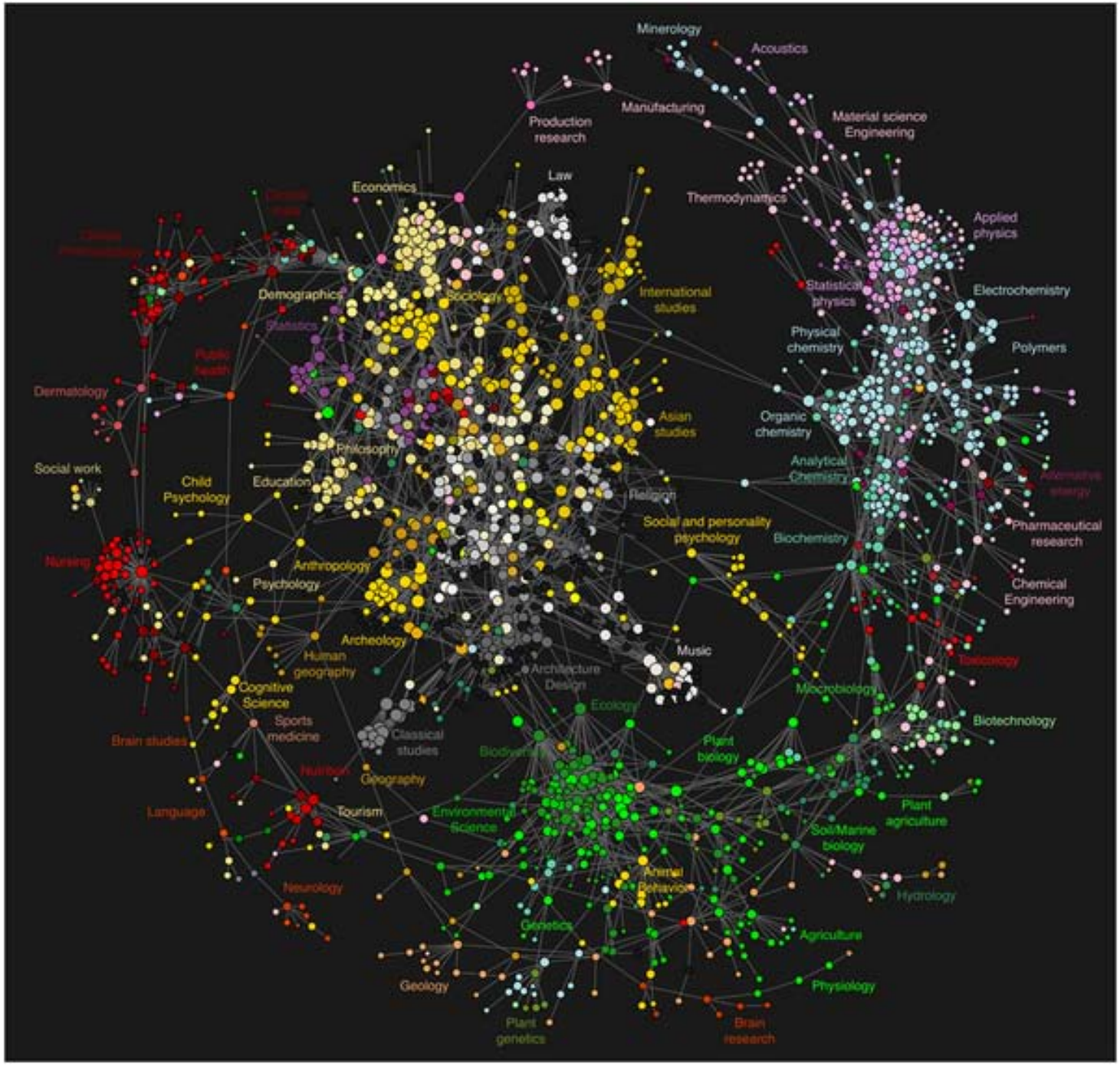
Using knowledgebases of structure-activity-data, receptor-site and protein structural similarity to generate new matter ideas

Steven Muskal, Ph.D.

Chief Executive Officer
Eidogen-Sertanty, Inc.

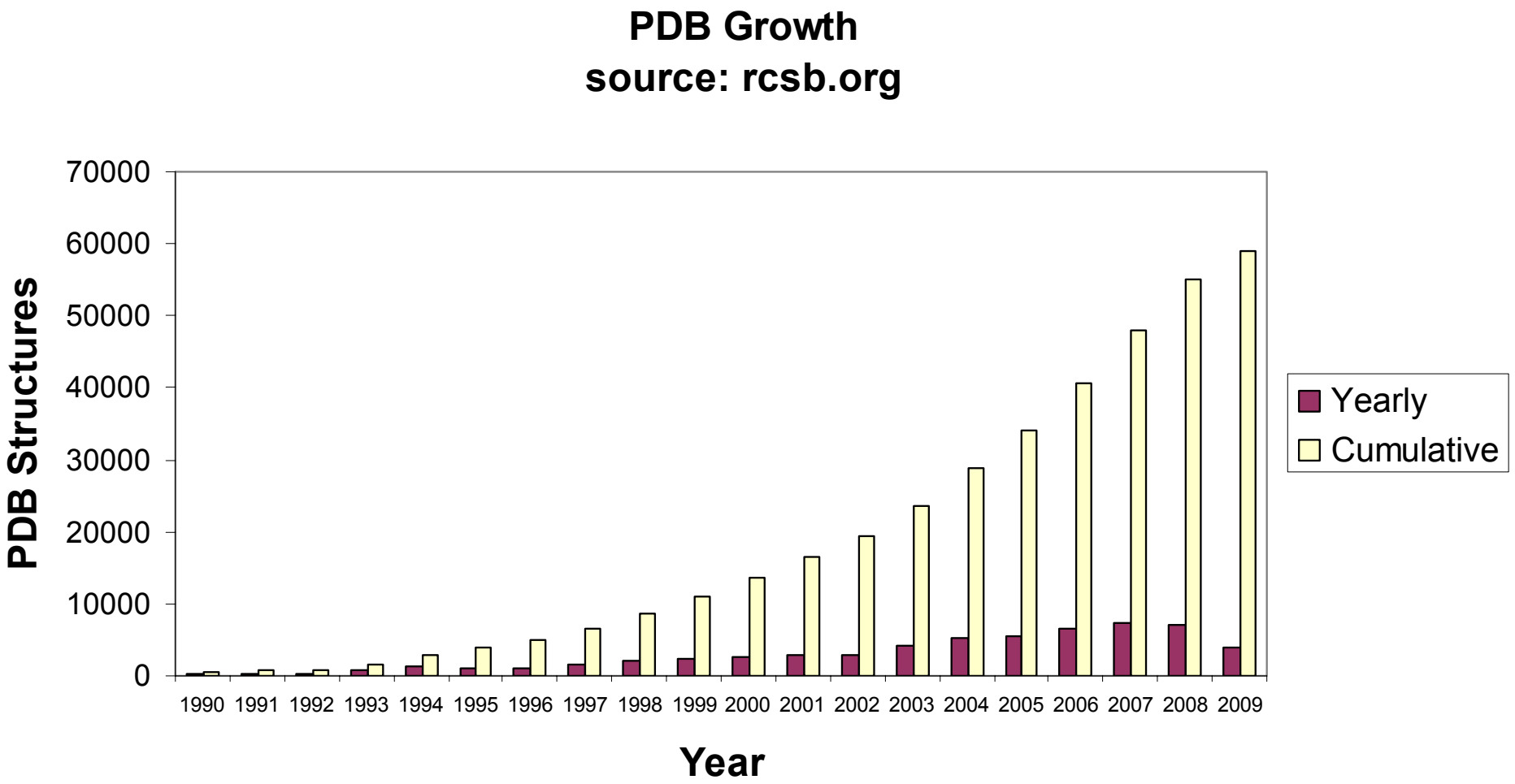
smuskal@eidogen-sertanty.com

Bollen J (March 2009) - Clickstream Data Yields High-Resolution Maps of Science



Protein Structure Growth Continues

~ 60K Structures/co-complexes (July-2009)
> 600 deposits per month → **>150/week!**

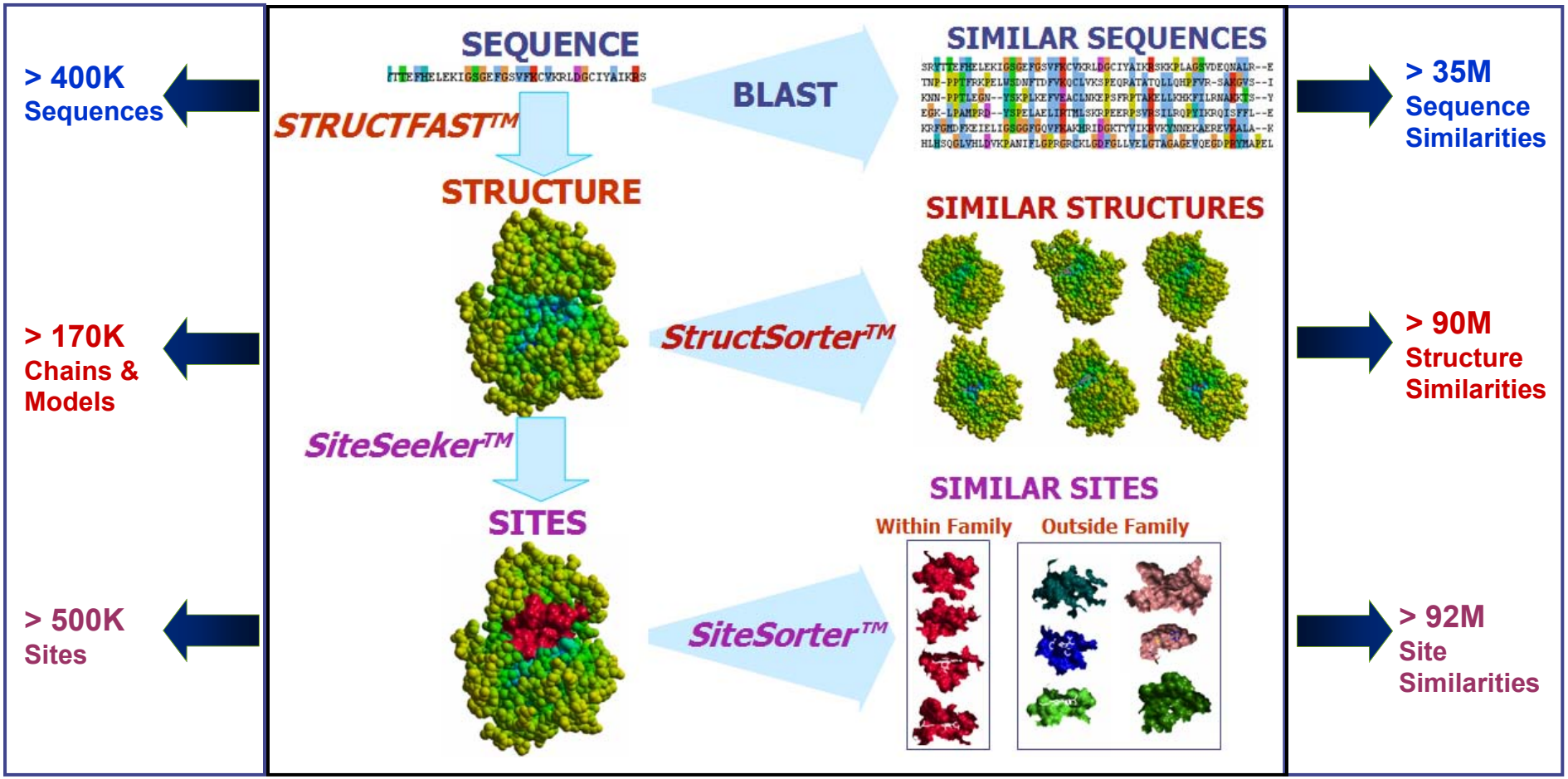


Drugs Developed using Structural Knowledge

Inhibitor/Drug	Disease	Company(s)	Protein targeted	Enzyme Family
STI-571/Gleevec	Chronic Myeloid Leukemia	Novartis	c-Abl kinase	Tyrosine kinase
Fluoroquinolone/Ciprofloxacin	Bacterial infection	Bayer	Gyrase	ATP Hydrolase
Saquinavir/Invirase, Ritonavir/Norvir, Indinavir/ Crixivan, Nelfinavir/Viracept, Amprenavir/Agenerase, Fosamprenavir/Lexiva,	AIDS	Roche, Abbott, Agouron, Merck, Vertex	HIV-1 Protease	Aspartylprotease
Trusopt	Glaucoma	Merck	Carbonic Anhydrase	Lyase
Thymitaq	Cancer	Agouron	Thymidylate synthase	Methyl transferase
Celecoxib/Celebrex, Rofecoxib/Vioxx	Inflammation, rheumatoid arthritis	Searle, Merck	Cox-2	Oxidoreductase
AG3340/Prinomastat	Cancer	Agouron	Matrix metalloprotease	Metalloprotease
Oseltamivir phosphate/Tamiflu, Zanamivir/Relenza	Influenza	Roche	Neuraminidase	Glycosidase

Source: http://www.active-sight.com/science/sbdd.html

TIP Content and Algorithm Engine

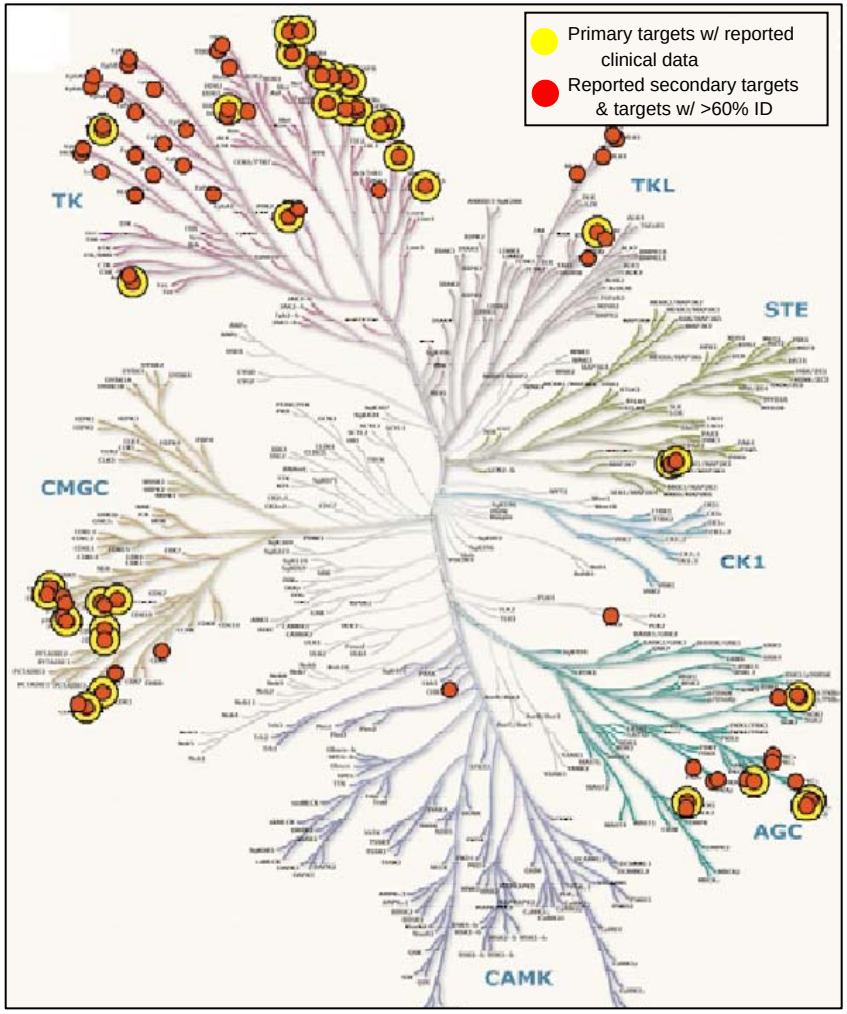


- Interrogating the druggable genome with structural informatics **MolecularDiversity (2006)**
- STRUCTFAST: Protein Sequence Remote Homology Detection and Alignment Using Novel Dynamic Programming and Profile-Profile Scoring **Proteins. 2006 64:960-967**
- StructSorter: A Method for Continuously Updating 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.**

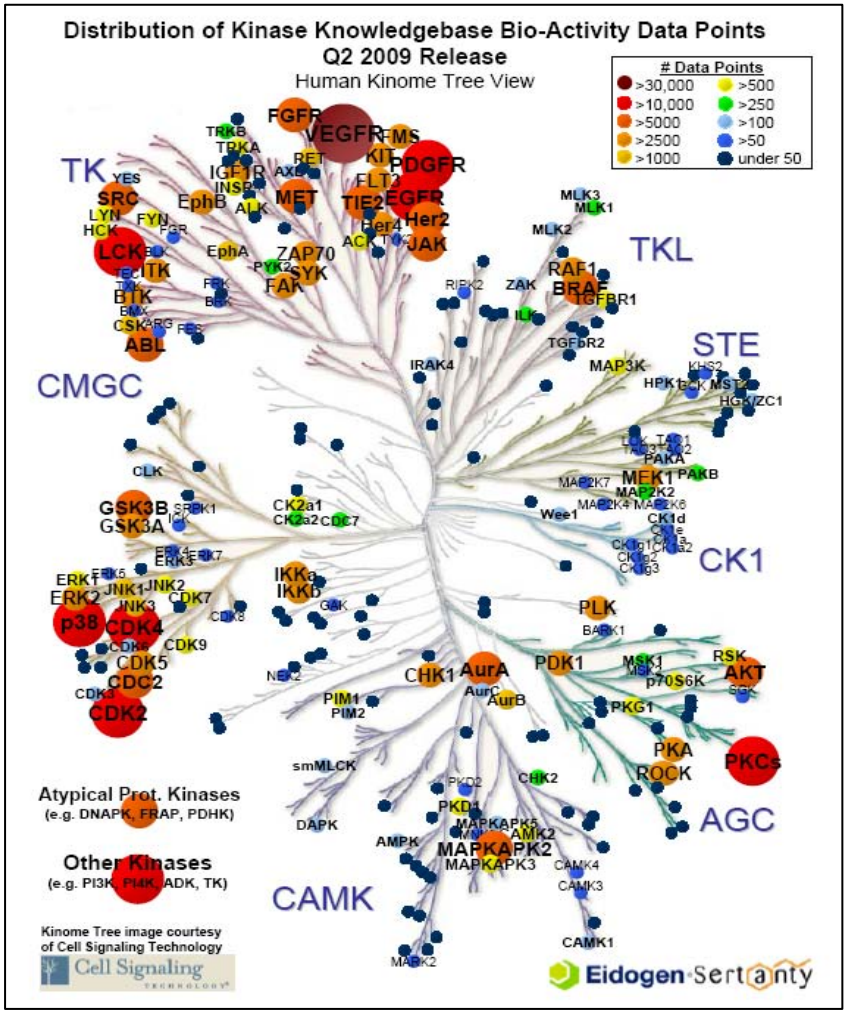
Eidogen-Sertanty

Kinase SAR Knowledgebase (KKB) – Hot Targets

Kinase Targets of Clinical Interest
from Vieth *et al*. *Drug Disc. Today* **10**, 839 (2005).



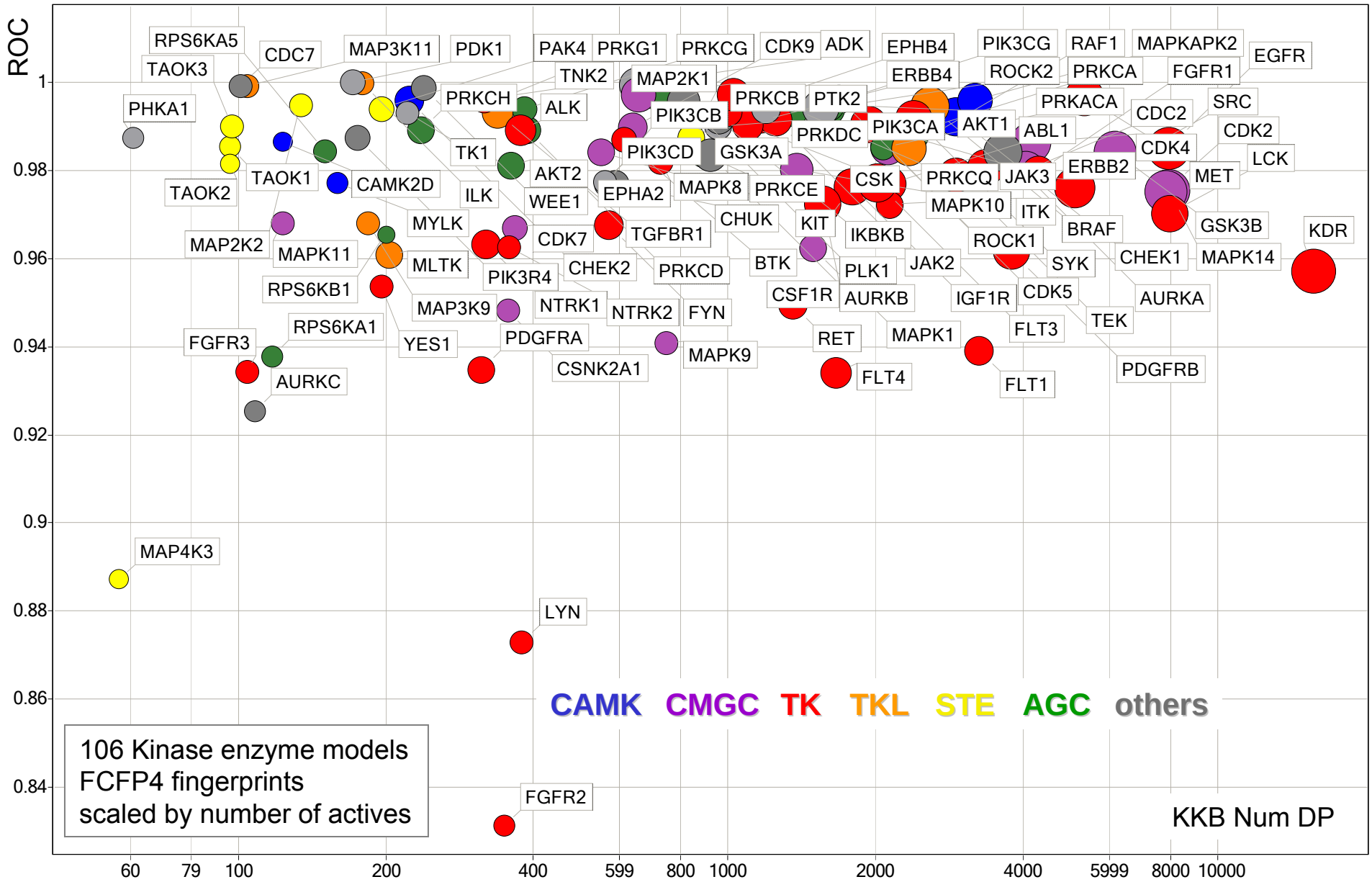
Eidogen-Sertanty KKB
SAR Data Point Distribution



> 415,000 SAR data points curated from
> 5800 journal articles and patents

Eidogen Sertanty

Kinase SAR Naïve Bayse Models



Tools for Patent Intelligence

Define query conditions for patents

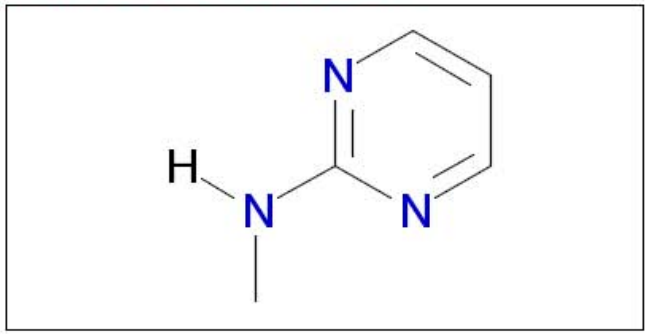
Assay Target:

GSK3B

Select Patent Assignee:

Thouin, Eryk
VERTEX PHARMACEUTICALS INC.
VERTEX PHARMACEUTICALS INCORPORATED
Vertex Pharmaceutical Incorporated
Vertex Pharmaceuticals Incorporated
YUYU Inc.

and optionally filter by **substructure**:



Search

Tools for Patent Intelligence

15 Patent(s) for the queried conditions:

Assay target = GSK3B and Patent assignee in (VERTEX PHARMACEUTICALS INC. , VERTEX PHARMACEUTICALS INCORPORATED , Vertex Pharmaceutical Incorporated , Vertex Pharmaceuticals Incorporated)

Patent	assignee	#molecules	Include in SAR analysis	Representative Compounds
Thiazole Compounds Useful as Inhibitors of Protein Kinase	VERTEX PHARMACEUTICALS INC.	95	include	Cluster 1 (30) Cluster 2 (15) Cluster 3 (10) Cluster 4 (6) Cluster 5 (4)
Pyrazole Compounds Useful as Protein Kinase Inhibitors: Triazole Compounds Useful as Protein Kinase Inhibitors	Vertex Pharmaceuticals Incorporated	270	include	Cluster 1 (45) Cluster 2 (28) Cluster 3 (19) Cluster 4 (14) Cluster 5 (13)
Pyrazole Compounds Useful as Protein Kinase Inhibitors	VERTEX PHARMACEUTICALS INCORPORATED	261	include	

Heterocyclic Inhibitors of ER and Uses Ther

Compositions Useful as Inhib

Heteroaryl Compounds Us as Inhibitors of GSK-3

SAR ana

OPEN TABLE OF PROTOCOLS | Export to SDfile | Export to Accord adf

SAVE PDF FILE | NEW SEARCH BY REFERENCE IDENTIFIER

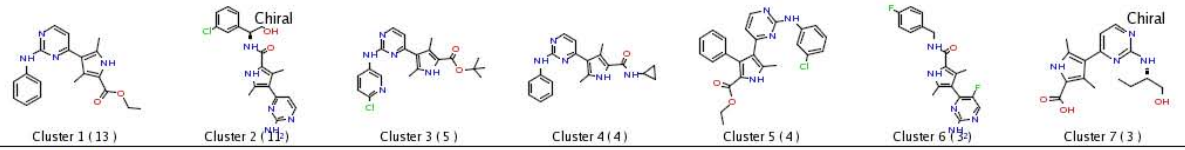
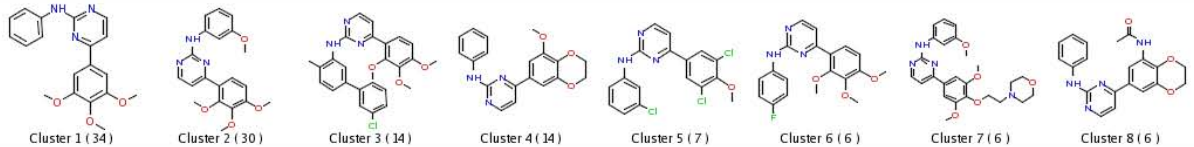
Search for Reference Identifier = 10000061

Reference Information	
PATENT	Thiazole Compounds Useful as Inhibitors of Protein Kinase
PATENT_ABSTRACT	The thiazole compounds of present invention are inhibitors of protein kinases, particularly inhibitors of GSK3, Aurora2, and Syk mammalian protein kinases. The invention also provides pharmaceutically acceptable compositions comprising the compounds of the invention and methods of utilizing those compounds and compositions in the treatement of various protein kinase mediated disorders.
PATENT_KEYWORD	Aurora2, GSK3, Syk, AURKA,thiazol derivatives, serine/threonine kinase
PATENTNUMBER	US20030119856A106/26/2003,WO02096905A112/05/2002
INVENTORS	John Cochran(4 Philips Farm Road, Marshfield, MA 02050 (US).),Suganthini Nanthakumar(253 Spiers Road, Newton, MA 02459 (US).),Edmund Harrington(460 East 8th Street, South Boston, MA 02127 (US).),Jian Wang(25 South Point Drive, Boston, MA 02125 (US).)
ASSIGNEE	VERTEX PHARMACEUTICALS INC.(Andrews S. Marks, 130 Waverly Street, Cambridge, MA02139-4242 (US).)

#_Molecule	mrid	Molecule	MW	Protocol Specification	Biological Activity	Assay Type	Assay Species	Assay CellLines	Assay Targets	Assay Group
1	4193766		396.41	GSK3B Kinase Inhibition (10000121) SYK Kinase Inhibition Assay (10000118) Aurora2 Kinase Inhibition Assay (10000117)	Ki < 0.1 uM IC50 < 0.5 uM IC50 < 0.5 uM	Enzyme Assay Enzyme Assay Enzyme Assay			GSK3B SYK AURKA	
2	4194262		344.43	GSK3B Kinase Inhibition (10000121)	Ki < 0.1 uM	Enzyme Assay			GSK3B	
3	4194335		398.40	GSK3B Kinase Inhibition (10000121)	Ki < 0.1 uM	Enzyme Assay			GSK3B	
4	4195808		298.36	GSK3B Kinase Inhibition (10000121)	Ki < 0.1 uM	Enzyme Assay			GSK3B	

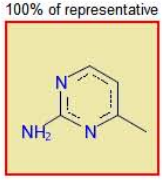
Drug Discovery Informatics Section

Tools for Patent Intelligence

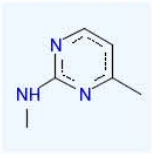
Patent	assignee	#molecules	Representative Compounds in selected patents
Compositions Useful as Protein Kinase Inhibitors	Vertex Pharmaceuticals Incorporated	67	 Cluster 1 (13) Cluster 2 (11) Cluster 3 (5) Cluster 4 (4) Cluster 5 (4) Cluster 6 (3) Cluster 7 (3)
Inhibitors of c-Jun N-Terminal Kinases (JNK) and Other Protein Kinases	Vertex Pharmaceuticals Incorporated	165	 Cluster 1 (34) Cluster 2 (30) Cluster 3 (14) Cluster 4 (14) Cluster 5 (7) Cluster 6 (6) Cluster 7 (6) Cluster 8 (6)

Select among one of the detected common scaffolds for these representative compounds or enter your own scaffold query for SAR analysis:

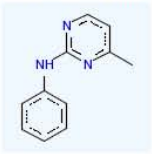
100% of representative



80% of representative



60% of representative



Scaffold query:

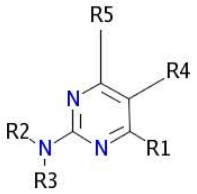
SAR analysis

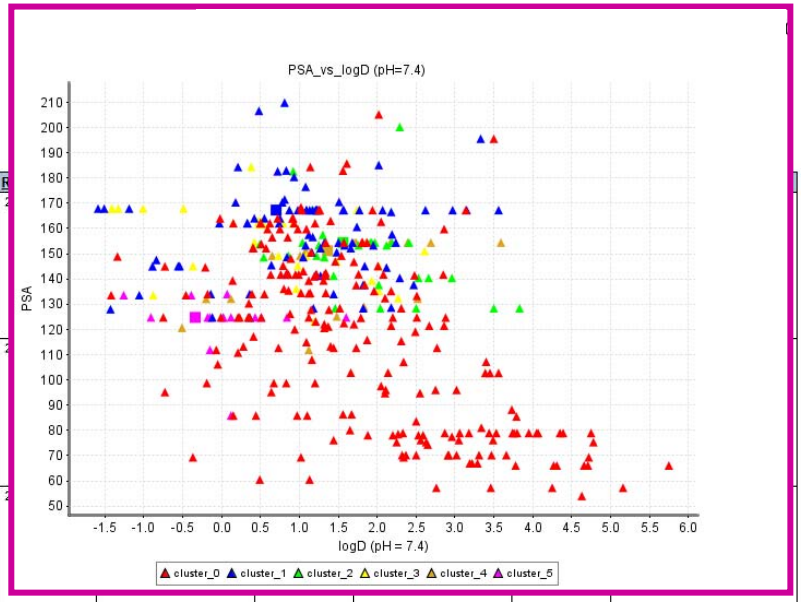
Tools for Patent Intelligence

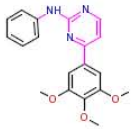
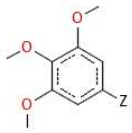
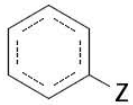
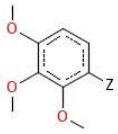
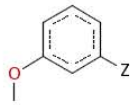
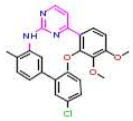
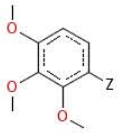
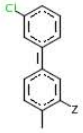
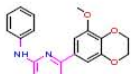
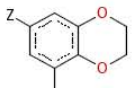
SAR Table for core in patents

223 compounds in patent with this scaffold series structure

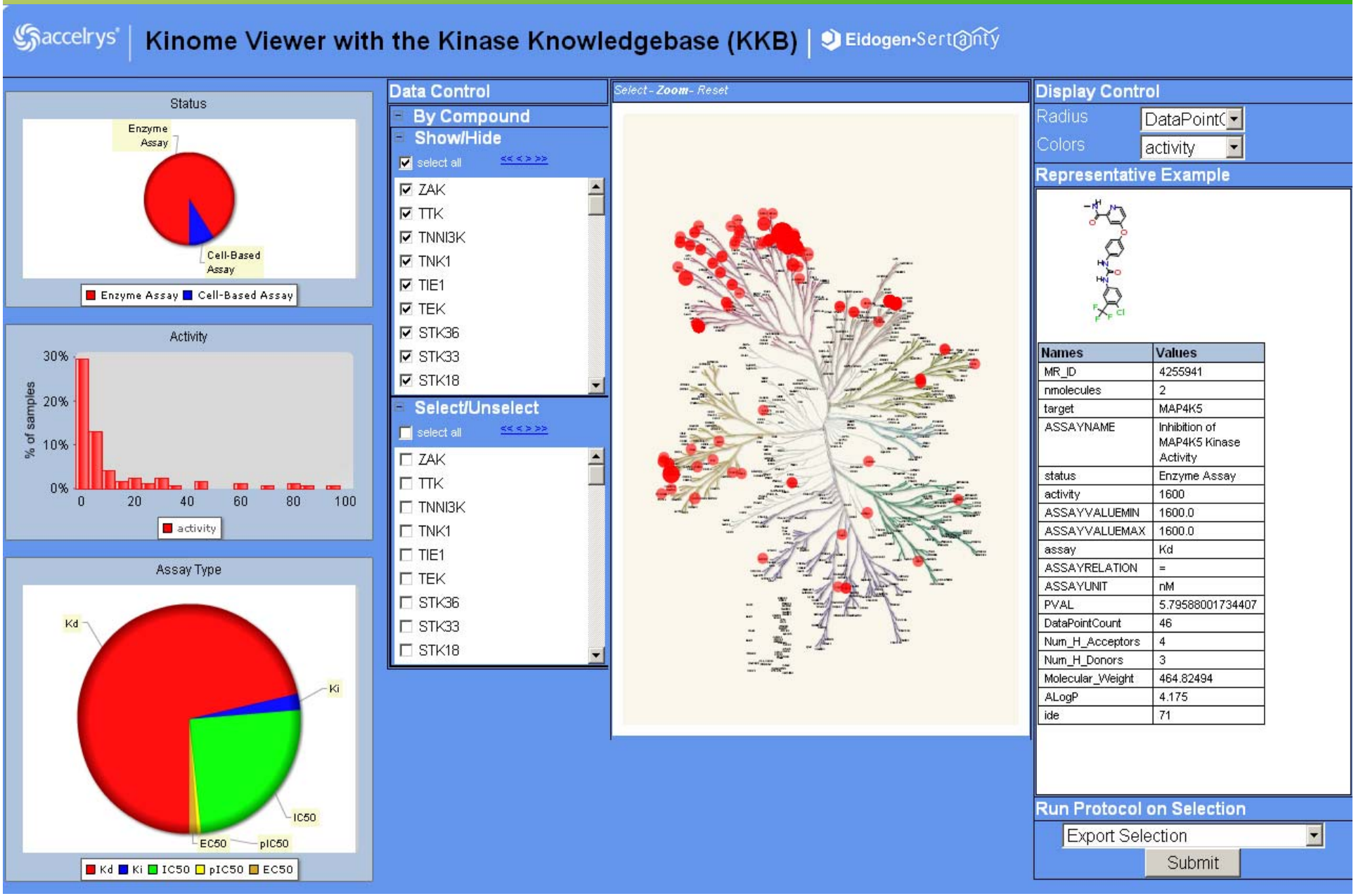
Core Structure





Record	Representative	Cluster	activity	Parent Molecule	R1_Frequency	R1_Group	R2_Frequency	R2_Group	R					
1	YES (34)	1	Ki < 0.1 uM Ki > 5 uM		47		17		2					
2	YES (30)	2	Ki between 0.1 and 1 uM		63		11		2					
3	YES (14)	3	Ki between 0.1 and 1 uM Ki > 5 uM		63		1		2					
4	YES (14)	4	Ki < 0.1 uM inhibition between 24 and 66 % (@ 2 uM)		9		17		220		210		211	

Accelrys Kinome Viewer with Eidogen-Sertanty KKB



Science

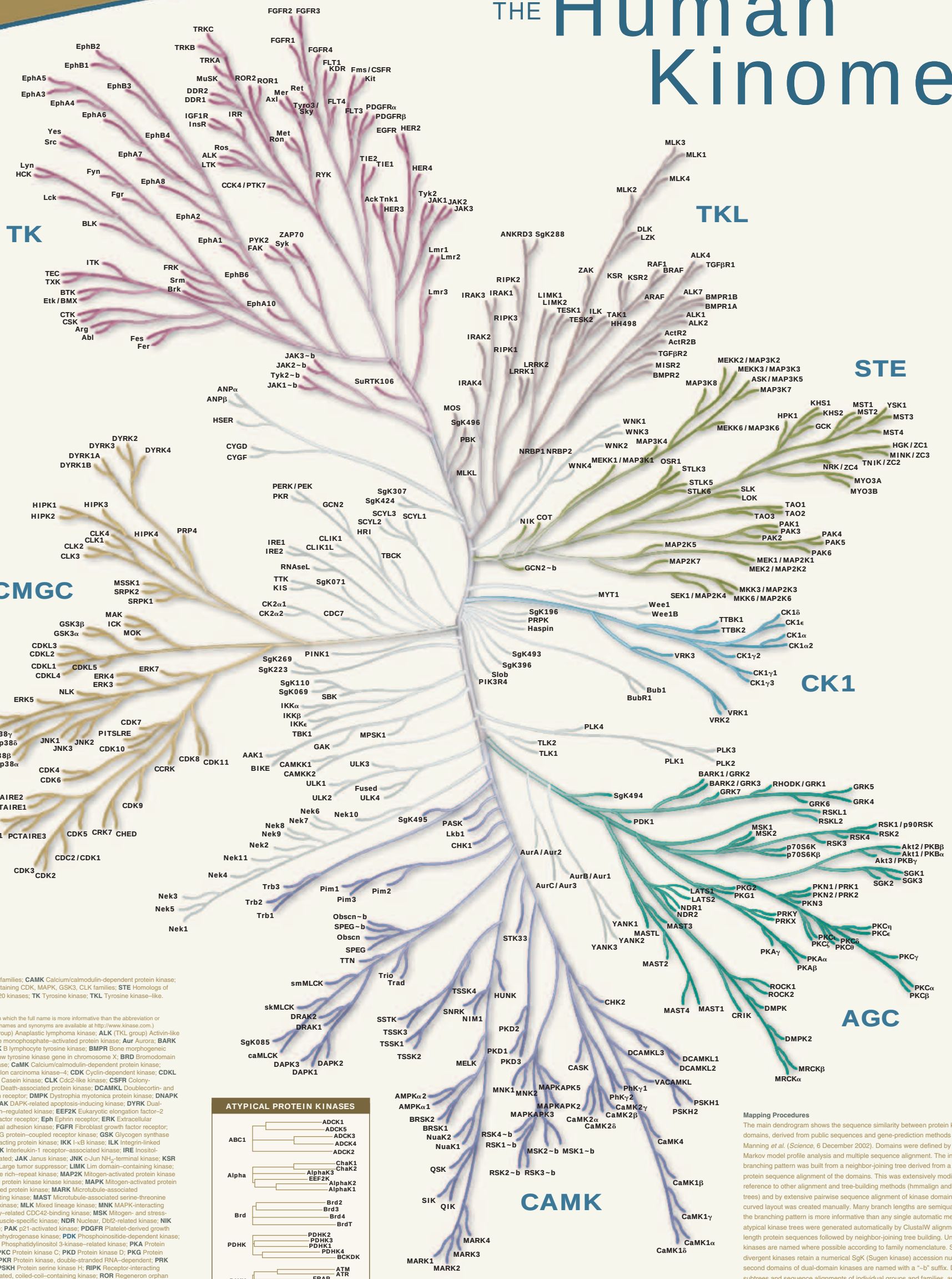
Science's STKE
SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT
WWW.STKE.ORG

This phylogenetic tree depicts the relationships between members of the complete superfamily of human protein kinases. Protein kinases constitute one of the largest human gene families and are key regulators of cell function. The 518 human protein kinases control protein activity by catalyzing the addition of a negatively charged phosphate group to other proteins. Protein kinases modulate a wide variety of biological processes, especially those that carry signals from the cell membrane to intracellular targets and coordinate complex biological functions.

Most protein kinases belong to a single superfamily of enzymes whose catalytic domains are related in sequence and structure. The main diagram illustrates the similarity between the protein sequences of these catalytic domains. Each kinase is at the tip of a branch, and the similarity between various kinases is inversely related to the distance between their positions on the tree diagram. Most kinases fall into small families of highly related sequences, and most families are part of larger groups. The seven major groups are labeled and colored distinctly. Other kinases are shown in the center of the tree, colored gray. The relationships shown on the tree can be used to predict protein substrates and biological function for many of the over 100 uncharacterized kinases presented here.

The inset diagram shows trees for seven atypical protein kinase families. These proteins have verified or strongly predicted kinase activity, but have little or no sequence similarity to members of the protein kinase superfamily. A further eight atypical protein kinases in small families of one or two genes are not shown.

THE Human Kinome



Group names
AGC Containing PKA, PKG, PKC families; **CAMK** Calcium/calmodulin-dependent protein kinase; **CK1** Casein kinase 1; **CMGC** Containing CDK, MAPK, GSK3, CLK families; **STE** Homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases; **TK** Tyrosine kinase; **TKL** Tyrosine kinase–like.

Kinase names
(A selective list includes those cases in which the full name is more informative than the abbreviation or acronym shown on the tree. Other full names and synonyms are available at http://www.kinase.com.)
ActR Activin receptor; **ALK** (TK group) Anaplastic lymphoma kinase; **ALK** (TKL group) Activin-like receptor kinase; **AMPK** Adenosine monophosphate–activated protein kinase; **Aur** Aurora; **BARK** β-adrenergic receptor kinase; **BLK** B lymphocyte tyrosine kinase; **BMPR** Bone morphogeneic protein receptor; **BMX** Bone marrow tyrosine kinase gene in chromosome X; **BRD** Bromodomain kinase; **BRSK** Brain-selective kinase; **CaMK** Calcium/calmodulin-dependent protein kinase; **CAMKK** CaMK kinase; **CCK-4** Colon carcinoma kinase–4; **CDK** Cyclin-dependent kinase; **CDKL** Cyclin-dependent kinase–like; **CK** Casein kinase; **CLK** Cdc2-like kinase; **CSFR** Colony-stimulating factor receptor; **DAPK** Death-associated protein kinase; **DCAMKL** Doublecortin- and CaMK-like; **DDR** Discoidin domain receptor; **DMPK** Dystrophia myotonica protein kinase; **DNAPK** DNA-activated protein kinase; **DRAK** DAPK-related apoptosis-inducing kinase; **DYRK** Dual-specificity tyrosine phosphorylation–regulated kinase; **EEF2K** Eukaryotic elongation factor–2 kinase; **EGFR** Epidermal growth factor receptor; **Eph** Ephrin receptor; **ERK** Extracellular signal–regulated kinase; **FAK** Focal adhesion kinase; **FGFR** Fibroblast growth factor receptor; **FRK** Fos-regulatory kinase; **GRK** G protein–coupled receptor kinase; **GSK** Glycogen synthase kinase; **HIPK** Homeodomain-interacting protein kinase; **IKK** I-κB kinase; **ILK** Integrin-linked kinase; **InsR** Insulin receptor; **IRAK** Interleukin-1 receptor–associated kinase; **IRE** Inositol-requiring; **IRR** Insulin receptor–related; **JAK** Janus kinase; **JNK** c-Jun NH_2-terminal kinase; **KSR** Kinase suppressor of Ras; **LATS** Large tumor suppressor; **LIMK** Lim domain–containing kinase; **LMR** Lemur kinase; **LRRK** Leucine rich–repeat kinase; **MAP2K** Mitogen-activated protein kinase kinase; **MAP3K** Mitogen-activated protein kinase kinase kinase; **MAPK** Mitogen-activated protein kinase; **MAPKAPK** MAPK–activated protein kinase; **MARK** Microtubule-associated protein/microtubule affinity–regulating kinase; **MAST** Microtubule-associated serine-threonine kinase; **MLCK** Myosin light chain kinase; **MLK** Mixed lineage kinase; **MNK** MAPK-interacting kinase; **MRCK** Myotonic dystrophy–related CDC42-binding kinase; **MSK** Mitogen- and stress-activated protein kinase; **MuSK** Muscle-specific kinase; **NDR** Nuclear, Dbf2-related kinase; **NIK** Nuclear factor κB–inducing kinase; **PAK** p21-activated kinase; **PDGFR** Platelet-derived growth factor receptor; **PDHK** Pyruvate dehydrogenase kinase; **PDK** Phosphoinositide-dependent kinase; **PhK** Phosphorylase kinase; **PIKK** Phosphatidylinositol 3-kinase–related kinase; **PKA** Protein kinase A; **PKB** Protein kinase B; **PKC** Protein kinase C; **PKD** Protein kinase D; **PKG** Protein kinase G; **PKN** Protein kinase N; **PKR** Protein kinase, double-stranded RNA–dependent; **PRK** Protein kinase C–related kinase; **PSKH** Protein serine kinase H; **RIPK** Receptor-interacting protein kinase; **ROCK** Rho-associated, coiled-coil–containing kinase; **ROR** Regeneron orphan receptor; **RSK** Ribosomal protein S6 kinase; **RSKL** RSK-like; **SgK** Sugen kinase; **SGK** Serum- and glucocorticoid-regulated kinase; **SRPK** Serine-arginine splicing factor protein kinase; **SYK** Spleen tyrosine kinase; **TAK** Transforming growth factor–β–activated kinase; **TEC** Tyrosine kinase expressed in hepatocellular carcinoma; **TESK** Testis-specific kinase; **TGFβR** Transforming growth factor–β receptor; **TIE** Tyrosine kinase with immunoglobulin and EGF repeats; **TIF1** Transcriptional intermediary factor 1; **TLK** Tousled-like kinase; **TSSK** Testis-specific serine

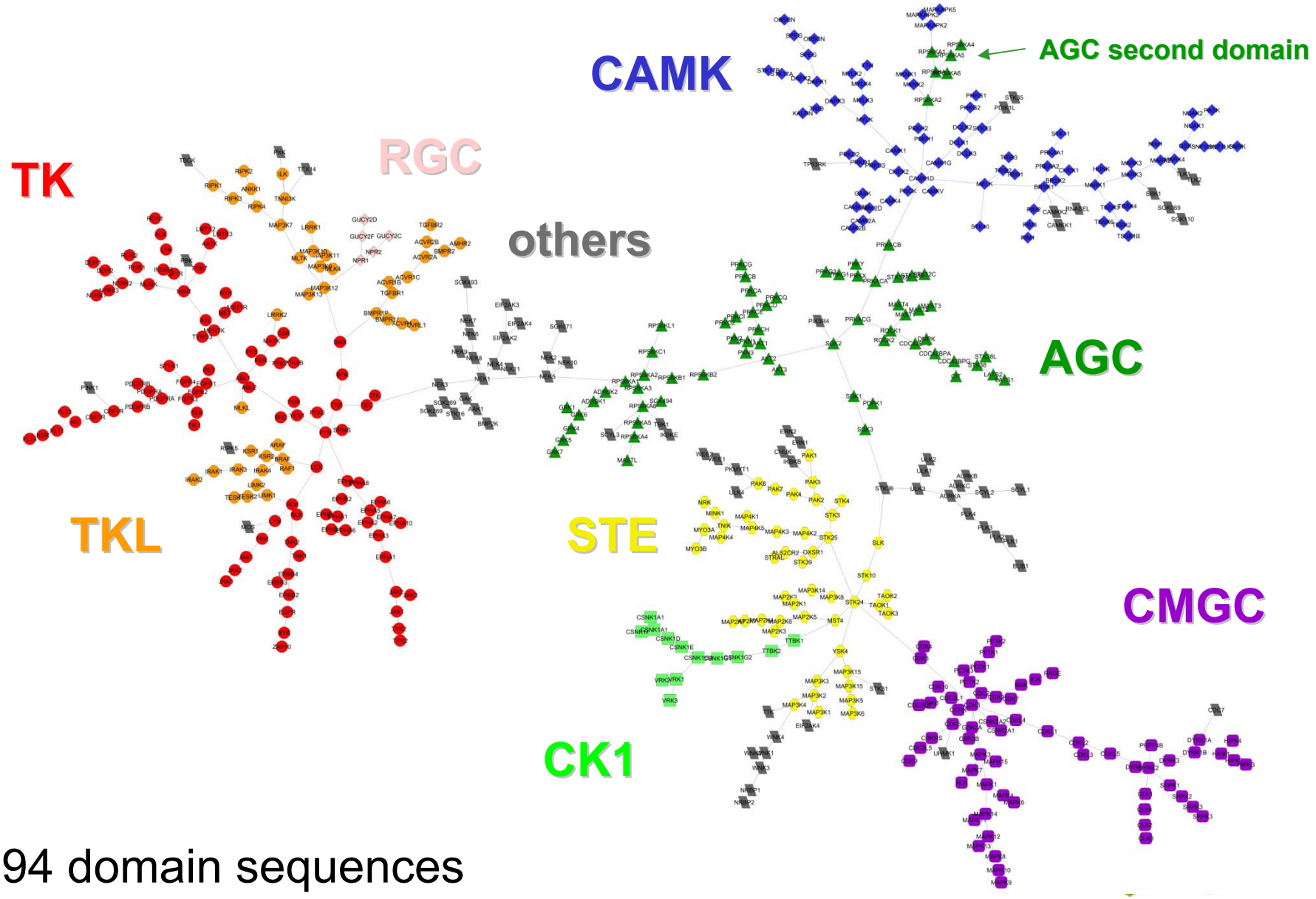
ATYPICAL PROTEIN KINASES

Mapping Procedures
The main dendrogram shows the sequence similarity between protein kinase domains, derived from public sequences and gene-prediction methods detailed in Manning *et al.* (*Science*, 6 December 2002). Domains were defined by hidden Markov model profile analysis and multiple sequence alignment. The initial branching pattern was built from a neighbor-joining tree derived from a ClustalW protein sequence alignment of the domains. This was extensively modified by reference to other alignment and tree-building methods (hmmalign and parsimony trees) and by extensive pairwise sequence alignment of kinase domains. The curved layout was created manually. Many branch lengths are semiquantitative, but the branching pattern is more informative than any single automatic method. The atypical kinase trees were generated automatically by ClustalW alignment of full-length protein sequences followed by neighbor-joining tree building. Unpublished kinases are named where possible according to family nomenclature. Some divergent kinases retain a numerical SgK (Sugen kinase) accession number. The second domains of dual-domain kinases are named with a "~b" suffix. Detailed subtrees and sequence alignments of individual groups and families, and comparative genomic trees are available at http://www.kinase.com. Information on regulation and substrates of many of these kinases is available at http://www.cellsignal.com.

Authors Gerard Manning, David B. Whyte, Ricardo Martinez, Sucha Sudarsanam, Sugen Inc., South San Francisco, CA, USA; Tony Hunter, Salk Institute, La Jolla, CA, USA. This poster accompanies the paper "The Protein Kinase Complement of the Human Genome" (Manning *et al.*, *Science*, 6 December 2002).

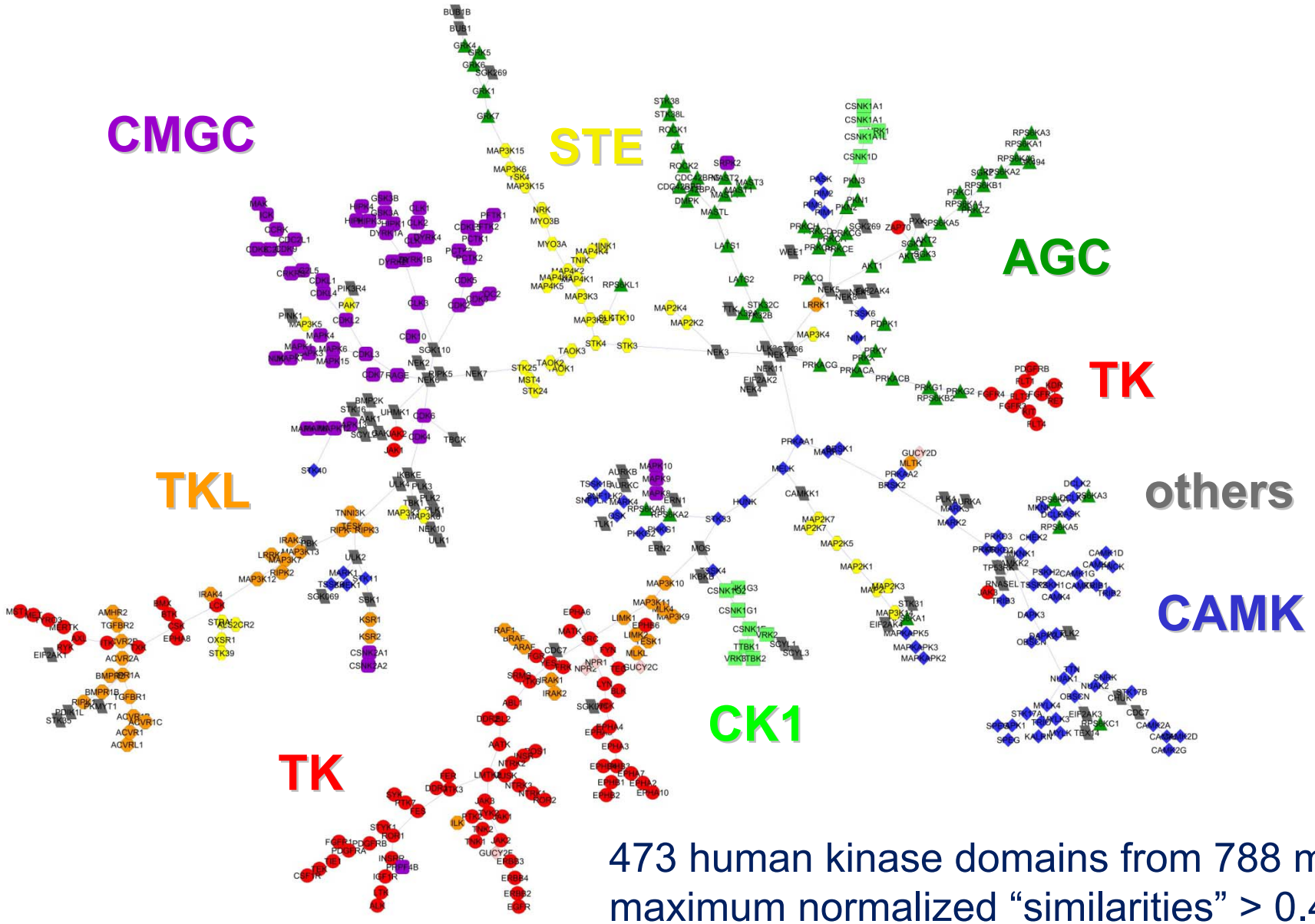
Science coordinator L. Bryan Ray
Design and production Anne Ashley, David Comb, Michael Melnick, C. Faber Smith, J. White, David M. Tompkins, Marcus Spiegler
Copyeditor Harry Jach
Sponsored by Cell Signaling Technology, Inc. and Sugen, Inc.

Kinase Domain Sequence Similarities - MST



494 domain sequences

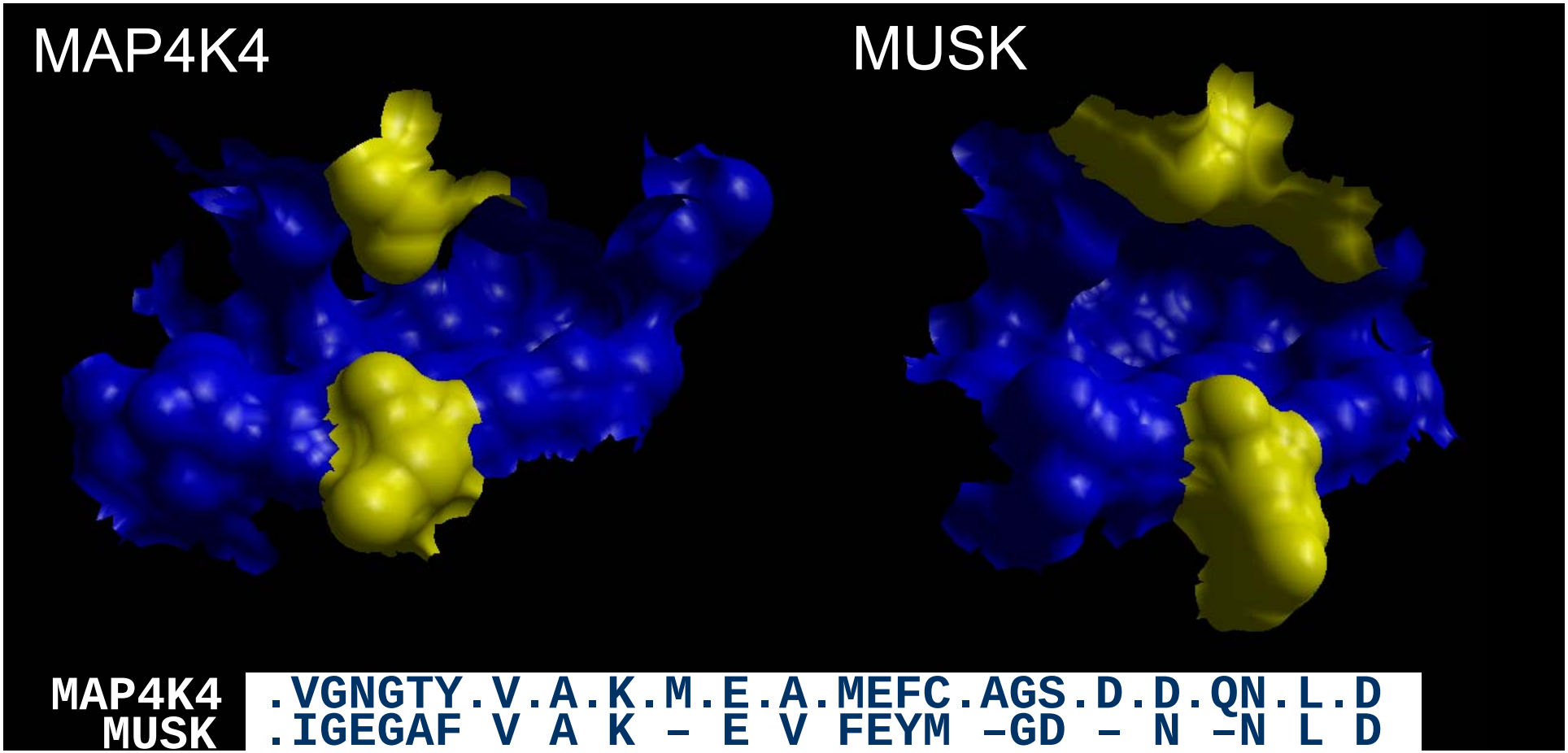
Local site similarity – MST



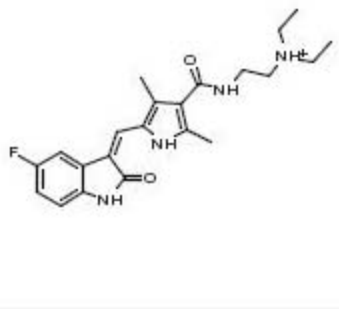
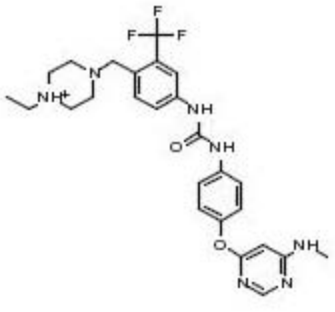
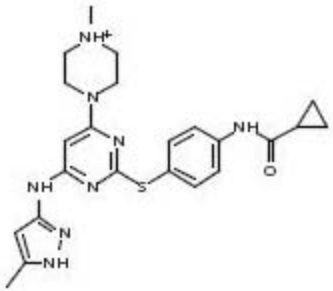
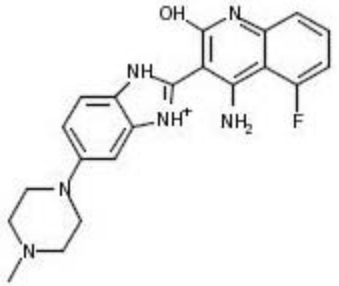
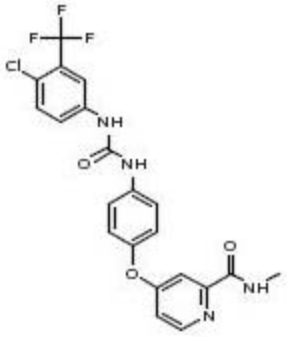
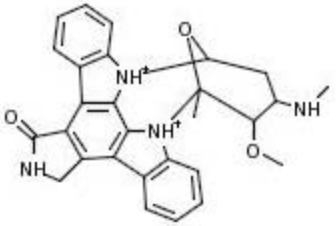
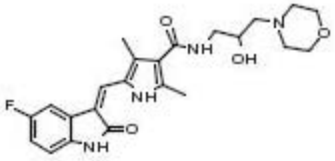
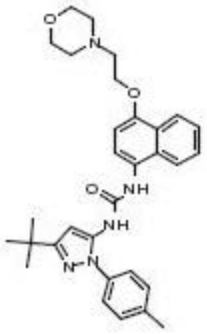
473 human kinase domains from 788 models
maximum normalized “similarities” > 0.45

Example: PhysChem SiteSim vs. 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



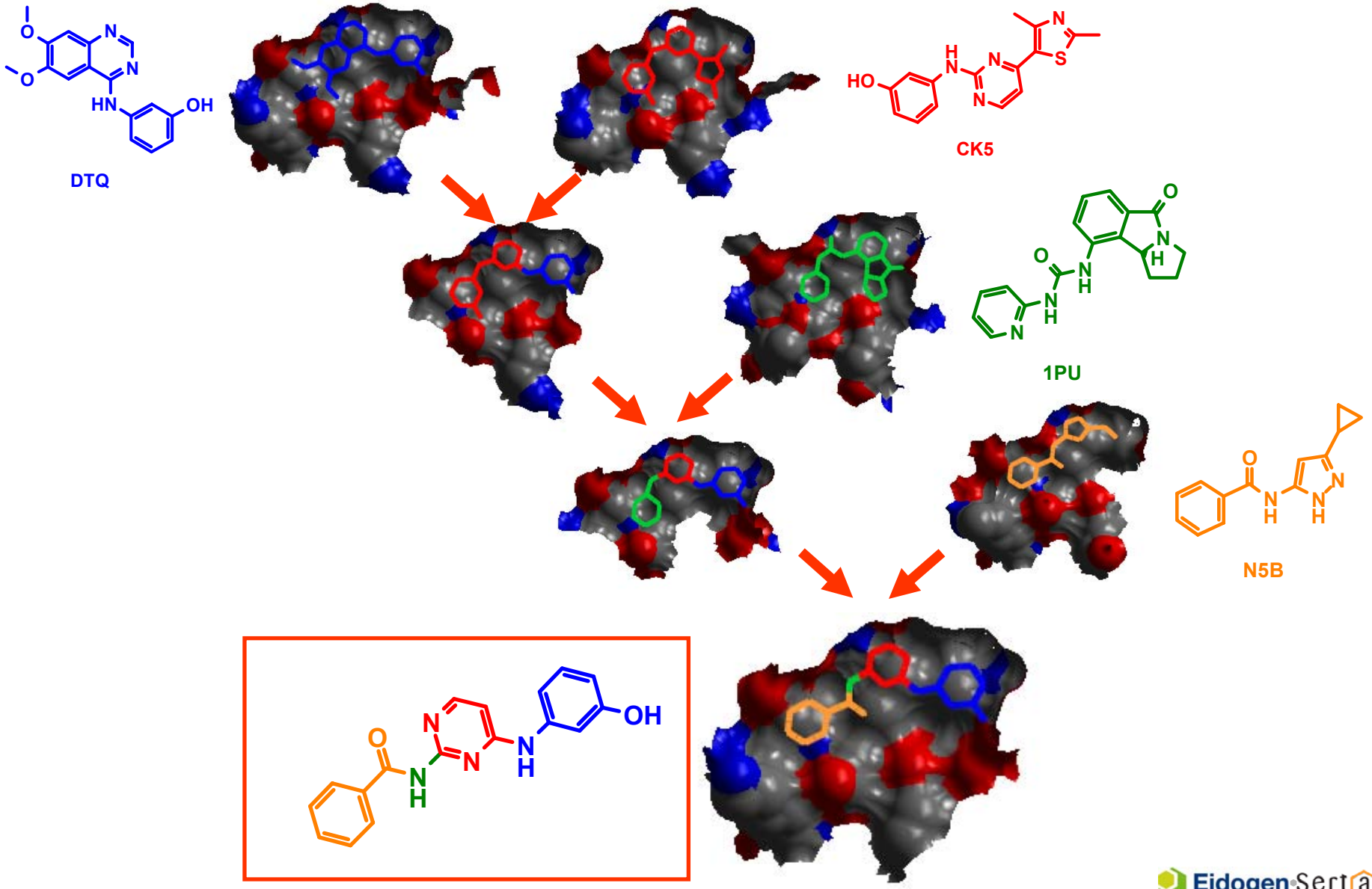
MAP4K4 and MUSK Small Molecule Inhibitors

 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

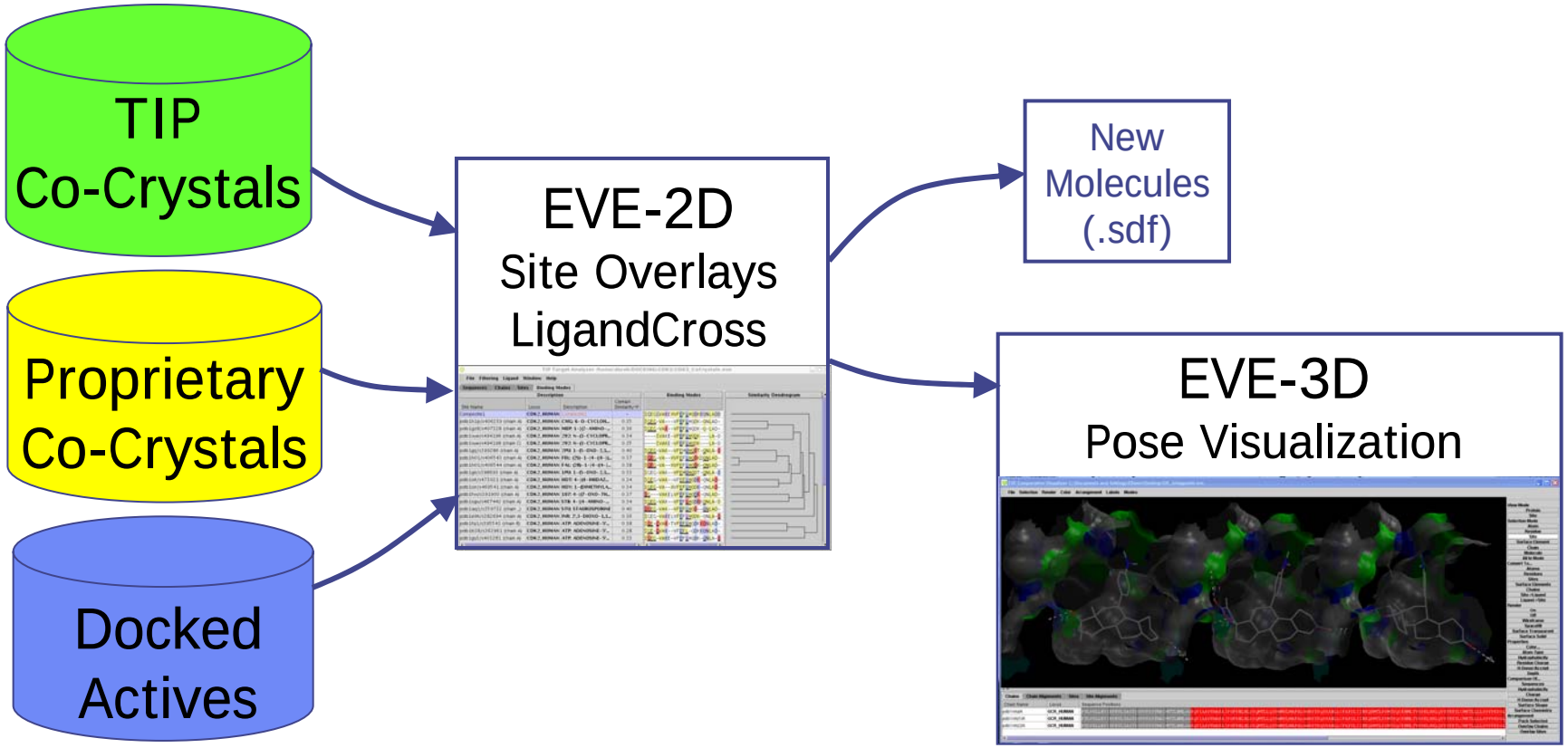
Eidogen-Sertanty

LigandCross: Shuffling Ligand Functionality

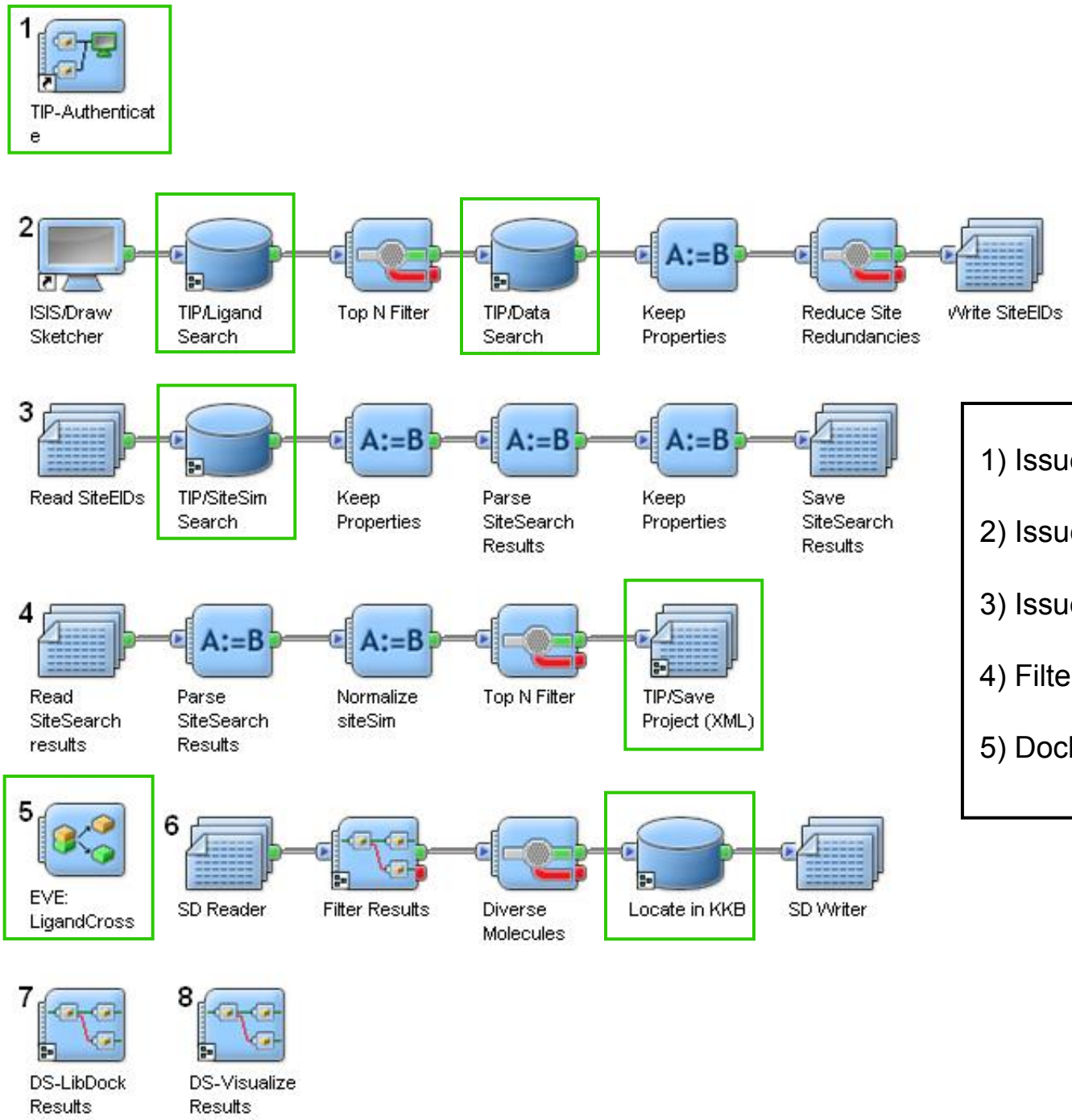
Similar to Vertex's BREED: J. Med. Chem. **47**, 2768 (2004)



LigandCross Workflow



New Molecules via LigandCross



1) Issue TIP/LigandSearch

2) Issue TIP/SiteSimSearch

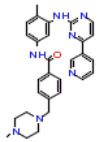
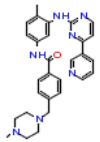
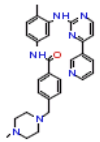
3) Issue LigandCross

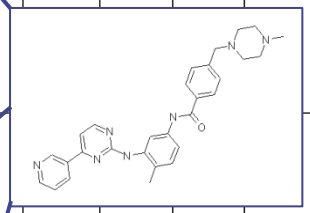
4) Filter and locate results in KKB

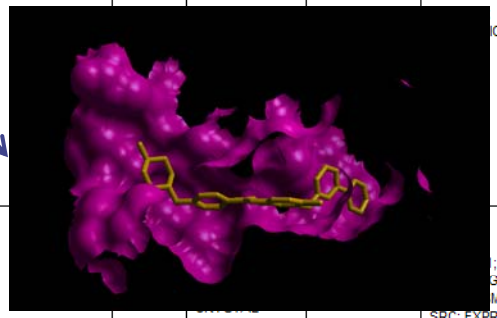
5) Dock and visualize results

Eidogen-Sertanty

Step 1: Find Co-complexes and Sites from Ligand-Structure-Search

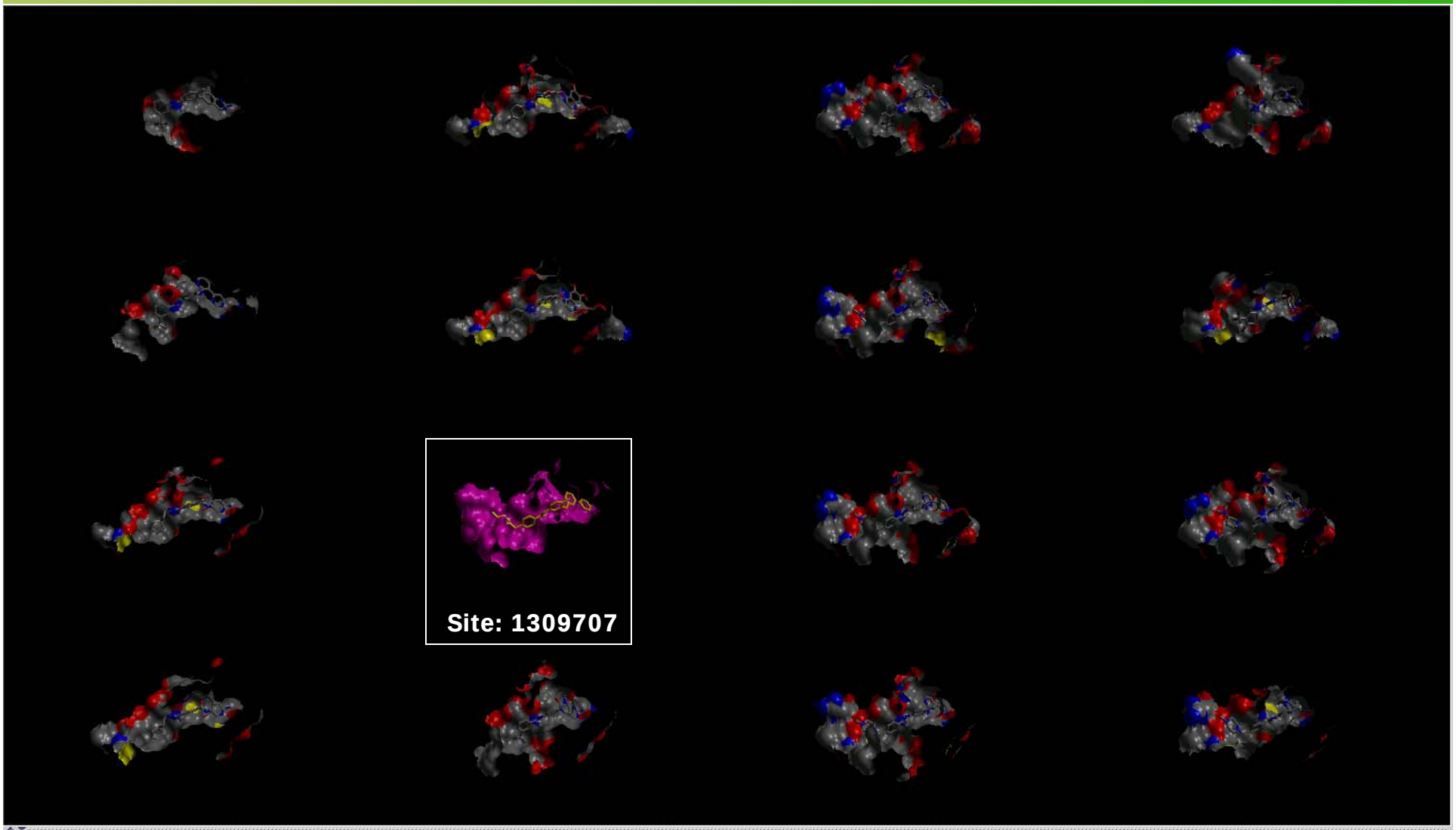
Molecule	ligname	similarity	pdbcode	siteeid	FourCode	pdbID	pdbBnxNumber	proteinId	title	classification	source	compound	releaseDate	journalTitle	journalReference	exptype
	STI	1	2pl0A	1309707	2pl0	2pl0	1305799	42526	LCK BOUND TO IMATINIB	TRANSFERASE	MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: HUMAN; GENE: LCK; EXPRESSION_SYSTEM: SPODOPTERA FRUGIPERDA; EXPRESSION_SYSTEM_COMMON: FALL ARMYWORM; EXPRESSION_SYSTEM_VECTOR_TYPE: ION_SYSTEM_PLASMID:	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	2oiqA	1146914	2oiq	2oiq	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 (STI571, GLIVEC)	TRANSFERASE	MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: HUMAN; GENE: ABL1; EXPRESSION_SYSTEM: SPODOPTERA FRUGIPERDA; 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 80 2007	XRAY DIFFRACTION





Eidogen-Sertanty

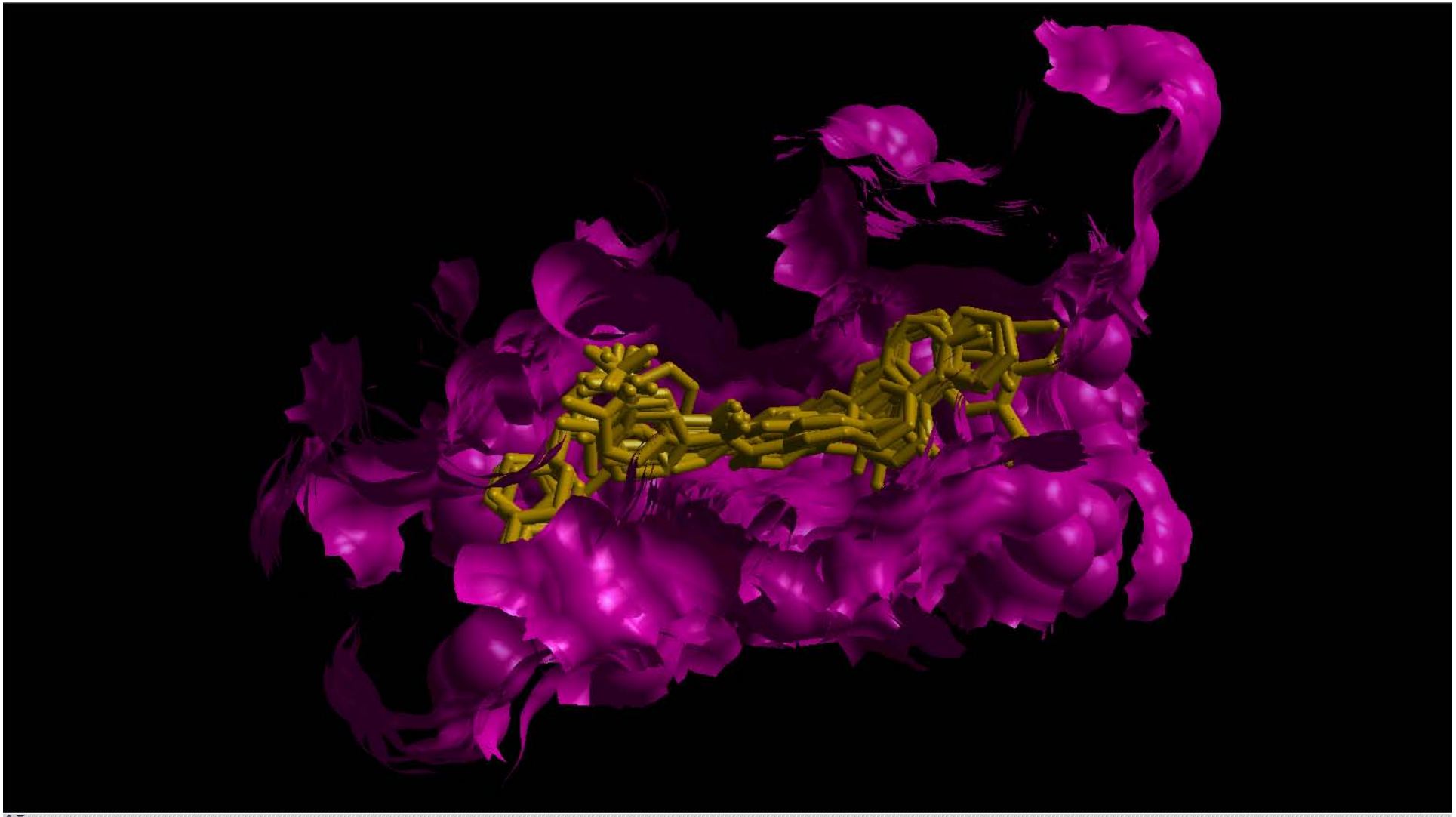
Step 2: Find Other Receptor Sites from Site-Similarity Search



Chains | Chain Alignments | Sites | Site Alignments

Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2pl0/s1309707 (chain A)	LCK	STI	100	.L.V.AVK.E.LM.L.LV.I.TEYM.GS.I.YIHR.L.IADF
pdb2ofw/s916548 (chain B)	LCK	242	100	.L.V.AVK.E.LM.L.LV.I.TEYM.G.I.Y.H.L.IADF.I
pdb2rl5/s1396160 (chain A)	-	2RL	100	.LG.V.AVK.L.E.LL.L.VV.V.TEFCKFGN.L.CIH.L.ICDF
pdb2e2b1/s1284639 (chain B)	ABL	406	100	.L.Y.V.A.K.E.VM.L.LV.I.TEFMT.G.L.FIHRD.L.VADF

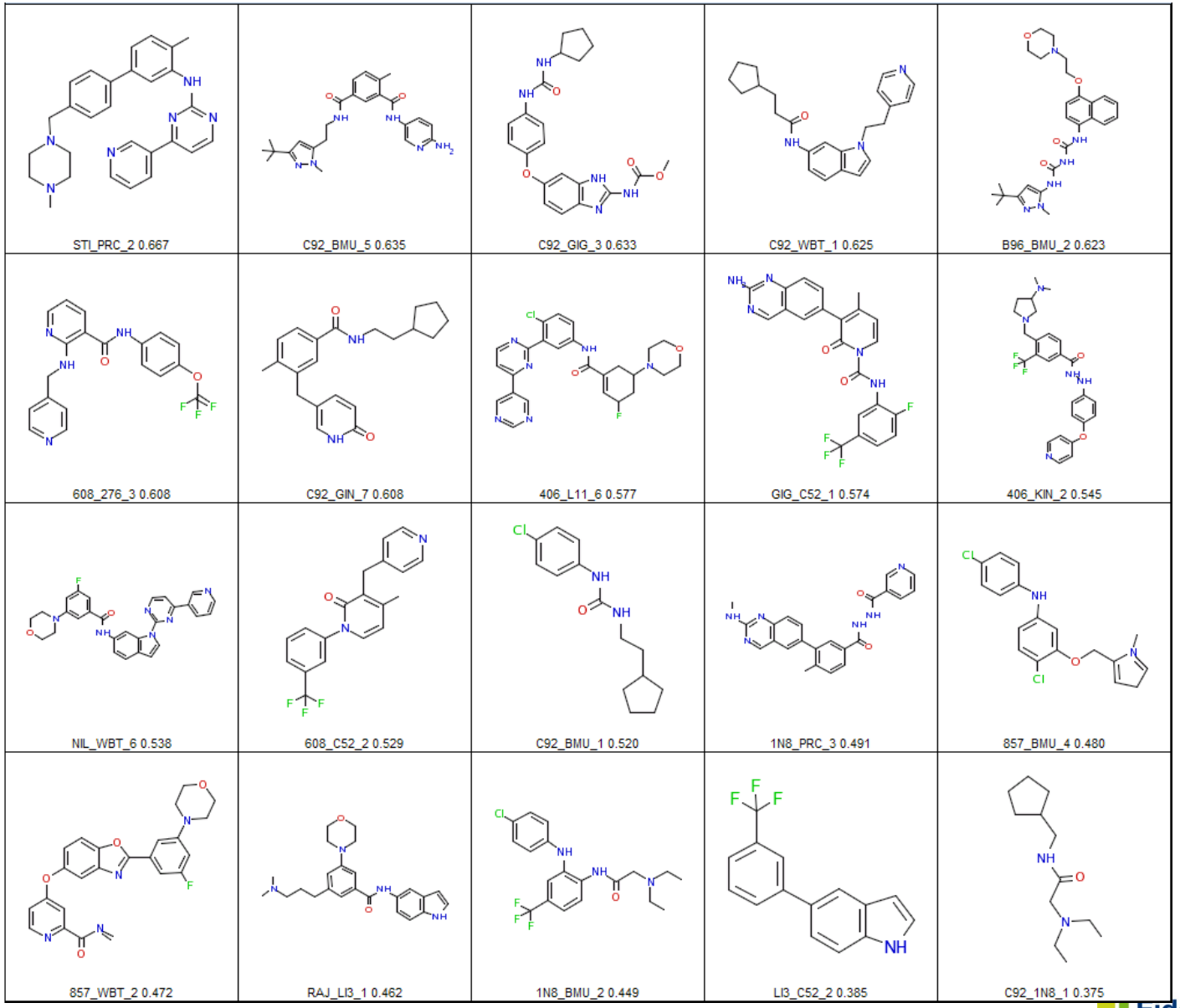
Step 3: LigandCross – Mixing Ligand Features from Aligned Sites



Chains | Chain Alignments | Sites | Site Alignments

Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2pl0/s1309707 (chain A)	LCK	STI	100	.L.V.AVK.E.LM.L.LV.I.TEYM.GS.I.YIHR.L.IADF
pdb2of/s916648 (chain B)	LCK	242	100	.L.V.AVK.E.LM.L.LV.I.TEYM.G.I.Y.H.L.IADF.I
pdb2rl5/s1396160 (chain A)	-	2RL	100	.LG.V.AVK.L.E.IL.I.VV.V.TEFCKFGN.L.CIH.L.ICDF
pdb2e2b1/s1284639 (chain B)	ABL	406	100	.L.Y.V.A.K.E.VM.I.LV.I.TEFMT.G.L.FIHRD.L.VADF

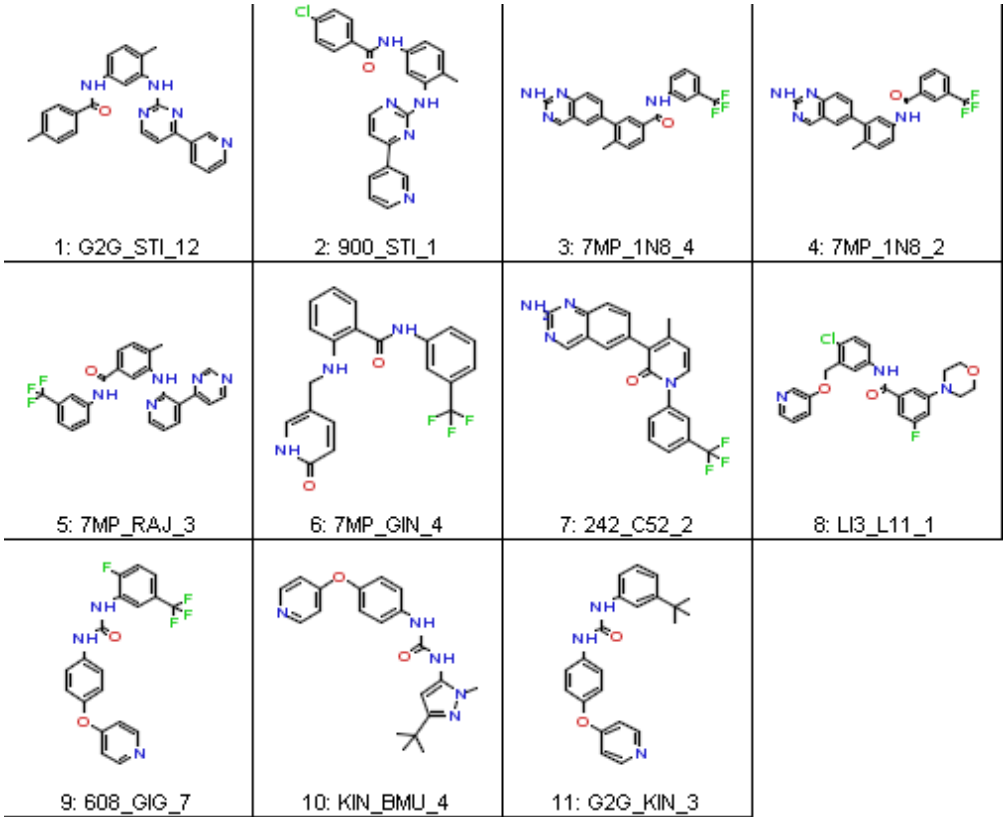
Example LigandCross Results



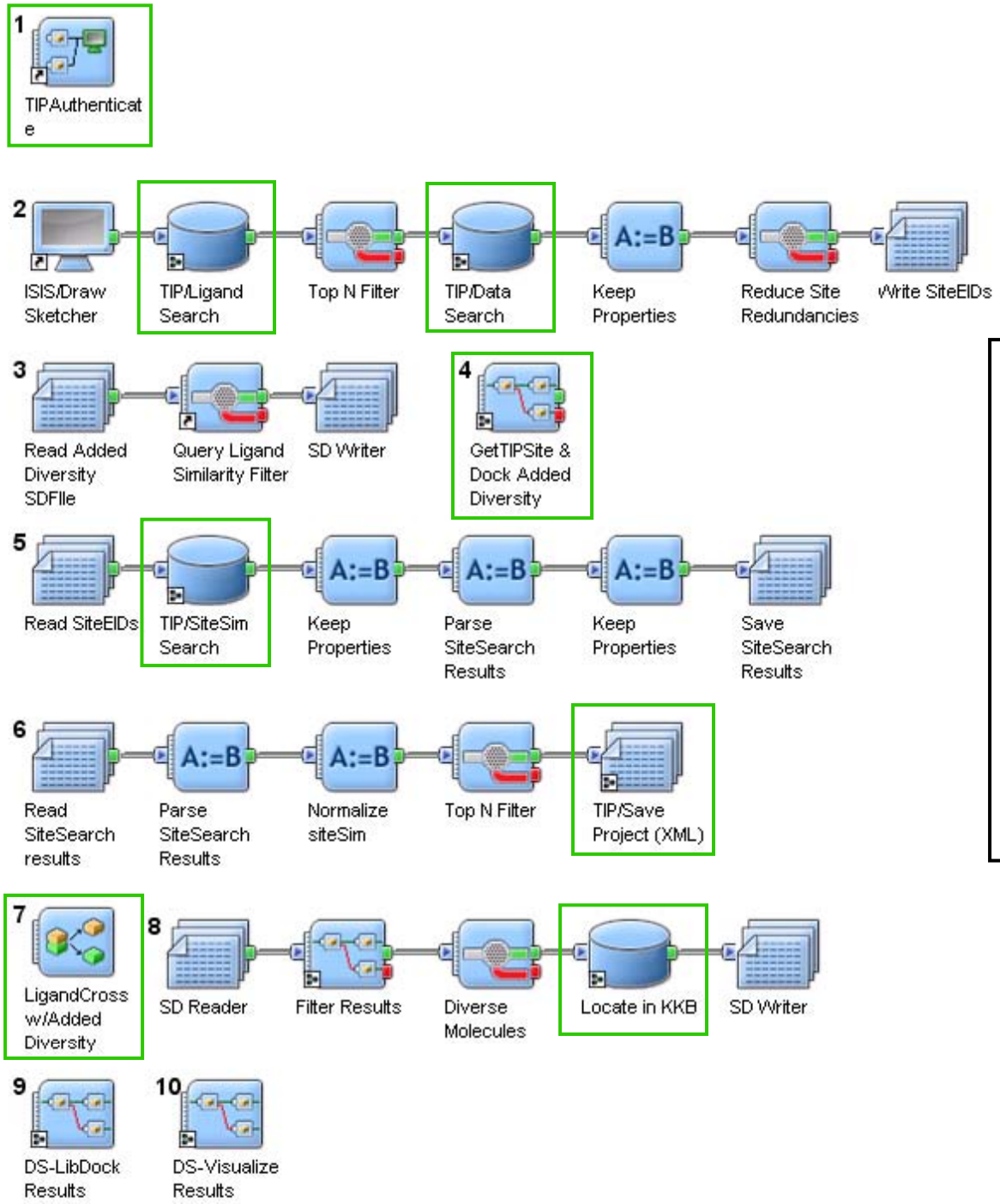
Eidogen-Sertanty

Step 4: LigandCross Ligands with Reported Biological Activity

	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



Eidogen-Sertanty



> Issue TIP/LigandSearch

> Identify/Dock “AddedDiversity”

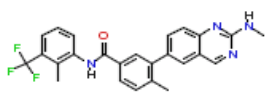
> Issue TIP/SiteSimSearch

> LigandCross w/AddedDiversity

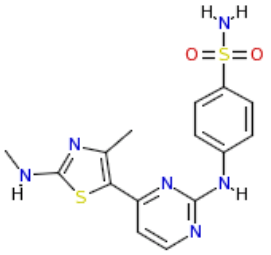
> Filter and locate results in KKB

> Dock and visualize results

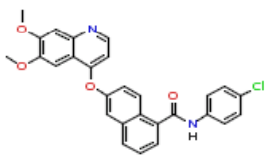
Example Potent Kinase Inhibitors ("Added Diversity")



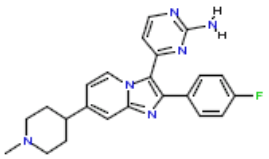
4336533 LCK pval: 11.00



4302493 CDK9 pval: 10.54

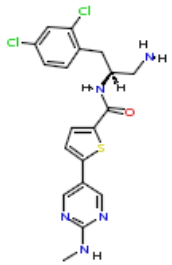


4332561 KDR pval: 10.52

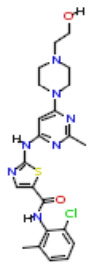


4318145 PKG pval: 10.40

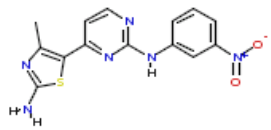
Chiral



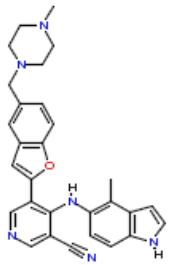
4336686 PKA pval: 10.00



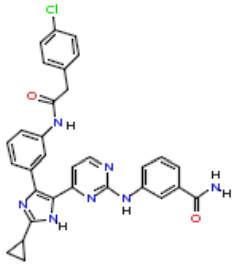
4272835 ABL1 pval: 10.00



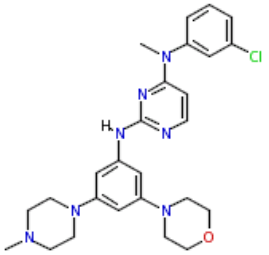
894611 CDK2 pval: 9.70



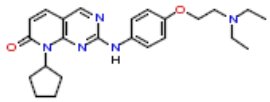
4358565 PRKCQ pval: 9.70



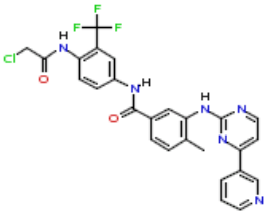
4363734 RAF1 pval: 9.30



4369892 EPHB4 pval: 9.24

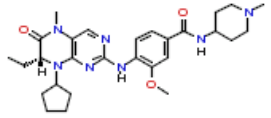


809 CDK4 pval: 9.15

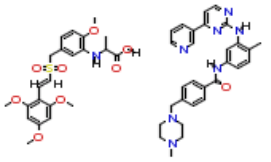


4374385 PDGFRA pval: 9.14

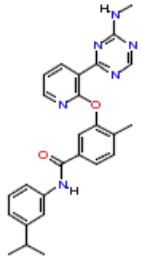
Chiral



4366691 PLK1 pval: 9.10

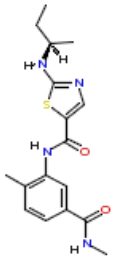


4301886 BCR_ABL pval: 9.08

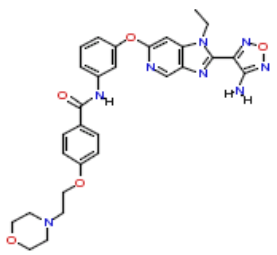


4307551 TEK pval: 9.00

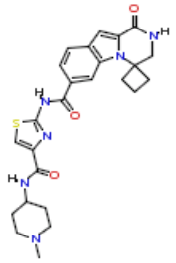
Chiral



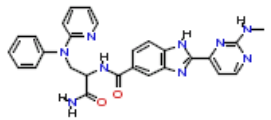
4363016 MAPK11 pval: 8.82



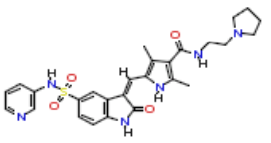
4343448 ROCK1 pval: 8.74



4363247 MAPKAPK2 pval: 8.70

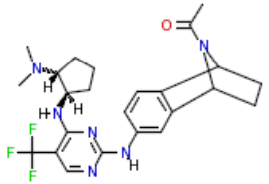


4291996 IKB pval: 8.70



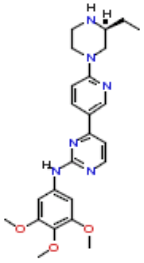
4208857 FAK2 pval: 8.22

Chiral

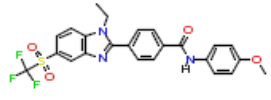


4373725 PTK2B pval: 8.22

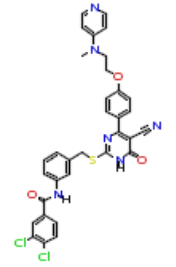
Chiral



1788 ZAP70 pval: 8.10

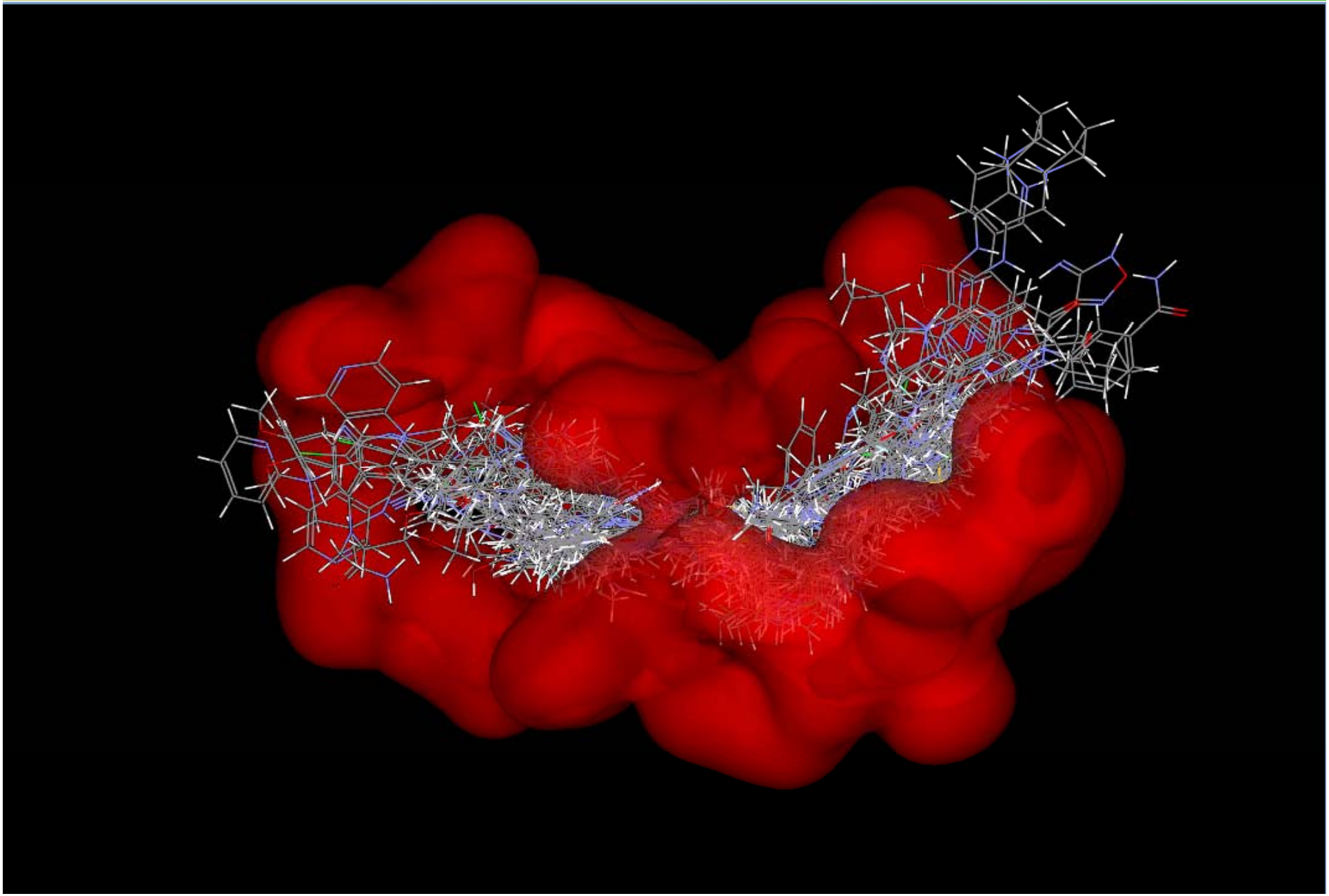


2425813 PTPN9 pval: 5.96

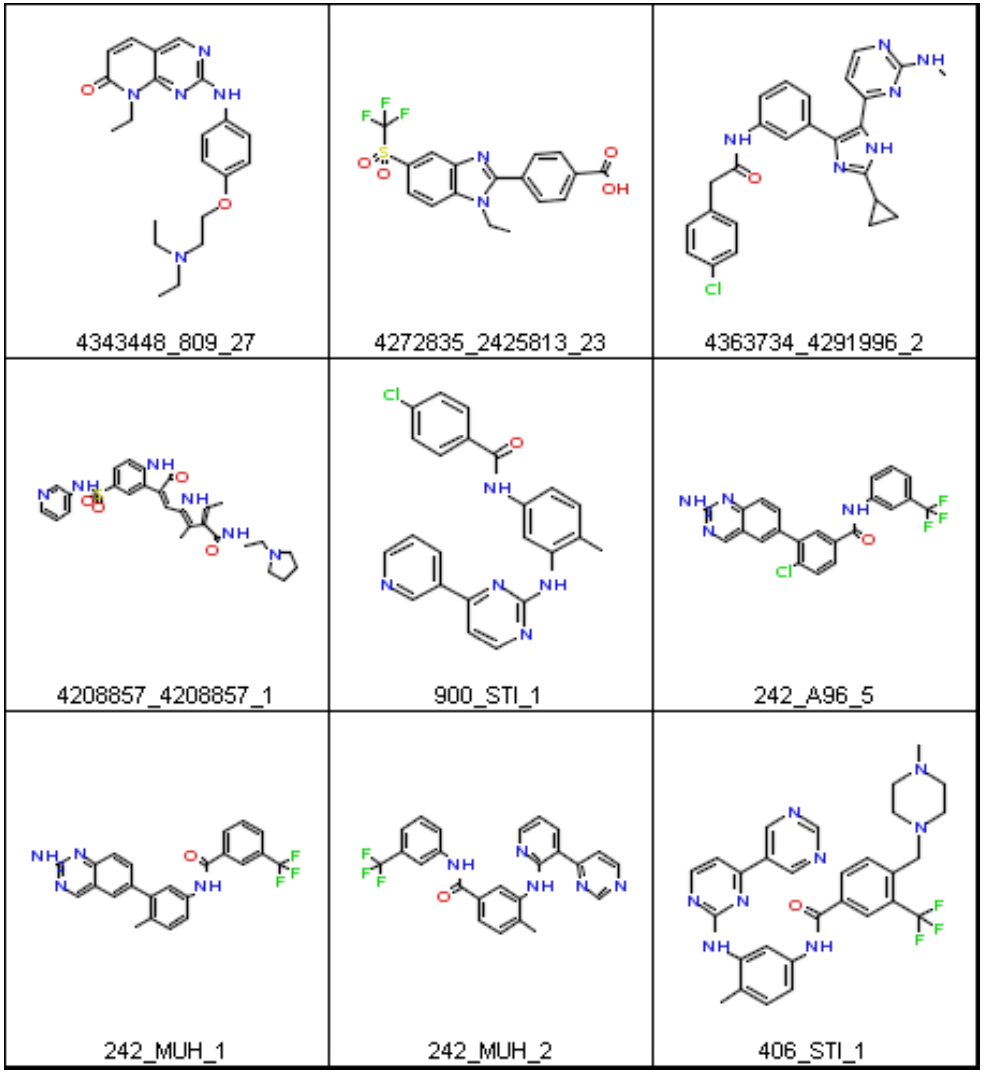


4303129 MAP3K2 pval: 4.70

Potent Kinase Inhibitors Docked (s1309707)



LigandCross Examples using “Added Diversity”



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:
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

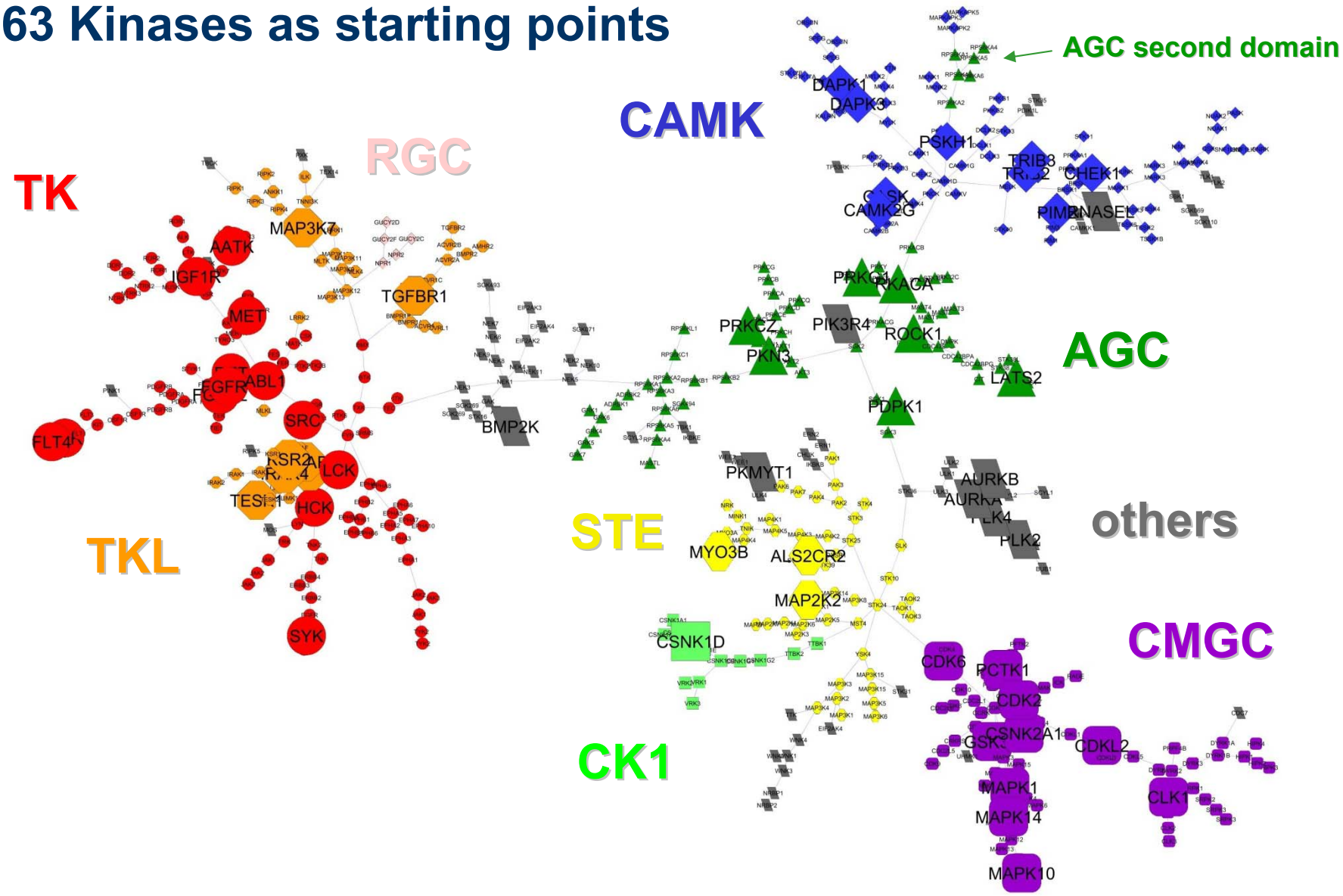
242_MUH_1:
LCK: 9.40 TEK: 7.68 KDR: 8.22 MAPK14: 9.00 JAK3: 6.81

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

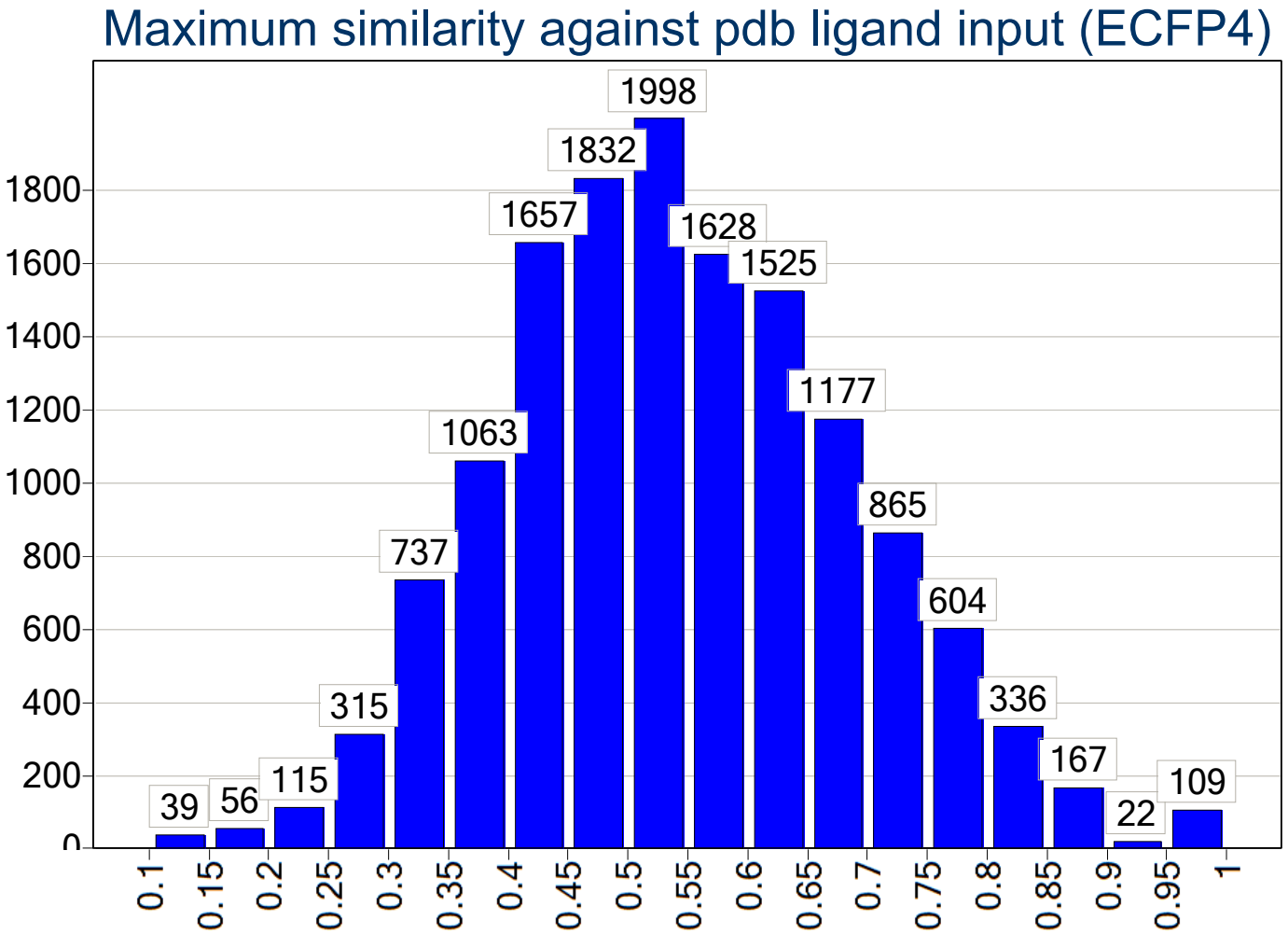
Ligand Functionality Shuffling across the Kinome

63 Kinases as starting points



Generates novel compounds

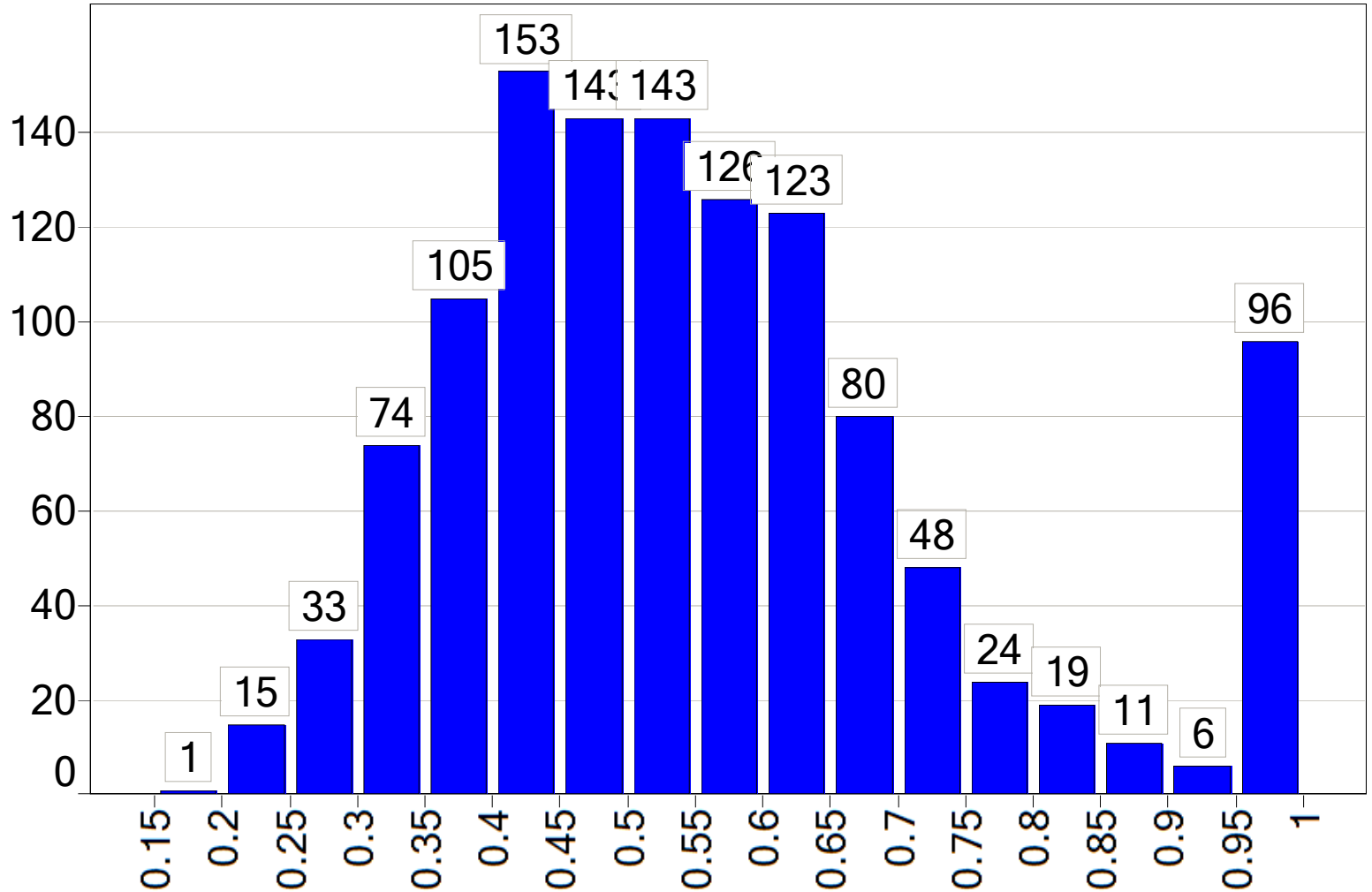
- From 280 Unique Kinase co-Crystal ligands
- > 14,000 new unique structures are generated (HTS filter)



Drug- / lead-like novel compounds

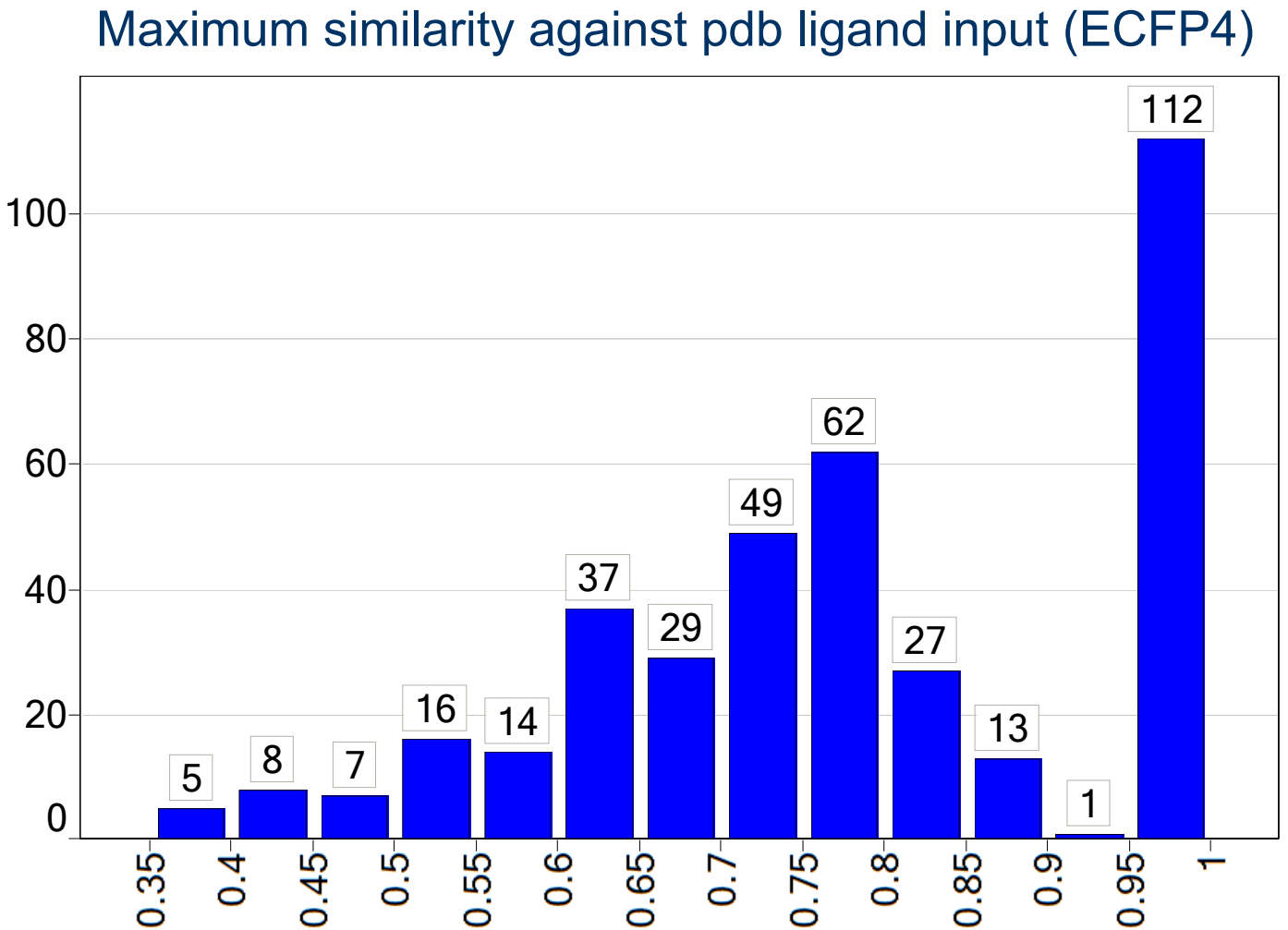
- Strict filtering (drug-/ lead-like; functional groups, properties)

➢2,153 unique compounds (64 pdb ligands pass the same filters)



Novel active compounds

- 380 reported kinase inhibitors are generated
- 268 are novel (not seen as input into the protocol)



Conclusions

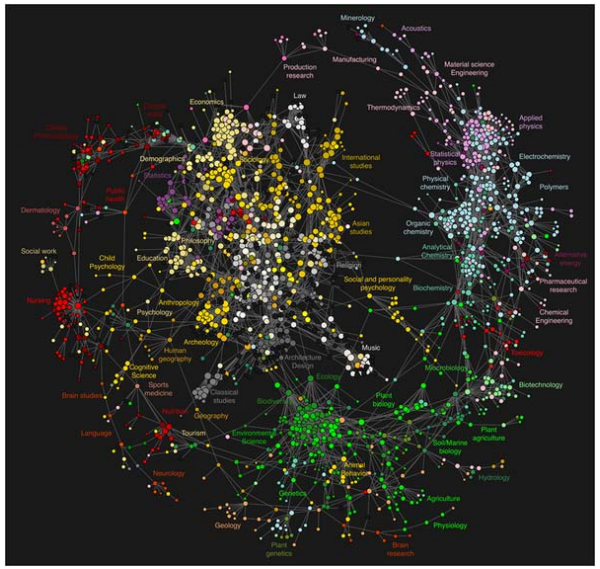
- Significant receptor-site similarities exist within and across target families

- 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.

Acknowledgements

- University of Miami
  - Stephan Schurer / Dusica Vidovic
- CNIO-Spain
  - Julen Oyarzabal/Obdulia Rabal
- Accelrys
  - Mori Masayuki
  - Shikha Varma / Ton van Daelen
  - Kush Kapila / Guillaume Paillard
  - June Snedecor



BollenJ (March 2009) -ClickstreamData Yields High-Resolution Maps of Science

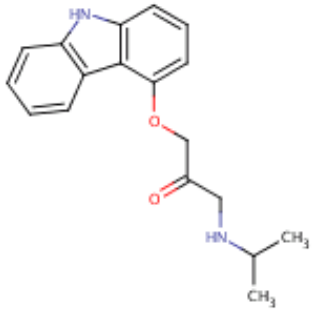
Eidogen-Sertanty

LigandCross CAU/siteSimilar Ligands w/Docked BioActives

- Extract CAU/2RH1 Site (siteEID: 1276810) from TIP
- Issue siteSimilarity Search against siteEID: 1276810
- LibDock 20-30 Bioactive Molecules (beta1-adrenergic active+ CAU similar)
- LigandCross CAU and siteSimilar Ligands w/Docked BioActives

TIP/SiteSimilarity Search – siteEID: 1276810 (CAU/2RH1)

Site Name	s1276810
Source	PDB Co-crystal
Confidence	100%
Resolution	XRAY
Description	CAU: (2S)-1-(9H-Carbazol-4-yloxy)-3-(isopropylamino)propan-2-ol
Parent	2rh1A



Query: 2rh1A/s1276810

Total Sites Found: 399

Same-Fold Sites:	358	(89.7%)	
Dissimilar-Fold Sites:	41	(10.3%)	
Ligands:	Predicted	354	(88.7%)
	MGD	15	(3.8%)
	NAD	11	(2.8%)
	ATP	4	(1.0%)
	Other	15	(3.8%)

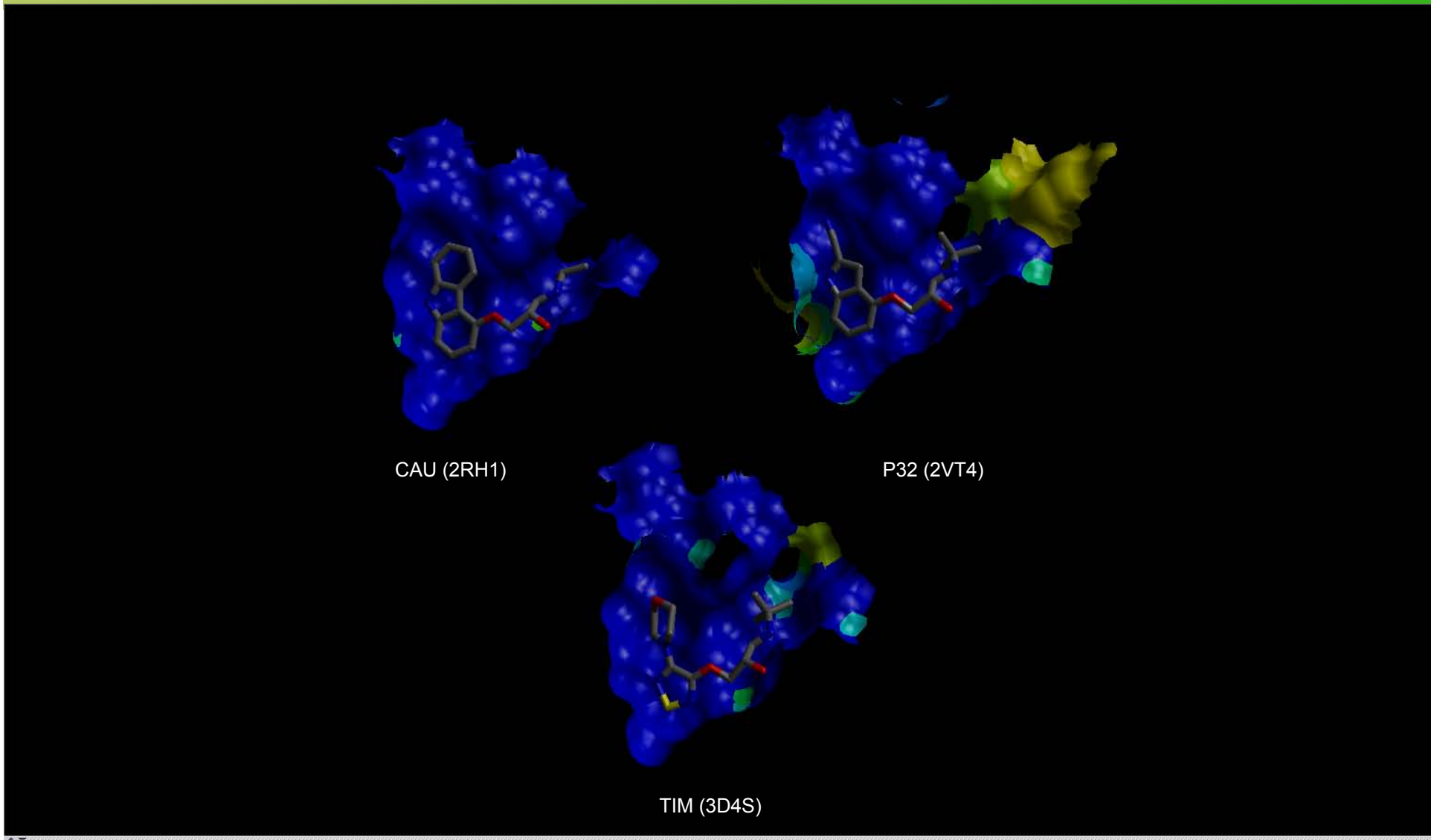
Index	Superfamily Clusters	Family Clusters	Site Clusters	Sites	Description	Ligand Information	SiteSorter Range	%ID Range	Contact Range
1.	»	3	298	358	Beta-1 adrenergic receptor	Predicted: 350, P32: 4, RET: 3, TIM: 1	60-107	14-100	0.00-0.77
2.	»	1	2	15	PERIPLASMIC NITRATE REDUCTASE	MGD: 15	60-66	14-14	0.05-0.09
3.	»	1	2	4	PYRUVATE KINASE	ATP: 4	61-66	14-14	0.10-0.12
4.	»	1	2	2	4S-LIMONENE SYNTHASE	F3P: 1, FPG: 1	61-65	14-14	0.12-0.14
5.	»	1	11	11	PROTEIN (S-ADENOSYLHOMOCYSTEINE	NAD: 11	61-64	14-14	0.05-0.06
6.	»	1	1	2	BIOTIN BIOSYNTHESIS CYTOCHROME P4	Predicted: 2	63-64	14-14	0.00-0.00
7.	»	1	3	3	ALPHA-2,3/2,6-SIALYLTRANSFERASE/SIA	CSF: 3	60-63	14-19	0.09-0.12
8.	»	1	1	2	ARSENITE OXIDASE	Predicted: 2	61-62	14-14	0.00-0.00
9.	»	1	1	1	ASPARTATE AMINOTRANSFERASE	1ahgC: 1	62-62	14-14	0.06-0.06
10.	»	1	1	1	ACYL-COA OXIDASE	FAD: 1	60-60	14-14	0.06-0.06

Eidogen-Sertanty

Sequences | Chains | Sites | Binding Modes

Description						Site Residue Conservation
Site Name	Locus	Description	%Conf	SiteSorter	%ID	Non-polar Polar H-bond Polar/H-bond
pdb2rh1/s1276810 (chain A)	-	**CAU: (2S)-1-(9H-...**	100	-	-	.WT.DV.VT.F.T.YA.SS.SF.W.FF.N.Y.N.Y
pdb2vt4/s1421547 (chain B)	-	**P32: 4-[(2S)-3-(t...**	100	93.03	95	WT DV VT F T YA SS Sf W FF N F N Y
pdb2vt4/s1421552 (chain C)	-	**P32: 4-[(2S)-3-(t...**	100	93.81	95	WT DV VT F T YA SS Sf W FF N F N Y
pdb2vt4/s1421564 (chain C)	-	Predicted Site	78	104.37	95	WT DV VT F T YA SS SF W FF N F N Y
pdb3d4s/s1420913 (chain A)	-	Predicted Site	64	102.08	100	WT DV VT F t YA SS SF W FF N y N Y
pdb3d4s/s1420907 (chain A)	-	**TIM: (2S)-1-(tert-...**	100	97.93	100	WT DV VT F T YA SS SF W FF N Y N Y
pdb2vt4/s1421543 (chain A)	-	**P32: 4-[(2S)-3-(t...**	100	93.08	95	WT DV VT F T YA SS Sf W FF N F N Y
pdb2vt4/s1421560 (chain D)	-	**P32: 4-[(2S)-3-(t...**	100	91.69	95	WT DV VT F T YA SS Sf W FF N F N Y
pdb2vt4/s1421562 (chain A)	-	Predicted Site	64	101.26	95	WT DV VT F T YA SS Sf W FF N F N Y
pdb2vt4/s1421566 (chain D)	-	Predicted Site	65	99.89	95	WT DV VT F T YA SS Sf W FF N F N Y
pdb2z73/s1406997 (chain B)	**RHO**	**RET: RETINAL**	100	61.13	14	YG GG GF f y ni MF gF W YA A a V K
pdb2z73/s1406994 (chain A)	**RHO**	**RET: RETINAL**	100	61.76	14	YG GG GF f y ni MF gF W YA A a V K
pdb2z73/s1408905 (chain B)	**RHO**	Predicted Site	64	77.09	14	YG GG GF f y nI MF GF W YA a a V K
pdb2z73/s1408902 (chain A)	**RHO**	Predicted Site	77	70.81	14	YG GG GF f y ni MF GF W YA a a V K
pdb2ziy/s1406739 (chain A)	**RHO**	**RET: RETINAL**	100	62.56	14	YG GG GF f y ni MF GF W Ya a a V K
pdb3eml/s1491370 (chain A)	-	**ZMA: 4-{2-[(7-ami...**	100	49.29	14	ia VL Tq f p mv nf cv W LH N M I h
pdb2cbj/s824988 (chain B)	**NAGJ**	Predicted Site	71	-	-	
pdb2cbj/s824984 (chain A)	**NAGJ**	Predicted Site	92	-	-	
pdb2cbj/s824980 (chains A, B)	**NAGJ**	**OAN: O-(2-ACETA...**	100	54.72	19	W- NV V- - - -K -S -- - -- - - D V
pdb2cbj/s824983 (chains A, B)	**NAGJ**	**OAN: O-(2-ACETA...**	100	53.44	19	W- NV V- - - -K -S -- - -- - - D V
pdb2qk8/s1319712 (chain A)	-	**MTX: METHOTRE...**	100	53.66	14	-- EL VK L - -- -- -- M -F N - - T
pdb3e0b/s1504298 (chain B)	-	**N22: 5-[3-(2,5-di...**	100	45.21	14	-- EL VK L - -- -- -- M -F N - - -
pdb3e0b/s1504295 (chain A)	-	**N22: 5-[3-(2,5-di...**	100	43.71	14	L- EL V- L - -- -- -- M -F N - - -
pdb3e0b/s1504297 (chain B)	-	**NAP: NADP NICO...**	100	-	-	
pdb3e0b/s1504296 (chain A)	-	**NAP: NADP NICO...**	100	-	-	
pdb3e0b/s1504300 (chain B)	-	Predicted Site	70	-	-	
pdb3e0b/s1504299 (chain A)	-	Predicted Site	66	-	-	
pdb2qk8/s1319713 (chain A)	-	Predicted Site	76	-	-	
pdb1ith/s416096 (chain B)	**HBF1**	**HEM: PROTOPOR...**	100	45.54	14	-Q -V -L F - Y- LF -- K H- I M - -
pdb1ith/s416099 (chain B)	**HBF1**	Predicted Site	77	-	-	
pdb1ith/s416098 (chain A)	**HBF1**	Predicted Site	61	44.86	14	-Q -V -L F - Y- LF -- K H- - M - -
pdb1ith/s416094 (chain A)	**HBF1**	**HEM: PROTOPOR...**	100	43.93	14	-Q -V -L F - Y- LF -- K H- - M - -
pdb1lht/s1143974 (chain A)	**MB**	**HEM: PROTOPOR...**	100	47.88	14	-H -V -L F T -Y IL -- L H- I - - -
pdb1lhs/s1143884 (chain A)	**MB**	**HEM: PROTOPOR...**	100	47.69	14	-H -V -L F T -Y IL -- - H- I - - -
pdb1lhs/s1143885 (chain A)	**MB**	Predicted Site	94	46.66	14	-H -V -L F T -Y IL -- - H- I - - -
pdb1lht/s1143976 (chain A)	**MB**	Predicted Site	91	-	-	
pdb2nrm/s1298120 (chain A)	**MB**	Predicted Site	65	-	-	
pdb2nrm/s1298115 (chain A)	**MB**	**HEM: PROTOPOR...**	100	46.88	14	-- -V -L F T -- IF -- L H- - - - -
pdb1qkn/s484704 (chain A)	**ESR2**	**RAL: RALOXIFENE**	100	41.73	14	-- -L AT F I -H -L -- W M- G - - -

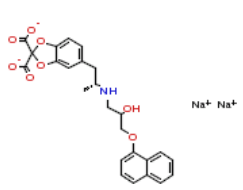
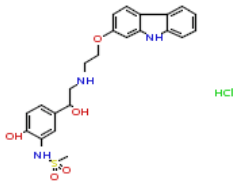
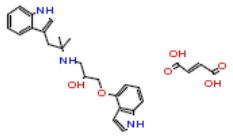
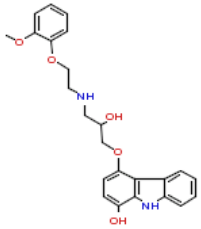
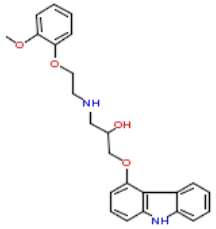
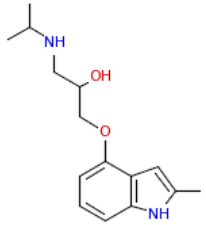
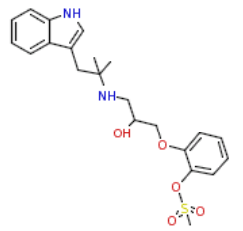
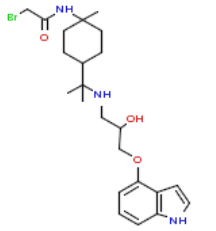
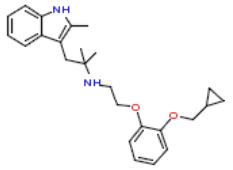
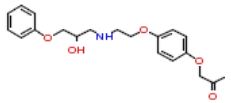
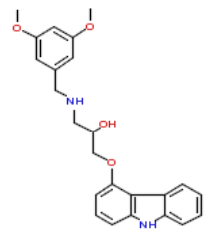
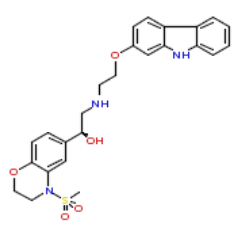
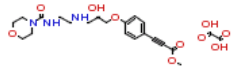
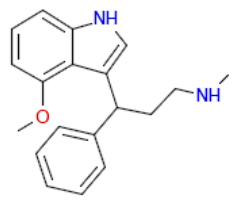
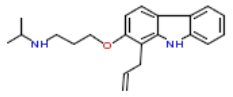
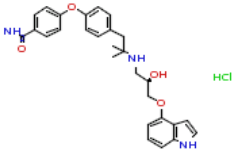
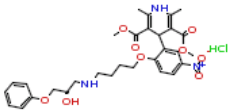
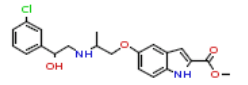
CAU (2RH1), P32 (2VT4), and TIM (3D4S)

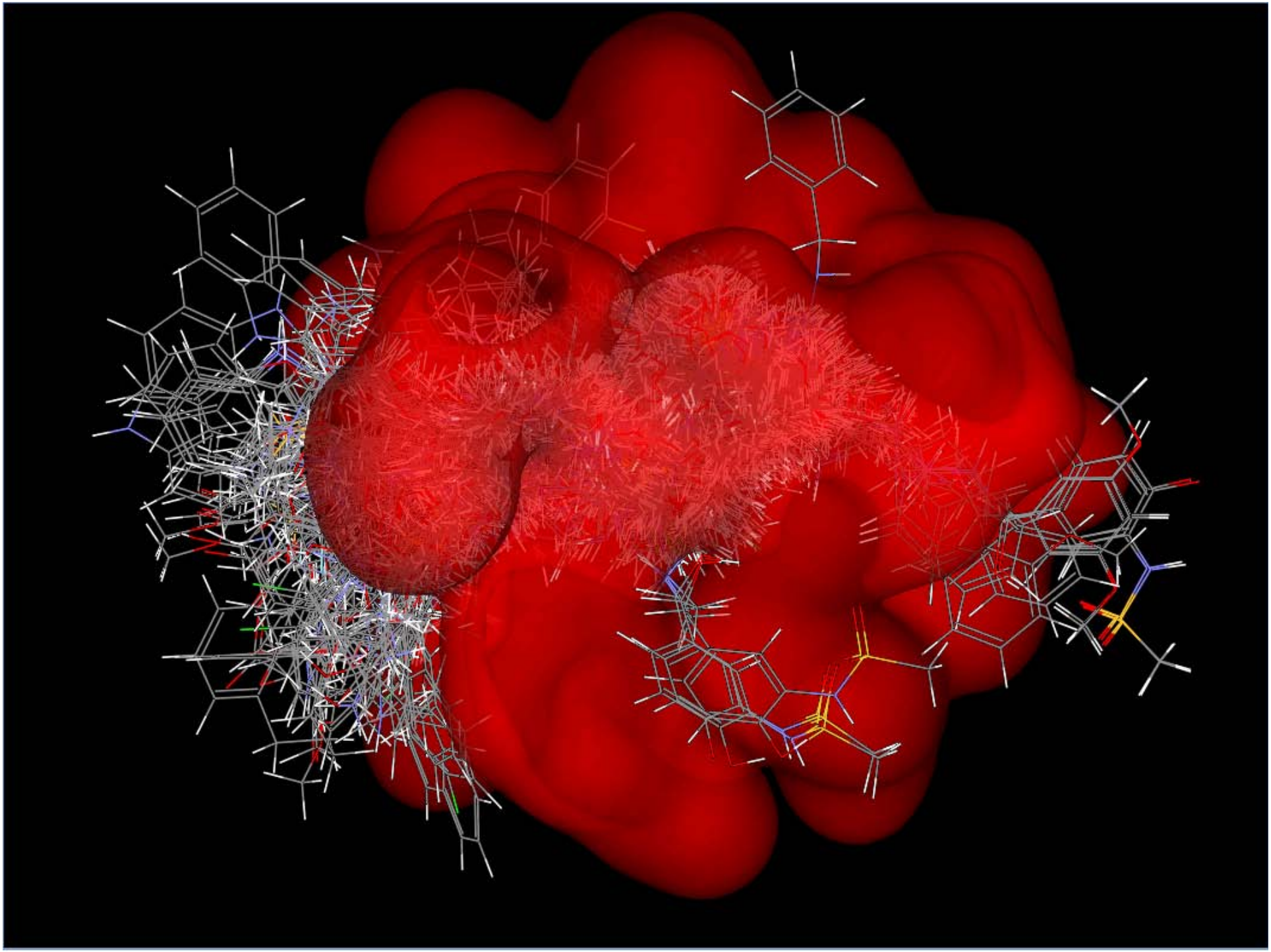


Chains | Chain Alignments | Sites | Site Alignments

Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2rh1/s1276810 (chain A)	-	**CAU**	100	.WT.DV.VT.F.T.YA.SS.SF.W.FF.N.Y.N.Y
pdb2vt4/s1421547 (chain B)	-	**P32**	100	.L.WT.DV.VT.S.C.F.T.YA.SS.S.W.FF.N.F.N.Y
pdb3d4s/s1420907 (chain A)	-	**TIM**	100	.WT.DV.VT.C.F.T.YA.SS.SF.W.FF.N.Y.N.Y

Example beta1-adrenergic blocker and CAU similar

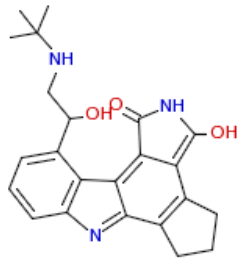
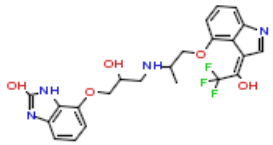
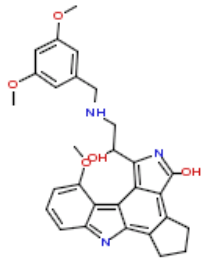
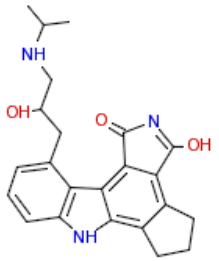
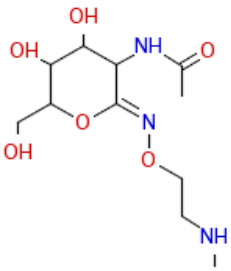
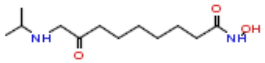
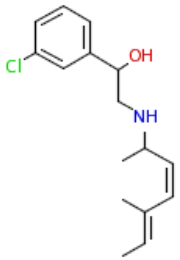
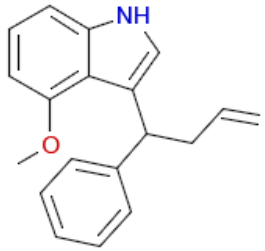
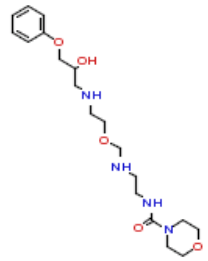
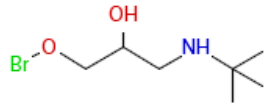
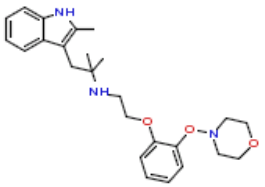
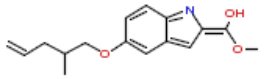
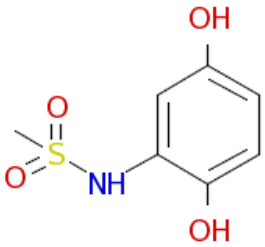
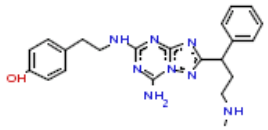
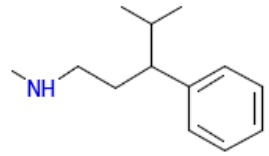
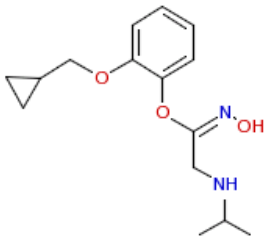
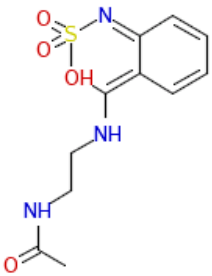
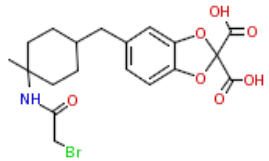
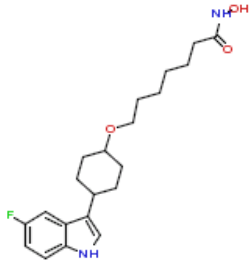
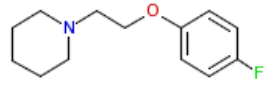
Docked into siteEID: 1276810 (Bioactive Molecules)					
Groups					
NumQueries: 1					
Chiral 					
232315 Antiobesity and antidiabetic agent, a specifically claimed compound within a series of benzodioxole dicarboxylates that exerts its action by virtue of its beta3-adrenoceptor-agonist activity, as demonstrated by stimulati	254159 Agent for the treatment of obesity, diabetes and hyperlipidemia, a beta3-adrenoceptor agonist, as demonstrated by stimulation of cAMP in CHO cells transfected with the human beta3-adrenocepto r (EC50 = 0.18 nM) and by st	135771 Antihypertensive agent, which decreases stimulating effect of isoproterenol and norepinephine in anesthetized dogs and is active in spontaneously hypertensive rats; another related indoyloxyp ropanolamine is: 141474	217630 A human metabolite of carvedilol that shows more potent antioxidant activity than parent compound in various models including iron-catalyzed lipid peroxidation in rat brain homogenates (IC50 = 0.30 mcmol/l vs. 8.1 mcmol/	090701 Antihypertensive, beta1-adrenergic blocker with additional alpha1-blocking properties. INDICATION - Essential hypertension. PRESENTATION - Tablets, 25 mg.	159104 Transdermal delivery system for mepindolol via TSD patch that appeared to be well tolerated and to induce pharmacodynamic equivalents of stable, prolonged and quite pronounced beta-adrenergic blockade within 1 week after
					Chiral 
175562 Oral antidiabetic agent with insulinomimetic and hypoglycemic activities. In vitro it activated glucose uptake by lipidic cells (epididymic lipid tissue of male rats) by 256% at a concentration of 1 mcM. After i.p. admin	156471 Anxiolytic and antimigraine agent that strongly and selectively binds to 5-HT1A-receptors (IC50 = 0.71 +/- 0.02 nM, against [3H]-8-OH-DPAT binding), with remarkable specificity in comparison to other 5-HT-receptors, e.g.	226668 alpha1-Adrenoceptor antagonist whose activity was evaluated in a receptor binding assay using rat submaxillary gland and liver membrane preparations, as well as in functional assays in rat is olated aorta and rat isolate	172623 Antiobesity agent having thermogenic activity in standard tests, giving an ED50 value for GDP binding in mitochondrias of rat interscapular brown adipose tissue of 0.55 mg/kg s.c. or p.o. Com pound induces very slight ca	450766 pSTAT3/IL-6 inhibitor that suppressed the growth of NCI-H460 (GI50 = 0.78 mcM), HCT 116 (GI50=8.5 mcM) and MDA-MB-231/U-251 (GI50 = 5 mcM) cells and inhibited IL-6 production in NCI-H460 cells (IC50 = 5 mcM) and in mice	291422 Agent for the treatment of obesity, diabetes and hyperlipidemia, a beta3-adrenoceptor agonist, as demonstrated by stimulation of cAMP in CHO cells transfected with the human beta3-adrenocepto r (EC50 = 12 nM; 69% activat
Chiral 			Chiral 	Chiral 	
174322 beta1-Adrenergic blocker with high selectivity, short-term activity and low toxicity. Remedy for angina pectoris, myocardial infarction, cardiac insufficiency,hypertension and arrhythmia. Inh ibition was shown to be 51.5	414953 Monoamine (5-HT and norepinephrine) reuptake inhibitor that displayed pKi values of 8.90 and 8.19, respectively, when tested for affinity at human 5-HT (SERT) and norepinephrine (NET) transporters. It is considered to ha	263881 Hypolipidemic agent, a potent inhibitor of squalene synthase (IC50 = 32 nM in human hepatoma HepG2 cells). It reduced blood cholesterol levels in rats fed a cholesterol-enriched diet by 50% at a dose of 50 mg/kg p.o. Ano	250527 Agent for the treatment of obesity and type II diabetes, a selective beta3-adrenoceptor agonist with minimal or no beta1- and beta2-adrenoceptor-stimulatory effects. Other specifically claime d compounds include the foll	146217 Calcium antagonist and selective beta1-adrenergic blocker, the 2S-isomer of YM-16151-1. When orally administered to rats (3 - 30 mg/kg) and dogs (0.3 - 10 mg/kg), it produces dose-dependent hypotension without increasing	239356 Antidiabetic and antiobesity agent that acts by virtue of its beta3-adrenoceptor-agonist activity (EC50 = 85 nM for increasing cAMP levels in CHO cells expressing human receptor). Its activity was also determined by meas

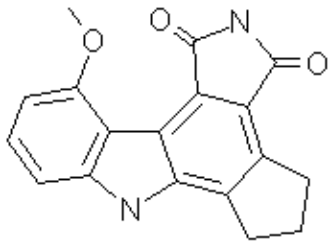


LigandCross Diverse Examples

Groups

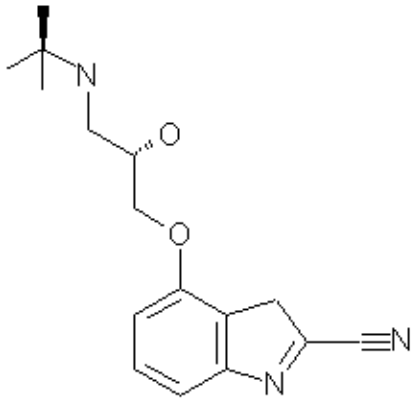
NumQueries: 1

 396707_P32_20 LC: 0.818 DL: 16.41	 263341_257094_35 LC: 0.674 DL: 16.53	 450766_396707_3 LC: 0.818 DL: 13.02	 396707_257094_15 LC: 0.614 DL: 15.30	 414953_OAN_11 LC: 0.465 DL: 19.22
 328954_CAU_7 LC: 0.500 DL: 15.74	 239356_RET_7 LC: 0.511 DL: 14.83	 263881_414953_2 LC: 0.659 DL: 10.75	 172623_174322_26 LC: 0.313 DL: 21.04	 156471_TIM_4 LC: 0.452 DL: 14.33
 226668_TIM_3 LC: 0.400 DL: 15.82	 239356_387297_42 LC: 0.619 DL: 10.22	 254159_263341_3 LC: 0.558 DL: 11.24	 414953_ZMA_1 LC: 0.568 DL: 10.66	 414953_164327_1 LC: 0.568 DL: 10.45
 226668_164327_7 LC: 0.500 DL: 11.02	 284771_229728_2 LC: 0.457 DL: 10.29	 232315_156471_2 LC: 0.386 DL: 11.58	 281697_328954_1 LC: 0.383 DL: 10.75	 263341_RAL_11 LC: 0.205 DL: 13.50



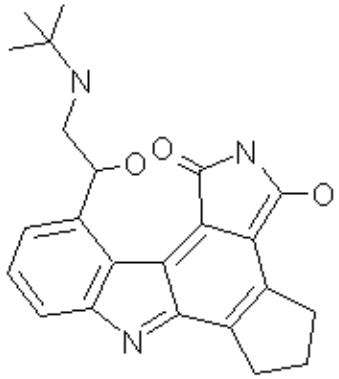
396707

Potent and selective poly(ADP-ribose) polymerase-1 (PARP-1) and PARP-2 inhibitor (IC50 = 20 and 6 nM, respectively) that demonstrated chemosensitizing properties at a concentration of 1 mcM when combined with temozolomid



P32 (2VT4)

2VT4: TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND CYANOPINDOLOL



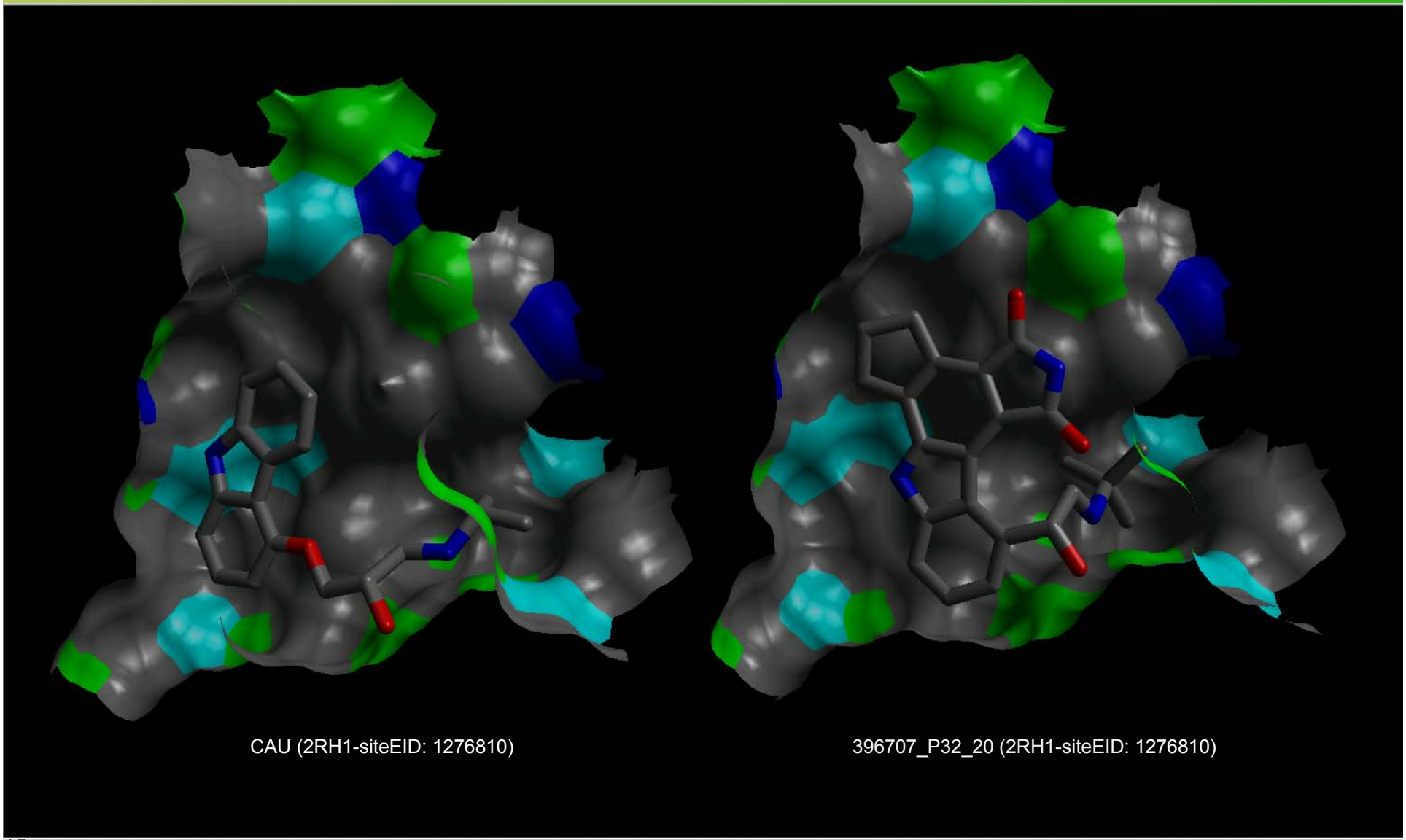
396707_P32_20

LigandCross result for siteEID:2RH1-siteEID: 1276810

Sequences | Chains | Sites | Binding Modes

Description				Binding Modes
Site Name	Locus	Description	Contact	Non-polar Polar H-bond Polar/H-bond
pdb2rh1/s1276810 (chain A)	-	**CAU: (2S)-1-(9H-Carba...**	-	.WT.DV.VT.F.T.YA.SS.SF.W.FF.N.Y.N.Y
pdb2rh1/s1276810_7 (chain A)	-	450766_396707_3	0.79	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_286 (chain A)	-	450766_396707_3	0.79	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_1 (chain A)	-	450766_396707_3	0.79	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_280 (chain A)	-	450766_396707_3	0.79	WT DV VT F T YA SS SF W FF N Y N Y
pdb2vt4/s1421543 (chain A)	-	**P32: 4-{[(2S)-3-(tert-bu...**	0.77	WT DV VT F T YA SS S- W FF N F N Y
pdb2vt4/s1421560 (chain D)	-	**P32: 4-{[(2S)-3-(tert-bu...**	0.77	WT DV VT F T YA SS S- W FF N F N Y
pdb2rh1/s1276810_13 (chain A)	-	396707_257094_15	0.74	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_292 (chain A)	-	396707_257094_15	0.74	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_67 (chain A)	-	414953_ZMA_1	0.73	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_346 (chain A)	-	414953_ZMA_1	0.73	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_18 (chain A)	-	263341_257094_35	0.73	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_31 (chain A)	-	414953_ZMA_1	0.73	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_297 (chain A)	-	263341_257094_35	0.73	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_310 (chain A)	-	414953_ZMA_1	0.73	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_17 (chain A)	-	414953_ZMA_1	0.72	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_296 (chain A)	-	414953_ZMA_1	0.72	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_107 (chain A)	-	263341_257094_35	0.71	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_386 (chain A)	-	263341_257094_35	0.71	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_42 (chain A)	-	263341_257094_35	0.71	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_321 (chain A)	-	263341_257094_35	0.71	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_3 (chain A)	-	450766_396707_3	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_24 (chain A)	-	450766_396707_3	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_282 (chain A)	-	450766_396707_3	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_303 (chain A)	-	450766_396707_3	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_19 (chain A)	-	396707_257094_15	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_27 (chain A)	-	396707_257094_15	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_29 (chain A)	-	396707_257094_15	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_298 (chain A)	-	396707_257094_15	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_306 (chain A)	-	396707_257094_15	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_308 (chain A)	-	396707_257094_15	0.70	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_164 (chain A)	-	226668_164327_7	0.69	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_211 (chain A)	-	263881_414953_2	0.69	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_443 (chain A)	-	226668_164327_7	0.69	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_490 (chain A)	-	263881_414953_2	0.69	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_138 (chain A)	-	239356_387297_42	0.68	WT DV VT F T YA SS SF W FF N Y N Y
pdb2rh1/s1276810_417 (chain A)	-	239356_387297_42	0.68	WT DV VT F T YA SS SF W FF N Y N Y
pdb3d4s/s1420907 (chain A)	-	**TIM: (2S)-1-(tert-butyla...**	0.67	WT DV VT F T YA SS SF W FF N Y N Y

Example Docked LigandCross Result (siteEID: 1276810)

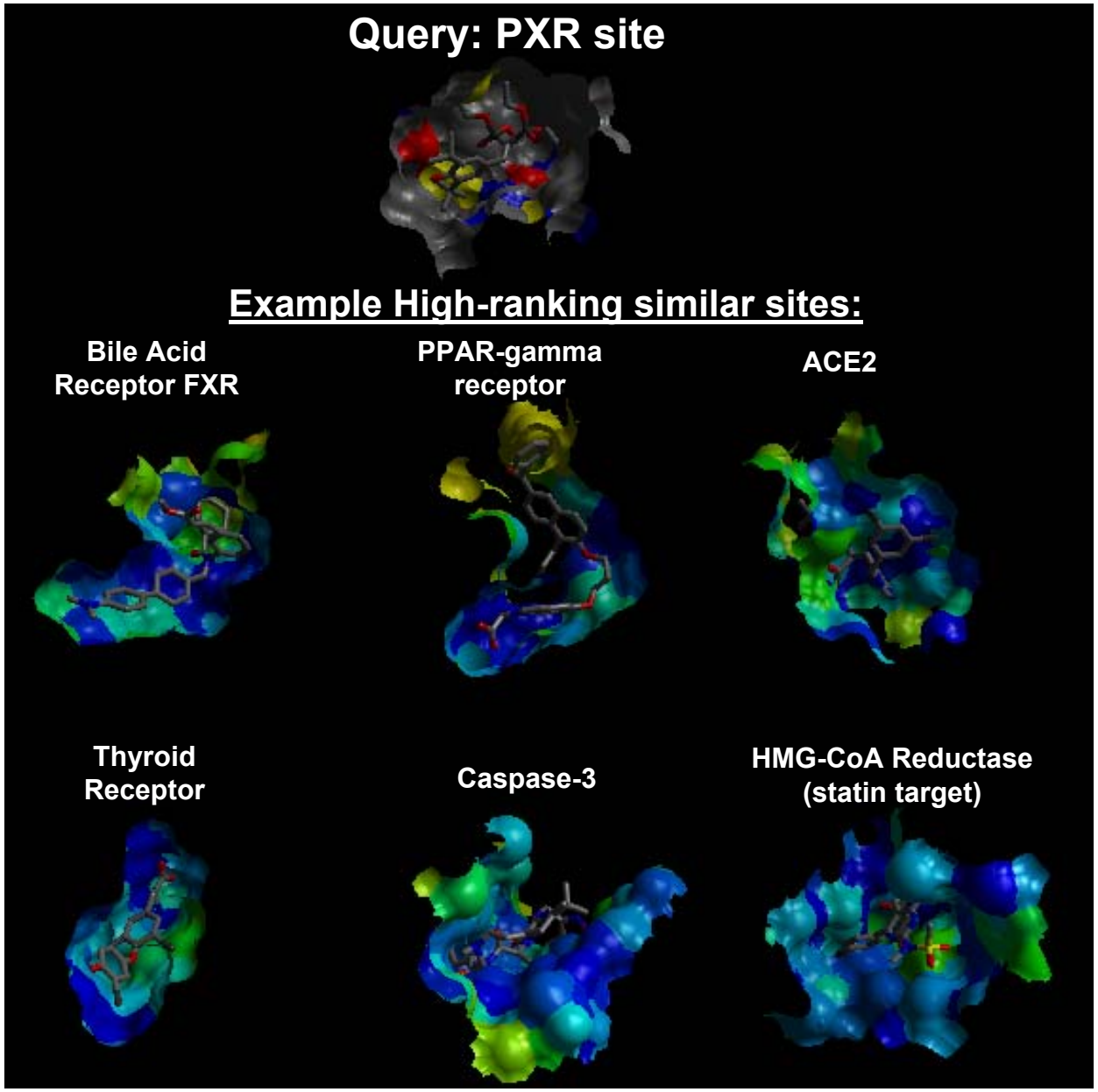


Chains | Chain Alignments | Sites | Site Alignments

Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2rh11/s1276810 (chain A)	-	CAU	100	.WT.DV.VT.F.T.YA.SS.SF.W.FF.N.Y.N.Y
pdb2rh1/s1276810_6 (chain A)	-	396707_P32_20	100	.WT.DV.VT.F.T.YA.SS.SF.W.FF.N.Y.N.Y

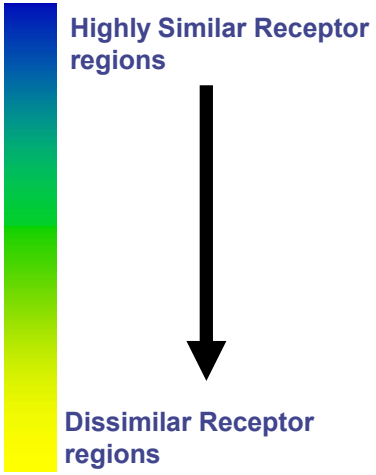
Add’l slides

Nature Exploits Site Similarity...



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-ãantagonist, xenobiotics and endobiotics…

Site Similarity Coloring



Eidogen-Sertanty

Multi-Kinase Inhibitors

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

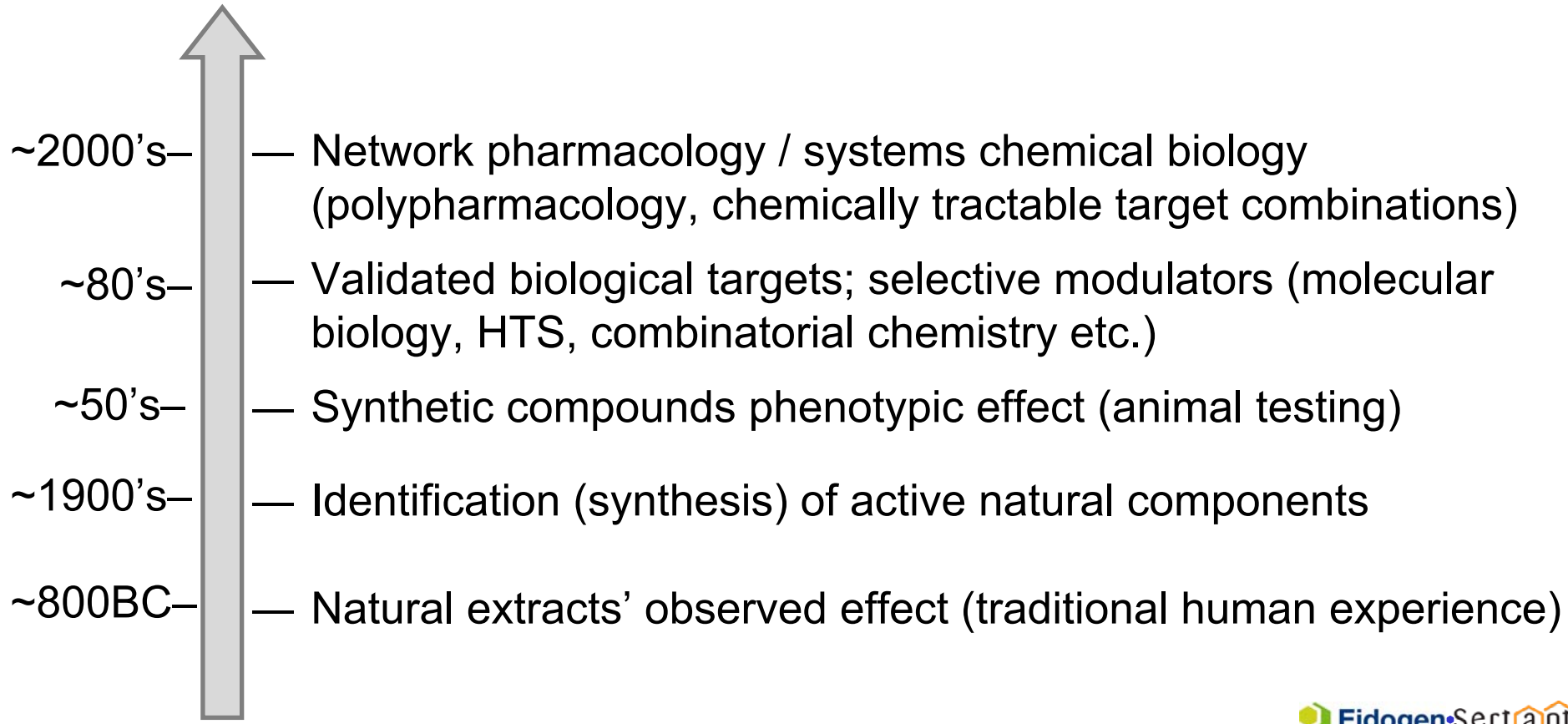
*Indication given for highest phase; all drugs are also in lower phase clinical trials for other oncology indications. ALL, acute lymphoblastic leukaemia; BMS, Bristol–Myers Squibb; CML, chronic myeloid leukaemia; EGFR, epidermal growth factor receptor; GSK, GlaxoSmithKline; NSCLC, non-small-cell lung 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**

Rationalizing Drug Discovery

From: Individual biological target → "Selective" compounds

To: Target combinations → Multi-target compound (combinations)



Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from

~ 5800 journal articles/patents

- **KKB Content Summary (Q2-2009):**

  # of kinase targets: **~400**
  # of SAR Data points: **~415,000**
  # of **unique** kinase molecules with SAR data: **~150,000**
  # of annotated assay protocols: **~19,000**
  # of annotated chemical reactions: **~2,300**
  # of unique kinase inhibitors: **~495,000** **(~340K enumerated from patent chemistries)**

- **KKB Growth Rate:**

  - Average **15-20K** SAR data points added per quarter
  - Average **20-30K** unique structures added per quarter

Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from

Kinase Validation Set

Three sizable datasets freely available to the research community

http://www.eidogen-sertanty.com/kinasednld.php

Eidogen-Sertanty

About Eidogen-Sertanty

- Knowledge-Driven Solutions Provider
  - Sertanty established in 2003, acquired Libraria assets
  - Sertanty acquired Eidogen/Bionomix in 2005→ Eidogen-Sertanty
  - $20M invested: Libraria ($6M), Eidogen/Bionomix ($12M), Sertanty/ES ($2M)
  - 14 distributed FTE's (4 US and 10 India)
  - Worldwide (bio)pharmaceutical customer base
  - Cash-positive since 2006

- Databases & Software – Annual Subscriptions
  - *TIP*™ - Protein Structural Informatics Platform
  - *KKB*™ - Kinase SAR and Chemistry Knowledgebase
  - CHIP™ - Chemical Intelligence Platform

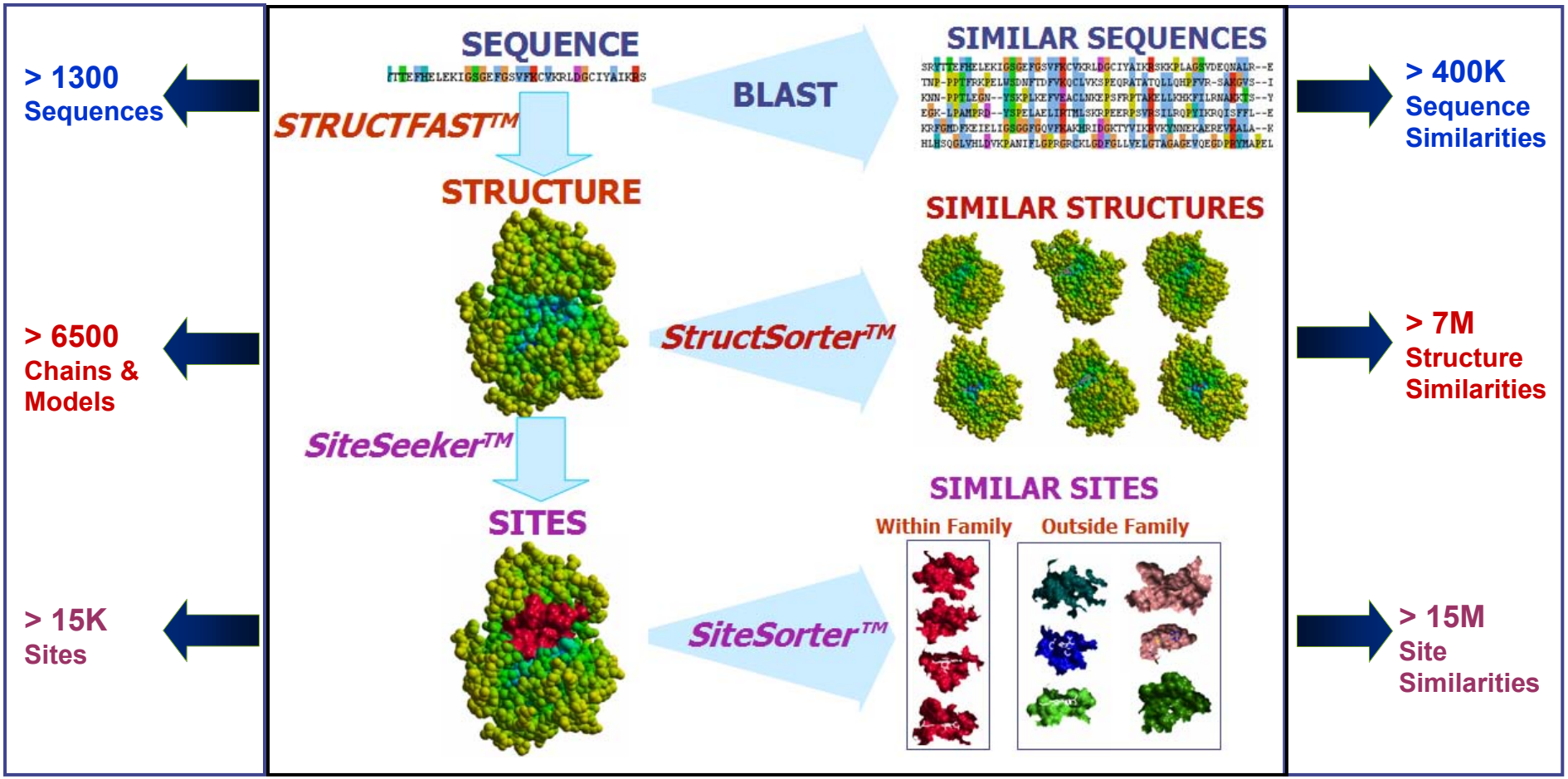
- DirectDesign™ Fee-For-Service
  - In Silico Target Screening ("Target Fishing" and Repurposing)
  - Target and compound prioritization services
  - Fast Follower Design: Novel, Patentable Leads

Eidogen·Sertanty

TIP/Kinase – 2009 Promotional Bundle

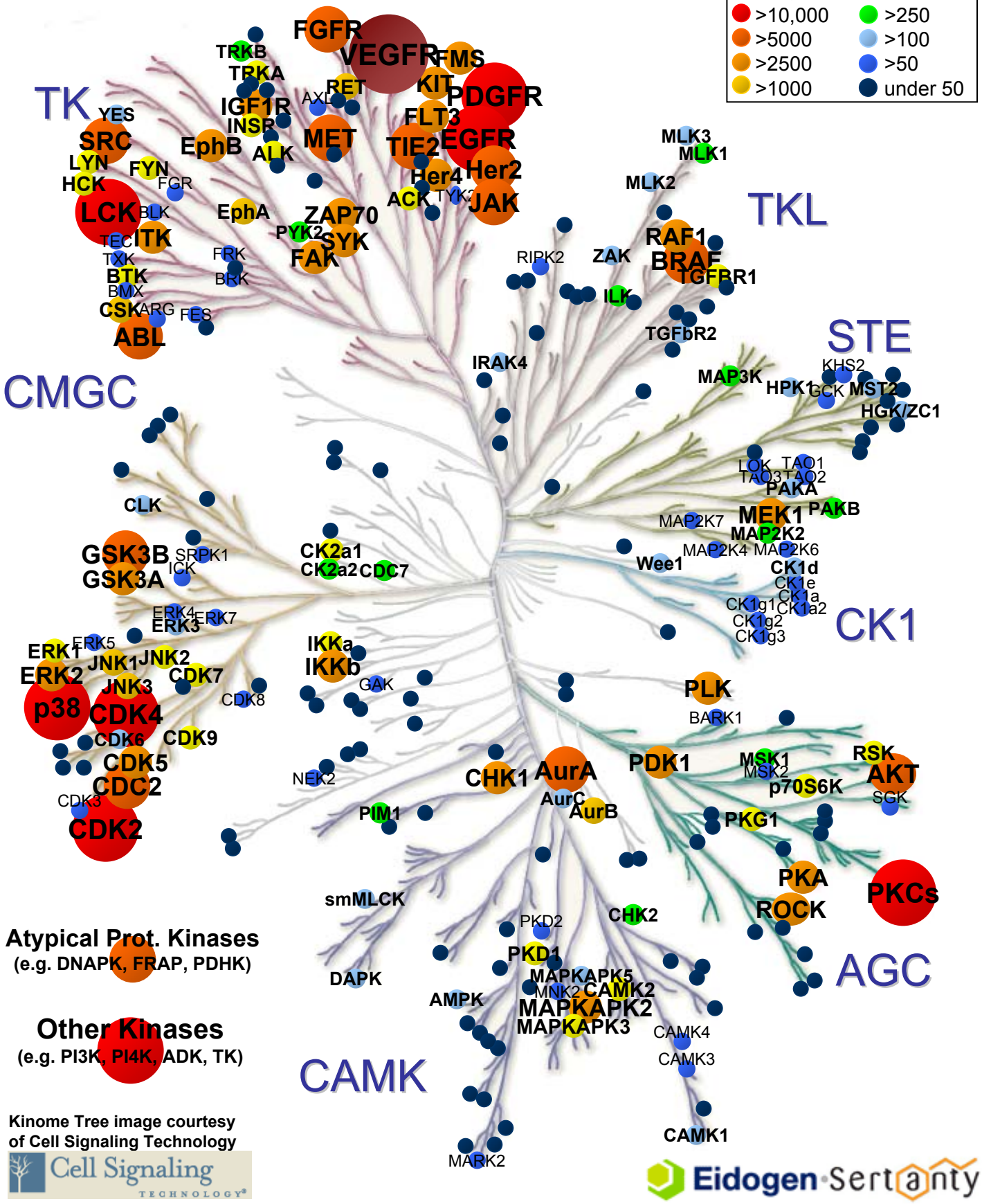
- **TIP/Workgroup technology**
  - Behind-the-firewall with web interface and commandline utilities
  - TIP database creation, administration, and update capabilities
  - Optionally available module: TIP/Webservices
- **TIP Kinase Family Database**
  - Over 1300 sequences (~500 human) modeled w/multiple templates
  - Over 4000 models derived from over 1200 PDB templates
  - Over 7M structure and over 15M siteSimilarities
  - Over 620 co-complexed ligands
- **One-year subscription to Kinase Knowledgebase (KKB exports)**
  - Over 402,000 SAR datapoints from over 5,500 articles/patents

TIP/Kinase Content



Reference: Interrogating the druggable genome with structural informatics, MolecularDiversity (2006)

Distribution of Kinase Knowledgebase Bio-Activity Data Points
Q1 2009 Release
Human Kinome Tree View



Atypical Prot. Kinases
(e.g. DNAPK, FRAP, PDHK)

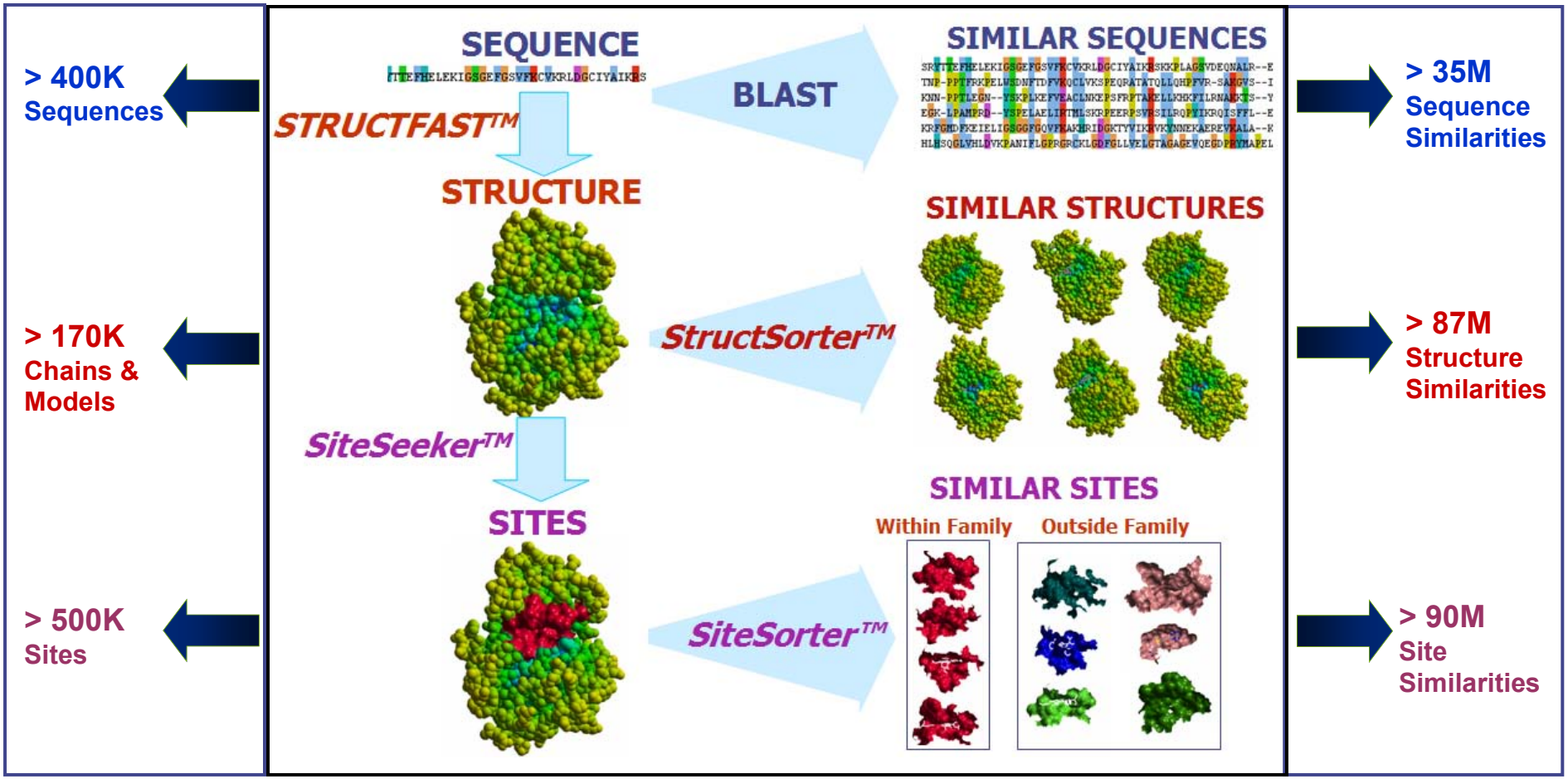
Other Kinases
(e.g. PI3K, PI4K, ADK, TK)

Kinome Tree image courtesy of Cell Signaling Technology

Cell Signaling TECHNOLOGY

Eidogen-Sertanty

TIP Content and Algorithm Engine



- 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.**

Eidogen-Sertanty

Kinome Viewer - KKB and TIP

Data Integration : Kinase Knowledgebase

Application Integration : Target Informatics Platform

accelrys®

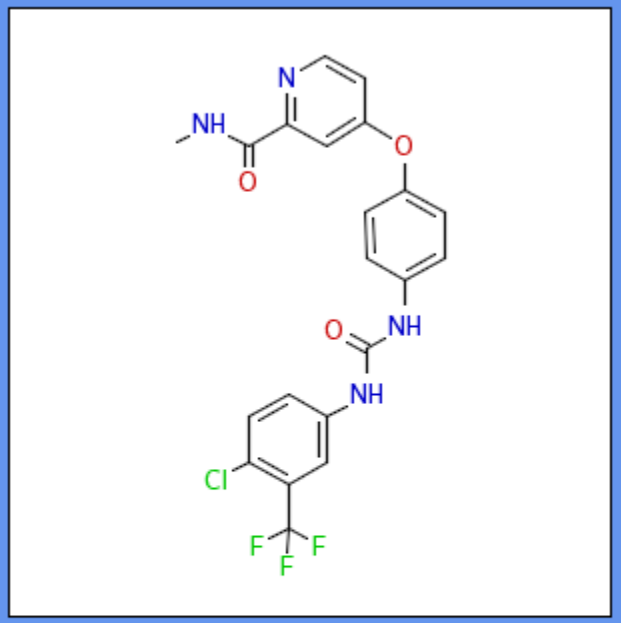
Eidogen•Sertanty

Kinome Viewer with the Kinase Knowledgebase (KKB)

Protocols available for KV-KKB (kkbq109 - 10% sample)

- KKB-Search by Chemical Structure

Kinase Knowledgebase (kkbq109 - 10% sample)
Results Displayed in Accelrys Kinome Viewer

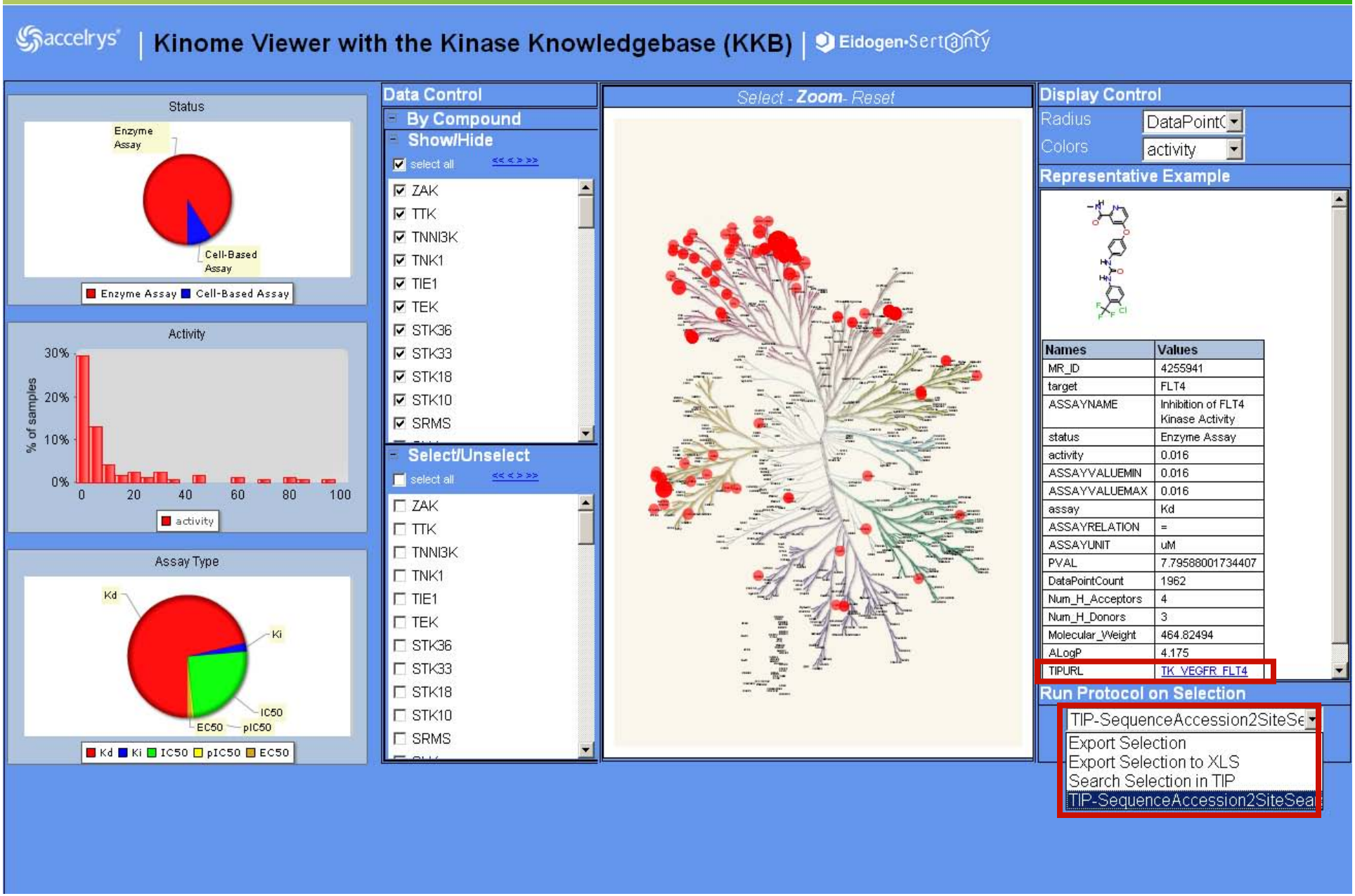


- (•) SSS
- ( ) Exact
- ( ) Sim

80

Search

Kinome Viewer (KV): Data and Application Integration



TIP Access: One-Click Away

Sequence 158401

Add To Project
Protein Search

"FLT4 kinase catalytic domain (Aliases: PCL,VEGFR3)"

Sequence Headers

Accession	Locus	Database	Enzyme Family	Species
TK_VEGFR_FLT4	FLT4_Hsap	kinbase		Homo sapiens

Chains

Source	Start	Stop	Identifier	Knowledge Level	#Sites	Primary Template
STRUCTFAST	1	325	182419	3	2	1ywnA 62%
STRUCTFAST	1	325	198501	3	2	1oecA 44%
STRUCTFAST	1	325	238701	3	1	1t46A 47%
STRUCTFAST	1	325	296001	3	2	1fgiA 43%
STRUCTFAST	1	325	429801	3	3	1ir3A 32%
STRUCTFAST	1	325	504101	3	3	2i1mA 45%

Sequence Domains

Pfam ID	Description	Start	Stop

Amino Acid Sequence

```
  1    LHLGR VLGYG AFGKV VEASA FGIHK GSSCD TVAVK MLKEG ATASE HRALM SELKI LIHIG    60
 61    NHLNV VNLLG ACTKP QGPLM VIVEF CKYGN LSNFL RAKRD AFSPC AEKSP EQRGR FRAMV   120
121    ELARL DRRRP GSSDR VLFAR FSKTE GGARR ASPDQ EAEDL WLSPL TMEDL VCYSF QVARG   180
181    MEFLA SRKCI HRDLA ARNIL LSESD VVKIC DFGLA RDIYK DPDYV RKGSA RLPLK WMAPE   240
241    SIFDK VYTTQ SDVWS FGVLL WEIFS LGASP YPGVQ INEEF CQRLR DGTRM RAPEL ATPAI   300
301    RRIML NCWSG DPKAR PAFSE LVEIL                                           325
```

Eidogen•Sertanty

Extracting Kinase Data Sets

- Only enzymatic (homogeneous) assays with defined target
- Only high quality data (IC50, Ki, Kd)
- Standardizing chemical structures (salt forms, stereochemistry, E/Z geometry, tautomers, ionization)
- Kinase target Entrez Gene names and SwissProt accessions
- Aggregate data by structure first in an individual experiment and then globally by unique kinase and structure

➢ 189,119 unique (structure target) data points (366 kinases)

➢ 93,121 unique structures

Relating Kinase Targets by Compound Activity

- "ACTivity similarity" for compounds tested in common - which are active for one (or both) target(s)

$$ACTsim_{ij} = 1 - \frac{1}{N}\sum_{k=1}^{N>2}\frac{\left|pIC50_{ki} - pIC50_{kj}\right|}{\max pIC50_{diff}}$$

Vieth et.al. "Kinomics" Biochim Biophys Acta **2004** 243

- Activity cutoff pVal ≥ 6.5; minimum 20 actives per kinase pair
- Compute Minimum spanning tree (Kruskal)
  - Visualization as network tree (Cytoscape)

Side note: "Activity fingerprint" (for a comprehensive activity matrix)

Bamborough et.al. J Med Chem **2008**, 7898

Eidogen·Sertanty

Relating Kinase Targets by SARsim 'Features'

- Laplacien-modified Naïve Bayesian models using FCFP_4 fingerprints
  - Measure contribution of a bit in a fingerprint for a specific outcome
  - Assume all variables are independent
  - A compound is scored by summing the weights of its fingerprint bits
- Kinase models compared by the Pearson correlation coefficient of the vector of the probabilistic weights (log of Avidon weights) of all fingerprint bits

  Adopted from Schuffenhauer *Org Biomol Chem* **2004** 3256
- Activity cutoff pIC50 > 6.5; all other compounds negative
- Select models with ROC > 0.8 and minimum 20 actives
- Compute the correlation matrix

Kinase Target Similarity by ACTsim/SARsim