Maximum Symmetrical Split of Molecular Graphs. Application to Organic Synthesis Design

Philippe Vismara,*,†,§ Yannic Tognetti,† and Claude Laurenço*,†,‡

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), UMR 5506
CNRS/Université Montpellier II, 161 rue Ada, 34392 Montpellier Cedex 5, France, Laboratoire des Systèmes
d'Information Chimique (LSIC), UMR 5076 CNRS/Ecole Nationale Supérieure de Chimie de Montpellier,
8 rue de l'Ecole Normale, 34296 Montpellier Cedex 5, France, and Ecole Nationale Supérieure d'Agronomie
de Montpellier (ENSA-M), place Pierre Viala, 34060 Montpellier Cedex 1, France

Received September 23, 2004

Whereas the potential symmetry of a molecule may be a feature of importance in synthesis design, this one is often difficult to detect visually in the structural formula. In the present article, we describe an efficient algorithm for the perception of this molecular property. We have addressed this problem in terms of graph theory and defined it as the *Maximum Symmetrical Split* of a molecular graph. A solution is obtained by deleting in such a graph a minimum number of edges and vertices so that the resulting subgraph consists of exactly two isomorphic connected components that correspond to a pair of synthetically equivalent synthons. In view to reduce the search space of the problem, we have based our algorithm on CSP techniques. In this study, we have found that the maximum symmetrical split is an original kind of Constraint Satisfaction Problem. The algorithm has been implemented into the RESYN_Assistant system, and its performance has been tested on a set of varied molecules which were the targets of previously published synthetic studies. The results show that potential symmetry is perceived quickly and efficiently by the program. The graphical display of this perception information may help a chemist to design reflexive or highly convergent syntheses.

INTRODUCTION

Short solutions are best—this common sense principle has found obvious applications in organic synthesis design. Ideally, a synthesis should give a target molecule from readily available starting materials in one step and in quantitative yield.[1] However, such an ideal synthesis is unlikely to be practical and remains a formidable challenge in the case of complex molecules. As most syntheses are in fact multistep processes, the practitioners want to minimize the number of steps so as to approach the ideal synthesis, and any progress in synthesis strategy occurs when they are able to devise new plans shorter or simpler than the previous ones.[2]

Velluz et al.[3] have stated that two types of multistep synthesis, linear and convergent, must be considered from the economical viewpoint. In the former, a target molecule is assembled by stepwise adding small building blocks in a linear reaction sequence. In the latter, the two or more direct precursors of a target are built independently, and they react together in the last step of the synthesis to give the desired molecule. The reaction sequence of a convergent synthesis is branched, and its longest path from starting materials to the target molecule is necessarily shorter than the one of the reaction sequence of a linear synthesis of the same number of steps (Figure 1). This has an impact on the cost of synthesis in terms of yield, time, and manpower.

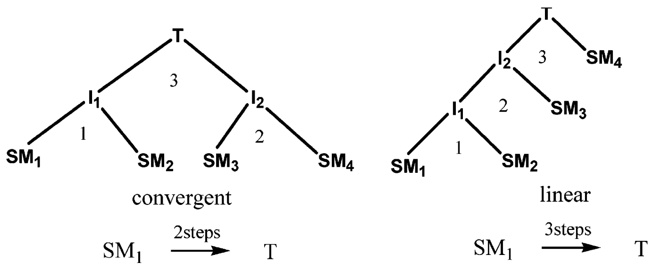


Figure 1. Convergent vs linear synthetic plans.

Convergency was formalized by Hendrickson,[4] who has introduced a graph-theoretical index to evaluate and compare linear, convergent, and intermediate synthetic plans. The simplest way to develop a convergent plan is to cleave retrosynthetically bonds in the middle of the target structure. Such disconnections divide the structure into two fragments—synthons[5]—which may be similar in size, composition, constitution, or structure. An optimal result occurs when the target structure is symmetric. In this case both fragments are identical and do not require separate preparations. A convergent synthesis which takes advantage of this symmetry by reducing the work at bench has been called a *reflexive synthesis* by Bertz.[6] For example, the C_2-symmetric structure of β-carotene may be disconnected at its central double bond, according to disconnection **a** in Figure 2. This disconnection suggests a coupling strategy ($C_{20} + C_{20}$) involving two identical fragments which correspond to vitamin A_1 derivatives. The route reported by McMurry and Fleming[7] is an illustration of the effectiveness of this strategy. It is interesting to note that the symmetric structure of β-carotene allows other strategies to be devised, in particular ones related to

* Corresponding author phone: +33-499-612650; fax: +33-467-521427; e-mail: vismara@lirmm.fr (P.V.); phone: +33-467-144333; fax: +33-467-144353; e-mail: laurenco@enscm.fr (C.L.).
† LIRMM.
‡ LSIC.
§ ENSA-M.

10.1021/ci049713h CCC: $30.25
Published on Web 04/19/2005

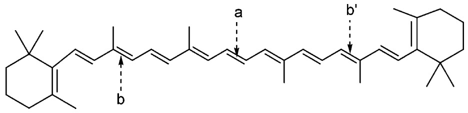


Figure 2. Two strategies based on symmetry for β-carotene synthesis.

simultaneous disconnection of two symmetric bonds of the polyenic chain. Many combinations of three fragments have been studied. As an example, a triply convergent synthesis based on a (C_{14} + C_{12} + C_{14}) strategy, according to disconnections **b** and **b′** in Figure 2, has recently been reported by Vaz et al.[8]

In his seminal article, "General Methods for the Construction of Complex Molecules", Corey[9] has pointed out that the analysis of potential as well as actual molecular symmetry can be of importance in solving a synthetic problem. As defined by Corey, a molecule may be said to possess *potential symmetry* when it can be disconnected to give a symmetrical structure or two or more synthetically equivalent structures. Syntheses which take advantage of potential symmetry in this way are clearly reflexive.[6] A classical example is given by usnic acid, which is unsymmetrical but which can be prepared by an *o,p*-coupling of two phenoxy radicals, both deriving from a same phenol (Figure 3) as in Barton's synthesis.[10]

Potential symmetry is frequently present in natural products but is not so easily perceived. For example, designing a reflexive synthesis of carpanone like Chapman's[11] is not obvious (Figure 4).

An attempt was made in 1986 to incorporate a strategy based on potential symmetry into the LHASA retrosynthetic analysis program.[12,13] The new strategy module was derived from a preceding one developed to incorporate a starting-material oriented strategy.[14] That new module was able to propose changes, in terms of retrosynthetic goals such as carbon−carbon bond disconnections and functional group transformations, which were needed to symmetrize the molecular graph of a target structure. Nevertheless, the program was confronted with a combinatorial problem in the case of complex targets, erythronolide A for example (cf. Figure 20), and then generated a large number of results without finding any acceptable solutions. The main reason for this pitfall was that no algorithms were devised specifically to solve the potential symmetry problem. Instead, the implemented algorithms were adapted from those developed for mapping the starting material and target structures,[15] which is a related but different problem. Later, a genetic algorithm was used to solve the *Maximum Common Subgraph* (MCS) problem, and a few examples were shown of its application in potential symmetry perception.[16] Such a nondeterministic approach may give good results but cannot guarantee that the optimum solution will be obtained.[17] This has led us to reexamine this problem and to solve it by following a new deterministic way based on the CSP (Constraint Satisfaction Problem). The algorithm we devised has been implemented into our RESYN_Assistant system, a program for computer-aided understanding of organic synthesis problems.[18] For a given molecule, as the potential symmetry of substances has its origin mainly in dimerizations of others, this algorithm allows the system to display the bond disconnections and atom deletions that would give pairs of the largest equivalent synthons. Several maximum solutions may exist, each resulting from splitting the molecular graph of the target into two isomorphic parts.

If we consider a molecular graph $G = (V, E)$, labeled on the vertices and edges with atom and bond types respectively, this *Maximum Symmetrical Split* problem can be defined as follows:[19]

Problem 1. Delete in G a minimum number of edges and vertices so that the resulting subgraph consists of exactly two isomorphic connected components.

A major point of this definition is that the two components resulting from the split must be connected since they represent synthons. Although maximal unconnected splits larger than a maximum connected one can be found, there is no relationship between these two kinds of splits, and the former are generally not significant from the viewpoint of retrosynthesis. Another point is that the isomorphism between the two components should preserve the labeling of the molecular graph. This last constraint, however, may limit the size of the components, and it has to be loosened in part to get larger ones. These two points are illustrated in Figure 5. In this example, the split is done on an abstract molecular graph in which all bond types are symbolized by a general label, precisely a single line; stereochemistry is omitted too. Of course, in such a case, refunctionalizations will have to be performed in order to take advantage of reflexivity, i.e., to find a single precursor of the target from the pair of synthons. In fact, an evaluation of the solutions may apply other criteria than the sizes of the components. The number and nature of the refunctionalizations as well as the number, types, and positions of the bonds being disconnected in the molecular graph can be taken into account. Then, a smaller solution may be preferred when these criteria are more favorable to it. On the other hand, if these criteria are unfavorable to reflexive solutions, particularly as refunctionalizations look impracticable or too complex, the maximum symmetrical split of the molecular graph will be useful to find convergent solutions, i.e., the ones that lead to pairs of similar precursors.

It is important to notice that the symmetrical split problem is absolutely different from the perception of exact topological symmetry, which has been widely investigated in chemistry, generally through the automorphism group of the molecular graph.[20] Graph drawing is another domain in which symmetry has been extensively studied. Chen et al.[21] have addressed the problem of finding the *maximum symmetric subgraph* that can be computed from a graph by deleting vertices or edges and contracting edges. The *Isomorphic Subgraphs* problem has been formalized by Bachl[22] as finding the two largest disjoint isomorphic subgraphs of a given graph such that they are a copy of each other. The "edge-induced" version of this problem is similar to but more general than the maximum symmetrical split problem: the resulting subgraphs can be unconnected. Since the connectivity cannot be used to prune the search space, the combinatory of the problem is very large. More recently, Bachl and Brandenburg[23] have proposed a greedy heuristic for the approximation of large isomorphic subgraphs, but this does not necessarily lead to an optimal solution. In analogy with disjoint isomorphic subgraphs, Bertz and Sommer[24] have defined the notion of disjoint isomorphic (DI) substructures and described its use in pruning the retrosynthetic search

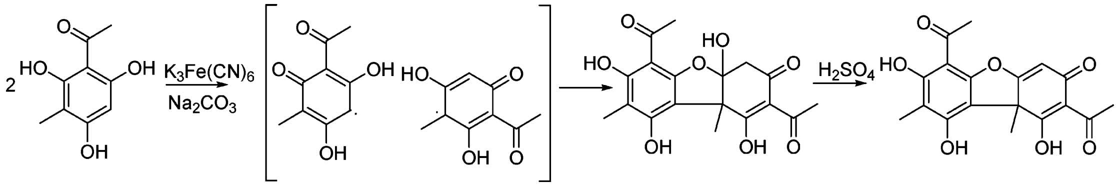


Figure 3. Barton's synthesis of usnic acid.

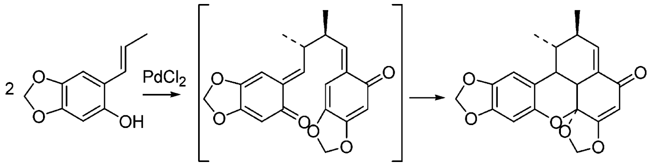


Figure 4. Chapman's synthesis of carpanone.

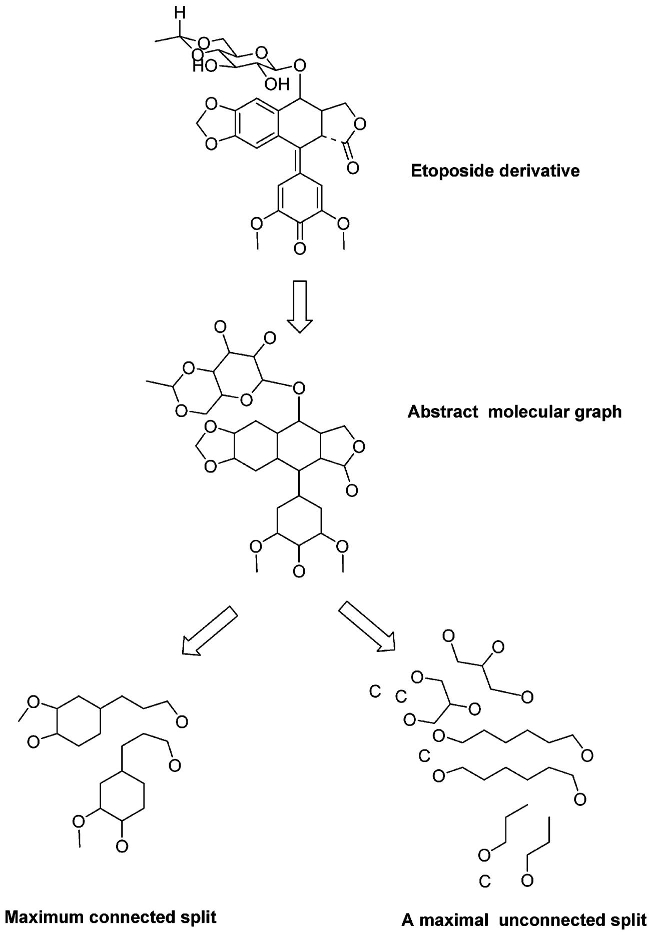


Figure 5. Connected and unconnected maximum symmetrical splits of a molecular graph.

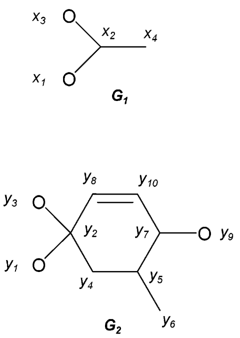
tree. The method for finding DI substructures allows one to discover reflexive routes for target molecules. However, it is a brute-force approach which gives many solutions, because all maximal substructures of size $k = |n|/2$, where n is the set of non-H atoms of the target structure, are enumerated and all pairs of DI substructures are retained. If DI substructures of size k are not found, the process is repeated with $k = k - 1$, and so on until a limit is reached. In our approach, the problem is defined differently, the aim being to directly find maximal solutions.

In a sense, solving Problem 1 is similar to finding a matching from a part of the molecular graph to the remainder. Hence, it is interesting to consider the problem of graph matching. Although the graph isomorphism problem can be solved in polynomial time for molecular graphs,[25] finding a matching from a molecular graph to another one is a NP-Complete problem, that is, no polynomial time algorithm exists to solve it. Several backtrack algorithms have been proposed for this subgraph isomorphism problem, one of the best-known being Ullman's.[26] It has been shown that significant improvements can be achieved in modeling and solving graph matching as a CSP.[27] In the next section we briefly present this model, and then we show that the maximum symmetrical split problem can be described as an original kind of CSP.

GRAPH MATCHING AS A CONSTRAINT SATISFACTION PROBLEM

A constraint satisfaction problem is described by a constraint network defined as a triple whose elements are a set of variables $x_1, x_2, ..., x_k$, a set of values for each variable, and a set of constraints among variables. For instance, a binary constraint specifies which pairs of values can be assigned to a pair of variables. Figure 6 describes a constraint network for testing if a molecular graph G_1 can be matched to a molecular graph G_2: (i) each vertex of G_1 is a variable of the CSP; (ii) a *variable* x can be assigned to any vertex of G_2 whose atom type label is the same as that of x; the set of values which the variable x can take is called the *domain* of x and denoted by $D(x)$; (iii) for each edge (x_1, x_2) in G_1, a *binary constraint* is defined between x_1 and x_2 to represent that the vertices assigned to x_1 and x_2 must be adjacent; therefore, a pair of values $(y_1, y_2) \in D(x_1) \times D(x_2)$ is allowed by the constraint if the edge (y_1, y_2) belongs to G_2; for exact matching search, the binary constraint also ensures that if two edges are matched together, they have the same bond type label (simple, double, ...), e.g., edge (x_2, x_4) cannot match edge (y_8, y_{10}); but, for *potential symmetry* search, as mentioned earlier, it is best to ignore the mismatch between bond type labels; (iv) a constraint of difference is defined on the variables to ensure that they all take different values. Of course, any solution for this constraint network is a matching of G_1 to G_2.

The usual approach to solve a CSP starts by trying to reduce the domains of the variables. The aim of this filtering step is to delete values that cannot be assigned to the variables. For instance, in Figure 6, a value y is deleted from the domain of a variable x if the degree of vertex y is smaller than that of vertex x. Clearly, x cannot be matched with a vertex which has less neighbors. Thus, the domain of variable x_2 is reduced to three values: y_2, y_5, and y_7. Another major filtering method is Arc-Consistency. For any binary constraint between x_i and x_j, the domain of x_i is reduced to the values that have at least one corresponding consistent assignment to x_j. In Figure 6, the value y_5 can be deleted from the domain of x_2 because y_5 cannot be assigned to any value according to the constraints $\{x_1, x_2\}$ and $\{x_3, x_2\}$.



variables	domains
x_1	y_1, y_3, y_9
x_2	$y_2, y_4, y_5, y_6, y_7, y_8, y_{10}$
x_3	y_1, y_3, y_9
x_4	$y_2, y_4, y_5, y_6, y_7, y_8, y_{10}$

constraints of difference
$x_1 \neq x_2 \neq x_3 \neq x_4$

binary constraints	
(x_1, x_2)	(x_2, x_4)
(y_1, y_2)	(y_7, y_{10})
(y_9, y_7)	(y_{10}, y_7)
(y_3, y_2)	(y_7, y_5)
	(y_8, y_2)
(x_3, x_2)	(y_5, y_4)
	(y_5, y_7)
(y_1, y_2)	(y_5, y_6)
(y_9, y_7)	(y_6, y_5)
(y_3, y_2)	(y_2, y_8)
	(y_4, y_5)
	(y_2, y_4)
	(y_4, y_2)

Figure 6. Definition of a CSP to search a matching from graph G_1 to graph G_2.

Increasingly efficient algorithms were developed to test Arc-Consistency: AC-4,[28] AC-6,[29] and AC-7.[30]

Generally, filtering methods dramatically reduce the size of the search space of the CSP. For instance, the previous examples of filtering applied to the CSP of Figure 6 give the following results:

variables	domains
x_1	y_1, y_3, y_9
x_2	y_2, y_7
x_3	y_1, y_3, y_9
x_4	$y_2, y_4, y_5, y_6, y_7, y_8, y_{10}$

constraints of difference
$x_1 \neq x_2 \neq x_3 \neq x_4$

binary constraints		
(x_2, x_4)	(x_3, x_2)	(x_1, x_2)
(y_7, y_{10})	(y_1, y_2)	(y_1, y_2)
(y_7, y_5)	(y_9, y_7)	(y_9, y_7)
(y_2, y_8)	(y_3, y_2)	(y_3, y_2)
(y_2, y_4)		

When the initial filtering step is achieved, a backtracking search is realized to instantiate the variables of the CSP. There is a vast literature on methods which improve this brute-force approach, such as partial filtering after each instantiation, intelligent backtracking, or maintaining Arc-Consistency during the backtrack.[27,29] For constraints of difference, Régin[31] has proposed an efficient filtering method which amounts to finding a maximal matching in the bipartite graph between the set of variables and the set of values. Recently, Larrosa et al.[32] have introduced a specialized filtering method for the CSP formulation of the subgraph isomorphism problem.

A KIND OF MAX-CSP PROBLEM

The problem of maximum symmetrical split can be easily reformulated in terms of pattern matching (Figure 7):

Problem 2. Find two maximum subsets of vertices V_1 and V_2, a minimum subset of edges E_D and a bijective matching ϕ from V_1 to V_2 such as the following:

- V_1 and V_2 are disjoint

- each subgraph G_1 or G_2, respectively induced by V_1 or V_2 in $G' = (V, E \backslash E_D)$, is connected

- ϕ preserves labeling and adjacency of G'

Of course, Problems 1 and 2 are equivalent.

Solving Problem 2 is very similar to searching a specific matching of the molecular graph G onto itself. From a CSP point of view, one can consider V_1 as the set of variables, V_2 as the set of values, and ϕ as an instantiation of the variables that satisfies most of the adjacency constraints of G. As previously stated, each edge of the graph can correspond to a constraint in the CSP. For each solution of Problem 2, E_D is the set of edges that must be deleted in order to obtain the isomorphic subgraphs G_1 and G_2. Hence, these edges can be considered as violated constraints in the CSP, and so Problem 2 is quite similar to a Max-CSP Problem, i.e., a CSP for which no instantiation exists that satisfies all the constraints. Such a problem is solved by finding an instantiation of all the variables that violates the fewest constraints. Nevertheless, Problem 2 is not a classic Max-CSP especially because it requires the distribution of the vertices of G into V_1 and V_2, i.e., into the set of variables and the set of values. Another major difference is that the constraints cannot be satisfied or violated independently, they must respect some specific rules. For instance, if the binary constraint for adjacency between two vertices x and y of G_1 is satisfied by a solution, then the constraint between the vertices of G_2 assigned to x and y must be satisfied too. Otherwise, G_1 and G_2 would not be isomorphic.

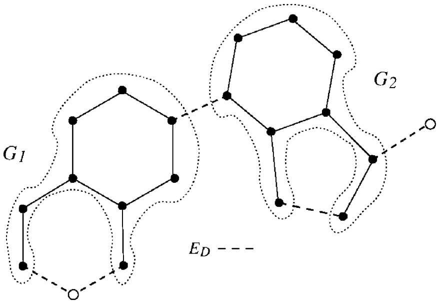


Figure 7. An illustration of Problem 2.

Consequently, it is almost impossible to benefit from the techniques developed to solve Max-CSP[33−36] since these assume that the constraints are considered independently. Faced with a new kind of CSP, we have developed an original approach based on the following main points: (i) searching only connected subgraphs is a good way to prune the search space; (ii) while computing a solution, a variable $v \in V_1$ should be instantiated only if at least one of its neighbors has been already instantiated; and (iii) only the constraints that involve instantiated variables must be checked.

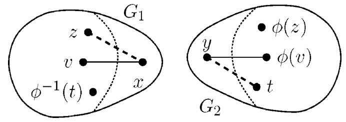


Figure 8. Dashed edges must be removed.

ALGORITHM

Our algorithm for solving Problem 2 applies a classic backtrack principle. It attempts to extend the current solution by assigning to a variable one of the values in its domain. When no new assignment is possible, it backtracks and changes the previous assignment. To ensure that the resulting subgraphs are connected, we propose to dynamically build the constraint network. Our algorithm starts with any vertex of the graph taken as the only element of the set *Variables*. The domain of this first variable is the set of all vertices whose labels are the same as that of the variable. A new vertex will be added to *Variables* only if it is adjacent to an already instantiated variable. Consequently, the resulting subgraphs will actually be connected. The constraint network is built and partially solved at the same time. During this computation, subgraphs G_1 and G_2 are both growing since they include the instantiated variables and the assigned values, respectively. Even if all the vertices are potential variables, at most half of them will really become variables of the constraint network. So the network size is considerably reduced. Dynamically building the constraint network also provides an efficient filtering method to reduce the sizes of the domains. For every value y in the domain of any variable x, we check that at least one instantiated variable v exists such as vertex x is adjacent to vertex v and vertex y is adjacent to vertex $\phi(v)$. This filtering method is easy to implement since all the noninstantiated variables are adjacent to at least one already instantiated variable.

Let us now describe in detail the different steps to building and solving the CSP.

Step 1. Choose a variable x from the set *Variables* which is not yet instantiated. Vertex x is added to G_1. If x is adjacent to any vertex z in G_2, the edge (x, z) is removed, i.e., (x, z) is added to E_D.

Step 2. Choose a value y in $D(x)$ and assign y to x, that is, $\phi(x) = y$. Vertex y is added to G_2. Since G_1 and G_2 must be disjoint, any edge joining y and a vertex of G_1 is removed, i.e., added to E_D. Figure 8 illustrates that some edges incident to x or y must be removed to ensure that G_1 and G_2 will be isomorphic. More precisely, to E_D is added any edge joining x and $z \in G_1$ such as $\phi(z) \notin \Gamma(y)$, and any edge joining y and $t \in G_2$ such as $\phi^{-1}(t) \notin \Gamma(x)$, where $\Gamma(x)$ and $\Gamma(y)$ denote the sets of neighbors of x and of y, respectively. To ensure that G_1 and G_2 are both connected graphs, we assume that for any $y \in D(x)$, there is at least one vertex $v \in \Gamma(x)$ such as $\phi(v) \in \Gamma(y)$. Step 4 shows how to maintain this property.

Step 3. Remove x and y from the domain of any unassigned variable. Consequently, if the domain of any variable v becomes empty, v is deleted from *Variables* and added to the set V_D. This deletion could be provisional if vertex v has some neighbors which are not yet in G_1 or G_2. If one of these neighbors is instantiated afterward, then v will be added to *Variables* again if its domain is not empty.

Step 4. As vertex x is now in G_1, any neighbor of x which is not in G_1, G_2, or V_D can be added to *Variables*. For any neighbor w_i, the domain of variable w_i, denoted by $D(w_i)$, is the set of neighbors of $y = \phi(x)$ which are not in G_1 or G_2, which are different from w_i, and which belong to *S*(w_i), i.e., the set of all vertices whose labels are the same as that of vertex w_i. Figure 9 illustrates this definition, showing that a vertex w_2 can be a variable and a value in the domain of another variable at the same time. The way of adding values to the variable domains ensures that the resulting graphs G_1 and G_2 are both connected graphs. Indeed, a value k is added to $D(w_i)$, only if $w_i \in \Gamma(x)$ and $k \in \Gamma(y)$. Hence, all the values in the domains are vertices adjacent to at least one vertex of the current set V_2.

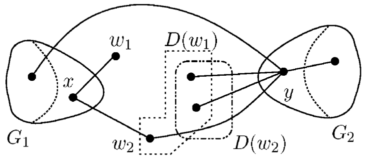


Figure 9. New variables w_1 and w_2 and their domains in step 4.

Step 5. If *Variables* is not empty, go recursively to step 1 in order to instantiate a new variable. Otherwise, save the current solution if it is maximum. Then, backtrack to Step 2 and choose a new value for the current variable.

The main recursive procedure of this algorithm is summarized in Figure 10, and the initialization program is given in Figure 11.

As mentioned in Step 4, due to dynamic building of the constraint network, a vertex can be added to and removed from *Variables* repeatedly. Such a case arises when a previously explored vertex v is reached again but from an adjacent vertex which had not yet been taken as a variable. To avoid circular search, we have to check that any value added to the domain of variable v was not explored for v before. This is done by using the set *OldD*(v), which stores the values already checked for variable v.

Finally, to prevent the algorithm from finding the same solution several times, we have made use of the vertex numbering. As the vertices of the molecular graph must be identified, they are labeled with integers from 1 to $|V|$. Then, the search from a starting vertex r is restricted to the vertices whose labels are greater than that of r. Hence, when the search procedure has been applied to a vertex r, all the edges adjacent to r can be added to E_D (see line 5 of the initialization procedure in Figure 11).

RESULTS AND DISCUSSION

Our algorithm has been implemented into the RESYN_Assistant system. We are developing this program as a tool to help in the understanding of organic synthesis problems. A given molecule is recognized by RESYN_Assistant as a member of different chemical categories according to its features.[37] From this perception, the program builds a new structured representation of the molecule which combines several viewpoints and hierarchical levels of abstraction. This representation can provide a basis for reasoning in problem solving.[38,39] The initial domain of RESYN_Assistant has been extended to organic reactions, and afterward the program has been used in studies of knowledge extraction from reaction databases via data mining techniques.[40,41] Graphic interfaces allow molecules or reactions to be edited

```
Search($Variables, V_1, V_2, D, E_{\mathcal{D}}, OldD, r$)
    $Variables' \leftarrow Variables$;   $\mathcal{E}_{\mathcal{D}} \leftarrow \varnothing$
    While $Variables' \neq \varnothing$ Do
        choose a variable $x$ and delete it from $Variables'$
        Add $x$ to $V_1$
        For each $y \in D(x)$ Do
            $\Phi(x) \leftarrow y$; // assign $y$ to variable $x$
            Add $y$ to $V_2$
            $\mathcal{E}'_{\mathcal{D}} \leftarrow \varnothing$ // edges to add in $E_{\mathcal{D}}$
            For each $t \in \Gamma(y)$ Do
                If $t \in V_1$ Or ($t \in V_2$ And $\Phi^{-1}(t) \notin \Gamma(x)$) Then add $(y, t)$ to $\mathcal{E}'_{\mathcal{D}}$
            $Variables'' \leftarrow Variables'$;
            For each $v \in Variables'$ Do
                $D'(v) \leftarrow D(v) \setminus \{x, y\}$
                If $D'(v) = \varnothing$ Or $v = y$ Then delete $v$ from $Variables''$ End If
            For each $z \in \Gamma(x)$ such that $z > r$ Do
                If $z \in V_1$ Then
                    If $\Phi(z) \notin \Gamma(y)$ And $(x, z) \notin E_{\mathcal{D}}$ Then add $(x, z)$ to $\mathcal{E}'_{\mathcal{D}}$End If
                Else If $z \in V_2$ Then
                    add $(x, z)$ to $\mathcal{E}'_{\mathcal{D}}$
                Else
                    $\Delta \leftarrow \{k \in \mathcal{S}(z) \cap \Gamma(y)$ such that $k \notin V_1, k \notin V_2, k \neq z, k > r$ and $k \notin OldD(z)\}$
                    If $z \in Variables$ Then $D'(z) \leftarrow D'(z) \cup \Delta$
                    Else If $\Delta \neq \varnothing$ Then
                        $D'(z) \leftarrow \Delta$;   Add $z$ to $Variables''$
                    Else add $(x, z)$ to $\mathcal{E}'_{\mathcal{D}}$
            If $|Variables''| > 0$ And current solution may become larger than current maximum Then
                Search($Variables'', V_1, V_2, D', (E_{\mathcal{D}} \cup \mathcal{E}_{\mathcal{D}} \cup \mathcal{E}'_{\mathcal{D}}), OldD, r$)
            Else
                save current solution if it is maximum
            Delete $y$ from $V_2$
        Delete $x$ from $V_1$
        $OldD(x) \leftarrow OldD(x) \cup D(x)$
        $\mathcal{E}_{\mathcal{D}} \leftarrow \mathcal{E}_{\mathcal{D}} \cup \{(x, z),\ z \in \Gamma(x) \cap V_1\}$
    For $x \in Variables$ Do $OldD(x) \leftarrow OldD(x) \setminus D(x)$ End For
```

Figure 10. Algorithm for computing the maximum symmetrical split of a molecular graph.

```
Main($G = (V, E)$)
    $E_{\mathcal{D}} \leftarrow \varnothing$
    For $r = 1$ to $|V|$ Do
        $D(r) \leftarrow \{k \in \mathcal{S}(v)$ such that $k > r\}$
        Search($\{r\}, \varnothing, \varnothing, D, \varnothing, \varnothing, r$)
        $E_{\mathcal{D}} \leftarrow E_{\mathcal{D}} \cup \Gamma(r)$
        If $(|E| - |E_{\mathcal{D}}|) <$ current_Max Then break;
```

Figure 11. Initialization procedure.

and results to be displayed. Written in Java, RESYN_Assistant runs on a large range of platforms.

Results obtained in the detection of potential symmetry in various molecular structures by using our algorithm are shown in Figures 12–20. Most examples have been chosen because they were the subject of published synthetic studies that took advantage of this property. The proposed disconnections of bonds—in bold in the figures—suggest retrosynthetic strategies. A chemist will evaluate their relevance by searching for appropriate transforms and possible starting materials corresponding to the generated synthons. These bond disconnections are not ranked, but the ones which separate the two components will be easily distinguished from those which make the components isomorphic by removing hanging parts. The former can be classified as main disconnections and the latter as adjustment disconnections.

When solving the problem of usnic acid precursor **1**, only one solution is obtained (Figure 12). The proposed disconnections leaves two isomorphic components to which we can relate the single starting material **2**. On the other hand, in the case of usnic acid **3** itself, we get two solutions because the oxygen atom in ring B matches two other oxygen atoms linked to carbon atom 9 in ring A and carbon atom 3 in ring C, respectively (Figure 13). In each solution the oxygen atom which remains unmatched needs to be removed. Thus, disconnections **a** and **b** imply disconnection **a′**, and disconnections **b** and **c** imply disconnection **c′**. To decide between the two solutions, a chemist would have to look for methods for the preparation of ethers in similar structural contexts.

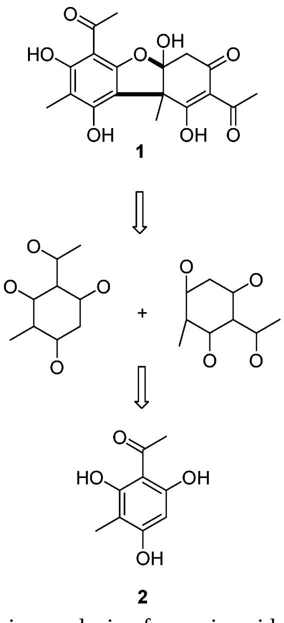


Figure 12. Reflexive analysis of a usnic acid precursor.

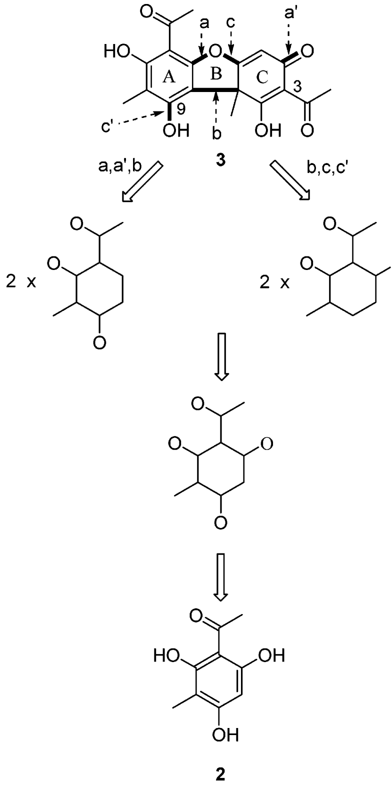


Figure 13. Reflexive analysis of usnic acid.

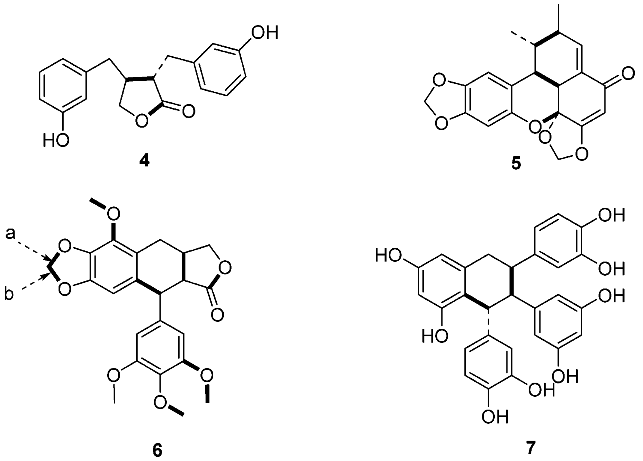


Figure 14. Potential symmetry perception in some lignan and aryltetralin structures; for structure **6**, two solutions are found which differ by only one bond disconnection: the alternative to bond **a** is bond **b**.

A Michael-like addition of a phenate followed by elimination of a hydroxide as described by Barton et al.[10] meets disconnection **c**. Moreover, superimposing the generated subgraphs shows a common supergraph which brings us back to the previous problem and the same starting material **2**. Similar cases where a set with an odd number of atoms has to match onto itself may be frequently encountered. Results obtained with other natural products from various families are presented below.

The lignan class is rich in examples (Figure 14). One of the simplest is enterolactone **4** whose potential symmetry is obvious. The split of its molecular graph by disconnecting the two bonds in bold gives a pair of equivalent synthons related to hydrocinnamic acid derivatives. A total synthesis of this molecule was described, the key step of which was an oxidative coupling of 3-methoxyhydrocinnamic acid dianion.[42] As said before, the other lignan carpanone **5** is a much more complex case. Nevertheless, the set of bond disconnections corresponding to the known reflexive synthesis[11] is found quickly by applying our algorithm. The cyclolignan subgroup contains many other instances, including *β*-peltatin-A methyl ether **6**. For this molecule, the program provides two solutions. Both involve the same main bond disconnections to split the cyclolignan skeleton and lead to precursors related to cinnamic acid. However, the adjustment disconnections point to difficult refunctionaliza-

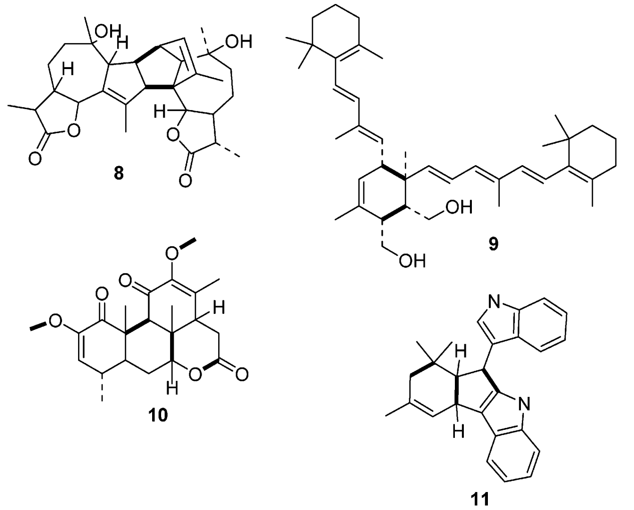


Figure 15. Potential symmetry-based strategy can meet Diels−Alder transform-based tactic.

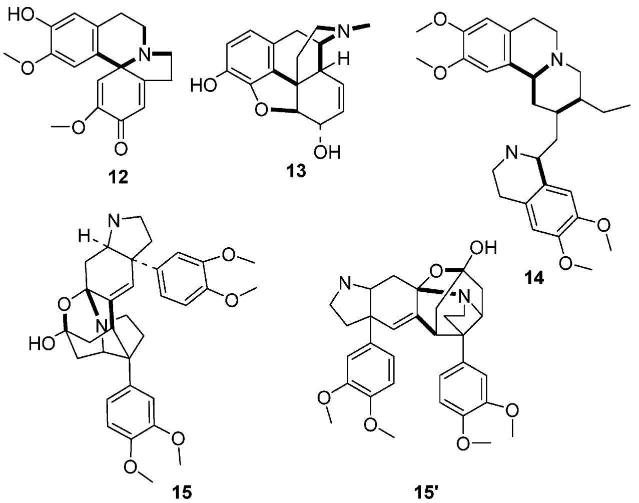


Figure 16. Potential symmetry perception in some alkaloid structures.

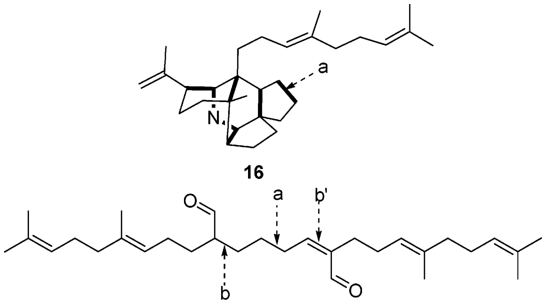


Figure 17. The *proto*-daphniphylline case.

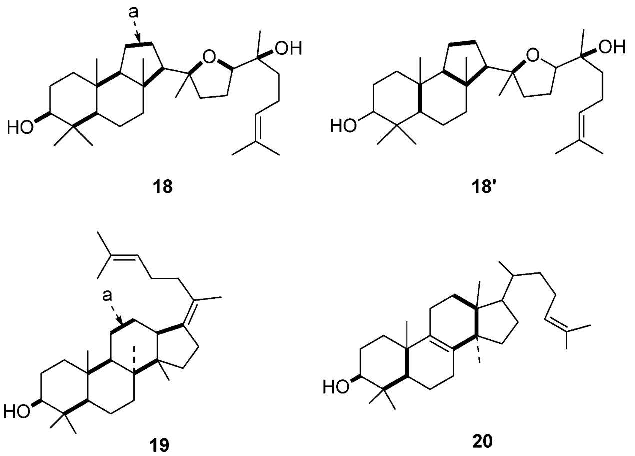


Figure 18. Potential symmetry perception in some triterpenoid structures.

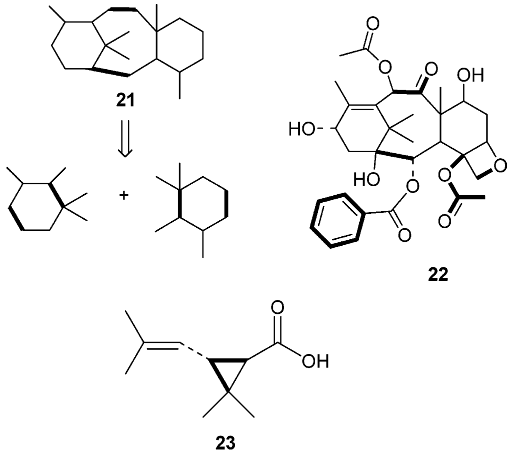


Figure 19. Potential symmetry perception in some mono and diterpene structures.

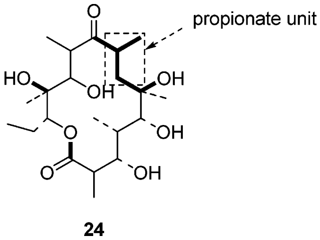


Figure 20. The erythronolide A case.

tions that are needed to get a single precursor in a reflexive synthetic plan. In fact, the known synthesis of **6** is convergent, starting from well substituted derivatives of cinnamic alcohol and phenylpropiolic acid.[43] Other aryltetralin derivatives, such as resformicol A **7**,[44] possess a similar potential symmetry (Figure 14). In the syntheses of compounds **6** and **7**, Diels−Alder reactions have formed the key steps. Reactions of this type, intra- or intermolecular, are often met in synthetic studies of natural products with potential symmetry. It is the case, for instance, of terpenes such as absinthin **8**,[45] kitol **9**,[46] and quassin **10**,[47] or of alkaloids such as yuehchukene **10**[48] (Figure 15). For each of them, our algorithm gives only one solution, which suggest that one could make use of a Diels−Alder reaction to carry out a reflexive synthesis.

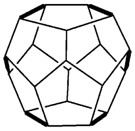
Besides yuehchukene, a number of alkaloids possess potential symmetry. Erysodienone **12**, morphine **13**, and emetine **14** are typical members of the isoquinoline alkaloid family (Figure 16). The solutions proposed for the two first molecules correspond to known convergent syntheses[49,50] but not for the last one.[51] Though synthesis of channaine **15** has yet to be described, this example is interesting because the set of bond disconnections shows that **15** is probably a dimer of *N*-demethylmesembrenone.[52] It should be noted that potential symmetry is not easily recognized at a glance from the two channaine formulas **15** and **15′** depicted in Figure 16.[53]

Such a visual perception is even more difficult in the case of *proto*-daphniphylline **16** (Figure 17). The solution our algorithm proposes is a set of disconnections which concern the four carbon−carbon bonds and the two carbon−nitrogen bonds formed by the biomimetic process of pentacyclization discovered by Heathcock et al.[54] during the synthesis of this molecule. The sixth bond disconnection of the set, marked with **a**, subsequently generates two equivalent synthons related to farnesyl derivatives, but it does not fit in with Heathcock's route. In this one, squalene derivative **17**, whose pentacyclization gave **16**, was obtained according to a (C_{13} + C_4 + C_{13}) strategy based on the bond disconnections marked with **b** and **b′**.

Squalene is a key intermediate in the biosynthesis of steroids and triterpenoids.[55] In the last class of structures, the rings often hide the symmetry of possible precursors, but our algorithm may reveal it (Figure 18). For example, in the first solution proposed for malabaricanediol **18**, the disconnections of cyclic bonds generate two equivalent synthons related to farnesyl derivatives as above, the disconnection marked with **a** splitting symmetrically the squalene skeleton.[56] The second solution is also maximum, but the bond disconnections break the structural isoprene rule and the induced symmetry. So far, the first solution is the only one that has been successfully experimented.[57] The same relationships between farnesyl, squalene, and a triterpene are found in the example of protosterol **19** whose C-20 cation is the key intermediate in the enzymatic conversion of 2,3-oxidosqualene to lanosterol.[55] On the other hand, these relationships are not found with lanosterol **20** itself. The two methyl migrations from C-14 to C-13 and from C-8 to C-14 break the initial symmetry of the squalene skeleton, and then the maximum symmetrical split of the molecular graph of this substance is not based on the structural isoprene rule.

As for taxane skeleton **21**, it is a regular diterpene framework as shown in Figure 19. The proposed disconnections of two bonds of the eight-membered ring correspond to the strategy applied by Nicolau et al.[58] in the synthesis of the baccatin III moiety **22** of taxol. The two resulting synthons have both the same monoterpene skeleton, including a six-membered ring, which can be split into two isoprene units. In a general way, monoterpenes give such a split; chrysanthemic acid **23** is a typical example[59] (Figure 19).

Other examples of potential symmetry can also be found in the class of macrocyclic compounds. These examples are often complex. For instance, our algorithm proposes 10 different maximum symmetrical splits for the macrolide erythronolide A **24**. One of them, which generates two carbohydrate-related synthons and a propionate unit, corresponds to the synthetic strategy proposed by Hanessian and Rancourt[60] (Figure 20).



25

Figure 21. Maximum symmetrical split of the dodecahedrane graph.

Of course, the maximum symmetrical split also applies to molecular graphs which present actual symmetry. The results may be useful in the design of syntheses of complex structures such as of dodecahedrane **25**[61] (Figure 21).

The present version of the algorithm is effective enough to quickly analyze the molecular graphs of most organic compounds. The execution times we have observed on different PCs are generally lower than 1 s. However, some improvements could be made. For example, a particular order on the vertex set which speeds up the process could be researched. As explained before, the way the problem space is searched depends on vertex numbering, and therefore the execution time may vary substantially with the numbering order. It would be interesting to study how numbering modifications, according to some properties of the vertices, change the speed of the algorithm. Some of such vertex features are already perceived by RESYN_Assistant. These are, in particular, atom type, nature of neighbors, functionality, stereochemical, and topological status. In addition, it should be useful to calculate the maximum size that graph G_1 can reach when starting from vertex r and exploring only vertices whose numbering labels are greater than that of r, because it is useless to continue the search if this size is smaller than that of the current maximum solution. Such an evaluation could even be performed dynamically during the exploration, so as to know more precisely the maximum sizes that G_1 or G_2 can reach according to the distribution of the remaining vertices.

In the above examples, the largest isomorphic connected substructures were searched ignoring the mismatches of bond type labels. A worthwhile extension of the program could be the implementation of an interface allowing the user to modify unary constraints induced by the labeling on vertices and edges. Thus, the algorithm would be easily set up to take into account more or less specific information on atoms and bonds, i.e., to extend or limit the domains. For instance, the domain of each aromatic-type edge could be limited to the set of the other edges of the same type, or it could be extended to the set of the other ring-type edges, or of the other edges without regard to their type. We think that the exploration could start with very specific information and be gradually continued with more generic one. This interface implementation should be concomitant with the development of an evaluation function. This one could take into account the sizes of the solutions and the cost of the corresponding symmetrical splits according to the topological, functional, and stereochemical changes they involve. For instance, a solution would be penalized by aromatic bond disconnections as well as by uncertain refunctionalizations or by impracticable stereochemical modifications. The proximity of the components of every solution to known starting materials could be used as another parameter of the function. However, tuning all the parameters of such a function can be difficult because the heuristics needed for this task are often specific to every family of compounds.

CONCLUSION

Symmetry is a major molecular feature to which an organic chemist has to pay attention at every stage of synthesis design.[62] The present work particularly concerns the potential symmetry of synthetic targets, the perception of which can allow retrosynthetic strategies to be developed for planning reflexive or highly convergent syntheses. This perception is a problem we have formalized in graph-theoretical terms as the maximum symmetrical split of a molecular graph. To solve this problem, we have devised an algorithm based on CSP techniques which can drastically reduce the size of its search space. In fact, we have found that the maximum symmetrical split problem can be described as a new kind of CSP. Our algorithm is intended as a tool for helping to make decisions during retrosynthesis. That is why it has been implemented as a new function of RESYN_Assistant, a system for computer-aided understanding of synthesis problems, whose aim is to perceive and represent target molecules according different synthetic viewpoints. This program, by itself, is not a procedure for generating synthetic pathways, but it could be incorporated into a larger system for organic synthesis design. The current version of the algorithm gives good results with varied synthetic targets, as shown by about 20 examples. For each of them, the program has quickly found all the pairs of largest similar synthons and the minimum sets of bond disconnections to get them; most of the proposed solutions correspond to published synthetic studies. Obviously, relevant results can be obtained only with structures that lend themselves to reflexive or highly convergent syntheses. For instance, applied to the structure of taxol the algorithm gives only unrealistic solutions, whereas it proposes a pertinent split of the taxane skeleton (Figure 19). Then, during a retrosynthesis, potential symmetry shall be searched as much in precursors as in the target structure. The display of results will reveal significant clues helping a chemist to develop short streamlined retrosynthetic routes.

ACKNOWLEDGMENT

One of the authors (C.L.) is indebted to Pr. A. P. Johnson for old but enlightening discussions.

REFERENCES AND NOTES

(1) According to Wender this operation should also be safe, environmentally acceptable, and resource-effective: Wender, P. A.; Handy, S. T.; Wright, D. L. Towards the ideal synthesis. *Chem. Ind.* **1997**, 765−769.

(2) Deslongchamps, P. Le concept de stratégie en synthèse organique. *Bull. Soc. Chim. Fr.* **1984**, *N° 9−10, II*, 349−361.

(3) Velluz, L.; Valls, J.; Mathieu, J. Spatial Arrangement and Preparative Organic Synthesis. *Angew. Chem., Int. Ed. Engl.* **1967**, *6*, 778−789.

(4) Hendrickson, J. B. Systematic Synthesis Design. 6. Yield, Analysis and Convergency. *J. Am. Chem. Soc.* **1977**, *99*, 5439−5450.

(5) Synthons have been defined by Corey as “structural units within molecules which can be formed and/or assembled by known or conceivable synthetic operations”.[9] Because of its general definition, this term has been (mis)interpreted in various ways, causing some confusion. Originally, it was used by Corey in the context of retrosynthesis to refer to both structural features of a target to which the transforms may be keyed (these substructures were later renamed *retrons*) and structural fragments that could result from bond disconnections. In this paper, it is used according to the second meaning. This is consistent with the synthon approach to organic synthesis as described in Warren’s classic book: Warren, S. *Organic Synthesis: The Disconnection Approach*; Wiley: Chichester, U.K., 1982.

(6) Bertz, S. H. The Role of Symmetry in Synthesis Analysis. The Concept of Reflexivity. *J. Chem. Soc., Chem. Commun.* **1984**, 218−219.

(7) McMurry, J. E.; Fleming, M. P. A New Method for the Reductive Coupling of Carbonyls to Olefins. Synthesis of β-Carotene. *J. Am. Chem. Soc.* **1974**, *96*, 4708−4709.

(8) Vaz, B.; Alvarez, R.; de Lera, A. R. Synthesis of Symmetrical Carotenoids by a Two-Fold Stille Reaction. *J. Org. Chem.* **2002**, *67*, 5040−5043.

(9) Corey, E. J. General Methods for the Construction of Complex Molecules. *Pure Appl. Chem.* **1967**, *14*, 19−37.

(10) Barton, D. H. R.; Delflorin, A. M.; Edwards, O. E. Synthesis of Usnic Acid. *J. Chem. Soc.* **1956**, 530−534.

(11) Chapman, O. L.; Engel, M. R.; Springer, J. P.; Clardy, J. C. The Total Synthesis of Carpanone. *J. Am. Chem. Soc.* **1971**, *93*, 6696−6698.

(12) Laurenço, C.; Johnson, A. P. Unpublished results.

(13) For an overview of the LHASA project, see: Corey, E. J.; Long, A. K.; Rubenstein, S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408−418 and http://lhasa.harvard.edu.

(14) Johnson, A. P.; Marshall, C.; Judson, P. N. Some recent progress in the development of the LHASA computer system for organic synthesis design: Starting-material-oriented retrosynthetic analysis. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 310−316.

(15) Johnson, A. P.; Marshall, C. Starting Material Oriented Retrosynthetic Analysis in the LHASA Program. 2. Mapping the SM and Target Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 418−425.

(16) Wagener, M.; Gasteiger, J. The Determination of Maximum Common Substructures by a Genetic Algorithm: Application in Synthesis Design and for the Structural Analysis of Biological Activity. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1189−1192.

(17) Raymond, J. W.; Willett, P. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 521−533.

(18) RESYN_Assistant is the acronym for REtroSYNthesis Assistant. The development of this system was undertaken in 1996 at LIRMM, in the framework of GDR 1093 of CNRS “Traitement Informatique de la Connaissance en Chimie Organique” (1993−2000, director: C. Laurenço), with a financial support from Sanofi Chimie and the TIIM Pole of the Region Languedoc Roussillon, and with an outstanding contribution of Dr. P. Jambaud.

(19) The problem of the maximum symmetrical split of graphs and molecular graphs was defined and first studied in Tognetti's thesis.[37] The principle and some preliminary results were presented at the RFIA'00 conference: Tognetti, Y.; Vismara, P. A Max-Var-CSP algorithm for the symmetrical split problem. In *Proceedings of the 12ème Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RIFA'00)*; Paris, 2000; pp 449−458.

(20) Ivanciuc, O. Canonical Numbering and Constitutional Symmetry. In *Handbook of Chemoinformatics*; Gasteiger, J., Ed.; Wiley-VCH: Weinheim, 2003; pp 139−160.

(21) Chen, H.-L.; Lu, H.-I.; Yen, H.-C. On Maximum Symmetric Subgraphs. *Lecture Notes in Computer Science* **2001**, *1984*, 372−383.

(22) Bachl, S. Isomorphic Subgraphs. *Lect. Notes Comput. Sci.* **1999**, *1731*, 286−296.

(23) Bachl, S.; Brandenburg, F.-J. Computing and Drawing Isomorphic Subgraphs. *Lect. Notes Comput. Sci.* **2002**, *2528*, 74−85.

(24) Bertz, S. H.; Sommer, T. J. The role of isomorphism in synthetic analysis. Pruning the search tree by finding disjoint isomorphic substructures. *Chem. Commun.* **2003**, 1001−1001.

(25) Foulon, J.-L. Isomorphism, Automorphism Partitioning, and Canonical Labeling Can Be Solved in Polynomial-Time for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 432−444.

(26) Ullman, J. R. An algorithm for subgraph isomorphism. *J. ACM* **1976**, *23*, 31−42.

(27) McGregor, J. J. Relational Consistency Algorithms and Their Application in Finding Subgraph and Graph Isomorphisms. *Inf. Sci.* **1979**, *19*, 229−250.

(28) Mohr, R.; Henderson, T. C. Arc and Path Consistency Revisited. *Artif. Intell.* **1986**, *28*, 225−233.

(29) Bessière, C. Arc-Consistency and Arc-Consistency Again. *Artif. Intell.* **1994**, *65*, 179−190.

(30) Bessière, C.; Freuder, E.; Régin, J.-C. Using Inference to Reduce Arc Consistency Computation. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*; Morgan Kaufmann: San Mateo, CA, 1995; Vol. 1, pp 592−598.

(31) Régin, J.-C. A Filtering Algorithm for Constraints of Difference in CSPs. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*; AAAI Press: Menlo Park, CA, 1994; pp 362−367.

(32) Larrosa, J.; Valiente, G. Constraint satisfaction algorithms for graph pattern matching. *Math. Struct. Comput. Sci.* **2002**, *12*, 403−422.

(33) Freuder, E. C.; Wallace, R. J. Partial Constraint Satisfaction. *Artif. Intell.* **1992**, *58*, 21−70.

(34) Wallace, R. J.; Freuder, E. C. Conjunctive Width Heuristics for Maximal Constraint Satisfaction. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI-93)*; AAAI Press: Menlo Park, CA, 1993; pp 762−768.

(35) Verfaillie, G.; Lemaître, M.; Schiex, T. Russian Doll Search for Solving Constraint Optimization Problems. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96)*; AAAI Press: Menlo Park, CA, 1996; pp 181−187.

(36) Larossa, J. Algorithms and Heuristics for Total and Partial Constraint Satisfaction Problems. Ph.D. Thesis, Technical University of Catalonia, Barcelona, Spain, 1998.

(37) Tognetti, Y. Contribution à la modélisation des systèmes d'information chimique par la théorie et l'algorithmique de graphes. Ph.D. Thesis, University of Montpellier II, Montpellier, France, 2002.

(38) Vismara, P. Reconnaissance et représentation d'éléments structuraux pour la description d'objets complexes. Application à l'élaboration de stratégies de synthèse en chimie organique. Ph.D. Thesis, University of Montpellier II, Montpellier, France, 1995.

(39) Vismara, P.; Laurenço, C. An abstract representation for molecular graphs. In *Discrete Mathematical Chemistry*; Hansen, P., Fowler, P., Zheng, M., Eds.; DIMACS series in discrete mathematics and theoretical computer science, American Mathematical Society & DIMAC: 2000; Vol. 51, pp 343−366.

(40) Berasaluce, S. Fouille de données et acquisition de connaissances à partir de bases de données de réactions chimiques. Ph.D. Thesis, University of Nancy I, Nancy, France, 2002.

(41) Laurenço, C.; Berasaluce, S.; Jauffret, P.; Napoli, A.; Niel, G. Fouille de données dans les bases de données de réactions: Extraction de connaissances sur les méthodes de synthèse. In *Proceedings of “Chimiométrie 2003”*, Paris, France, Dec 3−4, 2003; pp 63−66; http://www.chimiometrie.org.

(42) Belletire, J. L.; Fremont, S. L. Oxidative Coupling. II. The Total Synthesis of Enterolactone. *Tetrahedron Lett.* **1986**, *27*, 127−130.

(43) Kashima, T.; Tanoguchi, M.; Arimoto, M.; Yamaguchi, H. Studies on the Constituents of the Seeds of *Hermandia ovigera* L. VIII. Syntheses of (±)-Desoxypodophyllotoxin and (±)-β-Peltatin-A Methyl Ether. *Chem. Pharm. Bull.* **1991**, *39*, 192−194.

(44) Li, X.-M.; Huang, K.-S.; Lin, M.; Zhou, L.-X. Studies on formic acid-catalyzed dimerization of isorhapontigenin and of resveratrol to tetralins. *Tetrahedron* **2003**, *59*, 4405−4413.

(45) Beauhaire, J.; Fourrey, J.-L.; Vuilhorgne, M.; Lallemand, J.-Y. Dimeric Sesquiterpene Lactones: Structure of Absinthin. *Tetrahedron Lett.* **1980**, *21*, 3191−3194.

(46) Ghosh, M. C.; Rahman, M.; Ghosh, S. Preparation of Kitol by Oxidative Esterification of Retinol. *Ind. J. Biochem. Biophys.* **1973**, *10*, 289−290.

(47) Mandell, L.; Lee, D. E.; Courtney, L. F. Toward the Total Synthesis of Quassin. *J. Org. Chem.* **1982**, *47*, 610−615.

(48) Cheng, K. F.; Kong, Y. C.; Chan, T. Y. Biomimetic Synthesis of Yeuhchukene. *J. Chem. Soc., Chem. Commun.* **1985**, 48−49.

(49) Gervay, J. E.; McCapra, F.; Money, T.; Sharma, G. M. Phenol Oxidation. A model for the Biosynthesis of the *Erythrina* Alkaloids. *J. Chem. Soc., Chem. Commun.* **1966**, 142−143.

(50) Toth, J. E.; Fuchs, P. L. Total Synthesis of *dl*-Morphine. *J. Org. Chem.* **1987**, *52*, 473−475.

(51) Kametani, T. The Total Syntheses of Isoquinoline Alkaloids. In *The Total Synthesis of Natural Products*; ApSimon, J., Ed.; Wiley: New York, 1977; Vol. 3, pp 1−272.

(52) Jeffs, P. W.; Redfearn, R.; Wolfram, J. Total Syntheses of (±)-Mesembrine, (±)-Joubertinamine, and (±)-N-Demethylmesembrenone. *J. Org. Chem.* **1983**, *48*, 3861−3863.

(53) These formulas were obtained by queries to SciFinder Scholar.

(54) Heathcock, C. H.; Piettre, S.; Ruggeri, R. B.; Ragan, J. A.; Kath, J. C. *Daphniphyllum* Alkaloids. 12. A Proposed Biosynthesis of the Pentacyclic Skeleton. *proto*-Daphniphylline. *J. Org. Chem.* **1992**, *57*, 2554−2566.

(55) Abe, I.; Rohmer, M.; Prestwich, G. D. Enzymatic Cyclization of Squalene and Oxidosqualene to Sterols and Triterpenes. *Chem. Rev.* **1993**, *93*, 2189−2206.

(56) Biellmann, J.-F.; Ducep, J.-B. Synthèse du squalène par couplage queue à queue. *Tetrahedron Lett.* **1969**, *42*, 3707−3710.

(57) Sharpless, K. B. *d,l*-Malabaricandiol. The First Cyclic Natural Product Derived from Squalene in a Nonenzymic Process. *J. Am. Chem. Soc.* **1970**, *92*, 6999−7001.

(58) Nicolaou, K. C.; Yang, Z.; Liu, J. J.; Ueno, H.; Nantermet, P. G.; Guy, R. K.; Claiborne, C. F.; Renaud, J.; Couladouros, E. A.; Paulvannan, K.; Sorensen, E. J. Total Synthesis of Taxol. *Nature* **1994**, *367*, 630−634.

(59) Krief, A.; Provins, L. Stereoselective Synthesis of Methyl *trans*-Chrysanthemate and Related Derivatives. *Tetrahedron Lett.* **1998**, *39*, 2017−2020.

(60) Hanessian, S.; Rancourt, G. Carbohydrates as chiral intermediates in organic synthesis. Two functionalized chemical precursors comprising eight of the 10 chiral centers of erythronolide A. *Can. J. Chem.* **1977**, *55*, 1111−1113.

(61) Alvarez, S.; Serratosa, F. Symmetry Guidelines for the Design of Convergent Syntheses. On Narcissistic Coupling and La Coupe du Roi. *J. Am. Chem. Soc.* **1992**, *114*, 2623−2630.

(62) Ho, T.-L. *Symmetry: a basis for synthesis design*; Wiley-Interscience: New York, 1995.

CI049713H