Crank and Databases

Steven Ness

Leiden University

The Netherlands

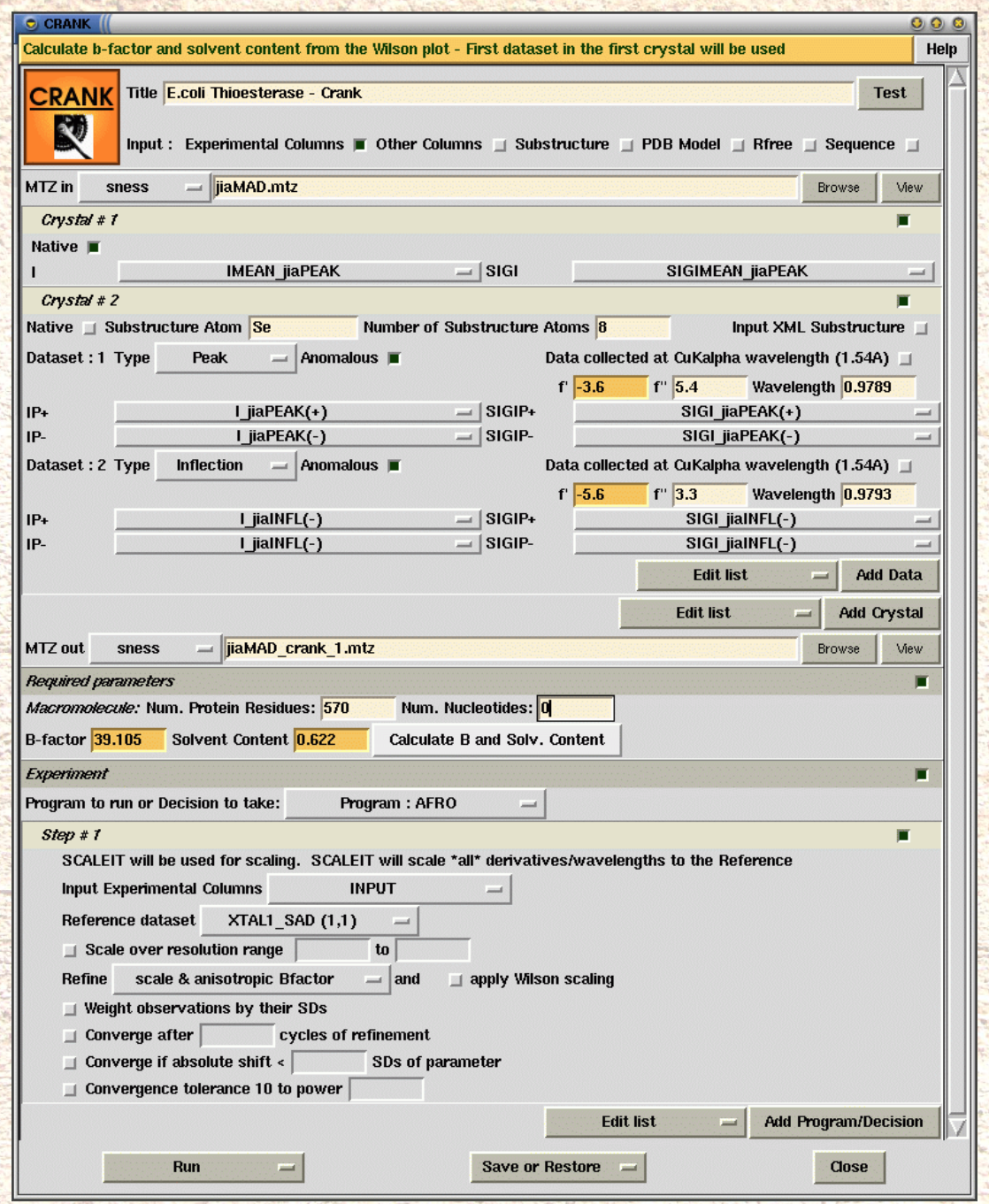
Crank

- 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		



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 frlow 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