MM-PBSA Validation Study

Trent E. Balius

Department of Applied Mathematics and Statistics

AMS 535

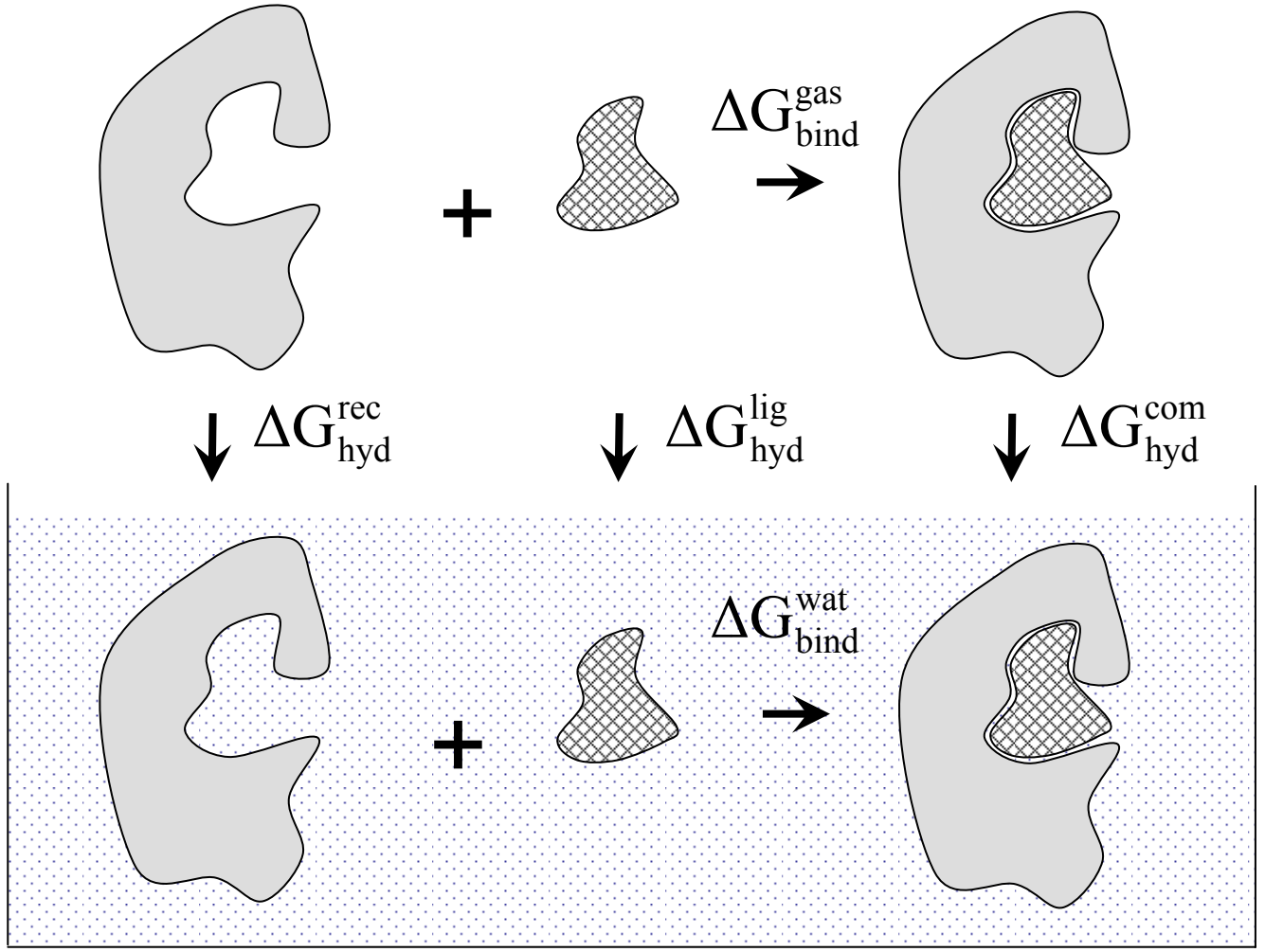
11-26-2008

Overview

- MM-PBSA
  - Introduction
  - MD ensembles
  - one snap-shots relaxed structures
- Enrichment
- Computational vs. experimental activities
- ROC curves
- Validating MM-PBSA with virtual screening

Introduction to Molecular Mechanics Poisson Boltzmann Solvent Accessible Surface Area (MM-PBSA)

Thermodynamic cycle



MM-PBSA/GBSA Equations

$$G = E_{MM} + G_{PBSA/GBSA} - TS$$

such that

$$E_{MM} = E_{bond} + E_{angle} + E_{tors} + E_{vdw} + E_{es}$$

TS is calculated using quasi harmonic analysis normal mode analysis

$$G_{PBSA/GBSA} = G_{polar} + G_{nonpolar}$$

where the polar and the nonpolar terms are defined in the following way

G_{polar} - is defined by solving the PB set of differential equations

or by using the GB equation.

$$G_{nonpolar} = \alpha \cdot SA + \beta$$

$$\Delta G = G_{complex} - G_{protein} - G_{ligand}$$

MM-PBSA/GBSA Equations

$$G = E_{MM} + G_{PBSA/GBSA} - TS$$

such that

$$E_{MM} = E_{bond} + E_{angle} + E_{tors} + E_{vdw} + E_{es}$$

TS is calculated using quasi harmonic analysis normal mode analysis

$$G_{PBSA/GBSA} = G_{polar} + G_{nonpolar}$$

where the polar and the nonpolar terms are defined in the following way

G_{polar} - is defined by solving the PB set of differential equations

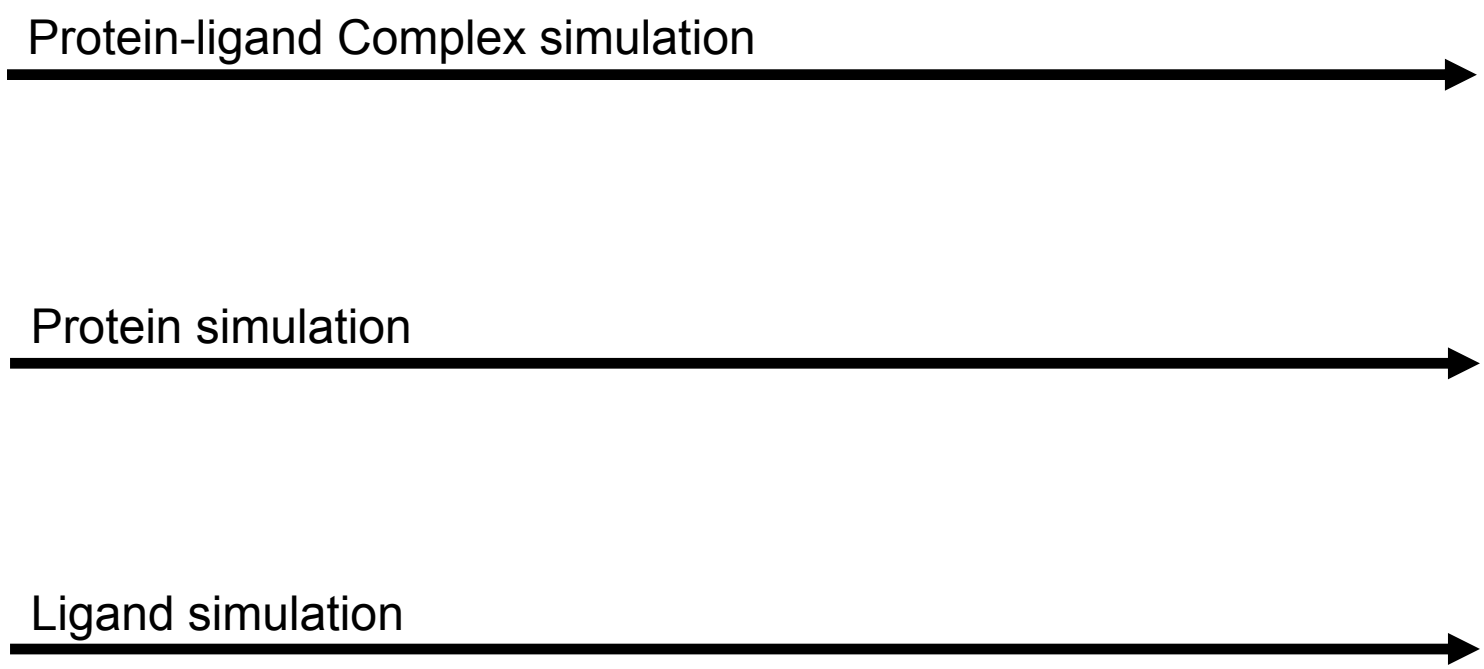
or by using the GB equation.

$$G_{nonpolar} = \alpha \cdot SA + \beta$$

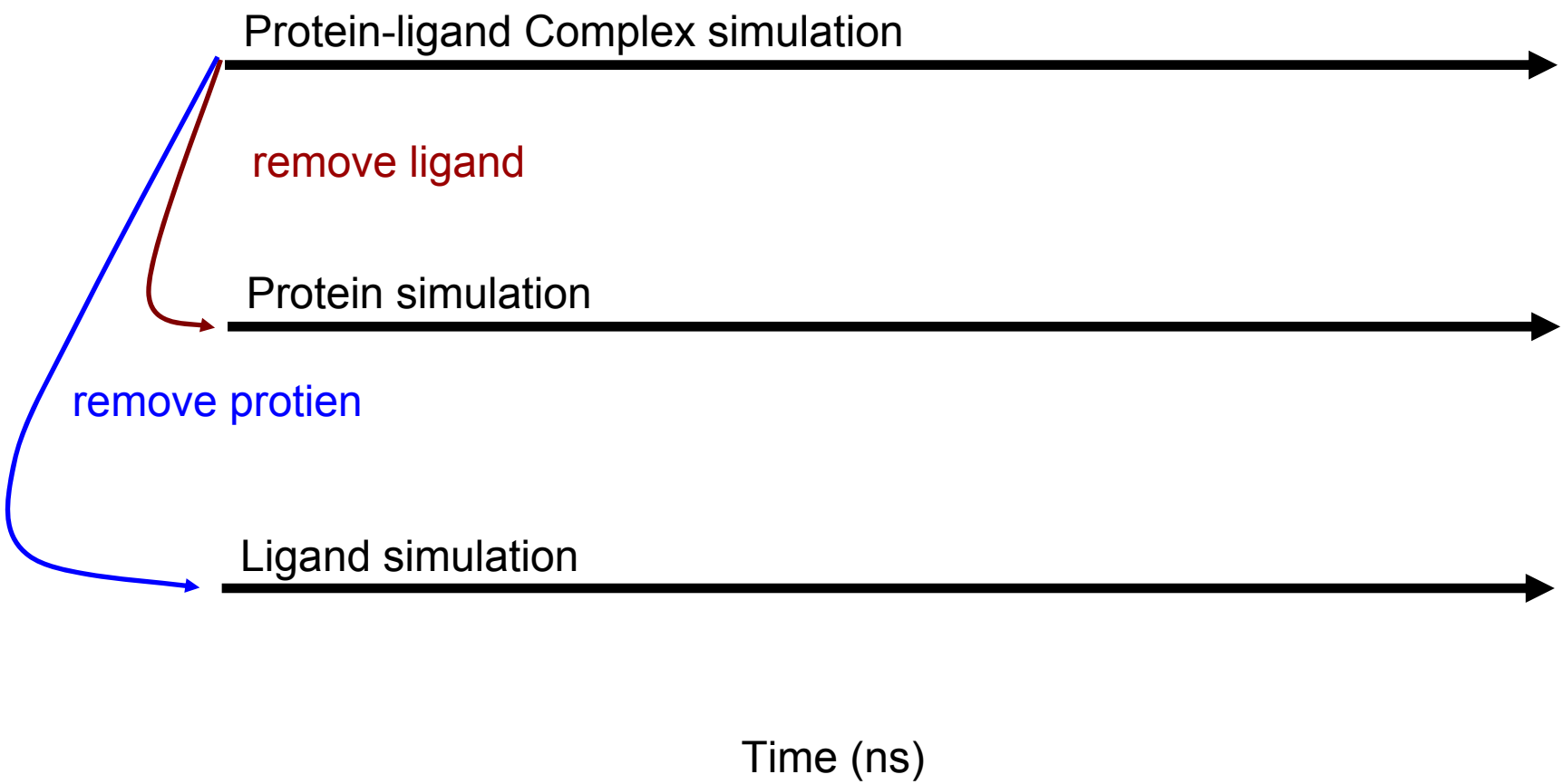
$$\Delta G = G_{complex} - G_{protein} - G_{ligand}$$

Using Molecular Dynamics generated ensembles

Run 3 independent Simulations



Run 1 Simulations



Metrics for Determining Good Agreement

Predictive Index (PI)

- prediction accuracy for ranking different inhibitors
- *E(i)* -- experimental binding free energy
- *P(i)* -- calculated energy value of a ligand *i*,
- w_{ij} --weighting term difference between the experimental values of the two ligands that are compared
- PI = 1, if always right
- PI = -1, if always wrong
- note that E is a IC50 value and P is an energy score

$$PI = \frac{\sum_i \sum_{j>i} w_{i,j} C_{i,j}}{\sum_i \sum_{j>i} w_{i,j}}$$

$$w_{i,j} = \left| E(i) - E(j) \right|$$

$$C_{i,j} = \begin{cases} 1 & \text{if } \left[E(i) - E(j)\right] / \left[P(i) - P(j)\right] < 0 \\ -1 & \text{if } \left[E(i) - E(j)\right] / \left[P(i) - P(j)\right] > 0 \\ 0 & \text{if } \left[P(i) - P(j)\right] = 0 \end{cases}$$

Correlation Coefficient -- Probability Theory

covariance →
$$\operatorname{cov}[X,Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

variance →
$$\operatorname{var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

expectation →
$$E[X] = \sum_{i=1}^{n} x_i \, p(x_i)$$

mean →
$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
← uniformly distributed

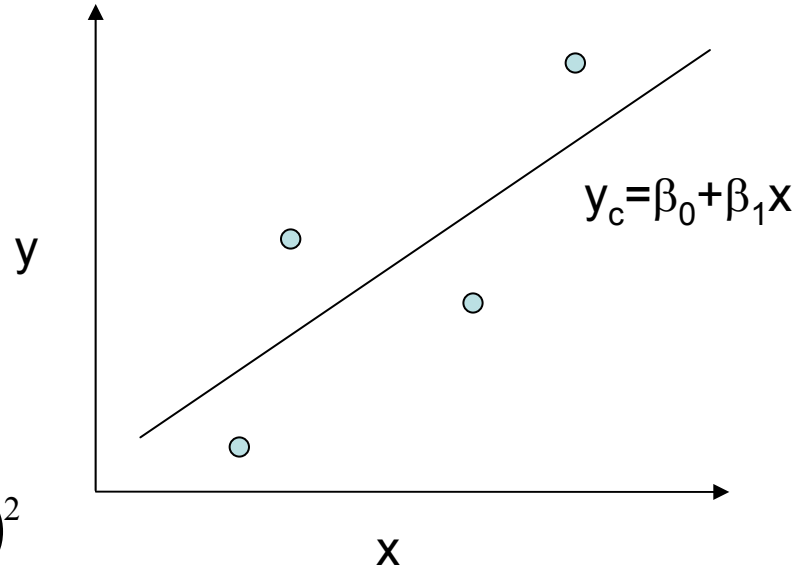
D.P. Bertsekas, J.N. Tsitsiklis. Introduction to Probability, 2nd Ed.

Correlation Coefficient

$$\rho(X,Y) = \frac{\operatorname{cov}[X,Y]}{\sqrt{\operatorname{var}[X]\operatorname{var}[Y]}}$$

$$R = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \hat{\beta}_1 \frac{s_x}{s_y}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}), \quad S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$



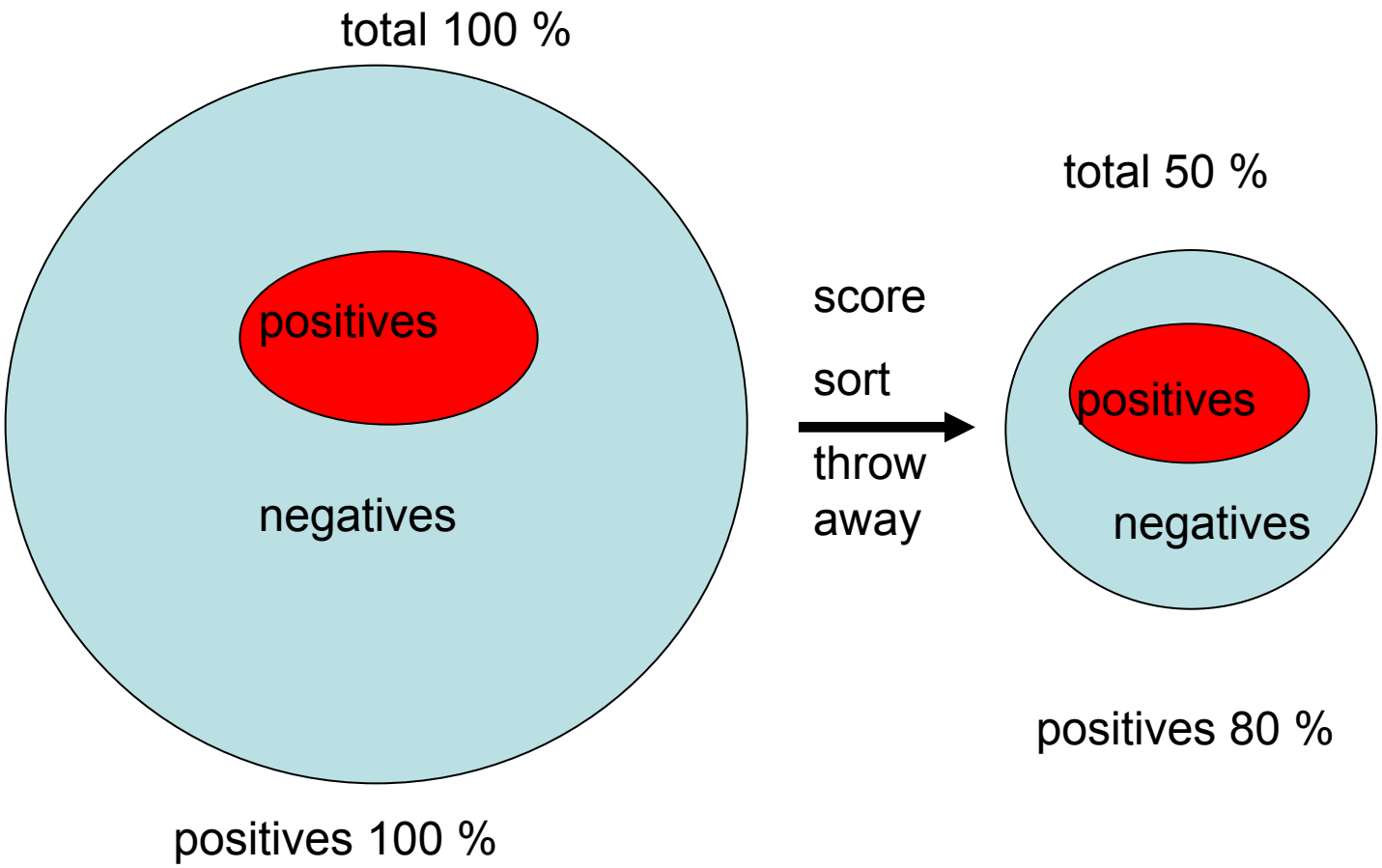
$$s_x = \sqrt{\frac{S_{xx}}{n-1}}$$ ← estimated population standard deviation for uniform distribution

A.C. Tamhane, D.D. Dunlop. Statistics and Data Analysis: From Elementary to Intermediate

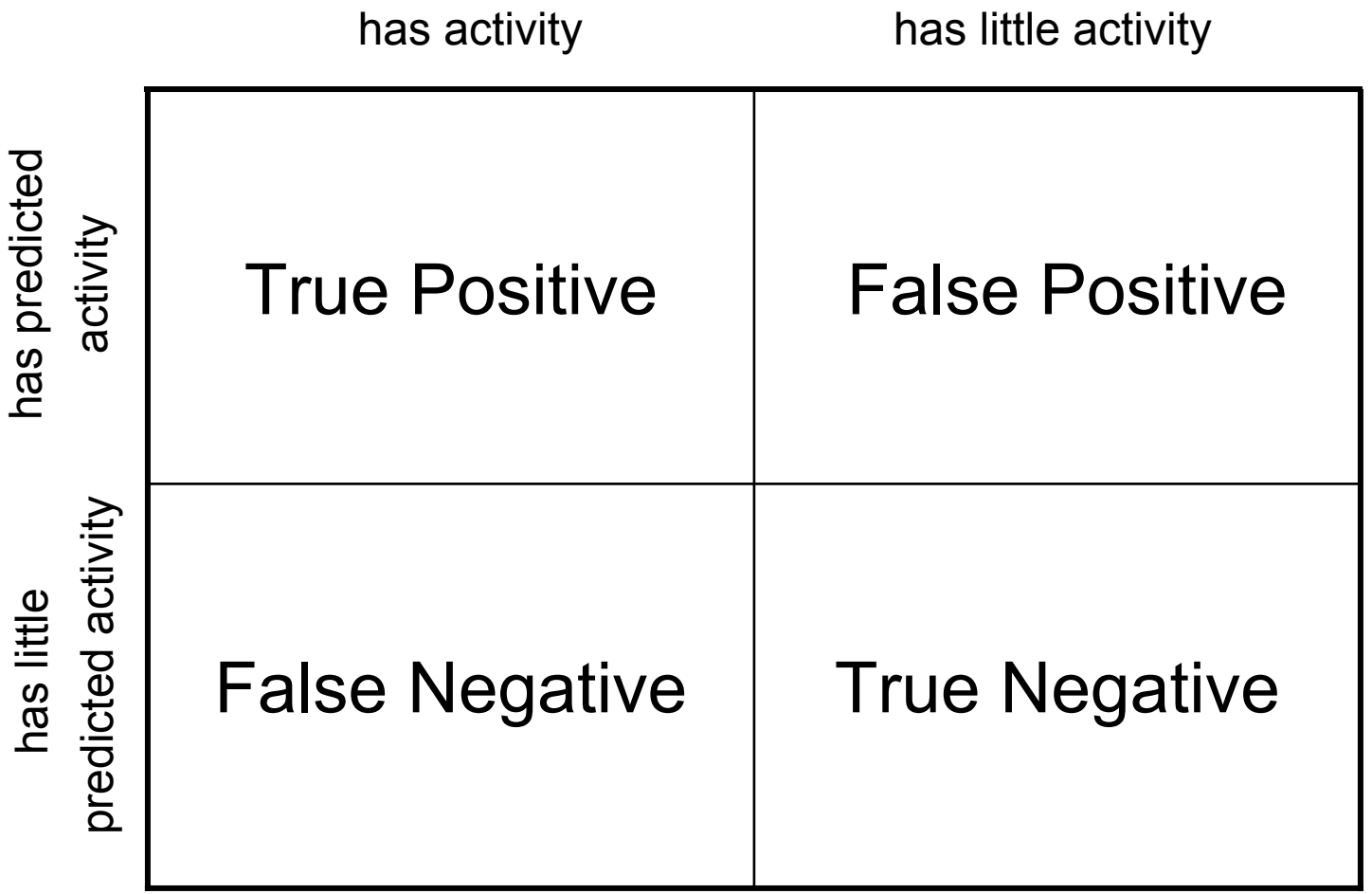
D.P. Bertsekas, J.N. Tsitsiklis. Introduction to Probability, 2nd Ed.

Virtual Screening and Receiver Operating Characteristic (ROC) Curves

Virtual Screening (Enrichment)

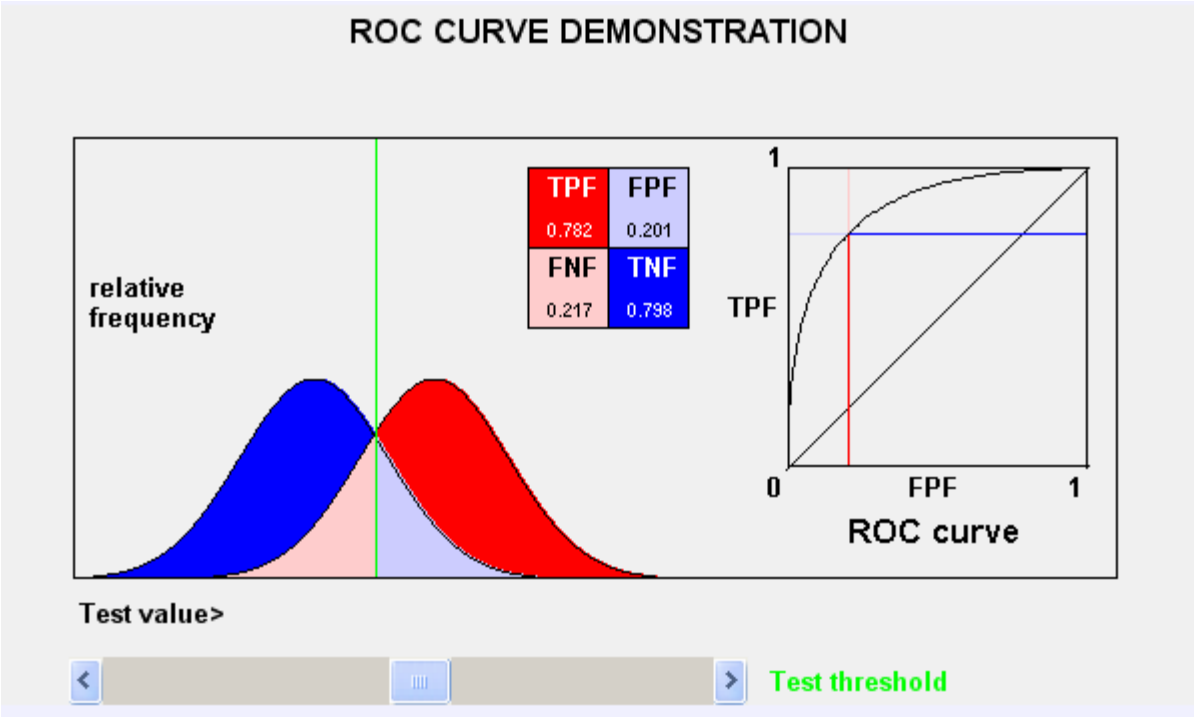


Computational Prediction vs. Experimental Evidenced



ROC curves

- ROC -- Receiver Operating Characteristic



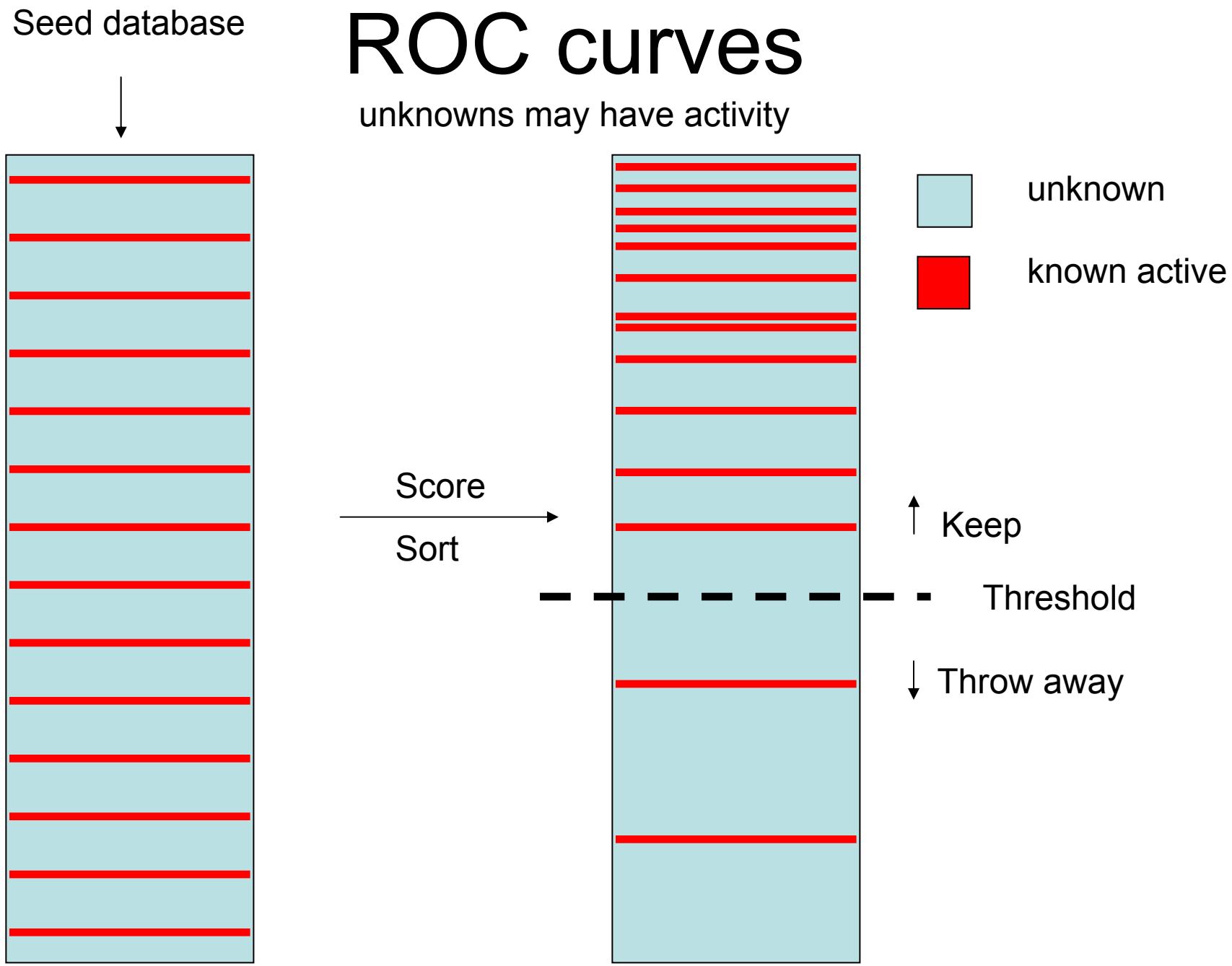
http://www.anaesthetist.com/mnm/stats/roc/Findex.htm

ROC curves

- Active and inactive is not known
  - Why not do the experiments on whole population?
    - expensive
    - takes time
    - multiple levels of experiments (needs to comparing type of experiments e.g.. HTPS)
  - Seed the population with know active compounds
  - see how many bubble to the top.

ROC curves

unknowns may have activity



Paper Figures and Analysis

Paper Nomenclature

- MM-RDIEL -- molecular mechanics energy function including distance dependent dielectric
- MM-PBSA -- single relaxed structure
- MD-PBSA -- molecular dynamics ensemble
- MD-PBSA* -- minimized staring point

Computational Details

- Force Fields (*antechamber* of Amber 7*)*
  - small molecules
    - AM1-BCC charges
    - GAFF (failed in 10%)
    - MAB* -- MAB (united atom) and GAFF
  - Protein -- FF94
  - 24 Å sphere of TIP3P water and neutralized by adding counter ions
- MD simulations -- Amber 6 for MD T = 300, Δt = 1.5fs, shake on
  - equilibration: 150 ps
  - snapshots every 5 ps for 50 ps
  - 10 structures ensemble
- Minimization -- Related structure
  - minimized for 1000 steps

Journal of Medicinal Chemistry, 2005, Vol. 48, No. 12

Computational Details

- MM-PBSA calculations
  - PB calculated using MEAD program package
  - Normal mode analysis -- mfebd module of MOLOC
  - SA -- in-house program XSAE

Energy Comparison

Table 1. Validation of the MAB* and GAFF Force Fields for Test Sets of Conformational Energies and Intermolecular Interaction Energies[a]

	MMFF94s	MM3*	CVFF	MAB*	GAFF
Conformational Energies I (19 comparisons)[38]					
rmsd	0.74	0.78	2.86	**1.70**	**0.50**
max. dev.	1.45	2.27	5.78	**3.58**	**0.87**
Conformational Energies II (37 comparisons)[37]					
rmsd	0.38	0.72	2.36	**1.93**	**1.17**
max. dev.	0.99	2.57	6.11	**4.27**	**2.88**
Intermolecular Interaction Energies (66 comparisons) [37]					
rmsd	0.76	3.81	4.36	**2.09**	–
max. dev.	2.52	13.31	16.39	**7.17**	–

[a] The data for the MMFF94s, MM3*, and CVFF force fields for these test sets are taken from the literature.[37] rmsd stands for the root-mean-square deviation of the calculated energies relative to the experimental or ab initio reference values and max. dev. denotes the maximal deviation. The force fields used in this study are shown in bold. No nonbonded cutoff and a dielectric constant of 1 were used.

$$\mathrm{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(E_i - E_i'\right)^2}$$

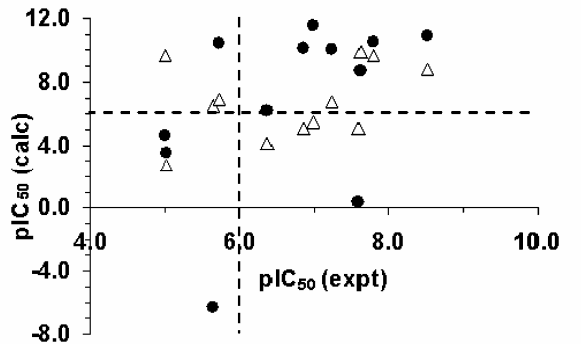
$$\mathrm{max.dev.} = \max\left(\left|E_i - E_i'\right|\right)$$

Journal of Medicinal Chemistry, 2005, Vol. 48, No. 12

Table 2. Performance Comparison in Ranking Similar Ligands for Three Different Test Sets Using MM-PBSA, MD-PBSA, and MM-RDIEL with the MAB* and GAFF Force Fields[a]

	MM-PBSA MAB*	MD-PBSA MAB*	MD-PBSA* MAB*	MM-PBSA GAFF	MD-PBSA GAFF	MM-RDIEL MAB*	FlexX ScreenScore
			Avidin (8 ligands)				
PI	0.81	0.80	0.91	0.90	0.86	0.99	0.49
R^2	0.63	0.68	0.87	0.87	0.73	0.59	0.24
			p38 MAP Kinase (test set II, 12 ligands)				
PI	0.53	0.31	0.27	0.51	0.37	−0.06	0.31
R^2	0.21	0.11	0.07	0.20	0.17	0.00	0.12
			p38 MAP Kinase (test set III, 16 ligands)				
PI	0.27	0.31	−0.22	−0.01	0.15	−0.14	−0.10
R^2	0.04	0.12	0.01	0.01	0.03	0.03	0.00

[a] MD-PBSA* indicates free energy averaging over minimized MD snapshots. PI stands for the predictive index defined in eqs 4−6, and R^2 is the correlation coefficient between experiment and computation. A previous MD-PBSA study on the avidin system using a considerably more demanding computational setup yielded PI = 0.99 and R^2 = 0.92.[16]



calculated vs. experimental of 12 Roche p38 MAP kinase inhibitors.
MM-PBSA using the MAB* force field (filled circles) with FlexX/ScreenScore (empty triangles)
Dashed lines at pIC50 = 6.0 indicate a threshold of IC50 = 1 μM.

Predicting Correct Binding Modes

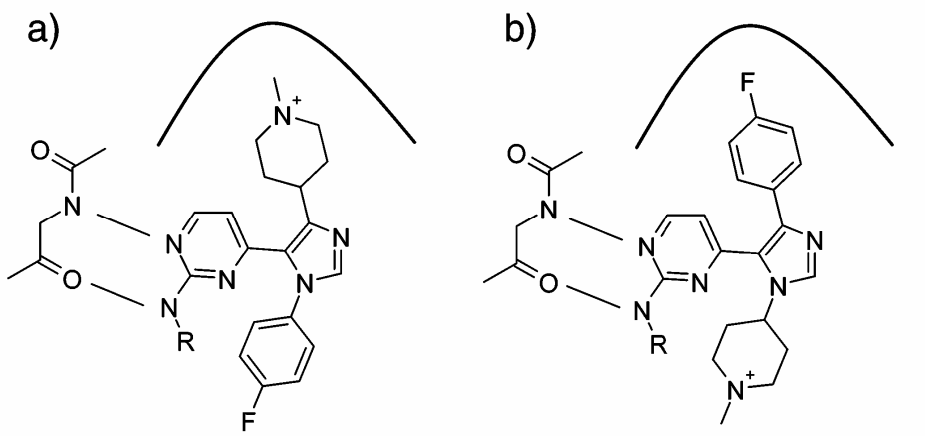
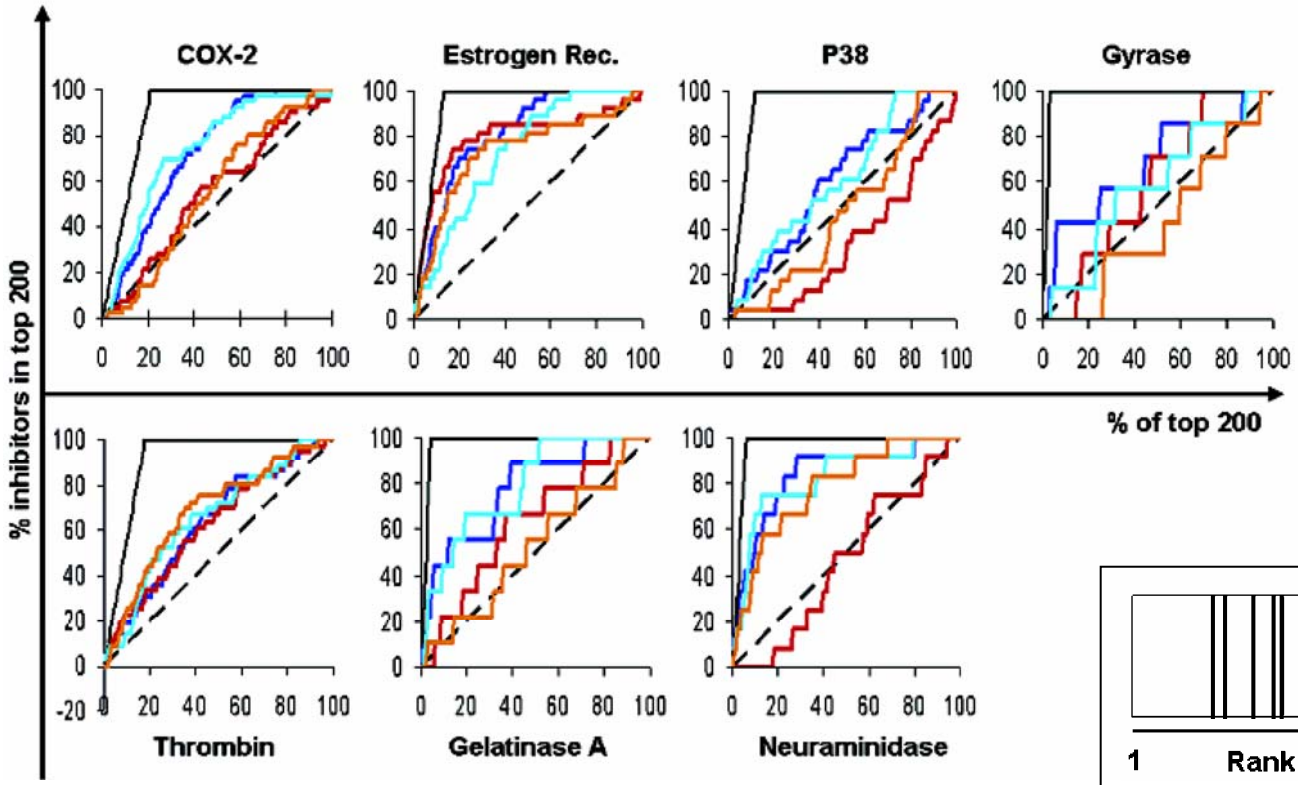
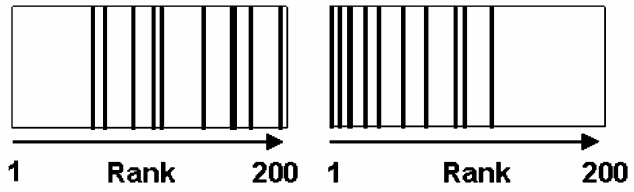


Illustration of handling of solvation effects. (a) Binding mode suggested by FRED/ChemScore (Rank 7); corresponding MM-PBSA result (Rank 122, ΔGbind) +7.8 kcal/mol). (b) X-ray binding mode obtained by rotation around the pyrimidine imidazole bond (MM-PBSA: Rank 57, ΔGbind = -2.6 kcal/mol). R = CH2-phenyl.

Journal of Medicinal Chemistry, 2005, Vol. 48, No. 12

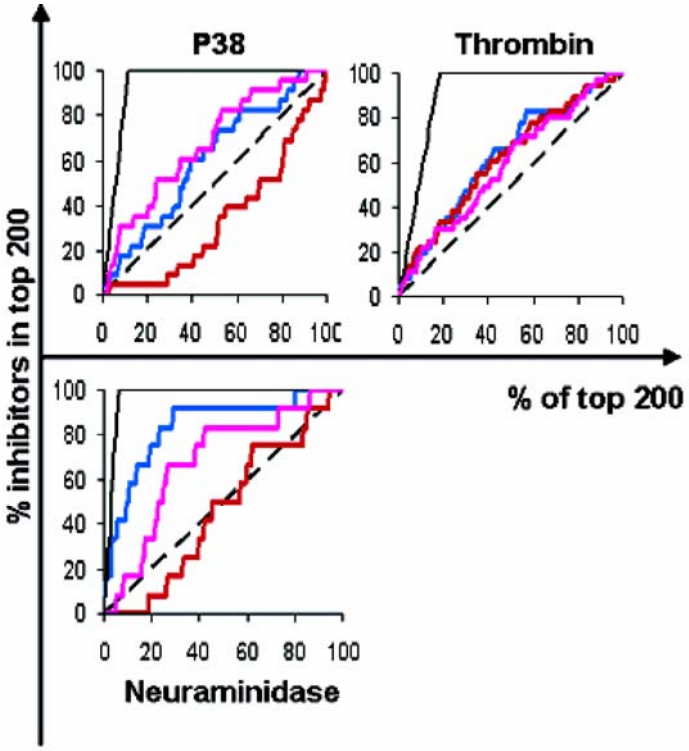


The enrichment curves show the percentage of inhibitors retrieved as a function of the percentage of total molecules in the database.
random selection (black, dashed)
ideal performance (black, solid)
FRED/ChemScore ranking (**red**)
MM-RDIEL ranking with MAB* force field (**orange**)
MM-PBSA ranking with MAB* force field (**blue**)
MM-PBSA ranking with GAFF force field (**cyan**).

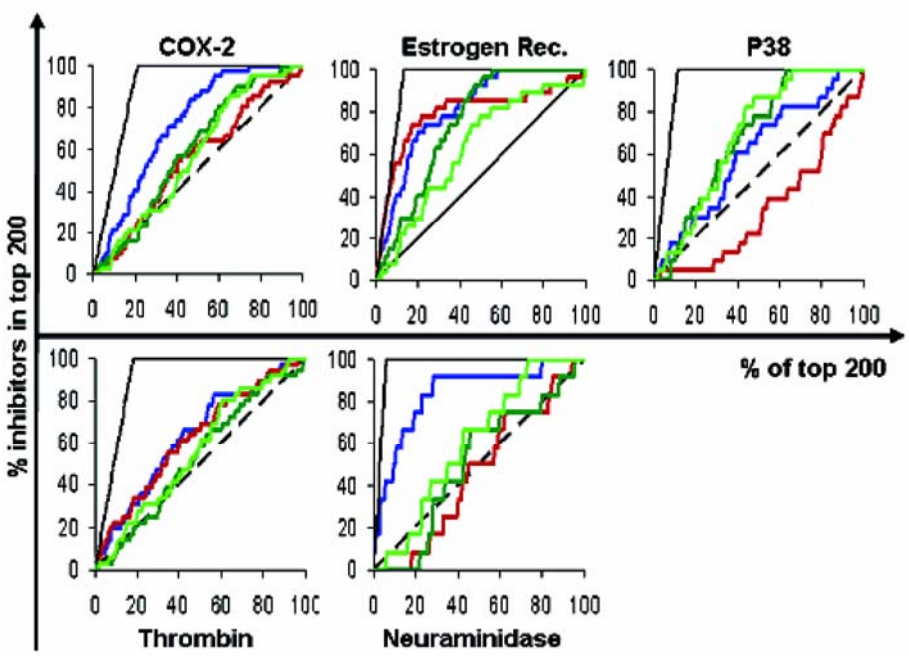


Ranking of p38 MAP kinase inhibitors with correct binding mode (*N* = 11). The diagrams show the ranking obtained with FRED/ChemScore (left) and MM-PBSA (right).

Journal of Medicinal Chemistry, 2005, Vol. 48, No. 12

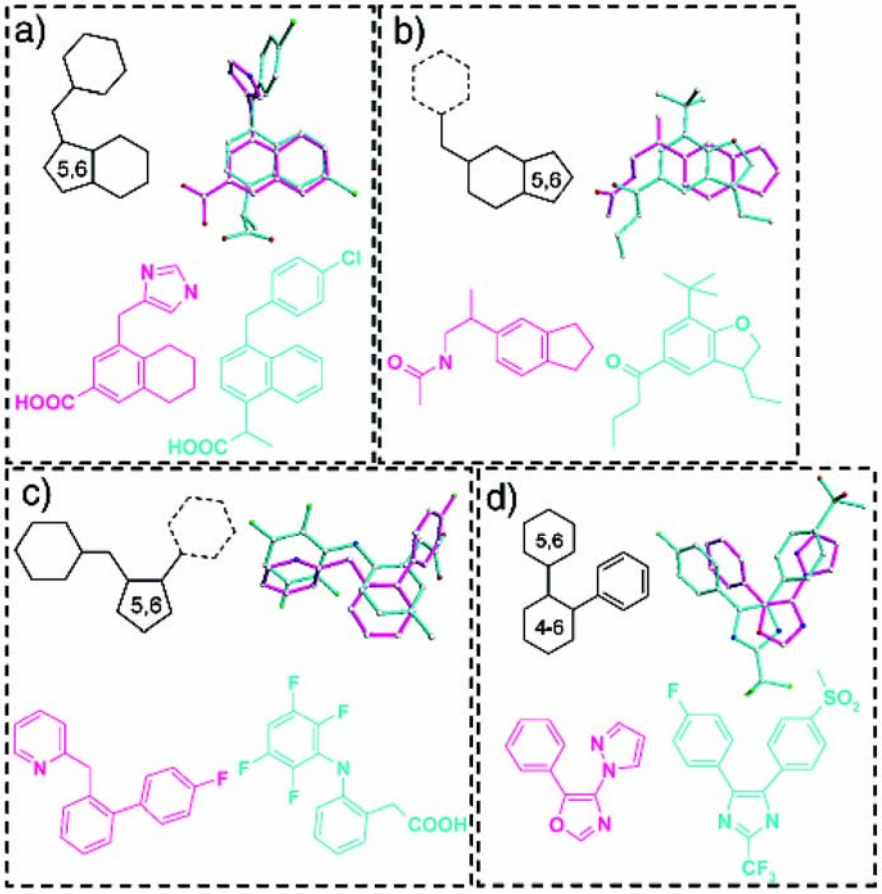


Assessment of multipose MM-PBSA.
random selection (black, dashed)
ideal performance (black,solid)
FRED/ChemScore ranking (**red**)
MM-PBSA (MAB*) ranking using the top scored docking pose (**blue**)
MMPBSA (MAB*) ranking using the three highest scored docking poses (**magenta**)



Comparison of MM-PBSA vs MD-PBSA performance for five different proteins.
random selection (black, dashed)
ideal performance (black, solid)
FRED/Chem-Score ranking (**red**),
MM-PBSA (MAB*) ranking (**blue**),
MDPBSA ranking with the MAB force field (**green**)
MD-PBSA ranking with GAFF (**light green**)

Lead Optimization \ De Novo Design



Superimposed pairs of modeled COX-2 ligands (cyan) and molecules designed by Skelgen (magenta). 2D representations of the general inhibitor topologies are displayed in black. Numbers inside the rings indicate that active ligands with alternative ring sizes are known. The MM-PBSA ranking of the Skelgen structures are: (a) 2, (b) 3, (c) 5, (d)11.

Journal of Medicinal Chemistry, 2005, Vol. 48, No. 12

Findings of the Paper

- MM-PBSA on a single structure:
  - post docking filter which enrich virtual screening results
  - tool to rank de novo design solutions
  - distinguisher between strong and weak binders
    - ΔpIC50 ≥ 2-3
    - not small free energy differences
- MD-PBSA did not improve ranking

Journal of Medicinal Chemistry, 2005, Vol. 48, No. 12