Applying the Constraint Satisfaction Problem Paradigm to Structure Generation

Jean-Marc Nuzillard*,†

Laboratoire de Pharmacognosie, Faculté de Pharmacie, Unité Associée au CNRS No. 492, 51, Rue Cognacq-Jay, F-51096 Reims Cedex, France

Wady Naanaa‡ and Simone Pimont‡

Laboratoire d'Ingéniérie des Systèmes d'Information, Université Claude Bernard Lyon 1 - Bât. 710, 43, Boulevard du 11 Novenbre 1918, F-69622 Villeurbanne Cedex, France

Received July 26, 1995⊗

A new approach to structure generation is described in this paper. It substitutes the widely used fragment assembly technique and exploits prospectively and simultaneously any kind of analytical data. This was accomplished by the expression of a structure generation problem in terms of a constraint satisfaction problem (CSP). The CSP techniques such as constraint propagation, intelligent backtracking, and ordered searching have proven to be efficient for accelerating the generation process. The implementation of the underlying concepts relies on object-oriented formalism.

1. INTRODUCTION

Molecular structure generation is one of the tasks of analytical chemistry that can be automatized through the use of intelligent computer programs. The goal of such programs is the derivation of a set of possible structural formula for a chemical compound of unknown chemical structure on the basis of structural constraints. These constraints are inferred from different spectral experiments and are usually expressed by molecular fragments believed to be present (or absent) in the target molecule. The structure generator is an essential part of all computer assisted structure elucidation systems (CASE) and provides a combinatory process consisting in the assembly of the molecular fragments in all possible ways to ensure an exhaustive analysis. However, care must be taken in order to avoid redundancy caused by the eventual symmetry of the fragments.

The DENDRAL algorithm,[1] the first structure generator of the DENDRAL CASE system, was originally designed to generate acyclic structures only. The STRGEN generator extended the domain of tractable problems to cyclic compounds by linking the DENDRAL algorithm to catalogues containing both cyclic and acyclic molecular graphs. CONGEN[2] follows STRGEN. It introduced the concept of *superatom*, i.e., a fragment which is regarded as a large multivalent atom. As a result the construction process is accelerated by the use of building block units of different sizes. The latest DENDRAL generator to date is GENOA.[3] It is able to manipulate ambiguous constraints by the use of alternative substructures and allows the fragments to share parts (i.e., to overlap). The generator was subsequently extended to cope with 2D-INADEQUATE information.[4] COCOA, the structure generator of the SESAMI CASE system,[5] is quite different. It operates on the basis of uniformly sized, precisely defined atom centered fragments ACFs, predicted to be present in the unknown structure. These ACSs are selected from a predefined list of small fragments according to their compatibility with the corresponding NMR, IR data, and the molecular formula. Other substructural inferences, mainly from 2D NMR experiments, are retrospectively used to constrain the generation process. The CHEMICS method[6] is similar to that of COCOA. A set of predefined substructures, called components by the authors, are stored in a file. For a given problem the relevant components are selected by reference to ^{13}C, ^{1}H NMR and IR spectra. The assembly process is based on specially designed logic and connectivity stack. The derived structures are filtered by applying human knowledge and ^{13}C NMR data prediction. The CHEMICS authors have introduced a program that applies 2D-INADEQUATE data for additional filtering of the already assembled structure.[7] Recently a novel approach which emphasizes the use of 2D NMR spectra was described.[8] The method is based on a connectivity matrix containing one entry for each atomic free valence. This matrix is filled in a predetermined order according to 2D NMR data transformed into one- or two-bond carbon−carbon connectivities.

In the EPIOS system,[9] structure generation is carried on in parallel at the spectral and structural levels. Hence, a more intensive exploitation of the spectral information is obtained. But even so the EPIOS algorithm would lead to a combinatorial explosion when the number of subspectra to be considered is high.

Bangov[10] has introduced a method that shifts the attention of bonding sites instead of single atoms so that the structural fragments are to be treated as *superatoms*. As in the case for CONGEN *superatoms* are not allowed to share domains in Bangov's program. This constraint severely diminishes the advantages gained by the use of fragments. Recent work by the same author was dedicated to avoiding redundant structures.[11]

The first version of the LSD program[13] also gives priority to data derived from various NMR experiments (i.e., 1D ^{1}H NMR, 1D ^{13}C NMR, and 2D NMR) and to atom properties such as the type, the hybridization state, and the number of attached hydrogens. IR and UV spectroscopy may be used to derive functionalized fragments which are retrospectively used to filter the assembled structures. The generator is fairly

† E-mail address: jm.nuzillard@univ-reims.fr.
‡ E-mail address: naanaa,pimont@lisisun.univ-lyon1.fr.
⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1995.

0095-2338/95/1635-1068$09.00/0

inefficient when it is confronted with a lack of data of NMR origin which can arise from low sensitivity or a weak resolution of the spectra. In other words the efficiency of the program is highly dependent on the amount and origin of the available data. In order to avoid these drawbacks, we have developed a program that prospectively and simultaneously treats any kind of constraining information, whatever the source. This approach has proved to be more efficient than a multiphase program that considers separately each type of structural constraint.

Two sources of data are usually exploited by a structure generator: first, the molecular formula (MF), often unequivocally available from high resolution mass spectroscopy and, second, the substructures which may most generally be defined as mono- or polyatomic fragments of a molecule which can be fully recognized by molecular spectroscopic methods. The intricacy in assembling substructures resides in the fact that we are not able to make explicit correspondence between the atoms involved in substructures and those of the MF. Atoms contained in substructures are often generic, i.e., we might know their type or their hybridization state, but we cannot precisely identify them before elucidating the structure.

2. DEFINITIONS

A constraint satisfaction problem CSP[14] can be defined as follows: "Given a set of n variables each with an associated value domain and a set of constraining relations each involving a subset of the variables, find all possible n-tuples such that each n-tuple is an instantiation of the n variables satisfying the relations."

Backtracking is the classic algorithm used for solving CSPs. It is a search procedure that assigns a value to a variable during each stage. When it fails to find a value for a variable consistent with all the previous assignments, backtracking takes place, and another value is assigned to the immediately preceding variable. The search continues in a similar way until all the possible assignments have been tried.

Whenever a value is assigned to a variable, restrictions of value domains for future variables are computed and recorded. This process has been referred to as *constraint propagation*.

Labeled graphs are used to represent molecular structures. A graph can be viewed as a collection of elements (vertices) and binary relations (edges) between these elements. In molecular graphs, labeled vertices (e.g., C, O, N, F, ...) correspond to the individual atoms, and labeled edges (e.g., single, double, or triple) correspond to the bonds between them.

3. METHOD

Piecing together substructures in a number of different ways is a time-consuming combinatory process. The complexity of this task could be significantly reduced if explicit mapping between the atoms contained in fragments and those of the MF was somehow available. Unfortunately this is not the case, because spectroscopic techniques are usually unable to reveal the relative positions of fragments in the molecular structure. The problem of assembling substructures could then be stated as follows: find all the possible mappings between the atoms of the available substructures, on the one hand, and the atoms of the MF, on the other hand. This is the key idea of our program.

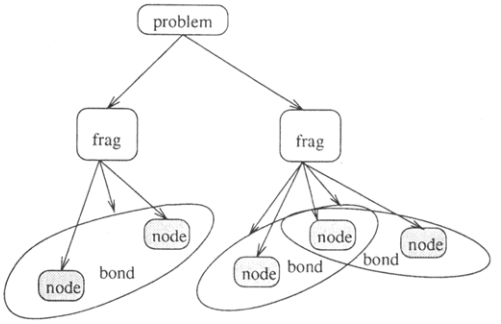

Figure 1. LSD problem structure.

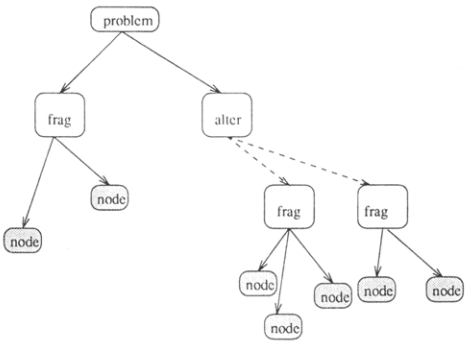

Figure 2. A general LSD problem structure.

We have expressed the structure generation problem in terms of a CSP. The atoms involved in the substructures will be referred to by *nodes* and will correspond to the problem variables. The atoms of the MF will be regarded as the values that will be assigned to variables. Obviously chemical bonds contained in the substructures will correspond to the constraining relations. An LSD problem is defined by a set of variables (*nodes*) and a set of values (*atoms*). The *nodes* are usually encapsulated in fragments with the related *bonds*. The domains of the variables are subsets of values. A primitive LSD problem structure is depicted in Figure 1. The goal of the generator is to find all the consistent instantiations of the *node* variables. It proceeds in stages. A single stage consists of the instantiation of one variable. The assignment of *atom* values to all *node* variables in all consistent ways leads indirectly to an exhaustive assemblage of the fragments. The candidate molecular graphs are constructed edge by edge; an edge is created each time a pair of *node* variables linked by a *bond* relation is instantiated by a pair of *atom* values.

The LSD program provides also a mechanism for manipulating ambiguous data. Because the interpretation of spectra is by nature an ambiguous task, i.e., for a given spectral signature one or more substructures may be plausible. A particular variable named *alternatives* is used. Its domain is a set of fragments. The presence of only one fragment is required at a time. The generator will assign a *fragment* value to an *alternatives* variable exactly as it does for a *node* variable. The structure of a more general LSD problem is shown in Figure 2. For a sake of clarity the *bond* relations are omitted.

3.1. Complementary Constraints. The LSD program exploits prospectively many other types of constraints: Unless it is explicitly specified by the chemist, it is possible that more than one *node* variable is assigned the same *atom*

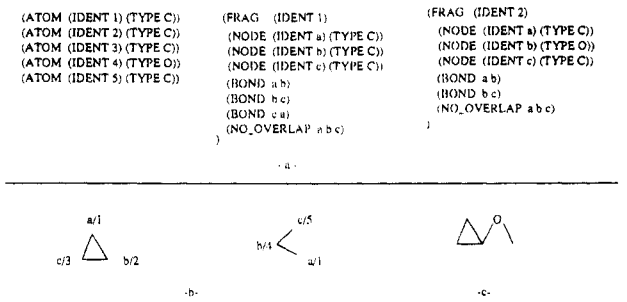

Figure 3. (a) An LSD problem, (b) a consistant instantiation, and (c) the resulting molecular graph.

value. This fact guarantees that all the possibilities of overlap among fragments are tried. When the chemist is sure that a given subset of *node* variables does not share values, he has to indicate that constraint to the generator by means of the *no_overlap* relation. A sentence (NO_OVERLAP i_0, i_1, i_2, ..., i_n) when introduced to the system ensures that the referenced *node* variables will be assigned distinct *atom* values.

The *saturation* relation is a second example of widely used constraining relations. Its role is to limit the number of direct neighbors to atoms. We attach an *sp2* relation to all the atoms that are known to be in the sp^2 hybridization state. An *sp2* relation causes any partial solution to be abandoned if it contains an sp^2 saturated atom (i.e., with no free valence) exclusively linked to sp^3 atoms. This is because the generator is not able to create a double bond emanating from that atom. The connectivity of the final molecular graph is also ensured. This can be easily verified by a linear time algorithm when the assembling of structures is completed. However in some cases, during the generation process, partial structures cannot be extended to connected complete structures. This deduction can be made in two situations: first, when the partial structure contains a saturated component, i.e., a fragment containing exclusively saturated atoms, but with less atoms than the MF and, second, when the number of rings contained in a partial structure exceeds the cyclomatic number of the target molecular graphs. This number is given by the well-known graph theoretical equality $\gamma = m - n + 1$ where m is the number of edges (bonds) and n is the number of vertices (atoms). This number cannot be precisely known because the multiplicity and the hybridization state of the atoms may represent some ambiguities that affect the computation of the number of chemical bonds m (see eq 1). However an upper limit on γ could be imposed. Detecting a future disconnected structure at an early stage saves time spent in running the connectivity check algorithm for each complete structure plus the time spent in completing the disconnected partial structures themselves. The connectivity checking is run by the *connex* relation every time an edge is added to a partial molecular graph.

$$m = 1/2 \sum_{i \in \text{atoms}} \text{valence}_i + \text{hybridization}_i - \text{multiplicity}_i - 3 \quad (1)$$

Figure 3a shows a fictitious LSD problem. A consistent instantiation is shown in Figure 3b, and the resulting molecular graph is depicted in Figure 3c.

3.2. Exploiting 2D NMR Data. With the advent of 2D NMR techniques new types of constraining information, imposing topological distance constraints among carbon

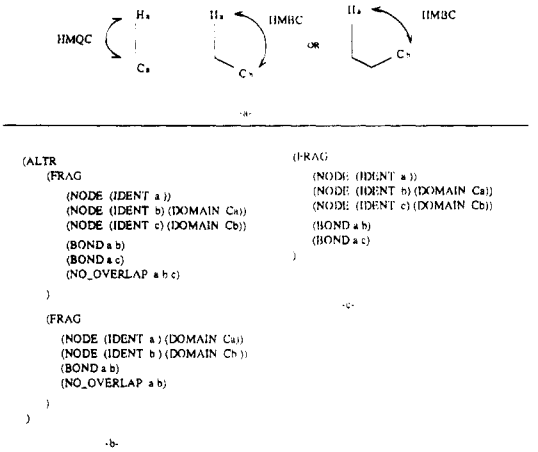

Figure 4. (a) HMQC and HMBC correlations, (b) the associated LSD codification, and (c) an optimized codification.

atoms, are now available. This source of information has not been efficiently exploited by the few programs in which it is incorporated. The 2D NMR experiments are powerful enough to reveal carbon−carbon connectivities. The 2D-INADEQUATE experiment[15] or the $^1H-^1H$ COSY[16] combined with a $^{13}C-^1H$ HMQC[17] correlation reveal the existence of direct bonds between carbon atoms. The combination of the HMBC[18] and the HMQC experiments yields long range carbon−carbon connectivities. In the usual case the number of intervening bonds between the two correlating nuclei is two or three (see Figure 4a), but it is in some cases four or even five! This topological distance constraint is translated to a partially identified chain fragment (i.e., two *node* variables are already mapped to the correlating atoms) whose length is chosen according to the maximum number of allowable intervening bonds and regarded by the generator as a substructural fragment. A straightforward codification of the constraint inferred from the correlations described above is shown in Figure 4b. A more efficient codification of the same constraint that saves on *alternatives* variable is depicted in Figure 4c. *Node* a is allowed to share the same value with *nodes* b and c. Hence the chain fragment may be of length one or two. Note that the generator ignores loop formation.

To illustrate the use of NMR data, we show here how to solve the structure of gibberellic acid, on organic compound of natural origin with molecular formula $C_{19}H_{22}O_6$. Direct and long range atom connectivities are inferred from COSY, HMQC, and HMBC correlations (see Table 1). Direct connectivities 1−20 and 1−22 are inferred from the chemical shifts of carbons 1 and 2. These constraints are expressed by chain fragments with variable length as explained above. The problem allows two solutions consisting of two partial structures which are subsequently completed so that each atom saturates all its free valences. The correct structure is shown in Figure 5d. Usually the saturation procedure is time consuming, especially when we do not have enough constraints. Many bonds are left free, and their saturation leads to a considerable number of molecular graphs which are frequently highly strained. Although a quick procedure[19] for eliminating strained anti-Bredt structures can be used, unlikely structures without any misplaced double bond, are withheld.

The actual routine used for completing partial structures is an ordered depth-first search that appeals to the hierarchi-

Table 1. Observed HMQC (O) and HMBC (×) Correlations for the *Gibberellic Acid* Problem

atom no.	4	5	6	9	11	12	13	15	16	17	19	22
1					×						×	
2					×	×						
3			×						×	×		
4	O											
5		O										
6			O						×			
7	×	×										
8			×						×	×		
9	×	×		O	×						×	
10	×				×	×					×	
11		×			O	×					×	
12					×	O			×			
13							O					
14						×	×		×			
15						×		O	×			
16			×			×	×		O			
17							×			O		
19											O	
22	×											O

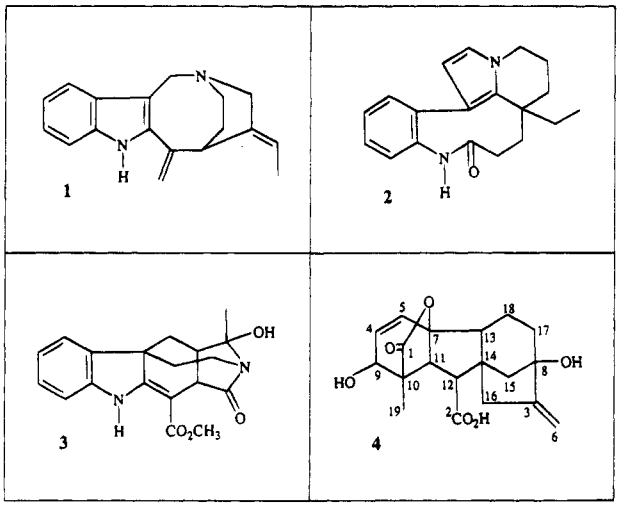

Figure 5. The correct structures for compounds **1**−**4**.

cally ordered extended connectivities procedure *HOC*[20] for finding graph orbits. The search is ordered to avoid redundant structures. Redundancy is caused by the well known graph isomorphism problem. More details about this procedure will be reported in a future paper.

3.3. Variable and Value Ordering Heuristics. Section 3 explains that the main task of the structure generator is to assign values to variables. Usually more that one variable is a candidate for instantiation at any time. The generator must take a decision about which variable is to be instantiated first. The solution to this problem is of great interest and has been widely studied in the domain of CSP, because the variable instantiation order can have far-reaching implications on the size of the problem search space. Several heuristics have been proposed. In ref 21 the order is established prior to exploring the search space. The variable participating in the highest number of constraints is selected first. The *Freuder* ordering[22] is also computed before enumerating the problem solutions. The criterion used is relative to the global problem structure. In refs 23 and 24 the order is dynamically determined. The first algorithm favors the variable with the smallest domain, while the second chooses randomly the first variable and then iterates by choosing as the next variable the one that is linked to the highest number of variables among those that are already instantiated. For the structure generation problem, an ordering heuristic based on a single criterion is insufficient. Because of the molecular graph which is being constructed, several other criteria might be taken into account. Our technique is based on the calculation of a score for each variable. This score is given by the function: $score = w_0c_0 + w_1c_1 + ... + w_pc_p$, and the variable totalizing the maximal score is instantiated first. The c_is are relevant numeric criteria attached to variables. The domain size of the variables, the number of attached constraints, and the number of free valences of the atoms instantiating the neighboring variables are some criteria worth mentioning. The last quoted criterion is very relevant because instantiating a *node* variable surrounded by almost saturated atoms will cause these atoms to reach their saturation state, and they will not be allowed to participate in the formation of any new bond. By means of *constraint propagation* operations they will be discarded from the domain of a lot of *node* variables.

The w_is are weights computed by a tuning operation similar to the *Perceptron* learning algorithm.[25] Initially the weights are intuitively set to values ranging from −1 to 1. A first series of runs is performed on a set of relatively small problems. The average size of the search spaces is recorded, and the weights are randomly modified. Then, other series of runs are undertaken, but the weights are corrected according to the variation of the average size of the search spaces. The procedure stops when the average size of the search spaces reaches a minimum. When harder problems are encountered, the program behaves much better with learned weights, because an appropriate order enhances the pruning capability of *constraint propagation* operations.

The ordering of the values in each variable domain is also exploited in our program. It increases the likelihood that the most probable solution will be found early on. Values ordering is based on an approximate distance matrix computed statically before the enumeration. The approximate distance matrix is inferred from 2D NMR data as follows. First, a correlation matrix C is initialized:

$$c_{ij} = \begin{cases} 1 & \text{if a direct connectivity between atoms } i \text{ and } j \text{ is inferred} \\ 2 & \text{if a long-range connectivity between atoms } i \text{ and } j \text{ is inferred} \\ \infty & \text{otherwise} \end{cases}$$

The approximate distance matrix is the *transitive closure* of matrix C. The generator favors the creation of bonds between neighbor atoms (according to the approximate distance matrix) and a rapid generation of the more plausible solutions is frequently guaranteed. Applying this heuristic is justified by the fact that the approximate distance matrix will not drastically separate atoms that are neighbors in reality.

Ordering values is very helpful when manipulating loosely constrained problems. In this case the time necessary to generate all the solutions is very long, and the program is interrupted when the number of generated solutions exceeds a manageable value. Although the program would have terminated abnormally, the most plausible structure could be looked for among the surviving structures, since by ordering values they will be the closest to the correct one.

Table 2. Natural Compounds Treated by the LSD Program

no.	molecular formula	nb. const	nb. var
1	$C_{18}H_{20}N_2$	20	29
2	$C_{19}H_{22}N_2O$	22	42
3	$C_{19}H_{20}N_2O_4$	25	20
4	$C_{19}H_{22}O_6$	25	33
5	$C_{18}H_{18}O_6$	24	28
6	$C_{20}H_{22}N_2O$	23	32
7	$C_{20}H_{20}O_6$	26	40
8	$C_{22}H_{22}O_8$	30	46

Table 3. Ordered Search Performance (in Seconds)

	total run time			correct solution generation time	
no.	VSDFH	MCH	MCH and value ord	MCH	MCH and value ord
1	6	3	3	2	2
2	10	6	4	5	3
3	31	7	5	6	4
4	130	20	6	8	4
5	85	35	9	20	7
6	156	39	11	27	8
7	125	31	12	30	7
8	155	45	15	32	9

4. RESULTS

The work of our generator is exemplified here with some results. Table 3 shows the improvement obtained in problem-solving performance for the compounds in Table 2.[26-32] Two ordering heuristics are compared. The first is the multicriteria heuristic (MCH) adopted by our program, and the second is *the variable with the smallest domain first* heuristic (VSDFH). When the MC heuristic is used, the benefit outweights the cpu time consumed in computing the numeric criteria and the weighting function even for the smallest problem. The fourth column of Table 3 shows the performance of a version of the program that adopts both variable and value ordering. An improvement in the total execution time is observed because additional constraints are dynamically inserted in the CSP. The fact that the most plausible value is used first for the instantiation of a given variable increases the pruning power of the newly added constraints. Moreover value ordering noticeably speeds up the generation of the correct structures. Figures 5 and 6 show the correct structures of the compounds from Table 2.

5. IMPLEMENTATION

We have used the object oriented formalism which provides a high degree of flexibility. The LSD program was written in Common Lisp version 4.0.1 under UNIX; it is implemented on a SUN-4 sparc station with 16 Mo of memory. 1D and 2D NMR spectra were recorded on a BRUKER AC300 or AMX500 spectrometer using standard microprograms.

6. CONCLUSION

The concepts of the CSP paradigm perfectly match those encountered in structure generation problems. Our program integrates the exploitation of all types of data and treats in an elegant way the unavoidable ambiguities of the domain. Moreoever, we have chosen an object oriented internal representation of the data and procedures, that has enhanced the flexibility of the program. We plan to extend the generator by a knowledge based module, in order to avoid the generation of rare and highly strained molecular structures.

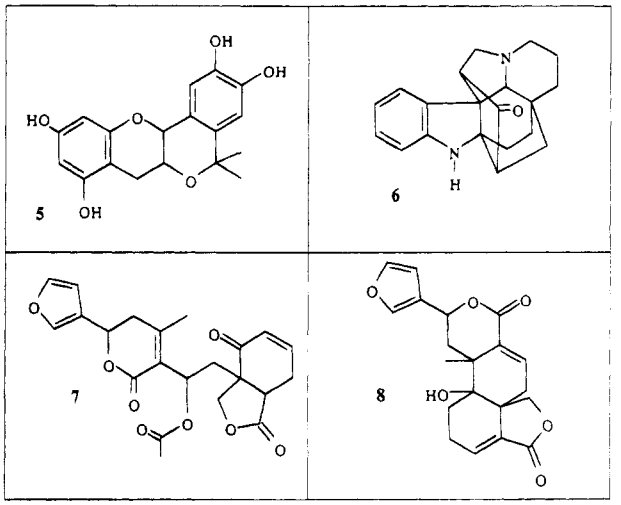

Figure 6. The correct structures for compounds **5**–**8**.

ACKNOWLEDGMENT

We thank J. D. Connolly, B. Richard, and J. Vercauteren for providing compounds. One of us (W.N.) thanks the Région Champagne-Ardenne (France) for financial support.

REFERENCES AND NOTES

(1) Gray, N. A. B. Dendral and Meta-Dendral: The Myth and the Reality. *Chem. Intell. Lab. Sys.* **1988**, *5*, 11–32.
(2) Masinter, L. M.; Sridharan, N. S.; Lederberg, J.; Smith, D. H. Applications of Artificial Intelligence for Chemical inference. 12. Exhaustive Generation of Cyclic and Acyclic Isomers. *J. Am. Chem. Soc.* **1974**, *96*, 7702–7714.
(3) 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.
(4) Lindley, M. R.; Shoolery, J. N.; Smith, D. H.; Djerassi, C. Application of Artificial Intelligence for Chemistry. 43. Application of the Program GENOA and 2D NMR Spectroscopy to Structure Elucidation. *Org. Magn. Reson.* **1983**, *21*, 405–411.
(5) Christie, B. D.; Munk, M. E. The Role of Two-Dimensional Nuclear Magnetic Resonance Spectroscopy in Computer-Enhanced Structure Elucidation. *J. Am. Chem. Soc.* **1991**, *113*, 3750–3757.
(6) Sasaki, S.; Abe, H. Computer-assisted structure Elucidation System—Functions That "Chemics" System Can Perform. *Computer Application in Chemistry*; Elsevier Science Publisher B. V.: Amsterdam, 1983.
(7) Funatsu, K.; Susuta, Y.; Sasaki, S. Introduction of Two-Dimensional NMR Spectral Information to an Automated Structure Elucidation System. CHEMICS. Utilisation of 2D-INADEQUATE Information. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 6–11.
(8) Peng, C.; Zheng, C.; Hui, Y. Efficient Application of 2D NMR Correlation Information in Computer-Assisted Structure Elucidation of Complex Natural Products. *J. Chem. Inf. Comput. Sci.* **1994**, *23*, 805–813.
(9) Dubois, J. E.; Carabedian, M.; Dagane, I. Computer-Aided Elucidation of Structures by CARBON-13 Nuclear Magnetic Resonance. *Anal. Chim. Acta* **1984**, *158*, 217–233.
(10) Bangov, I. P.; Kanev, K. D. Computer assisted Structure Generation from a Gross Formula. 2. Multiple Bond Unsaturated and Cyclic Compounds. Employment of the Fragments. *J. Math. Chem.* **1988**, *2*, 31–48.
(11) Bangov, I. P. Structure Generation from a Gross Formula. 7. Graph Isomorphism: A Consequence of the Vertex Equivalence. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 318–324.
(12) Bangov, I. P.; Simova, S.; Cabrol-Bass, D.; Laude, I. Computer-Assisted Structure Generation from a Gross Formula. 6. Reducing the Structural Redundancy by Employment of 2D NMR Spectral Information. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 546–557.
(13) Nuzillard, J. M.; Massiot, G. Logic for Structure Determination. *Tetrahedron* **1990**, *47*, 3655–3664.
(14) Montanari, H. Network of Constraints: Fundamental Properties and Application to Picture processing. *Inf. Sci.* **1974**, *7*, 95–132.

(15) Bax, A.; Freeman, R.; Frenkiel, T. A. An NMR technique for Tracing Out the Carbon Skeleton of an organic molecule. *J. Am. Chem. Soc.* **1981**, *103*, 2102−2104.
(16) Aue, W. P.; Bartholdi, E.; Ernst, R. R. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **1976**, *64*, 2229−2246.
(17) Bax, A.; Subramanian, S. J. Sensitivity-Enhanced Two-Dimensional Heteronuclear Shift Correlation NMR Spectroscopy. *J. Magn. Reson.* **1986**, *67*, 565−569.
(18) Bax, A.; Summers, M. F. J. ^{1}H and ^{13}C Assignments from Sensitivity-Enhanced Detection of Heteronuclear Multiple-Bond Connectivity by 2D Multiple Quantum NMR. *J. Am. Chem. Soc.* **1986**, *108*, 2093−2094.
(19) Nuzillard, J. M. Quick Method for Anti-Bredt structure detection. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 723−724.
(20) Baladan, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC procedures). I. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, *6*, 538−551.
(21) Dechter, R.; Meiri, I. Experimental Evaluation of Preprocessing technique in Constraint Satisfaction Problems. In *Proc. IJCAI*; Detroit, MI, 1989.
(22) Freuder, E. C. A Sufficient Condition of Back-free Search. *J. ACM* **1982**, *29*, 24−32.
(23) Nudel, B. Solving the General consistant labeling problem: Two Algorithms and their Expected complexities. In *Proc. AAAI*; Washington D.C., 1983.
(24) Zabih, R. Some Application of Graph Boundwidth to Constraint Satisfaction Problem. In *Proc. AAAI*; 1990.
(25) Minsky, M.; Papert, S. *Perceptrons*: an Introduction to computational Geometry; MIT Press: MA, 1969.
(26) Joule, J. A.; Ohashi, M.; Gilbert, B. Alkaloids studies-LIII.: The Structure of Nine Alkaloid from *Aspidosperma dasycarpon* A. Dc. *Tetrahedron* **1965**, *21*, 1717−1734.
(27) De Silva, K. T.; Ratcliffe, A. H.; Smith, G. F.; Smith, G. N. Rhazilam, a Neutral Alkaloidal Artefact from *Rhazia stricta* D. *Tetrahedron Lett.* **1972**, 913−916.
(28) Nkiliza, J. Total Synthesis of 19-hydroxytubotaiwine and Preparation of *Iboga*-like Indoloazepines. Ph. D. Thesis, Université de Bordeaux II-France, 1990.
(29) Artsuck, J. A.; Lipscomb, W. Molecular and Crystal Structure of the Di-*p*-bromobenzoate of Methyl Ester of Gibberellic Acid. *J. Am. Chem. Soc.* **1963**, *85*, 3414−3419.
(30) Ngnokam, D.; Massiot, G.; Nuzillard, J. M.; Tsamo, E. (+)-7,7′-dimethyl-5-hydroxy-2*R*,3*S*-*trans*-pubeschin from *Entandophragma cylindricum*. *Phytochemistry* **1994**, *37*, 529−531.
(31) Ferreira, J. M.; Gilbert, B.; Kitagawa, M.; Leme, L. A.: Durham, L. J. Four Heptacyclic Alkaloids from *Aspidosperma* Species. *J. Chem. Soc. C* **1966**, 1260−1266.
(32) Almanza, G.; Baldemara, L.; Labbe, C.; Lavaud, C.; Massiot, G.; Nuzillard, J. M.; Connolly, J. D.; Rycroft, D. S. Clerodane Diterpenoids and Ursane Triterpenoids from *Salvia Haenkei*. *Phytochemistry* Submitted to publication.

CI950220+