

Efficient Enumeration of Stereoisomers of Outerplanar Chemical Graphs Using Dynamic Programming

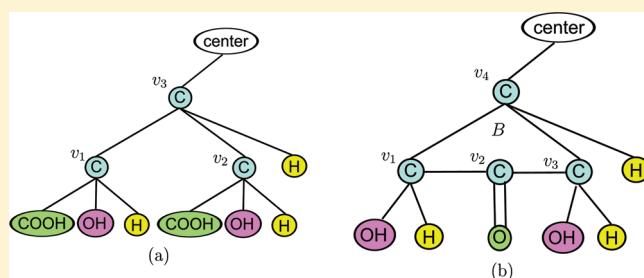
Tomoki Imada,[†] Shunsuke Ota,[†] Hiroshi Nagamochi,^{*,†} and Tatsuya Akutsu^{*,‡}

[†]Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Yoshida, Kyoto 606-8501, Japan

[‡]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

 Supporting Information

ABSTRACT: Exhaustive and nonredundant generation of stereoisomers of a chemical compound with a specified constitution is an important tool for molecular structure elucidation and molecular design. It is known that many chemical compounds have outerplanar graph structures. In this paper we deal with chemical compounds composed of carbon, hydrogen, oxygen, and nitrogen atoms whose graphical structures are outerplanar and consider stereoisomers caused only by asymmetry around carbon atoms. Based on dynamic programming, we propose an algorithm of generating all stereoisomers without duplication. We treat a given outerplanar graph as a graph rooted at its structural center. Our algorithm first recursively computes the number of stereoisomers of the subgraph induced by the descendants of each vertex and then constructs each stereoisomer by backtracking the process of computing the numbers of stereoisomers. Our algorithm correctly counts the number of stereoisomers in $O(n)$ time and space and correctly enumerates all of the stereoisomers in $O(n^3)$ time per stereoisomer on average and in $O(n)$ space, where n is the number of atoms in a given structure.



INTRODUCTION

Enumeration of molecules is one of the oldest problems in chemoinformatics and has attracted chemists and mathematicians for more than a century.^{1,2} It is also important from a practical viewpoint because it plays an important role in experimental structure elucidation (e.g., by NMR or MS) and molecular design.³ Among various problems for enumeration of molecules, we focus here on enumeration of stereoisomers for a given chemical structure. Stereoisomers are defined as isomers of identical constitutions (i.e., atom–bond relationships) that differ in the three-dimensional arrangement of their atoms. Since stereoisomers exhibit different chemical properties, enumeration of stereoisomers is an important subproblem in the enumeration of molecules.

Although various types of stereoisomers can be considered,^{4–6} we focus here on stereoisomers caused only by asymmetry around carbon atoms, which includes cis–trans asymmetry caused by double bonds, because these are considered to be fundamental and practically important. For enumeration of such isomers, several methods have been proposed,^{7–9} which mostly follow the work by Nourse.¹⁰ Most of these methods adopt an approach of combination of identification of m possible stereocenters (represented by a set of 2^m m -tuples), creation of possible stereoisomers, and deletion of identical stereoisomers using permutation groups called configuration groups. However, it seems that the worst case time complexity of such methods remains $\Omega(2^m)$, since the algorithm repeatedly generates a new

m -tuple which does not belong to any of the previously constructed equivalence classes until no such m -tuple remains. As different approaches, Gugisch and Rücker proposed a method using an orientation function⁴ following a suggestion by Dress et al.,¹¹ and Tratch et al. proposed a method using the concept of ladder of combinatorial objects.¹² Though these approaches can handle more general situations than those by Nourse and others, it seems that these methods are less efficient. It is also to be noted that mathematical proofs for the correctness of some of the existing methods are not fully provided. This situation motivated Rücker et al. to provide manually created examples for checking the validity of existing programs.¹³

In our previous work, we developed an efficient algorithm for enumerating stereoisomers for tree structured molecules (i.e., acyclic molecules) along with a proof of its correctness.¹⁴ Different from the existing approaches based on configuration groups, the algorithm is based on dynamic programming. Dynamic programming is a general technique for efficiently solving various kinds of problems in computer science and mathematics. It recursively breaks the original problem into subproblems and then combines the solutions of the subproblems to obtain an overall solution. Many subproblems are often identical, and thus tables are usually employed to store the solutions in order to eliminate recalculation of identical subproblems. In our previous

Received: February 17, 2011

Published: August 17, 2011

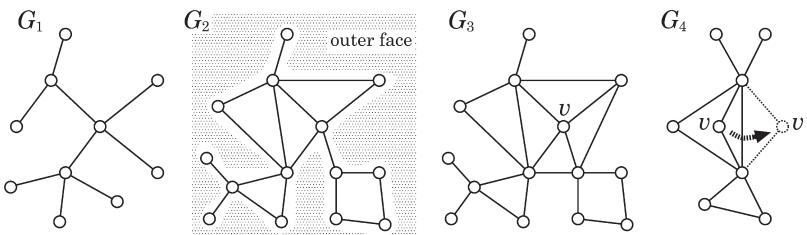


Figure 1. Tree, outerplanar graph, and nonouterplanar graph. G_1 is a tree, which is always an outerplanar graph. G_2 is an outerplanar graph. In G_2 , all vertices are adjacent to the outer face (the unbounded region surrounding the graph). G_3 is a planar graph but is not an outerplanar graph because vertex v is not adjacent to the outer face. G_4 is an outerplanar graph because v can be moved to v' so that the vertex is adjacent to the outer face.

algorithm, an input tree structured molecule is recursively broken into subtrees, and the numbers of stereoisomers of smaller subtrees are used to compute the number of stereoisomers of larger subtrees. It is shown that the algorithm correctly enumerates all stereoisomers without duplication in $O(n)$ space and in $O(n)$ time per stereoisomer, where n is the number of atoms in a given tree. In this paper, we extend the work from the class of tree-structured chemical compounds to a wider class of chemical compounds. We employ the class of chemical compounds that have outerplanar graph structures. A graph is outerplanar if it can be drawn on a plane in such a way that all vertices lie on the outer face without crossing of edges (see Figure 1), where the outer face is the unbounded exterior region. Although there may exist multiple embeddings (i.e., drawings on a plane) of an outerplanar graph, it is known that one embedding can be computed in linear time when only structural information of a graph is given.¹⁵ Outerplanar graphs are a well-studied class of graphs in computer science, for which many efficient algorithms have been developed. Furthermore, it is suggested that 94.3% of chemical compounds in the NCI chemical database have outerplanar graph structures.¹⁶ Therefore, it is reasonable to select the class of outerplanar graphs as the target class for enumeration of stereoisomers. It is to be noted that although outerplanar graphs are a subclass of planar graphs, Rücker and Meringer reported that most organic chemical compounds have planar graph structures, whereas most graphs are nonplanar even if the maximum degree is bounded by 4.¹⁷

In this paper, we present an algorithm for enumerating stereoisomers of a given outerplanar chemical graph with guaranteed computational complexity. Similar to our previous work,¹⁴ our algorithm is based on dynamic programming, although it is far from a simple extension due to the existence of cycles. For this, we treat a given outerplanar chemical graph as a graph rooted at its structural center and derive recursive formulas for the number of stereoisomers of rooted subgraphs which have a parent-child relationship. However, it is nontrivial to represent stereoisomers with a mathematically consistent form, without which such recursive formulas cannot be derived. The main contribution of this paper is to give a mathematical representation for stereoisomers and to design a dynamic programming algorithm that counts the total number of stereoisomers, K , of a given outerplanar chemical graph by integrating the number of the stereoisomers of each of the rooted subgraphs according to the derived recursive formulas and a traceback algorithm that constructs the k -th stereoisomer of the graph for each $k = 1, 2, \dots, K$ by backtracking the integration process to identify the stereoisomer of each rooted subgraph corresponding to the k -th stereoisomer. Assuming that each of the four arithmetic operations can be done in constant time, our algorithm correctly counts the number of

stereoisomers, K , in $O(n)$ time and space and correctly enumerates all K stereoisomers without duplication in $O(n)$ space and in $O(n^3)$ time per stereoisomer on average, where n is the number of atoms in a given graph. The time complexity for counting is optimal. It is to be noted that other existing approaches (other than our dynamic programming-based approach) do not yield the guaranteed polynomial time complexity for counting or for enumeration per stereoisomer, although our approach is limited to a class of tree-like chemical graphs¹⁴ or a class of outerplanar chemical graphs (this paper). By limiting a class of chemical graphs, the dynamic programming technique can be effectively applied, and thus we can obtain the polynomial time complexity. We conducted computational experiments to evaluate the practical computation time of the proposed algorithm. The results confirm that our proposed algorithm is significantly fast in practice for both counting and enumeration.

A preliminary nonrefereed version of this paper appeared as part of the master thesis of the first author.¹⁸

■ PRELIMINARY AND PROBLEM FORMULATION

Problem Definition. In this paper, we only consider outerplanar graphs. We also assume that an outerplanar embedding of a graph is given. That is, a graph is given with its drawing on a plane so that all vertices belong to the outer face and no two edges cross. It is known that the planar embedding of an outerplanar graph can be computed in $O(n)$ time.¹⁵

A cutvertex of a connected graph is a vertex whose removal disconnects the graph. In the outerplanar graph in Figure 2a, vertices v , u , w_1 , w_2 , w_3 , and u_c are cutvertices. A graph is 2-connected if it is connected and has no cutvertex. The 2-connected components of a graph are its maximal 2-connected subgraphs. In a graph, we call a 2-connected component with at least three vertices a block and a 2-connected component consisting of two vertices a bridge. For example, in the outerplanar graph in Figure 2a, the subgraph surrounded by the dashed circle is a block, and the edge e and its end vertices form a bridge.

In this paper, a chemical graph $G = (\nu, E)$ is defined to be a multigraph whose vertex set V is partitioned into $V_H = \{\nu | \nu \text{ is a hydrogen atom}\}$, $V_C = \{\nu | \nu \text{ is a carbon atom}\}$, $V_O = \{\nu | \nu \text{ is an oxygen atom}\}$, and $V_N = \{\nu | \nu \text{ is a nitrogen atom}\}$, where a double (respectively, triple) bond between two atoms is represented by a set of two (respectively, three) multiple edges joining the corresponding vertices in G . It is to be noted that all atoms in a chemical structure appear as vertices of G , that is, G is H-explicit. Given a multigraph G , let $S(G)$ denote the simple graph obtained from G by ignoring its multiplicity of each edge. Conversely, given a simple graph G' , let $\mathcal{M}(G')$ denote the set of chemical graphs represented by multigraphs G such that $G' = S(G)$. Thus a

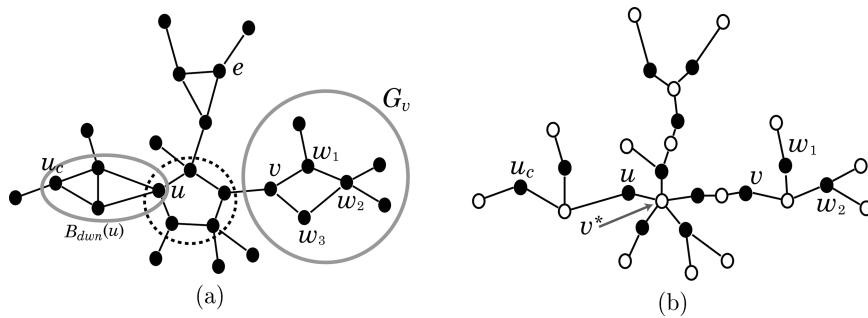


Figure 2. (a) An outerplanar graph G_v , where the center of G_v is surrounded by the dashed circle. The subgraphs surrounded by the gray circles are the subgraph G_v rooted at vertex v , and the block $B_{dwn}(u)$ whose root vertex is u . The counter vertex of $B_{dwn}(u)$ is u_c . The block surrounded by the dashed circle (i.e., the center of G_v) is the block $B_{up}(u)$. (b) The BC tree of G_v , where B-nodes are drawn as white circles, C-nodes are drawn as empty circles, and v^* is the centroid of the BC tree.

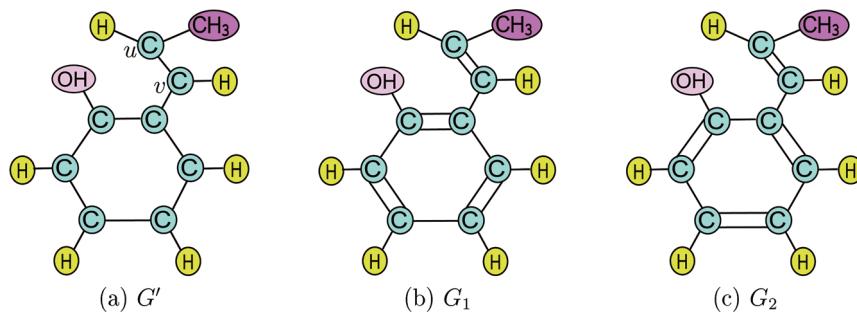


Figure 3. (a) A simple graph G' . (b and c) Two different chemical graphs $G_1, G_2 \in \mathcal{M}(G')$.

simple graph G' represents a collection of chemical graphs $G \in \mathcal{M}(G')$. For example, the two multigraphs G_1 in Figure 3b and G_2 in Figure 3c belong to $\mathcal{M}(G')$ for the simple graph G' in Figure 3a. Notice that there may be an edge e in G' such that the multiplicity of edge e is common to all chemical graphs in $\mathcal{M}(G')$. For example, if e is a bridge in G' , then the multiplicity of e in a chemical graph $G \in \mathcal{M}(G')$ is determined uniquely. In Figure 3a, the bridge between vertices u and v denoting carbon atoms always has a multiplicity of 2 in any multigraph $G \in \mathcal{M}(G')$. It is also possible that the multiplicity of an edge e in G' takes distinct values in different chemical graphs in $\mathcal{M}(G')$. For example, if a compound contains a benzene ring, as shown in Figure 3, then the multiplicities of ring edges are not unique in our multigraph representation of chemical graphs.

In this paper, we take into account the multiplicity of bridges, ignoring the multiplicity of edges within blocks, and focus on stereoisomers that arise only from the structure of simple graphs together with multiplicities of bridges. We take such an approach in order to establish the first mathematical formulation of a certain class of stereoisomers before we investigate more complicated definitions of stereoisomers. In this paper, for a given multigraph G , we show how to generate all of the stereoisomers of G . Note that, in our definition, two chemical graphs G_1 and G_2 will have the same set of stereoisomers if they induce the same simple graph G' , i.e., $G_1, G_2 \in \mathcal{M}(G')$.

The formal definition of the problem considered is given as follows:

- Input: An outerplanar multigraph $G = (V, E)$ and a vertex number $n: V \rightarrow \{1, 2, \dots, |V|\}$, by which each vertex is numbered from 1 to $|V|$.
- Output: All of the stereoisomers of G that can be generated by asymmetry around carbon atoms that are not in rings (the exact

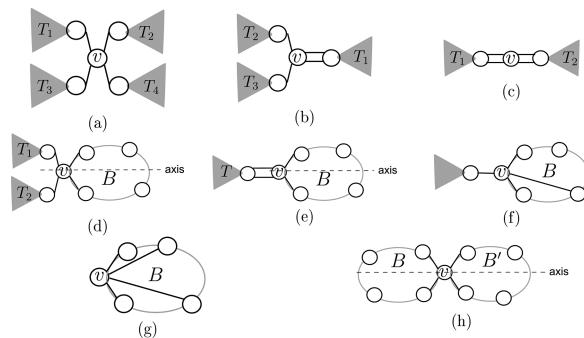


Figure 4. Patterns around v giving rise to two different three-dimensional arrangements.

definition of stereoisomers in this paper and its sketch will be given in the sections Formal Definition of Stereoisomers and Sketch of Our Definition of Stereoisomers, respectively).

In the following, a graph means a multigraph unless otherwise stated.

Stereoisomer Modeling. Informally, the stereoisomers in this paper are generated by two different three-dimensional arrangements around a carbon atom v satisfying one of the following seven cases labeled (1-i) to (1-iii) and (2-i) to (2-iv), where it is allowed that substructures (T_1, \dots, T_4) contain cycles:

- (1) For $v \in V_C$ which is not in a block,
 - (1-i) v is adjacent to four different substructures (T_1, \dots, T_4) in Figure 4a).
 - (1-ii) v is adjacent to both a substructure T_1 with a double bond and two different substructures T_2 and T_3 with

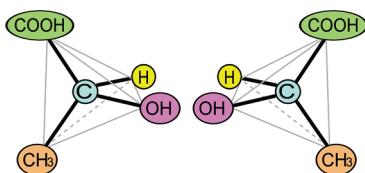


Figure 5. Three-dimensional arrangements around the asymmetric carbon atom in lactic acid. There are two different three-dimensional arrangements around the asymmetric carbon atoms (carbon atom at the center of the tetrahedron).

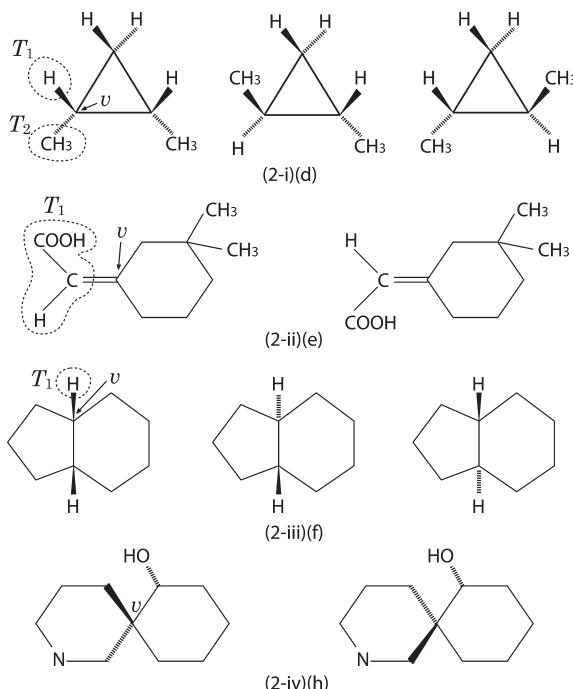


Figure 6. Stereoisomers caused by asymmetric carbon atoms corresponding to cases (d–f) and (h) of Figure 4.

single bonds, and T_1 is asymmetric along the double bond (Figure 4b).

- (1-iii) ν is adjacent to two substructures T_1 and T_2 with double bonds, each asymmetric along the double bond (Figure 4c).
- (2) For $\nu \in V_C$ which is in a block B ,
 - (2-i) ν is adjacent to two different substructures T_1 and T_2 with single bonds; B is asymmetric along the axis passing through ν and disjoint with T_1 and T_2 (Figure 4d).
 - (2-ii) ν is adjacent to a substructure T with a double bond; B and T are asymmetric along the axis passing through ν , and B is disjoint with T (Figure 4e).
 - (2-iii) ν is adjacent to three or four different atoms in B with single bonds (Figure 4f and g, respectively).
 - (2-iv) ν is a cutvertex between B and another block B' , which are asymmetric along the axis passing through ν (see Figure 4h).

For example, the asymmetric carbon atom in lactic acid (Figure 5) satisfies case (1-i), resulting in two different three-dimensional arrangements around the carbon atom, and (1-ii) is case of cis–trans stereoisomers, such as terpenes. For other examples, the asymmetric

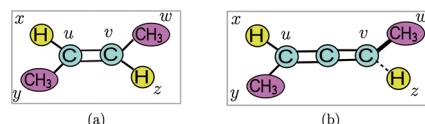


Figure 7. Three-dimensional arrangements around a chain of double bonds between two carbon atoms u and v . Each rectangle shows the plane that contains the two left substituents x and y . Thick lines indicate the edges on the front side of the plane, and dashed lines indicate the edges on the back side of the plane. In Figure (a), two carbon atoms u and v and the double bond between them form a carbon circuit.

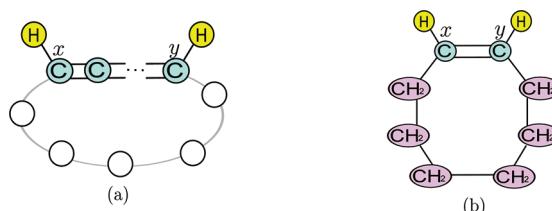


Figure 8. (a) An example of a chain of double bonds between two carbon atoms in a cycle. (b) Cyclooctene, an example for our definition of configuration.

carbon atoms satisfying conditions (2-i) to (2-iv) are shown in Figure 6, where some structures or stereoisomers may not be stable, and we could not find a realistic example for case (g) in Figure 4 (though it is graph theoretically possible). It is to be noted that in Figure 6, three stereoisomers are shown for cases (d) and (f) because there exist other asymmetric carbon atoms.

In the three-dimensional space, we assume that a double bond between two carbon atoms graphically consists of two distinct edges, as shown in Figure 7. Thus, a carbon atom ν with three adjacent atoms still has four edges connected to ν if it is adjacent to another carbon atom with a double bond. In this paper, we make the following assumption on the three-dimensional arrangement of a chain of double bonds between two carbon atoms u and v , such that u is adjacent to two atoms x and y by single bonds and v is adjacent to two atoms w and z by single bonds, as shown in Figure 7: For the number k of double bonds between u and v : (i) x, y, w , and z are on the same plane when k is odd; and (ii) x, y, w , and z are not on the same plane when k is even.

For example, Figure 7a and b illustrates the chains of k double bonds for $k = 1$ and 2, respectively.

As already stated, any multiplicity in a block will be ignored in our definition of stereoisomers because it would be very complicated to give a mathematical definition and develop a dynamic programming algorithm for covering the multiplicity in a block, and some stereoisomers caused by a double bond in a ring might not be stable if the length of the ring is very small. (A more complicated definition of stereoisomers is treated in another one of our ongoing preliminary studies¹⁸ wherein the length of a cycle is taken into account to determine whether distinct stereoisomers arise around the cycle or not). Hence even if the double bonds between two carbon atoms x and y in a block of a chemical graph are uniquely determined, as in the Figure 8a example, we assume that there is only one three-dimensional arrangement around x and y .

To represent a three-dimensional relationship among the substructures around a carbon atom ν , we define “configuration” around a carbon atom ν as follows: Suppose that the substructures around ν give four graphs $S_j, j = 1–4$ in a graph G such that

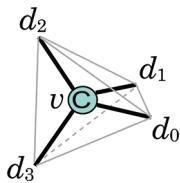


Figure 9. The four directions d_0, d_1, d_2 , and d_3 around a carbon atom v in three-dimensional space.

any two of these graphs share no vertex other than v (where v is called the root of each S_j) and that no two of these graphs are isomorphic under any correspondence such that the roots of both graphs match each other. Thus, v is a typical case of a carbon stereocenter. We introduce four directions d_i ($i = 0-3$) around a carbon atom v , as shown in Figure 9, and let $e_j = (v, u_j)$, $j = 1-4$ denote the edge joining the carbon atom v and its neighbor u_j in the graph S_j . A configuration around v is a correspondence between directions d_i ($i = 0-3$) and edges (v, u_j) , $j = 1-4$, where there are 24 possible assignments and we do not distinguish two correspondences if one can be obtained from the other by shifting the indices of d_i according to a rotation of the tetrahedron. For example, we regard that $\{(d_0, e_1), (d_1, e_2), (d_2, e_3), (d_3, e_4)\}$ and $\{(d_2, e_1), (d_0, e_2), (d_1, e_3), (d_3, e_4)\}$ represent the same configuration, since the second one can be obtained from the first one by the rotation $d_0 \rightarrow d_2 \rightarrow d_1 \rightarrow d_0$ along direction d_3 . On the other hand $\{(d_0, e_1), (d_1, e_2), (d_2, e_3), (d_3, e_4)\}$ and $\{(d_0, e_1), (d_1, e_2), (d_2, e_4), (d_3, e_3)\}$ represent different configurations, and there is no configuration other than these two. We define configuration analogously as long as no two of S_j , $j = 1-4$ are stereochemically isomorphic. The definition of configuration is then extended to the case where some two edges $e_i = (v, u_i)$ and $e_j = (v, u_j)$ belong to the same block or form a double bond (i.e., $u_i = u_j$). For example, each of the carbon atoms x and y in cyclooctene (Figure 8b) has only one configuration. The definition of stereoisomorphism (isomorphism of stereoisomers) will be given in the section Isomorphism on $\{+, -\}$ -Labeled Graphs, and the exact relationship between configurations and stereoisomers will be examined in the section Configuration around Each Carbon Atom Corresponding to Its Label.

Structural Isomorphism of Outerplanar Graphs. The sets of vertices and edges of a graph G are denoted by $V(G)$ and $E(G)$, respectively. Each connected graph G has a unique block-cutvertex tree (BC tree), in which each 2-connected component (block or bridge) of G is called a B-node, and each cutvertex is called a C-node. There is an edge between B-node v and C-node u if and only if u belongs to the block corresponding to v in G (see Figure 2). Every leaf is a B-node, and there is no edge between two B-nodes or between two C-nodes. It is known that all of the 2-connected components of a graph G and its BC tree can be obtained in $O(|V| + |E|)$ time.¹⁹ The BC tree of G is denoted by $BC(G)$.

We define the centroid of a BC tree using the following result which specifies a structurally unique vertex or edge. That is, this theorem states that every BC tree has either a center vertex or a center edge.

Theorem 1 (Jordan's Theorem).²¹ For any tree of $n \geq 1$ vertices, exactly one of the next two statements holds: (i) There exists a unique vertex v^* such that each of the subtrees obtained by removing v^* contains at most $(n - 1)/2$ vertices, and (ii) there exists a unique edge e^* such that each of the two subtrees obtained by removing e^* contains $n/2$ vertices.

Such a vertex v^* and an edge e^* are called the unicentroid and bicentroid of the tree, respectively. We call the unicentroid and

bicentroid the centroid of the tree. The root of a tree is defined by the vertex or vertices in its centroid. A BC tree must have a unicentroid. The unicentroid of the BC tree in Figure 2b is denoted by v^* . We define the center of G to be the set of the vertices which belong to the B- or C-node corresponding to the unicentroid when G is expressed as its BC-tree. For each vertex $v \in V$, we define the depth of v to be the distance from the unicentroid to the closest B-node whose biconnected component contains v in the BC tree. For example, the depth of vertices u, v , and w in Figure 2a is 0, 2, and 4, respectively (note that white circles are B-nodes in Figure 2b). We define the children of a vertex $v \in V$ to be the vertices which belong to the same 2-connected components as v and whose depth are greater than v . For example, vertex v in Figure 2a has three children w_1, w_2 , and w_3 . We denote the set of the children of v by $Ch(v)$. For each vertex $v \in V$, we define the graph G_v rooted at v to be the graph induced from G by v and all descendants of v . For example, Figure 2a illustrates the subgraph G_v for a vertex v . Every block B of an outerplanar graph has a unique Hamilton cycle $H(B)$. A linear time algorithm to find the Hamilton cycle of a 2-connected outerplanar graph is known.¹⁵

In the following, we consider that a bridge corresponding to a double bond between two vertices $x, y \in V_C$ forms a Hamilton cycle, which we call a carbon circuit (see Figure 7a). We call blocks and carbon circuits cyclic structures. Among cyclic structures containing a vertex v which does not belong to the center, we denote the one nearest to the center by $B_{up}(v)$ (if any) and the other one by $B_{dwn}(v)$ (if any). Furthermore, for a vertex v which belongs to the center, if the center does not consist of v , among cyclic structures containing v , we denote the one corresponding to the unicentroid B-node by $B_{up}(v)$ and the other one by $B_{dwn}(v)$ (if any) (see Figure 2a). For a noncentral block/bridge B , we define the root vertex v_r of B to be the vertex $v \in V(B)$ with smallest depth (see Figure 2a), and we define the graph G_B rooted at B to be the graph induced from G by $V(B)$ and all descendants of $V(B) \setminus \{v_r\}$. When $|V(B)| = m > 2$ is even, let $V(B) = \{v_r, v_1, \dots, v_{m-1}\}$, where v_r, v_1, \dots, v_{m-1} appear along $H(B)$ in this order, and we define the counter vertex of B to be the vertex $v_{m/2}$ (see Figure 2a). If a vertex $v \in V$ is adjacent to four different vertices in $B_{up}(v)$ with single bonds, then v is called an inner spiro atom (see Figure 4g).

Two multigraphs G_1 and G_2 are called isomorphic if their simple graphs $S(G_1) = (V_1, E_1)$ and $S(G_2) = (V_2, E_2)$ have a bijection $\psi: V_1 \rightarrow V_2$, such that (i) $(u, v) \in E_1$ holds if and only if $(\psi(u), \psi(v)) \in E_2$ holds; and (ii) each vertex in G_1 corresponds to a vertex in G_2 of the same atom type in the isomorphism represented by ψ . Two subgraphs G_u and G_v rooted at vertices $u, v \in V(G)$ of a multigraph G are rooted-isomorphic, denoted by

$$G_u \approx_{r} G_v$$

if they admit an isomorphism ψ with $\psi(u) = v$. Two subgraphs G_B and $G_{B'}$ rooted at blocks B and B' of a multigraph G are rooted isomorphic, denoted by

$$G_B \approx_{r} G_{B'}$$

if they admit an isomorphism ψ with $\psi(v) = v'$, $v' \in B'$ for all $v \in B$.

Indices of subgraphs G_v rooted at vertices $v \in V$ are defined to be non-negative integers $\sigma(v, G_v)$ such that

$$\sigma(v, G_v) = \sigma(u, G_u) \Leftrightarrow G_v \approx_{r} G_u$$

From our previous study,²² we have the following theorem. This theorem means that an index (i.e., integer number) can be

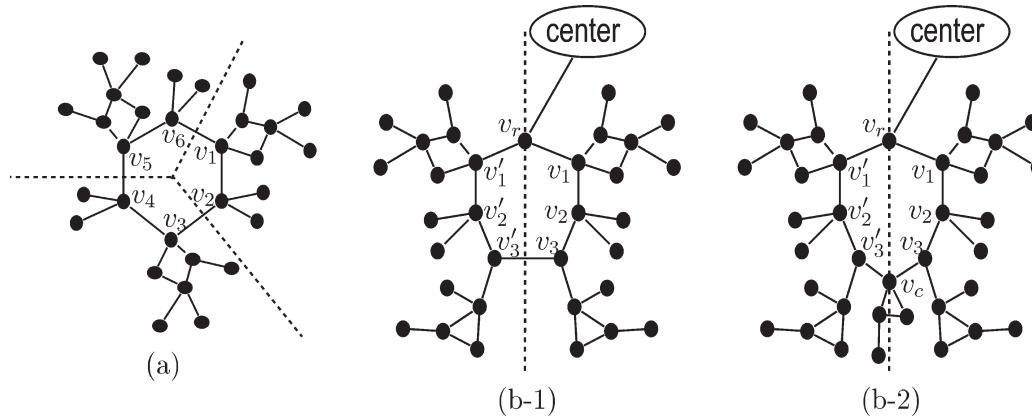


Figure 10. (a) An example of rotational symmetry in a subgraph rooted at the center block B of G . (b) Example of an axial symmetry in a noncenter block B . The broken line represents the symmetric axis. (b-1) $|V(B)|$ is odd, and (b-2) $|V(B)|$ is even.

efficiently computed for every vertex so that two vertices have the same index if and only if the subgraphs rooted at these vertices are isomorphic.

Theorem 2. There is a choice of indices such that the indices of all rooted subgraphs of a given rooted outerplanar simple graph can be computed in linear time and space.

Note that indices $\sigma(\nu, G_\nu)$ are independent of any vertex numbers $n(\nu)$ assigned to a chemical graph. In the rest of this paper, we write $\sigma(\nu, G_\nu)$ as $\sigma(\nu)$ if G_ν is clear from the context.

We formally define structural symmetry and orientation of the Hamilton cycle of each block in outerplanar graphs in Section S3, Supporting Information. Informally, a block B of a rooted outerplanar graph G is defined to have a rotational (respectively, axial) symmetry if a subgraph G_B has an outerplane drawing which displays the rotational (respectively, axial) symmetry in a geometric sense, if we ignore the structure of the subgraph G_{ν_c} rooted at the counter vertex ν_c of B . See Figure 10 for an example. By definition, a subgraph G_B rooted at noncentral block B never has a rotational symmetry. Informally, if a block B of a rooted outerplanar graph G has no symmetry, then we define the topological orientation $O_t(B)$ of the Hamilton cycle $H(B)$ by the structure of G_B . If a block B of a rooted outerplanar graph G has a rotational symmetry, then we define the topological orientation $O_t(B)$ by the vertex numbering. By the definitions and the result of a previous study by Manning and Atallah,²³ we have the following lemma, which means that the rotational symmetry (i.e., the minimum rotation by which two drawings can be superposed) of the root block and the axis symmetries (i.e., all symmetric axes) for all blocks can be computed efficiently.

Lemma 3. We can compute the symmetries of all subgraphs rooted at blocks in a given simple graph $G = (V, E)$ in $O(|V|)$ time and space.

Sketch of Our Definition of Stereoisomers. In the previous sections, we have seen the properties of an outerplanar graph, such as index and symmetry. In the section Formal Definition of Stereoisomers, we define a stereoisomer and examine some properties of it. In this section, for readers not interested in detailed definitions, we see just a sketch of our definition of stereoisomers.

At first, we see that a stereoisomer in this paper is defined recursively along the BC tree of G from the leaves to the root by showing some examples given in Figure 11. First, we consider an example shown in Figure 11a. Informally, each carbon atom

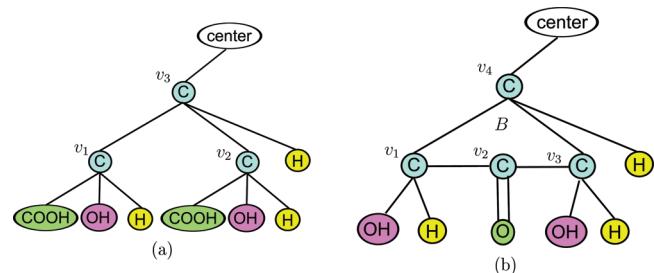


Figure 11. Examples of rooted subgraphs for showing a sketch of the proposed algorithm.

generates two different configurations when it has an asymmetric structure. In particular, any carbon atom that does not belong to any cyclic structure generates two different configurations if its four adjacent substructures are ‘different.’ Then ν_1 in Figure 11a always generates two different configurations because graph structures of all of its children are different. Similarly, ν_2 in Figure 11a always generates two different configurations. By contrast, ν_3 may seem to generate only one configuration because it has two children ν_1 and ν_2 , such that G_{ν_1} and G_{ν_2} are rooted isomorphic. Actually, if the three-dimensional arrangements of G_{ν_1} and G_{ν_2} are identical, then ν_3 generates only one configuration. However, ν_3 generates two different configurations if the three-dimensional arrangements of G_{ν_1} and G_{ν_2} are different. That is to say, as shown in this example, the number of three-dimensional arrangement of substituents around each carbon atom depends on the three-dimensional arrangements of its descendants. It is to be noted that ν_3 generates four different configurations in total (see also Figure 29).

We consider another example shown in Figure 11b. If a carbon atom has a block $B_{up}(\nu)$ but does not have a $B_{dwn}(\nu)$, then it generates two different three-dimensional arrangements if it has two different children and $B_{up}(\nu)$ does not have any symmetric axes passing through ν . Then ν_1 and ν_3 in Figure 11b always generate two different configurations, respectively. If a carbon atom ν has a block $B_{dwn}(\nu)$ but does not have a $B_{up}(\nu)$, then it generates two different arrangements if it has a child not in $B_{dwn}(\nu)$ and $B_{dwn}(\nu)$ does not have any symmetric axes passing through ν . Then, similar to the example above, ν_4 generates two different configurations if the three-dimensional arrangements

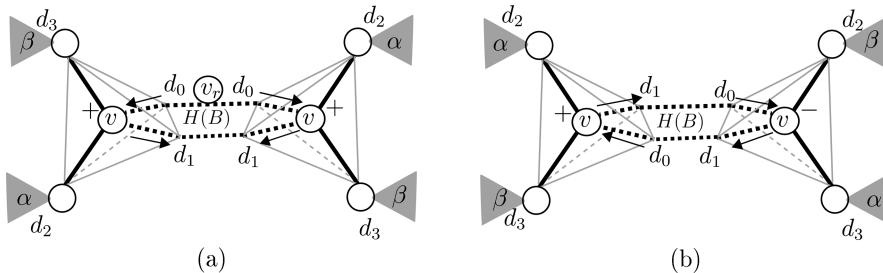


Figure 12. Configurations around a vertex v_0 . The dotted cycle shows Hamilton cycle $H(B)$. B is identical to ‘reverse.’ (a) G_B has no rotational symmetry, and v_r is defined by graph by structures of G_B . (b) G_B has a rotational symmetry.

of G_{v_1} and G_{v_3} are different, although it generates only one configuration if they are identical. Again, v_4 generates four different configurations in total.

Next, we present a sketch of our definition of stereoisomers. To define stereoisomers of an outerplanar graph G , we first introduce a label $l(v)$ for each carbon atom $v \in V_C$, where $l(v)$ takes one of $+$, $-$, or nil ($l(v) = \text{nil}$ means that v has a unique configuration around v). For every vertex $v \in V_O \cup V_N \cup V_H$ $l(v) = \text{nil}$. We introduce a total order among these labels by setting:

$$+ > - > \text{nil}$$

We next introduce a representation I of a given graph G as a set of pairs of vertex numbers $n(v)$ and labels $l(v)$ over all vertices $v \in V$. That is

$$I = \{(n(v), l(v)) | v \in V\}$$

Let $\mathcal{R}(G)$ denote the set of all representations I of G , where $|\mathcal{R}(G)| = 3^{|V_G|}$ holds. Similarly, for each vertex $v \in V$, we introduce a representation I_v of the rooted subgraph G_v as

$$I_v = \{(n(u), l(u)) | u \in V(G_v)\}$$

Let $\mathcal{R}(G_v)$ denote the set of all representations I_v of a rooted graph G_v . For $Ch(v) = \{x_1, \dots, x_k\}$, I_v is written by

$$I_v = \{(n(x_i), l(x_i))\} \cup I_{x_1} \cup \dots \cup I_{x_k}$$

where $I_{x_i} \in \mathcal{R}(G_{x_i})$ ($i = 1, \dots, k$). Similarly, for each block/bridge B of G , we define a representation I_B of the subgraph rooted at B as

$$I_B = \{(n(u), l(u)) | u \in V(G_B)\}$$

and let $\mathcal{R}(G_B)$ denote the set of all representations I_B of a rooted graph G_B .

For a given outerplanar graph G and a representation $I \in \mathcal{R}(G)$, a $\{+, -\}$ -labeled graph (G, I) is defined to be a graph such that each vertex $v \in V(G)$ receives labels according to I . As will be shown in the section Proper Representations, only representations which satisfy a certain condition, called proper representations, define stereoisomers in this paper.

Let $\mathcal{T}(G)$ and $\mathcal{T}(v)$ denote the maximal set of proper representations in which no two representations are stereochemically equivalent for an outerplanar chemical graph G and a subgraph G_v rooted at a vertex $v \in V(G)$ of B . $f^*(G) := |\mathcal{T}(G)|$ and $f(G, v) := |\mathcal{T}(v)|$. See Definitions 13, 14, and 15 for the detail of the definitions. The number of stereoisomers which we count is $f^*(G)$, and all of the elements in $\mathcal{T}(G)$ are the stereoisomers which we enumerate. In the section Algorithm, we give an algorithm that outputs each element I of $\mathcal{T}(G)$ without duplication.

■ FORMAL DEFINITION OF STEREOISOMERS

This section gives a formal definition of stereoisomers considered in this paper. Readers not interested in precise definitions can skip this and the next sections and go to the section Fundamental Idea of Our Enumeration Algorithm.

To define stereoisomers of an outerplanar multigraph G , we use the notations introduced in the end of the previous section: a label $l(v)$ for each carbon atom, a representation I of a given graph G , a representation I_v of the rooted subgraph G_v , and a representation I_B of the subgraph rooted at block/bridge B .

Isomorphism on $\{+, -\}$ -Labeled Graphs. As explained in the previous section, for a given outerplanar multigraph G and a representation $\mathcal{T} \in \mathcal{R}(G)$, a $\{+, -\}$ -labeled graph (G, \mathcal{T}) is defined to be a graph such that each vertex $v \in V(G)$ receives labels according to I , and only representations which satisfy a certain condition, called proper representations, define stereoisomers. In a $\{+, -\}$ -labeled graph represented by a proper representation, informally, the correspondence between labels $+$ or $-$ of a vertex v in some block B and its spatial configuration is defined as follows: (i) If G_B has no rotational symmetry, then the root vertex v_r of B is defined by the graph structure of G_B . For the four directions d_i , $i = 0-3$ from a carbon atom $v \in V_C$ as in Figure 9, we assume that the edge of $H(B)$ heading for v_r is in direction d_0 and that another edge of $H(B)$ is in direction d_1 . If the substructure α in direction d_2 and β in d_3 are asymmetric, then v receives a label $+$ or $-$. We define one configuration such that α is in d_2 and β is in d_3 to correspond to the label $+$ and vice versa for the label $-$ (β is in d_2 and α is in d_3). As shown in Figure 12a, if the three-dimensional arrangement of G_B is identical with its reverse, then the two labels of a vertex and its counterpart are identical. (ii) If G_B has a rotational symmetry, then the root vertex of B is defined only by the vertex numbering, not by the graph structure of G_B . The topological orientation $O_t(B)$ is defined by the vertex numbering. For the four directions d_i , let v_0 and v_1 be the vertices in directions d_0 and d_1 , respectively. We assume that $O_t(B)$ is defined to be $v_0 \rightarrow v \rightarrow v_1$. If the substructure α in direction d_2 and β in d_3 are asymmetric, then v receives a label $+$ or $-$. We define one configuration such that α is in d_2 and β is in d_3 to correspond to the label $+$ and vice versa for the label $-$ (β is in d_2 , and α is in d_3). As shown in Figure 12b, if the three-dimensional arrangement of G_B is identical with its reverse, then two labels of a vertex and its counterpart are different.

From the correspondence between labels and configurations, we define the isomorphism on $\{+, -\}$ -labeled graphs as follows.

Definition 4. Two subgraphs $(G_v, I_v), I_v \in \mathcal{R}(G_v)$ and $(G_u, I_u), I_u \in \mathcal{R}(G_u)$, respectively, rooted at vertices $v, u \in V(G)$ of a

$\{+, -\}$ -labeled graph (G, I) , $I \in \mathcal{R}(G)$ are called rooted isomorphic if there is a bijection $\psi: V((G_o, I_o)) \rightarrow V((G_u, I_u))$ such that: (i) $(\nu_1, \nu_2) \in E((G_o, I_o))$, if and only if $(\psi(\nu_1), \psi(\nu_2)) \in E((G_u, I_u))$; (ii) the types of atoms of $\omega \in V((G_o, I_o))$ and $\psi(\omega) \in V((G_u, I_u))$ are identical; (iii) the labels $l(\omega)$ of atoms of $\omega \in V((G_o, I_o))$ and $l(\psi(\omega))$ of $\psi(\omega) \in V((G_u, I_u))$ are identical; and it holds that $\psi(\nu) = u$.

In order to define an isomorphism on the entire $\{+, -\}$ -labeled graph, first we define the l -reverse drawing of a given outerplane drawing of a $\{+, -\}$ -labeled graph (G, I) , $I \in \mathcal{R}(G)$ with its central block B and an axis l as follows.

Definition 5. Let δ denote an outerplane drawing of a $\{+, -\}$ -labeled graph (G, I) , $I \in \mathcal{R}(G)$ with its central block B and an axis l . We define the l -reverse drawing of a given drawing δ as follows:

- (Case 1) G_B has no rotational symmetry: we define the l -reverse drawing to be the drawing obtained by reversing positions of all vertices around l in a geometric sense in the drawing δ .
- (Case 2) G_B has a rotational symmetry: we define the l -reverse drawing to be the drawing made by the following two steps:
 - Step 1: Reverse positions of all of the vertices except for the descendants of vertices through which the straight line l passes in a geometric sense in the drawing δ .
 - Step 2: Replace all labels of $\nu \in V(B)$ by

$$l(\nu) := \begin{cases} l(\nu), & \text{if } \nu \text{ is an inner spiro atom or } l(\nu) = \text{nil} \\ +, & \text{if } l(\nu) = - \\ -, & \text{if } l(\nu) = + \end{cases}$$

Then we define isomorphisms of $\{+, -\}$ -labeled graphs as follows.

Definition 6. Two $\{+, -\}$ -labeled graphs (G, I_1) and (G, I_2) , $I_1, I_2 \in \mathcal{I}(G)$ of a multigraph G are called isomorphic if there are some outerplane drawings δ_1 of (G, I_1) and δ_2 of (G, I_2) such that: (i) If the topological-orientation $O_t(B)$ of the central block B is defined, then the orientations $O_t(B)$ in δ_1 and δ_2 are clockwise; and (ii) δ_1 is identical to δ_2 or there exists some axis l such that δ_1 is identical with the l -reverse drawing of δ_2 .

Asymmetry Around a Carbon Atom. In a $\{+, -\}$ -labeled graph (G, I) , we define an asymmetry around each carbon atom $\nu \in V_C$. Note that a cyclic structure is a block or a carbon circuit, as defined in the section Structural Isomorphism of Outerplanar Graphs, and each carbon atom $\nu \in V_C$ belongs to at most two cyclic structures. First, we define an axial asymmetry of a rooted subgraph (G_B, I_B) , $I_B \in \mathcal{R}(G_B)$ as follows.

Definition 7. The subgraph (G_B, I_B) , $I_B \in \mathcal{R}(G_B)$ rooted at a cyclic structure B with its root vertex ν_r of a $\{+, -\}$ -labeled outerplanar graph is defined to have an axial asymmetry along an axis l if one of the following conditions is satisfied:

- (i) There is no cf-drawing of (G_B, I_B) which displays axial symmetry with an axis passing through ν_r in a geometric sense.
- (ii) There is some cf-drawing of (G_B, I_B) which displays axial symmetry with an axis l passing through ν_r in a geometric sense (we define such a subgraph to be cf-symmetric), and there exists the counter vertex ν_c of B which satisfies one of the following conditions:
 - (ii-1) ν_c is adjacent to two substructures T_1 and T_2 that are not rooted isomorphic to each other, where T_i denotes an adjacent substructure of ν_c rooted at an adjacent vertex ν_i which does not belong to B (Figure 13a).

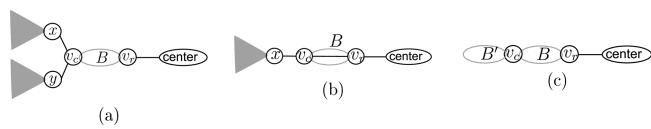


Figure 13. Graph structures around the counter vertex ν_c .

- (ii-2) ν_c has four adjacent vertices, and just one adjacent vertex does not belong to B (Figure 13b).
- (ii-3) ν_c is the root vertex of another cyclic structure B' such that a subgraph $(G_{B'}, I_{B'})$, $I_{B'} \in \mathcal{R}(G_{B'})$ has an axial asymmetry along an axis l passing thorough ν_c (Figure 13c).

Note that an axial asymmetry of a subgraph (G_B, I_B) , $I_B \in \mathcal{R}(G_B)$ is defined by considering the condition (ii-3) recursively.

Definition 8. We define a carbon atom $\nu \in V_C$ in a $\{+, -\}$ -labeled graph (G, I) to have an asymmetry if ν satisfies the following condition:

- (Case 1) If ν does not belong to any cyclic structures, then is adjacent to four substructures T_i , $i = 1-4$ such that no two of substructures T_i , $i = 1-4$ are rooted isomorphic, where T_i denotes an adjacent substructure of ν rooted at a vertex ν_i adjacent to ν (Figure 4a).
- (Case 2) If ν belongs to just one cyclic structure B , then is either an inner spiro atom (Figure 4g) or the following two conditions hold:
 - (i) (G_B, I_B) , $I_B \in \mathcal{R}(G_B)$ does not have an axial symmetry along an axis l passing through ν when regarding ν as the root vertex of B .
 - (ii) ν has four incident edges (note that two of the edges form a part of $H(B)$) and it satisfies one of the following conditions:
 - (ii-1) There is exactly one adjacent vertex of which does not belong to B (Figure 4f).
 - (ii-2) ν is adjacent to two substructures T_i , $i = 1, 2$ that are not rooted isomorphic to each other, where T_i denotes an adjacent substructure of ν rooted at an adjacent vertex ν_i which does not belong to B (Figure 4b and d).
- (Case 3) If ν belongs to two cyclic structures B and B' (see Figures 4c, e, and h), then both (G_B, I_B) , $I_B \in \mathcal{R}(G_B)$ and $(G_{B'}, I_{B'})$, $I_{B'} \in \mathcal{R}(G_{B'})$ have axial asymmetries along axes passing through ν by regarding ν as the root vertex of B and B' .

When considering asymmetry around a carbon atom ν , the symmetry of each cyclic structure containing ν is detected recursively.

Treatment of Three-Dimensional Arrangements of Inner Spiro Atoms. In Definitions 7 and 8, we define an inner spiro atom $\nu \in V_C$ to have an asymmetry, although a block B such that its counter vertex ν_c is an inner spiro atom is defined to be symmetric if B has a symmetric cf-drawing. Instinctively, it may seem that such a block B has asymmetry, though we do not define it as such. Here we show the reason why we do not define such a block B to have asymmetry.

Figure 14 shows an example of three-dimensional arrangements of a block such that its counter vertex ν is an inner spiro atom and such that it has a symmetric cf-drawing. As three-dimensional arrangement around ν forms a regular tetrahedron, these two three-dimensional arrangements are regarded as the same one. It means that the block is symmetric. By contrast, as in the example in Figure 15, if the length of the Hamilton cycle of

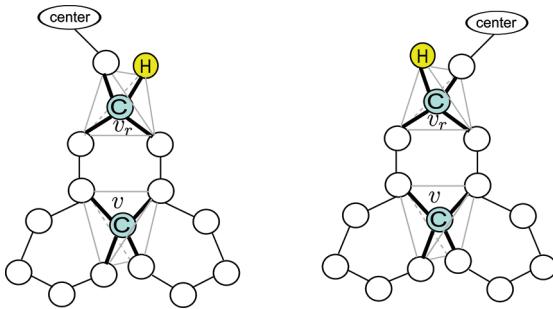


Figure 14. An example of three-dimensional arrangements of a block such that counter vertex ν is an inner spiro atom, and it has a symmetric cf-drawing. The three-dimensional arrangement around ν forms a regular tetrahedron.

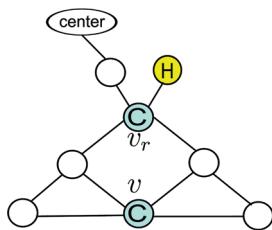


Figure 15. An example of a block such that its counter vertex ν is an inner spiro atom. It has a symmetric cf-drawing, and the three-dimensional arrangement around ν forms a twisted tetrahedron.

the block is short, then the three-dimensional arrangement around the inner spiro atom forms a twisted tetrahedron. It means that the block is asymmetric.

As can be seen, the symmetry of a block B such that its counter vertex ν_c is an inner spiro atom and B has a symmetric cf-drawing depends on the size of the block. For simplicity, we define that such a block B has an asymmetry in any case.

Proper Representations. We define *proper* representations as follows.

Definition 9. A representation $I \in \mathcal{R}(G)$ is called proper if the label $l(\nu)$ of each carbon atom $\nu \in V_C$ satisfies the following conditions:

- (Case 1) If ν has an asymmetry in the $\{+, -\}$ -labeled graph (G, I) , then it holds $l(\nu) \in \{+, -\}$.
- (Case 2) If ν does not have any asymmetry in the $\{+, -\}$ -labeled graph (G, I) , then it holds $l(\nu) = \text{nil}$.

Definition 10. A representation $I_\nu \in \mathcal{R}(G_\nu)$ is called proper if it is a subset of a proper representation $I \in \mathcal{R}(G)$.

Only asymmetric carbon atoms generate two different configurations, and then a proper representation has enough information for representing each three-dimensional arrangement of a given multigraph G . However, a proper representation still has redundant information. For example, in the chain of double bonds shown in Figure 7b, there are three carbon atoms with asymmetry. However, there exists only two three-dimensional arrangements around this carbon chain, whereas there exists 2^3 combinations of labels of them. In general, in a chain of asymmetric but cf-symmetric cyclic structures, there exists only two three-dimensional arrangements, although all cut and end vertices of the chain have asymmetry. See Figure 16 for an example. Then we define simplified proper representations as follows.

Definition 11. A proper representation $I \in \mathcal{R}(G)$ is called simplified proper if for each of the chains of asymmetric but cf-symmetric cyclic structures, only the root vertex ν of the nearest

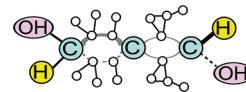


Figure 16. An example of chain of asymmetric but cf-symmetric cycle structures. There exists three asymmetric carbon atoms, although only two three-dimensional arrangements around the chain $C-C-C$.

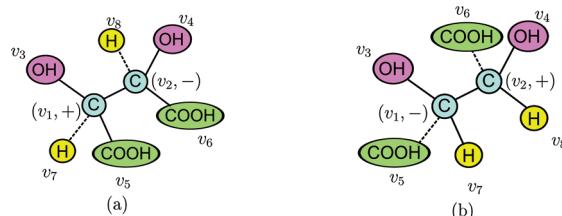


Figure 17. An example of the compound that has distinct proper representations which are stereoisomorphic.

cyclic structure to the center receives the label $l(\nu) \in \{+, -\}$, and the other cut vertices between two cyclic structures in the chain and the end vertices of the chain receive the labels nil.

As will be discussed in the section Configuration around Each Carbon Atom Corresponding to Its Label, a simplified proper representation $I_\nu \in \mathcal{R}(G_\nu)$ realizes a set of configurations around carbon atoms in G_ν and is considered as a rooted stereoisomer of G_ν . Similarly we consider a simplified proper representation $I \in \mathcal{R}(G)$ to be a stereoisomer of G .

Definition 12. Two simplified proper representations $I, I' \in \mathcal{R}(G)$ [respectively, $I_\nu \in \mathcal{R}(G_\nu)$ and $I_u \in \mathcal{R}(G_u)$] are defined to be stereoisomorphic (respectively, rooted stereoisomorphic) if and only if they represent isomorphic (respectively, rooted isomorphic) $\{+, -\}$ -labeled graphs. If two simplified proper representations $I_\nu \in \mathcal{R}(G_\nu)$ and $I_u \in \mathcal{R}(G_u)$ are rooted stereoisomorphic, then we write this as

$$I_\nu \approx I_u$$

However, two simplified proper representations $I \in \mathcal{R}(G)$ and $I' \in \mathcal{R}(G)$ may be isomorphic. In the next section, we determine one isomorphic simplified proper representation as the “canonical form” of the corresponding stereoisomer.

Canonical Form of Simplified Proper Representations. Here we consider the example given in Figure 17, where each of COOH and OH is regarded as a single vertex for simplicity. For the graph G in Figure 17, we write the vertex ν whose vertex number is $n(\nu) = i$ as ν_i ; ν_1 and ν_2 are the bicentroid of G , and there are representations $I_a, I_b \in \mathcal{R}(G)$ with

$$I_a = \{(1, +), (2, -), (3, \text{nil}), (4, \text{nil}), (5, \text{nil}), (6, \text{nil}), (7, \text{nil}), (8, \text{nil})\}$$

$$I_b = \{(1, -), (2, +), (3, \text{nil}), (4, \text{nil}), (5, \text{nil}), (6, \text{nil}), (7, \text{nil}), (8, \text{nil})\}$$

(Figure 17a and b, respectively). Figure 17 shows

$$G_{\nu_1} \approx G_{\nu_2}, G_{\nu_3} \approx G_{\nu_4}, G_{\nu_5} \approx G_{\nu_6}, G_{\nu_7} \approx G_{\nu_8}$$

and no two of $G_{\nu_1}, G_{\nu_3}, G_{\nu_5}$, and G_{ν_7} are rooted isomorphic. Then we assume that

$$\begin{aligned} s_1 &= \sigma(\nu_1) = \sigma(\nu_2) > s_2 = \sigma(\nu_3) = \sigma(\nu_4) > s_3 \\ &= \sigma(\nu_5) = \sigma(\nu_6) > s_4 = \sigma(\nu_7) = \sigma(\nu_8) \end{aligned}$$

for integers $s_i, i = 1-4$. Note that I_a and I_b are distinct as sets, although I_a and I_b are stereoisomorphic as can be seen in Figure 17.

Then we define the canonical form of simplified proper representations $I \in \mathcal{R}(G)$ as follows.

Definition 13. Let $L(I)$ (respectively, $L(I_u)$) be a nondecreasing sequence of the elements $(n(\nu), l(\nu))$ in a set $I = \{n(\nu), l(\nu)\} \in V\}$ (respectively, $I_u = \{(n(\nu), l(\nu))|\nu \in V(G_u)\} \in \mathcal{R}(G_u), u \in V$) according to the given numbering of the vertices in V . Note that $L(I)$ (respectively, $L(I_u)$) now reflects the given numbering on the vertex set V (recall that stereoisomorphism is determined independently of the vertex numbering): (i) The canonical form of all simplified proper representations in $\mathcal{R}(G)$ that are stereoisomorphic is defined to be the simplified proper representation $I \in \mathcal{R}(G)$ with the lexicographically maximum $L(I)$ among them; and (ii) For each vertex $u \in V$, the canonical form of simplified proper representations in $\mathcal{R}(G_u)$ which are rooted stereoisomorphic is defined to be the simplified proper representation $I_u \in \mathcal{R}(G_u)$ with the lexicographically maximum $L(I_u)$ among them.

For the example in Figure 17, we have

$$L(I_a) = [(1, +), (2, -), (3, \text{nil}), (4, \text{nil}), (5, \text{nil}), (6, \text{nil}), (7, \text{nil}), (8, \text{nil})]$$

$$L(I_b) = [(1, -), (2, +), (3, \text{nil}), (4, \text{nil}), (5, \text{nil}), (6, \text{nil}), (7, \text{nil}), (8, \text{nil})]$$

and we define I_a to be the canonical form of these stereoisomorphic representations.

Definition 14. For an outerplanar multigraph $G = (V, E)$, we define the number $f^*(G)$ of stereoisomers of G to be the number of all canonical forms of simplified proper representations in $\mathcal{R}(G)$. Similarly, for each vertex $\nu \in V$, we define the number $f(G, \nu)$ of stereoisomers of G to be the number of all canonical forms of simplified proper representations in $\mathcal{R}(G_\nu)$.

Definition 15. For an outerplanar multigraph $G = (V, E)$, let $\mathcal{I}(G)$ denote a set of simplified proper representations in $\mathcal{R}(G)$ such that $|\mathcal{I}(G)| = f^*(G)$, and no two representations in $\mathcal{I}(G)$ are stereoisomorphic. Similarly, for each vertex $\nu \in V$, let $\mathcal{I}(\nu)$ denote a set of simplified proper representations in $\mathcal{R}(G_\nu)$ such that $|\mathcal{I}(\nu)| = f(G, \nu)$, and no two representations in $\mathcal{I}(\nu)$ are stereoisomorphic. Similarly, for each block B in G , let $\mathcal{I}(B)$ denote a set of simplified proper representations in $\mathcal{R}(G_B)$ such that no two representations in $\mathcal{I}(B)$ are stereoisomorphic.

In the section Algorithm, we give an algorithm that outputs each element I of $\mathcal{I}(G)$ without duplication. The choice of $\mathcal{I}(G)$ and $\mathcal{I}(\nu), \nu \in V$ is determined by the order created for the branching process (see the section *Bijections for Fast Generation*).

■ CONFIGURATION AROUND EACH CARBON ATOM CORRESPONDING TO ITS LABEL

This section describes how the configuration around each carbon atom ν is determined based on its label $l(\nu)$. For that, we define signatures of stereoisomers. For each vertex ν or noncenter block/bridge B of G , the signature $\sigma_s(I_\nu, G_\nu)$ of a representation $I_\nu \in \mathcal{R}(G_\nu)$ or $\sigma_s(I_B, G_B)$ of a representation $I_B \in \mathcal{R}(G_B)$ is defined recursively as follows. Note that the σ_s is independent of the given numbering n of vertices $\nu \in V$.

(i) For a leaf $\nu \in V$, we define

$$\sigma_s(I_\nu, G_\nu) = (\sigma(\nu), l(\nu))$$

(ii) For a representation I_B of the subgraph G_B rooted at a noncenter block/bridge B of G with root vertex ν and

$V(B) = \{\nu, \nu_1, \dots, \nu_k\}$, where ν, ν_1, \dots, ν_k appear along the stereo-orientation $O_s(B)$ (the orientation defined in the latter part of this section which is detected uniquely by the graph structure of G_B and signatures $\sigma_s(I_{\nu_i}, G_{\nu_i})$, $i = 1, 2, \dots, k$). If $O_s(B)$ is not defined, then we use an arbitrary orientation. We can denote $I_B = \{(n(\nu), l(\nu))\} \cup I_{\nu_1} \cup I_{\nu_2} \cup \dots \cup I_{\nu_k}, I_{\nu_i} \in \mathcal{I}(G_\nu)$ ($i = 1, 2, \dots, k$). We define

$$\sigma_s(I_B, G_B) = [\sigma_s(I_{\nu_1}, G_{\nu_1}), \sigma_s(I_{\nu_2}, G_{\nu_2}), \dots, \sigma_s(I_{\nu_k}, G_{\nu_k})]$$

(iii) For a representation I_ν of the subgraph G_ν rooted at a nonleaf vertex $\nu \in V$ with blocks/bridges B_1, B_2, \dots, B_k rooted at ν where $\sigma_s(I_{B_1}, G_{B_1}), \sigma_s(I_{B_2}, G_{B_2}), \dots, \sigma_s(I_{B_k}, G_{B_k})$, $I_{B_i} \in \mathcal{I}(G_{B_i})$ ($i = 1, 2, \dots, k$) are sorted in a lexicographically nondecreasing order, denote $I_\nu = \{(n(\nu), l(\nu))\} \cup I_{B_1} \cup I_{B_2} \cup \dots \cup I_{B_k}$. We define

$$\sigma_s(I_\nu, G_\nu) = [(\sigma(\nu), l(\nu)), \sigma_s(I_{B_1}, G_{B_1}), \sigma_s(I_{B_2}, G_{B_2}), \dots, \sigma_s(I_{B_k}, G_{B_k})]$$

Lemma 16. Two simplified proper representations $I_u \in \mathcal{I}(G_u)$ and $I_\nu \in \mathcal{I}(G_\nu)$ of subgraphs respectively rooted at vertices $u, \nu \in V$ of an outerplanar multigraph $G = (V, E)$ are rooted stereoisomorphic if and only if $\sigma_s(I_u, G_u) = \sigma_s(I_\nu, G_\nu)$ holds.

Proof. Since $\sigma_s(I_\nu, G_\nu)$ is uniquely determined from the graph structure of G_ν and the representation $I_\nu \in \mathcal{I}(G_\nu)$, rooted stereoisomorphic stereoisomers always produce the same value of σ_s . Conversely, since the structure of G_ν and the representation $I_\nu \in \mathcal{I}(G_\nu)$ can be fully recovered from $\sigma_s(I_\nu, G_\nu)$, nonrooted stereoisomorphic stereoisomers always produce different values of σ_s .

Next, we define stereo-orientation $O_s(B)$ of a Hamilton cycle of a cyclic structure B of a $\{+, -\}$ -labeled graph, including a carbon circuit.

We define the root vertex of a central carbon circuit to be the vertex with a smaller vertex number; the counter vertex is the remaining vertex. We define a stereo-orientation $O_s(B)$ of a cyclic structure B only if one of the following cases holds. Otherwise, no stereo-orientation is defined for B . Let ν_r and ν_c denote the root vertex and the counter vertex of B , respectively. (Note that ν_c is defined only when $|V(B)|$ is even).

(Case 1) $O_t(B)$ is defined: we define $O_s(B)$ to be $O_t(B)$.

(Case 2) $O_t(B)$ is not defined and a $\{+, -\}$ -labeled graph $(G_B, I_B), I_B \in \mathcal{I}(B)$ is not cf-symmetric: let $V(B) = \{\nu_r, \nu_1, \nu_2, \dots, \nu_m\}$, where $\nu_r, \nu_1, \dots, \nu_m$ appear along $H(B)$ in this order. Then it holds that

$$(\sigma_s(I_{\nu_1}), \sigma_s(I_{\nu_2}), \dots, \sigma_s(I_{\nu_m})) \neq \\ (\sigma_s(I_{\nu_m}), \sigma_s(I_{\nu_{m-1}}), \dots, \sigma_s(I_{\nu_1}))$$

If $(\sigma_s(I_{\nu_1}), \sigma_s(I_{\nu_2}), \dots, \sigma_s(I_{\nu_m}))$ is lexicographically greater than $(\sigma_s(I_{\nu_m}), \sigma_s(I_{\nu_{m-1}}), \dots, \sigma_s(I_{\nu_1}))$, then we define

$$O_s(B) = \nu_r \rightarrow \nu_1 \rightarrow \dots \rightarrow \nu_m$$

Otherwise $(\sigma_s(I_{\nu_1}), \sigma_s(I_{\nu_2}), \dots, \sigma_s(I_{\nu_m}))$ is lexicographically smaller than $(\sigma_s(I_{\nu_m}), \sigma_s(I_{\nu_{m-1}}), \dots, \sigma_s(I_{\nu_1}))$, and we define $O_s(B)$ reversely by $O_s(B) = \nu_r \rightarrow \nu_1 \rightarrow \dots \rightarrow \nu_m$.

(Case 3) $O_t(B)$ is not defined, a $\{+, -\}$ -labeled graph $(G_B, I_B), I_B \in \mathcal{I}(B)$ is cf-symmetric, and there exists the

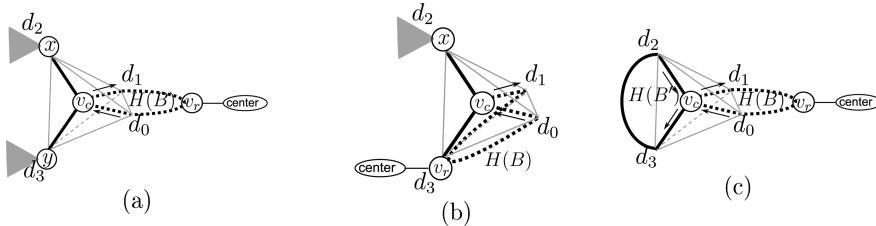


Figure 18. The stereo-orientation of a cyclic structure, where d_0, d_1, d_2 , and d_3 are the directions from v_c . Dotted cycles indicate Hamilton cycle $H(B)$.

counter vertex v_c which satisfies one of the following three conditions:

- (i) v_c has two children $x, y \in V(B)$ such that they are not in the same cyclic structure (see Figure 13a) and $\sigma_s(I_x) > \sigma_s(I_y)$ holds: For the four directions d_0, d_1, d_2 , and d_3 of carbon atom v_c (Figure 9), x and y are assumed to be in directions d_2 and d_3 , respectively. Then we define $O_s(B)$ to be

$$d_0 \rightarrow v_c \rightarrow d_1$$

(Figure 18a).

- (ii) v_c has one child $x \notin V(B)$ and is adjacent to three different vertices in $V(B)$ (Figure 13b): From assumption, B has a symmetric axis along v_r and v_c . Then v_c is adjacent to v_r . For the four directions d_0, d_1, d_2 , and d_3 of carbon atom v_c (Figure 9), x and v_r are assumed to be in directions d_2 and d_3 , respectively. Then we define $O_s(B)$ to be

$$d_0 \rightarrow v_c \rightarrow d_1$$

(Figure 18b).

- (iii) v_c is a root vertex of another cyclic structure $B' = B_{\text{down}}(v_c)$ (Figure 13c) and $O_s(B')$ is defined: For the four directions d_0, d_1, d_2 , and d_3 of carbon atom v_c (Figure 9), the atoms in $V(B')$ are assumed to be in directions d_2 and d_3 such that $O_s(B')$ is already given as $d_2 \rightarrow v_c \rightarrow d_3$. Then we define $O_s(B)$ to be

$$d_0 \rightarrow v_c \rightarrow d_1$$

(Figure 18c).

Now we define the correspondence configurations and labels of carbon atoms. By the definition of labels, the configuration around is unique if a carbon atom v receives label $l(v) = \text{nil}$. In what follows, we consider a carbon atom with $l(v) \neq \text{nil}$.

(Case 1) v does not belong to any cyclic structures: In this case v is adjacent to four vertices and signature σ_s of every child of v is different each other because $l(v) \in \{+, -\}$. If v is one vertex in a center consisting of two vertices, then we treat the other vertex in the center as the parent of v .

- (i) v has a parent: For the four directions $d_i, i = 0-3$ from v , as in Figure 9, we assume without loss of generality that the parent of v appears in direction d_0 , and the child u of v with the maximum σ_s appears in direction d_1 . Then each of the two configurations around v is determined by a placement of the rest of adjacent vertices x and y in directions d_2 and d_3 such that either

$$l(v) = + \Leftrightarrow \sigma_s(I_x) > \sigma_s(I_y)$$

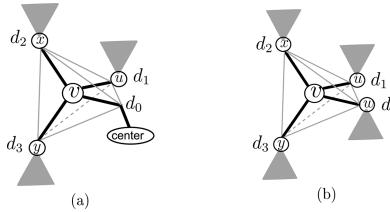


Figure 19. Configurations around a carbon atom $v \in V_C$ which is not in any cyclic structure and is adjacent to four atoms. (a) The case where v has a parent. It holds $\sigma_s(I_u) > \sigma_s(I_x) > \sigma_s(I_y)$ if and only if $l(v) = +$. It holds $\sigma_s(I_u) > \sigma_s(I_y) > \sigma_s(I_x)$ if and only if $l(v) = -$. (b) The case where v has no parent. It holds $\sigma_s(I_w) > \sigma_s(I_u) > \sigma_s(I_x) > \sigma_s(I_y)$ if and only if $l(v) = +$. It holds $\sigma_s(I_w) > \sigma_s(I_u) > \sigma_s(I_y) > \sigma_s(I_x)$ if and only if $l(v) = -$.

or

$$l(v) = - \Leftrightarrow \sigma_s(I_x) < \sigma_s(I_y)$$

holds (Figure 19a).

- (ii) v has no parent (i.e., v is a cut vertex and the corresponding C-node is the unicentroid of the BC tree): For the four directions $d_i, i = 0-3$ from v , as in Figure 9, we assume without loss of generality that the child u of v with the maximum σ_s appears in direction d_0 and that the child w of v with the second maximum σ_s appears in direction d_1 . Then each of the two configurations around v is determined by a placement of the rest of the adjacent vertices x and y in directions d_2 and d_3 such that either

$$l(v) = + \Leftrightarrow \sigma_s(I_x) > \sigma_s(I_y)$$

or

$$l(v) = - \Leftrightarrow \sigma_s(I_x) < \sigma_s(I_y)$$

holds (Figure 19b).

(Case 2) There exists a cyclic structure $B = B_{\text{up}}(v)$:

(Case 2-1) B is the center of G and B is a rotational symmetric block: In this case the topological orientation of $H(B)$ is defined by vertex numbers of $u \in V(B)$. Then for the four directions $d_i, i = 0-3$ from v , as in Figure 9, we assume without loss of generality that $O_s(B) = O_t(B)$ is defined to be $d_0 \rightarrow v \rightarrow d_1$.

- (i) v is an inner spiro atom: We denote by u_1 and u_2 the vertices, which are adjacent to v with edges which are not in $H(B)$, and by x_1 and x_2 the vertices, which are adjacent to v with edges which are in $H(B)$. We assume without loss of generality that it holds that $O_t(B) = x_1 \rightarrow v \rightarrow x_2 \rightarrow \dots \rightarrow u_2 \rightarrow \dots \rightarrow u_1 \rightarrow \dots \rightarrow x_1$ (Figure 20a). From assumptions, for the four directions $d_i, i = 0-3$ from v , as in Figure 9, x_1 and x_2 are in

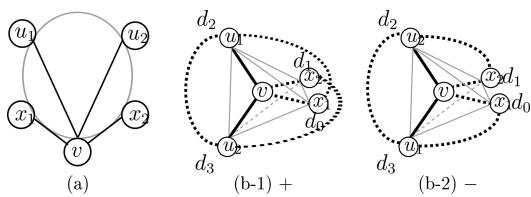


Figure 20. A graph structure and configurations around an inner spiro atom ν . (a) A graph structure around ν . The gray circle shows $H(B)$. (b) Configurations around ν . The dotted cycle shows $H(B)$; (b-1) $l(\nu) = +$ and (b-2) $l(\nu) = -$.

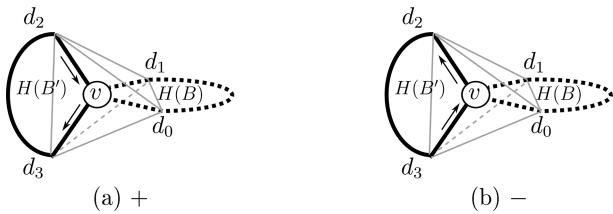


Figure 21. Configurations around ν with both $B_{\text{up}}(\nu)$ and $B_{\text{dwn}}(\nu)$. The dotted cycle shows $H(B)$. The arrows show the stereo-orientation of $H(B')$; (a) $l(\nu) = +$ and (b) $l(\nu) = -$.

directions d_0 and d_1 , respectively. Then each of the two configurations around ν is determined by a placement of u_1 such that either

$$l(\nu) = + \Leftrightarrow u_1 \text{ appears in direction } d_2$$

or

$$l(\nu) = - \Leftrightarrow u_1 \text{ appears in direction } d_3$$

holds (Figure 20b).

- (ii) Otherwise: We consider the following three subcases.
(ii-1) ν is a root vertex of another cyclic structure B' = $B_{\text{dwn}}(\nu)$: $O_s(B)$ is defined because $l(\nu) \in \{+, -\}$ holds. Then each of the two configurations around ν is determined by a placement of ν such that either

$$l(\nu) = + \Leftrightarrow O_s(B') \text{ is defined to be } d_2 \rightarrow \nu \rightarrow d_3$$

or

$$l(\nu) = - \Leftrightarrow O_s(B') \text{ is defined to be } d_2 \leftarrow \nu \leftarrow d_3$$

holds (Figure 21).

- (ii-2) ν has two children $x, y \notin V(B)$ which are not in the same cyclic structure: It holds that $\sigma_s(I_x) \neq \sigma_s(I_y)$ because $l(\nu) \in \{+, -\}$ holds. Assume without loss of generality that x and y appear in directions d_2 and d_3 , respectively. Then each of the two configurations around ν is determined by a placement of x and y such that either

$$l(\nu) = + \Leftrightarrow \sigma_s(I_x) > \sigma_s(I_y)$$

or

$$l(\nu) = - \Leftrightarrow \sigma_s(I_x) < \sigma_s(I_y)$$

holds (Figure 22).

- (ii-3) ν has one child $x \notin V(B)$ and is adjacent to three different vertices in $V(B)$: Each of the two configurations around ν is determined by a placement of x

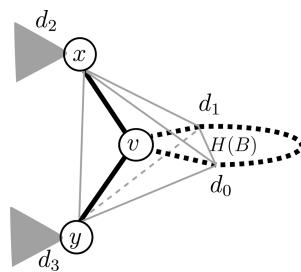


Figure 22. A configuration around ν with $B_{\text{up}}(\nu)$ and its children x and y , which are not in the same cyclic structure. The dotted cycle shows $H(B)$. If $\sigma_s(I_x) > \sigma_s(I_y)$ holds, then $l(\nu) = +$. If $\sigma_s(I_y) > \sigma_s(I_x)$ holds, then $l(\nu) = -$.

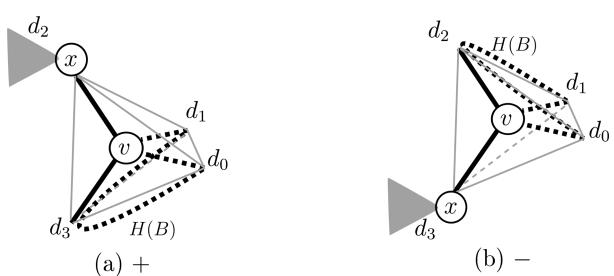


Figure 23. Configurations around ν with $B_{\text{up}}(\nu)$ and its one child x . The dotted cycle shows $H(B)$; (a) $l(\nu) = +$ and (b) $l(\nu) = -$.

such that either

$$l(\nu) = + \Leftrightarrow x \text{ appears in direction } d_2$$

or

$$l(\nu) = - \Leftrightarrow x \text{ appears in direction } d_3$$

holds (Figure 23).

- Case 2-2 Otherwise (i.e., B is not the center of G , or B is the center of G such that G_B does not have rotational symmetry):

- (i) ν is an inner spiro atom: We denote by u_1 and u_2 the vertices, which are adjacent to ν with edges which are not in $H(B)$, and by x_1 and x_2 the vertices, which are adjacent to ν with edges which are in $H(B)$. We assume without loss of generality that $\nu, x_1, \dots, u_1, \dots, u_2, \dots, x_2$ appear along $H(B)$ (Figure 20a).

- (i-1) ν is the root vertex or the counter vertex of B : We assume without loss of generality that x_1 is nearer to the root vertex than x_2 along $H(B)$ and that x_1 and x_2 are in directions d_0 and d_1 , respectively. Then each of the two configurations around ν is determined similarly to case 2-1 (i).

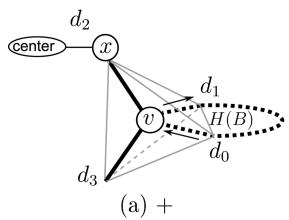
- (i-2) ν is the root vertex or the counter vertex of B :

- (i-2-a) $H(B)$ has the stereo-orientation which is defined by the graph structure and $\sigma_s(u)$, $u \in V(B) - \{\nu\}$: Each of the two configurations around is determined similarly to case 2-1 (i).

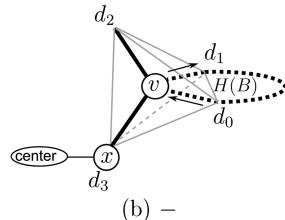
- (i-2-b) Otherwise: We determine the two configurations around ν similarly to case (i-1), by choosing x_1 and x_2 arbitrarily.

- (ii) Otherwise:

- (ii-1) ν is not the root vertex or the counter vertex of B : for the four directions d_i , $i = 0-3$ from ν , as in Figure 9,



(a) +



(b) -

Figure 24. Configurations around ν with only one $B = B_{\text{dwn}}(\nu)$. The dotted cycle shows $H(B)$. The arrows show the stereo-orientations of $H(B)$; (a) $l(\nu) = +$ and (b) $l(\nu) = -$.

we assume without loss of generality that the edge of $H(B)$ heading for the root vertex is in direction d_0 and that another edge of $H(B)$ is in direction d_1 . Then each of the two configurations around ν is determined similarly to Case 2-1 (ii).

- (ii-2) ν is the root vertex or the counter vertex of B : In this case $O_s(B)$ is defined by the graph structure of B and $\sigma_s(u), u \in V(B) - \{\nu\}$ because $l(\nu) \in \{+, -\}$ holds. Then for the four directions $d_i, i = 0 - 3$ from ν , as in Figure 9, we assume without loss of generality that $O_s(B)$ is defined to be $d_0 \rightarrow \nu \rightarrow d_1$. Then each of the two configurations around ν is determined similarly to Case 2-1 (ii).

Case 3 There does not exist any cyclic structure $B_{\text{up}}(\nu)$, and there exists some cyclic structures B containing ν and its children. In this case, if belongs to the center, then the center consists of only one vertex ν .

- (i) ν is not the center of G : In this case $O_s(B)$ is defined because $l(\nu) \in \{+, -\}$ holds. Then for the four directions $d_i, i = 0, 1, 2, 3$ from ν , as in Figure 9, we assume without loss of generality that $O_s(B)$ is defined to be $d_0 \rightarrow \nu \rightarrow d_1$. We write the parent vertex of ν as x . Then each of the two configurations around ν is determined by a placement of x such that either

$$l(\nu) = + \Leftrightarrow x \text{ appears in direction } d_2$$

or

$$l(\nu) = - \Leftrightarrow x \text{ appears in direction } d_3$$

holds (Figure 24).

- (ii) ν is the center of G :

- (ii-1) There exist two distinct cyclic structures B and B' containing ν and its children: In this case both $H(B)$ and $H(B')$ have stereo-orientations because $l(\nu) \in \{+, -\}$. Then for the four directions $d_i, i = 0 - 3$ from, as in Figure 9, we assume without loss of generality that $O_s(B)$ is defined to be $d_0 \rightarrow \nu \rightarrow d_1$. Then each of the two configurations around ν is determined by a placement of B' such that either

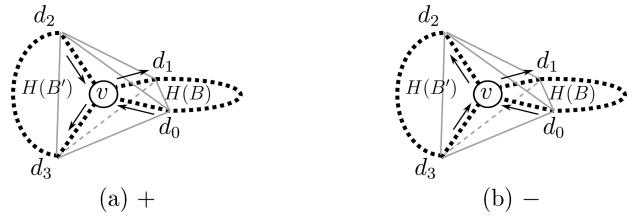
$$l(\nu) = + \Leftrightarrow O_s(B') \text{ is defined to be } d_2 \rightarrow \nu \rightarrow d_3$$

or

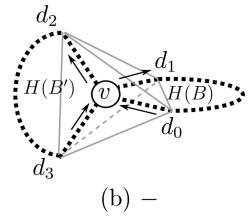
$$l(\nu) = - \Leftrightarrow O_s(B') \text{ is defined to be } d_2 \leftarrow \nu \leftarrow d_3$$

holds (Figure 25).

- (ii-2) There exists only one cyclic structure B containing ν and its children: In this case $O_s(B)$ is defined because $l(\nu) \in \{+, -\}$. Then for the four directions

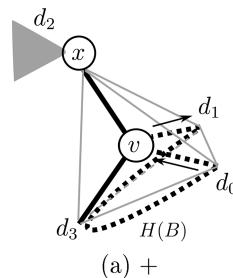


(a) +

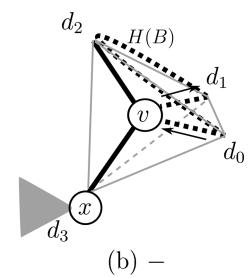


(b) -

Figure 25. Configurations around the center ν with two $B_{\text{dwn}}(\nu)$. The dotted cycles show $H(B)$ and $H(B')$. The arrows along d_0 and d_1 show the stereo-orientation of $H(B)$, and the arrows along d_2 and d_3 show the stereo-orientation of $H(B')$; (a) $l(\nu) = +$ and (b) $l(\nu) = -$.



(a) +



(b) -

Figure 26. Configurations around the center ν with only one $B_{\text{dwn}}(\nu)$, where ν has one child x , which is not in $V(B)$. The dotted cycle shows $H(B)$. The arrows show the stereo-orientation of $H(B)$; (a) $l(\nu) = +$ and (b) $l(\nu) = -$.

$d_i, i = 0 - 3$ from ν , as in Figure 9, we assume without loss of generality that $O_s(B)$ is defined to be $d_0 \rightarrow \nu \rightarrow d_1$.

- (ii-2-a) ν has one child $x \notin V(B)$: ν is adjacent to four distinct vertices because $l(\nu) \in \{+, -\}$. Each of the two configurations around is determined by a placement of x such that either

$$l(\nu) = + \Leftrightarrow x \text{ appears in direction } d_2$$

or

$$l(\nu) = - \Leftrightarrow x \text{ appears in direction } d_3$$

holds (Figure 26).

- (ii-2-b) ν has two children $x, y \notin V(B)$: It holds that $\sigma_s(I_x) \neq \sigma_s(I_y)$ because $l(\nu) \in \{+, -\}$. Assume without loss of generality that x and y appear in directions d_2 and d_3 , respectively. Then each of the two configurations around is determined by a placement of x and y such that either

$$l(\nu) = + \Leftrightarrow \sigma_s(I_x) > \sigma_s(I_y)$$

or

$$l(\nu) = - \Leftrightarrow \sigma_s(I_x) < \sigma_s(I_y)$$

holds (Figure 27).

Given a simplified proper representation $I \in \mathcal{I}(G)$, we can determine the set of configurations around all carbon atoms represented by I , which are determined along the BC tree from the leaves to the root. (For vertices which are in the same cyclic structure B with the root vertex ν_r and the counter vertex ν_c , we determine their configurations along the order $\nu \in V(B) - \{\nu_c, \nu_r\} \rightarrow \nu_c \rightarrow \nu_r$). Conversely, given a set of configurations of

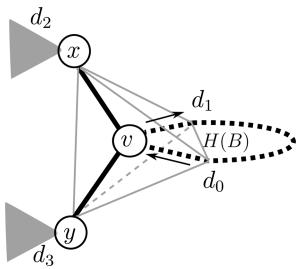


Figure 27. A configuration around the center ν with only one $B_{\text{dwn}}(\nu)$, where has two children x and y which are not in $V(B)$. The dotted cycle shows $H(B)$. The arrows show the stereo-orientation of $H(B)$. If $\sigma_s(I_x) > \sigma_s(I_y)$ holds, then $l(\nu) = +$. If $\sigma_s(I_y) > \sigma_s(I_x)$ holds, then $l(\nu) = -$.

all carbon atoms of a stereoisomer of G , we can construct the simplified proper representation I corresponding to the structure along the BC tree of G from the leaves to the root.

FUNDAMENTAL IDEA OF OUR ENUMERATION ALGORITHMS

To compute $f(G,\nu)$, $\nu \in V$ which is in a block, we define $g(G,\nu)$, $h(G,\nu)$, and $h'(G,\nu)$ as follows. Informally, $h(G,\nu)$ (respectively, $g(G,\nu)$) is the number of combinations of stereoisomers of G_x over all children x of ν such that ν has (respectively, does not have) an asymmetry induced by its descendants, while $h'(G,\nu)$ is the 0–1 value denoting whether ν is an inner spiro atom or not. In the following definition, we consider the symmetry of $\{+, -\}$ -labeled graphs, which is defined in the section Asymmetry around a Carbon Atom. Informally, a $\{+, -\}$ -labeled graph of a subgraph rooted at a block B is counter-free symmetric (cf-symmetric for short) if it has an outerplane drawing which displays the axial symmetry in a geometric sense by ignoring the structure of the subgraph G_{ν_c} rooted at the counter vertex ν_c of B , and it is symmetric if the $\{+, -\}$ -labeled graph is cf-symmetric and the structure of the $\{+, -\}$ -labeled graph G_{ν_c} is symmetric when considering its three-dimensional arrangements.

Definition 17. For each vertex $\nu \in V$ which is in a cyclic structure, we define $g(G,\nu)$, $h(G,\nu)$, and $h'(G,\nu)$ as follows:

Case 1 ν has no children: We define $h(G,\nu) = 0$,

$$g(G,\nu) = \begin{cases} 0 & \text{if } \nu \text{ is an inner spiro atom} \\ 1 & \text{otherwise} \end{cases}$$

and

$$h'(G,\nu) = \begin{cases} 1 & \text{if } \nu \text{ is an inner spiro atom} \\ 0 & \text{otherwise} \end{cases}$$

Case 2 ν has some children: We define $h'(G,\nu) = 0$, $g(G,\nu)$: the number of combinations of stereoisomers of G_x over all children x of ν such that

- (i) When $B_{\text{up}}(\nu)$ does not exist or ν is not the counter vertex of $B_{\text{up}}(\nu)$, it holds that $l(\nu) = \text{nil}$; or
 - (ii) When ν is the counter vertex of $B_{\text{up}}(\nu)$, it holds that $l(\nu) = \text{nil}$ and a $\{+, -\}$ -labeled graph of $G_{B_{\text{up}}(\nu)}$ is symmetric if it is cf-symmetric, and
- $h(G,\nu)$: the number of combinations of stereoisomers of G_x over all children x of ν such that

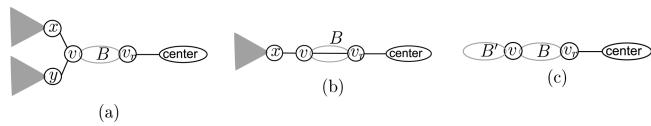


Figure 28. Graph structures around a vertex ν .

- (i) When $B_{\text{up}}(\nu)$ does not exist or ν is not the counter vertex of $B_{\text{up}}(\nu)$, it holds that $l(\nu) \in \{+, -\}$; or
- (ii) When ν is the counter vertex of $B_{\text{up}}(\nu)$, it holds that $l(\nu) \in \{+, -\}$ or a $\{+, -\}$ -labeled graph of $G_{B_{\text{up}}(\nu)}$ is asymmetric if it is cf-symmetric.

In the rest of this paper, we write $f(G,\nu)$, $g(G,\nu)$, $h(G,\nu)$, and $h'(G,\nu)$ as $f(\nu)$, $g(\nu)$, $h(\nu)$, and $h'(\nu)$, respectively. These definitions of $f(\nu)$, $g(\nu)$, and $h(\nu)$ are cycle-enabled extensions of our previous study,¹⁴ which treats only the case when G is a tree.

For $\nu \in V_C$ which is in a noncentral block, we compute $f(\nu)$, $g(\nu)$, $h(\nu)$, and $h'(\nu)$ using the following lemmas. These Lemmas give the conditions for each vertex having chirality (in our sense) and a procedural definition of $f(\nu)$, $g(\nu)$, and $h(\nu)$, where Lemmas 18 and 19, respectively, treat the cases where ν belongs to a nonroot block as a lower vertex and where ν belongs to a bridge as a lower vertex.

Lemma 18. Let $\nu \in V_C$ be a carbon atom with a noncentral cyclic structure $B = B_{\text{up}}(\nu)$.

- (i) ν receives a label $l(\nu) \in \{+, -\}$ in a combination of stereoisomers of its children if and only if one of the following conditions holds:
 - (1) ν is not the counter vertex of B and ν satisfies one of the following three conditions:
 - (a) ν has two children x and y such that G_x and G_y are not rooted isomorphic to each other (see Figure 28a).
 - (b) ν is adjacent to four distinct vertices and has just one child x (see Figure 28b).
 - (c) ν is the root vertex of another cyclic structure B' such that a $\{+, -\}$ -labeled graph of $G_{B'}$ is asymmetric (see Figure 28c).
 - (2) ν is the counter vertex of B , a $\{+, -\}$ -labeled graph of G_B is not cf-symmetric, and ν satisfies one of the above conditions (a), (b), or (c).
 - (3) ν is an inner spiro atom.
- (ii) If ν is the counter vertex of B and a $\{+, -\}$ -labeled graph of G_B is cf-symmetric, then it is asymmetric if and only if ν satisfies one of the above conditions (a), (b), or (c).
- (iii) Let x_1, \dots, x_m , $m > 0$, be the children of ν . Then for two sets

$$\mathcal{I}_g(\nu) = \{I_{x_1} \cup \dots \cup I_{x_m} | \nu \text{ does not satisfy any of the above conditions (a), (b), or (c)}, I_{x_i} \in I(x_i), i = 1, \dots, m\}$$

and

$$\mathcal{I}_h(\nu) = \{I_{x_1} \cup \dots \cup I_{x_m} | \nu \text{ satisfies one of the above conditions (a), (b), or (c)}, I_{x_i} \in I(x_i), i = 1, \dots, m\}$$

we have

$$g(\nu) = |\mathcal{I}_g(\nu)|, h(\nu) = |\mathcal{I}_h(\nu)|$$

Proof of Lemma 18 is given in Section S2.1, Supporting Information.

Lemma 19. Let $\nu \in V_C$ be a carbon atom with four vertices adjacent to ν which does not have any cyclic structure $B_{\text{up}}(\nu)$ and has a cyclic structure $B = B_{\text{dwn}}(\nu)$.

- (i) ν receives a label $l(\nu) \in \{+, -\}$ for a combination of stereoisomers of its children if and only if a $\{+, -\}$ -labeled graph of G_B is asymmetric.
- (ii) Let x_1, \dots, x_m be the children of ν . Then for two sets:

$$\begin{aligned}\mathcal{I}_g(\nu) &= \{I_{x1} \cup \dots \cup I_{xm} \mid a\{+, -\}\text{-labeled graph of } G_B \\ &\quad \text{is symmetric, } I_{xi} \in I(x_i), i = 1, \dots, m\}\end{aligned}$$

and

$$\begin{aligned}\mathcal{I}_h(\nu) &= \{I_{x1} \cup \dots \cup I_{xm} \mid a\{+, -\}\text{-labeled graph of } G_B \\ &\quad \text{is asymmetric, } I_{xi} \in I(x_i), i = 1, \dots, m\}\end{aligned}$$

$\mathcal{I}(\nu)$ is given by

$$\begin{aligned}\mathcal{I}(\nu) &= \{I \cup \{(n(\nu), +)\}, I \cup \{(n(\nu), -)\} \\ &\quad | I \in \mathcal{I}_h(\nu)\} \cup \{I \cup \{(n(\nu), \text{nil})\} | I \in \mathcal{I}_g(\nu)\}\end{aligned}$$

and we have $g(\nu) = |I_g(\nu)|$, $h(\nu) = |I_h(\nu)|$, and $f(\nu) = |I(\nu)| = g(\nu) + 2h(\nu)$.

Proof of Lemma 19 is given in Section S2.2, Supporting Information.

ALGORITHM

In this section we present an algorithm for enumerating all stereoisomers of a given outerplanar graph G . The first phase, called the counting phase, computes $f^*(G)$, the number of stereoisomers, by dynamic programming. Using the information calculated by the counting phase, the second phase, called the output phase, constructs each stereoisomer one by one.

Counting Phase. The counting phase computes $f(\nu)$, $g(\nu)$, $h(\nu)$, and $h'(\nu)$ for every vertex $\nu \in V$ along the BC tree of G from the leaves to the root. When we reach the centroid of G , we are ready to compute $f^*(G)$. All of the recursive formulas are given in Section S5, Supporting Information.

We show a simple example of computing $f(\nu)$, $g(\nu)$, and $h(\nu)$ for vertices $\nu \in V$. Consider the example given in Figure 11b. By Lemma 18, we have $g(\nu_1) = 0$ and $h(\nu_1) = 1$ because ν_1 has two nonrooted isomorphic subgraphs OH and H as its children, which satisfies the condition (a) in Lemma 18. Similarly, we have $g(\nu_3) = 0$ and $h(\nu_3) = 1$. By contrast, we have $g(\nu_2) = 1$ and $h(\nu_2) = 0$ because ν_2 does not satisfy any of the conditions (a), (b), or (c) in Lemma 18. After that, at vertex ν_4 , we compute $g(\nu_4)$, $h(\nu_4)$, and $f(\nu_4)$ by Lemma 19 as follows. The number of combinations of stereoisomers of the children ν_1, ν_2 and ν_3 of ν_4 is $\prod_{i=1}^3(g(\nu_i) + 2h(\nu_i))$. As the subgraph G_{ν_1} and G_{ν_3} are rooted isomorphic, the $\{+, -\}$ -labeled graph of G_B is asymmetric if and only if the $\{+, -\}$ -labeled graph of G_{ν_1} and G_{ν_3} are not rooted isomorphic. Then we have $g(\nu_4) = (g(\nu_1) + 2h(\nu_1))(g(\nu_2) + 2h(\nu_2)) = 2$, $h(\nu_4) = (g(\nu_2) + 2h(\nu_2)) = 1$, and $f(\nu_4) = g(\nu_4) + 2h(\nu_4) = 4$.

In the following description of our algorithm, “parents” and “children” mean what are defined in the simple graph $S(G)$ of a given graph G , not in the BC tree $\text{BC}(G)$.

Algorithm Counting Phase.

Input: An outerplanar multigraph $G = (V, E)$ whose vertex set consists of carbon, hydrogen, oxygen and nitrogen atoms along with vertex numbers.

Output: The number of stereoisomers $f^*(G)$ and $f(v), g(v), h(v), h'(v)$ for every vertex $v \in V$ except for a vertex v such that the center of G consists of v .

Compute all 2-connected components of G and construct the BC tree of G using a graph traversal method¹⁹;

Find the unicentroid of the BC tree according to Theorem 1, and let the center be the set of vertices which belong to the B-node or C-node corresponding to the unicentroid;

Let the root of G be the center;

Compute indices of rooted subgraphs G_v for all $v \in V$;

Compute the symmetries of subgraphs rooted at all blocks in G using the method by Imada et al.²²;

Initialize the scanning queue $Q := \phi$;

for each leaf $v \in V$ with no children do

 Compute $g(v), h(v)$ and $h'(v)$ as described in Section S4.1;

 Let v be “scanned”;

 /* Let u be the parent of v . */

 if all the children of u are “scanned” and u is not in the center of G then

 ENQUE(Q, u);

 end if

end for;

while $Q \neq \phi$ do

$v := \text{DEQUE}(Q)$;

 Compute $f(v), g(v)$ and $h(v)$ as described in Section S4.1;

 Let v be “scanned”;

 /* Let u be the parent of v . */

 if all the children of u are “scanned” and u is not in the center of G then

 ENQUE(Q, u);

 end if

end while;

Compute $f^*(G)$ as described in Section S4.2, Supporting Information.

In general, the number of stereoisomers increases exponentially as the number of atoms increases. In the following, we assume that each of addition, subtraction, multiplication, and division over integers can be executed in a unit time. We get the following theorem.

Theorem 20. For an outerplanar multigraph $G = (V, E)$ with $|V| = n$, the counting phase computes the number of stereoisomers $f^*(G)$ correctly and can be implemented to run in $O(n)$ time and space.

Proof of Theorem 20 is given in Section S2.3, Supporting Information.

Output Phase. The output phase constructs proper representations for stereoisomers by using $f^*(G), f(\nu), g(\nu), h(\nu)$, and $h'(\nu)$ computed in the counting phase. For $i = 1, 2, \dots, f^*(G)$, we output the proper representation for the i -th stereoisomer of G by backtracking the computation process of the counting phase. When we compute the k -th rooted stereoisomer of G_ν , we detect the corresponding $l(\nu)$ and calculate k_u for every child u of ν , and we trace the computation process recursively to the leaves of G . When this backtrack process completes, we get one proper representation generated by the calculated set of labels $l(\nu)$ for all $\nu \in V$.

Here we consider the example given in Figure 11a (see also Figure 29). When the output phase processes ν_3 , we have received an instruction from the parent of ν_3 to “choose the k_{ν_3} -th rooted stereoisomer of G_{ν_3} ”. It holds $1 \leq k_{\nu_3} \leq 4$ because the counting phase computed $f(\nu_3) = 4$. We order rooted stereoisomers of G_{ν_3} as follows:

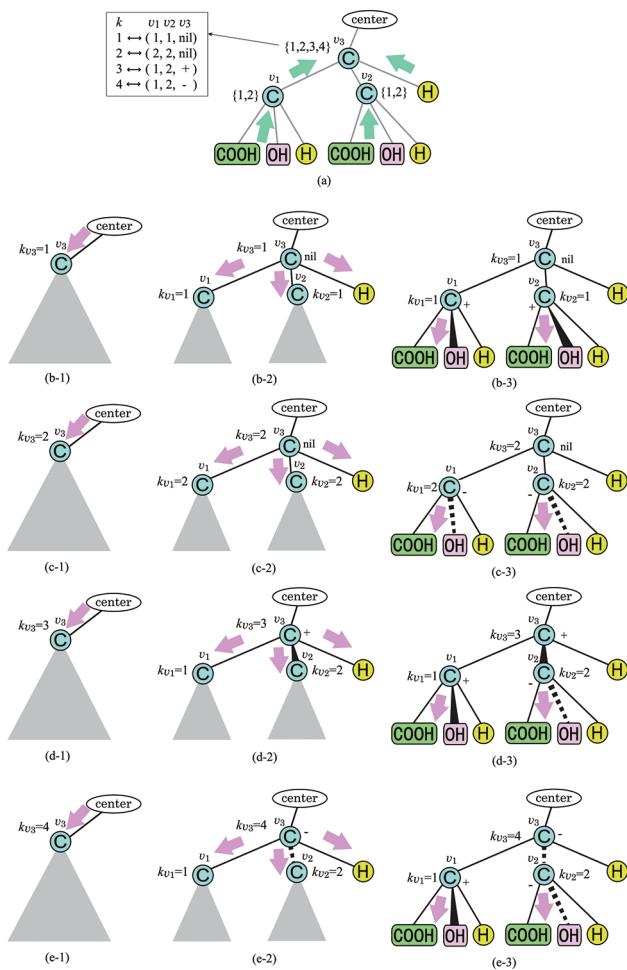


Figure 29. Explanation of the proposed algorithm: (a) counting phase and (b-l-e-3) output phase.

- If $k_{\nu_3} = 1$ hold, then we have $l(\nu_3) = \text{nil}$, and the rooted stereoisomer of G_{ν_3} is composed of the first stereoisomer of G_{ν_1} and the first stereoisomer of G_{ν_2} (i.e., $k_{\nu_1} = k_{\nu_2} = 1$ hold).
- If $k_{\nu_3} = 2$ holds, then we have $l(\nu_3) = \text{nil}$, and the rooted stereoisomer of G_{ν_3} is composed of the second stereoisomer of G_{ν_1} and the second stereoisomer of G_{ν_2} (i.e., $k_{\nu_1} = k_{\nu_2} = 2$ holds).
- If $k_{\nu_3} = 3$ holds, then we have $l(\nu_3) = +$, and the rooted stereoisomer of G_{ν_3} is composed of the first stereoisomer of G_{ν_1} and the second stereoisomer of G_{ν_2} (i.e., $k_{\nu_1} = 1$ and $k_{\nu_2} = 2$ holds).
- If $k_{\nu_3} = 4$ holds, then we have $l(\nu_3) = -$, and the rooted stereoisomer of G_{ν_3} is composed of the first stereoisomer of G_{ν_1} and the second stereoisomer of G_{ν_2} (i.e., $k_{\nu_1} = 1$ and $k_{\nu_2} = 2$ holds).

We compute $k_{\nu_1} = 1$, k_{ν_2} , and $l(k_{\nu_1} = 1 \text{ and } \nu_3)$ from the given k_{ν_3} , using information of $g(\nu_3)$, $h(\nu_3)$, and $f(\nu_3)$ computed in the counting phase.

In the remainder of this section, we first define bijections between a set of tuples and combinations of the elements in tuples, after which we give a general description of the algorithm and analyze the time complexity of the output phase.

Bijections for Fast Generation. Recall that we do not generate any tables of (rooted) stereoisomers during the counting phase. However, for a given k , the output phase needs to find the k -th combination of numbers k_u of all children of u . To design an efficient algorithm for finding a desired combination of such numbers k_u this subsection defines bijections between a set of tuples and combinations of the elements in tuples.

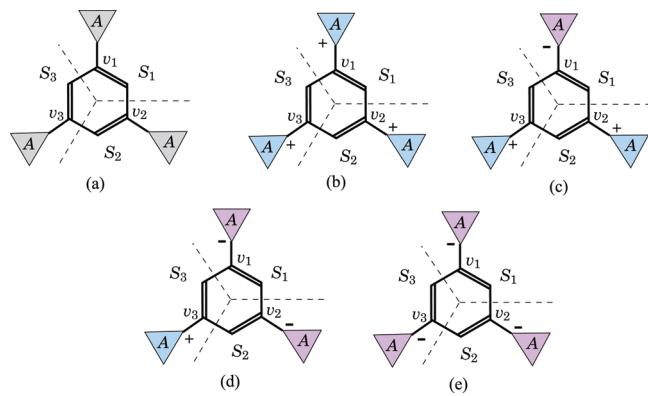


Figure 30. Explanation of the processing of the center block in the output phase. (a) The benzene ring is the center block whose order of the rotational symmetry is 3, where the multiplicity of double bonds is ignored. (b–e) For this example, the output phase examines $2^3 = 8$ tuples from $[1,1,1]$ to $[2,2,2]$ and selects minimal tuples $[1,1,1]$, $[1,1,2]$, $[1,2,2]$, and $[2,2,2]$ which correspond to stereoisomers (b–e), respectively.

Definition 21. For positive integers M_1, M_2, \dots, M_p , define the set $D(M_1, M_2, \dots, M_p)$ of tuples by

$$D(M_1, M_2, \dots, M_p) := \{[k_1, k_2, \dots, k_p] | k_i \in \{1, 2, \dots, M_i\}, i = 1, 2, \dots, p\}$$

Let $D(; M_1, M_2, \dots, M_p)$ denote a bijection between the set $\{1, 2, \dots, M_1 M_2 \dots M_p\}$ of integers and $D(M_1, M_2, \dots, M_p)$. Let $D(k; M_1, M_2, \dots, M_p)$ denote the tuple $[k_1, k_2, \dots, k_p] \in D(M_1, M_2, \dots, M_p)$ corresponding to $k \in \{1, 2, \dots, M_1 M_2 \dots M_p\}$.

Note that choice of such a bijection $D(; M_1, M_2, \dots, M_p)$ is not unique. It is not difficult to see that there exists a bijection $D(; M_1, M_2, \dots, M_p)$ such that we can compute $D(k; M_1, M_2, \dots, M_p)$ in $O(p)$ time and space for any integer $k \in \{1, 2, \dots, M_1 M_2 \dots M_p\}$ (see our previous study¹⁴ for details).

Definition 22. For positive integers n and p , define the set $C_{n,p}$ of tuples by

$$C_{n,p} := \{[k_1, k_2, \dots, k_p] | k_i \in \{1, 2, \dots, n\}, i = 1, 2, \dots, p, k_j \neq k_{j'}, 1 \leq j < j' \leq p\}$$

Let $C_{n,p}()$ denote a bijection between the set

$$\left\{1, 2, \dots, \binom{n}{p}\right\}$$

of integers and $C_{n,p}$. Let $C_{n,p}(k)$ denote the tuple $[k_1, k_2, \dots, k_p] \in C_{n,p}$ corresponding to

$$k \in \left\{1, 2, \dots, \binom{n}{p}\right\}$$

Again, choice of such a bijection $C_{n,p}()$ is not unique. For $p \leq 4$, there exists a bijection $C_{n,p}()$ such that we can compute $C_{n,p}(k)$ in $O(1)$ time and space¹⁴ for any integer

$$k \in \left\{1, 2, \dots, \binom{n}{p}\right\}$$

Description of the Output Phase. This subsection gives a general description of the output phase and analyzes its time complexity. Note that $A(\nu, k_\nu)$ is the pair $[l(\nu), k_\nu]$ of label $l(\nu)$ and the integer k_ν that stands for the k_ν -th combination of the stereoisomers of the children of ν in the k -th stereoisomer of G_ν , which is defined in Definition S6 in Section S5, Supporting Information.

Algorithm Output Phase.

Input: An outerplanar multigraph $G = (V, E)$ whose vertex set consists of carbon, hydrogen, oxygen and nitrogen atoms along with vertex numbers, the center of G , $f(v), g(v), h(v)$ and $h'(v)$ for all vertices $v \in V$ except for a vertex v such that the center of G consists of v , indices of all rooted-subgraphs $G_v, v \in V$, and $f^*(G)$.

Output: All of the elements of $I \in \mathcal{I}(G)$ without duplication.

Let the center of G be composed of vertices v_1, v_2, \dots, v_i ;

if the center of G is not a block B such that G_B has a rotational symmetry **then**

```
for each  $k = 1, 2, \dots, f^*(G)$  do
    for each  $v \in V$  do
         $l(v) := \text{nil}$ 
    end for;
    for each  $v_j$  ( $j = 1, \dots, i$ ) do
        Compute  $l(v_j)$  and  $k_{v_j}$  as described in Section S5.2;
        Recursive( $v_j, G_{v_j}, k_j$ )
    end for;
    Output  $I = \{(n(v), l(v)) \mid v \in V\}$  as the  $k$ -th stereoisomer
end for;
```

else

```
/* The center of  $G$  is a block  $B$  such that  $G_B$  has a rotational symmetry */
Let  $m$  be the order of the rotational symmetry of  $G$ ;
Divide  $G$  into  $m$  isomorphic subgraphs  $S_1, S_2, \dots, S_m$  by removing some edges in the
block  $B$ ;
Let  $V(B) \cap V(S_1) = \{v_1^1, v_2^1, \dots, v_{|V(B)|/m}^1\}$ , and let  $M = \prod_{j=1}^{|V(B)|/m} (g(v_j^1) + 2h(v_j^1) +
2h'(v_j^1))$ ;
 $r := 1$ ;
for each  $k = 1, 2, \dots, M^m$  do
    if  $r > f^*(G)$  then
        break;
    else
        Let  $[k_1, k_2, \dots, k_m] := D(k; M, M, \dots, M)$ ;
        if the list  $[k_1, k_2, \dots, k_m]$  is lexicographically minimum among
             $[k_{m-(j-1)}, k_{m-(j-2)}, \dots, k_m, k_1, \dots, k_{m-j}], j = 0, 1, \dots, m-1$  then

```

Figure 30 explains how the center block is processed in the output phase. As described in the caption of the figure, the output phase examines 8 tuples, selects minimal tuples $[1,1,1]$, $[1,1,2]$, $[1,2,2]$, and $[2,2,2]$, and outputs the corresponding stereoisomers.

Similar to the time complexity of the counting phase, in the following, we assume that each of addition, subtraction, multiplication, and division over integers can be executed in a unit time. We get the following theorem.

Theorem 23. For an outerplanar multigraph $G = (V, E)$ with $|V| = n$, the output phase enumerates all of the stereoisomers $I \in \mathcal{I}(G)$ without duplication in $O(n^3)$ time on average per isomer and in $O(n)$ space.

Proof of Theorem 23 is given in Section S2.4, Supporting Information.

■ EXPERIMENTAL RESULTS

We implemented our algorithm and conducted computational experiments to evaluate its practical performance. In addition, in order to validate our algorithm, we compared its outputs with the results of Rücker et al.¹³ (experiment 1) and Razinger et al.²⁴ (experiment 2). Rücker et al. manually counted the number of stereoisomers of several chemical compounds.¹³ Razinger et al. constructed a program for exhaustive, nonredundant stereoisomer

Compute $k_{v_j}, j = 1, 2, \dots, i$ as described in Case-1 in S5.2;
if the list $[k_{v_1}, k_{v_2}, \dots, k_{v_i}]$ is lexicographically minimum among the lists corresponding to stereoisomers obtained by reversing the stereoisomer corresponding to the given list along all of the symmetric axes of G_B **then**

```
/* Then output this stereoisomer */
for each  $v \in V$  do
     $l(v) := \text{nil}$ 
end for;
for each  $v \in V(B)$  do
     $[l(v), k_v] := A(v, k_v)$  and Recursive( $v_j, T_{v_j}, k_j$ )
end for;
Output  $I = \{(n(v), l(v)) \mid v \in V\}$  as the  $r$ -th stereoisomer;
 $r := r + 1$ 
end if
end if
end for
end if.
```

Procedure Recursive(v, G_v, k)

Input: $v \in V$, a rooted-subgraph G_v , and positive integer k .

Output: $l(u)$ for all of the vertices $u \in G_v$.

```
if  $v$  has some children then
    if  $l(v) = \text{nil}$  holds then
        Compute  $l(v)$  as described in S5.1
    end if.
/* Let  $v_1, \dots, v_i$  be children of  $v$  */
Compute  $k_j$  ( $j = 1, \dots, i$ ) as described in S5.1;
for each  $v_j$  ( $j = 1, \dots, i$ ) do
    Recursive( $v_j, G_{v_j}, k_j$ )
end for
end if.
```

generation using the idea of Nourse¹⁰ and tested the program with various compounds.²⁴ We report experimental results performed on a PC with a Intel(R) Core(TM) i5 650 CPU (3.20 GHz) running under the Ubuntu 10.04 Linux operating system, and our algorithm was implemented using the C language. Our program receives a MDL Mol file format as an input, which was created by KegDraw distributed by KEGG (Kyoto Encyclopedia of Genes and Genomes).²⁵

Experiment 1. Rücker et al.¹³ manually counted the number of stereoisomers of some oligoinositols. Figure S8, Supporting Information shows graph structures of oligoinositols for which Rücker et al. counted the number of stereoisomers. Table 1 shows the experimental results of our algorithm. For each compound in Figure S8, Supporting Information, CPU times for the counting phase are less than 0.01 s and the numbers of stereoisomers $f^*(G)$ that our program outputs are the same as those of Rücker et al. Furthermore, we counted and enumerated all of the stereoisomers of oligoinositols shown in Figures S9–S11, Supporting Information. The counting phase again required less than 0.01 s, and CPU times for the output phase and the numbers of stereoisomers $f^*(G)$ that our program outputs are shown in Table 1.

Experiment 2. We chose some instances (Figures S12–S14, Supporting Information) used by Razinger et al.,²⁴ which are

composed of carbon, hydrogen, oxygen, fluorine, chlorine, and bromine atoms. The current versions of our algorithm can treat only the compounds which are composed of carbon, hydrogen, oxygen, and nitrogen atoms. Therefore, we created instances by replacing fluorine, chlorine, and bromine atoms in the instances

Table 1. Experimental Results on Stereoisomer Enumeration for the Chemical Graphs Shown in Figures S8–S11 of the Supporting Information^a

input:		$f^*(G)$	t (s)
$C_6O_6H_{12}$	Figure S8a	9	less than 0.01
$C_{12}O_{11}H_{22}$	Figure S8b	528	less than 0.01
$C_{18}O_{16}H_{32}$	Figure S8c-1	32 896	0.05
	Figure S8c-2	32 768	0.04
	Figure S8c-3	16 512	0.03
$C_{24}O_{21}H_{42}$	Figure S9a	4 194 304	7.77
	Figure S9b	2 098 176	3.96
	Figure S9c	2 098 176	4.05
	Figure S9d	2 097 152	3.90
	Figure S10a	2 097 152	3.84
	Figure S10b	2 097 152	4.00
	Figure S10c	1 048 576	2.00
	Figure S11a	524 800	1.00
	Figure S11b	349 568	1.08

^aWhere $f^*(G)$ is the number of stereoisomers that our algorithm outputs, and t is the computation time for the output phase.

used by Razinger et al.²⁴ with substructures (i) $-\text{NH}_2$, (ii) $-\text{CH}_2-\text{NH}_2$, and (iii) $-\text{CH}_2-\text{CH}_2-\text{NH}_2$, respectively. As each compound which we choose does not include any of the substructures (i), (ii) or (iii), the results are expected to be the same. Figures S12–S14, Supporting Information show graph structures of these instances and the numbers of stereoisomers $f^*(G)$ that our algorithm computed. If $f^*(G)$ is different from the numbers N of stereoisomers that Razinger et al.²⁴ show, then we also write the number N beside $f^*(G)$ in Figures S12–S14, Supporting Information. For each compound, CPU times for the counting and output phases are less than 0.01 s.

For most compounds, the number of stereoisomers $f^*(G)$ is the same as that (N) of Razinger et al.²⁴ However, there exist some exceptions such that $f^*(G) = N$ does not hold. We consider the reasons why such exceptions occur.

For the compound shown in Figure S13d, Supporting Information, $f^*(G) < N$. The reason is that the multiplicity within a block of a chemical graph is ignored in our definition of stereoisomers, and thus we do not distinguish the chemical structures of Figure 31c and d. The order (which is defined in Section S3.1, Supporting Information) of a rotational symmetry in the chemical graph in Figure S13d, Supporting Information is regarded as four, although the actual order is two as in the case of Razinger et al. Therefore, for the compound shown in Figure S13d, Supporting Information, Razinger et al. enumerated all seven stereoisomers shown in Figure 31, whereas we regarded Figure 31c and d as the same stereoisomer, and thus only six stereoisomers were enumerated.

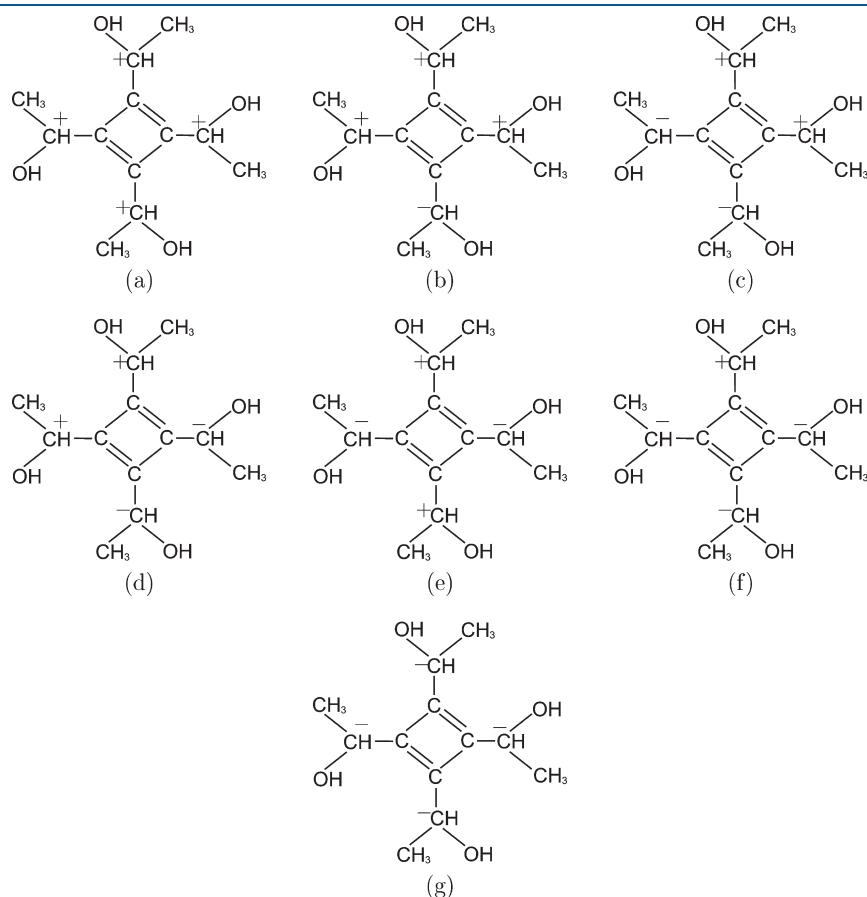


Figure 31. Stereoisomers of compounds shown in Figure S13d, Supporting Information.

For the compound shown in Figure S13h, Supporting Information, $f^*(G) < N$. The reason is again that we ignore the multiplicity within a block of the chemical graph shown in Figure S13h, Supporting Information, and thus cis–trans stereoisomers are not distinguished. However, it is known that if the length of the cycle is long enough, cis–trans isomers arise around the double bond in the chemical graph shown in Figure S13h, Supporting Information. Then, Razinger et al. obtained $N = 2$. By contrast, we compute $f^*(G) = 1$ because we ignored the multiplicity of the double bond in the chemical graph shown in Figure S13h, Supporting Information.

For the compound shown in Figure S13o, Supporting Information, $f^*(G) < N$. The reason is different from that of the above two exceptions. In this paper, we count stereoisomers generated by two different arrangements around only carbon atoms. However, in the compound shown in Figure S13h, Supporting Information, the two double bonds between a carbon atom and a nitrogen atom generate two different three-dimensional arrangements. Then it holds that $f^*(G) < N$.

For the compound shown in Figure S13n, Supporting Information, $f^*(G) = 7 > N = 5$. Razinger et al. reported that the numbers of enantiomers and diastereomers are 2 and 3. On the other hand, our proposed method outputted four enantiomers [(a) and (b), (c) and (d)] and three diastereomers as shown in Figure S15, Supporting Information, which is consistent with our definition. Though the reason of this discrepancy is unclear, we postulate that Razinger et al. might identify some stereoisomers based on mechanisms such as ring flipping.

CONCLUSION

In this paper, we designed an algorithm for generating stereoisomers of outerplanar chemical graphs based on dynamic programming. For this, we defined representations of stereoisomer, by attaching a suitable label to each vertex. For a graph with n vertices, our algorithm correctly counts the number of stereoisomers in $O(n)$ time and space and correctly outputs all possible stereoisomers in $O(n)$ space and in $O(n^3)$ time per stereoisomer on average. To our knowledge, it is the first algorithm other than our previous work¹⁴ for counting and enumerating stereoisomers with guaranteed computational complexity, though the application is limited to outerplanar chemical graphs. Furthermore, the algorithm is optimal for the problem of counting. We also conducted computational experiments to evaluate the practical performance of the algorithm. The results showed that it is significantly fast in practice. An executable program for x86/64bit Linux platforms is available at http://sunflower.kuicr.kyoto-u.ac.jp/tools/enumol/enum_outer/.

Our definition of stereoisomers is similar to that of Nourse¹⁰ in a sense that all possible configurations around asymmetric carbon atoms are to be enumerated. However, differently from his approach, our method guarantees its computational complexity by using dynamic programming and Burnside's lemma.

In this paper, we considered stereoisomers caused only by asymmetry around carbon atoms. However, the proposed techniques might be extended to enumeration of other types of stereoisomers for which stereochemical configurations depend only on local substructures. For example, if we only consider phosphorus atoms connected with four atoms (three using single bonds and one using a double bond), the method presented in this paper may be applied with a minor modification. However, there exists other configurations of phosphorus atoms. In particular,

if we consider phosphorus atoms connected with five atoms, its spatial configuration would be more complicated, and thus major modification of the method would be required. Development of recursive formulas for such configuration is far from a simple task. Therefore, extension of the proposed method to other types of atoms is left as future work.

We ignored the multiplicity within blocks of chemical graphs to define their stereoisomers in this paper. However, by considering the multiplicity within blocks, we can consider the cis–trans isomers around a chain of double bonds between two carbon atoms in cycles (see Figure 8a), though we cannot distinguish the two patterns of a resonant structure (see Figure 3). Thus the treatment of multiple bonds in cycles is another important future work, and currently partial results have been obtained by our ongoing study.¹⁸ Furthermore, it is left as a future work to extend our algorithms to a wider class of graphs, such as partial k -trees, as well as to extend to other types of stereoisomers since the dynamic programming-based approach proposed in this paper might work for some graph classes which admit compact and unique hierarchical decomposition of a graph. Another future work includes visualization and classification of chemical property of molecules generated by our algorithms.

ASSOCIATED CONTENT

S Supporting Information. Proofs of some lemmas and theorems, formal definitions of structural symmetry and orientation in outerplanar graphs, recursive formulas for the counting phase, computation process for the output phase, and figures of compounds for computational experiments are provided for describing the details. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: nag@amp.i.kyoto-u.ac.jp; takutsu@kuicr.kyoto-u.ac.jp.

ACKNOWLEDGMENT

This work was supported in part by Grant-in-Aid for Scientific Research (KAKENHI) no. 22240009 from MEXT, Japan.

REFERENCES

- (1) Cayley, A. On the Mathematical Theory of Isomers. *Philos. Mag.* **1874**, *47*, 444–446.
- (2) Pólya, G. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen. *Acta Math.* **1937**, *68*, 145–254.
- (3) Faulon, J.-L.; Visco, D. P., Jr.; Roe, D. Enumerating Molecules. *Rev. Comput. Chem.* **2005**, *21*, 209–286.
- (4) Gugisch, R.; Rücker, C. Unified Generation of Conformations, Conformers, and Stereoisomers: A Discrete Mathematics-based Approach. *MATCH Commun. Math. Comput. Chem.* **2009**, *61*, 117–148.
- (5) Mislow, K.; Siegel, J. Stereoisomerism and Local Chirality. *J. Am. Chem. Soc.* **1984**, *106*, 3319–3328.
- (6) O'Loane, J. K. Optical Activity in Small Molecules, Nonenantiomorphous Crystals, and Nematic Liquid Crystals. *Chem. Rev.* **1980**, *80*, 41–61.
- (7) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S. Generation of Stereoisomeric Structures Using Topological Information Alone. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216–219.
- (8) Contreras, M. L.; Alvarez, J.; Guajardo, D.; Rozas, R. Algorithm for Exhaustive and Nonredundant Organic Stereoisomer Generation. *J. Chem. Inf. Model.* **2006**, *46*, 2288–2298.

- (9) Wieland, T.; Kerber, A.; Laue, R. Principles of the Generation of Constitutional and Configurational Isomers. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 413–419.
- (10) Nourse, J. G. The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification, and Enumeration. *J. Am. Chem. Soc.* **1979**, *101*, 1210–1216.
- (11) Dress, A.; Dreiding, A.; Haegi, H. Classification of Mobile Molecule by Category Theory. *Stud. Phys. Theor. Chem.* **1983**, *23*, 39–58.
- (12) Tratch, S. S.; Molchanova, M. S.; Zefirov, N. S. A Unified Approach to Characterization of Molecular Composition, Connectivity, and Configuration: Symmetry, Chirality, and Generation Problems for the Corresponding Combinatorial Objects. *MATCH Commun. Math. Comput. Chem.* **2009**, *61*, 217–266.
- (13) Rücker, C.; Gugisch, R.; Kerber, A. Manual Construction and Mathematics- and Computer-Aided Counting of Stereoisomers. The Example of Oligoinositols. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1654–1665.
- (14) Imada, T.; Ota, S.; Nagamochi, H.; Akutsu, T. Efficient Enumeration of Stereoisomers of Tree Structured Molecules Using Dynamic Programming. *J. Math. Chem.* **2011**, *49*, 910–970.
- (15) Mitchell, S. L. Linear Algorithms to Recognize Outerplanar and Maximal Outerplanar Graphs. *Inform. Proc. Lett.* **1979**, *9*, 229–232.
- (16) Horvath, T.; Akutsu, T.; Wrobel, S. A Refinement Operator for Outerplanar Graphs. In *Proceedings of the 16th International Conference of Inductive Logic Programming*, Technical Report; University of Coruna: Coruna, Spain, 2006; pp 95–97.
- (17) Rücker, C.; Meringer, M. How Many Organic Compounds are Graph-Theoretically Nonplanar? *MATCH Commun. Math. Comput. Chem.* **2002**, *45*, 153–172.
- (18) Imada, T. *Enumeration of Stereoisomers of Outerplanar Chemical Graphs Using Dynamic Programming*, Master Thesis; Graduate School of Informatics, Kyoto University: Kyoto, Japan; February 10, 2011.
- (19) Aho, A. V.; Hopcraft, J. E.; Ullman, J. D. *The Design and Analysis of Computer Algorithms*; Addison-Wesley: Reading, MA, 1974.
- (20) Burnside, W. *Theory of Groups of Finite Order*; Cambridge University Press: Cambridge, U.K., 1897.
- (21) Jordan, C. Sur Les Assemblages De Lignes. *J. Reine Angew. Math.* **1869**, *70*, 185–190.
- (22) Imada, T.; Nagamochi, H. *Indexing All Rooted Subgraphs of a Rooted Graph*, Technical Report 2011-02; Graduate School of Informatics, Kyoto University: Kyoto, Japan, 2011; http://www-or.amp.i-kyoto-u.ac.jp/members/nag/Technical_report/TR2011-002.pdf.
- (23) Manning, J.; Atallah, M. J. Fast Detection and Display of Symmetry in Outerplanar Graphs. *Discrete Appl. Math.* **1992**, *39*, 13–35.
- (24) Razinger, M.; Balasubramanian, K.; Perdih, M.; Munk, M. E. Stereoisomer Generation in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 812–825.
- (25) Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for Representation and Analysis of Molecular Networks Involving Diseases and Drugs. *Nucleic Acids Res.* **2010**, *38*, D335–D360.