Virtual Screening

Hugo Kubinyi
Germany

E-Mail kubinyi@t-online.de
HomePage www.kubinyi.de

Drug Research is ....

the Search for a Needle in a Haystack
The Sceptical Chemist
R. Lahana, Drug Discovery today 4, 447-448 (1999)

„How many leads have we got from combinatorial chemistry and high-throughput screening so far? - None!“ Wrong

„When trying to find a needle in a haystack, the best strategy might not be to increase the size of the haystack“ True

„Combinatorial chemistry has certainly failed to meet early expectations. Does this mean the technology has failed? Or does the problem lie in the manner in which the technology has been applied?“

M. Ashton and B. Moloney, Curr. Drug Discov. 2003 (8), 9-11

The Medicinal Chemistry Space

Virtual Screening Reduces the Size of the Haystack by Selecting:

**Compounds or libraries that are either**
- lead-like, or
- drug-like, or have the potential of oral bioavailability,
- or are similar to a lead,
**by**
- rules (e.g. Lipinski bioavailability rules),
- neural nets (e.g. drug-like character),
- pharmacophore analyses,
- similarity analyses,
- scaffold hopping, or
- docking and scoring

Favourable Drug Properties

High affinity and selectivity
Synthetic accessibility
No chemically reactive groups (garbage filter)
Oral bioavailability
  - Lipinski (Pfizer) „Rule of Five“: MW < 500, log P < 5, H donors < 5, H acceptors < 10
Favourable pharmacokinetics
Metabolism (e.g. no first pass)
Elimination pathway/s
Lack of side effects and toxicity

„A Hit is no Ligand is no Lead is no Drug“
**Intestinal Absorption and Bioavailability of Drugs**

Intestinal Absorption

a) transcellular pathway (passive diffusion)

b) paracellular pathway

c) carrier-mediated transport

d) transcytosis

**Lipinski (Pfizer) „Rule of Five“**

No good absorption to be expected if
MW > 500, log P > 5, H-bond donors > 5,
H-bond acceptors > 10

C. A. Lipinski et al., Adv. Drug Delivery Res. 23, 4-25 (1997);

---

**Lead-Like and Drug-Like Structures**

In their optimization to clinical candidates, leads most often become large and lipophilic.
Thus, leads should have

- molecular weight < 350 (450)
- low lipophilicity
- several positions for chemical variation
- sufficient affinity and selectivity

Drugs should have

- molecular weight < 500
- lipophilicity in the range log P = -2 to +5
- few hydrogen bond donors and acceptors

Neural Net Characterisation of the „Drug-like“ Character of Organic Compounds

Input layer $x_i$
Chemical descriptors

Weights $w_{i,k}$

Hidden-to-output weights $v_k$

Output layer $y = \text{"drug-likeness"}$
Drugs = 0.9, Nondrugs = 0.1

Training set: 5,000 WDI, 5,000 ACD compounds
Test set: 38,416 WDI, 169,331 ACD compounds
Filters: WDI duplicates, reactive compounds

Chemical descriptors:
120 Ghose-Crippen parameters
6 x C.ar, 5 x H(-C.ar), 1 x Cl(-C.ar)

J. Sadowski and H. Kubinyi, A Scoring Scheme for Discriminating Between Drugs and Non-Drugs, J. Med. Chem. 41, 3325-3329 (1998);
"Drug-like" Character

100 Top-Selling Drugs, 1997

% Compounds

Training sets

Test set predictions

"Drug scores" of top-selling drugs (year 1994)

<table>
<thead>
<tr>
<th>Drug</th>
<th>Score</th>
<th>Drug</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranitidine</td>
<td>0.78</td>
<td>Lovastatin</td>
<td>0.89</td>
</tr>
<tr>
<td>Enalapril</td>
<td>0.82</td>
<td>Diltiazem</td>
<td>0.73</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>0.53</td>
<td>Cimetidine</td>
<td>0.72</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>0.80</td>
<td>Cefaclor</td>
<td>0.67</td>
</tr>
<tr>
<td>Co-amoxiclav</td>
<td>0.80</td>
<td>Estrogenes</td>
<td></td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>0.68</td>
<td>Estrone</td>
<td>0.62</td>
</tr>
<tr>
<td>Clavulanic Acid</td>
<td>0.40</td>
<td>Equilin</td>
<td>0.73</td>
</tr>
<tr>
<td>Diclofenac</td>
<td>0.40</td>
<td>Ceftriaxon</td>
<td>0.97</td>
</tr>
<tr>
<td>Omeprazole</td>
<td>0.85</td>
<td>Cyclosporin</td>
<td>0.84</td>
</tr>
<tr>
<td>Ciprofloxazin</td>
<td>0.93</td>
<td>Famotidine</td>
<td>0.65</td>
</tr>
<tr>
<td>Nifedipine</td>
<td>0.76</td>
<td>Beclometason</td>
<td>0.65</td>
</tr>
<tr>
<td>Captopril</td>
<td>0.82</td>
<td>Salbutamol</td>
<td>0.65</td>
</tr>
<tr>
<td>Aciclovir</td>
<td>0.64</td>
<td>Sertraline</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Acetylsalicylic acid Score = 0.30!
Elimination of certain therapeutic classes

Filters for Virtual Screening

Garbage filter 90%
Druglike / Non-druglike 60%
Bioavailability 40%
Cytotoxicity :
hERG channel inhibition :
Antitargets :
  α1a (orthostatic hypotension) :
  D2 (extrapyramidal syndrome) :
  5-HT2c (obesity) :
  musc. M1 (hallucinations, memory) :
CYP inhibition (3A4, 2C9, 2D6) 0%?
Combinatorial Chemistry Sublibrary Selection

A library with 2 sites of chemical variation:
e.g. 7,262 carboxylic acids and 1,761 aldehydes
= 13×10^6 compounds

Problem:
Select a sublibrary with optimum balance of
good diversity,
high percentage of drug-like compounds
and cheap building blocks.
→ 10^{82} possible 15×15 sublibraries

Solution: Selection by a genetic algorithm

Sublibrary Selection by Genetic Algorithms

J. Sadowski, in:
Virtual Screening for Bioactive Molecules,
H.-J. Böhm and
G. Schneider, Eds.,
Wiley-VCH, 2000

Fitness function Φ =
f(diversity,
drug score,
price)
Feature Tree Similarity Searches

Query in 175,000 compounds in 1.5 cpu minutes

MTX vs. DHF

FTree Similarity

- more powerful than linear but about as efficient to compute
- not 3D but capturing 3D characteristics
- optimal matching computationally feasible
- a solution is more than a mere similarity value

Feature Tree Query Results for COX-2 Inhibitors

known inhibitors          plausible hits          known inhibitor related to the plausible hit

Feature Tree Query Results for H1 Antagonists and Antidepressants


TOPAS (TOPology-Assigning System)

Virtual Screening, Analogy and Optimization

**Template**

\[ K_i \text{ hKv1.5} = 0.11 \mu M \]

**2D Similarity Search**

\[ K_i \text{ hKv1.5} = 9.5 \mu M \]

**Chemical Analogy**

\[ \text{IC}_{50} \text{ hKv1.5} = 4.8 \mu M \]

**Optimization**

\[ \text{IC}_{50} \text{ hKv1.5} = 0.16 \mu M \]


---

Virtual Screening of \( \alpha_4\beta_1 \) Integrin Antagonists
With CATALYST

**Lead**

\[ R = -\text{Leu-Asp-Val-OH} \]

\[ \text{IC}_{50} = 0.6 \text{ nM} \]

**Best Results**

\[ \text{IC}_{50} = 67 \text{ nM} \]
\[ \text{IC}_{50} = 58 \text{ nM} \]
\[ \text{IC}_{50} = 1.3 \text{ nM} \]

A Virtual Screening Success Story

Comparison of the performance of high-throughput screening and virtual screening of potential leads of protein tyrosine phosphatase 1B (PTP1B):

a) High throughput screening of 400,000 compounds from a corporate collection → 300 hits < 300 µM,
   85 validated hits with IC\textsubscript{50} <100 µM
   = 0.021 % hit rate (many violate Lipinski rules)

b) Virtual screening of 235,000 commercially available compounds, using DOCK, version 3.5 → 365 high-scoring molecules,
   127 with IC\textsubscript{50} <100 µM
   = 34.8% hit rate (hits are more drug-like)


Shape-Based Virtual Screening for Type I TGFβ receptor (TßRI) Kinase Inhibitors

3D structure of SB 203 580 (IC\textsubscript{50} TßRI = 30 µM), in complex with p38, served as template
Pharmacophore search in 200,000 commercially available compounds: shape-based, two H bond acceptors, three out of four aromatic ring systems

87 hits, e.g.
IC\textsubscript{50} = 27 nM, K\textsubscript{d} = 5 nM

**Virtual Screening of Cdk2 Inhibitors**

Flexible docking of about 50,000 commercially available compounds (program LIDAUS) into cdk2 3D structure,

![Chemical structure](image1)

- Biological testing of 200 hits

![Chemical structure](image2)

- Chemical optimization (increase in affinity but decrease in selectivity)

S. Y. Wu et al., Structure 11, 399-410 (2003);

**Stepwise Virtual Screening**

- Aventis in-house compound repository
- 22,950 compounds
- Docking into an $\alpha_{1A}$ receptor model (GOLD, PMF)
- 300 top-scoring compounds
- Clustering, diversity selection
- 80 compounds tested, 37 hits with $K_i < 10 \mu M$

![Chemical structure](image3)

$\alpha_{1A}$ adrenergic receptor antagonist, $K_i = 1.4 \text{ nM}$

Stepwise Virtual Screening

250,251 NCI compounds (3D database)
- 3D pharmacophore search
- 6,727 hits
- docking into four conformational clusters of a D₃ receptor homology model
- 2,478 potential ligands
- elimination of known chemotypes by similarity
- 20 compounds tested, 8 hits with $K_i < 0.5 \, \mu M$

Dopamine D₃ receptor antagonist, $K_i = 11 \, nM$

Stepwise Virtual Screening

826,952 compounds from 8 structural databases
- MW, rot-bond filter, presence of pharmacophoric groups
- 131,967 hits
- 2D and 3D pharmacophore searches plus excluded volumes
- 11,109 hits
- FlexX-Pharm docking, DrugScore
- 1,000 highest-scoring compounds visually inspected, 7 selected

NK₁ neurokinin antagonist, $K_i = 251 \, nM$
A. Evers and G. Klebe, Angew. Chem. Int. Ed. 43, 248–251 (2004);
Stepwise Virtual Screening

259,747 ACD compounds
- Ro5 filter with MW < 350 and rot-bond < 9,
12,545 candidates
- presence of -COO⁻ or equivalent
1,261 hits
- 3D pharmacophore search (derived from binding site analysis)
- FlexX docking into 0.66 Å aldose reductase 3D structure
216 highest-scoring compounds, after clustering and visual inspection:
- 9 hits for biological testing

aldose reductase inhibitor, IC₅₀ = 2.4 µM


Virtual Screening of Carbonic Anhydrase Inhibitors

98,850 compounds (LeadQuest and Maybridge libraries)
- filter for Zn²⁺-binding anchor groups
5,904 hits
- 2D and 3D pharmacophore searches (derived from binding site analysis)
3,314 hits
- FlexS superposition with dorzolamide, followed by FlexX docking of 100 hits into carbonic anhydrase binding site
13 hits

X = S  
Kᵢ = 0.9 nM  
X = SO₂  
Kᵢ = 0.8 nM

Virtual Screening of TGT Inhibitors

800,000 compounds (from eight different databases)

- MW < 450 and rot-bond filter
- About 400,000 molecules
- Pharmacophore searches, followed by binding-site derived volume constraints
- Flexible docking (program FlexX) into two different binding site conformations
- 872 hits
- 9 hits, all biologically active

\[
\text{tRNA-guanine transglycosylase (TGT) inhibitor, } K_i = 0.25 \text{ \mu M}
\]


Stepwise Virtual Screening

560,000 compounds (subsection of AstraZeneca repository)

- MW, rot-bond filter, presence of hinge region binding motif
- FlexX-Pharm docking into ATP binding site
- 199,000 hits
- 250 highest-scoring hits
- Visual inspection for unrealistic conformations
- 103 compounds tested, 36 hits in the range 110 nM to 68 \mu M

\[
\text{Checkpoint kinase 1 (Chk-1) inhibitor, } IC_{50} = 450 \text{ nM}
\]

Virtual Screening of CK2 Inhibitors

- **400,000 compounds** (Novartis in-house compound collection)
- Docking into CK2 homology model (from CK2 of Zea mays)
- Highest-scoring hits
- 12 hits
- Filters: binding to hinge region, favorable conformations
- Biological testing
- 4 hits with > 50% inhibition at 10 µM

Protein kinase CK2 (casein kinase II) inhibitor, IC$_{50}$ = 80 nM


---

Virtual Screening of Akt1 (PKBα) Inhibitors

- **50,000 ChemBridge compounds**
- Flexible docking (program FlexX)
- Top 2,000 molecules
- Ranking by CSCORE, DrugScore, GoldScore, ChemScore (top 100-200 hits)
- Re-ranking (top 700 hits), visual inspection
- 100 top-scoring compounds

**IC$_{50}$ = 4.5 µM**

**IC$_{50}$ = 2.6 µM**

Virtual Screening: The Screensaver Project

coordinated by W. G. Richards, University of Oxford.

launched in April 2001, now >1.5 million PC's in >200 countries are connected to a virtual 65-teraflop machine, so far >100,000 h CPU time.

Cancer Project: 3.5 billion compounds docked to 12 potential antitumor targets (RAS, VEGF, SOD, Insulin receptor tyrosine kinase, COX-2, BCR-ABL, FGFR, CDK2, RAF, FPT, PTP1B and VEGFR1).

Anthrax Project: 3.5 billion compounds tested as potential YWWL tetrapeptide mimetics (run time: 24 days; results reported to UK and US government).

Summary and Conclusions

Virtual screening is a powerful tool to enrich libraries and compound collections.

A proper preprocessing of the compound database is of utmost importance.

Further experimental data and theoretical investigations are needed for better $pK_a$ estimations and better scoring functions.

Stepwise procedures (filters, pharmacophore searches, docking and scoring, visual inspection) are most efficient.

Fragment-based approaches are a promising new strategy in lead structure search and optimization.

References


