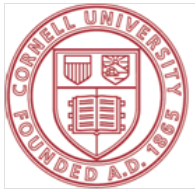
RAPD - Rapid Automated Processing of Data

Jonathan P. Schuermann

Cornell University / NE-CAT

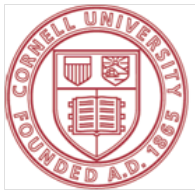
RAPD

- RAPD (Rapid Automated Processing of Data)
- Software Suite for Automated Macromolecular Crystallographic Data Analysis in Real Time
- Initial goal: Quickly provide automatic autoindexing and data collection strategies
- Morphed into much larger suite with additional pipelines for data processing and structure determination
- PI can monitor progress from home
- Results stored in database for improving RAPD or database mining



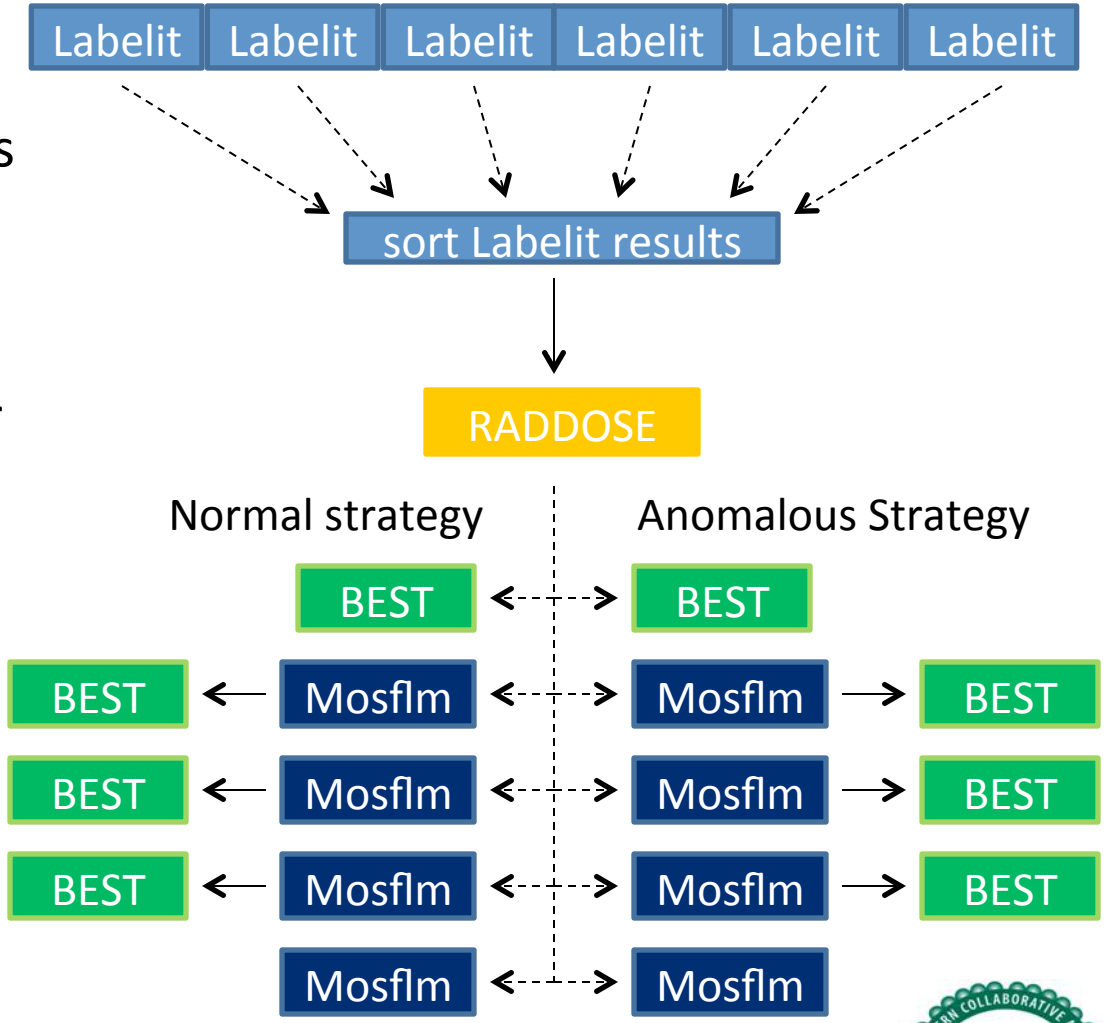
RAPD Pipelines

- **Autoindexing and strategy**
- STAC
- Data (re)processing and merging (David Neau)
- Data analysis
- SAD, MAD, SIRAS
- Molecular replacement



RAPD Autoindex/Strategy

- Multiple Labelit runs with different peak picking settings for Distl
- Sort by highest symmetry
- Raddose
- BEST strategies calculated for regular and anomalous data collections
- If BEST fails, Mosflm strategy calculated
- Highly parallelized code with several layers of error detection/correction
- Run time: ~25s on an Intel Core i7



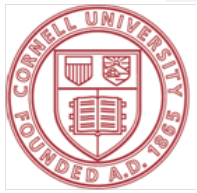
Distl Settings

ADSC Q315

Labelit run number	1	2	3	4	5	6
distl.minimum_spot_area (10)		6		7	8	6
distl.minimum_spot_height (3.5)			6			
distl.minimum_signal_height (1.5)		4.3		1.2		
distl_highres_limit						5

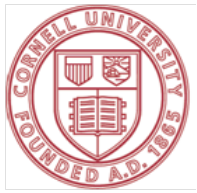
Dectris Pilatus

Labelit run number	1	2	3	4	5	6
distl.minimum_spot_area (5)		3		2	3	2
distl.minimum_spot_height (3.5)			6			
distl.minimum_signal_height (2.5)		4		2.3		
distl_highres_limit						5

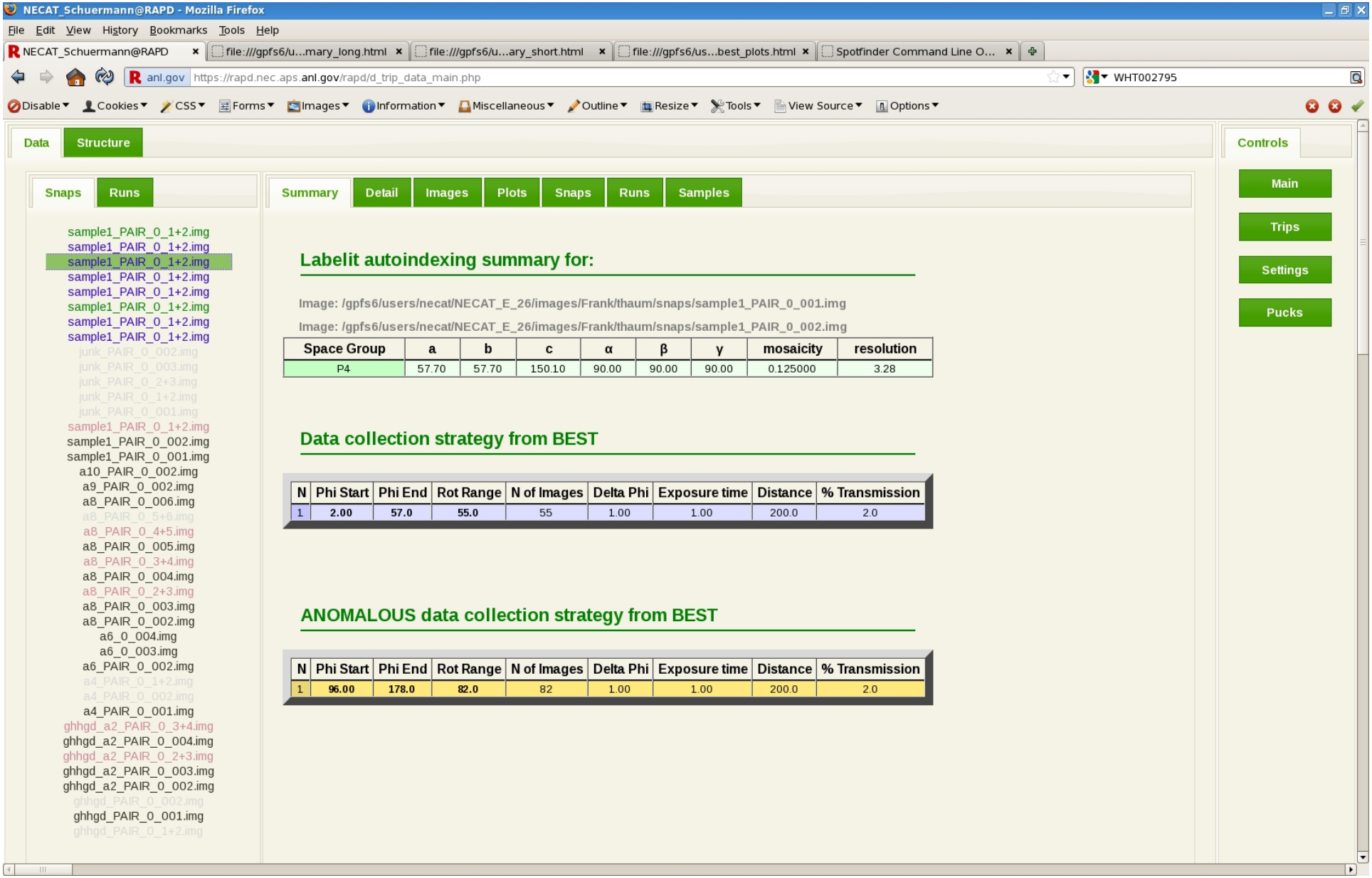


Labelit Error Checking

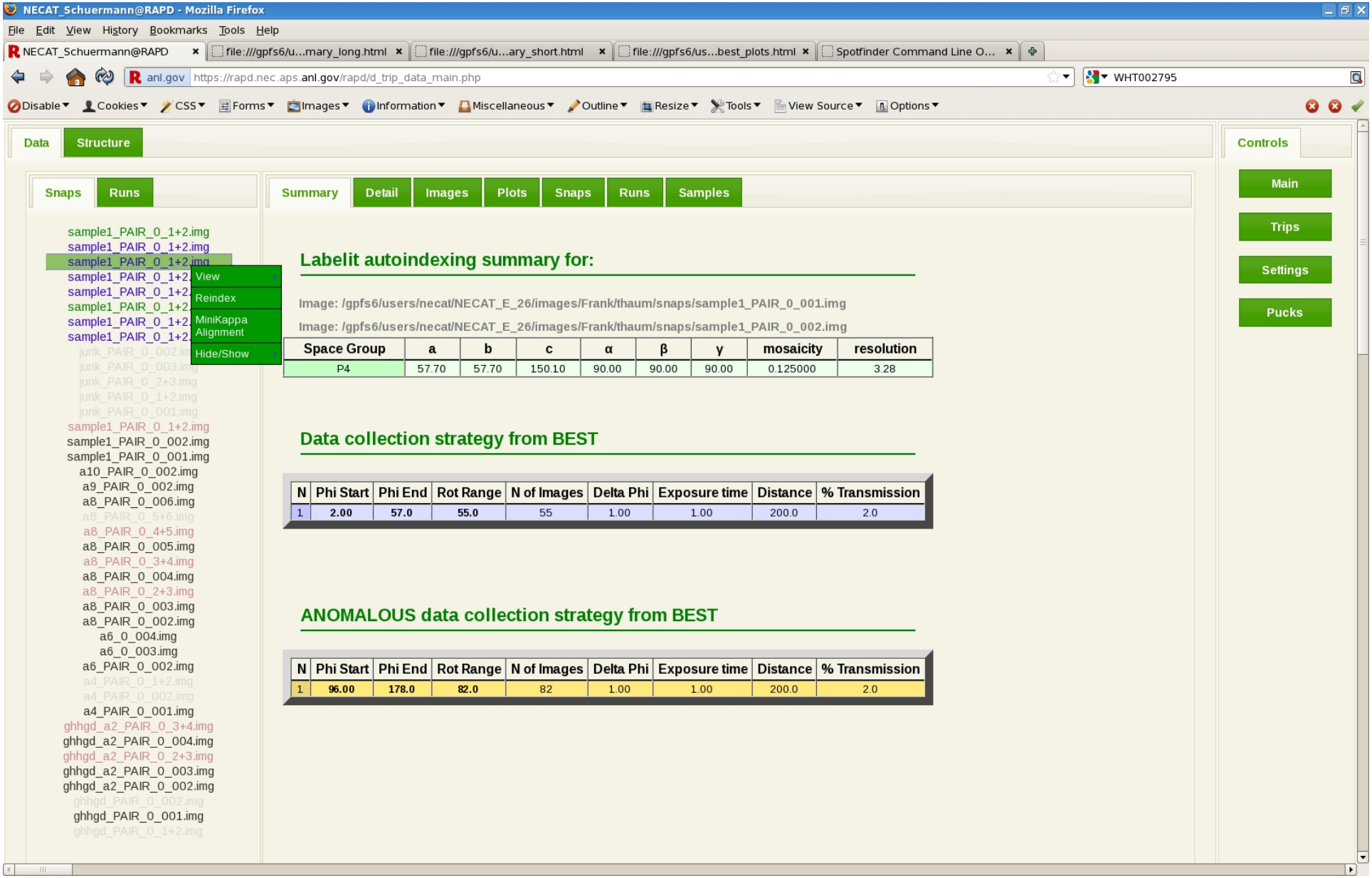
- The code will attempt to fix common autoindexing errors and rerun Labelit
  - If too few spots, resets min # of spots required
  - If unit cell/SG (specified by user) fails, resets without cell/SG restraints
  - If Mosflm step fails, lowers resolution limit
  - And any combination of the previous errors



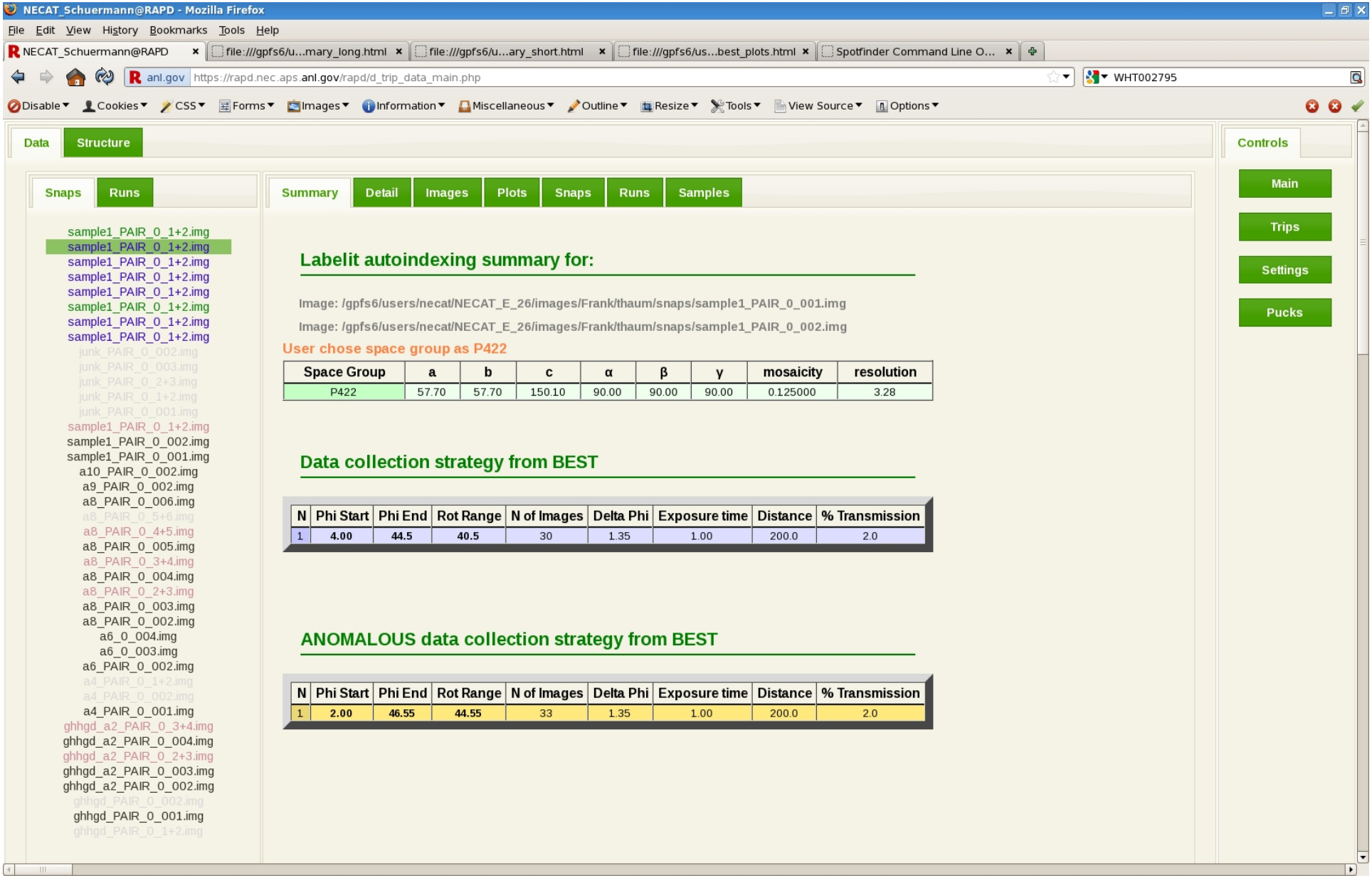
RAPD Autoindex and Strategy



RAPD Autoindex and Strategy

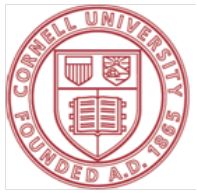


RAPD Autoindex and Strategy

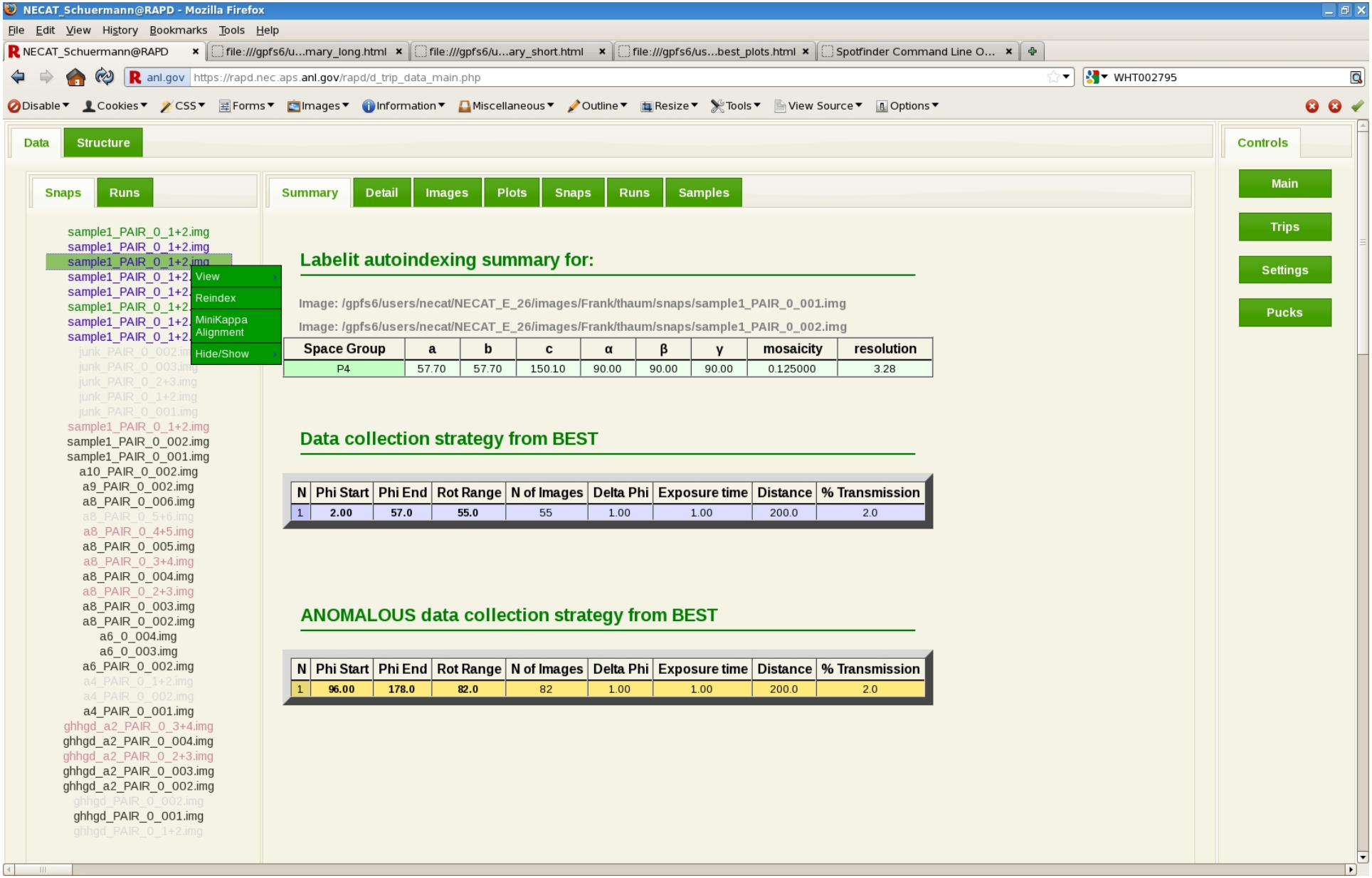


RAPD Pipelines

- Autoindexing and strategy
- **STAC**
- Data (re)processing and merging (David Neau)
- Data analysis
- SAD, MAD, SIRAS
- Molecular replacement

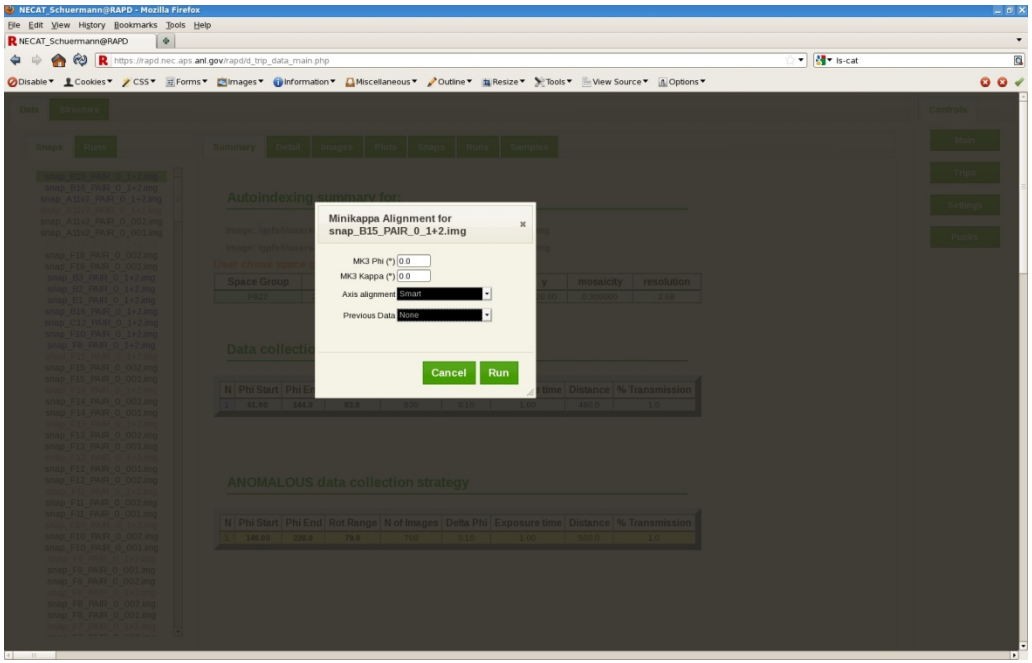


RAPD Autoindex and Strategy

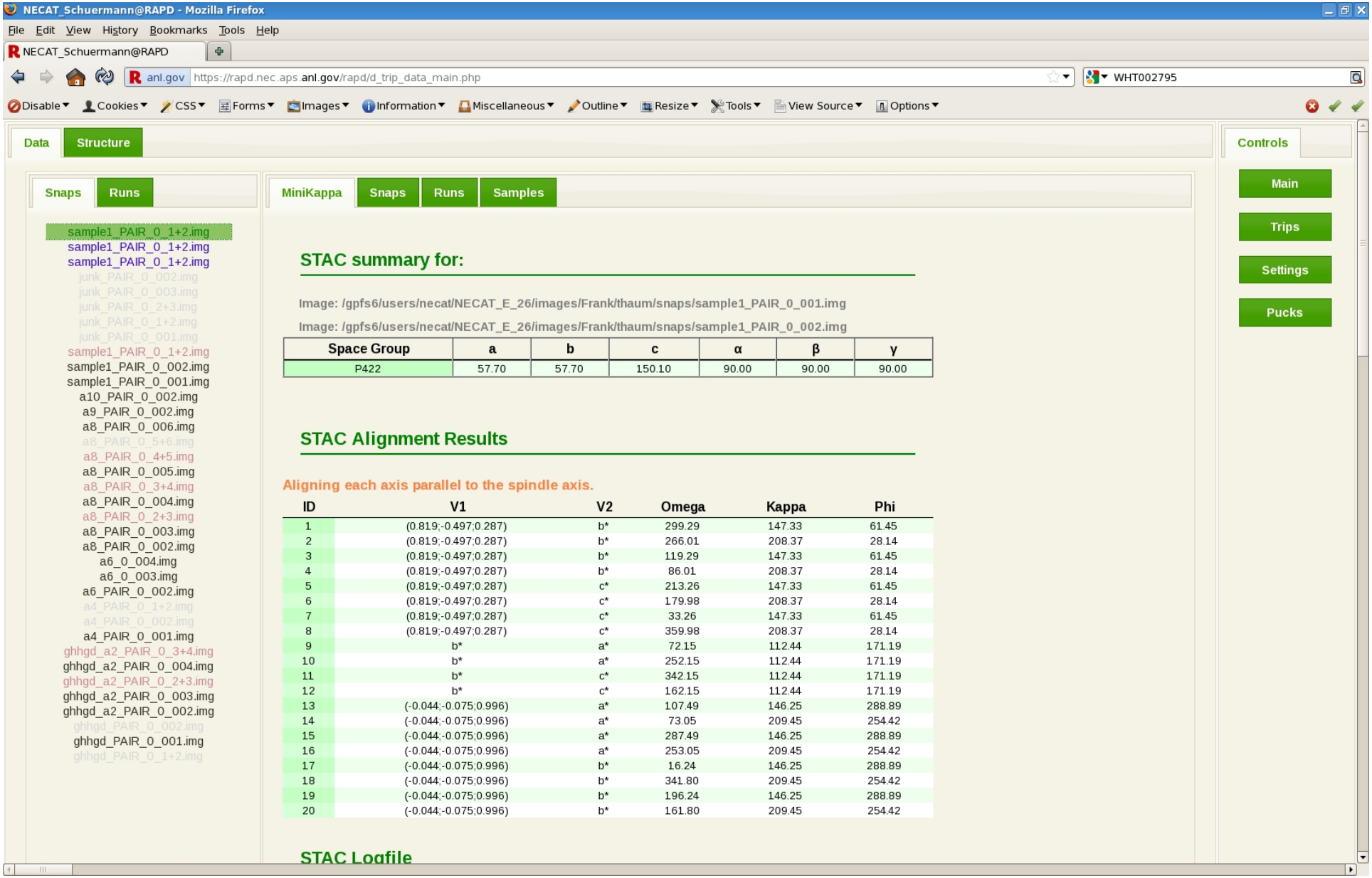


STAC Options

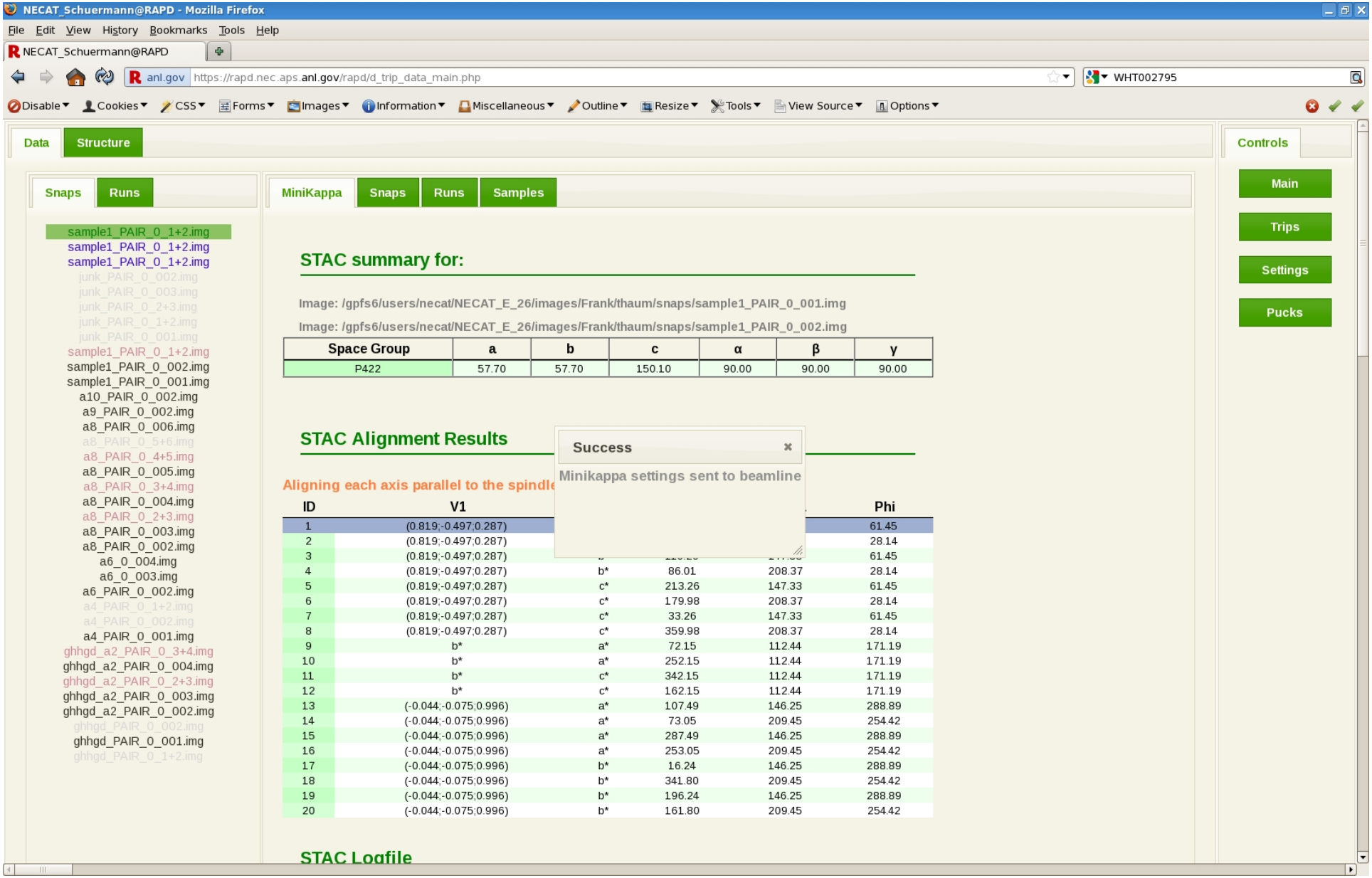
- Alignment options:
  - smart, long, anomalous, a*,b*,c*, split a*b*, split a*c*, split b*c*, all
  - Can align crystal to previously collected crystal



STAC Alignment Results

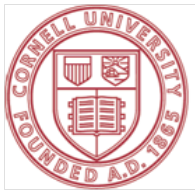


Send STAC Alignment to Beamline



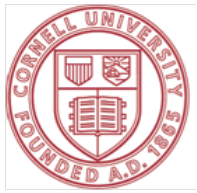
RAPD Pipelines

- Autoindexing and strategy
- STAC
- Data (re)processing and merging (David Neau)
- **Data analysis**
- SAD, MAD, SIRAS
- Molecular replacement

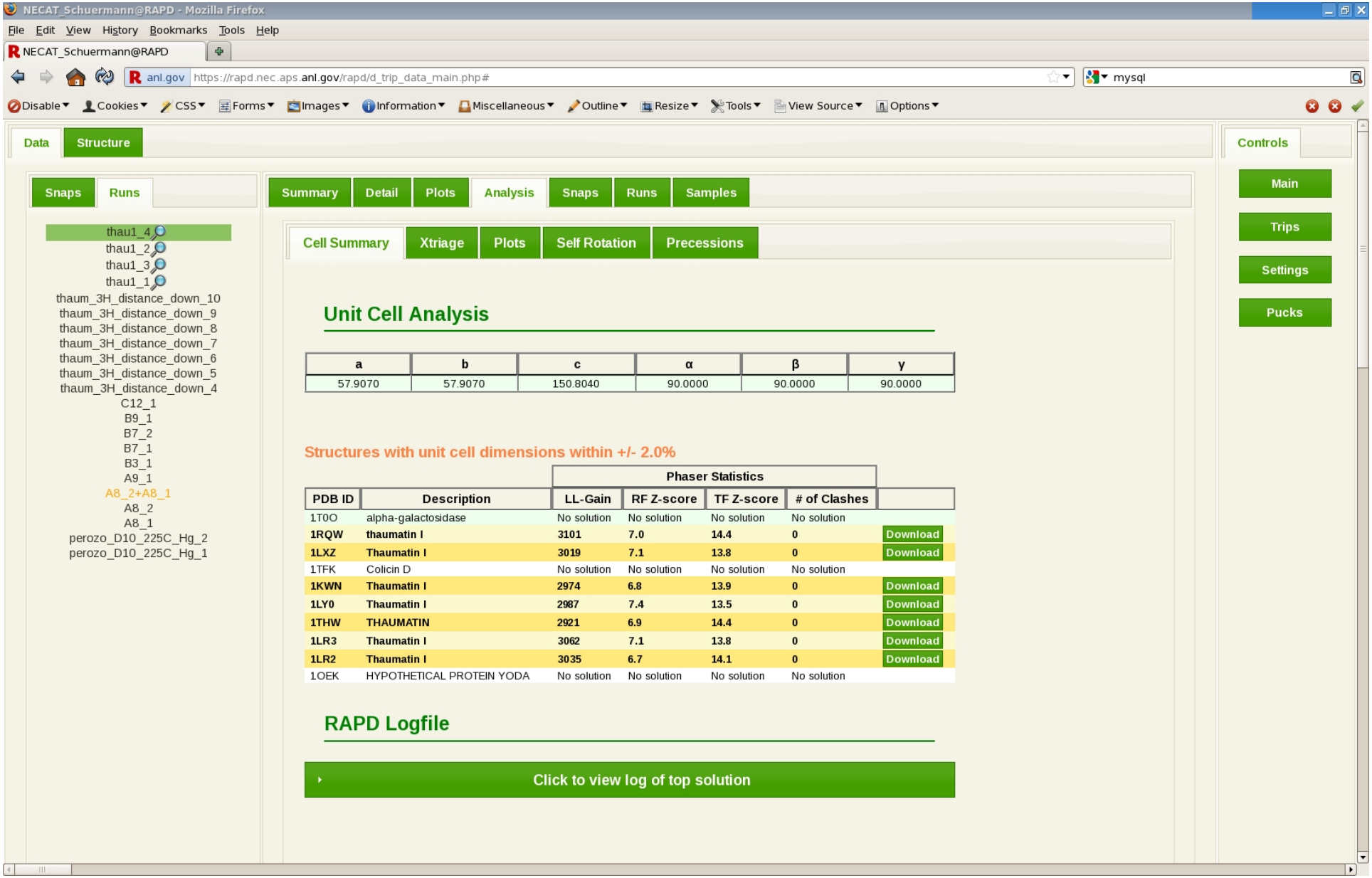


Data Analysis Pipeline

- Unit cell analysis
- Xtriage
- Xtriage plots
- Molrep self-rotation function
- labelit.precession_photo

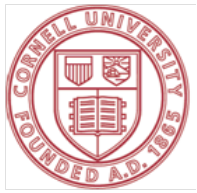


Unit Cell Analysis



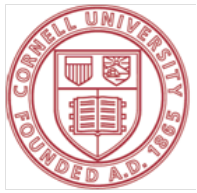
RAPD Pipelines

- Autoindexing and strategy
- STAC
- Data (re)processing and merging (David Neau)
- Data analysis
- **SAD, MAD, SIRAS**
- Molecular replacement

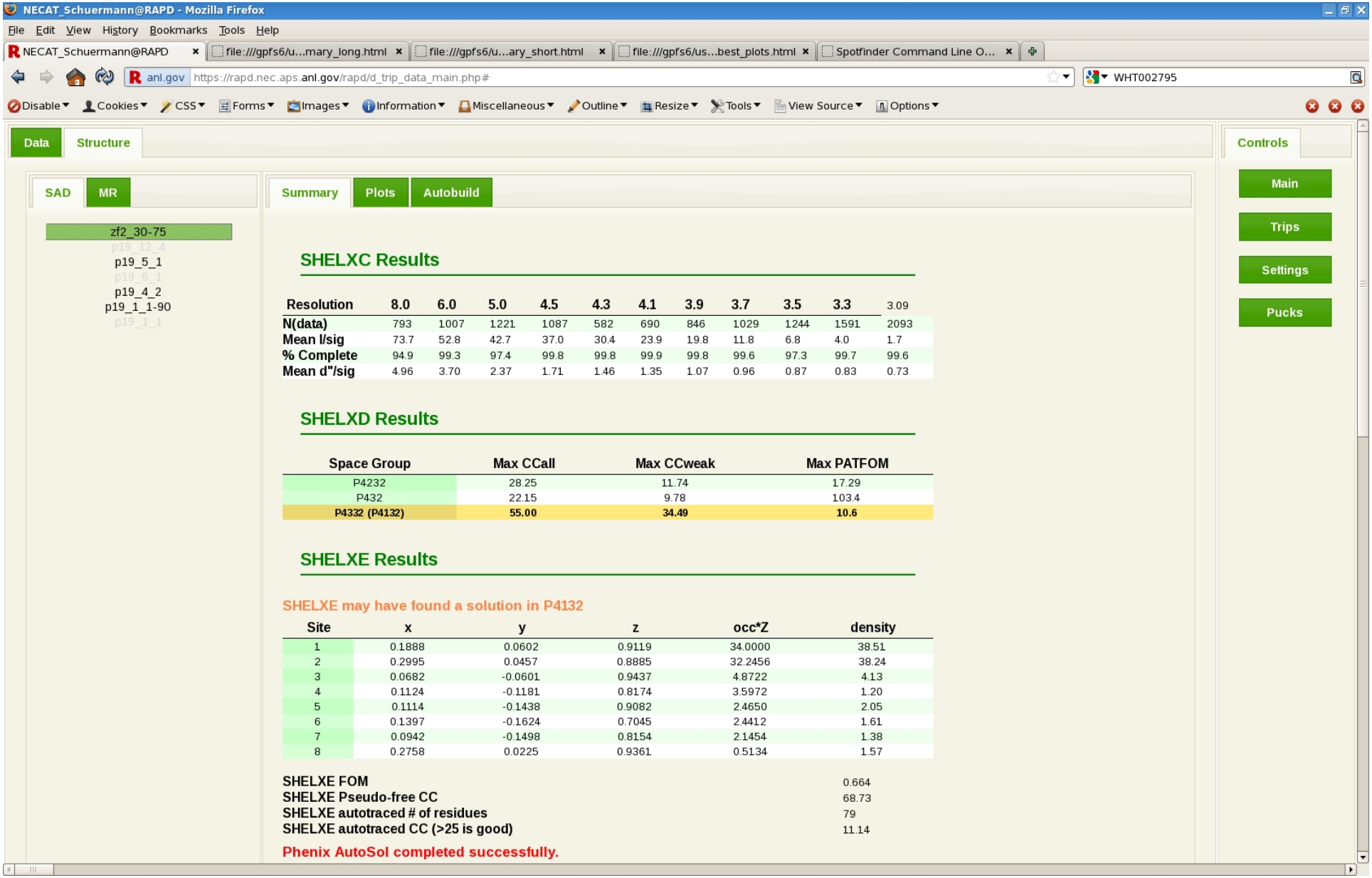


SAD, MAD, SIRAS Pipelines

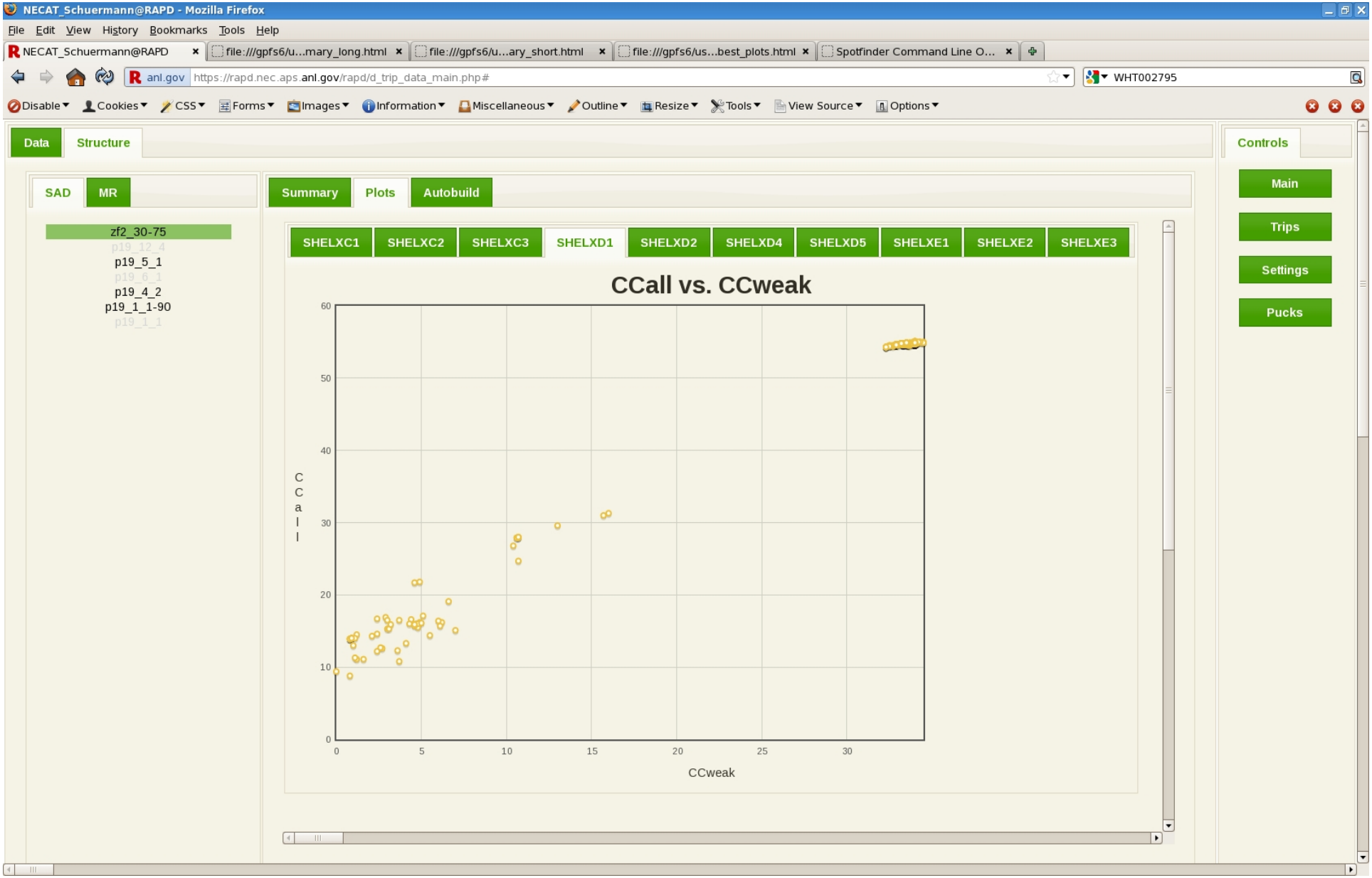
- SHELX(CDE)
- Run all possible space groups and cell permutations simultaneously
- Code written to parallelize SHELXD similar to SGXPro and new code written by Sheldrick to parallelize further
- Results: 1024 trials in all SG's ~1 minute
- Autotracing performed by SHELXE, AutoSol, and Buccaneer
- 'Hand' and SG determined automatically
- User can rerun changing the resolution cutoff, number of sites, or number of SHELXD trials
- All results and plots presented on RAPD GUI and are downloadable



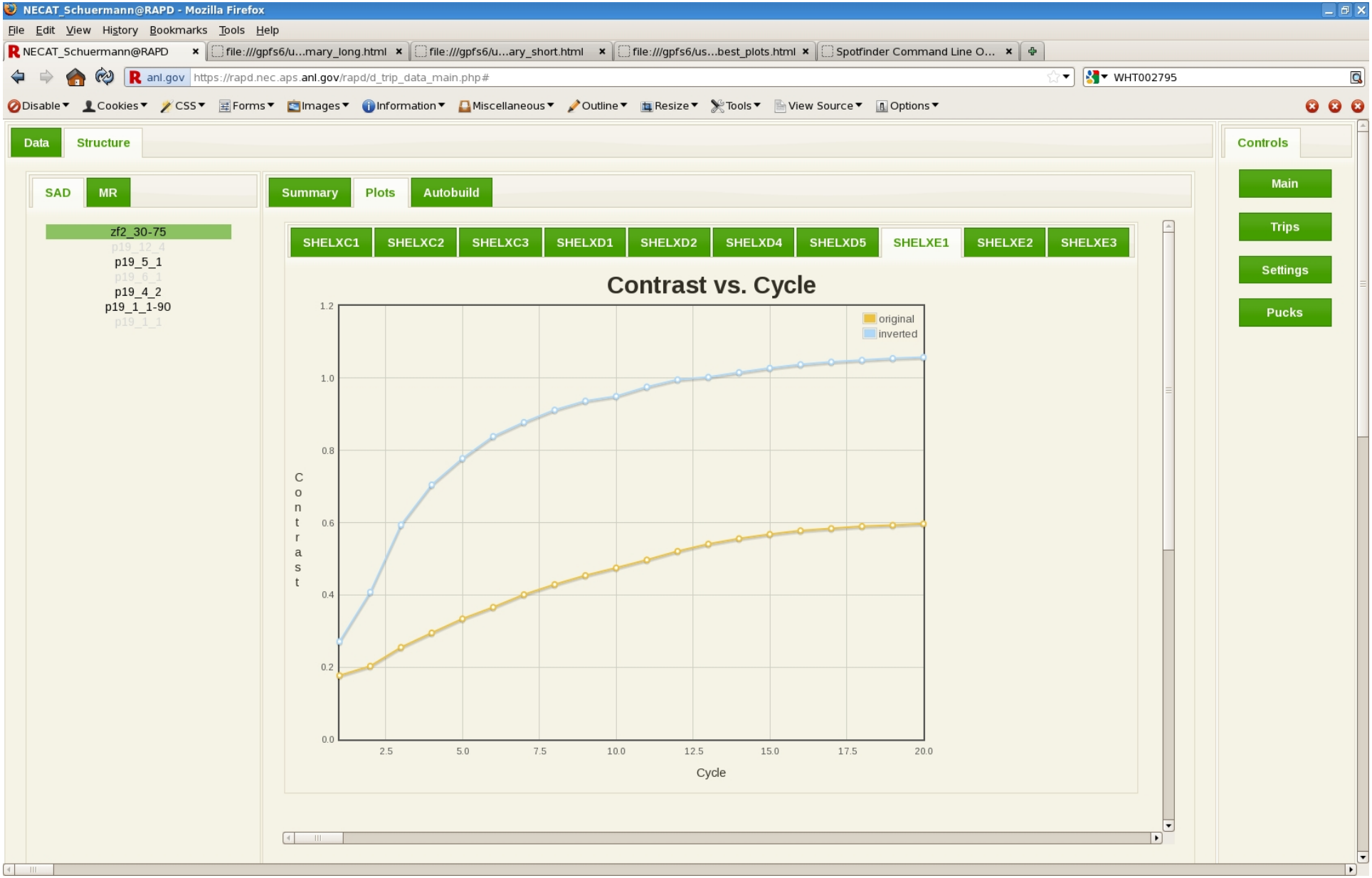
SAD, MAD, SIRAS Pipeline



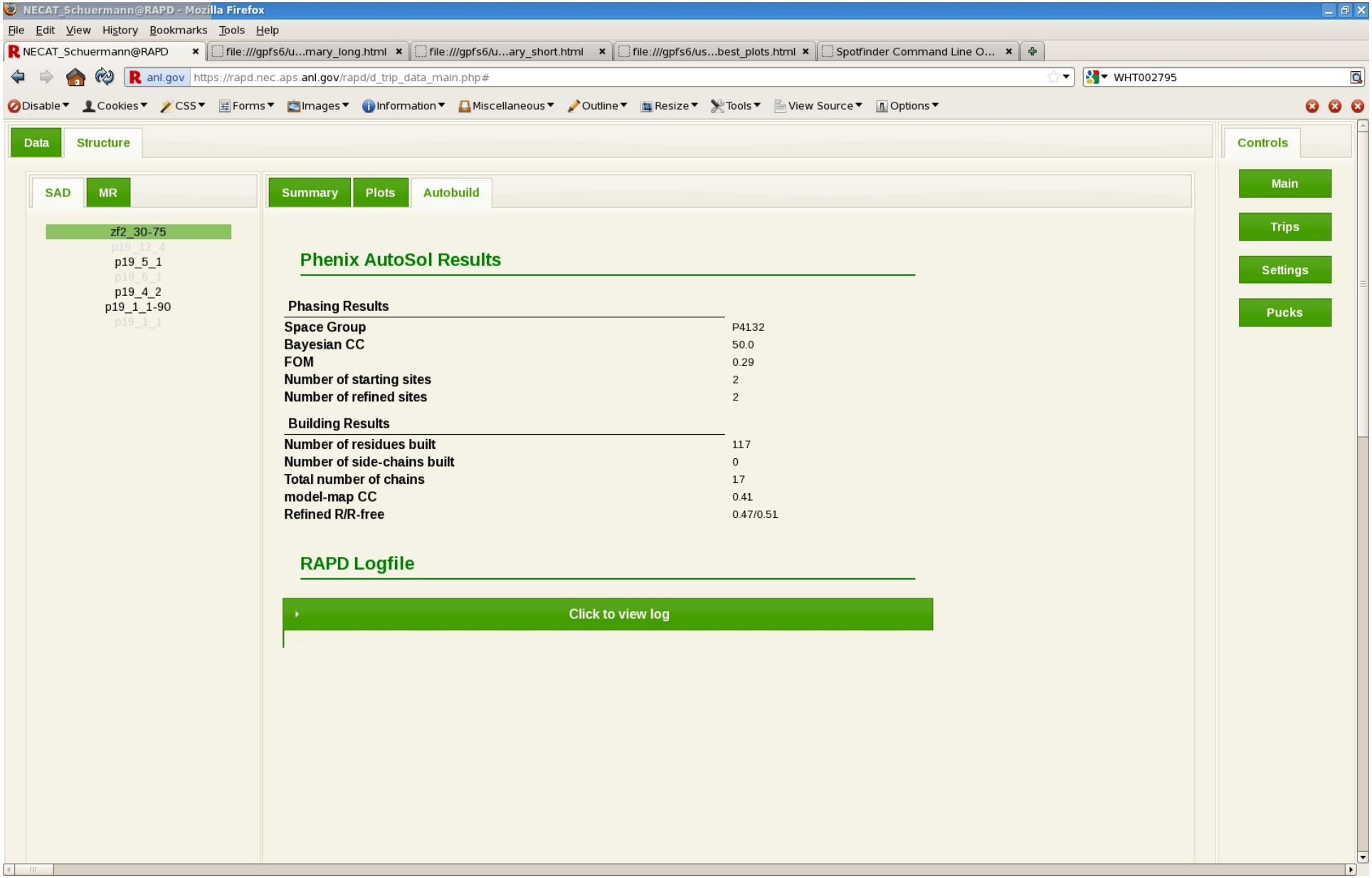
SHELXD Plots



SHELXE Plots

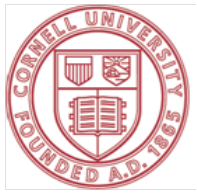


Autosol Results



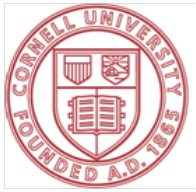
RAPD Pipelines

- Autoindexing and strategy
- STAC
- Data (re)processing and merging (David Neau)
- Data analysis
- SAD, MAD, SIRAS
- **Molecular replacement**

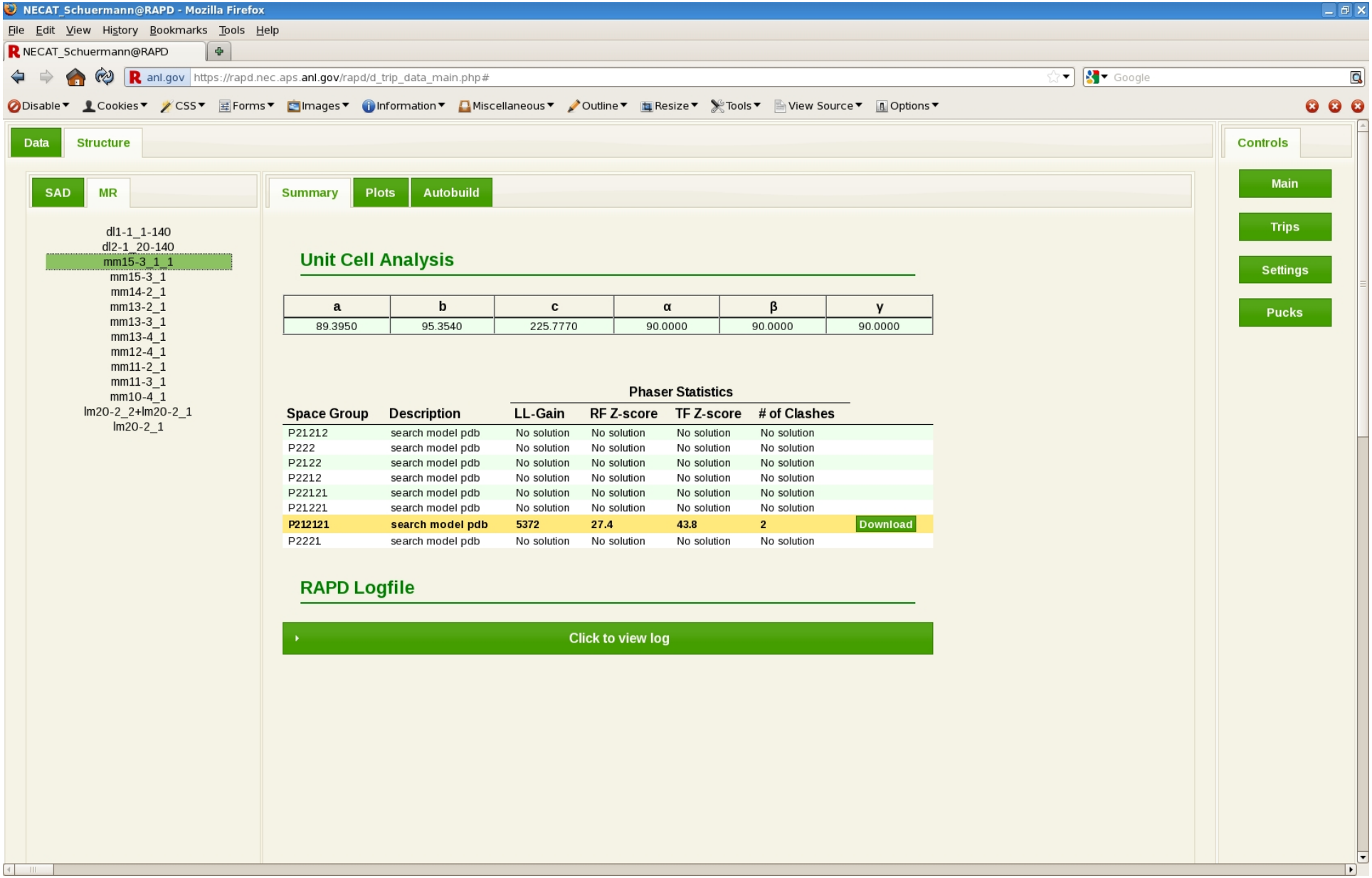


Molecular Replacement

- Run in all possible SG's and cell permutations simultaneously (actually twice)
  - first run: the top peak has to be > 4 sigma RF and 6 sigma TF
  - second run: default Phaser parameters but only the top peak
- User can input PDB code or file
- Matthew's run to determine number of molecules in the AU automatically (user can override)
- Statistics presented in RAPD GUI
- Results are downloadable

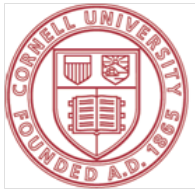


MR Pipeline Results



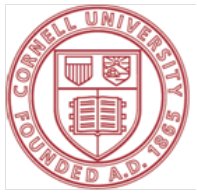
Current Capabilities of RAPD

- GUI can be viewed and results downloaded from an internet browser
- Automatic autoindex/strategy in ~20s
- Automatic data processing in ‘real’ time
- Automatic data analysis
- Automatic MR if similar structure in PDB
- Pipelines for SAD, MAD, SIRAS, and MR
- Datasets can be reprocessed and merged
- Results stored in users directory
- Details of each result stored in MySQL database
- So far:
  - 357,477 autoindexing/strategy
  - 30,575 datasets processed
  - 1,884 SAD jobs
  - 1123 MR jobs
  - 88,019 unit cell analysis runs (9,781 solutions with LL-gain > 100)

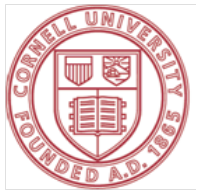


Future Goals/Wishes of RAPD

- Speed… much more speed (threading detector modules)
- XDS mosaicity calculated from small wedges of ‘snapshots’
- Shift libraries to CCTBX for compatibility
- More database mining
- Run auto-beam center calculation in background
- MAD and MRSAD pipelines
- New GUI
- Better hooks for wrapping
- Multiple lattice integration
- Better support for multi-wedge data collection



Acknowledgments

- Frank Murphy – Core and user interface
- David Neau – data (re)processing
- Kay Perry, Surajit Banerjee

- RAPD
  - https://rapd.nec.aps.anl.gov/rapd
- Wiki
  - https://rapd.nec.aps.anl.gov/wiki/main_page