

Crank and Databases



Steven Ness

Leiden University
The Netherlands

Crank

- Crank - Suite for automated structure solution
- Simple design – XML based
 - Input, Run, Output
- Designed to:
 - Teach beginners
 - Enable experts
- Variety of user interfaces
- Arbitrary user-designed pipelines
- Visualization and database storage of results
- High throughput tools for the individual scientist
- Working on adding Grid support to Crank

CRANK

User Interface	E/FA value calculation	Substructure Determination	Substructure Refinement and Phasing	Density Modification	Model Building/Refinement
CCP4i	AFRO	CRUNCH2	BP3	SOLOMON	RESOLVE
Web	DREAR	SOLVE	SHARP	RESOLVE	FFFEAR
Script	SHELXC	SHELXD	MLPHARE	SHELXE	MAID
XML	ECALC	RANTAN		DM	REFMAC
				PIRATE	ARP/wARP*
Validation	Viewing	Tools			
PROCHECK	CCP4mg	Emma	Truncate		
SFCHECK	O	FHSCAL	CAD		
	Xfit	Scaleit	SFTOOLS		
	PyMol				

CRANK

Calculate b-factor and solvent content from the Wilson plot - First dataset in the first crystal will be used

CRANK Title: E.coli Thioesterase - Crank

Input: Experimental Columns Other Columns Substructure PDB Model Rfree Sequence

MTZ in: sness | jiaMAD.mtz

Crystal # 1

Native | I: IMEAN_jiaPEAK | SIGI: SIGIMEAN_jiaPEAK

Crystal # 2

Native Substructure Atom Se Number of Substructure Atoms Input XML Substructure

Dataset : 1 Type: Peak Anomalous | Data collected at CuKalpha wavelength (1.54A)

f' -3.6 f'' 5.4 Wavelength 0.9789

IP+ I_jiaPEAK(+) | SIGIP+ SIGI_jiaPEAK(+)

IP- I_jiaPEAK(-) | SIGIP- SIGI_jiaPEAK(-)

Dataset : 2 Type: Inflection Anomalous | Data collected at CuKalpha wavelength (1.54A)

f' -5.6 f'' 3.3 Wavelength 0.9793

IP+ I_jiaINFL(-) | SIGIP+ SIGI_jiaINFL(-)

IP- I_jiaINFL(-) | SIGIP- SIGI_jiaINFL(-)

MTZ out: sness | jiaMAD_crank_1.mtz

Required parameters

Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: 0

B-factor 39.105 Solvent Content 0.622 Calculate B and Solv. Content

Experiment

Program to run or Decision to take: Program : AFRO

Step # 1

SCALEIT will be used for scaling. SCALEIT will scale *all* derivatives/wavelengths to the Reference

Input Experimental Columns: INPUT

Reference dataset: XTAL1_SAD (1,1)

Scale over resolution range to

Refine: scale & anisotropic Bfactor and apply Wilson scaling

Weight observations by their SDs

Converge after cycles of refinement

Converge if absolute shift < SDs of parameter

Convergence tolerance 10 to power

Run Save or Restore Close

Types of input data

Experimental
Data input

Required Parameters

Pipeline of programs

Crank database

- 3_crank/workdb
- Stores all information needed by each step
- Currently a directory with files
- File name encodes
 - Program “step”
 - Type of data
 - e.g. “crank.out.3_BP3.mtz” or
“crank.in.2_CRUNCH2.coords.xml”

MTZ column labels

- Symbolic column names
- All CCP4i user input column labels are renamed to avoid known problems (e.g. CAD/SFTOOLS)
- Examples
 - INPUT1_X1_D2_F_PLUS
 - 1_AFRO_F_COLUMNS_F
 - 3_BP3_PHASE_COLUMNS_PHIB
- This also works for other kinds of user input columns from the CCP4i interface

Other types of input data

- Sequence
- Substructure
- List of Substructures
- Protein Model
- List of Protein Models
- Map
- Rfree Column
- Many more to be added

Crank XML

- Generated either directly by programs or by wrappers to convert logfiles to XML
- Stores all information generated by programs
- Main purpose : Decisions
 - These are the way that the user can direct program/information flow in their pipeline
- Secondary purpose : Data mining

Our Needs

- Way to access any given column in an MTZ file
- Storage of
 - Sequence, Substructure, Protein Models, Maps, Rfree columns, many more types.
- Access via
 - API (Python, Tcl, C, C++)
 - Filesystem

Acknowledgements and Program availability:

Navraj Pannu

RAG de Graaff

Pavol Skubak

Irakli Sikharulidze

Jan Pieter Abrahams



<http://www.bfsc.leidenuniv.nl/software/crank>