

SIRS-SS: A System for Simulating IR/Raman Spectra.[†] 1. Substructure/Subspectrum Correlation

Jianhua Yao,^{‡,§} Botao Fan,^{*,‡} Jean-Pierre Doucet,[‡] Annick Panaye,[‡] Shengang Yuan,[§] and Jianfeng Li[§]

Institut de Topologie et de Dynamique des Systèmes, CNRS UPRES-A 7086, Université Paris 7, 1, rue Guy de la Brosse, 75005 Paris, France, and Shanghai Institute of Organic Chemistry, 354 Fenglin Road, 200032, Shanghai, China

Received February 2, 2001

An IR/RAMAN spectra simulation system is reported. The development of this software was based on the substructure/subspectrum relationships established for four different structural classes: small molecules, special fragments, atom-centered FRELs, and bond-centered FRELs (FREL: Fragment centered on an Environment which is Limited). Four corresponding knowledge-bases (now, at a pilot stage) are constructed from usual correlation charts or data analyses of large populations of compounds using data mining techniques.

INTRODUCTION

There is a close relation between a chemical substance and its spectra. "Structural elucidation" procedures allow chemists to select a structure corresponding to given spectrum. Conversely, it is also possible to construct the expected spectrum « logically » of a molecule via "spectral simulation" approach. Great efforts were devoted to develop computer-aided structural elucidation systems and, to a lesser extent, spectral simulation systems.

After the first attempts based on mass spectrometry,^{1–3} interest was largely focused on ¹³C NMR by using the one to one correspondence between carbons and resonance signals, which can be supplemented by connectivity information from 2D experiments. Infrared spectra, which allow the identification of bonds or structural fragments combining information from several spectral regions, were also largely used with many artificial intelligence methods [ref 4 and references therein].

Among the well-known systems proposed for structural elucidation, one can cite the following: DENDRAL,^{1–3} ASSIGNER,^{5,6} PAIRS,⁷ SPEKTREN,⁸ CRISE,⁹ STREC,¹⁰ CASE,¹¹ CHEMICS,^{12,13} KISIK,¹⁴ SEAC,¹⁵ EPIOS,^{16–18} SpecInfo,¹⁹ 2D-NMR,²⁰ NMR-SAMS,²¹ and CHORDS.²² We can also cite the following for spectral simulation: ACD/NMR,²⁰ SCNMR,²² based on magnetic resonance and TeleSpec^{23,24} and SIRS²⁵ for infrared spectrum prediction.

Quantum chemistry programs such as Cerius²-IR/Raman,²⁶ MOPAC,²⁷ and GAUSSIAN^{28,29} can also be used to simulate IR/Raman spectra.

Generally speaking, the main methods of IR or Raman spectral simulation are based on different approaches.

1. Calculation of Normal Modes of Vibration via Semi-empirical or *ab Initio*^{29,30} Quantum Methods (for Example, with Cerius²-IR/Raman, MOPAC, or GAUSSIAN Softwares). This approach can often give accurate results but requires the use of a high level basis set and the introduction of a scaling factor to adjust calculated and observed wavenumbers of the normal modes.³¹ The very large time-consuming and memory space requirements heavily limits practical applications, particularly in the cases where one needs to simulate the spectra of a large population of compounds. Other difficulties arise when the analyzed structure is a mixture of conformers.

2. Searching Spectral Similarities from Structural Similarities. Structural similarity is evaluated by nearest neighbor analysis in a structure library, and the predicted spectra are constructed from those of the structurally nearest neighbors.^{32,33} Alternatively, neural networks have been used^{23,24} in the TeleSpec system. The IR spectrum is computed by a "counter-propagation" (CGP) artificial neural network³⁴ at a fixed length descriptors encoding 3D structural features of the target molecule. This method needs to establish a series of models for different categories of compounds. However the similarity permits a more precise description of band shapes with an extended spectral range, whereas only indicating the wavenumbers of characteristic bands. Moreover, the uncharacterized skeletal vibrations may be also taken into account.

3. Substructure/Subspectrum Relationships. Another approach proposed by us in an exploratory version of SIRS²⁵ relies on the association of characteristic frequencies to relevant structural elements extracted from the investigated target molecule. This IR/Raman spectra simulation system SIRS-SS is therefore based on an algorithmic partition of the target molecule into (overlapping) fragments and application of substructure/subspectrum correlation (relationships between an extended set of about 700 molecular fragments and corresponding characteristic IR/Raman wavenumbers, via linked databases). In an alternative approach³¹ the spectra are reconstructed from a substructure search

* Corresponding author phone: +33-1-44274412; fax: +33-1-44276814; e-mail: fan@paris7.jussieu.fr.

[†] This work was supported by Ministry of Foreign Affairs of France (AFCRST PRA M98-03), National Scientific Research Center (CNRS/ASC 7491), and University Paris 7-Denis Diderot in the framework of French-Chinese Scientific Cooperation Program.

[‡] Université Paris 7.

[§] Shanghai Institute of Organic Chemistry.

within a set of elementary fragments or functional group and a neural treatment derives the predicted spectrum from this structural input observation. Layered neural networks have been used also, in a reverse way, for interpretation of IR spectra by recognition of functional groups^{4,35} thanks to their characteristic bands.

Although IR spectroscopy does not give comprehensive structural information (as compared with NMR for instance), its sensitivity and the fact that it can apply to very different types of samples (gases, films, pellets...) and the rather low cost of spectrometers make it a very attractive structural analysis method.

Recently, the needs for IR simulation are rapidly increasing in several areas, such as in environmental chemistry where the IR information is an important parameter in pollutant identifications in toxicity studies. This calls for a fast IR simulation system able to construct spectra for hundreds of molecules in an acceptable time. However, compared with NMR-structural elucidation and simulation systems, few IR simulation systems are available. This prompted us to develop a new IR/Raman spectra simulation system SIRS-SS based on substructure/subspectrum correlation. For the time being, SIRS-SS takes advantage of the information capitalized in the common "correlation charts" published in spectroscopy textbooks. But its knowledge can be enlarged by data mining methods applied to IR databases to refine relationships between structural fragments and corresponding characteristic bands. Compared with the exploratory version of SIRS,²⁵ this new issue gives a better and more complete estimate of coupled vibrations and deformation modes which also convey important structural information. This also gives information about Raman spectra, a feature not considered in the previous version of SIRS.

The SIRS-SS system consists of two main parts: (1) construction of interconnected structural and spectral databases for exploiting substructure/subspectrum relationships and (2) simulation of IR/Raman spectra.

In this first paper, we will present the principle of the SIRS-SS system and the construction of related databases.

1. PRINCIPLE OF SUBSTRUCTURE/SUBSPECTRUM CORRELATION

Generally speaking, IR and Raman spectra reflect the vibrations of bonds in a molecule. They are closely dependent on the molecular structure, mainly on functional groups. Therefore in the study of substructure/subspectrum correlation, we can handle molecular structures from fragments centered on bonds. This description is convenient for stretching vibrations. However, it may be difficult to apply to coupled modes (for instance in-phase or out-of-phase stretching vibrations for C–X₂ or C–X₃ groups). As for bending modes, they generally concern two bonds stemming from a common atom (as in scissoring modes). In both cases, bond-focused fragments do not seem to be the most suited for such an analysis. An easier description is obtained using fragments centered on an atom.

Such structural descriptions are easily carried out with the concept of **Fragment Reduced to an Environment** which is **Limited (FRELs)**,^{36,37} formerly introduced in the framework of the DARC system.^{36–38} Suffice it to say here that the structural organization of a fragment is described by con-

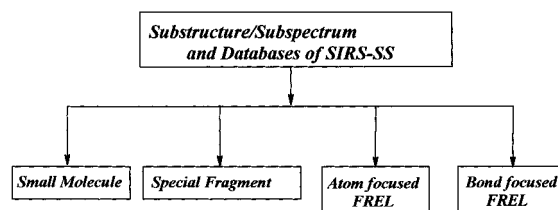


Figure 1. The databases of SIRS-SS,

centric layers stemming from a central element (bond or atom) bearing the relevant information and chosen as the "focus". These points will be developed below.

However, in some cases, vibrations are not localized on a particular bond (or a couple of bonds) and do not affect larger fragments, as the breathing mode of benzene, or the CO stretching vibrations of anhydrides. A description in terms of FRELs would require extended environments. This would weight down the treatment for all applications although only necessary for some rare cases. It seems therefore easier, for such limited examples, to define "special fragments" to be considered as a whole.

The necessity of introducing aliphatic special fragments can be explained through the example of azides $\text{--N}^1\text{=N}^2\text{=N}^3$. If we consider only the vibrations ν_a and ν_{as} , a description with a FREL focused on the central nitrogen is sufficient. Nevertheless it will be difficult to include the influences due to the presence of some groups far from the central nitrogen. This would need to extend the FREL description to three layers from the focus (FREL C). Such an approach would weight down the searching procedures for the whole set of data. This means an unworthy sacrifice for only some rare specific cases. In our azide example, to indicate the presence of supplementary frequencies toward 2400 and 2200 cm^{-1} when a group like CO, CN, or a conjugated bond is bonded to N¹ atom (Fermi resonance), it is more convenient to define this chain as a special fragment.

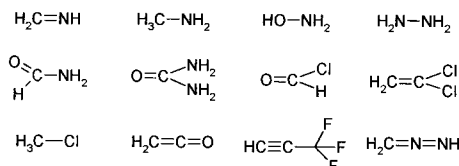
Finally, for small molecules, the wavenumbers of normal modes may be significantly shifted with respect to heavier homologues. A high degree of symmetry can also intervene (as in CCl₄). Therefore, it is not useful to divide them into substructures. They are better dealt with as unique entities.

Subsequently, the substructure/subspectrum correlations can be classed into four categories related to (1) small molecules; (2) special fragments; (3) atom-focused FRELs; and (4) bond-focused FRELs (see Figure 1). According to these four categories, the spectral information in SIRS-SS is therefore organized in four corresponding subdatabases.

The Chemical Abstract Service (CAS) reported statistics about the occurrence (frequencies of presence) of atoms or cycles in compounds. The results published in 1988³⁹ provide some indications for collecting and classifying data. Another element for this classification is the extent of the associated spectral information. Accordingly, no relevant characteristic frequency of analytical interest can be assigned to the C–C stretching vibrations, despite the very frequent existence of this type of bonds in organic molecules, since these frequencies are largely dependent on the organization of the molecular skeleton.

In following sections, we describe these four classes of correlations and their linked pilot databases.

Molecule:	NH₂-NH₂	
IR Data:	3200	NH . Stretching
	1630	NH ₂ Scissoring
	1110	N-N valence
	880	
Stored as:	3210 - 3190	NH . Stretching
	1640 - 1620	NH ₂ Scissoring
	1120 - 1100	N-N valence
	890-870	

Figure 2. Examples of small molecules treated as entities.**Figure 3.** Set of "small molecules" in the database of SIRS-SS.

II. DATABASE OF SMALL MOLECULES

As previously indicated, it is more convenient to consider small molecules as unique entities rather than dividing them into substructures. Two examples for such molecules are shown in Figure 2 and below. In the databases, molecules are recorded as complete structures, and their characteristic IR/Raman bands are stored.

Molecule: H₂C=N=NH

IR Data:	3095	C-H Asym. Stretching
	2880	C-H Sym. Stretching
	2100	CNN Asym. Stretching
	1420	CH ₂ Scissoring
	1170	CNN Sym. Stretching
	900	CH ₂ Wag

Spectra may slightly differ with sources and experimental conditions. To maintain an identical data-structure independently of their provenance for all databases, band positions are defined and stored as possible regions, shown as below:

3105 - 3085	C-H Asym. Stretching
2890 - 2870	C-H Sym. Stretching
2110 - 2090	CNN Asym. Stretching
1430 - 1410	CH ₂ Scissoring
1180 - 1160	CNN Sym. Stretching
910 - 890	CH ₂ Wag

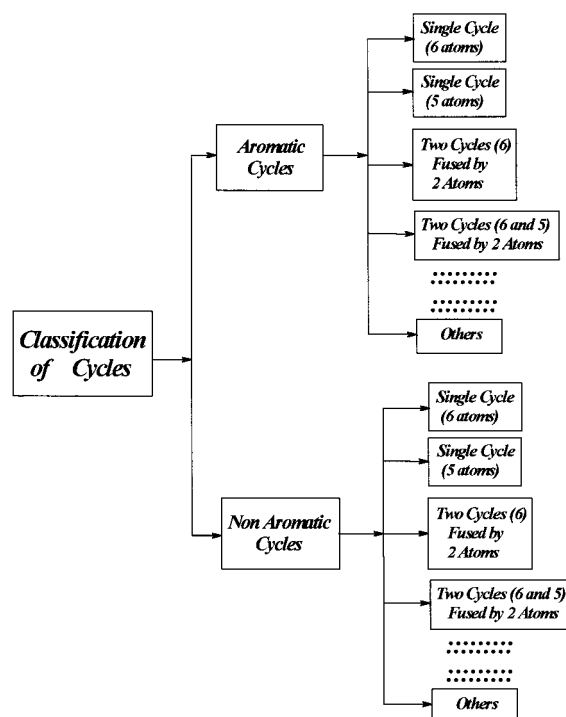
At present, this pilot set is reduced to 12 small molecules, shown in Figure 3.

This set of data in Figure 3 may be easily enlarged to other structures. But this base is not intended to be too largely increased since the aim of the SIRS-SS system is not to systematically retrieve a spectrum from a library but rather to reconstruct it from structural fragments which have been extracted from the target structure.

III. DATABASE OF SPECIAL FRAGMENTS

"Special fragments" concern structural moieties where vibrations are not localized on one or two bonds (e.g. benzene ring) or where a description in terms of FRELs would require an enlarged in depth description (as for anhydrides). These fragments are treated on the whole as unique entities. In SIRS-SS, two distinct subdatabases have been built up to store separately the cyclic and acyclic special fragments.

According to the statistical results reported by CAS, about 89% of known compounds contain one or several cycles. It is, therefore, necessary to consider cyclic special fragments

**Figure 4.** Classification of ring blocks.**Table 1.** Definitions of the Characteristics of a Ring Block

a_i ($1 \leq i \leq 10$)	definitions
A_1 ($a_1 = 0$)	unused
A_2 ($a_2 = 0$)	unused
a_3 ($a_3 = 0$)	unused
a_4 ($0 \leq a_4 \leq 9$)	total of fused atoms
a_5 ($a_5 = 0$ ou 1)	aromatic characteristics: 0 = nonaromatic, 1 = aromatic
a_6 ($0 \leq a_6 \leq 9$)	total of three member cycles
a_7 ($0 \leq a_7 \leq 9$)	total of four-member cycles
a_8 ($0 \leq a_8 \leq 9$)	total of five-member cycles
a_9 ($0 \leq a_9 \leq 9$)	total of six-member cycles
a_{10} ($0 \leq a_{10} \leq 9$)	total of seven-member cycles

as an important specific class in the database. These fragments are further organized according to the characteristics of the ring system (ring-block). These can be summarized in a single "cyclic block property (CBP)" number, calculated according to eq 1, and leads to the ordering indicated in Figure 4

$$\text{CBP} = a_1 \cdot 10^9 + a_2 \cdot 10^8 + a_3 \cdot 10^7 + a_4 \cdot 10^6 + a_5 \cdot 10^5 + a_6 \cdot 10^4 + a_7 \cdot 10^3 + a_8 \cdot 10^2 + a_9 \cdot 10^1 + a_{10} \cdot 10^0 \quad (1)$$

where the a_i ($1 \leq i \leq 10$) represents some ring characteristics. Definitions for a_i ($1 \leq i \leq 10$) are listed in Table 1. The three first parameters (a_1 – a_3) are not yet used but, in the future, may encode other characteristics.

Three examples are shown in Figure 5. The CBP values are CBP = 0000000010 (for (a)); CBP = 0002100020 (for (b)); and CBP = 0002001010 (for (c)), respectively.

Two hundred five cyclic "special fragments" are collected in the database. Some examples of monocyclic aromatic fragments are shown in Figure 6.

The acyclic special fragments are not further classified. This database contains actually 30 records. Some typical examples are shown in Figure 7.

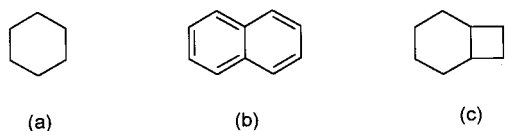


Figure 5. Examples of ring block: (a) CBP = 0000000010; (b) CBP = 0002100020; and (c) CBP = 0002001010.

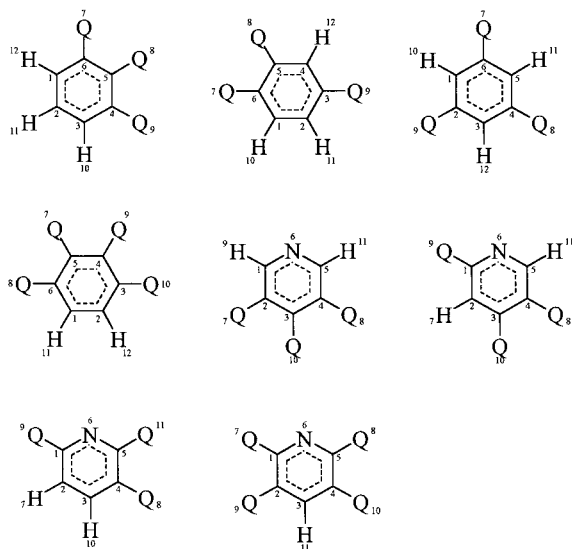


Figure 6. Examples of monocyclic aromatic special fragments (Q: any heavy atom, excluding hydrogen atoms, see below).

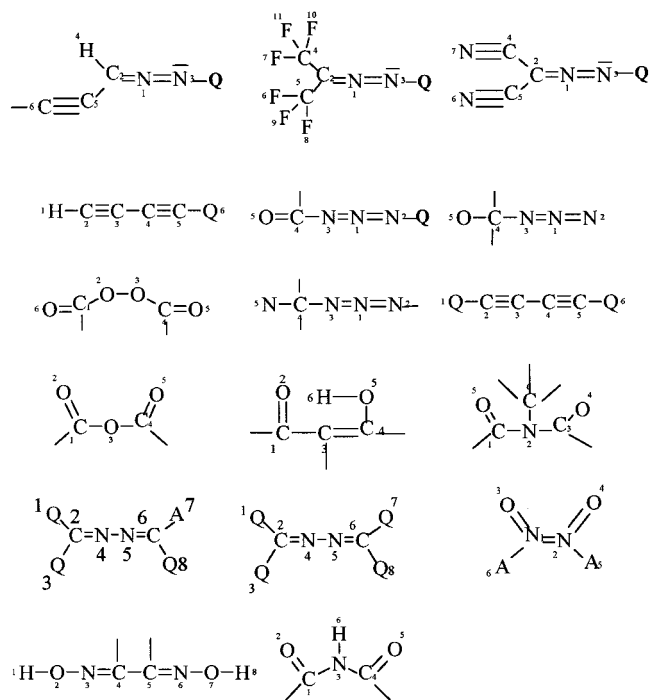


Figure 7. Examples of aliphatic special fragments (Q: any heavy atom, excluding hydrogen atoms; A: aromatic group).

It must be pointed out that ordering is not very important for cyclic special fragments, because we have the ring descriptors. But it becomes necessary for the acyclic ones since the properties of one or several of their atoms may be not specified. Moreover a smaller fragment may be included in a bigger one. Therefore the ordering of fragments in the database is carried out according to the following rules: (1) fragments with highest number of atoms and bonds are

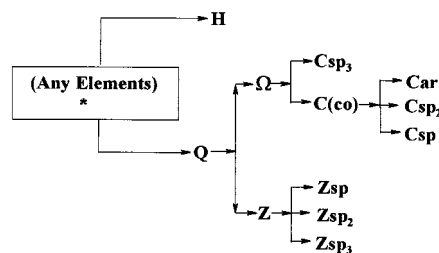


Figure 8. Classification of atoms.

placed on the head of the database and (2) fragments are ordered according to the number of strictly specified characteristics (number of atoms → triple bond → double bond → heteroatom, ...).

IV. DATABASES OF ATOM- AND BOND-CENTERED FRAGMENTS

IV-1. Atom- and Bond-Centered Fragments: FRELs.

FRELs (Fragment Reduced to an Environment which is Limited) and ELCO (Environment which is Limited, Concentric and Ordered) are basic concepts in the DARC system^{36,38} and have been extensively used in topological treatments and structure/property (or activity) relationships.^{37,40} Suffice it to say here that a FREL describes the structural environment of an atom or a bond (considered as the focus of the investigated property) by concentric layers. Within each layer, rules specify the preferential order of occupation of the atomic sites. Depending on the sensitivity of the investigated property to environment influences, this description may be extended to layers more or less apart from the focus (rank A: one layer of atoms; rank B: two layers, and so on ...). Similarly, the nature of the atoms and bonds in the environment of the focus may be completely specified, let undefined or split into equivalence classes ("banalisation", means bleaching the atom chromatism). This flexibility in the treatment of FRELs somewhat relies on the concept of fuzzy logic already introduced in chemistry.⁴¹⁻⁴³

In our case which solely concerns the prediction of wavenumbers for IR/Raman characteristic bands, the description may be limited to FRELs A or B because of the definition of "special fragments". This permits us to take into account more remote influences.

However, the sites in the environment of a bond (or atom) focus do not always affect the band position in a same manner. Furthermore, information extracted from the usual correlation charts is given with more or less details as to structural influences. This leads us to use varied levels of structural description to specify more or less precisely the nature of the atoms. Figure 8 summarizes these various levels of description.

In Figure 8, Q represents any "heavy" element (H excluded); Ω: any type of carbon and Z any heteroelement (C and H excluded); Csp3 means that the element is carbon with hybridization sp³; C(co) includes three situations: Car (aromatic), Csp2 (carbon with sp² hybridization), and Csp (carbon with sp hybridization). Similar levels are managed for Z (any heteroatom, C excluded). Table 2 gives the representations of atoms and their hybridization degrees in a fragment.

Furthermore, because of the various levels of fragments to be processed (special fragments, atom-FRELs, bond-

Table 2. Presentation of Atoms

Csp ₃	$\begin{array}{c} \text{*} \\ \diagup \\ \text{C} \\ \diagdown \\ \text{*} \end{array}$
Csp ₂	$\begin{array}{c} \text{*} \\ \diagup \\ \text{C} \\ \diagdown \\ \text{*} \end{array}$
Csp	$\begin{array}{c} \text{*} \\ \diagup \\ \text{C} \\ \diagdown \\ \text{*} \end{array}$ or $\begin{array}{c} \text{*} \\ \diagup \\ \text{C} \\ \diagdown \\ \text{*} \end{array}$
Zsp ₃	
Zsp ₂	$\begin{array}{c} \text{*} \\ \diagup \\ \text{Z} \\ \diagdown \\ \text{*} \end{array}$ (only one double bond, no triple bond)
Zsp	$\begin{array}{c} \text{*} \\ \diagup \\ \text{Z} \\ \diagdown \\ \text{*} \end{array}$ or $\begin{array}{c} \text{*} \\ \diagup \\ \text{Z} \\ \diagdown \\ \text{*} \end{array}$ (two double bonds or one triple bond)
C(fuzzy)	C-* (no enough information to express the value of hybridization)
Z(fuzzy)	z-* (no enough information to express the value of hybridization)

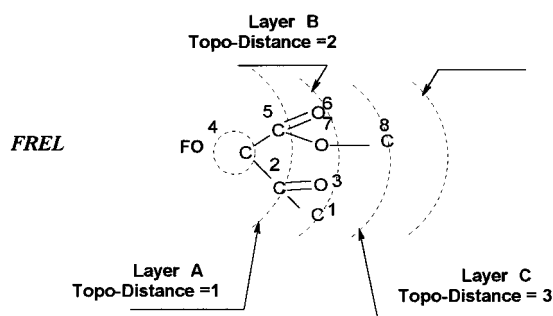
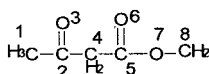
Substructure
(Fragment)

Figure 9. Example of atom-focused FRELs.

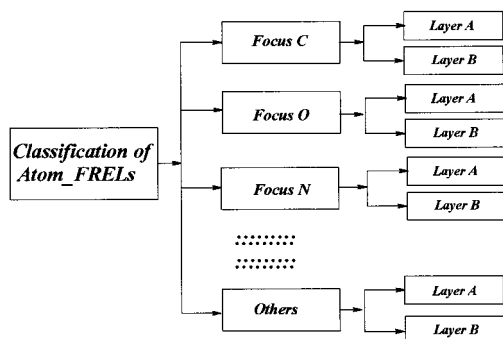


Figure 10. Classification of atom-focused FRELs.

FRELs), it is essential not to consider again a treated element at lower levels. This is carried out by putting up flags on atoms or bonds already considered.

IV-2. Database of Atom-Focused FRELs (Atom FRELs).

Figure 9 shows an example of atom-focused FRELs. In general, the IR/Raman band of the vibration of a bond depends only on the environment of topological distances on one or two layers. The database of atom-focused FRELs is further ordered according to the element constituting the focus and its environment. Figure 10 gives the classification of atom-FRELs.

For FRELs with a same depth (same number of layers), that bearing the most specific "atoms" should be examined before the FRELs containing more generic "atoms". This rule orders the organization of the databases. Examples are given in Figure 11, for FRELs focused on an oxygen atom.

In the corresponding SIRS-SS databases, 129 fragments are described as atom-focused FRELs.

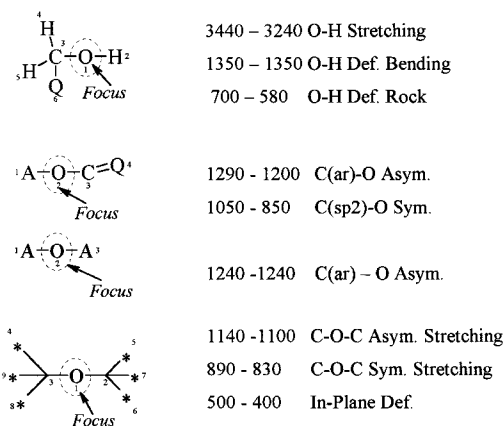


Figure 11. Examples of atom FRELs (focus atom = O) with their A-environment and corresponding wavenumbers of characteristic vibrations (cm⁻¹). *: any atom (including hydrogen); Q: any heavy atoms (excluding hydrogen); A: aromatic carbon; Def: deformation mode; sym, asym: symmetric, asymmetric mode.

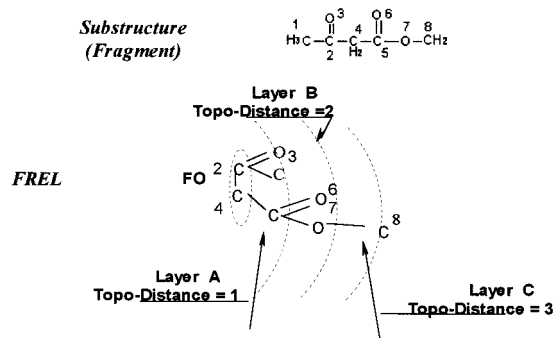


Figure 12. Example of bond-focused FREL.

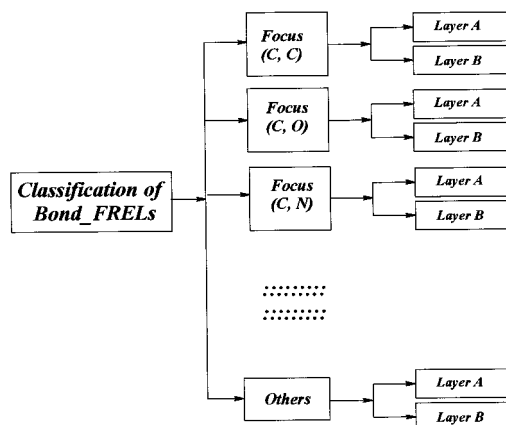


Figure 13. Classification of bond-FRELs.

IV-3. Database of Bond-Focused FRELs (Bond FRELs).

Similar to the atom FRELs base, the database of bond FRELs is further classified according to the nature of atoms constituting the focus bond and the environment of that bond (limited as previously to topological distance 2 from the focus bond). Figure 12 shows an example of bond-focused FREL, and Figure 13 indicates the corresponding classification.

The SIRS-SS database of bond-FRELs gathers for the time being 135 fragments, ordered in the same way as in the database of atom-FRELs. In Figure 14, we show some examples from these databases. They show that the characteristics of atoms located in layer A are completely specified

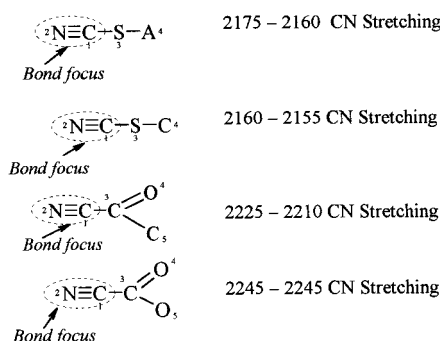


Figure 14. Examples of bond-FREs (focus: C≡N and their A or B environments).

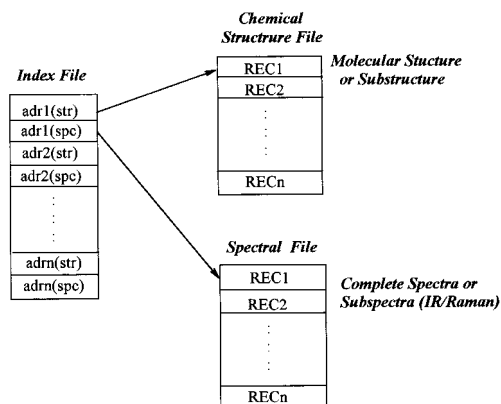


Figure 15. Structure of the SIRS-SS databases.

(fixed), whereas those in layer B (terminal atoms) may be less accurate.

V. STRUCTURE OF DATABASES

The four types of databases in SIRS-SS are constituted with a same data-structure, incorporating three files (index file, chemical structure file, spectral information file) as shown in Figure 15.

V-1. Structure of the Spectral Files.^{44–48} The spectral information involved in the SIRS-SS system is listed as follows: (1) position of the band (in general, expressed as an interval of wavenumbers, because FREs or special fragments have some generic characteristics, and may be found in several different molecules); (2) intensity (IR, Raman), indication of “broad” bands, polarization (or not) for Raman diffusion; and (3) identification (if possible) of the corresponding vibration mode.

A more precise comment can be added for a subspectrum if necessary.

To summarize, the intensities of IR bands and the polarization ratios of the Raman bands closely depend on the variation of the dipolar moment or of the polarization tensor during the vibration. Polarity of bonds and symmetry of vibration modes play an important role. The comparison of intensities of IR/Raman bands and examination of the polarization in Raman are therefore important to identify and attribute simulated and experimentally observed bands.

Band intensities (IR or Raman) are encoded using seven labels: vs (very strong), s (strong), m (middle), w (weak), vw (very weak), v (variable), and x (not specified in correlation charts).

The structure of a data record in this file is organized as indicated below:

```
IRINF>No IRS CS
MAX MIN int-IR shape-IR int-Ram. shape-Ram. Pol. COM(y/n) MODE
....
....
....
COMMENT Information
```

In line 1: “**IRINF>**” is the starting label; **No**, the number of the record; **IRS**, the total number of registered band; and **CS** is the total number of comment information. Line 2. gathers: position of the band (in fact an interval between **MAX**, **MIN** wavenumbers to cope with the (more or less) generic character of the FREs), **IR intensity**, **IR band shape**, **Raman intensity**, **Raman band shape**, **polarization** (Raman), existence of a **comment** (COM= yes/no), and (possibly) indication of the **mode** of vibration (MODE). The last lines of the record are dedicated to possible comments.

V-2. Structure of the Chemical Structure File. In the databases of SIRS-SS, the chemical structures are stored as substructures (except for the database of « simple molecules »). We can only represent a complete molecular structure by a graph according to usual methods. However, these methods are not adapted to represent a fragment. In a fragment, all bonds are represented precisely, but the characteristics of the end atoms may be not completely specified. Such fuzziness can be easily handled thanks to the flexibility of the FREs treatment.^{43,49}

The atom characteristics we consider are as follows: (1) element (atomic number in the periodic table); (2) charge; (3) degree of hybridization; (4) connectivity; (5) abnormal valence; (6) radical; (7) aromatic characteristic; (8) total number of attached hydrogen; and (9) cyclic features.

V-3. Structure of Index File. The technique of index file is widely used in database management. In SIRS-SS system, the index file is used to store information about addresses of corresponding records in the structural and spectral files. The structure of the records in the index file is the following:

Starting Address of a record in the structural file
Starting Address of the corresponding record in the spectral file

A program has been written in SIRS-SS to directly establish this index file when building the structural and spectral databases.

CONCLUSION

We present here the principle of a revised version of our IR/Raman simulation system and the construction of the relevant databases (Substructure/Subspectrum databases) which are up to now at a pilot scale. The design of the system and the construction of these bases rely on (1) statistical results from CAS,³⁹ (2) data mining; (3) FREs structural description; and (4) fuzzy graphs.

The databases of atom-focused FREs and bond-focused FREs are constructed and ordered according to the nature of the focus (atom or bond). This choice for categorizing the structural data is aimed to provide high performance and simplify computer process.

The pilot databases are open systems. This allows easy updating and extension of the knowledge, which leads to an increase in efficiency of simulation. The user can thus complete these databases by introducing his own substructure/subspectrum correlation data.

We will report soon the simulation procedures, the performance of the system, and the results of some comparative studies.

ACKNOWLEDGMENT

Authors are thankful to Dr. J. M. El Hage Chahine for his help with the English language.

REFERENCES AND NOTES

- (1) Lederberg, J.; Sutherland, G. L.; Buchanan, B. G.; Feigenbaum, E. A.; Robertson, A. V.; Duffield, A. M.; Djerassi, C. Applications of Artificial Intelligence for Chemical Inference. I. The Number of Possible Organic Compounds. Acyclic Structures Containing C, H, O, and N. *J. Am. Chem. Soc.* **1969**, *91*, 2973–2981.
- (2) Buchanan, B. G.; Smith, D. H.; Whiter, W. C.; Gritter, R. J.; Feigenbaum, E. A.; Lederberg, J.; Djerassi, C. Applications of Artificial Intelligence for Chemical Inference. 22. Automatic Rule Formation in Mass Spectrometry by Means of the Meta-DENDRAL Program. *J. Am. Chem. Soc.* **1976**, *98*, 6168–6178.
- (3) Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Crandell, C. W. The DEBTRAL Project: Recent Advances in Computer-Assisted Structure Elucidation. *Anal. Chim. Acta* **1981**, *133*, 471–491.
- (4) Robb, E. W.; Munk, M. E. A neural network approach to infrared spectrum interpretation. *Mikrochim. Acta* **1990**, *1*, 131–155.
- (5) Farkas, M.; Markos, J.; Szepesvary, P.; Bartha, I.; Szalontai, G.; Simon, Z. A Computer-Aided System for Organic Functional Group Determinations. *Anal. Chim. Acta* **1981**, *133*, 19–29.
- (6) Szalontai, G.; Simon, Z.; Csapo, Z.; Farkas, M.; Pfeifer, G. Use of IR and ¹³C NMR Data in the Retrieval of Functional Groups for Computer-Aided Structure Determination. *Anal. Chim. Acta* **1981**, *133*, 31–40.
- (7) Woodruff, H. B.; Smith, G. M. Computer Program for the Analysis of Infrared Spectra. *Anal. Chem.* **1980**, *52*, 2321–2327.
- (8) Zippel, M.; Mowitz, J.; Köhler, I.; Opferkuch, H. J. SPEKTREN – A Computer System for the Identification and Structure Elucidation of Organic Compounds. *Anal. Chim. Acta* **1982**, *140*, 123–142.
- (9) Visser, T.; Van Der Maas, J. H. Systematic Computer-Aided Interpretation of Vibrational Spectra. *Anal. Chim. Acta* **1981**, *133*, 451–456.
- (10) Gribov, L. A.; Elyashberg, M. E.; Serov, V. V. Computer System for Structure Recognition of Polyatomic Molecules by I. R., N. M. R., U. V.; M. S. Methods. *Anal. Chim. Acta* **1977**, *95*, 75–96.
- (11) Shelley, C. A. CASE- A Computer Model of the Structure Elucidation Process. *Anal. Chim. Acta* **1981**, *133*, 507–516.
- (12) Abe, H.; Yamasaki, T.; Fujiwara, I.; Sasaki, S. Computer-Aided Structure Elucidation Methods. *Anal. Chim. Acta* **1981**, *133*, 499–506.
- (13) Funatsu, K.; Sasaki, S. Recent Advances in the Automated Structure Elucidation System, CHEMICS. Utilization of two-dimensional NMR Spectral Information. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 190–204.
- (14) Zupan, J.; Penca, M.; Razinger, M.; Barlic, B. KISK – A Combined Chemical Information System for a Minicomputer. *Anal. Chim. Acta* **1980**, *122*, 103–115.
- (15) Debska, B.; Duliban, J.; Guzowska-Swider, B.; Hippe, Z. Computer-Aided Structural Analysis of Organic Compound. *Anal. Chim. Acta* **1981**, *133*, 303–318.
- (16) Dubois, J. E.; Carabédian, M.; Ancian, B. Elucidation Structurale Automatique par RMN du Carbone-13: Méthode DARC-EPIOS. I. Recherche d'une Relation Discriminante Structurale-Déplacement Chimique. *C. R. Acad. Sci. Paris, Ser.* **1980**, *290*, 369–372.
- (17) Dubois, J. E.; Carabédian, M.; Ancian, B. Elucidation Structurale Automatique par RMN du Carbone-13: Méthode DARC-EPIOS. II. Description de l'Elucidation Progressive par Intersection Ordonnée de Sous-Structures. *C. R. Acad. Sci. Paris, Ser.* **1980**, *290*, 383–386.
- (18) Carabédian, M.; Dagane, I.; Dubois, J. E. Elucidation by Progressive Intersection of Ordered Substructures from Carbon-13 Nuclear Resonance. *Anal. Chem.* **1988**, *60*, 2186–2192.
- (19) Barth, A. SpecInfo: An Integrated Spectroscopic Information System. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 52–58.
- (20) <http://www.acdlabs.com>.
- (21) <http://www.specres.com>.
- (22) <http://www.sioc.ac.cn>.
- (23) <http://www2.ccc.uni-erlangen.de>.
- (24) Schuur, J.; Gasteiger, J. Infrared Spectra Simulation of substituted Benzene Derivatives on the Basis of 3D Structure Representation. *Anal. Chem.* **1997**, *69*, 2398–2405.
- (25) 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.
- (26) <http://www.msi.com/solutions/materials/ir.html>.
- (27) <http://www.schrodinger.com/mopac.html>.
- (28) <http://www.gaussian.com>.
- (29) Foresman, J. B.; Frish, A. *Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian*; Gaussian Inc.: 1993.
- (30) Doucet, J. P.; Weber, J. *Computer-Aided Molecular Design: Theory and Applications*; Academic Press: London, 1996.
- (31) Weigel, U. M.; Herges, R. Simulation of infrared spectra using artificial neural networks based on semiempirical and empirical data. *Anal. Chim. Acta* **1996**, *331*, 63–74.
- (32) Baumann, K.; Clerc, J. T. Computer-assisted IR spectra prediction-linked similarity searches for structures and spectra. *Anal. Chim. Acta* **1997**, *348*, 327–343.
- (33) Baumann, K.; Affolter, C.; Pretsch, E.; Clerc, J. T. Numerical structure representation and IR spectra prediction. *Mikrochim. Acta /Suppl/* **1997**, *14*, 275–276.
- (34) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists -An Introduction*; VCH Press: Weinheim, 1993.
- (35) Munk, M. E.; Madison, M. S.; Robb, E. W. Neural networks models for infrared spectrum interpretation. *Mikrochim. Acta* **1991**, *II*, 505–514.
- (36) Dubois, J. E.; Laurent, D.; Viellard, H. Système DARC. Principes des Recherches de Corrélations et Equation Générale de Topo-Information. *C. R. Acad. Sci. Paris Ser C* **1967**, *264*, 1019–1022.
- (37) Dubois, J. E.; Mercier, C.; Panaye, A. DARC Topological System and Computer Aided Design. *Acta Pharm. Jugosl.* **1986**, *36*, 135–169.
- (38) 1. Dubois J. E.; Viellard, H. Système DARC. VII. Théorie de Génération-Description. I. Principes Généraux. *Bull. Soc. Chim.* **1968**, 900–904. 2. Dubois J. E.; Viellard, H. Système DARC. VIII. Théorie de Génération-Description. II. Etablissement du Descripteur Uniligne d'un Segment A_i-B_j: le DEL. *Bull. Soc. Chim.* **1968**, 905–912. 3. Dubois J. E.; Viellard, H. Système DARC. IX. Théorie de Génération-Description. III. Description Générale des Structures par le DEL. *Bull. Soc. Chim.* **1968**, 913–919.
- (39) Stobaugh, R. E. Chemical Abstracts Service Chemical Registry System. 11. Substance-Related Statistics: Update and Additions. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 180–187.
- (40) Dubois, J. E.; Laurent, D.; Viellard, H. Mathématique Appliquée. Système DARC. Description par Propagation Concentrique d'un Environnement Limité. Méthode de Représentation Moléculaire et Uniligne d'un Graphe Systématique. *C. R. Acad. Sci. Paris Ser. A* **1970**, *270*, 228–230.
- (41) Rouvray, D. H. *Fuzzy Logic in Chemistry*; Academic Press: San Diego, 1997.
- (42) Debska, B.; Guzowska-Swider, B. Fuzzy Definition of Molecular Fragments in Chemical Structures. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 325–329.
- (43) Dubois, J. E.; Panaye, A.; Attias, R. DARC Systems: Notions of Defined and Generic Substructures: Filiation and Coding of FREL Substructure(SS) Classes. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 74–82.
- (44) Roeges, N. P. G. *A Guide to The Complete Interpretation of Infrared Spectra of Organic Structures*; Wiley: New York, 1997.
- (45) Daimay, L. V.; Colthup, N. B.; Fatelry, W. G.; Grasselli, J. G. *The handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*; Academic Press Inc.: Boston, 1991.
- (46) Socrates, G. *Infrared Characteristic Group Frequencies, Tables And Charts*, 2nd ed.; Wiley: New York, 1994.
- (47) Pretsch, E.; Clerc, T.; Simon, W. *Tables of Spectral Data for Structure Determination of Organic Compounds*, 2nd ed.; Springer-Verlag: 1989; Rev. translation of *Tabellen zur Strukturaufklärung organischer Verbindungen mit spektroskopischen Methoden*.
- (48) Smith, B. *Infrared Spectral Interpretation – A systematic Approach*; CRC Press: Boca Raton, 1999.
- (49) Panaye, A.; Doucet, J. P.; Fan, B. T. Topological Approach of ¹³C NMR Spectral Simulation: Application to Fuzzy Substructures. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 258–265.