

**SimBioSys Inc.**

---

**Use of training techniques in eHiTS  
improves score-RMSD and score-IC50  
correlations in *in-silico* high throughput  
screening.**

Danni L. Harris, **Orr Ravitz**, Antony Williams,  
Zsolt Zsoldos

SimBioSys Inc.

## 2. Abstract

---

### **Abstract:**

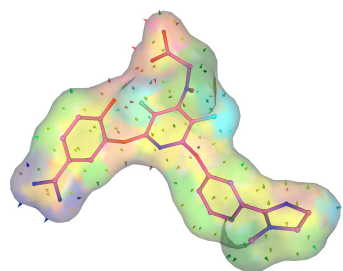
Screening of compounds has become a prevalent step in a variety of biologically related applications including drug discovery, predictive metabolism and toxicity prediction. The docking paradigm consists of two interrelated steps, namely pose prediction, and scoring. While most docking approaches are capable of predicting poses consistent with known structural solutions, they do not generally score these as the top ranking poses. Furthermore, correlations between docking scores and low RMSD values or bioactivity given in terms of  $\ln K_d$  or  $IC_{50}$  values are often limited (J. Med. Chem, 49:5912-5931). Such correlations are crucial if docking is to play a reliable role in either prioritizing prospective ligands for synthesis or in ranking protein-target interactions in metabolic and toxicity studies. We describe in detail the mixed physical and informatics approach of the scoring function of eHiTS (Electronic High Throughput Screening), and demonstrate by comparing to quantum mechanical results how it is well equipped to capture subtle interactions such as Pi-cation, non-conventional hydrogen bonding, and Pi-stacking. Good score-based low RMSD discrimination of biochemically and pharmacologically relevant poses as well as score- $IC_{50}$  correlations are shown with illustrations from several systems: nicotinic acetylcholine receptors and their surrogate binding proteins (AChBP), kinases, and cytochrome P450s. We demonstrate how these features may be further enhanced using eHiTS' unique training utility, and discuss this as a method to improve discrimination of ligand-target recognition.

# 3. The Problem

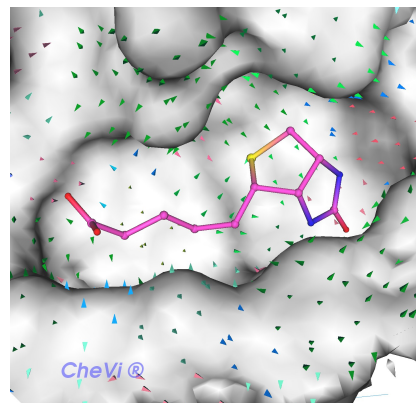
---

- The bound conformation of ligands in proteins (receptors) is a function of diverse protein-ligand interactions types such as:
  - Salt-bridge (electrostatic)
  - conventional and non-classical hydrogen bonds
  - $\pi$ - $\pi$  stacking
  - $\pi$ -cation
  - $\pi$ -hydrogen bonding-H
  - ion-induced dipole
  - dispersion
- While pose prediction with an accuracy of  $< 2.0 \text{ \AA}$  RMSD is not uncommon, achieving sub  $1\text{-\AA}$  RMSD is less typical.
- In many docking approaches, poor correlation between low RMSD poses and good docking scores is common even when pose prediction is of high performance.
- Correlations between docking scores and bioactivity or affinity are typically low.

# 4. Solution: A Trainable & Diverse Scoring Function Including Physics and Informatics Components.



CheVi®

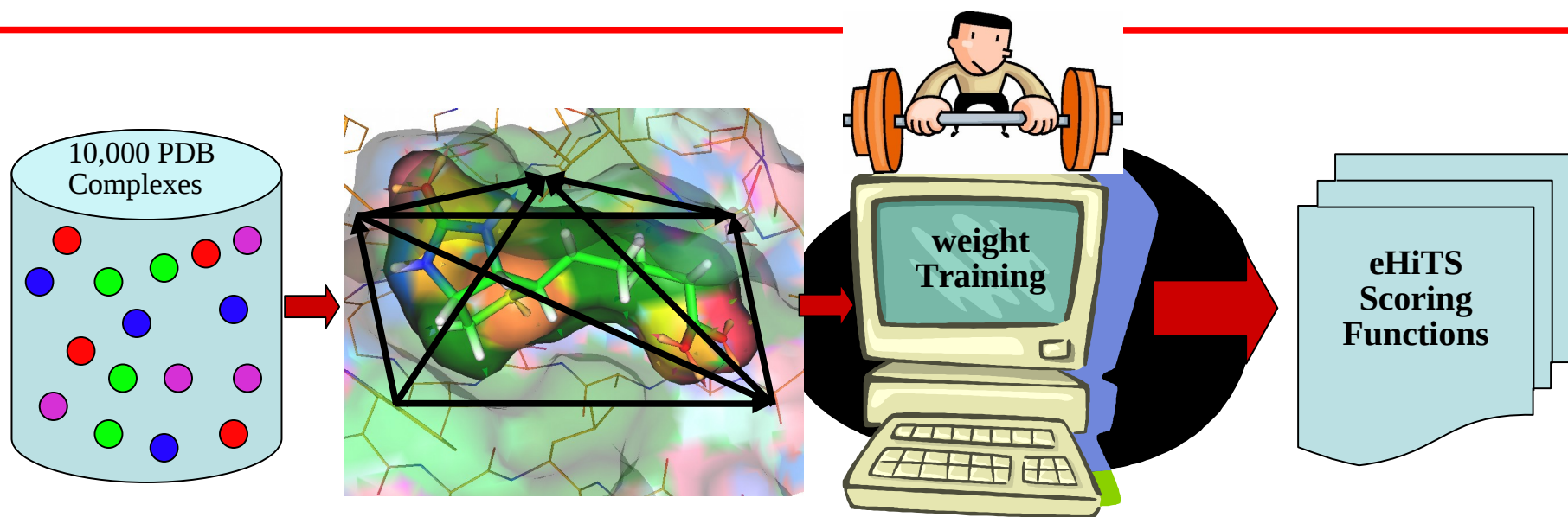


CheVi®

**23 Surface point types: all pairs considered => 23x23 matrix**

- METAL
- CHARGED\_HPLUS
- PRIMARY\_AMINE\_HLP
- HDONOR
- WEAK\_HDONOR
- CHARGED\_LONEPAIR
- ACID\_LONEPAIR
- LONEPAIR
- AMBIVALENT\_HLP
- ROTATABLE\_H
- ROTATABLE\_LP
- WEAK\_LONEPAIR
- PI\_SP2\_POLAR
- PI\_SP2\_CARBON
- HALOGEN
- SULFUR
- HYDROPHOB
- H\_AROM\_EDGE
- WS\_LIPO
- NEUTRAL
- PI\_AROMATIC
- PI\_RESON\_POLAR
- PI\_RESON\_CARBON

# 5. Family Training



**1.** 10,000 PDB complexes chosen to represent a wide range of protein families

**2.** Complexes are clustered automatically into ~500 protein families plus one default, global set based on amino acid ligand-component distance metrics.

**3.** eHiTS training utility optimizes scoring functions (weights) for each family

**4.** Scoring functions for each family are output and used as default scoring functions of eHiTS

# 6. Result: Interaction-scoring Matrix.

Re/Li	DonH+	Amine	Don-H	PO3--	AcidL	AccLp	WS-Lp	Ambiv	Rot-H	RotLp	CLipo	AromH	WSlip	Neutr	AromP	Res+	Res_C	Sp2+	Sp2_C	Halog	Join.	
METAL	-9.99	1.78	0.02	2.18	2.12	0.57	0.43	5.54	0.32	0.12	-4.8	-4.18	-3.19	0.27	-4.11	0.15	-1.49	0.98	-1.71	-6.46	4.39	METAL 0
DonH+	-5.15	-6.13	-5.38	2.57	3.8	1.14	-0.49	-0.79	-2.52	0.1	-2.95	-1.86	-1.23	-3.04	-5.31	-0.51	-1.69	-6.4	-6.6	-1.71	-1.46	DonH+ 1
Amine	-7.94	0.24	-5.21	2.13	1.37	1.47	-0.4	-0.75	0.3	1.43	-3.64	-1.72	-0.85	-5.5	-3.87	-0.75	-3.32	-1.94	-3.78	-1.89	0.91	Amine 2
Don-H	-9.99	-0.26	-1.78	2.86	2.18	3.27	2.36	0.49	-0.59	1.78	-1.7	-1.9	-0.72	-0.15	-3.95	0.01	-2.32	-2.68	-3.27	-0.65	-1.21	Don-H 3
WSdon	-9.99	-2.93	-5.56	-0.21	-0.16	-0.09	2.83	0.64	0.12	0.62	-0.64	-1.29	-0.97	-0.12	-0.74	-0.79	-2.48	-3.36	-3.21	-0.37	-0.53	WSdon 4
PO3--	-0.92	1.26	2.86	-0.72	-1.31	-1.7	-1.08	-0.58	3.77	-0.52	-1.28	-1.64	1.34	0.2	-4.65	-0.72	-0.65	-1.53	-3.57	-3.52	-0.89	PO3-- 5
AcidL	3.58	3.11	2.34	-0.66	-1.64	-0.72	-3.87	0.52	3.94	0.37	-0.99	-1.65	2.24	0.23	-5.04	-0.92	-2.45	0.12	-0.13	-0.86	-0.76	AcidL 6
AccLp	3.05	1.67	3.09	-3.8	-2.39	-1.7	-2.98	-0.01	0.51	-2.26	-0.29	0.45	0.6	0.8	-3.51	-2.42	0.14	-1.07	-0.97	-1.41	0.12	AccLp 7
Ambiv	-3.31	-0.98	2.1	3.02	1.41	0.9	0.65	4.8	1.08	1.63	-1.94	0.09	0.03	0.74	-2.79	-0.46	-2.73	1.23	-4.65	-1.02	0.68	Ambiv 9
Rot-H	-0.45	-9.99	0.24	3.15	3.78	0.89	-0.24	2.47	-4.02	-0.2	-0.61	0.05	0.19	0.54	-4.43	0.53	0.24	-3.71	-9.99	0.12	-1.09	Rot-H 10
RotLp	3.75	-1.25	2.63	0.01	0.06	-2.09	-4.05	3.75	-0.01	-2.46	-0.65	0.18	-0.98	1.34	-6.19	-1.37	-0.83	-0.79	-2.15	-1.46	1.35	RotLp 11
CLipo	-5.5	-3.79	-2.97	-2.53	-2.48	-0.87	-0.46	-1.75	-1.66	-1.25	0.83	0.78	-0.12	-0.02	1.91	0.34	0.59	0.06	1.27	1	-1.27	CLipo 12
AromH	-9.22	-3.12	-3.76	-1.88	-1.33	-1.11	0.11	-2.76	-1.94	-1.47	-0.01	0.61	-0.14	-0.11	1.28	1.08	1.87	0.03	0.4	0.39	-2.58	AromH 13
WSlip	0.01	-0.26	-1.25	-0.09	-1.07	0.12	0.16	-0.14	-1.05	-0.7	-0.09	0.27	-0.11	-0.1	1.58	2	1.62	0.35	0.62	0.76	-0.84	WSlip 14
Neutr	-9.99	-2	-0.44	0.03	0.2	-0.64	0.67	-0.63	-0.06	0.57	-0.38	-0.14	-0.68	-0.13	-0.27	0.42	0.47	0.81	-1.08	-0.26	-0.33	Neutr 15
AromP	-9.99	0.23	-3.21	-6.67	-4.18	-2.75	-1.83	0.14	-2.14	-1.67	3.61	3.12	3.75	3.29	4.88	3.14	4.56	4.6	3.66	2.61	0.79	AromP 16
Res+	-1.56	0.16	0.46	-1.96	-2.88	-2.12	-1.42	-0.52	0.02	0.21	2.86	3.09	2.54	2.97	4.05	3.35	5.08	4.69	3.84	4.22	-4.38	Res+ 17
Res_C	-4.02	-2.1	-1.04	-2.78	-4.97	-1.94	-3.65	-2.12	-1.71	0.38	2.49	3.37	2.05	2.5	4.86	2.78	3.85	1.4	3.1	4.14	-1.28	Res_C 18
Sp2+	-9.99	-9.04	-3.58	-2.43	-1.62	-2.64	-0.48	-2.43	-0.78	0.11	2.38	2.79	1.2	1.03	5.25	2.36	4.17	2.7	3.58	4.32	-2.88	Sp2+ 19
Sulfu	-0.03	-0.43	-1.03	-3.09	-3.24	-3.65	-0.98	-9.99	-1.24	-3.38	0.03	1.83	0.09	0.36	1.21	1.29	0.94	-0.16	1.98	0.07	1.04	Sulfu 22
Re/Li	DonH+	Amine	Don-H	PO3--	AcidL	AccLp	WS-Lp	Ambiv	Rot-H	RotLp	CLipo	AromH	WSlip	Neutr	AromP	Res+	Res_C	Sp2+	Sp2_C	Halog	Join.	

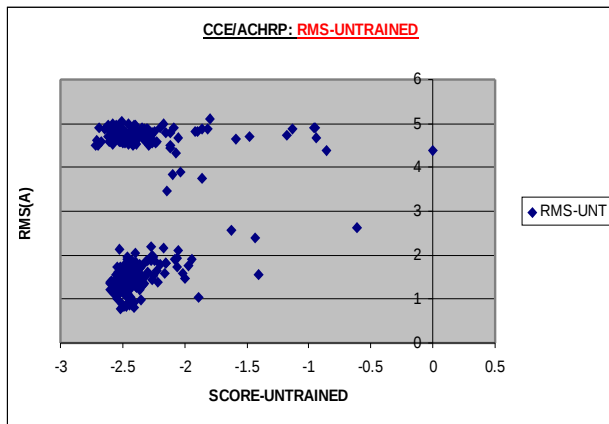
polar/electrostatic

$\pi$ - $\pi$  interactions

$\pi$ /aromatic--cation/H+

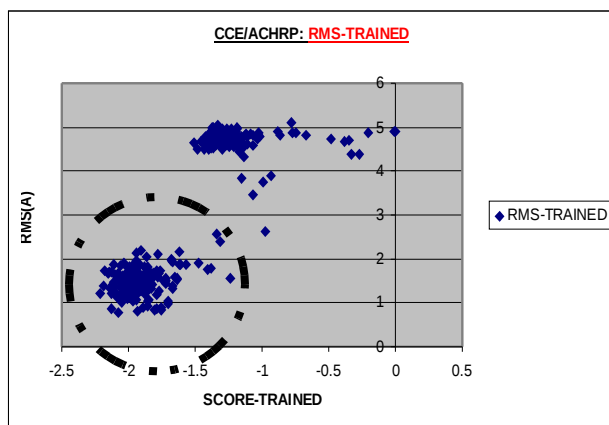
hydrophobic

# 7. Family Training Leads to Improved Pose Prediction Accuracy and Score-RMSD correlation



## Untrained Scoring:

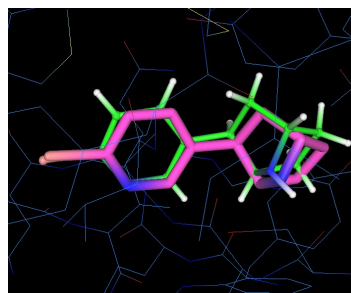
Many low RMSD solutions that cannot be distinguished from high RMSD ones.



## Trained Scoring:

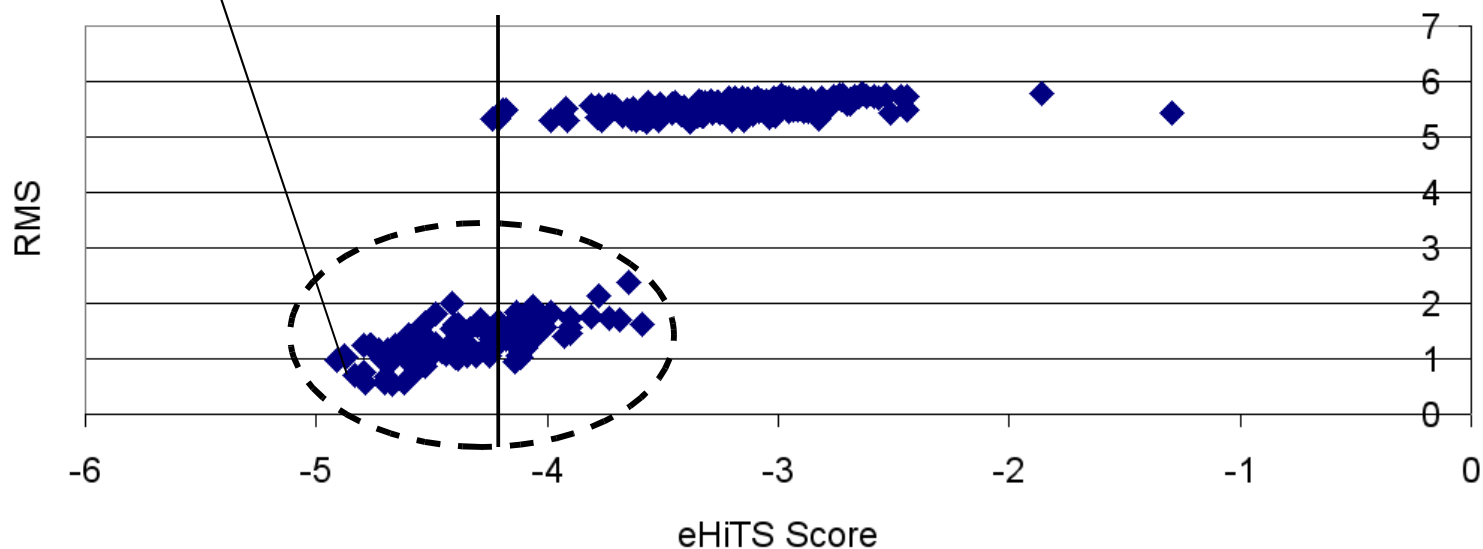
Dramatic score-separation of the 'correct' pose RMSD-regime (circled).

# 8. Best Docking Score Correlates with low RMSD pose reproduction.



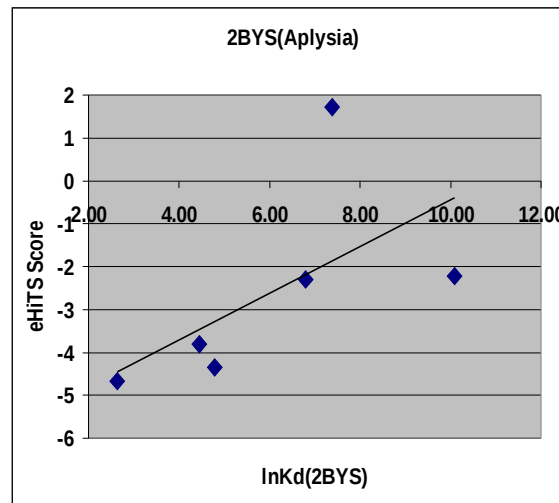
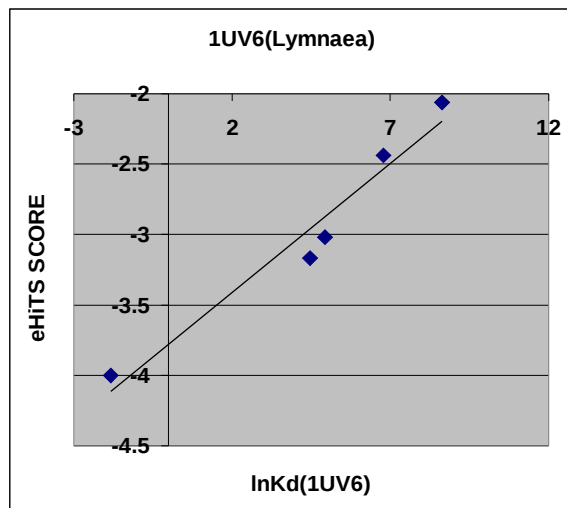
xray  
docked pose

Epibatidine/2BYQ-  
Score-RMS Separation: trained-Scoring FUnction



Low scored region is significantly enriched with low RMSD poses.

## 9. Family Trained Scoring Functions Show Score- ln(IC50) correlations in AChBP.



- Two cases of an acetylcholine binding surrogate protein(AChBP) show fair monotonic correlations between ligand docking scores and affinity.
- Score - log(Kd) correlation allow using the virtual screening scores to prioritize synthesis or in-vitro testing of compounds.

# 10. Good Pose Prediction in Proteins Outside the Default Trained Sets

**Table 1. Initial Survey of Targets and Native Ligands.**

Target	PDB code	Ligand Code (Abbr)	Closest pose		Top Ranked RMSD (Å)
			RMSD (Å)	rank	
FAAH	1mt5-S	Methyl Arachidonyl Fluorophosphonate(MAY)	1.45	7/60	9.74
Adenosine Kinase	2i6b-S	89I: Acetylinic inhibitor	1.01	15/60	1.48
Glutamate Carboxypeptidase	2pvw	G88: 2-PMPA	0.87	4/60	5.20
	2or4	QUS: QUISQUALATE	0.53	2/60	1.77
	2pvv	OSE: O-SULFO-L-SERINE	0.88	36/60	1.65
	3bhx	BHX:	1.00	5/60	1.39
	3bi1-S	3BI: Methotrexate-GLU	0.75	11/60	1.43
Aurora Kinase A	2np8-S	CC3	0.64	17/60	1.00
MMP-2/13	1g05-S	BBH	1.41	18/60	4.04
GSK-3B	1q41-S	IXM	0.92	3/60	2.47
nNOS	1vag-S	ARR:Thiophecarboxamidine	1.20	6/60	1.40
eNOS	3nos-S	ARG	1.04	8/60	6.30
	4nos	ITU	0.64	1/60	0.64
	1m9m-S	INI	0.73	2/60	4.76
iNOS	1nsi-S	INI-ARG	0.52	1/60	0.52
DDAH1	2jaj-S	D20	1.08	7/60	1.74
β2-adrenergic	2rh1-S	(S)-CARAZOLOL	0.81	1/60	0.87

Default scoring function generates good poses and ranking for proteins in families that were not included in the training of eHiTS' default scoring function.

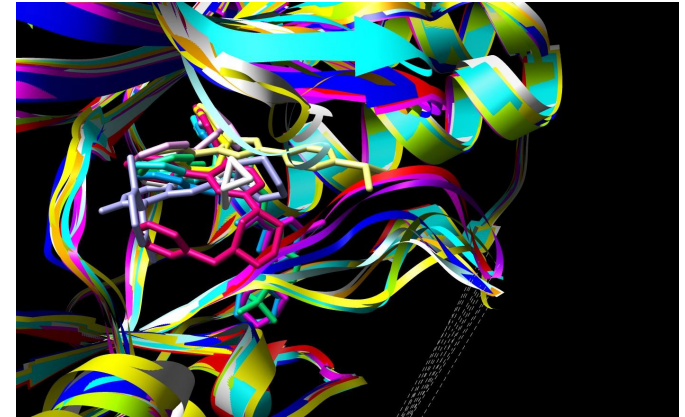
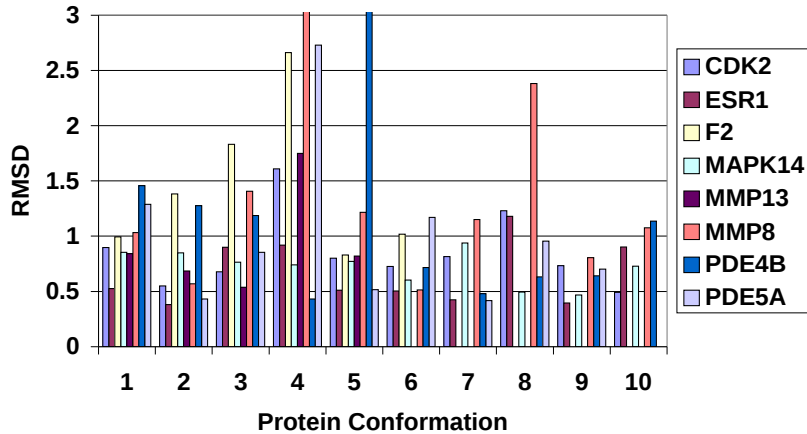
# 11. Exploring the Cross-Ligand-Protein Landscape in DeNovo Design and Predictive Toxicology.

	MMP-1g05	eNOS-1M9M	FAAH	iNOS	GSK	nNOS	AK:2I6B	DDAH1	AuroraKin	B2Adren	GlutCarb	eNOS:3NOS	eNOS:4NOS
MMP:1G05	<b>-5.36</b>	-1.13	-5.89	0.37	-6.75	-1.59	-3.56	-2.19	-0.56	-4.04	-0.3	0.32	
eNOS:1M9M	-3.5	<b>-4.28</b>	-3.25	-2.94	-3.76	-4.21	-3.87	-2.5	-2.19	-2.74	-3.74	-3.36	-2.19
FAAH	-2.13	-0.74	<b>-5.07</b>	-0.88	-5.48	-4.27	-3.39	-3.21	-1.94	-0.61	0.42	-1.36	-1.19
iNOS	-3.47	-5.03	-3.31	<b>-4.75</b>	-3.18	<b>-4.98</b>	-3.31	-4.6	-2.76	-3.17	-3.79	-4.7	<b>-4.41</b>
GSK	-5.75	-3.36	-5.15	-2.85	<b>-7.2</b>	-3.5	-5.7	<b>-4.83</b>	-2.13	<b>-4.36</b>	-1.65	-2.99	
nNOS	-4.93	-3.96	-4.75	-4.23	-4.49	<b>-5.89</b>	-4.73	-4.3	<b>-4.36</b>	-2.85	-1.88	-3.53	-0.52
AK:89i	-3.02	-2.97	-2.38	-2.79	-5.84	-2.14	<b>-3.38</b>	-0.5	<b>-4.11</b>	-3.8	-1.59	-1.17	
DDAH1	<b>-5.83</b>	-4.69	-3.54	-4.72	-4.36	<b>-5.28</b>	-3.52	<b>-4.59</b>	-3.2	-2.44	-3.44	-3.49	-1.97
AuroraKin	-4.69	-5.46	<b>-6.58</b>	0.77	-5.68	-3.62	<b>-7.21</b>	-4.14	<b>-2.57</b>	-2.99	-1.95	-4.05	
B2Adren	-4.1	-2.28	<b>-5.64</b>	-0.65	<b>-5.81</b>	-3	<b>-6.48</b>	-3.63	-3.87	<b>-5.69</b>	-0.07	-1.89	-1.71
GlutCarb	-1.7	1.2	-0.68	1.25	-3.44		-3.57	-1.06	-1.44	-1.01	<b>-8.58</b>	0.82	
eNOS:3NOS	-3.44	-4.3	-3.33	-4.16	-3.32	-5.15	-3.38	-3.71	-3.1	-2.78	<b>-5.22</b>	<b>-3.95</b>	-3.15
eNOS:4NOS	-2.87	<b>-4.49</b>	-2.6	<b>-5.34</b>	-3.55	-5.5	-3.2	<b>-4.22</b>	-2.14	-2.6	-1.87	<b>-4.73</b>	<b>-5.14</b>

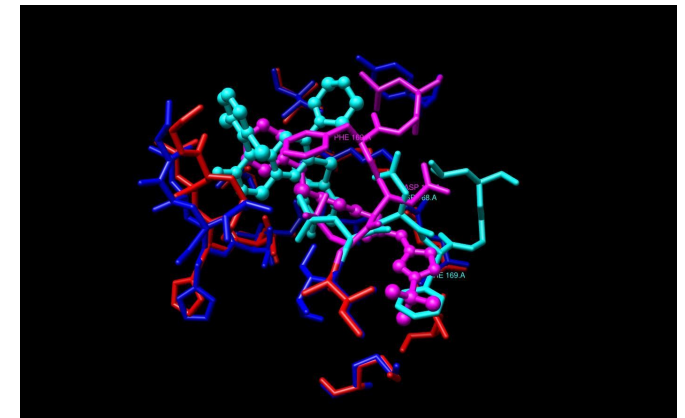
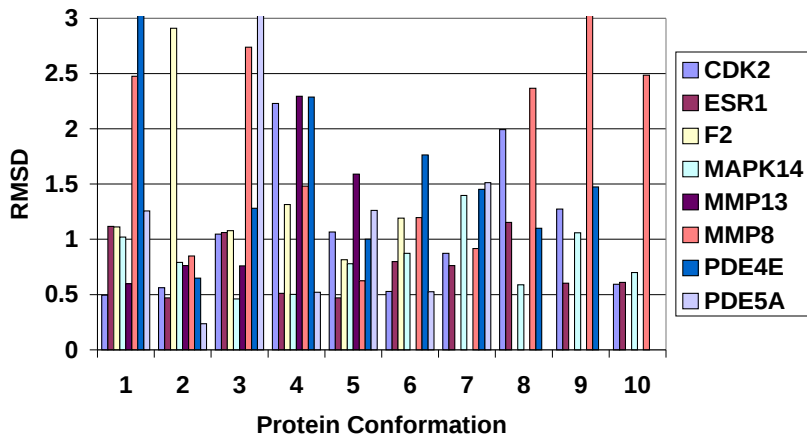
- Cross docking scores: each of the 'native' ligands in a particular protein family was docked against the other receptors.
- The score for each ligand against its original target (diagonal elements) is amongst the 'best' scores.
- eHiTS' score-log(Kd) correlations allow specificity evaluation.

# 12. Flexible Proteins: Family Trained Scoring Yields Conformation-Dependent Low RMSD Poses

Protein Conformational Sensitivity  
RMSD eHiTS 6.2



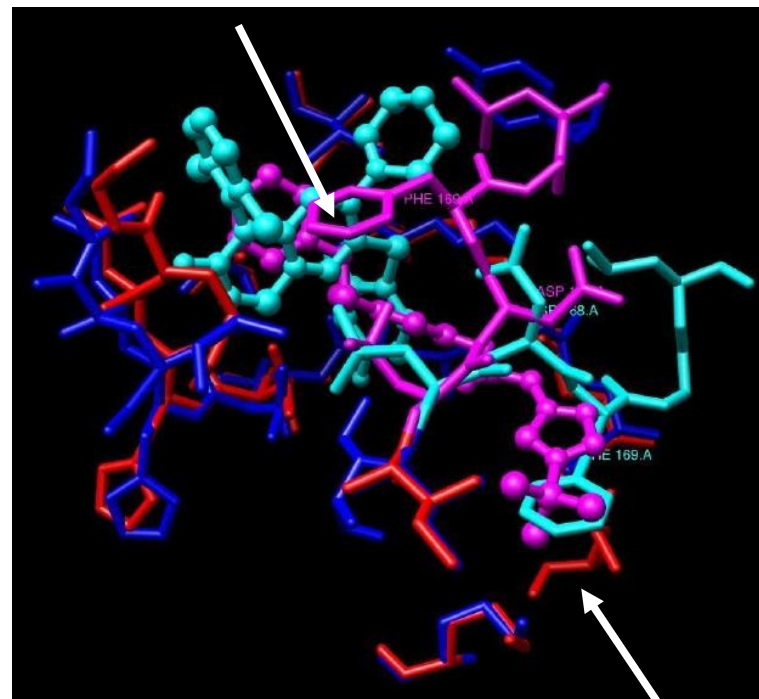
Protein Conformational Sensitivity  
RMSD (eHiT Lighting)



### 13. The eHiTS score 1) low RMSD, 2) InKd correlations provide a firm foundation for examination of the energy landscape for ligand libraries for flexible protein targets.

PROTEIN-CONF	CONF1	CONF2	CONF3	CONF4	CONF5	CONF6	CONF7	CONF8	CONF9	CONF10
LIGAND										
0	-8.9	-4.7	-6.4	-8.7	-7.1	-7.9	-7.9	-6.2	-8.1	-6.3
1		-5.6	-5.6	-6.4	-4.8	-10.2	-10.2		-8.4	
2		-3.6	-4.3	-7.4	-1.0	-4.1	-4.1	-0.9	-3.3	-1.4
3	-7.6	-3.2	-5.7	-8.1	-4.9	-8.5	-8.5	-5.0	-7.2	-4.8
4	-6.2	-3.6	-5.4	-5.6	-4.9	-7.5	-7.5	-6.3	-6.0	-3.8
5		-4.9	-5.2	-8.1	-4.7	-5.6	-5.6	-5.3	-5.6	-4.7
6	-5.4	-4.4	-5.0	-6.3	-3.7	-4.9	-4.9	-4.0	-3.7	-5.4
7	-4.2	-3.7	-5.6	-7.6	-4.7	-8.1	-8.1	-5.9	-6.7	-5.6
8	-5.9	-3.3	-3.5	-4.7	-3.3	-4.4	-4.4	-3.9	-5.2	-3.1
9	-5.8	-4.6	-3.9	-5.0	-4.1	-6.2	-6.2	-5.2	-5.9	-3.5
10	-6.0	-4.0	-3.8	-5.8	-4.4	-5.7	-5.7	-5.6	-5.0	-4.9
11	-6.2	-4.7	-5.2	-7.1	-4.2	-7.3	-7.3	-5.0	-5.7	-5.2
12	-6.4	-6.4	-6.4	-8.3	-6.8	-8.2	-8.2	-5.6	-6.8	-5.5
13	-5.8	-4.8	-5.3	-7.8	-4.7	-6.0	-6.0	-2.6	-6.2	-4.4
14		-4.1	-4.1	-5.7	-1.8	-7.8	-7.8	-6.3	-4.6	-4.3
15		-2.7	-4.1	-4.3	-6.0	-8.7	-8.7		-7.1	-0.3
16	-5.3	-3.1	-5.5	-5.8	-3.8	-6.1	-6.1	-2.0	-4.9	-2.6
17	-3.8	-3.7	-3.2	-3.2	-3.4	-4.0	-4.0	-3.2	-3.7	-3.8
18	-3.6	-4.9	-4.1	-8.1	-5.6	-5.4	-5.4	-3.1	-6.4	-4.2
19		-4.9	-5.4	-6.3	-6.2	-6.9	-6.9	-3.3	-6.7	-1.8
20	-2.3	-5.1	-5.8	-7.6	-3.9	-5.8	-5.8	-4.0	-5.7	-3.9
21	-6.3	-4.1	-5.5	-7.6	-4.9	-7.7	-7.7	-6.8	-6.3	-5.5
22	-3.9	-4.0	-4.7	-5.7	-4.3	-6.0	-6.0	-5.9	-6.4	-4.1
23	-4.6	-2.7	-4.1	-4.2	-3.6	-5.1	-5.1	-3.8	-4.1	-3.8
24	-1.7	-5.4	-5.6	-6.9	-4.7	-7.2	-7.2	-2.9	-6.8	-4.9

Color coded score profile of 25 ligands against 10 different conformations of the MAPK14 target highlight the sensitivity of affinity of the ligand to protein conformation



Two crystal structures of MAPK with native ligands showing large conformational changes and preferred ligand occupancy  
 Conf. 1: red protein/magenta-ligand  
 Conf. 2: blue protein/cyan ligand

# 14. Conclusions.

---

- The eHiTS scoring function captures a very diverse spectrum of ligand-protein interaction types and gives:
- Good score-Low RMSD correlations in pose prediction.
- Good score-Observed  $\log(K_d)$  correlations.
- The scoring function provides a firm basis for an initial screen of the ligand/protein-target energy landscape in cross docking profiles of diverse ligands and targets in HTS or DeNovo design.
- The initial examination of the eHiTS pose prediction accuracy for flexible protein paradigms indicates promising performance.
- The HTS docking and scoring in eHiTS Lightning gives comparable accuracy to eHiTS 6.2 with docking rates of 10 seconds per ligand on a \$400 PS3 platform.