The Discovery Process

- Gene
- Protein target
- Screen and identify lead
- Lead optimisation
- Chemical diversity (compound library)
- Test safety & efficacy in animals and humans

Targets → Hits → Leads → Candidates → Drugs → Products
Drug discovery and computational chemistry
Computation in Drug Discovery: the good news!

- *Only Two* questions
  - What (protein/pathway) to target?
  - And what molecule to make?

- *And One* topic for today
  - Molecule Design
Types of Molecules studied by GSK comp chem

- Small molecules
- Peptides
- Protein therapeutics
- Bioconjugates

- Materials

- Enzymes
The Lead Optimisation cycle

“Lead”

Analyse SAR

Design

Make

Test

“Candidate”

“Screening Cascade”

*in vitro*

Binding

Selectivity

Function

Safety Hazard

*in vivo*

PK/PD

Disease

Decreasing throughput
Surely Computation can help? “Rational” drug design

- Most design methodologies are aimed at reducing the number of cycles in lead optimisation- ideally to 1!

- All design methodologies, to date, have had limited success in this regard
A multi-objective optimisation

Traditional Way: Sequential Process, Costly, Lengthy
A multi-objective optimisation

Desired- faster navigation through multi-dimensional space, by reducing the cycles
Can computation help? Yes

Structure based drug discovery

Solvation

Potency

Solvability

Safety

Absorption

Metabolic stability

PC1
Structure-Based Optimization of Naphthryridones into Potent ATAD2 Bromodomain Inhibitors

Figure 1. From hit to micromolar ATAD2 inhibitor. pIC<sub>50</sub> values of compounds 1–3 in ATAD2 peptide- and ligand-based TR-FRET assays and against BRD4 BD1.
Computational analysis suggests ether linkage optimal for accessing RVF shelf

Data for C3’-Substituents in the C5-H Naphthyridone Series

<table>
<thead>
<tr>
<th>Compound</th>
<th>R</th>
<th>Enantiomer</th>
<th>ATAD2 pIC50</th>
<th>ATAD2 LE</th>
<th>BRD4 BD1 pIC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>H</td>
<td>-</td>
<td>4.8</td>
<td>0.33</td>
<td>5.4</td>
</tr>
<tr>
<td>9</td>
<td>OCH3</td>
<td>Enantiomer 1</td>
<td>5.0</td>
<td>0.33</td>
<td>5.2</td>
</tr>
<tr>
<td>10</td>
<td>O</td>
<td>Racemate</td>
<td>4.2</td>
<td>0.25</td>
<td>4.6</td>
</tr>
<tr>
<td>11</td>
<td>O</td>
<td>Racemate</td>
<td>4.6</td>
<td>0.27</td>
<td>5.2</td>
</tr>
<tr>
<td>12</td>
<td>O</td>
<td>Racemate</td>
<td>4.8</td>
<td>0.27</td>
<td>5.2</td>
</tr>
<tr>
<td>13</td>
<td>O</td>
<td>Racemate</td>
<td>5.5</td>
<td>0.28</td>
<td>5.3</td>
</tr>
</tbody>
</table>
Introduction of polarity to gain selectivity via differences in shelf region

![Chemical Structure](image)

<table>
<thead>
<tr>
<th>R</th>
<th>Chirality</th>
<th>Compound</th>
<th>ATAD2</th>
<th>ATAD2</th>
<th>ATAD2</th>
<th>BRD4 BD1</th>
<th>Δ</th>
<th>Solubility</th>
<th>LogD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pIC50</td>
<td>LE</td>
<td>pIC50</td>
<td>pIC50</td>
<td>pIC50</td>
<td>µM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.6</td>
<td>0.28</td>
<td>5.4</td>
<td>5.4</td>
<td>0.2</td>
<td>220</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Racemate</td>
<td>5.6</td>
<td>0.26</td>
<td>6.0</td>
<td>5.2</td>
<td>0.4</td>
<td>≥309</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enantiomer 1</td>
<td>4.3</td>
<td>0.20</td>
<td>5.3</td>
<td>5.3</td>
<td>-1.0</td>
<td>≥531</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enantiomer 2</td>
<td>6.1</td>
<td>0.29</td>
<td>6.3</td>
<td>4.9</td>
<td>1.2</td>
<td>≥421</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

![Diagram](image)
Adding back optimised C5 substituent

### Table 7. Summary of Properties of 38 and 46

<table>
<thead>
<tr>
<th>Property</th>
<th>38</th>
<th>46</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAD2 pIC&lt;sub&gt;50&lt;/sub&gt; (TR-FRET)</td>
<td>6.9</td>
<td>6.5</td>
</tr>
<tr>
<td>ATAD2 pIC&lt;sub&gt;50&lt;/sub&gt; (Peptide FRET)</td>
<td>6.9</td>
<td>6.6</td>
</tr>
<tr>
<td>ATAD2 pIC&lt;sub&gt;50&lt;/sub&gt; (Itochrome)</td>
<td>7.3</td>
<td>7.0</td>
</tr>
<tr>
<td>ATAD2 pK&lt;sub&gt;i&lt;/sub&gt; (Itochrome)</td>
<td>7.1</td>
<td>6.7</td>
</tr>
<tr>
<td>ATAD2 pK&lt;sub&gt;i&lt;/sub&gt; (Itochrome)</td>
<td></td>
<td>7.7</td>
</tr>
<tr>
<td>BRD2 BD1/BD2 pIC&lt;sub&gt;50&lt;/sub&gt; (TR-FRET)</td>
<td>4.5/&lt;4.5</td>
<td>&lt;4.3/&lt;4.3</td>
</tr>
<tr>
<td>BRD5 BD1/BD2 pIC&lt;sub&gt;50&lt;/sub&gt; (TR-FRET)</td>
<td>&lt;4.3/&lt;4.3</td>
<td>&lt;4.3/&lt;4.3</td>
</tr>
<tr>
<td>BRD2 BD1/BD2 pIC&lt;sub&gt;50&lt;/sub&gt; (TR-FRET)</td>
<td>4.3/&lt;4.3</td>
<td>&lt;4.3/&lt;4.3</td>
</tr>
<tr>
<td>Chrom LogD (pH 7.4)</td>
<td>1.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Antimicrobial permeability (nm/s, pH 7.4)</td>
<td>&lt;5</td>
<td>&lt;10</td>
</tr>
<tr>
<td>CLND solubility (μM)</td>
<td>179</td>
<td>≥341</td>
</tr>
</tbody>
</table>

![Diagram A](#)

![Diagram B](#)
Design approach for ATP site fragment set

- Fragment set for screening against kinases
- Target mainly the purine hinge-binding site, which is small and flat

- Aromatic ring and 1-3 hydrogen bonds
- 3D pharmacophore
- 2D representations also suitable
Design approach for ATP site fragment set

- Fragment set for screening against kinases
- Target mainly the purine hinge-binding site, which is small and flat

Available solids → Reactivity filters
3D pharmacophore → Kinase SMARTS
Cluster & visualise → LCMS / NMR QC

Final set: 936 fragments with good property profile
Kinase fragment activity profiles

- Kinase fragments have selectivity profiles, just like larger molecules
- This may be unexpected, since they bind in the most conserved part of the domain

Kinase activity profile similarity
e.g. Tanimoto coefficient
Kinase profile similarity tracks structural similarity

- 2D descriptors
- 3D similarity (e.g. Openeye ROCS/ EON, Cresset FieldScreen…)

- 3D similarity metrics capture some of the features of fragments important for selectivity

As the 3D fields of fragments become more similar, so their kinase inhibition profiles* also tend to become more similar

*Tanimoto activity profile similarity

A good choice for fragment similarity or diversity

“Selectivity of Kinase Inhibitor Fragments” Bamborough, Brown, Christopher, Chung, Mellor
Theoretical considerations when working with Protein Structures

- What you see on a screen is not real
  - Proteins *move*
  - Hydrogens are often assigned not observed

- How is solvation being treated?
  - especially when looking to make h-bonds
- Am I taking entropy into account?
- How are vDW energies being calculated?
  - remember $r^{12}$!

- Ligand conformational energies are as important as protein-ligand interaction energies
- Ligand based modelling methods are *very* effective when you also have a protein structure
Example: FXa/Thrombin inhibitors

\[
\text{Thrombin} = 31 \text{ nM}
\]

\[
\text{FXa} = 5 \text{ nM}
\]

\[
K_i(\text{Thrombin}) = 367 \text{ nM}
\]

\[
K_i(\text{Thrombin}) = 17 \text{ nM}
\]

Sulfonamide-related conformational effects and their importance in structure-based design

Dedicated to Prof. Dr. em. Günther Maier (Justus-Liebig-Universität Giessen, Germany) on the occasion of his 75th birthday.

Stefan Senger, Chuen Chan, Maire A. Convery, Julia A. Hubbard, Gita P. Shah, Nigel S. Watson, Robert J. Young
Do they have a different binding mode in thrombin???

No
Relaxed scan of the dihedral angle [C=C-S=O] (in 10° increments)
The calculations were performed with Gaussian98 at the B3LYP/6-31G* level of theory. The orange circle is placed at the dihedral angle found when the ligands are bound to thrombin.
Virtual Screening

- The *in silico* equivalent of “wet” assays/HTS
- Requires a method of predicting active compounds
  - protein structure (+docking)
  - 3D pharmacophore
  - QSAR equation
- Capable of screening many more molecules than can be made or tested in reality
- This “high throughput” use of comp. chem. requires us to make the most approximations
Chemical Descriptors

- Molecular fields
- Reduced Graphs
- 3D Pharmacophores
- "3-point" Pharmacophores
- Fragments
- Atom pairs and paths
- Shape fingerprint
- Reference Shapes
- "2D" fingerprints
- "3D" fingerprints
Chemical Descriptors

- Which ones to use?
  - Which are “best”?

- GSK
  - 2D fingerprints for diversity
  - Reduced graphs, 3D pharmacophores. 3D fields and shape for knowledge based work

- The most important ingredient for success is the quality of the computational chemist, not the software (or descriptors) they have to hand (assuming a base level of scientific validation!)
Pharmacophores

- A reductionist approach to bioactivity
  - "A **pharmacophore** is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" *IUPAC*

- Neglect of entropy, solvation and anything subtle!
Example: 3D pharmacophore lead hopping

Endothelin antagonists
Molecular Field methodology - Cresset
Use of a special forcefield to better simulate electrostatics

IsoStar- Cambridge Crystallographic Data Centre
Field searching example—Oxytocin antagonists

219 compounds

Discovery and optimization of highly ligand-efficient oxytocin receptor antagonists using structure-based drug design

Benjamin R. Beilenich*, Nicholas P. Barton*, Amanda J. Emmons*, Jag P. Heer*
“2D Pharmacophores”

- Use pharmacophore concept, but not 3D distances or arrangements
- Encode pairs of pharmacophore points separated by a certain number of bonds (typically from 1 to ~10)
- These are used to produce a vector, or fingerprint, which can then be used to compute molecular similarity (cf. in 2D chemical database systems)
- Advantage
  - Don’t have to worry about active conformations
- Disadvantage
  - May be less selective than 3D methods (i.e. may need to screen more molecules)
Reduced Graphs

– A powerful “2D” pharmacophore description
– CB1 antagonist example

The Reduced Graph Descriptor in Virtual Screening and Data-Driven Clustering of High-Throughput Screening Data

G. Harper, G. S. Bravi, S. D. Pickett, J. Hussain, and D. V. S. Green
GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY, United Kingdom

DOI: 10.1021/ci049860f
CB1 Antagonists

- **Pfizer**
  - 2D FP: 0.468
  - RG: 0.815

- **Solvay**
  - 2D FP: 0.468
  - RG: 1

- **Astra**
  - 2D FP: 0.420
  - RG: 1

- **Bayer**
- **Va. Commonwealth**
- **Solvay**
- **Eli Lilly**
- **Aventis**
- **Merck**
- **U. Connecticut**
Reaching Beyond the Fog in HTS

– Data-driven analysis finds motifs, Reduced Graphs, frameworks etc that are enriched in the active population

---

The Reduced Graph Descriptor in Virtual Screening and Data-Driven Clustering of High-Throughput Screening Data

G. Harper , G. S. Bravi, S. D. Pickett, J. Hussain, and D. V. S. Green
GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY, United Kingdom


DOI: 10.1021/ci049860f
Physical Properties

Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity

Alan P. Hill¹ and Robert J. Young²

¹Department of Analytical Chemistry, GlaxoSmithKline Medicines Research Centre, Stevenage, Hertfordshire SG1 2NY, UK
²Department of CSC Medicinal Chemistry, GlaxoSmithKline Medicines Research Centre, Stevenage, Hertfordshire SG1 2NY, UK
Simple metrics can be effective and highly competitive with first principles computation

Proposed: “Solubility Forecast Index”

SFI = clog D$_{7.4}$ + #Ar

or

SFI = Chrom log D$_{7.4}$ + #Ar

Measured CLND solubility < 30µM; 30-200µM >200µM
Can computation help? Sometimes

Potency

Solubility

Metabolic stability

Safety

Absorption

Toxicology prediction
Off target effects

Prediction of permeation
Active transport

Metabolite prediction

PC1
Physical Properties and flatness together

Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity

Robert J. Young¹, Darren V.S. Green², Christopher N. Luscombe² and Alan P. Hill³
# PFI and probability of success

<table>
<thead>
<tr>
<th>Assay / target value</th>
<th>PFI = mChrom log D_{447,4} + #Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;3</td>
</tr>
<tr>
<td>Solubility &gt;200 μM</td>
<td>89</td>
</tr>
<tr>
<td>%HSA &lt;95%</td>
<td>88</td>
</tr>
<tr>
<td>2C9 pIC&lt;5</td>
<td>97</td>
</tr>
<tr>
<td>2C19 pIC&lt;5</td>
<td>97</td>
</tr>
<tr>
<td>3A4 pIC&lt;5</td>
<td>92</td>
</tr>
<tr>
<td>Cl&lt;3 &lt;3 ml/min/kg</td>
<td>79</td>
</tr>
<tr>
<td>Papp &gt;200 nm/s</td>
<td>20</td>
</tr>
</tbody>
</table>

IPFI = mChrom log P + #Ar

| hERG pIC<5 (±1 charge) | 86 | 93  | 88  | 70  | 54  | 36  | 29  | 21   | 11  |
| Promiscuity <5 hits with pIC>5 | 85 | 78  | 74  | 65  | 49  | 30  | 20  | 13   | 7   |

*Colouring refers to the % chance of achieving benchmark value in that PFI bin: green, >67%; yellow, 34-67%; and red, <33%.
Adapted from:
“Lipophilicity in Drug Discovery”
M. J. Waring
BioDig – Why?

– Common problem in medicinal chemistry..

**Key Issue:** High intrinsic clearance

"Standing on the shoulders of giants" Isaac Newton

‘Why don’t you change it for an oxadiazole – it worked for us in Prog X’.
Mining our data

- How?

- How do we analyse our SAR?
  - One way is to look for pairs of compounds which only differ by a single change

The change in activity can then be attributed to the structural change
BioDig ADME database

- A database has been created where we have collected together the compound pairs (and associated structural changes) in our ADME data

<table>
<thead>
<tr>
<th>Property</th>
<th>Number of compounds</th>
<th>Number of matched molecular pairs</th>
<th>Number of transforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearance</td>
<td>54k</td>
<td>9M</td>
<td>7M</td>
</tr>
<tr>
<td>Solubility</td>
<td>267k</td>
<td>243M</td>
<td>217M</td>
</tr>
<tr>
<td>Log D</td>
<td>328k</td>
<td>316M</td>
<td>279M</td>
</tr>
<tr>
<td>hERG*</td>
<td>175k</td>
<td>98M</td>
<td>76M</td>
</tr>
<tr>
<td>PP Binding</td>
<td>248k</td>
<td>137M</td>
<td>117M</td>
</tr>
<tr>
<td>P450 3A4</td>
<td>262k</td>
<td>193M</td>
<td>167M</td>
</tr>
<tr>
<td>PGPefflux</td>
<td>6k</td>
<td>367k</td>
<td>306k</td>
</tr>
</tbody>
</table>

Represents a huge amount of knowledge to tap into

*old assay format
BioDig ADME

The information in the database can be used to answer other questions

Substructure X: “What possible replacements have been tried and which improve clearance?”
## Surely Computation can Help? What have we learnt?

<table>
<thead>
<tr>
<th><strong>Strengths</strong></th>
<th><strong>Threats</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Small, well-connected community</td>
<td>Small, aging, community</td>
</tr>
<tr>
<td>Our People</td>
<td>Pharma Industry uncertainty</td>
</tr>
<tr>
<td>Data</td>
<td>Hype (again)</td>
</tr>
<tr>
<td>Moore’s law/GPUs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Weaknesses</strong></th>
<th><strong>Opportunities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow progress in fundamental science</td>
<td>“Principles of Computational Drug Design”</td>
</tr>
<tr>
<td>Poor alignment of Academia and Industry</td>
<td></td>
</tr>
<tr>
<td>Performance Plateau of our techniques</td>
<td>PPPs</td>
</tr>
</tbody>
</table>

*Computer-aided molecular design under the SWOTlight*
Darren V. S. Green, Andrew R. Leach, Martha S. Head
Surely Computation can Help? What have we learnt?

**Strengths**
- Small, well-connected community
- Our People
- Data
- Moore’s law/GPUs

**Weaknesses**
- Slow progress in fundamental science
- Poor alignment of Academia and Industry
- Performance Plateau of our techniques

**Threats**
- Small, aging, community
- Pharma Industry uncertainty
- Hype (again)

**Opportunities**
- “Principles of Computational Drug Design”
- PPPs

*Computer-aided molecular design under the SWOTlight*
Darren V. S. Green, Andrew R. Leach, Martha S. Head
Can we predict this? A molecule that binds to the protein mostly through the water network!
Sampl1, Docking Mode Prediction

Cross-Docking Pose Prediction

Number of Methods

- < 0.5
- 0.5-1.0
- 1.0-1.5
- 1.5-2.0
- > 2.0

Median (RMSD - DPI)

- INK3
- Urokinase

Eric-Dock and Marti-Dock!
Surely Computation can Help? What have we learnt?

**Strengths**
Small, well-connected community
Our People
Data
Moore’s law/GPUs

**Threats**
Small, aging, community
Pharma Industry uncertainty
Hype (again)

**Weaknesses**
Slow progress in fundamental science

*Poor alignment of Academia and Industry*
Performance Plateau of our techniques

**Opportunities**
“Principles of Computational Drug Design”
PPPs

*Computer-aided molecular design under the SWOTlight*
Darren V. S. Green, Andrew R. Leach, Martha S. Head
**Compound Quality:**
Are you looking for a “probe”, a “tool”, a drug “publication”.... or a “publication”....

**REPROVIS-DB: A Benchmark System for Ligand-Based Virtual Screening Derived from Reproducible Prospective Applications**

Peter Ripphausen, Anne Mai Wassermann, and Jürgen Bajorath*

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms Universität, D¨uhlmannstr. 2, D-53113 Bonn, Germany

**ABSTRACT:** Benchmark calculations are essential for the evaluation of virtual screening (VS) methods. Typically, classes of known active compounds taken from the medicinal chemistry literature are divided into reference molecules (search templates) and potential hits that are added to background databases assumed to consist of compounds not sharing this activity. Then VS calculations are carried out, and the recall of known active compounds is determined. However, conventional benchmarking is affected by a number of problems that reduce its value for method evaluation. In addition to often insufficient statistical validation and the lack of generally accepted evaluation standards, the artificial nature of typical benchmark settings is often criticized. Retrospective benchmark calculations generally overestimate the potential of VS methods and do not scale with their performance in prospective applications. In order to provide additional opportunities for benchmarking that more closely

Many promiscuous/interference chemotypes published *time and time again*:
http://blogs.sciencemag.org/pipeline/archives/2012/06/01/return_of_the_rhodanome

None of these papers move the science of computational chemistry forward
The world view of an industrial computational chemistry group

...despite the real limitations in our underlying methodology, computational chemistry still makes a significant impact on drug discovery. Much of the reason can be ascribed to the fact that very rarely are computational chemistry methods used without any intervention and guidance from a scientist often with significant knowledge of their target and often many years of experience in the application of different computational chemistry methods.

Computer-aided molecular design under the SWOTlight
Darren V. S. Green, Andrew R. Leach, Martha S. Head
Confidence limits, error bars and method comparison in molecular modeling. Part 1: The calculation of confidence intervals

A. Nicholls

Received: 13 January 2014/Accepted: 14 May 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Computational chemistry is a largely empirical field that makes predictions with substantial uncertainty. And yet the use of standard statistical methods to quantify this uncertainty is often absent from published reports. This means “everything”, e.g. it could mean all members of a set, or all (infinite) repeats of an experiment. When we test predictive software we hope the average over a set of systems represents what we might get from the population.
Will computation improve? Yes

- Structure based drug discovery
  - Simulations, QM etc

- Crystal form prediction
  - Solvation energies

Diagram:
- PC1
- Solubility
- Potency
- Safety
- Absorption
- Metabolic stability
Will computation help? Maybe

- Safety
- Potency
- Toxicology prediction
- Off target effects
- Absorption
- Prediction of permeation
- Active transport
- Solubility
- Metabolic stability
- Metabolite prediction
- PC1
Summary

– Computational chemistry is a key component of modern drug discovery
  – With great reliance on experience & know how of individual computational chemists

– Computation is beginning to find application for a wide variety of problems, not just small molecule discovery

– We have a long way to go before we can replicate what is done in other industries
  – But it is a journey we must begin
Acknowledgements

– Paul Bamborough
– Chris Luscombe
– Stephen Pickett
– Jameed Hussain
– Ceara Rea
– Rob Young
– Alan Hill
– Eric Manas
– Marti Head
Thank you for the invitation!