DATA REDUCTION FROM SINGLE CRYSTAL DATA COLLECTION: INDEXING, INTEGRATION AND SCALING

A data collection using synchrotron radiation yielded diffraction images corresponding to rotations of the crystal along the φ , axis, perpendicular to the incident radiation. Data will be indexed and integrated using the Mosflm software [1], under the CCP4 suite interface [2]. After integration, systematic absences will be analyzed to identify the space group, manually or using the Pointless software [3]. Scaling in the correct space group will be performed using the software Aimless [4] of the CCP4 suite.

After opening the software CCP4i2 (new graphical interface), first a new project should be created (1). The name of the project will be selected (2) and the user should choose the working directory where input and output files will be saved by the software (3).

CCP4-7.1.000 Project Viewer: Lisozim File Edit Utilities Projects Help	
Task menu Export project Run Clone job Help Bi	lography Export MTZ Show log file New project
Job list Project directory Filter: Only show jobs containing text typed here	Fiter: Only show tasks containing text typed here
Job/File	Evaluation Integrate X-ray images
	Create a New Project ? Name of project/folder Lisozimal By default all projects go in the 'CCP4I2_PROJECTS' directory in your home area - click 'Select directory' to choose an alternative. 2 Hint to organise your projects: in the 'Manage projects' window you can use a project as a folder and drag other projects into it 2 Description of project
	Choose tag Choose tag C
<	Create project Select directory Cancel Help Cancel

Indexing and integration of diffraction spots.

In the window of the previously created project, the Mosflm software is selected from the list on the right (under "Integrate X-ray images") (4). If the task menu is not available in the main project window, the user can recall it at any time by pressing the "Task menu" button on the upper bar, left side. Once the software has been selected by double-click, to lunch it click "Run" (5) on the upper bar.





A new window will open, the main window of the Mosflm software. Images to be processed can be selected using the "Add images" icon (6). If the images are stored in a single directory, selecting a single image will be sufficient and the software will automatically upload all the images of the data collection. In the list that will appear in the main window, each image is followed by the $\Delta \phi$ interval (7). At the same time, a second window will automatically open allowing to manually analyze the diffraction images (8). In the image window, the user can see the selected spots using in the following processing (small red crosses) and the predicted position of the spots obtained during indexing, squares of different colors: blue for the full spots, yellow for the partials, red for spots discarded due to overlapping, and green for spots discarded due to too wide $\Delta \phi$ interval. In this window, the user can also manually correct the position of the main beam (large purple cross) and set the resolution limits for data analysis (blue circles).

iMosflm version	7.3.0, 2 6 st 2019				
Session Settings					
	140.18 125.73 + 100.10				m
	Images				
Images	Lattice 1	Unknown			^
	* Spacegroup	Unknown			
· · · ·	Mosaicity	0.00			
Indexing	Mosaic block size	100 (7)			
indexing	□ < Sector lyso2_rac_2_#####.cbf	φ:35.26->143.26			
	Q Matrix	Unknown			
	- Image 1	$\varphi(\mathbf{r}):35.26 - 35.56$			
Strategy	- Image 2	φ(r):35.56 - 35.86			
0.61	- Image 3	$\varphi(r):35.86 - 36.16$			
1	- Image 4	$\varphi(\mathbf{r}): 36.16 - 36.46$			
Cell Refinement	Image 5	$\varphi(r):36.46 - 36.76$			
	- Image 6	$\varphi(\mathbf{r}):36.76 - 37.06$			
	Image 7	$\varphi(r):37.06 - 37.36$			
AIIIIN	- Image 8	$\varphi(r):37.36 - 37.66$			
Integration	Image 9	$\varphi(r):37.66 - 37.96$			
	- Image 10	φ(r):37.96 - 38.26			
	Image 11	φ(r):38.26 - 38.56			
History	- Image 12	φ(r):38.56 - 38.86			
	- Image 13	φ(r):38.86 - 39.16			
	- Image 14	$\varphi(r): 39.16 - 39.46$			
	Image 15	$\varphi(\mathbf{r}):39.46 - 39.76$			
	- Image 16	$\varphi(\mathbf{r}):39.76 - 40.06$			
	Image 17	$\varphi(\mathbf{r}):40.06 - 40.36$			
	Image 18	$\varphi(\mathbf{r}):40.36 - 40.66$			
	Image 19	$\varphi(\mathbf{r}):40.66 - 40.96$			
	Image 20	φ(r):40.96 - 41.26			
	Image 21	$\varphi(\mathbf{r}):41.26 - 41.56$			
	Image 22	$\varphi(\mathbf{r}):41.56 - 41.86$			
	Image 23	$\varphi(\mathbf{r}):41.86 - 42.16$			
	Image 24	$\varphi(\mathbf{r}):42.16 - 42.46$			
	- Image 25	$\varphi(r):42.46 - 42.76$			~
			Groon	warnin	0.00



In the following tab of the software, selected from the left column (9), some images can be selected (10) for the initial determination of the unit cell. Usually, 10 images are sufficient, ideally selected at different diffraction angles. The software automatically selects the diffraction spots from the images (11). When present, rings due to the presence of crystalline ice can be excluded from analysis, by clicking the specific button from the main window (12). Presence of ice rings is indicative of a non-ideal cryoprotection of the crystal before freezing. The "index" button allows the start of the indexing process (13).

iMosflm version	17.3.0, 21st August 2	019								12)		×
Session Settings	Help												
🗋 🗁 🖬 🛤 1	40.08 🔂 125.8	0 😝 100.10	\$ 5.0	00 ()10.0	0.86	1.03	t0.00	l 🔔 2	0 🚱 🔩	23 23	81 0	2.50	
	Autoindex	ing		1	0								
Images	so2_rac_2_###	##.cbf:1-5,	301-305					•	•	1	Inde	× (13) 🖫	n
	Image	φ range		Auto	Man	De	1	> $I/\sigma(I)$	Find	Use ^			
	1	35.26 - 35	.56	594	0		0	388	۲		(11)		
	2	35.56 - 35	.86	581	0		0	381	۲		\bigcirc	. Time V	
	4 3	35.86 - 36	.16	608	0		0	377	۲		2	SALAR BARNA E.	
	4 4	36.16 - 36	.46	545	0		0	342	۲		2. 12 12 12 12 12 12 12 12 12 12 12 12 12	法公司 公	
Strategy	- 5	36.46 - 36	.76	610	0		0	380	۲				
0-8 ⁻¹	\$ 301	125.26 - 1	25.56	389	0		0	227	۲		12		
6-0	\$ 302	125.56 - 1	25.86	402	0		0	246	۲			The Market	
Cell Refinement	4 303	125.86 - 126.16		390	0		0		•			2.4.16	
	Total			4909	0		0	3011					
										~			
Integration	(14)1		1	1				1	(1	1	1	
	Stion	Lat.	Pen.	a	b	С	α	β	Y	σ(x,y)	Nref	δ beam	^
	H 1 (ref)	aP	0	37.1	79.0	79.2	89.7	89.5	89.6	0.33	1714	0.03 (0.	0)
History	⊞ 1 2 (ref)	aP	2	37.1	79.0	79.2	90.3	90.5	89.6	0.33	1/14	0.03 (0.	0)
	I 3 (ref)	mc	3	112.5	111.7	37.1	90.0	90.6	90.0	0.29	1682	0.02 (0.	0)
	E (ref)	mP	5	27.0	70 0	79.1	90.0	90.2	90.0	0.34	1710	0.03 (0.	0)
	ET 6 (ref)	mC	6	112 5	111 7	37 1	90.0	90.2	90.0	0.35	1682	0.03 (0.	0)
	1 7 (ref)	mP	6	37.0	78.9	78.9	90.0	89.9	90.0	0.33	1720	0.01 (0.	0)
	1 8 (ref)	oC	8	111.7	112.3	37.0	90.0	90.0	90.0	0.30	1691	0.03 (0.	0)
	1 9 (ref)	oP	8	37.0	78.8	79.0	90.0	90.0	90.0	0.34	1721	0.02 (0.	0)
	🗄 🎦 10 (ref) tP	11	78.9	78.9	37.0	90.0	90.0	90.0	0.34	1726	0.01 (0.	0)
	(rea) mC	110	162.6	37.1	79.1	90.0	90.2	90.0	-	-		
	Lattices: Sh	ow									Start	beam search	Show
	Spacegroup: P4	▼ Pri	ior cell:										
	Mosaicity:	0.55 Es	stimate										
Calculating predic	cted reflections											Green warni	nas 0

At the end of the indexing, the software shows a list of possible unit cells (14), with the respective Bravais lattice (a = triclinic, m = monoclinic, o = orthorhombic, t = tetragonal, h = hexagonal, c = cubic; P = primitive, C = base-centered, F = face-centered, I = body centered, R = rhombohedral centering), cell parameters, and penalty. The penalty value indicates how many spots do not match the proposed unit cell. When selecting the

unit cell, the user should prefer higher symmetry cells (lower in the list) among those marked by the software with a green color, indicating a satisfactory indexing. Structure solution can be performed also in a lower symmetry option, but a higher number of parameters would reduce the refinement quality. If a wrong higher symmetry is selected at this stage, the scaling will yield a high R_{merge} value. In this second case, the indexing should be done again decreasing the symmetry of the unit cell.

In the diffraction image window, the user can limit the resolution using the button (**15**) and moving the external circle to modify the highest resolution value (to move the circle, the arrow button on the same bar should be selected).



Once the images have been indexed and the unit cell parameters have been selected from the list, the image window shows the predicted positions of the diffraction spots (**16**), with the mentioned color code.

Once the indexing is complete, the tab "Cell refinement" (17) can be selected from the main window. This window allows to refine the values of the unit cell parameters using a larger number of diffraction images (18). After the process is started with the suitable button (19), the software starts to calculate optimized values of parameters in order to reduce the Root Mean Square residuals ("RMS res.") between measured and calculated spot positions. The 20 and 21 panels allow to observe the variation of



each parameter while the optimization is carried on. Graphs **22** and **23** show a graphical representation of the change of some parameters along with image number: "Tilt" e "Twist" in panel **22** are detector positioning parameters, while panel **23** shows mosaicity and orientation of the crystal. During the refinement, the spot position determined from indexing is optimized to cover the maximum intensity registered on the diffraction image (**24**). Finally, graph **25** shows the behavior of the RMS residues for the selected images.



In the following "Integration" tab (26), the software allows integration of the diffraction data. In this window the user can select the output file name (27), containing the integrated diffraction intensity corresponding to each reflection identified from Miller indexes hkl, and the number of images to be integrated (28). If no radiation damage is present during data collection, usually all collected images are integrated. When

integration process starts (29), the software automatically performs a new optimization of the image parameters using a smaller subset of subsequent images, showing the same parameter variations as in the previous cell refinement (panels 30 and 31), relative graphs for subsequent images (panel 32 and 33), and the integration profiles of diffraction spots with the respective background (panels 34 and 35). Panel 36 offers some indexes to monitor the behavior of the integration, among which the value of $\langle I/\sigma_I \rangle$. The same value for the single images is also reported in a graph (37) for both the full (red line) and partial (yellow line) reflections. A significant reduction of the $\langle I/\sigma_I \rangle$ value may indicate a signal deterioration, often due to the presence of radiation damage. Finally, graph 38 allows to evaluate the value of $\langle I/\sigma_I \rangle$ in the resolution shells, for both full and partial reflections.



At the end of the integration process, the software provides an *mtz* output file containing intensity values and standard deviations for each of the integrated diffraction spots. In addition, the software provides a second file, with extension *lp*, containing the details of the processing and the results for each image. The *lp* file is very useful when the integration process did not yield the expected results and the user should analyze it carefully before re-running the protocol.

After integration, the Mosflm software can be closed. Data are saved in a subfolder of the main folder of the project (selected at the beginning).

Analysis of systematic absences.

Before starting the scaling protocol, it is advisable to analyze the file containing integrated reflections (with *mtz* extension). This file can be opened from the main CCP4i2 interface by clicking on the specific job in the left panel of the main window to open the list of output and input files of the job (**39**). Double-click on the *mtz* file opens it.

CCP4-7.1.000 Project Viewer: Lisozir	ma			Ū.	×
File Edit Utilities Projects Help					
Task menu Export project 🛛 😵 🥥 🔞	Bibliography Export MTZ Show log file	New project			
Job list Project directory		Job 1: Integrate images with Mosfim The job is Finished			
Filter: Only show Jobs containing text typed here	\sim	Input Results Comments			
La La Marca	- Provide mail in the	Run			
Job/File	Evaluation				
5 6 Data reduction	San=P 43 21 2 res= 1.00 Rmeas=.	No internet of the found form Man flat			
5 Data reduction		No megration data round from mostim		18:09-23-M	ay-2020
> 1 MOLREP	R=0.55 RFree=0.57	Input Data			
3 Data reduction	Sgp=P 41 21 2 res=1.00 Rmeas=.				
1 Integrate images with Mosfim		Output Data			
₽ lyso2_rac_2_00001.mtz					
session.mos		Unmerged reflections lyso2_rac_2_00001.mtz	i i		
			1		
		► Job run details			
		1			
<	>	Data reduction			

The file is opened by another software of the CCP4 suite, ViewHKL (40), that allows visualization and analysis of the contents of the *mtz* file. Two additional buttons in the window of the ViewHKL software allow to visualize the list of reflections (41 and 43) and the sections of the reciprocal lattice showing the intensities of the nodes (42 and 44). In the latter window, a single section can be selected and analyzed with the buttons on the right (45).

h ViewHKL 1.11 04/02/20	016 [lyso2_rac_2_00001_c.mtz] [40] [
File Edit View Help		
Open Print PDF/PS	Copy General Summary HKL List HKI Zones Batches	0 Exit
General data		
Path	C:/Users/Rita/Documents/LAVORO/DIDATTICA/Biocristallografia/esercitazioni/CCP4_JOBS/job_2/lyso2_rac_2_00001_c.r	ntz
Туре	Unmerged MTZ	
Space group	P 4	
Space group confidence	ze X (confidence flag is not set)	
Cell	78.76 78.76 36.94 90 90 90	
Resolution low	22.24	
Resolution high	0.9999	
Number of lattices	1	
Number of reflections	1155551	
Number of datasets	1	
✓ Dataset #1: New/New/I	New	
Cell	78.76 78.76 36.94 90 90 90	
Wavelength	0.7	
Batches	1-360	
<		>

) Dpen	 → Pri) nt Pl	DF/PS	Сору	(Seneral	Summary	P HKL List	HKL Z	pnes (Batches							
	(4:	3)							Tit	tle: U	ntitlec	1						
	h	k	I.	Orig. H	Orig. K	Orig. L	M/ISYM	BATCH	T	SIGI	IPR	SIGIPR	FRACTIONCALC	XDET	YDET	ROT	WIDTH	LP
ſ	44	63	8	-44	-63	8	261	1	-22.85	56.87	-6.09	56.87	0.60	1089.69	1154.09	35.52	0.59	0.5
	44	62	8	-44	-62	8	261	1	-90.72	57.14	-75.18	57.14	0.44	1086.78	1146.85	35.24	0.59	0.50
	44	62	9	-44	-62	9	261	1	256.30	60.43	124.50	60.43	0.07	1101.76	1142.38	35.74	0.56	0.5
	44	61	9	-44	-61	9	261	1	205.78	61.22	77.68	61.22	0.71	1098.80	1135.28	35.47	0.56	0.5
	44	60	9	-44	-60	9	261	1	12.19	59.96	35.57	59.96	0.27	1095.91	1128.30	35.18	0.56	0.5
	44	60	10	-44	-60	10	261	1	-11.07	62.18	39.52	62.18	0.35	1110.66	1123.99	35.61	0.54	0.5
	44	59	10	-44	-59	10	261	1	167.15	63.53	165.07	63.53	0.67	1107.72	1117.14	35.32	0.53	0.5
	44	58	10	-44	-58	10	261	1	30.36	62.79	36.54	62.79	0.01	1104.85	1110.41	35.02	0.53	0.5
	44	59	11	-44	-59	11	261	1	189.14	65.20	106.15	65.20	0.16	1122.42	1112.97	35.68	0.51	0.5
0	44	58	11	-44	-58	11	261	1	-53.28	65.00	-34.10	65.00	0.79	1119.44	1106.23	35.39	0.51	0.5
1	44	57	11	-44	-57	11	261	1	-84.29	64.80	-26.34	64.80	0.07	1116.54	1099.62	35.09	0.51	0.5
2	44	58	12	-44	-58	12	261	1	-4.40	66.58	66.58	66.58	0.14	1134.12	1102.19	35.69	0.49	0.5
3	44	57	12	-44	-57	12	261	1	-2.18	66.95	50.08	66.95	0.81	1131.11	1095.56	35.40	0.49	0.5
4	44	56	12	-44	-56	12	261	1	57.08	36.62	78.08	36.62	0.07	1128.17	1089.05	35.10	0.49	0.5
5	44	57	13	-44	-57	13	261	1	11.79	37.04	-7.30	37.04	0.26	1145.82	1091.63	35.63	0.48	0.5
6	44	56	13	-44	-56	13	261	1	24.99	37.21	53.32	37.21	0.76	1142.78	1085.10	35.34	0.47	0.5
7	44	55	13	-44	-55	13	261	1	3.85	37.38	15.39	37.38	0.00	1139.81	1078.69	35.04	0.47	0.5
flec	tion sta	rting r	number										1			Find r	eflection	
mbr	er of ref	lection	ns	-									100		H	63	K	L



In this exemple, reflections with h = 0 and k = 0 are present only when l is a multiple of 4 (46). This kind of systematic absence is indicative of the presence in the unit cell of an element of roto-translational symmetry 4_1 or 4_3 along the unique axis of the tetragonal system (*c* axis).

A similar analysis can be performed for systematic absences on the other 2 reciprocal axes, h (with k = 0 and l = 0, **47**) and k (with h = 0 and l = 0, **48**), showing the presence of axes of roto-translational symmetry of order 2, namely 2_1 , along axes a and b in the real space.



This analysis can be performed automatically using the software Pointless [3] of the CCP4 suite. The user should be aware, however, of the possible limitations of the selection, such as the intrinsic ambiguity between enantiomorphic space groups.

Scaling and merging of diffraction data.

The next step in image analysis is the scaling of diffraction data, the Aimless software can be selected from the main CCP4 interface (**49**). When the software is selected, the input file can be specified (**50**), together with the name of the dataset (**51**). The user can select further scaling options, such as resolution limits of the data to be scaled (**52**) and the space group (**53**), when previously known. Once parameters are selected, the software can be started with the "Run" button (**54**).



CCP4-7.1.000 Project Viewer: Lisozima	- 🗆 🗙
File Edit Utilities Projects 54	
Task menu Export project Run Clone job Heip Bibliography Export MTZ	Show log file New project
Job list Project directory	Job 3: Data reduction - AIMLESS The job is Pending
Filter: Only show jobs containing text typed here	Input Results Comments
Job/File Evaluation	Input Data Important Options Additional Options
3 Data reduction	Job title Data reduction
 2 Integrate images with Mosfim 1 Integrate images with Mosfin 	Use data from job 2 Integrate images with Mosfim as input below Show list Select unmerged data files 2 lyso2_rac_200001.mtz 50 Crystal name uyso dataset name uyso1 51 Resolution range (Å) to Maximum resolution in files 0.00Å 52 use explicit resolution range in symmetry determination as well as in scaling Options for symmetry determination Determine Laue group and space group 53 Detional input data 1. Reference data to resolve indexing ambiguity and space group use reference data in analysis against Batch after scaling Reference data are Reflection list and is optionally defined in next line Reflectionsis not used C. Optional existing FreeR set, define to copy or extend if necessary
<u> </u>	

Once the scaling process is complete, the software provides the output reflection file, together with the space group selected by the Pointless software (automatically run by the Aimless protocol) (55) and the scaling statistics. For the lysozyme crystal analyzed in this example, the selected space group ($P 4_1 2_1 2$) has an enantiomorphic group ($P 4_3 2_1 2$). The choice of the correct space group is not possible at the scaling level, because the two possible groups have the same symmetry relations in the reciprocal space. When enantiomorphic space groups are present, it is advisable to perform in parallel two data reductions using the alternative groups and test the correctness of each solution in the following phasing step.

CCP4-7.1.000 Project Viewer: Lisozima				×
File Edit Utilities Projects Help				
Task menu Export project Run Clone job Help Bibliography Export MTZ	Show log file New project.			
Job list Project directory	Job 3: Data reduction - AIMLESS	The job is Finisher	d	
Filter: Only show jobs containing text typed here	Input Results Comments			
Lab/Ella Evoluation	Headline Summary SG details 1	ergingGraph SDanalysis 1ergingDetail	Istats Biblio	Run
A Data andustion Care D 41 31 3 and				
S 2 Integrate images with Mosfim A 1 Integrate images with Mosfim				
	▼ Key summary			
	Selecting space group P 41 21 2 as solutions are emantiomorphic Solution probability: 0.882, Confidence 0.854 (F Key statistics for Dataset: Lisozima/Lyso/Lyso/ Unit cell: 78.760 78.760 36.940 90.000 Resolution of input data: 1.00Å, Anistorpic limits: - In hk plane 1.00Å Remas: overall 0.056, inner bin 0.035 In outer bin: Mean():sdl 3.0 CC(1/2 0). Overall filtered Mean(chi/2): 1.02 Anomalous CC(1/2) in inner bin -0.362 No significant anomalous signal detect SD correction information: WARMNO: SD correction parameters ar Asymptotic 1/sigl, Isa: Run 1: Isa = 15.4	morphic ambiguity later high resolution limit for symmetry testing 1.000Å 1 90.000 90.000, wavelength 0.700000Å resolution estimate: beyond 1.00Å ° CC(1/2), 1.00Å 1/σ - Along I axis 1.00Ű CC(1/ 860 : ed resomewhat outside expected limits, see SD analy 4	å) 2), 1.00Å 1/σ ysis panel below for more details	
		1 . 11 12 .		~
	<			>
	Expert MR - PHASER	MOLREP	CRANK2	
< >>	SHELX			

The software provides statistical values that allow to evaluate the quality of the scaling processing. These values are summarized in a table (56). In addition, analysis of the graphs of statistical parameters against resolution shells (57) is crucial to determine the resolution limit of the data collection, to be used in the following steps. Statistical analysis against the progress of data collection (58) may indicate radiation damage accumulating in the crystal.





References

- [1] T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell and A.G.W. Leslie, *"IMosflm: a new graphical interface for diffraction-image processing with MOSFLM"*. Acta Cryst. 2011; D67, 271-281.
- [2] M. D. Winn et al., "Overview of the CCP4 suite and current developments". Acta Cryst. 2011; D67, 235-242
- [3] P.R. Evans, "An Introduction to Data Reduction: Space-Group Determination, Scaling and Intensity Statistics". Acta Cryst. 2011; D67(Pt 4), 282-92.
- [4] P.R. Evans, and G.N. Murshudov, *"How Good Are My Data and What Is the Resolution?"*. Acta Cryst. 2013; D69(Pt 7), 1204-14.