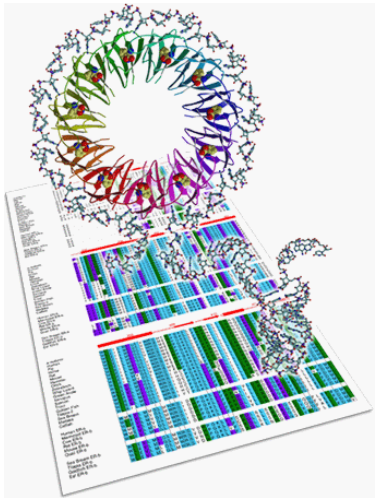
Escuela Complutense de Verano
Especialista en Bioinformática

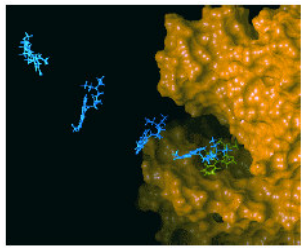
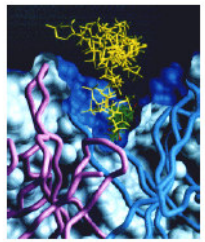
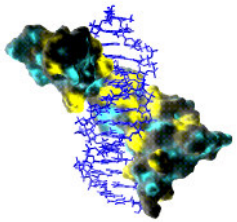


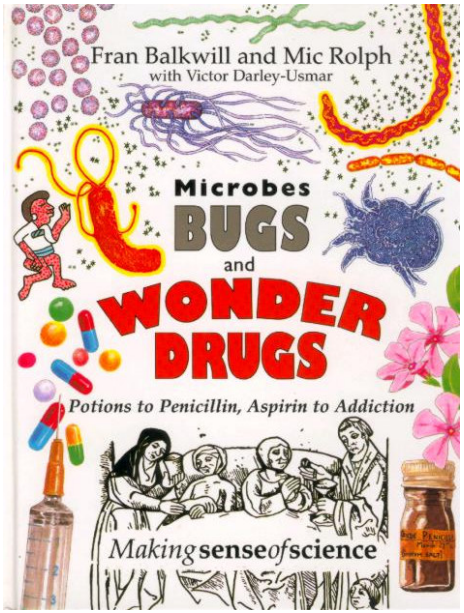
Interacciones entre proteínas y pequeños ligandos (II)

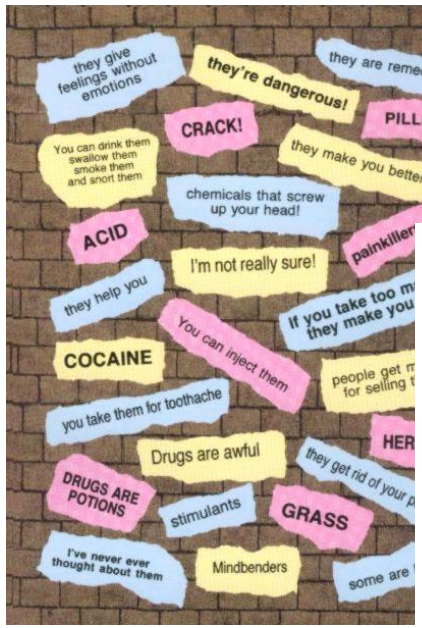
Federico Gago
Departamento de Farmacología
Universidad de Alcalá, Madrid

Major Application of Bioinformatics in 21st century: Designing Drugs

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





The word 'drug' means different things to different people. In fact, a drug is any chemical that can change the way a living creature functions. Drugs can be simple chemicals or complicated ones. They may be gas, liquid or solid. Drugs can be inhaled, swallowed, absorbed through the body surface, or injected. But all drugs work in the same way. They get very close to natural chemicals in living cells and alter what they do.

"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

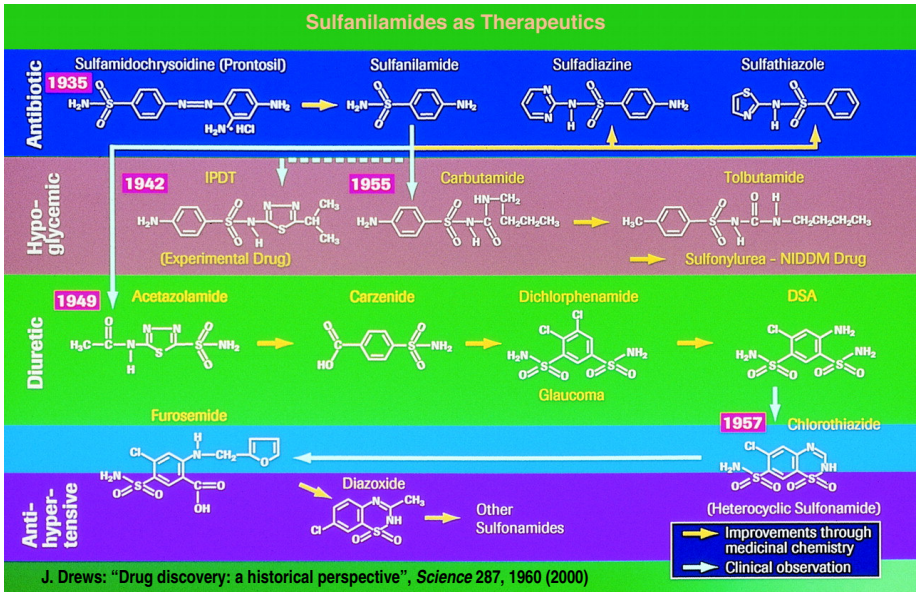
"Pharmacological Analysis of Drug-Receptor Interaction", 1987

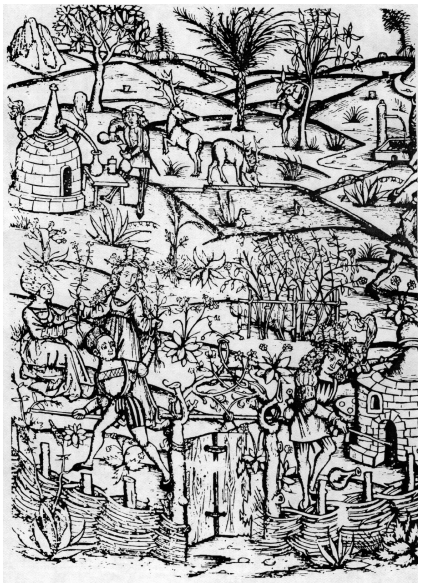
Drug Discovery 100 years ago

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."

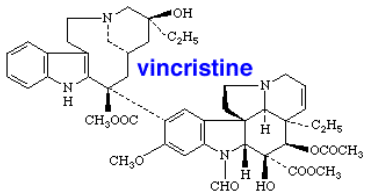
Drugs that originated from sulfanilamide

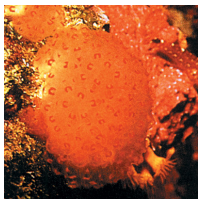


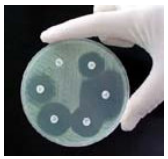


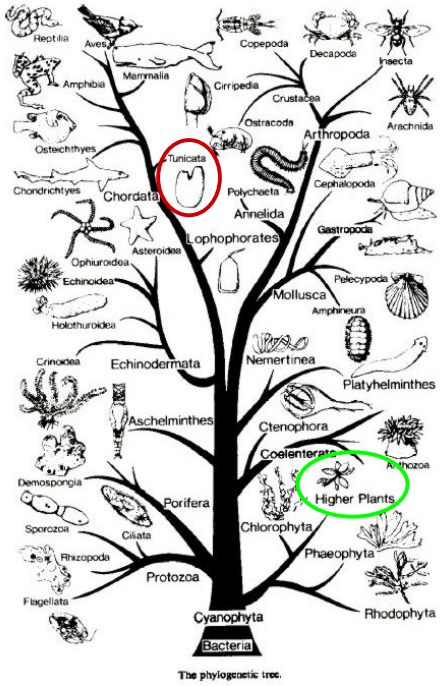
"Distilling in a Medicinal Garden" (1512)

Catharantus roseus

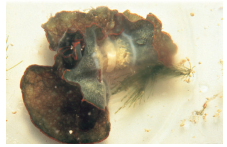








The phylogenetic tree.



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.

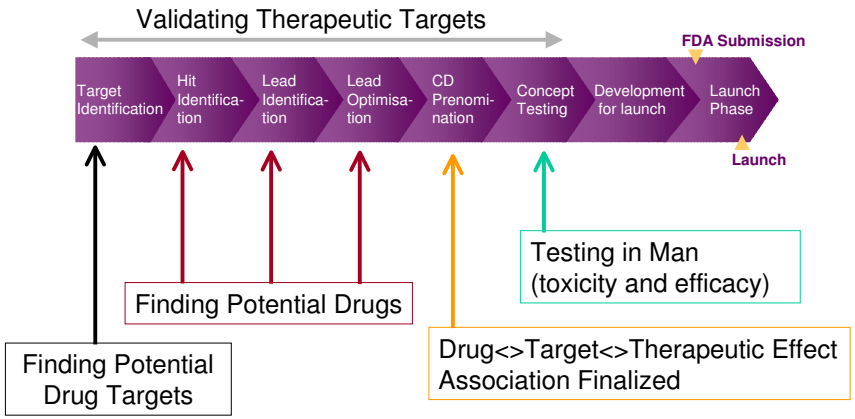
Hit: a molecule with confirmed concentration-dependent activity in a screen, and known chemical structure. The output of most screens.

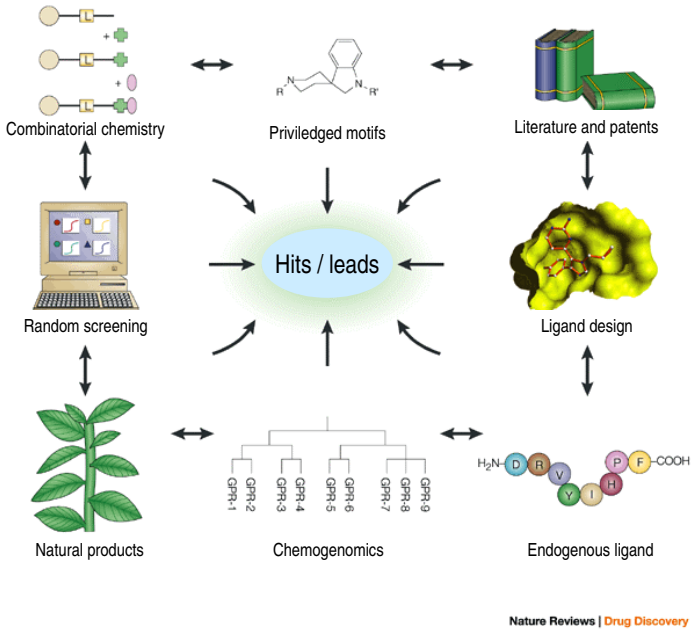
Progressible hit: a representative of a compound series with activity via acceptable mechanism of action and some limited structure-activity relationship information.

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.

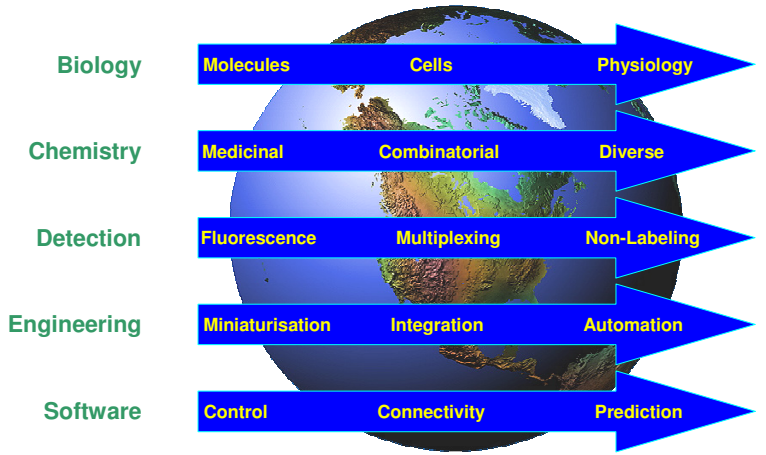
Pharmacophore: minimal structure with essential features for activity

The Drug Discovery Pipeline





The technology drivers of change

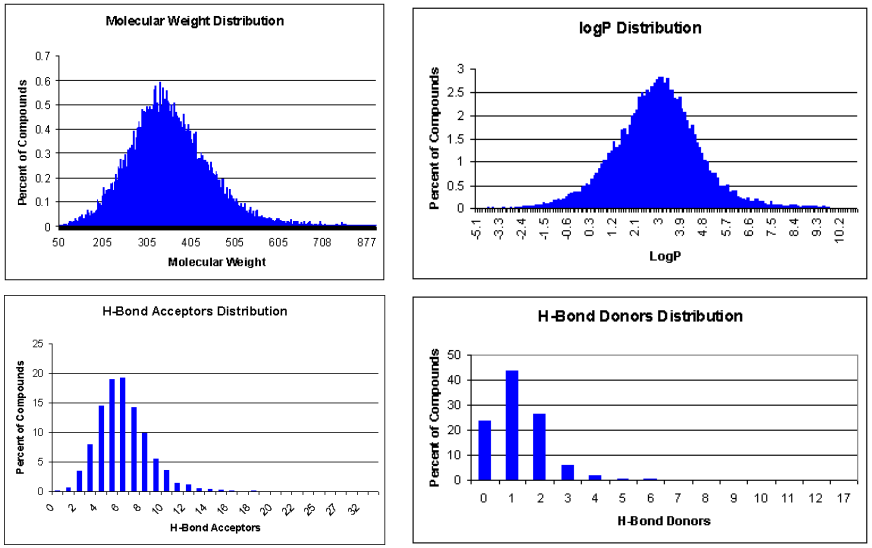


We must be able to understand:

- 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 developmental settings. *Advanced Drug Deliv. Rev.* **1997**, *23*, 3-29.

LIPINSKI's "rule of five"

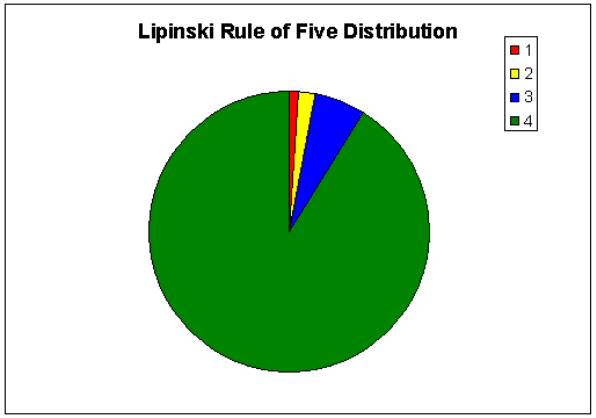
H-bond donors <5

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.



1 - compounds which satisfy 1 requirement - 1% of all compounds

2 - compounds which satisfy 2 requirements - 2% of all compounds

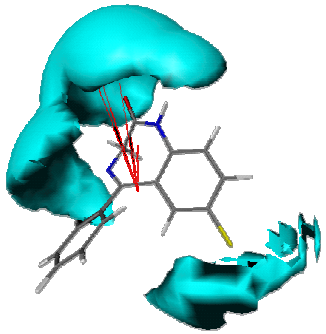
3 - compounds which satisfy 3 requirements - 6% of all compounds

4 - compounds which satisfy 4 requirements - 91% of all compounds

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

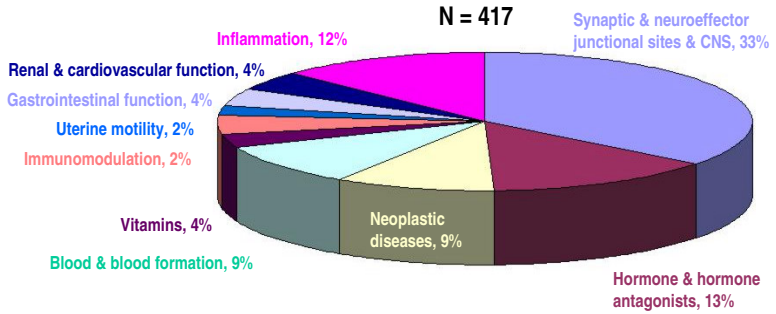
Molecular Discovery Ltd

- The basic idea of VolSurf 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.
- VolSurf descriptors are specifically designed for the optimization of pharmacokinetic properties.

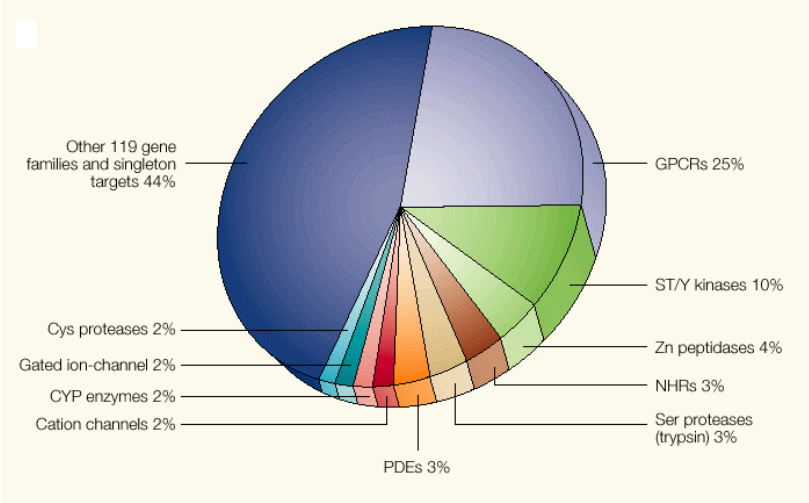


Silvio Mecucci, Gianluca Sforna, Manuel Pastor & Gabriele Cruciani

DRUG THERAPY TARGETS

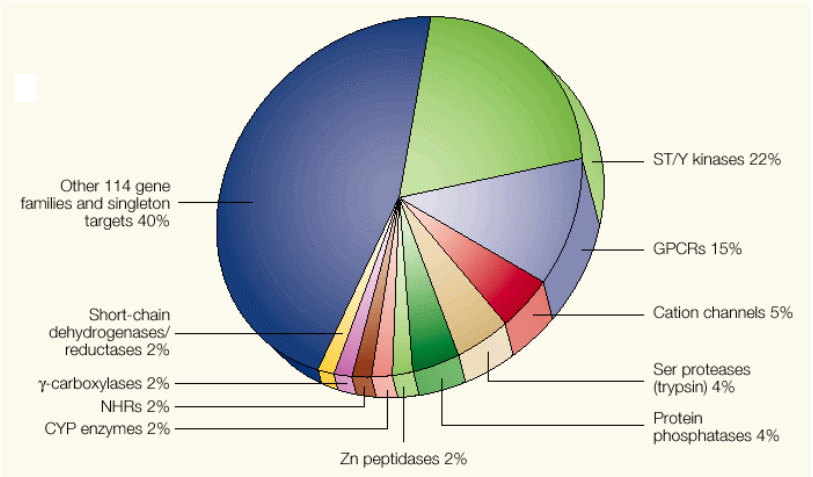


Goodman and Gilman. *The Pharmacological Basis of Therapeutics*, Edn. 9 (1996)



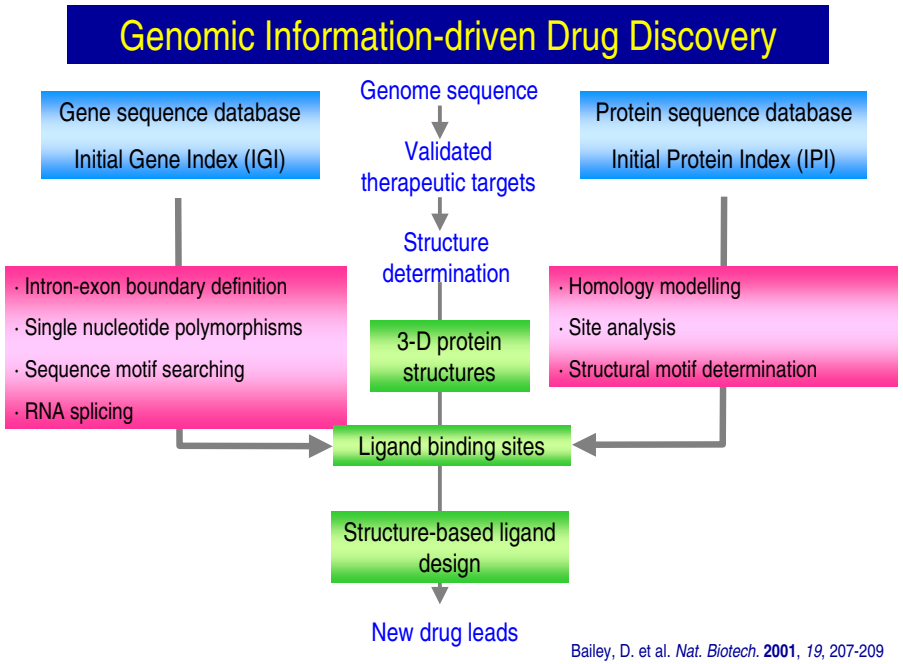
Gene-family distribution of the molecular targets of current rule-of-five-compliant experimental and marketed drugs

(~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)



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

SUCCESS OF GENETIC SCREENS IN ELUCIDATING THE PRINCIPLES OF BIOLOGICAL PROCESSES SUCH AS:

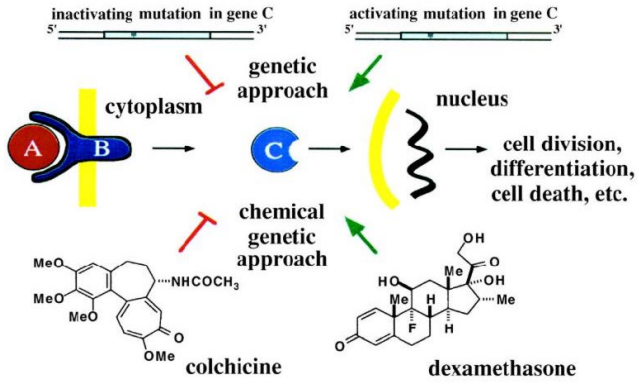
- CELL DIVISION in the budding yeast *Saccharomyces cerevisiae*

- PROGRAMMED CELL DEATH AND ORGAN DEVELOPMENT in the nematode *Caenorhabditis elegans*

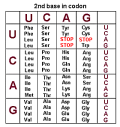
- EARLY EMBRYONAL DEVELOPMENT in the fruitfly *Drosophila melanogaster*

CHEMICAL GENETICS

Natural products and natural product-like compounds are used to understand and control protein function in signal transduction pathways, cells or whole organisms



FORWARD-GENETIC APPROACHES



1. The genome of the chosen organism is mutagenized in a random fashion

↓

2. The desired phenotype is detected within the large collection of generated mutants

↓

3. The mutated gene that underlies the phenotype is identified

1. The organism or cell type of interest is treated with a diverse collection of small molecules

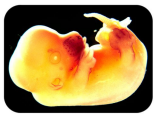
↓

2. The desired phenotype is detected within the large collection of compound-treated organisms or cells

↓

3. The cellular molecule targeted by the bioactive compound is identified

REVERSE-GENETIC APPROACHES



3. The function of the protein is elucidated

↑

2. The phenotype is analyzed

↑

1. A site-specific/'knock-out' mutation is introduced that affects the activity/expression of the chosen protein

3. The function of the protein is elucidated

↑

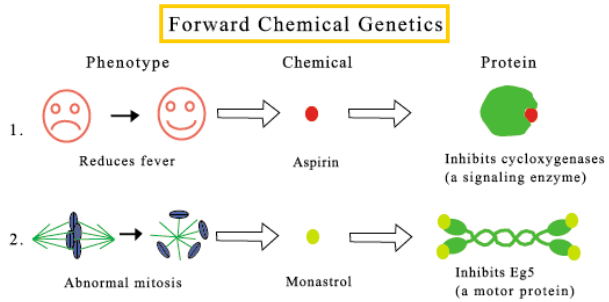
2. The phenotype is analyzed

↑

1. Search for bioactive ligands out of a large pool of small molecules

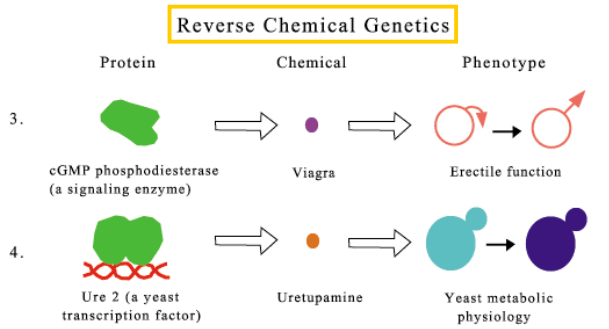
Small molecules are found that cause specific phenotypes in cells and organisms.

The protein target of the chemicals is then determined.



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: 6.0210^{23} mol^{-1}

Googol: 10^{100}

Googolplex: 10^{googol}

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

COMBINATORIAL ORGANIC CHEMISTRY

BROAD SCREENING

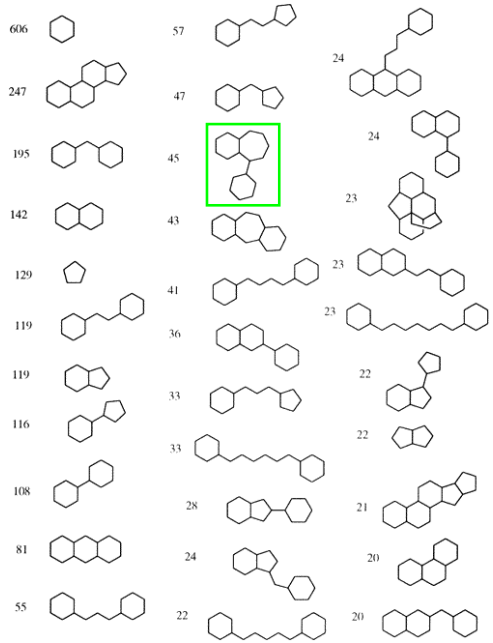
- Huge size library
- Broadest structural diversity
- No special initial structure goal
- Any building blocks
- Undefined order of reaction
- Flexible synthetic strategy
- Site of tether not crucial
- Ligand possibly uncouplable
- Single selection evolution

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

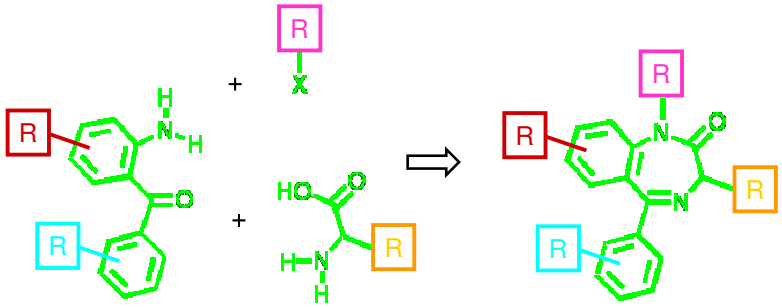
(numbers indicate frequency of occurrence in a total of 2548 drugs)



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)

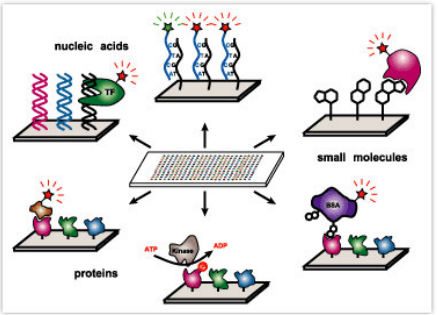


ASSAY PROCEDURES

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

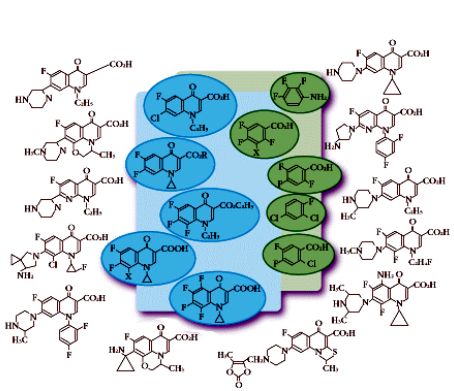
Assay formats:

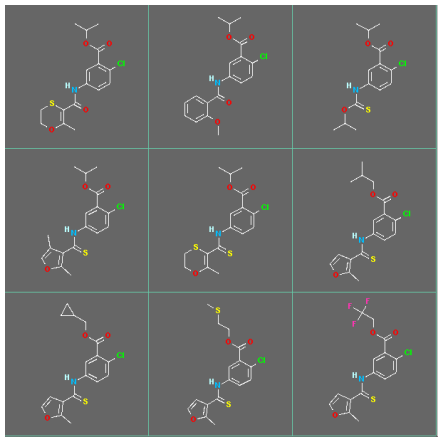
- Affinity purification with an immobilized target receptor
- A labeled soluble receptor binds to tethered ligands
- Soluble compounds are tested for activity: competition binding, enzyme inhibition, or cell-based bioassays.



Implementation of HTS

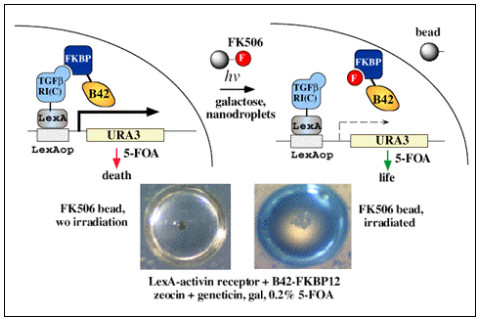
1) suitable libraries of compounds: in-house collections ($5x10^5$ - 10^6), specialist companies, combichem….

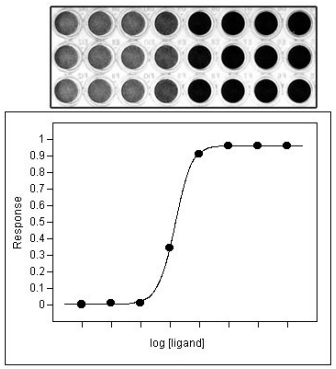


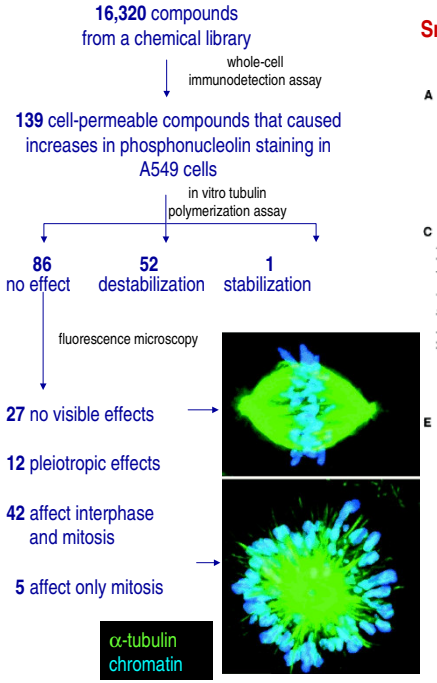


Implementation of HTS

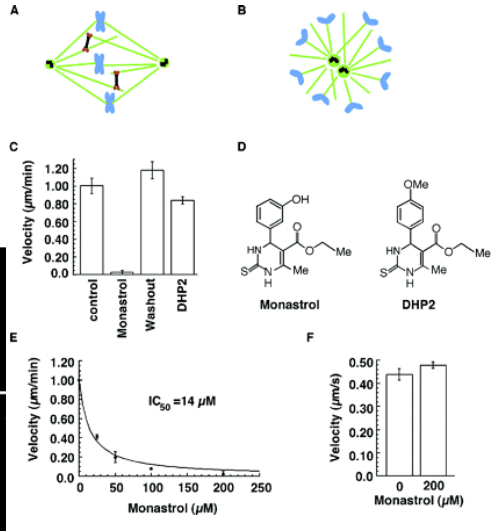
2) assay method configured for automation: radioligand binding assays, cell-based fluorescence and radiotracer assays, melanophore assays, reporter gene assays, cell viability assays, cell proliferation assays…







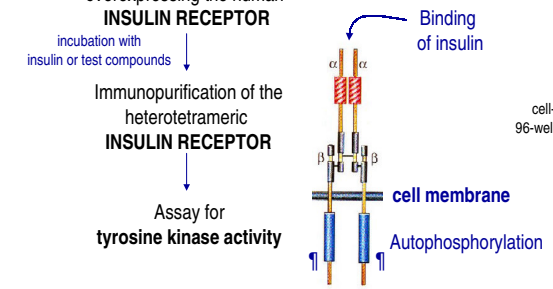
Small Molecule Inhibitor of Mitotic Spindle Bipolarity Identified in a Phenotype-Based Screen

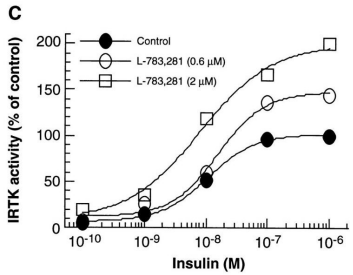


Eg5-driven microtubule motility

conventional kinesin-driven microtubule motility

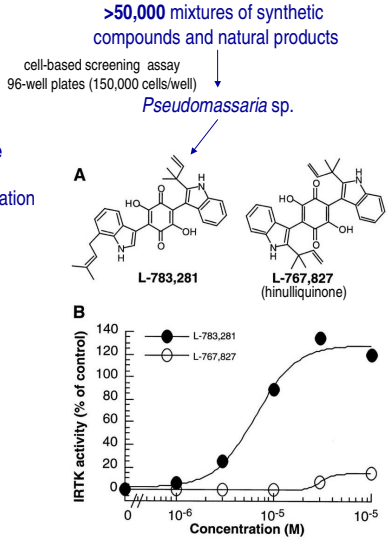
Mayer et al. *Science* **1999**, *286*, 971-974





Enhancement of insulin-stimulated tyrosine kinase activation

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



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

Implementation of HTS

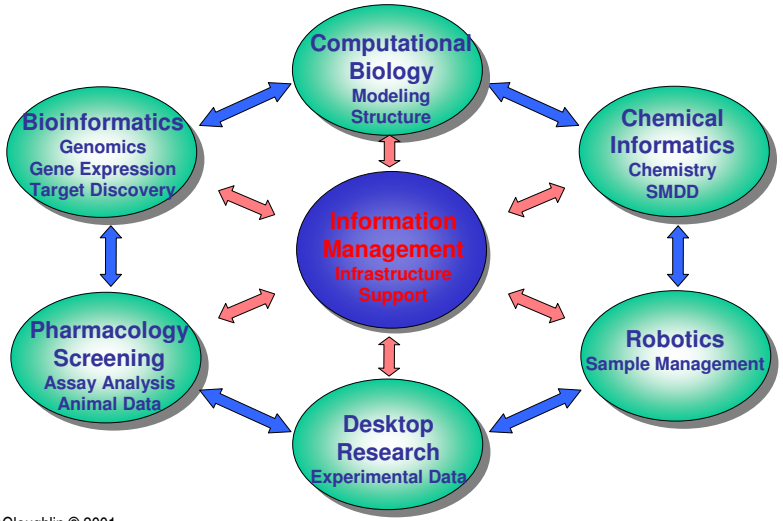
3) robotics workstation: full automation, 24 h continuous operation, more efficient and economical.

multi-well format

microarray format

Implementation of HTS

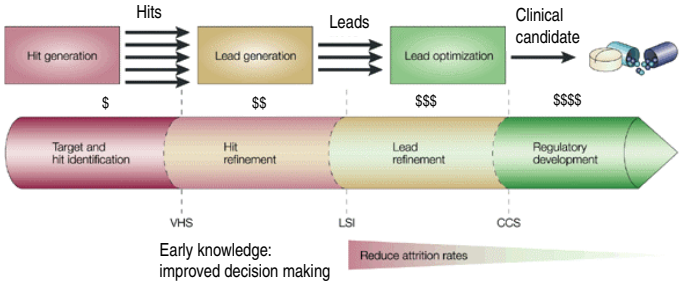
4) computerised data handling system: accurate and reproducible.



‘Needle in a Haystack’ Syndrome

- Estimated 10^{200} compounds could be made[1]
- 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`

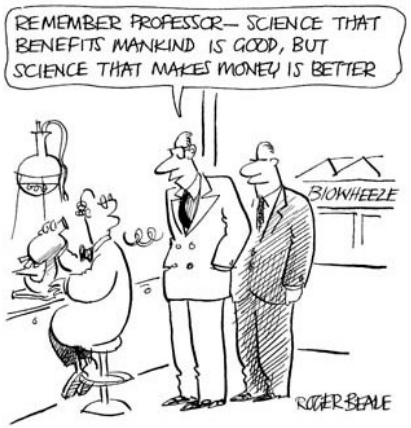


Nature Reviews | Drug Discovery

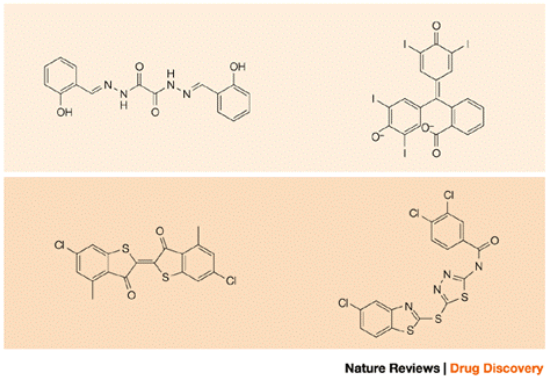
lots of hurdles

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

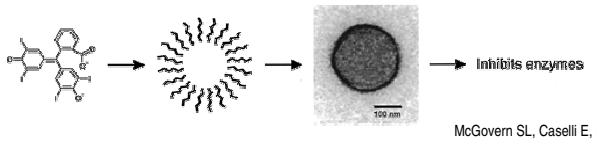
talking about $$$$….



Frequent hitters: “promiscuous binders”



Nature Reviews | Drug Discovery



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

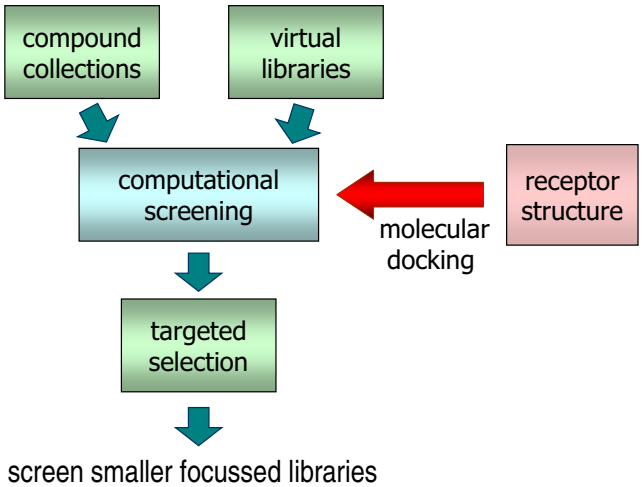
Also *insilico* (?)

Successful ligand-receptor pairings

1. Orphan receptor strategy	2. Reverse pharmacology
Five novel peptides/peptide families	**Pairing of 6 peptides to orphan R**
➳ nociceptin/orphanin FQ (N/OFQ)	➳ melanin concentrating hormone
➳ hypocretins/orexins (Hcrts/Oxs)	➳ urotensin II
➳ prolactin releasing peptide (PrRP)	➳ neuromedin U
➳ apelin	➳ motilin
➳ ghrelin	➳ neuropeptide FF
	➳ neuropeptide AF
Pairing of 3 known peptides to corresponding orphan receptor	**Pairing of 4 lipids to orphan R**
➳ melanin concentrating hormone	➳ sphingosine 1-phosphate
➳ urotensin II	➳ lysophosphatic acid
➳ neuromedin U	➳ leukotriene B_4, C_4 and D_4
	➳ sphingosylphosphorylcholine
	Pairing of non-lipid/non-peptide
	➳ histamine ➠ histamine H_3 receptor
	➳ UDP-glucose ➠ KIAA0001 receptor

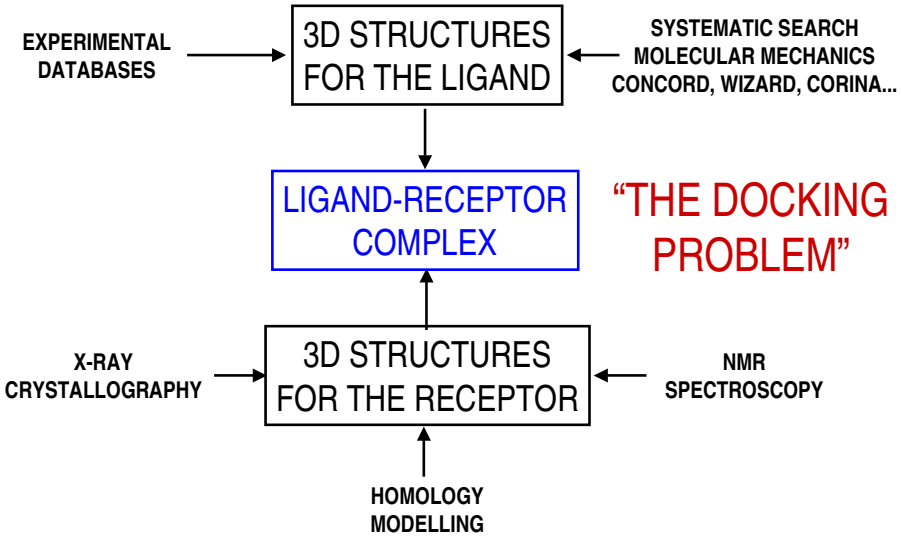
- **HTS methods are now routine procedure**
  - **Pros**
    - » allow detection of possible lead molecules and structural classes
    - » rapidly generate a provisional SAR relationship
    - » effectively utilise in-house historical compound archive
  - **Cons**
    - » HTS laboratories are expensive 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



"THE DOCKING PROBLEM"

1.- SITE/LIGAND REPRESENTATION
(treatment of H atoms?)

2.- JUXTAPOSITION OF THE LIGAND AND SITE FRAMES OF REFERENCE

3.- 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.

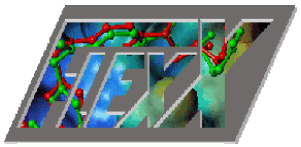
Virtual ("in silico") screening

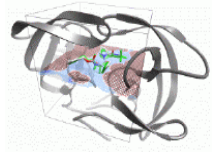
Docking/scoring programs

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)

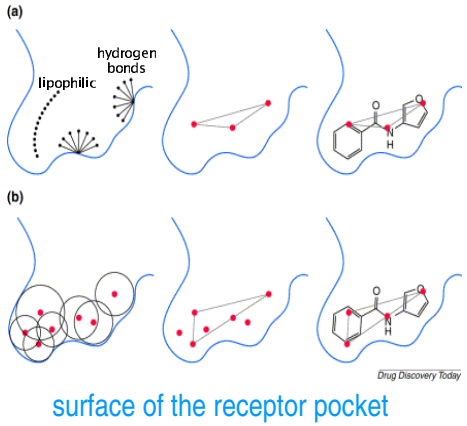
Simulated annealing (AutoDock 2.4)

Monte Carlo simulations (MCDOCK)

Genetic algorithm (AutoDock 3.0, GOLD, GAMBLER)

FlexX algorithm

DOCK algorithm



surface of the receptor pocket

FlexX matches triangles of interaction sites onto complementary ligand atoms.

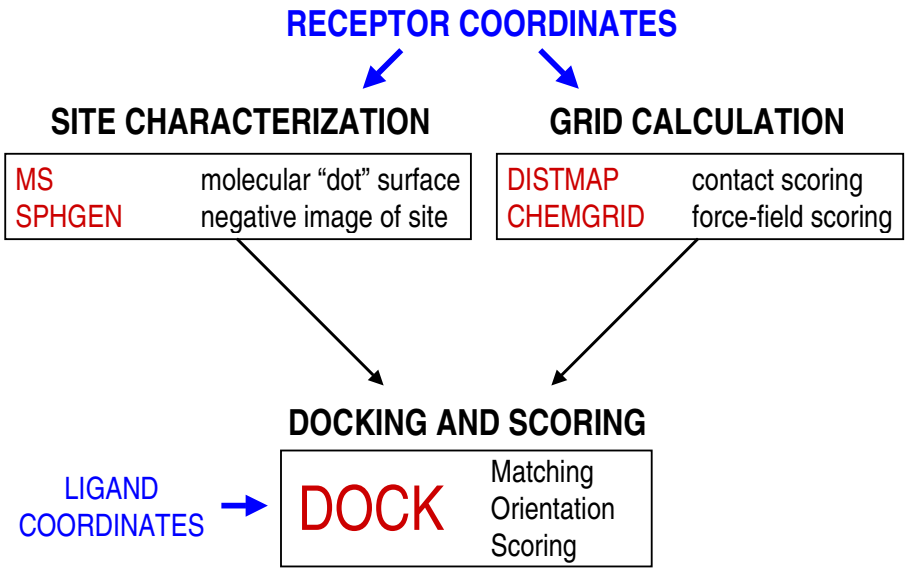
DOCK fills the binding site with spheres, and sphere centers are then matched to the ligand atoms to determine plausible ligand–receptor complexes.

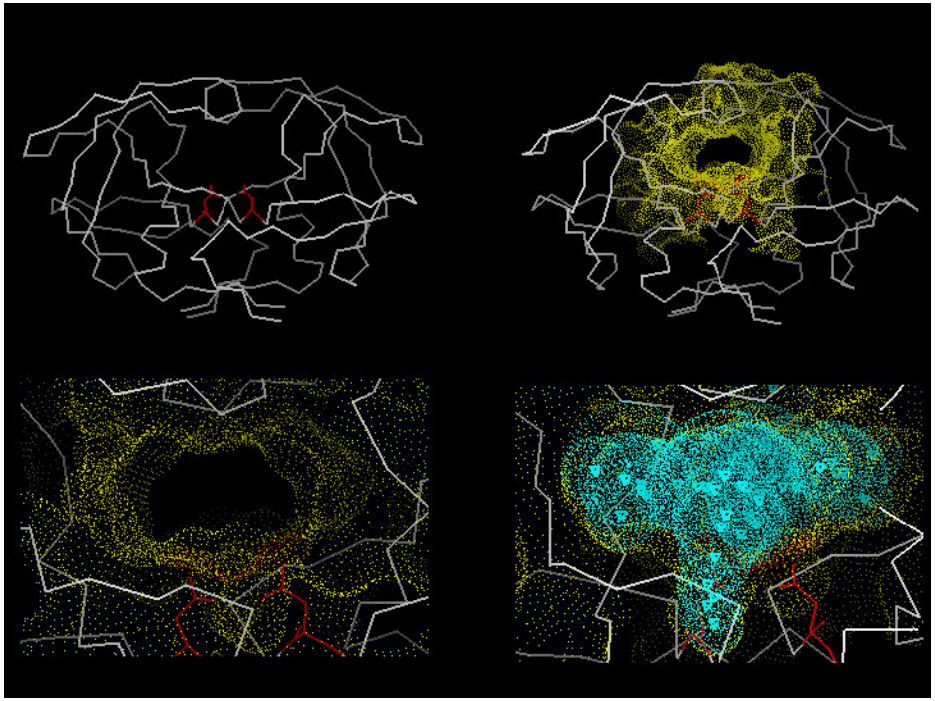
PROGRAM **DOCK**

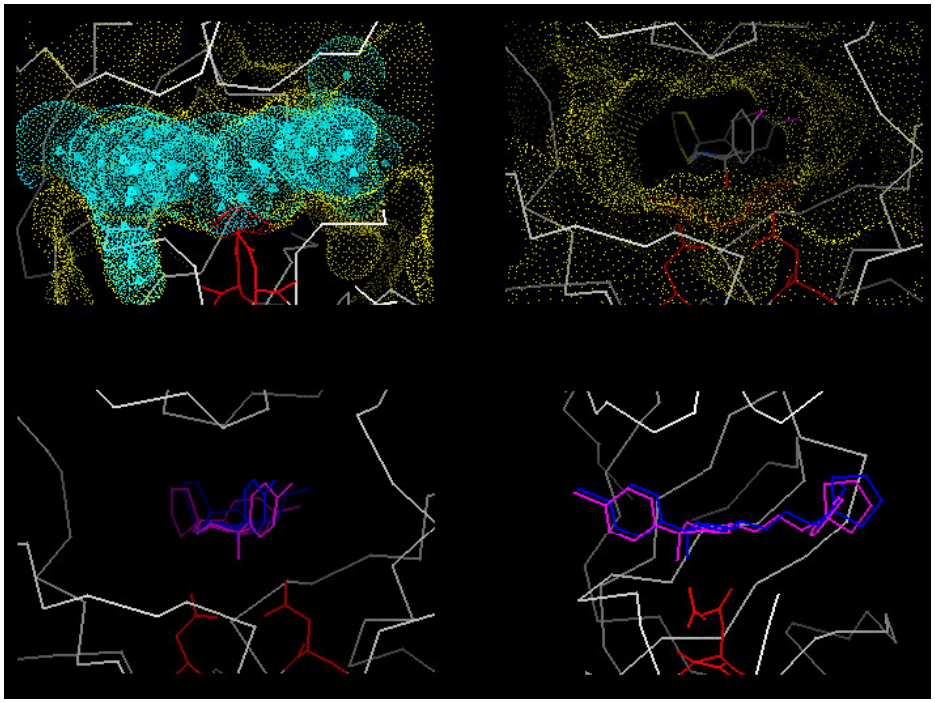
"A Geometric Approach to Macromolecule-Ligand Interactions"
I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, T. E. Ferrin
J. Mol. Biol. 161, 269-288 (1982)

"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)







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-hydroxymethy-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl]-5-methyl-*1*H-pyrimidin-2,4-dione	9.62	13.30	2.33
(North)-methanocarbathymidine	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

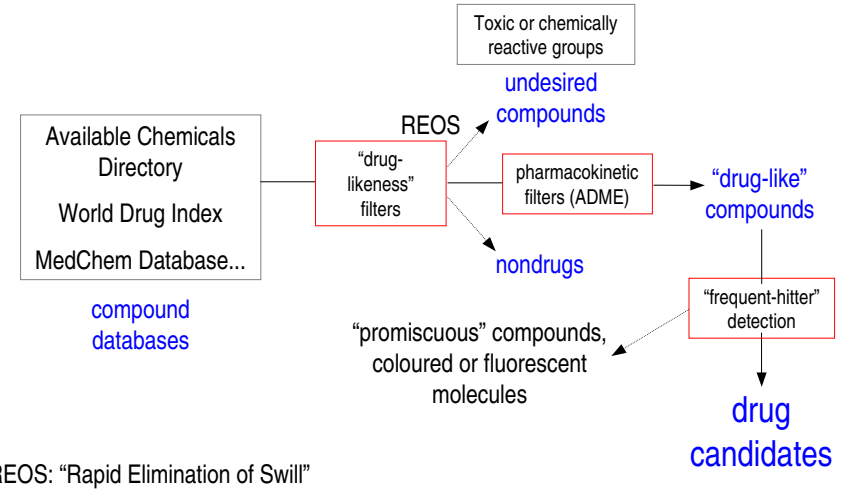
Knowledge-based: statistical analysis of 3D complex structures to derive a sum of *potentials of mean force* between receptor and ligand atoms

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

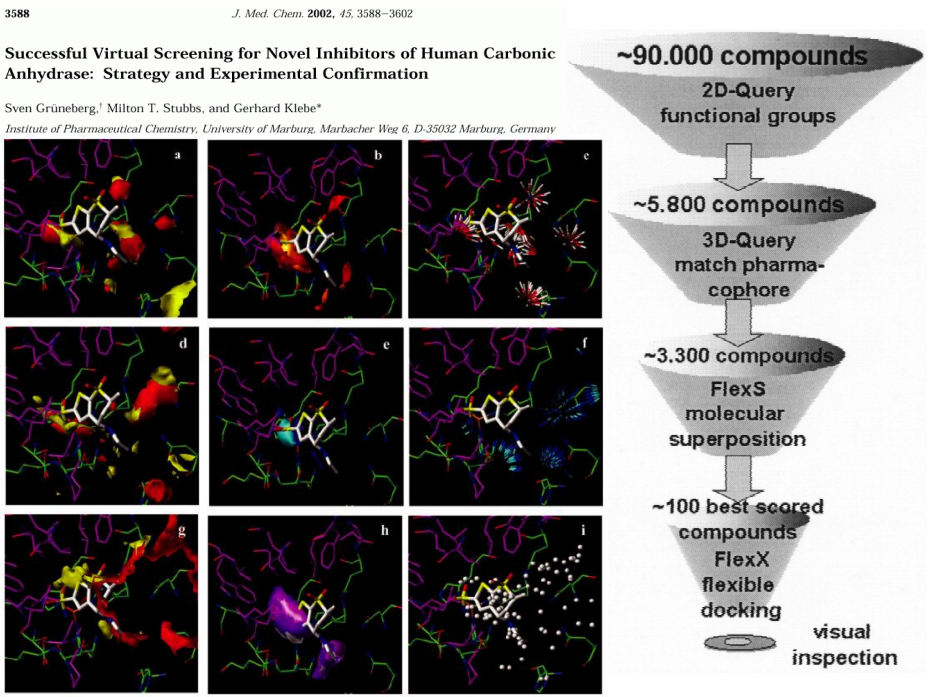
Empirical: the binding free energy is broken down into a number of different *weighted* contributions (supposed to be additive: number of hydrogen bonds, ionic interactions, apolar contacts, entropy penalties...)

In silico VIRTUAL SCREENING and FOCUSED LIBRARY DESIGN

Near-perfect structures in an imperfect world



REOS: "Rapid Elimination of Swill"



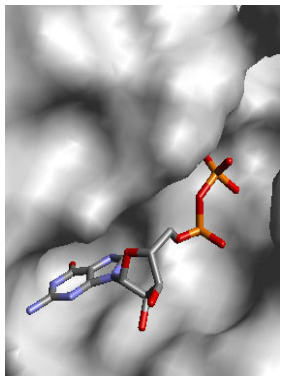
Prof.W.Graham Richards

Screensaver Lifesaver

searching for anti-cancer drugs by distributed computational chemistry

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

NATIONAL FOUNDATION FOR CANCER RESEARCH
research for a cure

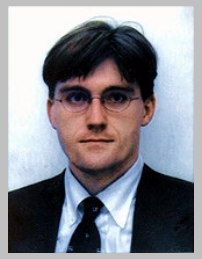


Superoxide dismutase	Vascular Endothelial Growth Factor
RAS proteins	Insulin Tyrosine Kinase
Cyclooxygenase (COX-2)	c-ABL Tyrosine Kinase
Fibroblast Growth Factor Receptor	CDK-2
RAF	Farnesyltransferase
Protein-Tyrosine-Phosphatase 1B	VEGFr1

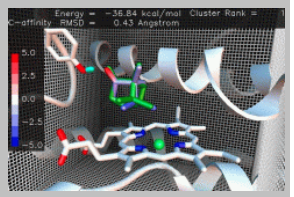
fightAIDS@home the Olson laboratory
computing toward a cure
powered by ENTROPIA

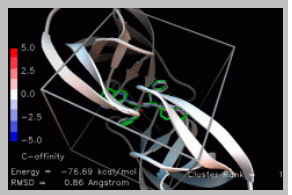
THE SCRIPPS RESEARCH INSTITUTE

http://www.FightAidsathome.org



Dr.GarrettMorris





Applications of Ligand-Protein Docking in Drug Design

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

Compound Database
Compound 1
...
Compound n

↓ Protein

↓ Successfully docked compounds as putative ligands

Science (1992) *257*: 1078

New method
Given a ligand, find potential potein targets from a protein database

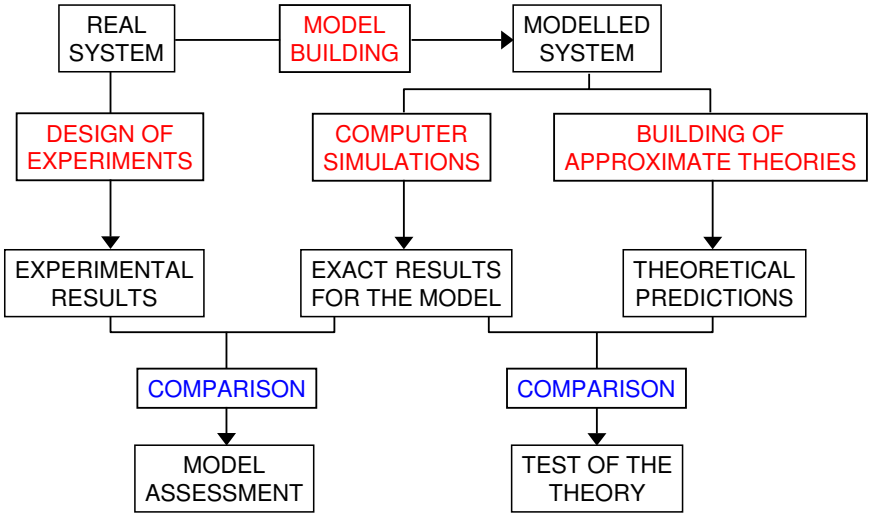
Protein Database
Protein 1
...
Protein n

↓ Ligand

↓ Successfully docked proteins as putative targets

Proteins (2001) *43*: 217

CONNECTION BETWEEN EXPERIMENT, THEORY AND COMPUTER SIMULATION



FORCES THAT DETERMINE LIGAND-RECEPTOR INTERACTIONS

Favourable forces

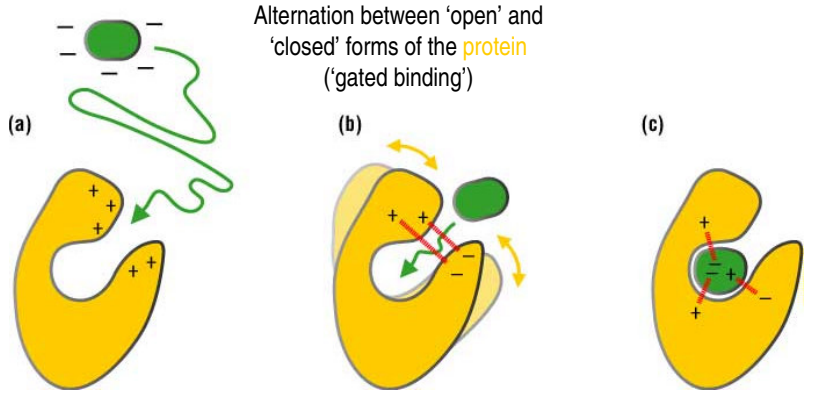
- electrostatic interactions
- hydrogen bonds
- hydrophobic effect
- van der Waals interactions
- desolvation of receptor and ligand

Unfavourable forces

- loss of translational and rotational entropy
- loss of internal rotations in ligand (*entropic*)
- loss of solvation energy of receptor and ligand (*enthalpic*)
- conformational changes in receptor

P. G. Strange *TiPS* 17, 238 (1996)

Electrostatic interactions in ligand binding



Alternation between 'open' and 'closed' forms of the protein ('gated binding')

Steering the ligand into its binding site on the protein.

Salt links and hydrogen bonds between ligand and protein contribute to:
- affinity and specificity
- orientation of the ligand in the binding site
- the structure of the complex formed.

1st SAR study:

ON THE

CONNECTION

BETWEEN

CHEMICAL CONSTITUTION

AND

PHYSIOLOGICAL ACTION.

PART I.

ON THE PHYSIOLOGICAL ACTION OF THE SALTS OF THE AMMONIUM BASES, DERIVED FROM STRYCHNIA, BRUCIA, THEBAIA, CODEIA, MORPHIA, AND NICOTIA.

BY

DR. A. CRUM BROWN AND DR. THOMAS R. FRASER

FROM THE
TRANSACTIONS OF THE ROYAL SOCIETY OF EDINBURGH, Vol. XXV.

EDINBURGH:
PRINTED FOR THE SOCIETY BY NEILL AND COMPANY.

MDCCCLXVIII.

Almost 100 years later:

"ρ–σ–π Analysis, A Method for the Correlation of Biological Activity and Chemical Structure"

C. Hansch & T. Fujita
J. Am. Chem. Soc. ***86***, 1616 (1964)

"A Mathematical Contribution to Structure-Activity Studies"

S. M. Free, Jr. & J. W. Wilson
J. Med. Chem. ***7***, 395 (1964)

QSAR: Quantitative Structure-Activity Relationships

Physiological activity $\Phi = f(C)$ (Brown & Fraser, 1868)

$\Delta\Phi = f(\Delta C)$

Biological activity = $f(a_i X_i, m)$ Linear Free Energy Relationships

$B.a. = \mu + \Sigma a_{ij} X_{ij}$ de novo model (X_{ij} = 1, 0)
μ = overall mean of b.a. values (Free & Wilson, 1964)

μ = b.a. of unsubstituted parent molecule (Fujita & Ban, 1971)

Biological activity = $\log(1/C) = k_1 (X_H) + k_2 (X_E) + k_3 (X_S) + \varepsilon$ parametric model
(Hansch & Fujita, 1964)

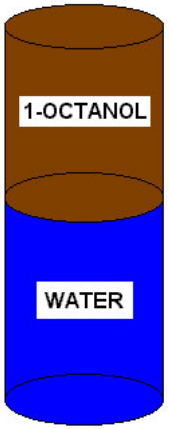
MOLECULAR PARAMETERS USED IN QSAR:

electronic: σ constants (ΔpKa values), NMR chemical shifts, atomic charges, MO indices, frontier orbital energies, superdelocalizability indices, electrostatic potential...

hydrophobic: π values (Δlog P values), HPLC log k' ...

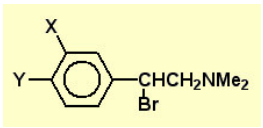
molecular shape/geometry: Taft' s parameters, Kier' s molecular connectivity indices, Verloop' s sterimol parameters...

Hydrophobicity



"shake flask" experiment

- **Measured as Water / Octanol Partition Coefficient (P).**
- **Log P_A = Log $\left[\frac{[A]_{1\text{-octanol}}}{[A]_{water}}\right]$**
- **log P > 0 : lipid phase**
  log P < 0 : water phase



$$\log(1/ED_{50}) = -0.301[m\text{–F}] + 0.27[m\text{–Cl}] + 0.434[m\text{–Br}] + 0.579[m\text{–I}] + 0.454[m\text{–Me}] + 0.340[p\text{–F}] + 0.768[p\text{–Cl}] + 1.020[p\text{–Br}] + 1.429[p\text{–I}] + 1.256[p\text{–Me}] + 7.821$$

$$n = 22,\ r^2 = 0.94,\ s = 0.194,\ F = 17.0$$

A negative coefficient indicates that the presence of that group is unfavourable to activity.

A positive coefficient indicates that the presence of that group is favourable to activity.

"A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based Upon Molecular Shape Analysis"

A. J. Hopfinger
J. Am. Chem. Soc. 120, 7196 (1980)

"Molecular Graphics and QSAR in the Study of Enzyme-Ligand Interactions. On the Definition of Bioreceptors"

C. Hansch & T. E. Klein
Acc. Chem. Res. 19, 392 (1986)

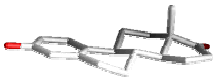
"**Comparative Molecular Field Analysis (CoMFA).** 1. Effect of Shape on Binding of Steroids to Carrier Proteins"

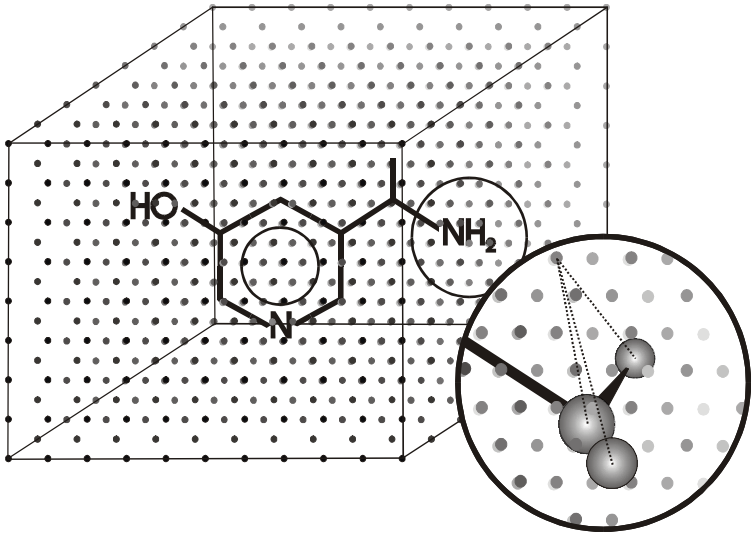
R. D. Cramer, III, D. E. Patterson & J. D. Bunce
J. Am. Chem. Soc. 110, 5959 (1988)

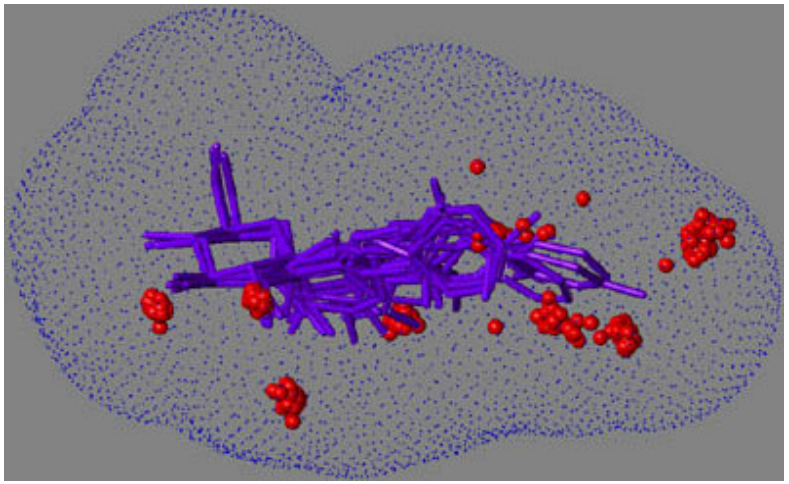
"Prediction of Drug Binding Affinities by **Comparative Binding Energy Analysis**"

A. R. Ortiz, M. T. Pisabarro, F. Gago & R. Wade
J. Med. Chem. 38, 2681 (1995)

Introducing the 3rd dimension: 3D QSAR (CoMFA)



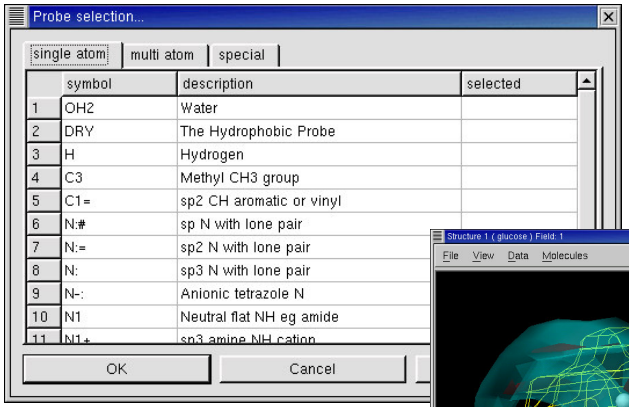




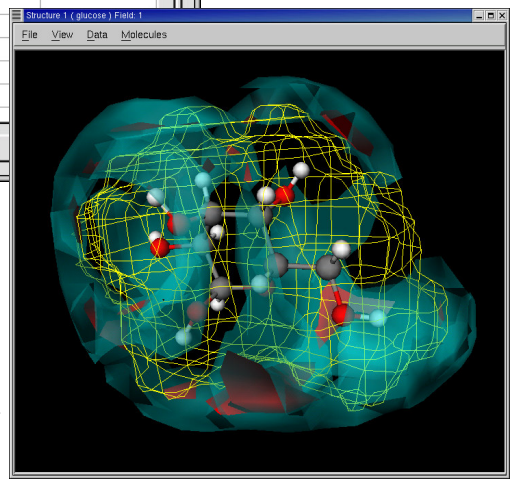
Manuel Pastor, Gabriele Cruciani, Kimberly Watson

"A strategy for the incorporation of water molecules present in a ligand binding site into a 3D QSAR analysis"

J.Med.Chem. 40, 4089-4102 (1997)



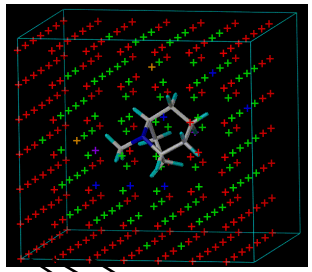
GRID v. 21

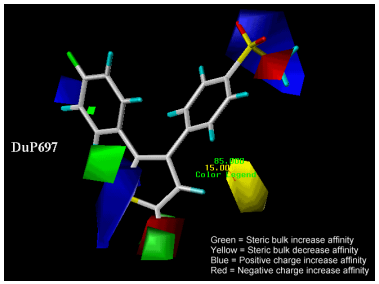


Molecular Discovery

http://www.moldiscovery.com/

CoMFA is a (3D-Q)SAR method





Contour Maps

Predictions

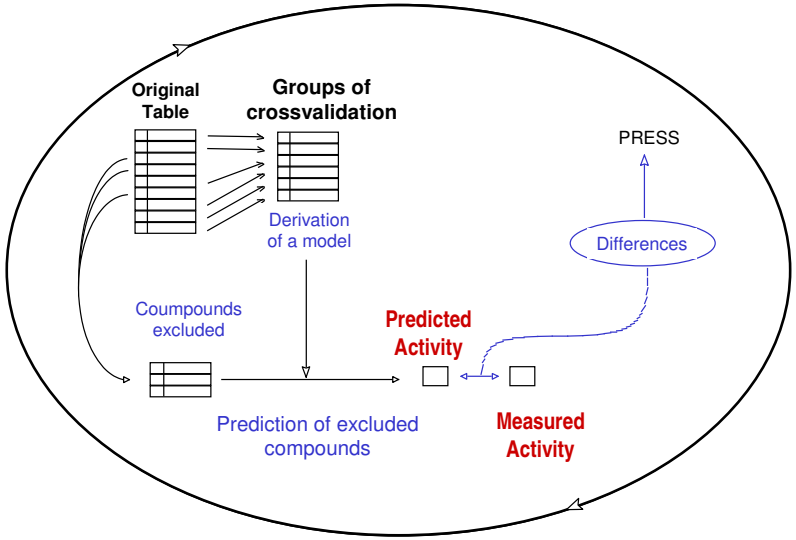
Bio

PLS

QSAR equation

QSAR Table = SYBYL MSS

Cross-validated PLS analyses

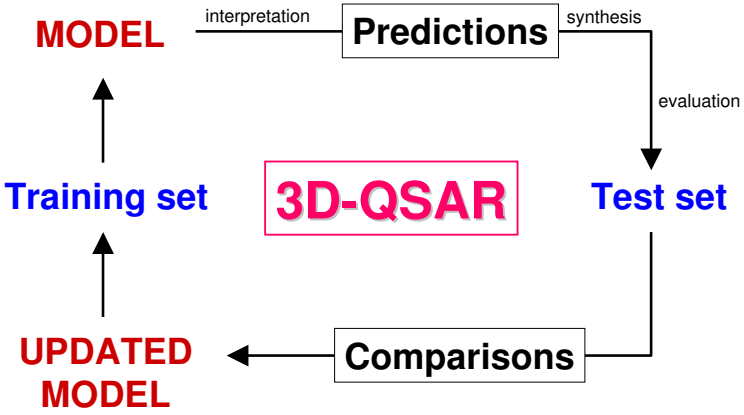


TRADITIONAL QSAR Disadvantages:	3D-QSAR Advantages:
- Congeneric series	- Mixed series
- Missing physicochemical parameter values	- No parameters must be estimated
- Lack of 3D structural information	- 3D structural information included
- Results expressed only as a numerical equation	- Results can be graphically displayed in 3D
- Collinearity of parameters must be avoided	- Energy fields can be collinear
- Inadequate description of steric effects	- Good description of steric effects
- Inadequate description of hydrogen bonding	- Good description of hydrogen bonding

K. H. Kim, in *'3D QSAR in Drug Design. Theory, Methods and Applications'* (1993)

TRADITIONAL QSAR Advantages:	3D-QSAR Disadvantages:
- Simplicity and speed	- More complicated to run
- No bioactive conformation required	- A bioactive conformation must be assumed
- No alignment needed	- Superposition rules and alignment problems
- May extrapolate into unexplored region with care	- Difficult to extrapolate into unexplored regions
- Results summarized in a simple equation	- Results not usually summarized in an equation
- Useful information is provided by the coefficients in the correlation equation	- Less useful information from the coefficients obtained in the correlation equation
- No weighting of parameters is necessary	- Many adjustable parameters involved
- Simple use of indicator variables	- Use of indicator variables is not straightforward

K. H. Kim, in ' *3D QSAR in Drug Design. Theory, Methods and Applications'* (1993)



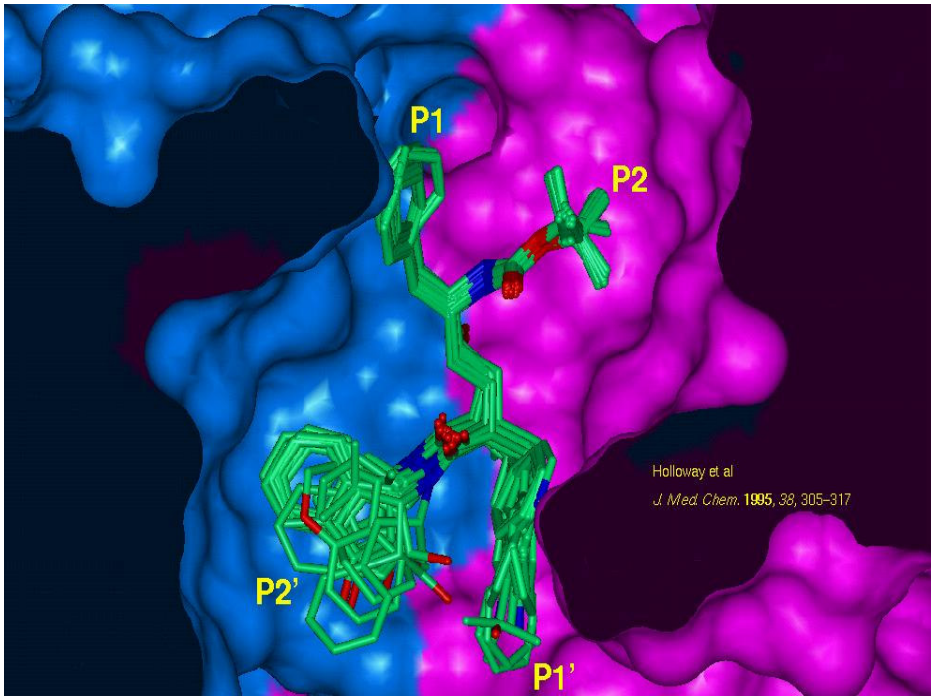
Performance

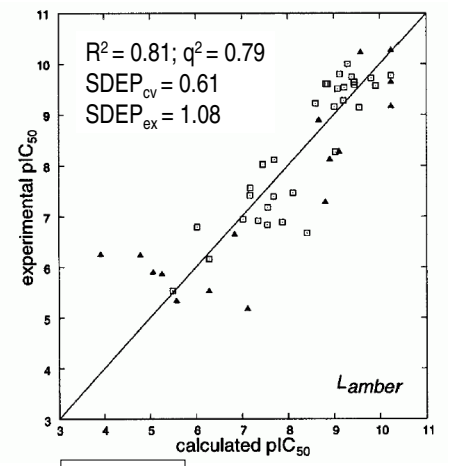
Standard Deviation of Error in Predictions:

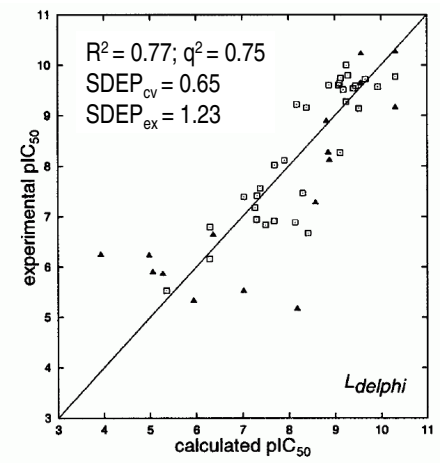
$$SDEP = \sqrt{\sum \frac{(Yexp(i) - Ypred(i))^2}{N}} = \sqrt{\frac{PRESS}{N}}$$

Correlation Coefficient in Cross-Validation:

$$Q^2 = 1 - \left[\frac{\sum (Yexp(i) - Ypred(i))^2}{\sum (Yexp(i) - \langle Yexp \rangle)^2} \right]$$





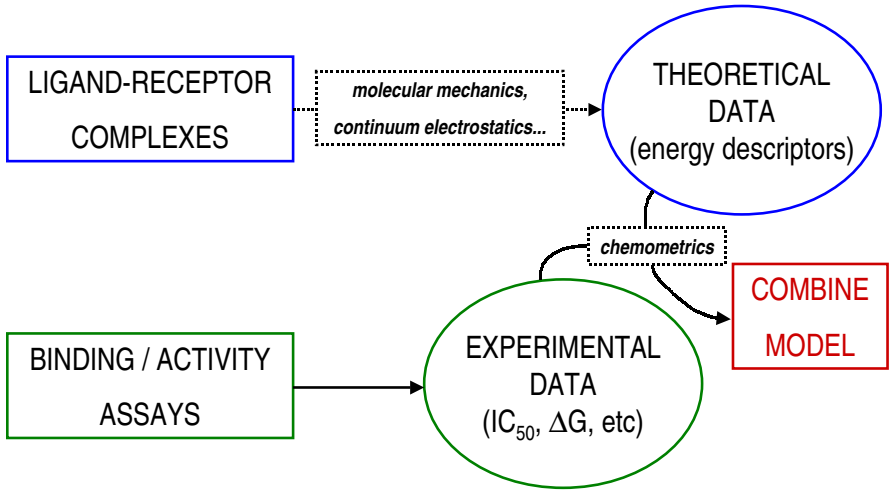


C. Pérez, M. Pastor, A. R. Ortiz & F. Gago
J. Med. Chem. ***41***, 836 (1998)

Multiple regression analysis:

$Activity = a\ (E_{inter}) + b$

Comparative Binding Energy Analysis



COMBINE Analysis

"Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis"
A. R. Ortiz, M. T. Pisabarro, F. Gago & R. Wade
J. Med. Chem. ***38***, 2681 (1995)

"Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation As a Powerful Tool in Receptor-Based Drug Design"
C. Pérez, M. Pastor, A. R. Ortiz & F. Gago
J. Med. Chem. ***41***, 836 (1998)

"Comparative Binding Energy Analysis"
Wade, R. C., Ortiz, A. R. & Gago, F.
Perspectives in Drug Discovery and Design, 9/10/11, 19-34 (1998)

MODELLING PHASE

unbound receptor (R)

n unbound ligands (L)

n (R:L) complexes

REFINEMENT STAGE

energy minimization

n refined (R:L) complexes

desolvation energy terms (?)

ENERGY CALCULATION AND PARTITIONING / MATRIX PRETREATMENT

$\Delta U = E_{LR} - (E_L + E_{LR})$ → energy decomposition

MODEL DERIVATION

Activity = $\sum_{i=1}^{n} w_i \Delta u_i^{sel} + C$ ← *Partial Least Squares (PLS)*

MODEL VALIDATION:
- cross-validation
- permutation of activity data (*scrambling*)
- random numbers

PREDICTIONS: error assessment

Energy decomposition in COMBINE analysis

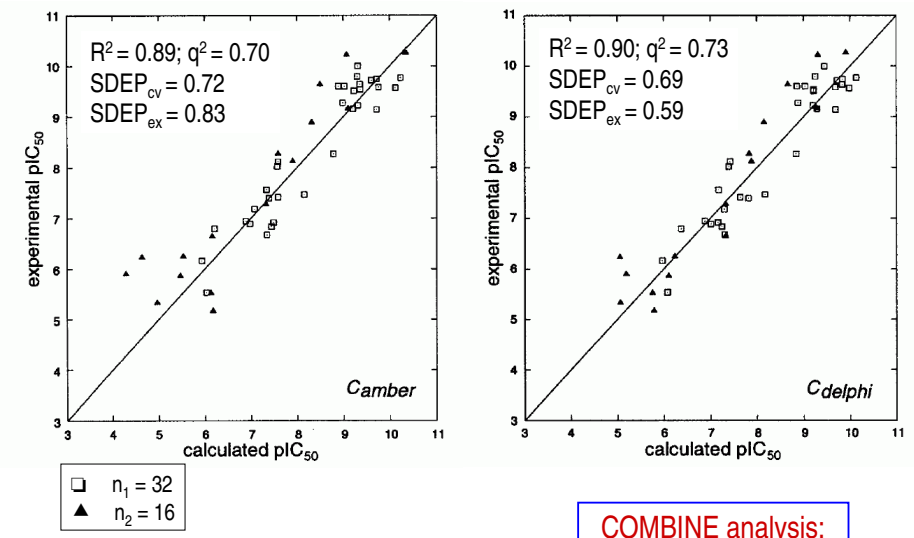
LIGAND → n_l fragments (*optional*)

RECEPTOR → n_r fragments (e.g. amino acids)

$$\Delta U = \sum_{i=1}^{n_l}\sum_{j=1}^{n_r} u_{ij}^{vdW} + \sum_{i=1}^{n_l}\sum_{j=1}^{n_r} u_{ij}^{ele} +$$

$$\sum_{i=1}^{n_l} \Delta u_i^{B,L} + \sum_{i=1}^{n_l} \Delta u_i^{A,L} + \sum_{i=1}^{n_l} \Delta u_i^{T,L} + \sum_{i<i'} \Delta u_{ii'}^{NB,L} +$$

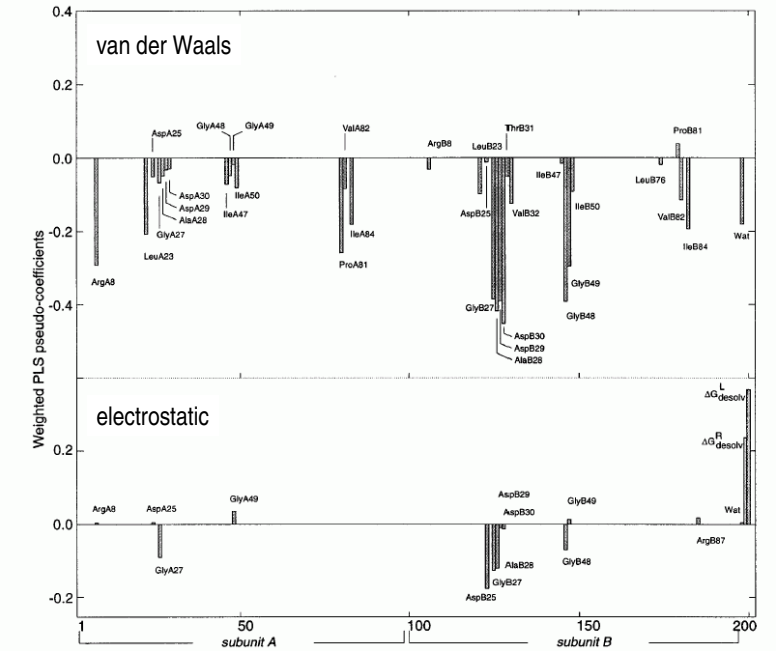
$$\sum_{j=1}^{n_r} \Delta u_j^{B,R} + \sum_{j=1}^{n_r} \Delta u_j^{A,R} + \sum_{j=1}^{n_r} \Delta u_j^{T,R} + \sum_{j<j'} \Delta u_{jj'}^{NB,R}$$

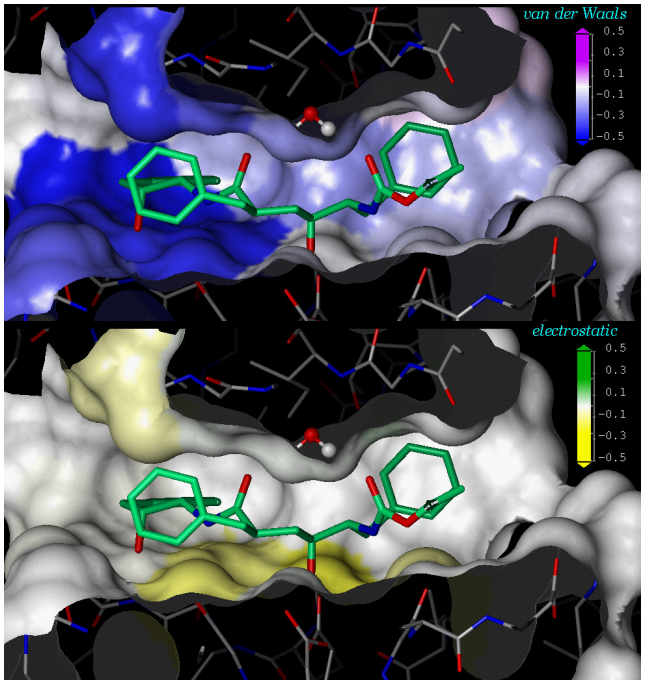


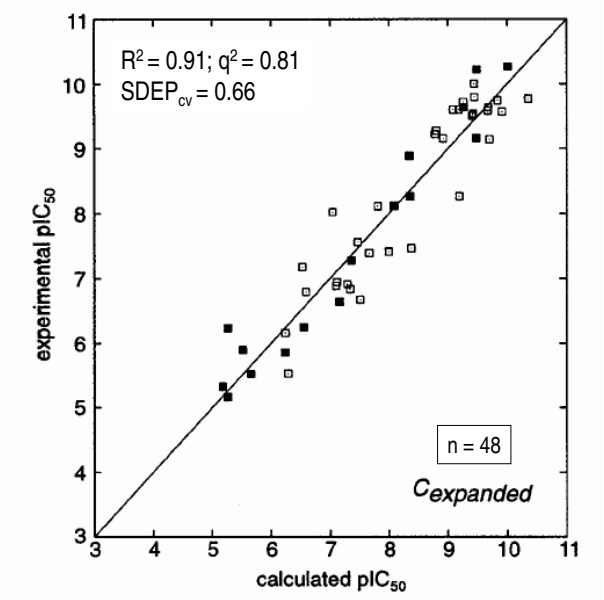
C. Pérez, M. Pastor, A. R. Ortiz & F. Gago
J. Med. Chem. ***41***, 836 (1998)

COMBINE analysis:

$$\sum_{i=1}^{n} w_i \Delta u_i^{sel} + C$$

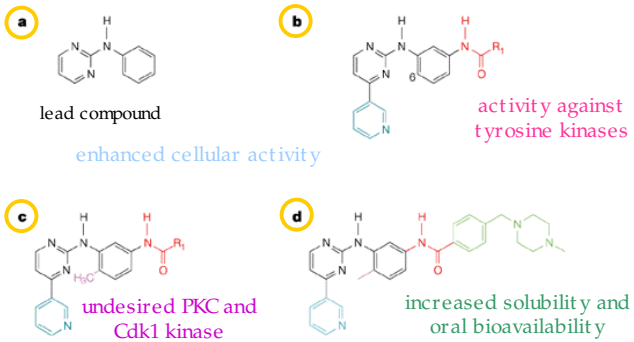




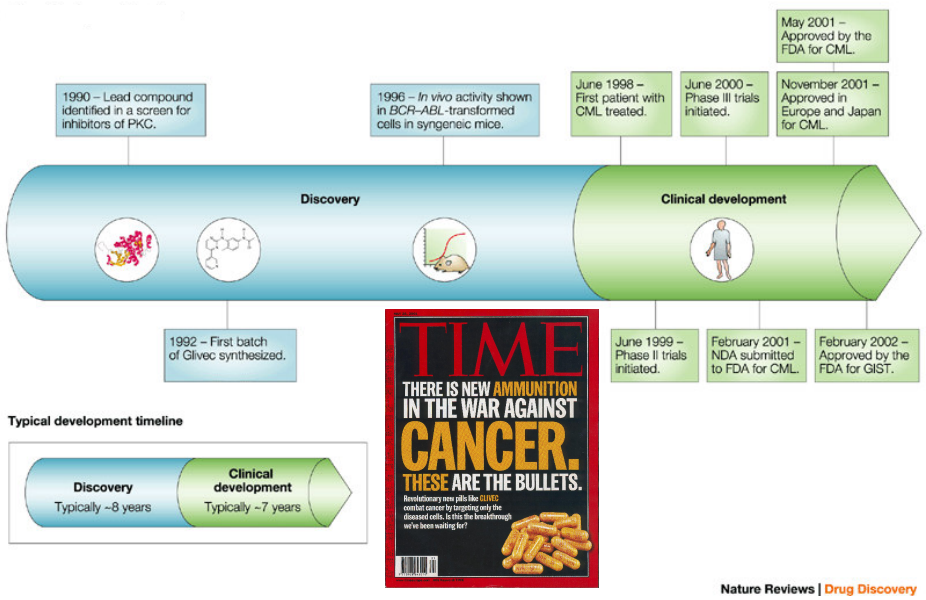


Glivec (STI571/ Imatinib):

a rationally developed, targeted anticancer drug

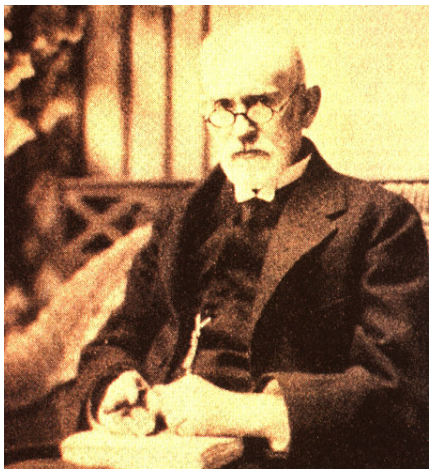


Glivec development timeline



CONCLUSIONS

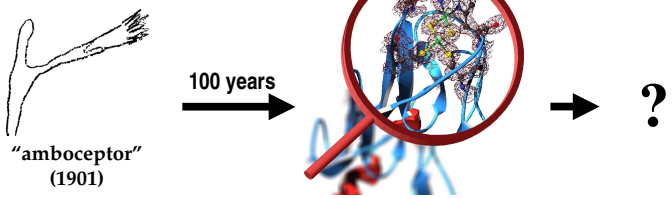
- High-throughput technologies alone are not likely to improve the productivity of drug discovery research greatly.
- The integration of diverse discovery technologies is expected to have an increasingly important role.
- High fall-out rates of clinical candidates present a major problem for the pharmaceutical industry at present.
- There is a clear trend in the field to take 'downstream' compound characteristics beyond potency into account as early as possible during the discovery process (especially ADME parameters).
- Excellent opportunities for chemoinformatics to interface with experimental discovery programmes: complementarity of VS and HTS efforts in the early phases of drug discovery research.
- VS has a natural tendency to aim at 'rational' reduction in the number and magnitude of experiments.

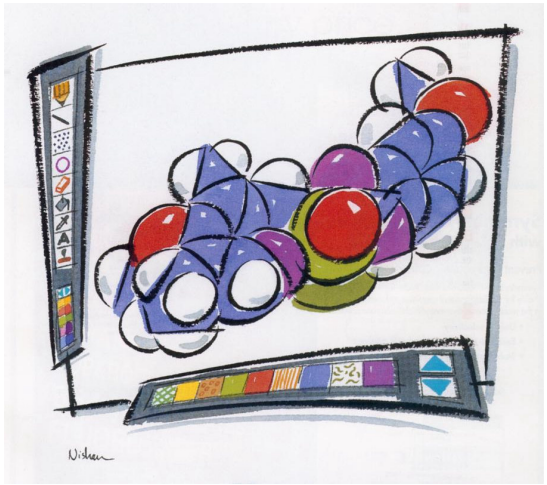


"Considering the enormous number of chemical combinations which are taken into consideration in a struggle with diseases, it will always be a caprice of chance, or fortune, or of intuition, which decides which investigator gets into his hands the substances which turn out to be the best materials for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





PREGUNTAS, POR FAVOR

E-mail: federico.gago@uah.es