SpecInfo: An Integrated Spectroscopic Information System

Andreas Barth

STN International, FIZ Karlsruhe, D-7500 Karlsruhe, FRG

Received August 18, 1992

The online version of the SpecInfo system consists of a factual database with spectroscopic information and a set of special tools for the spectral data analysis and interpretation. It has been designed to support experts in laboratories in their analytical work. The database contains high-quality spectral data of various types, experimental conditions, chemical substance information including the connection tables, and the bibliographic source data. In the first release, the database comprises approximately 70 000 ^{13}C NMR, 17 000 infrared, and 6000 NMR spectra of other nuclei (^{19}F, ^{15}N, ^{17}O, ^{31}P). Several other types of spectra are planned to be included in SpecInfo for future releases. The online database offers many different access capabilities to the factual data, and some illustrative examples are discussed in this paper. In addition, there are four special programs available in the SpecInfo system in dealing with ^{13}C NMR data: CHESS for a search of identical or similar chemical structures, COUPCAL for the calculation of coupling constants, GETSPEC for a similarity research of ^{13}C NMR spectra, and SPECAL for the estimation of a ^{13}C NMR spectrum from a query structure. In addition, there is a spectrum editor (EDSPEC) available to support the building of spectral queries. At the online host STN International, the SpecInfo system is embedded in a cluster of chemical databases, and it is possible to navigate through this system using the various substance registry numbers connecting the different databases. Finally, some future development plants for SpecInfo are presented.

INTRODUCTION

Molecular spectroscopy has become a very important tool for the analysis of chemical substances. The interpretation of spectra of unknown compounds is mostly done by human experts, and it requires excellent knowledge of the given spectroscopic method, i.e., the correspondence between spectral features and the associated structural fragments. It is possible, however, to perform this work to a large extent with the use of computer programs. This is especially important since the number of new or unknown compounds is very large and rapidly increasing. Since the late 1970s, various approaches to automate the interpretation of spectra have been described in the literature (see, e.g., refs 1-3). In most cases these techniques have been implemented at universities and did not become commercially available. An example for such a system is DENDRAL (*Dendr*itic *Al*gorithm), an expert system for the analysis of mass spectra. Although this system had a considerable influence on the development of expert systems in general and of spectroscopic information systems in particular, it did not appear as a commercial product in the market.

A number of spectral data collections are available worldwide.[4-6] The most important spectral databases include ^{13}C NMR, infrared, and mass spectroscopy. An overview of the available data collections is given in ref 4; the various retrieval systems have been discussed in ref 7. Among these sources, SpecInfo contains by far the largest number of spectra of different types. The spectroscopic information system SpecInfo was originally developed inside the BASF Co.[3] for in-house use. The Chemical Concepts Co. is now the owner of the system and is responsible for the distribution. SpecInfo is now available both as an in-house system through Chemical Concepts and for online use on STN International. In this paper the focus lies on the architecture and design of the online system.

CONTENT, ARCHITECTURE, AND DESIGN OF SPECINFO

The spectroscopic information system SpecInfo is both a collection of spectral data and chemical structures and a powerful tool to support the human expert in his tasks to identify new or unknown compounds. The database contains different types of spectra, e.g., NMR, IR, and mass spectra, describing complementary information of chemical compounds. A major principle of SpecInfo is the idea of **multidimensional spectroscopy**.[3,8,9] This concept implies the combined application of spectroscopic methods for the computer-assisted elucidation of chemical structures. Each spectroscopic method contributes partially to the solution of a structure elucidation problem. From infrared spectroscopy one can deduce the characteristic vibrations of certain functional groups; from mass spectroscopy the kind and position of molecular fragments can be concluded; and from the combination of both methods it may be possible to infer the structure of the underlying molecule. As a result of this analysis, SpecInfo provides a set of proposed chemical structures, and it is up to the expert to decide which structure most probably occurs in his experiment. If there is any uncertainty about the interpretation of the spectra, the expert may display or calculate additional spectra of the compounds in question and compare these spectra to the experimental spectra. After two or three steps of interpretation, inference of information, and performance of experiments, the analyst should be able to identify the compound with a very high degree of reliability. The strength of SpecInfo is both the retrieval of spectroscopic information and the interactive computer-assisted identification of chemical substances. A possible procedure for the elucidation of an unknown structure is depicted in the flow chart in Figure 1.

In the online version of SpecInfo (see also ref 10), the database contains approximately 85 000 chemical structures

0095-2338/93/1633-0052$04.00/0

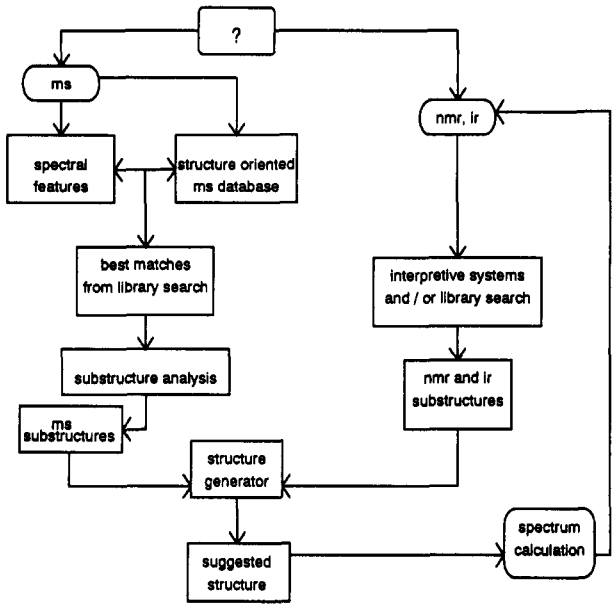


Figure 1. Flow chart containing steps for the elucidation of an unknown chemical structure. This figure has been taken from ref 8 with the kind permission of Dr. Neudert from BASF Ludwigshafen.

and 125 000 spectra of various types (see Table I). The major types of spectra are ^{13}C NMR (68 116 spectra; 11 317 coupling data) and infrared spectra (15 164 spectra). A number of NMR spectra of other nuclei (^{15}N, ^{17}O, ^{19}F, ^{31}P) are also present in the file, but they are of minor importance for the elucidation of chemical structures. Currently, there are 35 503 substances to which a Chemical Abstracts Registry Number (RN) has been assigned. The database is growing rather fast, and new types of spectra will be added very soon. In particular, a large number of mass spectra will be added by the end of 1992. It is also intended to register all the substances in the database in the near future by both Chemical Abstracts Service and the Beilstein Institute.

A number of special programs for the spectral data analysis and interpretation are also available for SpecInfo online. These programs can be invoked and executed online using the so-called RUN command of the Messenger query language. EDSPEC is an editor for spectral queries. This program allows the user to create new or edit existing spectral queries. The input for the EDSPEC program may be

- an existing query generated by EDSPEC
- an uploaded experimental spectrum which is in the JCAMP (*J*oint *C*ommittee on *A*tomic and *M*olecular *P*hysical Data) format (e.g., see refs 11–14)
- a spectrum that has been calculated using the spectrum calculation facility (see below)
- an existing spectrum from the database using the SpecInfo Registry Number or the number of the related spectrum.

The output from EDSPEC can be used to search for similar spectra in the database using the program GETSPEC. Currently, this function is only implemented for the search of ^{13}C NMR spectra. GETSPEC searches the database and produces a list of accession numbers and spectrum sequence numbers together with a similarity percentage between the query spectrum and the spectrum found in the database. In addition to the output from EDSPEC, the calculated spectra from the SPECAL program (see below) are also accepted as input.

It is also possible to calculate the spectrum for a given chemical structure using the program SPECAL (currently only for ^{13}C NMR spectra). The input for this program may be

Table I. Content of SpecInfo Database

type of data	no. of entries
chemical structures	84 000
with CAS Registry Numbers	35 503
^{13}C NMR spectra	68 116
^{13}C NMR coupling data	11 317
^{15}N NMR spectra	841
^{15}N NMR coupling data	133
^{17}O NMR spectra	673
^{17}O NMR coupling data	37
^{19}F NMR spectra	1 753
^{19}F NMR coupling data	957
^{31}P NMR spectra	1 941
^{31}P NMR coupling data	628
infrared spectra	15 164

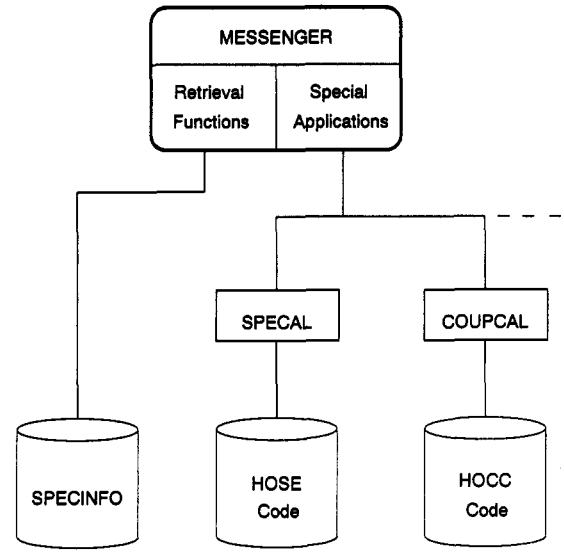


Figure 2. Architecture of the SpecInfo system on STN International.

generated using the Messenger command language (STRUCTURE command) or by uploading a structure from a front-end software, e.g., STN Express. SPECAL calculates a spectrum based on a comparison of structural fragments from the query and the database structures. As a result of this calculation, one obtains a spectrum where the individual lines have been evaluated on the basis of the comparison of the structural fragments. For each line the standard deviation is also given.

Coupling data are available for a subset of the chemical substances (see Table I). It is possible, however, to calculate the coupling constants for a particular chemical substance. The COUPAL program is able to calculate and display coupling constants for any given structure using a structure query as input. In order to perform the calculation of the coupling constants, the so-called HOCC (*h*ierarchically *o*rdered *c*oupling *c*onstants) codes for all substances are stored in a file. For a new structure, COUPCAL calculates the corresponding HOCC codes of possible coupling paths (for a given coupling nucleus) and compares them to the stored values. The values are extracted, and the arithmetic mean value is built. As output one obtains the coupling constants, standard deviation, number of coupling codes, number of structures, and range of coupling constants that have contributed to the mean value.

In general, it is not possible to perform a standard Messenger substructure search in SpecInfo (see below). There is, however, the capability to execute a special fragment based substructure search. The CHESS program is based on HOSE (*h*ierarchically *o*rdered *s*pherical description of *e*nvironment) or HORD (*h*ierarchically *o*rdered *r*ing *d*escription) codes describing the chemical environment of individual atoms and

Table II. Design of Substance Identification Data

field qualifier	field name	type of data	example
SRN	SpecInfo Registry Number	alphanumeric	STCC-25914-159
RN	CAS Registry Number	alphanumeric	69-65-8
RN.CNT	CAS Registry Number counter	numeric	5
CN	chemical name	alphanumeric	mannitol (D)
CNS	chemical name segments	alphanumeric	mannitol
MF	molecular formula	alphanumeric	C6 H14 O6
symbol for elements, e.g., C, B	single element counts	numeric	6, 14, 6, etc.
ATC	atom count	numeric	26
ELC	element count	numeric	3
ELS	element symbols	alphanumeric	C, H, O
ELR.XY[a]	element ratios	numeric	0,428574
PG	periodic group	alphanumeric	A6
MW	molecular weight	numeric	182.17

[a] X and Y = element symbols.

ring systems (ref 15). As a result from such a substructure search, one obtains a set of structures containing the requested fragment(s) and/or ring(s). This function can be used to locate chemical substances which have a special chemical environment and, hence, show the same spectroscopic features as a given substance.

Since SpecInfo consists of a database and a set of programs for special functions, the architecture of the Messenger retrieval system must allow for the integration of these additional functions. A sketch of the SpecInfo architecture is given in Figure 2. All the standard information retrieval queries are handled exclusively by the Messenger system, and the corresponding information is stored in the SpecInfo database. Additional functions, e.g., the program SPECAL or COUPCAL, are processed through the RUN command interface in the Messenger architecture. These programs may access their own data, and these data are stored in separate files. When Messenger receives a query, it has to decide whether it is a standard information retrieval request or a special function which has to be handled by the application software through the RUN command. Of course, it is necessary that the query language is identical in both cases. The results of the RUN command processing must be compatible with Messenger, i.e., answer sets that stem from an application software (e.g., GETSPEC) should be usable in a subsequent Messenger search.

The data in the SpecInfo database consists of three different groups of information: substance identification data, spectroscopic information, and bibliographic references. The latter type of information is rather conventional and will not be discussed in this context. A substance can be identified by a chemical structure, a chemical name, a registry number, or a nonunique secondary key field like the molecular formula or the total number of atoms present in a substance. In Table II the most important substance identification fields are listed together with the field qualifier, a classification of the content as being numeric or alphanumeric, and an example for the content of the field.

The design of spectral data is summarized in a similar way in Table III for ^{13}C NMR spectra where the physical unit is given in an additional column. A general list of spectral data search fields for the various types of NMR spectroscopy is given in Table IV. The term in brackets may be substituted by the symbol of any element that may be the source of NMR data. Infrared data are a little different than NMR data and, therefore, have different types of fields (see Table V). The structure of the design, however, is the same for the different types of spectral data. All search field qualifiers consist of a main qualifier and a subqualifier. The main qualifier is associated with the type of spectral information, and the subqualifier describes the particular search field.

Table III. Design of Spectral Data for ^{13}C NMR

field qualifier	field name	type of data	unit	example
CNMRS.PP	peak position	numeric	ppm	28.05
CNMRS.MUL	multiplicity	alphanumeric		Q (=quartet)
CNMRS.IN	intensity	numeric	1	1
CNMRS.UCG	uncertainty group	alphanumeric		
CNMRS.RT	relaxation time	numeric	s	1.20
CNMR.CC	coupling constant	numeric	Hz	12.10
CNMR.NUC	coupling nucleus	alphanumeric		P

Table IV. Spectral Data Search Fields for NMR Spectroscopy

field name	field qualifier
atom number	<X>NMRS.ATN
line	<X>NMRS.LIN
multiplicity	<X>NMRS.MUL
peak position	<X>NMRS.PP
relaxation time	<X>NMRS.RT
relative intensity	<X>NMRS.IN
uncertainty group	<X>NMRS.UCG
conversion shift	<X>NMR.CS
coupling constant	<X>NMR.CC
coupling nucleus	<X>NMR.NUC
instrument	<X>NMR.INS
note	<X>NMR.NTE
reference	<X>NMR.REF
solvent	<X>NMR.SOL
spectrum number	<X>NMR.NR
standard	<X>NMR.STD
temperature	<X>NMR.T
VCH number	<X>NMR.VCH

Table V. Spectral Data Search Fields for Infrared Spectroscopy

field name	field qualifier	field name	field qualifier
band width	IRS.BDW	quality index	IR.QI
intensity	IRS.IN	reference	IR.REF
peak position	IRS.PP	resolution	IR.RES
instrument	IR.INS	sample form	IR.SPF
intensity (absolute)	IR.IN	spectrum number	IR.NR
note	IR.NTE	VCH number	IR.VCH

In order to obtain a display of information in SpecInfo, one has the choice of various options. Besides the custom formats that refer to individual fields, there are also a number of predefined formats containing groups of data that belong together. A list of these predefined formats is given in Table VI. Users with a graphical terminal emulation software may request the spectra in graphical form using the display formats <X>NMRS.GSP or IRS.GSP. An example of a SpecInfo document is presented in Figure 3 containing the substance

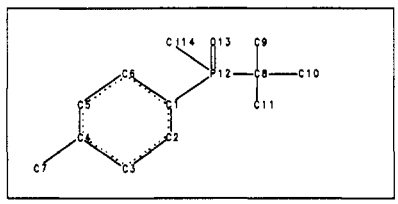
Table VI. Display Formats for Database

display format	format name
<X>NMR	<X>NMR spectral data
<X>NMRS	<X>NMRS spectrum
<X>NMRS.GSP	<X>NMR graphical spectrum
COUP<X>	<X>NMR coupling data
all<X>	all <X>NMR spectral data available
IR	IR spectral data
IRS	IR spectrum
IRS.GSP	IR graphical spectrum
ALLIR	all infrared spectral data available
HIT	format containing all hit terms
TRIAL	SRN, MF, MW, RN, FA
OCC	table containing display fields and the corresponding number of occurrences of hit terms

=> D L7 1 IDE IRS.GSP

L7 ANSWER 1 OF 16

SPECINFO Reg. No. (SRN)	: STCC-97458-121
CAS Registry No. (RN)	: 41308-89-8
Molecular Formula (MF)	: C11 H16 Cl1 O1 P1
Molecular Weight (MW)	: 230.66
Chemical Name (CN)	: T-BUTYL-(4-TOLYL)-PHOSPHINYL CHLORIDE CHLORO-T-BUTYL-4-TOLYL-PHOSPHINE OXIDE



L7 ANSWER 1 OF 16

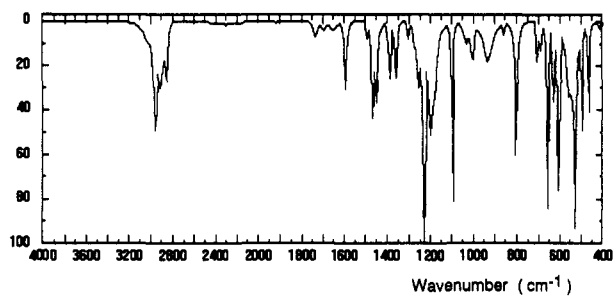


Figure 3. Example document of SpecInfo for the substance identification and spectral data.

identification data including the chemical structure and a spectrum in graphical form.

SEARCH AND RETRIEVAL CAPABILITIES

To access the information in SpecInfo, the standard Messenger commands can be used. A list of the commands is available as an online help message, and it is described in detail in the standard STN documentation. The first step for a novice user searching for spectroscopic information is a search in the NUMERIGUIDE database. This file contains the information about the availability of numeric data in STN databases; it is thus a referential file 'pointing' to the online sources of numeric data. Users can search for the preferred property names or synonyms, or may use the powerful thesaurus capability and perform any type of up- or down-search in the hierarchy of the classification scheme. An example for a simple search in NUMERIGUIDE is given in Figure 4. As a result, one obtains a table of online numeric data sources together with some additional useful information.

```
=> s infrared spectrum/ph
L1        1 INFRARED SPECTRUM/PH
=> d

L1  ANSWER 1 OF 1

AN      5153 NUMERIGUIDE
PPN     IR spectrum
UF      IR
UF      IR data
UF      IR spectrum (band spectrum line)
UF      infrared identification
UF      ***infrared spectrum***
UF      spectrum, IR
UF      spectrum, infrared
```

FQ	FQ Type S=search D=display	Default Search Unit(*)	File
IRS	S,D	cm**-1	BEILSTEIN
IRS	D		GMELIN, HODOC, SPECINFO
IRS.BDW	S	cm**-1	SPECINFO
IRS.IN	S	%	SPECINFO
IRS.PP	S	cm**-1	HODOC, SPECINFO
IRS.SOL	S		GMELIN
IRS.T	S	C	GMELIN
IRS.W	S	cm**-1	GMELIN
CT	S		BEILSTEIN

(*) Original file units. For more information use the HELP UNITS message in the specific file.

Figure 4. Example for a search of spectral data in NUMERIGUIDE.

In SpecInfo one may search for substance information using the search fields and features that are well-known from databases like REGISTRY or BEILSTEIN. The only exception is the Messenger substructure search capability, which is not implemented in SpecInfo. A more exciting feature is the numeric search for spectroscopic data, i.e., peak positions, intensities, and multiplicities. In Figure 5 an example for such a search is shown together with an answer document. The numeric values in the search query are given as value ± tolerance instead of a numeric range. Both possibilities for the input of a numeric query are accepted by the Messenger command language.

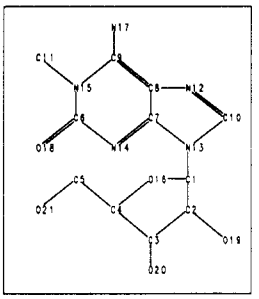
A spectrum that has been obtained from a previous database search can be used as a basis for a subsequent similarity search. Let us assume that we have found a set of spectra from a library search, and we are interested in all substances that have a similar spectrum to one of the spectra in the answer set. In this case it is necessary to produce a query for the similarity search program (GETSPEC). This is done by invoking the spectrum editor (EDSPEC) and using the spectrum number as input. After completion of the EDSPEC program, a query is generated and stored in an L-number. This type of query generation is similar to the structure query modeling that is available for the REGISTRY or BEILSTEIN database. In the next step, the query is used as input in the similarity search program GETSPEC (see Figure 6).

It has been stated already that it is not possible to perform a Messenger substructure search in SpecInfo. There are, however, several possibilities for structure searches in Messenger that can be used in connection with SpecInfo. A structure query can be built using the Messenger STRUCTURE command or front-end software like STN Express. The structure query must be created in the SpecInfo bonding conventions (see Table VII) since the programs SPECAL and CHESS require the specification of bonds that have no equivalent in Messenger, e.g., aromatic bonds. In addition to this type of structure-based processing, it is also possible to perform a substructure search in another structure file, e.g., REGISTRY or BEILSTEIN, and crossover to SpecInfo using the registry numbers. In this case, the structures must be built in the standard Messenger conventions. The diagram in Figure

```
=> s (30+-0.5 and 109+-1 and 152+-1 and 154+-1 and 87.5+-1)/cnmrs.pp
      8658      30+-0.5/CNMRS.PP
      4616      109+-1/CNMRS.PP
      5406      152+-1/CNMRS.PP
      5033      154+-1/CNMRS.PP
      1986      87.5+-1/CNMRS.PP
L1       1 (30+-0.5 AND 109+-1 AND 152+-1 AND 154+-1 AND 87.5+-1)
                                                        /CNMRS.PP
=> d ide cnmrs cnmr

L1   ANSWER 1 OF 1

SPECINFO Reg. No. (SRN)      :STCC-99941-105
Molecular Formula (MF)       :C11 H15 N5 O5
Molecular Weight (MW)        :297.28
Chemical Name (CN)           :1-METHYL-ISOGUANOSINE
```



```
13C-NMR Spectrum :
```

Atom No. .ATN	Peak Pos. .PP (ppm)	Multipl. .MUL	Intensity .IN	Relax. Time .RT (sec)
1	87.60	d	1	0.2
2	72.90	d	1	0.2
3	70.70	d	1	0.2
4	86.00	d	1	0.2
5	61.80	t	1	0.1
6	153.80	s	1	7.7
7	152.10	s	1	7.0
8	108.90	s	1	8.1
9	151.50	s	1	2.5
10	138.00	d	1	0.2
11	30.00	q	1	0.5

```
Spectrum Number (.NR)        :CNCC-82266-457
VCH Number (.VCH)            :458410
Instrument (.INS)            :Jeol FX-60
Solvent (.SOL)               :DMSO-D6
Standard (.STD)              :TMS
Temperature (.T)             :305 K
Literat. Ref. (.REF)         :R.S.NORTON, R.P.GREGSON, R.J.QUINN,
                              CHEM.COMMUN.339(1980)
```

Figure 5. Example for the search of spectroscopic information in SpecInfo.

7 shows the possibilities for structure building and processing in SpecInfo.

An interesting capability of SpecInfo is the possibility to calculate a spectrum that is not in the database. The program SPECAL allows the calculation of a complete or partial ^{13}C NMR spectrum using a structure query (in SpecInfo conventions) as input. In Figure 8 an example for such a spectrum calculation is shown together with the result of this calculation. For each atom the following information is given: mean value, standard deviation, multiplicity, and number of lines. SPECAL generates structure fragments (HOSE codes) and compares them to fragments stored in the database. For each fragment found in the file there is a number of spectral lines which correspond to this particular fragment in different chemical environments, and from these line values the mean value and the standard deviation are calculated. If the standard deviation is too large, it is possible to recalculate the mean value for this atom by restricting the calculation to smaller intervals of the line values in the file.

INTEGRATED CHEMICAL INFORMATION SYSTEM

SpecInfo is an integrated information system consisting of a database and a set of special application software packages. As an online system it is also embedded in the other databases of STN International. The overall architecture and concept of STN is based on the idea of an **integrated chemical information system** which consists of the following components:

```
=> run edspec c

 *** SPECINFO SPECTRUM EDITOR ***

: 26
: 33
: 59
: 65
: 70
: 81
: 165
: 172

:save

L1   RUN STATEMENT CREATED

=> run getspec l1 c

 *** SPECINFO SPECTRUM SIMILARITY SEARCH ***

PRE SEARCH LIMIT :                73.1
NO OF REFERENCE BITMASKS :        1494
NO OF REFERENCE SPECTRA :         14425
```

User Input of Spectral Lines

ANSWER	ACC NO	SPECTRUM	SIMILARITY (%)
1	44852293	1	96
2	57120485	1	57
3	78274353	2	48
..	..	..	..
..	..	..	..
19	31030097	1	22

```
L2   RUN STATEMENT CREATED

SORTED L#S ARE NOT ALLOWED FOR SEARCH

=> d cnmrs

L2   ANSWER 1 OF 96

13C-NMR Spectrum :
```

Atom No. .ATN	Peak Pos. .PP (ppm)	Multipl. .MUL	Intensity .IN
1	64.70	s	1
2	70.00	d	1
3	80.60	d	1
4	58.90	s	1
5	165.25	s	1
6	25.90	q	1
7	33.00	q	1
8	172.40	s	1

```
=> d srn cn cnmr

L2   ANSWER 1 OF 96

SPECINFO Reg. No. (SRN)      : STCC-44852-293
Chemical Name (CN)           : 6,6-DIBROMO-PENICILLIN

13C-NMR Data :

Spectrum Number (.NR)        : CNCC-83029-625
VCH Number (.VCH)            : 463790
Instrument (.INS)            : Var. CFT-20
Solvent (.SOL)               : Acetone-D6
Standard (.STD)              : TMS
Temperature (.T)             : ambient
Literat. Ref. (.REF)         : C.R.HARRISON,P.HODGE,
                                J.CHEM.SOC.PERKIN
                                I,1772(1976)
```

Figure 6. Example for the use of the similarity search program GETSPEC.

(1) multiple file environment
It is possible to search several databases together in a single search and eliminate the duplicate answers from the answer set.

(2) link between the individual databases
Information can be crossed between the different databases (cross-file searching). The registry numbers play a central role in this system since they provide the possibility to start a search in one file and continue the search in another file. Based on this key, it is even possible in some cases to access a second file while the user is still working in the first file. An example for this feature is the display of CA documents from the REGISTRY

Table VII. Bond Values Allowed in SpecInfo

value	definition	graphic symbol	text symbol	example
SE	single exact (default)	---	---	BON 1 2 SE
DE	double exact	═	==	BON 2 3 DE
T	triple	≡	#	BON 7 8 T
AR	aromatic	···	***	BON R 1 2 AR
CO	coordination bond between a central atom and another ligand	···	>>>	BON 3 4 CO
CP	coordination bond between a central atom and an electron cloud (π-system) defined by a group of atoms	not shown	not shown	BON 1 (2 6) CP

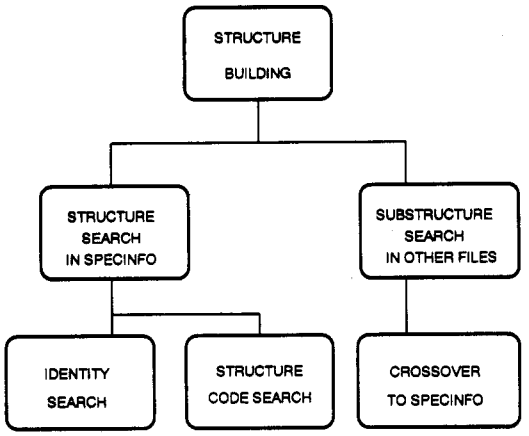
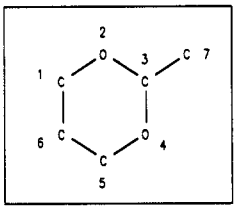


Figure 7. Procedure for structure building and searching in SpecInfo.

```
=> str
: gra r6, gra 3 c1, nod 2 o, nod 4 o, dis
```



```
: end
L7              STRUCTURE CREATED -- BUILT BY 'SPECINFO'
                CONVENTIONS
=> run sp all

        *** CALCULATION OF NMR SPECTRUM LINE VALUES ***

ATOM      MEAN      STD. DEV      MULTIPL.      HITS
NO.       (PPM)     (PPM)

1         67.17     0.88          T             6 LINES
3         100.57    1.05          D             3 LINES
5         67.17     0.88          T             6 LINES
6         26.52     2.38          T             15 LINES
7         21.11     0.34          Q             16 LINES

L8  RUN STATEMENT CREATED

ENTER NODE NUMBER (END) :7

ATOM NUMBER                     : 7
MULTIPLICITY                    : Q
HOSE CODE                       : C(OO/C,C/C,&)
NO. OF LINE VALUES              : 16
NO. OF STRUCTURES               : 10
MEAN, STANDARD DEVIATION        : 21.1  +/-  0.3 PPM
RANGE                           : 20.4  -  22.0 PPM
ENTER SD,DR,CL,SE (END) :sd

          INTERVAL    HITS
            (HZ)
          ----------  -------
          20 - 21       6
          21 - 22      10
```

Figure 8. Example for a calculation of a ^{13}C NMR spectrum.

file. STN will provide a similar feature for SpecInfo and BEILSTEIN in such a way that users will be able to display graphical spectra from SpecInfo while they are still working in the BEILSTEIN file. The connection between some

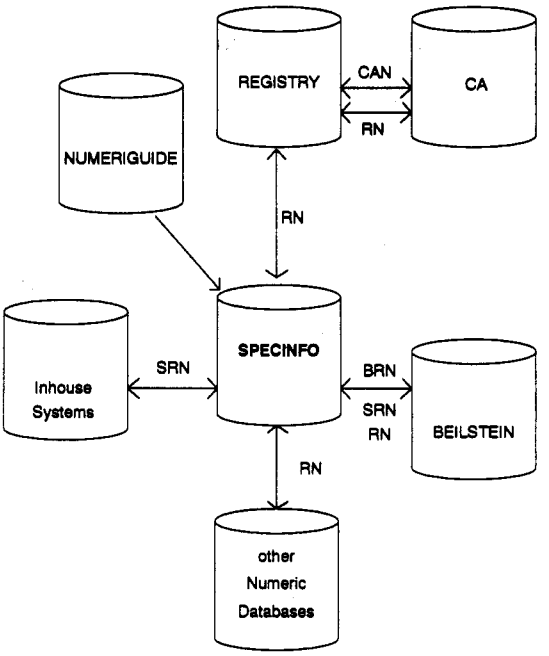


Figure 9. Links between SpecInfo and other databases on STN International.

databases on STN is outlined in Figure 9 where the focal point is lying on the SpecInfo file.

(3) integration of software components

An important concept of the Messenger architecture is the possibility to integrate application software packages that perform special tasks which are outside the scope of normal information retrieval systems. By using the interface of the RUN command, it is possible to implement a new software package in a rather short time. A number of such programs has been implemented already, e.g., the SpecInfo programs that have been mentioned above.

(4) support by decentralized user software

The online service of STN is augmented by a set of front-end and back-end programs that can be installed at the user's personal computer or workstation. STN Express is a front-end communication software with a comfortable structure editor dedicated to the special needs of chemists. Another program, STN PFS, is a personal file management system that has been designed to store, manage, and retrieve bibliographic documents which have been downloaded from an online host. Both programs are distributed by the STN service centers. Other programs like ALCHEMY III from Tripos Associates can be accessed easily from both STN Express or STN PFS. Additional interfaces will be provided in the near future for the software from Molecular Design Limited, e.g., ISIS or REACCS.

SUMMARY AND FUTURE PLANS

The information basis of SpecInfo is updated frequently, and new types of spectra are added if available. It is intended to add the following types of spectra until the beginning of 1993:

15 000–20 000 ^{13}C NMR spectra
15 000 IR spectra
40 000 MS spectra
other types of <X> NMR spectra.

The implementation of a capability for similarity and identity searches of NMR, MS, and IR spectra is also planned for the near future.

In summary, it can be stated that SpecInfo is a rather complex information system which offers a lot of different possibilities for the retrieval of spectroscopic information. In addition, there are special tools available which support the experts in their analytical work to evaluate the structures of unknown chemical substances. The various functions for structure elucidation, spectrum calculation, and similarity searches provide capabilities that are quite beyond the 'standard' retrieval possibilities. In order to make full use of the SpecInfo system, the user should be an expert in spectroscopy.

ACKNOWLEDGMENT

Bernard Matz has prepared some of the tables and figures and most of the examples in this paper. His contributions are gratefully acknowledged. The funding of this work by the Federal German Ministry for Research and Technology is also gratefully acknowledged.

REFERENCES AND NOTES

(1) Heller, S. R. The chemical information system and spectral databases. *J. Chem. Inf. Comput. Sci.* **1985**, *25* (3), 224–231.
(2) Damen, H.; Henneberg, B.; Weimann, B. SISCOM—a new library search system for mass spectra. *Anal. Chim. Acta* **1978**, *103* (4), 289–302.
(3) Bremser, W.; Fachinger, W. Multidimensional Spectroscopy. *Magn. Reson. Chem.* **1985**, *23* (12), 1056–1071.
(4) Warr, W. A.; Suhr, C. *Chemical Information Management*; VCH Verlagsgesellschaft: Weinheim and New York, 1992; pp 261.
(5) Warr, W. A. Spectral databases. *Chemom. Intell. Lab. Syst.* **1991**, *10* (3), 279–292.
(6) Wiggins, G. *Chemical Information Sources*; McGraw-Hill: New York, 1991; pp 352.
(7) Heller, S. R. Numerical data acquisition, dissemination and retrieval systems. *Anal. Chim. Acta* **1988**, *210* (1), 81–87.
(8) Neudert, R.; Bremser, W. Multidimensional spectroscopy. In *Software-Enwicklung in der Chemie 1. Proceedings des Workshops "Computer in der Chemie"*, Hochfilzen/Tirol, 1986; Springer-Verlag: Berlin, 1987; pp 127–135.
(9) Bremser, W.; Grzonka, M. SpecInfo—a multidimensional spectroscopic interpretation system. *Mikrochim. Acta* **1991**, *2*, 483–491.
(10) Hellenbrandt, M.; Canzler, D. SpecInfo—The Spectroscopic Information System on STN International. *Fresenius Z. Anal. Chem.* **1992**, *344*, 167–172.
(11) IUPAC Commission on Molecular Structure and Spectroscopy, Physical Chemistry Division. JCAMP-DX, a standard format for exchange of infrared spectra in computer readable form. *Pure Appl. Chem.* **1991**, *63* (12), 1781–92.
(12) Davies, A. N.; Hillig, H.; Linscheid, M. JCAMP-DX, a standard? In *Software Development of Chemistry 4, Proceedings of Workshop "Computer in Chemistry"*; Gasteiger, J., Ed.; Springer: Berlin, 1990; pp 147–156.
(13) Gasteiger, J.; Hendriks, B. M. P.; Hoever, P.; Jochum, C.; Somberg, H. JCAMP-CS: a standard exchange format for chemical structure information in computer-readable form. *Appl. Spectrosc.* **1991**, *45* (1), 4–11.
(14) McDonald, R. S.; Wilks, P. A., Jr. JCAMP-DX: a standard form for exchange of infrared spectra in computer readable form. *Appl. Spectrosc.* **1988**, *42* (1), 151–162.
(15) Bremser, W. HOSE—a novel substructure code. *Anal. Chem. Acta* **1978**, *103* (4), 355–365.