

## Reaction Indexing for Reaction Databases

JAMES B. HENDRICKSON\* and TODD M. MILLER

Department of Chemistry, Brandeis University, Waltham, Massachusetts 02254-9110

Received July 4, 1990

A program is described to translate the reaction entries in the REACCS and SYNLIB databases into the same terms as those employed in the SYNGEN program for synthesis design. This serves as an interface to allow rapid retrieval of closely matching literature references for any construction reaction generated by SYNGEN. In a similar fashion the refunctionalization reactions may also be organized and searched. Almost all the reactions in these databases are readily amenable to indexing by the SYNGEN system in a very simple fashion.

In the development of our SYNGEN program for synthesis design, we generated the reactions in the synthesis routes from a basis of chemical logic, specifying the *net structural change* for a reaction.<sup>1</sup> While this frees us from deriving particular reactions out of a given literature database contained in the program, it also leaves us at the end of the generation with routes of reactions which are not linked to any literature precedent. However, this lack can be rectified since the same logic that is used in SYNGEN to describe reactions can also be used to organize and index any existing database of reactions. In this way we can locate literature precedents for any reaction after it has been generated by SYNGEN.

Since SYNGEN has now reached a working state for export and we are interested in having it used in the organic synthesis community, we realized that a link from the theory of the program to the practical base of literature precedents for the laboratory was required. Therefore, we undertook to organize and index two popular databases, REACCS and SYNLIB, so that users possessing these programs could be provided with a direct retrieval from SYNGEN.

This organization of reactions must focus only on those atoms which change their connectivity from reactant to product, and not on the whole structures. The changing atoms alone constitute the *net structural change* in the reaction. Such an organization of reactions can then be overlaid on the database as an index system to be used for rapid retrieval of close matches. In this paper we show that almost all reactions in two commercial databases can be efficiently indexed and rapidly and reliably retrieved, by applying this simple organizing system.

## 1. REACTION CLASSIFICATION

The basis for reaction classification is simple and digital, but fundamental and rigorous, in the characterization of structures and reactions developed for SYNGEN. This consists of describing four synthetically important kinds of bonds on any skeletal carbon:

- H for hydrogen, electron pair or electropositive atom
- R for  $\sigma$ -bond to carbon
- $\Pi$  for  $\pi$ -bond to carbon
- Z for any bond ( $\sigma$  or  $\pi$ ) to electronegative atom

To describe a structure, each carbon may then be separately described by the *numbers* of each kind of bond:  $h$ ,  $\sigma$ ,  $\pi$ , and  $z$ , respectively, with a sum of 4. In this way the skeletal level of any carbon is its  $\sigma$  value ( $\sigma = 1, 2, 3, 4$ : primary, secondary, tertiary, quaternary, respectively), while its functionality is expressed by the two digits  $z$  ( $=0-4$ ) and  $\pi$  ( $=0-2$ ). The value of  $h$  is obtained then by subtraction:  $h = 4 - (\sigma + \pi + z)$ . The oxidation state of any carbon is then  $x = z - h$ .

Reactions are now described by the nature of the change occurring at each changing carbon of the skeleton. Here we

characterize a *unit reaction* as a unit exchange of attachments at each changing carbon, i.e., as a gain of H, R,  $\Pi$ , or Z with a loss of H, R,  $\Pi$ , or Z. Therefore, 16 unit reactions are possible at any one changing carbon, and these are annotated with a two-letter symbol, the first the bond made, the second the bond broken. These 16 symbols must describe every possible change at any one carbon. They directly order themselves into common chemical families:

HH, ZZ	simple substitutions
HZ, ZH	reductions and oxidations, respectively
$\Pi$ H, $\Pi$ Z, $\Pi$ R	eliminations
H $\Pi$ , Z $\Pi$ , R $\Pi$	additions
RH, RZ, R $\Pi$	constructions
HR, ZR, $\Pi$ R	fragmentations

The RR change is characteristic of the migrating carbon in a rearrangement, and the  $\Pi\Pi$  change is that of the central carbon in an allylic shift. Any carbon with an R or  $\Pi$  symbol must involve an adjacent carbon with the same change in R or  $\Pi$ , so that an actual full reaction may involve several carbons changing, each with a symbol linked to its neighbor. In this way a unit oxidative addition to a double bond is Z $\Pi$ .Z $\Pi$ , denoting both changing carbons; each is changing by  $\Delta z = +1$  and  $\Delta \pi = -1$ . Rearrangements have three changing carbons, as in ZR.RR.RZ, and an allylic substitution will be Z $\Pi$ . $\Pi\Pi$ . $\Pi$ Z, also with three carbons. The reactions which do not make or break C-C  $\sigma$ -bonds, i.e., which do not alter the carbon skeleton, are *refunctionalization* reactions. It is rare for any unit refunctionalization reaction to involve more than three changing carbons.

The construction reactions are central to synthesis and to SYNGEN, since they create the C-C  $\sigma$ -bonds of the target skeleton. A general view of construction reactions is shown in Figure 1 with up to three carbons changing, out from each end of the construction bond. These skeletal carbons are labeled as  $\alpha$ ,  $\beta$ , and  $\gamma$  out from the bond formed on each side. Thus, any construction reaction may be seen as a combination of two *half-reactions*, involving as many as three carbons each which change their functional groups during the construction.

There are six possible unit half-reactions for construction; these may be formulated as shown below in Figure 1, and a calculation of  $\Delta x$  for each of the six half-reactions shows three oxidative ones of  $\Delta x = +1$ , which are nucleophilic, and three reductive ones of  $\Delta x = -1$ , which are electrophilic. Each has a simple substitution, involving only one carbon, an addition to a  $\pi$ -bond, involving two carbons, and an allylic substitution on three carbons. Combining these all ways yields nine full construction reactions which are *isohypsic*,<sup>2</sup> i.e., neither reductive or oxidative overall.

Actual constructions are not all described by these combinations because some are composite reactions with both a refunctionalization and a construction combined in one step. These composites are recognized in SYNGEN in three common

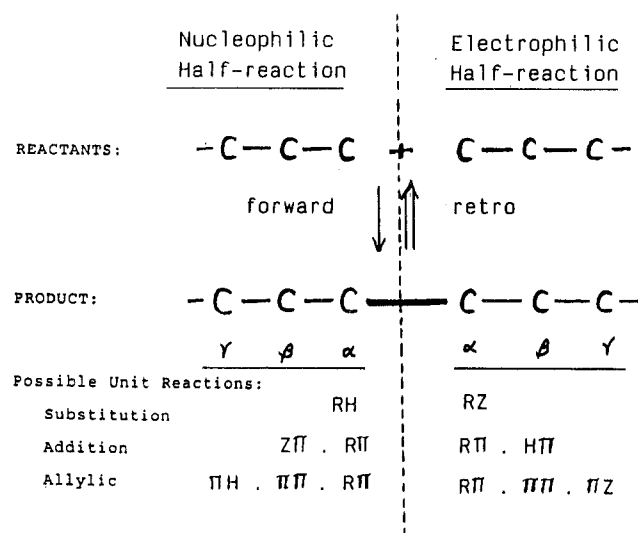


Figure 1. Generalized form of construction reactions.

categories: (1) prior reduction (cf., HZ) to form a carbanion, followed by construction (RH) at the  $\alpha$  carbon, as in the Grignard reaction; (2) elimination ( $\Pi$ H. $\Pi$ Z or  $\Pi$ Z. $\Pi$ Z) following construction (as in the aldol or Wittig reaction); or (3) tautomerization before or after construction. The list of possible construction half-reactions in SYNGEN was then expanded from the six in Figure 1 to incorporate these composites. With some further subdivision into common chemical families, we defined 16 nucleophilic and 9 electrophilic construction half-reactions, identified by simple two-character labels.<sup>3</sup> These 25 half-reactions are shown in Table I.

A construction half-reaction may be characterized as well by the changes taking place in the functionality on the  $\alpha$ ,  $\beta$ , and  $\gamma$  carbons. Here we create a  $z\pi$ -list of carbons  $\alpha$ ,  $\beta$ , and  $\gamma$  (i.e.,  $z_\alpha\pi_\alpha z_\beta\pi_\beta z_\gamma\pi_\gamma$ ) for both the reactant and product, and then subtraction of product from reactant affords a  $\Delta z\pi$ -list. These  $\Delta z\pi$ -lists then define the overall net functionality change for each of the 25 possible half-reactions. To identify a construction reaction in a database, we must then be able to assign to each of its two  $\alpha\beta\gamma$  strands a  $\Delta z\pi$ -list and recognize each as one of the defined 25 half-reactions.

## 2. INDEXING THE REACCS DATABASES

The terms outlined above can now be overlaid on the given databases to classify their construction reactions. There are three REACCS databases: Organic Syntheses, Theilheimer, and the Current Literature File (CLF), all documented alike.<sup>4</sup> The data in these files contain connectivity tables for both reactant and product structures. In most entries there is also a complete mapping table which identifies each reactant atom with its counterpart atom in the product. It is this correlation that allows a direct tabulation of the bonds made and broken in a reaction and so the assignment of the changes on each skeletal atom required by the above treatment. The databases were originally compiled without this correlation, but an extensive program for automatic atom-atom mapping has since been incorporated.<sup>5</sup> For a number of reasons this automatic mapping is incomplete in some entries, and these are inaccessible to our search; the reasons for nonmapping are discussed in ref 5. For our purposes it is sufficient that *all* the skeletal atoms in the *product* be mapped to the reactants. A *complete* mapping of all atoms in the reactant is then not required. Thus, if, for example, a tosylate in the reactant is a leaving group and its atoms do not appear in the product, this will not affect the classification. However, if some reagent, annotated only over the arrow (cf. ethyllithium), delivers skeletal atoms to the product, these will not be found to be

mapped to the reactant, and classification is impossible.

In the initial indexing for construction reactions, we first sorted out those mapped entries in the REACCS files in which only one C-C  $\sigma$ -bond is made. For each such entry all the  $\alpha\beta\gamma$  strands out from each end of the formed bond are identified in reactant and product, and the one strand at each end which changes is selected. For each such strand, the atom attachments on the  $\alpha$ ,  $\beta$ , and  $\gamma$  are then converted into  $z,\pi$  terms and the  $z\pi$ -list of the product subtracted from that of the reactant to create a half-reaction  $\Delta z\pi$ -list identity. If this matches one of the lists for the 25 standard half-reactions, it is so marked; the same procedure is followed for each of the two half-reactions in the full construction.<sup>6</sup>

When any full construction is identified as containing one nucleophilic and one electrophilic half-reaction in our defined list of 25, then the reaction is indexed by entering its registry number into a  $9 \times 16$  matrix of *electrophile*  $\times$  *nucleophile* for rapid retrieval. In this way construction reactions generated by SYNGEN may be quickly matched against constructions of the same definition, or family, from the databases, in order to locate literature precedents for them.

The matrix form of indexing construction reactions represents a cogent and efficient form of storage for fast and accurate retrieval of these reactions. There are about 86 500 reactions in the REACCS databases, of which some 70 500 (about 81%) are adequately mapped for our examination. The missing, unclassifiable reactions are generally those described above with flawed mapping between products and reactants.<sup>5</sup> Of those which are correctly mapped, some 16 500 are single constructions, from which about 12 000 can be classified as two half-reactions and entered in the  $N \times E$  storage matrix.

Of these reactions, the most common nucleophilic half-reactions are the enolates (25%), Grignard reactions (23%), and  $\pi$ -nucleophiles (17%, usually in aromatic substitution). Following these are the enol ether/enamine family (12%), heteroatom-stabilized carbanions (10%), and the Wittig nucleophiles (5%). The remaining eight nucleophilic half-reaction types are present in less than 10% of the cases. Among the electrophilic half-reactions the most common by far are the carbonyl additions without subsequent elimination of water (27%), with elimination (including Wittig reactions) 15%; then alkylations (16%), acylations (15%), and conjugate additions (11%). The remaining electrophiles were only 15% of the cases. These frequencies create some perspective on the nature of the widely used construction reactions in synthetic chemistry.

The constructions which are not classified into the matrix are those with at least one half-reaction which does not fit any of the standard 25 half-reaction  $z\pi$ -lists. About 500 of these exhibit changes on more than one  $\beta\gamma$  strand, usually involving concurrent changes unrelated to the construction itself, as with attendant aromatization. With the other 4000 that are not classified, one of the two half-reactions is almost always a standard one. Indeed in the most common nonclassifiable constructions both half-reactions are standard ones, but both are nucleophilic; these are oxidative couplings and are now not accepted in SYNGEN but they can easily be incorporated.

Some other composite half-reactions also turned up in this nonclassifiable group; some of these will also be added to the set of those used in SYNGEN to generate syntheses.<sup>7</sup> Examples include construction (RZ) to a carboxyl group derivative ( $z = 3$ ) followed by reduction (HZ) to  $z = 1$ , or RH reaction followed by oxidation. Many of these are characterized by subsequent refunctionalization attendant on aromatizing a product; a common case is that of a keto form changing to enol after construction. Some others involve a reactive strand of *four* carbons (generally as a 1,4-addition to a diene, cf. RII.III.III.ZII); very few exhibit strands longer than four



carbons.<sup>6</sup> Overall even these "nonclassifiable" half-reactions are usually found on inspection to be no more complex than just two unit reactions together, i.e., composites not in our standard list.

If we regard skeletal atoms as including atoms other than carbon, then the "construction" of skeletal C-N, C-S, and C-O bonds may also be examined in the same way. We limited these to heteroatoms which show at least two  $\sigma$ -bonds to carbon in the product skeleton, the others being regarded simply as functional groups. Their inclusion as skeletal is possible because the net structural change in constructions to these heteroatoms is identical with that on carbon as long as they are characterized with the same  $h$ ,  $\sigma$ ,  $\pi$ , and  $z$  values as defined for carbon.<sup>3</sup> Each of these heteroatom "constructions" is then entered in its own separate matrix (for C-N, C-S, and C-O, with totals of 4200, 900, and 1500, respectively). These same reactions will be found again when refunctionalizations are discussed below, since in that definition only carbon atoms are accepted as skeletal.

### 3. ORGANIZING THE SYNLIB DATABASE

The SYNLIB database<sup>8</sup> is a second large commercial library of reactions, which apparently has little overlap with the collection in REACCS.<sup>9</sup> However, SYNLIB uses a different format from REACCS; it contains a specific connectivity table for the reaction product but not for the reactant. The reactant structure is then generated by applying a "manipulation table", which is basically a record of the changes made by the person entering the reaction in the first place. Since these operations allow the addition and deletion of atoms, the ultimate mapping of reactant to product may be obscured and some atom correspondences lost.

In order to circumvent the atom mapping problem and to use the same interface with our program which was developed above for REACCS, we first created a file (RD-file) of the SYNLIB entries *converted to REACCS format*. Generalized atom types (cf. R, Z, A, X) in SYNLIB had to be replaced by specific atom types before the entries were acceptable to REACCS format: since in our classification system these are generalized again anyway, such replacement causes no problem. The routine provided by REACCS to create and map new reactions automatically into these RD-files was applied to the reaction data contained in SYNLIB. This in effect duplicated the SYNLIB reaction set in the form of mapped REACCS entries. Of the original 58 000 SYNLIB reactions, some 40 000 were found to be adequately mapped and their complete REACCS-type RD-files were stored as a new data set for access from SYNGEN. Following now the same indexing procedure as for REACCS, we located about 10 400 single C-C constructions and could classify 6500 into our  $9 \times 16$  matrix format for retrieval (about 2200 C-Z constructions also). Once classified<sup>6</sup> these reactions can be accessed from SYNGEN without the use of REACCS. From both databases, therefore, we have assembled a very extensive library of over 27 000 accessible construction reactions to use as a precedent base for SYNGEN.

### 4. PRUNING THE MATCHING ENTRIES

For any given full construction, the number of matches in the matrix entry is quite often too large for manageable inspection. In such cases we allow the operator to be selective as to which entries offer the closest precedents by searching for closer structural similarity between the given SYNGEN reaction and the matching examples in the database matrix. Structural similarities are based only on the reactive strands of the reactions under comparison, not on whole structures. This is done by successive queries of identity of each of the two reacting  $\alpha\beta\gamma$  strands, using these structural variables:

inter- vs intramolecular construction  
nature of the  $\alpha\beta\gamma$  atoms (C, N, S, O)  
 $\sigma$ -values of the two  $\alpha$  atoms  
 $z\pi$ -values of the two  $\alpha$  atoms  
presence of a ring incorporating each  $\alpha$  atom  
ring size of an intramolecular construction (cyclization)

Although listed in the default order, these characteristics may be individually selected in any desired order. If the  $\sigma$ ,  $z$ ,  $\pi$ , and ring values for the two  $\alpha$  atoms do not provide sufficient pruning of the possible matches, then the same characteristics may be addressed for the  $\beta$  and  $\gamma$  atoms in the strands as well. The user is first shown the number of matches in the matrix entry for his particular construction reaction. Then after each choice of an identity characteristic, the number of hits remaining is immediately displayed. In this way the operator may proceed until the number of close matches is deemed to be manageable, or he may revert back to a previous selection if he goes too far. When he has assembled an acceptable number of matches corresponding to his choice of criteria, they are then rapidly displayed for inspection, with options for printing as they appear.

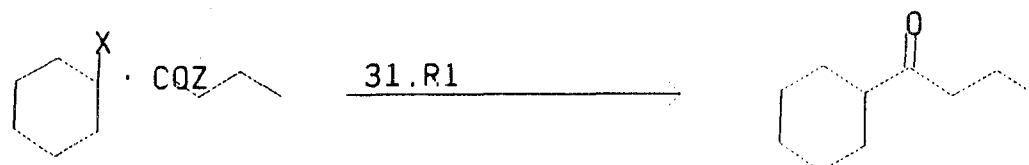
We find this procedure to be very flexible and effective for quickly narrowing the precedents before they are viewed according to the desired pruning criteria. Unlike other search systems for these databases, the preorganization of the matrix and the efficiency of the successive pruning steps make this method very fast; furthermore, the user sets his own criteria for matching and so knows exactly the structural nature of the precedents found. An example from SYNGEN is shown at the top of Figure 2; this construction turned up 157 matches in the CLF database, but the (default) pruning operation reduced this to only 6 examples, of which one is shown in Figure 2 at the bottom.

### 5. REFUNCTIONALIZATION REACTIONS

While the construction reactions are central to synthesis and to SYNGEN, the same reaction characterization will also serve to index all the other reactions. Most of the remaining reactions, by far, are refunctionalizations, which do not alter the skeleton. All possible refunctionalization unit reactions on 1-3 carbons can be easily derived by using the definitions above and excluding any with R changes; these are assembled in Figure 3. There are 14 fundamental categories, all of which fit neatly into familiar chemical classification terms, indicated on Figure 3, such as oxidative addition, reductive elimination, etc.<sup>10</sup> The two-letter symbols introduced above are used to show the changes at each changing carbon, e.g., ZII.ZII for oxidative addition. Simple chemical class abbreviations are also shown in brackets in Figure 3, with R for reduction, X for oxidation, E for elimination, etc. The allylic substitutions, on three carbons, are simply vinyls of the simple substitutions on one carbon, and so are labeled in parallel with primes.

To locate these in the databases, we proceeded as before to annotate the carbons with changing  $z$  and  $\pi$  values, omitting any which show any change in carbon  $\sigma$  values, i.e., skeletal alteration. We set up a code showing the number of carbons with each possible change in  $z$  or  $\pi$  ( $\pm 1$ , 2, or 3) and then counted the number of entries with each code. There are about 48 000 accessible refunctionalizations in the three REACCS databases (with all product carbons mapped). Those reactions which remain after the single constructions above and these refunctionalizations are only the fragmentations (and skeletal rearrangements) and some multiple constructions, numbering together only about 6000 entries.

The results of the reaction count for refunctionalization are presented in Table II, which shows the 21 most frequent reactions, with the number of carbons exhibiting change and the chemical classifications from Figure 3 shown in brackets. A



R1 31

Reacccs regno: 12762

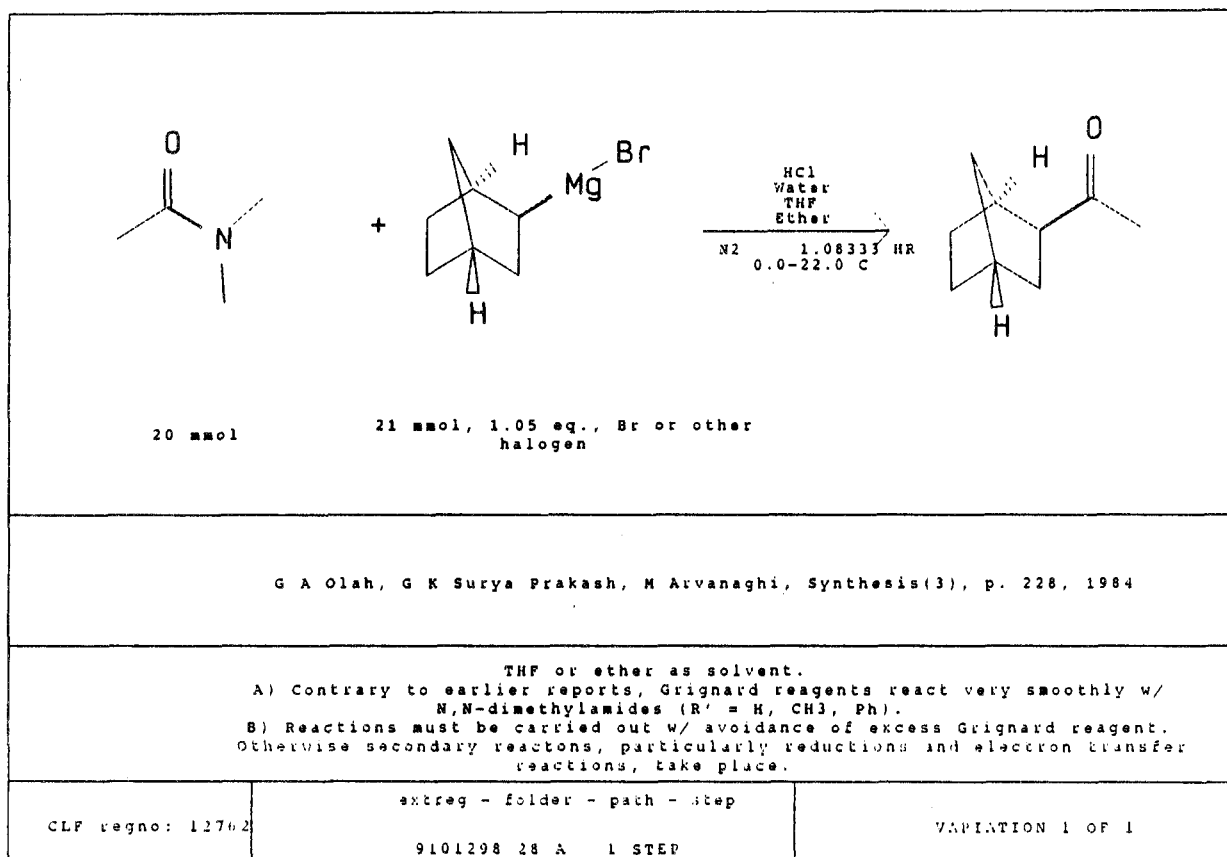


Figure 2. SYNGEN entry (top) and REACCS match.

$\Sigma \Delta x =$	-2	0	+2		
$\Sigma \Delta \pi =$	[RE] ΠZ,ΠZ	[E] ΠH,ΠZ	[XE] ΠH,ΠH	Elimination	(2 atoms)
+2					
	[R] HZ	[S <sub>H</sub> ] HH	[S <sub>Z</sub> ] ZZ	[X] ZH	Simple Substitution (1 atom)
0					
	[R'] HΠ,ΠΠ,ΠZ	[S' <sub>H</sub> ] HΠ,ΠΠ,ΠH	[S' <sub>Z</sub> ] ZΠ,ΠΠ,ΠZ	[X'] ZΠ,ΠΠ,ΠH	Allylic Substitution (3 atoms)
0					
	[RA] HΠ,HΠ	[A] HΠ,ZΠ	[XA] ZΠ,ZΠ	Addition	(2 atoms)
-2					
	Reductive	Isohypsic	Oxidative		

Figure 3. Refunctionalizations on 1-3 skeletal atoms.

summary of the code used to sort them by  $z$  and  $\pi$  changes is also included. The allylic changes are shown with  $\pm 1$  to indicate one carbon gaining and one losing  $z$  or  $\pi$ . The oxidation-state change in each case may be calculated from  $\Sigma \Delta x = 2 \Sigma \Delta z + \Sigma \Delta \pi$ . These 21 reactions represent some 93% of all the reactions in the REACCS databases, and they include all 14 of the simple unit reaction categories of Figure 3. The most common refunctionalizations by far, over half of the total, are the simple substitutions HH and ZZ, labeled [S<sub>H</sub>] and [S<sub>Z</sub>], which show no change at all in the  $z$ - and  $\pi$ -values. These have not yet been further organized in terms of what atom types are being substituted. This is not difficult and will

afford a useful kind of subclass for these many substitutions. With one exception all the reactions in the top 21 are simple in that they involve changes at only one or two carbons. The allylic substitutions only show a change at two carbons ( $\alpha$  and  $\gamma$ ) since the central carbon does not alter its value of  $\pi$  from reactant to product.

After the simple substitutions, the most common reactions in Figure 3 are the simple oxidations and reductions at one carbon, i.e., [X] and [R], followed by the isohypsic eliminations and additions, [E] and [A]. The reactions that are not just the unit reactions of Figure 3 are still simple, being merely combinations of two of the most common unit reactions in one conversion. The first of these is two successive reductions [R + R] at one carbon, as in the Wolff-Kishner reduction of ketones, and the next is the double oxidation at one carbon [X + X], e.g., CH<sub>2</sub>OH → COOH. The next to last is an oxidative addition followed by a further oxidation at one of the carbons, listed as [XA + X]. The rest are combinations of oxidations and reductions on two different carbons, listed as [X][X] and [R][R]. The one four-carbon strand case is just two eliminations, i.e., [E][E], to give two double bonds, as in dehydrohalogenation of a dichloro compound to a diene. The one reaction family which is surprising at first sight is [X][R], oxidation at one carbon and reduction at another. This is actually a disproportionation (coded in Table II as  $\Sigma \Delta z = \pm 1$ ) and involves no redox conditions. The common case

**Table II.** Frequency of Occurrence of Refunctionalization Reactions in REACCS

no. of carbons <sup>a</sup>	$\sum \Delta z^b$	$\sum \Delta \pi^b$	no. of entries	reaction <sup>c</sup>
	0	0	25 083	[S <sub>H</sub> ] and [S <sub>Z</sub> ]
1	+1	0	3 967	[X]
1	-1	0	3 245	[R]
2	-1	+2	2 784	[E]
2	+2	-2	1 847	[XA]
2	+1	-2	1 571	[A]
2	0	-2	944	[RA]
2	$\pm 1$	$\pm 1$	826	[S' <sub>Z</sub> ]
1	-2	0	702	[R + R]
2	-2	+2	619	[RE]
1	+2	0	576	[X + X]
2	0	+2	378	[XE]
2	$\pm 1$	0	353	[X][R]
2	+2	0	346	[X][X]
2	-	$\pm 1$	318	[S' <sub>H</sub> ]
4	-2	+4	262	[E][E]
2	-2	0	254	[R][R]
2	+1	$\pm 1$	191	[X']
2	+3	-2	141	[XA + X]
2	-1	$\pm 1$	138	[R']

<sup>a</sup>Number of carbons changing  $z$  and  $\pi$  values. <sup>b</sup>Total change in  $z$  and  $\pi$  with  $\pm 1$  to indicate plus at one carbon and minus at another. <sup>c</sup>Symbols from Figure 3; brackets include the change at one site.

is one of adjacent carbons for which a common intermediate enol changes the functional groups, as in hydrolysis of epoxide to ketone or aldo to keto sugars.

## 6. SUMMARY

The concept of the unit reaction, or unit exchange of attachments, which underlies the indexing discussed here, was derived initially<sup>10</sup> from theoretical considerations of synthesis logic. It was not clear to what extent real, practical reactions actually were simple unit reactions. The 12 000 constructions classified represent 24 000 half-reactions. Of these about three-quarters (17 500) are simple unit half-reactions; the rest are composites of only two unit reactions, one construction and one refunctionalization. The major composites are the reductions to form carbanions used as construction nucleophiles and the eliminations following carbonyl additions. Unit reactions are an even greater proportion of the 48 000 refunctionalizations sampled in Table II; nearly 42 000 (87%) are

simple unit reactions and of the remaining 6000 about two-thirds (4000) are just composites of two reactions, either sequential at the same site or changes at two sites. Refunctionalizations involving more than four skeletal carbons are less than 1% of the total, represented by examples such as the full hydrogenation of an aromatic ring, and even these are usually recognizably simple multiples of the 14 basic unit reactions.

Overall, our exploration of these databases shows that over 80% of the reactions in them are unit reactions, and most of the others are simple composites or multiples of these.<sup>11</sup> This observation serves to validate the use of this simple reaction description and to make possible an indexing system for all reactions which clarifies their net structural change and accurately affords retrieval by class and subclass, analogous to the Beilstein system for structures. This full database indexing is currently being assembled in our laboratory.

## ACKNOWLEDGMENT

We thank the National Science Foundation both for support of this work and for providing the MicroVAX computer on which it was done, through an instrumentation grant.

## REFERENCES AND NOTES

- (1) Hendrickson, J. B. *Acc. Chem. Res.* **1986**, *19*, 274.
- (2) Hendrickson, J. B. *J. Am. Chem. Soc.* **1971**, *93*, 6847.
- (3) Hendrickson, J. B.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 137.
- (4) Molecular Design Ltd., San Leandro, CA; Wipke, W. T.; Dill, J.; Hounshell, D.; Moock, T.; Grier, D. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower Publishing: Brookfield, VT, 1985.
- (5) Moock, T. E.; Nourse, J. G.; Grier, D. L.; Hounshell, W. D. In *Chemical Structures: The Universal Language of Chemistry*; Proceedings of the CSAC; Springer-Verlag: Berlin, 1988; p 303.
- (6) The identification and classification of the REACCS databases required a total of 12.5 h of cpu time on the MicroVAX 3500 computer; the SYNLIB database used 8 h.
- (7) Hendrickson, J. B.; Miller, T. M. *J. Am. Chem. Soc.* **1991**, in press.
- (8) Distributed by Chemical Graphics, Inc., Meadowbrook, PA; Chodosh, D. F. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower Publishing: Brookfield, VT, 1985.
- (9) Borkent, J. H.; Oukes, F.; Noordik, J. H. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 28.
- (10) Hendrickson, J. B. *J. Chem. Inf. Comput. Sci.* **1979**, *3*, 129.
- (11) Some few conversions of reactant to product will contain multiple unit reactions. The gross overall change may be dissected into a family of all the minimal sequences of unit reactions which amount to the overall conversion, cf.: Hendrickson, J. B.; Braun-Keller, E. *J. Comput. Chem.* **1980**, *3*, 323.