Molecular Obesity, Potency and other Addictions in Drug Discovery

Mike Hann
Bio-Molecular Structure
MDR Chemical Sciences
GSK Medicines Research Centre
Stevenage, UK

Mike.M.Hann@gsk.com
Patients are still in need of more effective drugs for many diseases.
Payers are increasingly only prepared to pay for innovative rather than derivative drugs.
Investors believe they can get a better return on investment elsewhere.
Legislators are demanding that only the very safest possible drugs are licensed.
Researchers are equally frustrated—despite all the accumulated knowledge from the new technologies, the challenge of successfully navigating through everything to find novel drugs seems to get harder rather than easier.
The fruit is possibly not as low hanging as before.
The consequence of all this is that the average cost is now $1.8bn.!

We owe it to patients and society to do better than this.

*How to improve R&D productivity: the pharmaceutical industry’s grand challenge.
Emergence of rules of thumb as guidance

**Permeability/solubility**: Pfizer analysis of existing drugs and oral absorption profile
- Mol Wt <500, LogP <5, OH + NH count <5, O + N count <10: 90% of oral drugs do not fail more than one of these rules.
- Lipinski *Rule of 5*

**Receptor Promiscuity**: AZ analysis of 2133 compounds in >200 Cerep Bioprint® assays
- cLogP < 3 decreases risk; > 4 increases risk; bases/quats>> neutrals > acids
- *Lipophilic Ligand Efficiency* LLE = pIC \textsubscript{50} – cLogP >5 for toxicity risk reduction
- AZ LLE >5 rule.

**Receptor Promiscuity**: Roche analysis of 213 compounds profiled at Cerep
- Pronounced promiscuity not observed below a threshold cLogP of 2. Increased promiscuity with increased calculated basicity.

**Toxicity**: Pfizer *in vivo* tolerability data on 245 compounds
- cLogP < 3 & TPSSA > 75 give 6-fold reduced *in vivo* toxicity vs. >3 & <75; 24-fold for bases
- Pfizer 3/75 rule

**ADMET**: GSK analysis of ~30,000 GSK compounds yielded simple rules of thumb for the effect of physchem parameters on solubility, permeability, bioavailability, volume of distribution, clearance, hERG inhibition, PGP efflux & P450 inhibition
- Mol Wt <400 & cLogP <4 reduces ADMET risks compared to >400 & >4
- GSK 4/400 rule
Some other things we have learnt

• **Size & permeability:**
  The larger a “small” molecule is, the more lipophilicity it is likely to need to permeate membranes.

<table>
<thead>
<tr>
<th>Molecular weight</th>
<th>AZ logD</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;300</td>
<td>&gt;0.5</td>
</tr>
<tr>
<td>300-350</td>
<td>&gt;1.1</td>
</tr>
<tr>
<td>350-400</td>
<td>&gt;1.7</td>
</tr>
<tr>
<td>400-450</td>
<td>&gt;3.1</td>
</tr>
<tr>
<td>450-500</td>
<td>&gt;3.4</td>
</tr>
<tr>
<td>&gt;500</td>
<td>&gt;4.5</td>
</tr>
</tbody>
</table>


• **Pre-clinical & clinical survival:**

Larger & more lipophilic molecules have reduced chances of survival in pre-clinical & clinical phases

Is MW or logP the source of promiscuity?

Graph showing series of pie charts in different cLogP and MW bins for a set of approximately 2500 compounds tested in more than 490 assays. The size of each pie chart represents the average number of hits for compounds in that pie, where a hit is defined as a pXC50 value of 5 or higher. The colours indicate the proportion of compounds within each pie having particular numbers of hits (red: <5; blue: 5-15; yellow: 15-25; black: >25), where a hit is defined as activity greater than 10mM in any of the ~490 assays.

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

RJ Young, DVS. Green, CN. Luscombe, AP Hill, Drug Discovery Today, 16 (17/18), 2011, 822-830

---

### Property Forecast Index (PFI) - a useful overall guide to where to look for developable compounds

**PFI = mChromLogD$_{pH7.4}$ + #Aromatic rings**

---

**TABLE 2**

Percentages of compounds achieving defined target values in the various developability assays categorised by PFI or iPFI bins

<table>
<thead>
<tr>
<th>Assay / target value</th>
<th>PFI = mChrom log D$_{pH7.4}$ + #Ar</th>
<th>iPFI = mChrom log D$_{pH7.4}$ + #Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solubility &gt;200 μM</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td>%HSA &lt;95%</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>2C9 pIC$_{50}$ &lt;5</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>2C19 pIC$_{50}$ &lt;5</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>3A4 pIC$_{50}$ &lt;5</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Cl$_{inh}$ &lt;3 ml/min/kg</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>Papp &gt;200 nm/s</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>hERG pIC$_{50}$ &lt;5</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>(+1 charge)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promiscuity &lt;5 hits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with pIC$_{50}$ &gt;5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Colouring refers to the % chance of achieving benchmark value in that PFI bin: green, >67%; yellow, 34–67%; and red, <33%.

---

**Sweet spot for permeability is in conflict with other desirable properties!**
What we have come to know (or rediscover!)

- Large and particularly lipophilic molecules are increasingly seen as bad - again!
- Cell penetration of larger molecules needs increasing lipophilicity
- Lipinski’s 500/5 for oral bioavailability is increasingly seen as far too lenient when it comes to the wider ADMET issues.
  - We should be thinking 400/4 or PFI <6 as better indicators of the space with highest probability of successfully developing a drug
  - and even smaller for leads as starting points!

<table>
<thead>
<tr>
<th>Average property values for the Sneader lead set, average change on going to Sneader drug set and percentage change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av # arom</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>1.3</td>
</tr>
<tr>
<td>Av # HBA</td>
</tr>
<tr>
<td>2.2</td>
</tr>
<tr>
<td>Av MW</td>
</tr>
<tr>
<td>272</td>
</tr>
</tbody>
</table>
Where have we come from and why has it got like this?

And what are the current attempts to try and improve?

In hindsight, the rush to numbers and “better/new tools” as a solution to productivity obscures our collective memory and experience!

The expanding “sciences” of Medicinal Chemistry and drug discovery
The curse of Molecular Obesity

- The tendency for drug discovery molecules to become too large and too lipophilic for their own good during lead optimisation through the quest for potency and specificity.
- It presents a high risk to the future “health” of the compound as a drug candidate.
- As with medical obesity, which is measured by Body Mass Index BMI, we now make use of indices such as Ligand Efficiency Index LE and Lipophilic Ligand Efficiency Index LLE to help identify and control the problem.

Indices as guideposts for life and drug discovery

- **Body Mass Index (BMI)** = human weight / height^2

- **Ligand Efficiency Index (LE)**
  - Potency in kcal/mol (= -1.37logKd) normalised by the number of heavy atoms
  - An ‘idealised’ compound with 1nm pIC50 and 30 heavy atoms has LEI = 0.42
  - An ‘okay’ compound with 10nm pIC50 and 38 heavy atoms (MW 500) has LEI = ca. .3

- **Ligand Lipophilicity Efficiency Index (LLE)**
  - Potency normalised by lipophilicity
  - LLE = pIC50 – clogP (typical good value are 5-7 for nanomolar potency)
  - During optimisation potency should increase more than just that due to bulk logP effects. Particularly true with membrane bound targets.

- **LLE_{Astex} = 0.11*ln(10)*RT(logP-log(K_d or pK_i or IC_{50}))/HA**
  - Lipophilic efficiency assessment for fragments
  - Scale fixed to be similar to LE so .3 is a base level number to aim for.

- **Binding Efficiency Index (BEI)**
  - Potency (pIC50) normalised for MW
  - An ‘idealised’ compound with 1nm pIC50 and MW of 0.333 kDA has BEI = 27

- **Surface Binding Efficiency Index (SEI)**
  - Potency normalised for Polar Surface Area
  - An ‘idealised’ compound with 1nm pIC50 and PSA of 50A^2 has SEI = 18
The link of potency and molecular obesity

- Potency can improve the therapeutic index, specificity and help reduce dosage
  - all good things but if we grow potency in the wrong way molecules can get very obese

- We can easily measure and optimise against potency
  - Potency results come back quickly and we react to them with decisions as to what to make next
  - It satisfies the “we are making progress” paradigm!
  - Unreasonable time pressures can make this seem like an end in its own right!

- Potency tends to correlate with increasing MW and logP for most series because we make more interactions.
  - Size needs lipophilicity to pass through membranes
  - Adding MW is easier than subtracting in synthetic chemistry!!
  - Most medicinal chemists are synthetic organic chemists!
**Organic synthesis and purification favours lipophilic molecules**


*Change in mean clog P*

**ΔLogP (mean, designed vs actual)**

**ΔLogP**

Difference in LogP of array of molecules designed and those that actually got made

**Different Arrays synthesised**

0.40 0.20 0.00 0.20 0.40 0.60 0.80 1.00

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Different Arrays synthesised
The link of potency and molecular obesity - cont’d?

Typical order of events in a drug discovery screening cascade

1. Biochemical assay
2. Cellular assay
3. in vivo assay

– We often start with isolated protein in a biochemical assay with none of the environment of more phenotypic assays to help balance the physicochemical properties.

– We look for early signs of cellular potency in our screening cascades - this needs both some intrinsic potency and cellular penetration

  – Both of these are very easily driven by increasing logP.

– Once we get cellular activity the damage may already be done if we do not revisit to look at how we got there.
The link of potency and molecular obesity - cont’d?

- **Structure based design** using crystal structures is a fantastic tool but it can easily draw you into the specifics of building potency rather than looking at the wider challenges at the same time.
The origins of potency – enthalpy and entropy

$$\Delta G = \Delta H - T \Delta S = -RT \ln K_d$$

- Measurements of Free Energy show that for synthetic ligands, potency correlates with buried apolar surface area (ie size of interface and it’s lipophilicity)
- Buried apolar surface area (lipophilicity) is an easier way to get potency than through buried polar surface area
- We need to be very careful that we are not drawn down the path of using too much lipophilicity as a quick fix for potency!

Fig. 2. The Gibbs free energy of binding for protein-ligand interactions correlates well with reduction in hydrated apolar surface area upon complex formation ($R^2=0.65$). The key indicates the proteins involved in each interaction. A linear least-squares fit to the data gives an intercept of 19.4±1.8 kJ mol$^{-1}$ and a slope of 0.049±0.005 kJ mol$^{-1}$ Å$^{-2}$. The thin dotted lines represent the 95% confidence intervals of the fit.

Fig. 3. (a) The apolar ($\Delta$CSA$_{apolar}$) and polar ($\Delta$CSA$_{polar}$) contributions to the total reduction in solvent ASA ($\Delta$CSA$_{total}$) upon complex formation diverge substantially with increasing extent of the binding interface. The interactions between proteins and synthetic ligands are represented by squares, whereas biological and other interactions are represented by circles. The intercept, slope and $R^2$ of $\Delta$CSA$_{apolar}$ are $-78.0\pm15.3$, $0.086\pm0.03$ and 0.93, whereas those of $\Delta$CSA$_{polar}$ are $78.0\pm15.3$, $0.14\pm0.03$ and 0.29, respectively. (b) Ligands themselves show similar but less marked trends in CSA$_{apolar}$ and CSA$_{polar}$.

---

Potency needs both entropic and enthalpic binding?!

- Very potent compounds seem to require very significant entropic contributions to the overall free energy.

- Broadly speaking enthalpy equates to polar interactions while a key contributor to entropy is lipophilic interactions => Molecular Obesity.

- Make sure you have got the most out of your polar enthalpies early in lead optimisation => Fragment approach.


- The challenge of medicinal chemistry – the role for nature and nurture in lead discovery and optimization
  M Hann and G Keseru. Accepted for Nature Reviews in Drug Discovery
Why is it so difficult?

1. Solvation accountancy is challenging 
2. Enthalpic interactions are directional, have more information content, and are harder to get right due to their complexity

The molecular complexity approach to thinking about interactions

Receptor + - - + + + - - + + + - - - 
Ligand    ++ - -

• Coping with complexity in molecular design, A.R. Leach and M.M. Hann chapter in “de novo Molecular Design” ed – G. Schneider, 2014 (Wiley-VCH)
Complexity and high information content

Receptor: + - - + + - + - + -

Ligand: + + - - - +

- = attractive primary interaction
or = attractive secondary interaction
or = repulsive secondary interaction

• High information content
• High Shannon entropy
• Difficult to shift < >
• Hard to get correct

Complexity and low information content

Receptor: + + + + + + + + + +

Ligand:

= attractive primary interaction
\ /  \ or= attractive secondary interaction

• Low information content
• Low Shannon entropy
• Easy to shift < >
• Easy to get correct
Information content *per unit surface area*

- **High information content**
  - E.g. amide
  - Directional interactions
  - Difficult

- **Adaptable information content**
  - E.g. aromatic
  - Intermediate

- **Polarisable**

- **Low information content**
  - E.g. Aliphatic
  - Lipophilic interactions
  - Easy!
More Molecular Obesity related issues

- Every increase of logP by one unit increases by one order of magnitude the amount of compound present in membranes or bound to lipophilic proteins, etc..
  - Home to key signalling proteins (GPCRs, ion channels, transporters etc) reside. Likely local high concentrations play havoc with them. => promiscuity

- Lipophilicity is the antithesis of solubility – relying on formulation to get insoluble compounds on board is only going to aggravate the body!
  - Your body can’t easily eliminate lipophilic compounds (they are too insoluble!) so it has to work harder to make them more polar with higher energy species => toxicity
Developability Classification System DCS

Increasing dose/decreasing solubility = increasing Volume to dissolve the dose

The developability classification system: Application of biopharmaceutics concepts to formulation development
Using knowledge of physchem properties & dose – the Developability Classification System DCS

Top 100 Oral Drugs

- DCS 1 51%
  - Good solubility and permeability

- DCS 2A 17%
  - Dissolution rate limited

- DCS 2B 6%
  - Solubility limited

- DCS 3 23%
  - Good solubility, poor permeability

- DCS 4 3%
  - Poor solubility and permeability

Dose/ FaSSIF solubility ratio - i.e. Volume in ml to dissolve dose

Ask yourself where your lead optimisation compounds fit on this plot?!!

If the solubility of compound is 0.05 mg/ml (100uM) then only 50mg will dissolve in this volume!

The developability classification system: Application of biopharmaceutics concepts to formulation development
Historically potency is not everything either!

How many drug targets are there? Overington, John P.; Al-Lazikani, Bissan; Hopkins, Andrew L. Nature Reviews Drug Discovery (2006), 5(12), 993-996
Pfizer three pillars analysis


Each pillar is necessary but not independently sufficient for efficacy
Typical "cell drop-off" effect compared to biochemical enzyme data - is your compound getting to the site of action?

What really matters at the end of the day is dosage!

– Hence the interest in Drug Efficiency which tells you how much of your dose actually is available in the biophase of interest.

\[
\text{DRUGeff} = \text{Biophase Concentration} \ast \frac{100}{\text{Dose}}
\]


– And more recently the use of Drug Efficiency Index as a strategy towards low therapeutic dose

\[
\text{DEI} = \log[\text{DRUGeff(%)}] + pKd
\]

DEI is a correction of the in vitro affinity by the in vivo pharmacokinetic potential.

It is a simple descriptor directly connected to efficacy and therapeutic dose with the potential to probe the balance between in vitro affinity and ADME properties.

Application of drug efficiency index in drug discovery: a strategy towards low therapeutic dose. Montanari, Dino; Chiarpin, Elisabetta; Gleeson, Matthew Paul; Braggio, Simone; Longhi, Raffaele; Valko, Klara; Rossi, Tino. Expert Opinion on Drug Discovery, Volume 6, Number 9, September 2011, pp. 913-920(8)
• Clozapine = 99.97% of drug is not being used for target engagement!
• What else is it getting up to???
• High affinity tempts low Drug$_{eff}$

And why do we waste compound?

- We make very potent and lipophilic compounds which probably have very low free concentration at site of action (ie low $K_{pu}$)
- We assume the “free drug hypothesis” will allow compound to get to the site of action
  - We measure blood concentration and then use AMPA/CACO2 measurements or logD models to guide our medicinal chemistry
- But technology now exists to measure actual cellular concentration and disposition in early discovery
  1. Incubate cells with compound, wash, rupture, extract, quantify by MS
  2. MALDI/SIMS Imaging

   Rapid Measurement of Intracellular Unbound Drug Concentrations. A. Mateus et al. Mol. Pharmaceutics 2013, 10, 2467–2478
2. MALDI imaging in rodent lung slices showing compound distribution A. West and P. Marshall. GSK
Moving from:

Biochemical assay

Cellular assay

in vivo assay

To:

Drug gets there? 1

Engages target? 2

Elicits response? 3

Each pillar is **sequentially** necessary for efficacy but they need to be built in **parallel**
Nobody said this was easy!
Distribution in LE/LLE space of a range of CCR5 antagonists

The role of ligand efficiency metrics in drug discovery. Andrew L. Hopkins, György M. Keserü, Paul D. Leeson, David C. Rees and Charles H. Reynolds. NRDD 13, 2014, 105-121
Summary

– Medchem is a discipline and we should be Rigorous and Disciplined in making sure we make the very best molecules we can.
  – Not necessarily the most potent
  – Not necessarily the easiest to make
  – Not necessarily the quickest to find

– Molecular Obesity has been killing us
  – We are addicted to quick wins – e.g. potency and its consequences

– We have an increasing understanding of why and how to separate out the drivers to let molecules survive.
  – Start slim and stay fit! Control the risks!
  – Know where your compound is going in lead optimisation when you can still do something about it!

– Known knowns, known unknowns, unknown unknowns and ……. 
The part that Donald Rumsfeld forgot

*Unknown knowns*

- Those things that are known but we don’t know ourselves
- Those things that are known but we have forgotten
- Those things that are known but we choose to ignore

- Lets try not to ignore the medchem knowledge that has been gained at very considerable expense over many years!
Acknowledgements

Many at GSK present and past:
Darren Green, Klara Valko,
Andy Brewster, Ian Churcher,
James Butler, Rob Young,
Alan Hill, Stephen Pickett,
Anthony Taylor, Chris Luscombe,
Andy West, Peter Marshall,
Gareth Wayne, Bill Leavens,
Simon Readshaw, Laurie Gordon
Shenaz Bunally, Ian Reid
Martin Bayliss, et al..

Andrew Leach, Anne Hersey,
John Bradshaw, Gavin Harper,
Paul Gleeson, Dino Montanari,
Paul Leeson, Andy Hopkins et al

Mike Waring, Chris Lipinski,
George Keseru
Per Artursson and Andre Mateus et al.