Workshop on Structure-based ligand design: *Lecture 2: COMBINE Analysis Overview*

Rebecca C. Wade

European Media Laboratory
Heidelberg

rebecca.wade@eml.villa-bosch.de
http://www.eml.org/english/Research/MCM

EML
MCM
EMBL

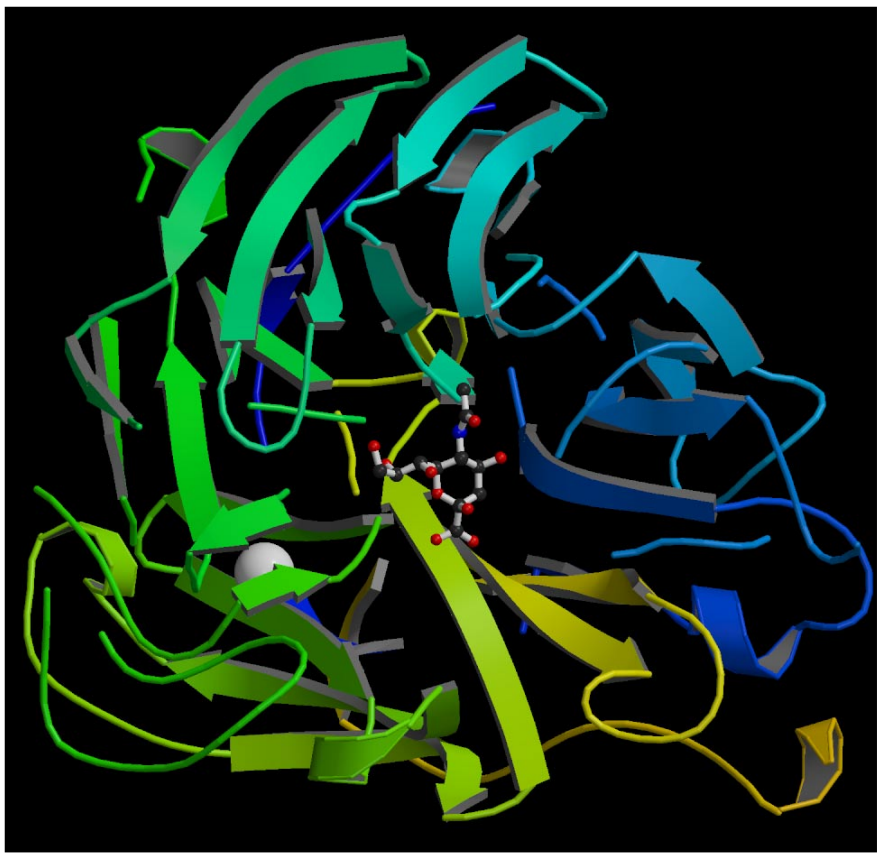
Workshop schedule

- **Lecture 1:** Introduction to Structure-based Drug Design, (GRID and 'flu)
- **Practical 1:** GRID
- **Lecture 2:** COMBINE Analysis overview
- **Lecture 3:** Molecular modeling for COMBINE
- **Practical 2:** COMBINE Analysis- molecular modeling
- **Lecture 4:** Chemometric analysis for COMBINE
- **Practical 3:** COMBINE Analysis- chemometrics
- **Lecture 5/Demo/Discussion/Practical 4**

Lecture 2: Overview

- What is COMBINE analysis?
- When can it be used?
- How does it work?
- What can it be used for?

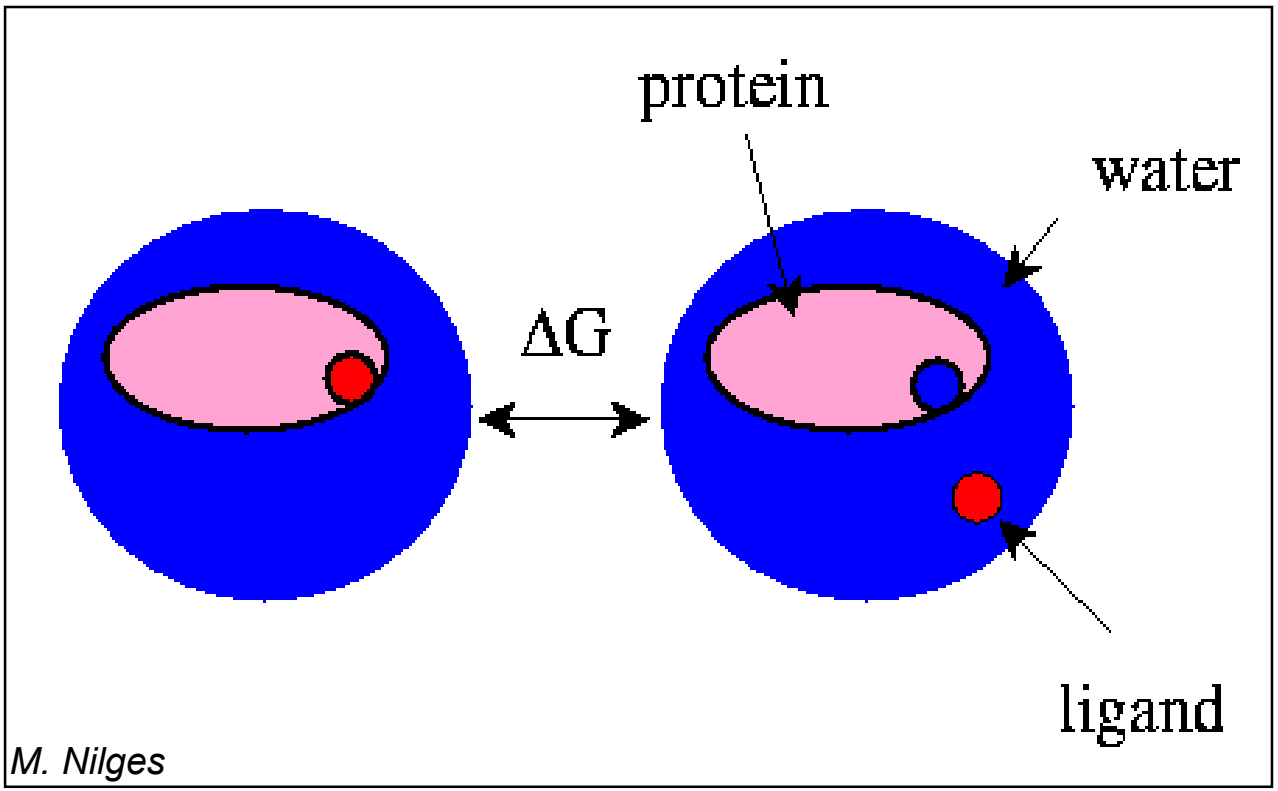
Influenza neuraminidase inhibitors



- 43 complexes:
  - 29 inhibitors: sialic acid TS and benzoic acid derivatives
  - N9 + N2 subtypes + mutants
- **Structures + Activities**
- **How can we estimate activity of new compounds?**

Binding Free Energy

Binding affinity: $K = \exp(-\Delta G/kT)$



Binding free energy is determined by differences in enthalpy and entropy of the free and bound states

Problem: How Can We Estimate Binding Affinities?

- "Absolute" binding free energies:
  - "First principles" physical models
    - MD FEP/TI with molecular mechanics model
    - MM-PB/SA
      - **Computationally demanding**
  - Empirical scoring functions
    - **Approximate**
- "Relative" binding free energies:
  - Quantitative structure-activity relationships (QSARs)
  - Use experimental binding data for deriving model
  - Specific, rather than general, model
    - **Comparative Binding Energy (COMBINE) analysis**
      - **Computationally efficient (1 structure/complex)**

Problem: How Can We Derive QSARs?

- Use experimental activity data for deriving model
- Classical QSAR:
  - Use information from formulae of ligands only
- 3D QSAR (CoMFA, CoMSIA, GRID/GOLPE, GRIND):
  - Use 3D structures of ligands only
- **Comparative Binding Energy (COMBINE) analysis:**
  - Use 3D structures of receptor-ligand complexes
  - Use physically meaningful variables that are directly relevant to binding

Comparative Binding Energy (COMBINE) Analysis

- COMBINE: data + techniques
  - 3D macromolecular structure + experimental binding data
  - Empirical molecular mechanics energies + chemometric PLS

$$\Delta G = \sum_i w_i \Delta u_i + C$$

$$\Delta G = \sum_i w_i^{vdw} u_i^{vdw} + \sum_i w_i^{ele} u_i^{ele} + C$$

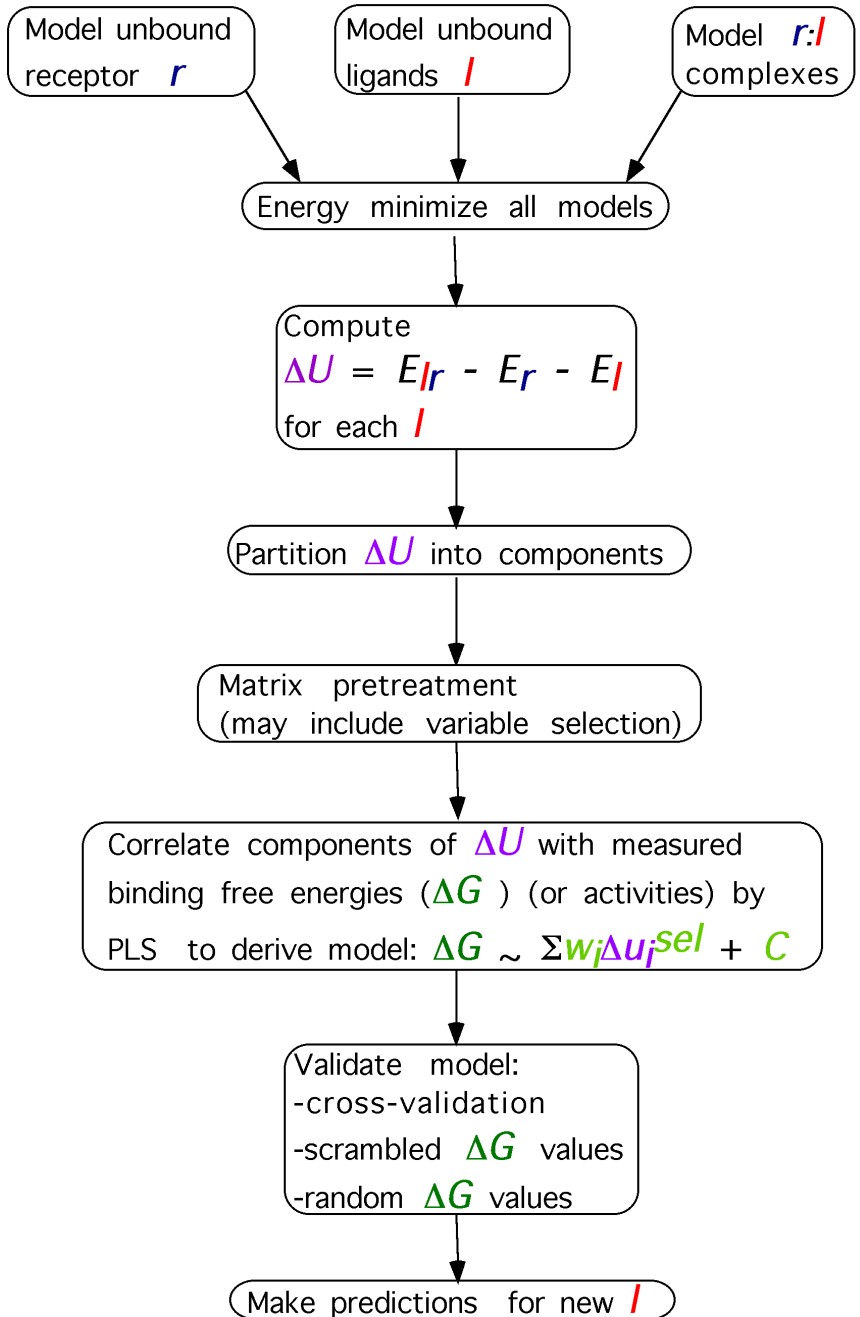
Ortiz,A.R., Pisabarro,M.T., Gago,F. Wade,R.C. *J. Med. Chem.* (1995) 38, 2681
Wade,R.C., Ortiz,A.R., Gago,F. *Persp. Drug. Disc. & Des.* (1998) 9, 19

Comparative Binding Energy (COMBINE) Analysis

- Decomposition of the energy change upon binding
  - Chemical and spatial location
  - Physical type

- Energy terms Δu_i:
  - **Intermolecular van der Waals/Lennard-Jones**
  - **Intermolecular electrostatic (Coulombic**/Poisson-Boltzmann)
  - Intramolecular non-bonded and bonded
  - Molecular desolvation (surface area/Poisson-Boltzmann)
  - Conformational entropy (side-chain rotamers)

Flowchart For Combine Analysis



Application of COMBINE Analysis

- What you need:

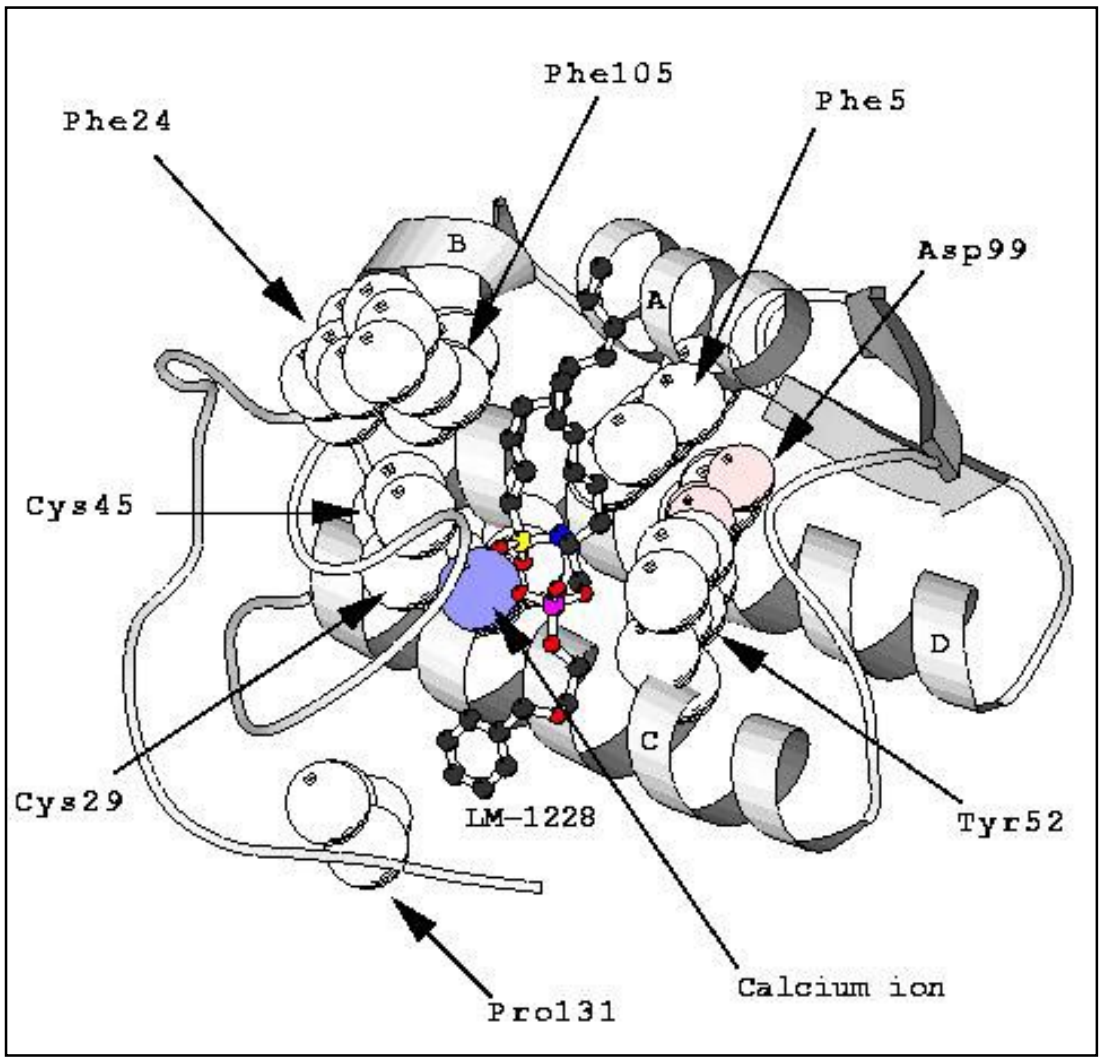
- To be able to generate energy-minimized structures of receptor-ligand complexes
  - Experimentally determined structures
  - Modelled structures
    - COMBINE Analysis can help distinguish correct models

- Experimental activity/binding data
  - Number of measurements is dependent on accuracy
  - Approx 10 usually permit a COMBINE model to be generated

Applications of COMBINE Analysis

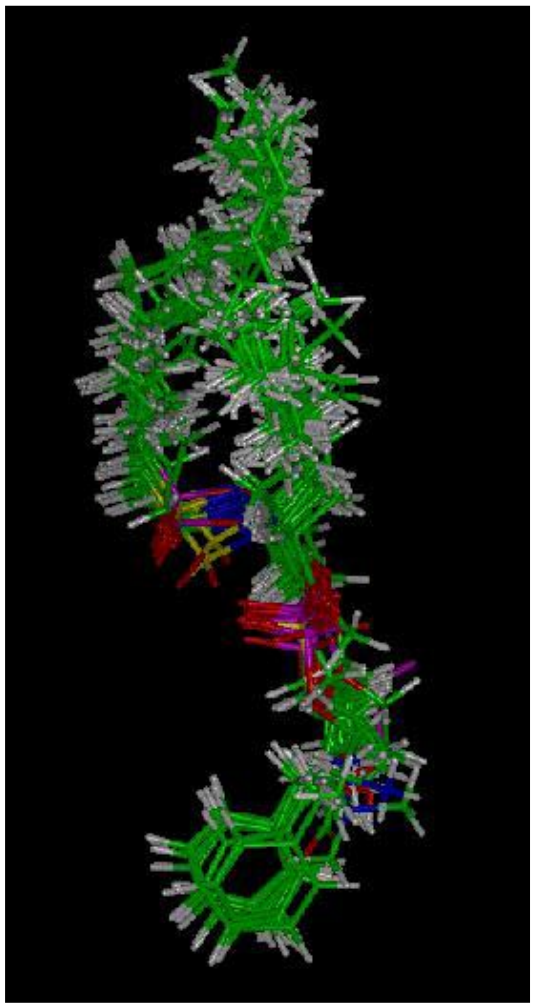
- Enzyme inhibition:
  - human synovial fluid phospholipase A_2
  - HIV-1 protease
  - glycogen phosphorylase
  - human neutrophile elastase
  - influenza neuraminidase
- Enzyme-substrate specificity:
  - cytochrome P450 1A2
  - haloalkane dehalogenase
  - lipase
- DNA-Transcription factor:
  - glucocorticoid nuclear receptor
- Receptor-peptide binding:
  - OppA-peptide

Human synovial fluid-phospholipase A_2 inhibitors



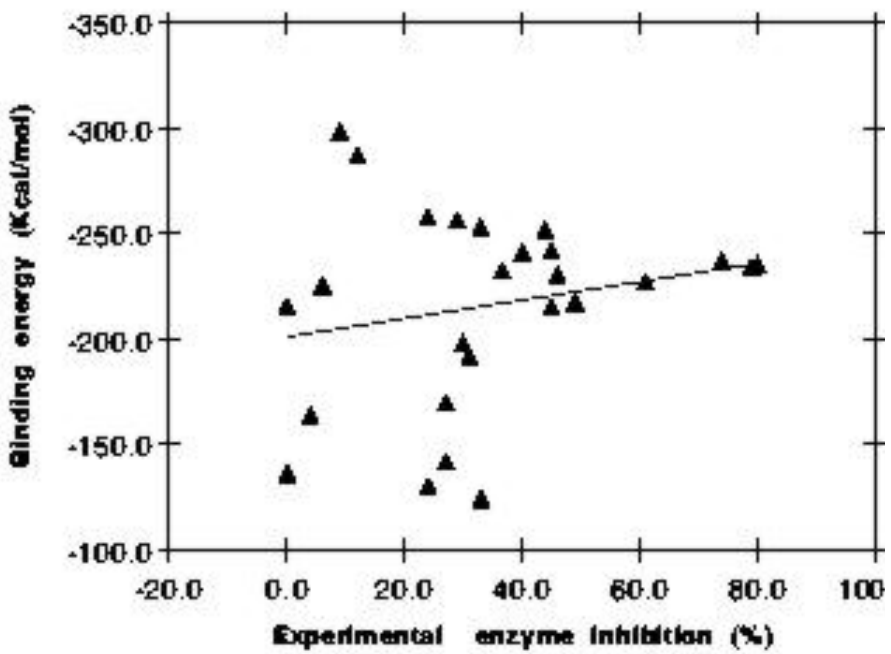
Ortiz,A.R., Pisabarro,M.T., Gago,F. Wade,R.C. *J. Med. Chem.* (1995) 38, 2681
Ortiz,A.R., Pastor, M., Palomer,A., Cruciani,G., Gago,F., Wade,R.C. *J. Med. Chem.* (1997) 4, 1136-1148,4168.

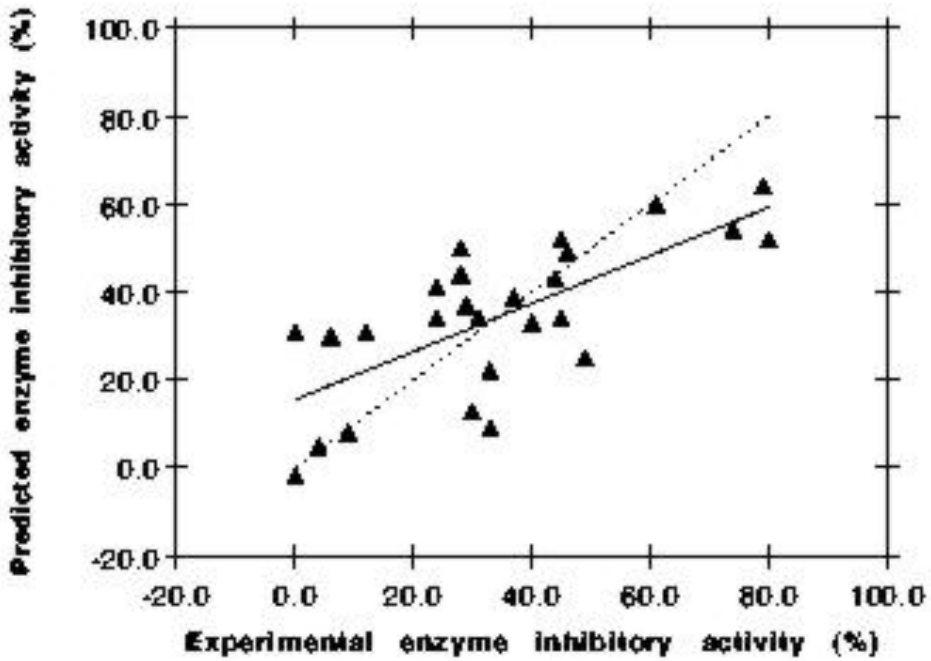
26 HSF-PLA2 inhibitors



- Inhibitors superimposed in active site
- Complexes and free protein and free ligands energy minimized
- Energy components computed

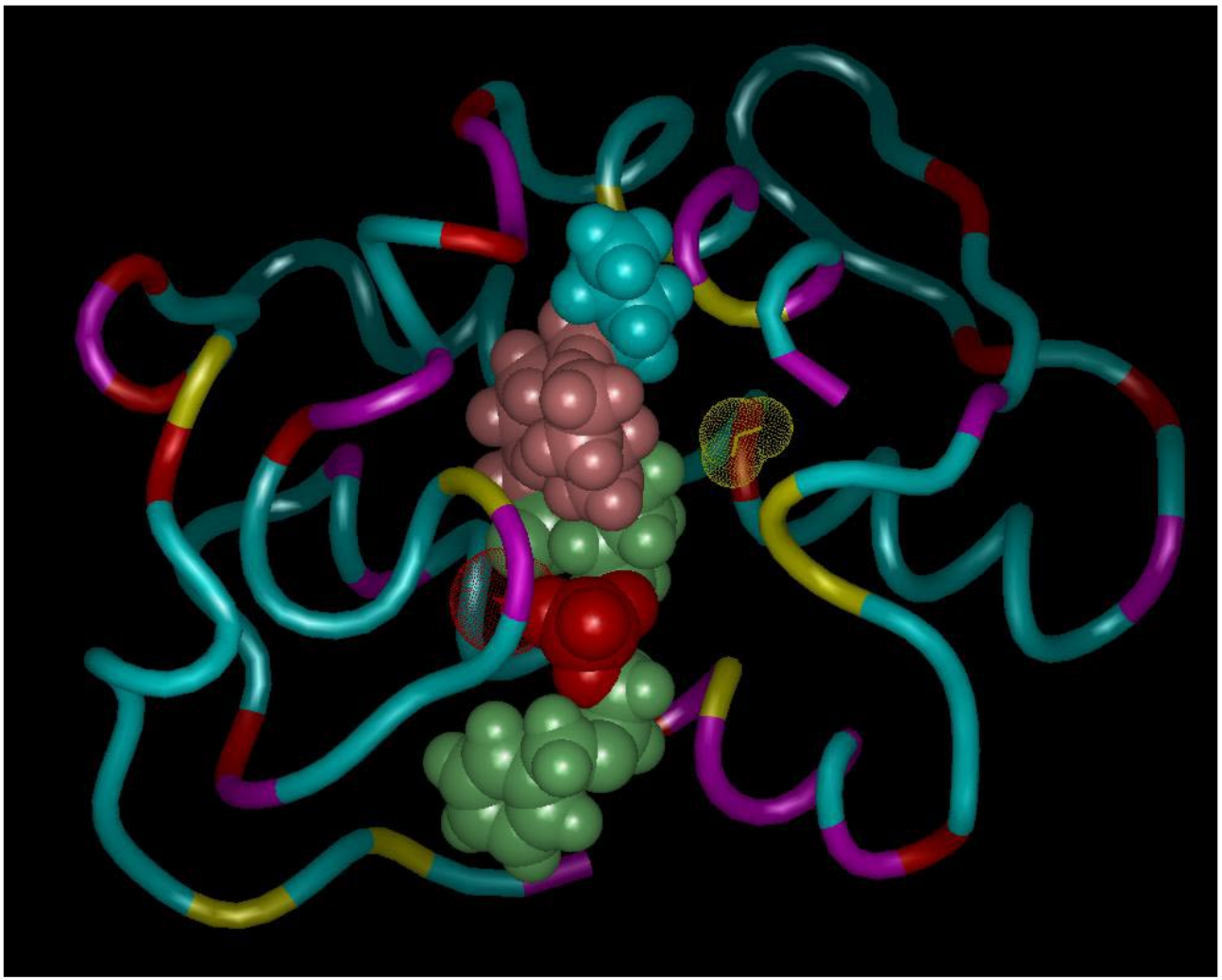
COMBINE model for HSF-PLA2 inhibitors





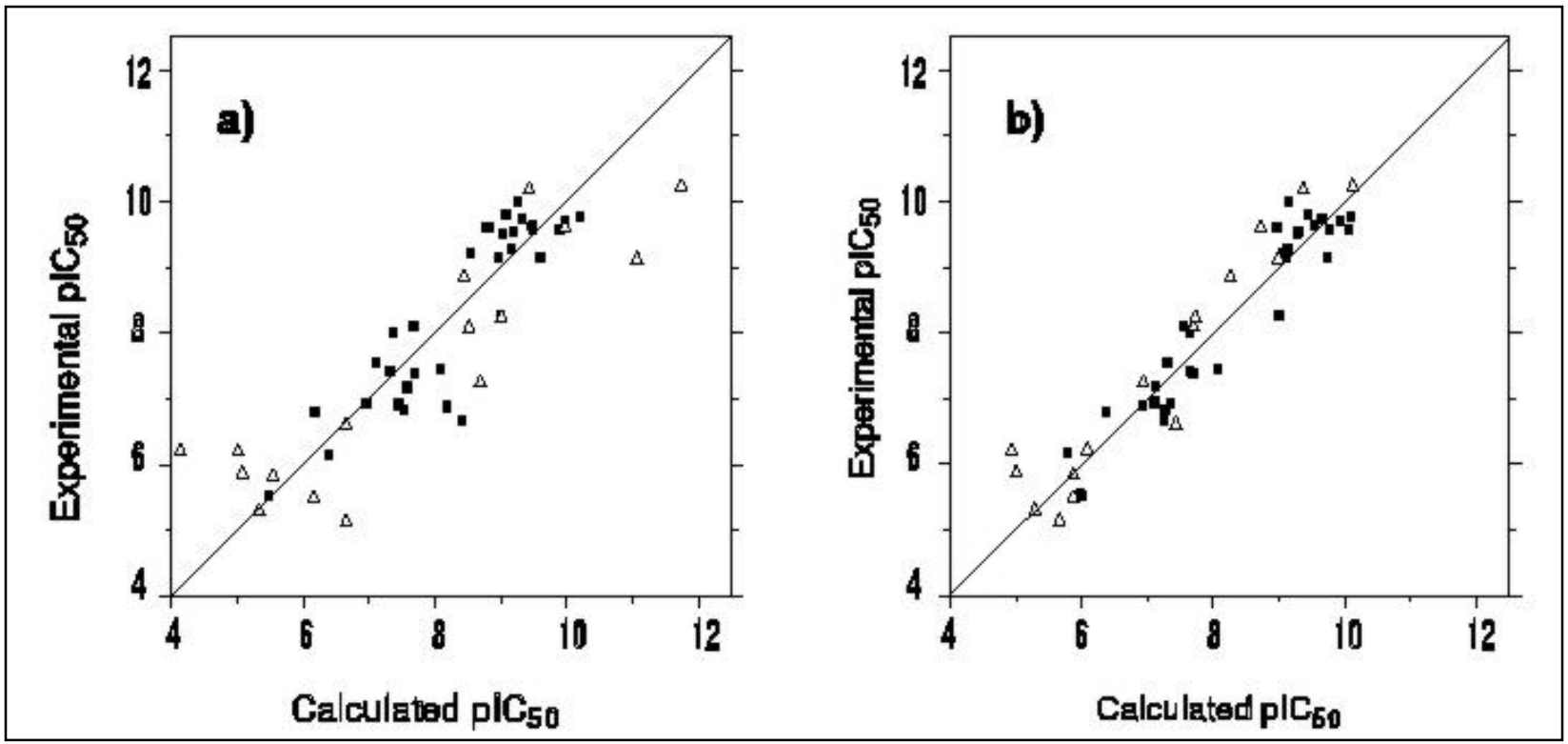
- Binding energy poorly correlated with enzyme inhibition (r=0.21)
- COMBINE → externally predictive model (r=0.71), q^2=0.82; q^2_{excess}=0.59

Important interactions for HSF-PLA2 inhibitors



HIV protease inhibitors

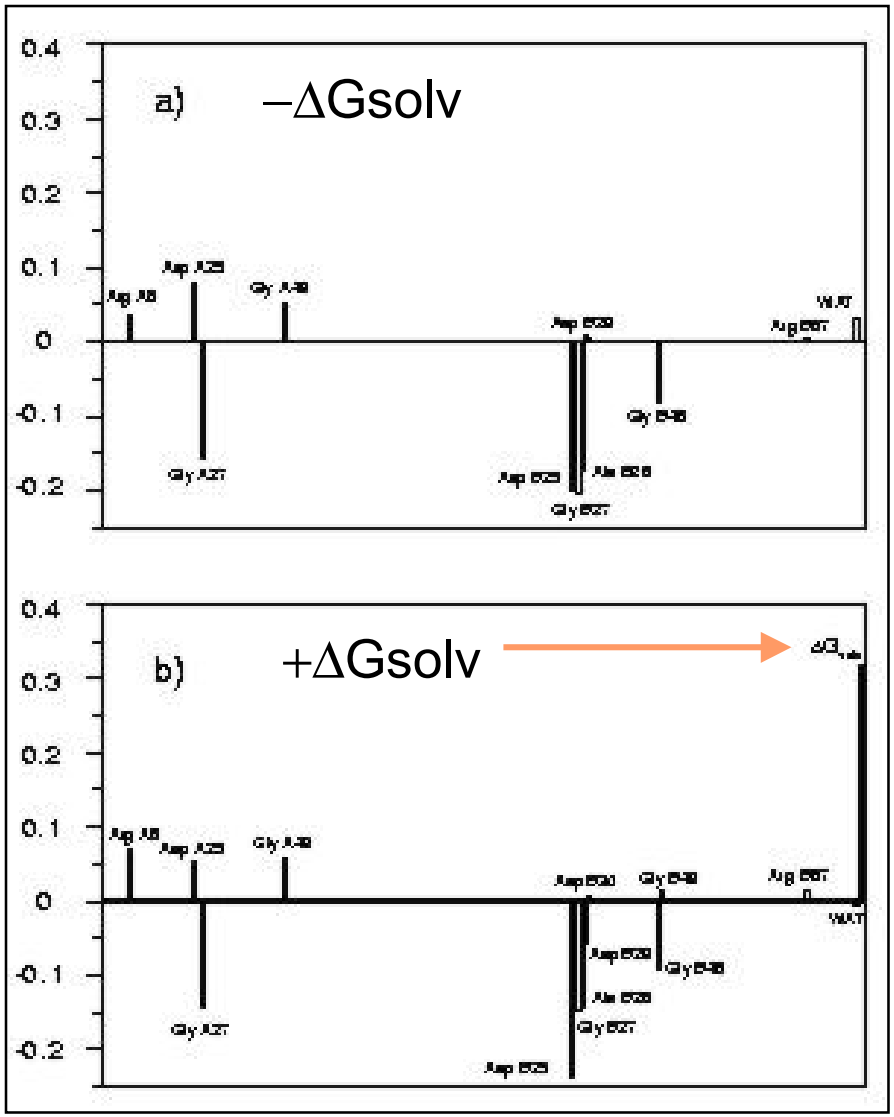
Merck dataset (Holloway et al.):33 training ■, 16 test Δ



- Total intermolecular energy:
- R2=0.79; Q2=0.77; SDEPext=1.08
- COMBINE, 2LV, continuum el.
- R2=0.90; Q2=0.72; SDEPext=0.62

Perez, C., Pastor, M., Ortiz, A.R., Gago, F. *J Med. Chem.* 1998

HIV protease inhibitors: important interactions

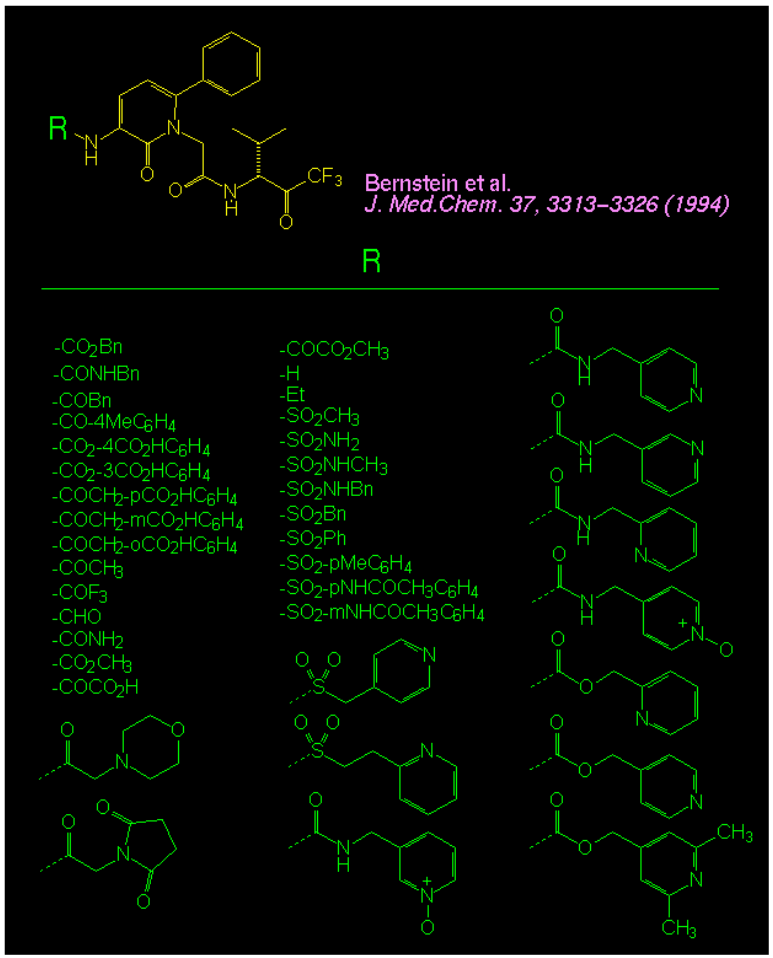


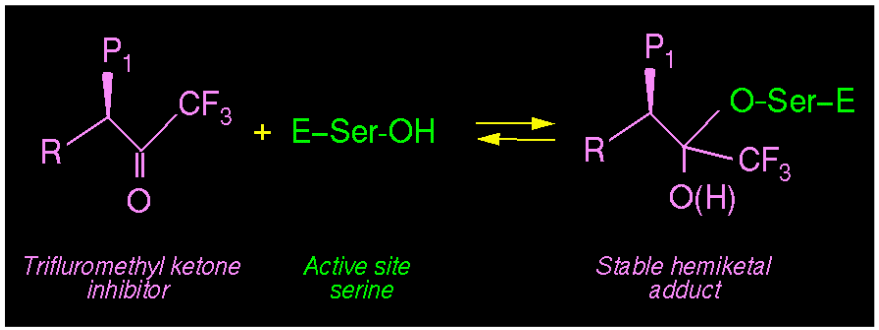
- ΔGsolv of inhibitor important: modulates interactions of charged residues
- Continuum treatment + desolvation -> modest improvement;
- Collinearities -> implicit representation

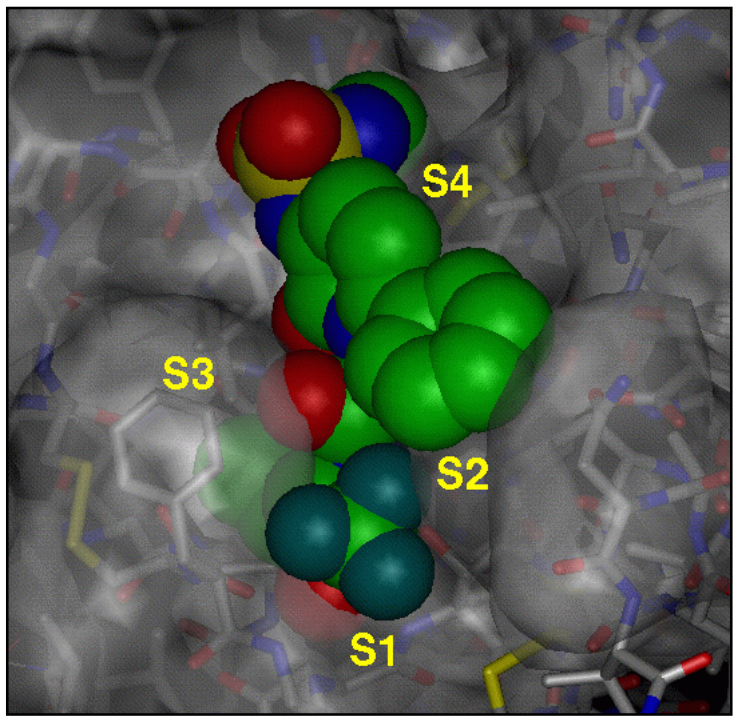
Perez, C., Pastor, M., Ortiz, A.R., Gago, F. *J Med. Chem.* 1998, 41: 836-52.

Pastor, M., Perez, C., Gago, F. *J Mol Graph Model* 1997, 15: 364-71, 389.

Human Neutrophil Elastase inhibitors (40)

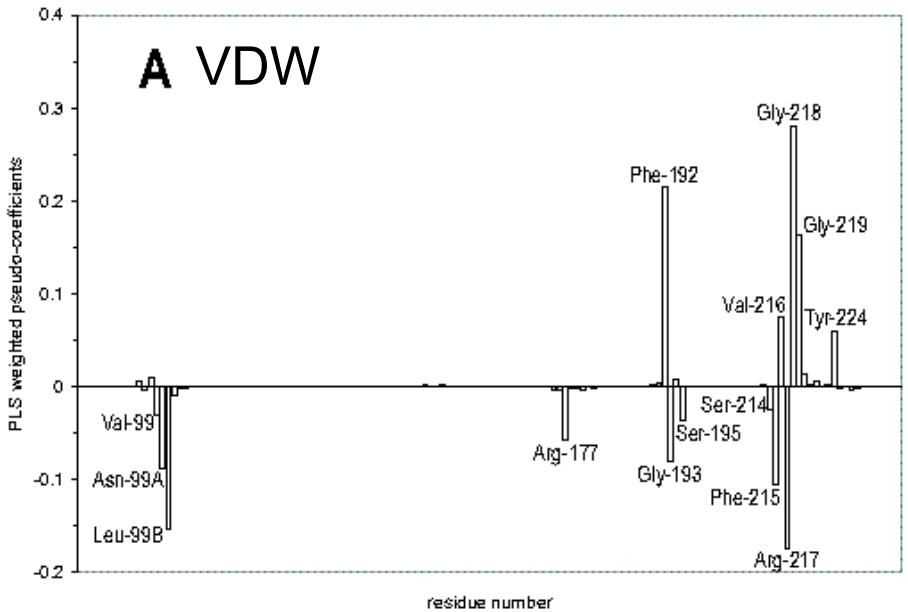


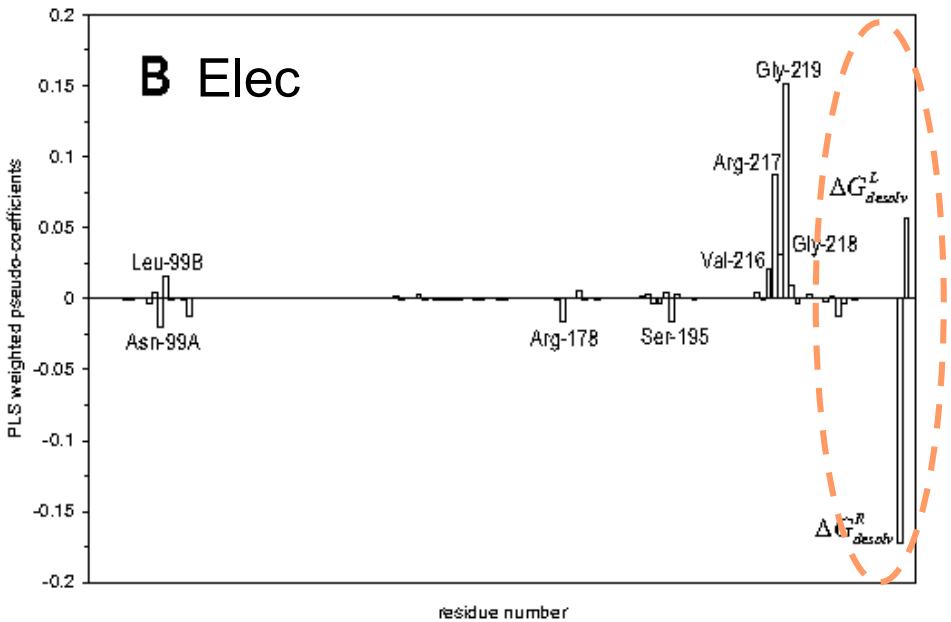




Cuevas, Pastor, Perez, Gago 2001, *Comb Chem High Throughput Screen*

Elastase inhibitors

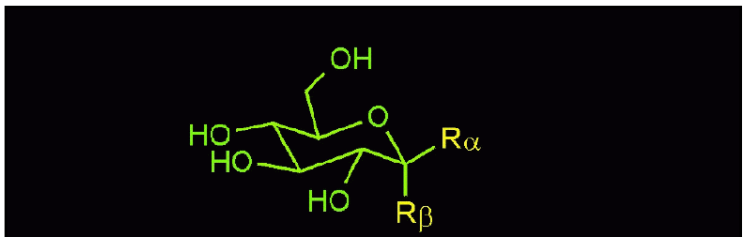




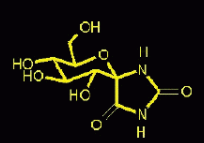
- Poor correlation of intermolecular energies with activities
- Good correlation with COMBINE, aided by incorporation of PB terms
- Electrostatic ΔGsolv of inhibitor and receptor important

Cuevas, Pastor, Perez, Gago 2000

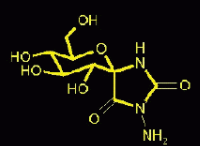
Glycogen phosphorylase inhibitors



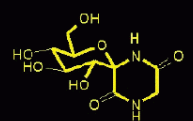
no.	Substituent at C1-position Rα	Rβ	pK_i(mM)
1	$C(=O)NH_2$	H	3.43
2	H	$C(=O)NH_2$	3.36
3	H	$COOCH_3$	2.55
4	H	CH_2CN	2.05
5	H	NHC(=O)NH2	3.85
6	$C(=O)NH_2$	$NHC(=O)OCH_3$	4.80
7	OH	H	2.77



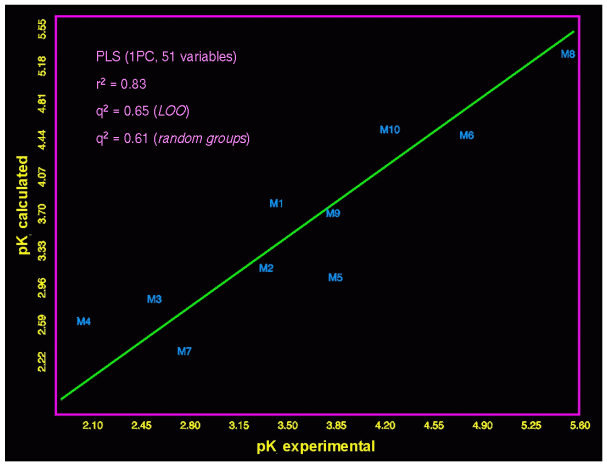
8 pK_i = 5.52 mM

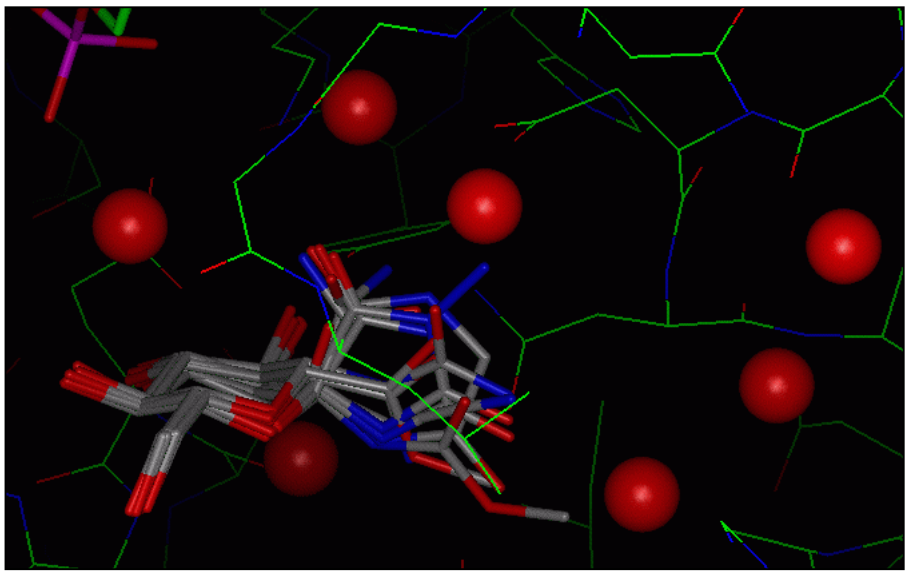


9 pK_i = 3.84 mM



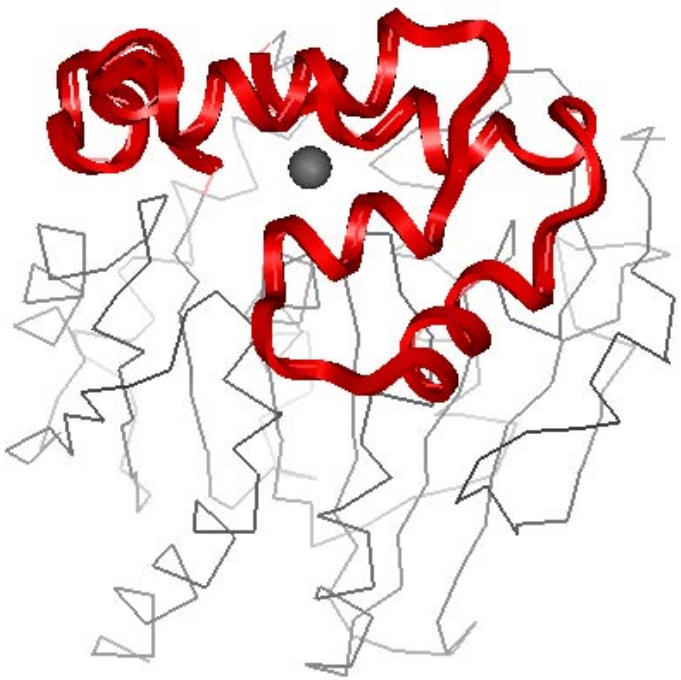
10 pK_i = 4.22 mM





Pastor,M., Gago,F. Cruciani, G. In “Molecular Modeling and Prediction of Bioactivity” Gundertofte, K., Jorgensen, F.S. (Eds.). Kluwer 2000, 329-330

Haloalkane dehalogenase-substrate binding

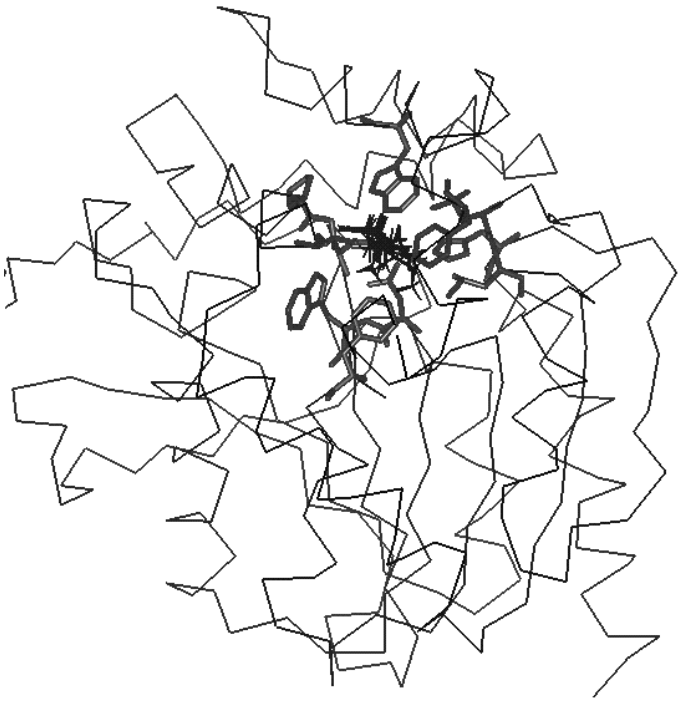


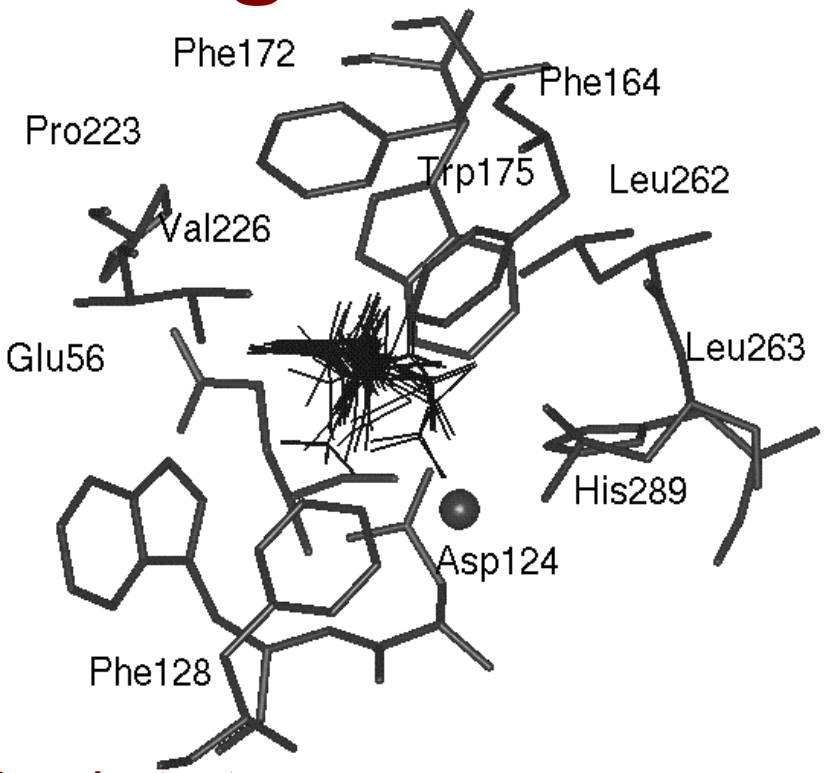
α/β-fold domain (α/β-hydrolases)
+ cap domain

- Microbial enzymes
- Degrade halogenated environmental pollutants
- Target for engineering for biotechnological applications
- COMBINE analysis to identify important interactions for substrate binding specificity

Kmunicek et al. *Biochemistry*, 2001.

Haloalkane dehalogenase-substrate binding





Xanthobacter autotrophicus GJ10 soil bacterium

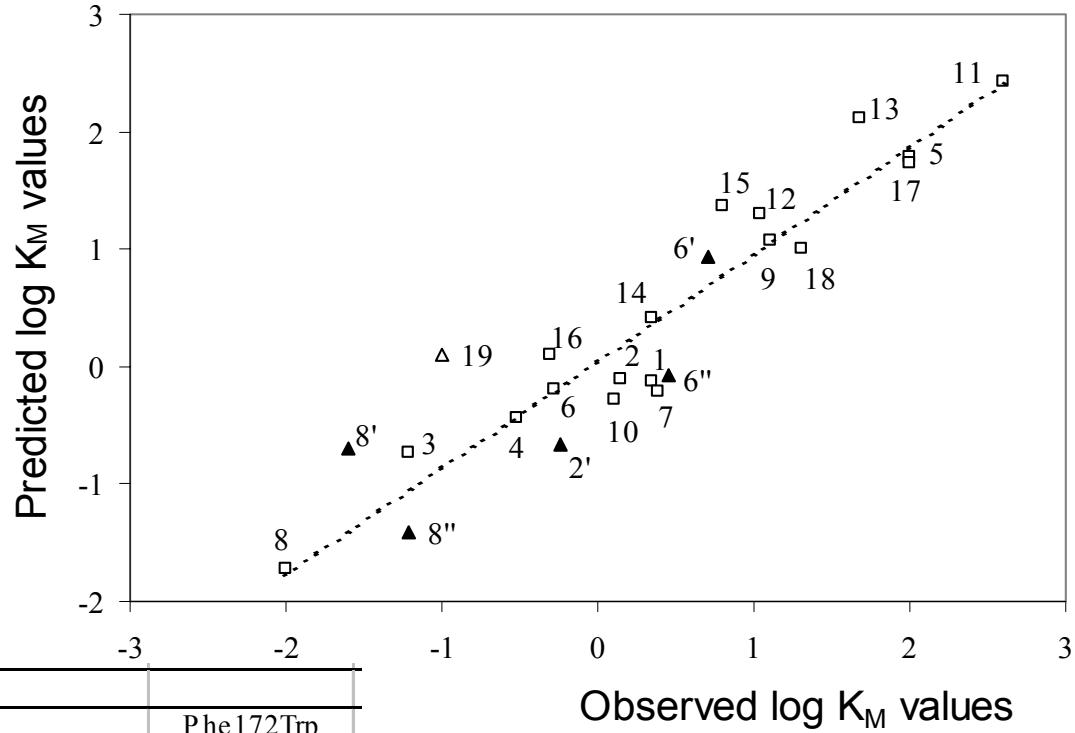
Kmunicek et al. *Biochemistry*, 2001.

18 substrates:

- 1-chlorobutane; 1-chlorohexane; dichloromethane; 1,2-dichloroethane; 1,2-dichloropropane; 2-chloroethanol; epichlorohydrine; 2-chloroacetonitrile, 2-chloroacetamide and their brominated analogs.
- Log(k_M) : -2 to +2

Haloalkane dehalogenase-substrate binding COMBINE model

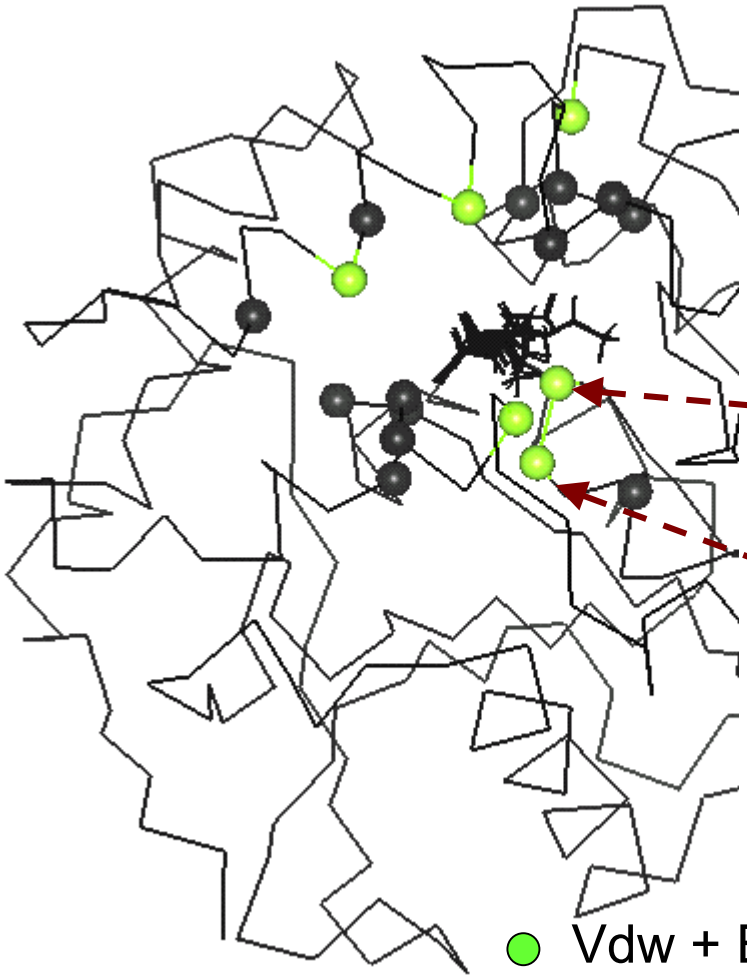
N=18, 4LV,
R^2=0.91, Q^2=0.77,
SDEP=0.59,
SDEPext =0.63

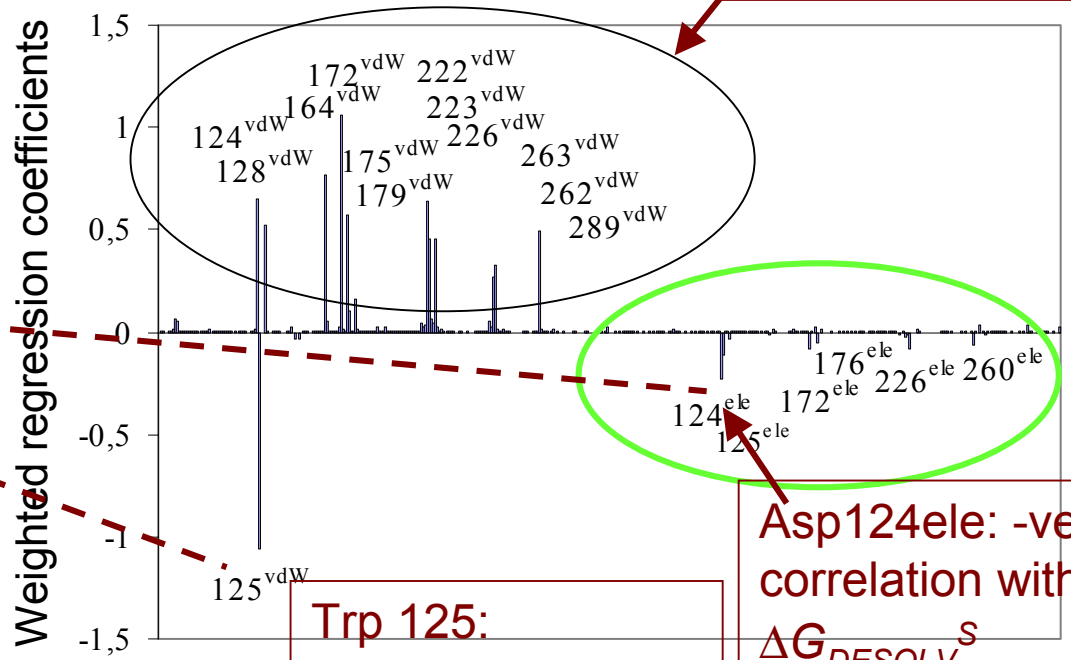


External Predictions				
		wt		Phe172Trp
no[a]	Substrate	experiment[b]	experiment[b]	prediction[d]
2'	1-chlorohexane	1,40	0,57	0,22
6'	1,2-dichloroethane	0,53	5,13	8,47
8'	1,2-dibromoethane	0,01	0,03	0,20
19'	1-bromo-2-chloroethane	0,07	0,10	1,28
		wt		Trp175Tyr
		experiment[c]	experiment[c]	prediction[d]
6"	1,2-dichloroethane	0,53	2,85	0,83
8"	1,2-dibromoethane	0,01	0,06	0,04

□ Training set
▲ Test set
‘ F172W
“ W175Y

Haloalkane dehalogenase-important interactions for substrate binding





Better VDW, Better binding

Asp124ele: -ve correlation with $\Delta G_{DESOLV}{}^{S}$

Trp 125: structural role, halide specificity

Vdw + Elec;

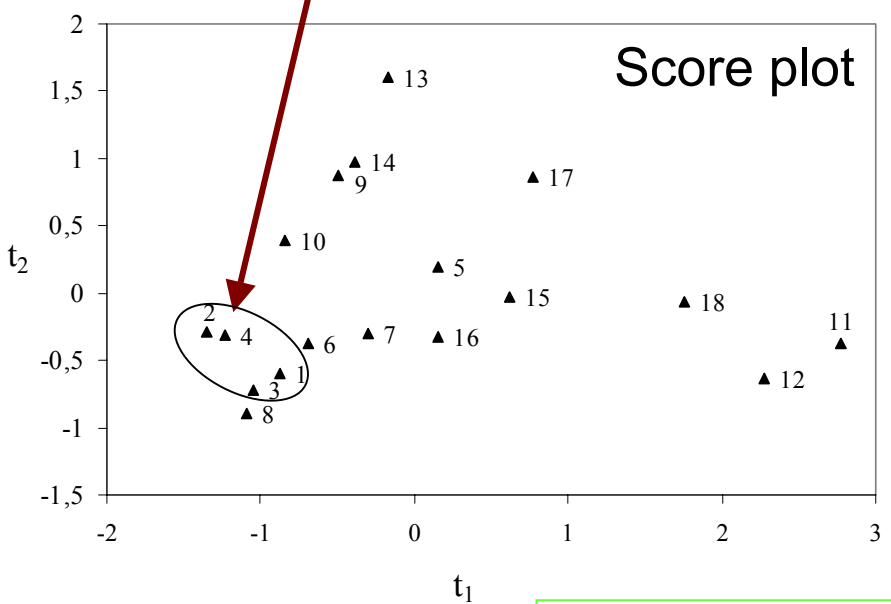
Vdw only

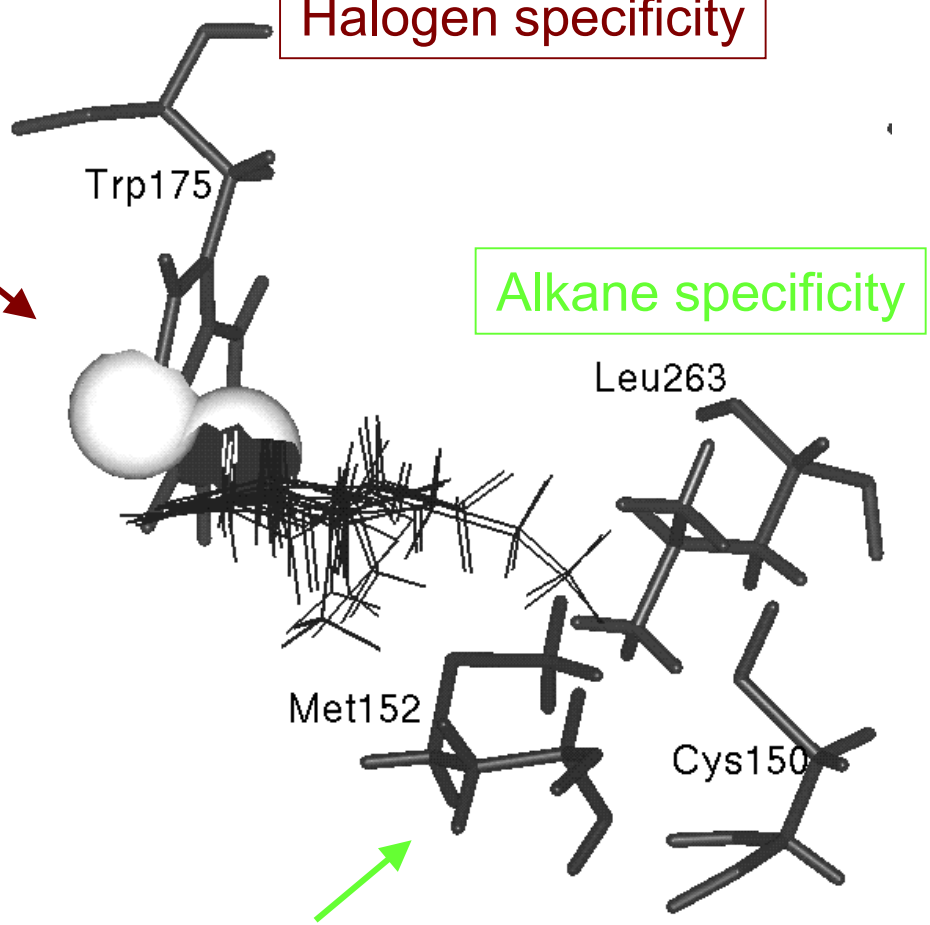
+ PB desolvation energy of ligand

Haloalkane dehalogenase-substrate binding

Halogen specificity

Unfavorable vdw interactions
With long-chain substrates

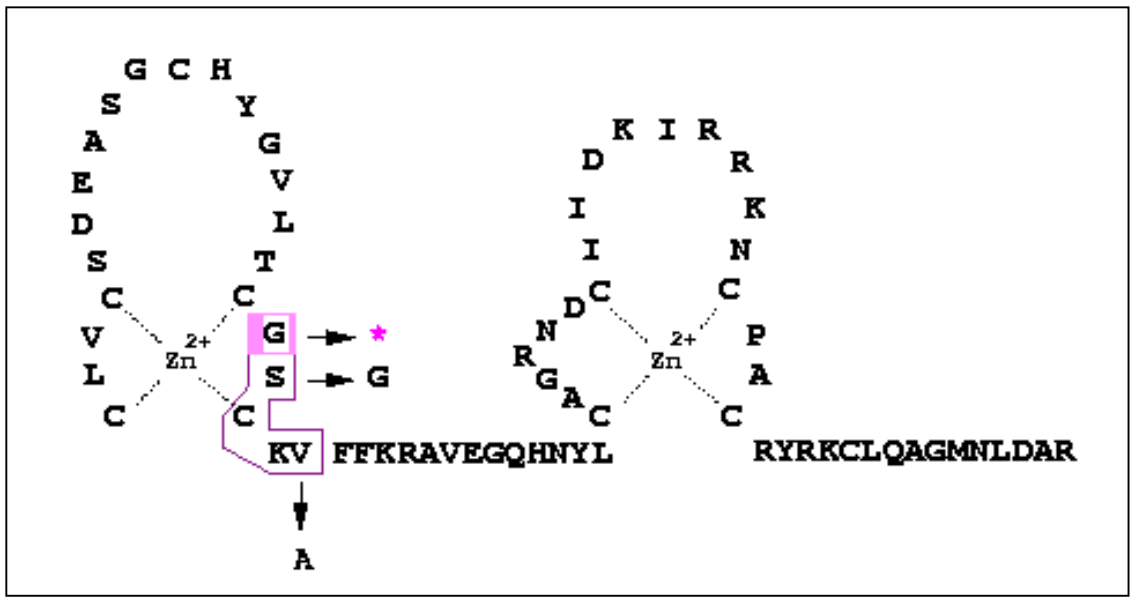


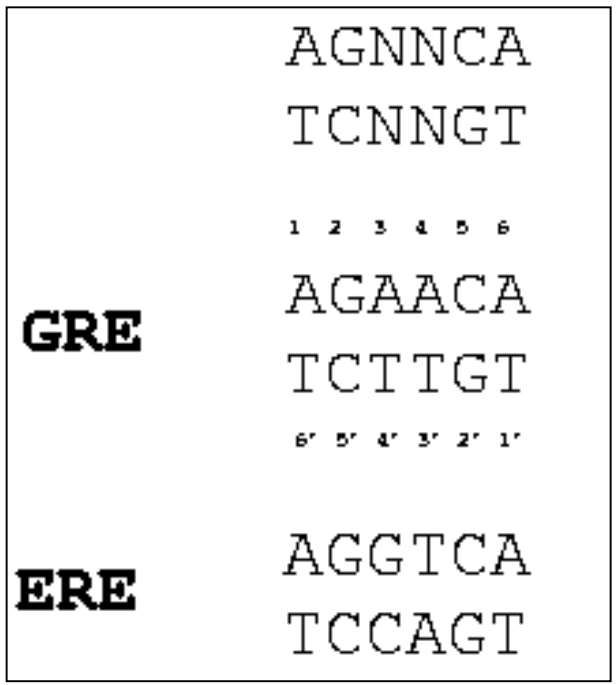


Alkane specificity

Favorable vdw interactions
With long-chain substrates

Glucocorticoid nuclear receptor DBD-DNA mutant binding:



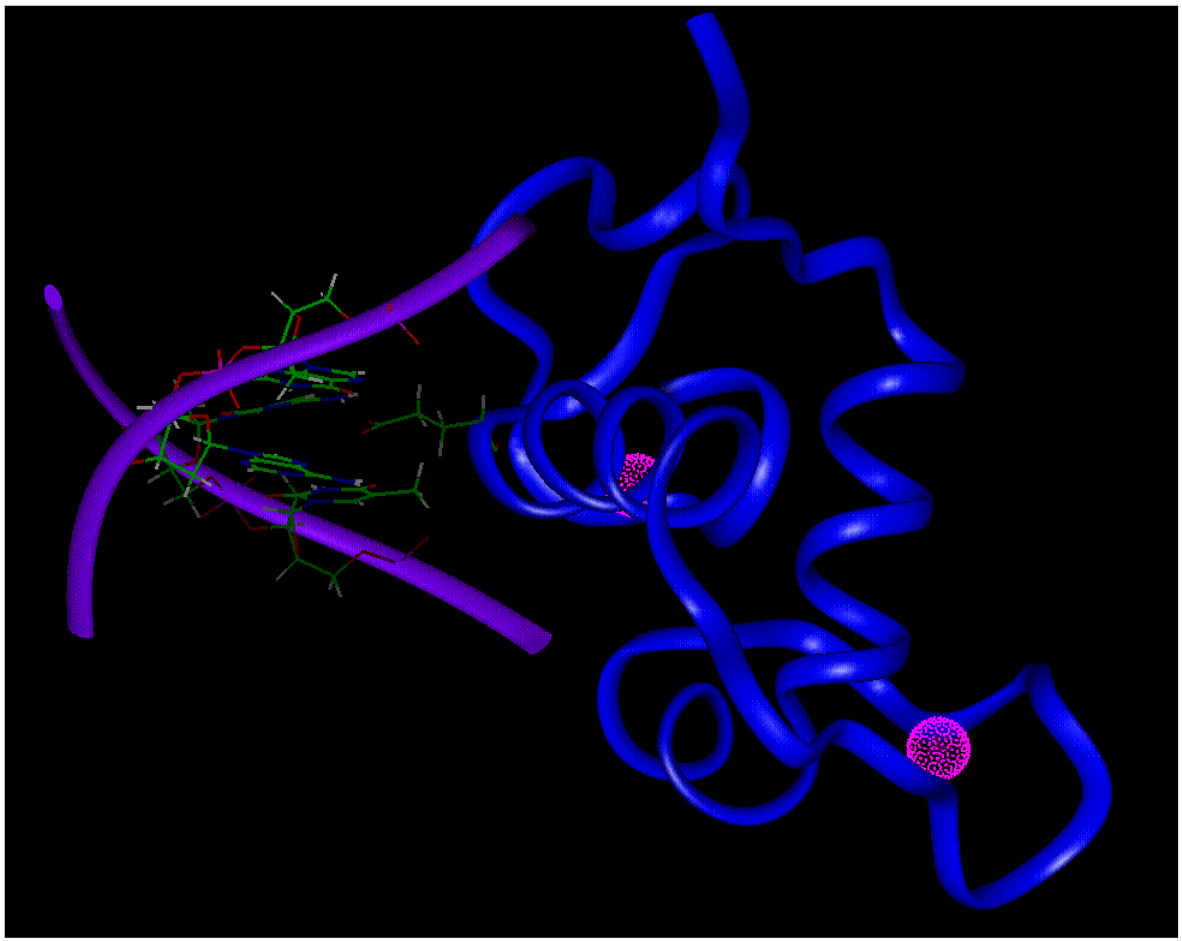


- S440G, V443A (estrogen receptor)
- G439 -> 20 natural AAs

- **Response elements:**
  - 3,4 -> 16 base-pair combinations
- **320 (20x16) complexes in total**

Transactivation assay:
dimeric DBD : 2 hexameric half-sites with 3-bp spacing (Zilliacus et al., 1995)

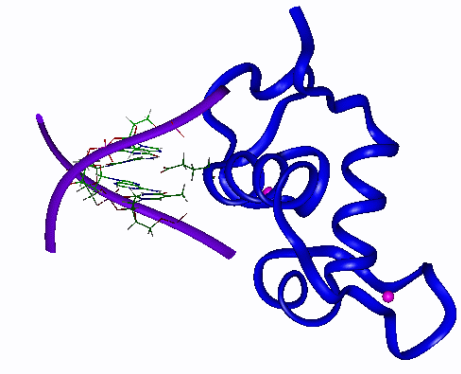
GR DBD-DNA complexes

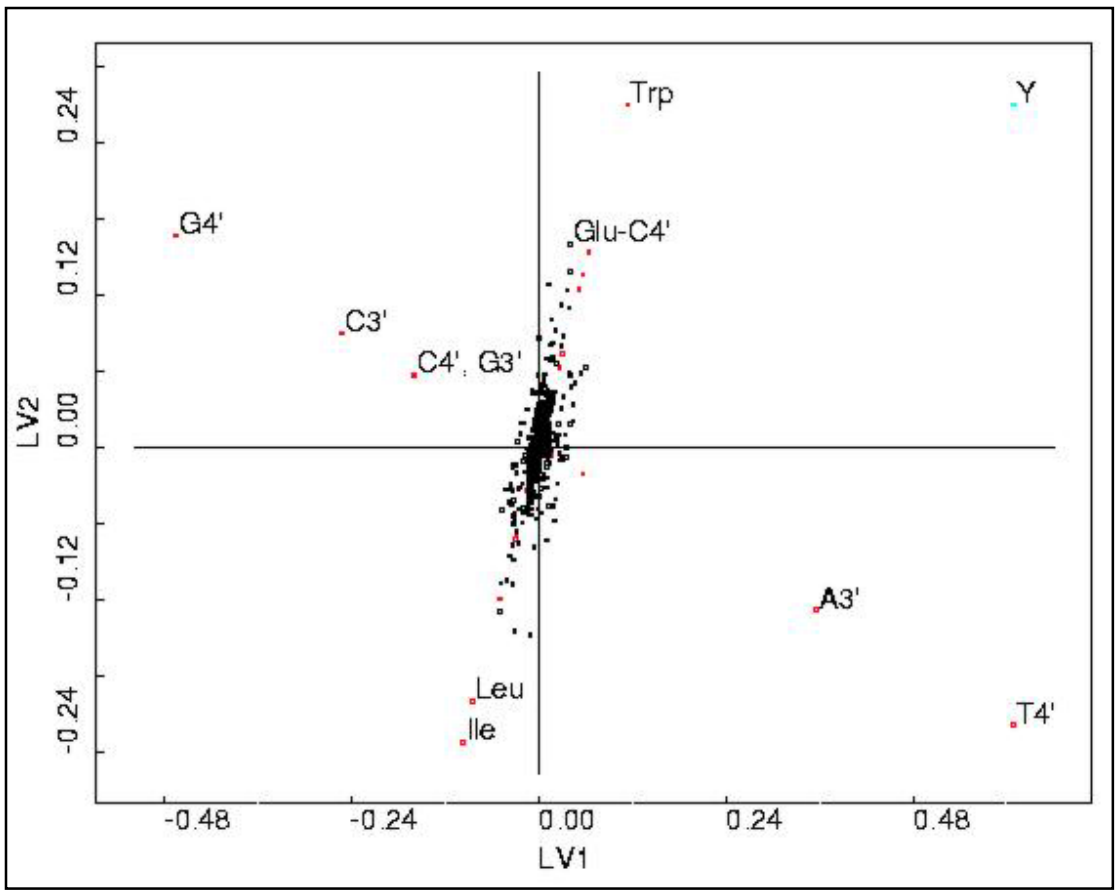


- Crystal structure of GR DBD-DNA complex (Luisi et al, 1991)
- 73-AA DBD
- 6-base pair RE
- Mutant models energy minimized with CHARMM and AMBER

Tomic,S., Nilsson,L., Wade,R.C. *J. Med. Chem.* (2000) 43, 1780-1792

Free-Wilson Analysis

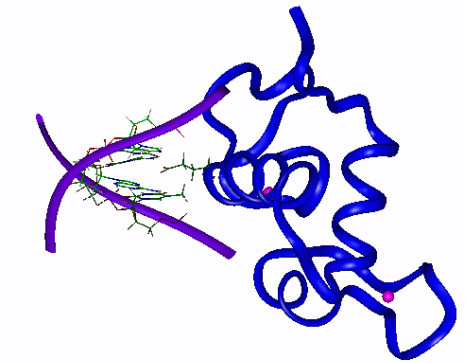




- Variables: 28 occupation + 480 association
- PLS –> additive model
- No 3D atomic structure

- Models: q^2 up to 0.5 and 1-2 LVs
- Comparable to Zilliacus et al. (1992) for different, smaller set

COMBINE Analysis

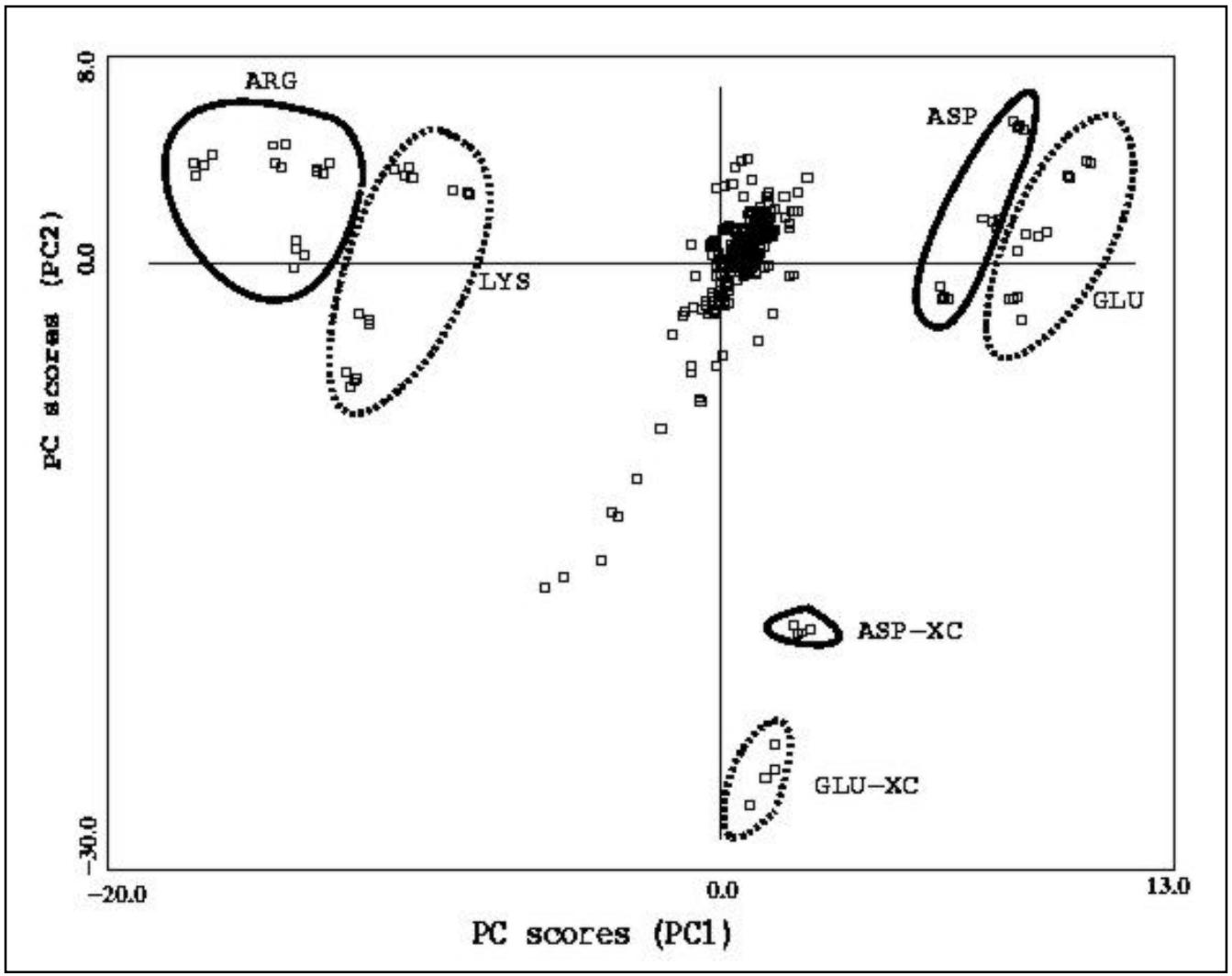


# objects	Q2: Free-Wilson	Q2: COMBINE	SDEP (320)
320	0.42	0.44	0.57
289	0.54	0.55	0.56
41	0.86	0.94	0.59
32*	~0.5	0.71	0.27

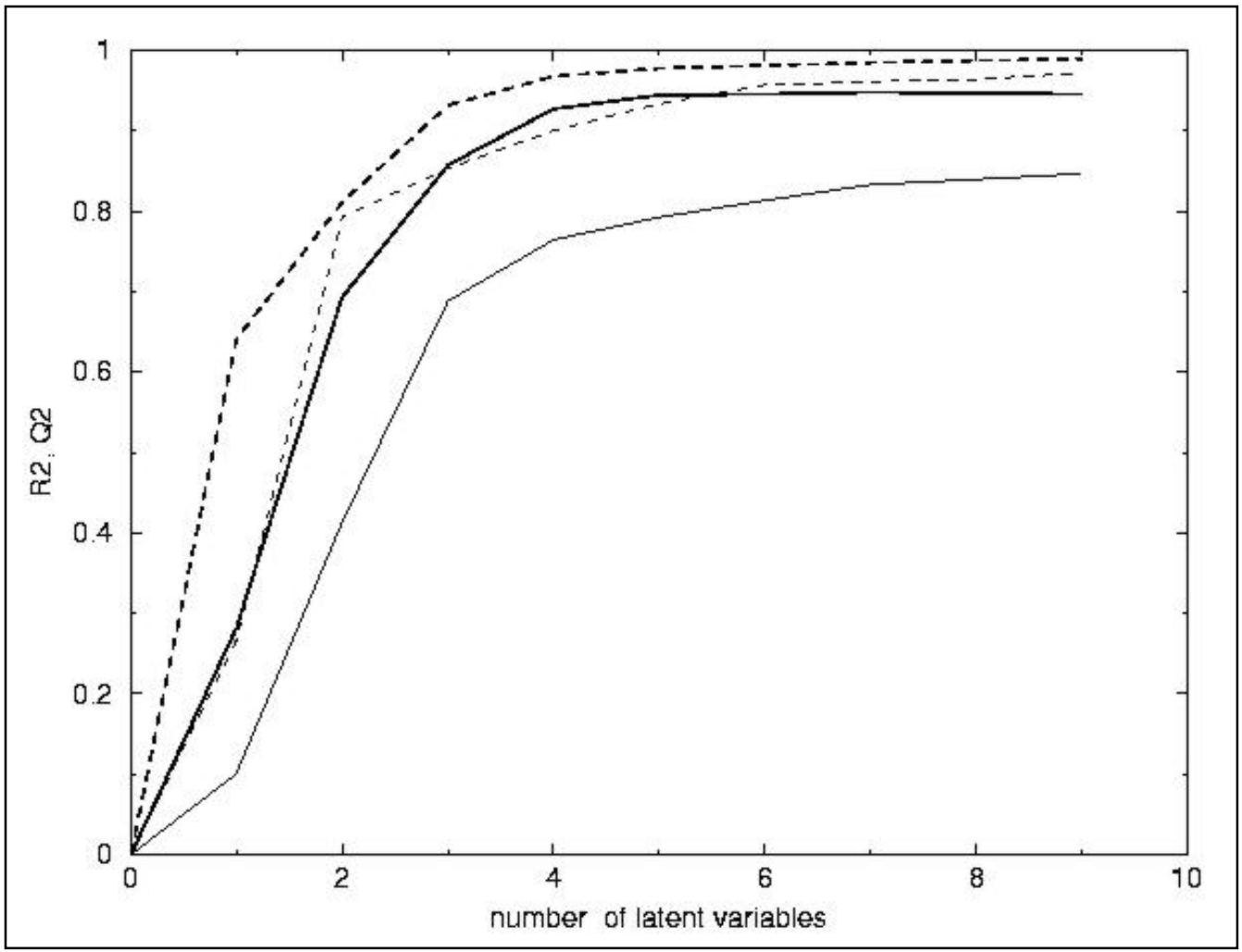
*GR-> ER: G439E, S440G, V443A; AA -> GT

Tomic,S. et al. *Med. Chem.* (2000), Croat. Chim. Acta (2001)

COMBINE analysis: principle components

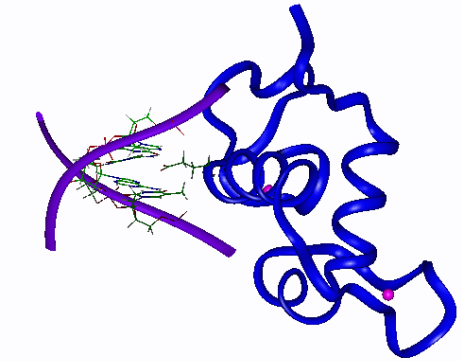


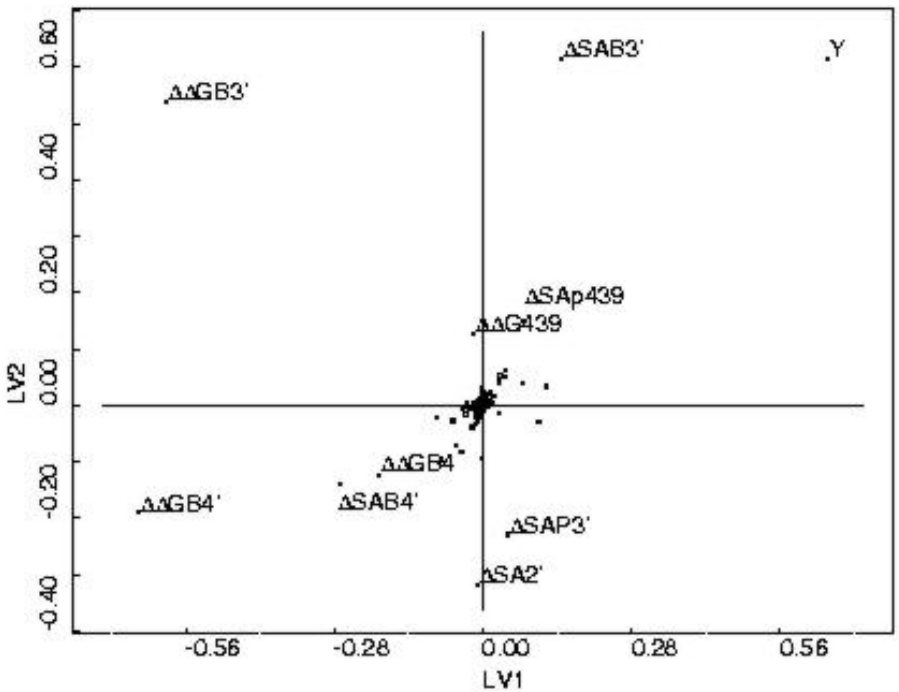
Derivation of COMBINE model

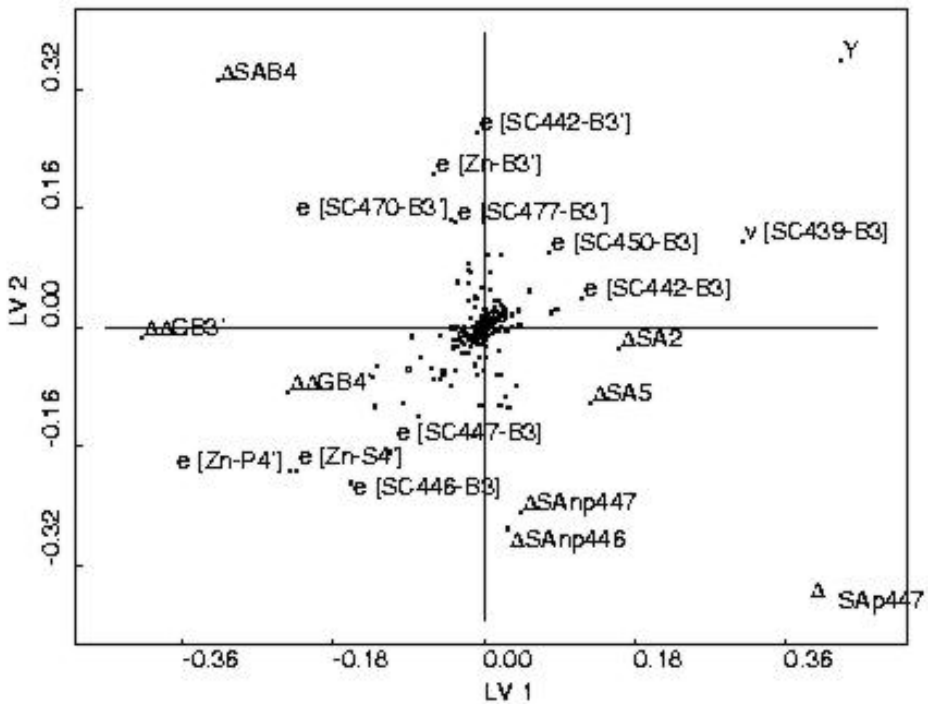


- Q^2-full
- R^2-dashed
- Variable selection: before - thin; after - bold
- for 41-object dataset

COMBINE:
intermolecular energy + ΔΔGhydr + SASA terms





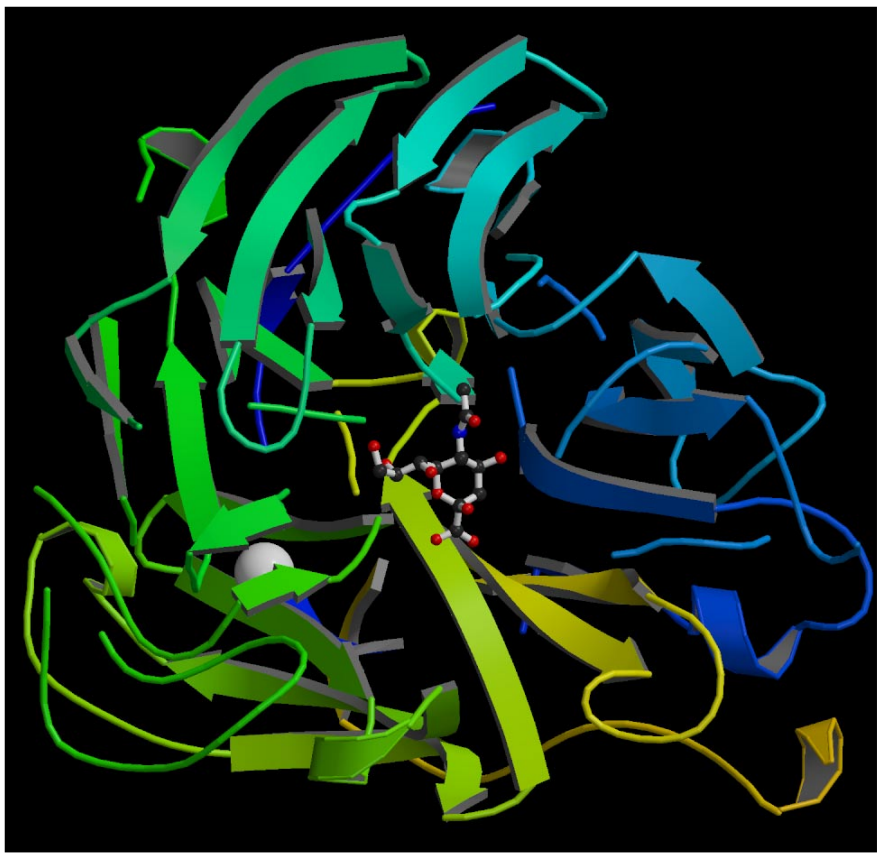


289 objects: before and after variable selection

Summary of DBD-DNA COMBINE analysis

- **Free-Wilson and COMBINE analyses consistent**
- **Best COMBINE model:**
  - intermolecular interaction energies
  - free energies of solvation
  - solvent accessible surface areas
- **Important properties for binding specificity**
  - change of solvation free energies of B3' and B4'
  - electrostatic interaction between mutated nucleotides (mostly bases) and mutated residue, Zn, several charged residues
  - changes of SASA of B4-B3' base-pair, bases 2,5 and the mutated residue
- **Additional features important:**
  - interfacial hydration, larger conformational differences

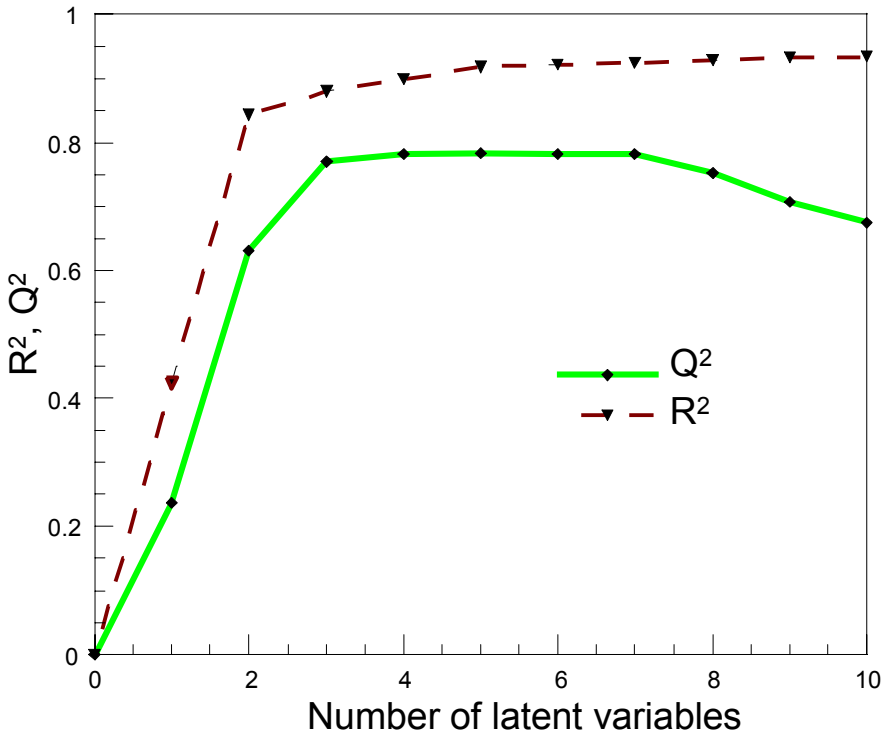
Influenza neuraminidase inhibitors

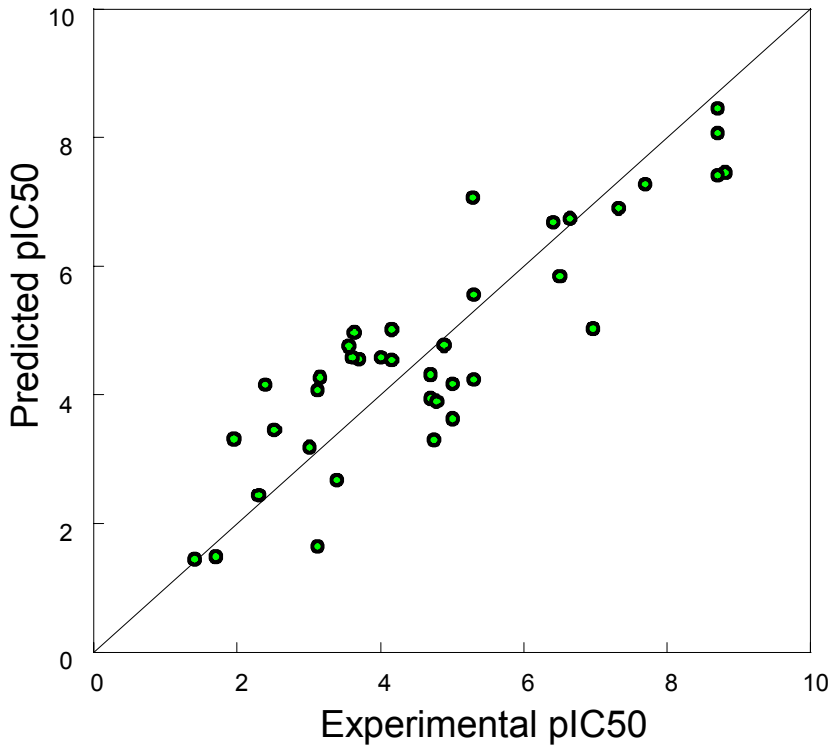


- 43 complexes:
  - 29 inhibitors: sialic acid TS and benzoic acid derivatives
  - N9 + N2 subtypes + mutants
- 32 crystal structures
- 11 docked (comparative/AUTODOCK)
- Energy minimize: AMBER

Wang,T., Wade,R.C. *J. Med. Chem.* (2001) **44**, 961-971

Influenza neuraminidase inhibitors COMBINE model

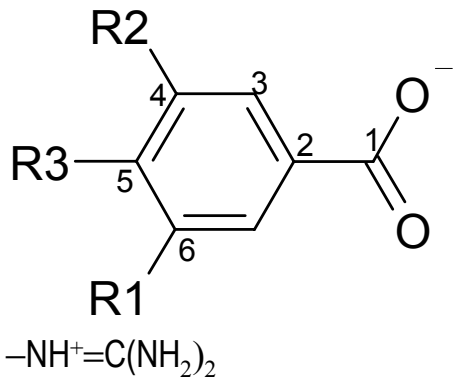


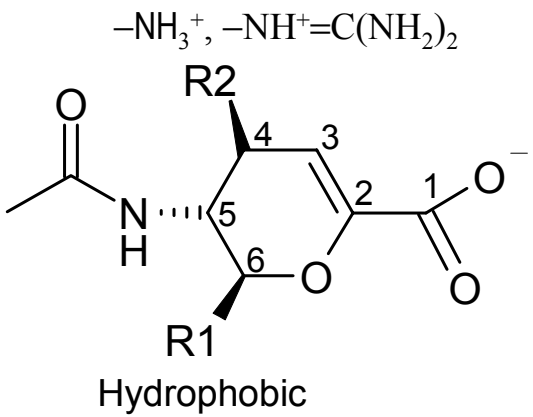


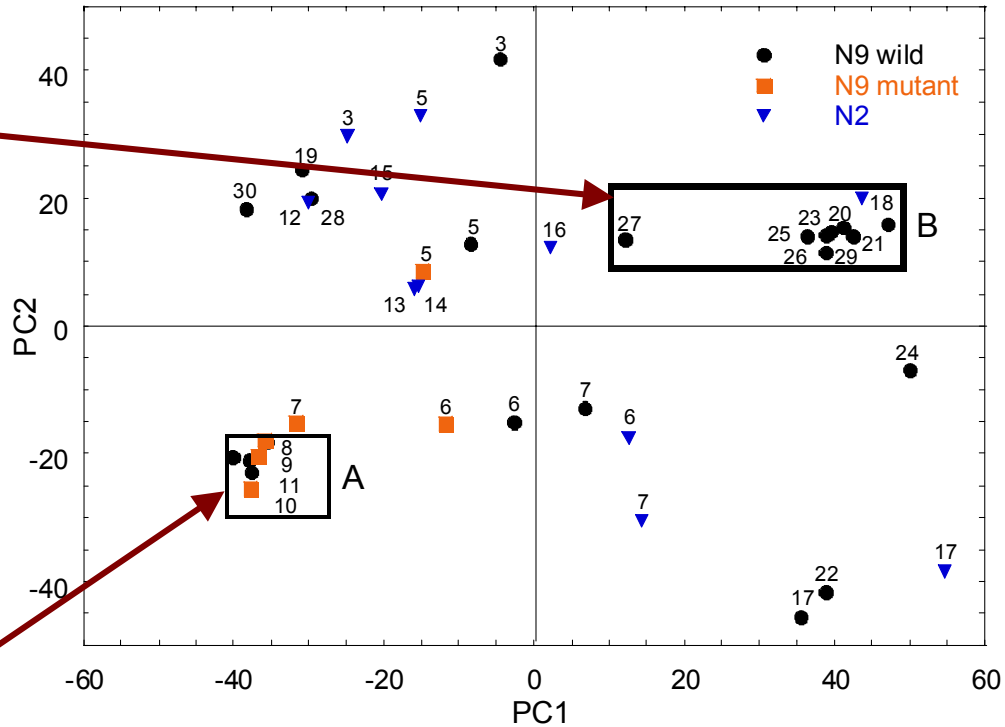
N=39 (N2 + N9), 3LV,
R^2=0.88, Q^2=0.77,
SDEP=0.96, SDEPext =1.2

Validation of docked ligand alignments

COMBINE model discrimination of influenza neuraminidase inhibitors

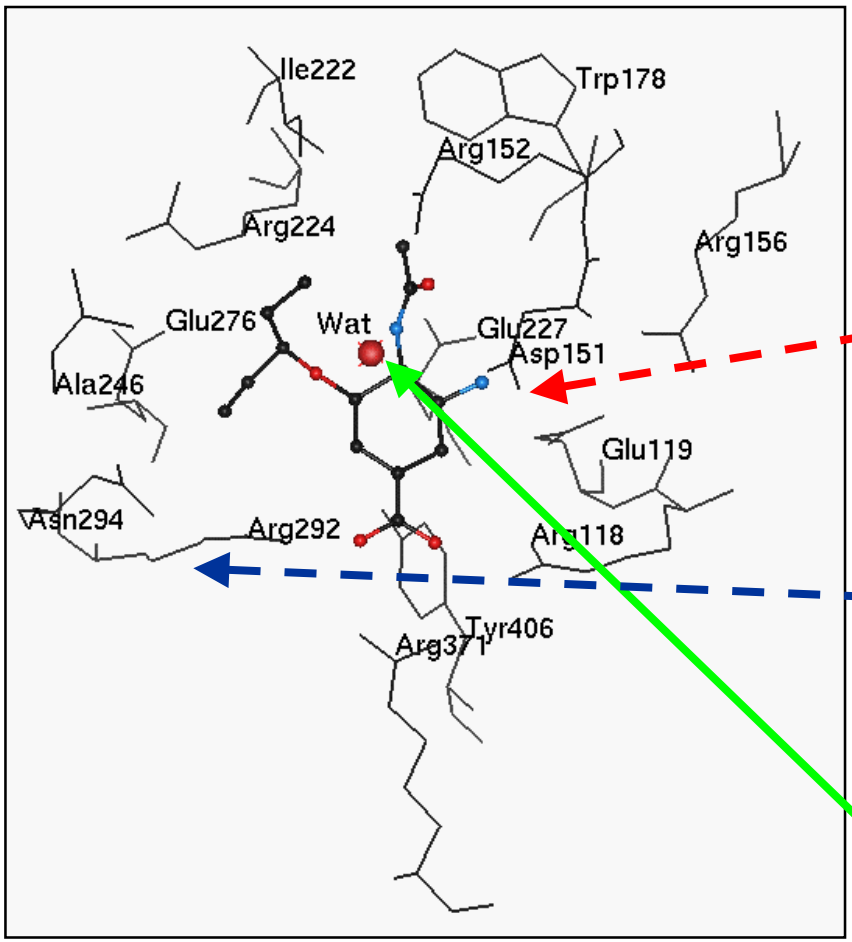


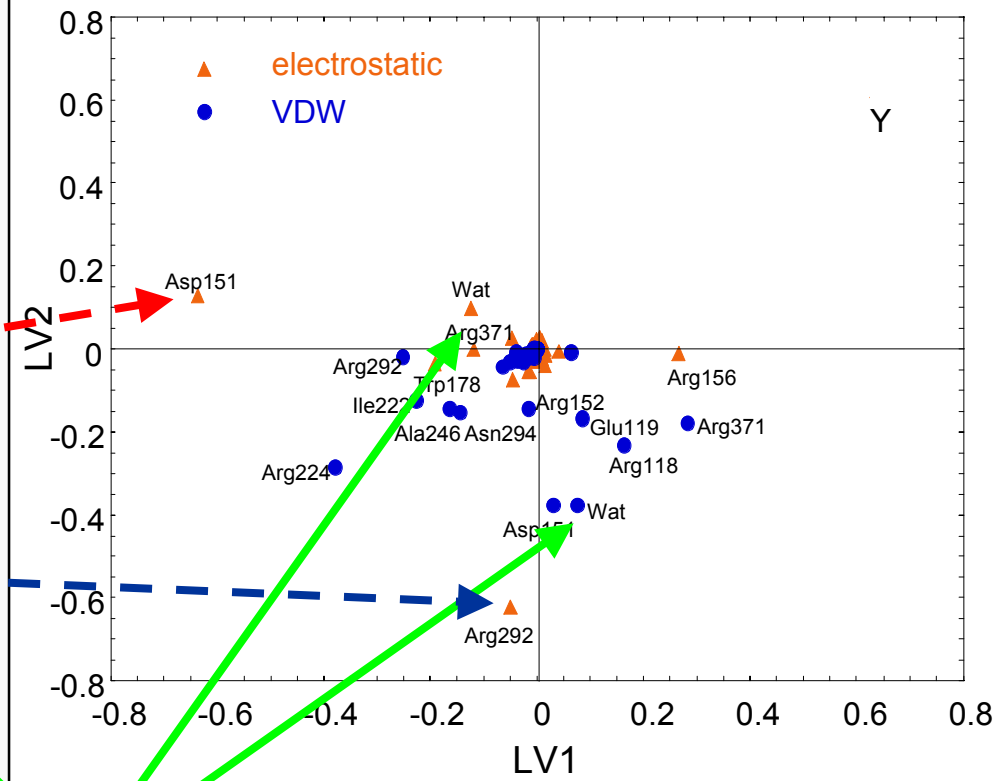




No obvious distinction between N2 and N9; N9 only model has better predictive statistics (Q^2=0.84. SDEP=0.75)

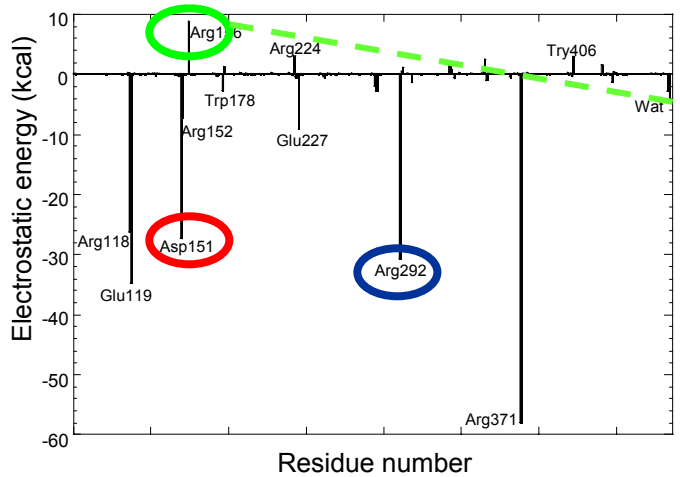
Important contributions to neuraminidase inhibition

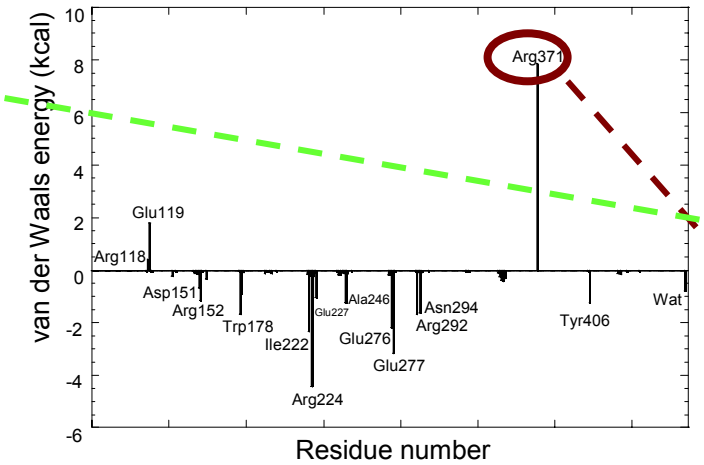


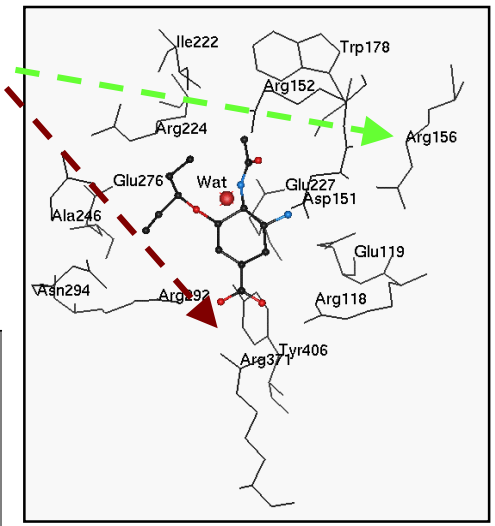


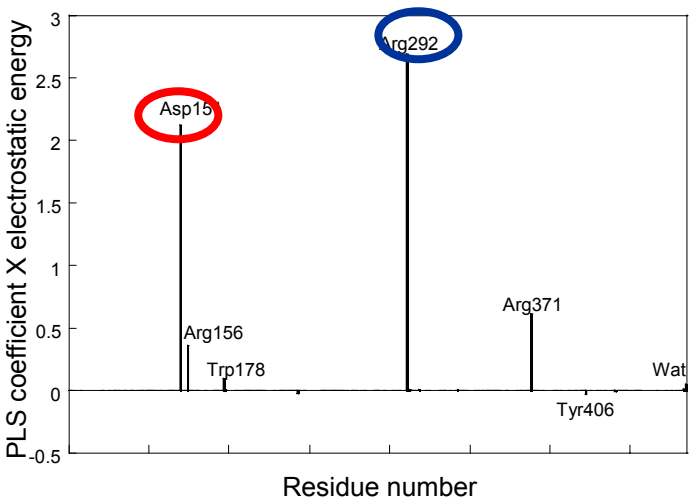
1 bridging water molecule

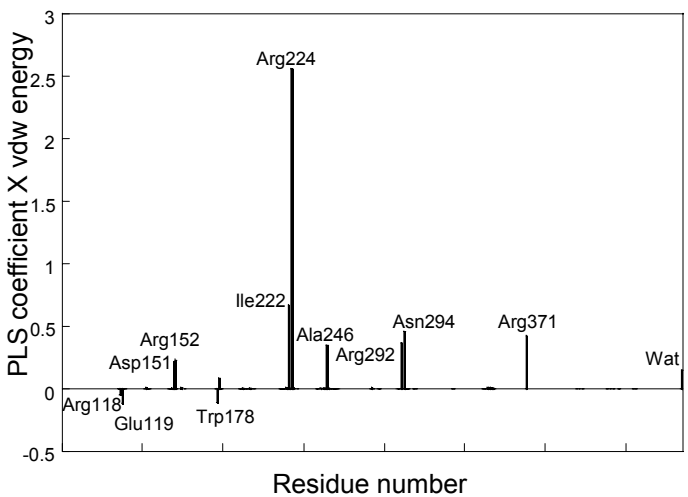
Influenza neuraminidase inhibitors COMBINE model









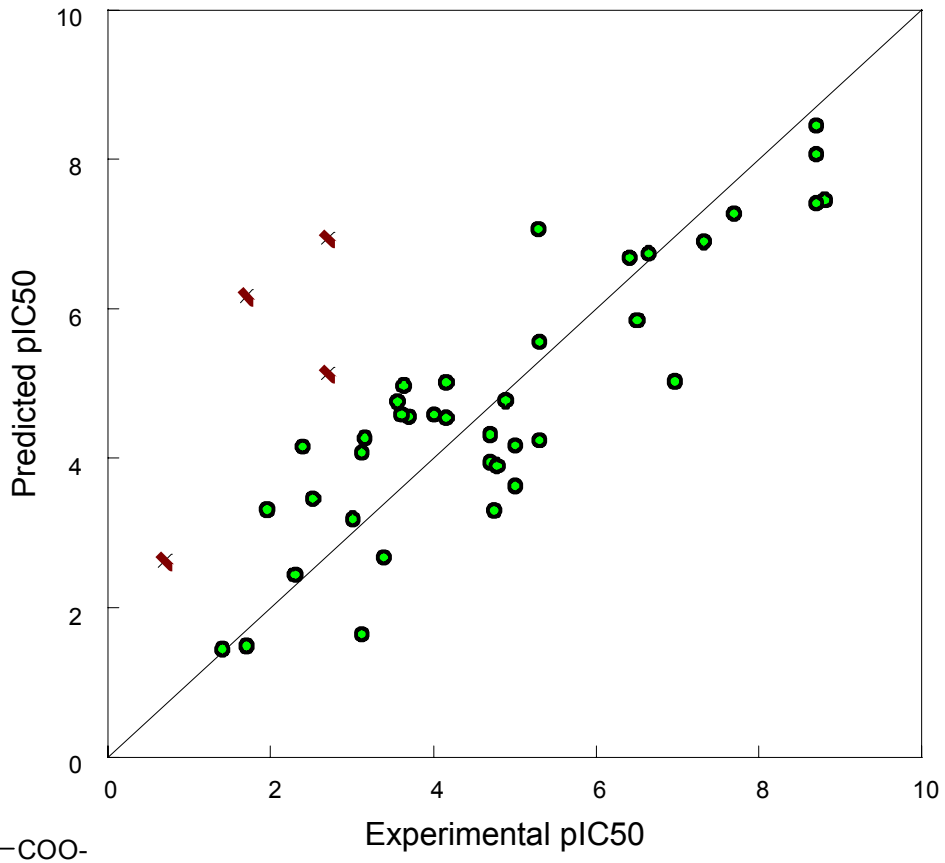


Influenza neuraminidase inhibitors COMBINE model outliers

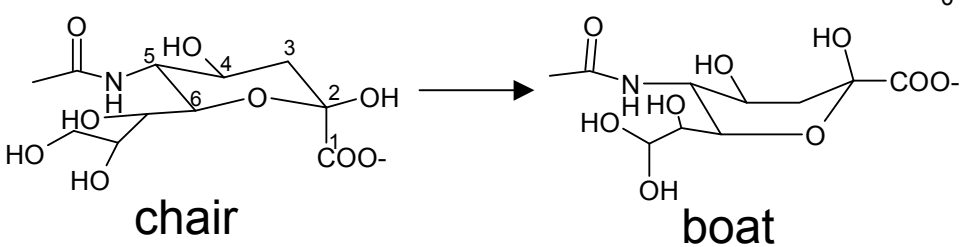
4 outliers overestimated binding: 3 sialic acid + axial-PANA complexes

Sialic acid:

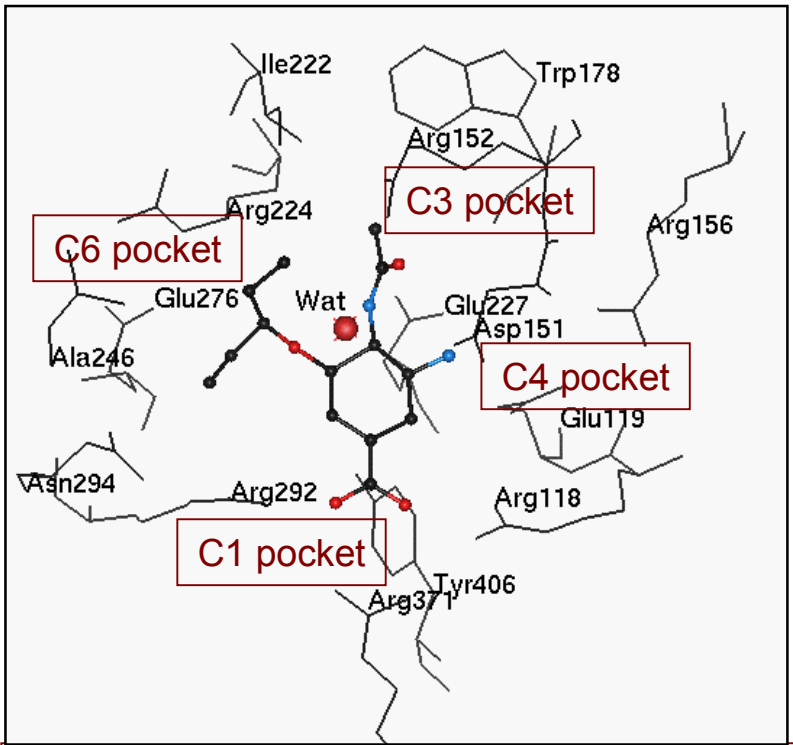
1. α: β anomers 1:10; α binds
2. Conformational change to boat form

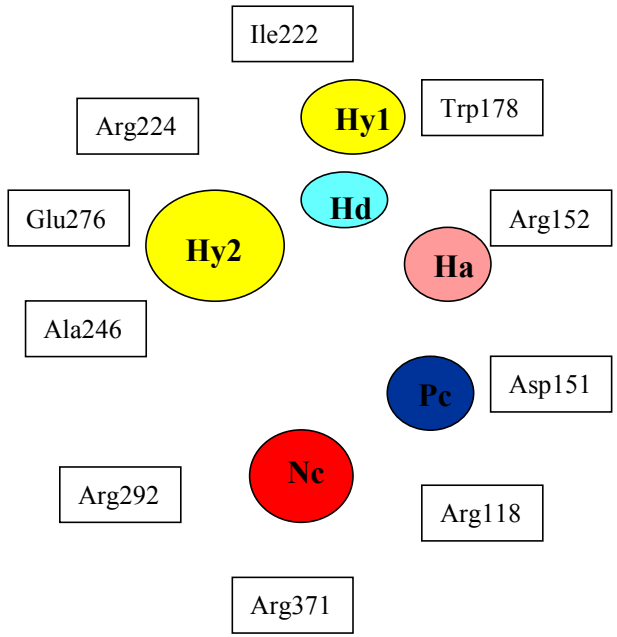


α-sialic acid



Design of new Influenza neuraminidase inhibitors





Characteristics of high affinity ligand

1. Novel frameworks possible
   E.g. bcx-1812- new inhibitor with 5-membered ring framework
   Experimental pIC_{50} ~ 8.85-10; predicted pIC_{50} ~ 8.4 (N9), 7.1 (N2)
2. Ligand binding affinity optimization to maximum number of subtypes with minimum resistance due to mutation. Avoid R292; target residues whose mutation is highly deleterious, e.g.151, 276, 406

Table 1. Applications of COMBINE analysis

Class	Protein	Ligand	No. of complexes	Energy terms						No. of LVs	Refs.
				vdw	Coulomb	PB	Intra	Solvation	SASA		
Enzyme-inhibitor	Human synovial phospholipase A_2	Derivatives of transition state analogues	26	x	x		x			2	1,6
	HIV protease	Merck set	33	x		x				2	7,8
	Glycogen phosphorylase	Glucose analogues	10	x						1	9
	Human neutrophile elastase	Pyridone-containing trifluoro-methyl ketones	40	x		x				3	Cuevas, Pastor, Perez, Gago, submitted
	Influenza neuraminidase (2 subtypes + 1 mutant)	Derivatives of sialic acid and benzoic acid	43	x	x					4	10
Enzyme-substrate	Cytochrome P450 2A1	Hetero-cyclic amines	12	x	x					2	11
	Haloalkane dehalogenase	Haloalkanes +other halogenated compounds	18	x	x					4	Kmunicek, Luengo, Gago, Ortiz, Wade, Damborsky, unpublished
Protein receptor-DNA	Glucocorticoid receptor DNA binding domain mutants	6-base-pair DNA mutants of glucocorticoid receptor response element	320	x	x			x	x	5	5
			32	x	x					4	Tomic, Wade, unpublished

LV: latent variable.; vdw: van der Waals; PB: Poisson-Boltzmann; SASA: Solvent accessible surface area

COMBINE Analysis

- Wide range of applications
- Predictive and robust models
- Mechanistic insights

- ***Solvation:***
  - Continuum Electrostatics
  - Explicit water molecules
- ***Conformations:***
  - Comparative modelling
  - Docking-*de novo*/ comparative
  - Conformational changes on binding
  - Entropy
- ***Mutations:***
  - selectivity