Using Sequence-, Structure- and Receptor-site Similarities to Generate New Matter Ideas within the Kinome

Steven Muskal, Ph.D.

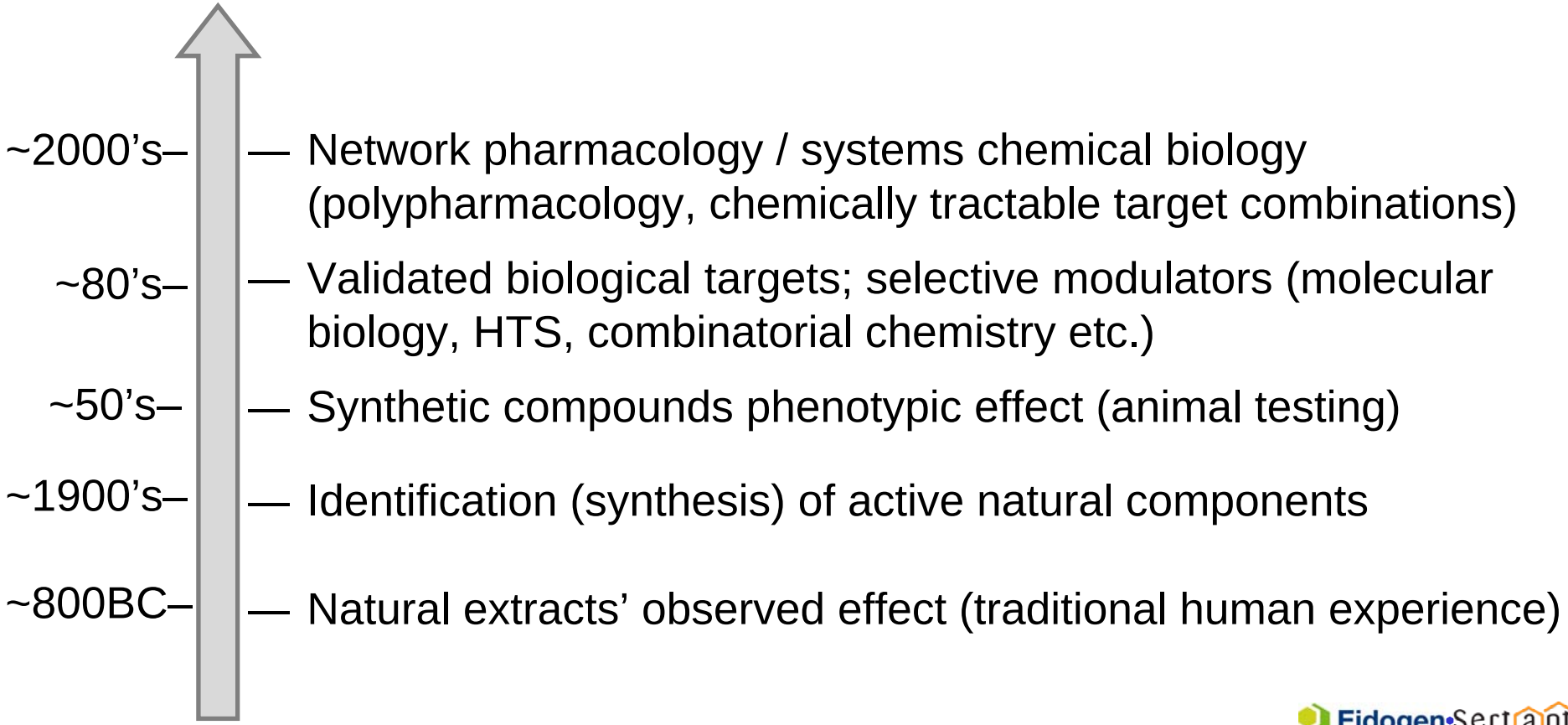
Chief Executive Officer
Eidogen-Sertanty, Inc.

smuskal@eidogen-sertanty.com

Motivation

From: Individual biological target → “Selective” compounds

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



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

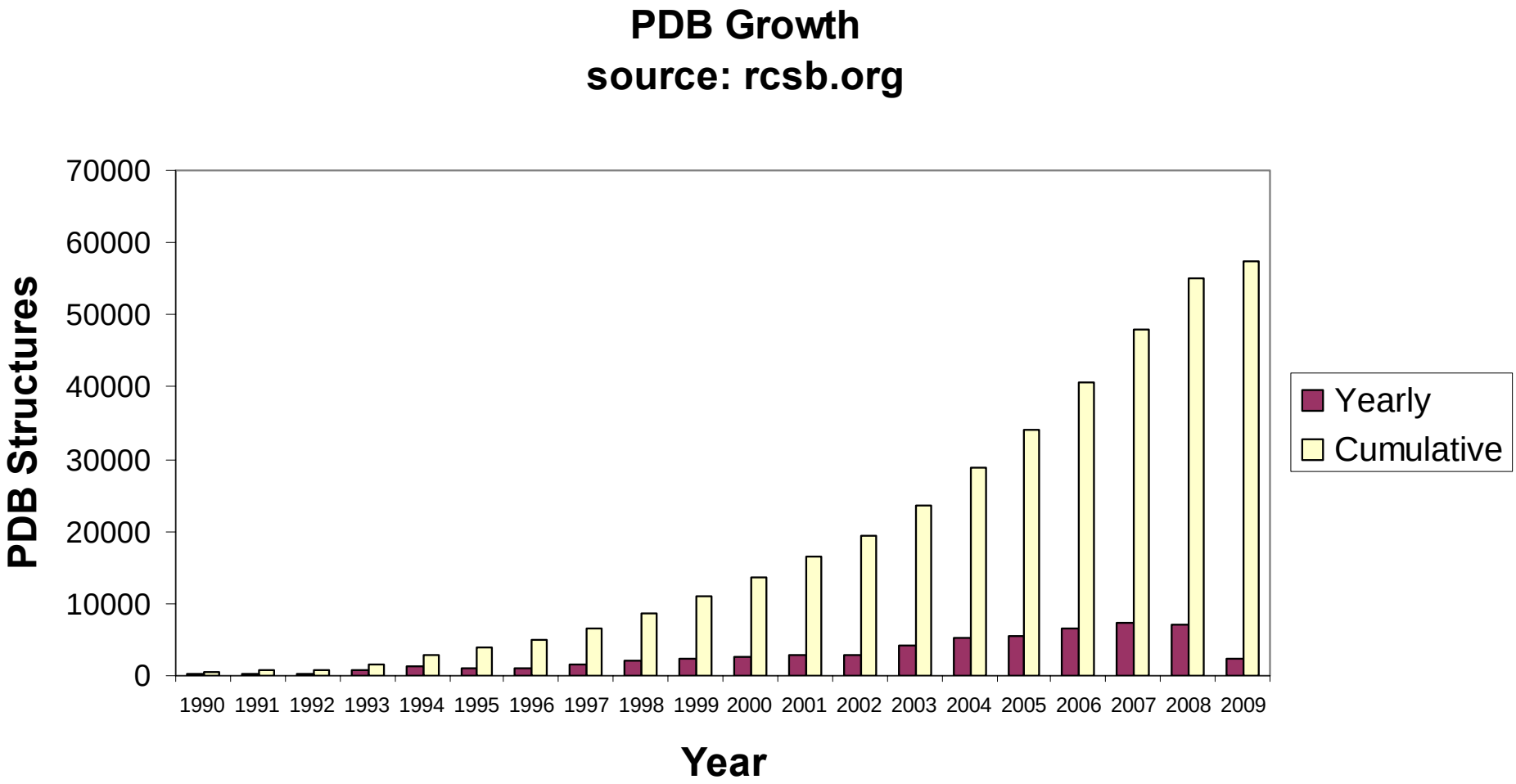
*Indication given for highest phase; all drugs are also in lower phase clinical trials for other oncology indications. ALL, acute lymphoblastic leukaemia; BMS, Bristol–Myers Squibb; CML, chronic myeloid leukaemia; EGFR, epidermal growth factor receptor; GSK, GlaxoSmithKline; NSCLC, non-small-cell lung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

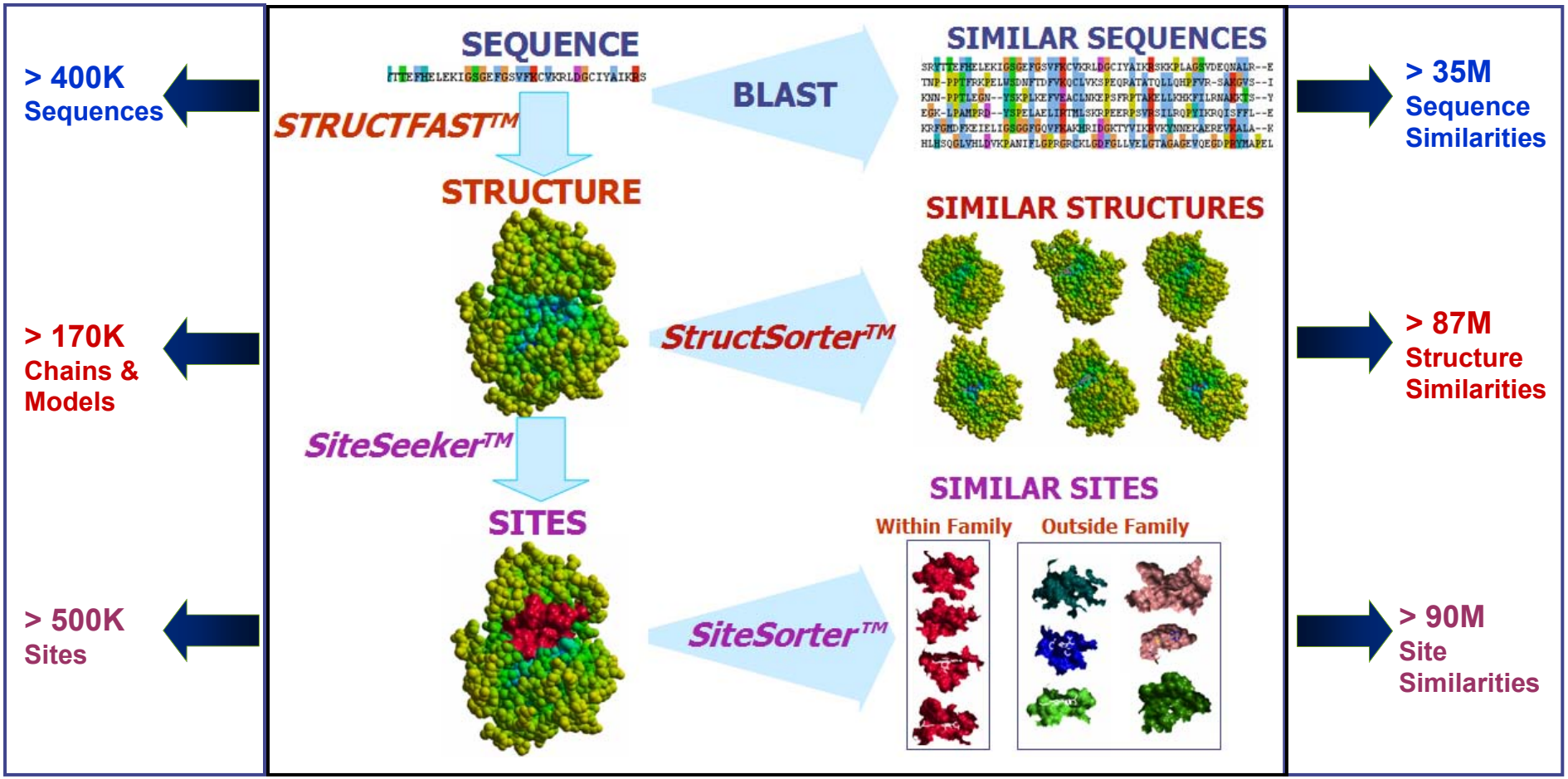
Eidogen•Sertanty

Protein Structure Growth Continues

> 50K Structures/co-complexes (Apr-2008)
> 600 deposits per month → **>150/week!**



TIP Content and Algorithm Engine

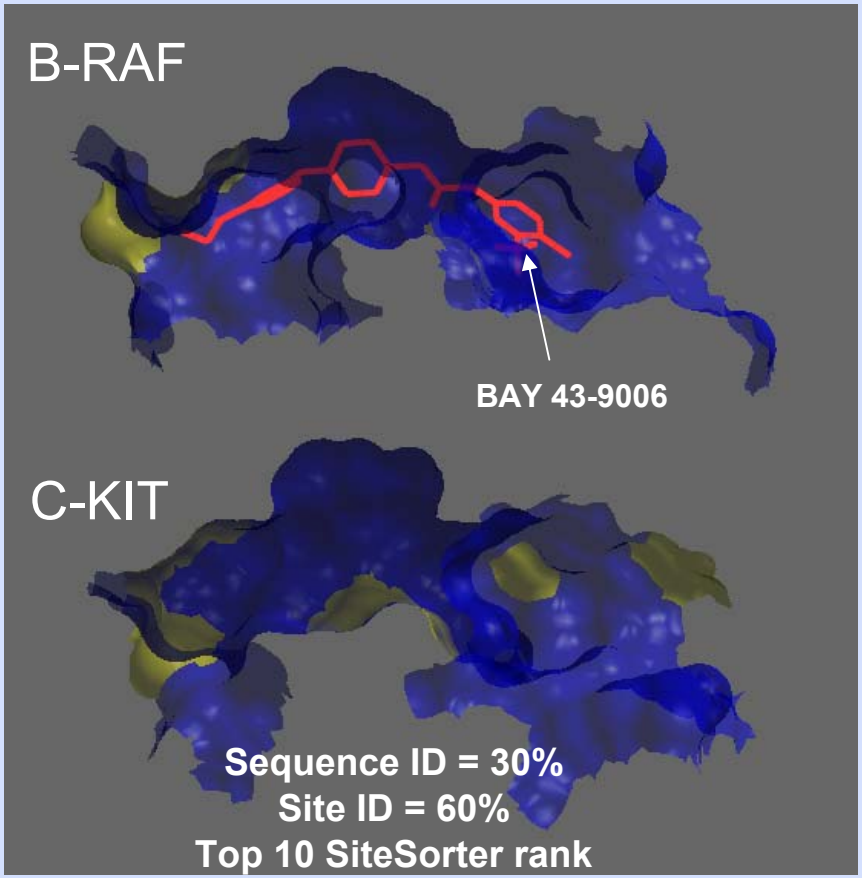


- Interrogating the druggable genome with structural informatics **MolecularDiversity (2006)**
- STRUCTFAST: Protein Sequence Remote Homology Detection and Alignment Using Novel Dynamic Programming and Profile-Profile Scoring **Proteins. 2006 64:960-967**
- StructSorter: A Method for Continuously Updating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

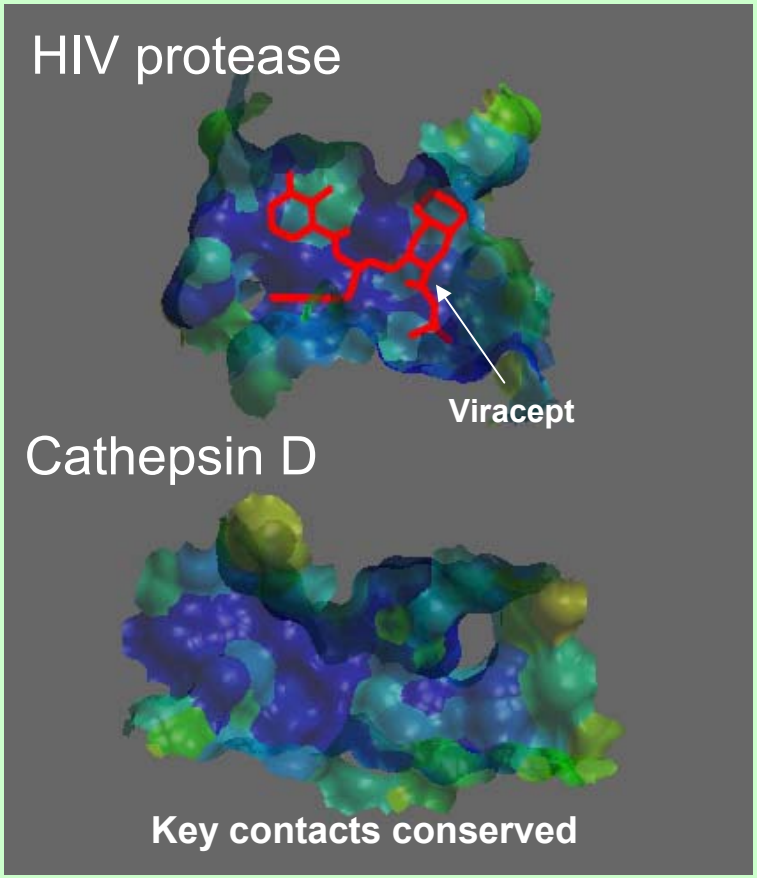
Off-Target Opportunities

Intra-Family Opportunities



B-RAF inhibitor BAY 43-9006 also inhibits C-KIT

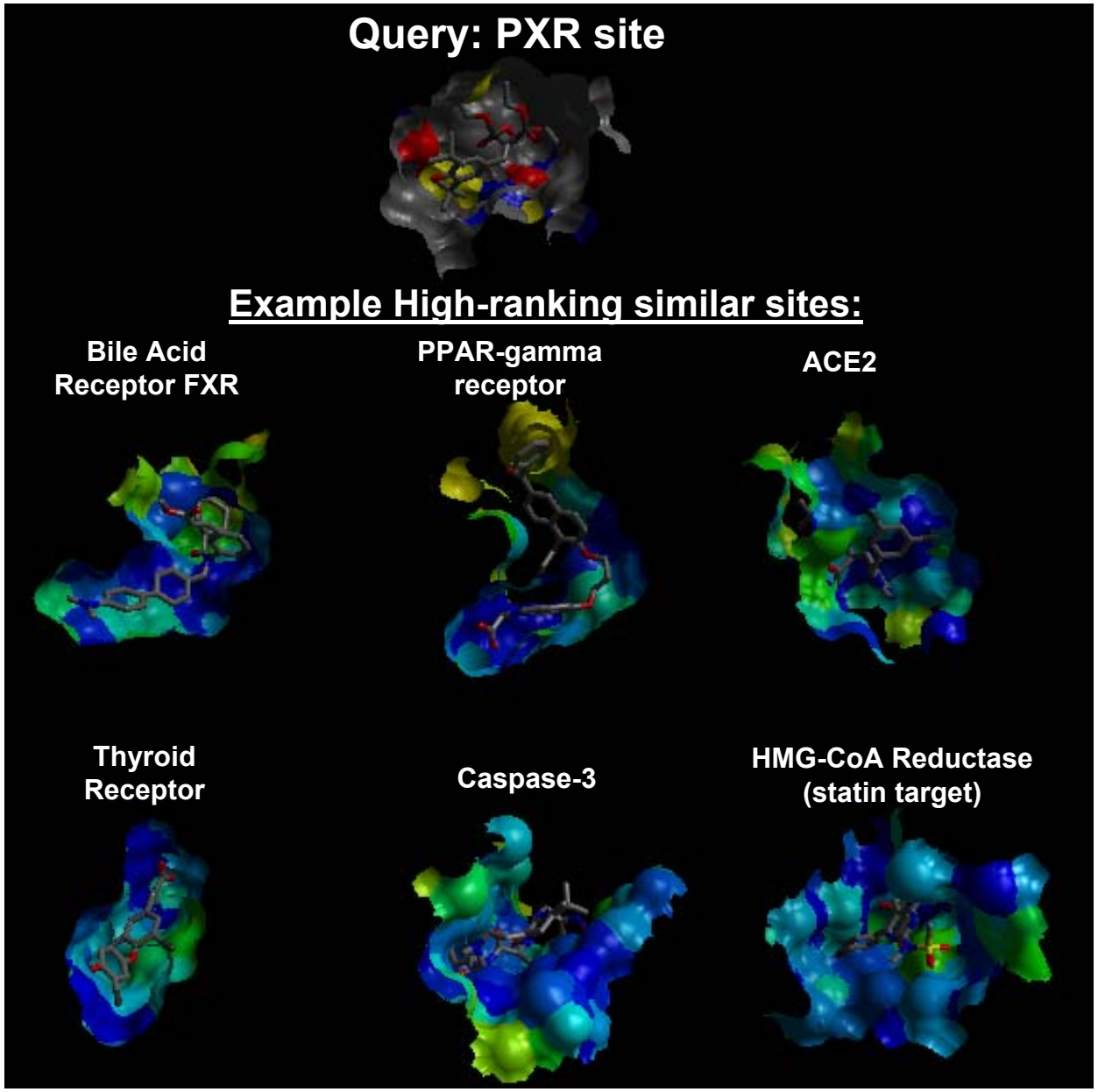
Inter-family Opportunities



Cathepsin D is inhibited by HIV protease inhibitors

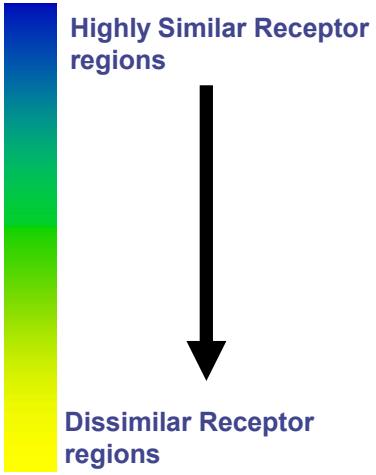
Eidogen-Sertanty

Nature Exploits Site Similarity...



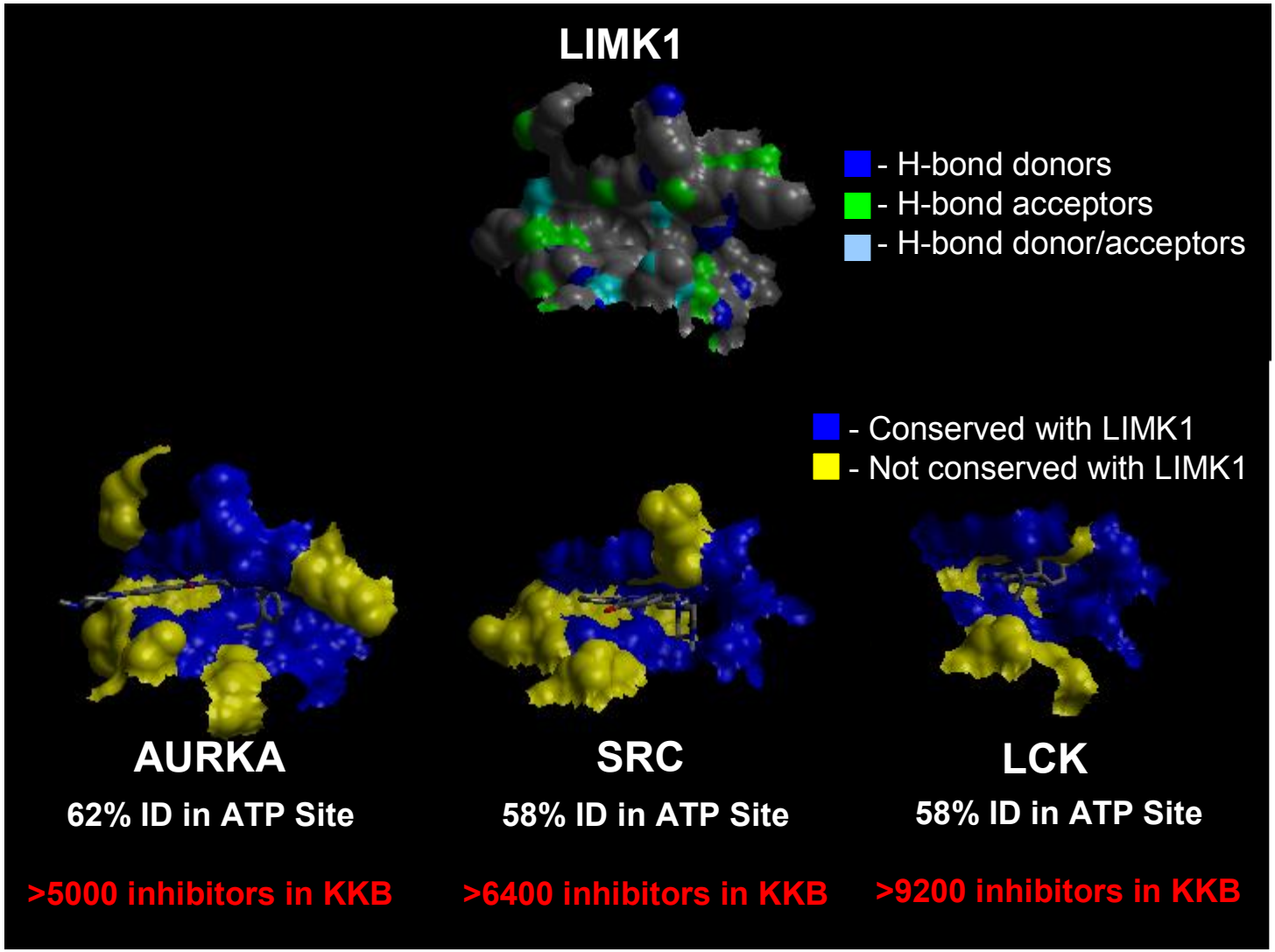
Pregnane X-receptor – PXR ("sensor)" →CYP3A4 ("executioner")
PXR Binds > 50% drugs
Including some bile acids, statins, herbal components, a selection of HIV protease inhibitors, calcium channel modulators, numerous steroids, plasticizers and monomers, organochlorine pesticides, a peroxisome proliferator-activated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

Site Similarity Coloring



Eidogen-Sertanty

Borrowing Matter Ideas using Site Similarity



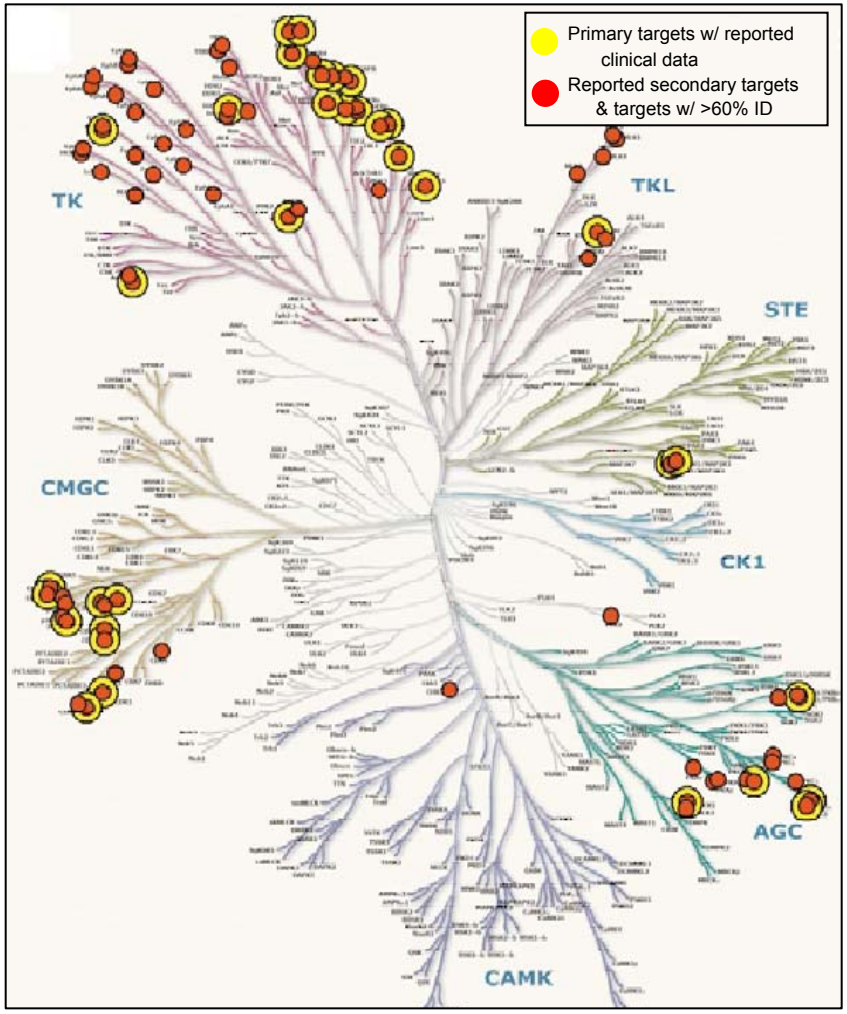
The ATP site of LIMK1 shares a high level of homology with several well-studied kinases

Eidogen Sertanty

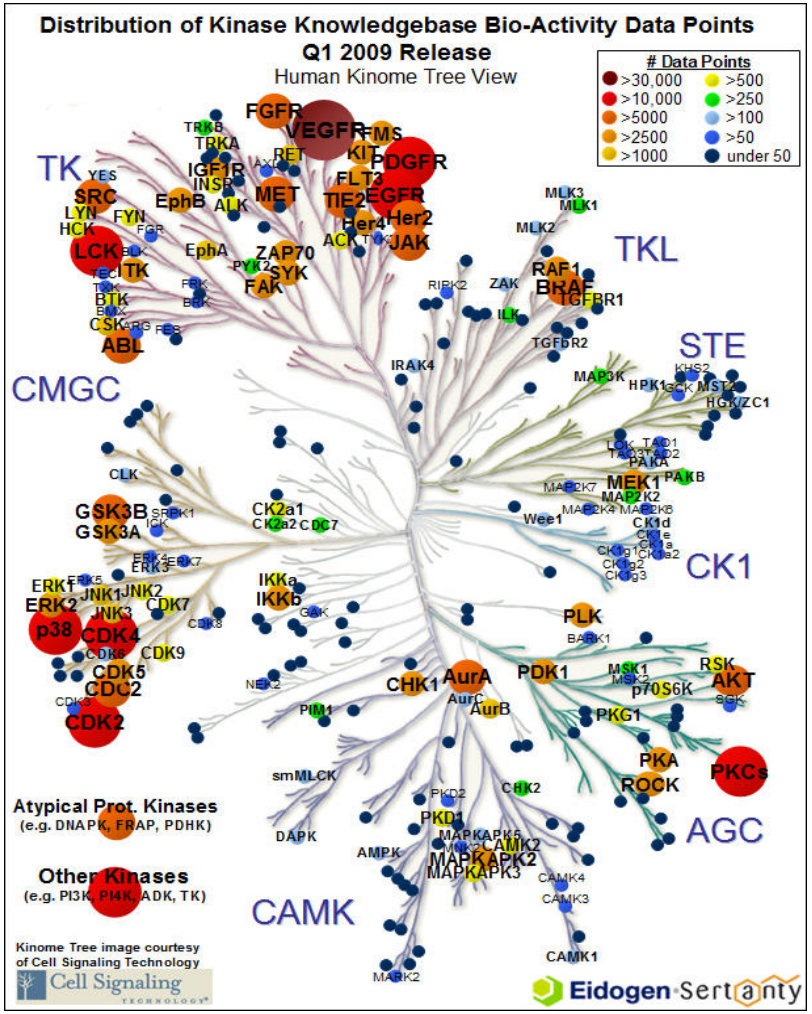
Kinase SAR Knowledgebase (KKB) – Hot Targets

Kinase Targets of Clinical Interest

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



Eidogen-Sertanty KKB
SAR Data Point Distribution



> 402,000 SAR data points curated from
> 5560 journal articles and patents

Eidogen Sertanty

Eidogen-Sertanty Kinase Knowledgebase

Summary Statistics – Q1 2009 Release

Articles covered:	1,616	(+ 51)
Patents and patent applications covered:	3,951	(+ 413)
Total Number of Bio-activity data points:	402,467	(+ 17,594)
Total Number of unique molecules:	486,711	(+ 8,907)
Total Number of unique molecules w/ assay data:	141,718	(+ 8,864)
Total Number of assay protocols:	18,357	(+ 722)

Targets with largest increase in Data Points in Q1-09

Target	# Data Points added
CDK4	3349
PIK3CG	2566
PIK3CA	848
CDK2	622
PIK3CB	596
MET	569
CHEK1	566
AURKA	529
ESR1	480
KDR	450
JAK3	425
CSF1R	422
PLK1	396
CDC2	373
PIK3CD	327
PDGFRB	312
AKT1	301
LCK	292
MAPK14	287
SRC	274
JAK2	250
AURKB	242
IKBKB	213
IGF1R	210
SYK	206
CHUK	195
BRAF	176
CDC7	148
TEK	145
CDK9	136
ROCK1	136
MAPK1	133
GSK3B	127
PDPK1	120
ZAP70	120
KIT	115
ERBB2	90
JAK1	87
ITK	86

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

Science

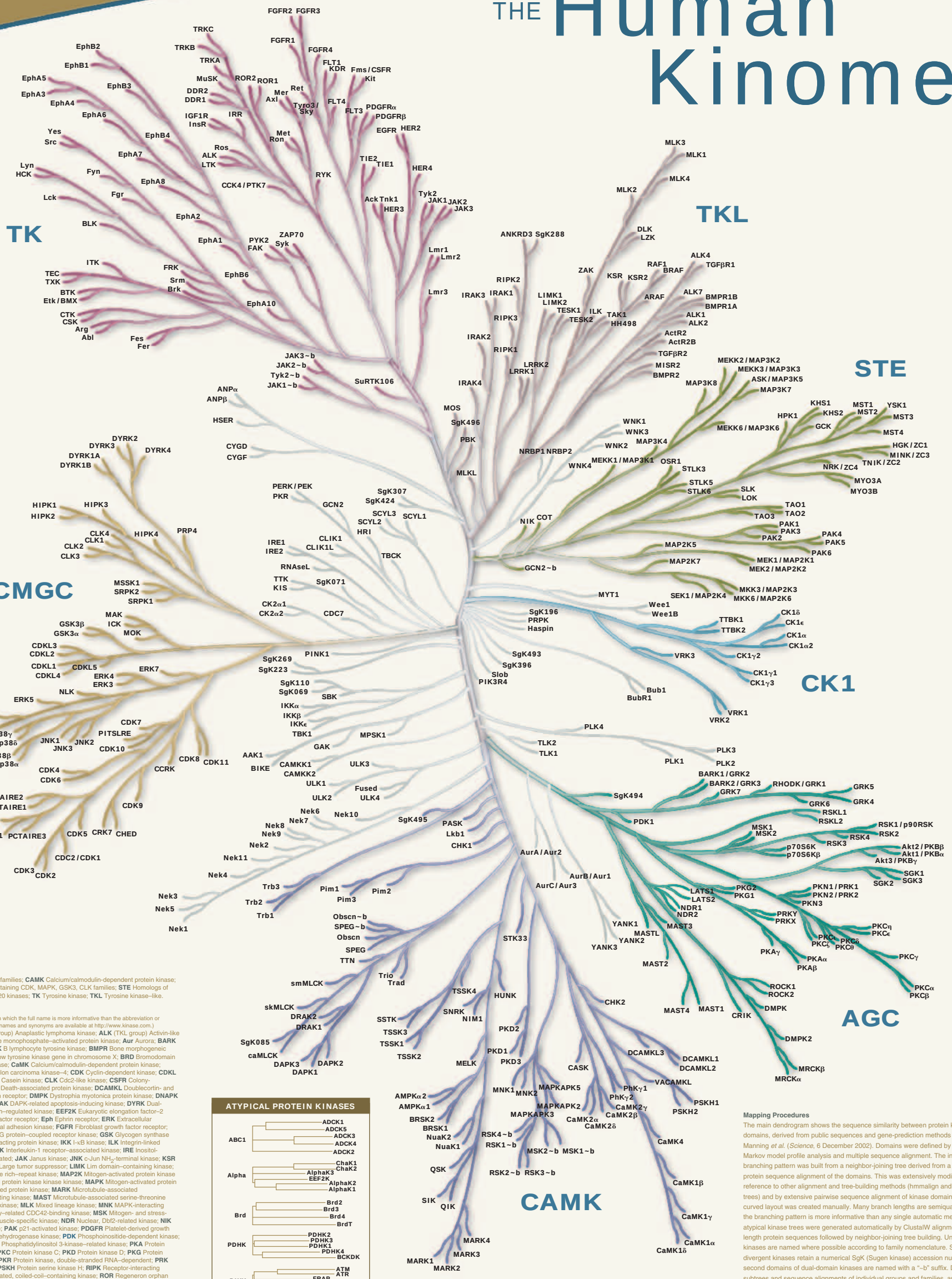
Science's STKE
SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT
WWW.STKE.ORG

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

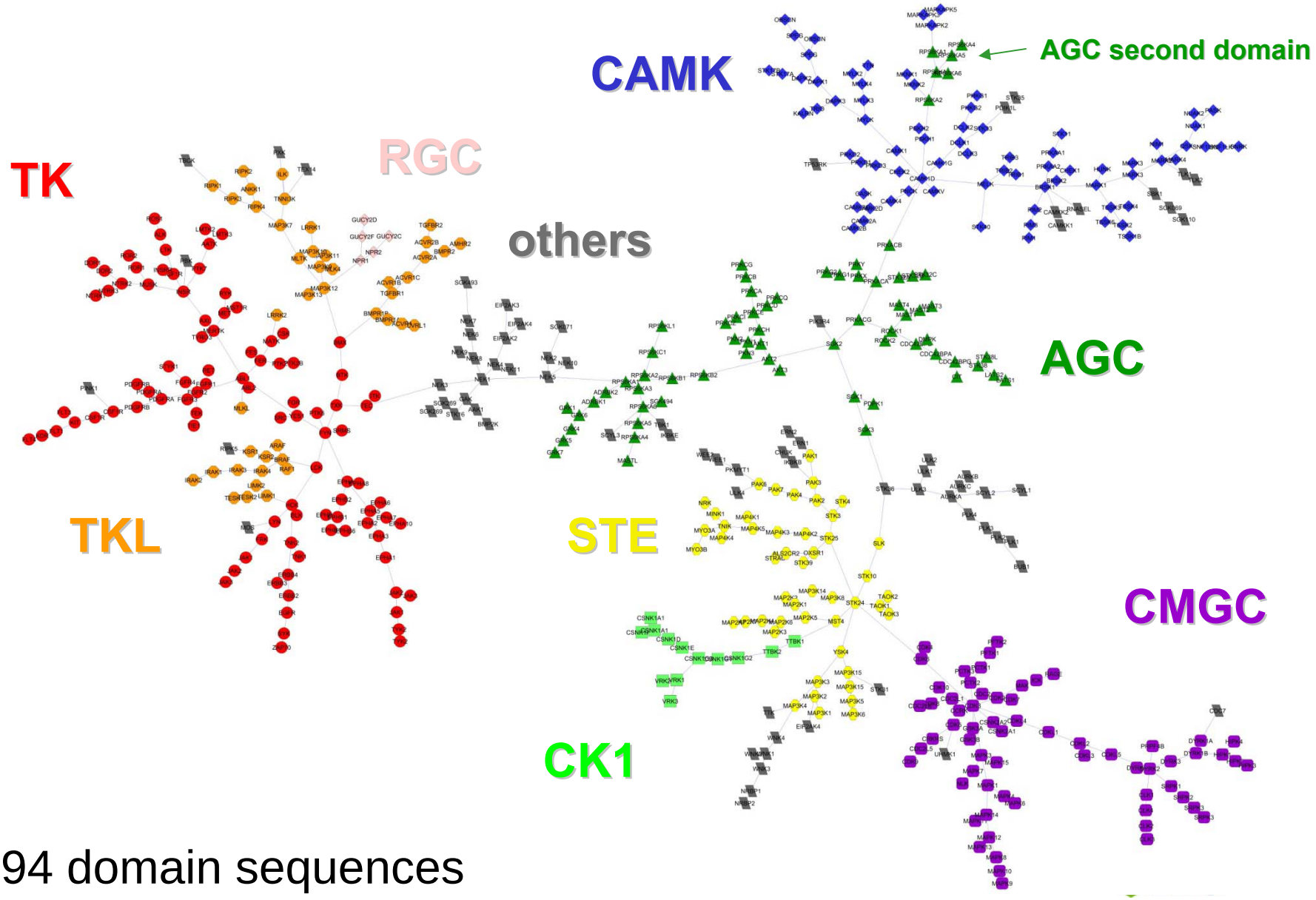
ATYPICAL PROTEIN KINASES

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).

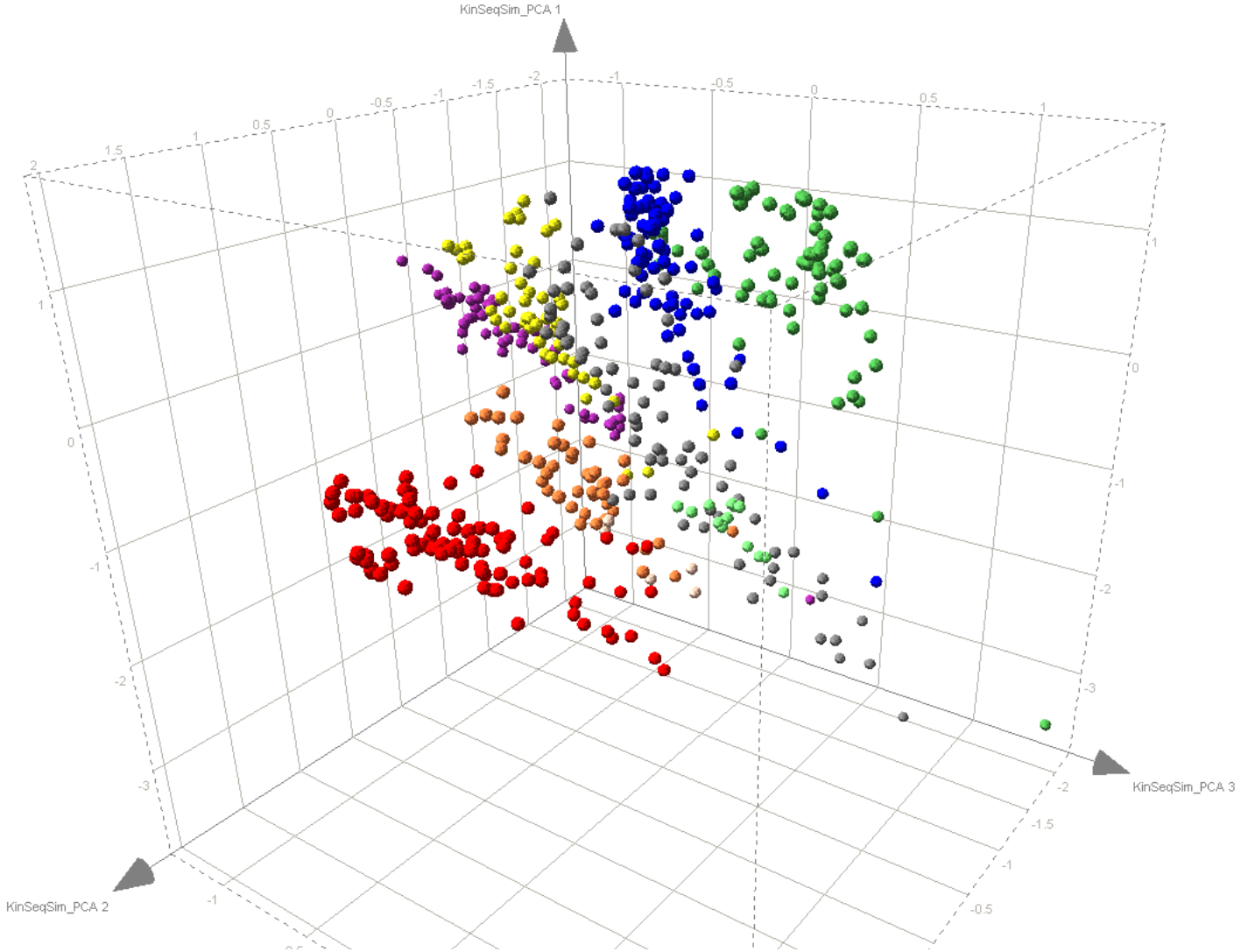
Science coordinator L. Bryan Ray
Design and production Anne Ashley, David Comb, Michael Melnick, C. Faber Smith, J. White, David M. Tompkins, Marcus Spiegler
Copyeditor Harry Jach
Sponsored by Cell Signaling Technology, Inc. and Sugen, Inc.

Kinase Domain Sequence Similarities - MST



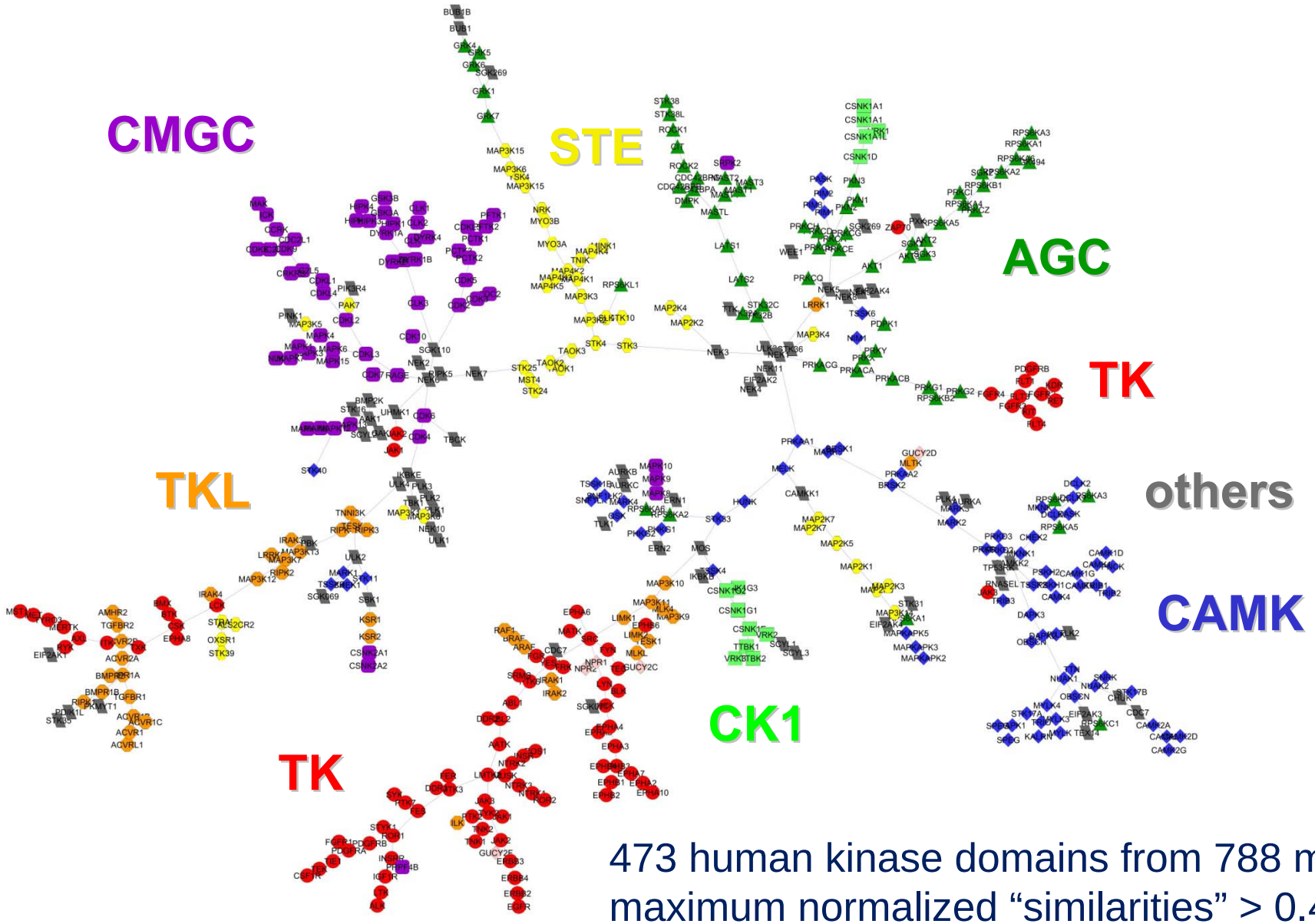
494 domain sequences

PCA View – All Pairwise Similarities



494 domain sequences; 3 PCA dimensions preserve 61 % variability

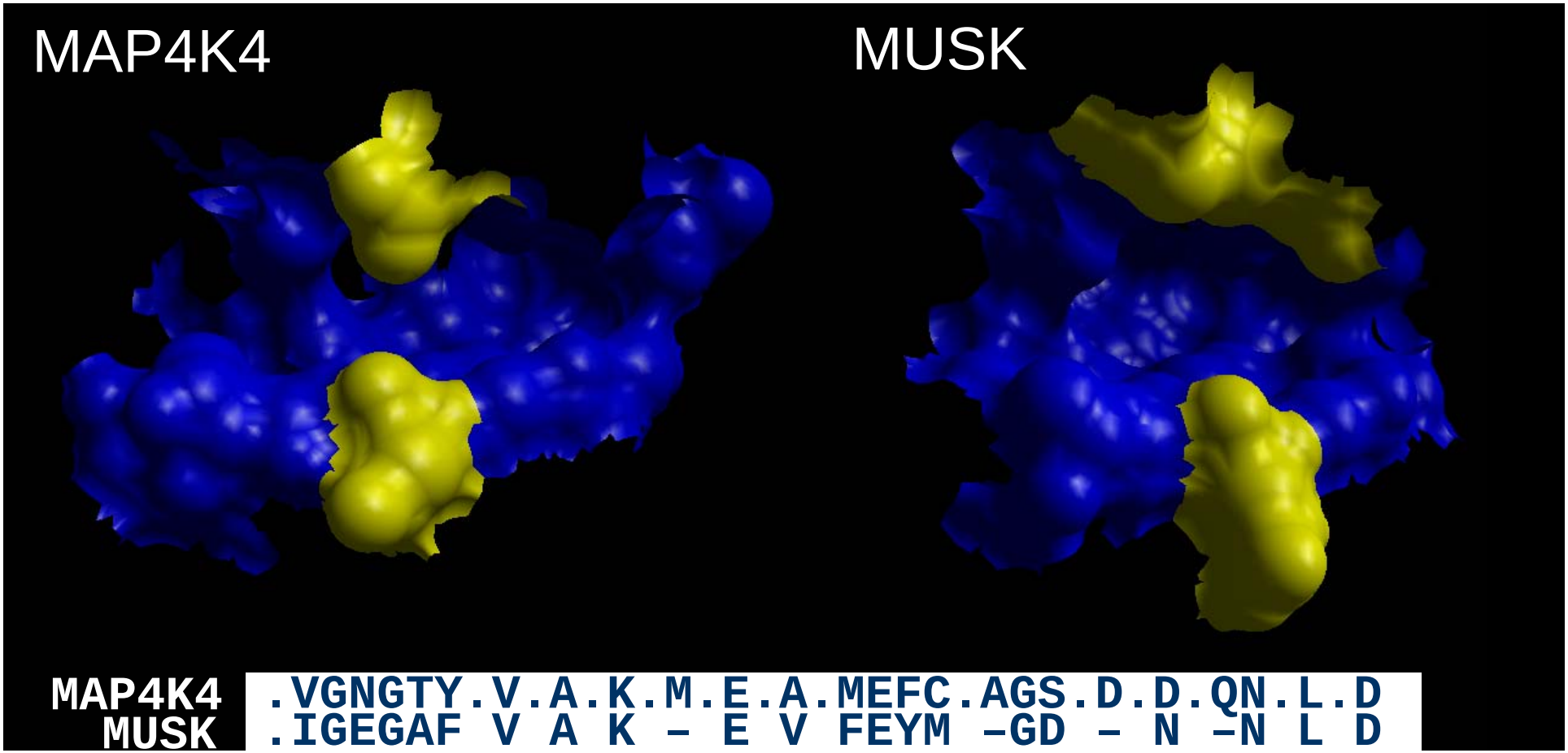
Maximum local site similarity – MST



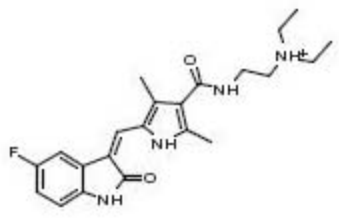
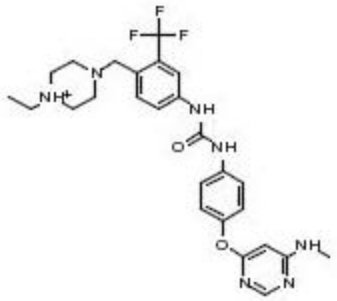
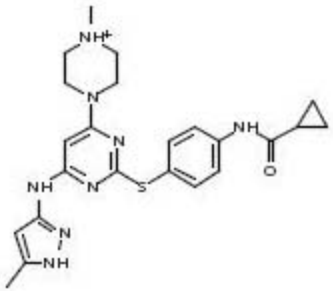
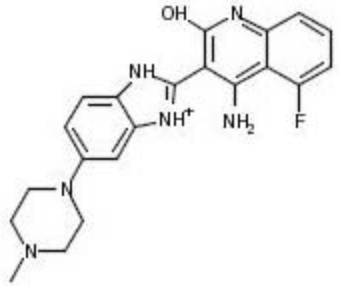
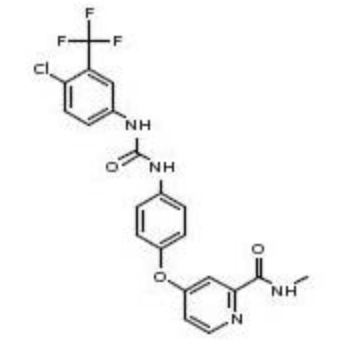
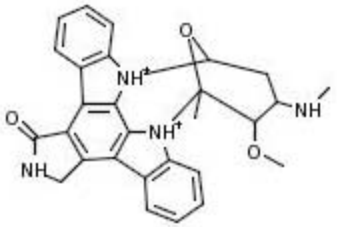
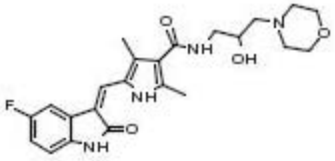
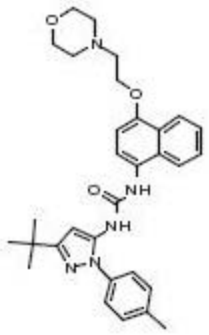
473 human kinase domains from 788 models
maximum normalized “similarities” > 0.45

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



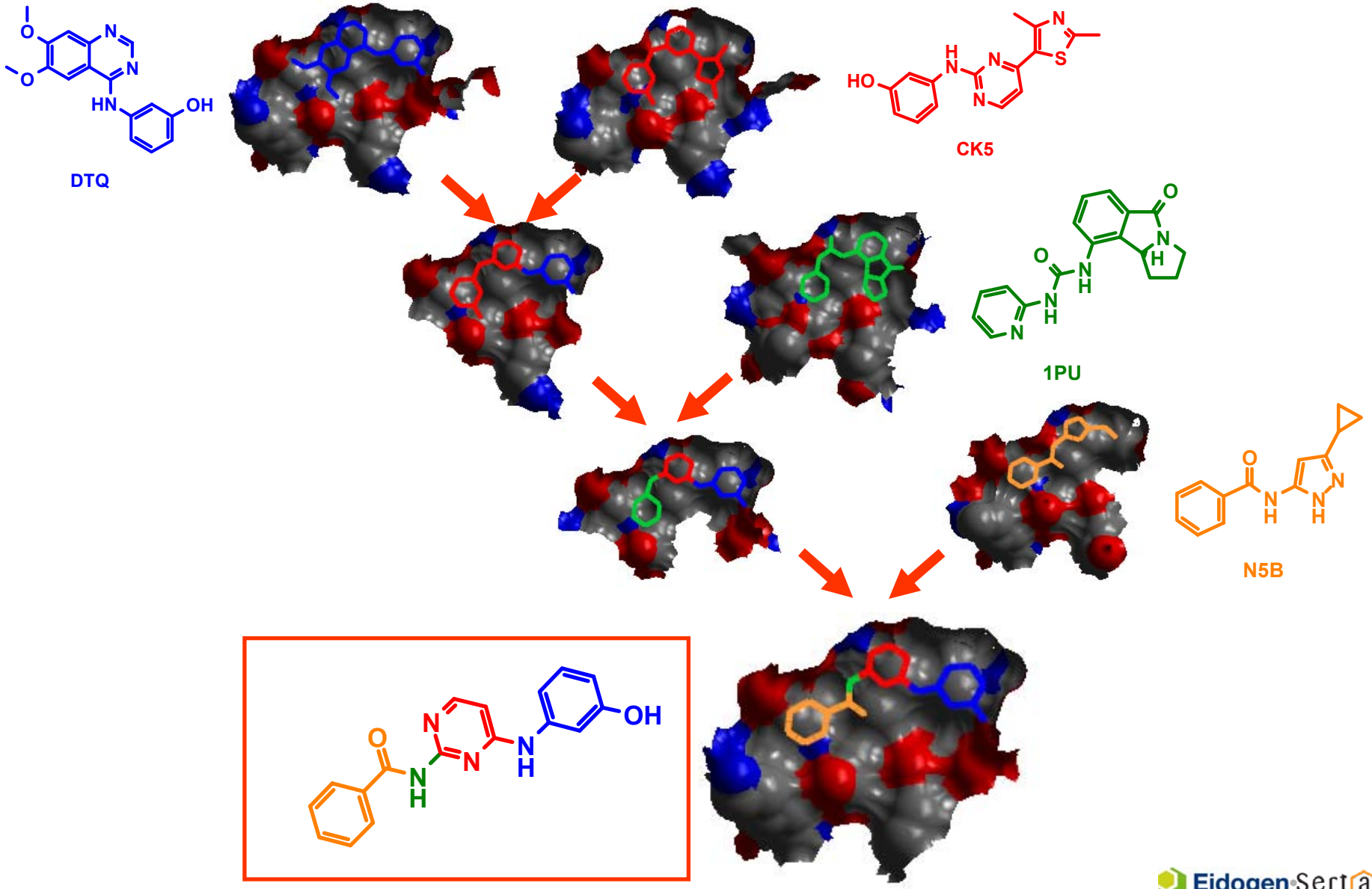
MAP4K4 and MUSK Small Molecule Inhibitors

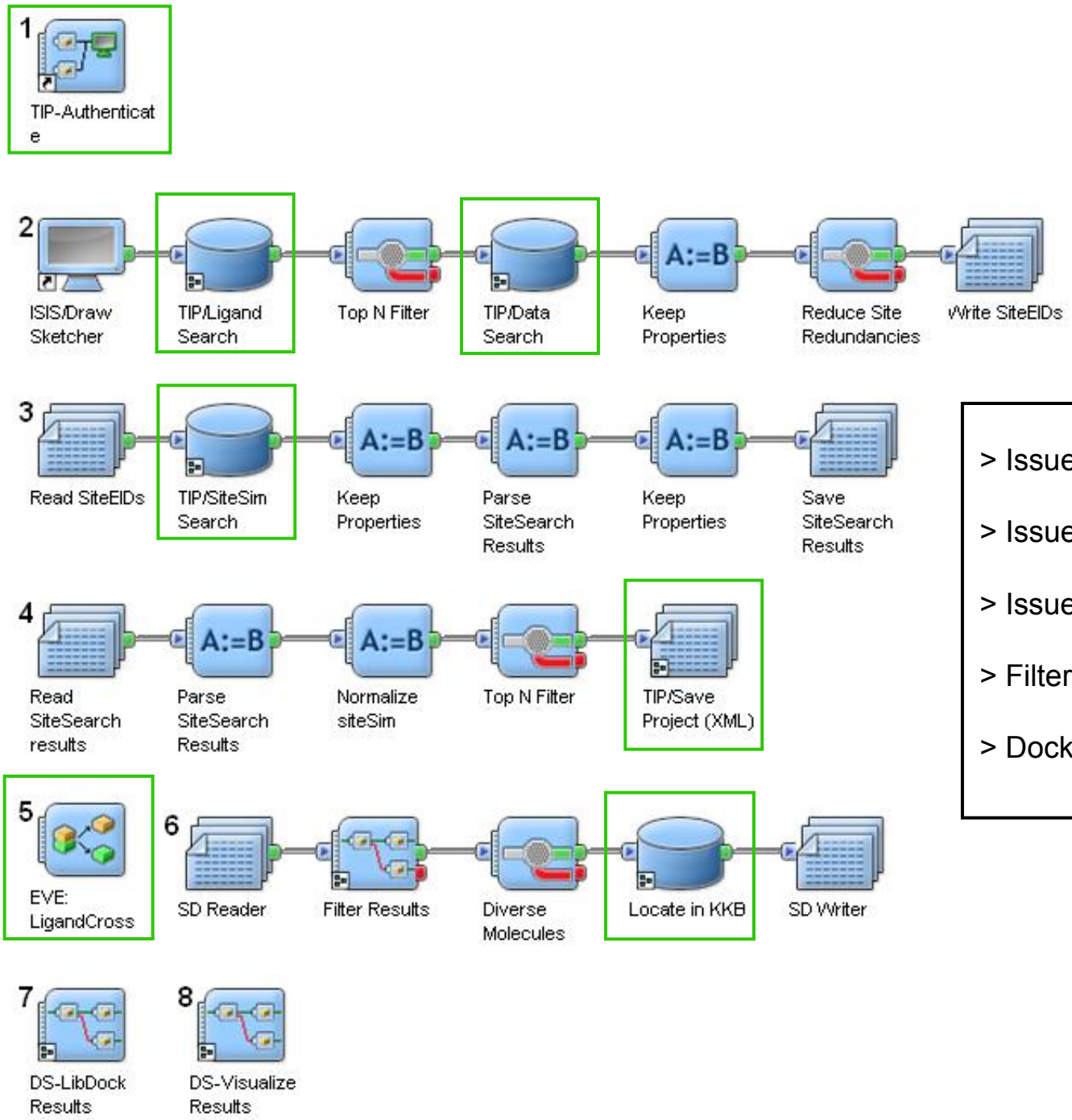
 MAP4K4 / 6.9 MUSK / 6.3	 MAP4K4 / 7.8 MUSK / 8.5	 MAP4K4 / 5.6 MUSK / 6.5	 MAP4K4 / 6.8 MUSK / 6.1
 MAP4K4 / 5.3 MUSK / 6.9	 MAP4K4 / 7.6 MUSK / 7.1	 MAP4K4 / 6.6 MUSK / 7.0	 MAP4K4 / 7.0 MUSK / 5.7

Eidogen-Sertanty

LigandCross: Shuffling Ligand Functionality

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





> Issue TIP/LigandSearch

> Issue TIP/SiteSimSearch

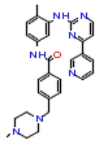
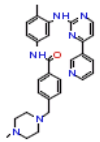
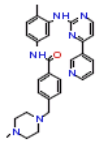
> Issue LigandCross

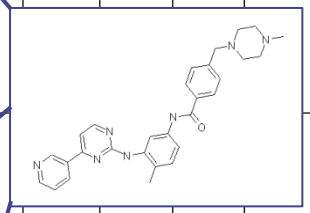
> Filter and locate results in KKB

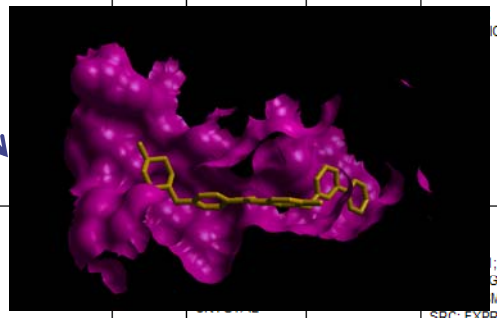
> Dock and visualize results

Eidogen Sertanty

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

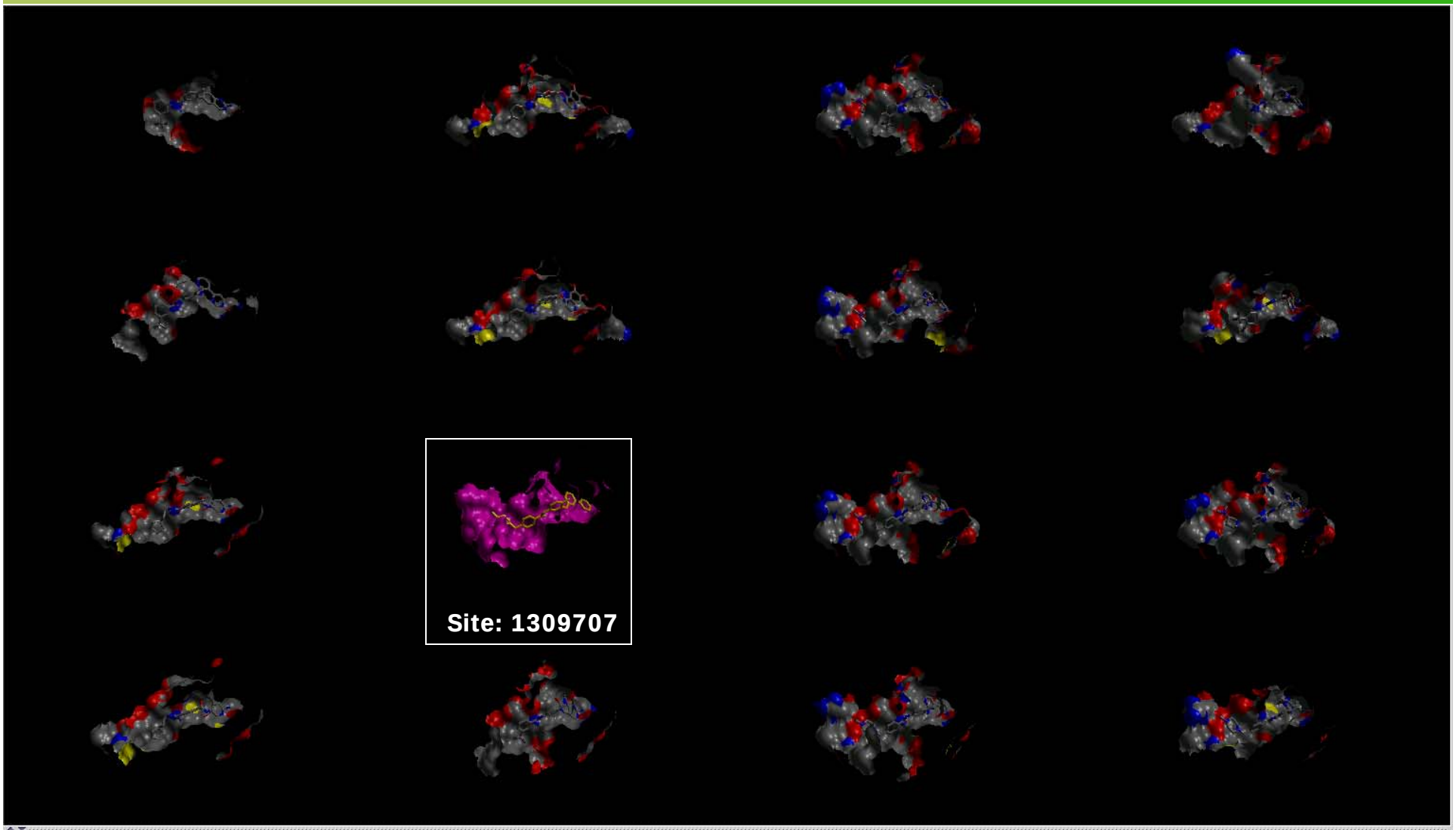
Molecule	ligname	similarity	pdbcode	siteeid	FourCode	pdbID	pdbBnxNumber	proteinId	title	classification	source	compound	releaseDate	journalTitle	journalReference	exptype
	STI	1	2pl0A	1309707	2pl0	2pl0	1305799	42526	LCK BOUND TO IMATINIB	TRANSFERASE	MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: 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

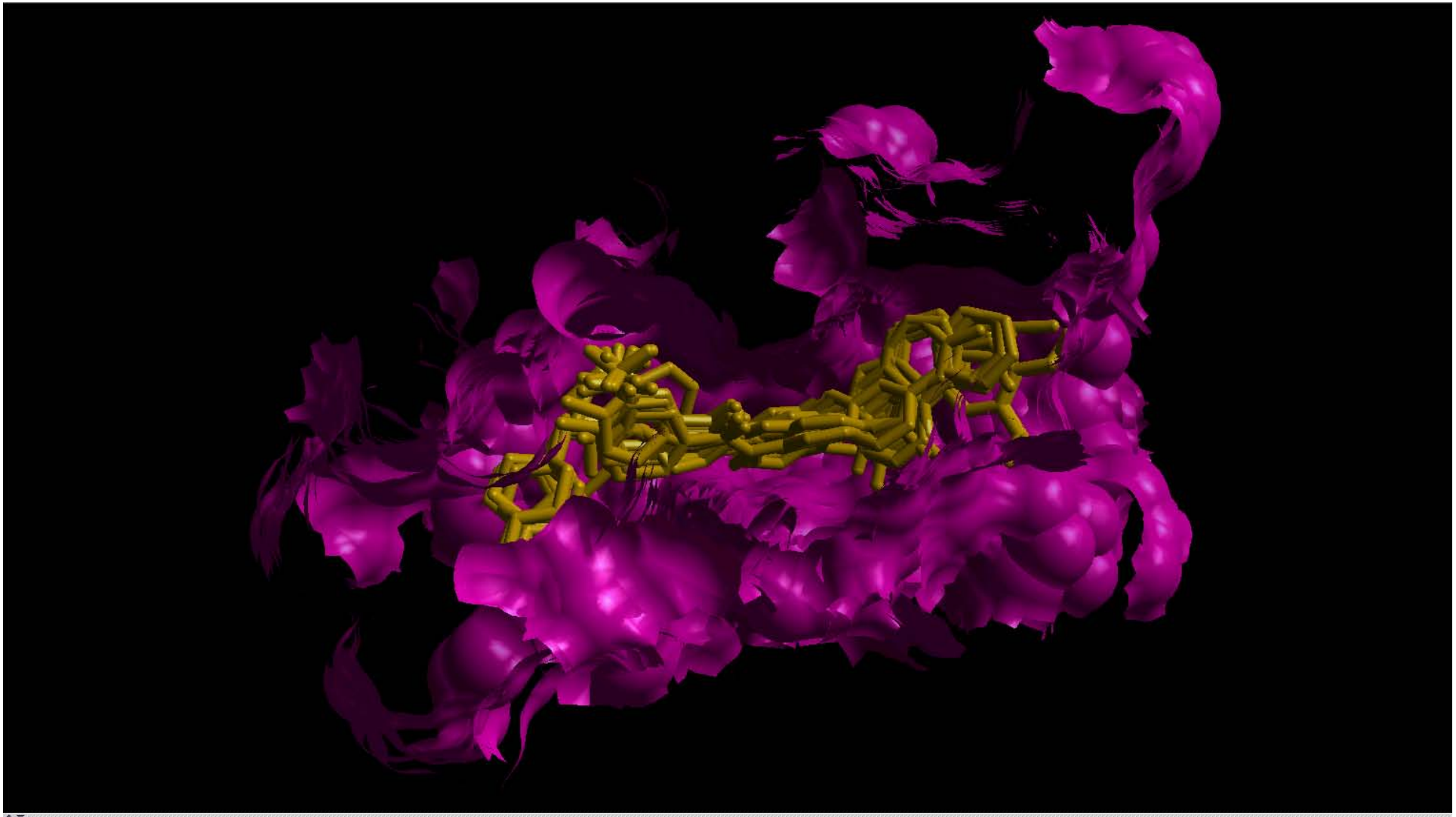
Step 2: Find Other Receptor Sites from Site-Similarity Search



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

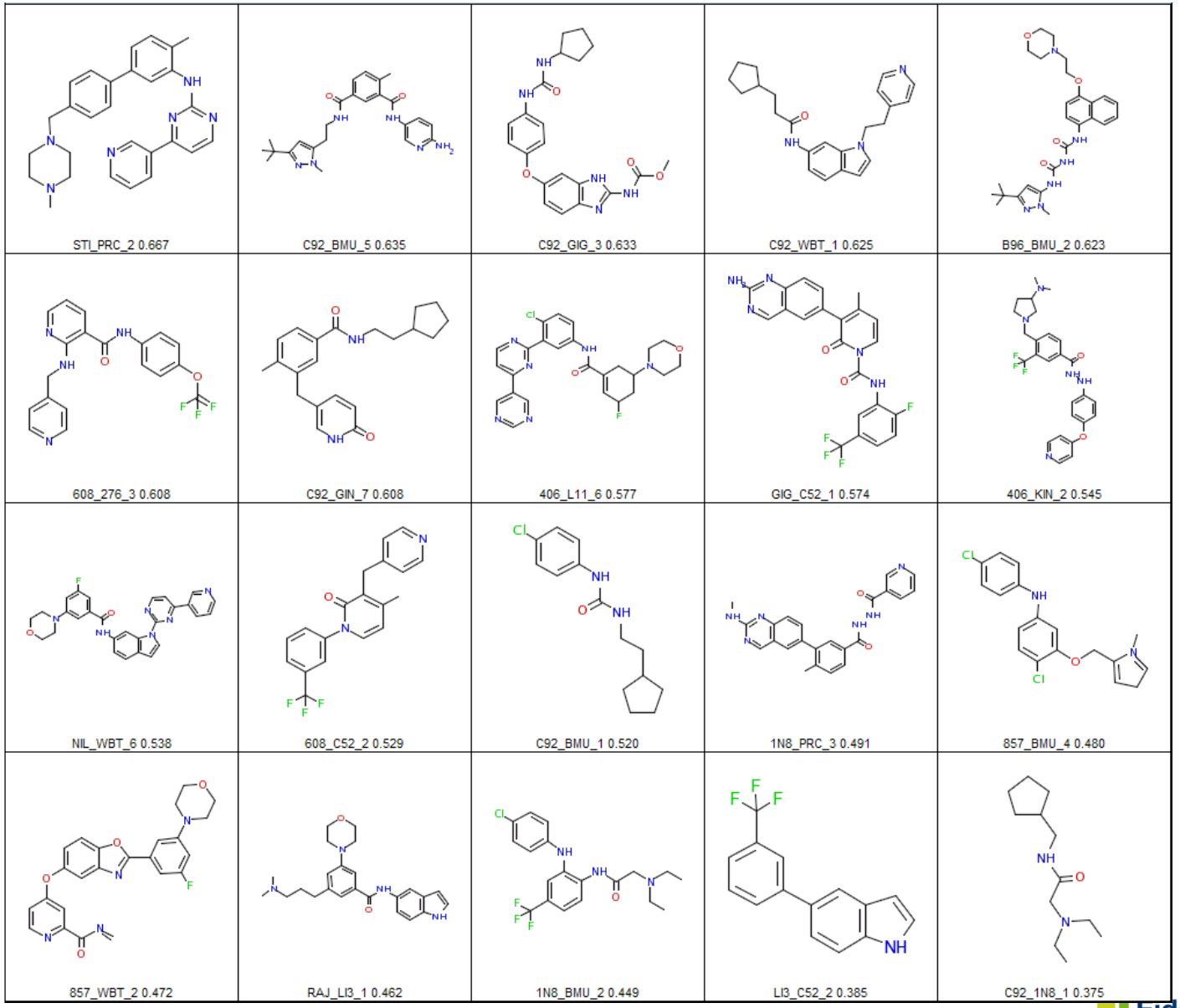
Step 3: LigandCross – Mixing Ligand Features from Aligned Sites



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

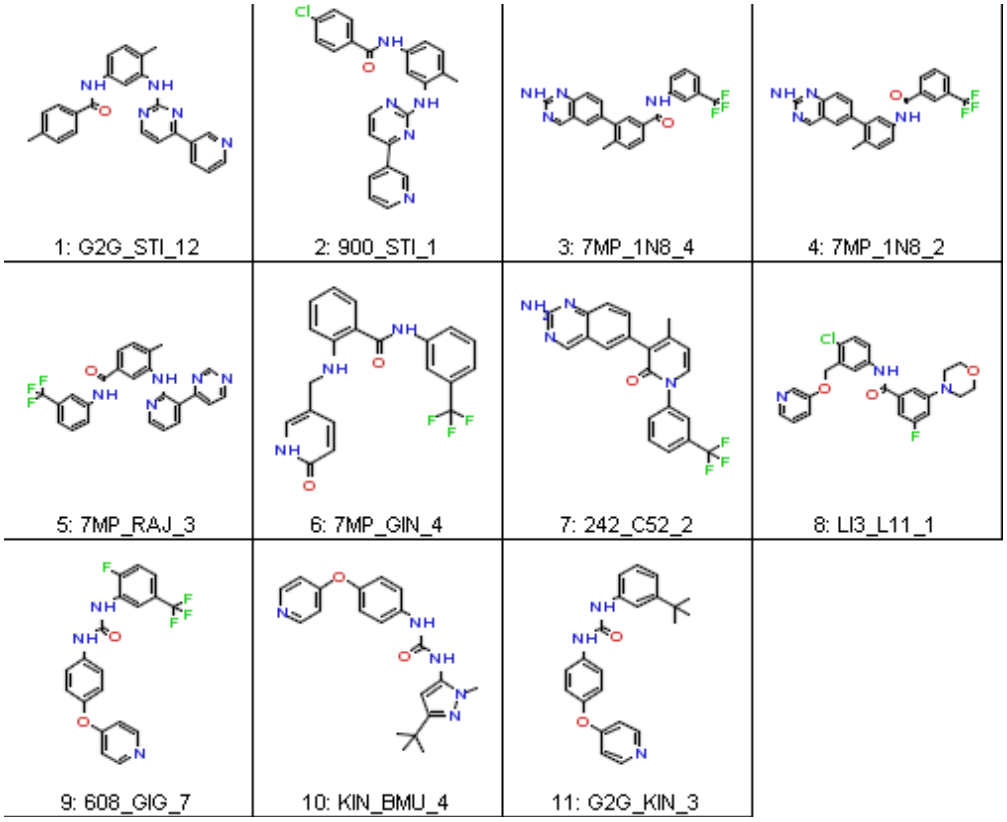
Example LigandCross Results



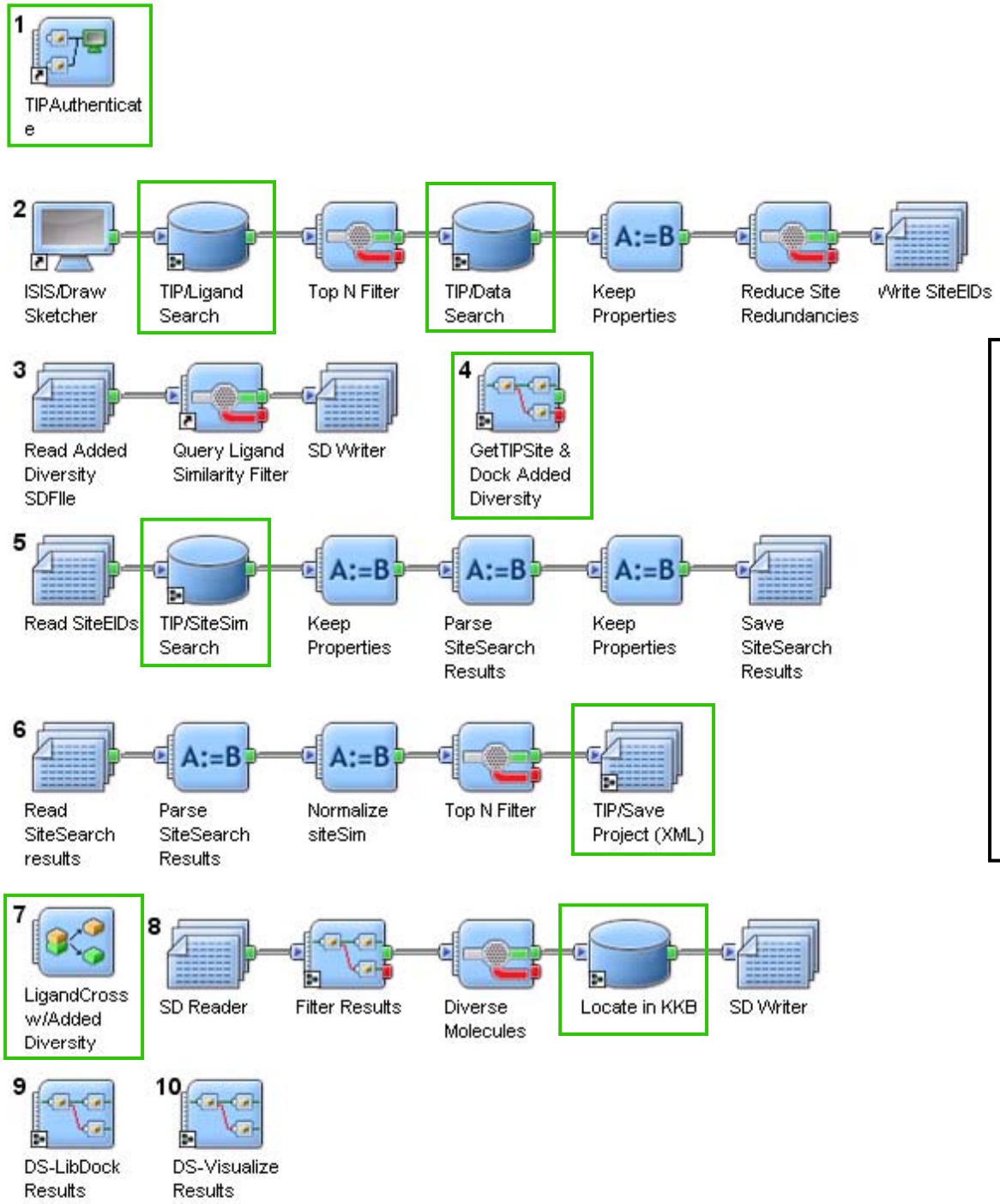
Eidogen-Sertanty

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”

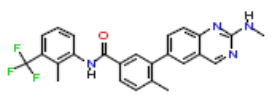
> Issue TIP/SiteSimSearch

> LigandCross w/AddedDiversity

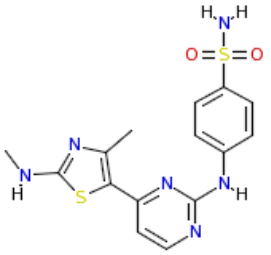
> Filter and locate results in KKB

> Dock and visualize results

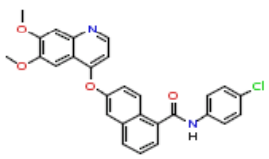
Example Potent Kinase Inhibitors ("Added Diversity")



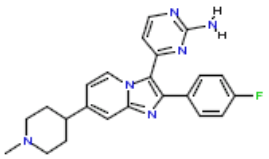
4336533 LCK pval: 11.00



4302493 CDK9 pval: 10.54

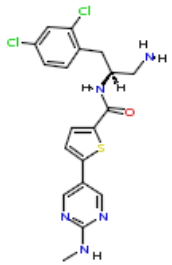


4332561 KDR pval: 10.52

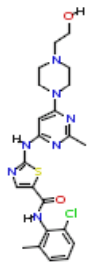


4318145 PKG pval: 10.40

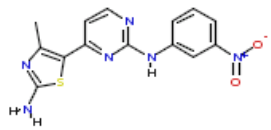
Chiral



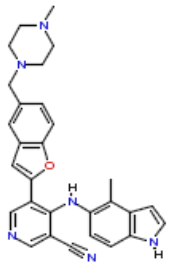
4336686 PKA pval: 10.00



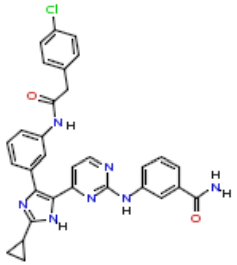
4272835 ABL1 pval: 10.00



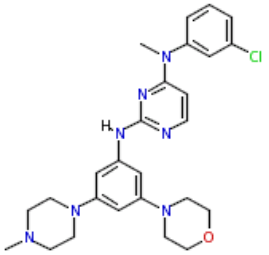
894611 CDK2 pval: 9.70



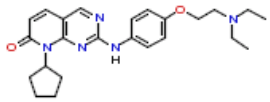
4358565 PRKCQ pval: 9.70



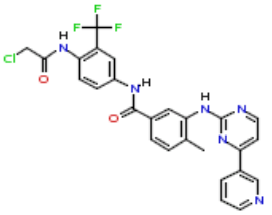
4363734 RAF1 pval: 9.30



4369892 EPHB4 pval: 9.24

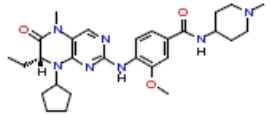


809 CDK4 pval: 9.15

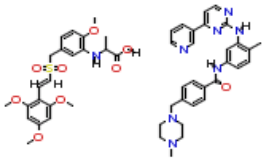


4374385 PDGFRA pval: 9.14

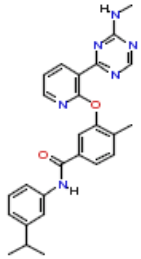
Chiral



4366691 PLK1 pval: 9.10

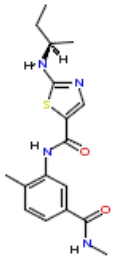


4301886 BCR_ABL pval: 9.08

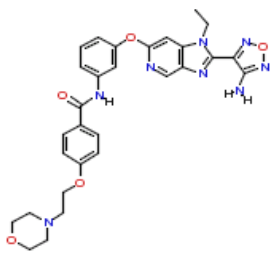


4307551 TEK pval: 9.00

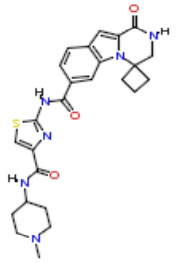
Chiral



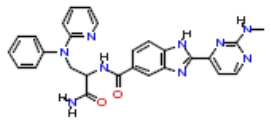
4363016 MAPK11 pval: 8.82



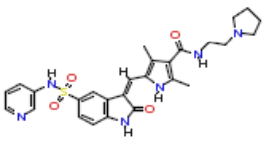
4343448 ROCK1 pval: 8.74



4363247 MAPKAPK2 pval: 8.70

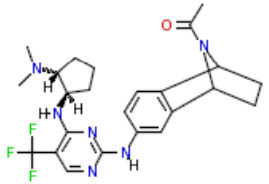


4291996 IKB pval: 8.70



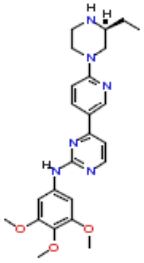
4208857 FAK2 pval: 8.22

Chiral

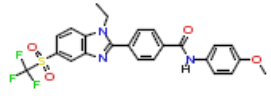


4373725 PTK2B pval: 8.22

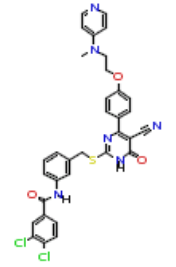
Chiral



1788 ZAP70 pval: 8.10

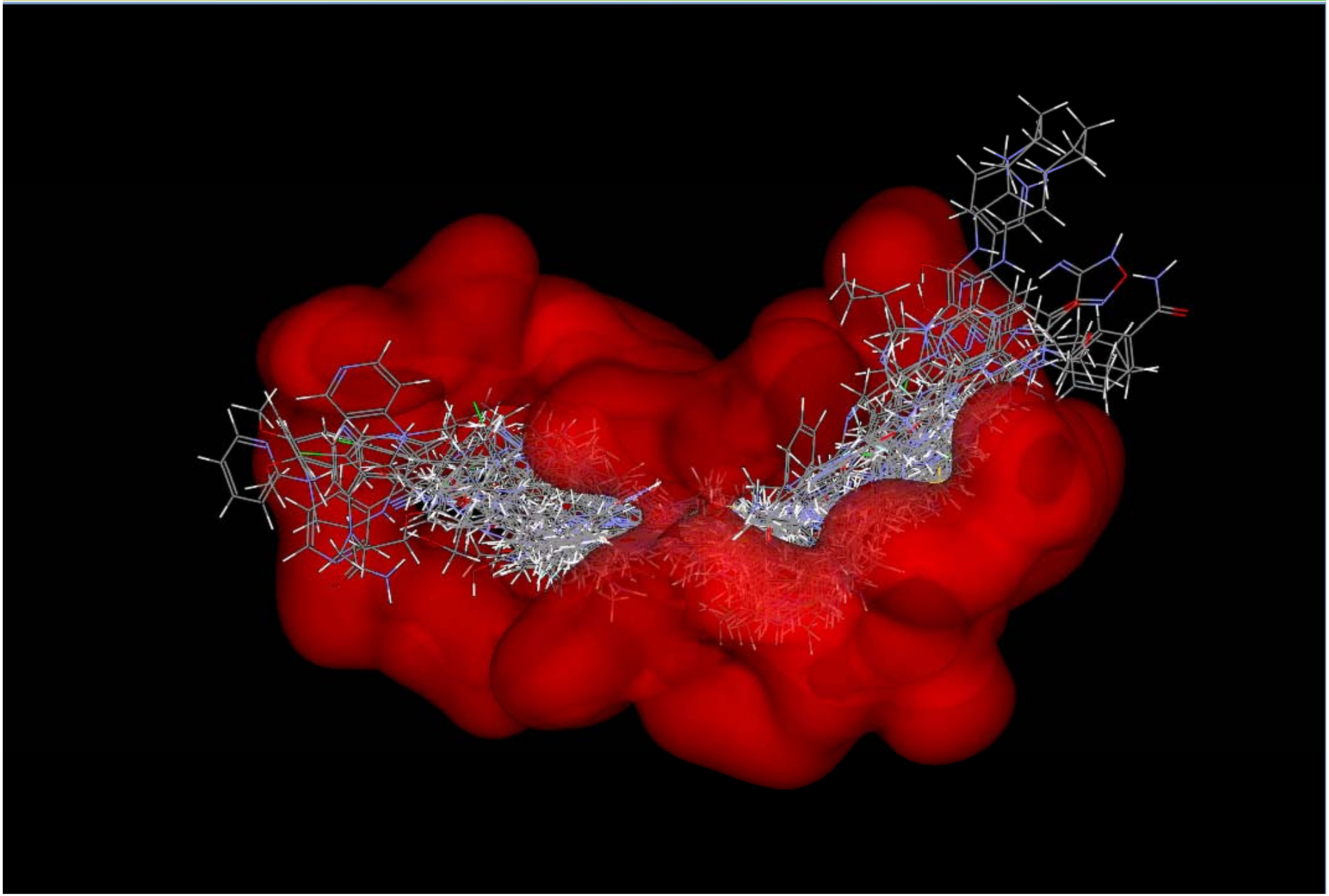


2425813 PTPN9 pval: 5.96

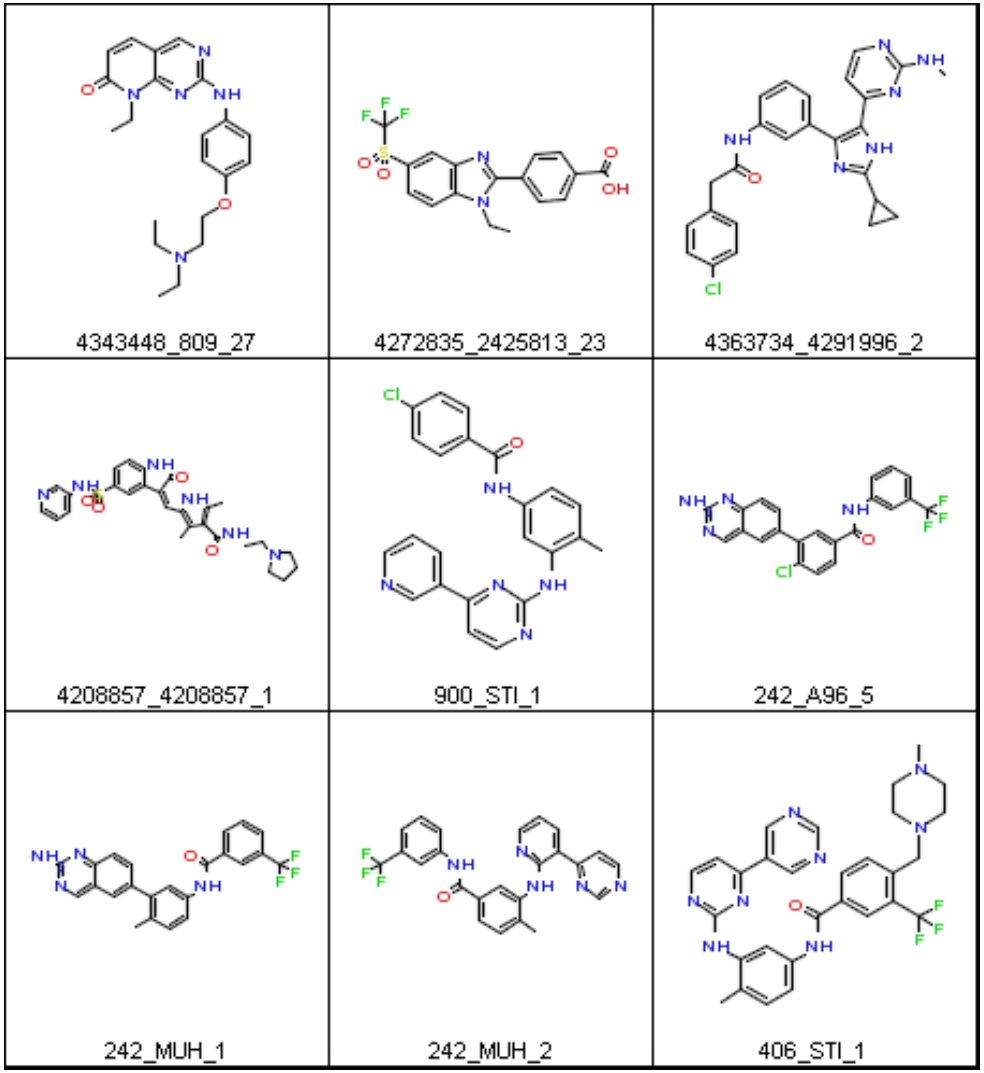


4303129 MAP3K2 pval: 4.70

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

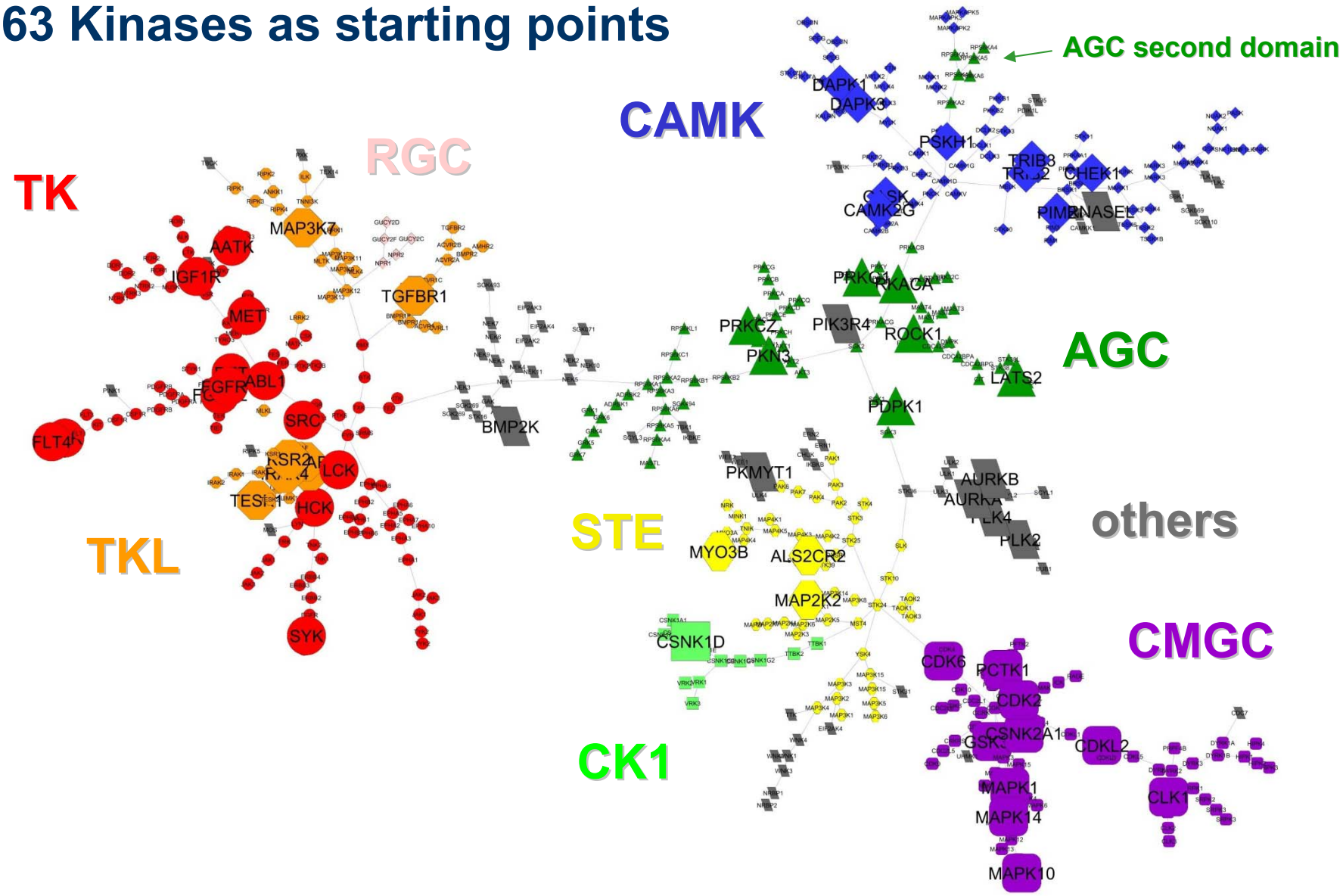
242_MUH_1:
LCK: 9.40 TEK: 7.68 KDR: 8.22 MAPK14: 9.00 JAK3: 6.81

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

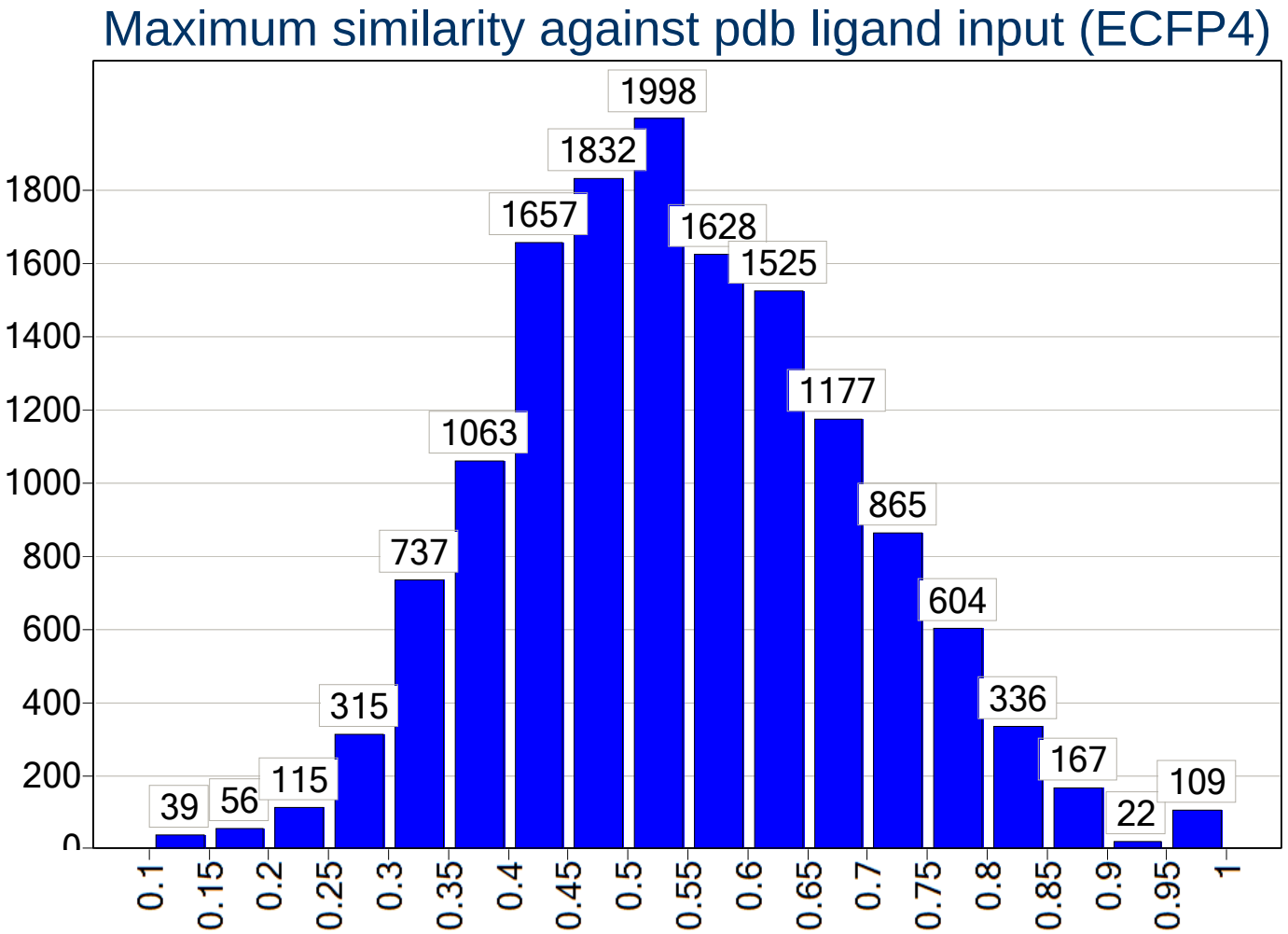
Ligand Functionality Shuffling across the Kinome

63 Kinases as starting points



Generates novel compounds

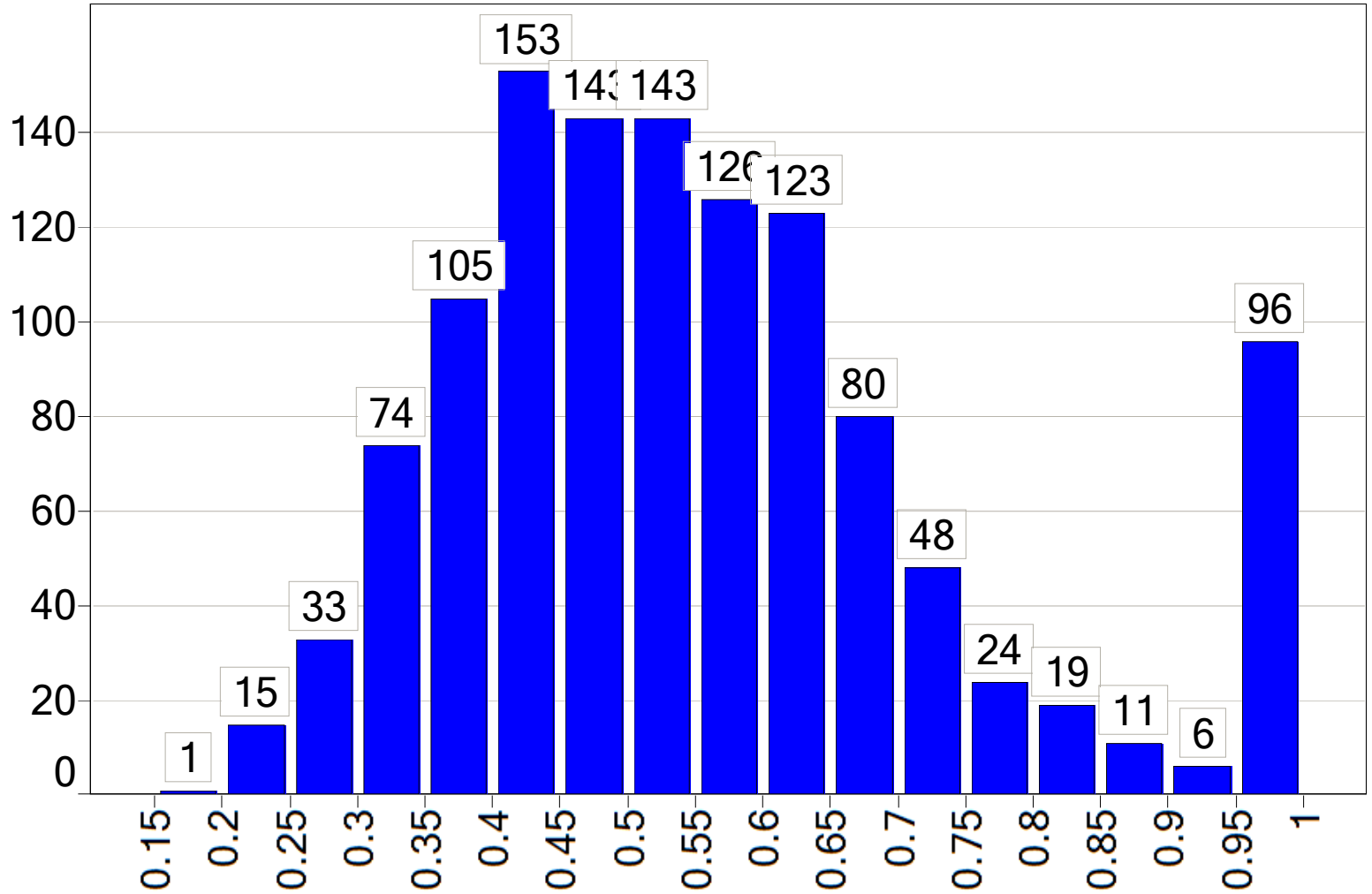
- From 280 Unique Kinase co-Crystal ligands
- > 14,000 new unique structures are generated (HTS filter)



Drug- / lead-like novel compounds

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

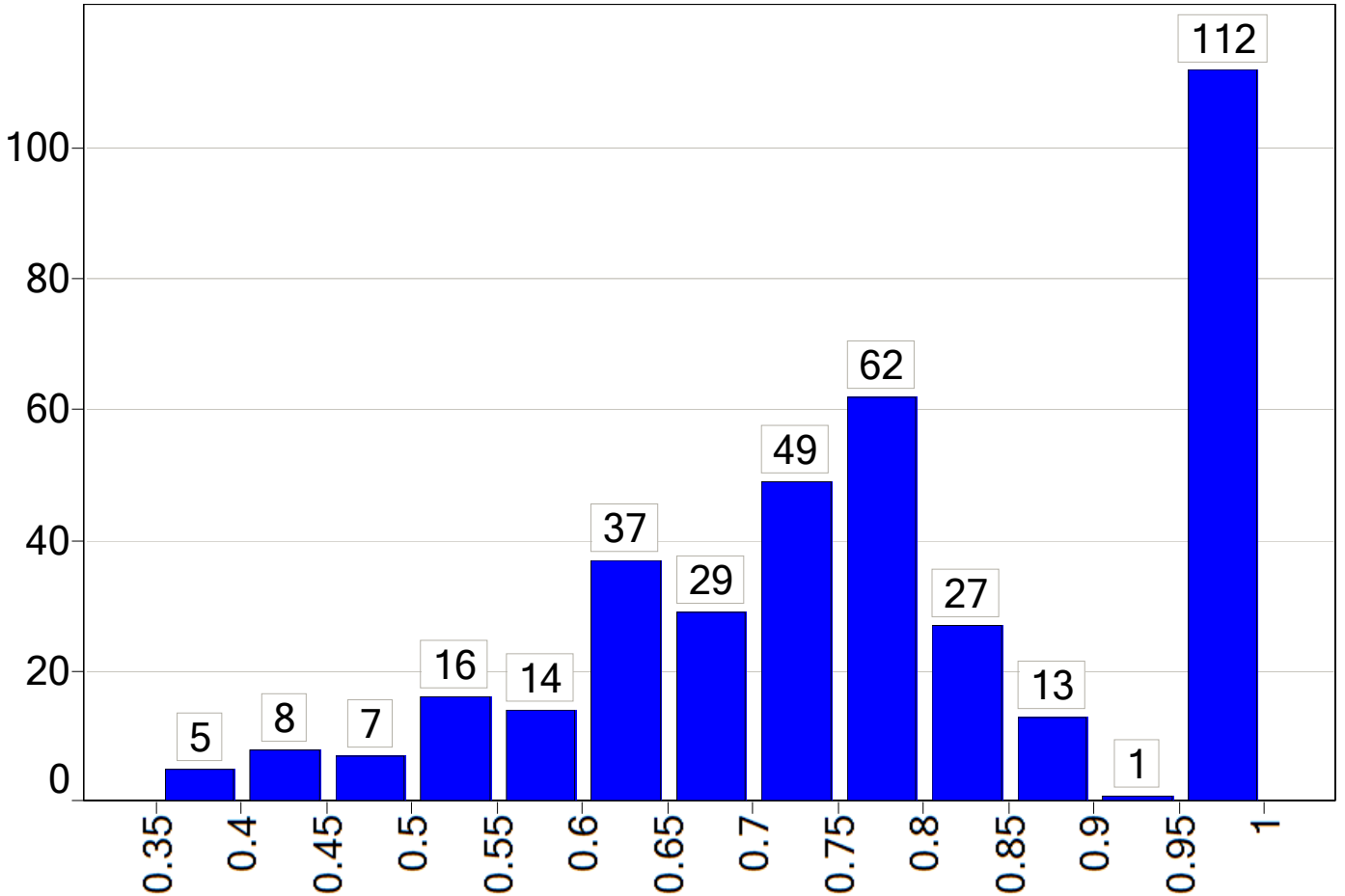
➢2,153 unique compounds (64 pdb ligands pass the same filters)



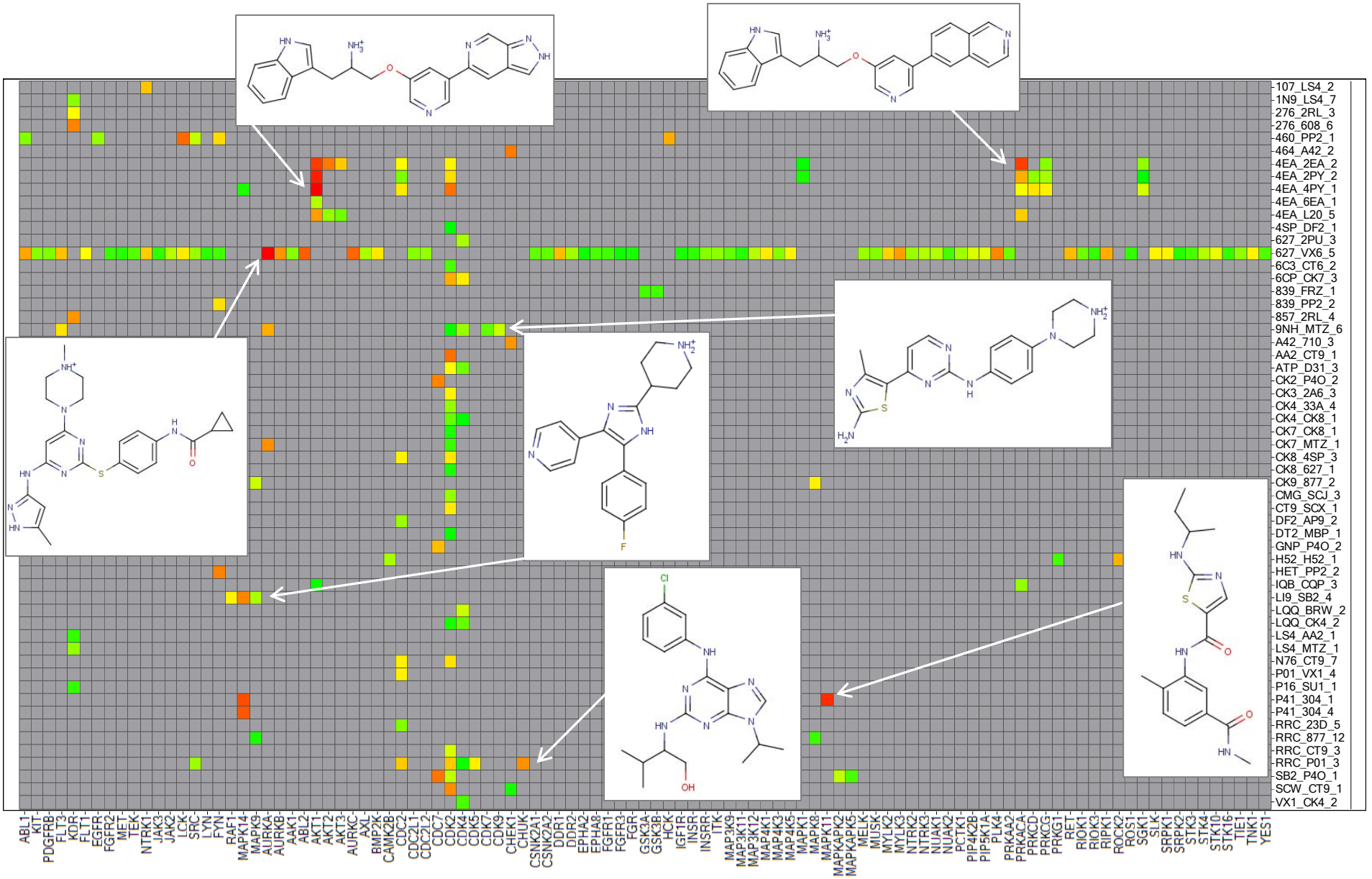
Novel active compounds

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

Maximum similarity against pdb ligand input (ECFP4)



Known drug/lead –like actives (less than 1 μM)



Conclusions

- Systematic modeling and analysis of both small molecule activity data and protein structure site similarities can reveal pharmacologically relevant insights and predict possible cross reactivity within (and across) target families
- Systematic analysis of protein site similarities is in many cases consistent with existing experimental SAR
- The structurally resolved and modelable proteome is a very rich source for new matter ideas
- LigandCross can be an effective strategy to generate novel, bioactive molecules from co-complex information
- There is synergy between protein structure information and small molecule SAR data

Acknowledgements

- Stephan Schürer
- Kevin Hambly
- Joe Danzer
- Brian Palmer
- Derek Debe
- Aleksandar Poleksic

- Accelrys/Scitegic - Shikha Varma-O'Brien/Ton van Daelen

Add’l slides

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.

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

Drugs Developed using Structural Knowledge

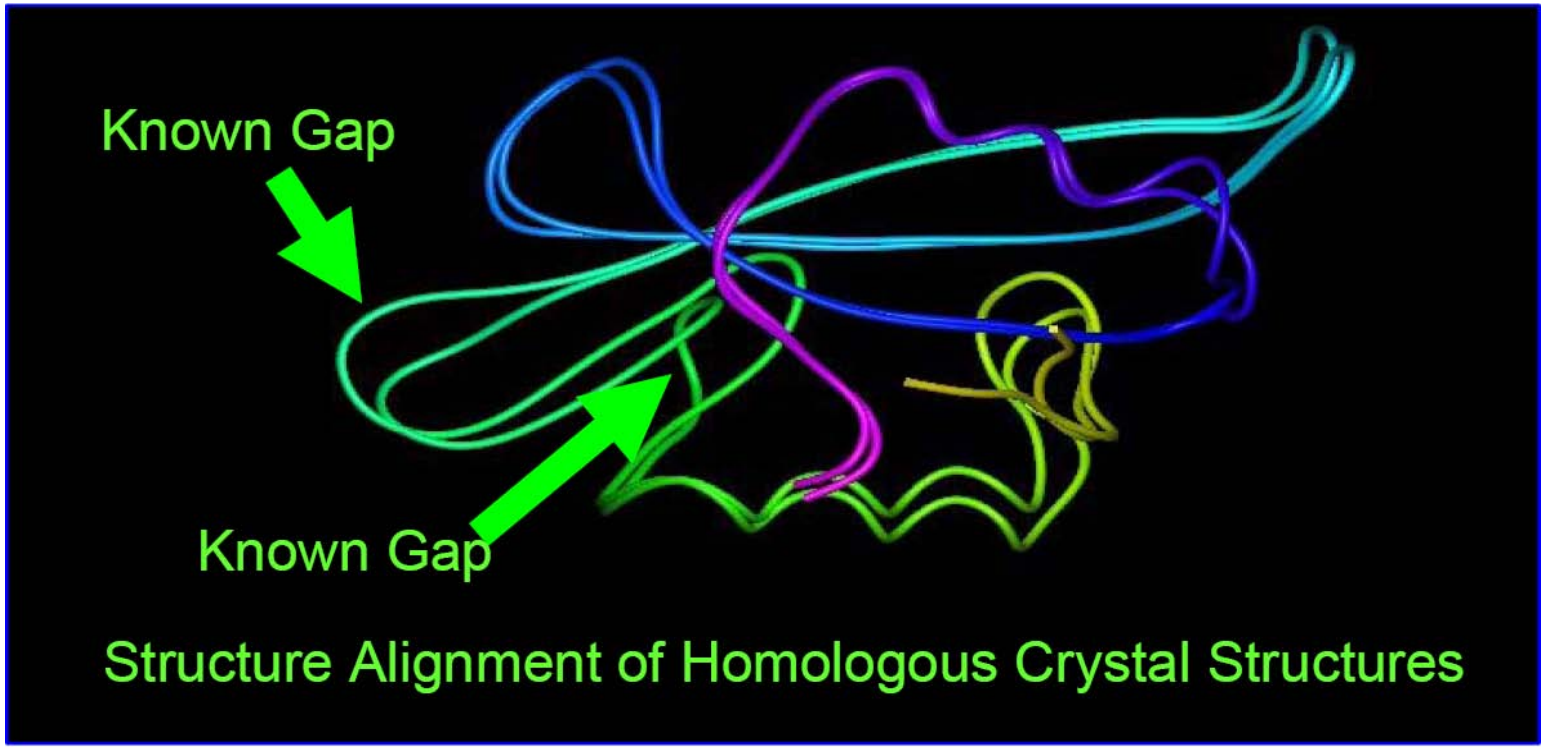
Inhibitor/Drug	Disease	Company(s)	Protein targeted	Enzyme Family
STI-571/Gleevec	Chronic Myeloid Leukemia	Novartis	c-Abl kinase	Tyrosine kinase
Fluoroquinolone/Ciprofloxacin	Bacterial infection	Bayer	Gyrase	ATP Hydrolase
Saquinavir/Invirase, Ritonavir/Norvir, Indinavir/ Crixivan, Nelfinavir/Viracept, 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

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

TIP (Eidogen-Sertanty) / Debe et. al. Proteins 2006, 960

Eidogen•Sertanty

STRUCTFAST™ Algorithm Comparison

Alignment	Scoring Methods	Gap Treatment	Examples
Sequence-Sequence	BLOSUM PAM GONET	Length Proportional Affine	BLAST FASTA Smith-Waterman Needleman-Wunsch
Sequence-Profile	PSSM HMM	Affine Position-Specific	PSI-Blast HMMer
Sequence-Structure	Threading potential	Affine Position-Specific	Raptor GenThreader
Profile-Profile	Dot-product Log Average **Analytic Statistics**	Position-Specific **Structural Family-based**	3D-PSSM FFAS **STRUCTFAST**

STRUCTFAST™ CASP6 Results

December 2004 CASP6 Total Comparative Modeling Results

of models placed in the top 20 according to the number of correctly aligned residues

Group Name (Servers in Red)	# of Models in the Top 20
KOLINKSI-BUJNICKI	79
Jones-UCL	69
GeneSilico-Group	60
STRUCTFAST	54
BAKER	53
Ginalski	51
TOME	51
Skolnick-Zhang	50
CBRC-3D	38
FISCHER	37
CHIMERA	34
SAM-T04-hand	29
SBC	28
Sternberg	27
CAFASP-Consensus	26
zhousp3	23
ZHOUSPARKS2	23
ACE	23
SBC-Pmodeller5	19

Other Notables:

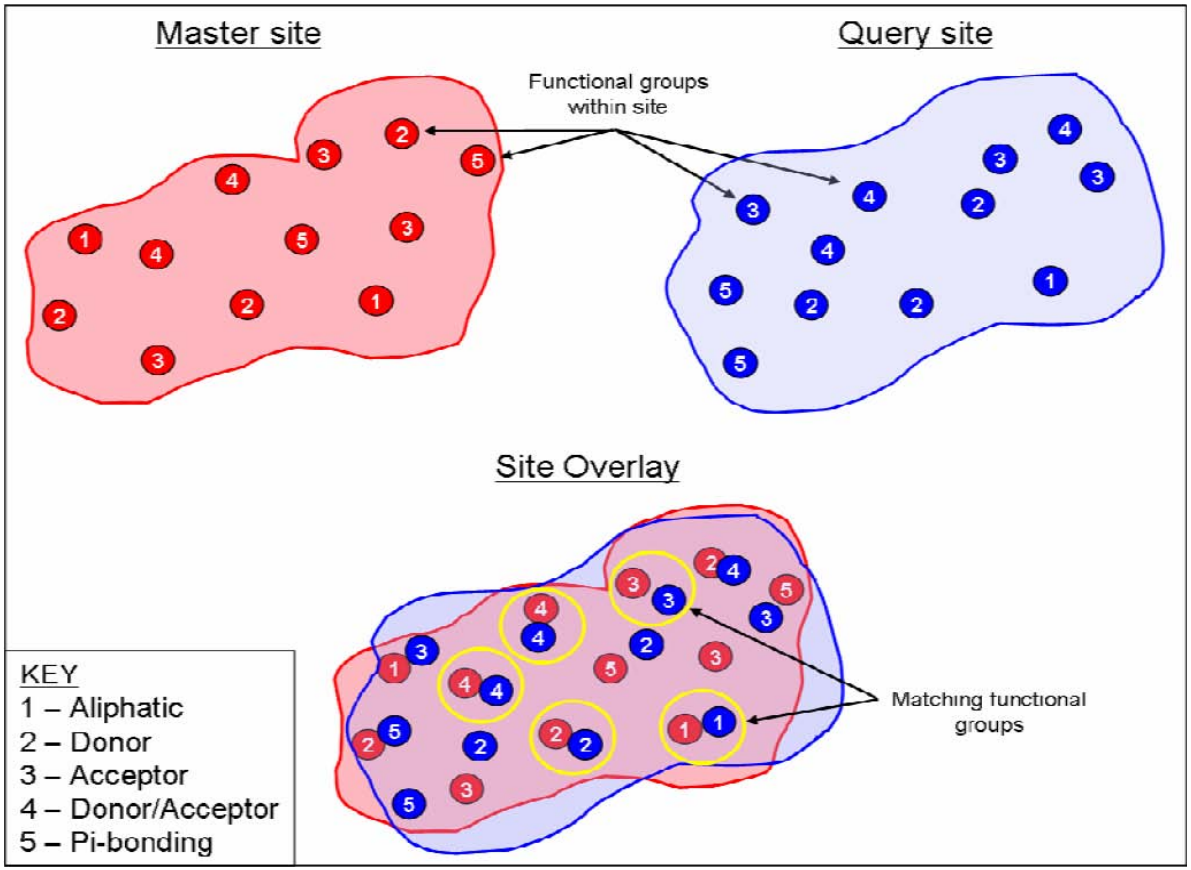
FAMS	15
Accelrys	4

STRUCTFAST had more than twice as many models in the top 20 compared to the second best automated server.

Only 3 of 124 hand modeling teams produced better alignments than ***STRUCTFAST***.

SiteSorter™ binding site comparison

Weighted Clique Detection Algorithm (importance of points related to conservation in multiple sequence alignment)



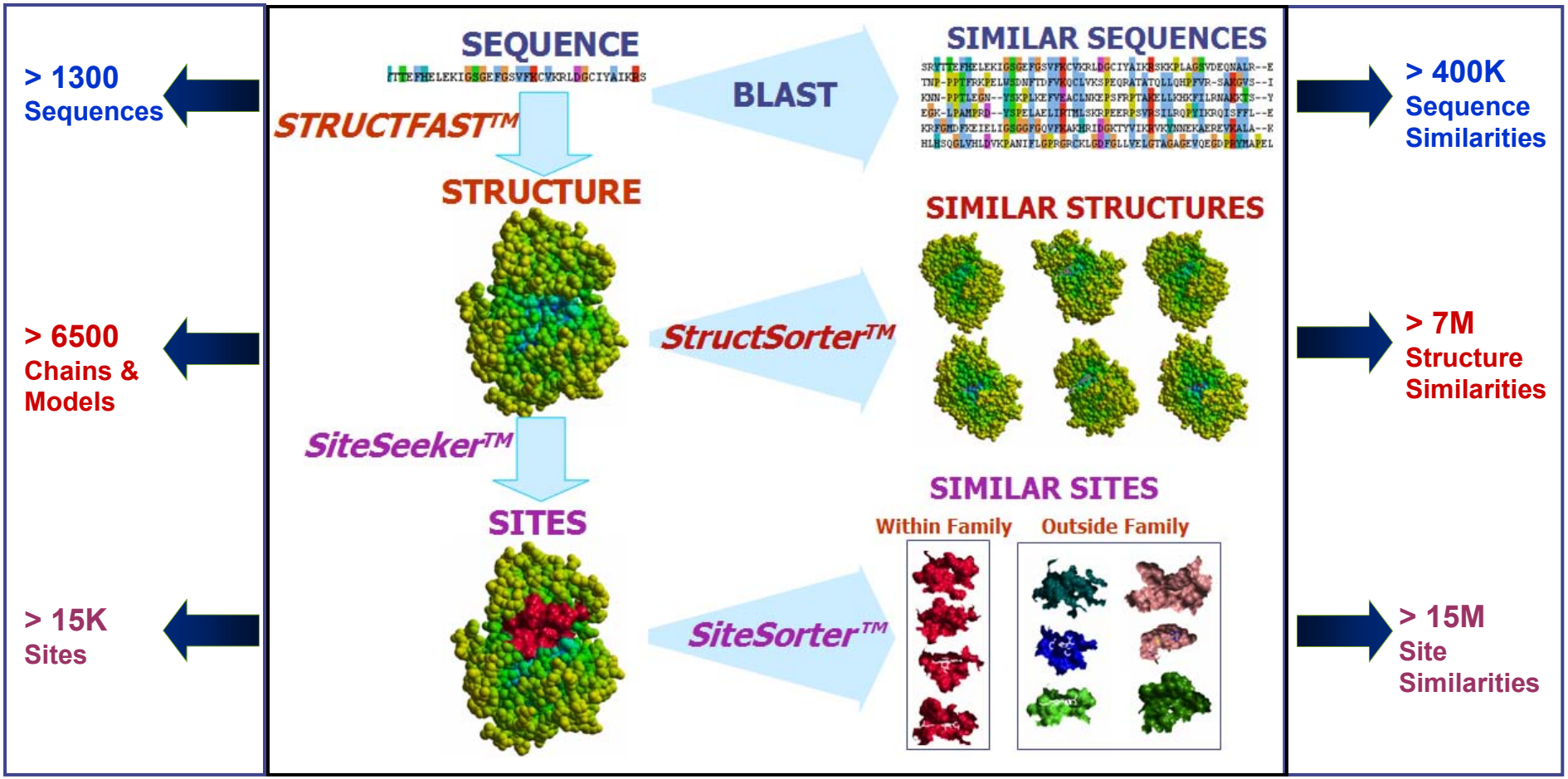
Surface atoms assigned one of 5 different chemical characters (pseudocenters); matching points increase the site similarity score

TIP (Eidogen-Sertanty) / Klebe et. al. J Mol Chem 2002

TIP/Kinase – 2009 Promotional Bundle

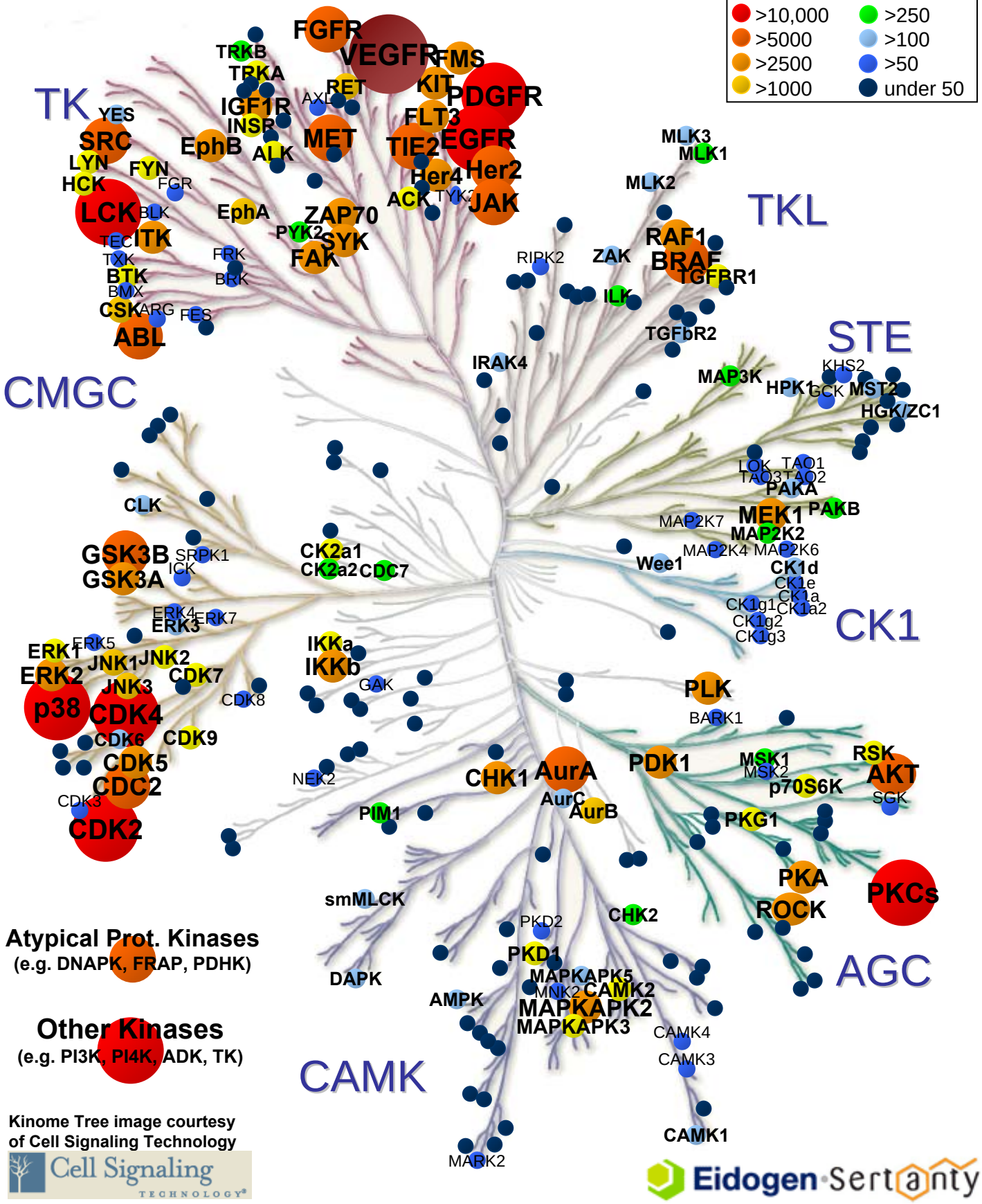
- **TIP/Workgroup technology**
  - Behind-the-firewall with web interface and commandline utilities
  - TIP database creation, administration, and update capabilities
  - Optionally available module: TIP/Webservices
- **TIP Kinase Family Database**
  - Over 1300 sequences (~500 human) modeled w/multiple 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

Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from

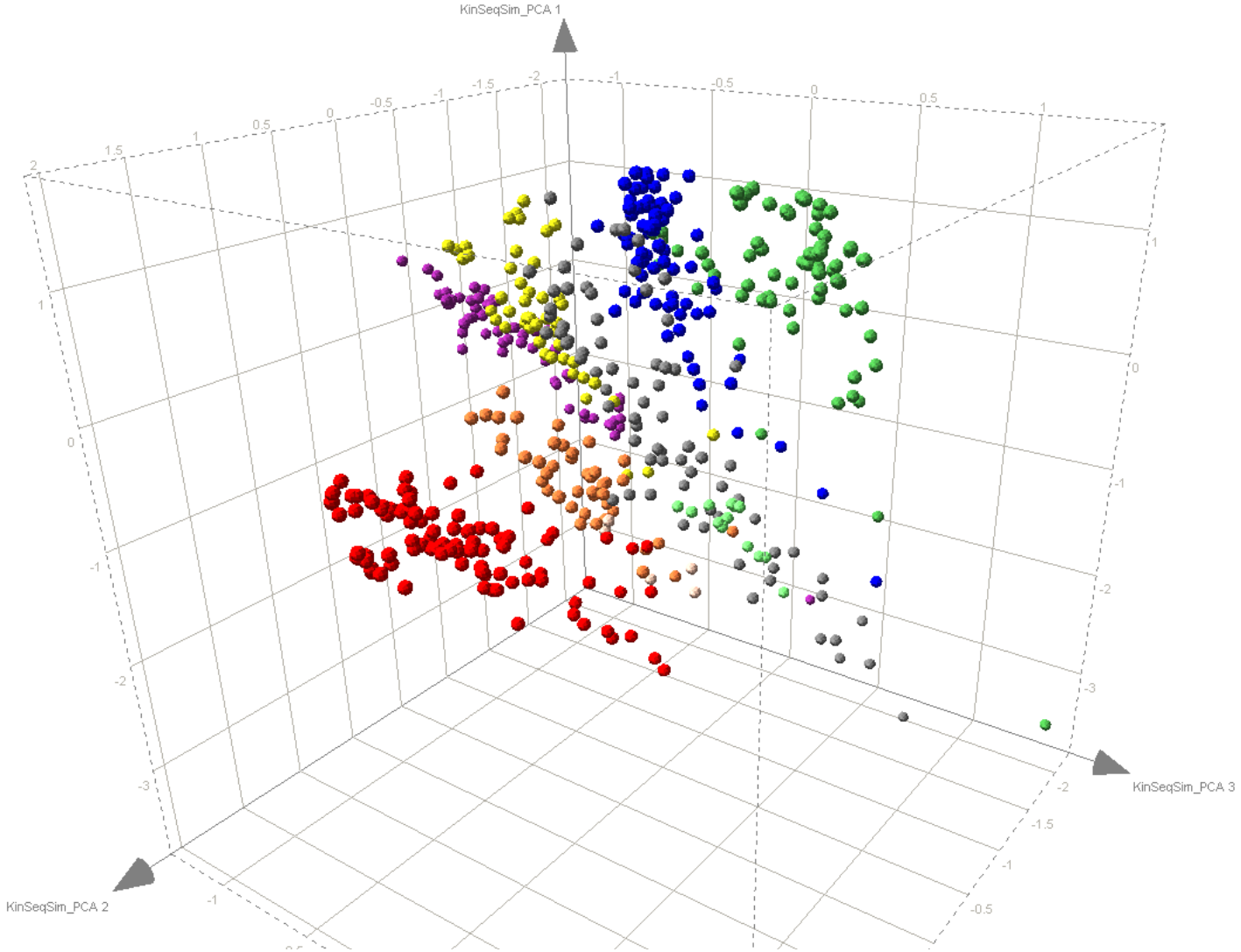
> 5567 journal articles/patents

- **KKB Content Summary (Q1-2009):**
  - # of kinase targets: **>390**
  - # of SAR Data points: **> 402,000**
  - # of **unique** kinase molecules with SAR data: **>141,000**
  - # of annotated assay protocols: **>18,350**
  - # of annotated chemical reactions: **>2,300**
  - # of unique kinase inhibitors: **>486,000** **(~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

Kinome by Sequence Similarity

PCA View – All Pairwise Similarities



494 domain sequences; 3 PCA dimensions preserve 61 % variability

Kinase Target Similarities by SAR

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

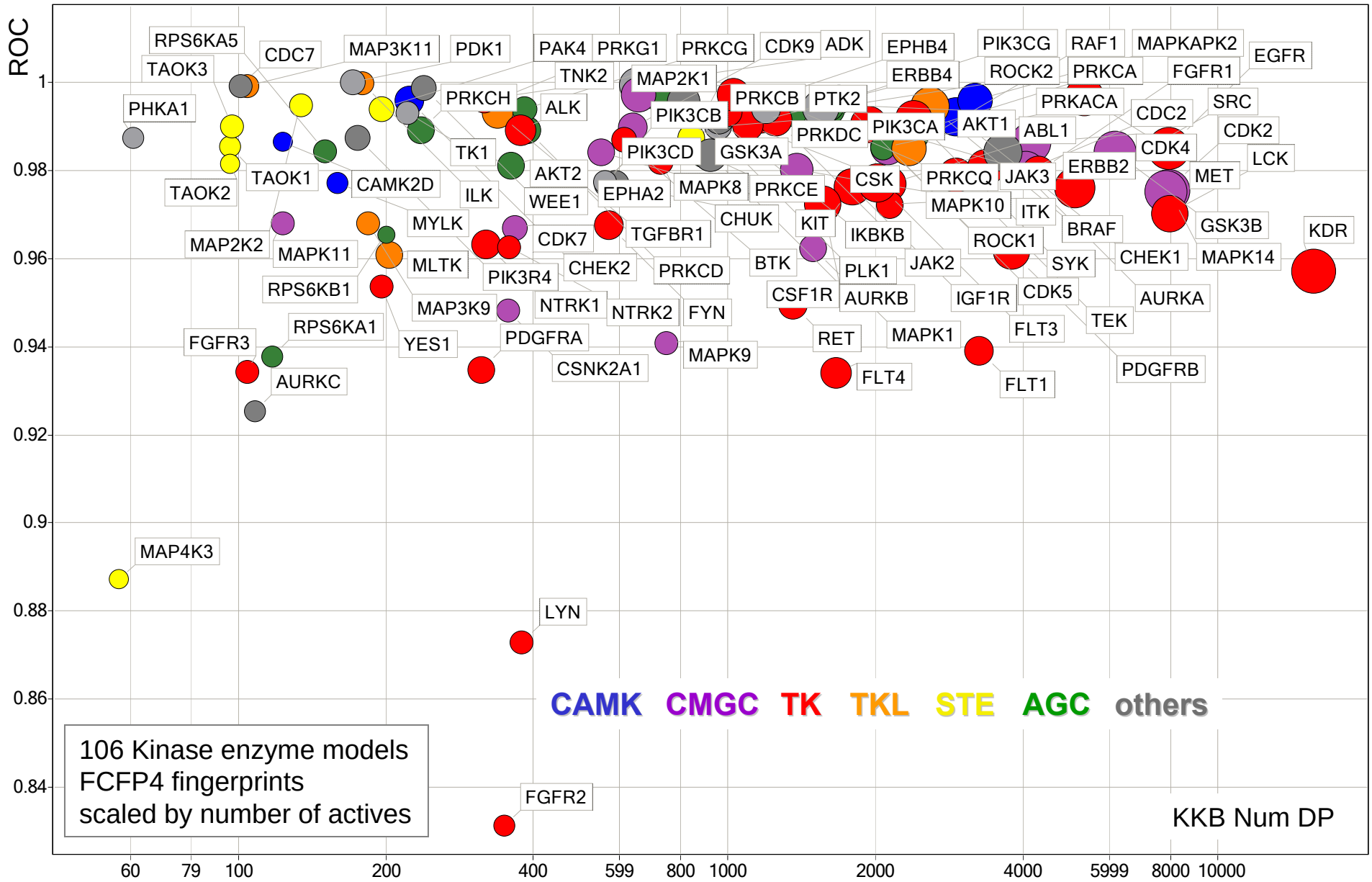
- Activity cutoff pVal ≥ 6.5; minimum 20 actives per kinase pair
- Compute Minimum spanning tree (Kruskal)
  - Visualization as network tree (Cytoscape)

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

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

Eidogen·Sertanty

Kinase SAR Naïve Bayse Models



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 SAR-based Similarities – Summary

- Growing body of accessible kinase inhibition data facilitates a more comprehensive analysis of kinase polypharmacology
- Evolving picture, currently still a sparse kinase – inhibitor matrix
- SAR similarity analysis supports a global intuitive trend: the more similar a kinase the more likely to bind to the same compound
- Phylogenetic kinase tree breaks down in activity space; many examples of compounds that bind to “distant” kinases

- Bayesian models are robust and tolerant to noise and false positives
- Considering “features” maybe less sensitive to the gaps in the accessible data and has the potential to predict cross reactivity for novel compounds
- Fairly robust wrt activity cutoff and fingerprints used

- Be aware of limitations of descriptor-based statistical modeling
- No consideration of how a compounds binds (DFG-in/ -out)

- Small molecules can in many cases be optimized to differentiate between very similar (sequence) kinases in many cases

Kinome by Local Structural Binding Site Similarities (physicochemical)

Kinases Comparison by ATP Site Similarity

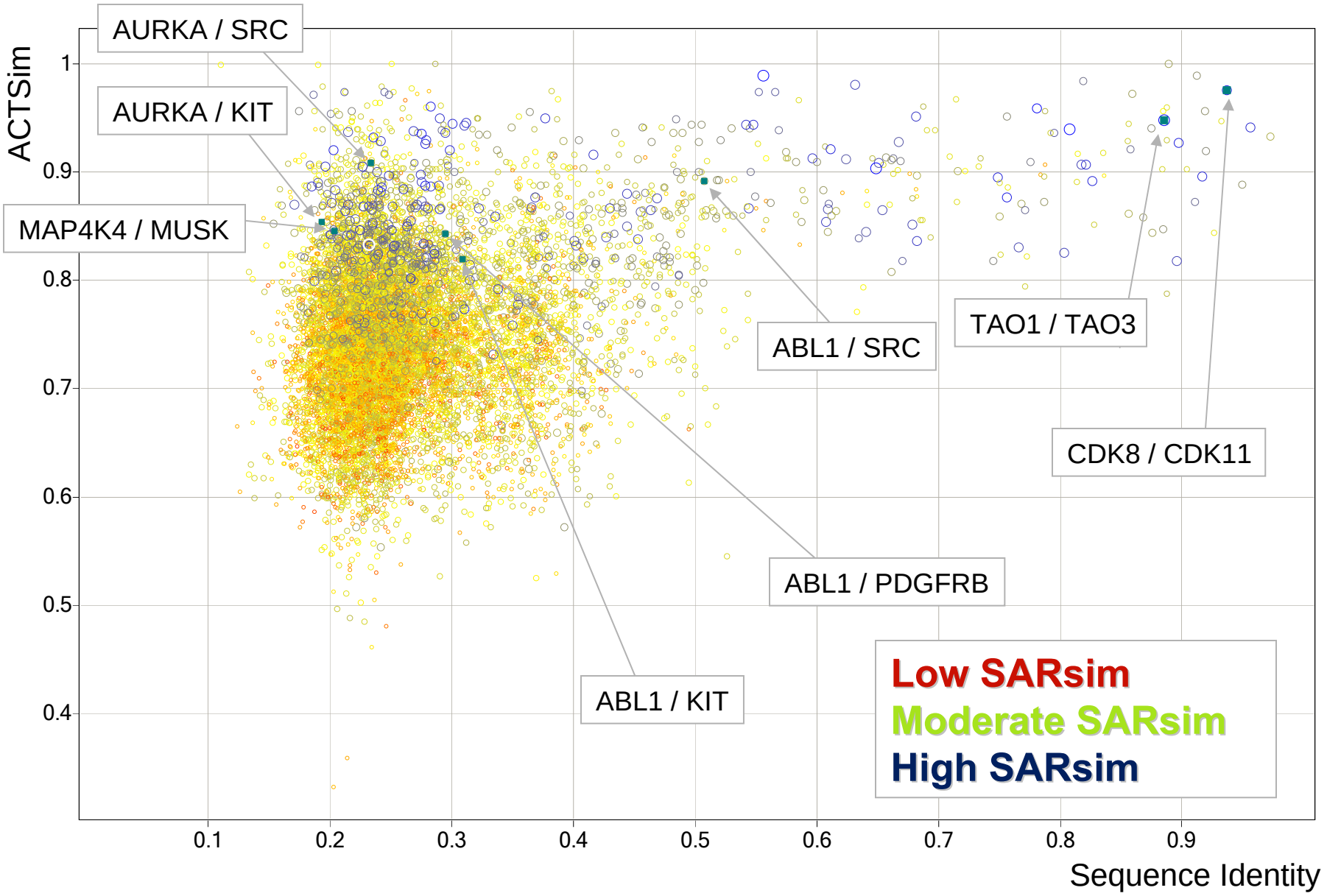
- Extract kinase domain sequences (Sugen, Swissprot, PFAM)
- Model almost the entire Kinome (501 sequences) using STRUCTFAST automated homology modeling (1,117 templates, > 5,000 models)

STRUCTFAST, Proteins **2006**, 960

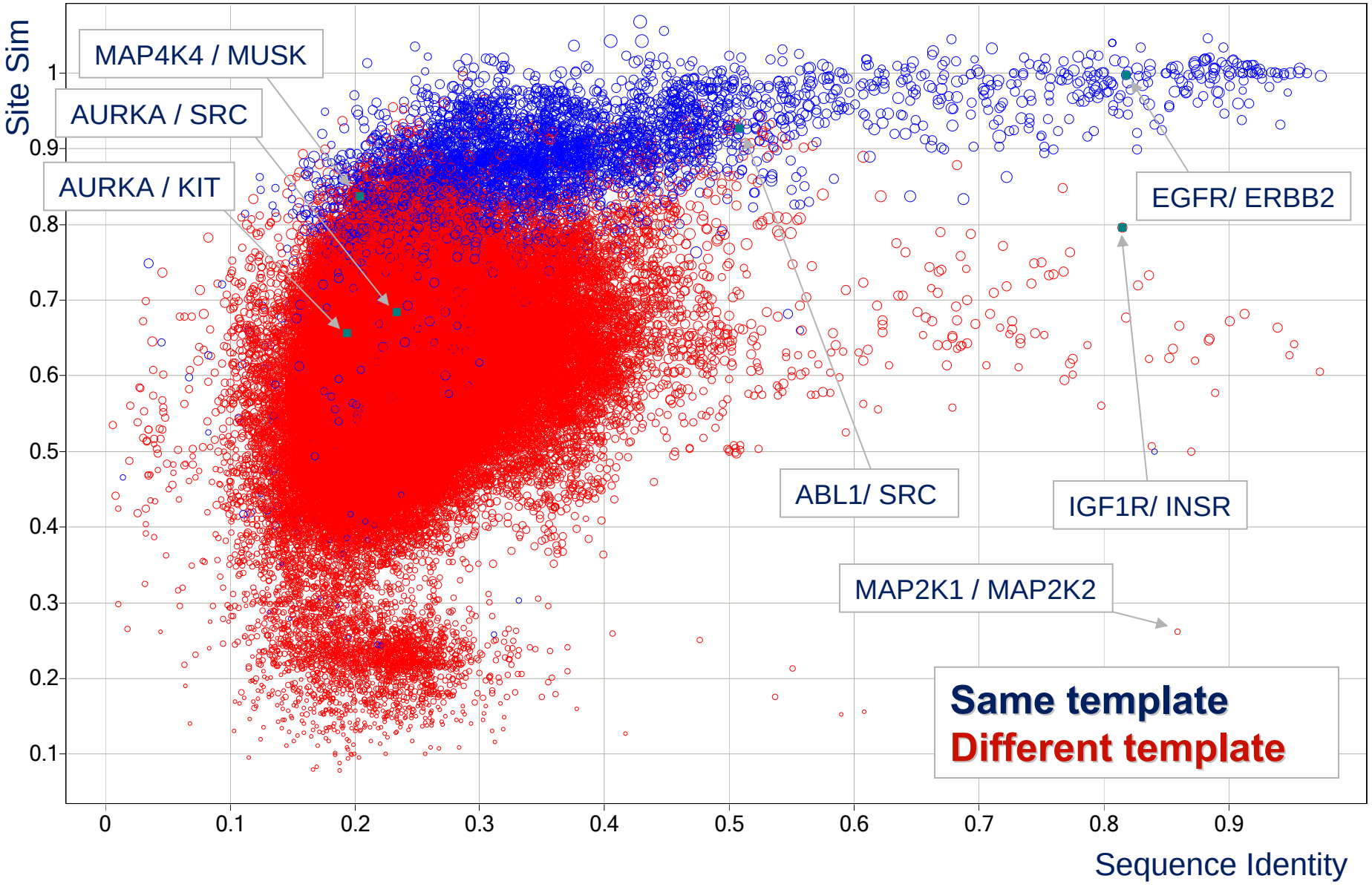
- Define ATP binding sites for all models (homology and predicted)
- Compute binding site similarities
  - Define binding site amino acid features
  - Construct a graph: nodes are all corresponding features of the two sites; edges exist if the spatial distance of the a feature pair is similar between the two sites
  - Compute a complete sub-graph by clique detection (~100 solutions)
  - Overlay sites of the clique solution and sum up the corresponding surface areas
- Compute scores for all site pairs and each site for itself
- Normalize Tanimoto-like: AB_Norm := AB / (AA + BB - AB)
- Analyze and visualize (MST, PCA, hierarchical clustering)

➢ Preliminary results reported (DFG-in only, homology sites only)

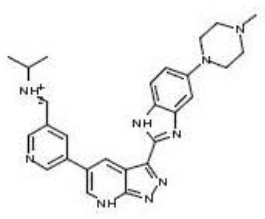
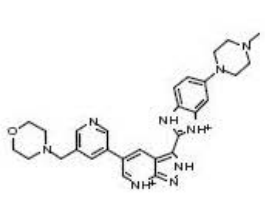
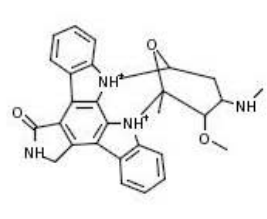
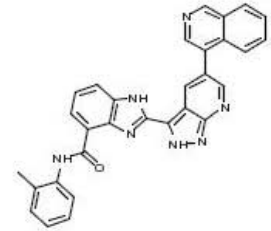
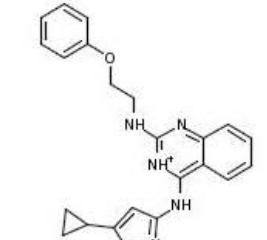
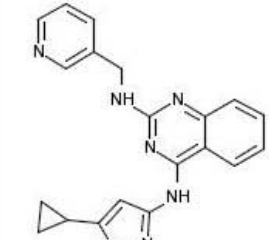
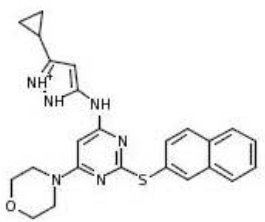
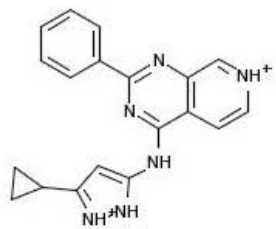
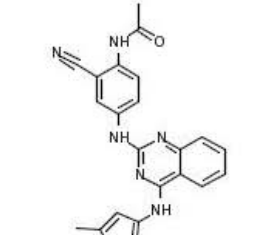
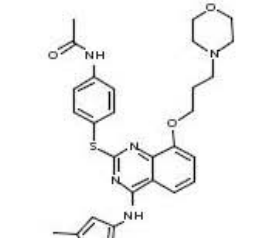
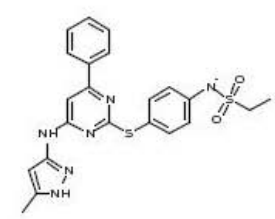
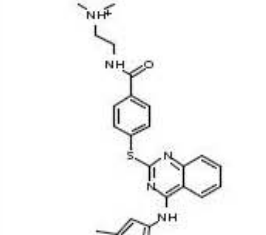
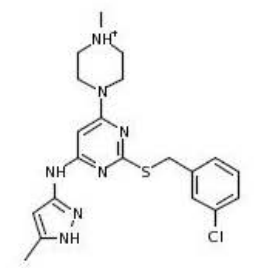
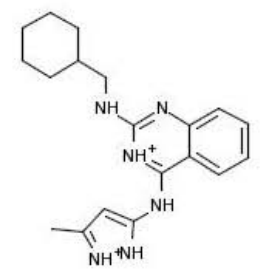
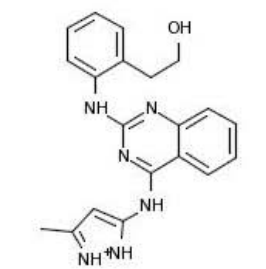
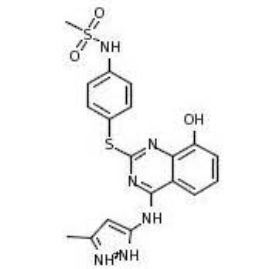
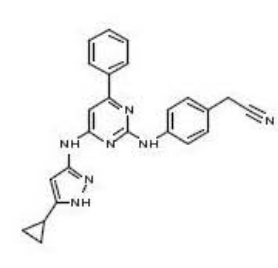
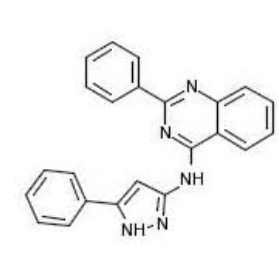
Kinase Target Similarity by ACTsim/SARsim



PhysChem SiteSim vs. Domain Sequence Identity



AURKA and SRC Kinase Dual Inhibitors

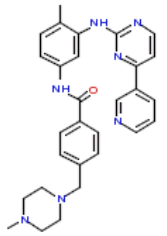
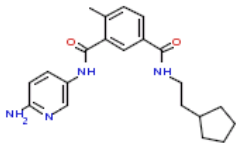
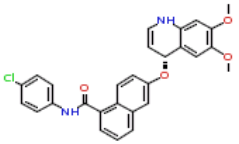
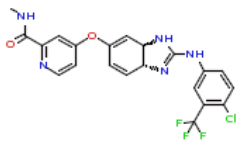
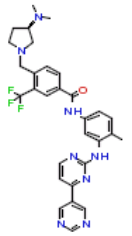
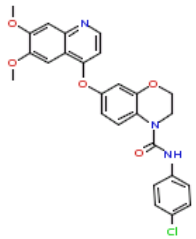
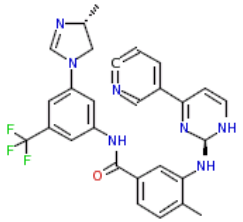
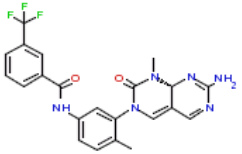
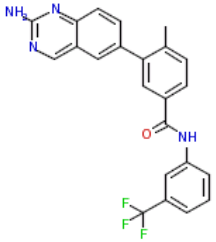
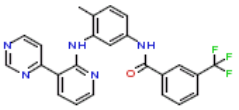
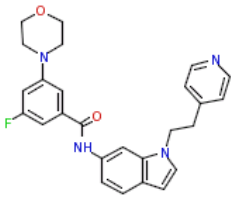
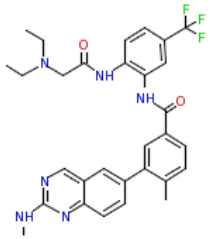
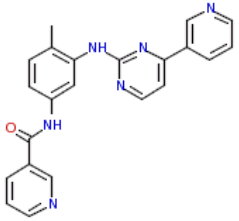
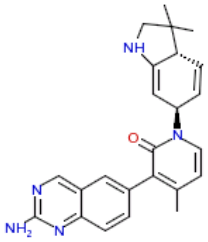
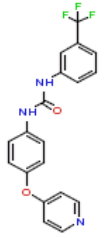
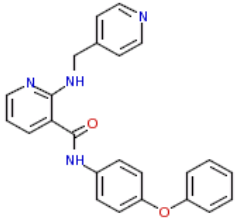
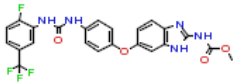
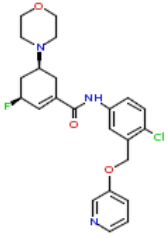
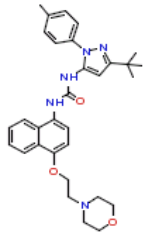
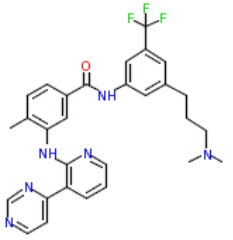
 AURKA / 6.6 SRC / 8.2	 AURKA / 7.1 SRC / 8.3	 AURKA / 7.0 SRC / 7.8	 AURKA / 6.6 SRC / 8.9	 AURKA / 7.0 (min) SRC / 7.0 (min)	 AURKA / 7.0 (min) SRC / 7.0 (min)
 AURKA / 7.0 (min) SRC / 7.0 (min)	 AURKA / 7.0 (min) SRC / 7.0 (min)	 AURKA / 7.0 (min) SRC / 6.5 (est)	 AURKA / 7.0 (min) SRC / 6.5 (est)	 AURKA / 7.0 (min) SRC / 6.5 (est)	 AURKA / 6.5 (est) SRC / 6.5 (est)
 AURKA / 7.0 (min) SRC / 7.0 (min)	 AURKA / 7.0 (min) SRC / 7.0 (min)	 AURKA / 6.5 (est) SRC / 6.5 (est)	 AURKA / 7.0 (min) SRC / 6.5 (est)	 AURKA / 7.0 (min) SRC / 7.0 (min)	 AURKA / 6.5 (est) SRC / 6.5 (est)

diverse subset

Kinome Site Similarities – Summary

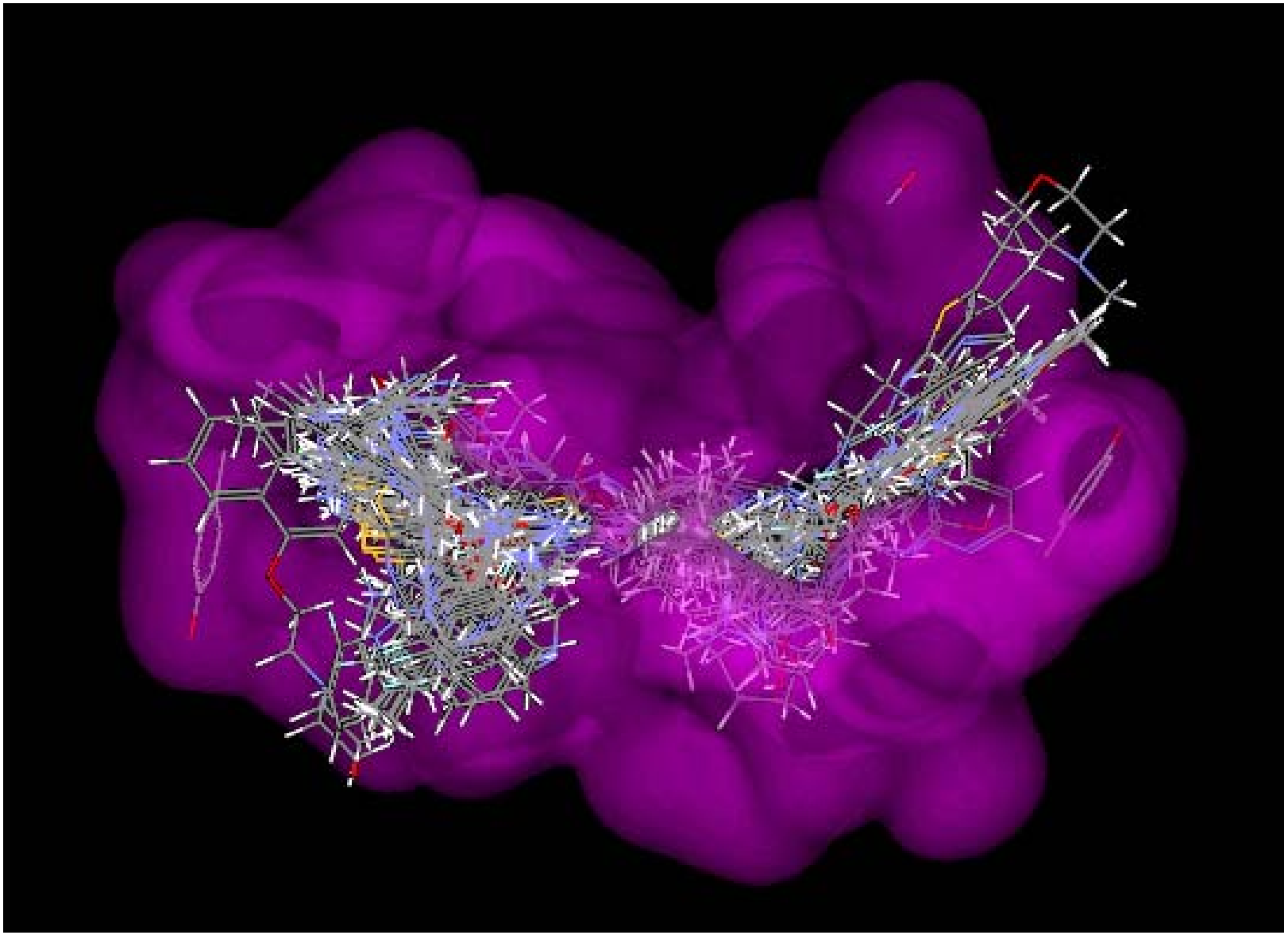
- Relating kinases by local binding site similarity may be meaningful for development of selective inhibitors or compounds with desired profiles
- Many experimental examples confirm the validity of this approach
- Results suggest an expected global trend that similar sequence results in structural- and physicochemical- similar binding sites
- Dissimilar sequences do not always result in different binding sites
- There are subtle differences in the kinase site relationships among groups and sub-types
- Strong template effect
  - only homology sites (from co-crystal templates) are used in the present analysis (similarities using entire solvent accessible ATP sites)
  - for many kinases no experimental structures exist, but they can be modeled
- Although almost all kinases are modelable; experimental coverage and quality of structures will likely influence results
- Growing body of structural information will optimize this picture (in particular co-crystal structures)

Example Ligands Extracted from Similar Sites

 STI	 C92	 900	 276	 406
 2RL	 NIL	 7MP	 242	 MUH
 WBT	 1N8	 PRC	 C19	 KIN
 608	 GIG	 L11	 B96	 RAJ

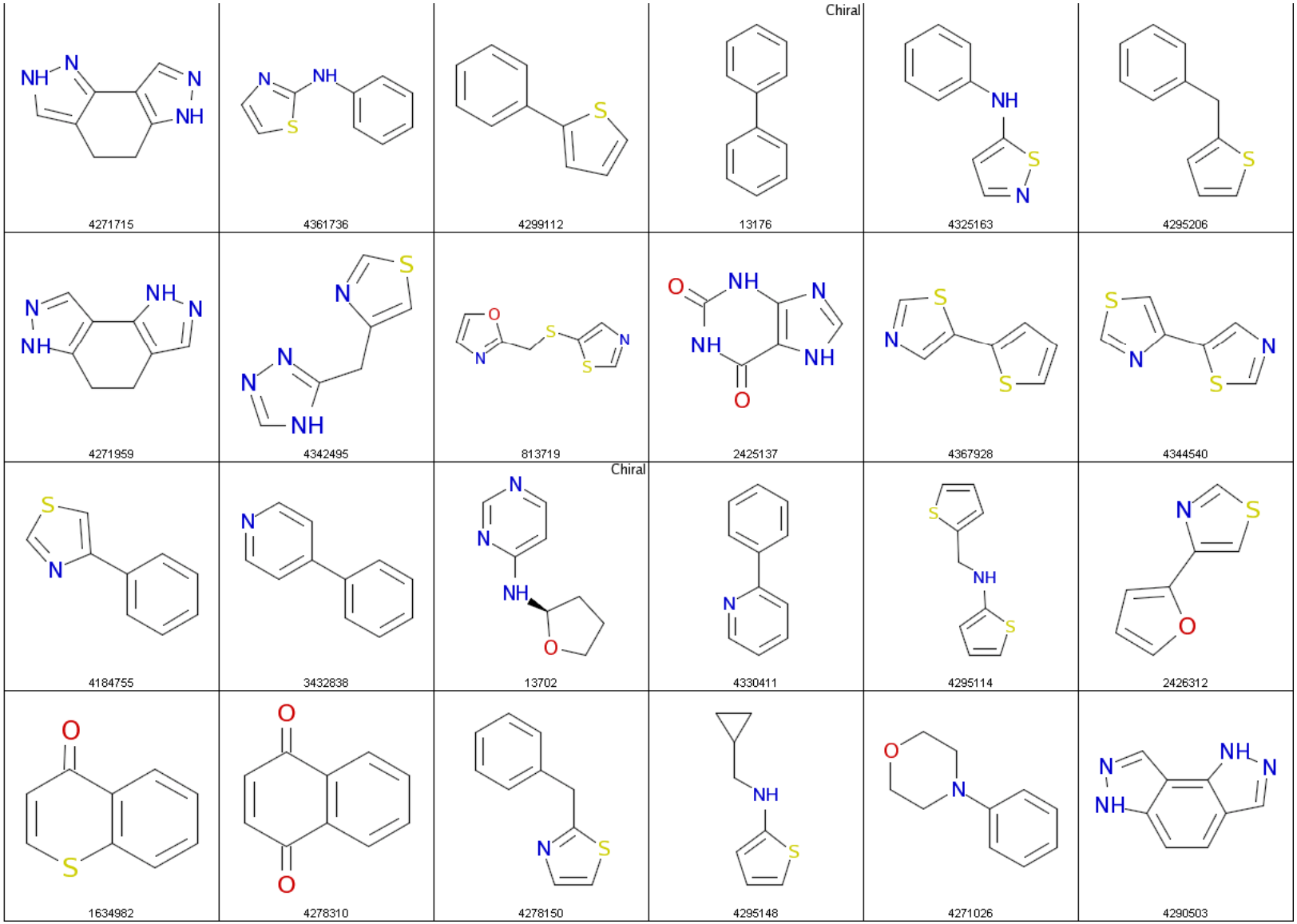
Eidogen-Sertanty

Step 5: LigandCross Ligands reDocked into s1309707



Ligand Functionality Shuffling Across the Kinome

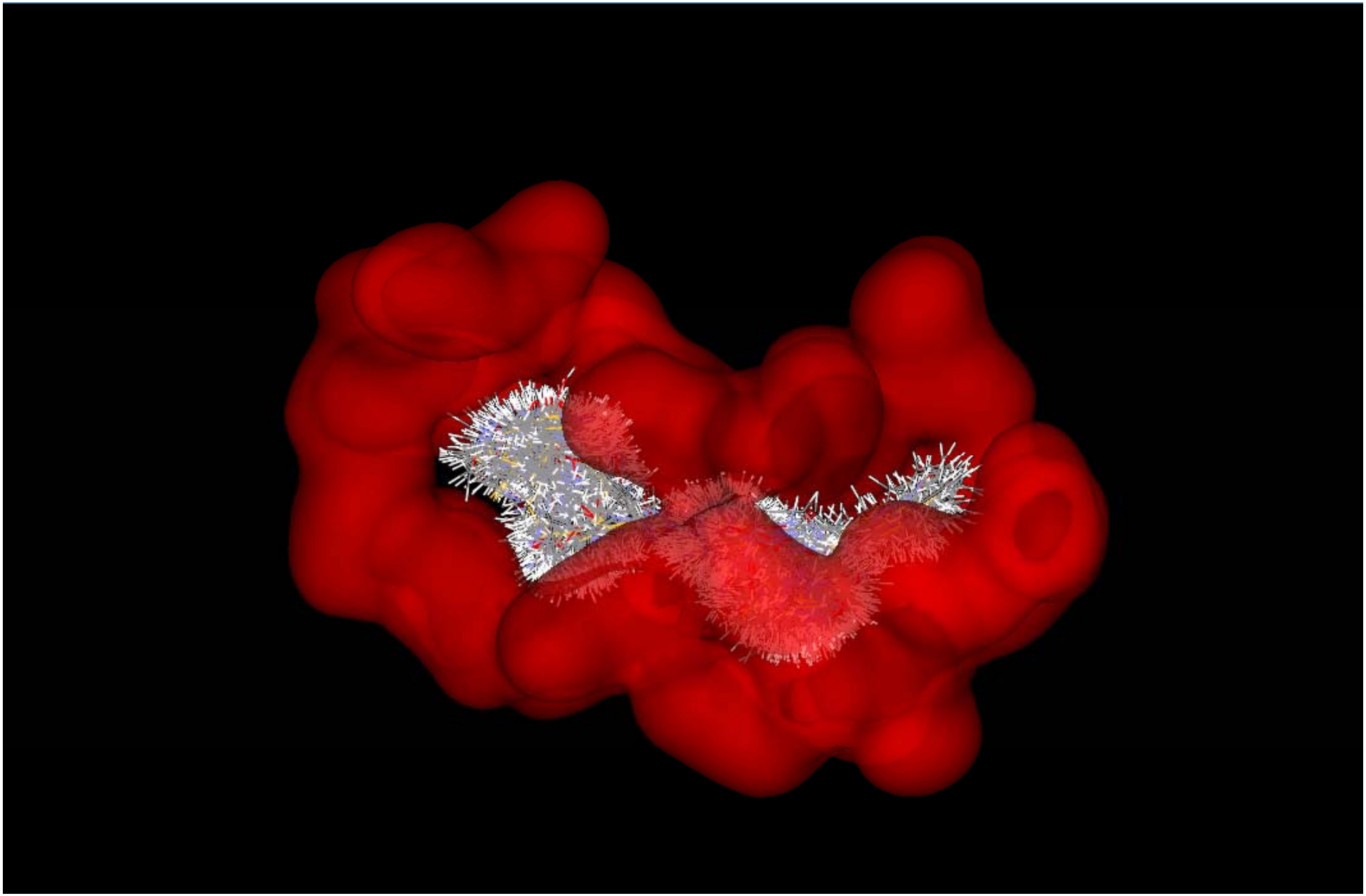
Murcko Assemblies Found in Kinase Inhibitors



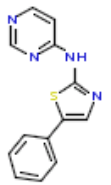
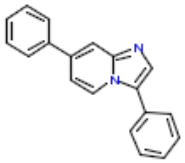
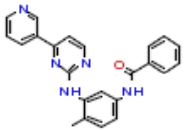
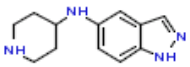
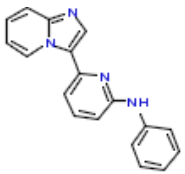
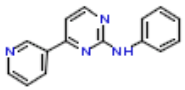
Murcko Assemblies: Contiguous ring systems plus chains that link two or more rings

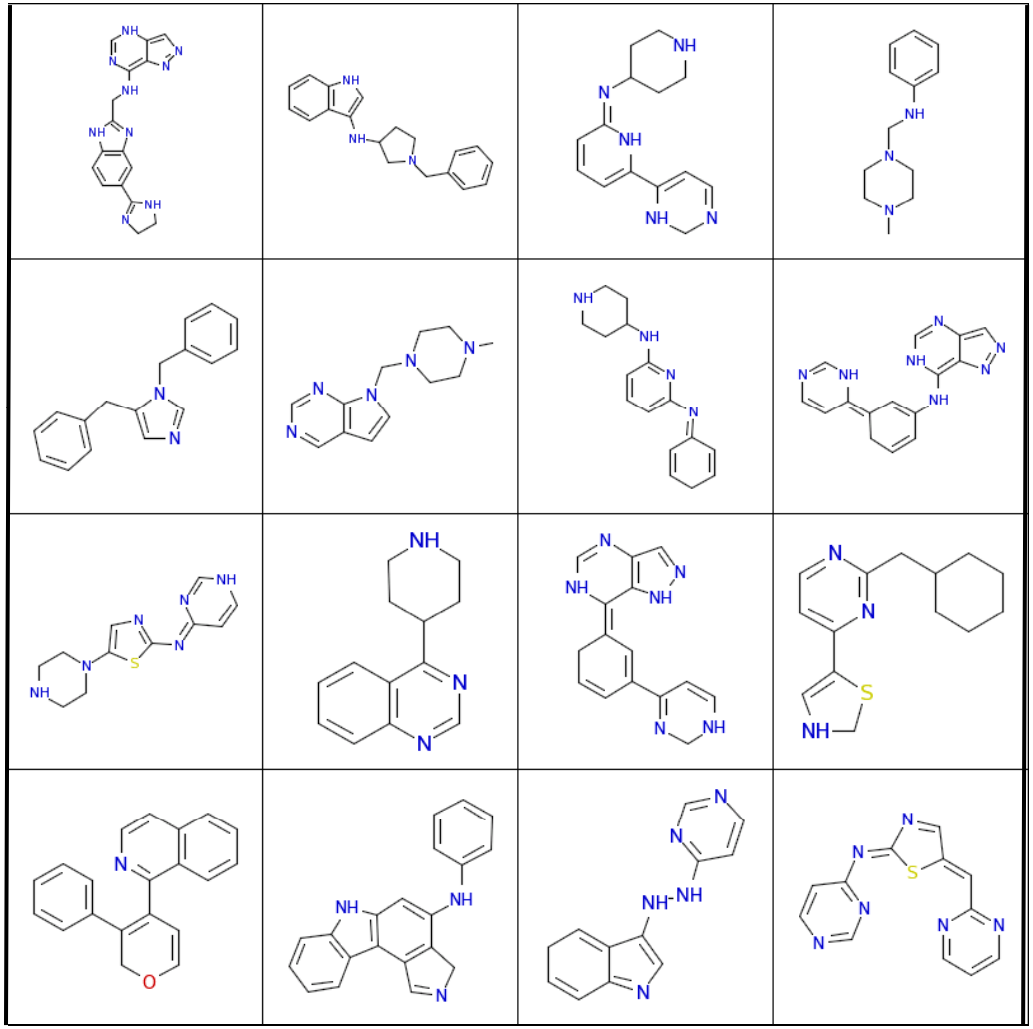
"The Properties of Known Drugs. 1. Molecular Frameworks", Guy W. Bemis and Mark A. Murcko, *J. Med. Chem.* 1996, 39, 2887-2893.

Positional Murcko Assemblies (parent inhibitors docked into s1309707)



LigandCross Results: Positional Murcko Assemblies
from docked Kinase inhibitors (s1309707)

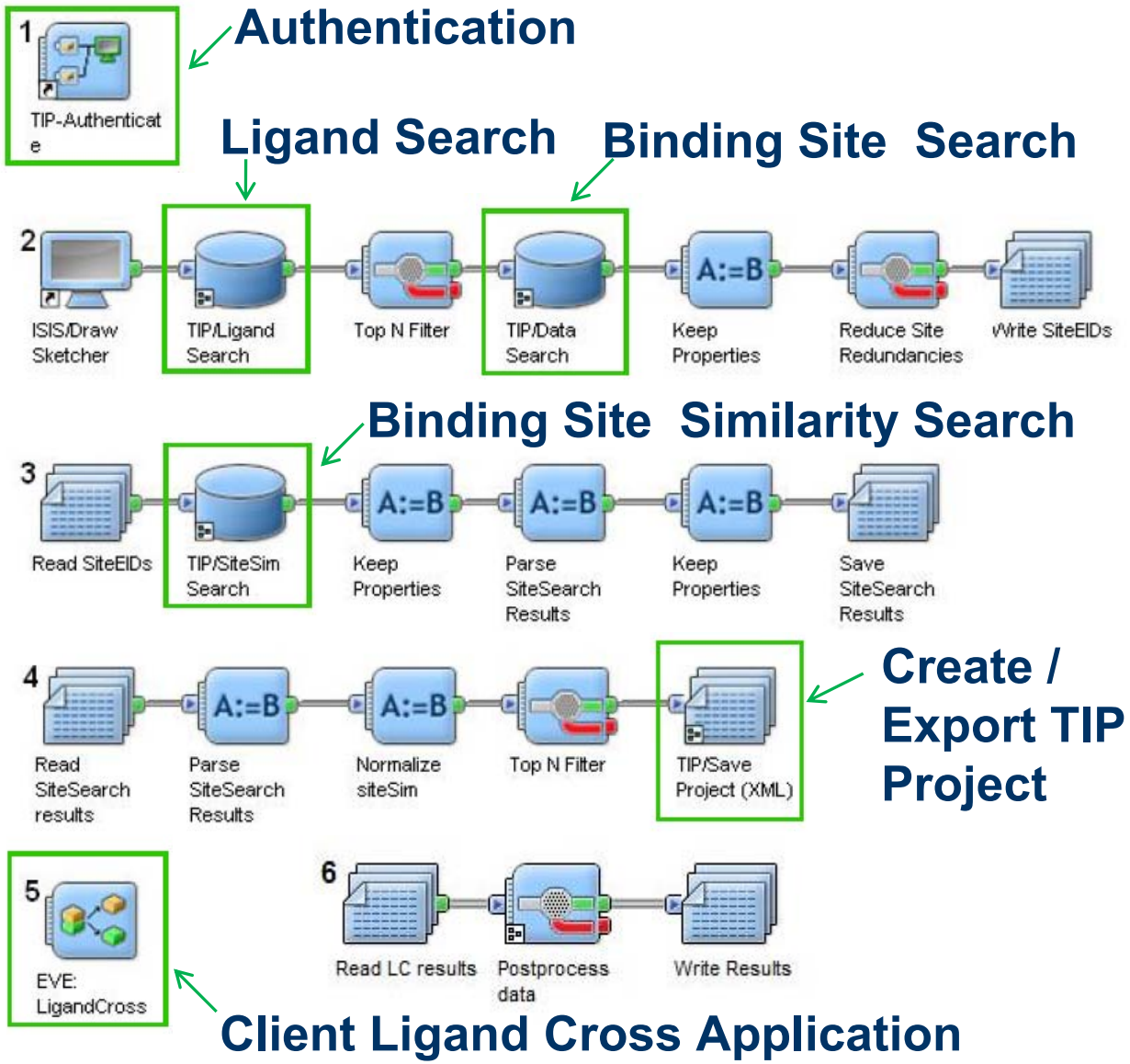
	KDR KDR	Enzyme Assay Enzyme Assay	7.4437 7.4437
	KDR	Enzyme Assay	7.0088
	PDGFR PRKCA PRKCA ABL EGFR PDGFR PDGFR PDGFRB ABL PDGFRB PDGFRB PDGFRA ABL PDGFRB	Enzyme Assay Enzyme Assay Enzyme Assay Enzyme Assay Enzyme Assay Enzyme Assay Enzyme Assay Cell-Based Assay Enzyme Assay Cell-Based Assay Enzyme Assay Enzyme Assay Enzyme Assay Cell-Based Assay	7 4.1427 4.1427 6.3979 4.1871 7 7 7.1871 6.3979 6.2218 7 5.2218 6.3979 6.7696
	ROCK ROCK1	Enzyme Assay Enzyme Assay	6.5421 6.5229
	IRAK4	Enzyme Assay	5.9370
	PRKCA PRKCD ABL EGFR	Enzyme Assay Enzyme Assay Enzyme Assay Enzyme Assay	4.9788 4.4089 5.7447 4

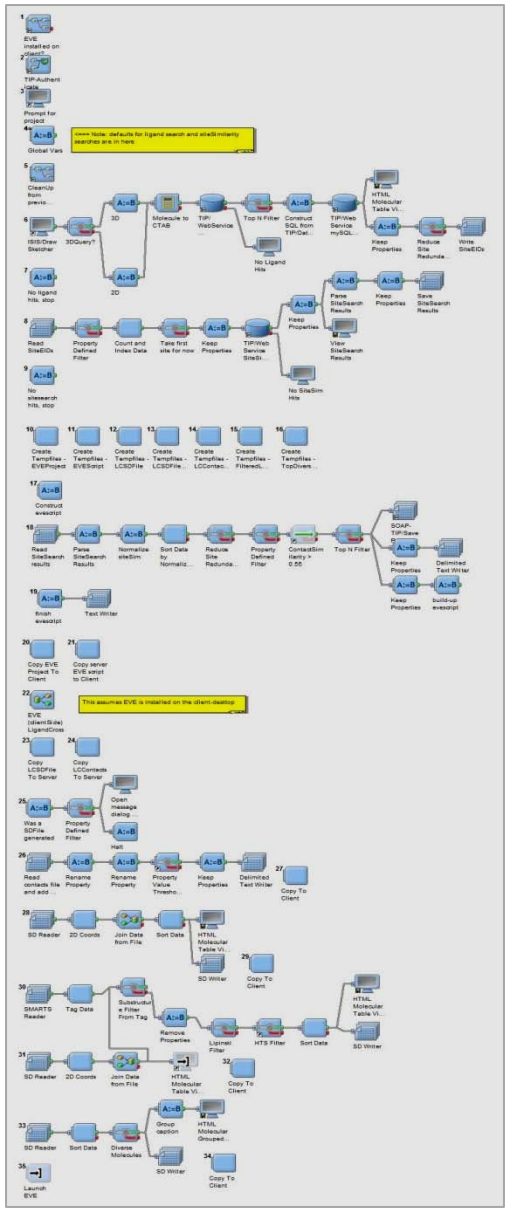


Kinase Activity ????

Eidogen-Sertanty

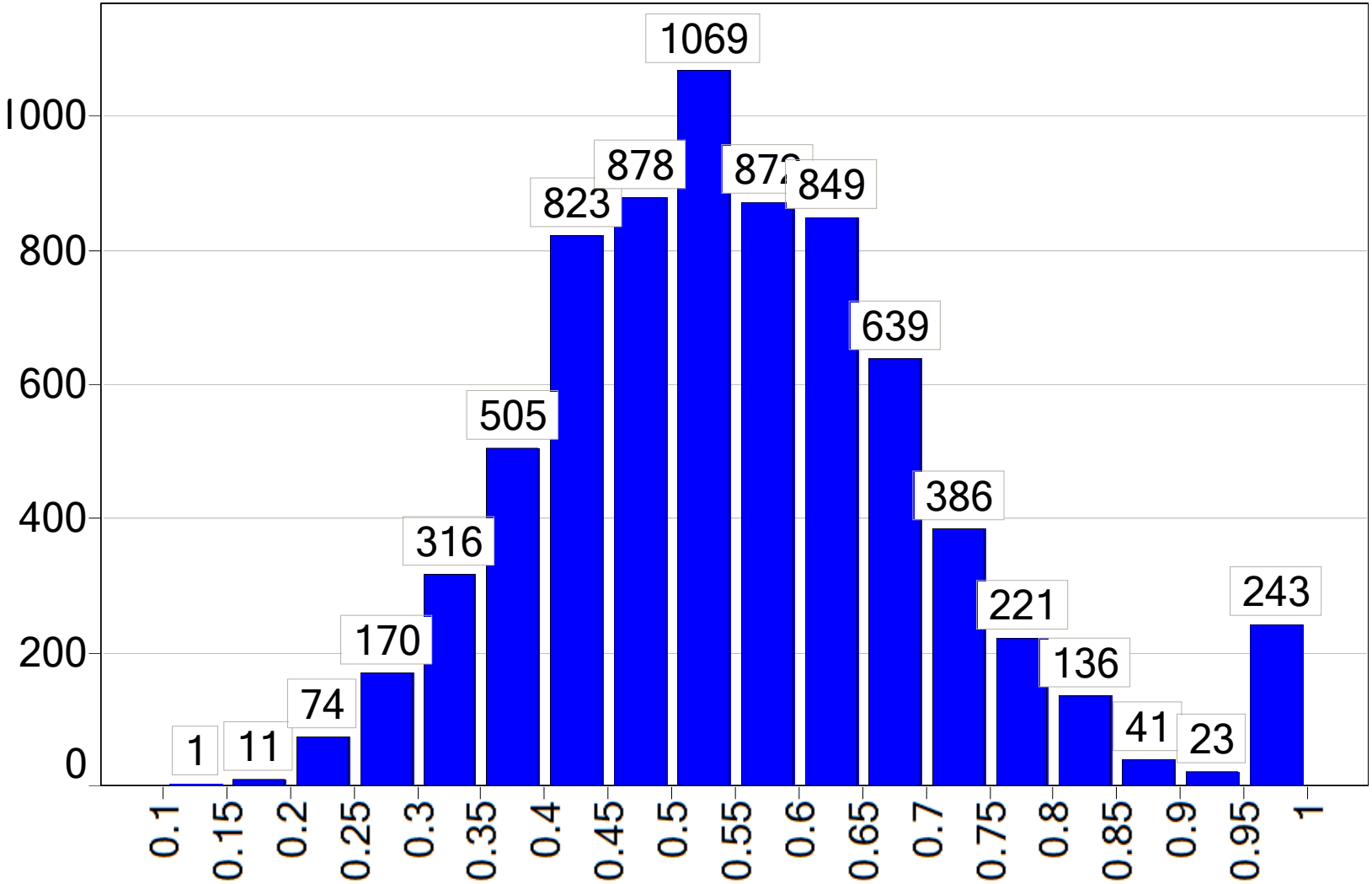
Integration and automation (Web Services via Scitegic)



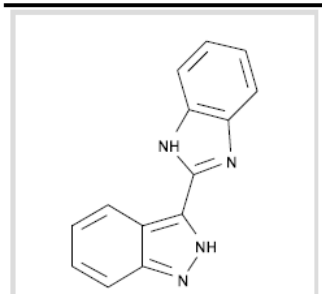


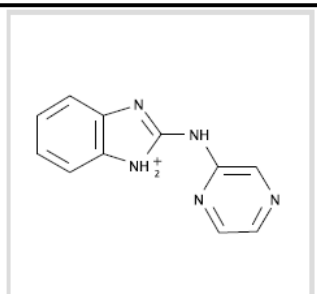
Novel scaffold diversity

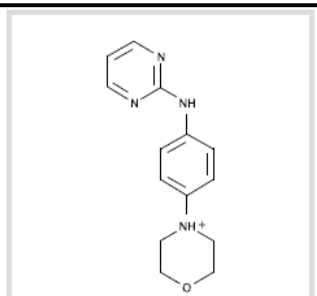
Maximum similarity of unique **Murcko Ring Assmblies**
of LC Results compared to pdb ligand input (ECFP4)

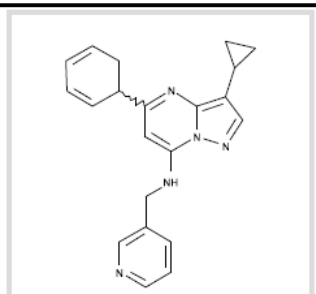


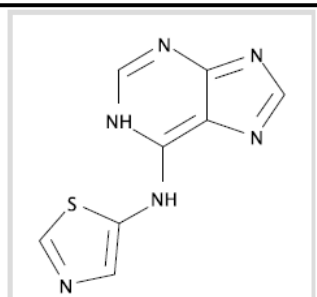
Diverse examples of novel scaffolds (drug/lead like)

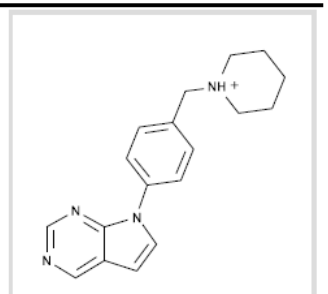

422_12C_2 \ 0.73

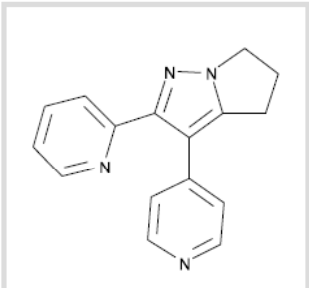

A42_12C_3 \ 0.40

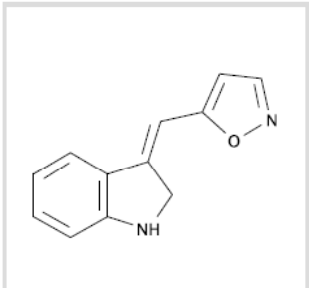

9NH_MTZ_4 \ 0.40

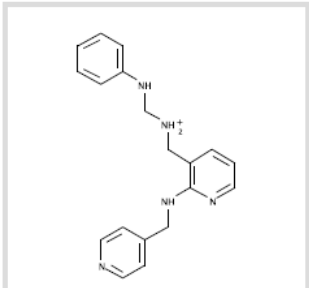

5SC_SCZ_4 \ 0.63

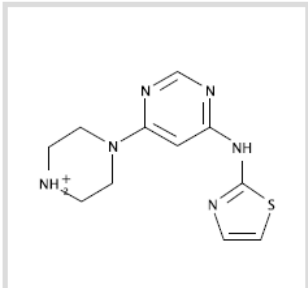

RRC_D42_1 \ 0.53

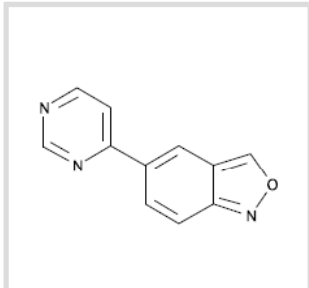

S03_QQ2_3 \ 0.63

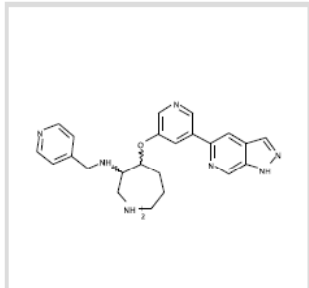

SB2_580_2 \ 0.85

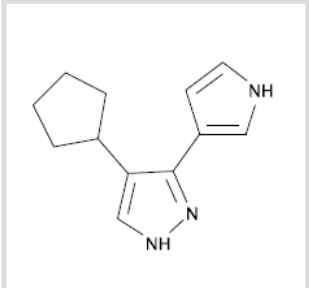

6C3_FMD_2 \ 0.57

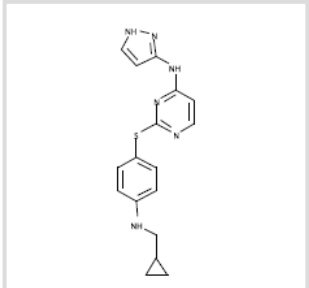

608_900_1 \ 0.60

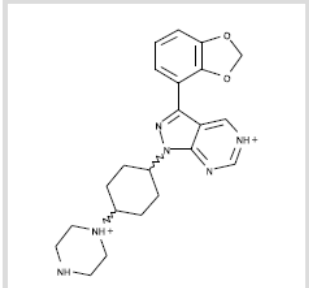

1N1_LQQ_1 \ 0.52

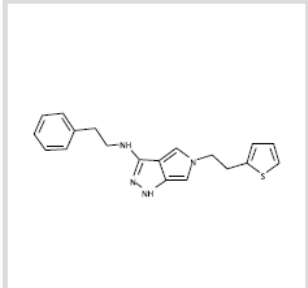

VX1_900_1 \ 0.49

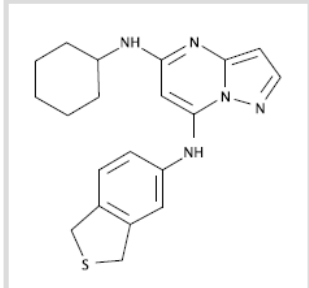

I01_4PY_2 \ 0.39

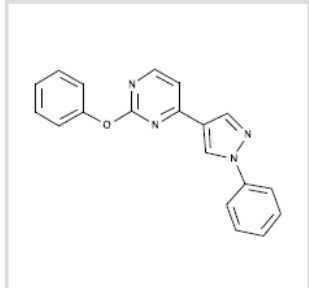

HET_19A_2 \ 0.50


JIN_VX6_7 \ 0.63


L1G_H8H_1 \ 0.36


626_D05_1 \ 0.58


LS3_CT9_2 \ 0.71


PQB_PGJ_1 \ 0.47