Exercising receptor-site similarity:

From Off-Target Identification to Scaffold Hopping

Steven Muskal, Ph.D.

Chief Executive Officer

Eidogen-Sertanty, Inc.

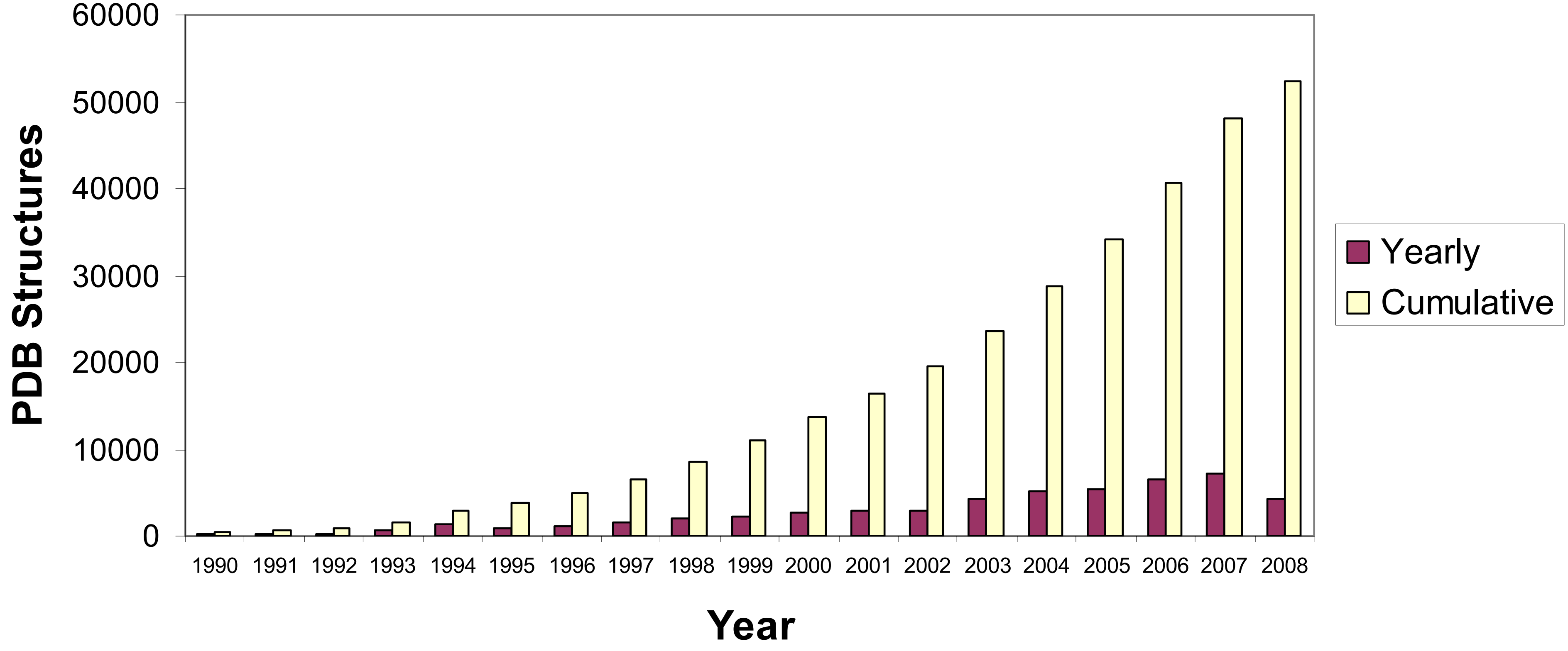
smuskal@eidogen-sertanty.com

Protein Structure Growth is Accelerating

> 50K Structures/co-complexes (Aug-2008)

> 600 deposits per month → **>150/week!**

PDB Growth
source: rcsb.org



Drugs developed using SBDD

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

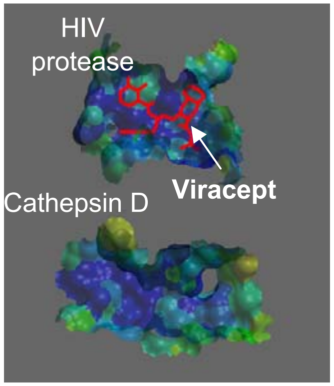
Source: http://www.active-sight.com/science/sbdd.html

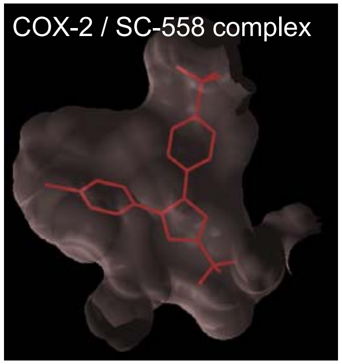
Bringing Proteomic Riches to All Scientists

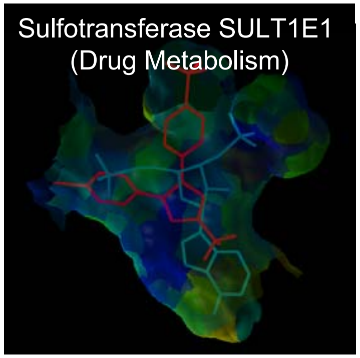
Automated Modeling and Proteomic Structural Mining

- **Sequence-to-structure; keyword-, sequence-, ligand-, protein structure-, receptor-site-searching**
- **Exploit Structural fold, receptor-site conservation, and ligand-based SAR**

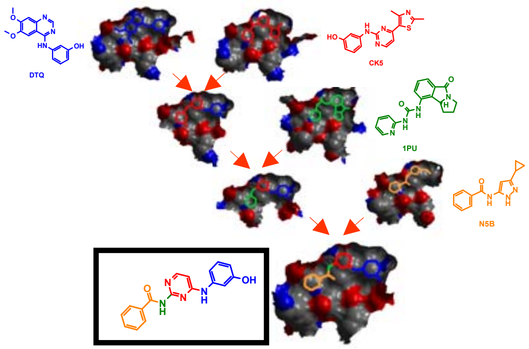
→Off-Target Identification (opportunities v. liabilities)



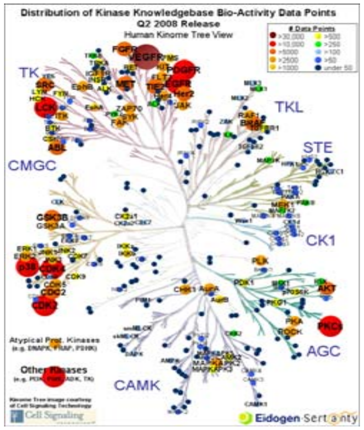




→Borrowing Matter Ideas from co-complexes, protein structures, and SAR



LigandCross



Kinase Knowledgebase
(KKB)

Eidogen-Sertanty

About Eidogen-Sertanty

- Knowledge-Driven Discovery Solutions Provider
  - Formed in March 2005 when Sertanty (Libraria→Sertanty 2003) acquired Eidogen (Bionomix 2000)
  - >$20M Invested in Technology Development
  - 12 FTE's
  - Worldwide Customerbase
  - Cash-Positive

- Chemogenomic Databases & Analysis Software
  - *TIP™* - Structural Informatics Platform
  - *KKB™* - Kinase SAR and Chemistry Knowledgebase
  - CHIP™ - Chemical Intelligence Platform

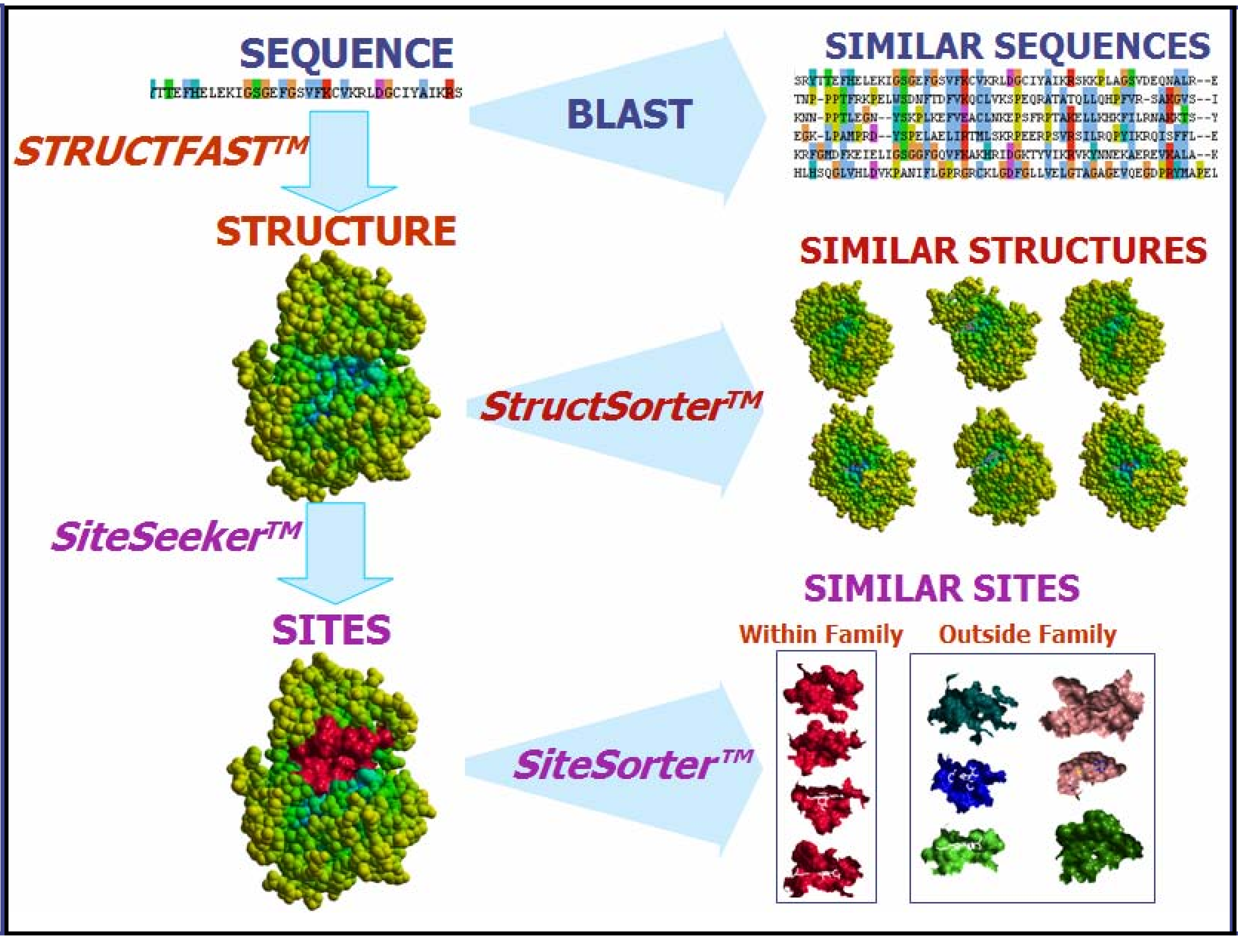
- DirectDesign™ Discovery Collaborations
  - In Silico Target Screening ("Target Fishing" and Repurposing)
  - Target and compound prioritization services
  - Fast Follower Design: Novel, Patentable Leads

TIP Algorithm Engine

> 400K
Sequences

> 158K
Chains & Models

> 388K
Sites



> 33M
Sequence Similarities

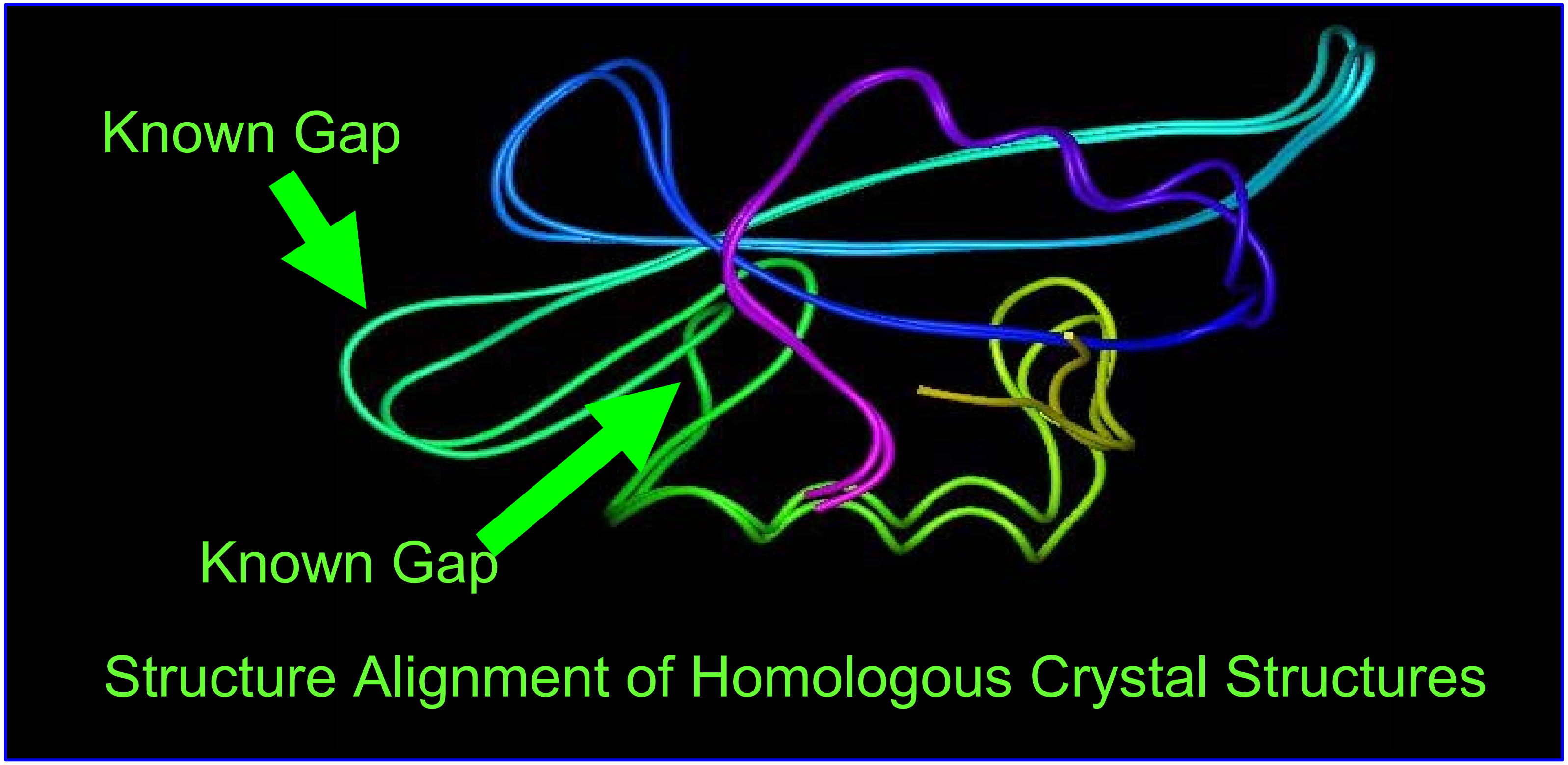
> 69M
Structure Similarities

> 62M
Site Similarities

STRUCTFAST™

STructure Realization Utilizing Cogent Tips From Aligned Structural Templates

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



Structure Alignment of Homologous Crystal Structures

Leverages experimental structure and structural alignment data to create better alignments

1) Convergent Island Statistics: A fast method for determining local alignment score significance. Bioinformatics, 2005, 21, 2827-2831

2) STRUCTFAST: Protein Sequence Remote Homology Detection and Alignment Using Novel Dynamic Programming and Profile-Profile Scoring
Proteins. 2006 64:960-967

SiteSeeker™

Geometric Site-Finding Algorithms Find Many Pockets

But they don't know which pockets are important!

Evolutionary Trace Approach

Can't clearly define site boundary

Not all conserved residues are functionally relevant

SiteSeeker combines *both* methods

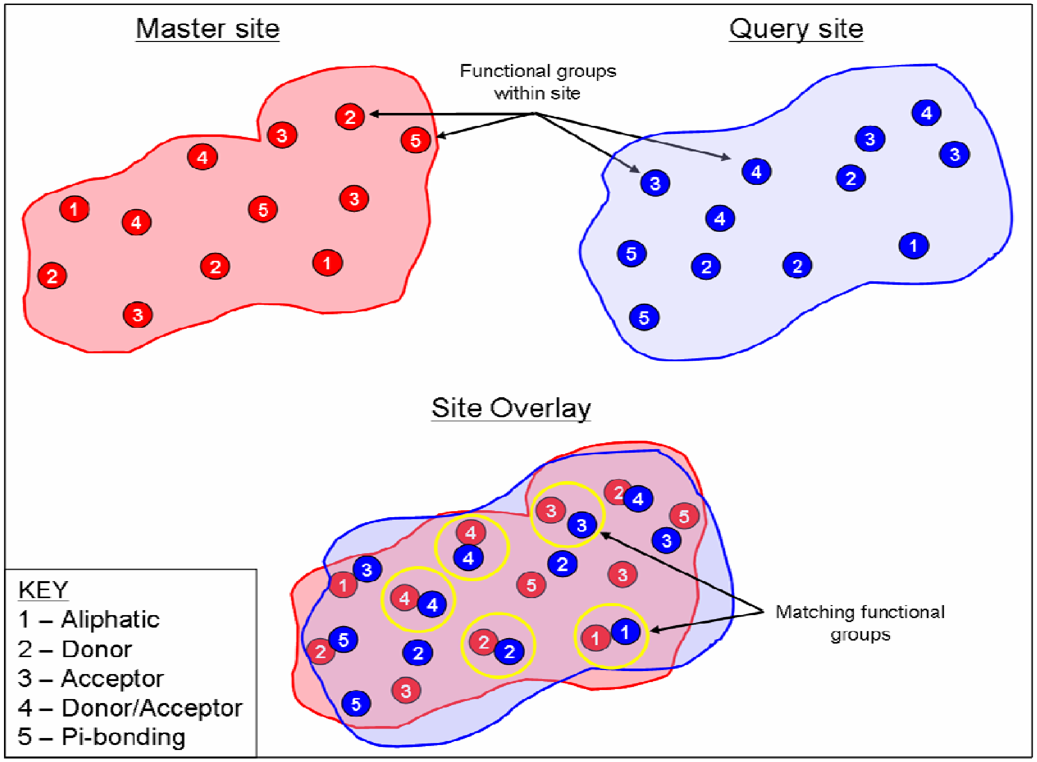
Reliability & Confidence

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

Allows us to map *SiteSeeker* score to predict confidence!
(e.g. At this *SiteSeeker* score, 80% are "real" co-crystal sites)
→ Sites with <60% confidence are not stored in TIP

Weighted Clique Detection Algorithm

Importance of Points Related To Conservation In Multiple Sequence Alignment



KEY
1 – Aliphatic
2 – Donor
3 – Acceptor
4 – Donor/Acceptor
5 – Pi-bonding

Surface Atoms Assigned One of 5 Different Chemical Characters

Matching points increase the *SiteSorter* similarity score

TIP Content

>75,000 Human Sequences

>33M Sequence Similarities

>116,000 Total PDB chains (~50K PDBs)
> 42,000 Homology Models

>69M Structural Similarities

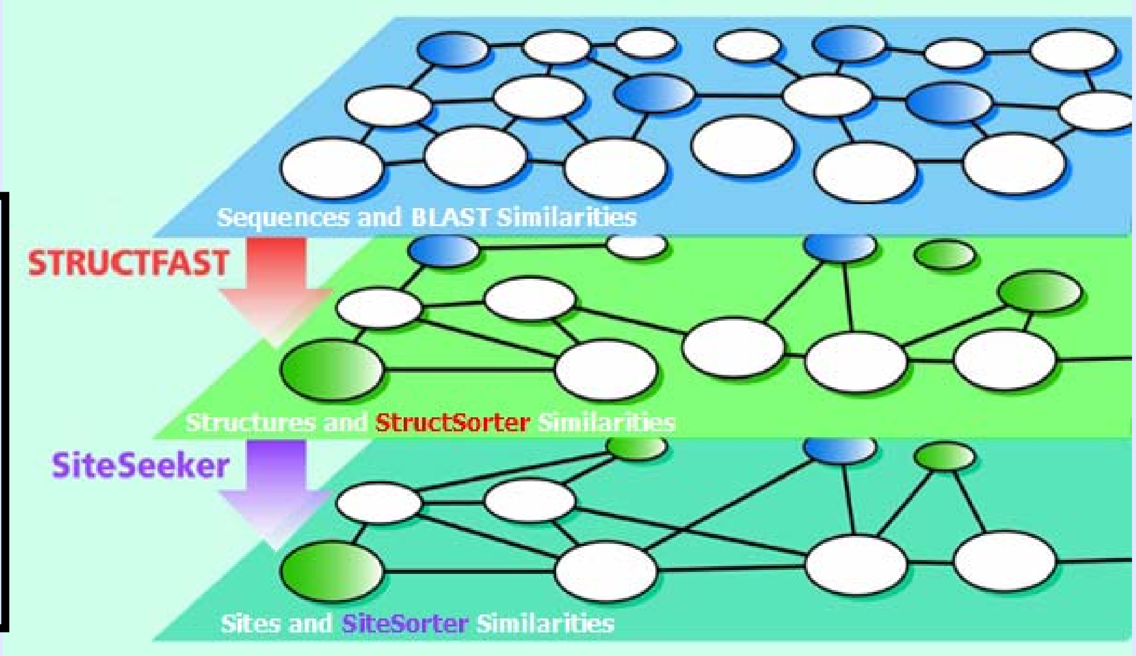
>194,000 PDB co-crystal sites
>190,000 Predicted Sites (on PDBs & Models)

>62M Site Similarities

Updated monthly with new PDBs and models:

e.g. March 2006:
- **→ 661 new PDBs added**
- **→ 447 new models built**
  - \- 153 had no previous structure in TIP
  - \- 294 had “better” models built

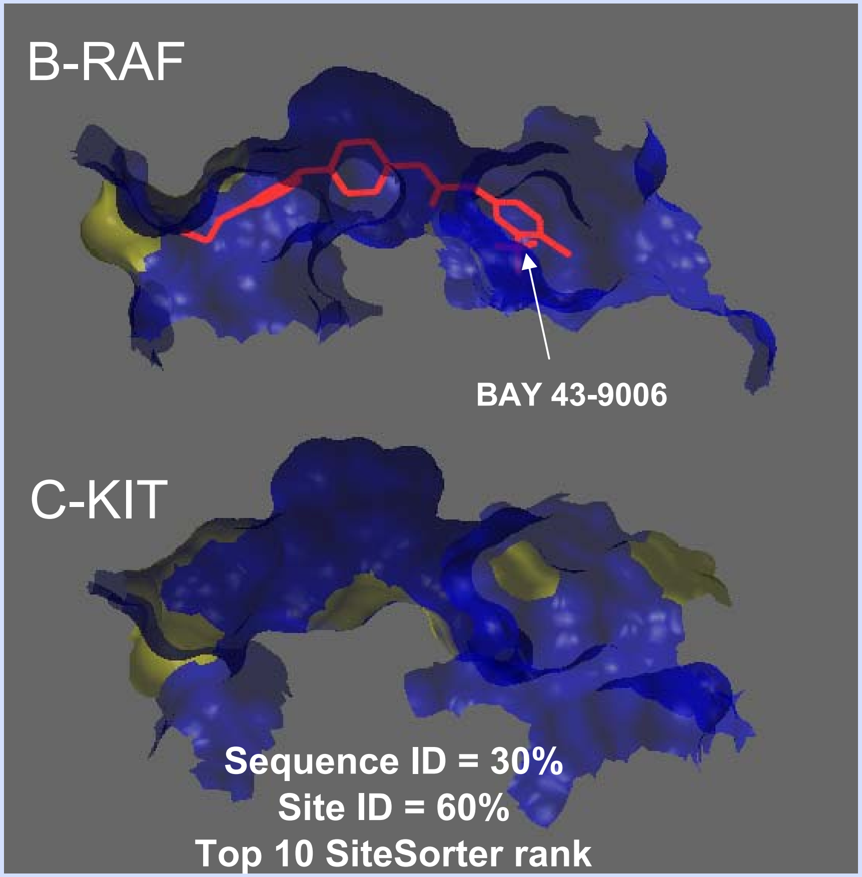
e.g. July 2008:
- **→ 576 new PDBs added**
- **→ 1045 new models built**



Automatically updated with new models as the PDB grows

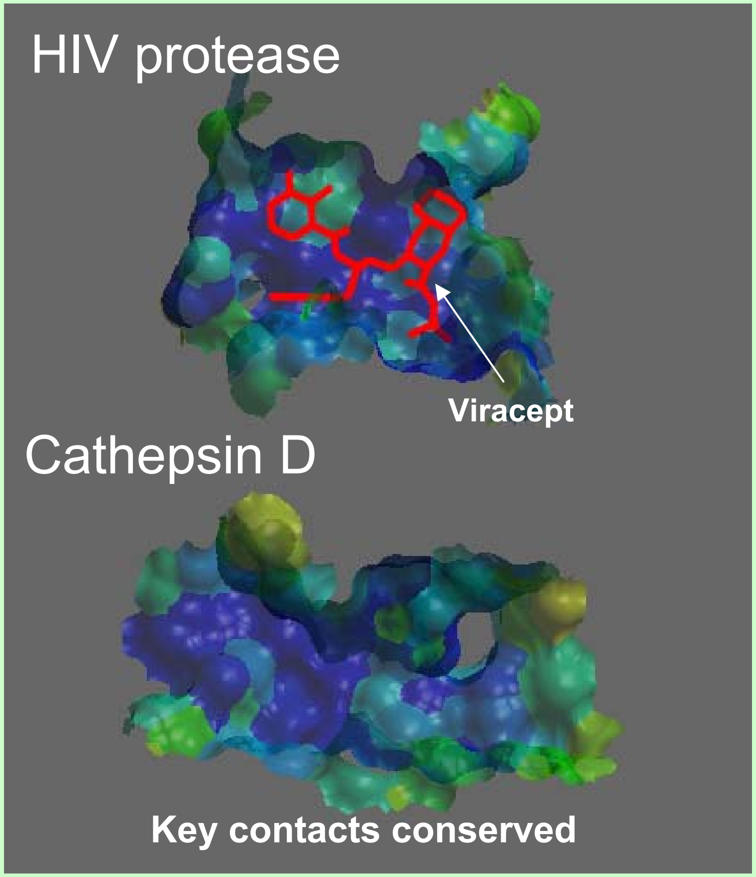
Off-Target Opportunities

Intra-Family Opportunities



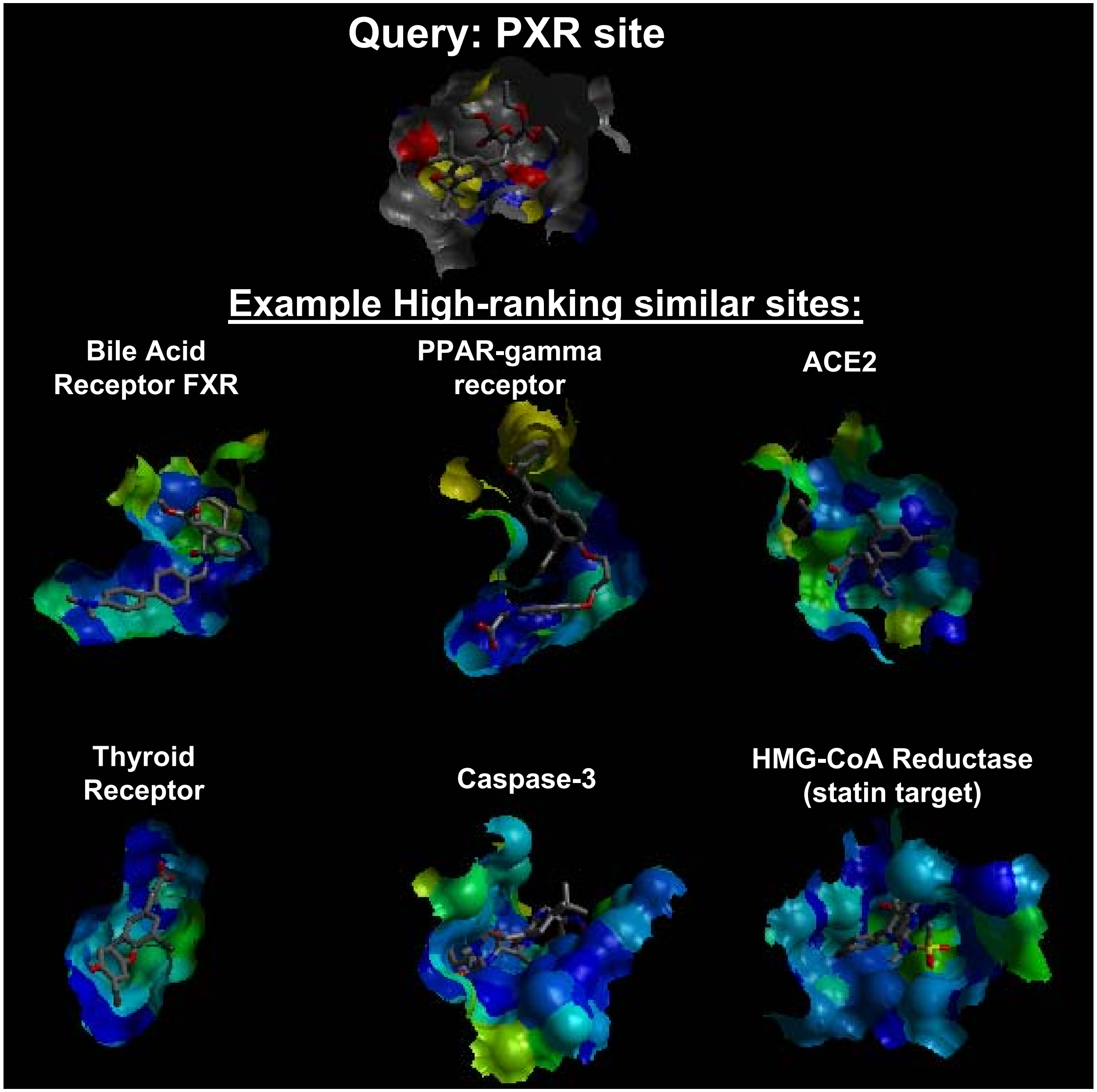
B-RAF inhibitor BAY 43-9006 also inhibits C-KIT

Inter-family Opportunities



Cathepsin D is inhibited by HIV protease inhibitors

PXR – Promiscuous Ligand-Binding Site

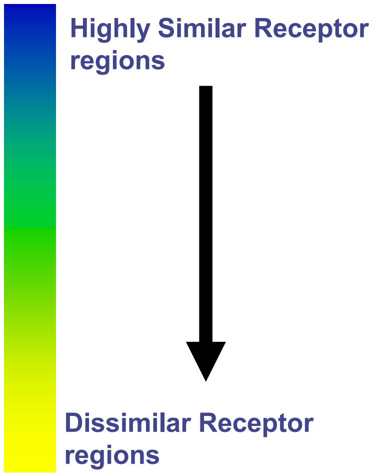


Pregnane X-receptor – PXR (“sensor)” →CYP3A4 (“executioner”)

<u>PXR Binds > 50% drugs</u>

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…

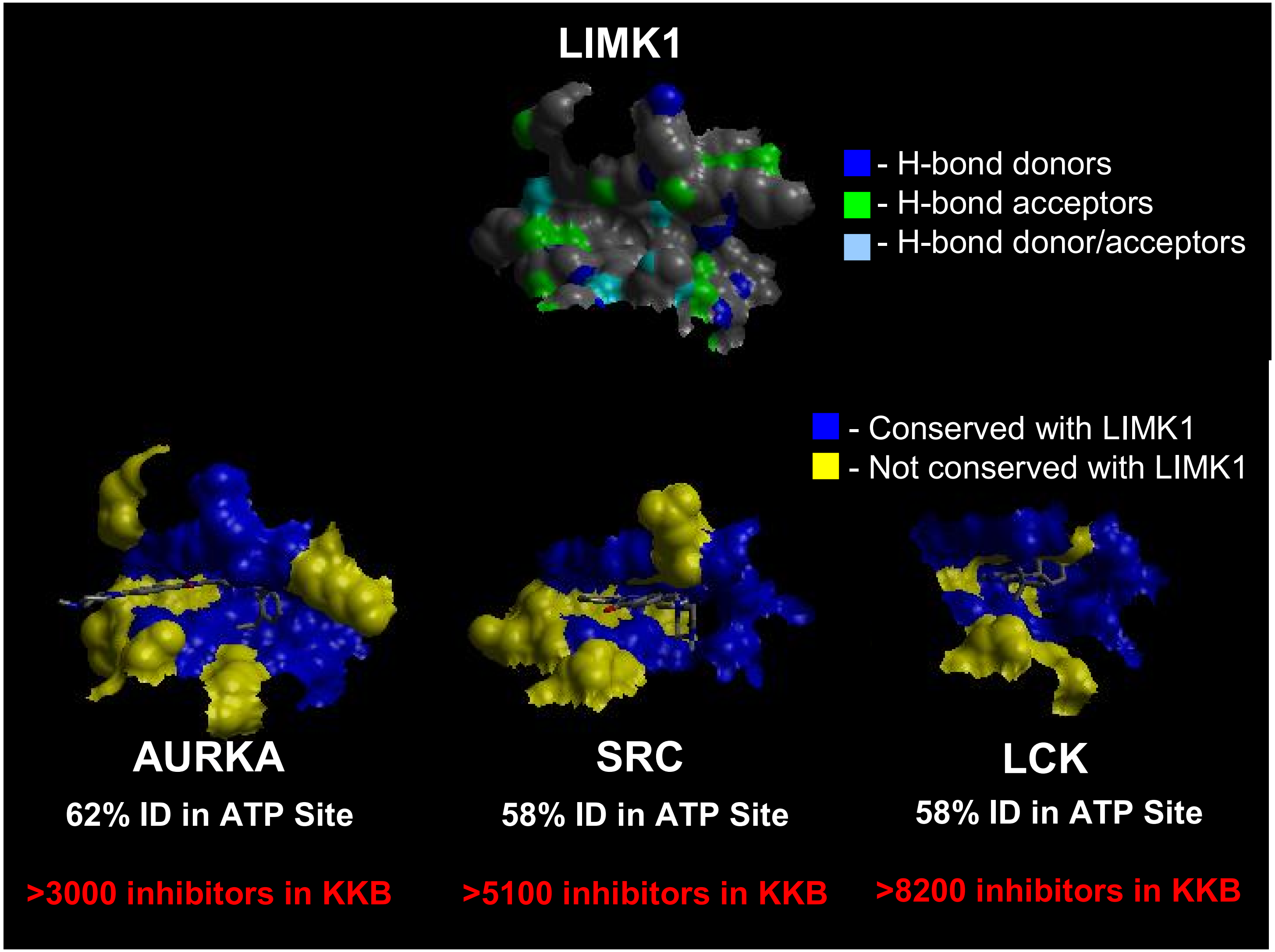
<u>Site Similarity Coloring</u>



Highly Similar Receptor regions

Dissimilar Receptor regions

LIMK1 – ATP binding site comparison



LIMK1

🟦 - H-bond donors
🟩 - H-bond acceptors
🟦 - H-bond donor/acceptors

🟦 - Conserved with LIMK1
🟨 - Not conserved with LIMK1

AURKA

62% ID in ATP Site

>3000 inhibitors in KKB

SRC

58% ID in ATP Site

>5100 inhibitors in KKB

LCK

58% ID in ATP Site

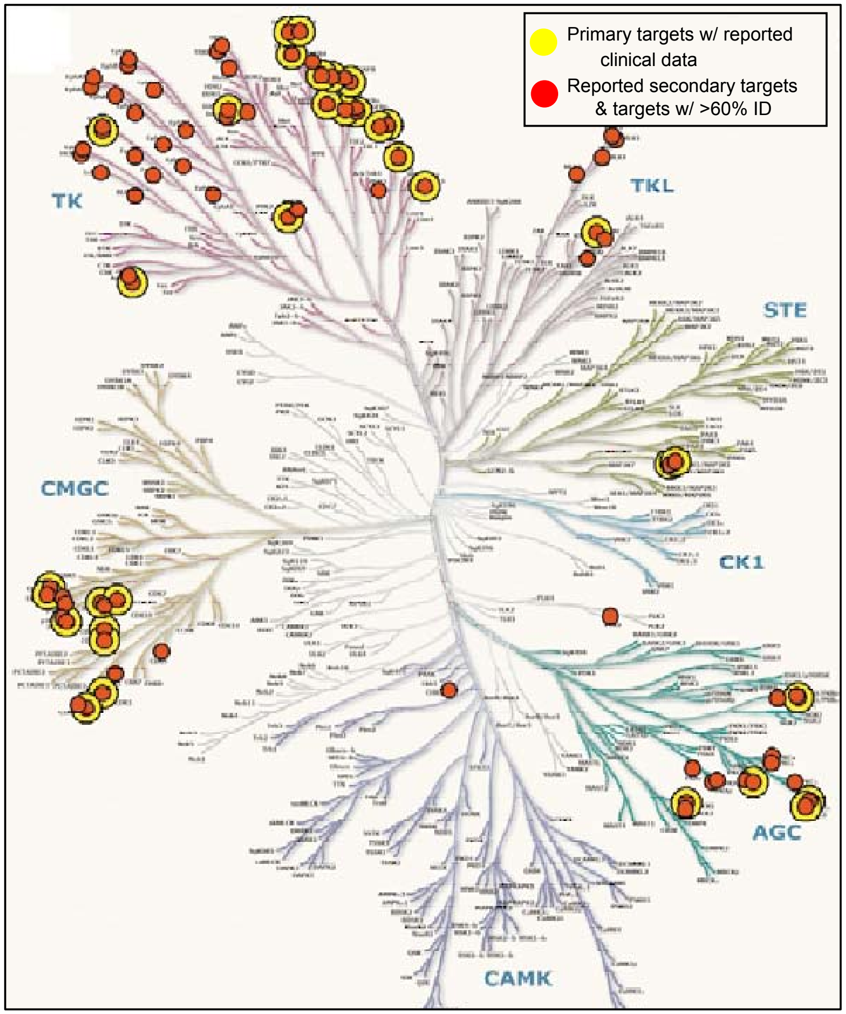
>8200 inhibitors in KKB

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

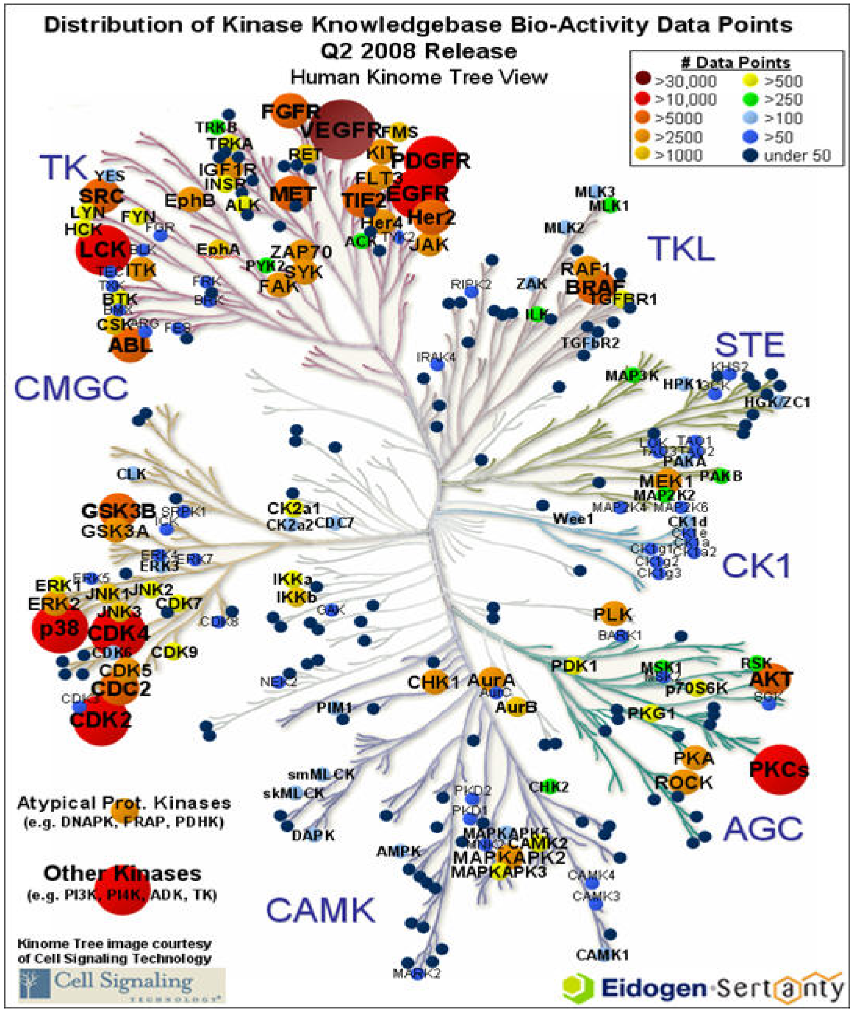
Kinase SAR Knowledgebase – Hot Targets

Kinase Targets of Clinical Interest

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



Eidogen-Sertanty KKB
SAR Data Point Distribution



>362,000 SAR data points curated from
>4,270 journal articles and patents
>130 Bayesian QSAR Models

Eidogen Sertanty

Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from

> 4278 journal articles/patents

- **KKB Content Summary (Q2-2008):**
  - # of kinase targets: **>390**
  - # of SAR Data points: **> 362,000**
  - # of **unique** kinase molecules with SAR data: **>120,000**
  - # of annotated assay protocols: **>16,000**
  - # of annotated chemical reactions: **>2,300**
  - # of unique kinase inhibitors: **>465,000** **(~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

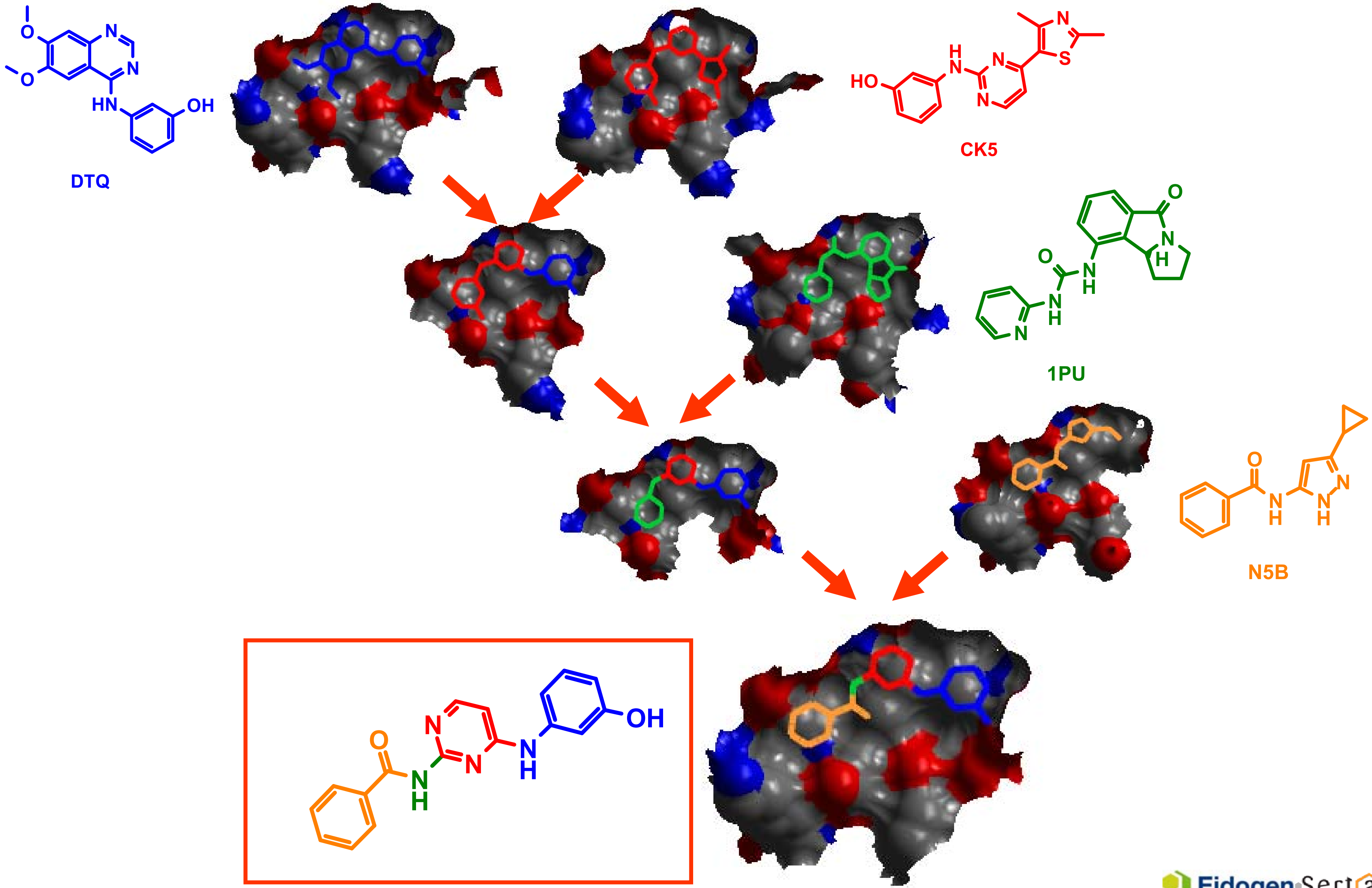
Kinase Validation Set

Three sizable datasets freely available to the research community

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

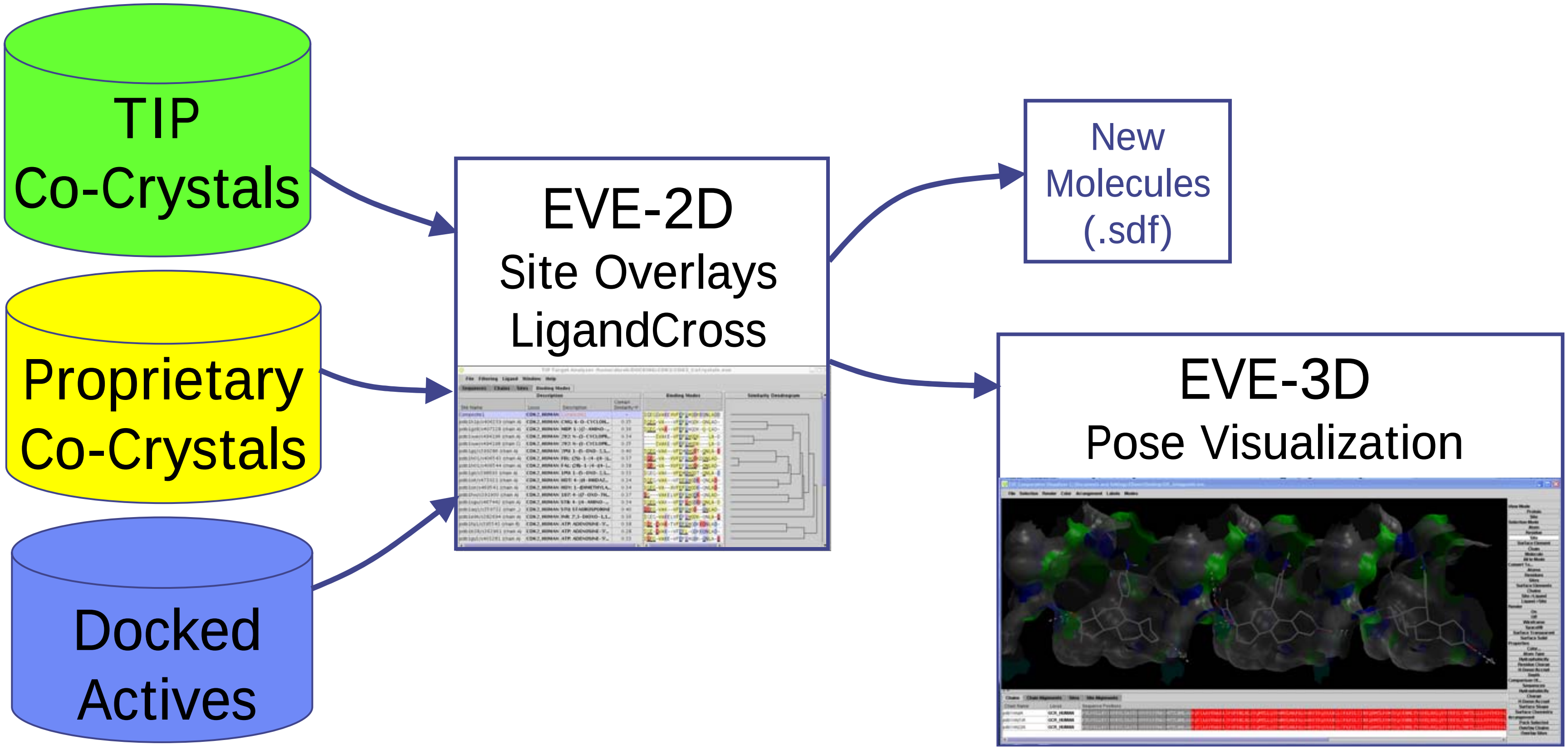
Lead Discovery: Knowledge-Based Design

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



Eidogen-Sertanty

LigandCross Workflow



TIP Co-Crystals

Proprietary Co-Crystals

Docked Actives

EVE-2D
Site Overlays
LigandCross

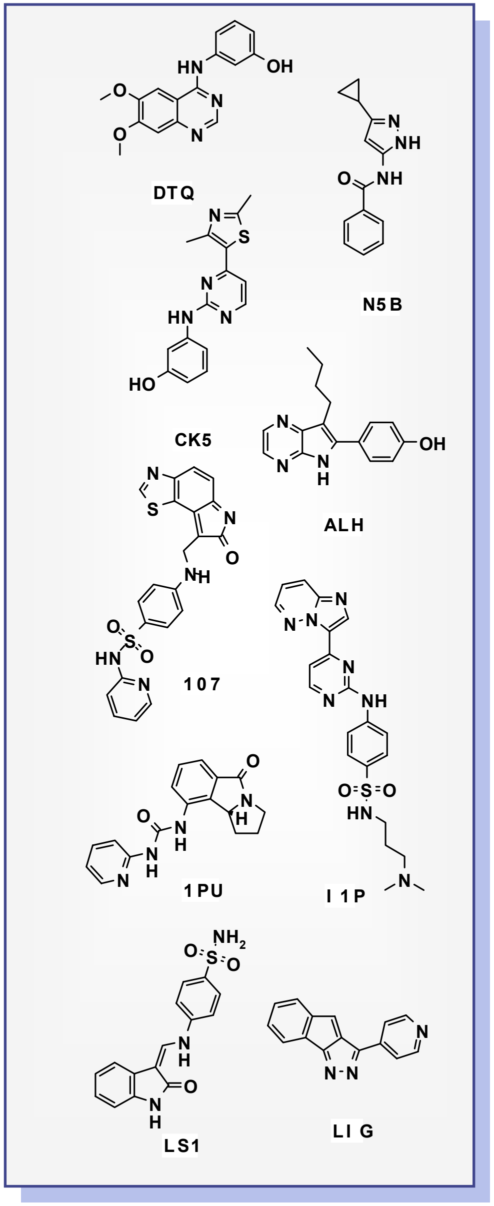
New Molecules (.sdf)

EVE-3D
Pose Visualization

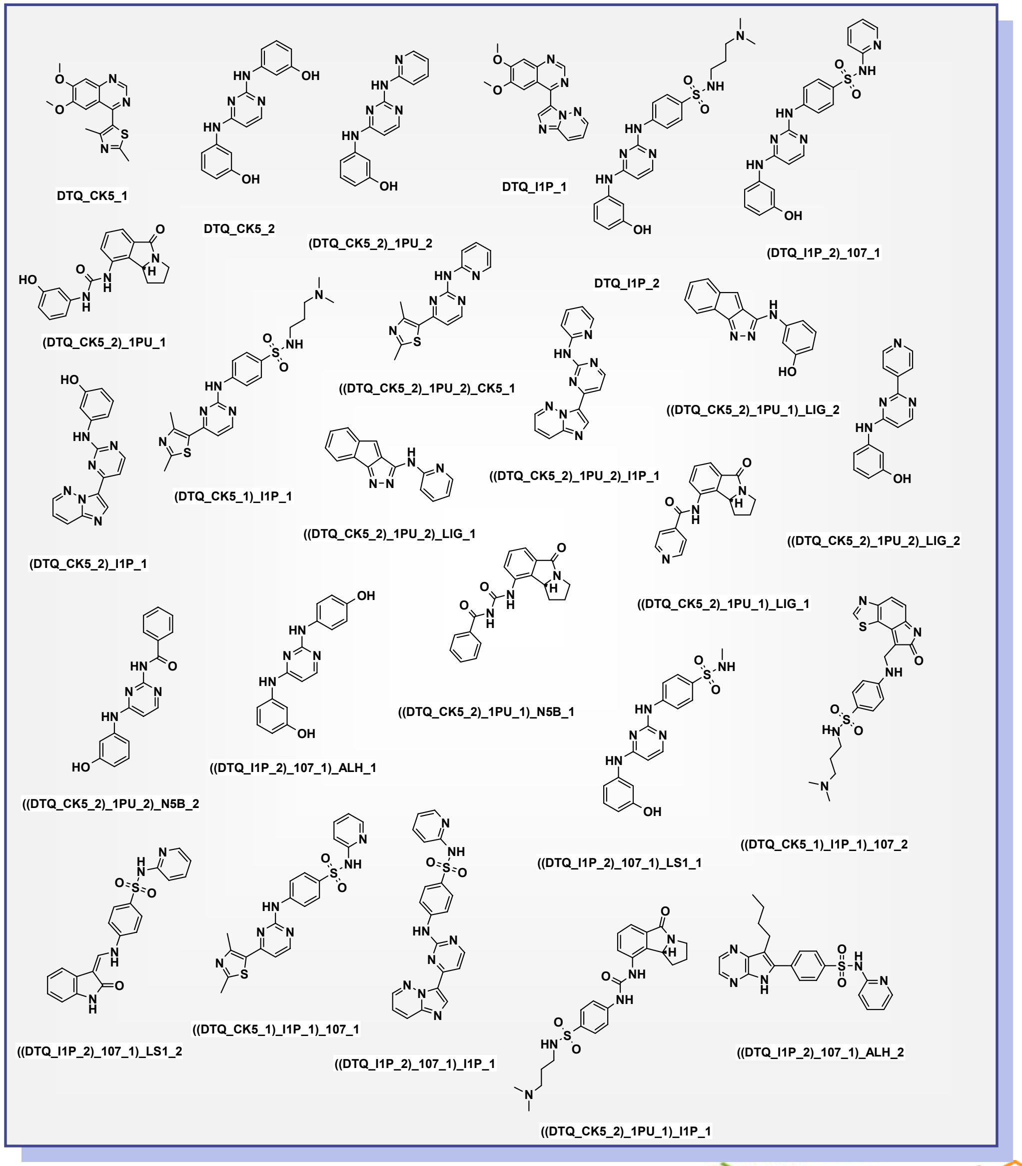
New Molecules via LigandCross

Novel Ligands via Ligand Crossover

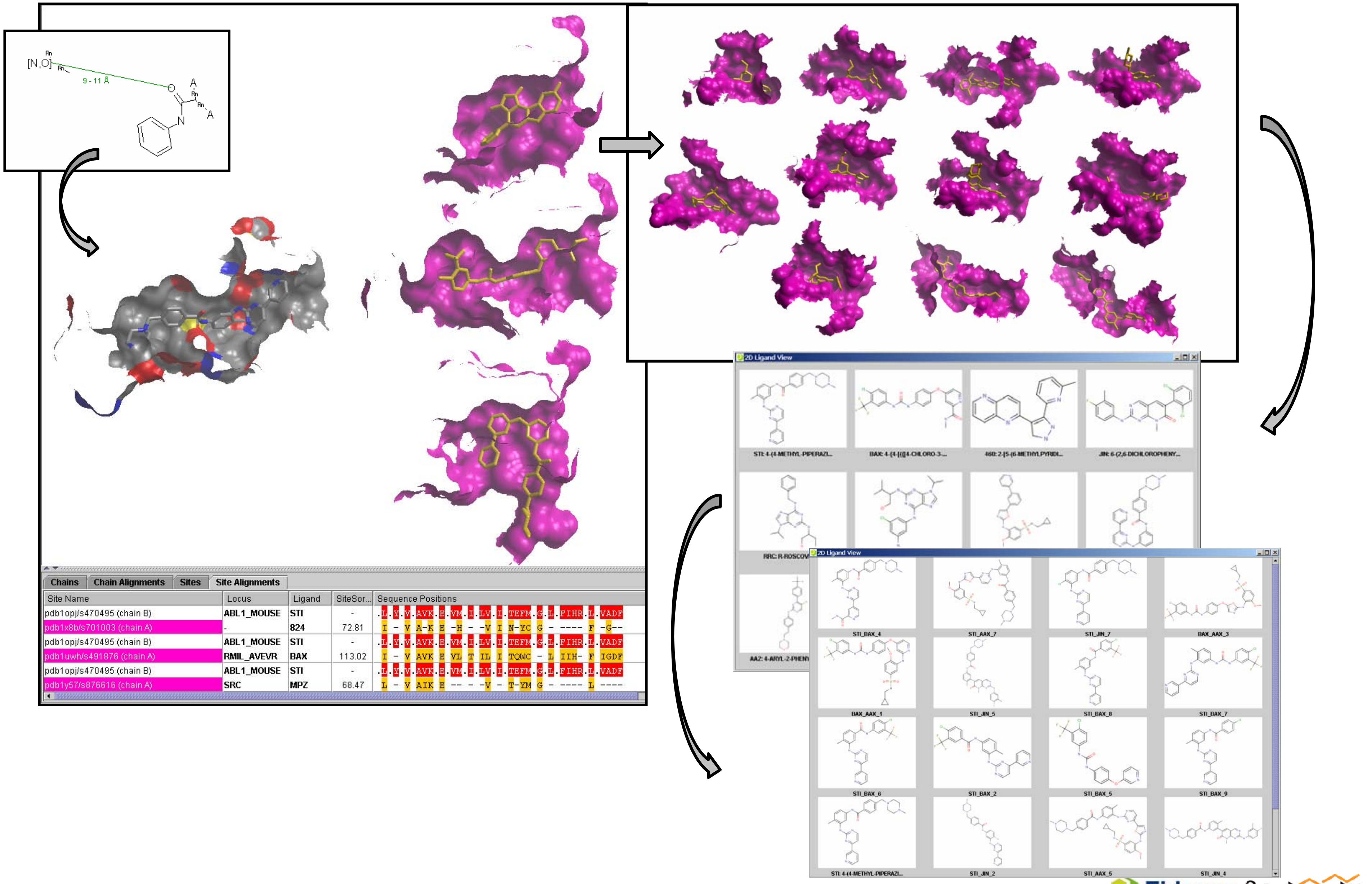
Starting ligands


DTQ
N5B
CK5
ALH
107
1PU
I1P
LS1
LIG

Hybridized product ligands

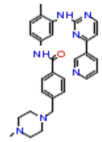
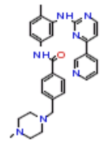
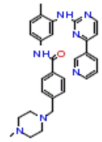

DTQ_CK5_1
DTQ_CK5_2
(DTQ_CK5_2)_1PU_2
DTQ_I1P_1
DTQ_I1P_2
(DTQ_I1P_2)_107_1
(DTQ_CK5_2)_1PU_1
((DTQ_CK5_2)_1PU_2)_CK5_1
((DTQ_CK5_2)_1PU_1)_LIG_2
(DTQ_CK5_1)_I1P_1
((DTQ_CK5_2)_1PU_2)_I1P_1
(DTQ_CK5_2)_I1P_1
((DTQ_CK5_2)_1PU_2)_LIG_1
((DTQ_CK5_2)_1PU_2)_LIG_2
((DTQ_CK5_2)_1PU_1)_LIG_1
((DTQ_CK5_2)_1PU_1)_N5B_1
((DTQ_I1P_2)_107_1)_ALH_1
((DTQ_CK5_2)_1PU_2)_N5B_2
((DTQ_CK5_1)_I1P_1)_107_2
((DTQ_I1P_2)_107_1)_LS1_1
((DTQ_CK5_1)_I1P_1)_107_1
((DTQ_I1P_2)_107_1)_LS1_2
((DTQ_I1P_2)_107_1)_I1P_1
((DTQ_I1P_2)_107_1)_ALH_2
((DTQ_CK5_2)_1PU_1)_I1P_1

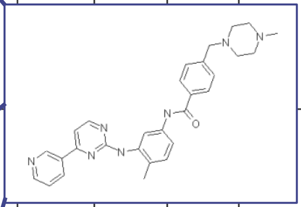
From Ligand Query to Sites to New Ligand Ideas

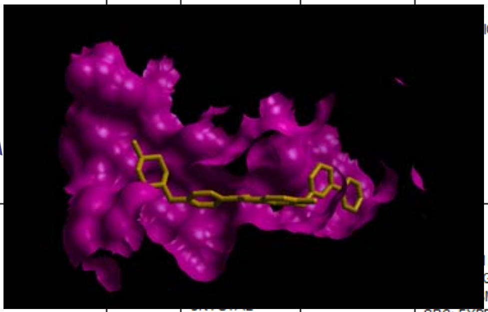


Site Name	Locus	Ligand	SiteSor...	Sequence Positions
pdb1opj/s470495 (chain B)	ABL1_MOUSE	STI	-	.L.Y.V.AVK.E.VM.I.LV.I.TEFM.G.L.FIHR.L.VADF
pdb1x8b/s701003 (chain A)	-	824	72.81	I - V A-K E -H - -V I N-YC G - ---- F -G--
pdb1opj/s470495 (chain B)	ABL1_MOUSE	STI	-	.L.Y.V.AVK.E.VM.I.LV.I.TEFM.G.L.FIHR.L.VADF
pdb1uwh/s491876 (chain A)	RMIL_AVEVR	BAX	113.02	I - V AVK E VL T IL I TQWC - L IIH- F IGDF
pdb1opj/s470495 (chain B)	ABL1_MOUSE	STI	-	.L.Y.V.AVK.E.VM.I.LV.I.TEFM.G.L.FIHR.L.VADF
pdb1y57/s876616 (chain A)	SRC	MPZ	68.47	L - V AIK E -- - -V - T-YM G - ---- L ----

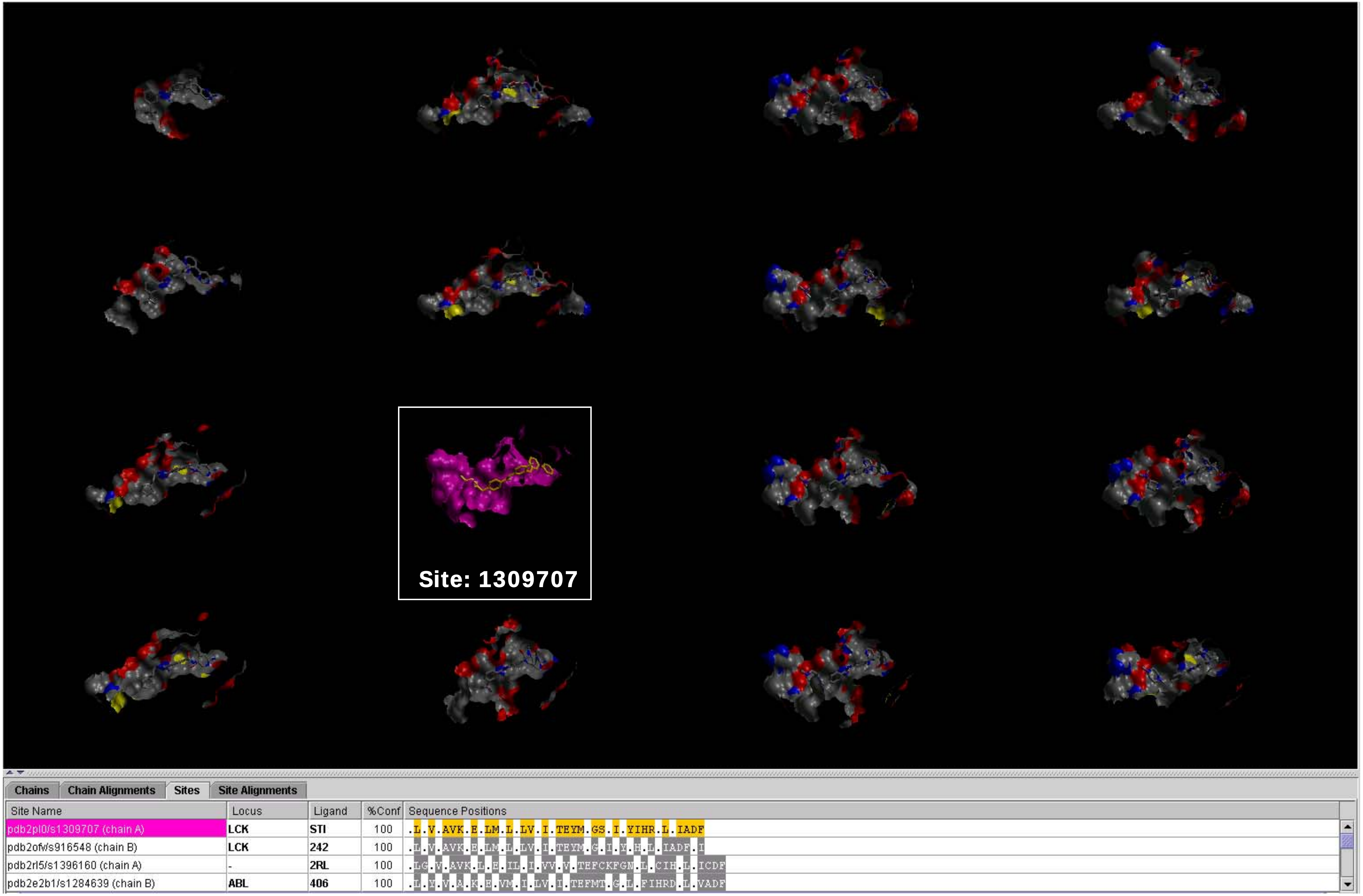
Step 1: Find Co-complexes and Sites from Ligand-Structure-Search

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: [illegible] 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	[illegible] STRUCTURE OF CHICKEN C-SRC KINASE DOMAIN IN COMPLEX WITH THE CANCER DRUG IMATINIB.	TRANSFERASE	[illegible]; ORGANISM_SCIENTIFIC: [illegible] GALLUS; [illegible] 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





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



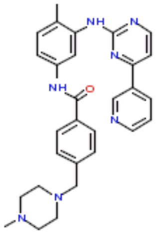
Site: 1309707

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
pdb2ofv/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.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

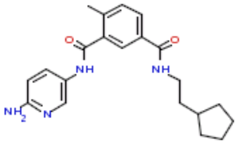
Example Site Similarity Results (Query: s1309707)

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

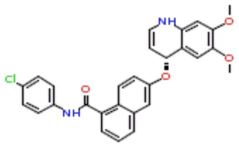
Example Ligands Extracted from Similar Sites



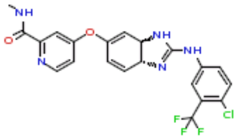
STI



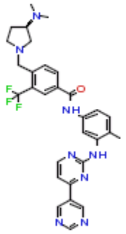
C92



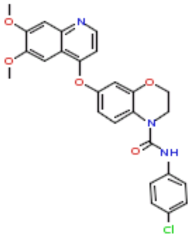
900



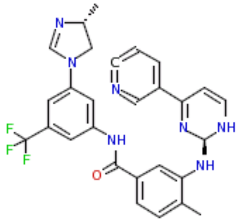
276



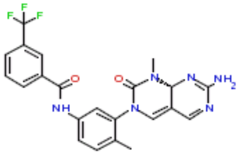
406



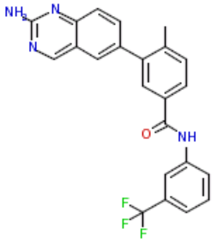
2RL



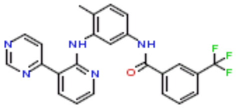
NIL



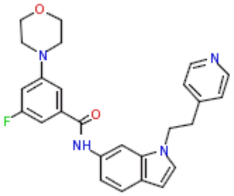
7MP



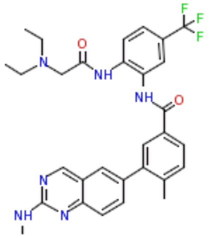
242



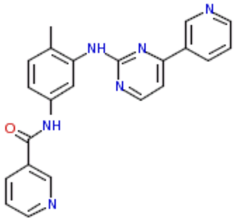
MUH



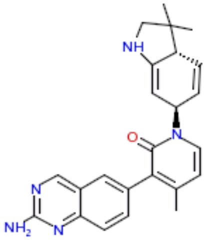
WBT



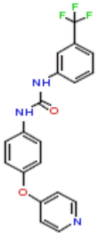
1N8



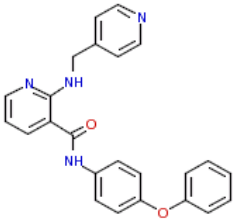
PRC



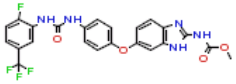
C19



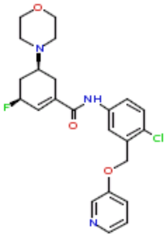
KIN



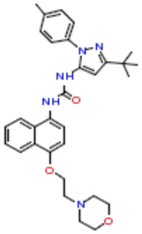
608



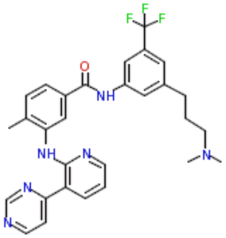
GIG



L11

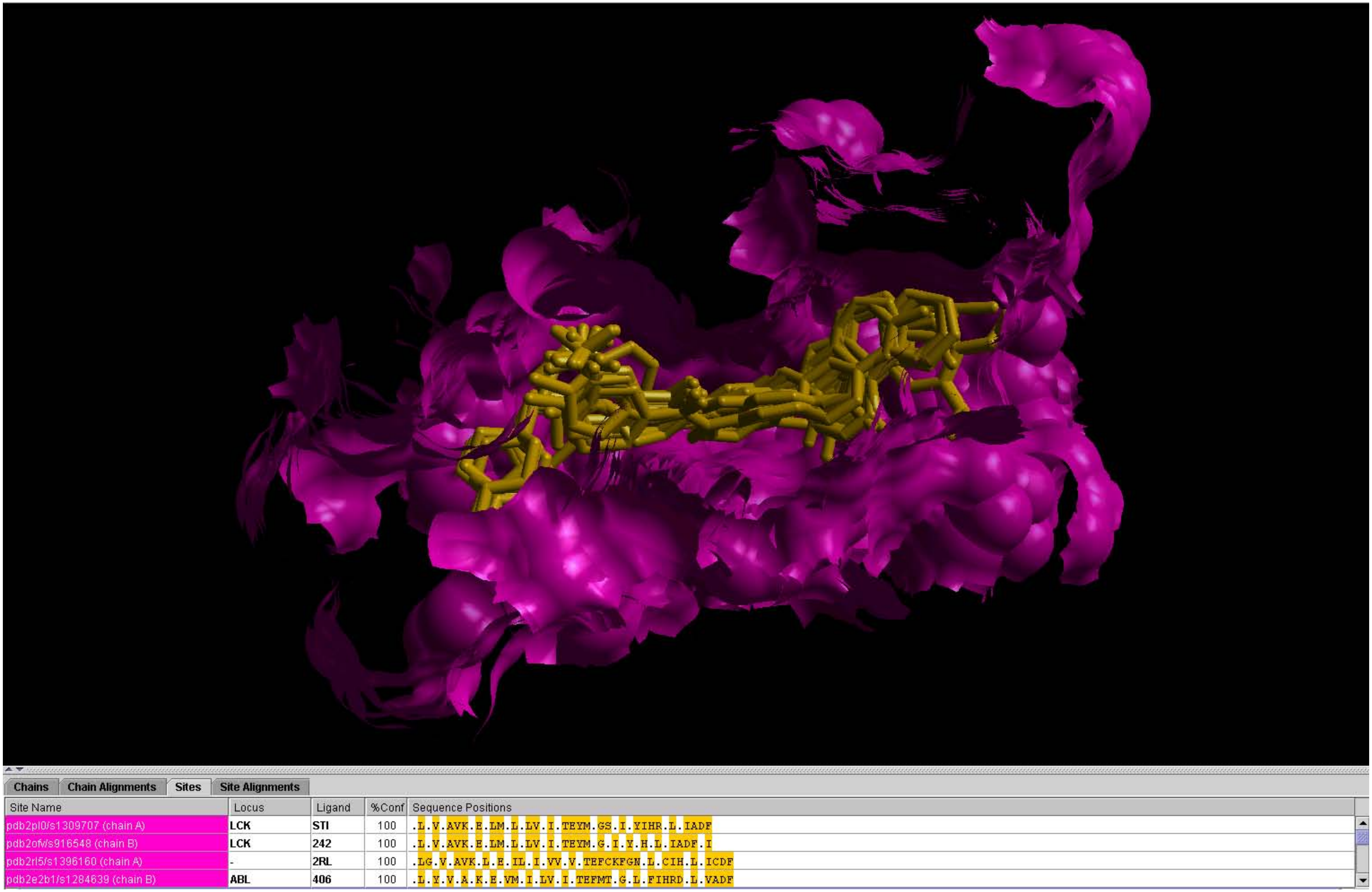


B96



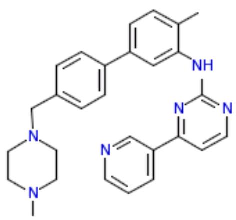
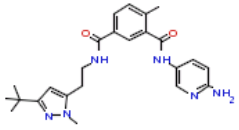
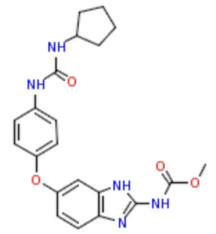
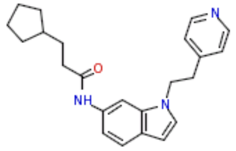
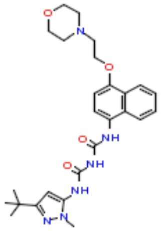
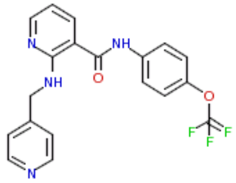
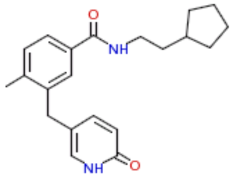
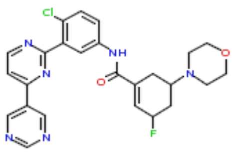
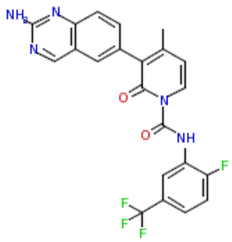
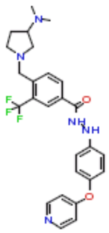
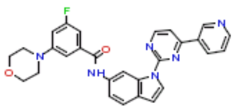
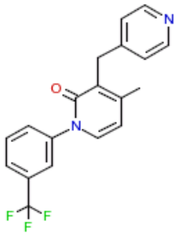
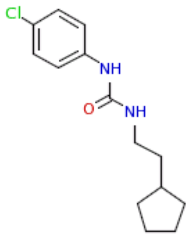
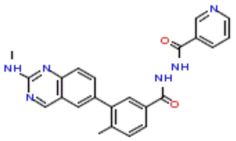
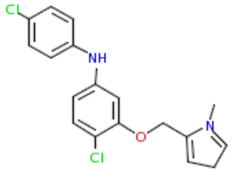
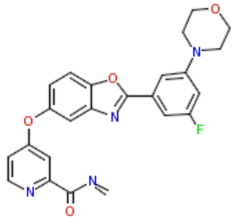
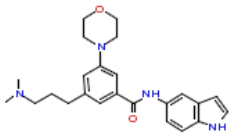
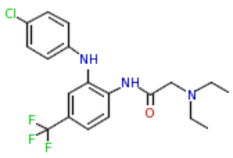
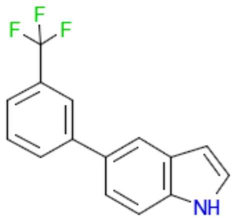
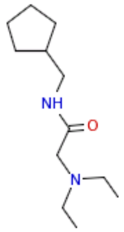
RAJ

Step 3: LigandCross – Mixing Ligand Features from Aligned Sites

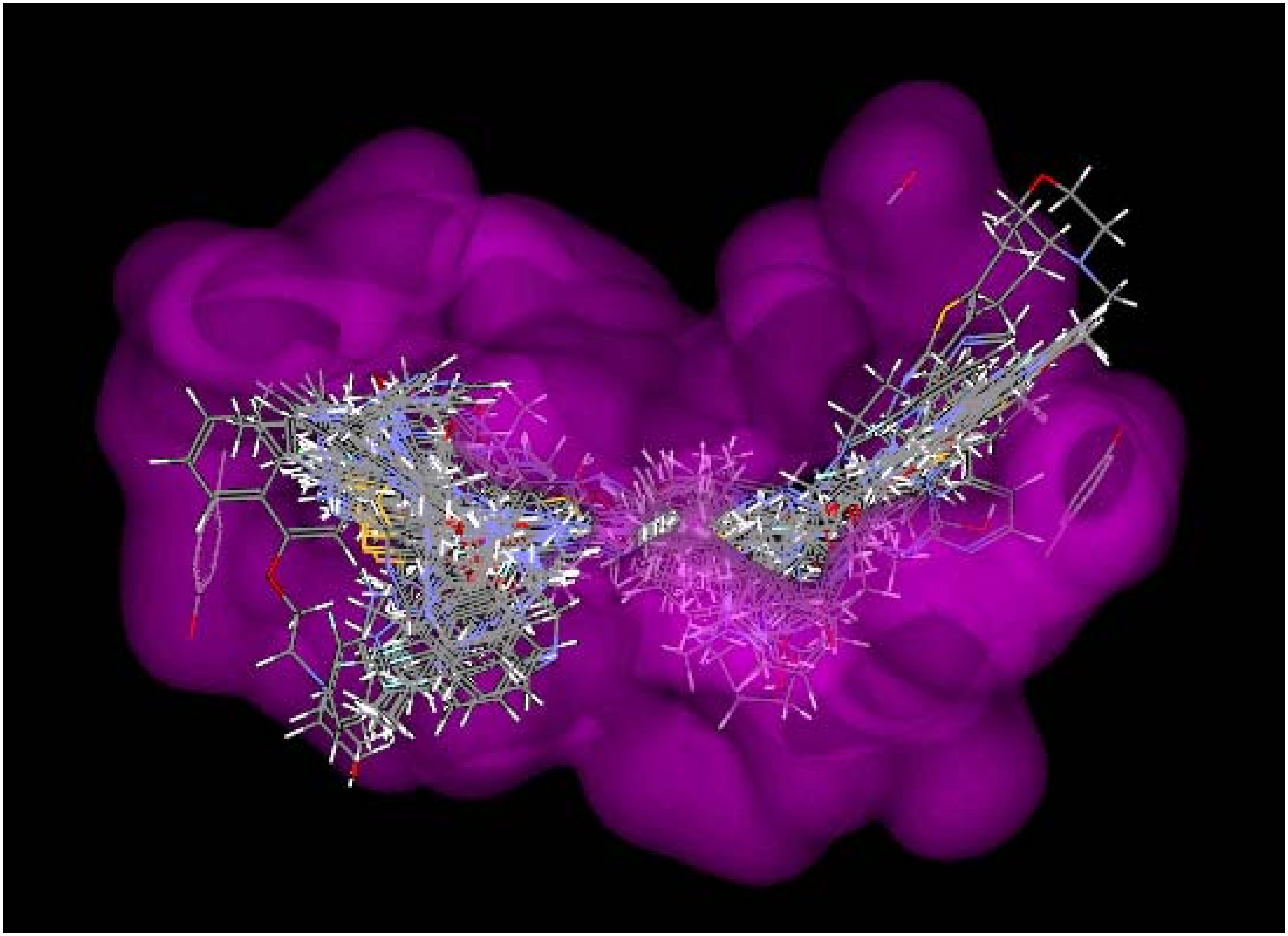


Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2pl0/s1309707 (chain A)	**LCK**	**STI**	100	.L.V.AVK.E.LM.L.LV.I.TEYM.GS.I.YIHR.L.IADF
pdb2ofv/s916548 (chain B)	**LCK**	**242**	100	.L.V.AVK.E.LM.L.LV.I.TEYM.G.I.Y.H.L.IADF.I
pdb2rl5/s1396160 (chain A)	-	**2RL**	100	.LG.V.AVK.L.E.IL.I.VV.V.TEFCKFGN.L.CIH.L.ICDF
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

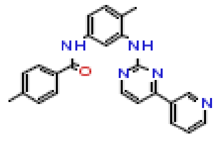
 STI_PRC_2 0.667	 C92_BMU_5 0.635	 C92_GIG_3 0.633	 C92_WBT_1 0.625	 B96_BMU_2 0.623
 608_276_3 0.608	 C92_GIN_7 0.608	 406_L11_6 0.577	 GIG_C52_1 0.574	 406_KIN_2 0.545
 NIL_WBT_6 0.538	 608_C52_2 0.529	 C92_BMU_1 0.520	 1N8_PRC_3 0.491	 857_BMU_4 0.480
 857_WBT_2 0.472	 RAJ_LI3_1 0.462	 1N8_BMU_2 0.449	 LI3_C52_2 0.385	 C92_1N8_1 0.375

Step 4: LigandCross Ligands reDocked into s1309707

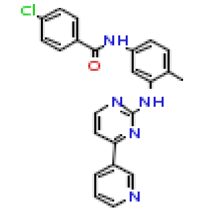


LigandCross Ligands with Reported Biological Activity

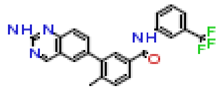
	Kinase Knowledgebase (pIC50)										Bayesian Model Predictions (PP)									
LC-ID	ABL	PDGFR	PDGFRB	JAK3	KDR	LCK	MAPK14	TEK	KIT	RAF1	ABL	PDGFR	PDGFRB	JAK3	KDR	LCK	MAPK14	TEK	KIT	RAF1
G2G_STI_12	6.7	8	8								0.40	0.90	0.76	0.81	0.59	0.15	0.89	0.45	0.70	0.37
900_STI_1	6.1	8	8								0.38	0.91	0.76	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



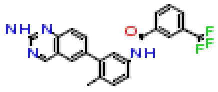
1: G2G_STI_12



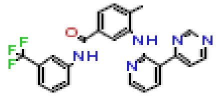
2: 900_STI_1



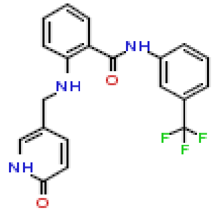
3: 7MP_1N8_4



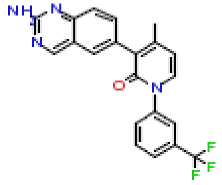
4: 7MP_1N8_2



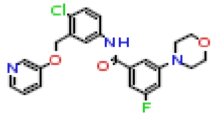
5: 7MP_RAJ_3



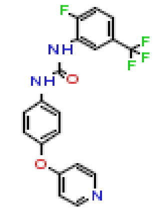
6: 7MP_GIN_4



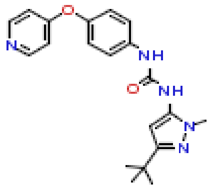
7: 242_C52_2



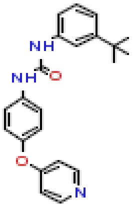
8: LI3_L11_1



9: 608_GIG_7



10: KIN_BMU_4



11: G2G_KIN_3

Eidogen·Sertanty

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 strategy to generate novel, bioactive molecules from co-complex information.

Acknowledgements

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

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

- ChIP: National Institute of Standards and Technology (NIST) –

ATP program: ‘Chemical Intelligence Platform for Rapid Discovery of DrugLeads’

Contact
Steven Muskal
Chief Executive Officer
smuskal@eidogen-sertanty.com

Eidogen·Sertanty