MaNDi Software requirements for April 1st 2013

MaNDi is a key part of an upcoming proposal in May 2013 from BSMD to BER/NIH, which seeks to establish a Biomedical Technology Research Center (BTRC) at Oak Ridge National Laboratory (ORNL). A key part of this proposal will be to demonstrate the scientific capabilities of MaNDi. To successfully do this we need several changes to be made to the ISAW data reduction package in order for us to add some preliminary data into the BTRC proposal.

Specifically the following statistics will need to be produced by the data reduction software. (see attached data reduction tables)

1) No of Unique Reflections	(Overall)
2) Resolution range	(Overall)
3) Multiplicity	(Overall and by Resolution Shell)
4) Mean ((I)/sd (I))	(Overall and by Resolution Shell)
5) R_{merge}	(Overall and by Resolution Shell)
6) R_{pim}	(Overall and by Resolution Shell)
7) Data Completeness	(Overall and by Resolution Shell)

Also it would be nice to see the program give us some data on rejected outliers. As a double check we will need to obtain statistics (1 - 7) for each Anger camera and for each orientation. When I did this on TOPAZ data a few years ago I noticed nearly all the reflections from some detectors had very bad statistics.

The Following attachments are included

1) Sample Data reduction tables (I & II)
2) Sample Publication data statistics
3) Draft report on Integration of protein data on TOPAZ (2009)

Sample Data Reduction Table I

N	1/d^2	Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
1	0.0313	5.65	0.027	-	0.027	0.050	255	59719	4632	12.9	2182	41.6	1302	369	120	-0.005	288
2	0.0625	4.00	0.019	0.036	0.021	0.028	549	116500	3772	30.9	5086	40.1	2584	687	159	0.009	590
3	0.0938	3.26	0.028	0.040	0.024	0.030	744	83538	4086	20.4	4288	33.2	3382	868	159	0.014	862
4	0.1251	2.83	0.028	0.050	0.025	0.040	892	36672	1769	20.7	2395	24.8	3939	995	158	0.029	961
5	0.1564	2.53	0.037	0.021	0.026	0.049	1025	20777	1161	17.9	1720	19.3	4399	1110	159	0.020	1084
6	0.1876	2.31	0.051	-	0.028	0.058	1138	13391	1021	13.1	1399	15.6	4796	1196	148	0.020	1299
7	0.2189	2.14	0.074	-	0.030	0.071	1227	9475	1174	8.1	1223	12.8	5061	1277	154	0.045	1349
8	0.2502	2.00	0.101	-	0.032	0.084	1313	6849	994	6.9	1142	10.0	5225	1325	142	0.026	1492
9	0.2814	1.88	0.151	-	0.035	0.122	1367	4159	916	4.5	981	7.2	5252	1332	127	0.014	1435
10	0.3127	1.79	0.241	-	0.037	0.177	1263	2105	713	3.0	814	4.5	4861	1237	105	-0.018	1318
$$																	

">For inline graphs use a Java browser</applet>

Overall:			0.037	0.039	0.037	0.043	9773	26401	2001	13.2	1850	17.0	40801	10396	1431	0.016	10678
	N	1/d^2	Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias

Sample Data Reduction Table II

Completeness and multiplicity, including reflections measured only once

N	1/resol^2	Dmin	Nmeas	Nref	Ncent	%poss	C%poss	MIplct	AnoCmp	AnoFrc	AnoMlt	Rmeas	Rmeas0	(Rsym)	Rpim	RpimO	PCV	PCV0
$$	$$																	
1	0.031	5.65	1418	433	158	94.1	94.1	3.3	90.7	92.4	2.1	0.037	0.059	0.027	0.024	0.031	0.039	0.066
2	0.063	4.00	2781	765	186	99.4	97.5	3.6	94.7	94.7	2.1	0.026	0.037	0.019	0.017	0.018	0.027	0.040
3	0.094	3.26	3643	968	190	99.8	98.5	3.8	95.6	95.6	2.1	0.037	0.045	0.028	0.025	0.022	0.039	0.049
4	0.125	2.83	4253	1115	185	99.5	98.9	3.8	95.7	95.7	2.1	0.037	0.054	0.028	0.024	0.027	0.039	0.060
5	0.156	2.53	4795	1266	188	99.2	99.0	3.8	94.6	94.8	2.1	0.049	0.067	0.037	0.032	0.033	0.053	0.075
6	0.188	2.31	5244	1375	180	99.1	99.0	3.8	94.7	95.1	2.1	0.068	0.085	0.051	0.044	0.042	0.072	0.096
7	0.219	2.14	5568	1472	183	98.8	99.0	3.8	94.4	95.0	2.1	0.099	0.113	0.074	0.065	0.057	0.105	0.129
8	0.250	2.00	5864	1576	178	97.7	98.7	3.7	92.3	93.7	2.0	0.134	0.141	0.101	0.088	0.071	0.144	0.164
9	0.281	1.88	6037	1670	176	97.1	98.5	3.6	89.3	91.1	2.0	0.200	0.209	0.151	0.131	0.106	0.214	0.238
10	0.313	1.79	5671	1633	153	91.8	97.5	3.5	78.9	83.6	2.0	0.321	0.322	0.241	0.210	0.162	0.340	0.360

→ 30Å 2.0

10 →

Over

Sample publication Data

Supplementary material

Table 1. Data Collection and Refinement Statistics

	Neutron diffraction at room temperature: apo *pd*-Toho-1 R274N/R276N	X-ray diffraction at room temperature: apo *pd*-Toho-1 R274N/R276N
Data collection		
Unit-cell parameters (Å)	a = 72.92, b = 72.92, c = 98.53 α = β = 90° and γ = 120°	a = 73.50, b = 73.50, c = 99.33 α = β = 90° and γ = 120°
Space group	$P3_221$	$P3_221$
No. of unique reflections	14,991	15,962
Resolution range (Å)	63.15-2.10 (2.21-2.10)	29.54-2.20 (2.32-2.20)
Multiplicity	7.2 (6.2)	7.5 (7.5)
Mean ((I)/sd(I))	10.1 (8.2)	31.3 (22.0)
R_{merge}[a] (%)	17.6 (21.0)	4.9 (8.3)
R_{pim}[a]	5.7 (7.6)	1.9
Data completeness (%)	83.7 (59.4)	98.6 (97.9)
Crystallographic refinement		
R_{factor}[b] (%)	22.5	13.4
R_{Free}[b] (%)	25.9	17.2
$RMSD_{Bonds}$[c] (Å)	0.004	0.009
$RMSD_{Angles}$[c] (°)	0.667	1.122
Atoms (nonhydrogen)	4,151	2,163
Solvent molecules	71	154
Hydrogen atoms	37	
Deuterium atoms	2,079	
Ramachandran plot[d]		
Outliers (%)	0.4	0.0
Favored (%)	95.7	98.1
Rotamer Outliers (%)	2.4	1.4

Highest resolution shell is shown in parentheses

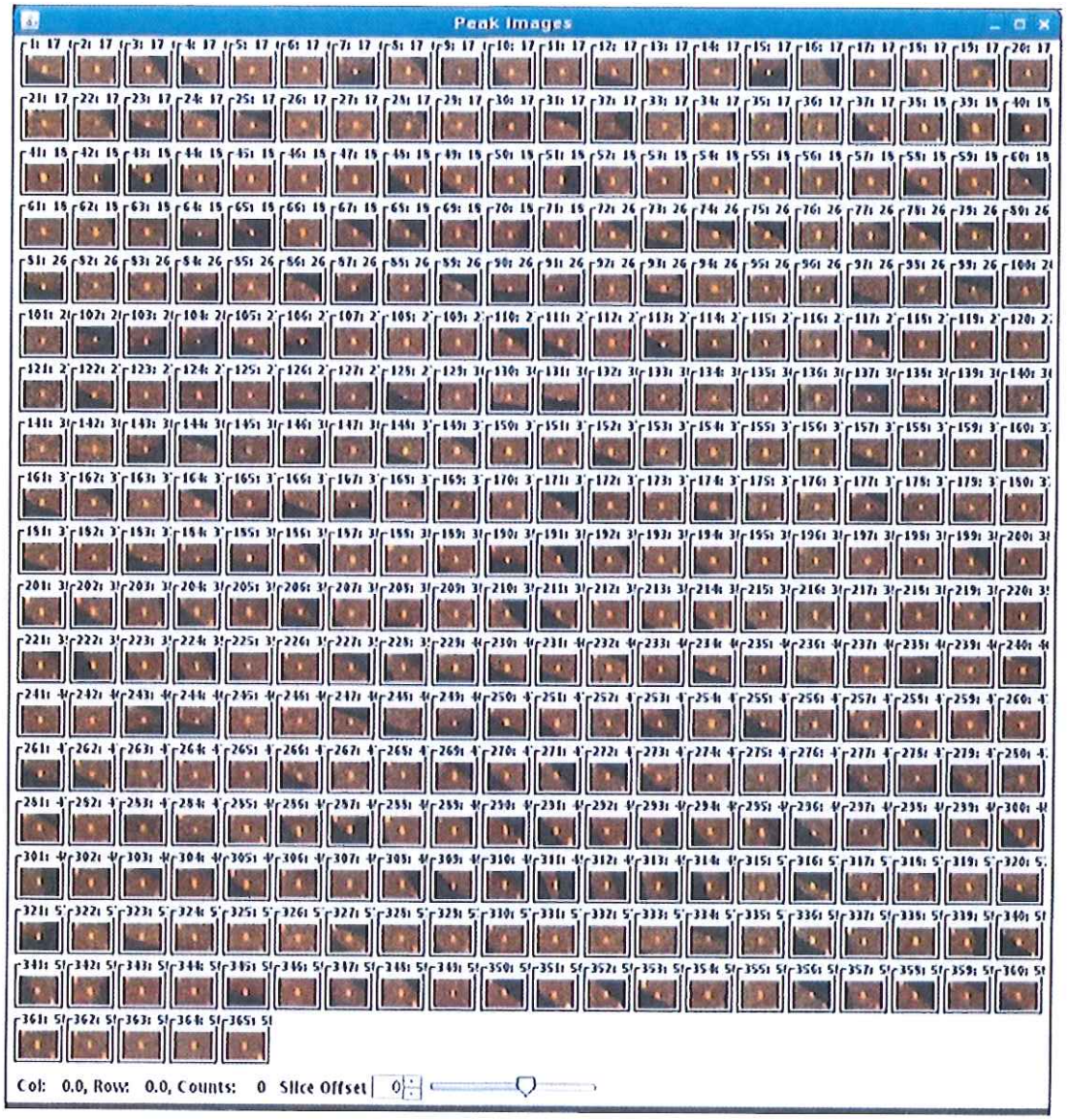
[a] $R_{merge} = 100 \times \left[\sum_{hkl} \sum_{i=1}^{n} \left| F^2(hkl) \right| - \left\langle F^2(hkl) \right\rangle_i \right] / \sum_{hkl} \sum_{i=1}^{n} F^2(hkl)$ where $F^2(hkl)$ is the intensity of the *hkl* reflection and $\left\langle F^2(hkl) \right\rangle_i$ is the mean value of *i* multiple measurements of the *n* equivalent reflections.

[b] $R_{value} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} |F_{obs}|$, R_{free} is the R_{value} computed using 5% randomly excluded from the structure determination data subset (where F_{calc} includes all scale factors and corrections for bulk-solvent).

[c] Root-mean-square deviations of bond lengths in Ångstrom and bond angles in degrees calculated with *phenix.refine* in the *PHENIX* program suite.

[d] Ramachandran plot quality assessment using *MolProbity*.

Integration of protein diffraction data on TOPAZ (draft)



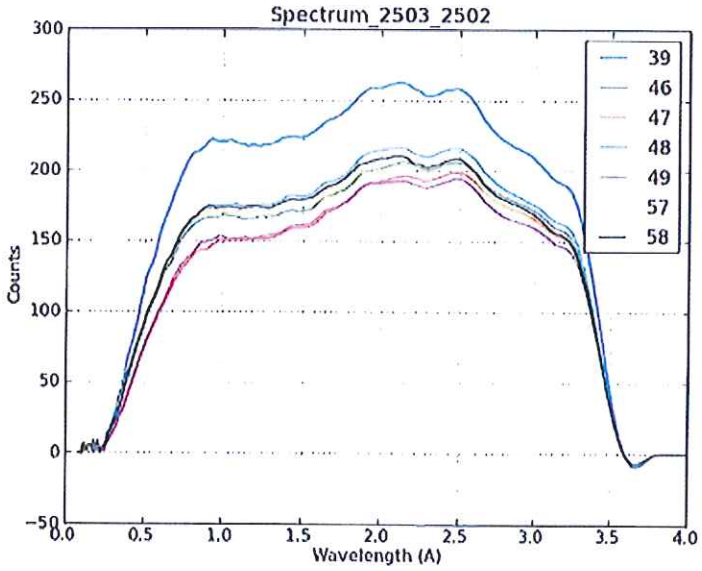
Leighton Coates and Art Schultz

Six diffraction images from an 8mm^3 fully perdeuterated protein crystal of TOHO-1 β lactamase were collected on the TOPAZ instrument using the second frame (3.6 to 7.2 Å). Table 1 below lists the sample and data collection parameters.

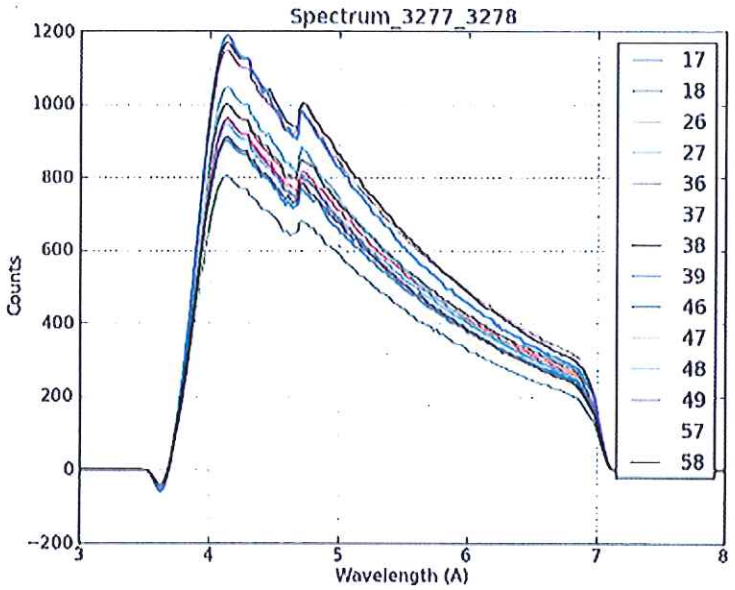
Table 1

Number of Images	6
Exposure time	~6 hours
Unit cell Dimensions (Å)	a=72.92, b=72.92, c=98.53 α=β=90° γ=120°
Spacegroup	$P3_221$ (number 154)
Rotation between settings	30°
Wavelength Range	3.6 to 7.3 Å

The incident neutron spectra for the first (0.5 to 3.50 Å) and the second frame (3.6 to 7.2 Å) are shown below.



(a) The wavelength spectra for the first frame on TOPAZ



(a) The wavelength spectra for the second frame on TOPAZ

Collecting protein data in the second frame (3.6 to 7.3 Å) is far from ideal as the peak flux occurs at 4 Å and the flux below 4 Å is very low. If one looks at Bragg's law it can be seen that the maximum resolution or lowest D_{min} spacing one can reach with a 4 Å neutron is half the wavelength or 2 Å. This would not be too bad however the corresponding two theta angle is 180° meaning that these reflections are inaccessible as they scatter back along the incident beam. This means that reasonable data completeness can only really be achieved to D_{min} values of around 2.3 to 2.5 Å. The six images were integrated using the Isaw and andrev packages the data statistics given below were generated using the WinGX program.

5014 reflections were integrated from the six images most of those were unique reflections however an R_{merge} of 30.4% could be generated from 1393 reflections which were merged to 671. The minimum value for D_{min} used was 3.0 Å with the maximum being 16 Å. The I/sigmal value at 3.0 Å was 4.81.

Conclusions

1) Protein data needs to be collected in the first frame. However due to the low flux in the first frame, the reflectivity of neutrons (λ^2) and wavelength dependent detector efficiency the exposure time increases by roughly a factor of eight compared to collecting data in the second frame.

2) It would be much better to collect protein data collect data using neutrons with wavelengths between 2.0-5.69 Å. However operating in this wavelength range enables the prompt pulse to

be seen. Further investigation into what is needed to operate in this wavelength band should be undertaken. It might be possible to exclude from data integration certain TOF values corresponding to the prompt pulse.

3) The R_{merge} indicates from such a small number of reflections signals that further work is needed on the integration software to give data of publishable quality.