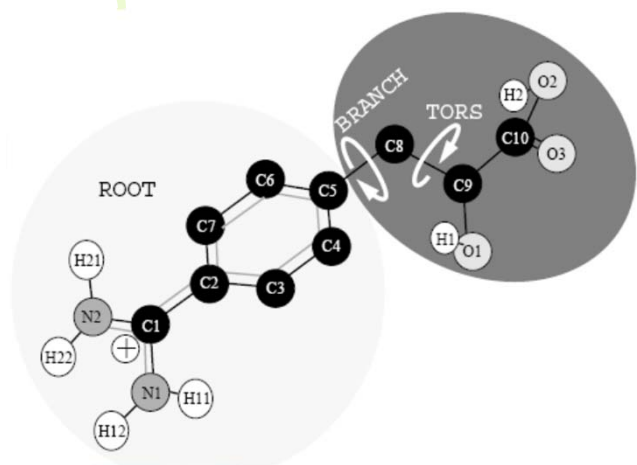# Principles of Molecular Dockings

Jung-Hsin Lin (林榮信)

Research Center for Applied Sciences &
Institute of Biomedical Sciences, Academia Sinica
School of Pharmacy, National Taiwan University
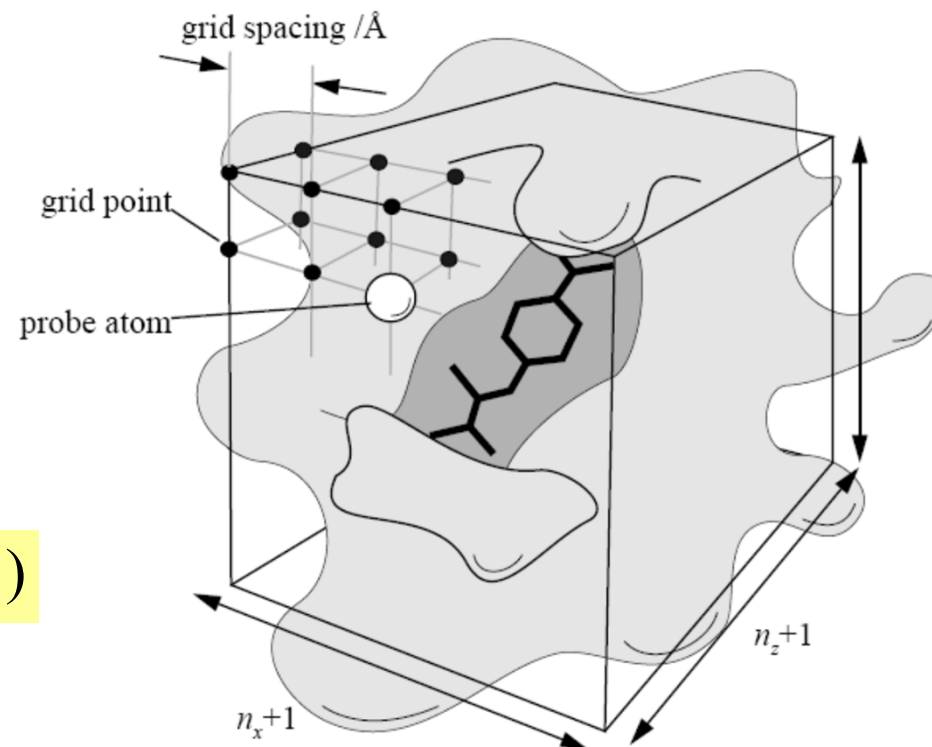College of Engineering, Chang Gung University

http://www.rcas.sinica.edu.tw/faculty/jhlin.html

The Flexible Docking Problem
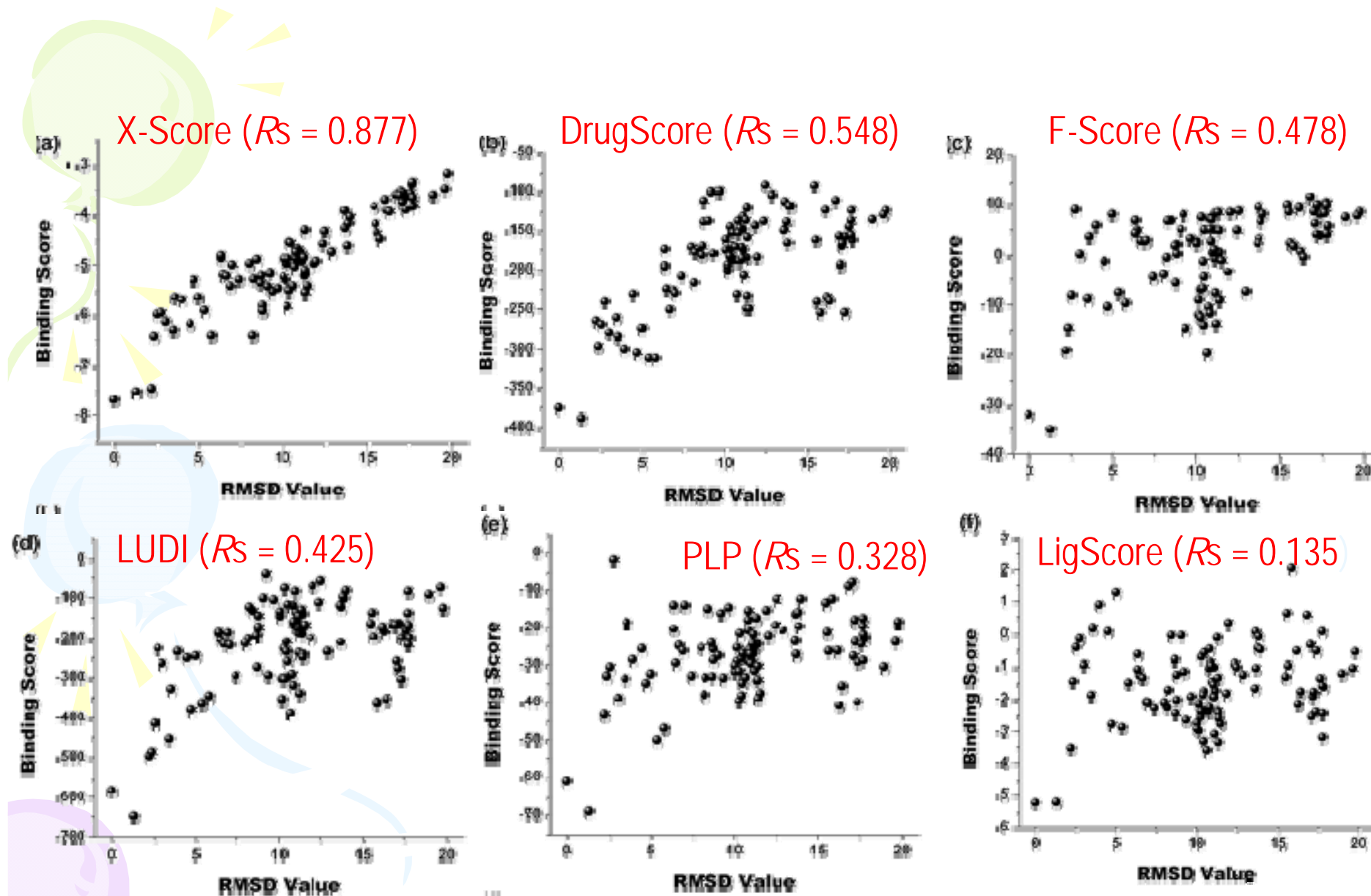

$$(x_{CM}, y_{CM}, z_{CM}, \phi, \theta, \psi, \chi_1, \chi_2, \cdots, \chi_k)$$

The Tacit Variational Principle:

$$\Delta G^{\exp} \cong \Delta G^{pred}\left[\rho^*(r)\right] \leq \Delta G^{pred}\left[\rho(r)\right]$$

X-Score (Rs = 0.877) DrugScore (Rs = 0.548) F-Score (Rs = 0.478)

LUDI (Rs = 0.425) PLP (Rs = 0.328) LigScore (Rs = 0.135)

Correlations between rmsd values (Å) and binding scores of the 101 docked conformations of PDB entry 1CBX .

RX Wang et al. J. Med. Chem. 46, 2287-2303 (2003)

Designing the Scoring Function

- Avoid steric contacts (where shape complementarity is the major factor for docking).

- Take chemical nature of each atom into account. Force field-based scoring.

- Map scoring function to binding free energies, if possible.

- Reproduce X-ray crystallographic or NMR results.

Automated Docking with Grid-Based Energy Evaluation

Elaine C. Meng, Brian K. Shoichet, and Irwin D. Kuntz*

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143-0446

The ability to generate feasible binding orientations of a small molecule within a site of known structure is important for ligand design. We present a method that combines a rapid, geometric docking algorithm with the evaluation of molecular mechanics interaction energies. The computational costs of evaluation are minimal because we precalculate the receptor-dependent terms in the potential function at points on a three-dimensional grid. In four test cases where the components of crystallographically determined complexes are redocked, the "force field" score correctly identifies the family of orientations closest to the experimental binding geometry. Scoring functions that consider only steric factors or only electrostatic factors are less successful. The force field function will play an important role in our efforts to search databases for potential lead compounds.
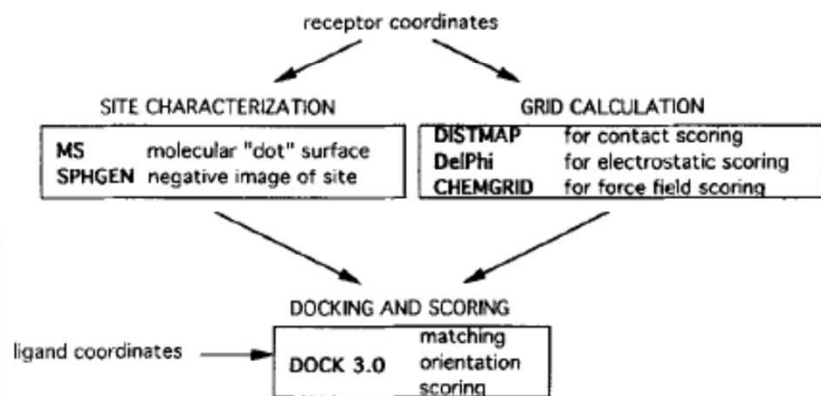
**Figure 1.** Programs involved in the use of DOCK. MS[18,19] and DelPhi[20,21] are distributed independently.

nealing approach. Our procedure is not carried out within a representation of Cartesian space; instead, it depends only on internal distance matching. Thus, there is no dependence on the starting locations of the molecules, and there are no effects due to steric hindrance or unfavorable charge–charge interactions en route to the site. Molecules may be docked successfully even when there is no low-energy pathway from the outside of the protein to the binding site.

Times cited: 956 (2018/3/17)

5

Dock Scoring Functions

- ## Contact Scoring
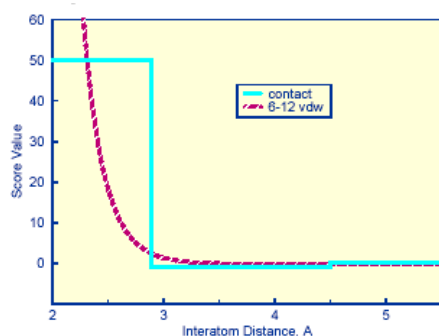
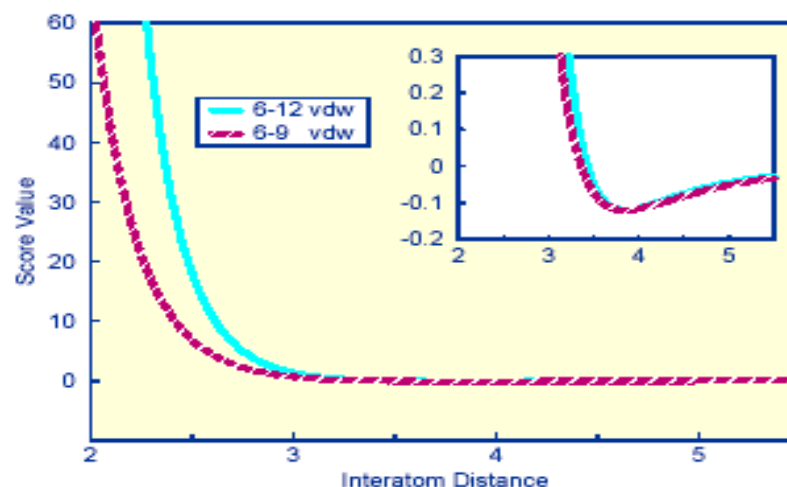|	
E	Intermolecular interaction energy
r_{ij}	Distance between atoms i and j
A_{ij} and B_{ij}	van der Waals repulsion and attraction parameters
a and b	van der Waals repulsion and attraction exponents
q_i and q_j	Point charges on atoms i and j
D	Dielectric function
332	Factor to convert electrostatic energy to kcal/mol.

- ## Energy Scoring

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_i q_j}{D r_{ij}} \right)$$
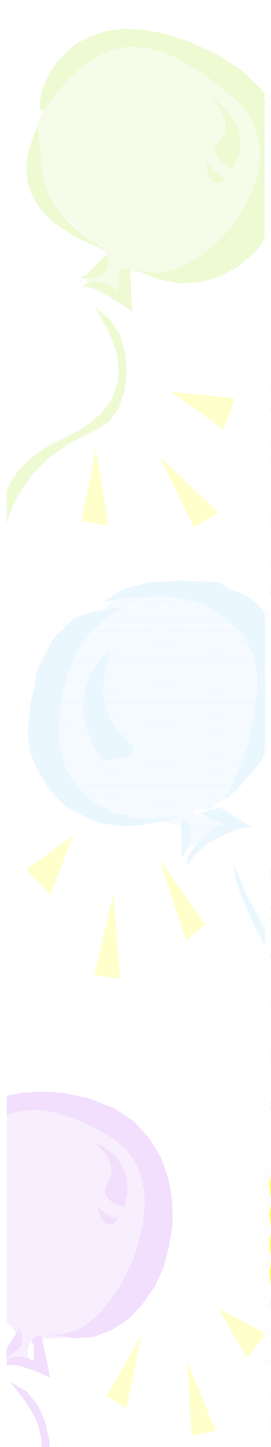

The radius and well depth used to generate this figure (1.925 A and 0.12 kcal/mol) represent a united-atom model of CH_2.

- ## Chemical Scoring

Based on energy score.
- To enhance recognition of chemical complementarity.
- To incorporate qualitative aspects of solvation.

Meng et al., J. Comput. Chem. **13:** 505 (1992)

Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function

GARRETT M. MORRIS,[1] DAVID S. GOODSELL,[1]
ROBERT S. HALLIDAY,[2] RUTH HUEY,[1] WILLIAM E. HART,[3]
RICHARD K. BELEW,[4] ARTHUR J. OLSON[1]

[1] Department of Molecular Biology, MB-5, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037-1000
[2] Hewlett-Packard, San Diego, California
[3] Applied Mathematics Department, Sandia National Laboratories, Albuqurque, NM
[4] Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA

ABSTRACT: A novel and robust automated docking method that predicts the bound conformations of flexible ligands to macromolecular targets has been developed and tested, in combination with a new scoring function that estimates the free energy change upon binding. Interestingly, this method applies a Lamarckian model of genetics, in which environmental adaptations of an individual's phenotype are reverse transcribed into its genotype and become heritable traits (*sic*). We consider three search methods, Monte Carlo simulated annealing, a traditional genetic algorithm, and the Lamarckian genetic algorithm, and compare their performance in dockings of seven protein–ligand test systems having known three-dimensional structure. We show that both the traditional and Lamarckian genetic algorithms can handle ligands with more degrees of freedom than the simulated annealing method used in earlier versions of AUTODOCK, and that the Lamarckian genetic algorithm is the most efficient, reliable, and successful of the three. The empirical free energy function was calibrated using a set of 30 structurally known protein–ligand complexes with experimentally determined binding constants. Linear regression analysis of the observed binding constants in terms of a wide variety of structure-derived molecular properties was performed. The final model had a residual standard error of 9.11 kJ mol^{-1} (2.177 kcal mol^{-1}) and was chosen as the new energy

7

AutoDock Scoring Function

- A free energy-based semi-empirical approach.

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{tor} + \Delta G_{sol}$$

$$\Delta G_{vdw} = W_{vdw} \times \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$$

$$\Delta G_{hbond} = W_{hbond} \times \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} - E_{hbond} \right)$$

$$\Delta G_{elec} = W_{elec} \times \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}$$

$$\Delta G_{tor} = W_{tor} \times N_{tor}$$

$$\Delta G_{sol} = W_{sol} \times \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{-r_{ij}^2 / 2\sigma^2}$$
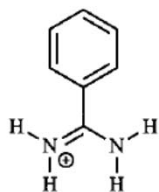
$$\Delta G_{obs} = RT \ln K_D$$

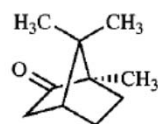|	
W_{vdw}	0.1485
W_{hbond}	0.0656
W_{elec}	0.1146
W_{tor}	0.3113
W_{sol}	0.1711

Morris *et al., J. Comput. Chem.* **19:** 1639-1662 (1998)
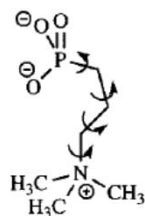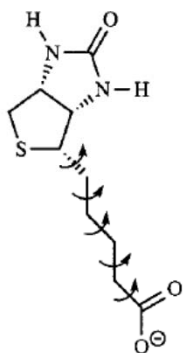
Times Cited: 8296 (2018/3/15)

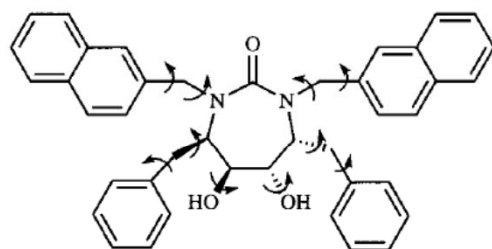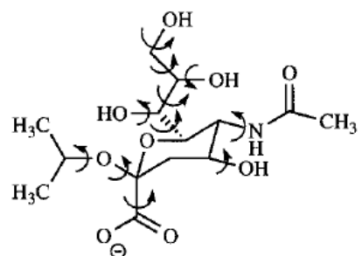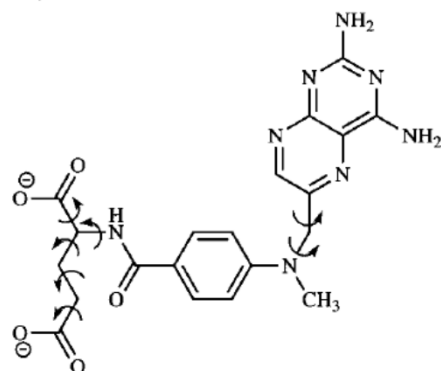(a) (b) (c) (d) (e) (f) (g)

Seven ligands as examples showing the rotatable bonds as curly arrows: (a) benzamidine; (b) camphor; (c) phosphocholine; (d) biotin; (e) HIV-1 protease inhibitor XK-263; (f) isopropylated sialic acid; and (g) methotrexate. Note that two ligands, (e) and (f), contain **hydroxyl rotors**, which are **not counted** in the total number of torsional degrees of freedom; note also that **cyclic rotatable bonds are excluded**.

9

TABLE I.

Protein–Ligand Complexes Used to Calibrate Empirical Free Energy Function, Along with Brookhaven Protein Data Bank (PDB) Accession Codes and Binding.

Protein–ligand complex	PDB code	$\text{Log}(K_i)$[a]
Concanavalin A / α-methyl-D-mannopyranoside	4cna	2.00
Carboxypeptidase A / glycyl-L-tyrosine	3cpa	3.88
Carboxypeptidase A / phosphonate ZAA=P=(O)F	6cpa	11.52
Cytochrome P-450$_{cam}$ / camphor	2cpp	6.07
Dihydrofolate reductase / methotrexate	4dfr	9.70
α-Thrombin / benzamidine	1dwb	2.92
Endothiapepsin / H-256	zer6	7.22
ε-Thrombin / MQPA	1etr	7.40
ε-Thrombin / NAPAP	1ets	8.52
ε-Thrombin / 4-TAPAP	1ett	6.19
FK506-binding protein (FKBP) / immunosuppressant FK506	1fkf	9.70
D-Galactose / D-glucose binding protein / galactose	2gbp	7.60
Hemagglutinin / sialic acid	4hmg	2.55
HIV-1 Protease / A78791	1hvj	10.46
HIV-1 Protease / MVT101	4hvp	6.15
HIV-1 Protease / acylpepstatine	5hvp	5.96
HIV-1 Protease / XK263	1hvr	9.51
Fatty-acid-binding protein / C$_{15}$COOH	2ifb	5.43
Myoglobin (ferric) / imidazole	1mbi	1.88
McPC603 / phosphocholine	2mcp	5.23
β-Trypsin / benzamidine	3ptb	4.74
Retinol-binding protein / retinol	1rbp	6.72
Thermolysin / Leu-hydroxylamine	4tln	3.72
Thermolysin / phosphoramidon	1tlp	7.55
Thermolysin / n-(1-carboxy-3-phenylpropyl)-Leu-Trp	1tmn	7.30
Thermolysin / Cbz-Phe-p-Leu-Ala (ZFpLA)	4tmn	10.19
Thermolysin / Cbz-Gly-p-Leu-Leu (ZGpLL)	5tmn	8.04
Purine nucleoside phosphorylase (PNP) / guanine	1ulb	5.30
Xylose isomerase / CB3717	2xis	5.82
Triose phosphate isomerase (TIM) / 2-phosphoglycolic acid (PGA)	2ypi	4.82

[a] Adapted from Böhm.[54]

Determining the optimal weighting factors

Calibration of Empirical Free Energy Function.

Model[a]	Residual standard error	Multiple R^2	ΔG_{vdw}[b]	ΔG_{estat}	ΔG_{hbond}	ΔG_{tor}	ΔG_{solv}
A	2.324	0.9498	0.1795 (0.0263)	0.1133 (0.0324)	0.0166 (0.0625)	0.3100 (0.0873)	0.0101 (0.0585)
B	2.232	0.9537	0.1518 (0.0269)	0.1186 (0.0246)	0.0126 (0.0382)	0.3548 (0.0890)	0.1539 (0.1050)
C	2.177	0.9559	0.1485 (0.0237)	0.1146 (0.0238)	0.0656 (0.0558)	0.3113 (0.0910)	0.1711 (0.1035)

[a] Models differ in the formulation of the solvation term and the hydrogen bonding term. Model A: full volume-based solvation term and standard 10–12 hydrogen bonding, as in Eq. (1). Model B: apolar ligand atoms only in the solvation term, and standard 10–12 hydrogen bonding. Model C: apolar ligand atoms only in the solvation term, and the standard 10–12 hydrogen less the estimated average, as in Eq. (2).
[b] Values for the model coefficients, with standard deviations in parentheses.

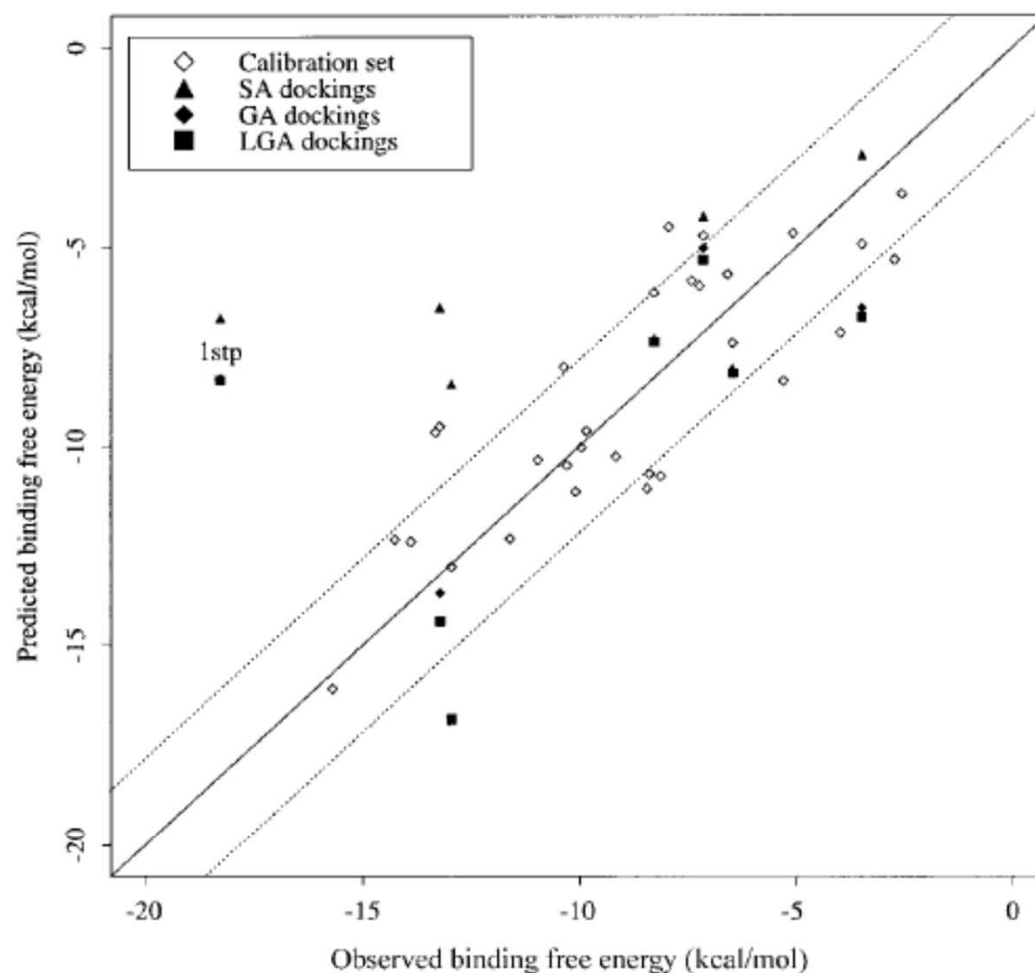*J. Comput. Chem.* **19:** 1639-1662 (1998)

FIGURE 3. Predicted versus observed binding free energies for the calibration set and the docking tests. The solid line shows a perfect fit, and the dotted lines show one standard deviation above and below this. Hollow diamonds show the 30 protein–ligand complexes used in fitting the terms of the binding free energy function. Solid triangles show the results of the simulated annealing (SA) dockings, solid diamonds show the genetic algorithm (GA) dockings, and the solid squares show the Lamarckian genetic algorithm (LGA) dockings. Note the outlying biotin–streptavidin complex (1stp), where it is believed there are significant contributions to the binding free energy due to protein rearrangements.
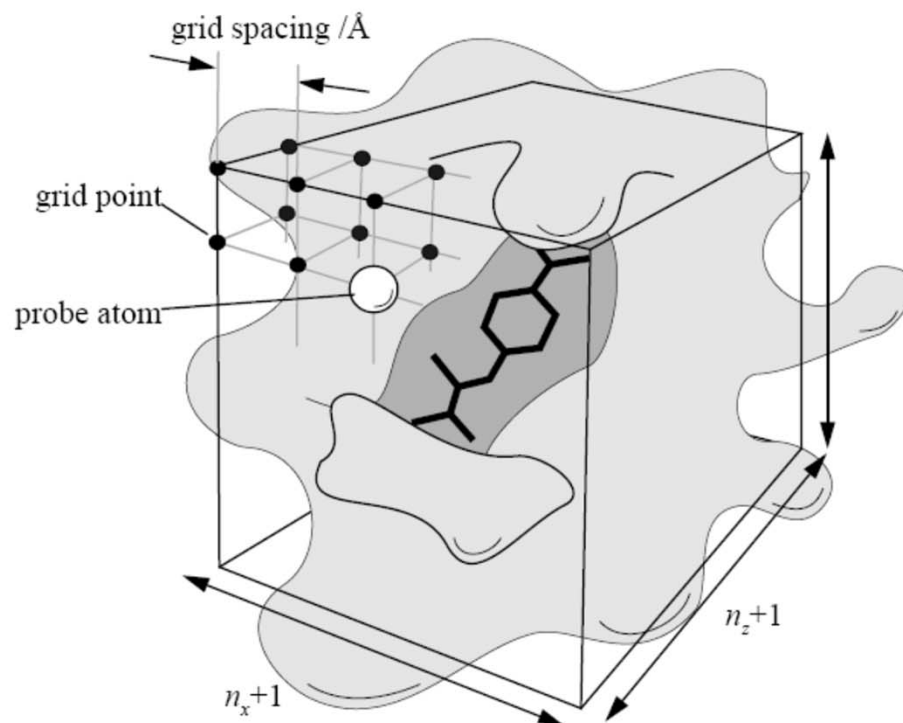
Hydrogen Bond
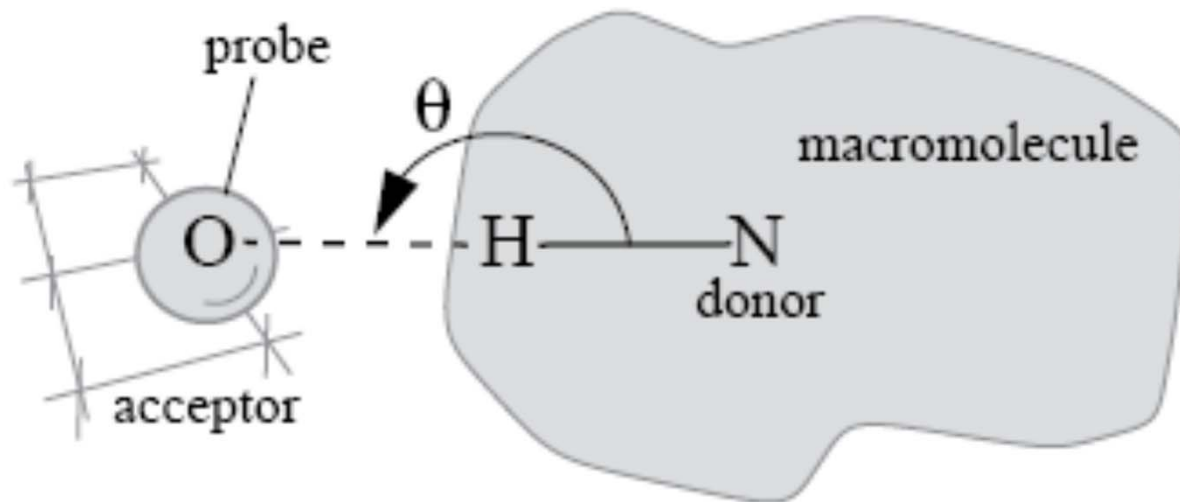
# Van der Waals Potential Energy
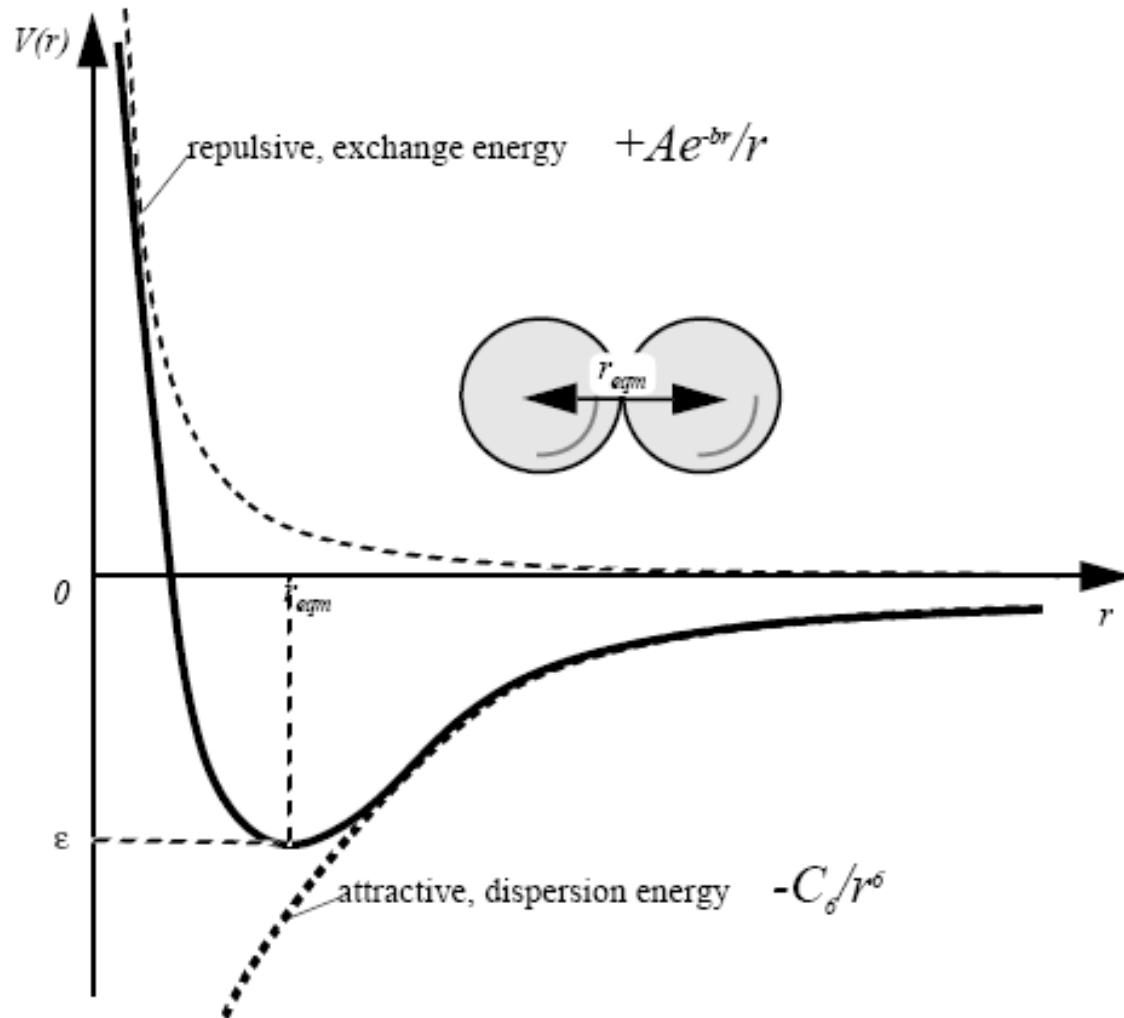
Grid Maps

A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules

P. J. Goodford

The Laboratory of Molecular Biophysics, The Rex Richards Building, University of Oxford, Oxford OX1 3QU, England. Received August 3, 1984

The interaction of a probe group with a protein of known structure is computed at sample positions throughout and around the macromolecule, giving an array of energy values. The probes include water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl. Contour surfaces at appropriate energy levels are calculated for each probe and displayed by computer graphics together with the protein structure. Contours at negative energy levels delineate regions of attraction between probe and protein and are found at known ligand binding clefts in particular. The contours also enable other regions of attraction to be identified and facilitate the interpretation of protein–ligand energetics. They may, therefore, be of value for drug design.

Torsional Energy

looking down the B-C bond
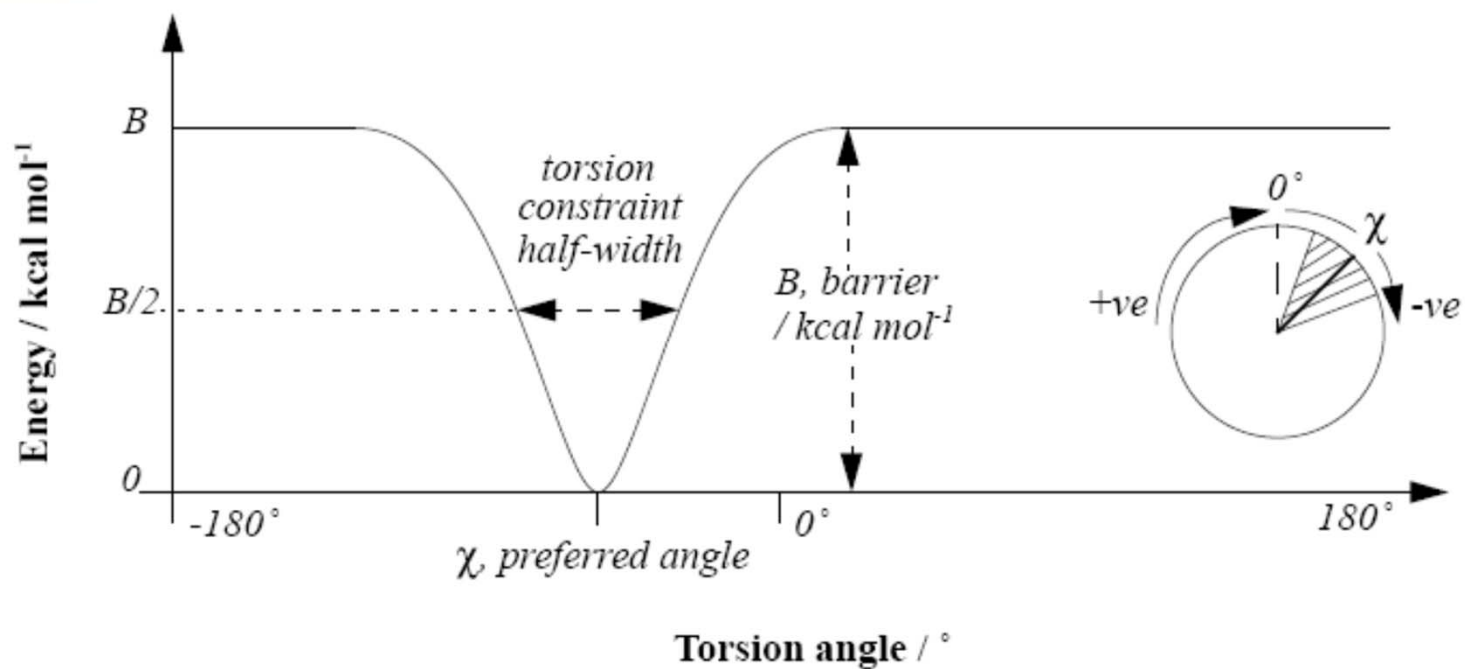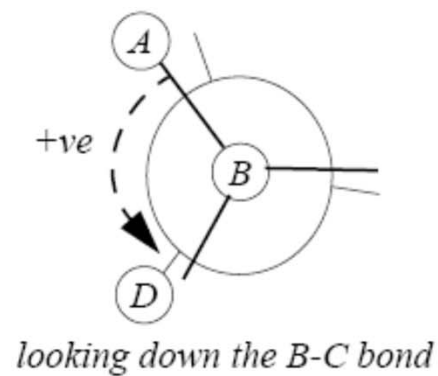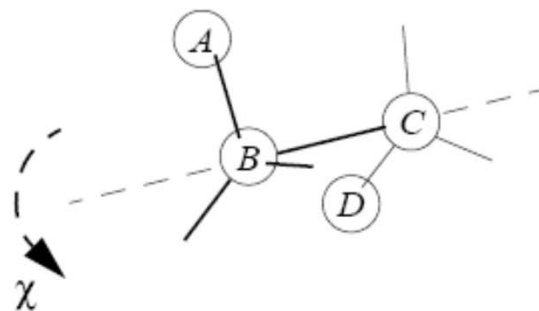
Prediction of binding pose could be challenging:
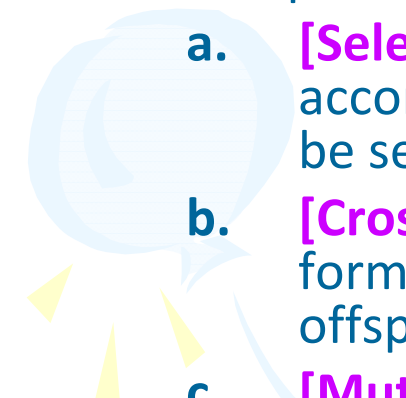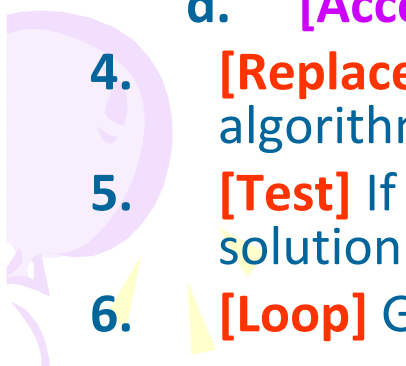## the case of P-glycoprotein with Paclitaxel (Taxol)
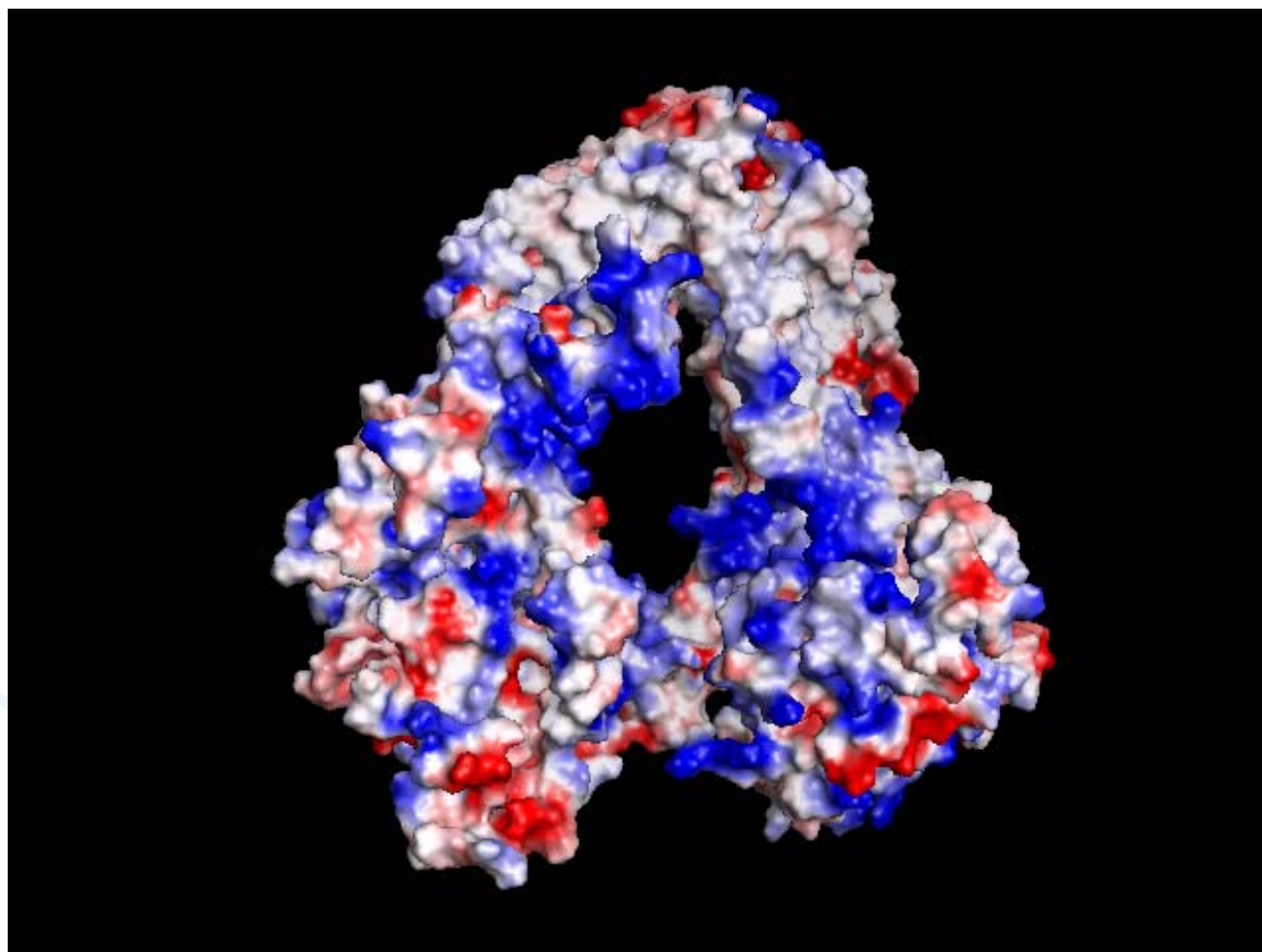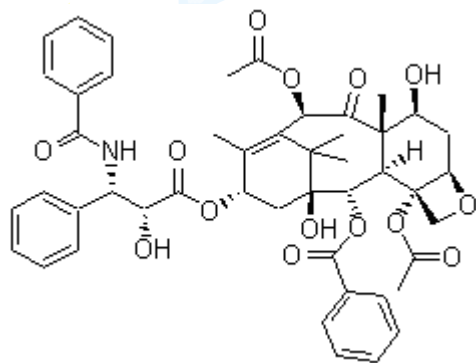
Characteristics of Biological Complex Problems

- The potential energy function is extremely rugged.
- The potential energy surface is usually highly asymmetric.
- The true global minimum is often surrounded by many deceptive local minima.
- The biological complex problems are mostly in the space of high dimensionality.
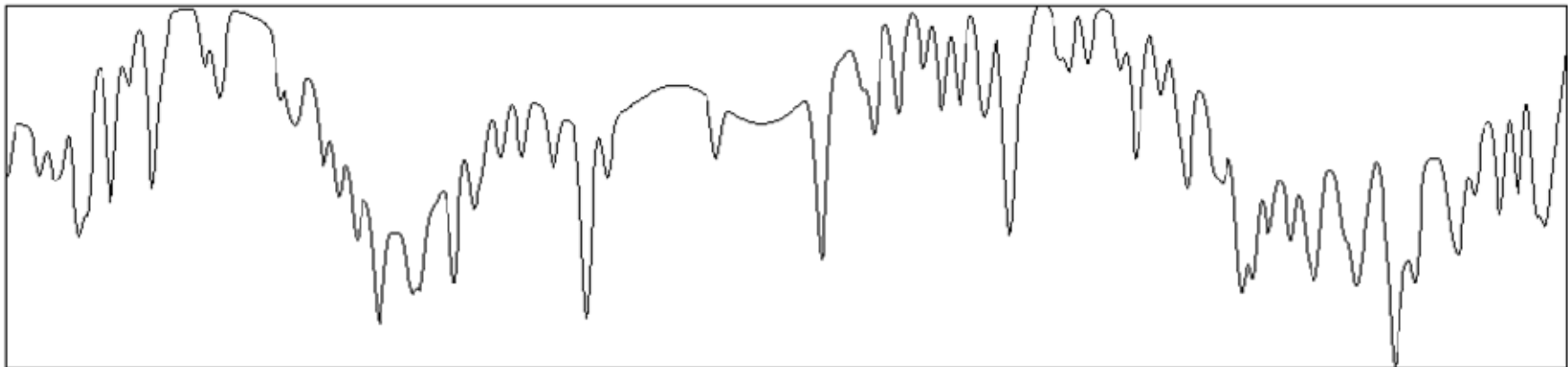
Genetic Algorithm

1. **[Start]** Generate random population of *n* chromosomes (suitable solutions for the problem)
2. **[Fitness]** Evaluate the fitness *f(x)* of each chromosome *x* in the population
3. **[New population]** Create a new population by repeating following steps until the new population is complete
   a. **[Selection]** Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected)
   b. **[Crossover]** With a crossover probability cross over the parents to form new offspring (children). If no crossover was performed, offspring is the exact copy of parents.
   c. **[Mutation]** With a mutation probability mutate new offspring at each locus (position in chromosome).
   d. **[Accepting]** Place new offspring in the new population
4. **[Replace]** Use new generated population for a further run of the algorithm
5. **[Test]** If the end condition is satisfied, **stop**, and return the best solution in current population
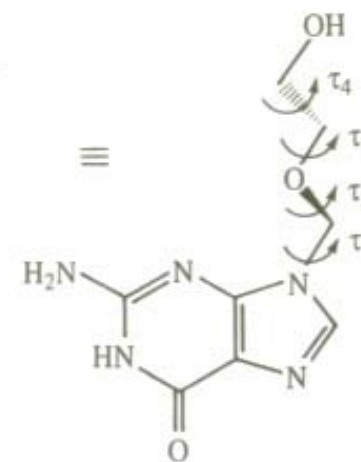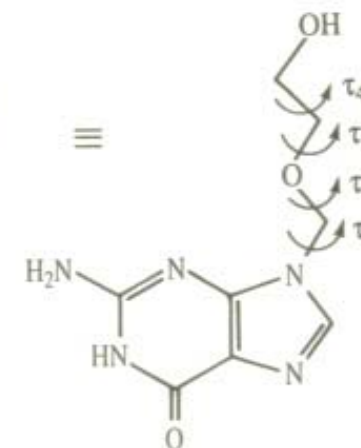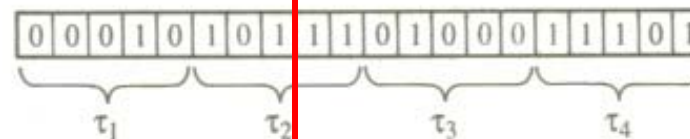6. **[Loop]** Go to step **2**

"Chromosomes" for Flexible-Ligand Docking
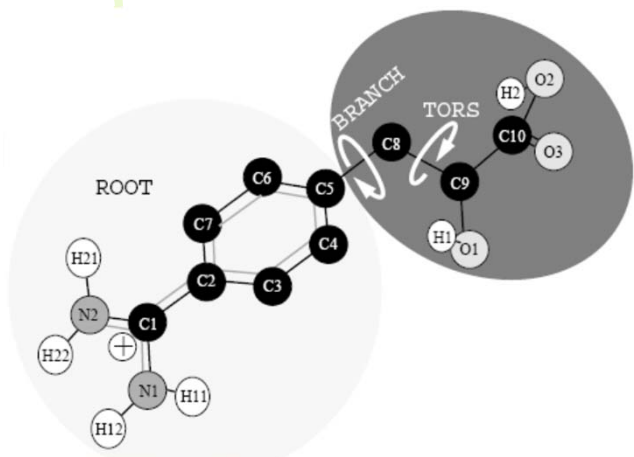
$$(x_{CM}, y_{CM}, z_{CM}, \phi, \theta, \psi, \chi_1, \chi_2, \cdots, \chi_k)$$
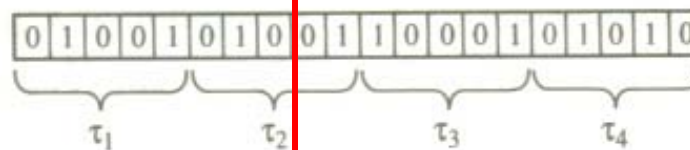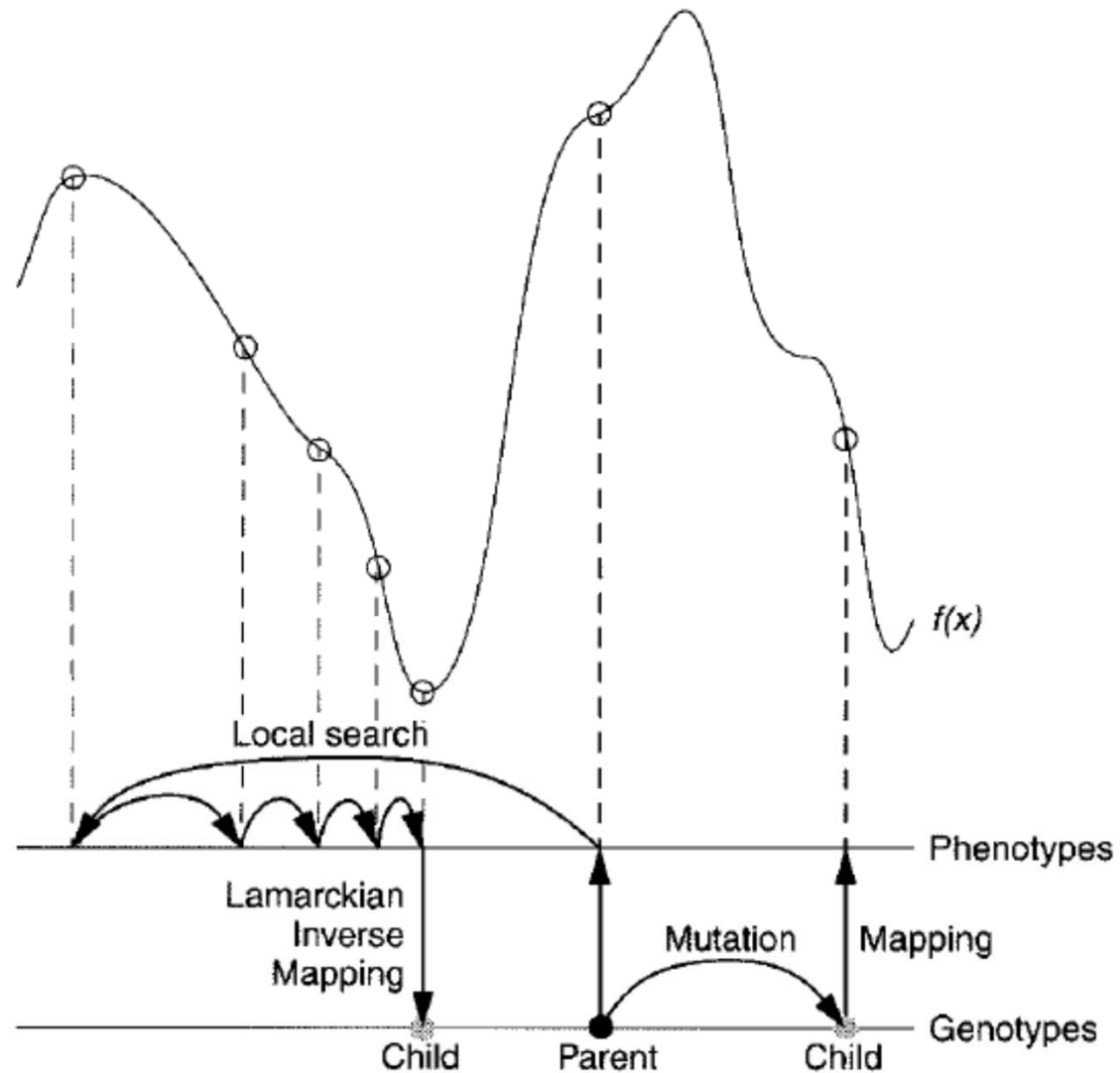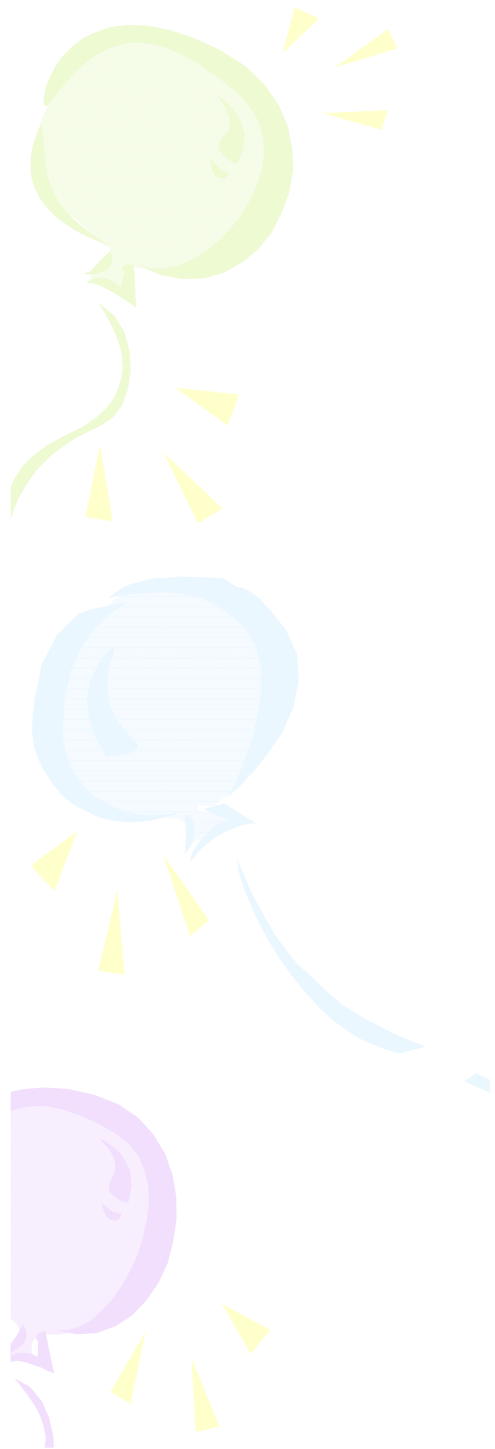
Crossover operation

Leach (2001)

FIGURE 1. This figure illustrates genotypic and phenotypic search, and contrasts Darwinian and Lamarckian search.[27] The space of the genotypes is

21

The Evolutionary Gaussians Algorithm

- n individuals, denoted by $s_1, s_2, ..., s_n$, are generated. Each s_i is a vector corresponding to a point in the domain of the objective function f. In order to achieve a scale-free representation, each component of s_i is linearly mapped to the numerical range of [0,1].

- The individuals in each generation of population are then sorted in the ascending order based on the values of the energy function on evaluated on these individuals. Let $t_1, t_2, ... t_n$ denote the ordered individuals and we have $f(t_1) < f(t_2) < f(t_n)$.

- n Gaussian distributions, denoted by $G_1, G_2, ... G_n$, are generated before the new generation of population is created. The center of each Gaussian distribution is selected randomly and independently from $t_1, t_2, ... t_n$, where the probability is not uniform but instead follows a discrete diminishing distribution, $n : n\text{-}1 : ... : 1$.

$$G_i(\mathbf{x}) = \left(\frac{1}{\sqrt{2\pi} \cdot \sigma_i} \right) \exp\left(-\frac{(\mathbf{x} - \mathbf{t}_k)^2}{2\sigma_i^2} \right)$$

$$\sigma_i^2 = \alpha + \frac{(\beta - \alpha)(k-1)}{n-1}$$
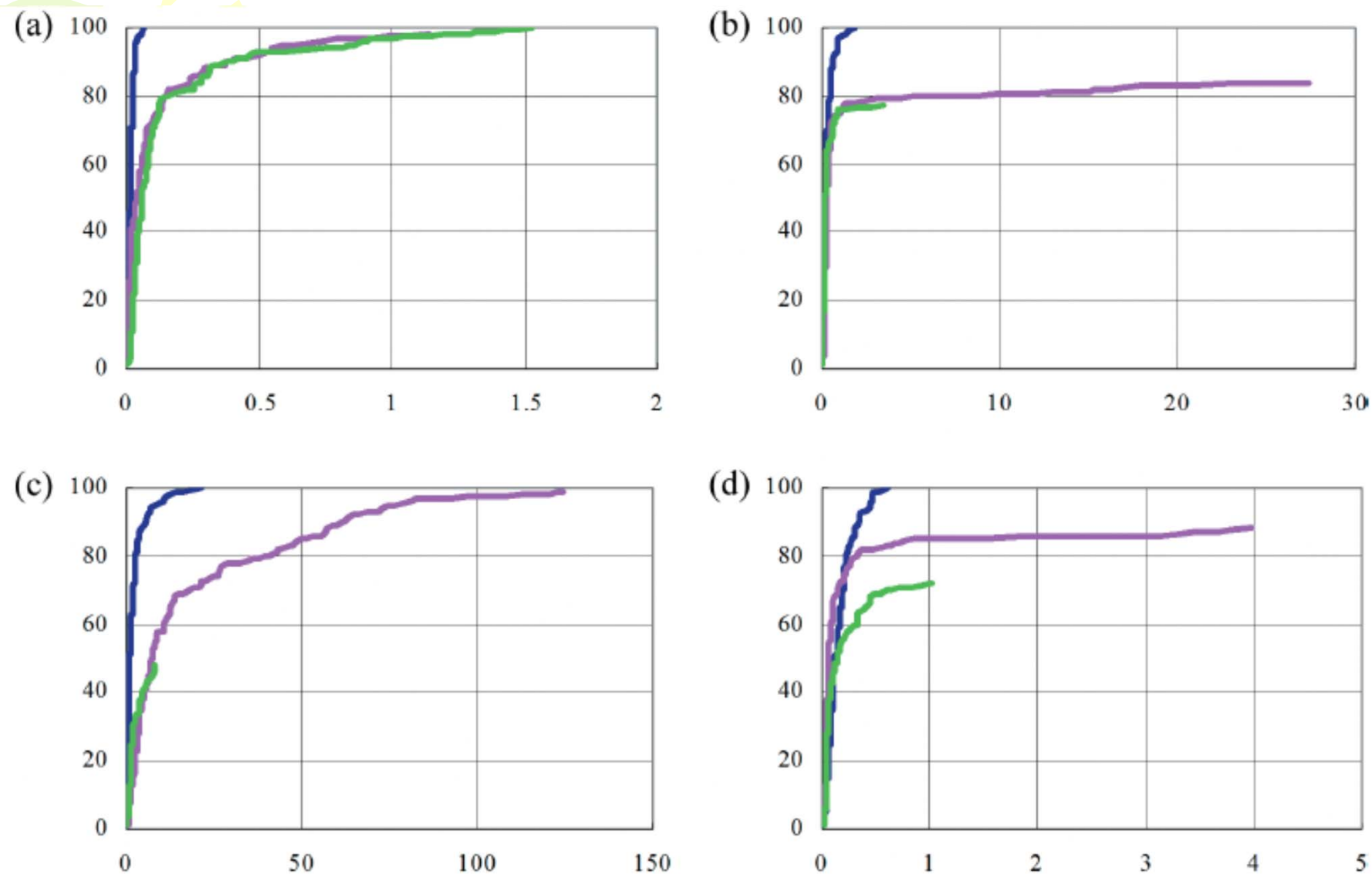
Figure 3. Number of runs to reach convergence versus the number of energy evaluations consumed (in units of 10^7): blue, MEDock results; magenta, LGA results (with parameters tuned); green, LGA results (with default parameters). (a) HIV-II protease complexed with its inhibitor L-735,524 (PDB ID: 1HSH); (b) FKBP-FK506, an immunophilin-immunosuppressant complex (PDB ID: 1FKF); (c) complex formed between phospholipase A2 and aspirin (PDB ID: 1OXR); and (d) TATA-box binding protein (YTBP) complexed with DNA containing a TATA-box (PDB ID: 1YTB).

Chang *et al. Nucleic Acids Research* 33: W233-W238 (2005)

A challenging scenario for most searching algorithms
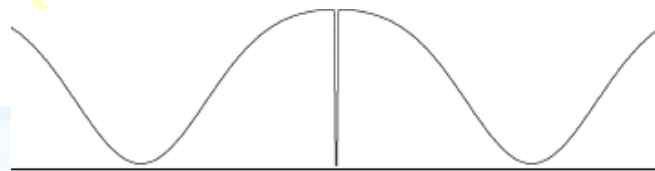
$$f(\mathbf{v}) = \sum_{i=1}^{21} \frac{w_i}{\left(\sqrt{2\pi}\cdot\sigma_i\right)^{20}} \exp\left(-\frac{\left(\mathbf{v}-\mu_i\right)^2}{2\sigma_i^2}\right), \text{ where}$$

$$\mu_1 = (0.5\pm0.1, 0.5, 0.5, \cdots, 0.5),$$
$$\mu_2 = (0.5, 0.5\pm0.1, 0.5, \cdots, 0.5),$$
$$\vdots$$
$$\mu_{20} = (0.5, 0.5, 0.5, \cdots, 0.5\pm0.1),$$
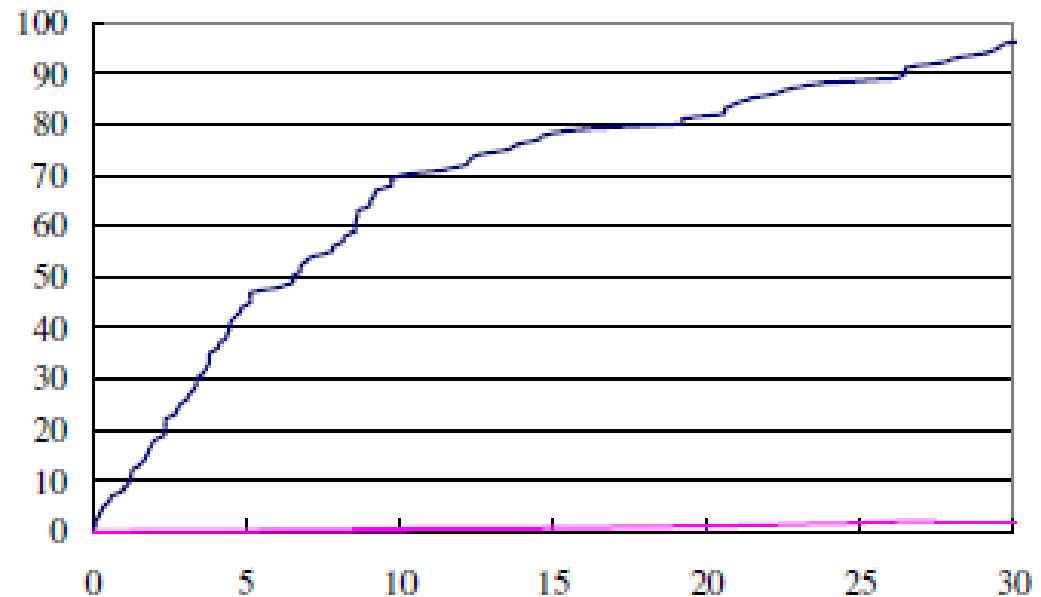$$\mu_{21} = (0.55, 0.55, 0.55, \cdots, 0.55),$$
$$\sigma_1 = \sigma_2 = \cdots = \sigma_{20} = 0.04, \sigma_{21} = 0.02$$
$$w_1 = w_2 = \cdots = w_{20} = -1, w_{21} = -0.55$$

Success rate of **EGA** versus **LGA** algorithm in locating the global minimum.

Chang *et al. Int J. Art. Intell. Tools* **19:** 267-280 (2010)

Design of benchmark energy functions in high dimensional space

Table 2. The ranges of the parameters in Eq. (2).

Parameter	Range
$\tilde{w}_i$	1.0
$\tilde{\sigma}_i$	$\dfrac{1}{\pi\sqrt[m]{k}}[0.9,1.1]$
$\tilde{c}_i$	$[0,1]$
$\hat{w}_i$	1/32
$\hat{\sigma}_i$	$\dfrac{1}{\pi\cdot\hat{k}^{1/m}}[0.9,1.1]$
$\hat{c}_i$	$[0,1]$
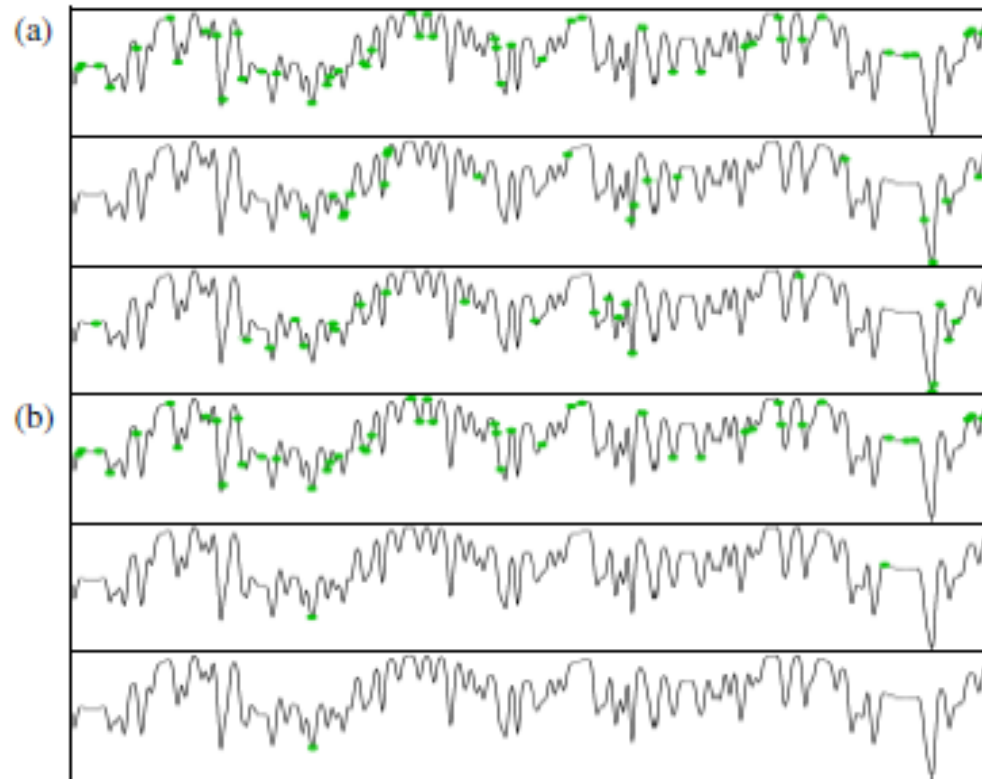
$$f(\mathbf{v})=\sum_{i=1}^{\tilde{k}}\frac{\tilde{w}_i}{\left(\sqrt{2\pi}\cdot\tilde{\sigma}_i\right)^m}\exp\left(-\frac{(\mathbf{v}-\tilde{\mathbf{c}}_i)^2}{2\tilde{\sigma}_i^2}\right)+\sum_{j=1}^{\hat{k}}\frac{\hat{w}_j}{\left(\sqrt{2\pi}\cdot\hat{\sigma}_j\right)^m}\exp\left(-\frac{(\mathbf{v}-\hat{\mathbf{c}}_j)^2}{2\hat{\sigma}_j^2}\right),$$

Dependence of searching efficiency on dimensionality
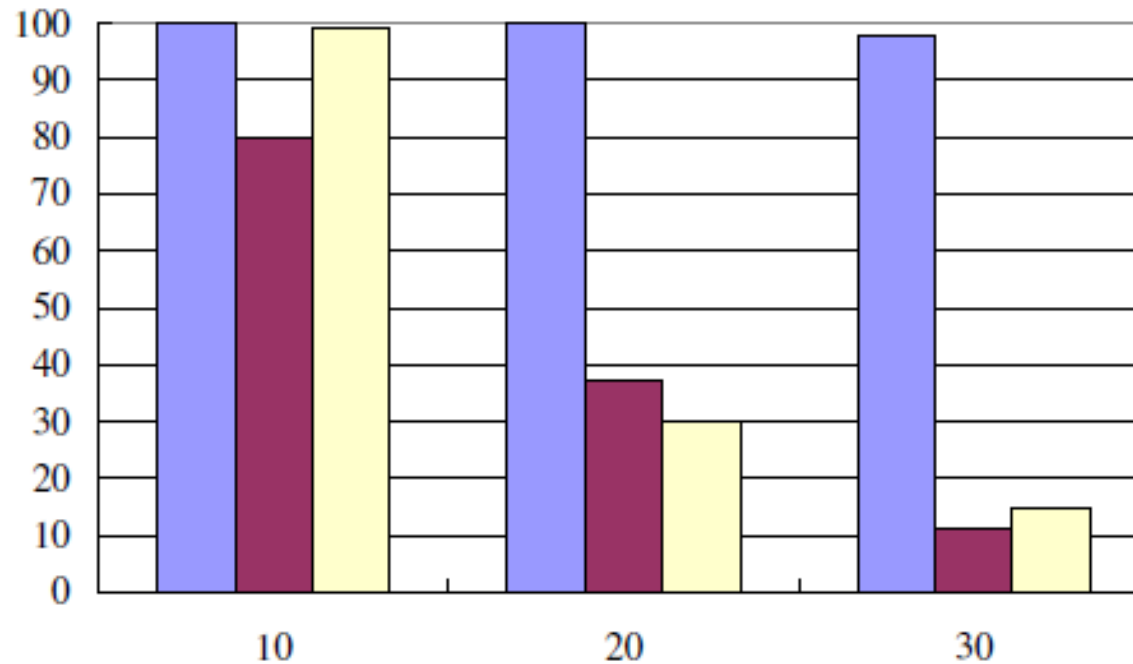
The x-axis corresponds to the dimension of the vector space and the y-axis corresponds to the numbers of simulation runs out of 100 independent runs in which the global minimum were successfully located by three different algorithms (EGA, LGA, RLGA).

The quest for a perfect scoring function

- The simplest forms of evaluating protein-ligand binding affinity are empirical scoring functions based on the quantitative structure-activity relationships (QSAR) approach pioneered by Hansch, or semi-empirical models with physics-based energetics.

- The physical-energy-based semi-empirical models have the advantages of easier rational interpretation of binding modes, and are more sensitive to protein conformational changes, which is important when protein dynamics and flexibility are to be accommodated.

X-Score

$$\text{HPScore} = C_0^1 + C_{vdw}^1 \cdot E_{vdw} + C_{RT}^1 \cdot N_{RT} + C_{HP} \cdot HP$$

$$\text{HMScore} = C_0^2 + C_{vdw}^2 \cdot E_{vdw} + C_{RT}^2 \cdot N_{RT} + C_{HM} \cdot HM$$

$$\text{HSScore} = C_0^3 + C_{vdw}^3 \cdot E_{vdw} + C_{RT}^3 \cdot N_{RT} + C_{HS} \cdot HS$$

$$\text{X-Score} = \frac{\text{HPScore} + \text{HMScore} + \text{HSScore}}{3}$$

Functional form of AutoDock4 scoring function

$$\Delta G_{bind} = W_{vdw} \times \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$$

The Gasteiger charge model was used.

$$+ W_{H-bond} \times \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$$

$$+ W_{estat} \times \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \quad ,$$

The desolvation energy is accounted for by calculating the surrounding volume of an atom (V_i), weighted by the atomic solvation parameter (S_i) and an exponential term with a distance weighting factor σ (0.35Å in AutoDock4).

$$+ W_{desol} \times \sum_{i,j} \left(S_i V_j + S_j V_i \right) e^{\left(-r_{ij}^2 / 2\sigma^2 \right)}$$

$$+ W_{tor} \times N_{tors}$$

$$S_i = (ASP_k + QASP \times |q_i|), \quad k = \text{C,A,N,O,S,H}$$

Huey *et al.*, *J. of Comput. Chem.* **28**: 1145-1152 (2007)

Times Cited: 1540 (2018/3/15)

Are all the charge models similar?

- The atomic charges of AutoDock4 were prepared by using the Gasteiger charge model, whose primary advantages lie in its simplicity and speed, but the charge determined by this model could be less accurate.

- For example, the dipole moment of the well-known polar molecule dimethyl sulfoxide (DMSO) in solution can be estimated as 4.7D according to its value in vacuum (3.96 D) in vacuum by experiment. The dipole moments estimated by different charge models are: 2.96 D (Gasteiger), 4.61 D (RESP) and 4.57 D (AM1-BCC), respectively.

Is it just to choose a good charge model?

- If RESP charges for ligands and parm99SB charges for protein were used with the original AutoDock4 scoring function, the root mean square of error (RMSE) for the LPDB data set will be 7.3 kcal/mol!
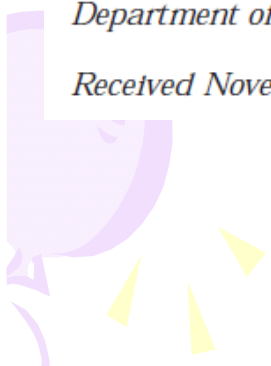
Ligand–Protein DataBase: Linking Protein–Ligand Complex Structures to Binding Data

Olivier Roche,[†] Ryuichi Kiyama,[‡] and Charles L. Brooks, III[*]

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received November 1, 2000

Ordinary least square (OLS) regression models with three charge combinations

OLS regression results of AutoDock4 scoring function with different charge models

Combinations			Coefficients of energetic terms					
Ligand	Protein	Size	W_{desolv}	W_{estat}	W_{hbond}	W_{tors}	W_{vdw}	RMSE
Gasteiger	Gasteiger	187	0.120	0.142	0.121	0.283	0.172	2.542
AM1-BCC	Amber99SB	187	0.093	0.125	0.077	0.279	0.167	2.523
RESP	Amber99SB	187	0.107	0.132	0.090	0.301	0.167	2.471

All RMSE values are in kcal/mol.

Calibrations were done with 187 complexes from LPDB.

Molecular cloning, characterization, and localization of a high-affinity serotonin receptor (5-HT$_7$) activating cAMP formation

(rat/Chinese hamster ovary cells/*in situ* hybridization)

MARTIAL RUAT*†, ELISABETH TRAIFFORT*, ROB LEURS*, JOEL TARDIVEL-LACOMBE*, JORGE DIAZ‡, JEAN-MICHEL ARRANG*, AND JEAN-CHARLES SCHWARTZ*

*Unite de Neurobiologie et Pharmacologie (U. 109) de l'Institut National de la Santé et de la Recherche Médicale, Centre Paul Broca, 2ter rue d'Alesia, 75014 Paris, France; and ‡Laboratoire de Physiologie, Faculte de Pharmacie, Universite Rene Descartes, 75006 Paris, France

K_I of metergoline = 63 $\pm$4 nM

Times cited: 579 (2018/3/15)

RTln(10)=1.38 kcal/mol for *T=300 K*

Molecular Cloning and Expression of a 5-Hydroxytryptamine$_7$ Serotonin Receptor Subtype*

Yong Shen‡, Frederick J. Monsma, Jr.‡, Mark A. Metcalf§, Pedro A. Jose¶, Mark W. Hamblin§‖, and David R. Sibley‡**

From the ‡Molecular Neuropharmacology Section, Experimental Therapeutics Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, the §Geriatric Research, Education, and Clinical Center, Seattle Veterans Affairs Medical Center and the Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington 98108, and the ¶Department of Pediatrics, Georgetown University School of Medicine, Washington, District of Columbia 20007

K_I of metergoline = 6.21$\pm$1 nM

Times cited: 546 (2018/3/15)

Both articles are in ACS Legacy Archive!

Expedited Articles

Synthesis and Antitumor Activity of Novel Water Soluble Derivatives of Camptothecin as Specific Inhibitors of Topoisomerase I

Michael J. Luzzio,[*,†] Jeffrey M. Besterman,[‡] David L. Emerson,[§] Michael G. Evans,[†] Karen Lackey,[†] Peter L. Leitner,[‡] Gordon McIntyre,[‡] Bradley Morton,[‡] Peter L. Myers,[†] Michael Peel,[†] Jay M. Sisco,[∥] Daniel D. Sternbach,[†] Wei-Qin Tong,[∥] Ann Truesdale,[‡] David E. Uehling,[†] Alan Vuong,[§] and Julie Yates[‡]

Departments of Medicinal Chemistry, Cell Biology, Pharmacology, and Pharmaceutics, Glaxo Research Institute, 5 Moore Drive, Research Triangle Park, North Carolina 27709

IC50 of campothecin = 679 nM

Times cited: 157 (2018/3/15)

$RTln(178)$=3.13 kcal/mol for *T=300 K*

Acylshikonin Analogues: Synthesis and Inhibition of DNA Topoisomerase-I

Byung-Zun Ahn,[*] Kyong-Up Baik, Gi-Ryang Kweon,[†] Kyu Lim,[†] and Byung-Doo Hwang[†]

Colleges of Pharmacy and Medicine, Chungnam National University, 220 Gungdong Youseongku, Taejon 305-764, Korea

IC50 of campotechin = 127 μM

Times cited: 139 (2018/3/15)

Ordinary least square (OLS) regression

Gauss, 1800

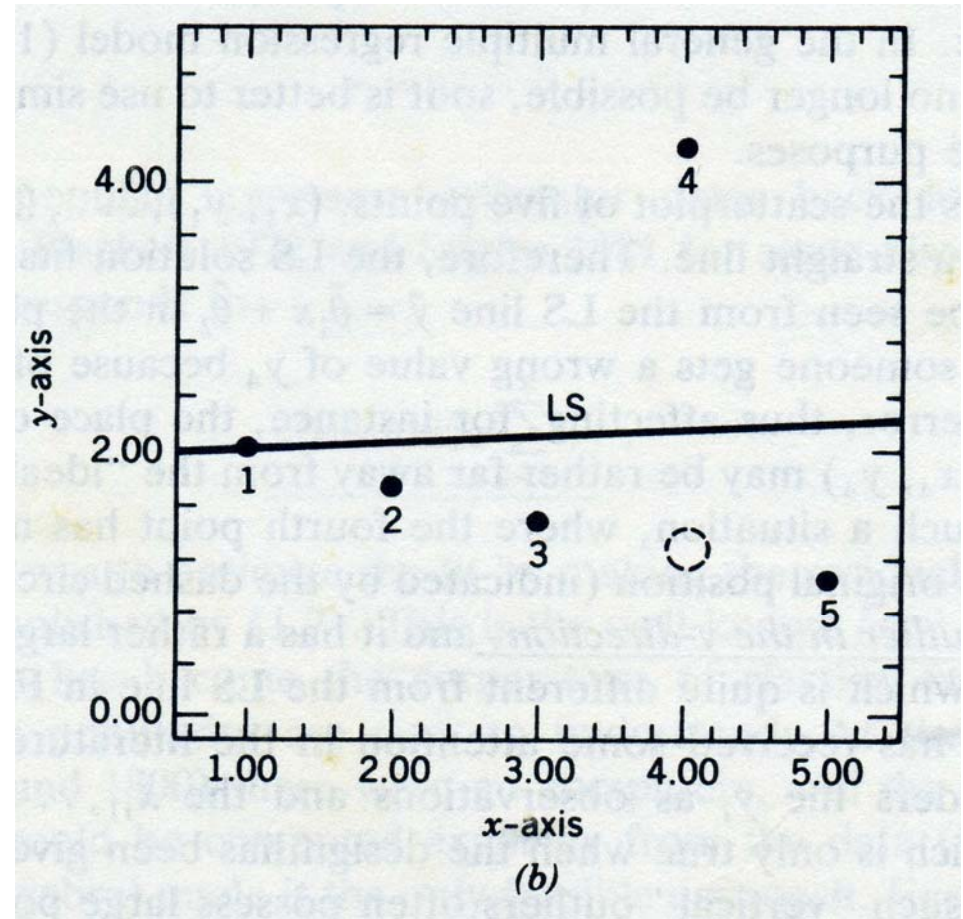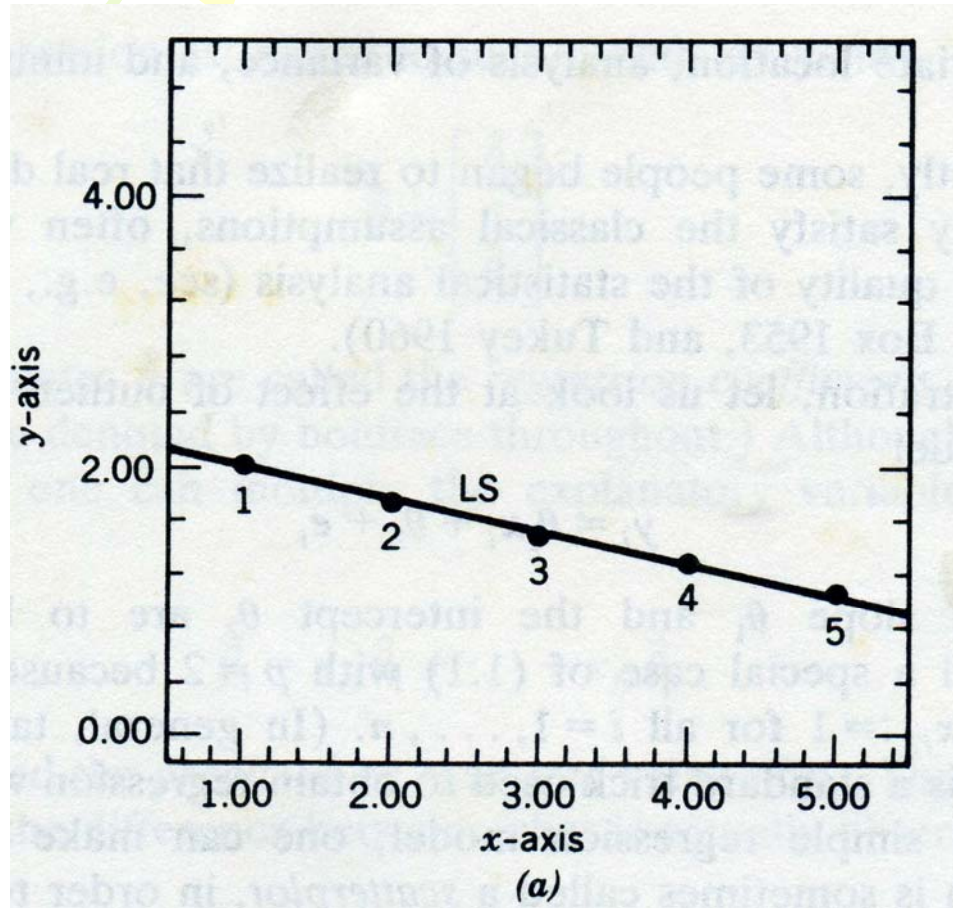$$y_i = w_1 x_{i1} + w_2 x_{i2} + \cdots + w_p x_{ip} + e_i$$

$$\mathbf{w} = \left(w_1, w_2, \cdots, w_p \right)$$

sample size

$$\underset{\mathbf{w}}{Minimize} \sum_{i=1}^{n} r_i^2$$

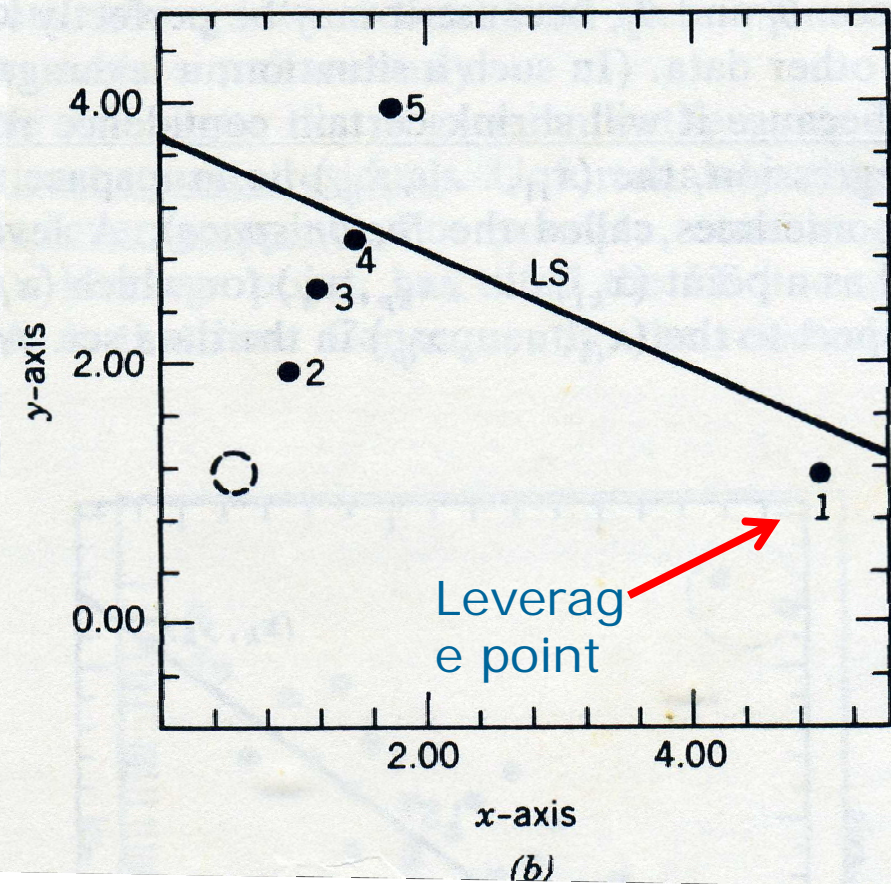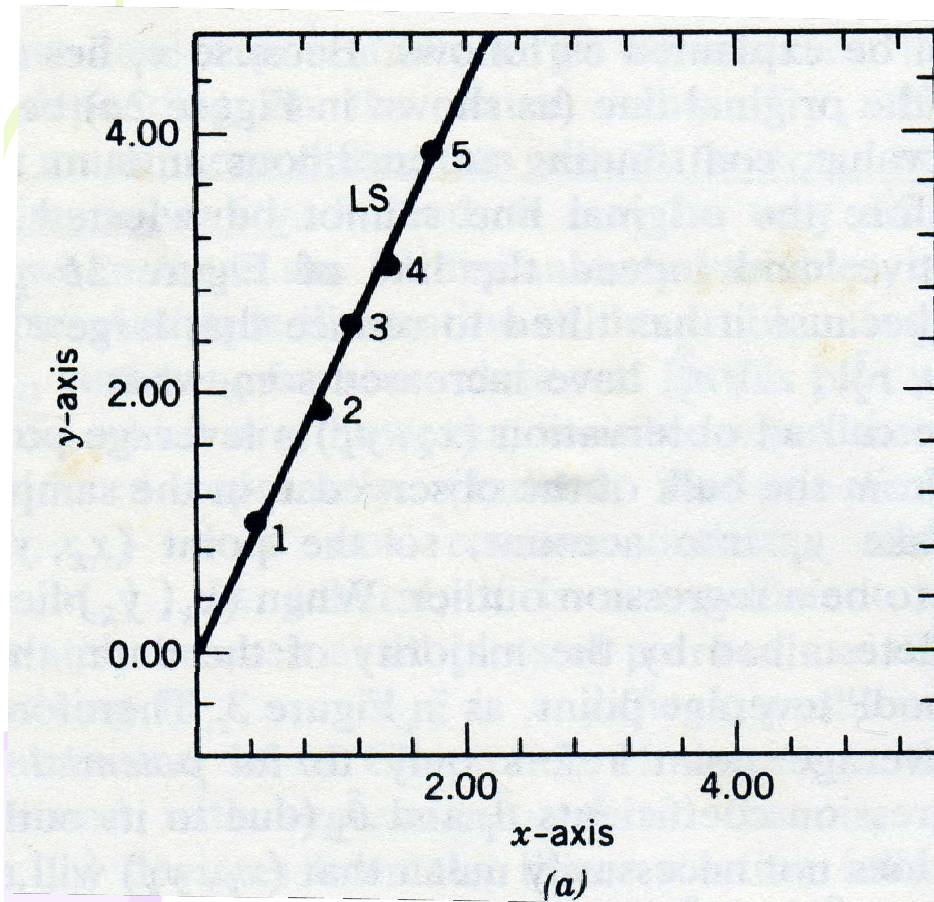*Summe d. Quadr. d. Differenz zw. beob. u. berechn. -> Min.*

p. 150

The problem of OLS regression

Rousseeuw (1985)

Outlier in the y-direction

The problem of OLS regression

Leverage point

Outlier in the x-direction

Rousseeuw (1985)
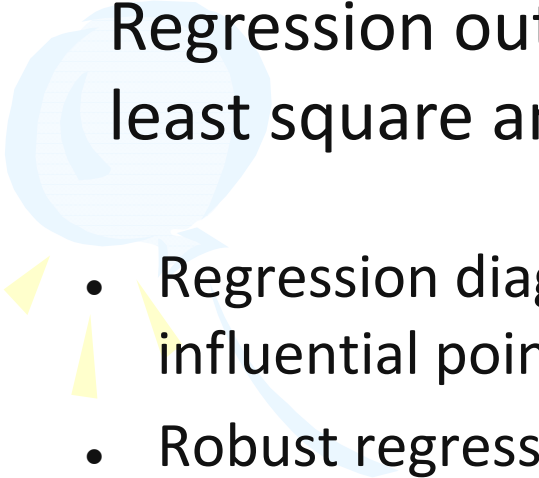
Regression diagnostics versus robust regression

Regression outliers pose a serious threat to standard least square analysis.

- Regression diagnostics: Use some quantity to pinpoint the influential points, remove the outliers, and then LS.

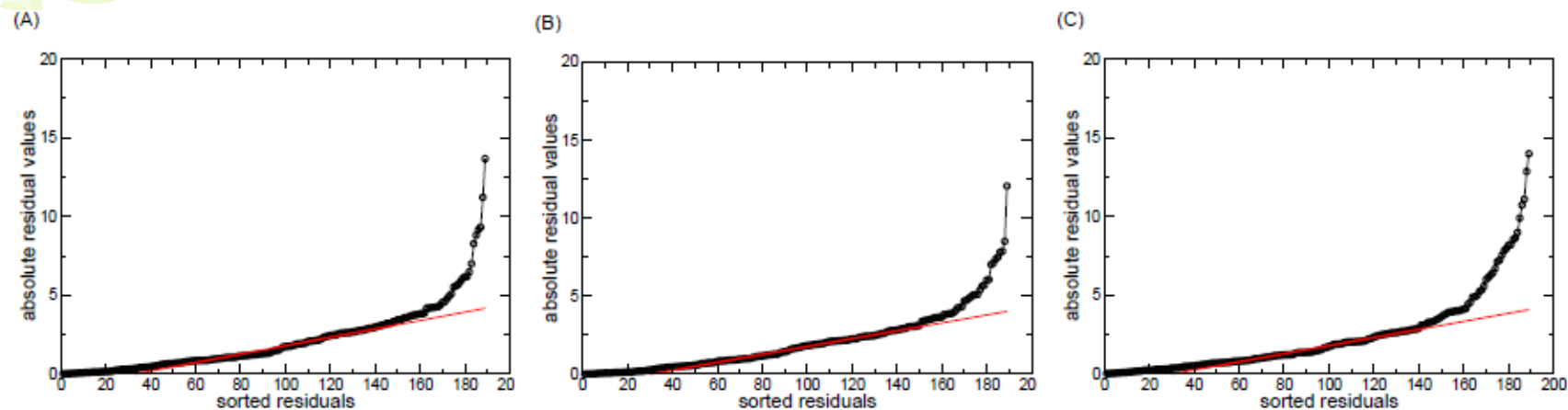- Robust regression: Devise estimators not so strongly affected by outliers. Fit to the majority of data.

Rousseeuw (1985)

Robust regression
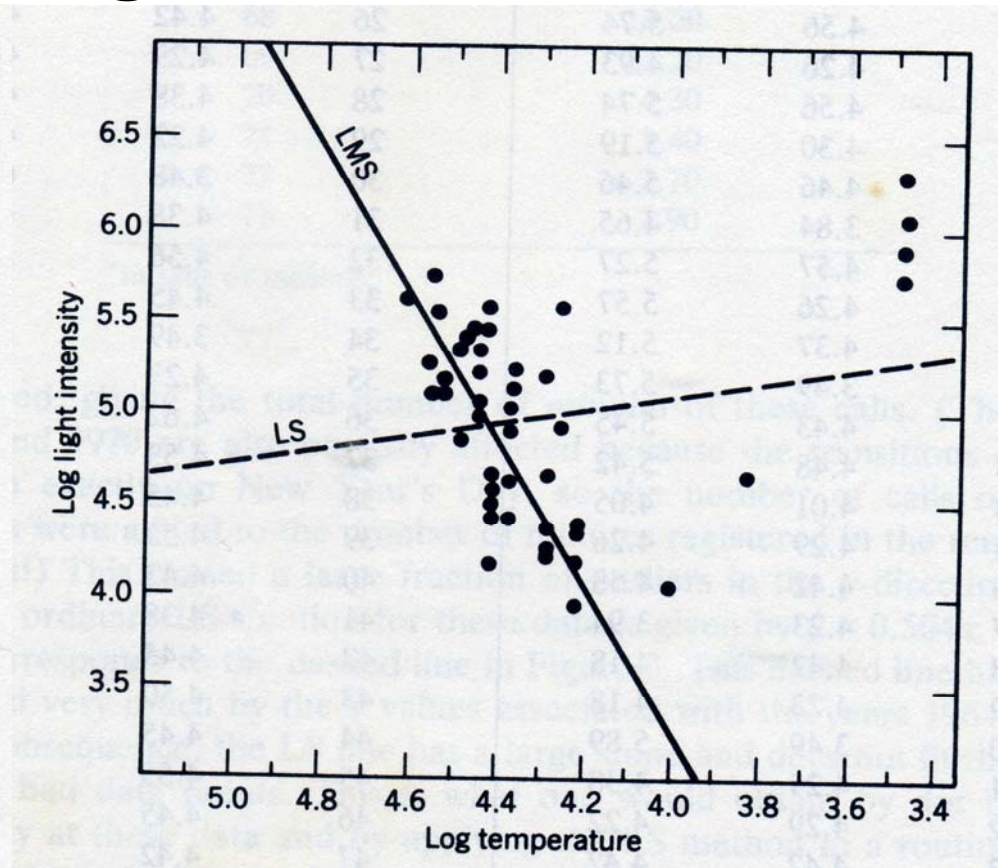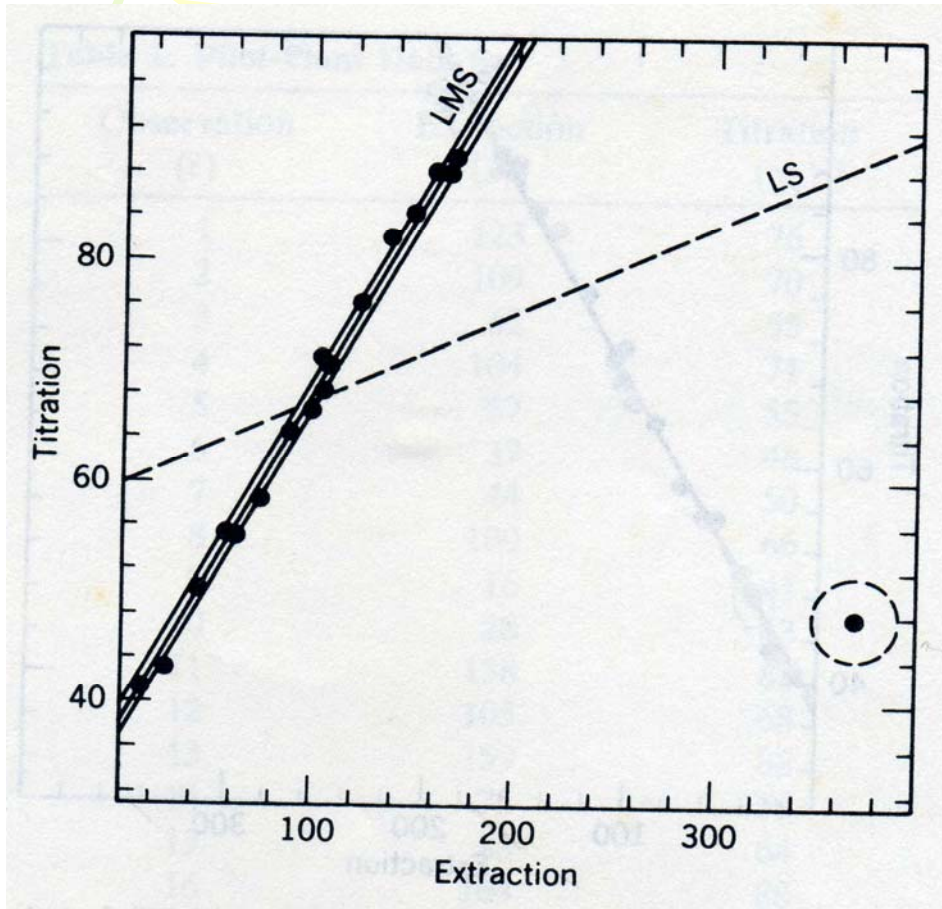
Building robust AutoDock4 scoring functions with quantum chemical charge models

Sorted absolute residuals based on robust regression analysis for the (A) RP and (B) AP and (C) GG charge combinations. Red lines are fitted to the residuals between top 25% and 75%.

Coefficients of the robust AutoDock4 scoring functions. The size of the clean set is 147

Models	Coefficients of different energetic terms					
	W_{desolv}	W_{estat}	W_{hbond}	W_{tors}	W_{vdw}	RMSE
AutoDock4[RGG]	0.0996	0.0241	0.1806	0.3594	0.1734	1.664
AutoDock4[RAP]	0.0993	0.0491	0.1565	0.3422	0.1736	1.637
AutoDock4[RRP]	0.0954	0.0661	0.1521	0.3618	0.1698	1.641

All RMSE values are in kcal/mol.

Cross-validations

leave-one-out cross-validation (LOO-CV)
leave-group-out cross-validation or Monte Carlo cross-validation (MCCV)

Scoring Function	LOO-CV		MCCV	
	S_{PRESSS}	q^2	S_{PRESS}	q^2
AutoDock4^{RGG}	1.732	0.675	1.782	0.657
AutoDock4^{RAP}	1.707	0.684	1.749	0.670
AutoDock4^{RRP}	1.711	0.683	1.755	0.668

$$S_{PRESS} = \sqrt{\frac{\sum_i \left(E_{i,\mathrm{pred.}} - E_{i,\mathrm{exp.}} \right)^2}{(N-k)}}$$

$$q^2 = 1 - \frac{\sum_i \left(E_{i,\mathrm{pred.}} - E_{i,\mathrm{exp.}} \right)^2}{\sum_i \left(E_{i,\mathrm{exp.}} - E_{i,\mathrm{mean}} \right)^2}$$

41

Wang et al., J. Chem. Inf. Model. **51:** 2528-2537 (2011)

Assessment with large external set of protein-ligand complexes

Performance of the robust AutoDock4 scoring functions and two other recent scoring functions tested with the PDBbind data sets

scoring function	N_{train}	N_{test}	R_p	R_s	SD	ME
AutoDock4^{RGG}	147	1427	0.604	0.615	1.61	1.26
AutoDock4^{RAP}	147	1427 v2009	0.606	0.617	1.60	1.25
AutoDock4^{RRP}	147	1427	0.595	0.610	1.62	1.26
original AutoDock4^{GG}	187	1427	0.562	0.594	1.66	1.31
sfc_290m	290	919	0.492	0.555		
sfc_229m	229	919 v2005	0.501	0.558		
sfc_frag	130	919	0.525	0.576		
PDSE-SVM	278	977 v2005	0.517	0.535	1.84	1.42

R_p: Pearson's correlation coefficient; R_s: Spearman's correlation coefficient
SD (standard error) and ME (mean error) are presented in the *pKd* unit. The binding free energy in kcal/mol at 298 K was converted to the *pKd* unit by dividing with the factor of -1.36.

Sotriffer *et al.*, *Proteins* **73:** 395-419 (2008)
Das *et al.*, *J. of Chemi. Inf. Model.* **50:** 298-308 (2010)
Wang *et al.*, *J. Chem. Inf. Model.* **51:** 2528-2537 (2011)

Assessment with a small and diverse external set

Performance of the robust AutoDock4 scoring functions and 17 scoring functions assessed in Cheng *et al*. on the 195 set of PDBbind v2007

(Cheng, T.J. *et al*., *J. Chem. Inf. Model.* 2009, 49, 1079-1093)

scoring function	N_{train}	N_{test}	R_p	R_s	SD
AutoDock4^{RGG}	147	195	0.654	0.696	1.81
AutoDock4^{RAP}	147	195	0.643	0.665	1.83
AutoDock4^{RRP}	147	194	0.615	0.634	1.88
original AutoDock4^{GG}	187	195	0.615	0.669	1.88
X-Score::HMScore			0.649	0.701	1.82
DrugScore^{CSD}			0.589	0.649	1.93
SYBYL::ChemScore			0.622	0.668	1.87
DS::PLP1			0.529	0.569	2.03
GOLD::ASP			0.518	0.558	2.04
SYBYL::G-Score			0.522	0.579	2.03
DS::LUDI3			0.477	0.478	2.10
DS::LigScore2			0.479	0.505	2.10
GlideScore-XP			0.555	0.556	2.01
DS::PMF			0.471	0.482	2.11
GOLD::ChemScore			0.528	0.553	2.05
by NHA			0.431	0.517	2.15
SYBYL::D-Score			0.388	0.443	2.20
DS::Jain			0.339	0.362	2.26
GOLD::GoldScore			0.329	0.386	2.26
SYBYL::PMF-Score			0.235	0.235	2.31
SYBYL::F-Score			0.238	0.208	2.31

43

Recognizing the "correct" binding poses from the decoys
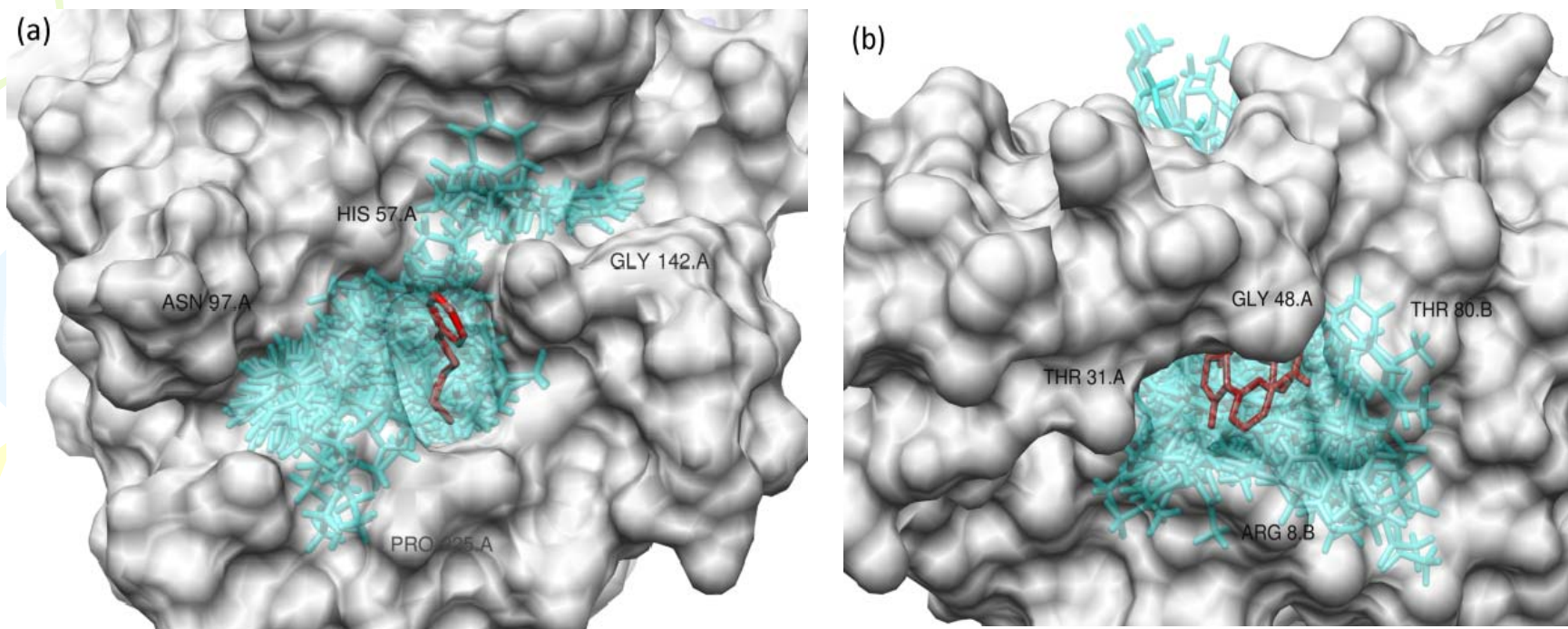
The decoys (cyan) and its complex with the native pose (red). (a) Trypsin bound with 4-phenylbutylamine (PDB ID: 1UTP; K_d: 36 mM). (b) HIV-1 protease bound with N-Aryl-oxazolidinone-5-carboxamides (PDB ID: 2I0D; K_i: 0.8 pM).

Wang, and Lin, *Curr. Pharm. Des. 19: 2174* (2013)

# Assessment of binding pose prediction with external decoys
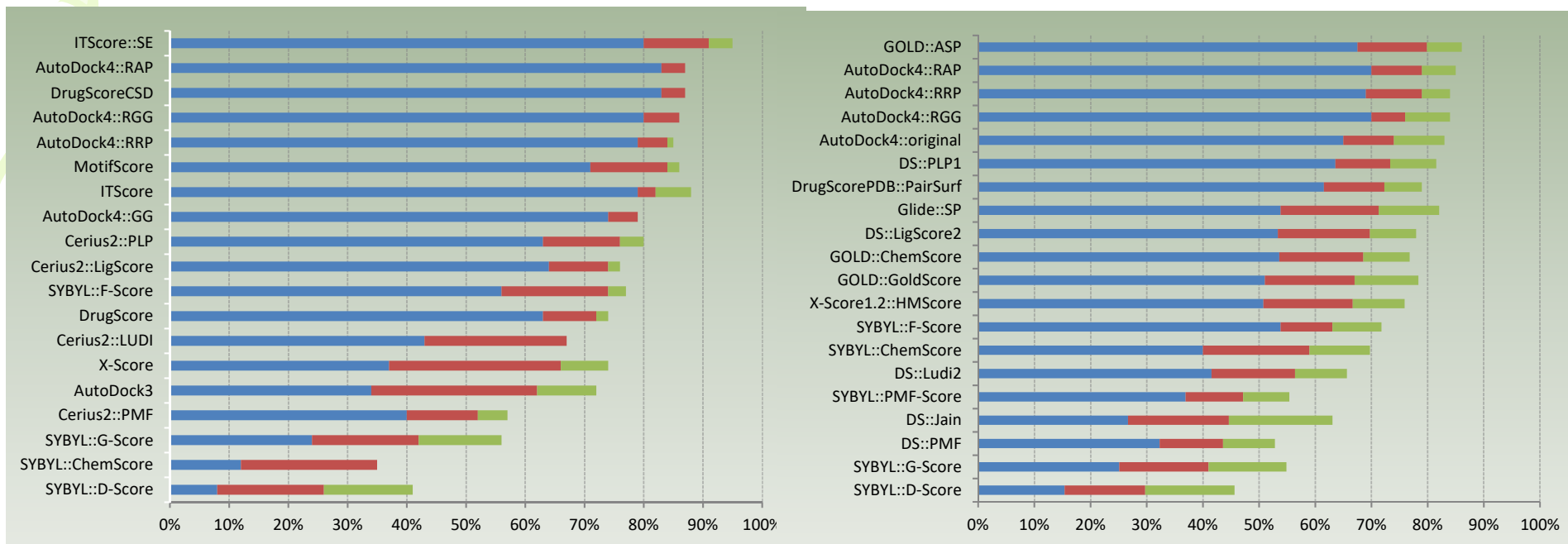
Comparison of the success rates of AutoDock4 scoring functions and other scoring functions on the decoys set of 100 complexes (Wang, R.X. et al., *J. Med. Chem.* 2003, 46, 2287-2303)

Comparison of the success rates of AutoDock4 scoring functions and other scoring functions on the decoys set of 195 complexes. (Cheng, T.J. *et al.*, *J. Chem. Inf. Model.* 2009, 49, 1079-1093)

blue: < 1 Å
red: < 2 Å
green: < 3 Å

Wang *et al., J. Chem. Inf. Model.* **51:** 2528-2537 (2011) [45]

Partial atomic occupancy in 1ajp
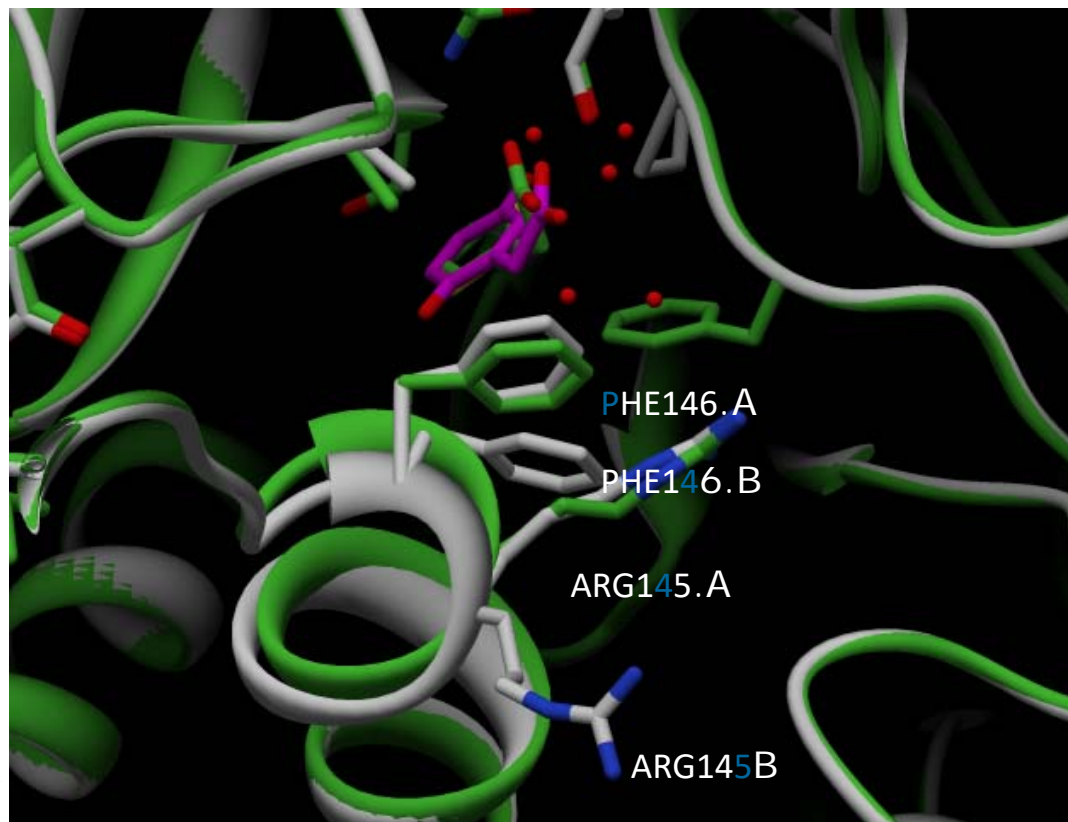
1ajp (grey) has two conformations of side chains on PHE146 and ARG145, whose atomic occupancy is 0.5. If the alternative conformations that are similar to 1ajq (green) were chosen, the estimated energy is only -1.87 kcal/mol due to the appeared clash (magenta ligand and PHE146.A).

# Poor fitting quality of the model structure to the electron density map for 1ajp
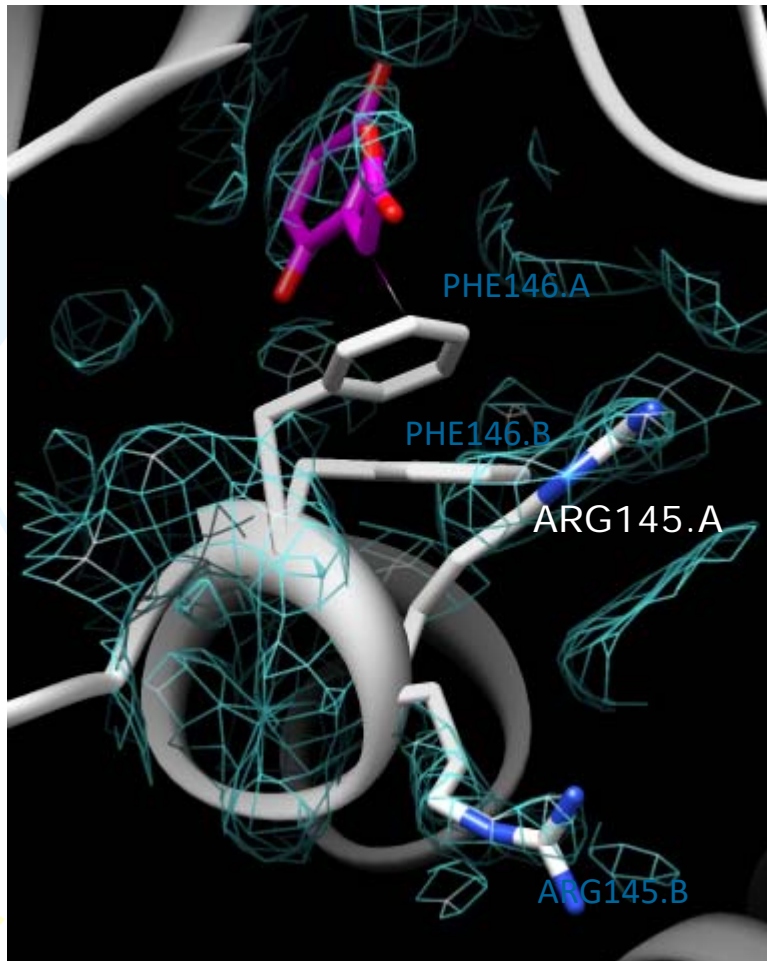
- The electron density was presented by Chimera with the setting of "Level = 0.426".
- Apparently, the side chains of PHE146 dost not map to the electron density shown in this figure.
- A close contact exists between the **ligand** and PHE146.A

Wrong biological assembly for 1tyr?
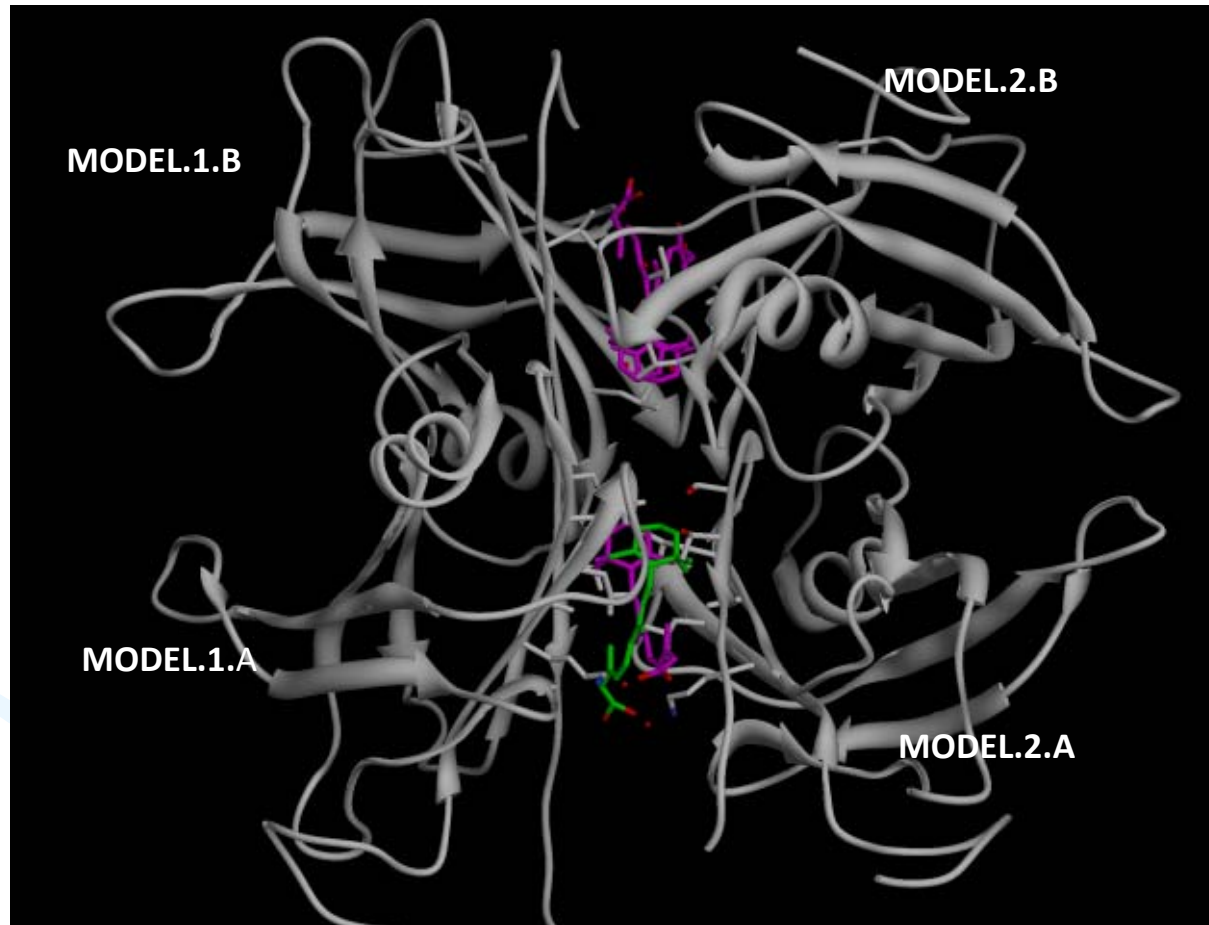
There two models in the biological assembly of 1tyr. The MODEL.1 with A and B chains was chosen in the set of Cheng *et al*.

Sometimes the water molecule still plays a critical role
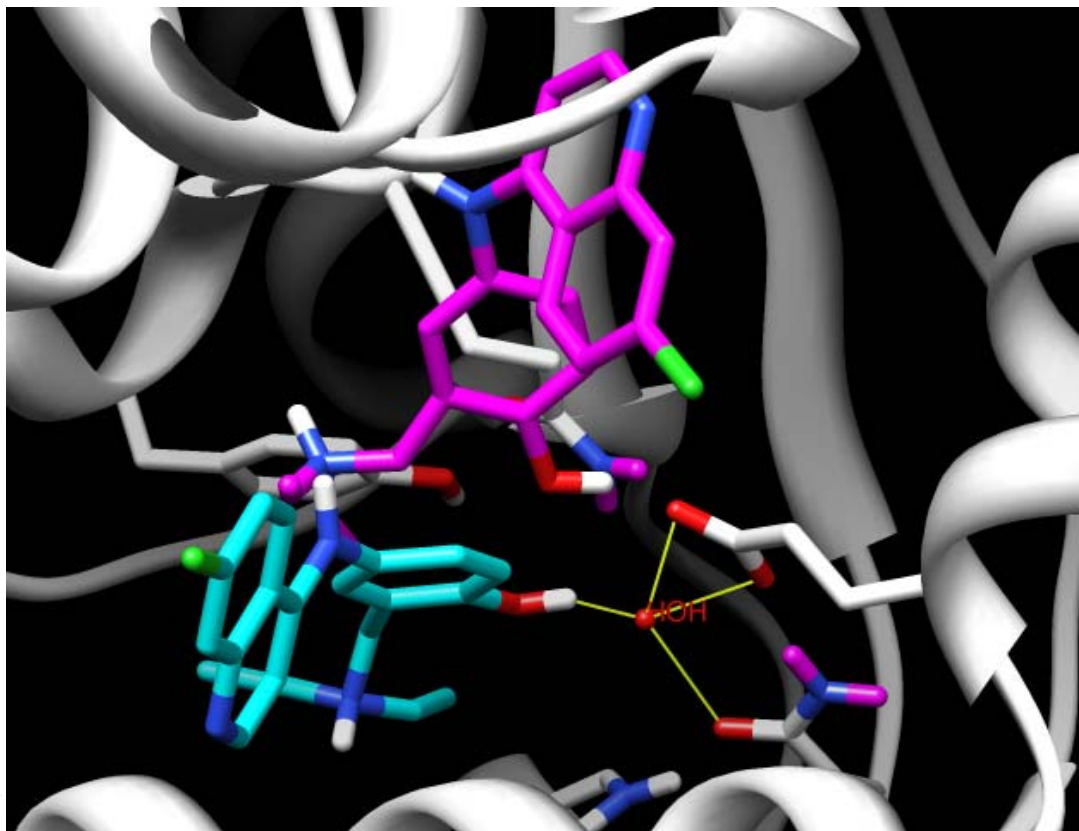
A water molecule HOH520 formed a tetra-coordinated hydrogen bonding network in the complex (**2aou**) of Histamine *N*-methyltransferase. However, neglect of this water resulted in the wrong ligand pose prediction (in magenta; native pose in cyan)

Performance of some scoring functions on weakly interacting complexes

Scoring functions	$K_{i/d} \geq 1$ mM	$K_{i/d} \geq 100$ μM	$K_{i/d} \geq 10$ μM
AutoDock4[RGG]	2.01	2.14	2.25
AutoDock4[RAP]	2.01	2.05	2.12
AutoDock4[RRP]	2.07	2.14	2.27
AutoDock Vina	2.17	1.92	1.90
X-Score	3.36	2.91	2.29

(RMSE in kcal/mol)

Class-dependence of robust scoring functions

	success rate (%; rmsd $\leq 2\text{Å}$)			
	overall	hydrophilic	mixed	hydrophobic
scoring function	(100)	(44)	(32)	(24)
AutoDock4^{RAP}	87	89	91	79
AutoDock4^{RGG}	86	86	91	79
AutoDock4^{RRP}	84	84	91	75
original AutoDock4^{GG}	79	77	81	79
Cerius2/PLP	76	77	78	71
SYBYL/F-Score	74	75	75	71
Cerius2/LigScore	74	77	75	67
DrugScore^{PDB}	72	73	81	58
Cerius2/LUDI	67	75	66	54
X-Score	66	82	59	46
AutoDock3	62	73	53	54
Cerius2/PMF	52	68	44	33
SYBYL/G-Score	42	55	34	29
SYBYL/ChemScore	35	32	34	42
SYBYL/D-Score	26	23	28	29

[a] Data were adopted from Wang *et al.*[23] except for AutoDock4 scoring functions.

[b] Scoring functions are sorted according to the overall success rates.

ZINC ^{12}

About   **Search**   **Subsets**   **Help**   **Social**   G+1 81

Quick Search Bar... **Go**

Please consider switching to ZINC15, which is superior to ZINC12 in most ways. If you prefer ZINC12 after trying ZINC15, we would like to know why @chem4biology so that we can get you to make the switch.

Welcome to ZINC, a free database of commercially-available compounds for virtual screening.   ZINC contains over 35 million purchasable compounds in ready-to-dock, 3D formats.   ZINC is provided by the Irwin and Shoichet Laboratories in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF).   To cite ZINC, please reference: Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model. 2012* DOI: 10.1021/ci3001277. The original publication is Irwin and Shoichet, *J. Chem. Inf. Model. 2005;45(1):177-82* PDF, DOI.   We thank NIGMS for financial support (GM71896).

Molecule of the Week   12414383
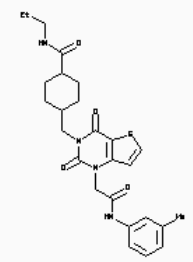
ZINC ID, Drug Name, SMILES, Catalog, Vendor Code, Target & m... **Go**

Structure/Draw   Physical Properties   Catalogs & Vendors   ZINC IDS   Targets   Rings   Combination

What's NEW?   Feedback   Like us
@chem4biology   Blog   RSS

ZINC Database

The Extended Relaxed Complex Scheme

- Multivalent drug design in a building-block fashion
- Computational analogue of "SAR by NMR"

1. Accommodating receptor flexibility by molecular dynamics

2. Assigning compound molecular properties using rigorous quantum chemical approaches

3. Rapid docking using efficient global search algorithms.

4. Ranking compounds by *binding free energy spectra*, instead of single binding free energy.

5. Molecular dynamics simulations of complexes with top ranked compounds.

6. Automatic free energy calculation with the double decoupling method or with the adaptive umbrella sampling scheme.

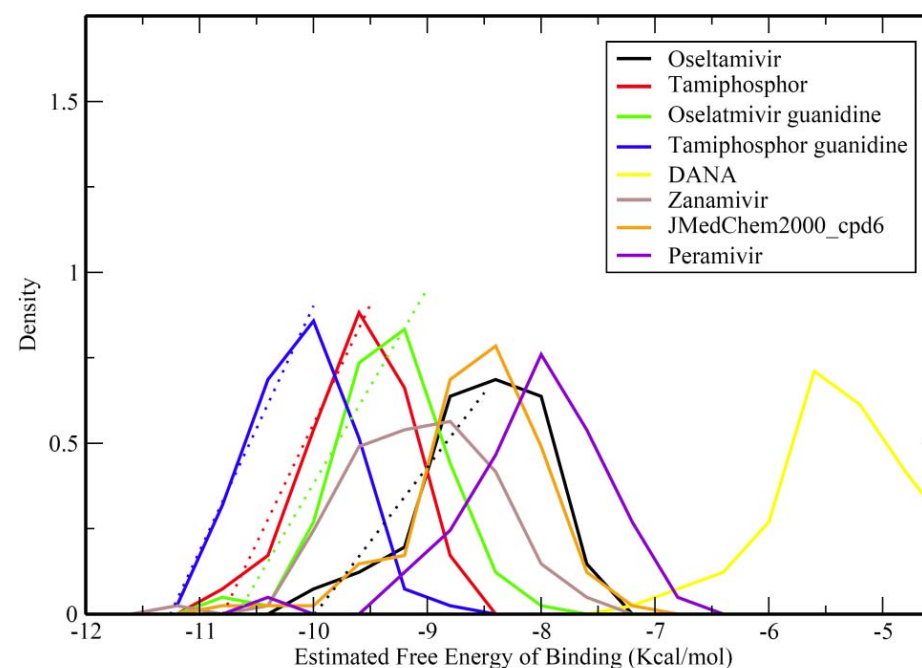Lin *et al. J. Am. Chem. Soc.,* **124**, 5632 (2002) , Lin *et al. Biopolymers,* **68**, 47 (2003)
Amaro *et al. J. Comput-Aid Mol. Des*, **22**, 639 (2008)
Lin, *Curr. Top. Med. Chem.* **11:** 171 (2011) , Lin, *Biopolymers,* **105**, 2 (2016)

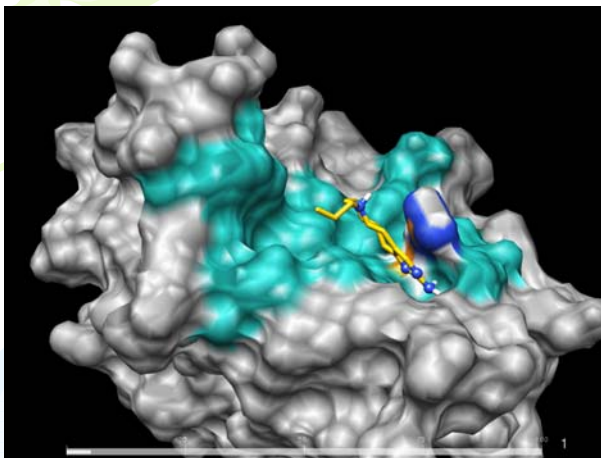# Binding free energy spectra with affinity propagation clustering
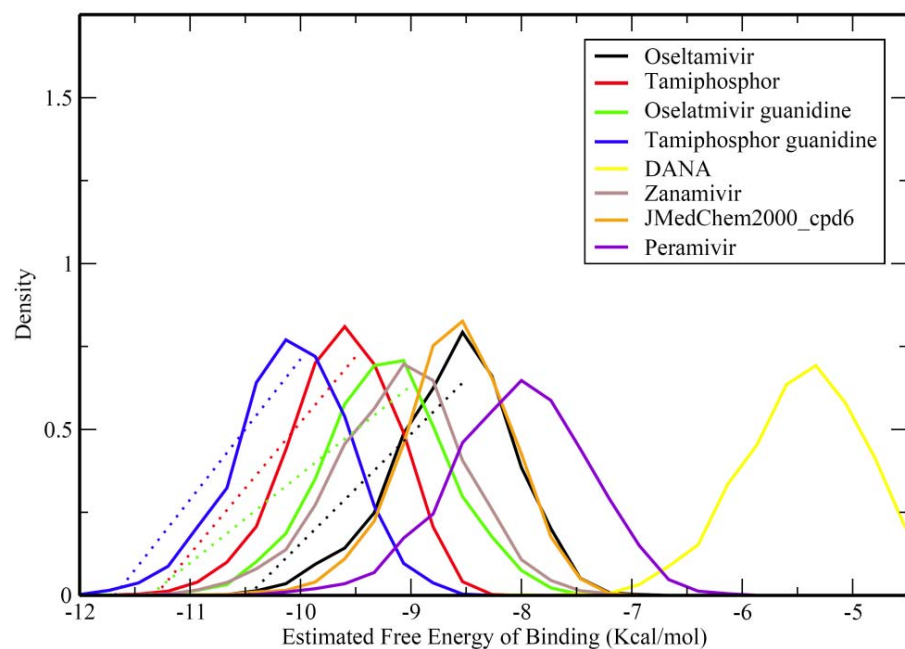
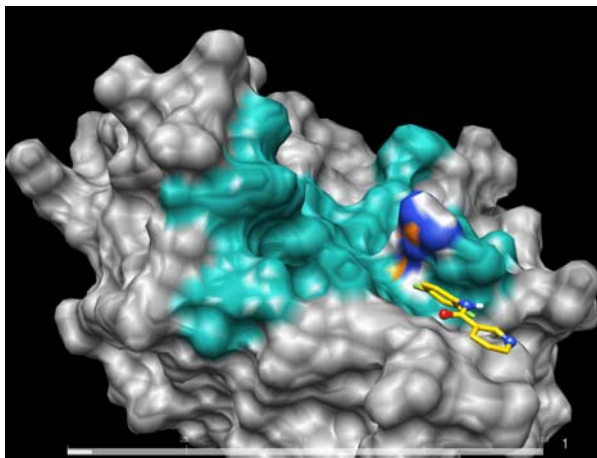RCS binding spectra from 15-ns explicit solvent MD 3000 snapshots, no clustering

RCS binding spectra from AP-clustered 100 MD snapshots
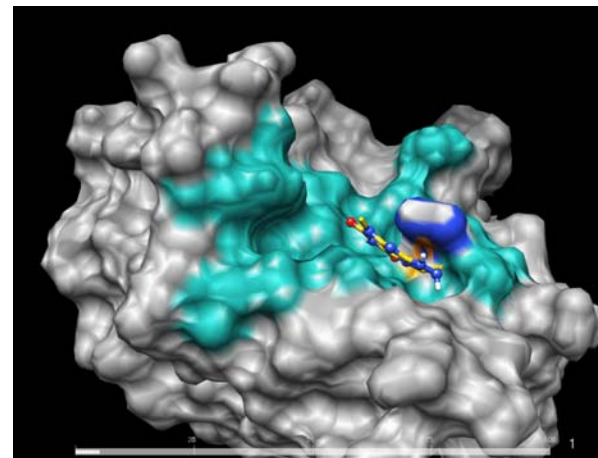
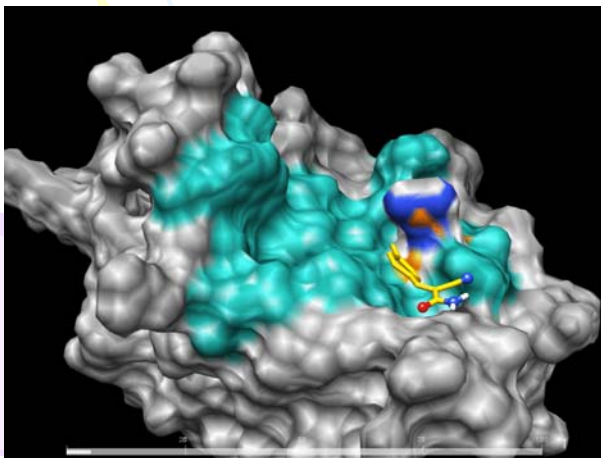Frey *et al. Science* **315**, 972(2007)

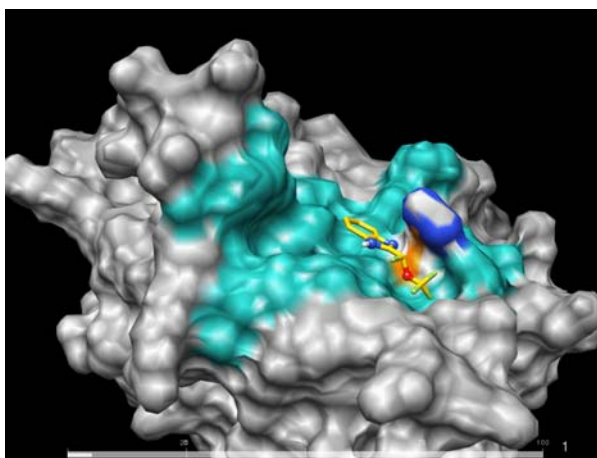## ZINC0000XX41
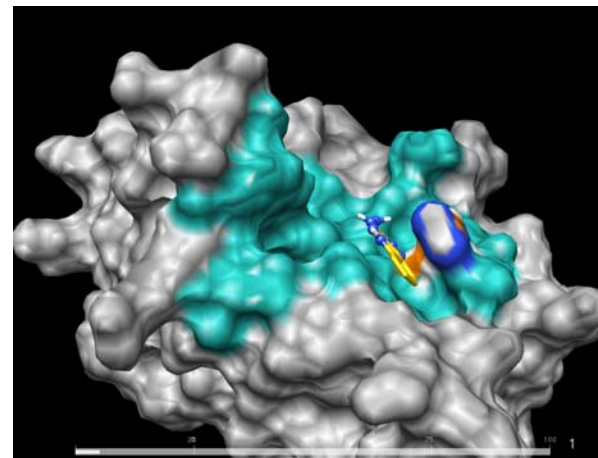

ZINC0002XX54

ZINC0003XX09

ZINC0004XX25

ZINC0004XX47

ZINC0005XX58

ZINC15

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 100 million purchasable compounds in ready-to-dock, 3D formats.

ZINC is provided by the Irwin and Shoichet Laboratories in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). We thank NIGMS for financial support (GM71896).

To cite ZINC, please reference: Sterling and Irwin, *J. Chem. Inf. Model,* *2015* http://pubs.acs.org/doi/abs/10.1021/acs.jcim.5b00559. You may also wish to cite our previous papers: Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model, 2012* DOI: 10.1021/ci3001277 or Irwin and Shoichet, *J. Chem. Inf. Model. 2005;45(1):177-82* PDF, DOI.

Getting Started

- Getting Started
- What's New
- About ZINC 15 Resources
- Current Status / In Progress
- Why are ZINC results "estimates"?

Explore Resources

Chemistry
Tranches, Substances, 3D

Ask Questions

You can use ZINC for **general** questions such as

- How many substances in current clinical trials have PAINS patterns? (150)
- How many natural products have names in ZINC and are not for sale? (9296) get them as SMILES, names and calculated logP
- How many endogenous human metabolites are there? (47319) and how many of these can I buy? (8271) How many are FDA approved drugs? (94)
- How many compounds known to aggregate are in current clinical trials? (60)

ZINC15 News

- 2016-03-14 - Massive purchasability update.
- 2016-03-14 - ZINC15 is going to ACS. See you in San Diego.
- 2015-09-28 - ZINC15 is released! Follow us on twitter @chem4biology Known limitations What's new

Caveat Emptor: We do not guarantee the quality of any molecule for any purpose and take no responsibility for errors arising from the use of this

Resource	Is a set of	Approximate Number	the answer to your question is: a) a list of these things b) one of these things, or c) things derived from a single one of these things.
substances ⧉	molecules	200,000,000	a) compounds you can buy [1] ⧉ b) ZINC ID 53 [2] ⧉ c) All genes hit by compound ZINC597013 [3] ⧉
catalogs ⧉	vendor and annotated catalogs	400	a) catalogs whose members are biogenic (but see also endogenous, metabolite) [4] ⧉ b) The ChEMBL20 catalog [5] ⧉ c) Items in the DrugBank FDA catalogs [6] ⧉ Interestingly, this shows clearly

Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease

Nanjie Deng,[*,†,‡] Stefano Forli,[§] Peng He,[†,‡] Alex Perryman,[§] Lauren Wickstrom,[‖] R. S. K. Vijayan,[†,‡] Theresa Tiefenbrunn,[§] David Stout,[§] Emilio Gallicchio,[⊥] Arthur J. Olson,[§] and Ronald M. Levy[*,†,‡]

[†]Center for Biophysics & Computational Biology/ICMS, [‡]Department of Chemistry, Temple University, Philadelphia, Pennsylvania19122, United States

[§]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, United States

[‖]Borough of Manhattan Community College, The City University of New York, Department of Science, New York, New York 10007, United States

[⊥]Department of Chemistry, Brooklyn College, the City University of New York, Brooklyn, New York, United States

■ INTRODUCTION

Molecular docking is widely used in rational drug discovery and structural biology for predicting the most favorable pose and for estimating the strength of ligand–receptor binding.[1,2] In a typical virtual screening application, a large library of compounds is docked against a receptor target site to generate plausible poses ranked by scoring functions. Such functions are typically designed to have a simple form for computational efficiency. While docking has matured into a powerful tool for pharmaceutical research after decades of development,[1–7] the accuracy of docking calculations continues to be limited by these relatively simple scoring functions which lack a complete treatment of desolvation and receptor reorganization.[8,9] Additionally, entropic factors are generally not captured well by scoring based on a single structure.[8,10] As a result, structure-based ligand screening by docking often generates a large number of false positive hits. As a recent example, Shoichet et al.[11] conducted a parallel study of docking and HTS to screen 197861 compounds against cruzain, a thiol protease with a relatively rigid binding pocket. Among the top 0.1% of the docking-ranked library, 97.5% of the hits were found to be false positives.[11]

Binding free energy methods are based on statistical mechanics and atomistic simulations and, in principle, can
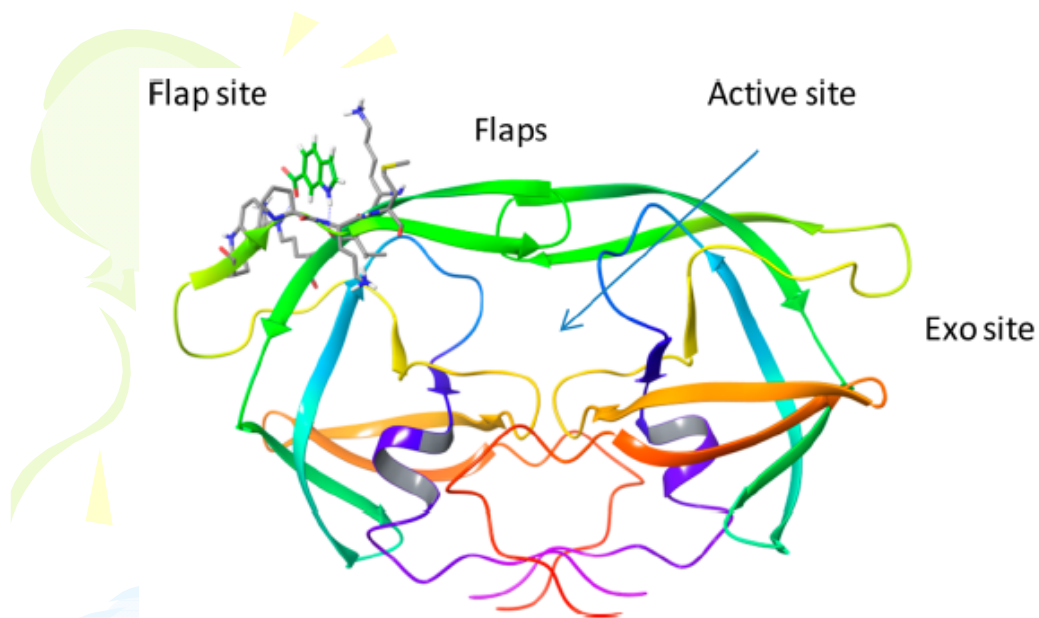
Figure 1. Crystal structure of HIV PR (pdb id: 3kfr) with its flap site occupied by the ligand 1F1 shown in green stick. The active site ligand is removed from the figure for clarity.
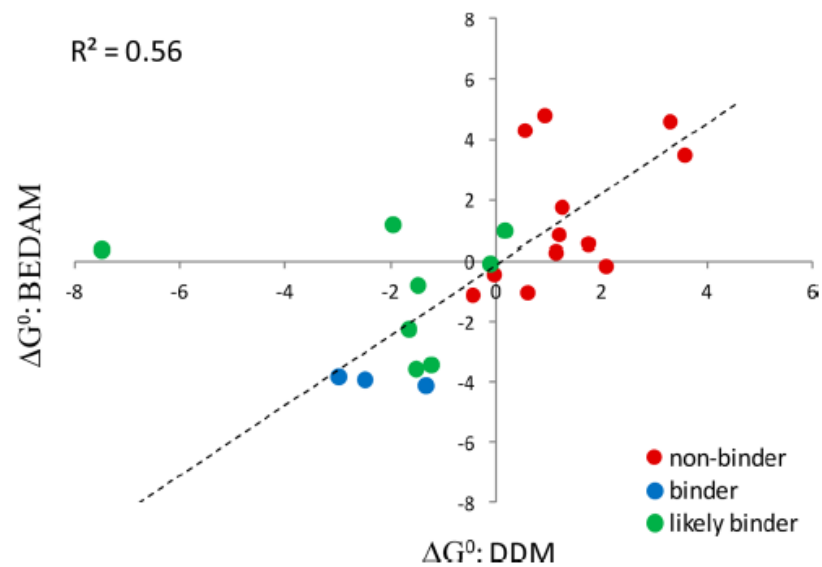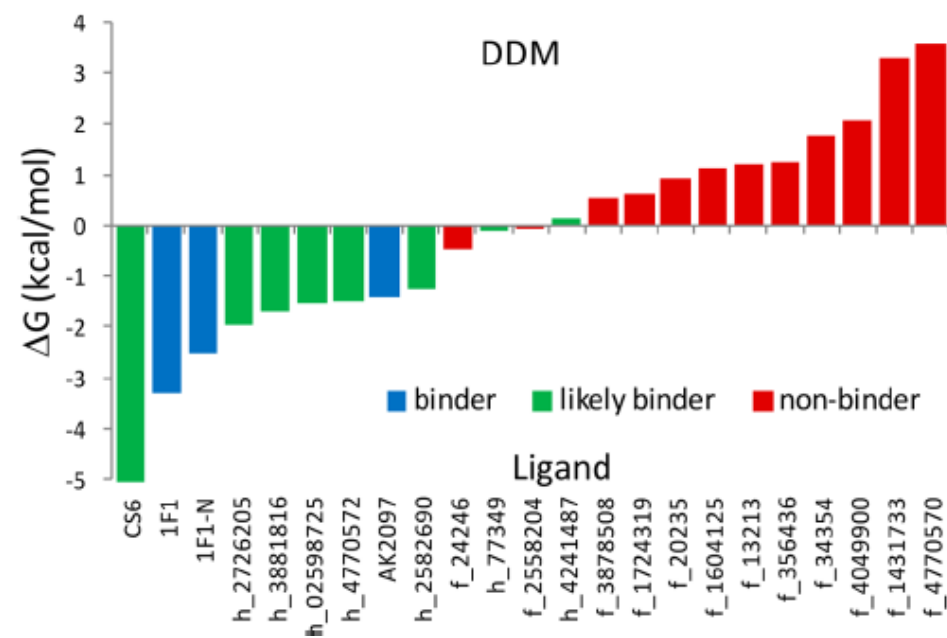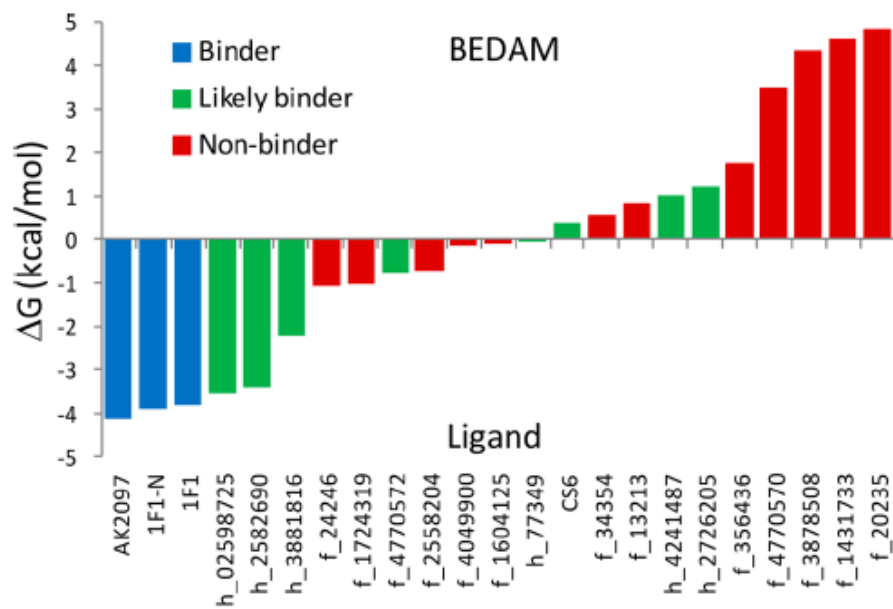

Figure 3. Correspondence between the binding free energies computed using BEDAM and DDM. Unit in kcal/mol. The ligand CS6 is excluded from the linear regression.

The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review

Michael K. Gilson,* James A. Given,* Bruce L. Bush,# and J. Andrew McCammon§

*Center for Advanced Research in Biotechnology, National Institute of Standards and Technology, Rockville, Maryland 20850-3479; #Department of Molecular Design and Diversity, Merck Research Laboratories, Rahway, New Jersey 07065; and §Departments of Chemistry and Biochemistry, and Pharmacology, University of California at San Diego, La Jolla, California 92093-0365 USA

ABSTRACT   Although the statistical thermodynamics of noncovalent binding has been considered in a number of theoretical papers, few methods of computing binding affinities are derived explicitly from this underlying theory. This has contributed to uncertainty and controversy in certain areas. This article therefore reviews and extends the connections of some important computational methods with the underlying statistical thermodynamics. A derivation of the standard free energy of binding forms the basis of this review. This derivation should be useful in formulating novel computational methods for predicting binding affinities. It also permits several important points to be established. For example, it is found that the double-annihilation method of computing binding energy does not yield the standard free energy of binding, but can be modified to yield this quantity. The derivation also makes it possible to define clearly the changes in translational, rotational, configurational, and solvent entropy upon binding. It is argued that molecular mass has a negligible effect upon the standard free energy of binding for biomolecular systems, and that the cratic entropy defined by Gurney is not a useful concept. In addition, the use of continuum models of the solvent in binding calculations is reviewed, and a formalism is presented for incorporating a limited number of solvent molecules explicitly.

Taiwan Pharmaceutical Databank
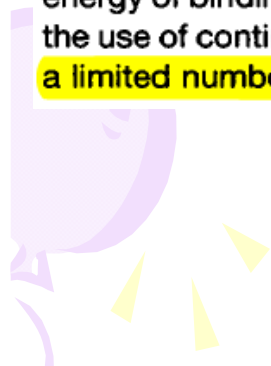
TPD Home

Browse TPD
search TPD
– by keyword
– by structure
searchReferences

List
– Class
– Family Name
– Scientific Name
– Journal
– Ref. Title
– PI

Links
Help & Doc
Contact Us

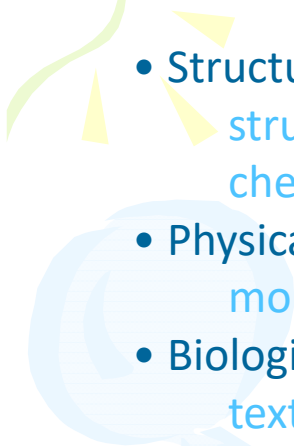Register
Login

User ID:

password:

Login

Welcome to Taiwan Pharmaceutical Databank

- Browse compound entries
- Search compound entries
    - by text:
        compound's name, compound's class
        family name or scientific name of natural sources
        title, author, or journal name of references
    - by structure:
        substructure or structural similarity
- List
    non-redundant list for quick search
- User registration and administration
    guest, user, advanced user, manager
- Deposit and edit compound entries
    edit data and entry status

Taiwan Pharmaceutical Databank

The information for each compound entry includes:

- Structure and Nomenclature
  structure (still image and interactive presentation), compound's class
  chemical name, IUPAC name, synonym, CAS Registry Number
- Physical and Chemical Properties
  molecular formula, molecular weight, melting point, solubility, store condition
- Biological Activity and Toxicity
  text and/or image
- Spectrum Analysis
  text and/or image
- Natural Sources
  family name, scientific name, part used, collected time and location
- Amount in the lab
- References
  reference information and links to PubMed and/or DOI if available
- External links
  Chemical Structure Lookup Service, ChemSpider

Taiwan Pharmaceutical Databank

Ordered by Molecular Formula 10 entries/page submit
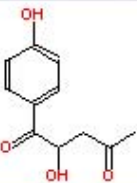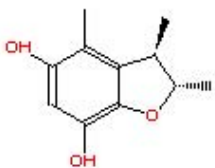
Jump to

1245 entries found in TPD   **Page 125 - 3**

#	CSN	Chemical Name / Molecular Formula / Class	Structure
21	151188792665	Cryptamygin A C11H12O3 Benzenoid	
22	91186201659	(-)-5-hydroxy-4-methoxy-1-tetralone C11H12O3 Naphthalenone	
23	121185757535	4-hydroxy-5-(p-hydroxyphenyl)pentane-2,5-dione C11H12O4 Phenolic	
24	101184292620	2,3,4-trimethyl-5,7-dihydroxy-2,3-dihydrobenzofuran C11H14O3 Phenolic	

idTarget: Identification of Biomolecular Targets of Small Chemical Molecules

http://idtarget.rcas.sinica.edu.tw/
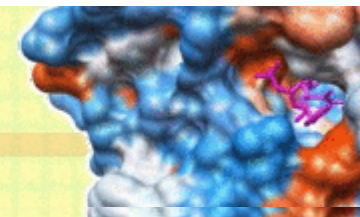
idTarget

Welcome to idTarget

A web server for identifying biomolecular targets of small chemical molecules with a divide-and-conquer docking approach

Identification of biomolecular targets of small chemical molecules is essential for unraveling the underlying molecular causes of actions. Often, natural products, i.e., compounds discovered from plants, animals, marine lives or other living organism, exhibit useful pharmaceutical effects, e.g., anti-inflammatory, anti-cancer, anti-viral effects, yet their molecular mechanisms remain elusive. On the other hand, many drugs are known to be accompanied with unpleasant adverse effects, but the molecular targets of such effects are largely unknown. In contrast, there are also some old drugs whose beneficiary effects are discovered recently, i.e., the anticancer effect of cholesterol-lowering drugs, statins, and their molecular mechanisms have become an intensive research subject. idTarget is a web server that can predict possible binding targets of a small chemical molecule via a divide-and-conquer docking approach, in combination with a recently recalibrated scoring function and a consensus scoring scheme, where the new scoring function was trained based on 7864 protein-ligand complexes. In the divide-and-conquer docking calculations, small overlapping grids are adaptively constructed to constrain the searching space and thereby achieving the convergence of docking results with better efficiency. idTarget has been shown to be able to reproduce known off-targets of drugs or drug-like compounds.

Wang et al. Nucleic Acids Research 40: W393-W399 (2012)

Your E-Mail: [_____]

E-Mail is not required. If a valid email is provided, you will be noticed when the submitted job is finished.

Structural Files 💡

Ligand: ⦿ Upload molecular file [_____] 瀏覽...

Which charge model was used for this uploaded molecule?

○ Gasteiger ○ AM1-BCC ○ RESP ⦿ unknown

Has the protonated state been determined in this molecular file?

○ No ○ Yes, no further adding polar hydrogens is needed

○ Draw a molecule [using on-line structure editor]

○ ligand A for example 1: **Dual inhibition of**

○ ligand B for example 1: Dual inhibition of H

○ ligand for example 2: **Dual inhibition of C**

Protein Set: ○ User's list [Edit list on-line]

⦿ idTarget validated Set 1 (**189 prot**

○ **Protein list** for example 1: Dual i

○ **Protein list** for example 2: Dual i
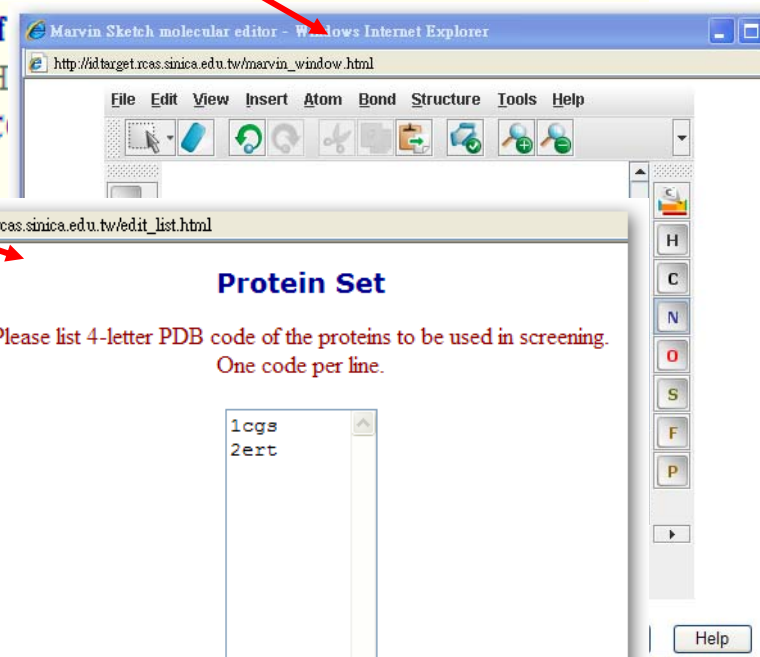
Parameters 💡

Charge model combination for Ligand / Protein

○ Gasteiger / Gasteiger

⦿ AM1-BCC / Amber PARM99SB

○ RESP / Amber PARM99SB

Docking Method ⦿ AutoDock4 ○ Vina + AutoDock4

[submit] [reset]

Marvin Sketch molecular editor - Windows Internet Explorer

http://idtarget.rcas.sinica.edu.tw/marvin_window.html

File Edit View Insert Atom Bond Structure Tools Help

http://idtarget.rcas.sinica.edu.tw/edit_list.html

Protein Set

Please list 4-letter PDB code of the proteins to be used in screening. One code per line.

```
1cgs
2ert
```

[done & back] [cancel & close]

H
C
N
O
S
F
P

[Help]

Some proteins could be very large :
the case of P-glycoprotein with Paclitaxel (Taxol)
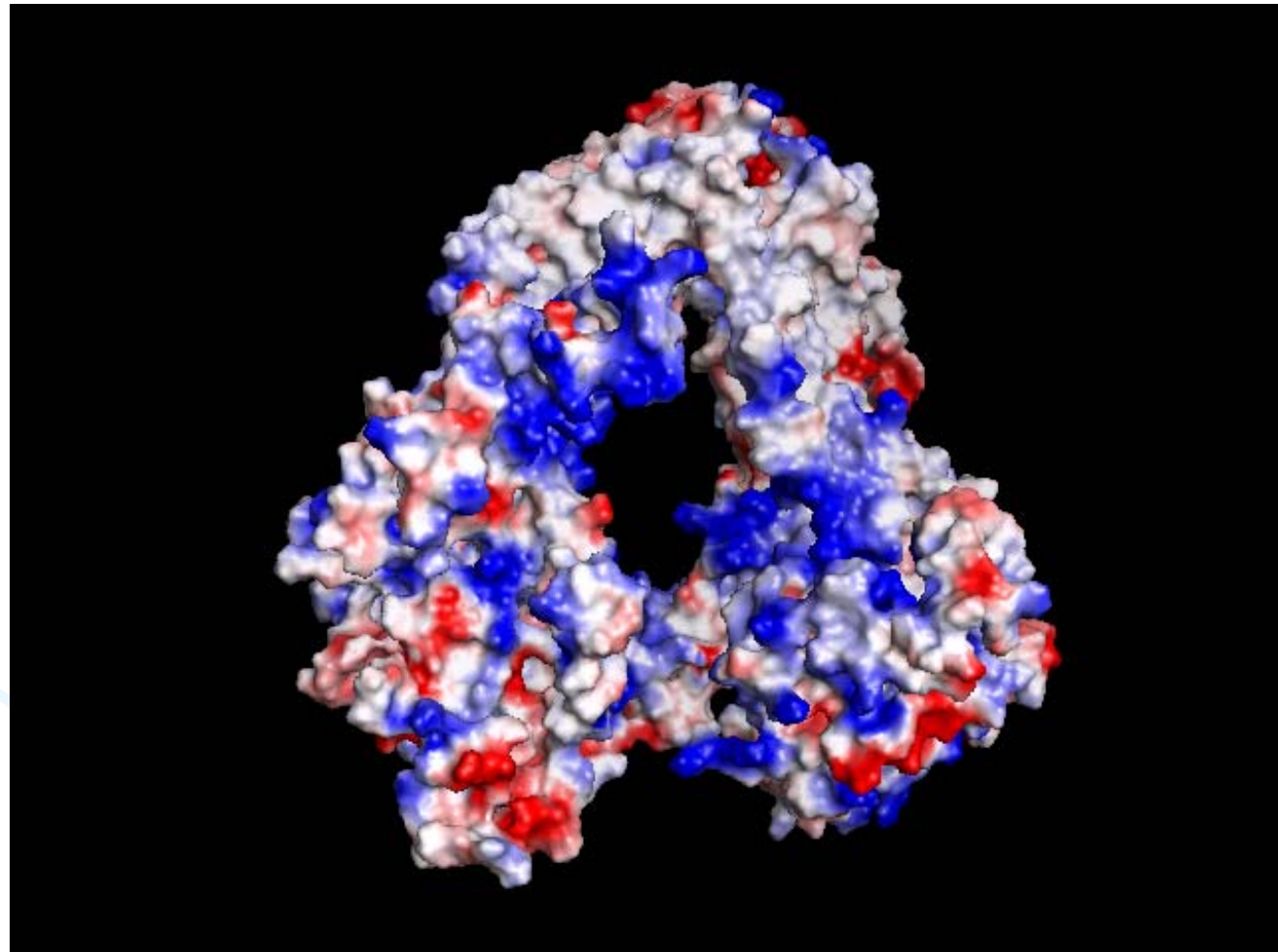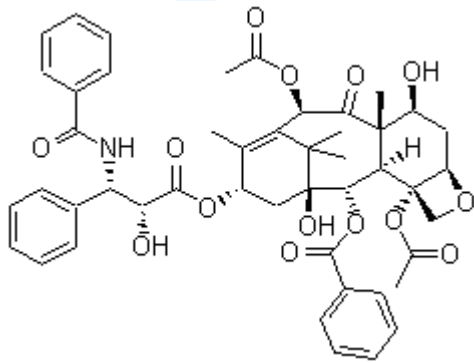
Divide-and-Conquer Docking
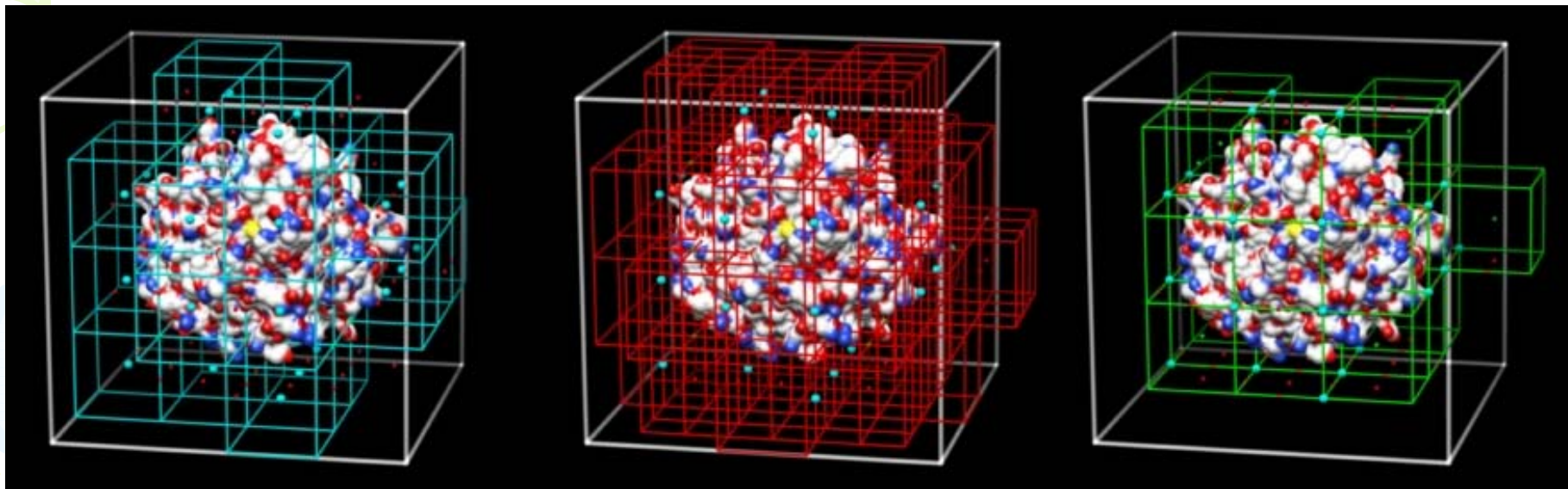
Firstly, a box with 10Å to each boundary of the receptor was drawn (white). This big box was subsequently divided into smaller boxes (cyan; the cyan dots are grid box centers) where the size box was dynamically determined according to the size of query ligand. The overlapped boxes (red and green dots are the grid box centers of overlapped boxes) should be taken into count. If a grid box is far away the receptor (no atom is within 1.42*length of the grid box to the center), it will be eliminated to reduce the computational cost. Finally, the boxes shown above are retained.

Example: Off-target screening for the inhibitor in 3MAX

idTarget

| Home | Services | Documentaion | Examples | Contact us |
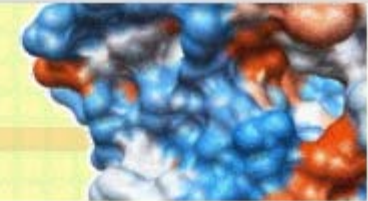
Result of 1293705965mj

3max

Jmol
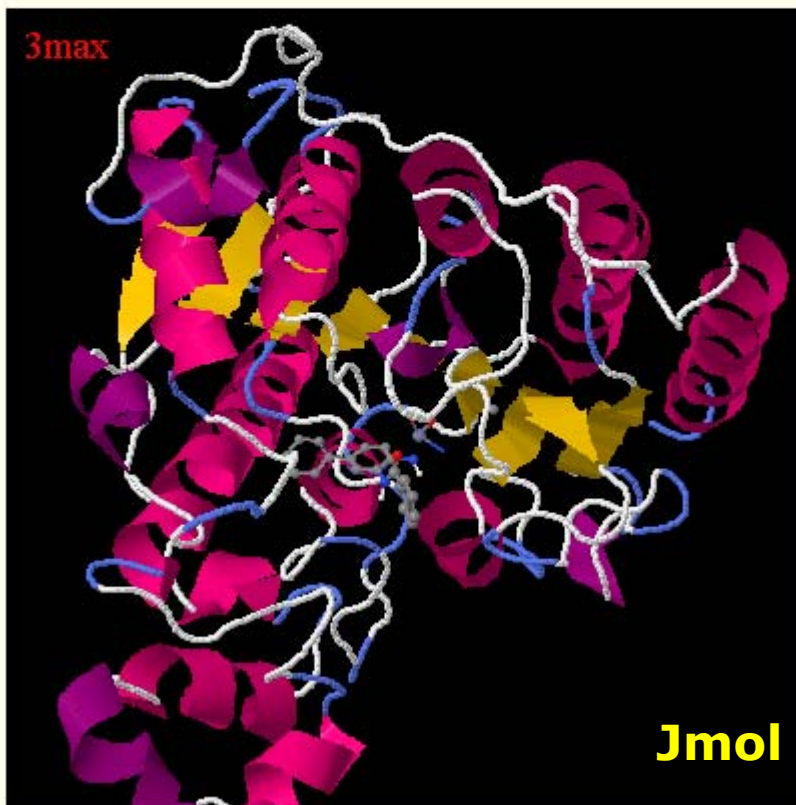
This idTarget job is finished. 10 proteins were screened. You may download the result.

Rank by Energy

#	PDB ID	Energy (kcal/mol)	XScore	ligand pose	PDB Link
1	3max	-15.01	7.70	Download	PDB
2	1hw9	-10.08	6.59	Download	PDB
3	1ajv	-8.37	6.94	Download	PDB
4	1cim	-8.21	6.34	Download	PDB
5	1tng	-8.00	6.20	Download	PDB
6	1hhi	-7.86	6.44	Download	PDB
7	1mcb	-7.75	7.19	Download	PDB
8	1epo	-7.13	6.52	Download	PDB
9	4hmg	-6.65	6.07	Download	PDB
10	1aee	-4.68	5.52	Download	PDB

Virtual screen of large chemical library in search of new drug candidates

By using 1056 cores, we are able to screen the entire over 10,000,000 compound library within 1 or 2 weeks. It will take more than 1 year for a typical lab which owns less than 100 cores.

Such application is especially useful when time is a critical issue, e.g., finding effective compounds against H5N1, H7N9, SARS, and other infectious disease.

In general, for new drug targets of any disease without any known effective therapeutic agents, virtual screen is an effective approach to find out drug candidates.

Virtual screening results in the search of new generation antibiotics. The hits are ranked according to the predicted binding affinity.