CORRESPONDENCE

SpecSolv: Artificially Intelligent or Artificially Innovative?

Michel Carabédian

Institut de Topologie et de Dynamique des Systèmes, Université Paris 7, associé au C.N.R.S. (URA 34), 1 Rue Guy de la Brosse, 75005, Paris, France

Received November 3, 1996

A recent article[1] reported the achievement of the spectroscopist's dream thanks to a new elucidation system "SpecSolv". This "artificially intelligent" system, based exclusively on ^{13}C-NMR, depends on fundamental concepts which are claimed to be "truly innovative". Unfortunately, it appears that the innovative character of the method proposed is the result of a very selective analysis of the literature which ignores some references likely to invalidate the thesis that this work is innovative. The similarity of the work described in this article to the concepts developed and published by our laboratory prompts us to write this note.

In the field of structure elucidation by ^{13}C-NMR, on which we have been working since 1978, the principle of the EPIOS system, whose name *E*lucidation by *P*rogressive *I*ntersection of *O*rdered *S*ubstructures expresses the originality of the strategy, was first published in 1980[2,3] (Figure 1). Its originality, compared to other systems of that time (DENDRAL, CASE, CHEMICS), was remarked in 1982 by Gray.[4] It represents an advantageous alternative to the strategy of these systems centered on isomer generators and therefore faced with serious combinatorial difficulties. With EPIOS three fundamental concepts which distinguish it from these precursors are introduced:

- The search is not enumerative, as with isomer generators, but becomes constructive by advancing with the progressive assembly of the partially overlapping $ELCO_b$ (*E*nvironment that is *L*imited, *C*oncentric and *O*rdered) primitives.
- The chemical shifts are not considered in isolation but in the form of pairs ($^{13}C_{Fo}$, $^{13}C_A i$) attributed to N carbon pairs (C_{Fo}, $C_A i$) making up the $ELCO_b$.
- The generation of candidate structures is synchronized with the attribution of the spectrum. A solution in EPIOS is a structure which has been completely validated by progressive attribution of the problem spectrum.

The redundancy of structural (atom pairs) and spectral (pairs of chemical shifts) information intrinsically linked to these concepts guides a search exclusively toward the valid solutions to a problem. The totality of the information deduced from the interpretation of a spectrum is organized in a generation graph. A solution is a path in this graph which is completely consistent with the data of the problem. Since the publication of the method, its practical application and its development have been described in the literature.[5]

In 1985 Bremser and Fachinger proposed an elucidation system, ACCESS,[6] which was inspired by the original principles of our system. Figure 2 compares its generation

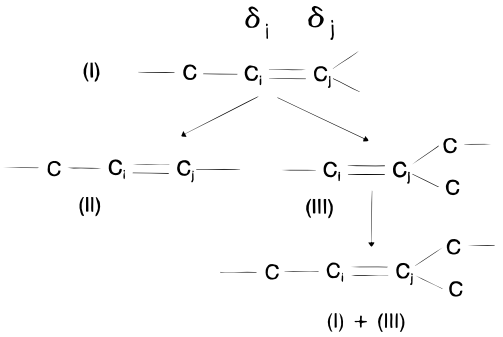

Figure 1. Principal of structural and spectral overlap in EPIOS.

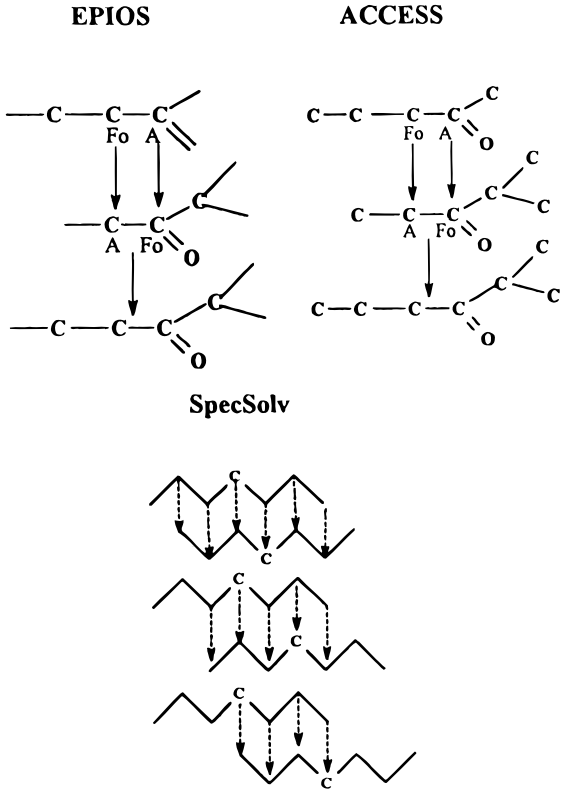

Figure 2. Comparison of structure generation in EPIOS, from $ELCO_b$, ACCESS, from two sphere substructures, and SpecSolv from three sphere substructures.

principle with that of EPIOS. Although the size of the primitives used is different, their assembly is based on partial and progressive overlap, as defined in EPIOS. However, this treatment only concerns the structural aspect of the problem. It does not use the redundancy of the chemical shifts assigned to the primitives, and the validity of the

S0095-2338(96)00033-9 CCC: $14.00

structures produced must be checked *a posteriori*. Despite the similarity of the concentric coding of the primitives, as in DARC[7] and the identity, at the structural level, of the principle of progressive assembly of the primitives with that of EPIOS, our work was not cited in the article published.[6]

The three principle "innovations" proposed in SpecSolv are

- −assembly of the substructures using an innovative approach that takes advantage of overlapping substructure information,
- −redundancy in structural description is the key to the new assembly process,
- −this instant validation of intermediate structures is one main difference from other generation programs, which typically validate structures only after generation of an entire molecule.

This approach based on the concept of overlap is termed "innovative" when this concept is in fact at the origin of the EPIOS method and even its name. The fundamental principles of the EPIOS system are not cited. Nonetheless, it appears from the presentation of SpecSolv that the assembly operation corresponds to a repetition of elementary overlaps as defined in EPIOS.

Although one of the authors of the present article (W. Fachinger) previously contributed to the ACCESS system, there is no explicit reference to this system, and SpecSolv is presented as original and without analogy to EPIOS or any relationship to ACCESS. It would, however, have been desirable to do a comparative analysis of this system, also presented as an innovation, with SpecSolv, in order to give readers an idea of the disadvantages of the predecessor and the real merits of the second.

REFERENCES AND NOTES

(1) Will, M.; Fachinger, W.; Richert, J. R. Fully Automated Structure Elucidation - A Spectroscopist's Dream Comes True. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 221−227.

(2) Dubois, J. E.; Carabédian, M.; Ancian, B. Automatic structural elucidation by carbon-13 NMR: DARC-EPIOS method. Description of progressive elucidation by ordered intersection of sub-structures *C. R. Acad. Sci. Paris, Ser. C* **1980**, *290*, 383−386.

(3) Dubois, J. E.; Carabédian, M.; Ancian, B. Elucidation structurale automatique par RMN du carbone-13: methode DARC-EPIOS. Recherche d'une relation discriminante structure-déplacement chimique. Automatic structural elucidation by carbon-13 NMR: DARC-EPIOS method. Search for a discriminant chemical structure-displacement relationship *C. R. Acad. Sci. Paris, Ser. C* **1980**, *290*, 369−372.

(4) Gray, N. A. B. Computer Assisted Analysis of Carbon-13 NMR Spectral Data. *Prog. NMR Spectrosc*. **1982**, *15*, 201−248.

(5) Carabédian, M.; Dagane, I.; Dubois, J. E. Elucidation by Progressive Intersection of Ordered Substructures from Carbon-13 Nuclear Magnetic Resonance. *Anal. Chem*. **1988**, *60*, 2186−2192.

(6) Bremser, W.; Fachinger, W. Multidimensional Spectroscopy. *Magn. Reson. Chem*. **1985**, *23*, 1056−1071.

(7) The Hose topological used to define the primitives is based on the same ordering function as the DARC Code (1966). The concept of Focus and Concentrically Ordered environment of DARC was gradually adopted in various structural databases and CAD systems.

CI960033R

SpecSolv−An Innovation at Work

Martin Will and Joachim Richert*

BASF Aktiengesellschaft, Main Laboratory, ZHV/S-B9, D-67056 Ludwigshafen, Germany

Received November 3, 1996

In response to our recent article,[1] Carabédian discusses above the innovative character of our structure elucidation system SpecSolv. This is a response to this discussion.

SpecSolv is a self-learning system based exclusively on ^{13}C-NMR chemical shift, intensity, and multiplicity information. Unlike earlier structure generation programs,[2,3] including the DARC/EPIOS[4] approach. SpecSolv requires neither the molecular formula of the unknown or additional input from other spectroscopic techniques. This is why we regard SpecSolv as a "true innovation" in the field of structure elucidation.

Automatic elucidation of chemical structures by SpecSolv entails four steps:

1. Acquisition of experimental ^{13}C- and DEPT-NMR spectra and extraction of chemical shift, intensity, and multiplicity information.
2. Subspectra search in a dedicated subspectra-substructure correlation (SSC) library and generation of a hit list.
3. Assembly of the retrieved substructures using a new approach that employs a flexible overlap (0−3 atomic spheres) of substructure information.
4. Rigorous validation steps for the intermediate substructures and the final result.

In their EPIOS system, Dubois et al.[5] also utilize overlapping substructure information, as had been done previously by others.[6] Overlapping of substructures during assembly is a fundamental procedure used by many algorithms, but the flexible overlap of substructures over 0−3 atomic spheres[7] is unique to SpecSolv. This feature even allows assembly and validation of substructures containing up to five adjacent heteroatoms as well as handling of systems in the database that lack a 1-sphere SSC (Figure 1 (parts e and f)). Figure 1 summarizes the conceivable types of overlaps used by SpecSolv.

Other systems which use overlapping information have been developed. These include GENOA,[6] CASE,[8,9] ACCESS,[10] and EPIOS.[4,5] Some of the reasons for their inability to elucidate more complex organic structures were (1) the small size of the substructures in their knowledge bases, which lead to combinatorial explosions with larger structures,

S0095-2338(97)00481-2 CCC: $14.00

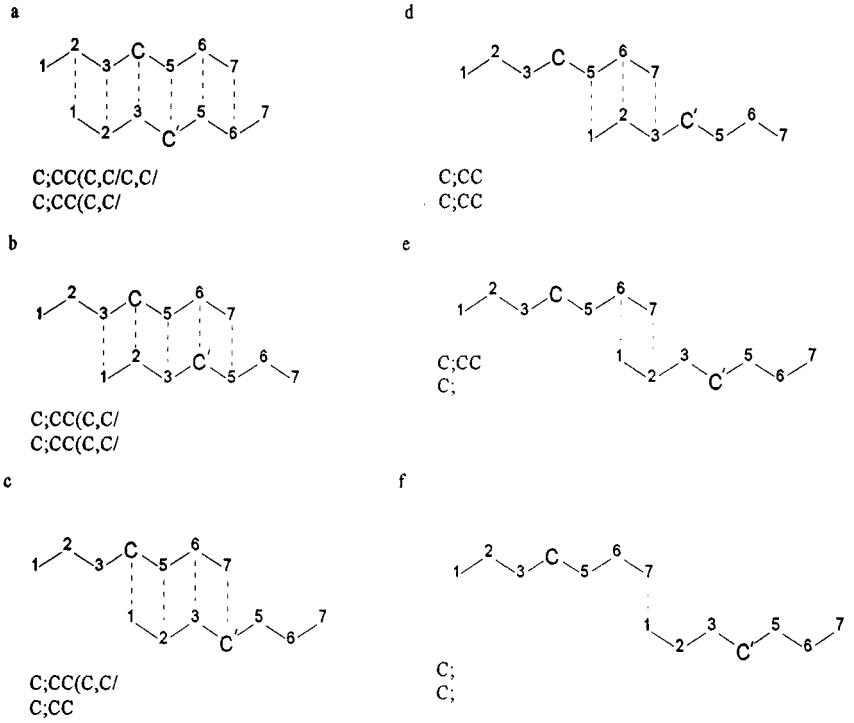

Figure 1. Classes of substructure overlaps in SpecSolv.

(2) the small overlap between substructures, typically only 1 or $1^1/_2$ spheres, as is used by EPIOS with its $ELCO_b$ coding,[4,5] and (3) the most critical deficiency, the lack of stringent verification tools for intermediate and final substructures. Likewise small structures and small overlaps lead to severe permutational problems, especially when one is confronted with the spectra of structures that contain several acyclic nitrogen atoms.[1] Further, most of these earlier systems neglected intensity information[4] which allows automated identification of symmetrical substructures or coincidentally identical chemical shifts.

In our original paper,[1] while referencing prior efforts[2,3] to automate the structure elucidation process including the EPIOS approach, we omitted explicit references to the work of Dubois et al., and we regret this omission. This nothwithstanding, we believe that SpecSolv is an innovative tool for the spectroscopist. It is the only system published to date which allows structure elucidation based only upon ^{13}C-NMR and which does not require knowledge of a molecular formula. It permits unsupervised operation and appears to be the fastest system currently available. It allows automated operation, accepting data at the spectrometer level, and, with no intervention, developing a final structure proposal which can be made available to the spectroscopist or a client of an analytical laboratory. SpecSolv contains an AI assembly algorithm as well as a self-learning database. With respect to the types of substructural overlap shown in Figure 1, it can be seen that the system is capable of generating news SSCs—those associated with the atoms between the two central atoms (C and C′) of the two overlapped structures. New SSCs can be extracted from the assigned experimental data and stored in the SpecSolv knowledge base, which is thus a self-learning database.

REFERENCES AND NOTES

(1) Will, M.; Fachinger, W.; Richert, J. R. Fully Automated Structure Elucidation - A Spectroscopist's Dream Comes True. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 221−227.
(2) Warr, W. A. Computer-Assisted Structure Elucidation. Part 1. Library Search and Spectral Data Collections. *Analyt. Chem*. **1993**, *65*, 1045A−1050A.
(3) Warr, W. A. Computer-Assisted Structure Elucidation. Part 2. Indirect Approaches and Established Systems. *Analyt. Chem*. **1993**, *65*, 1087A−1095A.
(4) Carabédian, M.; Dagane, I.; Dubois, J. E. Elucidation by Progressive Intersection of Ordered Substructures from Carbon-13 Nuclear Magnetic Resonance. *Analyt. Chem*. **1988**, *60*, 2186−2196.
(5) Dubois, J. E.; Carabédian, M.; Dagane, I. Computer-Aided Elucidation of Structures by Carbon-13 Nuclear Magnetic Resonance. *Analyt. Chim. Acta* **1984**, *158*, 217−233.
(6) Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. Genoa: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures. *J. Org. Chem*. **1981**, *46*, 1708−1718.
(7) In principle, SpecSolv's substructure description contains up to "$3^1/_2$" spheres, i.e., the number of hydrogens attached to atoms in the third sphere is known and used in the assembly and validation processes.
(8) Munk, M. E.; Shelley, C. A.; Woodruff, H. B.; Trulson, M. O. Computer-Assisted Structure Elucidation. Fresenius Z. *Analyt. Chem*. **1982**, *313*, 473−479.
(9) Lipkus, A.; Munk, M. E. Combinatorial Problems in Computer-Assisted Structural Interpretation of Carbon-13 NMR. *J. Chem. Inf. Comput. Sci*. **1985**, *25*, 38−45.
(10) Bremser, W.; Fachinger, W. Multidimensional Spectroscopy. *Magn. Res. Chem*. **1985**, *23*, 1056−1071.

CI970481E