GSB/FBDD

Macromolecular Crystallography: An Industrial Perspective

Stephen R. Wasserman

Aplicación de la Radiación Sincrotrón a la Caracterización de Materiales

October 28, 2010

Lilly

Answers That Matter.

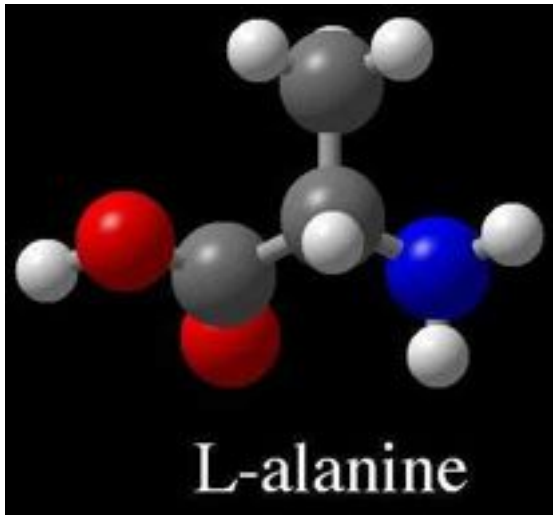
Outline

- Develop a qualitative, intuitive understanding of protein structure determination
- Little mathematics
  - Protein crystallographic analyses usually performed by biologists

- The Phase Problem
- The Beamline
- Automation
- Overview of Structure Based Drug Design
- An Example

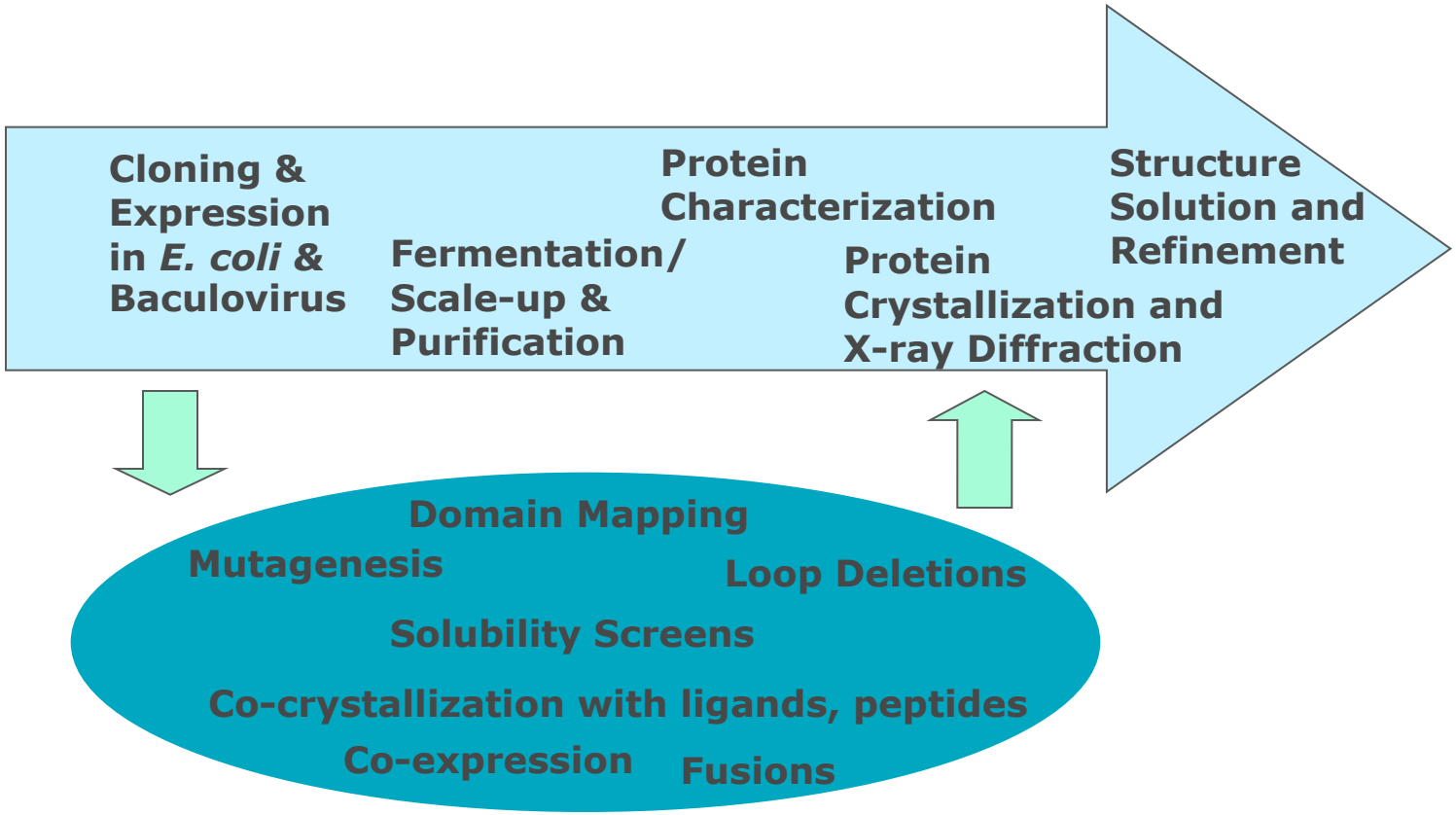
Proteins

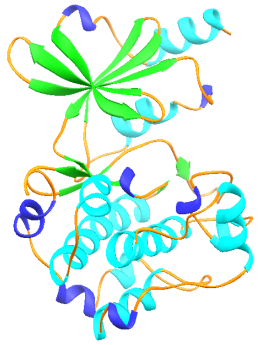
- The amino acids in proteins are chiral
  - Natural amino acids are L. Mirror image is D.
- Chirality restricts the symmetry.
  - Mirror planes are not allowed
  - Otherwise, both D and L forms of the amino acids would be in the protein
- Only 65 of the 230 space groups do not have a mirror plane



Target-to-Structure Processes

- Combination of biology, chemistry, physics and engineering
- Start with the DNA that codes for the protein

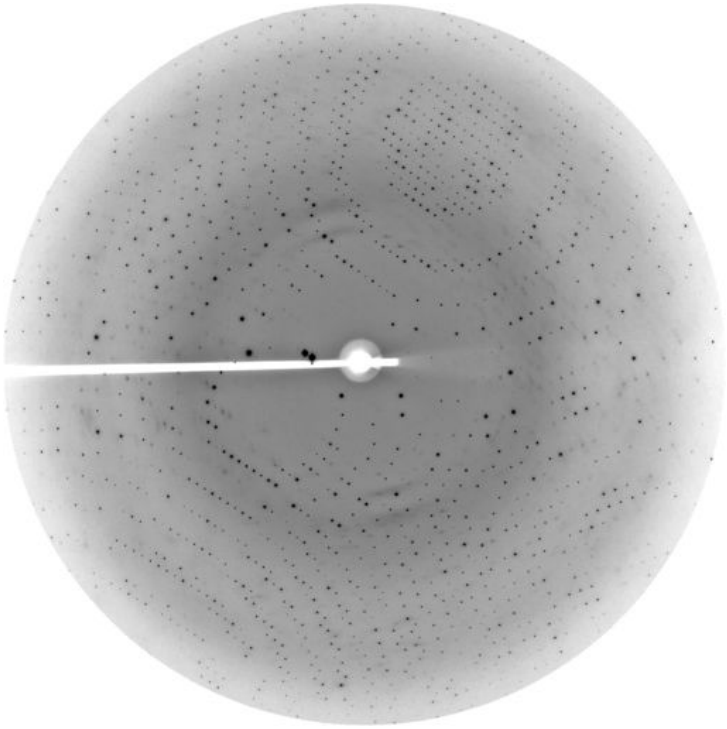




FASTEST KINASE STRUCTURE IN 28 DAYS

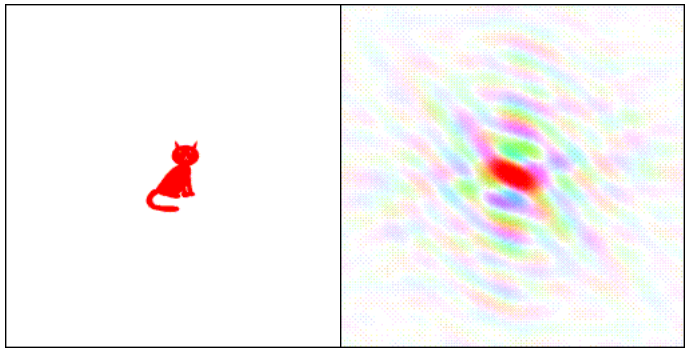
Diffraction Information

- Diffraction is the Fourier Transform of the electron density in the crystal
- In crystallographic image, measure amplitudes only
- Diffraction is a wave experiment
  - Uses phases and amplitudes of diffraction
- Can analysis use only amplitudes?

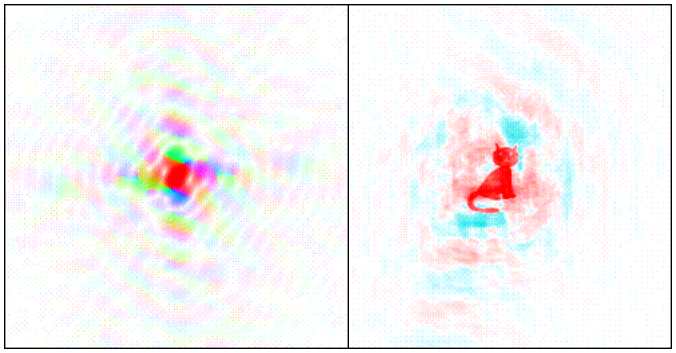


Phases vs. Amplitudes

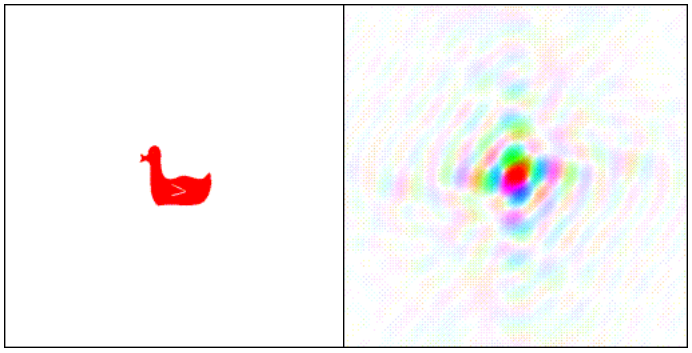
- Images from Kevin Cowtan's Book of Fourier
  - http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

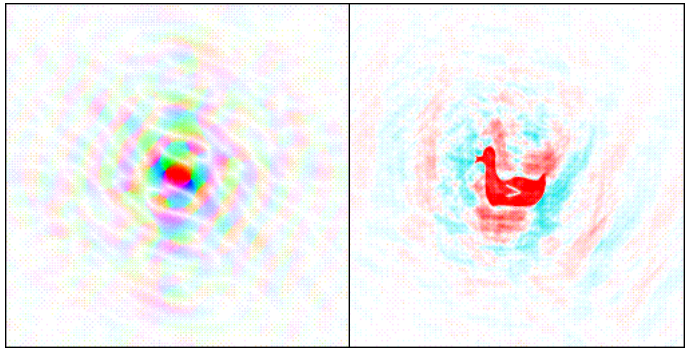


Fourier Transform



Cat Phase Duck Amplitude Fourier Transform





Duck Phase Cat Amplitude

Phases vs. Amplitudes

- Phases contain most of the information needed to solve the structure
- Same is true for EXAFS
- Methods are needed to determine the phases

- Phases measure distances between atoms

- The problem in crystallography: (at first) we do not have the most important information

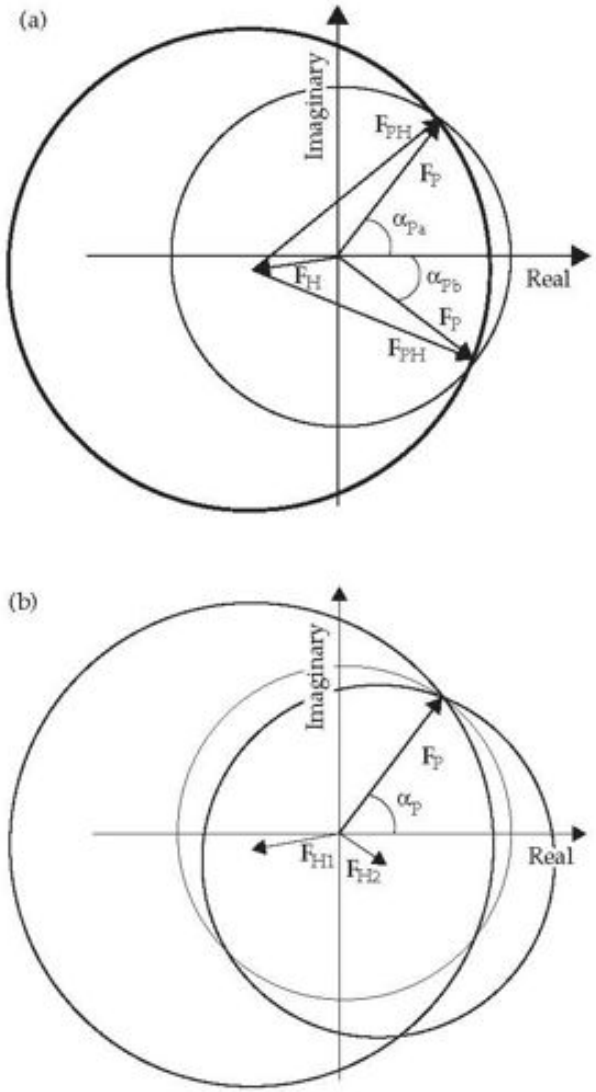
- Using a synchrotron is probably the most effective solution
  - Quality of data simplifies the analysis

Direct methods

- Phases determined from (probabilistic) analysis of measured amplitudes
- Used for small molecules
- Proteins usually too large for successful analysis

Isomorphous Replacement

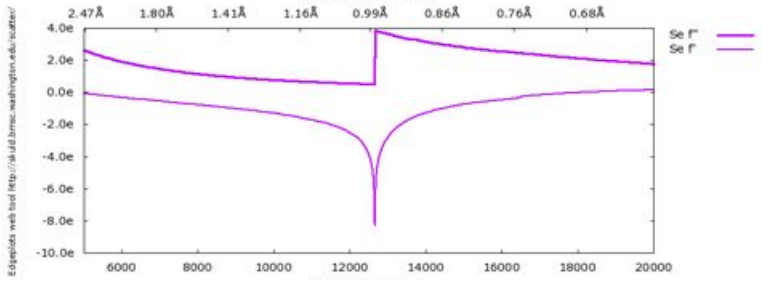
- Used before synchrotrons available
- Make derivatives by adding heavy atom to crystal
- Method requires that all crystals are isomorphous
  - Unit cell is unchanged by addition of heavy atom
- Changes in intensity of each reflection (ultimately) give the phase angle
- With one derivative, the sign of the phase is unknown
  - cos(x) = cos(-x)
- Multiple derivatives permit determination of the sign of the phase



Anomalous Dispersion

- Multi wavelength (MAD) or single wavelength (SAD) anomalous dispersion
- Usually used for new structures

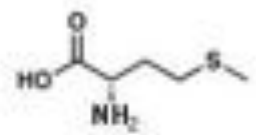
- Scattering of X-rays is usually independent of wavelength
- Exception is when the wavelength matches the binding energy of an electron in an element
- $f = f_0 + f' + if''$ f'' is the absorption spectrum
- f' can be calculated from f''
- Friedel's law is not true
  - $F_{hkl} != F_{-h-k-l}$
- In experiment, choose energy of X-ray to match the absorption edge of a element in the crystal



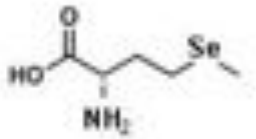
Anomalous Dispersion

- Use of anomalous scattering to phase structures moved protein crystallography data collection from the laboratory to the synchrotron
- Substitute the S (sulfur) atoms in methionine with Selenium

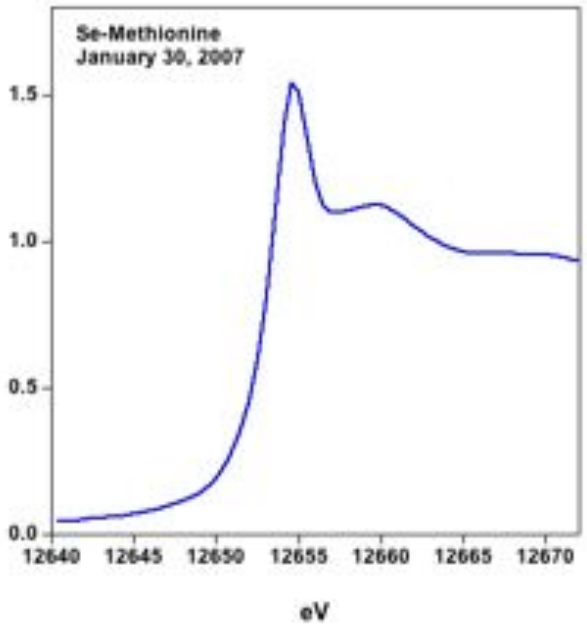
- Absorption edge of selenium, ~12.660 keV, well matched to energies available at a synchrotron
- XANES/EXAFS part of SAD/ MAD phasing
- Since know sequence of protein, the location of the selenium atoms in the proteins is also known (approximately)



Methionine



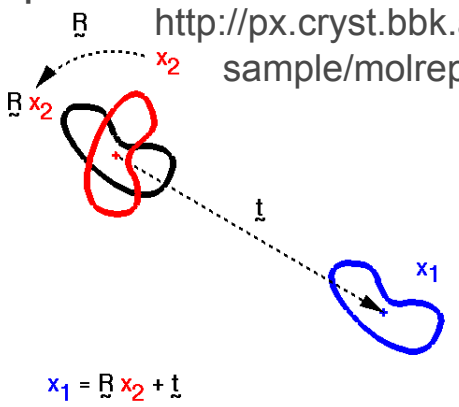
Se-methionine



Molecular Replacement

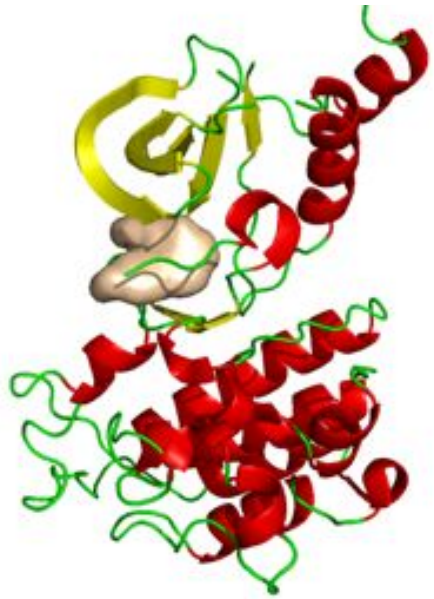
- Phases taken from a known protein
- Typical agreement (homology) is ~ 40% or greater
- Rotate and translate model protein to get maximum "agreement" between model and unknown protein
- For pharmaceutical research this is usually the most common phasing method
  - Solving structure of same protein (100% homology) with different compounds
  - "Rigid body refinement" of the ligand attached to the protein

http://px.cryst.bbk.ac.uk/03/sample/molrep.htm



Philosophy of LRL-CAT

- Integrate directly into drug discovery
- Mail samples to the beamline
- Automated
- Robust
- Rapid upgrades and maintenance
- Control processes
  - At beamline
  - Crystallization



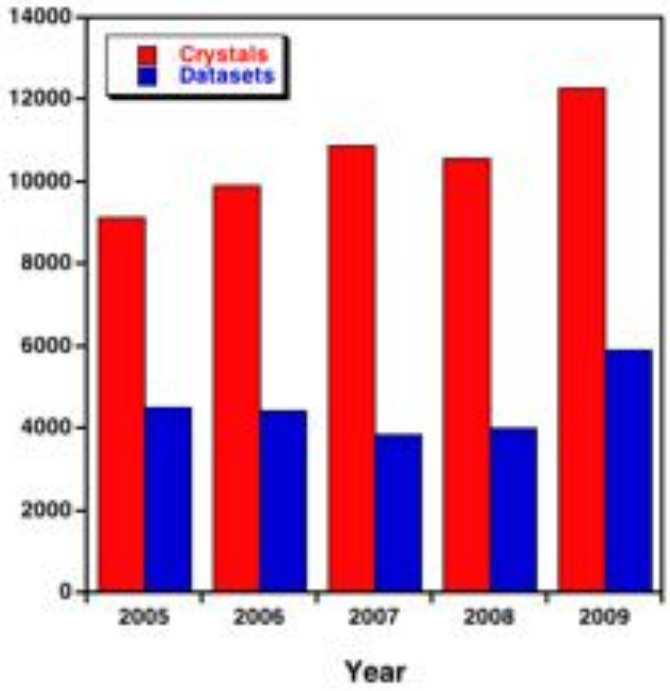
- Evolution from data collection to data management
- LRL-CAT aims to be virtually user-free

Beamline as a “Home Laboratory Source”

- APS beamline as a “home source”
  - “Walk-in” access
  - No advance scheduling

- Goal of automation to provide equivalent sample turnaround
- Lilly has decomissioned lab sources in San Diego and Indianapolis
- Available to all users

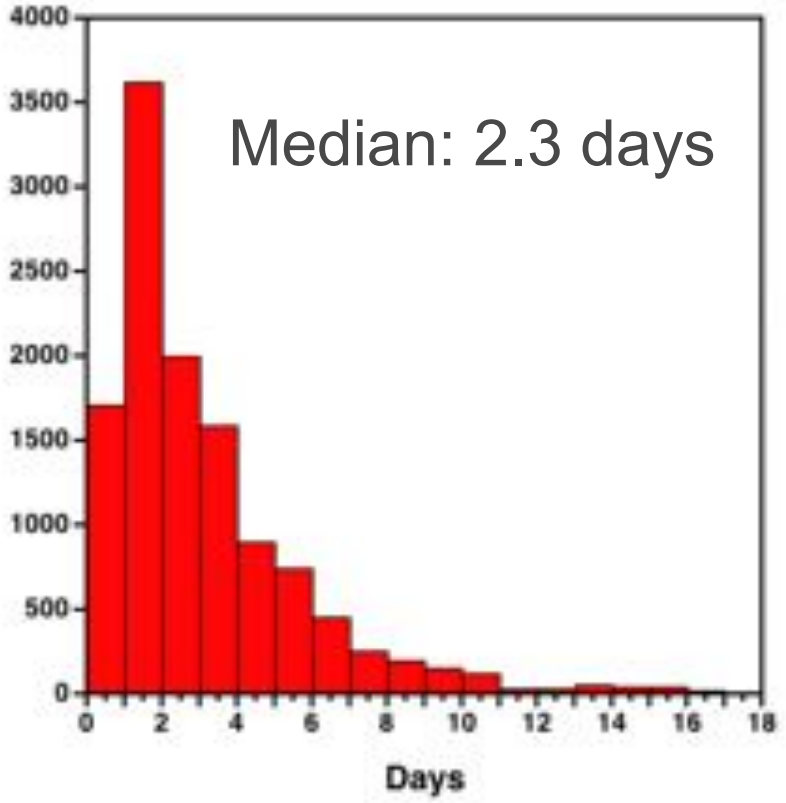
Statistics

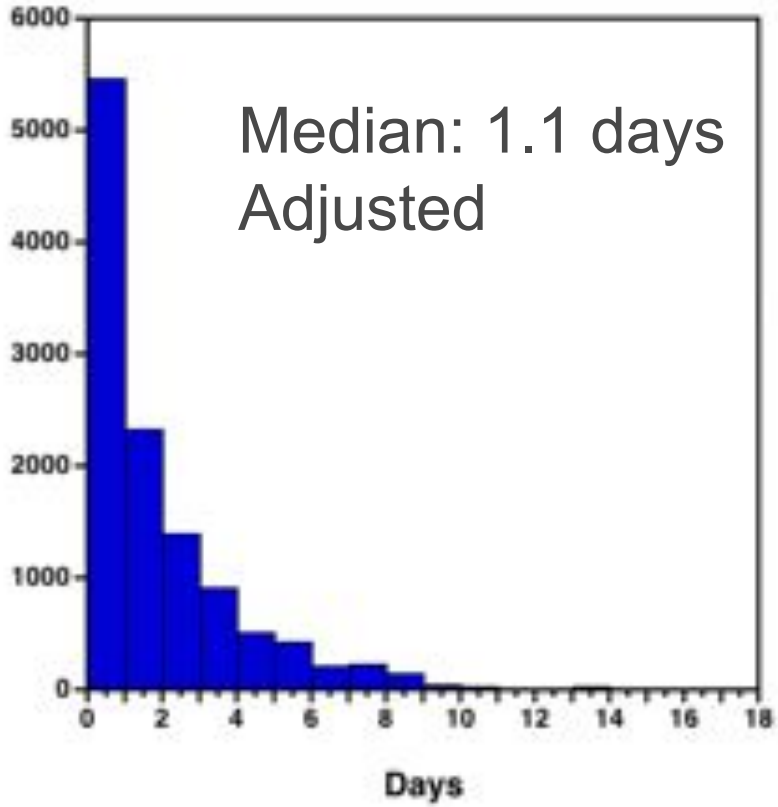


- 2009
  - 12,301 crystals
  - 5,932 datasets

- >80% of datasets scale and reach desired resolution limit
- Resolution for internal crystals (2009)
  - Median: 2.1 Å
  - Mean: 2.2 Å, Std. Dev. 0.7 Å

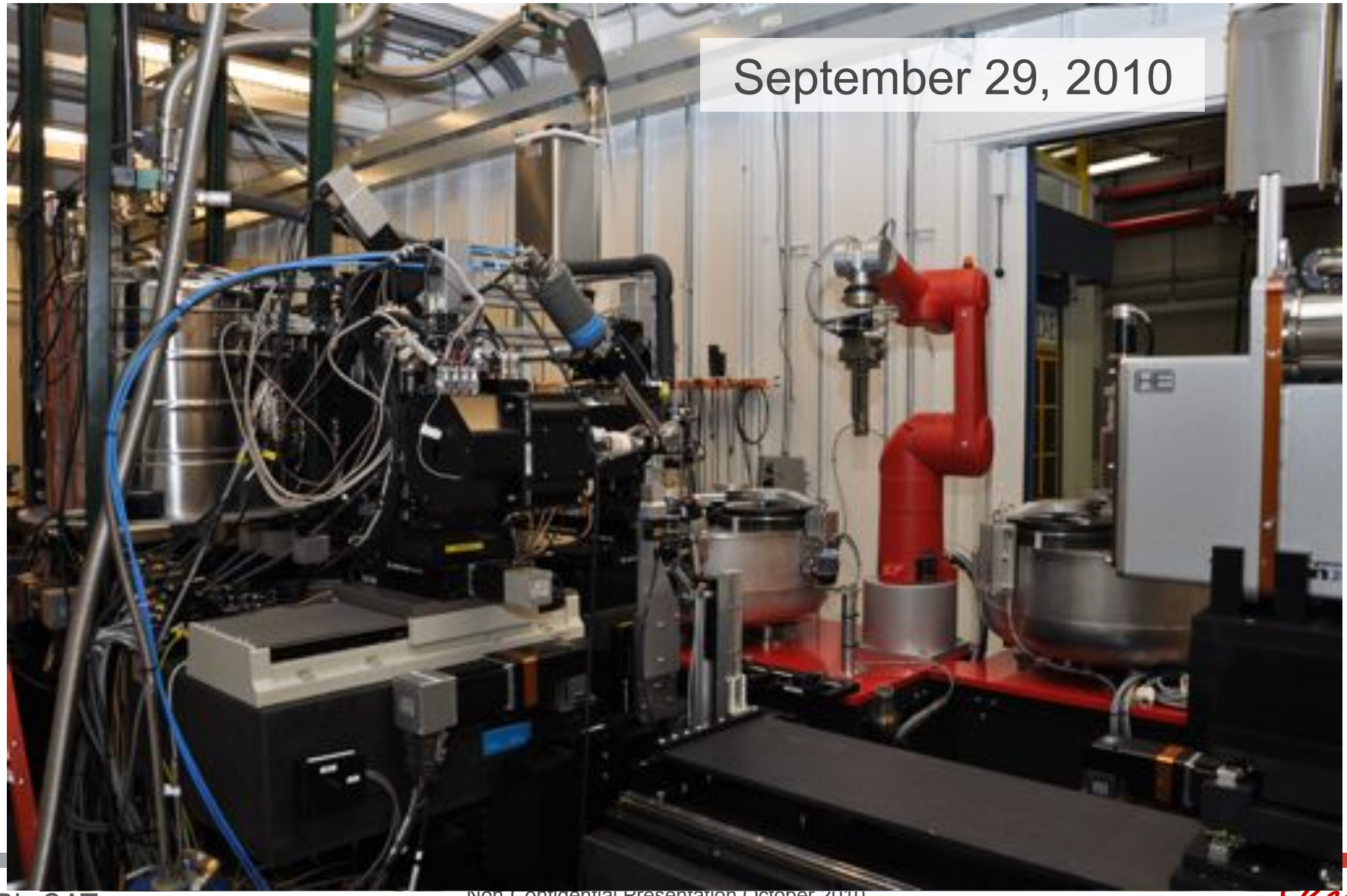
Sample Turnaround





- Red: Time from crystal harvest to completion of data collection
- Blue: Corrected for shipment and APS operations
- Transmission of processed datasets within 0.75 hours (median) of collection

LRL-CAT Experimental Station



Automated Processes

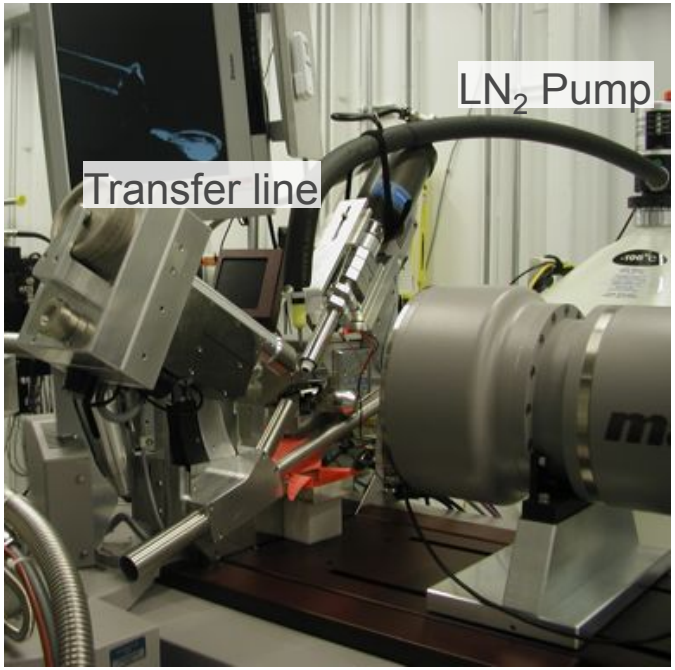
- Crystal loading
- LN_2 to remove surface ice
- Place crystal in x-ray beam (Loop centering)*
- Crystal screening
- Crystal quality score and resolution estimate*
  - Selection of best crystal within group of duplicates*
- Beam attenuation*
- Data collection*
- Data processing*
- Quality control*

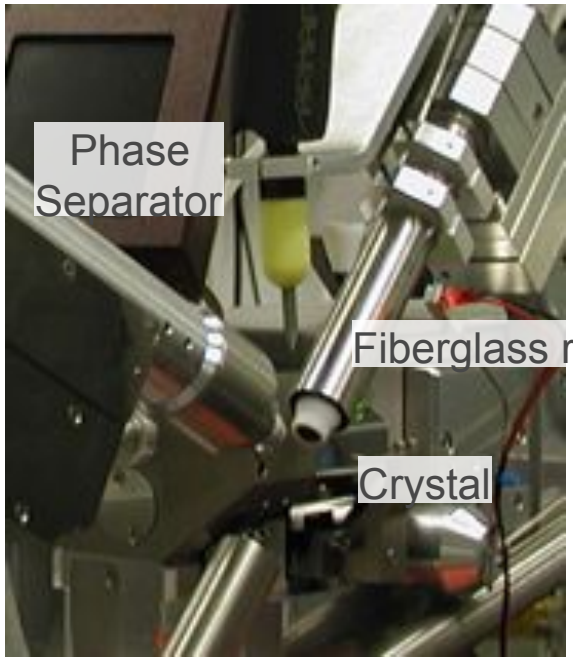
*Use automated decisions

- Approximately 70% of the data we collect is analyzed automatically by computer

Engineering: Ice Removal

- Diffraction from ice complicates diffraction from protein
- Ice prevents automatic vision systems from correctly putting crystal in X-ray beam
- Wash each crystal with a small amount of liquid nitrogen



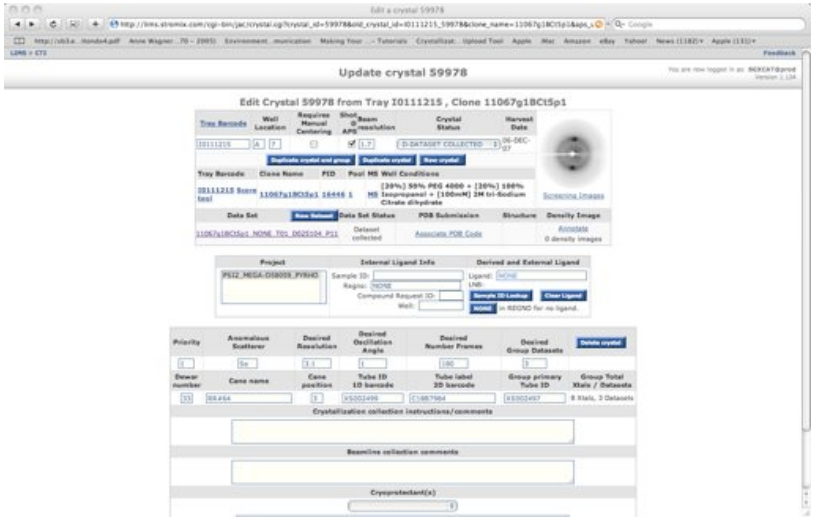


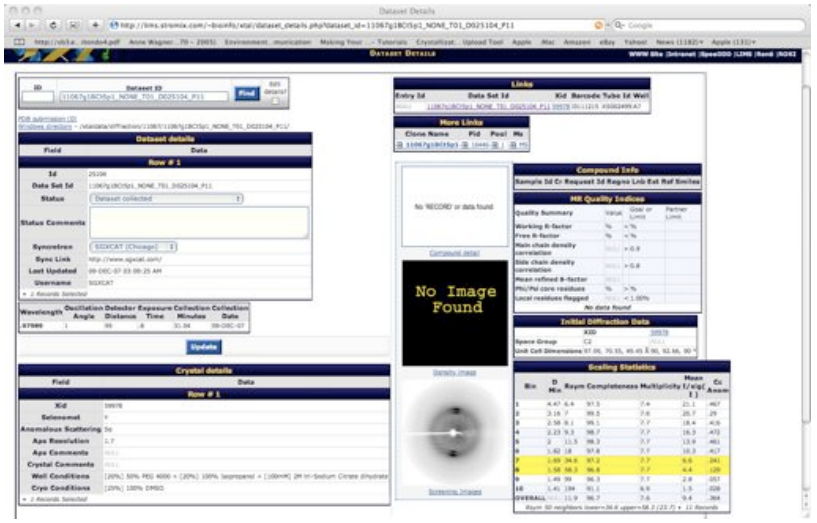
- Dehumify experimental enclosure (relative hum. ~ 20%)

Decisions Require Information

- Every sample tracked through barcodes
- All sample information and results stored in Oracle database
- Description of crystals provided electronically
  - Direct entry into database (Lilly/PSI2)

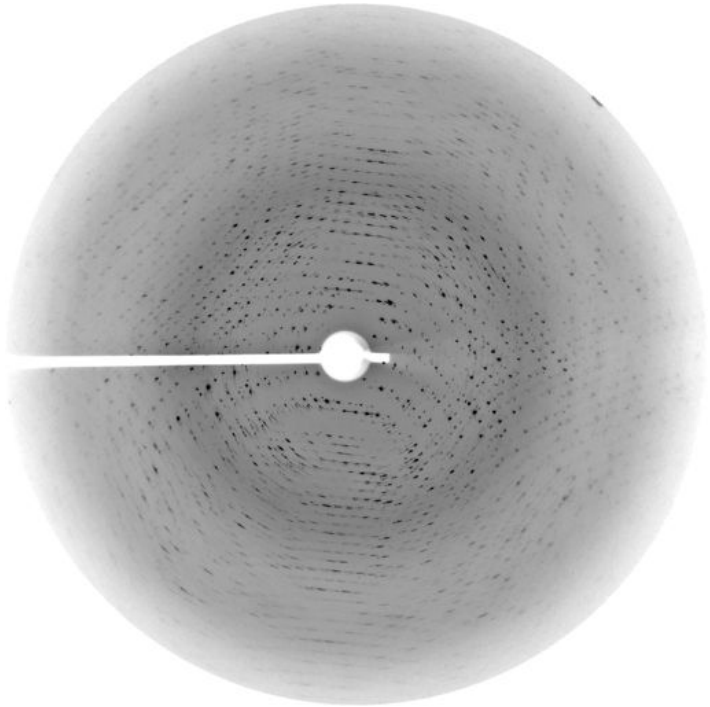
  or
  - Spreadsheet (Partners/General Users)
- Direct MPLS connection between LRL-CAT and Lilly-San Diego/Lilly-Indianapolis
- Database used simultaneously at the beamline and Lilly San Diego and Indianapolis locations
- Rapid access to results
  - Screens: 4 minutes
  - Datasets: 20 to 180 minutes





Crystal Evaluation

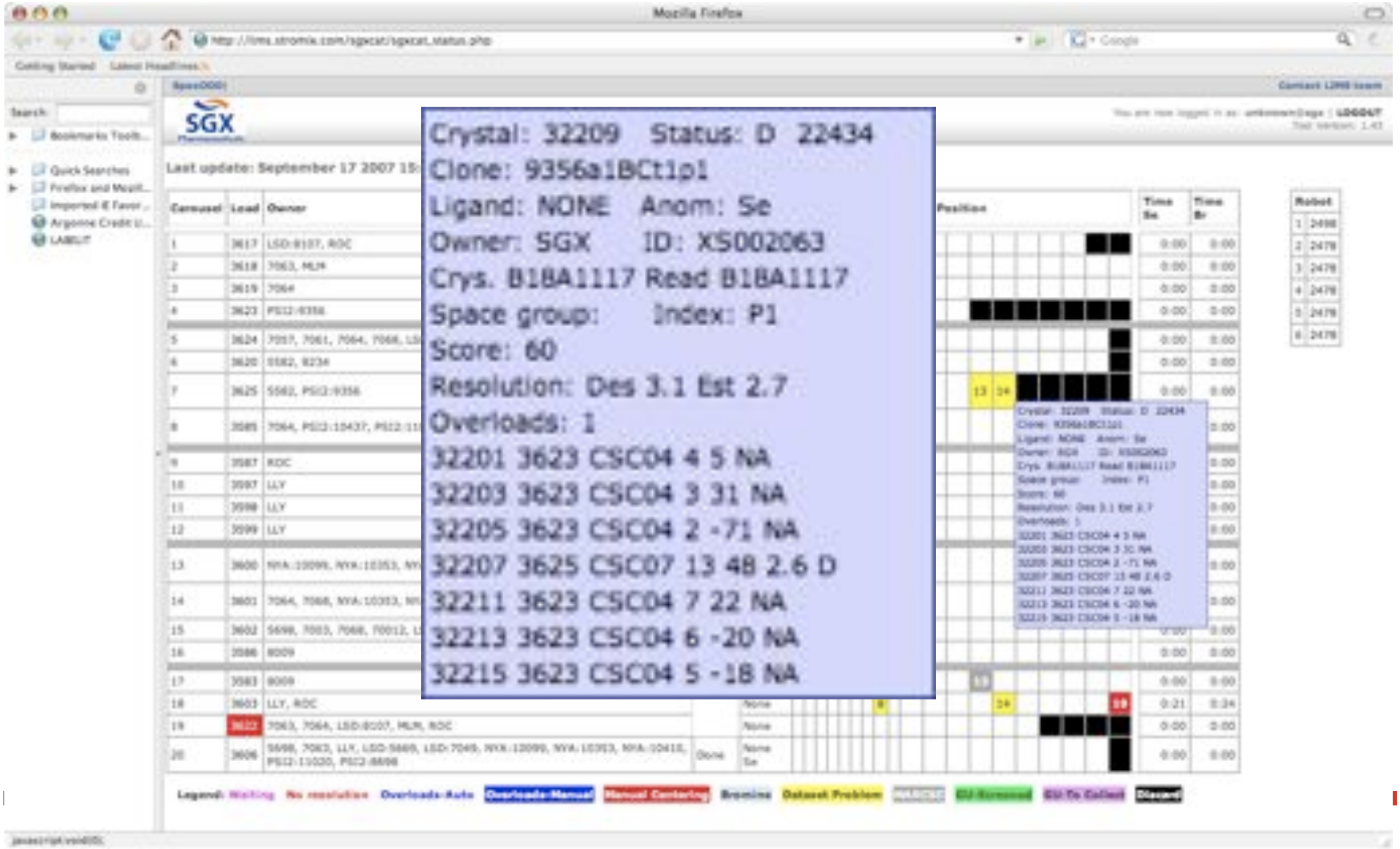
- Crystals automatically evaluated after screening
  - Crystal quality
  - Resolution limit of observed diffraction
- Score of quality
  - Scores based on 11 criteria including:
    - I/σ
    - Spot shape
    - Ability to index
    - Mosaicity
    - Ice rings
    - Percentage of spots indexed
- Resolution estimates
  - Usually within 0.3 Å of the actual resolution
  - Intentional overestimate
- System optimized for specific beamline
- Score, resolution and images loaded automatically into database
- 45,999 crystals scored at LRL-CAT (through 12/22/2009)



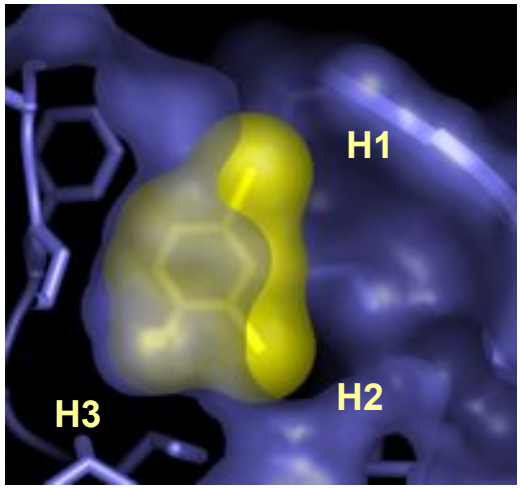
Status Monitor

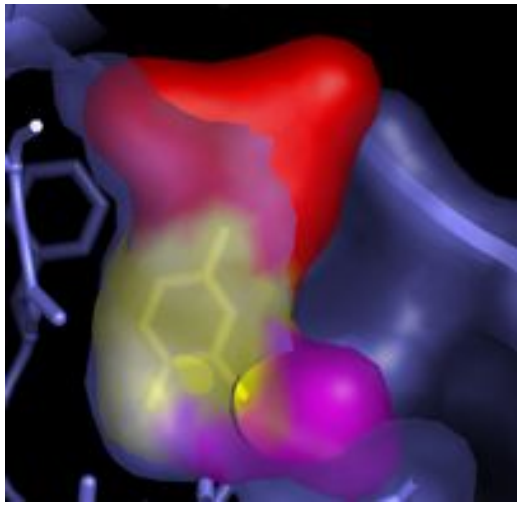
- Tracks up to 540 crystals
  - Row -> carousel
  - Box -> crystal
- ~45 fields for each crystal (total >20,000 fields)
- Updates every 15 minutes directly from database
- Decision: Selects crystals for collection
  - Crystal score (diffraction quality)
  - Resolution
  - Best of set of identical crystals (multiple datasets supported)
- Flags potential problems
  - Centering – Did the software place the crystal in the beam?
  - Overloads – reflections that saturate the detector

Status Monitor



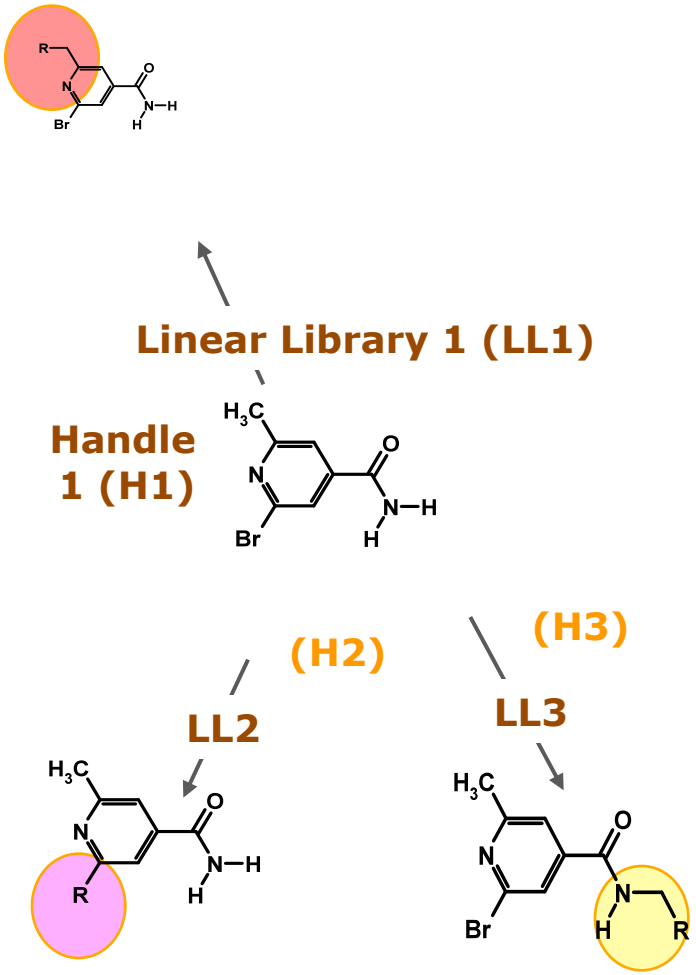
Design a molecule





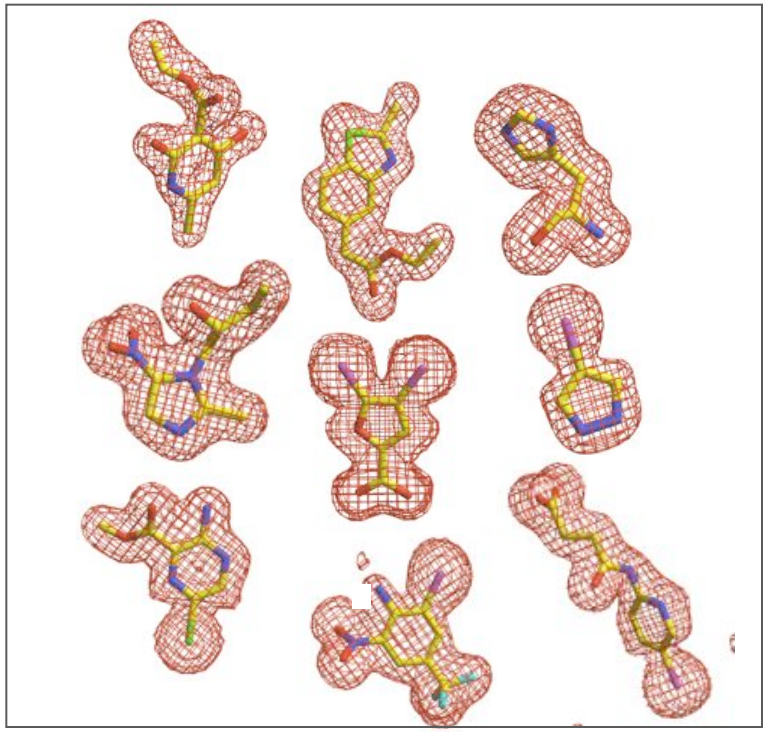
- **Optimize interaction of small molecule with protein**
- **Screen library of scaffolds/ fragments**
- **Literature precedents**
- **Process guided by crystal structures**
- **Pick “promising” opportunities**

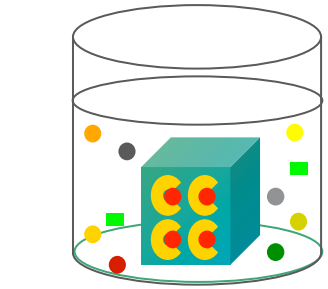
FAST Fragment Library Design



- ~1500 fragments (scaffolds)
- Fragments have 2-3 handles for auto synthesis
- ~26 different synthetic handles support >50 chemistries
- Sites in fragment bind to target active sites
- Library supports rapid synthesis of a new molecule with (hopefully) improved properties

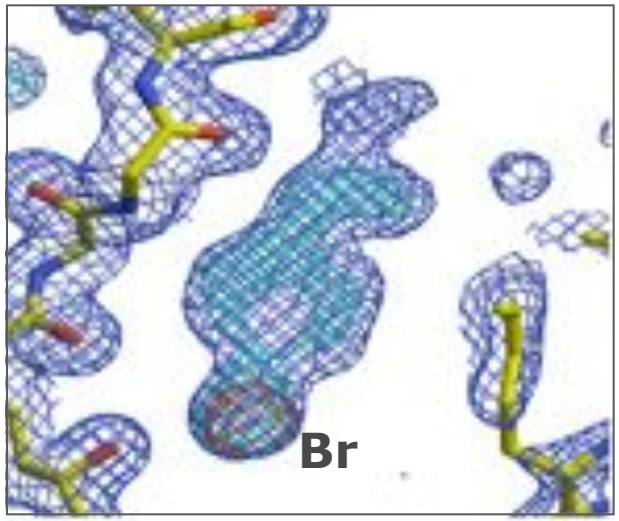
Crystallographic Fragment Screening





Crystal soaking and data collection

Hit rate=1-5%

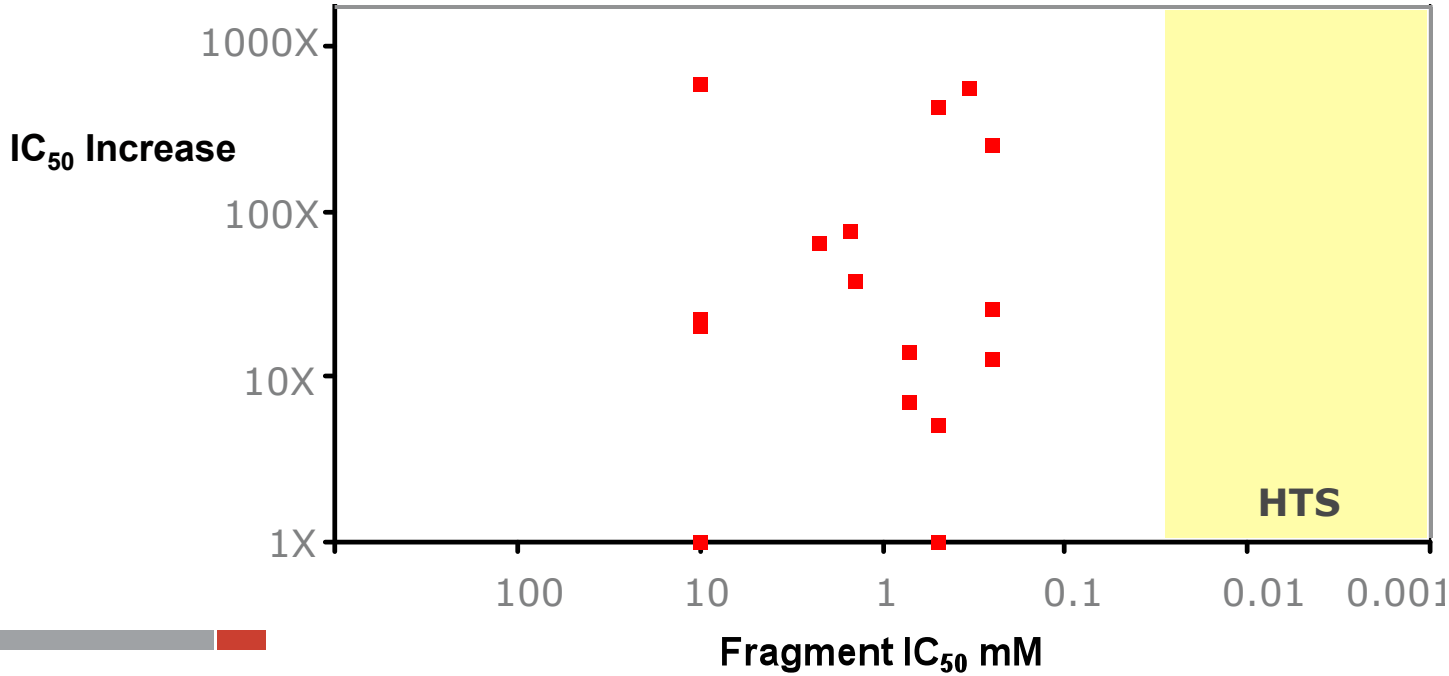


- Fragment library is divided into pools of 10 shape diverse compounds→soaked into preformed crystals
- Bound fragment is identified from shape of electron density and anomalous signal from bromine
- Complementary to other biophysical and biochemical analyses

mM Hits Can Be Optimized!

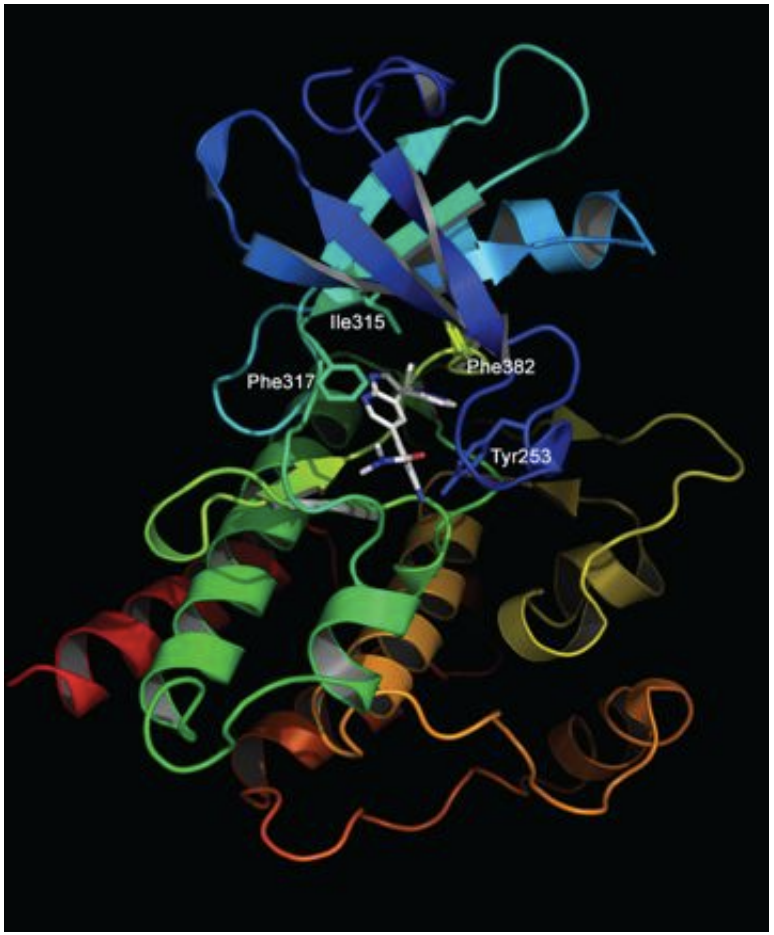
- Up to 550-fold activity increase from weak scaffolds in first iteration
- Scaffold screening detects *optimizable* hits that would be missed by HTS
- Starting potency doesn't correlate with increase

IC_{50} Increase vs. Starting Fragment IC_{50} for 5 Kinases



SGX393

- Structure of azaindole bound to BCR-Abl
- Further development led to SGX393
- SGX393 reduces number and range of mutations in leukemia cell lines
- Combination with another inhibitor preempts emergence of mutations
- Successful IND at FDA in mid-2008



T. O'Hare *et al.*, *PNAS*, **2008**, *105*, 5507-5512.

Why use the synchrotron?

- Upstream
  - Smaller crystals
    - Laboratory source
    - Other US synchrotrons
  - Savings in crystallization time
- At APS
  - Intense source
  - Tunable – anomalous experiments
  - High data rate
- Downstream
  - Data quality
  - Automatic data analysis

Acknowledgments

- John Koss
- Laura Morisco
- Sonal Sojitra
- Stephen Burley
- Horst Hemmerle

- Lilly Crystallography, Crystallization, Bioinformatics, and Automation Groups
- Use of the Advanced Photon Source is supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.