# SIRS-SS: A System for Simulating IR/Raman Spectra. 2. Procedures and Performance

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

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 June 24, 2001

This paper is devoted to the description of procedures used in our IR/RAMAN spectrum simulation system, based on substructure/subspectrum correlations established between linked databases. The search is performed in the following order: small molecules/specific fragments/atom centered FRELs (FREL: FRagment centered on an Environment which is Limited)/bond focused FRELs. Comparative study with several reported methods has been carried out to show good performance of this software both for IR and RAMAN spectrum simulation.

## INTRODUCTION

The purpose of SIRSS-SS is mainly to very rapidly recall, for a bench chemist, wavenumbers of characteristic vibrations according to usual correlation charts, largely used in preliminary structural analysis.

In our precedent paper,[1] we reported the principle of a computer aided simulation system, SIRS-SS, for predicting IR/RAMAN spectra. Based on the principle of substructure/subspectrum correlation, relationships have been established between an extended set of about 500 molecular fragments and corresponding characteristic IR/RAMAN wavenumbers, via linked databases. We focused attention on the building up of four pilot bases, Small Molecules (SM), Specific Fragments (SF), Atom-Centered FRELs (ACF), and Bond-Centered FRELs (BCF). In the present paper, we will present the general strategy, search procedures, and association substructure/subspectra used in this system, and we will evaluate its performance by carrying out a series of comparative studies with other methods and reported systems.

In fact, the spectral simulation procedure includes a number of molecular graph treatments, such as determination of rings, extraction of varied fragments either precise (specifying the nature of the atoms), or more generic (with possibly bleaching, that is taking no account on the nature of the atoms), homomorphism, isomorphism, etc. Ordering strategy for the search into the databases is also an important feature for the precision and the speed of the simulation process.

## SIMULATION

**1. General Strategy.** Before presenting the general strategy of SIRS-SS, we will briefly reiterate the structural fragments on which the system is based (details may be found in our previous paper). The IR and RAMAN spectra reflect the vibrations of bonds in a molecule. Therefore the Bond-Centered FRELs are normally convenient to describe the substructures associated with subspectra. However, the complexity of molecular structures leads to many special cases such as coupled modes (e.g. in-phase or out-of-phase stretching vibrations for $C-X_2$ or $C-X_3$ groups), vibration delocalized on several bonds, etc. Taking into account these specific situations besides the general behavior and with a view to simplifying the treatments, three types of databases have been constructed: Special Fragments (SF), Atom-Centered FRELs (ACF), and Bond-Centered FRELs (BCF). Finally a database "Small Molecules" was added in order to shorten the treatment when the target molecule is small and simple. This also allows for giving a better prediction in such cases where the characteristic wavenumber may be different from those of heavier homologues, or symmetry significantly modify the spectrum. The construction of these four databases has been reported in precedent paper.[1]

Now we must define the priority order of these four databases in the searching procedure. The principle of this step is to avoid any redundancy (even partial) and any loss of substructures. In the case of partial redundancy, the combination of structural elements (from an already treated substructure) with other fragments (not yet examined) can generate new (unwanted) substructures, resulting in supplementary unexpected spectral peaks**.** This can be avoided by masking, step by step, the already considered fragments, and by very carefully looking at the association substructre/subspectrum. Generally, complete structure scanning begins with the search for the target molecule in the Small Molecule database. If this first step is successful, we get the corresponding IR/Raman spectral information of the whole molecule from the associated spectral database, and the system will stop the search process. In contrast, if this step fails, the target structure will be mapped with particular fragments. As the database of special fragments consists of two subdatabases (database of special cyclic fragments and database of special acyclic fragments,) the searching order is first the cyclic fragments and then the acyclic fragments. Generally a special fragment contains more atoms (and bonds) than an Atom-Centered FREL or Bond-Centered FREL. The corresponding spectral information is a group of wavenumbers covering all atoms and bonds involved in the special fragment. The found fragment will be masked

* Corresponding author phone: +33-1-44274412; fax: +33-1-44276814; e-mail: fan@paris7.jussieu.fr.
‡ Université Paris 7.
§ Shanghai Institute of Organic Chemistry.

10.1021/ci010061w CCC: $22.00 
Published on Web 03/03/2002

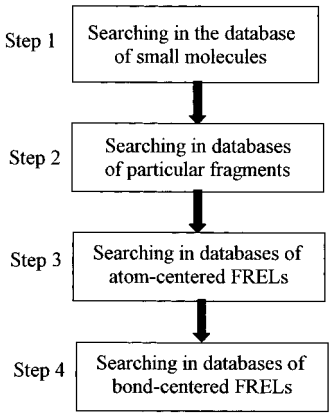



**Figure 1.** Priority order of search procedure in SIRS-SS system.

after this step, leading to a rapid decrease of the number of untreated atoms to be examined and then accelerating the search. On the other hand, as indicated in the preceding paper,[1] 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. If a special fragment is broken into several substructures (Atom-Centered FRELs or Bond-Centered FRELs), the simulated spectral peaks will be the result of simple addition of several subspectra asociated to the corresponding Atom-Centered FRELs or Bond-Centered FRELs. This is often wrong, because the delocalization of vibrations will be ignored. It is why the search of special fragments should be computed before examining Atom-Centered FRELs and Bond-Centered FRELs.

The next step is to deal with Atom-Centered FRELs and Bond-Centered FRELs. As the Atom-Centered FRELs convey information relative to several bonds (coupled modes in $CX_2$ or $CX_3$, deformation modes ...), it is therefore preferable to treat these particular FRELs before the more general cases of Bond-Centered FRELs. Now we can summarize the priority order, as shown in Figure 1.

It is possible that in some cases, it remains atoms or bonds not considered after the whole walk along the four databases, because there is no relevant spectral information about these fragments stored in the correlation charts, and so there is no corresponding records in our bases. The similarity search is then employed to reduce this type of damage. In SIRS-SS, the similarity searching is different from those reported by other authors.[2,3] We will describe this process later in this paper.

The priority order of search, described above, is very efficient. Joined to a step by step indication of the structural elements investigated, it allows for avoiding redundancy and mismapping.

**2. Structural Information.** In general, information related to atomic features can be directly extracted from the structural data file (the target molecule being described via a standard ISIS-DRAW file in MOL (MDL) format). For example, the following information, regarding atom properties, is directly obtained from the file:

1. Element;
2. Charge;
3. Radical character;
4. Valence;
5. Isotope

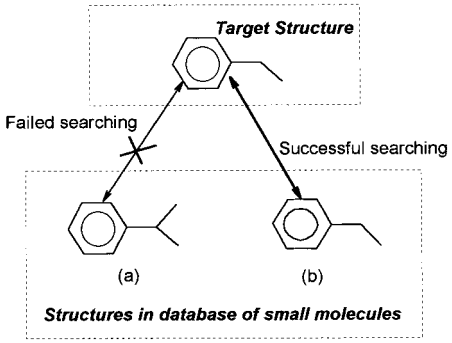



**Figure 2.** Example of homomorphism.

Other atom properties, such as hybridization, connectivity, total number of attached hydrogen atoms, and aromatic character, are also easily deduced from the structural data file by simple analyses. By contrast, some properties, for example cyclic structures, equivalent sites, etc., must be computed with specific routines. Fortunately, during the development of a series of spectral simulation systems, a number of molecular graph handling routines were developed based on algorithms proposed in our laboratory, for instance, the perception of the smallest set of smallest rings (SSSR),[4] the detection of equivalent sites.[5,6] These routines are applied to the structural analysis of the target molecule. For example, the detection of fused rings is used to define the cyclic character of an atom in the target structure, since some special fragments may contain such cyclic groups.

**3. Searching in Database of Small Molecules.** It is well-known that a 2D chemical structure can be considered as a graph in graph theory. The mapping of a target structure on a molecule of the database corresponds to an homomorphism of fully colored graphs. It means that for each atom and each bond of the target**,** we should find their rigorous and unique correspondences with elements of the database. This problem can be solved using canonic numbering methods based on the partitioning of atoms into classes of equivalent sites or by isomorphism procedures such as Ullmann's algorithm,[7] since the homomorphism can be considered as a particular case of isomorphism.

Ullmann's algorithm has been successfully used for several years in structure or substructure searching in chemistry.[8] It can be described as search for isomorphism of subgraphs. In the search for a substructure into a structure, the basic step can be expressed as the following: if an atom in the structure is candidate to correspondence with an atom in the substructure, the neighboring atoms in the substructure must have an equivalent in the molecule. The algorithm proceeds by building a correspondence matrix between putative candidates and updating it by examination of the atoms neighborhood.

It is also implemented in our SIRS-SS system. However, its uses are monitored by the system according to the types of databases. For the searching of target in database of small molecules, the method of exact mapping is applied (see Figure 2).

**4. Searching in Databases of Special Fragments.** In general, a large or complicated molecule cannot be matched in the database of small molecules, because there are only few small (simple) molecules in it. It is known that a

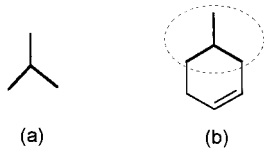



**Figure 3.** Fragment (b) includes fragment (a).

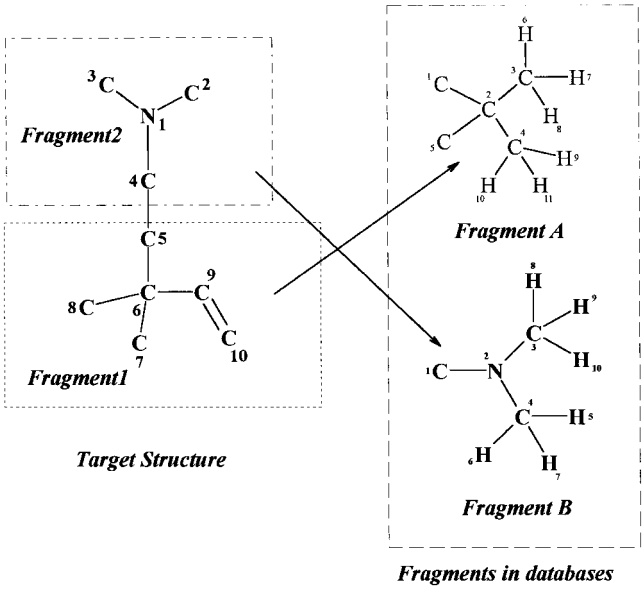



**Figure 4.** Example of searching in the database of special acyclic fragments.

molecule has its corresponding properties,[9] we can present the relation between a type of property, here the spectral property, with molecular structure by eq 1:

$$\text{Spectral Property} = f(\text{Structure}) \qquad (1)$$

Close relationships between substructures and IR/Raman subspectra have been investigated.[10,11] To extract the useful IR/Raman subspectra for target molecule from corresponding spectral databases, the structure should be divided into varied substructures by searching procedures. As we have discussed above, the first class of substructures to be considered is that of the special fragments, generally associated to vibration modes delocalized over several bonds. In SIRS-SS, there are two databases of special fragments: 1. cyclic special fragments and 2. acyclic special fragments. Searching is processed examining first cyclic fragments and then acyclic ones**.** This order allows us to avoid the misextraction of substructures, as shown on an example in Figure 3.

Ullmann's isomorphism algorithm[7] was applied to the search in databases of special fragments. An example is illustrated in Figure 4.

**5. Searching in Databases of Atom-Centered FRELs.** In general, special fragments are bigger than atom or bond FRELs (which are limited to rank A or B from the focus).

After scanning the databases of special fragments, the next step is the search in the database of atom-centered FRELs. Because of the different natures of central atoms (focus), heteroatom or carbon, varied hybridization degrees, it is more efficient to define a priority order for the search. Figure 5 shows this priority order. Furthermore, inside these classes,

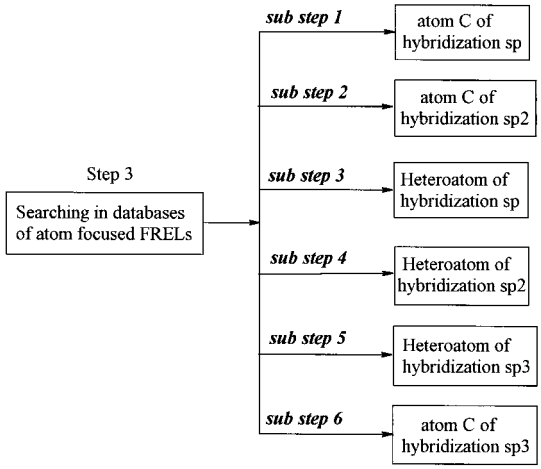



**Figure 5.** Searching rules for atom-centered FRELs.

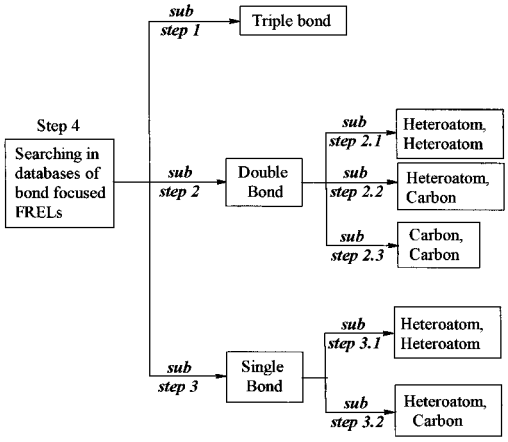


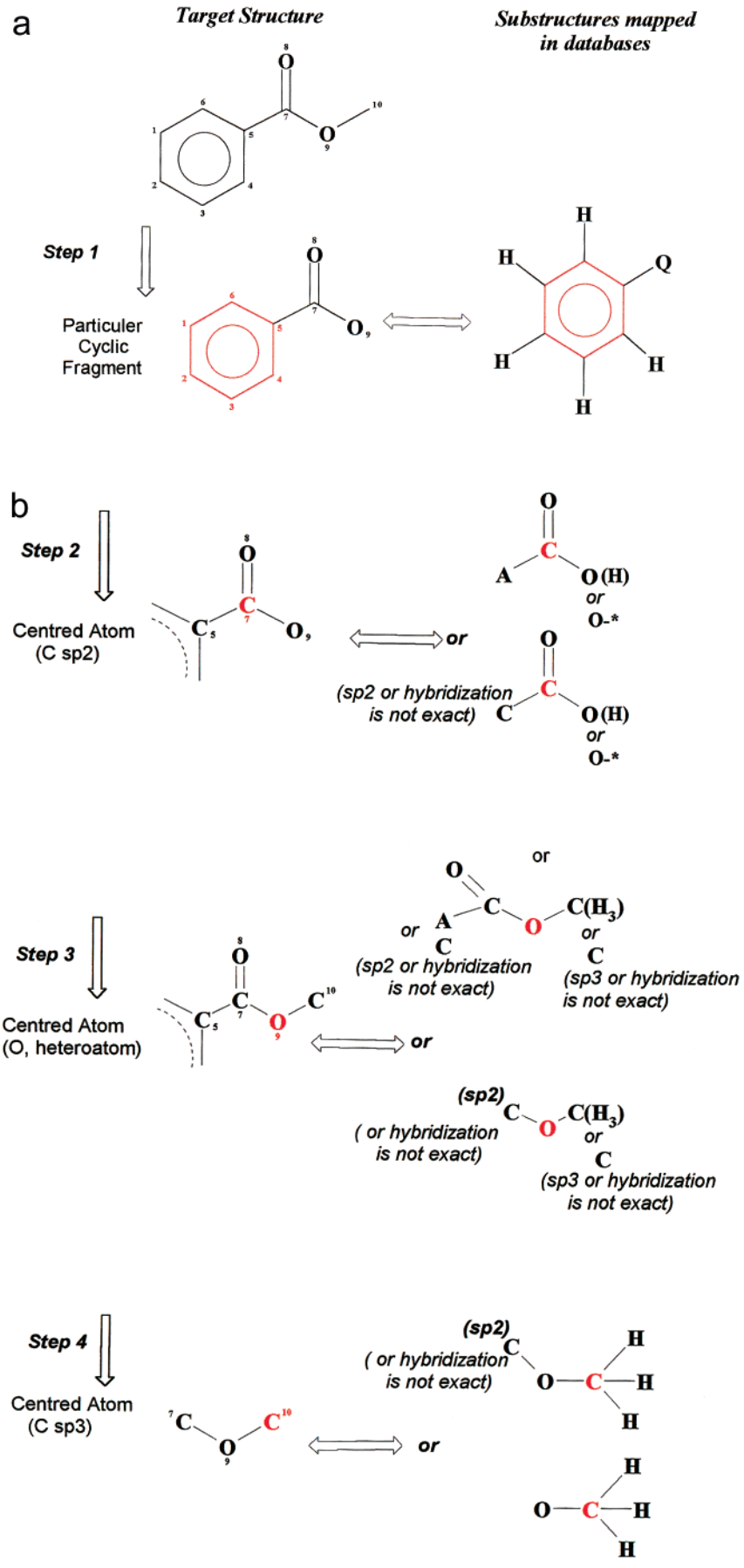

**Figure 6.** Searching rules for bond-centered FRELs.

FRELs are arranged so that more specific fragments are considered before more general ones, so that we get the most precise available information as to characteristic wavenumbers. Similar to special fragments, the search for atom-centered FRELs in the target molecule is also an isomorphism mapping. We can therefore use again Ullmann's algorithm for this procedure.[7]

**6. Searching in Databases of Bond-Centered FRELs.** The search in the database of bond-centered FRELs is similar to that of atom-centered FRELs. Only a new priority order should be defined, as shown in Figure 6.

We now illustrate these steps of simulation in SIRS-SS by an example shown in Figure 7a−c. The target molecule is not included in the "Small Molecules" database. So the search for special fragments is performed, leading to the identification of a monosubstituted benzene ring (Figure 7a). Then we look for atom-centered FRELs, successively focusing on sp2 carbons, heteroatoms (oxygen), and sp3 carbons, and consider either a precise description (with specified characters) or a more fuzzy one, not specifying hybridization for example (Figure 7b). A similar search is carried out with bond-centered FRELs (Figure 7c), and the final spectrum, simulated with SIRSS, is indicated in Figure 8 with a listing of the recognized fragments and associated information or a graphical display.

The simulated IR spectrum is shown in Figure 8.


a
Target Structure
Substructures mapped in databases
Step 1
Particuler Cyclic Fragment
b
Step 2
Centred Atom (C sp2)
or
(sp2 or hybridization is not exact)
Step 3
Centred Atom (O, heteroatom)
(sp3 or hybridization is not exact)
(sp2)
( or hybridization is not exact)
Step 4
Centred Atom (C sp3)

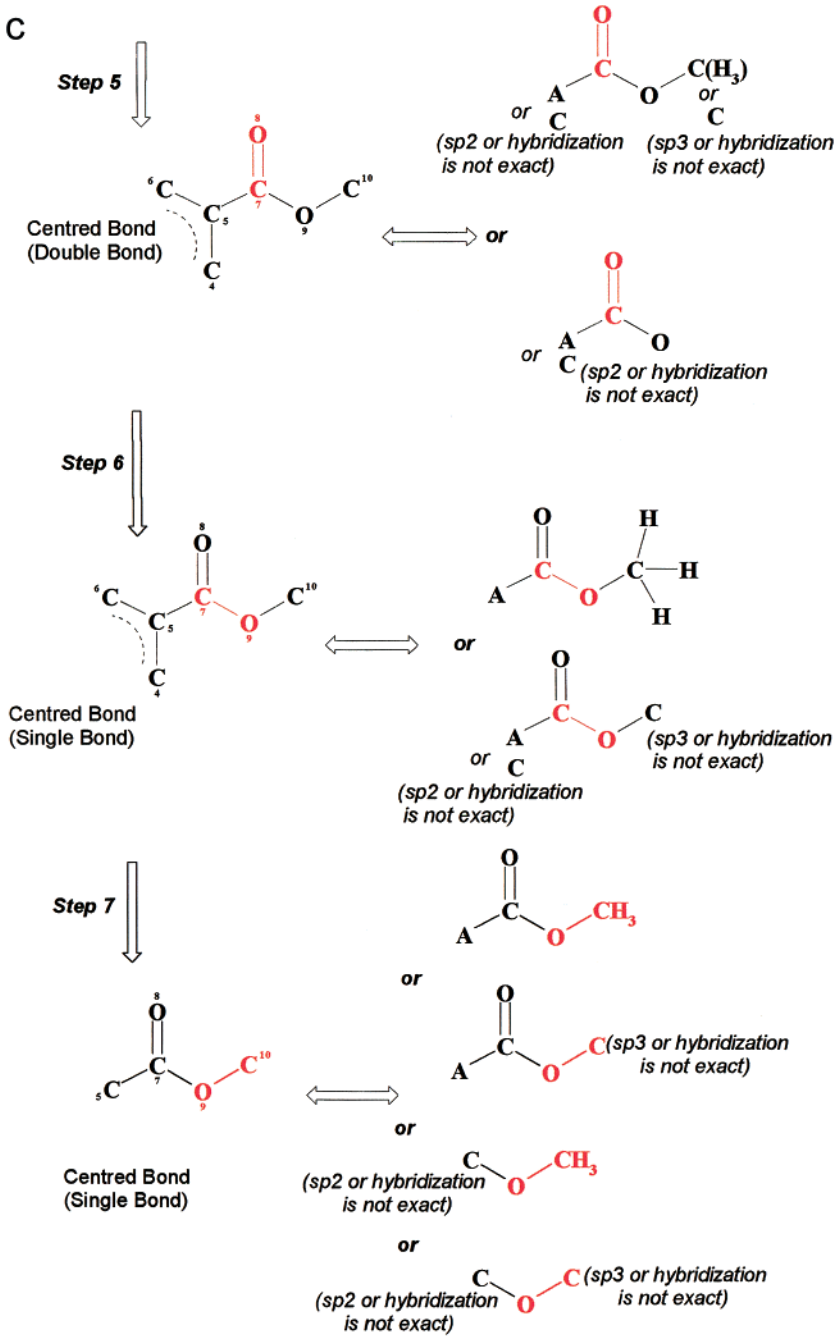



**Figure 7.** a. Simulation procedure of IR/Raman spectra: searching *for special fragments (note: there* is no candidate in the database of small molecules matching the target molecule). b. Simulation procedure of IR/Raman spectra: searching in databases of atom-centered FRELs (note: atom 5, C, in target structure has been already treated in step 1 and is not considered furthermore). c. Simulation procedure of IR/Raman spectra: searching in databases of bond-centered FRELs (note: atom 5, C, in target structure has been treated in step 1).

**7. Similarity Searching.** SIRS-SS is an open system. Users can construct their private pilot databases in order to supplement the original databases. But we can never be sure that the databases of atom-centered and bond-centered FRELs include relevant information for all FRELs. In the case where there is no exactly matching substructure, the similarity searching is computed in order to provide useful information. The system will try to find the atom-centered or bond-centered FREL in databases with the maximal similitude with the query. However, this procedure is controlled by some constraints. The similarity is limited to the environment around the central atom or bond, whereas the FREL focus (central atom or bond), extracted from the target molecule, must be identical to that of a FREL found in the databases. So, classes of equivalence are defined to decide on environment similarity: Cl or Br; O or S or N, aromatic C or $sp_2$ C or $sp_3$ C; any type of C...

An example of similarity searching is shown in Figure 9. In this example, fragment A is more similar to the target than fragment B, and the corresponding wavenumber for CO stretching is retained as a first approximation for the target.

The frequency of stretch vibration of C═O in $CH_3COCl$ is 1805 $cm^{-1}$, and that in $CH_3COBr$ is 1812 $cm^{-1}$.[12] In SIRS-SS, there is information about the C═O stretching of

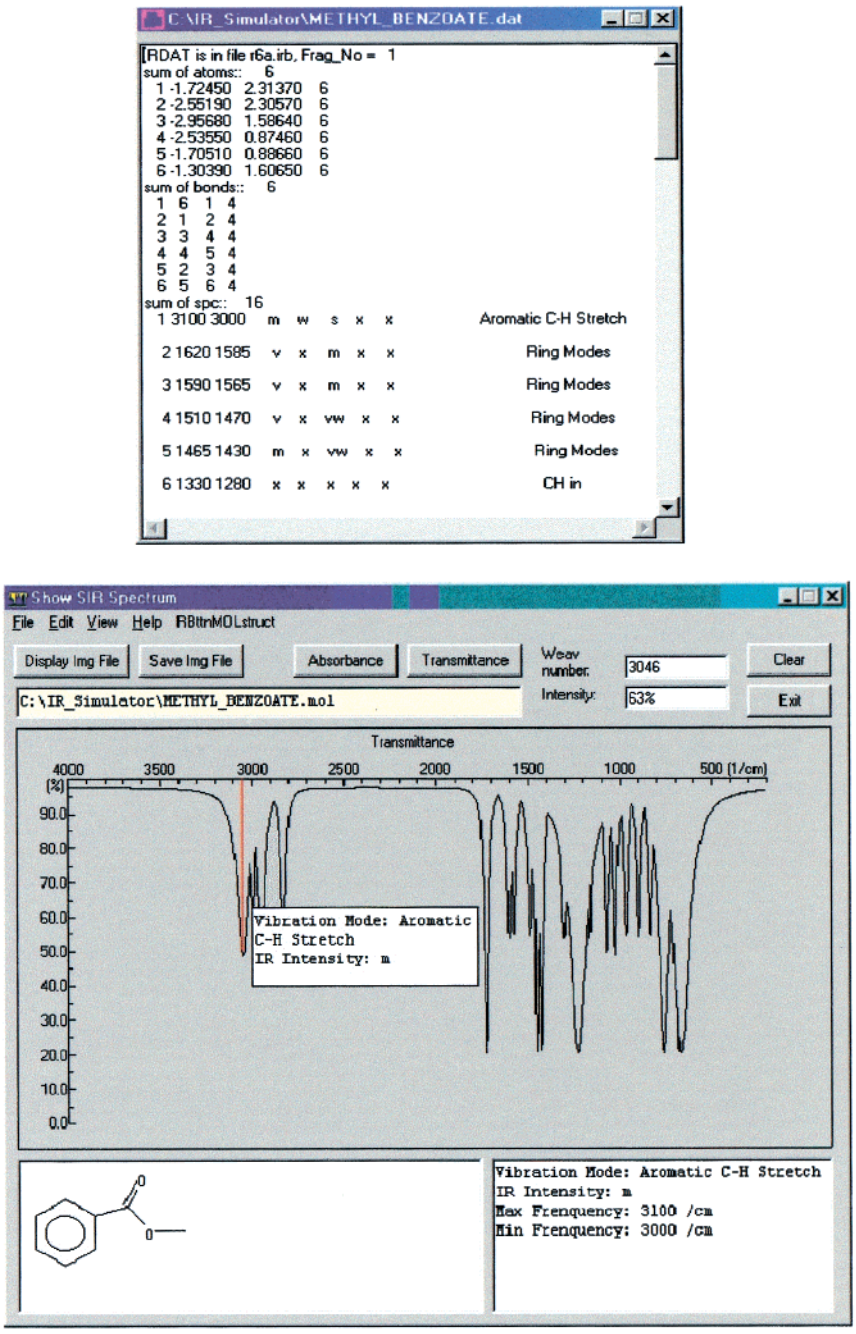



**Figure 8.** Simulated IR spectrum. Above: text representation (partial); below: reconstructed spectrum.

fragments C and D 1810−1770 $cm^{-1}$ (stretching, for fragment C), 1800−1750 $cm^{-1}$ (stretching, for fragment D).

## PERFORMANCE

Performances of SIRS-SS system can be evaluated by comparison of predicted spectra with experimental ones. However a fine quantification of these performances is rather difficult, insofar as one does not a priori know for a given structure the number of characteristic frequencies to be retrieved since bearing relevant information. Some of the (3N-6) normal modes may be inactive, or wavenumbers may be heavily dependent on the atomic framework (as C−C stretching modes for example), or intensities may be too weak for an unambiguous detection. This contrasts for example with $^{13}C$ NMR where there must be a one to one correspondence between the number of carbons in the structure and the number of peaks in the spectrum (unless chance coincidence or magnetic equivalence). However, from numerous examples[11] we can note the following:

An important characteristic is the high speed of SIRS-SS: Once the target molecule is input into the computer, simulating its IR/Raman spectra generally requires less than 1 s on a PC (for a medium-sized molecule of about 50 atoms).

All structural elements are explored once and only once thanks to the systems of flags put on elements already examined in the hierarchical treatment of fragments and FRELs. Relevant spectral information may be present or not, according to the more or less exhaustiveness of the correlation charts at the origin of our databases. But the ordered

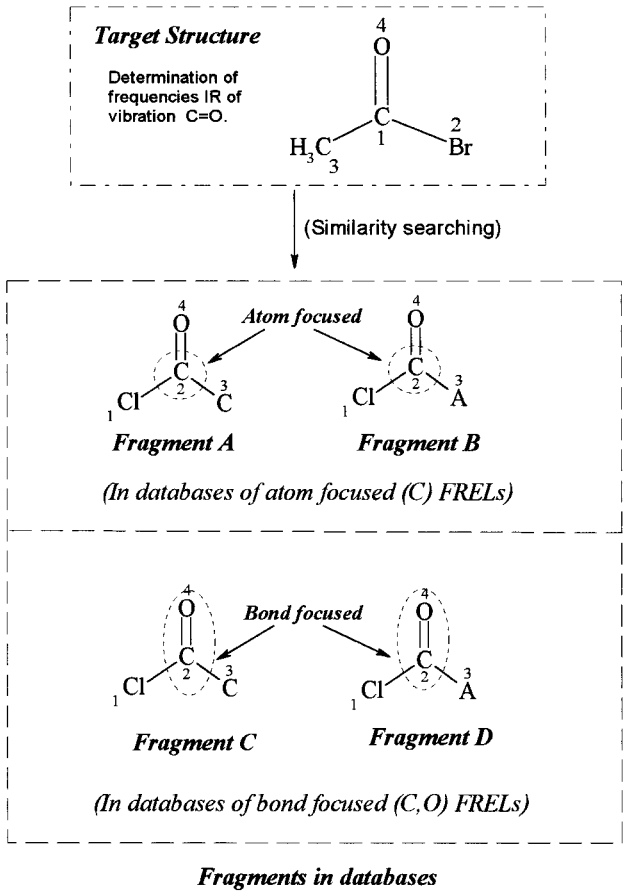



**Figure 9.** Example of similarity searching. C (fragment C) means sp3 carbon; A (fragment D): aromatic carbon.

exploration of these diverse databases guarantees that no silence nor noise may appear in the structural treatment.

In a very large majority of cases, there exists in the experimental spectrum a peak within the wavenumber range for predicted bands, possessing convenient characteristics (shape, depolarization ration in the Raman).

It is more difficult to compare SIRS-SS with other approaches of spectral simulation, since (at least for IR spectra) the aims of existing systems are clearly quite different.

Quantum chemistry programs (such as GAUSSIAN) not only allows for computing normal modes of vibration but also (and there is undoubtedly their major interest, and widest applications) gives information about electron populations, orbital energy, and so on. However the need of sophisticated basis sets required to get good predictions put this approach at disadvantage, due to the very high CPU time requirements.

On the other hand, simulations based on neural networks[13,14] are expected to give a more accurate description of the absorption profile, after learning from neighboring examples. This requires training various specialized networks devoted to more or less specific classes of chemical compounds.

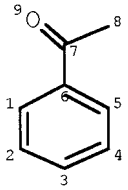
As already indicated, the aim of SIRS-SS is to very rapidly recall characteristic vibrations which may be useful for a bench chemist in preliminary structural analysis.

**1. Comparison with *ab Initio* Method (GAUSSIAN 94).** The following example has been computed with both GAUSSIAN 94 (6-31G*, scaling factor 0.8929) and SIRS-SS on a PC Pentium II (400 MHz, 128Mo RAM,Windows 98). The results are gathered in Table 1.

For another example, the results are given in Table 2.

A series of test molecules have been simulated for their IR/Raman spectra. The results show that with *ab initio* method (GAUSSIAN 94[15]), the calculations are much more time-consuming. Furthermore, high level basis sets (and a scaling factor) are necessary to hope for obtaining good results. This will increase again the computation time. Therefore it is impossible to widely apply this method in practical works, in particular for the spectral simulation of a huge population of compounds.

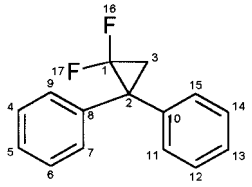
**Table 1.** Comparison of Results (G: GAUSSIAN 94; S.-SS: SIRS-SS)[a]



| experimental frequencies[6] | | | GAUSSIAN 94 | | | SIRS-SS | | | |
|---|---|---|---|---|---|---|---|---|---|
| freq | IR inten | Raman inten | freq | IR inten | Raman inten | freq (max.) | freq (min.) | IR inten | Raman inten |
| 371 | | 2.5 | | | | | | | |
| 588 | S | 2 | 586 | 16.85 | 1.3 | 630 | 605 | W | M |
| 618 | | 0.5 | | | | 630 | 605 | W | M |
| 692 | S | | 684 | 34.96 | 0.53 | 710 | 665 | S | X |
| 733 | | 13 | 709 | 0.88 | 11.26 | | | | |
| 781 | S | SH | 763 | 41.16 | 0.32 | 805 | 720 | S | X |
| 957 | M | 1.5 | 936 | 26.05 | 0.76 | 970 | 970 | X | X |
| 1000 | | 40 | 1004 | 1.94 | 13.25 | 1010 | 990 | VW | VS |
| 1028 | M | 15 | | | | 1040 | 1016 | M | M |
| 1076 | | 6 | | | | 1100 | 1000 | X | X |
| 1267 | S | 10 | 1246 | 176.45 | 22.54 | 1330 | 1280 | X | X |
| 1360 | S | | 1382 | 39.98 | 3.27 | 1380 | 1370 | W | M |
| 1600 | S | 25 | 1613 | 18.96 | 67.02 | 1620 | 1585 | V | M |
| 1685 | S | 13 | 1774 | 259.49 | 34.73 | 1720 | 1710 | VS | M |
| 3072 | | 9 | 3030 | 15.91 | 154.76 | 3100 | 3000 | M | S |

[a] CPU time: GAUSSIAN 94: 3 h 35 min. 47 s. SIRS-SS: 1.37 s. Intensity abbreviations: VS, very strong; S, strong; M, medium; W, weak; VW, very weak; SH, shoulder; X, not. Indicated in the literature (according to refs 44−48 in our preceding paper[1]).

**Table 2.** Comparison of Results (G: GAUSSIAN 94; S.-SS: SIRS-SS)[a]

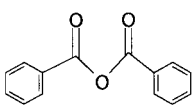



| experimental frequencies[6] | | | GAUSSIAN 94 | | | SIRS-SS | | | |
|---|---|---|---|---|---|---|---|---|---|
| freq | IR inten | Raman inten | freq | IR inten | Raman inten | freq (max.) | freq (min.) | IR inten | Raman inten |
| 82 | | 28 | | | | | | | |
| 263 | | 7 | | | | | | | |
| 285 | W | 5 | 276 | 0.95 | 3.99 | | | | |
| 425 | | 3.5 | 425 | 0.11 | 0.51 | | | | |
| 540 | S | 2 | 542 | 18.67 | 2 | | | | |
| 617 | | 4.5 | 605 | 0.37 | 6.12 | 630 | 605 | W | M |
| 703 | S | | 701 | 64.81 | 2.9 | 710 | 665 | S | X |
| 711 | | 11 | 693 | 2.55 | 11.82 | 710 | 665 | S | X |
| 762 | S | | 771 | 39.07 | 0.9 | 805 | 720 | S | X |
| 980 | S | 4 | 991 | 21.55 | 1.37 | 960 | 900 | S | V |
| 1003 | SH | 38 | 1013 | 11.82 | 15.27 | 1010 | 990 | VW | VS |
| 1029 | S | 4 | 1046 | 61.8 | 4.74 | 1030 | 990 | S | W |
| 1212 | S | SH | 1239 | 85.94 | 4.25 | 1200 | 1180 | M | V |
| 1368 | S | 1 | 1377 | 64.4 | 10.37 | 1420 | 1400 | S | W |
| 1453 | S | 1.5 | 1483 | 62.27 | 5.49 | 1465 | 1430 | M | VW |
| 1600 | M | 8.5 | 1614 | 9.79 | 20.17 | 1620 | 1585 | V | M |
| 3016 | M | 11 | 3020 | 66.09 | 34.57 | 3020 | 3000 | S | S |
| 3057 | | 14 | 3027 | 11.23 | 460.52 | 3100 | 3000 | M | W |

[a] CPU time: GAUSSIAN 94: 1 day 19 h 57 min. 48 s. SIRS-SS: 0.93 s. Intensity abbreviations: VS, very strong; S, strong; M, medium; W, weak; VW, very weak; SH, shoulder; X not. Indicated in the literature (according to refs 44−48 in our preceding paper[1]).

**2. Comparison with the Software TeleSpec.** The software TeleSpec[16] can simulate quickly the IR spectrum for a target molecule. It combines the search for known spectra in databases and the simulation for unknown spectra. When a target molecule is input, the system performs a search in all attached databases. If the target molecule is found, the system directly extracts the corresponding IR spectrum and displays it. If the search fails, the system simulates the IR spectrum thanks to artificial neural network algorithms. This strategy permits user to get the experimental spectra if they are available. The computation time is quite suitable for practical use. But unfortunately, up to now, it cannot provide any information on Raman bands.

Another shortcoming of this software is that some structures cannot be accepted (it may be completed now). For example, the following structure was not accepted by the system at the moment of our test:




But there is no problem for SIRS-SS to simulate the corresponding IR/Raman spectra results are shown in Table 3.

A number of tests have been carried out with largely diverse molecules. Compared with experimental spectra, results are generally fairly good both for IR and Raman spectra. Generally agreement is good for the position of peaks.

In building the spectral databases we focus attention on the usual spectral range commonly used in IR by a bench chemist (say 600−4000 $cm^{-1}$). Indeed correlation charts usually do not contain much information (if any) about the spectral domain out of this region. Such bands are so lacking in our simulation. Would the database contain characteristic bands in such a region, SIRS-SS will have no problem to recall them. In Tables 1−3, only the experimental data of these bands have been given, although they are less useful for the bench chemist in many cases.

More discrepancies may occur for intensities, since correlation charts frequently specify a range of wavenumbers but do not indicate the corresponding intensity. It may also occur that some bands are quoted as "variable intensity".

**Table 3.** Simulation Results for the above Structure[a]

| experimental frequencies[6] | | | SIRS-SS | | | |
|---|---|---|---|---|---|---|
| freq | IR inten | Raman inten | freq (max.) | freq (min.) | IR inten | Raman inten |
| 1776 | VS | | 1825 | 1815 | S | |
| 1771 | | 32 | | | | |
| 1715 | S | 13 | 1755 | 1745 | VS | |
| 1597 | S | 48 | 1620 | 1585 | VS | M |
| 1450 | S | 2 | 1510 | 1470 | VS | W |
| | | | 1590 | 1565 | | M |
| 1280 | | 12 | | | | |
| 1213 | VS | 5 | 1330 | 1280 | X | |
| 1170 | S | | 1180 | 1150 | X | |
| 1157 | | 5.5 | 1150 | 1100 | X | |
| 1043 | VS | | 1080 | 1065 | M | W |
| | | | 1040 | 1016 | M | M |
| 999 | | 43 | 1010 | 990 | | VS |
| 789 | | 4.5 | | | | |
| 777 | S | | 805 | 720 | S | |
| 707 | S | 5.5 | 710 | 665 | S | |
| 616 | S | 11 | 630 | 605 | W | M |
| 166 | | 12 | | | | |

[a] CPU time: SIRS-SS: 0.99 s. Intensity abbreviations: VS, very strong; S, strong; M, medium; W, weak; VW, very weak; SH, shoulder; X, not indicated in the literature (according to refs 44−48 in our preceding paper[1]).

This also induces some difficulty in the graphical display of the simulated spectrum, joint to the fact that encoding intensity on five levels only is an handicap for a precise representation in view of a visual comparison with an experimental spectrum.

From Tables 1−3, it can be remarked that for some simulated peaks, although they are few, the difference is significant comparing with experimental data. This phenomenon is probably due to the inaccuracy of data in pilot substructure/subspectrum databases. Moreover, it is better to examine the performance of a system in a large scope. To ameliorate our substructure/subspectrum databases and to make a precise comparison of performances of different systems, a statistical study based on a huge spectral database is being carried out using different data mining techniques. The results will be reported in our next paper.

## CONCLUSION

Taking into account the priority order of searching defined, SIRS-SS system sequentially examine the four databases: small molecules; special fragments; atom-centered FRELs; and bond-centered FRELs. Efficiency of the search (avoiding redundancy and mismapping) is increased thanks to flags preventing at each step to consider structural fragments already treated upstream. Knowledge management is extended by defining a hierarchy between atoms in classes of equivalence and carrying out an ordered exploration performed from specific to generic features.

The system SIRS-SS can run on a PC (with Windows 95 or upper) with great speed. This user-friendly system, of easy use, can be widely applied in varied areas of chemistry.

## ACKNOWLEDGMENT

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.

## REFERENCES AND NOTES

(1) Yao, J. H.; Fan, B. T.; Doucet, J. P.; Panaye, A.; Yuan, S. G.; Li, J. F. SIRS-SS: system of simulation of IR/Raman spectra. I. substructure/subspectra correlation. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1046−1052.
(2) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure−Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci*. **1985**, *25*, 64−73.
(3) Willett, P.; Winterman, V.; Bawden, D. Implementation of Nearest-Neighbour Searching in an Online Chemical Structure Search System. *J. Chem. Inf. Comput. Sci*. **1986**, *26*, 36−41.
(4) Fan, B. T.; Panaye, A.; Doucet, J. P.; Barbu, A. Ring perception: a new algorithm for finding directly the SSSR from connection table. *J. Chem. Inf. Comput. Sci*., **1993**, *33(5)*, 657−662.
(5) Fan, B. T.; Barbu, A.; Panaye, A.; Doucet, J. P. Detection of constitutionally equivalent sites from a connection table. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 654−659.
(6) Fan, B. T.; Panaye, A.; Doucet, J. P. Comment on Isomorphism, Automorphism Partitioning and Canonical Labeling can be Solved in Polynomial Time for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 630−631.
(7) Ullmann, J. R. An Algorithm for Subgraph Isomorphism. *J. Assoc. Computing Machinery* **1976**, *23*, 31−42.
(8) Barnard, J. M. Substructure Searching Methods: Old and New. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 532−538.
(9) Rouvray, D. H. *Fuzzy Logic in Chemistry*; Academic Press: San Diego, CA, 1997.
(10) Lambert, J. B.; Shurvel, H. F.; Lightner, D.; Cooks, R. G. *Introduction to Organic Spectroscopy*; Macmillan: New York, 1987.
(11) YAO, J. H. Thesis, University of Paris 7, 2000.
(12) Schrader, B. *Raman/Infrared Atlas of Organic Compounds*, 2nd ed.; VCH: New York, 1989.
(13) Schuur, J.; Gasteiger, J. Infrared Spectra Simulation of substituted Benzene Derivatives on the Basis of 3D Structure Representation. *Anal. Chem*. **1997**, *69*, 2398−2405.
(14) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists -An Introduction;* VCH Press: Weinheim, 1993.
(15) http://www.gaussian.com.
(16) http://www2.ccc.uni-erlangen.de.

CI010061W