EMDB Map Distribution Format Description

Version 1.0

Document created and published by the EMDataBank.org* team

***collaboration between PDBe, RCSB-PDB, NCMI**

http://emdatabank.org

JULY 23 2010

EM Data Bank Map Distribution Format

The EM Data Bank (EMDB) accepts and distributes 3D map volumes derived from several types of EM reconstruction methods, including single particle averaging, helical averaging, 2D crystallography, and tomography. Since its inception in 2002, the EMDB map distribution format has followed CCP4 definition [1], which is widely recognized by software packages used by the structural biology community. CCP4 map format is closely related to the MRC map format used in the 3DEM community [2]; CCP4 is slightly more restrictive, in that voxel positions are limited to a grid that includes the Cartesian coordinate origin (0,0,0).

In July 2010, all 765 released EMDB maps were remediated to improve the uniformity of CCP4 format header parameters across the archive, including correction of density statistics and voxel dimensions. The CCP4 map format description [1] is reproduced here along with the operational implementation for 3DEM map volumes employed in the remediation effort. Conventions noted below indicate parameter settings that are common to many (but not all) maps in the archive.

A CCP4 format map file consists of (i) a required metadata header consisting of 256 thirty-two-bit words (ii) an optional crystallographic symmetry table and (iii) a required contiguous block of voxels of an exact byte length defined by the header.

(i) HEADER SECTION

WORD	ID	DATA TYPE [3]	VALUE	CRYSTALLOGRAPHIC MAPS	Modified use for 3D EM MAPS	Notes
1-3	NC, NR, NS	signed int	>0	# of columns (fastest changing), rows, sections (slowest changing)	convention: NC=NR=NS	
4	MODE	signed int	0,1,2,3,4	voxel datatype	convention: 2	[4]
5-7	NCSTART, NRSTART, NSSTART	signed int		position of first column, first row, and first section (voxel grid units)		[5]
8-10	NX, NY, NZ	signed int	>0	intervals per unit cell repeat along X,Y Z	intervals per map length along X,Y,Z; convention: same as NC, NR, NS	[6]
11-13	X_LENGTH, Y_LENGTH, Z_LENGTH	floating pt	>0	Unit Cell repeats along X, Y, Z In Ångstroms	Map lengths along X,Y,Z in Ångstroms	[6]
14-16	ALPHA, BETA,GAMMA	floating pt	>0, <180	Unit Cell angles (degrees)	convention: 90, 90, 90	[7]
17-19	MAPC, MAPR, MAPS	signed int	1 (=X) 2 (=Y) 3 (=Z)	relationship of X,Y,Z axes to columns, rows, sections	convention: 1, 2, 3	
20-22	AMIN, AMAX, AMEAN	floating pt		Minimum, maximum, average density		
23	ISPG	signed int	1-230	space group #	1	[8]
24	NSYMBT	signed int	80n	# of bytes in symmetry table (multiple of 80)	0	[8]
25	LSKFLG	signed int	0,1	flag for skew matrix	0	
26-34	SKWMAT	floating pt		skew matrix-S11, S12, S13, S21, S22, S23, S31, S32, S33		[9]
35-37	SKWTRN	floating pt		skew translation-T1, T2, T3		[9]
38-52	EXTRA	32 bit binary		user-defined metadata		[9]
53	MAP	ASCII char	"MAP "	MRC/CCP4 MAP format identifier		[10]
54	MACHST	32 bit binary		machine stamp		[11]
55	RMS	floating pt		Density root-mean-square deviation		
56	NLABL	signed int	0-10	# of labels		[12]
57-256	LABEL_N	ASCII char		Up to 10 user-defined labels		[12]

(ii) SYMMETRY TABLE (OPTIONAL, CRYSTALLOGRAPHIC MAPS ONLY)

Symmetry operations for the map space group stored as text as in IUCr Tables, with operators grouped into lines of 80 characters. Example:

X, Y, Z
-X, -Y, Z
X+1/2, -Y+1/2, -Z
-X+1/2, Y+1/2, -Z

(iii) VOXELS

Voxels stored within the file are organized according to MAPC, MAPR & MAPS values. Voxel datatypes are defined by MODE. The byte size of this contiguous voxel dataset is equal to NC*NR*NS*(number of bits of the datatype defined by MODE divided by 8). Columns change fastest, followed by Rows, and then by Sections.

CONVERSION OF DEPOSITED MAPS to CCP4 FORMAT for DISTRIBUTION

EMDB accepts several map formats for upload. In all cases the map voxel lengths are set to author-provided input values by adjusting X_LENGTH, Y_LENGTH, and Z_LENGTH values.

When an MRC map is uploaded, NCSTART, NRSTART, and NSTART values are determined from floating point values for the (X,Y,Z) origin point encoded in MRC map header words 50-52. In order to avoid loss of precision in map placement upon conversion, MRC origin position values should be cleanly divisible by voxel sizes.

Spider maps are converted to CCP4 format using the recommended Spider script, which rotates the map according to the convention (Z → -Z, X→Y, Y→X) [13].

Voxel density values are not modified by EMDB, with one exception: if the density for the imaged macromolecule is found to be negative, a scale factor of (-1) is applied to make the density positive, following the convention recommended by Heymann et al. [14].

Voxel/axis order is retained from the uploaded map unless the handedness of the map must be changed; in this case the map is mirrored through the X-Y plane (Z→-Z).

EMDB does not interpolate maps onto alternate grid settings. For maps with point symmetry, symmetry axes should intersect at a single voxel point to allow placement at the physical-space origin, following recommended conventions [14, 15, 16]. Similarly, a helical symmetry axis should coincide with a line of voxels.

NOTES AND REFERENCES

[1] CCP4 map format: http://www.ccp4.ac.uk/html/maplib.html. Definition and maintenance of the CCP4 map format is the provenance of the CCP4 project executive committee.

[2] MRC map format: http://www2.mrc-lmb.cam.ac.uk/image2000.html.

[3] map header section word types:

signed int =Thirty-two (32) bit signed integer, ISO/IEC 10967

floating pt = Thirty-two (32) bit floating point number, IEEE 754

[4] MODE value establishes the voxel datatype. EMDB currently holds MODE=0,1,and 2 maps.

MODE = 0: 8 bits, density stored as a signed byte (range -128 to 127, ISO/IEC 10967)

MODE = 1: 16 bits, density stored as a signed integer (range -32768 to 32767, ISO/IEC 10967)

MODE = 2: 32 bits, density stored as a floating point number (IEEE 754)

MODE = 3: 32 bits, Fourier transform stored as complex signed integers (ISO/IEC 10967)

MODE = 4: 64 bits, Fourier transform stored as complex floating point numbers (IEEE 754)

MODES other than 2 and 0 may not work in CCP4 programs.

[5] The position of the first voxel is defined in grid units by NCSTART, NRSTART, and NSSTART. The center of the voxel with grid position (0,0,0) corresponds to the Cartesian coordinate origin.

[6] Lengths in Ångstroms for a single voxel are as follows:

Xvoxel = X_LENGTH/NX

Yvoxel = Y_LENGTH/NY

Zvoxel = Z_LENGTH/NZ

[7] By convention, cell angles (ALPHA, BETA, GAMMA) are 90 degrees for single particle or tomogram EM maps; they follow IUCr space group conventions for crystals.

[8] Space Group Numbers are defined by IUCr conventions (Table 12.3.4.1 Standard space-group symbols", pages 824-831, International Tables for Crystallography, Volume A, fifth edition). For 3D volumes of single particle or tomogram entries, ISPG=1 and NSYMBT=0.

[9] SKWMAT, SKWTRN, and EXTRA fields are not currently used by EMDB.

[10] MACHST is 0x44,0x41,0x00,0x00 (written/read as 4 hex byte sequence) for little endian machines and 0x11,0x11,0x00,0x00 (written/read as 4 hex byte sequence) for big endian machines.

[11] The "MAP " text string is terminated with a space and not with a NULL character.

[12] Following the 2010 remediation, maps distributed by EMDB now have a single label of form ":::EMDataBank.org::::EMD-1234::::".

[13] http://www.wadsworth.org/spider_doc/spider/docs/spider.html

[14] Heymann, J.B., Chagoyen, M., Belnap, D.M. (2005). Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. J. Struct. Biol. 151, 196-207. doi: 10.1016/j.jsb.2005.06.001

[15] Baldwin, P.R., Penczek, P. A. (2007). The transform class in SPARX and EMAN2. J. Struct. Biol. 157, 250-261. doi: 10.1016/j.jsb.2006.06.002

[16] Lawson, C.L., Dutta, S., Westbrook, J.D., Henrick, K., Berman, H.M. (2008). Representation of viruses in the remediated PDB archive. Acta Cryst. D 64, 874-882. doi: 10.1107/S0907444908017393