

SAMPL at first glance

So much data, so little time..

Geoff Skillman, Greg Warren,
& Anthony Nicholls



Data sets

- **Transfer energies – Peter Guthrie**
 - 63 new values & error bars from the literature
 - Multi-functional & flexible
- **JNK3 Kinase –Vertex Pharmaceuticals**
 - 52 structures in ~12 classes
 - Co-crystal structures with data
 - K_i calculated from 10 point curve (1-3 times)
 - K_i spanning 3.55 log units (5.15-8.7)
- **Urokinase – Abbott Laboratories**
 - 27 structures in ~9 classes
 - Co-crystal structures with uncertainty
 - K_i spanning 4.9 log units (4.3-9.2)



Responsibility & Philosophy

- Learn most from the data
- Let participants *evaluate* methods
 - File drawer
 - It's for them, not us
- Run our own experiments
- Answer bigger questions
- Save things for next year!

Invited...

- SAMPL-0, Winter 2007
- Relevant but conservative
- Nov 1, 2007 Invitation
 - 6 major competitors
 - 16 leading academic/government labs
 - 12 major pharmaceutical companies
- Word of mouth



Participation

- 54 active participants
 - $\sim 1/2$ acad-gov $\sim 1/3$ anonymous
- 205 predictions

Virtual Screen	JNK Kinase	Urokinase
-worse structure	29	21
-better structure	23	17
Pose Prediction		
-cross-docking	20	18
-self-docking	6	4
Affinity	20 (+5)	5 (+5)
Transfer Energy	31	

Who participated?

- **Classes**
 - Vendors
 - Academic developers
 - 3rd party users
- **Institutions**
 - Pharma
 - Academic
 - Government
- **OpenEye**
 - Only in Peter Guthrie's transfer energy portion
 - Roger Sayle, VS Control
- **Algorithm Types**
 - Ligand-based
 - Structure-based
 - QSAR
 - 2D,3D
 - Hybrid



Virtual Screening



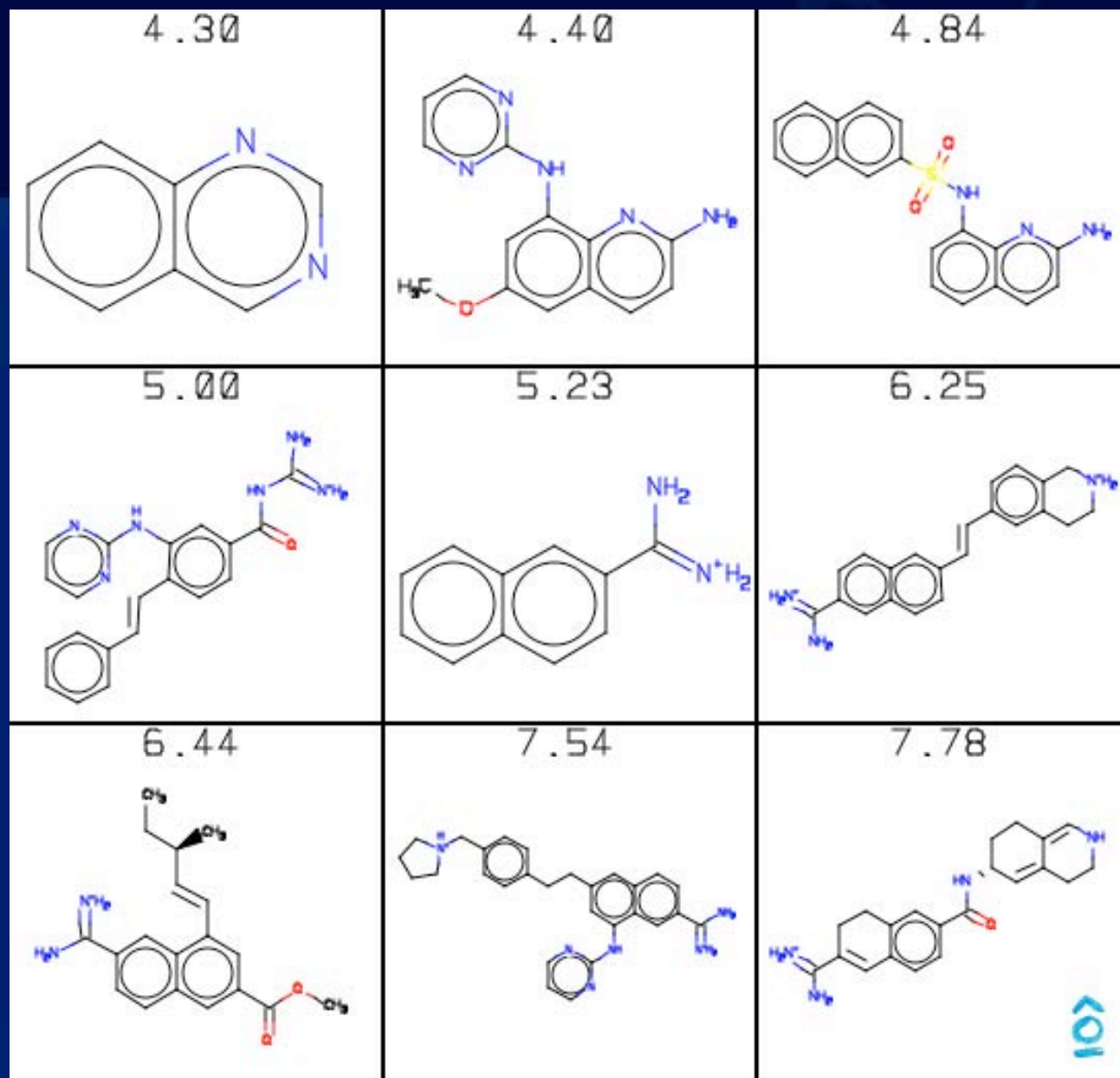
Experiments: Virtual screening

- Actives
 - All
 - Good
 - Class-based & Low activity
- Structures
 - JNK
 - 1jnk, Met 146 in, 2.3A
 - ATP analog
 - 1pmq,146 out, 2.2A
 - imidazole-pyrimidine
 - UK
 - 1owd, res 2.3, DPI 1.1
 - naphthalamidine
 - 1o5c, res 1.6, DPI 0.43,
 - benzimidazole amidine
- Decoys
 - Standard
 - Random
 - Rognan
 - Drug-like
 - Schrodinger
 - DUD-like
 - Specific, paired
 - Uncolored
 - Shape
 - Clash
 - 12K & 8K



Good

- Rank by affinity
- Take first
 - Remove similar
 - MACCS > 0.6
 - Lingo > 0.45
- Take next
 - repeat
- Prefer
 - Low affinity
 - Statistical Power



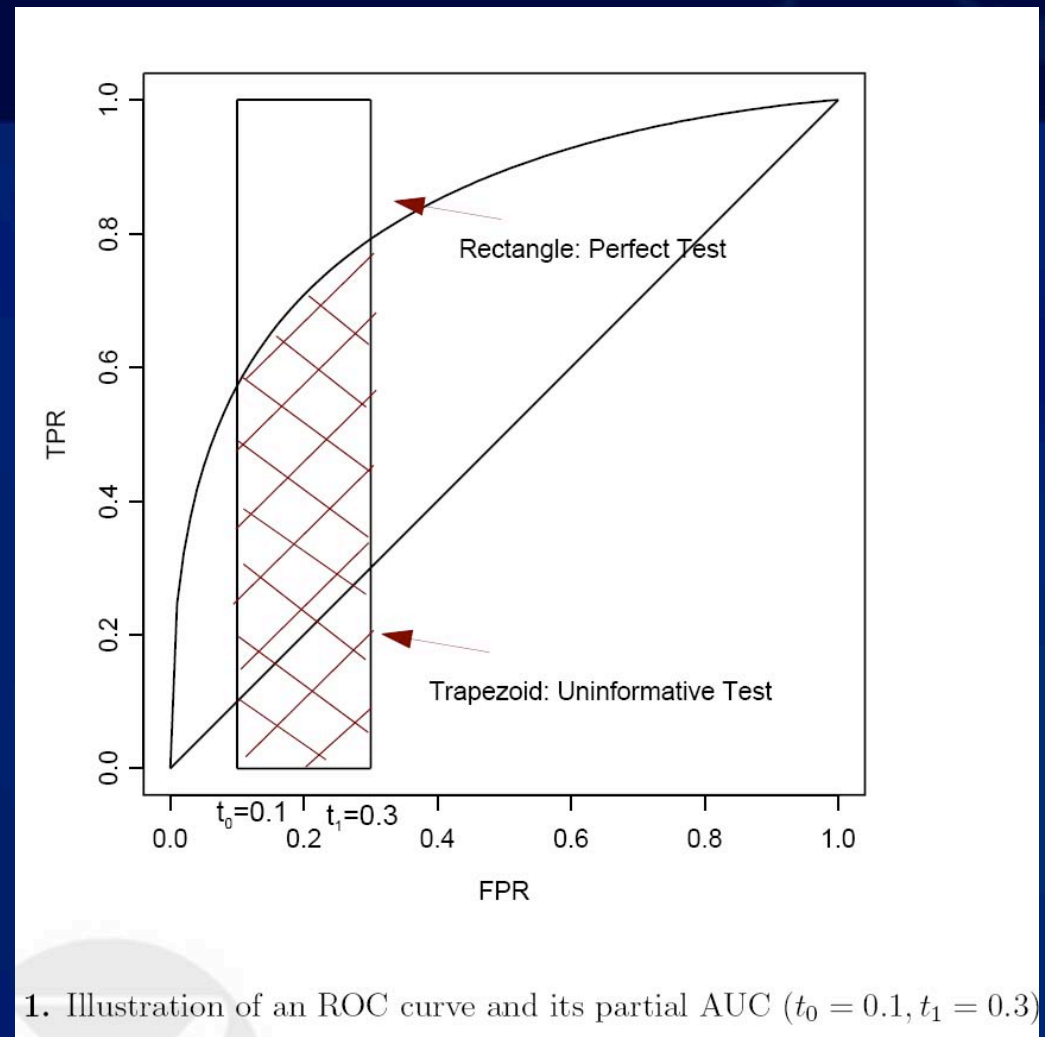
Assessment

- Resample bootstrap with replacement
 - N=2000
- 2, 5 and 10% (0.5, 1 and 2%)
- ROC enrichment @each point
 - 95% bootstrap interval
- Area under 1st 10% ROC curve
 - 95% bootstrap interval
- Good Actives, Drug-like Decoys



Partial AUC

- ROC curve
 - f_{active} , f_{decoy}
- $t_0=0.0$, $t_1=0.1$
- Normalized (0-1)
- Perfect = 1
- Random = 0.05



Dodd LE, Pepe MS. Partial AUC estimation and regression. *Biometrics*. 59(3):614-23, 2003

Within error of best, good-druglike

(alphabetical, no consistent winner, many good are also bad)

- **Jnk.vs.1**

- Anonymous1 (structure)
- FRED
 - cg3, constraint
- GLIDE, SP/HTVS
 - refine, const
- ROCS
- 2D control

- **Jnk.vs.2**

- FRED
- GLIDE, SP/HTVS
 - Refine & const
 - With/without ANP
- ROCS
- 2D control

- **UK.vs.1**

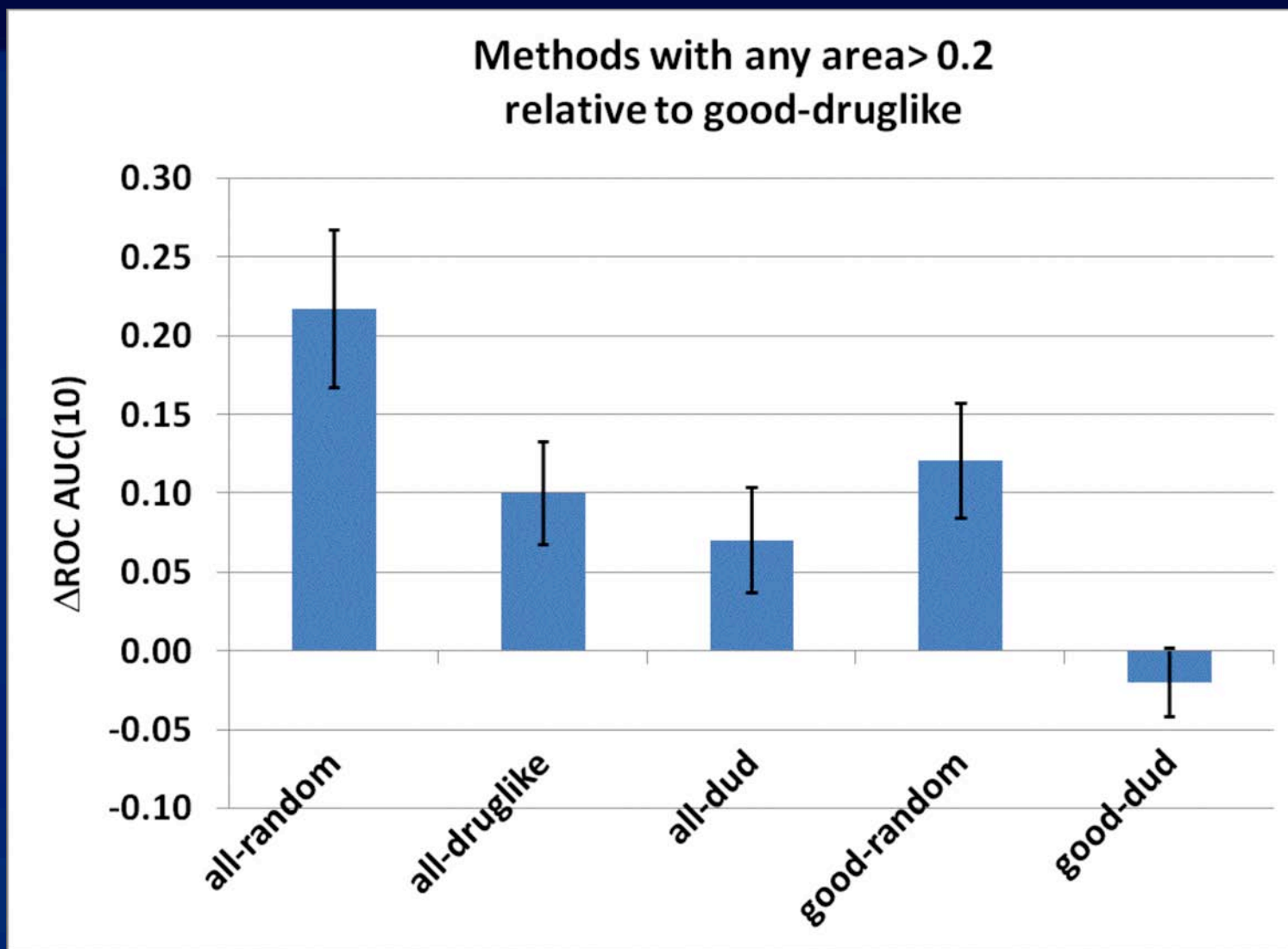
- Anonymous2 (structure)
- FRED
 - cg3, with/without const
- GLIDE, SP/HTVS
 - Refine
- GOLD
 - ASP, consensus
- ROCS, ROCS/EON
- 2D control

- **UK.vs.2**

- Anonymous3 (ligand)
- FRED
 - cg3, without const
- GOLD
- ROCS/EON
- 2D control



Active-Decoy performance



Consistency with Active/Decoy

- Not just more difficult

	All Rand	All D-Like	All DUD	Good Rand	Good D-like	Good DUD
All Rand	1.00					
All D-Like	0.94	1.00				
All DUD	0.90	0.96	1.00			
Good Rand	0.97	0.92	0.88	1.00		
Good D-Like	0.85	0.94	0.90	0.90	1.00	
Good DUD	0.79	0.88	0.91	0.85	0.95	1.00



Consistency with Active/Decoy

- Not just more difficult

	All Rand	All D-Like	All DUD	Good Rand	Good D-like	Good DUD
All Rand	1.00					
All D-Like	0.94	1.00				
All DUD	0.90	0.96	1.00			
Good Rand	0.97	0.92	0.88	1.00		
Good D-Like	0.85	0.94	0.90	0.90	1.00	
Good DUD	0.79	0.88	0.91	0.85	0.95	1.00



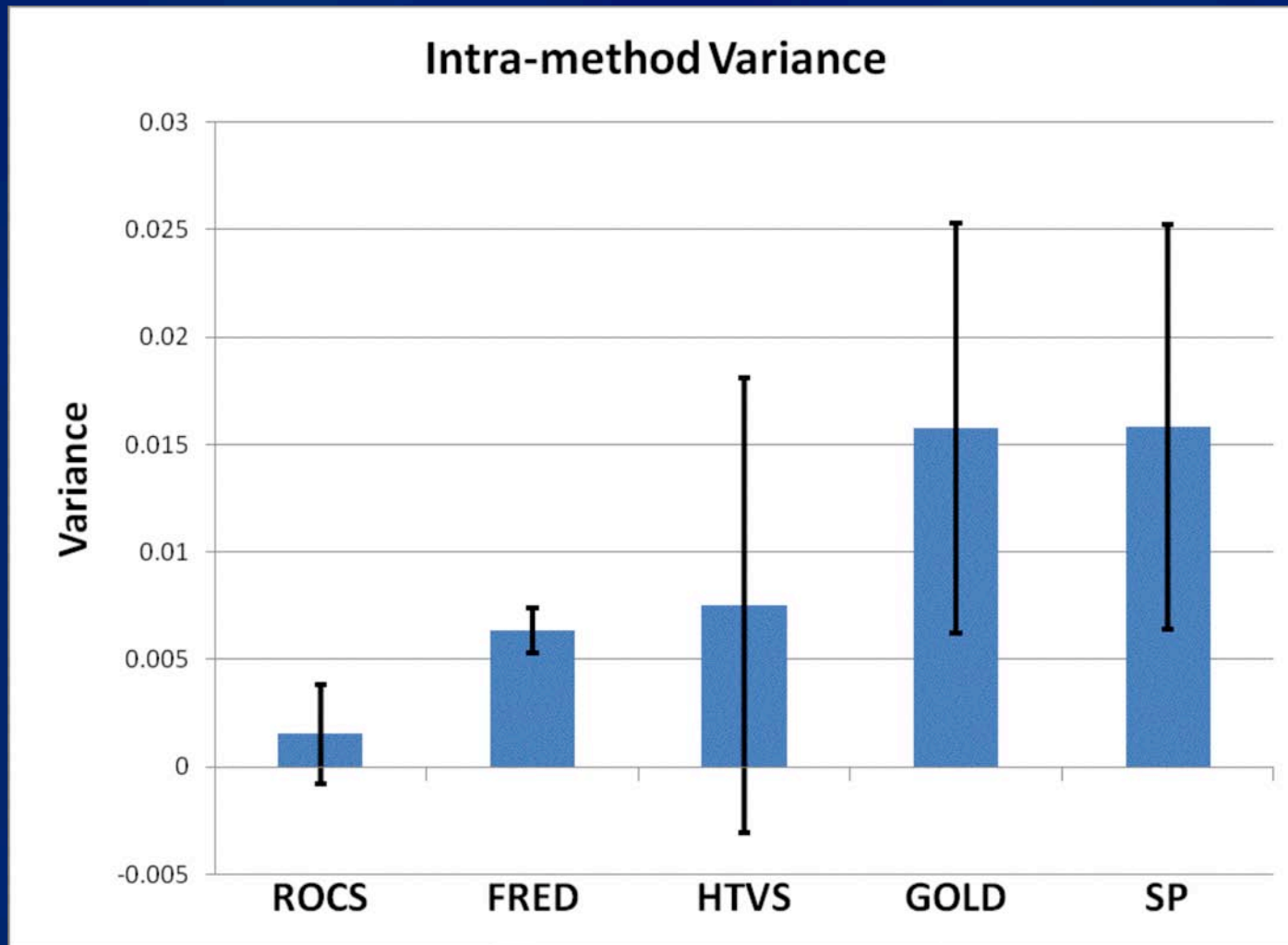
Ranking consistency

- Kendall Tau Rank Correlation

	All Rand	All D-Like	All DUD	Good Rand	Good D-like	Good DUD
All Rand	1.0					
All D-Like	0.79	1.0				
All DUD	0.75	0.86	1.0			
Good Rand	0.84	0.76	0.7	1.0		
Good D-Like	0.68	0.81	0.73	0.76	1.0	
Good DUD	0.62	0.73	0.75	0.71	0.85	1.0

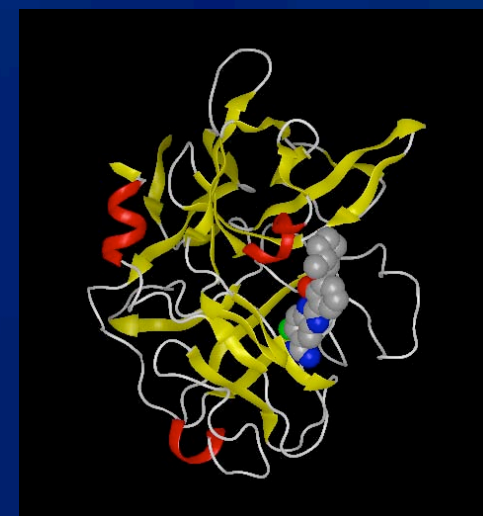
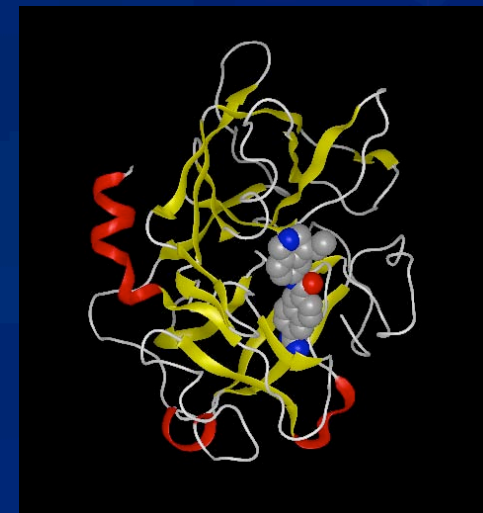


It's all in the ~~User~~ parameters?



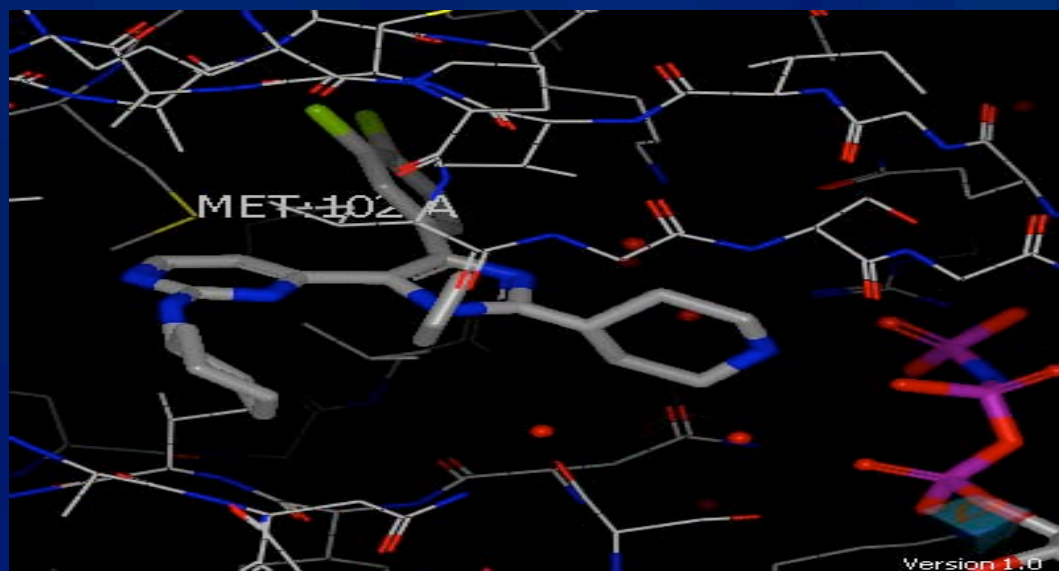
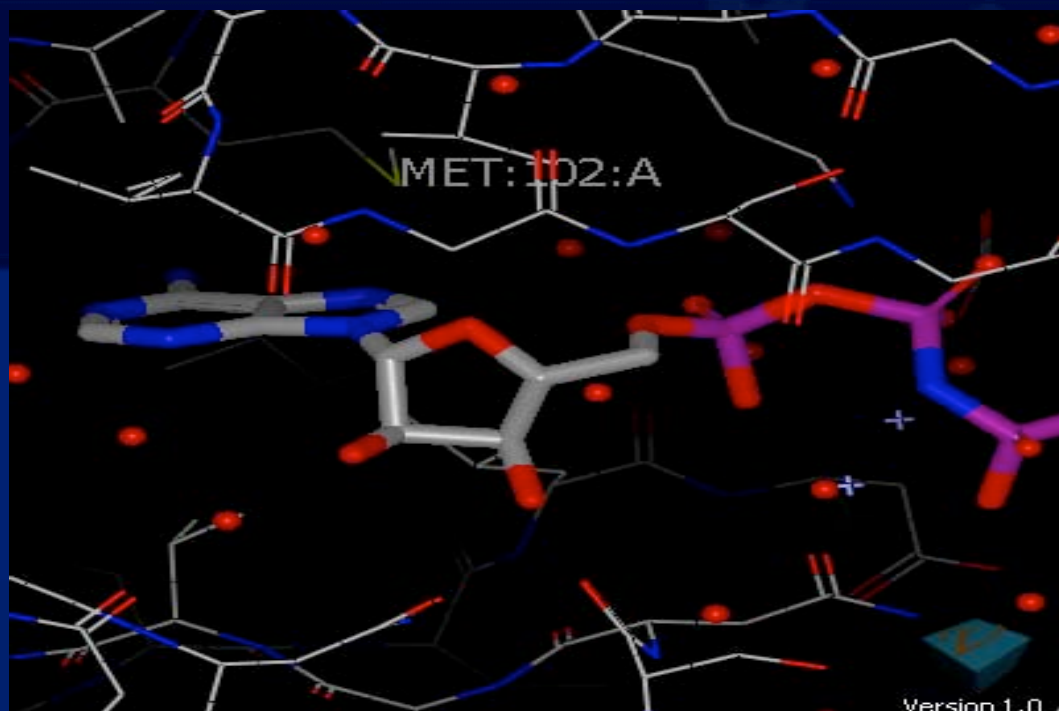
Urokinase 1 vs 2

- 1owd, res 2.3, DPI 1.1
 - naphthalamidine
- 1o5c, res 1.6, DPI 0.43,
 - benzimidazole amidine
- **NO difference in performance!**
 - In any of the 6 active/decoy pairs
 - With mean or median
- **Everyone did quite well**
 - 2D control

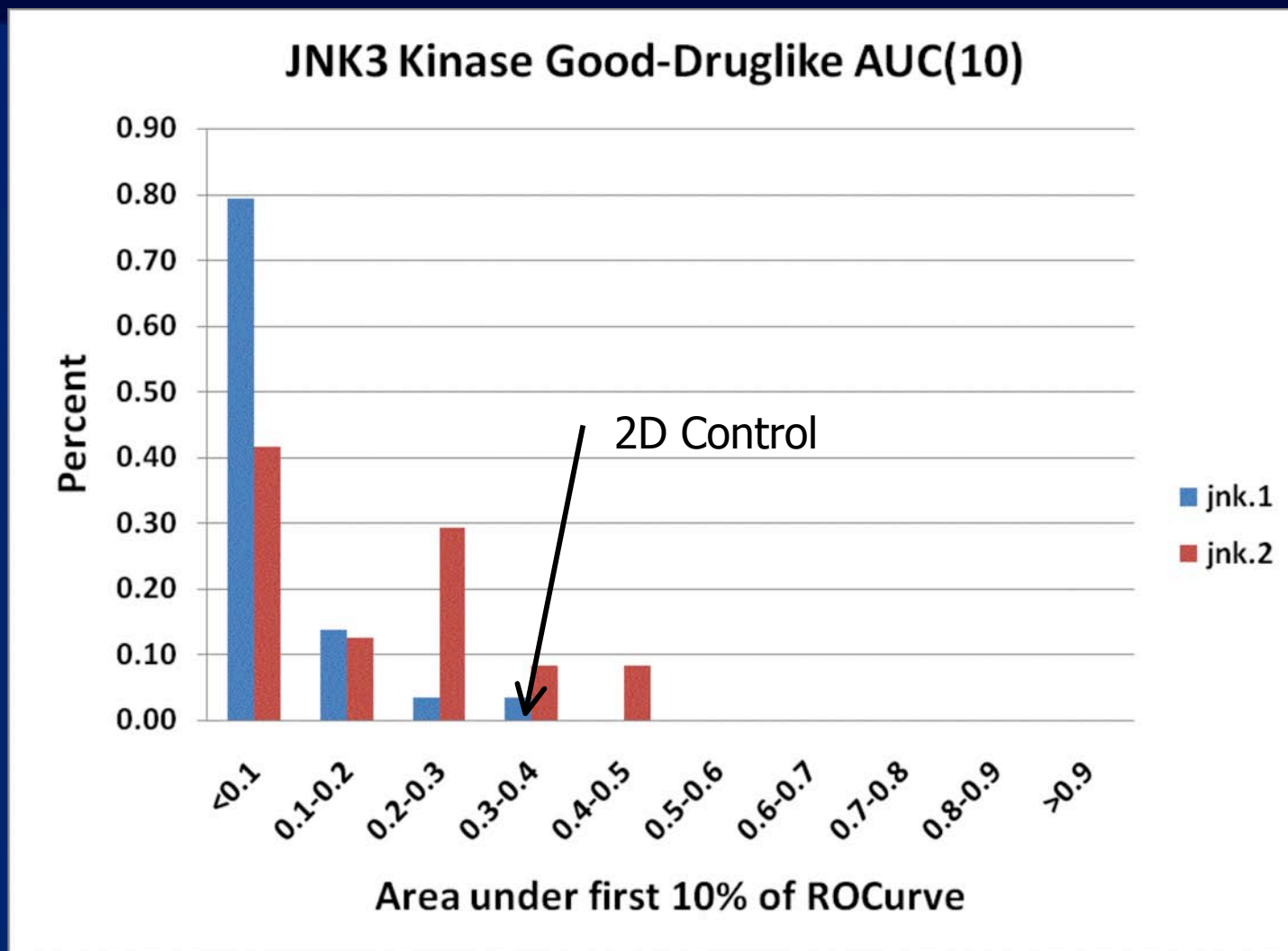


JNK sites

- 30 method pairs
- Require 1 AUC > 0.2
- vs.2 – vs.1
 - good-druglike
 - AUC(10)
- Mean 0.22!
 - 95%ci [0.15,0.29]



All Early JNK3 Data



Pose Prediction



Experiments: Pose Prediction

JNK3

- **Cross-docking**
 - MET 146 in
 - MET 146 out
 - 52 active ligands
 - 10 known inactive
- **Self Docking**
 - 62 protein-ligand pairs

Urokinase

- **Cross-docking**
 - 1 Abbott structure
 - 27 active ligands
 - 7 known inactive
- **Self-docking**
 - 34 protein-ligand pairs

Not enough Method Pairs

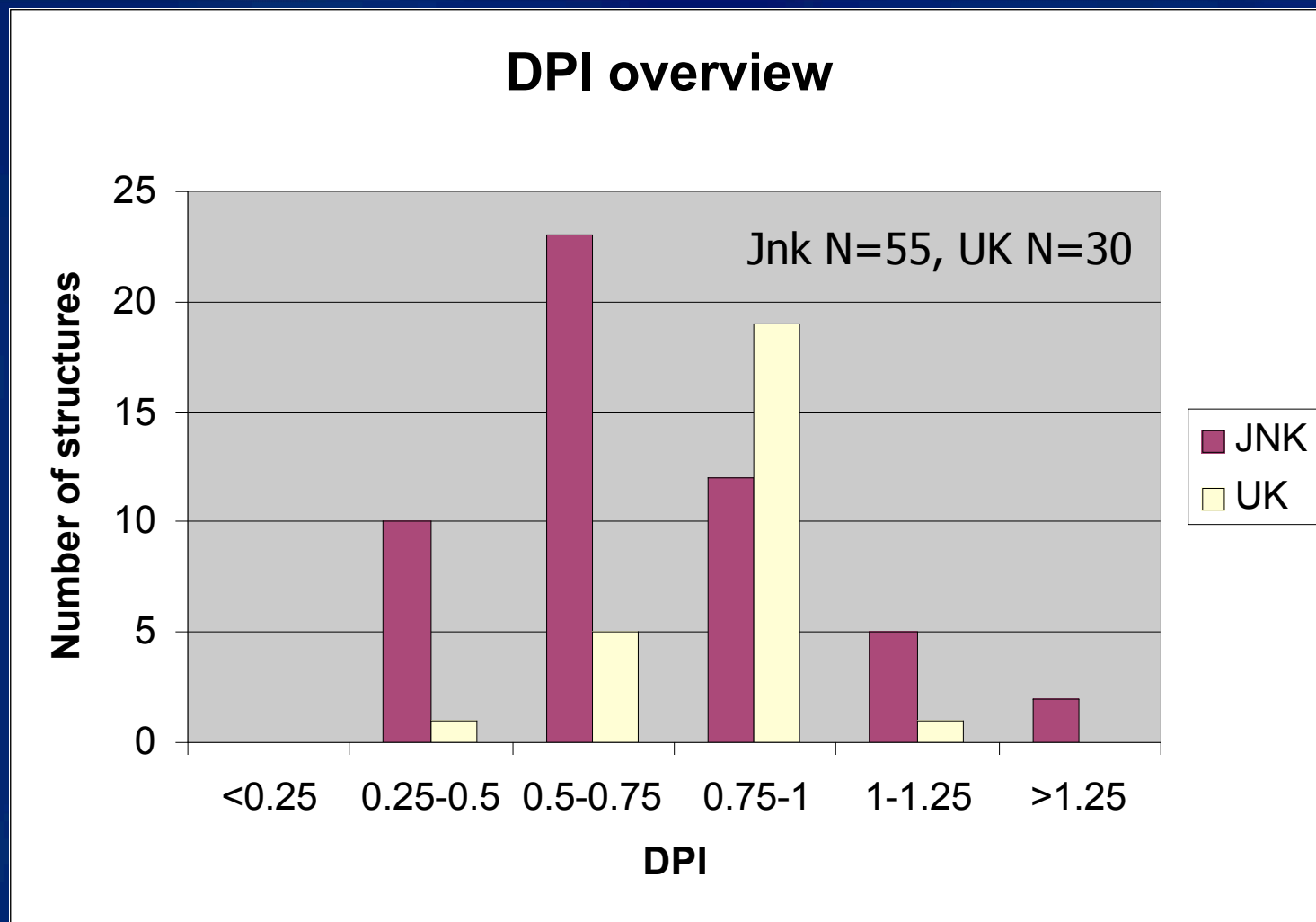


Pose Assessment

- Tin standard: RMSD
 - Unbounded
 - Omnibus problem
- Many aluminum alternatives
- Experimental error
 - DPI/Coordinate precision
- $\text{Max}(0.0, \text{rmsd-dpi})$

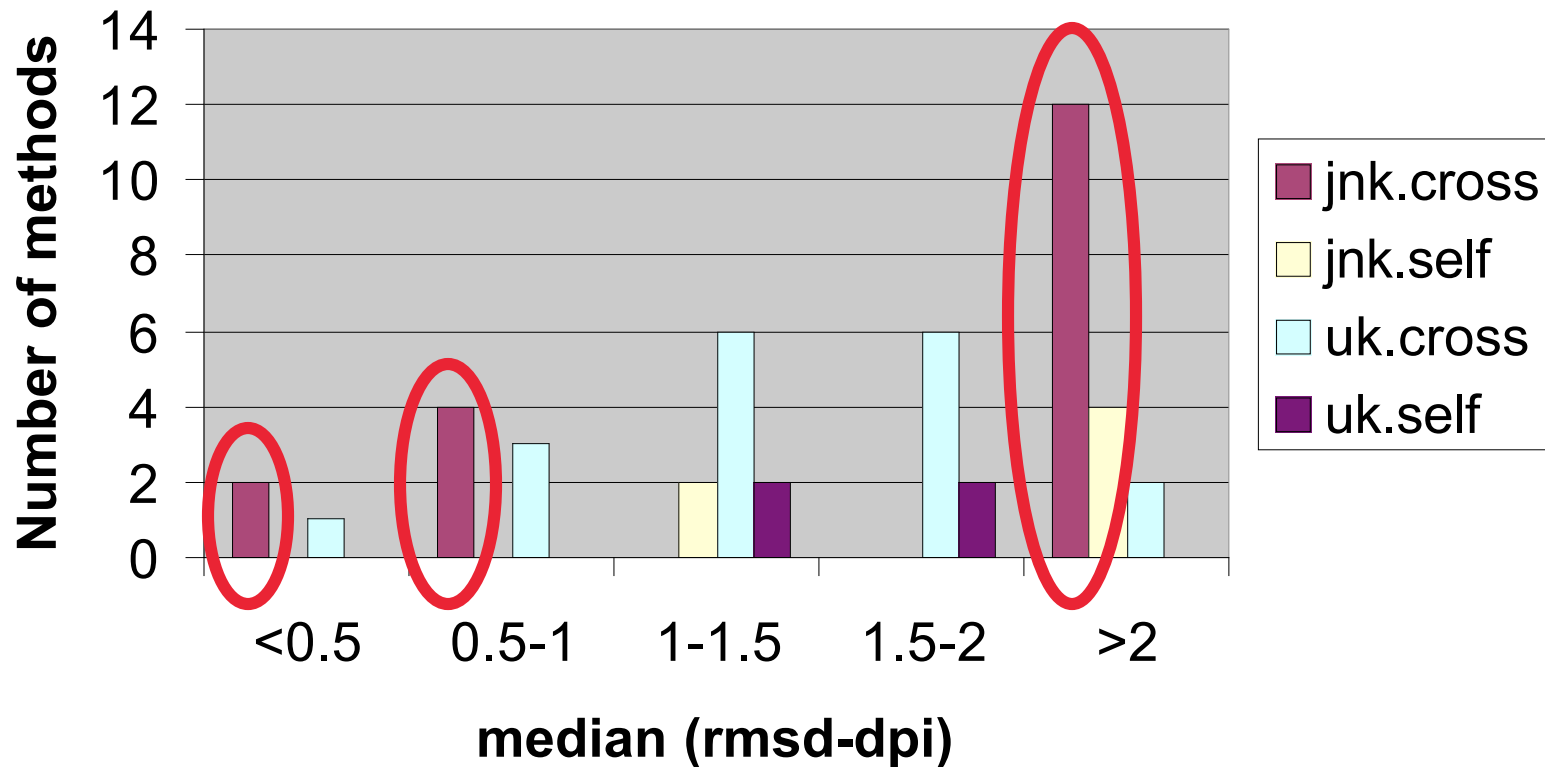


Estimated coordinate precision



Haves & have-nots

All methods pose prediction



Difficult ligands

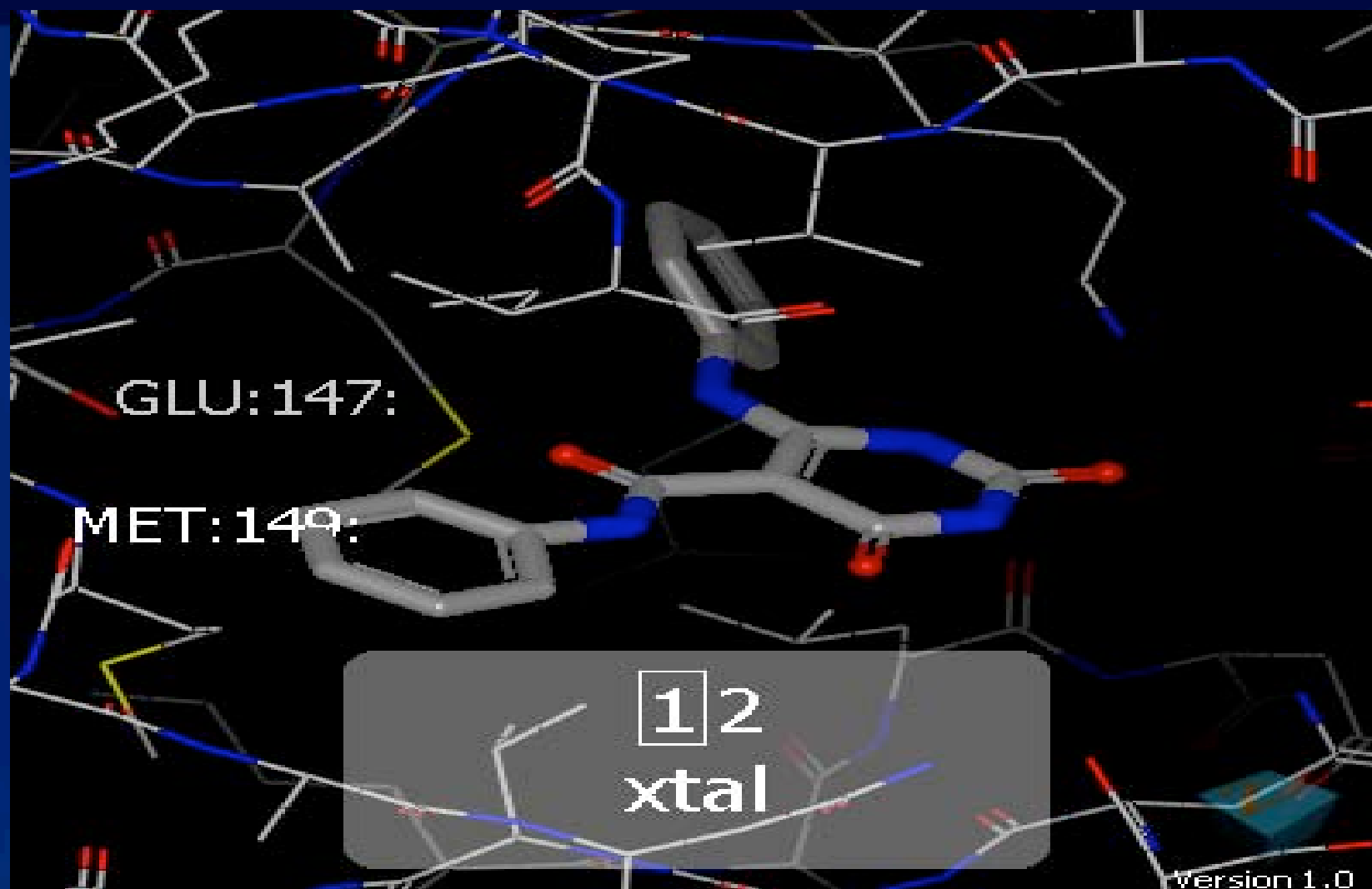
- JNK3 cross-docking
- All submission
- *Best* prediction
- *Worst* Ligands

- 5 different best predictions

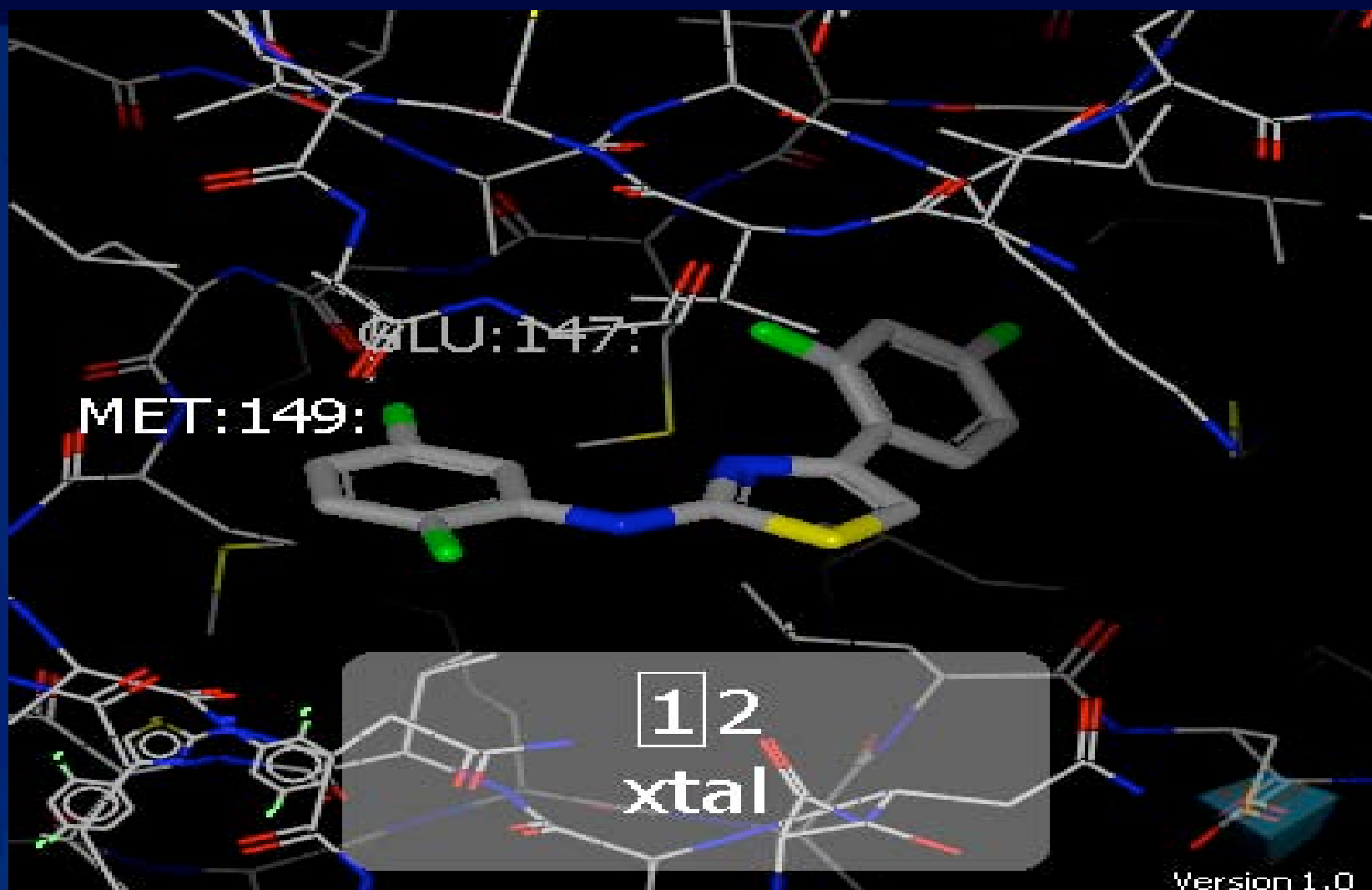
	DPI	RMSD-DPI
Jnk.pp.56	0.77	3.42
Jnk.pp.24	0.74	2.75
Jnk.pp.31	1.31	1.75
Jnk.pp.61	0.84	1.68
Jnk.pp.62	0.74	1.4
Jnk.pp.46	0.67	1.12



jnk.pp.56, 0.77, 3.42

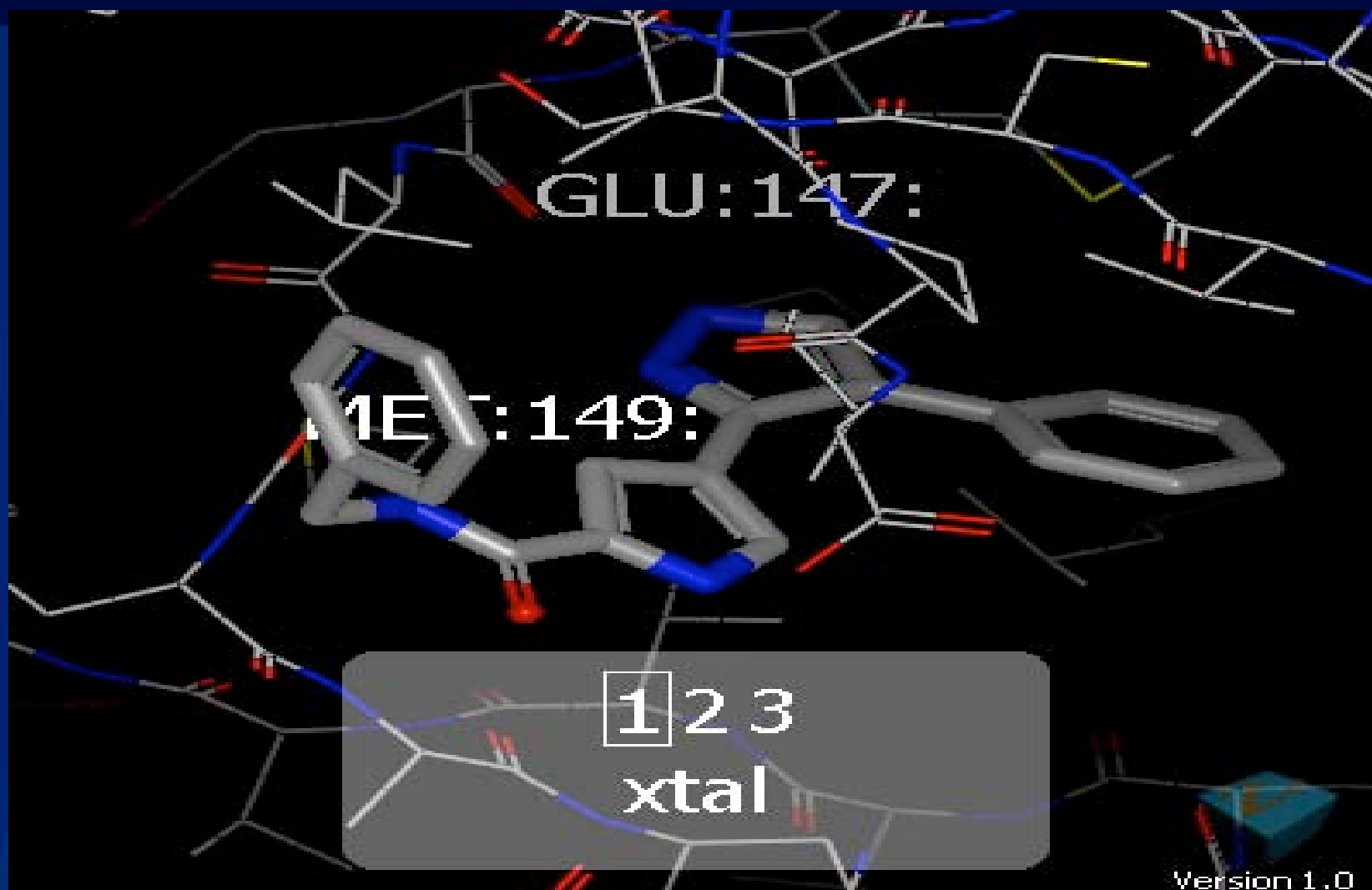


jnk.pp.24, 0.74, 2.75



Version 1.0

jnk.pp.61, 0.84, 1.68



Version 1.0

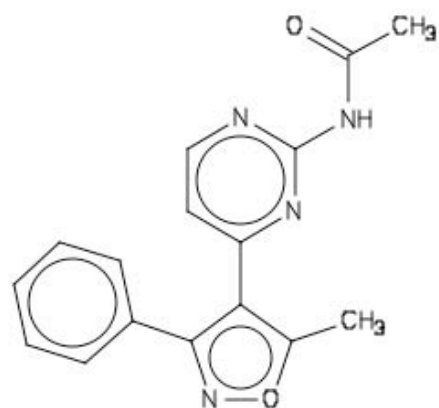
Now the Easy ligands...

- Look for ligands with a low *worst* prediction
- Every ligand had at least 1 prediction >6A!
- Look at 3rd quartile < 3A

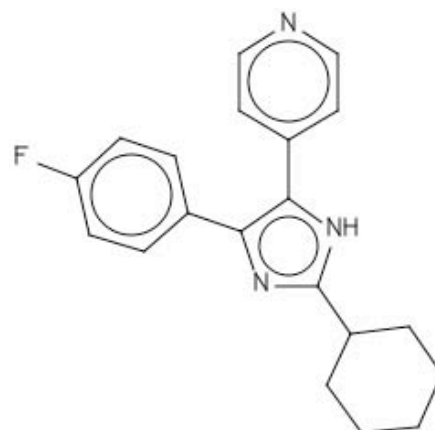
ID	DPI	RMDS-DPI
jnk.pp.38	0.61	2.31
jnk.pp.14	0.71	2.64
jnk.pp.7	0.81	2.73
jnk.pp.44	0.82	2.71

Rigid, small, kinase inhibitor-like

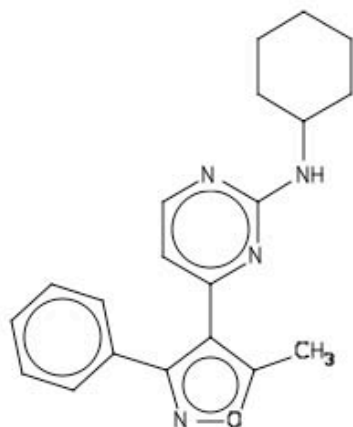
vertex_jnk3_structure32_lig



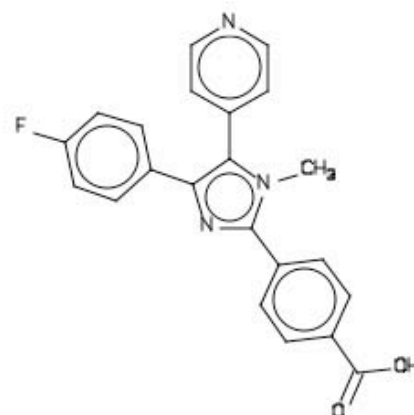
vertex_jnk3_structure8_lig



vertex_jnk3_structure22_lig



vertex_jnk3_structure29_lig



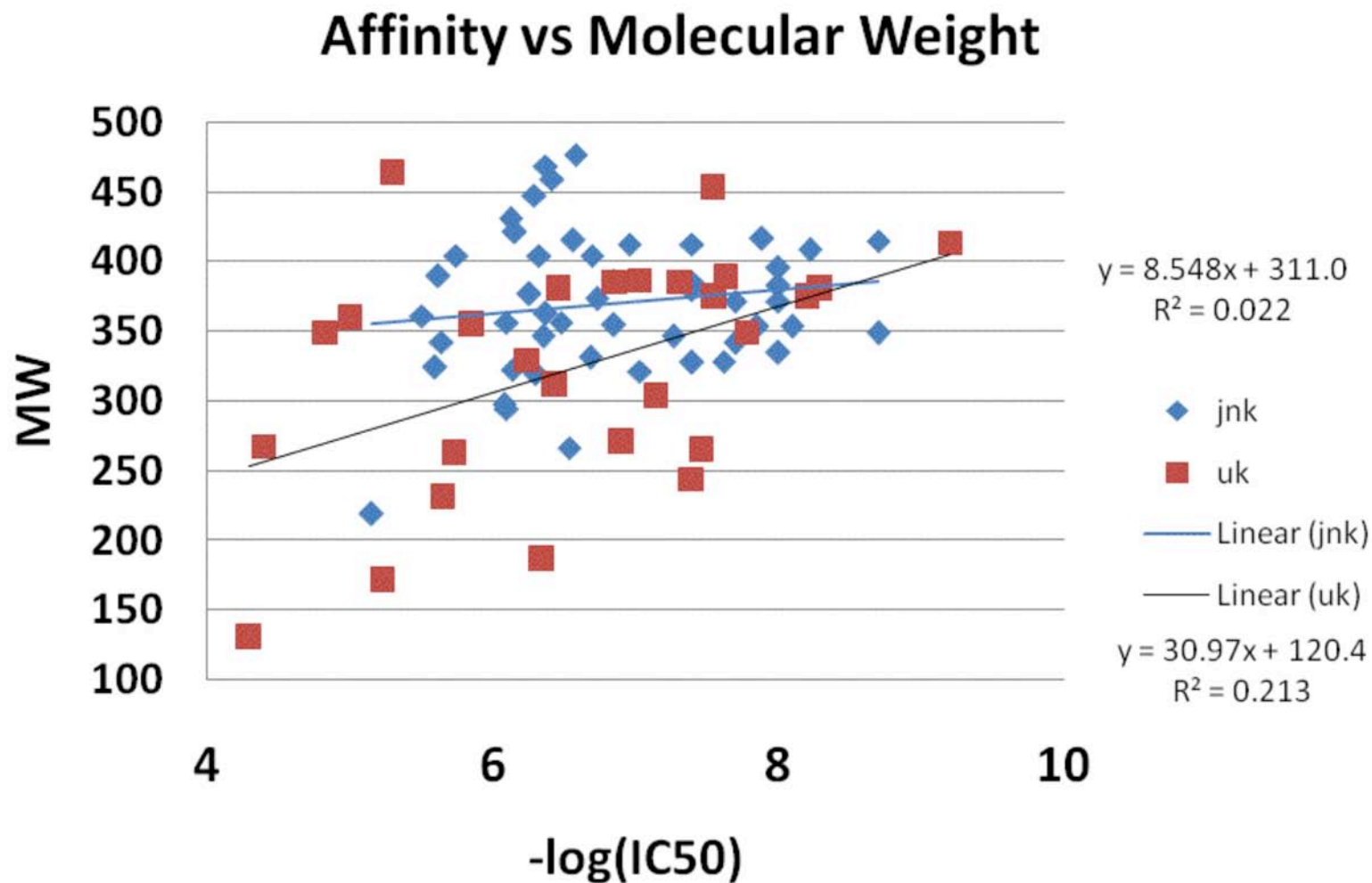
Affinity Prediction



Experiments: Affinity Prediction

- Series of protein-ligand models
 - Co-crystals
 - “Some ligands might not be active”
 - Naïve model
 - Known inactives, ROCS & MMFF, duplicate protein
- JNK
 - 52 co-crystals, IC50 & estimated errors
 - 10 model systems
- Urokinase
 - 27 co-crystals, IC50 & experimental errors
 - 7 model systems

First things first

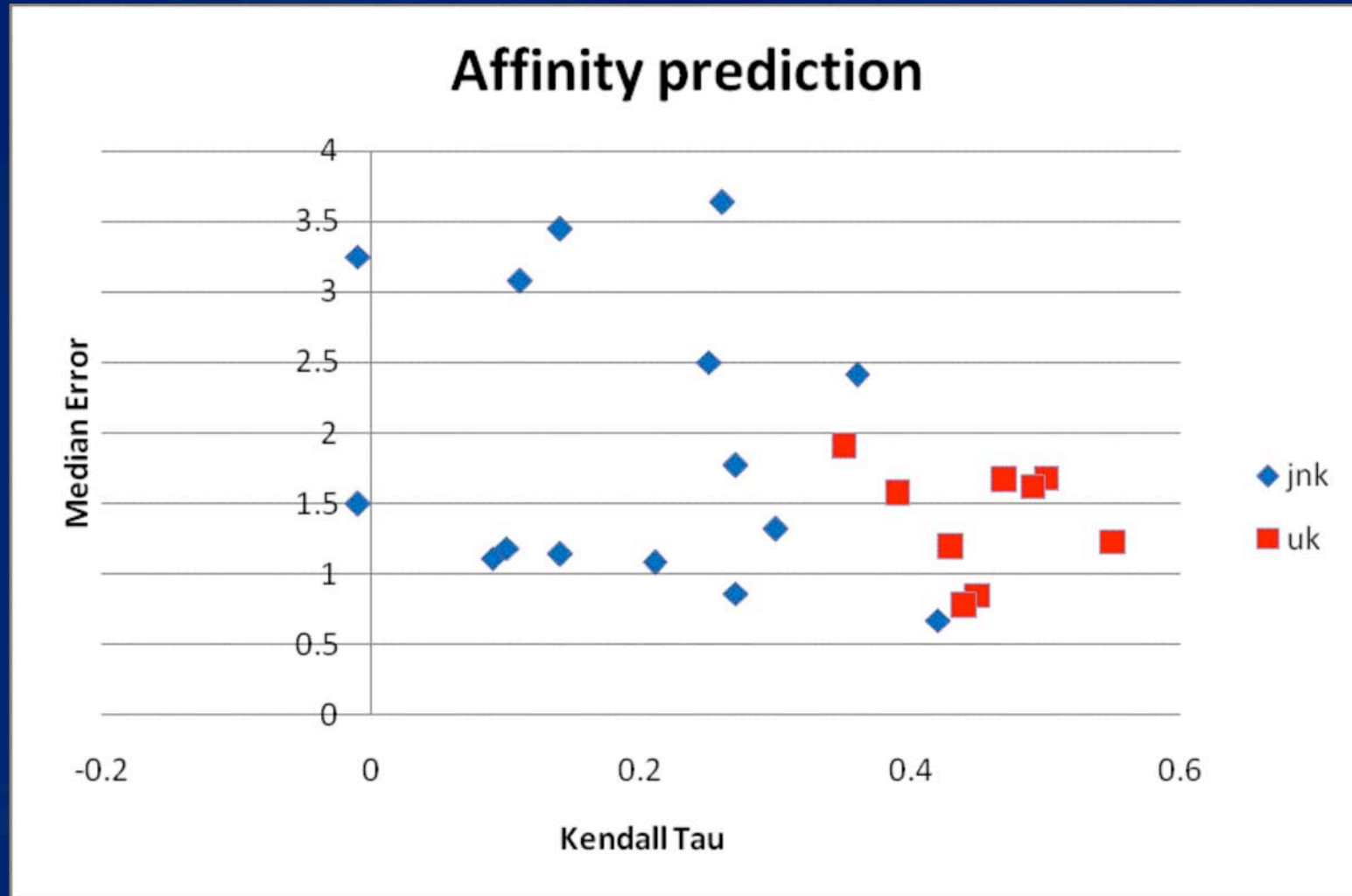


Affinity assessment

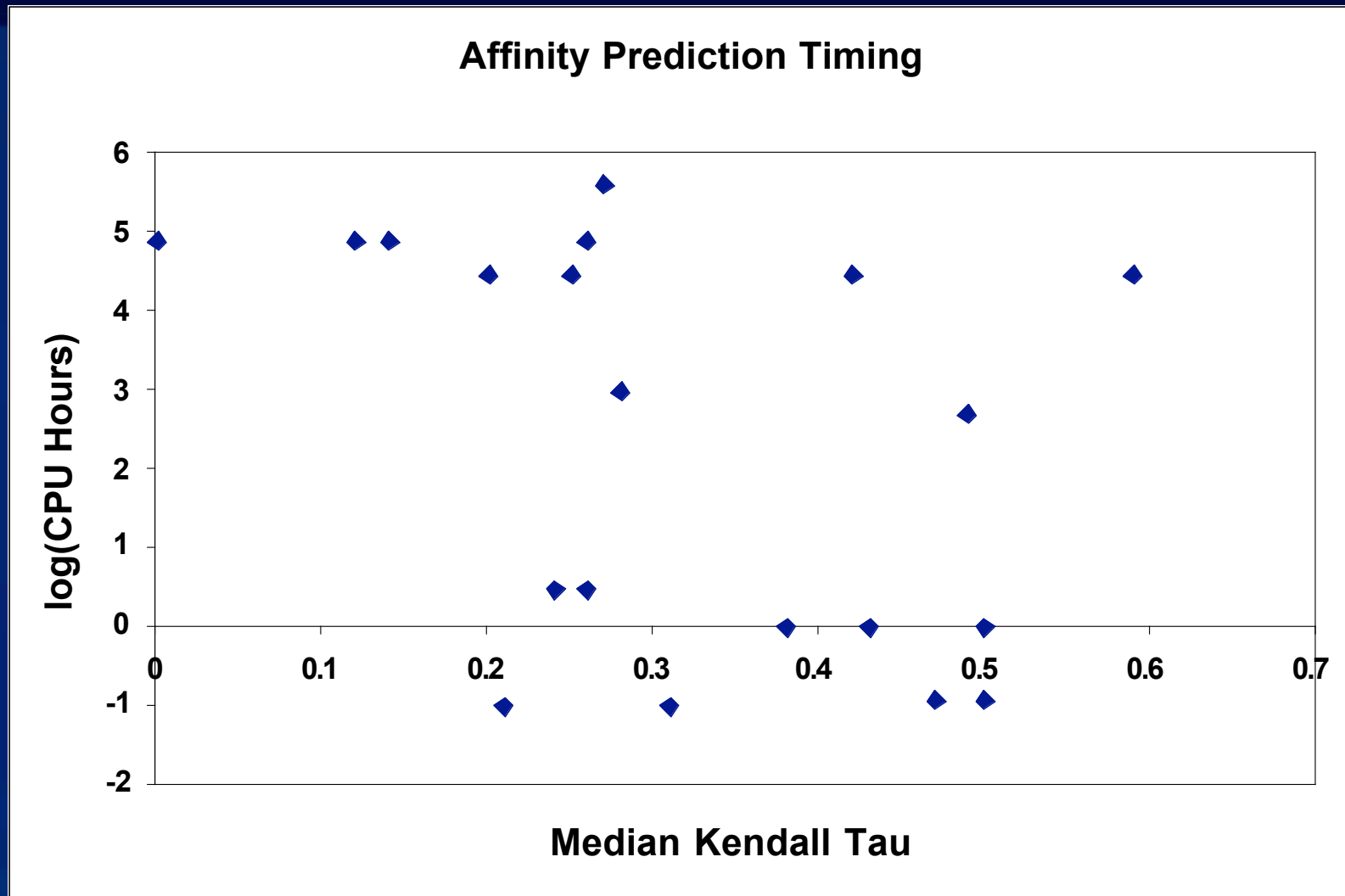
- 10,000 simulated experimental sets
 - Linear fit, Slope, Intercept, R^2
 - Rank Correlation - Kendall Tau
 - Median error
- Median value
- 95% bootstrap interval
- Experiment vs Experiment + Error
 - Upper bounds on performance



System dependent

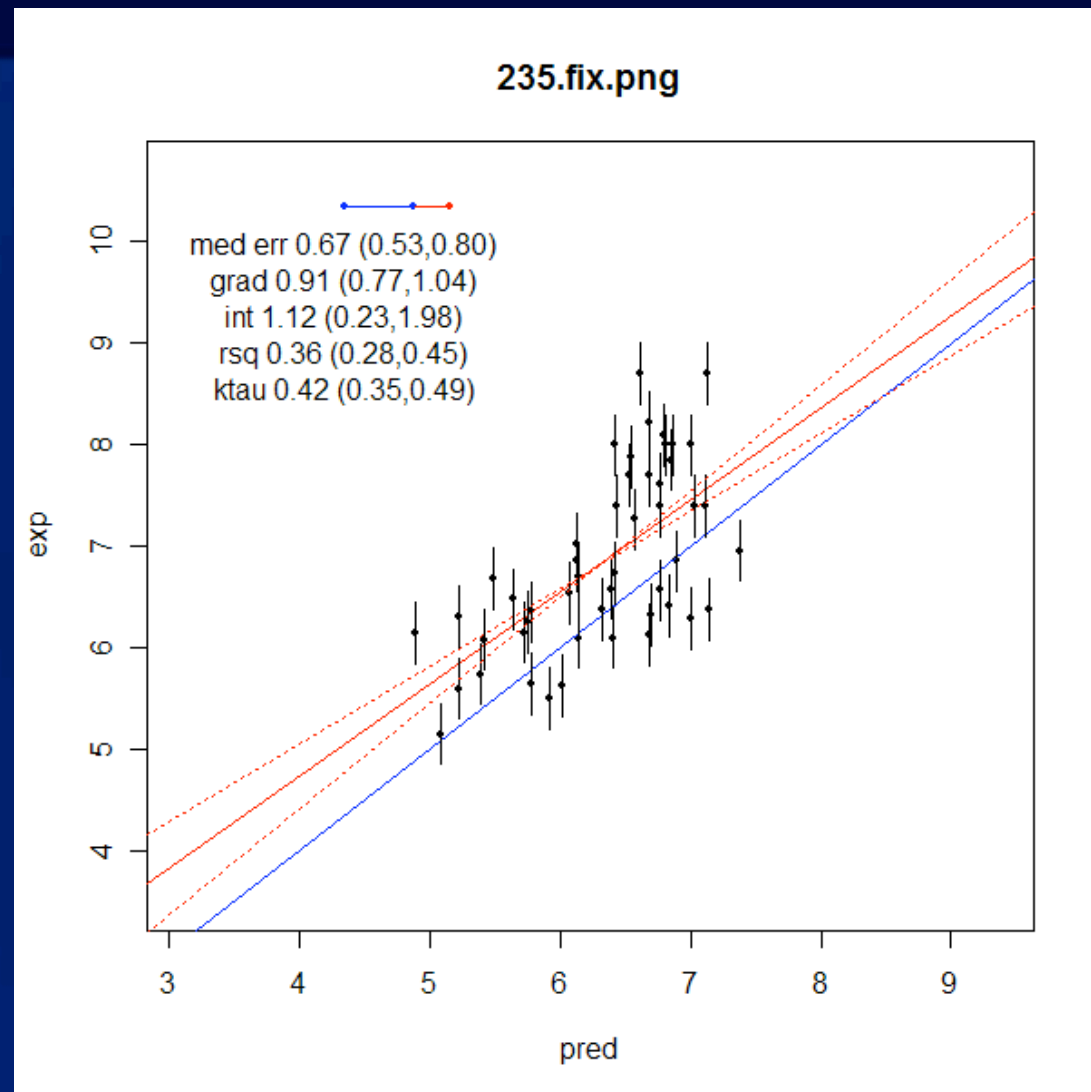


More is not always better.



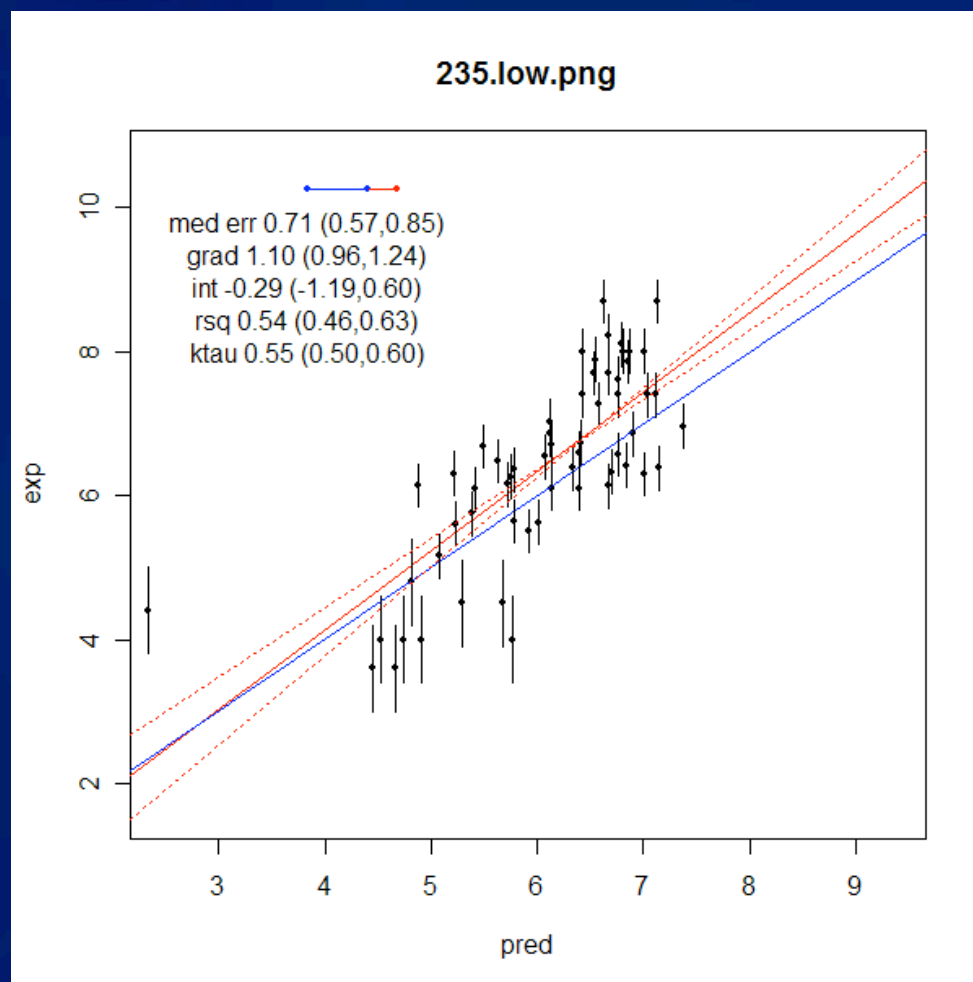
Getting some things right

- Scale
- Bias
- Slope
- Precision



Putting more pressure on system

- Spurious success
- Add lots of inactives
- 10 cpds
 - Selectivity studies
 - Poor data
 - K_i , IC_{50} , fuzzy data
- All $10 < 1\mu M$

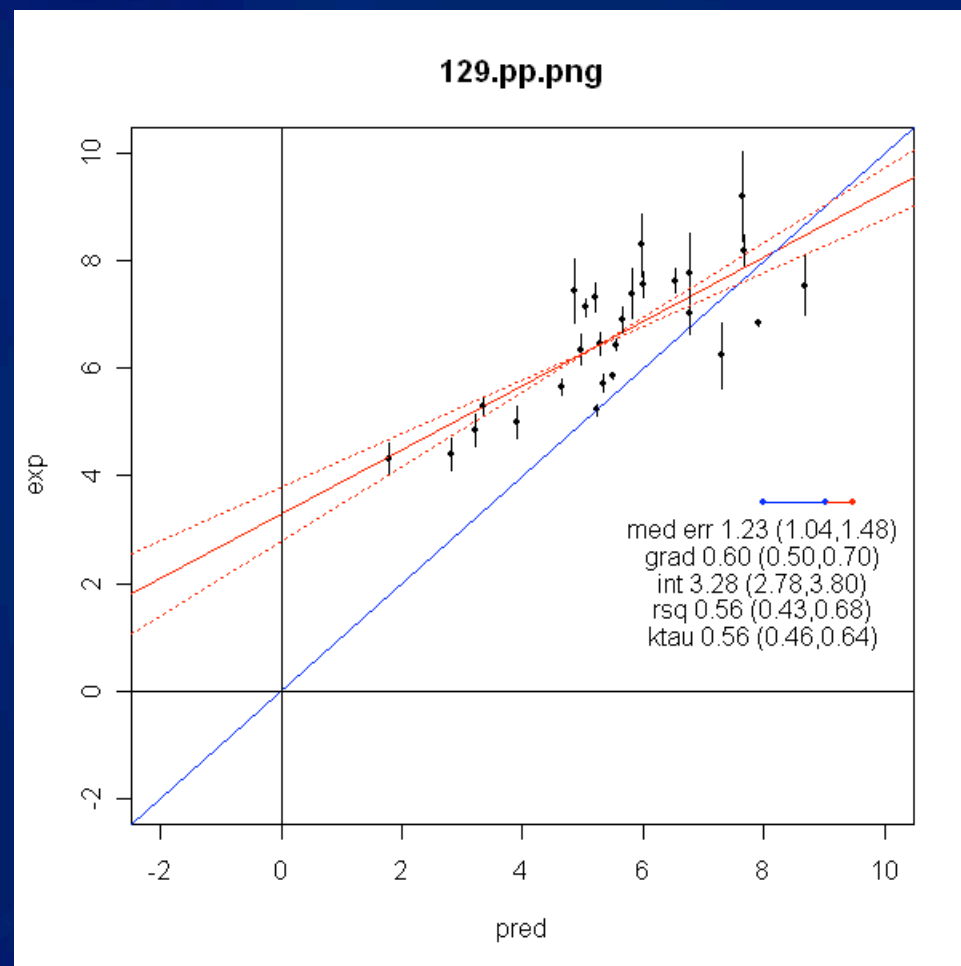


Putting more pressure on system

- Spurious success
- Add lots of inactives
- 10 cpds
 - Selectivity studies
 - Poor data
 - K_i , IC_{50} , fuzzy data
- All $10 < 1\mu M$
- 4 compounds
 - 20-30%
- R squared = 0.36
- Kendall tau = 0.42

Bold: QSAR-aided docking

- Cross-docking
- Monte-Carlo docking
- Flexible protein
- Physics-based enthalpy scoring
- PDBeBind Urokinase structures
- Dock & fit scores to those binding energies

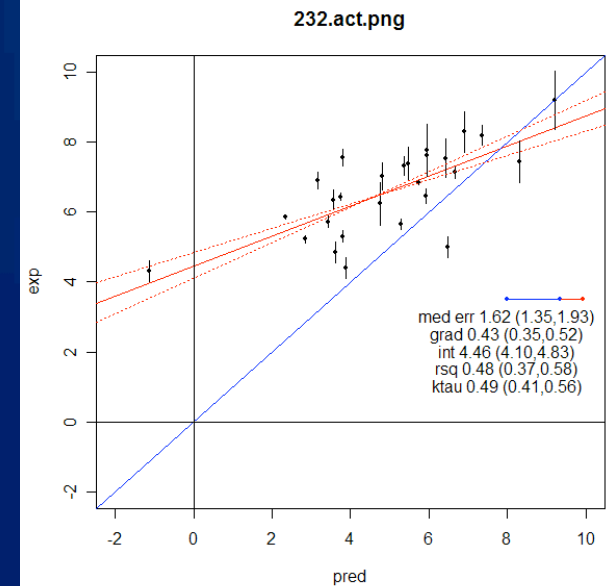
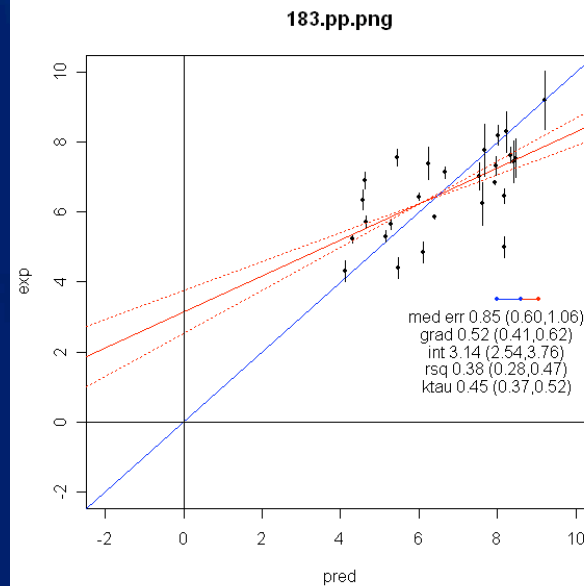
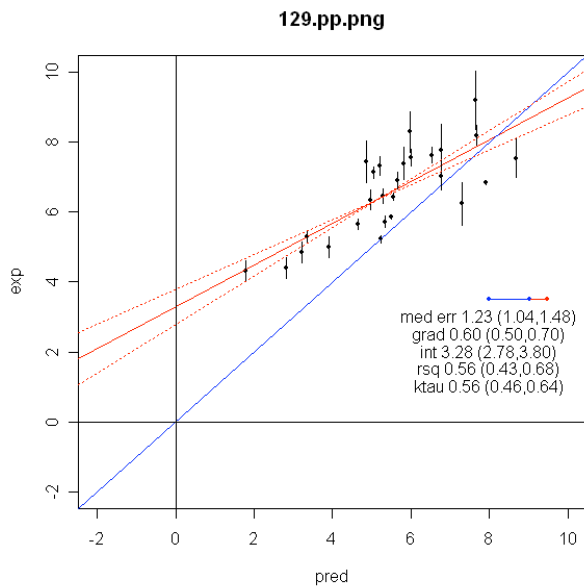


Self comparison on Urokinase

Cross-docking

Self-docking

Affinity



Kt,err=(0.56, 1.23)

(0.45, 0.85)

(0.49, 1.62)

rms-dpi = 2.31

1.43

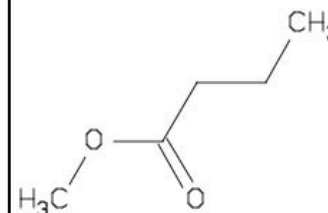
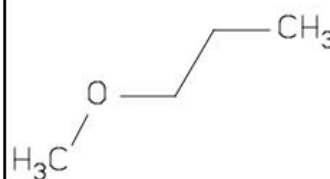
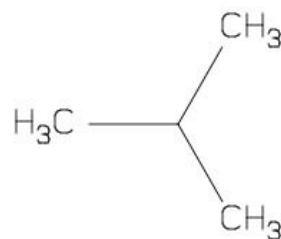
Transfer energies



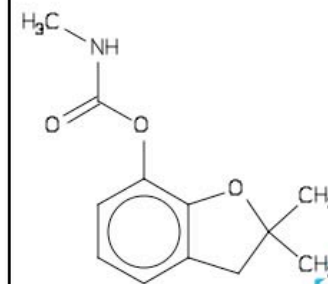
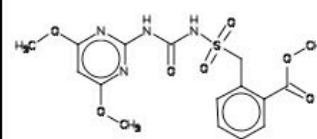
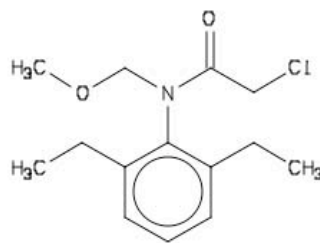
Transfer energy data

“Wow what a fantastic test set! A knock-down, drag-out, eye-gougin', ear-pullin', no-holds-barred test set!” -Chris Bayly

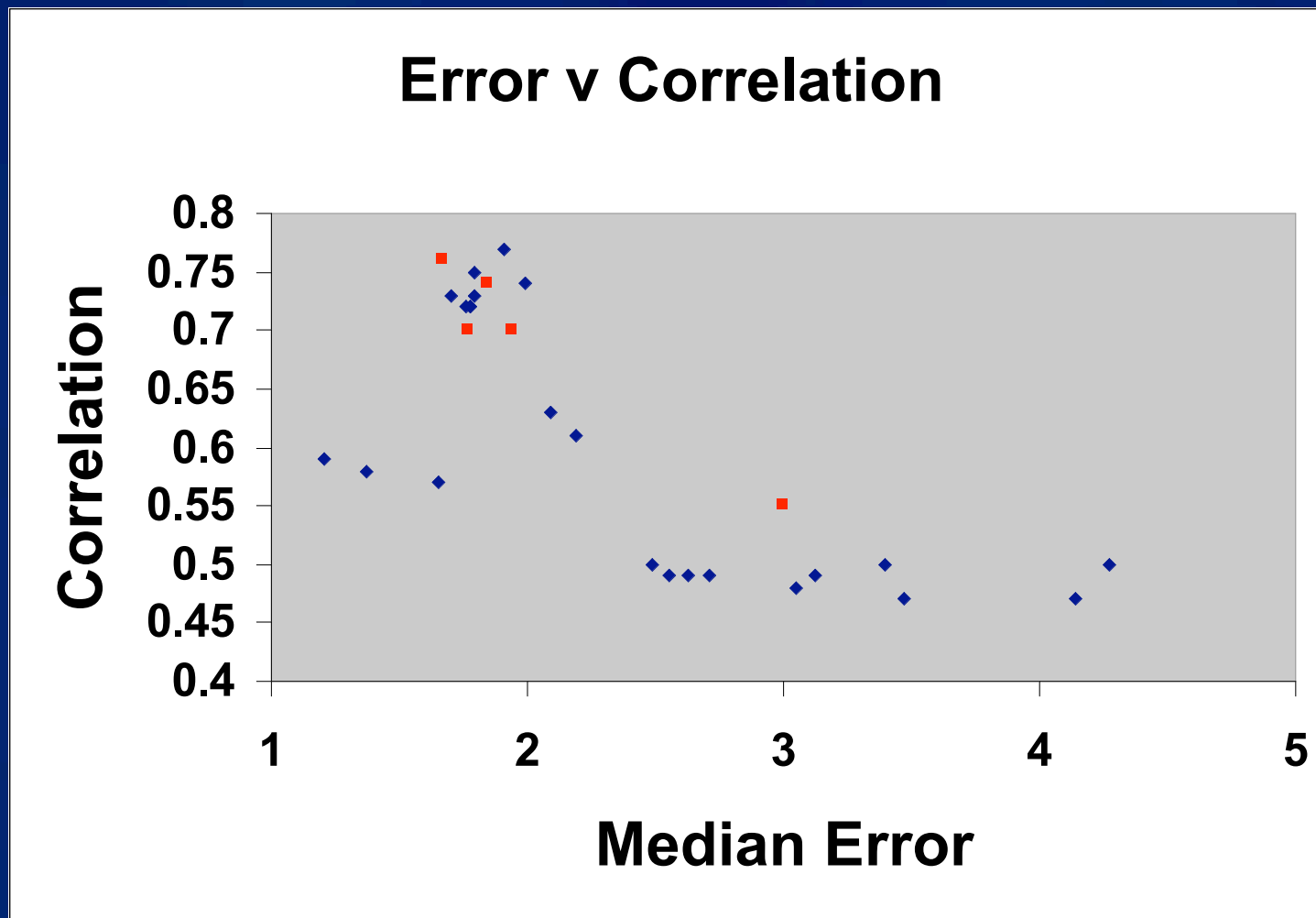
Typical data:



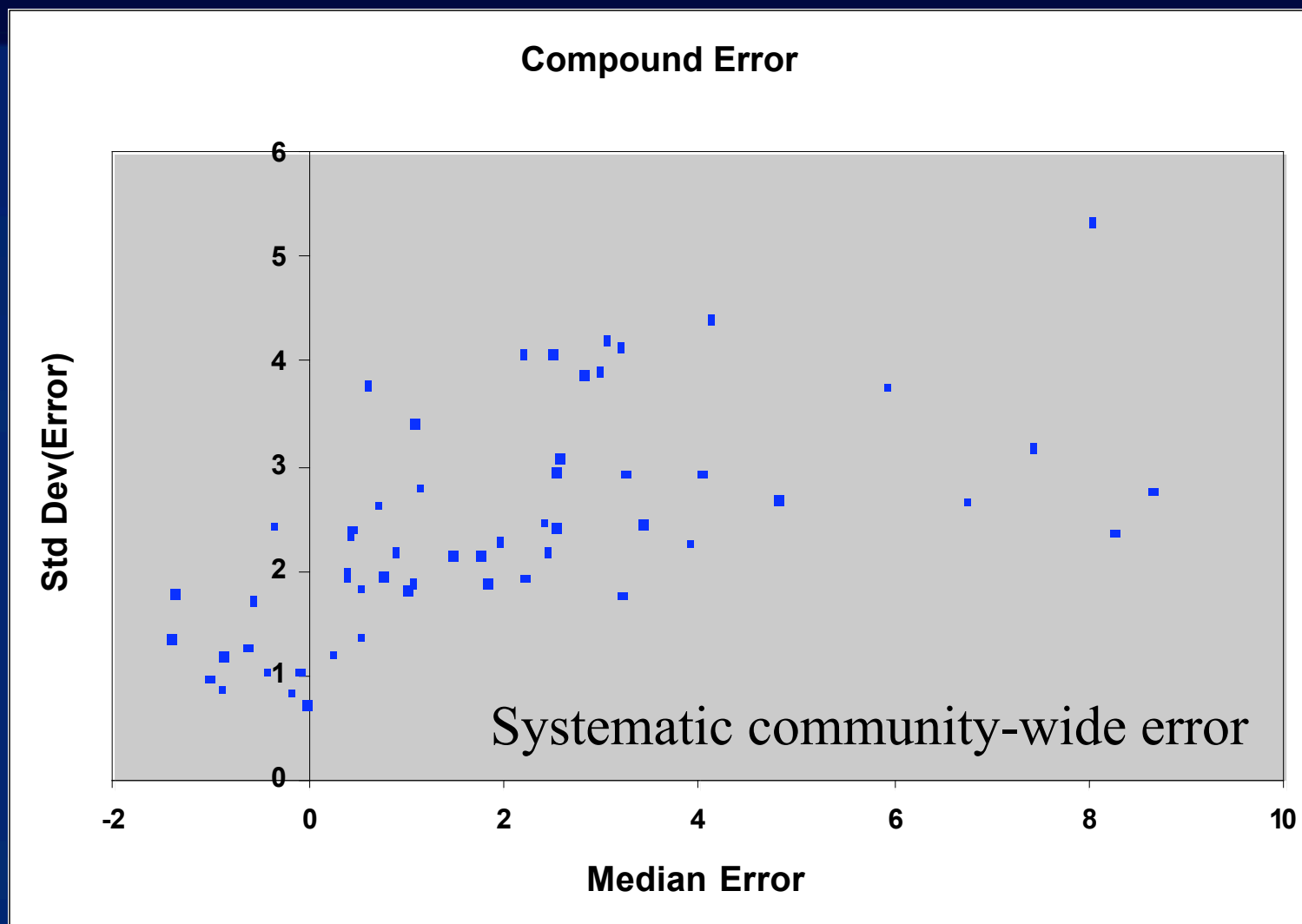
Peter Guthrie data:



Explicit vs Implicit Solvent

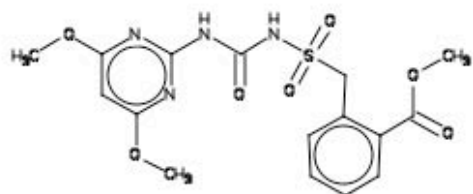


Per-compound Error Analysis

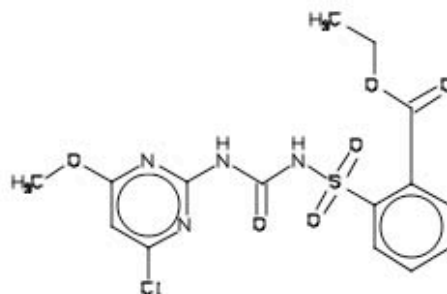


Difficult compounds

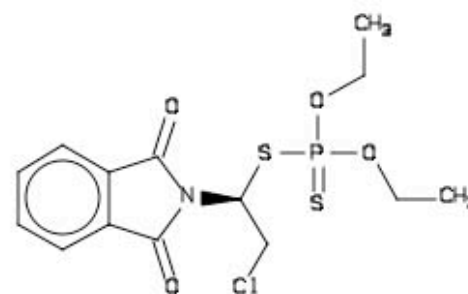
CUP08012.xyz



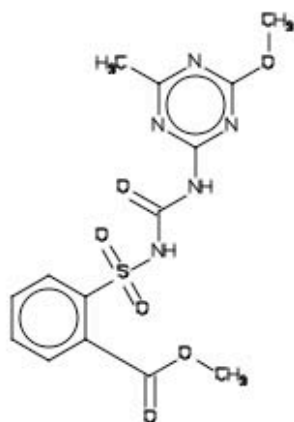
CUP08020.xyz



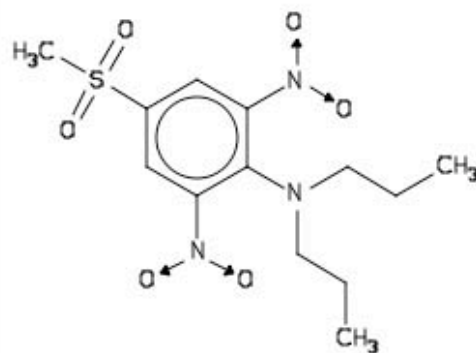
CUP08023.xyz



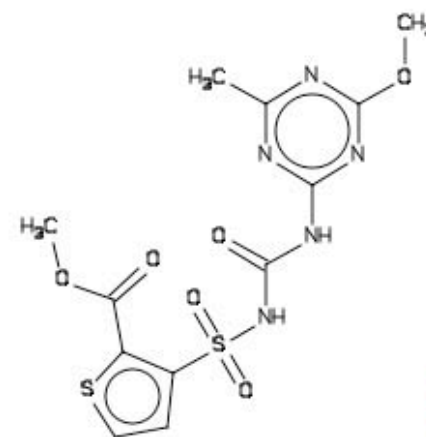
CUP08039.xyz



CUP08040.xyz



CUP08054.xyz



101



Conclusions & Conjectures

- It's an *evaluation*
- Active/Decoy choices matter for rank ordering VS methods as well as absolute difficulty
- $2 \ll 100$
 - But N methods also interesting
- Pose by hand & don't move the protein
- Affinity: There's more (to life) than MMPB/SA
- Xfer: Varied methods, same problems
- Not finished
- Even with lots of data, many points of interest are still anecdotal



Acknowledgements

- **Vertex Pharmaceuticals**
 - Pat Walters
 - Paul Charifson
- **Abbott Laboratories**
 - Steve Muchmore
 - Scott Brown
- **Peter Guthrie**
- **GSK**
 - Marti Head
- **Participants (present)**
 - Dave Mobley
 - Enrico Purisima
 - Eric Manas
 - Vijay Pande
 - Marti Head
 - Carleton Sage
 - Imran Haque
 - Matt Geballe
 - Roger Sayle
 - John Chodera
 - Guha Jayachandran
 - Anthony Nicholls
 - Stan Wlodek
 - Chris Bayly

