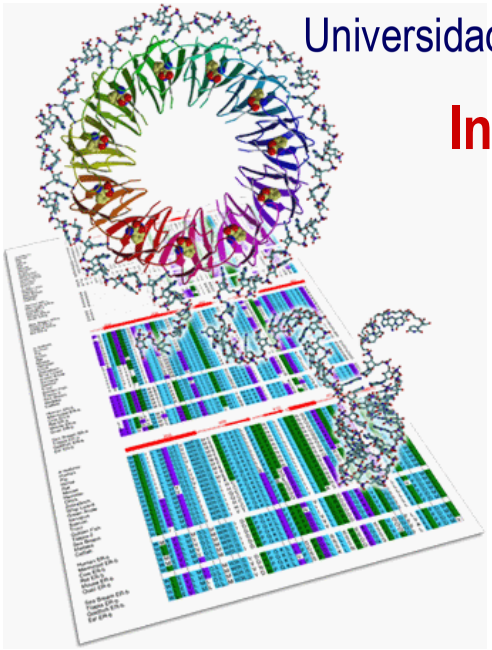
BioInfo 2004

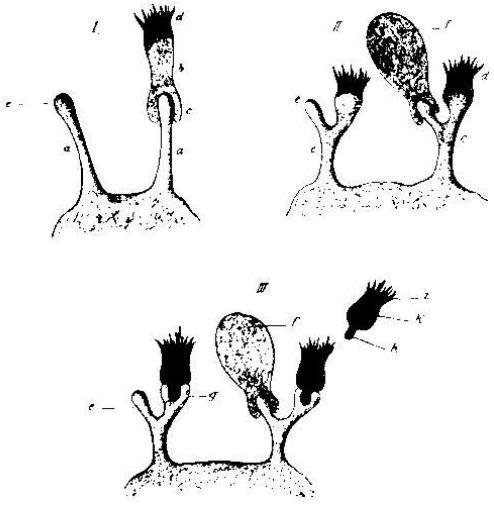
Curso de Doctorado: BIOINFORMÁTICA
Universidad Autónoma de Madrid. Marzo-Abril 2004



Interacciones entre proteínas y moléculas pequeñas (IV)

Federico Gago
Departamento de Farmacología
Universidad de Alcalá, Madrid

"Corpora non agunt nisi fixata"



Paul Ehrlich
"Address in Pathology on Chemotherapeutics:
Scientific Principles, Methods, and Results"
Lancet II, 445 (1913)

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)

1st QSAR 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.

(The Paper for which the Makdougall-Brisbane Prize was awarded; biennial period 1868-69.)

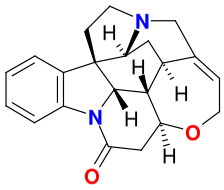
FROM THE

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

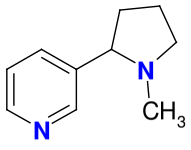
EDINBURGH:

PRINTED FOR THE SOCIETY BY NEILL AND COMPANY.

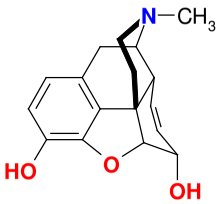
MDCCCLXVIII.



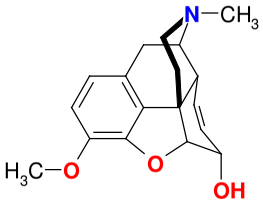
Estricnina



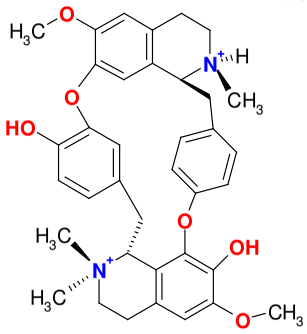
Nicotina



Morfina



Codeína



d-tubocurarina

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)

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)

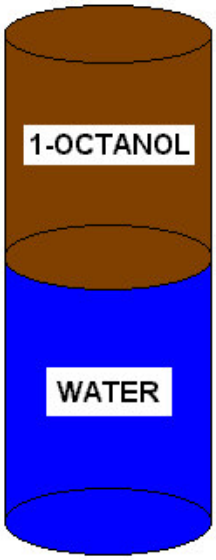
MOLECULAR PARAMETERS USED IN QSAR:

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'...

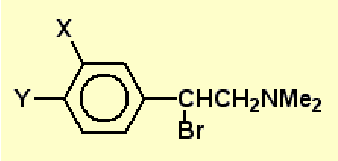
molecular shape/geometry: Taft's parameters, Kier's molecular connectivity indices, Verloop's sterimol parameters...

Hydrophobicity



"shake flask" experiment

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

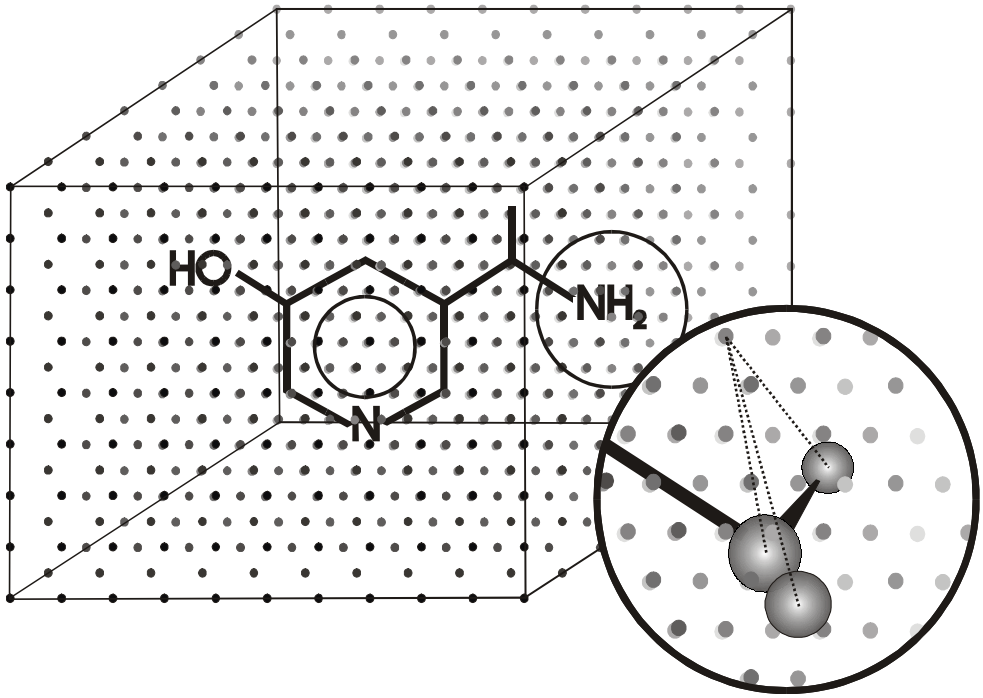
"Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins"

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)



THE ESSENCE OF 3D-QSAR IS:

* select a group of molecules, each possessing a measured biological response

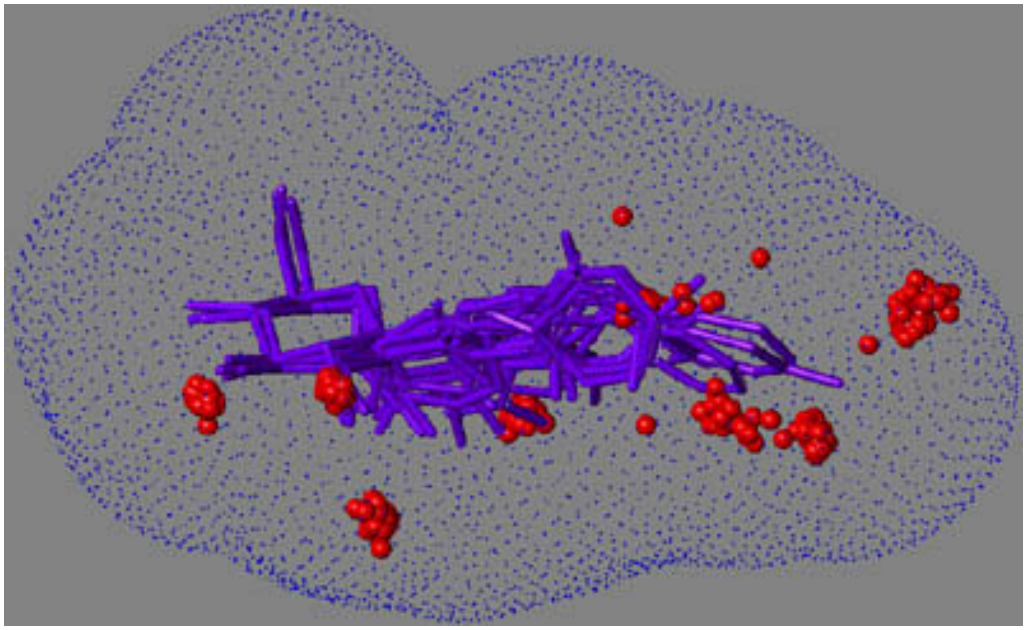
* align molecules according to some predetermined orientation rules

* calculate a set of spatially dependent parameters for each molecule determined in the receptor space surrounding the aligned series

* derive a function that relates each molecule's spatial parameters to their respective biological property

* establish self-consistency and predictive ability of the derived function

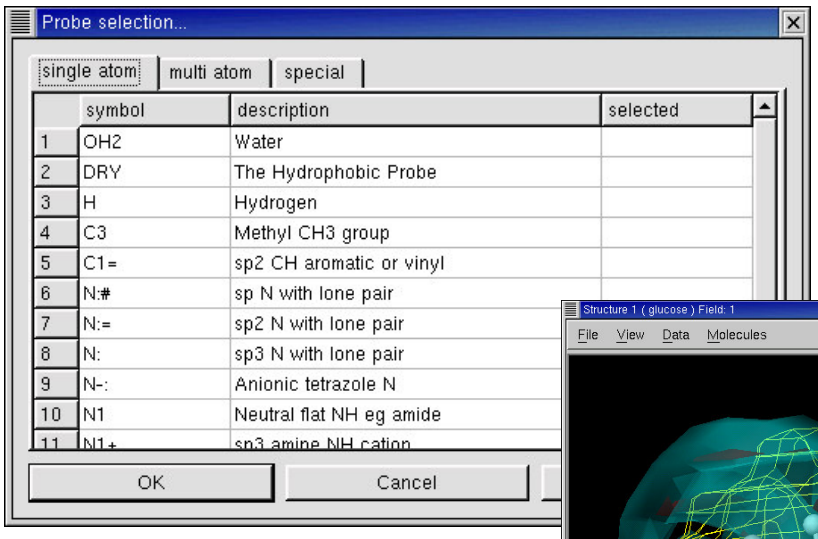
Green & Marshall, *TiPS 16, 285 (1995)*

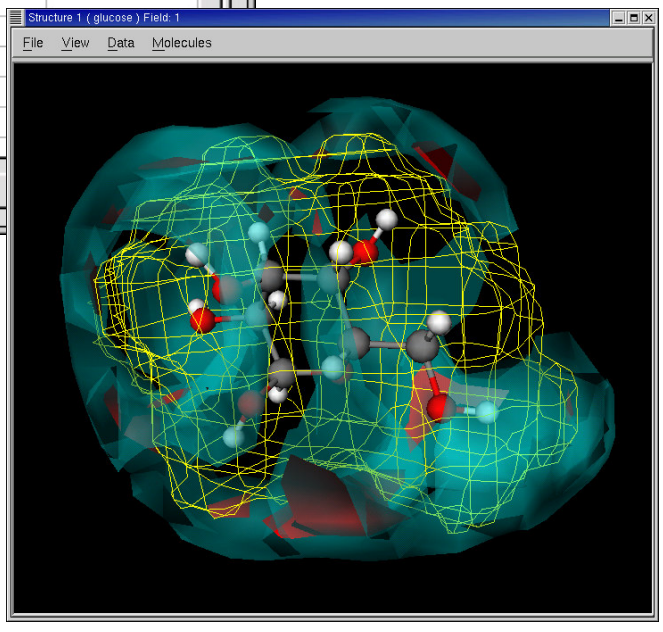


Manuel Pastor, Gabriele Cruciani, Kimberly Watson

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

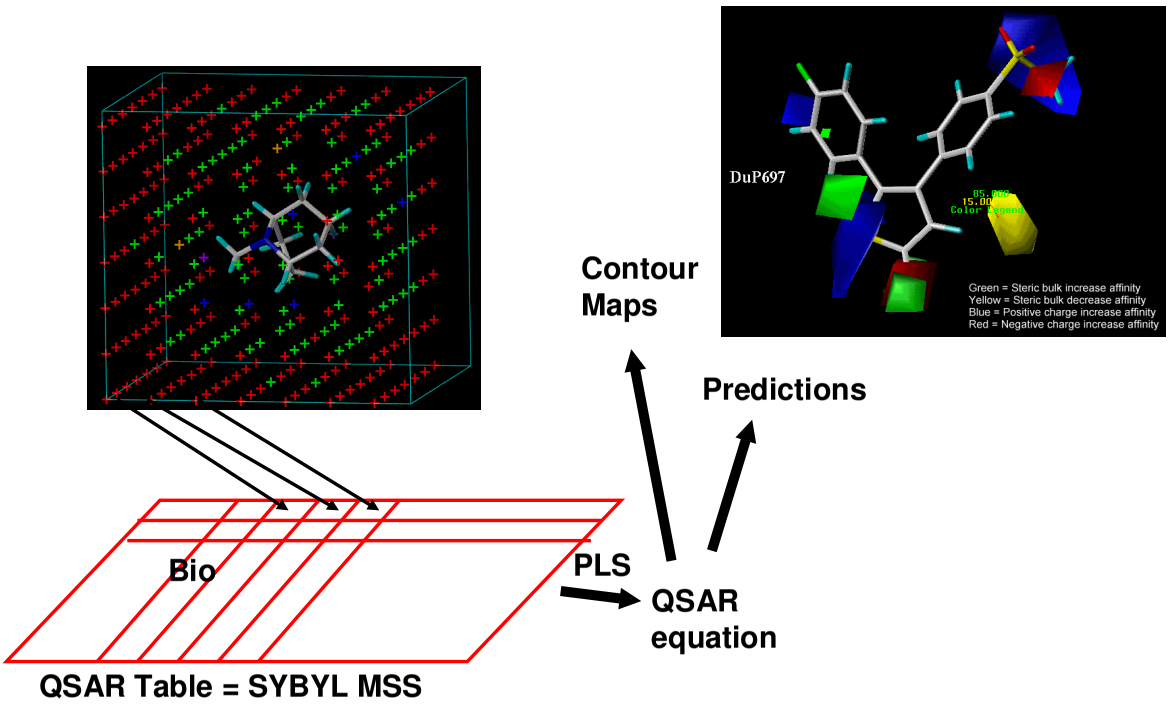
J.Med.Chem. 40, 4089-4102 (1997)



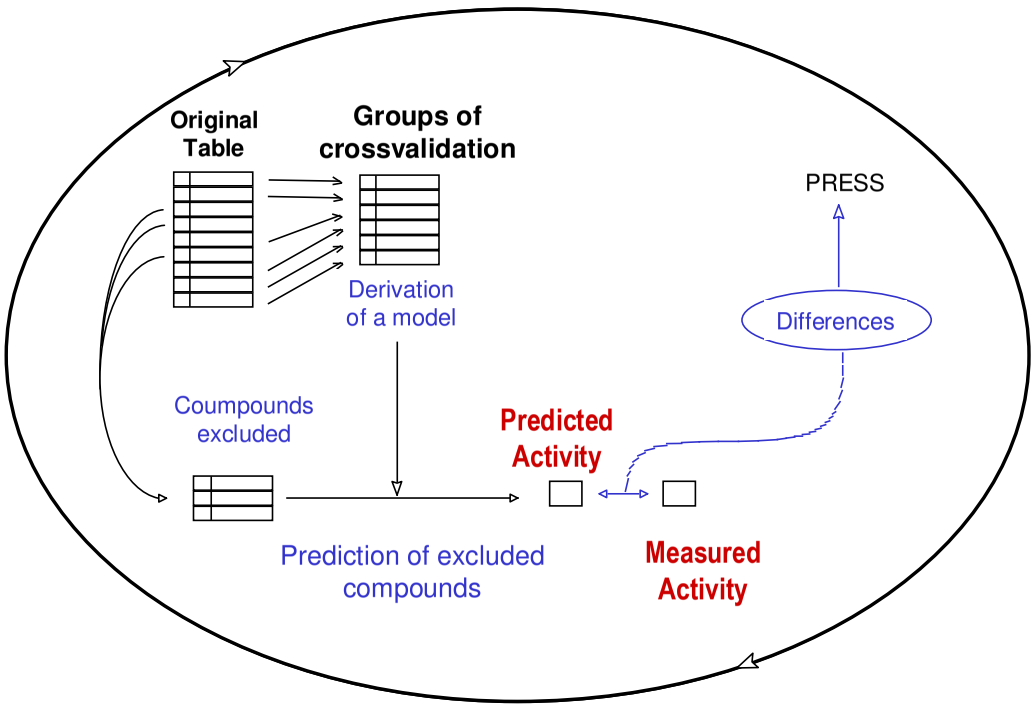


http://www.moldiscovery.com/

CoMFA is a (3D-Q)SAR method



Cross-validated PLS analyses



TRADITIONAL QSAR

Disadvantages:

- Congeneric series
- Missing physicochemical parameter values
- Lack of 3D structural information
- Results expressed only as a numerical equation
- Collinearity of parameters must be avoided
- Inadequate description of steric effects
- Inadequate description of hydrogen bonding

3D-QSAR

Advantages:

- Mixed series
- No parameters must be estimated
- 3D structural information included
- Results can be graphically displayed in 3D
- Energy fields can be collinear
- Good description of steric effects
- Good description of hydrogen bonding

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

TRADITIONAL QSAR

Advantages:

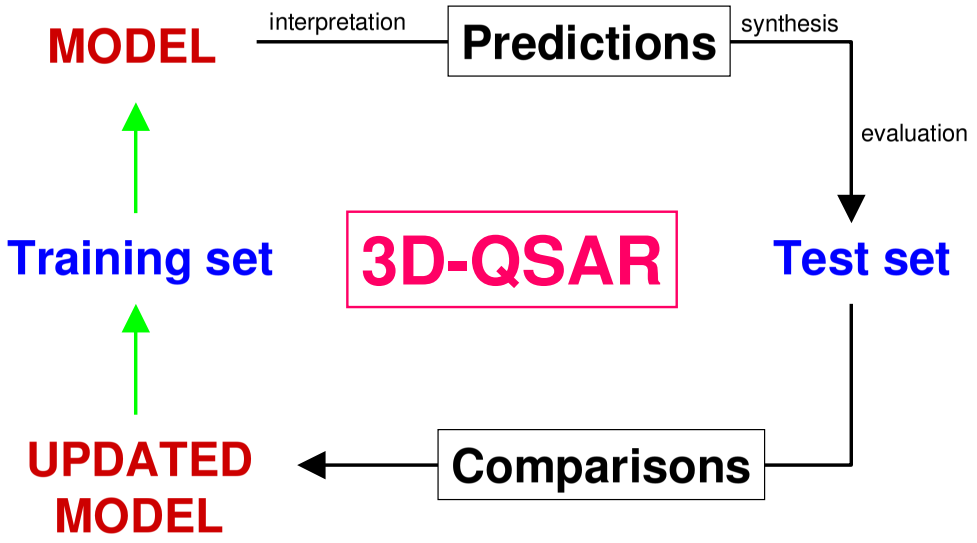
- Simplicity and speed
- No bioactive conformation required
- No alignment needed
- May extrapolate into unexplored region with care
- Results summarized in a simple equation
- Useful information is provided by the coefficients in the correlation equation
- No weighting of parameters is necessary
- Simple use of indicator variables

3D-QSAR

Disadvantages:

- More complicated to run
- A bioactive conformation must be assumed
- Superposition rules and alignment problems
- Difficult to extrapolate into unexplored regions
- Results not usually summarized in an equation
- Less useful information from the coefficients obtained in the correlation equation
- Many adjustable parameters involved
- Use of indicator variables is not straightforward

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



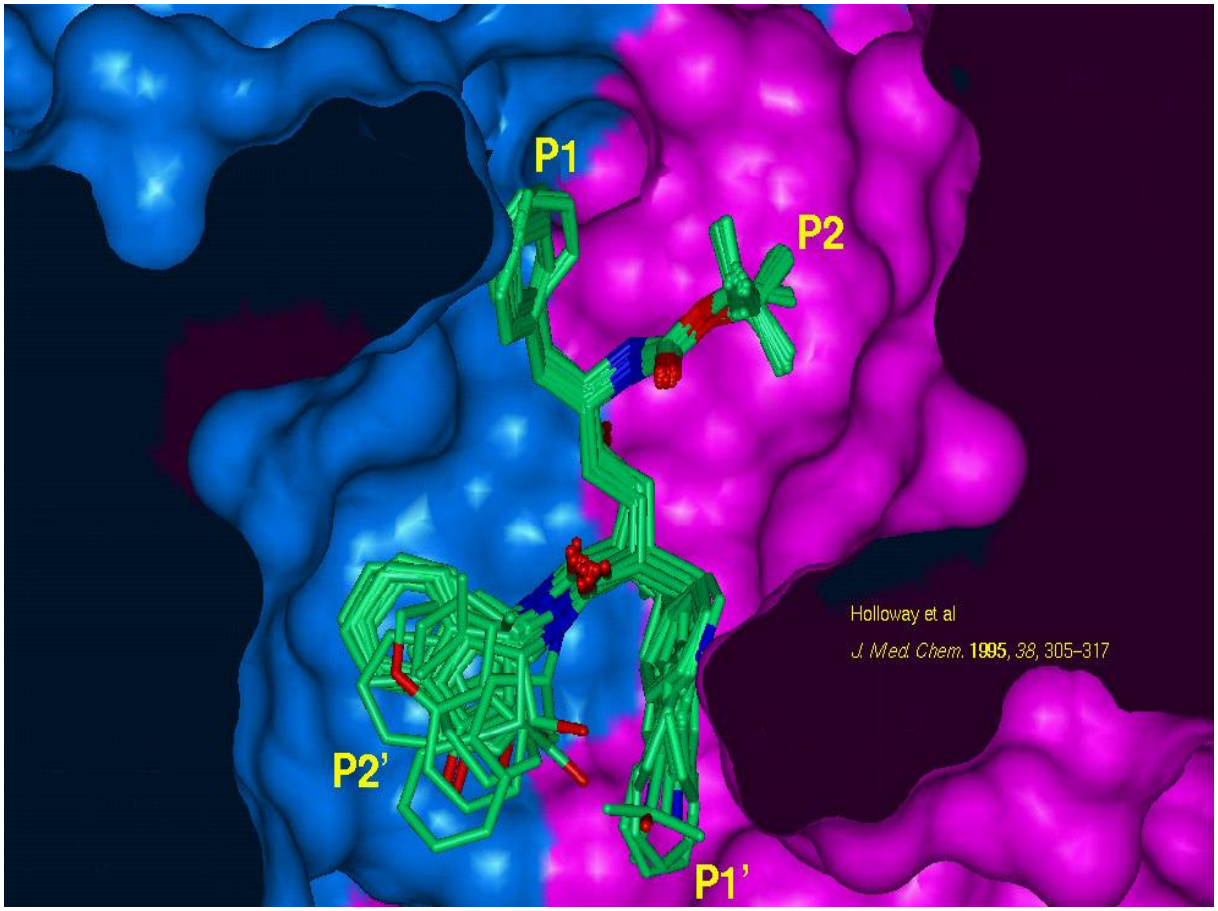
Performance

Standard Deviation of Error in Predictions:

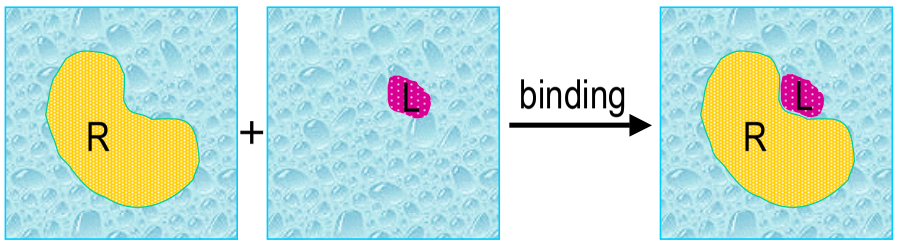
$$SDEP = \sqrt{\sum \frac{(Yexp(i) - Ypred(i))^2}{N}} = \sqrt{\frac{PRESS}{N}}$$

Correlation Coefficient in Cross-Validation:

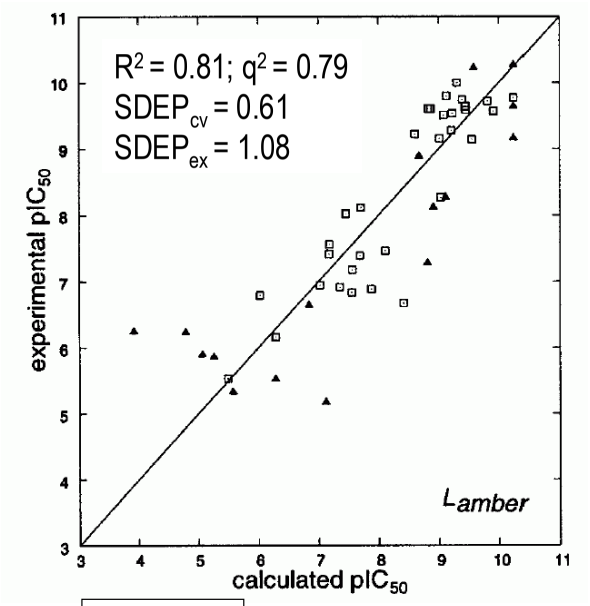
$$Q^2 = 1 - \left[\frac{\sum (Yexp(i) - Ypred(i))^2}{\sum (Yexp(i) - \langle Yexp \rangle)^2} \right]$$



ENERGETICS OF COMPLEX FORMATION

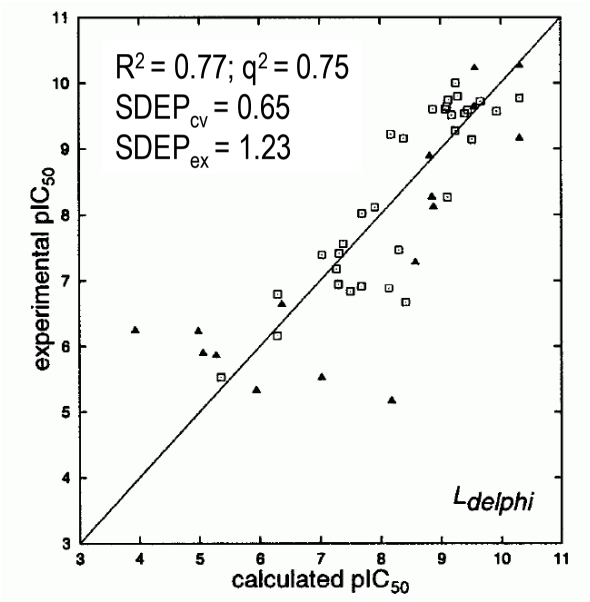


$$\Delta E_{binding} = E_{LR} - (E_R + E_L)$$



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

(COMBINE)

Análisis Comparativo de Energías de Unión

Comparative Molecular Field Analysis

(CoMFA)

Análisis Comparativo de Campos Moleculares

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

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

MODEL DERIVATION

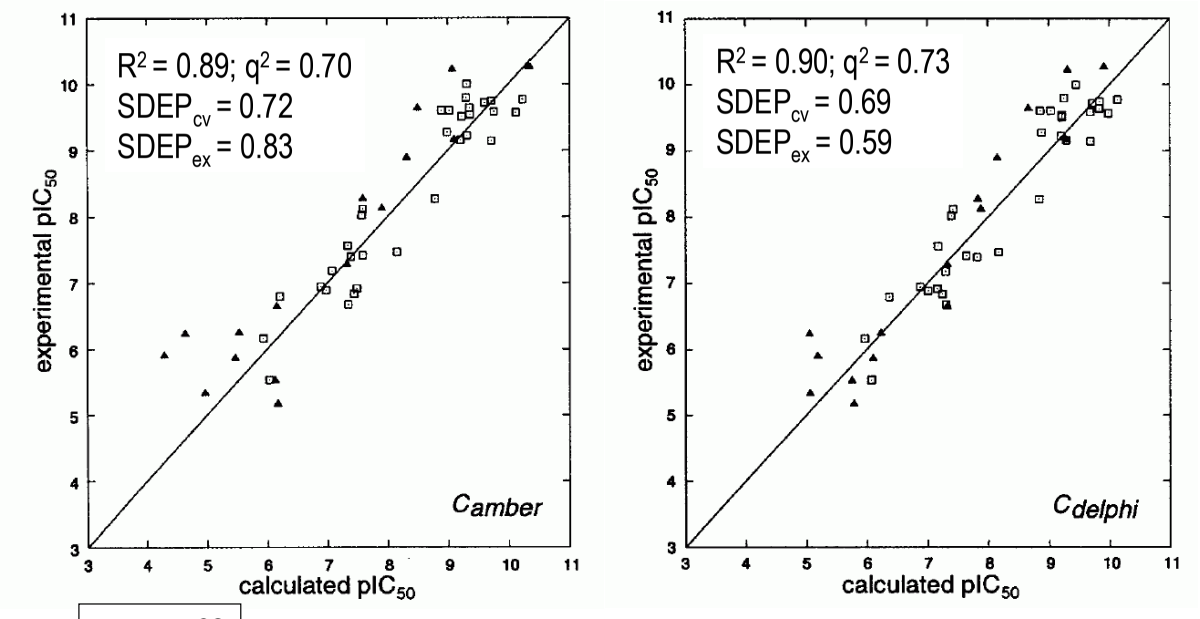
Activity = $\sum_{i=1}^{n} w_i \Delta u_i^{sel} + C$

Principal Component Analysis (PCA)

Partial Least Squares (PLS)

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

PREDICTIONS: error assessment

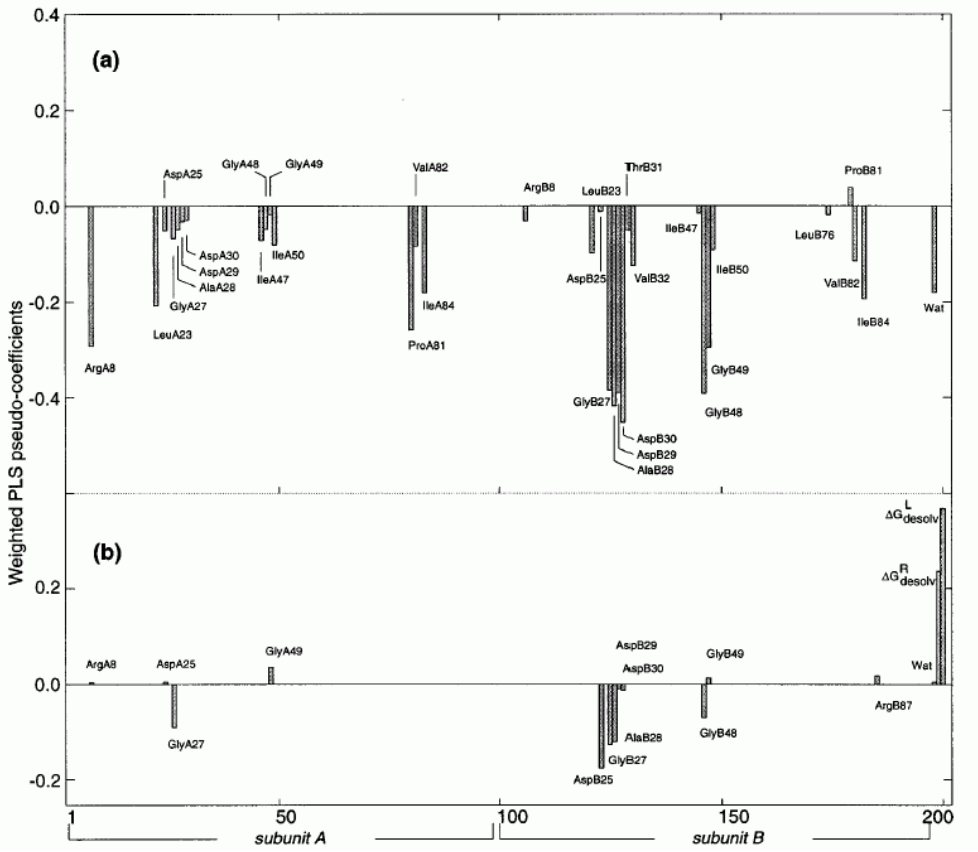


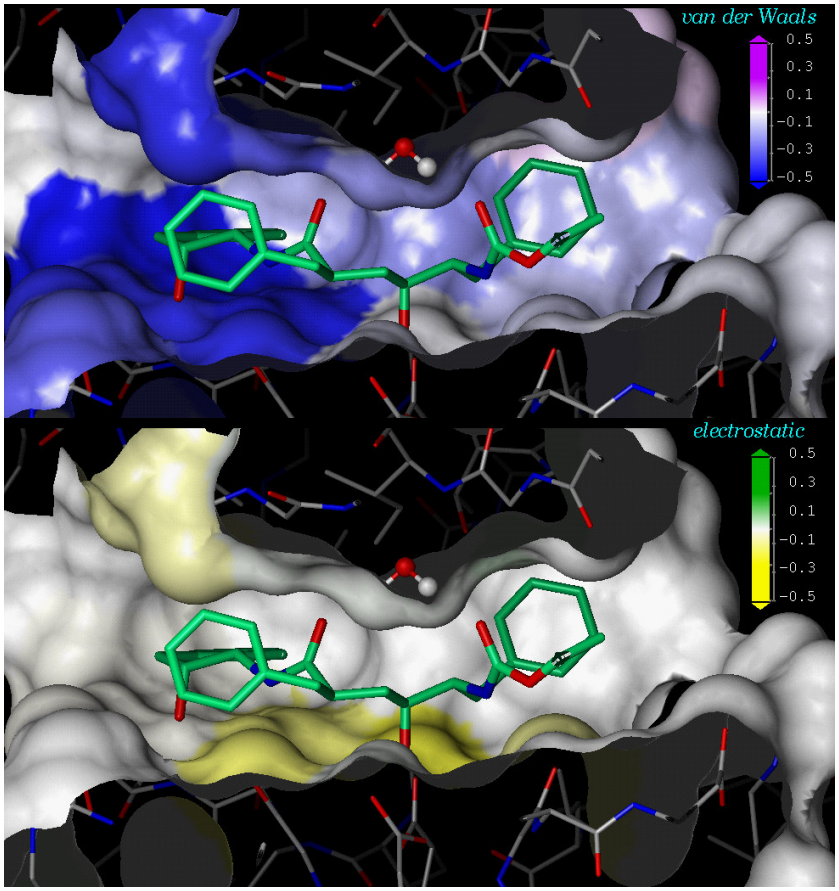
C. Pérez, M. Pastor, A. R. Ortiz & F. Gago

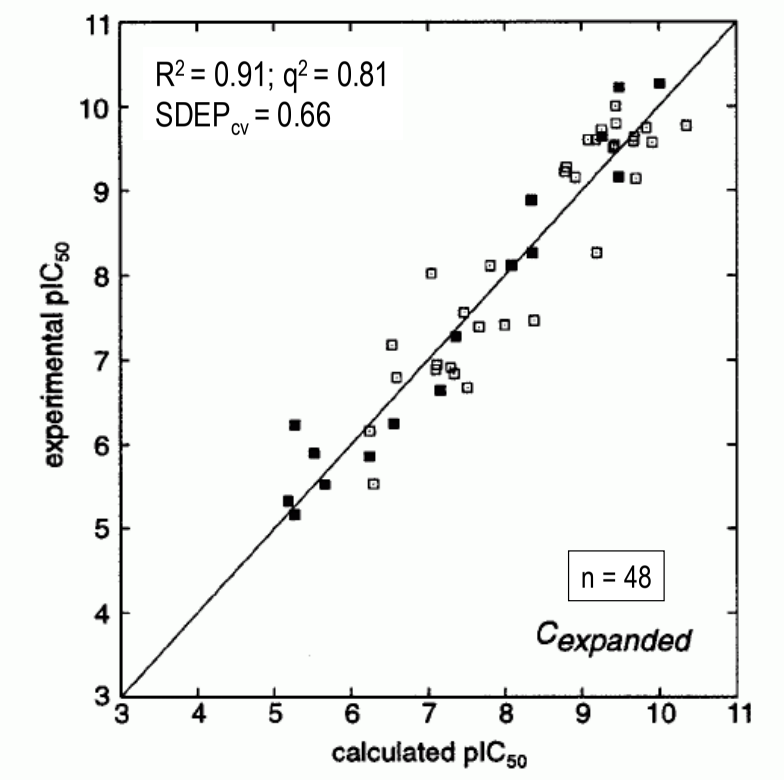
J. Med. Chem. ***41***, 836 (1998)

COMBINE analysis:

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


(a)
(b)
Weighted PLS pseudo-coefficients
0.4
0.2
0.0
-0.2
-0.4
ArgA8
AspA25
GlyA48
GlyA49
ValA82
ArgB8
LeuB23
ThrB31
ProB81
IleB47
LeuB76
AspA30
AspA29
AlaA28
GlyA27
LeuA23
IleA50
IleA47
IleA84
ProA81
AspB25
ValB32
IleB50
ValB82
IleB84
Wat
GlyB49
GlyB27
GlyB48
AspB30
AspB29
AlaB28
ΔG^{L}_{desolv}
ΔG^{R}_{desolv}
AspB29
AspB30
GlyB49
ArgB87
AlaB28
GlyB27
AspB25
1
50
100
150
200
subunit A
subunit B


van der Waals
0.5
0.3
0.1
-0.1
-0.3
-0.5
electrostatic
0.5
0.3
0.1
-0.1
-0.3
-0.5



Combinatorial Chemistry & High Throughput Screening, 2001, *4*, 627-642

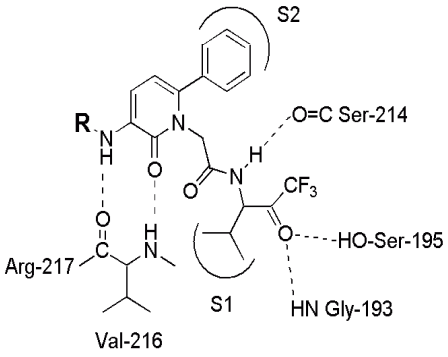
Comparative Binding Energy (COMBINE) Analysis of Human Neutrophil Elastase Inhibition by Pyridone-containing Trifluoromethylketones

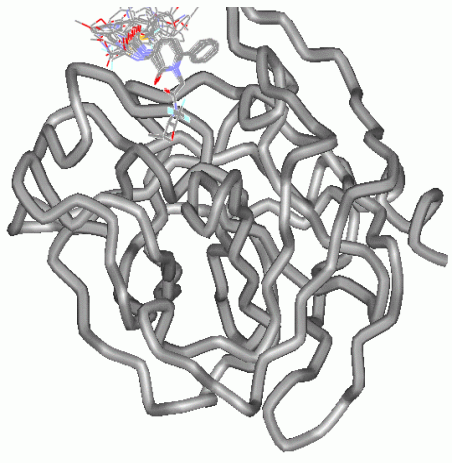
Carmen Cuevas,[†] Manuel Pastor,[#] Carlos Pérez[†] and Federico Gago[*]

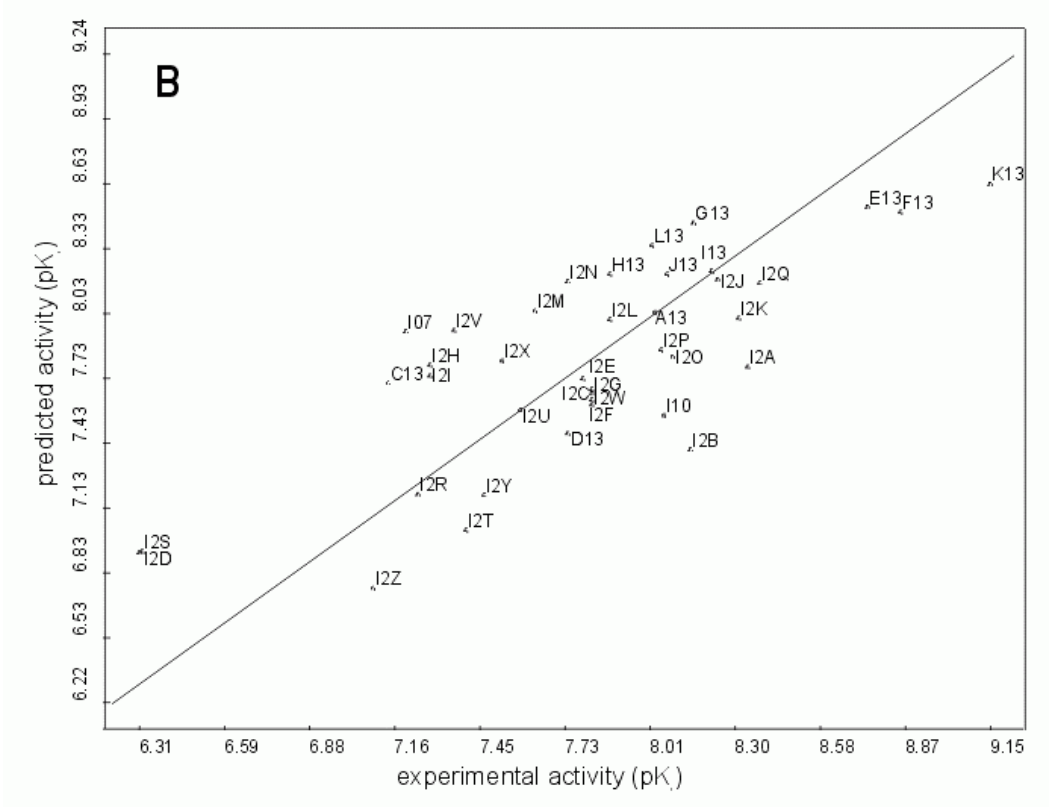
Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain

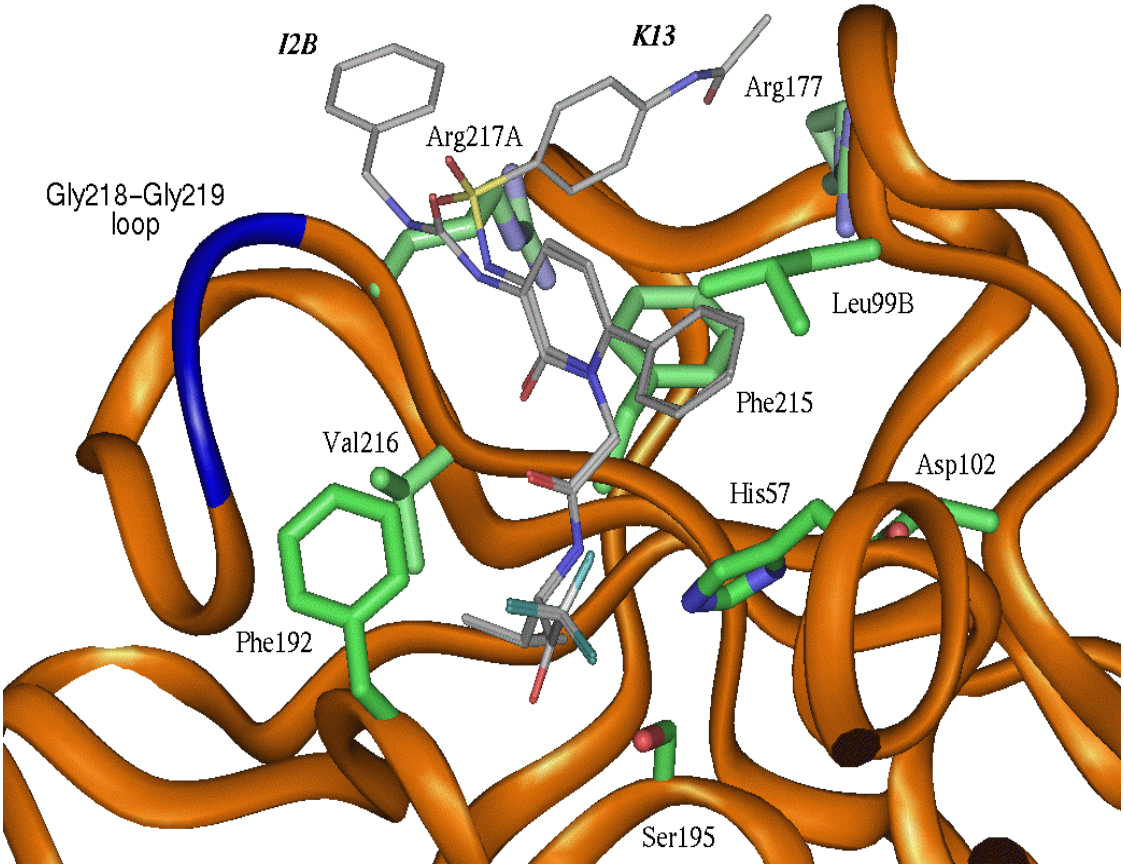
[†] *Present address: Pharma Mar S.A., Cantoblanco, 28760 Tres Cantos, Madrid, Spain*

[#] *Present address: Departament de Ciències Experimentals i de la Salut, Universitat Pompeu, Fabra, Dr. Aiguader 80, E-08003 Barcelona, Spain*





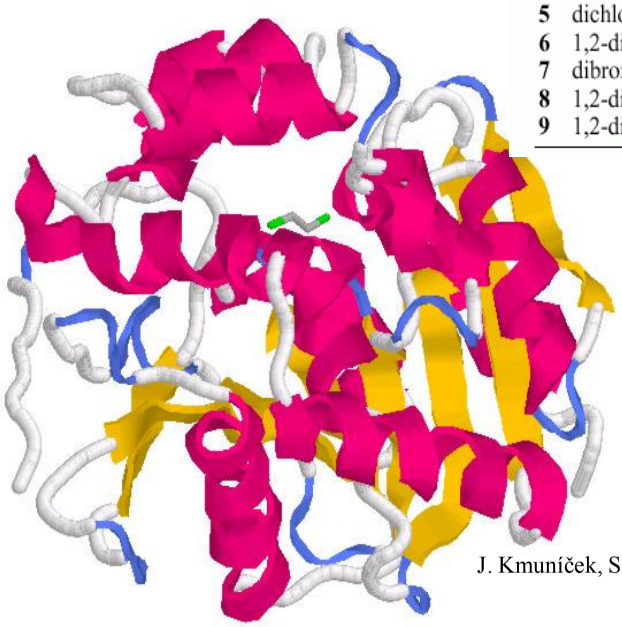

B
predicted activity (pKi)
experimental activity (pKi)
9.24
8.93
8.63
8.33
8.03
7.73
7.43
7.13
6.83
6.53
6.22
6.31
6.59
6.88
7.16
7.45
7.73
8.01
8.30
8.58
8.87
9.15
K13
E13
F13
G13
L13
I13
J13
H13
I2N
I2J
I2Q
I2M
I2L
A13
I2K
I07
I2V
I2P
I2H
I2X
I2O
I2A
C13
I2I
I2E
I2G
I2C
I2W
I2F
I10
I2U
D13
I2B
I2R
I2Y
I2T
I2S
I2D
I2Z

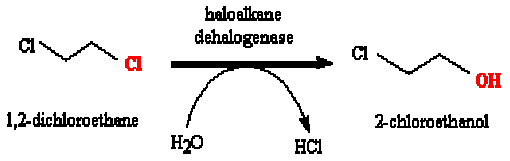

I2B
K13
Arg177
Arg217A
Gly218–Gly219
loop
Leu99B
Phe215
Val216
Asp102
His57
Phe192
Ser195

Haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10

Table 1: Steady-State Dissociation Constants of Haloalkane Dehalogenase[a]

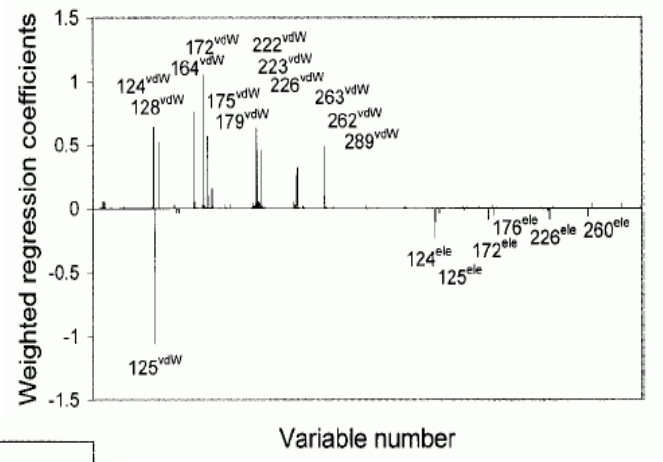
	compound	log K_m (mM)		compound	log K_m (mM)
1	1-chlorobutane	0.34	**10**	1,2-dibromopropane	0.11
2	1-chlorohexane	0.15	**11**	2-chloroethanol	2.60
3	1-bromobutane	−1.22	**12**	2-bromoethanol	1.04
4	1-bromohexane	−0.52	**13**	epichlorohydrine	1.68
5	dichloromethane	2.00	**14**	epibromohydrine	0.34
6	1,2-dichloroethane	−0.28	**15**	2-chloroacetonitrile	0.80
7	dibromomethane	0.38	**16**	2-bromoacetonitrile	−0.31
8	1,2-dibromoethane	−2.00	**17**	2-chloroacetamide	2.00
9	1,2-dichloropropane	1.11	**18**	2-bromoacetamide	1.30

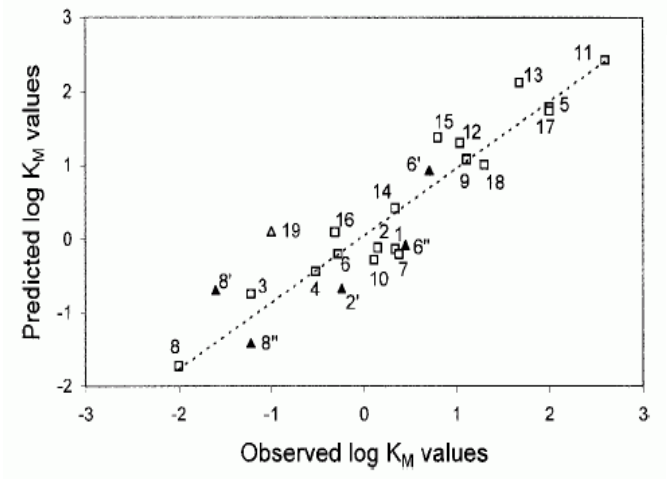




J. Kmuníček, S. Luengo, F. Gago, A.R. Ortiz, R.C. Wade & J. Damborský
Biochemistry, **40,** 8905-8917 (2001)

Selected energy contributions in the best COMBINE model





- □ training set
- ▲ prediction set
- ' Phe172Trp mutant enzyme
- " Trp175Tyr mutant enzyme
- △ new substrate + Phe172Trp mutant enzyme

STRUCTURE CONSTRUCTION

• Determine locations in space that are favourable for non-covalent interactions with the protein, e.g. *GRID*, *MCSS*.

• Connect these locations using spacer residues, e.g. *BUILDER*.

• Fill the active site with connected fragments drawn from a prechosen library, e.g. *LUDI*.

• Dock a drug molecule whose shape is complementary to that of the active site, e.g. *DOCK*.

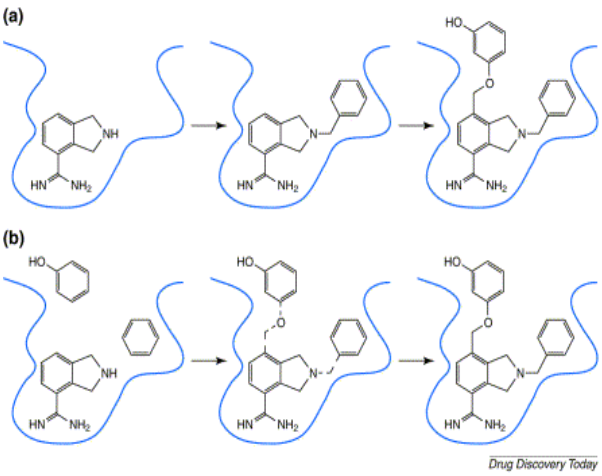
• Dynamic creation and breakage of bonds, permutation of building blocks over time and empirical energy function on all the system for decision making, e.g. *CONCEPTS*.

N.B. All tree search techniques imply chained choices.

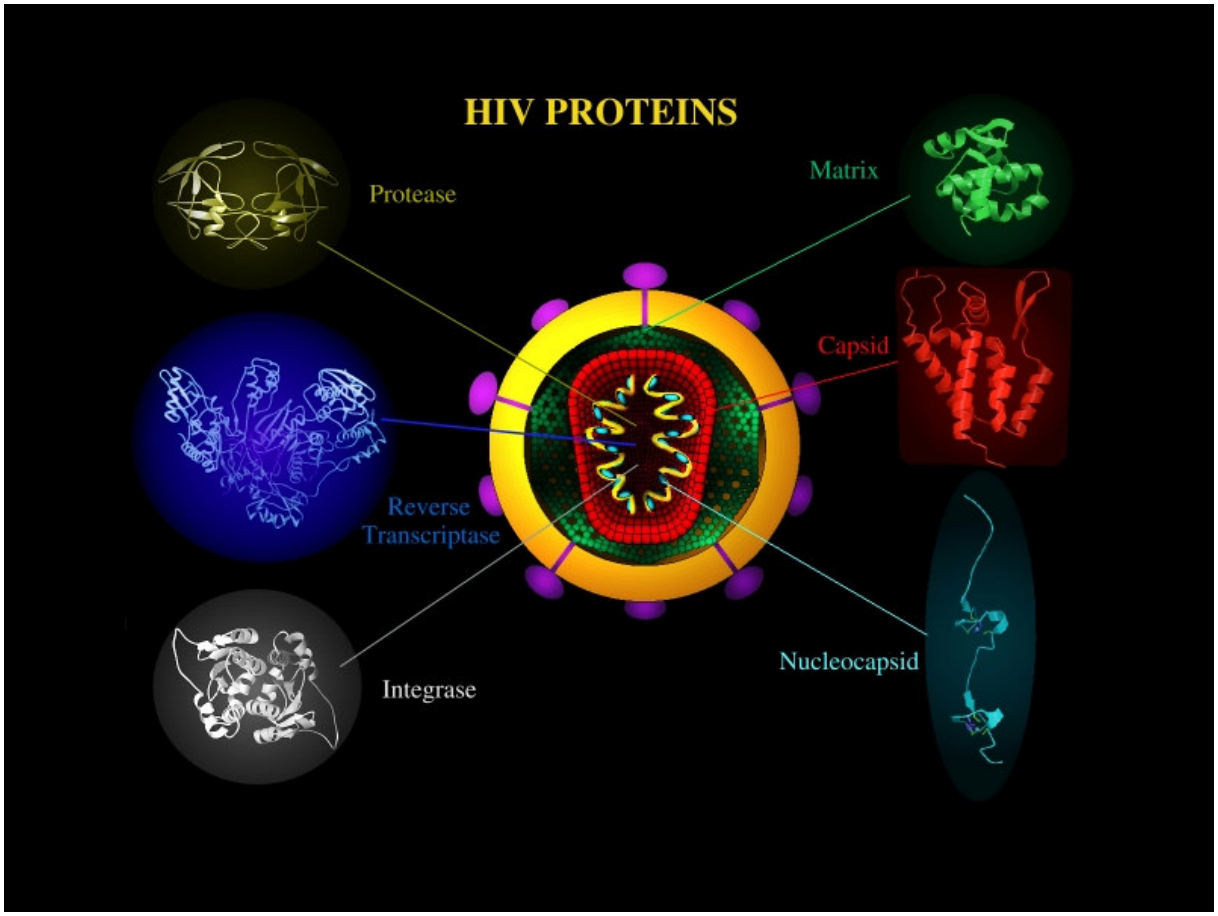
Two strategies for structure-based molecule assembly from fragments

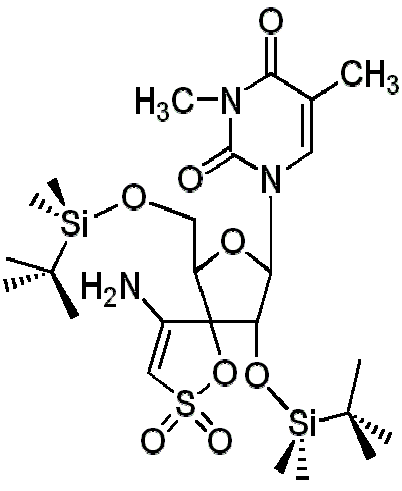
Sequential growth technique

Fragment-placing and linking



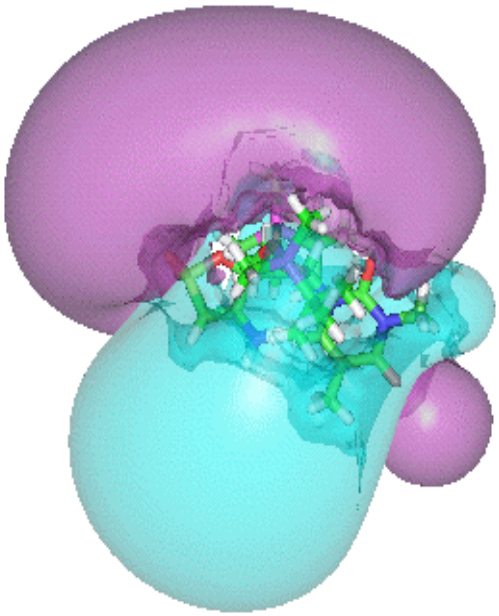
a ligand-binding pocket on the surface of a protein



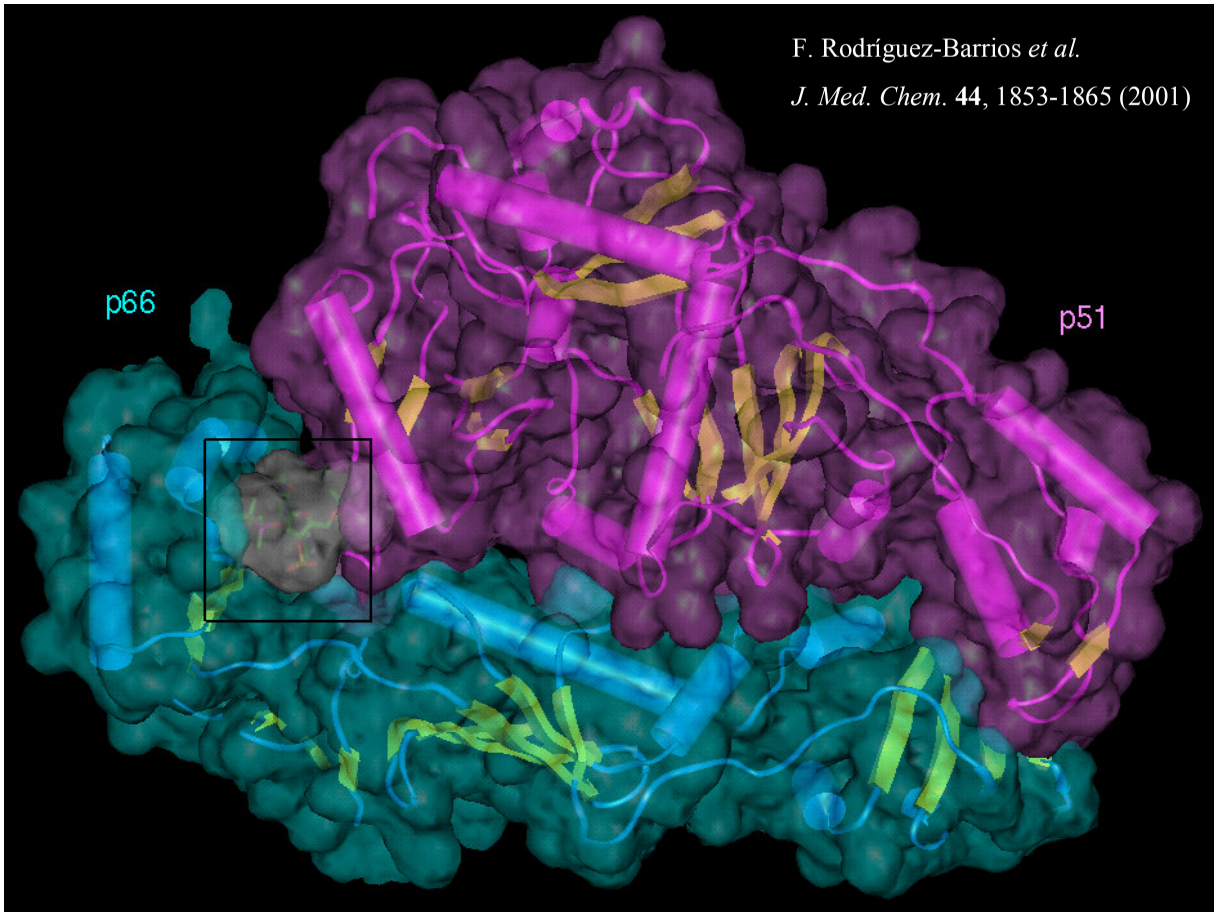


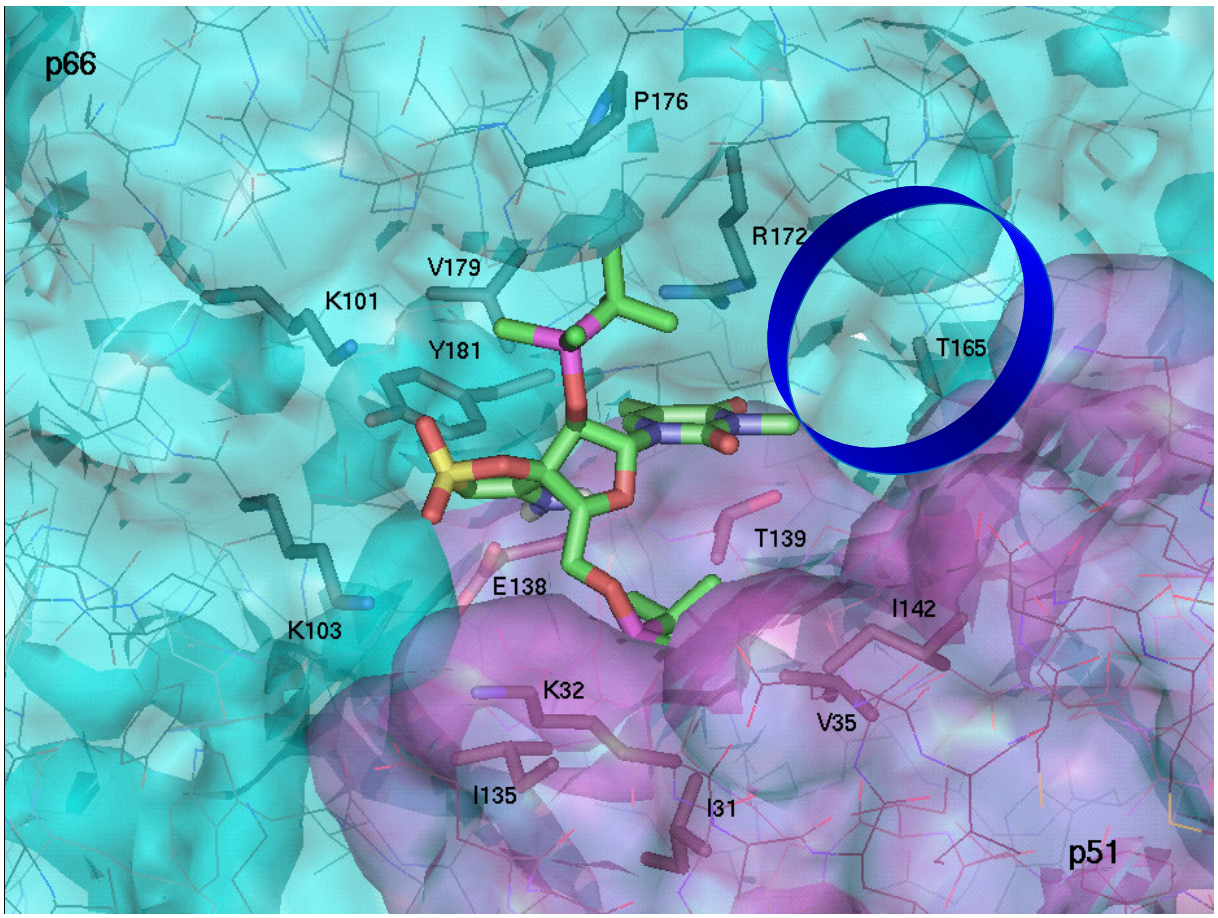
TSAO-m^{3}T

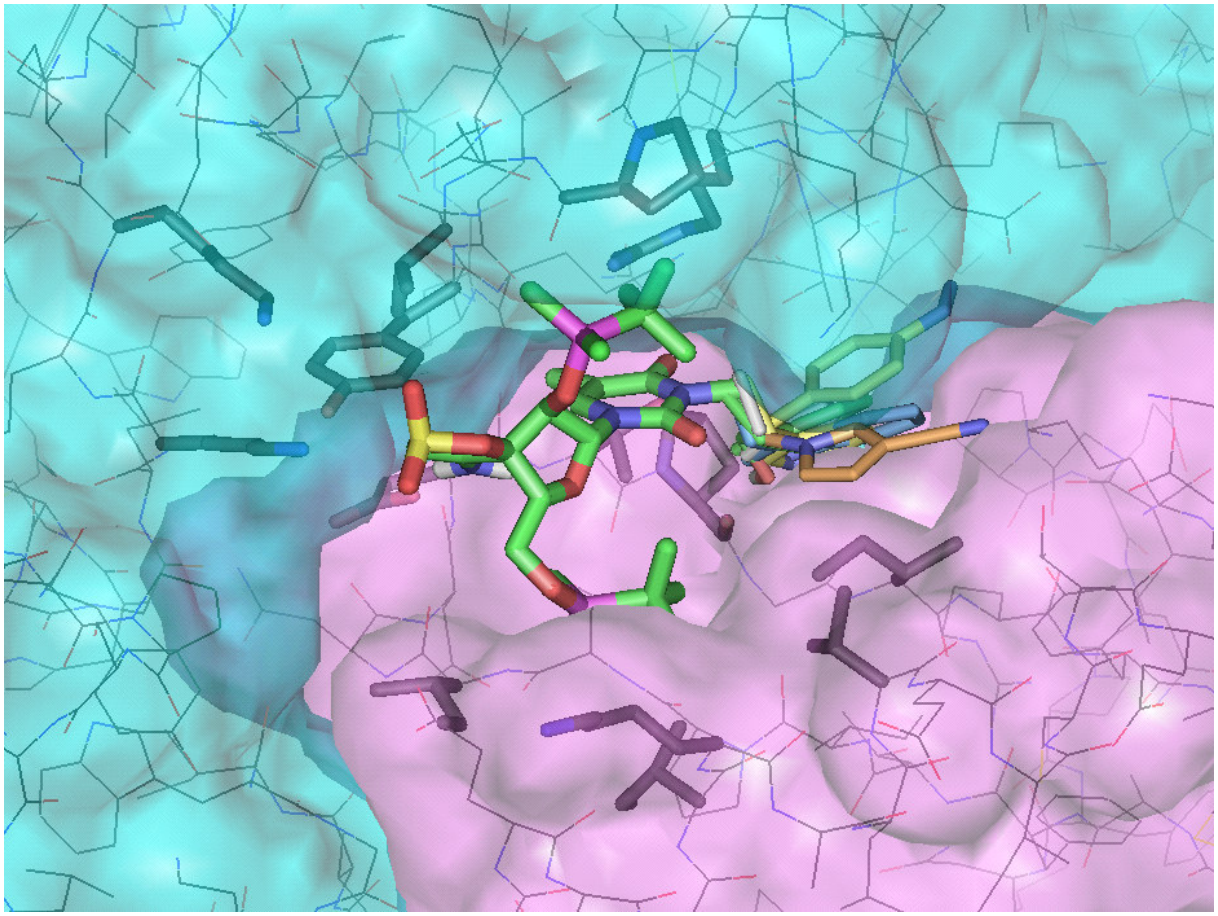
Camarasa *et al* (IQM, CSIC, Madrid)



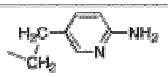
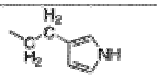
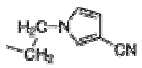
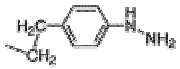
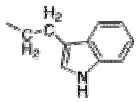
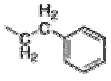
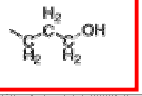
Molecular electrostatic potential
Negative: -0.1 — -3.2 kcal mol^{-1}
Positive: 0.1 — 10.3 kcal mol^{-1}

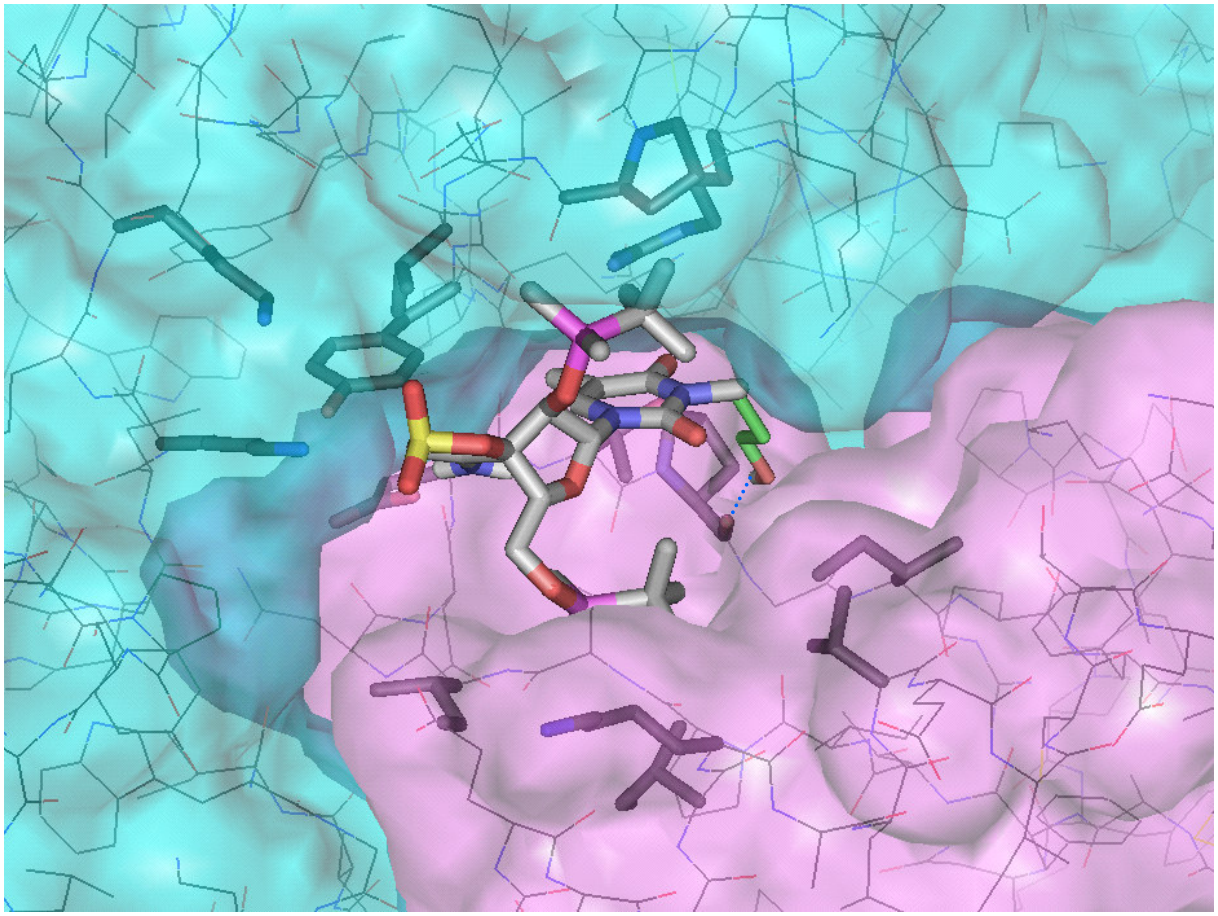

F. Rodríguez-Barrios *et al.*
J. Med. Chem. **44**, 1853-1865 (2001)
p66
p51


p66
P176
R172
V179
K101
Y181
T165
T139
E138
I142
K103
K32
V35
I135
I31
p51



Substituent Fragments Suggested by the LUDI Program for Attachment to the N3 Position of TSAO-T

id	Fragment	# hb[a]	Target[b]	id	Fragment	# hb[a]	Target[b]
I		1	CO Gly-B141	V		1	CO Pro-A140
II		1	NH_3 Lys-B49	VI		1	COO^- Glu-A169
III		1	CO Pro-A140	VII		0	Ile-B142
IV		0	Ile-B142	VIII		1	CO Pro-B140



Effect on Activity of Attaching a -CH_2-CH_2-CH_2OH Substituent to the N3 Position of TSAO-T

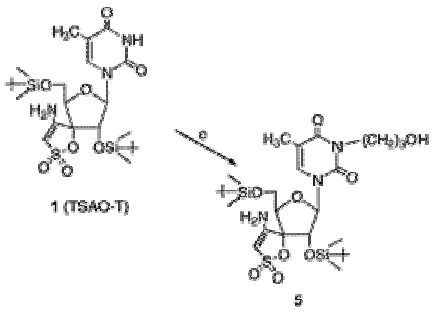
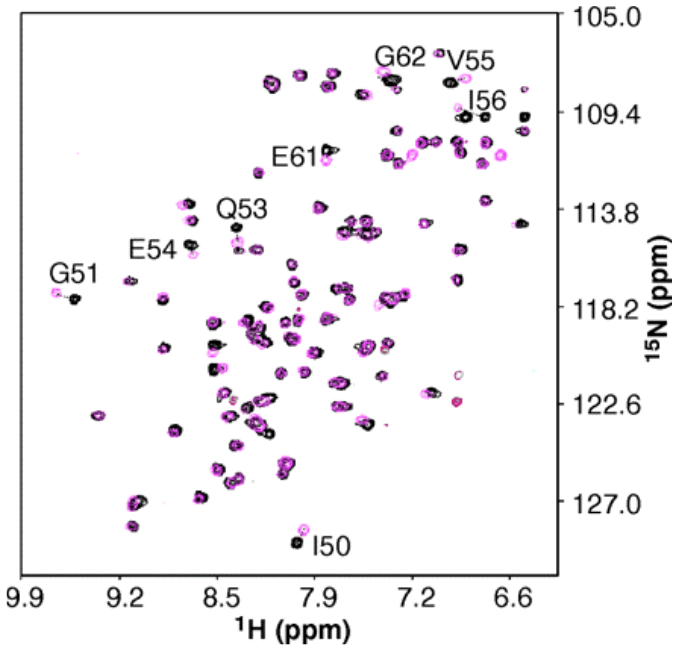


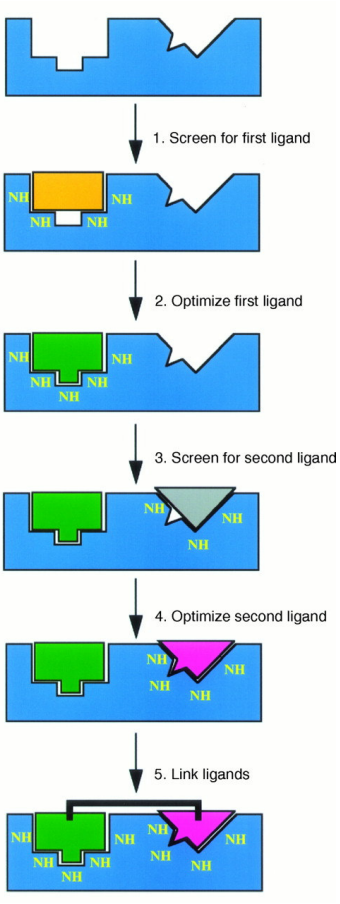
Table 4. Inhibitory Activity of TSAO Derivatives against HIV-1 and HIV-2 in CEM and MT-4 Cell Cultures

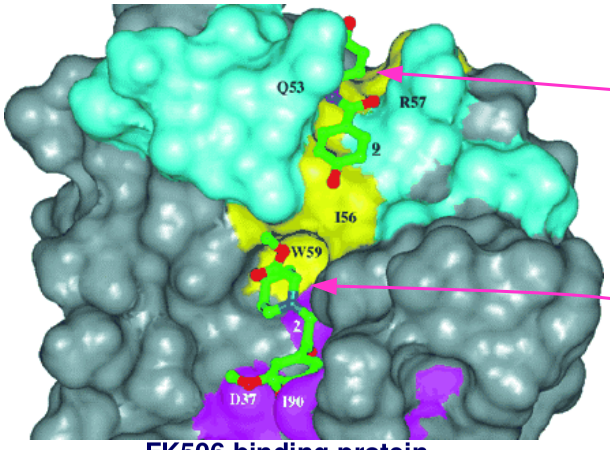
	EC_{50} (μM)[a]			
	MT-4		CEM	
compd	HIV-1	HIV-2	HIV-1	HIV-2
2	0.17 ± 0.05	>10	0.17 ± 0.05	>2
3	0.20 ± 0.03	>10	0.10 ± 0.0	>2
4	0.23 ± 0.06	>10	0.12 ± 0.0	>10
5	0.03 ± 0.003	>2	0.01 ± 0.006	>2
TSAO-T	0.06 ± 0.03	>20	0.06 ± 0.01	>20
TSAO-m^3T	0.05 ± 0.01	>200	0.06 ± 0.09	>200

[a] The 50% effective concentration, or the concentration required to protect 50% of the virus-infected cells against destruction by the virus.

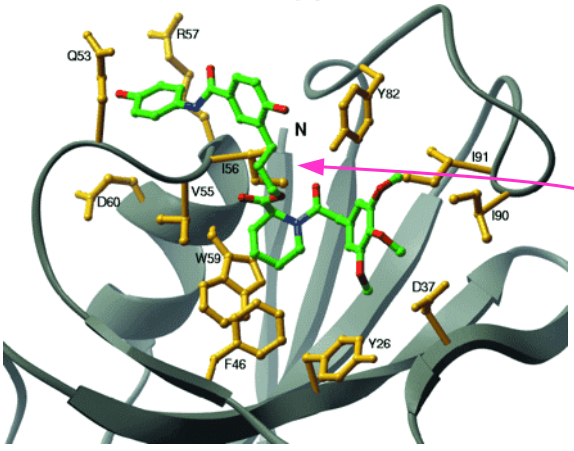
"SAR by NMR"



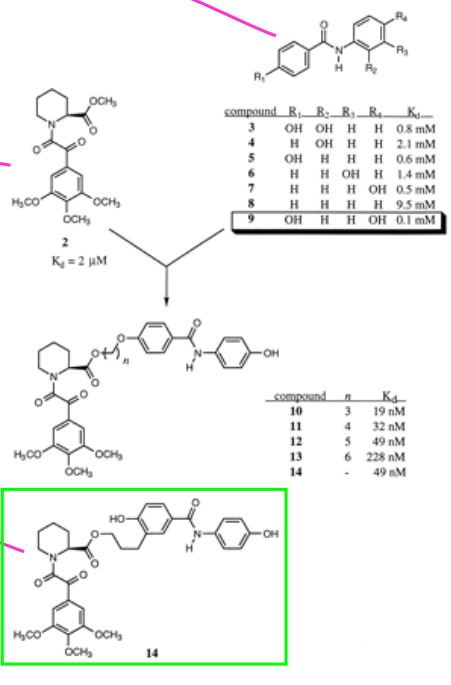




FK506 binding protein



Discovering High-Affinity Ligands for Proteins:
SAR by NMR



Schuker, S. B. et al. *Science* **1996**, *274*, 1531-1534

The life history of a successful drug

Drug discovery

↓

Initial characterisation

↓

Pre-clinical trials

↓

Regulatory approval sought to start trials in humans

↓

Clinical trials Phases I, II, III

↓

Submission of marketing/manufacturing authorisation application to regulatory authorities

↓

Regulatory authorities review information and grant (or refuse) licences

↓

Product goes on sale

↓

Post-marketing surveillance

Library of compounds

↓

In vitro screening: human/animal receptor/enzyme assay; reporter system

↓

HITS / LEAD

↓

Biochemical, tissue or animal model of function

↓

LEAD

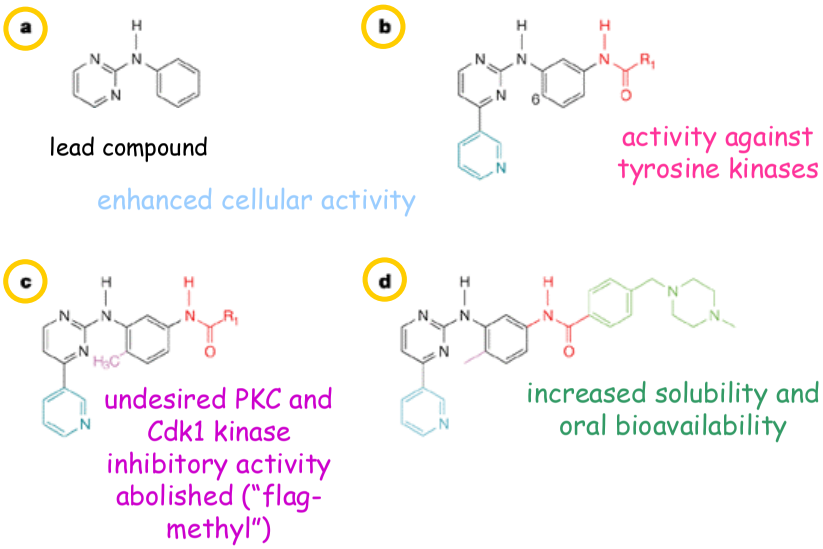
↓

Animal model of therapeutic target

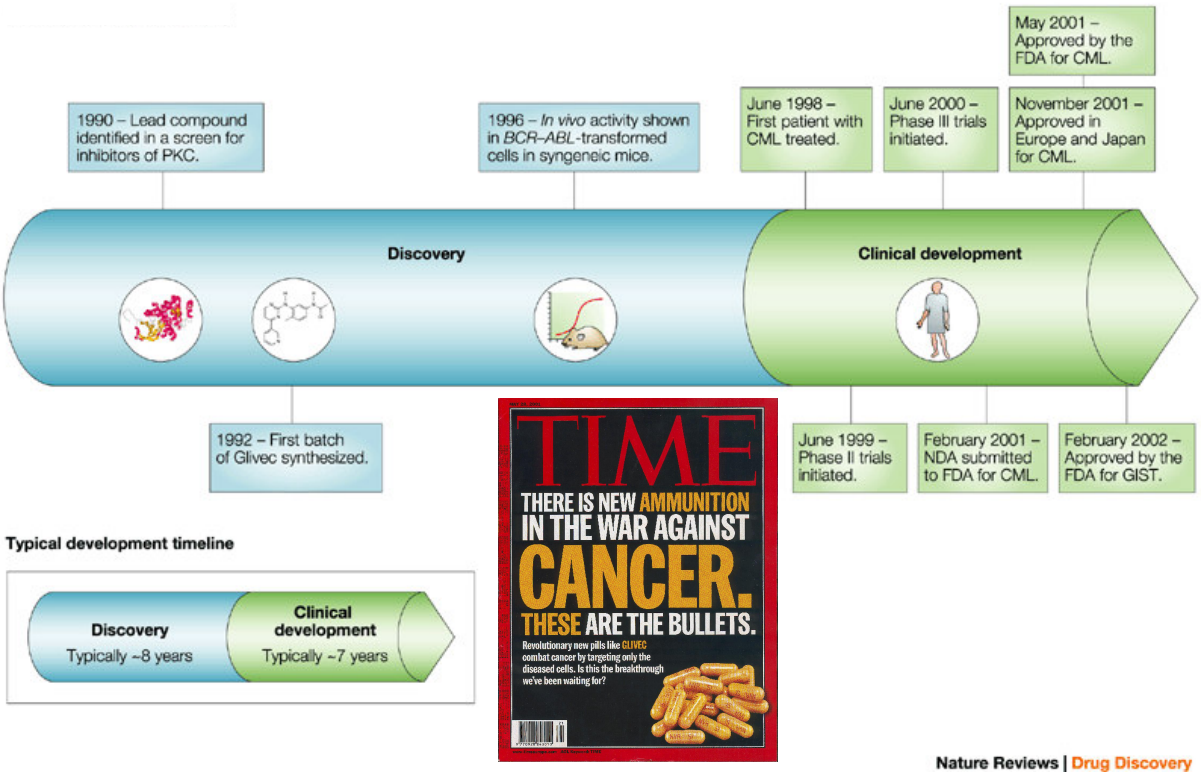
↓

ADME, formulation, acute toxicology

Glivec (STI571/ Imatinib):
a rationally developed, targeted anticancer drug



Glivec development timeline



Resistance to imatinib (STI571) therapy

1) Median levels of BCR-ABL transcripts were not significantly changed at the time of resistance but 7/55 patients showed a >10-fold increase in BCR-ABL levels;

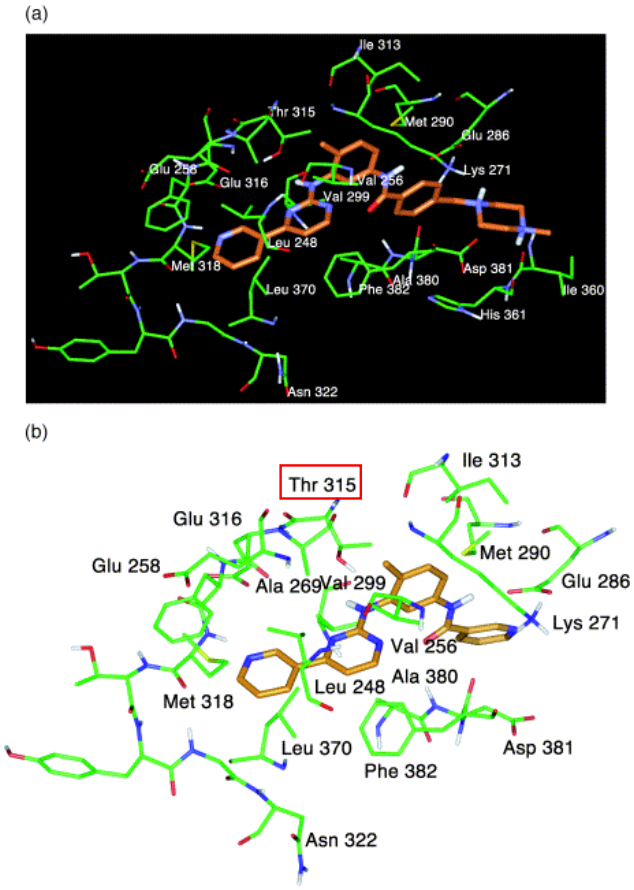
(2) genomic amplification of BCR-ABL was found in 2/32 patients evaluated by fluorescence *in situ* hybridization;

(3) additional chromosomal aberrations were observed in 19/36 patients;

(4) point mutations of the ABL tyrosine kinase domain resulting in reactivation of the BCR-ABL tyrosine kinase were detected in 23/66 patients.

Hochhaus et al. Molecular and chromosomal mechanisms of resistance to imatinib (STI571) therapy. *Leukemia* (2002) 16:2190-2196

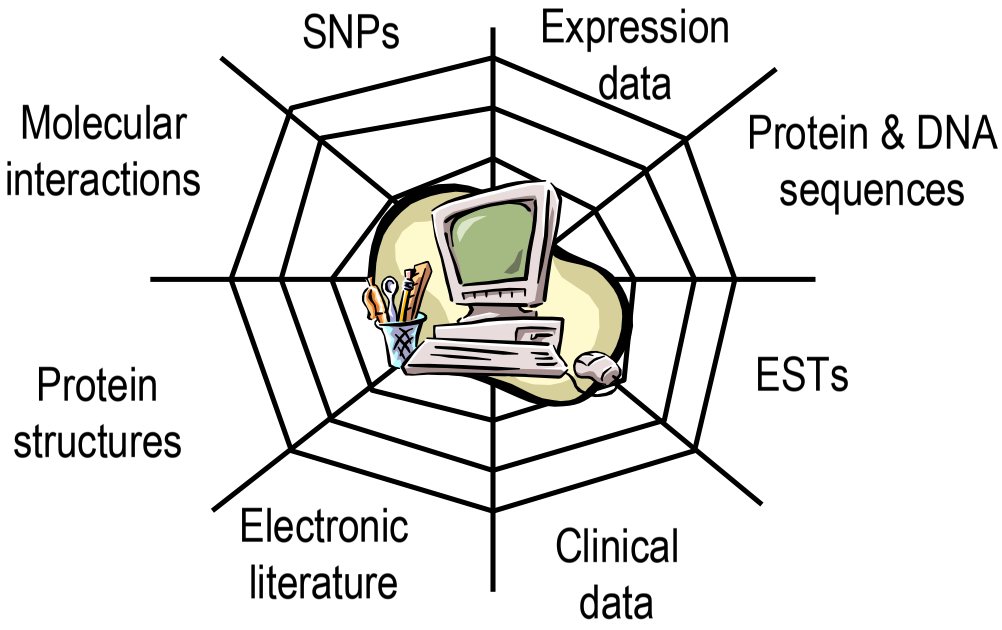
Interactions between STI571 and the kinase domain of human Abl

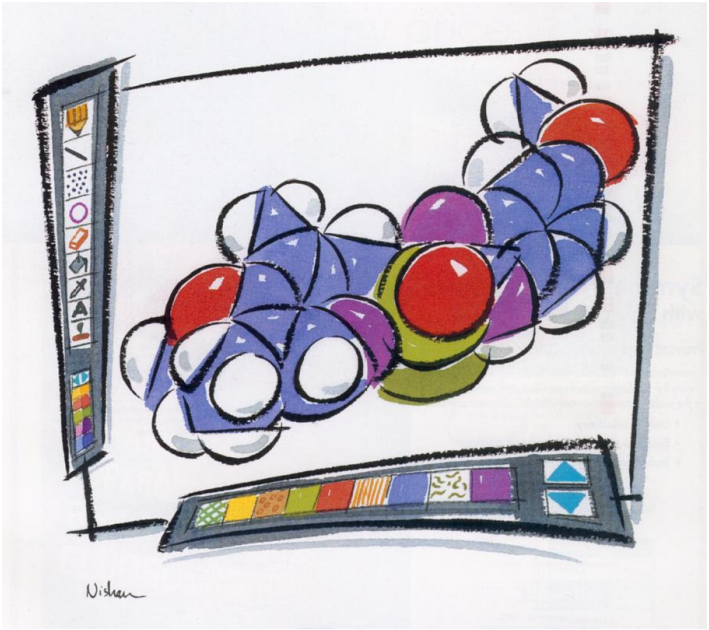


CONCLUSIONS

- 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.

The invisible web





PREGUNTAS, POR FAVOR

E-mail: federico.gago@uah.es