ELSEVIER

Journal of Molecular Structure 485–486 (1999) 249–254

Journal of
MOLECULAR
STRUCTURE

MOGADOC—a database for gasphase molecular spectroscopy and structure

J. Vogt*, B. Mez-Starck, N. Vogt, W. Hutter

Sektion für Spektren-und Strukturdokumentation, Universität Ulm, D-89069 Ulm, Germany

Dedicated to Prof. L.S. Bartell on the occasion of his 75th birthday

Received 7 December 1998; accepted 16 December 1998

Abstract

The MOGADOC database (Molecular Gasphase Documenation) deals with compounds which were studied in the gasphase by means of electron diffraction, microwave spectroscopy or radio astronomy. For the first time, the Windows™ version of this valuable retrieval tool with its new user-friendly graphical input interfaces and additional retrieval features is described. Presently the database contains over 25 000 bibliographic references for about 7600 inorganic, organic and organometallic compounds including numerical datasets for bond lengths and angles for about 4000 compounds.

Keywords: MOGADOC database; Molecular structure; Gas electron diffraction; Microwave spectroscopy; Radio astronomy

1. Introduction

For many years the MOGADOC database (Molecular Gasphase Documentation) has been a powerful tool for scientists all over the world to retrieve information about gasphase investigations. Covering literature on laboratory as well as atmospheric or astronomical work it provides the user with a fast and easy access to the literature of each field [1–4].

The information has been compiled worldwide from scientific journals, periodicals, books, and grey literature. The literature (see Fig. 1) is recorded back to:

- 1930 for gasphase electron diffraction,
- 1945 for microwave spectroscopy and
- 1960 for molecular radio astronomy.

The next update, which is scheduled for the end of 1998, will comprise over 25 000 bibliographic references for about 7600 inorganic, organic and organometallic compounds. Moreover, it will list numerical values for bond lengths, bond angles and dihedral angles for about 4000 compounds.

The purpose of this publication is to announce the availability of a new version of this database for MS Windows with new retrieval features. In the recent literature only the MS-DOS version was described.

2. Retrieval features

The MOGADOC database consists of two files, a bibliographic one (MGDLIT) for references and a file (MGDCOM) with compound descriptions. The following retrieval options are available:

* Corresponding author. Tel.: + 49-731-50-31050; fax: + 49-731-50-31059.

E-mail address: juergen.vogt@chemie.uni-ulm.de (J. Vogt)

0022-2860/99/$ - see front matter
PII: S0022-2860(99)00046-0

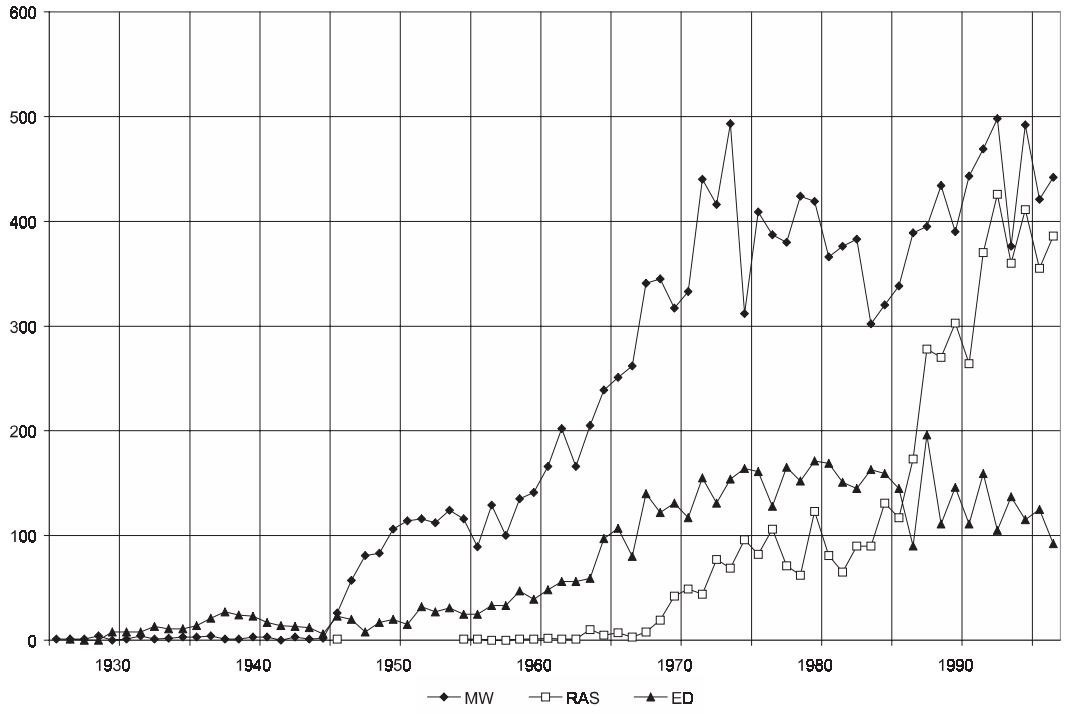


Fig. 1. Publication frequency of papers in the field of gas electron diffraction (ED), microwave spectroscopy (MW) and molecular radio astronomy (RAS).

- bibliographic search terms
- keywords (controlled hierarchical thesaurus)
- compound names and chemical name segments
- gross formula (Hill system)
- CAS Registry numbers
- element formula, total atom count and periodic groups
- individual element counts for C, H, N, O, P, S and F
- numeric values for bond lengths and angles
- nested Boolean expressions

The database is available as an in-house version on IBM compatible PCs under the Windows™ operating system and runs as an application of the Professional File System 2000™. Graphical input interfaces and user-friendly search forms provide an easy access to the data without requiring knowledge of a certain command syntax. Users familiar with the Messenger syntax (known from retrievals in on-line STN databases such as Chemical Abstracts, INSPEC, etc.) can, however, use a command line interface for sophisticated queries.

The information is stored in different search fields. In the search forms the explanations of these search fields are given in front of each input box (Fig. 2). Generally the search fields are combined by the Boolean operator ''AND''. However, the ''OR'' or ''NOT'' operator can be easily selected by clicking the corresponding drop-down menu. Within the numeric search fields numeric operators ($<$, $\leqq$, $>$, and $\geqq$) are allowed.

The last column gives the partial results for each search field in relation to the total database, whereas the last line informs how many documents fulfill the total logical requirement.

In order to retrieve similar expressions, variable beginnings and/or endings of search terms are allowed. In the example the truncation symbol? is standing for any number of characters, whereas the symbol # is a variable for zero or one character. Simultaneous right and left truncation is allowed.

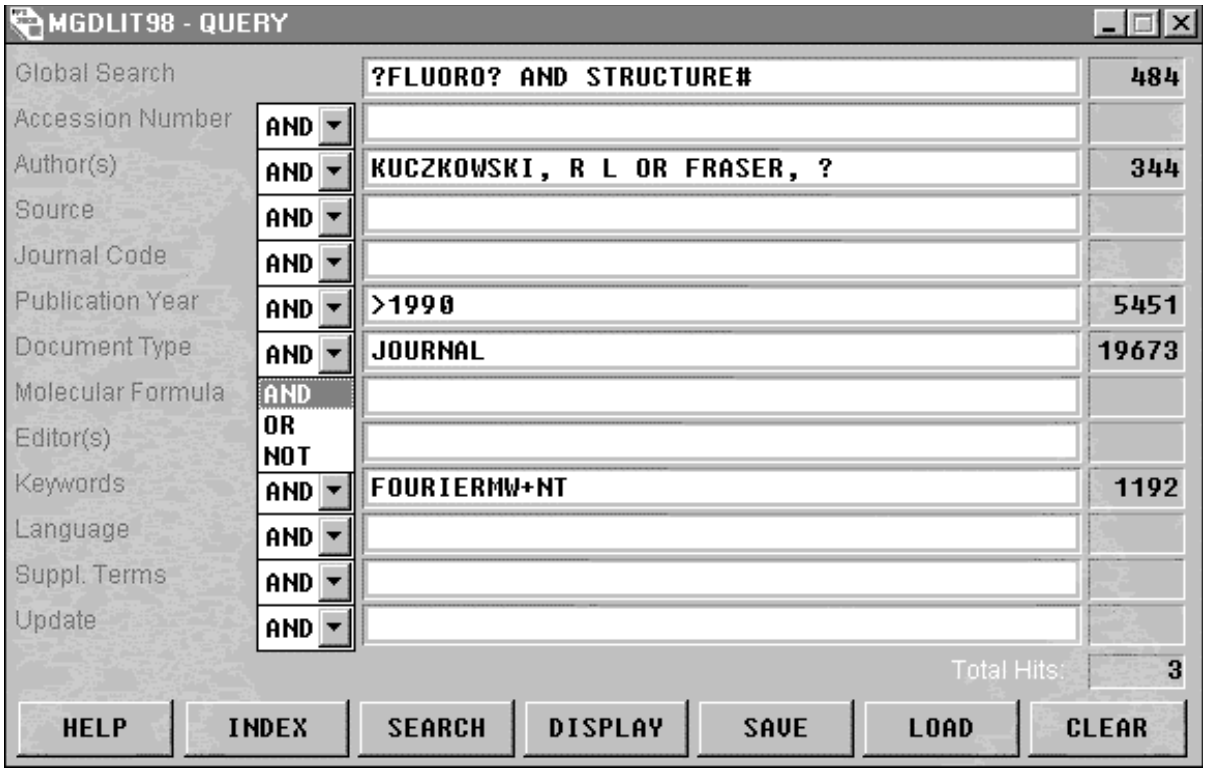


Fig. 2. Tabular form for the bibliographic file MGDLIT with a retrieval example.

Keywords, which were selected by reviewers, describe the content of the references in much more detail than titles can do. Therefore searches by means of keywords are more promising. These keywords come from an hierarchical thesaurus with a controlled vocabulary, which has been developed over many years and is the result of extensive and very valuable interactions with the community of prospective users. In the example in Fig. 2 the keyword ''FOURIERMW'' demands for references dealing with Fourier transform microwave spectroscopy. By adding the relationship code ''NT'' (for narrower terms) all documents with more specific keywords such as ''Fourier transform microwave spectroscopy of molecular beams'' are automatically included in the retrieval.

In order to facilitate the application of the thesaurus, which consists of almost 2000 controlled keywords, the retrieval program enables the user to interactively browse the keywords according to topics.

Last but not the least, a fuzzy search option will be introduced in the next program update (scheduled for the end of 1998) as an alternative to retrieve similar expressions. The amount of resulting documents can be controlled by allowing up to four deviations in each retrieved search term from a given input.

One of the resulting retrieved documents is shown in Fig. 3. The keywords are given as acronyms (which are also used as input for the retrieval) as well as the full text explanations. The group of keywords is attributed to the gross formula and the MOGADOC specific compound number of the compound studied in that paper. The underlined compound number contains a hyperlink to the compound file MGDCOM which enables the quick display of the identification of that compound including compound names, structural formula and Chemical Abstracts registry numbers.

The compound file MGDCOM also has a similar user-friendly search form which allows bibliographic (names, synonyms, name segments, gross formulas,

AN	70894		
AU	Xu, L. W.; Klausner, M. E.; Andrews, A. M.; Kuczkowski, R. L.		
TI	**Structure of hexafluorocyclobutene.**		
SO	J.Phys.Chem., 97, 10346-10348		
CO	49		
PY	1993		
DT	Journal		
LA	English		
IN 1	**CN**	**526266**	
	MF	**C4 F6**	
	KW	TWOFOLD	asymmetric top with twofold axis
		NEARSYM	nearly symmetric top
		C2V	C2v symmetry
		MW	microwave or millimeter wave spectroscopy
		FOURIERMW	Fourier transform microwave spectroscopy
		MWRMW	microwave region: 2 - 100 GHz
		RCEGRST	rovibrational constants: ground state
		RCOCFD	rotational constants (nonrigid rotor)
		RCOALL	all rotational constants determined
		CDQCEXSP	quartic centrifugal distortion constants determined
		RCISOT	rotational and rovibrational constants of isotopic species
		SEG	structure in ground state
		SI	moment of inertia equation used for structure determination
		DIST0	r(0) distance(s)
		DISTSK	r(s) distance(s) from Kraitchman equations
		ANGL0	r(0) angle(s)
		ANGLS	r(s) angle(s)
		GEOGIV	structural parameters given in MOGADOC
		COMPOMETH	structure compared with that by other methods
		ISOTOPEEFF	isotope effect used for structure determination
		DPEGRST	dipole moment in ground vibrational state
		DPESHI	dipole moment from the shifts of Stark components
		STKTEE	quadratic Stark effect
		13E	carbon 13 (enriched sample)

Fig. 3. Printout of one of the resulting bibliographic entries which fulfill the logical requirement in Fig. 2.

registry numbers, element occurrence, periodic groups, etc.) and numerical (number of atoms, individual element counts, internuclear distances, bond angles, etc.) retrievals. Fig. 4 shows the full display of the compound which is briefly mentioned in Fig. 3. By means of hyperlinks to the bibliographic file MGDLIT the ten most recent references can be displayed.

3. Hardware and software requirements

The MOGADOC database, which is updated once a year, is distributed worldwide as CD-ROM. The following system requirements must be fulfilled:

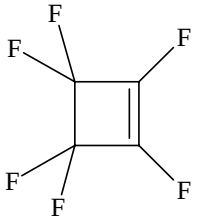
- IBM-compatible PC 486 or higher
- VGA equipment, minimum 800 × 600 resolution
- at least 4 MB RAM
- Windows™ 3.1 or higher
- CD-ROM drive

More information and a demo version with 50 bibliographic and 50 compound entries, which comes rather close to the new Windows™ user interface and which shows the functionality of the database, is available in the World Wide Web (http://www.uni-ulm.de/strudo/mogadoc/).

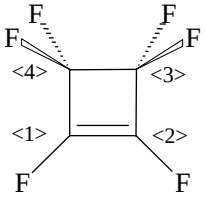
Acknowledgements

The database structure, computational details, and retrieval features have been developed over many years and are the result of extensive and very valuable interactions with the community of prospective users. Therefore the authors would like to express their sincere thanks to Prof. H.K. Bodenseh, Prof. I. Botskor, Dr. D. Boucher, Dr. B.N. Cyvin, Prof. S.J. Cyvin, Dr. J. Demaison, Prof. H. Dreizler, Prof. J.L.

CN 526266
MF **C4 F6**
RN 697-11-0
NA **Hexafluorocyclobutene**
LC 70894



Method of study : Microwave Spectroscopy



Refined geometrical parameters

Errors of values are parenthesized in units of the last digit.

Structure type for distances: r(s)

Parameter(s)	Value(s) (Angstroem)	Note(s)
C1-C2	1.3326	*
C2-C3	1.4799	
C3-C4	1.5542	

Note(s):
*) no error limits given

Structure type for angles: r(s)

Parameter(s)	Value(s) (degrees)	Note(s)
C1-C2-C3	94.29	*
C1-C4-C3	85.71	

Note(s):
*) no error limits given

Other relevant information:

C(2v) symmetry.

r(0) structure also given in that paper.

Structure determination by L.-W. Xu et al. (1993); (AN: 70894)

Fig. 4. Printout of one of the corresponding compound entries. The example lists the identification of the compound (gross formula in Hill system, compound name, Chemical Abstracts registry number) and numeric parameters for internuclear distances and bond angles.

Duncan, Prof. I. Hargittai, Prof. M. Hargittai, Prof. W. Hüttner, Prof. H. Jones, Dr. Z. Kisiel, Dipl. Chem. R. Kramer, Dipl. Math. A. Lohr, Dr. A.H. Lowrey, Dipl. Chem. R. Mutter, Prof. H. Oberhammer, Prof. H.D. Rudolph, Prof. A. Ruoff, Dr. K. Volka, Dr. D.G. Watson, Dr. B.P. Winnewisser, Prof. M. Winnewisser, Dr. G. Wlodarczak, and Dr. K.M.T. Yamada for their stimulating discussions and valuable contributions. The project was supported by Fachinformationszentrum in Karlsruhe (the European host of STN International) and the software house Kramer & Hofmann in Saarbrücken.

References

[1] A. Lohr, B. Mez-Starck, H.G. Schirdewahn, D.G. Watson, J. Mol. Struct. 97 (1983) 57.
[2] J. Vogt, Rev. Mex. Astron. Astrofis. 23 (1992) 119.
[3] J. Vogt, Struct. Chem. 3 (1992) 147.
[4] J. Vogt, J. Mol. Spectrosc. 155 (1992) 413.