

The description of organic reactions based on imaginary transition structures. A novel approach to the computer-oriented taxonomy of organic reactions

Shinsaku Fujita

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

Abstract - An imaginary ring that appears in an imaginary transition structure (ITS) is a descriptor for a ring-opening, ring-closure, or rearrangement reaction. An n -nodal subgraph of the ITS is a descriptor of a reaction-site change such as substitution, construction, cleavage, addition, and elimination. The n -nodal subgraph is subdivided into a reaction kernel and a terminal descriptor, the latter of which is useful to classify oxidation and reduction. The ITS that corresponds to an individual reaction is represented uniquely and unambiguously by a newly-defined canonical code.

INTRODUCTION

We have presented the concept of imaginary transition structure (ITS) for the computer-oriented representation of organic reactions (ref. 1). The ITS approach provides the comprehensive representation of: (1) individual reactions along with reaction types (ref. 1b,c,f,g), (2) the participation of substrates, reagents, catalysts, and products (ref. 1a), (3) skeletal changes (ref. 1i) and reaction-site changes (ref. 1d), (4) the dual character of a bond (ref. 1h), (5) the stoichiometry of a reaction, (6) the stereochemistry of a reaction (ref. 1h), and (7) multistep reactions (ref. 1e). The ITS representation thereby embraces conventional reaction equations as well as other alternatives proposed in the past (ref. 1j,2). This lecture deals with the computer-oriented taxonomy of organic reactions for our system (FORTUNITS: Fuji Organic Reaction Treating UNITY based on ITS).

IMAGINARY TRANSITION STRUCTURE (ITS)

An ITS is a kind of structure (or structural formula) in which the participant nodes are linked by par- (—), out- (—H—) and in-bonds (—⊙—) in accord with structural changes during a reaction (ref. 1a). For example, the ITS (1) represents the bromination of naphthalene with bromine. The corresponding starting stage is regenerated by projection to starting stage (PS) that is the deletion of all in-bonds from the ITS. Projection to product stage (PP) is the deletion of all out-bonds, which affords the product stage. The ITS contains a set of imaginary bonds (IB) selected from the 15 combinations of par-, out-, and/or in-bonds (Table 1). Each of the IBs is represented by a pair of integers (a b) called a complex bond number (CBN). The IBs and CBNs are the extended concepts of bonds and bond multiplicities, respectively, in the field of organic structure. The ITS is thereby extended structure that embraces organic structure. The internal representation of the ITS is a connection table (or its equivalent) that defines the connectivities between the nodes by means of the CBNs.

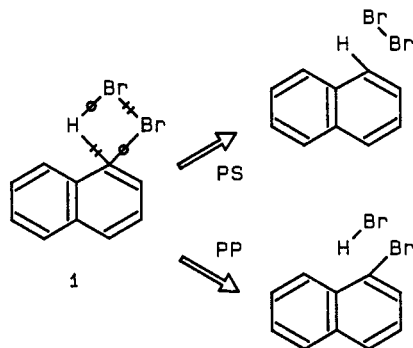


Table 1. Imaginary bonds and complex bond numbers

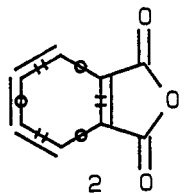
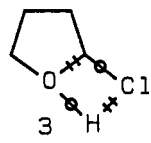
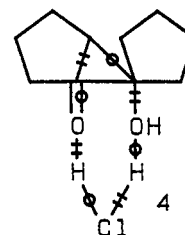
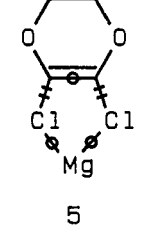
$b =$	-3	-2	-1	0	+1	+2	+3
$c = 1$			—H—	—	—⊙—		
			(1-1)	(1+0)	(0+1)		
$c = 2$		==	==H—	==	==⊙—	==	==
		(2-2)	(2-1)	(2+0)	(1+1)	(0+2)	
$c = 3$	≡≡	≡≡H—	≡≡	≡≡	≡≡⊙—	≡≡	≡≡
	(3-3)	(3-2)	(3-1)	(3+0)	(2+1)	(1+2)	(0+3)

$$c = a + (b + |b|)/2 \text{ for CBN } (a \ b)$$

IMAGINARY RINGS. DESCRIPTORS FOR SKELETAL CHANGES

A ring structure (an imaginary ring) that appears in an ITS (Table 2) is a descriptor for a ring opening, ring-closure, or rearrangement reaction (ref. 1a,i). The ITS (2) has a 6-membered BC_2 that indicates the formation of a 6-membered ring during the course of the Diels-Alder₂ reaction. The ring-opening of tetrahydrofuran is characterized by a 5-membered BO , that appears in the ITS (3). The pinacol rearrangement is designated by a 3-membered BR as illustrated in the ITS (4). The ITS (4) also contains a 5-membered BO_1 and a 6-membered BC_1 . The combination of the BO_1 , BC_1 , and BR describes the total feature of the reaction (a ring enlargement by the pinacol rearrangement). The ITS (5) contains another type of 6-membered ring that holds the skeleton but changes the functionality during the reaction. This type of ring is called a reactive intact ring. The essential set of essential rings (ESER) selects the 6-membered BC_2 but no 9-membered BC_2 from the ITS (2) (ref. 1k).

Table 2. Imaginary rings

Ring type (Bridge of:)	Size	No. of IBs ^a			Remarks
		Type A	Type B	Type C	
 2	r	r - m	m	0	$m \geq 1$
 3	r	r - n	0	n	$n \geq 1$
 4	r	r - 2	1	1	
 5	r	r	0	0	
intact ring ^b IR	r	r	0	0	
trivial ring ^c	r	r - m - n	m	n	$m + n > 2$, $mn \neq 0$

^aType A, B, and C bonds are IBs of (a b) in which $a + b \neq 0$ and $a \neq 0$; $a + b = 0$; and $a = 0$, respectively.

^bAn intact ring is classified into a reactive and a complete intact ring.

^cA trivial ring may be a reaction string (ref. 1a).

HIERARCHICAL N-MODAL SUBGRAPHS. DESCRIPTORS FOR REACTION-SITE CHANGES

The ITS (1) contains a three-nodal subgraph ($Br \text{---} C \text{---} H$) in which a carbon reaction center is attached by two terminal atoms (bromine and hydrogen) through colored bonds (nonparabonds). The subgraph is codified into $Br(0+1)C(1-1)H$ in terms of the CBNs. The subgraph and the code correspond strictly to the IUPAC name, bromo-dehydrogenation (ref. 3). Moreover, the PS and PP operations generate a reaction equation, $Br + C-H$ (PS) \rightarrow $Br-C + H$ (PP), which meets chemist's convention. When we consider such hierarchies of atoms as $Br \subset Hal$ (halogen) $\subset Z$ (electronegative atom) and $H \subset HH$ (hyperhydrogen), we obtain more generic descriptors, $Hal \text{---} C \text{---} HH$ (hologenation) and $Z \text{---} C \text{---} HH$ (oxidative substitution), in ascending order of generality.

The ITS (5) affords a four-nodal subgraph, $Cl \text{---} C \text{---} C \text{---} Cl$ (dichloro-elimination), which similarly provides more generic subgraphs (Table 3). In Table 3, the classification levels of biology (phylum, class, etc.) are applied to designate the reaction hierarchies. The variety of natural language terms corresponding to the respective levels (Table 3) shows that the ITS is capable of providing systematic descriptors in the form of four-nodal subgraphs.

In general, an n-nodal subgraph is a descriptor of a reaction-site change. The subgraph is subdivided into a reaction kernel (RK) and a terminal descriptor (TD). For example, the four-nodal subgraph $Cl \text{---} C \text{---} C \text{---} Cl$ consists of an RK ($K_e: C \text{---} C$) and a TD ($Cl \text{---} K \text{---} Cl$). The TD is useful to classify oxidation and reduction, since all even-nodal subgraphs provide a common set of TDs and since all odd-nodal subgraphs give another common set of TDs. These facts stem from the alternant character of reaction strings (ref. 1m). Table 4 collects the TDs of order level derived from even-nodal subgraphs.

Table 3. Systematic classification of organic reactions by subgraphs of ITSs

level ^a	descriptor	example	natural language term
phylum	stringity ^b	one-string	
class	no. of C atoms in an RK	C ₂	
division	reaction kernel (RK)	C=C (= Ke)	double-bond formation
section	terminal descriptor (TD) ^c	A-+-Ke-+-A	elimination
order	terminal descriptor (TD)	Z-+-Ke-+-Z	reductive elimination
family	terminal descriptor (TD)	Hal-+-Ke-+-Hal	dehalogenation (dihalo-elimination)
genus	terminal descriptor (TD)	Cl-+-Ke-+-Cl	dechlorination (dichloro-elimination)
species	ITS	ITS (5)	

^aIntermediate levels may be added if necessary. For example, a TD of subfamily level is derived in terms of such terminal nodes as H and as Hal. Compare this with the TD that contains HH and Hal as terminal nodes.

^bFor the definition of stringity, see ref. 1a.

^cThe capital A represents any atom.

Table 4. Terminal descriptors of order level for even-nodal subgraphs^{a,b}

Oxidative (super)elimination:	Reductive (super)addition:
HH-+-Ke-+-HH	HH-○-Ke-○-HH
Isohypsic (super)elimination: ^c	Isohypsic (super)addition:
HH-+-Ke-+-Z HH-+-Ke- <u>+-</u> Z	HH-○-Ke-○-Z HH-○-Ke- <u>○</u> Z
HH-+-Ke- <u>+-</u> Z	HH-○-Ke- <u>○</u> Z
Reductive (super)elimination:	Oxidative (super)addition:
Z-+-Ke-+-Z Z- <u>+-</u> Ke- <u>+-</u> Z	Z-○-Ke-○-Z Z- <u>○</u> Ke- <u>○</u> Z
Z-+-Ke- <u>+-</u> Z Z- <u>+-</u> Ke- <u>+-</u> Z	Z-○-Ke- <u>○</u> Z Z- <u>○</u> Ke- <u>○</u> Z
Z-+-Ke- <u>+-</u> Z Z- <u>+-</u> Ke- <u>+-</u> Z	Z-○-Ke- <u>○</u> Z Z- <u>○</u> Ke- <u>○</u> Z

^aThe terms superelimination and superaddition are the extended concepts of the elimination and addition that are concerned with four-nodal subgraphs. The TDs of family level can be obtained by the combination of HH, Hal and Het. Similarly, the TDs of genus level can be derived by using atomic symbols as illustrated in Table 3.

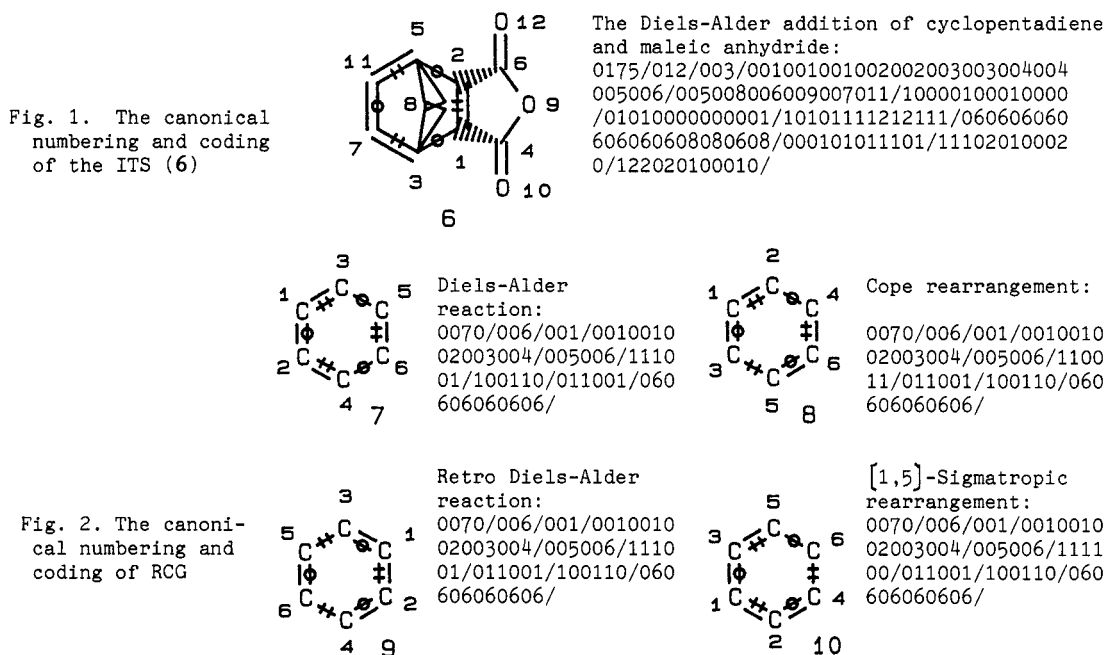
^bTerminal descriptors of order level for odd-nodal subgraphs can be introduced in the same line. See ref. 1d and 1m.

^cFor the term isohypsic, see J. B. Hendrickson, *J. Am. Chem. Soc.*, 93, 6847 (1971). The ITS approach has been compared with Hendrickson's coding system (ref. 1d).

NAMING OF INDIVIDUAL ORGANIC REACTIONS AND OF REACTION TYPES. CANONICAL CODES OF ITS'S AND OF REACTION-CENTER GRAPHS

An ITS stored in the form of an ITS connection table may vary due to the numbering of the nodes. The canonical numbering and coding of the ITS are necessary to construct an effective computer system. We present here a novel code (CANITS; the canonical name of an ITS) that is the linear combination of the following lists: (a) the length of the canonical name, (b) the number of nodes considered, (c) the number of rings contained in the ITS (d) FROM list, (e) RING CLOSURE list, (f) PAR-BOND list, (g) IN-BOND list, (h) OUT-BOND list, (i) ATOM list, (j) INTACT NODE list, (k) STARTING STEREO list, and (l) PRODUCT STEREO list. The CANITSs are obtained by two steps: (1) the partial partitioning of the nodes in the light of 4 kinds of extended connectivities and (2) the selection of the lexicographically smallest code. Fig. 1 exemplifies the canonical numbering of the ITS (6) of a Diels-Alder addition as well as the corresponding CANITS.

A reaction-center graph (RCG) is defined as a set of reaction centers and nonpar-bonds. The RCG is a more specific subgraph than an n-nodal subgraph and is a useful descriptor of a reaction type. For example, the RCG (7) derived from the ITS (6) corresponds to the generic Diels-Alder reaction. The CANITS algorithm is applicable to give RCGs the canonical codes that contain the lists (a-i). Fig. 2 shows several examples of the CANITSs of RCGs.



CONCLUSION

A compound can be characterized by the fragments contained within it. This means that the structure-substructure relationship has provided a sound basis for a retrieval and design system of organic compounds. The present formulation of ITSs and their substructures (subgraphs) has introduced an analogous but extended foundation to describe organic reactions. The manipulation of organic reactions has thereby been replaced by the handling of ITSs, which provides us with broader prospects than the conventional methods do.

REFERENCES

1. S. Fujita, *J. Chem. Inf. Comput. Sci.*, 26, (a) 205-212; (b) 212-223; (c) 224-230; (d) 231-237; (e) 238-242 (1986); 27, (f) 99-104; (g) 104-110; (h) 111-115; (i) 115-120; (j) 120-126 (1987); 28, (k) 1-9; (l) in press (1988). (m) S. Fujita, *J. Chem. Soc., Perkin 2*, in press (1988).
2. (a) S. Fujita, *Yuki Gosei Kagaku Kyokaiishi*, 44, 354-364 (1986). (b) P. Willett (ed.), *Modern Approaches to Chemical Reaction Searching*, Gower, Aldershot (1986). (c) J. E. Ash, P. A. Chuff, S. E. Ward, S. M. Welford, P. Willett, *Communication, Storage and Retrieval of Chemical Information*, Ellis Horwood, Chichester (1985).
3. (a) J. F. Bunnett, *Pure Appl. Chem.*, 53, 305-321 (1981). (b) J. March, *Advanced Organic Chemistry*, 3rd Ed., Wiley, New York (1985).