Fuzzy Definition of Molecular Fragments in Chemical Structures

Barbara Dębska* and Barbara Guzowska-Świder

Department of Computer Chemistry, Rzeszów University of Technology, 35-041 Rzeszów, Poland

This paper presents a methodology for seeking the relationships between chemical substructures (molecular fragments) and spectral parameters using a computer collection data of molecular spectra. To establish the spectrum−structure correlations, the program has to search the chemical structure base in order to find compounds containing a given molecular fragment in the molecule. There exists no sole definition of a substructure, as it always depends on the type of problem dealt with. In the problem of structural identification, fuzzy definitions of substructures are applied, and their forms are imposed by the spectral methods used.

1. INTRODUCTION

The origin of chemical information is the chemical compound, and the notion of their structure is fundamental in most application fields:[1−7] structure−activity relationships, reaction indexing, reaction synthesis, molecular modeling, substructure search, structure elucidation, etc. A number of computers programs were described to support the structure elucidation process.[8−29] A possible approach to spectra interpretation can be based on determination of functional groups, which may be present in an unknown compound. Several computer programs have been written in order to mimic the human process of interpreting molecular spectra of an unknown compound. To recognize the structure of a given compound, these programs[8−22] use correlation tables, i.e., lists of structural fragments (i.e., functional groups, molecular fragments) together with their spectral characteristics. Usually, these characteristics are taken from literature,[8−17] but they can also be generated from a computer library of molecular spectra.[18−22]

A number of programs have been developed for recognition of the presence or absence of substructures in chemical molecules without extracting the information about spectrum−structure correlations, e.g., as interpretation rules. These programs mainly apply neural network methods.[23−29]

The first step in developing programs for the structure elucidation process is to elaborate a list of substructures that can be recognized by a computer and to define a numerical representation of chemical structures and molecular fragments. In our approach, complete chemical structures and substructures are coded in the form of molecular graphs.

2. DEFINITION OF MOLECULAR GRAPH

The structure of each molecule can be defined as a molecular graph, whose nodes describe the atoms and whose edges represent chemical bonds linking the atoms. The definition of graphs applied in our research was presented by Fic.[30] The nodes were defined by three attributes describing chemical features of the subsequent atoms, such as A, the type of an atom; V, its value; and E, the number of free electrons. The edges of molecular graphs were coded by the attribute B (bond), whose value was 1, for a single chemical bond; 2, for a double bond; and 3, for a triple bond. Figures 1 and 2 depict molecular graphs of a linear molecule (pentan-1-ol) and of an aromatic one (propiophenone).

Every chemical structure can be considered as a combination of smaller fragments, called substructures. The program for structure elucidation requires searching through a structure database to select compounds that in their molecule contain the substructures sought, and for this reason the substructures have to be defined as molecular subgraphs.

3. FUZZY DEFINITION OF MOLECULAR SUBSTRUCTURE

The number of molecular substructures is practically indefinite; therefore they are often divided into characteristic classes to simplify the work on them. In the course of our research, the fuzzy substructure conception was introduced. A fuzzy substructure stands for a characteristic class of substructures and is defined in the form of a fuzzy molecular subgraph.[31] The characteristic features of subgraphs are multivalent attributes describing the nodes and the edges of a graph, which means that each fuzzy subgraph represents more than one substructure. Each substructure is defined by the following:

The substructure vector describing the global features of a substructure, such as the number of atoms (graph nodes), the number of bonds, the number of single, double, triple, or aromatic bonds, the number of atoms covered by ring systems, the minimum chemical formula of the substructure (containing information about types and quantity of atoms C, N, O, ..., excluding H atoms).

The set of configuration vectors coding for each atom the type of an atom, the number of free electrons, the location of the atom in the molecule (1 = not in the ring, 2 = in the ring, and 0 = any), the number of neighbors, the number of single, double, triple, and aromatic bonds a given atom is a part of.

The list of bonds representing the way the atoms are connected in and the type of bonds.

The substructure vector is used as a screening filter for fast and efficient checking whether the subsequent molecules

---

* Corresponding author. E-mail: bjdebska@prz.rzeszow.pl.

pentan-1-ol

$$\overset{1}{CH_3}-\overset{2}{CH_2}-\overset{8}{CH_2}-\overset{11}{CH_2}-\overset{12}{CH_2}-\overset{17}{OH}$$
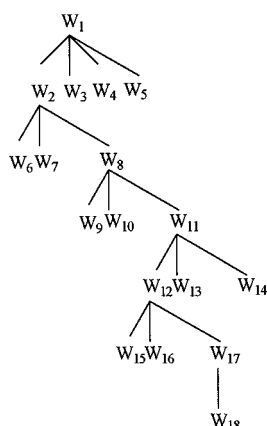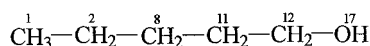
Attributes of nodes

W	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	C	C	H	H	H	H	H	C	H	H	C	C	H	H	H	H	O	H
V	4	4	1	1	1	1	1	4	1	1	4	4	1	1	1	1	2	1
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0

List of bonds

W_1	1	1	1	1	2	2	2	8	8	8	11	11	11	12	12	12	17
W_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 1. Molecular graph of pentan-1-ol.

propiophenone
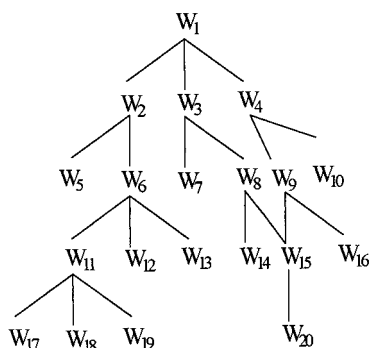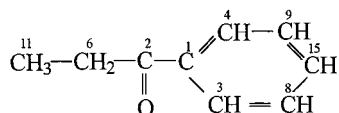
Attributes of nodes

W	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	C	C	C	C	C	O	C	H	H	C	H	C	H	H	H	C	H	H	H	H
V	4	4	4	4	2	4	1	1	4	1	4	1	1	1	4	1	1	1	1	1
E	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0

List of bonds

W_1	1	1	1	2	2	3	3	4	4	6	6	6	8	8	9	9	11	11	11	15
W_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
B	1	1	2	2	1	1	2	1	2	1	1	1	1	2	1	1	1	1	1	

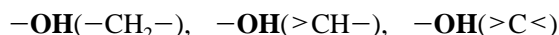**Figure 2.** Molecular graph of propiophenone.

from the structure database can contain the analyzed substructure. Then, the program checks the set of configuration vectors and the list of bonds to see whether the molecule really contains the examined substructure.

An example of a fuzzy substructure definition is shown in Figure 3. A set of substructures, which can be present in the neighborhood of −OH substructure, is coded as $. Figure 3a presents a fuzzy substructure, Figure 3b a fuzzy subgraph, and Figure 3c−e contains various examples of real sub-
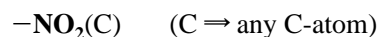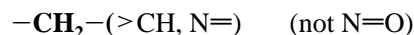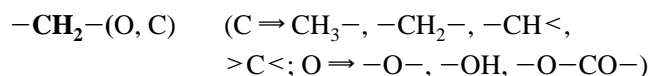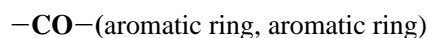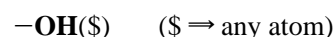
graphs. The definitions of subgraphs recognized by the identification system are stored in the computer database. The subgraph database of our system contains data on 222 substructures (e.g., 62 recognized only by IR spectroscopy, 132 by ^{1}H NMR, and 28 identified by both spectral methods).

The substructure database contains precise definitions of groups containing the C, N, O, and S elements and the halogens. The basic set of substructures is connected with various types of organic compounds, such as ketones −CO− ($,$), aldehydes −CHO($), carboxylic acids −COOH($), amines −NH−($,$), alcohols −OH($), aromatic benzene ring ($), heterocyclic rings, pyridine ring ($), etc. Each subset (ketones, aldehydes, etc.) includes substructures together with their precisely or fuzzily defined surroundings. The subset of a given substructure can contain

precisely defined molecular fragments

$$-\mathbf{OH}(-CH_2-),\quad -\mathbf{OH}(>CH-),\quad -\mathbf{OH}(>C<)$$

fuzzily defined molecular fragments

$$-\mathbf{OH}(\$)\qquad (\$\Rightarrow \text{any atom})$$

$$-\mathbf{OH}(\text{aromatic ring})$$

$$-\mathbf{CO}-(\text{aromatic ring, aromatic ring})$$

$$-\mathbf{CH_2}-(CH3\,,X)\qquad (X = F, Cl, I, Br)$$

$$-\mathbf{CH_2}-(O, C)\qquad (C \Rightarrow CH_3-, -CH_2-, -CH<,$$
$$>C<; O \Rightarrow -O-, -OH, -O-CO-)$$

$$-\mathbf{CH_2}-(>CH, N=)\qquad (\text{not } N=O)$$

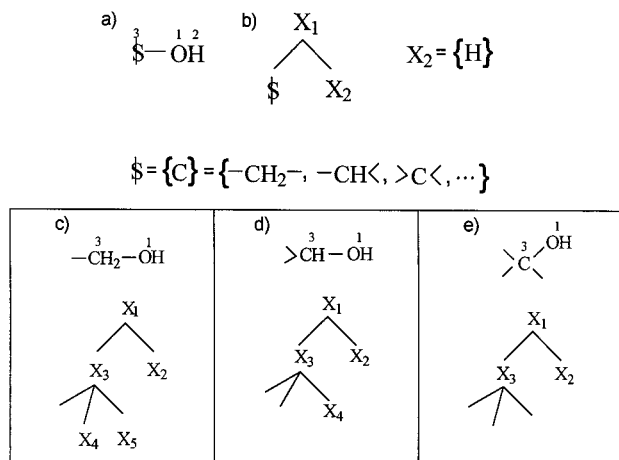$$-\mathbf{NO_2}(C)\qquad (C \Rightarrow \text{any C-atom})$$

The parentheses ( ) contain the alternative surroundings of the basic substructures (in bold).

The list containing some examples of substructure recognized by IR spectroscopy is presented in ref 22. The above-mentioned substructures are used in the SCANKEE system[32] to realize two tasks: (1) the generation of spectrum−structure correlations for selected substructures from a molecular database and automatic transformation of the correlations into a rule knowledge base, and (2) the identification of the substructures in the molecule of an unknown analyzed compound using the rule knowledge base and molecular spectra of the substance.

The rule knowledge base with the inference engine of the SCANKEE system constitutes the expert system for the recognition of substructures possibly present in the molecule of an unknown compound from its spectrum (spectra).

System SCANKEE (see Figure 4) contains various modules, but for this problem only three are used: the MDS module (multispectroscopy database system) for storage of spectral data, the MKF module (multistrategy knowledge formalizer) for rule knowledge base generation, and the SEK module (simulator of expert knowledge), an inference engine. The source spectral databases, discussed here, were developed using the MDS module and spectra registered in our university for organic compounds (e.g., 57% of compounds stored in the IR base contain an aromatic ring, 45% contain a methylene group, and 39% contain a methyl fragment).

Fuzzy Definition of Molecular Fragments

J. Chem. Inf. Comput. Sci., Vol. 40, No. 2, 2000 **327**

Figure 3. Definition and coding of a $-OH(\$)$ fuzzy substructure: (a) atoms' numeration; (b) fuzzy subgraph; (c) subgraph of $-OH(CH_2)$ substructure; (d) subgraph of $-OH(CH)$ substructure; (e) subgraph of $-OH(>C<)$ substructure.

Substructure vector: 3, 2, 2, 0, 0, 0, 0, 1C, 1O

Set configuration vectors

Atom No.	Type	Number of free electrons	Location of atom	Number of neighbors	Bonds		
					double	triple	aromatic
1	O	4	1	2	0	0	0
2	H	0	1	1	0	0	0
3	C	0	999	4	0	0	0

where 999 – any value

List of bonds

X_i	1	1
X_j	2	3
bond type	1	1

The MKF module can realize a process of searching for regularities in the spectral database selected by the user (see Table 1) to extract the spectrum−structure correlations for chosen substructures. The correlations are next converted into interpretation rules and together with the rules for controlling the inference engine form the rule knowledge base, which can be applied by the SEK module. The user introduces into the computer the spectral parameters of an unknown spectrum, and the inference engine accessing the rule knowledge base generates the list of substructures forming (with a probability index) a part of the molecule of the analyzed compound. In this way, for each selected spectral database the rule knowledge base can be created, and as a result of an application of the developed rule knowledge bases for interpretation of unknown spectra, the user obtains for each spectroscopy a separate list of recognized substructures. Soon the lists will be used by a GEN module for generation of structure candidates. The rule knowledge base generation was already presented in detail in our paper "Automatic Generation of Knowledge Base from Infrared Spectral Database for Substructure Recognition" preceding this paper in this issue.

4. SUBSTRUCTURE RECOGNITION

The SCANKEE system enables generation of a rule knowledge base for six spectral methods (^{13}C NMR, ^{1}H NMR, IR, MS, Raman, and UV) and consequently also recognition of substructures from these spectral data registered for an unknown substance whose chemical constitution has been found. There is a common list of substructures for all the spectral techniques presented above, but the correlations are generated only for the substructures which are active
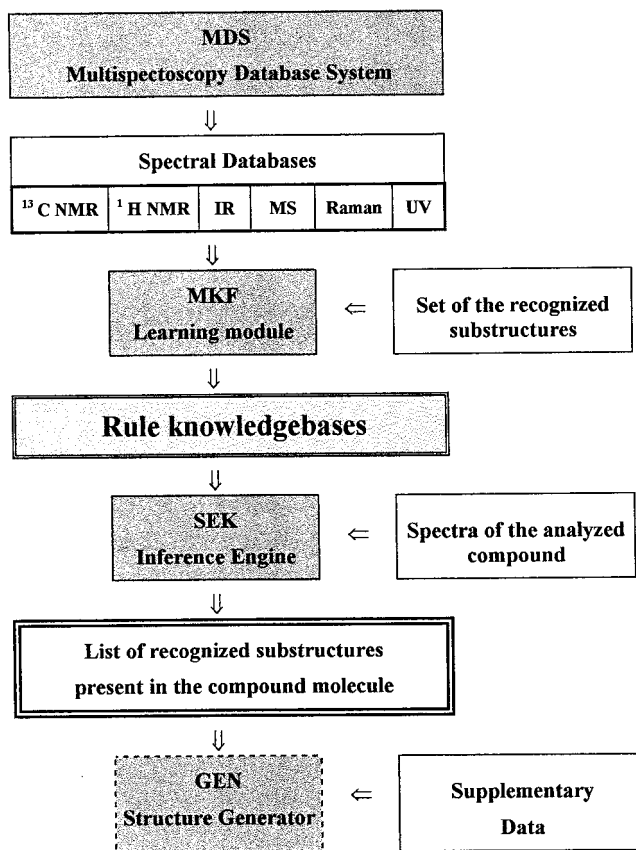

Figure 4. Block diagram of the spectral data processing in SCANKEE system.

in a given spectroscopy and which give the most characteristic correlations. The fuzzy molecular substructures can be recognized by several spectral techniques, whereas the precise determination of the environment of molecular fragments is possible usually only by one or two spectral methods. For example, the substructure $-CO-(\$,\$)$ is best recognized by IR spectroscopy, with minor reliability by Raman, MS, or UV spectroscopy, and it is unseen by the ^{1}H NMR method. On the other hand, the $-OH(\$)$ functional group is very well identified by IR spectroscopy, is quite well identified by Raman, MS, and ^{1}H NMR spectroscopies, and is completely inactive in ^{13}C NMR.

For each spectral method selected by the chemist, various rule knowledge bases can be created with respect to the set of substructures taken into consideration. Therefore, the chemist can generate a number of knowledge bases for the types of compounds that are of interest. The algorithm for generating the spectrum−structure correlations, the examples of created correlations, and the methodology of their transformation into a rule knowledge base are presented in one of our recent papers.[22]

5. RESULTS AND DISCUSSION

The usefulness of the fuzzy and precise substructure definition is illustrated by the process of structure recognition of 2-(dimethylamino)ethyl acetate with the use of IR and ^{1}H NMR rule knowledge bases and IR and ^{1}H NMR spectra. Tables 2 and 3 depict two spectra of the compound in the form of spectral parameters being the input data for the SCANKEE system. The results of substructure identification are shown in Tables 4−8. The recognition process of a CH_3-

Table 1. Spectral databases of SCANKEE System

no.	spectral databases	no. of spectra	spectral parameters	
1	^{13}C NMR	500	chemical shift (δ ppm)	intensity
2	^{1}H NMR	700	chemical shift (δ ppm)	intensity
			multiplicity	coupling constant (Hz)
3	IR	900	band location (cm^{-1})	intensity (% transmittance) (absorbance)
4	Raman	200	band location (cm^{-1})	intensity
5	UV	300	band location (cm^{-1})	intensity (extinction)
6	MS	3000	peak positio (m/e)	intensity

Table 2. IR Spectrum of 2-(Dimethylamino)ethyl Acetate (Spectral Parameters of Absorption Bands)

no.	band location (cm^{-1})	intensity (A)
1	2967.0	0.481
2	2817.0	0.422
3	2780.0	0.593
4	1748.0	1.521
5	1460.0	0.462
6	1370.0	0.586
7	1235.0	1.723
8	1181.0	0.113
9	1093.0	0.212
10	1042.0	1.238

Table 3. ^{1}H NMR Spectrum of 2-(Dimethylamino)ethyl Acetate (Spectral Parameters of Absorption Bands)

no.	chem shift (δ ppm)	intensity (H)	multiplicity	coupling const (Hz)
1	1.95	3	singlet	—
2	2.20	6	singlet	—
3	2.45	2	triplet	7.1
4	4.05	2	triplet	7.1

Table 4. Results of CH$_3$−($) Substructure Identification

spectroscopy	recognized substructure	
IR	**CH$_3$−($)**	$: any atom
^{1}H NMR		
band location	**CH$_3$−(CO−$)** or **CH$_3$−(N(C,C))** or −**CH$_2$−(C,CO-$)**	$: C, O−C; C: CH$_3$−, −CH$_2$−, −CH<, >C<
band location, intensity, and multiplicity	1 substructure **CH$_3$−(CO−$)** or 1 substructure **CH$_3$−(N(C,C))**	

Table 5. Results of CH$_3$−N−CH$_3$($) Substructure Identification

spectroscopy	recognized substructure	
IR	CH$_3$−N−CH$_3$ ($)	$: any atom
^{1}H NMR		
band location	**CH$_3$−(CO−$)** or **CH$_3$−(N(C,C))** or −**CH$_2$−(C,CO-$)**	$: C, O−C; C: CH$_3$−, −CH$_2$−, −CH<, >C<
band location, intensity, and multiplicity	2 substructure **CH$_3$−(CO−$)** or 2 substructure **CH$_3$−(N(C,C))** or	$: C, O−C; C: CH$_3$−, −CH$_2$−, −CH<, >C<
	3 substructure −**CH$_2$−(C,CO−$)**	$: C, O-C; C: >C<

Table 6. Results of First −CH$_2$−($,$) Substructure Identification

spectroscopy	recognized substructure	
IR	−CH$_2$−($,$)	$: any atom
^{1}H NMR		
band location	−**CH$_2$−(C,CO−$)** or −**CH$_2$−(C,N(C,C))**	$: C, H; C: CH$_3$−, −CH$_2$−, −CH<, >C<
band location, intensity, multiplicity	1 substructure −**CH$_2$−(CH$_2$,CO−$)** or 1 substructure −**CH$_2$−(CH$_2$, N (C, C))**	

Table 7. Results of Second −CH$_2$−($,$) Substructure Identification

spectroscopy	recognized substructure	
IR	-CH$_2$-($,$)	$: any atom
^{1}H NMR		
band location	−**CH$_2$−(C,(C)−O−CO−(C))**	C: CH$_3$−, −CH$_2$−, −CH<, >C<
band location, intensity, and multiplicity	1 substructure −**CH$_2$−(CH$_2$,(C)−O−CO−(C))**	

Table 8. Results of −CO−O−(C,C) Substructure Identification

spectroscopy	recognized substructure	
IR	−CO−O−(C,C)	C: CH$_3$−, −CH$_2$−, −CH<, >C<
^{1}H NMR	—	

($) fuzzy fragment is presented in Table 4. The infrared spectroscopy has identified the presence of CH$_3$−($) in the molecule, whereas the ^{1}H NMR spectrum has determined also the environment of the CH$_3$ group, coded as $. The precise definition of the substructure's neighborhood was led in two steps being a result of taking into consideration

subsequent spectral parameters of the ^{1}H NMR spectrum (chemical shift value, intensity, and multiplicity). The analysis of the CH$_3$−N−CH$_3$($) group (Table 5), the two −CH$_2$−($,$) groups (Tables 6 and 7), and the −CO−O−(C,C) group (Table 8) was performed in a similar way. The CH$_3$−N−CH$_3$($) fuzzy substructure was identified by IR spectroscopy, whereas ^{1}H NMR spectroscopy recognized a methyl group neighboring nitrogen atom connected with alkyl groups, but another surrounding group (CO−$) was also taken into consideration. One methylene group was recognized as fuzzy substructure −CH$_2$−($,$) by IR spectroscopy. The ^{1}H NMR rule knowledge base identified it at first as a methylene group linked with an alkyl substructure (CH$_3$−, −CH$_2$−, −CH−, or >C<) by one bond and by a second connected with a nitrogen atom (surrounding by an alkyl group) or CO−$ group. The multiplicity of the signal enabled determination of (−CH$_2$−) as an alkyl neighbor of the discussed methylene substructure (Table 6).

The second substructure −CH$_2$−($,$) recognized by IR spectroscopy as a fuzzy substructure was unequivocally recognized by ^{1}H NMR as a methylene group connected with another methylene group and oxygen atom, a part of a −O−CO−(C,C) fragment. The last molecular fragment was identified only by IR spectroscopy as an aliphatic ester substructure.

Fuzzy Definition of Molecular Fragments

J. Chem. Inf. Comput. Sci., Vol. 40, No. 2, 2000 **329**

A set of fragments recognized in the above-described manner must be then further analyzed by a chemist to select the substructures which are likely to enter into the composition of the molecule of an analyzed compound.

6. CONCLUSION

An exactly defined set of substructures, which would enable building of all chemical molecules, does not exist. The complexity and number of molecular fragments in a database depends on numerous objective factors (e.g., the goal of research) and on subjective factors (e.g., the experience of the program designer).

In our studies, spectral methods such as infrared, ultraviolet, Raman, and nuclear magnetic resonance (^{1}H, ^{13}C) spectroscopies were used to establish a set of substructures containing molecular fragments that are active in these spectroscopies. Since the natures of the mentioned spectral methods and their features in the field of substructure identification are different, a fuzzy definition of a substructure was introduced. The concept of such definition is based on a representation of a given set of substructures (molecular fragments) whose neighbors are precisely determined, which have the same framework but differ in some nodes (atoms) or in some edges (bonds).

The concept of fuzziness leads first of all to the recognition of the class of an unknown compound by various spectroscopies. What is important is that a very high probability marks such a recognition. In subsequent steps, the set of fuzzy substructures is analyzed and the chemical fragments linked to the recognized substructures are determined. This enables a considerable reduction in the number of structures candidates, which are finally created from the set of the identified molecular fragments.

REFERENCES AND NOTES

(1) Gray, N. A. B. *Computer-Assisted Structure Elucidation*; John Wiley & Sons: New York, 1986.
(2) Meyer, D. E., Warr, W. A., Love, R. A., Eds.; *Chemical Structure Software for Personal Computers*; ACS Professional Reference Book, American Chemical Society: Washington, DC, 1988.
(3) Dubois, J. E.; Mathieu, G.; Peguet, P.; Panaye, A.; Doucet, J. P. Simulation of Infrared Spectra: An Infrared Spectral Simulation Program (SIRS) Which Uses DARC Topological Substructures, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 290−302.
(4) Attias, R. Substructures Systems and Structural Retrieval Systems, In *Encyclopedia of Library and Information Science*; Kent, A., Ed.; Marcel Dekker: New York, 1992; Vol. 50, Suppl. 13, pp 308−363.
(5) Panaye, A.; Doucet, J. P.; Fan, B. T. Topological Approach of ^{13}C NMR Spectral Simulation: Application to Fuzzy Substructures, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 258−265.
(6) Laidboeur, T.; Laude, I.; Cabrol-Bass, D.; Bangov, I. P. Employment of Fuzzy Information Derived from Spectroscopic Data Toward Reducing the Redundancy in the Process of Structure Generation. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 171−178.
(7) Cabrol-Bass, D.; Laude, I.; Laidboeur, T.; Bangov, I. P. Fuzzy Molecular Graphs in a Decision Support System for Molecular Structure Elucidation. *Am. Inst. Phys.* **1995**, 610−617.
(8) Woodruff, H. B.; Smith, G. M. Generating Rules for PAIRS−A Computerized Infrared Spectral Interpreter. *Anal. Chim. Acta* **1981**, *133*, 545−553.
(9) Sasaki, S.; Fujiwara I.; Abe, H.; Yamasaki, Y. A Computer Program System−New Chemics−for Structure Elucidation of Organic Com-
pounds by Spectral and Other Structural Information. *Anal. Chim. Acta* **1980**, *122*, 87−90.
(10) Woodruff, H. B.; Smith, G. M. Computer Program for the Analysis of Infrared Spectra.−*J. Am. Chem. Soc.* **1980**, *52*, 2321−2327.
(11) Turlson, M. O.; Munk, M. E. Table-Driven Procedure for Infrared Spectrum Interpretation. *J. Am. Chem. Soc.* **1983**, *55*, 2137−2142.
(12) Moldoveanu, S.; Rapson, C. A. Spectral Interpretation for Organic Analysis Using an Expert System. *Anal. Chem.* **1987**, *59*, 1207−1212.
(13) Huixiao, H.; Xinquan, H. ESSESA: An Expert System for Elucidation of Structures from Spectra. 1. Knowledge base of Infrared Spectra and Analysis and Interpretation Programs. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 203-210.
(14) Gribov, L. A.; Dementyev, V. A.; Elyashberg, E. Z.; Yakapov Automation of Spectrochemical Investigations. *J. Mol. Struct.* **1974**, *22*, 161−172.
(15) Andreev, G. N.; Argirov, O. K. EXPIRS, an Expert System for Generation of Alternative Sets of Substructures, Derived by Infrared Spectra Interpretation. *Anal. Chim. Acta* **1996**, *321*, 105−111.
(16) Elyashberg, M. E.; Martirosian, E. R.; Karasev, Yu. Z.; Thiele, H.; Somberg, H. X-PERT: a User-Friendly Expert System for Molecular Structure Elucidation by Spectral Methods. *Anal. Chim. Acta* **1997**, *337*, 265−286.
(17) Schriber, H.; Pretsch, E. Rule-Based System to Derive Automatically Good-List and Bad-List Entries for Structure Generators from Spectra. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 884−891.
(18) Seil, J.; Koehler, I.; von der Lieth, C. V. Interpretation of Infrared Spectra Based on Statistical Approaches. *Anal. Chim. Acta* **1986**, *188*, 219−227.
(19) Passlack, M.; Bremser, W. IDIOTS-A Structure-Oriented Databank System for the Identification and Interpretation of Infrared Spectra. In *Computer-Supported Spectroscopic Databases*; Zupan, J., Ed.; Ellis Horwood: Chichester, England, 1986; pp 92−117.
(20) Luinge, H. J.; van der Mass, J. H. Artificial Intelligence for the Interpretation of Combined Spectral Data. Design and Development of the Spectrum Interpreter. *Anal. Chim. Acta* **1989**, *223*, 135−147.
(21) Ehrentreich, F.; Dietze, U.; Meyer, U.; Schulz, H.; Klötzer, H. M.; Abbas S.; Otto, M. "IR-Spectroscopy"−Computer Program for the Interpretation of Infrared Spectra. *Fresenius J. Anal. Chem.* **1996**, *354*, 829−832.
(22) Dębska, B.; Guzowska-Świder, B. The Methodology of Knowledge Acquisition from the Collection of IR and UV Spectra. *Fresenius J. Anal. Chem.* **1998**, *361*, 235−238.
(23) Ricard, D.; Cachet, C.; Cabrol-Bass, D.; Cachet, C. Neural Network Approach to Structural Feature Recognition from Infrared Spectra. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 202−210.
(24) Novic, M.; Zupan, J. Investigation of Infrared Spectra-Structure Correlation Using Kohonen and Counterpropagation Neural Network. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 454−466.
(25) Klawun, Ch.; Wilkins, Ch. L. Optimization of Functional Group Prediction from Infrared Spectra Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 69−81
(26) Munk, M. E.; Madison, M. S. The Neural Network as a Tool for Multispectral Interpretation. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 231−238.
(27) Schaller, R. B.; Munk, M. E.; Pretsch, E. Spectra Estimation for Computer-Aided Structure Determination. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 239−243.
(28) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. Deriving the 3D Structure of Organic Molecules from their Infrared Spectra. *Vib. Spectrosc.* **1999**, *19*, 151−164.
(29) Affolter, Ch.; Baumann, K.; Clerc, J. Th.; Schriber, H.; Pretsch, E. Automatic Interpretation of Infrared Spectra. *Mikrochim. Acta* **1997**, *14*, 143−147.
(30) Hippe, Z. S.; Fic, G.; Nowak, G. Representation of Common Sense in Chemical Syntheses by Means of Molecular Graphs. *Found. Comput. Decision Sci.* **1994**, *19*, 21−30.
(31) Nowak, G.; Fic, G. Implementation of Commonsense Reasoning in Organic Reaction Prediction System CSB. *Comput. Chem.* **1998**, *22*, 147−152.
(32) Dębska, B.; Guzowska-Świder, B. Application of Knowledge Engineering Environment Program System SCANKEE for Recognition of Structural Units in the Molecule of an Organic Compound. *J. Mol. Struct.* **1995**, *348*, 473−476.