

# Using Structural and Receptor-Site Similarity to Generate New Matter Ideas

Dr. Steven Muskal

Chief Executive Officer

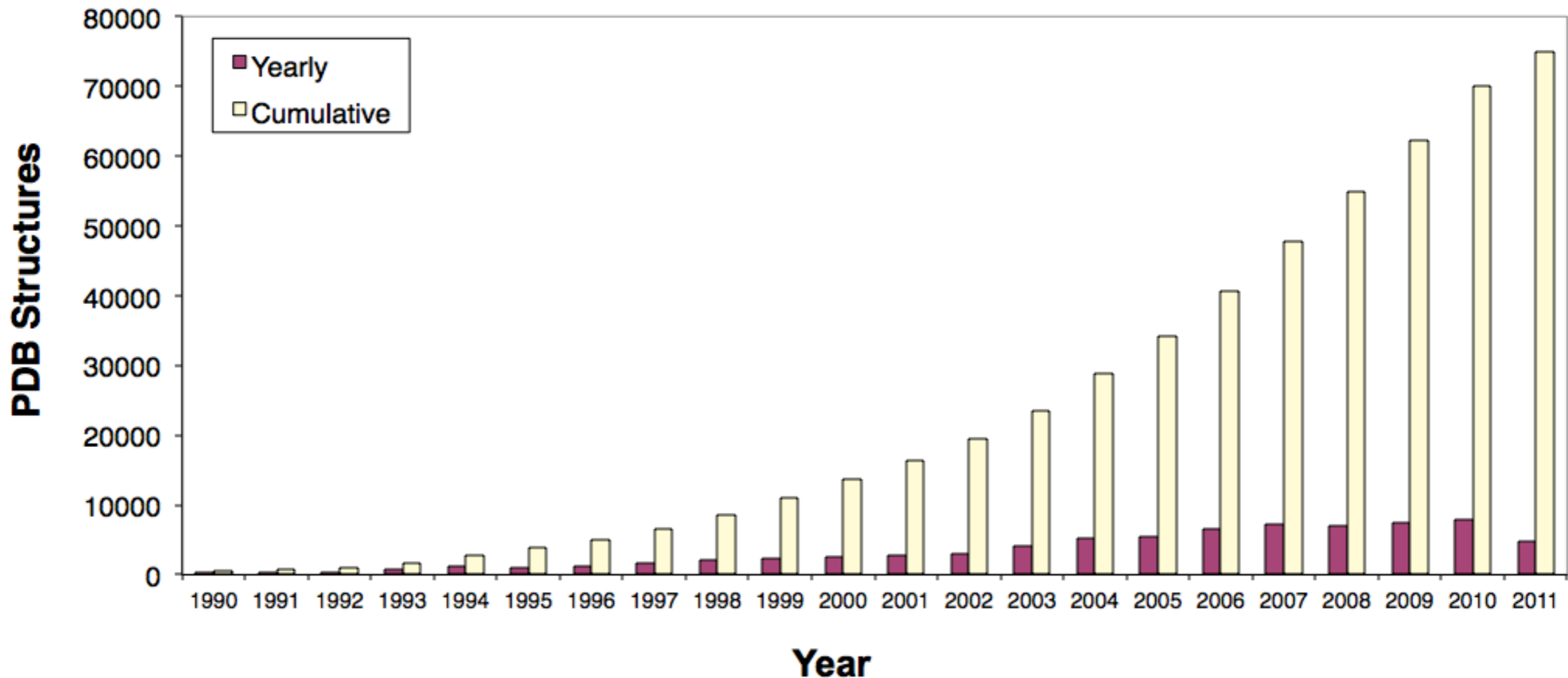
Eidogen-Sertanty, Inc

[smuskal@eidogen-sertanty.com](mailto:smuskal@eidogen-sertanty.com)

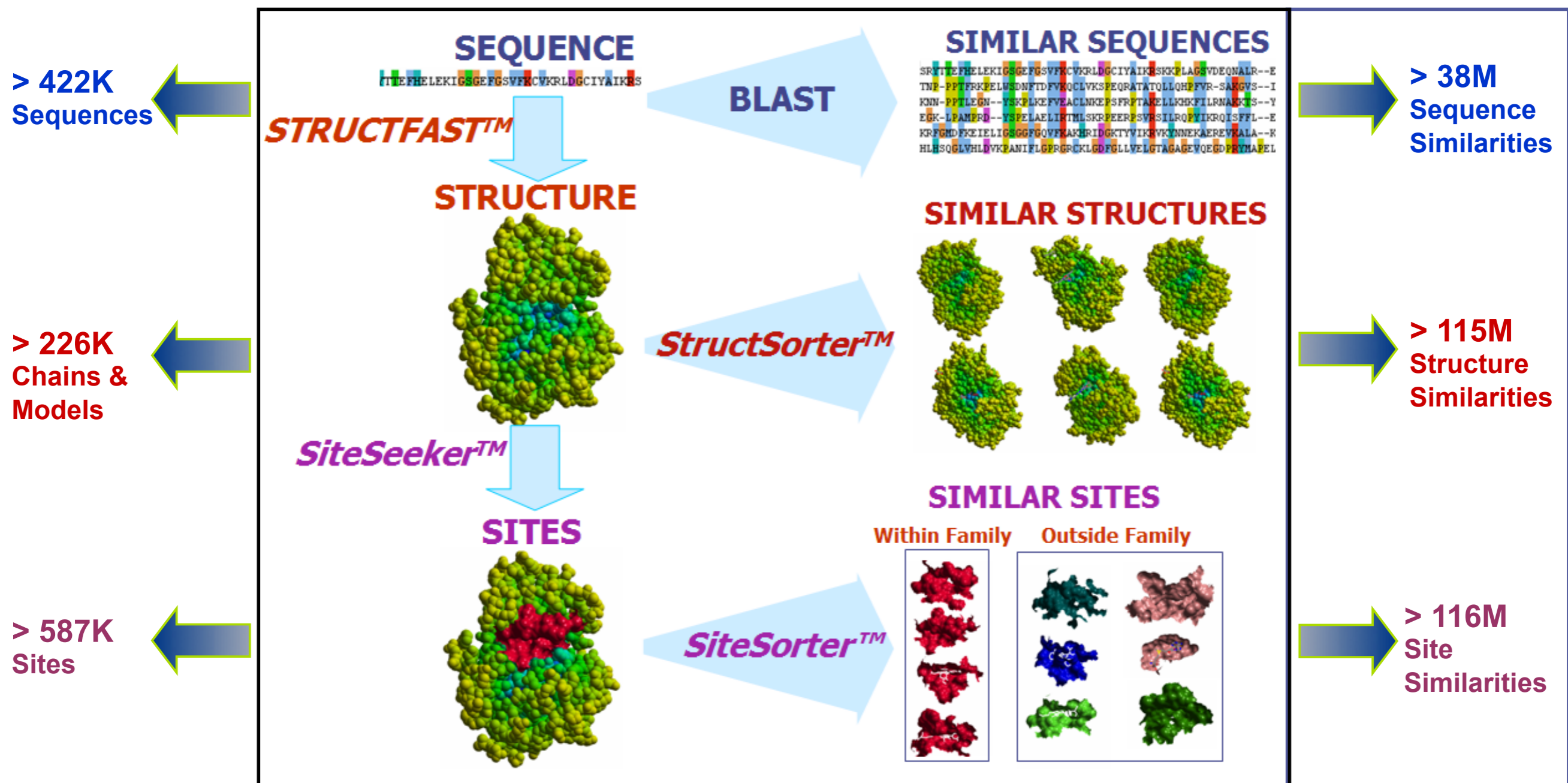
# Protein Structure Growth Continues

~ 75K Structures/co-complexes (Aug-2011)  
> 600 deposits per month → >150/week!

**PDB Growth**  
source: rcsb.org



# Target Informatics Platform (TIP)

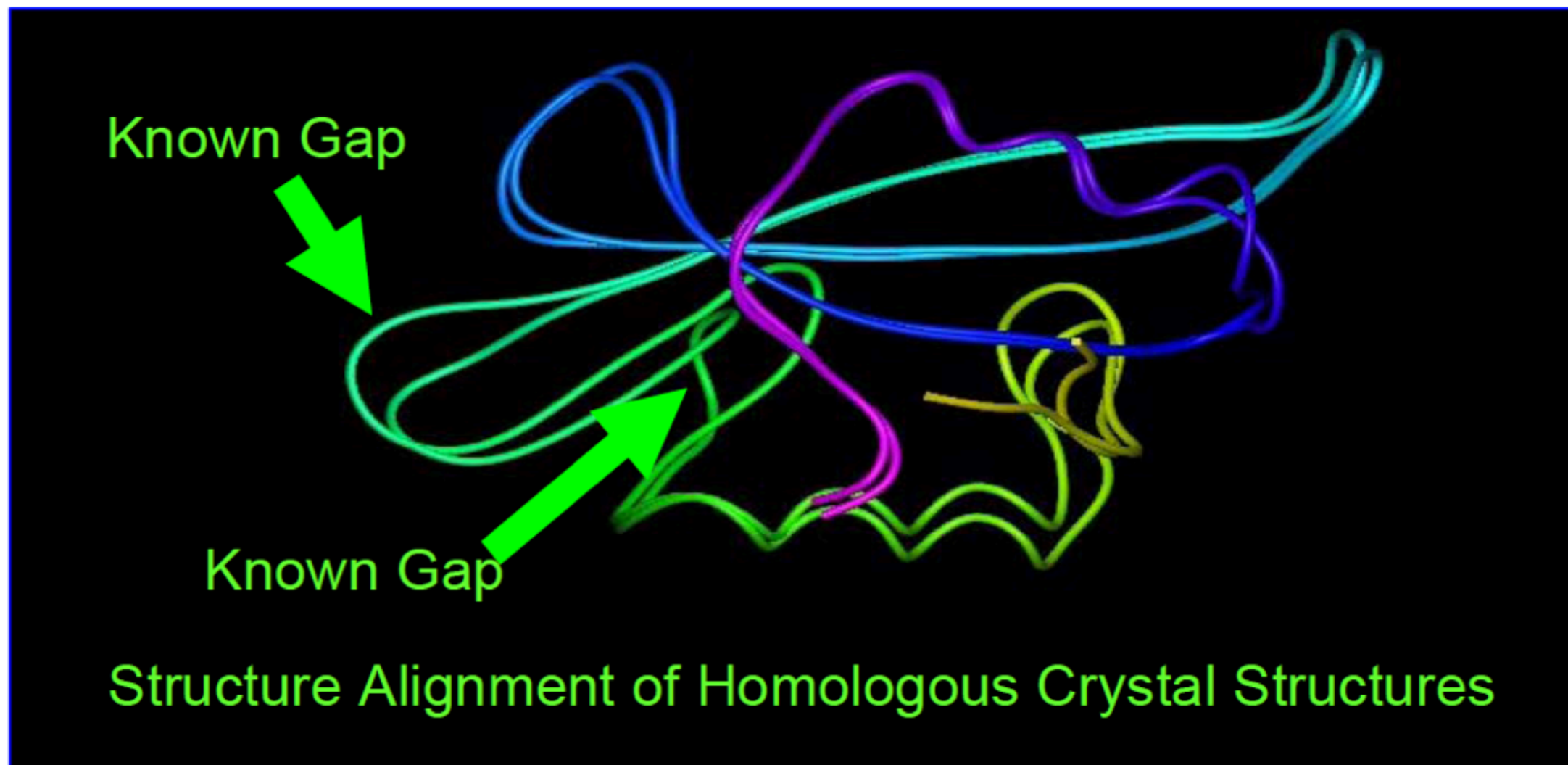


- Interrogating the druggable genome with structural informatics **MolecularDiversity (2006)**
- STRUCTFAST: Protein Sequence Remote Homology Detection and Alignment Using Novel Dynamic Programming and Profile-Profile Scoring **Proteins. 2006 64:960-967**
- StructSorter: A Method for Continuously Updating a Comprehensive Protein Structure Alignment Database **J. Chem. Inf. Model. 2006, 46, 1871-1876**
- Convergent Island Statistics: A fast method for determining local alignment score significance. **Bioinformatics, 2005, 21, 2827-2831.**

# STRUCTFAST

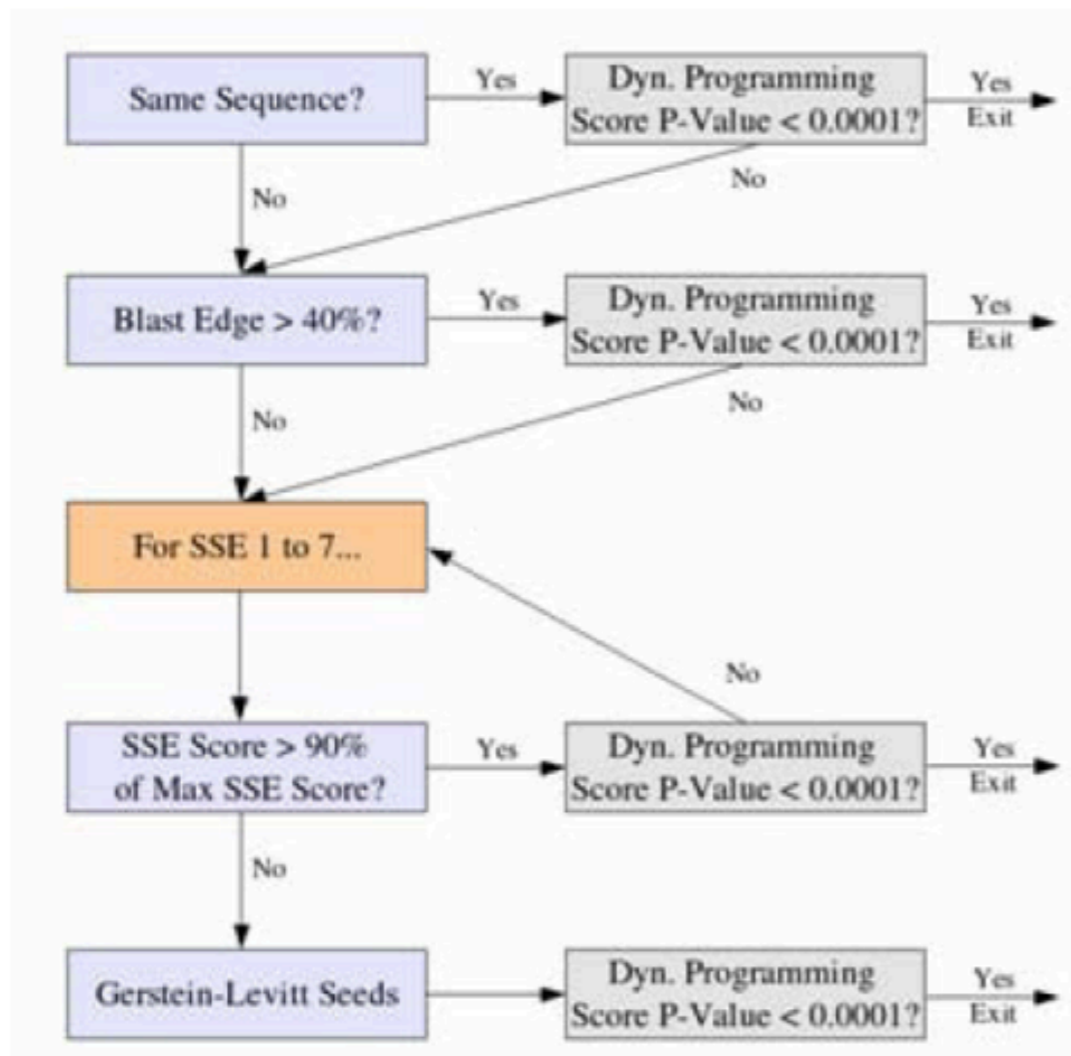
*Structure Realization Utilizing Cogent Tips From Aligned Structural Templates*

Basic Principle: Gaps known to exist should not be strongly penalized.



Leverages experimental structure and structural alignment data to create better alignments

# StructSorter



Various dynamic programming seeding methods are used in order to utilize as much information as is available.

Dynamic programming scores are fit to an EVD to assess alignment significance.

## Clustering Scheme and Hierarchical Protocol

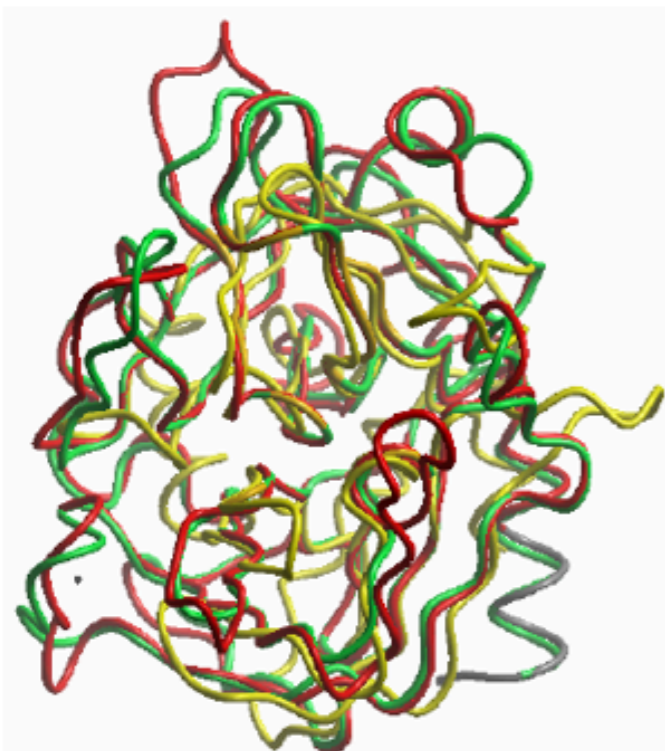
- 1) PDB sequences clustered at 90% identity and 95% coverage.
- 2) N-by-N comparison of one representative chain from each cluster

(All other chains are only compared to the representative's significant hits)

**Allows structural alignment database to be computed in 1.5 months instead of 2.5 years.**

# StructSorter Example

StructSorter computes and stores alignments between Rhinovirus Protease and other mammalian proteases in TIP, despite very low overall sequence and structural similarity



EVE Target Analyzer

File Export Filtering Ligand Window Help

Sequences Chains Sites Binding Modes

Chain Name	Locus	Organism	CRMS	%ID	Sites	Description
pdb1cqq/A	POLG_	H.rhinovirus	-	-	2	Chain A, MOL_ID: 1; MOLECULE: TYPE 2 RHINOVIRUS 3C F
pdb1q31/A	POLG_	T.etch virus	2.325	11	2	Chain A, NUCLEAR INCLUSION PROTEIN A
pdb1q31/B	POLG_	T.etch virus	2.324	11	2	Chain B, NUCLEAR INCLUSION PROTEIN A
pdb1f7z/A	TRY2_R_	B.taurus	2.753	11	1	Chain A, TRYPSIN II, ANIONIC
pdb1a0j/A	TRY3_S_	S.salar	2.581	11	5	Chain A, TRYPSIN
pdb1spj/A	KLK1	H.sapiens	2.704	10	5	Chain A, KALLIKREIN 1
pdb1mza/A	GRAK_	H.sapiens	3.043	9	2	Chain A, PRO-GRANZYME K
pdb1mzd/A	GRAK_	H.sapiens	3.530	10	2	Chain A, PRO-GRANZYME K
pdb1bio/_	CFAD_	H.sapiens	2.822	8	5	Chain _ COMPLEMENT FACTOR D
pdb1bru/P	EL2_PIG	S.scrofa	2.696	9	2	Chain P, ELASTASE
pdb1p57/B	HEPS_	H.sapiens	2.705	9	2	Chain B, SERINE PROTEASE HEPSIN
pdb1a0i/A	TRB2_	H.sapiens	2.687	10	2	Chain A, BETA-TRYPTASE
model3999_	MPN	H.sapiens	2.805	9	3	Pancreasin precursor (EC 3.4.21.-) (Marapsin) (Channel-act
pdb1ybw/A	-	H.sapiens	2.722	11	2	Chain A, HEPATOCYTE GROWTH FACTOR ACTIVATOR PRE
pdb1eaw/A	ST14_H_	B.taurus	2.713	7	1	Chain A, SUPPRESSOR OF TUMORIGENICITY 14
model5711_	PRSS12	H.sapiens	2.867	8	3	Neurotrypsin precursor (EC 3.4.21.-) (Motopsin) (Leydin)
pdb1lmw/B	UROK_	H.sapiens	2.924	10	5	Chain B, UROKINASE-TYPE PLASMINOGEN ACTIVATOR
pdb1lmw/D	UROK_	H.sapiens	2.846	10	5	Chain D, UROKINASE-TYPE PLASMINOGEN ACTIVATOR
model3428_	PROZ	H.sapiens	3.007	8	4	Vitamin K-dependent protein Z precursor
pdb1h1b/A	ELNE_H_	H.sapiens	2.672	9	8	Chain A, LEUKOCYTE ELASTASE
model2909_	HP	H.sapiens	3.375	5	3	Haptoglobin precursor
pdb1iau/A	GRAB_	H.sapiens	2.619	11	11	Chain A, GRANZYME B
pdb1131/A	MCT1_	H.sapiens	2.808	10	12	Chain A, CHYMASE
pdb1jpl/A	MCT1_	H.sapiens	3.298	10	6	Chain A, CHYMASE
pdb1azz/B	COGS_	C.pugilator	2.625	8	2	Chain B, COLLAGENASE
pdb1azz/A	COGS_	C.pugilator	2.615	8	2	Chain A, COLLAGENASE
pdb1au8/A	CATG_	H.sapiens	2.588	11	2	Chain A, CATHEPSIN G
pdb1gpz/A	C1R_H_	H.sapiens	3.379	11	12	Chain A, COMPLEMENT C1R COMPONENT
model2909_	HP	H.sapiens	3.047	5	5	Haptoglobin precursor
pdb1gpz/B	C1R_H_	H.sapiens	3.373	10	6	Chain B, COMPLEMENT C1R COMPONENT
pdb1sgf/A	KLK4_	M.musculus	3.614	8	3	Chain A, NERVE GROWTH FACTOR
pdb1sgf/K	KLK4_	M.musculus	3.623	8	4	Chain X, NERVE GROWTH FACTOR
pdb1wcz/A	STSP_	S.aureus	2.604	10	2	Chain A, GLUTAMYL ENDOPEPTIDASE
model1370_	PRSS11	H.sapiens	2.822	12	1	Serine protease HTRA1 precursor (EC 3.4.21.-) (L56)
pdb1agj/A	ETA_ST_	S.aureus	2.558	7	2	Chain A, EPIDERMOLYTIC TOXIN A
pdb1dxi/A	POLG_	H.c virus	2.677	11	3	Chain A, PROTEASE/HELICASE NS3 (P70)
pdb1bef/A	POLG_	D.virus type 2	2.739	9	1	Chain A, DENGUE VIRUS NS3 SERINE PROTEASE
pdb1bl7/_	POLG_	H.c virus	3.157	11	2	Chain _ NS3 SERINE PROTEASE
pdb1hpg/A	GLUP_	S.griseus	2.730	8	4	Chain A, GLUTAMIC ACID-SPECIFIC PROTEASE

Chain Alignments

Similarity Dendrogram

# SiteSeeker

## Geometric Site-Finding Algorithms Find Many Pockets

*But they don't know which pockets are important!*

## Evolutionary Trace Approach

*Can't clearly define site boundary*

*Not all conserved residues are functionally relevant*

## **SiteSeeker combines *both* methods**

### Reliability & Confidence

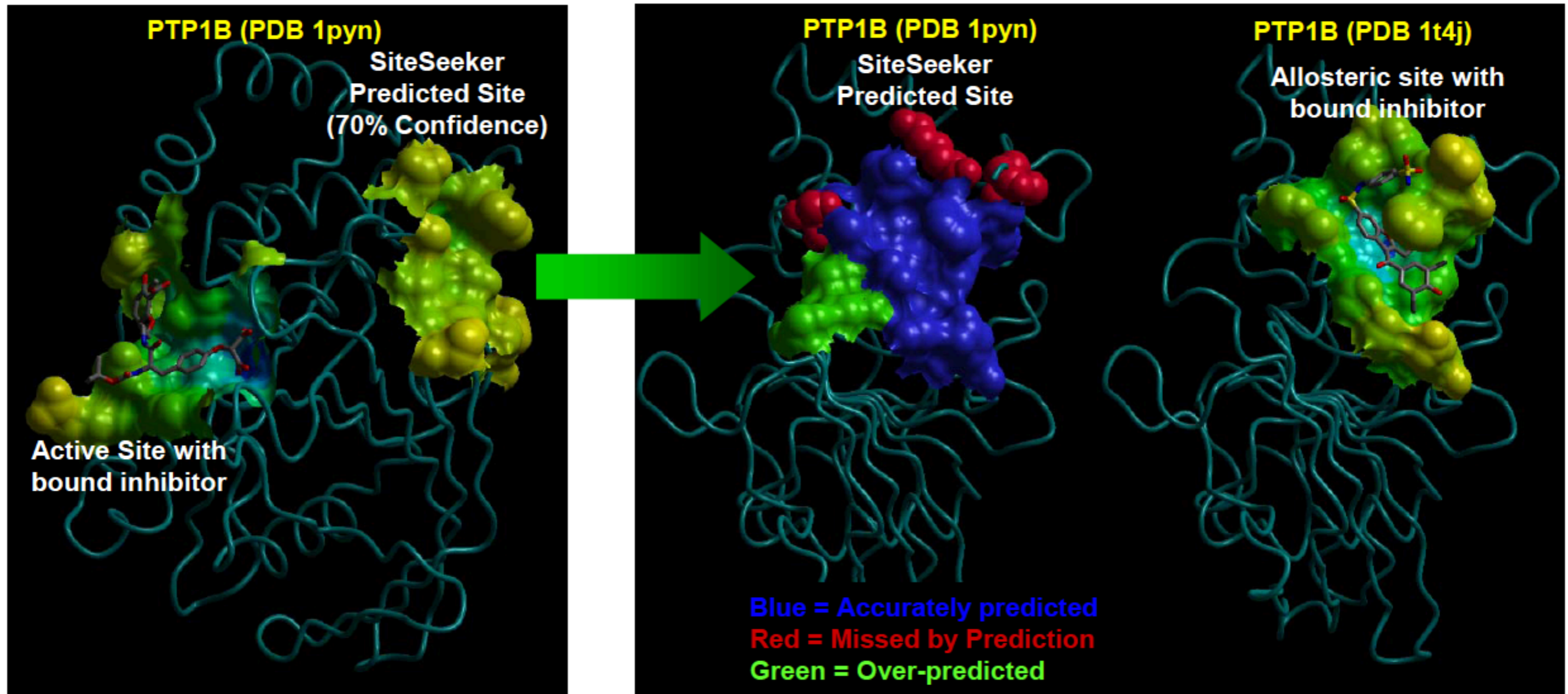
We use proteins with apo- & co-crystal structures in the PDB to test the accuracy & reliability of method

Allows us to map *SiteSeeker* score to predict confidence!

(e.g. At this *SiteSeeker* score, 80% are “real” co-crystal sites)

→ Sites with <60% confidence are not stored in TIP

# SiteSeeker Example



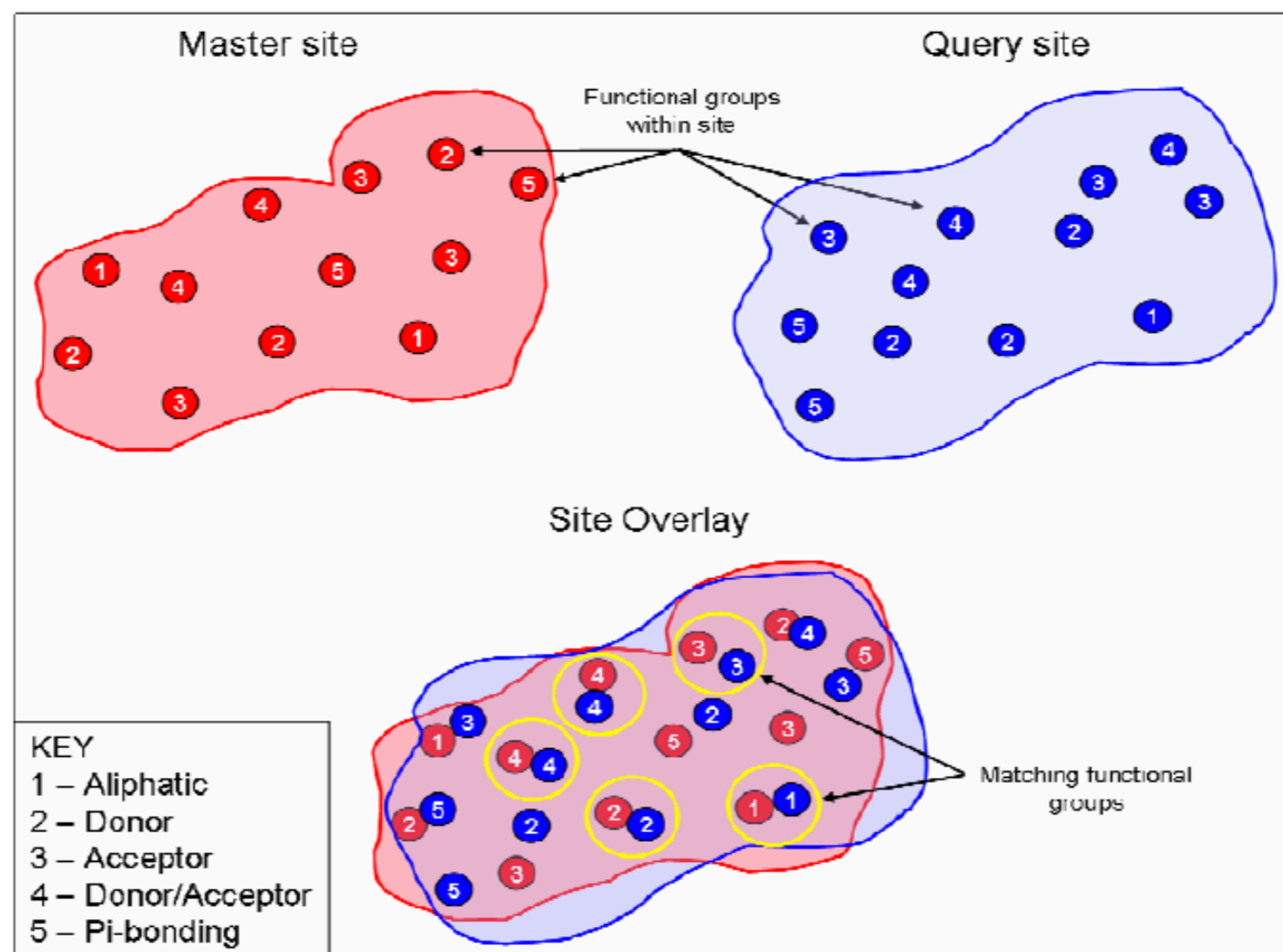
All structures in TIP are annotated with known and predicted binding sites, along with **confidence** levels for each annotation



# SiteSorter

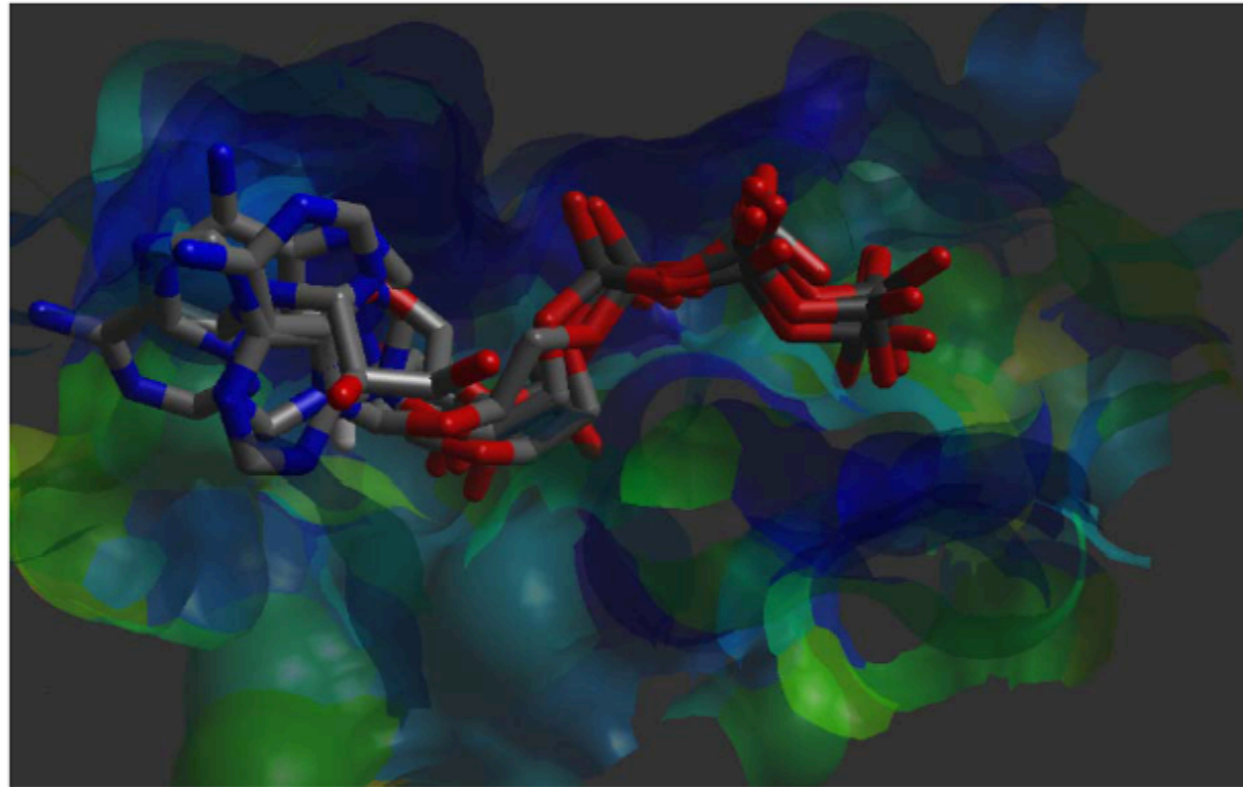
## Weighted Clique Detection Algorithm

Importance of Points (Weights) Related to their Similarity

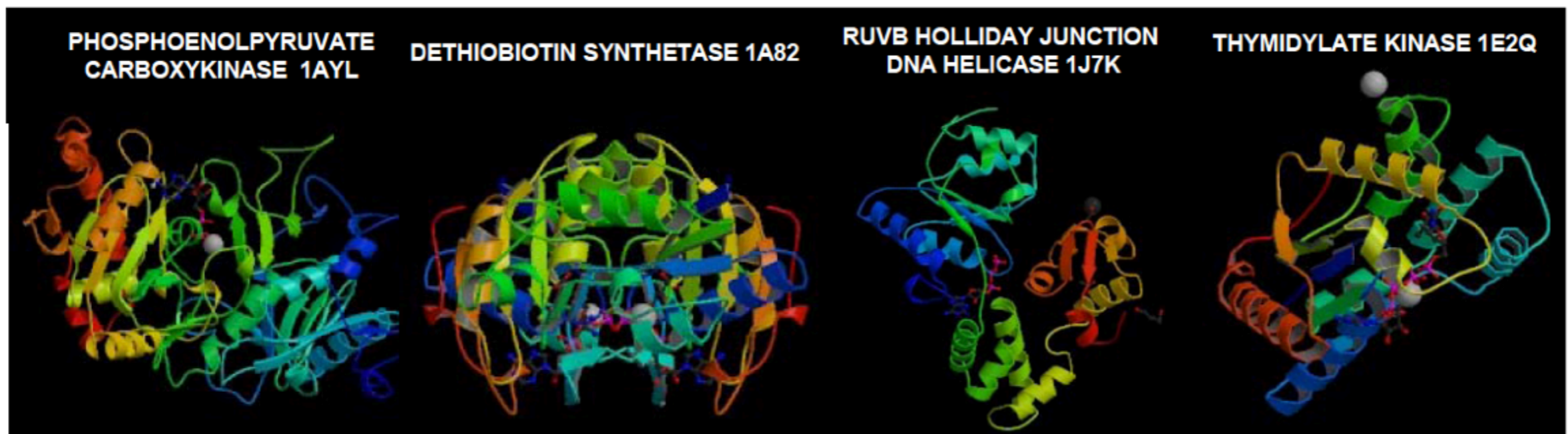


Surface Atoms Assigned One of 5 Different Chemical Characters  
Matching points increase the *SiteSorter* similarity score

# SiteSorter Example



Overlay of ATP binding sites from completely different folds



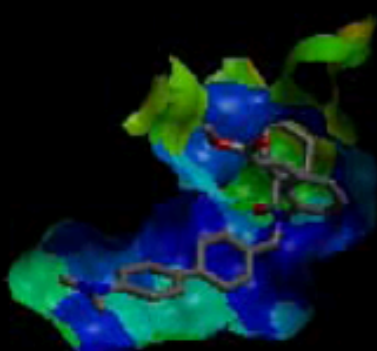
# Nature Exploits Site-Similarity

## Query: PXR site

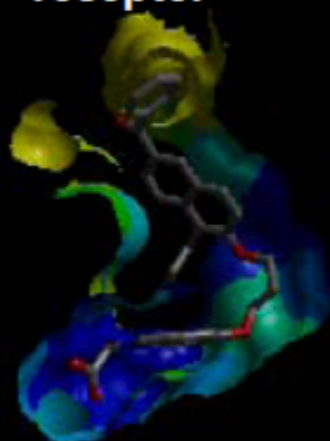


## Example High-ranking similar sites:

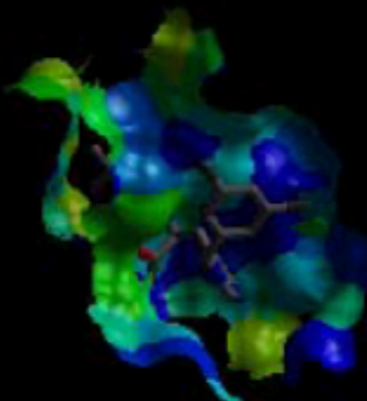
Bile Acid Receptor FXR



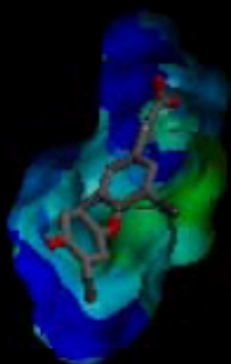
PPAR-gamma receptor



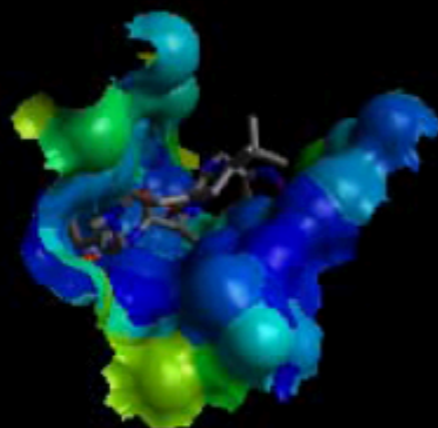
ACE2



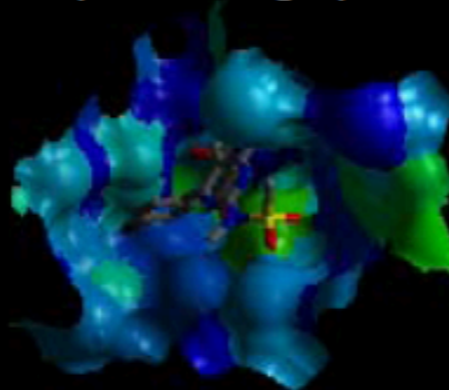
Thyroid Receptor



Caspase-3

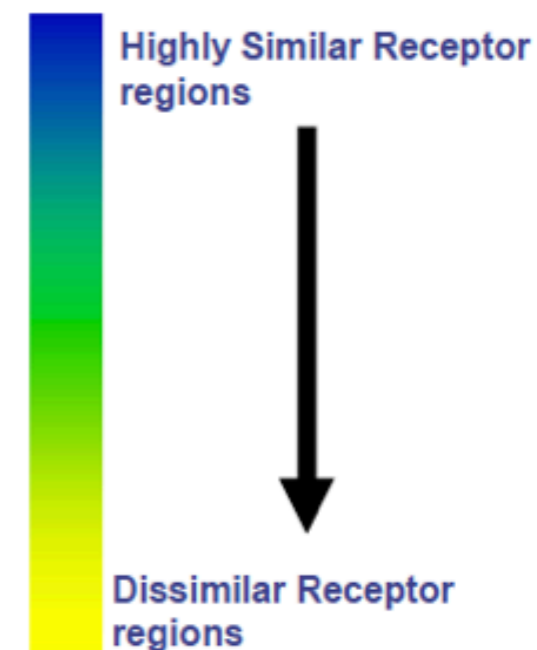


HMG-CoA Reductase (statin target)



Pregnane X-receptor –  
PXR (“sensor”) → CYP3A4  
 (“executioner”)  
PXR Binds > 50% drugs  
Including some bile acids,  
statins, herbal components, a  
selection of HIV protease  
inhibitors, calcium channel  
modulators, numerous  
steroids, plasticizers and  
monomers, organochlorine  
pesticides, a peroxisome  
proliferator-activated receptor-  
antagonist, xenobiotics and  
endobiotics...

## Site Similarity Coloring



# Leveraging SiteSimilarity: Kinases

Nature Reviews | Drug Discovery Vol 8 | February, 2009

Table 1 | **Selected multi-target kinase inhibitors**

Drug (company)	Target	Highest phase	Indication*
Sorafenib (Bayer and Onyx)	PDGFR, VEGFR2 and 3, FLT3, KIT, RET, RAF	Launched	Hepatocellular carcinoma, RCC, renal tumour
Dasatinib (BMS)	BCR-ABL, FYN, SRC, LCK, EPH	Launched	ALL, CML
Nilotinib (Novartis)	PDGFR, ABL, KIT	Launched	CML
Sunitinib (Pfizer)	PDGFR, VEGF2, FLT3, KIT	Launched	Gastrointestinal tumour, RCC
Motesanib (Amgen and Takeda)	PDGFR, VEGFR, KIT	Phase III	NSCLC
Vandetanib (AstraZeneca)	EGFR, VEGFR2, RET	Phase III	Thyroid tumour, NSCLC
Lestaurtinib (Cephalon)	JAK2, FLT3, TRKA	Phase III	Myeloid leukaemia
XL184 (BMS and Exelixis)	VEGFR2, MET, KIT, FLT3, RET, TEK	Phase III	Thyroid tumour
Pazopanib (GSK)	PDGFR, VEGFR1, 2 and 3, KIT	Phase III	Renal tumour, sarcoma

\*Indication given for highest phase; all drugs are also in lower phase clinical trials for other oncology indications. ALL, acute lymphoblastic leukaemia; BMS, Bristol-Myers Squibb; CML, chronic myeloid leukaemia; EGFR, epidermal growth factor receptor; GSK, GlaxoSmithKline; NSCLC, non-small-cell lung cancer; PDGFR, platelet-derived growth factor receptor; RCC, renal cell carcinoma; VEGFR, vascular endothelial growth factor receptor.

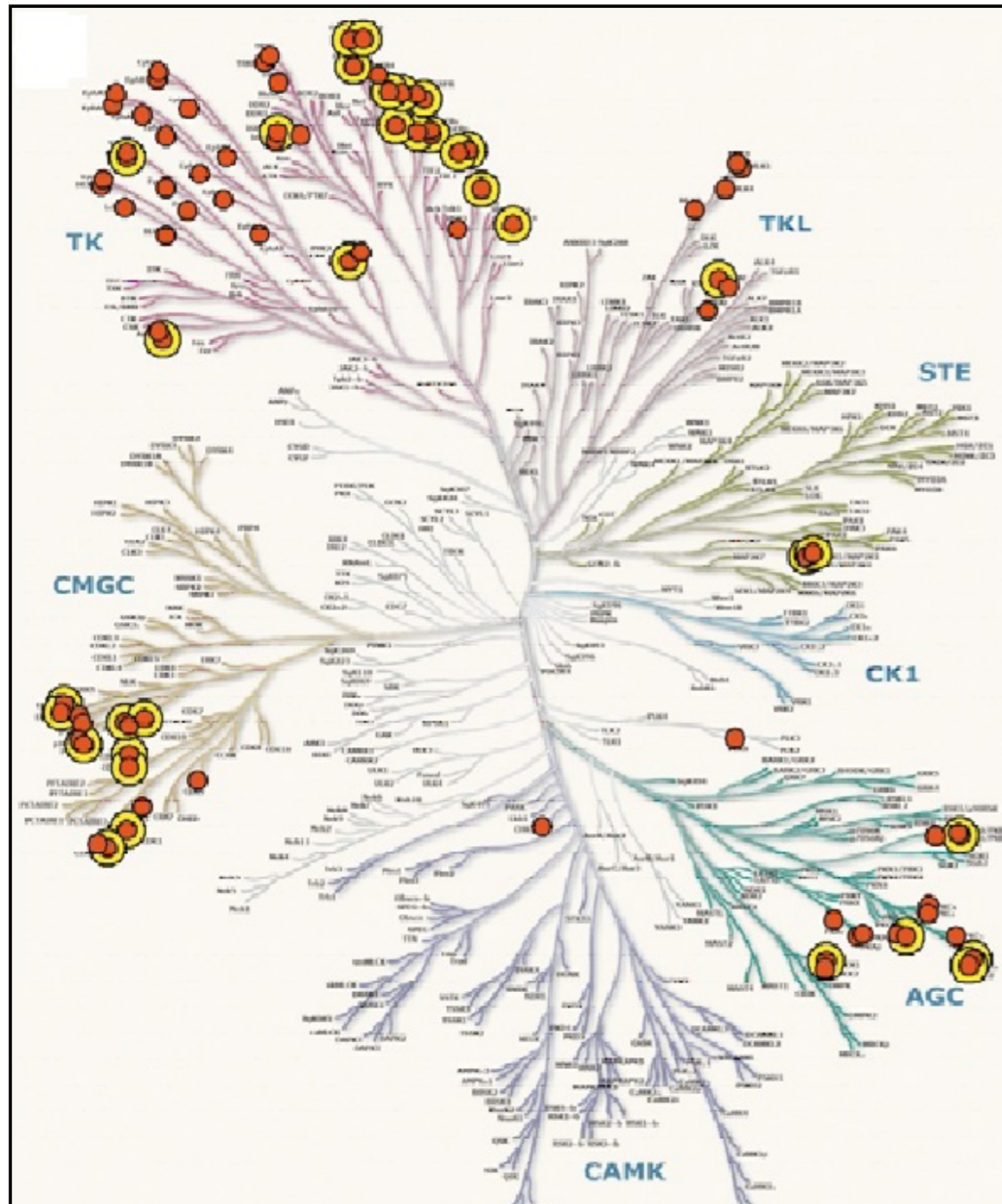
**Imatinib (Gleevec: Novartis)**      **ABL, PDGFR, KIT**      **CML, GIST**

**Gefitinib (Iressa: Astra Zeneca)**      **EGFR, (ERBB4,GAK,...)**      **NSCLC**

# Kinase Knowledgebase (KKB)

## Kinase Targets of Clinical Interest

from Vieth *et al. Drug Disc. Today* **10**, 839 (2005).

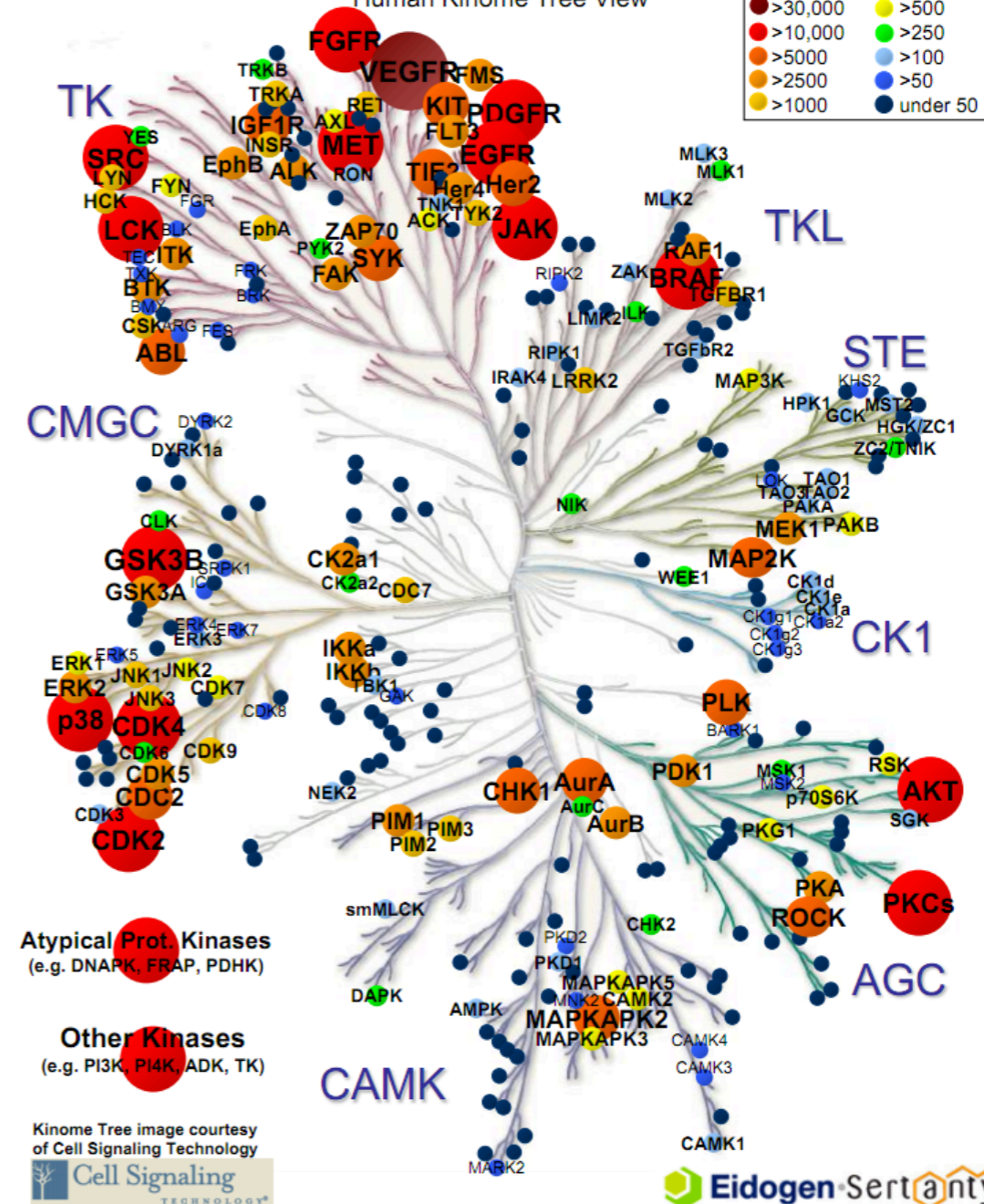
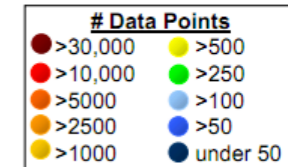


## Eidogen-Sertanty KKB SAR Data Point Distribution

Distribution of Kinase Knowledgebase Bio-Activity Data Points

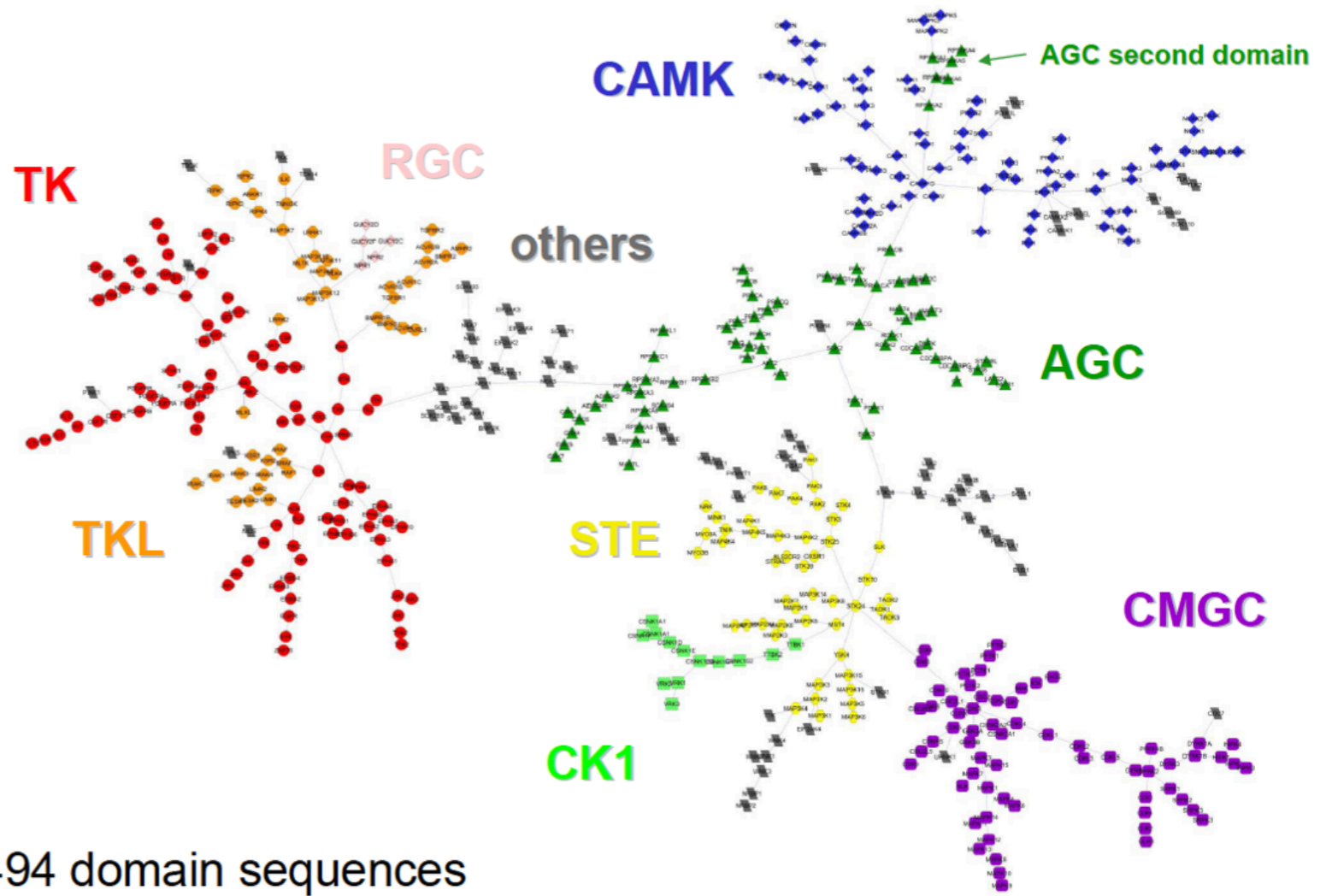
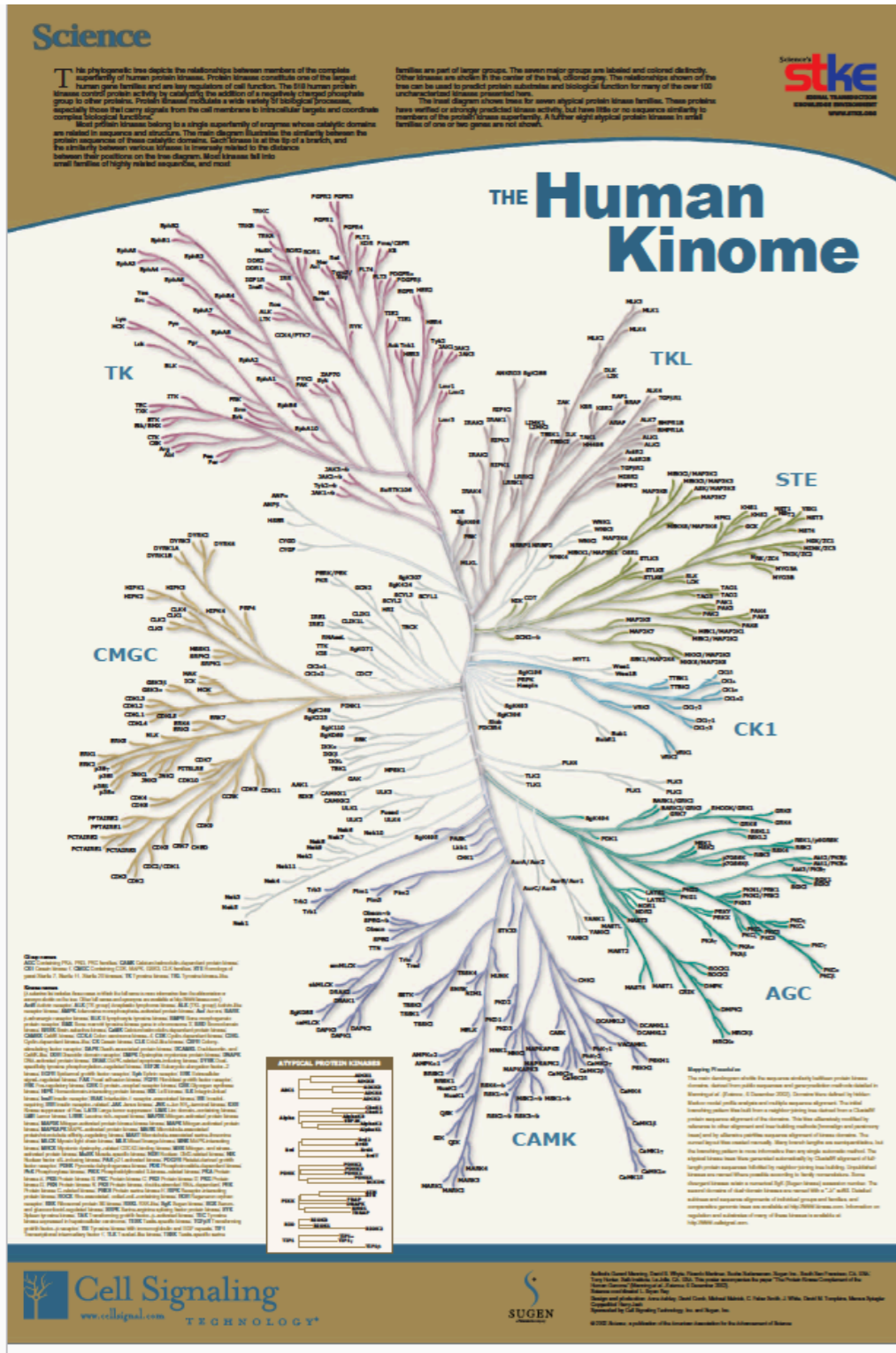
Q2 2011 Release

Human Kinome Tree View

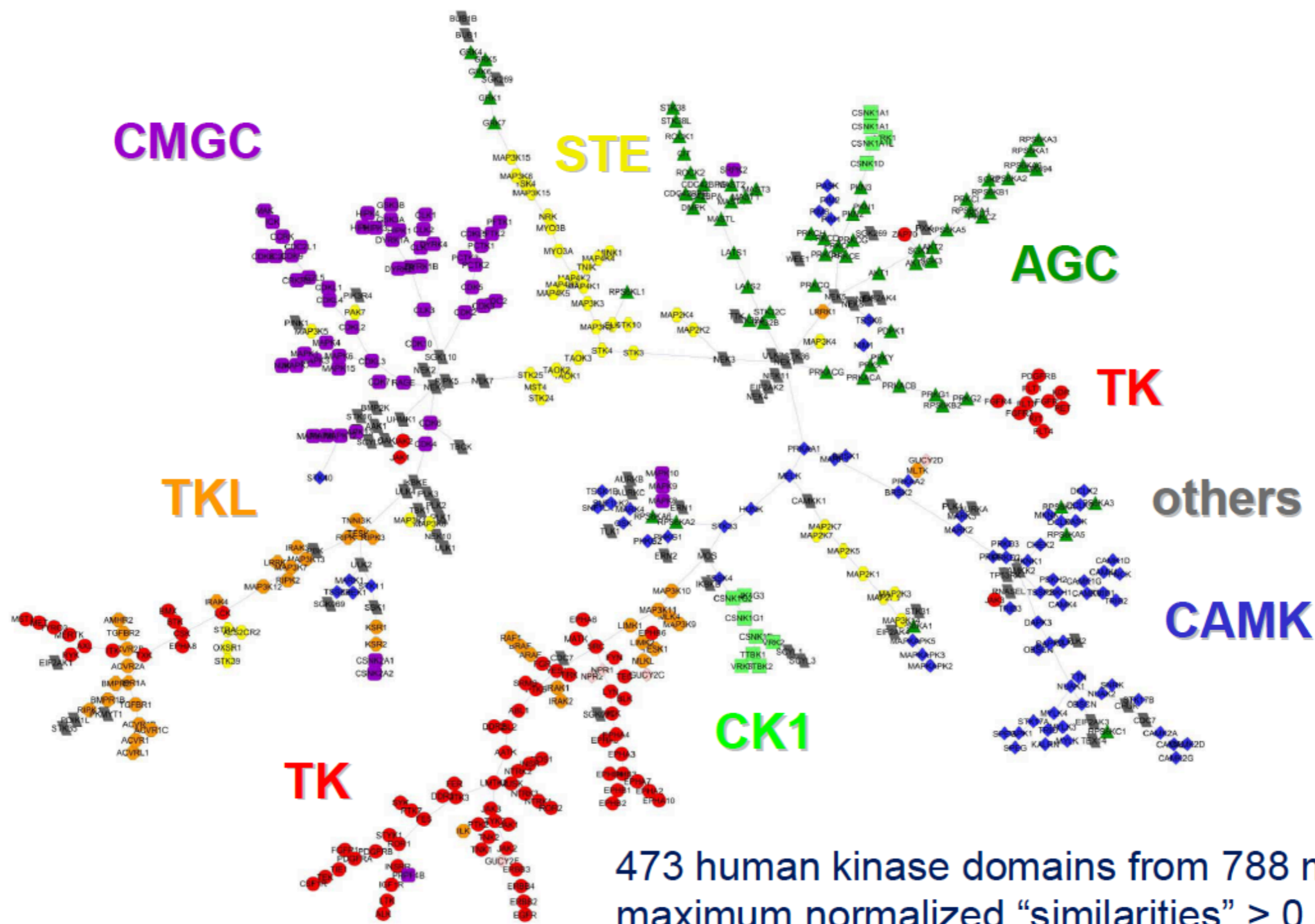


> 545,000 SAR data points curated from  
> 7428 journal articles and patents

# Kinome Trees



# Local Site Similarity - MST

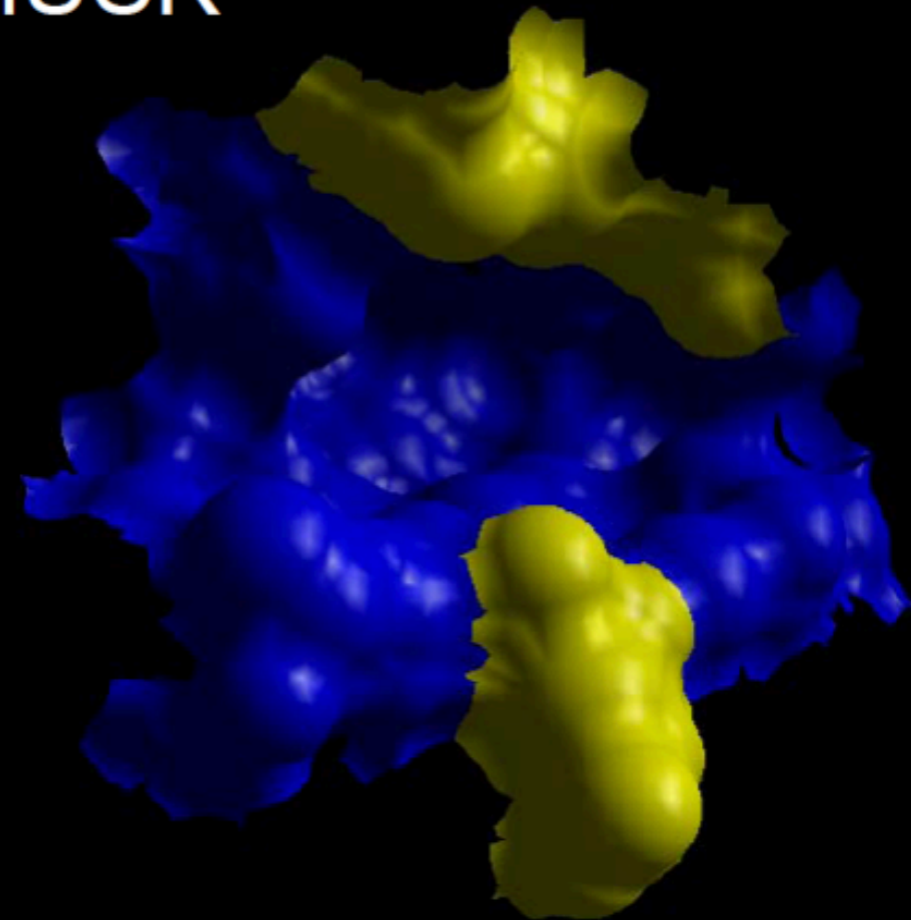
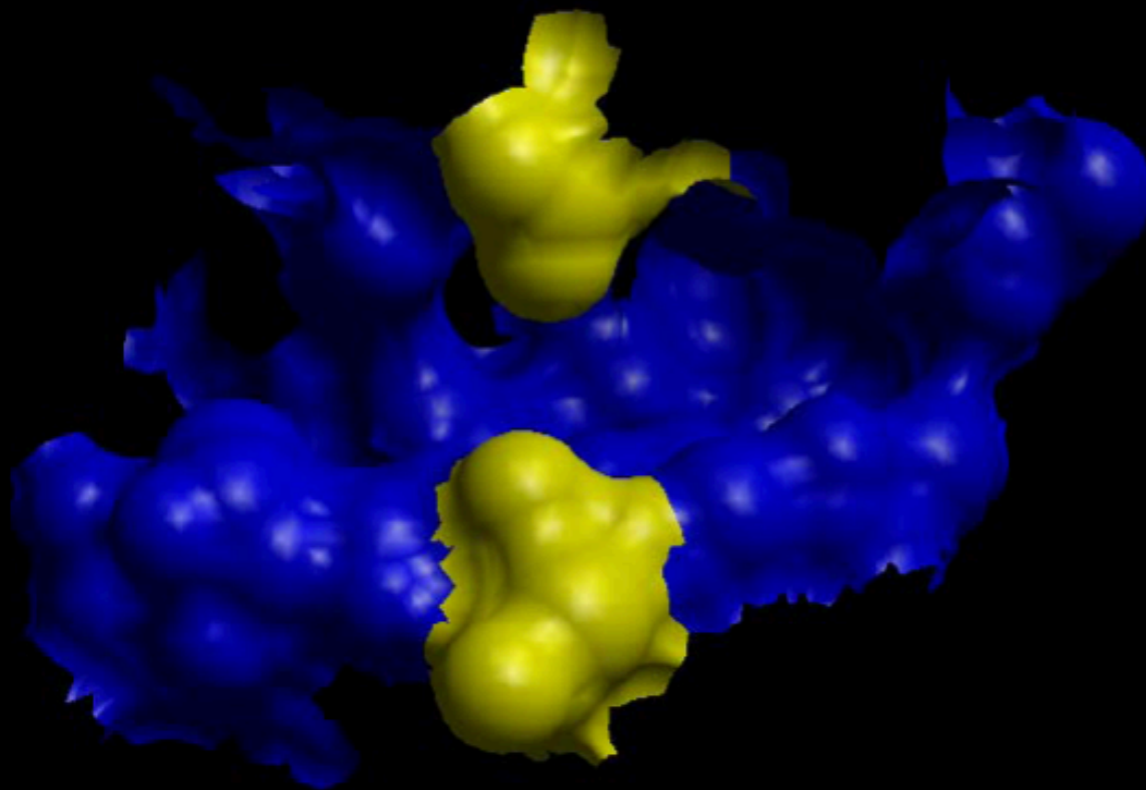


# PhysChem SiteSim v. Domain Seq ID

- STE\_STE20\_HGK (MAP4K4): template 1u5rA
- TK\_Musk\_MUSK (MUSK) : template 1ir3A
- Full Sequence identity: 0.22 Site Sequence identity: 0.55
- Normalized (physicochemical) site similarity: 0.84

MAP4K4

MUSK

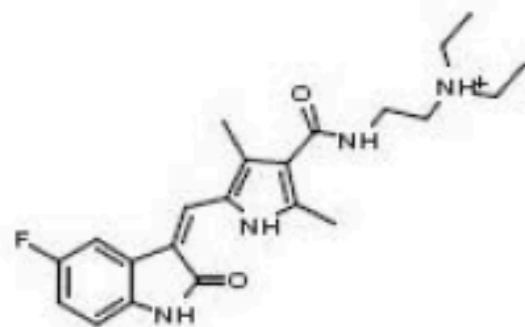


MAP4K4  
MUSK

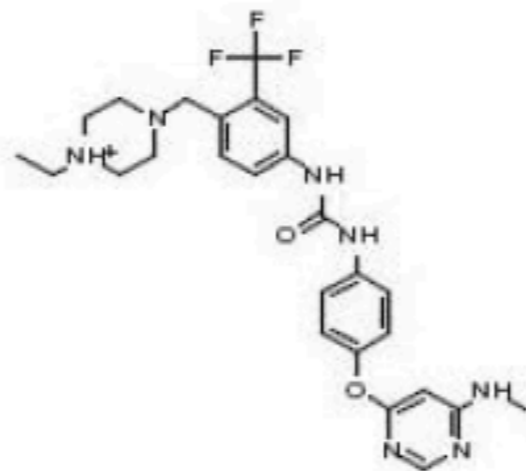
.VGNGTY.V.A.K.M.E.A.MEFC.AGS.D.D.QN.L.D  
.IGEGAF V A K - E V FEYM -GD - N -N L D



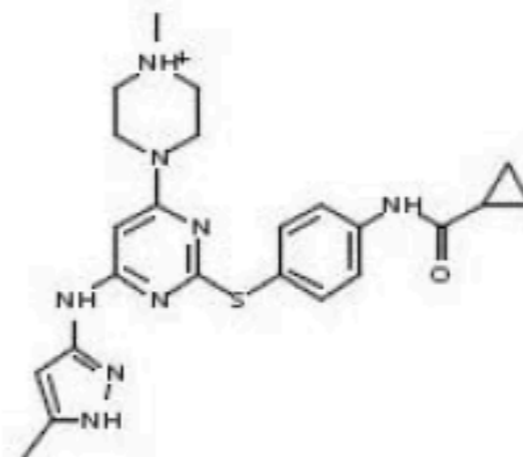
# MAP4K/MUSK Small Molecule Inhibitors



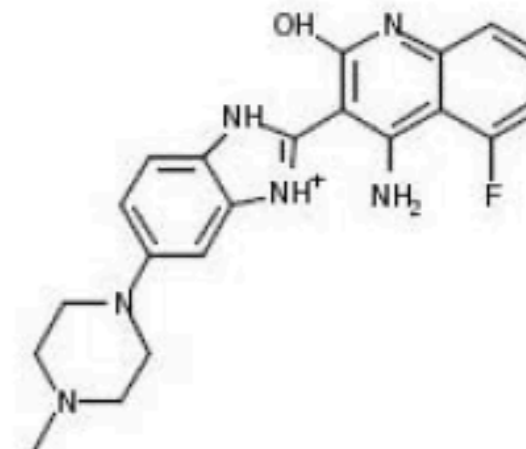
MAP4K4 / 6.9  
MUSK / 6.3



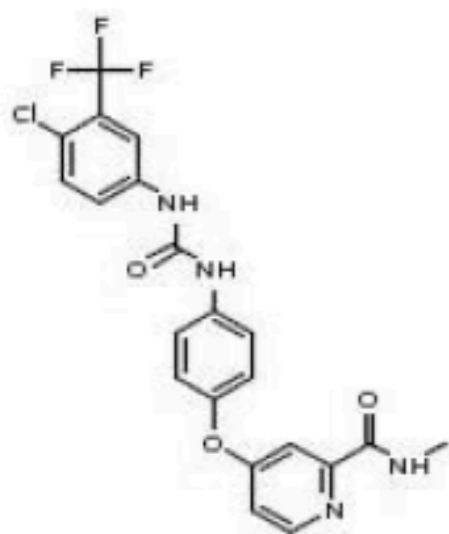
MAP4K4 / 7.8  
MUSK / 8.5



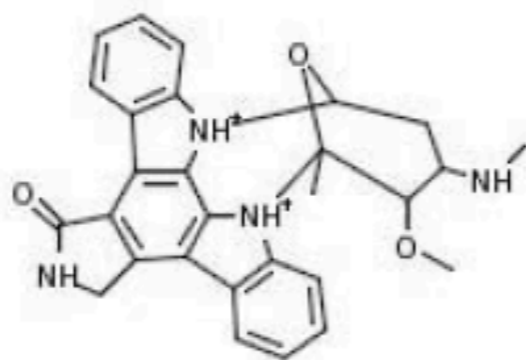
MAP4K4 / 5.6  
MUSK / 6.5



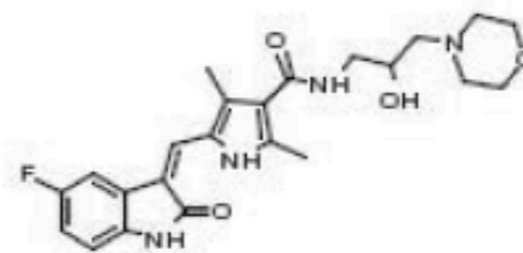
MAP4K4 / 6.8  
MUSK / 6.1



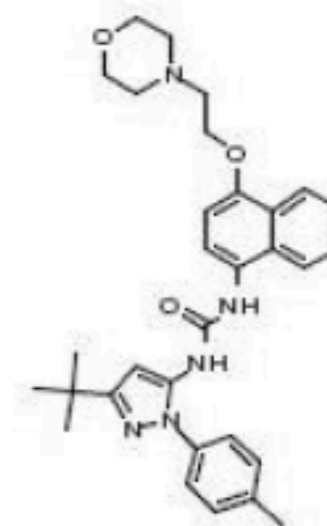
MAP4K4 / 5.3  
MUSK / 6.9



MAP4K4 / 7.6  
MUSK / 7.1



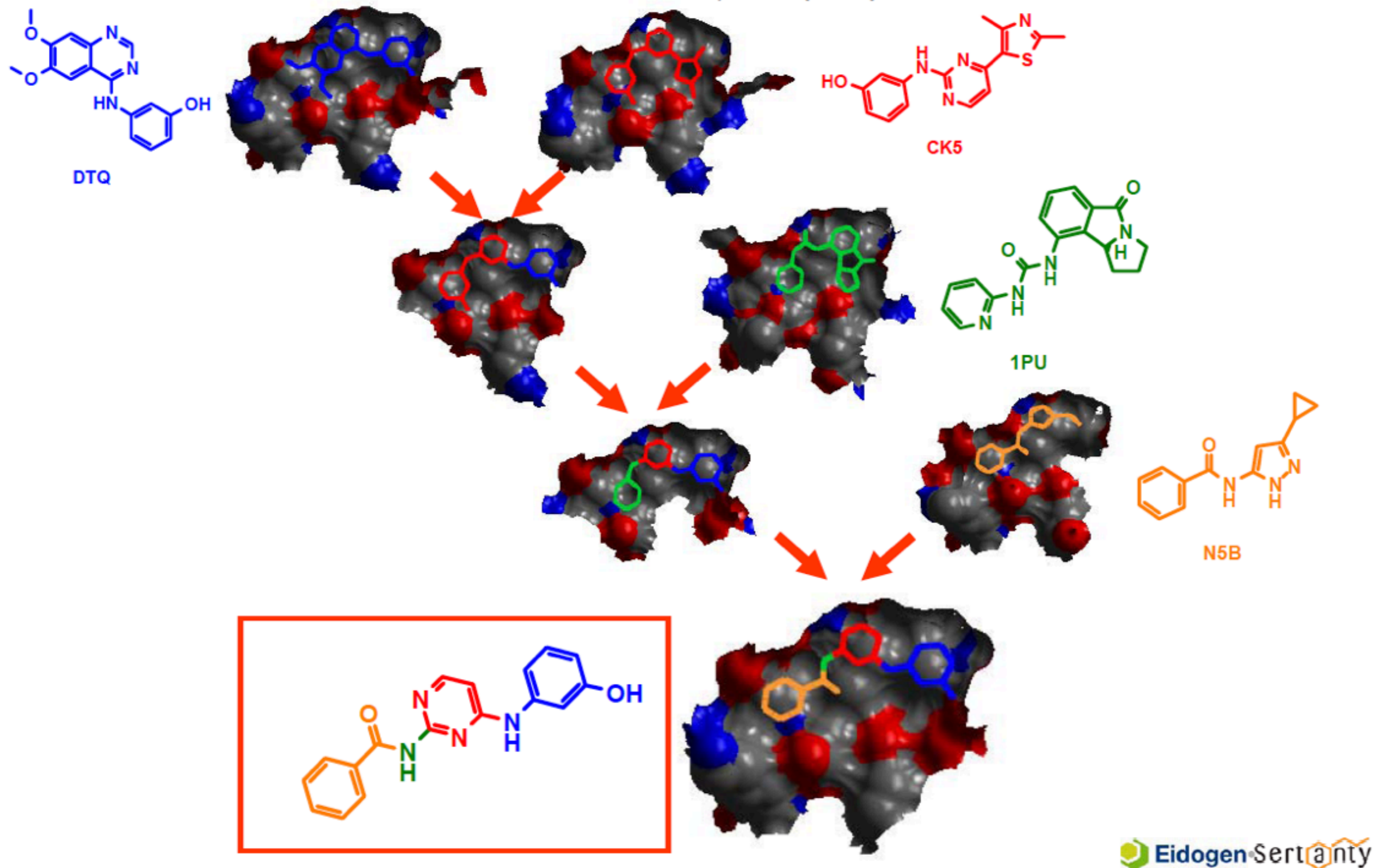
MAP4K4 / 6.6  
MUSK / 7.0



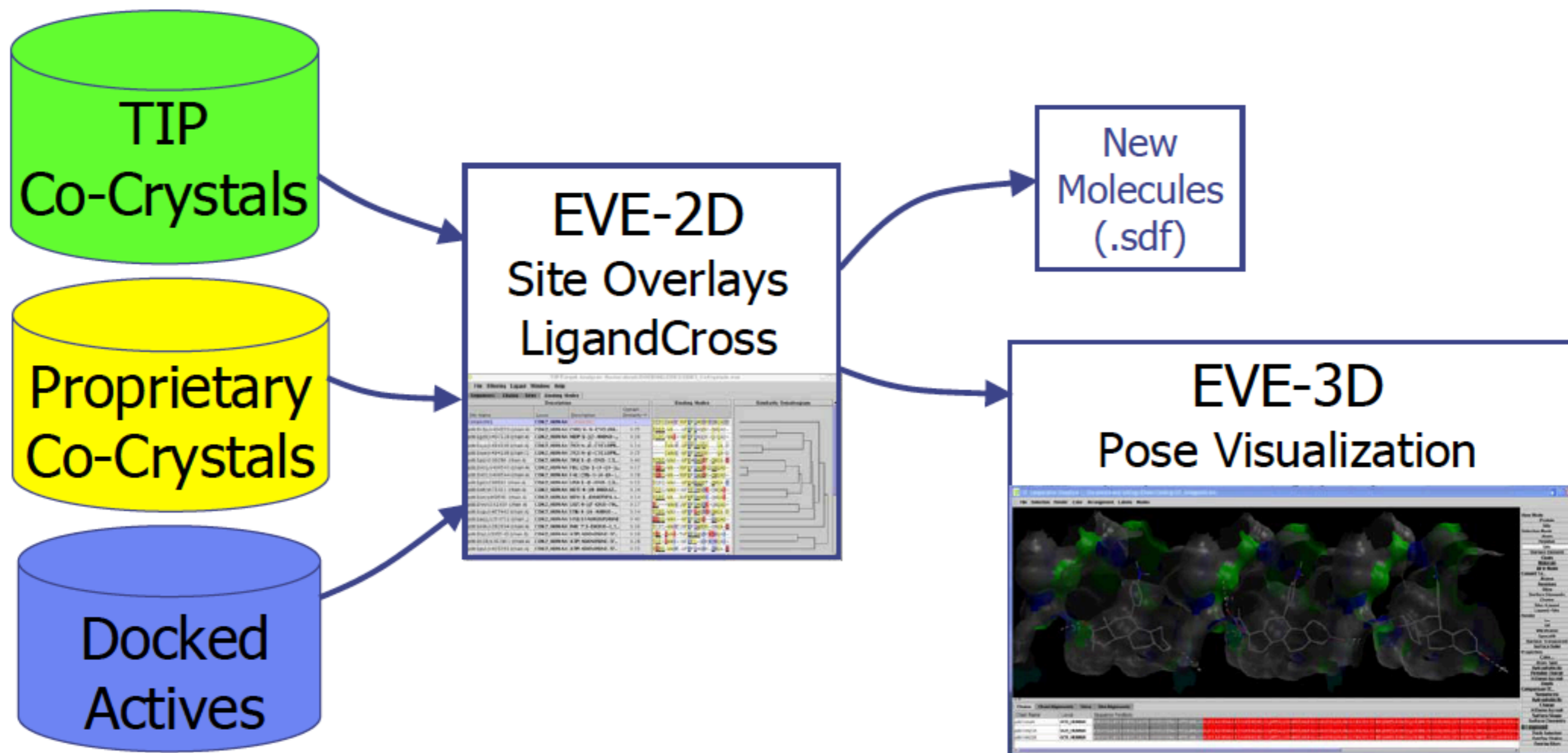
MAP4K4 / 7.0  
MUSK / 5.7

# LigandCross: Shuffling Ligand Functionality

Similar to Vertex's BREED: J. Med. Chem. **47**, 2768 (2004)

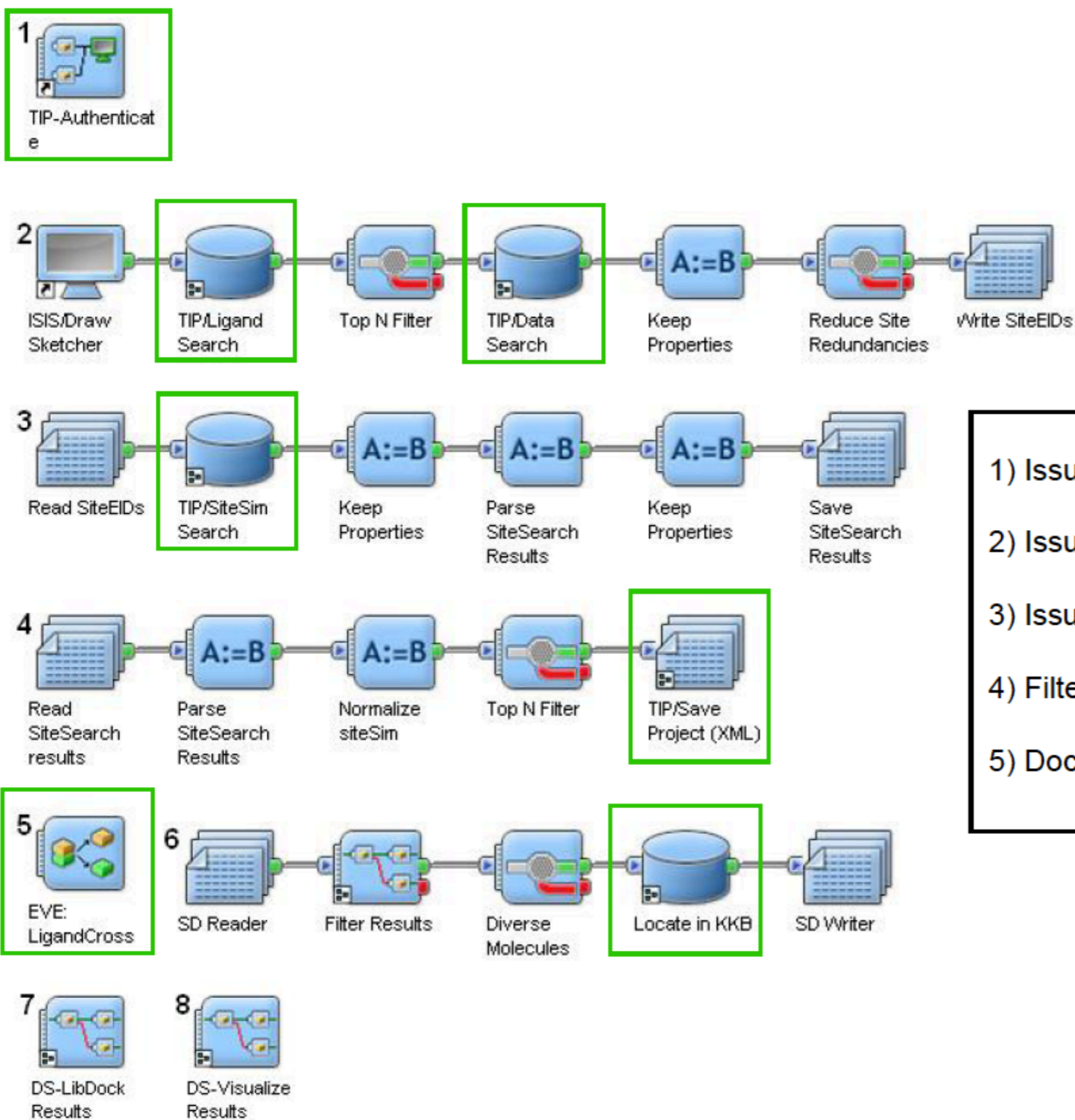


# LigandCross WorkFlow



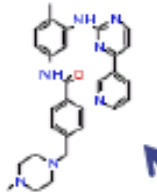
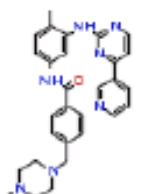
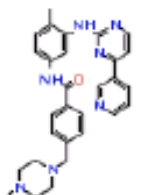
New Molecules via LigandCross

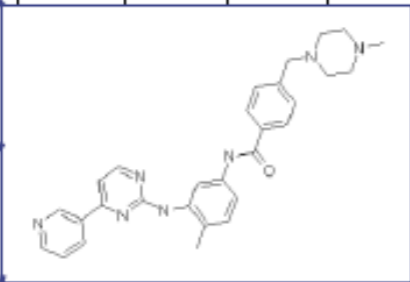
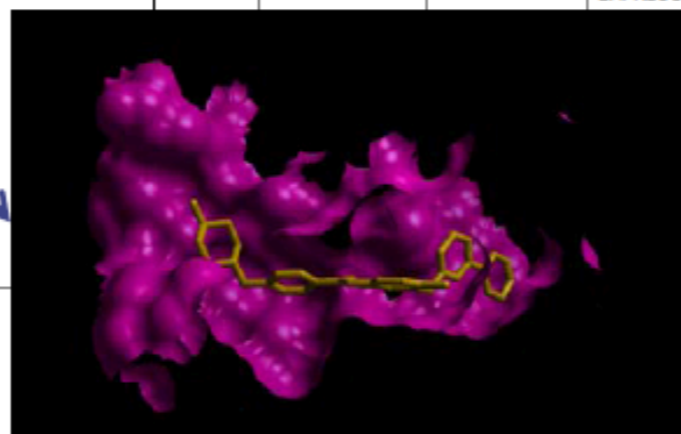
# LigandCross via PipelinePilot



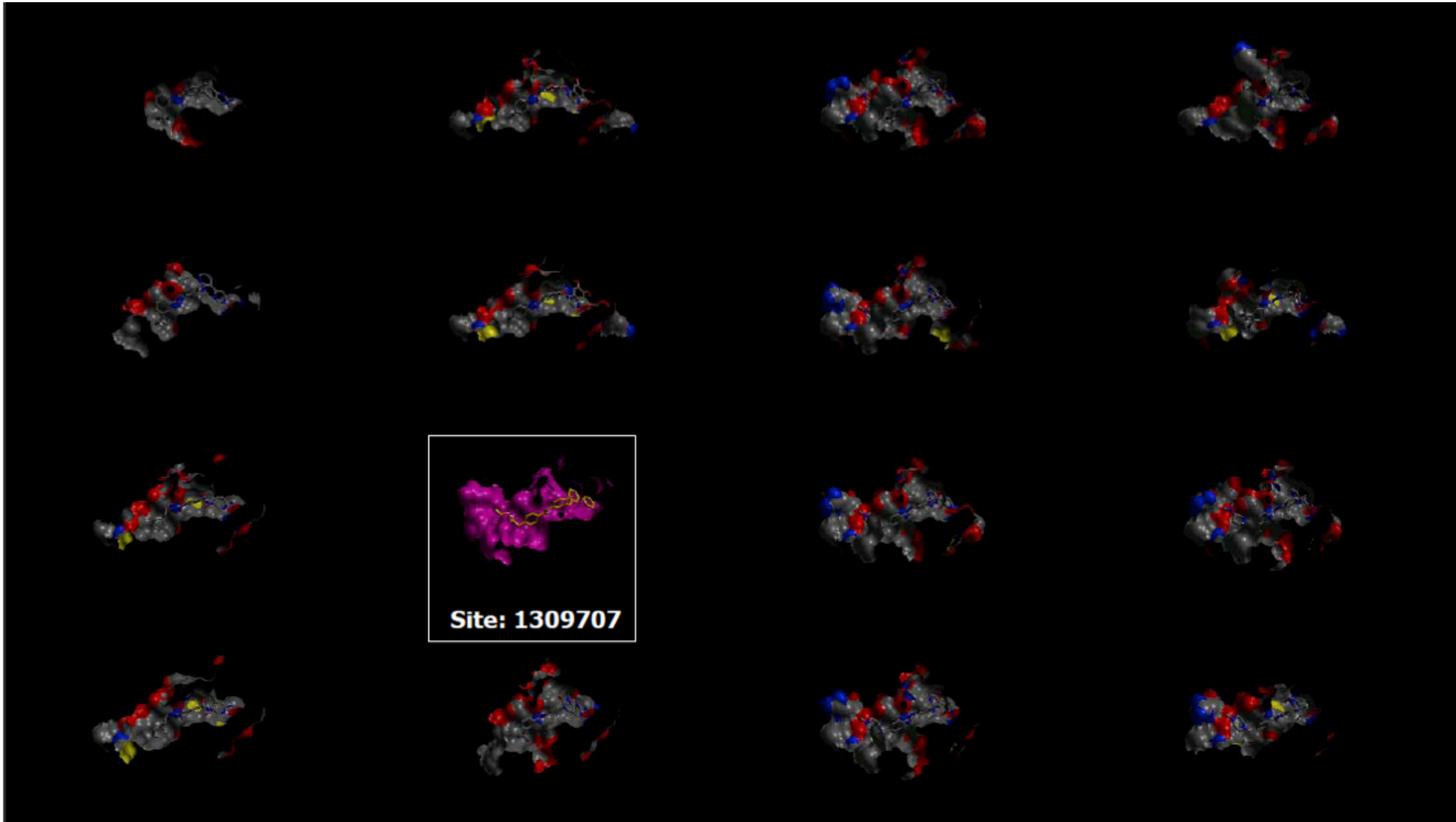
- 1) Issue TIP/LigandSearch
- 2) Issue TIP/SiteSimSearch
- 3) Issue LigandCross
- 4) Filter and locate results in KKB
- 5) Dock and visualize results

# Step 1: Find Co-complexes and Sites

Molecule	ligname	similarity	pdcode	siteid	FourCode	pdblD	pdBbnxNumber	proteinId	title	classification	source	compound	releaseDate	journalTitle	journalReference	exptype
	STI	1	2p0A	1309707	2p0	2p0	1305799	42526	LCK BOUND TO IMATINIB	TRANSFERASE	MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: HUMAN; GENE: LCK; EXPRESSION_SYSTEM: SPODOPTERA FRUGIFERDA; EXPRESSION_SYSTEM_COMMON: FALL ARMYWORM; EXPRESSION_SYSTEM_VECTOR_TYPE: ...	MOL_ID: 1; MOLECULE: PROTO-ONCOGENE TYROSINE-PROTEIN KINASE LCK; CHAIN: A; FRAGMENT: PROTEIN KINASE; SYNONYM: P56-LCK, LYMPHOCYTE CELL-SPECIFIC PROTEIN-TYROSINE KINASE, LSK, T CELL-SPECIFIC PROTEIN-TYROSINE KINASE; EC: 2.7.10.2; ENGINEERED: YES	09-OCT-07	CLASSIFYING PROTEIN KINASE STRUCTURES GUIDES USE OF LIGAND-SELECTIVITY PROFILES TO PREDICT INACTIVE CONFORMATIONS: STRUCTURE OF LCK/IMATINIB COMPLEX	PROTEINS 2007	XRAY DIFFRACTION
	STI	1	2o1qA	1148914	2o1q	2o1q	1125109	26318	STRUCTURE OF CHICKEN C-SRC KINASE DOMAIN IN COMPLEX WITH THE CANCER DRUG IMATINIB.	TRANSFERASE	... ORGANISM_SCIENTIFIC: ... GALLUS; ... M_COMMON: CHICKEN; GENE: SRC; EXPRESSION_SYSTEM: ESCHERICHIA COLI; EXPRESSION_SYSTEM_COMMON: BACTERIA; EXPRESSION_SYSTEM_STRAIN: BL21DE3; EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID; EXPRESSION_SYSTEM_PLASMID: PET28	MOL_ID: 1; MOLECULE: PROTO-ONCOGENE TYROSINE-PROTEIN KINASE SRC; CHAIN: A, B; FRAGMENT: KINASE DOMAIN; SYNONYM: P60-SRC, C-SRC, PP60C-SRC; EC: 2.7.10.2; ENGINEERED: YES	20-MAR-07	C-SRC BINDS TO THE CANCER DRUG IMATINIB WITH AN INACTIVE ABL/C-KIT CONFORMATION AND A DISTRIBUTED THERMODYNAMIC PENALTY.	STRUCTURE V. 15 299 2007	XRAY DIFFRACTION
	STI	1	2hyyA	918207	2hyy	2hyy	904013	16961	HUMAN ABL KINASE DOMAIN IN COMPLEX WITH IMATINIB (ST1571, GLIVEC)	TRANSFERASE	MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: HUMAN; GENE: ABL1; EXPRESSION_SYSTEM: SPODOPTERA FRUGIFERDA; EXPRESSION_SYSTEM_COMMON: FALL ARMYWORM	MOL_ID: 1; MOLECULE: PROTO-ONCOGENE TYROSINE-PROTEIN KINASE ABL1; CHAIN: A, B, C, D; SYNONYM: P150, C-ABL, ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1; EC: 2.7.10.2;	16-JAN-07	STRUCTURAL BIOLOGY CONTRIBUTIONS TO THE DISCOVERY OF DRUGS TO TREAT CHRONIC MYELOGENOUS LEUKAEMIA.	ACTA CRYSTALLOGR. SECT.D V. 63 60 2007	XRAY DIFFRACTION



# Step 2: Find Other Sites vis SiteSimilarity



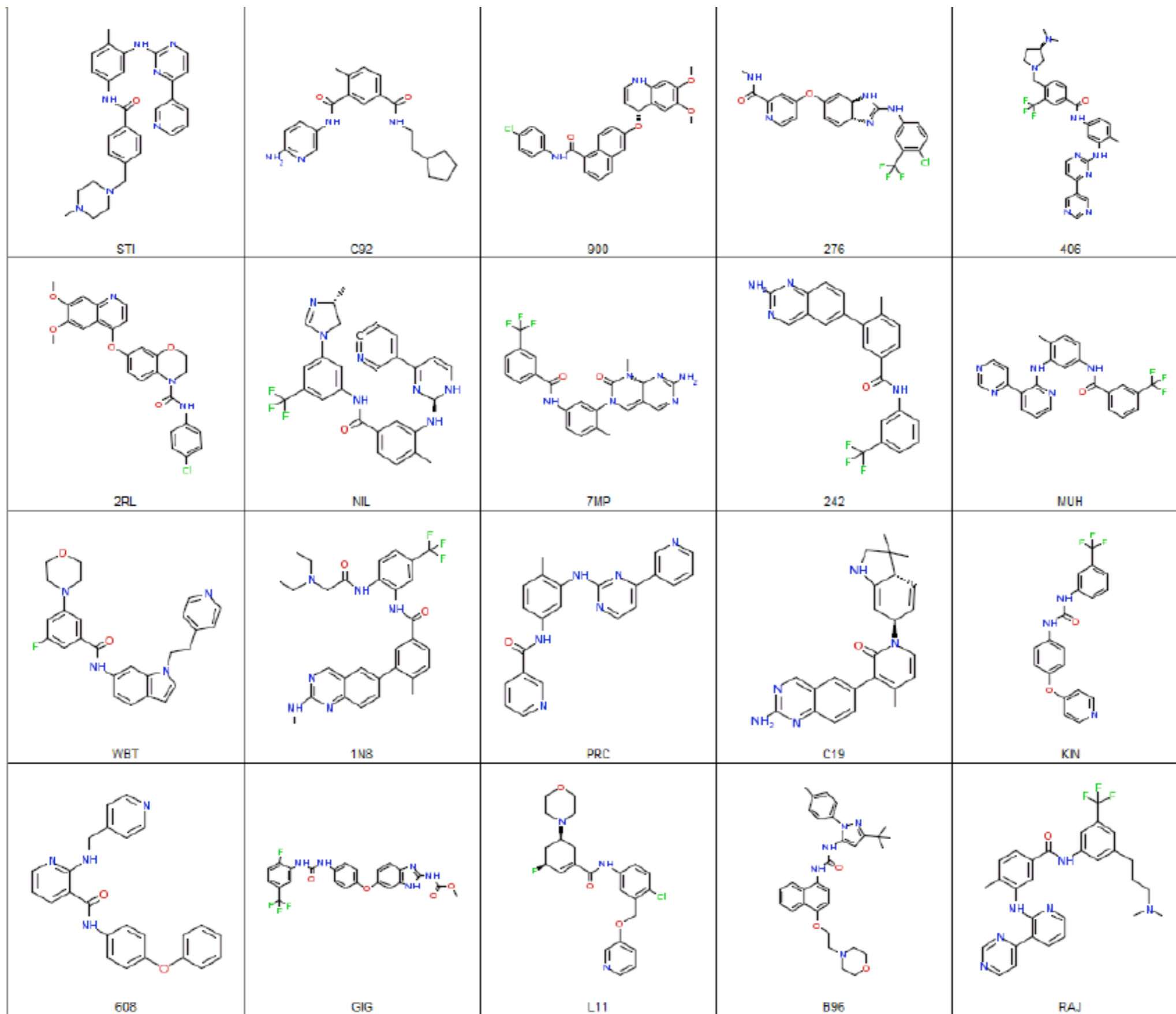
**Site: 1309707**

Chains	Chain Alignments	Sites	Site Alignments	
Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2pl0/s1309707 (chain A)	LCK	STI	100	.L.Y.AYK.E.LM.L.LV.I.TEYM.GS.I.YIHR.L.IADP
pdb2of/s10548 (chain B)	LCK	242	100	.L.V.AVK.E.LM.L.LV.I.TEYM.G.I.Y.H.L.IADP.I
pdb2t15/s1396160 (chain A)	-	2RL	100	.L.G.V.AVK.L.E.I.L.I.VV.V.TEYCKFGH.L.CIH.G.ICDF
pdb2e2b1/s1284639 (chain B)	ABL	406	100	.L.E.V.A.K.E.VM.I.LV.I.TEYMT.G.L.FIHRD.L.VADP

# Example Site Similarity Results (Query: s1309707)

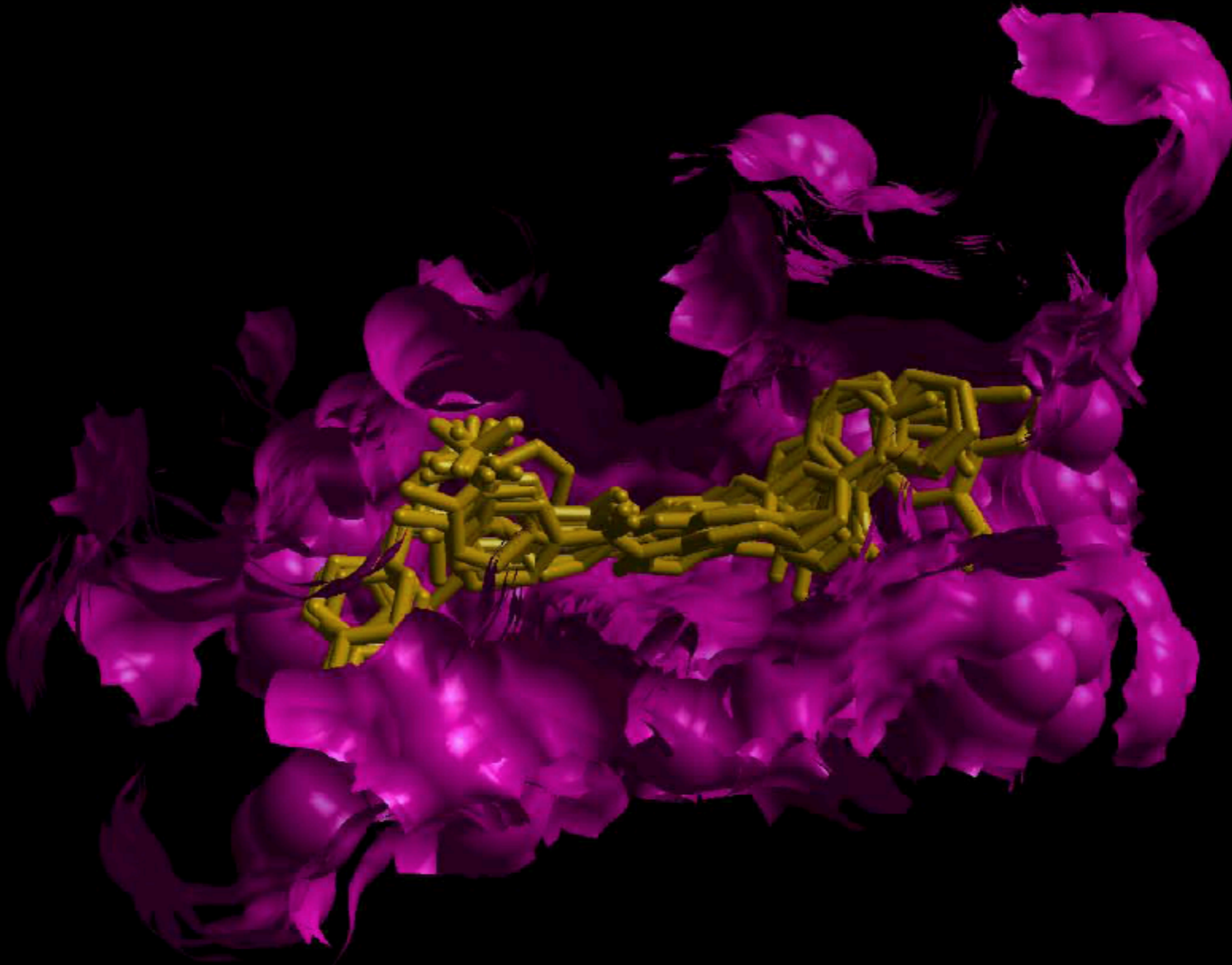
Site	SiteLigand	SiteProtein	SiteScore	ContactScore
1309707	STI	2pl0A	1000	1
1420904	C92	3cpbB	110.906	0.7
1384893	900	3b8qB	121.051	0.67
1322334	276	2qu5A	117.866	0.66
1284638	406	2e2bA	119.18	0.64
1396160	2RL	2rl5A	121.208	0.63
1400124	NIL	3cs9D	111.198	0.62
867405	7MP	2hiwA	101.948	0.61
916548	242	2ofvB	109.214	0.6
1147514	MUH	2oscA	104.115	0.6
776230	WBT	1wbtA	101.635	0.6
916805	1N8	2og8A	116.819	0.59
394066	PRC	1fpuB	107.297	0.57
1415780	C19	3cp9A	104.078	0.56
911671	KIN	2hznA	106.08	0.56
1148488	608	2p2iB	109.41	0.55
1300447	GIG	2oh4A	110.471	0.53
1320735	857	2qu6B	116.424	0.52
437653	B96	1kv2A	107.323	0.52
691631	L11	1w83A	101.268	0.52
1147212	RAJ	2oo8X	104.058	0.52
910098	GIN	2hz0B	108.713	0.51
1396708	P38	3bv2A	124.962	0.51
436174	BMU	1kv1A	88.568	0.5
1412158	G2G	2puuA	118.296	0.5
775147	LI3	1wbvA	85.135	0.5
1415688	C52	3cpcB	102.25	0.48
1431710	GK6	3d83A	104.164	0.48

# Example Ligands from Similar Sites



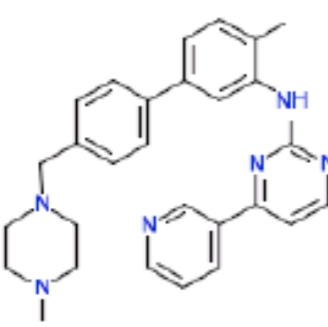
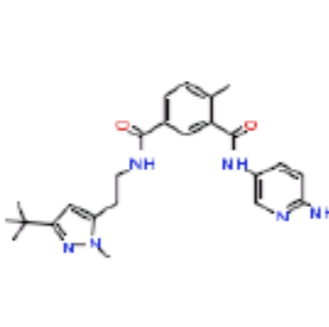
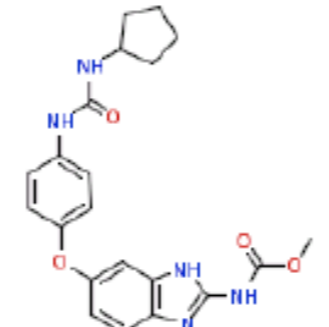
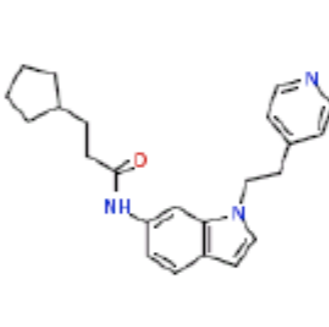
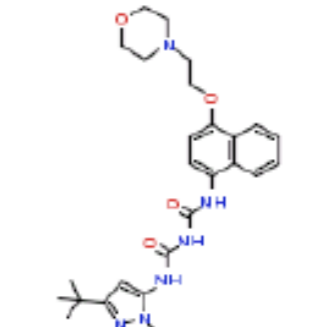
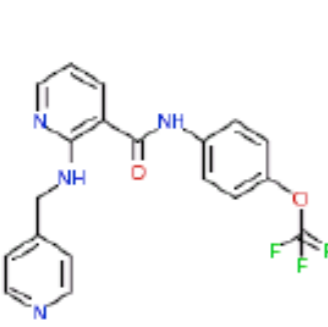
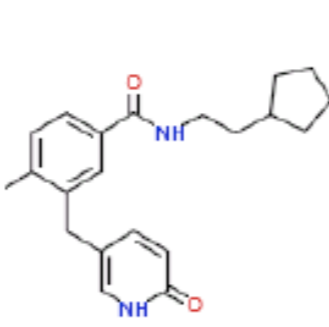
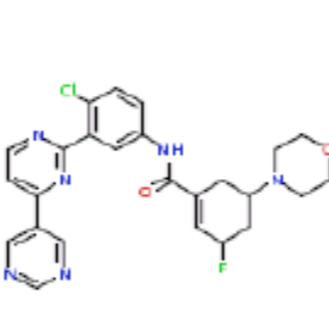
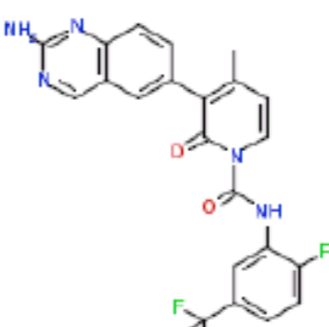
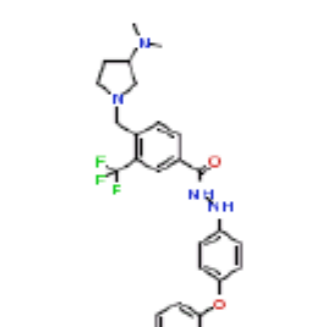
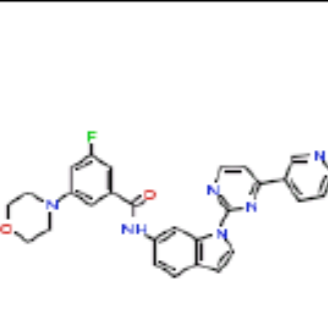
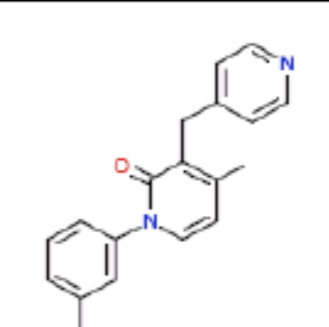
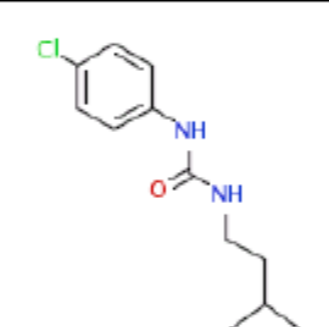
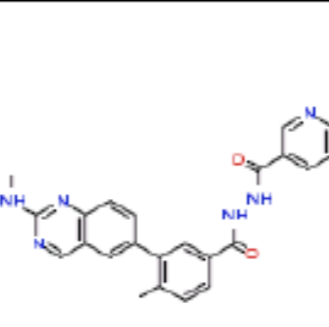
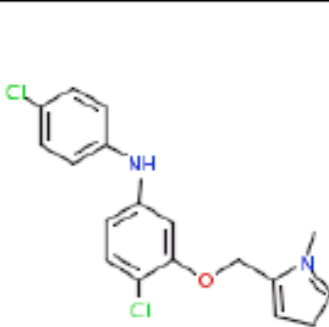
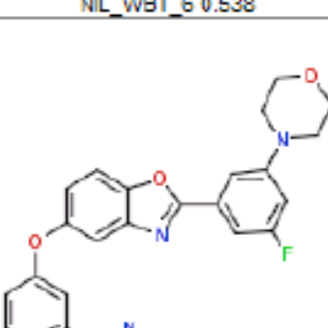
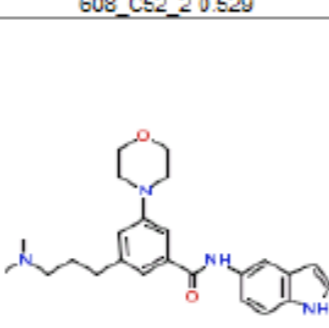
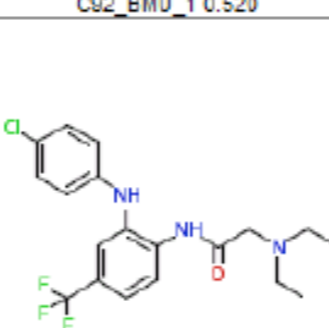
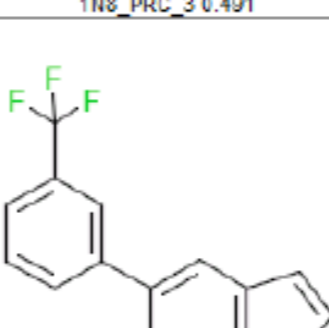
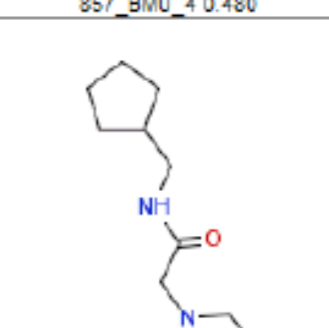


# Step 3: LigandCross: Shuffle Ligand Features from Aligned Sites



Chains	Chain Alignments	Sites	Site Alignments	
Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2a10/s1309707 (chain A)	LCK	STI	100	.L.V.AVE.E.LM.L.LV.I.TEYM.GS.I.YIHR.L.IADF
pdb2a10/s916548 (chain B)	LCK	242	100	.L.V.AVE.E.LM.L.LV.I.TEYM.G.I.V.H.L.IADF.I
pdb2d3/s1395160 (chain A)	-	2RL	100	.LG.V.AVE.L.E.II.I.VV.V.TEECKFGN.L.CIH.L.ICDF
pdb2e2b1/s1264639 (chain B)	ABL	406	100	.L.Y.V.A.K.E.VM.I.LV.I.TEEMT.G.L.PIHRD.L.VADF

# Example LigandCross Results

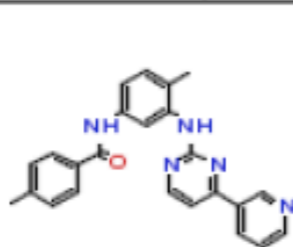
 <p>STI_PRC_2 0.667</p>	 <p>C92_BMU_5 0.635</p>	 <p>C92_GIG_3 0.633</p>	 <p>C92_WBT_1 0.625</p>	 <p>B96_BMU_2 0.623</p>
 <p>608_276_3 0.606</p>	 <p>C92_GN_7 0.606</p>	 <p>406_L11_6 0.577</p>	 <p>GIG_C52_1 0.574</p>	 <p>406_KIN_2 0.545</p>
 <p>NIL_WBT_6 0.538</p>	 <p>608_C52_2 0.529</p>	 <p>C92_BMU_1 0.520</p>	 <p>1N8_PRC_3 0.491</p>	 <p>857_BMU_4 0.480</p>
 <p>857_WBT_2 0.472</p>	 <p>RAJ_LB_1 0.462</p>	 <p>1N8_BMU_2 0.449</p>	 <p>LB_C52_2 0.385</p>	 <p>C92_1N8_1 0.375</p>

# LigandCross Validation

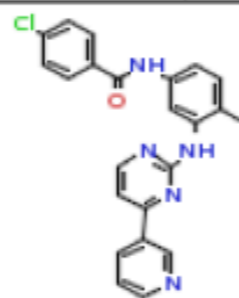
## Kinase Knowledgebase (pIC50)

## Bayesian Model Predictions (PP)

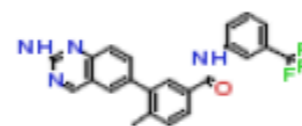
LC-ID	ABL	PDGFR	PDGFRB	JAK3	KDR	LCK	MAPK14	TEK	KIT	RAF1	ABL	PDGFR	PDGFRB	JAK3	KDR	LCK	MAPK14	TEK	KIT	RAF1
G2G_STI_12	6.7	8	8								0.40	0.90	0.76	0.81	0.59	0.15	0.89	0.45	0.70	0.37
900_STI_1	6.1	8	8								0.38	0.91	0.76	0.72	0.55	0.16	0.88	0.42	0.71	0.55
7MP_1N8_4				7.8	9	9.5	8.7				0.36	0.49	0.34	0.32	0.94	1.00	0.95	0.67	0.86	0.39
7MP_1N8_2				6.8	8.3	9.5	9				0.37	0.46	0.31	0.44	0.92	1.00	0.92	0.69	0.84	0.45
7MP_RAJ_3					8.4			8.4			0.35	0.73	0.50	0.49	0.92	0.81	0.86	0.94	0.74	0.37
7MP_GIN_4					7.6						0.16	0.50	0.40	0.82	0.95	0.67	0.70	0.41	0.76	0.51
242_C52_2									7.9		0.30	0.28	0.29	0.74	0.80	0.66	0.74	0.31	1.00	0.43
LI3_L11_1							7.2				0.31	0.73	0.55	0.84	0.74	0.69	0.62	0.36	0.76	0.85
608_GIG_7										6.1	0.28	0.61	0.57	0.69	0.93	0.50	0.60	0.68	0.85	0.50
KIN_BMU_4										6.1	0.31	0.43	0.45	0.78	0.75	0.57	0.77	0.33	0.81	0.25
G2G_KIN_3										6.1	0.25	0.51	0.52	0.75	0.89	0.59	0.64	0.43	0.84	0.43



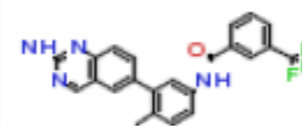
1: G2G\_STI\_12



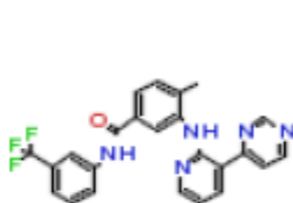
2: 900\_STI\_1



3: 7MP\_1N8\_4



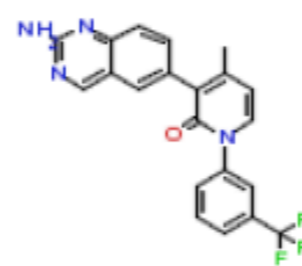
4: 7MP\_1N8\_2



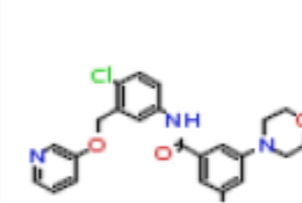
5: 7MP\_RAJ\_3



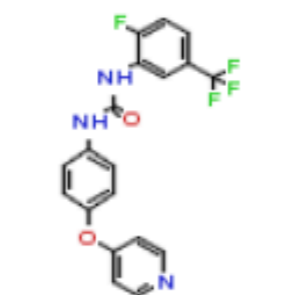
6: 7MP\_GIN\_4



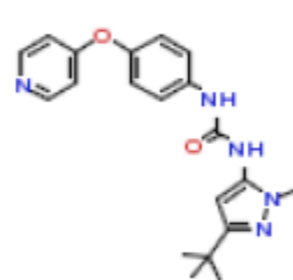
7: 242\_C52\_2



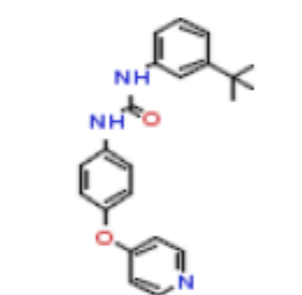
8: LI3\_L11\_1



9: 608\_GIG\_7

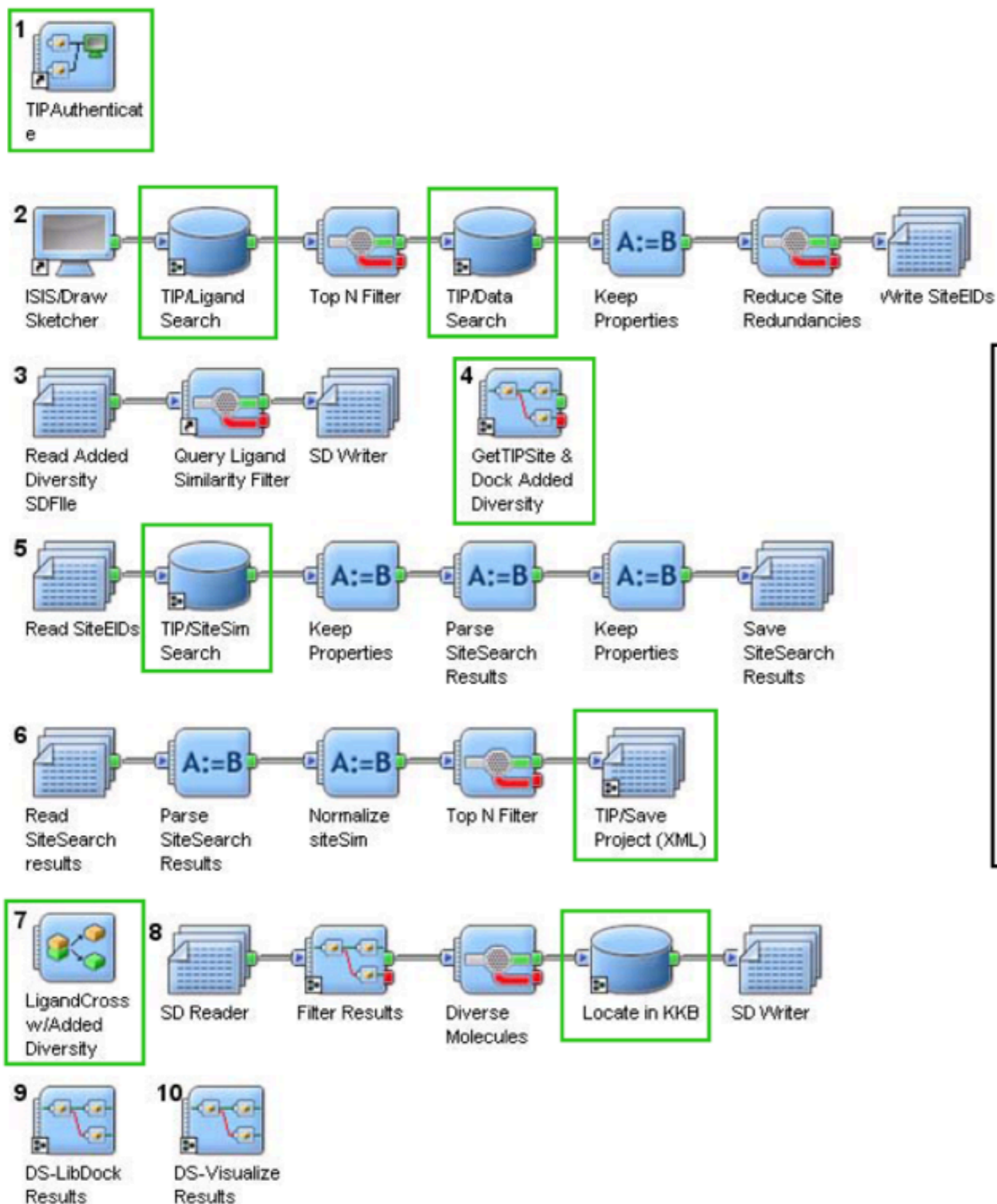


10: KIN\_BMU\_4



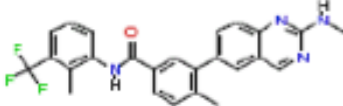
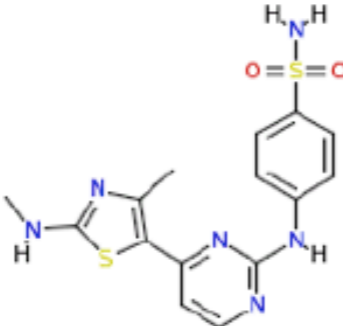
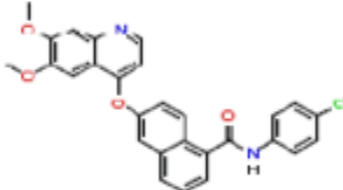
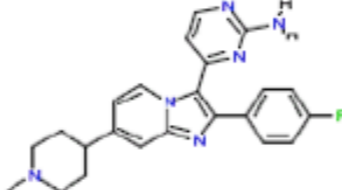
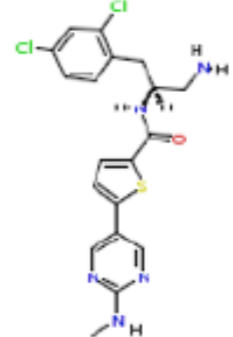
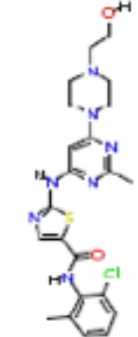
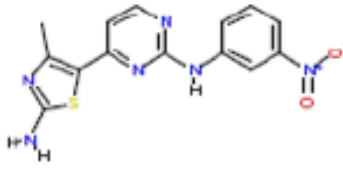
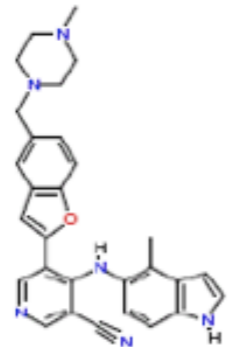
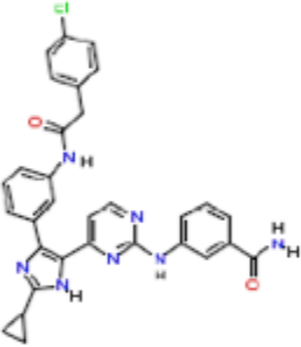
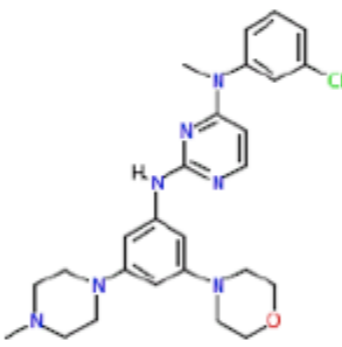
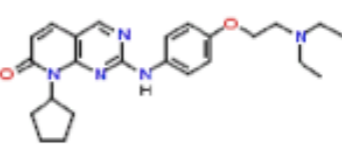
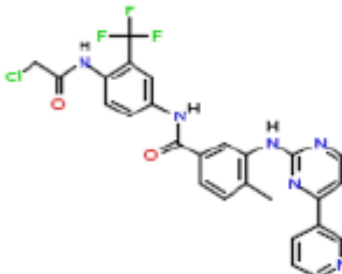
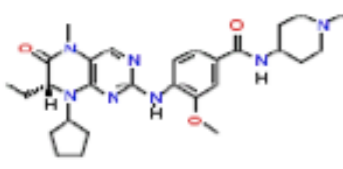
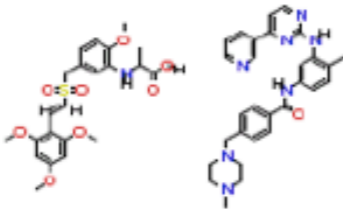
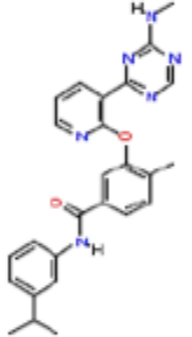
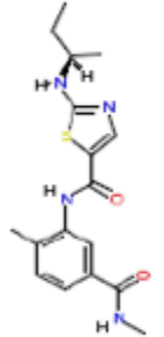
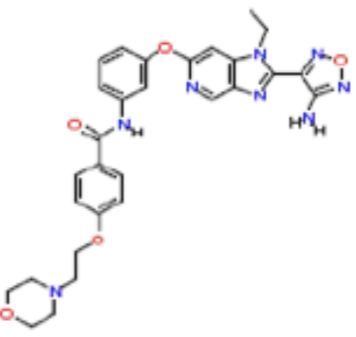
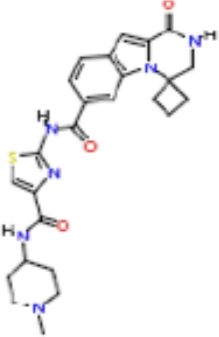
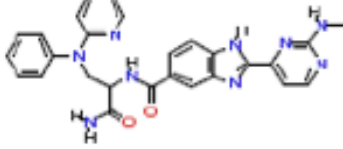
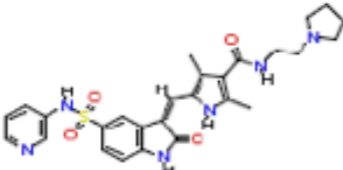
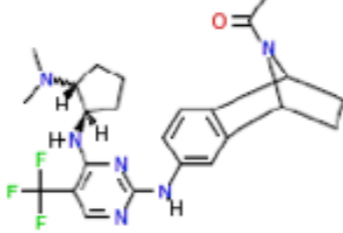
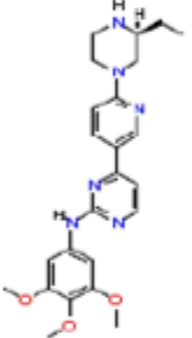
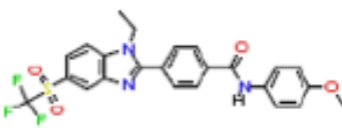
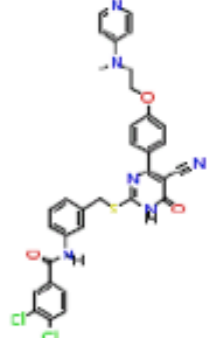
11: G2G\_KIN\_3

# Enhance LigandCross with Added Diversity

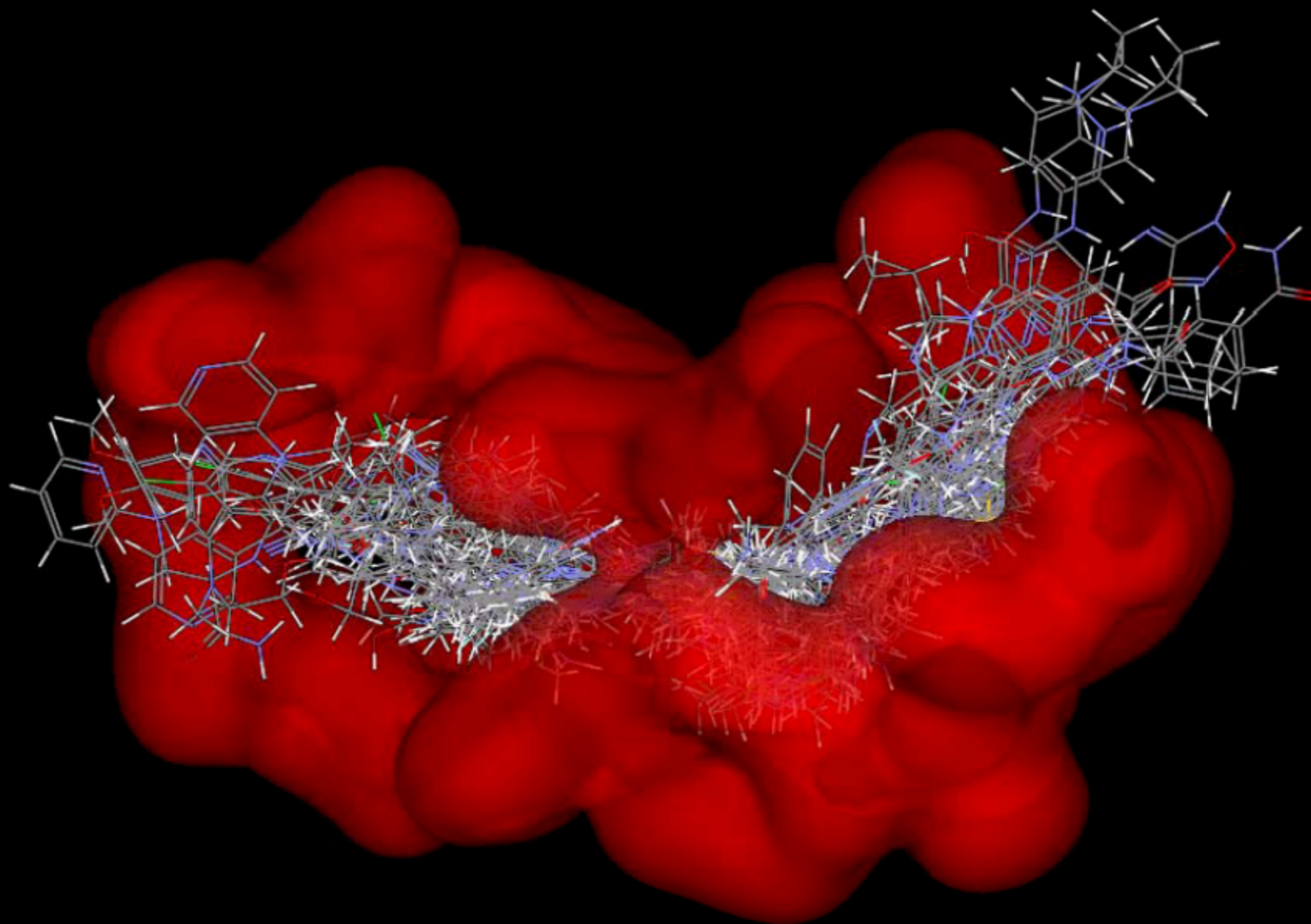


- > Issue TIP/LigandSearch
- > Identify/Dock "AddedDiversity"
- > Issue TIP/SiteSimSearch
- > LigandCross w/AddedDiversity
- > Filter and locate results in KKB
- > Dock and visualize results

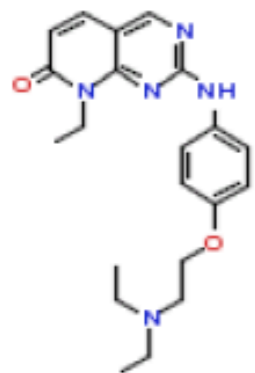
# Example Potent Kinase Inhibitors (From KKB)

 4336533 LCK pval: 11.00	 4302493 CDK9 pval: 10.54	 4332561 KDR pval: 10.52	 4318145 PKG pval: 10.40	 4336686 FKA pval: 10.00	 4272835 ABL1 pval: 10.00
 894611 CDK2 pval: 9.70	 4358585 PRKCG pval: 9.70	 4363734 RAF1 pval: 9.30	 4369892 EPHB4 pval: 9.24	 809 CDK4 pval: 9.15	 4374385 FDGFRA pval: 9.14
 4366691 PLK1 pval: 9.10	 4301886 BCR_AEL pval: 9.08	 4307551 TEK pval: 9.00	 4363016 MAPK11 pval: 8.82	 4343448 ROCK1 pval: 8.74	 4363247 MAPKAPK2 pval: 8.70
 4291996 IKB pval: 8.70	 4208857 FAK2 pval: 8.22	 4373725 PTK2B pval: 8.22	 1788 ZAP70 pval: 8.10	 2425813 PTPN9 pval: 5.96	 4303129 MAP3K2 pval: 4.70

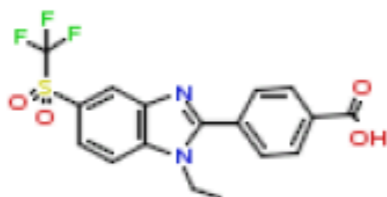
# Potent Kinase Inhibitors Docked (s1309707)



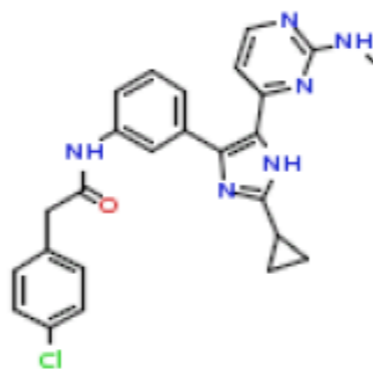
# Added Diversity LigandCross Validation



4343448\_809\_27



4272835\_2425813\_23



4363734\_4291996\_2

4343448\_809\_27:

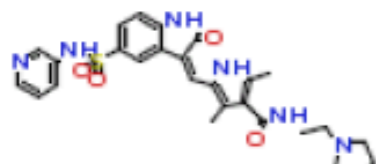
CDK4: 6.80 CDK2: 5.63 CDK2: 6.12 CDC2: 5.58 CSK: 5.99 CDK5: 6.81  
CDK4: 6.80 CDK2: 5.63 CDK2: 6.12 CDC2: 5.58 CDK4: 6.80

4272835\_2425813\_23:

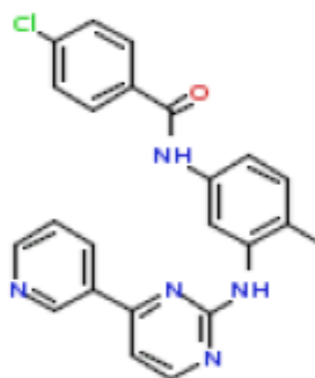
PTPN1: 4.24 PTPRA: 4.21

4363734\_4291996\_2:

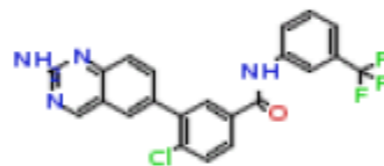
RAF1: 9.00 MAPK1: 5.29 BRAF: 8.05 BRAF: 8.52



4208857\_4208857\_1



900\_STI\_1



242\_A96\_5

4208857\_4208857\_1:

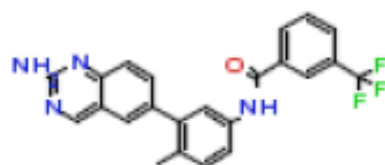
FAK2: 8.22 KDR: 5.86 PDGFRB: 4.90 EGFR: 4.17 ERBB2: 5.23

900\_STI\_1:

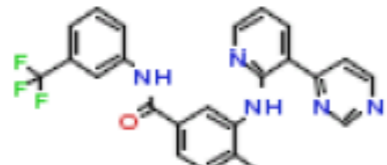
PDGFR: 8.00 PDGFR: 8.00 ABL: 6.10 PDGFRB: 8.00 PDGFR: 8.00  
ABL: 6.10

242\_A96\_5:

LCK: 9.40



242\_MUH\_1



242\_MUH\_2



406\_STI\_1

242\_MUH\_1:

LCK: 9.40 TEK: 7.68 KDR: 8.22 MAPK14: 9.00 JAK3: 6.81

242\_MUH\_2:

KDR: 8.40 TEK: 8.40 TEK: 8.40 KDR: 8.40 TEK: 8.40 KDR: 8.40

406\_STI\_1:

BCR\_ABL: 8.40 BCR\_ABL: 5.30 LYN: 8.06 ABL1: 8.07 ABL1: 8.40

# Conclusions

- Significant receptor-site similarities exist within and across target families
- The structurally resolved and modelable proteome is a very rich source for new matter ideas
- LigandCross can be an effective approach to generating novel, bioactive matter using co-complexes, known inhibitors, and/or fragment-based information.

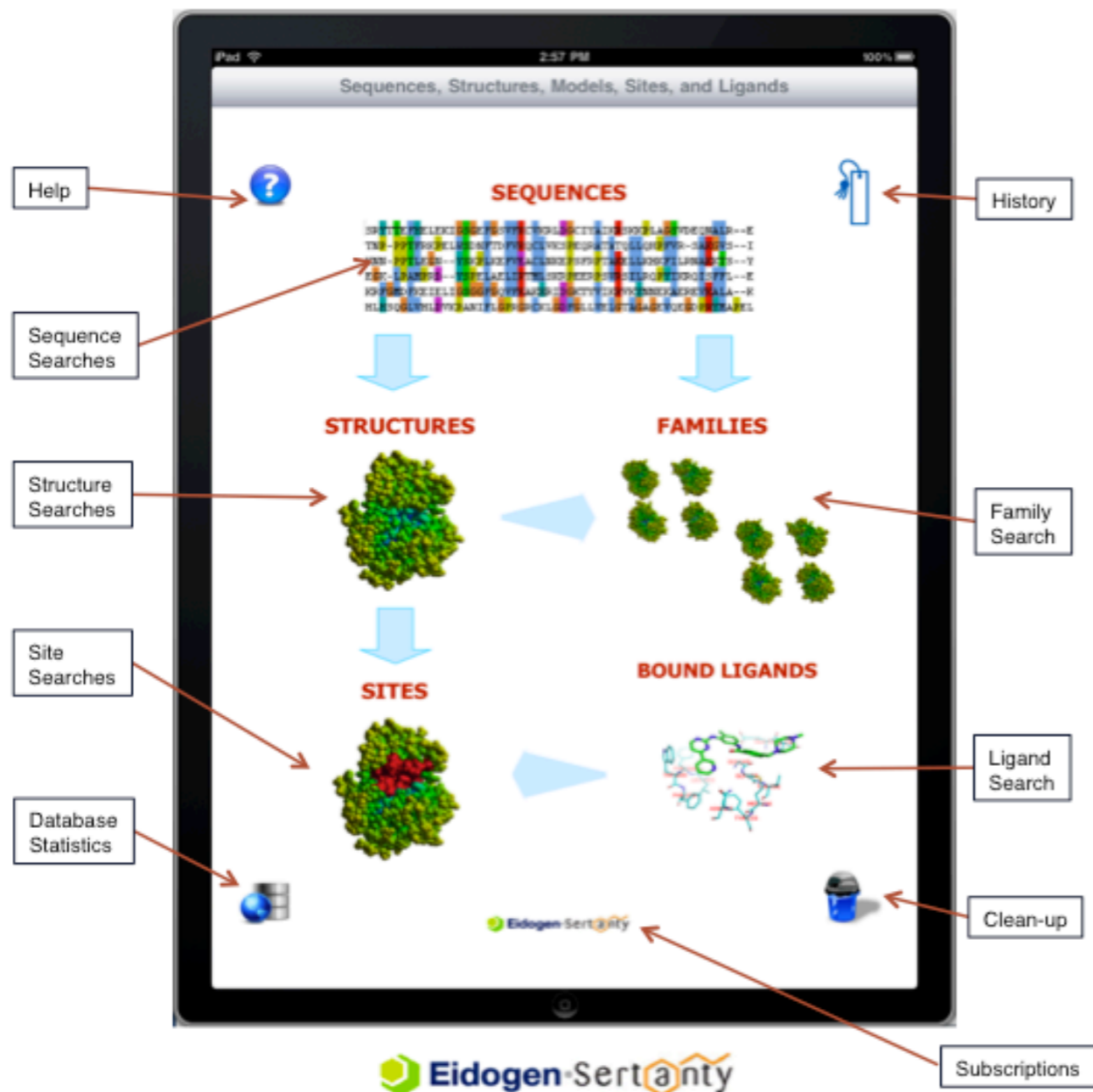


# iProtein - 8/10 (v1) ... 5/11 (v4)



## Accessing TIP content through iProtein

Through iProtein, the TIP database can be surveyed by Sequence, Structure/Model, Site, head-nodes (i.e. protein family), and by bound ligand structure searches. Simply click on any image in the main page to initiate a search. Future versions of iProtein may enable more complex searching – e.g. protein structure search, site-search, etc. So stay tuned....



## Structure/Model Searches (cont)

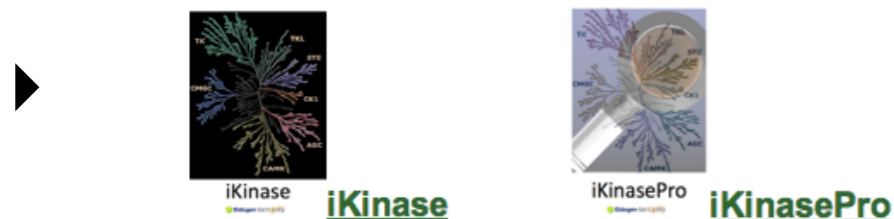
Click on the arrow (when present) to see more data. You can click the arrow (">") to see more detail. If there is a bound ligand, you can initiate ligand-based searches.



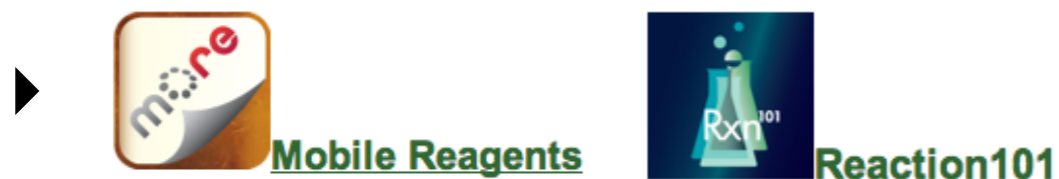


# Acknowledgements

- Dr. Rajan Sharma and Prof. Stephan Schurer



- Dr. Maurizio Bronzetti



- Dr. Alex Clark

- ▶ MMDSLlib:



- Dr. Tony Yuan

- ▶ Chemene:



# iProtein iPad Demo

