Virtual Screening - The Road to Success

Hugo Kubinyi

Germany

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

XIXth ISMC, Istanbul, Turkey, August 29-September 02, 2006
Strategies in Design

- no protein 3D structure, no ligands
  - combicheem, HTS, virtual screening
- protein 3D structure, no ligands
  - de novo design (protein flexibility !)
- no protein 3D structure, ligands
  - pharmacophores, (3D) similarity, (3D) QSAR
- protein 3D structure, ligands
  - structure-based design

LBFD  SBLD
Drug Research is ....

the Search for a Needle in a Haystack
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, or fit the binding site of a certain protein by rules (e.g. Lipinski bioavailability rules), neural nets (e.g. drug-like character), similarity analyses, pharmacophore analyses, scaffold hopping, or docking and scoring.
Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings

Christopher A. Lipinski*, Franco Lombardo, Beryl W. Dominy, Paul J. Feeney

Central Research Division, Pfizer Inc., Groton, CT 06340, USA

Received 9 August 1996; accepted 14 August 1996

Abstract

Experimental and computational approaches to estimate solubility and permeability in discovery and development settings are described. In the discovery setting 'the rule of 5' predicts that poor absorption or permeation is more likely when there are more than 5 H-bond donors, 10 H-bond acceptors, the molecular weight (MWT) is greater than 500 and the calculated Log P (CLogP) is greater than 5 (or MLogP > 4.15). Computational methodology for the rule-based Moriguchi Log P (MLogP) calculation is described. Turbidimetric solubility measurement is described and applied to known drugs. High throughput screening (HTS) leads tend to have higher MWT and Log P and lower turbidimetric solubility than leads in the pre-HTS era. In the development setting, solubility calculations focus on exact value prediction and are difficult because of polymorphism. Recent work on linear free energy relationships and Log P approaches are critically reviewed. Useful predictions are possible in closely related analog series when coupled with experimental thermodynamic solubility measurements.
"Drug-like" Character

100 Top-Selling Drugs, 1997

% Compounds

ACD

Training sets

WDI

Test set predictions

Drug Score
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% ?
Pharmacophore Generation and Searches

Catalyst (Accelrys)
   established tool for hypothesis generation and 3D searches

CATS topological pharmacophores (Roche)
   no 3D structures required

FTree (feature trees; BioSolveIT)
   no 3D searches required, ultrafast searches

LigandScout (inte:ligand)
   automated generation of bioactive pharmacophores from protein 3D structures
Problems in Pharmacophore Generation

Isomers, enantiomers, diastereomers

Superposition of flexible molecules

Ionisation and Dissoziation (Sadowski rules, ACS Boston, 2002)

Tautomeric and protomeric forms (program AGENT, ETH Zurich; ChemoSoft tautomer recognition, ChemDiv)

Acceptor properties of oxygen and sulfur atoms (esters, aromatic ethers, oxazoles, isoxazoles, thiazoles, etc.)
Pharmacophore Hypotheses - Histamine
Dissociation of Acids and Protonation of Bases

strong acids  CF$_3$COOH  
acids  arom. + aliph. COOH, CF$_3$SO$_2$NH$_2$, tetrazole  
weak acids  arom. OH, arom. SO$_2$NH$_2$  
neutral  aliph. -OH, -CONH$_2$  
weak bases  arom. NH$_2$, imidazole  
bases  aliph. NH$_2$  
strong bases  amidines, guanidines

$pK_a$ Values of Selected Organic Compounds

\[
\begin{align*}
\text{neutral} & \quad \text{pK}_a = 2.48 \\
pK_a &= 1.15 \\
pK_a &= 6.99 \\
pK_a &= 2.45 \\
pK_a &= 4.90
\end{align*}
\]

The Discovery of the DNA Double Helix

Summer 1952: Erwin Chargaff criticizes that Francis Crick and James Watson are ignorant about the structures of the bases.

Early 1953: Pauling publishes a DNA model with a phosphate core.
February 27, 1953: Jerry Donohue corrects the formulas of the bases.
February 28, 1953: Watson and Crick derive the correct DNA model.
April 02, 1953: Manuscript sent to Nature; published April 25, 1953.

Tautomeric Forms of an MMP-8 Inhibitor (1jj9)

"enantiomer I" \[ \text{prochiral form} \] "enantiomer II"

Glu 198 \[ \text{Ala161} \] Leu160

Ro 200-1770

H. Brandstetter et al., J. Biol. Chem. 276, 17405-17412 (2001)

Batimastat
Donor and Acceptor Properties of O and N
FTree (feature tree) Generation for Fast Similarity Searches

FTree Query Results for H1 Antagonists and Antidepressants

LigandScout (inte:ligand)

LigandScout (inte:ligand)

LigandScout (inte:ligand)

LigandScout Superposition: Zanamivir vs. GS 4071

1nnnc

2qwk
LigandScout Superposition: Zanamivir vs. GS 4071
LigandScout Superposition: Zanamivir vs. GS 4071

1nnc

2qwk
Computer-Aided Drug Design

H.-J. Böhm,
1992

LUDI
Problems in Docking and Scoring

Pre-processing of the protein
  lacking hydrogens, hydrogen bonds network,
  protonation states of his, lys, asp, glu

Pre-processing of the ligands
  protonation states, tautomers

Flexibility of the ligand (no serious problem)

Flexibility of the protein / binding site (the real problem)

Fuzzy scoring functions (the biggest problem)
HIV Protease, without a ligand

HIV Protease, with a ligand
Consideration of Water, Flexibility and Mobility

Diagram showing the interaction of a ligand with a receptor in an aqueous environment. The diagram includes terms such as $\Delta S_{rt}$, $\Delta S_{int}$, $\Delta H_{DW}$, and $\Delta H_{RW}$, representing changes in entropy and enthalpy, respectively. The ligand interacts with the receptor and is loosely associated with water molecules.
Factors to be Considered in Scoring Functions

Desolvation enthalpy and entropy (ligand and protein)
Protonation state of the ligand and the binding site
Distortion energy of the ligand and its binding site
Loss of translational and rotational degrees of freedom of the ligand
MEP + dielectric constant at the binding site
Dipole moment of the ligand and local dipole moment at the binding site
Binding enthalpy of the ligand-protein complex
Repulsive effects (e.g. -O⋯O-)
Inserted water molecules
Solvation enthalpy and entropy of the complex
Virtual Screening vs. High-Throughput Screening

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$_{50}$ < 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 are tested 127 with IC$_{50}$ < 100 µM = 34.8% hit rate (hits are more drug-like?)

Stepwise Virtual Screening

560,000 compounds (subsection of AstraZeneca repository)

- MW, rot-bond filter, presence of hinge region binding motif

199,000 hits

- FlexX-Pharm docking into ATP binding site

250 highest-scoring hits

- Visual inspection for unrealistic conformations

103 compounds tested, 36 hits in the range 110 nM to 68 µM

Checkpoint kinase 1 (Chk-1) inhibitor

IC$_{50} =$ 450 nM

Stepwise Virtual Screening

Aventis in-house compound repository

MW, rot-bond filter, 3D pharmacophore search

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 $K_i < 10 \, \mu M$

$\alpha_{1A}$ 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

259,747 ACD compounds

- Ro5 filter with MW < 350 and rot-bond < 9, presence of -COO⁻ or equivalent

12,545 candidates

- 3D pharmacophore search (derived from binding site analysis)

1,261 hits

- 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

Combinatorial Design of Carbonic Anhydrase Inhibitors

start structure

\[
\begin{align*}
\text{H}_2\text{N} & \quad \text{SO} \\
\text{S} & \quad \text{NH}_2 \\
\text{O} & \quad \text{O}
\end{align*}
\]

\(K_d = 120\ \text{nM}\)

optimized structure

\[
\begin{align*}
\text{H}_2\text{N} & \quad \text{SO} \\
\text{S} & \quad \text{O} \\
\text{NH} & \quad \text{N} \\
\text{CH}_3 & \quad \text{R}
\end{align*}
\]

\(R\) enantiomer, \(K_d = 30\ \text{pM}\)

(S enantiomer: \(K_d = 230\ \text{pM}\))

Program CombiSMoG, „best“ N-substituents from 100,000 candidates (20 scored by knowledge-based potentials)

B. A. Grzybowski et al., Acc. Chem. Res. 35, 261-269 (2002);
The Future: Combinatorial Drug Design

FlexX docking of MTX vs. X-ray structure
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

Further progress needed in the understanding and scoring of ligand-receptor interactions
References


