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)

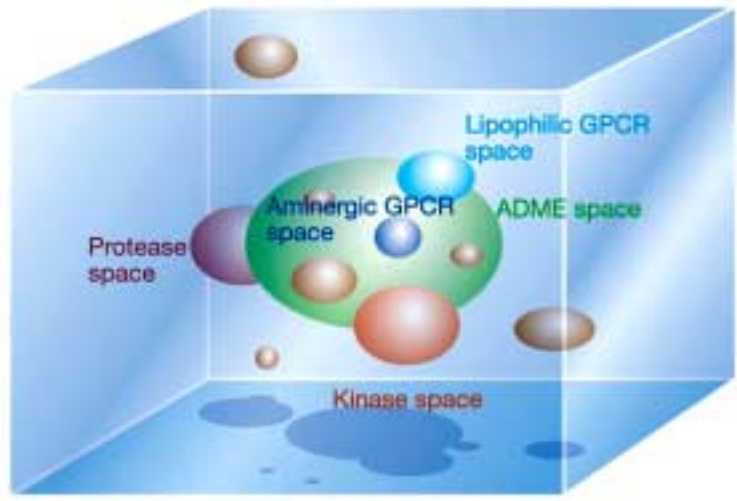
„How many leads have we got from combinatorial chemistry and high-throughput screening so far? - None !“ **Wrong**

„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



C. Lipinski and A. Hopkins, Nature 432, 855-861 (2004)

Virtual Screening Reduces the Size of the Haystack by Selecting:

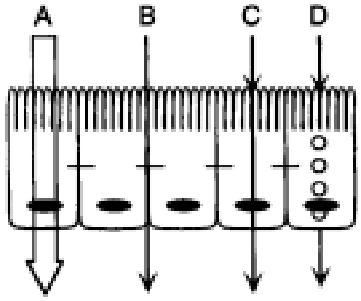
Compounds or libraries that are either
lead-like, or
drug-like, or have the
potential of oral bioavailability,
or are similar to a lead,
by rules (e.g. Lipinski bioavailability rules),
neural nets (e.g. drug-like character),
pharmacophore analyses,
similarity analyses,
scaffold hopping, or
docking and 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);
cf. D. F. Veber et al., J. Med. Chem. 45, 2615-2623 (2002).

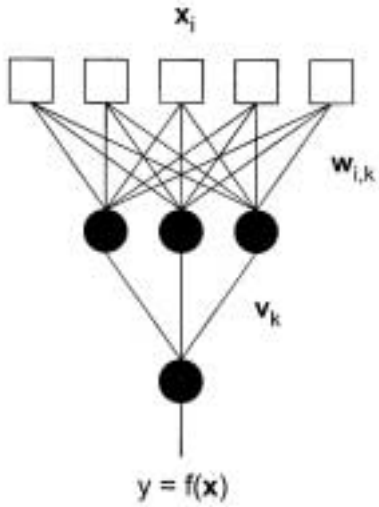
Lead-Like and Drug-Like Structures

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

S. J. Teague et al., Angew. Chem. Int Ed. Engl. 38, 3743-3748 (1999)
T. I. Oprea et al., J. Chem. Inf. Comput. Sci. 41, 1308-1315 (2001)
J. R. Proudfoot, Bioorg. Med. Chem. Lett. 12, 1647-1650 (2002)

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



input layer

hidden layer

output layer

Input layer x_i
Chemical descriptors

Weights $w_{i,k}$

Hidden-to-output weights v_k

Output layer y = ”drug-likeness”
Drugs = 0.9,
Nondrugs = 0.1

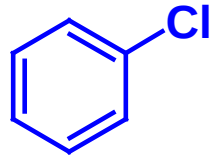
Hugo Kubinyi, www.kubinyi.de

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

Training set: 5,000 WDI, 5,000 ACD compounds

Test set: 38,416 WDI, 169,331 ACD compounds

Filters: WDI duplicates, reactive compounds

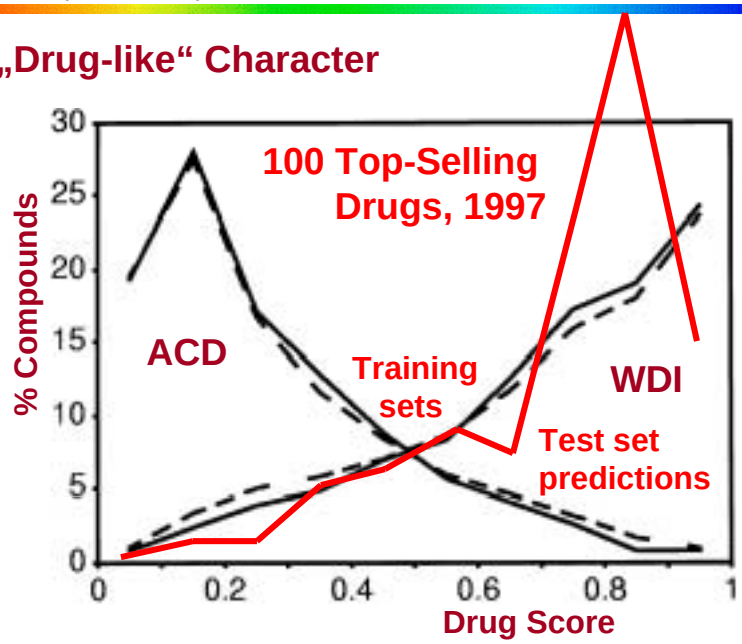


Chemical descriptors:
120 Ghose-Crippen parameters

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);
Ajay, W.P. Walters and M.A. Murcko, J. Med. Chem. 41, 3314-324 (1998)

„Drug-like“ Character

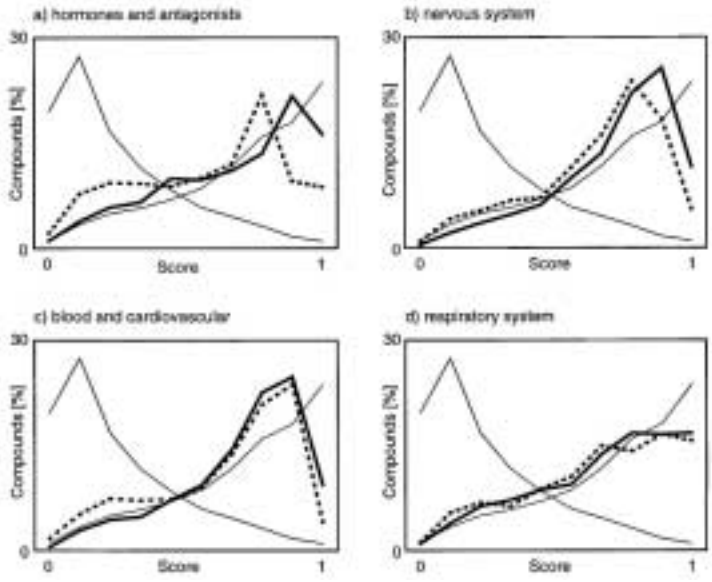


”Drug scores” of top-selling drugs (year 1994)

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

Acetylsalicylic acid Score = 0.30 !

Elimination of certain therapeutic classes



Filters for Virtual Screening	remaining
Garbage filter	90%
Druglike / Non-druglike	60%
Bioavailability	40%
Cytotoxicity	:
hERG channel inhibiton	:
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.

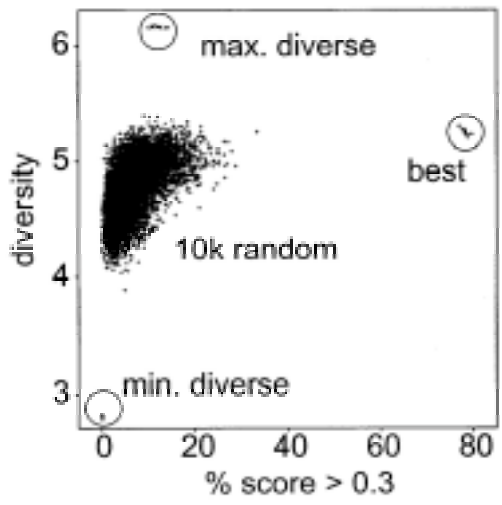
$\rightarrow$ 10^{82} possible 15x15 sublibraries

Solution: Selection by a genetic algorithm

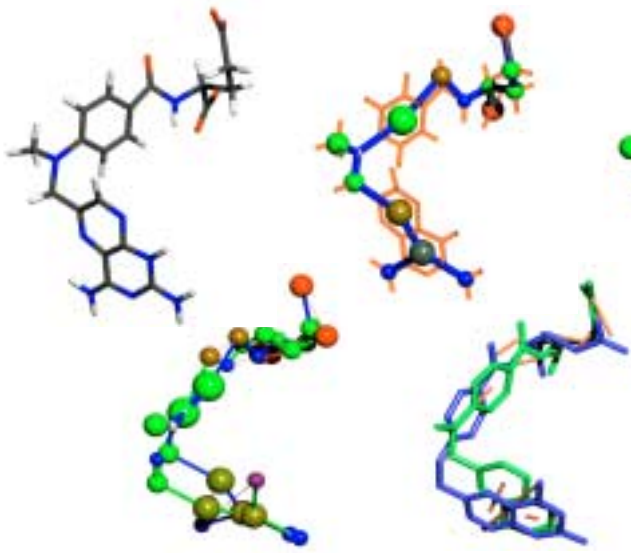
Sublibrary Selection by Genetic Algorithms

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



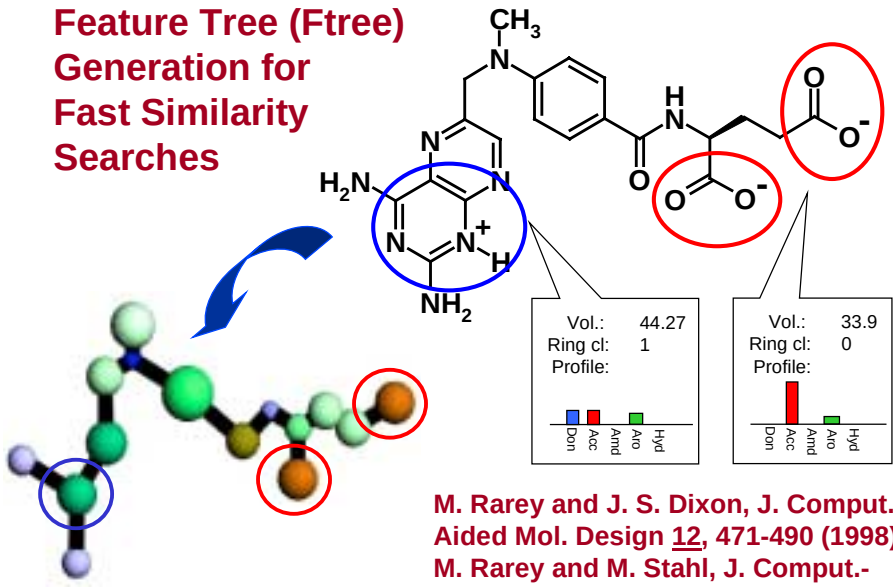
Query in 175,000 compounds in 1,5 cpu minutes

MTX vs. DHF

www.biosolveit.de

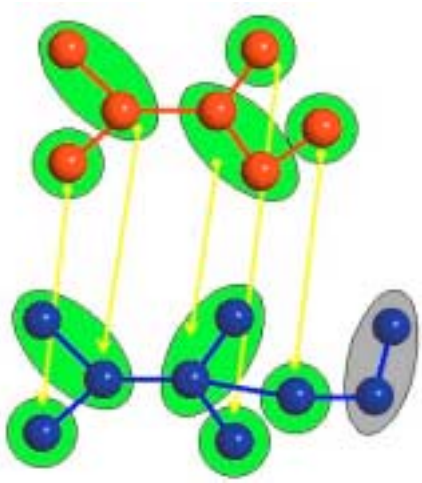
Hugo Kubinyi, www.kubinyi.de

Feature Tree (Ftree) Generation for Fast Similarity Searches



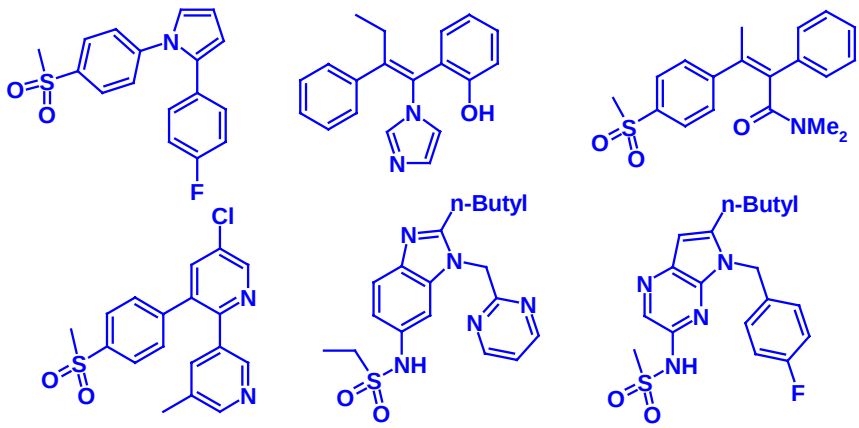
M. Rarey and J. S. Dixon, J. Comput.-Aided Mol. Design 12, 471-490 (1998); M. Rarey and M. Stahl, J. Comput.-Aided Mol. Design 15, 497-520 (2001)

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

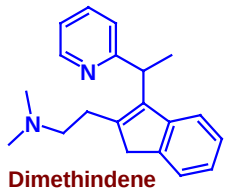
Feature Tree Query Results for COX-2 Inhibitors



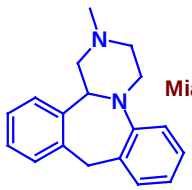
known inhibitors | plausible hits | known inhibitor related to the plausible hit

M. Rarey and M. Stahl, J. Comput.-Aided Mol. Design 15, 497-520 (2001)

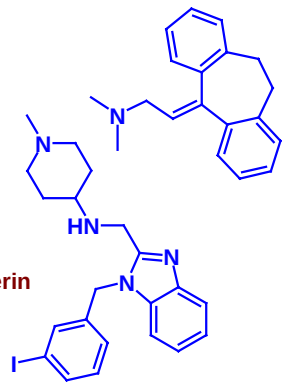
Feature Tree Query Results for H1 Antagonists and Antidepressants

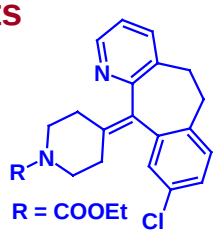


Dimethindene

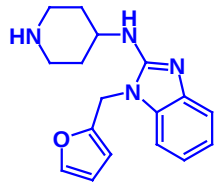


Mianserin





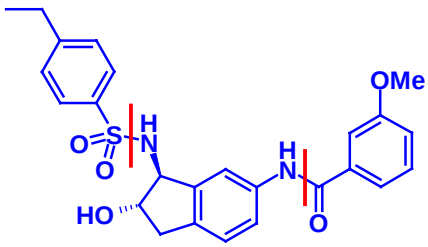
R = COOEt



known actives | plausible hits | known actives related to the plausible hit

M. Rarey and M. Stahl, J. Comput.-Aided Mol. Design 15, 497-520 (2001)

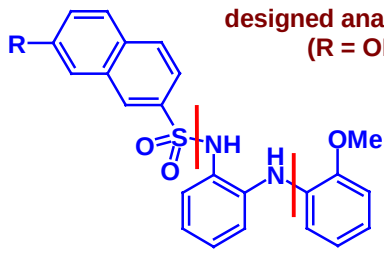
TOPAS (TOPology-Assigning System)



template

K_i hK channel 1.5 = 0.11 µM

designed analog (R = OMe)



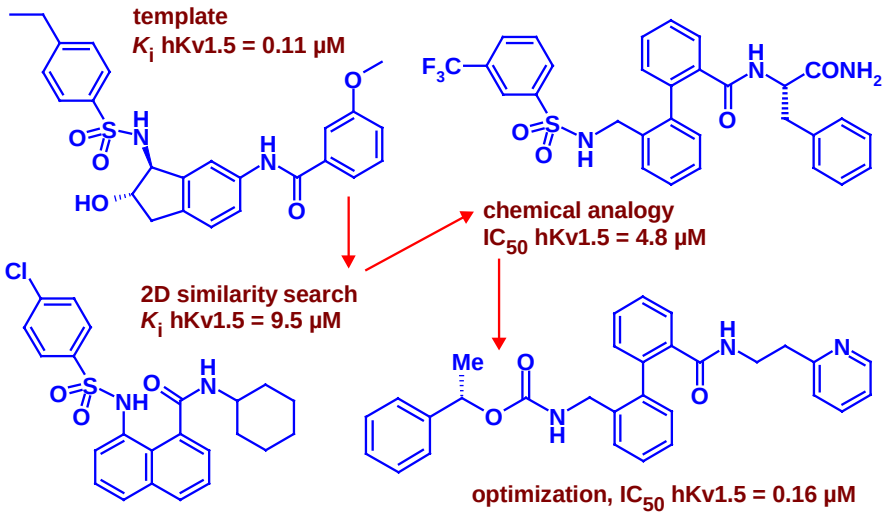
R = OMe: K_i hK channel 1.5 = 7.34 µM

R = H: K_i hK channel 1.5 = 0.47 µM

Scaffolds and building blocks from a RECAP process are re-assembled by their 3D similarity to the template („fragment-based evolutionary design“).

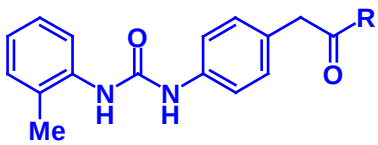
G. Schneider et al., Angew. Chem. Int. Ed. Engl. 39, 4130-4133 (2000)

Virtual Screening, Analogy and Optimization



S. Peukert et al., J. Med. Chem. 46, 486-498 (2003)

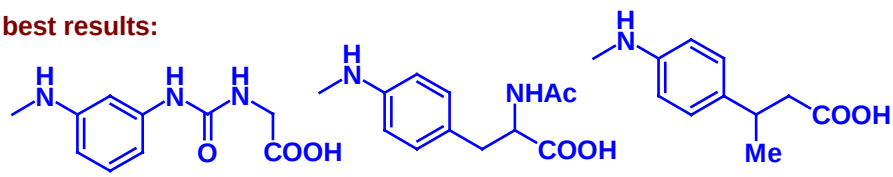
Virtual Screening of α4β1 Integrin Antagonists With CATALYST



3D structure of ligand modelled from vascular cell adhesion molecule-1 (VCAM-1) 3D structure; CATALYST search in 8,824 ACD molecules with free NH_2 or NO_2

lead: R = -Leu-Asp-Val-OH IC_{50} = 0.6 nM

best results:



IC_{50} = 67 nM IC_{50} = 58 nM IC_{50} = 1.3 nM

J. Singh et al., J. Med. Chem. 45, 2988-2993 (2002)

A Virtual Screening Success Story

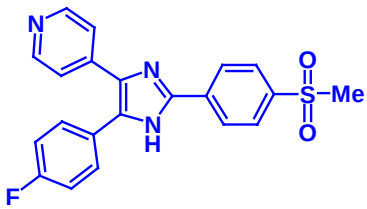
Comparison of the performance of high-throughput screening and virtual screening of potential leads of protein tyrosine phosphatase 1B (PTP1B):

a) High throughput screening of 400,000 compounds from a corporate collection → 300 hits < 300 µM, 85 validated hits with IC_{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_{50} <100 µM = 34.8% hit rate (hits are more drug-like)

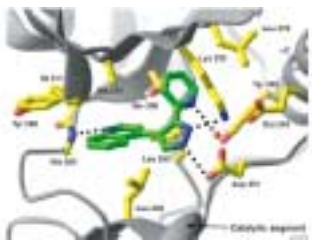
T. N. Doman et al., J. Med. Chem. 45, 2213-2221 (2002)

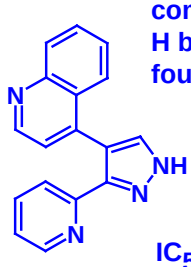
Shape-Based Virtual Screening for Type I TGFβ receptor (TßRI) Kinase Inhibitors



3D structure of SB 203 580 (IC_{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





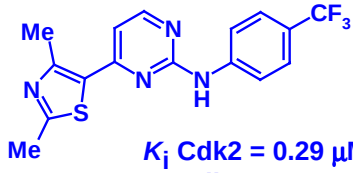
87 hits, e.g.

IC_{50} = 27 nM, K_d = 5 nM

J. Singh et al., Bioorg. Med. Chem. Lett. 13, 4355-4359 (2003)

Virtual Screening of Cdk2 Inhibitors

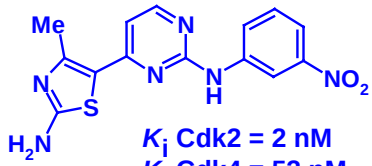
Flexible docking of about 50,000 commercially available compounds (program LIDAEUS) into cdk2 3D structure,



K_i Cdk2 = 0.29 µM
K_i Cdk4 > 20 µM

Biological testing of 200 hits

Chemical optimization (increase in affinity but decrease in selectivity)



K_i Cdk2 = 2 nM
K_i Cdk4 = 53 nM

S. Y. Wu et al., Structure 11, 399-410 (2003);
S. D. Wang et al., J. Med. Chem. 47, 1662-1675 (2004)

Stepwise Virtual Screening

Aventis in-house compound repository

↓ MW, rot-bond filter, 3D pharmacophore search

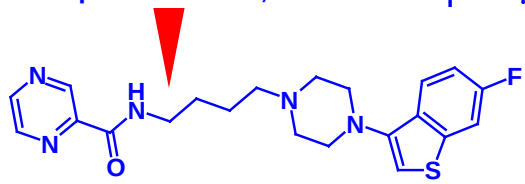
22,950 compounds

↓ docking into an α_{1A} receptor model (GOLD, PMF)

300 top-scoring compounds

↓ clustering, diversity selection

80 compounds tested, 37 hits with K_i < 10 µM



α_{1A} adrenergic receptor antagonist, K_i = 1.4 nM

A. Evers and T. Klabunde, J. Med. Chem. 48, 1088-1097 (2005)

Stepwise Virtual Screening

250,251 NCI compounds (3D database)

↓ 3D pharmacophore search

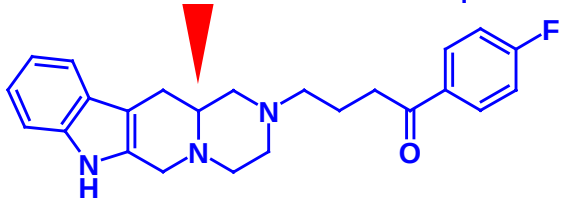
6,727 hits

↓ docking into four conformational clusters of a D_3 receptor homology model

2,478 potential ligands

↓ elimination of known chemotypes by similarity

20 compounds tested, 8 hits with K_i < 0.5 µM



dopamine D_3 receptor antagonist, K_i = 11 nM

J. Varady et al., J. Med. Chem. 46, 4377-4392 (2003)

Stepwise Virtual Screening

826,952 compounds from 8 structural databases

↓ MW, rot-bond filter, presence of pharmacophoric groups

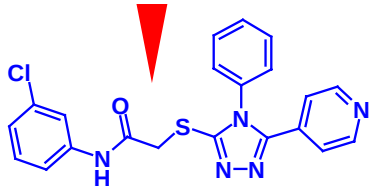
131,967 hits

↓ 2D and 3D pharmacophore searches plus excluded volumes

11,109 hits

↓ FlexX-Pharm docking, DrugScore

1,000 highest-scoring compounds visually inspected, 7 selected



NK_1 neurokinin antagonist
K_i = 251 nM

A. Evers and G. Klebe, Angew. Chem. Int. Ed. 43, 248–251 (2004);
A. Evers and G. Klebe, J. Med. Chem. 47, 5381-5392 (2004)

Stepwise Virtual Screening

259,747 ACD compounds

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

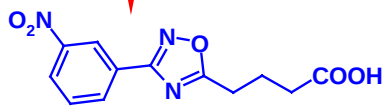
12,545 candidates

↓ 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, IC_{50} = 2.4 µM

O. Krämer et al., Proteins Struct. Funct. Genet. 55, 814–823 (2004)

Virtual Screening of Carbonic Anhydrase Inhibitors

98,850 compounds (LeadQuest and Maybridge libraries)

↓ filter for Zn^{2+}-binding anchor groups

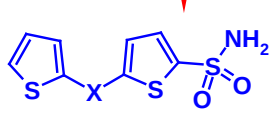
5,904 hits

↓ 2D and 3D pharmacophore searches (derived from binding site analysis)

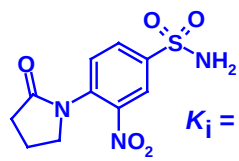
3,314 hits

↓ FlexS superposition with dorzolamide, followed by FlexX docking of 100 hits into carbonic anhydrase binding site

13 hits



X = S
K_i = 0.9 nM
X = SO_2
K_i = 0.8 nM



K_i = 0.6 nM

S. Grüneberg et al., Angew. Chem., Int. Ed. Engl. 40, 389-393 (2001);
S. Grüneberg et al., J. Med. Chem. 45, 3588-3602 (2002).

Virtual Screening of TGT Inhibitors

800,000 compounds (from eight different databases)

MW < 450 and rot-bond filter

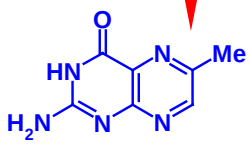
about 400,000 molecules

pharmacophore searches, followed by binding-site derived volume constraints

872 hits

flexible docking (program FlexX) into two different binding site conformations

9 hits, all biologically active



tRNA-guanine transglycosylase (TGT) inhibitor, K_i = 0.25 µM

R. Brenk et al., J. Med. Chem. 46, 1133-1143 (2003);
R. Brenk et al., J. Mol. Biol. 338, 55-75 (2004)

Stepwise Virtual Screening

560,000 compounds (subsection of AstraZeneca repository)

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

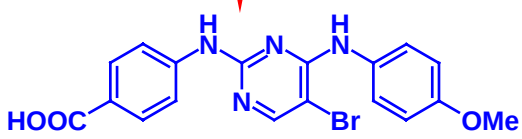
199,000 hits

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

P. D. Lyne et al., J. Med. Chem. 47, 1962-1968 (2004)

Virtual Screening of CK2 Inhibitors

400,000 compounds (Novartis in-house compound collection)

↓ docking into CK2 homology model (from CK2 of *Zea mays*)

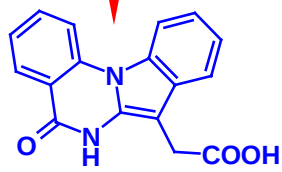
highest-scoring hits

↓ filters: binding to hinge region, favorable conformations

12 hits

↓ biological testing

4 hits with > 50% inhibition at 10 µM



Protein kinase CK2 (casein kinase II) inhibitor, IC_{50} = 80 nM

E. Vangrevelinghe et al., J. Med. Chem. 46, 2656-2662 (2003)

Virtual Screening of Akt1 (PKBα) Inhibitors

50,000 ChemBridge compounds

↓ flexible docking (program FlexX)

top 2,000 molecules

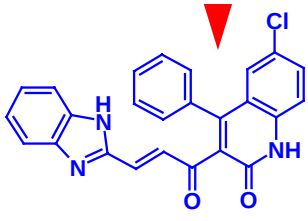
↓ ranking by CSCORE, DrugScore, GoldScore ChemScore (top 100-200 hits)

1 hit

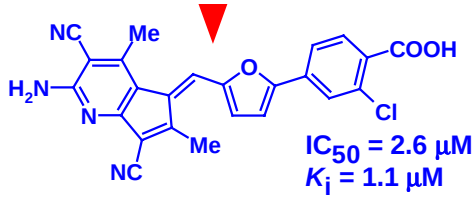
re-ranking (top 700 hits), visual inspection

↓

100 top-scoring compounds



IC_{50} = 4.5 µM
K_i = 3.9 µM



IC_{50} = 2.6 µM
K_i = 1.1 µM

M. Forino et al., J. Med. Chem. 48, 2278-2281 (2005)

Virtual Screening: The Screensaver Project

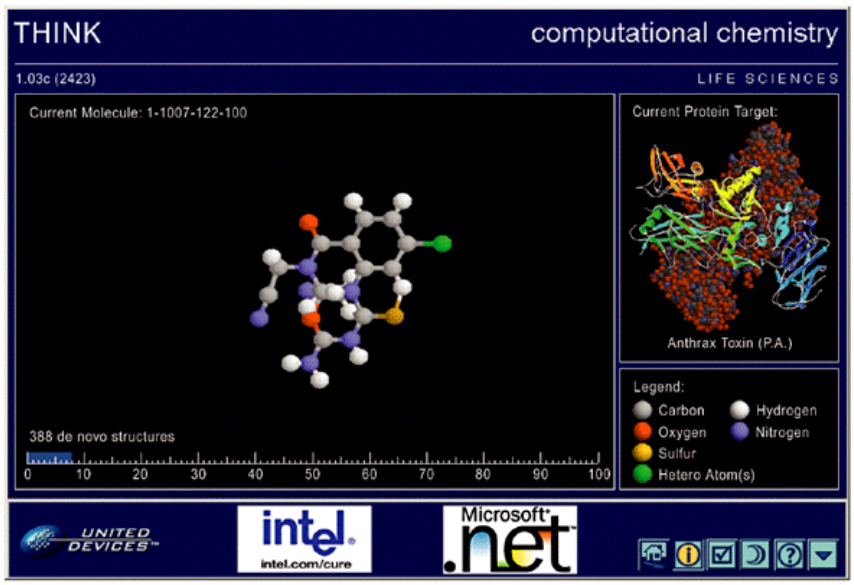
coordinated by W. G. Richards, University of Oxford.

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).

W. G. Richards, Nature Rev. Drug Discov. 1, 551-555 (2002)

Hugo Kubinyi, www.kubinyi.de



The Screensaver Surface (W. G. Richards, Nature Rev. Drug Discov.)

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

Böhm, H.-J., and Schneider, G., Eds., Virtual Screening for Bioactive Molecules (Methods and Principles in Medicinal Chemistry, Volume 10, Mannhold, R., Kubinyi, H., and Timmerman, H., Eds.), Wiley-VCH, Weinheim, 2000

Klebe, G., Ed., Virtual Screening: An Alternative or Complement to High Throughput Screening, Persp. Drug Discov. Design 20, 1-287 (2000)

Güner, O.F., Ed., Pharmacophore Perception, Development and Use in Drug Design, International University Line, La Jolla, CA, USA, 2000

Böhm, H.-J., and Schneider, G., Eds., Protein-Ligand Interactions. From Molecular Recognition to Drug Design, (Methods and Principles in Medicinal Chemistry, Volume 19, Mannhold, R., Kubinyi, H., and Folkers, G., Eds.), Wiley-VCH, Weinheim, 2003

Alvarez, J., and Shoichet, B., Eds., Virtual Screening in Drug Discovery, CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2005