Modeling Aromatic Nitration Reactions Using Graph-Theoretic Transforms

YI ZOU,† MARK A. JOHNSON,*,‡ and CHUN-CHE TSAI*,†

Department of Chemistry, Kent State University, Kent, Ohio 44242, and Computational Chemistry, Upjohn Laboratories, Kalamazoo, Michigan 49001

Received July 30, 1990

An algorithmic method of constructing graph-theoretical models of aromatic nitration reactions is presented. The model achieves simplicity and generality by replacing the traditional letter designations of the atoms and bonds in the chemical graph with multiple labels that differentiate atoms by their group, period, and charge and differentiate bonds by their σ and π bonding constituents. The method represents an application of a general formalism for modeling chemical transformation pathways. The relationships between the chemical knowledge and the mathematical model of aromatic nitration reactions are discussed.

INTRODUCTION

The relationship between structure and reactivity for electrophilic aromatic substitution reactions has been studied since about 1870.[1] The classification of substituents as activating or ortho-para-directing and deactivating or meta-directing has been known since those early studies. This classification is widely understood in terms of the qualitative theory by Ingold[2] and Remick[3] for predicting the product distributions of electrophilic aromatic substitution reactions.

There is a growing interest in computer assistance in predicting such product distributions. For example, CAMEO, an interactive computer program developed by Jorgensen et al.[4] predicts the products of organic reactions primarily through mechanistic reasoning. Elrod et al.[5,6] are predicting the product distributions by using neural networks, which are computer-based approaches that share numerous analogies to a model of human learning.

Whereas these approaches use quantitative data, our approach is modeled after developments in artificial intelligence (AI) approaches to "learning" transforms from a reaction database.[7-10] Reaction specificity can be achieved by associating reactivity factors with the atoms of the transform or molecule.[11,12] Another approach is to augment the changed bonds that define the reaction center with additional atoms and bonds that make the transform structurally more specific. These additional atoms may simply be the atoms immediately adjacent to the reaction center.[10] On the other hand, data on related reactions may be used to select these additional atoms.[7,9] We take this latter approach using the formalized concept of a transform. These transforms in essence specify a class of reactions characterized by a common structural change together with an associated substructural environment that must be present for the change to take place. This structural environment is obtained by examining the structural differences between compounds after prioritizing these differences in terms of their distances from the reaction center. We replace the traditional letter designations of the atoms and bonds in the chemical graph with a labeled subgraph which differentiates atoms by their charge and by their group number and period number in the periodic table, and which differentiates bonds by their σ and π bonding constituents. This enables us to derive a simple and highly predictive transform kit, composed of a set of transforms, that agrees with the traditional explanation as to why particular substituents are meta-directing.

GENERAL METHODOLOGY

Constructing an Extended Transform. Chemical reaction pathways may be represented as molecules which are pairwise connected by chemical transformations indicated by arrows drawn from the reactants to the possible products. These transformation arrows define a chemical transformation pathway (CTP). In this study, we convert the quantitative product percentages to "observed" or "unobserved" reaction pathways using a 10% criterion of the isomer percentage of the final overall product. Stated another way, if one isomer product percentage (o, m, p) of a nitration reaction is greater than or equal to 10%, we draw a transformation arrow from the reactant to that product; if the percentage is less than 10%, we do not draw a transformation arrow to that product. (See Figure 1.) For example, the nitration reaction of anisole shown in Figure 1a has three products (ortho, 34%; meta, 2%; para, 64%). After the quantitative product percentages are converted to transformation arrows by the 10% criterion, the nitration reaction pathway of anisole consists of an ortho-transformation arrow and a para-transformation arrow as shown in Figure 1b.

To compute the structural change of a chemical transformation in a CTP, we use the maximum common substructure approach.[9] (See ref 8 for another approach for computing the structural change.) This approach is illustrated for the para pathway of the anisole nitration reaction shown in Figure 1c. The reactant and the product are superimposed so as to match up as many atoms and bonds as possible. Following the notation of Fujita,[13] bonds of the reactant that cannot be matched up (deleted in the transformation) are marked with a double slash, while bonds of the product that cannot be matched up (added in the transformation) are marked with a small circle.

The core transform of a chemical transformation is obtained by deleting all the atoms and bonds of the common structure except those atoms of the common structure connected to a bond which is either broken or formed. The core transform defines the basic structural change in the reaction and usually corresponds to the reaction center. It occurs in various formalistic guises such as the R matrix in the Dugundji-Ugi theory,[14,15] the minimum reaction concept of Wilcox and Levinson,[9] and the reaction center graph of Fujita.[13]

The atoms and bonds deleted by the transform will be termed the leaving elements, and the atoms and bonds added by the transform will be termed the entering elements. In the example of anisole nitration reaction, the H atom and the C-H bond are the leaving elements, and the NO_2 group and the C-NO_2 bond are the entering elements. All elements of the transform that remain after excluding the leaving elements constitute the prepositioning structure of the transform. To see where and how a transform acts on a molecule, one completely superimposes the prepositioning structure on the chemical graph. Then one deletes the leaving elements and adds the entering elements to obtain the product of the re-

† Kent State University.
‡ Upjohn Laboratories.

0095-2338/90/1630-0442$02.50/0

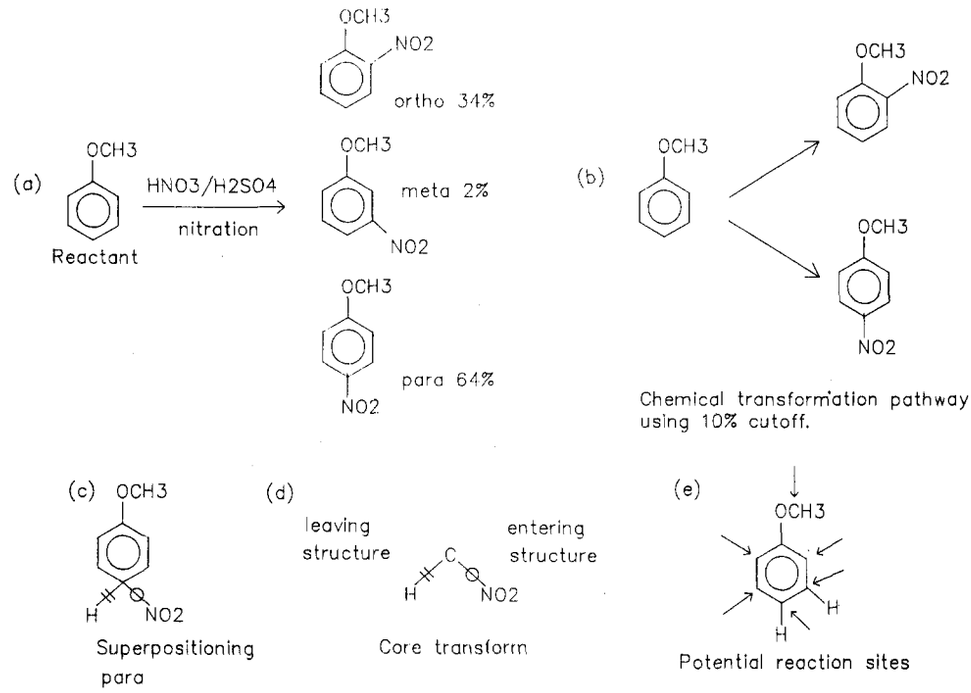

Figure 1. Core transform and its potential reaction sites.

action. If the prepositioning structure cannot be completely superimposed on the reactant, the transform does not act on that molecule.

Although the core transform in Figure 1d will "nitrate" the para position, it will also nitrate the meta and para positions and the methyl group. More specific transforms are obtained by deleting fewer atoms and bonds shared by the reactant and product. This in effect increases the number of atoms and bonds in prepositioning structure of the transform which must be shared by both the reactant and the product. The question is: Which ones should be deleted?

Figure 2 shows how the reaction data can be used to decide the issue. Anilinium nitrates at the meta position, but anisole does not. Chemical knowledge suggests that any structural explanation that explains why the two reactants yield different products is likely to reflect the structural elements nearest the reaction site that differentiate the two reactants. Thus, we shall seek a shortest, connected transform which contains the core transform as well as some structural elements of anilinium which are not shared with anisole. Transforms obtained in this manner will be called extended transforms. Those structural elements possessed by anilinium, but not by anisole, are called the differencing elements of the transform with respect to that reaction. (A brief glossary of definitions related to transforms is given in the Appendix.)

As the left side of Figure 2 illustrates, it is a trivial matter to construct an extended transform that acts at the meta position of anilinium but does not act on anisole. We simply let the charged nitrogen be one of the differencing elements of the extended transform.

Graph Representation of the Chemical Structures. Reactions proceed along different reaction pathways because of structural differences in the reactants. For example, the N^+ in anilinium (see Figure 2) favors nitration at the meta position, whereas the O at the corresponding position of anisole favors nitration at the ortho and para positions. If we, as chemists, examine the similarity between O and N^+, we know that the oxygen and nitrogen are very similar with respect to their reactivity such as electronegativity, size, etc. The real difference between O and N^+ that causes the different reactivities in nitration reaction comes from the positive charge. Chemical knowledge tells us that the positive charge on the nitrogen will be

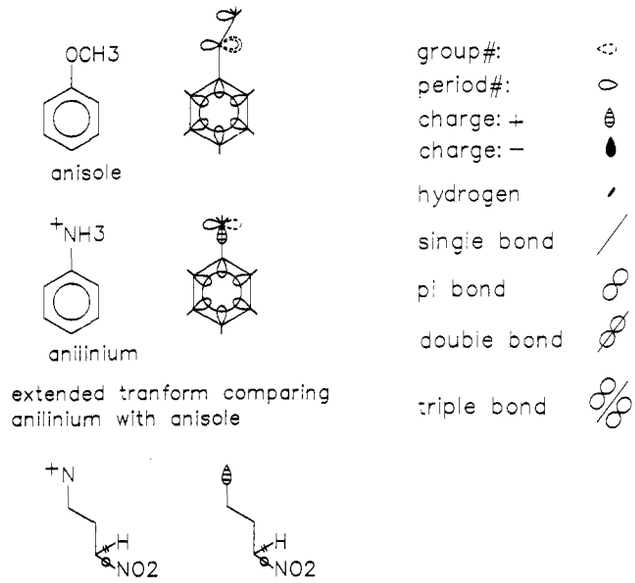

Figure 2. Extended transforms that nitrate anilinium in the meta position but do not act on anisole expressed in terms of the chemical graph and after replacing the atom and bond labels with corresponding typing labels.

meta-directing. If the computer is to generate analogous explanations, it must be able to distinguish differences as well as similarities related to the chemical interpretations of the symbols N^+ and O.

In this paper, the chemical structures are represented by modifications of chemical graphs. The chemical graph is a graph in which each vertex is assigned a single atomic label (N^+, O, etc.) and each edge is assigned a single bond label (single, double, etc.). In our study, the single atom and bond labels are replaced by multiple atom and bond feature labels defined as follows:

i. Group label: The group label of a vertex (atom) is the group number of the atom in the periodic table minus 4. For example, the group label of a C atom equals 0, an N atom equals 1, an O atom equals 2, and an F atom equals 3, etc. This label reflects the

number of valence lone-pair electrons as well as the relative electronegativity of the atom. It is represented in the modified chemical graph by dotted loops at the right of the atom vertex. (See Figure 2.)

ii. Period label: The period label of a vertex (atom) is the periodic number of the atom in the periodic table or the principal quantum number of the atom minus 1. This label reflects the relative size of the atom and is represented by solid loops at the left of the atom vertex.

iii. Charge label: If an atom in the molecule has a negative charge, we add a solid dark loop below the atom vertex; if the atom has a positive charge, we add a hatched loop label below the atom vertex.

iv. Hydrogen label: An atom–hydrogen bond and the hydrogen atom are treated differently from other kinds of atom–atom bonds in the model. Hydrogen is treated as a property associated with the atom (vertex) bonded with the hydrogen. The atom–hydrogen bond and hydrogen atom are represented by a short dark line.

v. Bond labels: Single bonds between atoms are represented by a single line, the same as in the chemical structural formula. Double bonds are treated as a combination of a σ bond and a π bond, which is represented by two adjoining circles. Triple bonds are treated as a combination of a σ bond and two π bonds. Aromatic bonds are treated as a combination of a σ bond and a delocalized π bond represented by a dashed line or by a circle if a complete aromatic ring is depicted.

In graph-theoretic representations of molecules, the concept of a substructure is the principle means of defining structural features shared by two molecules. By replacing the atom and bond labels with the preceding atom and bond feature labels, the substructure concept can better resolve the similarities and dissimilarities between molecules. For example, we see in the chemical graphs of Figure 2 that the charged nitrogen of anilinium cannot be superimposed on the oxygen of anisole as the N^+ and O are distinct labels. On the other hand, in the modified chemical graphs, the vertices corresponding to those two atoms can be superimposed, as can their period label and one of their group labels. However, the positive charge label and the three hydrogen labels of the nitrogen have no counterparts associated with the vertex of the oxygen atom.

This enhanced resolution has important consequences in the construction of extended transforms. For example, the middle of Figure 2 illustrates an extended transform that will nitrate anilinium at the meta position, but will not act on anisole. This extended transform has two distinct advantages over its chemical-graph counterpart. It will act on any reactant that its counterpart will act on, but not conversely. As we shall see, this greater generality usually leads to simpler models of a chemical-transformation pathway. In addition, extended transforms constructed from the modified chemical graph are more closely tied to mechanistic interpretations.

Graph Transform Kits. A graph transform kit (GTK) is often nothing more than a set of transforms. We have already indicated that a transform "acts" on a molecule if the transform's prepositioning structure can be completely superimposed on the chemical graph (or modified chemical graph) of that molecule. In this basic form, a GTK "acts" on a molecule if at least one of its transforms acts on that molecule. Since the prepositioning structure of a transform can often be superimposed in more than one way on a molecule, and since there are often many transforms in a GTK, it follows that a GTK can often act at a variety of positions or sites on a molecule.

It has been shown that there are chemical transformation pathways that cannot be exactly modeled by using only GTKs of this basic form.[16,17] However, by dividing the set of transforms of a GTK into a set of "inducing" transforms and a set of "blocking" transforms, one can obtain generalized GTKs that can exactly model any chemical transformation pathway. Although blocking transforms are not required for the particular modeling problem of this study, we will still refer to the transforms incorporated into the learned transform kit as inducing transforms.

MODELING THE META PATHWAY OF AROMATIC NITRATION REACTIONS

We selected as a learning dataset the 31 compounds asterisked in Table I that had no non-hydrogen atoms more than two bonds distant from the aromatic ring. Every compound in the learning set has three possible pathways (ortho, meta, and para). In this study, we model only the meta pathway. Thus, as indicated in Figures 3 and 4, we split the compounds into two sets: meta structures and non-meta structures.

Minimal Differencing Transforms. We seek a transform kit that will nitrate at the meta position those compounds that have meta-directing substituents, but will not act on any of those compounds without meta-directing substituents. In the first step of the modeling process, we find the simplest extended transforms which operate on a particular meta structure A_i of Figure 4, but not on a particular non-meta structure B_j. By simplest, we mean that if any structural element of the transform is deleted, then the transform ceases to be an extended transform that acts on A_i, yet fails to act on B_j. Such a transform is called a minimal differencing transform (MDT).

The following example illustrates why there can be more than one MDT associated with a pair of structures. Consider the minimal differencing transforms obtained by comparing meta structure A_1 (anilinium) and B_{11} (ethylbenzene) in Figure 5. First, we find that the lowest differencing level is at the first vertex (carbon) of the substituent group. Secondly, we find that at this vertex there are three types ($k = 1, 2, 3$) of different elements possessed by A_1 but not by B_{11}: the positive charge label, the lone-pair electron label, and the three hydrogen labels. The positive charge label and the lone-pair electron label each give rise to a MDT. Anilinium possesses three hydrogen labels at the lowest difference level. However, since ethylbenzene possesses two hydrogen labels, the corresponding MDT must possess three hydrogen labels in order to act on anilinium, but not ethylbenzene. Since anilinium and ethylbenzene possess the same number of period labels, there is no MDT in which the differencing elements are period labels. Note that if any structural element is removed from one of the three MDTs in Figure 5b, that transform fails to satisfy the previously defined extended transform property of being a shortest, connected transform that nitrates anilinium in the meta position, but does not act on anisole. Figure 5c gives all of the MDTs that are constructed by comparing anilinium to each of the non-meta structures in this manner. The upper part of Figure 6 gives the MDT sets for selected meta structures 1–3 and 15–17.

Minimal Inducing Transforms. Each MDT in Figure 5c constitutes a structural "explanation" as to why anilinium undergoes meta nitration and some other non-meta structure does not. Their diversity suggests that a combination of these structural features may constitute a better "explanation", one that explains why more non-meta structures are excluded. The next step constructs extended transforms that are in effect combinations of MDTs.

In this step, we find the simplest extended transforms which operate on a particular meta structure A_i but not on any of the non-meta structures B_j, $j = 1, \ldots, 12$. Such a transform

Table I. Nitration Reaction Data of Monosubstituted Benzene Compounds

compound	substituent	ortho, %	meta, %	para, %
diphenylamine	NHPh	71.0	0.0	29.0
aniline*	NH_2	50.0	0.0	50.0
biphenyl	Ph	58.0	0.0	42.0
cinnamic acid	CH=CHCOOH	33.0	0.0	67.0
fluorobenzene*	F	12.4	0.0	87.6
isopropylbenzene*	$CH(CH_3)_2$	14.0	0.0	86.0
phenoxide*	O^-	80.0	0.0	20.0
phenol*	OH	51.0	0.0	49.0
ethylbenzene*	CH_2CH_3	55.0	0.0	45.0
chlorobenzene*	Cl	29.6	0.9	69.5
bromobenzene*	Br	36.5	1.2	62.4
iodobenzene*	I	38.3	1.8	59.7
acetanilide	$NHCOCH_3$	19.0	2.0	79.0
anisole*	OCH_3	34.0	2.0	64.0
β-nitrostyrene	$CH=CHNO_2$	31.0	2.0	67.0
β-styrenesulfonyl chloride	$CH=CHSO_2Cl$	31.0	2.0	67.0
methyl phenethyl ether	$CH_2CH_2OCH_3$	41.0	3.0	56.0
3-phenylpropyl methyl ether	$CH_2CH_2CH_2OCH_3$	44.0	4.0	52.0
toluene*	CH_3	58.0	4.0	38.0
ethyl phenylpropiolate	$CCCOOCH_2CH_3$	36.0	6.1	57.9
phenylpropiolic acid	CCCOOH	27.0	8.0	65.0
tert-butylbenzene*	$C(CH_3)_3$	12.0	8.5	79.5
ethyl phenylacetate	$CH_2COOCH_2CH_3$	42.0	10.6	47.7
methyl benzyl ether	CH_2OCH_3	39.0	12.0	49.0
benzylsulfonate	$CH_2SO_2O^-$	33.5	13.7	52.8
benzyl cyanide	CH_2CN	17.0	14.0	69.0
chlorotoluene*	CH_2Cl	32.0	15.5	52.5
2-fluorotoluene*	CH_2F	28.0	18.0	54.0
benzyltrimethylphosphonium	$CH_2P^+(CH_3)_3$	13.1	19.4	67.5
benzylsulfonamide	$CH_2SO_2NH_2$	27.9	31.4	40.7
benzylmethylsulfonate	$CH_2SO_2OCH_3$	24.7	32.4	42.9
dichlorotoluene*	$CHCl_2$	23.3	33.8	42.9
TMS*	$Si(CH_3)_3$	25.5	39.8	30.2
anilinium*	N^+H_3	4.0	42.0	54.0
benzyl chlorosulfonate	CH_2SO_2Cl	16.9	50.8	33.1
α-nitrotoluene	CH_2NO_2	22.0	55.0	23.0
benzenesulfonate*	SO_3^-	20.0	60.0	20.0
trichlorotoluene*	CCl_3	6.8	64.5	28.7
ethyl benzoate	$COOCH_2CH_3$	28.3	68.4	3.3
N-methylanilinium*	$NH_2^+CH_3$	0.0	70.0	30.0
benzamide*	$CONH_2$	27.0	70.0	3.0
acetophenone	$COCH_3$	26.0	72.0	2.0
benzaldehyde*	CHO	19.0	72.0	9.0
dimethylanilinium*	$NH^+(CH_3)_2$	0.0	78.0	22.0
benzoic acid*	COOH	18.5	80.2	1.3
benzonitrile*	CN	16.0	82.0	2.0
diphenyliodonium	I^+Ph	7.5	82.5	10.0
benzyltrimethylammonium	$CH_2N^+(CH_3)_3$	2.0	88.0	10.0
nitrobenzene*	NO_2	6.4	93.2	0.4
4-(methylsulfonyl)benzoic acid*	$S^+(CH_3)_2$	0.0	95.0	5.0
methyl phenyl sulfone*	SOO_2CH_3	0.0	100	0.0
trifluorotoluene*	CF_3	0.0	100	0.0
phenyltrimethylammonium*	$N^+(CH_3)_3$	0.0	100	0.0
phenyltrimethylphosphonium*	$P^+(CH_3)_3$	0.0	100	0.0

will be called a minimal inducing transform with respect to A_i. By "simplest" we mean that if any structural element is removed, then the transform ceases to be an extended transform with the desired property of operating on A_i but not on any of the non-meta structures.

To define an algorithm for constructing minimal inducing transforms, we need to define reductions and extensions of transforms. A transform t is a reduction of another transform t' (and conversely, t' is an extension of t) if t and t' have identical core transforms and every structural element of t is also a structural element of t'. The minimal inducing transforms with respect to meta structure A_i can now be defined by the following algorithm:

i. Construct the set T_i of all MDTs constructed from meta structure A_i. This set is given for anilinium in Figures 5c and 7a.

ii. Determine the simplest transform $t^c(i)$ that is an extension of every transform in the minimal differencing transform set T_i. The transform $t^c(i)$ can be obtained by constructing the minimal superstructure of every minimal differencing transform in T_i. This transform is given for anilinium in Figure 7b. Note that $t^c(i)$ will always act on A_i. In particular, $t^c(1)$ given in Figure 7b acts on anilinium.

iii. Construct all the transforms that are reductions of $t^c(i)$ and are also extensions of at least one transform in T_i. Let T_i^c denote this set of constructed transforms. These transforms are partially ordered in Figure 7c according to the reduction/extension relationship between transforms.

iv. Select each transform in T_i^c that operates on A_i but does not operate on B_j for any j, $j = 1, ..., 12$. Let T_i^m denote this set of marked transforms. (See Figure 7d.)

v. Those transforms in T_i^m that are not extensions of any other transforms in T_i^m are the minimal inducing transforms with respect to A_i. The minimal inducing transforms for anilinium are given Figure 7e. The

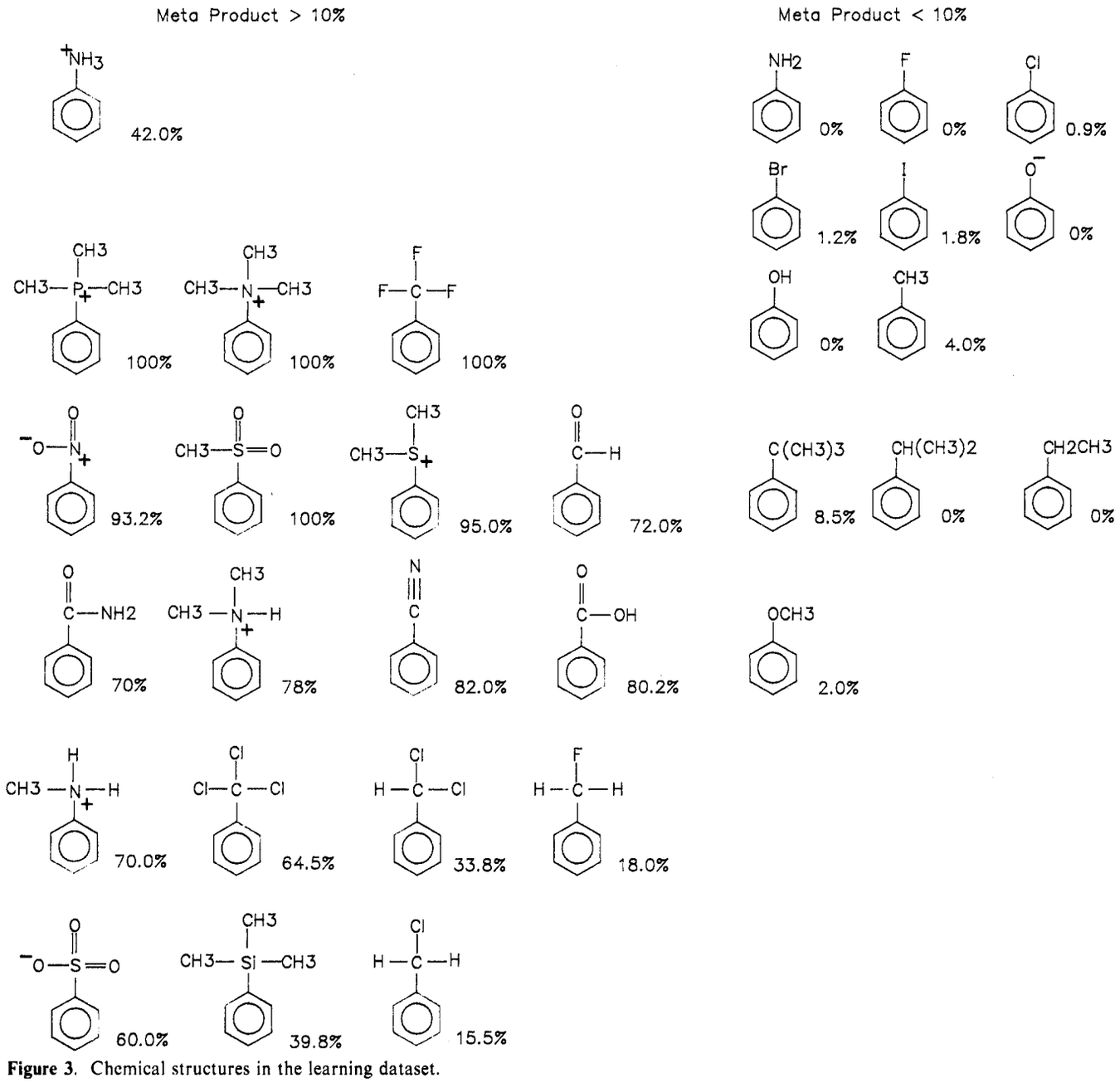

Figure 3. Chemical structures in the learning dataset.

minimal inducing transforms for selected meta structures are given in the lower half of Figure 6.

Graph Transform Kits. The minimal inducing transforms in Figure 6 provide combinations of structural features that must be present for a compound to undergo meta nitration. These combinations are sufficiently restrictive to guarantee that no non-meta structure in the learning dataset can satisfy them. However, each one is too restrictive to assure that every meta structure in the learning dataset will satisfy them. Thus, we might "explain" that a compound nitrates in the meta position if at least one of these transforms operates on its modified chemical graph.

This explanation is achieved via a graph transform kit. As defined previously, a GTK acts on a structure if one or more of its transforms acts on that structure. We seek a GTK that acts on every meta structure in the learning dataset, but does not act on any non-meta structure. A GTK consisting of all the minimal inducing transforms will satisfy our purpose. However, we may not need all of the transforms in the kit. The following algorithm will generate a minimal kit K that satisfies our purpose.

i. Let K consist of all minimal inducing transforms.
ii. Repeat this step for every transform in K. Remove a transform from K to form K'. Check if K' operates on every meta structure.
iii. If there exists a K' at step ii that operates on every meta structure, redefine K equal to one of those $K's$, and repeat step ii. Otherwise, K is a minimal GTK.

To illustrate, we begin with K equal to T^s in Figure 8a. In Figure 8c, we indicate the structures on which each transform in K operates. By construction, none of the transforms operates on the non-meta structures. At step ii, we see from Figure 8c that either t_2^s, t_5^s, or t_6^s can be removed. After these three transforms have been removed, no others can be removed if the resulting GTK is to operate on every meta structure. Thus, our minimal GTK is given by the three transforms in Figure 8b.

Depending on which K' is selected at step iii, one may end up with different GTKs, all of which would be minimal. Although the resulting GTKs may differ, each GTK would act on all of the meta structures. Moreover, if any additional transform is removed from one of these minimal kits, the

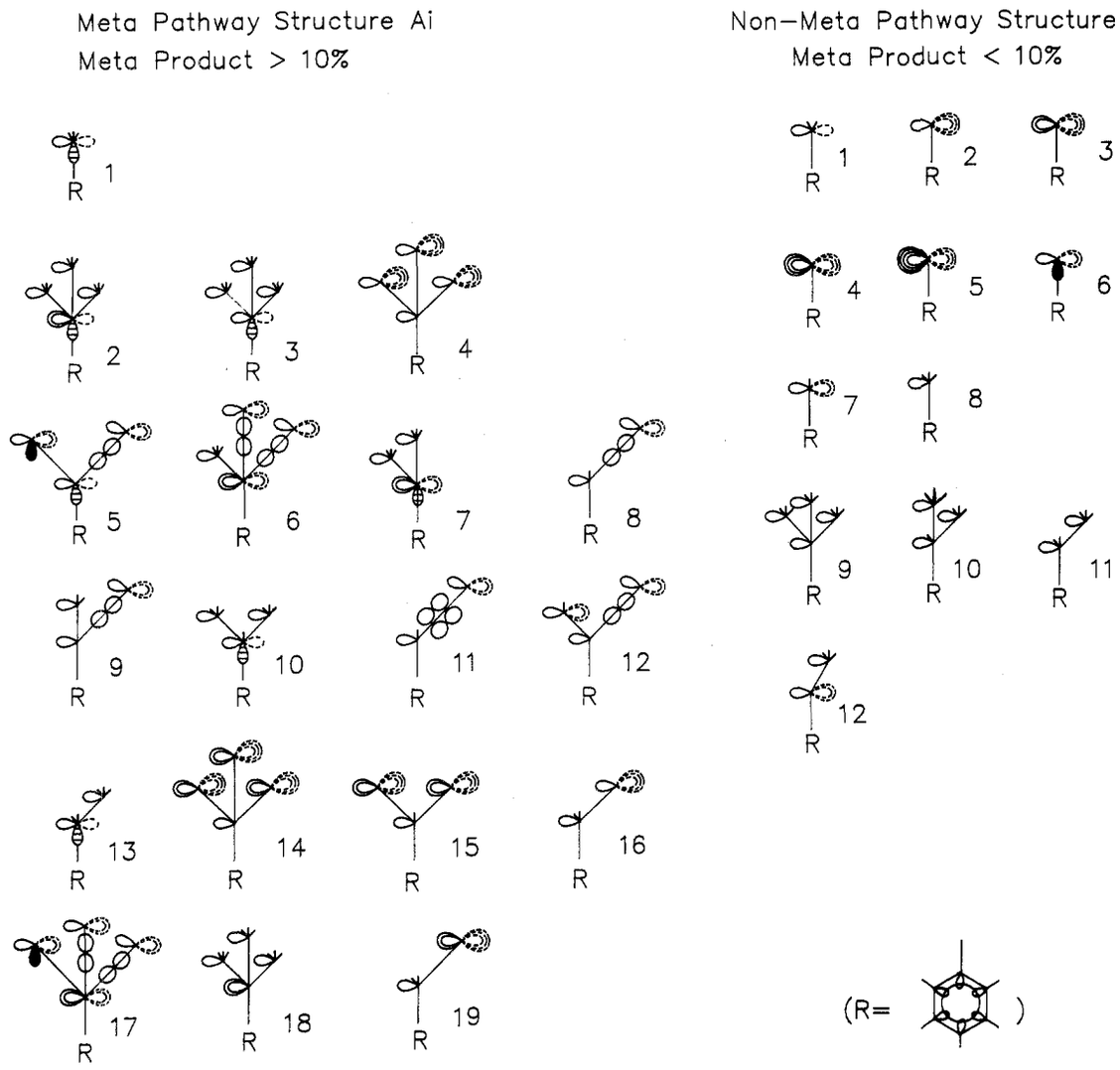

Figure 4. Modified chemical graph representations of the structures in the learning dataset.

resulting kit would fail to act on at least one meta structure. In our particular example, there was only one minimal GTK.

DISCUSSION

The simplicity and generality of the GTK in Figure 8b suggest the utility of such kits in modeling chemical reactions. Two of the transforms are directly interpretable. Transform t_1 indicates that a positive charge at the first atom level of the substituent is meta-directing, which agrees with our understanding that the positive charge destabilizes the transition states of the ortho and para pathways more than that of the meta pathway. Transform t_3 indicates that atoms in groups 5–7 of the periodic table, when at the second atom level, are also meta-directing. This agrees with the fact that electronegative atoms at the second level are electron-withdrawing groups and, consequently, meta-directing.

The absence of a transform with the group label at the first level suggests that electronegative atoms such as N, O, and F at the first level are not meta-directing. This is consistent with our understanding of the overriding resonance effects associated with the formation of the immonium, oxonium, and halonium intermediates.

The presence of transform t_2 in the kit illustrates how modeling efforts of this type alert us to data requiring closer inspection. In this case we lack an immediate interpretation of the transform based on our understanding of electrophilic aromatic reactions. However, comparing the three nitration reactions of chlorobenzene (meta, 0.9%), *tert*-butylbenzene (meta: 8.5%), and tetramethylsilane (TMS) (meta, 39.8%), we can see that t_2 is the simplest transform which will operate on the meta-pathway structure TMS but not on either of the non-meta structures chlorobenzene or *tert*-butylbenzene.

This closer examination can lead to a refinement in our chemical understanding. Transform t_2 reflects a particular combination of size and inductive effects of the substituent group on the meta product percentage. The second (or higher) period atom at the first vertex will increase the size of the substituent so as to increase the meta percentage. The single bond with another atom on the second level means that the first vertex atom can not belong to the halogen group which would increase the ortho-para product percentage.

Since the constructed GTK agrees so closely with our chemical understanding, we can expect it to be quite predictive. To quantitatively establish this point, we predicted for the remaining 23 nonasterisked compounds in Table I whether or not their percent metabolic product exceeded 10% or more. All of the compounds in the testing set are monosubstituted benzene derivatives with one or more non-hydrogen atoms at least three levels from the phenyl ring. There were 13 meta structures and 10 non-meta structures in this testing set. The constructed transform kit in Figure 8b correctly predicted all 10 non-meta reaction pathway products and 11 of the 13 meta reaction pathway products. The classification error is less than 10%, which is quite satisfactory. The two structures, ethyl phenylacetate and benzyl cyanide, which were incorrectly

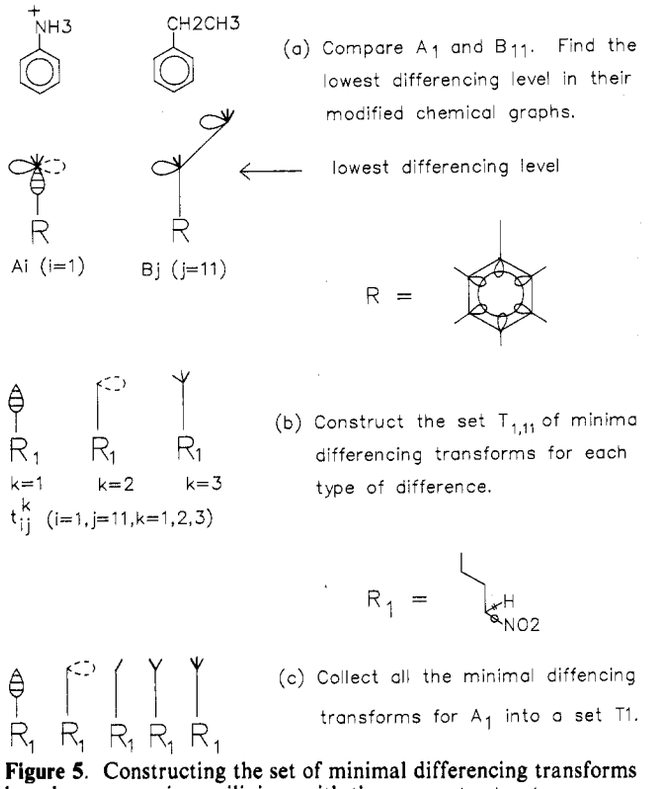

Figure 5. Constructing the set of minimal differencing transforms based on comparing anilinium with the non-meta structures.

classified as non-meta structures, had meta product percentages of 10.6% and 14%, which are near the 10% cutoff.

The algorithm for constructing the graph transform kit is based on qualitative modeling of a reaction pathway model. That is, the existence of reaction arrows are predicted, but quantitative reaction rates or relative product rates are not. Partly because of this qualitative nature, the algorithm can search over an almost unlimited range of transforms in modeling a chemical transformation pathway. This range is suggested by the variety of minimal differencing transforms and their extensions that were generated and examined in this study. However, more extent reaction rate and relative rate prediction require more quantitative models. There are a variety of approaches in which quantitative considerations might be introduced. One is to use more or less standard QSAR techniques in which the selected transforms become the chemical descriptors in a regression analysis or become the input to a neural net. We are currently exploring some of these possibilities.

As currently formulated, the learning algorithm takes no account of noise or variation in the reaction rate data that cannot or should not be fully explained away by the model. By introducing statistical considerations at step iv under Minimal Inducing Transforms when selecting the inducing transform and at steps ii and iii under Graph Transform Kit when selecting the minimal graph transform kit, one should be able to obtain methods of selecting GTKs that are less sensitive to inherently unpredictable variation in the data. We are also exploring some of these possibilities.

ACKNOWLEDGMENT

We thank Dave Elrod for providing the data in Table I and Gerald Maggiora, Director Computational Chemistry, Upjohn Laboratories, for making available partial support for this research in the form of a summer student employment program.

APPENDIX

Transform: A specification of a structural change common to a class of reactions along with a structural environment

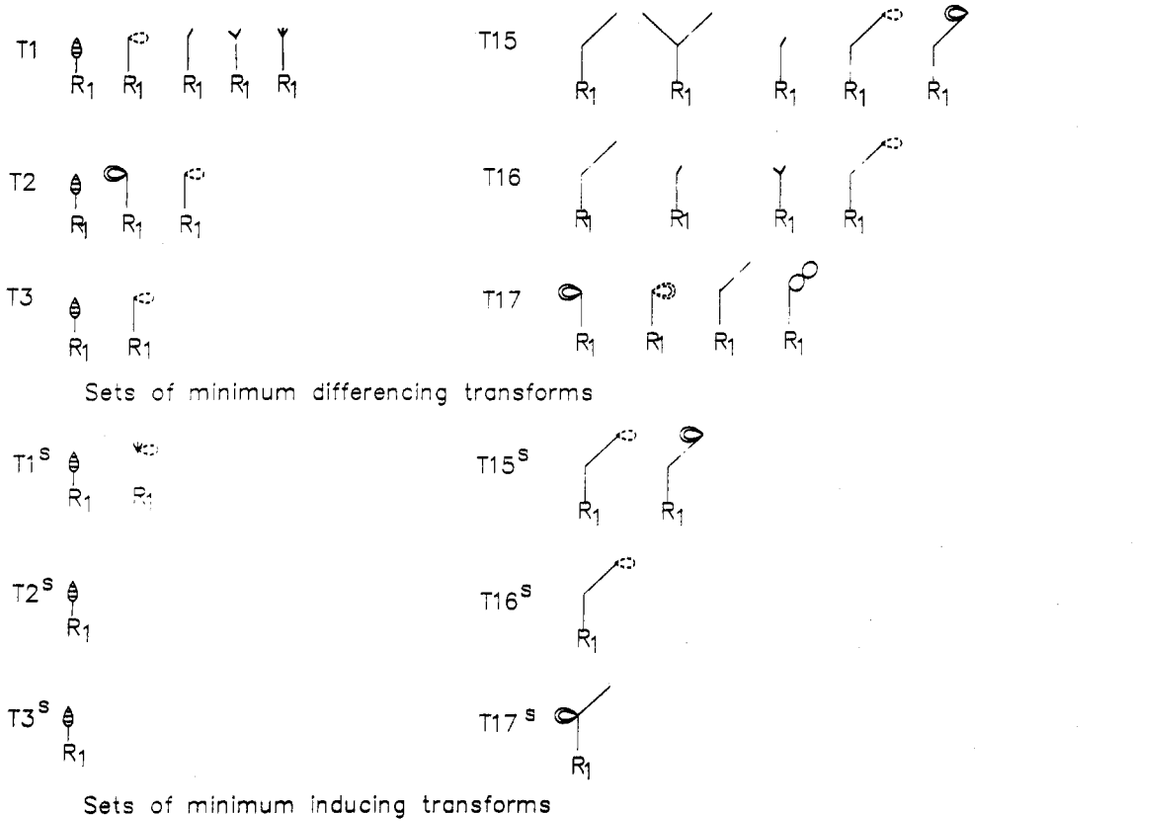

Figure 6. The set of minimal differencing transforms and the set of minimal inducing transforms constructed with respect to selected meta structures.

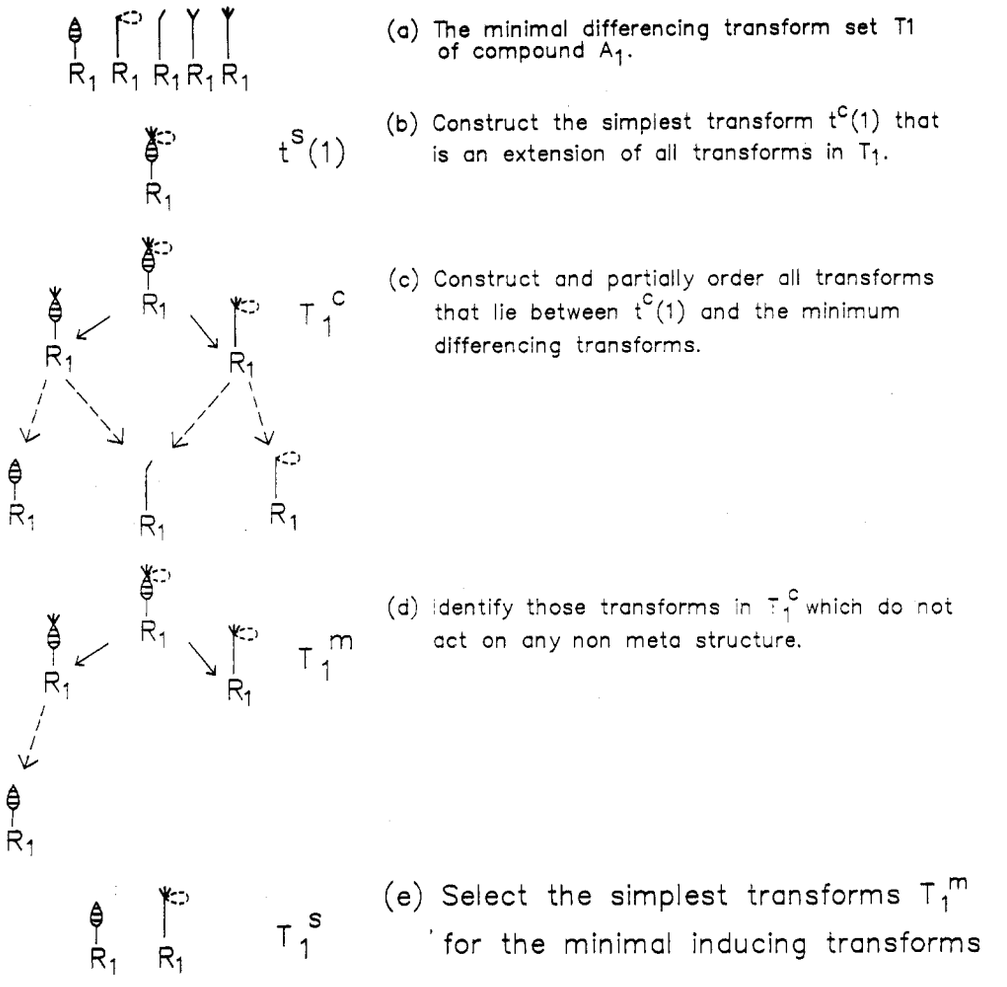

Figure 7. Constructing the minimal inducing transforms from the set of minimal differencing transforms of anisole.

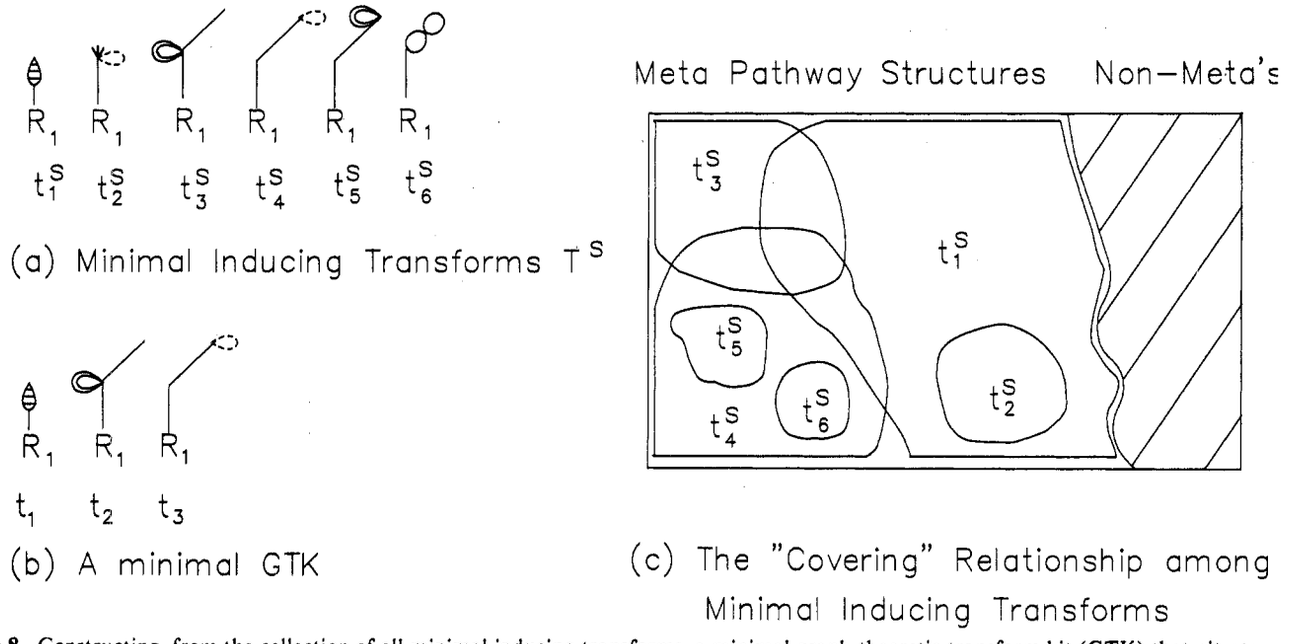

Figure 8. Constructing, from the collection of all minimal inducing transforms, a minimal graph-theoretic transform kit (GTK) that nitrates all meta structures in the meta position, but does not act on any of the non-meta structures in Figure 4.

that must be present for that change to occur.

Entering (leaving) elements of a transform: Those elements of a transform that are to be added to (deleted from) the reactant to form the product.

Prepositioning structure of a transform: All elements of the transform excluding the leaving elements.

Core transform: The least restrictive transform among all transforms generating a particular structural change.

Extended transform: A shortest, connected transform that will act at a desired site on one molecule and not at an undersired site on another molecule.

Graph transform kit (GTK): A set of transforms.

Inducing transform: A transform in a GTK that can generate a transformation as opposed to a "blocking" transform which can prevent an action of an inducing transform.

Minimum differencing transform: A least-restrictive extended transform that will act at a desired site of one molecule but will not act at an undesired site of another molecule.

Reduction (extension) of a transform: A transform t is a reduction of another transform t' (and conversely, t' is an extension of t) if t and t' have identical core transforms and every structural element of t is also a structural element of t'.

Minimum inducing transform: A least-restrictive transform that will act at a desired site of one molecule but will not act at any sites in a set of undesired sites.

Minimal GTK: A GTK which will act at all sites in a set of desired sites, but will lose this property if any transform of the kit is removed.

REFERENCES

(1) Carey, F. A.; Sunberg, R. J. *Advanced Organic Chemistry*, 2nd ed., Part A; Plenum Press: New York, 1984; pp 481–503.
(2) Ingold, C. K. *Structure and Mechanism in Organic Chemistry*; Bell: London, 1963.
(3) Remick, A. E. *Electronic Interpretation of Organic Chemistry*, 2nd ed.; Wiley: New York, 1959.
(4) Bures, M. G.; Roos-Kozel, B. L.; Jorgensen, W. L. Computer-Assisted Mechanistic Evaluation of Organic Reactions. 11. Electrophilic Aromatic Substitution. *J. Org. Chem.* **1985**, *50*, 4490–4498.
(5) Neural Network Applications in Chemistry Begin to Appear. *C&EN News* **1989**, *67*, *No. 4*, 24–28.
(6) Elrod, D. W.; Maggiora, G. M.; Trenary, R. G. Applications of Neutral Networks in Chemistry. 1. Prediction of Electrophilic Aromatic Substitution Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 477.
(7) Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*; McGraw-Hill Co.: New York, 1980.
(8) Weise, A. Computerized Analysis of Chemical Reaction Equations with Correction of Stoichiometry. *J. Prakt. Chem.* **1980**, *322*, 761–768.
(9) Wilcox, C. S.; Levinson, R. A. In *Artificial Intelligence Applications in Chemistry*; Pierce, T. H., Hohne, B. A., Eds.; ACS Symposium Series 306; American Chemical Society: Washington, DC, 1986.
(10) Yanaka, M.; Nakamura, K.; Kurumisawa, A.; Wipke, W. T. In *QSAR: Quantitative Structure–Activity Relationships in Drug Design*; Fauchere, J. L., Ed.; Alan R. Liss, Inc.: New York, 1989.
(11) Gasteiger, J.; Saller, H.; Low, P. Elucidating Chemical Reactivity by Pattern Recognition Methods. *Anal. Chim. Acta* **1986**, *191*, 111–123.
(12) Weise, A. Derivation of Organic Chemistry Reactions with the AH-MOS Simulation Program. *Z. Chem.* **1975**, *15*, 333–340.
(13) Fujita, S. "Structure–Reaction Type" Paradigm in the Conventional Methods of Describing Organic Reactions and the Concept of Imaginary Transition Structures Overcoming This Paradigm. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 120–126.
(14) Dugundji, J.; Ugi, I. An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs. *Fortschr. Chem. Forsch.* **1973**, *39*, 19–64.
(15) Ugi, I.; Wochner, M.; Fontain, E.; Bauer, J.; Gruber, B.; Karl, R. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley Interscience: New York, 1990.
(16) Johnson, M. A. In *Proceedings of the Sixth International Conference on the Theory and Applications of Graphs*; Alavi, Y., Chartrand, G., Oellermann, O. R., Schwenk, A. J., Eds.; John Wiley: New York, 1988.
(17) Johnson, M. A.; Gifford, E.; Tsai, C.-c. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley Interscience: New York, 1990.