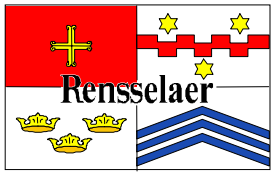
Quantitative Structure Property Relationship Models for Evaluating Mixed Mode Chromatographic Systems

Steven M. Cramer
Isermann Department of Chemical and Biological Engineering
Rensselaer Polytechnic Institute, Troy, NY

Presented at
MIXED MODE CHROMATOGRAPHY
SCI International Headquarters
London, UK

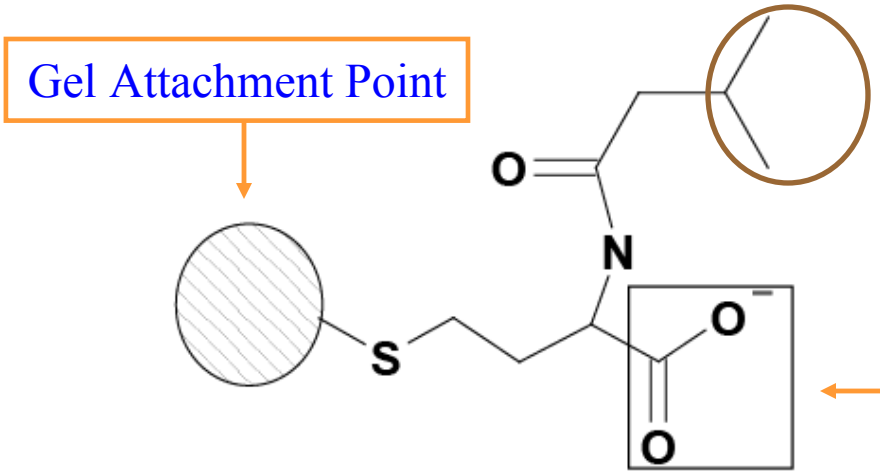


Outline

- Elution and SPR studies with mixed mode resins.
- QSPR investigation of mixed mode chromatographic systems.
- Other experimental and theoretical tools for studying these systems.

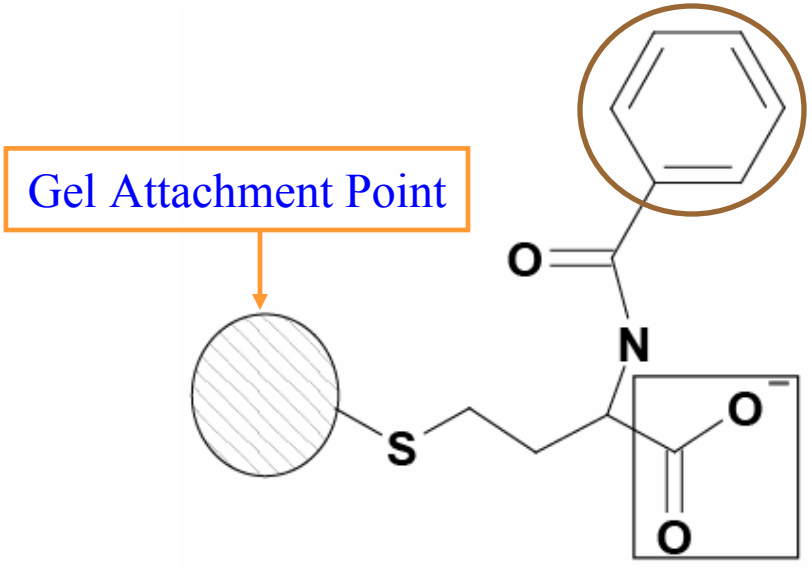
Rensselaer

Structures of Anionic High Salt Binding Ligands (obtained from GEHealthcare)



Prototype I: 0.085 mmol/ml
Prototype II: 0.16 mmol/ml

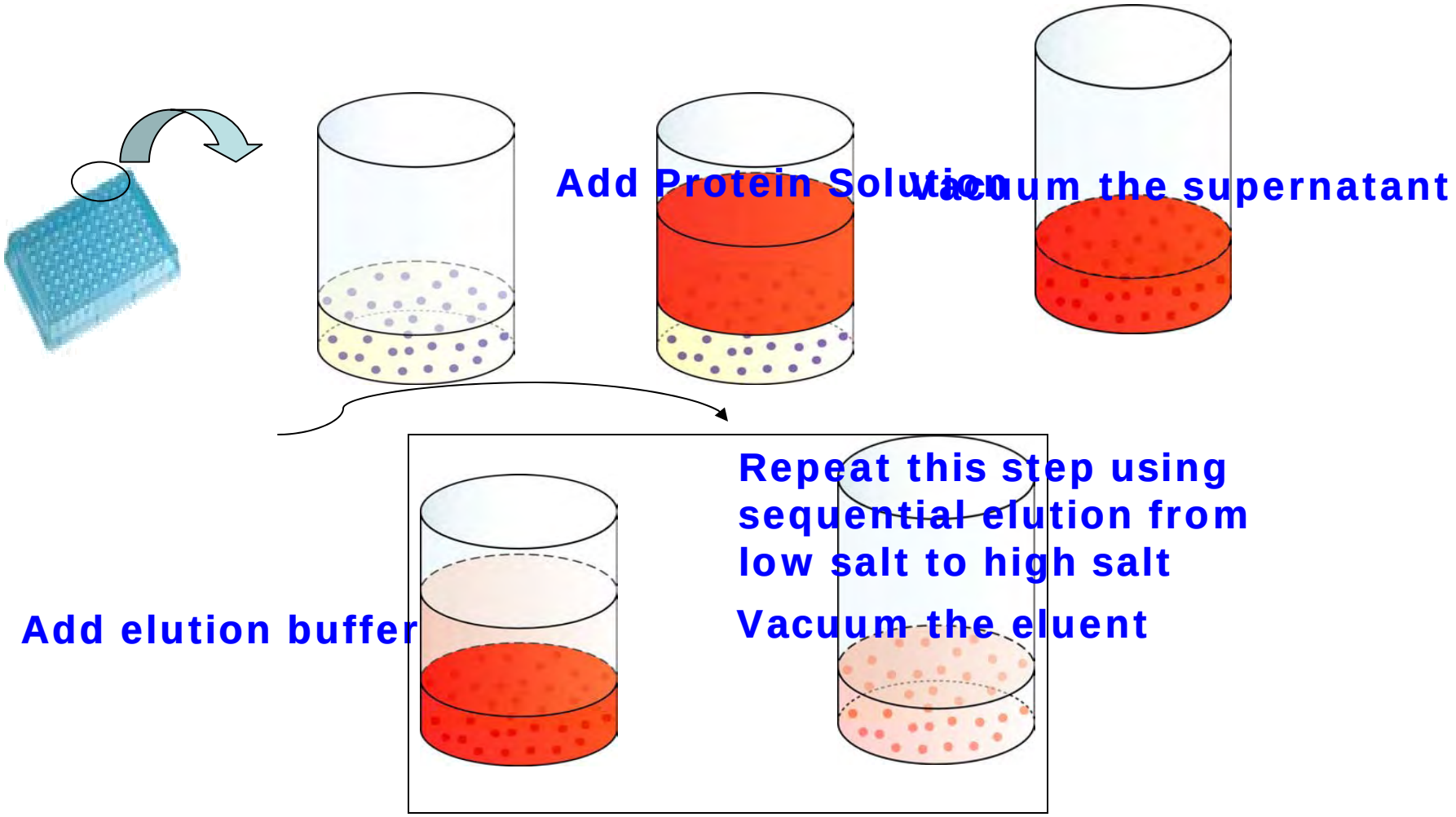
Charge Functional Group



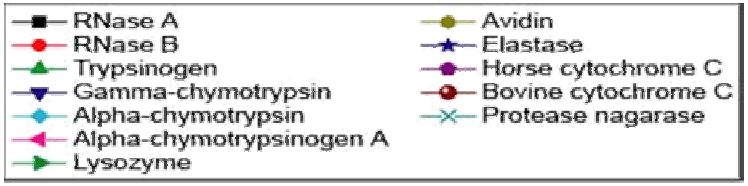
Prototype III: 0.086 mmol/ml
Prototype IV: 0.154 mmol/ml

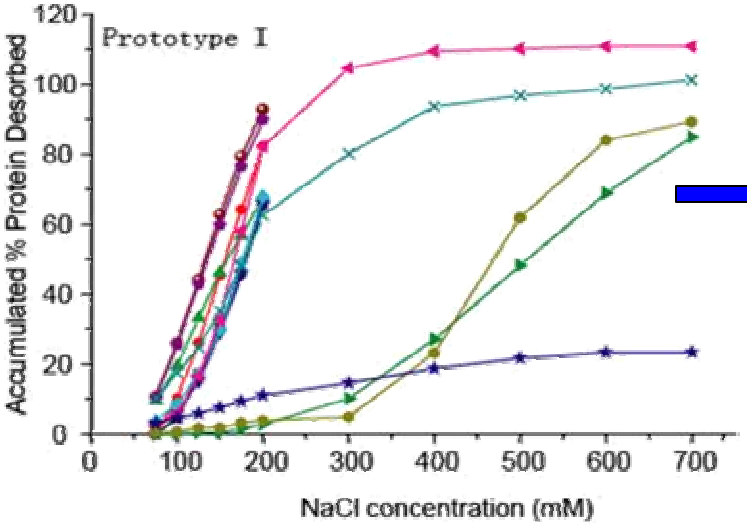
Charge Functional Group

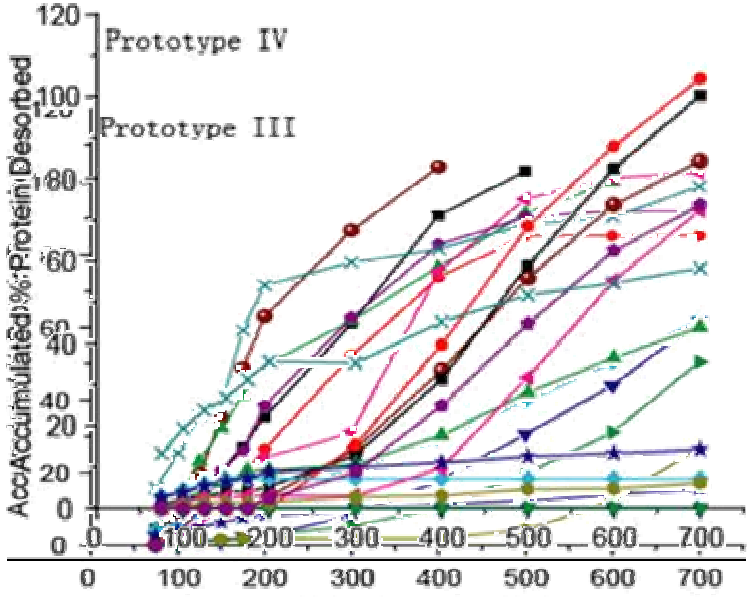
Batch Adsorption-Desorption Profiles

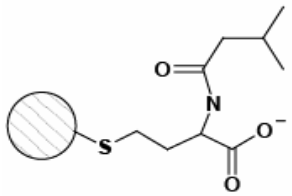


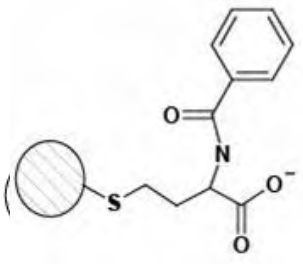
Desorption Profiles of Proteins







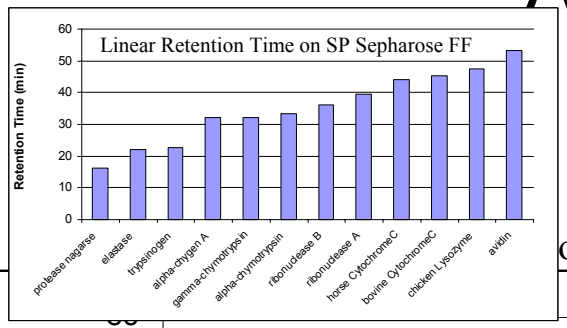


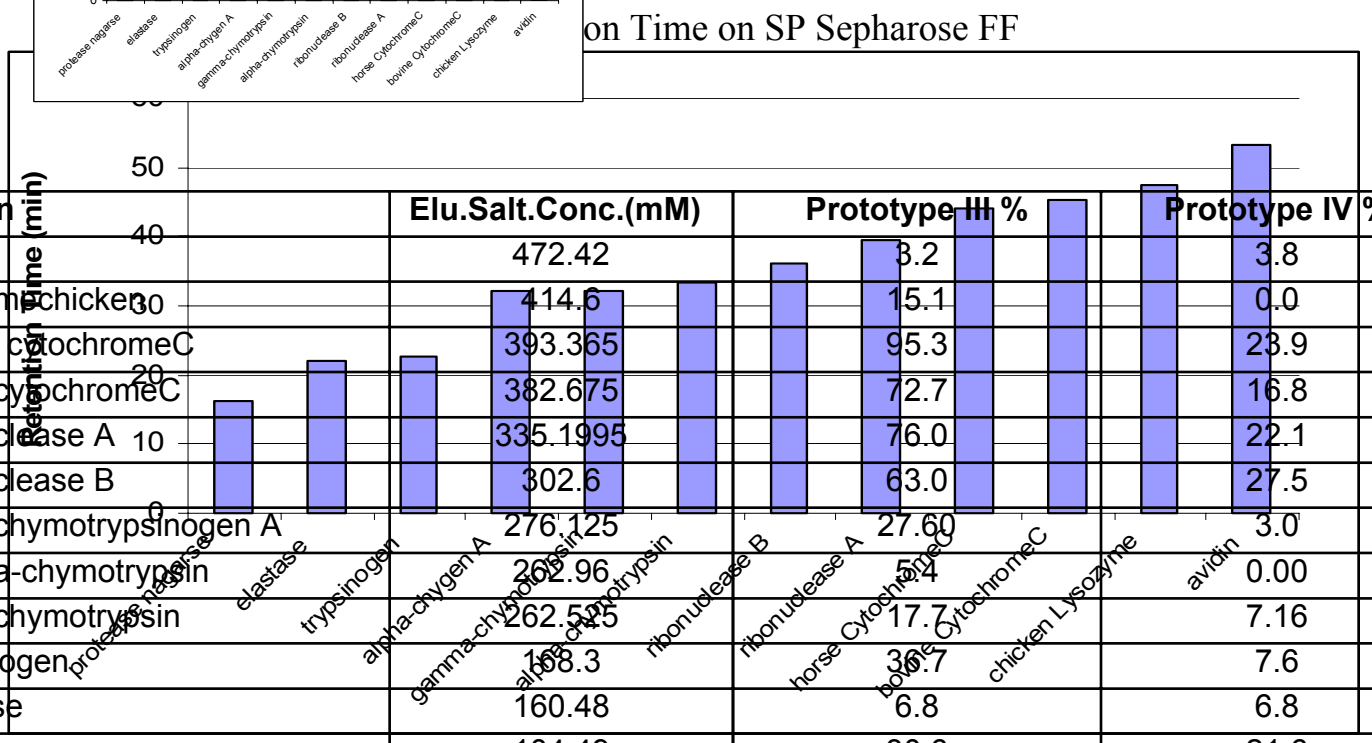


I.C. 0.085 mmol/ml

I.C. 0.854 mmol/ml

Effect of the Aromatic Ring on Protein Binding Affinity

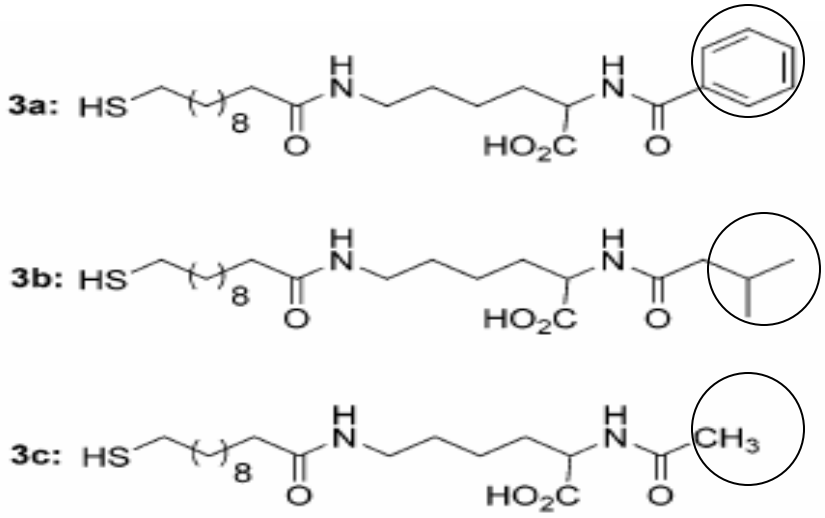




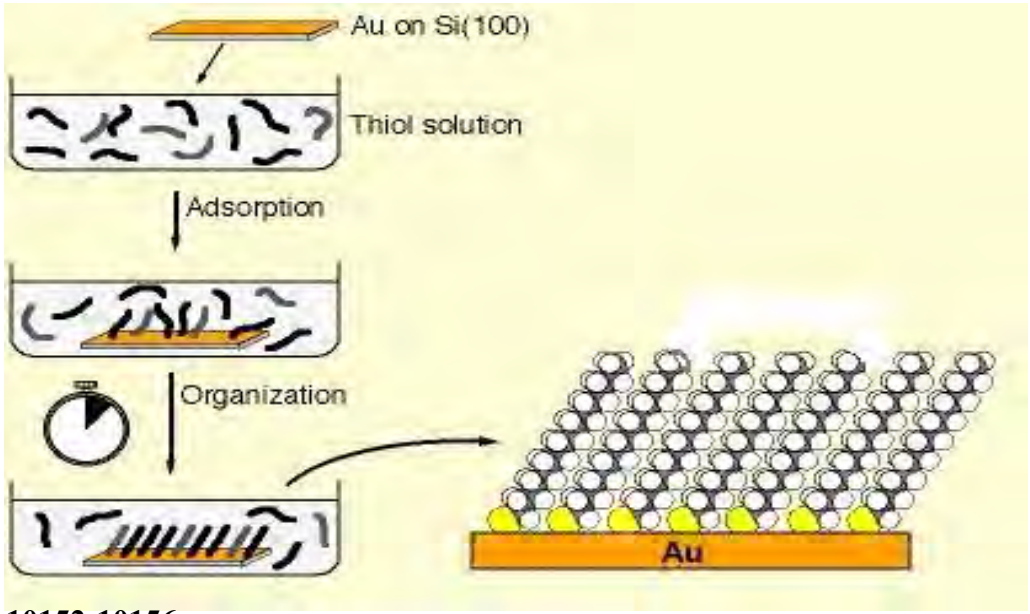
Protein	Elu.Salt.Conc.(mM)	Prototype III %	Prototype IV %
avidin	472.42	3.2	3.8
lysozyme chicken	414.6	15.1	0.0
bovine cytochromeC	393.365	95.3	23.9
horse cytochromeC	382.675	72.7	16.8
ribonuclease A	335.1995	76.0	22.1
ribonuclease B	302.6	63.0	27.5
alpha-chymotrypsinogen A	276.125	27.60	3.0
gamma-chymotrypsin	262.96	5.4	0.00
alpha-chymotrypsin	262.525	17.7	7.16
trypsinogen	168.3	30.7	7.6
elastase	160.48	6.8	6.8
protease nagarse	104.49	30.6	21.6

Binding Mechanism Investigation Using SPR

Hypothesis: Hydrophobicity plays important role in protein high salt binding



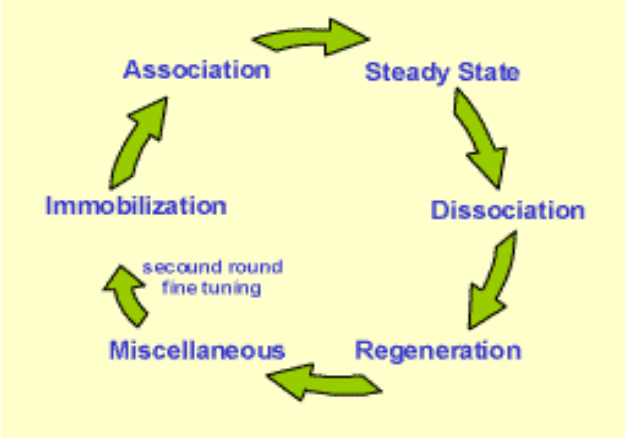
Hydrophobicity

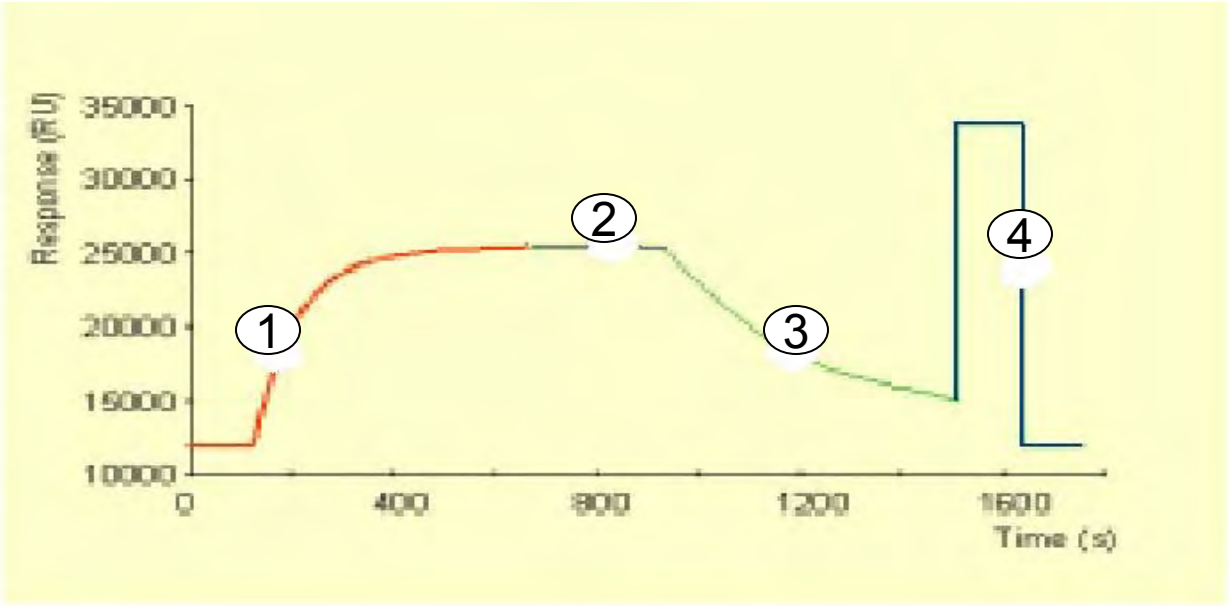


Kane et al. Langmuir 2006 *22*, 10152-10156

Biacore 3000 Spectrometer: Assay and Analysis

- SPR: Measures biomolecular interactions in a real time label free environment
- Biacore assay: Immobilization, Interaction analysis and Regeneration
- Sensogram: Plot of response against time





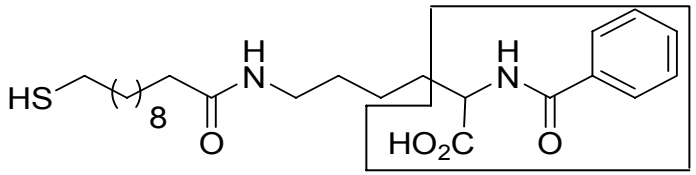
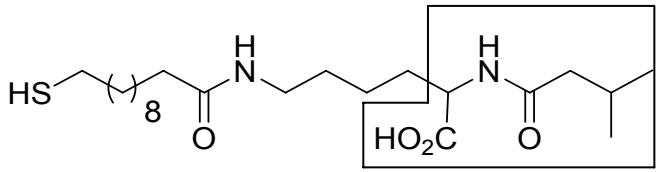
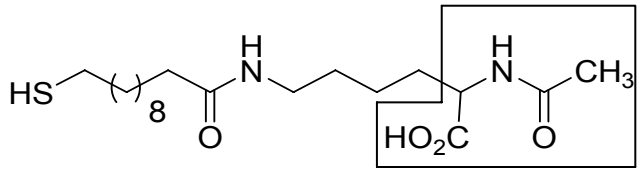
1: Association

2: Steady state

3: Dissociation

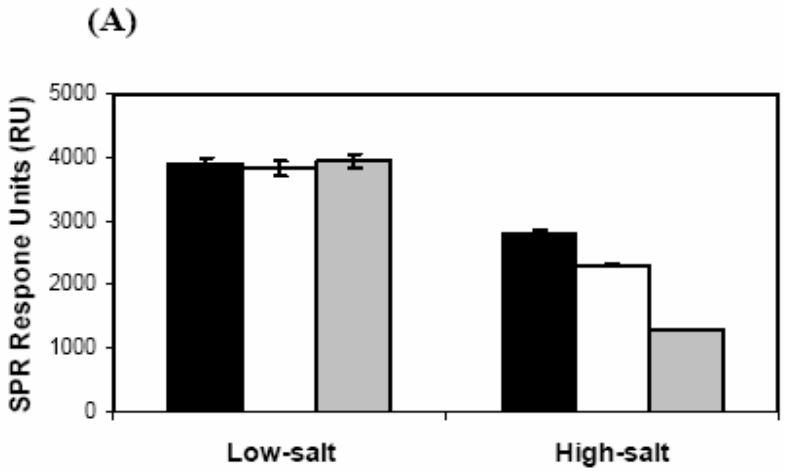
4: Regeneration

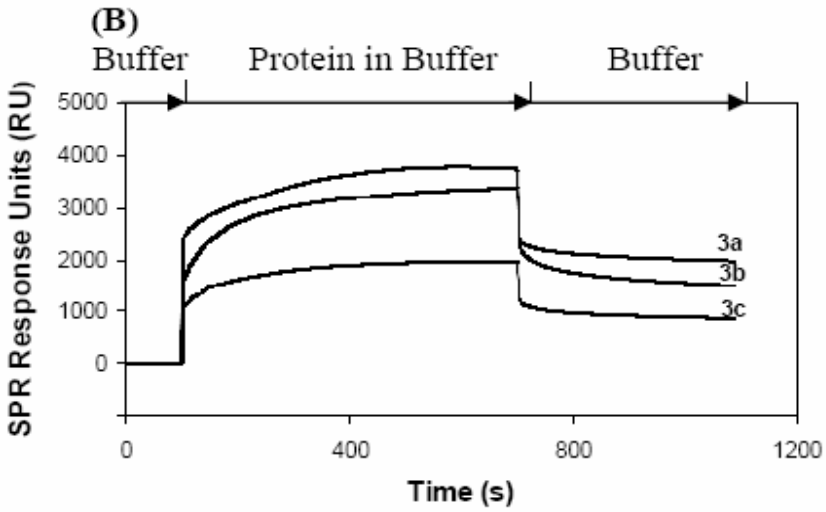
Multimodal surfaces: Structure and Characterization

Ligand	Log P	Contact Angle	Ellipsometric thickness (Å)	
			Estimated	Measured
	0.44	75.8 ± 2.7	25.1	24.5 ± 1.0
	-0.05	64.2 ± 1.4	23.1	22.5 ± 1.2
	-1.46	52.5 ± 1.5	23.0	21.9 ± 1.6

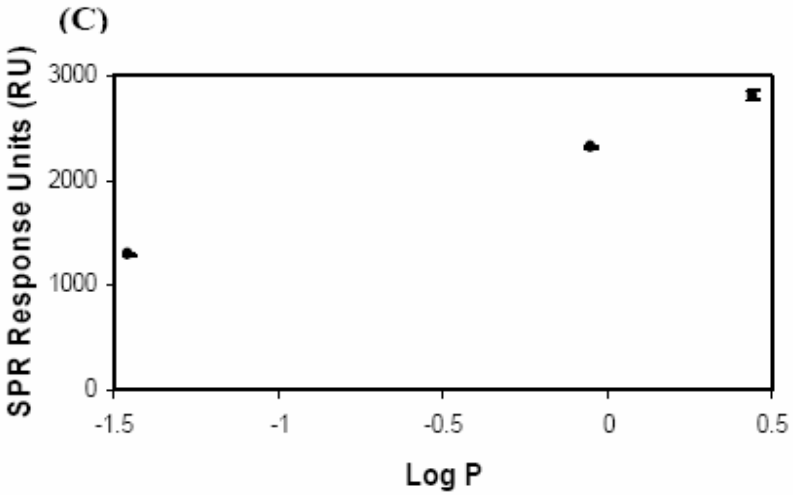
Interactions of Lysozyme with SAMs Presenting Multi-modal Ligands: Effect of Ligand Chemistry

3a: Benzoyl; **3b:** Isovaleryl ; **3c:** Acetyl



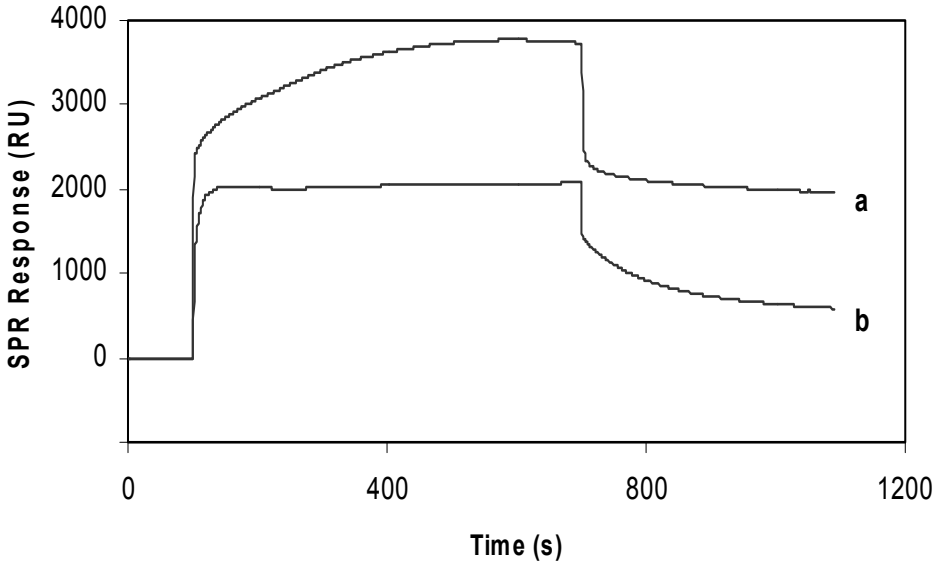


Adsorption of lysozyme at high salt conditions increases as head group hydrophobicity increases.

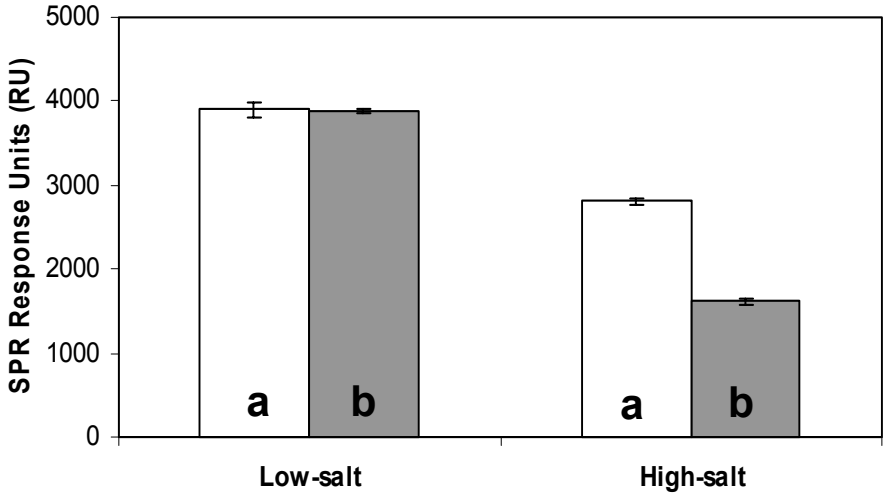


SPR experiments: Protein hydrophobicity

(A) SPR Sensograms



(B) Amount of protein adsorbed

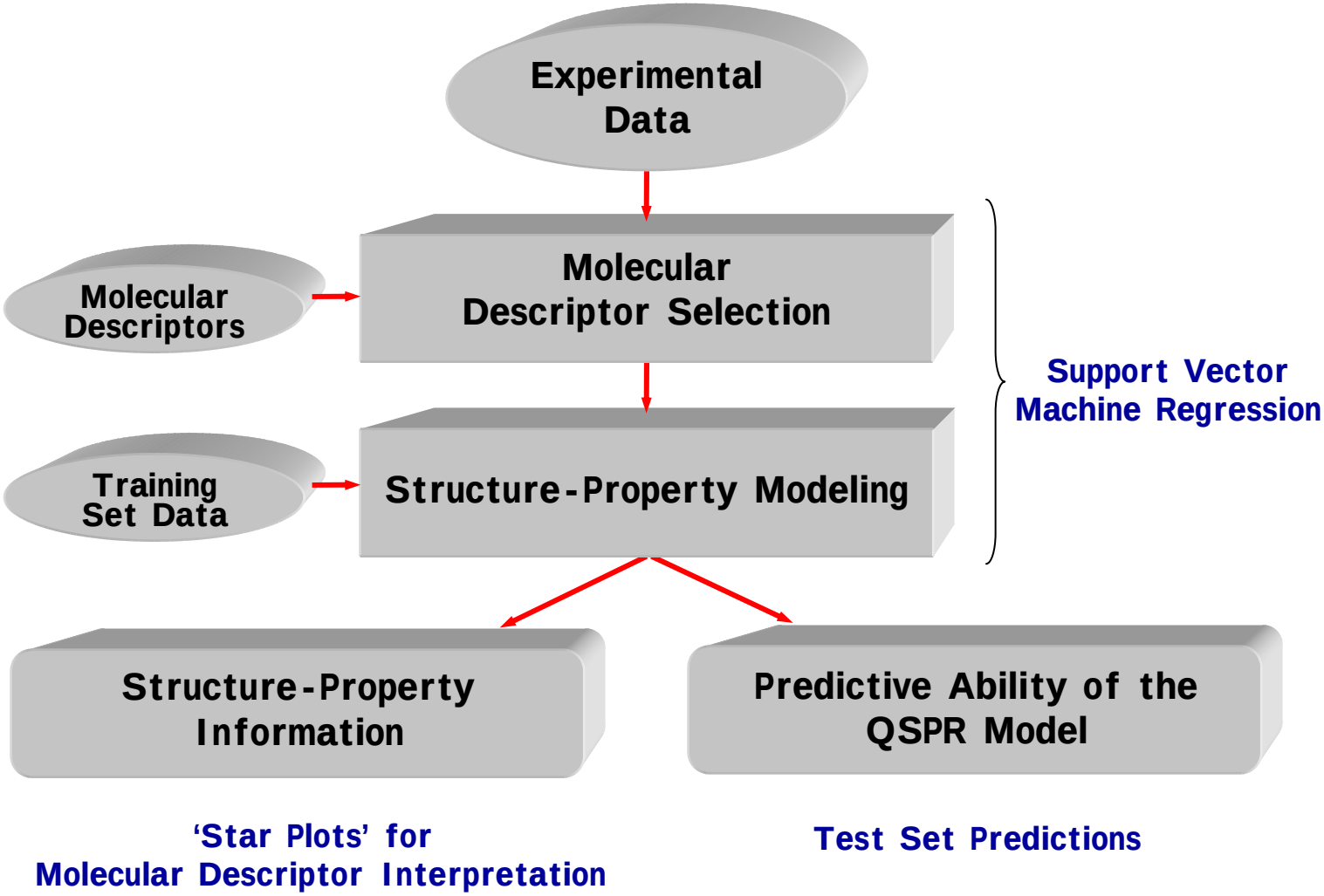


a: lysozyme

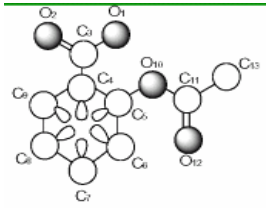
b: horse cytochrome *c*

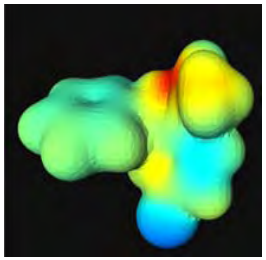
Greater hydrophobicity ⟶ Higher binding

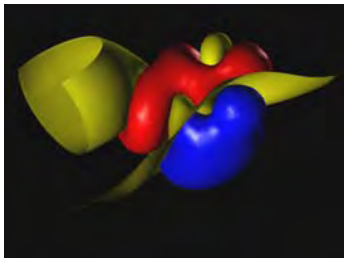
QSPR Modeling Flowchart

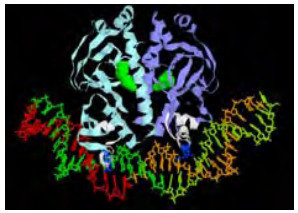


Encoding Structure : Descriptors









AAACCTCATAGGAAGCATACCA
GGAATTACATCA…

Structural Descriptors

Physiochemical Descriptors

Topological Descriptors

Geometrical Descriptors

Constitutional Descriptors

Electrostatic Descriptors

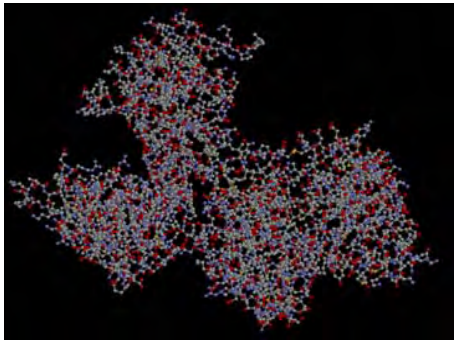
Quantum-chemical Descriptors

Thermodynamic Descriptors

Molecular Structures → Descriptors ↔ Model ↔ Affinity

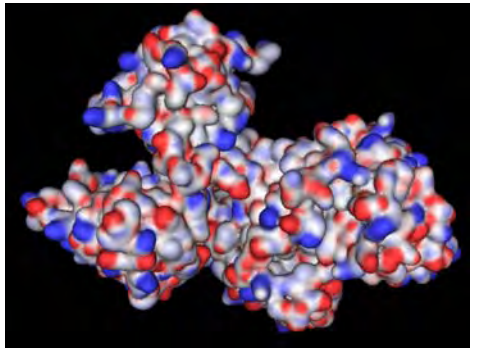
MOE Descriptors

- Classical physicochemical properties:
  - logP, molecular refractivity
- Pharmacophore features:
  - the number of H-bond donor/acceptor atom
  - polar or hydrophobic surface area
- Property-mapped subdivided surface area:



3D protein crystal geometry

map partial charge on molecular surface



blue: positive; red: negative

www.chemcomp.com

TAE/RECON Descriptors

EP	Electrostatic Potential	$EP(r) = \sum_\alpha Z_\alpha / \lvert r - R_\alpha \rvert - \int \rho(r') d(r') / \lvert r - r' \rvert$
Del(Rho)•N	Electron Density Gradient normal to electron density iso-surface	
G	Electronic Kinetic Energy	$G = -(\hbar/4m) \int \{\nabla\psi^* \cdot \nabla\psi\} d\tau$
K	Electronic Kinetic Energy	$K = -(\hbar/4m) \int \{\psi^* \nabla^2 \psi + \psi \nabla^2 \psi^*\} d\tau$
Del(K)•N	Gradient of K Electronic Kinetic Energy normal to surface	
Del(G)•N	Gradient of G Electronic Kinetic Energy normal to surface	
Fuk	Fukui F^+ function scalar value	$F^+(r) = \rho_{HOMO}(r)$
Lapl	Laplacian of the electron density	$\nabla^2 \rho(r) = G(r) - K(r)$
BNP	Bare Nuclear Potential	$BNP_j = \sum_{i=1}^{n} Z_i / r_{ij}$
PIP	Local Average Ionization Potential	$PIP(r) = \sum_i \rho_i(r) \cdot \lvert \varepsilon_i \rvert / \rho(r)$

1. Bader, R.F.W. *Atoms in Molecules: A Quantum Theory*; Oxford Univ. Press, 1994.

2. Breneman, C.M.; Rhem, M. *J. Comp. Chem.* 18, 182-197, 1997.

pKa Estimation Methods for Acidic and Basic Amino Acids

- **Null pKa (no microenvironment effects)**
- **Estimated pKa from various software.**

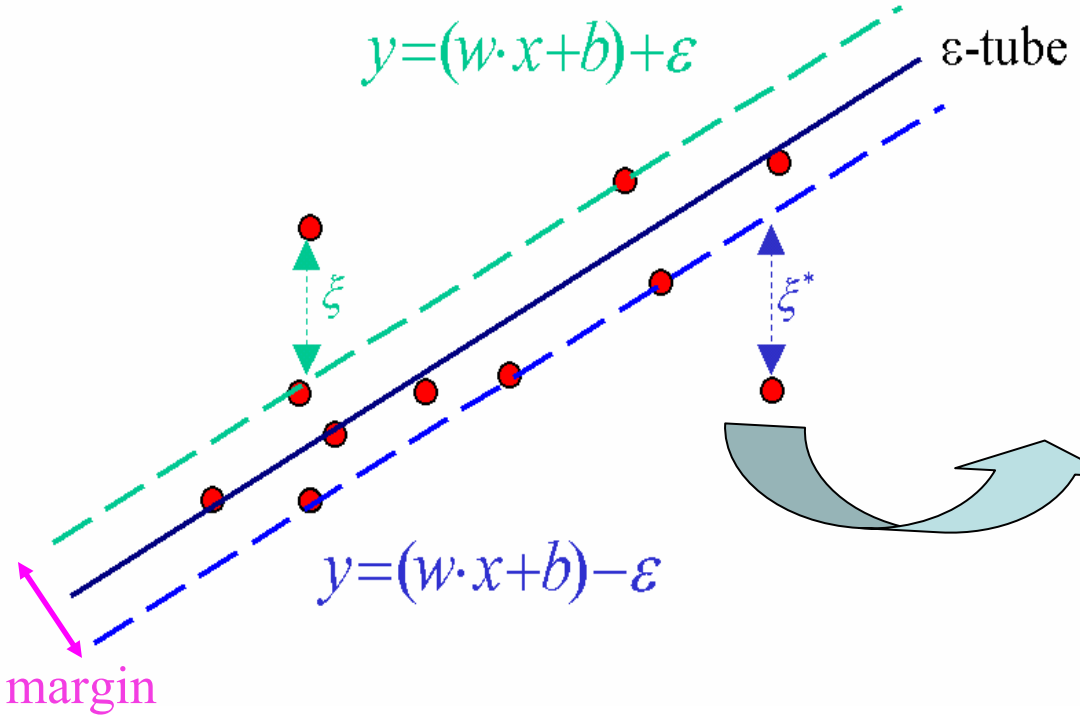
  (e.g. What If, PropKa, MM_SCP)

- **Once pKa is estimated, Henderson-Hasselbach equation is employed:**

  pKa=pH+log(protonated/deprotonated)

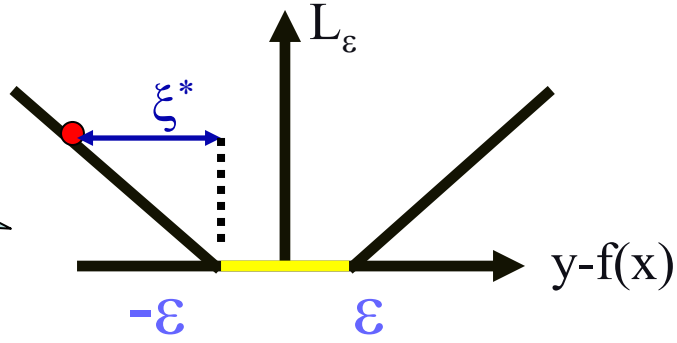
Support Vector Regression (SVR)

- Minimize the regularized empirical error:
  - training error + model complexity $\min_{w,b,\xi_i,\xi_i^*} C\sum_{i=1}^{l}(\xi_i + \xi_i^*) + \frac{1}{2}\|w\|^2$



ε-insensitive loss function:

$L_\varepsilon(y - f(x)) := \max(0, |y - f(x)| - \varepsilon)$



- Avoid overfitting by controlling the model complexity

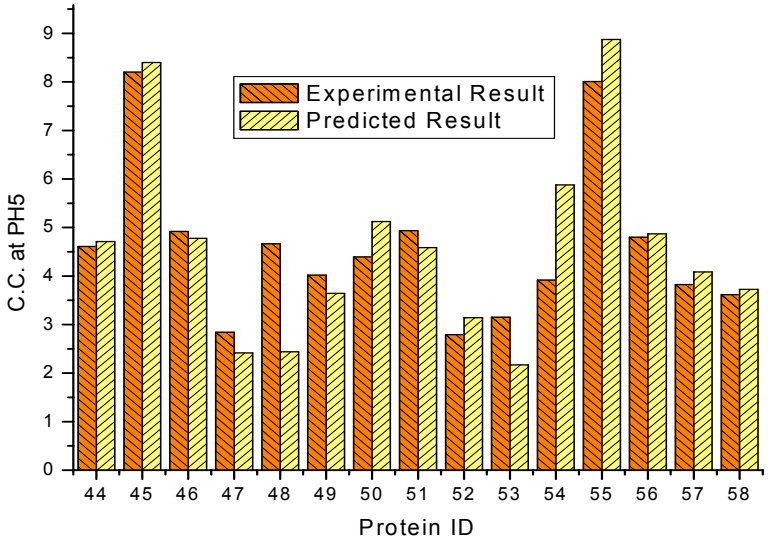
QSPR Prediction Map

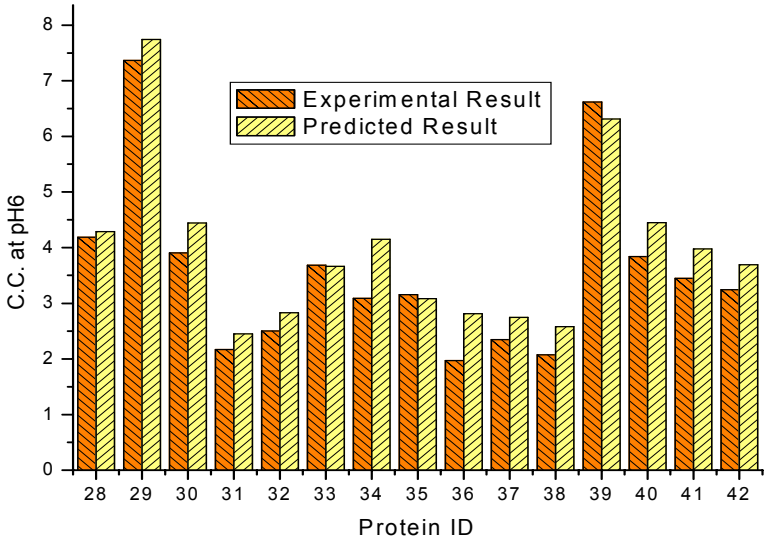
Experimental responses:

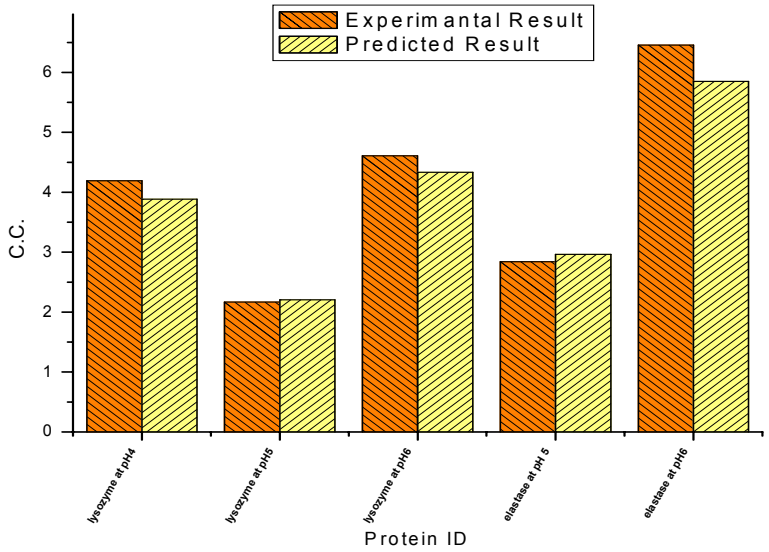
1. **Retention volume**
2. **SMA parameters: characteristic charge and Ksma**

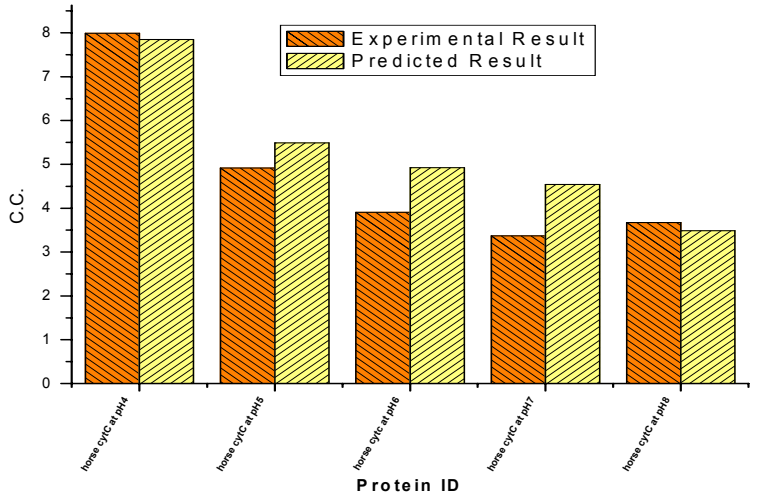
pH4	pH5	pH6	pH7	pH8
Mol.1	Mol.1	Mol.1	Mol.1	Mol.1
Mol.2	Mol.2	Mol.2	Mol.2	Mol.2
Mol.3	Mol.3	Mol.3	Mol.3	Mol.3
Mol.4	Mol.4	Mol.4	Mol.4	Mol.4
Mol.5	Mol.5	Mol.5	Mol.5	Mol.5
Mol.6	Mol.6	Mol.6	Mol.6	Mol.6
Mol.7	Mol.7	Mol.7	Mol.7	Mol.7
Mol.8	Mol.8	Mol.8	Mol.8	Mol.8
Mol.9	Mol.9	Mol.9	Mol.9	Mol.9
Mol.10	Mol.10	Mol.10	Mol.10	Mol.10
Mol.11	Mol.11	Mol.11	Mol.11	Mol.11
Mol.12	Mol.12	Mol.12	Mol.12	Mol.12
Mol.13	Mol.13	Mol.13	Mol.13	Mol.13
	Mol.14	Mol.14	Mol.14	
	Mol.15	Mol.15		
	Mol.16	Mol.16		

QSPR Prediction of C.C.





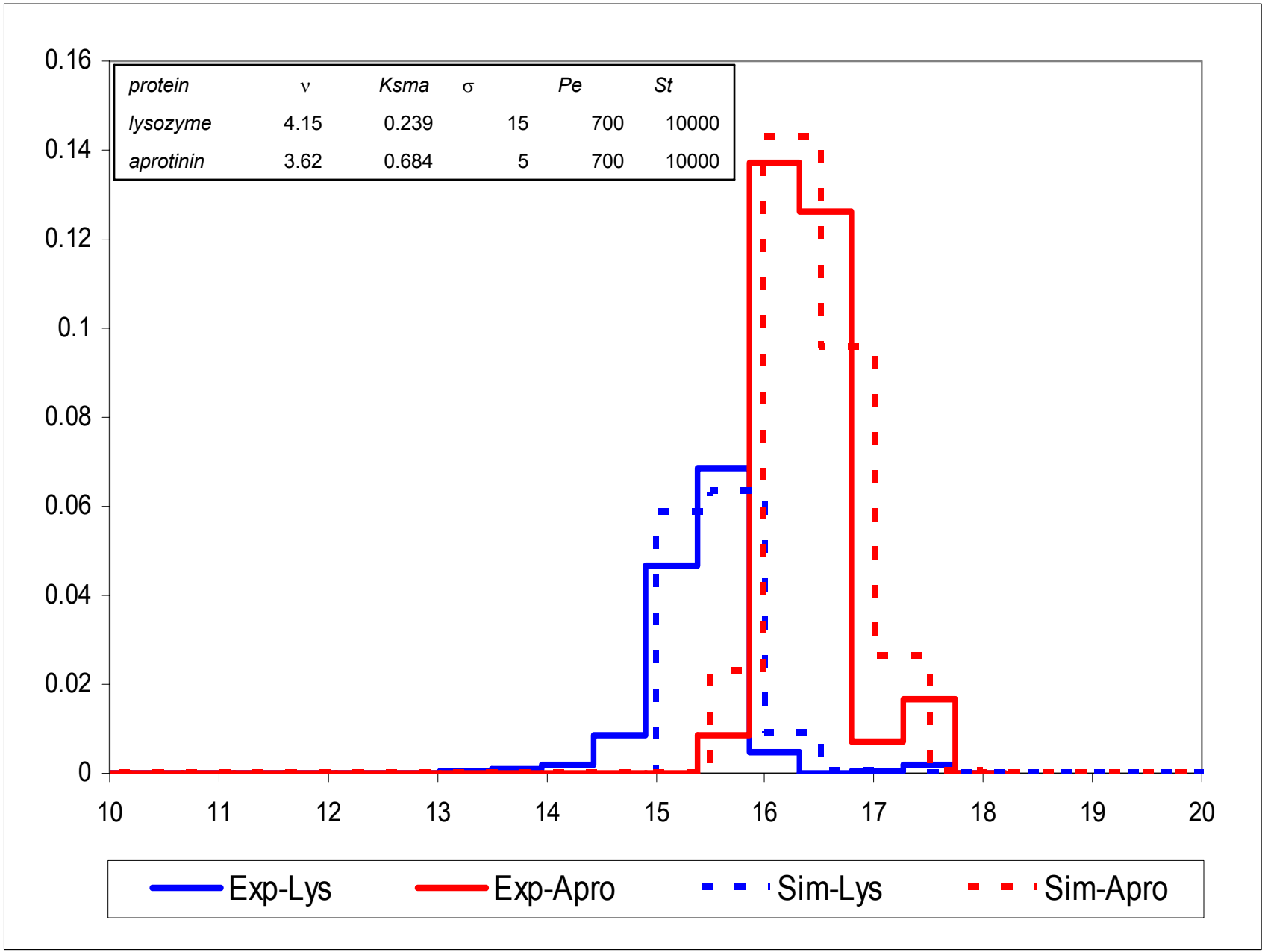




QSPR Prediction of SMA Parameters at Any Given pH

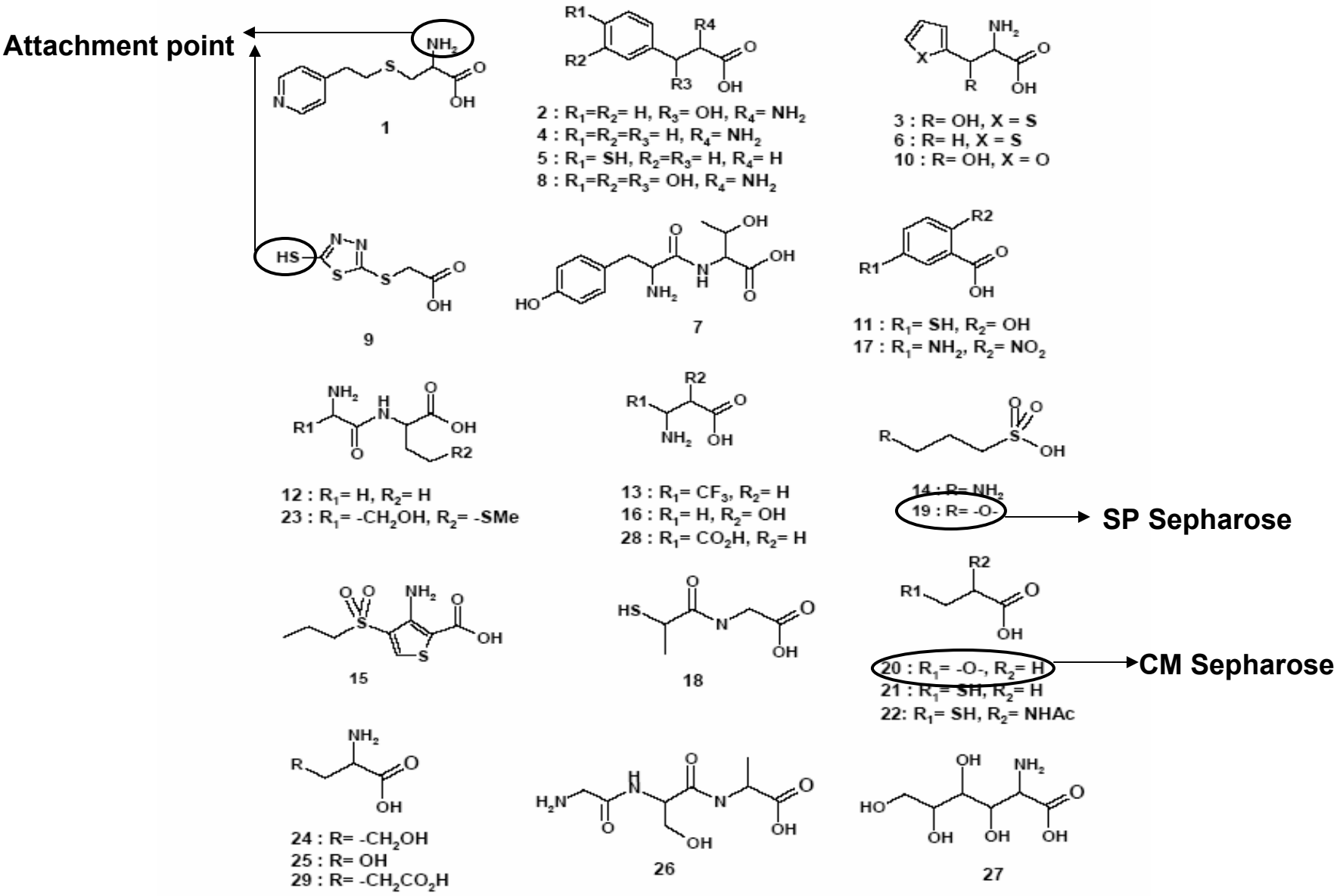
pH4	pH5	pH5	pH6	pH6.5	pH7	pH8
Mol.1	Mol.1	Mol.1	Mol.1	Mol.1	Mol.1	Mol.1
Mol.2	Mol.2	Mol.2	Mol.2	Mol.2 ?	Mol.2	Mol.2
Mol.3	Mol.3	Mol.3	Mol.3	Mol.3 ?	Mol.3	Mol.3
Mol.4	Mol.4	Mol.4	Mol.4	Mol.4	Mol.4	Mol.4
Mol.5	Mol.5	Mol.5	Mol.5	Mol.5	Mol.5	Mol.5
Mol.6	Mol.6	Mol.6	Mol.6	Mol.6	Mol.6	Mol.6
Mol.7	Mol.7	Mol.7	Mol.7	Mol.7	Mol.7	Mol.7
Mol.8	Mol.8	Mol.8	Mol.8	Mol.8	Mol.8	Mol.8
Mol.9	Mol.9	Mol.9	Mol.9	Mol.9	Mol.9	Mol.9
Mol.10	Mol.10	Mol.10	Mol.10	Mol.10	Mol.10	Mol.10
Mol.11	Mol.11	Mol.11	Mol.11	Mol.11	Mol.11	Mol.11
Mol.12	Mol.12	Mol.12	Mol.12	Mol.12	Mol.12	Mol.12
Mol.13	Mol.13	Mol.13	Mol.13	Mol.13	Mol.13	Mol.13
	Mol.14	Mol.14	Mol.14	Mol.14	Mol.14	Mol.14
	Mol.15	Mol.15	Mol.15	Mol.15	Mol.15	Mol.15
	Mol.16	Mol.16	Mol.16	Mol.16	Mol.16	Mol.16

Prediction of Protein Chromatographic Behavior at Any Given Condition



Mixed Mode Systems

Candidates of High Salt Binding Ligands



Johansson et al. Journal of Chromatography A 1016 (2003) 21 & 35

Functionality Screening-Linear Retention Time

Buffer A: 20 mM phosphate buffer, pH 6.8
Buffer B: 20 mM phosphate buffer+2M NaCl, pH 6.8

ILC	Elution conductivity (mS/cm)		
	Rib **pI=9.4**	Cyt *c* **pI=10.2**	Lys **pI=11**
1	29	35	143
2	42	43	89
3	46	48	98
4	47	53	120
5	66	78	150
6	40	45	97
7	34	42	71
8	37	37	87
9	29	36	109
10	32	39	62
11	32	38	65
12	26	35	53
13	24	33	47
14	20	30	39
15	26	34	56
16	23	29	37
17	26	33	51
18	23	31	39
19	19	32	32 → **SP Sepharose**
20	17	26	30 → **CM Sepharose**
21	25	33	44
22	19	29	34
23	21	29	36
24	20	28	38
25	20	29	33
26	20	31	31
27	20	27	33
28	28	33	33
29	24	32	32

Observations:

1. **Lysozyme generally has higher retention than RNAseA and Cyt C**
2. **RNAse and Cyt C which have no retention in HIC system shows higher binding affinity on aromatic ligands**

Johansson et al. Journal of Chromatography A 1016 (2003) 21 & 35

RNaseA_Elution Conductivity		
Ligand ID.	E.C.	Ring
20	17	N
19	19	N
22	19	N
14	20	N
24	20	N
25	20	N
26	20	N
27	20	N
23	21	N
16	23	N
18	23	N
13	24	N
29	24	N
21	25	N
12	26	N
15	26	Y
17	26	Y
28	28	N
1	29	Y
9	29	Y
10	32	Y
11	32	Y
7	34	Y
8	37	Y
6	40	Y
2	42	Y
3	46	Y
4	47	Y
5	66	Y

Cyt C_Elution Conductivity		
Ligand ID	E.C.	Ring
20	26	N
27	27	N
24	28	N
16	29	N
22	29	N
23	29	N
25	29	N
14	30	N
18	31	N
26	31	N
19	32	N
29	32	N
13	33	N
17	33	Y
21	33	N
28	33	N
15	34	Y
12	35	N
1	35	Y
9	36	Y
8	37	Y
11	38	Y
10	39	Y
7	42	Y
2	43	Y
6	45	Y
3	48	Y
4	53	Y
5	78	Y

Lysozyme_Elution Conductivity		
ligand ID	E.C.	Ring
20	30	N
26	31	N
19	32	N
29	32	N
25	33	N
27	33	N
28	33	N
22	34	N
23	36	N
16	37	N
24	38	N
14	39	N
18	39	N
21	44	N
13	47	N
17	51	Y
12	53	N
15	56	Y
10	62	Y
11	65	Y
7	71	Y
8	87	Y
2	89	Y
6	97	Y
3	98	Y
9	109	Y
4	120	Y
1	143	Y
5	150	Y

Observations:

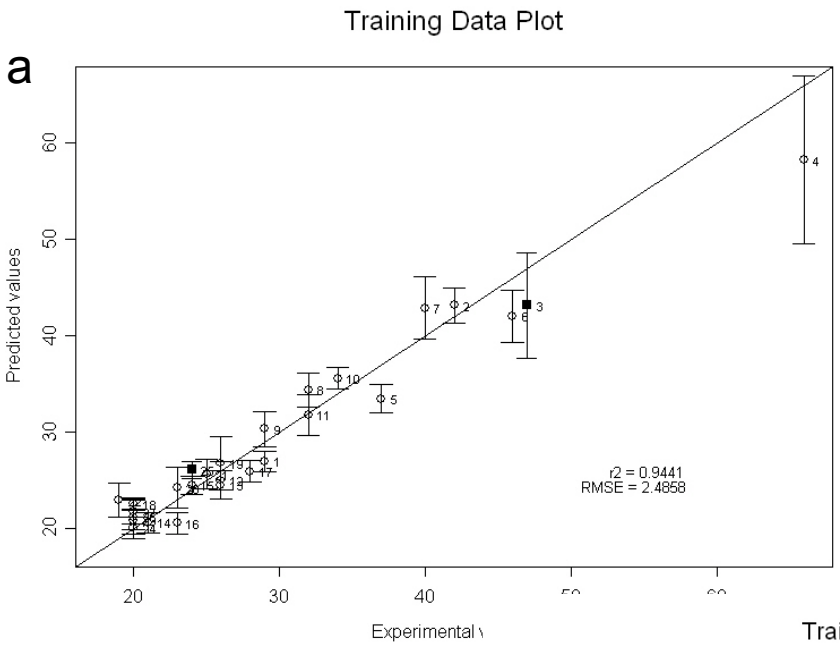
1. All the higher retention occurred on aromatic rings
2. Ligand 5 has the highest binding affinity for all three proteins
3. Ligand 1 shows low binding affinity for RNAse A and Cyt C but high affinity for hydrophobic protein lysozyme

Ligand Based QSPR Models

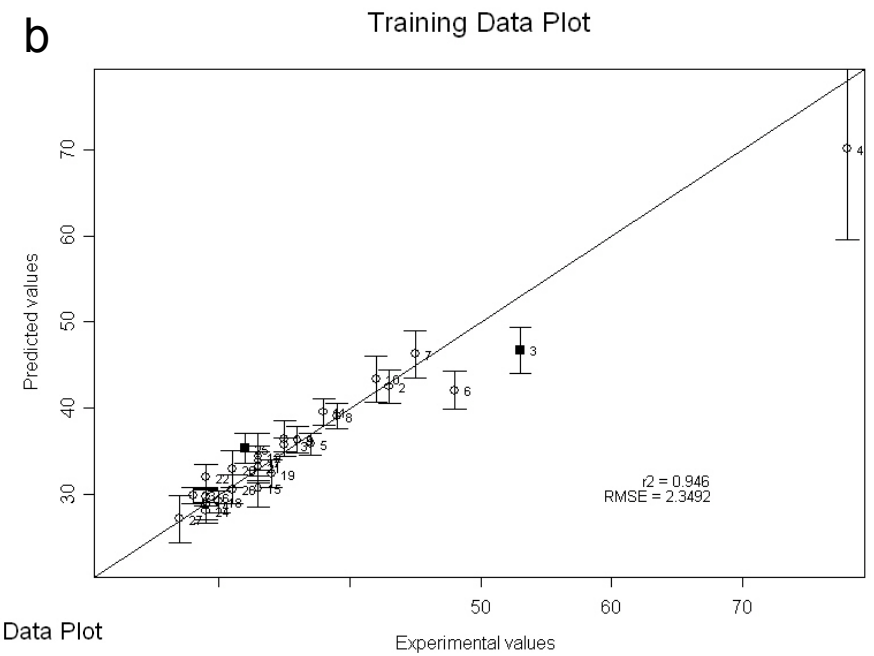
- Ligand structures were drawn in MOE and energy minimized.
- The ACD/p*Ka* DB package (Advanced Chemistry Development) was employed to compute the p*Ka*'s of the charge centers on the ligand molecules. These p*Ka* values were than used to assign the charges on the ligand molecules at pH 6.8.
- A set of 132 molecular descriptors were calculated based on the structures of these cation exchange ligands.
- The resulting descriptors were used to generate the QSPR models for predicting the elution conductivity of three test set proteins.
- Interpretation of the selected descriptors was employed to provide insight into the important physicochemical properties and structural characteristics required for protein binding under high salt conditions.

Predictions for Rnase A (a), Cyt C (b) and Lys (c)

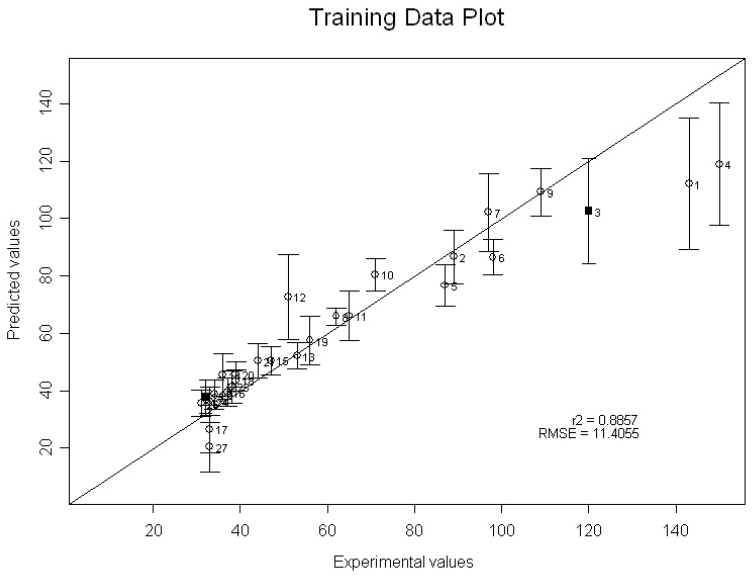
a



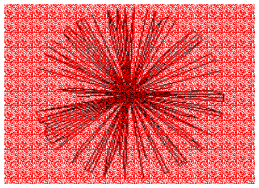
b



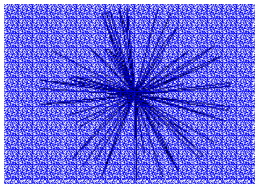
c



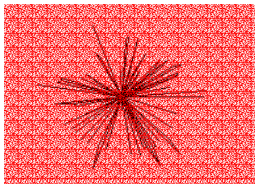
QSPR model for elution conductivity of RNase A



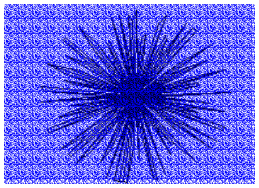
PEOE.VSA.1 +17.4%



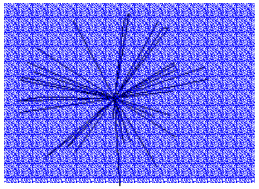
Q.VSA.FPOL -16.7%



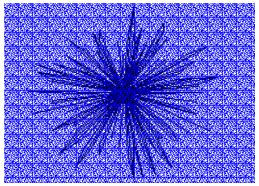
Q.VSA.FNEG +12%



SLOGP.VSA3 -20.9%



PEOE.RPC -10.6%



SLOGP.VSA4 -22.4%

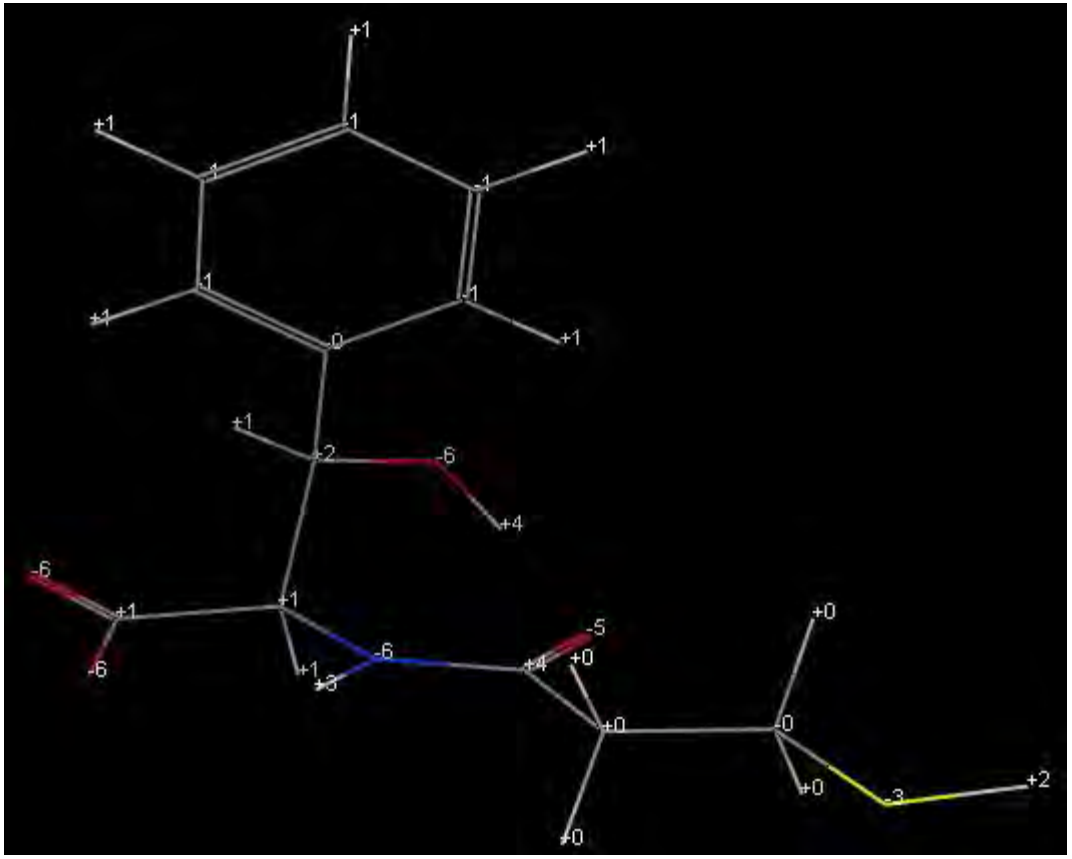
PEOE.VSA 1 and Q.VSA F NEG had positive contributions and represent the negative partial charge at specific bin levels.

PEOE.RPC which represents the positive partial charge of the molecule, showed a negative contribution to the model.

Q.VSA.FPOL represents the fractional polar surface area and had a negative contribution to the model. This is a surrogate for the presence of hydrogen bond donors on the ligands.

Descriptors SLOGP.VSA3 and SLOGP.VSA4 which represent intermediate levels of hydrophobicity were shown to be the most important negative contributors to the RNase A model.

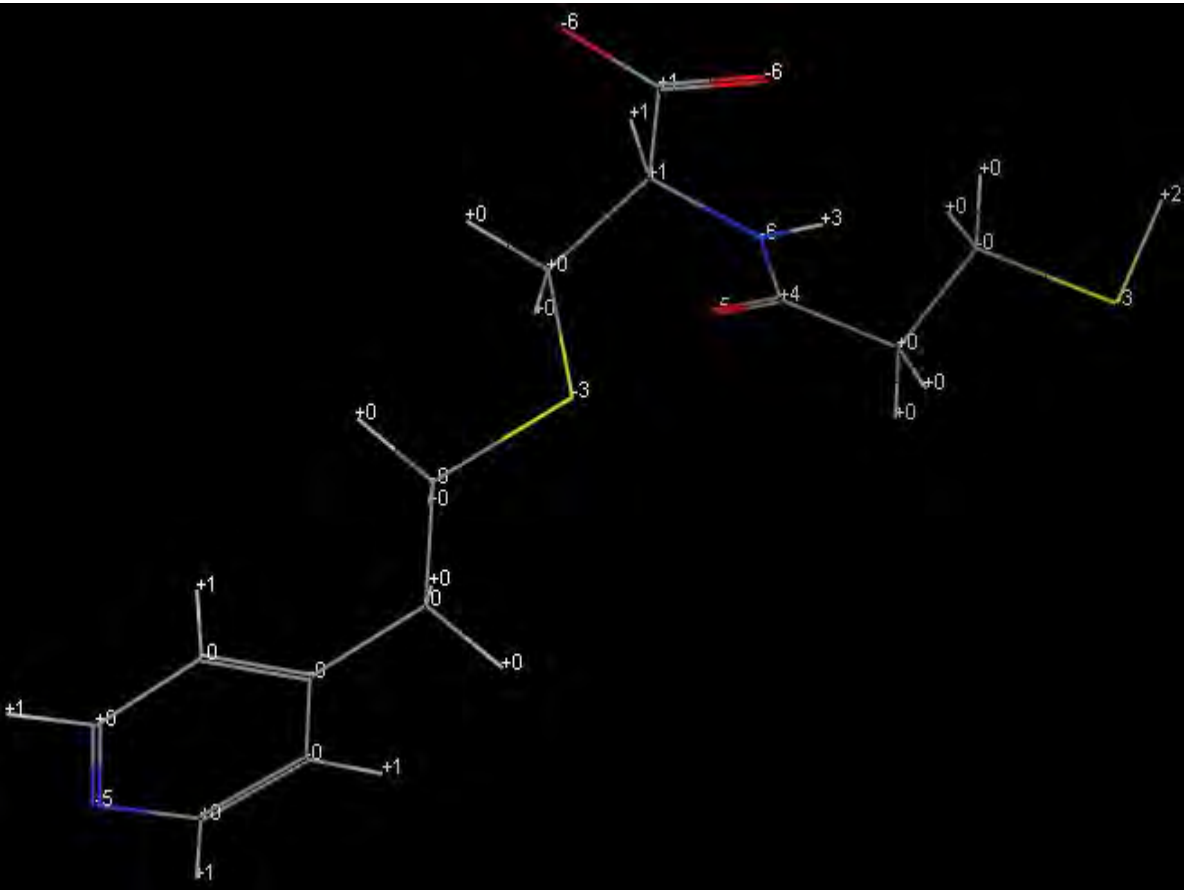
The structure of ligand 2



Red: oxygen atom
Blue: nitrogen atom
Yellow: sulfur atom

Descriptor PEOE.VSA.1 (PEOE.VSA-1) which was shown to be the most positive contributor to the Rnase A model was only assigned to the carbon atoms on the aromatic ring of the ligands.

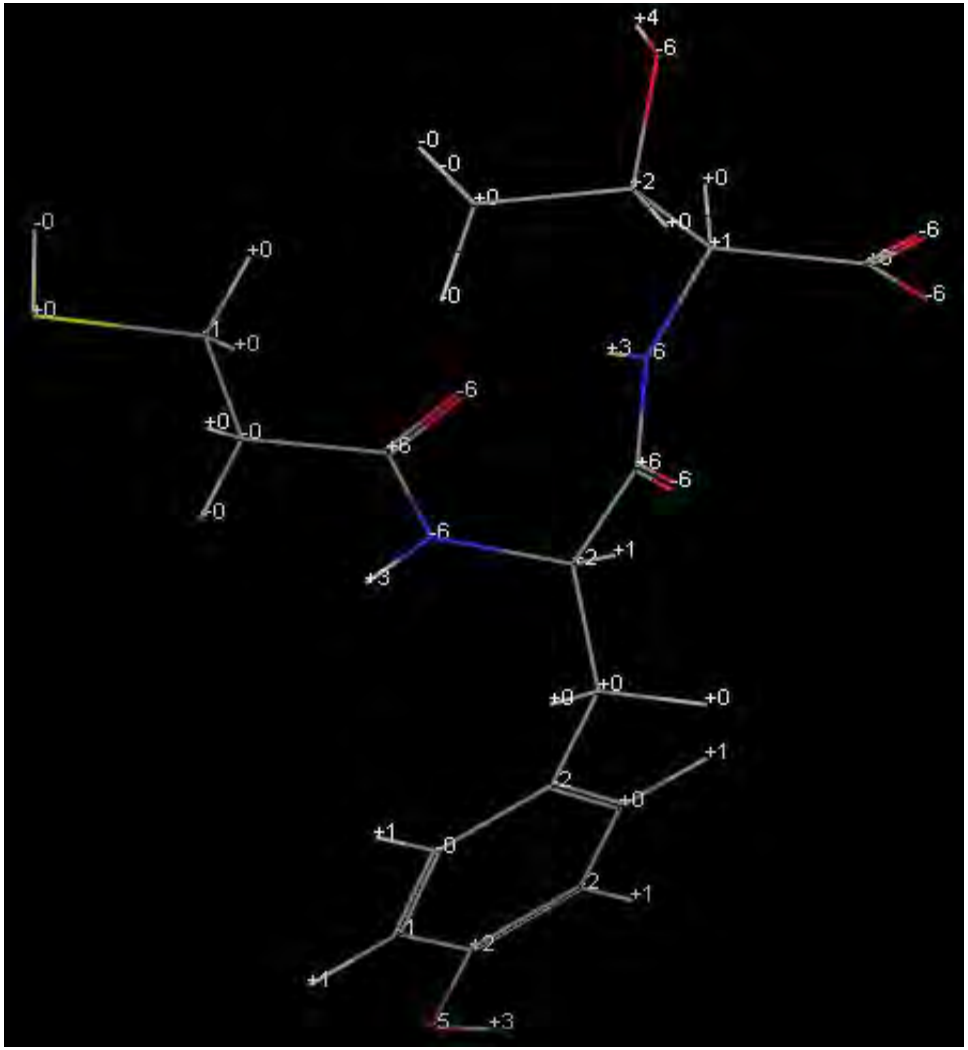
The structure of ligand 1



Red: oxygen atom
Blue: nitrogen atom
Yellow: sulfur atom

Not every aromatic containing ligand had carbon atoms with descriptor values of PEOE.VSA.1

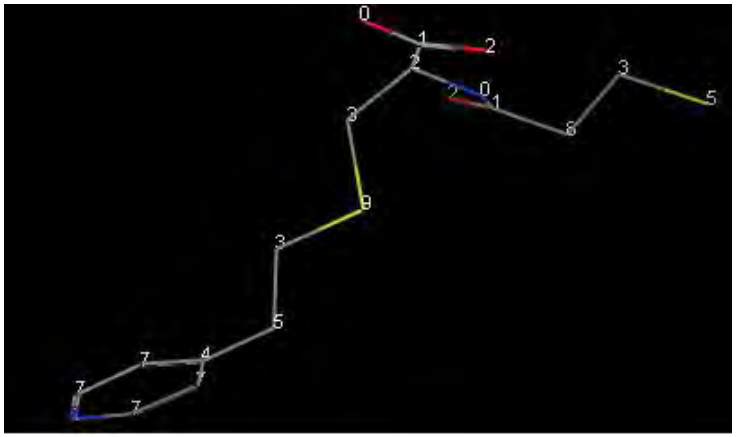
The structure of ligand 7



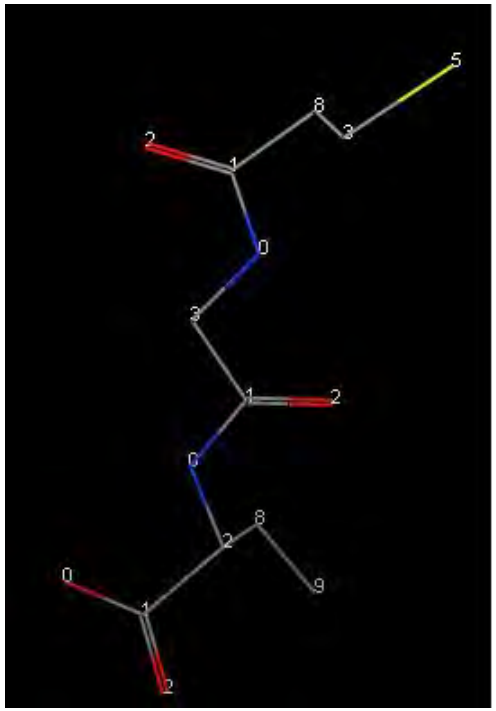
Red: oxygen atom
Blue: nitrogen atom
Yellow: sulfur atom

Q.VSA.FPOL which had a negative contribution to the RNAse model is a surrogate for the presence of hydrogen bond donors on the ligands (e.g. –NH and –OH) (i.e. atoms assigned with absolute numbers equal or greater than 4)

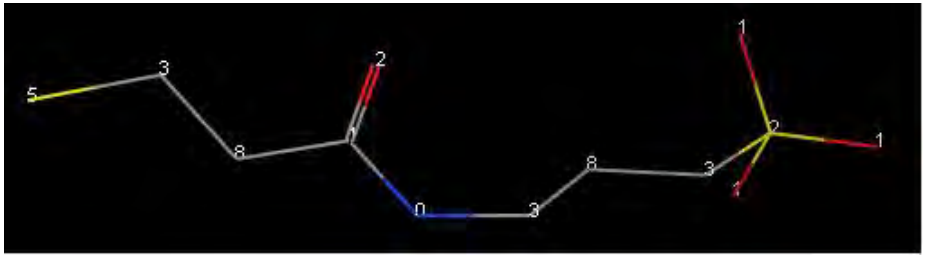
The structure of ligand 1(a), 12(b), 14(c)



(a)



(b)

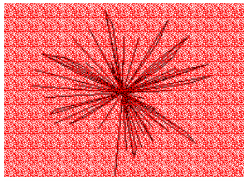


(c)

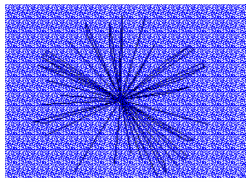
Red: oxygen atom
Blue: nitrogen atom
Yellow: sulfur atom

The numbers associated with the atoms indicate the categorized bin of SlogP (index of hydrophobicity) of the atoms. Descriptor SLOGP.VSA3 was mainly associated with aliphatic carbons or carbons which were adjacent to sulfur atoms.

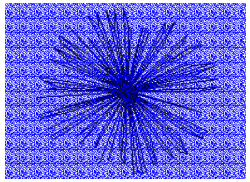
QSPR model for elution conductivity of horse cyt C



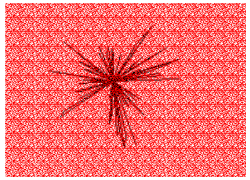
PEOE.VSA.FHYD +13.8%



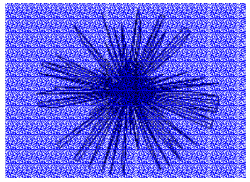
PEOE.RPC -15.1%



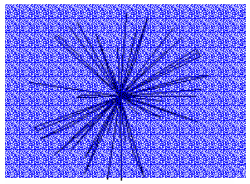
PEOE.VSA.3 -21%



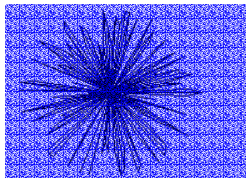
PM3.DIPOLE +8.6%



SLOGP.VSA3 -15.3%



Q.VSA.FPOL -7.5%



SLOGP.VSA4 -18.7%

PEOE.VSA.FHYD represents the fractional hydrophobic van der Waals surface area. (positive contribution)

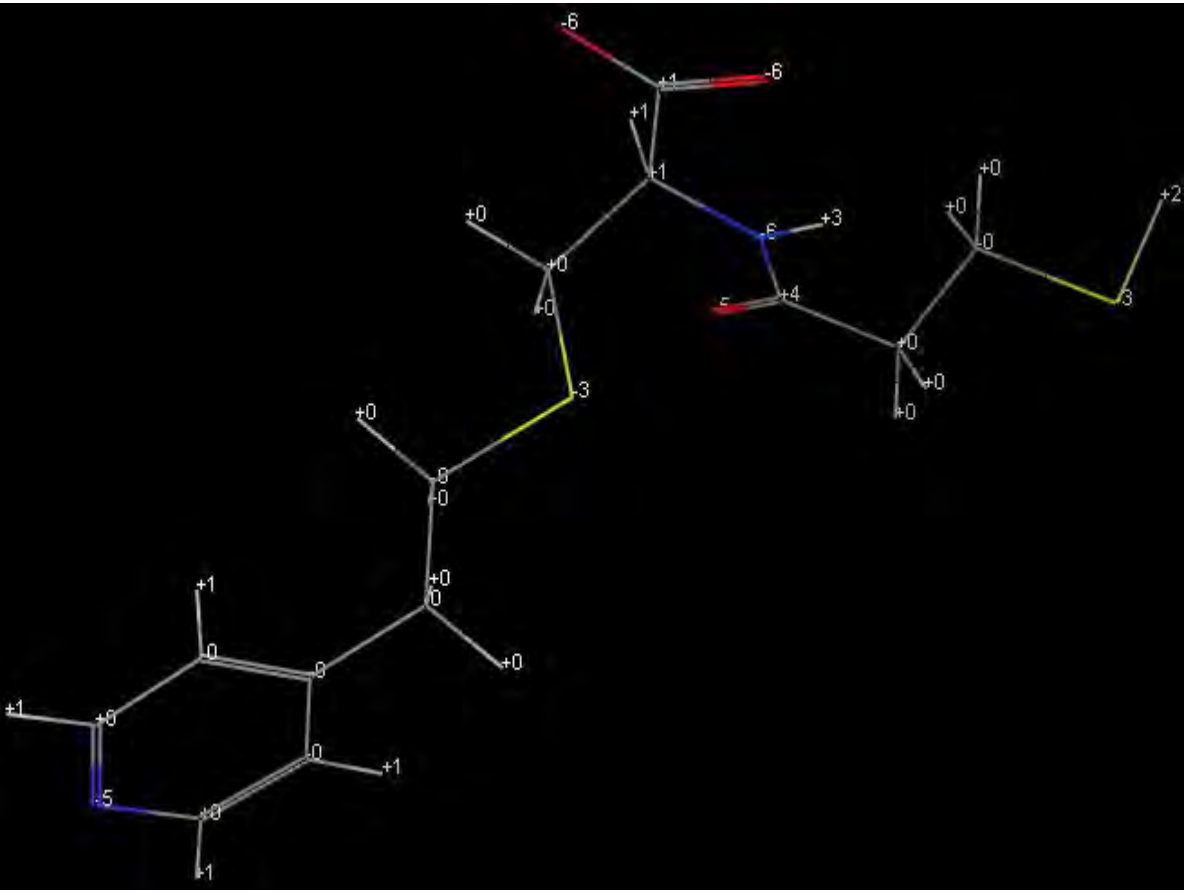
PM3.DIPOLE is the dipole moment of the molecule. (positive contribution)

Thus, both hydrophobic and electrostatic interactions are both important in this model.

PEOE.VSA.3 corresponds to the van der waals surface area with negative partial charge. (negative contribution)

Q.VSA.FPOL, PEOE.RPC, SLOGP.VSA3 and SLOGP.VSA4 showed negative contributions to the binding of cytochrome C (same as the RNase A model).

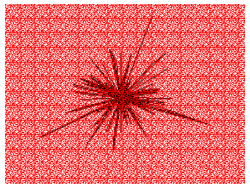
The structure of ligand 1



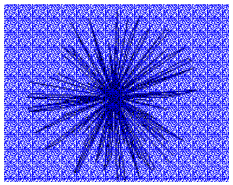
Red: oxygen atom
Blue: nitrogen atom
Yellow: sulfur atom

PEOE.VSA.3 is associated with sulfur and fluorine atoms in the aliphatic chains.

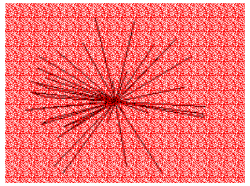
QSPR model for elution conductivity of lysozyme



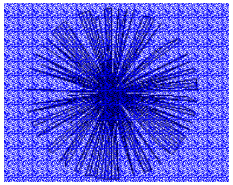
LOGP.O.W. +11.2%



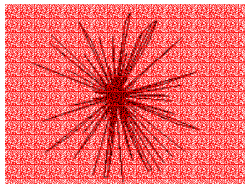
PEOE. VSA.FPNEG +14.3%



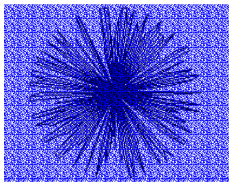
RADIUS +10.7%



BALABANJ -24.1%



PEOE.VSA.2 +10.4%



SLOGP.VSA4 -29.4%

LOG.O.W (the hydrophobicity of the ligand) was the most positive contribution (aromatic containing ligands had higher values of LOG.O.W. as compared to aliphatic containing ligands).

radius (a size related descriptor) had a positive contribution to the model (i.e. the larger the size of the ligand the stronger the binding).

PEOE.VSA.2 (a slightly negative partial charge) was a positive contributor. (note: aromatic substructures containing ligands having one or two associated sulfur atoms).

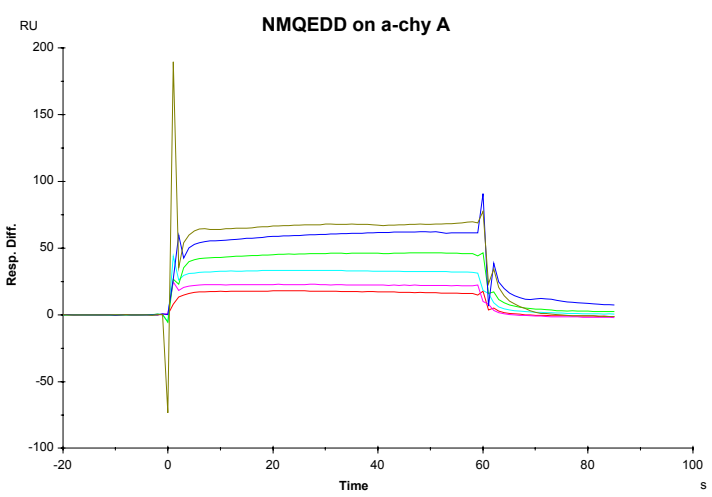
PEOE.VSA.FPNEG (fractional negative polar surface area) and was only assigned to structural components that could form hydrogen bonds with the protein.

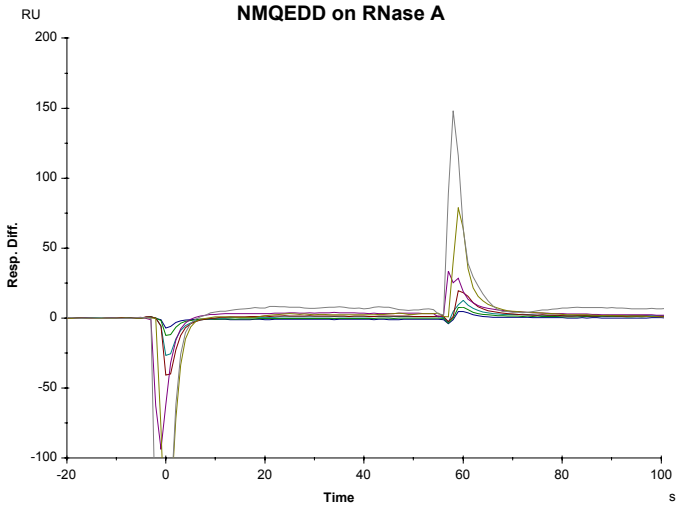
BALABANJ increases with branching and the number of carbon atoms. Non-aromatic ligands had larger values.

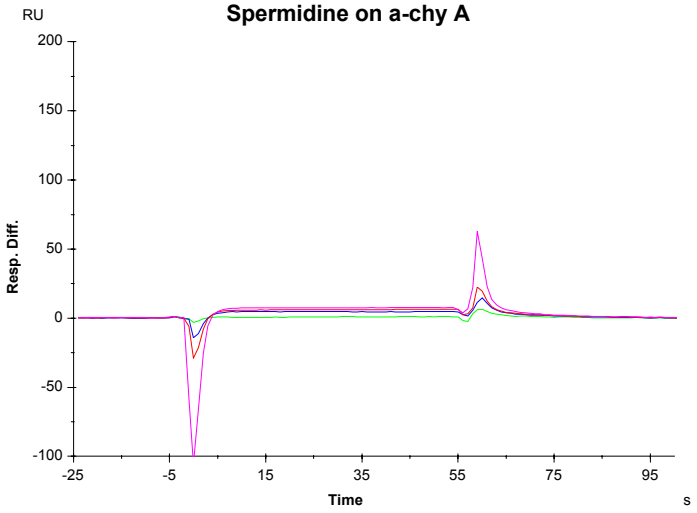
SLOGP.VSA4 associated with the carbon atoms on aliphatic chains

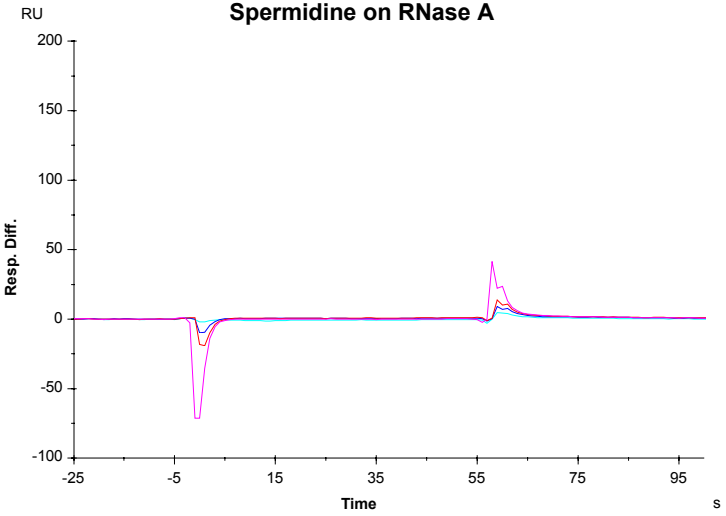
Other tools to study mixed mode systems.

SPR analysis of chemically selective displacer



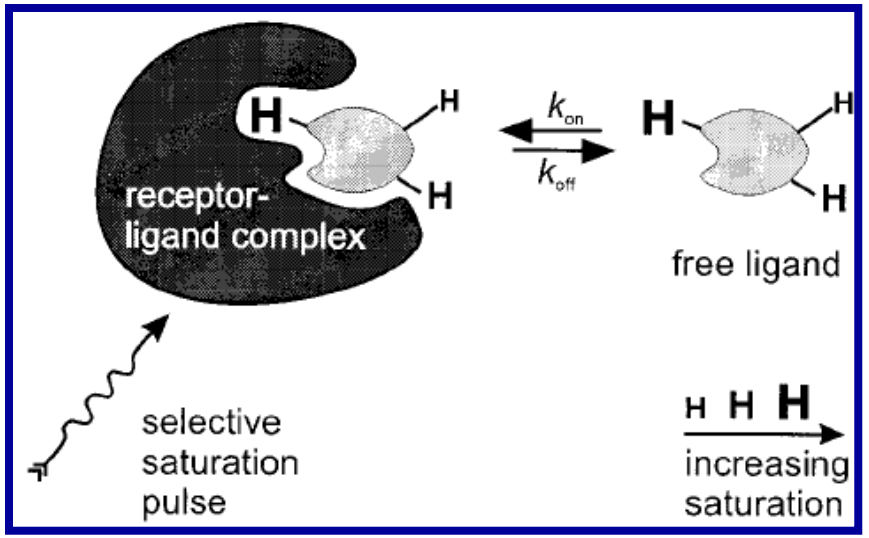




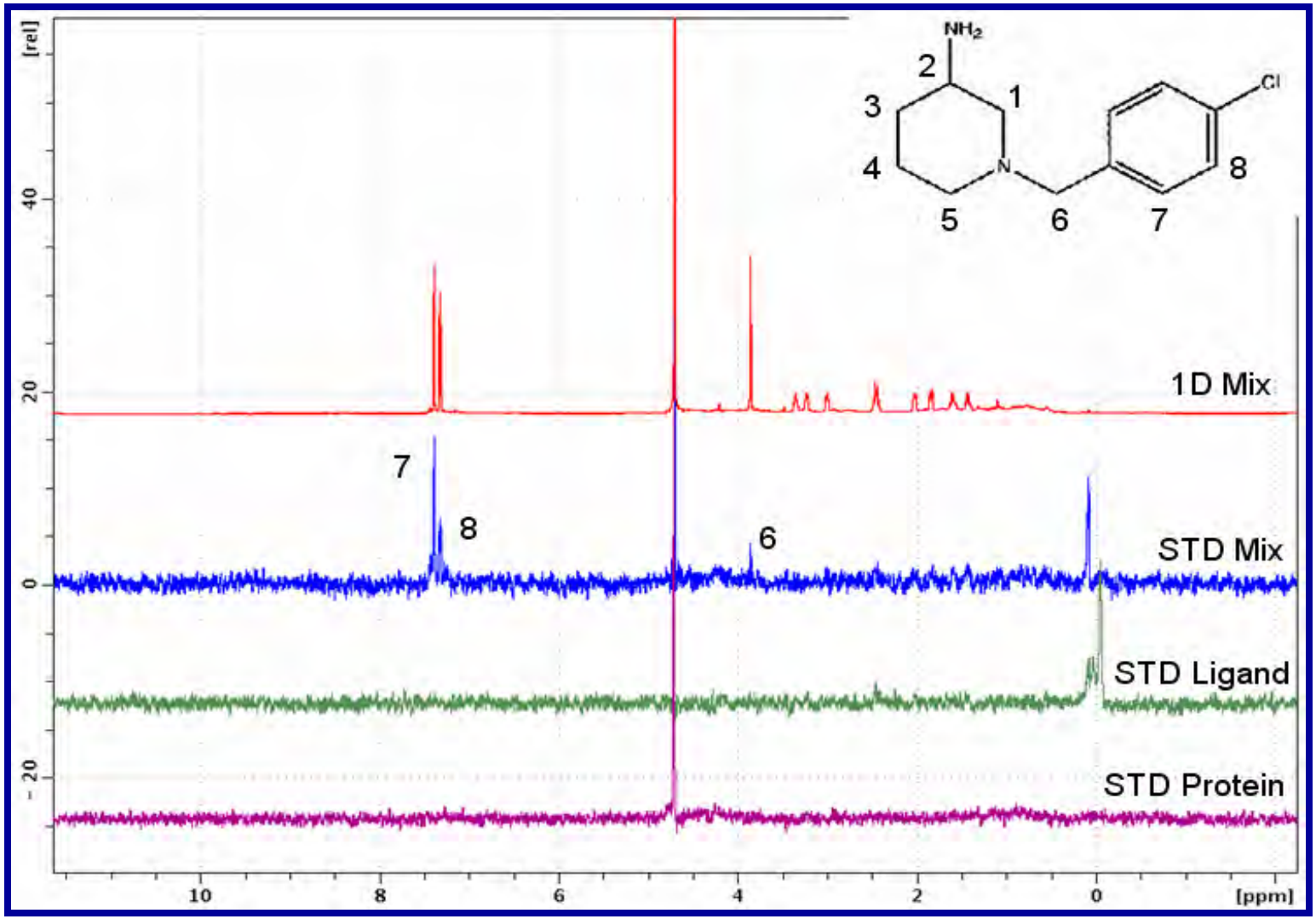


Saturation Transfer Difference (STD) NMR Experiment

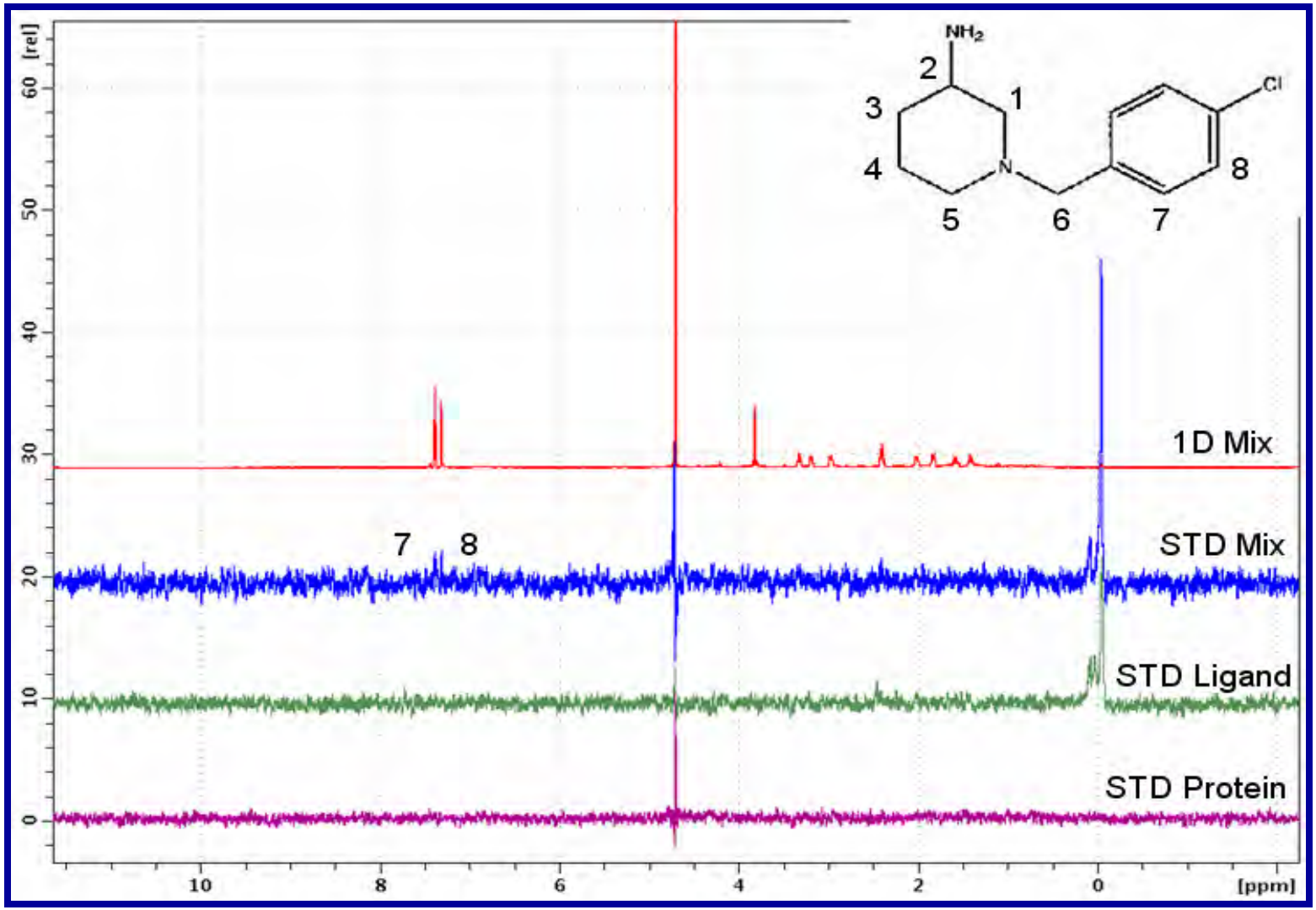
- Selectively saturate your protein
- Spin diffusion spreads magnetism throughout protein (bonds)
- Dipolar coupling spreads magnetism to contacting ligand (space)



STD NMR Result: 1-(4-Chlorobenzyl)piperidin-3-aminesulfate /Alpha Chy A

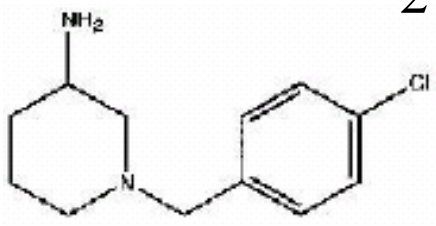


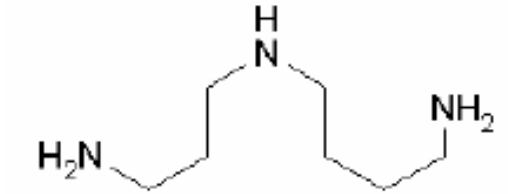
STD Result: 1-(4-Chlorobenzyl)piperidin-3-aminesulfate / Ribonuclease A

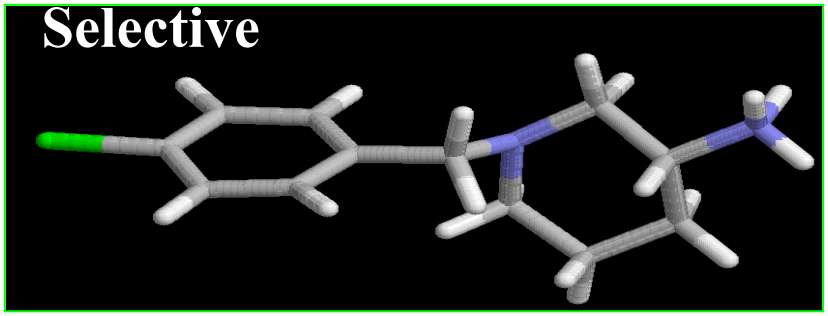


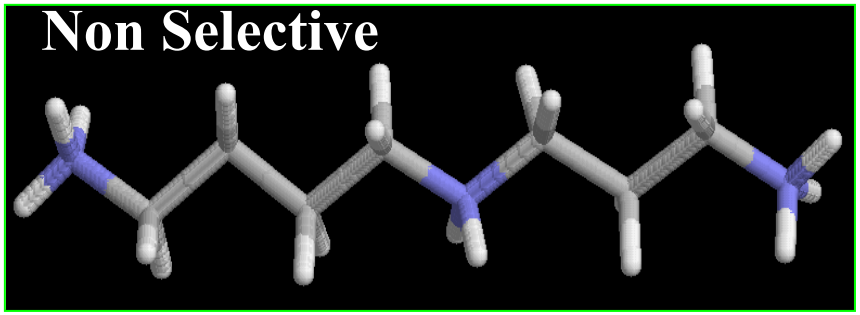
MD Simulation

2 proteins : α Chy and Rnase A









1-(4-Chlorobenzyl)piperidin-3-aminesulfate

Simulation Parameters:

Non Bond, Bond, Angle etc: AMBER

Charges: Not found in literature Chlorobenzene simulations, AMBER

Total Charge: +1

Spermidine

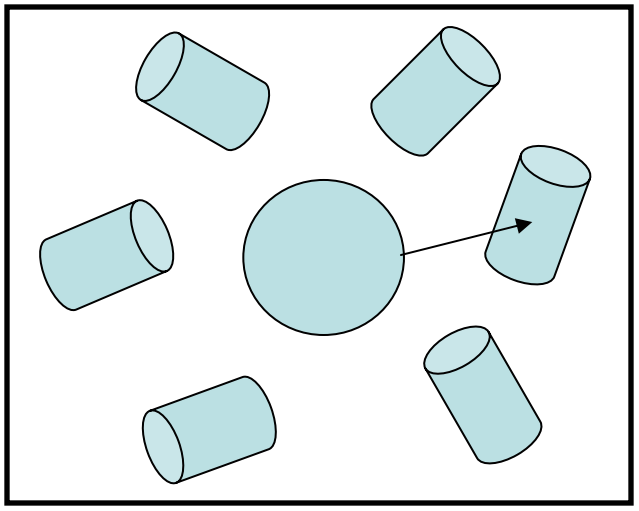
Simulation Parameters:

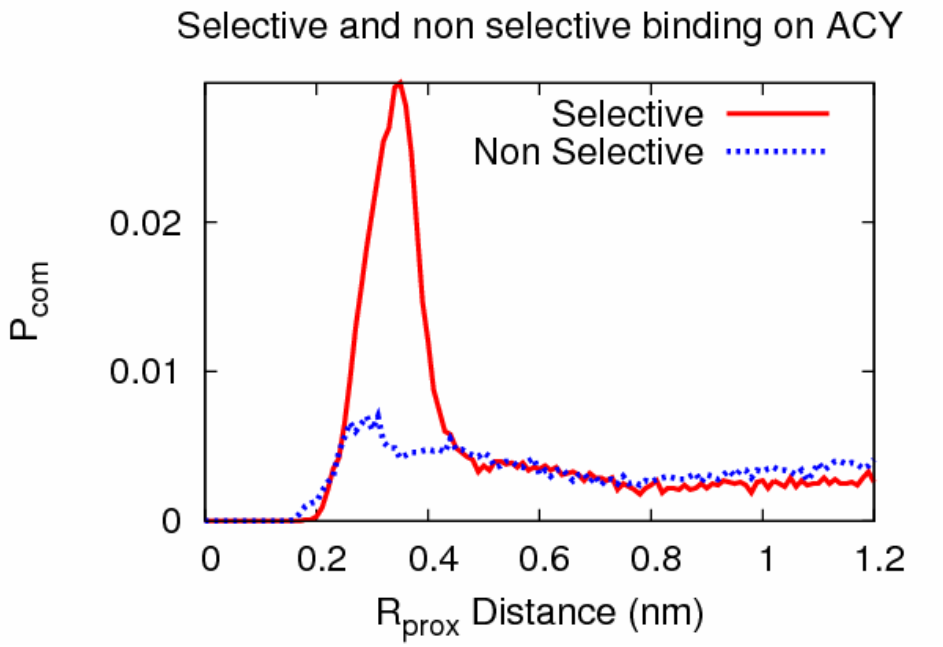
Non Bond, Bond, Angle etc: AMBER

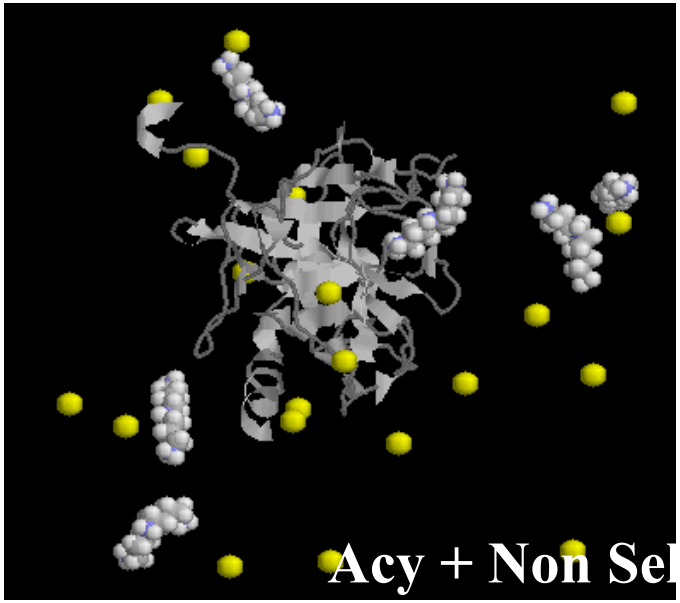
Charges: Korolev et al

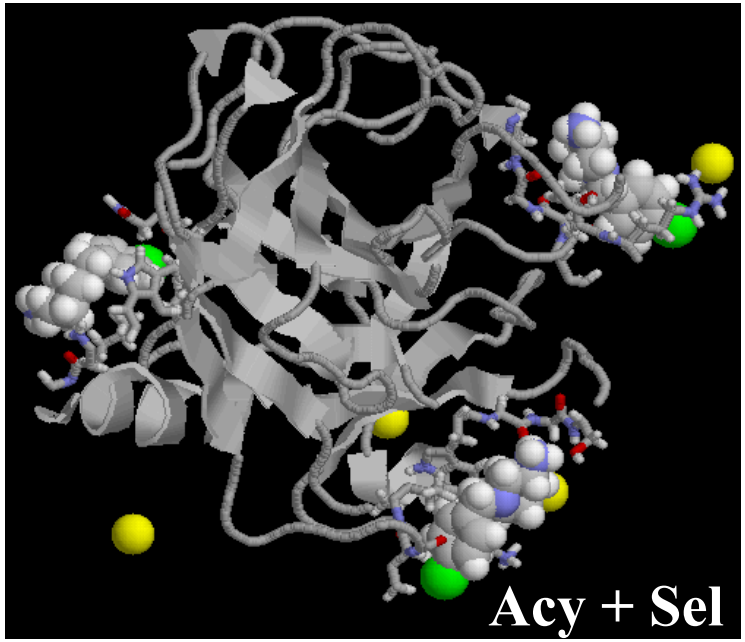
Total Charge : +3

Simulation Results

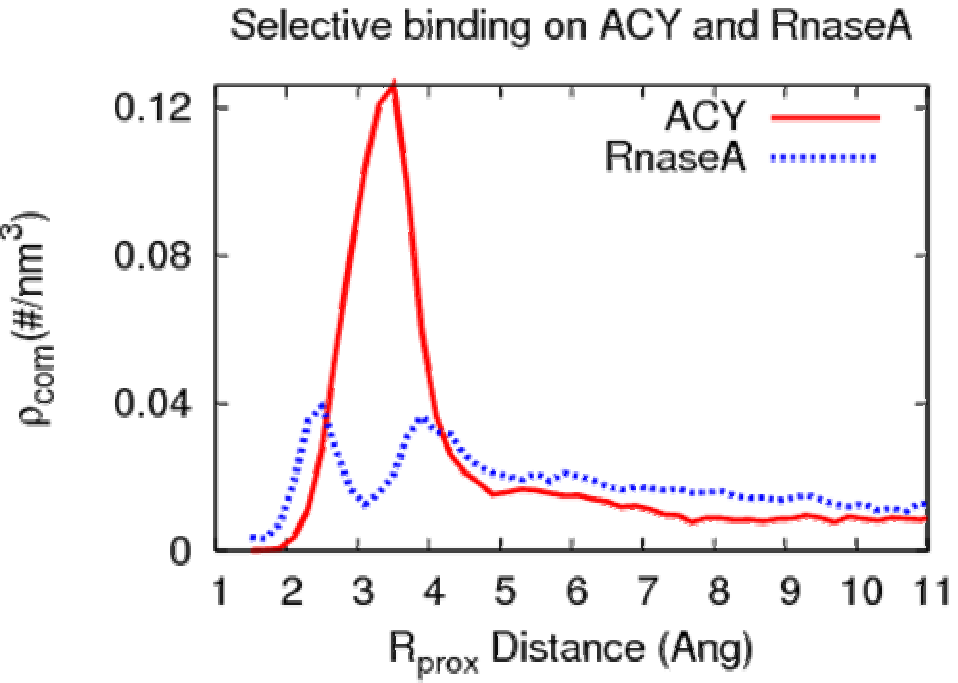


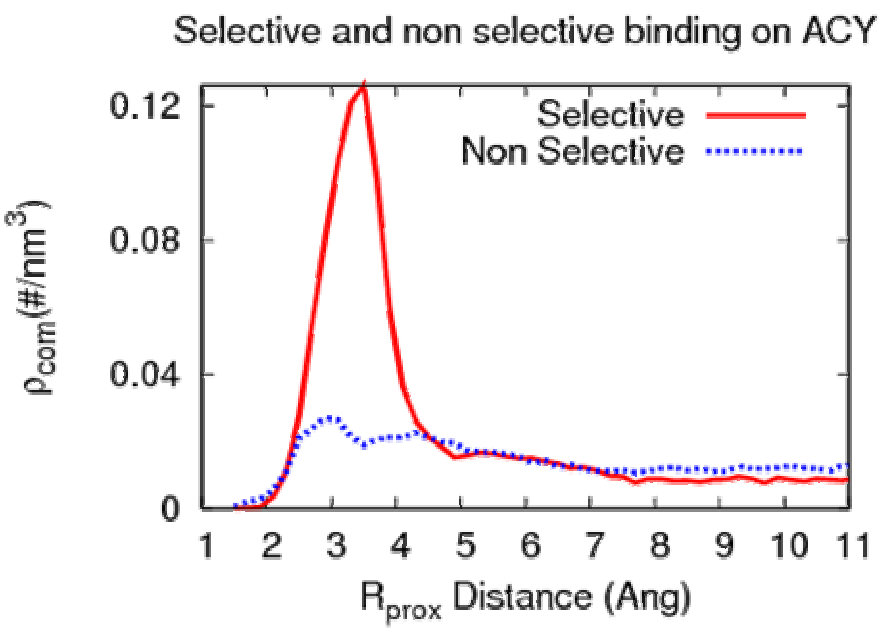


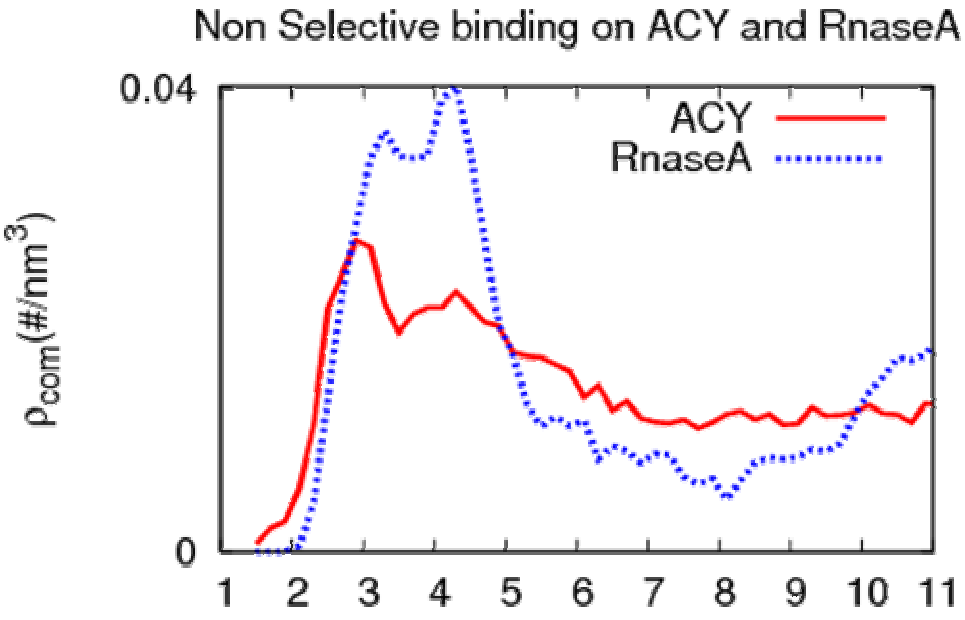




Displacer COM Density

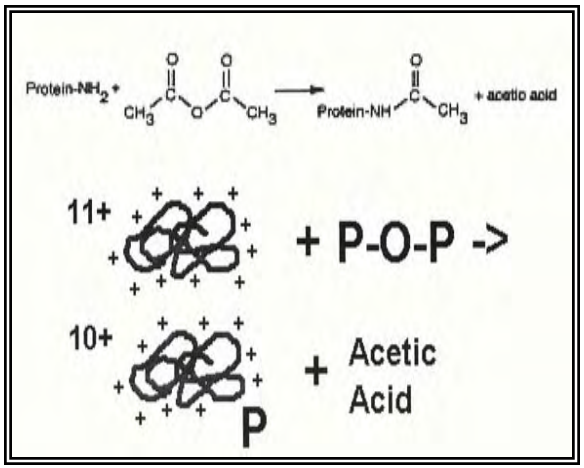




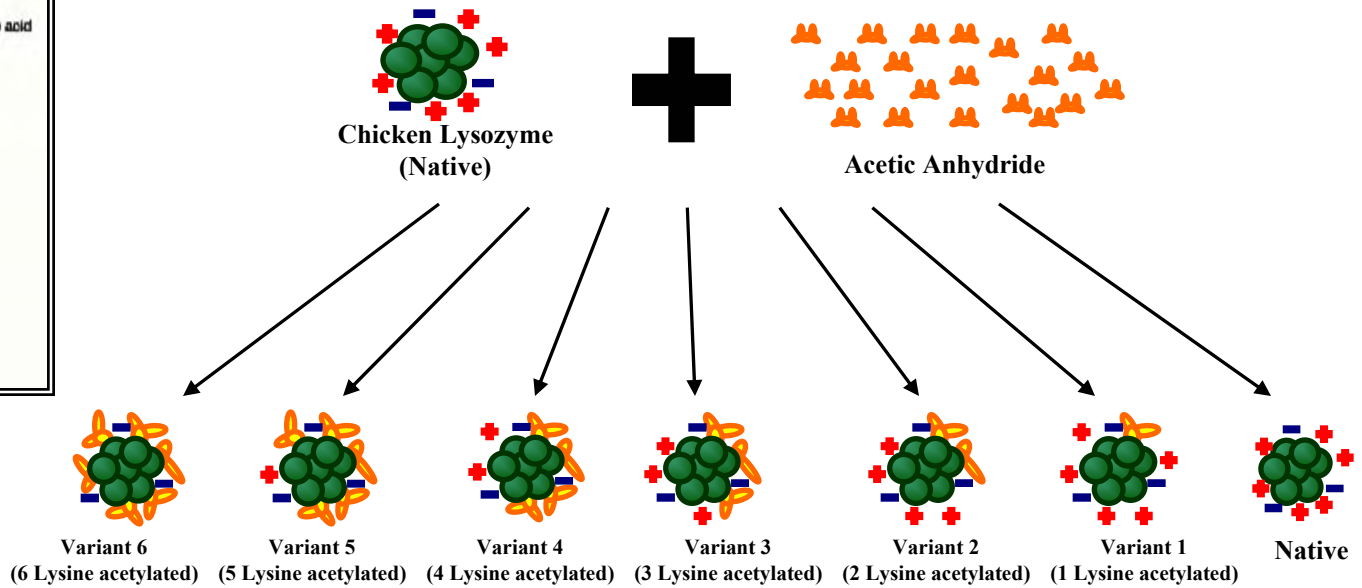


Collective statistics show selective displacer binds more strongly to one protein than other.

Protein Charge Ladders

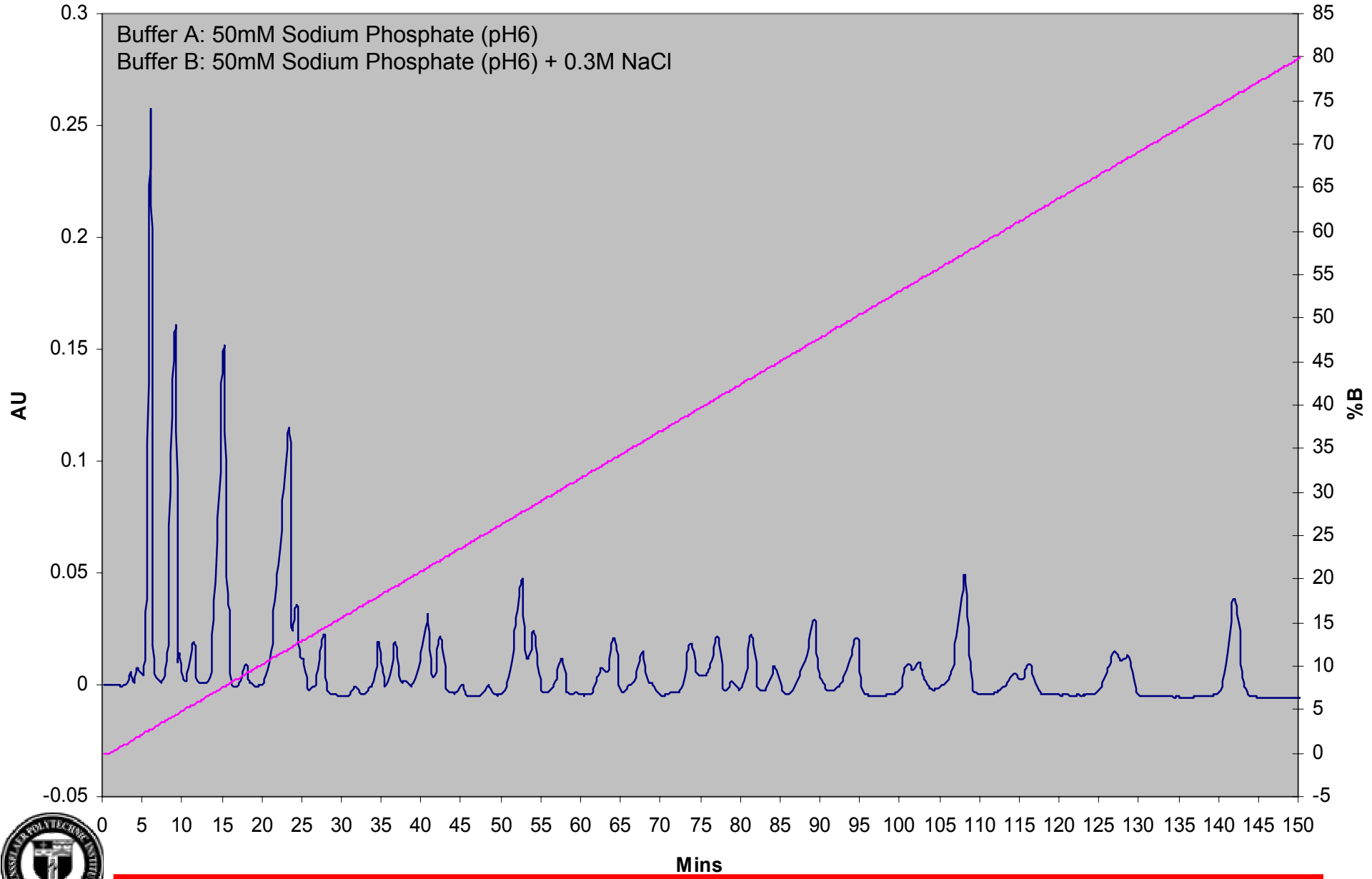


Reference: Menon et al (2000) Anal. Chem

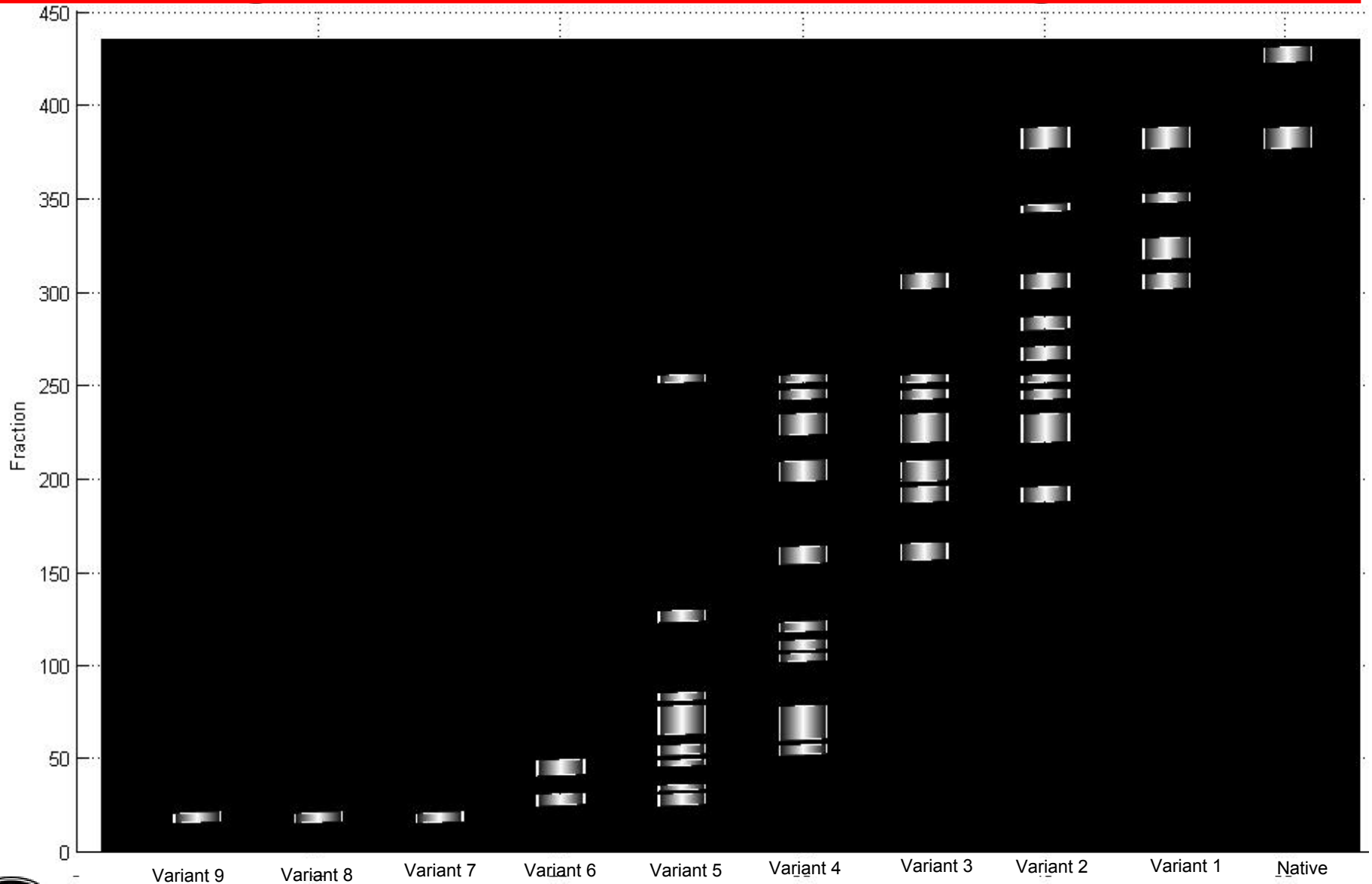


- Involves functionalization of specific amino acids on protein surface with chemical modifiers
- Resulting variants differ by single units of charge
- Gives a heterogeneous protein mixture with similar structure but varying charge distribution
- Acetic anhydride used to functionalize surface lysine residues (6) of chicken lysozyme

SCX Analysis of Lysozyme Ladder Mixture



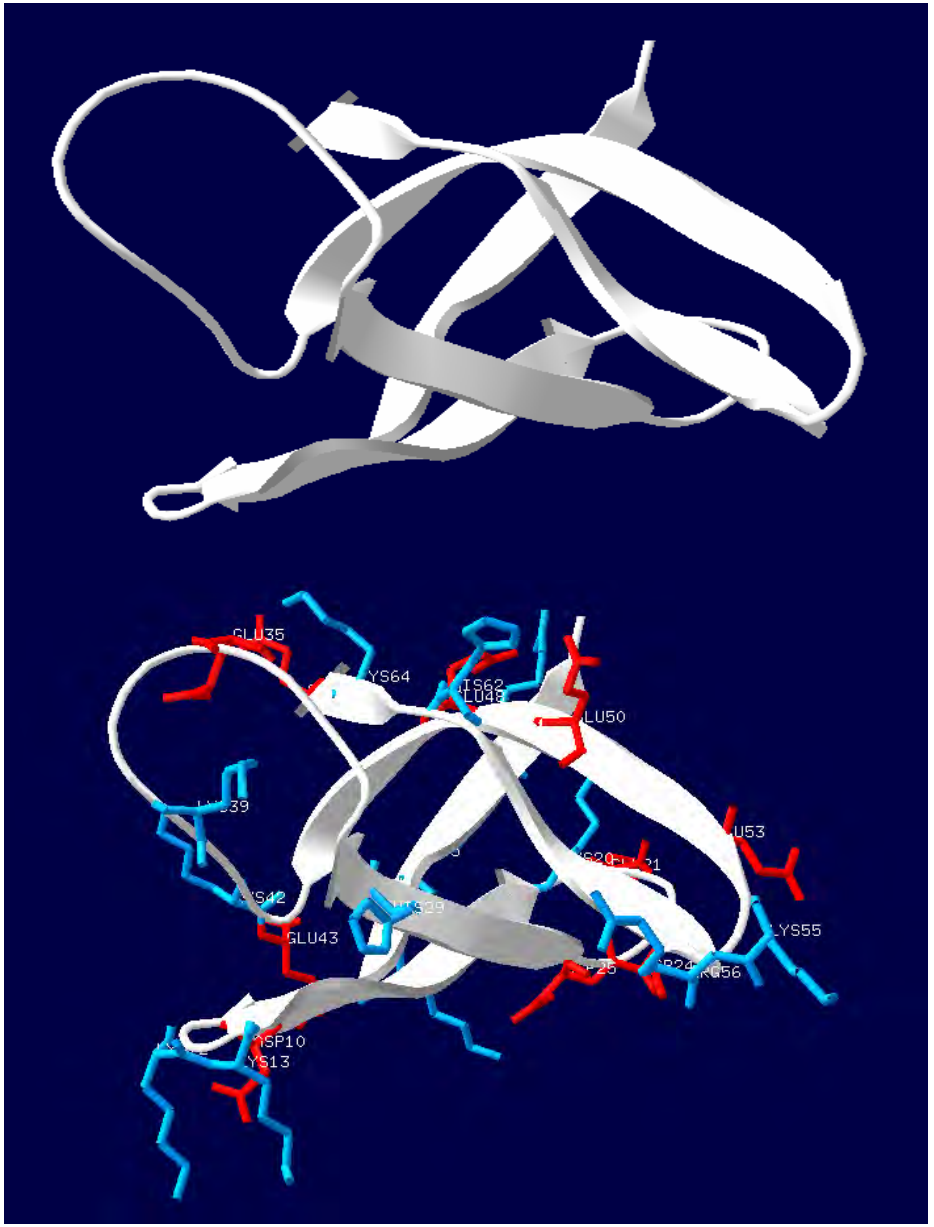
Charge Ladder Peak Assignment



Homologous Protein Library

Protein Selected:

- Cold Shock Protein B (CspB)
- Mw: 7.373 kDa
- No. of Amino Acids: 67
- Protein pI: 8.05 (Calculated)
- Extinction coefficient: 5,690 /M cm

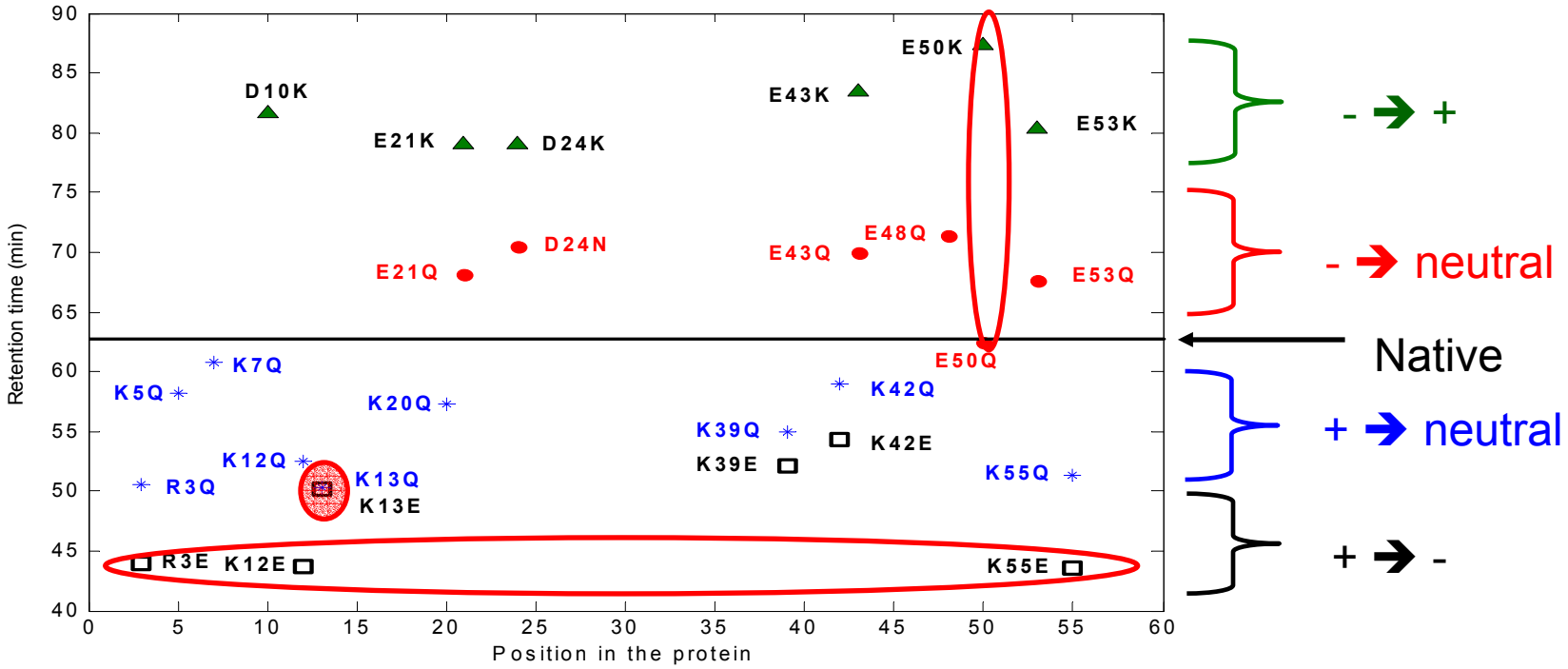


Homologous Protein Library

Wild Type Template: M36G

Positive ➔Negative	Positive ➔Neutral	Negative ➔Positive	Negative ➔Neutral
R3E	R3Q	D10K	E21Q
K12E	K5Q	E21K	D24N
K13E	K7Q	D24K	D25N
K39E	K12Q	D25K	E43Q
K42E	K13Q	E43K	E48Q
K55E	K20Q	E50K	E50Q
	K39Q	E53K	E53Q
	K42Q		
	K55Q		

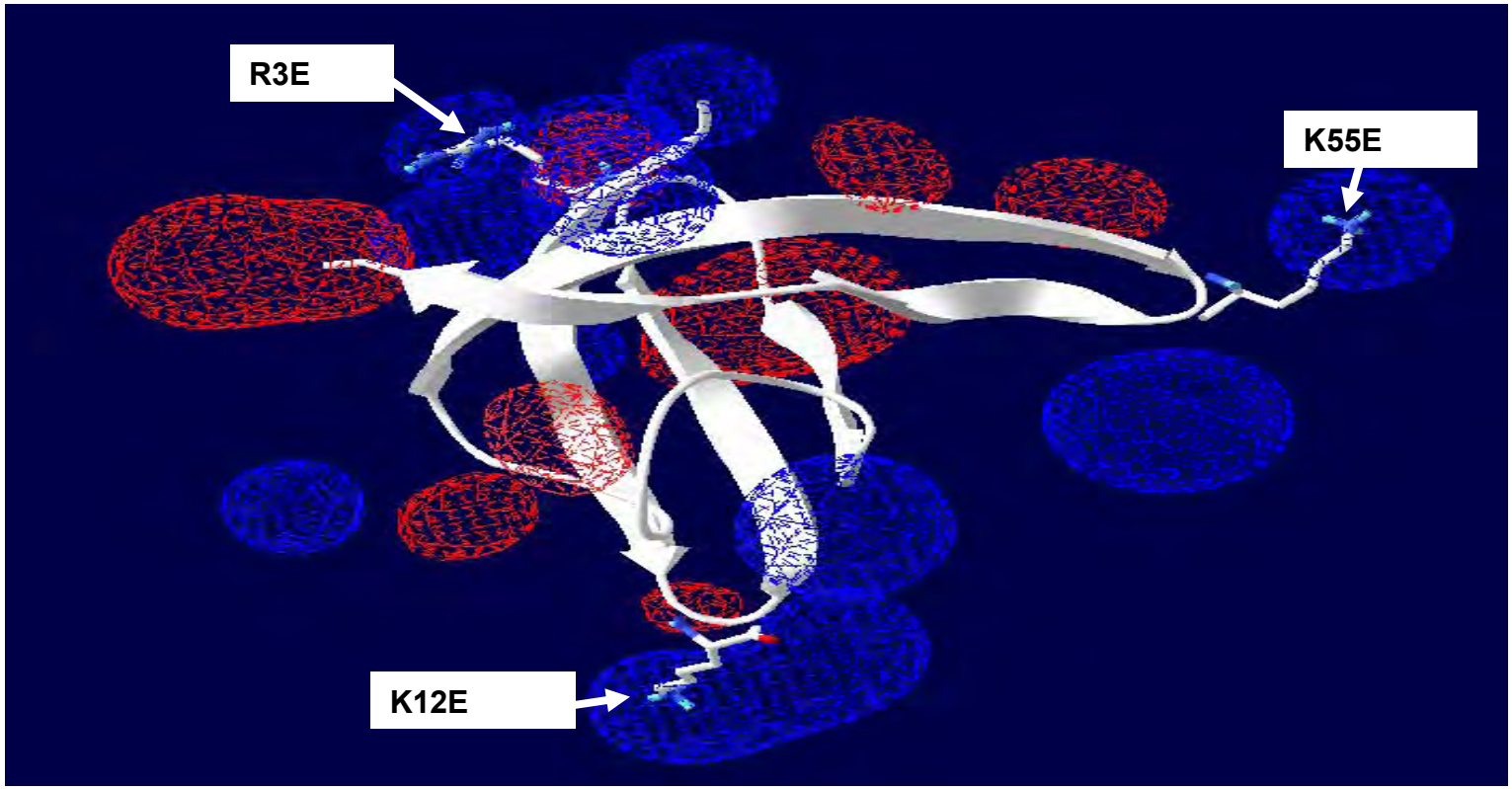
CEX Retention data for CspB mutants



- Neutralization of charge at position 50 had no effect on protein retention
- However alteration of the charge from negative to positive yielded a significant increase in retention time
- No appreciable difference was observed when the charge species at position 13 was neutralized or changed.
- Changing the charge species at positions 3, 12 and 55 reduced retention time to similar extents

Effects of Mutation on Protein Binding

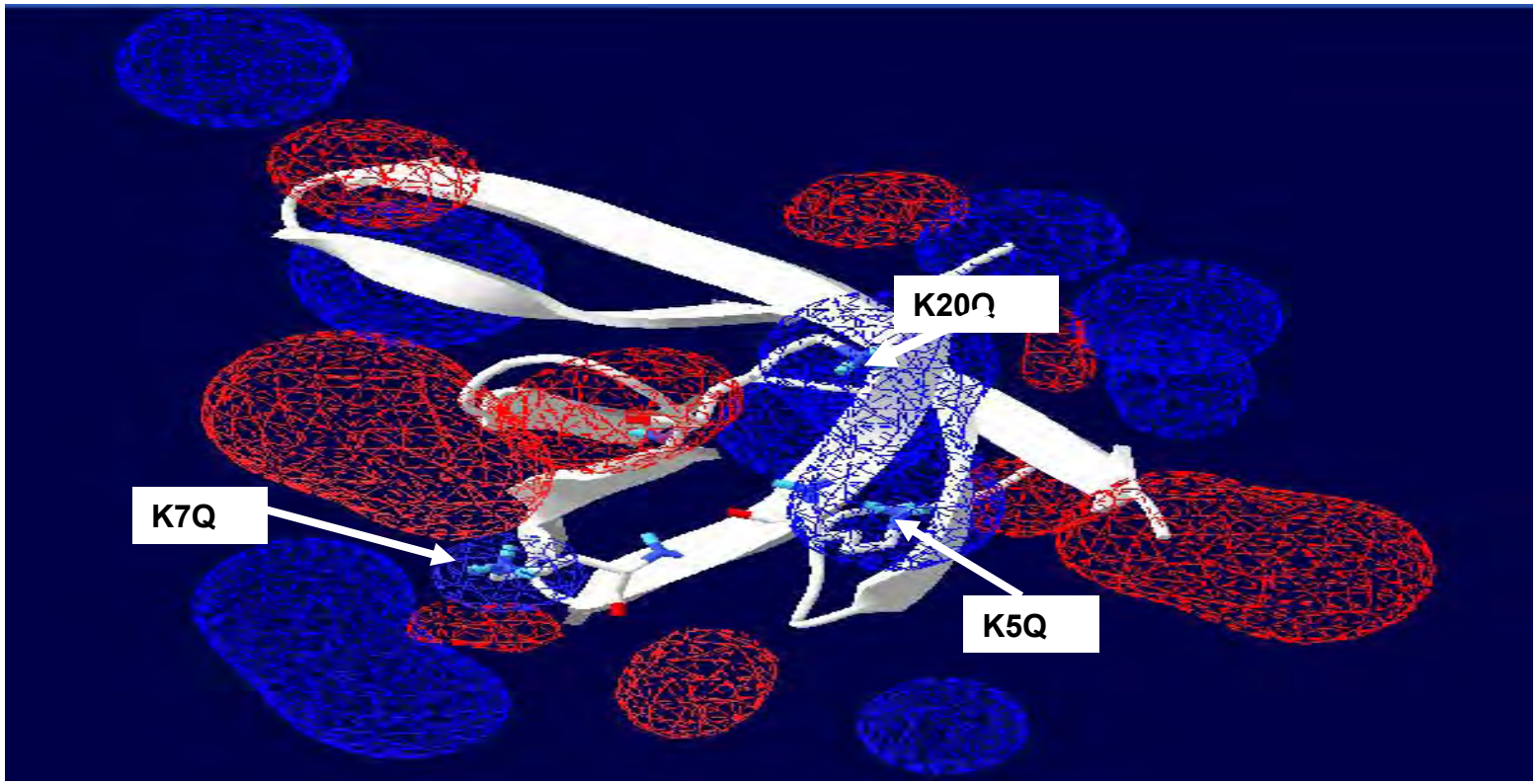
Positive ➔ Negative



- Positions 3, 12 and 55 are located at protruding regions of protein surface
- Allows for interactions with resin surface at larger distances
- Reversal of charge caused significant reduction in retention time

Effects of Mutation on Protein Binding

Positive ➔ Neutral



- Positions 5, 7 and 20 surrounded by major negative charge clusters
- Minimal interactions of these regions with resin

Summary

- QSPR can be employed to provide insight into the design of mixed mode chromatograhic systems.
- The selected features of these models illustrate the effect of multiple interaction mechanisms (charge, hydrophobicity and hydrogen bonding) on protein retention in multi-modal cation exchange systems.
- The aromatic ring can play an important role in promoting protein binding under high salt conditions.
- Regions of the ligands with negative partial charge also tend to promote high salt protein binding.

Summary (cont.)

- Moieties associated with intermediate hydrophobicity (e.g. aliphatic side chain) or the presence of hydrogen bond donors (e.g. NH and -OH) tend to suppress the binding.
- While the sulfur atoms were found to have a positive contribution to the lysozyme model, for the RNase A and cyt C models the sulfur atoms reduced binding affinity. Thus, while general trends can be observed for the design of high salt binding ligands, selective binding to various classes of proteins may require unique ligand design.
- Additional experimental and theoretical techniques such as protein ladders, mutant libraries, SPR, NMR, and MD simulations are potentially powerful tools for examining the behavior of mixed mode chromatographic systems.

Acknowledgements

- Students: Ting Yang, Matt Sundling, N. Sukumar, Wai Keen Chung, Chris Morrison
- Professors: Curt Breneman, Shekhar Garde, Ravi Kane
- Useful discussions: Gunnar Malmquist, Enrique Carredano
- Funding: NIH, NSF, GE Healthcare

Rensselaer