Computational Drug Discovery

Dr. Bryce, 6 lectures
Room 2.031, r.a.bryce@manchester.ac.uk

Aim

to provide you with an understanding of the role of molecular modelling in the process of rational design of small-molecule drugs, with an appreciation of its strengths and limitations

Objectives

By the end of this subsection, you should be able to:

1. describe briefly the sources from which **drug leads** are obtained

2. outline the concept of structure-based drug design and within this, the role of **molecular modelling**

3. discuss the key molecular modelling technique of **conformational searching**, which includes the concept of a potential energy surface; the force field method to score geometry through the calculation of energy; and three approaches to exploration of conformation (energy minimisation, Monte Carlo simulation, molecular dynamics)

4. describe **virtual screening** approaches based on molecular docking and/or a pharmacophore; consider the content of compound libraries used for virtual screening; discuss how some common problems in virtual screening are addressed, in regard to scoring function, receptor availability and receptor flexibility.

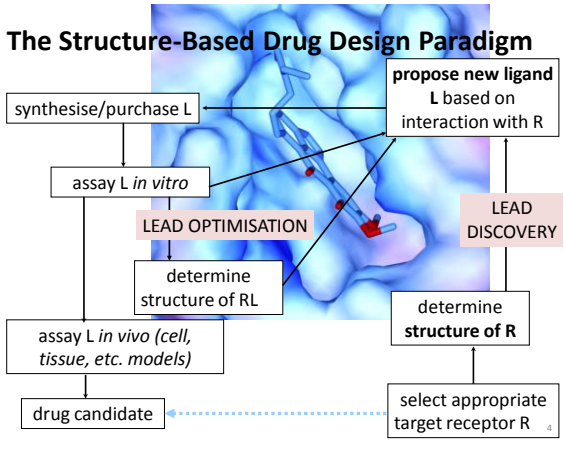
5. appreciate the principles, strengths and limitations of computer-based **de novo design methods** that use active site analysis and connection methods.

6. Discuss the **3D QSAR** approach to computational design.

The needle in the haystack

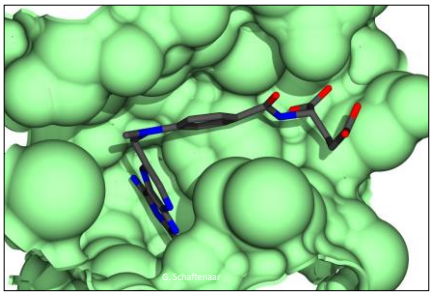
4. virtual screening

The Structure-Based Drug Design Paradigm



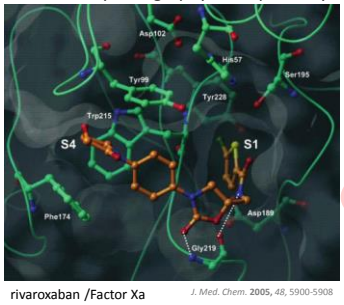
4.1. Docking

- How well does ligand **L** fit into active site?
  - what is the ligand's *bound pose*?



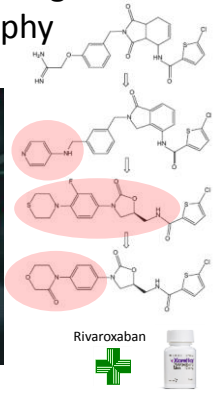
4.1.1. Experimental docking: use X-ray crystallography

- can use crystallography *retrospectively*



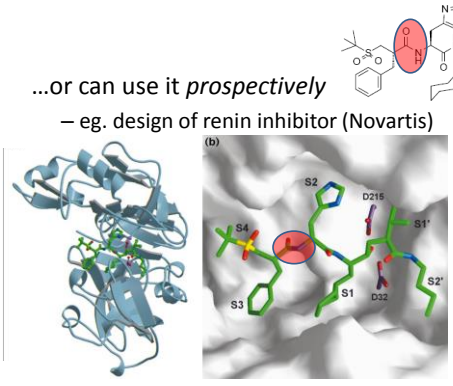
rivaroxaban /Factor Xa

J. Med. Chem. **2005**, *48*, 5900-5908



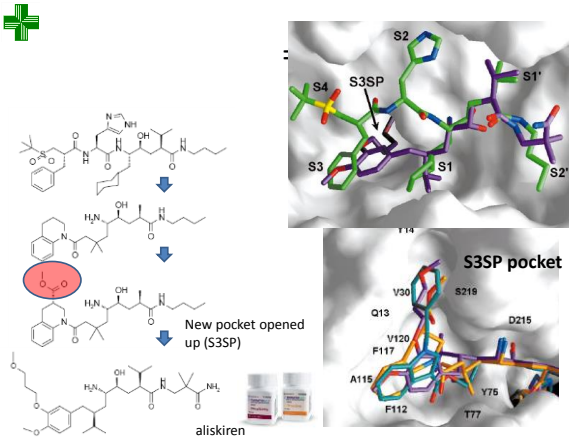
…or can use it *prospectively*

– eg. design of renin inhibitor (Novartis)

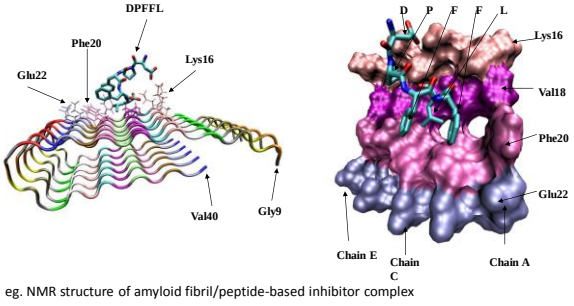


Chem. Biol. 2000, 7, 493

Get away from peptidic lead molecule - Design into S1-S3 groove, avoid S2-S4 peptide



Experimental docking: NMR spectroscopy



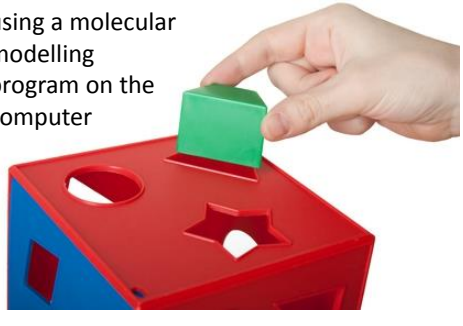
eg. NMR structure of amyloid fibril/peptide-based inhibitor complex

3D coordinates of X-ray/NMR protein structures are archived in the Protein Data Bank (PDB):

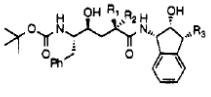
http://www.rcsb.org

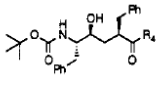
4.1.2. Manual docking

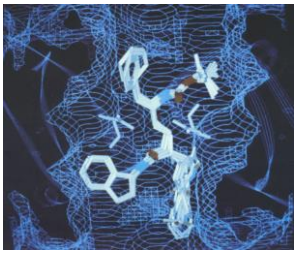
- using a molecular modelling program on the computer

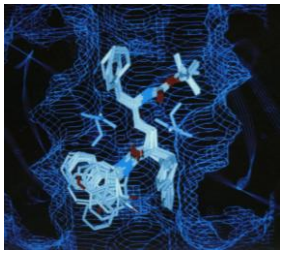


HIV-1 protease inhibitors



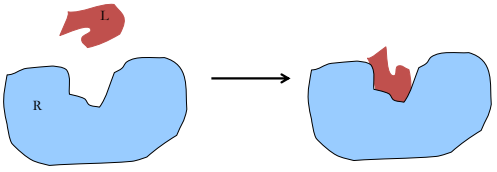






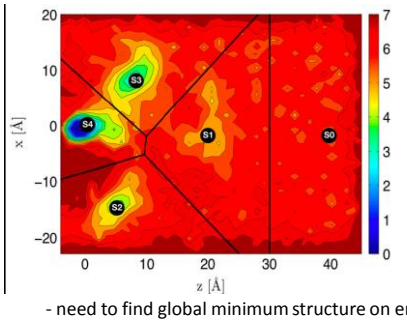
4.1.3. Automated docking

- much faster for predicting RL structure!
- for a given X-ray/NMR 3D structure of receptor R
  - predict how ligand L binds



ligand's correct bound geometry in receptor binding site =
binding mode or bound *pose*

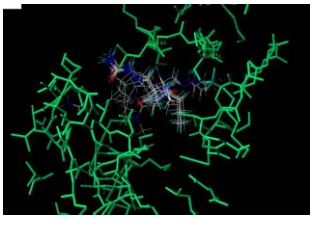
receptor-ligand docking



- need to find global minimum structure on energy surface (without knowing topology of surface in advance)

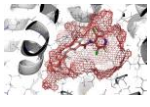
Conformational search

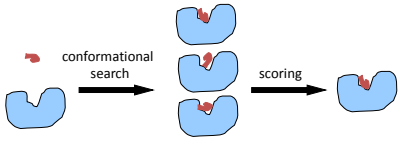
- possible techniques to generate bound conformations:
  - Monte Carlo
  - molecular dynamics
  - genetic algorithm
  - particle swarms
  - graph theory
- can also use 'local knowledge' to bias predictions
  eg. known coordination to a metal or conserved hydrogen bond



receptor-ligand docking

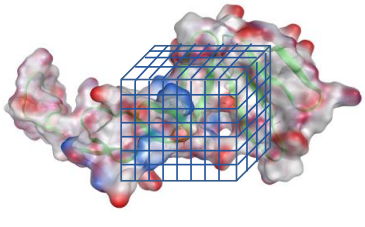
1. *Characterization* of protein active site
2. *Conformational search*
   - generation of orientations of ligand in active site
3. *Scoring* of ligand poses





4.1.3.1. Characterisation

- eg. AutoDock program
  - uses a grid to define active site region



4.1.3.2. Conformational search

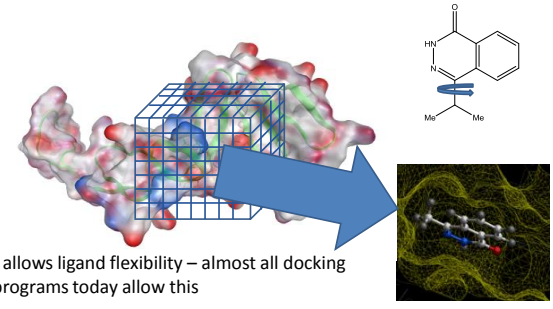
#1) make random change in ligand geometry x in the protein active site, to give new structure x+Δx

#2) compare energies of old and new structures

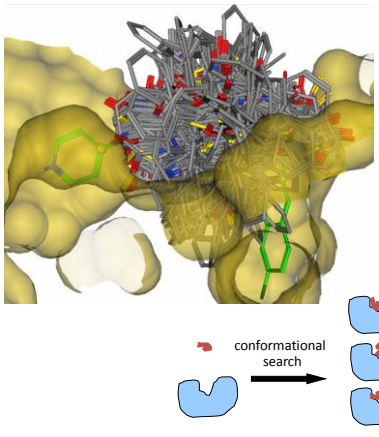
- accept new structure if its interaction energy is lower than that of the old structure: U(x+Δx) < U(x)
- accept new structure if Boltzmann factor exp[–{U(x+Dx) –U(x)}/RT] > random number between 0 and 1
- otherwise reject

#3) go to #1

Generate ligand geometries in protein active site using Monte Carlo



- allows ligand flexibility – almost all docking programs today allow this



4.1.3.3. Scoring

- require a "scoring function"
  - eg. a molecular mechanics force field

$$U_{tot}(r_{ij}) = \sum_{i,j}^{atoms} \left(\frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}} \right) + \sum_{i,j}^{atoms} \frac{q_i q_j}{r_{ij}} + \text{other terms}$$

  - often use difference in energy ΔU_{inter} rather than total energy U_{tot} of complex, for scoring:

$$\Delta U_{\text{inter}}(r_{ij}) = U_{tot}^{L-R} - U_{tot}^{L} - U_{tot}^{R}$$

Example docking programs

Program	Search strategy	Free for academia	Website
AutoDock (10)	GA/MC	Yes	http://autodock.scripps.edu
Dock (11)	IC	Yes	http://dock.compbio.ucsf.edu
FlexX (12)	IC	No	http://www.biosolveit.de/flexx
Glide (13)	Hybrid	No	http://www.schrodinger.com
Gold (14)	GA	No	http://www.ccdc.cam.ac.uk/products/life_sciences/gold
Surflex (15)	IC	No	http://www.tripos.com/index.php
ICM (16)	MC	No	http://www.molsoft.com/docking.html
LigandFit (17)	MC	No	http://accelrys.com/products/discovery-studio
eHiTS (18)	IC	No	http://www.simbiosys.ca/ehits/index.html

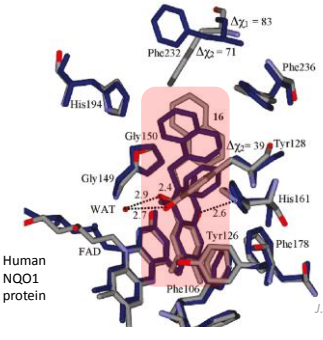
GA genetic algorithm, *MC* Monte Carlo, *IC* incremental construction

Go deeper: **Patrick, Introduction to Medicinal Chemistry, Ch. 17. Section 17.12**

Cheng et al. "Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review", The AAPS Journal, Vol. 14, No. 1, March 2012 (# 2012)

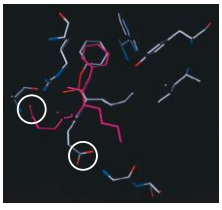
Successful docking

prediction = experiment (X-ray pose)

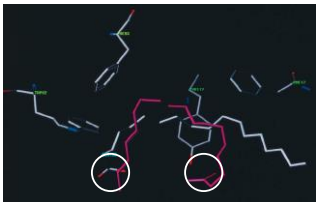


Human NQO1 protein

J. Med. Chem. **2009**, *52*, 7142–7156



not quite right



poor

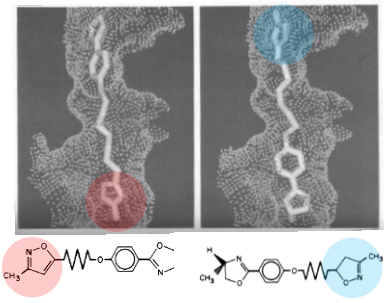
Docking predicts the true ligand pose as the top-scored pose about 70-80% of time

JMB **1997**, 267, 727

Ligand binding can be subtle

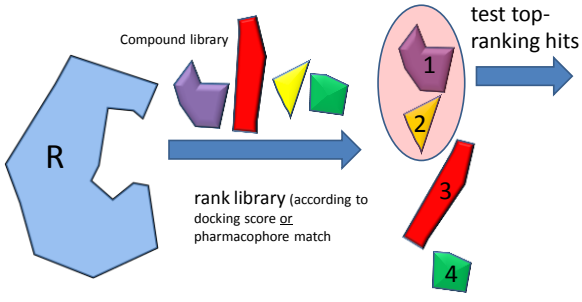
– eg. two different X-ray poses for very similar ligands

human rhinovirus 14



P. S. Charifson, Practical Application of Computer-Aided Drug Design (Taylor & Francis, 1997)

4.2. Receptor-based virtual screening



Library of 3D structure of compounds often ~10^5 - 10^6

4.2.1. Compound libraries for screening

- virtual compound libraries
  - databases of 3D structure of small molecule compounds
  - often ~10^5 - 10^6 compounds
- sources:
  1. in-house compound collections of pharmaceutical companies
  2. computationally generated set of compounds that could be synthesised *in principle.* Consider:
     - 10^4 commonly used chemical scaffolds
     - usually ~3 side-chains in molecules generally
     - 10^3 different side-chains in known drugs

$10^4 \times (10^3)^3 = 10^{13}$ compounds

Compound libraries for screening

3. open-access compound repositories
   - NCI, National Cancer Institute
     - dtp.nci.nih.gov
4. commercially-available reagent catalogues
   - ZINC, "ZINC Is Not Commercial"
     - blaster.docking.org/zinc
     - *eg. "All purchasable" set*
       » 17.8M compounds

Compound libraries

Database	Type	No. of compounds[a]	Website
PubChem	Public	30 million	http://pubchem.ncbi.nlm.nih.gov
ChEMBL	Public	1 million	https://www.ebi.ac.uk/chembldb/index.php
NCI Set	Public	140,000	http://dtp.nci.nih.gov/index.html
ChemSpider	Public	26 million	http://www.chemspider.com
CoCoCo	Public	7 million	http://cococo.unimore.it/tiki-index.php
TCM	Public	32,000	http://tcm.cmu.edu.tw
ZINC	Public	13 million	http://zinc.docking.org
ChemBridge	Commercial	700,000	http://www.chembridge.com
Specs	Commercial	240,000	http://www.specs.net
Asinex	Commercial	550,000	http://www.asinex.com
Enamine	Commercial	1.7 million	http://www.enamine.net
Maybridge	Commercial	56,000	http://www.maybridge.com
WOMBAT	Commercial	263,000	http://www.sunsetmolecular.com
ChemDiv	Commercial	1.5 million	http://www.chemdiv.com
ChemNavigator	Commercial	55.3 million	http://www.chemnavigator.com
ACD	Commercial	3,870,000	http://accelrys.com/products/databases/sourcing/available-chemicals-directory.html
MDDR	Commercial	150,000	http://accelrys.com/products/databases/bioactivity/mddr.html

[a] Approximate numbers

Cheng et al. "Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review", The AAPS Journal, Vol. 14, No. 1, March 2012 (# 2012)

ZINC libraries

- *"Drug-Like" subset*
  - Lipinski filter applied
  - 10.6M compounds

Lipinski's Rule of Five

Compounds may have problem with oral absorption unless they have:

- molecular weight ≤ 500 g/mol
- LogP/CLogP ≤ 5
- ≤ 5 H-bond donors (usually, sum of NH and OH)
- ≤ 10 H-bond acceptors (usually, sum of N and O)

- *"Drugs Now" subset*
  - delivery less than 2 weeks usually
  - excludes make-on-demand compounds
  - 6M compounds

31

Compound libraries: issues

- Does library contain desired range of compounds?
  - eg. does it contain too many examples of a particular type of molecule?
- Does library contain "swill"?
  - *ie*. reactive compounds (eg. aldehydes), non-Lipinski compounds, etc.?

32

Examples of swill

```
ELIMINATE_METALS
    Sc,Ti,V,Cr,Mn,Fe,Co,Ni,Cu,Zn,Y,Zr,Nb,M
    o,Tc,Ru,Rh,Pd,Ag,Cd

#specific, undesirable functional groups
RULE 0 Carbazides
RULE 0 Acid_anhydrides
RULE 0 Pentafluorophenyl_esters
RULE 0 Paranitrophenyl_esters
RULE 0 HOBT_esters
RULE 0 Triflates
RULE 0 Lawesson_s_reagent
RULE 0 Phosphoramides
RULE 0 Aromatic_azides
RULE 0 Beta_carbonyl_quart_nitrogen
#RULE 0 Acylhydrazide
RULE 0 Quarternary_C_Cl_I_P_or_S
RULE 0 Isonitrile
RULE 0 Triacyloxime
RULE 0 Cyanohydrins
RULE 0 Acyl_cyanides
RULE 0 Sulfonyl_cyanides
RULE 0 Cyanophosphonates
RULE 0 Azocyanamides
RULE 0 Azoalkanals
RULE 0 Polyenes
RULE 0 Saponin_derivatives
RULE 0 Cytochalasin_derivatives
RULE 0 Cycloheximide_derivatives
RULE 0 Monensin_derivatives
RULE 0 Cyanidin_derivatives
RULE 0 Squalestatin_derivatives
RULE 0 Phosphoranes
RULE 0 Chloramidines
RULE 0 Nitroso
RULE 0 P_S_Halides
RULE 0 Carbodiimide
```

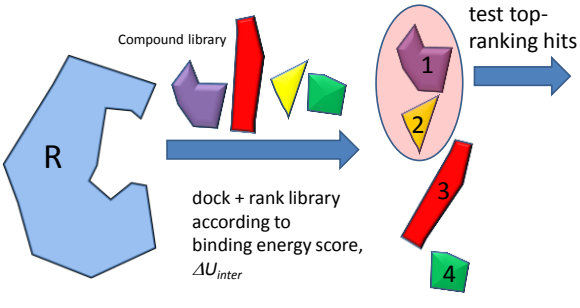
33

Compound libraries: issues

- How quickly can compounds be delivered?
- How much do compounds cost?
- Are compounds pure?

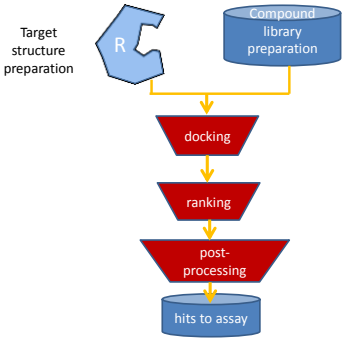
34

4.2.2. Docking-based virtual screening



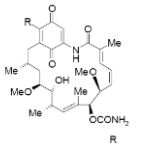
35

Virtual screening funnel

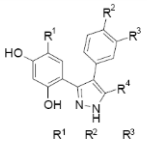


Example: VS against Hsp90

- molecular chaperone protein Hsp90 involved in cancer onset and progression
  - attractive oncology target
- Geldanamycin-derived inhibitor 17-AAG (**1b**) entered phase I trials
  - but poor solubility and bioavailability, toxicity and extensive metabolism



	R
1a 17-DMAG	$-NHCH_2CH_2NMe_2$
1b 17-AAG	$-NHCH_2CH{=}CH_2$



	R^1	R^2	R^3	R^4
4a G3129	-Et	$-O-CH_2-O-$		-COOH

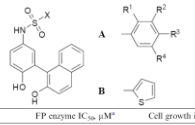
Bioorg. Med. Chem. Lett. 15 (2005) 5187–5191

Protocol

- rDOCK software
- 700,000 compounds screened
  - from 3.5M based on reactive groups and delivery time
- 9000 from docking
- Post-processing:
  - all X-ray structures show ligands donate hydrogen bond to Asp93 C=O and accept hydrogen bond from Hsp90-bound water molecule
  - use this "pharmacophore" to filter down hits
- pharmacophore =

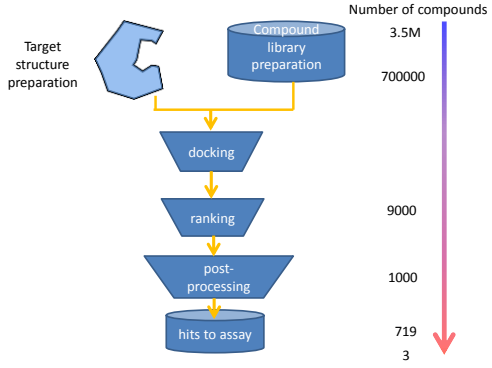
Results

- After applying pharmacophore and after reducing overrepresented chemical scaffolds, 1000 compounds selected for purchase
  - only 719 actually available
- From 719 tested, discovered 3 hits from new chemical scaffold:



Compound	X	R^1	R^2	R^3	R^4	FP enzyme IC_{50}, μM[a]	Cell growth inhibition, μM[b]
11	A	Cl	H	H	Cl	0.7	29.0

Virtual screening funnel



Results

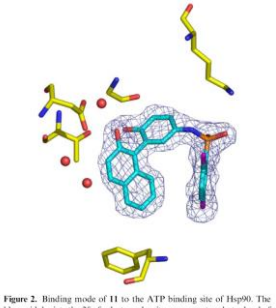


Figure 2. Binding mode of **11** to the ATP binding site of Hsp90. The blue grid depicts the 2fo-fc electron density map countered at a level of

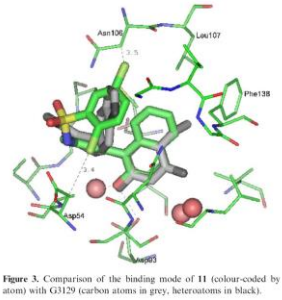
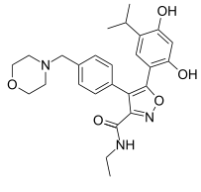


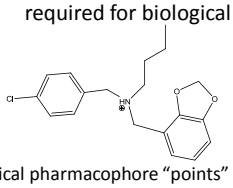
Figure 3. Comparison of the binding mode of **11** (colour-coded by atom) with G3129 (carbon atoms in grey, heteroatoms in black).

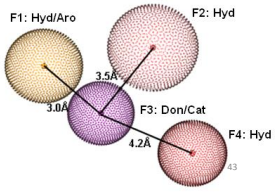
- From same lab (Vernalis and Institute of Cancer Research), a Hsp90 inhibitor called NVP-AUY922 is currently in clinical trials (conducted under licence by Novartis):



4.2.3. Pharmacophore-based virtual screening

pharmacophore = 3D arrangement of atoms/groups required for biological activity



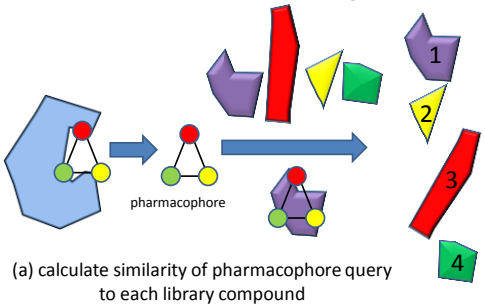


Typical pharmacophore "points"

- hydrogen-bond donor (Don)
- hydrogen-bond acceptor (Acc)
- anion (An)
- cation (Cat)
- aromatic ring (Aro)
- hydrophobic group (Hyd)

43

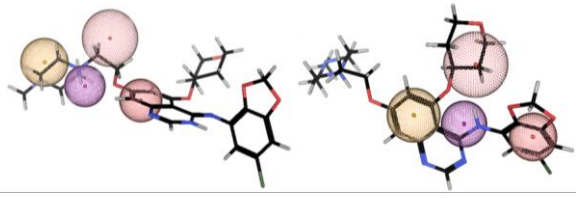
Pharmacophore-based virtual screening



(a) calculate similarity of pharmacophore query to each library compound
(b) rank compounds according to this similarity

44

Score: how well do ligands fit pharmacophore?

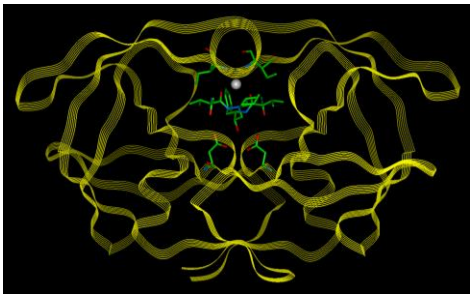


45

Example of pharmacophore-based virtual screening

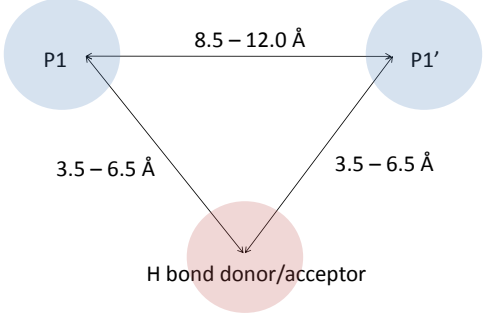
e.g. HIV protease (P.S. Lam et al., *Science* **1994**, *263*, 380)

X-ray structure shows bound water molecule at active site



46

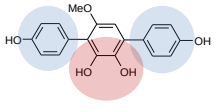
Pharmacophore hypothesis



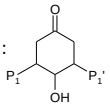
47

Pharmacophore-based virtual screening

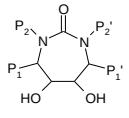
database "hit" molecule:



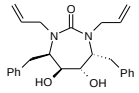
initial idea:



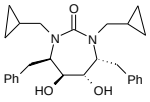
include urea and diol:



model right stereochemistry:



Ki 4.7 nM



Ki 2 nM

48

Pharmacophore-based virtual screening

Advantages

- very fast
- can also be used in the absence of a receptor structure

Limitations

- requires correct pharmacophore
  - need receptor and/or active ligand structures
  - can be subjective, assumes certain interactions to be key
- requires representative 3D conformations of small-molecule compounds in virtual library

49

Docking-based virtual screening

Advantages

- provides easily interpreted binding modes of protein-ligand
  - identify key interactions
- can identify unanticipated binding modes

Limitations

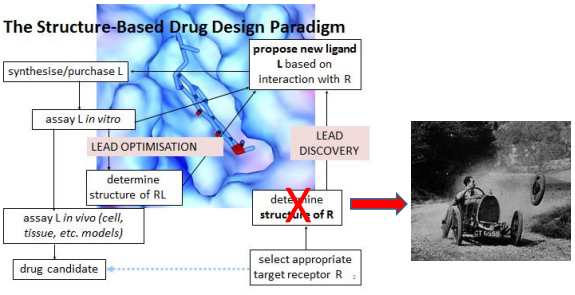
- requires a receptor structure
- software may not allow for ligand, protein side-chain or protein main-chain flexibility during docking
- slower than pharmacophore-based VS
  - but can use pharmacophore pre-screen to cut down number of compounds to dock
- scoring functions can fail to predict (a) correct docked pose of a ligand and/or (b) the correct ranking of different ligands
- requires representative 3D conformations of small-molecule compounds in virtual library

Flies in the ointment

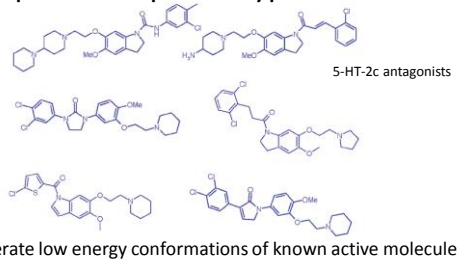
4.3. problems in virtual screening

51

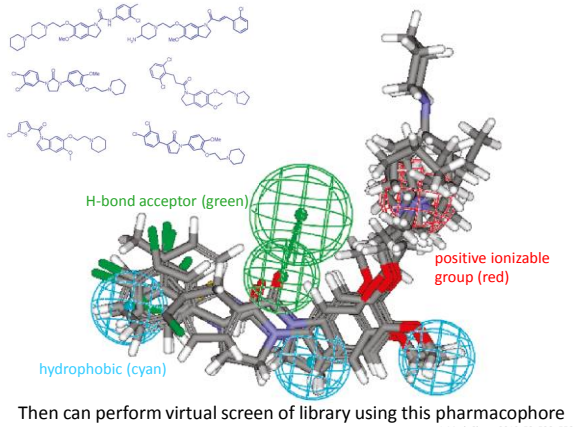
4.3.1. Problem #1 What if you don't have a protein structure?



4.3.1.1. Solution #1 - Use knowledge of active ligands: generate a pharmacophore hypothesis



- Generate low energy conformations of known active molecule
- Superimpose them on top of each other
- Look for common chemical features across molecules

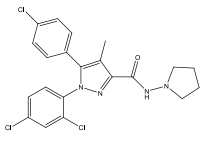


Then can perform virtual screen of library using this pharmacophore

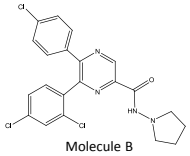
J. Med. Chem. 2010, 53, 539–558

4.3.1.2. Solution #2 - Use knowledge of active ligand: compare shape of active ligand to other (as yet untested) ligands

eg. How similar is A to B?



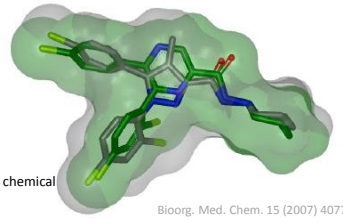
Molecule A



Molecule B

Compare space-filling models of A and B

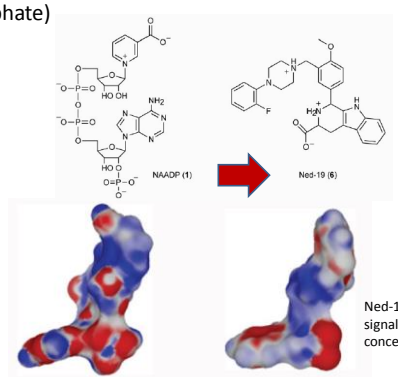
- maximise overlap volume of A and B (by translating/rotating molecules to maximise overlap)



ROCS program: rapid overlay of chemical structures

Bioorg. Med. Chem. 15 (2007) 4077–4084

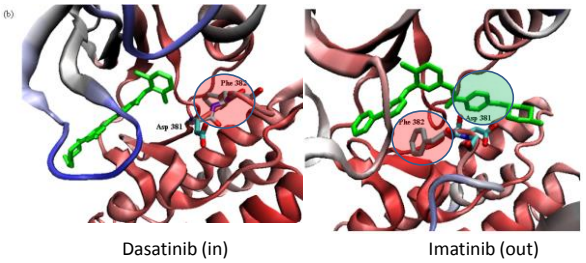
- eg. discovered an antagonist of Ca^{2+}-releasing second messenger NAADP (nicotinic acid adenine dinucleotide phosphate)



Ned-19 blocks NAADP signaling at nanomolar concentrations

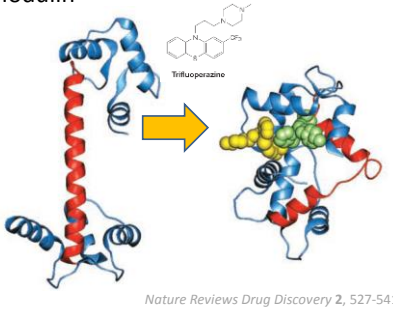
4.3.2. Problem #2 What if the protein is flexible?

- *in* and *out* conformations of c-Abl kinase



Dasatinib (in) Imatinib (out)

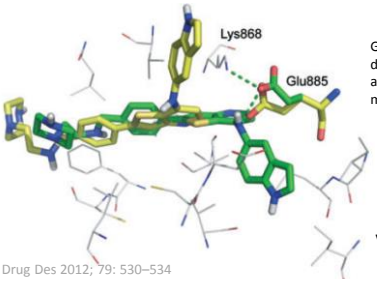
- calmodulin



Nature Reviews Drug Discovery **2**, 527-541 (July 2003)

4.3.2.1. Solution #1: Induced fit docking

- Allow protein to move during docking
  - Amino acid side-chain flexibility eg. AutoDock 4

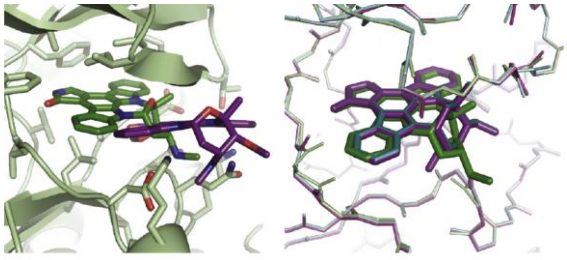


Green = correctly docked pose, which allows side-chain motion

VEGF receptor-2

Chem Biol Drug Des 2012; 79: 530–534

- Peptide backbone flexibility eg. ROSETTALIGAND
  - Monte Carlo sampling of ligand and receptor



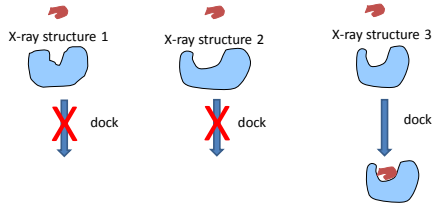
Docking without backbone flexibility Docking with backbone flexibility

CDK2/staurosporine

J. Mol. Biol. (2009) 385, 381–392

4.3.2.2. Solution #2: Ensemble-based docking

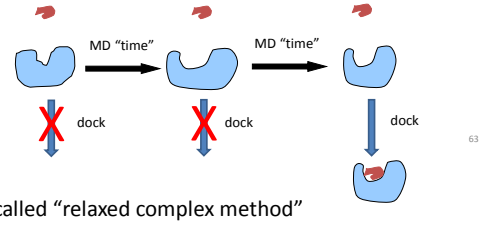
- If we know what the range of conformations are (eg. from different X-ray structures), then dock into each of these conformations



62

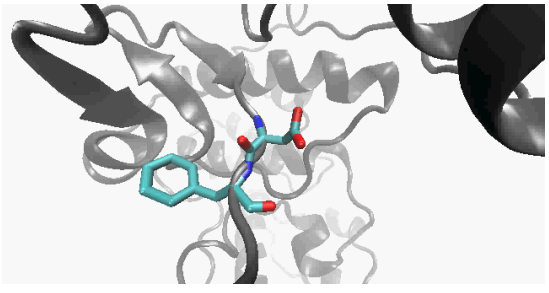
4.3.2.3. Solution #3: Relaxed complex method

- If we don't know what the conformations are
  - perform long MD simulation of protein
  - dock into different conformations observed during MD

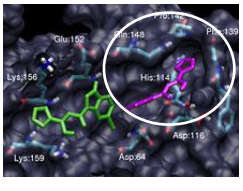


63

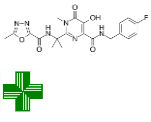
  - called "relaxed complex method"



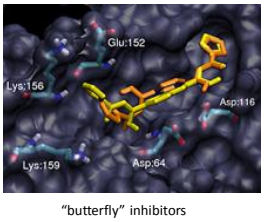
HIV integrase



Hidden trench (2004) guided Merck to design of raltegravir (Isentress), approved 2007

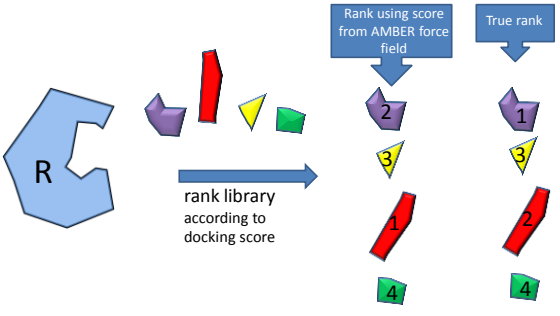


Advantage of MD – can uncover "cryptic" hidden pockets, not visible from X-ray structures



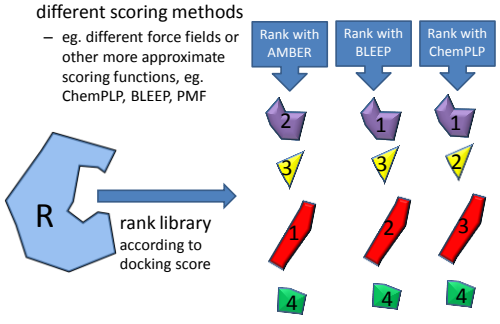
"butterfly" inhibitors

4.3.3. Problem #3 What if my scoring function is no good?



4.3.3.1. Solution: use “consensus scoring”

- re-rank using several different scoring methods
  - eg. different force fields or other more approximate scoring functions, eg. ChemPLP, BLEEP, PMF



Consensus scoring

Score each ligand in library using *several* scoring functions

- then calculate *overall rank* by adding ranks from each scoring function

	Ligand 1	Ligand 2	Ligand 3	Ligand 4
Experiment	**1**	**3**	**2**	**4**
AMBER	2	3	1	4
BLEEP	1	3	2	4
ChemPLP	1	2	3	4
Summed ranks	**4**	**8**	**6**	**12**

68

4.4. Virtual screening - Summary

- Faster and cheaper than experimental screening via HTS or NMR-based approaches
- VS predictions are tempered by problems in docking (bound geometry, ranking) or determination of correct pharmacophore
- Nevertheless, VS is a useful tool to reduce the number of compounds selected for subsequent HTS