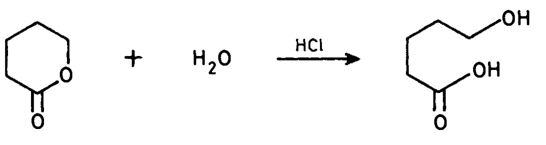
A Novel Approach to Systematic Classification of Organic Reactions. Hierarchical Subgraphs of Imaginary Transition Structures[1]

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

A new systematic classification of organic reactions is presented in terms of imaginary transition structures (ITS) and their hierarchical *n*-nodal subgraphs. The ITS corresponding to an individual organic reaction is a structural formula which has par-bonds (invariant bonds during the reaction), out-bonds (bonds appearing in the starting stage), and in-bonds (bonds appearing only in the product stage). The *n*-nodal subgraph indicates the corresponding reaction type and is divided into two parts: reaction kernels (RK) and terminal descriptors (TD). The RK is a set of adjacent carbon reaction centres to which out- and in-bonds are incident. The RK affords information on the changes of carbon skeletons, *e.g.*, 'substitution', 'construction (C–C bond formation)', 'cleavage', 'addition', and 'elimination'. The TD is a set of terminal non-carbon atoms and imaginary bonds incident to the terminals. Odd- and even-nodal subgraphs have common TDs, respectively, which afford information on the changes of oxidation states, *e.g.*, 'oxidative', 'reductive', and 'isohypsic'. This fact stems from the alternant character of reaction strings.

Systematic representation of organic reactions and logical characterization of the reaction features are of crucial importance in constructing computer systems both for retrieval and for synthetic design.[2] For the purpose of retrieval, unequivocal descriptors should be selected to designate organic reactions.

Using the hydrolysis of 5-pentanolide as an example we should extract a descriptor 'hydrolysis' (or the corresponding code which is a subclass of 'substitution') from the Scheme.



Scheme.

Moreover, another type of descriptor 'ring opening' (or the corresponding machine-readable code) should also be selected. Thus, there are at least two categories of descriptors for characterizing organic reactions. The first type of descriptors are concerned with *reaction-site changes*, *e.g.* substitution, construction (C–C bond formation), cleavage, addition, and elimination. They are focused on local changes of the participating molecules. On the other hand, the second type of descriptors indicate *skeletal changes* during the reactions, *e.g.*, ring opening, ring closure, ring construction, ring enlargment, and rearrangement. These descriptors are whole respect to the whole structure of the molecules. The two types of descriptors can be abstracted from a reaction diagram such as the Scheme only by chemical intuition. The process of recognition based on reaction diagrams is complex enough to be applied to computer manipulation. Hence, a new representation of an organic reaction should be so constructed that it can provide the two types of descriptors by a logical procedure.

Several representative methods can be reviewed from the viewpoint described above. In conventional methods, an organic reaction used to be represented by a reaction diagram such as the Scheme, which involves the structures of the starting material(s) and of the product(s) combined with an arrow. The two types of descriptors used to be extracted from the reaction diagram, where the difference between two or more structures is detected to be a reaction-site or a skeletal change. However, the detection of the difference is not so easy for a computer, since the correspondence of the starting nodes (atoms) to the product nodes is not unique in the reaction diagram.[2]

Ugi[3] overcame this difficulty by considering that the product stage (represented by an E matrix) is an isomer of the starting stage (B matrix). He formulated an organic reaction as an R matrix (R = E − B) which is concerned only with the above described reaction-site changes. Since the R matrix is a net change during the reaction, it is too simple to describe *exact* reaction-site changes. For example, the R matrix of the Diels–Alder addition is the same as those of the Claisen and the Cope rearrangement. Moreover, the R matrix provides no information on skeletal changes during organic reactions.

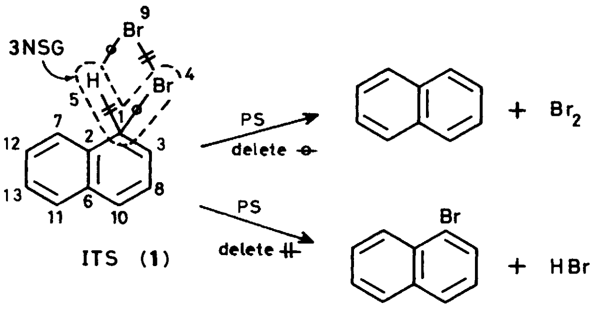
Kvasnicka[4] proposed reaction graphs which can be considered as a representation of Ugi's R matrix. He added an intact molecular subgraph to the reaction graph in order to remedy its oversimplification. Since his method focused only on reaction sites, skeletal changes such as ring opening and closure cannot be detected.

Roberts' coding system,[5] Arens' $(+-)_n$ method,[6] and Hendrickson's method[7] also relate only to reaction sites and have the same disadvantage as described above.

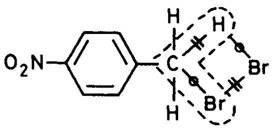
Wilcox and Levinson[8] reported a bond-centred labelled graph which represents a total reaction in itself. Their representation deals with organic compounds as well as organic reactions. However, the representation of an organic compound is remarkably different from the usual structural formula. Thus, a ring structure in this representation is not always a ring in the actual participating compounds. Hence, a rather complex algorithm is necessary to detect ring opening, ring closure, and so on.

Vladutz[9] reported the concept of superimposed reaction graphs (SRG), which contain no catalysts even if they participate in the corresponding reactions. Superimposed reaction skeleton graphs (SRSG) were also proposed to classify reaction-site changes of organic reactions. His method has a potential ability to describe skeletal changes which have never been discussed.

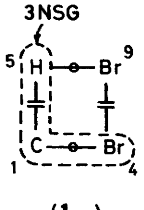
In the previous papers,[10] I have proposed imaginary transition structures (ITS) as unitary representations of organic reactions (Appendix). Descriptors of the first type for reaction-site changes can be extracted as subgraphs of ITS by simple



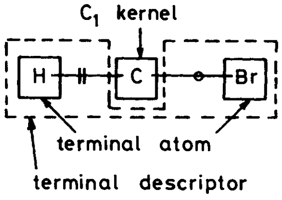
ITS (1)



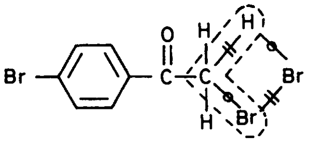
(2)



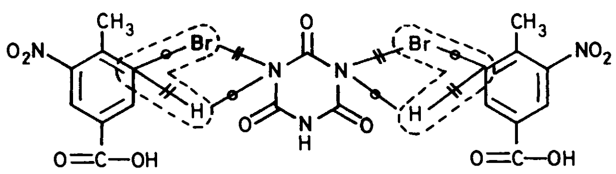
(1_{rc})



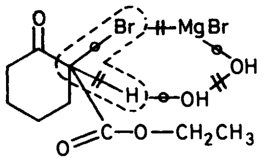
3NSG (1_{sg})



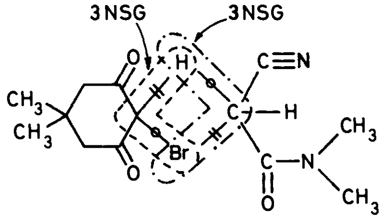
(3)



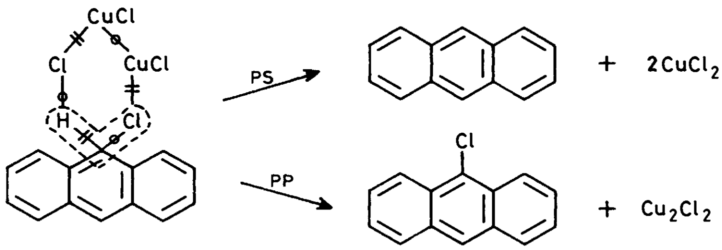
(4)



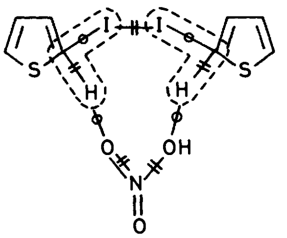
(5)



(6)



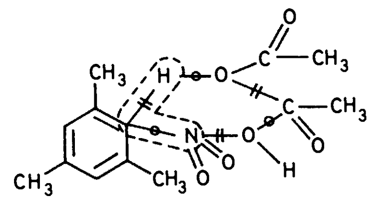
(7)



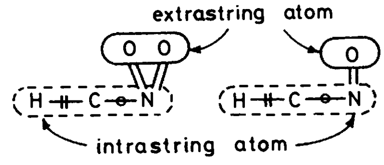
(8)

algorithms: reaction centre (RC) graphs,[10a,b] three-nodal and four-nodal subgraphs.[10d] The examination of ring structures appearing in ITS provides the descriptors of the second type for skeletal changes.[10a]

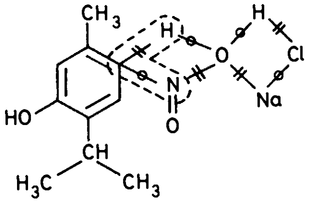
This paper deals with the systematic characterization of organic reactions based upon ITS. Detailed discussion especially on the three- and four-nodal subgraphs affords a reaction hierarchy which is effective in classifying reaction-site changes



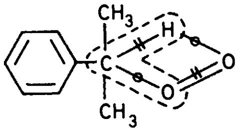
(9)



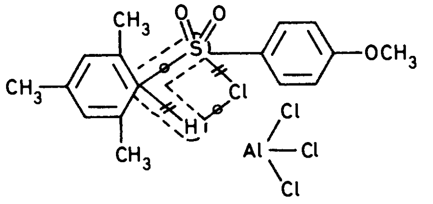
(9_{sg}) (10_{sg})



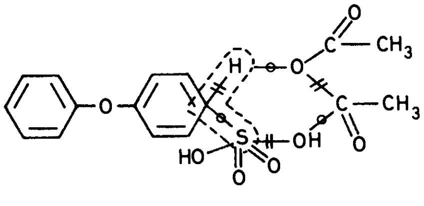
(10)



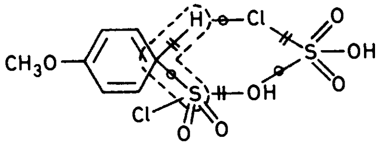
(11)



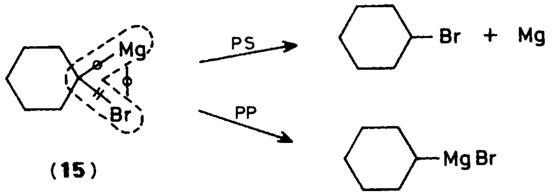
(12)



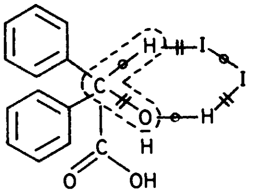
(13)



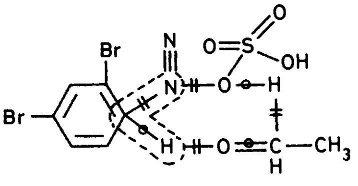
(14)



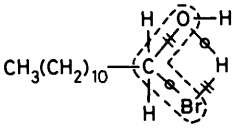
(15)



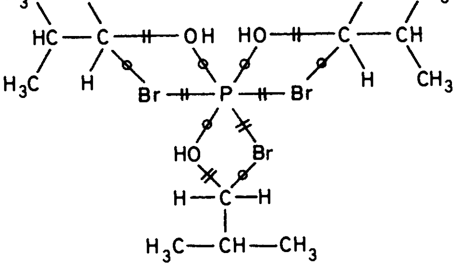
(16)



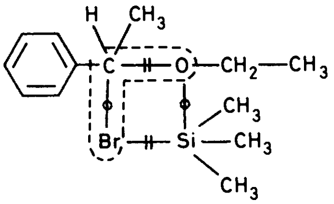
(17)



(18)



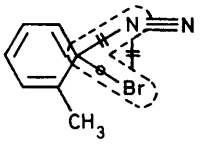
(19)



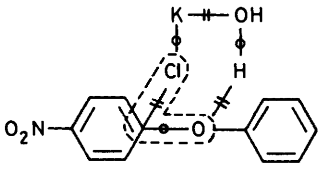
(20)

during organic reactions. The five- and six-nodal subgraphs are also discussed as extensions of the three- and four-nodal subgraphs, respectively. For this purpose, reaction kernels (RKs) and terminal descriptors (TDs) are proposed as key concepts.

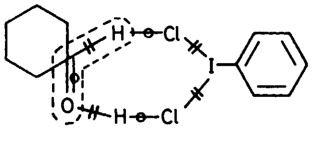
Imaginary Transition Structures (ITS), ITS Connection Tables, Projections to Starting and to Product Stage.—An imaginary transition structure (**1**) for bromination of naphthalene[11] is a kind of structural formula which contains



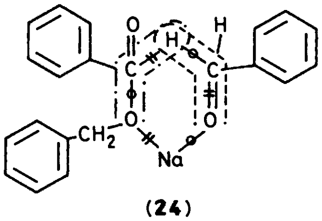
(21)



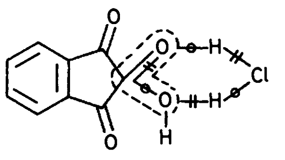
(22)



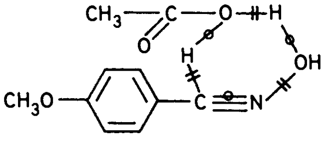
(23)



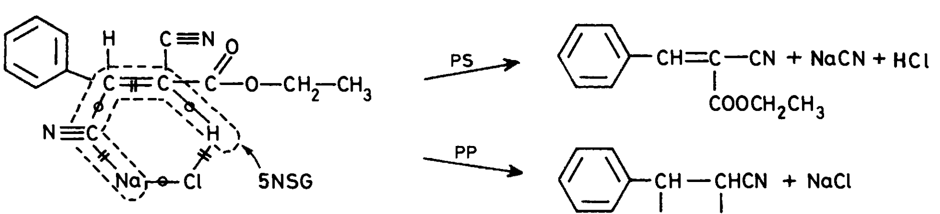
(24)



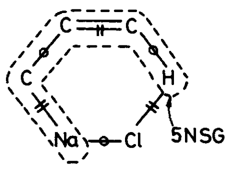
(25)



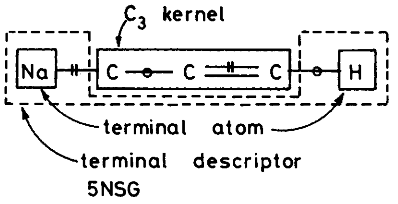
(26)



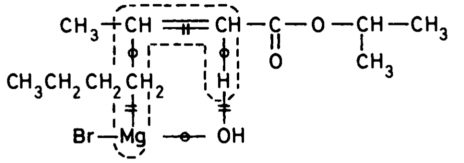
(27)



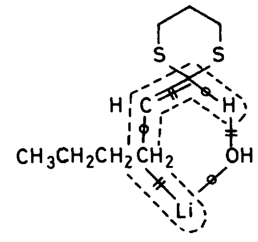
(27rc)



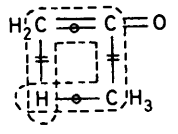
(27sg)



(28)



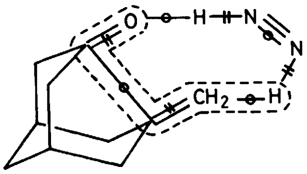
(29)



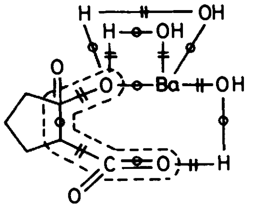
(30)

three types of bonds, out-bonds (⧺), in-bonds (⦶), and par-bonds (–). The ITS (**1**) is stored in the form of an ITS connection table (Table), in which each pair of nodes is connected with an imaginary bond (or ITS bond) of complex bond number (*a b*).

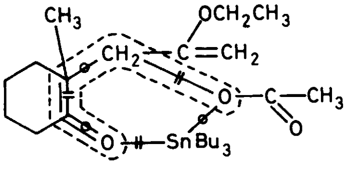
The starting molecules, *i.e.* naphthalene and bromine, can be regenerated by projection to starting stage (PS) which corresponds to the graphical operation of deleting in-bonds or to the computer operation of replacing each bond of (*a b*) by



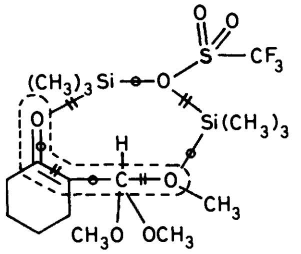
(31)



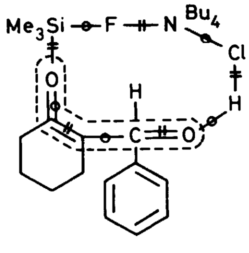
(32)



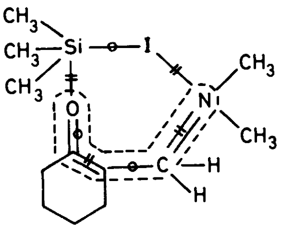
(33)



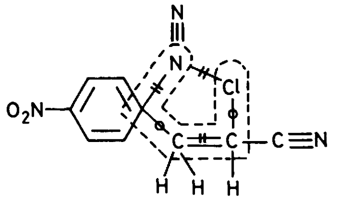
(34)



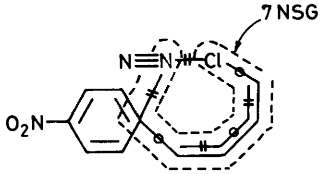
(35)



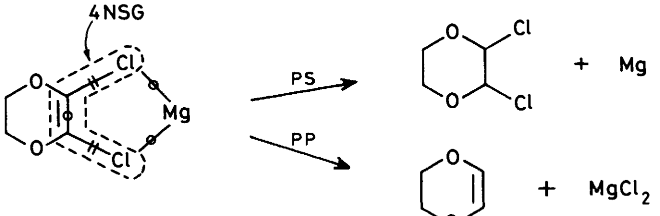
(36)



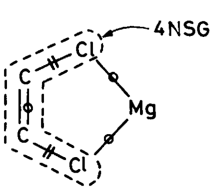
(37)



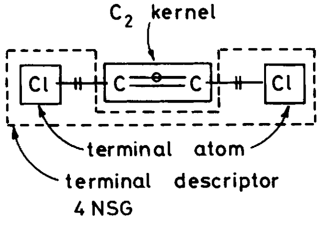
(38)



(39)

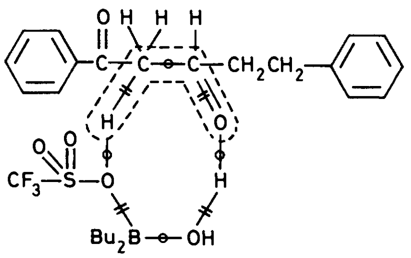


(39_{rc})

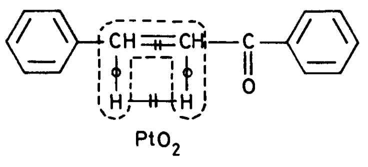


(39_{sg})

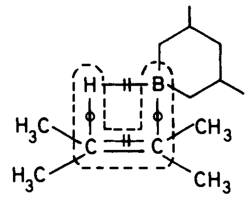
the usual bond of multiplicity *a*. The projection to product stage (PP) is the deletion of out-bonds and provides the product molecules, *i.e.* 1-bromonaphthalene and hydrogen bromide. The PP operation is the replacement of (*a b*) by the multiplicity *a* + *b*. The following discussions are based mainly on graphical expressions. However, the corresponding computer-readable expressions are easily obtained in terms of complex bond numbers.



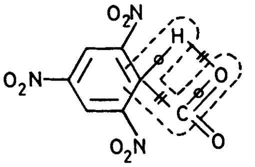
(40)



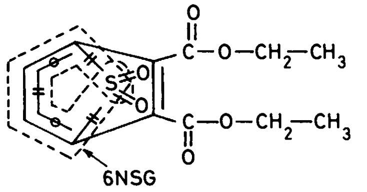
(41)



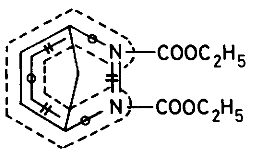
(42)



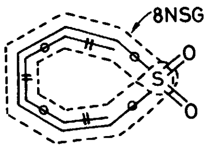
(43)



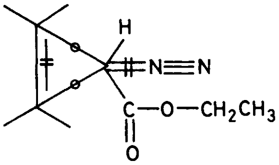
(44)



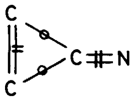
(45)



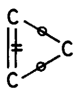
(46)



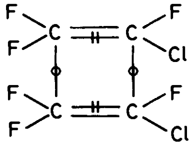
(47)



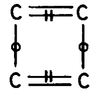
(47_{rc})



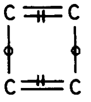
(47_{rk})



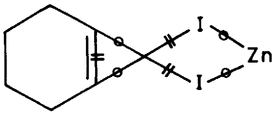
(49)



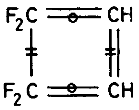
(49_{rc})



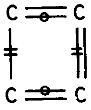
(49_{rk})



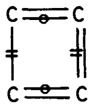
(48)



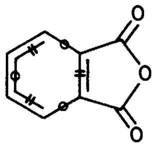
(50)



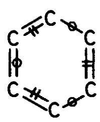
(50_{rc})



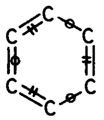
(50_{rk})



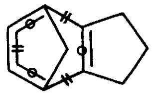
(51)



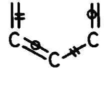
(51_{rc})



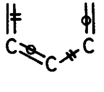
(51_{rk})



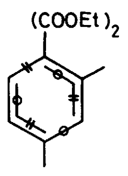
(52)



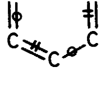
(52_{rc})



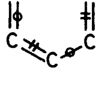
(52_{rk})



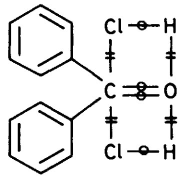
(53)



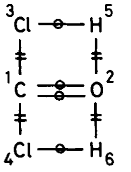
(53_{rc})



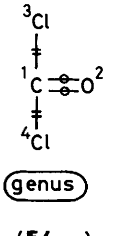
(53_{rk})



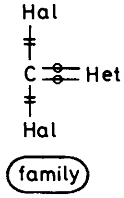
(54)



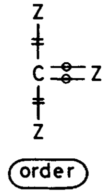
(54_{rc})



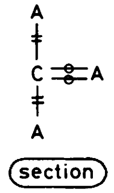
(54_{ge})



(54_{fa})

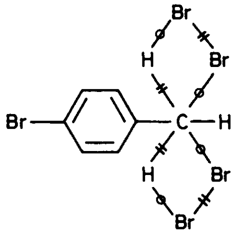


(54_{or})

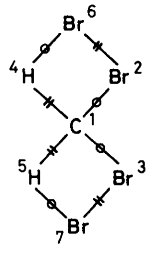


(54_{se})

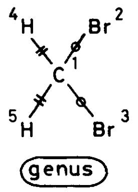
Organic chemists who used to think in the light of ordinary reaction diagrams may feel that the systematic characterization proposed here is artificial and strange. However, if they reconsider their own processes of recognizing organic reactions, they should become aware that they compare the starting structure and the product structure in order to differentiate the two sides of the diagram. It is to be emphasized that the ITS approach has formulated these processes of recognition and



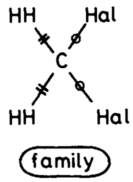
(55)



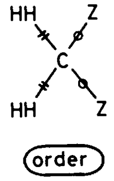
(55_{rc})



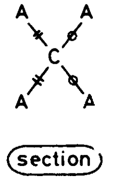
(55_{ge})



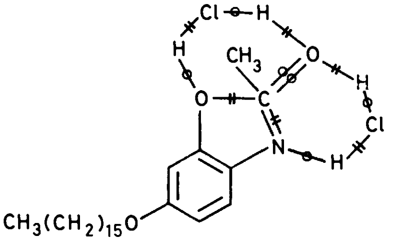
(55_{fa})



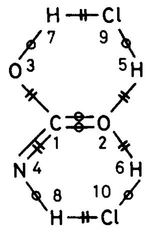
(55_{or})



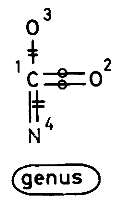
(55_{se})



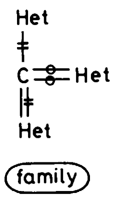
(56)



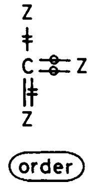
(56_{rc})



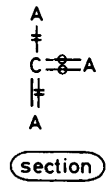
(56_{ge})



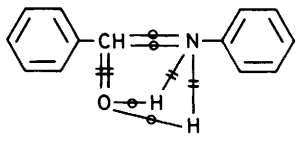
(56_{fa})



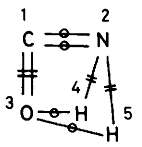
(56_{or})



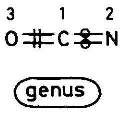
(56_{se})



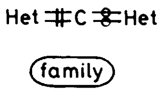
(57)



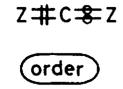
(57_{rc})



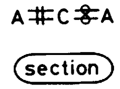
(57_{ge})



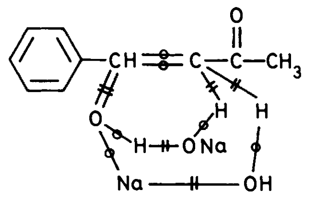
(57_{fa})



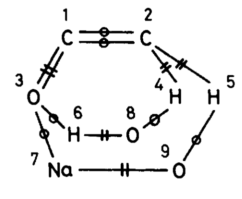
(57_{or})



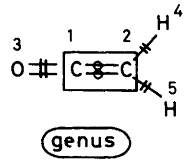
(57_{se})



(58)

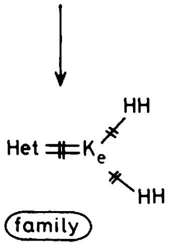


(58_{rc})

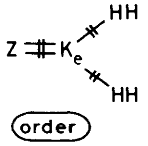


(58_{ge})

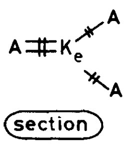
C_2 reaction kernel



(58_{fa})



(58_{or})



(58_{se})

provided clear expressions of reaction features so as to make computer manipulation easier. Moreover, the classification by ITS subgraphs described below affords a deeper insight into organic reactions which may be overlooked by the conventional methods. It should be noted that ITS is intended not to represent an exact mechanism but to give an effective methodology to classification of organic reactions.

Reaction Hierarchy based on Three-nodal Subgraphs.—When all reaction centres (RC) are collected from ITS (**1**), an RC

Table. Connection table of ITS (**1**)[a]

Node	Atom or group	Co-ordinate[b]		Neighbour 1		Neighbour 2		Neighbour 3		Neighbour 4	
		x	*y*	node	(*a b*)[c]	node	(*a b*)	node	(*a b*)	node	(*a b*)
1	C	0	0	2	(1 + 0)	3	(2 + 0)	4	(0 + 1)	5	(1 − 1)
2	C	−75	−50	1	(1 + 0)	6	(2 + 0)	7	(1 + 0)		
3	CH	75	−50	1	(2 + 0)	8	(1 + 0)				
4	Br	100	100	1	(0 + 1)	9	(1 − 1)				
5	H	−100	100	1	(1 − 1)	9	(0 + 1)				
6	C	−75	−150	2	(2 + 0)	10	(1 + 0)	11	(1 + 0)		
7	CH	−150	0	2	(1 + 0)	12	(2 + 0)				
8	CH	75	−150	3	(1 + 0)	10	(2 + 0)				
9	Br	0	200	4	(1 − 1)	5	(0 + 1)				
10	CH	0	−190	6	(1 + 0)	8	(2 + 0)				
11	CH	−150	−190	6	(1 + 0)	13	(2 + 0)				
12	CH	−230	−50	7	(2 + 0)	13	(1 + 0)				
13	CH	−230	−150	11	(2 + 0)	12	(1 + 0)				

[a] The numbering of nodes, which is not canonized, is given in ITS (**1**). [b] Stereochemical features can be described by addition of *z* co-ordinates. [c] For the definition of complex bond numbers, see Appendix.

graph ($\mathbf{1}_{rc}$) is obtained as a subgraph.[10b] This consists of a reaction string, in which out-bonds and in-bonds appear alternately. The RC graph ($\mathbf{1}_{rc}$) is represented also by a coded name |4L|H(1 − 1)C(0 + 1)Br(1 − 1)Br(0 + 1)H. The term |4L| denotes that the RC graph is looped to be a four-membered ring. This RC graph ($\mathbf{1}_{rc}$) is a descriptor of the bromination (or oxidative substitution with bromine) at a carbon atom. When we select a three-nodal subgraph (3NSG) having the central carbon atom and two terminal non-carbon atoms, we obtain a more generic descriptor (H#C•Br), or H(1 − 1)C(0 + 1)Br, corresponding to bromination. The term 'bromination' stems from comparison of the two terminal atoms (H and Br). This descriptor pays no attention whether bromination is effected by bromine or by other brominating agents. In general, an *n*-nodal subgraph is defined as a subgraph in which *n* − 2 of carbon reaction centres are linked to each other with in- or out-bonds and two non-carbon reaction centres are linked to either of carbon reaction centres.

The three-nodal subgraph ($\mathbf{1}_{sg}$) is divided into two parts. A maximum subgraph involving only adjacent carbon atoms to which out-bonds and in-bonds are incident is defined as a reaction kernel (RK). In this case, a C_1 reaction kernel denotes a substitution reaction at a carbon atom. A terminal descriptor (TD) is defined as a set of terminal atoms and imaginary bonds which combine a reaction kernel with terminals.

Brominations with bromine as reagent shown by ITS (**2**)—(**6**) have the same RC graph ($\mathbf{1}_{rc}$), three-nodal subgraph (3NSG) encircled with broken lines, and RK as those of (**1**). It should be emphasized that all these appear as the subgraphs in the corresponding ITSs. Detailed information on each reaction is obtained by examining environmental atoms (extrastring atoms) attached to the intrastring carbon atoms. This examination is again a substructure search of ITSs, since the ITSs involves the whole information on structural changes. For example, (**1**) is concerned with bromination of an aromatic hydrocarbon, which is perceived by the fact that the carbon atom of the reaction centre is a substructure of the aromatic nucleus.

Entry (**2**)[12] is α-bromination of an aromatic compound. Entry (**3**) is α-bromination of a carbonyl compound.[13] Entry (**4**) is bromination of an aromatic compound with dibromoisocyanuric acid.[14] Each RC graph of the separate two-string ITS (**4**) has the corresponding code |4L|H(1 − 1)C(0 + 1)Br(1 − 1)-N(0 + 1)H. ITS (**5**) is an example in which a brominating agent is generated *in situ* by oxidation of an appropriate bromide species ($MgBr_2$) with hydrogen peroxide.[15] The same 3NSG (H#C•Br) appears in this case. Entry (**6**) is an example of bromination by means of bromine exchange.[16] Usually 5,5-dimethylcyclohexane-1,3-dione is considered to be a substrate in this case. However, it can be regarded as a reagent (dehydrobrominating agent) from the opposite viewpoint. Hence, methods of representation of organic reactions must be flexible in case our points of view are changed in future. The ITS approach does recognize substrates and reagents as equal components of each ITS. Thus, we can extract 3NSGs, (H#C•Br) and (Br#C•H), from the ITS (**6**). The former 3NSG corresponds to a bromination and the latter to a dehydrobromination. This viewpoint would be reasonable in the light of essential features of chemical reactions. On the other hand, the conventional methods have some difficulties, since their initial classification to substrate class or to reagent class is rather fixed.

Although RC graphs vary in the brominating agent as shown in ITSs (**1**)—(**6**), a common 3NSG (H#C•Br) can be extracted from them. Thus, bromination (oxidative substitution by a bromine atom) is designated by a H#C•Br graphically or by the corresponding code H(1 − 1)C(0 + 1)Br. The graphical expression of the 3NSG is understandable intuitively, when the PS and PP operations take place on it (H–C + Br ⟶ H + C–Br).

Similarly, ITS (**7**) for chlorination of anthracene with copper(II) chloride[17] contains a 3NSG (H#C•Cl) as a subgraph. Iodination reactions[18] are characterized by the three-nodal subgraph (H#C•I) which, for instance, appears in ITS (**8**).

When Hal and HH represent respectively halogen atoms and hyperhydrogen atoms (hydrogen and hydrogen-like atoms), a more generic descriptor (HH#C•Hal) is obtained. The concept (HH#C•Hal) corresponds approximately to a natural language term 'halogenation (substitution by a halogen atom)' which is superior to H#C•F, H#C•Cl, H#C•Br, and H#C•I. As a result, the reaction hierarchy of halogenation reactions is obtained as shown in Figure 1. For the purpose of further discussion, the classification levels of biology (phylum, division, section, order, family, genus, and species) are applied to describe reaction hierarchies. Thus, we call a subgraph having terminal elements as a subgraph of the genus level. For example, 3NSG (H#C•Br) is a three-nodal subgraph of genus level. More generic descriptors such as HH#C•Hal are called 3NSGs of the family level.

Nitration of mesitylene[19] has an ITS (**9**), which involves 3NSG (H#C•N). Nitrosation of thymol[20] affords an ITS (**10**) having the same 3NSG (H#C•N). In order to distinguish the

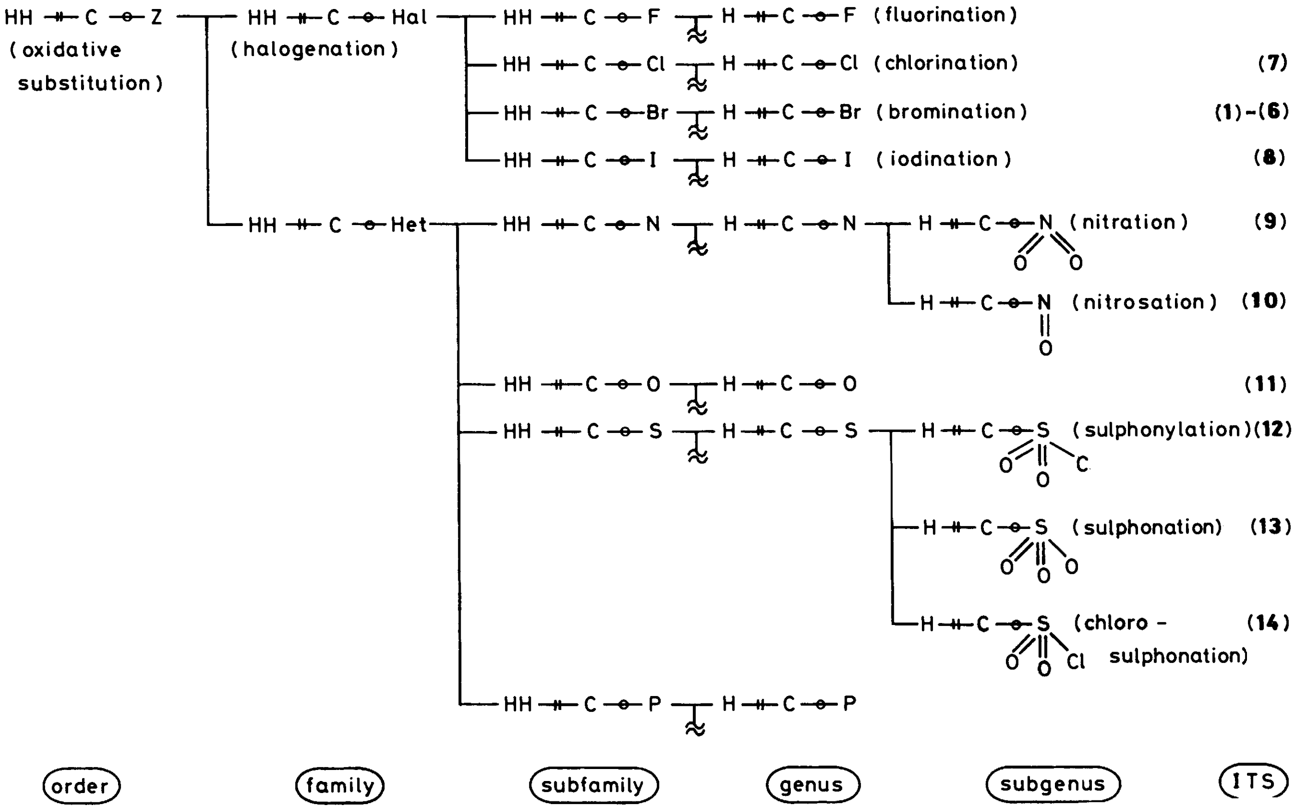


Figure 1. Reaction hierarchy of HH(1 − 1)C(0 + 1)Z order

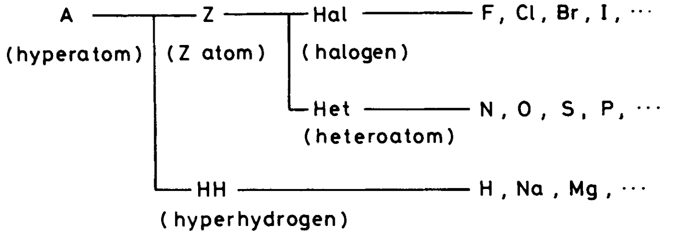


Figure 2. Hierarchy of elements

two reactions, we should consider extrastring atoms, *e.g.* H#C∘N(=*O*)$_2$ ($\mathbf{9}_{sg}$) for nitration and H#C∘N(=*O*) ($\mathbf{10}_{sg}$) for nitrosation. These subgraphs, the terminal atoms of which contain next neighbour extrastring atoms, are called descriptors of subgenus level. Air oxidation of cumene[21] is characterized by the 3NSG (H#C∘O) which is a subgraph of the corresponding ITS (**11**).

Friedel–Crafts acylation[22] leads to an ITS (**12**) which contains the three-nodal subgraph (H#C∘S). Sulphonation (**13**)[23] and chlorosulphonation (**14**)[24] give the same H#C∘S as a 3NSG. The three cases are distinguished by examination of extrastring atoms.

Let us consider a superior concept 'heteroatom (Het)' which embraces nitrogen, phosphorus, oxygen and sulphur atoms. We can designate the above reactions (**9**)—(**14**) in terms of HH#C∘Het. This descriptor of family level represents displacement of hydrogen (or a hydrogen-like atom) by a heteroatom (HH–C + Het → HH + C-Het). Even if the corresponding natural language term is absent, it is unnecessary to create a new word as long as we use the subgraphs of ITSs (or the corresponding codes). However, adequate new words corresponding to subgraphs of ITSs are sometimes convenient.

When we give the more generic name 'Z atom'* to a set containing halogen atoms and heteroatoms, we can obtain a generic term HH#C∘Z which is superior to HH#C∘Hal and HH#C∘Het. The descriptor of order level (HH#C∘Z) corresponds to the natural language term 'oxidative substitution'. The term 'oxidative' or 'oxidation' stems from the comparison of the terminal hyperatoms (HH or Z).

As we can see in the above discussion, we have obtained the reaction hierarchy concerned with HH#C∘Z order as shown in Figure 1. Figure 2 shows the hierarchy of atoms used in this paper. It should be emphasized that all the graphs collected in Figure 1 appear in ITSs as subgraphs. Thus, the characterization of reaction types (reaction-site changes) is accomplished by examination of the subgraphs of ITSs. A descriptor such as H#C∘N [or H(1 − 1)C(0 + 1)N] is called that of genus level. A superior one such as HH#C∘Het [or HH(1 − 1)C(0 + 1)Het] is designated as a descriptor of family level. An even more superior one HH#C∘Z [or HH(1 − 1)C(0 + 1)Z] is defined as a descriptor of order level.

Figure 3 shows the hierarchy of the three-nodal subgraphs (Z#C∘HH) which correspond to the reverse reactions of (HH#C∘Z) collected in Figure 1. For example, entry (**6**) provides a 3NSG of genus level (Br#C∘H) along with the reverse type. Formation of a Grignard reagent (**15**)[25] affords a 3NSG of genus level (Br#C∘Mg). When a magnesium atom is considered as a hyperhydrogen atom (HH), a 3NSG of family level (Hal#C∘HH) is obtained.

Reduction of benzilic acid with hydroiodic acid[26] leads to ITS (**16**), from which a 3NSG (O#C∘H) is abstracted. Reduction of a diazonium salt (**17**)[27] gives a heptagonal RC graph, which involves a 3NSG (N#C∘H). In the ITS approach, a diazonium function is represented in terms of the covalent form

---

* The definition of Z atom is the same as that of Hendrickson.[7]

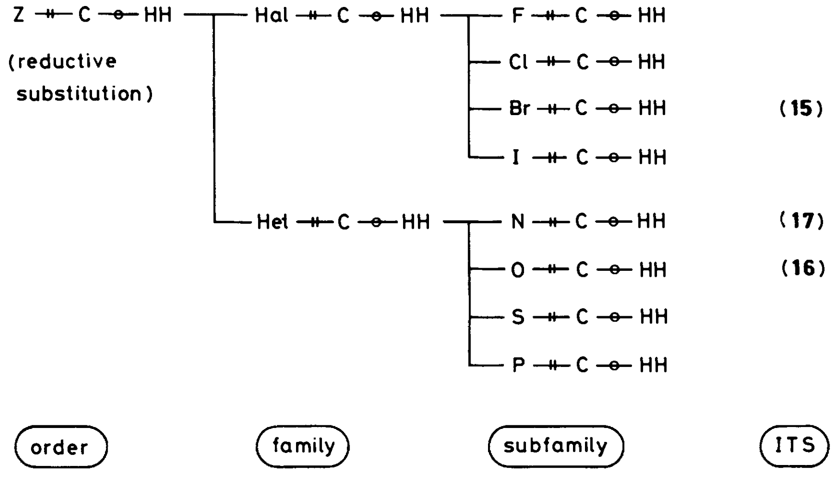


Figure 3. Reaction hierarchy of Z(1 − 1)C(0 + 1)HH order. Genus and lower levels are abbreviated

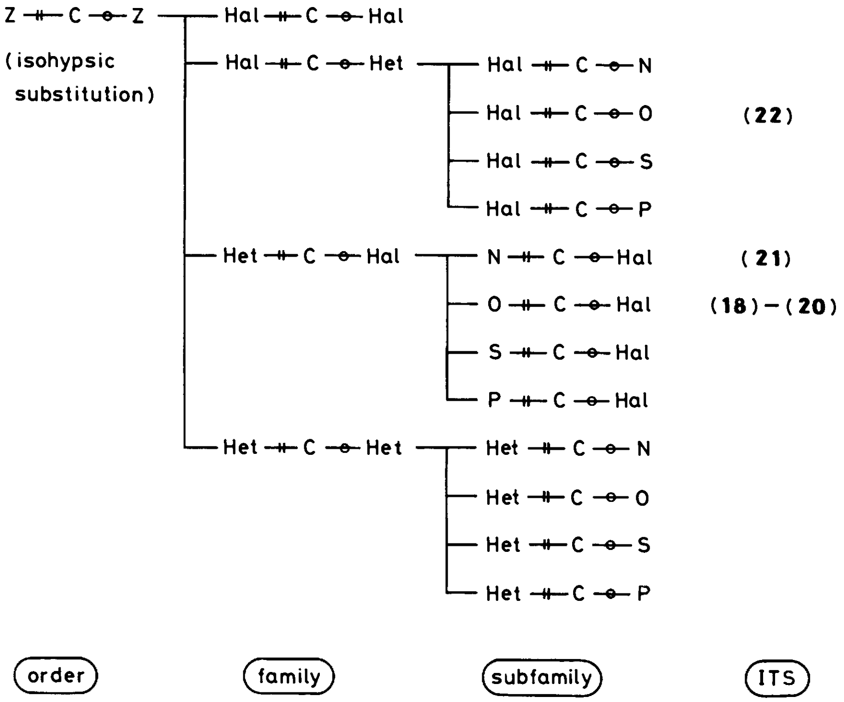


Figure 4. Reaction hierarchy of Z(1 − 1)C(0 + 1)Z order. Genus and lower levels are abbreviated

and hence the nitrogen atom has quinquevalency. This convention has been discussed elsewhere.[10c,h] By using the generic terms HH (hyperhydrogen atoms) for hydrogen and Het (heteroatoms) for O and N, we can obtain the corresponding 3NSG at the family level (Het#C∘HH). A 3NSG (Z#C∘HH) is a descriptor of order level which is superior to the 3NSGs, Hal#C∘HH and Het #C∘HH.

A reaction hierarchy of Z#C∘Z order is shown in Figure 4. The 3NSG (Z#C∘Z) is a descriptor of order level which denotes an isohypsic substitution.* The term ‘isohypsic’ can be derived by comparison of the terminal hyperatoms (Z or Z). Descriptors of family level are easily obtained by introducing subordinal hyperatoms (Hal and Het) instead of Z.

The reactions of Z(1 − 1)C(0 + 1) order consists of familiar organic reactions. For example, replacement of hydroxy by a bromine atom (**18**)[28] contains a 3NSG (O#C∘Br), which is subordinate to Het#C∘Hal. The ITS (**19**) affords the same 3NSG, although the brominating agent is different.[29] Bromotrimethylsilane is a reagent which cleaves an ethereal linkage.[30] This process is represented by ITS (**20**), which affords a 3NSG (O#C∘Br). It should be noted here that discrimination between (**19**) and (**20**) requires the additional description of extrastring atoms. The Sandmeyer reaction (**21**)[31] involves a 3NSG (N#C∘Br). Nucleophilic substitution of an aromatic compound (**22**)[32] gives this type of 3NSG (Cl#C∘O).

When an arbitrary atom is represented by A (Figure 5), the resulting section descriptor (A#C∘A is a more generic concept (substitution) than the above 3NSGs, HH#C∘Z, X#C∘HH, and Z#C∘Z, which correspond to oxidative, reductive, and iso-

---

* For the definition of the term ‘isohypsic’, see ref. 7.

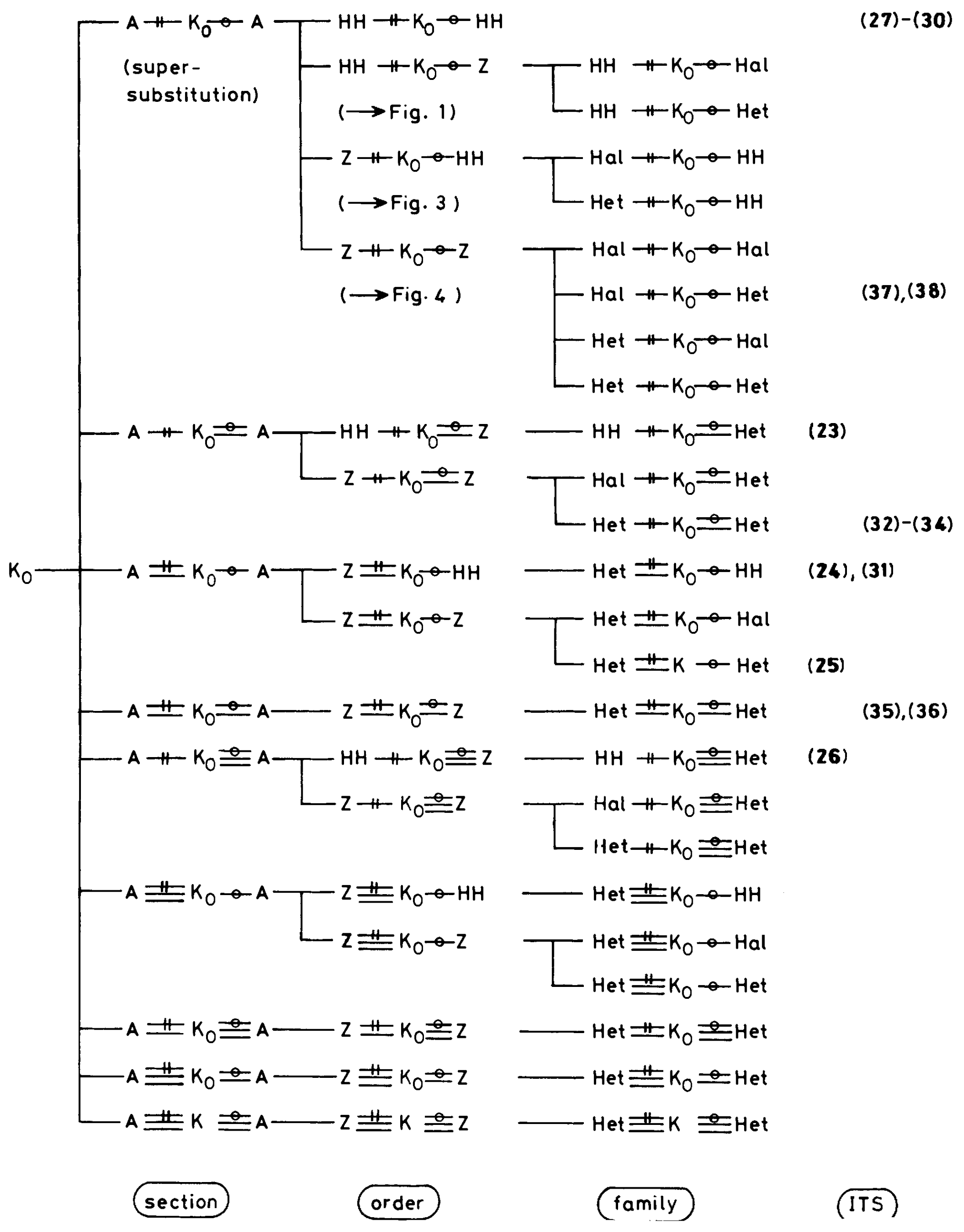


Figure 5. Terminal descriptors for odd-nodal subgraphs

hypsic (nucleophilic) substitution, respectively. The subgraphs such as A⧺C◦A are called descriptors of section level.

Oxidation of an alcohol[33] is represented by ITS (**23**). This ITS gives a 3NSG (H⧺C≗O) which is codified as H(1 − 1)-C(1 + 1)O. The corresponding superior concept is HH⧺C≗Het, which is, in turn, subordinate to HH⧺C≗Z.

The Cannizzaro reaction[34] is redox process which is represented by ITS (**24**). Two 3NSGs, O⧺C◦H [or O(1 − 1)-C(0 + 1)H] and H⧺C≗O [or H(1 − 1)C(1 + 1)O], are obtained for simultaneous reduction and oxidation of C=O. The first subgraph is of inferior level to Het⧺C◦HH, which is, in turn, a suborder of Z⧺C◦HH. The reductive character of this reaction-site change is perceived by comparison between Z and HH. The second 3NSG (H⧺C≗O) is generalized into HH⧺C≗Het (family), HH⧺C≗Z (order), and A⧺C≗A (section), successively. The oxidative character of the reaction is recognized at the order level (HH or Z). The term 'addition to C=O' is derived from the descriptor of section levels in terms of (1 − 1) and (1 + 1) bonds.

Geminal diol formation from ninhydrin (**25**)[35] is characterized by a 3NSG (O±C◦O), which is a subfamily of Het±C◦Het.

Dehydration of an aldoxime to a nitrile (**26**)[36] involves the 3NSG (H⧺C≡N). This corresponds to a more general expression (HH⧺C≡Het). It should be noted that the 3NSG (H⧺C≡N) represents a reductive elimination at the carbon centre, but the actual reaction is an isohypsic elimination to form a nitrile. This is because we consider only carbon reaction centres as three-nodal subgraphs.

Figure 5 summarizes a hierarchy of three-nodal subgraphs A(a − 1)K_0(a + 1)A, wherein K_0 is C. The term K_0 is called an odd-membered reaction kernel, and is discussed in the next section. It should be noted that each descriptor can be represented by a combination of a descriptor of section level and of a pair of terminal atoms. For example, HH#C•Het is replaced by a combination of A#C•A and HH/Het. In other words, a descriptor of section level represents a mode of bond switching during the reaction.

It is worth comparing here the ITS approach and Hendrickson's one.[7] His representation[7] is a net change at a carbon reaction centre which has a broader meaning. For example, Hendrickson's ZH corresponds to a set of HH#C•Z, HH#C≛Z, and HH#C≛Z in the ITS approach. Thus, the 3NSG notation based on ITS is more specific than Hendrickson's notation. When modification by single (–) or double par-bonds (=) is abbreviated, we can obtain a basic subgraph HH#C•Z which is strictly equivalent to ZH. In other words, Hendrickson's notation neglects modification by single or double par-bonds.

Reaction Hierarchy based on Odd-nodal Subgraphs which consist of Odd-membered Reaction Kernels and Terminal Atoms.— Addition of cyanide to an activated double bond[37] is represented by ITS (**27**). This reaction has one string and affords a hexagonal RC graph (**27**$_{rc}$), from which the corresponding five-nodal subgraph (5NSG), Na#C•C±C•H, is easily abstracted.

The 5NSG (**27**$_{sg}$) is represented by a combination of a C_3 kernel (K_0 = C•C±C) and a terminal atom relationship (Na#K_0•H). The latter terminal descriptor is considered to be A#K_0•A) in which the As are replaced by Na and H.

The terminal descriptors in 3NSG and in 5NSG can be discussed in common terms (Figure 5). Three-nodal subgraphs at the section level are represented in general by the formula (A#K_0•A), where K_0 is a C_1 kernel as discussed in the preceding section. The out-bond may be replaced by a bond having a complex bond number (*a b*) of *b* – 1 and the in-bond by a bond (*a b*) of *b* + 1. Similarly, five-nodal subgraphs have a general formula (A#K_0•A), where K_0 is a C_3 kernel. A terminal atom (A) is incident to the reaction kernel (K_0) by an out-bond and another terminal (A) by an in-bond. This relationship applies in odd-membered RK commonly and stems from the alternate character of a reaction string.[10a]

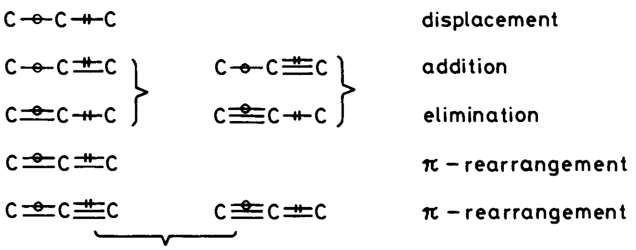


Figure 6. C_3 Reaction kernels. Each pair linked with a brace is a reaction pair

New natural language terms are introduced to describe the graphs collected in Figure 5. The graph A#K_0•A is called as 'supersubstitution' which is an extension of A#C•A (substitution). Similarly, HH#K_0•Z is an 'oxidative supersubstitution', which subsumes 'oxidative substitution' (H#C•Z). An 'isohypsic supersubstitution' corresponds to Z#K_0•Z, and so on.

A mode of terminal atoms and an RK can be discussed independently for characterization of reaction features. Thus, the C_3 reaction kernel (C#C•C) gives information about the carbon skeleton (C + C=C → C–C–C). The terminal descriptor (Na#K_0•H) of genus level or HH#K_0•HH of family level indicates the change of oxidation state (isohypsic).

Straight-chain C_3 reaction kernels are found in Figure 6. These RKs represent substitution reactions participated in by three carbons, additions of a carbon centre to a C=C double bond, eliminations to form a C=C double bond accompanied by a C–C bond cleavage, and π-rearrangements.

The following examples illustrate the above discussions. The ITS's (**28**)[38] and (**29**)[39] contain 5NSGs Mg#C•C±C•H and Li #C•C±C•H respectively, both of which can be considered as a subfamily of HH#C•C±C•HH. The generic 5NSG is perceived as a combination of a C_3 kernel (K_0 = C•C±C in Figure 6) and a terminal descriptor of order or family level (HH#K_0•HH in Figure 5). The C_3 kernel indicates an addition (C + C=C → C–C–C) and the latter terminal descriptor corresponds to 'isohypsic' (HH–K_0 + HH → HH + K_0–HH). Ketene formation (**30**)[40] is characterized by a 5NSG H#C≛C#C•H. In general, this is represented by C≛C#C (= K_0) and HH#K_0•HH. The former term K_0 indicated that entry (**30**) is an elimination (C–C–C → C=C + C) and the latter terminal descriptor corresponds to the 'isohypsic' term (HH–K_0 + HH → HH + K_0–HH).

Di-imide reduction accompanied by a transannular reaction (**31**)[41] has a 5NSG O±C•C±C•H, which is a combination of C•C±C (= K_0) and O±K_0•H. The former descriptor corresponds to an addition (C + C=C → C–C–C). The latter terminal descriptor is indicative of the term 'reductive' (Z = K_0 + HH → Z–K_0–HH).

Decarboxylative formation of a ketone (**32**)[42] involves a 5NSG O#C•C#C•O; the five-nodal subgraph can be divided into two descriptors, *i.e.* C•C#C (= K_0) and O#K_0•O.

Aldol-type condensations (**33**) and (**34**)[43,44] afford a common 5NSG O#C•C±C≛O. This subgraph indicates isohypsic (*i.e.*, Z#K_0≛Z) addition [*i.e.*, C•C±C (= K_0)]. An aldol condensation (**35**)[45] is characterized by the subgraph O±C•C±C≛O, which is a suborder of an ordinal descriptor Z±C•C±C≛Z. The related condensation (**36**)[46] leads to subgraph N±C•C± C≛O, which is subordinate to the descriptor Z±C•C±C≛Z. The ordinal subgraph is perceived as a combination of a C_3 kernel C•C±C (= K_0) and a terminal descriptor Z±K_0≛Z. This TD expresses another type of isohypsic reaction.

Arylation of an activated olefin with a diazonium salt (**37**)[47]

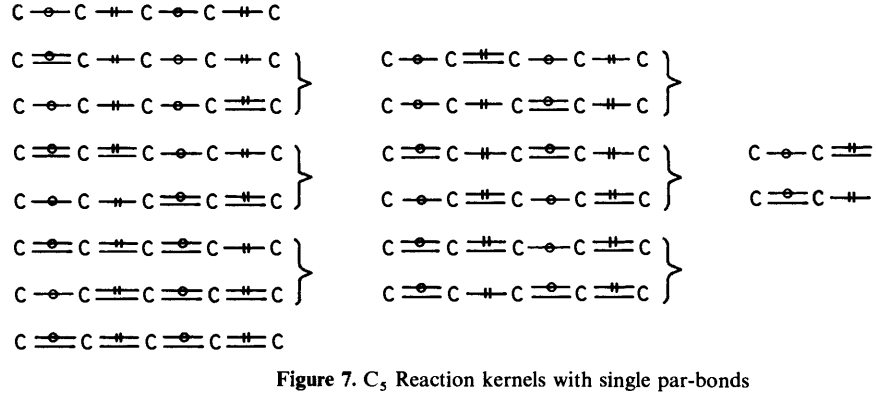


Figure 7. C_5 Reaction kernels with single par-bonds

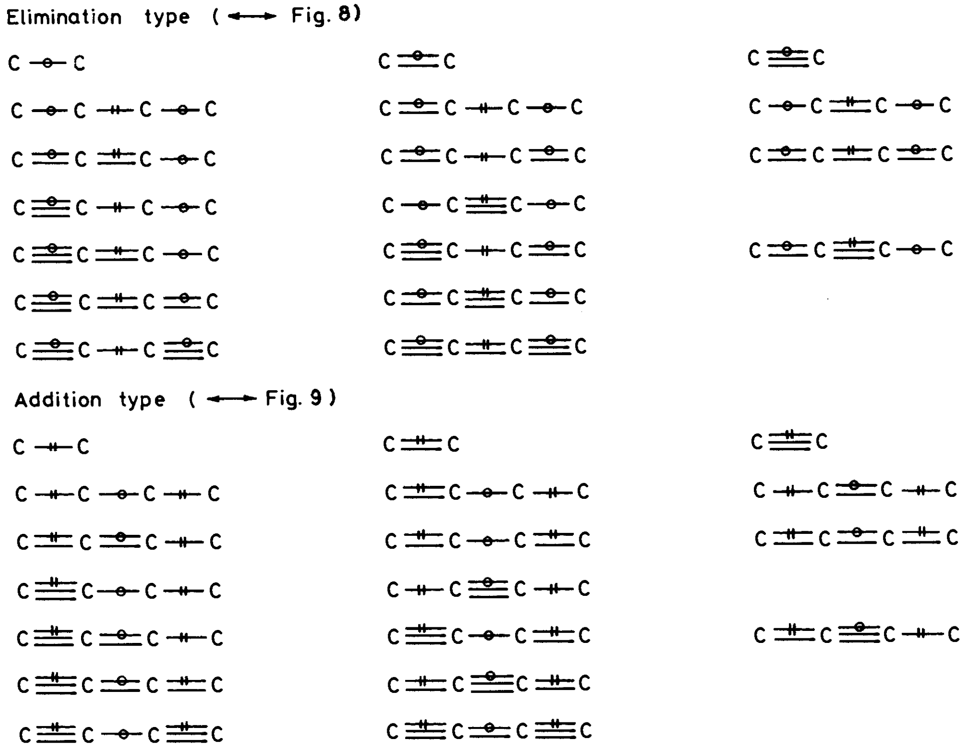


Figure 8. C_2 and C_4 reaction kernels

has N#C◆C±C◆Cl as a 5NSG. The reaction feature 'addition' is confirmed by the corresponding C_3 kernel (C◆C±C). The corresponding terminal descriptor, N#K_o◆Cl (in general, Het#K_o◆Hal or Z#K_o◆Z) provides information about a change of oxidation state during the reaction.

An example of a seven-nodal subgraph (7NSG) is derived from ITS (**38**)[48] which is a conjugated addition of an aryl group to butadiene. The 7NSG N#C◆C±C≗C±C◆Cl is divided into two descriptors, *i.e.* a C_5 reaction kernel C◆C±C≗C±C (= K_o) and a terminal descriptor N#K_o◆Cl (more generally Het#K_o◆-Hal or Z#K_o◆Z as collected in Figure 5). Enumeration of the C_5 reaction kernels can be effected easily by counting isomers of C◆C#C◆C#C with *m* double par-bonds and *n* single par-bonds. Figure 7 lists C_5 reaction kernels with single par-bonds. The terminal descriptors are selected from the set of descriptors collected in Figure 5.

Because of alternant characters of reaction strings,[10a] the classification scheme based on two terminal atoms (Figure 5) is effective for all odd-nodal subgraphs. The graphic expressions shown in Figure 5 contains more detailed information on reaction features than such natural language terms as 'oxidative', 'reductive', and 'isohypsic'. It should be emphasized that the latter terms can be derived from the graphs collected in Figure 5.

Reaction Hierarchy based on Even-nodal Subgraphs which consist of Even-membered Reaction Kernels and Terminal Atoms.—A C_2 reaction kernel is defined as two adjacent carbon atoms which are linked to one another by an imaginary bond of non-zero *b* and have no other carbon atoms linking to them by out- or in-bonds. For example, dechlorination of 2,3-dichloro-1,4-dioxane by magnesium[49] affords an ITS (**39**) which involves a pentagonal RC graph of level (1) (**39**$_{rc}$), |5L|Mg(0 + 1)-Cl(1 − 1)C(1 + 1)C(1 − 1)Cl(0 + 1)Mg. This RC graph also has one reaction string. The corresponding four-nodal subgraph (4NSG), Cl#C≗C#Cl or Cl(1 − 1)C(1 + 1)C(1 − 1)Cl, is a descriptor of genus level for an eliminative dechlorination. In this case, a C_2 reaction kernel is C≗C, or C(1 + 1)C, which, in turn, corresponds to a more general term 'formation of a double bond' (which implies 'elimination' in most cases). When we represent C_2 reaction kernel by K_e, the above 4NSG is converted into more generic expressions, Cl#K_e#Cl, Hal#K_e#Hal, and Z#K_e#Z, in a hierarchical order. These expressions are related to 'dechlorination', 'dehalogenation', and 'reductive elimination', respectively. The two terminal atoms are linked to K_e with two imaginary bonds of *b* −1.

Figure 8 collects two- or four-membered reaction kernels (RKs). An even-membered reaction kernel (K_e) is a broad descriptor to give information on a bond switching during a reaction. The kernel C±C corresponds to a natural language term 'disappearance of a double bond' (which implies 'addition' in most cases). This graph is a reverse graph of the kernel C≗C which indicates 'formation of a double bond' as described above. The raction kernels C◆C and C#C are descriptors of 'construction (C–C bond formation)' and 'cleavage', respectively. This correspondence is well explained, if the PS and PP operations take place onto the RKs. For instance, a four-membered reaction kernel (C±C◆C±C) is a unitary representation of (C=C + C=C → C–C–C–C).

Comparison of two atoms of terminal descriptor (A#K_e#A) describes the change of oxidation state during the reaction. Figures 9 and 10 collect hierarchies of terminal descriptors A(*a* − 1)K_e(*b* − 1)A and A(*a* + 1)K_e(*b* + 1)A, respectively. Even-nodal subgraphs are expressed by a combination of a descriptor of Figure 8 and that of Figure 9 (or 10).

Let us examine other examples of 4NSGs extracted from several one-string reactions. Aldol condensation (**40**)[50] is characterized by a 4NSG H#C◆C±O (genus level), or more generally HH#C◆C±Het (family level). The corresponding ordinal descriptor is HH#C◆C±Z, which is a combination of a C_2 kernel (C◆C) and a terminal descriptor (HH#K_e±Z). The graph (C◆C) corresponds to the term 'construction (C–C bond

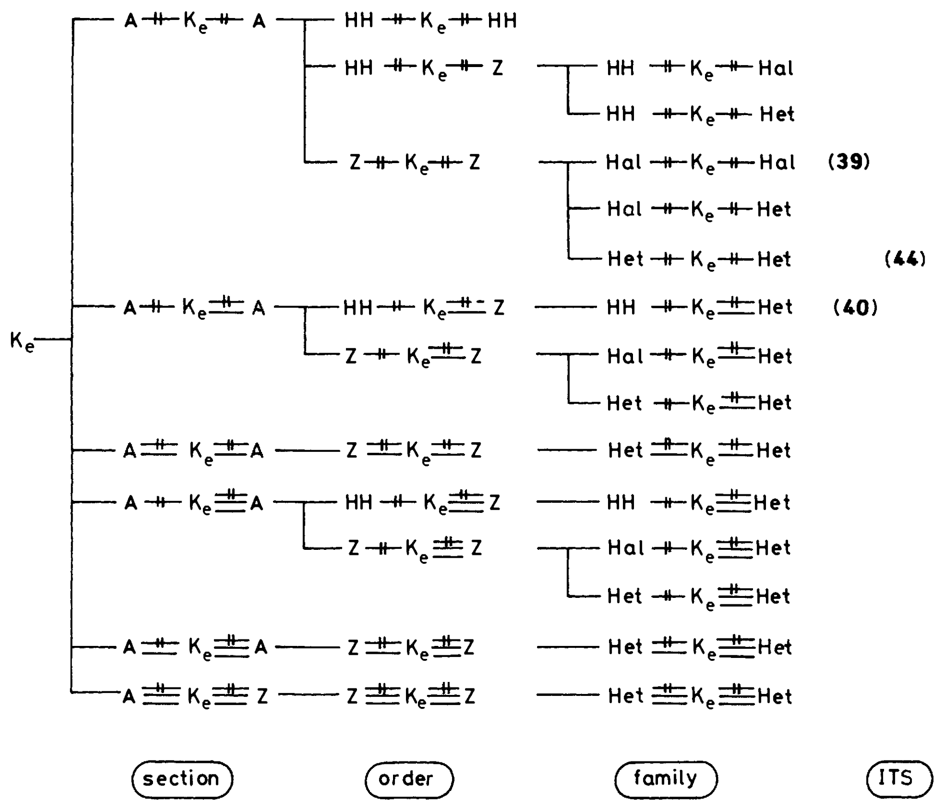


Figure 9. Terminal descriptors for even-nodal subgraphs

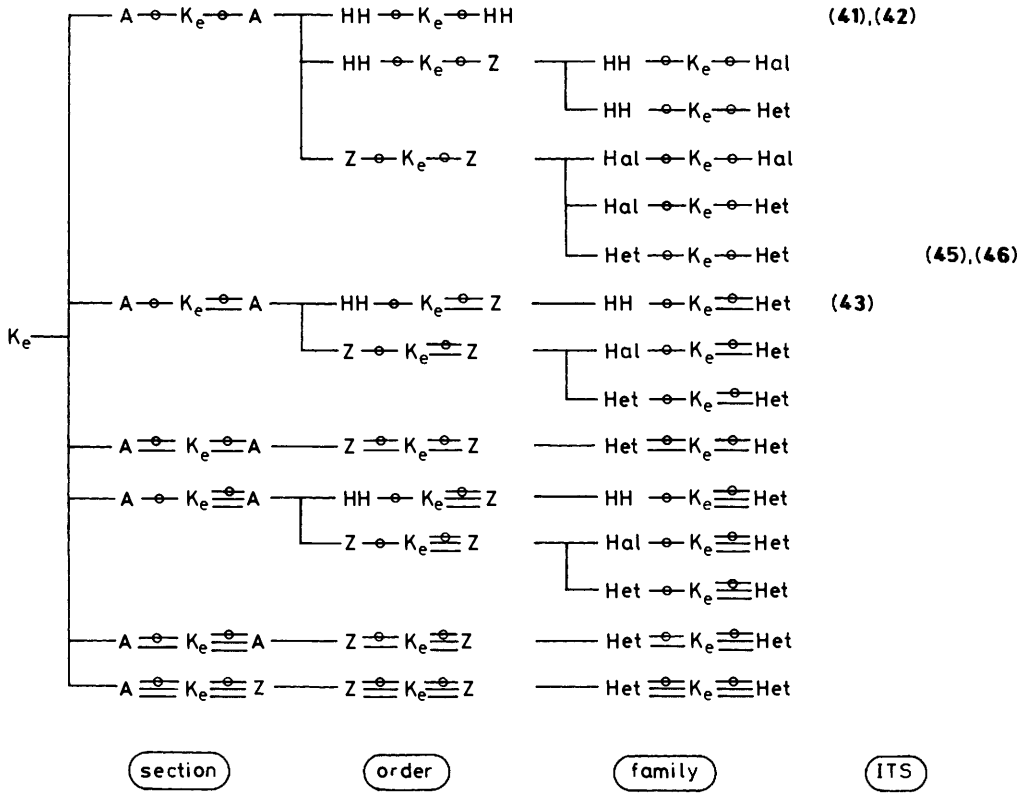


Figure 10. Terminal descriptors for even-nodal subgraphs

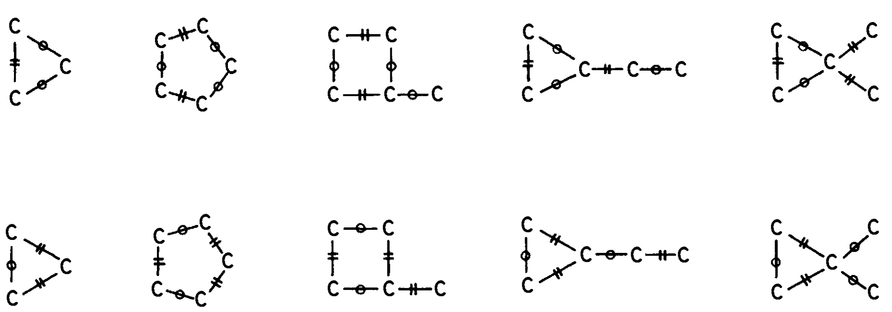


Figure 11. Odd-membered basic reaction kernels having a cyclic structure

formation)' and HH⧺K_e±Z is in accordance with the term 'isohyptic'. This example should be compared with entry (**35**).

Catalytic hydrogenation of chalcone[51] affords ITS (**41**). The corresponding 4NSG H⊖C±C⊖H is converted into K_e (C±C: 'addition') and HH⊖K_e⊖HH ('reduction'). The latter two descriptors characterize this reaction well.

Hydroboration (**42**)[52] affords 4NSG H⊖C±C⊖B (genus level) or HH⊖C±C⊖HH (family or order level). The latter descriptor is represented by the C_2 kernel (C±C) and the terminal descriptor (HH⊖K_e⊖HH).

In ITS (**43**) of decarboxylation,[53] the corresponding 4NSG H⊖C⧺C±̊O is perceived as a combination of Ke (C⧺C: 'cleavage') and H⊖K_e±̊O ('isohypsic').

A conjugated elimination[54] represented by ITS (**44**) is characterized by the corresponding six-nodal subgraph (6NSG) S⧺C±̊C±C±̊C⧺S. The C_4 kernel K_e (C±̊C±C±̊C) is a descriptor of conjugated elimination (C–C=C–C → C=C–C=C). The remaining terminal atoms (S⧺K_e⧺S) or the corresponding superior concepts, Het⧺K_e⧺Het and Z⧺K_e⧺Z, indicate that this reaction is a reductive process.

A Diels–Alder-type addition (**45**)[55] provides a 6NSG N⊖C±C ±̊C±C⊖N (genus level) or more generally Het⊖C±C±̊C ±C⊖Het (family level) or Z⊖C±C±̊C±C⊖Z (order level). This last subgraph is a reverse of the counterpart shown in ITS (**44**). Thus the C_4 kernel (C±C±̊C±C) corresponds to a conjugated addition. And the terminal descriptor (Z⊖K_e⊖Z) is for an oxidative process.

Other C_4 kernels are collected in Figure 8. They are combined with the terminal descriptors (Figures 9 and 10) to describe various levels of reaction features.

Addition of SO_2 to hexatriene (**46**)[56] provides an eight-nodal subgraph (8NSG) S⊖C±C±̊C±C±̊C±C⊖S. The C_6 kernel (C±C±̊C±C±̊C±C) indicates a conjugated addition to hexatriene and the terminal atoms (S⊖K_e⊖S, Het⊖K_e⊖Het, or Z⊖K_e⊖Z) shows that this reaction involves an oxidative process.

The terminal descriptors collected in Figures 9 and 10 are effective for all even-nodal subgraphs of one string. This advantage stems from the fact that a reaction string contains alternate out- and in-bonds. It should be emphasized that all procedures described above are based upon a substructure search of ITSs.

A natural language term often has several meanings.* For example, 'bromination' means (1) displacement of hydrogen by a bromine atom, (2) addition of bromination to a double bond or (3) to a triple bond, and sometimes (4) displacement of a hydroxy group by a bromine atom. On the other hand, the ITS approach distinguishes them as (1) H⧺C⊖Br, (2) Br⊖C±C⊖Br, (3) Br⊖C±̊C⊖Br, and (4) O⧺C⊖Br, respectively. And then, reactions (1) and (4) are easily classified as 'substitutions' by examination of the in- and out-bonds of 3NSGs. Reactions (2) and (3) are perceived as 'addition' by the subgraphs (reaction kernels) such as C±C and C±̊C. The oxidative character of the reactions (1)—(3) are indicated by comparison of the respective terminal atoms (HH or Z). An isohypsic character of reaction (4) is decided by comparing O and Br, because both are Z atoms.

Subgraphs based on Odd-membered Reaction Kernels having a Cyclic Structure.—Figure 11 shows odd-membered RKs having a cyclic structure. Each of them can be modified by *m* of double par-bonds (=) and *n* of single par-bonds (–) on its edges. The enumeration of such isomers of basic reaction kernels is essentially equivalent to that of reaction graphs reported previously.[10c]

Carbene-type additions to form cyclopropane rings are examples of this class.[57,58] The corresponding ITSs (**47**) and (**48**) contain a C_3 kernel of this type ($\mathbf{47}_{rk}$), which is characterized by substitution of a single par-bond.

Even-membered Cyclic Reaction Kernels without Terminal Atoms.—Dimerization of an olefin (**49**)[59] involves a tetragonal RC graph which is identical to a cyclic C_4 reaction kernel ($\mathbf{49}_{rk}$). This is substituted by two single par-bonds (*n* 2) and no double par-bonds (*m* 0). Ring opening of cyclobutene (**50**)[60] provides another cyclic C_4 kernel ($\mathbf{50}_{rk}$) which is modified by three single par-bonds (*n* 3) and no double par-bonds (*m* 0). Enumeration of the cyclic C_4 reaction kernels is equivalent to that of tetragonal reaction graphs.[10b]

The Diels–Alder addition (**51**)[61] affords the corresponding RC graph ($\mathbf{51}_{rc}$) which is identical to a C_6 reaction kernel ($\mathbf{51}_{rk}$). A retro-Diels–Alder reaction (**52**)[62] gives an RC graph ($\mathbf{52}_{rc}$) and reaction kernel ($\mathbf{52}_{rk}$) which are the same. These reaction kernels ($\mathbf{51}_{rk}$) and ($\mathbf{52}_{rk}$) are alternately coloured hexagons with four single par-bonds on their edges. The Cope rearrangement[63] (**53**) gives a C_6 reaction kernel ($\mathbf{53}_{rk}$) which is an isomer of ($\mathbf{51}_{rk}$) and ($\mathbf{52}_{rk}$). The enumeration of hexagonal (C_6) reaction kernels can be done by straightforward use of Polya's theorem. This problem is essentially equivalent to enumeration of hexagonal reaction graphs.[10b]

The C_1 Reaction Kernels of Two-string Reactions.—Hydrolysis of a *gem*-dichloride[64] (**54**) affords the corresponding RC graph ($\mathbf{54}_{rc}$) which consists of two reaction strings ($1 - 3 + 5 - 2 + 1$ and $1 - 4 + 6 - 2 + 1$). In order to characterize this reaction, we select an effective subgraph which has the maximum number of adjacent carbon reaction centres and has incident terminal non-carbon reaction centres. The effective subgraph of a multistring reaction is an extension of the three- or more-nodal subgraph of a one-string reaction. Thus, the subgraph of genus level is represented by ($\mathbf{54}_{ge}$). When we substitute the terminal atoms by the corresponding generic atoms, we obtain ($\mathbf{54}_{fa}$) (family level), ($\mathbf{54}_{or}$) (order level), and ($\mathbf{54}_{se}$) (section level) in accord with a reaction hierarchy. These generic subgraphs are based on a C_1 reaction kernel of two strings.

---

* *Note added in proof*: The IUPAC nomenclature for straightforward transformations settled this problem to a great extent (see ref. 69).

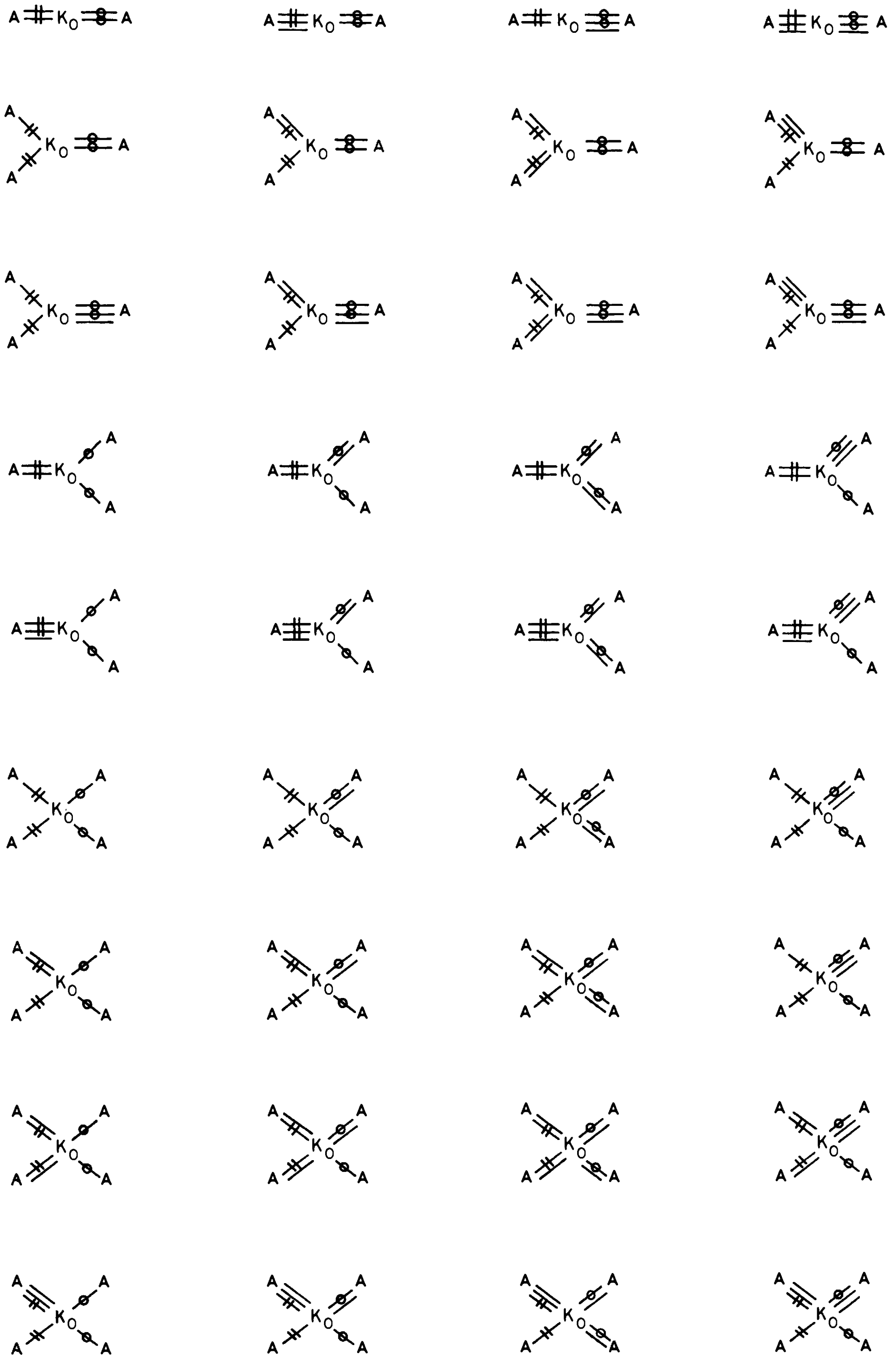


Figure 12. Representative terminal descriptors of section level for two-string reactions

In the light of the stringity (number of reaction strings), the generic subgraph ($\mathbf{54}_{ge}$) can be considered to contain two strings [Cl(3)⫲C(1)⦵O(2) and Cl(4)⫲C(1)⦵O(2)] fused at the CO bond. Thus the graph ($\mathbf{54}_{ge}$) is represented by a combination of a section descriptor ($\mathbf{54}_{se}$) and two pairs of atoms (ClO;ClO). In a similar way, the graph ($\mathbf{54}_{fa}$) is regarded as a

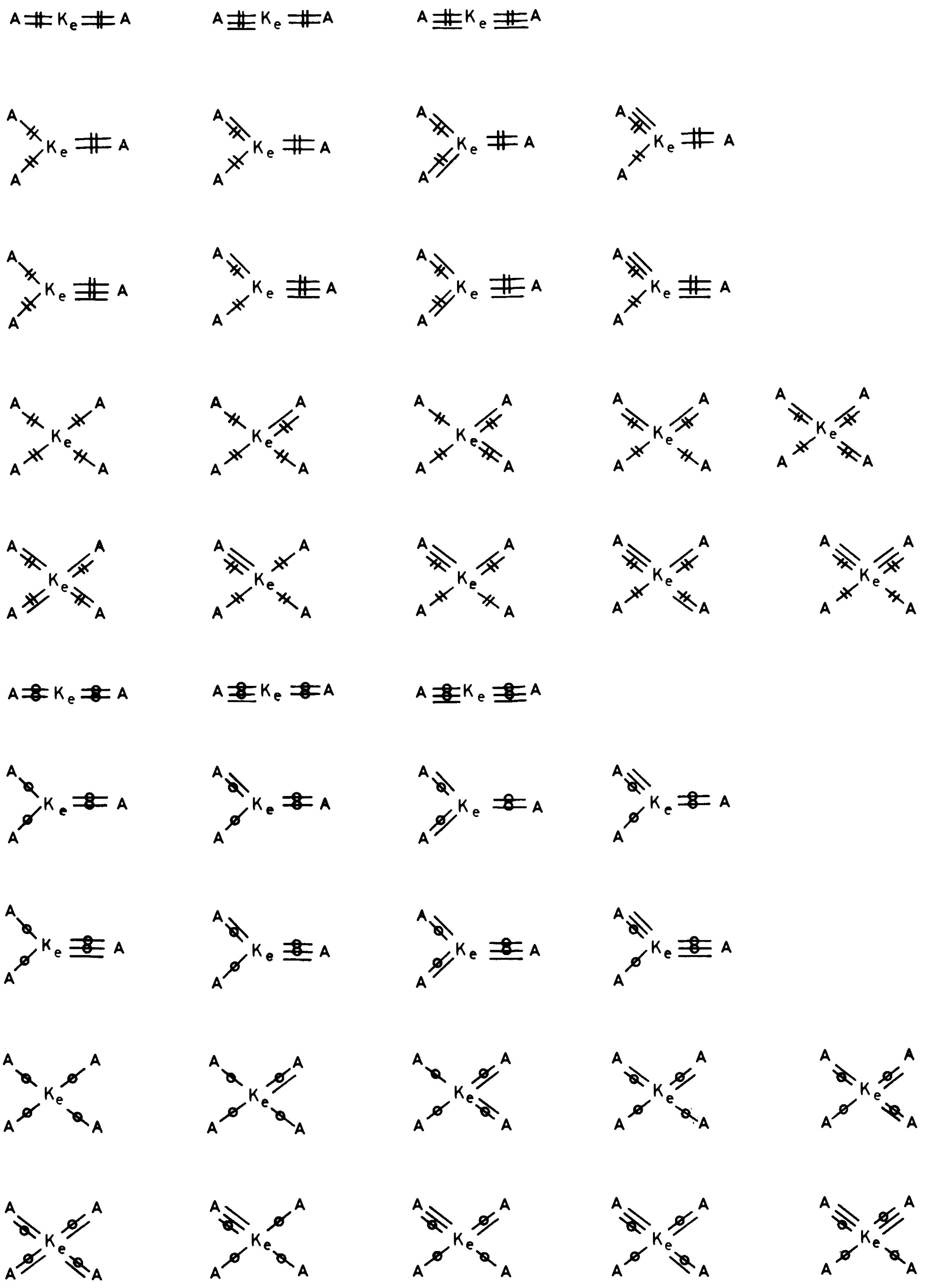


Figure 13. Representative terminal descriptors of section level for two-string reactions

combination of (**54**$_{se}$) and two pairs of extended atoms (HalHet;HalHet). A graph of higher level (**54**$_{or}$) is characterized by (**54**$_{se}$) and ZZ;ZZ.

Figure 12 shows various descriptors of section level, wherein K_0 may be C_1 or more odd reaction kernels. Atom A represents any atom. That is to say, a section descriptor clarifies numbers and types of imaginary bonds which are incident to the reaction kernel. This is applicable to the case that K_0 contains three or more (odd) carbons. But such reactions are rare to our knowledge.

Formation of a *gem*-dibromo derivative[65] affords ITS (**55**). Entry (**56**)[66] is hydrolysis of a benzoxazole. Formation of

level	clue	example	natural language term
phylum	stringity	one - string	
class	no. of C atoms in a reaction kernel	C_2	
division	reaction kernel	$C\overset{\circ}{=}C$ $(=K_e)$	formation of a double bond
section	terminal descriptor	$A \# K_e \# A$	elimination
order	terminal descriptor	$Z \# K_e \# Z$	reduction
family	terminal descriptor	$Hal \# K_e \# Hal$	dehalogenation
genus	terminal descriptor	$Cl \# K_e \# Cl$	dechlorination
species	ITS	ITS (**39**)	

Figure 14. Systematic classification of organic reactions by subgraphs of ITS

benzylidene-aniline[67] is represented by ITS (**57**). These examples are also two-string reactions. Their effective subgraphs of various levels are shown above and can be considered as derivatives of the section descriptors collected in Figure 11.

The C_2 *Reaction Kernels of Two-string Reactions.*—Condensation to form an α,β-unsaturated ketone (Claisen–Schmidt reaction)[68] represented by ITS (**58**) contains two reaction strings. This condensation is characterized by the effective subgraph of genus level ($\mathbf{58}_{ge}$) which is subdivided into $C\underset{\circ}{\overset{\circ}{-}}C$ and a descriptor of Figure 13, wherein atoms (A) are selected from a set of all atoms of the Periodical Table. Descriptors of family level ($\mathbf{58}_{fa}$) and of order level ($\mathbf{58}_{or}$) corresponds to those of Figure 13, where atoms (A) are selected from a set of (HH, Hal, and Het) or (HH and Z), respectively. The general reaction feature, 'condensation', is related to the C_2 kernel ($C\underset{\circ}{\overset{\circ}{-}}C$). The descriptor ($\mathbf{58}_{or}$) indicate the fact that the reaction is isohypsic. This perception is based on the comparison of the two pairs of terminal atoms (ZHH;ZHH) along the two respective reaction strings. The Knovenagel condensation and aldol-condensation–dehydration afford the same descriptor of genus level which reveals the close relationship among these reactions.

In general, C_2 reaction kernels of this type are $C\underset{\circ}{\overset{\circ}{-}}C$, $C\#C$, $C\underset{\circ}{\overset{\circ}{=}}C$, and $C\overset{\pm}{-}C$. They are combined with a section descriptor collected in Figure 13 for the purpose of classification of organic reactions.

Conclusions.—Figure 14 summarizes classification levels of organic reactions and clues for recognizing these levels. The stringity of a given reaction is a clue of a phylum level. For example, the reaction of ITS (**39**) has one reaction string. Thus, this is a one-string phylum. The number of adjacent carbon reaction centres (reaction kernel: RK) is a clue to the class level. Entry (**52**) belongs to a C_2 class. The RK (*e.g.*, $C\overset{\circ}{=}C$) shows that the reaction belongs to $C\overset{\circ}{=}C$ division ('elimination'). Examples of reaction kernels are shown in Figures 6—8. A section level is characterized by a descriptor ($A\#C\overset{\circ}{=}C\#A$). Because the type of C_2 kernel has been already indicated in the upper level, the section level can be characterized only by a terminal descriptor ($A\#K_e\#A$). The terminal descriptors are collected in Figures 5, 9, 10, 12, and 13. An order level is perceived by comparison of two terminal atoms (HH and Z). A family level is recognized by the examination of the terminal atoms (HH, Hal, and Het). Thus, ITS (**39**) belongs to $Hal\#K_e\#Hal$ family, which is '(super)dehalogenation. A descriptor of genus level is obtained by comparison of the terminal atoms of the lowest category of Figure 2. ITS (**39**) gives $Cl\#K_e\overset{\pm}{-}Cl$, which corresponds to the natural language term '(super)dechlorination'. In summary, all clues for characterization of organic reactions are obtained by examination of effective subgraphs of ITSs.

Appendix

Glossary

Imaginary Transition Structure (*ITS*).—A three-coloured structural formula in which all nodes appearing both in the starting stage of a given reaction and the product stage are connected with out-, in-, and par-bonds according to the change of bonds during the reaction.

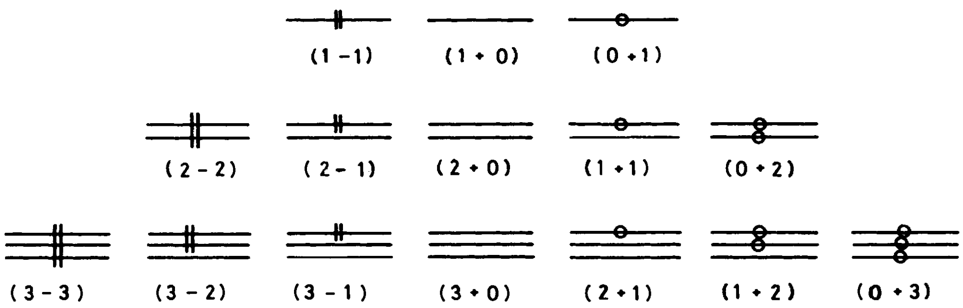
Out-bond.—A bond appearing only in the starting stage of a reaction.

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

Par-bond.—A bond which is invariant during a reaction.

Bond Colour.—A category of bonds appearing in ITS. There are three bond colours, *i.e.*, out, in, and par.

ITS Bond or Imaginary Bond.—A bond appearing in ITSs, which consists of out-, in-, and/or par-bonds. The types of ITS bonds are as follows:



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.

Connection Table of ITS.—A connection table, in which the connectivity is represented by complex bond numbers in accord with ITS.

Projection to the Starting Stage (PS).—Deletion of in-bonds from an ITS. This operation produces the corresponding starting stage.

Projection to the Product Stage (PP).—Deletion of out-bonds from an ITS. This operation produces the corresponding product stage.

Reaction Centre.—A node which is incident to in- and/or out-bonds.

Graph of Reaction Centres (RC Graph).—A graph which consists of, at least, reaction centres. An RC graph of level (**1**) is defined as an RC graph containing only reaction centres. An RC graph of level (**2**) is the one with reaction centres and the next neighbour nodes. RC graphs are the subgraphs of ITS and the descriptors of reaction types.

Reaction Graph (RC Graph of Level 0).—A graph in which every node is regarded as an abstract ball 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- 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 corresesonding reverse reaction.

Reaction Pair.—A pair of reaction graphs (or an RC graph of a given level) which is obtained by transformation to a 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 a reverse reaction.

Reaction String.—A structure which has alternate in- and out-bonds and can be modified by par-bonds. A reaction graph (or an R3 graph of a given level or an ITS) contains one or more reaction strings.

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

One-string Reaction.—A reaction which contains a single reaction string in the corresponding ITS (or RC graphs of various level). In a similar way, two-, three-,. . ., or multi-string reaction is defined.

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

Bridge of Ring Opening of Order p (BO_p).—An ITS ring in which p of 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.—An ITS ring in which p of ITS bonds have (ab) 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.—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.

References

1 Presented in part at the 192nd National Meeting of the American Chemical Society, Anaheim, 1986. See also J. Haggin, *Chem. Eng. News*, 1986, **64**, 75.
2 S. Fujita, *Yuki Gosei-Kagaku Kyokai-Shi*, 1986, **44**, 354, and references cited therein.
3 I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum, and W. Schubert, *Angew. Chem., Int. Ed. Engl.*, 1979, **18**, 111.
4 (*a*) V. Kvasnicka, M. Kratschvil, and S. Koca, *Collect. Czech. Chem. Commun.*, 1983, **48**, 2284; (*b*) V. Kvasnicka, *ibid.*, 1984, **49**, 1090.
5 D. C. Roberts, *J. Org. Chem.*, 1978, **43**, 1473.
6 J. F. Arens, *Recl. Trav. Chim. Pays-Bas*, 1979, **98**, 155.
7 J. B. Hendrickson, (*a*) *J. Chem. Inf. Comput. Sci.*, 1979, **19**, 129; (*b*) *J. Am. Chem. Soc.*, 1971, **93**, 6847.
8 C. S. Wilcox and R. A. Levinson, *ACS Symposium Series*, 1986, **306**, 209.
9 G. Vladutz, in 'Modern Approaches to Chemical Reaction Searching,' ed. P. Villet, Gower, 1986, p. 202.
10 S. Fujita, *J. Chem. Inf. Comput. Sci.*, 1986, **26**, (*a*) 205; (*b*) 212; (*c*) 224; (*d*) 231; (*e*) 238; 1987, **27**, (*f*) 99; (*g*) 104; (*h*) 111; (*i*) 115; (*j*) 120.
11 H. T. Clarke and M. R. Brethren, *Org. Synth.*, 1932, Coll. Vol. 1, 121.
12 G. H. Coleman and G. E. Honewell, *Org. Synth.*, 1943, Coll. Vol. 2, p. 243.
13 W. D. Langley, *Org. Synth.*, 1932, Coll. Vol. 1, 127.
14 A. R. Leed, S. Boettger, and B. Banem, *J. Org. Chem.*, 1980, **45**, 1098.
15 N. Inukai, H. Iwamoto, T. Tamura, I. Yanagisawa, Y. Ishii, and M. Murakami, *Chem. Parm. Bull. Jpn.*, 1976, **24**, 820.
16 M. Sekiya, K. Ito, and K. Suzuki, *Tetrahedron*, 1975, **31**, 231.
17 D. C. Nonhebel, *Org. Synth.*, 1973, Coll. Vol. 5, 206.
18 H. Y. Lew and C. R. Noller, *Org. Synth.*, 1963, Coll. Vol. 4, 545.
19 G. Powell and F. R. Johnson, *Org. Synth.*, 1943, Coll. Vol. 2, 449.
20 E. Kremers, N. Wakeman, and R. M. Hixon, *Org. Synth.*, 1932, Coll. Vol. 1, 511.

21 C. E. Frank, *Chem. Rev.*, 1950, **46**, 155.
22 G. Holt and B. Pagdin, *J. Chem. Soc.*, 1960, 2508.
23 C. M. Suter, *J. Am. Chem. Soc.*, 1931, **53**, 1114.
24 M. S. Morgan and L. H. Cretcher, *J. Am. Chem. Soc.*, 1948, **70**, 375.
25 R. Lespieer and M. Bouggnel, *Org. Synth.*, 1932, Coll. Vol. 1, 186.
26 C. S Marvel, F. D. Hayer, and E. C. Candle, *Org. Synth.*, 1932, Coll. Vol. 1, 224.
27 L. A. Bigelow, J. R. Johnson, and L. T. Sandborn, *Org. Synth.*, 1932, Coll. Vol. 1, 133.
28 E. E. Reid, J. R. Ruhoff, and R. E. Burnett, *Org. Synth.*, 1943, Coll. Vol. 2, 246.
29 C. R. Noller and R. Dinsmore, *Org. Synth.*, 1943, Coll. Vol. 2, 358.
30 H. R. Kricheldorf, G. Morber, and W. Regel, *Synthesis*, 1981, 383.
31 L. A. Bigelow, *Org. Synth.*, 1932, Coll. Vol. 1, 135.
32 R. Q. Brewster and T. Groening, *Org. Synth.*, 1943, Coll. Vol. 2, 445.
33 J. Wicka, A. Zarecki, and M. Kocor, *Tetrahedron Lett.*, 1973, 3635.
34 O. Kamm and W. F. Kamm, *Org. Synth.*, 1932, Coll. Vol. 1, 105.
35 L. F. Fieser, 'Experiments in Organic Chemistry,' Heath, Boston, 1955, 3rd edn., p. 129.
36 J. H. Hunt, *Chem. Ind.* (*London*), 1961, 1873.
37 A. Lapworth and W. Baker, *Org. Synth.*, 1932, Coll. Vol. 1, 451.
38 J. Munch-Peterson, *Org. Synth.*, 1973, Coll. Vol. 5, 762.
39 D. Seebach, *Synthesis*, 1969, 17.
40 C. D. Hurd, *Org. Synth.*, 1932, Coll. Vol. 1, 330.
41 K. Kimoto and M. Kawanisi, *Chem. Ind.* (*London*), 1971, 1174.
42 J. F. Thorpe and A. R. Kon, *Org. Synth.*, 1932, Coll. Vol. 1, 193.
43 B. M. Trost and E. Keinan, *Tetrahedron Lett.*, 1980, **21**, 2595.
44 S. Murata, M. Suzuki, and R. Noyori, *J. Am. Chem. Soc.*, 1980, **102**, 3248.
45 R. Noyori, K. Yokoyama, J. Sakata, I. Kuwajima, E. Nakamura, and M. Shimizu, *J. Am. Chem. Soc.*, 1977, **99**, 1265.
46 S. Danishefsky, M. Prisbulla, and B. Lipisko, *Tetrahedron Lett.*, 1980, **21**, 805.
47 E. Feltenstein, *Ann. Chim.* (*Paris*), 1957, 587.
48 G. A. Ropp and E. C. Coyner, *Org. Synth.*, 1963, Coll. Vol. 4, 727.
49 R. K. Summerkell and R. R. Umhoefer, *J. Am. Chem. Soc.*, 1939, **61**, 3019.
50 T. Mukaiyama and T. Inoue, *Chem. Lett.*, 1976, 559.
51 R. Adams, J. W. Kern, and R. L. Shriner, *Org. Synth.*, 1932, Coll. Vol. 1, 101.
52 H. C. Brown and B. Negishi, *J. Am. Chem. Soc.*, 1971, **93**, 3777.
53 H. T. Clarke and W. W. Hartman, *Org. Synth.*, 1932, Coll. Vol. 1, 549.
54 W. J. Bailey and E. W. Cummins, *J. Am. Chem. Soc.*, 1954, **76**, 1940.
55 P. G. Gassman and K. T. Mansfield, *Org. Synth.*, 1973, Coll. Vol. 5, 96.
56 W. L. Mock, *J. Am. Chem. Soc.*, 1967, **89**, 1281.
57 U. Schoellkopf and M. Reetz, *Tetrahedron Lett.*, 1969, 1541.
58 H. E. Simmons and R. D. Smith, *J. Am. Chem. Soc.*, 1958, **80**, 5323.
59 J. R. Lacher, G. Tonpkin, and J. D. Park, *J. Am. Chem. Soc.*, 1952, **74**, 1693.
60 J. L. Anderson, P. E. Putnam, and W. Shenkey, *J. Am. Chem. Soc.*, 1961, **83**, 382.
61 M. C. Kloetzel, *Org. React.*, 1948, **4**, 1.
62 R. B. Moffett, *Org. Synth.*, 1963, Coll. Vol. 4, 234.
63 S. J. Rhoads and N. R. Raulins, *Org. React.*, 1975, **22**, 1.
64 C. S. Marvel and W. M. Sperry, *Org. Synth.*, 1932, Coll. Vol. 1, 95.
65 G. H. Coleman and G. E. Honeywell, *Org. Synth.*, 1943, Coll. Vol. 2, 89.
66 (*a*) S. Fujita, K. Koyama, Y. Inagaki, and K. Waki, U.S. P. 4,055,428/1982; (*b*) S. Fujita, *Yuki Gosei-Kagaku Kyokai-Shi*, 1982, **40**, 176.
67 L. S. Bigelow and J. Eeatough, *Org. Synth.*, 1932, Coll. Vol. 1, 80.
68 E. P. Kohler and H. M. Chadwell, *Org. Synth.*, 1943, Coll. Vol. 1, 78.
69 J. F. Bunnett, *Pure Appl. Chem.*, 1981, **53**, 305.

Received 27*th January* 1987; *Paper* 7/136