with their anions, and the CTs have to be corrected as, according to the rules of the General Chemical Code, these ions are not compounds, as is the case in examples (b and c) and 7 (b). This leads to two coding sheets for each ion.

In example 6 the Br^- ion must be blocked for conversion. The coding of the salt is carried out in column 26 of the Ringcode. The same occurs in examples 3 and 7.

Example 8 is an adduct of two compounds. The CT contains the information of both under the same registry number. Two Ringcode codings with the same registry number are necessary.

Heterocycles, such as the one shown in example 9a, must be coded only in their enolized form (9b).

To consider remarks a–d accordingly, a structure analysis has to be undertaken before conversion. This will be carried out by some subroutine programs required.

THE PROGRAMS

The basis for the conversion programs was the newest edition of the Ringcode coding rules. Because the program cannot distinguish between general chemical structures and special compounds, as steroids, peptides, carbohydrates, and inorganic compounds, all these are converted into the general chemical code.

The program cannot handle compounds with more than nine-membered rings. These compounds are, according to the GREMAS rules[5] which are applied to the CT generation, not assigned to a so-called ring center. Without an atom-by-atom search, such compounds cannot be identified as rings and are handled as noncyclic structures. The conversion of ring systems with several endo bridged sometimes will lead to faulty codings in columns 2–10 and 25, the reason for this again being the different rules for ring definitions of Ringcode and GREMAS. Figure 4a shows the compound adamantane which contains according to GREMAS rules seven rings; according to Ringcode rules, only three rings. The compound shown in Figure 4b contains nine rings according to the first code and six according to the second.

If such codings are converted, the program prints, besides a warning message, the converted Ringcode positions additionally with the CT and the corresponding structure diagram.

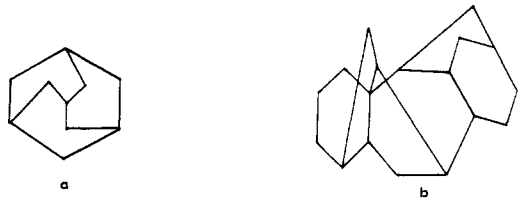

Figure 4.

A further coding error should be indicated. To determine the number of heteroatoms in rings, four counters are used: one for oxygen, one for nitrogen, one for sulfur, and one for the remaining heteroatoms. If a heterocyclic ring contains no N, O, or S-atoms but two other different heteroatoms, the fragment for "several-like X in rings" (pos. 7/11) is erroneously converted, instead of the fragment for "several different X in rings" (pos. 7/0). Because of the rareness of such structures, this error may be acceptable.

Another inaccuracy occurs during the conversion of the keto groups of tetracyclines. According to the Ringcode rules, the fragment "POLY" (pos. 25/12) for cyclic keto groups is not allowed. The described program cannot consider this, and generates pos. 25/12.

In consideration of the aforementioned details, there are probably no further restrictions for the application of this program package. Several error and warning messages are provided in the different subroutines which will lead to the printout of CTs, Ringcodes, and structure diagrams to cover those possibilities not considered during the programming work.

REFERENCES AND NOTES

(1) W. Nübling and W. Steidle, *Angew. Chem.*, **82**, 618 (1970).
(2) E. Meyer, Aufbau der BASF-Hauptmatrix, Ludwigshafen, BASF, 1969. BASF = Badische Anilin & Soda-Fabrik AG.
(3) E. Meyer, *Angew. Chem.*, **82**, 605 (1970).
(4) Ringcode, General Chemical Code, 5th ed., April, 1977, Pharma Dokumentationsring e.V. and Derwent Publications Ltd.
(5) S. Rössler and A. Kolb, *J. Chem. Doc.*, **10**, 128 (1970).

Algorithms for Unique and Unambiguous Coding and Symmetry Perception of Molecular Structure Diagram. I. Vector Functions for Automorphism Partitioning†

MASAHIRO UCHINO

Laboratory of Resources Utilisation, Tokyo Institute of Technology, 4259 Nagatsudacho, Midoriku, Yokohama, 227 Japan

Received June 4, 1979; Revised Manuscript Received January 21, 1980

Several vector functions which are useful for automorphism partitioning of the set of vertices in a chemical graph are proposed on the basis of mathematical theorems. Proofs are given for the theorems used and details of the methods to employ these functions are described. The present approach provides more systematic and powerful methods for a partitioning procedure than existing algorithms (Morgan, Corneil–Gotlieb).

INTRODUCTION

Problems such as unique and unambiguous coding,[1-11] automorphism partitioning,[12] and computation of symmetry groups[11,14] often require preliminary partitioning of vertices of a chemical graph. Morgan's connectivity value[5] approach and Corneil–Gotlieb's Algorithm I[12] are two well-known methods which can be used. These existing methods, however, are entirely ineffective for partitioning of the vertices of regular graphs and poorly effective for several types of graphs.

In this paper, a matrix method[7] has been applied to the partitioning problem, and a general equation useful for de-

† Presented at the Division of Chemical Information, ACS/CSJ Congress, Honolulu, Hawaii, April 3, 1979.

0095-2338/80/1620-0116$01.00/0

signing and comparing partitioning procedures is presented. Several specific vector functions devised from the general equation have been compared with the functions corresponding to Morgan–Corneil–Gotlieb-type procedures.

I. NUMBERING INDEPENDENT PROPERTIES AND THEIR REPRESENTATIONS

It is convenient to consider that a chemical graph consists of a set $V = (v_1, v_2, ..., v_n)$ of nodes, a map g from V to the set P of positive integers which associates to each node, a node value, and a map g′: $V \times V \rightarrow$ (the set N of nonnegative integers) associating to each pair of nodes a bond value between them. The bond value 0 implies no bond and nonzero values represent the bond multiplicities. A node value vector (or atomic vector) $\boldsymbol{a}$ and a bond matrix $\mathbf{B}$ can be used for the representation of the maps g and g′: $\boldsymbol{a}_i = g(v_i)$ ($i = 1, 2, ..., n$) and $\mathbf{B}_{ij} = g'(v_i,v_j)$ ($i = 1, 2, ..., n$; $j = 1, 2, ..., n$). Maps f ($V \rightarrow P$) and f′ ($V \times V \rightarrow N$) define other numbering-independent properties of node and node-pair, and can be represented by a vector and a matrix similarly defined.

The ways of assigning numbers to graph nodes can be related with the group F of n-element permutation vectors whose binary operation is defined by $\boldsymbol{f}\circ\boldsymbol{f}' = (\boldsymbol{f}_{f'_1}, \boldsymbol{f}_{f'_2}, ..., \boldsymbol{f}_{f'_n})$. The identity element of the group F is given by $\boldsymbol{e} = (1, 2, ..., n)$. The numbering corresponding to an element $\boldsymbol{f} \in F$ can be obtained by assigning the number i to the $\boldsymbol{f}_i$th node of a graph (or by assigning the number $\boldsymbol{f}_i^{-1}$ to the ith node). The successive renumbering process with $\boldsymbol{f}$ followed by $\boldsymbol{f}'$ is equivalent to the number by $\boldsymbol{f}\circ\boldsymbol{f}'$.

If a vector $\boldsymbol{p}$ and a matrix $\mathbf{E}$ are the representations of numbering independent properties of node and node pair, a numbering by $\boldsymbol{f}$ changes these representations respectively to $\boldsymbol{p}\circ\boldsymbol{f} = (\boldsymbol{p}_{f_1}, \boldsymbol{p}_{f_2}, ..., \boldsymbol{p}_{f_n})$ and $\mathbf{E}\circ(\boldsymbol{f};\boldsymbol{f})$, where the (i,j)th component of $\mathbf{E}\circ(\boldsymbol{f};\boldsymbol{f})$ is the $(\boldsymbol{f}_i,\boldsymbol{f}_j)$th component of $\mathbf{E}$.

Suppose we have a node property vector $\boldsymbol{p}$ and r matrices of node-pair properties, $\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)}$. A function $h(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})$ whose value is a n-element vector of positive integers defines the representation of a numbering independent property of node if the following holds:

$$h(\boldsymbol{p}\circ\boldsymbol{f}, \mathbf{E}^{(1)}\circ(\boldsymbol{f};\boldsymbol{f}), \mathbf{E}^{(2)}\circ(\boldsymbol{f};\boldsymbol{f}), ..., \mathbf{E}^{(r)}\circ(\boldsymbol{f};\boldsymbol{f})) = h(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})\circ\boldsymbol{f} \quad (1)$$

for any $\boldsymbol{f} \in F$. Note that if eq 1 holds, we can assign to each node a unique value, independently of the numberings of nodes, by assigning the ith component of the function value to the ith node.

A function $h(\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})$ will be called the cumulative vector function when the following holds: if $\boldsymbol{p}_i \neq \boldsymbol{p}_j$, then

$$h(\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})_i \neq h(\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})_j$$

Morgan's connectivity value can be given as the component of a vector function $h(\boldsymbol{p},\mathbf{C}) = \mathbf{C}\boldsymbol{p}$, where $\mathbf{C}$ is the adjacency matrix. This function is not cumulative and the use of nonculmulative function is related to the oscilatory behavior observed by Randić.[9] The design of good partitioning procedure is related to the device of a cumulative vector function satisfying eq 1 and being invariant under action of symmetry group.

II. INVARIANT PROPERTIES UNDER ACTION OF A SYMMETRY GROUP

An atomic vector $\boldsymbol{a}$ and a bond matrix $\mathbf{B}$ represent a chemical graph and the adjacency matrix $\mathbf{C}$ is obtained from the bond matrix $\mathbf{B}$ by replacing the nonzero elements of $\mathbf{B}$ with 1's. The symmetry characteristics of a chemical graph are defined by the subgroups of F, associated with $\boldsymbol{a}$, $\mathbf{B}$, and $\mathbf{C}$:

(i) Atomic symmetry $G^{(\mathbf{a})}$

$$G^{(\mathbf{a})} = \{\boldsymbol{f} \in F;\ \boldsymbol{a}\circ\boldsymbol{f} = \boldsymbol{a}\}$$

(ii) Bond symmetry $G^{(\mathbf{B})}$

$$G^{(\mathbf{B})} = \{\boldsymbol{f} \in F;\ \mathbf{B}\circ(\boldsymbol{f};\boldsymbol{f}) = \mathbf{B}\}$$

(iii) Topological symmetry $G^{(\mathbf{C})}$

$$G^{(\mathbf{C})} = \{\boldsymbol{f} \in F;\ \mathbf{C}\circ(\boldsymbol{f};\boldsymbol{f}) = \mathbf{C}\}$$

A node property vector $\boldsymbol{p}$ and a node-pair property matrix $\mathbf{E}$ will be said to be invariant under action of a symmetry group G if and only if $\boldsymbol{p}\circ\boldsymbol{g} = \boldsymbol{p}$ and $\mathbf{E}\circ(\boldsymbol{g};\boldsymbol{g}) = \mathbf{E}$ for any $\boldsymbol{g} \in G$. Note that a vector function $h(\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})$ satisfying the condition 1 is invariant under G if $\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)}$ are invariant under G.

III. rTH-ORDER ILLUMINATION FUNCTION FOR A SYMMETRY GROUP G

Let $\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)}$ be invariant under action of symmetry group G. The function $h(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)})$ defined below is called rth-order illumination function for the group G:

$$h(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)}) = b^q\boldsymbol{p} + b^{q-k_1}(\mathbf{E}^{(1)}\mathbf{H})^*\boldsymbol{b} + b^{q-(k_1+k_2)}(\mathbf{E}^{(2)}\mathbf{H})^*\boldsymbol{b} + ... + b^{q-(k_1+k_2+...+k_i)}(\mathbf{E}^{(i)}\mathbf{H})^*\boldsymbol{b} + ... + ... + b^0(\mathbf{E}^{(r)}\mathbf{H})^*\boldsymbol{b} \quad (2)$$

where $\mathbf{H}$ is $n \times n$ diagonal matrix defined by $\mathbf{H}_{ii} = \boldsymbol{p}_i$ ($i = 1, 2, ..., n$) and $\boldsymbol{k}$ is the r-element vector whose ith component $\boldsymbol{k}_i$ is the maximum number of nonzero elements in rows of $\mathbf{E}^{(i)}\mathbf{H}$, and $q = \sum_{i=1}^{r}\boldsymbol{k}_i$. A scalar b is an integer constant which is greater than the maximum component of the matrices $\mathbf{E}^{(i)}\mathbf{H}$ for $i = 1, 2, ..., r$ and

$$\boldsymbol{b} = \begin{bmatrix} b^{n-1} \\ b^{n-2} \\ \vdots \\ b^0 \end{bmatrix}$$

The asterisk * of $(\mathbf{E}^{(i)}\mathbf{H})^*$ means the ordered form of the matrix, and the ordered form $\mathbf{A}^*$ of a matrix $\mathbf{A}$ is obtained by ordering the components of each row of $\mathbf{A}$ in ascending order.

Theorem I. The rth-order illumination function for a symmetry group G is invariant under G.

Proof. By the definition, $\boldsymbol{p}, \mathbf{E}^{(1)}, \mathbf{E}^{(2)}, ..., \mathbf{E}^{(r)}$ are invariant under G and we need merely show that the illumination function satisfies condition 1. As can be understood directly, the following computing laws can be used for addition, multiplication, permutation, and ordering of $n \times n$ matrices $\mathbf{X}$, $\mathbf{Y}$ and n-element vectors $\boldsymbol{x}$, $\boldsymbol{y}$:

(a) $(\mathbf{XY})\circ(\boldsymbol{f};\boldsymbol{f}') = \mathbf{X}\circ(\boldsymbol{f},\boldsymbol{f}')\mathbf{Y}\circ(\boldsymbol{f}';\boldsymbol{f}')$

(b) $(\mathbf{X}\boldsymbol{y})\circ\boldsymbol{f} = \mathbf{X}\circ(\boldsymbol{f};\boldsymbol{e})\boldsymbol{y}$

(c) $((\mathbf{XY})\circ(\boldsymbol{f};\boldsymbol{f}'))^* = (\mathbf{XY})^*\circ(\boldsymbol{f};\boldsymbol{e})$

(d) $(\boldsymbol{x} + \boldsymbol{y})\circ\boldsymbol{f} = \boldsymbol{x}\circ\boldsymbol{f} + \boldsymbol{y}\circ\boldsymbol{f}$

where $\boldsymbol{f}$, $\boldsymbol{f}'$ and $\boldsymbol{f}'' \in F$, and $\boldsymbol{e}$ is the identity. By definition,

$$h(\boldsymbol{p}\circ\boldsymbol{f}, \mathbf{E}^{(1)}\circ(\boldsymbol{f};\boldsymbol{f}), ..., \mathbf{E}^{(r)}\circ(\boldsymbol{f};\boldsymbol{f})) = b^q\boldsymbol{p}\circ\boldsymbol{f} + ... + b^{q-(k_1+k_2+...+k_i)}(\mathbf{E}^{(i)}\circ(\boldsymbol{f};\boldsymbol{f})\mathbf{H}\circ(\boldsymbol{f};\boldsymbol{f}))^*\boldsymbol{b} + ...$$

By (a) and (c),

$$= b^q\boldsymbol{p}\circ\boldsymbol{f} + ... + b^{q-(k_1+k_2+...+k_i)}(\mathbf{E}^{(i)}\mathbf{H})^*\circ(\boldsymbol{f};\boldsymbol{e})\boldsymbol{b} + ...$$

By (b) and (d),

$$= (b^q\boldsymbol{p} + ... + b^{q-(k_1+k_2+...+k_i)}(\mathbf{E}^{(i)}\mathbf{H})^*\boldsymbol{b} + ...)\circ\boldsymbol{f}$$

$$= h(\boldsymbol{p},\mathbf{E}^{(1)}, ..., \mathbf{E}^{(r)})\circ\boldsymbol{f} \text{ for any } \boldsymbol{f} \in F$$

Theorem II. The rth-order illumination function is cumulative.

Proof. Only last k_i columns of $(\mathbf{E}^{(i)}\mathbf{H})^*$ contain nonzero values. Let $\mathbf{Q}^{(i)}$ be the last k_i columns of $(\mathbf{E}^{(i)}\mathbf{H})^*$. Then,

$$b^{q-(k_1+k_2+\cdots+k_i)}(\mathbf{E}^{(i)}\mathbf{H}^*\boldsymbol{b} = \mathbf{Q}^{(i)}\begin{bmatrix} b^{q-(k_1+k_2+\cdots+k_i)+k_i-1} \\ b^{q-(k_1+k_2+\cdots+k_i)+k_i-2} \\ \vdots \\ b^{q-(k_1+k_2+\cdots+k_i)} \end{bmatrix}$$

To simplify the notation, we define a matrix $\mathbf{Q} = [\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \ldots, \mathbf{Q}^{(r)}]$. Then, the jth component of the function is given by:

$$\mathrm{h}(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, \ldots, \mathbf{E}^{(r)})_j = b^q p_j + \mathbf{Q}_{j1}b^{q-1} + \mathbf{Q}_{j2}b^{q-2} + \ldots \mathbf{Q}_{jq}$$

Assume $p_j \neq p_{j'}$ and $\mathrm{h}(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, \ldots, \mathbf{E}^{(r)})_j = \mathrm{h}(\boldsymbol{p},\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, \ldots, \mathbf{E}^{(r)})_{j'}$.
Then

$$(p_j - p_{j'})b^q + (\mathbf{Q}_{j1} - \mathbf{Q}_{j'1})b^{q-1} + \ldots + (\mathbf{Q}_{jq} - \mathbf{Q}_{j'q}) = 0$$

Hence

$$|(p_j - p_{j'})b^q| \leq |(\mathbf{Q}_{j1} - \mathbf{Q}_{j'1})b^{q-1}| + \ldots + |\mathbf{Q}_{jq} - \mathbf{Q}_{j'q}|$$

and

$$|(p_j - p_{j'})b^q| \leq b^q - 1, \text{ since} |\mathbf{Q}_{jm} - \mathbf{Q}_{j'm}| < b$$

Since $\boldsymbol{p}$ is an integer vector, this is a contradiction. And hence the rth-order illumination function is cumulative.

A selection of invariant matrices $\mathbf{E}^{(1)}, \mathbf{E}^{(2)}, \ldots, \mathbf{E}^{(r)}$ for a symmetry group G provides a specific illumination function for this group by virtue of eq 2. For instance, if $r = 1$ and $\mathbf{E}^{(1)} = \mathbf{C}$, then $\mathrm{h}(\boldsymbol{p},\mathbf{C}) = b^q\boldsymbol{p} + (\mathbf{CH})^*\boldsymbol{b}$ is obtained as a first-order illumination function for topological symmetry $G^{(\mathbf{C})}$. With $\mathbf{C}$ fixed, we denote the function by $\mathrm{h}(\boldsymbol{p})$ and make a composition $\mathrm{h}(\ldots(\mathrm{h}(\mathrm{h}(\boldsymbol{p})))\ldots)$; then partitioning of nodes by the component values of the vector produced from this composite corresponds to the procedure(Algorithm I) of Corneil–Gotlieb.[12] The cumulativeness of the function which theorem II assures is an important property in relation with such composition of maps. When we neglect the first term of $\mathrm{h}(\boldsymbol{p},\mathbf{C})$ and put $\boldsymbol{b} = (1, 1, \ldots, 1)^T$, $\mathrm{h}(\boldsymbol{p}) = \mathbf{C}\boldsymbol{p}$ which corresponds to Morgan's connectivity value, but this is not cumulative and is no more an illumination function.

Obviously a composition fgh of illumination functions f, g, and h for a symmetry group G is cumulative and invariant under G. If f is an illumination function for bond symmetry $G^{(\mathbf{B})}$, then the value $\mathrm{f}(\boldsymbol{a})$ of atomic vector $\boldsymbol{a}$ is invariant vector under the intersection group $G^{(\mathbf{B})} \cap G^{(\mathbf{a})}$, and we describe in next section several efficient functions for symmetry group $G^{(\mathbf{B})}$.

IV. USEFUL INVARIANT MATRICES UNDER SYMMETRY GROUPS AND COMPUTATION OF INVARIANT VECTORS UNDER $G^{(\mathbf{B})}$

An adjacency matrix $\mathbf{C}$ is invariant under $G^{(\mathbf{B})}$, since symmetry group $G^{(\mathbf{B})}$ is a subgroup of $G^{(\mathbf{C})}$. Two types of the sequence of matrices which are invariant under $G^{(\mathbf{C})}$ and $G^{(\mathbf{B})}$ can be derived from an adjacency matrix $\mathbf{C}$ by the use of Boolean multiplication of $n \times n$ (0,1) matrices:

(i) Matrices of Open Walks, $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(m)}$. Suppose a sequence of $n \times n$ (0,1) matrices $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \ldots, \mathbf{W}^{(m)}$ defined as:

$$\mathbf{W}^{(1)} = \mathbf{C}, \qquad \mathbf{W}^{(2)} = (\mathbf{CW}^{(1)})^{\theta},$$
$$\ldots, \mathbf{W}^{(m)} = (\mathbf{CW}^{(m-1)})^{\theta}$$

Here the matrix multiplication is a Boolean product and the superscript θ indicates clearing of the prinicpal diagonal elements of the matrix. Then matrices $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(m)}, \ldots$ are defined by:

$$\mathbf{V}^{(1)} = \mathbf{W}^{(1)}, \mathbf{V}^{(2)} = \mathbf{W}^{(1)} \wedge \mathbf{W}^{(2)}, \ldots, \mathbf{V}^{(m)} = \mathbf{W}^{(m)} \wedge \overline{\mathbf{W}^{(m-1)}}$$

where the overbar indicates the negation, and the component-by component "anding" operation is indicated by $\wedge$.

When $\mathbf{V}_{ij}^{(m)} = 1$, this indicates the existence of certain open walks of length m from the ith node to the jth node. But this interpretation is not a requisite for the use of such matrices to construct the illumination function. We need merely show that these matrices are invariant under $G^{(\mathbf{C})}$.

Theorem III. Matrices $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(m)}$ are invariant under $G^{(\mathbf{C})}$.

Proof. It is obvious that the following computation laws can be used for the permutation of $n \times n$ (0,1) matrices $\mathbf{X}$ and $\mathbf{Y}$; $(\mathbf{X} \wedge \mathbf{Y})_\circ(\boldsymbol{f};\boldsymbol{f}) = \mathbf{X}_\circ(\boldsymbol{f};\boldsymbol{f}) \wedge \mathbf{Y}_\circ(\boldsymbol{f};\boldsymbol{f})$, $\bar{\mathbf{X}}_\circ(\boldsymbol{f};\boldsymbol{f}) = \overline{\mathbf{X}_\circ(\boldsymbol{f};\boldsymbol{f})}$, and $(\mathbf{XY}^{\theta})_\circ(\boldsymbol{f};\boldsymbol{f}) = (\mathbf{X}_\circ(\boldsymbol{f};\boldsymbol{f})\mathbf{Y}_\circ(\boldsymbol{f};\boldsymbol{f}))^{\theta}$, where $\boldsymbol{f}$ is any member of F. By the first and second laws, the matrices $\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(m)}$ are invariant if $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \ldots, \mathbf{W}^{(m)}$ are invariant. The matrix $\mathbf{W}^{(1)}$ is invariant under $G^{(\mathbf{C})}$ by definition, and if we assume $\mathbf{W}^{(m-1)}$ is invariant, then $\mathbf{W}^{(m)}$ is invariant by the third law.

(ii) Matrices $\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \ldots, \mathbf{D}^{(m)}$ of Distance. The matrices $\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \ldots, \mathbf{D}^{(m)}$ are $n \times n$ (0,1) matrices defined by $\mathbf{D}_{ij}^{(m)} = 1$ if and only if the distance from the ith node to the jth node is equal to m. It is easily shown by the method similar to (i) that these matrices are invariant under $G^{(\mathbf{C})}$.

We now present several methods to compute useful invariant vectors under $G^{(\mathbf{B})}$ by the composite of the specific illumination functions obtained from the matrices described in (i) and (ii) using the general equation (2).

Method I. Compute a composite of the first-order illumination functions $\mathrm{h}(\boldsymbol{p},\mathbf{B})$, $\mathrm{h}(\boldsymbol{p},\mathbf{V}^{(2)})$, $\mathrm{h}(\boldsymbol{p},\mathbf{V}^{(3)}), \ldots, \mathrm{h}(\boldsymbol{p},\mathbf{V}^{(m)})$, that is, $\mathrm{g}(\boldsymbol{p},m) = \mathrm{h}(\ldots\mathrm{h}(\mathrm{h}(\mathrm{h}(\boldsymbol{p},\mathbf{B}),\mathbf{V}^{(2)}),\mathbf{V}^{(3)}),\mathbf{V}^{(4)}),\ldots,\mathbf{V}^{(m)})$ for $\boldsymbol{p} = (1,1,1,1,\ldots,1)$ and $m = m^*$ determined by the following: If the number of distinct elements of $\mathrm{g}(\boldsymbol{p},m)$ does not increase and $\mathbf{V}^{(m)}$ does not become a zero matrix until $m = 8$, then $m^* = 7$. Otherwise, m^* is the minimum integer m such that $\mathbf{V}^{(m+1)}$ is a zero matrix or the number of distinct elements of $\mathrm{g}(\boldsymbol{p},m)$ is equal to that of $\mathrm{g}(\boldsymbol{p},m+1)$ and different from that of $\mathrm{h}(\boldsymbol{p},\mathbf{B})$.

Method II. Employ matrices $\mathbf{D}^{(2)}, \mathbf{D}^{(3)}, \ldots, \mathbf{D}^{(m)}$ instead of $\mathbf{V}^{(2)}, \mathbf{V}^{(3)}, \ldots, \mathbf{V}^{(m)}$ of method I.

Method III. Compute a composite $\mathrm{g}^{m^*}(\boldsymbol{p})$ of a single second-order illumination function $\mathrm{g}(\boldsymbol{p}) = \mathrm{h}(\boldsymbol{p},\mathbf{B},\mathbf{V}^{(2)})$, i.e., $\mathrm{g}^{m^*}(\boldsymbol{p}) = \mathrm{h}(\ldots\mathrm{h}(\mathrm{h}(\boldsymbol{p},\mathbf{B},\mathbf{V}^{(2)}),\mathbf{B},\mathbf{V}^{(2)}),\ldots,\mathbf{B},\mathbf{V}^{(2)})$, for $\boldsymbol{p} = (1,1,1,\ldots,1)$ and m^* defined by the following: m^* is the minimum integer m such that the number of distinct elements of $\mathrm{g}^m(\boldsymbol{p})$ is equal to that of $\mathrm{g}^{m+1}(\boldsymbol{p})$.

Method IV. Employ a first-order illumination function $\mathrm{h}(\boldsymbol{p},\mathbf{B})$ instead of $\mathrm{h}(\boldsymbol{p},\mathbf{B},\mathbf{V}^{(2)})$ of method III.

Method V. Employ the illumination functions of variable orders, i.e., $\mathrm{h}(\boldsymbol{p},\mathbf{B})$, $\mathrm{h}(\boldsymbol{p},\mathbf{D}^{(2)})$, $\mathrm{h}(\boldsymbol{p},\mathbf{D}^{(2)},\mathbf{D}^{(3)})$, $\mathrm{h}(\boldsymbol{p},\mathbf{D}^{(2)},\mathbf{D}^{(3)},\mathbf{D}^{(4)})\ldots$, $\mathrm{h}(\boldsymbol{p},\mathbf{D}^{(2)},\mathbf{D}^{(3)},\mathbf{D}^{(4)},\ldots,\mathbf{D}^{(m)})$ instead of the first-order functions of method I.

In methods I–V, we must compute the value of the composition, say $\mathrm{ff}'\ldots\mathrm{f}''(\boldsymbol{p})$ of functions f, f′, ..., f″. The component values of the invariant vector computed by such composites are sometimes too large for the range of the machine number representation, and in such a case, a simple function $\mathrm{r}(\boldsymbol{p})$ which changes the component values of $\boldsymbol{p}$ to the relative ranks in $\boldsymbol{p}$ should follow each illumination function, and in this way, we compute $\mathrm{rfrf}'\ldots\mathrm{rf}''(\boldsymbol{p})$ instead of $\mathrm{ff}'\ldots\mathrm{f}''(\boldsymbol{p})$. Function $\mathrm{r}(\boldsymbol{p})$ can be given by the following: Let $\hat{\boldsymbol{p}}$ denote the vector obtained from $\boldsymbol{p}$ by removing the duplicated elements and by ordering the remaining elements in ascending order. Then, $\mathrm{r}(\boldsymbol{p})_i = j$ such that $p_i = \hat{p}_j$, and $i = 1, 2, \ldots, n$.

J. Chem. Inf. Comput. Sci., Vol. 20, No. 2, 1980

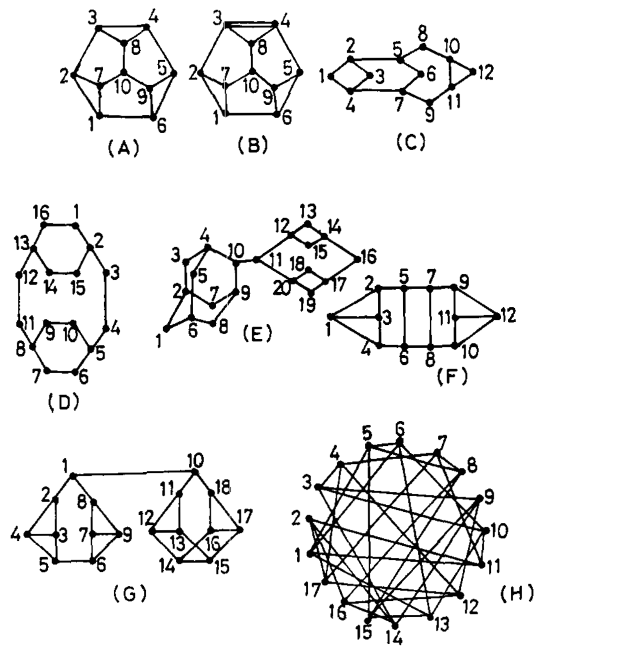

Figure 1. Some structures for which Morgan–Corneil–Gotlieb-type procedures fail to give the automorphism partitioning.

Table I. Invariant Vectors Produced by the Methods I, III, and V

meth-od	struc-ture[a]	m*	invariant vector under bond symmetry[b]
I	(A)	3	(1,1,1,1,1,1,2,2,2,3)
	(B)	3	(2,3,6,6,3,2,4,1,4,5)
	(C)	3	(2,6,2,6,7,4,7,3,3,5,5,1)
	(D)	3	(1,3,2,2,3,1,1,3,1,1,2,2,3,1,1,1)
	(E)	3	(3,6,4,8,4,6,4,4,8,9,10,7,1,5,1,2,5,1,1,7)
	(F)	2	(1,2,1,2,3,3,3,3,2,2,1,1)
	(G)	3	(8,4,1,1,2,2,1,4,1,7,6,3,3,5,5,3,3,6)
	(H)	3	(3,10,15,8,6,1,12,17,11,14,16,5,2,7,13,9,4)
III	(A)	2	(1,1,1,1,1,1,2,2,2,3)
	(B)	2	(1,3,6,6,3,1,2,4,2,5)
	(C)	2	(2,6,2,6,7,4,7,3,3,5,5,1)
	(D)	1	(1,2,1,1,2,1,1,2,1,1,1,1,2,1,1,1)[c]
	(E)	2	(3,7,4,8,4,7,4,4,8,10,9,6,1,5,1,2,5,1,1,6)
	(F)	1	(1,2,1,2,3,3,3,3,2,2,1,1)
	(G)	2	(7,3,1,1,2,2,1,3,1,8,6,4,4,5,5,4,4,6)
	(H)	2	(2,12,16,7,4,1,14,17,11,9,15,5,3,8,10,13,6)
V	(A)	3	(2,2,2,2,2,2,1,1,1,3)
	(B)	3	(4,3,6,6,3,4,2,1,2,5)
	(C)	3	(2,6,2,6,7,4,7,3,3,5,5,1)
	(D)	3	(1,3,2,2,3,1,1,3,1,1,2,2,3,1,1,1)
	(E)	3	(3,7,4,8,4,7,4,4,8,10,9,6,1,5,1,2,5,1,1,6)
	(F)	2	(1,2,1,2,3,3,3,3,2,2,1,1)
	(G)	3	(7,3,1,1,2,2,1,3,1,8,6,4,4,5,5,4,4,6)
	(H)	3	(3,13,16,8,4,1,11,17,10,12,15,5,2,7,9,14,6)

[a] See Figure 1. [b] Invariant vector is based on the numbering shown in Figure 1. [c] Incomplete partitioning.

Since the function r($\boldsymbol{p}$) is cumulative and invariant if $\boldsymbol{p}$ is invariant, the modification of the methods is permissible for the purpose of computing an invariant vector under $G^{(\mathbf{B})}$.

V. COMPARISON OF THE METHODS

Morgan's connectivity value approach[5] and Corneil–Gotlieb's Algorithm I[12] fail to give the automorphism partitioning for the structures shown in Figure 1, which includes some structures described by Carhart[16] and Mackay[17] as counter-examples for the reported procedures[10,15,16] of unique coding.

Methods I–V were all found to be more effective than the Morgan–Corneil–Gotlieb-type procedure. Method IV, the simplest of the five methods, is similar to the matrix formulation of the Corneil–Gotlieb's procedure, but is more effective for structures containing unsaturated bonds and provides automorphism partitioning for structure B, although this method was ineffective for other structures. Method I and method V provided automorphism partitioning for all structures in Figure 1, and, in fact, we know of no example for which these methods fail to give automorphism partitioning. Method III shows improved performance for regular graphs in comparison with method IV by only one additional Boolean matrix multiplication step, and can be used as a better substitute for Morgan–Corneil–Gotlieb-type procedures. Method II is less effective than method I and fails for structures A and H; it is instructive to compare these methods for structure A. In both methods, the central 10th node is first discriminated, but in method II this discrimination does not propagate since there is no node which is at distance 3 from the 10th node. This defect of method II, which uses the matrix of distance, is removed in method V. In principle, method V is the most powerful method of the five methods, but this requires more storage and computation time than method I which uses the matrix of open walks. Table I summarizes the results obtained by methods I, III, and V.

```
      ∇PARTITION1[⎕]∇
    ∇ HM←PARTITION1 B;N;C;K;BS;EA;HM1;EM
      ;WM;WMM;EX;V;SET
[1]   N←1↑ρC←B≠0
[2]   K←⌈/+/[2] C
[3]   BS←1+⌈/,B←(N,-K)↑ASTERISK B
[4]   BS←1+E0←ΔE←ρSET←SETSELECT HM←BS⊥⍉B
[5]   HM←SET[⍋SET]⍳HM
[6]   WMM←C
[7]   COUNT←2
[8]  LW:WM←,C∨.∧WMM
[9]   WM[(⍳N)+N×(¯1+⍳N)]←Nρ0
[10]  WM←(N,N)ρWM
[11]  EM←WM∧~WMM
[12]  WMM←WM
[13]  K←⌈/V←+/[2] EM
[14]  →(K=0)/0
[15]  EX←(~EX),[2] EX←(N,K)ρ1
[16]  EX←,(N,K)↑EX←VΦ[2] EX
[17]  V←(N,K)ρEX\(,EM)/(N×N)ρHM
[18]  V←ASTERISK V
[19]  HM1←(HM×BS*K)+BS⊥⍉V
[20]  EA←ρSET←SETSELECT HM1
[21]  →((ΔE=EA)∧(EA≠E0))/0
[22]  ΔE←EA
[23]  COUNT←COUNT+1
[24]  →((EA=E0)∧COUNT>7)/0
[25]  HM←SET[⍋SET]⍳HM1
[26]  BS←1+EA
[27]  →LW
    ∇
```

Figure 2. APL function for method I.

```
      ∇PARTITION3[⎕]∇
    ∇ HM←PARTITION3 B;N;V2;C;K1;K2;R1;BS;MAX;EX1
      ;EX2;HM;HM1; VV1;VV2;E0;EA
[1]   N←1↑ρC←B≠0
[2]   V2←,(~C)∧C∨.∧C
[3]   V2[(⍳N)+N×(¯1+⍳N)]←Nρ0
[4]   K2←⌈/VV2←+/[2](N,N)ρV2
[5]   BS←1+MAX←⌈/,R1←(,C)/,B
[6]   K1←⌈/VV1←+/[2] C
[7]   EX1←(~EX1),[2] EX1←(N,K1)ρ1
[8]   EX1←,(N,K1)↑EX1←VV1Φ[2] EX1
[9]   EX2←(~EX2),[2] EX2←(N,K2)ρ1
[10]  EX2←,(N,K2)↑EX2←VV2Φ[2] EX2
[11]  HM←NρΔE←1
[12] L:HM1←R1×(,C)/(N×N)ρHM
[13]  HM1←ASTERISK(N,K1)ρEX1\HM1
[14]  HM1←HM1,[2] ASTERISK(N,K2)ρEX2\V2/(N×N)ρHM
[15]  HM1←((BS*K1+K2)×HM)+BS⊥⍉HM1
[16]  SET←SETSELECT HM1
[17]  EA←ρSET
[18]  →(ΔE=EA)/0
[19]  ΔE←EA
[20]  HM←SET[⍋SET]⍳HM1
[21]  BS←1+EA×MAX
[22]  →L
    ∇
```

Figure 3. APL function for method III.

```
      ∇PARTITION5[⎕]∇
    ∇ HM←PARTITION5 B;N;C;K;K1;BS;EA;HM1
      ;DM;DA;D1;EX;V;V1;VM;SET;Q;I
[1]   N←1↑⍴DA←C←B≠0
[2]   K1←⌈/+/[2] C
[3]   BS←1+⌈/,B←(N,-K1)↑ASTERISK B
[4]   BS←1+E0←ΔE←⍴SET←SETSELECT HM←BS↓⍳⍴B
[5]   HM←SET[⍋SET]↓HM
[6]   COUNT←2
[7]   K←↓0
[8]   V←(0,N)⍴↓0
[9]   DM←(0,N,N)⍴↓0
[10] LO:D1←C∨.∧DA
[11]  T←D1∧~DA
[12]  DM←DM,[1] T
[13]  DA←D1∨DA
[14]  K1←⌈/V1←+/[2] T
[15]  V←V,[1] V1
[16]  K←K,K1
[17]  Q←+/K
[18]  HM1←HM×BS*Q
[19]  I←0
[20] LO:→((⍴K)<I←I+1)/NXT
[21]  K1←K[I]
[22]  →(K1=0)/LO
[23]  V1←V[I;]
[24]  T←DM[I;;]
[25]  EX←(~EX),[2] EX←(N,K1)⍴1
[26]  EX←,(N,K1)↑EX←V1⌿[2] EX
[27]  VM←(N,K1)⍴EX\(,T)/(N×N)⍴HM
[28]  VM←ASTERISK VM
[29]  HM1←HM1+(BS×(Q-+/I↑K))×BS↓⍳⍴VM
[30]  →LO
[31] NXT:EA←⍴SET←SETSELECT HM1
[32]  →((ΔE=EA)∧(EA≠E0))/0
[33]  ΔE←EA
[34]  COUNT←COUNT+1
[35]  →((EA=E0)∧COUNT>7)/0
[36]  HM←SET[⍋SET]↓HM1
[37]  BS←1+EA
[38]  →LO
      ∇
```

Figure 4. APL function for method V.

```
      ∇ASTERISK[⎕]∇
    ∇ R←ASTERISK M;I;N
[1]   N←1↑⍴M
[2]   I←0
[3]  L:→(N<I←I+1)/OUT
[4]   M[I;]←M[I;]⍋M[I;]
[5]   →L
[6]  OUT:R←M
      ∇


      ∇SETSELECT[⎕]∇
    ∇ Z←SETSELECT V
[1]   Z←((↓⍴V)=V↓V)/V←,V
      ∇
```

Figure 5. APL functions for the ordered form of matrix and set-selection.

VI. APPENDIX. PROGRAMS

The methods described in this paper have been implemented on an IBM 5110 and on a MELCOM COSMO 700II (Mitsubishi Electric Corp.) in APL. The APL programs for methods I, III, and V are shown respectively in Figures 2, 3, and 4.

These programs compute for a bond matrix **B** the invariant vector under bond symmetry $G^{(\mathbf{B})}$. Two simple functions ASTERISK and SETSELECT are required for these programs and are given in Figure 5. The function ASTERISK computes the ordered form **M*** of a matrix **M**. And the function SETSELECT removes the duplicated elements from a vector **V** and provides the vector of distinct elements.

Table II. Explanation of Arrays Used in the Programs

Partition1	
WM	$\mathbf{W}^{(m)}$
WMM	$\mathbf{W}^{(m-1)}$
EM	$\mathbf{V}^{(m)}$
HM	p
Partition3	
V2	$\mathbf{V}^{(2)}$
HM	p
Partition5	
DM	Three-dimensional array which accomodates $\mathbf{D}^{(2)}$, $\mathbf{D}^{(3)}$, . . ., $\mathbf{D}^{(m)}$
HM	p

In the programs of Figures 2–4, the computation of the term $(\mathbf{E}^{(i)}\mathbf{H})^*$ in the general equation (2) of section III is not carried out directly to avoid unnecessary multiplication by zero; only principal diagonal elements of **H** are used. The APL selection and expansion operations are used to multiply sparse matrixes. Arrays used in the programs are described in Table II.

ACKNOWLEDGMENT

The present work spans over several years, and the author acknowledges the encouragements given by many people: particularly Professors J. E. Dubois (University of Paris), Y. Yoneda (University of Tokyo), S. Fujiwara (University of Tokyo), and W. T. Wipke (University of California). He also wishes to thank Mr. R. Kojima for his assistance during the preparation of the manuscript.

REFERENCES AND NOTES

(1) W. T. Wipke and T. M. Dyott, *J. Am. Chem. Soc.*, **96**, 4834 (1974).
(2) J. E. Rush, *J. Chem. Inf. Comput. Sci.*, **16**, 202 (1976).
(3) M. F. Lynch, J. M. Harrison, W. G. Town, and J. E. Ash, Ed., "Computer Handling of Chemical Structure Information", Macdonald, London, 1971.
(4) W. T. Wipke, S. R. Heller, R. J. Feldmann, and E. Hyde, Ed., "Computer Representation and Manipulation of Chemical Information", Wiley, New York, N.Y., 1974.
(5) H. L. Morgan, *J. Chem. Doc.*, **5**, 107 (1965).
(6) Y. Kudo and S. Sasaki, *J. Chem. Doc.*, **14**, 200 (1974).
(7) M. Uchino, *Kagakusosetsu*, **18**, 47 (1978).
(8) Y. Kudo, *Kagakusosetsu*, **18**, 63 (1978).
(9) M. Randić, *J. Chem. Inf. Comput. Sci.*, **15**, 105 (1975).
(10) M. Randić, *J. Chem. Phys.*, **60**, 3920 (1974).
(11) M. Uchino, *J. Chem. Inf. Comput. Sci.*, following paper in this issue.
(12) D. G. Corneil and C. C. Gotlieb, *J. ACM*, **17**, 51 (1970).
(13) H. Enomoto, T. Katayama, and N. Yonezaki, *Johoshori*, **18**, 1209 (1977).
(14) L. M. Masinter, N. S. Srindharan, R. E. Carhart, and D. H. Smith, *J. Am. Chem. Soc.*, **96**, 7714 (1974).
(15) C. A. Shelly and M. E. Munk, *J. Chem. Inf. Comput. Sci.*, **17**, 110 (1977).
(16) (a) C. Jochum and J. Gasteiger, "Canonical Numbering and Constitutional Symmetry", *J. Chem. Inf. Comput. Sci.*, **17**, 113 (1977); (b) R. E. Carhart, "Erroneous Claims Concerning the Perception of Topological Symmetry", *ibid.*, **18**, 108 (1978); (c) C. Jochum and J. Gasteiger, "On the Misinterpretation of Our Algorithm for the Perception of Constitutional Symmetry", *ibid.*, **19**, 49 (1979); (d) R. E. Carhart, "Perception of Topological Symmetry", *ibid.*, **19**, 56 (1979). These papers indicate indirectly a necessity of mathematical method for precise treatment of the problem.
(17) (a) A. L. Mackay, *J. Theor. Biol.*, **54**, 399 (1975); (b) *J. Chem. Phys.*, **62**, 309 (1975).