

BioInfo 2004

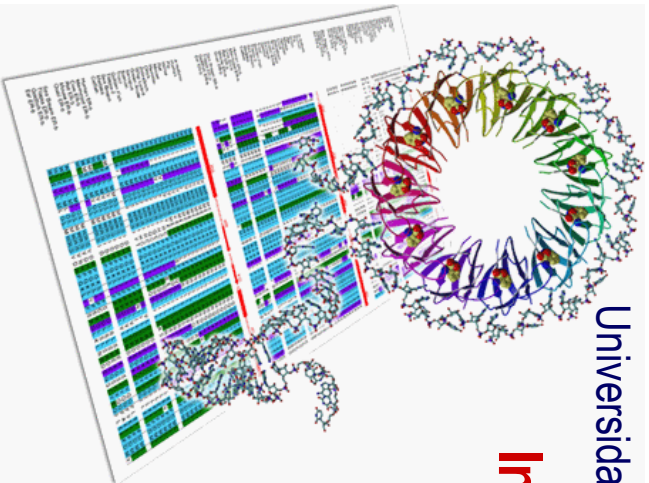
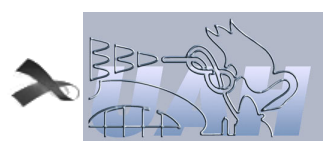
Curso de Doctorado: BIOINFORMÁTICA

Universidad Autónoma de Madrid. Marzo-Abril 2004

Interacciones entre proteínas y moléculas pequeñas (III)

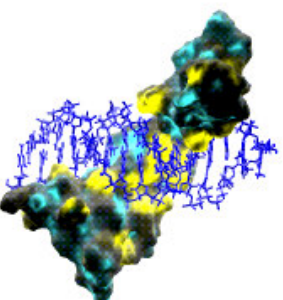
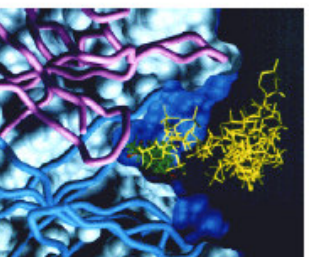
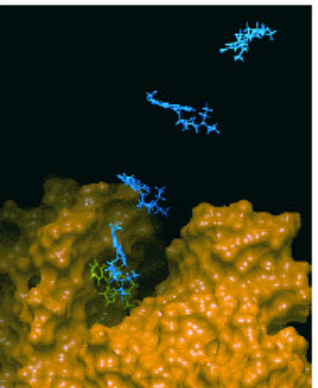
Federico Gago

Departamento de Farmacología
Universidad de Alcalá, Madrid



Major Application of Bioinformatics in 21st century: Designing Drugs

- Understanding how structures bind other molecules (function)
- Designing inhibitors
- Docking, structure modeling



"When a medicinal chemist synthesizes a compound that does something extraordinary to a biological system, this compound enters *an elite class of chemicals* and becomes classified as *a drug*."

T. P. Kenakin

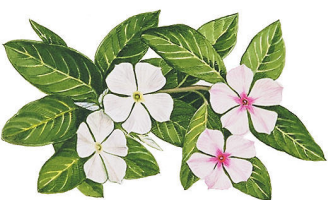
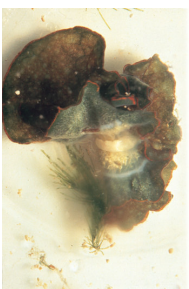
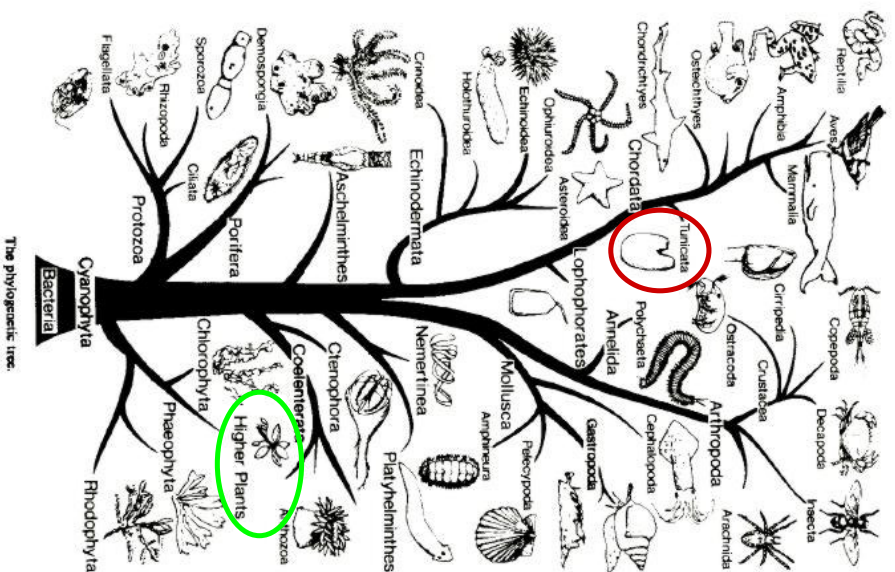
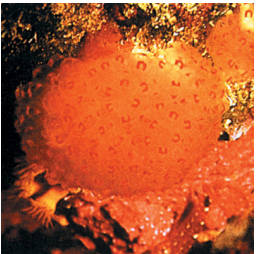
"Pharmacological Analysis of Drug-Receptor Interaction", 1987

Drug Discovery 100 years ago

P. Ehrlich (1909)



"The discoveries of those uncivilized peoples represented the sum of limitless testing of thousands of *natural materials*. By contrast with their selection of medicines by *pure chance*, we have to find first certain compounds, for example some arsenic derivatives, which show at least a low degree of therapeutic effect. Once this is done through more or less laborious tests, the purely *empirical screening* is replaced by preparing *chemical variations*, homologs and other derivatives whose efficacy has to be tested. But even at best chemical drugs are not *magic bullets*, and will not always hit only the center of the *target*, that is the disease-causing organisms. Moreover, nothing is as simple as to ascertain the lethal or the maximal well-tolerated dose, and the curative dose in a given animal species. In humans the determination of dosages is infinitely more difficult as one has to start with low doses and increase them gradually until they become therapeutically active. This is further aggravated by the occurrence of *congenital or acquired idiosyncrasies* from most medicines... and it cannot justly be demanded that a decision be made within a few months as to the merits or demerits of such new agents."

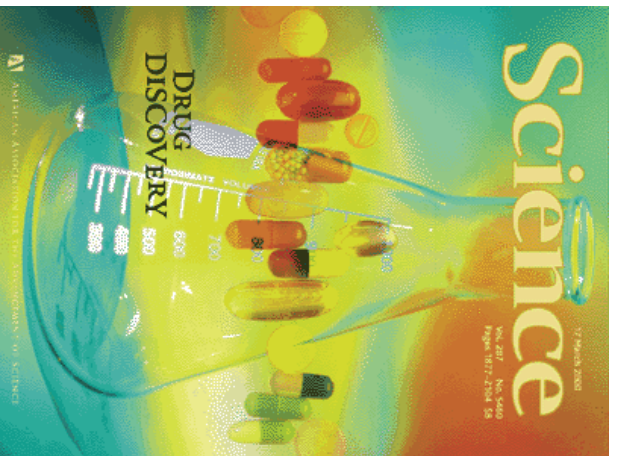


Drug discovery:

Iterative process of *make* and *test*

A new era due to the **synergy** of:

- ✓ Genomics & Proteomics
- ✓ Large collections of biologically active molecules
- ✓ High-throughput assays



Commonly used terms in drug discovery

High throughput screen: an optimised, miniaturised assay format that enables the testing of >100,000 chemically diverse compounds per day.

Assay: a test system in which biological activity can be detected.

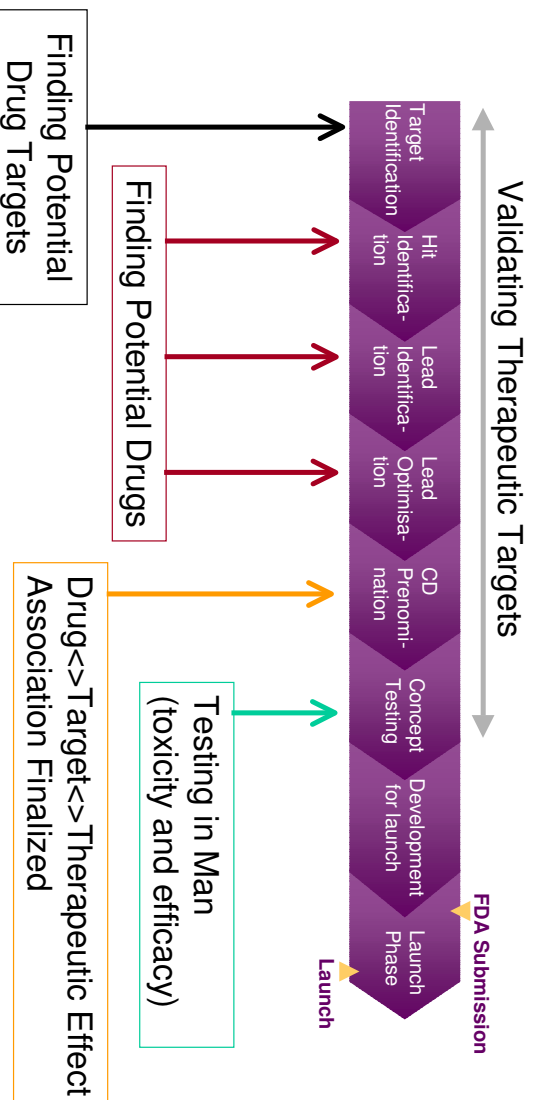
Hit: a molecule with confirmed concentration-dependent activity in a screen, and known chemical structure. The output of most screens.

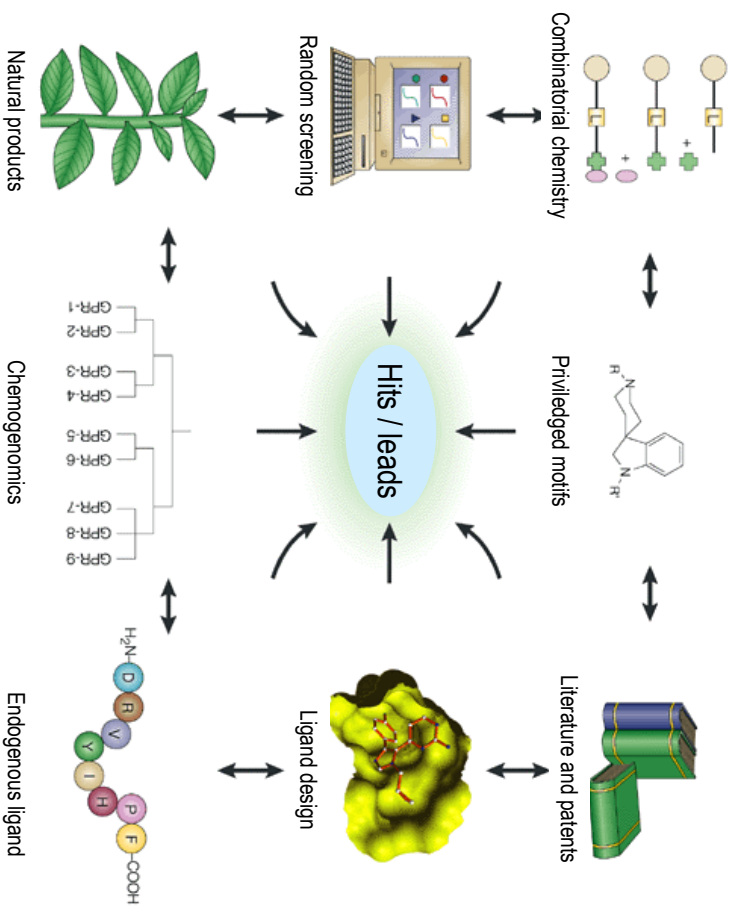
Progressible hit: a representative of a compound series with activity via acceptable mechanism of action and some limited structure-activity relationship information.

Lead: a compound with potential (as measured by potency, selectivity, physico-chemical properties, absence of toxicity or novelty) to progress to a full drug development programme.

Pharmacophore: minimal structure with essential features for activity

The Drug Discovery Pipeline





Nature Reviews | Drug Discovery

The technology drivers of change

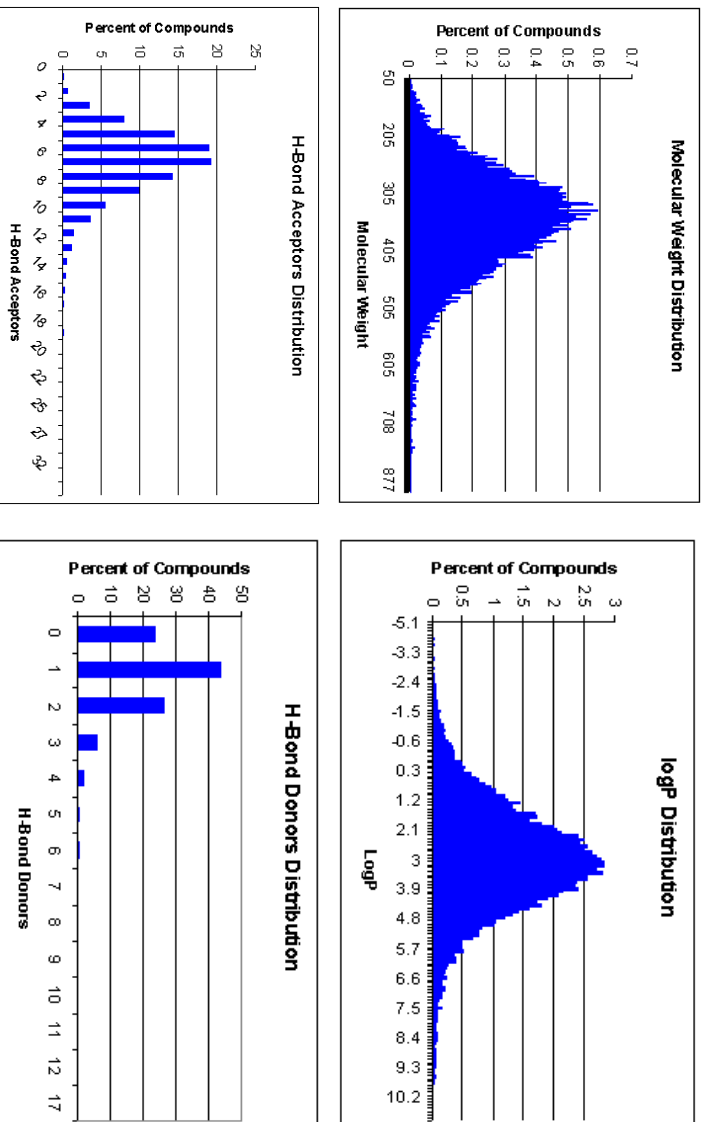


We must be able to understand:

- the properties that are required for
a good drug

- what makes **a good drug target**

Drug-like characteristics



Lipinski, C. A. *et al.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Deliv. Rev.* 1997, 23, 3-29.

LIPINSKI'S "rule of five"

H-bond donors <5

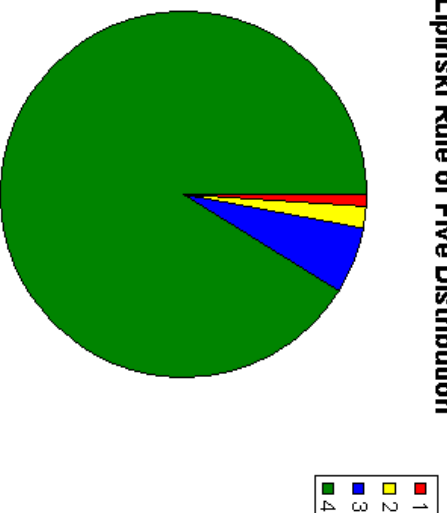
H-bond acceptors (N, O) <10

cLog P <5

Molecular Weight <500 Da

Lipinski, C. A. *et al.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and developmental settings. *Advanced Drug Deliv. Rev.* **1997**, 23, 3-29.

Lipinski Rule of Five Distribution

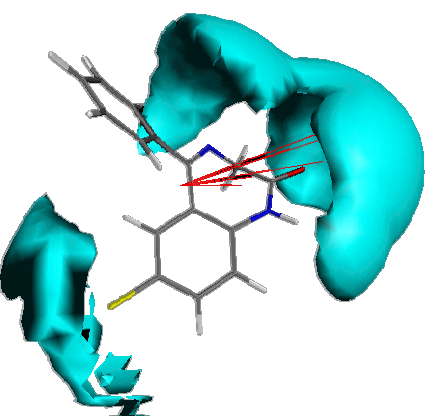


- 1 - compounds which satisfy 1 requirement - 1% of all compounds
- 2 - compounds which satisfy 2 requirements - 2% of all compounds
- 3 - compounds which satisfy 3 requirements - 6% of all compounds
- 4 - compounds which satisfy 4 requirements - 91% of all compounds



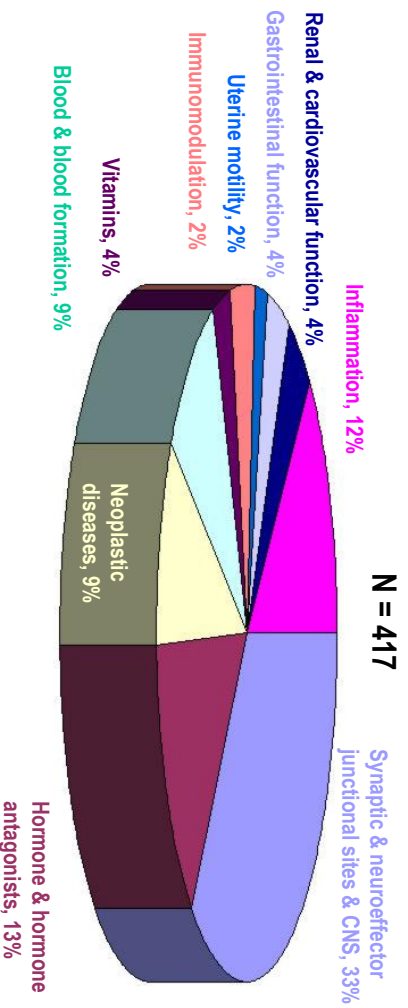
: a computational procedure to produce 2D molecular descriptors from 3D interaction energy grid maps

- ✓ The basic idea of VoISurf is to compress the information present in 3D maps into a few **2D numerical descriptors** which are very simple to understand and to interpret.
- ✓ The inherent information is summarized and interpreted in **physicochemical terms**.
- ✓ VoISurf descriptors are specifically designed for the **optimization of pharmacokinetic properties**.

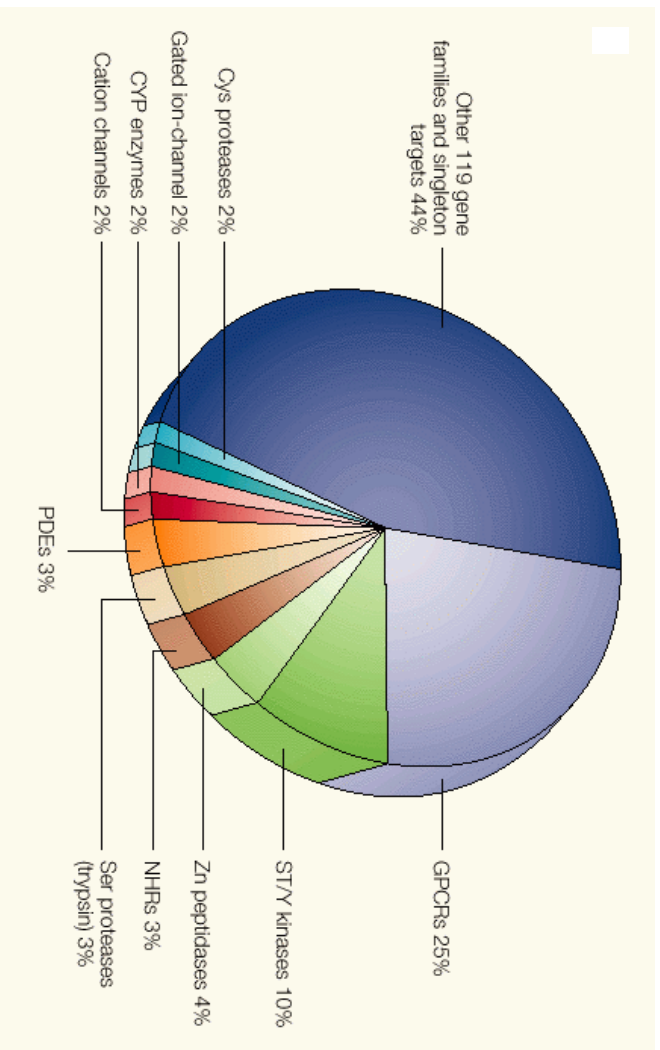


Silvio Mecucci, Gianluca Storma,
Manuel Pastor & Gabriele Cruciani

DRUG THERAPY TARGETS

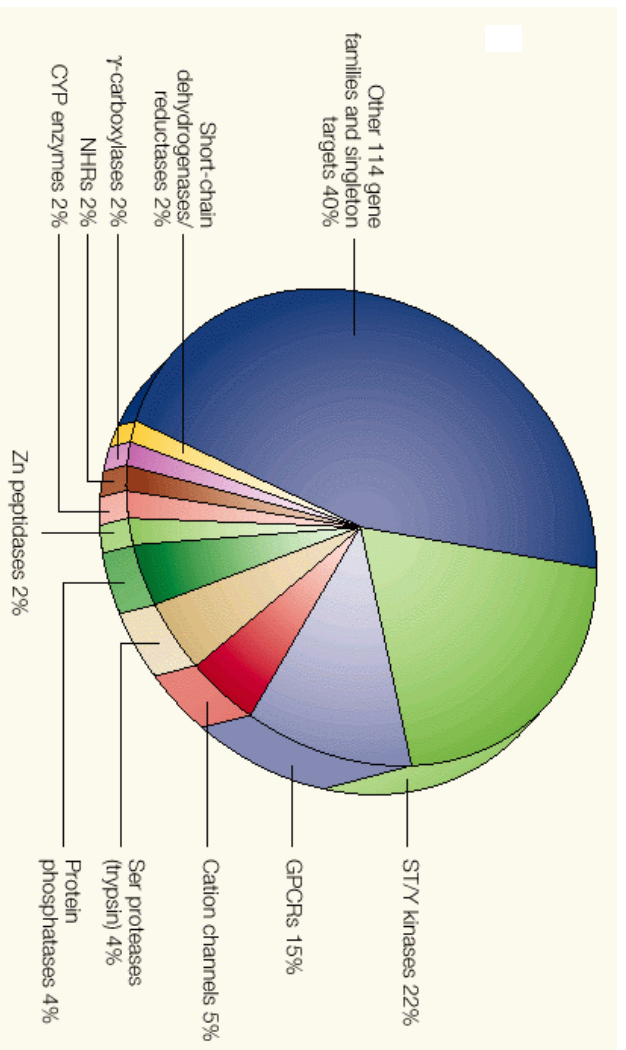


Goodman and Gilman. *The Pharmacological Basis of Therapeutics*, Edn. 9 (1996)



Gene-family distribution of the **molecular targets** of current **rule-of-five-compliant experimental and marketed drugs**

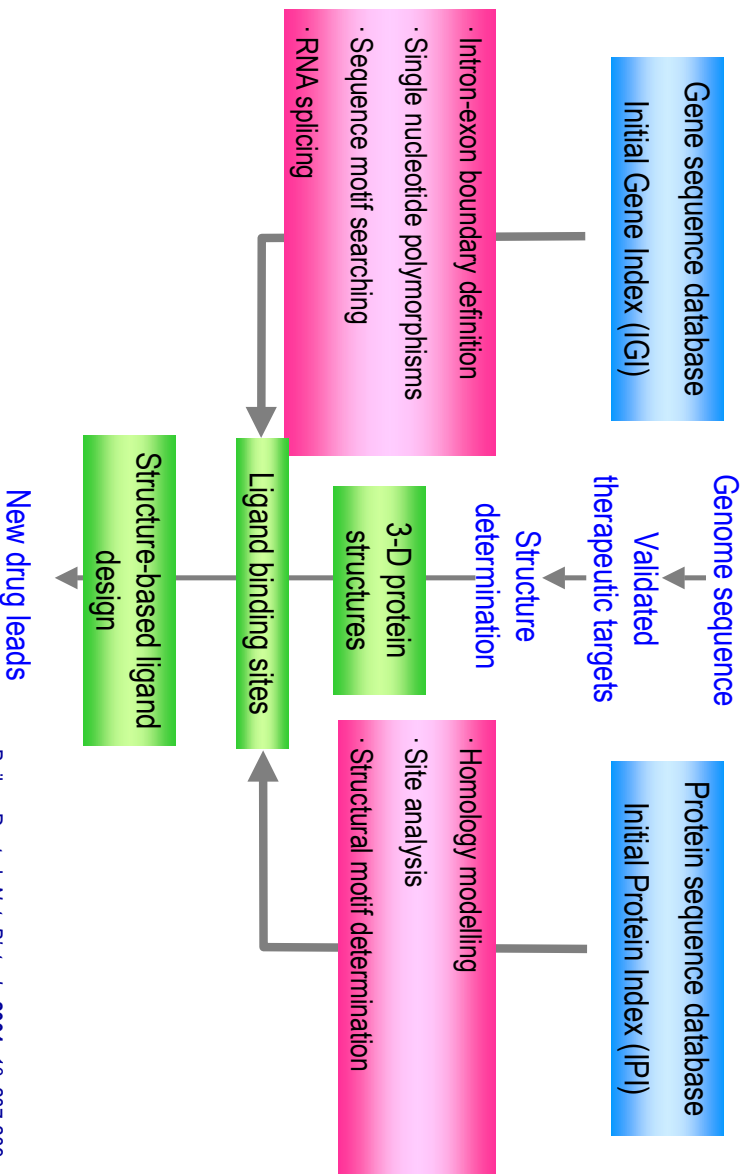
(~130 families; ~400 non-redundant molecular targets)



Gene-family distribution of **the druggable genome**

(based on known numbers of genes in the same families where members have been shown to be modulated by small-molecule drugs: ~3,000 genes)

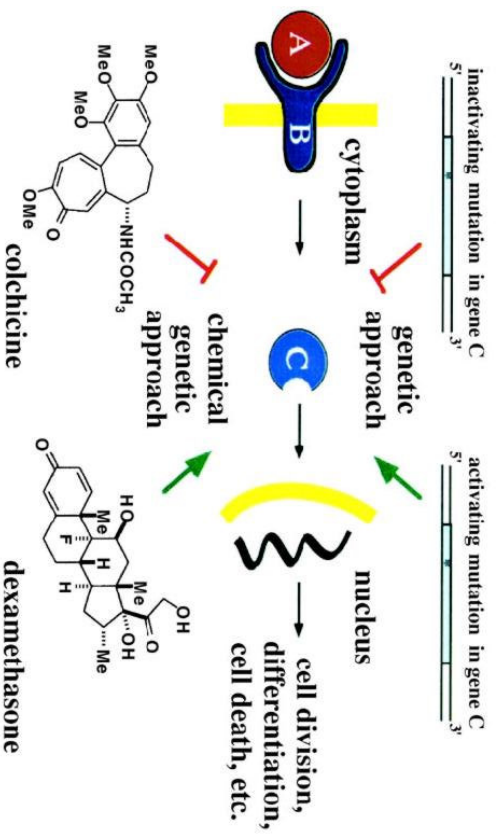
Genomic Information-driven Drug Discovery



Bailey, D. et al. *Nat. Biotech.* 2001, 19, 207-209

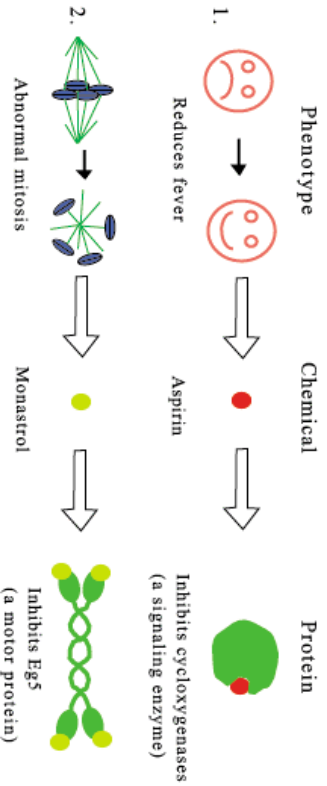
CHEMICAL GENETICS

Natural products and natural product-like compounds are used to **understand and control** the cellular and physiological functions of proteins



Forward Chemical Genetics

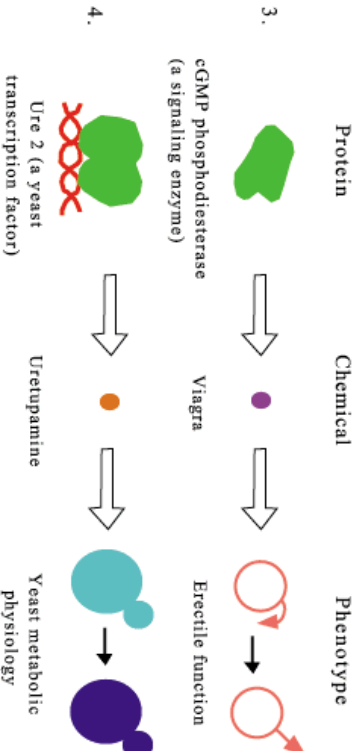
Small molecules are found that cause specific **phenotypes** in cells and organisms.



The protein **target** of the chemicals is then determined.

Reverse Chemical Genetics

Small molecules are found that **bind to, and/or disrupt** the function of, pure proteins **in vitro**.



The chemicals are then used to study the **effects** of deleting the function of the protein in a cell or organism.

Creating chemical diversity from a basis set of building blocks

Units	Library entities
Basis Set of 20 (e.g. natural amino acids)	
20^3	8,000
20^4	160,000
20^5	3,200,000
Basis Set of 100	
100^3	1,000,000
100^4	100,000,000
100^5	10,000,000,000
Basis Set of 1000	
1000^3	1,000,000,000
1000^4	1,000,000,000,000
1000^5	1,000,000,000,000,000

Combinatorial Chemistry: a googol of molecules

Avogadro's number: $6.023 \cdot 10^{23} \text{ mol}^{-1}$

Googol (*): 10^{100}

Googolplex: $10^{10^{100}}$



(*) Word coined in 1938 by Milton Sirota, the 9-year-old nephew of the American mathematician Edward Kasner (1878-1955).

COMBINATORIAL ORGANIC CHEMISTRY

BROAD SCREENING

- Huge size library
- Broadest structural diversity
- No special initial structure goal
- Any building blocks
- Undefined order of reaction
- Flexible synthetic strategy
- Site of tether not crucial
- Ligand possibly uncouplable
- Single selection evolution



CHEMICAL ANALOGING/OPTIMIZATION

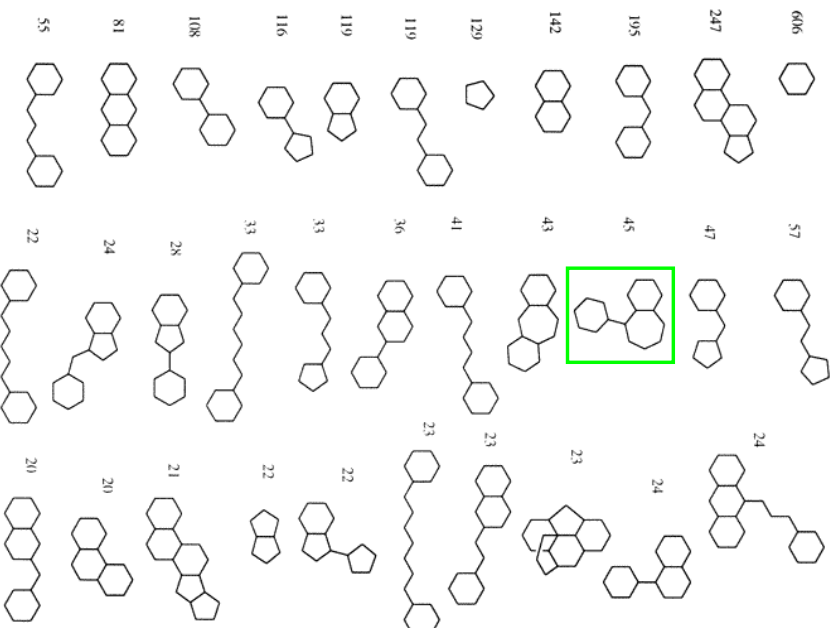
- Modest size library
- Relatively narrow structural diversity
- Specific structural goal
- Specific retrocombinatorial building blocks
- Specific order of combination
- Well defined synthetic strategy
- Tether crucial-build in redundancy
- Ligand should be releasable
- Cumulative selection evolution



32 GRAPH FRAMEWORKS

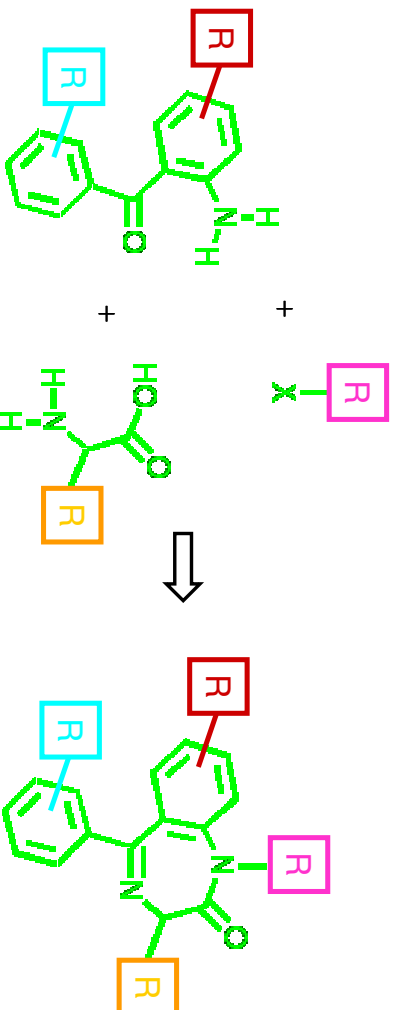
for Compounds in the
*Comprehensive Medicinal
Chemistry Database*
as classified by Connectivity Triangles

(numbers indicate frequency of
occurrence in a total of 2548 drugs)



G. W. Bemis & M. A. Murcko
(*J. Med. Chem.* **1996**, 39, 2887-2893)

COMPONENTS OF A BENZODIAZEPINE LIBRARY (one of Medicinal Chemistry's most notable pharmacophores)

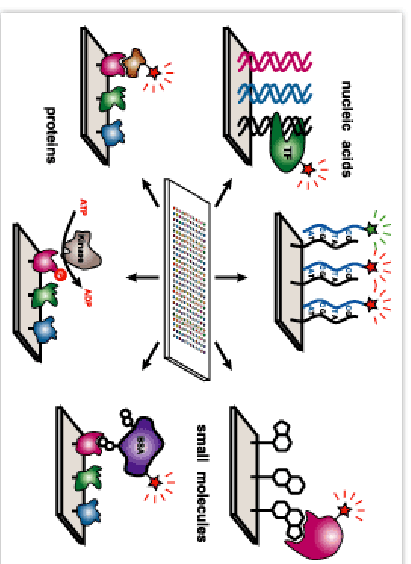


ASSAY PROCEDURES

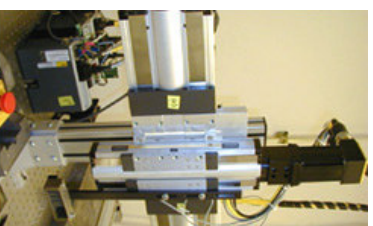
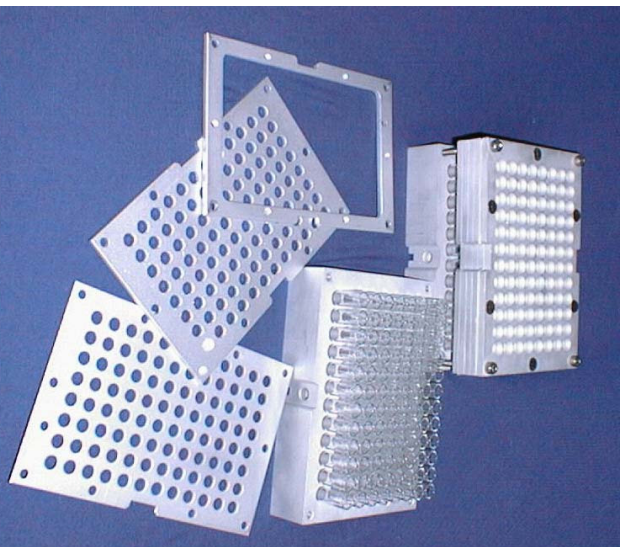
Successful use of combinatorial libraries is highly dependent on the sensitivity and specificity of the assays that are used to identify and characterize ligands

[Assay formats:](#)

- Affinity purification with an immobilized target receptor
- A labeled soluble receptor binds to tethered ligands
- Soluble compounds are tested for activity: competition binding, enzyme inhibition, or cell-based bioassays.

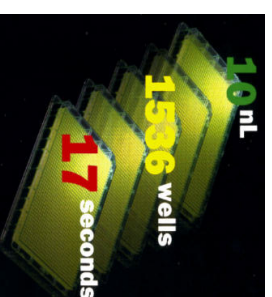


HIGH-THROUGHPUT SCREENING



automation

miniaturization

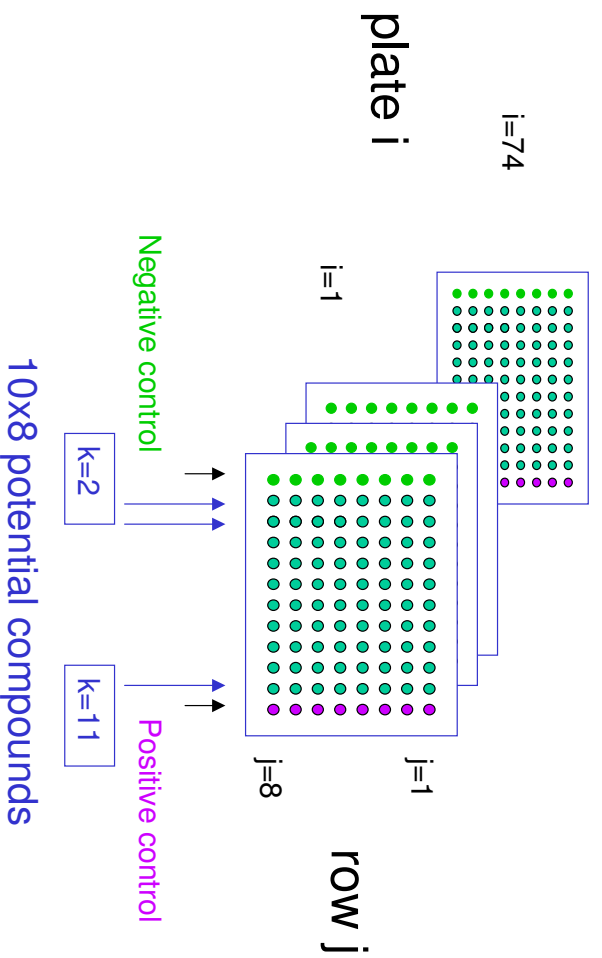


speed

From the 96-well format reactor to the 384- and the 1536-well microplate

High Throughput Screening of chemical compounds

- **Purpose:** at early stages of drug development, screen a large number of potential chemical compounds, in order to find any interaction with a given class of compounds (a "hit")
- The classes may be substructures of libraries of compounds involving up to 10^5 members.
- Each potential compound interaction with class member is tested once and only once



Implementation of HTS

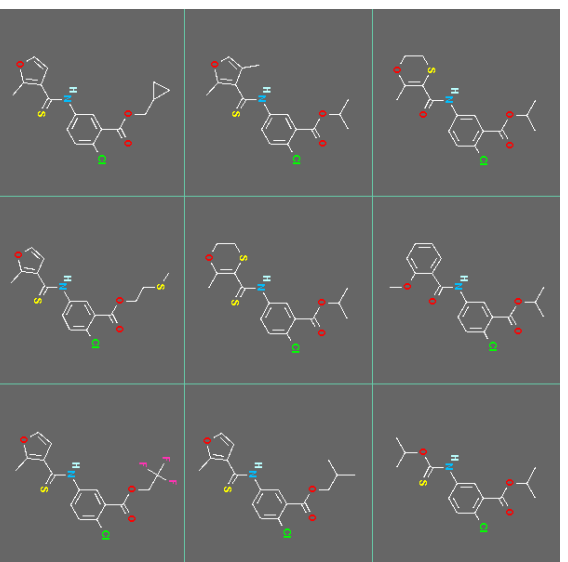
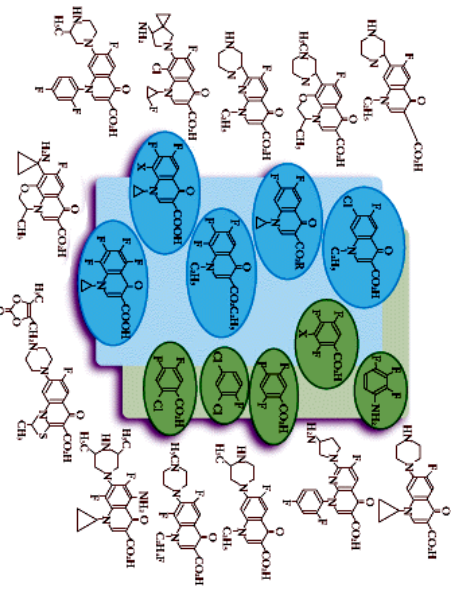
- 1) **suitable libraries of compounds:** in-house collections (5×10^5 - 10^6), specialist companies, combchem....
- 2) **assay method configured for automation:** radioligand binding assays, cell-based fluorescence and radiotracer assays, melanophore assays, reporter gene assays, cell viability assays, cell proliferation assays...

3) **robotics workstation** (multi-well formats): full automation, 24 h continuous operation, more efficient and economical.

4) **computerised data handling system:** accurate and reproducible.

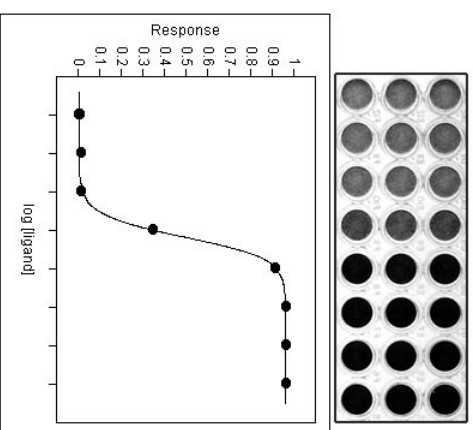
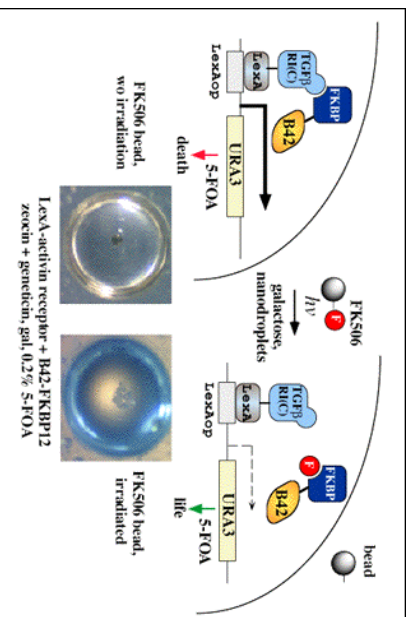
Implementation of HTS

- 1) **suitable libraries of compounds:** in-house collections (5×10^5 - 10^6), specialist companies, combchem....



Implementation of HTS

2) assay method configured for automation: radioligand binding assays, cell-based fluorescence and radiotracer assays, melanophore assays, reporter gene assays, cell viability assays, cell proliferation assays...



16,320 compounds from a chemical library
 ↓ whole-cell immunodetection assay

139 cell-permeable compounds that caused increases in phosphonucleolin staining in A549 cells

in vitro tubulin polymerization assay

86 no effect 52 destabilization 1 stabilization

fluorescence microscopy

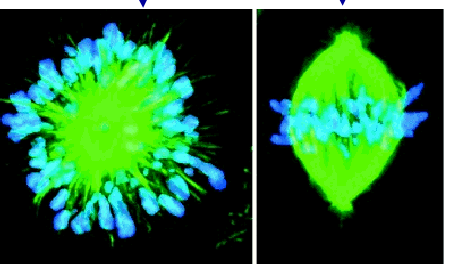
27 no visible effects

12 pleiotropic effects

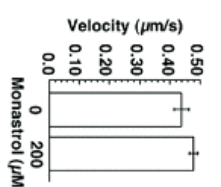
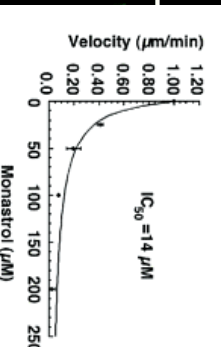
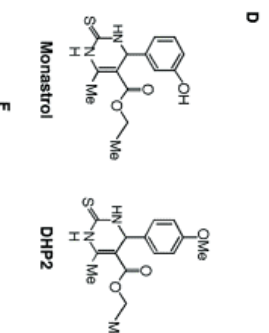
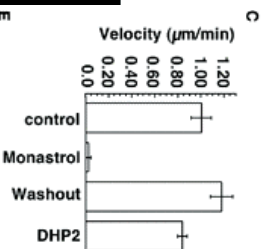
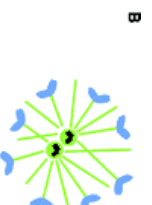
42 affect interphase and mitosis

5 affect only mitosis

α -tubulin
chromatin



Small Molecule Inhibitor of Mitotic Spindle Bipolarity Identified in a Phenotype-Based Screen



Eg5-driven microtubule motility

conventional kinesin-driven microtubule motility

Chinese Hamster Ovary cells
overexpressing the human
INSULIN RECEPTOR

**Discovery of a Small Molecule Insulin Mimetic
with Antidiabetic Activity in Mice**

Incubation with
insulin or test compounds

Immunopurification of the
heterotetrameric
INSULIN RECEPTOR

Binding
of insulin

>50,000 mixtures of synthetic
compounds and natural products

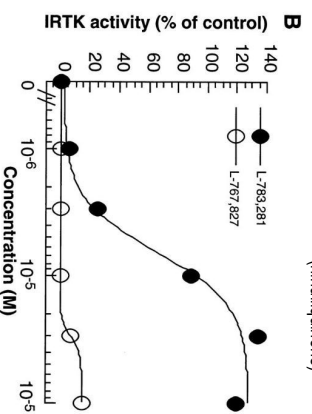
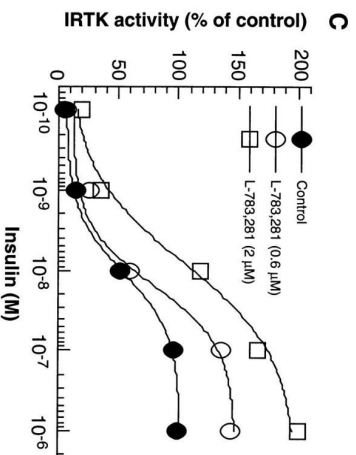
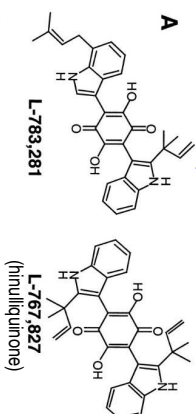
cell-based screening assay
96-well plates (150,000 cells/well)

Pseudomassaria sp.



Assay for
tyrosine kinase activity

cell membrane
Autophosphorylation

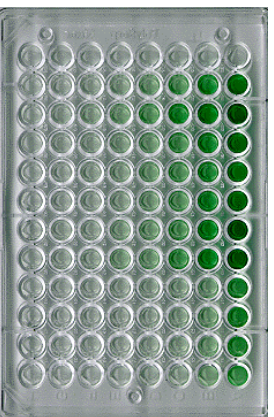


Enhancement of insulin-stimulated tyrosine kinase activation

Zhang et al. *Science* 1999, 284, 974-977

Implementation of HTS

3) **robotics workstation**: full automation, 24 h continuous operation,
more efficient and economical.



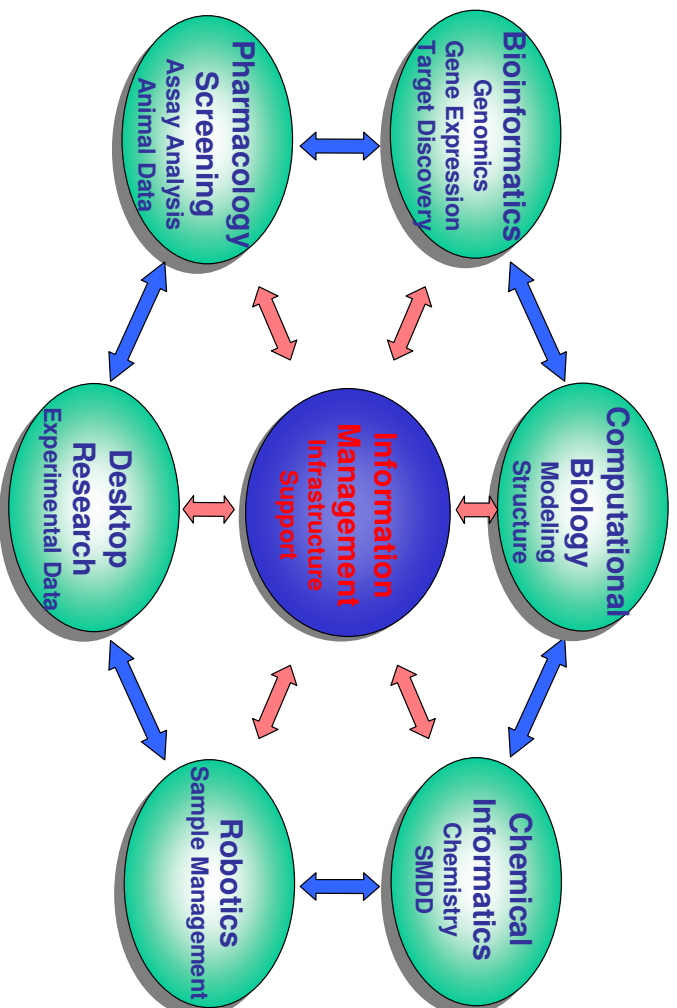
multi-well format



microarray format

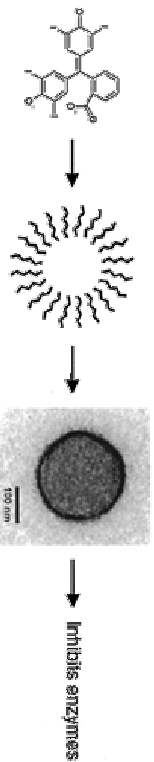
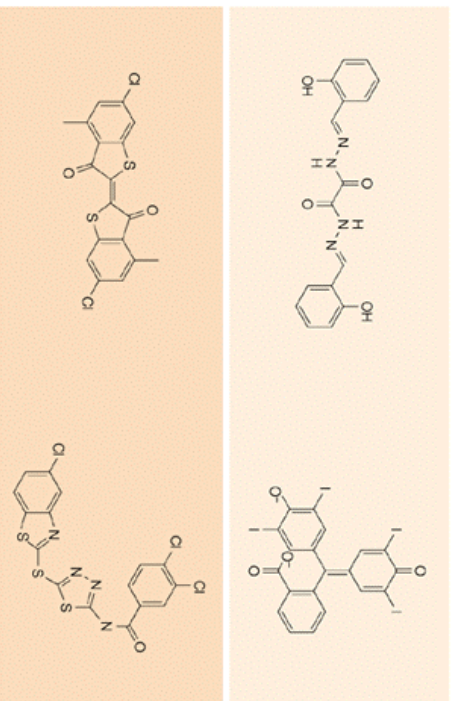
Implementation of HTS

4) computerised data handling system: accurate and reproducible.



Martina McGloughlin © 2001

Frequent hitters: “promiscuous binders”



McGovern SL, Caselli E, Grigorieff N, Shoichet BK.
A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening.
J. Med. Chem. (2002) 45:1712-1722

Also *in silico* (?)

Successful ligand-receptor pairings

1. Orphan receptor strategy

Five novel peptides/peptide families

- ✓ nociceptin/orphanin FQ (N/O/FQ)
- ✓ hypocretins/orexins (Hcrts/Oxs)
- ✓ prolactin releasing peptide (PRRP)
- ✓ apelin
- ✓ ghrelin

Pairing of 3 known peptides to corresponding orphan receptor

- ✓ melanin concentrating hormone
- ✓ urotensin II
- ✓ neuromedin U

2. Reverse pharmacology

Pairing of 6 peptides to orphan R

- ✓ melanin concentrating hormone
- ✓ urotensin II
- ✓ neuromedin U
- ✓ motilin
- ✓ neuropeptide FF
- ✓ neuropeptide AF

Pairing of 4 lipids to orphan R

- ✓ sphingosine 1-phosphate
- ✓ lysophosphatic acid
- ✓ leukotriene B₄, C₄ and D₄
- ✓ sphingosylphosphorylcholine

Pairing of non-lipid/non-peptide

- ✓ histamine → histamine H₃ receptor
- ✓ UDP-glucose → KIAA0001 receptor

Strategies for hit identification

Random screening: All possible drug molecules screened against target.

Estimated no. of possible drug molecules is $\pm 10^{40}$!!! - *Simply not possible.*

Focused screening: A limited number of compounds are pre-selected for screening.

- Has proved successful as a hit generation strategy - Useful when 3D structure of target is known (e.g. crystal structure of a receptor).
- use computer modelling to predict optimal structure to interact with target
 - use known ligand to construct 3D pharmacophore

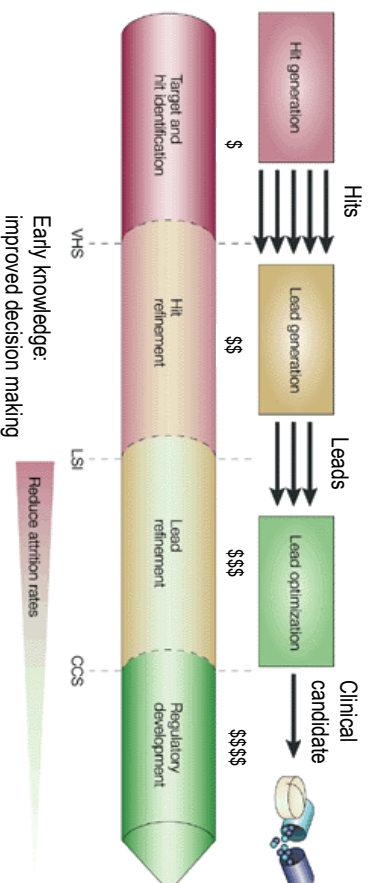
Diversity screening: The aim is to synthesize, access and test all the molecules that could be drug candidates - How many diverse samples??

'Needle in a Haystack' Syndrome

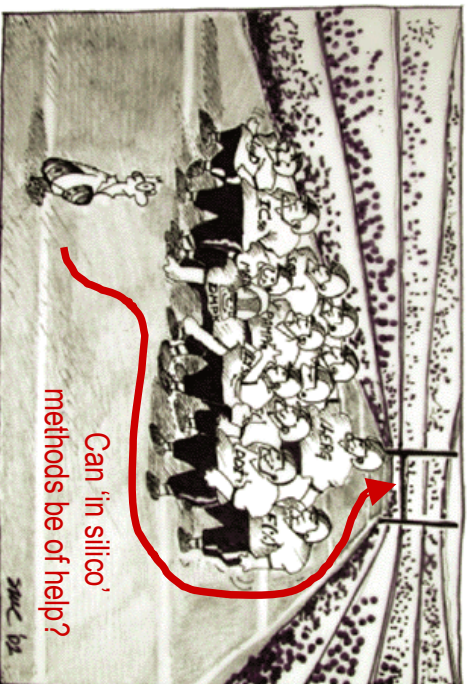


- Estimated 10^{200} compounds could be made¹
- 28 million compounds currently registered (CAS)
- Drug company biologists screen up to 1 million compounds against target using ultra-high throughput technology
- Chemists select 50-100 compounds for follow-up
- Chemists work on these compounds, developing new, more potent compounds
- Pharmacologists test compounds for pharmacokinetic and toxicological profiles
- 1-2 compounds are selected as potential drugs

¹ See http://www.daylight.com/meetings/mug98/Nicholls/The_Hitch_Hiker.html



Nature Reviews | Drug Discovery



Can 'in silico'
methods be of help?

lots of hurdles

Nature Reviews Drug Discovery 2: 369-378 (2003)
HIT AND LEAD GENERATION: BEYOND HTS

- **HTS methods are now routine procedure**

- **Pros**

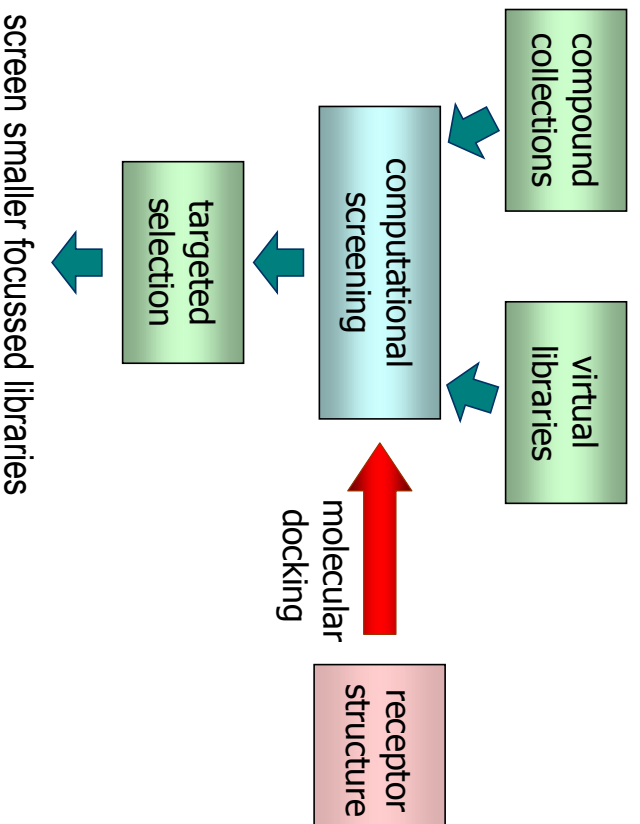
- » allows detection of possible lead molecules and structural classes
- » rapidly generate a provisional SAR relationship
- » effectively utilises in-house historical compound archive

- **Cons**

- » HTS laboratories are expensive to maintain
- » typical screens cost \$100,000 -> \$1,000,000
- » consumes valuable physical compound archive
- » commercially available compound libraries are expensive
- » patent minefield

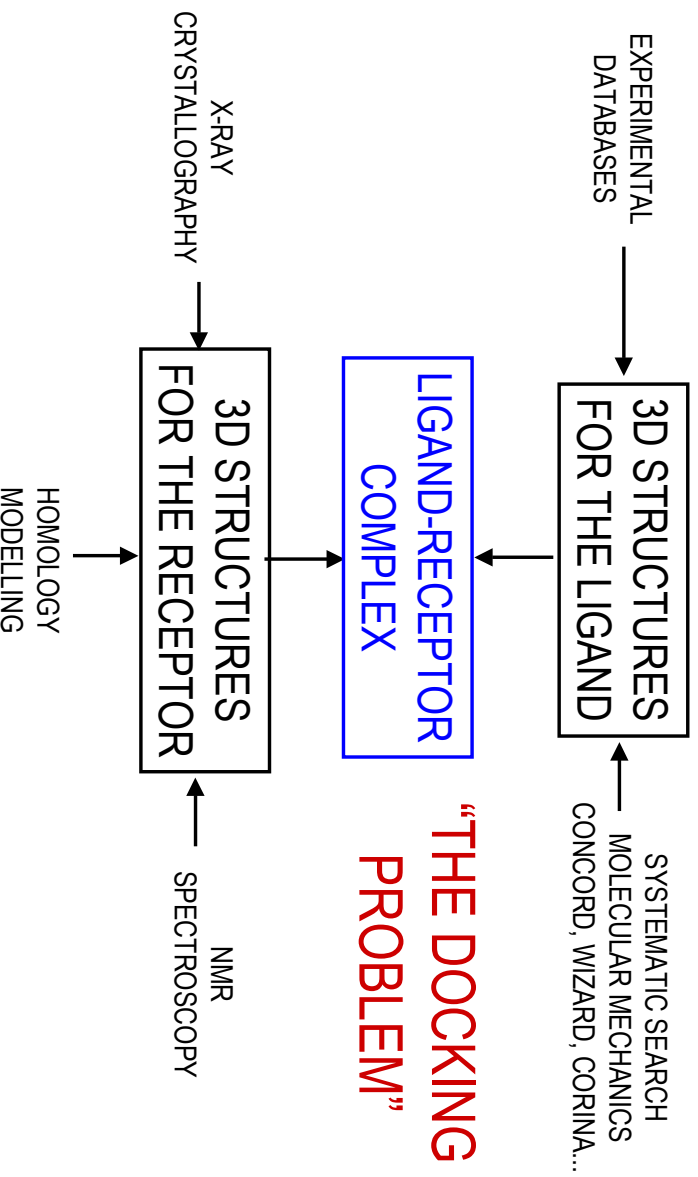
- ***in-silico* methods are therefore attractive**

Virtual Screening



Why Use Molecular Docking?

- Most detailed representation of binding site
 - overcomes simplifications of pharmacophores
 - identifies both conservative and novel solutions
 - provides impetus for *de novo* design/optimisation
- Broad range of analyses applicable
 - diverse scoring/selection criteria
- Quality/throughput of available methods
 - good enough, despite technical limitations



SITE/LIGAND REPRESENTATION
(treatment of H atoms?)

“THE DOCKING PROBLEM”
JUXTAPOSITION OF THE LIGAND AND
SITE FRAMES OF REFERENCE

EVALUATION OF COMPLEMENTARITY
(scoring functions)

OBJECTIVE: Obtain the lowest free energy structure(s) for the receptor-ligand complex.

MOLECULAR DOCKING

□ SYSTEMATIC SEARCH (*brute force algorithm*):

All binding orientations of all conformers of the ligand and the receptor (impractical for most situations).

□ AUTOMATED SEARCH:

GEOMETRIC METHODS: Matching of ligand and receptor site descriptors (descriptors, grids, fragments...).

FORCE FIELD METHODS: Minimizing the ligand-receptor interaction energy - Molecular dynamics and Monte Carlo simulations.

Virtual (“in silico”) screening

- Search a database of putative ligands for new leads. *
- Rank the selected ligands in terms of their interaction energy with a particular receptor.
- Calculate the differential binding of a ligand to two different macromolecular targets.
- Study the geometry of a particular complex.
- Propose modifications of a lead molecule to optimize interactions.

* Success at lead identification.

False positives are accepted and false negatives are not recognized.

Virtual (“in silico”) screening

Docking/scoring programs

Docking engines: search the conformational space
in the binding site

Scoring functions: discrimination of correctly docked
from misdocked conformations

Examples of docking algorithms

Rigid ligand:

Fast shape matching (DOCK)



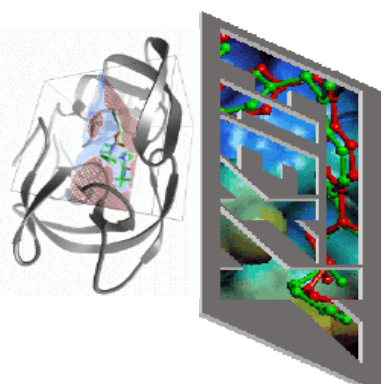
Flexible ligand:

Fast shape matching (DOCK 4.0)

Incremental construction (FlexX)

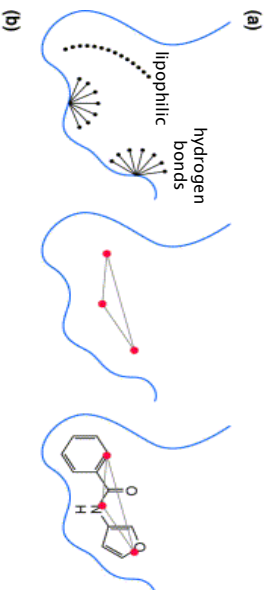
Simulated annealing (AutoDock 2.4)

Monte Carlo simulations (MCDOCK)

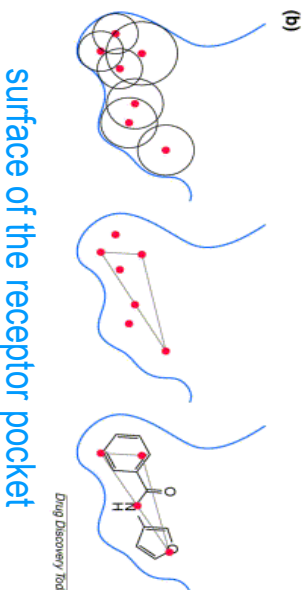


Genetic algorithm (AutoDock 3.0, GOLD, GAMBLER)

FlexX
algorithm



DOCK
algorithm



Drug Discovery Today

FlexX matches triangles of interaction sites onto complementary ligand atoms.

DOCK fills the binding site with spheres, and sphere centers are then matched to the ligand atoms to determine plausible ligand–receptor complexes.

PROGRAM DOCK

"A Geometric Approach to Macromolecule-Ligand Interactions"

I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin
J. Mol. Biol. 161, 269-288 (1982)

"Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure"

R. L. DesJarlais, R. P. Sheridan, G. L. Seibel, J. S. Dixon, I. D. Kuntz, R. Venkataraghavan
J. Med. Chem. 31, 722-729 (1988)

"Automated Docking with Grid-Based Energy Evaluation"

E. C. Meng, B. K. Soichet, I. D. Kuntz
J. Comp. Chem. 13, 505-524 (1991)

RECEPTOR COORDINATES

SITE CHARACTERIZATION

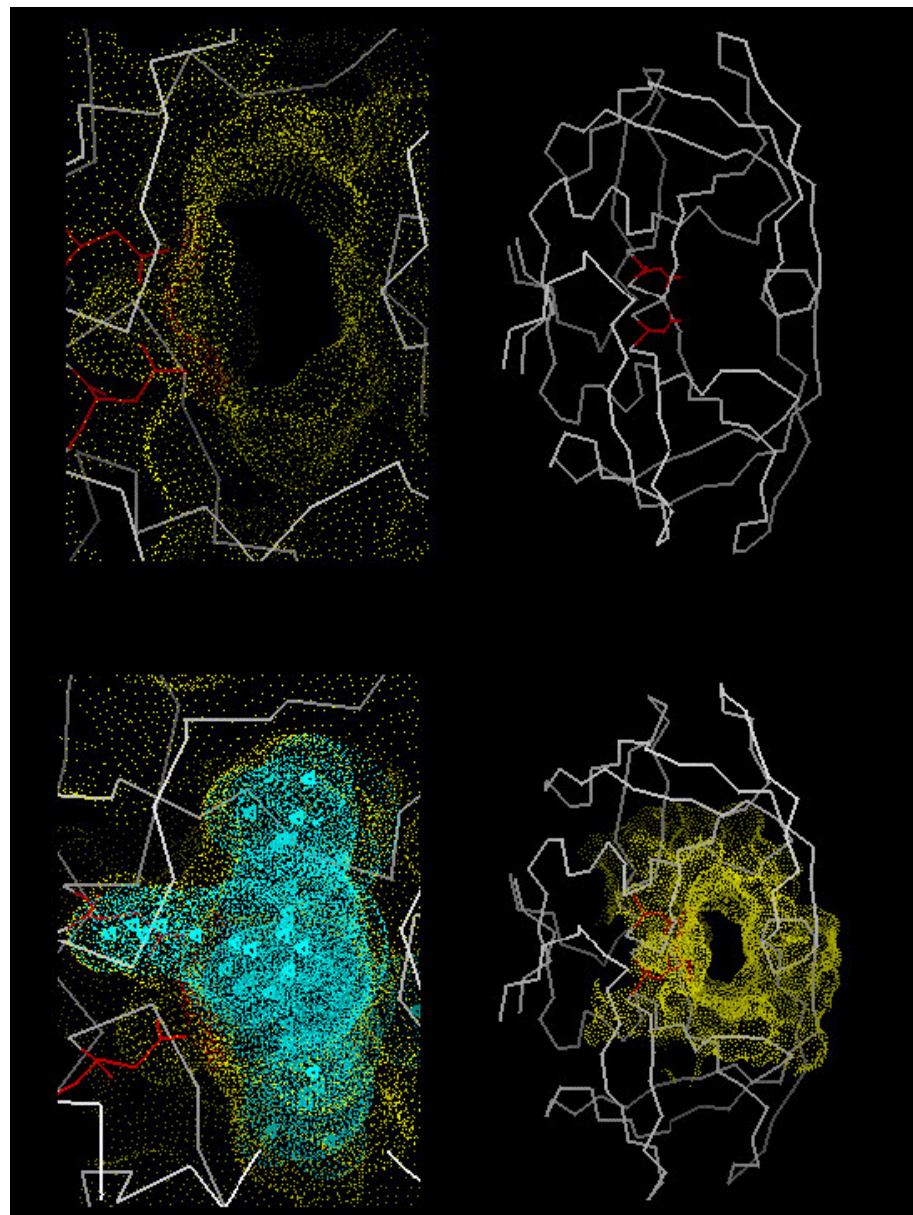
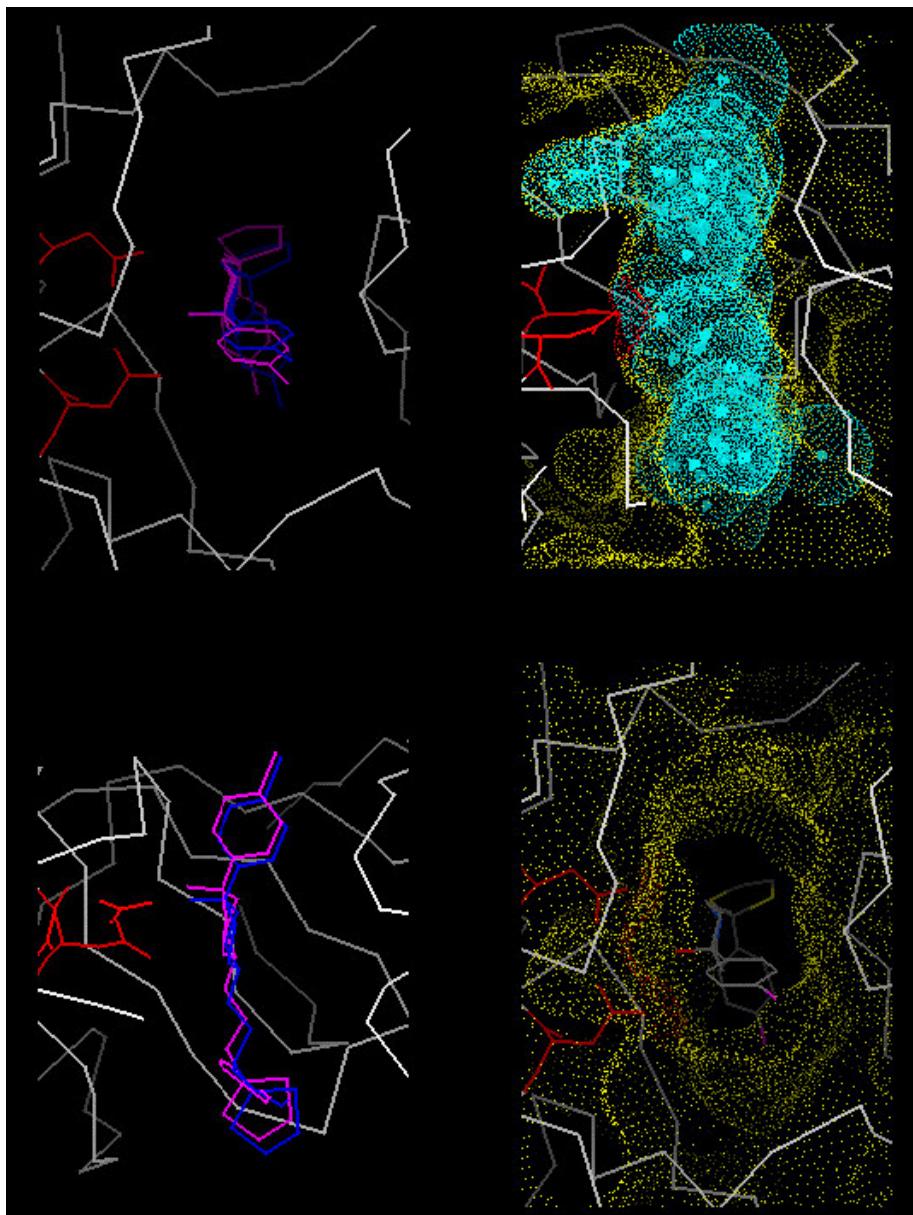
MS molecular "dot" surface
SPHGEN negative image of site

GRID CALCULATION

DISTMAP contact scoring
CHEMGRID force-field scoring

DOCKING AND SCORING

LIGAND COORDINATES → **DOCK** Matching Orientation Scoring



Docking accuracy

[Rms deviations (non hydrogen atoms, in Å) from the X-ray pose]
(top solution of each docking tool)

Docking method			
ligand	DOCK	FlexX	GOLD
deoxythymidine	0.82	0.78	0.72
5-iododeoxyuridine	9.33	1.03	0.77
5-iodouracil-anhydrohexitol	1.16	0.88	0.63
dhbt (not publicly available)	2.02	3.65	0.93
6-(3-hydroxy-propyl-thymine)	1.02	4.18	0.49
6-[6-hydroxymethyl-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl]-5-methyl-/H-pyrimidin-2,4-dione	9.62	13.30	2.33
(North)-methanocarbothymidine	7.56	1.11	1.19
aciclovir	3.08	2.71	2.74
ganciclovir	3.01	6.07	3.11
penciclovir	4.10	5.96	3.01

Only one set of protein (TK) coordinates used: pdb code 1kim

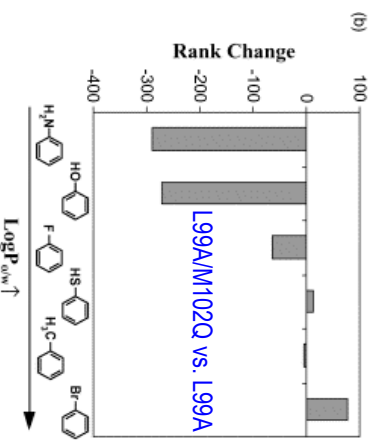
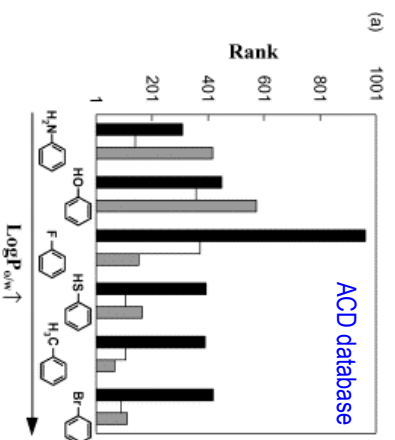
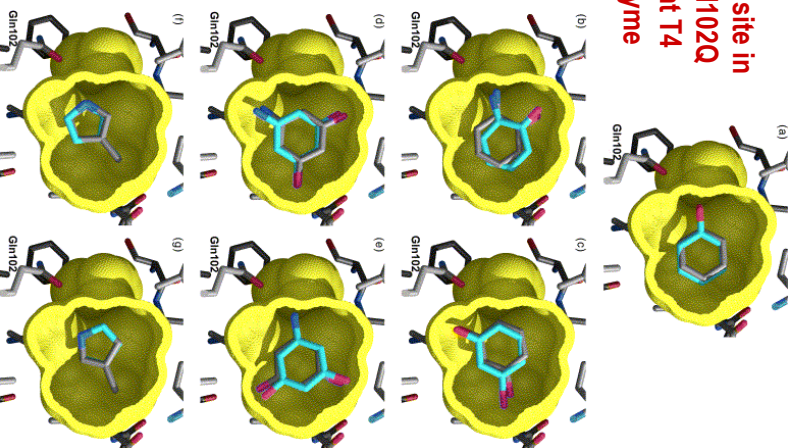
Scoring functions

Knowledge-based: statistical analysis of 3D complex structures to derive a sum of *potentials of mean force* between receptor and ligand atoms

Force field-based: calculation of van der Waals and electrostatic interaction energies between the receptor and the ligand atoms

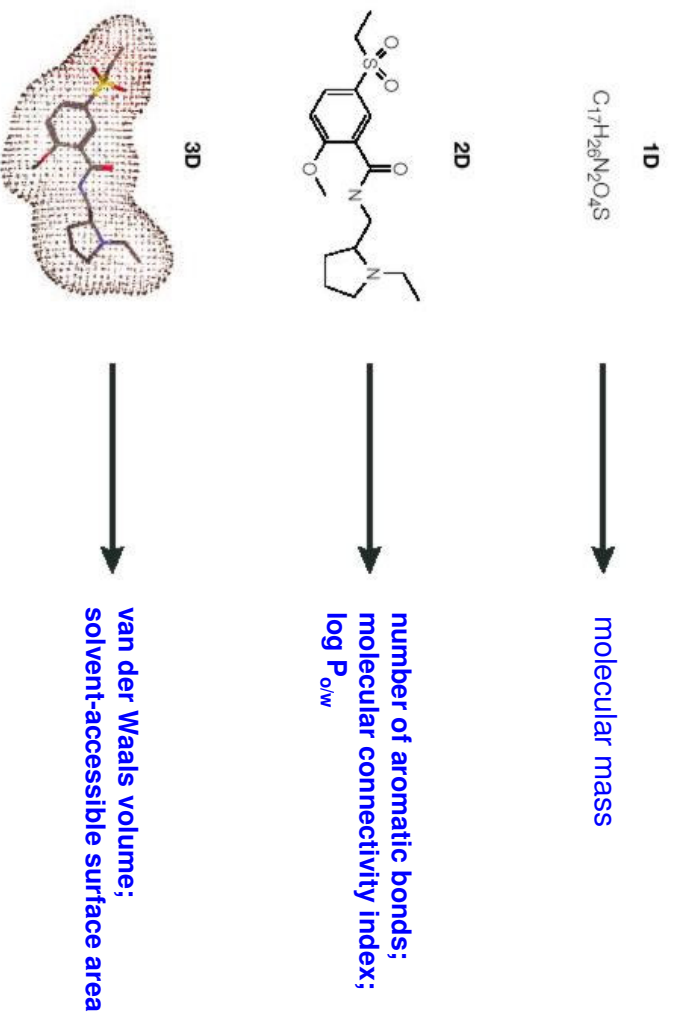
Empirical: the binding free energy is broken down into a number of different *weighted* contributions (supposed to be additive: number of hydrogen bonds, ionic interactions, apolar contacts, entropy penalties...)

**Binding site in
L99A/M102Q
mutant T4
lysozyme**

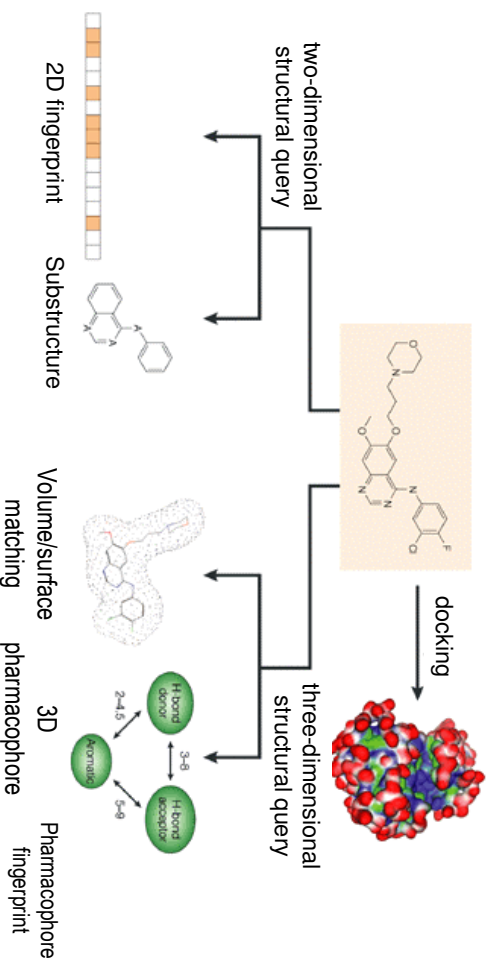


Wei BQ, Baase WA, Weaver LH, Matthews BW, Shoichet BK.
A model binding site for testing scoring functions in molecular docking.
J. Mol. Biol. (2002) 322:339-355

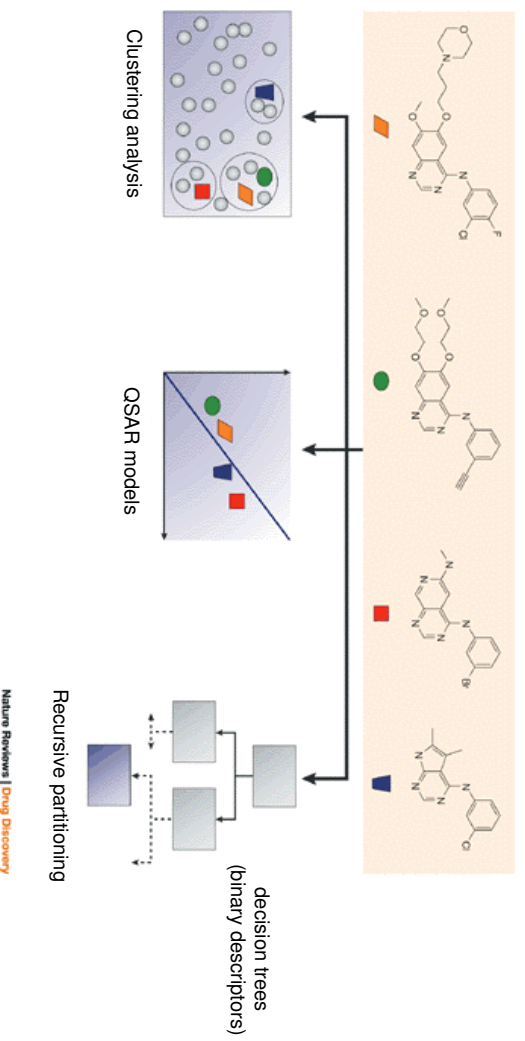
Representative molecular descriptors



Methods and tools for virtual screening: 1 single molecule as input

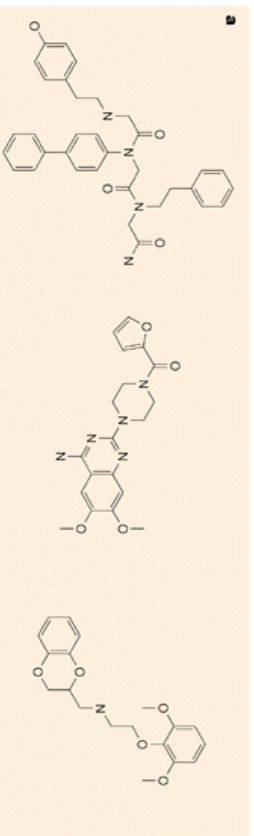


Methods and tools for virtual screening: Multiple molecules as input

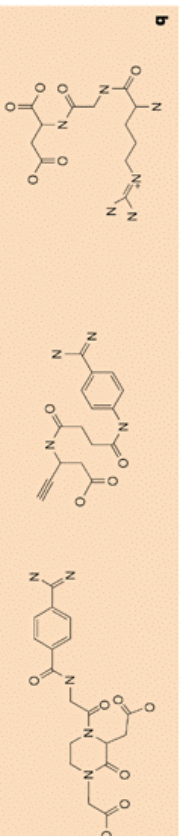


Recognition of remote-similarity relationships: “molecular fingerprints”

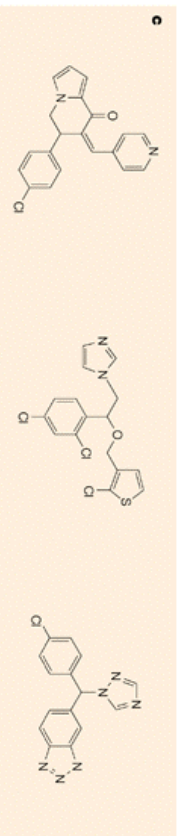
α_1 -adrenoceptor ligands



RGD-peptidomimetics

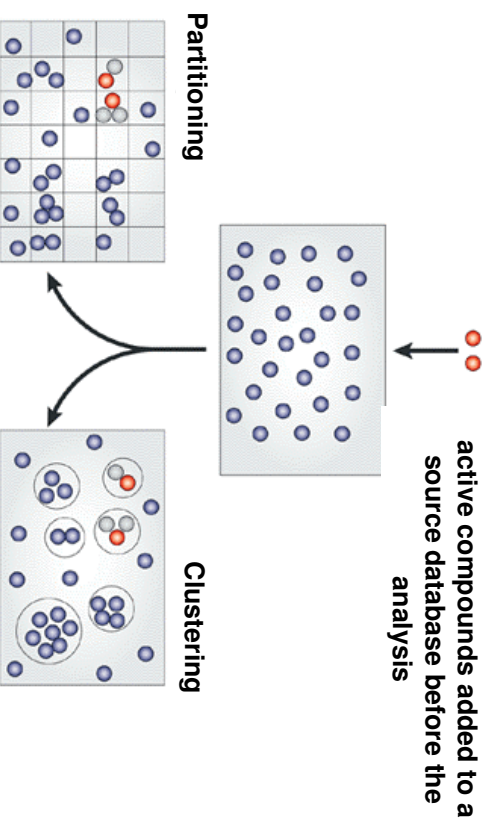


aromatase inhibitors



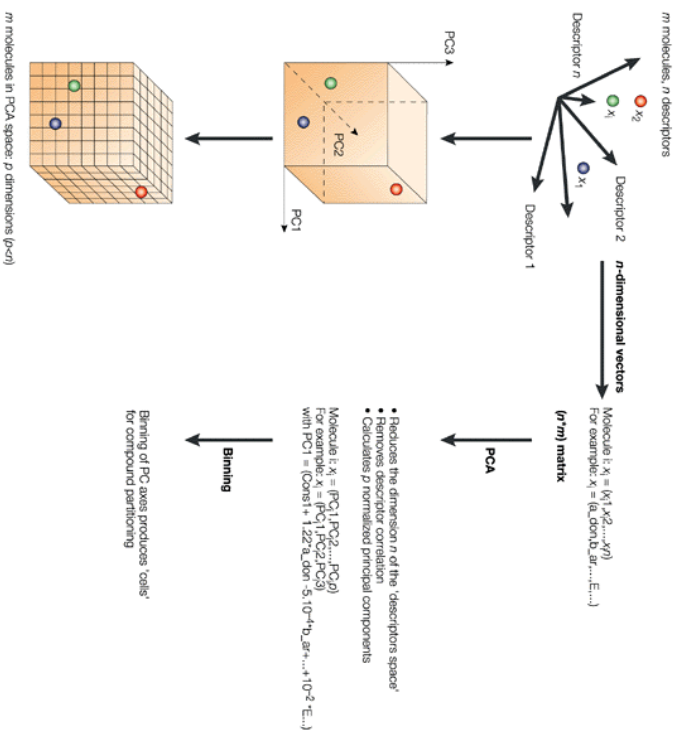
Nature Reviews | Drug Discovery

Clustering versus partitioning: methodological differences



Nature Reviews | Drug Discovery

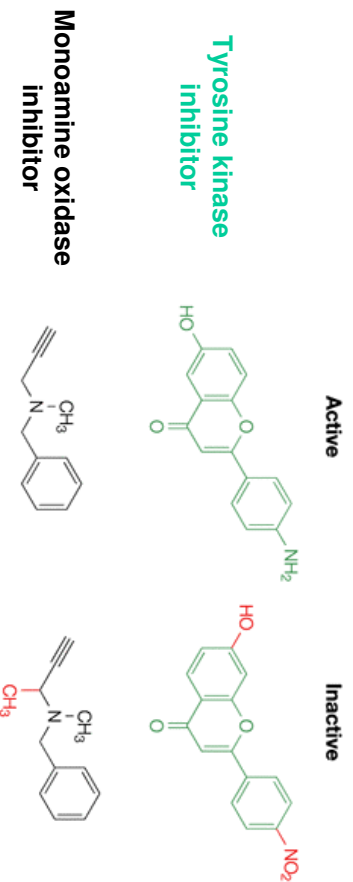
Generation of low-dimensional chemical spaces for cell-based partitioning: PRINCIPAL COMPONENT ANALYSIS



Nature Reviews | Drug Discovery

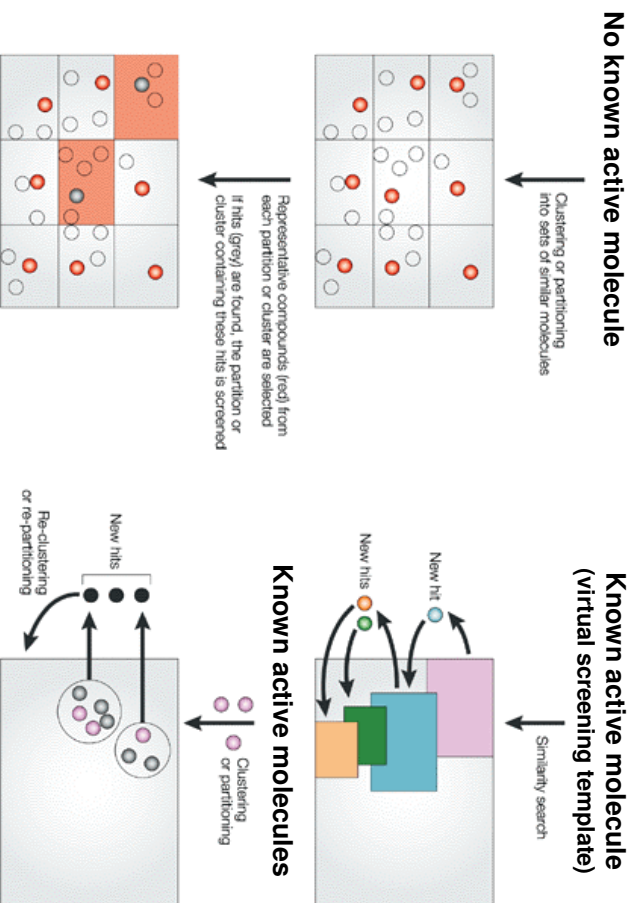
Structural similarity versus biological activity:

minor structural modifications can render some drugs inactive



Can virtual screening tell the difference?

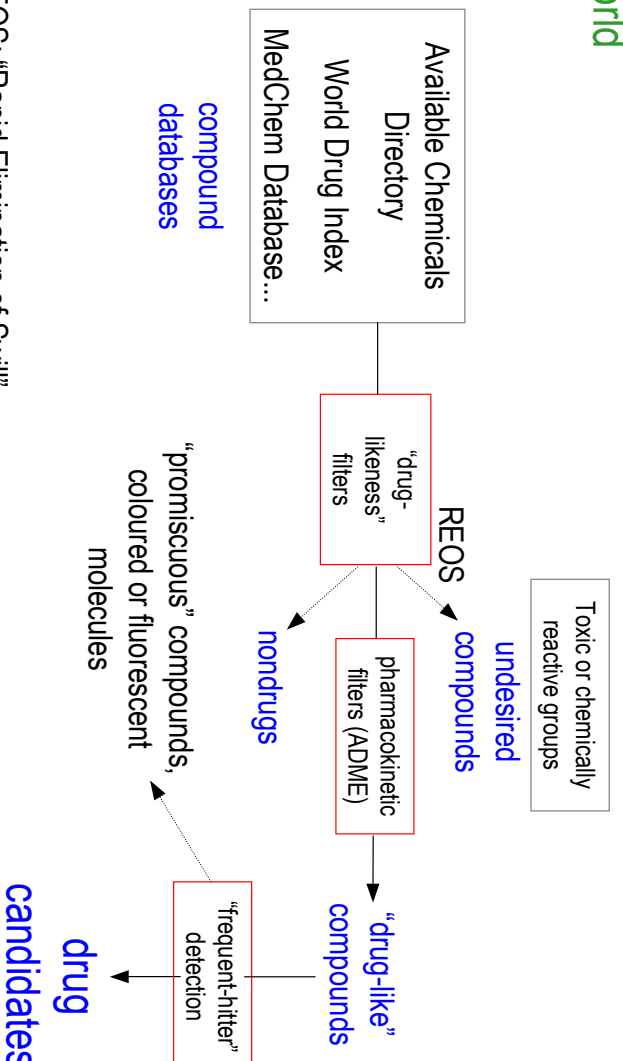
Strategies for sequential screening



Nature Reviews | Drug Discovery

In silico VIRTUAL SCREENING and FOCUSED LIBRARY DESIGN

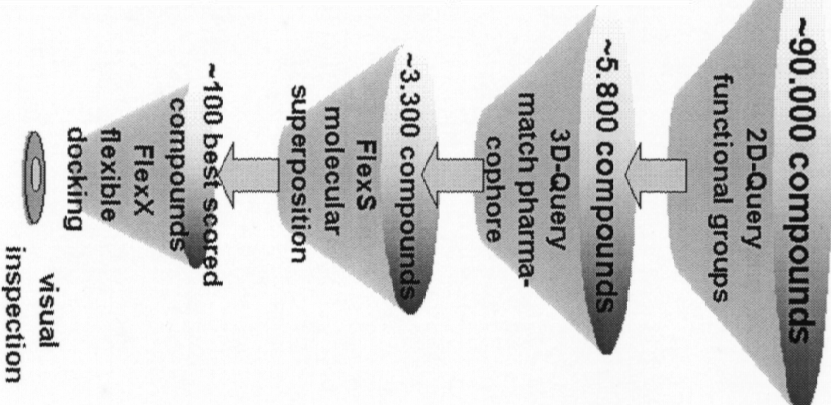
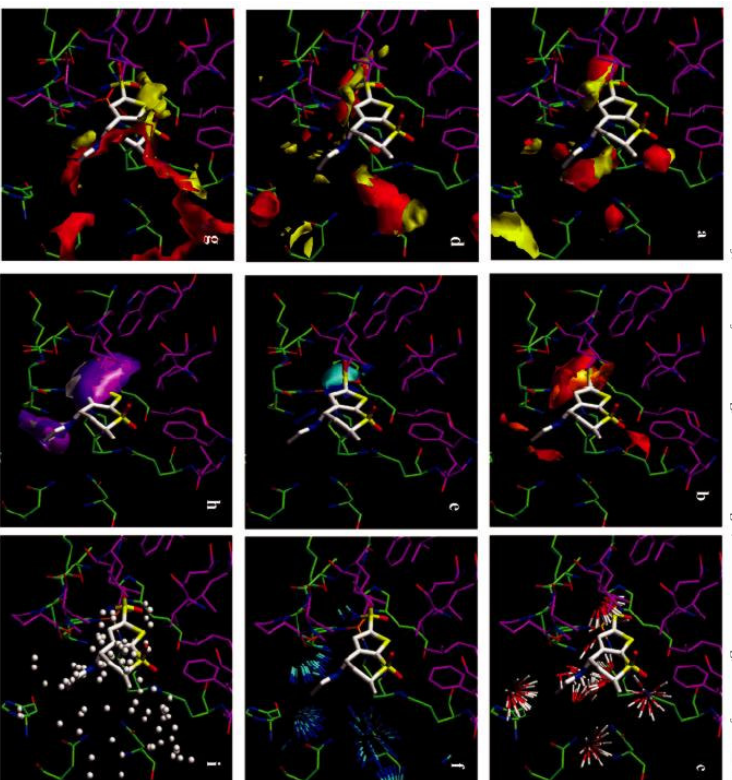
Near-perfect structures in an imperfect world



Successful Virtual Screening for Novel Inhibitors of Human Carbonic Anhydrase: Strategy and Experimental Confirmation

Sven Grunenberg,¹ Milton T. Stubbs, and Gerhard Klebe*

¹Institute of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany

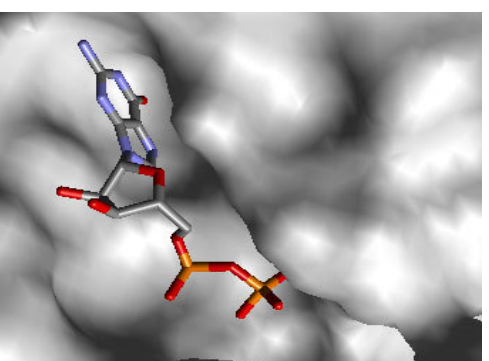


University of Marburg
ScreenSaver Lifesaver
SEARCHING FOR anti-cancer drugs by distributed computational chemistry

<http://www.chem.ox.ac.uk/ccdd/ccdd.html>




Prof. W. Graham Richards

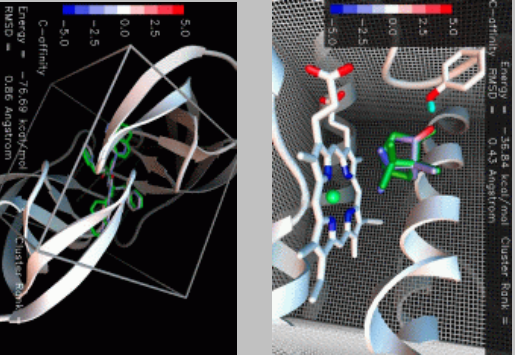


Superoxide dismutase	Vascular Endothelial Growth Factor
RAS proteins	Insulin Tyrosine Kinase
Cyclooxygenase (COX-2)	c-ABL Tyrosine Kinase
Fibroblast Growth Factor Receptor	CDK-2
RAF	Farnesyltransferase
Protein-Tyrosine-Phosphatase 1B	VEGF11

<http://www.FightAidsathome.org>



Dr. Garrett Morris



Applications of Ligand-Protein Docking in Drug Design

Existing methods

Given a protein, find potential binding ligands from a chemical database

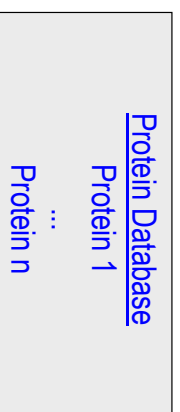


Successfully docked compounds as putative ligands

Science (1992) 257: 1078

New method

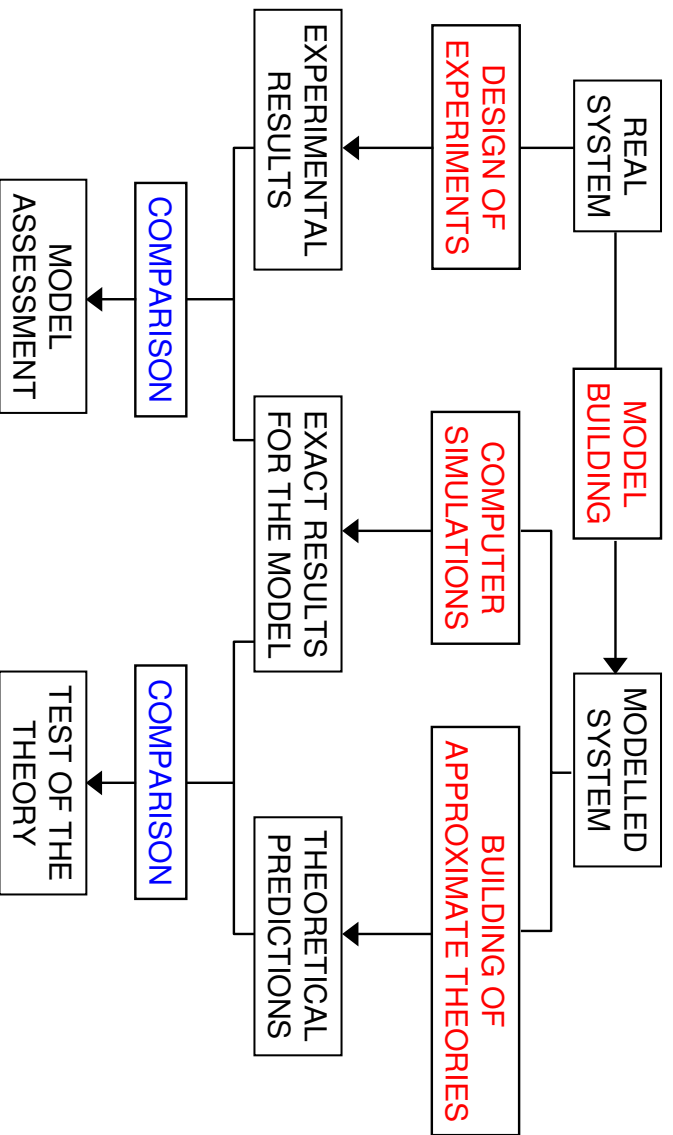
Given a ligand, find potential protein targets from a protein database



Successfully docked proteins as putative targets

Proteins (2001) 43: 217

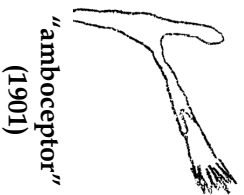
CONNECTION BETWEEN EXPERIMENT, THEORY AND COMPUTER SIMULATION



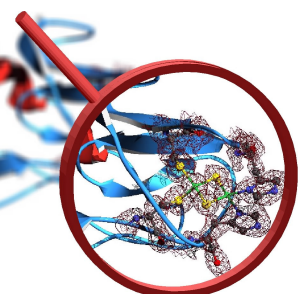
"Considering the enormous number of chemical combinations which are taken into consideration in a struggle with diseases, it will always be a caprice of **chance**, or **fortune**, or of **intuition**, which decides which investigator gets into his hands the substances which turn out to be the best materials for fighting the diseases or the basal substances for the discovery of such"

Paul Ehrlich

"Address in Pathology on Chemotherapeutics: Scientific Principles, Methods, and Results" *The Lancet*, 445 (1913)



100 years





**“he that is afraid to shake the dice
never throws a six”**

(Chinese proverb)