readJDX: Import data in the JCAMP-DX format

Package Version 0.2.3
December 21, 2016

Bryan A. Hanson
DePauw University
Department of Chemistry & Biochemistry
Greencastle Indiana USA
e-mail: hanson@depauw.edu
github.com/bryanhanson/readJDX
CRAN.R-project.org/package=readJDX

The JCAMP-DX format was developed as an manufacturer-independent means of sharing spectroscopic data (http://www.jcamp-dx.org). The standard is described in a series of publications.[1–8] JCAMP-DX was developed during a time when data storage was expensive, and hence makes extensive use of compression schemes. The original application was to IR spectroscopy, but the standard has evolved over time to accommodate other spectroscopies.

File Structure

JCAMP-DX files consist of two parts:

1. A more-or-less human readible set of metadata which is need to understand the data and carry out the decompression. Besides required basic information about the data itself, most files contain instrument and manufacturer-specific parameters in the metadata.
2. A data table, compressed in various ways.

Challenges When Reading Files

The JCAMP-DX standard allows a lot of flexibility and instrument manufacturers have written widely varying export functions. Some of the challenges in reading a JCAMP-DX file include:

1. JCAMP-DX files can contain different kinds of data, including non-spectroscopic data[9] and more than one type of spectroscopic data.
2. JCAMP-DX files can contain more than one spectrum in the file.
3. Instruments may be configured to use . or , as the decimal point when writing files. This is generally a geographical / cultural nuance.
4. Numbers may be written using E to signify exponent, but only in some compression formats.
5. The data table can be presented in several possible formats.
6. Some manufacturers take liberties with the required format.

Supported Formats

1. Data tables can be presented in several different formats. The supported formats are:
   (a) XYDATA=(X++(Y..Y)) Each line starts with an x value, and is followed by as many y values as can conveniently fit within the 80 character limit. Subsequent x values are incremented according to the x resolution and the number of y values that fit on the previous line (which in turn depends upon the compression scheme). Each x value is used as a check point to verify that the correct number of y values were read.
   (b) DATA TABLE=(X++(R..R)) As above. The real data from an NMR spectrum.
   (c) DATA TABLE=(X++(I..I)) As above. The imaginary data from an NMR spectrum.
2. Within a data table, several different compression schemes can be employed. The following are supported:
   (a) AFFN: ASCII numbers separated by at least one space.

(b) PAC: Numbers separated by exactly one space, + or -.

(c) SQZ: Delimiter, leading digit and sign are replaced by a pseudo-digit.

(d) DIF: DIF uses a SQZ pseudo-digit for the first y value, but subsequent y entries are differences between each data value after the first. Sometimes referred to as SQZDIF.

(e) DUP: Not a format, but a method of signifying repeated values.

(f) DIFDUP: A combination of DIF and DUP. Widely used, as it permits the greatest amount of compression.

Formats That are Not Supported

1. Mixed spectroscopic types and non-spectroscopic entries (such as structures) are not supported by `readJDX` and are unlikely to be so in the future.
2. Compound files: JCAMP-DX files may contain more than one spectrum in the file. These types of files are not currently supported (however, NMR data are supported, even though they contain two spectra). The following JCAMP-DX standards require a compound file and are not supported:
   (a) EMR, EPR, ESR spectroscopy[7]
   (b) CD spectroscopy[8]
3. Data tables of simple x, y pairs (denoted XY..XY) are not supported. Your pull requests to deal with this format are welcomed.
4. `readJDX` is geared toward raw spectral data. Therefore data table formats representing derived information like PEAK TABLE and PEAK ASSIGNMENTS are not supported (but your pull requests are welcomed!).

Practical Matters

`readJDX` tries its best to deal with all these options. If you have a file that you believe should be supported but gives an error, please e-mail or file an issue. Be sure to attach the file that is giving you problems. Contributions to improve or expand the package, including pull requests, are always welcome!

Before release, `readJDX` is tested against a large collection of files with varying formats. A few of these files were obtained locally. Others were collected from publically available sources (e.g. http://www.jcamp-dx.org/testdata.html). These files are not included with the package to save space, and in addition, while they are publically available, for many of them the licensing status is unclear (i.e. the OWNER entry).

Known Issues

1. Files containing "S" representing a DUP digit of 1 are not supported and will stop with a message. Such files seem to be fairly rare.
2. Line number reporting (line numbers in the input file) during debugging is slightly off for NMR files that have lines which are solely composed of a comment. Such lines start with "$$" and are deleted during processing.
3. Some test files in the author's possession fail when carrying out the y value check, for unknown reasons.

Bibliography

[1] R. McDonald and P. Wilks, "JCAMP-DX, A Standard Format for Exchange of Infrared-Spectra in Computer Readible Form," *Applied Spectroscopy*, vol. 42, pp. 151–162, JAN 1988.

[2] J. Grasselli, "JCAMP-DX, A Standard Format for Exchange of Infrared-Spectra in Computer Readible Form," *Pure and Applied Chemistry*, vol. 63, pp. 1781–1792, DEC 1991.

[3] A. Davies and P. Lampen, "JCAMP-DX for NMR," *Applied Spectroscopy*, vol. 47, pp. 1093–1099, AUG 1993.

[4] P. Lampen, H. Hillig, A. Davies, and M. Linscheid, "JCAMP-DX for Mass Spectrometry," *Applied Spectroscopy*, vol. 48, pp. 1545–1552, DEC 1994.

[5] P. Lampen, J. Lambert, R. Lancashire, R. McDonald, P. McIntyre, D. Rutledge, T. Frohlich, and A. Davies, "An Extension to the JCAMP-DX Standard File Format, JCAMP-DX V.5.01 (IUPAC Recommendations 1999)," *Pure and Applied Chemistry*, vol. 71, pp. 1549–1556, AUG 1999.

[6] J. Baumbach, A. Davies, P. Lampen, and H. Schmidt, "JCAMP-DX. A Standard Format for the Exchange of Ion Mobility Spectrometry Data - (IUPAC recommendations 2001)," *Pure and Applied Chemistry*, vol. 73, pp. 1765–1782, NOV 2001.

[7] R. Cammack, Y. Fann, R. Lancashire, J. Maher, P. McIntyre, and R. Morse, "JCAMP-DX for electron magnetic resonance(EMR)," *Pure and Applied Chemistry*, vol. 78, pp. 613–631, MAR 2006.

[8] B. Woollett, D. Klose, R. Cammack, R. W. Janes, and B. A. Wallace, "JCAMP-DX for circular dichroism spectra and metadata (IUPAC Recommendations 2012)," *Pure and Applied Chemistry*, vol. 84, no. 10, pp. 2171–2182, 2012.

[9] J. Gasteiger, B. M. P. Hendricks, H. P., J. C., and S. H., "JCAMP-CS: A Standard Exchange Format for Chemical Structure Information in a Computer-Readible Form," *Applied Spectroscopy*, vol. 45, no. 1, pp. 4–11, 1991.