Pure Appl. Chem., Vol. 84, No. 10, pp. 2171–2182, 2012.
http://dx.doi.org/10.1351/PAC-REC-12-02-03

JCAMP-DX for circular dichroism spectra and metadata (IUPAC Recommendations 2012)*

Benjamin Woollett[1], Daniel Klose[2], Richard Cammack[3], Robert W. Janes[2,‡], and B. A. Wallace[1,‡]

[1]Department of Crystallography, School of Biological Sciences, Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, UK; [2]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK; [3]Pharmaceutical Sciences Research Division, King's College, University of London, London, UK

Abstract: Circular dichroism (CD) spectroscopy is a widely used technique for the characterisation of proteins. A number of CD instruments are currently on the market, and there are more than a dozen synchrotron radiation circular dichroism (SRCD) beamlines in operation worldwide. All produce different output formats and contents. In order for users of CD and SRCD data to be able simply to compare and contrast data and the associated recorded or unrecorded metadata, it is essential to have a common data format. For this reason, the JCAMP-DX-CD format for CD spectroscopy has been developed, based on extensive consultations with users and senior representatives of all the instrument manufacturers and beamlines, and under the auspices of IUPAC, based on the Joint Committee on Atomic and Physical Data Exchange protocols. The availability of a common format is also important for deposition to, and access from, the Protein Circular Dichroism Data Bank, the public repository for CD and SRCD data and metadata. The JCAMP-DX-CD format can be read by standard JCAMP programs such as JSpecView. We have also created a series of parsers, available at the DichroJCAMP web site (http://valispec.cryst.bbk.ac.uk/formatConverter/dichroJCAMPDX-CD.html), which will enable the conversion between instrument and beamline formats and the JCAMP-DX-CD format.

Keywords: circular dichroism spectroscopy; data standards; IUPAC Physical and Biophysical Chemistry Division; JCAMP-DX-CD; Protein Circular Dichroism Data Bank (PCDDB); synchrotron radiation circular dichroism spectroscopy.

1. INTRODUCTION

The JCAMP-DX file format was first published in 1988 and was designed as a standard form for exchanging data collected using infrared spectroscopy [1]. This was the culmination of the efforts of the Task Force on Spectral Data Portability proposed at the Pittcon (Pittsburgh Conference) in 1983. The name is an acronym for Joint Committee on Atomic and Molecular Physical data and the group Data eXchange and has been, since 1995, the responsibility and intellectual property of the International Union of Pure and Applied Chemistry (IUPAC). Versions of JCAMP-DX are already available for

*Sponsoring bodies: IUPAC Physical and Biophysical Chemistry Division; IUPAC Committee on Printed and Electronic Publications (CPEP): see more details on p. 2180.

‡Corresponding authors: E-mail: b.wallace@mail.cryst.bbk.ac.uk, r.w.janes@qmul.ac.uk

infrared spectra, chemical structures, nuclear magnetic resonance, mass spectroscopy, ion mobility spectrometry, and electron magnetic resonance data [1–6]. The addition of the JCAMP-DX version for circular dichroism (CD) spectroscopy brings it in line with those of other spectroscopic techniques and allows for the users of this technique to take advantage of available JCAMP-DX software and facilitate cross-talk between research communities.

CD spectroscopy is used globally in many biology and chemistry laboratories. Lab-based CD instruments are available as a standard piece of equipment in most life science departments. In the past three years, the technique has been utilised in more than 5000 published papers. A more recent development has been the establishment of synchrotron radiation circular dichroism spectroscopy (SRCD) beamlines [7–9], which take advantage of the high light flux available in the ultraviolet and vacuum ultraviolet wavelength ranges at synchrotrons to produce spectra covering broader spectral ranges and with higher signal-to-noise levels.

At present there are a number of commercial CD instrument manufacturers and more than a dozen SRCD beamlines producing CD data, all of which use different formats for spectral and associated metadata. At a meeting of the International Technical Advisory Board (consisting of instrument company and beamline experts, see acknowledgments) of the Protein Circular Dichroism Data Bank (PCDDB) [10,11] in July 2010, agreement was reached to accept the use of the IUPAC JCAMP-DX format as a standard medium of data exchange. The version of JCAMP-DX presented in this paper (JCAMP-DX-CD) has been developed to address the need for a community-approved, machine-independent, standard format for protein CD and SRCD spectral data and associated metadata. With the growing availability of online resources for the archiving and analysis of this data, such as the PCDDB [10,11] and DichroWeb [12], the advantages of a seamless means for data transfer between resources have never been so apparent.

2. SCOPE

This document is designed as a reference for those using CD and SRCD spectroscopy as a tool for research and protein characterisation, for those engaged in the development of CD and SRCD instrumentation and analysis software, and to ensure uniformity and conformity with the standards used by the PCDDB and the available JCAMP-DX visualisation tools such as JSpecView [13]. A concise overview of the special characters, structure, and data types involved is presented in order to provide the context for the description of how the data and metadata types in a complete PCDDB entry have been incorporated into the JCAMP-DX format. This is not a comprehensive description of all available features provided by the JCAMP-DX formats in general, and hence should be read in conjunction with previous publications available from the IUPAC web site (http://www.jcamp-dx.org).

3. COPYRIGHT

4. THE JCAMP-DX-CD FORMAT

4.1 Data types

The following data types are used with the labels defined in a JCAMP-DX-CD file to enable it to be read by both humans and/or machines:

TEXT To be read by humans; not intended to be read by computers.

STRING Intended to be parsed by both computers and humans.

AFFN ASCII Free Format Numeric. e.g., '-5.067320E+02' or '3', for computer parsing.

4.2 Use of special characters in JCAMP-DX-CD

JCAMP-DX files are fed into parsing software as a sequence (or string) of ASCII characters. To provide structure to the file, certain "special characters" are used by parsing software as markers to splice different sections of this string into different categories. These categories may in turn be parsed into subcategories using different rules specific to the type of data associated with that category. To parse a specific piece of data from a file involves navigating via these markers (indicated using special characters) to identify the location in the substring of characters that represent the desired information. The special characters listed below were defined in the publication of JCAMP-DX version 4.24 [1].

Referred to as a data label flag and indicates the start of a data label.

= The first instance of a '=' after the '##' indicates the end of a data-label, and the text between indicates the type of data which will come after the '='.

If '=' is the last character on a line then it indicates that the String data type on that line continues on the next line without break (i.e., ignore the new line characters).

$$ Everything after this on the line is a comment to be read by a human and should be ignored by any parsing software.

**$** The first character in a user-defined label. To ensure flexibility, JCAMP-DX was designed to allow users to add their own labels for data types not covered by reserved data labels. These take the form '##$labelName='.

? Indicates that a required value is either out of range or unavailable.

<> Used instead of quotation marks to delimit strings of characters.

() Used to delimit data groups which include strings of characters.

JCAMP-DX files were intended to be readable by humans and parsed by computer software. The location on a line provides context to some special characters, meaning that new line tags are in effect unspoken special characters. Each line is restricted to a length of 80 characters, and all data are described as being 'free form', meaning that the column (position along the line) is not relevant, differing from other formats where the column number is used to identify the meaning of some of the symbols present.

As a data label flag is parsed, all spaces, dashes, and underlines are removed and all lower case is converted to upper case, making the following, in effect, equivalent: '##TI T LE=', '##ti tle__=', '##tI/_t_le='.

4.3 Structure and contents of a JCAMP-DX-CD file

A JCAMP-DX file is described as either a compound file or a simple file. A compound file contains the textual content of multiple simple files. In the context of a compound file, each of these are referred to as a data BLOCK (a simple file can be thought of as a file containing a single data BLOCK). These data are not simply concatenated, but are placed within an outer BLOCK referred to as a link BLOCK which can be thought of as analogous to a folder in a file system (although you cannot have a link BLOCK within a link BLOCK). Each BLOCK (both the link and data varieties) begins with the data label '##TITLE=' and ends with the data label '##END=' (both must be at the beginning of a line) with the data to be stored and associated labels on the lines in between. The term 'labeled data record' (LDR) is defined to describe both the data label and the 'data set' (the data associated with that data label). Specifications state that only a single LDR can be present per line (consistent with the fact that a data label must occur at the beginning of the line) and that it can continue over an indefinite number of lines until the next data label. Just one of each type of LDR is permitted per BLOCK. Each data BLOCK in a compound file should be a complete simple file in its own right (if, for example, it were cut and pasted to a blank file), and for this reason LDRs present in each BLOCK can only relate to the data within this BLOCK. In other words, all LDRs relating to a data BLOCK must be contained within that particular BLOCK. The LDRs contained within a BLOCK are defined as belonging to one of two main types, core and note LDRs.

A compound file with two data BLOCKs will be of the structure shown below:

##TITLE= a title for the link BLOCK.
$$ Core LDRs required for link BLOCKs.
##TITLE= a title in free form text.
$$ Core LDRs required for all data BLOCKs.
$$ Note LDRs containing the metadata associated with abscissae values in data table.
$$ Core LDRs required to aid parsing of abscissa values in data table.
$$ Data table.
##END=
##TITLE= a title in free form text.
$$ Core LDRs required for all data BLOCKs.
$$ Note LDRs containing the metadata associated with abscissa values in data table.
$$ Core LDRs required to aid parsing of abscissa values in data table.
$$ Data table.
##END=
##END=

An '##END=' tag on a line will close the BLOCK opened most recently (on the nearest line above), meaning that the first '##TITLE=' which opens the link BLOCK is closed by the last '##END=', embedding the two data BLOCKs in between.

4.4 Core LDRs

Fixed header information is required in all blocks (except where otherwise stated) and must be provided on successive lines in the order they are listed below:

##TITLE= (TEXT) A description of the contents of the BLOCK.
##JCAMP-DX= (STRING) For CD this is always: '5.01'. This LDR conveys to any parsing software the version of JCAMP-DX the following lines will conform to.

##DATATYPE= (STRING) If the BLOCK in question is a link BLOCK then the predefined spelling is 'LINK'. For infrared spectra, for example, the predefined spellings include 'INFRARED SPECTRUM', 'INFRARED PEAK TABLE', etc.

For CD data the predefined spellings are: 'FINAL PROCESSED SPECTRUM PCD', 'NET SPECTRUM PCD','AVERAGE SPECTRUM PCD', 'AVERAGE BASELINE PCD', 'CALIBRATION SPECTRUM PCD', 'RAW SAMPLE PCD', 'RAW BASELINE PCD','RAW SAMPLE HT PCD', 'RAW BASELINE HT PCD'.

The characters 'PCD', short for 'protein circular dichroism', are included to differentiate it from similar techniques such as linear dichroism or vibrational circular dichroism, and from other sample types such as nucleic acids, small molecules, and polysaccharides, which would give rise to a different set of metadata fields. However, this format is designed so that in the future it could be modified to include other related techniques and/or types of samples.

##DATA CLASS= (STRING) Not present in the header of link BLOCKs, this LDR defines the type of data expected in a data BLOCK. In the case of CD files this will always consist of the string 'XYDATA'.

##ORIGIN= (TEXT) Name of organisation and contact details of individuals who generated the data.

##OWNER= (TEXT) This field should either have the string "PUBLIC DOMAIN" or if copyrighted this should ideally be expressed in the following form: 'COPYRIGHT(C)<year>by<name>'.

Although not always referred to as core data, the following two LDRs are required in their respective contexts: A link BLOCK must have the '##BLOCKS=' LDR with an integer (AFFN) indicating the number of BLOCKs contained within. All data inside a link BLOCK must also have a '##BLOCK ID=' which is an integer from 1 to n (AFFN), where n is the total number of BLOCKs, identifying which data BLOCK this is (no two data BLOCKs within a link BLOCK will have the same integer). So if a data BLOCK is removed from a link BLOCK the IDs of the other BLOCKs must be altered to reflect this (unless it is the last one in the sequence), the '##BLOCKS=' value would also need amending.

4.4.1 Core LDRs required to aid parsing of abscissa values in data table

The following LDRs must be included on the lines preceding the table containing the spectral data:

##XUNITS= (STRING)	In the protein CD version of JCAMP-DX this is always measured in units of 'nanometers'.
##YUNITS= (STRING)	Predefined spellings used for the Y-axis units for CD are: 'Delta Epsilon', 'Mean Residue Ellipticity', 'Millidegrees (theta)', 'yy units', 'DRS units', 'Molar Ellipticity', 'Delta Absorbance', 'Arbitrary' - for use with the high tension or dynode voltage.
##XFACTOR= (AFFN) **##YFACTOR=** (AFFN)	The X and Y factor values are required for use in the spectral data compression process used in JCAMP-DX (described, in brief, in a later section). During this process two values are found which if multiplied by every Y or X value will turn that value into an integer. This data compression method was deemed unnecessary for use with the CD version, so these values will always be 1.
##FIRSTX= (AFFN)	The first X value in the data table.
##LASTX= (AFFN)	The last X value in the data table.
##MAXY= (AFFN)	The maximum Y value in the data table.
##MINY= (AFFN)	The minimum Y value in the data table.
##NPOINTS= (AFFN)	Number of abscissa pairs.
##FIRSTY= (AFFN)	The first Y value in the data table.

4.5 The data table

This is a requirement in all data BLOCKS. The data table contains the data in tabular form, with each tab representing a different column and the lines representing rows. The data table starts on the line below the data label '##XYDATA=' and ends on the line above '##END='. On the same line as the '##XYDATA=' label is a predefined expression (STRING) indicating the type of table used for data storage; for protein CD data this is always the following string '(X++(Y..Y))'. Below is a sample of a data table with just the required core LDRs associated with the data and the data table itself (with all but two data points removed from the 105 originally present).

##XUNITS= nanometers
##YUNITS= Millidegrees (theta)
##XFACTOR= 1
##YFACTOR= 1
##FIRSTX= 175
##LASTX= 279
##MINY= –1.87522
##MAXY= 2.03978
#NPOINTS= 105
##FIRSTY= –0.03666
##XYDATA= (X++(Y..Y))
175.0 -0.03666
$$ middle section of data trimmed to save space
279.0 –0.0459442
##END=

4.6 JCAMP-DX-CD reserved notes

A restriction for JCAMP-DX-CD to make it compatible with a PCDDB deposition is that only a single set of abscissa value pairs can be represented per JCAMP-DX file or BLOCK. Since a fully populated PCDDB deposition involves multiple raw sample and raw baseline spectra, an average baseline and average sample spectrum, a net spectrum, a final processed spectrum, and a calibration spectrum, multiple data BLOCKs are required. Each data BLOCK must contain all the information of a complete stand-alone simple JCAMP-DX-CD file, and different data types have different associated metadata. The metadata present in a BLOCK where '##DATATYPE= CALIBRATION SPECTRUM PCD' will, for example, not have the same metadata as the BLOCKs where '##DATATYPE= AVERAGE BASELINE PCD' which will not have the same as BLOCKs where '##DATATYPE= RAW SAMPLE PCD'. But the final processed spectrum ('##DATATYPE= FINAL PROCESSED SPECTRUM PCD') contains all metadata types since it is generated using the abscissae values of all three.

Each raw sample spectrum and raw baseline spectrum is also accompanied by an additional set of abscissa value pairs relating the high tension (HT) voltage (otherwise known as dynode voltage) at each wavelength. A link is therefore required from each raw sample BLOCK to its respective HT BLOCK. The following LDR is to be included in every raw baseline, raw HT and raw sample BLOCK (where '##DATATYPE=' is 'RAW BASELINE PCD', 'RAW BASELINE HT PCD', 'RAW SAMPLE HT PCD' or 'RAW SAMPLE PCD', respectively).

##REPEAT NUMBER= (AFFN)

By this means, the order of scans is recorded and the association of each HT BLOCK and its respective raw sample/baseline BLOCK is made clear.

The following section lists the reserved labels that are incorporated into JCAMP-DX-CD along with the type of data expected to accompany the label. When this information alone is not deemed self-explanatory, additional notes are provided.

4.6.1 The use of 'STRING' data type

LDRs with associated data that are described as belonging to the '(STRING)' type (which as stated in the definitions for version 4.24 [1] are intended to be readable by humans and computers) are presented either with restrictions of the form that data provided should take or with a list of acceptable predefined spellings. While the use of predefined spellings can be of great value for automated analyses, these lists are not future proof. For this reason, the predefined spelling lists are not intended to be restrictive and free-form text is acceptable if a predefined spelling of the necessary information is not available.

4.6.2 Sample data LDRs

These LDRs can be included in all BLOCKs where the abscissae values are generated via measurements of the sample under investigation. These include the final processed spectrum, the net spectrum, the average sample spectrum, all raw sample spectra repeats, and the HT measured during each raw sample repeat.

##LOW WAVELENGTH CUTOFF= (AFFN) Lowest wavelength where spectral data is deemed reliable.

##CRITERIA FOR LOW WAVELENGTH CUTOFF= (TEXT) Reason for selection of low wavelength cutoff.

##BUFFER CONTENTS AND CONCENTRATIONS= (TEXT)

##PROTEIN PURITY PERCENTAGE= (AFFN)

##PURITY QUANTIFICATION METHOD= (STRING) 'QAA', 'SDS gels', 'Mass Spectrometry', 'Activity assay', 'Supplier value', 'none'.

##PROTEIN CONCENTRATION= (AFFN) In units of mg/ml.

##CONCENTRATION QUANTIFICATION METHOD= (STRING) 'QAA', 'A280', 'Extinction Coefficient', 'Gravimetric', 'Nitrogen Determination', 'Nanodrop/Microspectrophotometer', 'Supplier value'.

##SEQUENCE= (STRING) Amino acid sequence using the same single letter notation as UniProt, spaces are permitted and both upper and lower case are permissible. Spaces are permitted and both upper and lower case are permissible.

##NUMBER OF RESIDUES= (AFFN)

##MEAN RESIDUE WEIGHT= (AFFN) In Daltons.

##MOLECULAR WEIGHT= (AFFN) In Daltons.

##MACROMOLECULAR PARTNERS AND CONCENTRATION OR RATIO= (TEXT)

##LIGANDS PRESENT AND CONCENTRATION OR RATIO= (TEXT)

##EXPRESSION TAGS= (STRING) These should be provided in the form 'N,xxxx' or 'C,xxxx' where N and C denote the terminus where the expression tag is located and 'xxxx' denotes the amino acid sequence of the expression tag. Expression tags should only be included when they have not been cleaved prior to spectral analysis.

##EXPRESSED AS= (STRING) 'Wild-type', 'Mutant', 'Mutant (single site)', 'Mutant (multi site)'.

##EXPRESSION SYSTEM OR NATURAL SOURCE= (STRING) 'E. coli', 'Baculovirus', 'P. pastoris', 'S. cerevisiae'.

##PROTEIN NAME= (STRING) Name deemed most scientifically correct and hence most likely to yield relevant information upon searches of the scientific literature. Ideally this matches the name present in the associated UniProt sequence record.

##ALTERNATIVE PROTEIN NAMES= (STRING) Alternative names likely to yield relevant information from the scientific literature, including abbreviations. Names should be comma separated.

##SOURCE ORGANISM= (STRING) Most scientifically correct name of organism from which the original sequence was obtained. In the case of synthetically generated sequences, the word 'none' should be entered.

##UNIPROT ID= (STRING) Multiple codes should be comma separated. All A–Z characters should be upper case.

##PDB ID= (STRING) Upper and lower case accepted. Multiple codes should be comma separated.

##ENZYME CLASSIFICATION= (STRING) Four integers period ('.') separated.

##CATH CLASSIFICATION= (STRING) Integers period ('.') separated.

##MEDLINE ENTRY= (STRING) Positive integer.

##TYPE OF PROTEIN= (STRING) 'membrane', 'soluble globular', 'soluble fibrous'.

##PROTEIN SUPPLIER= (TEXT) Information such as company name (plus catalogue number where applicable) or name of laboratory and/or group, etc.

##MUTATION DETAILS= (TEXT) Descriptive details regarding the differences between the wild-type and the sample from which the spectrum was taken. Use of commonly used mutation nomenclature such as 'H55del' or 'Q5_W6insC' is recommended.

The following LDRs are reserved for information relating to publications which make use of this experiment and key words or phrases which may be relevant for this spectrum upon a search of the PCDDB. The inclusion of these LDRs in JCAMP-DX-CD permits an entire PCDDB entry to be stored, downloaded, or deposited in JCAMP-DX-CD format.

##PUBLICATION AUTHORS= (TEXT)
##PUBLICATION YEAR= (TEXT)
##PUBLICATION JOURNAL= (TEXT)
##PUBLICATION TITLE= (TEXT)
##PUBLICATION VOLUME= (TEXT)
##PUBLICATION PAGES= (TEXT)
##KEYWORD PHRASE A= (TEXT)
##KEYWORD PHRASE B= (TEXT)
##KEYWORD PHRASE C= (TEXT)
##KEYWORD PHRASE D= (TEXT)
##KEYWORD PHRASE E= (TEXT)
##KEYWORD PHRASE F= (TEXT)
##KEYWORD PHRASE G= (TEXT)
##KEYWORD PHRASE H= (TEXT)
##KEYWORD PHRASE I= (TEXT)
##KEYWORD PHRASE J= (TEXT)

4.6.3 Final processed spectrum specific LDRs

The following LDRs should only be included in the final processed spectrum data BLOCK ('##DATA TYPE= FINAL PROCESSED SPECTRUM PCD').

##WAVELENGTH RANGE FOR ZEROING= (STRING) Should take the form value dash value; '255-258' with no spacing. Values in nanometers. If no zeroing has been performed this LDR should not be included.

##SMOOTHING TECHNIQUE= (STRING) 'Savitsky-Golay', 'Fourier Transform', 'Moving Average', or 'Cubic Spline'. If no smoothing was performed this LDR should be absent.

##NUMBER OF SMOOTHING POINTS= (AFFN) Refers to the window size used by the smoothing algorithm used (not applicable for Fourier smoothing). This value will be an integer. If no smoothing was performed this LDR should also be absent.

##SMOOTHING PERFORMED= (STRING) Boolean; 'YES' or 'NO'.

##FINAL SPECTRUM CALIBRATED= (STRING) Boolean; 'YES' or 'NO'. The term calibration refers to the instrument calibration carried out using the calibration spectrum.

4.6.4 LDRs not to be included in calibration spectrum

As these variables should be identical for both the measurement of the baseline and the sample under investigation, they are to be included in all JCAMP-DX-CD BLOCKs with the exception of the BLOCK containing the abscissae values for the calibration spectrum ('##DATA TYPE= CALIBRATION SPECTRUM PCD').

##WAVELENGTH INTERVAL= (AFFN) Units in nanometers.

##DWELL OR AVERAGING TIME= (AFFN) Units in seconds.

##SAMPLE CHAMBER ATMOSPHERE= (STRING) 'Nitrogen', 'Vacuum'.

##SAMPLE CELL COMPOSITION= (STRING) 'CaF2', 'LiF', 'MgF', 'Quartz', 'Suprasil'.

##SAMPLE CELL TYPE= (STRING) 'Cylindrical', 'Cylindrical-Demountable', 'Rectangular', 'Rectangular-Demountable', 'Flow Cell'.

##SAMPLE CELL PATHLENGTH= (AFFN) Units in centimetres.

##EXPERIMENTAL TEMPERATURE= (AFFN) Units in degrees Celsius.

##CONTINOUS OR STEPPED SCAN= (STRING) 'Continuous', 'Stepped', 'Unknown'.

##CELL PATHLENGTH CALIBRATION METHOD= (STRING) 'Interferometry', 'Dilution Standards', 'Manufacturers Spec'.

4.6.5 LDRs to be included in all data BLOCKs

##INSTRUMENT OR BEAMLINE= (STRING) Predefined spelling; 'Applied Photophysics', 'Aviv', 'Jasco', 'Olis', 'BESSY II', 'Brookhaven NSLS UB9', 'Brookhaven NSLS UB11', 'BSRF', 'Daresbury 3.1', 'Daresbury 12.1', 'Diamond B23A', 'Diamond B23B', 'HiSOR', 'ISA CD1', 'ISA UV1', 'NSRRC', 'Soleil DISCO'.

##DEPOSITION DATE= (STRING) In the form YYYY-MM-DD. This LDR is reserved for data downloaded from the PCDDB and refers to the date of deposition. Inclusion of LDR definitions such as this permits a complete PCDDB entry to be stored in the JCAMP-DX-CD format.

##CD OR SRCD= (STRING) 'CD', 'SRCD'

4.6.6 LDRs for inclusion within BLOCKs where the abscissae values are influenced by the average sample spectrum and/or the average baseline spectrum

For inclusion in 'FINAL PROCESSED SPECTRUM PCD', 'NET SPECTRUM PCD', and 'AVERAGE SPECTRUM PCD' BLOCKs.

##NUMBER OF RAW SAMPLE REPEATS= (AFFN) Number of raw sample spectra involved in the generation of the averaged abscissae values.

For inclusion in 'FINAL PROCESSED SPECTRUM PCD', 'NET SPECTRUM PCD', and 'AVERAGED BASELINE PCD' BLOCKs.

##NUMBER OF RAW BASELINE REPEATS= (AFFN) Number of raw baseline spectra involved in the generation of the averaged abscissae values.

4.6.7 LDRs specific to the calibration spectrum

These LDRs are only of relevance for BLOCKs containing abscissae values influenced by the calibration spectrum. These are the calibration spectrum itself ('##DATA TYPE= CALIBRATION SPECTRUM PCD') and the final processed spectrum ('##DATA TYPE= FINAL PROCESSED SPECTRUM PCD').

##CSA OR ACS= (STRING) 'CSA', 'ACS'. CSA is camphour sulphonic acid and ACS is ammonium camphour sulphonate.

##CSA ACS EXPERIMENT TEMPERATURE= (AFFN) Units in degrees Celsius.

##CD SIGNAL AT 290NM= (AFFN) Units in millidegrees.

##CSA ACS RATIO 192 AND 290NM= (AFFN) The ratio of the absolute value at 192 nm divided by the value at 290 nm.

##CSA ACS ZEROED AT= (STRING) Wavelength range used for zeroing. This should take the form; value dash value (for example, 255–258) with no spacing (values in nanometers). If calibration spectrum has not been zeroed this LDR should be absent.

##CSA ACS PATHLENGTH= (AFFN) Units of centimetres.

##CSA ACS STANDARD CONCENTRATION= (AFFN) In units of mg/ml.

5. SUMMARY

We have created a new JCAMP-DX format for CD and SRCD data and metadata after consultations with the user community, manufacturers, and beamline scientists. We have also created a web site (DichroJCAMP) which includes parsers for all common existing instrument output formats and popular CD data processing program formats, as well as for data input and output from the PCDDB. This work is aimed at enabling and facilitating data sharing and data comparisons of CD spectra produced by various means, and with data produced by other spectroscopic means for which JCAMP-DX formats already exist.

6. MEMBERSHIP OF SPONSORING BODIES

Membership of the IUPAC Physical and Biophysical Chemistry Division for the period 2012–2013 is as follows:

7. ACKNOWLEDGMENTS

This work was supported by grants from the International Union of Pure and Applied Chemistry (to BAW), from the Biotechnology and Biological Research Council (Strategic Tools and Resources Development Fund) (to RWJ) and from the Biotechnology and Biological Research Council (Bioinformatics and Biological Resources Fund) (to BAW and RWJ). We thank Drs. Lee Whitmore and Andrew Miles of Birkbeck College for helpful discussions, Drs. Robert Lancashire and Tony Davies from the IUPAC Joint Committee on Atomic and Molecular Physical Data Exchange for advice, and the members of the International Technical Advisory Board of the Protein Circular Dichroism Data Bank for working with us to make this a reality.

Membership of the International Technical Advisory Board is as follows: P. Baumgartel (BESSY II Synchrotron, Germany), P. Boxrud (Olis Inc., USA), S. Cave (Jasco U.K. Ltd., U.K.), L. Cole (Applied Photophysics Ltd., U.K.), K. Gekko (Hiroshima University, Japan), S.V. Hoffmann (Institute for Storage Ring Facilities (ISA), Denmark), Y.Lin (National Synchrotron Radiation Research Center, Taiwan), D. Moss (ANKA Synchrotron, Germany), G. Ramsay (Aviv Biomedical, Inc., USA), G. Siligardi (Diamond Light Source, U.K.), J. Sutherland (East Carolina University, USA), Y. Tao (Chinese Academy of Sciences, China), F. Wien (Soleil Synchrotron, France).

8. REFERENCES

1. (a) R. S. McDonald, P. A. Wilks Jr. *Appl. Spectrosc*. **42**, 151 (1988); (b) J. G. Grasselli. *Pure Appl. Chem.* **63**, 1781 (1991).
2. J. Gasteiger, B. M. P. Hendriks, P. Hoever, C. Jochum, H. Somberg. *Appl. Spectrosc.* **45**, 4 (1991).
3. (a) A. N. Davies, P. Lampen. *Appl. Spectrosc.* **47**, 1093 (1993); (b) A. N. Davies, J. Lambert, R. J. Lancashire, P. Lampen, W. Conover, M. Frey, M. Grzonka, E. Williams, D. Meinhart. *Pure Appl. Chem*. **73**, 1749 (2001).
4. P. Lampen, H. Hillig, A. N. Davies, M. Linscheid. *Appl. Spectrosc.* **48**, 1545 (1994).
5. J. I. Baumbach, A. N. Davies, P. Lampen, H. Schmidt. *Pure Appl. Chem.* **73**, 1765 (2001).
6. R. Cammack, Y. Fann, R. J. Lancashire, J. P. Maher, P. S. Mcintyre, R. Morse. *Pure Appl. Chem.* **78**, 613 (2006).
7. B. A. Wallace. *J. Synch. Rad.* **7**, 289 (2000).
8. B. A. Wallace. *Q. Rev. Biophys.* **42**, 317 (2009).
9. B. A. Wallace, K. Gekko, S. V. Hoffmann, Y.-H. Lin, J. C. Sutherland, Y. Tao, F. Wien, R. W. Janes. *Nucl. Inst. Meth. Phys. Res., Sect. A* **649**, 177 (2011).

10. L. Whitmore, B. Woollett, A. J. Miles, R. W. Janes, B. A. Wallace. *Structure* **18**, 1267 (2010).
11. L. Whitmore, B. Woollett, A. J. Miles, D. P. Klose, R. W. Janes, B. A. Wallace. *Nucleic Acids Res.* **39**, D480 (2011).
12. L. Whitmore, B. A. Wallace. *Biopolymers* **89**, 392 (2008).
13. R. J. Lancashire. *Chem. Central J.* **1**, 31 (2007).