Pure & Appl. Chem., Vol. 67, No. 4, pp. 593–596, 1995.
Printed in Great Britain.

INTERNATIONAL UNION OF PURE
AND APPLIED CHEMISTRY

PHYSICAL CHEMISTRY DIVISION
COMMISSION ON MOLECULAR STRUCTURE AND SPECTROSCOPY*
WORKING PARTY ON COMPUTERIZED NMR DATABASES†

GUIDELINES ON NUCLEAR MAGNETIC RESONANCE COMPUTERIZED DATABASES

(IUPAC Recommendations 1995)

Prepared for publication by

CHARLES L. WILKINS

University of California, Riverside, CA 92521, USA

*Membership of the Commission during the preparation of this report (1985–94) was as follows:

Chairman: J. R. Durig (USA; 1983–89); C. J. H. Schutte (RSA; 1989–93); J. E. Bertie (Canada; 1993–97); *Secretary*: H. A. Willis (UK; 1983–87); J. G. Grasselli (USA; 1987–89); J. F. Sullivan (USA; 1989–95); *Titular Members*: A. M. Bradshaw (FRG; 1983–91); R. D. Brown (Australia; 1987–93); B. G. Derendjaev (Russia; 1983–91); R. K. Harris (UK; 1989–93); A. M. Heyns (RSA; 1993–97); E. Hirota (Japan; 1985–93); R. Janoschek (Austria; 1993–97); P. Klaeboe (Norway; 1993–97); S. Tsuchiya (Japan; 1993–97); B. P. Winnewisser (FRG; 1991–93); J. F. J. Todd (UK; 1983–89); *Associate Members*: A. M. Bradshaw (FRG; 1991–93); B. G. Derendjaev (Russia; 1991–93); P. Granger (France; 1989–93); R. K. Harris (UK; 1981–89); S. Leach (France; 1987–91); G. J. Martin (France; 1983–89); J. P. Maier (Switzerland; 1989–93); L. N. Masalov (Russia; 1987–91); A. Oskam (Netherlands; 1991–95); W. B. Person (USA; 1983–89); C. J. H. Schutte (RSA; 1983–89); J. F. J. Todd (UK; 1989–91); S. Tsuchiya (Japan; 1987–93); C. Zhang (China; 1989–93); *National Representatives*: J. E. Collin (Belgium; 1985–93); S. M. Cabral de Menezes (Brazil; 1991–93); S. L. Spassov (Bulgaria; 1987–93); J. Lu (China; 1983–89); E. Hadjoudis (Greece; 1991–93); G. Varsányi (Hungary; 1985–93); M. Chowdhury (India; 1986–93); Z. Luz (Israel; 1983–93); S. Califano (Italy; 1989–93); S. Ng (Malaysia; 1983–93); A. Oskam (Netherlands; 1989–91); B. Jezowska-Trzebiatowska (Poland; 1983–89); J. J. C. Teixeira-Dias (Portugal; 1991–93); Y. S. Lee (Republic of Korea; 1989–93); T. A. Ford (RSA; 1987–91); C. Pascual (Spain; 1987–91); D. Escolar (Spain; 1991–93); M. Cebe (Turkey; 1987–91); S. Suzer (Turkey; 1991–93); T. Ast (Yugoslavia; 1989–93).

†Membership of the Working Party:

Chairman: C. L. Wilkins (USA; 1985–94); *Members*: W. Bremser (FRG; 1987–94); B. G. Derendjaev (Russia; 1987–94); J. E. Dubois (France; 1987–94); M. Elder (UK; 1985–87); S. R. Heller (USA; 1987–94); O. Mills (UK; 1985–87); J. R. Mooney (USA; 1985–94); W. Robien (Austria; 1987–94); S.-I. Sasaki (Japan; 1987–94); Shelley (USA; 1985–87); N. Zumbulyadis (USA; 1985–87).

Guidelines on nuclear magnetic resonance computerized databases (IUPAC Recommendations 1995)

Synopsis

With the widespread use and availability of computers for the acquisition and storage of nuclear magnetic resonance spectra, particularly proton and carbon spectra, it is clear that assembly and distribution of computer-readable nmr databases is highly desirable. To maximize the utility of such efforts, it is desirable to define a standardized format for information exchange and guidelines for information to be included in such databases. An equally important concern is the need for nmr databases (indeed, all spectral databases) to include information which will allow users to determine the quality of the spectra. In 1984, at a workshop on the subject of computerized data bases held in the United States at the National Bureau of Standards (now the National Institute of Standards and Technology) , it was concluded that a consensus body, such as the IUPAC, should organize and establish recommendations for such databases.

As one result of that recommendation, a "Working Party on NMR Computerized Databases" was established, with the charge of producing recommendations of guidelines for such databases. The present document is the result of the working party's efforts, is the consensus of its members, and constitutes the recommendation of the working party.

Background and Recommendation

It was the position of the committee members from the outset that, to the degree possible, the guidelines developed should take advantage of the previous work in the field, particularly with respect to recommended primary database format. Because an extensive international effort in carbon NMR data exchange has been underway for over ten years (1), it seemed logical to begin with the consideration of the database format adopted for that project, a format which has come to be known as the "Bremser format" (3,4). That effort resulted in acceptance of the basic premise that, for greatest generality and maximum data portability, an ASCII character-oriented format should be used. Such an approach maximizes portability and ease of data transport from one computer system or NMR spectrometer to another.

Subsequent data transportability efforts in related fields by the Joint Committee on Atomic and Physical Data (JCAMP) (5) and collaborators (particularly in the areas of infrared spectroscopy and mass spectrometry) have resulted in an emerging consensus that an excellent vehicle for such exchange is the JCAMP-DX computer-readable exchange format originally defined in detail for infrared spectral data exchange. That format, defined as JCAMP-DX 4.24 (6) includes all the necessary features to specify the format of the "Notes" portion of NMR data files which would include necessary spectral identification information, physical properties, and related information, as already delineated for infrared spectral data (modified of course to accommodate the differences between NMR and infrared spectrometry experiments). Similarly, an organic structural information exchange format, called the Standard Molecular Data (SMD) format is currently being developed by the IUPAC Committee on Chemical Databases(CCDB) and others in collaboration with the United States of America standards organization ASTM. A preliminary version of this structure format has been published (7). In the next year or two it is expected that this format will undergo some modification and will then be adopted by the major organizations in the field, such as Chemical Abstracts, the Beilstein Institute, and Molecular Design, Ltd. Accordingly, it is our recomendation that this one evolving standard, SMD, when completed and approved by CCDB for IUPAC, be used for the structural portion of NMR data exchange. Finally, the JCAMP Executive Council has now approved a JCAMP-DX NMR (8), which specifically addresses NMR data exchange within the context of the JCAMP standard, version 5.0, (a modified format based upon JCAMP 4.24), to be released in the near future (9). Three major NMR manufacturers, Bruker, Varian, and Jeol, are committed to implementing the standard in their software. The Working Party has reviewed these standards and concludes that all of the requisite elements now exist for a versatile NMR database standard, implementing the JCAMP framework, as modified by use of JCAMP-CS and JCAMP-NMR. Accordingly, it is our recommendation that the IUPAC officially endorse use of these standards for NMR database data exchange. It should be noted that the IUPAC Interdivisional Committee on Nomenclature and Symbols approved JCAMP-DX for publication in 1991(10).

General Principles

1. General Remarks - Although it is recognized that for particular applications (e.g. spectral display or computerized spectral searches) highly compact and efficient secondary database structures are required, the archival or "exchange" format should be character based and as complete as possible, allowing individuals to utilize those elements of the information required for their particular applications, without compromising the quality of the data or placing unnecessary constraints upon possible applications. It is absolutely essential that the database contain information regarding the quality of the sample from which each spectrum is derived.

1.1 Inclusion of Compounds - One should be as complete as possible when spectra are drawn from a particular literature reference (i.e. include all spectral information from the reference). In general, spectra should contain, at a minimum, 50% of the expected information (i.e. line frequencies, relaxation times, etc.). Spectra from compounds with unverified structure should not be included. Ideally, sample authenticity should be verified by independent spectral measurements such as infrared and/or mass spectrometry. When that is done, that fact should be included in the database entry.

1.2 Data Input - Information should be coded in the general character-oriented format (ASCII) described in the JCAMP-DX standard. Where appropriate (e.g. compound names, authors, and journal references intended for eventual output) upper and lower case representations conforming to normal usage should be included.

2. Tape Format - No label, fixed record length of 80 bytes, unblocked, ASCII code (upper and lower case)

3. Notes Section of JCAMP Files

3.1 Names - The name should be entered in upper and lower case. Compounds with unknown names, in general, should not be included. If they are, descriptive information should be placed here.

3.2 Formulas - Molecular formulas (Hill system) composed of 1-2 element symbols, 1-3 digit numbers of atoms, and separating blanks should be included. For example, C15 H17 Br5 O1. Upper and lower case should be used, as appropriate. Note that a single atom must be coded (e.g. O1).

3.3. Registry Numbers and Molecular Weights - Chemical Abstracts registry number and molecular weight (in units of 0.01 dalton, mass spectral value for composition of molecule comprised of the most abundant isotopes).

3.3 References - All authors with initials for all first names, upper and lower case, as appropriate. Journal citations following Chemical Abstracts, however with year at end. For example, P.A. Gorin, M.Mazurek, Can.J.Chem. 53,1212 (1975). If unpublished, that fact should be entered.

3.4 Comments - For alphanumeric comments on compound (configuration), solvent (pH, sample concentration), experiment (sample preparation, instrument calibration (how and when)), similar references, etc. Characterize reference numbers with #, optical centers with *, and cross references with # or < >. Use Upper and lower case, as appropriate. Examples:
*17R,16S
Cf.Ref.#BAS00047 for other isomer.
<S 0041>
Use as many comment records as appropriate.

3.5 Experimental Notes - Instrument , solvent , concentration, original standard, conversion shift , temperature in kelvin (if reported), quality index (QI, per Bremser (4)).

References

1. D.L.Dalrymple, C.L.Wilkins, G.W.A. Milne, and S.R. Heller, *Org. Magn. Reson.*, **11**, 535-540 (1978)
2. C.W.v.d.Lieth, J.Seil, I. Kohler, and H.J.Opferkuch, *Magn. Reson. Chem.*, **23**, 1048-1055 (1985).
3. W. Bremser and W. Fachinger, *Magn. Reson. Chem.*, **23**, 1056-1071 (1985).
4. W. Bremser and W. Fachinger, *Magn. Reson. iChem.*, **24**, 183 (1986).
5. Chairman: Edward G. Brame, Jr., The CECON Group, Inc., 242 N. James Street, Wilmington, DE 19804.
6. R.S. McDonald and P.A. Wilks, Jr., *Appl. Spectrosc.*, **42**, 151 (1988).
7. J.M., Barnard, *J. Chem. Infor. Comp. Sci.*, 30, 81 (1990).
8. A. N. Davies and P. Lampen, "JCAMP-DX for NMR", *Appl. Spectrosc.*, **47**, 1093 (1993).
9. R.S. McDonald, "Generic JCAMP-DX 5.0, Draft 1.0 February 28, 1991", JCAMP-DX Subcommittee, 9 Woodside Drive, Burnt Hills, NY 12027, USA.
10. J.G. Grasselli, "JCAMP-DX, A Standard Format for Exchange of Infrared Spectra in Computer Readable Form", *Pure & Appl. Chem.*, **63**, 1781 (1991).