UNIVERSITY of VIRGINIA

HKL-3000

Toward the future of protein crystallography

Argonne Users Meeting, May 2008

Long and winding road (with traps)

Data reduction

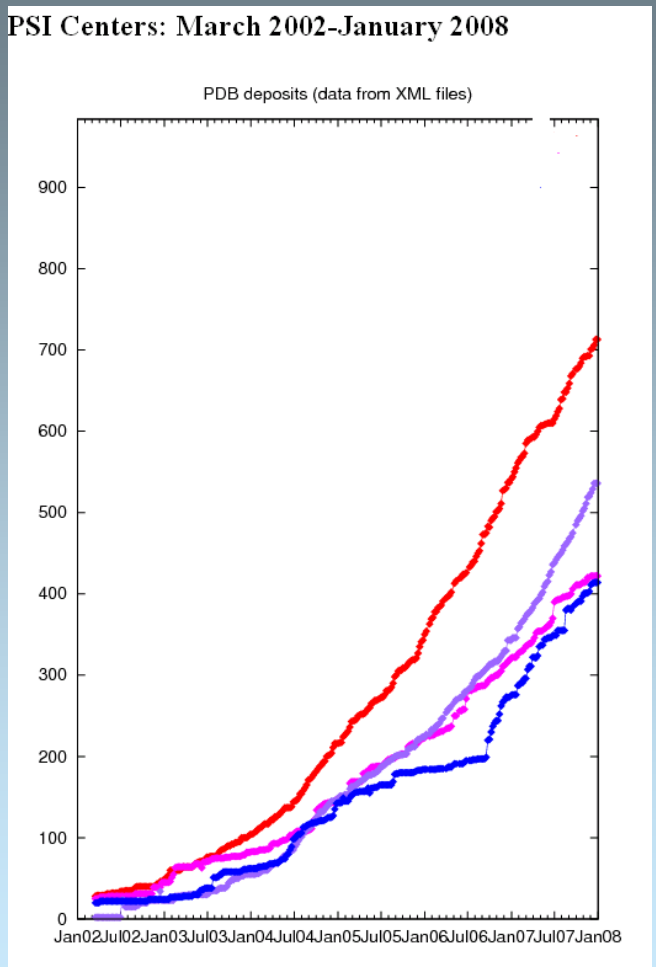
Phasing

Data collection

Project start

Protein production
crystallization

SG dynamics



In every PSI Large Center

Roughly

new structure every 48 hours

Current situation

W0308

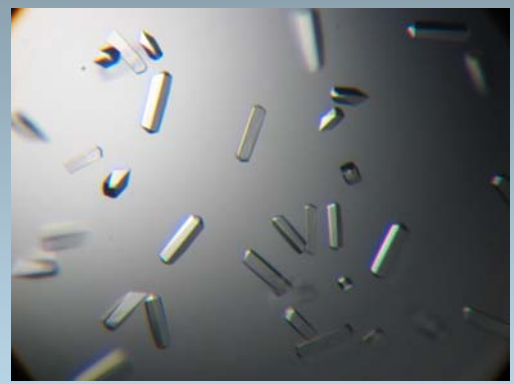
Teaching Elves to Collect Data: An Analysis of the Last Million Diffraction Images from ALS 8.3.1.
James Holton, Physical Biosciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

Most X-ray data sets collected at synchrotron sources do not produce usable results. An analysis of data collected in 2003 at the ALS beamline 8.3.1 shows that 2346 datasets were collected and 41 structures were deposited in the PDB. Although it is understandable that not every dataset leads to a published structure, it is troubling that ~98% of them do not. This large gap between collected data and useful results is not unique to 8.3.1. The 28 operating American PX beamlines collect ~100,000 datasets/year. This suggests that a great deal of improvement in scientific productivity can be attained if the reasons for failed projects are better understood.

57 datasets/deposit !!!!
Worldwide < 40 hours/deposit
~ 40-150 crystals/structure

Priceless crystals

Star of Africa



Crystallography lab

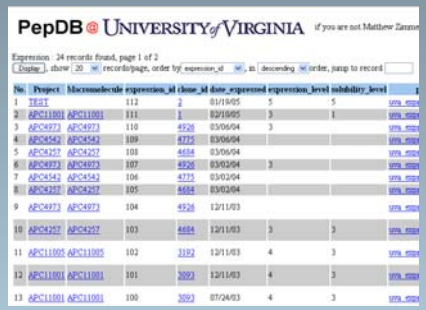
How to define high throughput at the beamline ?

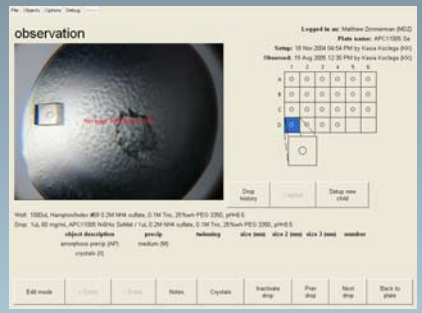
Integration
HKL-3000 is a pipeline component

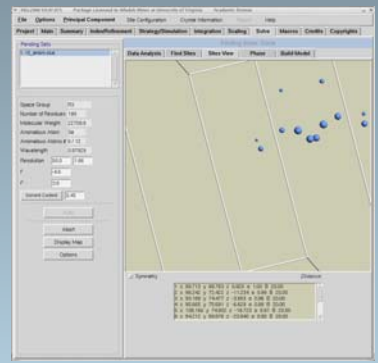
Pepdb
expression & purification

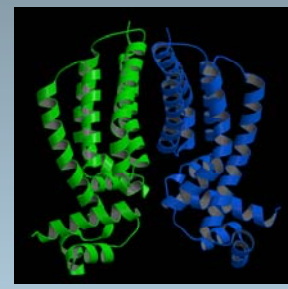
Xtaldb
crystallization

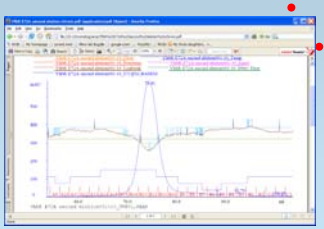
HKL-3000
Data collection,
diffraction & structure solution

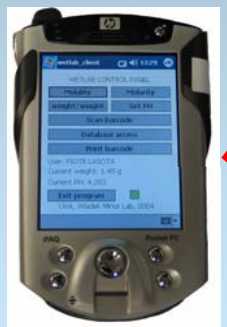






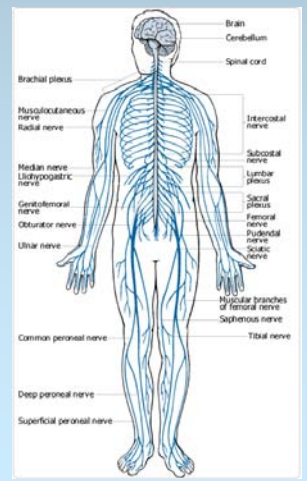




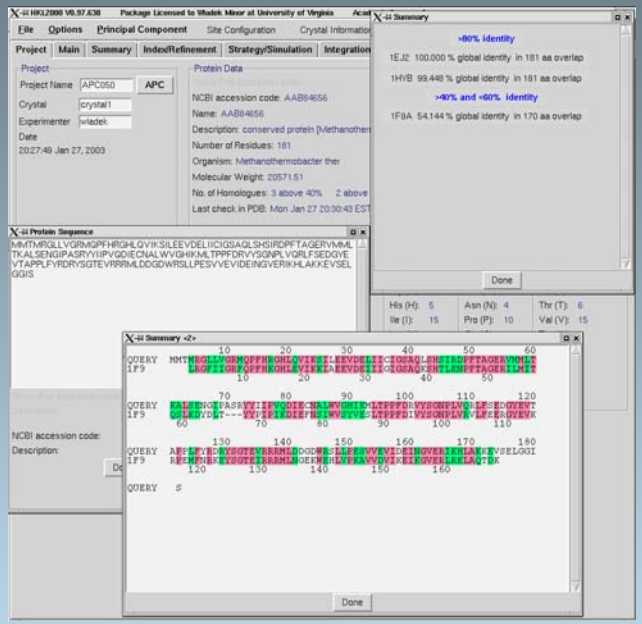


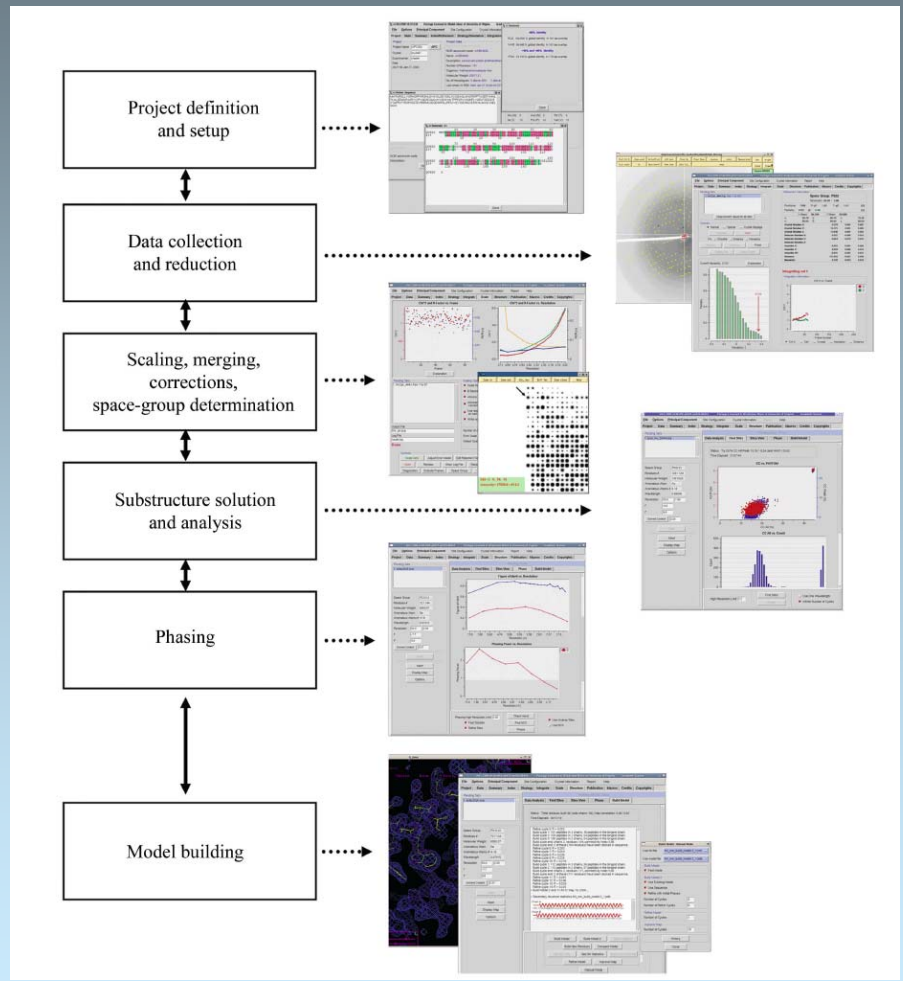
Wetlab
chemicals & solutions

Other systems provide additional contextual up- and downstream information



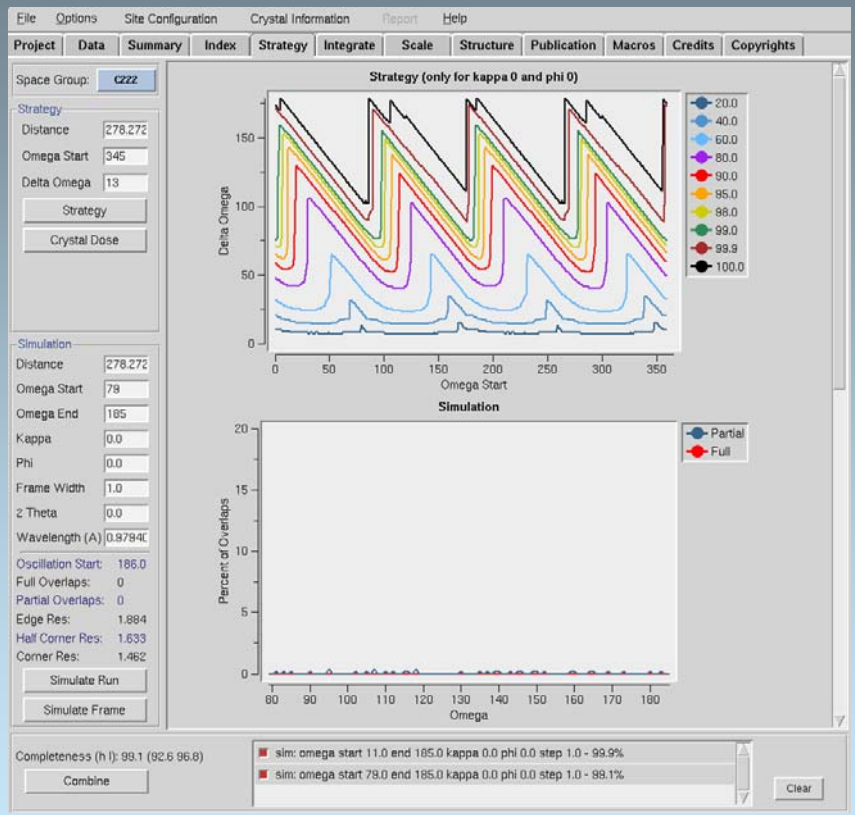
Project
Check against DB during data collection

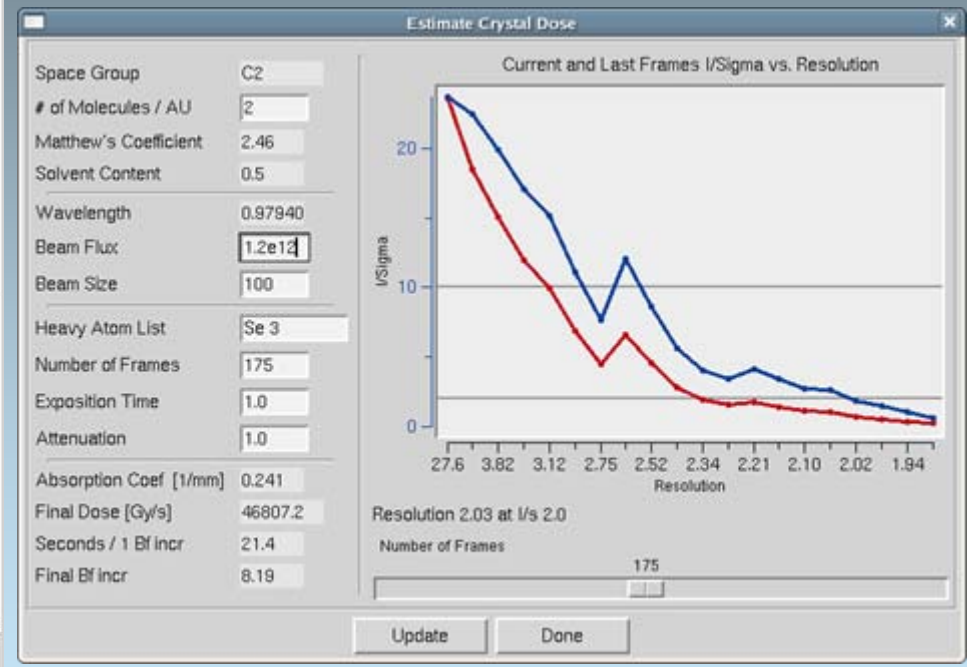




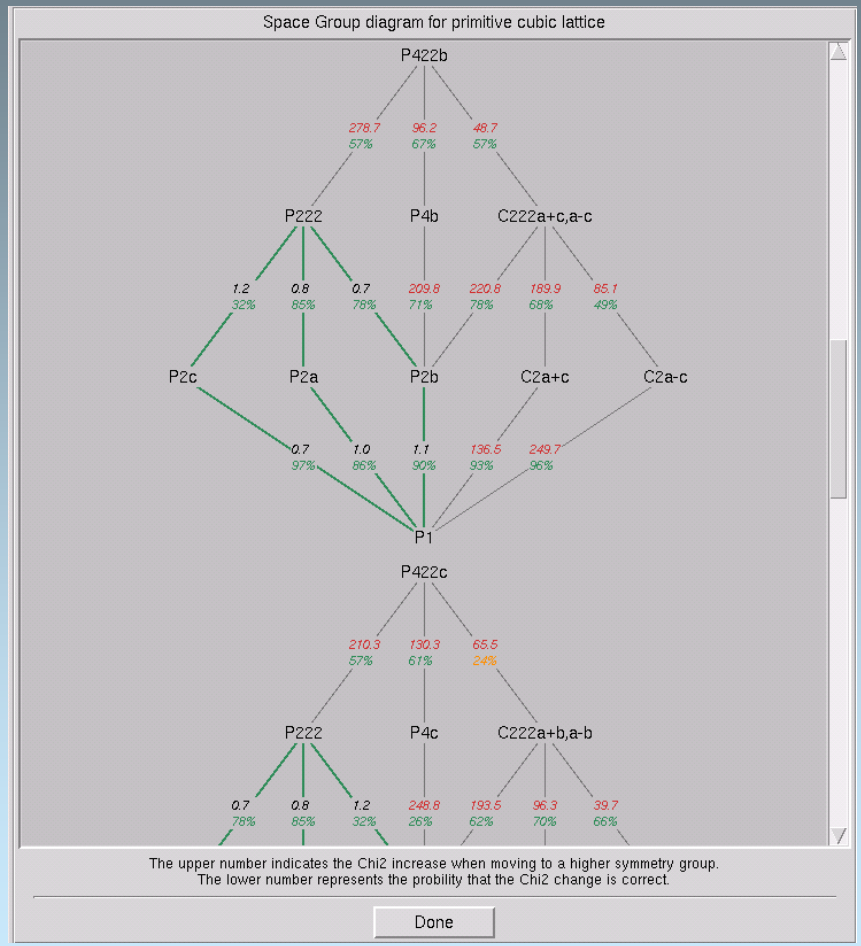
HKL-2000
SHELXD, SHELXE
CCP4
SOLVE, RESOLVE
ARP/WARP
O, COOT,CCP4

Completeness – overlaps
Multi-crystal strategy



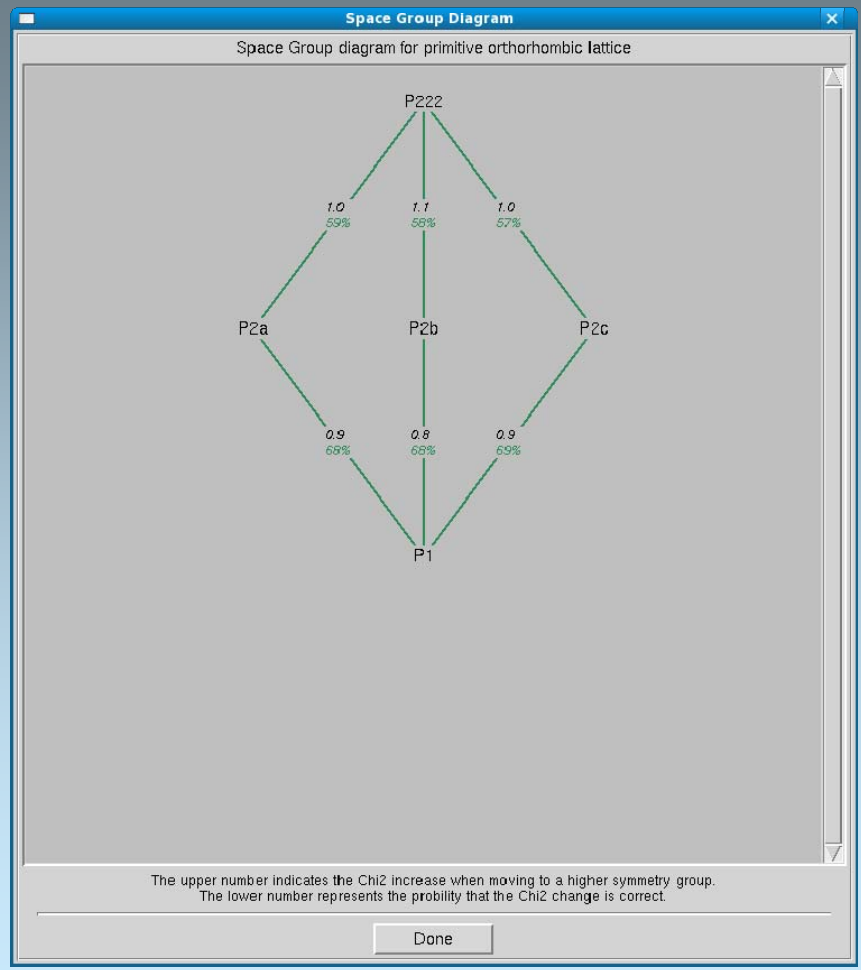


Space Group Determination

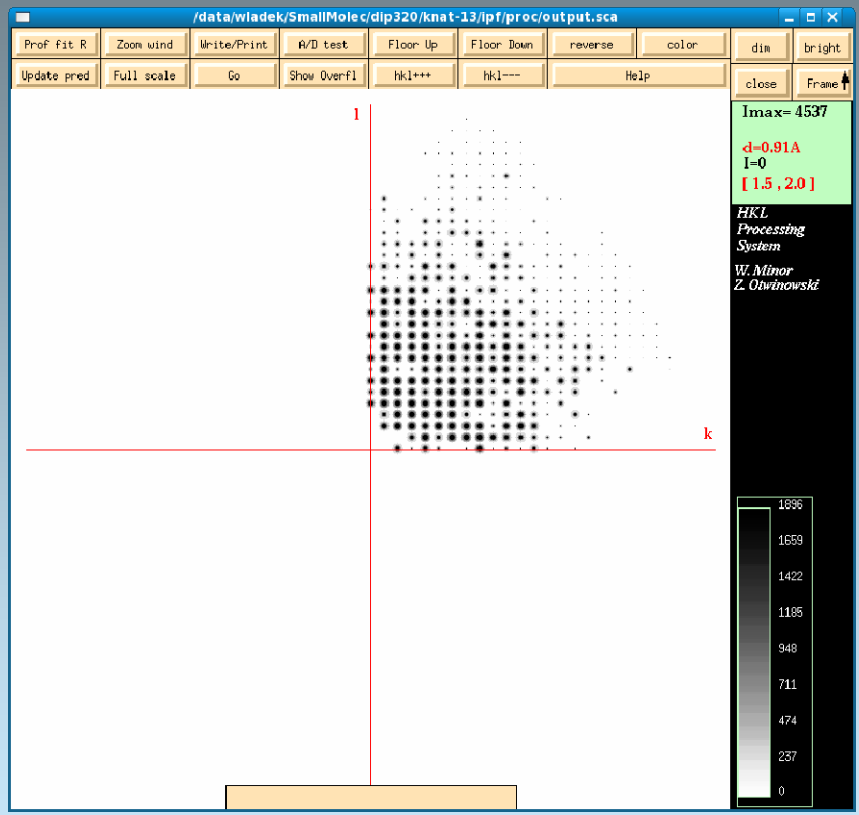


Argonne Users Meeting, May 2008

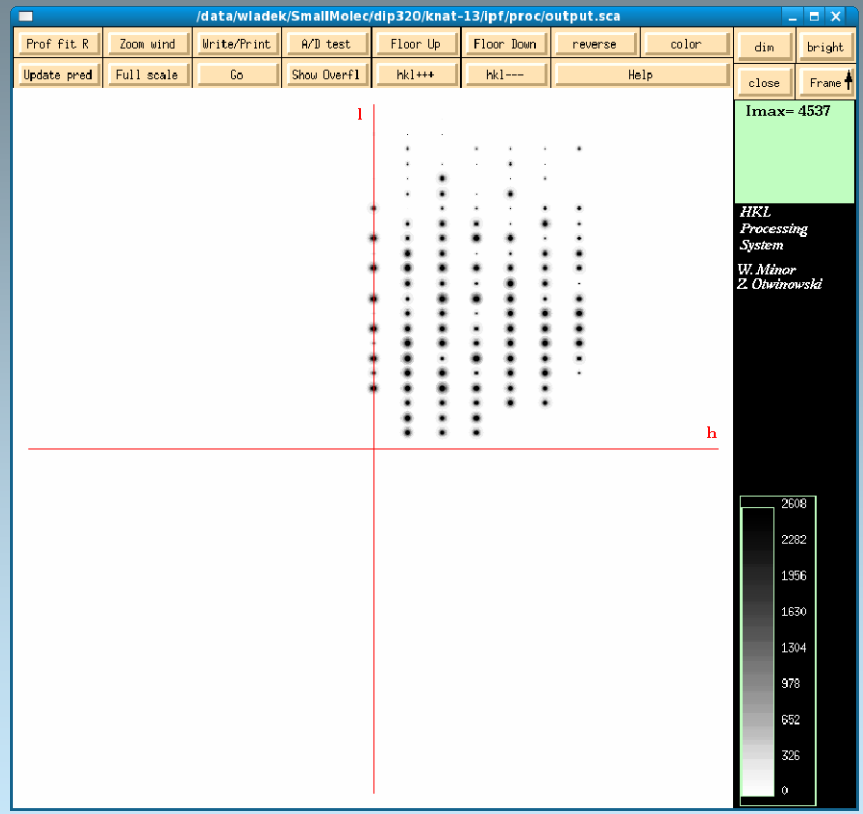
Reliable Space Group Determination



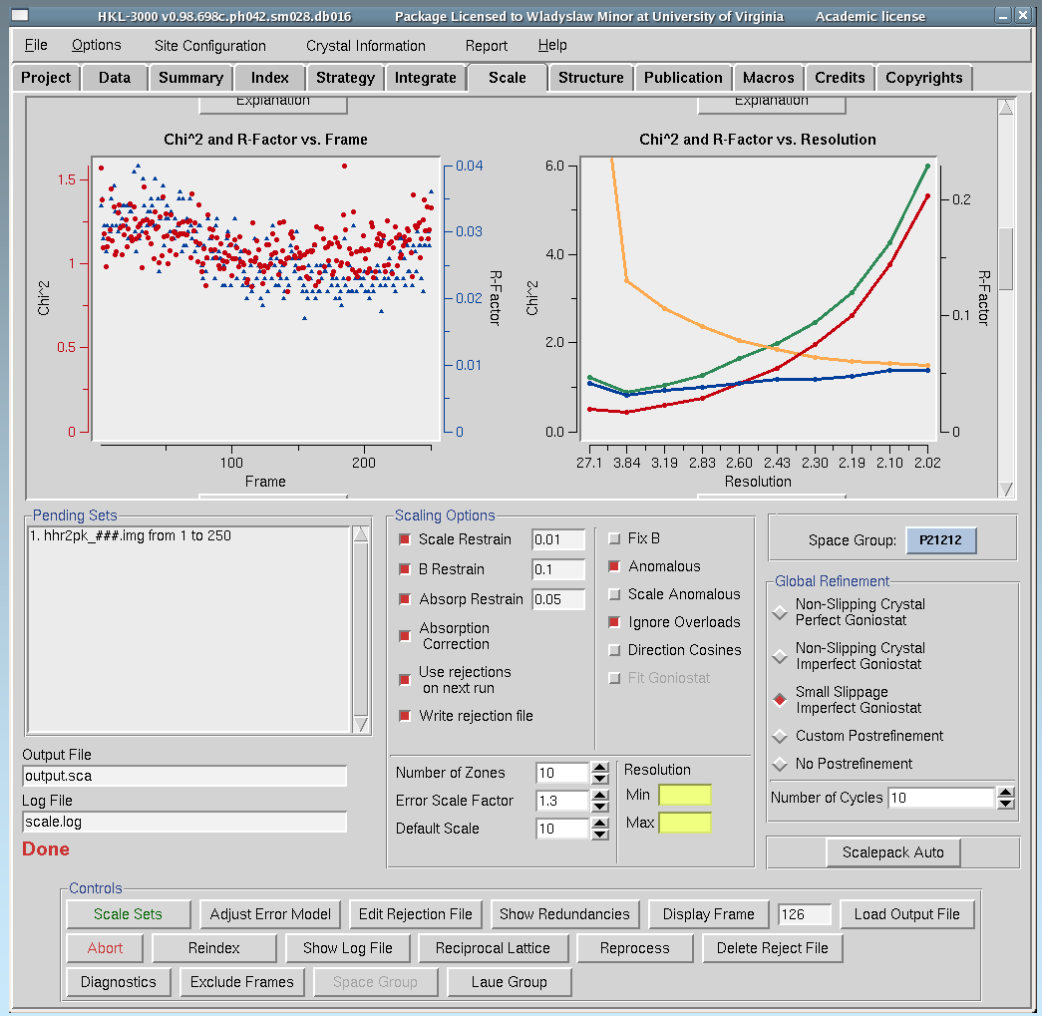
P2?2?2?

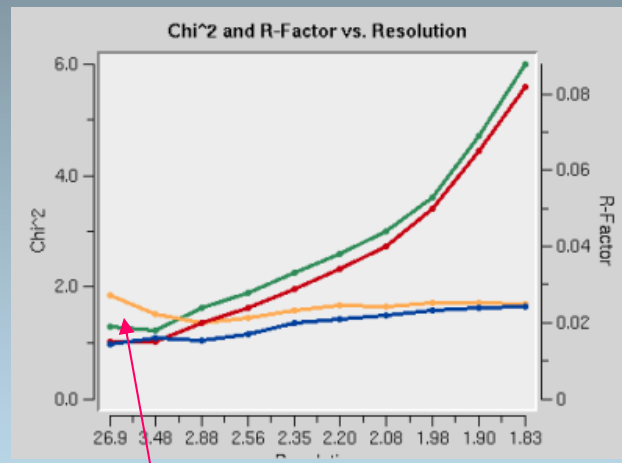

/data/wladek/SmallMolec/dip320/knat-13/ipf/proc/output.sca
Prof fit R
Zoom wind
Write/Print
A/D test
Floor Up
Floor Down
reverse
color
dim
bright
Update pred
Full scale
Go
Show Overfl
hkl+++
hkl---
Help
close
Frame
Imax= 4537
d=0.91A
I=0
[1.5 , 2.0]
HKL
Processing
System
W. Minor
Z. Otwinowski
l
k
1896
1659
1422
1185
948
711
474
237
0

P2?2?2?


/data/wladek/SmallMolec/dip320/knat-13/ipf/proc/output.sca
Prof fit R
Zoom wind
Write/Print
A/D test
Floor Up
Floor Down
reverse
color
dim
bright
Update pred
Full scale
Go
Show Overfl
hkl+++
hkl---
Help
close
Frame
Imax= 4537
HKL
Processing
System
W. Minor
Z. Otwinowski
l
h
2608
2282
1956
1630
1304
978
652
326
0

Scaling

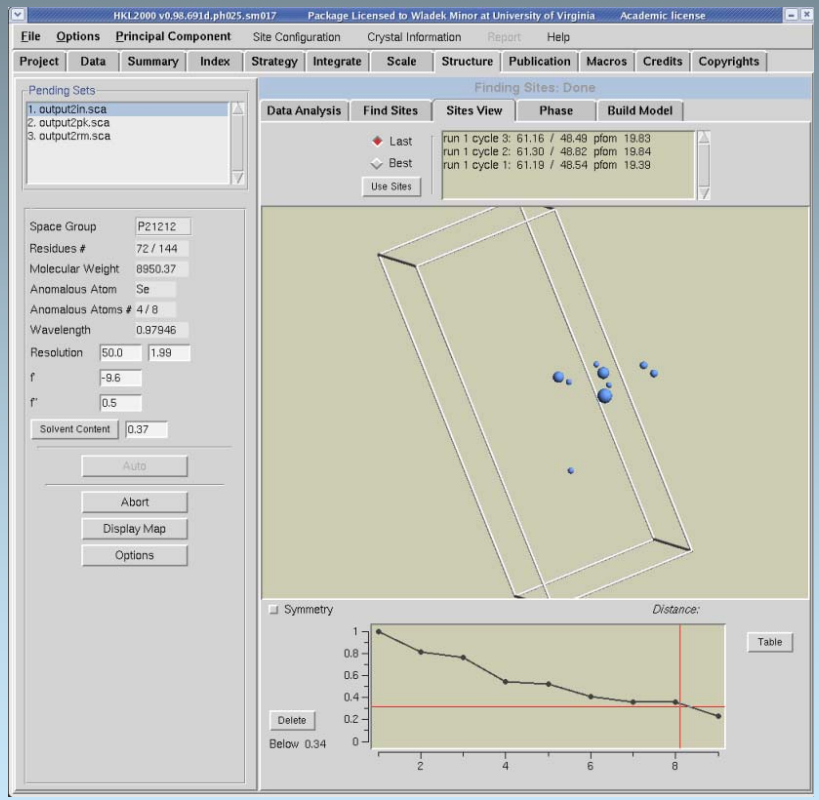


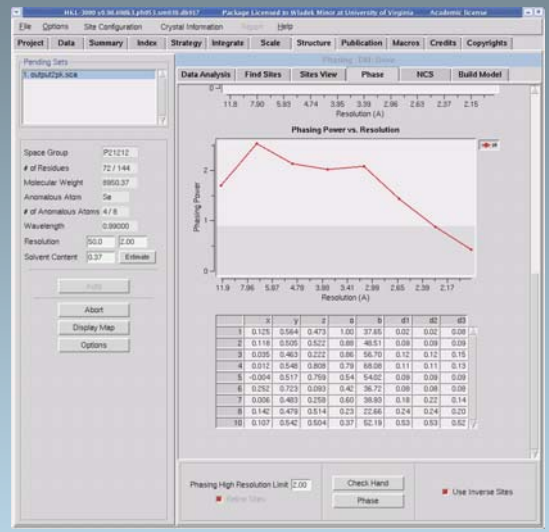


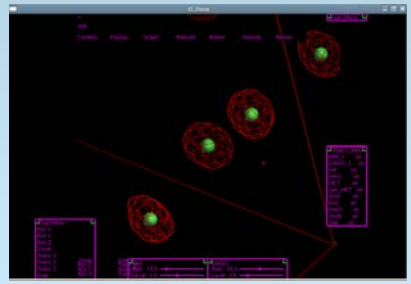
Signal from Sulfur

Substructure solution

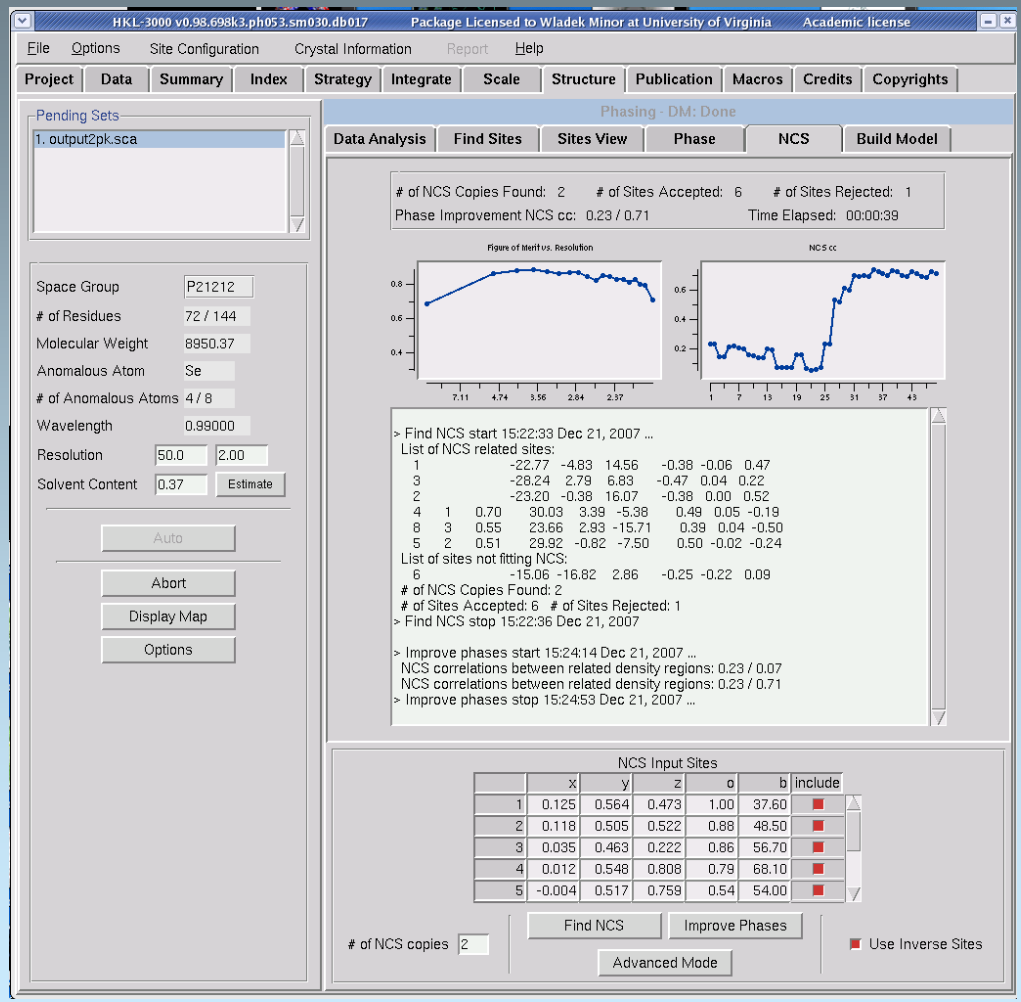
- Data management in HKL-3000 - substructure solution



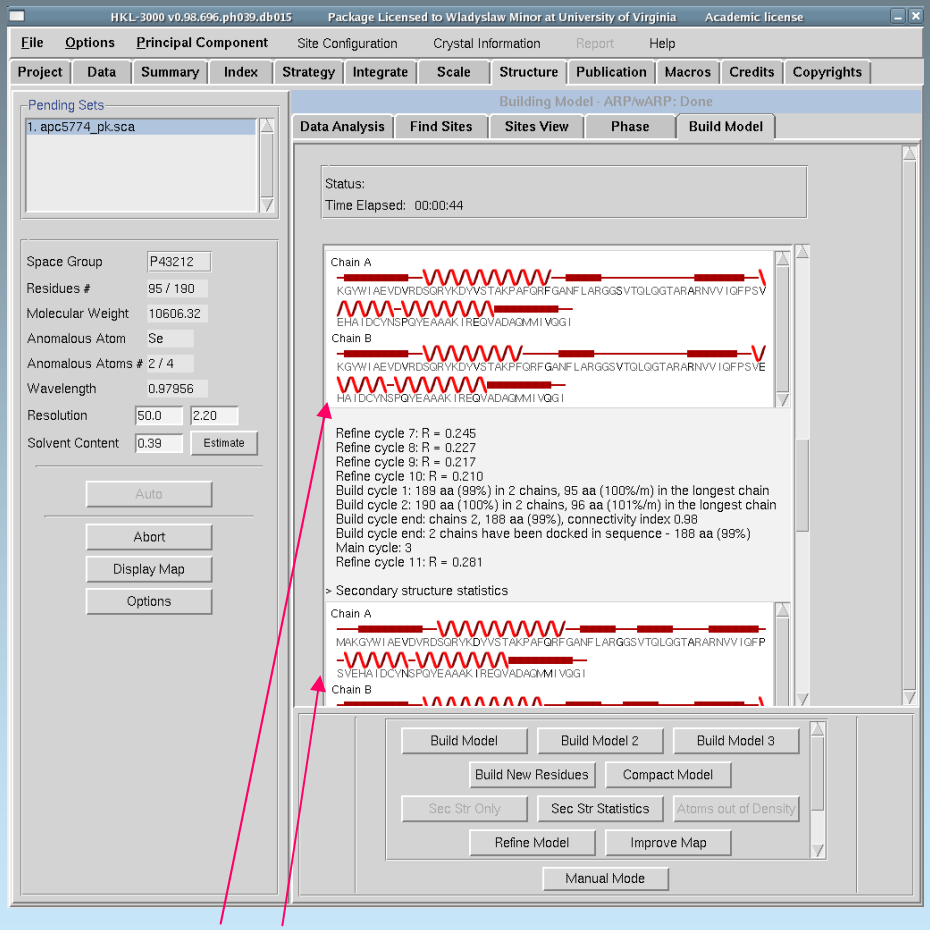


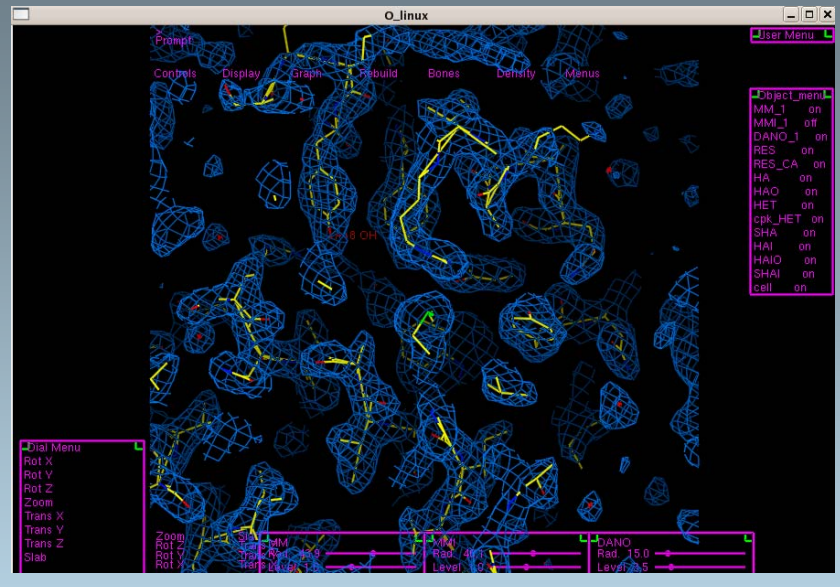


NCS handling



Model building and preliminary refinement (10 minutes, 190aa)

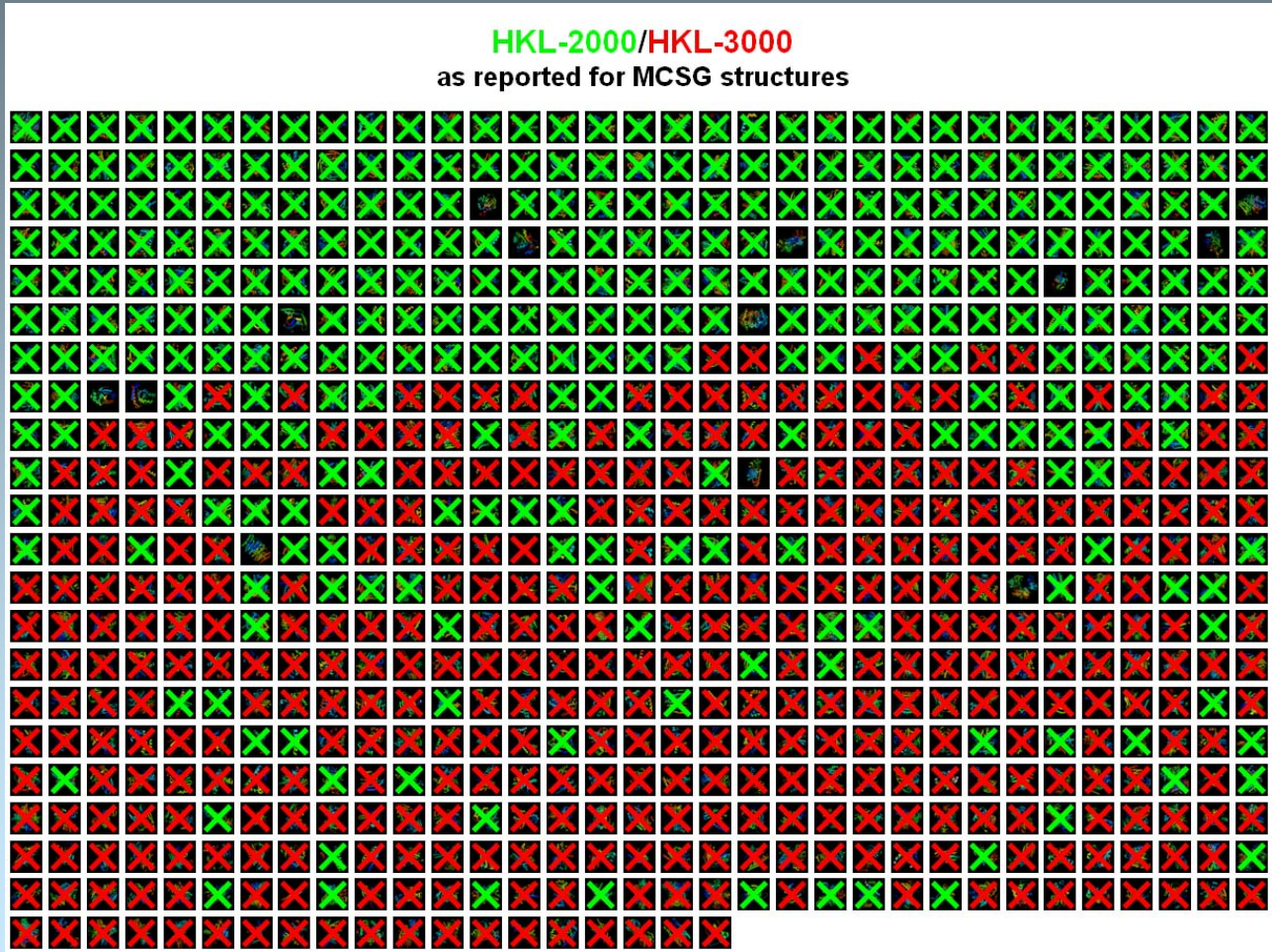




Unidentified density -> Xtaldb

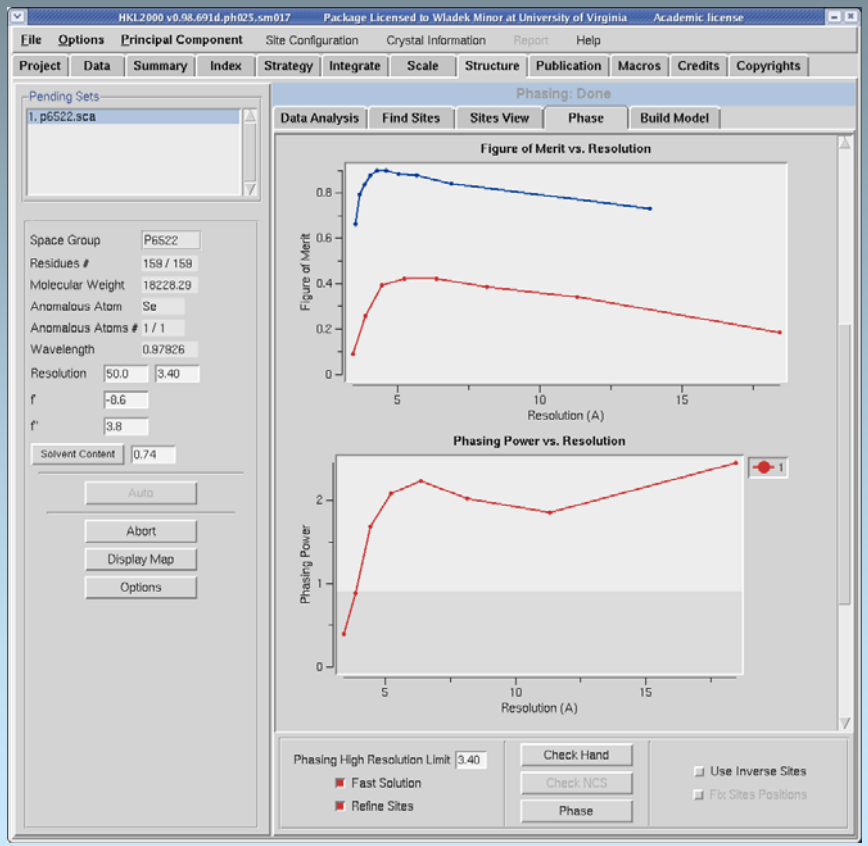
Secondary structure after each cycle

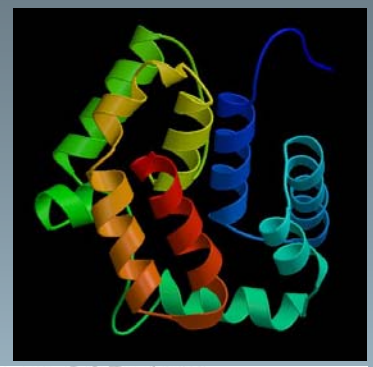
MCSG 712 [HKL-3000]

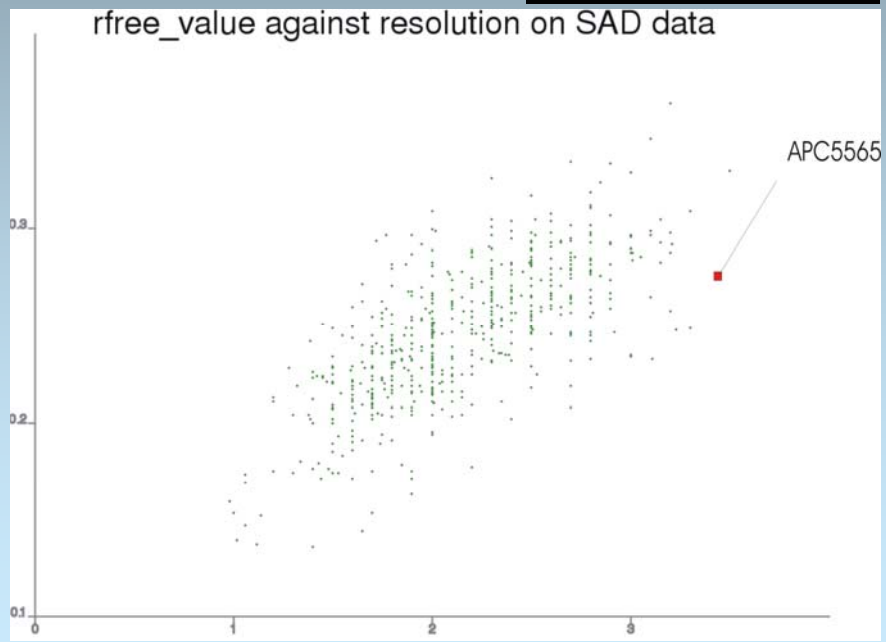


Low resolution SAD phasing difficult structure – no

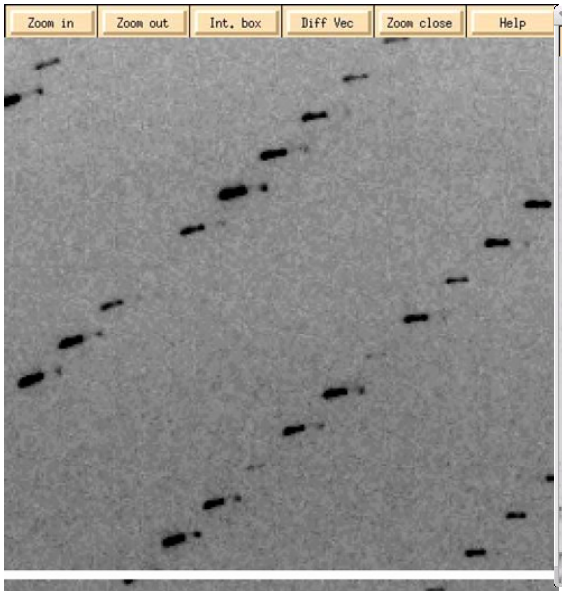
1ze0

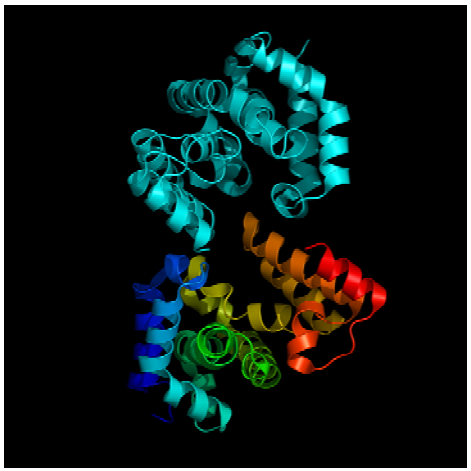




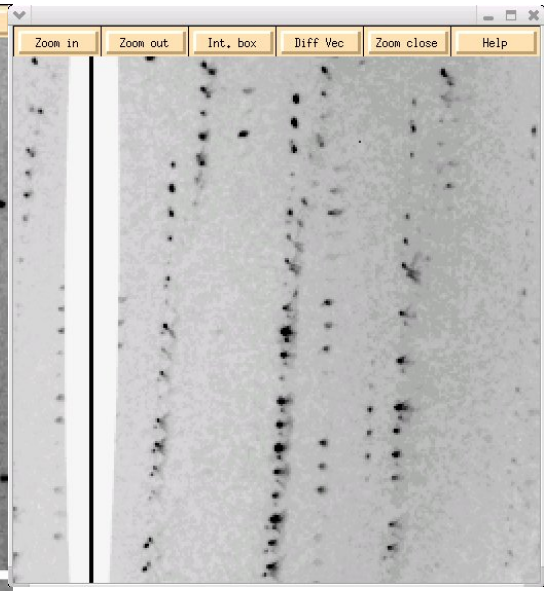


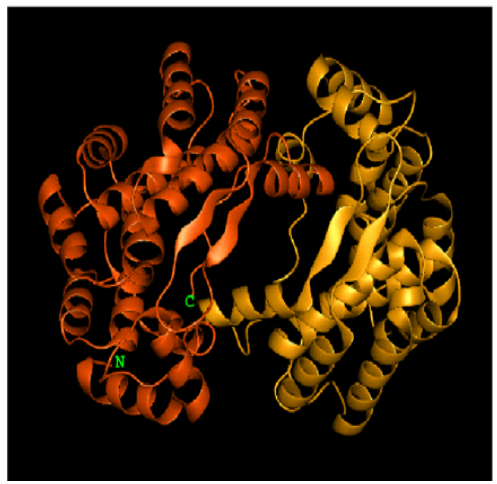
APC5905



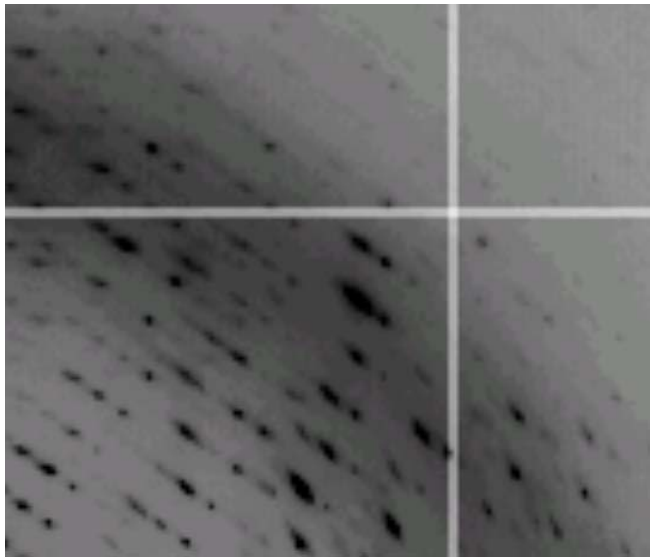


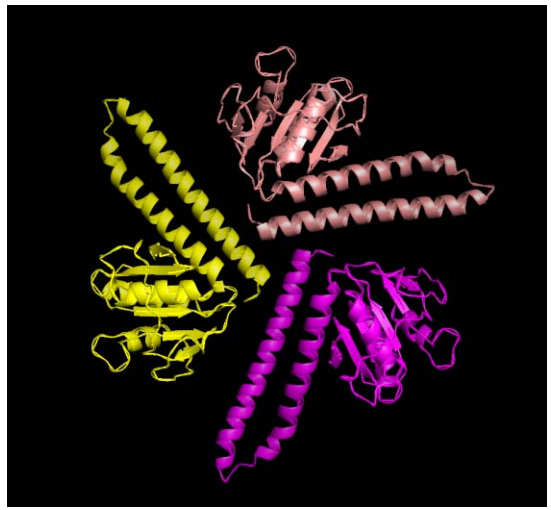
APC24929





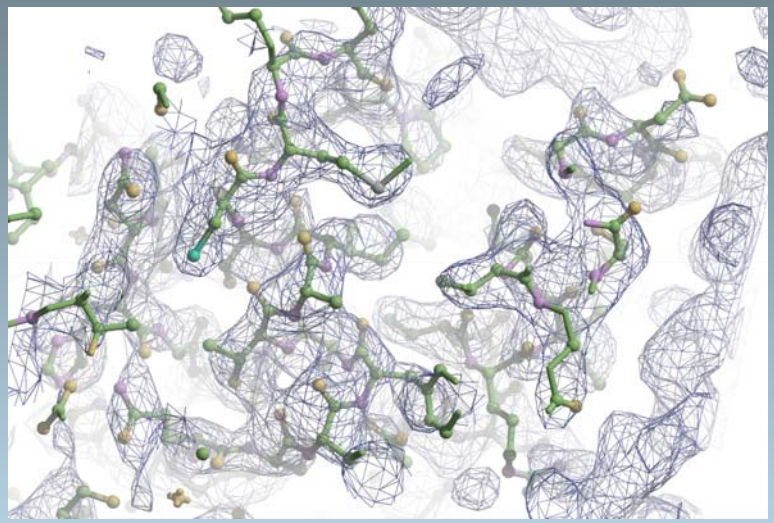
APC25179



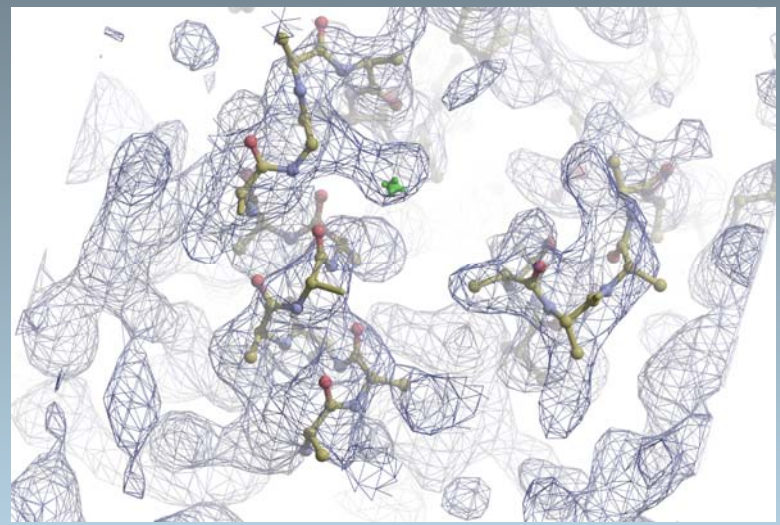


Structures solved in wrong target

APC81521 (ChangSoo)



Final model with right target
APC81521



Auto-build (resolve) model with wrong target
APC81501

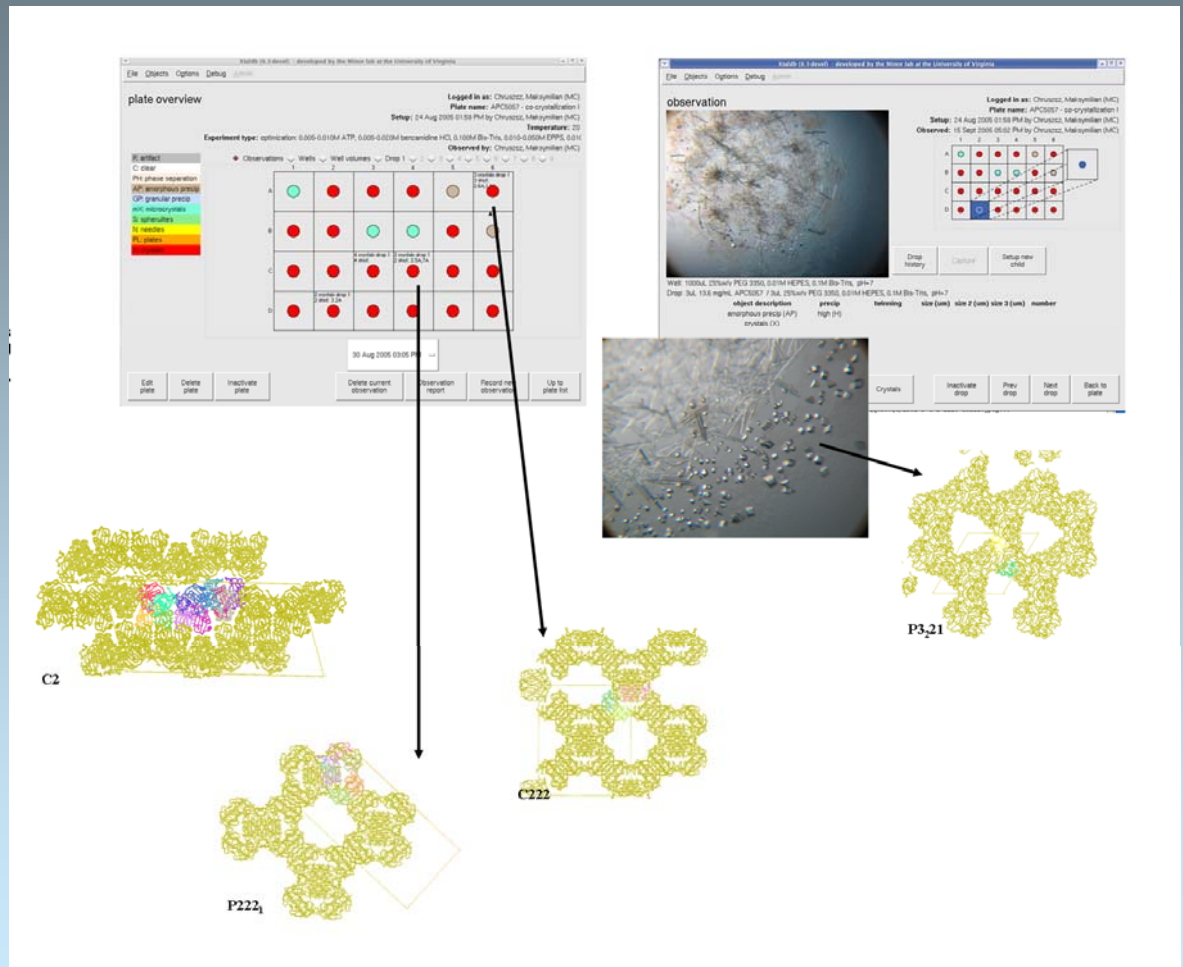
To figure out the right target at the resolution of 2.5Å

1. Manually build helixes and stands;
2. Submit this model to dali to find out several similar structures in PDB;
3. Search targetDB(MCSG) using the sequence of these structures to find out the right target.

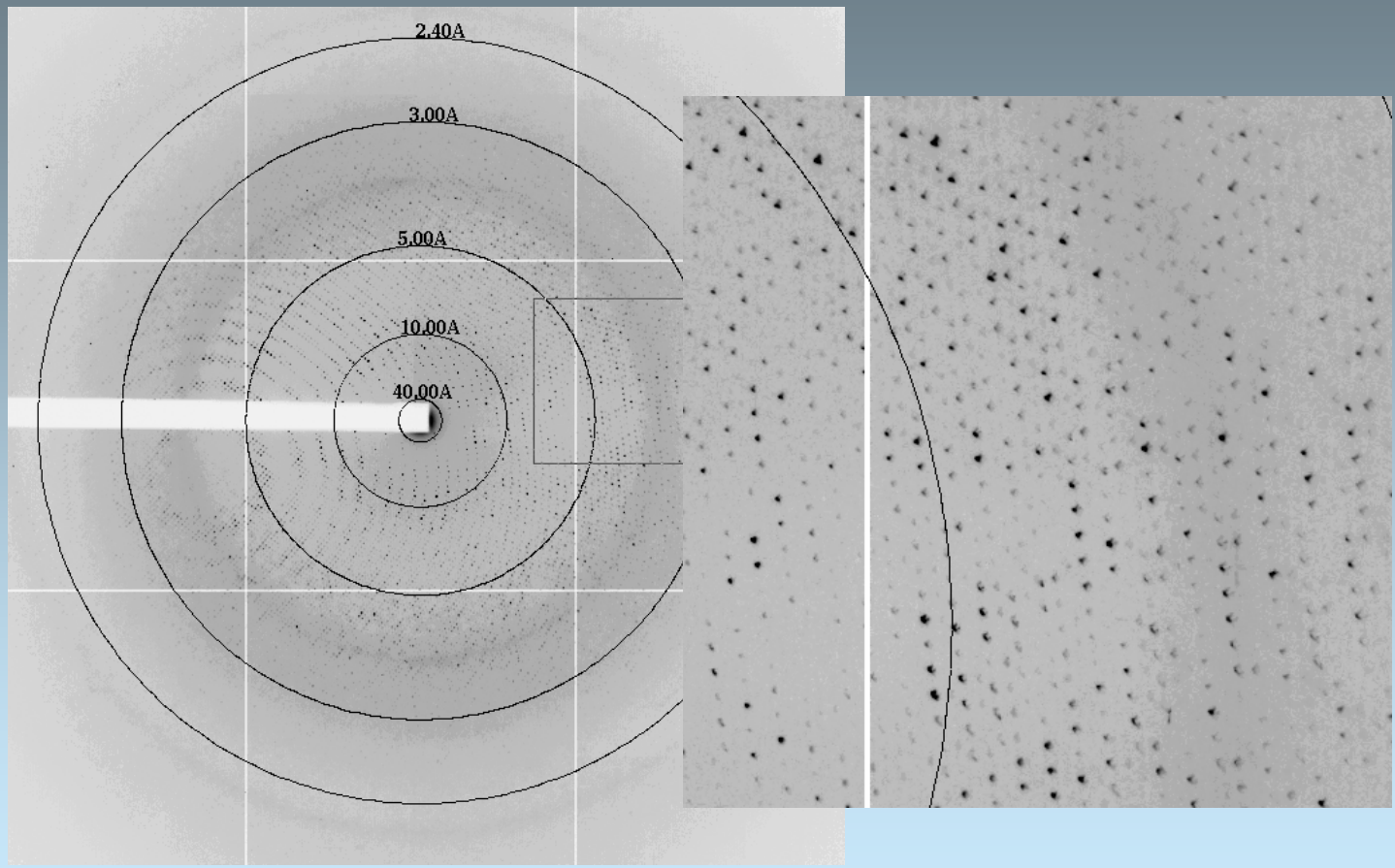
Twinning

Argonne Users Meeting, May 2008

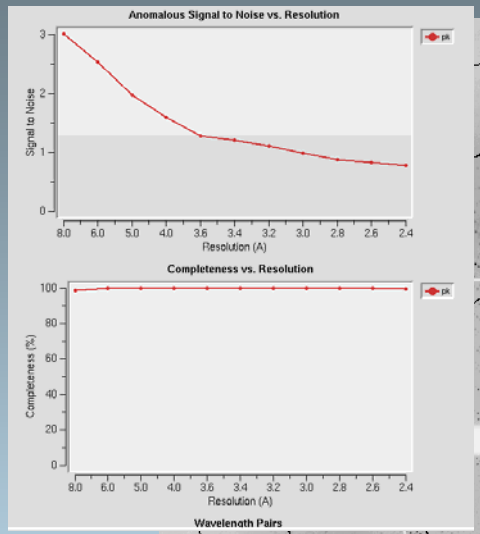
Xtaldb – crystallization database

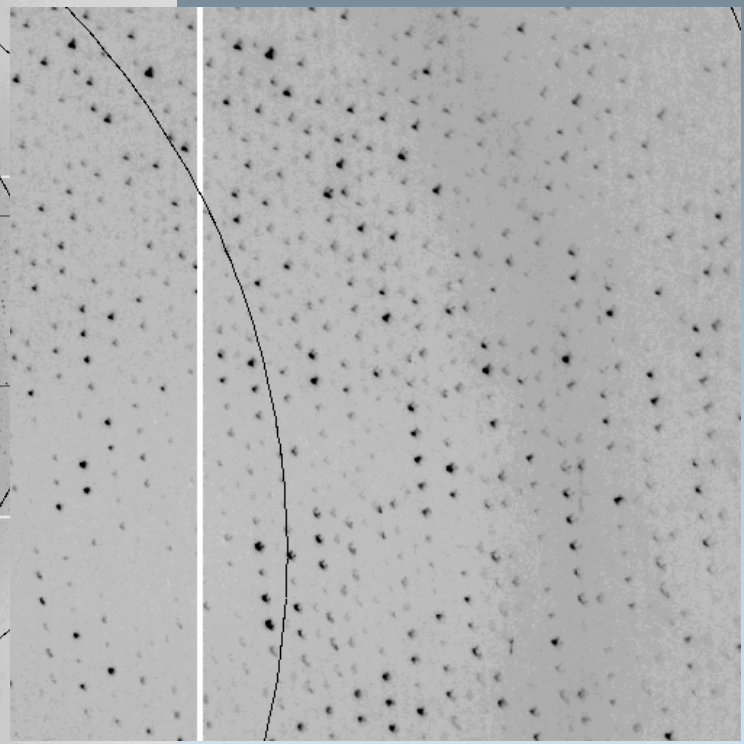


TM1086 (APC4579)

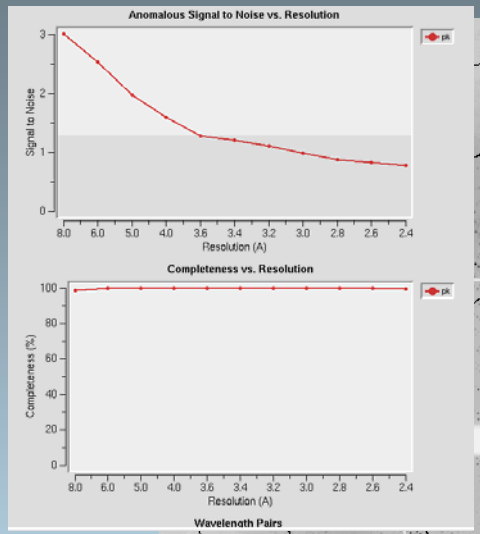


TM1086 (APC4579)



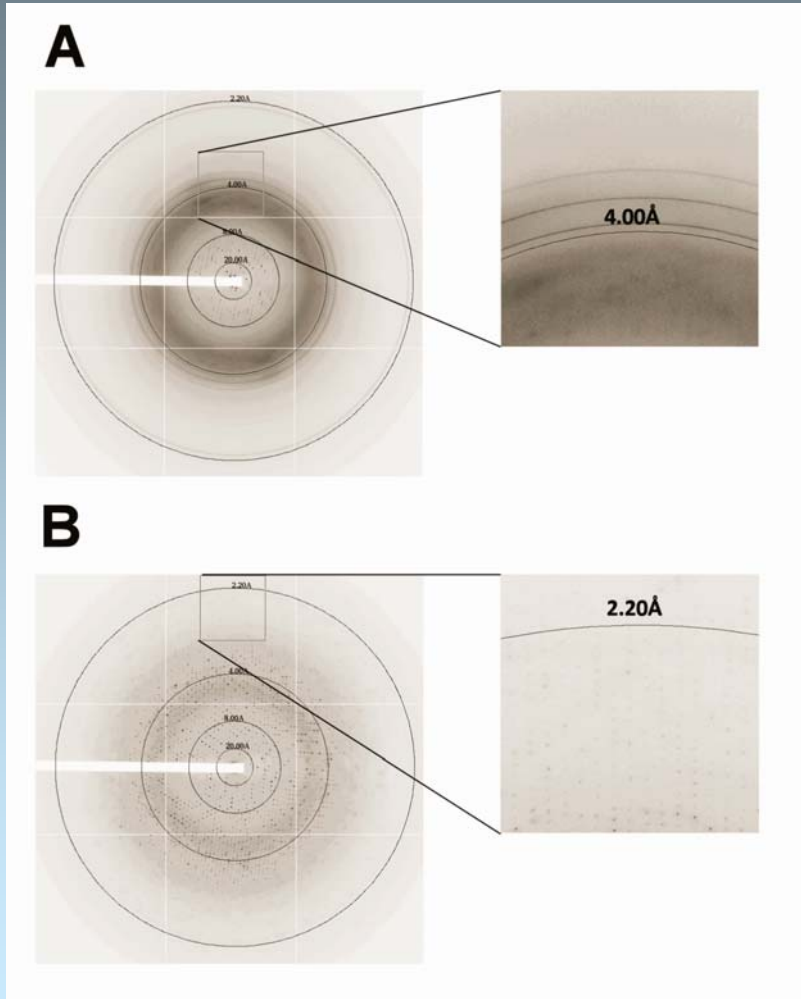


TM1086 (APC4579)



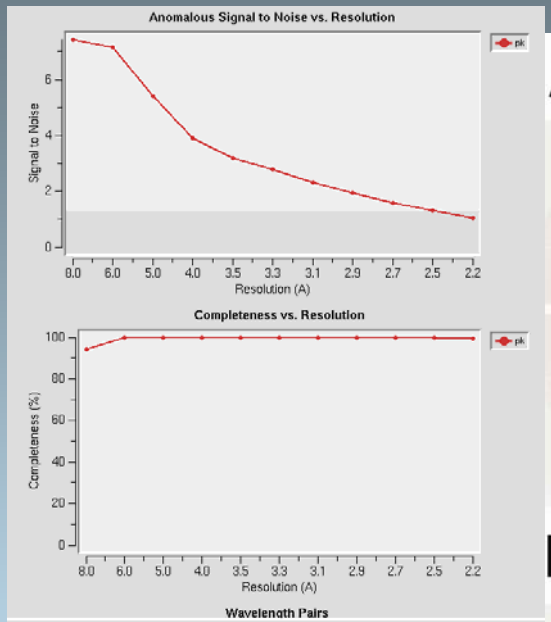
10 Se-Met/282 aa
35.1kDa
I222
a= 290Å, b= 300Å, c=316Å
~81% solvent content = 17 molecules in AU
~38% solvent content = 57 molecules in AU

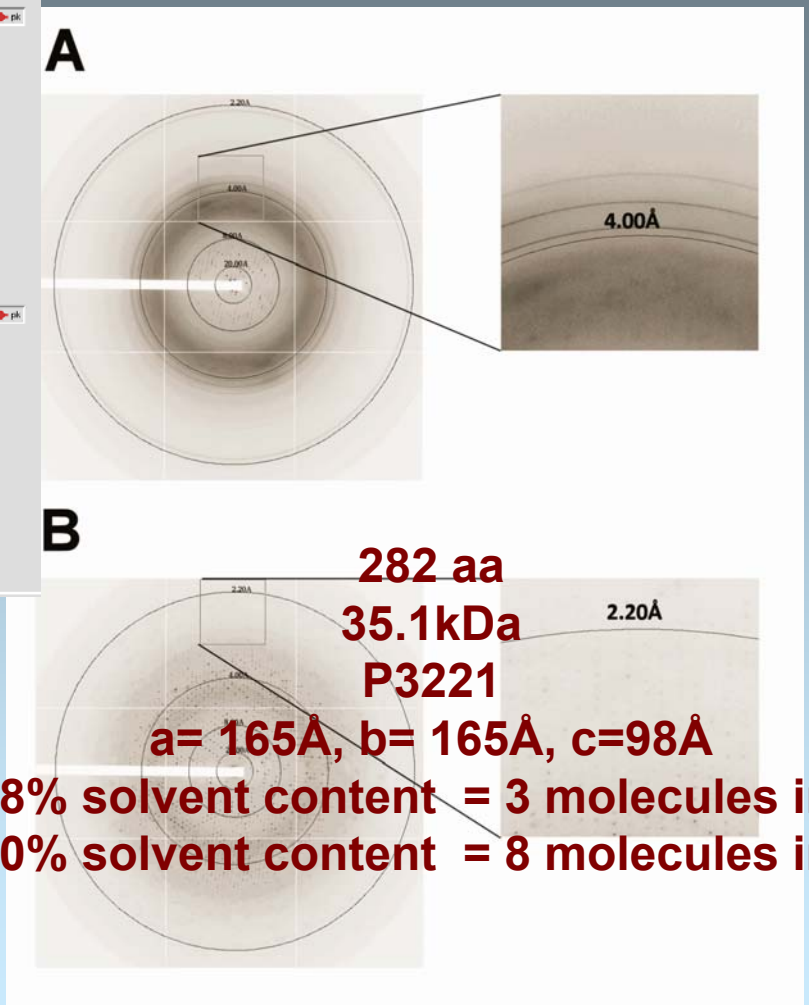
TM1086 (APC4579)



Argonne Users Meeting, May 2008

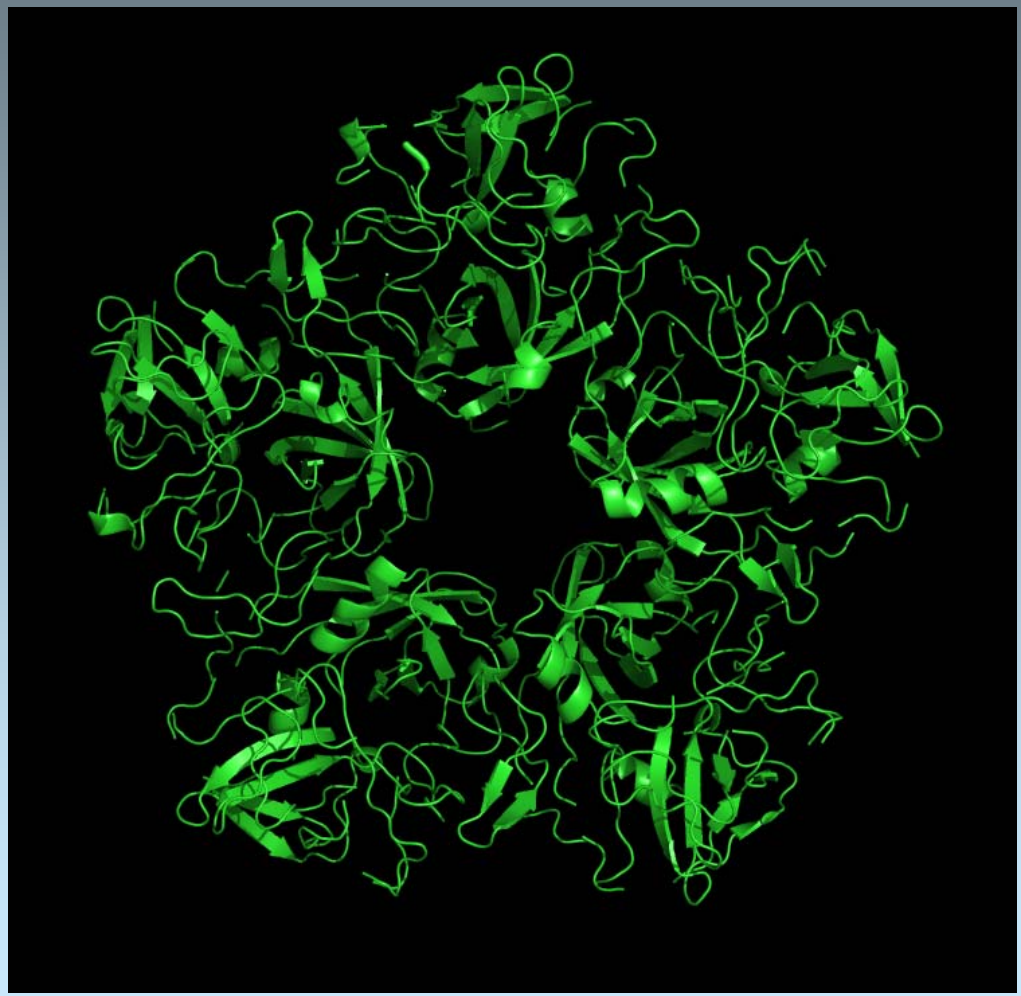
TM1086 (APC4579)





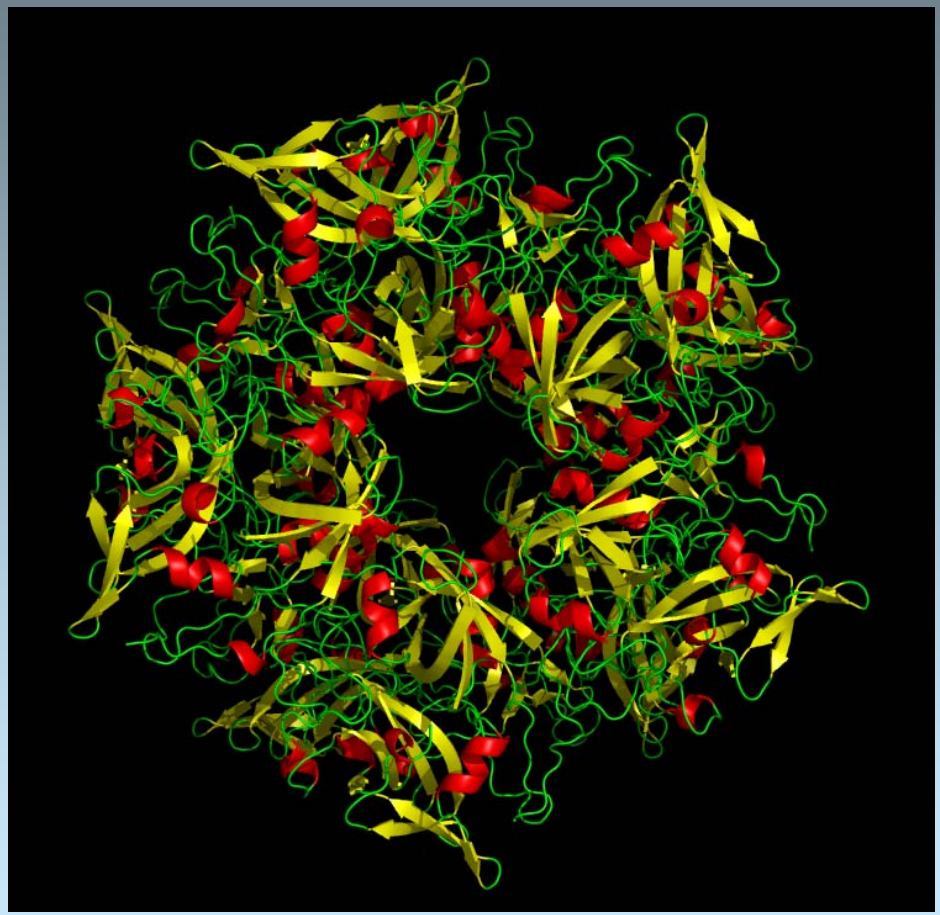
282 aa
35.1kDa
P3221
a= 165Å, b= 165Å, c=98Å
~78% solvent content = 3 molecules in AU
~40% solvent content = 8 molecules in AU

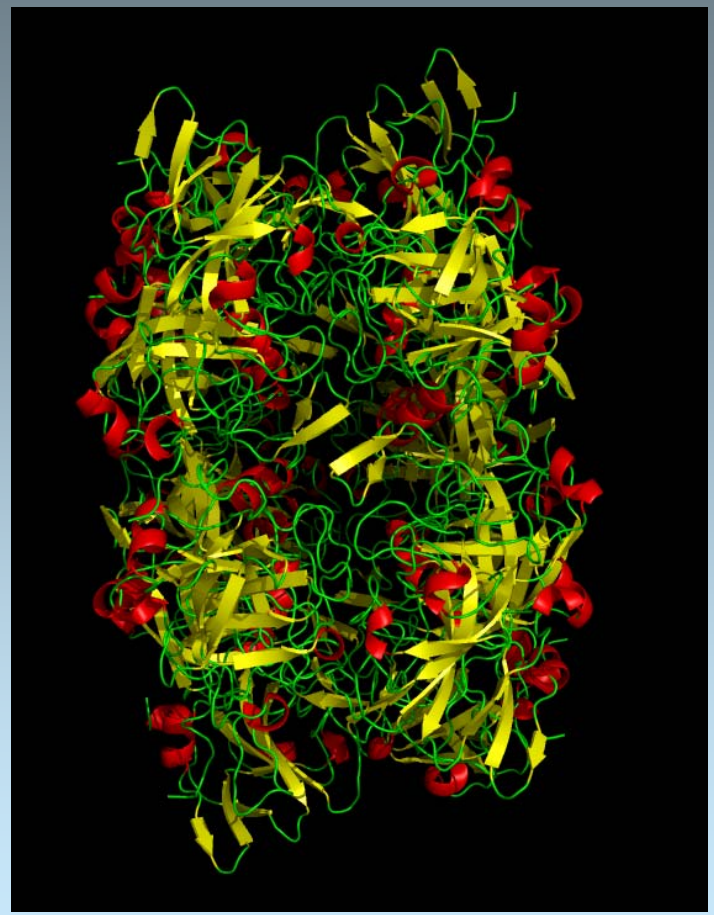
TM1086 (APC4579)



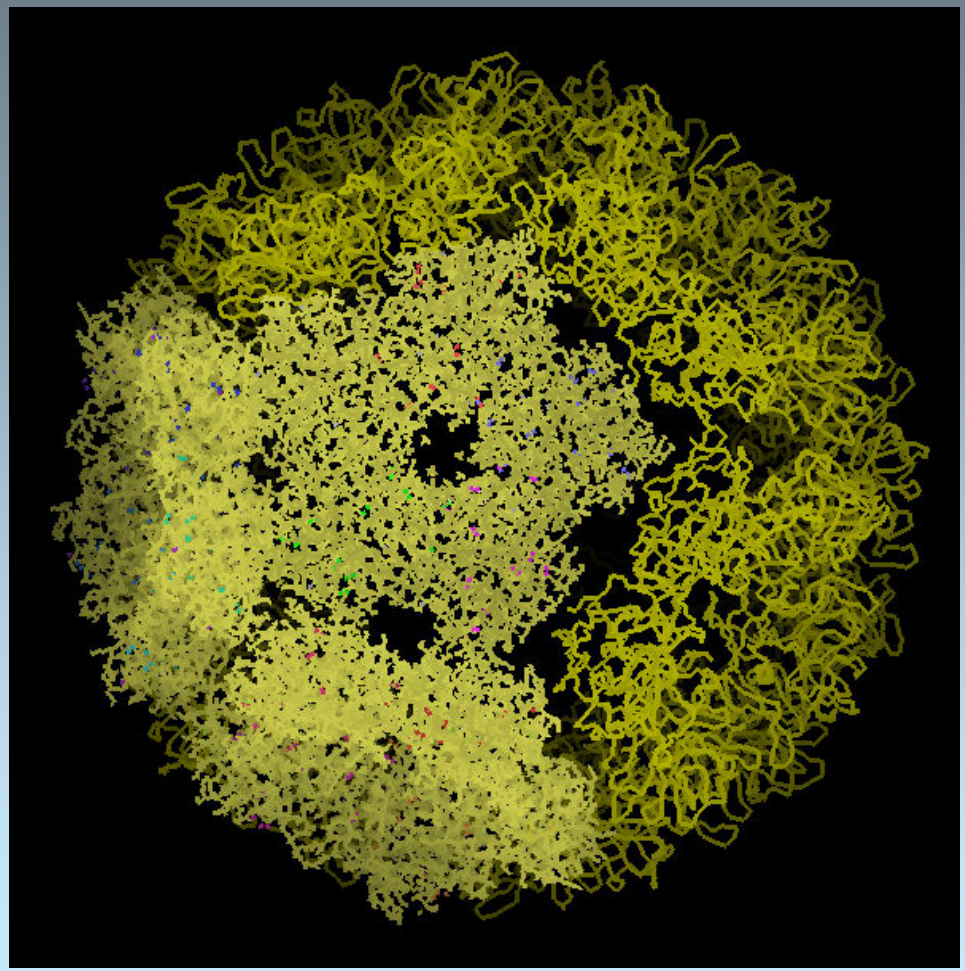
Argonne Users Meeting, May 2008

TM1086 (APC4579)



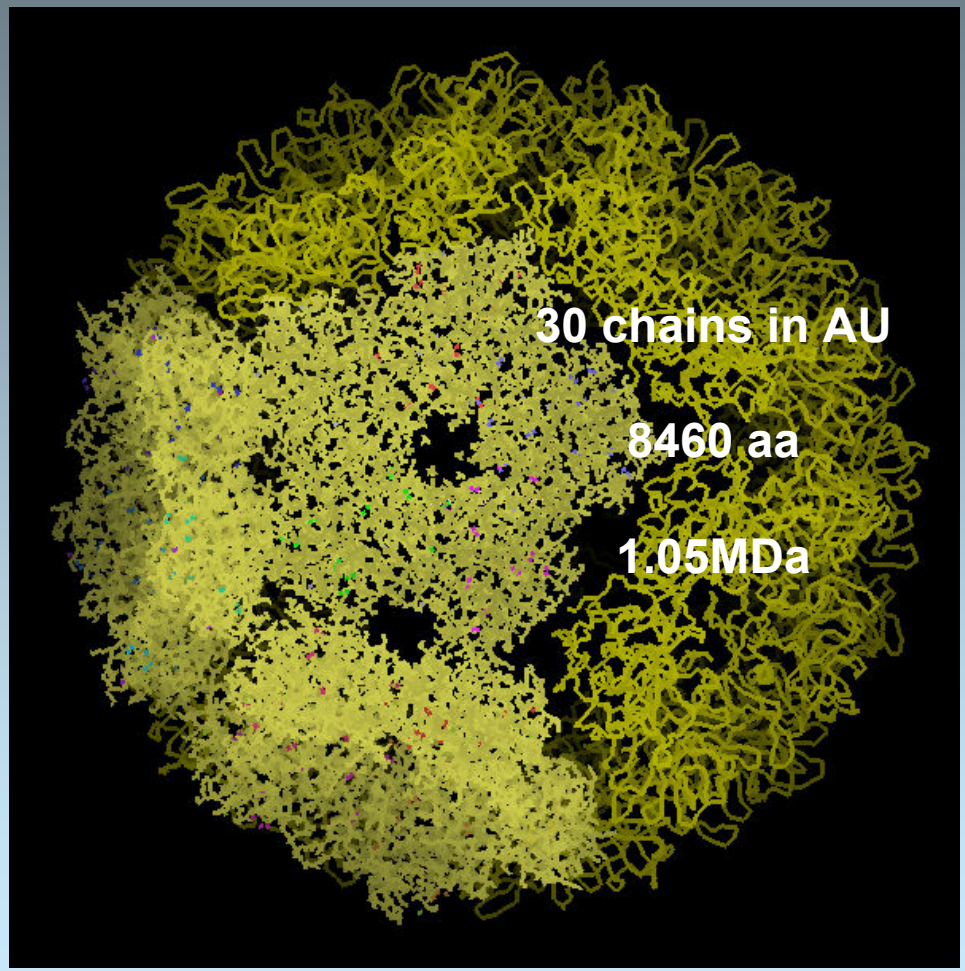


I222 crystal form



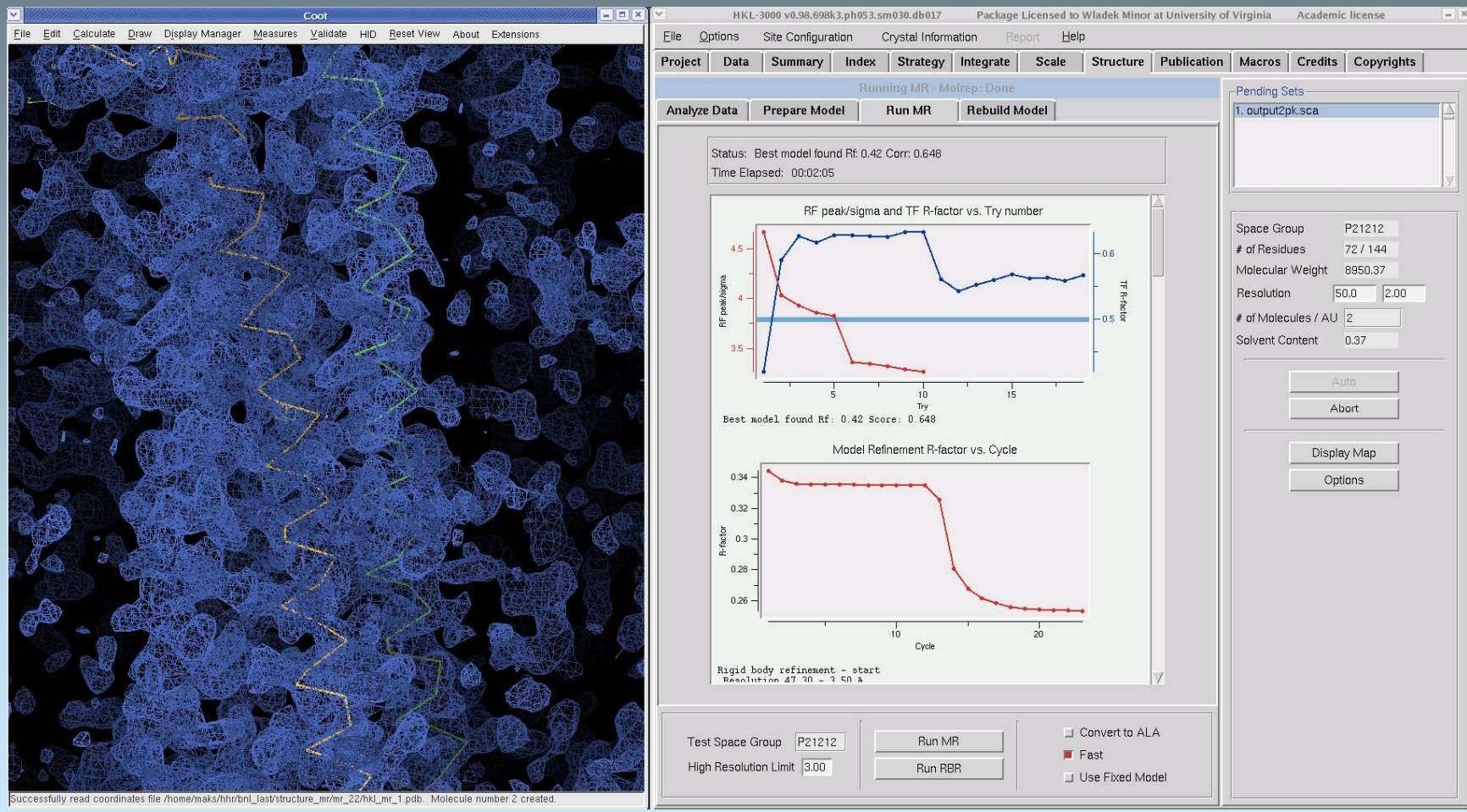
~210 Å

I222 crystal form



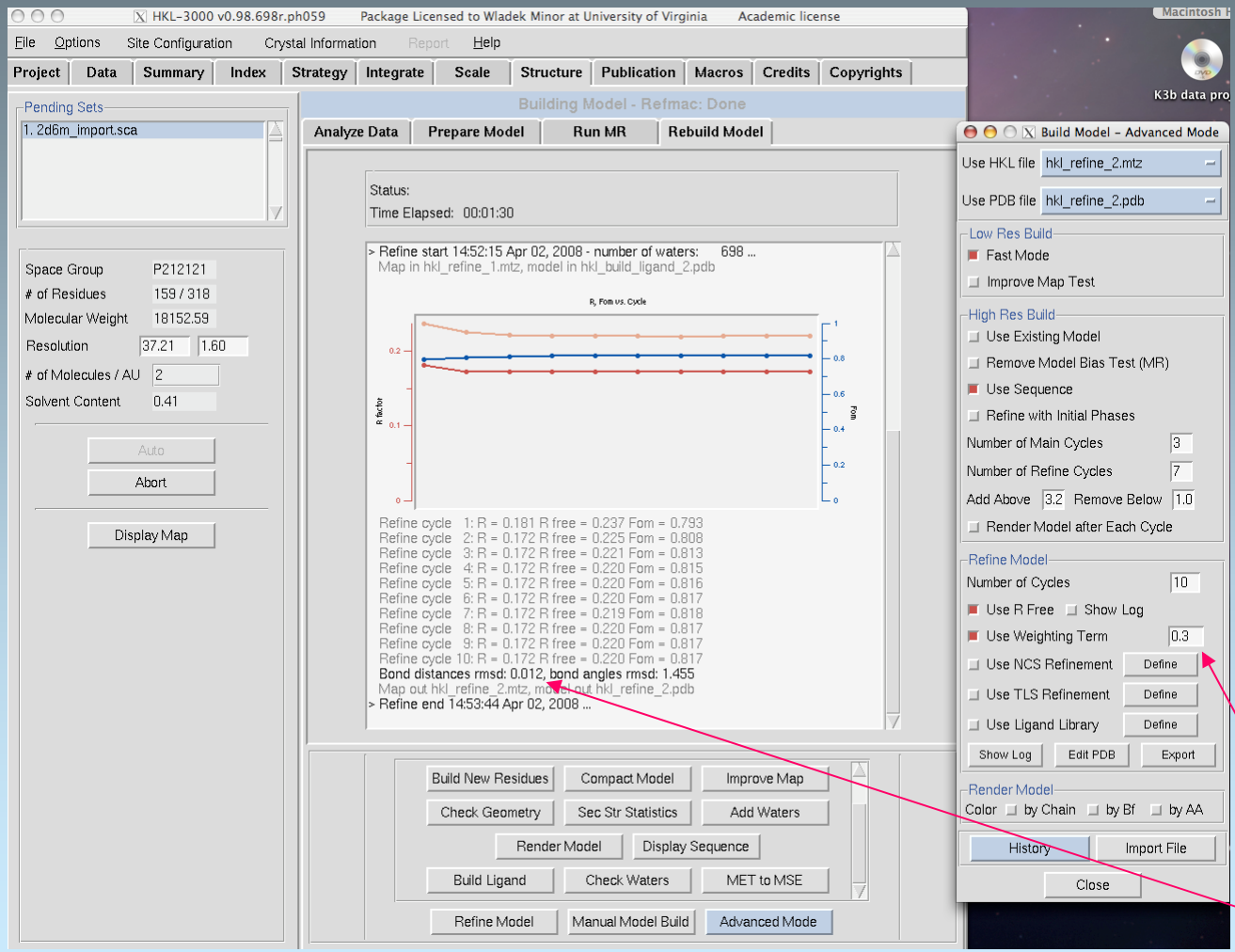
~210 Å

Molecular Replacement



Argonne Users Meeting, May 2008

Fast refinement (full control)



Weighting scheme

Unidentified densities ?

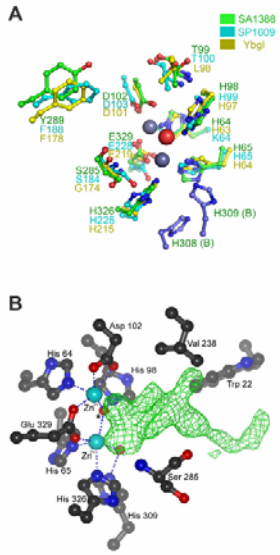


Figure 5
The putative active site of SA1388. (**A**) Conserved residues at the putative active sites of SA1388 (green), SP1609 (cyan) and *E. coli* ybgI (yellow). The two zinc ions of SA1388 and the bridging water molecule are shown as spheres and the equivalent residues in the three structures are labelled in their respective colors. The two histidine residues of SA1388 that are contributed by the adjacent subunit are colored in navy blue. (**B**) Electron density of the endogenous ligand (Fo-Fc map contoured at 3.0 σ) that is directly ligated to the active site zinc ions in SA1388. Residues interacting with the two zinc atoms are shown and the metal coordinating interactions are indicated by dotted lines. The bridging water molecule between the two Zn atoms is indicated by an asterisk (*).

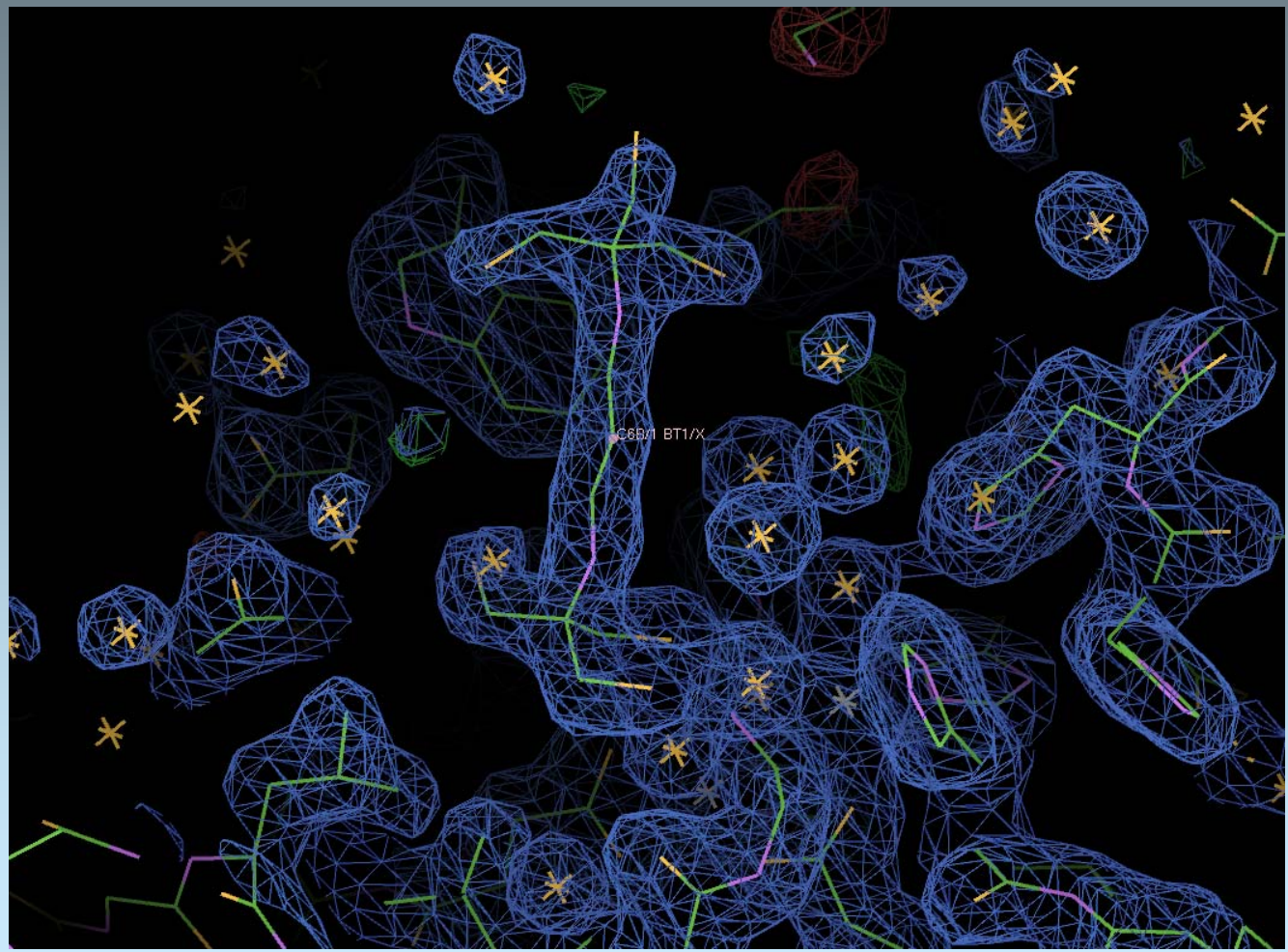
cryoprotection and therefore conclude that the observed ligand must in fact be an endogenously bound ligand that was co-purified after recombinant overexpression. The contours of the difference electron density map (Figure 5B) after final refinement suggests the presence of a head group directly ligated to the metal ions and a mostly aliphatic tail that has few specific polar interactions with the surrounding protein residues (Figure 5B). Two aromatic residues, a tyrosine (Y289, shown in Figure 5A) and a tryptophan (W22) lie in close proximity to the bound ligand and may be involved in ligand binding.

Discussion

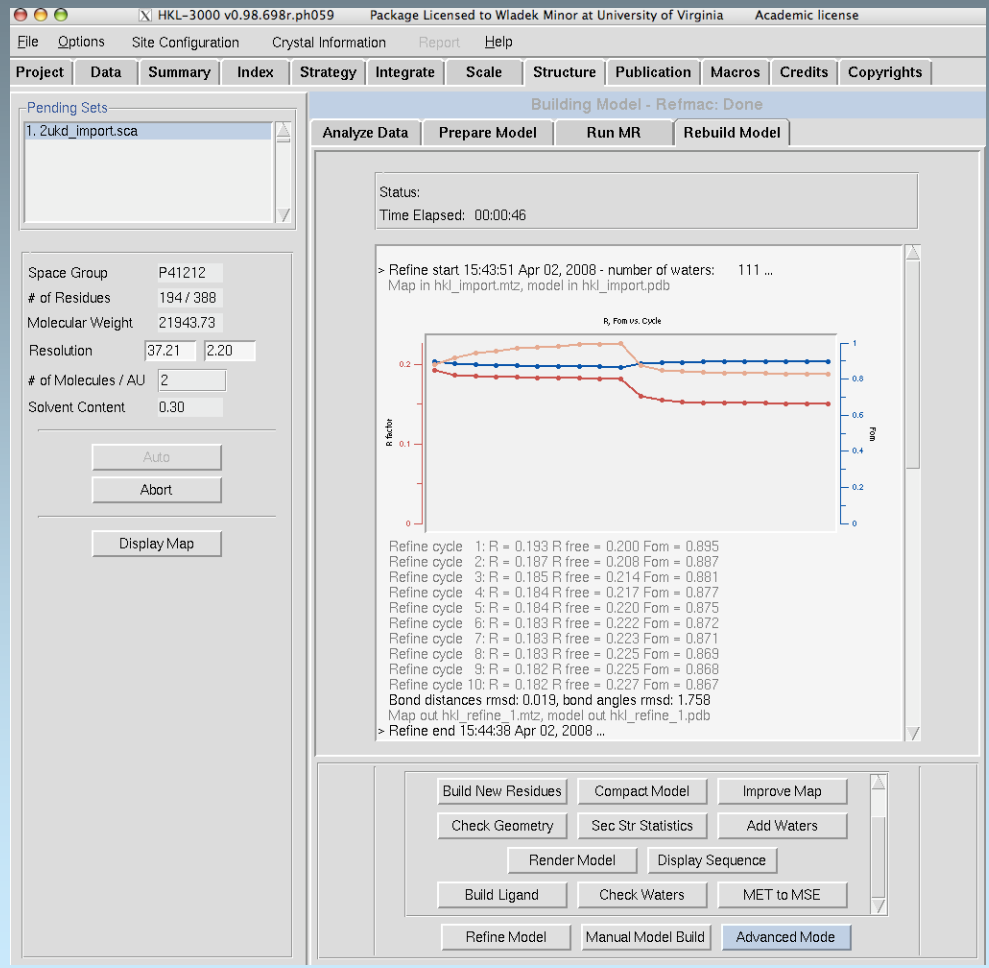
Although homologs of both PII and NIF3 proteins are found ubiquitously in all three kingdoms of life, the function of SA1388 and its homologs that contain both PII-like and NIF3-like domains remain completely unknown. The fusion of the two proteins in one peptide chain indicates a functional coupling of these two proteins. The structural features revealed in the present study, such as the cage-like hexameric toroid structure with its NIF3 domains as walls and the two PII-like domain trimers as lids, the dinuclear metal site, and the intrinsically bound ligand, may provide certain clues of its potential function. The primary role of the nitrogen regulatory PII proteins is to integrate various intracellular carbon and nitrogen signals by regulating enzymes involved in nitrogen assimilation [10]. PII exerts its regulatory effects by undergoing different post translational modifications, such as uridylylation [24] and phosphorylation [25] by various modifying enzymes in response to the primary cellular nitrogen signal glutamine. Structures of several PII and PII-like proteins have been solved including glnB (from *E. coli, T. thermophilus, Synecococcus* and *Herbaspirillum*), GlnA (*E. coli*), glnK (*E. coli*), HisG (*E. coli* and *M. tuberculosis*), CutA1 (human, rat and *T. maritima*), and a hypothetical protein from COG1993. Effectors of PII include transcription factors [10,11], signalling proteins e.g., histidine kinases [26], and metabolic enzymes like glutamine synthase [27]. Apart from its primary signalling modulator glutamine, PII proteins are known to bind a range of small molecule effectors such as ATP, UMP, and 2-ketoglutarate, which affects its function antagonistically to glutamine (reviewed in [10]). It is highly likely that the trimeric PII domain of SA1388 also plays a ligand induced signalling role and probably regulates the function of the NIF3-like domains.

Unlike PII domains, the NIF3-like domains have only recently begun to be structurally characterized, and their structure-function relationship remains sketchy. NIF3-like proteins are ubiquitously conserved from bacteria to higher eukaryotes [28]. They have been defined in uniprot database as the uncharacterized protein family UPF0135 that has 64 homologs [29]. Several homologs of this fam-

Unidentified densities

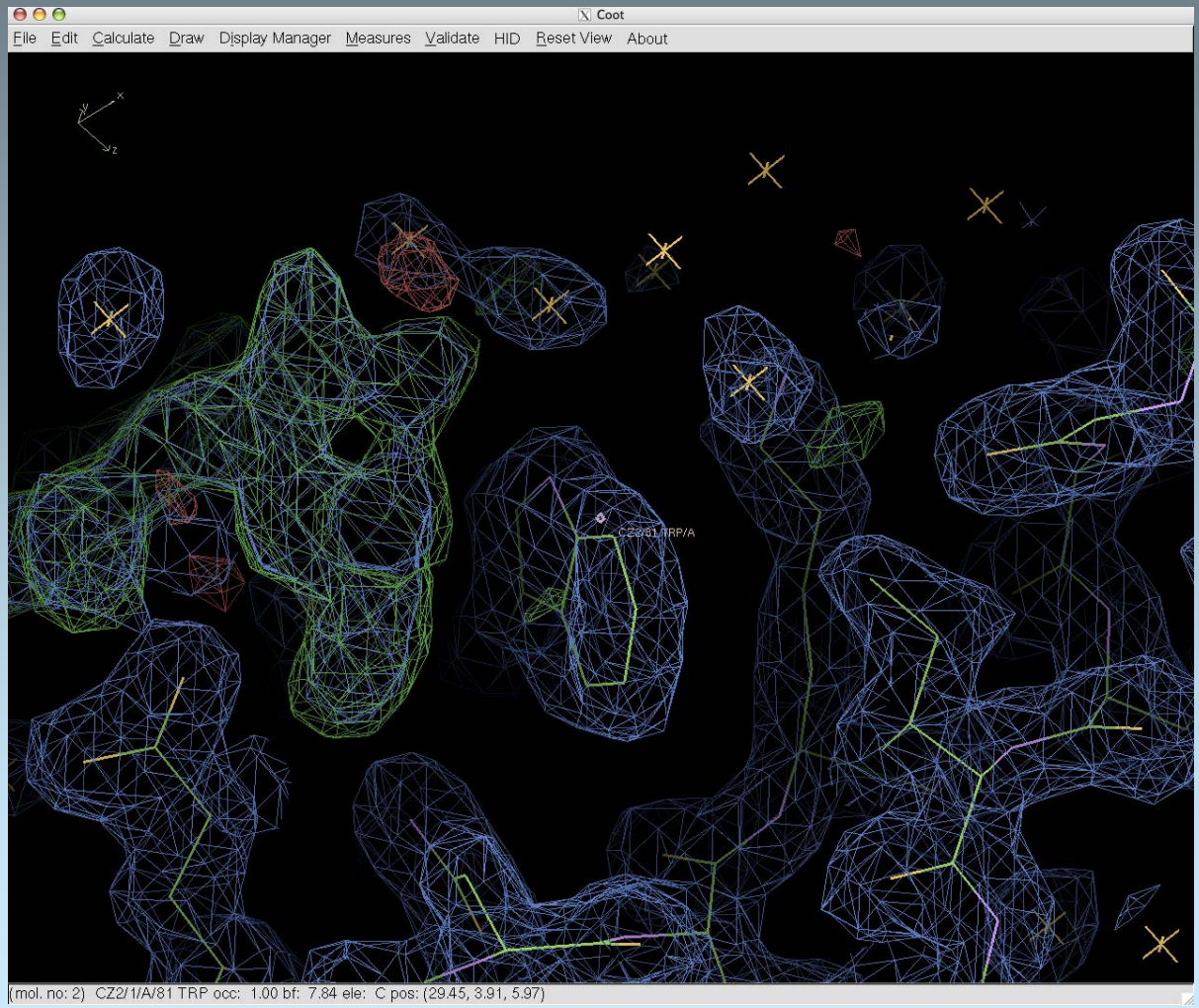


Fast refinement



Argonne Users Meeting, May 2008

Unidentified density

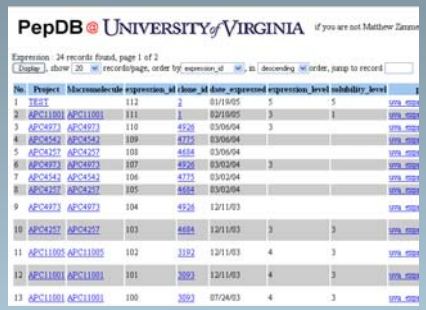


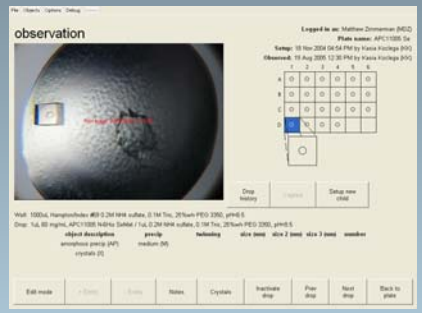
Integration
HKL-3000 is a pipeline component

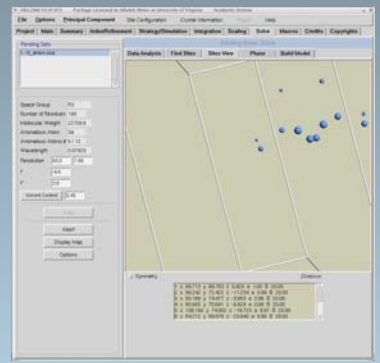
Pepdb
expression & purification

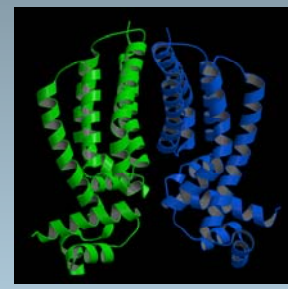
Xtaldb
crystallization

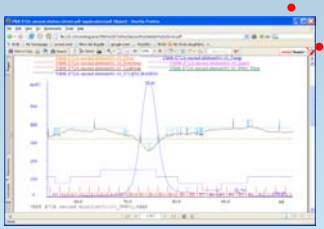
HKL-3000
Data collection,
diffraction & structure solution

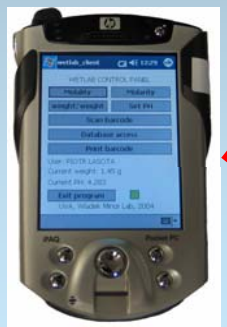






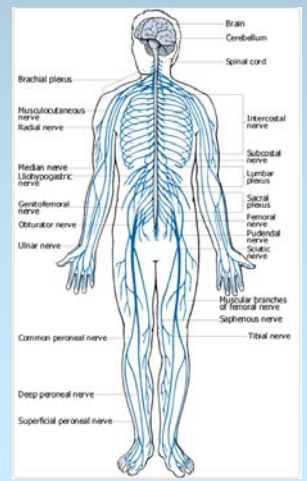




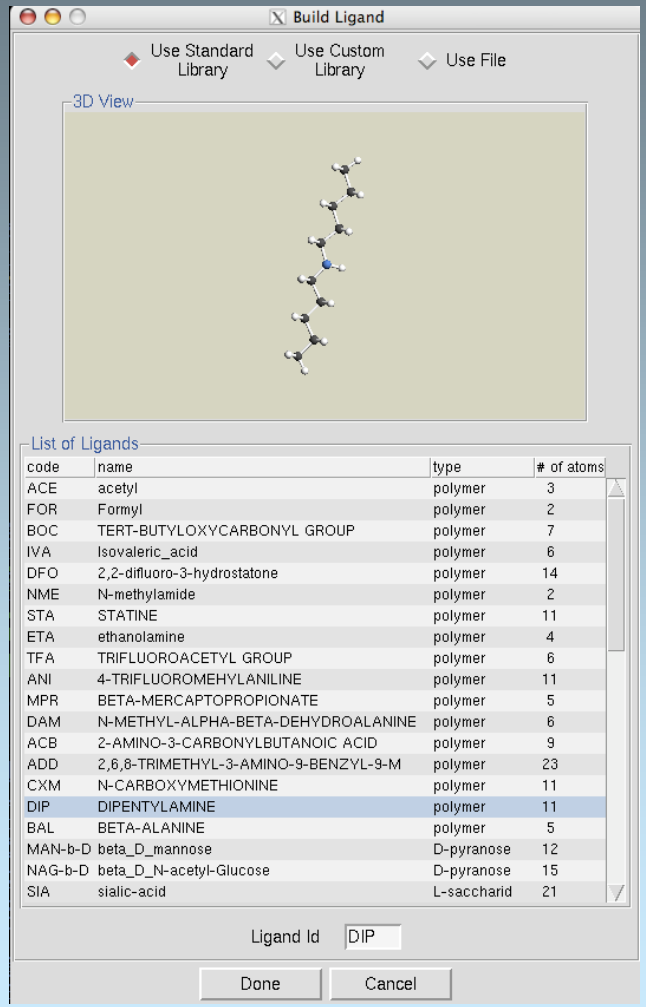


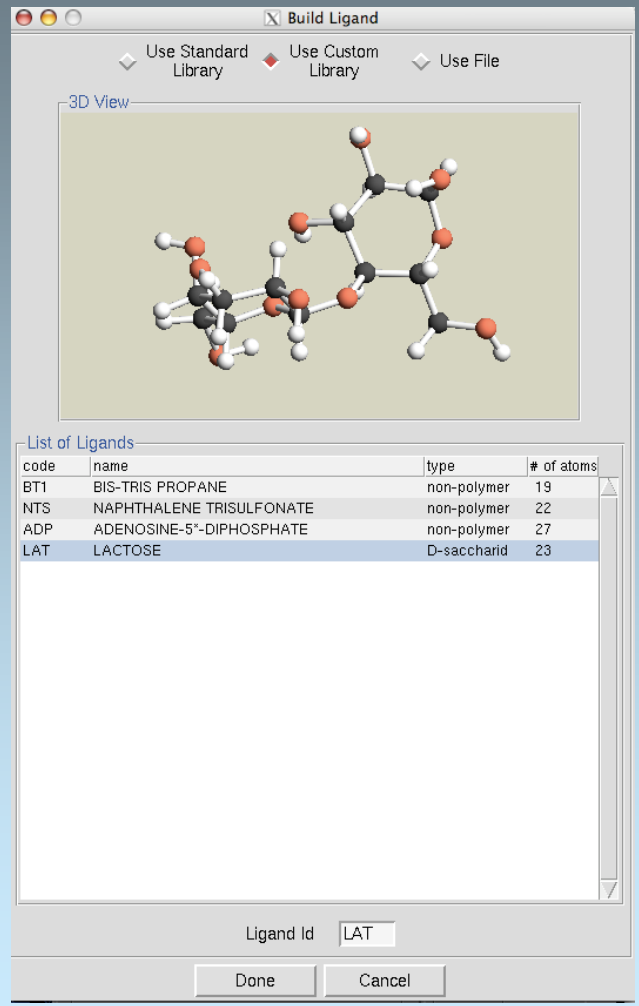
Wetlab
chemicals & solutions

Other systems provide additional contextual up- and downstream information

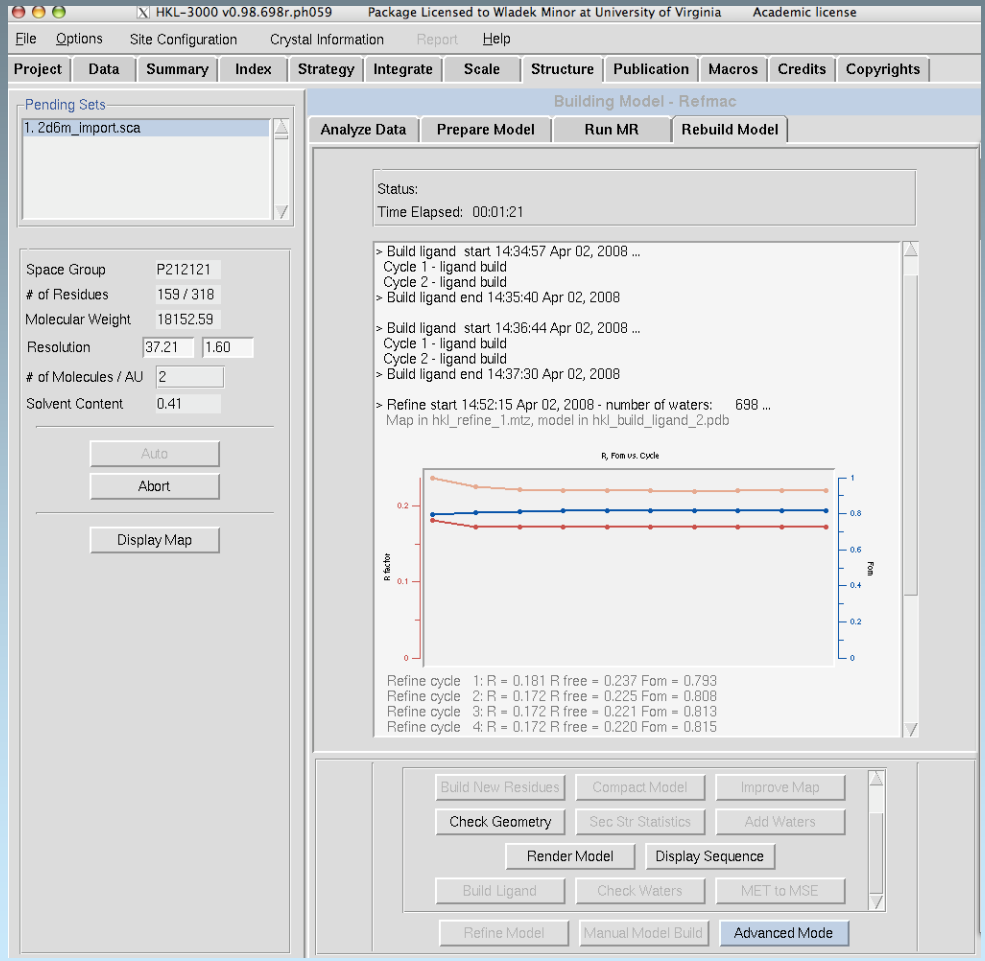


Ligand libraries

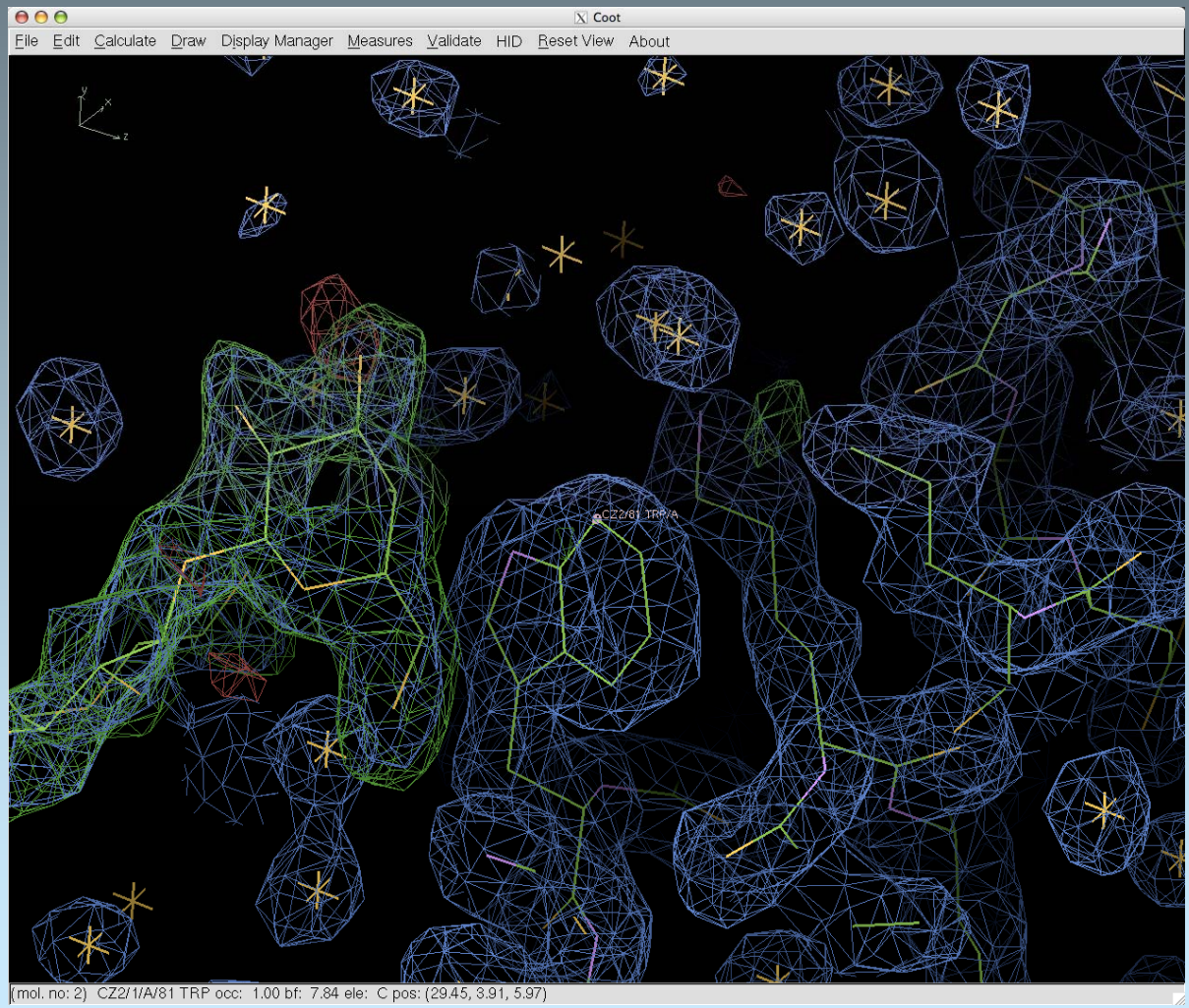




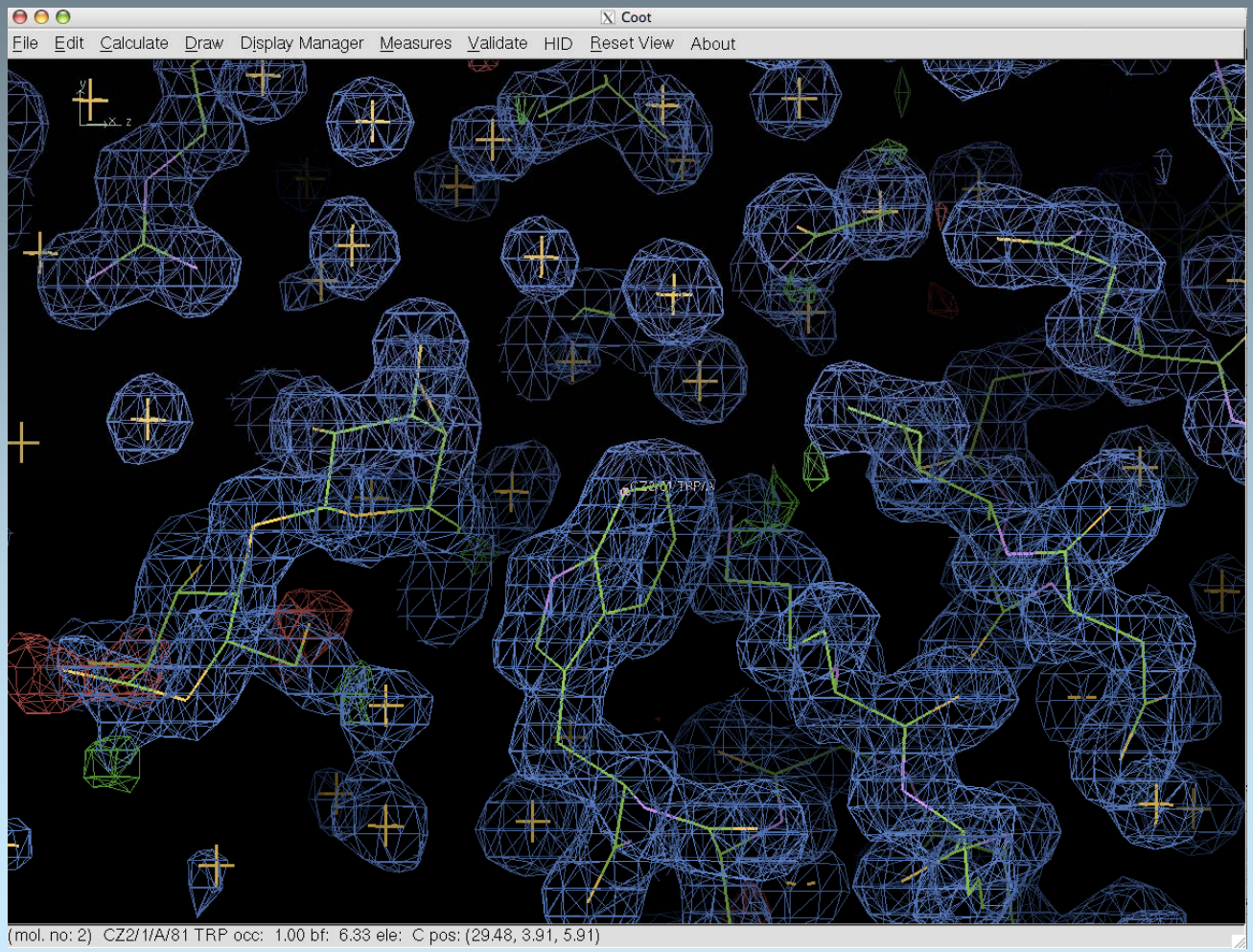
Automatic ligand building and refinement



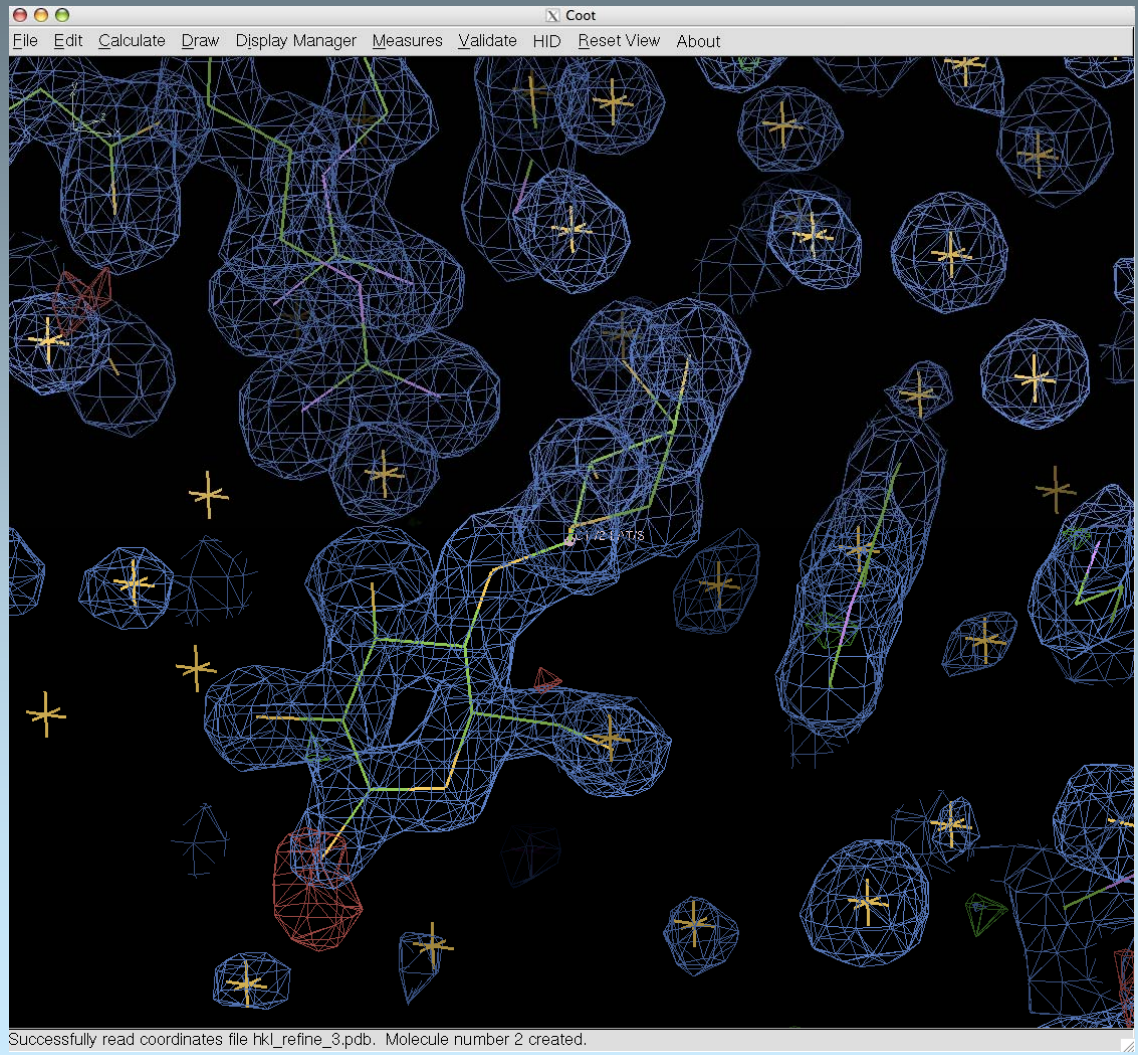
Unidentified densities



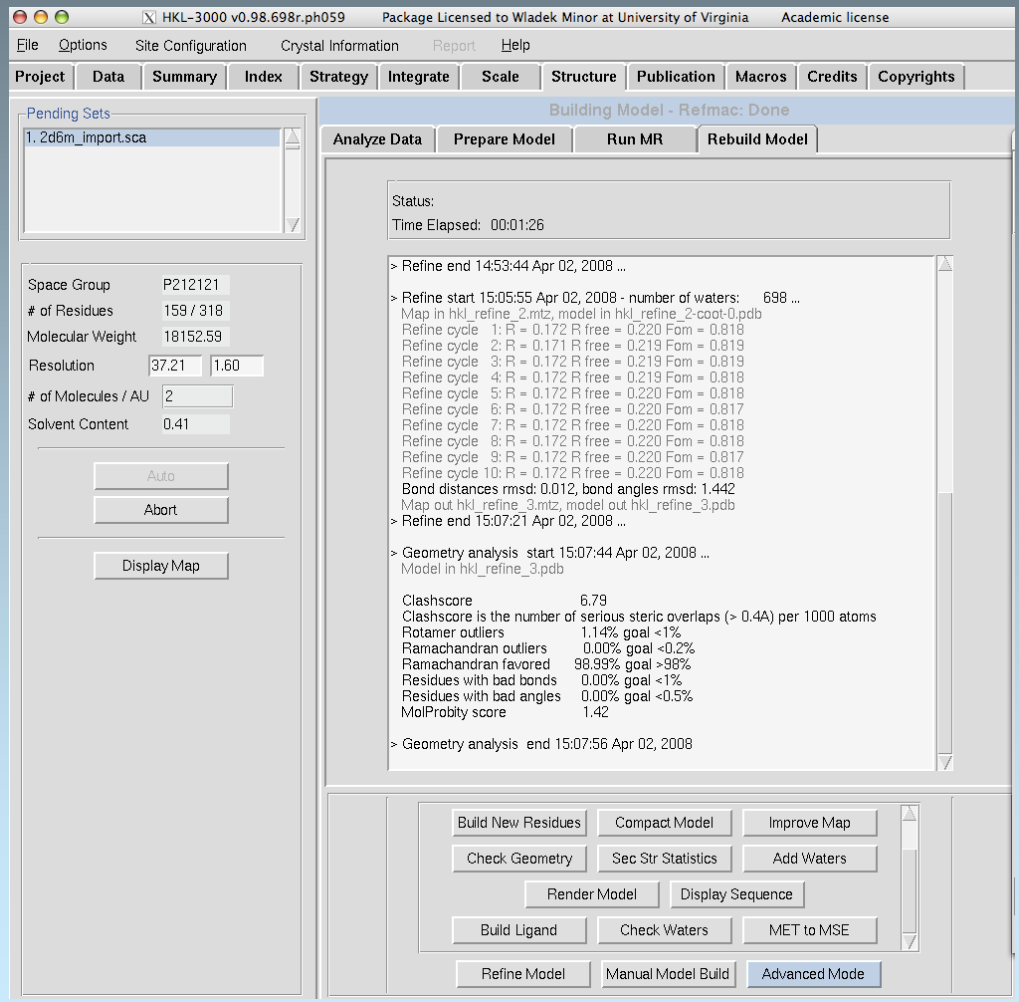
Automatic building



Fully refined ligand

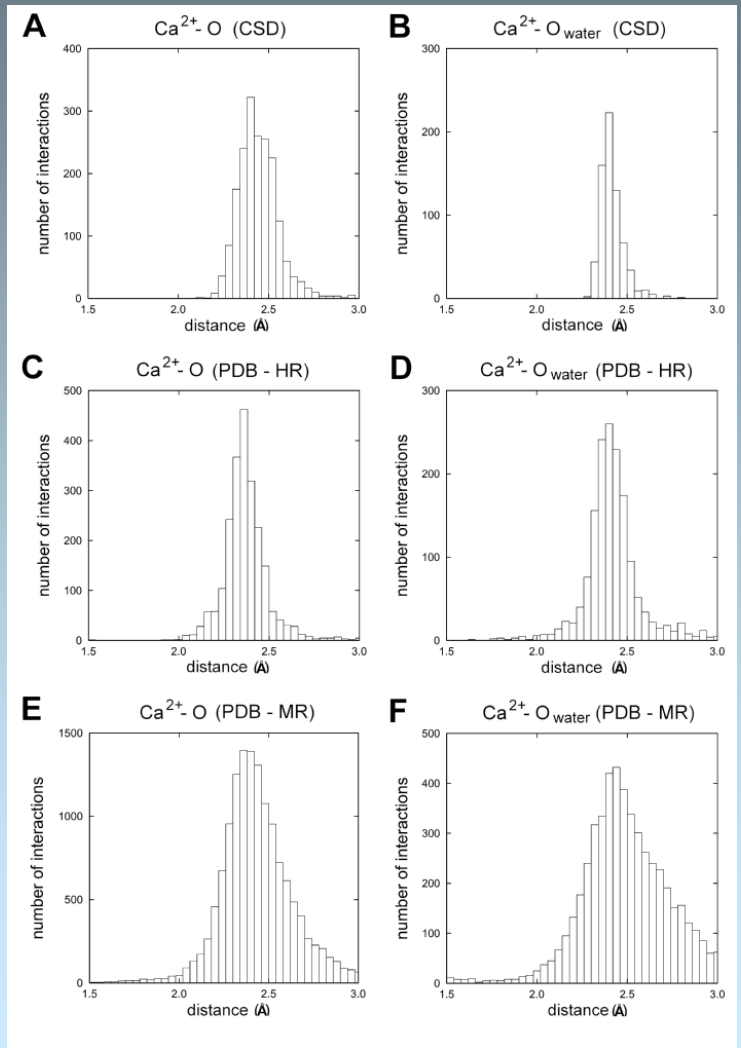


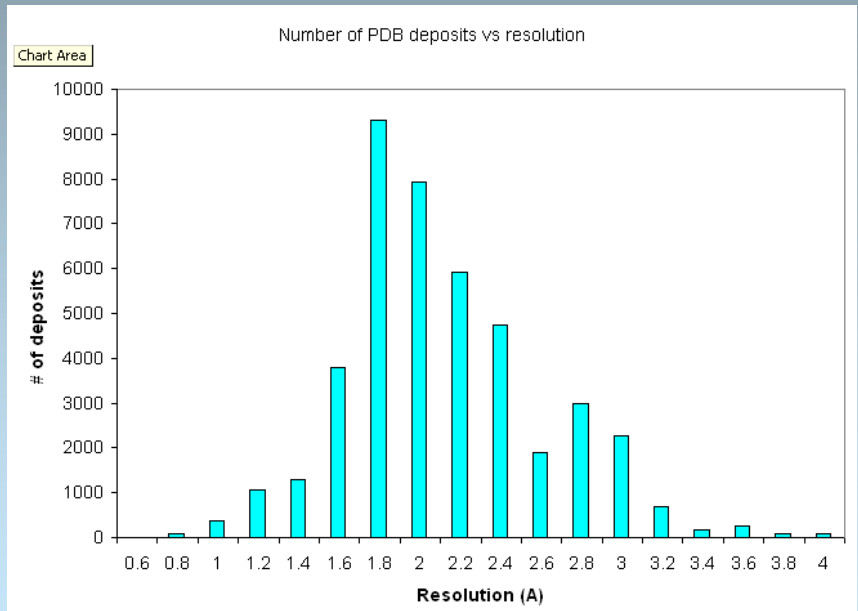
Structure validation



Argonne Users Meeting, May 2008

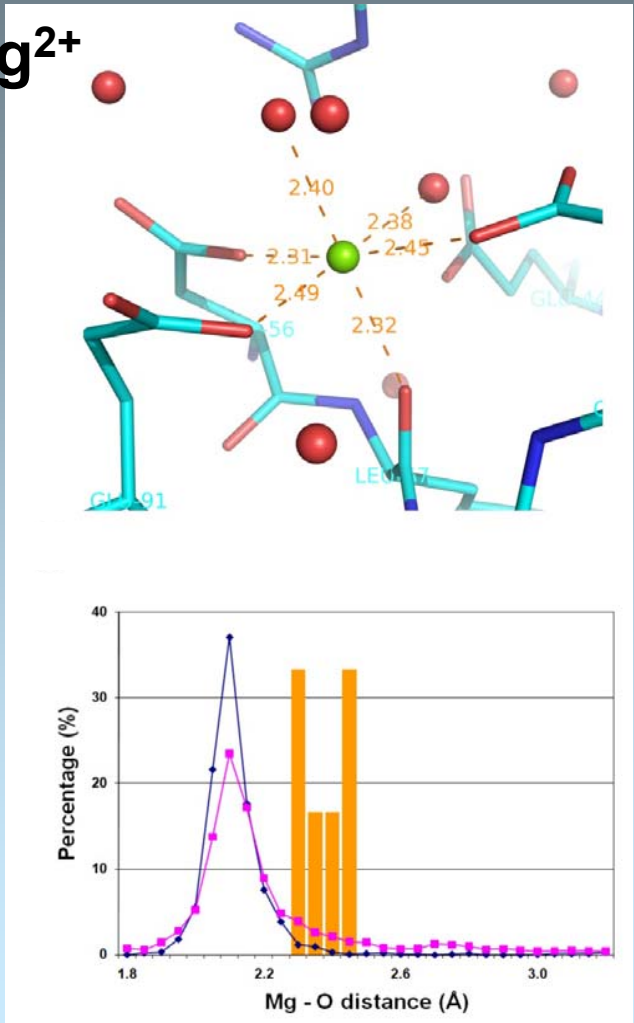
Ca-O distances



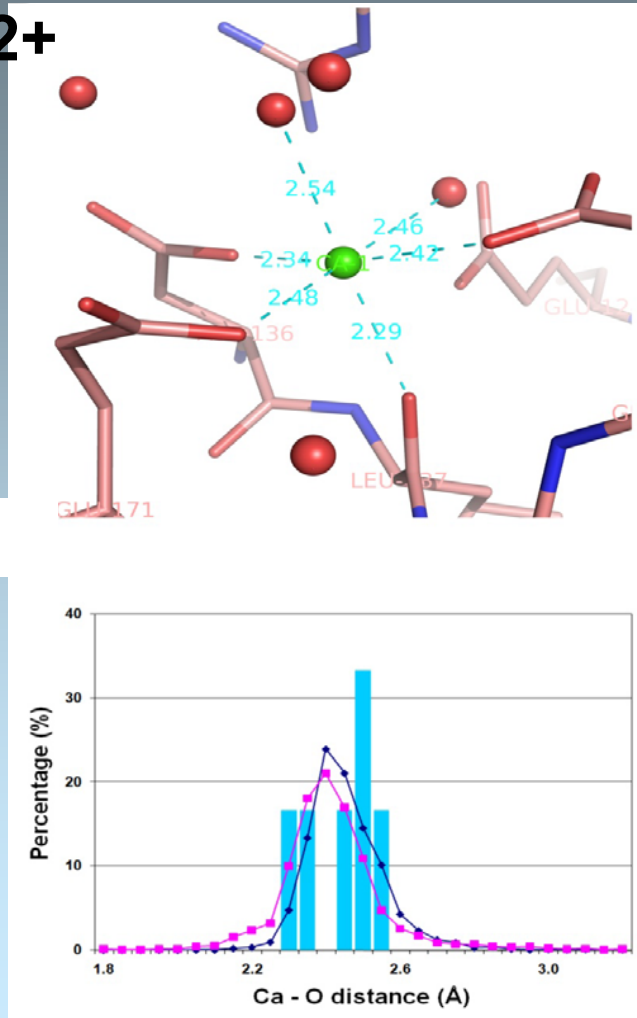


Metal ion in Der p 1

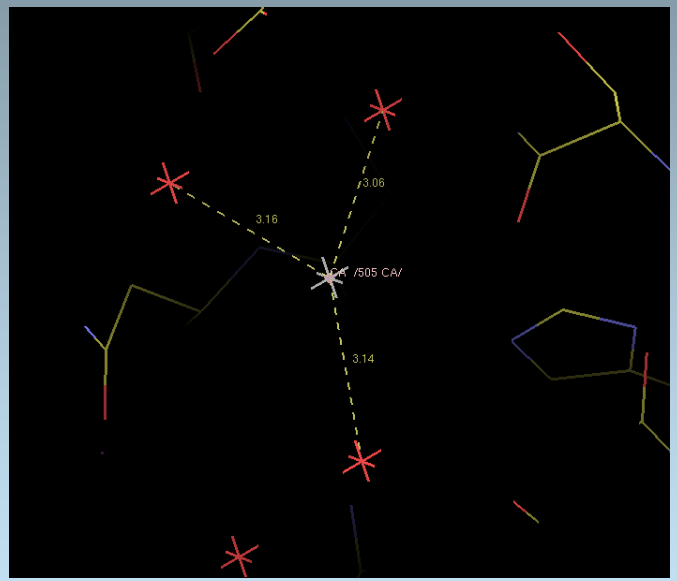
Mg^{2+}

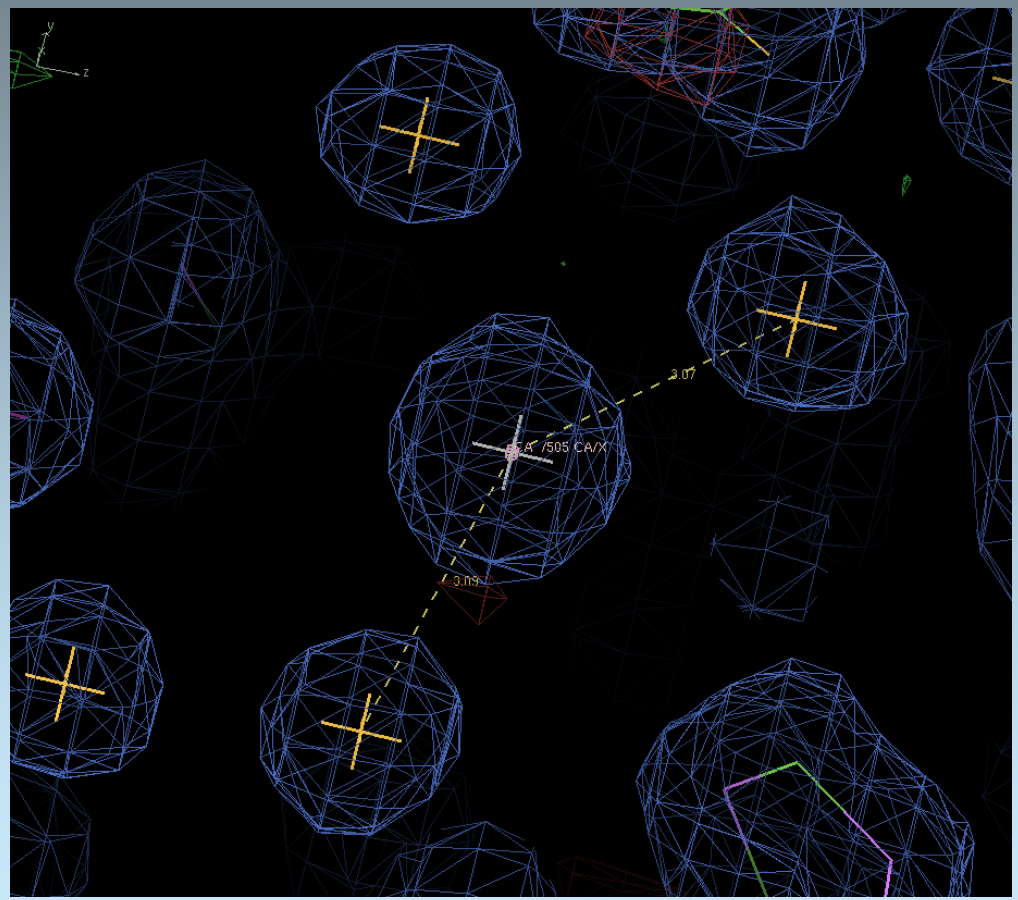


Ca2+

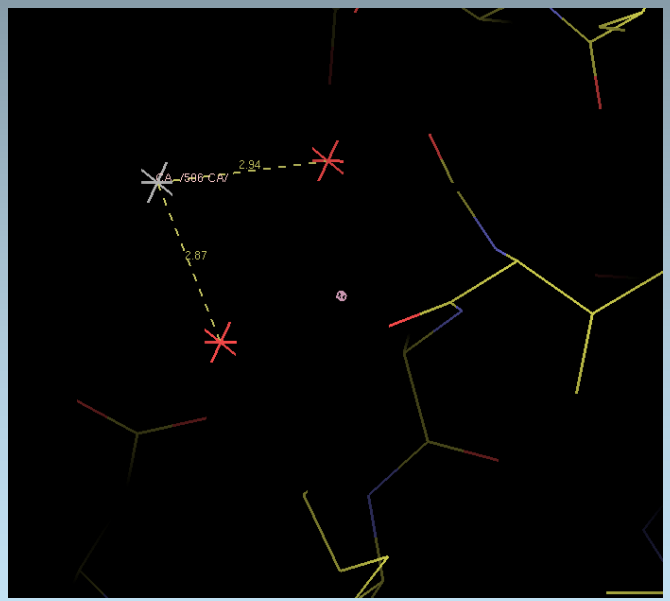


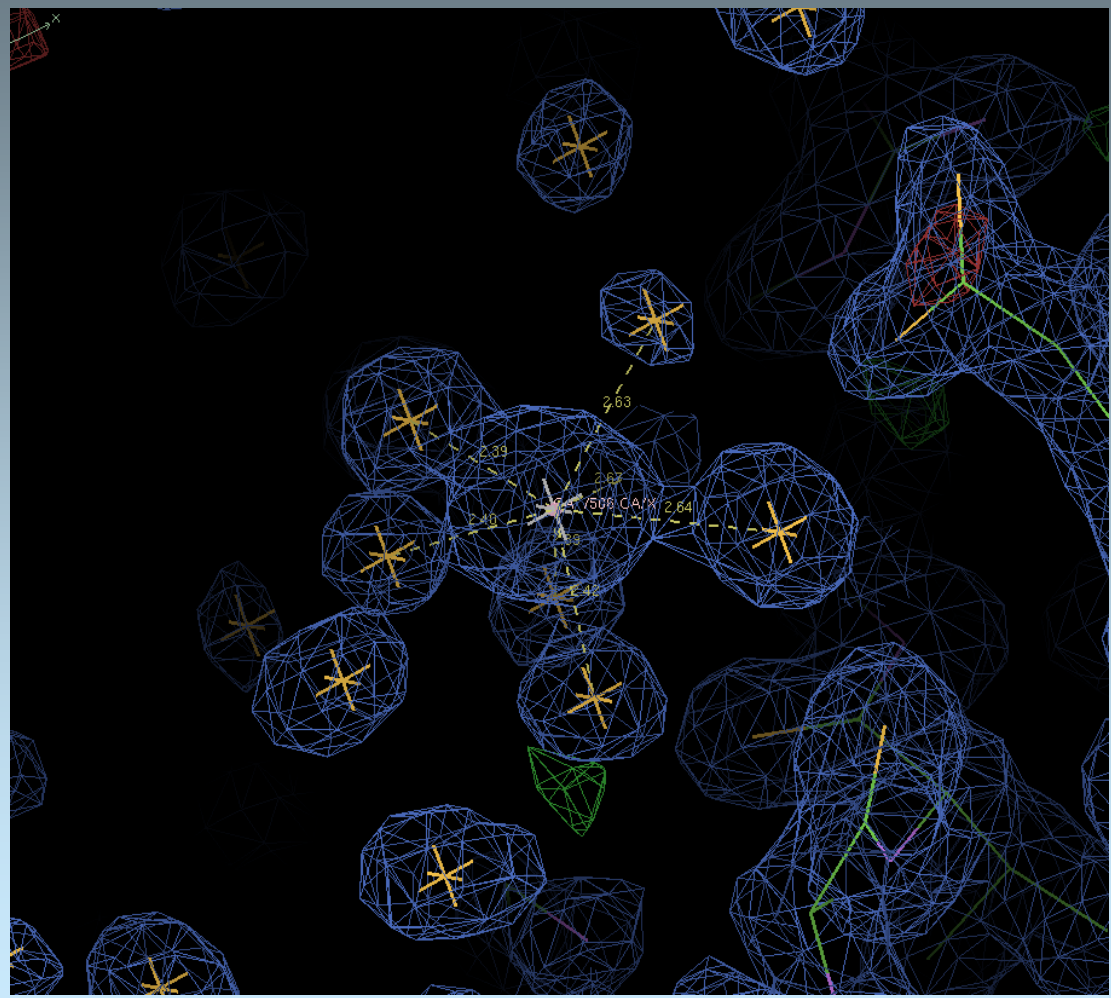
Incorrect assignments 1kyt



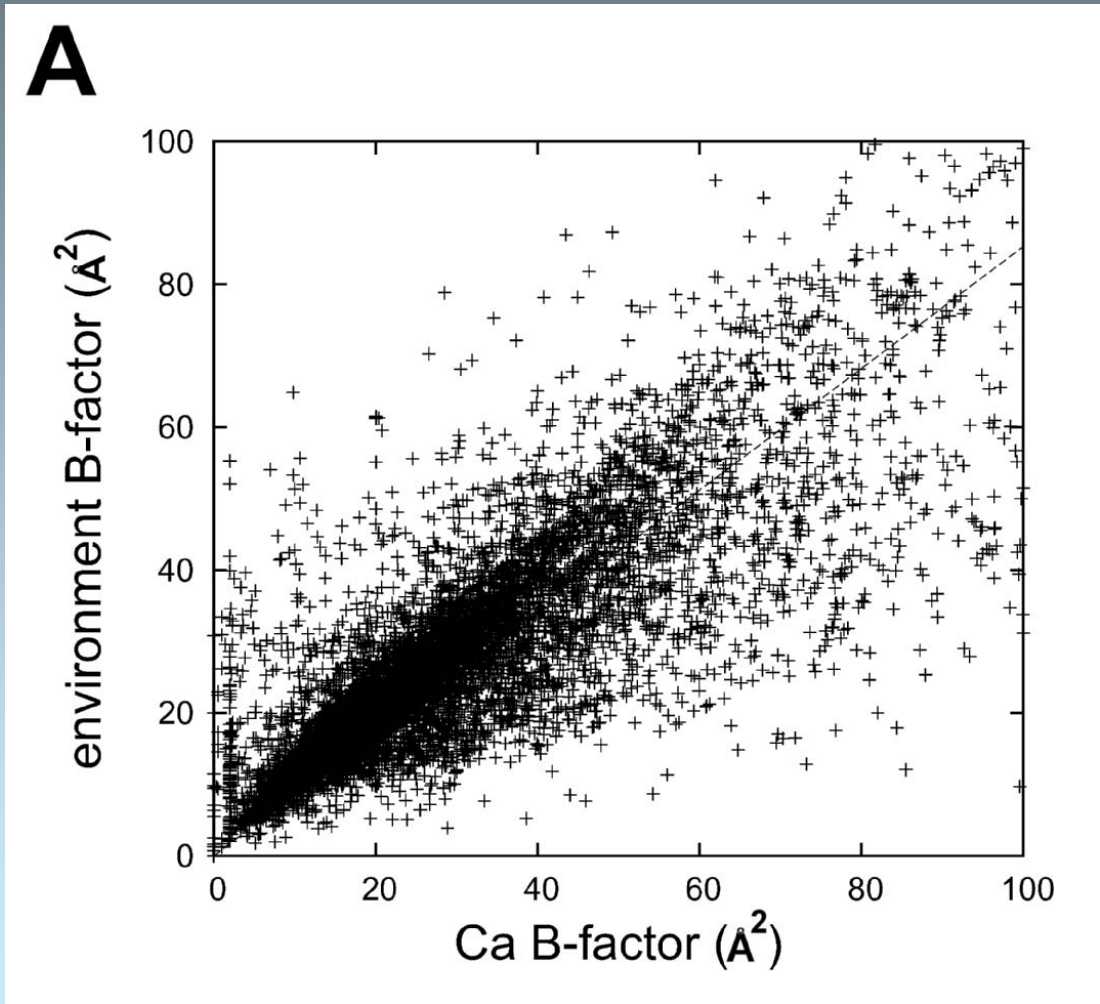


Correct assignment suboptimal refinement

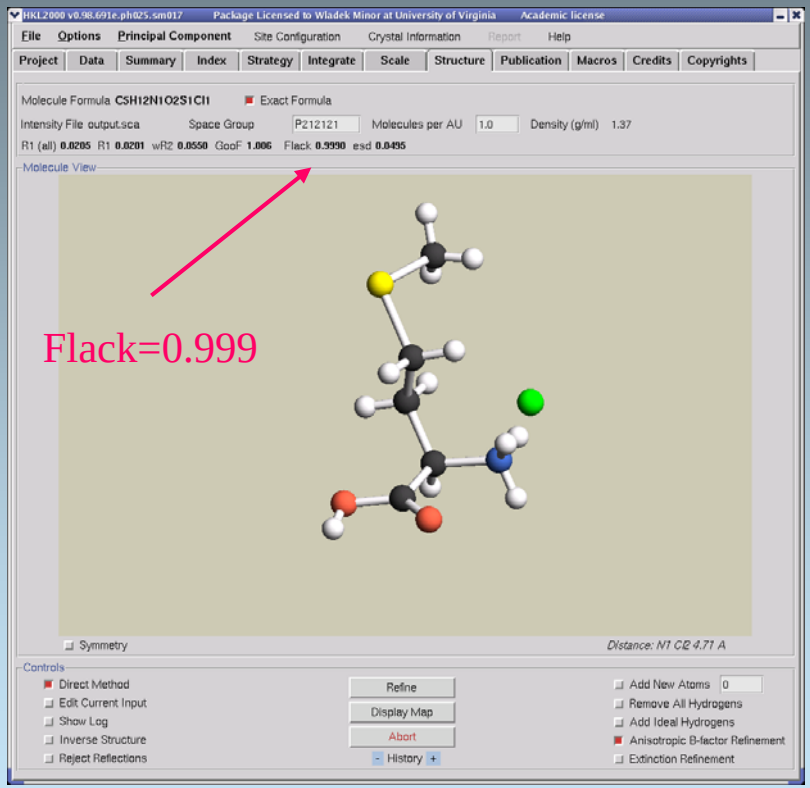


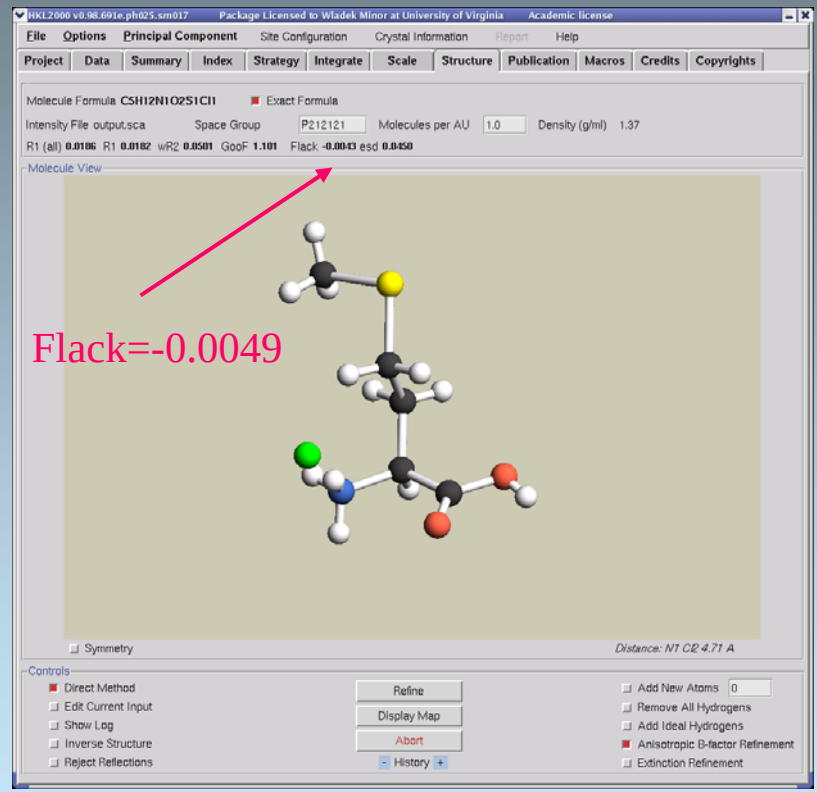


B factors



Absolute configuration and small molecule agents
HKL-3000SM





People involved

Wladek Minor
- Maks Chruszcz
- Marcin Cymborowski
- Matt Zimmerman
- Heping Zheng
- Marek Grabowski
- Haijun Qiu
- Olga Kirillova
- John Raynor

Zbyszek Otwinowski
- Dominika Borek
- Andrzej Kudlicki
- Roger Ljungberg

Andrzej Joachimiak
- Rongguang Zhan
- Marianne Cuff
- Chris Lazarski
- SBC staff

Zbyszek Dauter

Grants:
NIH GM53163, GM 62414,
DOE, NCI
HKL Research