Experimental Phasing with CCP4

- Speaker: Charles Ballard
- Program Authors:
  - Raj Pannu, Pavol Skubak
  - Randy Read, Airlie McCoy
  - Rudolf de Graaff, Jan Abrahams
  - Kevin Cowtan
  - Garib Murshudov
  - Phil Evans
  - (George Sheldrick)
- “Every program is becoming an experimental phasing program” - Kevin Cowtan
Data and Resolution

- Spacegroup
  - Translation symmetry
  - Fixed origin
  - No special positions
  - P21 21 21 ideal
- Redundancy
- Scala
  - Completeness
  - $\langle I \rangle / \langle \sigma I \rangle$
  - Inter dataset anomalous correlation – MAD ($> 0.3$)
  - Intra dataset anomalous correction - SAD
Data and Resolution II

- $<\text{Dano}>/\text{<sigDano>}$
  - $>1.2$
  - Output by AFRO
  - Also SHELXC via crank.
- George Sheldrick's rule of thumb
  - Hi Res + 0.5
  - 3.0 A
Running CRANK
http://www.bfsc.leidenuniv.nl/software/crank/

- Takes Merged data
- Experiment type
- Atom types plus $f'$ and $f''$
- Customize pipeline
- Heart of CRANK
  - AFRO
  - CRUNCH2
  - BP3
Flow inside CRANK

**GCX**

AFRO, SHELXC

CRUNCH2, SHELXD

BP3, SHELXE

SOLOMON, PARROT, PIRATE, RESOLVE

ARP/wARP+REFMAC, BUCCANEER+REFMAC RESOLVE

$F_A$ estimation

Substructure detection

Phasing

Density Modification

Model Building
AFRO: Multivariate SAD equation for $F_A$ estimation

- $E(|F_A;|F^+|,|F^-|) = \frac{\iiint|F_A|P(|F_A|,\alpha_A,|F^+|,\alpha^+,|F^-|,\alpha^-)d|F_A|d\alpha_Ad\alpha^+d\alpha^-}{\iiint P(|F_A|,\alpha_A,|F^+|,\alpha^+,|F^-|,\alpha^-)d|F_A|d\alpha_Ad\alpha^+d\alpha^-}$

- Direct methods very sensitive to FA values

- The multivariate FA estimation leads to more substructures being determined than $\Delta F = ||F^+| - |F^-||$
Crunch2: substructure detection

- Algebraic approach based on rank reduction of Karle/Hauptman matrices
- Includes higher order collections of reflections, not just triplets
- Standalone via crank interface
Scoring the substructure solution

- **CRUNCH2**: $\text{FOM} > 1.0$
- **SHELXD**: $\text{CC}_{\text{weak}} > 30\%$
- Both conservative criteria. Solutions may be missed

- **BP3** can be run in “PHAS” mode to verify
BP3: substructure refinement

- Multivariate substructure refinement
  - Includes effect of model and measurement errors
- SAD, MAD, SIR(AS) and MIR(AS)
- Refines atomic and error parameters
- Outputs H-L coeffs, FOM and PHIB
- Normal and fast phasing (PHAS)
Solution validation

- FOM
- SAD - Anomalous Luzatti Parameter
  - > 0.7 strong solution
  - 0.3 unlikely
- Build
Density Modification

- Map improvement
  - SOLOMON, (SHELXE)
  - PIRATE – statistical in reciprocal space
  - PARROT – MLHL target function, pairwise NCS
- SAD Hand selection
  - SOLOMON/SHELXE:
    - contract local map density
      - higher better
  - PARROT:
    - sigmaa value – higher better
      (DM: Real Space Free R, lower better)
- Tolerences about 0.05
Refinement - Refmac

- Relevant refinement targets
  - SAD (and SIRAS coming)
    - Greatly improves maps
  - MLHL
    - Prior phase information used indirectly in the form of H-L coefficients
    - Assume prior phase information is independent of refined phases
    - Generally applicable
Model Building - Buccaneer

- Statistical model building. Likelihood function based on conserved density
  - Requires phase probabilities
- **Finding Growing**: Cα environment
- **Sequencing**: Cβ environment

![Chemical structures of ALA, CYS, HIS, MET, THR]
GerE SAD – a challenging example solved by default

- Example data with CCP4. Originally MAD+native
- 2.7Å SAD peak data with 12 Se (F`-3.6, F” 6.0)
- Early Crank version failed
- Crank v1.3 builds 70% by default
- Crank v1.4 builds 93%, and 70% at inflection point.
Phaser

- Multivariate SAD target
- Refines coordinates, temperature, occupancy
- Iterative atom location
  - Improves substructure and sensitivity
- Maps
  - Electron density
  - Log likelihood gradient (by atom type)
Phaser – log likelihood gradient

- Peak height dependent on f'/f'' ratio
  - May distinguish atom types
  - Can be fooled if normal scatter present at site
- Very sensitive to detail
  - Image shows anisotropic
MR-SAD

- Input: partial structure/MR solution
- Possible atom types

Better maps

Heavy atom parameters

$H^+, H^-$ estimates

SAD-function
Other Programs

- Acorn
  - Direct method based density modification
  - Free lunch
- OASIS
  - SAD ambiguity breaking
- Venerable
  - RANTAN, MLphare
- External
  - SHELX, HKL2MAP
  - Autosharp