SAND2010-8182P

Tuning molecular specificity for design of medical countermeasures

Diana Roe, Ph.D. (PI), Joe Schoeniger, M.D., Ph.D. (co-PI),
Peter Anderson, Vincent De Sapio, Sid Elmer, Yooli Light
Sandia National Labs

Controlling Binding Specificity

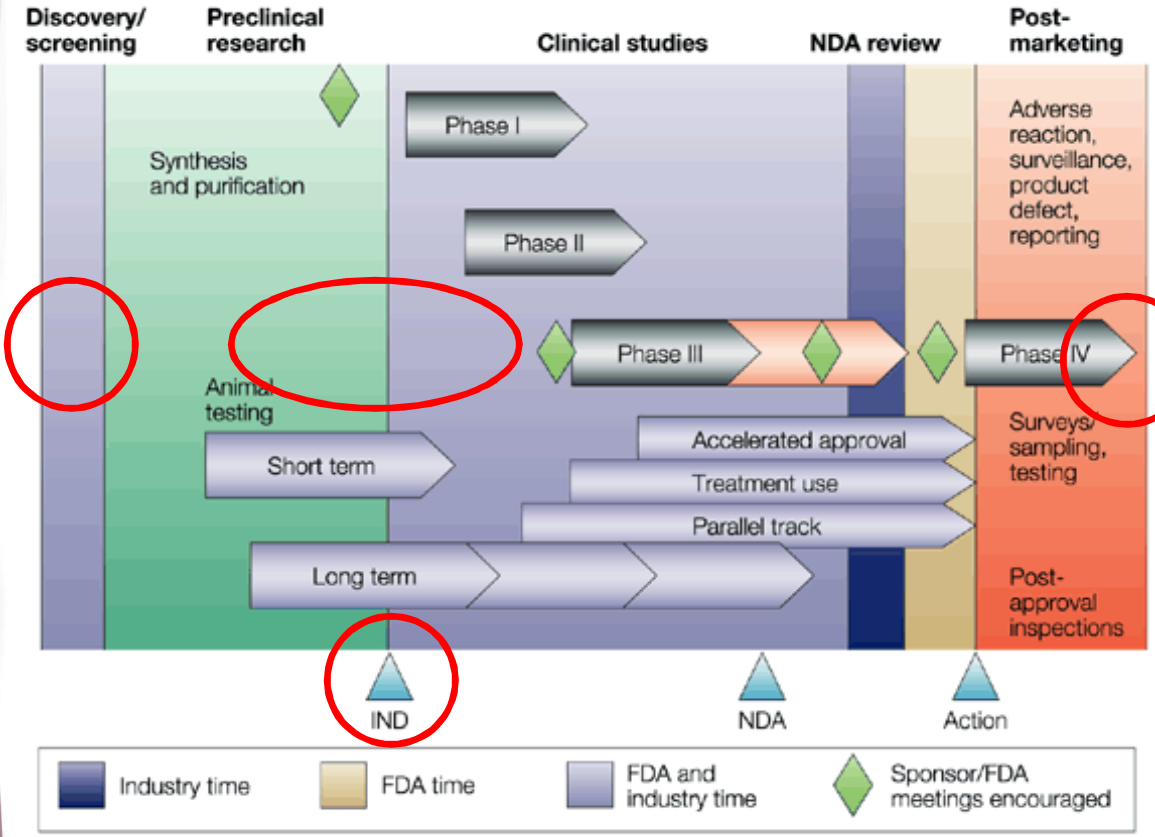
Our work is focused on classifying, understanding, predicting and controlling specificity of binding interactions across large classes of proteins and drugs.

- Combines computational, statistical, and experimental approaches to identify *Specificity Determining Features (SDFs)* of the molecular interaction interface.
- Can be used to answer questions like:
  - Can I find compounds that maintain antimicrobial activity but do not bind off-target host receptors?
  - How effective will my compounds be against new or genetically engineered strains?
  - Is resistance likely to emerge quickly?

Highlights: Specificity Profiling

- Developed pipeline for specificity profiling as part of an ongoing DTRA basic research project in studying molecular recognition.
- Integrates bioinformatics, structure-based drug design, experimental high-throughput screening, mutational analysis and next-generation sequencing.
- Produces intuitive, predictive models
- Key concept: probing specificity with ligand binding profiles
  - sequence- or structure-based predictions do not correlate well with experimental binding data.

Where should systematic specificity profiling be in the drug design cycle?



***Nature Reviews Drug Discovery* 2, 71-74 (January 2003)**

- To expand/tailor the specificity profile of a validated drug.
- *In development:* to explore salvage strategies.
- At the IND stage: to prioritize candidates with low off-target and resistance potential
- *In discovery*: to explore the possible breadth of activity across a target family.

Early use of specificity profiling can help determine which targets and lead compounds to move forward

Specificity Profiling: Finding specificity determining features (SDFs)

Binding Data (Ki, Kd)

Ligands

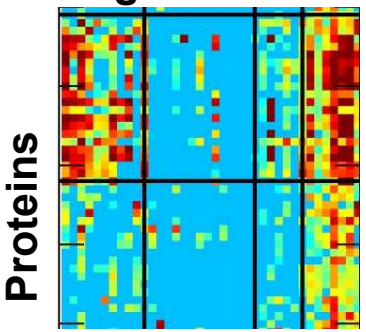
Proteins

Table of Binding Data

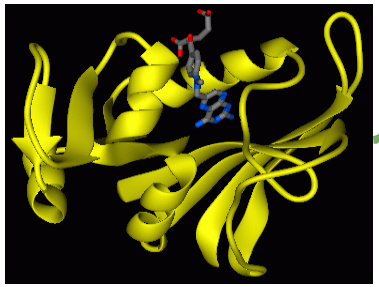
Classify

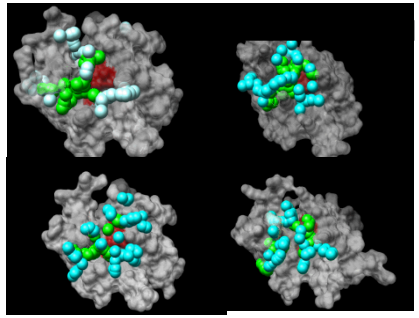
Ligands

Proteins

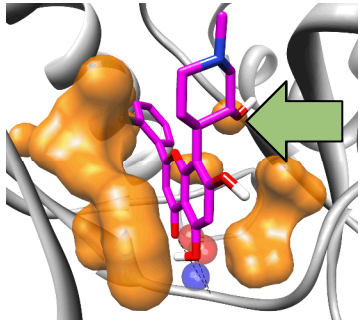


Simulation (Docking, MD)

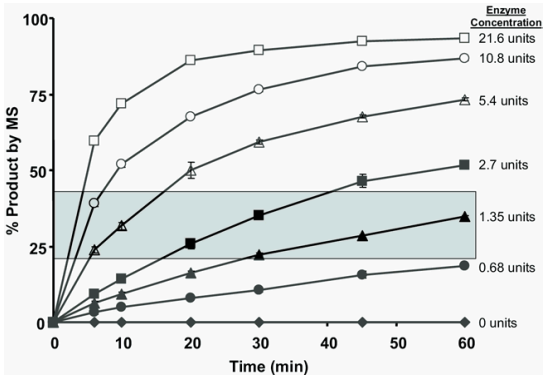




Extract SDFs



New predictions

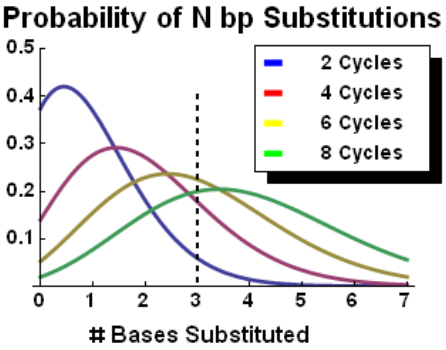


Experimental Validation

Mutational Profiling: Accelerated in-vitro approach to generate and evaluate ALL Triple Mutants of Binding Site

PCR out gene segment around the binding site (~ 70-130 bp)

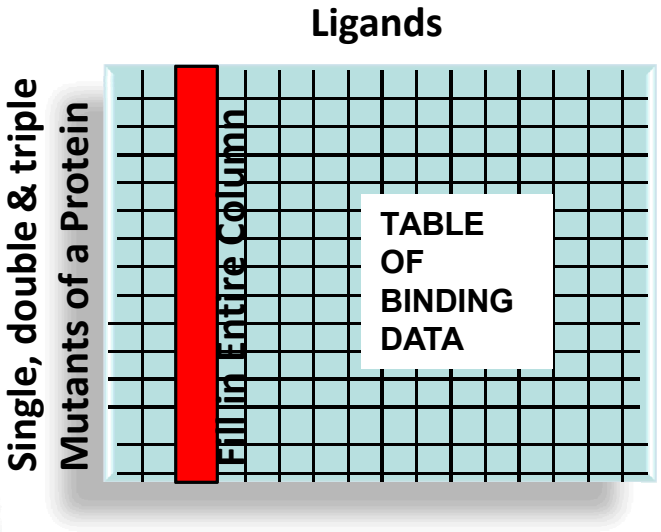
Amplify with Error-prone PCR to get all single, double and triple mutants



Reintegrate into gene & Express in phage protein display system (<= 1 protein copy per phage) to produce a library of triple (bp not aa) mutants.

Affinity chromatography versus immobilized ligands sorts out weak, medium and strong binders

Illumina Sequencing of each fraction determines which mutants are in it. 1-2 Million reads (~$200) provides >4x coverage of ALL triple mutants. Since coverage is complete, can be repeated for additional ligands.

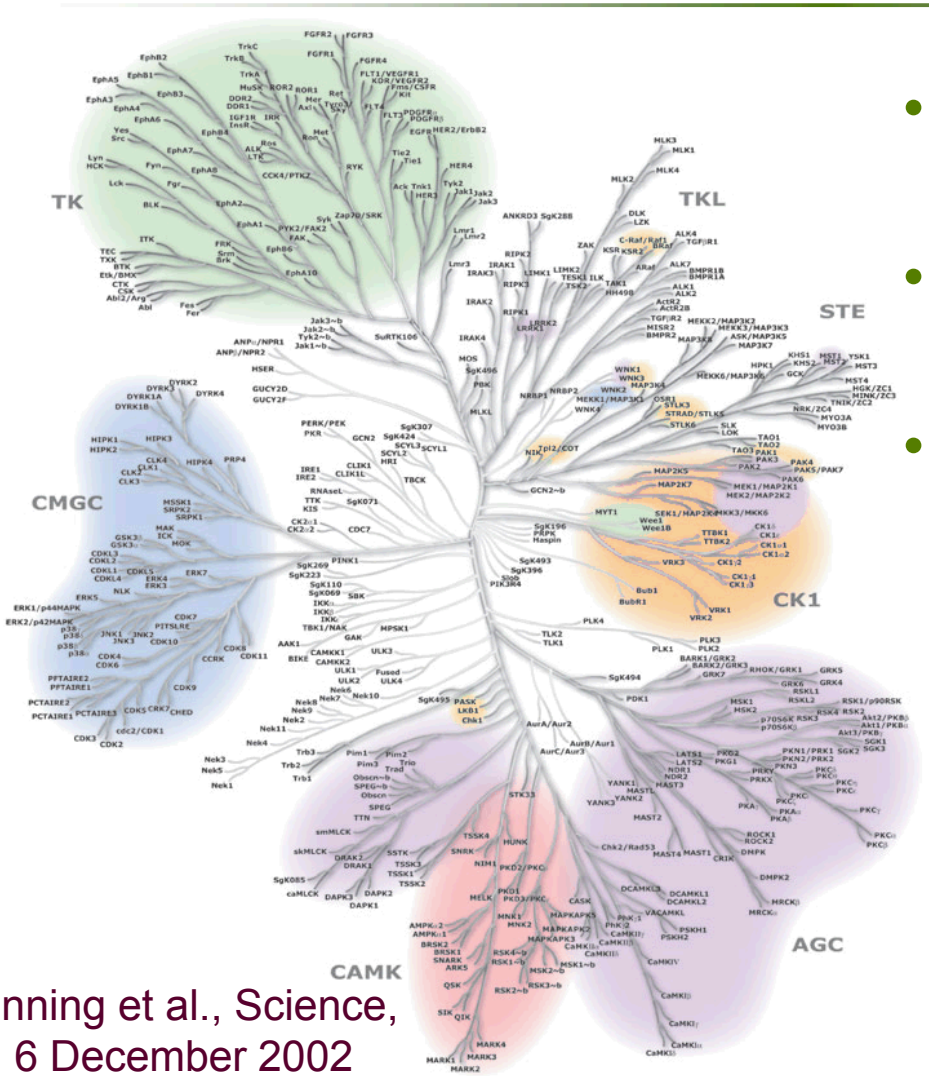


Current Test Systems

- **GOAL: Find specificity-determining features (SDFs) across protein target and ligand spaces**

Test System	Enzyme Source	Experimental Ligand Binding Data Available in literature
Protein Kinases	Human	>40,000
DHFR	Bacterial / fungal / protist	> 4000
HIV / HCV Proteases	Viral	>14,000 / >300

Specificity profiling protein kinases

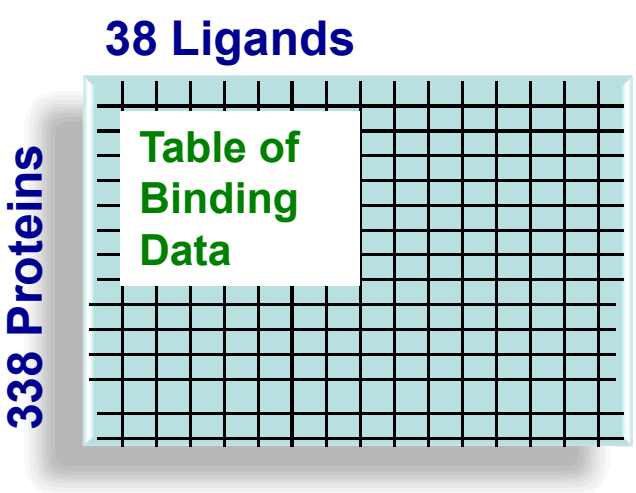


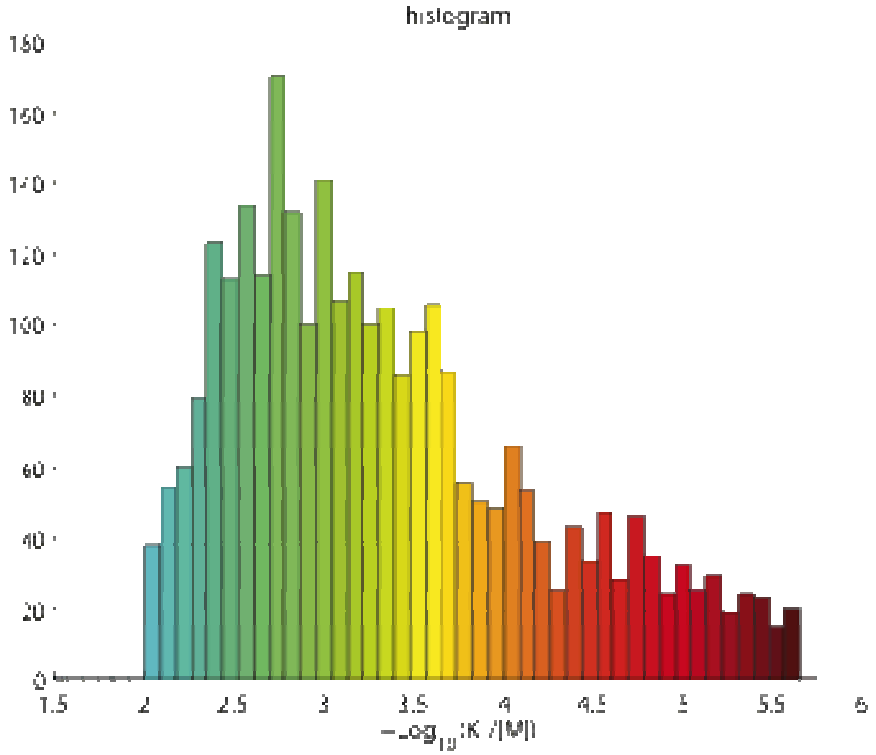
- Human kinome: over 518 proteins
- Most within cluster of >30% homology
- Known inhibitors can bind to the native fold (“type I”) or induced fit hydrophobic subpocket (“type II”)

Manning et al., Science, 6 December 2002

Starting TBD for the human kinome

Values for Kinase/Ligand TBD taken from a comprehensive experimental study in the literature.



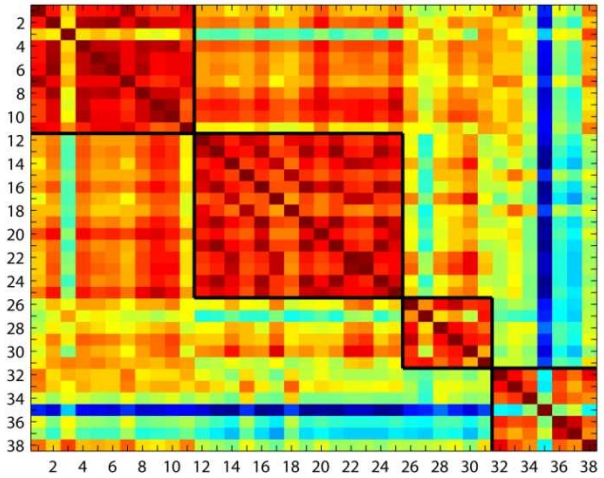


Karaman MW, et al, Nat. Biotechnol, 2008. 26 127-132.

Human Kinome Results: Cluster by ligand binding data

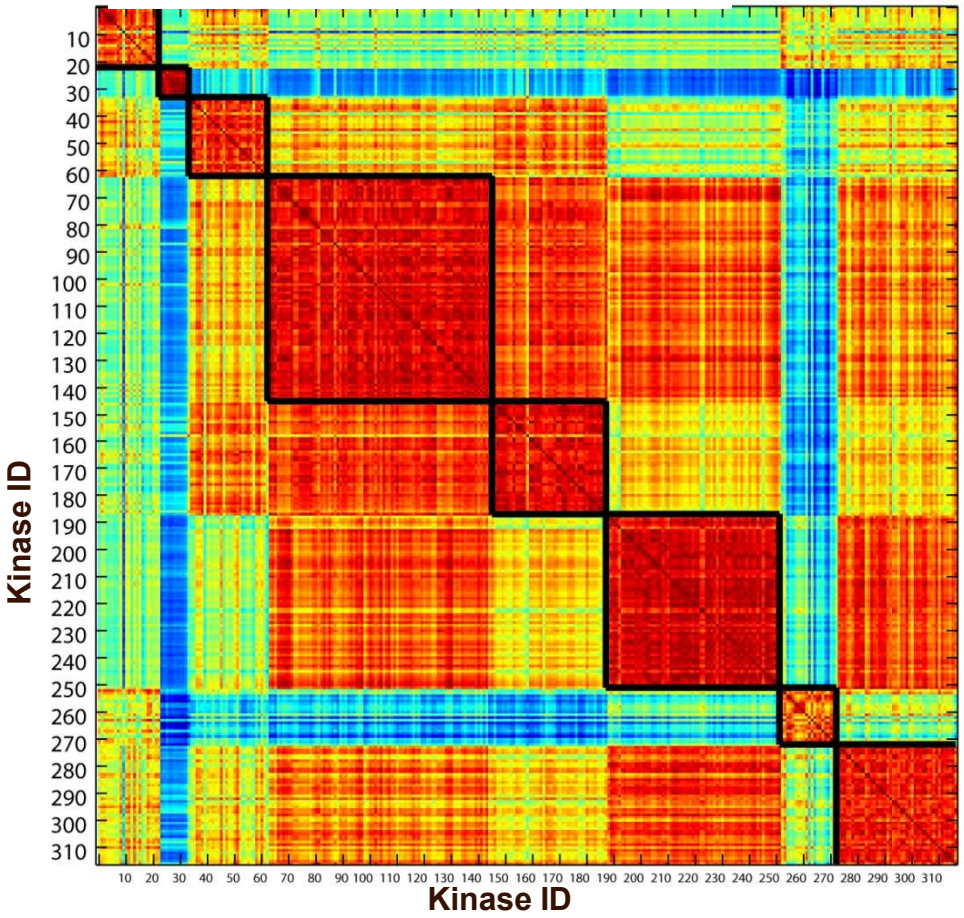
Ordered Heatmap showing kcenters clusterings

Ligand Clustering



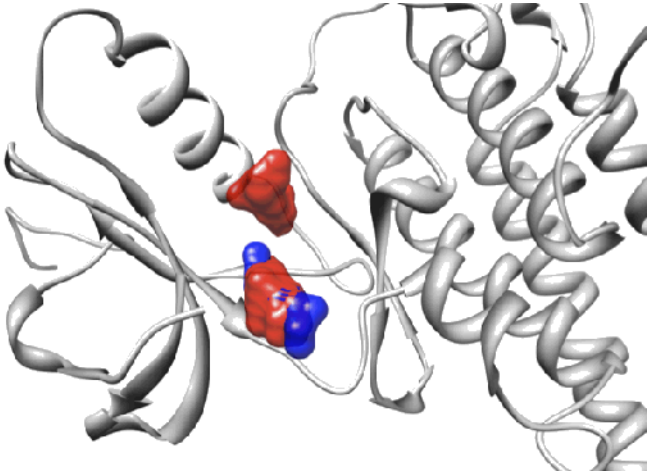
- "type-2" inhibitors in ligand cluster 1
- broad binders in ligand cluster 4

Protein Clustering

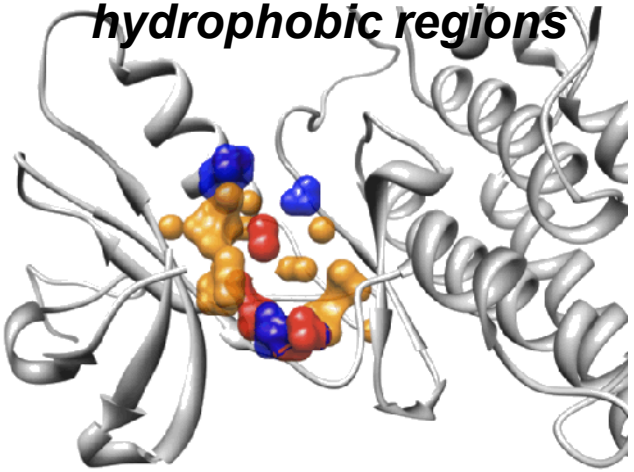


SDFs: Broad Binding Features (common among all clusters)

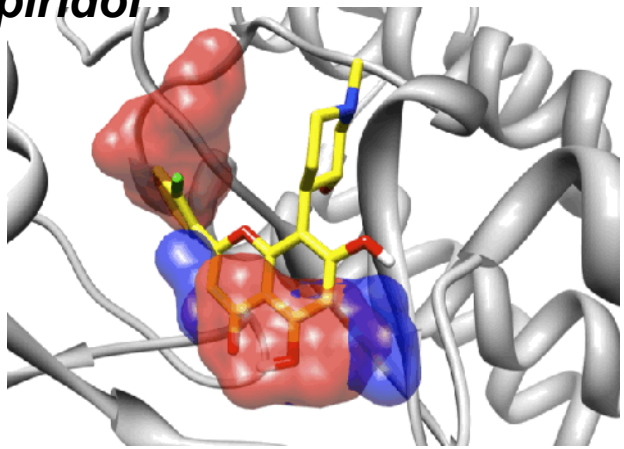
Ligand-space hbond regions



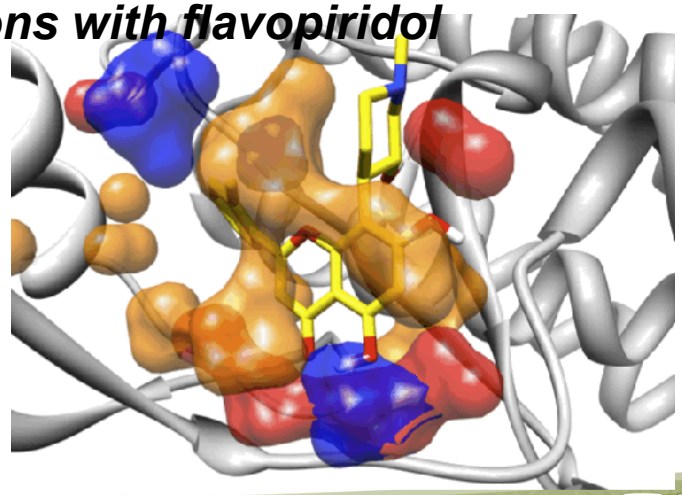
Protein-space hbond and hydrophobic regions



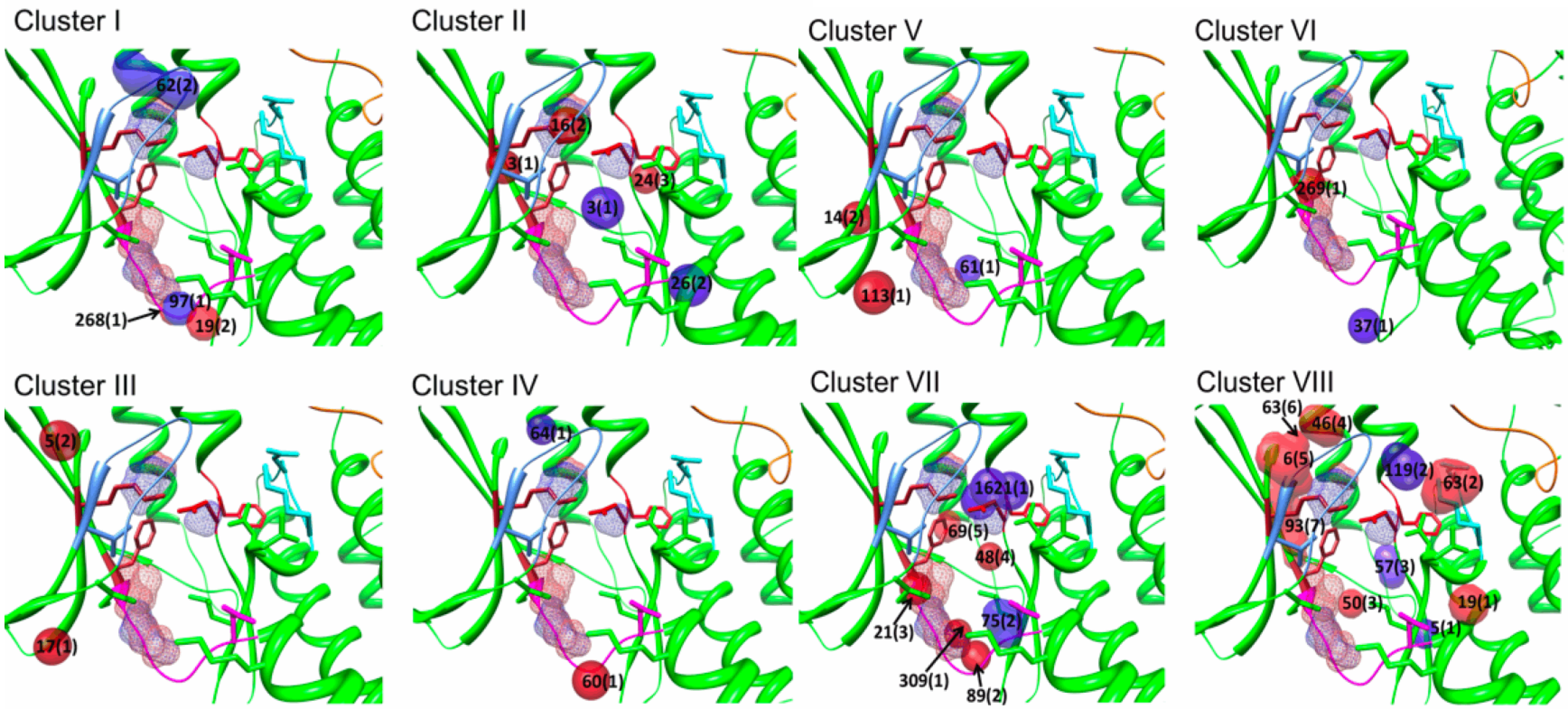
Ligand-space hbond regions with flavopiridol



Protein-space hbond and hydrophobic regions with flavopiridol



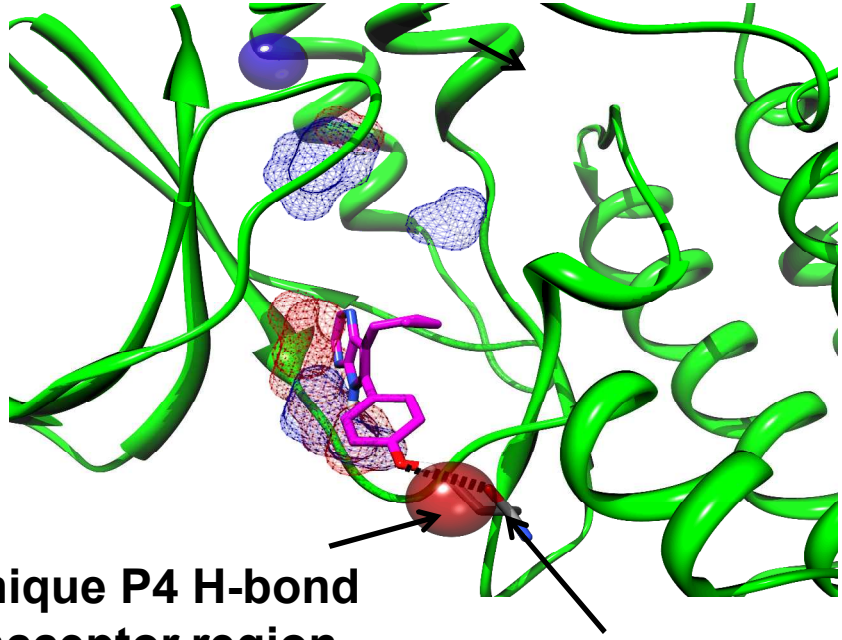
SDFs Unique to a Cluster



Note: individual ligands will have 1-2 SDFs from within this set which represents a superset of ligands

SDFs predictive for other datasets

CDK2 (IC_{50} = 880 nM)



Unique P4 H-bond acceptor region

H-bond acceptor for aloisine

- A prior study using traditional QSAR (Sheridan, 2009) produced different models for each dataset
- Preliminary data from SDF approach produces common SDF models between the Federov and Karaman data sets
- Example: unique H-bond acceptor SDF found in both datasets for cluster containing CDK2 & ZAP70. Compounds binding specifically to these proteins make this interaction.

Summary of Kinase Study

- Using ligand binding data is a robust way to cluster proteins and ligands and useful patterns of binding emerge from these clusterings.
- We can turn combine these clusters with docked poses to extract SDFs
- These SDFs match specificity features in ligands outside our initial data set.
- Predictive capability validated with ligands outside our dataset
- Experimental validation of novel predictions ongoing

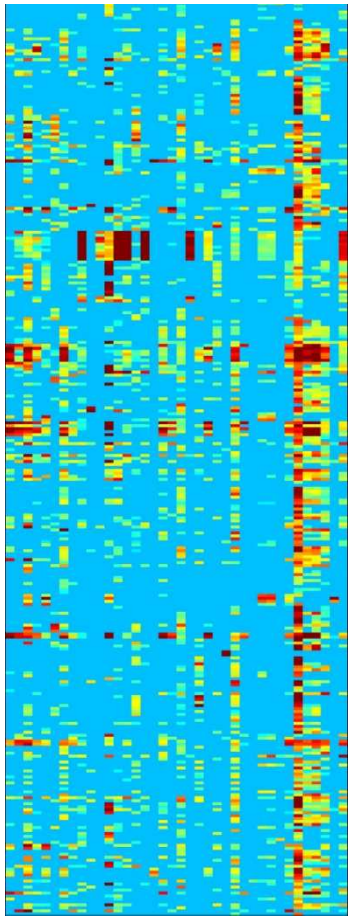
How can specificity profiling be used in potential TMT projects?

- *Emerging threats*: NDM-1 metallo β-lactamase – characterize structural features responsible for its resistance - mutational profiling to identify future potential variants, identification drug variants broadly binding to all variants (with Carol Zhou, Adam Zemla, LLNL)
- *Viruses*: Alphaviruses (or Flaviviruses) - identify broad spectrum inhibitors to viral proteases that are robust against resistance/genetic engineering, validated by experimental mutational profiling (with Stan Langevin, SNL; CDC)
- *Decision Tool :*Prioritize IND candidates with known structure-based targets (eg: viral proteases)
- *Host-directed therapeutics:* Immunopotentiation – RIG-1 pathway of innate immune system is a broad responder to many viral families. Find specific inhibitors to LGP2 to upregulate RIG-1 pathway (with Steve Branda, SNL)

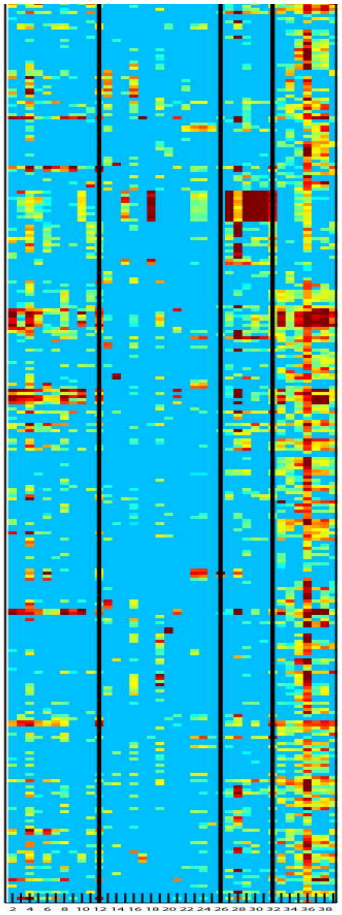
Extra Slides

Human Kinome Results: Binding Data Ordered by Clusters

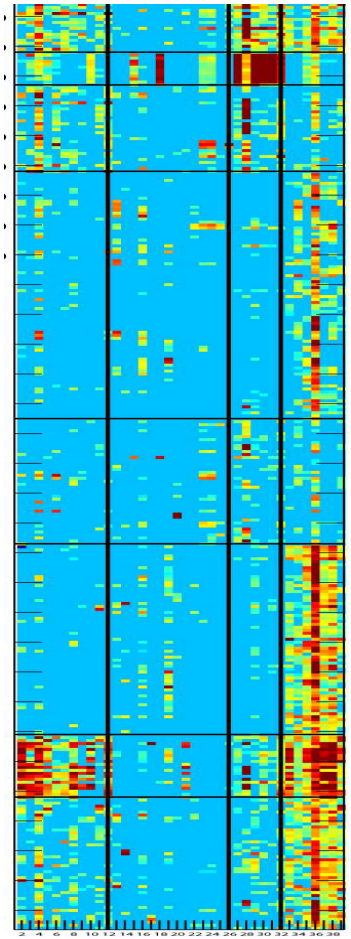
Unordered Binding Matrix



Binding Matrix Ordered by Ligand Clusters



Binding Matrix Ordered by Ligand and Protein Clusters

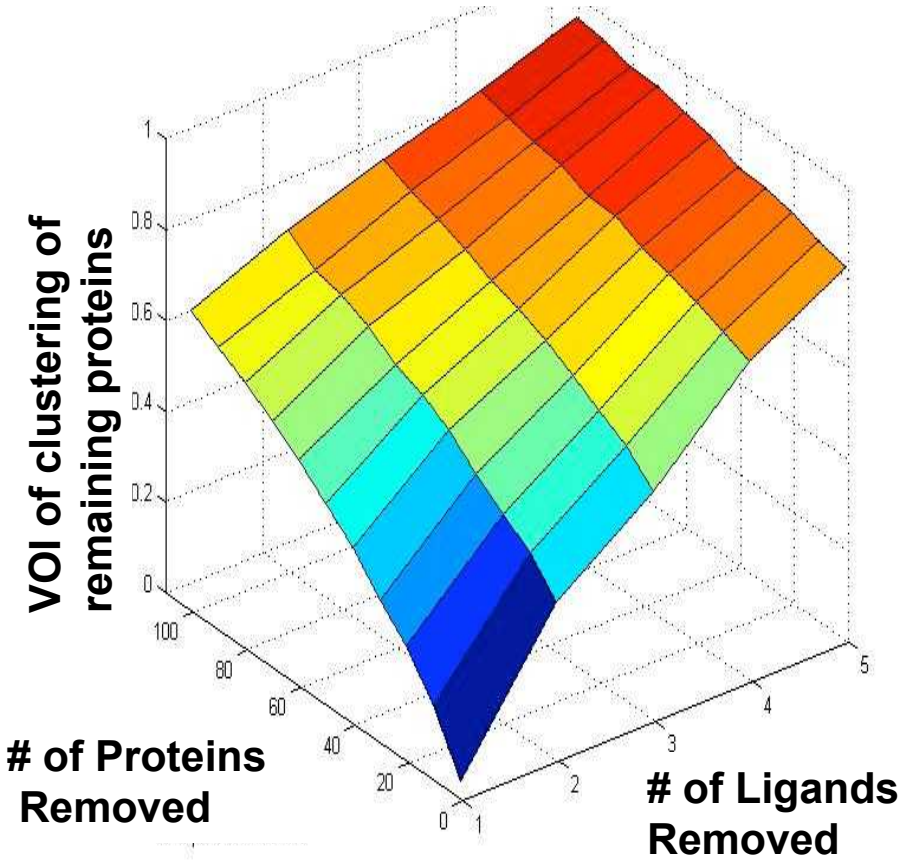


Robustness of Classifications

Leave 1-out analysis shows clustering robust for both ligands and proteins

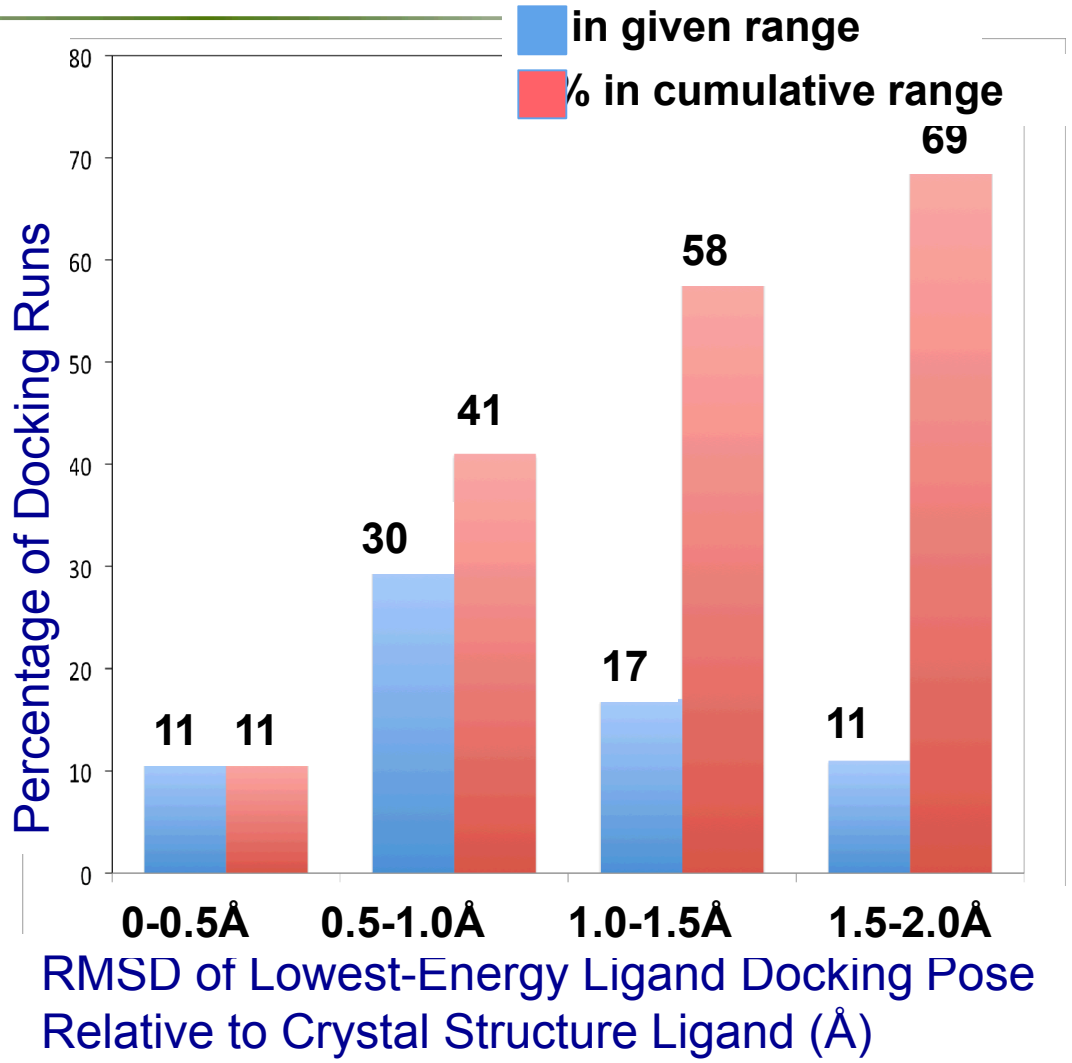
- Variation of information (VOI) Mathematical method to measure distance between 2 clusterings.
- *Clustering by sequence or structure do not capture the patterns in experimental data.*
  - VOI of random cluster is 3.7
  - VOI for clustering by sequences is 2.57
  - VOI for clustering by structure motifs is 2.73

Cluster Degradation with respect to protein and ligand removal



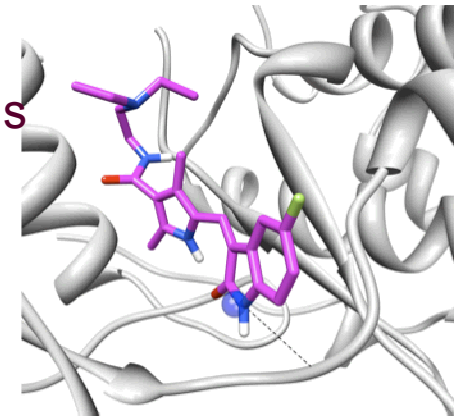
Docking to kinases and extracting specificity determining features (SDFs)

- Docked 38 ligands to 113 kinase structures using autodock 4 with flexible ligands
- Validated docking poses with crystallographic ones for those with co-crystals (figure)
- Features (h-bonds, polar, hydrophobic) extracted from docked poses using experimentally determined clusterings.
- Statistical approach to feature extraction– insenstive to “noise” from mis-docked features

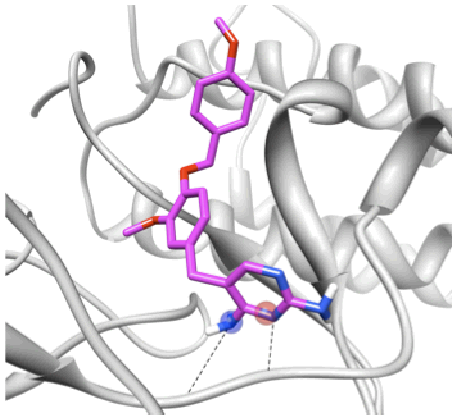


SDFs predictive for ligands outside our dataset

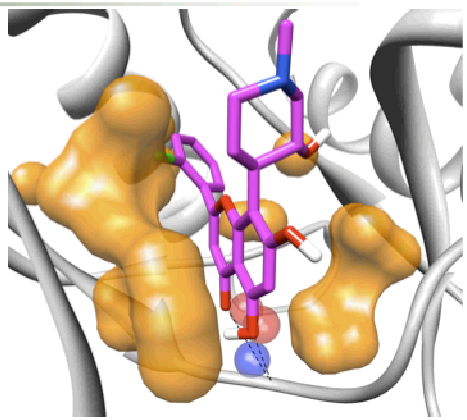
Top row: docked ligands from dataset



JNK1, sunitinib

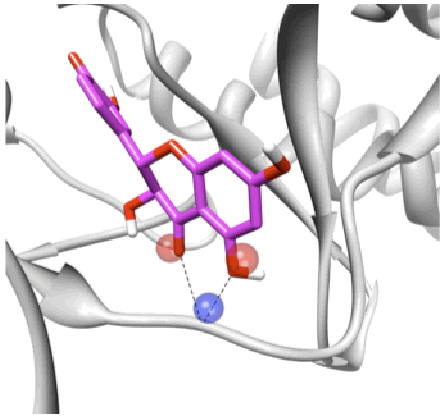


LYN, GW-2580

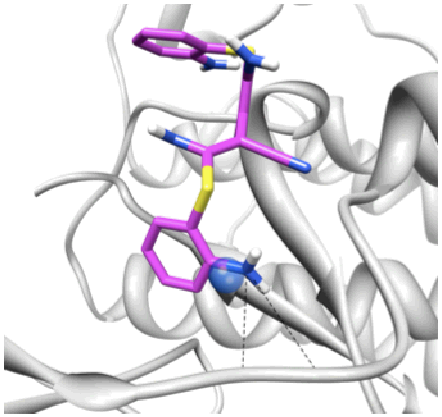


CLK1, flavopiridol

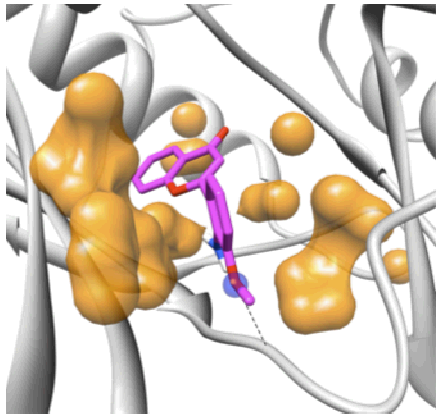
Bottom row: docked ligands from outside dataset



ITK, quercetin



ITK, U0126



KIT, PD98059

How to use SDFs in specificity tuning?

- To generate broad binding:
  - Design/optimize ligands to bind to globally present SDFs
  - Accelerated mutational analysis used to generate SDFs will help ensure ligands robust against simple strain variations
- To generate narrow binding:
  - Design ligands to bind to cluster-specific SDFs.
  - “Drill down” within cluster to find SDFs unique to target.
  - Use other members of cluster both computationally and experimentally
- To prioritize drug candidates: use SDFs to compare candidate’s specificity to desired specificity range (ie desired range of proteins).