Bioisosteres and Scaffold Hopping in Medicinal Chemistry

Nathan Brown
Group Leader, In Silico Medicinal Chemistry
Cancer Research UK Cancer Therapeutics Unit
Division of Cancer Therapeutics
The Institute of Cancer Research, London

Chemoinformatics Strasbourg Summer School 2014
Thursday 26th June 2014

Making the discoveries that defeat cancer
In Silico Medicinal Chemistry

What is a Bioisostere?

Bioisosteres

- Structural moieties with broadly similar shape and function
- Function should be biological but modulate other properties
- **Bioisosteric replacement**: replacement of functional groups

Molecular Scaffolds

- *Subset of bioisosterism*
- Identification of the core functional or structural element
- **Scaffold hopping**: replacement of core element

The *molecular interactions* must be maintained

- Important to mimic **shape** and **function**

Why Bioisosteres?

Many properties can be modulated with appropriate bioisosteres:

- Improved selectivity
- Fewer side effects
- Decreased toxicity
- Improved pharmacokinetics: solubility/hydrophobicity
- Increased metabolic stability
- Simplified synthetic routes
- Patented lead compounds

Drug Design is Inherently a Multiobjective Optimisation Problem

Why Bioisosteres?

- Bioisosteres
- Known Medicinal Chemistry Space
- Less Interesting Chemistry Space
- Potential False Positives

Biological Activity

Chemical Structure Similarity
Why Bioisosteres?

- Bioisosteres
- Known Medicinal Chemistry Space
- Less Interesting Chemistry Space
- Potential False Positives

Biological Activity

Chemical Structure Similarity
Why Bioisosteres?

![Diagram showing the relationship between chemical structure similarity and biological activity, with Bioisosteres and Known Medicinal Chemistry Space on one axis, and Less Interesting Chemistry Space and Potential False Positives on the other.]
Why Bioisosteres?

The octet theory of valence indicates that if compounds having the same number of atoms have also the same total number of electrons, the electrons may arrange themselves in the same manner. **In this case the compounds or groups of atoms are said to be isosteric.** Such compounds should show remarkable similarity in physical properties, that is, in those properties which do not involve a separation of the atoms in the molecule.

---

<table>
<thead>
<tr>
<th>Type</th>
<th>List of Isotera.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>H-, He, Li+</td>
</tr>
<tr>
<td>2.</td>
<td>O−, F−, Ne, Na+, Mg++, Al+++</td>
</tr>
<tr>
<td>3.</td>
<td>S−, Cl−, A, K+, Ca++</td>
</tr>
<tr>
<td>4.</td>
<td>Cu+, Zn++</td>
</tr>
<tr>
<td>5.</td>
<td>Br−, Kr, Rb+, Sr++</td>
</tr>
<tr>
<td>6.</td>
<td>Ag+, Cd++</td>
</tr>
<tr>
<td>7.</td>
<td>I−, Xe, Cs+, Ba++</td>
</tr>
<tr>
<td>8.</td>
<td>N2, CO, CN−</td>
</tr>
<tr>
<td>9.</td>
<td>CH4, NH3</td>
</tr>
<tr>
<td>10.</td>
<td>CO2, NO, N2+, CNO−</td>
</tr>
<tr>
<td>11.</td>
<td>NO3−, CO3−</td>
</tr>
<tr>
<td>12.</td>
<td>O2</td>
</tr>
<tr>
<td>13.</td>
<td>HF, OH−</td>
</tr>
<tr>
<td>14.</td>
<td>ClO−, SO3−, PO4−</td>
</tr>
<tr>
<td>15.</td>
<td>ClO2−, SO4−, PO4−</td>
</tr>
<tr>
<td>16.</td>
<td>SO2, PO2−</td>
</tr>
<tr>
<td>17.</td>
<td>SO3−, PO4−</td>
</tr>
<tr>
<td>18.</td>
<td>SiO4−, PbO4−</td>
</tr>
<tr>
<td>19.</td>
<td>MnO−, CrO−</td>
</tr>
<tr>
<td>20.</td>
<td>SeO4−, AsO4−</td>
</tr>
</tbody>
</table>
Friedman first coined the term bio-isosteric in 1951:

“We shall term compounds “bio-isosteric” if they fit the broadest definition for isosteres and have the same type of biological activity.”

Isosterism and Molecular Modification in Drug Design

By C. W. Thornber

Imperial Chemical Industries Limited, Pharmaceuticals Division, Mereside, Alderley Park, Macclesfield, Cheshire, SK10 4TG

The element of a molecule being modified may have one or more of the following roles.

(i) **Structural.** If the moiety has a structural role in holding other functionalities in a particular geometry, parameters such as size and bond angle will be important. The moiety may be buried deep in the molecule and have little contact with the external medium.

(ii) **Receptor interactions.** If the moiety to be replaced is concerned with a specific interaction with a receptor or enzyme its size, shape, electronic properties, $pK_a$, chemical reactivity, and hydrogen bonding will be the important parameters.

(iii) **Pharmacokinetics.** The moiety to be replaced may be necessary for the absorption, transport, and excretion of the compound. In this case lipophilicity, hydrophilicity, hydrogen bonding, and $pK_a$ are likely to be important.

(iv) **Metabolism.** The moiety may be involved in blocking or aiding metabolism. In this case chemical reactivity will be an important parameter. For example chloro and methyl substituents on a benzene ring may be interchangeable for certain purposes but the toluene derivative can be metabolized to a benzoic acid and may therefore have a shorter half-life or unexpected side effects.

(A) A given molecular modification may allow some, but probably not all of the parameters (a)–(h) to be kept the same.  
(B) Whether the same or a different biological activity results from the replacement will be governed by the role(s) which that moiety fulfils in the molecule and whether parameters affecting that role have been disturbed.  
(C) From (A) and (B) it follows that what proves to be a good bioisosteric replacement in one series of compounds will not necessarily be useful in another.

| Table 1 |
|---|---|---|---|---|---|---|---|---|
| 1) Univalent atoms and groups | F | OH | NH₂ | Me | Cl |
| | SH | Ph₂ | I | Br | F |
| 2) Bivalent atoms and groups | O | CO₂R | S | COSR | Se | CH₃ | CONH₂ | H |
| 3) Tervalent atoms and groups | N≡ | –CH≡ | –As≡ |
| 4) Quadivalent atoms | –C≡ | –Si≡ |
| 5) Ring equivalents | –CH=CH= | S | e.g. benzene, thiophene |
| | –C≡ | –N≡ | e.g. benzene, pyridine |

(a) Size.  
(b) Shape (bond angles, hybridization).  
(c) Electronic distribution (polarizability, inductive effects, charge, dipoles).  
(d) Lipid solubility.  
(e) Water solubility.  
(f) $pK_a$.  
(g) Chemical reactivity (including likelihood of metabolism).  
(h) Hydrogen bonding capacity.
Exploration versus Exploitation

**Exploration**

“... includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation.”

**All Exploration:** “...the costs of experimentation without any of its benefits.” Undeveloped ideas, little distinctive competence.”

**Exploitation**

“... includes such things as refinement, choice, production, efficiency, selection, implementation, execution.”

**All Exploitation:** “Locked-in to suboptimal equilibria (local maxima). Can’t adapt to changing circumstances.”

Feedback to exploitation occurs much more quickly. Increasing returns can lead to lock-in at a suboptimal equilibrium.

“...these tendencies to increase exploitation and reduce exploration make adaptive processes potentially self-destructive.”

Exploration versus Exploitation
Exploration versus Exploitation
Exploration versus Exploitation
Exploration versus Exploitation

Exploration Enabled Through Introduction of ‘Controlled Fuzziness’ of Bioisosteric Transformations and Descriptors
Methods to Identify Bioisosteres

- **Databases**
  - BIOSTER
  - ChEMBL – Matched Molecular Pairs
  - Cambridge Structural Database (CSD)

- **Descriptors**
  - Physicochemical properties
  - Molecular Topology
  - Molecular Shape
  - Protein Structure
BIOSTER Database – István Ujváry

- Database of ~26,000 bioisosteric transformations
- Bio-analogous pairs mined from the literature:
  - Systematic abstracting since 1970
- Compound pairs represented as hypothetical reactions
  - ‘bioisosteric transformations’
  - Compatible with most reaction-searching software

2. Distributed by Digital Chemistry: [http://www.digitalchemistry.co.uk](http://www.digitalchemistry.co.uk)
Matched Molecular Pairs

- Identification of molecules that differ in only one position
- Can suggest structural changes to modulate biological or physicochemical properties

MMP Transformation: $H \nrightarrow CF_3$
Bioisosteric Similarity Methods

Physicochemical Properties

Molecular Topology

Molecular Shape

Protein Structure

Peter Ertl

James Mills

ROCS

USR

Cresset

CATS

Hopfen

Similog

radius

atoms
Case Study: Bioisosteric Replacement

X = Br, Cl, CN, CF₃ equivalent
X = H: 10 to 50 fold weaker
X = large group: inactive

Benzyl-type linker optimal

Solvent accessible

Butressed against hinge
*Ortho* substitution poor
*Meta* tolerated but weaker

320 Compounds already made: What is the learning?
Unbiased and objective analysis
Focus on enzyme potency and cell penetration
Generation of a Virtual Library

- Preferred $R_2$ and $R_3$ groups from Free-Wilson analysis.
  - Introduce other ideas from bioisosteric replacements
  - $X = \text{Cl}, R_2 = 54, R_3 = 49$
  - > 2600 possible compounds
- Filter to remove compounds that:
  - Have > 1 basic centre
  - Have TPSA > 100
  - Have $A\log P$ > 3.5
  - Have MW > 520 Da.
  - Have > 2 HBD
  - 1500 compounds remaining

Easy to generate ideas: Picking which ones to make is much harder
Alexander Crum Brown defined the following relationship between:

- $\Phi$, the physiological action, and
- $C$, the chemical constitution of a molecule

$$\Phi = f(C)$$

Predictive Modelling

Predictive Modelling

Build naïve Bayesian model
FCFP_6 fingerprint molecular descriptors
Active threshold set at:
• 10 nM for enzyme IC$_{50}$
• 300 nM for cell IC$_{50}$
Training set and test set ($n = 320$)

Molecules scored by predicted activity/inactivity
• Partition dataset into training and test sets
• Derive statistical models

Predict Activity for 1500 Virtual Molecules

Prioritise the best molecules to make first
Predictions on Virtual Compounds

Make some of the preferred compounds first
Example of Multiobjective Prioritisation Using Bioisosteric Replacements

Ligand

Chemical Tool

Potential Drug

Optimal combination of $R_2$ and $R_3$ delivers desired profile

MW = 497
AlogP = 2.9
Aurora A = 42 nM
MLM unstable

MW = 551
AlogP = 4.1
FLT3 = 4 nM
Aurora A = 15 nM
MLM = 60% remaining
F = 100% mouse
HLM = 18% remaining
hERG IC50 = 3 uM

MW = 456
logD = 3.8
FLT3 Ki = 6 nM
Aurora A Ki = 7 nM
MLM = 70% remaining
F = 100% (mouse)
HLM = 90% remaining
hERG IC50 > 30 uM
Case Study: Scaffold Hopping

Why do we need a definition?

- Scaffolds are often the synthetic invariant in lead optimization
- Library Analysis
  - Scaffold diversity
- Scaffold Hopping
  - Subset of bioisosteric replacement

What do we need in a definition?

- Objective and invariant
  - Their definition derives solely from information in the molecule

Case Study: Scaffold Hopping

- Scaffold Hopping
- Known Medicinal Chemistry Space
- Less Interesting Chemistry Space
- Potential False Positives

Chemical Structure Similarity vs. Biological Activity
Case Study: Scaffold Hopping

IC\textsubscript{50} = 25.2 μM
LE = 0.48

IC\textsubscript{50} = 25.1 μM
LE = 0.48

53% inhibition at 325 μM

IC\textsubscript{50} < 25 nM
LE = 0.28

IC\textsubscript{50} = 33.31 μM
LE = 0.31

IC\textsubscript{50} = 4.21 μM
LE = 0.28

Case Study: Scaffold Hopping

IC\textsubscript{50} = 25.2 µM
LE = 0.48

53% inhibition at 325 µM

IC\textsubscript{50} = 25.1 µM
LE = 0.48

Potential False Positives
Known Medicinal Chemistry Space

Less Interesting Chemistry Space

Scaffold Hopping

% Inhibition at 40 µM

Tanimoto Similarity

IC\textsubscript{50} < 25 nM
LE = 0.28

IC\textsubscript{50} = 4.21 µM
LE = 0.31

IC\textsubscript{50} = 33.31 µM
LE = 0.31

Nine active SHv3 compounds have been soaked with TTK apo crystals

Structures determined from X-ray crystallography

Four active SHv3 compounds have been confirmed with co-crystal structures

X-ray Co-crystal Structures

- Nine active SHv3 compounds have been soaked with TTK apo crystals
- Structures determined from X-ray crystallography
- Four active SHv3 compounds have been confirmed with co-crystal structures

IC$_{50}$ = 8.27µM
LE = 0.53

PDB: 4BHZ
Resolution = 2.85 Å

Conclusions

Bioisosterism has seen more than a century of innovation
• Remains a difficult concept to define accurately, however...
• Databases of bioisosteric transforms routinely available
• Molecular descriptors allow for the exploration and validation of structurally disparate replacements

Scaffold Hopping is a subset of bioisosteric replacement
• Ability to successfully move away from problematic scaffolds
• Important to maintain exit vector geometries
Acknowledgements
Bioisosteres and Scaffold Hopping

- Principles of bioisosteres
- Scaffolds: Identification, Representation, Diversity, and Navigation
- Data Mining
- Methods
  - **Bioisosteres**: Physicochemical, Topology, Shape, Protein
  - **Scaffold Hopping**: CATS, Molecular Interaction Fingerprints
- Case Studies
  Abbott, AstraZeneca, BMS, CCDC, Cresset, Digital Chemistry, EBI, Eli Lilly, ETH-Zurich, GSK, ICR, MRCT, Novartis, Pfizer, UCB Celltech, Bonn, Cambridge, Manchester, Sheffield, Strasbourg, Vanderbilt