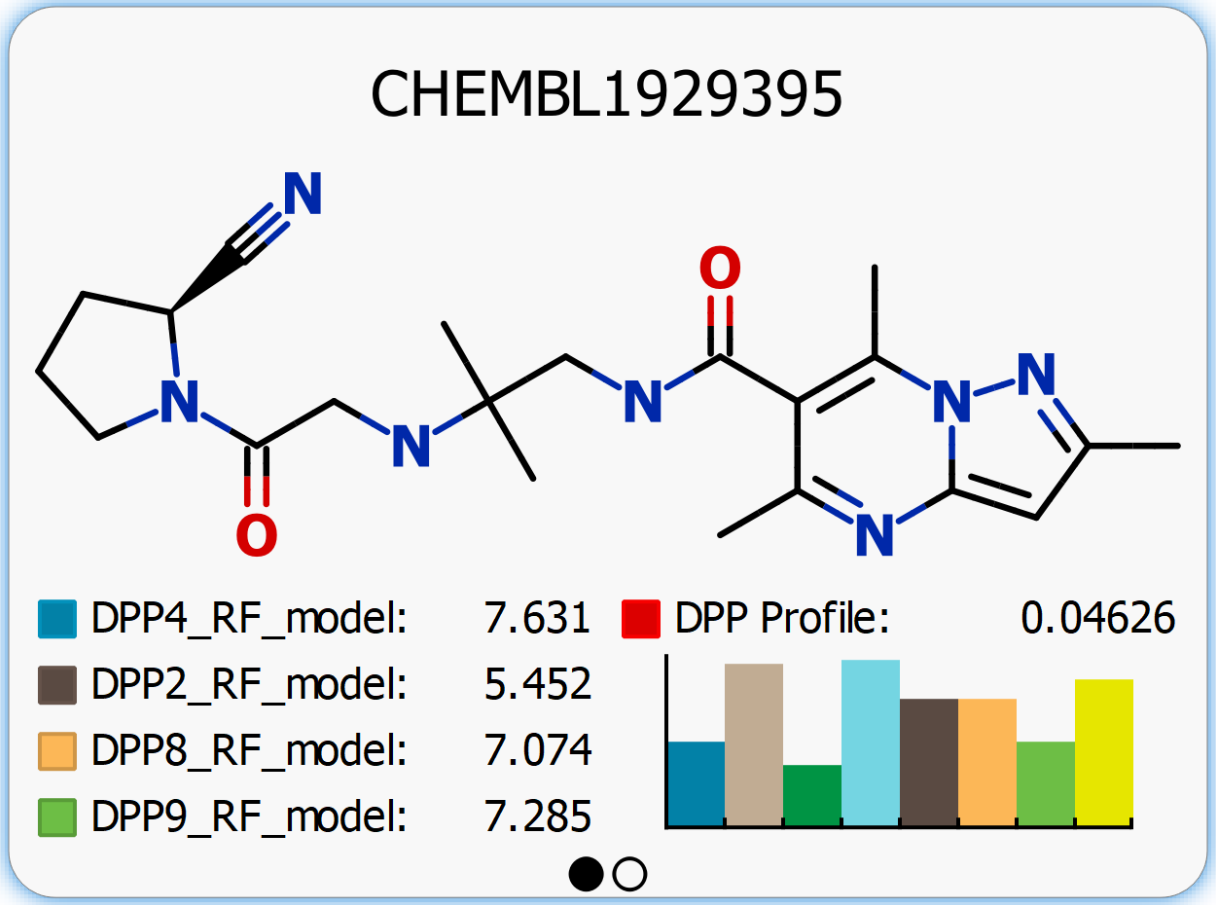
Capturing the Optimisation Objectives

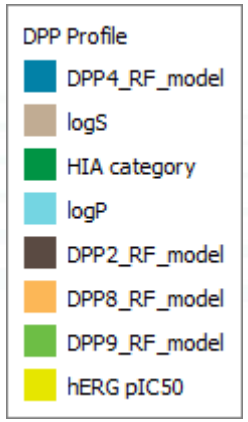


Starting Point
CHEMBL1929395



Starting Point
CHEMBL1929395

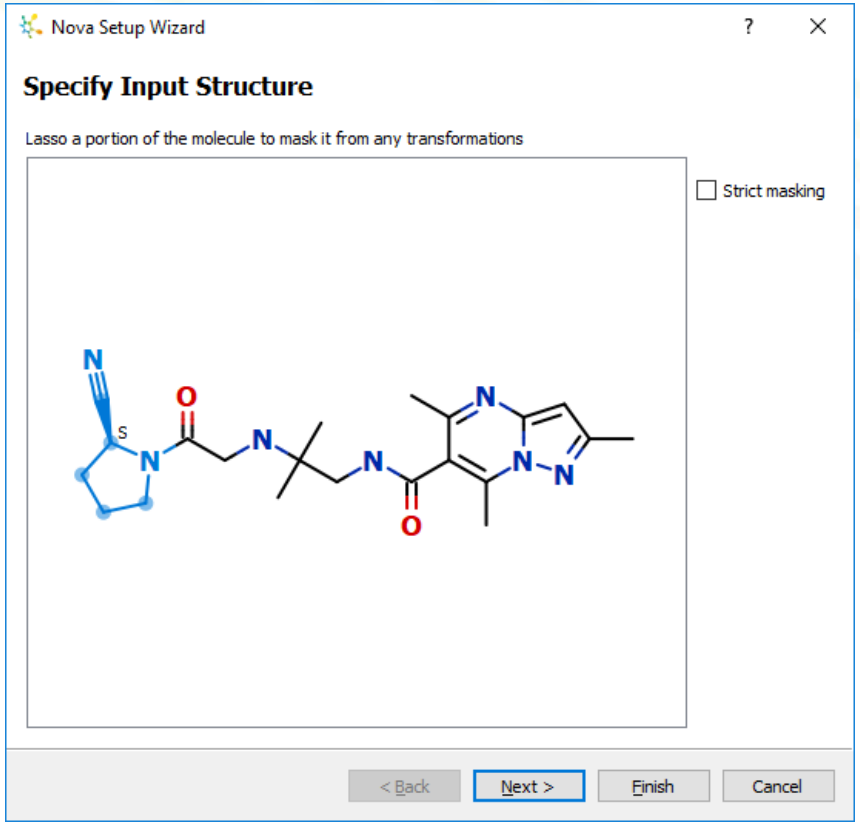


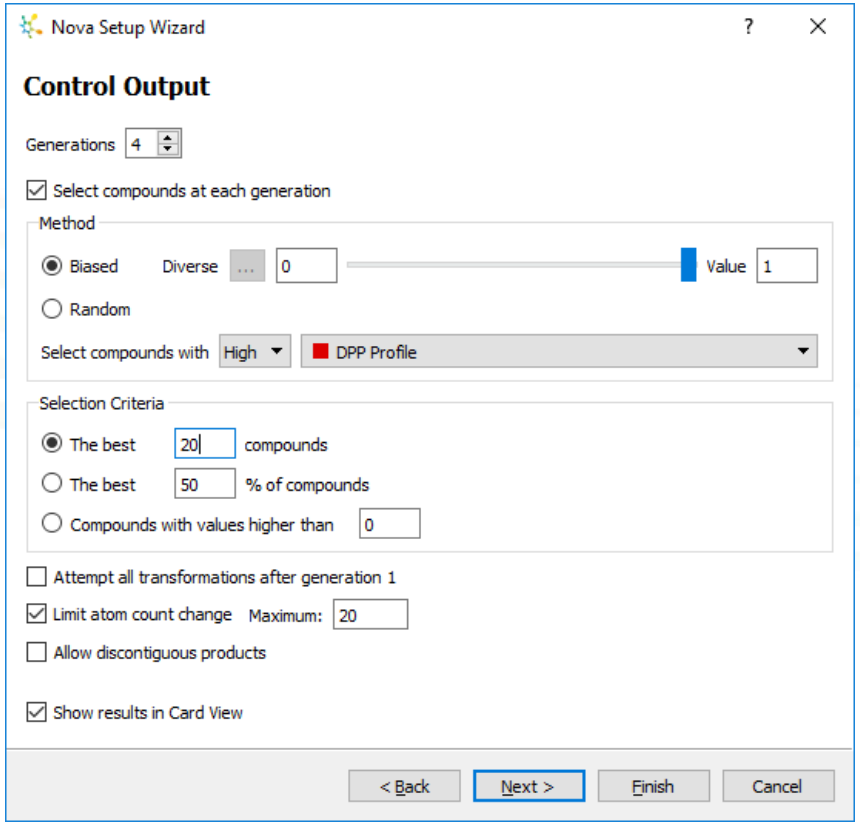


Guided Optimisation

Applying captured knowledge

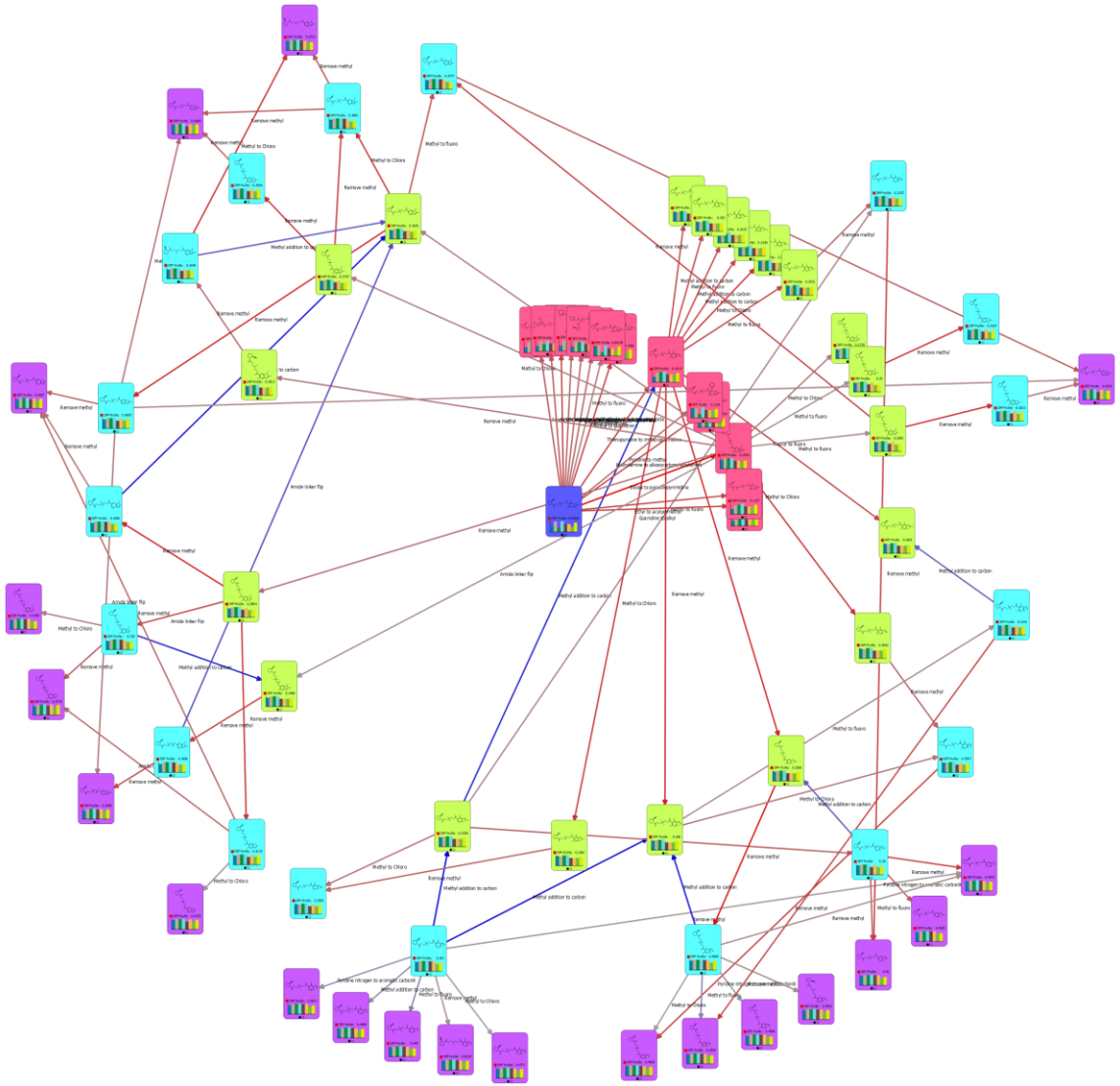
- Cyanopyrrolidine conserved
- Selection applied at each generation based on multi-parameter profile





Results

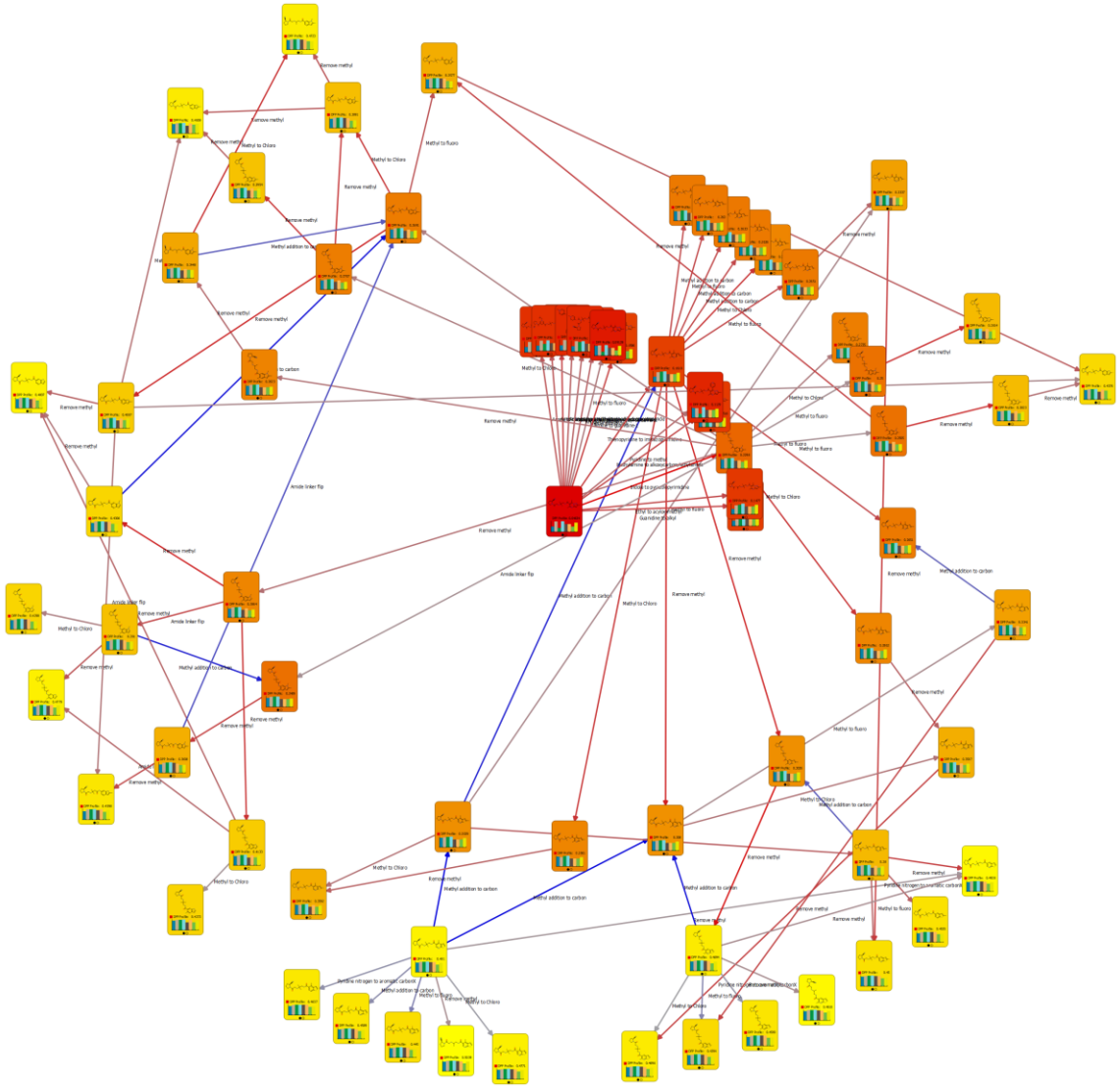
~16,000 compounds explored



- Input
- Generation 1
- Generation 2
- Generation 3
- Generation 4

Results

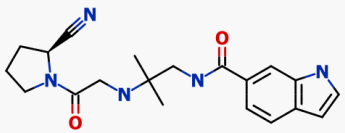
~16,000 compounds explored



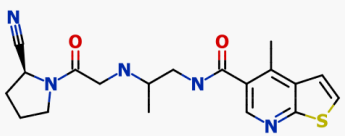
DPP-4 Profile Score

Low High

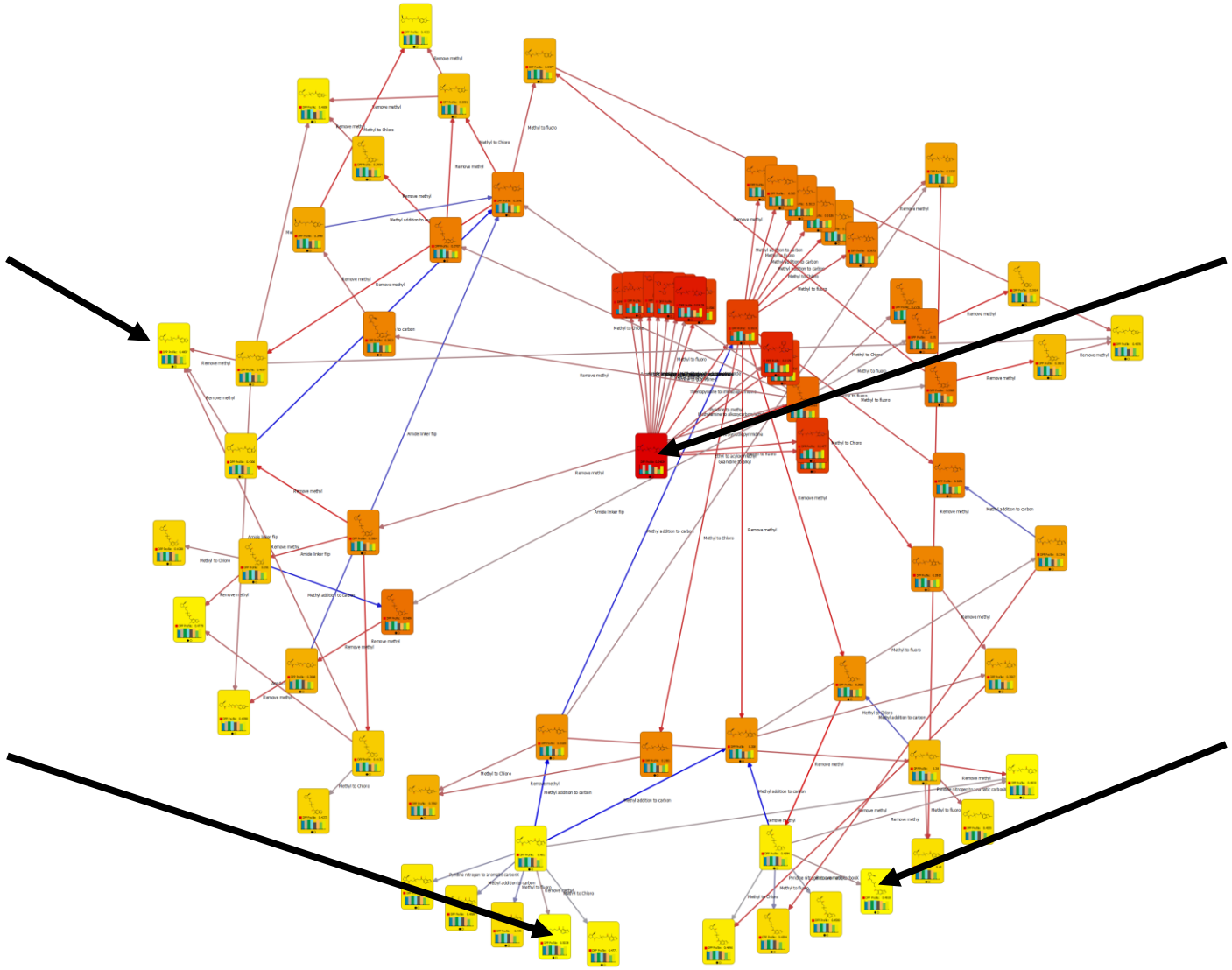
Results
Example compounds



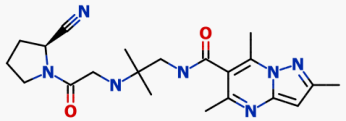
DPP Profile: 0.4807



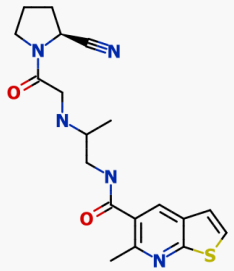
DPP Profile: 0.5039



CHEMBL1929395

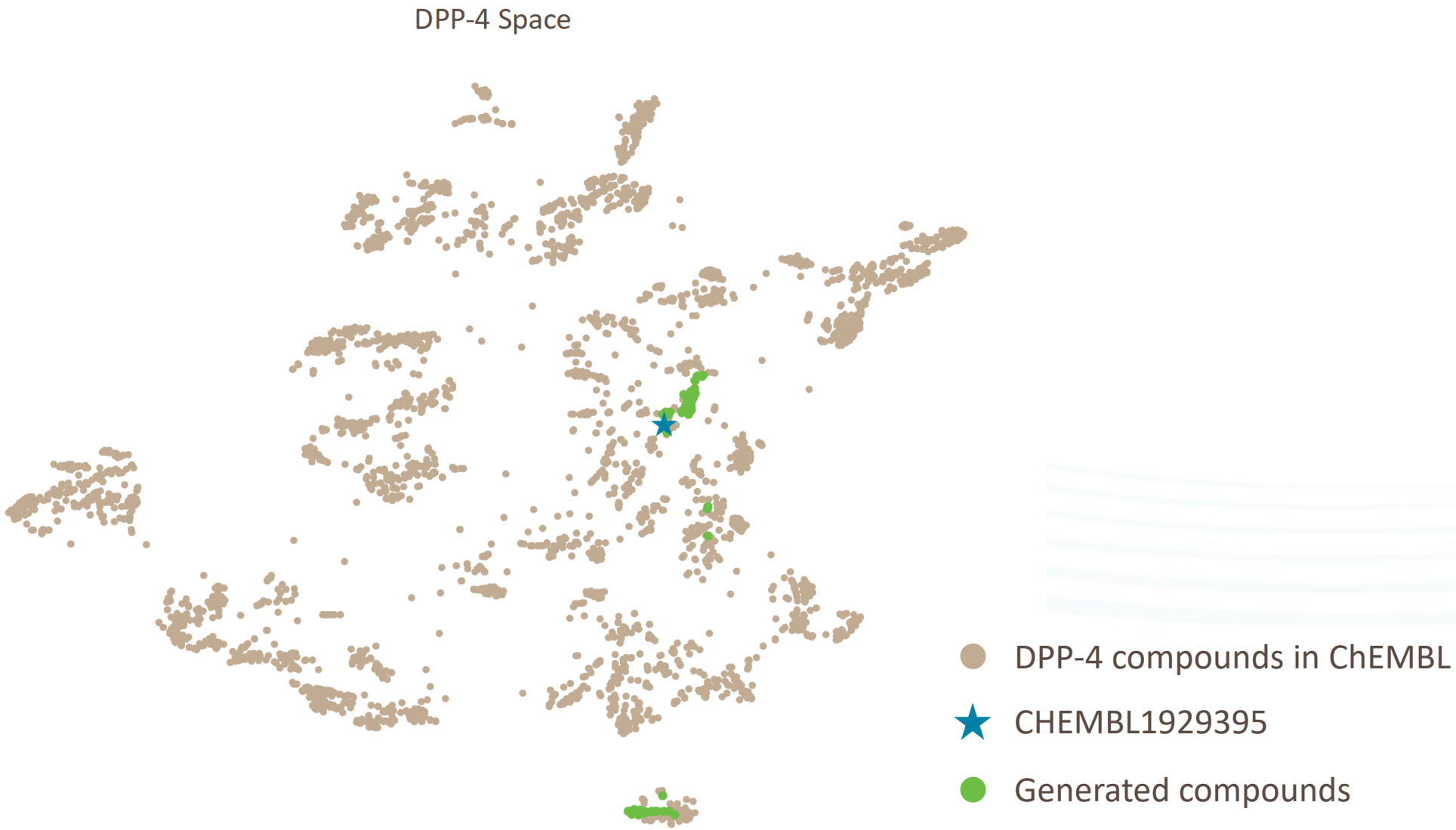


DPP Profile: 0.04634



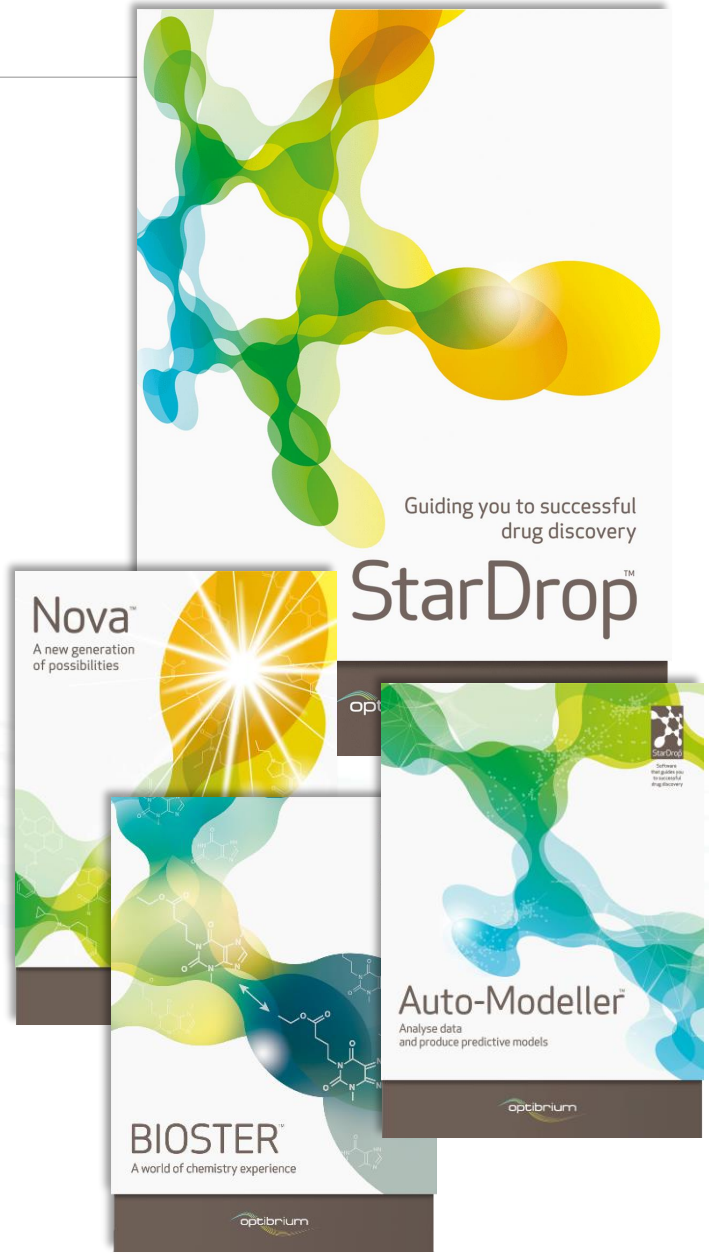
DPP Profile: 0.4918

Exploration of Chemical Space



Conclusions

- Capturing and applying knowledge computationally enables broad exploration of new optimisation strategies
  - ‘Think outside the box’
  - Rigorous investigation of possibilities
- Chemistry knowledge
  - Transformations representing optimisation strategies used in the past
- SAR knowledge
  - Machine learning to build QSAR models trained on relevant data
- Project knowledge
  - Expertise of project team from different disciplines as a scoring profile
- Integrated and applied using an evolutionary algorithm
- For more information and references, please visit:
  - www.optibrium.com/stardrop/
  - www.optibrium.com/community/



Acknowledgements

- The Optibrium team, including
  - Chris Leeding
  - James Chisholm
  - Alex Elliott
  - Fayzan Ahmed
  - Coran Hoskin
  - Aishling Cooke

- Istvan Ujváry, Julian Hayward, John Barnard
  - BIOSTER