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



User Interface	E/FA valu calculatio		ture Re	ostructure finement d Phasing	Density Modification	Model Building/ Refinement
CCP4i	AFRO	CRUNCH	2 E	BP3	SOLOMON	RESOLVE
Web	DREAR	SOLVE	Sŀ	IARP	RESOLVE	FFFEAR
Script	SHELXC	SHELXD	MLF	HARE	SHELXE	MAID
XML	ECALC	RANTAN			DM	REFMAC
					PIRATE	ARP/wARP*
Validatio	on	Viewing	Тс	ols		
PROCHECK		CCP4mg	Emma	Truncate		
SFCHECK		0	FHSCAL	CAD		
		Xfit PyMol	Scaleit	SFTOOLS		

Crystal # 1 Image: Crystal # 1 Image: Crystal # 2 I Image: I	∋ CRANK ((Calculate b-	factor and solvent content from the Wilson plot - First dataset in the first crystal will be use		<mark>) (</mark> H
Input : Experimental Columns © Other Columns Substructure PDB Model Rfree Sequence TZ in sness BaMAD.ntz Browse Vev Crystel # 1 Netive # 1 MEAN jaPEAK SIGI SIGIMEAN jaPEAK Crystel # 2 Native Substructure Atom Se Number of Substructure Atoms 8 hput XML Substructure P Dataset : 1 Type Peak Anomalous © Data collected at OuKalpha wavelength (1.54A) p. LjaPEAK(-) SIGIP Dataset : 2 Type Inflection Anomalous © Data collected at OuKalpha wavelength (1.54A) p. LjaPEAK(-) SIGIP Dataset : 2 Type Inflection Anomalous © Data collected at OuKalpha wavelength (1.54A) p. LjaINFL(-) SIGIP Dataset : 2 Type Inflection Anomalous © Data collected at OuKalpha wavelength (1.54A) p. LjaINFL(-) SIGIP F S.6 (* 3.3 Wavelength 0.3793 P JaINFL(-) SIGIP Edit list Add Data Edit list Add Data F S.6 (* 3.1 Wavelength 0.3793 Solvent Content 0.522 Calculate B and Solv. Content Stop # 1 SIGI jaINFL(-) SiGI jaINFL(-) SiGI jaINFL(-) Solvent Content 0.522 Calculate B and Solv. Content Stop # 1 Solvent Content 0.522 Solvent Content 0.522 Solvent Content 0.522 Solvent Content 0.522 Calculate B and Solv. Content Stop # 1 Solvent Content 0.522 Solvent Content 0.525 Solvent Content 0.				
MTZ in sness jiaMAD.mtz Browse Mew Crystal # 1 IMEAN_jiaPEAK SIGI SIGIMEAN_jiaPEAK I Native # IMEAN_jiaPEAK SIGI SIGIMEAN_jiaPEAK I Orystal # 2 Imean jiaPEAK SIGI SIGIMEAN_jiaPEAK Imput XML Substructure Dataset : 1 Type Peak Anomalous Imput XML Substructure Anomalous Imput XML Substructure Imput XML	CRANK	Title E.coli Thioesterase - Crank	Test	
Cryski # 1 IMEAN_jiaPEAK SIGI SIGIMEAN_jiaPEAK Native J IMEAN_jiaPEAK SIGI SIGIMEAN_jiaPEAK Cryski # 2 Imput XML Substructure Atoms 8 Input XML Substructure Atoms 9 Input XML Substructure Atoms 9 Input XML Substructure Atoms 9 Data collected at CuKalpha wavelength (1.54A) Imput XML Substructure Atoms 9 Data collected at CuKalpha wavelength (1.54A) Imput XML Substructure Atoms 9 Imput XML Substructure JiaPEAK(-) SIGIPA SIGI jiaPEAK(-) SIGI jiaPEAK(-) Data collected at CuKalpha wavelength (1.54A) Imput XML Substructure Atoms 9 Data collected at CuKalpha wavelength (1.54A) Imput XML Substructure Atoms 9 Data collected at CuKalpha wavelength (1.54A) Imput XML Substructure Atoms 9 Data collected at CuKalpha wavelength (1.54A) Imput XML Substructure Atoms 9 IpiaINFL(-) SIGIPA SIGI jiaINFL(-) SIGI jiaINFL(-) SIGI jiaINFL(-) Pre- IjiaINFL(-) SIGIPA SIGI jiaINFL(-) SIGI jiaINFL(-) SIGI jiaINFL(-) SIGIPA SIGI jiaINFL(-) SIGIPA SIGI jiaINFL(-) Pactor 33.105 Solvent Content 0.622 Calcutate B and Solv. Content Scale Atoms 9 Scale Atoms 9 Scale Atoms 9	X	Input : Experimental Columns 🔳 Other Columns 🔟 Substructure 🔲 PDB Model 📃 Rf	ree 🔟 Sequence 📋	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 Atom Se Number of Substructure Atoms © Input XML Substructure Paatset : 1 Type Peak Atom Se Number of Substructure Atoms © Input XML Substructure Preak Anomalous # Data collected at Curkalpha wavelength (1.54A) # Crystal # 2 Preak JiaPEAK(+) SIGIP- SIGI jaPEAK(+) SIGIP- Dataset : 2 Type Inflection Anomalous # Data collected at Curkalpha wavelength (1.54A) # Crystal # 2 Preak JiaPEAK(-) SIGIP- Dataset : 2 Type Inflection Anomalous # Data collected at Curkalpha wavelength (1.54A) # Pre JiaPEAK(-) SIGIP- SIGI jaINFL(-) SIGIP- SIGI jaINFL(-) SIGIP- SIGI jaINFL(-) SIGIP- Edit list Add Crystal Edit list Add Crystal MIZ out sness jiaMAD_crank_1.mtz Browse Vev Required parameters # Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: 0 D-factor 39.105 Solvent Content 0.622 Calculate B and Solv. Content Experiment # SCALEIT will be used for scaling. SCALEIT will scale *all* derivatives/vavelengths to the Reference Input Experimental Columns INPUT Reference dataset XTAL1_SAD (1,1) Scale over resolution range to Refine scale & anisotropic Bfactor SD of refinement Converge after cycles of refinement Converge after cycles of refinement Converge if absolute shift < SD of anameter Converge of the scale 10 to power Edit list Add Program/Decision	MTZ in	sness 🖵 jiaMAD.mtz	Browse	w
1 IMEAN jiaPEAK SIGI SIGIMEAN jiaPEAK Crysted # 2 Imput XML Substructure Atom Se Number of Substructure Atom Se Input XML Substructure P Dataset : 1 Type Peak Anomaious Impact Atom Se Input XML Substructure P Dataset : 1 Type Peak Anomaious Impact Atom Se Input XML Substructure P IP+ I jiaPEAK(-) SIGIP SIGI jiaPEAK(-) SIGI jiaPEAK(-) Dataset : 2 Type Inflection Anomaious Impact Atom Atom Atom Atom Atom Atom Atom Ato	Crystal #	1) III (
Crystel # 2 Native Substructure Atom Se Number of Substructure Atoms 8 Input XML Substructure 2 Dataset : 1 Type Peak Anomalous T Data collected at CuKalpha wavelength (1.54A) P+ LjiaPEAK(+) SIGIP- SIGI jaPEAK(-) SIGIP- Dataset : 2 Type Inflection Anomalous T P+ LjiaINFL(-) SIGIP- Data collected at CuKalpha wavelength (1.54A) r * * SiGI jaPEAK(-) Dataset : 2 Type Inflection Anomalous * P+ LjiaINFL(-) SIGIP- Data collected at CuKalpha wavelength (1.54A) r * * SiGI jaINFL(-) Data collected at CuKalpha wavelength (1.54A) r * * SiGIP- Bata collected at CuKalpha wavelength (1.54A) r * * SiGIP- SiGI jaINFL(-) SIGIP- * SiGI jaiNFL(-) * * Add Orystal #TZ out sness jiaMAD_crank_1.mtz Browrae Vew Beopuired parameters	Native 🔳			
Halive Substructure Atom Se Number of Substructure Atoms B Input XML Substructure Dataset : 1 Type Peak Anomalous Data collected at CuKalpha wavelength (1.54A) P+ JjaPEAK(+) SIGIP- P- IjaPEAK(-) SIGIP- Data collected at CuKalpha wavelength (1.54A) P+ JjaINFL(-) SIGIP- SIGI jiaINFL(-) SIGIP- SIGI jiaINFL(-) SIGIP- Edit list Add Oxystal MTZ out sness jiaMAD_crank_1.mtz Browse Vew Required parameters Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: Stap # 1 SCALEIT will be used for scaling. SCALEIT will scale *all* derivatives/wavelengths to the Reference Input Experimental Columns INPUT Reference dataset XTALI_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 Converge if absolute shift < SDs of parameter	1	IMEAN_jiaPEAK SIGI SIGIMEAN_jia	PEAK -	-
Dataset : 1 Type Peak Anomalous Data collected at CuKalpha wavelength (1.54A) P+ LjiaPEAK(+) \$IGI_jiaPEAK(+) \$IGI_jiaPEAK(+) Data collected at CuKalpha wavelength (1.54A) r *.4 Wavelength (1.54A) P+ LjiaPEAK(-) SIGI_P Data collected at CuKalpha wavelength (1.54A) P+ LjiaINFL(-) SIGIP+ P- LjiaINFL(-) SIGIP+ SIGI_jiaINFL(-) SIGI_jiaINFL(-) P- LjiaINFL(-) SIGIP+ SIGI_jiaINFL(-) P- LjiaINFL(-) SIGIP+ SIGI_jiaINFL(-) Browse View Mitz out sness J-actor 33.105 Solvent Content 0.622 Calculate B and Solv. Content Experiment 0 Stop #1 <td< td=""><td>Crystal #</td><td>2</td><td>) III (</td><td></td></td<>	Crystal #	2) III (
P+ LjiaPEAK(+) \$IGP JiaPEAK(-) \$IGP SIGP JiaPEAK(-) \$IGP SIGI jiaPEAK(+) Dataset : 2 Type Inflection Anomalous P+ LjiaINFL(-) \$IGP Data collected at CuKalpha wavelength (1.54A) f f 5.6 f' 3.3 Wavelength (0.5793 F' 5.6 P+ LjiaINFL(-) \$IGP JiaINFL(-) \$IGP SIGI jaINFL(-) \$IGI jaINFL(-) Edit list Add Crystal Macromolecule: Num. Nucleotides: @acromolecule: Scale @acromolecule: Num. Nucleotides: @acromolecule: Scale @acromolecule: Num. Nucleotides:	Native 🔲 S	Substructure Atom Se Number of Substructure Atoms 8 Input	XML Substructure	
P+ 1 jiaPEAK(+) SIGIP+ P- 1 jiaPEAK(-) SIGIP- Dataset : 2 Type Inflection Anomalous P+ 1 jiaINFL(-) SIGIP- Dataset : 2 Type Inflection Anomalous P+ 1 jiaINFL(-) SIGIP- Data collected at CuKalpha wavelength (1.54A) f r 5.6 r' 3.3 Wavelength (0.9733) SIGI jiaINFL(-) P+ 1 jiaINFL(-) SIGIP- SIGI jiaINFL(-) SIGIP- Edit list Add Data Edit list Add Crystal MTZ out sness #Convecter Imm. Nucleotides: Ø	Dataset : 1	Type Peak — Anomalous 🔳 Data collected at CuKalpha way	velength (1.54A) 📋	
P. L jaPEAK(-) SIGIP SIGIP SIGI jaPEAK(-) Dataset : 2 Type Inflection Anomalous Data collected at CuKalpha wavelength (1.54A) P+ L jiaINFL(-) SIGIP- SIGIP- SIGI jiaINFL(-) P- L jiaINFL(-) SIGIP- SIGI jiaINFL(-) Edit list Add Data IP- L jiaINFL(-) SIGIP- SIGIP- SIGI jiaINFL(-) Edit list Add Data IP- L jiaINFL(-) SIGIP-		f' <mark>-3.6 f'' 5.4</mark>	avelength 0.9789	
Dataset : 2 Type inflection Anomalous Data collected at CutKalpha wavelength (1.54A) P- LjiaINFL(-) SIGIP- P- LjiaINFL(-) SIGIP- Figure 1 JiaINFL(-) SIGIP- Edit list Add Data Edit list Add Data Edit list Add Crystal WTZ out sness JjiaMAD_crank_1.mtz Browse Vew Required parameters Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: B-factor 39.105 Solvent Content 0.622 Calculate B and Solv. Content Experiment Program to run or Decision to take: Program : AFRO Stap # 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 if absolute shift < SDs of parameter Edit list Add Program/Decision	IP+	I_jiaPEAK(+)SIGIP+SIGI_jiaPEA	K(+)	-
P+	P-	I_jiaPEAK(-)	K(-)	-
P+ LjialNFL(-) SIGIP+ SIGIP+ SIGI jialNFL(-) P- LjialNFL(-) SIGIP- SIGIP- SIGI jialNFL(-) Edit list Add Data Edit list Add Crystal WTZ out sness jiaMAD_crank_1.mtz Prowse View Required parameters Image: State of the	Dataset : 2	Type Inflection 😑 Anomalous 🔳 Data collected at CuKalpha way	velength (1.54A) 📋	
p. I jialNFL(-) SIGIP- SIGIP- SIGI jialNFL(-)		f' <mark>-5.6 f'' 3.3 stars</mark> Wa	avelength 0.9793	1
Edit list Add Data Edit list Add Crystal WITZ out sness jjjaMAD_crank_1.mtz Browse View Required parameters Image: State of the state of t	P+	I_jialNFL(-) = SIGIP+ SIGI_jialNF	L(-)	-
Edit list Add Crystal MIZ 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 to run or Decision to take: Program : AFRO Step # 1 SCALEIT will be used for scaling. 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 converge if absolute shift < SDs of parameter	IP-	I_jialNFL(-) — SIGIP- SIGI_jialNF	L(-)	-
MTZ out sness jjaMAD_crank_1.mtz Browse View Required parameters Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: 0 Ba-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. Scale over resolution range to Converge after cycles of refinement Converge if absolute shift < SDs of parameter Convergence tolerance 10 to power Edit list Add Program/Decision		Edit list	💻 🛛 Add Dat	a
Required parameters Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: 0 3-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 Edit list Add Program/Decision		Edit list	Add Crysta	al
Macromolecule: Num. Protein Residues: 570 Num. Nucleotides: 9 3-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 Edit listAdd Program/Decision	MTZ out	sness	Browse Vie	w
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 Edit list Add Program/Decision	Required pa	rameters		
Experiment Image: Seperation of take: Program : AFRO Image: Seperation of take: Program : AFRO Step # 1 Image: Scale for scaling. SCALEIT will be used for scaling. SCALEIT will scale *all* derivatives/wavelengths to the Reference Input Experimental Columns INPUT Image: Scale cover resolution range Image: Scale cover resoluticover range Image: Scale cov	Macromoled	zule: Num. Protein Residues: 570 Num. Nucleotides: 0		
Program to run or Decision to take: Program : AFRO	3-factor 39	.105 Solvent Content 0.622 Calculate B and Solv. Content		
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 weight observations by their SDs Converge after cycles of refinement Converge if absolute shift <	Experiment			
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 Refine scale & anisotropic Bfactor Weight observations by their SDs and Converge after cycles of refinement Converge if absolute shift <	Program to	run or Decision to take: Program : AFRO 💴		
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 Weight observations by their SDs converge after cycles of refinement Converge if absolute shift <	Step # 1			
Input Experimental Columns INPUT		IT will be used for scaling. SCALEIT will scale *all* derivatives/wavelengths to the Referen	nce	
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 <				
Scale over resolution range to Refine scale & anisotropic Bfactor andapply Wilson scaling Weight observations by their SDs Converge after cycles of refinement Converge if absolute shift <		-		
Refine scale & anisotropic Bfactor and apply Wilson scaling Weight observations by their SDs Converge after cycles of refinement Converge if absolute shift <				
Weight observations by their SDs Converge after cycles of refinement Converge if absolute shift <				
Converge after cycles of refinement Converge if absolute shift < SDs of parameter Convergence tolerance 10 to power Edit list Add Program/Decision				
Converge if absolute shift < SDs of parameter Convergence tolerance 10 to power Edit list Add Program/Decision		-		
Convergence tolerance 10 to power Edit list Add Program/Decision				
Edit list Add Program/Decision				
	Con	vergence tolerance 10 to power		
Dum Sava ar Destara				
		Edit list 🛁 🕴	Add Program/Decisio	n

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

N7