NIST Standard Reference Database 35

NIST/EPA Gas Phase Infrared Library
(in "JCAMP-DX" Format)

Users' Guide

The NIST Mass Spectrometry Data Center:

Carol Clifton
Jean Gallagher
Abdullah Shamin*
Stephen Stein
Hussein Zohdi*

*Guest scientist at NIST

May 1992 (Original Data)
December 2004 (CD-Rom and Documentation update)
April 2007 (Documentation update)

U.S. Department of Commerce
National Institute of Standards and Technology
Standard Reference Data Program
Gaithersburg, MD 20899

The National Institute of Standards and Technology (NIST) uses its best efforts to deliver a high quality copy of the Database and to verify that the data contained therein have been selected on the basis of sound scientific judgement. However, NIST makes no warranties to that effect, and NIST shall not be liable for any damage that may result from errors or omissions in the Database.

CONTENTS

INTRODUCTION

This data collection contains 5,228 gas phase infrared spectra of different compounds along with chemical structures for most of them. Spectra are provided on a CD-ROM or in a 'zip' file in JCAMP-DX (Joint Committee for Atomic and Molecular Physical Data "Data Exchange") format [1]. Chemical structures are provided in the MOL-file format [2]. The IR data originated from two sources, from the so-called "EPA Vapor-Phase IR Library" and from NIST laboratories. The data have been sub-divided in two ways, 1) as concatenated JCAMP and SDF files (concatenated MOL-files) and 2) in individual files where each spectrum and structure is provided in a separate JCAMP and MOL file, using file names containing the CAS registry number of the compound.

The data originally derived from the EPA Vapor-Phase IR Library itself (3,108 spectra) and the spectra measured at NIST (2,120 spectra) are provided in separate directories. All spectra are given as normalized absorbance and empirical formulas and CAS Registry Numbers are given for all compounds. About 80% of these compounds (and nearly all in the NIST collection) have mass spectra included in the NIST/EPA/NIH Mass Spectral Database [3].

The EPA-IR collection has already been widely distributed [4] and now exists in a number of differently edited versions on a variety of commercial data systems. The original EPA-IR collection contained 3,300 spectra measured by GC-IR under EPA contract [5]. These were obtained at 4 cm^{-1} resolution from 450 to 4000 cm^{-1}, although values below 500 cm^{-1} and above 3900 cm^{-1} often show extraneous signals. The sequence number of each spectrum in the original EPA-IR Database is provided. Each spectrum was inspected at NIST and, where possible, compared to published spectra [6] or to spectra generated at NIST. 145 erroneous or low quality spectra were excluded and 47 spectra were replaced by spectra of higher quality obtained at NIST. A complete original set of "header" information has not been located, so analytical conditions are not given. Further, we cannot be certain of the correctness of the absolute values of absorbance.

NIST spectra were obtained on an integrated capillary GC-MS-IR instrument [7] at 8 cm^{-1} resolution, most between 550 and 4,000 cm^{-1}. Maximum percent transmittance generally fell between 85% and 96%. Approximately 400 spectra present in the EPA file were remeasured at NIST; agreement was excellent in all cases. When spectra were equivalent in quality, the EPA spectrum was selected for inclusion. In all cases where a single GC run generated acceptable IR and Mass Spectra, the identification number (NIST/EPA spectrum number) of the compound in the NIST/EPA/NIH Mass Spectral Database was assigned.

Spectra included in the NIST/EPA IR data file have been baseline corrected using a single algorithm for all spectra and all have been converted to exact 8.0 cm^{-1} resolution. This was done to ensure homogeneity of the data. EPA spectra are given in the range 450 - 3966 cm^{-1}, and NIST spectra were acquired from 550 - 3846 cm^{-1}.

It is important to note that measurements by GC/IR do not provide molar absorbance values, so this data cannot be used for quantitative purposes. They are intended for use in compound identification as a source of relative absorbance information at different wavelengths for a given substance. Note also the gas phase spectra contained in this collection are generally quite different from condensed phase spectra.

IR JCAMP 2

CD Contents

cas-index.txt	*List of spectra sorted by CAS reg. no.*
formula-index.txt	*List of spectra sorted by chemical formula*
name-index.txt	*List of spectra sorted by species name*

IR — *Data folder*

- \Combined_Files
  - epa_ir_jcamp_dx.txt
  - nist_ir_jcamp_dx.txt
  - epa_ir_struct.sdf
  - nist_ir_struct.sdf
- \Individual_Files
  - \EPA — *EPA spectra (* = CAS rn)*
    - \DX — *contains 3108 *.DX files*
    - \MOL — *contains 3055 *.MOL files*
  - \NIST — *NIST spectra*
    - \DX — *contains 2120 *.DX files*
    - \MOL — *contains 2104 *.MOL files*

The spectra and structures are provided in two ways, as individual files and as combined files. The individual files contain one spectrum or structure per file, The combined files contain the same information except that multiple spectra or structures have been combined into a single file.

The individual spectra (.DX) and structure (.MOL) files have names based on the species' CAS registry number. Index files sorted by CAS registry number, chemical formula, and species name are provided to help identify the data contained in individual files. The index files are plain text files which information about the species and the relevant spectra and structure file names.

REFERENCES

1] R.S. McDonald and P.A. Wilks, Jr., A Standard Form for Exchange of Infrared Spectra in Computer Readable Form, Applied Spectroscopy, Vol. **42** (1988), 151-162

[2] Dalby A., Nourse J.G., Hounshell W.D., Gushurst A.K.I., Grier D.L., Leland B.A., Laufer J., Description Of Several Chemical-Structure File Formats Used By Computer-Programs Developed At Molecular Design Limited, Journal Of Chemical Information And Computer Sciences 32 (3): 244-255 May-Jun 1992

[3] http://www.nist.gov/srd/nist1a.htm

[4] M.D. Erickson, Applied Spectroscopy, Vol. **35**, 181 (1981).

[5] EPA-commissioned spectra were obtained at Sadtler Research Laboratories on Digilab Instruments.†

[6] "The Aldrich Library of FT-IR Spectra, Vol. **3**, Vapor Phase", Charles J. Pouchert, editor, The Aldrich Chemical Company (1989).

[7] All NIST spectra were obtained on a Hewlett-Packard FT-IR (IRD-5965).†

†References to these instruments are not to be construed as an endorsement for the products by NIST or by the Standard Reference Data Program.

Appendix A

FILE IN JCAMP-DX FORMAT

```
##JCAMPDX=4.24
##DATA TYPE=INFRARED SPECTRUM
##ORIGIN=EPAIR VAPOR PHASE LIBRARY
##OWNER=NIST OSRD
##CAS REGISTRY NO=99081
##MOLFORM=C 7 H 7 N O 2
##DELTAX=6.000000
##XUNITS=1/CM
##YUNITS=ABSORBANCE
##XFACTOR=1.000000
##YFACTOR=1.000000
##FIRSTX=500.000000
##LASTX=3980.000000
##NPOINTS=580
##XYDATA=(X++(Y..Y))
500.000 53 50 35 40 41 31 30 37 34 34
560.000 28 31 34 25 30 28 25 30 30 33
620.000 30 29 32 31 49 76 137 250 378 407
680.000 319 160 86 63 102 260 542 760 1423 456
740.000 183 80 54 57 76 107 191 397 1047 1417
800.000 1105 557 125 49 47 53 54 50 48 53
860.000 63 91 124 156 156 185 216 221 195 140
920.000 128 83 65 48 41 43 39 37 36 39
980.000 38 39 43 44 44 43 38 39 47 57
1040.000 75 84 99 123 173 224 246 297 338 348
1100.000 272 154 74 43 38 41 46 45 47 52
1160.000 50 48 41 39 43 46 49 55 56 54
1220.000 48 47 45 45 48 56 69 94 132 187
1280.000 260 298 308 265 229 210 244 338 567 1024
1340.000 1863 3168 3631 3082 1462 570 319 239 211 169
1400.000 142 124 115 106 115 136 149 163 163 165
1460.000 168 222 329 392 425 401 360 357 415 625
1520.000 1176 2231 3533 3992 3961 3047 2113 1610 1146 775
1580.000 619 540 451 340 245 174 129 98 88 67
1640.000 68 69 60 57 54 54 52 37 38 30
1700.000 31 45 51 52 47 31 28 33 34 40
1760.000 44 45 43 45 45 51 60 55 46 32
1820.000 22 18 20 24 29 34 35 29 22 24
1880.000 29 45 48 42 28 16 16 17 20 30
1940.000 39 41 40 29 18 14 14 14 15 13
2000.000 15 12 10 9 10 6 8 7 7 9
2060.000 11 8 9 9 10 6 9 9 6 9
2120.000 9 8 11 8 8 11 12 12 13 13
2180.000 12 11 10 7 8 8 8 7 6 5
2240.000 9 11 14 12 13 12 10 14 16 18
```

2300.000 19 19 20 26 28 27 25 25 23 20
2360.000 16 11 14 17 15 14 12 8 9 8
2420.000 10 14 16 21 21 21 18 14 10 7
2480.000 5 6 6 5 2 3 1 0 1 0
2540.000 1 0 3 5 11 16 16 15 13 14
2600.000 14 11 11 9 8 8 9 12 9 9
2660.000 11 11 14 17 16 20 21 19 20 15
2720.000 12 8 12 17 22 25 17 14 14 11
2780.000 17 20 23 22 23 24 27 32 37 44
2840.000 51 59 74 102 138 175 195 193 177 164
2900.000 163 173 194 243 312 350 342 292 231 179
2960.000 144 119 101 94 103 119 138 154 164 163
3020.000 180 218 248 243 199 153 150 192 240 246
3080.000 222 148 113 83 61 38 24 16 8 10
3140.000 5 8 4 8 8 5 6 7 6 7
3200.000 10 10 7 8 8 11 8 10 7 9
3260.000 11 9 6 7 8 10 10 11 10 9
3320.000 9 11 7 10 11 9 10 7 7 11
3380.000 7 7 11 8 8 11 9 11 11 9
3440.000 9 12 13 10 8 8 10 10 9 9
3500.000 10 8 11 11 5 9 10 12 13 11
3560.000 11 10 9 11 12 8 13 14 9 10
3620.000 14 11 10 11 0 12 12 15 14 9
3680.000 13 10 13 16 14 16 16 16 16 10
3740.000 2 3 9 16 15 17 15 12 12 7
3800.000 7 11 8 10 9 11 10 4 8 5
3860.000 4 3 4 9 11 13 11 8 10 10
3920.000 9 8 7 7 9 15 12 10 12 12
##END=

Appendix B

CONTACTS

If you have any questions or comments, the Standard Reference Data Program would like to hear from you. Please contact:

Joan Sauerwein
National Institute of Standards and Technology
Standard Reference Data
Mail Stop 2300
Gaithersburg, MD 20899-2300
Phone: (301) 975-2208
FAX: (301) 926-0416
Internet: http://srdata.nist.gov/contact/contact
http://www.nist.gov/srd/nist35.htm