

**UK QSAR
& Chemoinformatics**

Spring Meeting

14th April 2005

UniS
University of Surrey
Guildford

idbs



MERCK
Research Laboratories

Multivariate Analysis of GPCRs

UK-QSAR Spring Meeting

14th April 2005

Philip Evans

University of Portsmouth

What are GPCRs?

- G Protein-Coupled Receptors
- 7 Trans-Membrane Helices (aka 7TMs)
- Conformational change upon activation leads to 2nd messenger pathways cAMP, cGMP, IP3
- Perhaps as many as 40 - 50 % of drugs on market act on < 10% of GPCRs
 - Zantac
 - Seretide
 - Imigran



G-Proteins bind to intracellular loops

Research Aims

- To **identify** which **positions** on the trans-membrane helices are **involved** in ligand binding
- To identify the **properties** of the amino acids at these positions that are **important** for ligand **binding**

Why?

- For thorough and rapid analysis of large and sparse GPCR datasets
- To identify likely positions of ligand binding
- To generate hypotheses for the nature of the binding
 - based on the identification of side chain properties and their possible interactions with the ligand
- To help modelling teams and chemists reduce the time for development of prospective drug candidates

Concept

- Any given ligand will have different affinities for different receptors
 - Receptor specificity
- This is due to the variation in sequence between receptors
- Regression is used to explain variation in observed data
- So multivariate regression analysis of the sequence variation could explain the differences in the affinities

Sequence Data Required

Receptors (rows)	Y	T	L	L	A	N	Y	A	V	V	D	T	Q	F	I	D	V	C	T	H	M	S	I	G	T	Y	T	S	T	F	L	P	L	A	F	F	W	F	I	T	G	I	N	W	G	N	N
	W	L	L	I	T	N	Y	S	V	V	D	T	Q	W	L	D	I	C	T	H	F	S	I	S	V	Y	T	S	T	F	T	P	L	S	F	F	W	F	I	L	F	F	T	W	G	N	N
	L	L	L	I	T	N	Y	S	V	V	D	T	Q	W	L	D	I	C	T	H	L	S	I	S	V	Y	T	S	T	F	S	P	L	S	F	F	W	F	I	P	F	F	T	W	G	N	N
	E	I	L	I	T	N	Y	S	V	V	D	T	Y	W	L	D	M	C	T	H	L	S	I	S	V	Y	T	S	T	F	L	G	L	E	F	F	W	F	I	S	A	L	T	W	G	N	N
	S	V	L	L	T	N	Y	S	V	V	D	T	Q	W	L	D	I	C	T	H	L	S	I	S	V	S	T	S	T	F	L	N	L	E	F	F	W	F	I	E	S	L	A	W	G	N	N
	E	S	T	V	T	N	T	S	V	L	D	A	S	W	I	D	V	S	T	H	P	S	I	S	V	F	V	G	S	F	L	V	I	N	F	F	W	F	V	G	L	F	V	W	G	S	N
	K	A	I	V	T	N	T	A	V	V	D	A	L	W	L	D	V	S	T	H	P	A	I	S	V	F	M	G	S	F	L	V	I	N	F	F	W	F	V	Q	L	F	V	W	G	S	N
	V	P	I	I	T	N	A	S	V	V	D	A	R	W	I	D	V	S	T	H	P	S	V	S	V	F	V	G	S	F	L	V	I	N	F	F	W	F	V	E	L	F	V	W	G	C	N
	F	I	L	L	T	N	H	S	V	V	D	S	R	W	I	D	V	C	T	N	L	S	I	S	T	Y	A	S	T	F	L	P	L	E	F	F	W	F	L	A	K	F	L	W	G	N	N
	G	V	L	V	T	N	N	A	V	V	D	S	R	W	T	D	V	C	S	N	L	S	A	A	A	F	V	A	S	F	S	A	I	N	F	F	W	F	L	P	F	L	T	W	G	N	N
	E	I	L	I	T	N	T	V	V	V	D	A	H	F	I	D	V	C	T	T	L	T	I	S	V	Y	T	S	T	F	M	P	T	S	F	F	W	F	I	L	E	F	L	W	G	N	N
	H	Y	L	L	I	N	L	V	V	V	D	A	R	F	V	D	V	C	T	N	L	S	I	S	V	F	V	S	S	F	F	I	I	H	F	F	W	F	V	P	Y	F	T	W	G	N	N
	P	Y	Y	L	I	N	L	V	V	V	D	A	R	F	V	D	V	C	T	N	L	S	V	A	V	F	V	S	S	F	F	T	V	H	F	F	W	F	V	P	Y	T	T	W	G	N	N
	Q	A	G	L	V	N	S	F	V	L	D	A	P	L	M	D	V	C	T	N	L	A	V	S	T	Y	V	S	S	F	C	A	I	H	F	F	W	F	V	P	V	V	T	W	G	N	N
	W	L	M	L	T	N	Y	Y	C	V	D	S	R	W	L	D	Y	C	T	N	I	Y	L	A	V	F	L	A	S	F	F	A	I	M	T	Y	W	F	I	D	Y	S	F	W	L	N	N
	R	V	L	F	A	N	L	S	V	L	D	A	R	W	A	D	V	C	T	S	L	S	I	S	V	Y	A	S	S	F	L	S	P	L	F	F	W	F	V	D	F	V	F	W	G	N	N
W	M	M	I	I	N	H	G	V	M	D	A	N	W	T	D	V	V	T	T	I	S	T	S	V	Y	A	S	S	F	L	V	I	N	F	F	W	F	I	K	Y	L	N	W	G	N	N	

- Matrix is potentially very large
- Reduce matrix size
 - exclude conserved residues
 - exclude loops
 - select positions considered to be most likely to be involved in binding, 35 in this research

Property Data Required

Amino acid	E Vol	E Pot	Hyd
E	54.03	-126.77	-3.5
D	39.75	-124.72	-3.5
M	74.97	-11.83	1.9
N	41.56	-7.28	-3.5
Q	69.06	-5.64	-3.5
C	40.54	-5.62	2.5
F	101.69	-3.77	2.8
H	69.79	-0.41	-3.2
T	35.18	-0.06	-0.7
S	28.96	1	-0.8
P	38.12	5.92	-1.6
A	14.04	7.3	1.8
V	34	7.47	4.2
I	58.46	8.2	4.5
L	58.48	8.2	3.8
G	5.76	8.78	-0.4
Y	103.04	24.52	-1.3
W	157.46	29.72	-0.9
R	133.43	117.97	-4.5
K	60.92	136.59	-3.9

- 3 properties to describe each amino acid
 - ellipsoidal volume
 - electrostatic potential
 - hydrophathy

Affinity Data Required

	Ligand 1	Ligand 2	Ligand 3
Receptor 1	6.48	*	8.54
Receptor 2	8.01	7.71	8.26
Receptor 3	6.49	7.05	7.43
Receptor 4	5.2	5.19	5.09
Receptor 5	5.61	6.37	7.16
Receptor 6	8.15	6.4	5.91
Receptor 7	7.83	8.5	6.9
Receptor 8	6.63	6.42	5.53
Receptor 9	7.15	6.51	5.22
Receptor 10	9.08	6.13	5.89
Receptor 11	7.9	6.57	6.14
Receptor 12	*	7.02	6.8
Receptor 13	*	*	6.85
Receptor 14	7.26	*	5.64
Receptor 15	5.56	*	5.78

- Affinity data is the observed variable
- * denotes data not available, sparse database
- In this research full-rank matrices were used

Matrix Construction

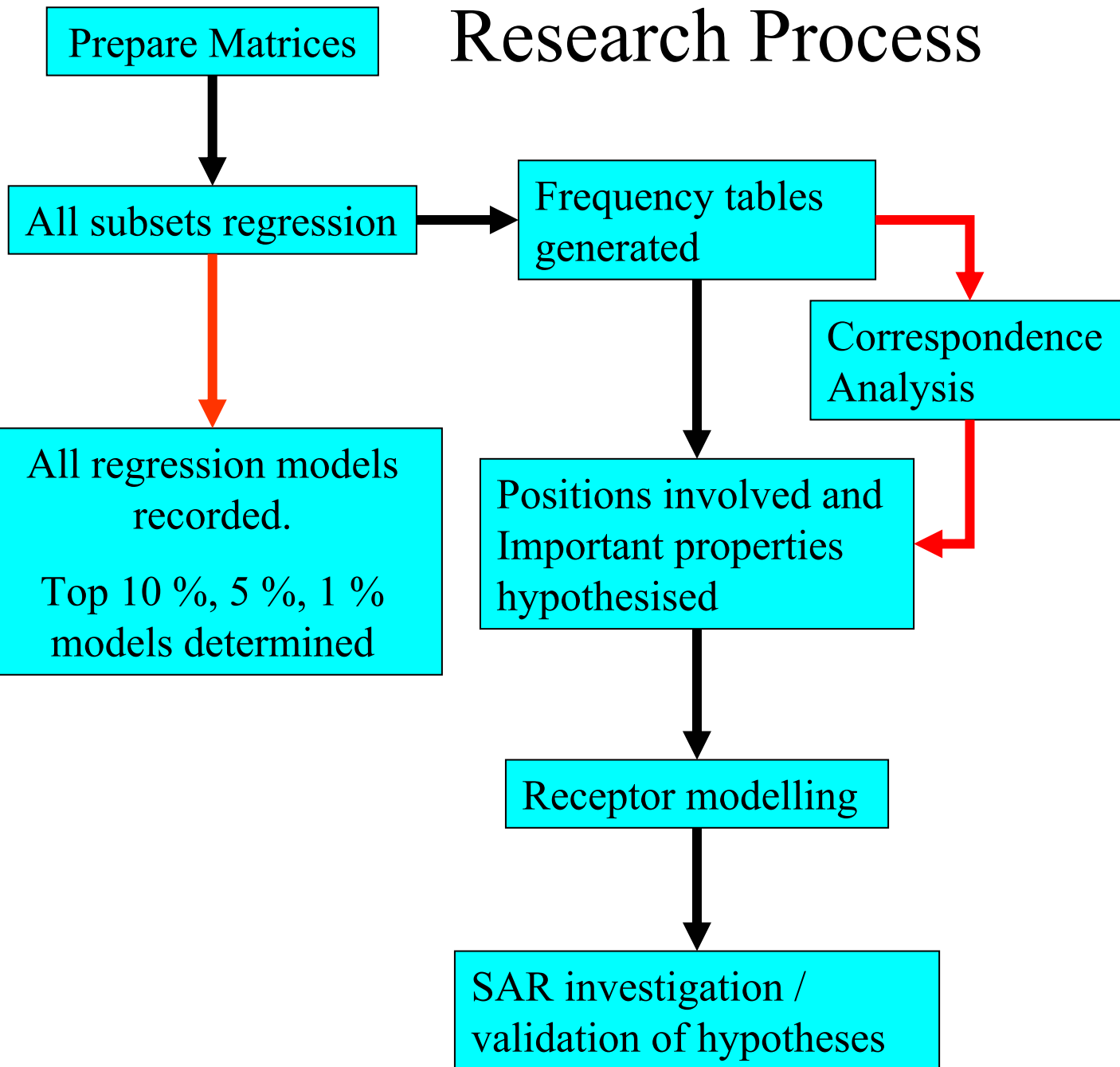
Ligand affinity data

Sequence – Property Data Matrix

105 columns x 15 rows

(35 positions x 3 properties)

Research Process



Statistical Methods: SSR

- All subset regression (SSR)
- Exhaustive
- Information from a number of regression models
- Prediction from a consensus of all components hypothesised to be important

$$Y = b_0 + b_1X_1 + b_2X_2 \dots + b_iX_i$$

Number of models:

35 positions, 3 properties = 105 independent variables (n).

$${}_n C_r$$

where r = No. terms in regression model.

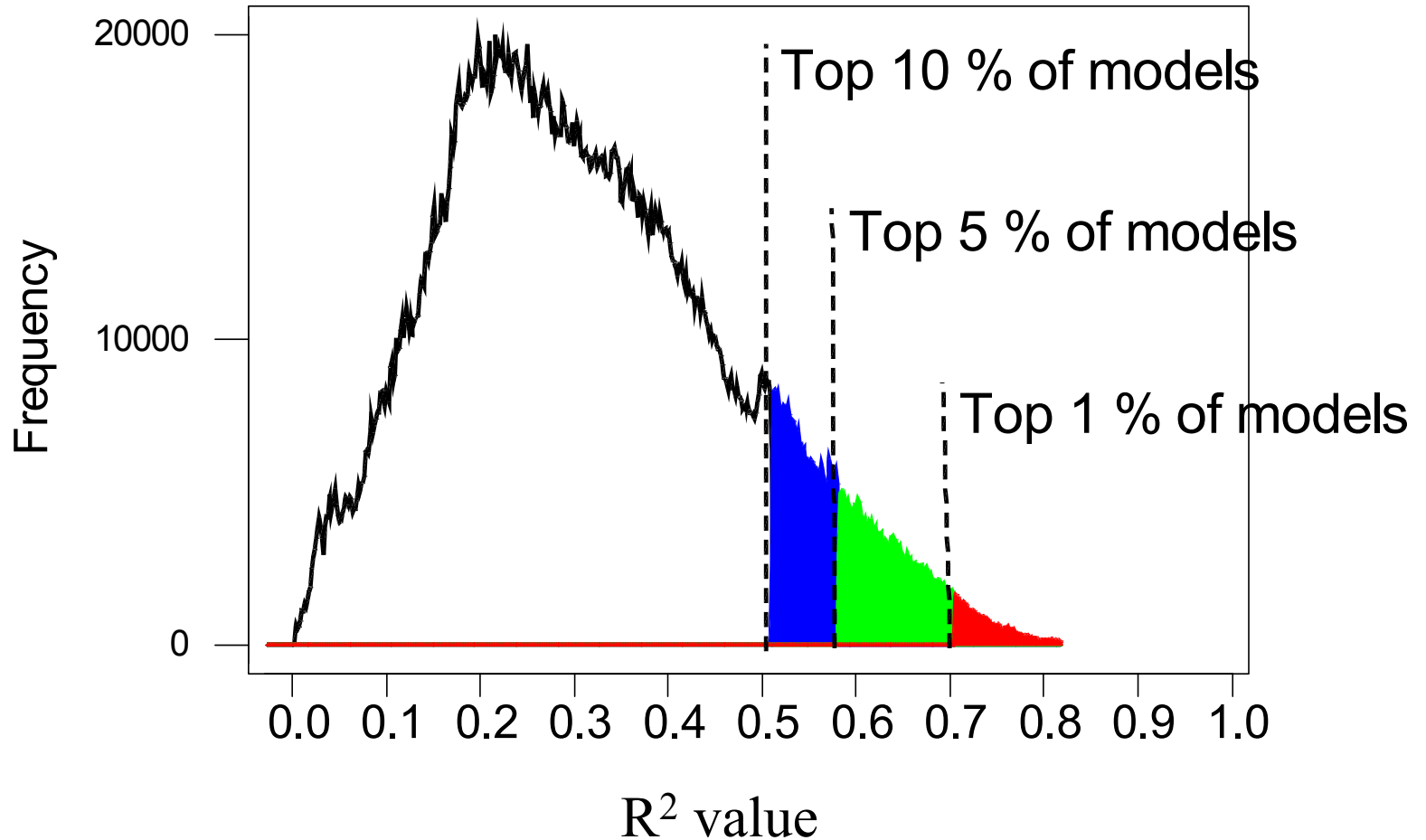
$${}_{105} C_2 = 5,460$$

$${}_{105} C_3 = 187,460$$

$${}_{105} C_4 = 4,780,230$$

Statistical Methods: SSR

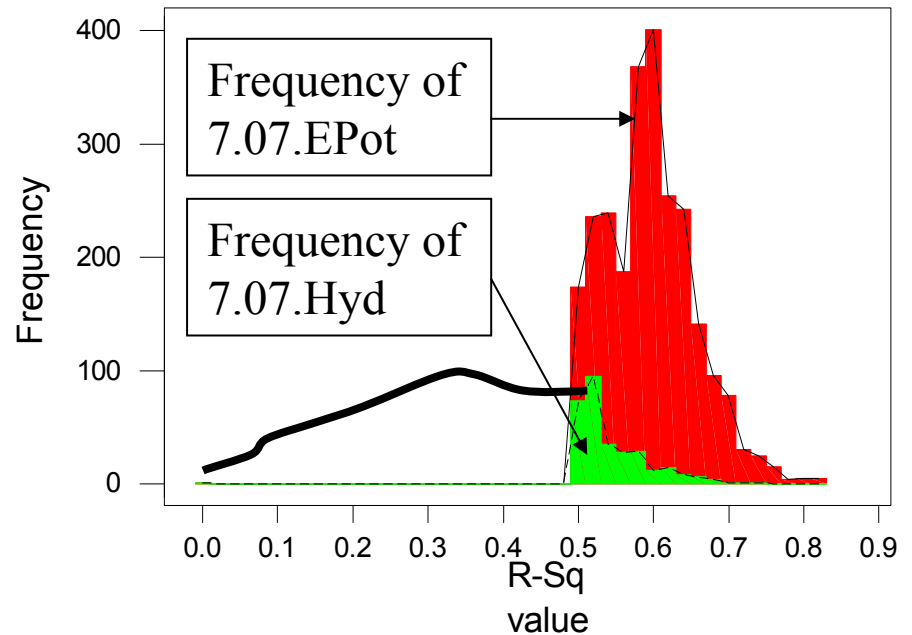
Distribution of 4-variable models



Statistical Methods: SSR

- The distribution of variables among the regression models is different for each variable
- Assumption:
 - The greater the frequency of a variable, the more likely it is to be involved in ligand binding
- 7.07.EPot is more likely to be involved in binding than 7.07.Hyd

Frequency of 7.07.EPot & 7.07.Hyd in Top 5% of Regression Models

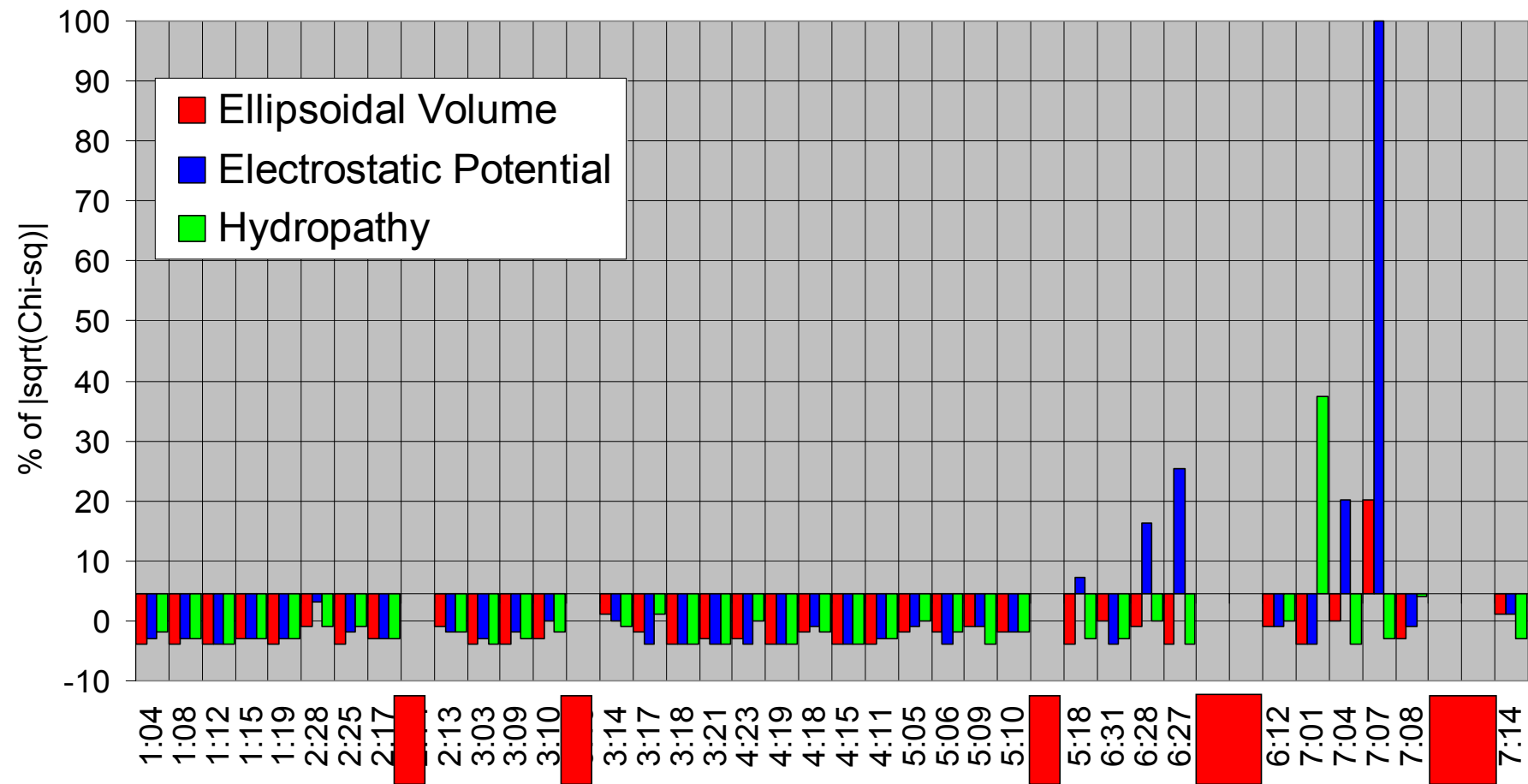


RESULTS: SB-656249

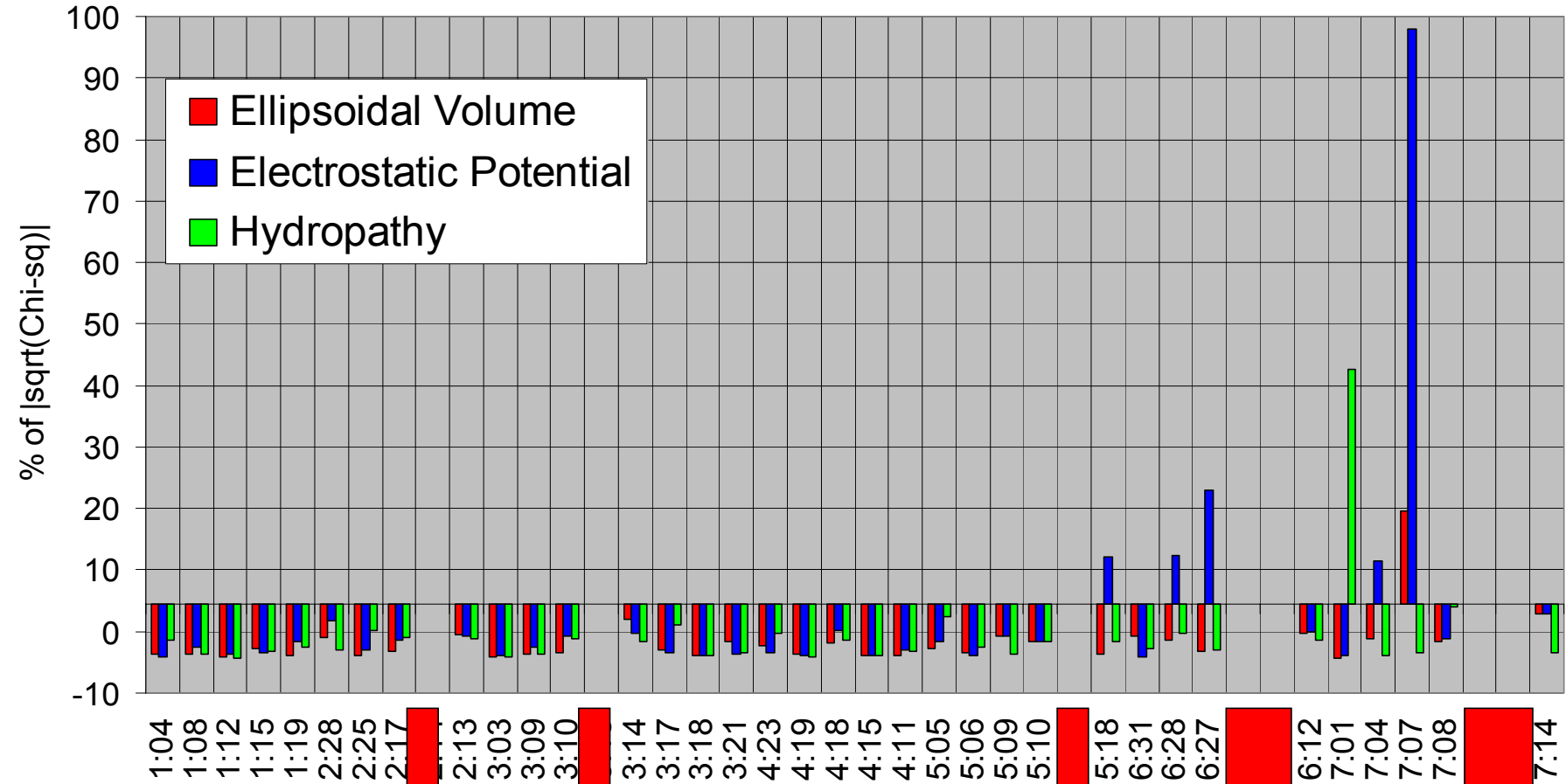
- Frequency graphs
- Positions on receptor
- Brief analysis of dominant variable

Ian T. Forbes, et al. SB-656104-A: A Novel 5-HT₇ Receptor Antagonist with Improved In Vivo Properties. *Bioorganic & Medicinal Chemistry Letters* 12 (2002) 3341-3344.

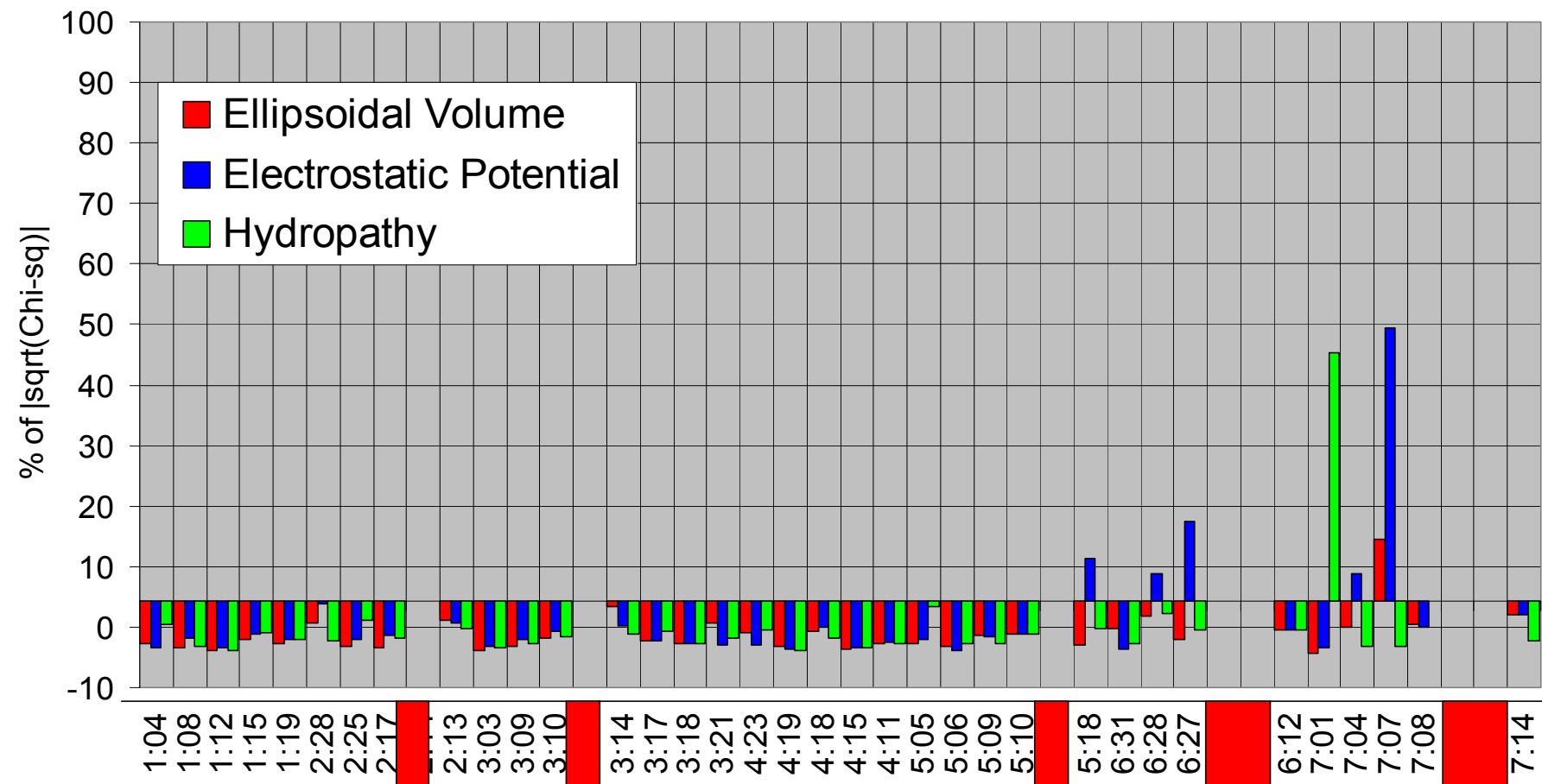
Top 5% of 2-variable models



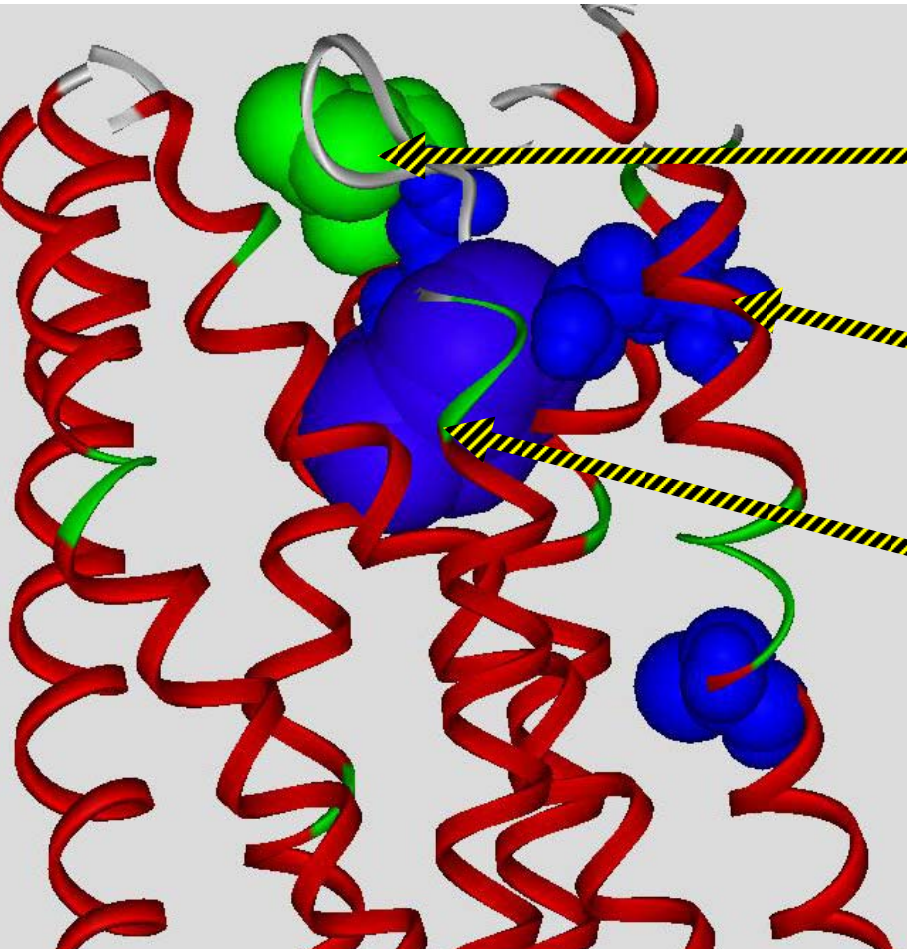
Top 5% of 3-variable models



Top 5% of 4-variable models



Positions Important for SB-656249

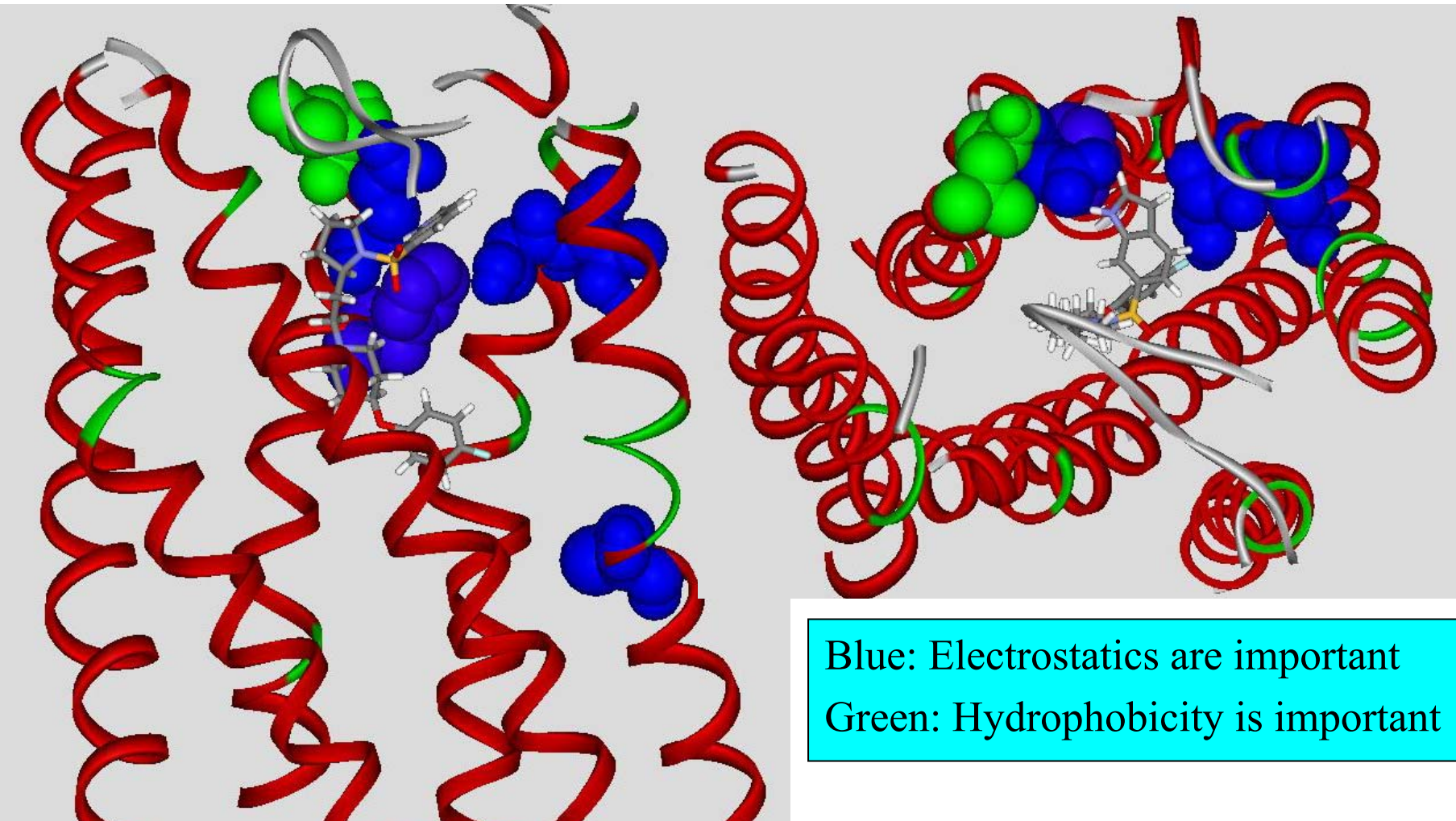


Green: Hydrophobicity important at this position

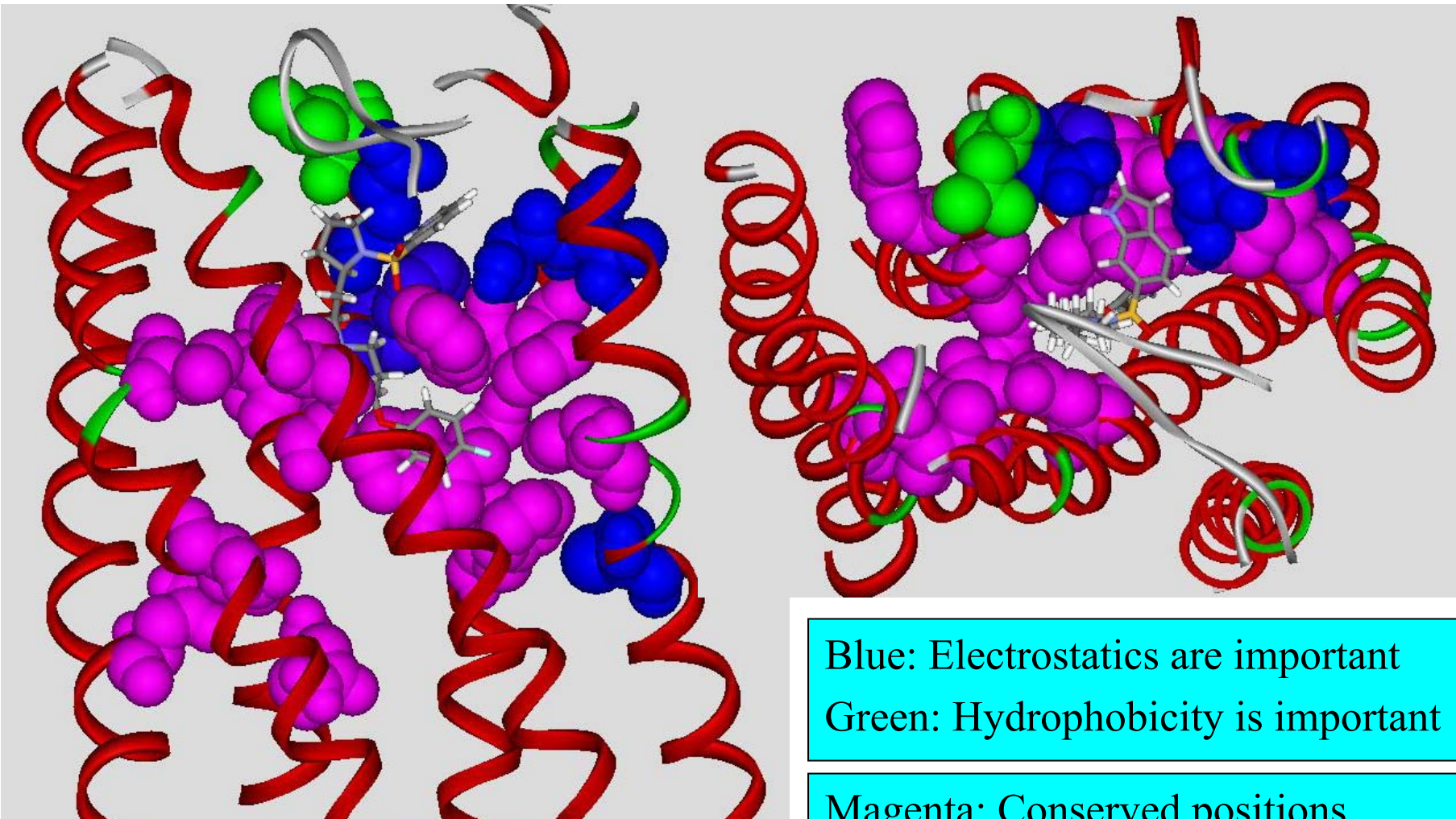
Blue: Electrostatics important at these positions

Electrostatics and volume important at this position

SB-656249 Docked into 5HT7



SB-656249 Docked into 5HT7

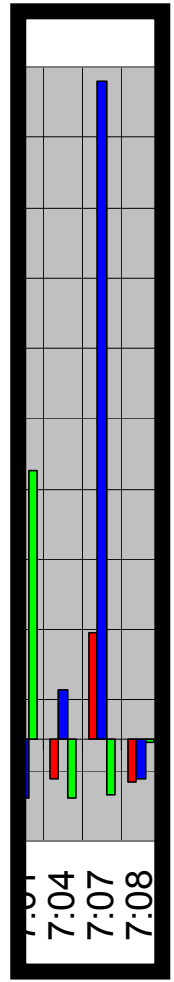


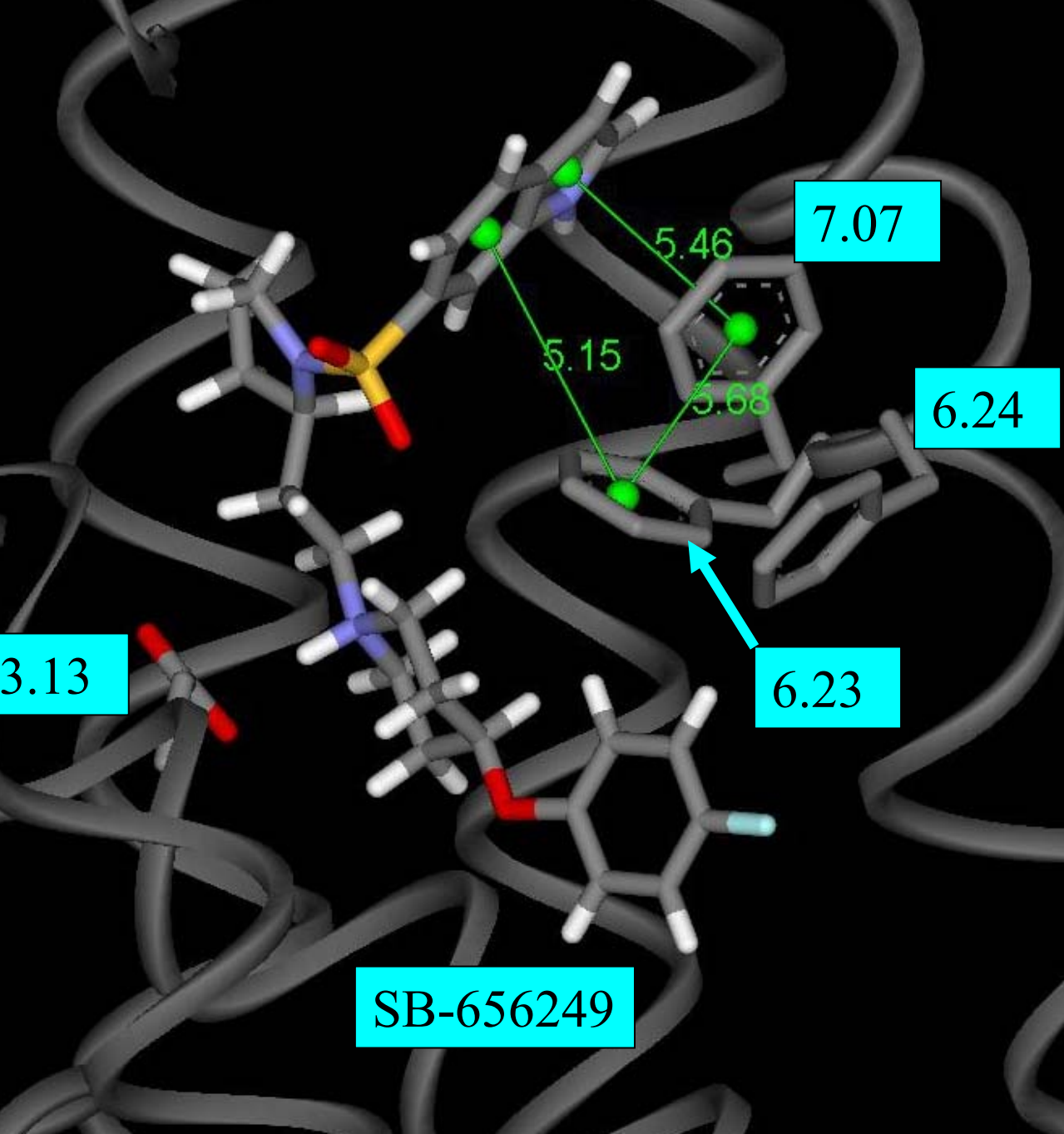
Blue: Electrostatics are important

Green: Hydrophobicity is important

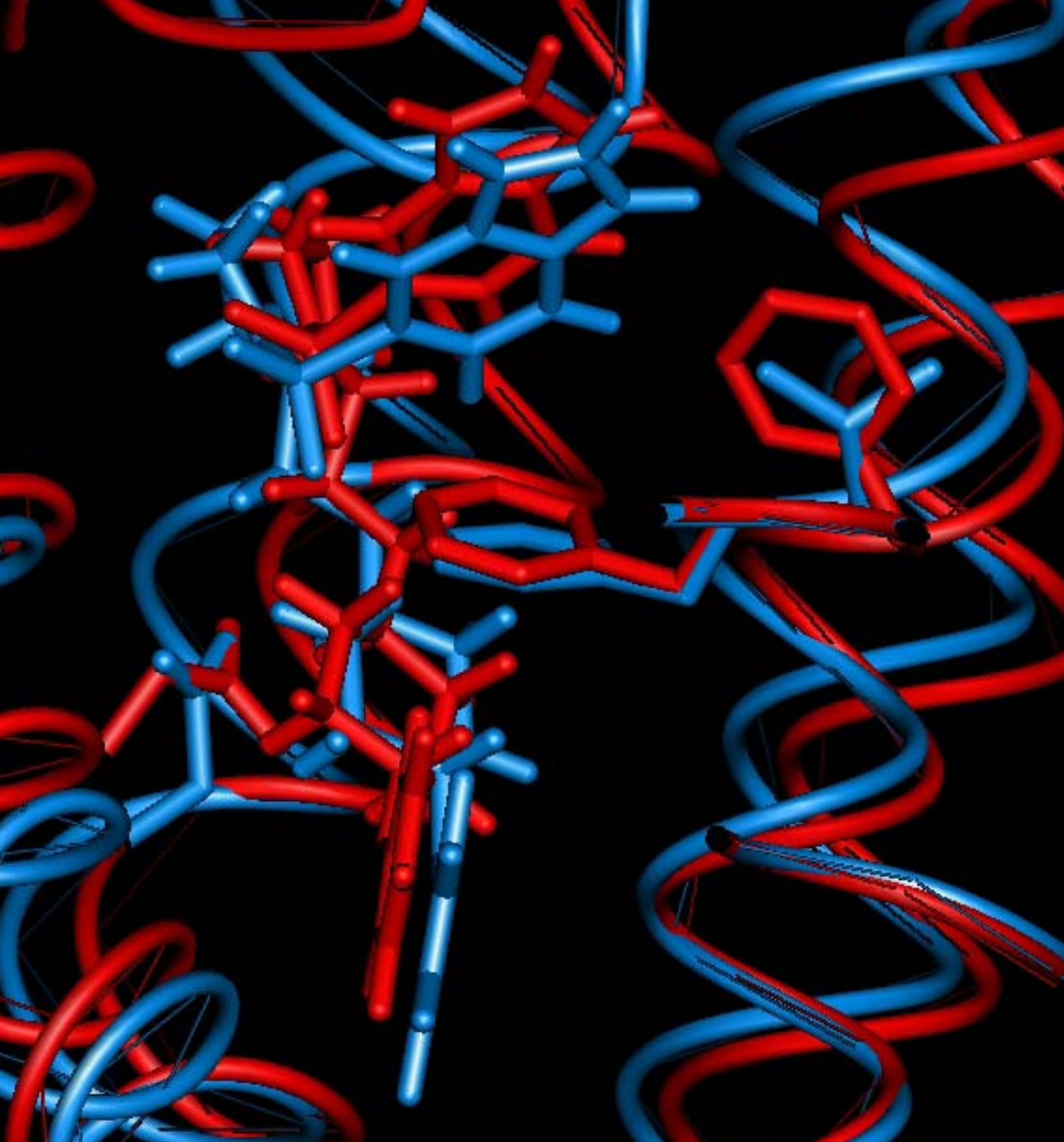
Magenta: Conserved positions

Top 5% of 3-variable models





- Edge - Face interaction network between indole moiety of ligand, 7.07 F and 6.23 F
- Supported by 6.24 F



Receptors with F
at 7.07

Receptors with V
at 7.07 (or I / L)

- Assuming the same binding mode for all receptors:
- Pi – Pi interaction network not present without 7.07 F
- Offers part of an explanation for greater ligand affinity observed at receptors with 7.07 F

Conclusion

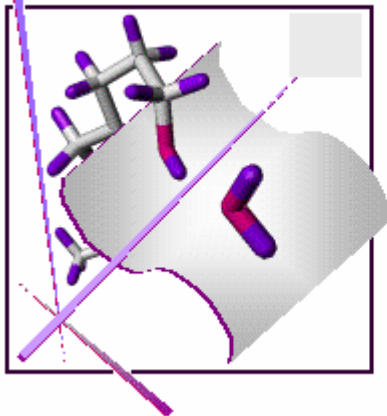
- SSR can provide interesting and perhaps unexpected hypotheses about which residue positions and the properties of those residues are important
- These can be tested, for example through ligand docking / compound design or SDM

Acknowledgements

- Martyn Ford
- Dave Salt
- Stephen Garland
- David Whitley
- Ben Tehan
- Frank Blaney
- BBSRC
- GSK

Contact details

- philip.evans@port.ac.uk
- david.salt@port.ac.uk
- martyn.ford@port.ac.uk



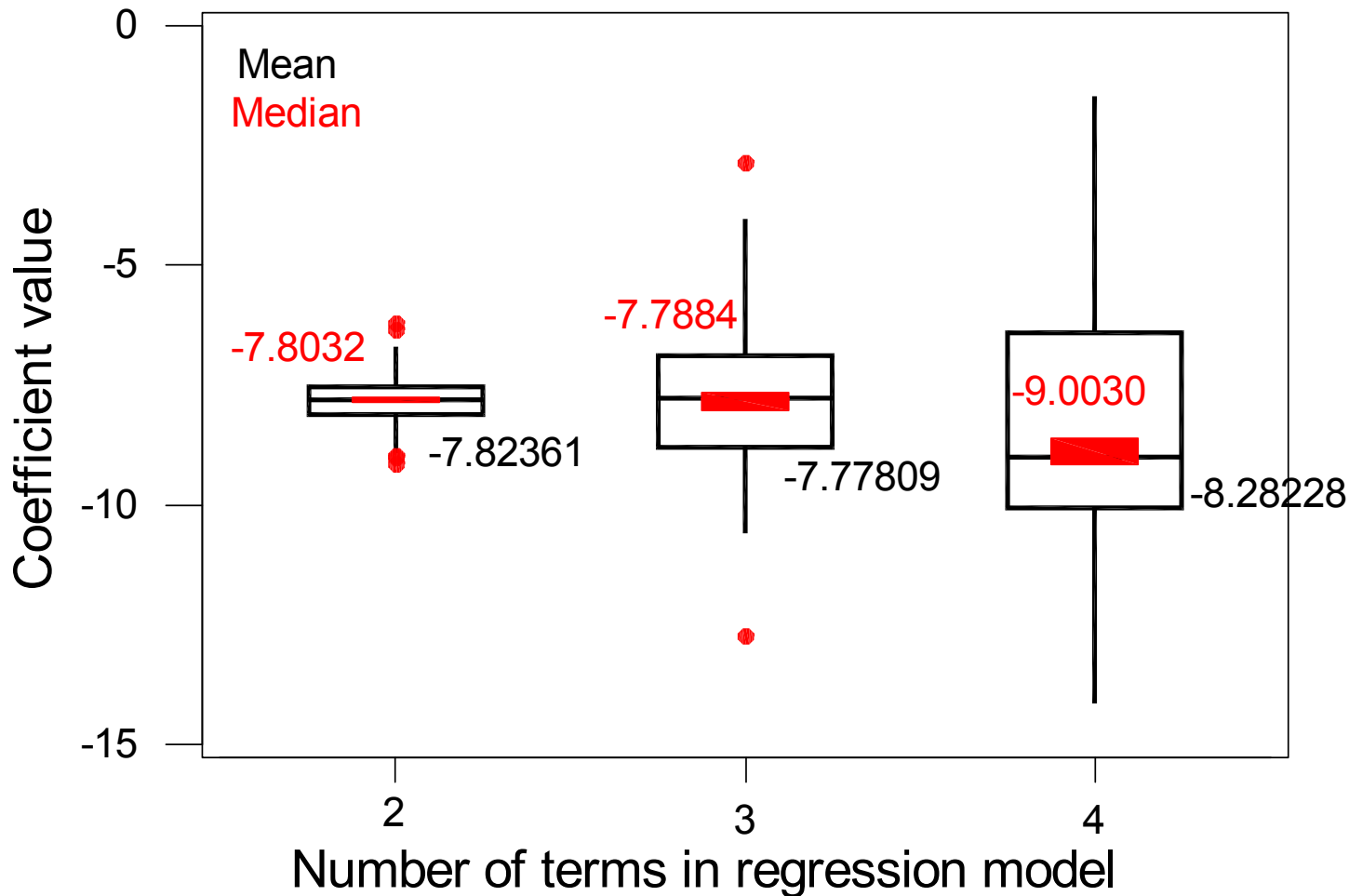
**UK QSAR
& Chemoinformatics**

Spring Meeting

14th April 2005

Coefficients of 7:07:Electrostatic

Distributions of Coefficients for variable 7:07:EPot

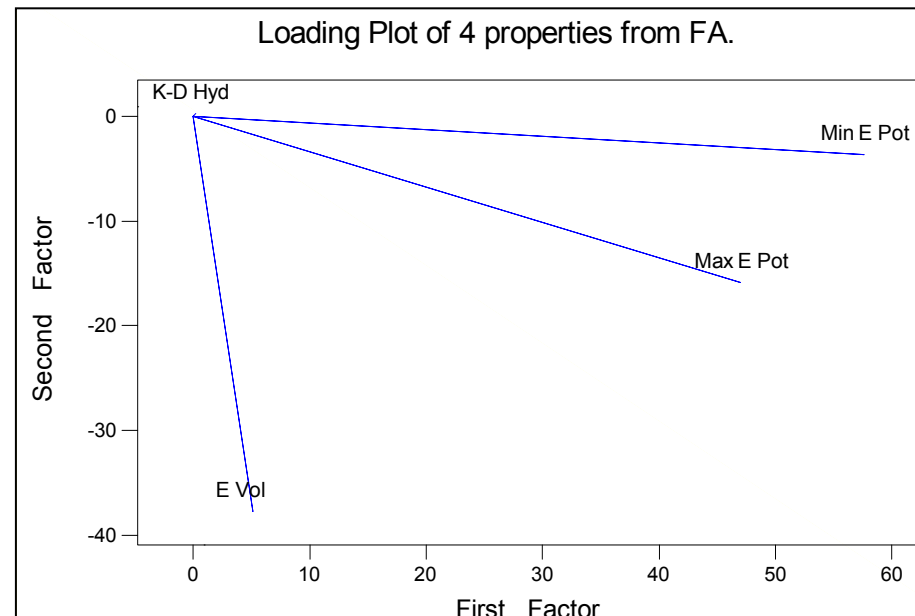


Electrostatic Property

- Maximum Electrostatic Potential (Max EPot) and Minimum Electrostatic Potential (Min EPot) are correlated

	E Vol	Max E Pot	Min E Pot
Max E Pot	0.423		
	0.063		
Min E Pot	0.161	0.661	
	0.498	0.001	
K-D Hyd	-0.177	-0.357	0.320
	0.456	0.123	0.169

Cell Contents: Pearson correlation
P-Value



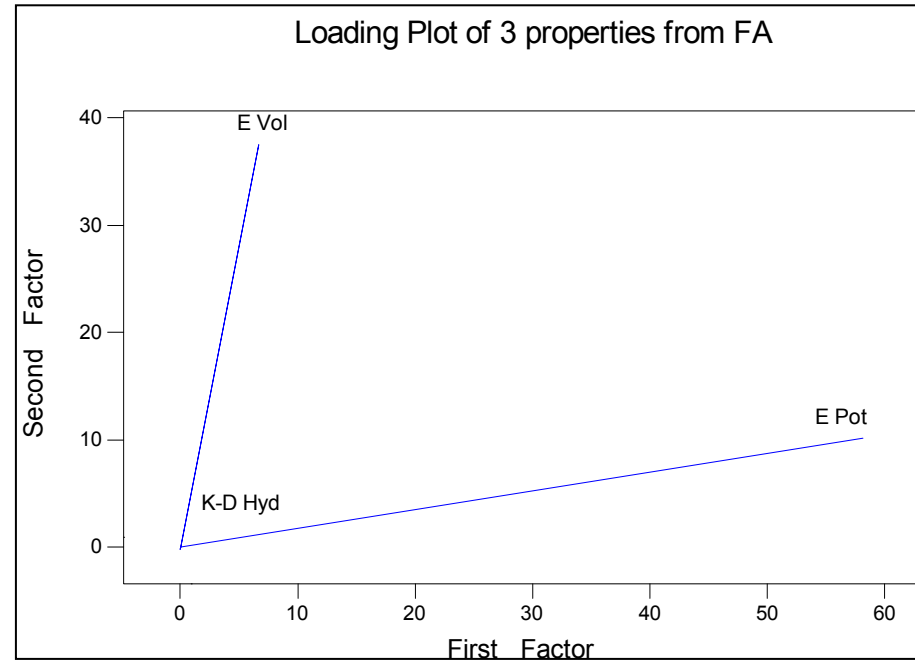
Electrostatic Property

- $\text{Abs}(\text{Max EPot}) - \text{Abs}(\text{Min EPot}) = \text{EPot}$
- The 3 properties are orthogonal

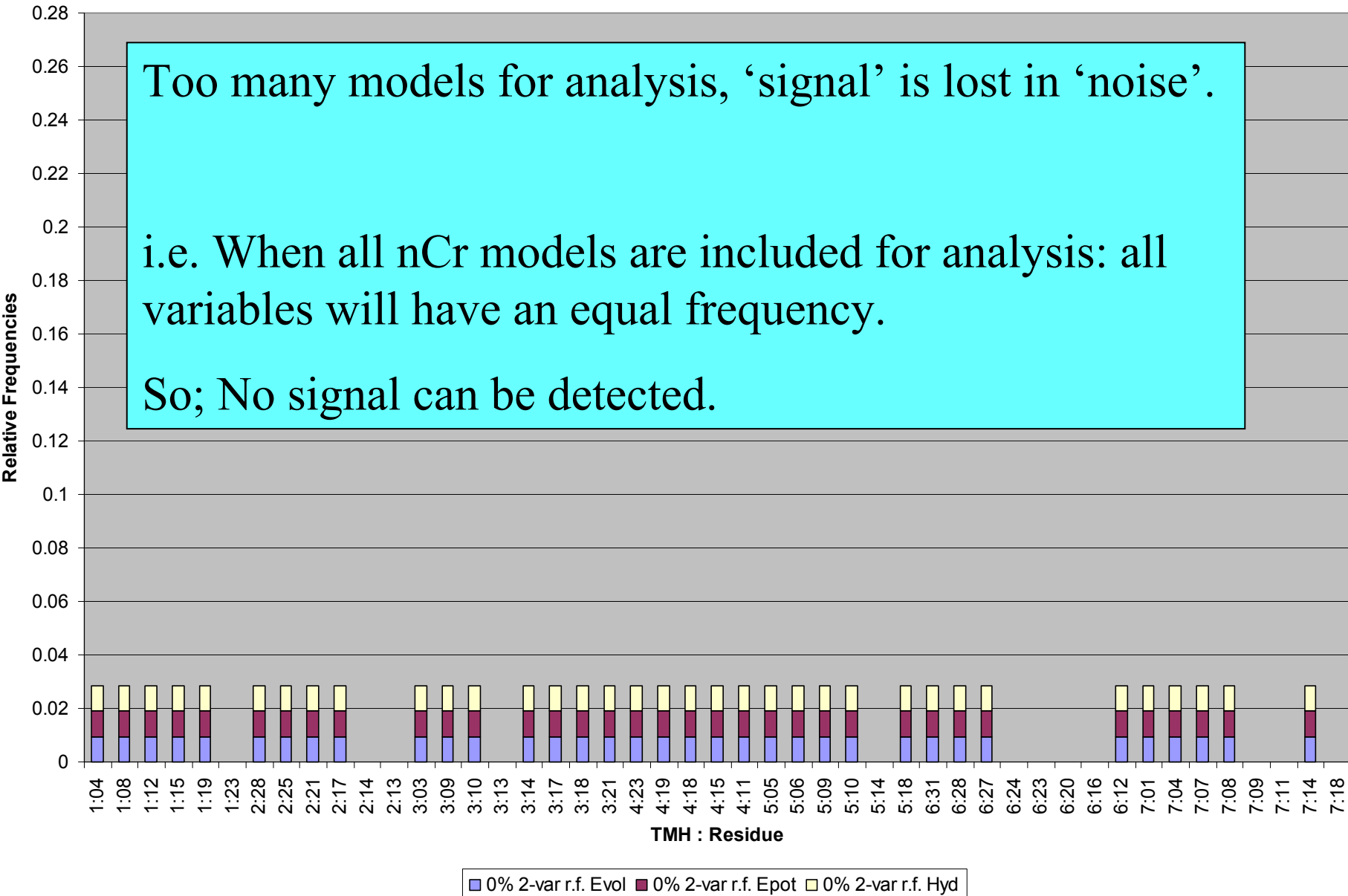
Correlations: Evol, EPot, Hyd

	Evol	EPot
EPot	0.341	0.141
Hyd	-0.177	-0.033
	0.456	0.891

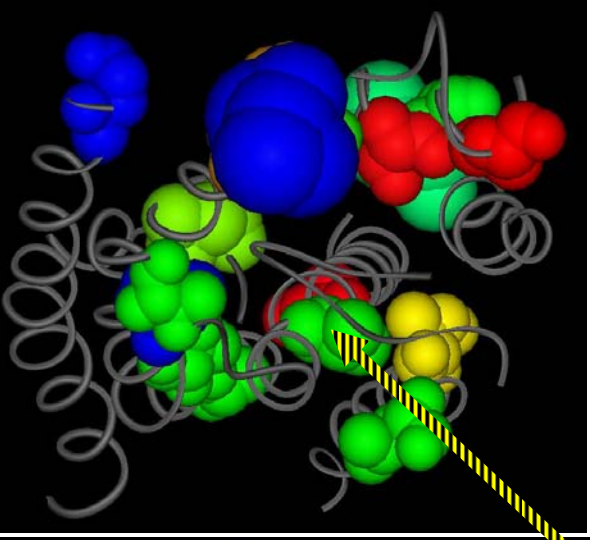
Cell Contents: Pearson correlation
P-Value



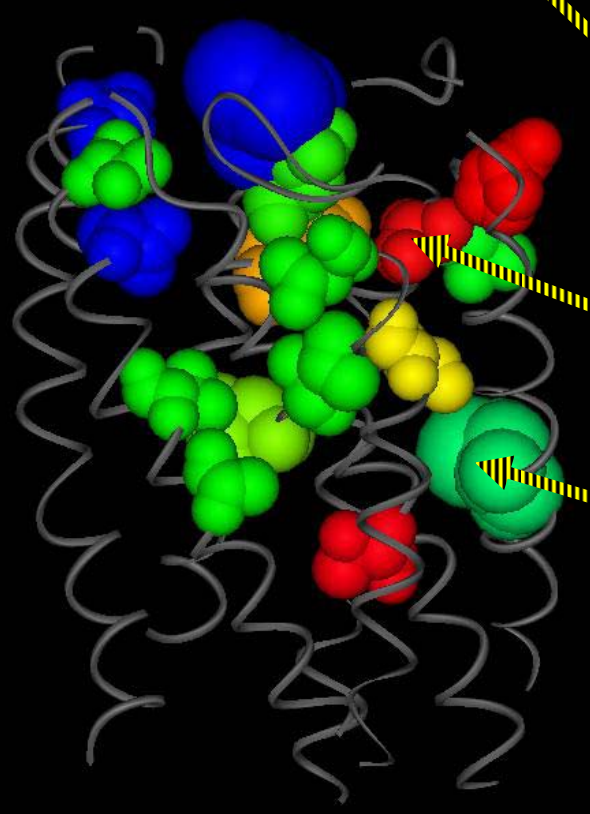
Example when all models analysed



Positions on the Receptor



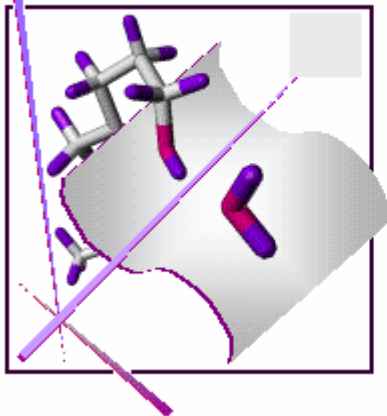
- Positions are sized to represent affect on regression equation Coloured according to property
 - Red: Volume
 - Green: Electrostatics
 - Blue: Hydrophobicity



Electrostatic is important at this position

Volume is important at this position

Both hydrophobicity and electrostatics are important for this position



UK QSAR & Chemoinformatics

Spring Meeting

14th April 2005