Journal of Molecular Structure, 97 (1983) 57—60
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

MOGADOC (MOLECULAR GAS PHASE DOCUMENTATION) - AN INTERACTIVE COMPUTERISED SEARCH/RETRIEVAL SYSTEM

A. LOHR[1], B. MEZ-STARCK[2], H.-G. SCHIRDEWAHN[1], and D.G. WATSON[3]
[1]Rechenzentrum der Universität Freiburg, D-7800 Freiburg (F.R.G.)
[2]Sektion für Strukturdokumentation der Universität Ulm, D-7900 Ulm (F.R.G.)
[3]Chemistry Department, University of Cambridge (U.K.)

ABSTRACT

For 22 years the Section for Structure Documentation (SSD), University of Ulm, has provided a documentation service covering the literature of molecules studied in the gas phase by electron diffraction, microwave spectroscopy and other techniques. Much of the information which has been accumulated over these years has now been keyboarded to constitute a computerised database. An interactive search/retrieval system, MOGADOC, has been written using the SIMULA* language implemented on the UNIVAC 1100/82 computer at the University of Freiburg. MOGADOC enables the user to search the database on the basis of bibliographic, chemical and physical search terms.

In recent years the SSD has increasingly been asked questions which are difficult and time-consuming to answer using the traditional manual card index system. As examples, can you provide literature references for papers describing the joint use of gas ED and MW spectroscopy; isotope effect in ED studies; computer programs used in ED work; hot nozzles? Such queries can really only be tackled using a computerised search/retrieval system. To this end the MOGADOC database system was established (ref.1).

As shown on the next page, MOGADOC consists of three principal components - literature, investigations and compounds. In fact there is a fourth component, topology, which contains chemical connectivity records for the compounds and which can be linked to the compounds component. The topology component will allow us to search for complete or partial structures, eg. C-N-C, on the basis of chemical connectivity. The software necessary for topological searching is in the final testing stages.

It should be noted that one paper may contain more than one investigation - thus the structures of 2 molecules in a paper would define 2 investigations.

---

* SIMULA is claimed to be the only general purpose high level programming language which is reasonable wide spread and allows direct implementation of abstract data types.

0022-2860/83/$03.00

LITERATURE (lit)	INVESTIGATION (inv)	COMPOUNDS (com)
AUTHOR (aut)	PHYS.DESCRIPTOR(S) (phy)	COMPOUND NAME
INITIALS	COMPOUND NUMBER	SYNONYM(S)
REFERENCE	PUBLICATION NUMBER	CHEM. FORMULA (for)
JOURNAL (jou)	INVESTIGATION NUMBER (inu)	CHEM.DESCRIPTORS (che)
YEAR (yea)		COMPOUND NUMBER (cnu)
PUBLICATION TYPE (pty)		CHEM.DIAGRAM NUMBER
EDITOR (edi)		
PUBLICATION NUMBER (pnu)		

In the above table the broken lines show how the 3 components are linked. Terms which can be searched are underlined, the others can be displayed or printed.

The physical results reported in a paper are represented by physical descriptors, each one identified by a mnemonic, eg. ED for electron diffraction. In total there are some 1600 physical descriptors, grouped into 24 classes: -

0 MOLECULE	12 LARGE AMPLITUDE MOTION
1 METHODS	13 OTHER TECHNIQUES MAINLY LASER
2 SPECTRUM AND ASSIGNMENT	14 ED-THEORY, PROCEDURE, REVIEWS
3 ROT.CONSTANTS AND ROT-VIBR.INTERACTIONS	15 ED-TECHNIQUES
4 STRUCTURE	16 THEORETICAL MOLECULAR CALCULATION
5 FINE STRUCTURE	17 SPECIFIC ISOTOPES
6 ZEEMAN EFFECT (DIAMAGNETIC)	18 LINEAR PARAMAGNETIC MOLECULES
7 MICROWAVE TECHNIQUE	19 NONLINEAR PARAMAGNETIC MOLECULES
8 DIPOLE MOMENT (ALL METHODS)	20 ZEEMAN EFFECT (PARAMAGNETIC)
9 STARK EFFECT	21 PRODUCTION OF UNSTABLE SPECIES
10 COLLISIONAL EFFECT	22 ASTROPHYSICAL PROCESSES
11 INTENSITY MEASUREMENTS, LINE SHAPE	23 FORCE CONSTANTS

In addition to its name and formula the chemical structure of a compound is represented by fragment codes known as chemical descriptors, each one identified by a mnemonic, eg. RH1 for 1 hetero-atom in a ring. Nearly all of these chemical descriptors can be computer-generated from the topological descriptions of molecules. There are some 700 chemical descriptors grouped in 8 classes: -

1 MOLECULAR UNIT	5 BOND-CENTRED FRAGMENTS
2 ELEMENTAL CONSTITUTION	6 GROUPS, LIGANDS, SUB-UNITS
3 SIMPLE COMPOUNDS ($\leq$7 ATOMS)	7 METAL COMPLEXES
4 RINGS, CAGES, CLUSTERS	8 INORG.ACIDS, SALTS, RELATED CPDS.

Both the physical and chemical descriptor lists are hierarchical in nature. Thus, for example, physical list 15 and chemical list 3 contain: -

⋮	
15.5 diffraction chamber	
15.5.1 registration	
15.5.1.1 photographic	
15.5.1.2 non-photographic	
⋮	

⋮	
	3.2 triatomic
	3.2.1 X_3
	3.2.2 XY_2
	3.2.3 XYZ
	⋮

The examples which follow illustrate the interactive dialogue between user (U) and computer (C): -

1. Find literature citations for papers published after 1975 by Almenningen
   U: Search literature where author = Almenningen and year >= 1975
   C: 24 publications found
   U: display (or print)
   C: publications are shown by
   publication number, authors, reference, title
   To save keystrokes the user can simply type the first 3 letters of certain terms (see below)
2. Find literature citations for ED studies of molecules having no multiple rings but which contain a 5-membered ring with 1 hetero-atom.
   U: sea com whe che = R5 and che = RH1 and not che = RM and phy = ED
   C: 35 compounds found (query 1)
   U: display (or print)
   C: compounds are shown by
   compound number, sum formula, chem.diagram no., name(s)
   U: sea lit whe cnu = 22372 or cnu = 93714
   C: 5 publications found (query 2)
   U: display (or print)
   C: publications are displayed or printed with publication number, title etc.
3. I would like to compare structures for which empirical correlations have been reported.
   U: sea inv whe phy = TCSEMPCOR
   C: 39 investigations found (query 1)
   U: print
   C: publication number, investigation number, compound number
   U: sea lit whe result
   C: 32 publications found (query 2)

U: print

C: publications are printed with publication number, title etc.

U: sea com whe 1

C: 37 compounds found (query 3)

U: print

C: compounds are printed with compound number etc.

With these three print-outs the user will be able to choose what he wants to compare.

The current status of the contents of the 3 main MOGADOC components is: -

lit: total 27000; ED is current, MW and other techniques to 1981

inv: total 3800; ED is current (ref.2)

com: total 4200; ED is current, MW to 1981, chem.descriptors for ED, MW to 1981

Organic topology records are in computer form and inorganic will be added later. A further 6800 papers have been indexed - 2500 MW, 3000 force constants, 1300 others. The indexing of ED, MW, force constants, diatomics, quantum chem. structures, mean amplitudes will continue and efforts will be made to find indexers for high-resolution IR / rotational Raman studies.

ACKNOWLEDGEMENT

We would like to thank coworkers and indexers.

We would finally like to thank, for financial support, the German Ministry for Research and Technology which has sponsored this work via the German Society for Information and Documentation, Frankfurt in cooperation with the Fachinformationszentrum Energie, Physik, Mathematik at Karlsruhe.

REFERENCES

1 H.-G. Schirdewahn and B. Mez-Starck, Data for Science and Technology, 227 (1981), Proceeding of 7th Int. CODATA Conf., Kyoto, 1980, P.S. Glaeser, Ed., Pergamon Press, Oxford.

2 I. Buck, E. Maier, R. Mutter, U. Seiter, C. Spreter, B. Starck, I.Hargittai, O. Kennard, D.G. Watson, A. Lohr, T. Pirzadeh, H.G. Schirdewahn, Z. Majer, Bibliography of gas phase electron diffraction 1930 - 1979, Physics Data, Nr. 21-1 (1981), Fachinformationszentrum Energie, Physik, Mathematik GmbH, Karlsruhe, D-7514 Eggenstein-Leopoldshafen 2.