

2. A Taxonomy of JCAMP-DX Files

true

2021-09-20

Contents

1 Compound Files	1
2 Simple Files	1
2.1 Single Spectra	2
2.1.1 XYDATA= (X++(Y..Y))	2
2.1.2 XYPOINTS= (XY..XY)	2
2.1.3 PEAK TABLE= (XY..XY)	2
2.2 Multiple Spectra in NTUPLES Format	2
2.2.1 1D NMR	2
2.2.2 2D NMR	3
2.2.3 LC-MS or GC-MS	3
3 Number Formats	4
4 Program Flow	4
References	5

This vignette is based on `readJDX` version 0.6.1.

This vignette describes a taxonomy of JCAMP-DX files from the point of view of the `readJDX` package. It's based on reading and re-reading of the cited references, as well as inspection of files found in the wild and files shared with the author. No single publication includes all the needed information as the standard has evolved over time. This viewpoint may not be a perfect reflection of the actual standards and may include errors. Nevertheless, this is how the author looks at the structure of the files at this point in time, especially aspects that determine how the files should be processed. A table of the program flow is included as well.

At the most basic level, JCAMP-DX files are either simple or compound.

1 Compound Files

Compound files contain more than one spectrum or type of information (McDonald and Wilks 1988, sec. 3.3.2). `readJDX` does not support the processing of compound files. However, a utility function, `splitMultiblockDX`, is provided that will split such files into separate files which can be processed with `readJDX`.

2 Simple Files

Simple files contain the data for one experiment (McDonald and Wilks 1988, sec. 3.3.1). This experiment could be a single spectrum. However, it can also be a more complex experiment, in which case the NTUPLES tag is used. Examples of more complex experiments are NMR spectra in which both the real and imaginary points are reported, 2D NMR spectra, and LC-MS or GC-MS results.

2.1 Single Spectra

A single spectrum could be any basic spectroscopic experiment: IR, processed NMR, UV-Vis, MS, Raman, ESR, CD etc.

2.1.1 XYDATA= (X++(Y..Y))

This format assumes equal spacing along the x-axis. The notation (X++(Y..Y)) means that each line begins with an x-value and is followed by as many y-values as will fit on the line (McDonald and Wilks 1988, sec. 5.1.1 & 6.4.1). There are a number of checks built into this format.

```
##XYDATA= (X++(Y..Y))
```

[followed by x, y1, y2, y3... data in AFFN or ASDF]

2.1.2 XYPOINTS= (XY..XY)

This format assumes arbitrary spacing along the x-axis, but the data is essentially continuous in nature (McDonald and Wilks 1988, sec. 6.4.2). This variable list is handled the same as the next one, as they both come down to processing series of x,y points.

```
##XYPOINTS= (XY..XY)
```

[followed by x, y data in AFFN]

2.1.3 PEAK TABLE= (XY..XY)

This format assumes arbitrary spacing along the x-axis, and implies some curation of the data has occurred, e.g. only counting peaks above a threshold (McDonald and Wilks 1988, sec. 6.4.3). Data is fundamentally discontinuous. This variable list is handled the same as the previous one, as they both come down to processing series of x,y points.

```
##PEAK TABLE= (XY..XY)
```

[followed by x, y data in AFFN]

2.2 Multiple Spectra in NTUPLES Format

2.2.1 1D NMR

A processed 1D NMR spectrum composed of the real data points would be reported in one of the formats for simple files. NTUPLES is used when both the real and imaginary data are reported (Davies and Lampen 1993, sec. 7).

```
##NTUPLES= NMR
```

```
.
```

```
.
```

```
.
```

```
##PAGE= N=1
```

```
##NPOINTS= xx
```

```
##DATA TABLE= (X++(R..R)), XYDATA
```

[followed by x, y1, y2, y3... data in AFFN or ASDF]

```
.
```

```
.
```

```
.
```

```
##PAGE= N=2
```

```
##NPOINTS= yy
```

```
##DATA TABLE= (X++(I..I)), XYDATA
[followed by x, y1, y2, y3... data in AFFN or ASDF]
##END NTUPLES= NMR
##END=
```

2.2.2 2D NMR

The author is unaware of official documentation of this format, though unofficial documents can be found on the web. It is analogous to the spectral series used for LC-MS data, among others

```
##NTUPLES= nD NMR SPECTRUM
.
.
.
##PAGE= F1= x
##FIRST= x, y, z
##DATA TABLE= (F2++(Y..Y)), PROFILE
[followed by x, y1, y2, y3... data in AFFN or ASDF]
.
.
.
##PAGE= F1= x
##FIRST= x, y, z
##DATA TABLE= (F2++(Y..Y)), PROFILE
[followed by x, y1, y2, y3... data in AFFN or ASDF]
.
.
.
##PAGE= F1= x
##FIRST= x, y, z
##DATA TABLE= (F2++(Y..Y)), PROFILE
[followed by x, y1, y2, y3... data in AFFN or ASDF]
##END NTUPLES= nD NMR SPECTRUM
##END=
```

2.2.3 LC-MS or GC-MS

In this case the NTUPLES tag is used to report a page for each time format, with the spectral data in (XY..XY) format (referred to as a “spectral series” in Lampen et al. 1994, sec. 5.3.3). Processing these files requires dealing with the paging structure, but each page can be processed like an (XY..XY) data set.

```
##NTUPLES= MASS SPECTRUM
.
.
.
##PAGE= T=1
##NPOINTS= x
##DATA TABLE= (XY..XY), PEAKS
[followed by x, y data in AFFN]
.
.
.
##PAGE= T=2
##NPOINTS= x
##DATA TABLE= (XY..XY), PEAKS
```

```
[followed by x, y data in AFFN]
.
.
.
##PAGE= T=n
##NPOINTS= x
##DATA TABLE= (XY..XY), PEAKS
[followed by x, y data in AFFN]
##END NTUPLES= MASS SPECTRUM
##END=
```

3 Number Formats

- AFFN (ASCII Free Format Numeric) is x, y pairs separated by ; or any amount of white space (McDonald and Wilks 1988, sec. 4.5.3). “A field that starts with a +, -, decimal point or digit is treated as a numeric field. . . E is the only other allowed character. A numeric field is terminated by E, comma or blank. If E is followed immediately by either + or - and a two- or three-digit integer, it gives the power of 10 by which the initial field must be multiplied.”
- ASDF (ASCII Squeezed Difference Form) is a particular means of compressing data in the sequence x, y1, y2, . . . (covered in detail in McDonald and Wilks 1988, sec. 5).

4 Program Flow

readJDX is coded in such a way that it should be easy to add features. Contributions to improve or expand the package, including pull requests, are always welcome! Figure 1 shows the overall flow of the function calls. Only a couple of these functions are exported, so take a look at the source code for documentation. Be sure to check out the *MiniDIFDUP_1* and *MiniDIFDUP_2* vignettes for additional information about the JCAMP-DX file structure and how readJDX functions extract the data.

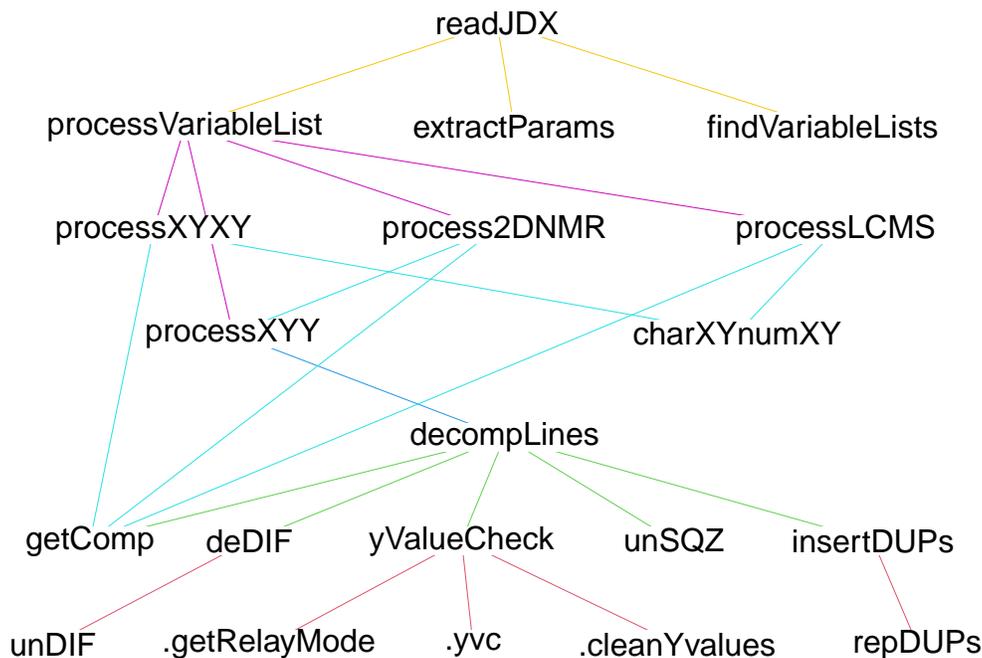


Figure 1: Functions & the flow of file processing.

References

- Davies, AN, and P Lampen. 1993. "JCAMP-DX for NMR." *Applied Spectroscopy* 47 (8): 1093–9. <https://doi.org/10.1366/0003702934067874>.
- Lampen, P, H Hillig, AN Davies, and M Linscheid. 1994. "JCAMP-DX for Mass Spectrometry." *Applied Spectroscopy* 48 (12): 1545–52.
- McDonald, RS, and PA Wilks. 1988. "JCAMP-DX, A Standard Format for Exchange of Infrared-Spectra in Computer Readable Form." *Applied Spectroscopy* 42 (1): 151–62. <https://doi.org/10.1366/0003702884428734>.