

Supplementary Data

Multicomponent mixtures for cryoprotection and ligand solubilization.

Lidia Ciccone^{1,2}, Laura Vera^{1§}, Livia Tepshi^{1,2}, Lea Rosalia^{1,2},

Armando Rossello², and Enrico A. Stura^{1*}

¹CEA, iBiTec-S, Service d'Ingénierie Moléculaire des Protéines, Laboratoire de Toxinologie Moléculaire et Biotechnologies, Gif-sur-Yvette, F-91191, France.

²Dipartimento di Farmacia, Università di Pisa, Via Bonanno 6, 56126 Pisa, Italy.

Description of the database – The data collection database is a series of HTML formatted files. The organization is based on two frames (“menu” and “main”) (Supplementary Fig. 1). The “menu” frame on the left contains the individual synchrotron sessions and the “main” frame the details of each flash-cooled crystal. On clicking on the left menu the details of the individual sessions can be visualized in the main frame. The HTML formatting codes for a table where each table cell represents a flash-cooled crystal contained in the dewar and each line represents a carousel of vials containing the cryo-cooled crystals. If two dewars are used during the synchrotron session, two tables follow one after the other. Various details of the crystal are easily obtained either directly on the same HTML page or by following one of the HTML links. The entire database, which includes raw data storage, extends to over 8 Tb.

X-ray data and processing statistics – The first item of each cell of the table is the codename for the co-crystal. (TTR-LiC131_1: TTR = transthyretin; LiC = Lidia Ciccone; 131 = compound 131; _1 = first crystal.) It has an associated HTML link that points to the location where the frames are stored on disk and where the processed data and statistics can be found. The ligand name or code is reported in the line below for co-crystals and soaked crystals. An image for the ligand is often present on the same page (not shown). For soaks: the soak details are given on the “cryo:” line.

Resolution – Since the primary metric for evaluation of success of a data collection experiment is the resolution limit to which each crystal diffracts to, this data item is reported directly in the cell corresponding to each crystal. The resolution limit is reported only when the data has been indexed and merged correctly. In the past, the resolution limit was evaluated from images collected, for example from a marCCD X-ray detector diffraction image (Supplementary Fig. 1). Evaluating data collected on the fast PILATUS 2M or 6M detectors visually is difficult, since very few spots are visible, although the data quality is excellent. Evaluation of the diffraction quality can only be done after the data has been processed and scaled together. The resolution in the HTML file is updated each time that the data is reprocessed with the latest software, XDS [2] or MOSFLM [3]. The resolution limit is set with a correlation coefficient $CC(1/2) > 50^1$ [4] and a mean $I/\sigma(I)$ of about 1

1 $CC(1/2)$ = percentage of correlation between intensities from random half-datasets. Correlation significant at the 0.1% level is marked by an asterisk in CORRECT.LP in XDS [1].

($0.8 < I/\sigma(I) < 2.5$). The resolution is associated to an HTML link that points to the first image of the data set, which allows for an independent visual evaluation of the data.

Images – The number of images collected are reported in the cell. With the fast PILATUS 2M or 6M detectors a strategy calculation is rarely carried out since the data collection terminated before humans can chose and set up the strategy.

Cryo-soaking – Since there is not enough space in the table cell to describe even in an abbreviated form the details of the cryosoaking, each cryo-soaking solution is given a name and the exact composition and details on how it has been carried out is reported below the dewar tables, under the header cryosolutions. The 40% cryomix, 50% precipitant, 10% buffer is used in all cases when a cryomix (CM) is mentioned. The composition of the cryomixes is in another table.

FRZ081 : 20 February 2014 - 16 - 8:00 ESRF ID23-2 Enrico & Lidia

DEWAR1	POSITIONS									
CAROUSEL	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-
457 #1
205 #2
453 #3
454 #4
206 #5

DEWAR1	POSITIONS									
CAROUSEL	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-
456 #6
482 #7
498 #8
483 #9
484 #10

DEWAR2	CAROUSEL
457 #1	-- 1 --

Synchrotron sessions Link to collected frames Link to crystallization database

Cryosolutions:
M8Cryo4: C4, 12% PEG 10K, 100mM PCTP 80/20
M8Cryo7: CM31, 18% MPEG 2K, 100mM MMT 50/50
M12Cryo11: CM24, 25% PEG 6K, 100mM TRIS HCl, pH 8.0
M12Cryo13: CM30, 25% PEG 6K, 100mM TRIS HCl, pH 8.0
M12Cryo15: CM24, 25% PEG 6K, 100mM TRIS HCl, pH 8.0
M8Cryo5: CM15, 10% PEG 20K, 100mM AAB 80/20
M13Cryo6: CM33, 10% PEG 20K, 100mM AAB 80/20
TTRCryo1: CM29, 25% MPEG 5K, 100mM AAB 60/40
TTRCryo4: CM17, 25% MPEG 5K, 100mM PCTP 65/25
TTRCryo5: CM25, 25% MPEG 5K, 100mM AAB 80/20
TTRCryo6: CM20, 25% MPEG 5K, 100mM CHC 80/20
TTRCryo7: CM34, 25% MPEG 5K, 100mM AAB 80/20
TTRCryo8: CM19, 25% MPEG 5K, 100mM AAB 80/20
UTRryo5: C6 (5E/10M/5P/5S), 25% MPEG 5K, 25% MPEG 550, 100mM PC
TTRCryo2: CA1, 12.5% DMSO + 37.5% DMSO + 30% 1,4-dioxane

Link to cryosolution Link to first image

CryoMixes	Concentrations
CM1	12.5 % di-ethylene glycol + 12.5 % MPD + 37.5 % 1,2-propanediol + 12.5 % DMSO
CM2	25 % di-ethylene glycol + 25 % glycerol + 25 % 1,2-propanediol
CM3	12.5 % di-ethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5mM NDSB 201
CM4	25 % di-ethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol
CM5	12.5 % di-ethylene glycol + 25 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 mM NDSB 201
CM6	12.5 % ethylene glycol + 25 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO
CM7	12.5 % di-ethylene glycol + 12.5 % ethylene glycol + 12.5 % glycerol + 25 % 1,2-propanediol + 12.5 % DMSO
CM8	12.5 % di-ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5 mM NDSB 201
CM9	12.5 % diethylene glycol + 12.5 % MPD + 25 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO
CM10	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO
CAEBE11	12.5 % diethylene glycol + 12.5 % MPD + 37.5 % 1,2-propanediol + 12.5 % 1,4-dioxane
COCM12	25 % diethylene glycol + 12.5 % glycerol + 25 % 1,2-propanediol + 12.5 % 1,4-dioxane
COCM13	12.5 % diethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane + 12.5 mM NDSB 201
CDICM14	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % 1,4-dioxane
CECM15	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % 1,4-dioxane + 12.5 mM NDSB 201
CEFCM16	12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane
COCM17	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane
CHCM18	12.5 % diethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane + 12.5 mM NDSB 201
CLICM19	12.5 % diethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane
CIUCM20	12.5 % diethylene glycol + 12.5 % MPD + 37.5 % 2,3-butanediol + 12.5 % DMSO
CIUCM21	25 % diethylene glycol + 25 % glycerol + 25 % 2,3-butanediol
CIUCM22	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO + 12.5 mM NDSB 201
CIUCM23	12.5 % ethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol
CIUCM24	12.5 % diethylene glycol + 25 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 mM NDSB 201
OSICM25	12.5 % ethylene glycol + 25 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO
CFICM26	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % glycerol + 25 % 2,3-butanediol + 12.5 % DMSO
CIUCM27	12.5 % diethylene glycol + 12.5 % MPD + 25 % glycerol + 12.5 % 2,3-butanediol + 25 % DMSO + 12.5 mM NDSB 201
OSICM28	12.5 % diethylene glycol + 12.5 % MPD + 25 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO
CAICM29	25 % diethylene glycol + 12.5 % MPD + 37.5 % 2,3-butanediol + 12.5 % 1,4-dioxane
BIUCM30	25 % diethylene glycol + 12.5 % glycerol + 25 % 2,3-butanediol + 12.5 % 1,4-dioxane
CCICM31	12.5 % diethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO + 12.5 mM NDSB 201
CDICM32	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % 1,4-dioxane
CEICM33	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % 1,4-dioxane + 12.5 mM NDSB 201
CFICM34	12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane
CIUCM35	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane
CHICM36	12.5 % diethylene glycol + 12.5 % MPD + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane + 12.5 mM NDSB 201
CIUCM37	12.5 % diethylene glycol + 12.5 % ethylene glycol + 12.5 % glycerol + 12.5 % 2,3-butanediol + 12.5 % DMSO + 12.5 % 1,4-dioxane
SMICM38	12.5 % diethylene glycol + 12.5 % glycerol + 12.5 % 1,2-propanediol + 25 % DMSO + 25 % 1,4-dioxane
SMICM39	12.5 % ethylene glycol + 12.5 % glycerol + 12.5 % 1,2-propanediol + 25 % DMSO + 25 % 1,4-dioxane
SMICM40	25 % diethylene glycol + 25 % ethylene glycol + 25 % glycerol + 25 % 1,4-dioxane
SMICM41	12.5 % diethylene glycol + 25 % ethylene glycol + 12.5 % 1,2-propanediol + 25 % DMSO + 25 % 1,4-dioxane
SMICM42	12.5 % ethylene glycol + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % 2,3-butanediol + 25 % DMSO + 25 % 1,4-dioxane
SMICM43	12.5 % diethylene glycol + 12.5 % glycerol + 12.5 % 1,2-propanediol + 12.5 % 2,3-butanediol + 25 % DMSO + 30 % 1,4-dioxane

Crystal size is not considered important and there is no note of this parameter. Many small crystals used on microfocus beamlines give excellent data, at times better than larger ones. The main problem is getting them well centred. The implementation of mesh and line scans in MxCuBE [5] provides a good aid for difficult cases. Parameters like mosaicity and other details are not stored in the data collection database, but accessible via the HTML link to the stored data. The space group details for all solved structures are stored in a separate structure within the database. All polymorphs

are listed in another HTML table as an aid to indexing data.

Crystallization & ligand solubilization – Protein production, ligand solubilization, protein-ligand complexation details are accessible via the HTML link to the crystallization drop, which opens in the “main” frame the crystallization tray.

Connection to other databases – In addition, for data collected at the ESRF in Grenoble the HTML database stores all the details needed to connect to the ISPyB database [5].

The main advantage of an HTML database is that it can be used over various platforms, Windows, Mac and Linux with the simple use of a navigator. By keeping images small, a subset of the entire data collection, crystallization, and solved structure databases, without access to the raw and processed data fits on an 8 Gb. USB key. The subset contains sufficient details useful for collecting new data, indexing data, retrieving crystallization details and viewing the PDB files of all solved structures with COOT [6] or PyMOL [7] (no maps nor MTZ files). Details can be extracted from the database using Linux commands like “grep”. HTML is user-friendly and all new students can navigate the web. Unfortunately, at present it is difficult to mine the data using a computer program since the database is not homogeneous nor sufficiently well structured for program use.

References:

- [1] W. Kabsch, Integration, scaling, space-group assignment and post-refinement., *Acta Crystallogr. D. Biol. Crystallogr.* 66 (2010) 133–44. doi:10.1107/S0907444909047374.
- [2] W. Kabsch, XDS, *Acta Crystallogr. D. Biol. Crystallogr.* 66 (2010) 125–32. doi:10.1107/S0907444909047337.
- [3] A.G.W. Leslie, The integration of macromolecular diffraction data., *Acta Crystallogr. D. Biol. Crystallogr.* 62 (2006) 48–57. doi:10.1107/S0907444905039107.
- [4] P.A. Karplus, K. Diederichs, Linking crystallographic model and data quality., *Science.* 336 (2012) 1030–3. doi:10.1126/science.1218231.
- [5] J. Gabadinho, A. Beteva, M. Guijarro, V. Rey-Bakaikoa, D. Spruce, M.W. Bowler, et al., MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments., *J. Synchrotron Radiat.* 17 (2010) 700–7. doi:10.1107/S
- [6] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot., *Acta Crystallogr. D. Biol. Crystallogr.* 66 (2010) 486–501. doi:10.1107/S0907444910007493.
- [7] W.L. DeLano, The PyMOL Molecular Graphics System, in: L. Schrödinger, New York Editor, 2010.