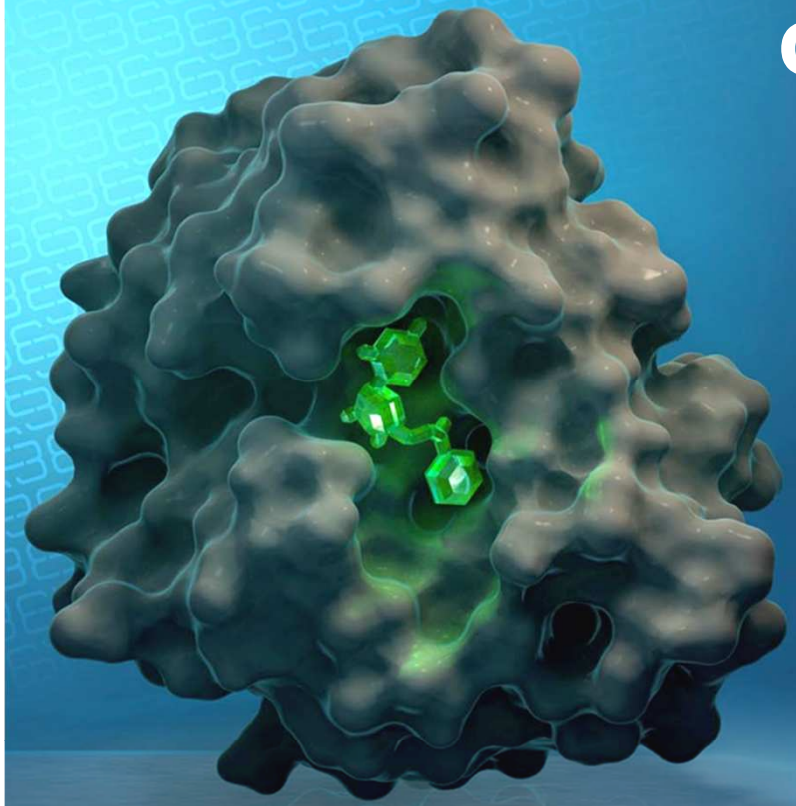
EMERALD BIOSTRUCTURES



Predicting Success For Crystallographic Fragment Screens

Doug R. Davies
Sr. Director of Structural Biology

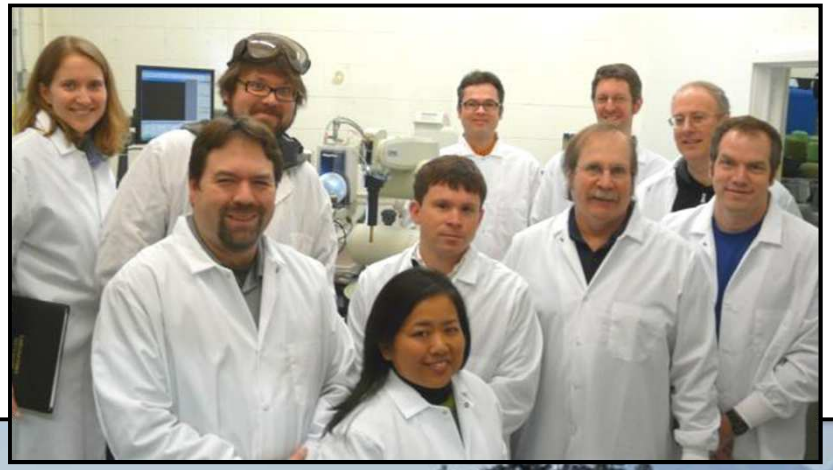
FBLD 2010
October 12, 2010

Outline

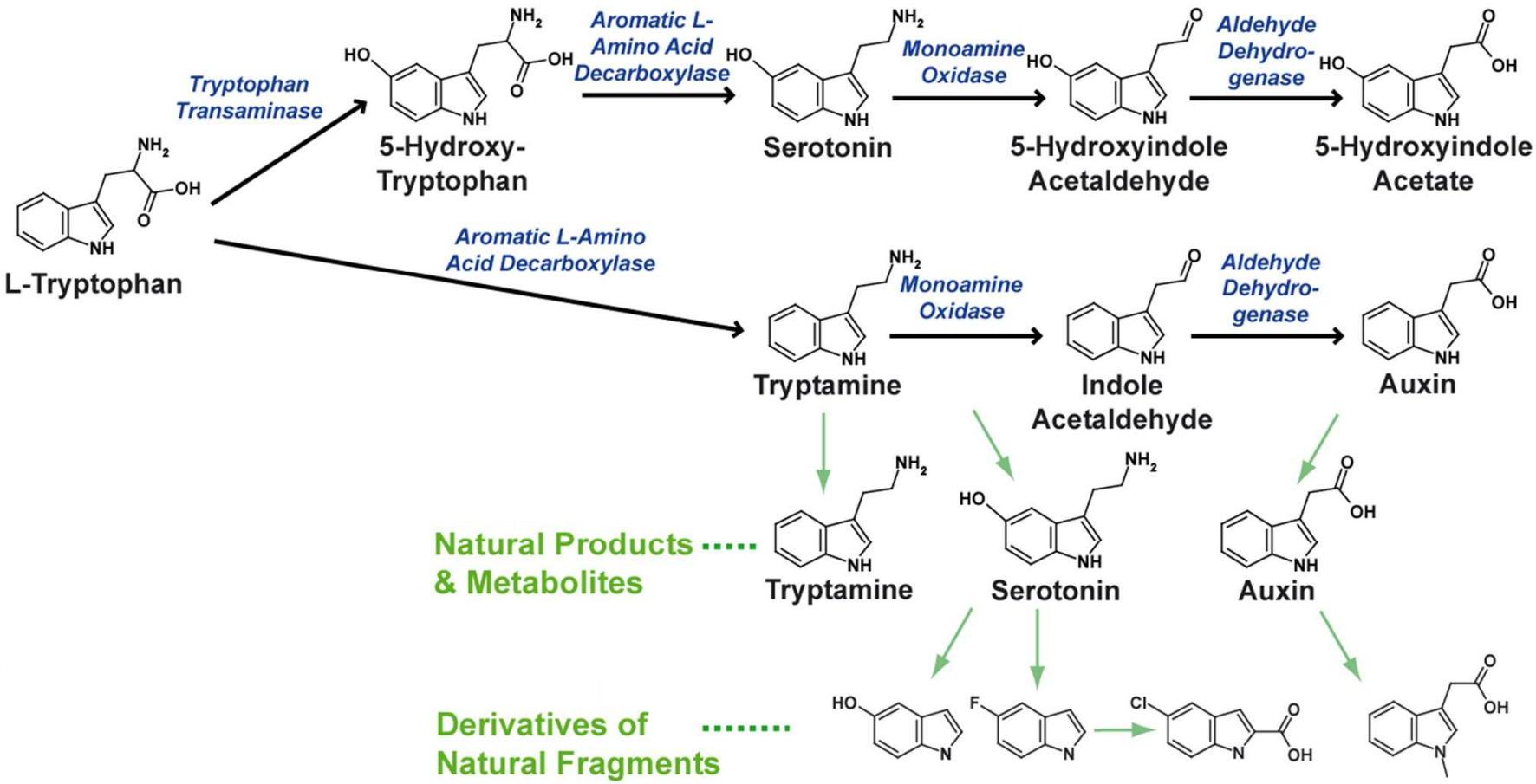
- **(Re)Introduction to Emerald BioStructures**
- **Fragments of Life™ library—rationale and design**
- **Seattle Structural Genomics Center for Infectious Disease (SSGCID)**
- **Fragment Screening Review: Predicting Successful Fragment Targets From Initial Crystal Structures**

Emerald BioStructures: A Collaborative Research Organization With Gene-Structure-Lead-IND Capabilities

- Formerly deCODE biostructures, now independent, privately-owned
- Located on Bainbridge Island, near Seattle, WA
- 55 Employees, 10 Ph.D. crystallographers
- Over 13 years experience in structural biology collaboration
- High throughput pipeline: >1500 crystals/month, >400 structures/year



Fragments of Life: Natural Products & Derivatives



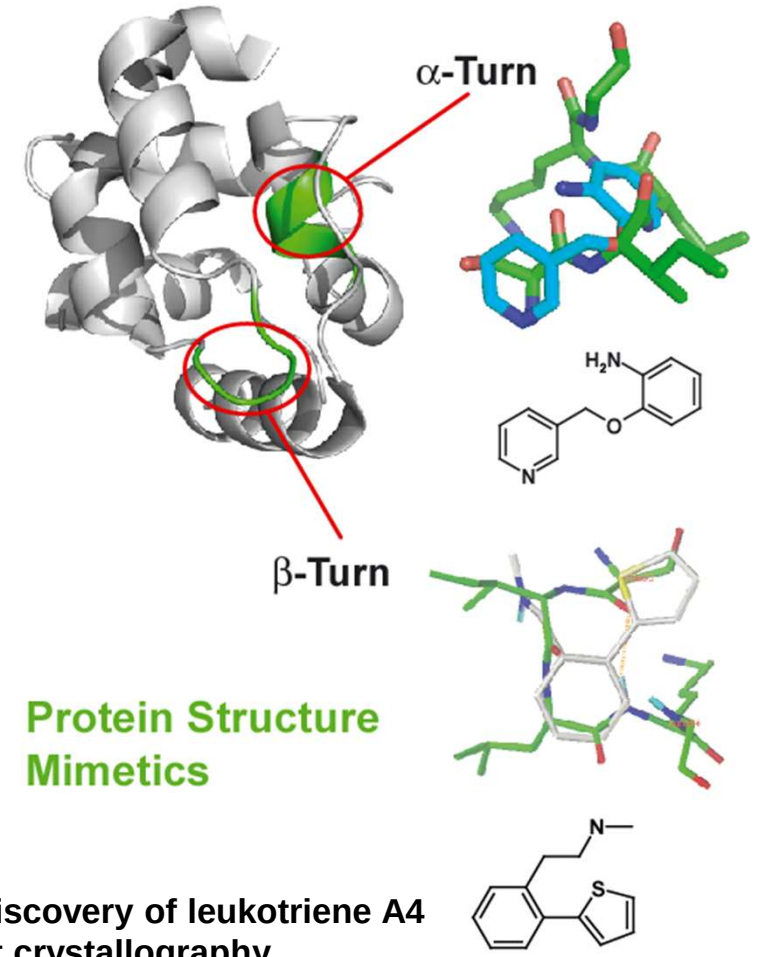
D.R. Davies, et al. (2009). *J Med Chem*. 52(15):4694-715. Discovery of leukotriene A4 hydrolase inhibitors using metabolomics biased fragment crystallography.

EMERALD BIOSTRUCTURES

Fragments of Life: Protein Structure Mimetics

- Inspiration from recent papers describing small molecule mimics of protein structure.
  - Biros, *et al*. (2007) Heterocyclic alpha-helix mimetics for targeting protein-protein interactions. *Bioorg. Med. Chem. Lett.* **17**, 4641.
  - Robinson (2008) Beta-hairpin peptidomimetics: design, structures and biological activities. *Acc. Chem. Res*. **41**, 1278.

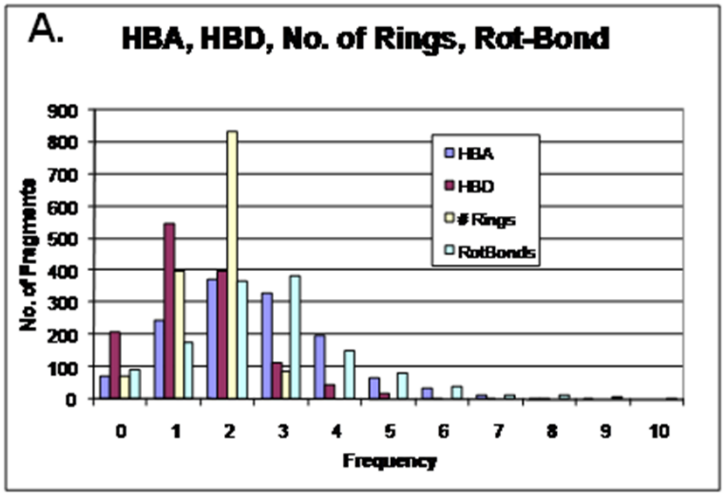
- Biaryl compounds screened *in silico* for energy-minimized conformations that match α-, β-, and γ-turns.

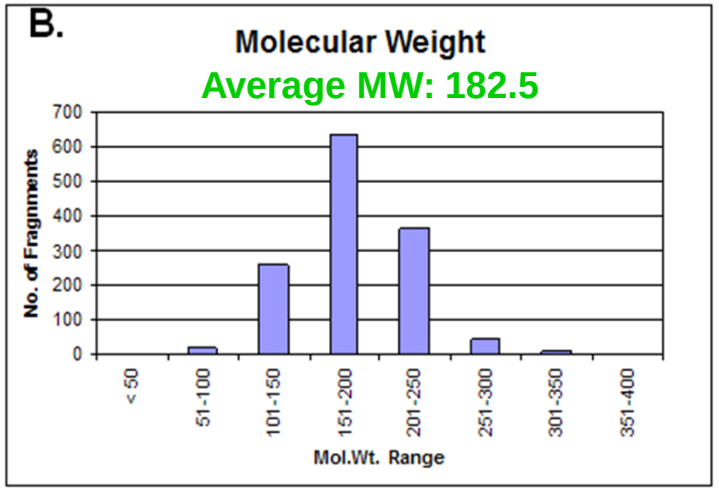


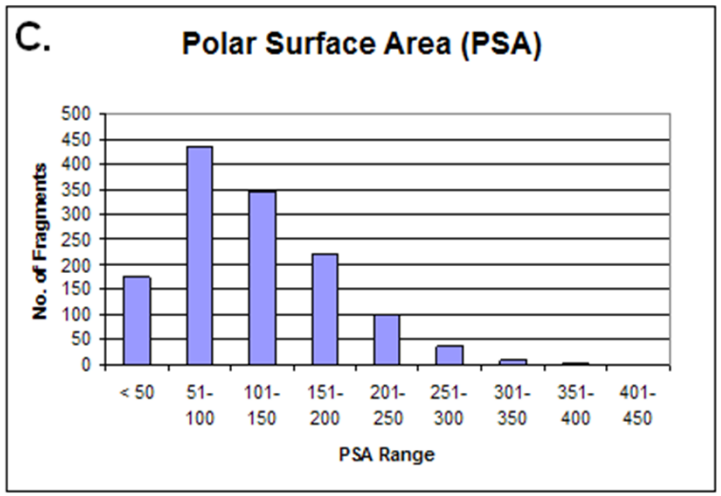
D.R. Davies, et al. (2009). *J Med Chem*. 52(15):4694-715. Discovery of leukotriene A4 hydrolase inhibitors using metabolomics biased fragment crystallography.

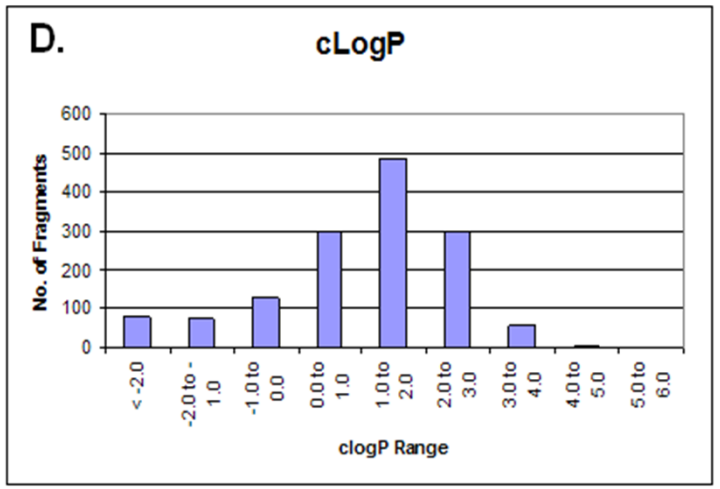
EMERALD BIOSTRUCTURES

Fragments of Life Library Physical and Chemical Properties (1444 Compounds)









RULE of 3

MW <300 Da

- **HBD ≤ 3**
- **HBA ≤ 3**
- **clogP ≤ 3**
- **TPSA ≤ 60**

Congreve *et al*. (2003) *DDT* 8: 876-877.

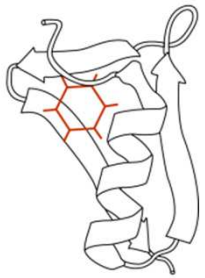
EMERALD BIOSTRUCTURES

Seattle Structural Genomics Center for Infectious Disease (SSGCID)

- **Role**: Emerald is contractor providing all X-ray services, some construct design, expression and purification
- **Targets**: NIAID Category A-C agents; (re)emerging infectious diseases
- **Goal:** Produce 500+ structures of novel infectious disease targets
- **250th Structure milestone marked September, 2010**

- ***Emerald is the only CRO that collaborates on such a large scale*.**

SEATTLE STRUCTURAL GENOMICS CENTER FOR INFECTIOUS DISEASE



EMERALD BIOSTRUCTURES

UNIVERSITY OF WASHINGTON

Seattle BioMed

Pacific Northwest NATIONAL LABORATORY

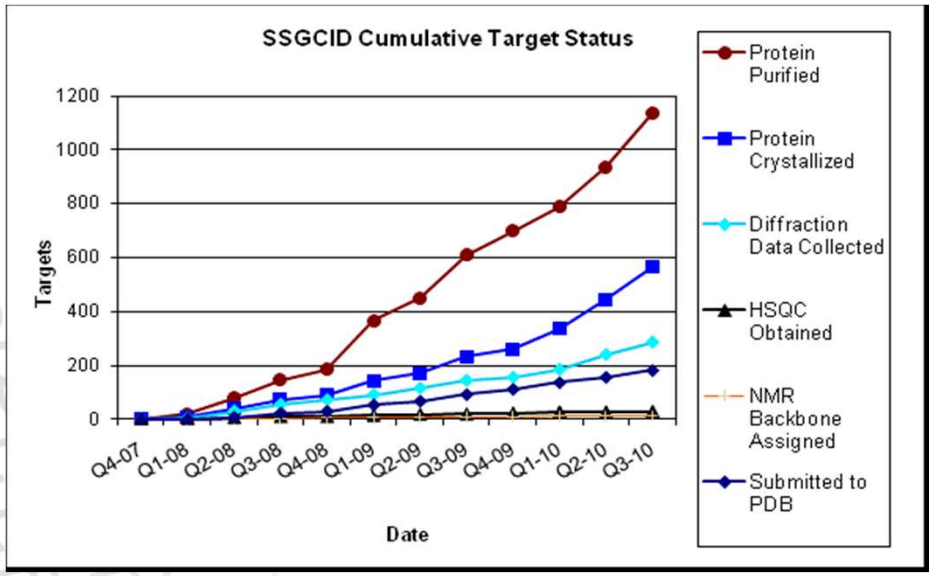
Proudly Operated by Battelle Since 1965

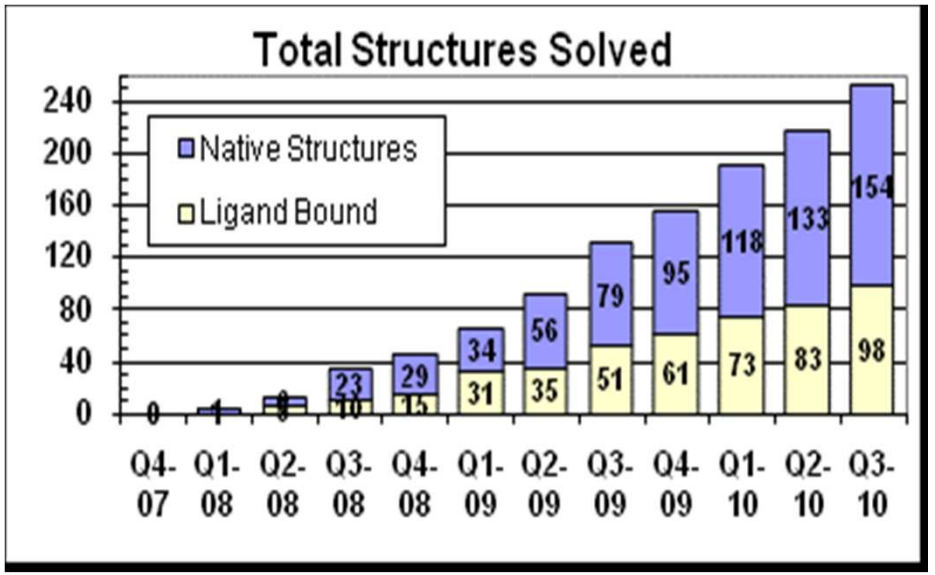
National Institute of Allergy and Infectious Diseases
National Institutes of Health

EMERALD BIOSTRUCTURES

Productivity of SSGCID

- **253 structures, 183 unique targets solved to date**
- **On pace to exceed 500 structures by 2012**





EMERALD BIOSTRUCTURES

New Concepts in Structural Genomics

PSI I-II Concepts (2000-2005):

1. Establish HT structure determination centers, technology development
   -Reduce time and expense per target

2. Coverage of “fold space”
   -Enable prediction of 3D structure from primary sequence

Emerging Concepts in SSGCID:

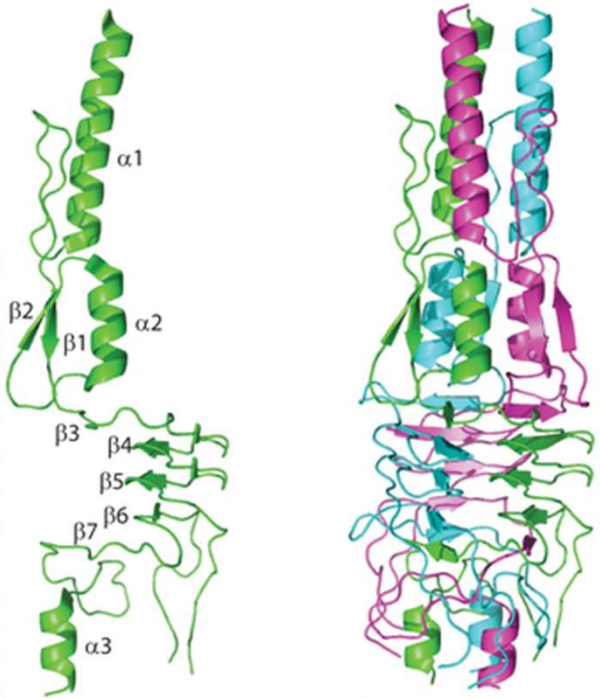
1. Engaging the Scientific Community

2. Ensembles of Structures/ Pilot Screening Projects

SSGCID: Engaging The Scientific Community

Community Requests Submitted Via Web:

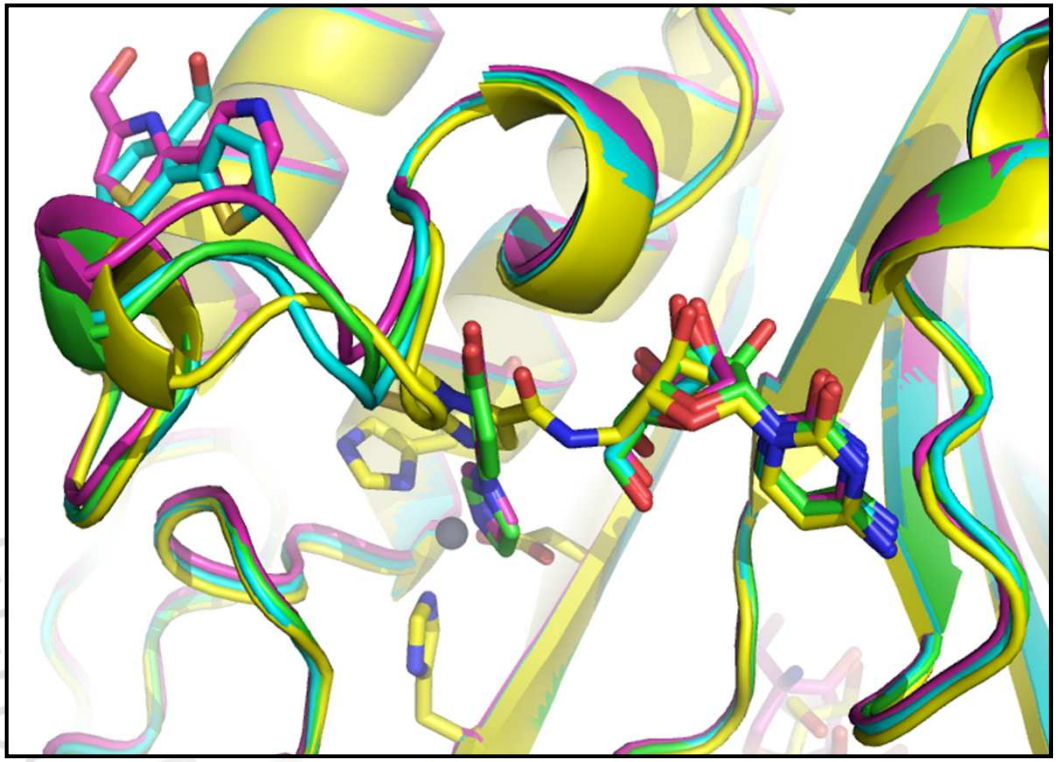
- http://apps.sbri.org/SSGCIDCommTargReq/
- **489 requests have entered SSGCID pipeline**
- **22 structures from 13 targets solved to date**



Edwards TE, Phan I, Abendroth J, *et al*. 2010 1.35 Å Structure of a *Burkholderia pseudomallei* Trimeric Autotransporter Adhesin Head. PLoS ONE 5(9): e12803.

EMERALD BIOSTRUCTURES

SSGCID: Ensembles of Ligand-Bound Structures ➔Pilot Screening



Superposition of four ligand (fragment) bound structures of *Burkholderia pseudomallei* IspF

Why Ensembles?

- Economical use of crystallized protein
- Reveal structural dynamics
- Identify multiple binding sites
- Blueprint for SBDD

EMERALD BIOSTRUCTURES

Fragment Screening of SSGCID Targets

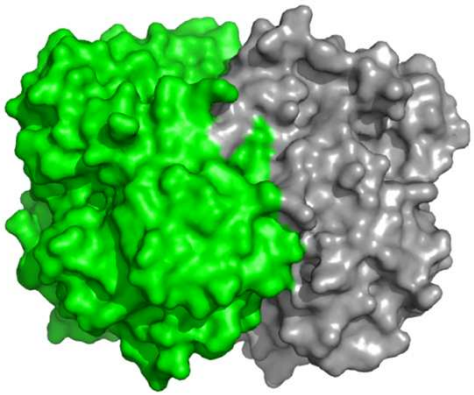
- Fragments of Life library used for screening of some SSGCID Targets
- Crystallography used as primary screen
    - Emerald has high throughput in-house X-ray system
        - Two Rigaku X-ray generators
        - Four detectors, two ACTOR robots
        - 24 hour unattended data collection possible
        - Semi-automated scripts for data processing, solution
        - Primary screen on amenable crystals in < 1 week
- Targets chosen for desirable crystal properties (diffraction, symmetry)
- Some were “productive”: yielding hits that could inspire lead development
- Some were unexpectedly “non-productive”: no hits, or crystal artifacts

EMERALD BIOSTRUCTURES

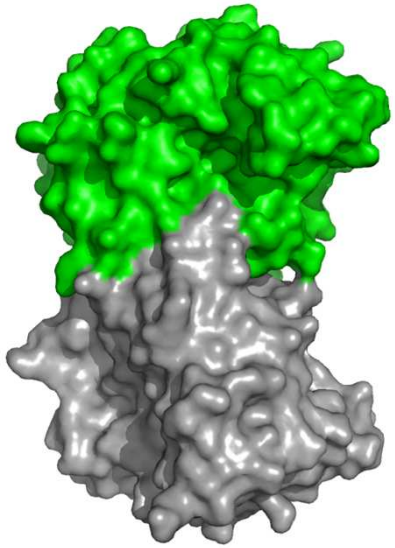
Why Are Some Crystals Non-Productive?

- Can we predict success of crystallographic screens?
- Crystal system is layer of complexity on top of “ligandability” problem
- Some intractable crystals can be identified by experienced researcher
  - Active site blocked by crystal contacts
  - Active site blocked by binding of component of crystallization solution
- Can we identify other factors that correlate with “productivity”?
  - **Survey conditions of 18 fragment screening campaigns**
    - 12 “productive” targets
      - 5 SSGCID
      - 1 internal drug discovery target
      - 4 literature examples
      - 2 examples from proprietary work
    - 6 “non-productive” targets
      - 4 SSGCID
      - 2 examples from proprietary work

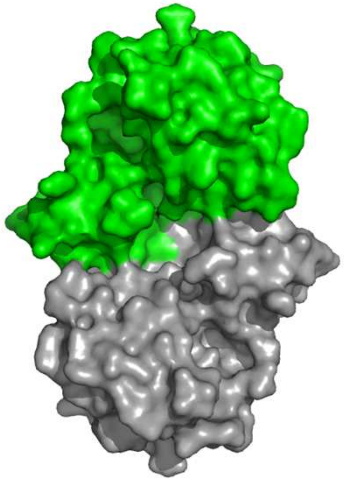
Non-Productive Fragment Targets



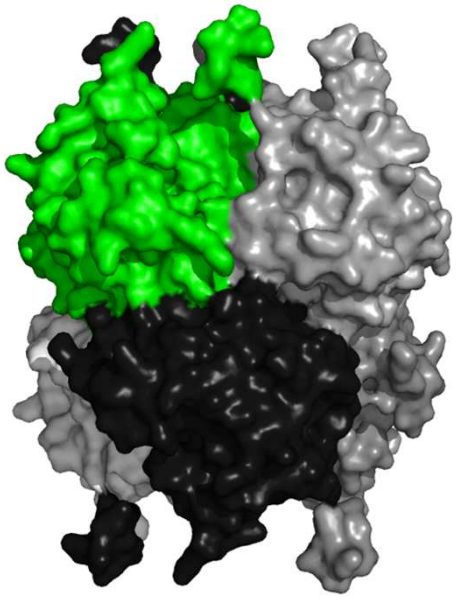
Burkholderia pseudomallei
3-Oxoacyl-(acyl-carrier-protein) Synthase III
3GWA



Mycobacterium tuberculosis
Triosephosphate Isomerase
3GVG

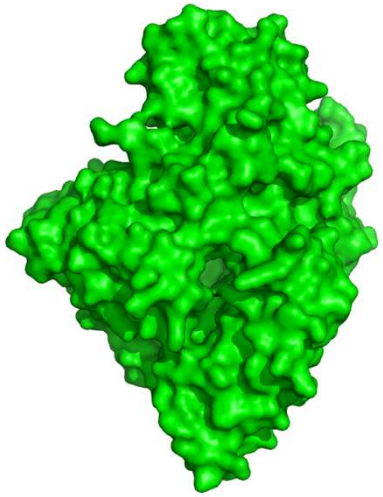


Large C-terminal Domain of Polymerase Basic Protein 2 from Influenza Virus
3KHW

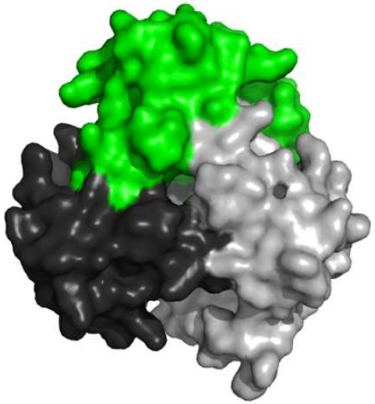


Burkholderia pseudomallei
Inorganic Pyrophosphatase
3EIZ

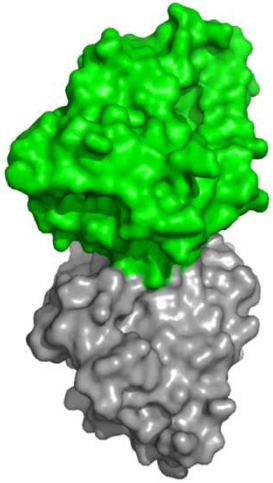
Productive Fragment Targets



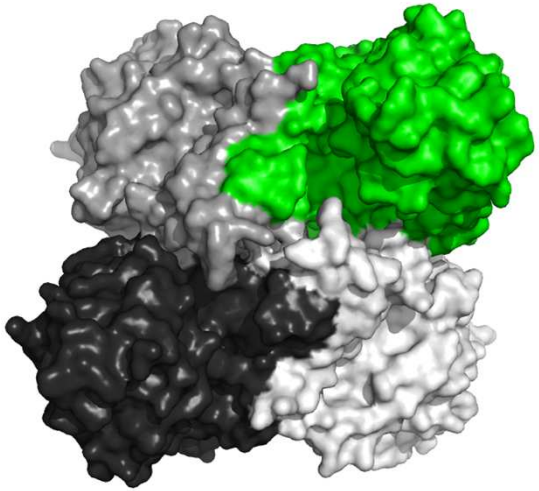
H. sapiens **Leukotriene A4 Hydrolase**
1H19



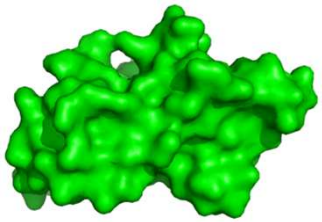
B. pseudomallei **2C-methyl-D-erythritol 2,4-cyclodiphosphatase synthase (IspF)**
3F0D



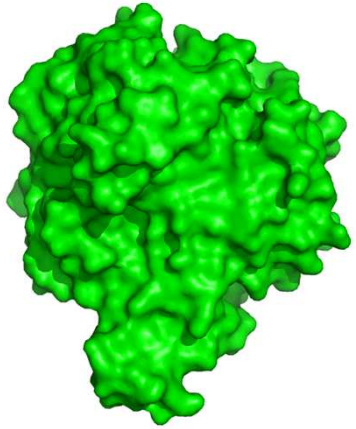
B. pseudomallei **Phosphoglyceromutase**
3EZN



B. pseudomallei **Glutaryl-CoA Dehydrogenase**
3EOM



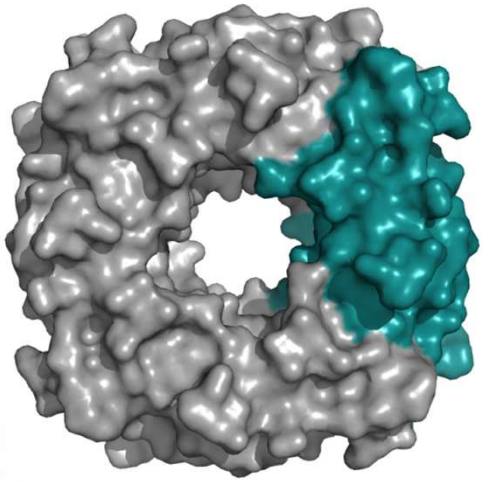
Burkholderia pseudomallei **RisS sensor protein**
3LR0



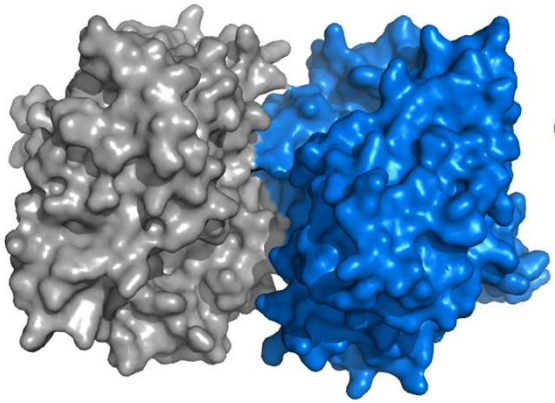
B. melitensis **Beta-Ketoacyl Synthase**
3LRF

EMERALD BIOSTRUCTURES

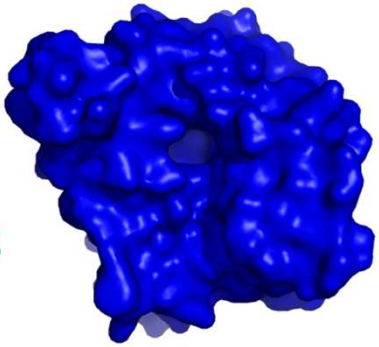
Productive Fragment Targets (Other Published Case Studies)



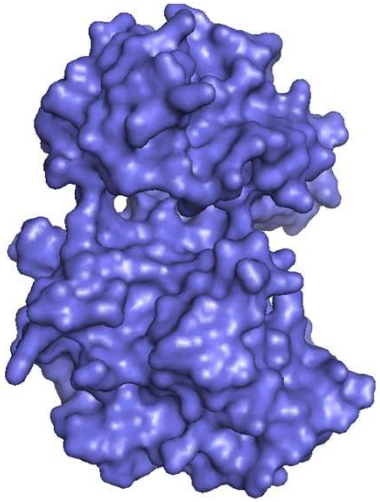
***S. aureus* 7,8-dihydroneopterin aldolase (DHNA)**
Abbott Laboratories
1RRW



***H. sapiens* PDE4**
Plexxikon
1Y2B



***H. sapiens* HSP90**
Vernalis
2WI1



***H. sapiens* P38α**
Astex
1WBO

Factors With No Significant Correlation To Tractability of Crystal for Fragment Screening:

- **Enzyme class**
- **Size**
- **Oligomeric State**
- **Precipitant**
- **pH**
- **Solvent Content**
- **Solvent Channel Size**

	Productive	Non-Productive
pH Range	4.2 - 8.5	5.4 - 8.5
Mean pH	7.03	7.08

Solvent Channel Analysis

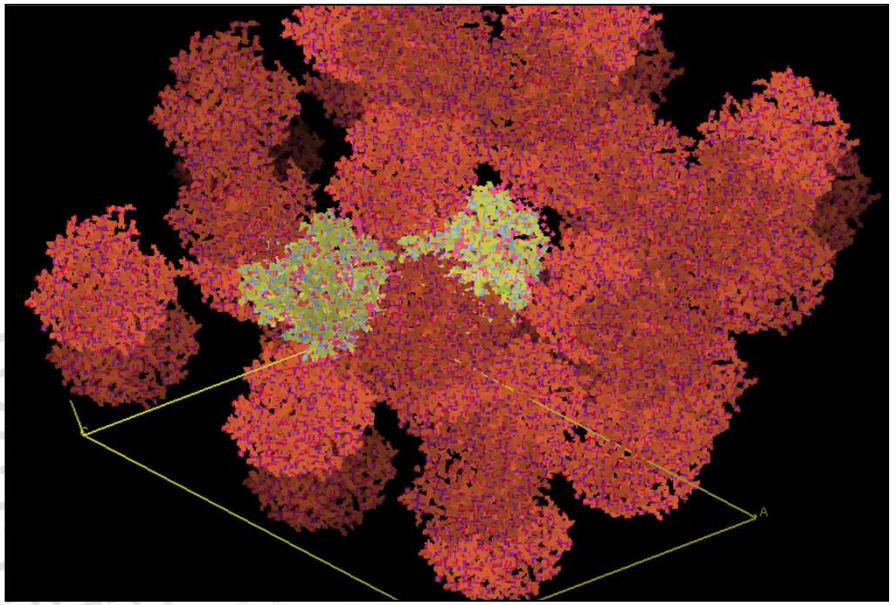
Modified program AREAIMOL (CCP4) to measure largest solvent channel

- Symmetry operators applied to create model of crystal lattice
- Probe sphere size increased until protein monomer at “center” inaccessible
- Biggest sphere that contacts central monomer ~ solvent channel diameter

	Productive	Non-Productive
Solvent Content Range	39% - 62%	39% - 50%
Mean Solvent Content	53.3%	44.2%
Solvent Channel Diameter	18 - 34	15 - 31
Mean Channel Diameter	25.8	25.0

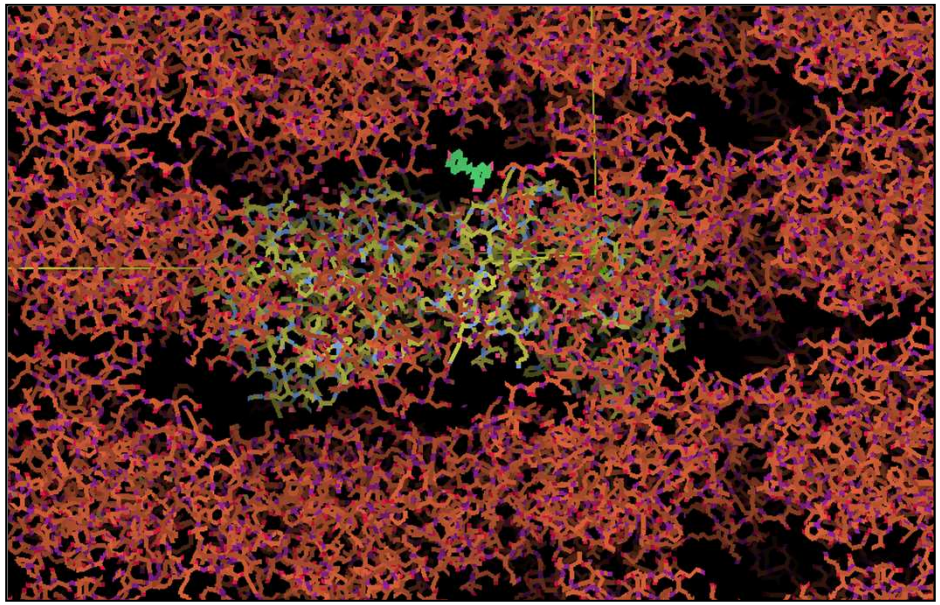
Solvent Channel Analysis

Productive Target



3F0D: solvent content: 45 %, diameter 34 Å

Non-Productive Target



3KHW: solvent content: 43 %, diameter 15 Å

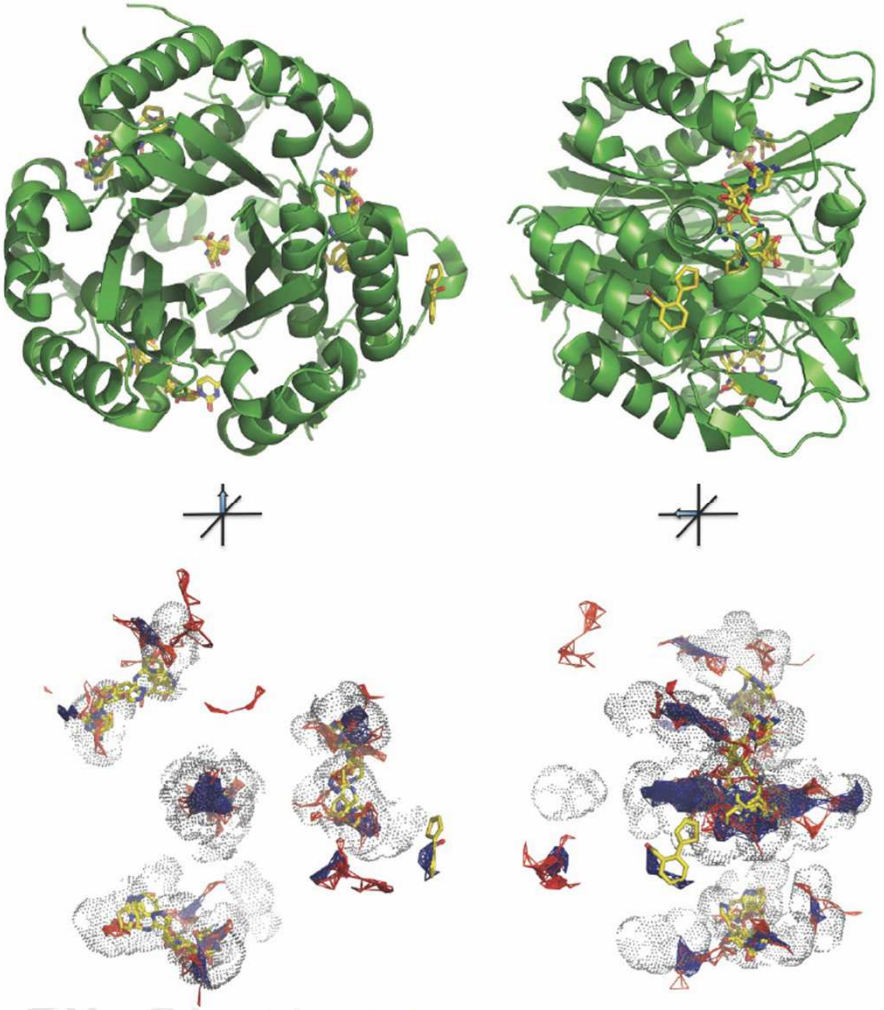
Pocket-Finding Software

- **Pocket Finder (P-Pocket)** http://www.modelling.leeds.ac.uk/pocketfinder/
  - spatial/directional algorithm based on LIGSITE which effectively pushes a neutral sphere through a protein structure while taking note of its surroundings.
  - Hendlich, M., et al. (1997). *J Mol Graph Model.* **15,** 359-63, 389.

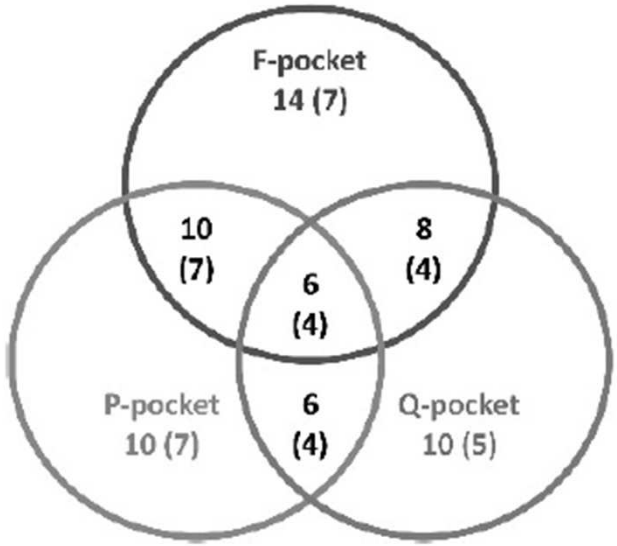
- **Q-Sitefinder (Q-Pocket)** http://www.modelling.leeds.ac.uk/qsitefinder/
  - Method positions and clusters methyl (CH_3) probes to the protein surface followed by calculation and ranking of likely pockets based on predicted binding energies for such hydrophobic clusters.
  - Laurie AT, Jackson RM (2005). *Bioinformatics,* **21:** 1908-1916

- **F-Pocket** http://bioserv.rpbs.univ-paris-diderot.fr/fpocket/
  - Detects protein cavities by scanning, categorizing, clustering and ranking sets of alpha spheres which can be drawn within a 3D protein structure
  - Le Guilloux, V., et al. (2009). *Bioinformatics*. **10,** 168

Pocket Prediction for IspF

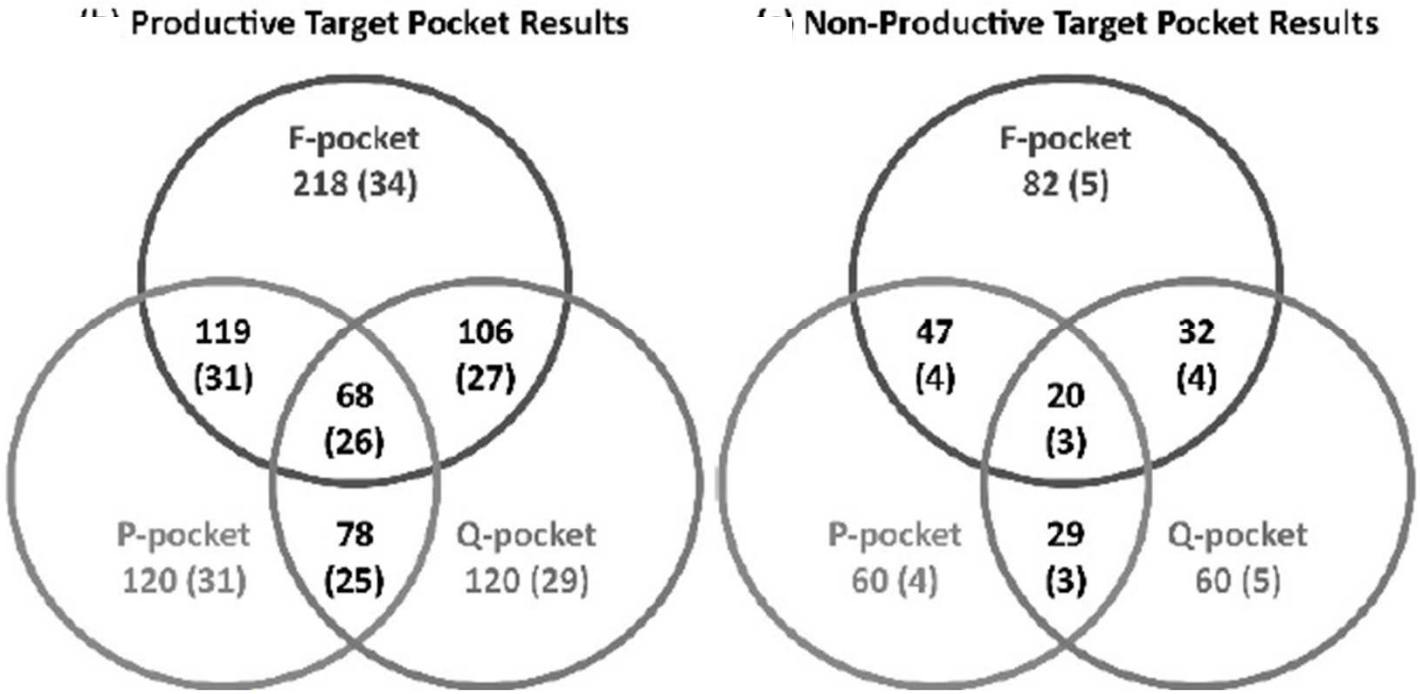


Bound Ligands
P-Pockets
Q-Pockets
F-Pockets



EMERALD BIOSTRUCTURES

Consensus Prediction Gives Best Agreement With Experimental Results



Numbers of predicted pockets (numbers of predicted pockets shown experimentally to bind ligands)

Consensus (C-Pocket): Determined by visual inspection of overlap of pockets-consensus determined when >50% overlap between P,Q and F.

EMERALD BIOSTRUCTURES

Pocket Factor Analysis and "Pocket Factor"

PRODUCTIVE					
PDB	known sites	total F-Pockets	total C-Pockets	Pocket Factor*	C-Pockets w/ ligands
3lr0	1 (2)	10	5	26	1
1wbo	1	15	7	20	1
PRP4	N.A.	11	4	14	N.A.
3f0d	8	14	6	14	4
1h19	5	19	8	13	4
PRP1	N.A.	13	5	13	N.A.
1rrw	1 (4)	28	6	13	4
3ezn	1 (2)	13	5	11	2
3eom	3 (12)	44	8	11	7
1y2b	2 (4)	19	7	11	2
2wi1	1	4	2	10	1
3lrf	1	22	5	6.6	0
	SUM:	212	68		26
	MIN:	4	2	7	
	MAX:	44	8	26	
	MEAN:	18	6	14	
	STDEV:	10	2	5	

NON-PRODUCTIVE					
PDB	known sites	total F-Pockets	total C-Pockets	pocket factor*	C-Pockets w/ ligands
3khw	1	10	3	16	0
3eiz	3 (18)	18	8	7.7	1
PRP3	N.A.	12	3	7.4	N.A.
3gvg	1 (2)	12	3	6.3	2
PRP2	N.A.	4	1	3.9	N.A.
3gwa	0	22	2	3.1	0
	SUM	78	20		3
	MIN	4	1	3	
	MAX	22	8	16	
	MEAN	13	3	7	
	STDEV	6	2	5	

- Pocket factor: total number of C-pockets per volume of protein (in 100 nm^3) as calculated by Q-Pocket.
- Average pocket factor for Productive proteins was twice that of Non-productive, at a 98% confidence level (p = 0.024)

Conclusions

- **Fragment screening has a place in structural genomics efforts**
- **Tractability of crystal system to fragment screening is impossible to predict with certainty**
- **“Pocket Factor” shows statistically significant correlation**
  - ***Not* absolutely predictive**
  - **may be a useful metric for prioritizing multiple targets for fragment screening**

Acknowledgements

- **Crystallography /FOL Team (Emerald)**

← Darren Bagley

Robert Hartley →

← Tom Edwards

Jan Abendroth →

← Jeff Christensen

Jess Leonard →

← Michele Dieterich

Bart Staker, Sr. Dir. →

← Alex Burgin, COO

Lance Stewart, CEO →

- **SSGCID**
  - Dr. Peter Myler (SBRI)
  - Dr. Gabriele Verani (UW)
  - Dr. Gary Buchko (PNNL)
  - Dr. Wes Van Voorhis (UW)
  - Robin Stacy (SBRI)
  - Bart Staker (Emerald)
  - Alberto Napuli (UW)

- **SSGCID Funding from NIAID HHSN272200700057C**

EMERALD BIOSTRUCTURES