Introduction to experimental phasing and Structure solution with Crank

Biophysical Structural Chemistry,
Leiden University, The Netherlands

http://www.bfsc.leidenuniv.nl/software/crank/
http://www.ccp4.ac.uk/

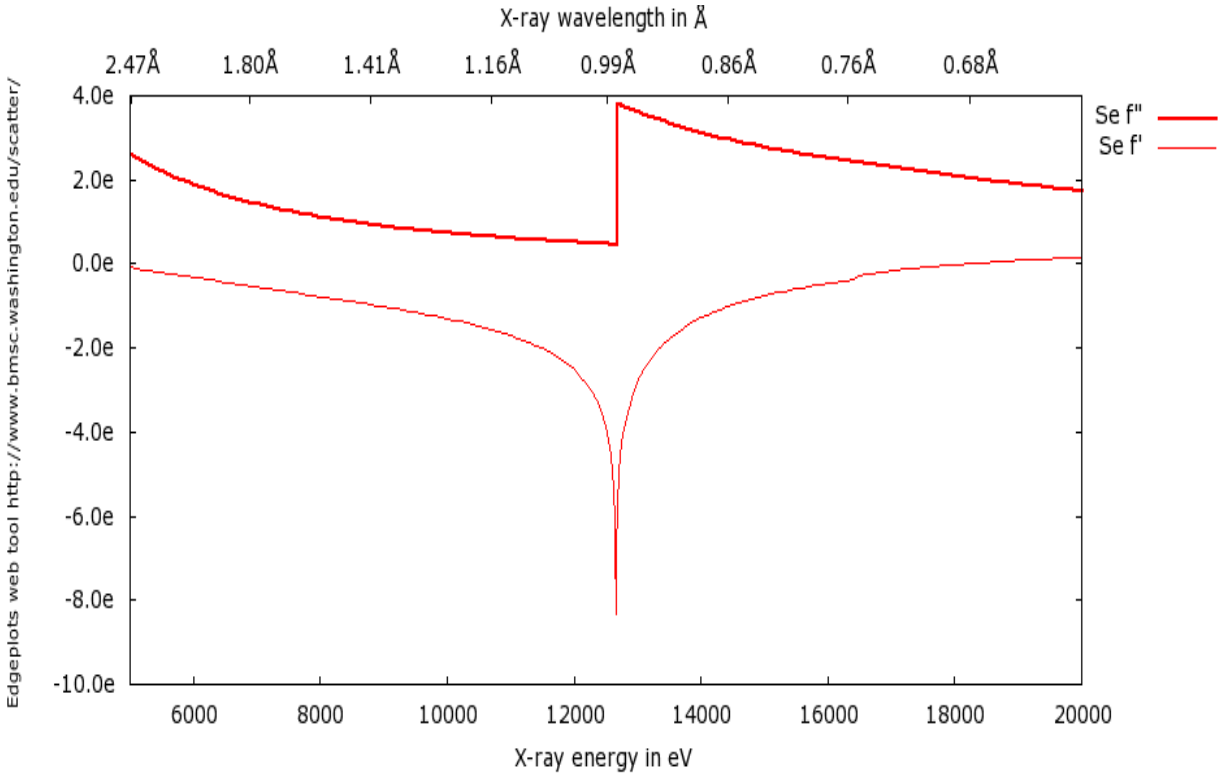
Solving phases: using the anomalous signal

- Anomalous scattering: X-rays are scattered with a phase shift caused by their absorption by electrons
- The strength of the anomalous effect depends on the atom type and X-ray wavelength
- Anomalous scattering causes very small changes in intensity between F(+) and F(-) through addition of f” phase shift term to the atomic scattering factor:

$$\boldsymbol{F^{+}(h)} = \sum (f_j + f_j' + i\, f_j'') \exp(2\pi i\, h \cdot x_j)$$

$$\boldsymbol{F^{-}(h)} = \sum (f_j + f_j' - i\, f_j'') \exp(2\pi i\, h \cdot x_j)$$

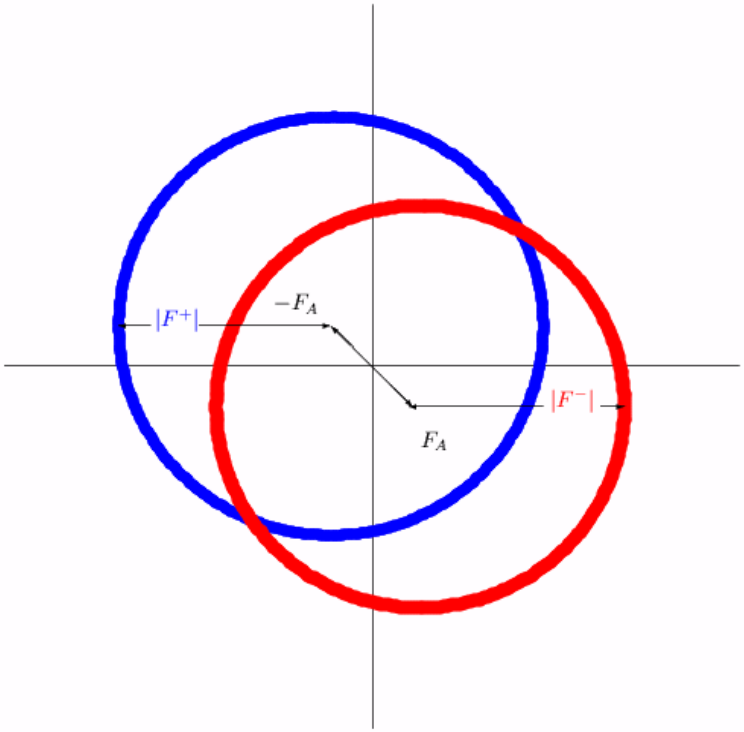
Wavelength dependency of anomalous scattering



$$\boldsymbol{F^{+}(h)} = \sum (f_j + f_j' + i\, f_j'') \exp(2\pi i\, h \cdot x_j)$$

$$\boldsymbol{F^{-}(h)} = \sum (f_j + f_j' - i\, f_j'') \exp(2\pi i\, h \cdot x_j)$$

Single-wavelength Anomalous Diffraction



Solving structures using Friedel pairs collected at one wavelength from a crystal that contains an anomalously scattering substructure.

$F_A = \sum_j f''_j \exp(2\pi i h \cdot x_j)$

$F = F^+ - F_A$, $F = F^- + F_A$

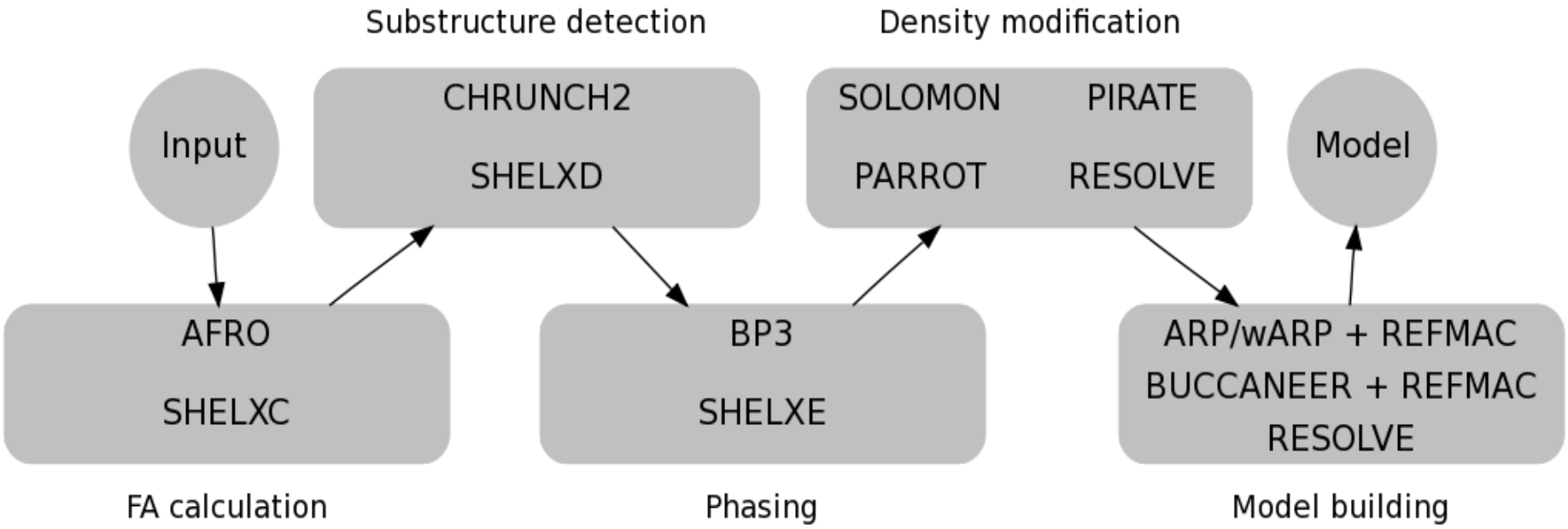
MAD – multiple wavelength anomalous diffraction

- Collect diffraction data at multiple wavelengths, significantly differing by their anomalous scattering coefficients
- Multiple measurements can resolve the phase ambiguity and significantly improve phase estimates
- At synchrotron sites, the wavelength is tunable
- Radiation damage becomes an even greater danger for MAD, as the same crystal is exposed for longer periods

Crank

- Crank is a software for automated structure solution using SAD, MAD or SIRAS data.
- It requires minimal input, but is highly configurable.
- User friendly gui/pipelines incorporating our latest developments in substructure detection, phasing, density modification and model building & refinement as well as plugins to externally developed programs.

Structure solution from experimental phases with Crank.



Other Crank developers

Navraj Pannu

Willem-Jan Waterreus

RAG de Graaff

Finding heavy atom sites

Patterson methods:

By examining the map of interatomic distances, heavy atoms can be located. A Patterson map can be calculated without any phases.

Direct methods:

Use mathematical methods from small molecule crystallography that exploit relations between structure factors

Hybrid (Patterson/Direct methods)

Combination of both - both SHELX and CRUNCH2

F_A estimation

- Direct methods need to first estimate "F_A" or "substructure factor amplitude"
- The programs are very sensitive to F_A values.
- Improving the estimates can improve hit rates of direct methods and solve things that can not previously be solved.
- The simplest estimation of F_A for SAD data is
  $\Delta F = \; | \, |F^+| - |F^-| \, |$
- Available programs in Crank: SHELXC and AFRO

AFRO: Multivariate SAD equation for F_A estimation

- Giacovazzo previously proposed multivariate F_A estimation, with an implementation assuming Friedel phases are equal.
- An equation can be obtained without the equal phase assumption requiring only one numerical integration.
- The multivariate F_A calculation leads to more substructures determined (by default) over ΔF.

CRUNCH2: A program for substructure detection.

- Algebraic approach based on rank reduction of Karle-Hauptman matrices
- Takes into account triplets/tangent formula as well as higher order relationships between reflections
- de Graaff *et al.* (2001) *Acta Cryst.* D57, 1857-1862.

Important parameters in substructure detection

- The number of trial cycles to run.
- The number of atoms to search for.
  - Should be within 10-20% of actual number
  - A first guess uses Matthew's coefficient
- The resolution cut-off:
  - For MAD, a good guess comes from anomalous difference correlation.
  - For SAD, a first guess is high resolution limit + 0.5A

Substructure determination validation

- If substructure coordinates are found, often all positions are correctly determined.
- Indicators of a correct solution:
  - CCweak > 30% in SHELXD
  - FOM > 1.0 in CRUNCH2

(conservative criteria for a correct solution)

- Running BP3 in a fast “Check” mode provides another way to detect a correct solution without running all substructure detection cycles

Substructure refinement and phasing

- Refinement of the substructure parameters and error parameters
- The refined parameters are used for initial estimation of phases
- Available programs in Crank: BP3, SHELXE, REFMAC

Substructure refinement and phasing by BP3 and Refmac

- Can be used for SAD, SIR(AS), MAD (BP3 only)
- Using multivariate SAD phasing function
- Outputs the “best” phases PHIB, the estimate of their reliability FOM and four Hendrickson-Lattman (HL) coefficients “encoding” the phase distribution to MTZ file
- Two “modes” of operation of BP3: normal and check (fast phasing)
- BP3 is usually faster for a few substructure atoms,

  REFMAC for tens or hundreds substructure atoms (FFT vs direct summation)

SAD functions in heavy atom refinement before BP3

- Earlier heavy atom refinement programs use a Gaussian (or least squares) function in Friedel differences ($\Delta F = |F^{+}|- |F^{-}|$) (North, 1965), (Matthews, 1966).
- The calculated Friedel difference is determined based on a assumed value of F and α and the heavy atom structure factor model.

Multivariate distribution for a SAD experiment

- Include effect of model and measurement errors and correlation between observed and calculated Friedel pairs.
- Required multivariate joint probability distribution

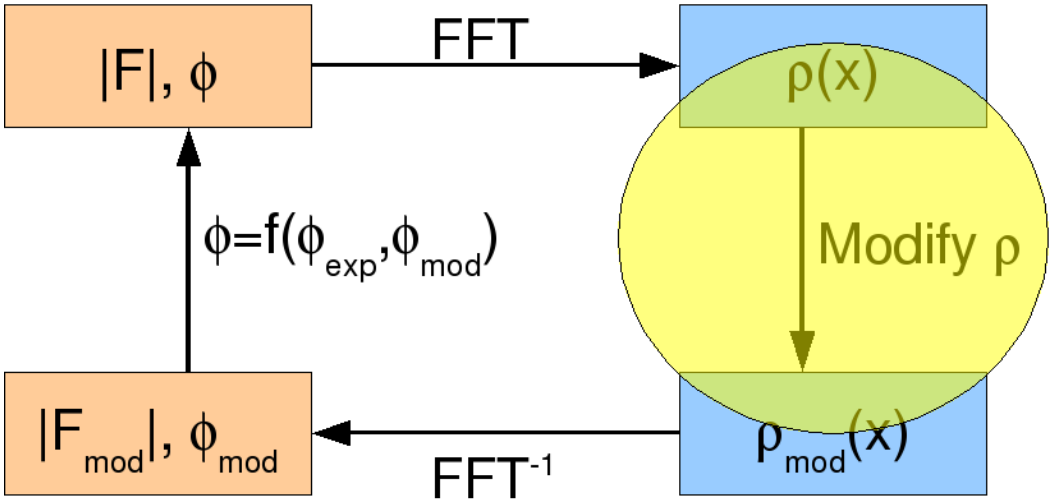
$$P_{ph}\left(F_o^+, F_o^- \middle| F_H^+, \varphi_H^+, F_H^-, \varphi_H^-\right)$$

- The function can be further extended for phase combination and refinement in density modification and model building stages

After we have initial phases: Density modification

- Improving phases of the initial map by incorporation of prior information about protein map features into it
- Available programs in Crank:
  - Parrot
  - Solomon
  - SHELXE
- In case of SAD or SIRAS, Parrot and Solomon can use their built-in phase combination or external multivariate combination by Multicomb or Refmac

Density modification



Traditional density modification techniques:

- Solvent flattening
- Histogram matching
- Non-crystallographic symmetry (NCS) averaging

Assessing quality of the map

- Statistics from substructure phasing:
  - FOM (0.3 or higher usually leads to a solution, 0.15 or lower usually not)
  - skewness of the map and related statistics - outputted by MAPRO utility
- Statistics from density modification:
  - FOM (only useful if bias reduction was employed!)
  - “contrast” of the map
- Does it look like a protein? (visualization)

Automatic hand determination

- The hand is not known - either the found substructure or its inverse is correct
- Lazy but slow approach: try building with both hands
- Crank chooses the hand before building, assuming that a correctly handed substructure provides a better map
- Criteria used: combination of skewness of the map after phasing with correlation.contrast after “fast” density modification
- In tests on almost 150 datasets, the wrong hand is chosen in 2 cases (none of which can be built due to very weak anomalous signal)

Detection of additional anomalous scatterers

- The substructure found by SHELX or CRUNCH2 may be incomplete or smaller anomalous scatterers are present (ions, sulphurs)
- Crank tries to find any additional sites from difference anomalous maps
- For SAD, likelihood gradient maps are available in BP3 and REFMAC, providing slightly better distinction between correct and false peaks than the traditional anomalous difference maps

Further improving the map

- Adjusting solvent content can improve the map after density modification. (Since the number of monomers is usually not known beforehand, neither is the solvent content.)
- If BP3 was run in fast mode, or SHELXE was run, a better map may result if BP3 is run in “default” mode.
- Try to determine NCS manually - occasionally, the automatic NCS detection in Parrot does not succeed
- Try to find additional anomalous scatterers (the peak picking treshold in Crank is conservative)

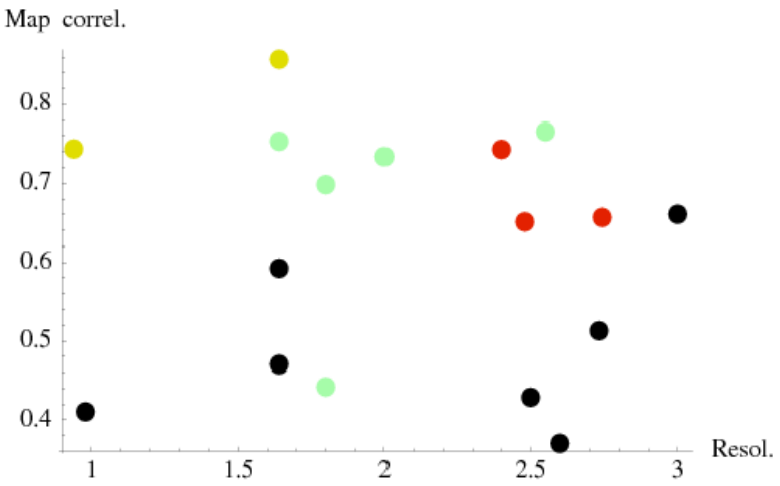
Automatic model building

- Available programs in Crank:
  - ARP/wARP (after separate installation)
  - Buccaneer (comes with CCP4)
- in both cases, model building is iterated with model refinement by Refmac

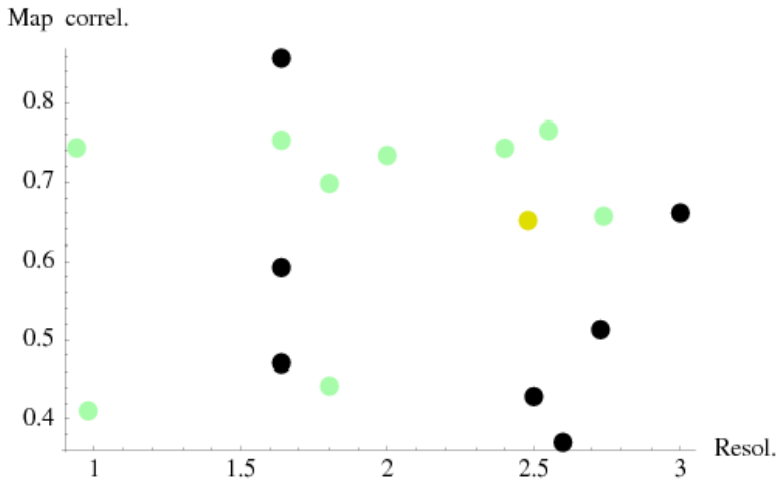
SAD and SIRAS functions in model refinement

- Refinement functions in REFMAC:
  - Rice: no prior phase information
  - MLHL: incorporating phase information from phasing indirectly via HL coefficients - general but assumes independence of phase information from model
  - SAD, SIRAS: phase information derived directly using all actually available information
  - SADDM: SAD function taking into account information from actual density modified map

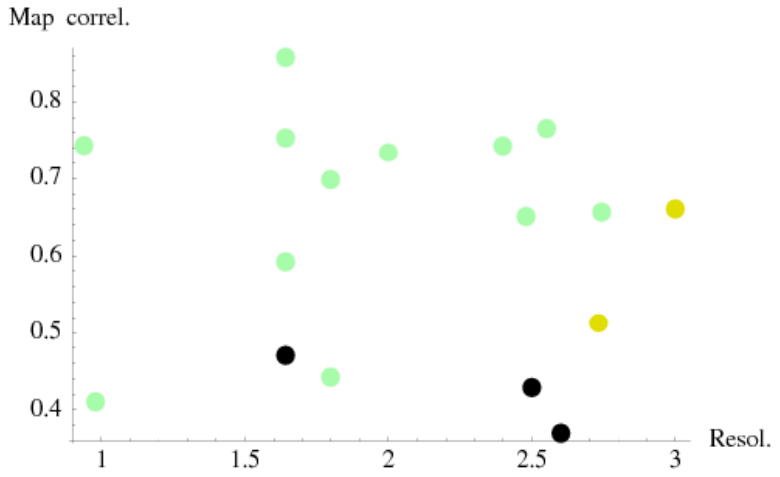
Results from SAD function



Rice



MLHL function



SAD function

Green: 80 – 100% built

Yellow: 50 – 80% built

Red: 20 – 50% built

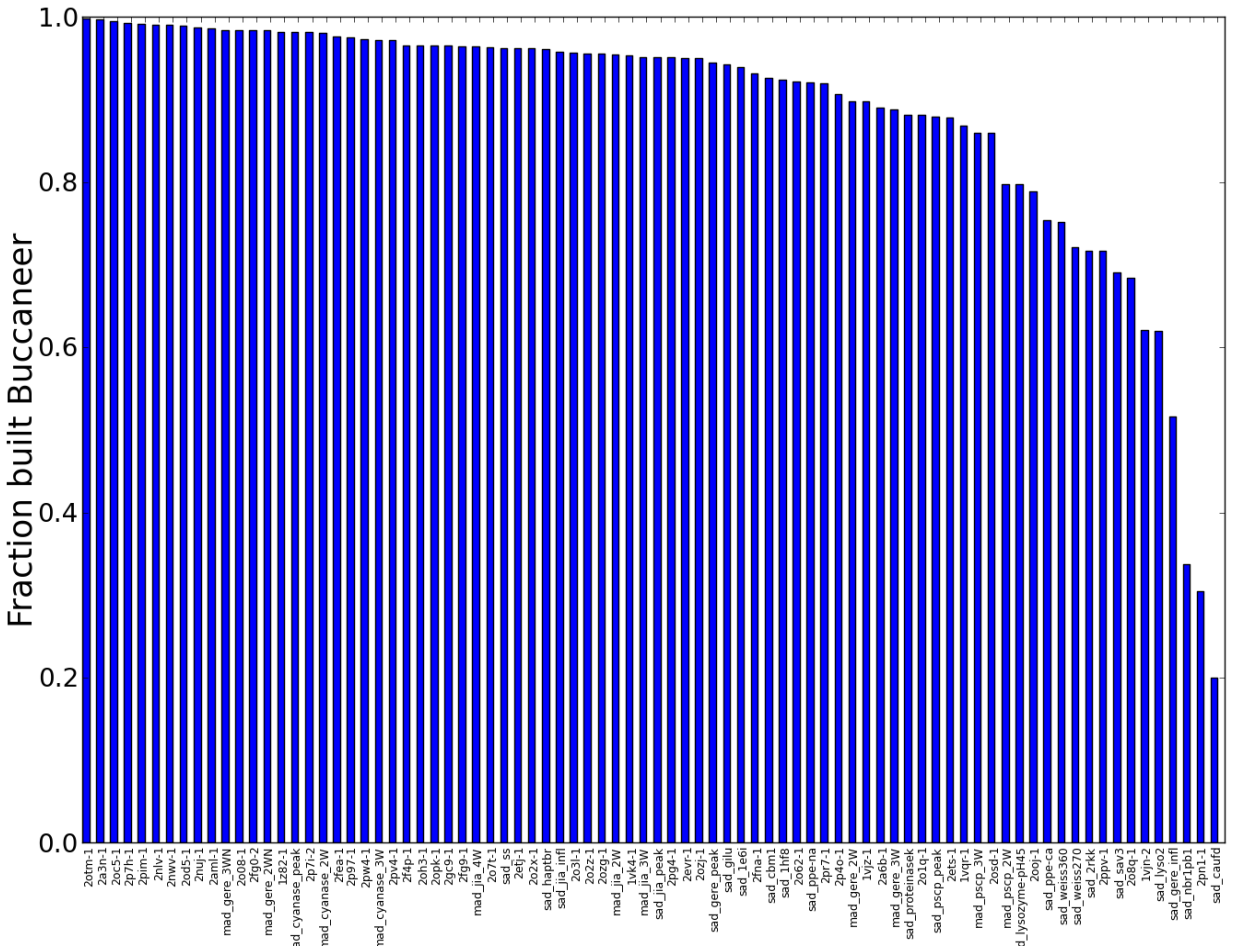
Black: 0 - 20% built

Assessing robustness of Crank

- A test system has been built of over 150 SAD, SIRAS and MAD data sets with a range of phasing quality and resolution.

- Over 10% are not solvable by data sets authors.

Crank pipeline in default mode: afro/crunch2/bp3/solomon/buccaneer



A challenging problem solved by default: GerE with SAD data

- GerE data set is distributed with CCP4 and originally solved by MAD + native.
- 2.75 Angstrom data with 12 seleniums
- Could not be built with earlier Crank versions.
- Crank version 1.4 builds over 90% using peak SAD data and over 70% using inflection point SAD data.
- Crank version 1.5 (not released yet) builds over 90% using either of the SAD datasets.

Conclusions/Remarks

- With a sufficient anomalous signal and resolution, structures can be solved automatically.
- When structures can not, first determine which step has failed: Crank attempts to make re-running steps easier.

Availability & Documentation

- Crank works under Linux, MAC OS, Windows and is free software.
- Crank is available in CCP4 version 6.2.x
- ***Please* use version 1.4 (or higher)!**
- Crank wiki page is available:
  - http://ccp4wiki.org/
  - tested on undergraduates with no previous knowledge of crystallography/phasing

Acknowledgements

- All dataset contributors (JCSG, SSGCID, Z. Dauter, M.Weiss, A.Sharma)
- Garib Murshudov, Kevin Cowtan, George Sheldrick, Victor Lamzin, Charles Ballard, Francois Remacle, Peter Briggs, Norman Stein, Martyn Winn
- http://www.bfsc.leidenuniv.nl/software/crank/

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Cyttron

CCP4