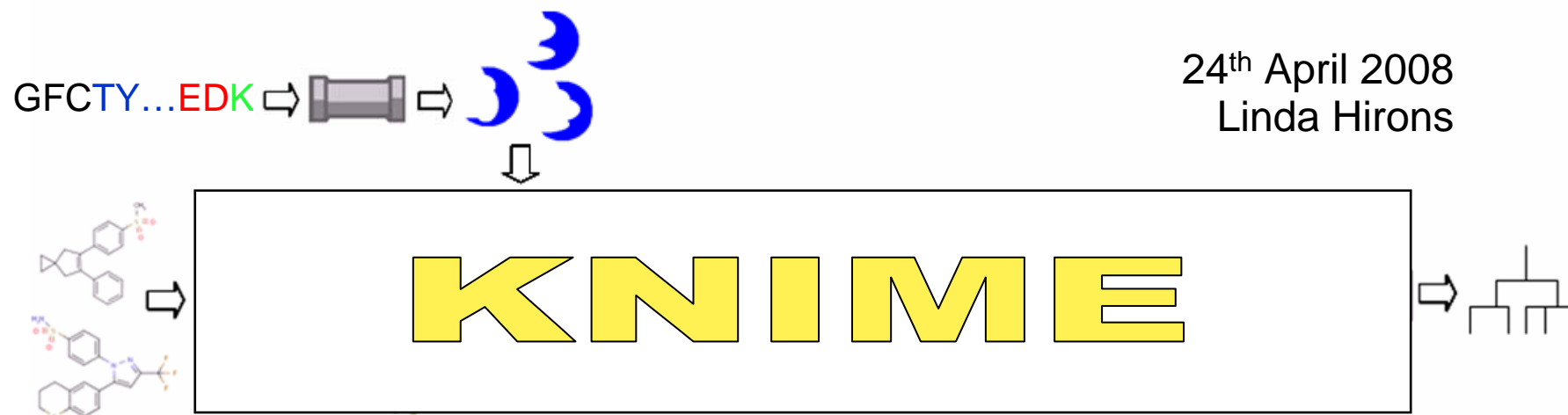

Inverse Structure Based Drug Design

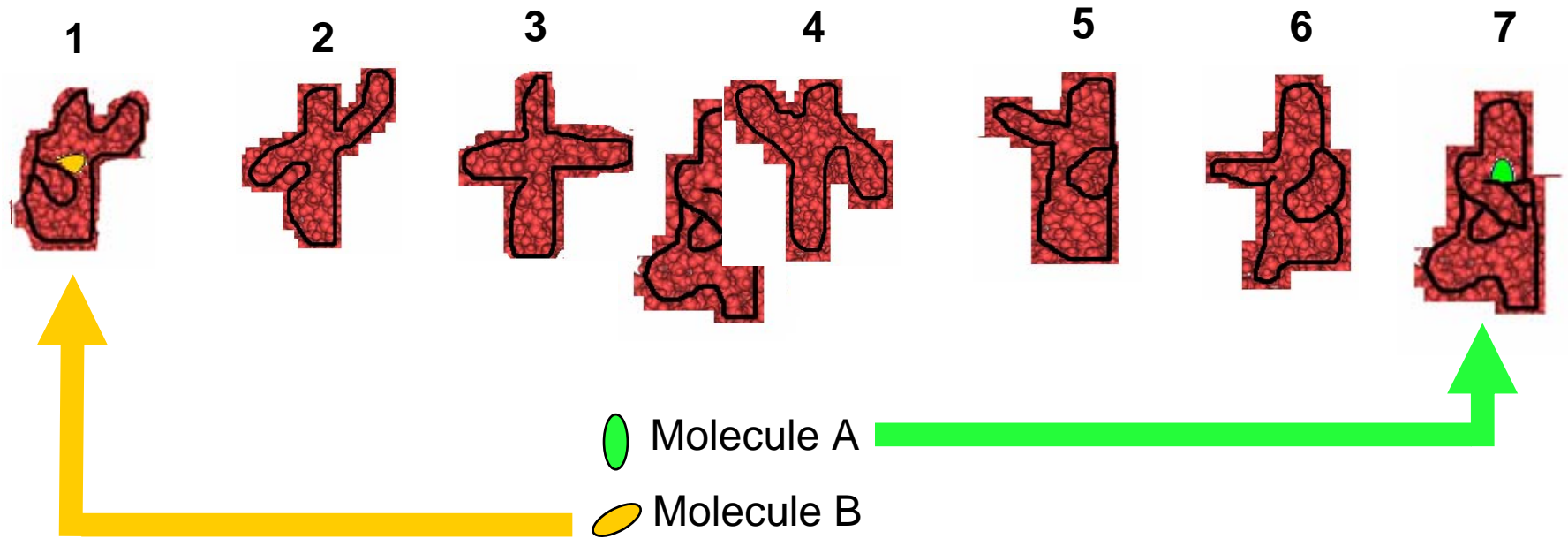


Lilly

Answers That Matter.

“Dynamic Personalities of Proteins”

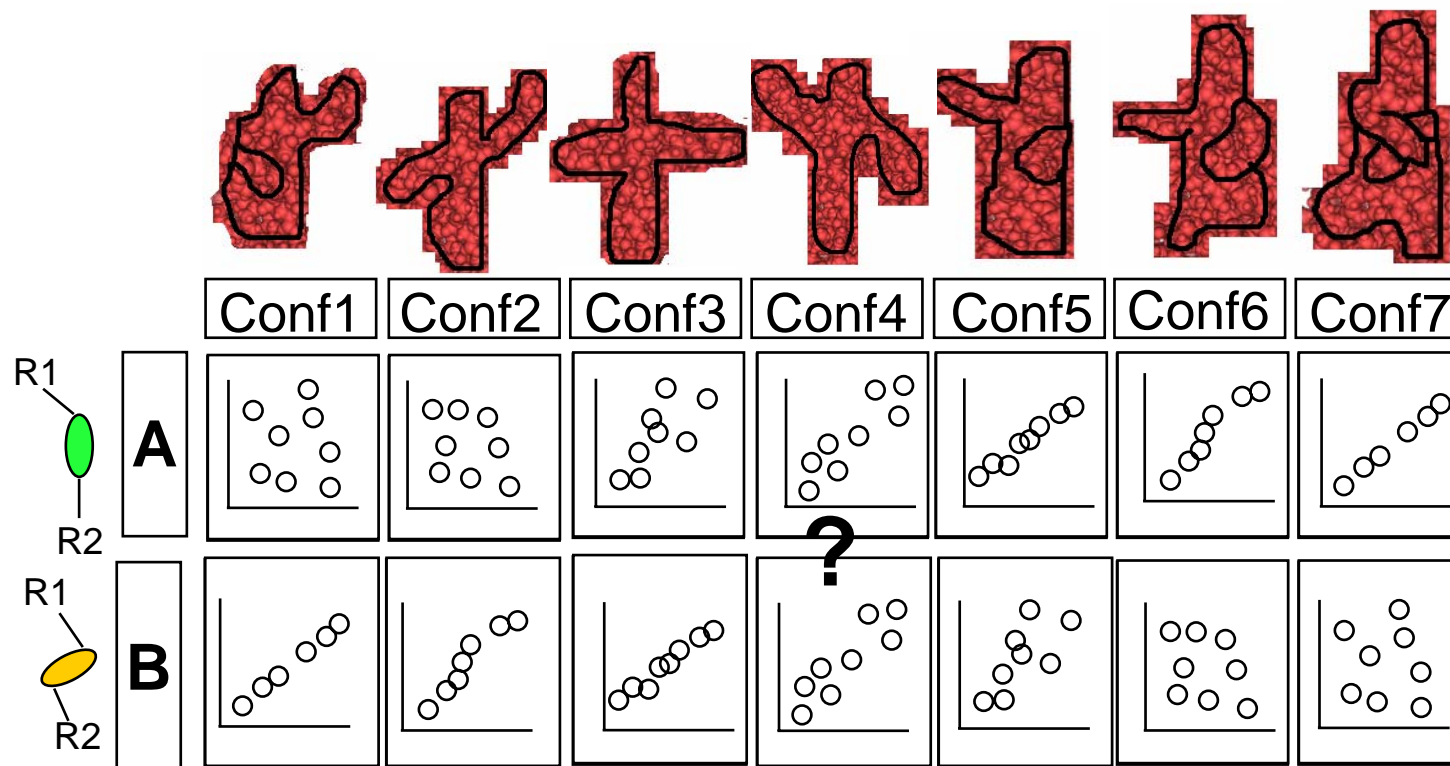
- “Enzyme engages in a dynamic dance even before its catalytic date shows up” *Caught in the act: The dynamic dance of enzymes Science Daily (Dec 13, 2007)*



Henzler-Wildman et Kern (2007) Nature, 450, 964-972

Objective

- Use SAR data to pull out representative protein model



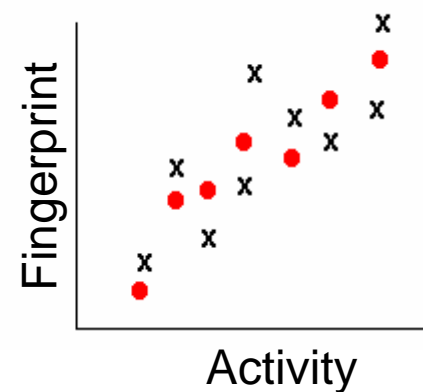
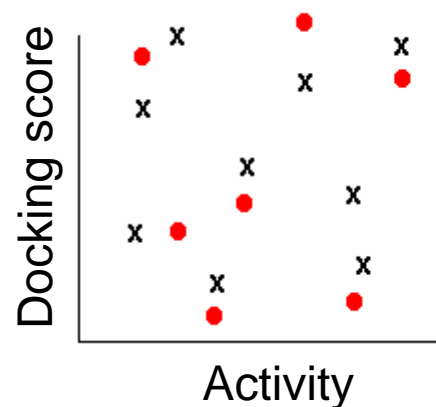
Docking Scores

Objective

- Docking successful at generating correct binding modes

Warren J. Med. Chem. **2006**, 49, 5912-5931

- PROBLEM:

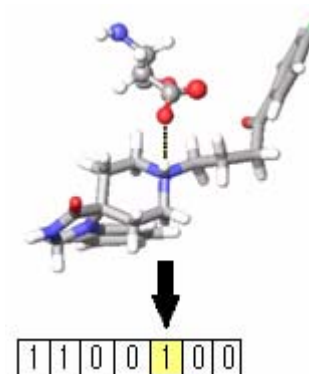


- x Top scoring pose
- Crystal structure

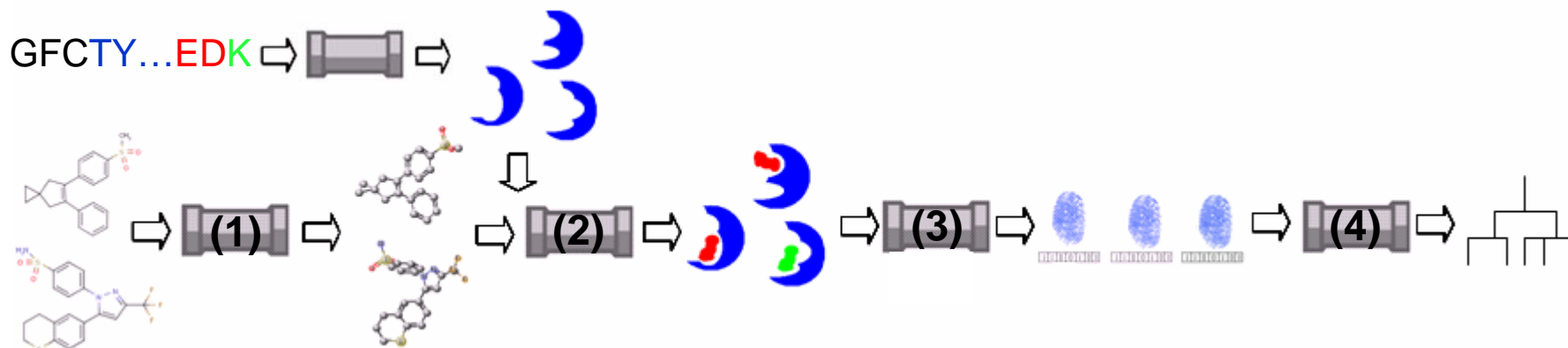
- SOLUTION:

Structural Interaction Fingerprints (SIFTs)

Singh et al., Chem. Biol. Drug Des. **2006**, 67, 5-



Protein Prediction Pipeline



1. 2D structure → 3D structure
2. Perform dockings
3. Calculate fingerprints
4. Build decision trees

- Molconvert www.chemaxon.com
- GOLD www.ccdc.cam.ac.uk
- SILVER www.ccdc.cam.ac.uk
- Weka nodes www.cs.waikato.ac.nz/ml/weka

Building the Pipeline into KNIME

File Reader



Protein conformers

File Reader



Ligands

MolConverter



GoldDock



Silver Descriptors



Fingerprint Expander



Numeric Binner



J48 (Weka)



Decision Tree Predictor



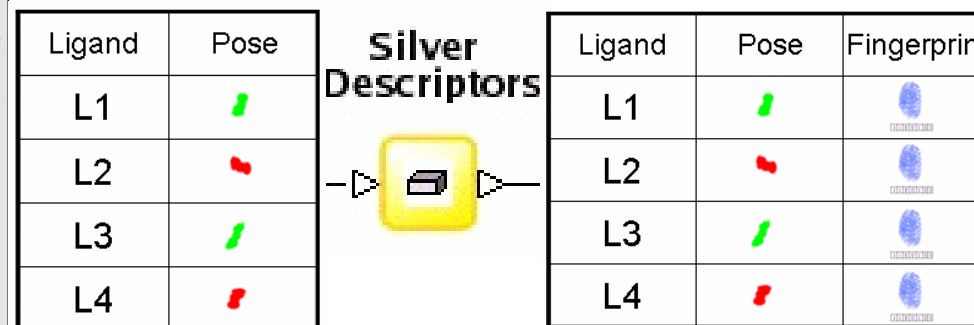
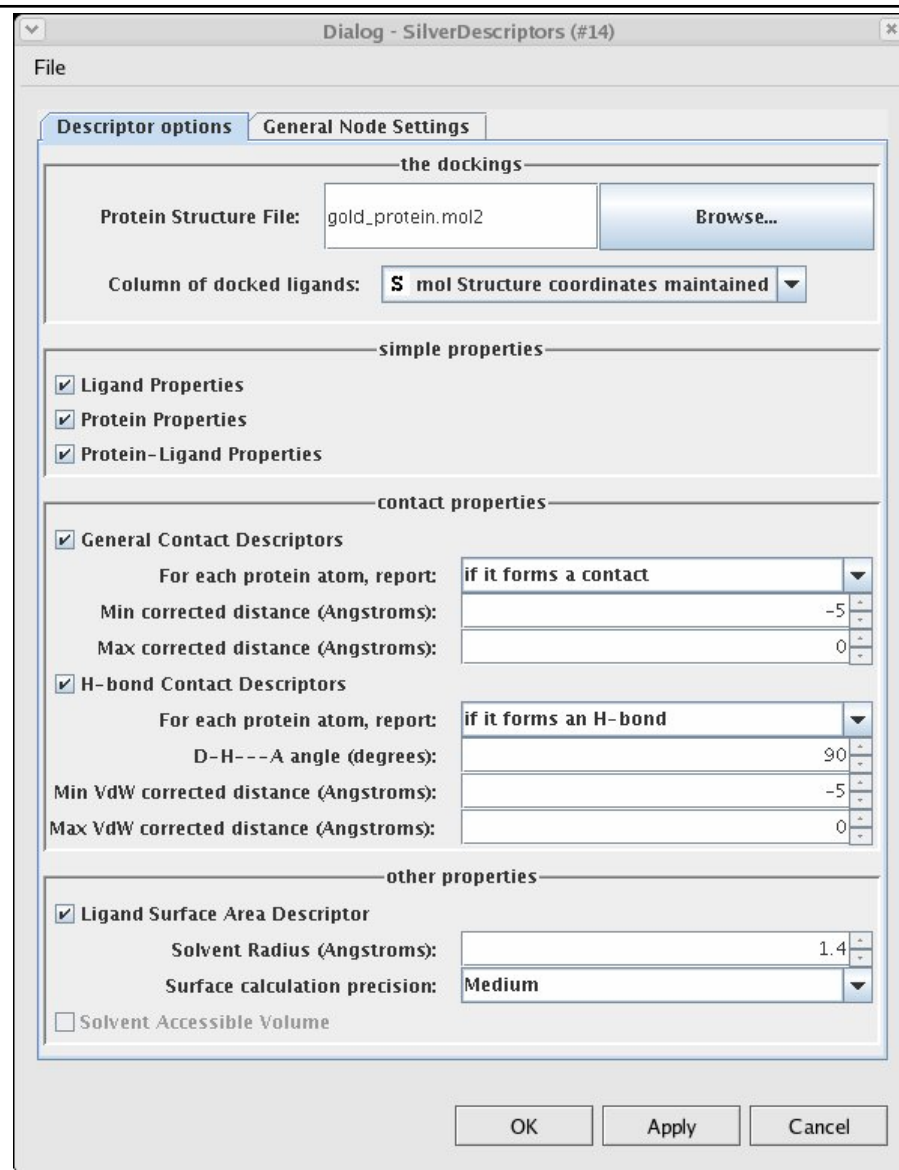
Validation



KNIME Node Development

Neighbourgrams, SVMs, NN, regression, bayesian...

Building the Pipeline into KNIME



GUI split into four panels on the Descriptor options tab:

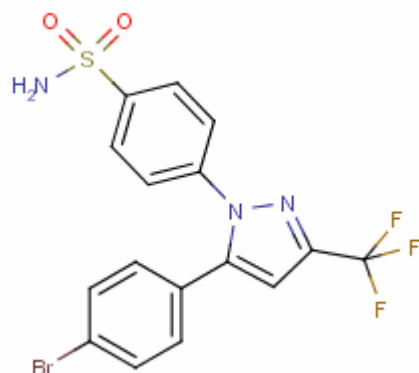
- The dockings
- Simple properties
- Contact properties
- Other properties

COX-2 Case Study

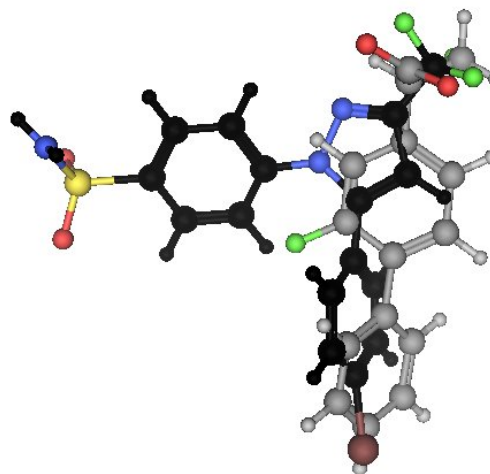
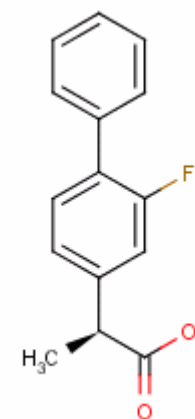
- 2 Crystal structures of COX-2 with bound inhibitors

Kurumbail et al. 1996 *Nature*, **384**, 644-648

1CX2: SC-558 bound



3PGH: Flurbiprofen bound



- Ligand dataset of 322 COX-2 Inhibitors with SC-558 scaffold

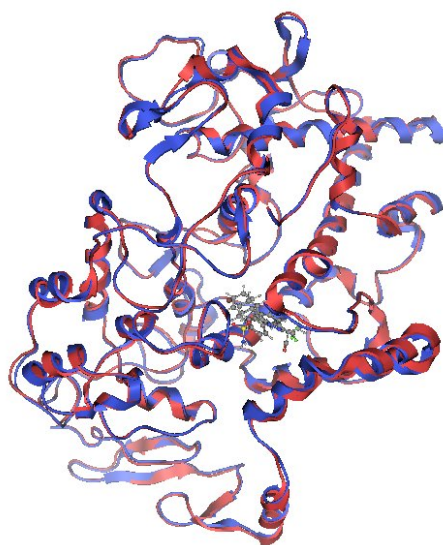
Sutherland et al. 2004 *J. Med. Chem.*, **47**, 5541-5554



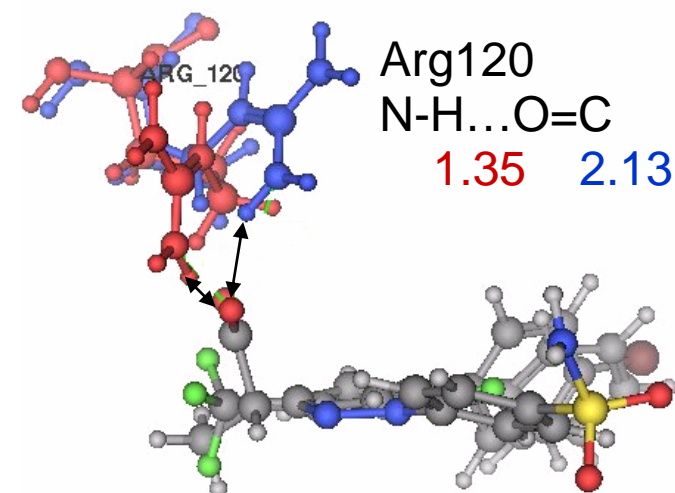
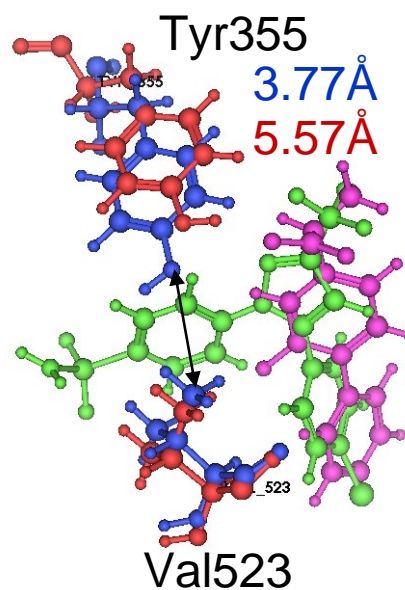
Protein Conformers: 1CX2 & 3PGH



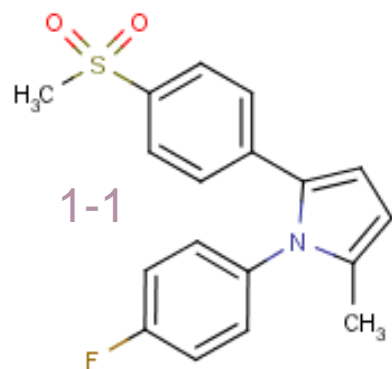
Answers That Matter.



$C\alpha$ RMSD = 0.456Å
All atom RMSD = 0.896Å



Docking of 1-1 to 1CX2 & 3PGH

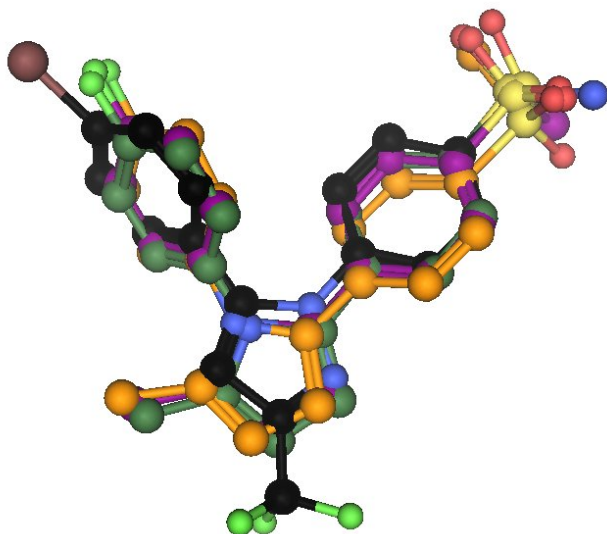


- Docking pose 1
- Docking pose 2
- Docking pose 3

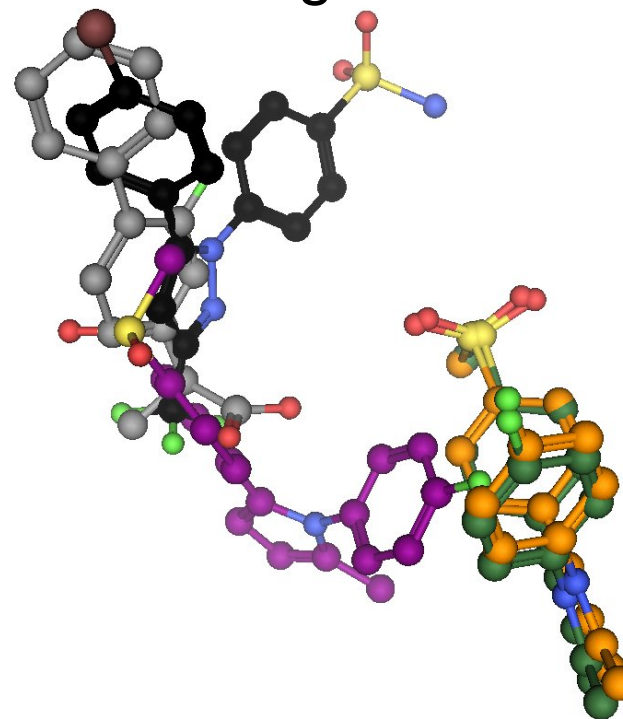
■ SC-558 docking to 1CX2

■ Flurbiprofen docking to 3PGH

1-1 Docking to 1CX2



1-1 Docking to 3PGH

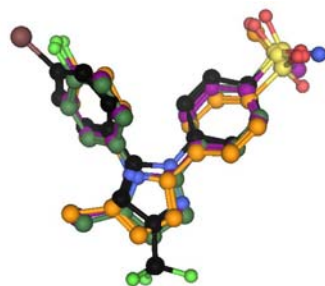
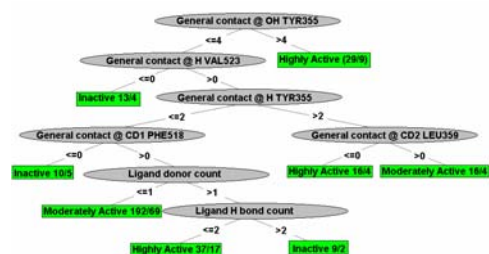


COX2 Decision Trees

S
I
N
G
L
E
L
I
G
A
N
D
S
E
R
I
E
S

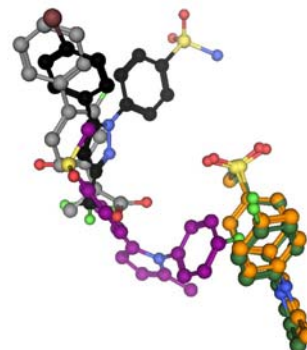
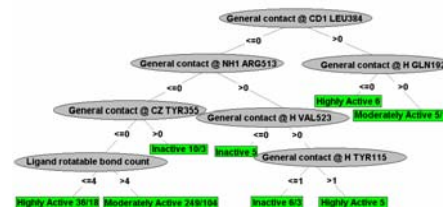
1CX2

Accuracy = 0.65
Recall = 0.65
Precision = 0.63
 $F_{\text{actives}} = 0.64$
Random $F_{\text{actives}} = 0.09$
100 x 5-fold CV = 0.52



3PGH

Accuracy = 0.60
Recall = 0.36
Precision = 0.62
 $F_{\text{actives}} = 0.46$
Random $F_{\text{actives}} = 0.13$
100 x 5-fold CV = 0.49



Accuracy:

Fraction of molecules correctly classified

Recall (r):

Fraction of actives that are correctly predicted to be active

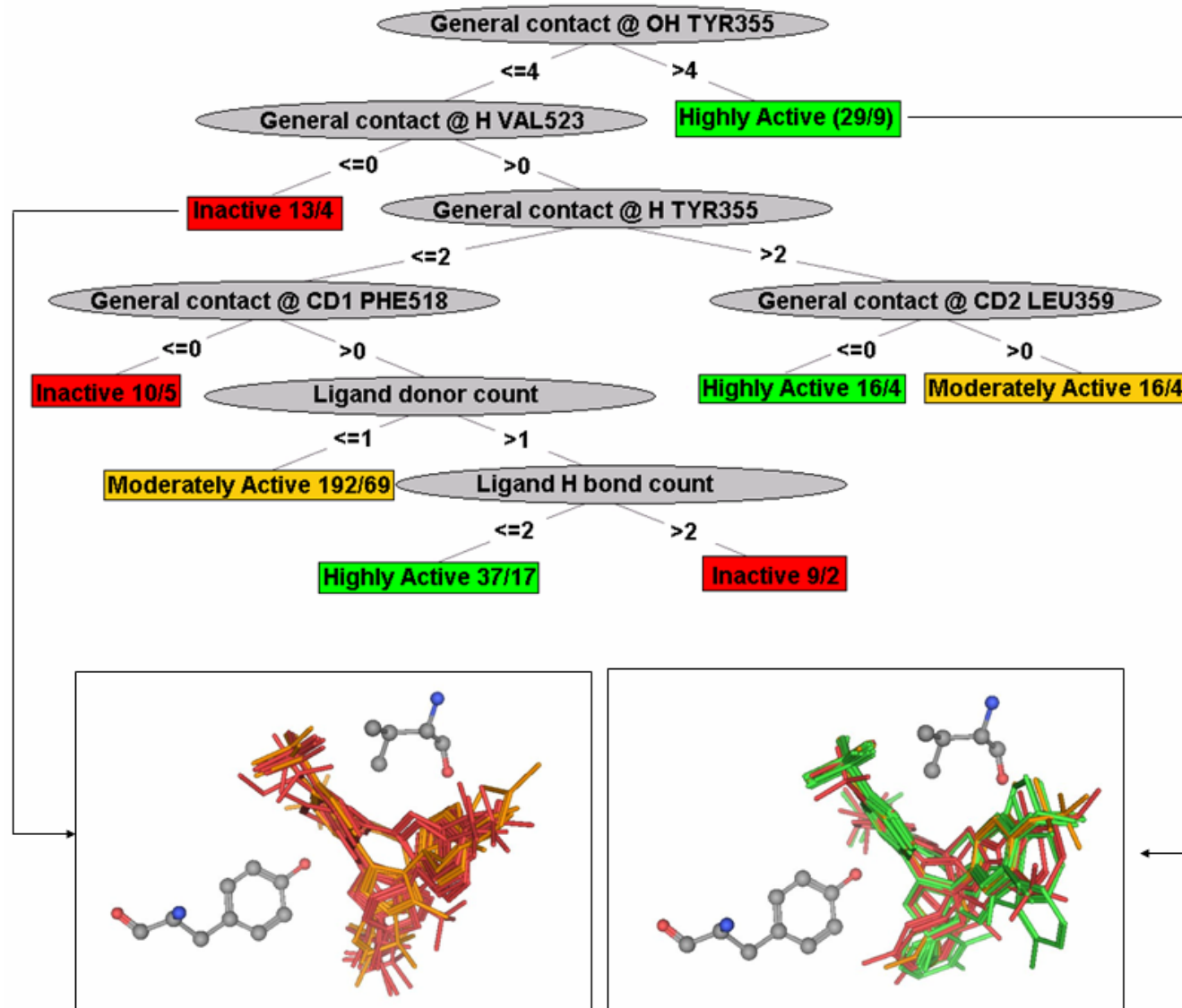
Precision (p):

Fraction of those molecules predicted to be active that actually are active

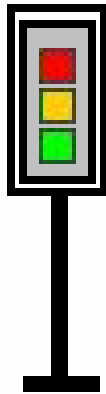
F-measure:

$$(2rp)/(r + p)$$

Decision Tree for COX-2 inhibition

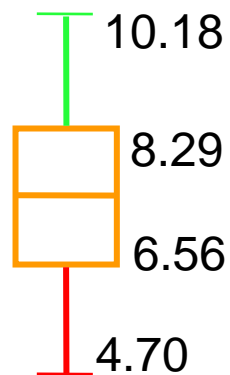


In House Dataset



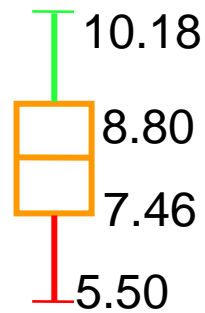
Inactive
Moderately active
Highly active

SERIES A + B
158 ligands

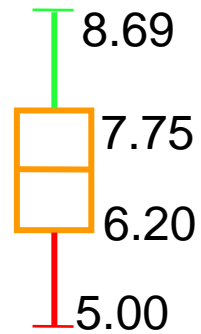


Units of activity = pIC50

SERIES A
80 ligands



SERIES B
78 ligands

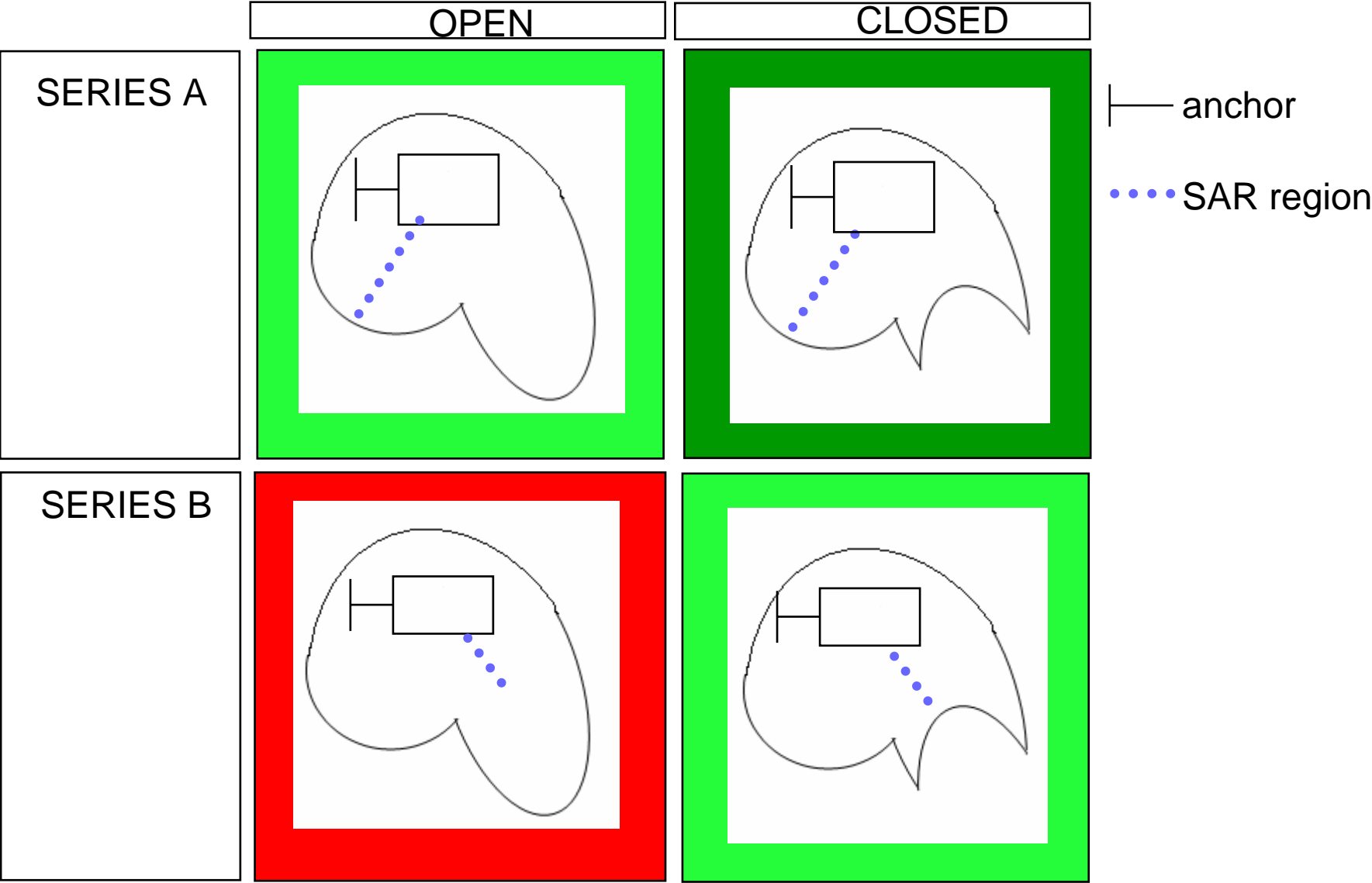


	OPEN	CLOSED
SERIES A 80 ligands		
SERIES B 78 ligands		

Decision Trees

	OPEN	CLOSED
SERIES A	<p>Accuracy = 0.81 Recall = 0.55 Precision = 0.85 $F_{\text{actives}} = 0.67$ Random $F_{\text{actives}} = 0.57$ 100 X 5-fold CV = 0.50</p>	<p>Accuracy = 0.88 Recall = 0.90 Precision = 0.82 $F_{\text{actives}} = 0.857$ Random $F_{\text{actives}} = 0.64$ 100 X 5-fold CV = 0.43</p>
SERIES B	<p>Accuracy = 0.65 Recall = 0.33 Precision = 0.86 $F_{\text{actives}} = 0.48$ Random $F_{\text{actives}} = 0.56$ 100 X 5-fold CV = 0.43</p>	<p>Accuracy = 0.85 Recall = 0.89 Precision = 0.89 $F_{\text{actives}} = 0.89$ Random $F_{\text{actives}} = 0.61$ 100 X 5-fold CV = 0.47</p>

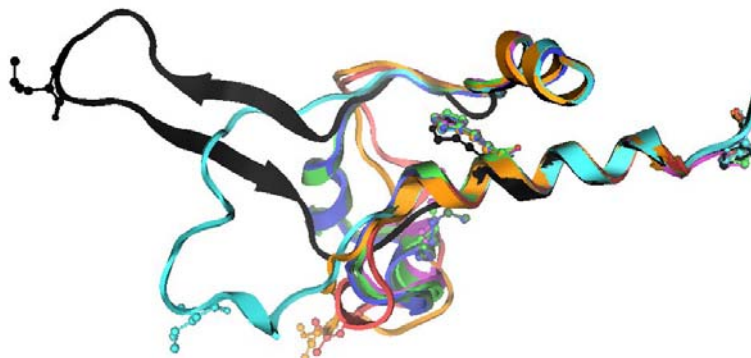
Decision Trees



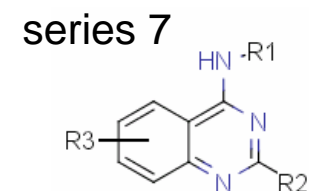
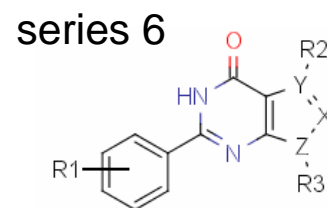
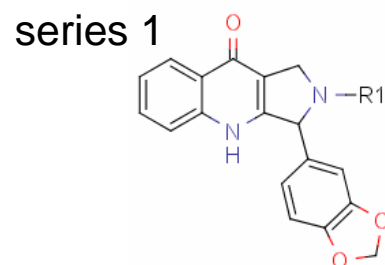
PDE5 Case Study

Exploring how variations in the H-loop effect SAR across different ligand series

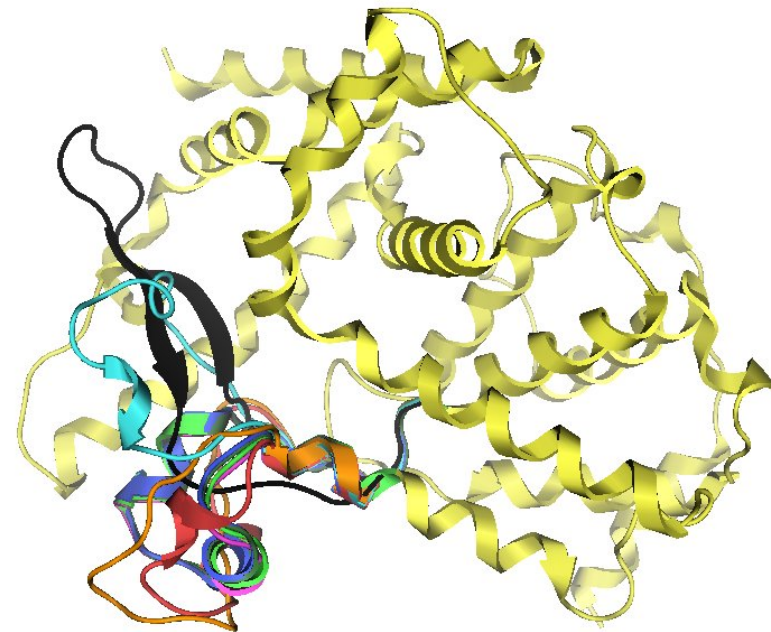
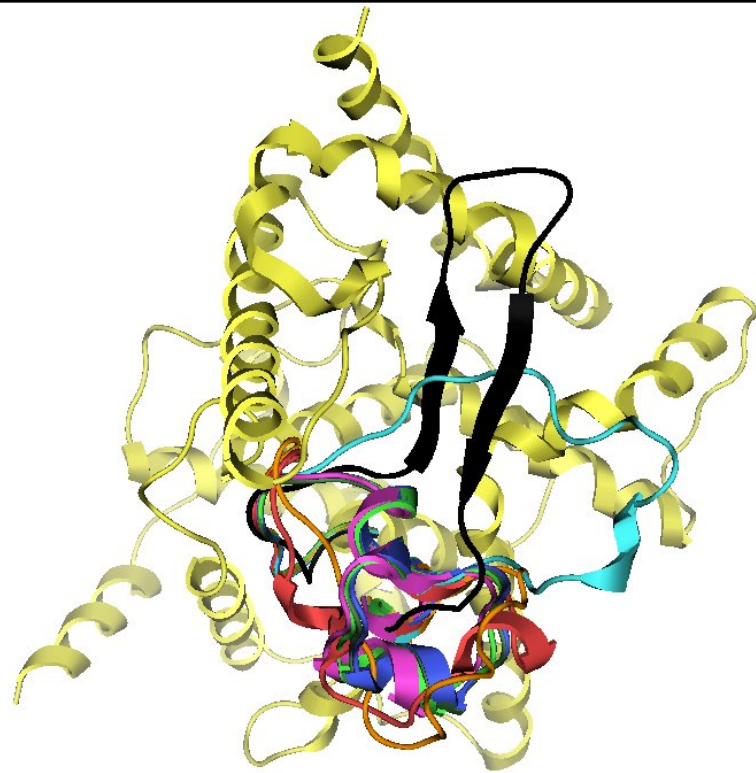
- 8 PDE5 H-loop conformations



- 3 ligand series



Eight PDE5 crystal structures from PDB



■ =2H40 no

■ ligand

■ =1RKP IBMX

■ =2H42 sildenafil

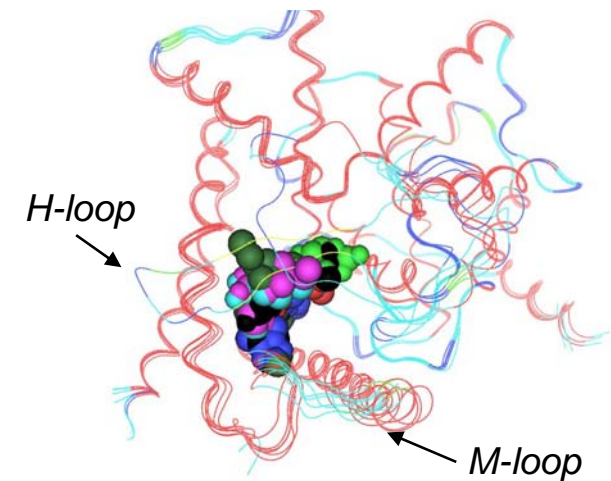
■ =2H44 icarisd II

■ =1T9S GMP

■ =1XOZ tadalafil

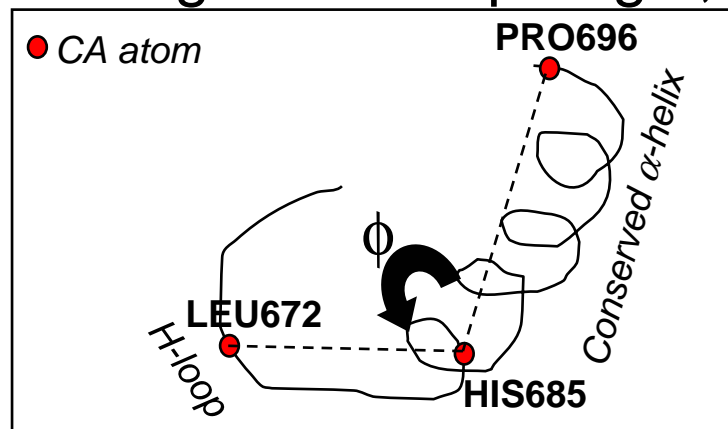
■ =1XP0 vardenafil

■ =2CHM 3P4

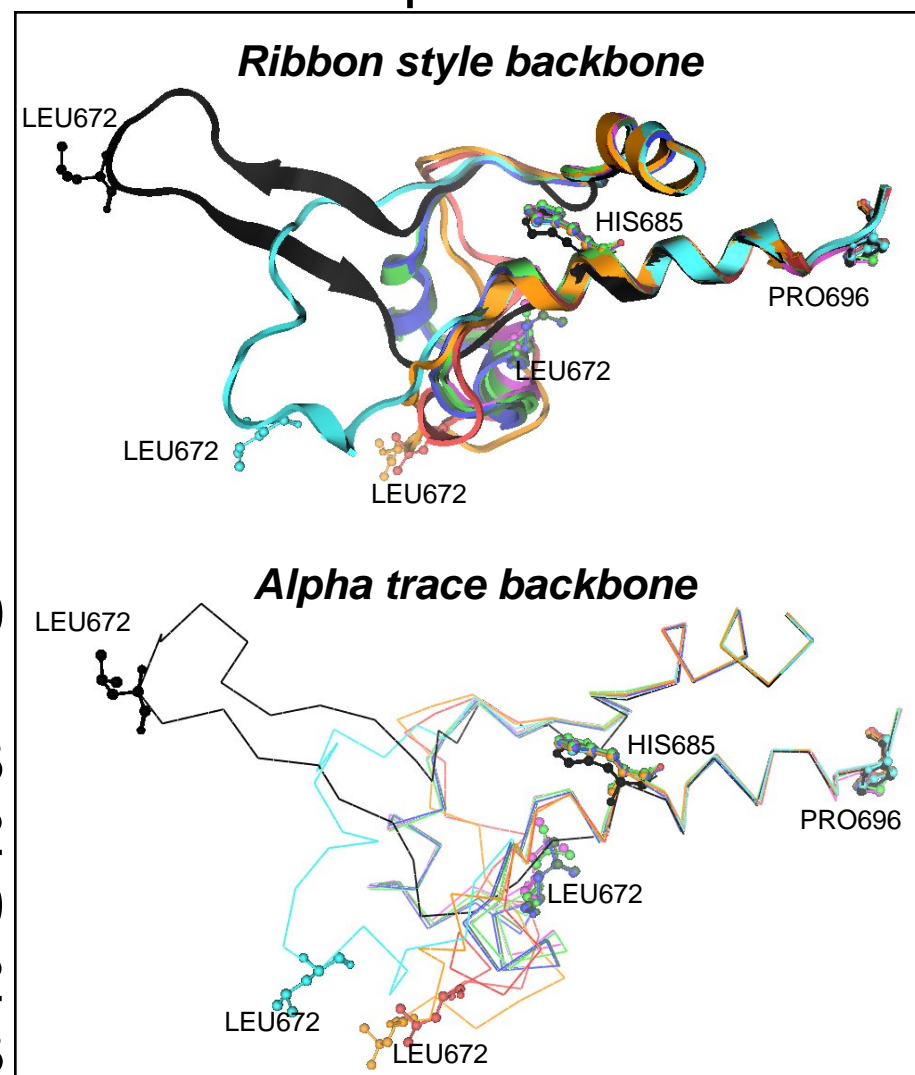


How open is your H-loop?

Defining the H-loop angle, ϕ

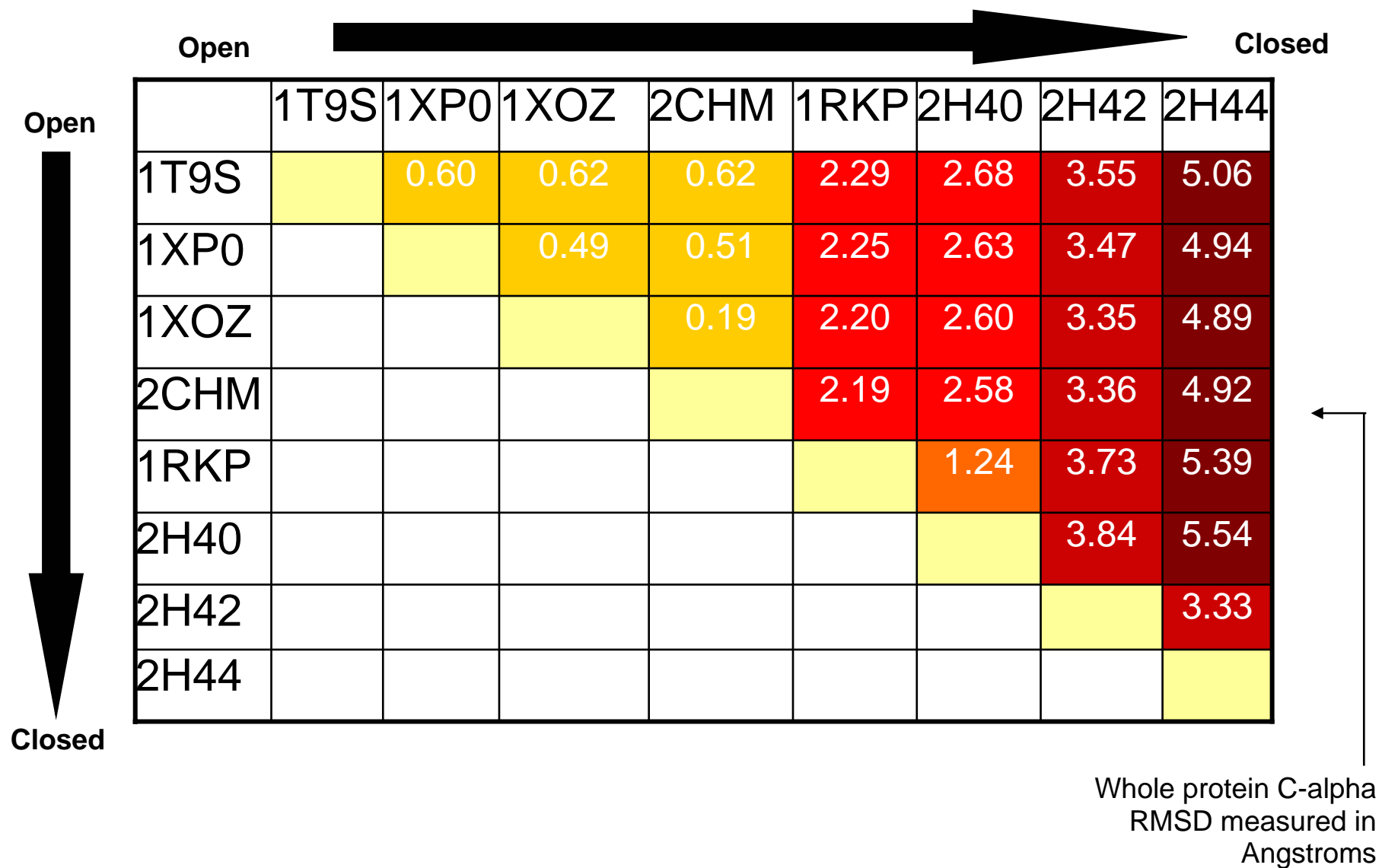


The 8 H-loop conformations

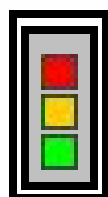


		$\phi/^\circ$
<p>Open</p> <p>Closed</p>	1T9S GMP	248.1
	1XP0 vardenafil	248.0
	1XOZ tadalafil	247.1
	2CHM 3P4	246.8
	1RKP IBMX	234.2
	2H40 no ligand	233.9
	2H42 sildenafil	209.2
	2H44 icarisd II	166.3

Protein RMSD matrix



Ligand series



Inactive

Moderately active

Highly active

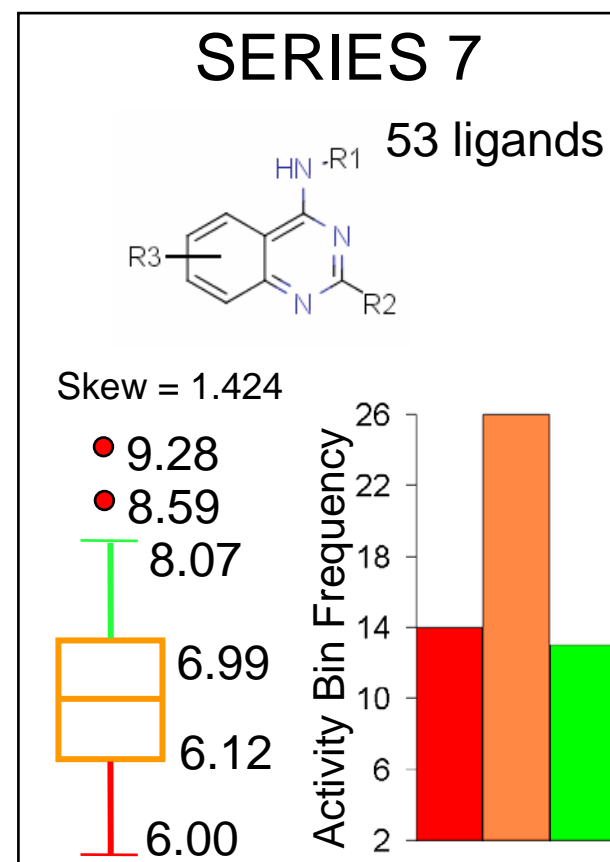
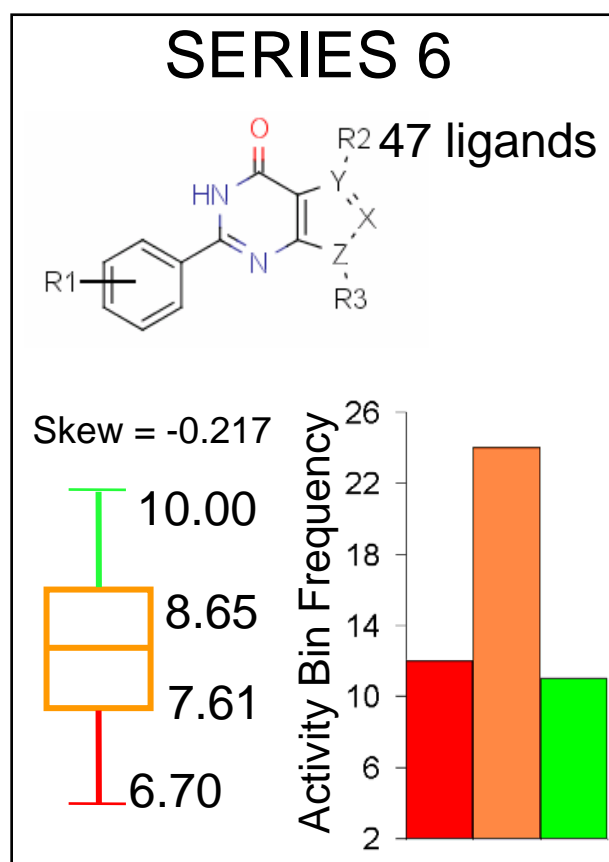
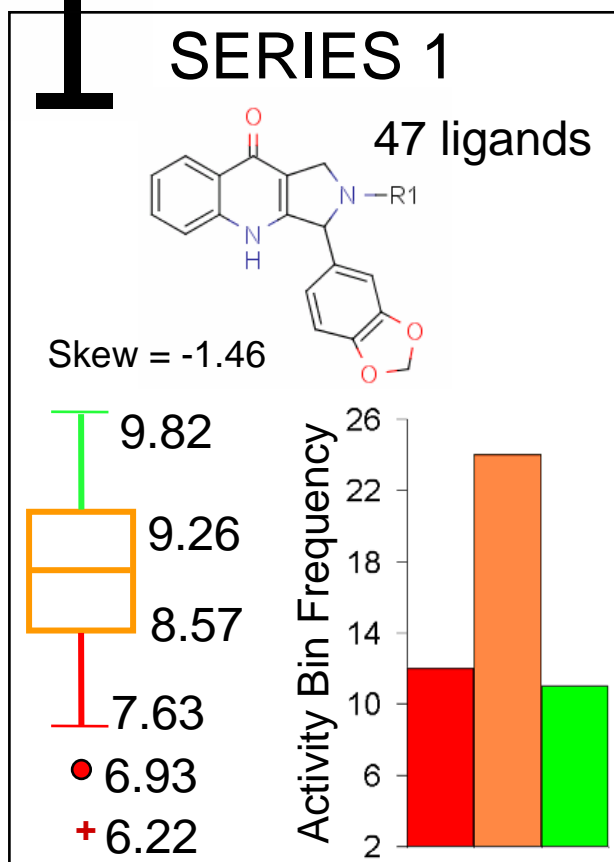
- Mild outliers
- + Extreme outliers

•Williams, C. (2006) *Molecular Diversity*, **10**, 311-322

•8 series - 3 of which chosen for this analysis

•Units of activity = pIC50

$$skew = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{s} \right)^3$$



Ligand Similarity

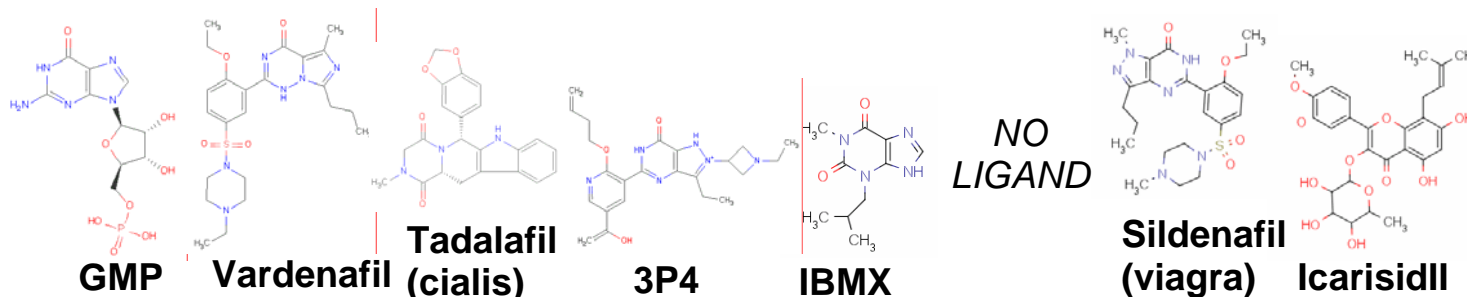
MACCS key tanimoto similarity of ligands to ligand bound in protein conformation



Open

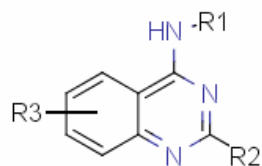
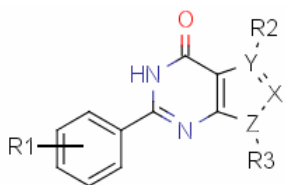
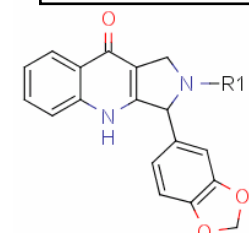


Closed



1T9S
1XPO
1XOZ
2CHM
1RKP
2H40
2H42
2H44

1	0.18 0.24	0.09 0.20	0.36 0.86	0.07 0.12	0.11 0.16		0.16 0.26	0.01 0.21
6	0.13 0.23	0.18 0.48	0.11 0.20	0.15 0.21	0.16 0.37		0.16 0.47	0.00 0.00
7	0.06 0.25	0.03 0.08	0.06 0.22	0.03 0.08	0.02 0.09		0.04 0.18	0.00 0.09



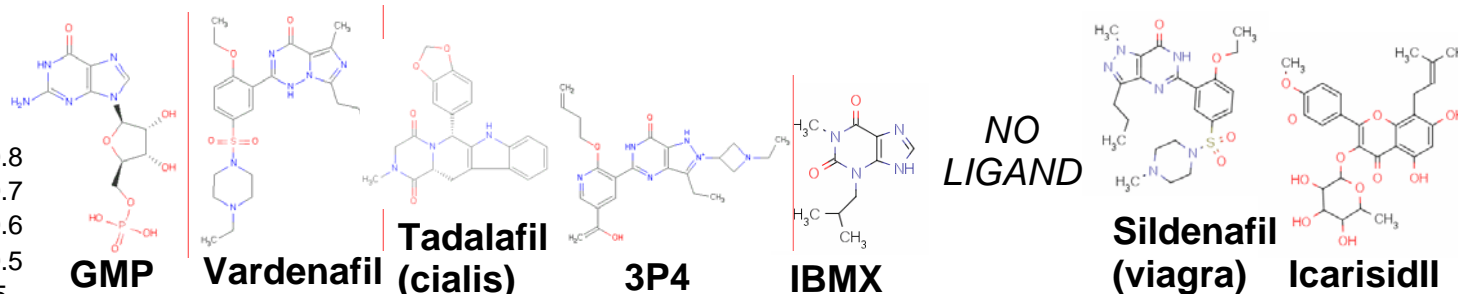
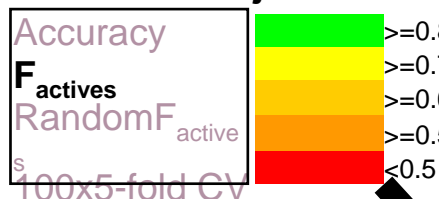
Interaction Tree SAR results

Medium Pruning

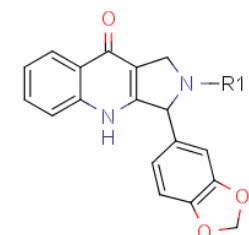


Open Closed

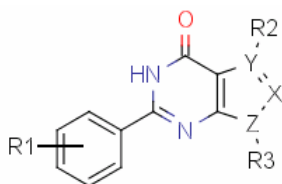
Confidence factor=0.1
minNumObj=4



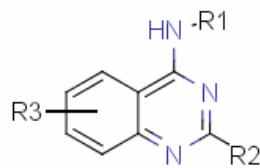
1T9S 1XPO 1XOZ 2CHM 1RKP 2H40 2H42 2H44



1	0.81	0.83	0.79	0.79	0.70	0.72	0.75	0.77
	0.78	0.82	0.80	0.76	0.50	0.73	0.53	0.64
	0.69	0.61	0.58	0.62	0.63	0.63	0.61	0.66
	0.38	0.42	0.41	0.52	0.45	0.42	0.49	0.51



6	0.91	0.87	0.79	0.87	0.83	0.83	0.89	0.81
	0.90	0.83	0.40	0.78	0.77	0.67	0.87	0.63
	0.69	0.63	0.64	0.64	0.64	0.61	0.59	0.61
	0.62	0.59	0.63	0.55	0.55	0.62	0.59	0.59



7	0.85	0.79	0.68	0.83	0.81	0.79	0.72	0.81
	0.88	0.67	0.00	0.77	0.89	0.76	0.67	0.63
	0.71	0.62	0.68	0.69	0.62	0.63	0.64	0.64
	0.40	0.48	0.41	0.43	0.33	0.37	0.45	0.44

Concluding Remarks

INVERSE STRUCTURE BASED DRUG DESIGN

- Novel pipeline strategy
- Fingerprints
- Interaction trees – predictability and interpretability

FUTURE WORK

- Regression trees
- Incorporate protein conformer generation
- GPCR data

Acknowledgements

Thank you to...

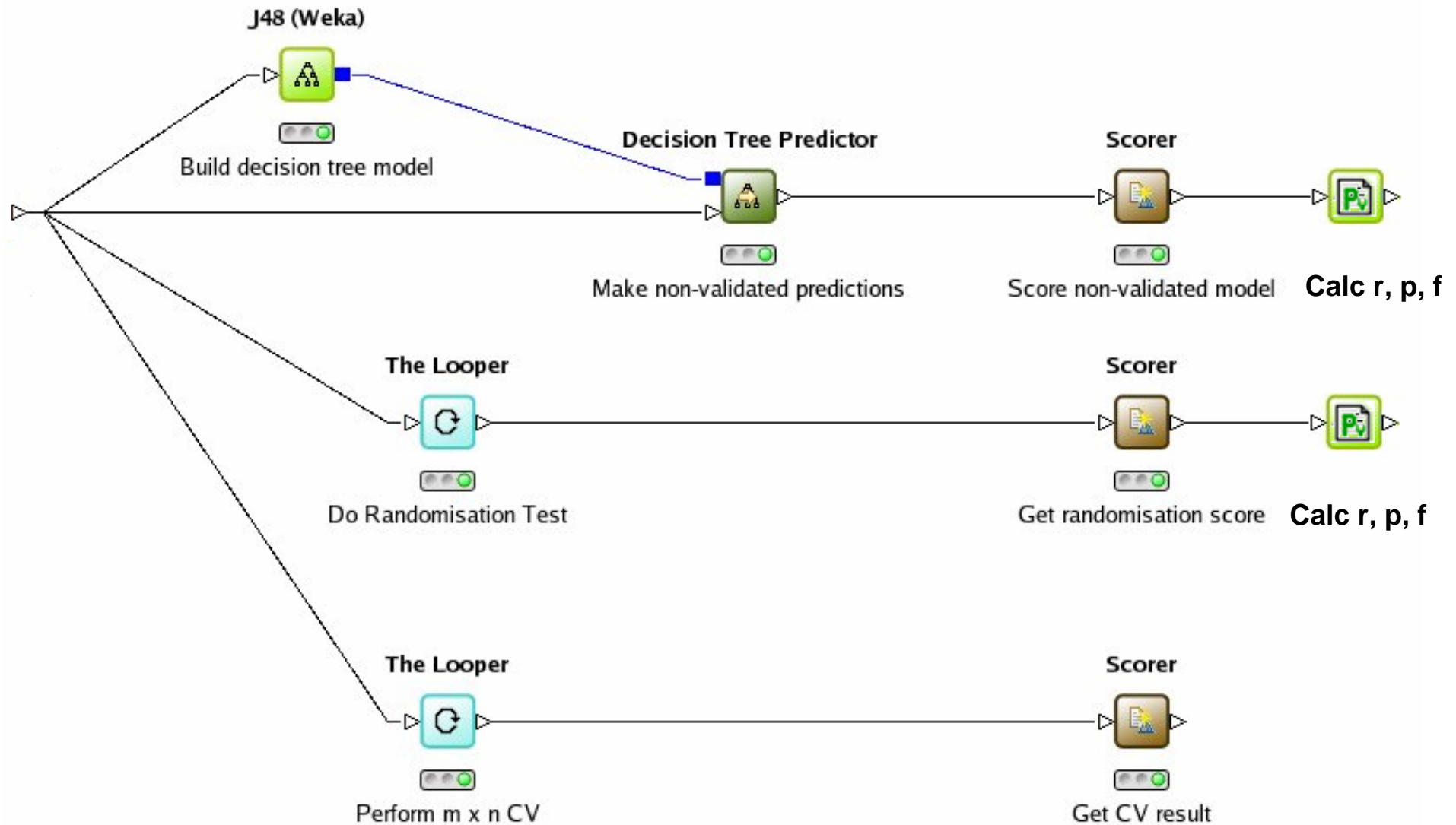
Mike Bodkin

Dave Evans

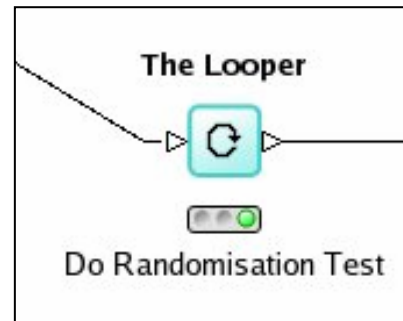
Dave Thorner

MOE, KNIME, CCDC, Schrodinger

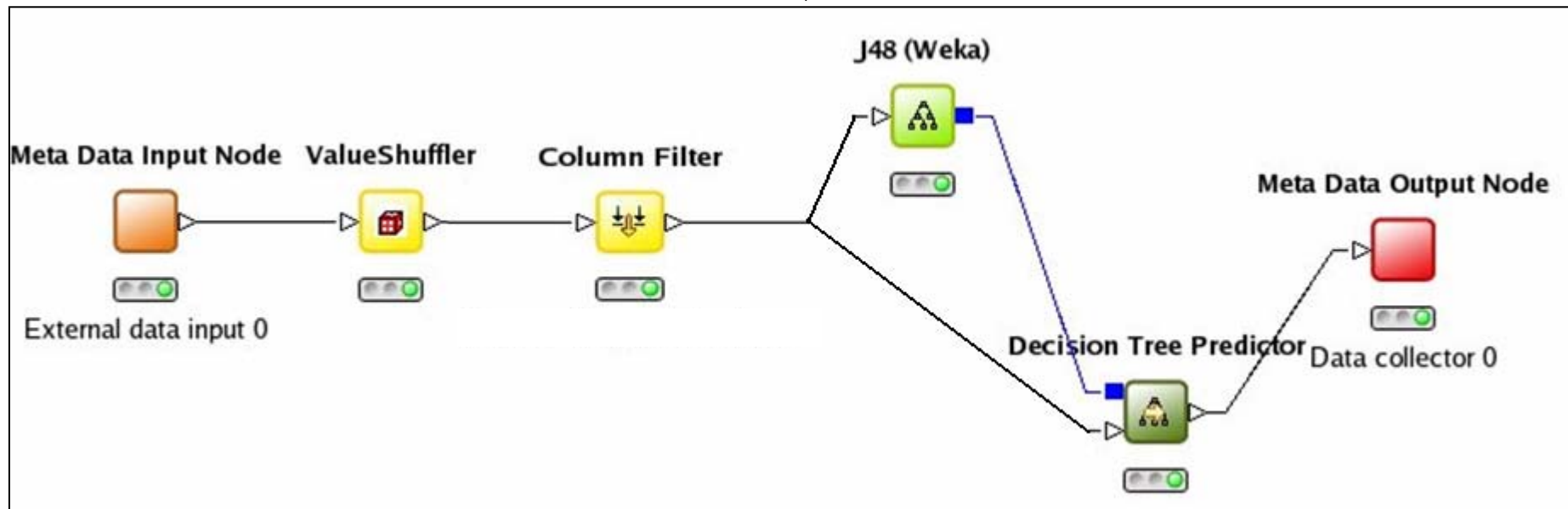
Appendix: Validation Workflows



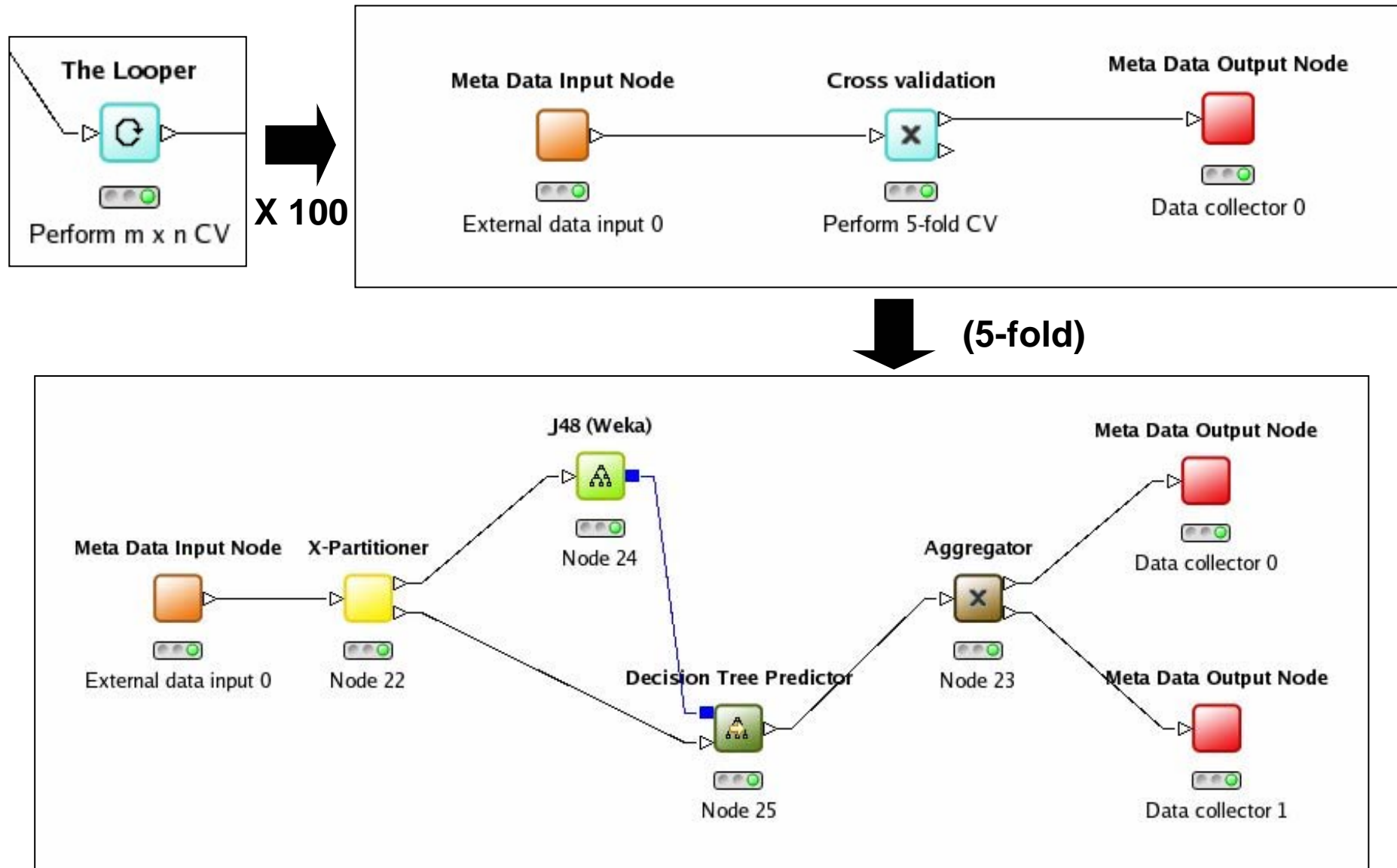
Appendix: Validation Workflows



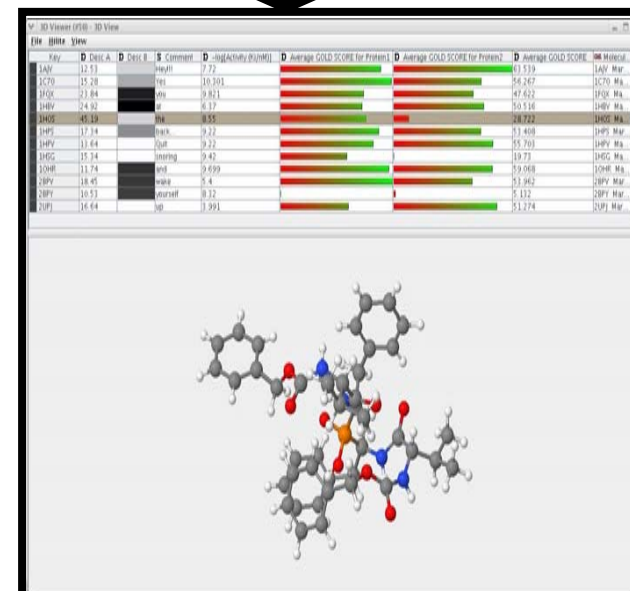
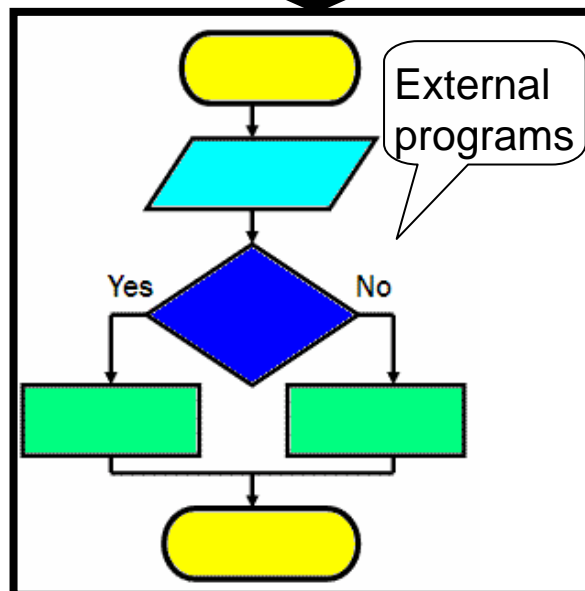
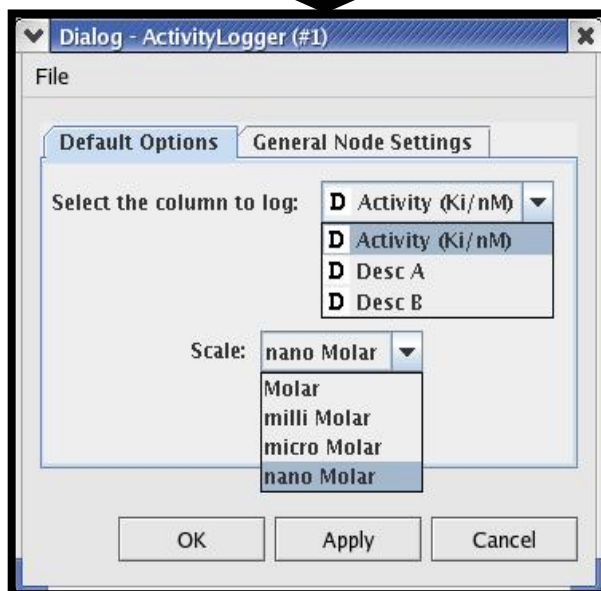
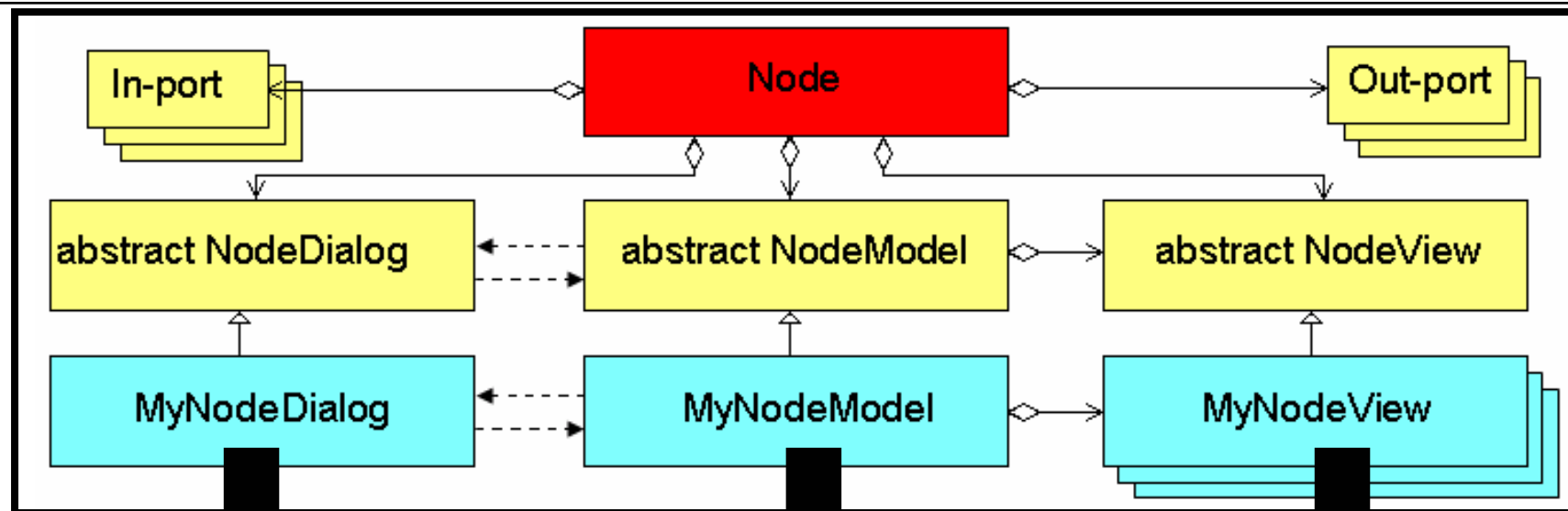
X 100



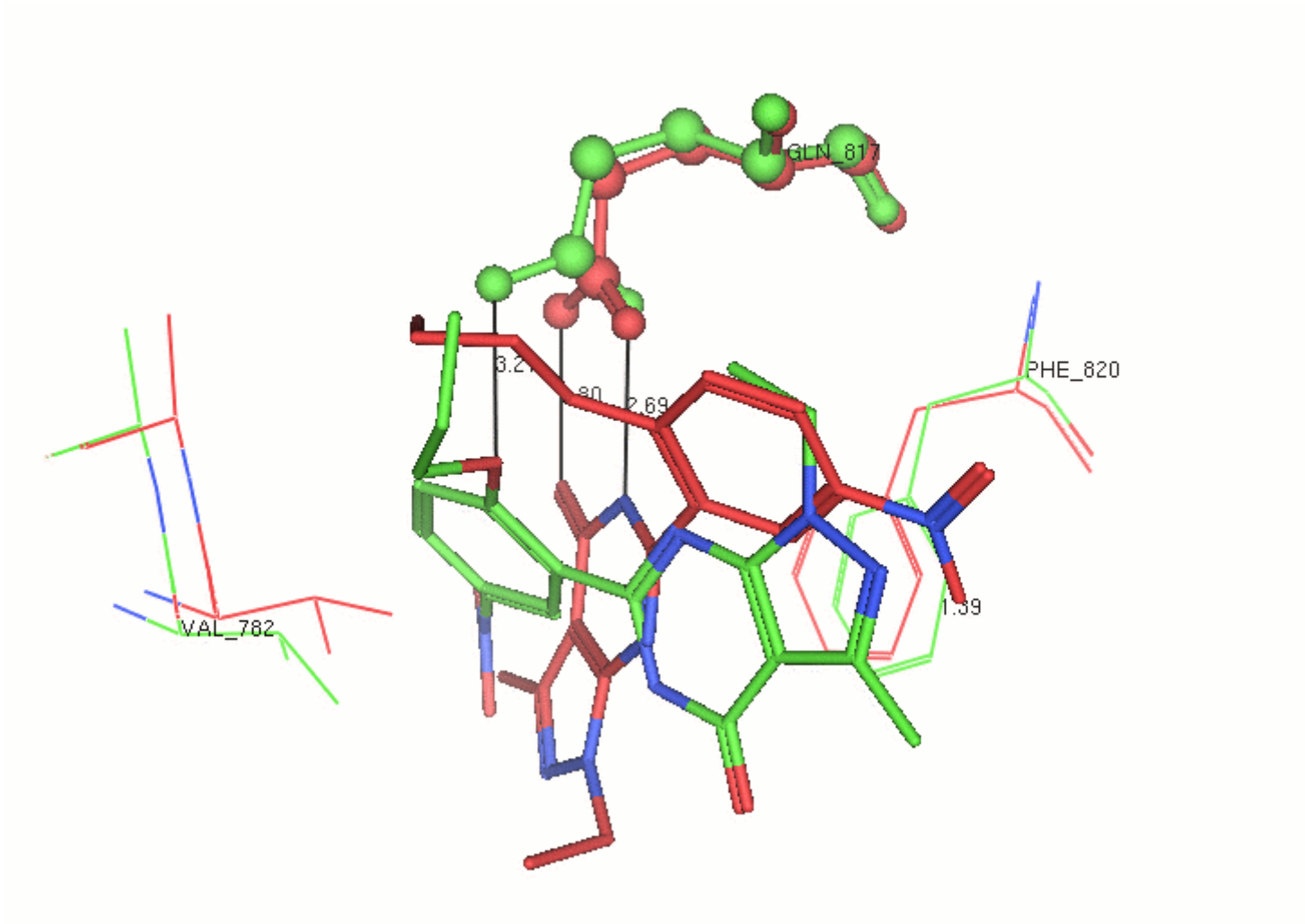
Appendix: Validation Workflows



Node Architecture



Docking of lead from series 6 to 1XPO and 1XOZ



Protein RMSD and ligand similarity matrix

