Exercising receptor-site similarity:

From Off-Target Identification to Scaffold Hopping

Steven Muskal, Ph.D.

Chief Executive Officer

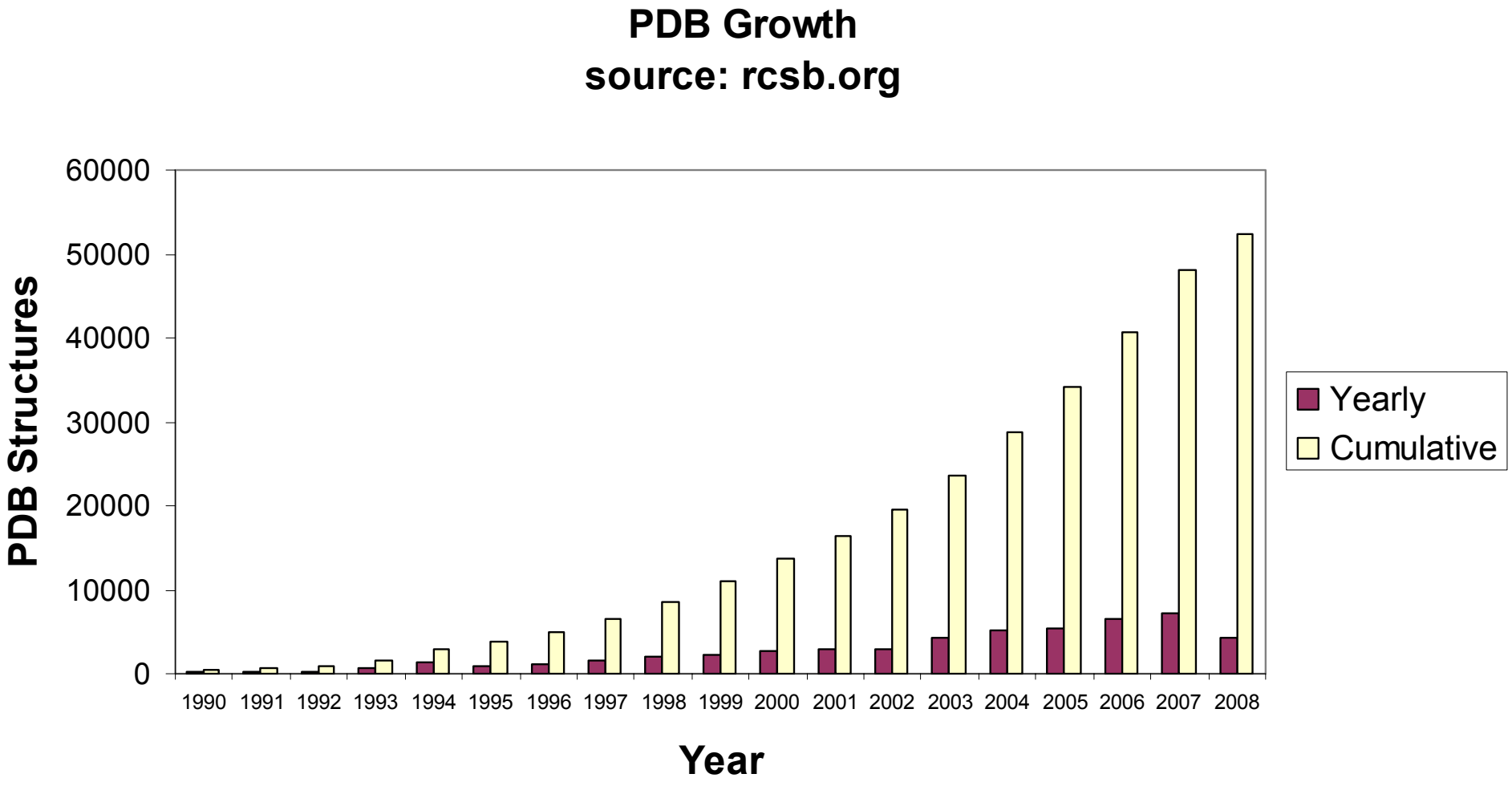
Eidogen-Sertanty, Inc.

smuskal@eidogen-sertanty.com

Eidogen·Sertanty

Protein Structure Growth is Accelerating

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



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

Industrializing an Information Rich Craft

- Pharma cost reductions (reductions in jobs, spending, etc.)

  → Fewer IT specialists with less resource, supporting more people

- One-at-a-time computational efforts are bottlenecks

  → Many proteomic riches remain untapped

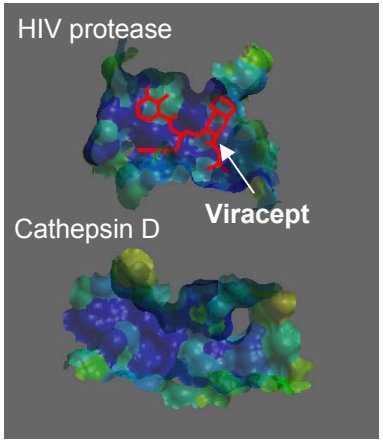
Non-specialist driven workflow-apps have become necessary!

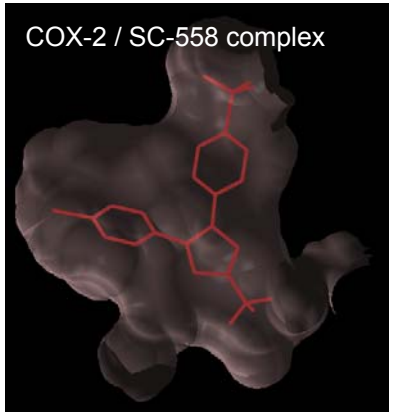
Bringing Proteomic Riches to Non-Specialists

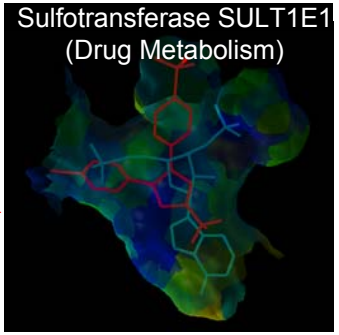
Automated Modeling and Proteomic Structural Mining

- ➢Sequence-to-Structure Calculation Cascade
- ➢Search-by: KeyWord, Sequence, Ligand, Protein Structure, Receptor-Sites, etc.
- ➢Exploit Structural fold and receptor-site conservation

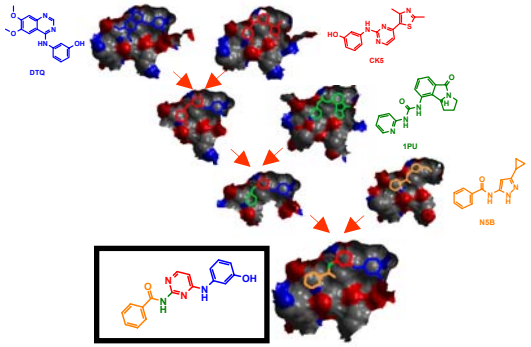
→Off-Target Identification (opportunities v. liabilities)







→Borrowing Matter Ideas from co-complexes and protein structures



Eidogen-Sertanty

About Eidogen-Sertanty

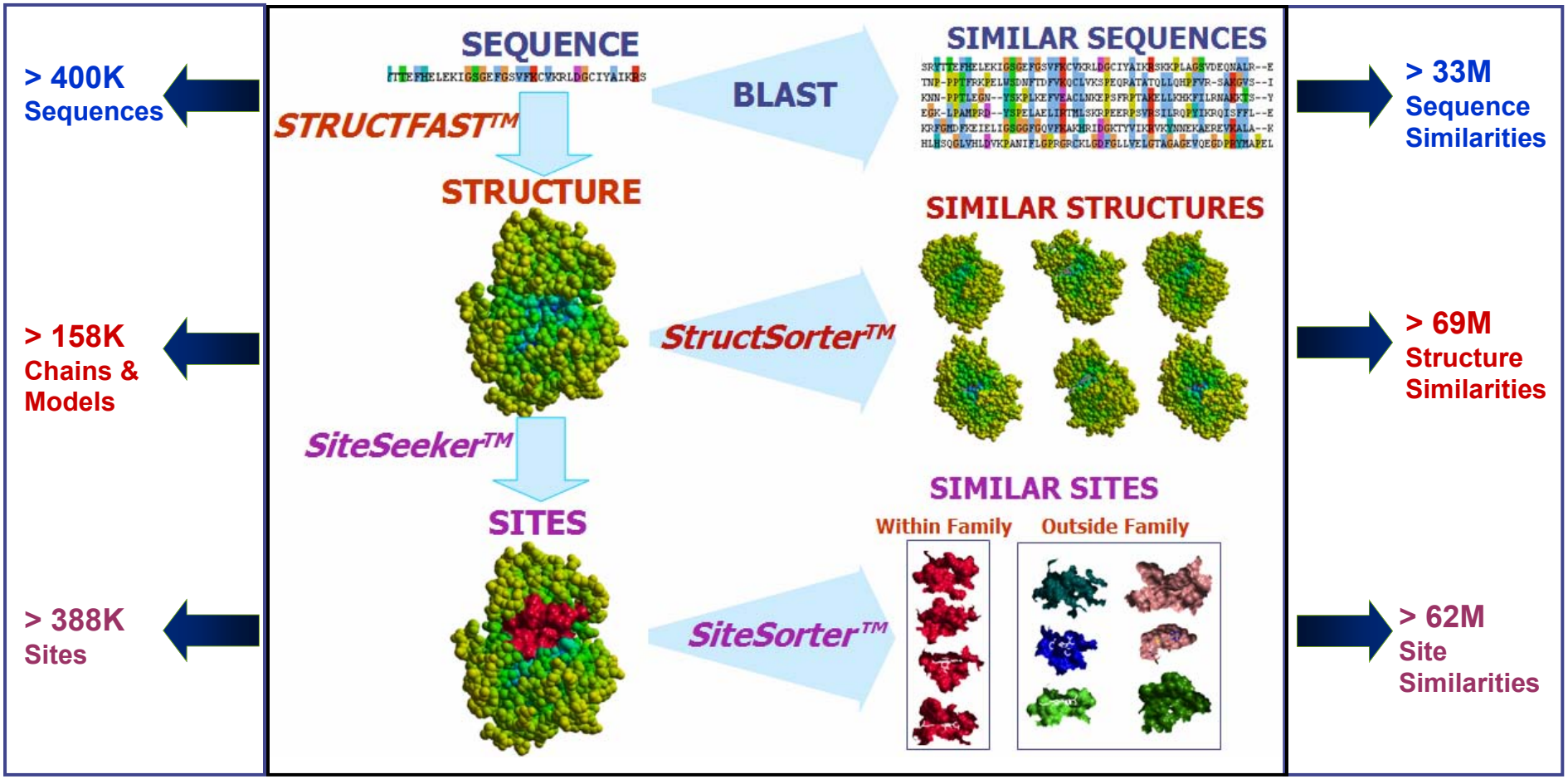
- Knowledge-Driven Discovery Solutions Provider
  - Formed in March 2005 when Sertanty (Libraria→Sertanty 2003) acquired Eidogen (Bionomix 2000)
  - >$20M Invested in Technology Development
  - 12 FTE's
  - Worldwide Customerbase
  - Cash-Positive

- 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

Eidogen·Sertanty

TIP Algorithm Engine

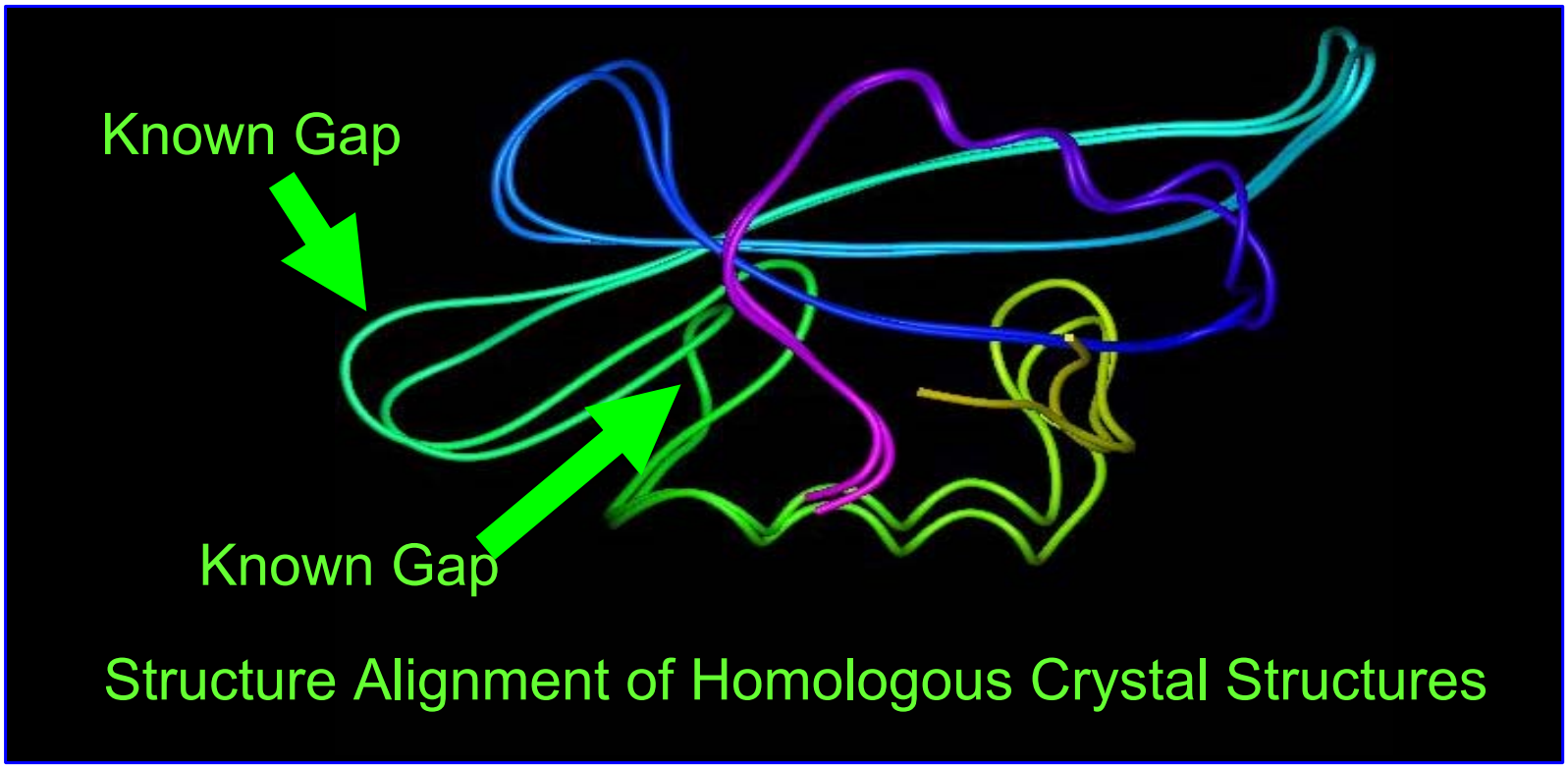


Eidogen-Sertanty

STRUCTFAST™

ST ructure 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

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

Eidogen-Sertanty

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

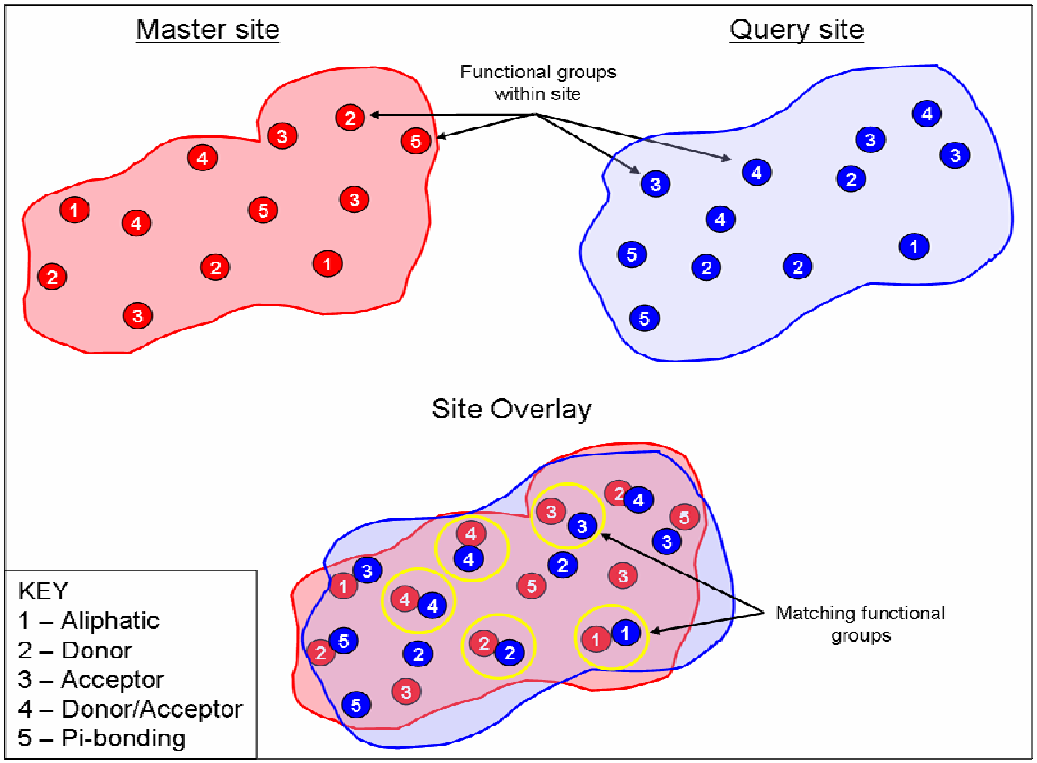
We use proteins with apo- & co-crystal structures in the PDB to test the accuracy & reliability of method

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

SiteSorter™

Weighted Clique Detection Algorithm

Importance of Points Related To Conservation In Multiple Sequence Alignment



Surface Atoms Assigned One of 5 Different Chemical Characters

Matching points increase the *SiteSorter* similarity score

Eidogen-Sertanty

TIP Content

>75,000 Human Sequences

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

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

>33M Sequence Similarities

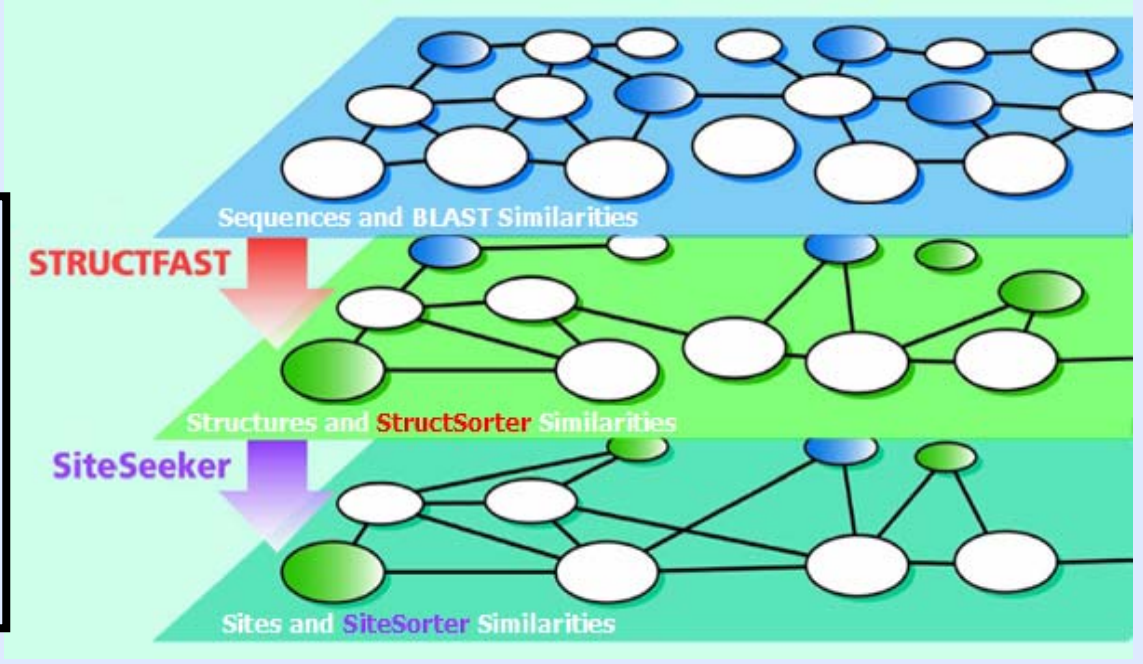
>69M Structural Similarities

>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

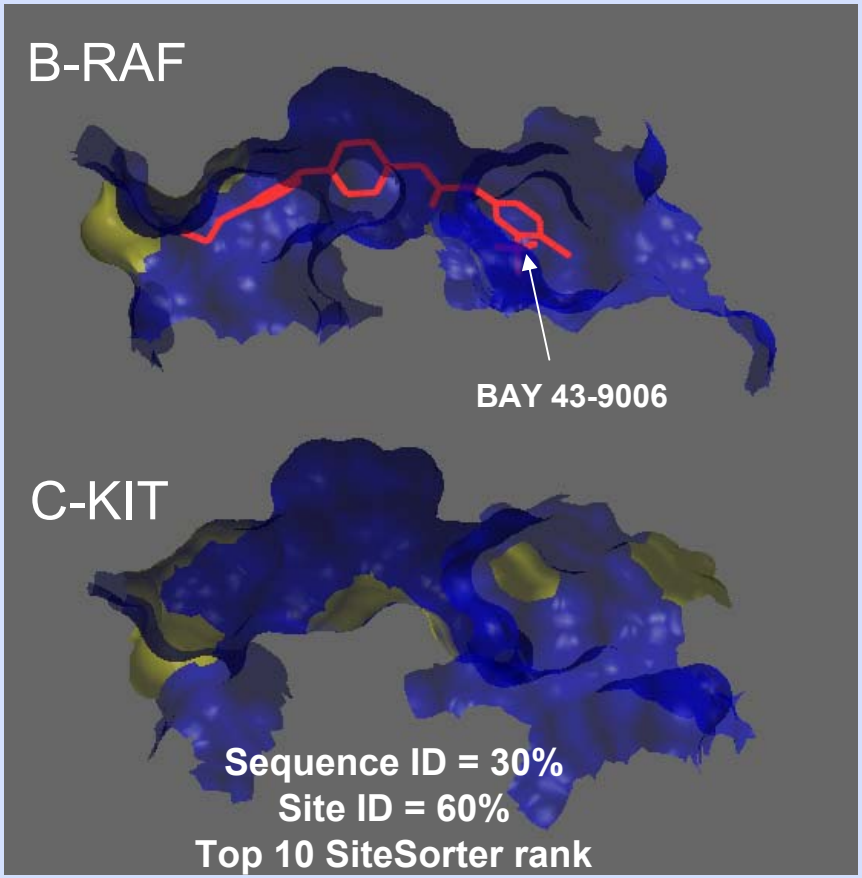


Automatically updated with new models as the PDB grows

Eidogen-Sertanty

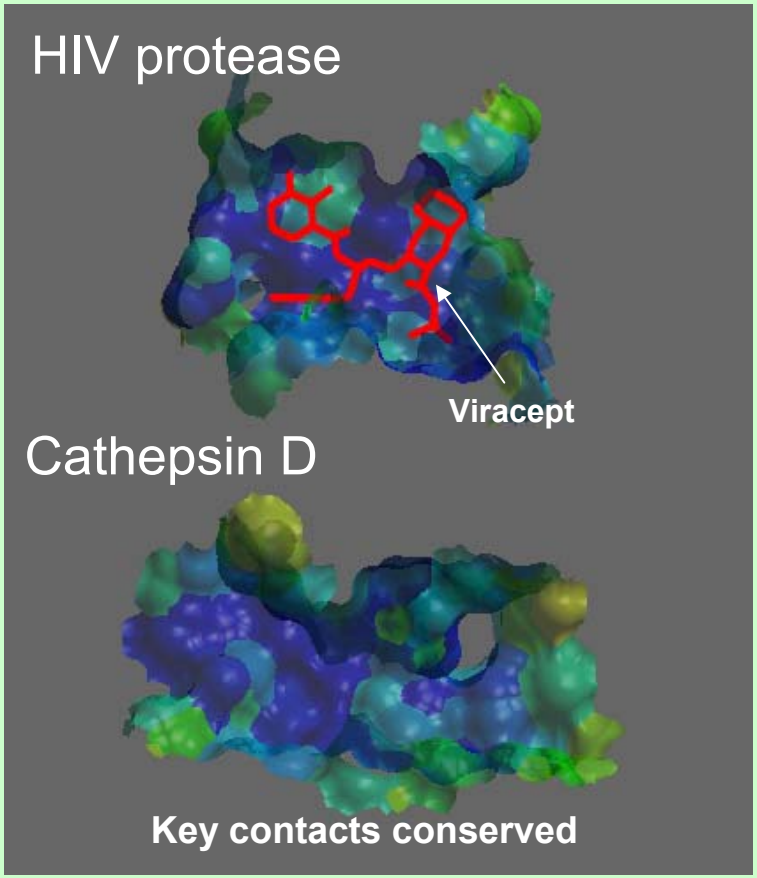
Off-Target Opportunities

Intra-Family Opportunities



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

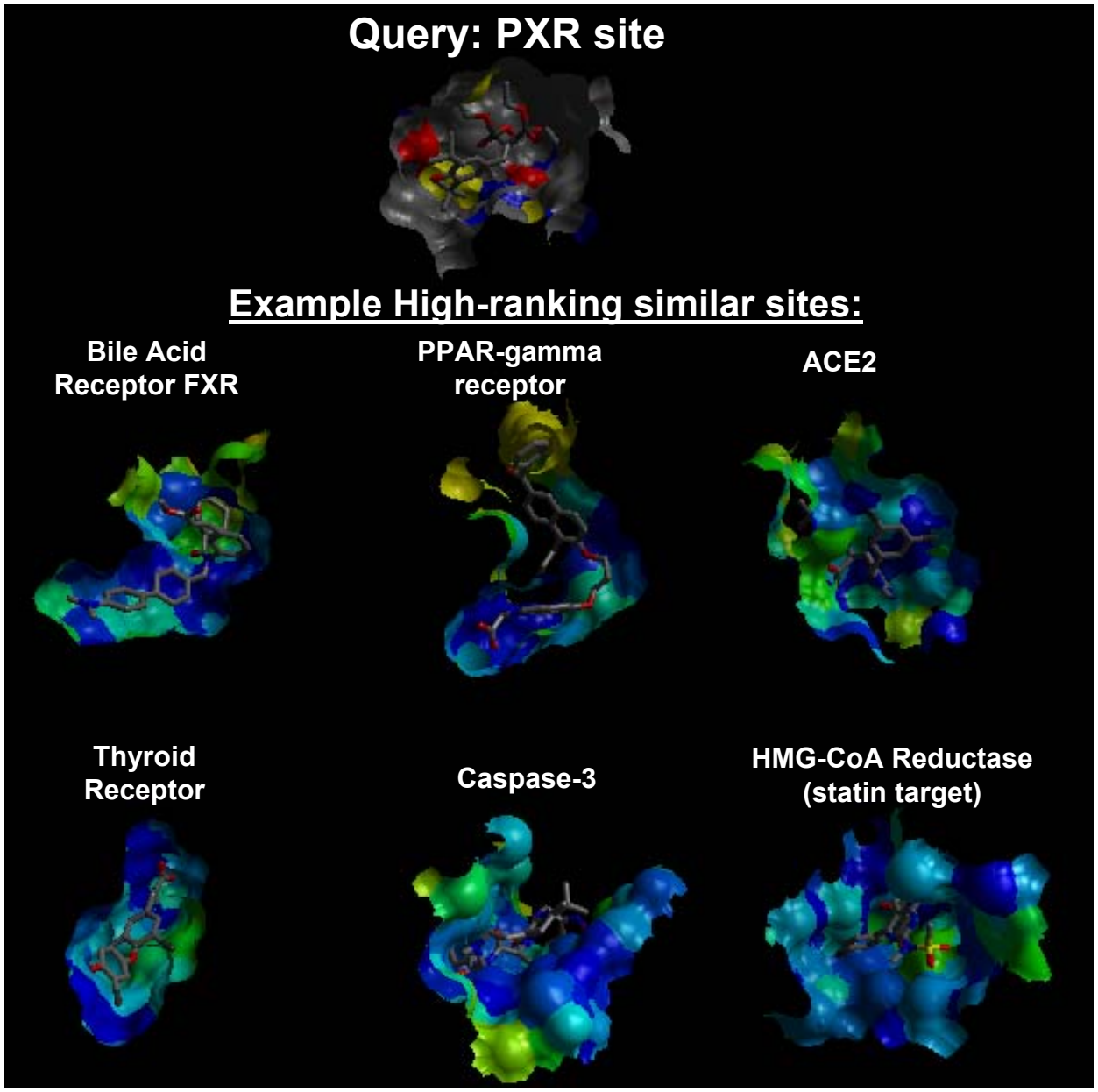
Inter-family Opportunities



Cathepsin D is inhibited by HIV protease inhibitors

Eidogen-Sertanty

PXR – Promiscuous Ligand-Binding Site

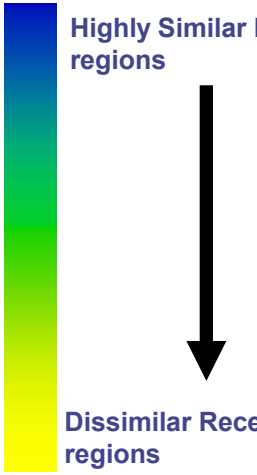


Pregnane X-receptor – PXR ("sensor)" →CYP3A4 ("executioner")

PXR Binds > 50% drugs

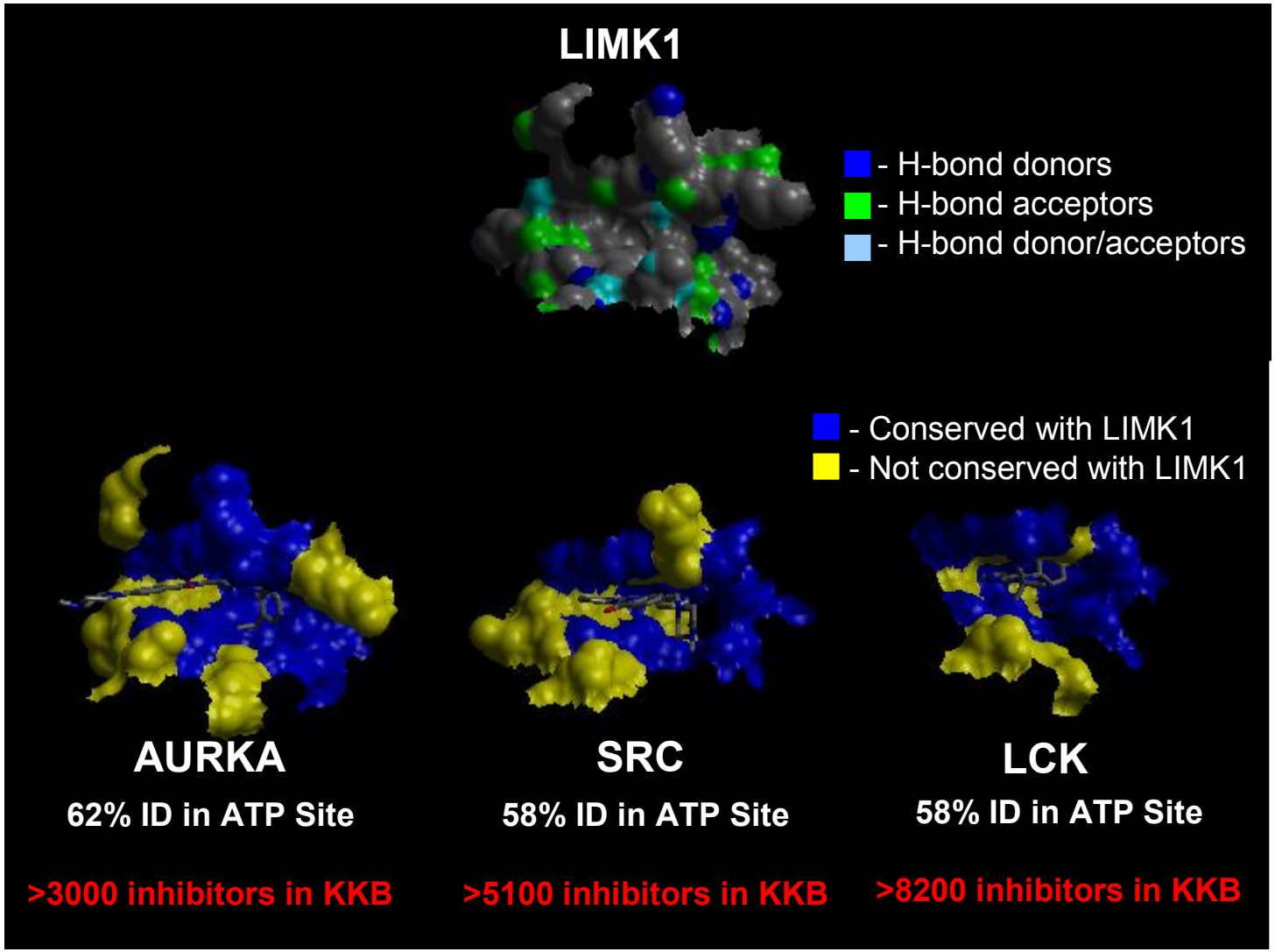
Including some bile acids, statins, herbal components, a selection of HIV protease inhibitors, calcium channel modulators, numerous steroids, plasticizers and monomers, organochlorine pesticides, a peroxisome proliferator-activated receptor-ãantagonist, xenobiotics and endobiotics…

Site Similarity Coloring



Eidogen-Sertanty

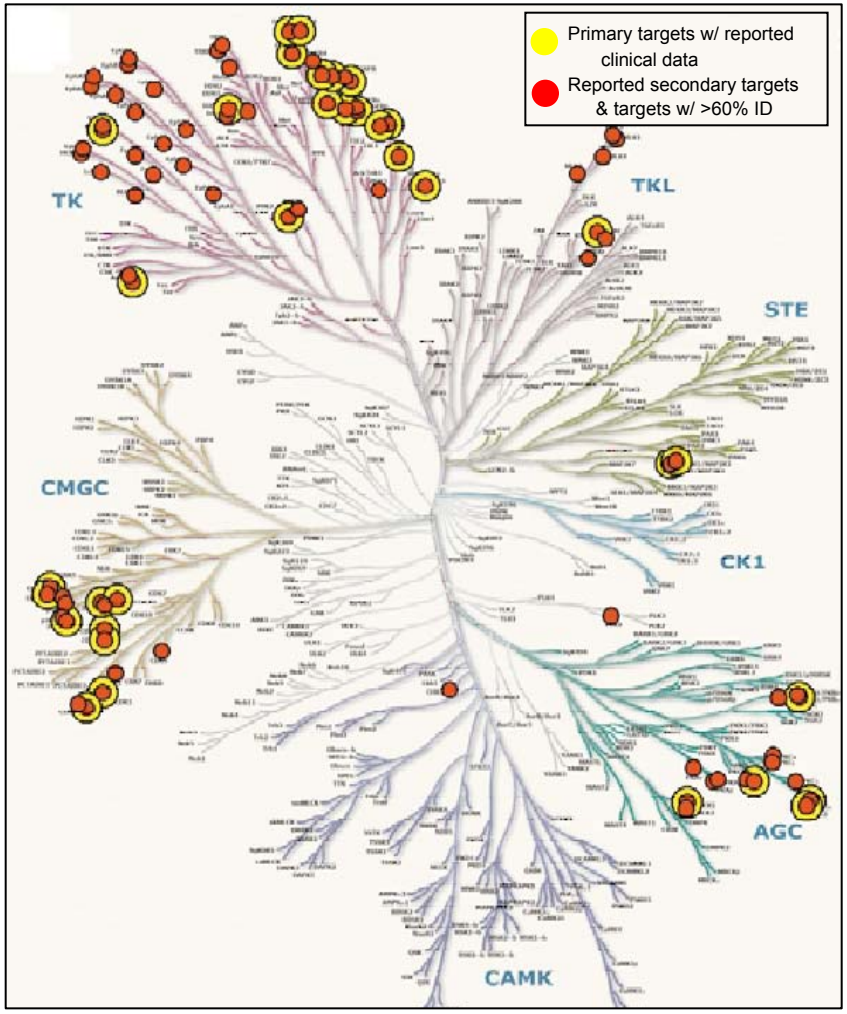
LIMK1 – ATP binding site comparison



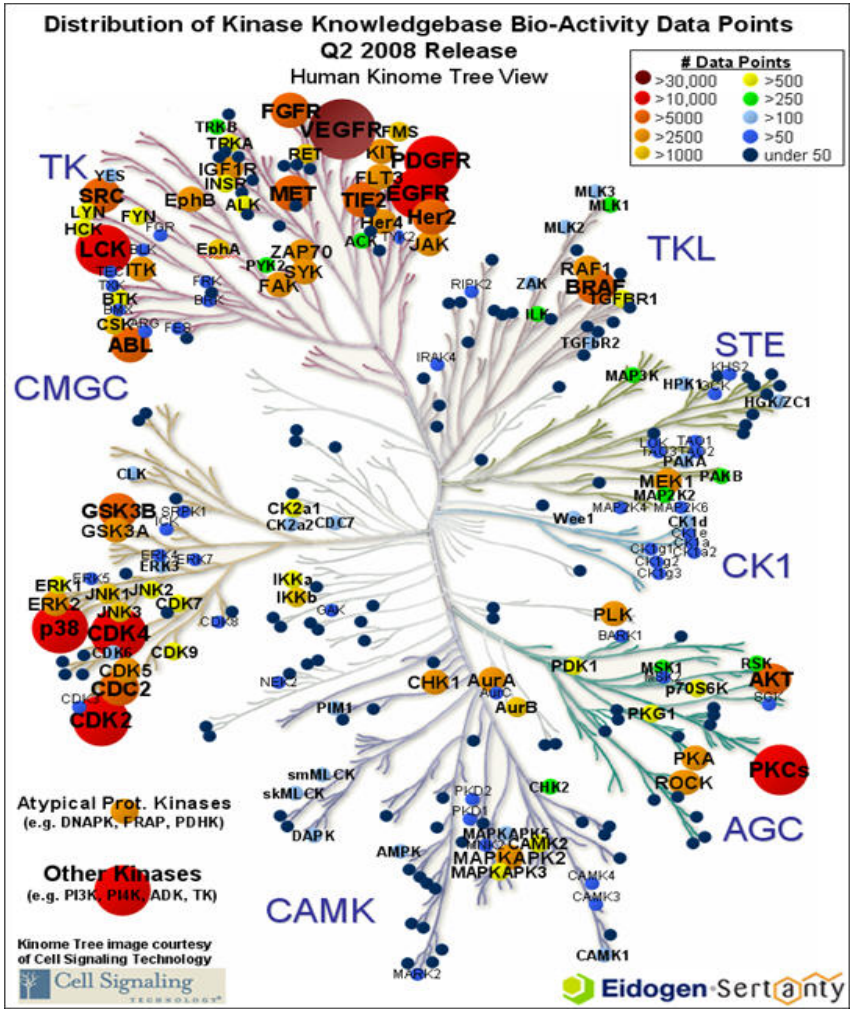
The ATP site of LIMK1 shares a high level of homology with several well-studied kinases

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

Eidogen·Sertanty

Kinase Knowledgebase (KKB)

Kinase inhibitor structures and SAR data mined from

Kinase Validation Set

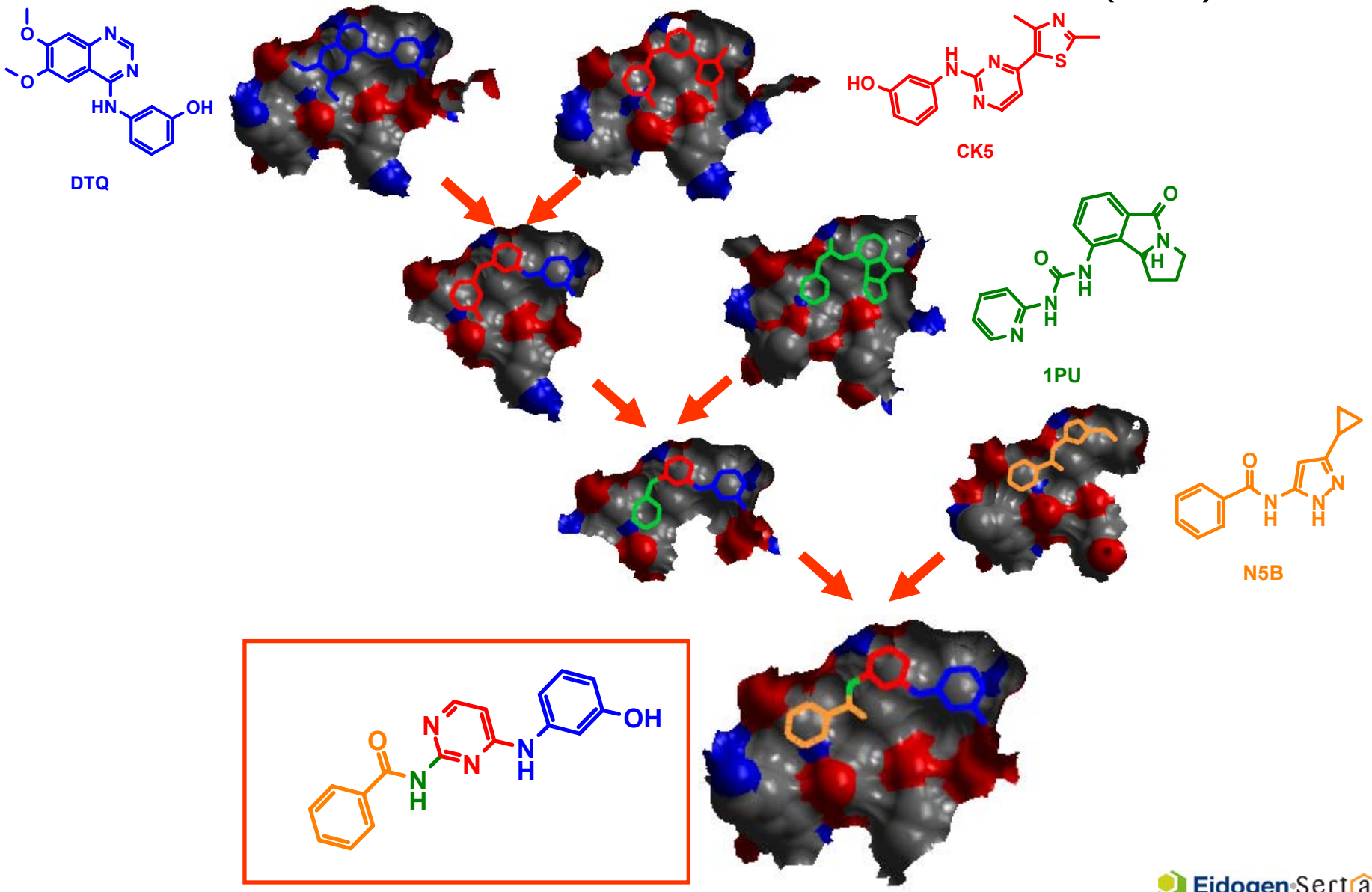
Three sizable datasets freely available to the research community

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

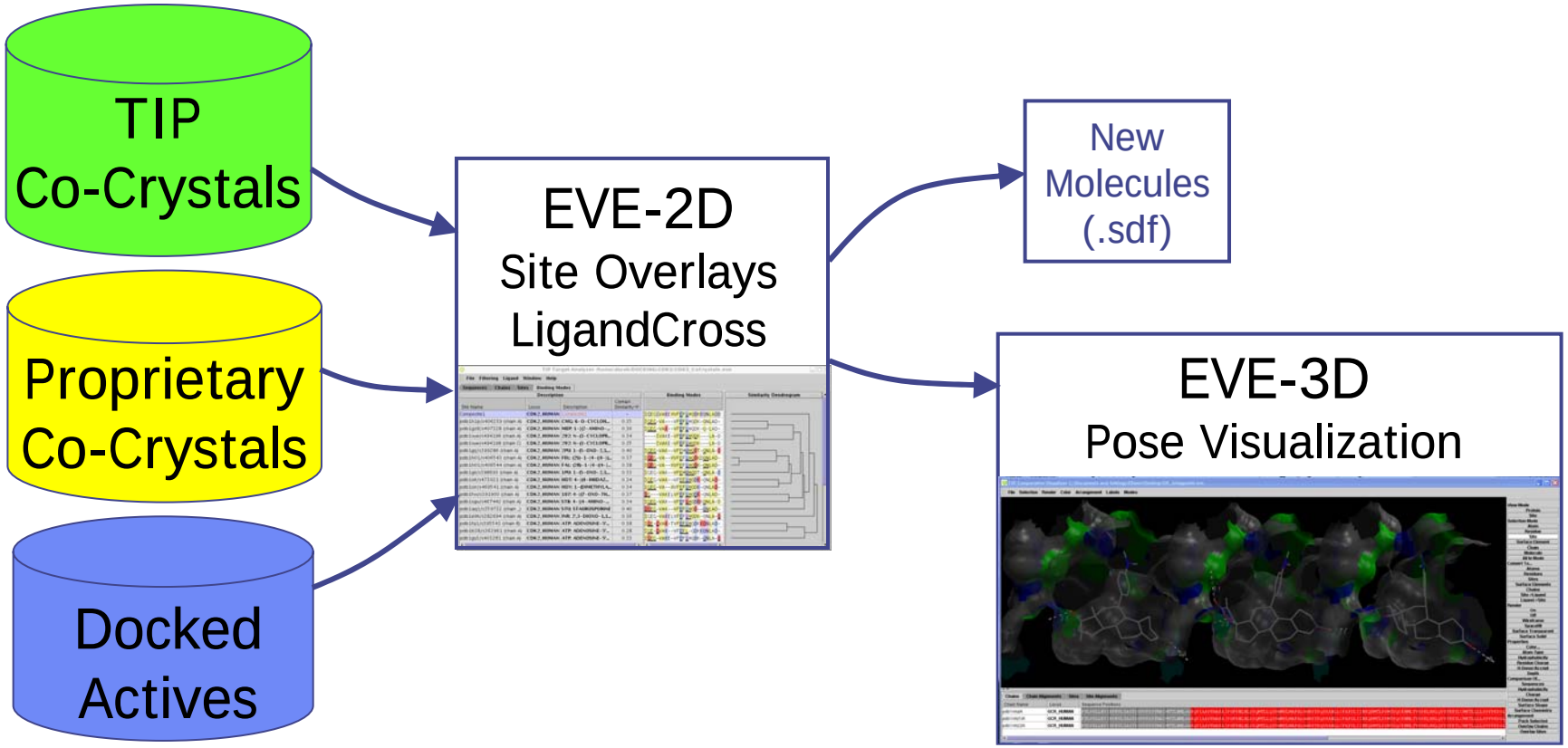
Eidogen-Sertanty

Lead Discovery: Knowledge-Based Design

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



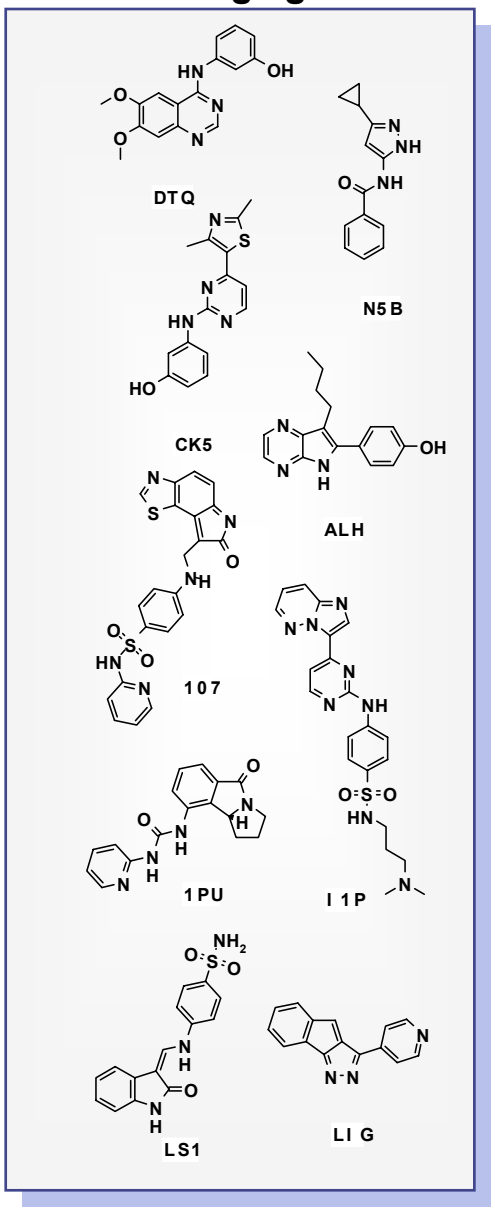
LigandCross Workflow



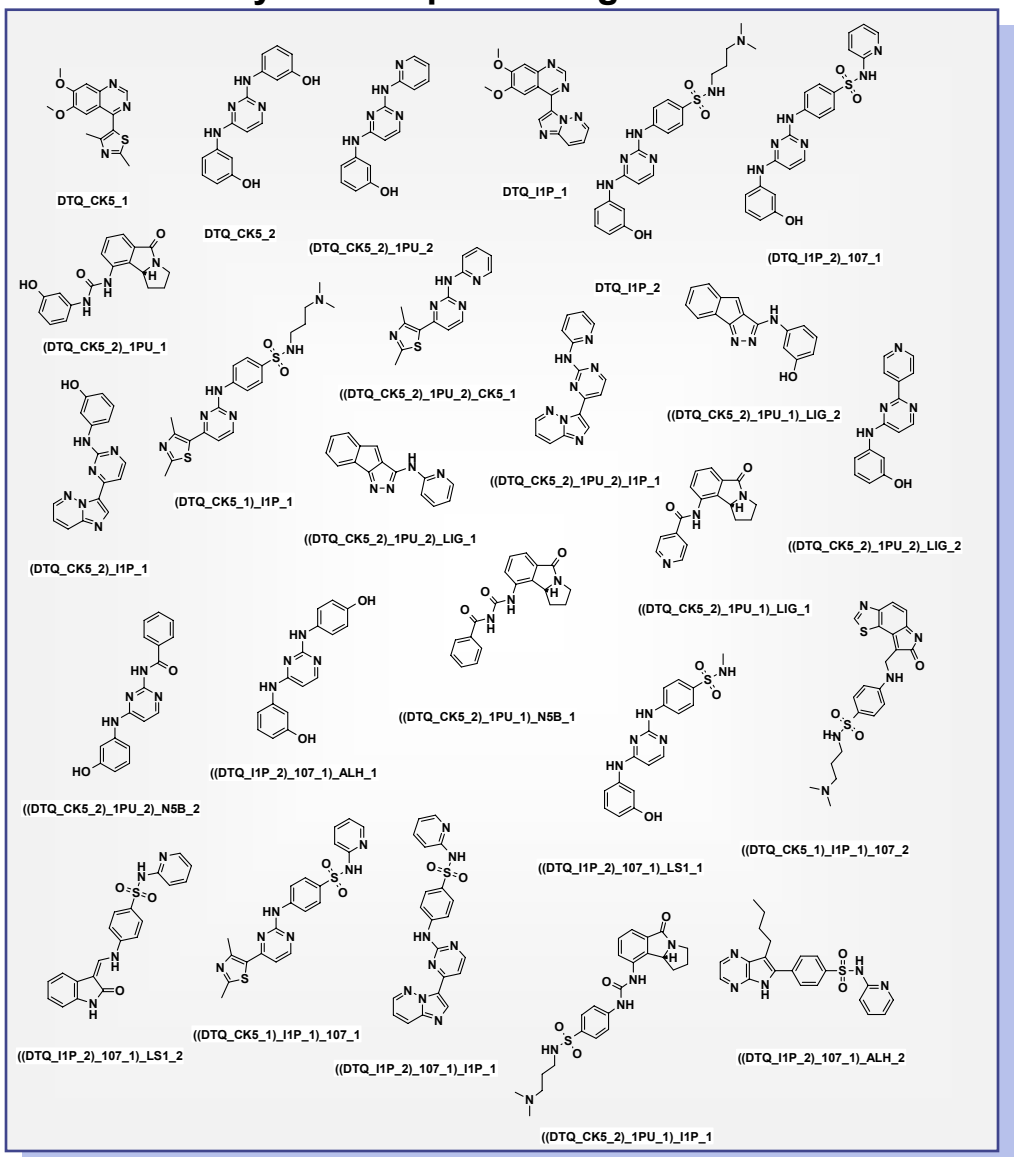
New Molecules via LigandCross

Novel Ligands via Ligand Crossover

Starting ligands

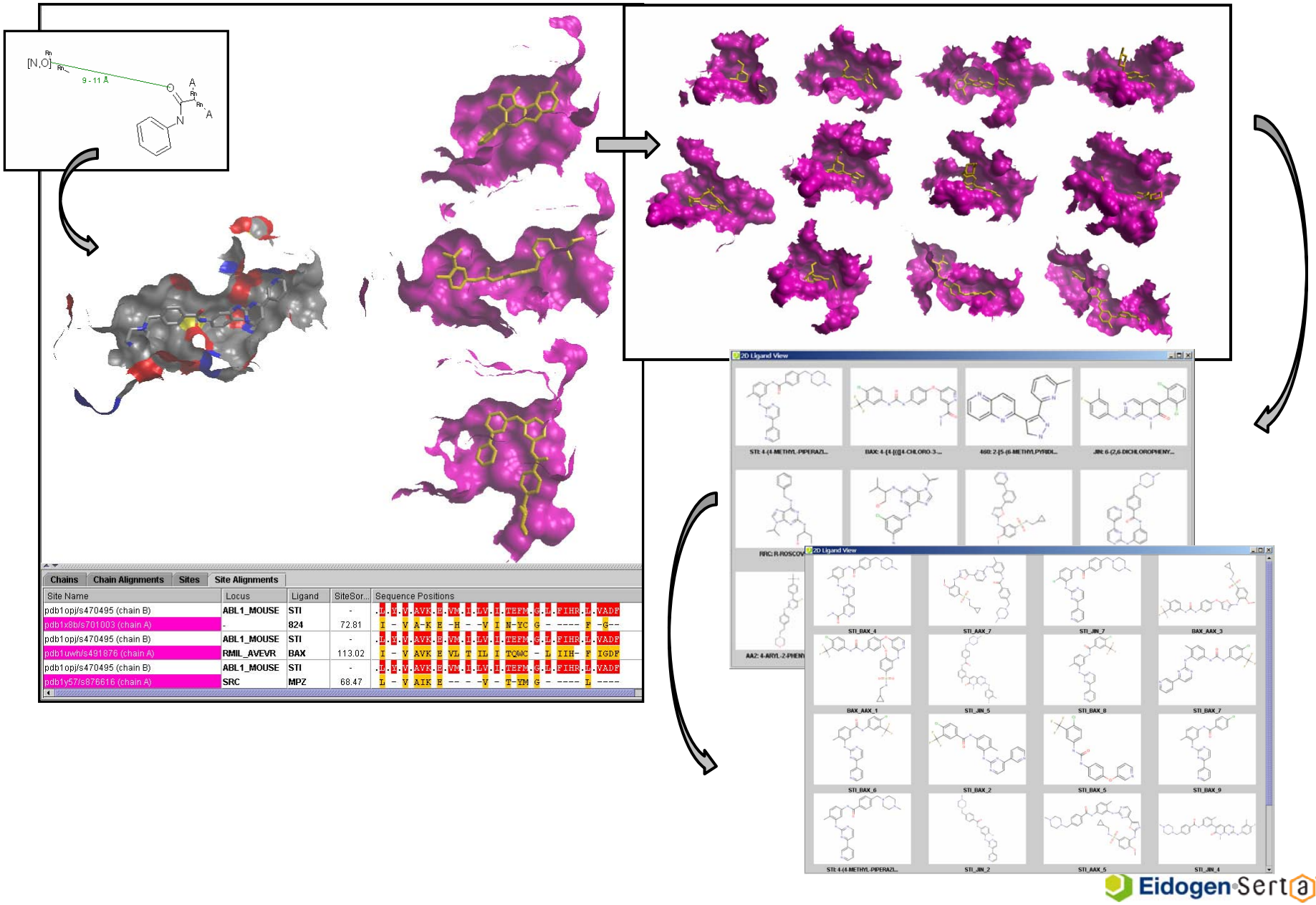


Hybridized product ligands



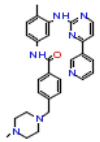
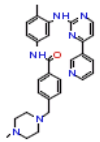
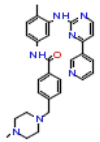
Eidogen-Sertanty

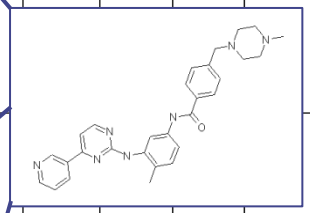
From Ligand Query to Sites to New Ligand Ideas

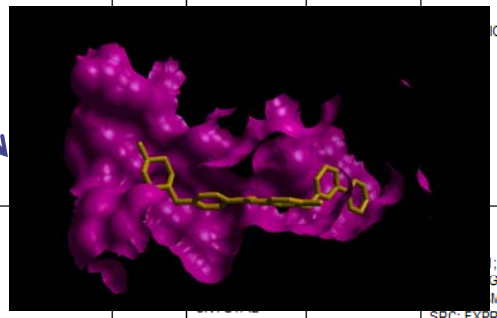


Eidogen-Sertanty

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

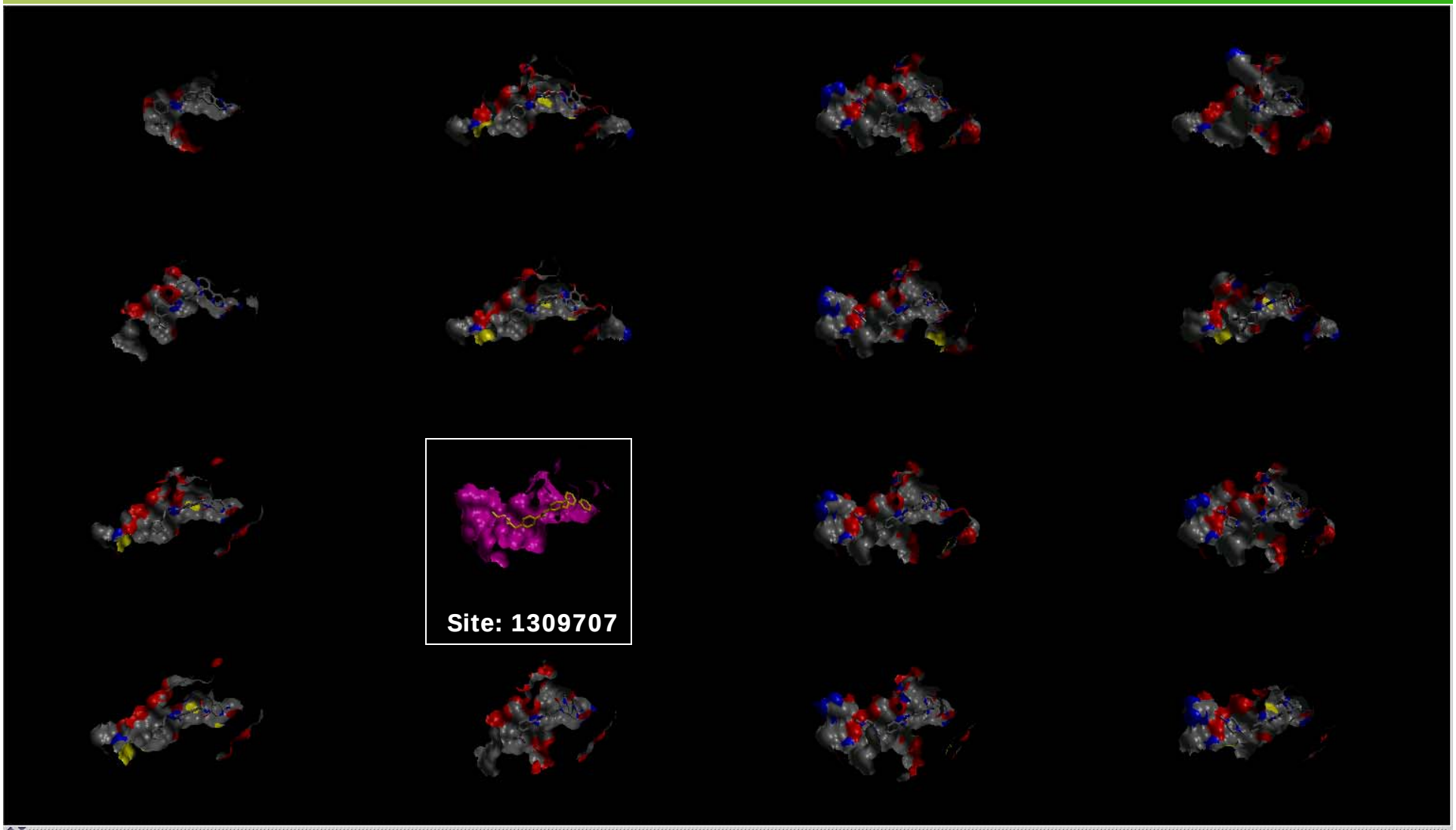
Molecule	ligname	similarity	pdbcode	siteeid	FourCode	pdbID	pdbBnxNumber	proteinId	title	classification	source	compound	releaseDate	journalTitle	journalReference	exptype
	STI	1	2pl0A	1309707	2pl0	2pl0	1305799	42526	LCK BOUND TO IMATINIB	TRANSFERASE	MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: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



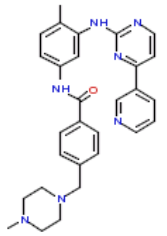
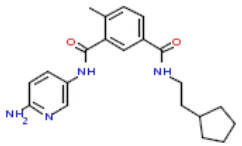
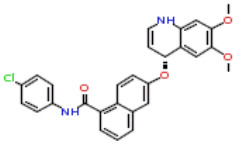
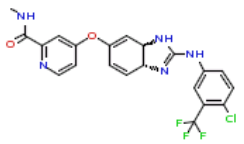
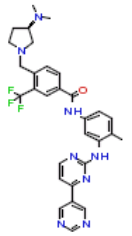
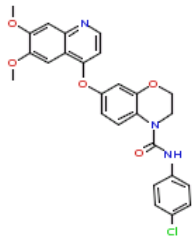
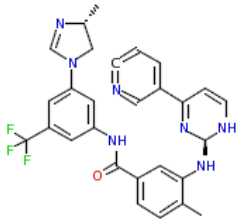
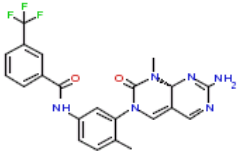
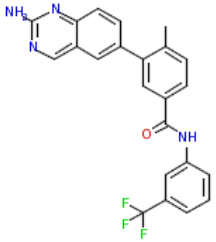
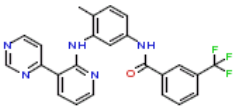
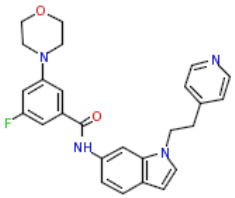
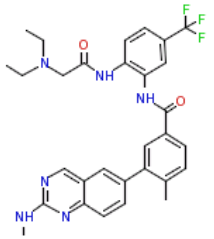
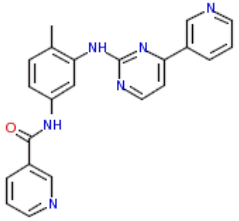
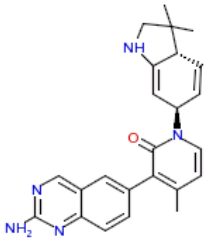
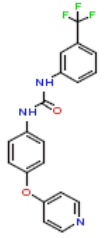
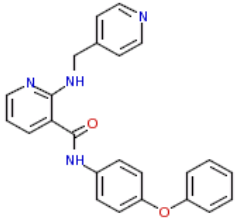
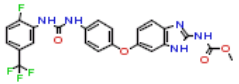
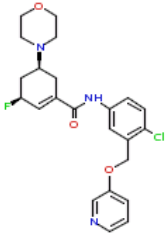
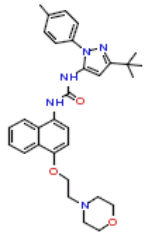
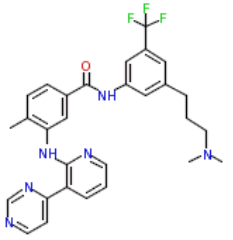
Chains | Chain Alignments | Sites | Site Alignments

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

Example Site Similarity Results (Query: s1309707)

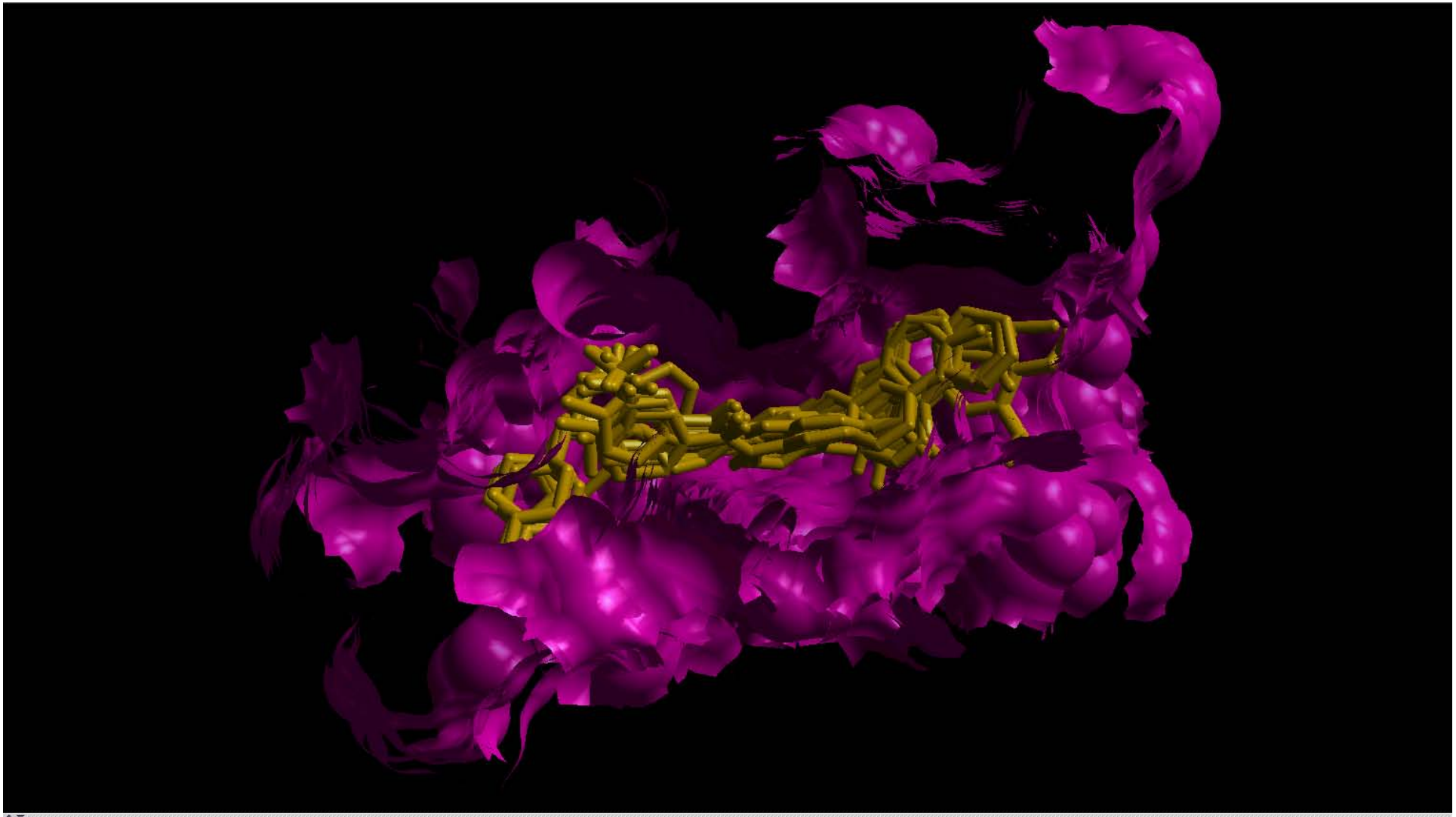
Site	SiteLigand	SiteProtein	SiteScore	ContactScore
1309707	STI	2pl0A	1000	1
1420904	C92	3cpbB	110.906	0.7
1384893	900	3b8qB	121.051	0.67
1322334	276	2qu5A	117.866	0.66
1284638	406	2e2bA	119.18	0.64
1396160	2RL	2rl5A	121.208	0.63
1400124	NIL	3cs9D	111.198	0.62
867405	7MP	2hiwA	101.948	0.61
916548	242	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

Eidogen-Sertanty

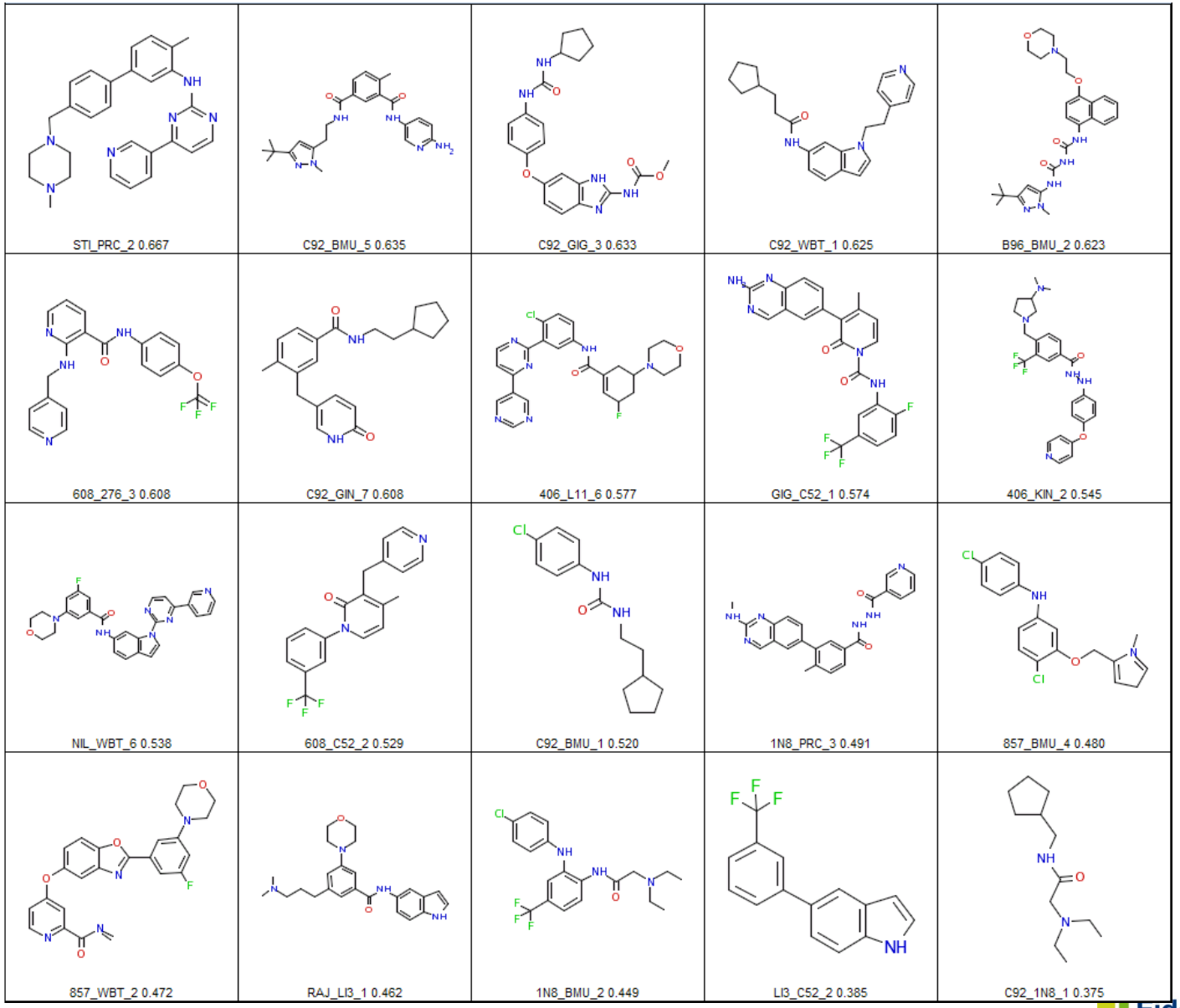
Step 3: LigandCross – Mixing Ligand Features from Aligned Sites



Chains | Chain Alignments | Sites | Site Alignments

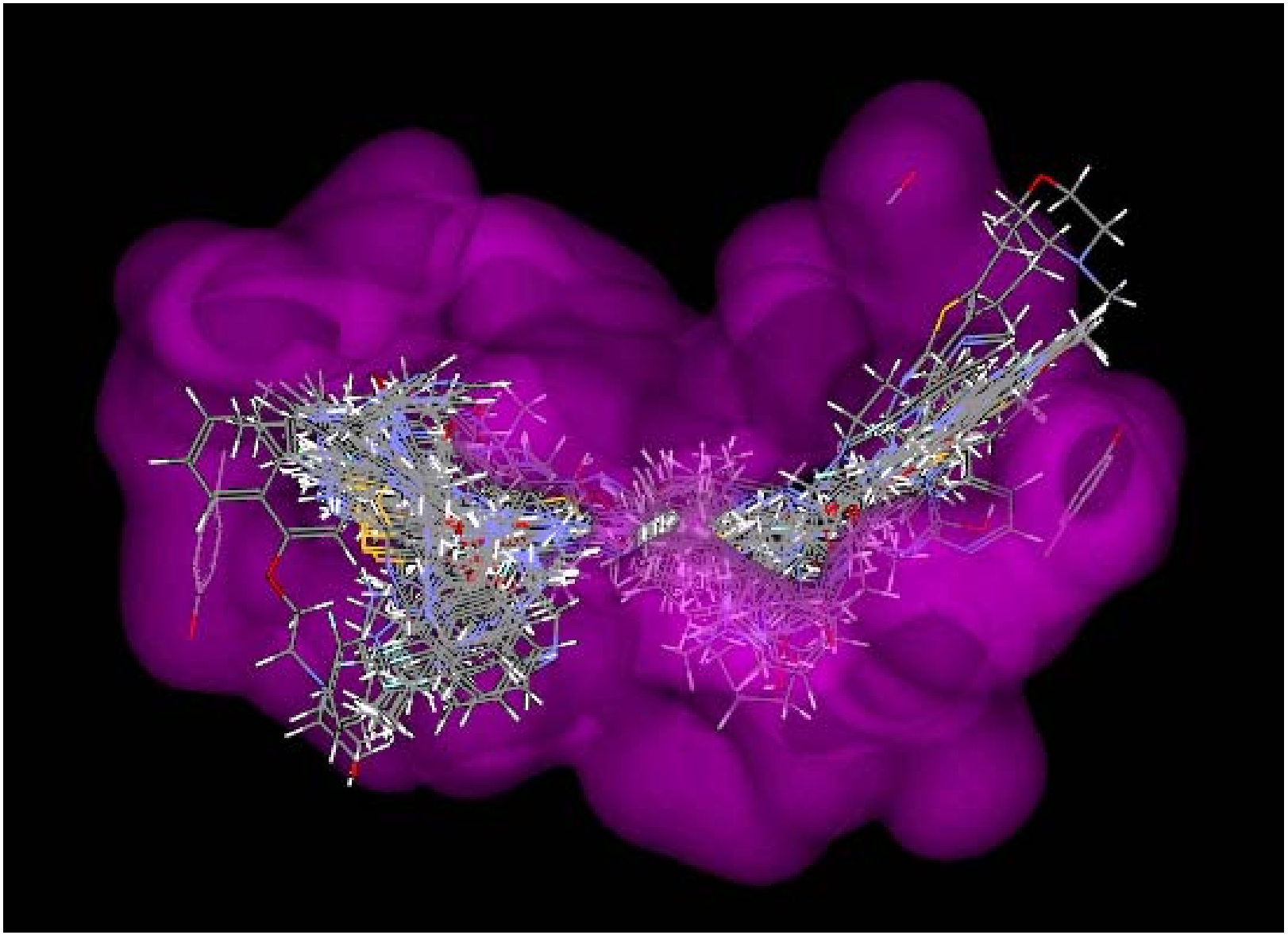
Site Name	Locus	Ligand	%Conf	Sequence Positions
pdb2pl0/s1309707 (chain A)	LCK	STI	100	.L.V.AVK.E.LM.L.LV.I.TEYM.GS.I.YIHR.L.IADF
pdb2of/s916648 (chain B)	LCK	242	100	.L.V.AVK.E.LM.L.LV.I.TEYM.G.I.Y.H.L.IADF.I
pdb2rl5/s1396160 (chain A)	-	2RL	100	.LG.V.AVK.L.E.IL.I.VV.V.TEFCKFGN.L.CIH.L.ICDF
pdb2e2b1/s1284639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

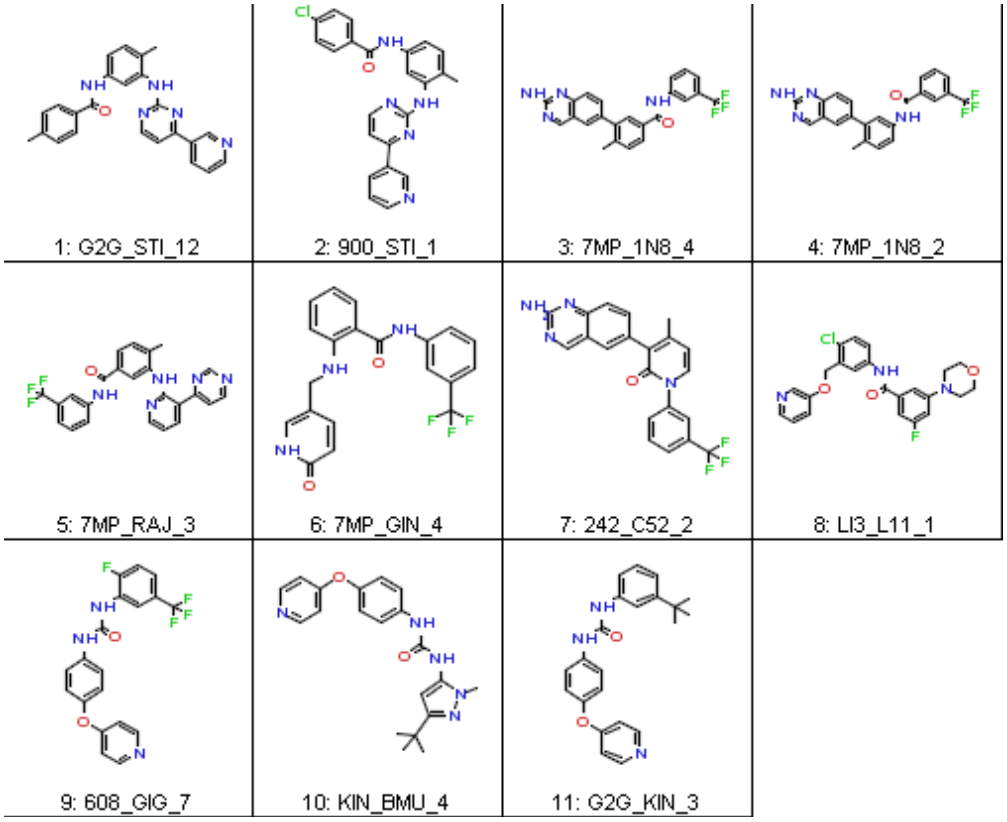
Step 4: LigandCross Ligands reDocked into s1309707



Eidogen-Sertanty

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

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

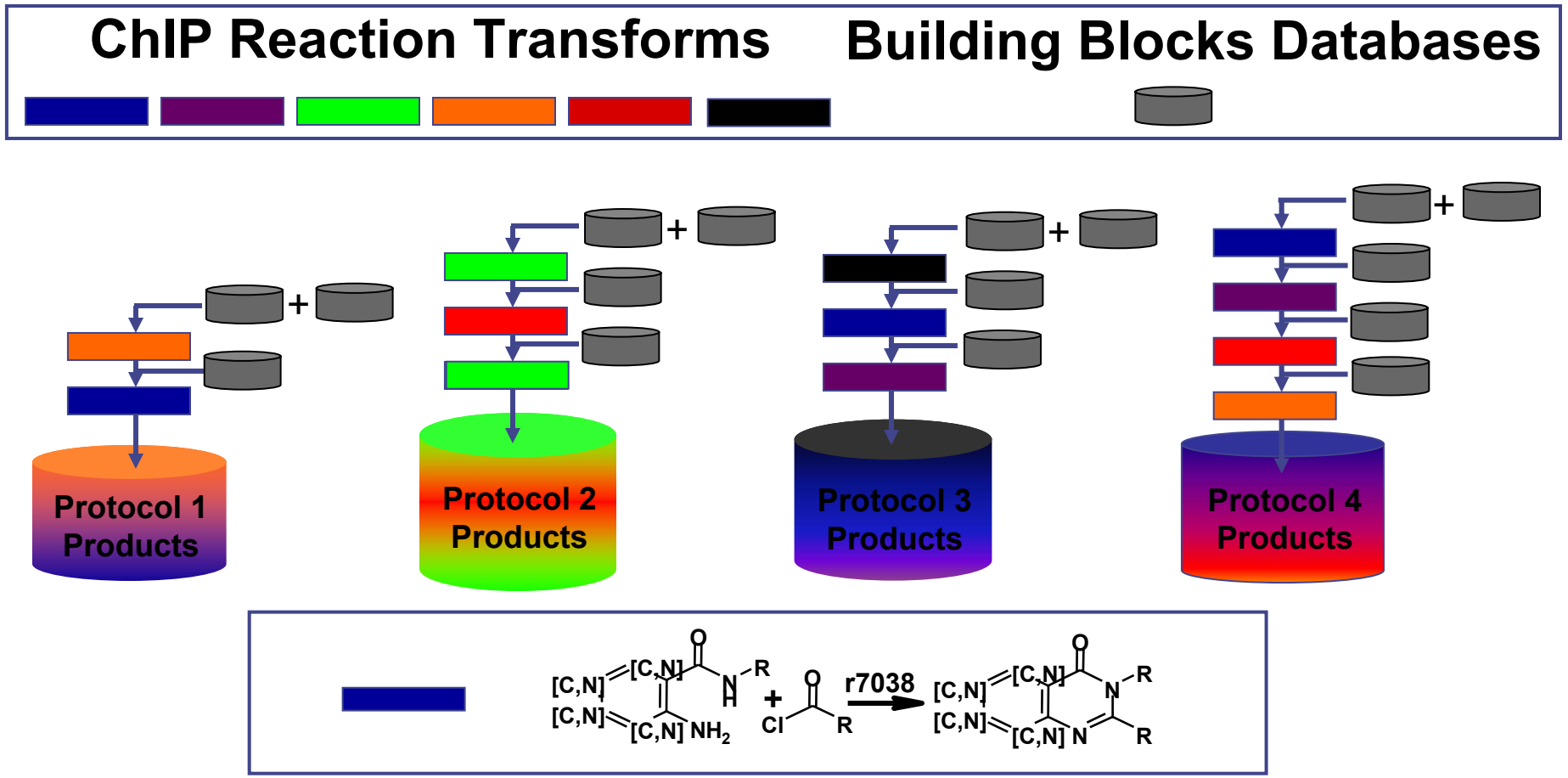
Chemical Intelligence Platform (ChIP™)

Eidogen-Sertanty

ChIP: Navigating Accessible Synthetic Space

SC Schurer, P Tyagi, SM Muskal, *J. Chem. Inf. Model.* **45**:239-48 (Mar-Apr, 2005).

Development funded by a $2.5M NIST-ATP Grant

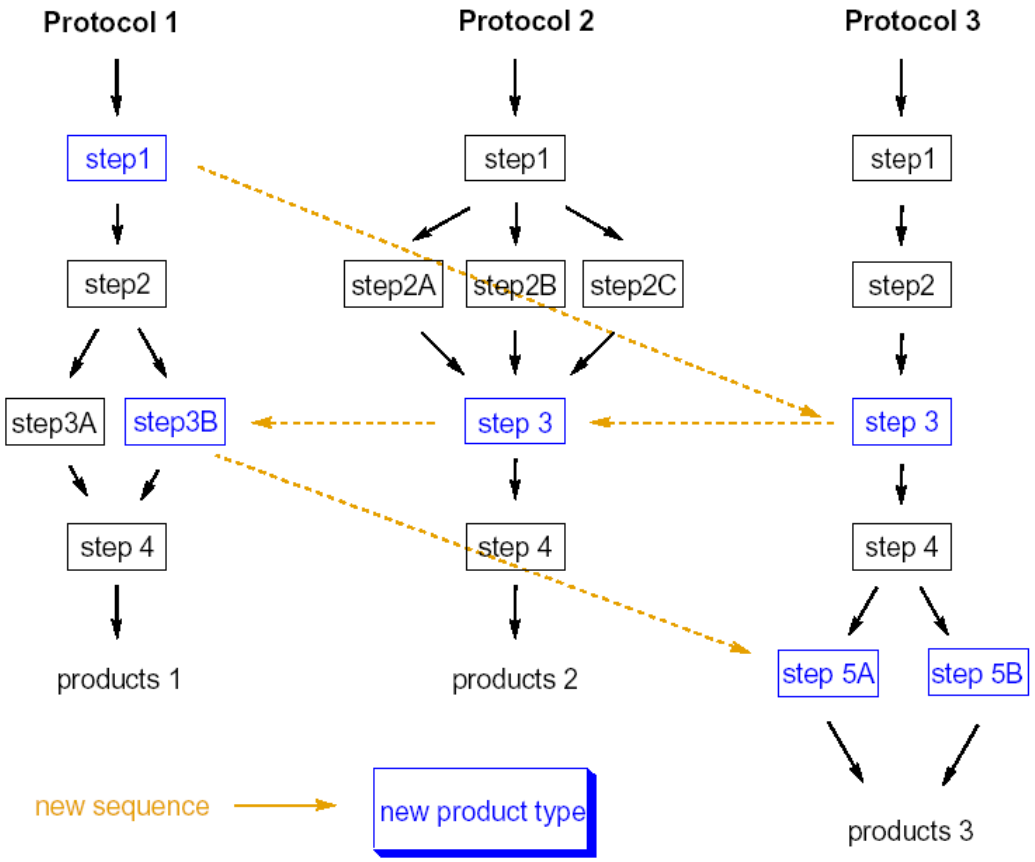


Building Block Sources May Be Changed,
Enabling Diversity Oriented or Focused Synthesis

Eidogen Sertanty

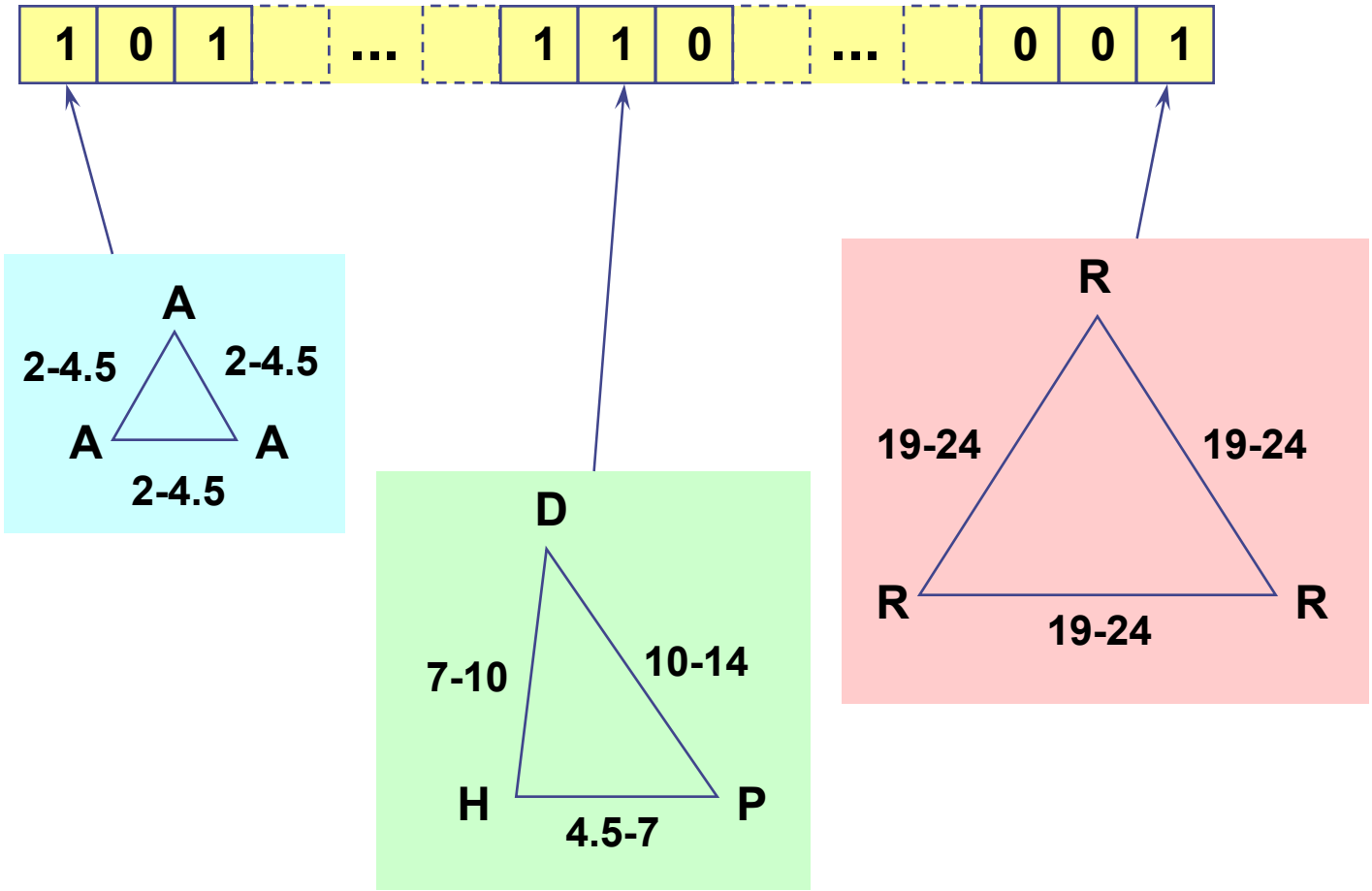
ChIP: Protocol Shuffling Chemistries

ChIP™ mixes and matches reaction methods so that novel scaffolds are generated *with* their synthetic road-maps



- S1: Nucleophilic Aromatic Substitution Reactions
- S2: Pd-Catalyzed Aromatic Substitutions
- S3: Functional Group Transformations
- S4: Amine Acylation Reactions (Amides/Carbamates)
- S5: Amine Acylation Reactions (Ureas/Thioureas)
- S6: Formation of Diverse Heterocyclic Systems
- S7: Formation of Thioimidazoles
- S8: Standard Deprotection Steps

Ligand pharmacophoric potential

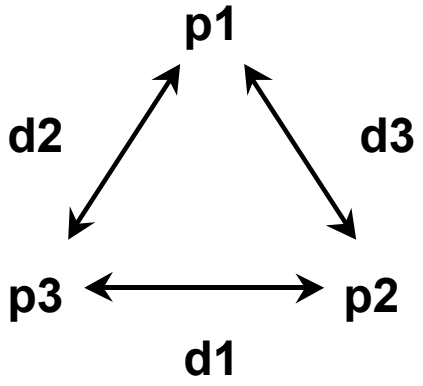


Journal of Chemical Information and Computer Sciences, 1999, 39 (3) : 569-74

Pharmacophoric Feature

7 pharmacophore types :

- H-bond acceptor (A) & donor (D)
- negative (N) & postive charges (P)
- aromatic (R), hydrophobic (H)
- other (X)

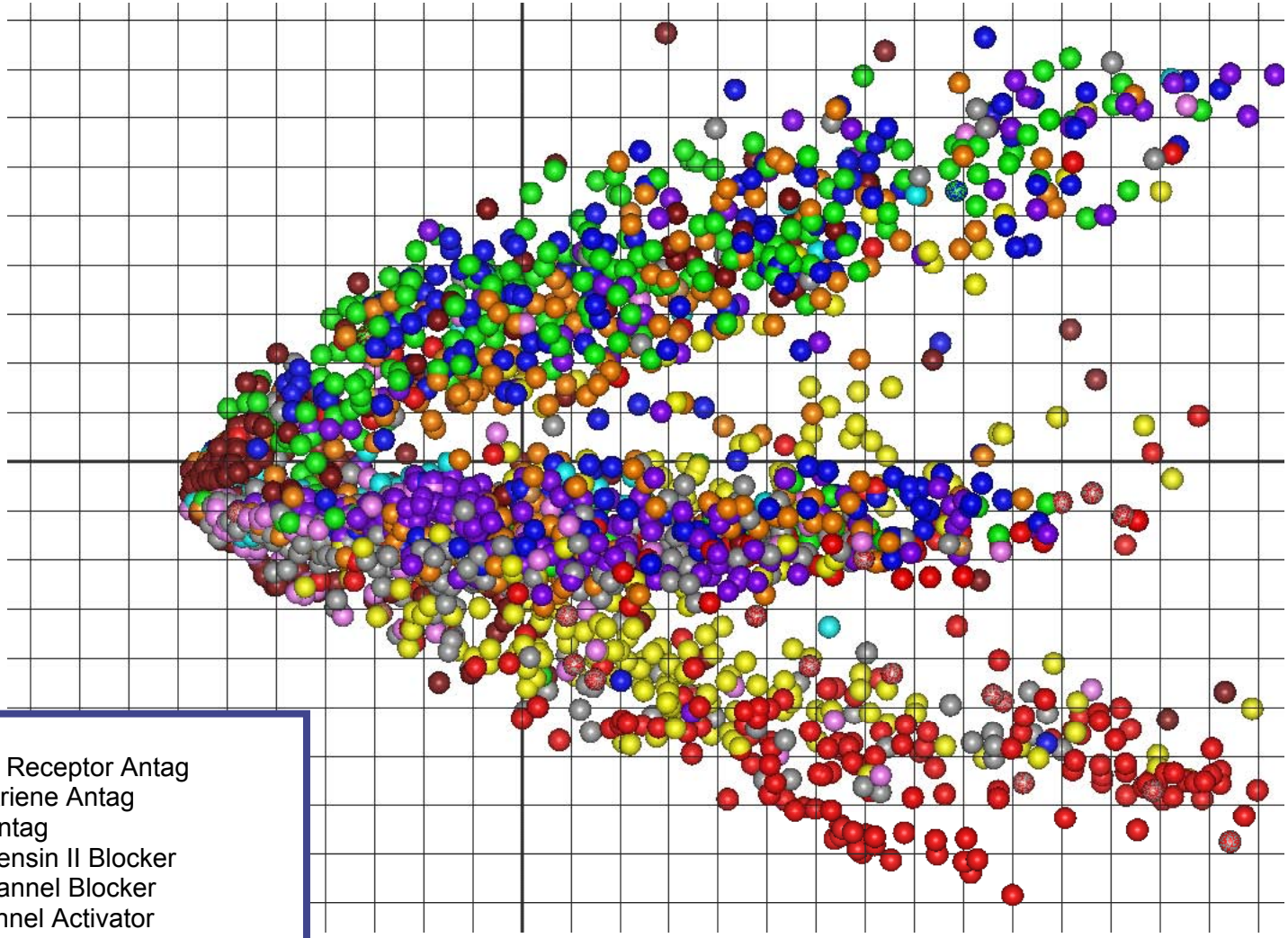


6 distance ranges :

2-4.5, 4.5-7, 7-10, 10-14, 14-19, 19-24 Å

Enumerate 3-point pharmacophores
-> 10,549

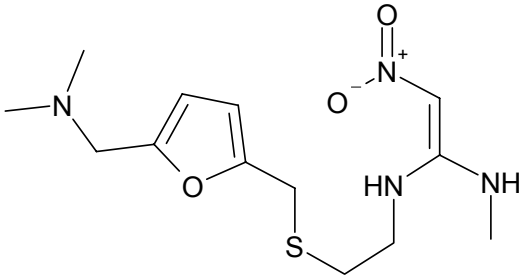
Target specific ligands share pharmacophoric feature

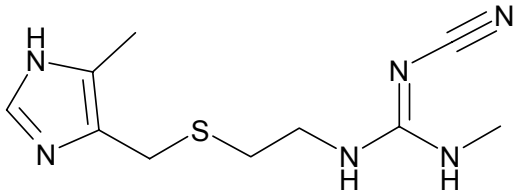


- NMDA Receptor Antag
- Leukotriene Antag
- PAF Antag
- Angiotensin II Blocker
- Ca Channel Blocker
- K Channel Activator
- Substance P Antag
- ACAT Inhibitor
- Cyclooxygenase Inhib
- Lipoxygenase Inhib

Non-obvious Me-Too's

MDLSim: 46.8/100.0
DaySim: 0.32/1.0
PFPSim: 0.88/1.0





MWT: 314.4; LogP: 0.27; pKa [2.30, 8.20]

Target: Histamine H2-antagonist.

Oral Avail.: 52% (±11)
Urinary Excretion: 69% (±6)
Plasma Bound: 15% (±3)
Clearance: 730 mL/min (±80)
Half-Life: 2.1 hr (±0.2)
Effective Conc.: 100 ng/mL

RANITIDINE (Zantac)
"Antiulcer"

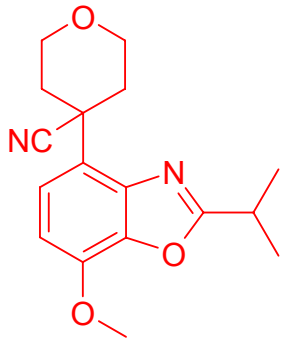
MWT: 252.3; LogP: 0.40; pKa: [6.80]

Target: Histamine H2-antagonist.

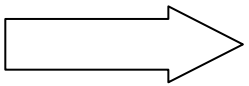
Oral Avail.: 62% (±6)
Urinary Excretion: 62% (±20)
Plasma Bound: 19%
Clearance: 540 mL/min (±130)
Half-Life: 1.9 hr (±0.3)
Effective Conc.: 800 ng/mL

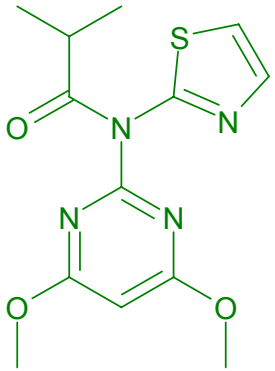
CIMETIDINE (Tagamet)
"Antiulcer"

ChIP-ing Towards Me-Too's



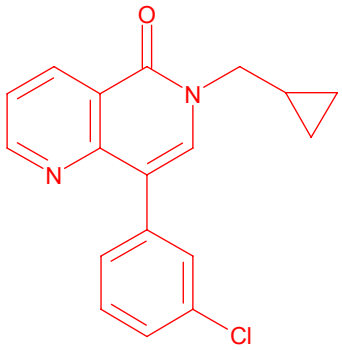
ChIP'd Me2



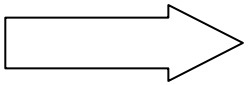


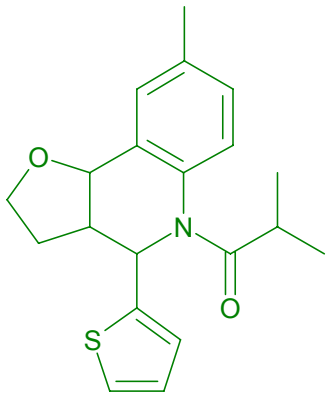
PFPSim: 0.94
MDLSim: 33.1/100.0

Known PDE-IV Inhibitors

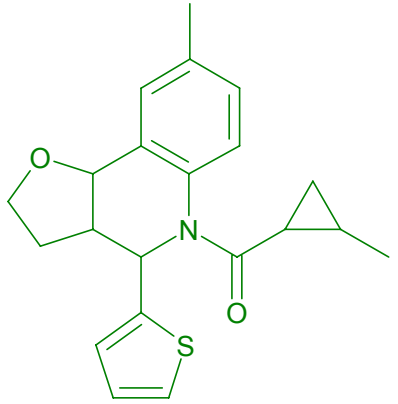


ChIP'd Me2





PFPSim: 0.92
MDLSim 44.4/100.0

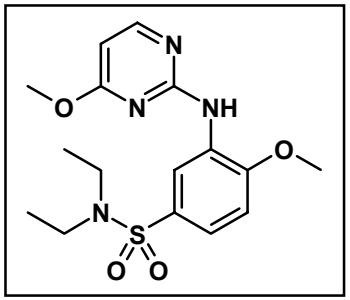


PFPSim: 0.92
MDLSim 50.6/100.0

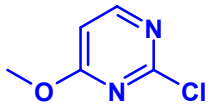
McGregor, M. J.; Muskal, S. M. Pharmacophore Fingerprinting. 1. Application to QSAR and Focused Library Design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 569-574.

Eidogen-Sertanty

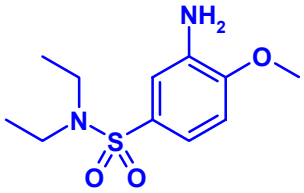
Example ChIP Generated Synthetic Road-Maps



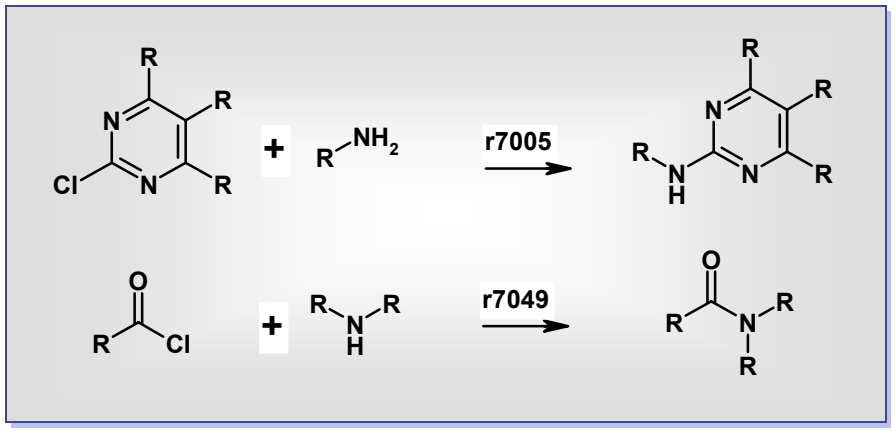
PFP Sim 0.8623

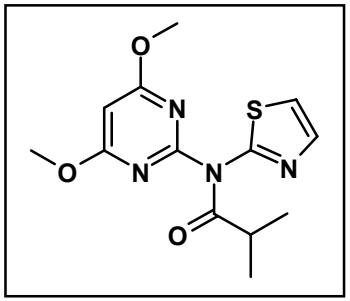


MFCD00194055

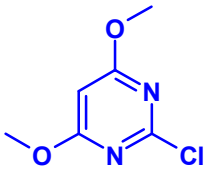


MFCD00009045

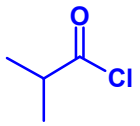




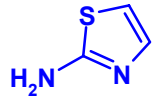
PFP Sim 0.9375



MFCD00274530



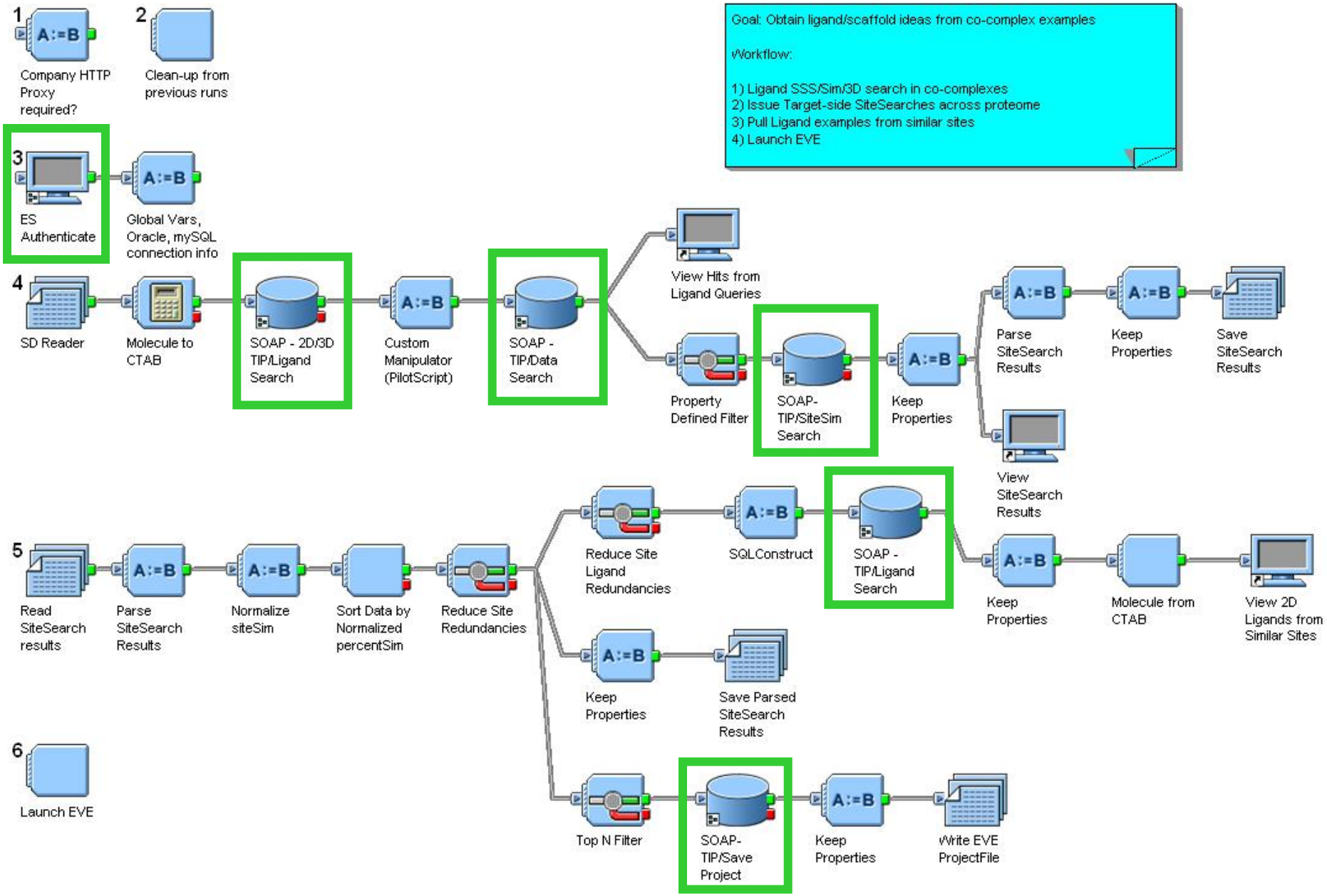
MFCD00000717



MFCD00005325

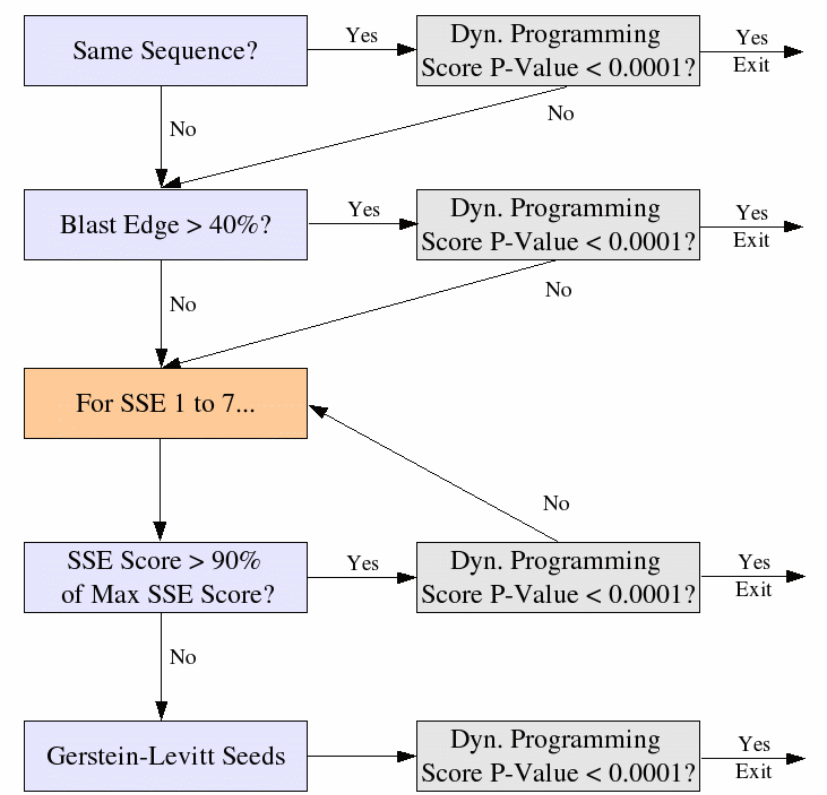
Eidogen-Sertanty

e.g. Non-specialist Proteomic Mining



StructSorter™

Pairwise StructSorter



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

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

Database-wide StructSorter

Clustering Scheme and Hierarchical Protocol

1) PDB sequences clustered at 90% identity and 95% coverage.

2) N-by-N comparison of one representative chain from each cluster

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

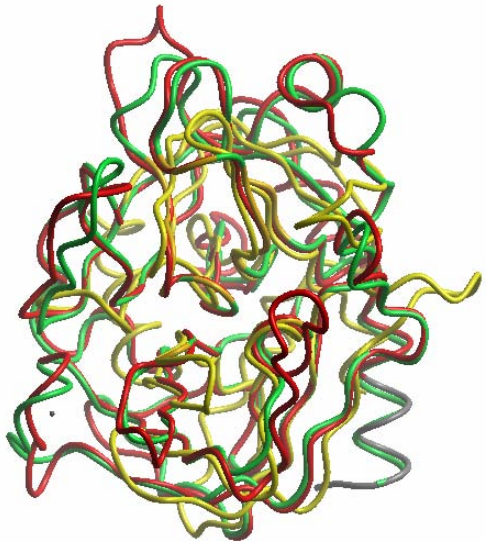
Allows structural alignment database to be computed in 1.5 months instead of 2.5 years.

StructSorter: A Method for Continuously Updating a Comprehensive Protein Structure Alignment Database J. Chem. Inf. Model. 2006, 46, 1871-1876

Eidogen-Sertanty

StructSorter Example: Rhinovirus protease

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



EVE Target Analyzer

File Export Filtering Ligand Window Help

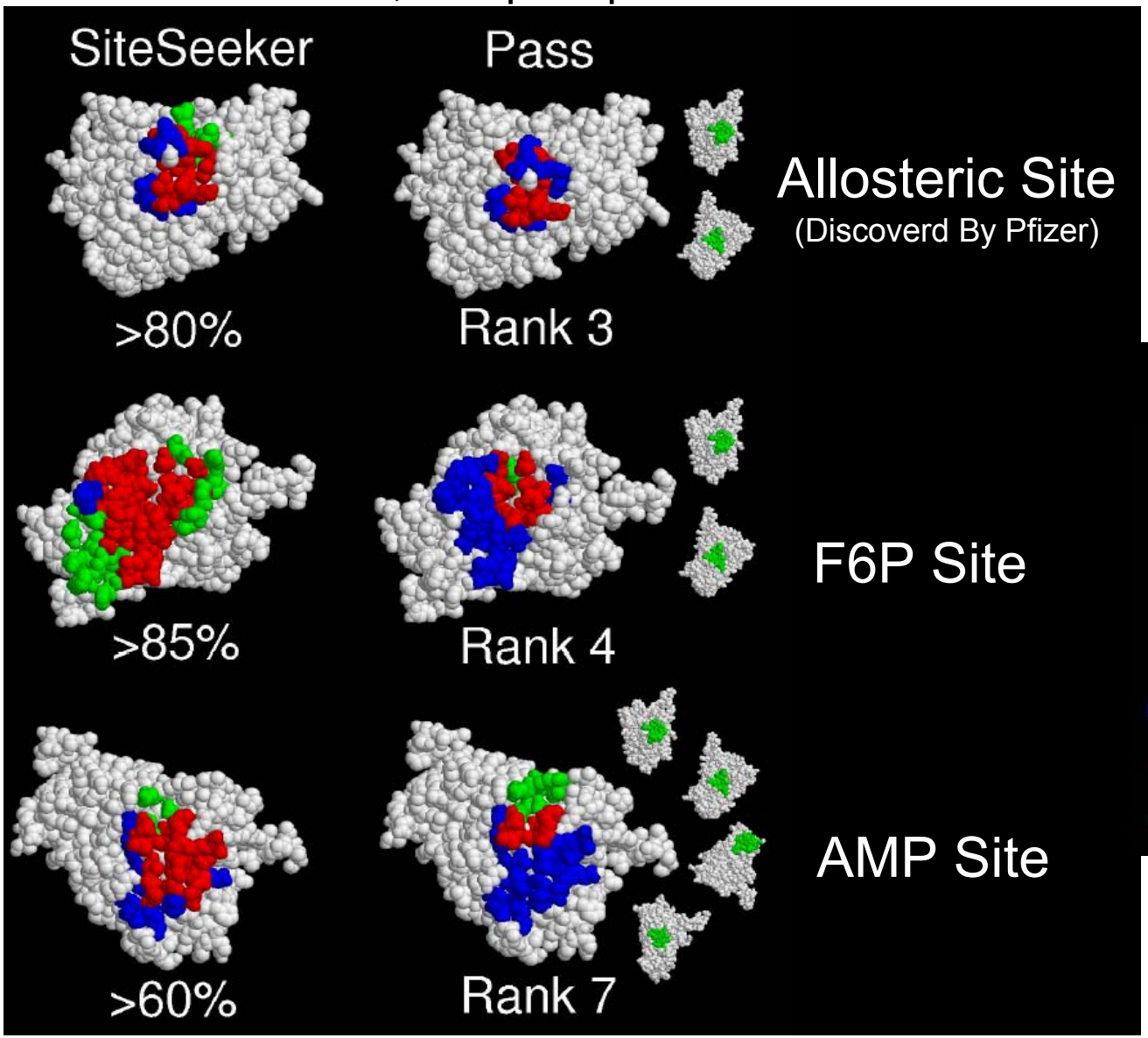
Sequences | Chains | Sites | Binding Modes

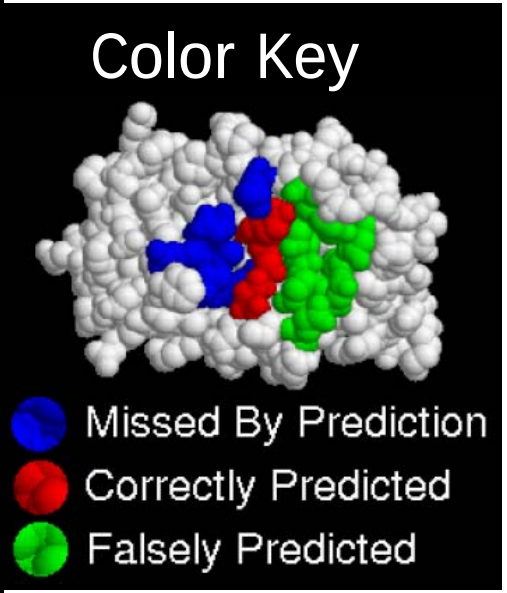
Description							Chain Alignments	Similarity Dendrogram
Chain Name	Locus	Organism	CRMS	%ID	Sites	Description		
pdb1cqq/A	POLG_...	H.rhinovirus 2	-	-	2	Chain A, MOL_ID: 1; MOLECULE: TYPE 2 RHINOVIRUS 3C F		
pdb1q31/A	POLG_...	T.etch virus	2.325	11	2	Chain A, NUCLEAR INCLUSION PROTEIN A		
pdb1q31/B	POLG_...	T.etch virus	2.324	11	2	Chain B, NUCLEAR INCLUSION PROTEIN A		
pdb1f7z/A	TRY2_R...	B.taurus	2.753	11	1	Chain A, TRYPSIN II, ANIONIC		
pdb1a0j/A	TRY3_S...	S.salar	2.581	11	5	Chain A, TRYPSIN		
pdb1spj/A	KLK1	H.sapiens	2.704	10	5	Chain A, KALLIKREIN 1		
pdb1mza/A	GRAK_...	H.sapiens	3.043	9	2	Chain A, PRO-GRANZYME K		
pdb1mzd/A	GRAK_...	H.sapiens	3.530	10	2	Chain A, PRO-GRANZYME K		
pdb1bio/_	CFAD_...	H.sapiens	2.822	8	5	Chain _, COMPLEMENT FACTOR D		
pdb1bru/P	EL2_PIG	S.scrofa	2.696	9	2	Chain P, ELASTASE		
pdb1p57/B	HEPS_...	H.sapiens	2.705	9	2	Chain B, SERINE PROTEASE HEPSIN		
pdb1a0l/A	TRB2_...	H.sapiens	2.687	10	2	Chain A, BETA-TRYPTASE		
model3999...	MPN	H.sapiens	2.805	9	3	Pancreasin precursor (EC 3.4.21.-) (Marapsin) (Channel-act		
pdb1ybw/A	-	H.sapiens	2.722	11	2	Chain A, HEPATOCYTE GROWTH FACTOR ACTIVATOR PRE		
pdb1eaw/A	ST14_H...	B.taurus	2.713	7	1	Chain A, SUPPRESSOR OF TUMORIGENICITY 14		
model5711...	PRSS12	H.sapiens	2.867	8	3	Neurotrypsin precursor (EC 3.4.21.-) (Motopsin) (Leydin)		
pdb1lmw/B	UROK_...	H.sapiens	2.924	10	5	Chain B, UROKINASE-TYPE PLASMINOGEN ACTIVATOR		
pdb1lmw/D	UROK_...	H.sapiens	2.846	10	5	Chain D, UROKINASE-TYPE PLASMINOGEN ACTIVATOR		
model3428...	PROZ	H.sapiens	3.007	8	4	Vitamin K-dependent protein Z precursor		
pdb1h1b/A	ELNE_H...	H.sapiens	2.672	9	8	Chain A, LEUKOCYTE ELASTASE		
model2909...	HP	H.sapiens	3.375	5	3	Haptoglobin precursor		
pdb1iau/A	GRAB_...	H.sapiens	2.619	11	11	Chain A, GRANZYME B		
pdb1t31/A	MCT1_...	H.sapiens	2.808	10	12	Chain A, CHYMASE		
pdb1pjp/A	MCT1_...	H.sapiens	3.298	10	6	Chain A, CHYMASE		
pdb1azz/B	COGS_...	C.pugilator	2.625	8	2	Chain B, COLLAGENASE		
pdb1azz/A	COGS_...	C.pugilator	2.615	8	2	Chain A, COLLAGENASE		
pdb1au8/A	CATG_...	H.sapiens	2.588	11	2	Chain A, CATHEPSIN G		
pdb1gpz/A	C1R_H...	H.sapiens	3.379	11	12	Chain A, COMPLEMENT C1R COMPONENT		
model2909...	HP	H.sapiens	3.047	5	5	Haptoglobin precursor		
pdb1gpz/B	C1R_H...	H.sapiens	3.373	10	6	Chain B, COMPLEMENT C1R COMPONENT		
pdb1sg0/A	KLK4_...	M.musculus	3.614	8	3	Chain A, NERVE GROWTH FACTOR		
pdb1sg0/X	KLK4_...	M.musculus	3.623	8	4	Chain X, NERVE GROWTH FACTOR		
pdb1wcz/A	STSP_...	S.aureus	2.604	10	2	Chain A, GLUTAMYL ENDOPEPTIDASE		
model1370...	PRSS11	H.sapiens	2.822	12	1	Serine protease HTRA1 precursor (EC 3.4.21.-) (L56)		
pdb1agj/A	ETA_ST...	S.aureus	2.558	7	2	Chain A, EPIDERMOLYTIC TOXIN A		
pdb1dxp/A	POLG_...	H.c virus	2.677	11	3	Chain A, PROTEASE/HELICASE NS3 (P70)		
pdb1bef/A	POLG_...	D.virus type 2	2.739	9	1	Chain A, DENGUE VIRUS NS3 SERINE PROTEASE		
pdb1bt7/_	POLG_...	H.c virus	3.157	11	2	Chain _, NS3 SERINE PROTEASE		
pdb1hpg/A	GLUP_...	S.griseus	2.730	8	4	Chain A, GLUTAMIC ACID-SPECIFIC PROTEASE		

Eidogen-Sertanty

SiteSeeker Example

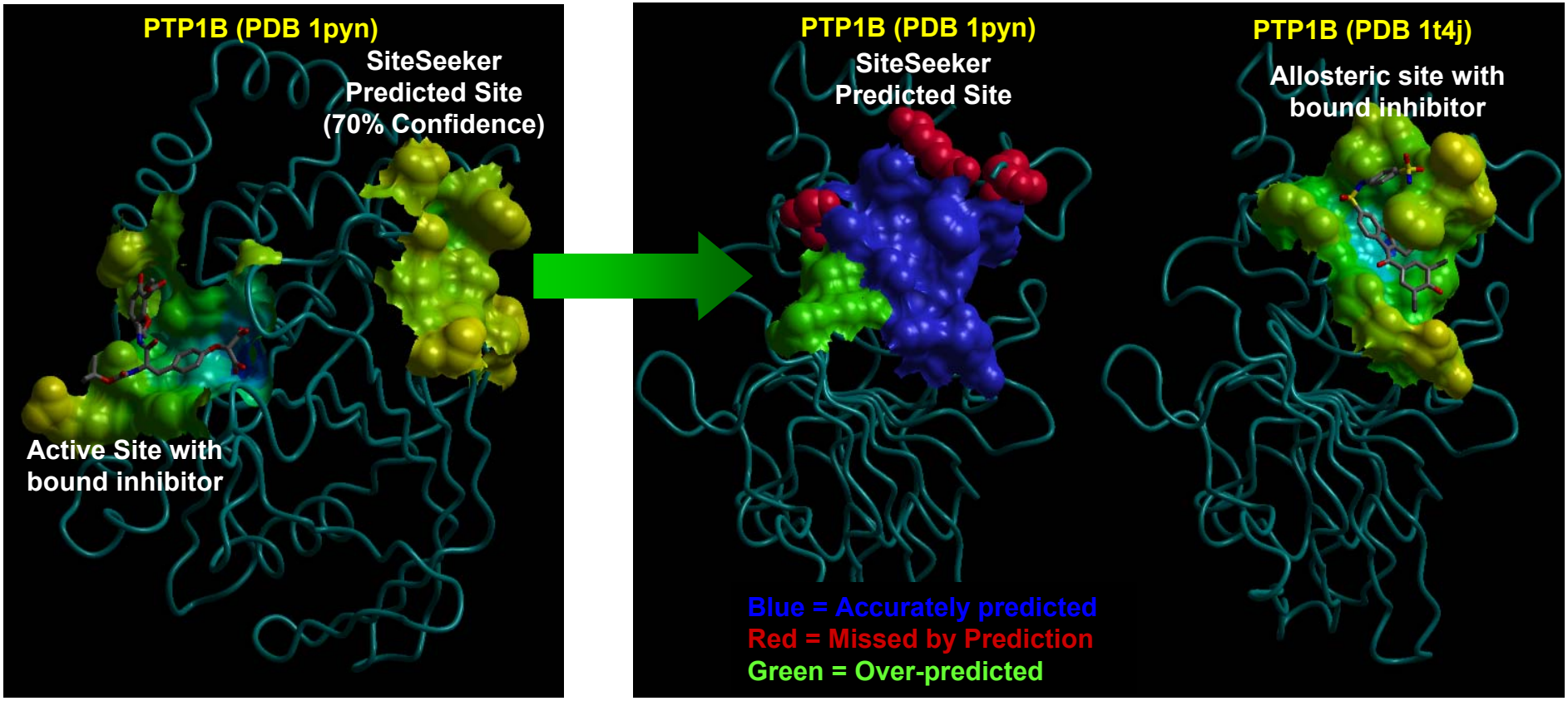
Fructose-1,6-bisphosphatase FBPase





Eidogen-Sertanty

SiteSeeker Example: PTP1B Allosteric Site



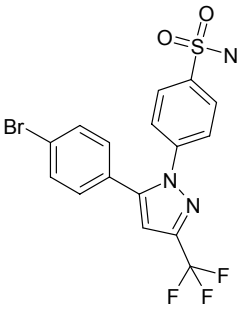
All structures in TIP are annotated with known and predicted binding sites, along with **confidence** levels for each annotation

Eidogen Sertanty

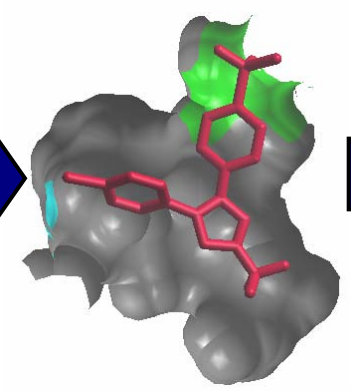
Virtual Target Screening Example: COX-2

Example: Identifying potential "off-targets" for COX-2 inhibitors

COX-2 inhibitor



COX-2 Active Site



Use site to query TIP and rank similar binding sites

EVE Target Analyzer

File Export Filtering Ligand Window Help

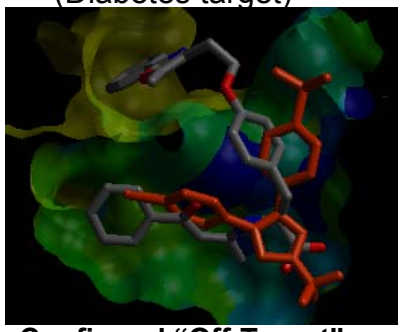
Sequences | Chains | Sites | Binding Modes

Description

Site Name	Locus	Description	%Conf	SiteSorter Similarity
pdb1cx2/s3748...	PGH2_MOUSE	S68: 1-PHENYL	100	-
pdb2prg/s4720...	PPAT_HUMAN	Predicted Site	97	74.57
pdb1efh/s3812...	SUHA_HUMAN	Predicted Site	97	73.70
pdb1ojd/s4695...	AOFB_HUMAN	Predicted Site	98	72.92
pdb1d1s/s381...	ADH7_HUMAN	Predicted Site	78	71.59
pdb1deh/s384...	ADHB_HUMAN	Predicted Site	80	71.25
pdb1ivh/s4135...	IVD_HUMAN	FAD: FLAVIN-A...	100	69.09
model41234_1...	HSPCB	Predicted Site	89	68.99
model171728_...	LYN_HUMAN	Predicted Site	98	68.77
model220469...	TR2A_HUMAN	Predicted Site	85	68.45
pdb1hy3/s4121...	SUOE_HUMAN	Predicted Site	67	68.23
model30570_8...	HSD11B2	NAP: NADP NI...	100	67.93
pdb1v4s/s5011...	HXK4_HUMAN	Predicted Site	98	67.63
model5407_67...	PDE3A	Predicted Site	87	67.58
model12228_3...	HSD11B1	Predicted Site	66	66.25
pdb3grt/s4054...	GSHR_HUMAN	FAD: FLAVIN-A...	100	66.23
pdb1og5/s467...	CPC9_HUMAN	Predicted Site	97	65.97
pdb1xu9/s5232...	HSD11B1	Predicted Site	91	65.69
model13939_2...	TUBB	GDP: GUANOS...	100	65.59
pdb1pq2/s477...	CPC8_HUMAN	HEM: PROTOP...	100	65.57
pdb1dqa/s378...	HMDH_HUMAN	NAP: NADP NIC...	100	65.48
model5988_3_...	SLC25A4	Predicted Site	98	65.31
model14830_1...	ADH1C	Predicted Site	76	65.10
pdb1nzx/s4604...	DHAM_HUMAN	Predicted Site	89	64.91
pdb1jk8/s4263...	HA24_HUMAN	C: 1jk8C	100	64.08
pdb1xn2/s6848...	BAE1_HUMAN	F: 1xn2F	100	63.95
pdb1q9m/s466...	CN4D_HUMAN	Predicted Site	67	63.92
pdb1gwx/s403...	PPAS_HUMAN	433: 2-(4-[3-[1-...	100	63.78
model2826_42...	ACE	Predicted Site	70	63.78
pdb1t2a/s5006...	GMDS_HUMAN	NDP: NADPH DI...	100	63.74
model33343_3...	ACHE	Predicted Site	95	63.23

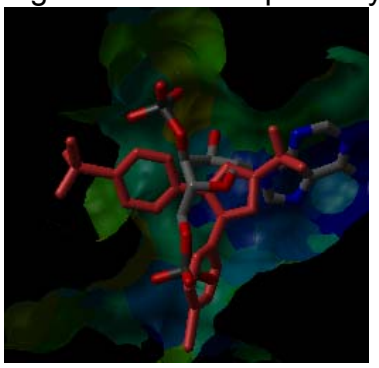
Prioritize "off-targets" based on Site-Ligand Contact analysis and/or biological relevance

Rank #1: PPAR-Gamma
(Diabetes target)



- **Confirmed "Off-Target" for COX-2 inhibitors**

Rank #10: Estrogen Sulfotransferase
(Estrogen Metabolism pathway)



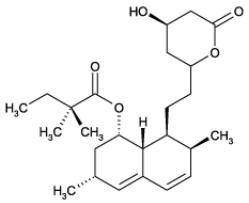
- **Possible "Off-Target" for COX-2 inhibitors**

Eidogen-Sertanty

Virtual Target Screening Example: Statins

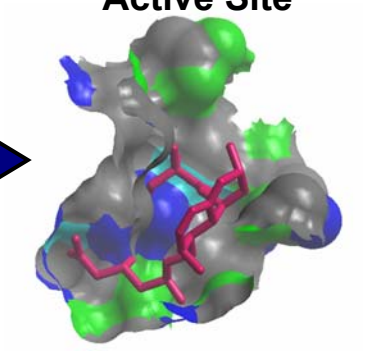
Example 1: Searching for off-targets to explain *pleiotropic* effects of statins

Simvastatin (Zocor)



Identify binding site

HMG-CoA Reductase Active Site



Perform Site Similarity Search in TIP

Visualize Site alignments in EVE

EVE Target Analyzer

File Export Filtering Ligand Window Help

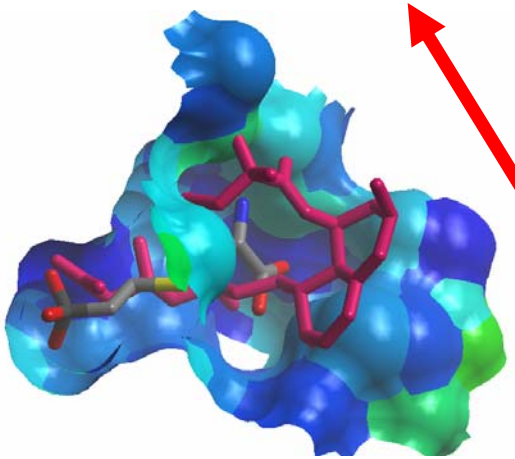
Sequences | Chains | Sites | Binding Modes

Description

Site Name	Locus	Description	%Conf	SiteSorter Similarity	%ID
pdb1hw9/s41...	HMDH...	SIM: SIMVASTATIN	100	-	-
pdb1rkp/s483...	CN5A_...	Predicted Site	97	82.34	8
pdb1bwc/s35...	GSHR_...	FAD: FLAVIN-ADE...	100	75.91	17
pdb1ege/s38...	ACDM...	FAD: FLAVIN-ADE...	100	71.02	8
pdb1ojd/s469...	AOFB_...	Predicted Site	98	68.11	8
pdb1p0i/s451...	CHLE_...	Predicted Site	98	66.74	13
pdb1yet/s496...	HS9A_...	Predicted Site	76	63.26	13
model17170_...	CYP51...	Predicted Site	98	62.98	8
pdb1hy3/s412...	SUOE_...	Predicted Site	89	62.57	8
pdb1j8h/s418...	2DRA_...	C: 1j8hC	100	62.15	8
pdb1lfl/s4392...	HBA_H...	Predicted Site	73	61.48	8
pdb1qb0/s47...	MPI2_...	Predicted Site	77	60.77	8
pdb1myp/s44...	PERM_...	HEM: PROTOPOR...	100	60.43	8
pdb1xu7/s522...	HSD11...	NDP: NADPH DIH...	100	60.39	8
pdb1jvd/s427...	UAP1_...	UD1: URIDINE-DIP...	100	60.17	8
model21075_...	ALOX5	Predicted Site	72	60.00	13
pdb1nnl/s446...	SERB_...	Predicted Site	60	59.94	8
model29287_...	TUBB5	GDP: GUANOSIN...	100	59.56	8
pdb1e28/s38...	1B51_...	C: 1e28C	100	59.30	8
pdb1nut/s456...	NMA3_...	APC: DIPHOSPH...	100	58.77	8
pdb1gwx/s40...	PPAS_...	Predicted Site	84	58.61	8
pdb1f83/s390...	BXB_C...	B: 1f83B	100	57.87	8
model2028_1	TMPR...	I: 1z8qI	100	57.39	8
pdb1wvb/s72...	ARG1	S2C: S-2-(BORO...	100	56.80	8
pdb1r1h/s518...	1NEP_...	BIR: N-[3-[(1-AMI...	100	55.86	8
pdb1cgl/s375...	MM01...	C: 1cglC	100	55.84	8

Active site of human arginase

Overlay of simvastatin in Human Arginase site:



Similar Surface patches — Dissimilar Surface patches

Supporting Evidence

1) Statins have anti-atherosclerosis effects *independent* of HMG-CoA Reductase inhibition
***(Circulation.* 2003;108:1368)**

2) Statins *increase* NO synthesis and bioavailability
***(J Am Soc Nephrol* 15:1098-1100, 2004)**

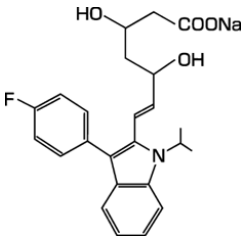
3) Arginase is a novel therapeutic target for atherosclerosis, as it *decreases* the bioavailability of nitric oxide (NO), leading to endothelial dysfunction
(*Curr Hypertens Rep*. 2006 Apr ;8:54-9)

Eidogen-Sertanty

Virtual Target Screening Example: Statins

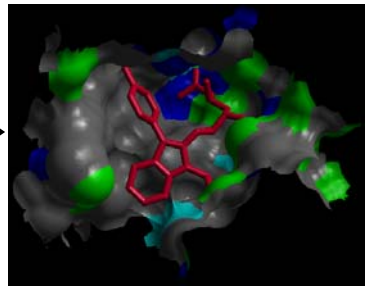
Example 2: Searching for off-targets to explain *adverse* effects of statins

Fluvastatin (Lescol)



Identify binding site

HMG-CoA Reductase Active Site



Perform Site Similarity Search in TIP

Visualize Site alignments in EVE

EVE Target Analyzer

File Export Filtering Ligand Window Help

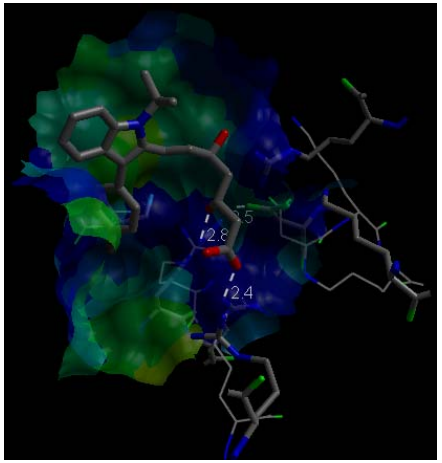
Sequences | Chains | Sites | Binding Modes

Description

Site Name	Locus	Description	...	SiteSorter Similarity	%ID
pdb1hwi/s41...	HMDH_...	115: 7-[3-(4-FLUO...	...	-	-
pdb1rkp/s48...	CN5A_...	Predicted Site	...	84.12	12
pdb1uyl/s49...	HS9A_...	Predicted Site	...	76.18	12
pdb1v4s/s50...	HXK4_...	Predicted Site	...	72.51	12
model46657...	HMGCR	Predicted Site	...	72.46	42
pdb1dgb/s37...	CATA_...	HEM: PROTOPOR...	...	70.13	12
model12228...	HSD11...	NDP: NADPH DIH...	...	68.33	4
pdb1ksw/s4...	SRC_H...	NBS: N6-BENZYL ...	...	68.15	15
pdb1a27/s36...	DHB1_...	Predicted Site	...	65.47	12
pdb1hlg/s41...	LIPG_H...	Predicted Site	...	65.08	8
model21075...	ALOX5	Predicted Site	...	64.88	15
pdb1q6p/s47...	PTN1_...	213: 4'-((2S)-2-(1...	...	64.65	12
pdb1nnl/s44...	SERB_...	Predicted Site	...	64.46	12
pdb1fm6/s39...	PPAT_...	Predicted Site	...	63.00	12
pdb1p0i/s45...	CHLE_...	Predicted Site	...	62.07	12
pdb1pk0/s47...	CYAA_...	EMA: (ADENIN-9-...	...	61.74	12
model5988_...	SLC25...	CXT: CARBOXYA...	...	60.31	12

Carboxyatractyloside-binding site of human mitochondrial Adenosine Nucleotide Translocator 1 (ANT1)

Overlay of fluvastatin in human ANT1 site



Highly similarly charged pockets
(conserved Arg/Lys/Asp residues)
and similar H-bond contacts

Supporting Evidence

1) Statins induce apoptosis in muscle tissue (myotoxicity) via unknown mechanism
(Pharmacol Exp Ther. 2005 Sep;314(3):1032)

2) Statins activate mitochondrial pathway of apoptosis, and this is mediated by the Mitochondrial Permeability Transition Pore (MPTP)
(Br J Pharmacol. 2004 Nov;143(6):715-24.)
(Toxicology. 2006 Feb 15;219(1-3):124-32.)

3) ANT1 is an essential component of the MPTP, and is activated by carboxyatractyloside to permeabilize the mitochondria and induce apoptosis
(Science. 1998 Sep 25;281(5385):2027-31.)
(Cell Biol. 1999 Dec 27;147(7):1493-502.)

Eidogen-Sertanty

Supplemental Slides

EidoSert Products & Services

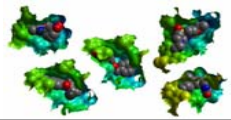
Data | **Knowledge-Based Solution**

TIP™ EVE™

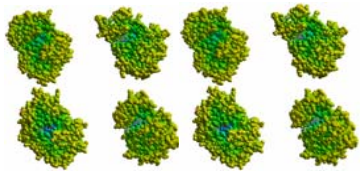
Protein Sequences

Experimental Protein Structures

Ligand-binding sites and binding modes



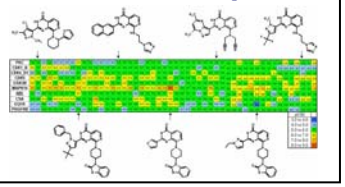
Proteome-wide homology modeling and structural alignment calculations



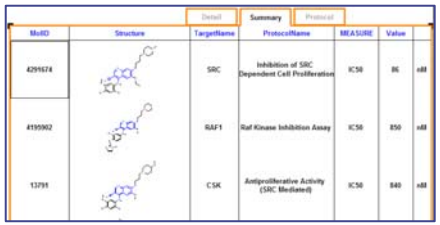
Correlating binding site and binding mode similarities for selectivity & cross-reactivity prediction

KKB™ ARK™

Small molecule ligands and Structure-Activity Relationships

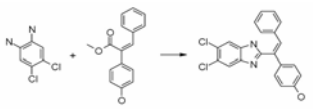


Kinase database and infrastructure for managing SAR information

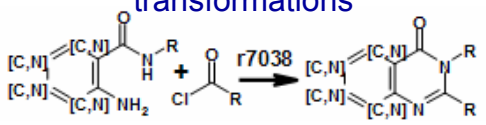


ChIP™

Available chemical building blocks and reactions



Guided *in-silico* focused library design using programmed reaction transformations



DirectDesign™ Collaborations

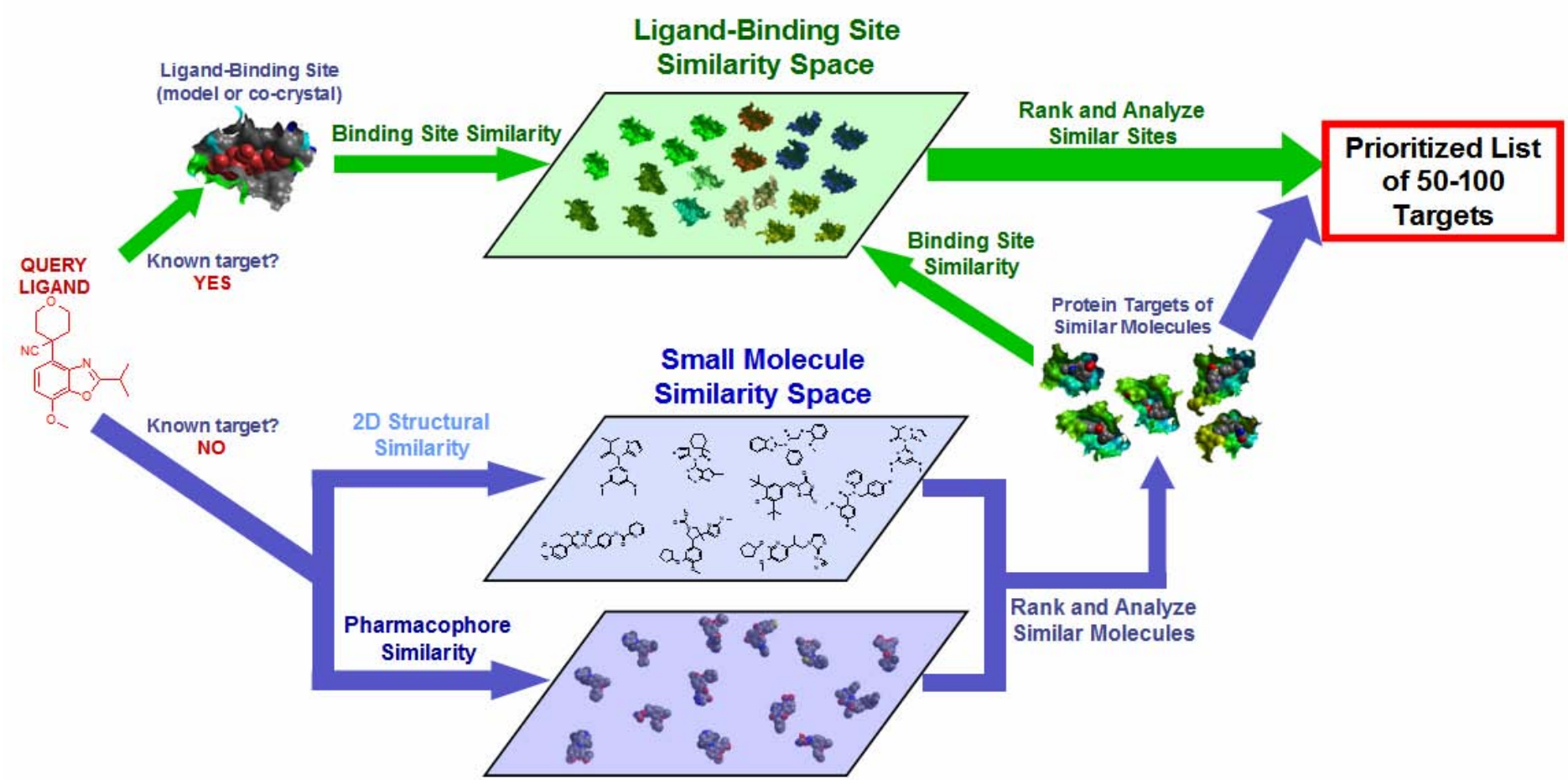
EidogenSertanty

Target Informatics Platform (TIP™)

Comparative Visualizer (EVE™)

Eidogen-Sertanty

In Silico Target Screening ("Target Fishing")



Example Reference: Interrogating the Druggable Genome with Structural Informatics
Kevin Hambly*, Joseph Danzer, Steven Muskal, and Derek A. Debe.
Molecular Diversity, 2006.

Eidogen·Sertanty

TIP Druggable Genome Coverage

5,715 Sequences
From 157 Druggable PFAM Domains

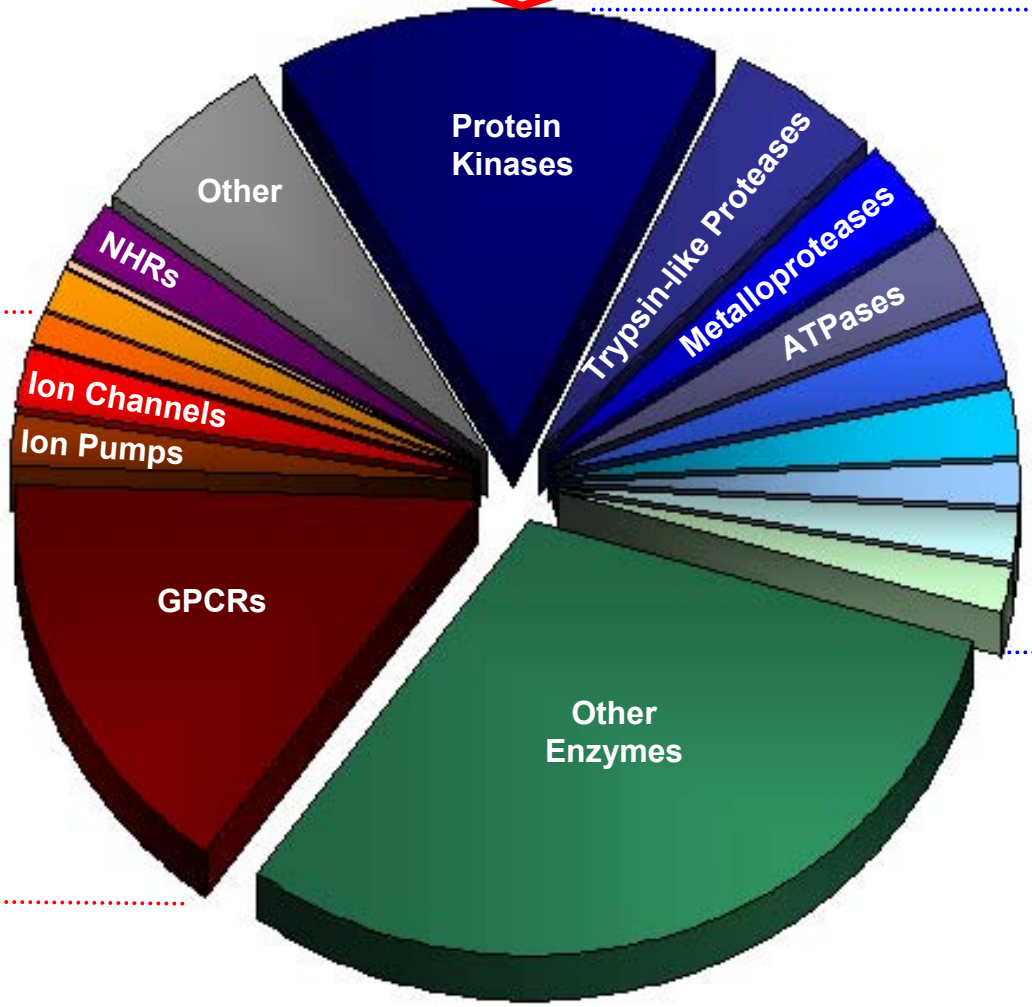
→

4,679 Druggable Targets with one or more structures (81.9% Coverage)

→

1,263 with PDB's
3,416 with STRUCTFAST Models

Major Membrane Protein Targets
983 with TIP structure
(65% coverage)
→2.7% PDBs
→97.3% Models



Major Enzyme Targets
1767 with TIP structure
(95% coverage)
→23.5% PDBs
→76.5% Models

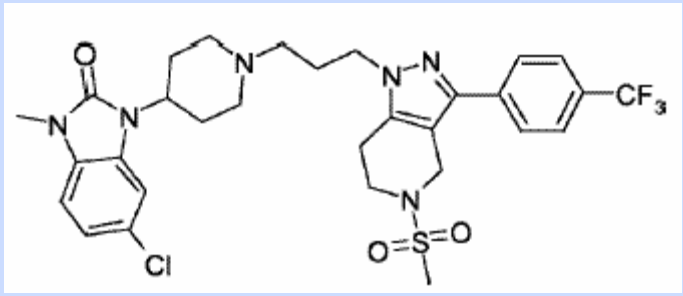
KEY Description	%Structural Coverage
Protein Kinases	98.0%
Trypsin-like Proteases	98.8%
Metalloproteases	91.0%
ATPases	80.1%
Short chain dehydrogenases	95.6%
P450s	96.9%
Phospholipases	82.1%
Cyclophilin-type isomerases	100.0%
Protein Tyr. Phosphatases	100.0%
Other Enzymes	88.7%
GPCRs, Family 1	78.2%
Ion Pumps	33.9%
Ion Channels	69.1%
GPCRs, Family 2	5.3%
Mitochondrial transporters	69.7%
Other Cell Surf. Receptors	62.1%
Other Membrane Transporters	4.5%
Nuclear Hormone Receptors	100.0%
Other Drug Target Domains	78.9%

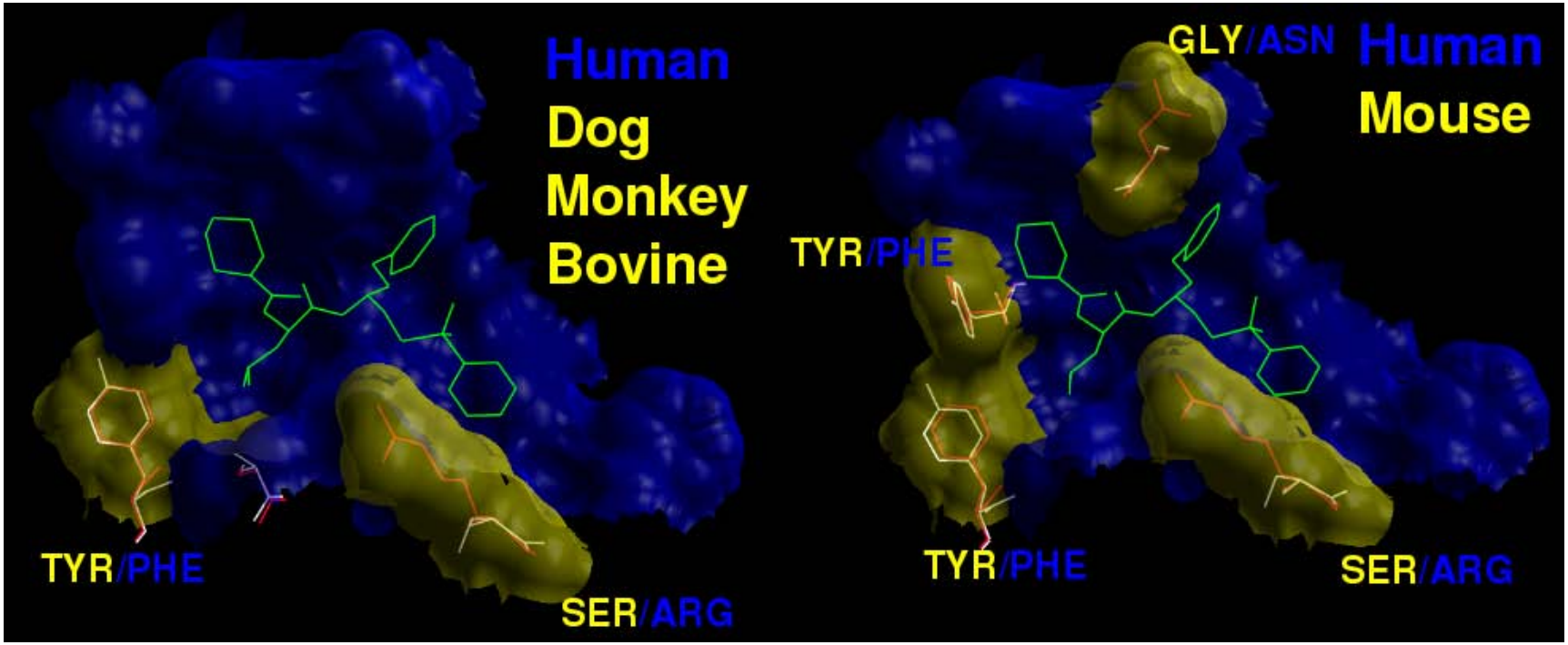
Eidogen·Sertanty

Animal Model Suitability

Cathepsin S Inhibition by JNJ 10329670

Human:	34nM
Dog:	124nM
Monkey:	266nM
Bovine:	411nM
Mouse:	2364nM





R.L. Thurmond, S. Sun, C.A. Sehon et al. *J. Pharmacol. Exp. Ther.* **308**:269-276 (2004).

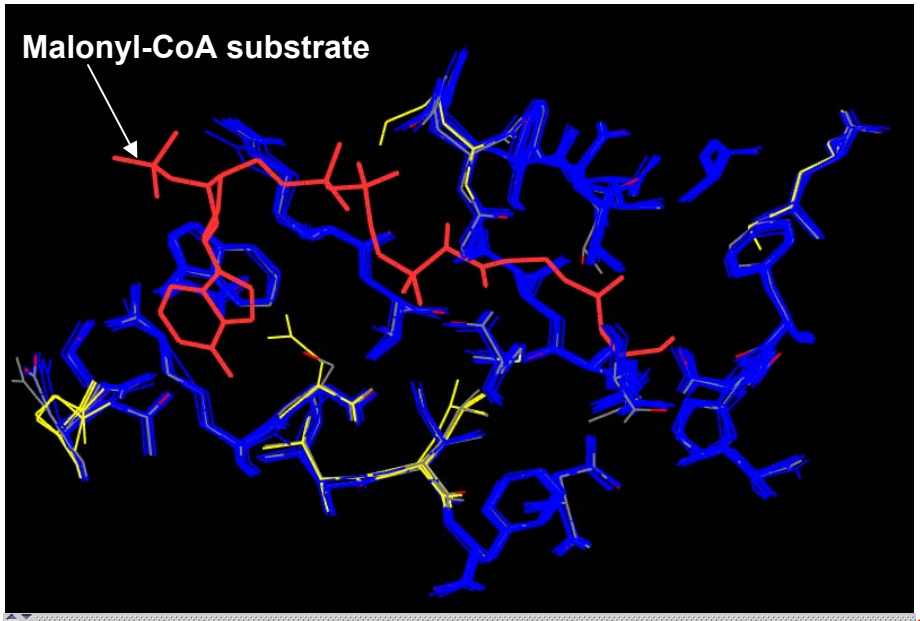
Anti-Infective Spectrum

Comparison of fabH from Several Pathogens

Sequences | Chains | Sites | Site-Ligand Contacts

Description					Site Residue Conservation	Similarity Dendrogram
Site Name	Locus	Description	SiteSorter Similarity	%ID		
pdb1hnj/s4162...	FABH_ECOLI	MLC: MALON...	-	-	.DTS.W.RT.C.RG.IIF.L.M.GN.VF.A.L.H.AN.RI.N.FG	
model147283_...	FABH_SALTY	MLC: MALON...	138.07	96	ETS W RT C RG IIF L M GN VF A L H AN RI N FG	
model147282_...	FABH_SALTI	MLC: MALON...	139.82	96	ETS W RT C RG IIF L M GN VF A L H AN RI N FG	
model147285_...	FABH_SHIFL	MLC: MALON...	140.52	100	DTS W RT C RG IIF L M GN VF A L H AN RI N FG	
model147305_...	FABH_YERPE	MLC: MALON...	141.24	96	DTS W RT C RG ILF L M GN VF A L H AN RI N FG	
model147254_...	FABH_HELPJ	MLC: MALON...	140.76	96	DTS W RT C RG ILF L M GN VF A L H AN RI N FG	
model147239_...	FABH_CHLPN	MLC: MALON...	136.96	82	DTS W RT C RN VLF L M GK VF A M H AN RI N FG	
model147252_...	FABH_HAEIN	MLC: MALON...	137.27	82	DTS W RS C RS VLF L M GN TF A L H AN RI N FG	

Visualize and Overlay



Chains | Chain Alignments | Sites | Site Alignments

Site Name	Locus	Ligand	SiteSorter Sim...	Sequence Positions
pdb1hnj/s416201 ...	FABH_ECOLI	MLC	-	.DTS.W.RT.C.RG.IIF.L.M.GN.VF.A.L.H.AN.RI.N.FG
model147239_9_...	FABH_CHLPN	MLC: MA...	136.96	DTS W RT C RN VLF L M GK VF A M H AN RI N FG

For broad spectrum inhibition, avoid interactions with non-conserved regions in *C. pneumonia* fabH

Eidogen-Sertanty

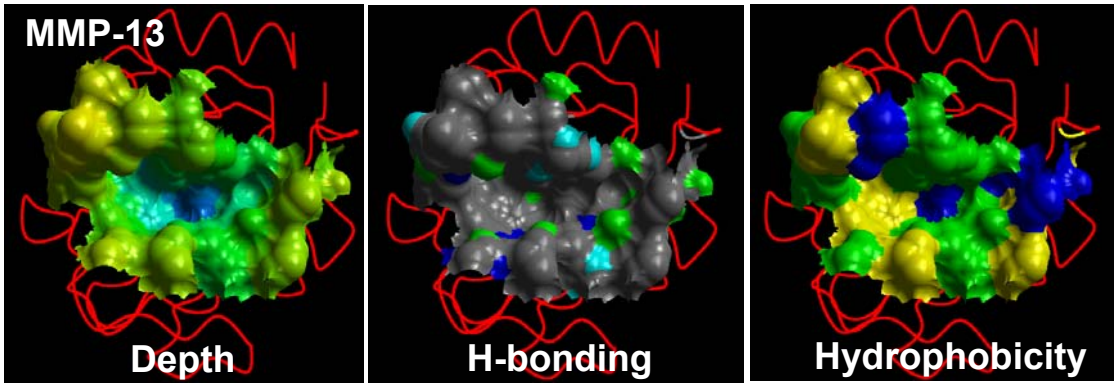
Druggability and Selectivity Analysis

MMP Substrate Site Similarity

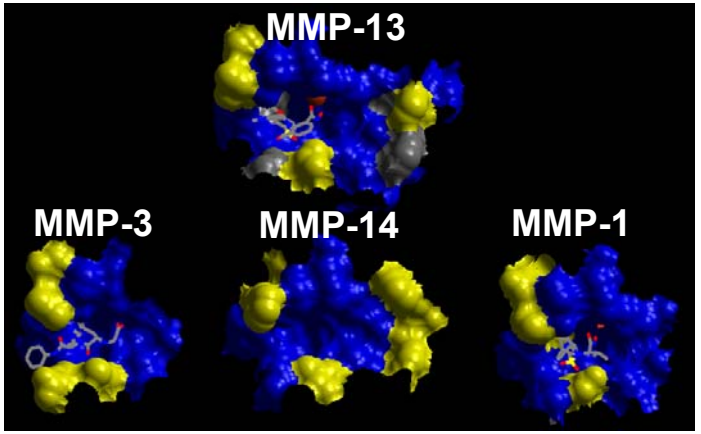
Sequences | Chains | Sites | Site-Ligand Contacts

Description						Site Residue Conservation	Similarity Dendrogram
Site Name	Locus	Description	%Conf	SiteSorter Similarity	%ID		
pdb1sln/s48...	**MM03_HUMAN**	**INH: N-(R-C...**	100	-	-	.GNVLAHA.E.T.LV.HE.H.FH.AL.YPLYHS	
model25148...	**MM12_HUMAN**	HTA: N-[3-(...	100	94.98	63	gGILAHA e T LT HE H gH av FPTYky	
model11126...	**MM15_HUMAN**	HTA: N-[3-(...	100	89.78	63	gGFLAHA e N LV HE H eH ai APFYqw	
pdb1mmq/s4...	**MM07_HUMAN**	**RRS: N4-HY...**	100	101.42	63	GNTLAHA e I YA HE H gH av YPTYgn	
model28047...	**MM20_HUMAN**	HTA: N-[3-(...	100	94.03	63	rGTLAHA e F TV HE H aH al YPTYky	
pdb2tcl/s492...	**MM01_HUMAN**	**RO4: [[1-[N-...**	100	96.38	67	GGNLAHA e Y RV HE H sH al YPSYt-	
pdb1mmb/s4...	**MM08_HUMAN**	**BAT: 4-(N-H...**	100	98.82	67	NGILAHA e Y LV HE H aH aL YPNYaf	
model15220...	**MM24_HUMAN**	BAT: 4-(N-H...	100	98.21	63	GGFLAHA e N LV HE H eH aI APFYq-	
pdb1rm8/s48...	**MM16_HUMAN**	**BAT: 4-(N-H...**	100	97.33	63	GGFLAHA e N LV HE H eH aI APFYq-	
pdb1q3a/s48...	**MM10_HUMAN**	**NGH: N-ISO...**	100	93.65	88	gHSLAHa e t LV HE H fH aL YPLYns	
pdb1fm1/s39...	**MM13_HUMAN**	**WAY: N-HYD...**	100	87.43	63	sGLLAHa e y LV HE H dH aL FPIYTy	
pdb1how/s41...	**MM02_HUMAN**	**I52: N-{4-[(1...**	100	101.51	63	dGLLAHA e y LV HE H eH AL APiYTy	
model14614...	**MM09_HUMAN**	Predicted Site	61	68.88	67	dgllAha e y LV hE h dh Al ypmyrf	
model30578...	**MM14_HUMAN**	Predicted Site	81	67.42	63	ggflaha e n LV HE h eh AI Apfyqw	

Druggability



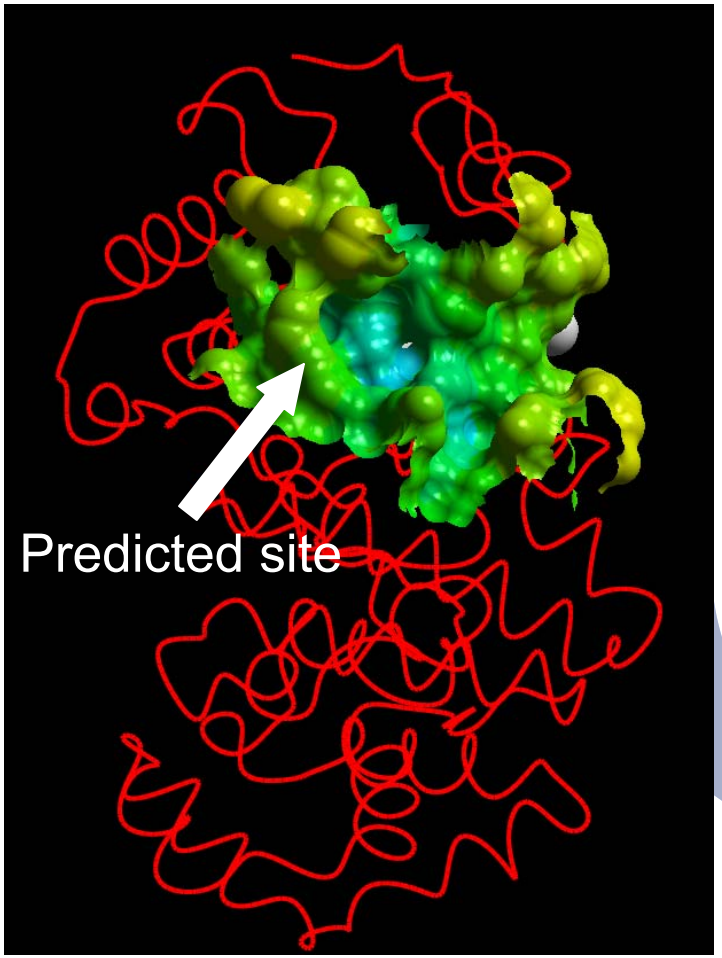
Selectivity



Eidogen-Sertanty

Allosteric Site Opportunities

Allosteric site on p38, behind ATP site



Other kinases with same allosteric site

File Filtering Window Help

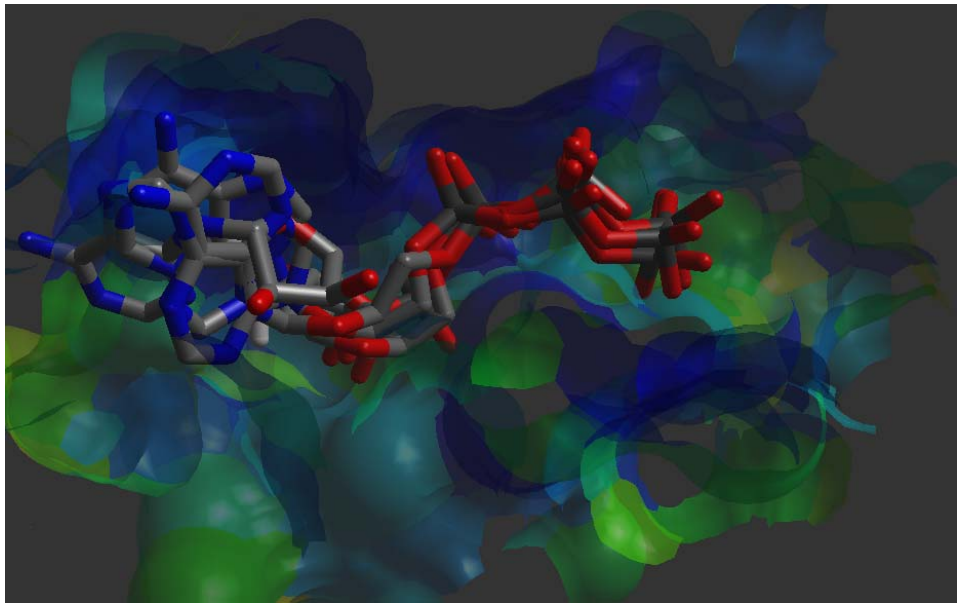
Sequences | Chains | Sites | Site-Ligand Contacts

Description						Site Residue Conservation	Similarity Dendrogram
Site Name	Locus	Description	%Conf	SiteSorter Similarity	%ID		
pdb1bl6/...	MK14_HUMAN	Predicted Site	84	-	-	.R.TGLR.R.LKHMKHENVIGLLD.VTHLMG.Q.R.Y.AVNE.ELK.	
model30...	MK03_HUMAN	Predicted Site	67	145.31	47	r rktr q LLRFRHENVIGIRd vqDLMe Q R Y LINt DLK	
model36...	PASK_HUMAN	Predicted Site	62	111.48	26	k knKE a lsrvehaNIIKvLd vmEKhg q s y ViAe TIK	
model19...	FER_HUMAN	Predicted Site	78	144.42	23	v -ktS k LKQYdHPNIvkLIG IMELVs d a Y LVGE VLK	
model13...	FES_HUMAN	Predicted Site	60	121.81	28	- dntL r LKQYsHPnivrLIG VMELvq d a Y lvtE VlK	
model35...	MPK7_HUMAN	Predicted Site	70	117.37	21	- tghV v LksHDCpyIvqcfg AmELMg a k y lLDE qiK	
pdb1jkt/s...	DAK1_HUMAN	Predicted Site	68	143.09	30	y tglQ s LKEIqHPNVITLHE ILELVA Q n Y mLld RIK	
model39...	S17B_HUMAN	Predicted Site	68	126.06	23	f tgqE a LelakCPRvINLHe ILEyAg q e Y LLSS dIK	
pdb1gjo/...	FGR2_HUMAN	Predicted Site	95	146.27	26	k IavT e MKmIGHKnIiNLLG IVEyAs q r Y lvtE vmK	
pdb1oec/...	FGR2_HUMAN	Predicted Site	85	147.27	26	k IavT e MKMIGHKNIiNLLG IVEYAs q r y lvtE vmK	
model36...	FGR3_HUMAN	Predicted Site	67	108.17	30	r ipvT e MKmIkHKnIiNLLG LvEyaa q r y lvtE vmK	
pdb1agw...	FGR1_HUMAN	Predicted Site	63	146.39	30	r LvTK e MKmgkHKNIiNLLG IVeyAs q r y lVte VmK	
pdb1ksw/...	SRC_HUMAN	Predicted Site	61	122.40	26	- ---R q mkkLrHEKLvqLYa vGEYMs q s y lVGE vCK	
pdb1k3a/...	IG1R_HUMAN	Predicted Site	79	147.31	23	K peTr s MKEFNCHHVvrLLG IMELMt e D Y mvaE tVK	
pdb1irk/s...	INSR_HUMAN	Predicted Site	87	151.50	23	K ieTR s MkgFTCHHVvrLLG VMELMa E D Y Mvah TVK	
model13...	IRR_HUMAN	Predicted Site	88	147.67	21	q esTp s MKAFKCHHVvrLLG IMELMt e D Y mvsq tvK	
pdb1mqb...	EPA2_HUMAN	Predicted Site	66	147.89	21	C keVp g MgQFSHHNIIrLEg ITEYMe g a Y lVNS vcK	
model24...	M3K1_HUMAN	Predicted Site	61	116.62	26	- tgtL r MshLNHpnIIrmLg fiEwMa q r Y lIDS r-L	
model76...	KG3A_HUMAN	Predicted Site	68	160.67	23	a trEL q MRkLDHCNIVrLRY VlEYVP Q R y LVDp VLK	
model13...	CYGF_HUMAN	Predicted Site	74	116.96	28	s lgDF e mkdlrheNInplLg vTEFCs d k y VvDG vLK	

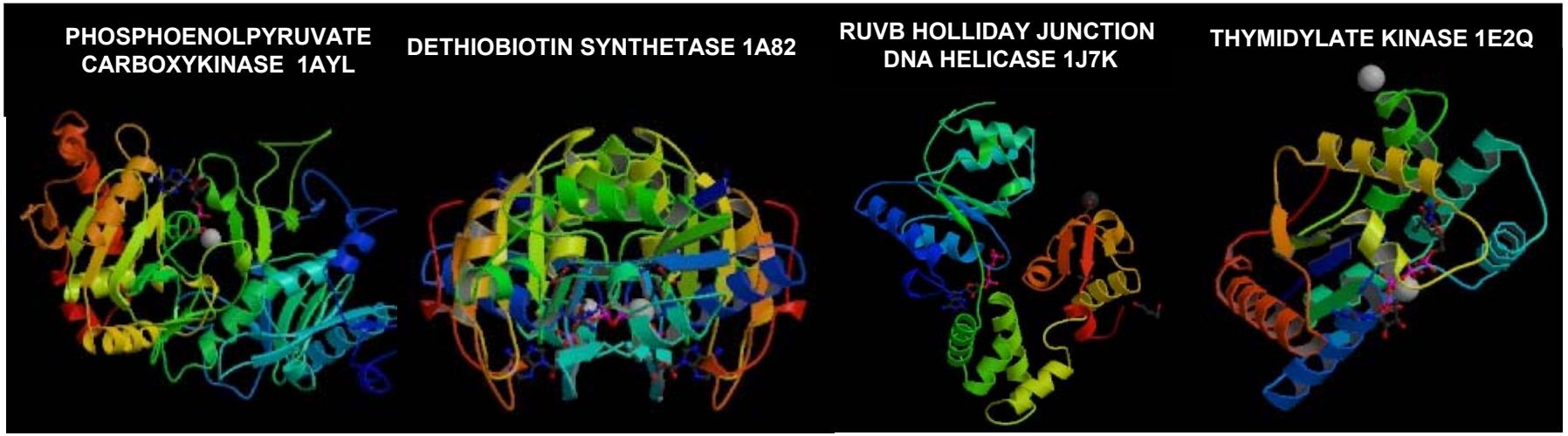
DATABASE SEARCH

Eidogen-Sertanty

SiteSorter Example: ATP Sites

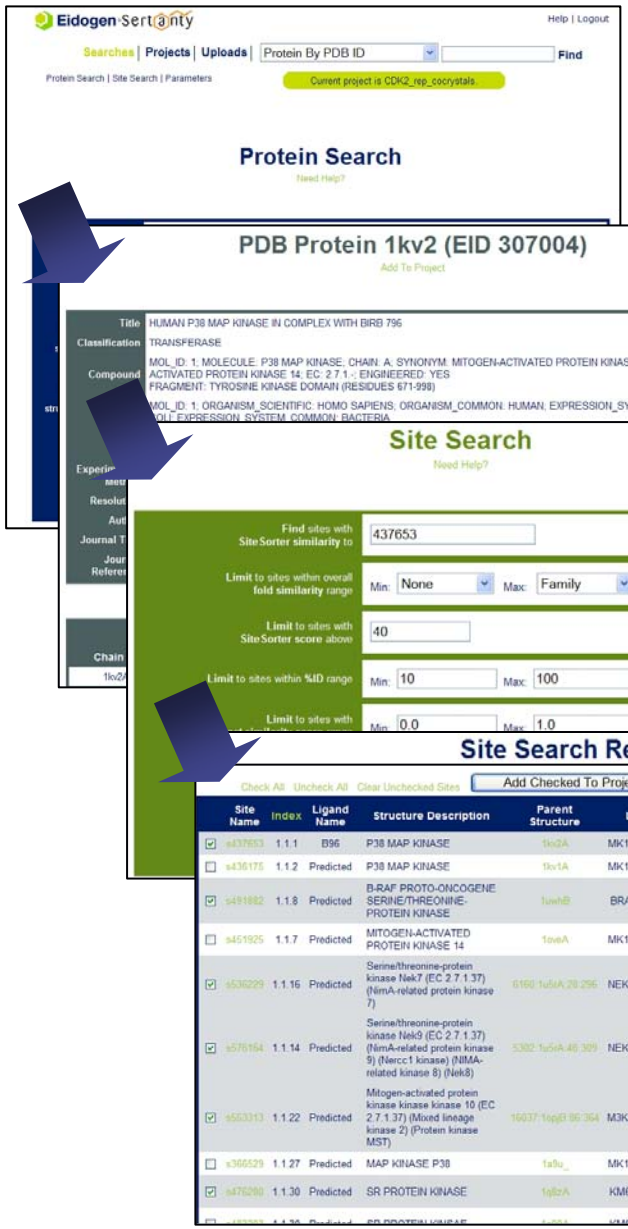


Overlay of ATP binding sites from completely different folds



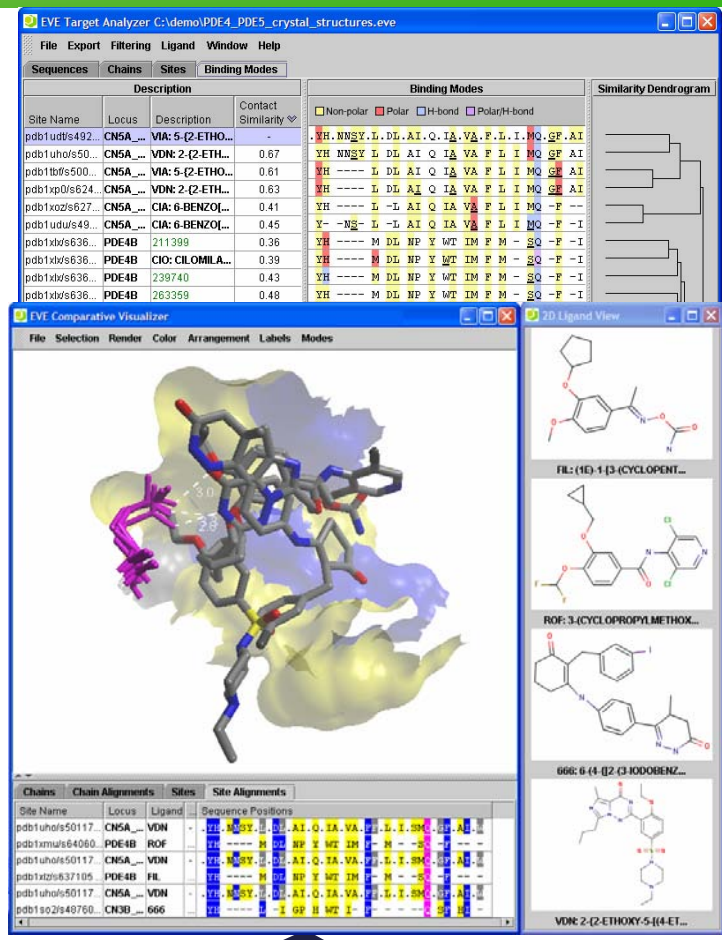
Eidogen Sertanty

EVE: TIP's Fully Integrated Analysis Tool



EVE-2D

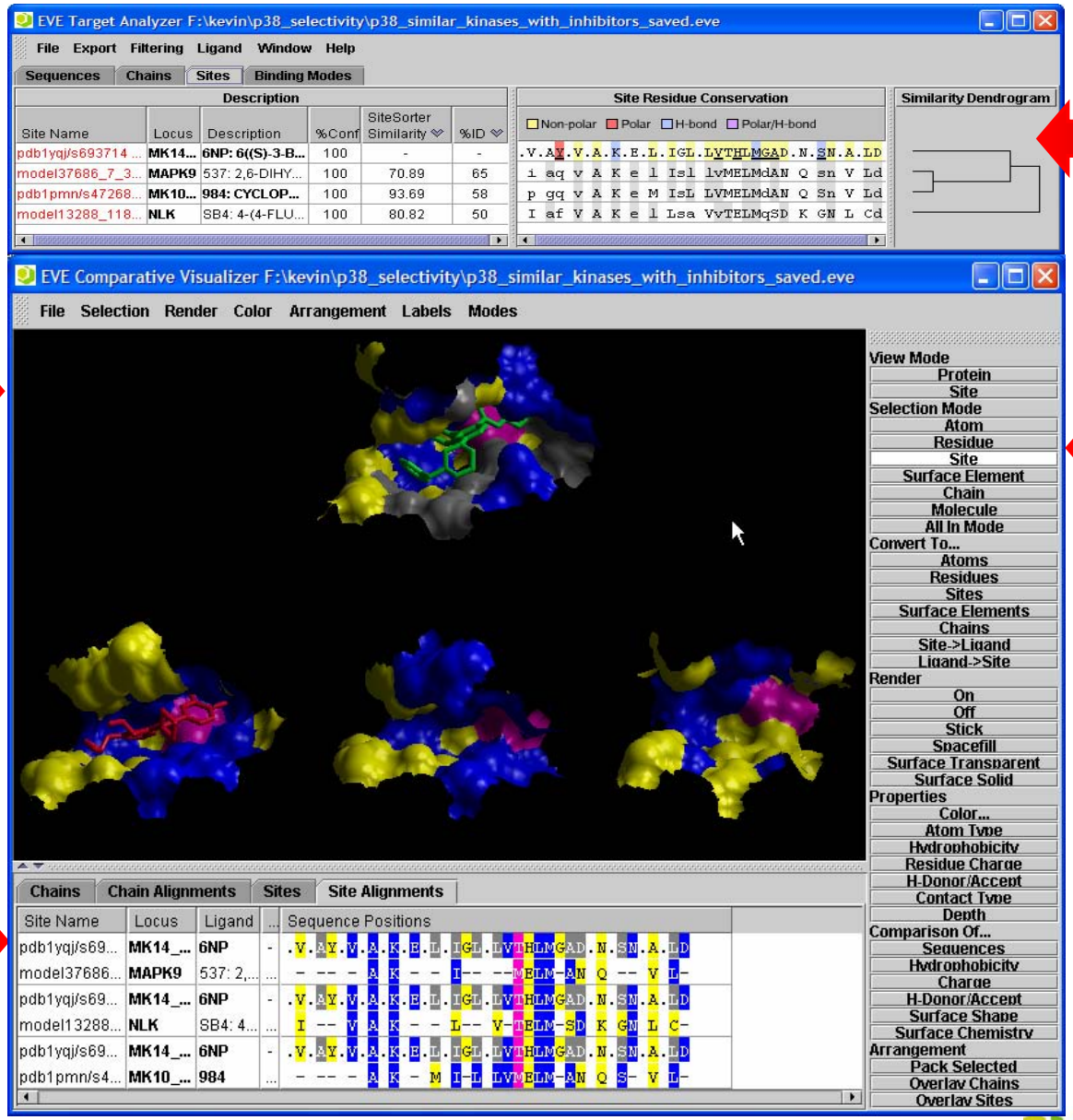
EVE-3D



TIP Project Data

Eidogen-Sertanty

EVE Comparative Visualizer Layout



Similarity Clustering View

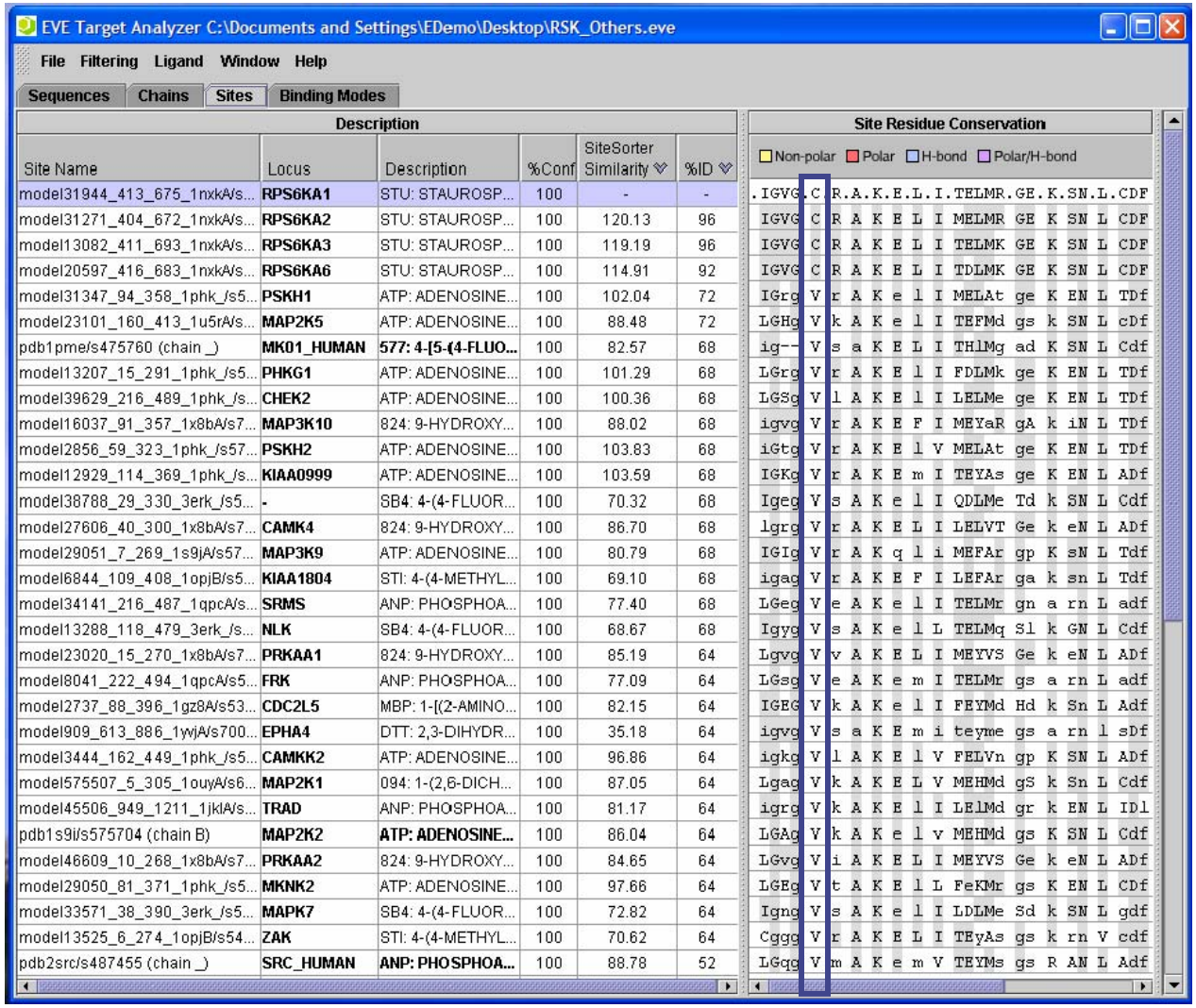
Multiple Structure/Site viewing

Customizable Selection & Analysis Toolbar

Interactive Structure/Site Alignment Window

Eidogen Sertanty

Selectivity Opportunities



Site Name	Locus	Description	%Conf	SiteSorter Similarity	%ID
model31944_413_675_1nxkA/s...	**RPS6KA1**	STU: STAUROSP...	100	-	-
model31271_404_672_1nxkA/s...	**RPS6KA2**	STU: STAUROSP...	100	120.13	96
model13082_411_693_1nxkA/s...	**RPS6KA3**	STU: STAUROSP...	100	119.19	96
model20597_416_683_1nxkA/s...	**RPS6KA6**	STU: STAUROSP...	100	114.91	92
model31347_94_358_1phk_/s5...	**PSKH1**	ATP: ADENOSINE...	100	102.04	72
model23101_160_413_1u5rA/s...	**MAP2K5**	ATP: ADENOSINE...	100	88.48	72
pdb1pme/s475760 (chain _)	**MK01_HUMAN**	**577: 4-[5-(4-FLUO...**	100	82.57	68
model13207_15_291_1phk_/s5...	**PHKG1**	ATP: ADENOSINE...	100	101.29	68
model39629_216_489_1phk_/s...	**CHEK2**	ATP: ADENOSINE...	100	100.36	68
model16037_91_357_1x8bA/s7...	**MAP3K10**	824: 9-HYDROXY...	100	88.02	68
model2856_59_323_1phk_/s57...	**PSKH2**	ATP: ADENOSINE...	100	103.83	68
model12929_114_369_1phk_/s...	**KIAA0999**	ATP: ADENOSINE...	100	103.59	68
model38788_29_330_3erk_/s5...	**-**	SB4: 4-(4-FLUOR...	100	70.32	68
model27606_40_300_1x8bA/s7...	**CAMK4**	824: 9-HYDROXY...	100	86.70	68
model29051_7_269_1s9jA/s57...	**MAP3K9**	ATP: ADENOSINE...	100	80.79	68
model6844_109_408_1opjB/s5...	**KIAA1804**	STI: 4-(4-METHYL...	100	69.10	68
model34141_216_487_1qpcA/s...	**SRMS**	ANP: PHOSPHOA...	100	77.40	68
model13288_118_479_3erk_/s...	**NLK**	SB4: 4-(4-FLUOR...	100	68.67	68
model23020_15_270_1x8bA/s7...	**PRKAA1**	824: 9-HYDROXY...	100	85.19	64
model8041_222_494_1qpcA/s5...	**FRK**	ANP: PHOSPHOA...	100	77.09	64
model2737_88_396_1gz8A/s53...	**CDC2L5**	MBP: 1-[(2-AMINO...	100	82.15	64
model909_613_886_1ywjA/s700...	**EPHA4**	DTT: 2,3-DIHYDR...	100	35.18	64
model3444_162_449_1phk_/s5...	**CAMKK2**	ATP: ADENOSINE...	100	96.86	64
model575507_5_305_1ouyA/s6...	**MAP2K1**	094: 1-(2,6-DICH...	100	87.05	64
model45506_949_1211_1jklA/s...	**TRAD**	ANP: PHOSPHOA...	100	81.17	64
pdb1s9i/s575704 (chain B)	**MAP2K2**	**ATP: ADENOSINE...**	100	86.04	64
model46609_10_268_1x8bA/s7...	**PRKAA2**	824: 9-HYDROXY...	100	84.65	64
model29050_81_371_1phk_/s5...	**MKNK2**	ATP: ADENOSINE...	100	97.66	64
model33571_38_390_3erk_/s5...	**MAPK7**	SB4: 4-(4-FLUOR...	100	72.82	64
model13525_6_274_1opjB/s54...	**ZAK**	STI: 4-(4-METHYL...	100	70.62	64
pdb2src/s487455 (chain _)	**SRC_HUMAN**	**ANP: PHOSPHOA...**	100	88.78	52

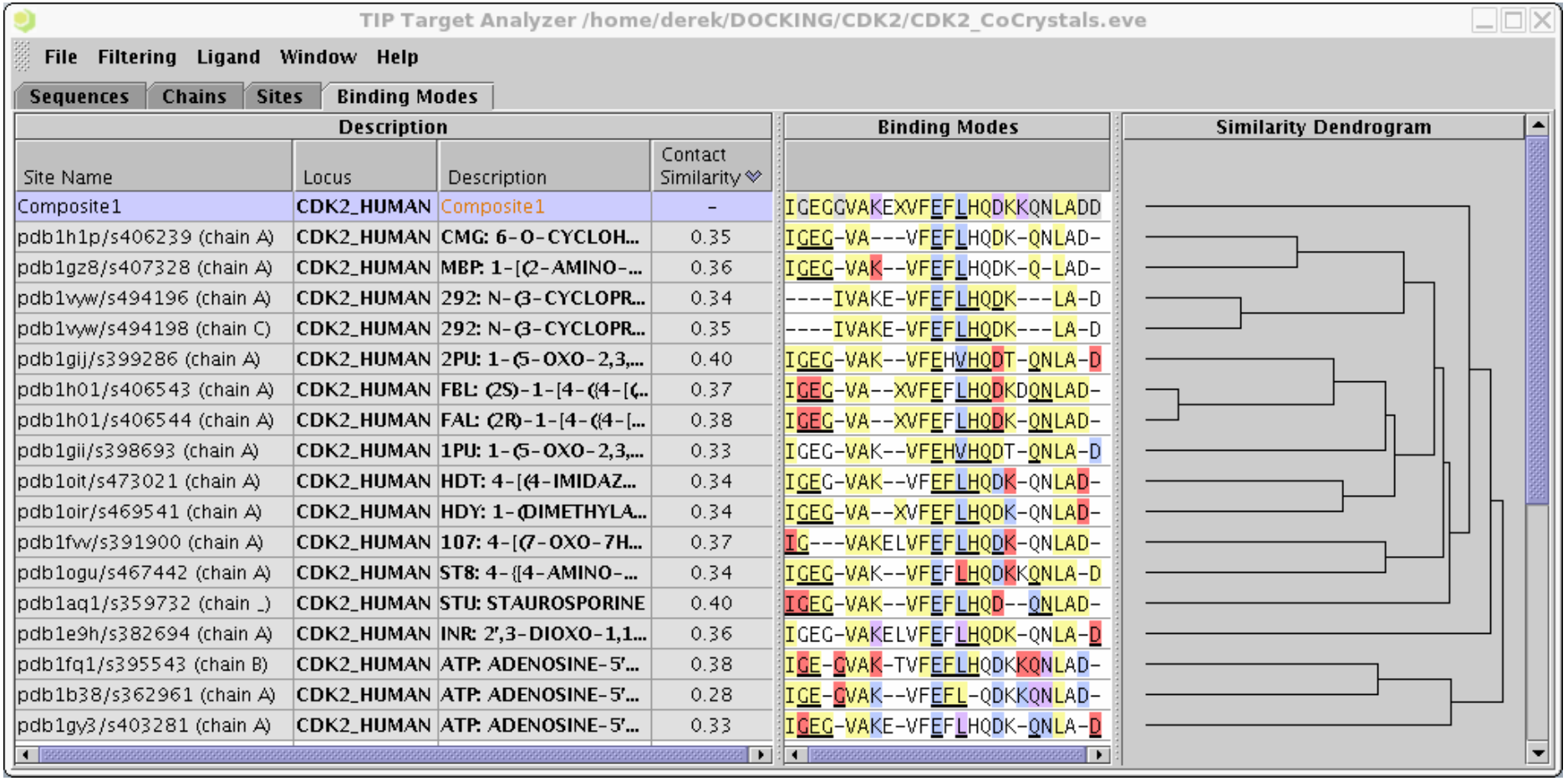
Finds the basis for selectivity in RSK's (p90 ribosomal S6 Kinases)

M.S. Cohen, C. Zhang, K.M. Shokat, J. Taunton, Structural Bioinformatics-Based Design of Selective, Irreversible Kinase Inhibitors, *Science* **308**:1318-1321.

Eidogen Sertanty

cSLiC Binding Mode Analysis

Similar to pSIFT approach developed by Jus Singh's Group At Biogen
see *J. Med. Chem.* **47**, 337 (2004) & *J. Med. Chem.* **48**, 121 (2005).



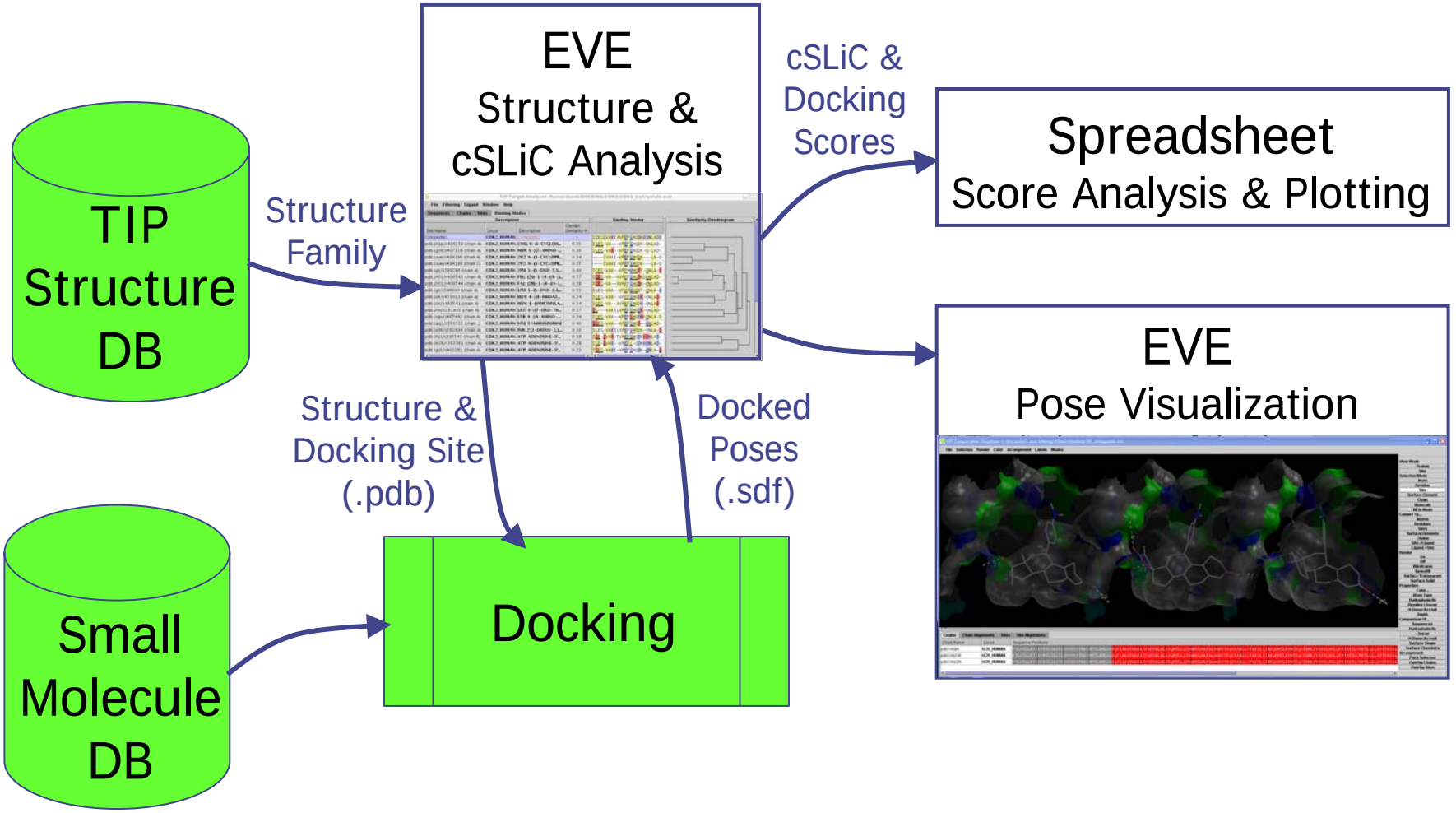
Site Name	Locus	Description	Contact Similarity
Composite1	CDK2_HUMAN	Composite1	-
pdb1h1p/s406239 (chain A)	CDK2_HUMAN	CMG: 6-O-CYCLOH...	0.35
pdb1gz8/s407328 (chain A)	CDK2_HUMAN	MBP: 1-[(2-AMINO-...	0.36
pdb1vyw/s494196 (chain A)	CDK2_HUMAN	292: N-(3-CYCLOPR...	0.34
pdb1vyw/s494198 (chain C)	CDK2_HUMAN	292: N-(3-CYCLOPR...	0.35
pdb1gij/s399286 (chain A)	CDK2_HUMAN	2PU: 1-(5-OXO-2,3,...	0.40
pdb1h01/s406543 (chain A)	CDK2_HUMAN	FBL: (2S)-1-[4-((4-[(...	0.37
pdb1h01/s406544 (chain A)	CDK2_HUMAN	FAL: (2R)-1-[4-((4-[...	0.38
pdb1gii/s398693 (chain A)	CDK2_HUMAN	1PU: 1-(5-OXO-2,3,...	0.33
pdb1oit/s473021 (chain A)	CDK2_HUMAN	HDT: 4-[(4-IMIDAZ...	0.34
pdb1oir/s469541 (chain A)	CDK2_HUMAN	HDY: 1-(DIMETHYLA...	0.34
pdb1fvv/s391900 (chain A)	CDK2_HUMAN	107: 4-[(7-OXO-7H...	0.37
pdb1ogu/s467442 (chain A)	CDK2_HUMAN	ST8: 4-{[4-AMINO-...	0.34
pdb1aq1/s359732 (chain _)	CDK2_HUMAN	STU: STAUROSPORINE	0.40
pdb1e9h/s382694 (chain A)	CDK2_HUMAN	INR: 2',3-DIOXO-1,1...	0.36
pdb1fq1/s395543 (chain B)	CDK2_HUMAN	ATP: ADENOSINE-5'...	0.38
pdb1b38/s362961 (chain A)	CDK2_HUMAN	ATP: ADENOSINE-5'...	0.28
pdb1gy3/s403281 (chain A)	CDK2_HUMAN	ATP: ADENOSINE-5'...	0.33

Binding Modes
IGEGGVAKEXVFEFLHQDKKQNLADD
IGEG-VA---VFEFLHQDK-QNLAD-
IGEG-VAK--VFEFLHQDK-Q-LAD-
----IVAKE-VFEFLHQDK---LA-D
----IVAKE-VFEFLHQDK---LA-D
IGEG-VAK--VFEHVHQDT-QNLA-D
IGEG-VA--XVFEFLHQDKDQNLAD-
IGEG-VA--XVFEFLHQDK-QNLAD-
IGEG-VAK--VFEHVHQDT-QNLA-D
IGEG-VAK--VFEFLHQDK-QNLAD-
IGEG-VA--XVFEFLHQDK-QNLAD-
IG---VAKELVFEFLHQDK-QNLAD-
IGEG-VAK--VFEFLHQDKKQNLA-D
IGEG-VAK--VFEFLHQD--QNLAD-
IGEG-VAKELVFEFLHQDK-QNLA-D
IGE-GVAK-TVFEFLHQDKKQNLAD-
IGE-GVAK--VFEFL-QDKKQNLAD-
IGEG-VAKE-VFEFLHQDK-QNLA-D

cSLiC: Composite Site-Ligand Contacts

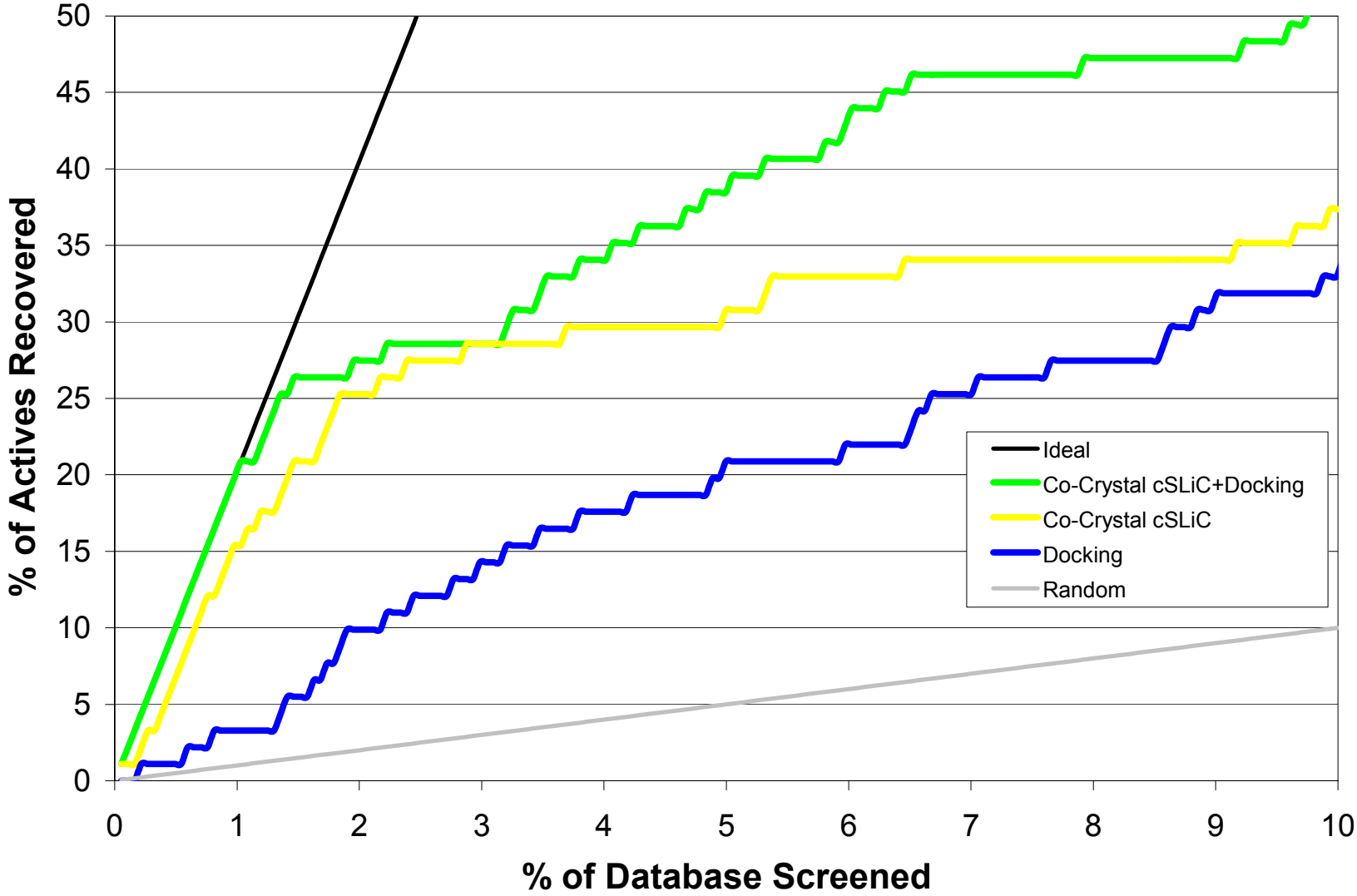
Eidogen Sertanty

cSLiC Binding Mode Analysis



Docking Analysis & Rescoring

cSLiC: CDK2 Enrichment



Dramatically enhances docking-based screening

TIP and EVE Flexibility

TIP Database

Upload new sequences (or genomes)

Define and query custom binding sites

Build STRUCTFAST models from multiple templates

EVE Comparative Visualizer

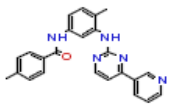
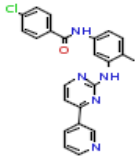
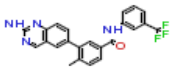
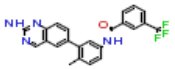
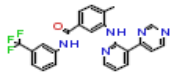
Import your own docked ligands

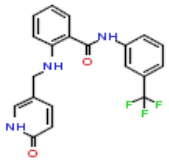
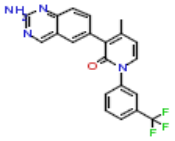
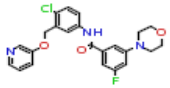
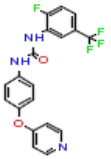
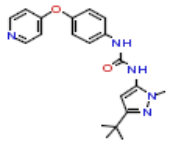
Import your own homology models

Import your own crystal and co-crystal structures

Public data is made available to all users.

Your data and calculation results stay *private*

LC-ID	KKB-ID	Structures	ABL	PDGFR	PDGFRB	JAK3	KDR	LCK	MAPK14	TEK
G2G_STI_12	2082		6.7	8	8					
900_STI_1	2083		6.1	8	8					
7MP_1N8_4	4336542					7.8	9	9.5	8.7	
7MP_1N8_2	4336547					6.8	8.3	9.5	9	
7MP_RAJ_3	4307626						8.4			8.4

LC-ID	KKB-ID	Structures	ABL	PDGFR	PDGFRB	JAK3	KDR	LCK	MAPK14	RAF1
7MP_GIN_4	4344221						7.6			
242_C52_2	4360910									
LI3_L11_1	4198240								7.2	
608_GIG_7	4360457									6.1
KIN_BMU_4	4360452									6.1

Kinase Knowledgebase™ (KKB™)

Eidogen-Sertanty

eScreen QSAR Models

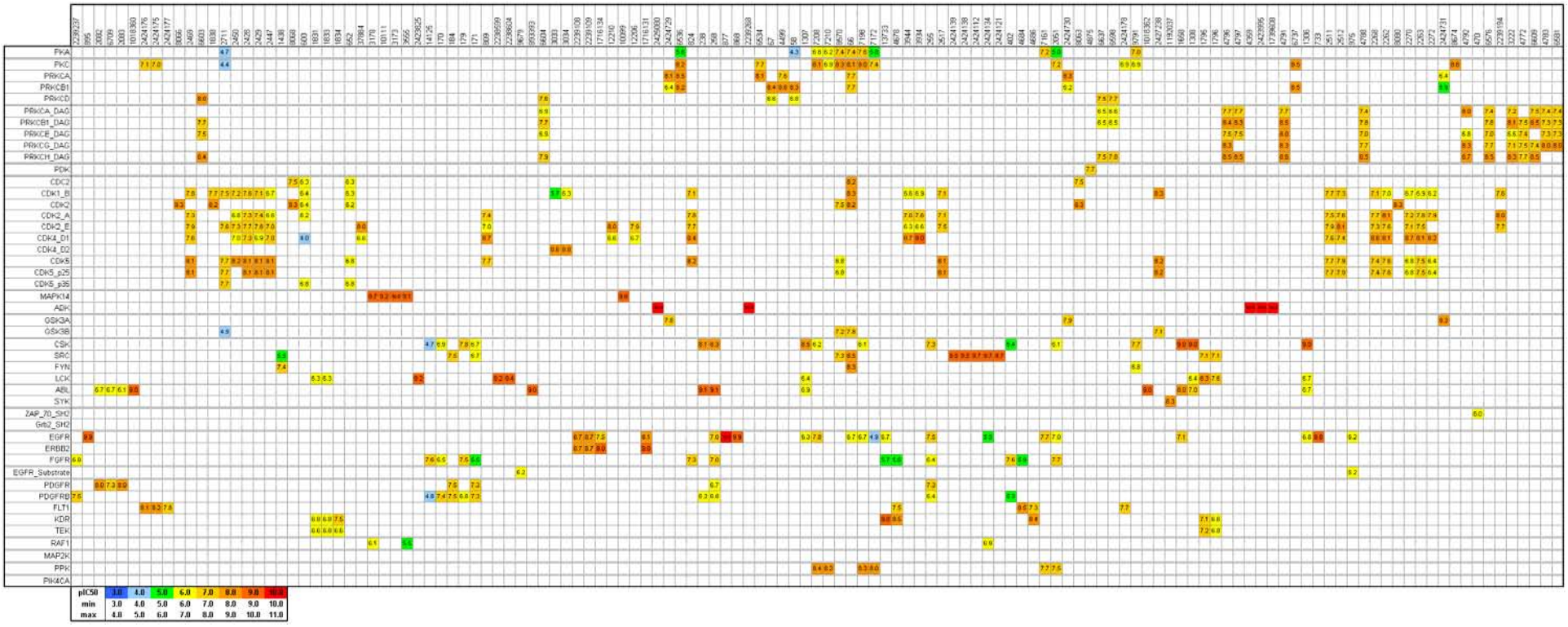
3D-QSAR models built from data annotated in KKB

eScreen Training Set Criteria (all data mined from KKB)

- **Modes of ligand interaction with biological target?**
  - •PKC: ATP site, phosphatidyl serine site, diacyl glycerol site, substrate site
  - •SRC: ATP site, substrate site, SH2-binding domain
- **Substrate for the enzyme?**
  - •Peptide, Carbohydrate, Pyruvate, ADP, Creatine
- **What kind of assay was run?**
  - • Ligands only grouped from purified enzyme assays
- **What is the size and shape of the ligands in the assay protocol?**
  - • Peptides, Small molecules, Heterocycles, unnatural amino acids, peptidomimetics
- **Experimental assay conditions (ATP concentration 5-20 uM)**
- **Is the information suspicious in the paper?**

eScreen Enrichment of Bio-Activity Space

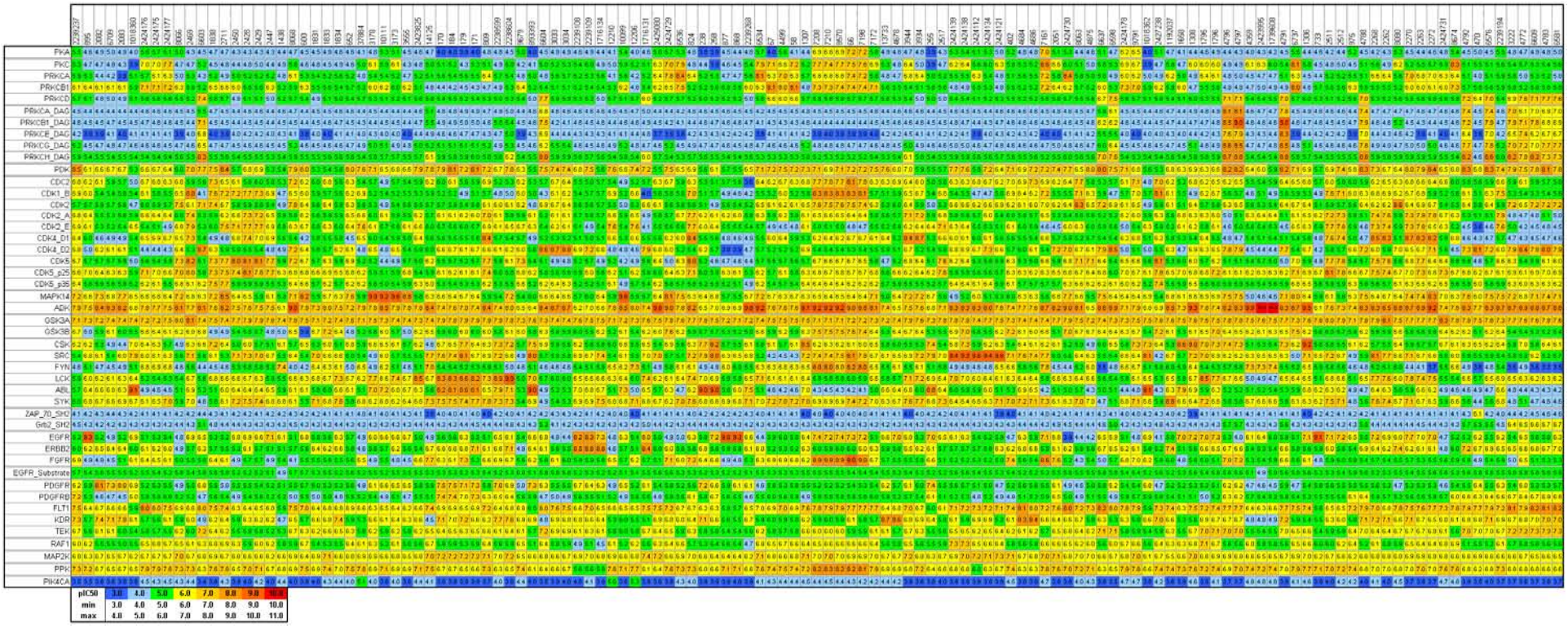
SAR-Based Activity Matrix



EidogenSertanty

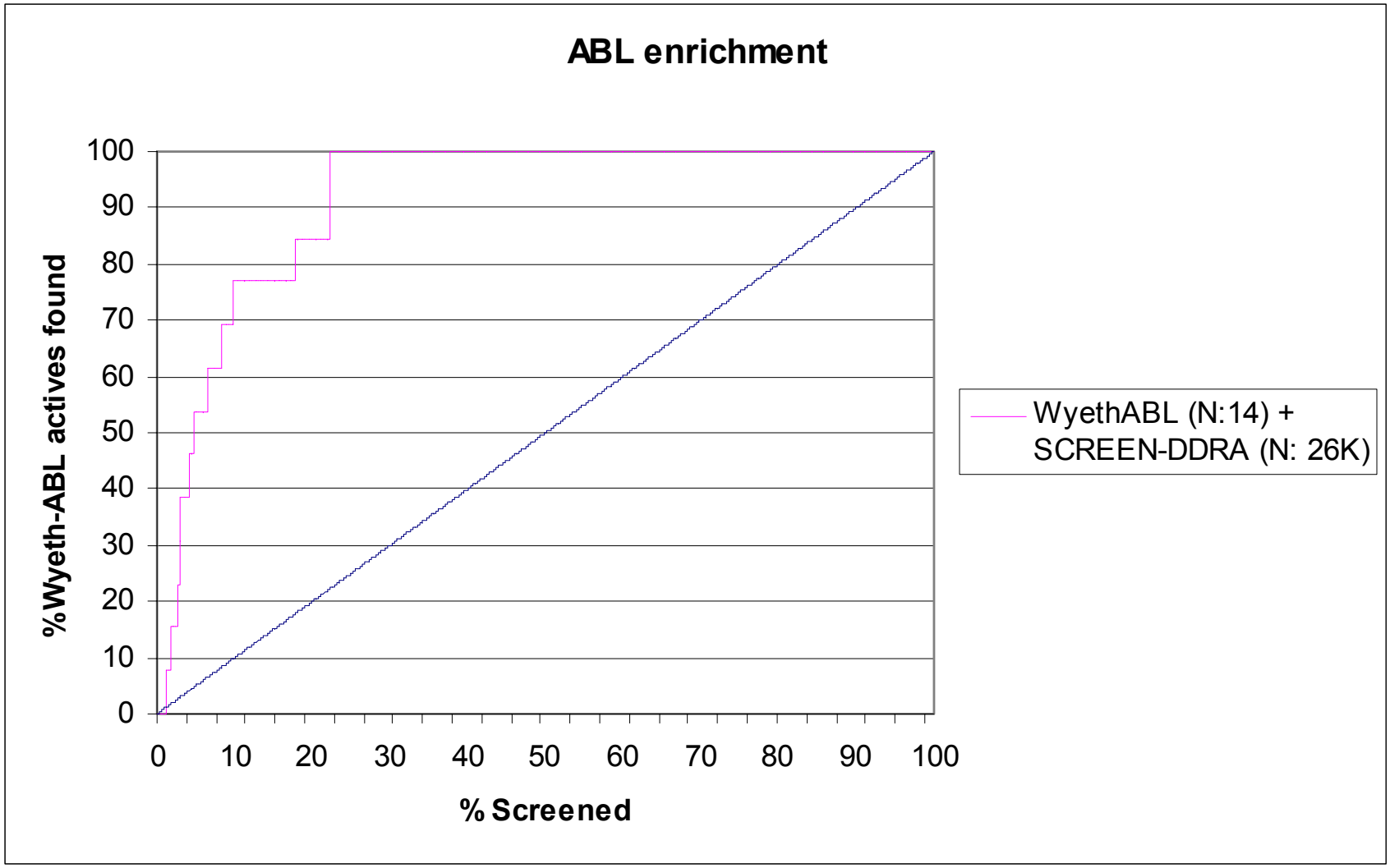
eScreen Enrichment of Bio-Activity Space

eScreen Amplified Activity Matrix



Eidogen-Sertanty

Example Enrichment Study - WyethABL



Example Enrichment Detail - WyethABL

Enrichment analysis for 14 "new" ABL active compounds published by Wyeth (Diane H. Boschelli,* Yanong D. Wang, Steve Johnson, Biqi Wu, Fei Ye, Ana Carolina Barrios Sosa, Jennifer M. Golas, and Frank Boschelli "7-Alkoxy-4-phenylamino-3-quinolinecarbonitriles as Dual Inhibitors of Src and Abl Kinases" J. Med. Chem. 2004, 47, 1599-1601). None used in eABL eScreen development.

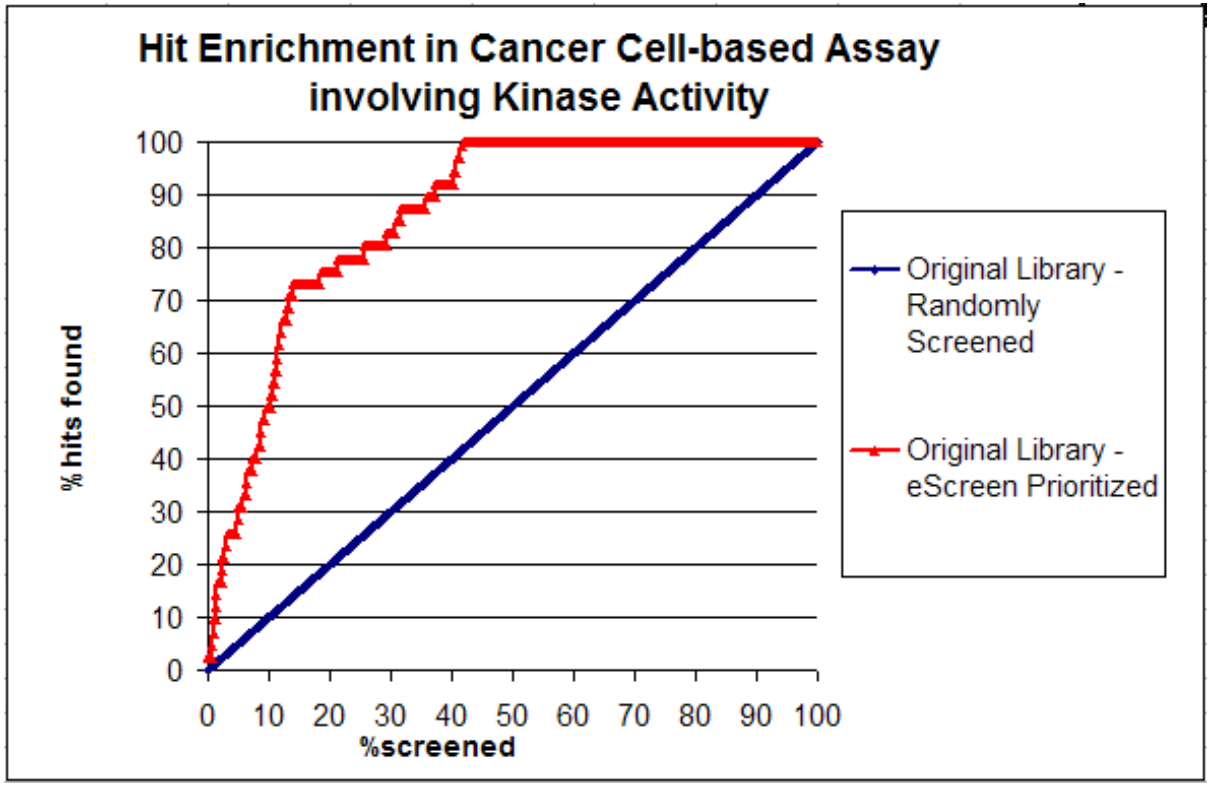
Enrichment greater than 7-fold. 13 of the 14 had Observed pIC50's >= 6.0 and were termed "WyethABL active" and thrown into the SCREEN-DDRA pot of compounds previously described. 1 compound overlap (molid: 13783, SCREEN-DDRAId: 301966) between the 13 WyethABL compounds and the 26K SCREEN-DDRA set (actually denoted as a Src protein kinase inhibitor).

Several known kinase actives in SCREEN-DDRA, many of which seem to bubble to the top in a general eABL ranking. In the rank-ordered compounds, for example, several of the top compounds are noted PDGF, FGF, and SRC inhibitors. Also, no less than 45 of the top-300 in this set are classified as Antineoplastic.

10 of 13 are identified when screening through just 10% of the set and all 13 are identified before screening through 40% of the set.

eScreen Example

- Prioritized 51,000 compound library using a collection of eScreens
- Compounds screened in lung cancer **cell-based assay***



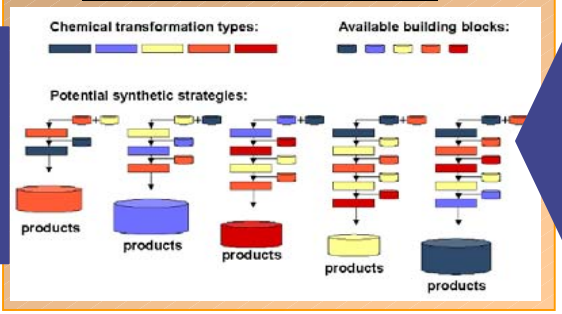
~5-fold enrichment after screening 10% of the library.

***Screening data courtesy of Hakim Djaballah, Memorial Sloan-Kettering Cancer Center**
3,169 of the 51,000 compounds were "active" in at least one of 5 cell-based assays.

Eidogen-Sertanty

ChIP Smart Library Generation

AMPLIFICATION



KNOWLEDGE

Corporate synthetic expertise

Reaction knowledgebase

Target Prioritization

TIP

Tractable Targets

Target Interrogation

ChIP

High Potential Compounds

Activity and Relevancy Assessment

TIP + ARK

Selectivity and Suitability Assessment

Compound Candidates

High Probability Leads and Targets For Selectivity Assay

Novel, Tractable Compound Libraries

Guided chemistry evolution directed by family, target, prior-art
Consistent with accessible synthesis methods and commercially available starting materials.

Eidogen-Sertanty

Reaction Content – Enumerated From Published Reactions

- ~1,300 journal articles from 80 journals and ~80 patents
  - Publications cover parallel, solid-phase and solution-phase methods
- ~15,000 generic reactions → ~1.8 million products
  - Reactions with their starting materials correspond to ~2.8 million specific reactions
  - Reaction exports available in RDFile/SMIRK and building-blocks/product molecules in SDFile/SMI formats
- Hierarchically organized chemistry content covers a variety of synthetic methodology
  - Solution phase reactions
  - Solid-phase reactions
  - Polymer-supported solution-phase reactions

Enhanced Enumeration-Ready ("ChIP-able") Reaction Content

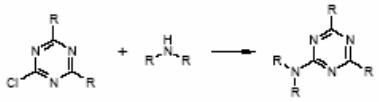
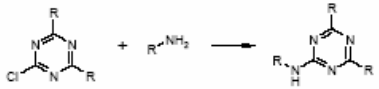
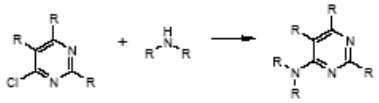
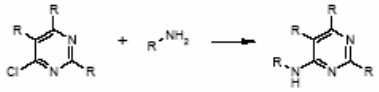
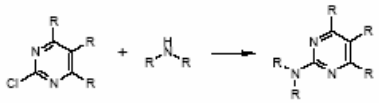
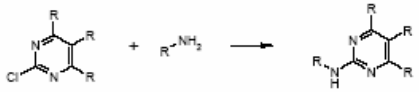
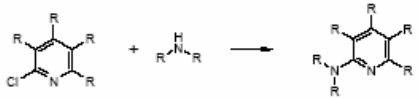
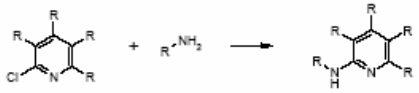
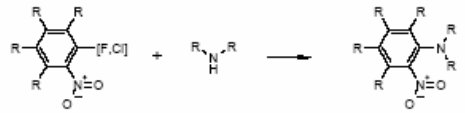
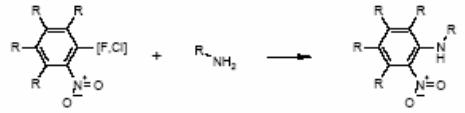
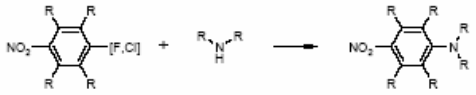
- Reaction transforms – generic reactions in SMIRKS format
  - Associated incompatibility SMARTS filters ("required" and "exclude")
  - Introspective reactivity filters are included in the SMIRKS representation

- Example reaction types

  Nucleophilic Aromatic Substitution Reactions
  Pd-Catalyzed Aromatic Substitutions
  Functional Group Transformations
  Amine Acylation Reactions (Amides / Carbamates)
  Amine Acylation Reactions (Ureas / Thioureas)
  Formation of Diverse Heterocyclic Systems
  Michael addition, cyclo-condenzation (formation of 4-quinolinones)
  Formation of Thioimidazoles
  Standard Deprotection Steps
  Robinson Annulation
  Diels-Alder Reaction (26 representations)
  Fisher-Indole Synthesis

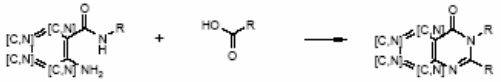
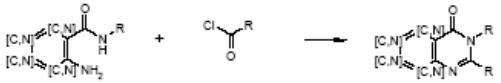
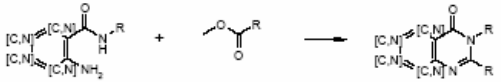
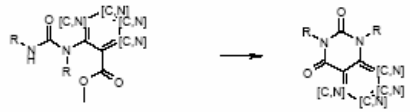
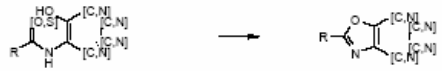
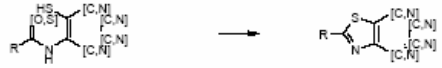
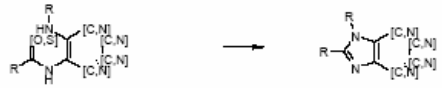
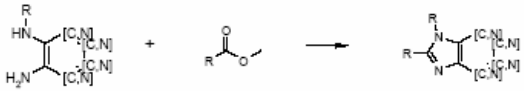
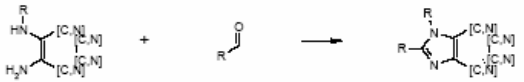
- Filters of reactive functionalities/undesired motifs in SMARTS

ChIP-able Nucleophilic Aromatic Substitutions

7000		aliphatic primary or secondary amine
7001		aromatic primary amine
7002		aliphatic primary and secondary amine
7003		aromatic primary amine
7004		2-chloropyrimidine, but not 4-chloropyrimidine aliphatic primary or secondary amine
7005		2-chloropyrimidine, but not 4-chloropyrimidine primary aromatic amine
7006		aliphatic primary or secondary amine
7007		aromatic primary amine
7008		aliphatic primary or secondary amine
7009		aromatic primary amine
7010		amine aliphatic primary or secondary

Eidogen·Sertanty

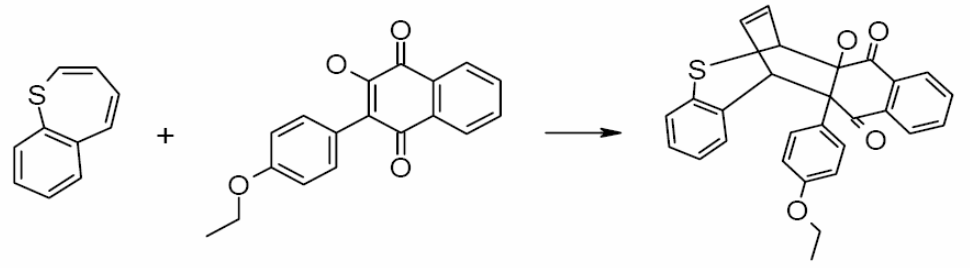
ChIP-able Diverse Heterocycles

7037		any 2-amino arylamide not in a ring; any carboxylic acid; excludes acylators, alkylators, nucleophiles, other acids, activated aryl halides, aryl bromides/iodides
7038		any 2-amino arylamide not in a ring; carboxylic acid chloride, no chloroformate, carbamoyl chloride, etc.; excludes other acylators, alkylators, nucleophiles, acids, activated aryl halides, aryl bromides/iodides
7039		any 2-amino arylamide not in a ring; carboxylic acid methyl ester, does not allow any other ester, which is too restrictive; no differentiation between different ester reactivity; excludes acylators, alkylators, nucleophiles, acids, activated aryl halides, aryl bromides/iodides
7040		any 2-ureido aryl methyl carboxylate not in a ring; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides
7041		any 2-ureido aryl carboxylic acid not in a ring; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides
7066		any 2-hydroxyaryl-(NH)-amide, thioamide, urea, thiourea, etc.; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides
7067		any 2-mercaptoaryl-(NH)-amide, thioamide, urea, thiourea, etc.; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromides/iodides
7068		any 2-aminoaryl-(NH)-amide, thioamide, urea, thiourea, etc.; not amido aryl amide; excludes, acylators, alkylators, acids, nucleophiles, activated aryl halides, aryl bromide/iodide
7069		1,2-diamino aryl (amino-2-arylamine); carboxylic acid methyl ester, does not allow any other ester, which is too restrictive; no differentiation between different ester reactivity; exclude, nucleophiles, acids, acylators, alkylators, activated aryl halides, aryl bromide/iodide, aldehydes, other esters
7070		1,2-diamino aryl; aldehyde; exclude, nucleophiles, acids, acylators, alkylators, activated aryl halides, aryl bromide/iodide

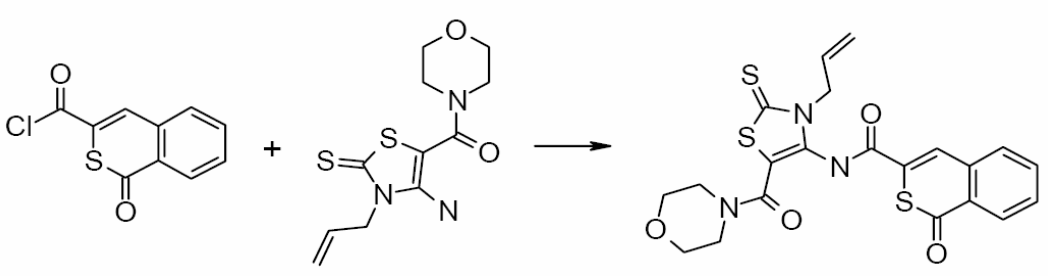
Eidogen-Sertanty

Leveraging Building Block Complexity

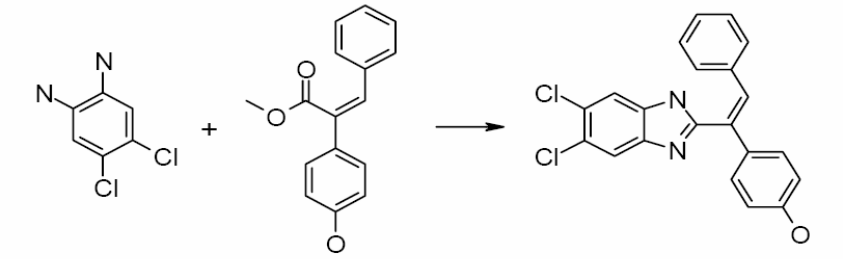
Diels-Alder



Amide Formation

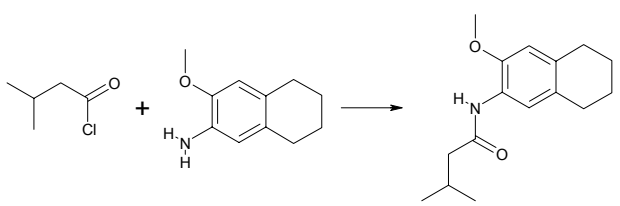


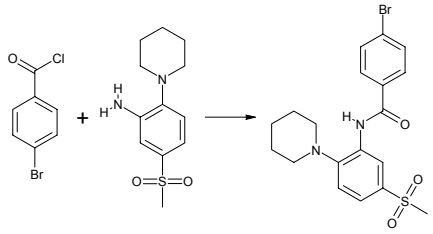
Benzimidazole Formation

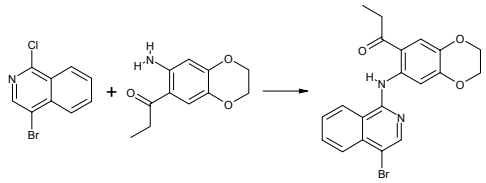


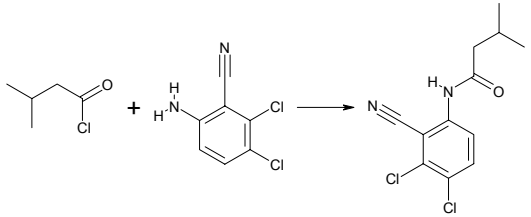
A landslide of interesting pharmacophores, all just one Rxn step away

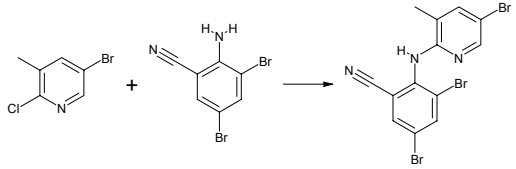
Substantial One-Step Diversity

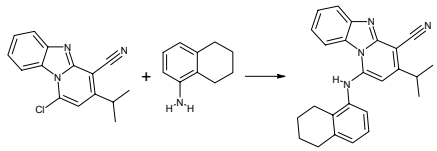


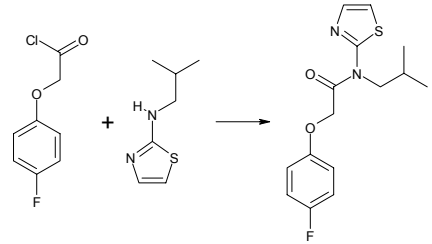


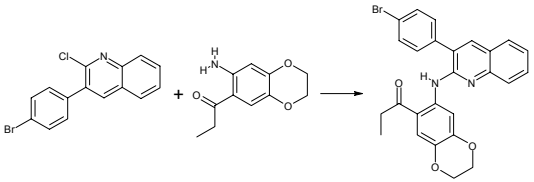


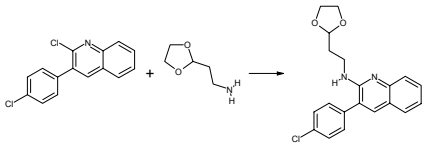


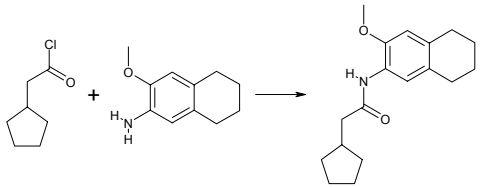


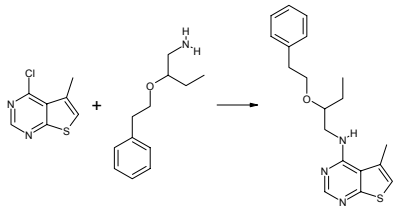


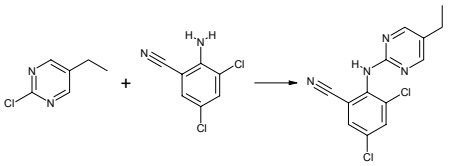


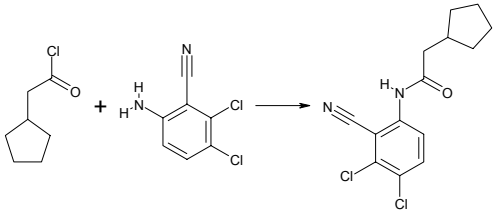


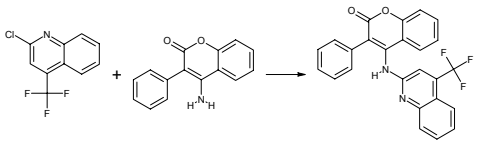


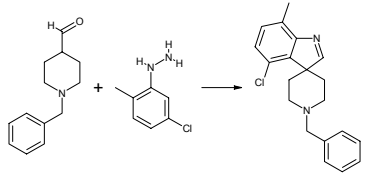












Eidogen-Sertanty

Good Filtering Is Critical…

Each reaction transform is encoded with the information necessary to prevent undesirable products.

Side reactions are avoided using…

Incompatibility Filters to analyze available starting materials to exclude multiple reactive functional groups

Synthetic infeasibility is avoided using…

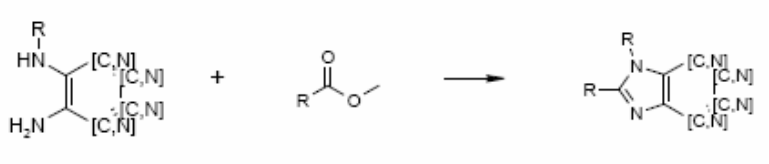
Introspective filters to analyze the reaction center environment to exclude reactants with groups that prevent successful synthesis, e.g. nucleophilic amine, activated aromatic chloride,...

Reactive product molecules are removed using…

Global "bad-frag" filters to avoid reactive functionality or undesired motifs, e.g. acylators, undesired elements/isotopes...

Filtering Example

How The Chemist Thinks



Required

1,2-diaminoaryl (amino-2-arylamine); carboxylic acid methylester

Not Allowed

nucleophiles, acids, acylators, alkylators, activated arylhalides, aryl bromide/iodide, aldehydes, any other esters (no differentiation between reactivities of different esters)

How The Computer Thinks

RXN Smirk with “introspective filter”

[N;$([N;!H0](c1[#6,#7][#6,#7][#6,#7][#6,#7]c1[N;!H0;!H1])[A,a]);!$([N+]);!$(NC=,#[!#6]);!$(NC=,#[#6]);!$(N[!#6;!#1]):4]([c:11]1[c:10]([N;!H0;!H1:3]([H])[H])[a:7][a:5][a:6][a:8]1)([A,a:1])[H].[CH3]O[C;$(C(O[CH3])(=O)[A,a]);!$(C(=O)(O[CH3])[!#6]):9](=O)[A,a:2]>>[n:3]1[c:9]([n:4]([c:11]2[c:10]1[a:7][a:5][a:6][a:8]2)[A,a:1])[A,a:2]

e.g. Building Block Required Filters (r7069.1, r7069.2)

c(c([NH2;v3])[a])([a])[N;!H0;!$([N+]);!$(NC=,#[!#6]);!$(NC=,#[#6]);!$(N[!#6])]

[C;$(C(=O)O[CH3]);!$(C(=O)(O[CH3])[!#6])]

e.g. Building Block Exclude Filters (r7069.1)

([N;!H0;$(NC);!$([N+]);!$(NC=,#[!#6]);!$(NC=,#[#6]);!$(N[!#6]);!$(Nc)]) OR ([N;!H0;$(N[N;$(N[#6]);!$(NC=,#[!#6])]);!$(NC=,#[!#6])])

OR (C([NH2])=[NH]) OR ([S;$([SH]),$([S-])])

([S;$(S(=O)[OH])]) OR ([C;$(C(=O)[OH]);!$(C(=O)([OH])[!#6])])

([C;$(C(=[O,N,S])[O,S]C(=[O,N,S]))]) OR ([C;$(C(=[O,N,S])[F,Cl,Br])]) OR ([C;$(C(=[O,S])=N)]) OR ([S;$(S(=O)[Cl,Br,F,I])])

([C;$(C[Br,I]);!$(C=,#[A])]) OR ([C;$(COS(=O)(=O))])

([c;$(c1([Cl,Br,F,I])nc[n,c][c,n][c,n]1)]) OR ([c;$(c1([F,Cl])c([N+](=O)[O-])cccc1),$(c1([F,Cl])ccc([N+](=O)[O-])cc1)])

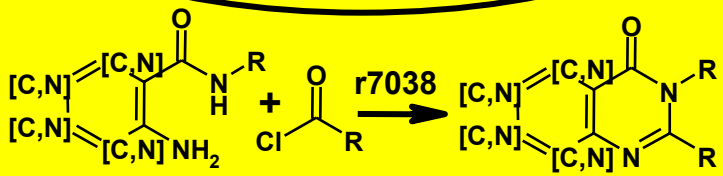
[c;$(c[Br,I])]

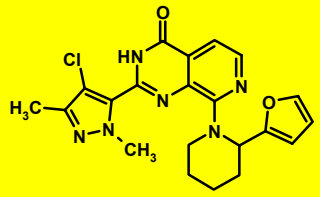
[C;$(C(=O)O);!$(C(=O)(O)[!#6]);!$(C(=O)O[!#6]);!$(C(=O)([OH]));!$(C(=O)([O-]));!$(C(=O)OC=,#[!#6])]

[C;!H0;$(C=O);!$(C(=O)[!#6]);!$(C(=O)=[A])]

De-"Know"-vo Drug Design

ChIP™
Chemical Intelligence Platform



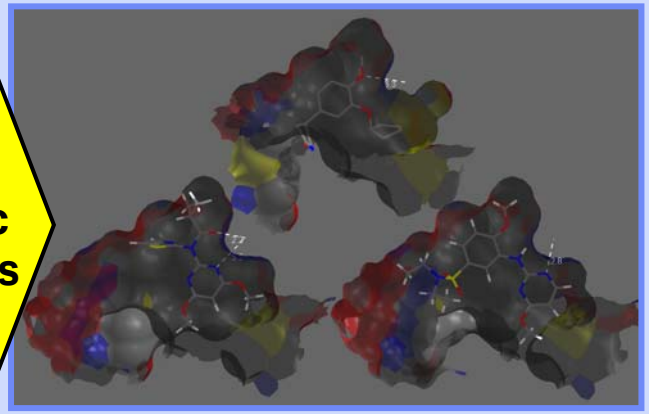


Active Seeds From SAR Database
K_i = 27nM

Synthetically Focused, Active Guided Lead Design

Target Specific Libraries

KKB™ / TIP™
Kinase Knowledgebase Target Informatics Platform



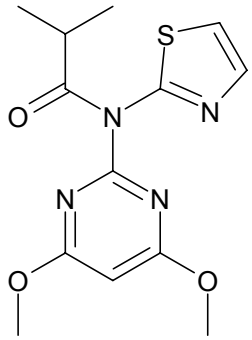
Ligand and Target Analyses

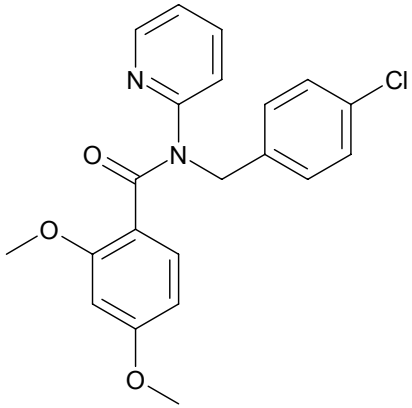
Design Candidates

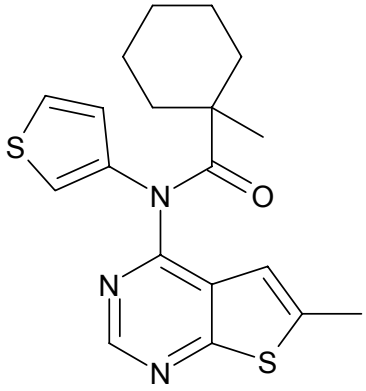
Synthetic, Activity, and Structural Knowledge Leveraged Simultaneously

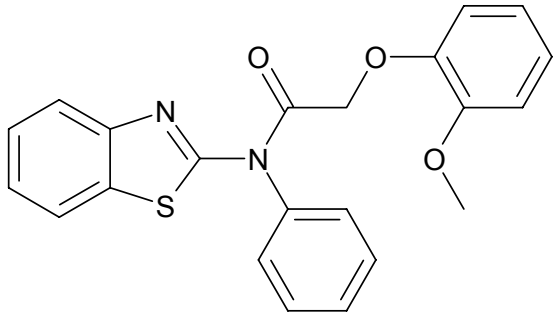
EidogenSertanty

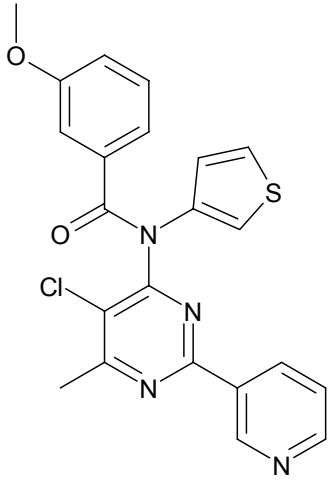
Other ChIP Generated PDE-IV “Me-Too’s”

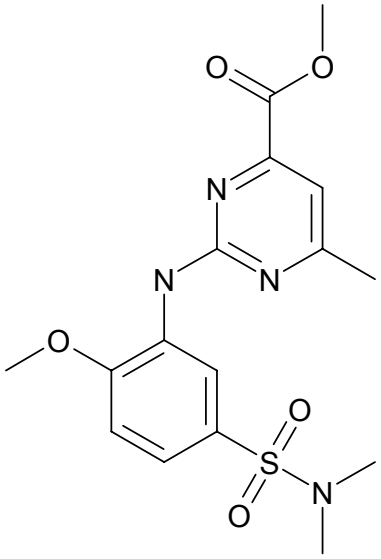


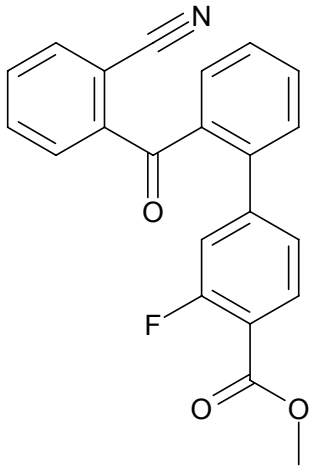


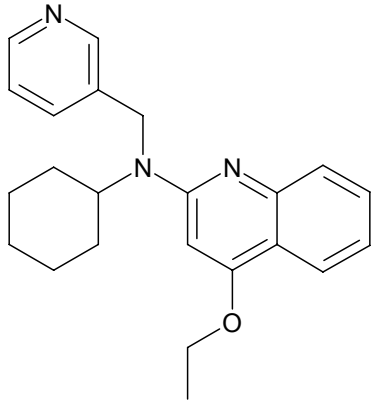










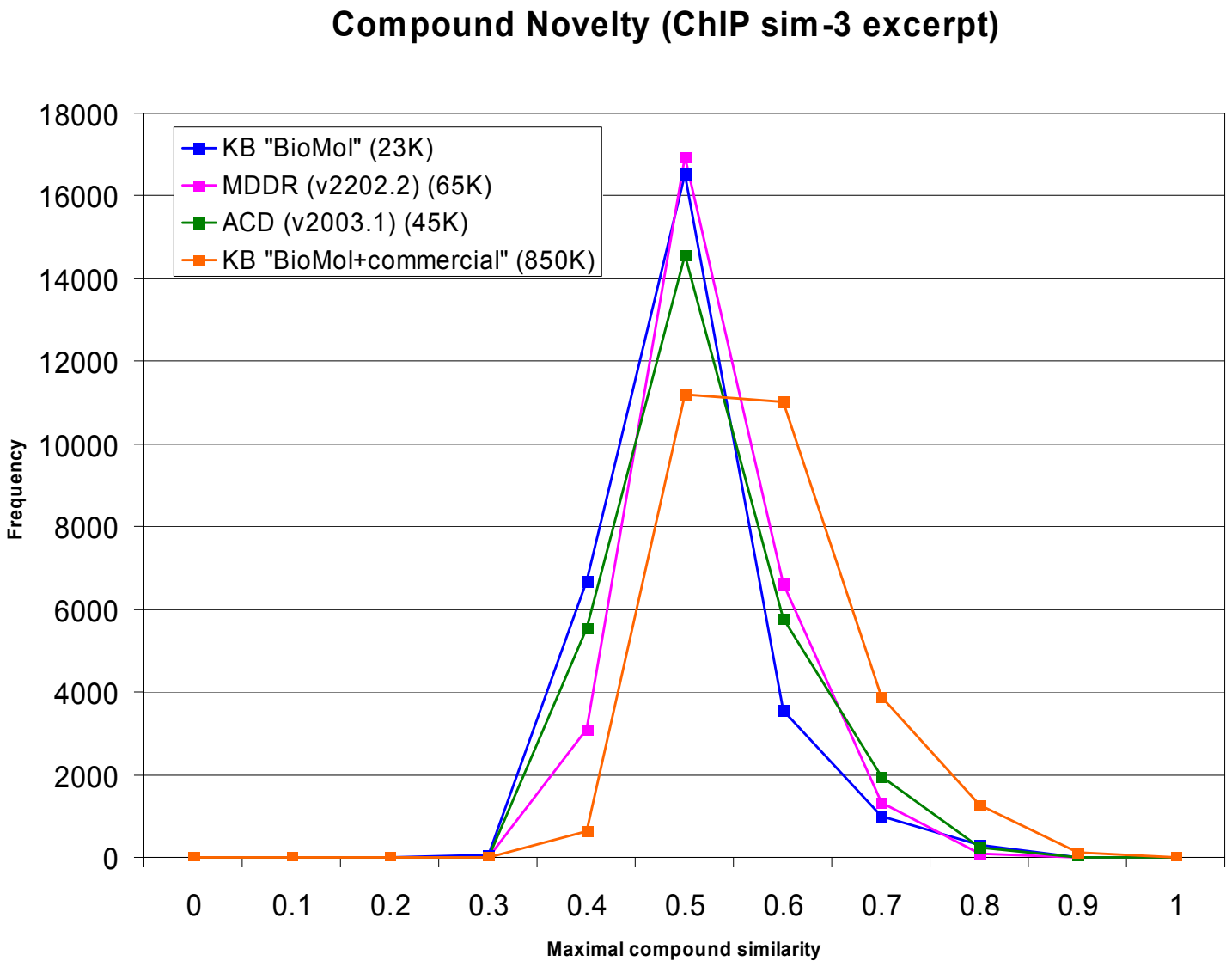


Eidogen-Sertanty

ChIP – “Diversity” Example Simulation

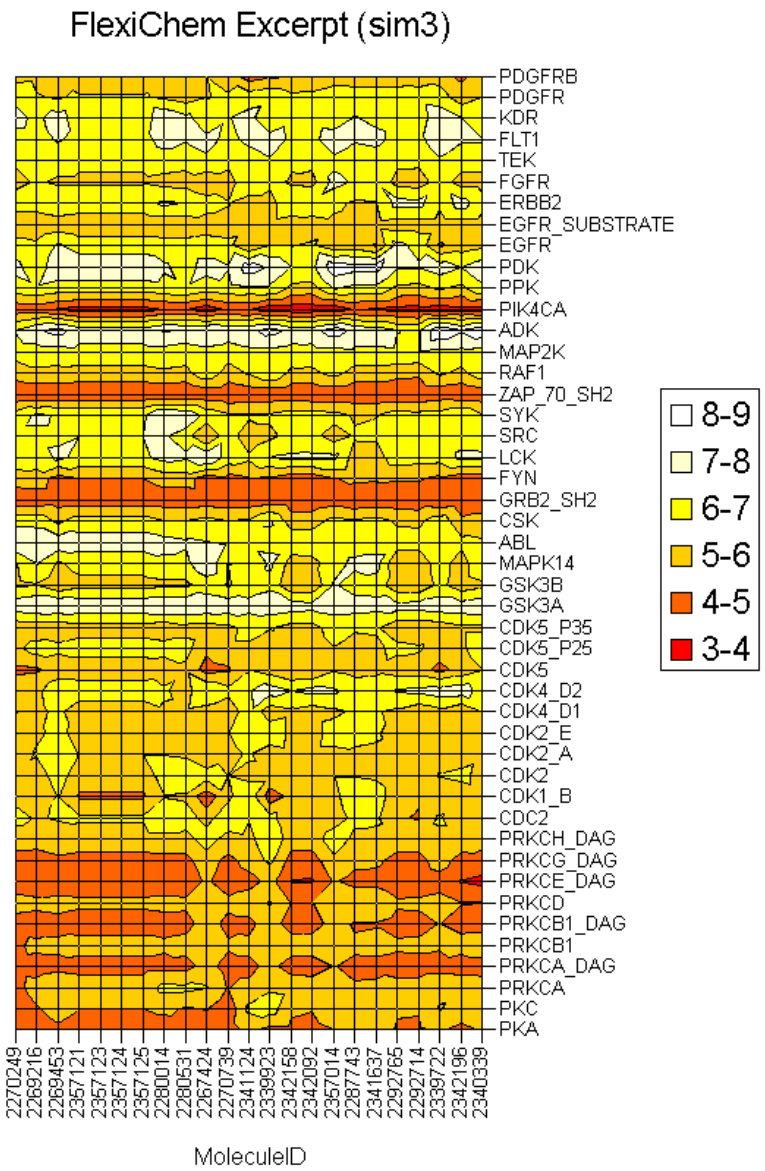
- Start with ~ 40 generic reactions with “introspective” filters
- Generate virtual protocols by graph traversal algorithm based on compatibility of generic representations
- Enumerate virtual protocols using commercially available starting materials as input
- Eliminate structures that overlap with database TDTFile (e.g. KKB)
- Apply Lipinski and structural filters (> 90)
  - MWT, HBA/HBD, ClogP, TPSA, rotBond
  - reactive functionalities like alkylators / acylators, electrophiles, nucleophiles, etc.
  - Undesired motifs (non-standard elements, >2 halo or >1 nitro per aryl, thioesters / ureas, un-branched chains, etc.
- Excerpted set: 28K novel compounds

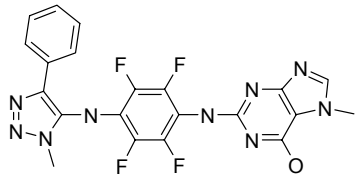
Example Simulation Results



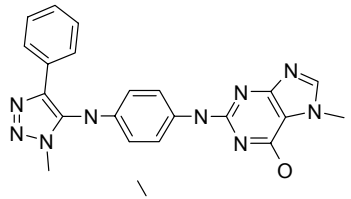
Biogen Sertanty

Excerpted ePotency & eSelectivity (sim3)

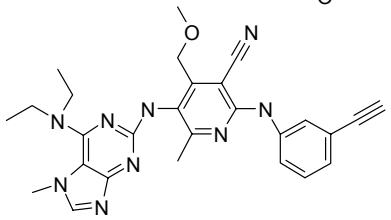




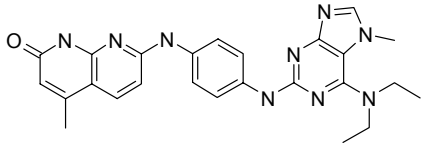
2270249 eABL: 7.96; eSRC: 6.83



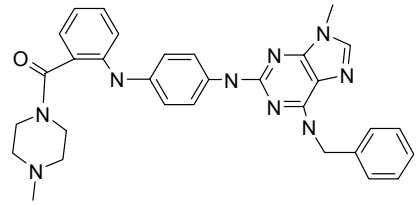
2269216 eABL: 7.75; eSRC: 6.82



2357124 eABL: 7.51; eSRC: 6.99



2280014 eABL: 7.36; eSRC: 7.12

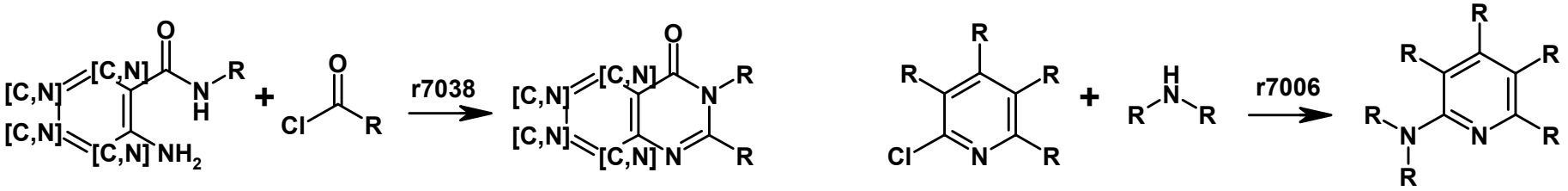


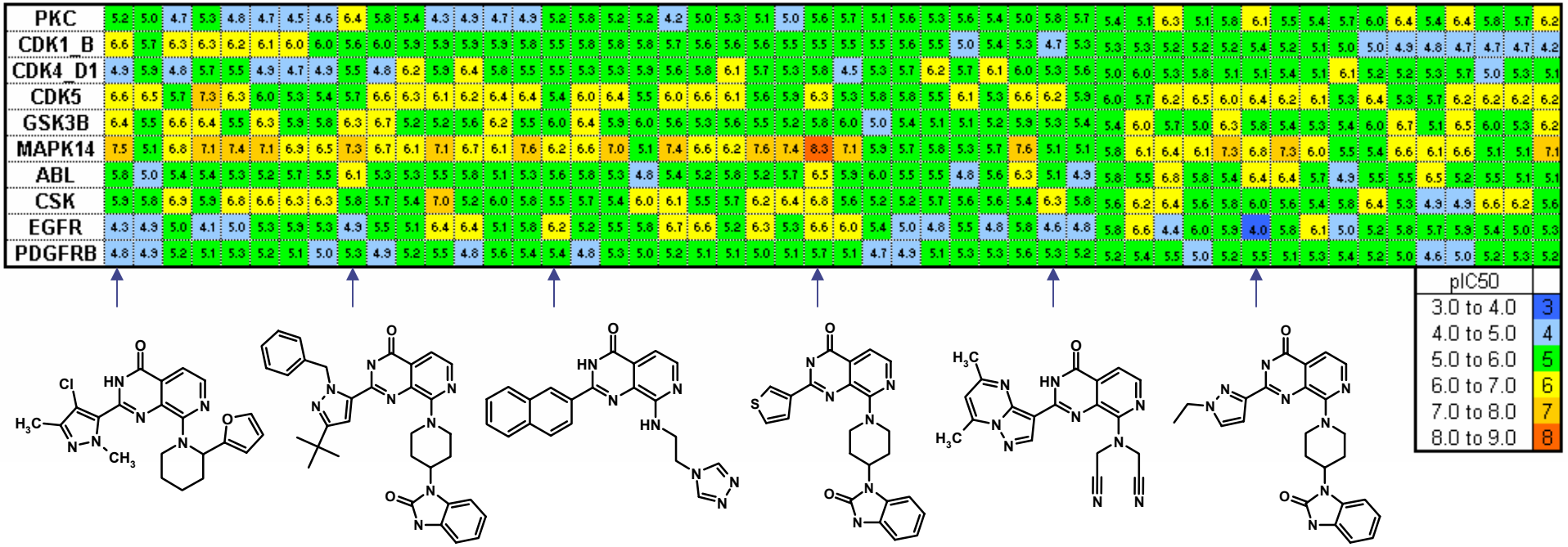
2267424 : eABL: 7.04; eSRC: 4.94

Eidogen-Sertanty

P38 ATP-Site Directed Simulation

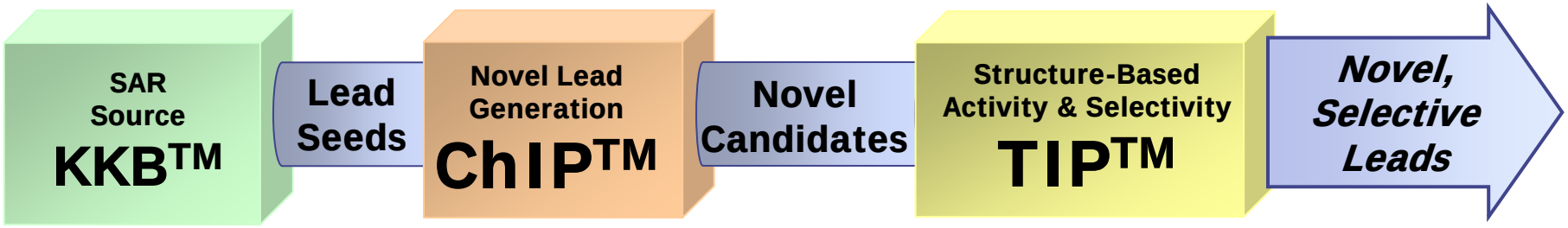
Activity Screen: MAPK14 / p38alpha Pharmacophore Model
High Scoring Reaction Products: 4-Aminopyridopyrimidinones





Eidogen-Sertanty

De-”Know”-vo Design Summary



- Diversity and/or focused library evolution through tractable chemistry with corresponding available building blocks

- Flexible evolution direction and scoring: e.g. PFPSim, (Q)SAR, Target Structure-directed, etc.

- Project-based collaboration engagements and/or several components individually licensable (e.g. ARK, Reaction Content, EVE, KKB, TIP, command line enumeration, etc.)