

# Logical Perception of Ring-Opening, Ring-Closure, and Rearrangement Reactions Based on Imaginary Transition Structures. Selection of the Essential Set of Essential Rings (ESER)<sup>1</sup>

SHINSAKU FUJITA

Research Laboratories, Ashigara, Fuji Photo Film Co., Ltd., Minami-Ashigara, Kanagawa, Japan 250-01

Received February 9, 1987

Ring structures appearing in imaginary transition structures (ITS) are classified into five categories: bridges of ring opening (BO), bridges of ring closure (BC), bridges of rearrangement (BR), invariant rings (IR), and trivial rings. From all the rings detected in a given ITS, the essential set of essential rings (ESER) is selected to afford necessary and sufficient information on changes of the ring structures. The ESER contains rings other than tied, multi-tied, or dependent rings. The tied rings are defined as rings that have one transannular type A (TATA) bond attached by at least one par-subring. The multi-tied rings contain two or more TATA bonds, at least one of which has a par-subring. Dependent rings are covered rings that are covered by tied rings under covering conditions introduced. An algorithm for selection of the ESER is given. This algorithm is applicable also to the detection of the ESER of an organic compound, which is more effective than conventional methods.

In previous papers,<sup>2</sup> I discussed odd- and even-nodal subgraphs of imaginary transition structures (ITS). The subgraphs are potential clues for characterizing reaction-site changes during organic reactions. Thus, natural language terms, such as "substitution", "construction", "cleavage", "elimination", and "addition" for indication of bond switching and "oxidation" and "reduction" for information on changes of oxidation states, can be redefined by a reaction hierarchy based upon the subgraphs.<sup>2d,k</sup>

The ITS's also contain another type of information that is concerned with skeletal changes. Thus, ring opening, ring closure, and conservation of rings can be recognized by examination of ring structures appearing in ITS's (ITS rings): bridges of ring opening (BO) and of ring closure (BC) and invariant rings (IR).<sup>2a</sup> In addition, bridges of rearrangement (BR) are a kind of ITS ring that are useful to perceive "rearrangement".<sup>2j</sup> However, adoption of all rings detected from an ITS affords unnecessary duplication of information on such skeletal changes.

The present paper discusses selection of essential ITS rings that are necessary and sufficient to describe skeletal changes during organic reactions.

## BACKGROUND

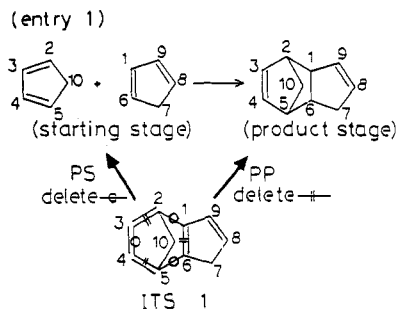
**Changes of Ring Structures As Defined in Terms of a Reaction Diagram.** Many methods have been proposed to detect rings from an organic compound and to select an appropriate number of said rings that afford necessary and sufficient information of the ring structure.<sup>3-6</sup> The smallest set of the smallest rings (SSSR)<sup>3</sup> is one of the representative methods that has been adopted by most systems of retrieval of organic compounds.

For the purpose of design of synthetic pathways, various proposals have appeared to select synthetically important rings.<sup>4-6</sup> However, their algorithms contain such ad hoc rules as to select all  $\leq 6$ -membered rings, etc. Moreover, all the methods have dealt with ring structures of organic compounds statically, even if they direct toward organic synthesis. In other words, they have never treated *changes of ring structures per se*. In this sense, they may be called methods of *compound basis*.

For the purpose of constructing an integrated system for retrieval of organic reactions and synthetic design,<sup>21</sup> methods

must be developed so as to afford necessary and sufficient information on *changes* of the ring structures. In other words, methods of *reaction basis* are desired.

Let us illustrate the situation described above by examining dimerization of cyclopentadiene (entry 1). Each molecule

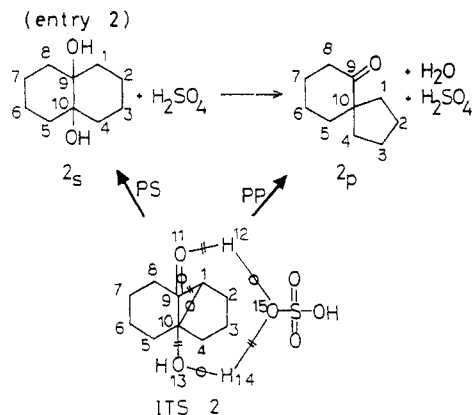


of the starting stage has a 5-membered ring. The product has three 5-membered rings, and 6-, 8-, and 9-membered rings in all, from which the SSSR algorithm selects only three 5-membered rings. The two 5-membered rings (2-3-4-5-10-2 and 1-6-7-8-9-1) are invariant rings as can be seen by comparison of both sides of the reaction diagram. The remaining 5-membered ring (1-2-10-5-6-1) is a ring formed during the dimerization. The perception of the fact that the ring (1-2-10-5-6-1) is absent in the starting stage and is present in the product stage requires strict correspondence of the participating nodes.

In this case, organic chemists would recognize formation of a 6-membered ring (1-2-3-4-5-6-1) along with that of the 5-membered ring (1-2-10-5-6-1) described above. However, we used to regard formation of the 8- and 9-membered rings as unnecessary to perceive. How do we realize this complex selection by logical procedure?

For further discussion, let us examine a pinacol rearrangement for decaline-9,10-diol to spiro[5.4]decan-1-one (entry 2)<sup>7</sup> from the viewpoint of ring perception. The starting material (**2<sub>s</sub>**) is perceived as two 6-membered rings fused. A 10-membered ring is also involved in the compound (**2<sub>s</sub>**), but it can be regarded as a combination of the two 6-membered rings. The product (**2<sub>p</sub>**) contains 6- and 5-membered rings.

Although ring opening of the 6-membered ring and formation of the 5-membered ring can be detected by comparing



the starting and product stages, taking further information such as "rearrangement" or "ring contraction" requires more elaborate examination. If this perception procedure is effected in terms of a reaction diagram, multiple reference to both sides of the diagram is necessary.

**Redefinition of Changes of Ring Structures Based on ITS.** The Diels-Alder reaction of entry 1 is represented by ITS 1, which involves three types of colored bonds according to the bond change during the reaction, i.e., out-bonds ( $-//$ ), in-bond ( $-○-$ ), and par-bonds ( $-$ ).<sup>8-10</sup> This formulation transforms detection of structural changes to examination of subgraphs of the ITS.<sup>2a</sup> As a result, *perception of changes of ring structures are transformed to detection of ITS rings (ring structures appearing in ITS)*. Thus, changes of ring structures are represented by six rings detected in ITS (Figure 1). Two 5-membered rings 1-2 and 1-3 are invariant rings (IR) as defined below and show that these are intact during the reaction. A remaining 5-membered ring (1-1) is a bridge of ring closure of order 2 ( $BC_2$ ), which corresponds to ring closure with formation of two bonds. A 6-membered BC (1-4) is also detected to show formation of a 6-membered ring. These rings are necessary and sufficient to reveal changes of ring structures during the reaction of entry 1.

An 8-membered ring (1-5) is a bridge of ring closure of order 2 ( $BC_2$ ). But this is unnecessary once the rings 1-2 ( $BC_2$ ) and 1-4 (IR) are selected. Similarly, a 9-membered ring (1-6) of  $BC_2$  can be replaced by the combination of the ring 1-3 ( $BC_2$ ) and 1-4 (IR). The above-described problem to select necessary and sufficient information from entry 1 is redefined as that of developing a novel algorithm which selects 1-1 to 1-4 from ITS 1 and omits the rings 1-5 and 1-6.

Entry 2 is represented by ITS 2,<sup>8-10</sup> from which the following four rings should be selected for necessary and sufficient information: a 3-membered ring (1-9-10-1; BR), a 5-membered ring (1-2-3-4-10-1;  $BC_1$ ), a 6-membered ring (5-6-7-8-9-10-5; IR), and another 6-membered ring (1-2-3-4-10-9-1;  $BO_1$ ). It should be noted that the term "rearrangement" corresponds to the 3-membered BR. Which algorithm do we apply to the problem to select these four rings?

In manipulation of usual structural formulas, the concept of SSSR (the smallest set of smallest rings)<sup>3</sup> is fruitful to restrict the number of rings detected to an appropriate size. However, in the present cases of imaginary transition structures, the SSSR is insufficient to describe all features of organic reactions. Thus, the SSSR for ITS 1 contains three 5-membered rings (1-1, 1-2, and 1-4) and misses the 6-membered ring 1-3 ( $BC_2$ ). Obviously, the 6-membered  $BO_1$  cannot be detected in ITS 2 by the SSSR algorithm.

**Imaginary Transition Structures as Extension of Structural Formulas.** A set of ITS's contains that of usual structures as a subset (Figure 2). A usual structural formula of an organic compound can be recognized as an ITS that contains only

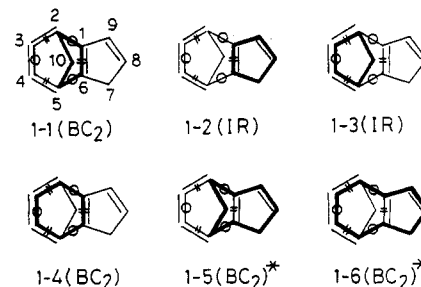


Figure 1. Rings in ITS 1. Rings with an asterisk are tied rings.

Organic reaction	⊃	Organic compound
ITS	⊃	Structure
ITS connection table	⊃	Connection table
Imaginary bond	⊃	Bond
Complex bond number	⊃	Bond multiplicity
ITS ring	⊃	Ring

Figure 2. Extended concepts in ITS's.

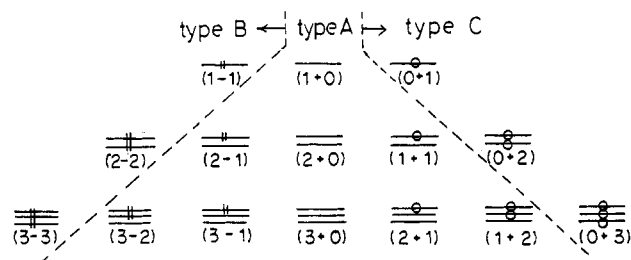


Figure 3. Imaginary bonds classified to type A, B, and C bonds.

par-bonds ( $-$ ) and has neither out-bonds ( $-//$ ) nor in-bonds ( $-○-$ ). In other words, an (unchanged) organic compound can be regarded as a kind of organic reaction in which the starting and the product stages are the same.

In accordance with this fact, other concepts used in usual structural formulas are subsumed within the corresponding concepts derived from ITS's (Figure 2). One concept of machine-readable expressions is an ITS connection table, wherein connectivities of nodes (atoms or groups) are represented by means of complex bond numbers. Each of the complex bond numbers is a pair of integers ( $a\ b$ ), where  $a$  is bond multiplicity of the starting stage and  $b$  is the difference between bond multiplicity of the starting stage and that of the product. Single, double, and triple bonds in a usual structural formula are represented by complex bond numbers (1+0), (2+0), and (3+0), respectively. Thus, the concept of "complex bond numbers" subsumes that of "bond multiplicities".

As a result, the concept of ITS rings (ring structures appearing in ITS) is an extended concept of rings defined for usual structural formulas. Ring structures appearing in usual structural formulas (for organic compounds) can be regarded as invariant rings (IR) from the viewpoint that ITS's embrace usual structural formulas. In other words, a set of ITS rings includes that of usual rings as a subset.

Th above discussions indicate that the desired algorithm for selecting essential rings must be applicable to ITS's as well as to usual structural formulas. Thus, it must solve the remaining problems in the methods of compound basis.<sup>3g</sup> At the same time, it must give an answer for the new problem concerned with ITS rings.

## DEFINITION

**Imaginary Bonds As Classified to Type A, B, and C Bonds.** Various types of imaginary bonds (ITS bonds) appearing in

Table I. Ring Types Appearing in ITS's

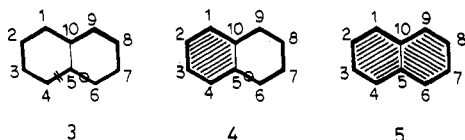
ring type	size	no. of imaginary bonds			remarks
		type A	type B	type C	
$BO_m$	$r$	$r - m$	$m$	0	$m \geq 1$
$BC_n$	$r$	$r - n$	0	$n$	$n \geq 1$
BR	$r$	$r - 2$	1	1	
IR	$r$	$r$	0	0	
trivial	$r$	$r - m - n$	$m$	$n$	$m + n > 2, mn \neq 0$

ITS's are collected in Figure 3, wherein they are represented by combinations of out-bonds, in-bonds, and/or par-bonds. Each imaginary bond is denoted by a complex bond number as shown in Figure 3. We classify these bonds into three categories. A *type A bond* is a bond that has a complex bond number ( $a b$ ) of  $a \neq 0$  and  $a + b \neq 0$ . Each *type B bond* has a complex bond number ( $a b$ ) of  $a + b = 0$ . A *type C bond* is characterized by the equation  $a = 0$  in the complex bond number ( $a b$ ).

Let us represent an imaginary bond by a combination of three integers ( $o, i, p$ ), alternatively, where integer  $o$  is the number of out-bonds in said imaginary bond,  $i$  is that of in-bonds, and  $p$  is that of par-bonds. If said bond is also represented by a complex bond number ( $a b$ ), the relationship between ( $o, i, p$ ) and ( $a b$ ) is as follows:  $a = p + o$ ,  $b = i - o$ , and  $oi = 0$ . Type A, B, and C bonds are distinguished by ( $o, i, p$ ) as follows: type A bond,  $p \neq 0$  or  $a + b \neq 0$  and  $a \neq 0$ ; type B bond,  $p + i = 0$  or  $a + b = 0$ ; and type C bond,  $p + o = 0$  or  $a = 0$ .

**Ring Types: Bridges of Ring Opening, of Ring Closure, and of Rearrangement, Invariant Rings, and Trivial Rings.** Rings appearing in a given ITS are classified by the number of type A, B, or C bonds as shown in Table I. The ITS rings,  $BO_m$ ,  $BC_m$ , BR and IR are important to describe changes of ring structures during reactions. A trivial ring may be equal to a *reaction string* in some cases with which I will deal elsewhere.

**Transannular Type A Bonds (TATA Bonds).** A transannular type A bond (TATA bond) is defined as a type A bond that links directly two nonadjacent nodes of a given ring. Let us examine substructure 3 from which a 10-membered ring is



abstracted. A bond between nodes 5 and 10 is a TATA bond as defined above. Similarly, a 10-membered ring, 4 or 5, has a TATA bond between nodes 5 and 10.

**Par-Subrings.** Let  $A$  be the set of all bonds in a given ring ( $R$ ) that has one or more TATA bonds. It should be noted here that the TATA bonds are not members of  $A$ . Let  $A_1$  be a subset of  $A$ , and then let the combined set of  $A_1$  and one of the TATA bonds construct an ITS ring. This ring is defined as a par-subring if the following conditions are fulfilled: (1) All elements (bonds) of  $A_1$  are type A bonds. (2) The ring contains no other TATA bonds than those defined above.

For example, a 10-membered ring of 3 has a TATA bond but no par-subring. In structure 4, the bond between nodes 5 and 10 is a TATA bond and the hatched 6-membered ring is a par-subring with respect to the 10-membered ring. Structure 5 has a TATA bond between nodes 5 and 10 and two 6-membered par-subrings (represented by hatching) with respect to the 10-membered ring.

**Tied Rings and Multi-Tied Rings.** A tied ring is defined as a ring that is not trivial and has a par-subring. A multi-tied ring is defined as a ring that has two or more par-subrings. For example, the 10-membered ring of 3 is not a tied ring. This ring (BR) is necessary to show reaction features. On the

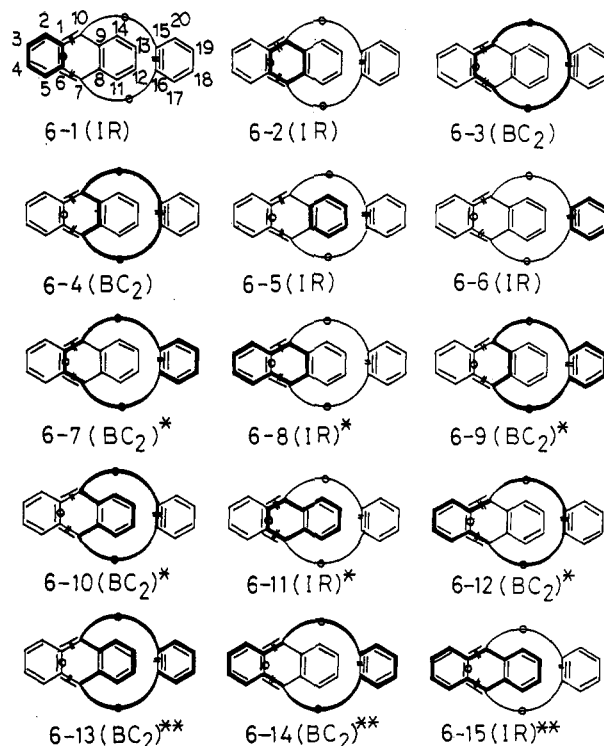
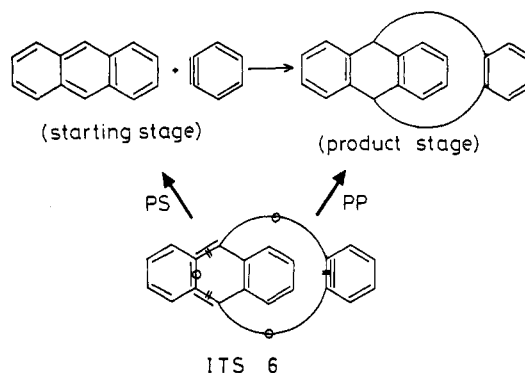


Figure 4. Rings in ITS 6.

other hand, the 10-membered ring of 4 is a tied ring that is a nonessential  $BC_1$  ring, since this can be replaced by a combination of a 6-membered  $BC_1$  and a 6-membered invariant ring. In ITS 1 of Figure 1, the 1-5 ( $BC_2$ ) and 1-6 ( $BC_2$ ) are tied rings.

The Diels-Alder reaction of anthracene with benzyne is represented by ITS 6, from which 15 rings are detected as



shown in Figure 4: i.e., six 6-membered rings (6-1 to 6-6), six 10-membered rings (6-7 to 6-12), and three 14-membered rings (6-13 to 6-15). All the 10-membered rings are tied rings (marked with an asterisk) in light of the above criterion. Three 14-membered rings are multi-tied rings (marked with double asterisks).

It is noted that tied rings and multi-tied rings of ITS's (for reactions) are an extension of the counterparts in usual structural formulas (for compounds). In a usual structural formula, a transannular bond is a clue for detection of a tied ring or a multi-tied ring. On the other hand, the corresponding TATA bond cannot be a clue for detection of a tied ring or a multi-tied ring in ITS's, where a par-subring is used for their definition instead of a TATA bond.

**Heterogeneity and Abnormality of Rings.** In order to bring chemist's intelligence into the perception of nontrivial ITS rings, non-hydrogen atoms that construct a nontrivial ITS are classified into three classes, i.e., (1) carbon atoms, (2) heteroatoms (N, O, S, and P), and (3) abnormal atoms (other

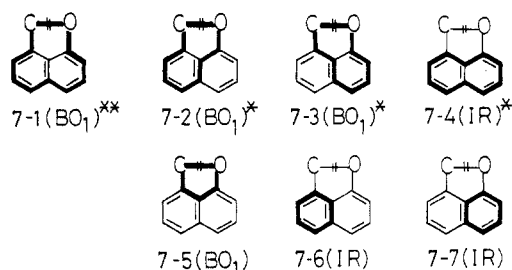


Figure 5. ITS rings in ITS 7: (\*) tied ring; (\*\*) dependent ring.

than the above). An index of heterogeneity (IH) of a ring is defined as the number of heteroatoms in said ring. An index of abnormality of a ring is the number of abnormal atoms in said ring. We classify rings into three categories, i.e., (1) carbocyclic rings that contain neither heteroatoms nor abnormal atoms (IH = 0 and IA = 0), (2) heterocyclic rings that involve at least one heteroatom and no abnormal atom (IH > 0 and IA = 0), and (3) abnormal rings that contain at least one abnormal atom (IA > 0). This classification provided chemist's intelligence to computer perception of rings to some extent. This aspect will be discussed in detail elsewhere.

**Dependent Rings.** Let  $B$  be the set of all imaginary bonds in a given nontrivial ring ( $R$ ). Let  $T$ , be the set of all tied rings.<sup>11</sup> The ring  $R$  is defined as a dependent ring when it is covered by a set ( $T$ ) of tied rings<sup>11</sup> and the following covering conditions are fulfilled. A dependent ring is a nonessential ring.

(1) Let  $T$  be a subset of  $T_r$ ;  $T$  covers all imaginary bonds in  $B$ .

(2) All  $S (\in T)$  are the same as or smaller than said ring ( $R$ ) in their ring sizes.

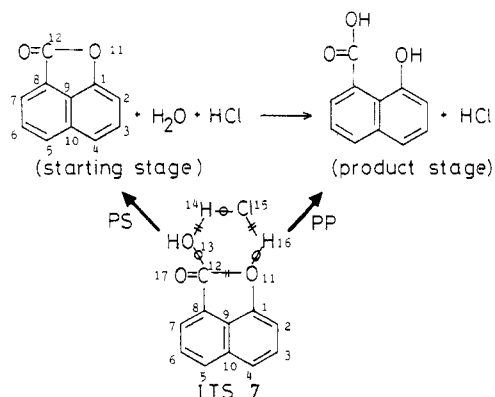
(3) The intersection of all  $S (\in T)$  and  $R$  involves not less than half the bonds of  $S$ .

(4) All  $S (\in T)$  are of the same class (i.e., carbocyclic, heterocyclic, or abnormal) as said ring ( $R$ ).

(5) All  $S (\in T)$  have the same or smaller value of IH (in the case of heterocyclic rings) or IA (in the case of abnormal rings) as compared with that of said ring ( $R$ ).

(6) All  $S (\in T)$  belong to the same ring type (BO, BC, or BR) as  $R$  or are of IR.

The concept of dependent rings is illustrated by an example of a lactone hydrolysis shown in Figure 5. Two 9-membered rings (7-2 and 7-3) are of  $BO_1$  and are classified as tied rings



(marked with an asterisk) in light of the criterion described above. An 11-membered ring (7-1) is a bridge of ring opening of order 1 ( $BO_1$ ). This ring is unnecessary to adopt chemically once a more essential ring of  $BO_1$  (7-5) is selected as an essential ring. The ring 7-1 is neither a tied nor a multi-tied ring and cannot be rejected unless the concept of dependent rings is introduced. The ring 7-1 is a dependent ring (marked with a double asterisk) because it can be covered by two 9-membered tied rings of  $BO_1$  (7-2 and 7-3) in light of the

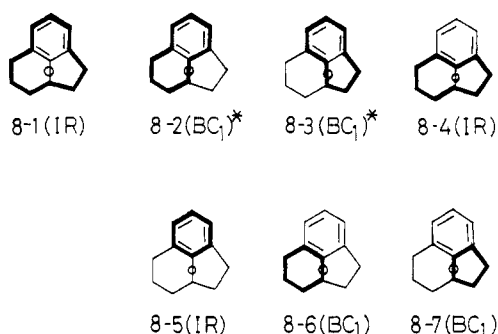
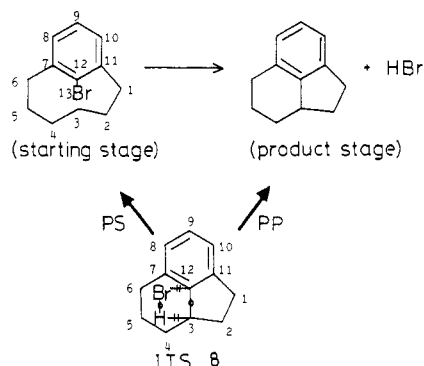


Figure 6. ITS rings in ITS 8: (\*) tied ring.

covering conditions described above. The covering condition (6) forbids the ring 7-4 to be selected for covering the ring 7-1.

Let us work out another example, i.e., dehydrobromination represented by ITS 8.<sup>12</sup> Nontrivial rings in ITS 8 are listed



in Figure 6. Nine-membered rings ( $BC_1$ ) 8-2 and 8-3 are tied rings, as can be seen easily. On the other hand, 9-membered ring 8-4 is not a tied ring, since it has neither a TATA bond nor a par-subring. An 11-membered ring (8-1) of Figure 6 is not covered by tied rings, 8-2 and 8-3, since the ring classes are different from each other. Thus, we can select five essential rings, 8-1, 8-4, 8-5, 8-6, and 8-7, from the substructural ITS of Figure 6.

#### ESSENTIAL SET OF ESSENTIAL RINGS IN AN IMAGINARY TRANSITION STRUCTURE

The above-defined tied, multi-tied, dependent rings as well as trivial rings are all regarded as nonessential. The essential set of essential rings (ESER) in an imaginary transition structure (ITS) is defined as a set of rings other than non-essential rings. When this definition of ESER is applied to the case of ITS 1 in Figure 1, the rings 1-5 ( $BC_2$ ) and 1-6 ( $BC_2$ ) are tied rings. The ESER of ITS 1 contains 1-1 (IR), 1-2 ( $BC_2$ ), 1-3 ( $BC_2$ ), and 1-4 (IR).

The ESER of the case of Figure 4 contains six 6-membered rings (6-1 to 6-6). Other rings of Figure 4 are either tied or multi-tied rings.

The example of Figure 5 affords the rings 7-5 ( $BO_1$ ), 7-6 (IR), and 7-7 (IR) as ESER. These three rings are sufficient to show features of this reaction concerned with the ITS rings.

From the rings shown in Figure 6, the rings 8-1 (IR), 8-4 (IR), 8-5 (IR), 8-6 ( $BC_1$ ), and 8-7 ( $BC_1$ ) are selected as ESER.

#### AN ALGORITHM FOR ESER IN ITS

The above-described strategy for detection of ESER is summarized as the following algorithm.

**Step 1.** Detect all rings.<sup>13</sup>

**Step 2.** Set the rings in ascending order of ring size.

**Step 3.** Count type A, type B, and type C bonds, respectively, involved in each ring. Classify the ring according to the criterion shown in Table I. Trivial rings are not considered.

**Step 4.** Calculate IH and IA for each ring.

**Step 5.** Set MULTI = 0 and MLT = 0 for each ring. Detect TATA bond(s) for each ring. Set MULTI = MULTI + 1 for each detection. Examine whether each TATA bond has a par-subring or not. If a par-subring is detected, said ring is regarded as a tied ring. Set MLT = 1 for said ring. Detect multi-tied ring(s) similarly, set MLT = 1 for a multi-tied ring, and then exclude tied and multi-tied rings as nonessential rings.

**Step 6.** If a remaining ring is carbocyclic, compare this with carbocyclic tied rings (IH = IA = 0). If said ring is covered by the terms of the above-described covering conditions, set MLT = 1 for said ring and exclude this. Then go to step 9.

**Step 7.** If a remaining ring is heterocyclic, compare this with heterocyclic tied ring (IH > 0 and IA = 0). If said ring is covered by these as above, set MLT = 1 and exclude said ring. Then go to step 9.

**Step 8.** If a remaining ring is an abnormal ring, compare this with abnormal tied rings (IA > 0). If said ring is a covered ring as defined above, set MLT = 1 and exclude said ring. Then go to step 9.

**Step 9.** If all remaining rings are not examined, go to step 5. When the examination is complete, all rings of MLT = 0 (i.e., other than tied, multi-tied, and dependent rings) are selected as ESER.

#### ESER IN THE STARTING STAGE OR THE PRODUCT STAGE

A usual structural formula for an organic compound can be regarded as an ITS that contains par-bonds only as discussed above. In other words, a set of ITS's embraces usual structural formulas as a subset. As a result, the above-described ESER algorithm for ITS is applicable exactly to usual structural formulas. In a usual structural formula, each TATA bond has a par-subring necessarily. Hence, some simplification of the above-described algorithm is permissible.

An algorithm for detection of ESER in a starting stage is as follows. After the PS operation (cf. Appendix I), the following steps are effected.

**Step 1' to Step 4'.** The same operations as shown in step 1 to step 4.

**Step 5'.** Set ER-S = 0 for each ring. Unless the ring type of an examined ring is BO or IR, set ER-S = -1 and go to the next ring. If so, detect TATA bond(s) from said ring. When one TATA bond is detected, regard said ring as a tied ring and set ER-S = 1. When two or more TATA bonds are detected, regard said ring as a multi-tied ring and add 1 to ER-S for each detection.

**Step 6' to Step 9'.** The same operations as described in step 6 to step 9, respectively, except replacement of MLT = 1 by ER-S = 1.

After the operation of this algorithm, rings in which ER-S = 0 belong to ESER in the starting stage.

An algorithm for detection of ESER in the product stage is also given. Thus, rings of ER-P = 0 are members of ESER in the product stage.

#### IMPLEMENTATION AND RESULTS

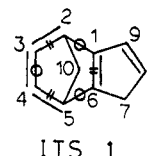
The above algorithm was programed in FORTRAN 77 and implemented on a VAX 11/750 (Digital Equipment Co.). The output of the operation to ITS 1 is shown in Figure 7 as an example. The value of MULTI corresponds to the number of TATA bonds detected from a given ring. Rings of MLT = 0 are of ESER. Other rings are nonessential rings. When ER-S = 0, the ring is an essential ring in the starting stage.

REACTION NAME:  
[TEST12] DIELS-ALDER ADDN; CYCLOPENTADIENE TO DIMER (FULL)  
NUMBER OF RINGS: 6

\*\*\*\*\*  
SSSITS RINGS  
\*\*\*\*\*

\*\*\*LIST OF RINGS\*\*\*

RING # 1: SIZE = [ 5 ]  
MEMBERS: 1 2 10 5 6  
RING # 2: SIZE = [ 5 ]  
MEMBERS: 1 6 7 8 9  
RING # 3: SIZE = [ 5 ]  
MEMBERS: 2 3 4 5 10  
RING # 4: SIZE = [ 6 ]  
MEMBERS: 1 2 3 4 5 6  
RING # 5: SIZE = [ 8 ]  
MEMBERS: 1 2 10 5 6 7 8 9  
RING # 6: SIZE = [ 9 ]  
MEMBERS: 1 2 3 4 5 6 7 8 9



ITS 1

\*\*\*RING TYPE\*\*\*

RING #	SIZE	A	B	C	RING TYPE	MULTI	ESER	MLT	ER-S	ER-P
1	[ 5 ]	3	0	2	BC 2	0	0	0	-1	0
2	[ 5 ]	5	0	0	INVAR	0	0	0	0	0
3	[ 5 ]	5	0	0	INVAR	0	0	0	0	0
4	[ 6 ]	4	0	2	BC 2	0	0	0	-1	0
5	[ 8 ]	6	0	2	BC 2	1	1	1	-1	1
6	[ 9 ]	7	0	2	BC 2	1	1	1	-1	1

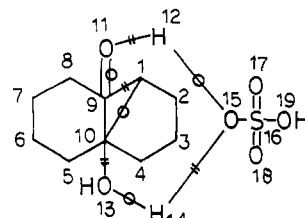
Figure 7. ESER detection of ITS 1. A ring of MLT = 0 is a member of the ESER.

REACTION NAME:  
[TEST32] A PINACOL REARRANGEMENT OF DECALINE-DIOL  
NUMBER OF RINGS: 10

\*\*\*\*\*  
SSSITS RINGS  
\*\*\*\*\*

\*\*\*LIST OF RINGS\*\*\*

RING # 1: SIZE = [ 3 ]  
MEMBERS: 1 9 10  
RING # 2: SIZE = [ 5 ]  
MEMBERS: 1 2 3 4 10  
RING # 3: SIZE = [ 6 ]  
MEMBERS: 1 2 3 4 10 9  
RING # 4: SIZE = [ 6 ]  
MEMBERS: 5 6 7 8 9 10  
RING # 5: SIZE = [ 7 ]  
MEMBERS: 1 9 8 7 6 5 10  
RING # 6: SIZE = [ 7 ]  
MEMBERS: 9 10 13 14 15 12 11  
RING # 7: SIZE = [ 8 ]  
MEMBERS: 1 9 11 12 15 14 13 10  
RING # 8: SIZE = [ 10 ]  
MEMBERS: 1 2 3 4 10 5 6 7 8 9  
RING # 9: SIZE = [ 11 ]  
MEMBERS: 5 6 7 8 9 11 12 15 14 13 10  
RING # 10: SIZE = [ 11 ]  
MEMBERS: 1 2 3 4 10 13 14 15 12 11 9



ITS 2

\*\*\*RING TYPE\*\*\*

RING #	SIZE	A	B	C	RING TYPE	MULTI	ESER	MLT	ER-S	ER-P
1	[ 3 ]	1	1	1	BR 1 2	0	0	0	-1	-1
2	[ 5 ]	4	0	1	BC 1	0	0	0	-1	0
3	[ 6 ]	5	1	0	BO 1	0	0	0	0	-1
4	[ 6 ]	6	0	0	INVAR	0	0	0	0	0
5	[ 7 ]	5	1	1	BR 1 6	1	1	1	-1	-1
6	[ 7 ]	2	3	2						
7	[ 8 ]	1	4	3						
8	[ 10 ]	9	1	0	BO 1	1	1	1	1	-1
9	[ 11 ]	6	3	2						
10	[ 11 ]	5	4	2						

Figure 8. ESER detection of ITS 2. A ring of MLT = 0 is a member of the ESER.

When ER-S = 1, the ring is a nonessential ring in the starting stage. When ER-S = -1, the ring is absent in the starting stage. The counterparts of the product stage are distinguished by the values of ER-P.

As a matter of course, rings of the ESER detected (MLT = 0) were identical with the list of Figure 1. Rings 2 (1-2) and 3 (1-3) had zero ER-S value, which showed the corresponding rings were essential rings in the starting stage. Other rings were absent in the starting stage (ER-S = -1). Similarly, rings 1-4 (1-1 to 1-4 of Figure 1) were essential rings in the product stage since ER-P = 0. Other rings (5 and 6) were present in the product stage but nonessential in terms of the present algorithm.

Figure 8 shows the result of the ESER effected toward ITS 2. The ESER (MLT = 0) involved four rings, i.e., nos. 1 (BR), 2 (BC<sub>1</sub>), 3 (BO<sub>1</sub>), and 4 (IR), which were identical with those discussed above. They corresponded to the natural language terms rearrangement, ring closure to a 5-membered ring, and ring opening of a 6-membered ring. The term ring contraction can be derived by a combination of the BR, BC<sub>1</sub>, and BO<sub>1</sub>, which will be discussed elsewhere. The 7-membered BR (no. 5) can be considered a combination of rings 1 (BR) and 4 (IR) and therefore is not a member of the ESER. A 10-membered

BO<sub>1</sub> (no. 8) was nonessential since it was replaced by a combination of rings 3 (BO<sub>1</sub>) and 4 (IR). The other rings were all trivial rings. However, an 8-membered ring (no. 7) is a reaction string that should be detected by another algorithm.

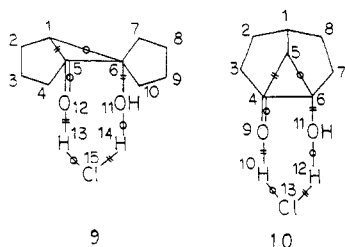
Two 6-membered rings corresponding to rings 3 and 4 were essential rings in the starting stage of ITS 2 (ER-S = 0). Rings 2 (5-membered) and 4 (6-membered) were essential rings in the product stage (ER-P = 0).

When the ESER algorithm was applied to ITS 6, six 6-membered rings (6-1 to 6-6) were detected in accordance with the list of Figure 4. The two BC<sub>2</sub> (6-3 and 6-4) indicate the features of this reaction. The substructure 6-3 can be perceived as a reaction string by another algorithm. Other rings (6-7 to 6-15) were nonessential rings in ITS 6. In the starting stage, the rings corresponding to 6-1, 6-2, 6-5, and 6-6 were essential rings (ER-S = 0). In the product stage, rings corresponding to 6-1 to 6-6 were essential rings (ER-P = 0).

Twelve rings were detected from ITS 7, from which nontrivial rings were selected and shown in Figure 5. Three rings, 7-5, 7-6, and 7-7, were of the ESER in light of this algorithm. In the starting stage, the counterparts of these three rings were essential rings. In the product stage, two 6-membered rings (the counterparts of 7-6 and 7-7) were selected as the ESER.

Twelve rings in all were detected in ITS 8 by the ESER algorithm. Among them, seven rings were nontrivial as shown in Figure 6. Then five rings (8-1, 8-4 to 8-7 of Figure 6) were selected as the ESER of ITS 8. Contrary to ITS 7 (Figure 5), additional rings, 8-1 and 8-4, were adopted as essential invariant rings in this case. In the starting stage, a 6-membered ring (corresponding to 8-5), a 10-membered ring (corresponding to 8-4), and a 12-membered ring (corresponding to 8-1) were essential rings. On the other hand, two 6-membered rings (corresponding to 8-5 and to 8-6) and a 5-membered ring (corresponding to 8-7) were selected as the ESER in the product stage. It should be noted that the 11-membered ring of the product is not selected as a member of the ESER since it is a dependent ring in the product stage.

A pinacol rearrangement<sup>14</sup> corresponds to ITS 9. This ITS afforded seven rings, from which a 3-membered ring (1-5-6-1, BR), a 5-membered ring (1-2-3-4-5-1, BO<sub>1</sub>), another 5-membered ring (6-7-8-9-10-6, IR), and a 6-membered ring (1-2-3-4-5-6-1, BC<sub>1</sub>) were extracted as the ESER. The other three rings are all trivial rings.



In the starting stage of ITS 9, the counterparts of two 6-membered rings were perceived as the ESER (ER-S = 0). In the product stage, the 5-membered ring (6-7-8-9-10-6) and the 6-membered ring (1-2-3-4-5-6-1) were the ESER.

Another pinacol rearrangement represented by ITS 10 afforded 12 rings, from which 7 rings were selected as nontrivial rings. The output in the case of ITS 10 is shown in Figure 9. Rings that have no assignment in the column "ring type" were trivial rings. The symbol BR 1 2 (no. 1) corresponded to a 1,2-migration reaction. It is noted that ring 9 (8-membered ring, BR) is a tied ring (the bond 4-6 is a TATA bond having a par-subring 4-6-7-8-1-2-3-4) and, therefore, is a nonessential ring.

A carbene addition afforded ITS 11. A 3-membered ring (1-6-7-1, B<sub>2</sub>) and a 6-membered ring (1-2-3-4-5-6-1, IR) were of the ESER. A 7-membered ring (1-2-3-4-5-6-7-1,

REACTION NAME:  
[TEST2] PINACOL REARR: BICYCLO[3.2.1] TO [3.2.1] (FULL)  
NUMBER OF RINGS: 12

\*\*\*\*\*  
\*\*\*\*\*  
\*\*\*\*\*

\*\*\*LIST OF RINGS\*\*\*

RING # 1: SIZE = [ 3 ]

MEMBERS: 4 5 6

RING # 2: SIZE = [ 5 ]

MEMBERS: 1 5 6 7 8

RING # 3: SIZE = [ 5 ]

MEMBERS: 1 2 3 4 5

RING # 4: SIZE = [ 6 ]

MEMBERS: 1 2 3 4 6 5

RING # 5: SIZE = [ 6 ]

MEMBERS: 1 5 4 6 7 8

RING # 6: SIZE = [ 7 ]

MEMBERS: 1 2 3 4 6 7 8

RING # 7: SIZE = [ 7 ]

MEMBERS: 4 6 11 12 13 10 9

RING # 8: SIZE = [ 8 ]

MEMBERS: 4 5 6 11 12 13 10 9

RING # 9: SIZE = [ 8 ]

MEMBERS: 1 2 3 4 5 6 7 8

RING # 10: SIZE = [ 11 ]

MEMBERS: 1 2 3 4 9 10 13 12 11 6 5

RING # 11: SIZE = [ 11 ]

MEMBERS: 1 5 4 9 10 13 12 11 6 7 8

RING # 12: SIZE = [ 12 ]

MEMBERS: 1 2 3 4 9 10 13 12 11 6 7 8

\*\*\*RING TYPE\*\*\*

RING #	SIZE	A	B	C	BOND TYPE	RING TYPE	MULTI	ESER	MLT	ER-S	ER-P
1	[ 3 ]	1	1	1	BR	1 2	0	0	0	-1	-1
2	[ 5 ]	4	0	0	1 BC	1	0	0	0	-1	0
3	[ 5 ]	4	1	0	BO	1	0	0	0	0	-1
4	[ 6 ]	5	0	1	BC	1	0	0	0	-1	0
5	[ 6 ]	5	1	0	BO	1	0	0	0	0	-1
6	[ 7 ]	7	0	0	INVAR		0	0	0	0	0
7	[ 7 ]	2	3	2							
8	[ 8 ]	1	4	3							
9	[ 8 ]	6	1	1	BR	1 7	2	1	-1	-1	-1
10	[ 11 ]	5	3	3							
11	[ 11 ]	5	4	2							
12	[ 12 ]	7	3	2							

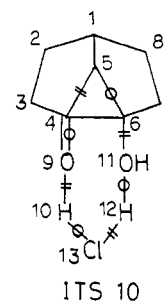
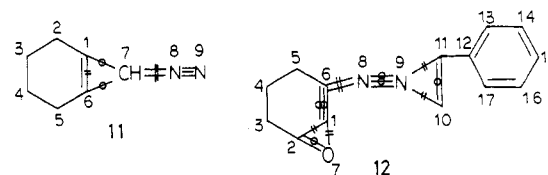


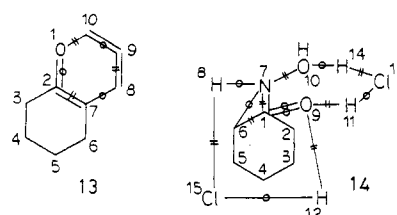
Figure 9. ESER detection of ITS 10. A ring of MLT = 0 is a member of the ESER.

BC<sub>2</sub>) was nonessential, since it can be regarded as a combination of the above 3- and 6-membered rings.



Eschenmoser's epoxyhydrazone reaction<sup>15</sup> has ITS 12. From this ITS, five rings were detected as the ESER: a 3-membered ring (1-2-7-1, BO<sub>2</sub>), another 3-membered ring (9-10-11-9, BO<sub>2</sub>), a 6-membered ring (1-2-3-4-5-6-1, BO<sub>1</sub>), another 6-membered ring (12-13-14-15-16-17-12, IR), and a 7-membered ring (1-6-5-4-3-2-7-1, BO<sub>1</sub>). It is noted that the 7-membered ring was a fused ring (the 3- plus the 6-membered rings) in some sense but was adopted as one of the ESER in the present paper. Suppose that a bond between nodes 1 and 2 was cleaved at first. The adoption of the 7-membered BC<sub>1</sub> was permitted under this situation.

The Claisen rearrangement<sup>16</sup> represented by ITS 13 had three rings, from which two 6-membered rings (BR and IR) were selected as the ESER. The 10-membered ring (BR) was nonessential and regarded as a combination of the former two rings.



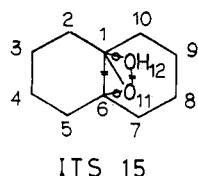
The Beckmann rearrangement<sup>17</sup> (ITS 14) contained 10 rings, three of which were nontrivial and also of the ESER in light of this algorithm: a 3-membered ring (1-6-7-1, BR), a 6-membered ring (1-2-3-4-5-6-1, BO<sub>1</sub>), and a 7-membered ring (1-2-3-4-5-6-7-1, BC<sub>1</sub>).

REACTION NAME:  
[TEST7] CREEGE REARR; DECALINE TO BRIDGED COMPO (FULL)  
NUMBER OF RINGS: 10

#####  
#####  
#####

\*\*\*LIST OF RINGS\*\*\*

RING # 1: SIZE = [ 3 ]  
MEMBERS: 1 6 11  
RING # 2: SIZE = [ 3 ]  
MEMBERS: 1 11 12  
RING # 3: SIZE = [ 4 ]  
MEMBERS: 1 6 11 12  
RING # 4: SIZE = [ 6 ]  
MEMBERS: 1 6 7 8 9 10  
RING # 5: SIZE = [ 6 ]  
MEMBERS: 1 2 3 4 5 6  
RING # 6: SIZE = [ 7 ]  
MEMBERS: 1 10 9 8 7 6 11  
RING # 7: SIZE = [ 7 ]  
MEMBERS: 1 2 3 4 5 6 11  
RING # 8: SIZE = [ 8 ]  
MEMBERS: 1 10 9 8 7 6 11 12  
RING # 9: SIZE = [ 8 ]  
MEMBERS: 1 2 3 4 5 6 11 12  
RING # 10: SIZE = [ 10 ]  
MEMBERS: 1 2 3 4 5 6 7 8 9 10

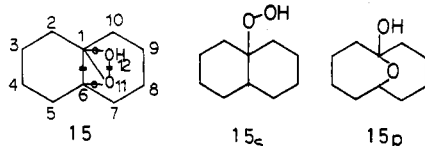


\*\*\*RING TYPE\*\*\*

RING #	SIZE	A	B	C	RING TYPE	MULTI	ESER MLT	ER-S	ER-P
1	[ 3 ]	1	1	1	BR 1 2	0	0	-1	-1
2	[ 3 ]	1	1	1	BR 1 2	0	0	-1	-1
3	[ 4 ]	0	2	2					
4	[ 6 ]	5	1	0	BO 1	0	0	0	-1
5	[ 6 ]	5	1	0	BO 1	0	0	0	-1
6	[ 7 ]	6	0	1	BC 1	0	0	-1	0
7	[ 7 ]	6	0	1	BC 1	0	0	-1	0
8	[ 8 ]	5	1	2					
9	[ 8 ]	5	1	2					
10	[ 10 ]	10	0	0	INVAR	0	0	1	0

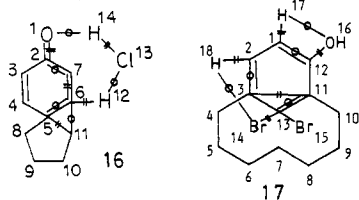
Figure 10. ESER detection of ITS 15. A ring of MLT = 0 is a member of the ESER.

Figure 10 shows the output of examination of ITS 15, which represents the Criegee rearrangement.<sup>18</sup> It is noted that a



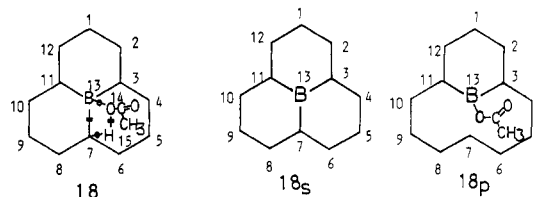
10-membered ring (no. 10) was an essential ring in the ITS and in the product stage (15<sub>p</sub>) but was a nonessential ring in the starting stage (15<sub>s</sub>). The 10-membered ring of 15<sub>p</sub> cannot be selected by the SSSR algorithm<sup>3</sup> or by other algorithms.<sup>4,6</sup> This perception is important to understand that this reaction is a transannular reaction.

A dienone-to-phenol rearrangement<sup>19</sup> represented by ITS 16 afforded 10 rings on the present examination. Among them, six rings were nontrivial and four rings were of the ESER: a 3-membered ring (5-6-11-5, BR), a 5-membered ring (5-8-9-10-11-5, BO<sub>1</sub>), and two 6-membered rings (5-6-11-10-9-8-5, BC<sub>1</sub>; 2-3-4-5-6-7-2, IR).



A bond-switching reaction<sup>12</sup> represented by ITS 17 had 19 rings, 6 of which were nontrivial and were selected as the ESER: a 3-membered ring (3-11-13-3, BO<sub>1</sub>), a 5-membered ring (1-2-3-11-12-1, BO<sub>1</sub>), a 6-membered ring (1-2-3-13-11-12-1, IR), a 9-membered ring (3-4-5-6-7-8-9-10-11-3, BO<sub>1</sub>), a 10-membered ring (3-4-5-6-7-8-9-10-11-13-3, IR), and a 12-membered ring (1-2-3-4-5-6-7-8-9-10-11-12-1, IR).

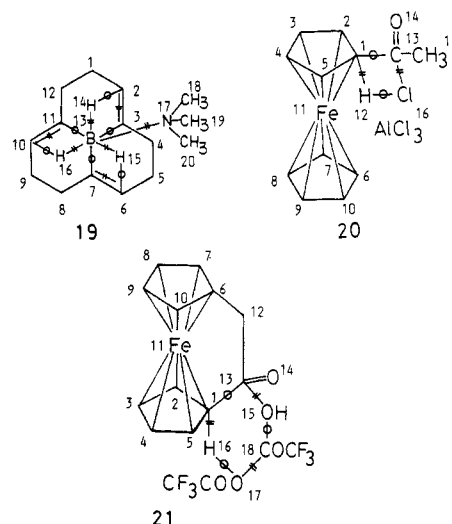
A ring-opening reaction of a boron-containing compound<sup>20</sup> (ITS 18) afforded 12 rings on the present examination. Among them, seven rings were nontrivial rings, five of which, in turn, were of the ESER. The ESER contained two 6-membered BO<sub>1</sub>'s (3-4-5-6-7-13-3 and 7-8-9-10-11-13-7), a 6-membered IR (1-2-3-13-11-12-1), a 10-membered IR (3-4-5-6-7-8-9-10-11-13-3), and a 12-membered IR



(1-2-3-4-5-6-7-8-9-10-11-12-1). The situation is the same as in the case of ITS 8.

However, the ESER of the starting and the product levels was different between ITS's 8 and 18. A 12-membered ring of the starting stage (18<sub>s</sub>) was selected as a member of the starting ESER. This should be compared with the fact that the 11-membered ring of the starting material (in ITS 8) was not a member of the starting ESER. This difference stems from the abnormality as defined above.

In the case of ITS 19,<sup>20</sup> 31 rings were detected. Among them, seven rings were nontrivial and were members of the ESER. Three 10-membered rings (BC<sub>2</sub>) belonged to the ESER. Other rings of the ESER were three 6-membered rings (BC<sub>2</sub>) and a 12-membered ring (IR).



The Friedel-Crafts reaction of ferrocene is represented by ITS 20. From 43 rings detected, 40 rings were nontrivial. The ESER contains 12 rings. Then 10 3-membered rings (IR) and 2 5-membered rings (IR) were selected as the ESER. The latter 5-membered rings cannot be selected by the conventional SSSR algorithm.<sup>21</sup> The result stems from the introduction of abnormality described above.

An intramolecular Friedel-Crafts acylation (ITS 21) afforded 205 rings, from which 10 3-membered rings (IR), 2 carbocyclic 5-membered rings (IR), and an iron-containing 5-membered ring (BC<sub>1</sub>) were selected as the ESER.

## CONCLUSION

A novel concept of the essential set of essential rings (ESER) was proposed to afford necessary and sufficient information on changes of ring structures during organic reactions. Tied rings and multi-tied rings were defined in terms of the presence of par-subring(s). Dependent rings were covered by tied rings under the covering conditions in which ring types, heterogeneity, and abnormality of rings were considered. Then the essential set of essential rings was defined as a set of rings other than tied, multi-tied, or dependent rings. An algorithm for selection of the ESER was afforded. This algorithm was applicable to usual structural formulas.

## APPENDIX I. GLOSSARY

**Imaginary Transition Structure (ITS).** A three-colored structural formula in which all nodes appearing both in the



starting stage of a given reaction and in the product stage are connected with out-, in-, and par-bonds according to the change of bonds during the reaction.

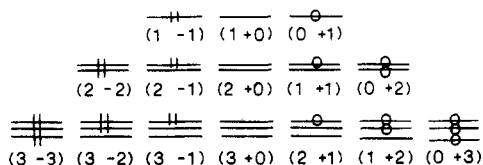
**Out-Bond.** A bond appearing only in the starting stage of a reaction.

**In-Bond.** A bond appearing only in the product stage of a reaction.

**Par-Bond.** A bond that is invariant during a reaction.

**Bond Color.** A category of bonds appearing in ITS's. There are three bond colors, i.e., out, in, and par.

**ITS Bond or Imaginary Bond.** A bond appearing in ITS's that consists of out-, in-, and/or par-bonds. The types of ITS bonds are collected below.



**Complex Bond Number.** A pair of integers ( $a\ b$ ), wherein the integer  $a$  is the bond multiplicity of the starting molecule of a given reaction and  $b$  is the difference in the bond multiplicity between the product and the starting stage. The sum  $a + b$  is the bond multiplicity of the corresponding bond of the product.

**Color-Bond Number.** A combination of three integers ( $o, i, p$ ), wherein  $o$  is the number of out-bonds contained in a given imaginary bond,  $i$  is the number of in-bonds, and  $p$  is the number of par-bonds. The following relationship between a complex bond number ( $a\ b$ ) and the corresponding color-bond number ( $o, i, p$ ) is derived:  $a = p + o$ ,  $b = i - o$ , and  $oi = 0$ .

**Connection Table of ITS's (ITS Connection Table).** A kind of connection table in which the connectivity is represented by complex bond numbers in accordance with ITS's.

**Projection to the Starting Stage (PS).** Deletion of in-bonds from an ITS. This operation produces the corresponding starting stage. This operation is replacement of each complex bond number ( $a\ b$ ) by usual multiplicity  $a$ .

**Projection to the Product Stage (PP).** Deletion of out-bonds from an ITS. This operation produces the corresponding product stage. This operation is replacement of each complex bond number ( $a\ b$ ) by usual multiplicity  $a + b$ .

**Reaction Center.** A node that is incident to in- and/or out-bonds.

**Graph of Reaction Centers (RC Graph).** A graph that consists of, at least, reaction centers. An RC graph of level 1 is defined as an RC graph containing only reaction centers. An RC graph of level 2 is one with reaction centers and the next-neighbor nodes. RC graphs are the subgraphs of ITS's and the descriptors of reaction types.

**Reaction Graph (RC Graph of Level 0).** A graph in which all nodes are regarded as balls in an abstract fashion and combined by ITS bonds ( $a\ b$ ) of  $b \neq 0$ . A reaction graph is also a descriptor of reaction type in the most abstract level.

**Transformation to Reverse Reaction.** An operation in which all in-bonds and out-bonds of a reaction graph (or an RC graph of a given level) are exchanged with each other. This operation gives the reaction graph of the corresponding reverse reaction.

**Reaction Pair.** A pair of reaction graphs (or of RC graphs of a given level) that is obtained by transformation to reverse reaction. The two reaction graphs of a reaction pair have a common skeleton of par-bonds.

**Self-Reaction Pair.** An invariant reaction graph with respect to transformation of reverse reaction.

**Par-Bond Skeleton.** A set of par-bonds that is involved in a reaction graph. Respective par-bond skeletons of a reaction pair are identical with each other. In other words, a par-bond

skeleton is invariant on transformation to reverse reaction.

**Single Par-Bond.** A par-bond that attaches to or modifies an out- or in-bond. These bonds constitute an imaginary bond.

**Double Par-Bond.** A combination of two par-bonds that attach to or modify an out- or in-bond. These bonds constitute an imaginary bond.

**Reaction Graph Counting Polynomial.** A polynomial of  $x$  and  $y$  [ $G(x, y)$ ] in which the coefficient of a term  $x^m y^n$  is the number of reaction graphs (isomers) with  $m$  double par-bonds and  $n$  single par-bonds.

**Reaction Pair Counting Polynomial.** A polynomial of  $x$  and  $y$  [ $P(x, y)$ ] in which the coefficient of a term  $x^m y^n$  is the number of reaction pairs (isomers) with  $m$  double par-bonds and  $n$  single par-bonds.

**Reaction String.** A structure that has alternate in-bonds and out-bonds and can be modified by par-bonds. A reaction graph (or an RC graph of a given level or an ITS) contains one or more reaction strings.

**Stringity.** Number of reaction strings appearing in a reaction graph, etc.

**One-String Reaction.** A reaction that contains a single reaction string in the corresponding ITS (or RC graphs of various levels). Two-, three-, or multi-string reactions are defined similarly.

**ITS Ring.** A ring structure appearing in an ITS.

**Bridge of Ring Opening of Order  $p$  ( $BO_p$ ).** An ITS ring in which  $p$  ITS bonds have ( $a\ b$ ) of  $a + b = 0$  and all other bonds have ( $a\ b$ ) of  $a + b \neq 0$  and  $a \neq 0$ . Appearance of this bridge corresponds to a ring-opening reaction.

**Bridge of Ring Closure of Order  $p$  ( $BC_p$ ).** An ITS ring in which  $p$  ITS bonds have ( $a\ b$ ) of  $a = 0$  and all other bonds have ( $a\ b$ ) of  $a + b \neq 0$  and  $a \neq 0$ . This is a descriptor of a ring-closure reaction.

**Bridge of Rearrangement (BR).** An ITS ring in which one ITS bond has ( $a\ b$ ) of  $a + b = 0$ , another has ( $a\ b$ ) of  $a = 0$ , and all other ITS bonds have ( $a\ b$ ) of  $a + b \neq 0$  and  $a \neq 0$ . This corresponds to a rearrangement.

**Intrastring Atom.** An atom that is a member of a reaction string.

**Extrastring Atom.** An atom that is not a member of a reaction string.

**Three-Nodal Subgraph.** A subgraph (of an ITS) in which a central carbon atom and adjacent two non-carbon terminal atoms are linked with an out-bond and an in-bond. The two bonds may be modified by single par-bond(s) or double par-bond(s).

**Four-Nodal Subgraph.** A subgraph (of an ITS) in which two intrastring adjacent carbon atoms are linked to two intrastring terminal non-carbon atoms.

**Reaction Kernel.** A subgraph (of an ITS) that consists of adjacent carbon atoms linked by out- and/or in-bonds. The three-nodal subgraph has a  $C_1$  reaction kernel. The four-nodal subgraph has a  $C_2$  reaction kernel.

**Odd-Nodal Subgraph.** A subgraph (of an ITS) in which an odd-membered reaction kernel is linked to two non-carbon atoms at its terminal positions by an out-bond and an in-bond.

**Even-Nodal Subgraph.** A subgraph (of an ITS) in which an even-membered reaction kernel is linked to two non-carbon atoms at its terminal positions by two out-bonds or by two in-bonds.

**Terminal Descriptor.** A subgraph that is a part of an odd- or even-nodal subgraph other than the corresponding reaction kernel. In other words, an odd- or even-nodal subgraph consists of a reaction kernel and a terminal descriptor.

## REFERENCES AND NOTES

- (1) Presented in part at the 192nd National Meeting of the American Chemical Society, Anaheim, CA, September 1986. See also: *Chem. Eng. News* 1986, 64(39), 75-76.



- (2) (a) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, 26, 205. (b) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, 26, 212. (c) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, 26, 224. (d) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, 26, 231. (e) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, 26, 238. (f) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 99. (g) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 104. (h) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 111. (i) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 115. (j) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1987**, 27, 120. (k) Fujita, S. *J. Chem. Soc., Perkin Trans. 2* (in press). (l) Fujita, S. *Yuki Gosei Kagaku Kyokaiishi* **1986**, 44, 354.
- (3) (a) Frèrejacque, M. *Bull. Soc. Chim. Fr.* **1939**, 6, 1008. (b) Plotkin, M. *J. Chem. Doc.* **1971**, 11, 60. (c) Zamora, A. *J. Chem. Inf. Comput. Sci.* **1976**, 16, 40. (d) Bersohn, M. *J. Chem. Soc., Perkin Trans. 1* **1973**, 1239. (e) Esak, A. *J. Chem. Soc., Perkin Trans. 1* **1975**, 1120. (f) Schmidt, B.; Fleischhauer, J. *J. Chem. Inf. Comput. Sci.* **1978**, 18, 204. (g) Gasteiger, J.; Jochum, C. *J. Chem. Inf. Comput. Sci.* **1979**, 19, 43. (h) Roos-Kozel, B. L.; Jorgensen, W. L. *J. Chem. Inf. Comput. Sci.* **1981**, 21, 101. (i) Hendrickson, J. B.; Grier, C. L.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1984**, 24, 195.
- (4) (a) Corey, E. J.; Wipke, W. T. *Science (Washington, D.C.)* **1969**, 166, 178. (b) Corey, E. J.; Petersson, G. A. *J. Am. Chem. Soc.* **1972**, 94, 460.
- (5) Fujamann, R.; Dölling, U.; Nickelsen, H. *Angew. Chem., Int. Ed. Engl.* **1967**, 6, 723.
- (6) Wipke, W. T.; Dyott, T. M. *J. Chem. Inf. Comput. Sci.* **1975**, 15, 140.
- (7) Hüchel, W.; Danneel, R.; Schwartz, A.; Gerke, A. *Liebigs Ann. Chem.* **1929**, 474, 121.
- (8) For the construction of an ITS, see ref 1 and 2a.
- (9) The ITS is stored in a computer in terms of an ITS connection table in which connectivity between each pair of nodes is represented by a complex bond number (*a b*) or a color-bond number (*o, i, p*). See Appendix I.
- (10) The starting molecules are generated by the PS operation (projection to starting stage), which is deletion of in-bonds in a graphical expression. The products are reproduced by the PP operation (projection to product stage), which deletes out-bonds in a graphical expression. See Appendix I.
- (11) Multi-tied rings are not adopted here because of the need to save various arrays that are concerned with rings detected. However, another algorithm that considers both tied and multi-tied rings for covering conditions would be adopted for some purposes.
- (12) Hirano, S.; Hara, H.; Hiyama, T.; Fujita, S.; Nozaki, H. *Tetrahedron* **1975**, 31, 2219.
- (13) An algorithm for detecting all rings has been reported. See: Kudo, Y. in *Organic Synthesis and Computers* CMC Press: Tokyo, 1983; p 49. See also ref 6.
- (14) Meiser, W. *Ber.* **1899**, 32, 2055.
- (15) Felix, D.; Muller, R.; Horn, U.; Schreiber, R.; Eschenmoser, A. *Helv. Chim. Acta* **1972**, 55, 1276.
- (16) Tarbell, D. S. *Org. React.* **1944**, 2, 1.
- (17) Donaruma, L. G.; Heldt, W. Z. *Org. React.* **1960**, 11, 1.
- (18) Criegee, R.; Kaspar, R. *Liebigs Ann. Chem.* **1948**, 560, 127.
- (19) von Auwers, K.; Ziegler, K. *Liebigs Ann. Chem.* **1921**, 425, 217.
- (20) (a) Fujita, S.; Nozaki, H. *Bull. Chem. Soc. Jpn.* **1971**, 44, 2827. (b) Fujita, S.; Nozaki, H. *Yuki Gosei Kagaku Kyokaiishi* **1972**, 30, 679.
- (21) (a) Ryan, A. W.; Stobaugh, R. E. *J. Chem. Inf. Comput. Sci.* **1982**, 22, 22. (b) Freeland, R. G.; Funk, S. A.; O'Korn, L. J.; Wilson, G. A. *J. Chem. Inf. Comput. Sci.* **1979**, 19, 94.

## Automated Classification of Candidate Structures for Computer-Assisted Structure Elucidation

ALAN H. LIPKUS<sup>1</sup> and MORTON E. MUNK\*

Department of Chemistry, Arizona State University, Tempe, Arizona 85287

Received March 12, 1987

In computer-assisted structure elucidation, a large number of candidate structures for the unknown compound can be generated. A computer program designed to aid in the recognition of significant differences between these structures is described. The goal of the program is to group the candidates into meaningful classes with a minimum of input from the user; each class is distinguished by a substructure its members have in common. A computational approach is developed based on the combinatorial problem of set covering. The program evaluates its results using an information-theoretical criterion. An application of the program to a real-world structure problem is presented.

### INTRODUCTION

The field of computer-assisted structure elucidation now comprises many different computer programs designed to facilitate the process by which a chemist deduces the structure of an unknown compound from spectroscopic and chemical data.<sup>2</sup> Of central importance to this field are programs capable of generating all the molecular structures satisfying certain given constraints that reflect the current state of knowledge about the unknown compound. These so-called structure generators allow the chemist to examine all plausible candidates for the unknown at any stage of the elucidation process. Several of these programs, such as ASSEMBLE,<sup>3</sup> CONGEN,<sup>4</sup> GENOA,<sup>5</sup> and COCOA<sup>6</sup> are highly sophisticated and offer considerable breadth in the range of constraining information they can accept from the chemist. The recently developed program COCOA is designed to receive most of its structural input directly from spectrum-interpreting programs. Some other more specialized structure generators<sup>7,8</sup> developed earlier also function in this manner.

When the available information is insufficient to narrow the set of generated structures to one, this set must be examined for the purpose of finding additional information with which to constrain structure generation more severely. The result of this examination is often the planning of new experiments. This step in the computer-assisted structure elucidation process

requires the chemist to perceive chemically or spectroscopically meaningful differences between the candidate structures. This may be impractical, however, if the number of candidates generated is very large. For this reason, structure-editing programs have been devised to assist the chemist in examining large sets of candidate structures.<sup>9,10</sup> These programs allow the creation of subsets of candidates possessing certain substructural features, or combinations of features, selected by the user. Organizing the candidates in this manner can facilitate the perception of experimentally meaningful structural differences.

In the present work, a different approach is explored to the problem of examining the potentially large number of candidate structures generated for an unknown compound. A computer program is described that, like a structure editor, places candidates into different groups on the basis of selected substructural features. Unlike an editor, however, this program aims to select these features largely by itself, with a minimum of input from the user.

### CLASSIFICATION SCHEME FOR CANDIDATE STRUCTURES

A method of classifying candidate structures is proposed that is both useful to the chemist and amenable to computational treatment. In this classification scheme all the candidate