A Simple Code for Improving the Retrieval of Information Associated with Keto-Enol Tautomers*

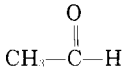
CHARLES H. DAVIS
Graduate Library School, Indiana University, Bloomington, Indiana

Received August 1, 1966

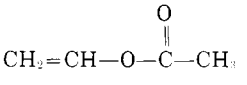
Scattering of information associated with tautomers can occur in a file whether the basic arrangement is by compound name or molecular formula. The "parent compound" concept which is used in the Subject Indexes of *Chemical Abstracts* can be used as the basis for a code which will improve the retrieval of such information. A code based solely on structural considerations may ultimately provide a better system for relating tautomers, but using a well-established nomenclature system is both quick and easy.

"Acetaldehyde" and "vinyl alcohol" are specific examples of names that are used to designate different tautomeric forms of the same compound. These names are used as main headings in the Subject Indexes of *Chemical Abstracts*, which has its own vocabulary control and nomenclature system. Anyone interested in all the information available on this compound must look at both main headings, because the indexing policy of *Chemical Abstracts* states that "such a tautomeric compound is usually indexed as an aldehyde or ketone as the more stable form"; however, it is named as an alcohol "if the enol form is individually mentioned" (*1*). In other words, indexing depends on which form the author of the original paper mentions, and no attempt is made to index the compound in all of its tautomeric forms.

It is true that the various tautomeric forms of a given compound have identical molecular formulas. However, the problem of scattering of information in files and/or indexes based on the alphabetic arrangement of names is not solved by setting up a file based on molecular formulas, because derivatives of these compounds will file at different formulas and the number of possible derivatives is infinite.

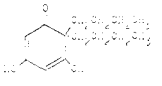


Acetaldehyde
C_2H_4O



Vinyl acetate
$C_4H_6O_2$

Phloroglucinol
$C_6H_6O_3$



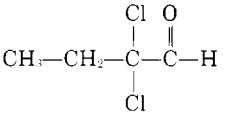
2,4-Cyclohexadiene-1-one,
6,6-dibutyl-3,5-dihydroxy-
$C_{14}H_{22}O_3$

* This paper is based on an M. A. thesis submitted to the Graduate School of Indiana University, August, 1966.

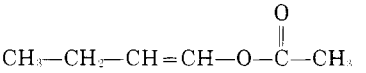
Compounds of the type shown above are related to each other in a way which makes them mutually interesting. That is, a chemist interested in information about one might very well be interested in information about the other. Conventional card files of chemical compounds, which characteristically consist of 3 × 5 index cards arranged either by molecular formula or alphabetically by name, do not lend themselves to the retrieval of such information. Edge-notched cards offer one means by which the storage and retrieval of tautomeric information can be improved, because they can be coded in a manner which is independent of their sequence in the file.

SEMANTIC LEVELS OF TAUTOMERIC INFORMATION

It is not immediately obvious that compounds such as those shown below can be of mutual interest, because they have been stabilized in different tautomeric forms. They can be of mutual interest, however, because butyraldehyde and 1-buten-l-ol are keto-enol tautomers. In other words, the parent compounds (those designated by that part of the chemical name appearing before the comma of inversion) should be considered when establishing whether two or more compounds of this type are mutually interesting.



Butyraldehyde,
2,2-dichloro-



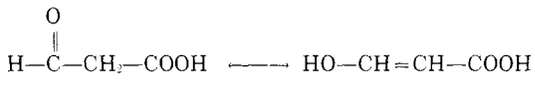
1-Buten-l-ol, acetate

However, when parent compounds are considered instead of specific compounds, then a different level of meaning has been introduced. The names "butyraldehyde" and

"1-buten-l-ol" are really generic to the specific names of the compounds shown above, and they represent a class intermediate to specific tautomers and the class of all keto-enol tautomers.

In using direct coding of tautomeric information as an adjunct to a subject file, this intermediate class should be most useful, because to set up a one-to-one relationship only between specific tautomers would be to ignore the potential mutual interest which exists between the various derivatives of tautomers, and to code directly for the entire class of keto-enol tautomers would plainly make the system too general to be of much use.

In a nomenclature system such as that used by *Chemical Abstracts* in its subject indexes, names are dependent on an order of precedence of functions in the cases where a compound contains more than one functional group (2). Therefore, keto-enol tautomers which contain functional groups higher in the order of precedence than hydroxyl or keto groups should be named as derivatives of the highest function present, but they should be coded as though they were named as alcohols, phenols, aldehydes, or ketones. This will ensure that a one-to-one relationship is established at the proper coordinate level.



Malonaldehydic acid (acetaldehyde, carboxy-)	Acrylic acid, 3-hydroxy- (vinyl alcohol, 2-carboxy-)

Thus, for the compound shown above, the entries should be filed alphabetically at "malonaldehydic acid" and "acrylic acid, 3-hydroxy-" but should be assigned the code for the "vinyl alcohol... acetaldehyde" tautomeric system. This enables a person interested in derivatives of acetaldehyde and vinyl alcohol to find them without examining "malonaldehydic acid," "acrylic acid," and all the other headings of all the functions higher in the order of precedence than aldehydes or alcohols. Naturally, ketones and phenols can be treated in a completely analogous manner.

CODING

Since the object of coding is simply to establish a one-to-one relationship between different tautomeric forms of the same parent compound, the most straightforward approach using edge-notched cards would be to assign one unique hole to each distinct class of tautomers. By assigning a different hole to each class, the operator of the system establishes a unique relationship between the tautomers so that only one needle-sort is required to obtain all the information in the card file about a given class. Cards exist which have as many as 172 holes, arranged in four rows of 43 each, on one edge of each card (3). However, if this number is too small to represent the total number of primary tautomeric systems in the file, then a more efficient code from the standpoint of space should be used, such as the 7-4-2-1 code described by Bourne (4).

Coding should be done systematically in order to avoid using the same code for more than one primary tautomeric system. It would not be necessary to keep a complete record of compound names, numbers, etc., because a simple tally of the number of classes which had been coded would reveal which code numbers had already been used. The numbers should, of course, be assigned sequentially. In practice, the user would simply examine the card file in order to establish the code number for a given tautomeric system.

Before adding a new card to the file, the operator should examine the structure to determine which class of tautomers is represented, if any, and then he should check the file to see whether representatives of this class are already present. If they are, then the incoming card should be assigned the code of the cards already in the file. Otherwise, a new code number must be assigned. It is possible that the file will contain no entries for a given tautomeric form—that is, a new entry for the file might have tautomeric forms which are not already represented by cards. In such a case, the operator should supply a "dummy" card having the name of the parent compound and should code this card so that it corresponds to the code of the incoming card at the other tautomeric form. These "dummy" cards could be removed and discarded whenever a "real" entry was made at the heading involved.

DATA COLLECTION

The Volume 57 Subject Index of *Chemical Abstracts* was used as the source of data for a study of the feasibility of this method. Only compounds capable of keto-enol tautomerism were selected, and the entire index was scanned so that the compound names would be well distributed throughout the alphabet. In addition, names were chosen so that a variety of structural types would be obtained; and also so that examples of aldehydes, ketones, alcohols, and phenols would all be included whether they were named as such or not. For the sake of simplicity, the subheadings or "modifications" were ignored, because they would have little or no bearing on the feasibility of the method. Sixty-five compound names were selected in this manner, and as each was chosen, its structure was drawn and a code number was assigned on the basis of the primary tautomeric system present—*i.e.*, on the basis of the parent compounds present in each tautomeric system. If the name of a tautomeric structure representing a compound selected did not appear in the index, it was supplied in upper case type. There were 60 of these names.

Table I lists the names alphabetically as they would appear in a subject index or card file, with the code number of the primary tautomeric system opposite each name. Table II lists the names alphabetically opposite the numbers which are the codes for the primary tautomeric systems present in those structures. In other words, Table II shows what information would be retrieved by hypothetical needle-sorts of Table I. Table III shows the structures involved in each primary tautomeric system opposite the code number for that system.

CONCLUSION

Examination of the system numbers in Table I will reveal the extent to which the tautomeric information

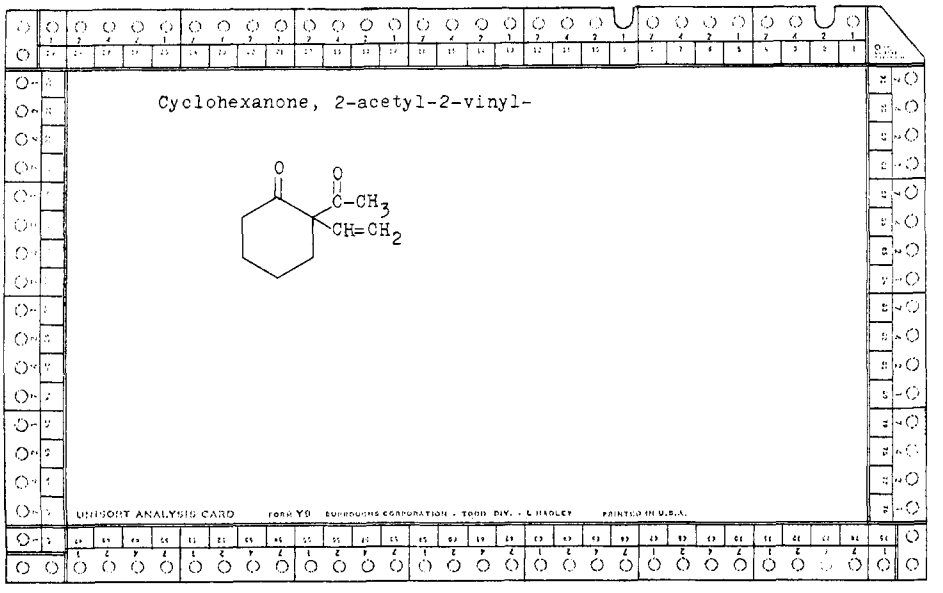


Figure 1. Edge-notched card coded for two tautomeric systems.

Table I.

Compound Name	System Number
1-Acenaphthenone, 3-*sec*-butyl-	1
1-ACENAPHTHYLENOL, 3-*sec*-BUTYL-	1
Acetaldehyde	2
Acetamide, *N*-(3-formylpropyl)-	3
ACETAMIDE, *N*-(4-HYDROXY-3-BUTENYL)-	3
Acetoacetic acid	4
Acetone	4
Acetophenone	2
ACRYLONITRILE, 2-HYDROXY-	2
9-Anthrol	5
Anthrone	5
Benzoic acid, *m*-hydroxy-	8
Benzonitrile, 3-acetyl-5-nitro-	2
BENZONITRILE, 3-(1-HYDROXYVINYL)-5-NITRO-	2
2*H*-1-BENZOPYRAN-4-OL, 2-PHENYL-	13
BENZYL ALCOHOL, α-BUTYLIDENE-	16
Benzyl alcohol, α-methylene-	2
BENZYL ALCOHOL, α-PROPYLIDENE-	3
1,3-Butadiene-1,4-diol, 2-*tert*-butyl-	6
1-Butanone, 4-chloro-1-(2-thienyl)-	6
2-Butanone	20
3-BUTENAL, 2-*tert*-BUTYL-4-HYDROXY-	6
3-BUTENAL, 3-*tert*-BUTYL-4-HYDROXY-	6
3-BUTENOIC ACID, 4-CYCLOPENTADIENYL-4-HYDROXY-2-METHYL-	3
3-BUTENOIC ACID, 4-HYDROXY-	17
1-BUTEN-1-OL	3
1-BUTEN-2-OL	20
2-BUTEN-2-OL	20
Butyraldehyde	3
Butyric acid, 2-oxo-	17
Butyrophenone	3
CINNAMIC ACID, α-HYDROXY-β-METHYL-*o*-NITRO-	17
CROTONIC ACID, 2-HYDROXY-	17
Crotonic acid, 3-hydroxy-	4
Cyclobutaneacetic acid, 2,2-dimethyl-3-oxo-	7
Cyclobutanone	7
2-CYCLOBUTENE-1-ACETIC ACID, 3-HYDROXY-4,4-DIMETHYL-	7
1-CYCLOBUTEN-1-OL	7
2,4-Cyclohexadien-1-one	8
2,4-CYCLOHEXADIEN-1-ONE, 2,3-DIMETHYL-	8
2,4-CYCLOHEXADIEN-1-ONE, 5,6-DIMETHYL-	8
Cyclohexanone, 2-acetyl-2-vinyl-	2, 9
CYCLOHEXANONE, 2-(1-HYDROXYVINYL)-2-VINYL-	2, 9
1-CYCLOHEXEN-1-OL, 6-(1-HYDROXYVINYL)-6-VINYL-	2, 9
Cyclopentadienebutyric acid, α-methyl-γ-oxo-	3
Cyclopentaneheptanoic acid, 3-hydroxy-2-(3-hydroxy-1-octenyl)-5-oxo-	10
Cyclopentanone	10
1-CYCLOPENTENE-1-HEPTANOIC ACID, 2,4-DIHYDROXY-5-(3-HYDROXY-1-OCTENYL)-	10
2-CYCLOPENTENE-1-HEPTANOIC ACID, 2,4-DIHYDROXY-5-(3-HYDROXY-1-OCTENYL)-	10
1-Cyclopenten-1-ol	10
Decanal, 10-hydroxy-	11
1-DECENE-1,10-DIOL	11
1-DECEN-1-OL	11
1,4-Dioxaspiro[4.5]decan-8-one	12
1,4-DIOXASPIRO[4.5]DEC-7-EN-8-OL	12
1-DODECEN-1-OL	19
1,2-Ethenediol	2

Table I. (*Continued*).

Compound Name	System Number
Flavanone	13
Fumaric acid, hydroxy-	2
3-FURANOL, 5-METHYL-	14
3(2*H*)-Furanone, 5-methyl-	14
Glycolaldehyde	2
1,6-Heptadien-4-one	15
4-Heptanone	15
1,3,6-HEPTATRIEN-4-OL	15
3-HEPTEN-4-OL	15
Hexanedioic acid, 2-oxo-4-phenyl-	3
Hexanoic acid, 2-oxo-	16
3-Hexanone	22
2-HEXENEDIOIC ACID, 2-HYDROXY-4-PHENYL-	3
2-HEXENOIC ACID, 2-HYDROXY-	16
2-HEXEN-3-OL	22
3-HEXEN-3-OL	22
HYDROCINNAMIC ACID, β-[2-HYDROXY-4-(1-NAPHTHYL)-1-BUTENYL]-	22
HYDROCINNAMIC ACID, β-[2-HYDROXY-4-(1-NAPHTHYL)-2-BUTENYL]-	22
Hydrocinnamic acid, β-methyl-*o*-nitro-α-oxo-	17
1,3-Indandione, 2-phenyl-	18
INDONE, 3-HYDROXY-2-PHENYL-	18
KETONE, 2-HYDROXY-1-VINYL-2-CYCLOHEXEN-1-YL METHYL	2, 9
Lauraldehyde	19
Levulinic acid	20
Maleic acid, hydroxy-	2
Malonic acid, (3-hydroxy-2-heptenylidene)-	21
MALONIC ACID, (3-HYDROXY-3-HEPTENYLIDENE)-	21
MALONIC ACID, (3-OXOHEPTYLIDENE)-	21
1-Naphthaleneheptanoic acid, δ-oxo-β-phenyl-	22
2-Norbornanecarbonitrile, 2-acetonyl-3,3-dimethyl-	4
2-NORBORNANECARBONITRILE, 2-(2-HYDROXYALLYL)-3,3-DIMETHYL-	4
2-NORBORNANECARBONITRILE, 2-(2-HYDROXYPROPENYL)-3,3-DIMETHYL-	4
1,3-OCTADIEN-4-OL	21
1,4-OCTADIEN-4-OL	21
Octanoic acid, 8-(4-methoxy-*m*-tolyl)-5-oxo-	15
4-OCTANONE	21
4-OCTENOIC ACID, 5-HYDROXY-8-(4-METHOXY-*m*-TOLYL)-	15
5-OCTENOIC ACID, 5-HYDROXY-8-(4-METHOXY-*m*-TOLYL)-	15
3-OCTEN-4-OL	21
4-OCTEN-4-OL	21
1-OCTEN-4-ONE	21
Oxalacetic acid	2
1-Pentalenol, octahydro-	23
1(2*H*)-Pentalenone, hexahydro-	23
3-Pentenoic acid, 4-hydroxy-	20
4-PENTENOIC ACID, 4-HYDROXY-	20
1-PENTEN-1-OL	16
Phenol	8
PHTHALIMIDE, *N*-(2-HYDROXY-1-CYCLOHEXEN-1-YL)-	9
PHTHALIMIDE, *N*-(2-HYDROXY-2-CYCLOHEXEN-1-YL)-	9
Phthalimide, *N*-(2-oxocyclohexyl)-	9
1-Propen-1-ol	17
1-Propen-2-ol	4
Propionaldehyde	17
Pyruvonitrile	2
Salicylic acid	8
Spiro[2.4]heptan-4-one	24
SPIRO[2.4]HEPT-4-EN-4-OL	24
SUCCINALDEHYDE, 2-*tert*-BUTYL-	6
Succinaldehydic acid	17
2-THIOPHENEMETHANOL, α-(3-CHLOROPROPYLIDENE)-	3
Valeraldehyde	16
Valerophenone	16
Vinyl alcohol	2
2,3-Xylenol	8
2-Undecanone	25
1-UNDECEN-2-OL	25
2-UNDECEN-2-OL	25

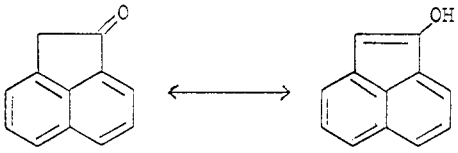
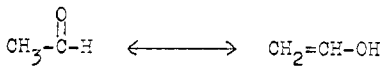
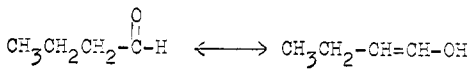
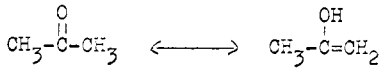
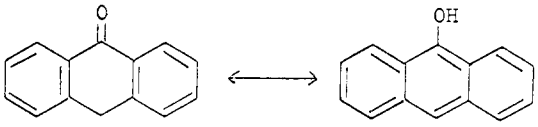
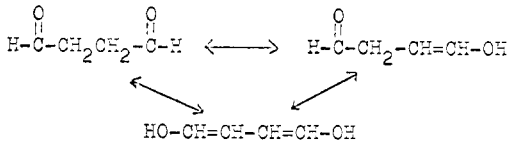
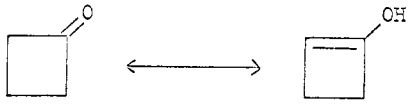
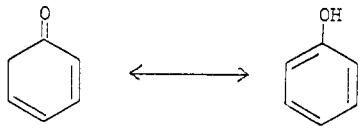
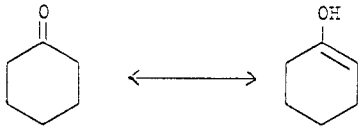
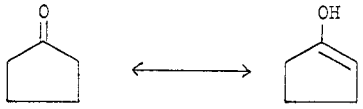
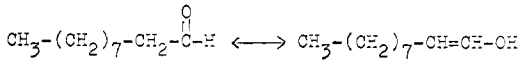
Table II.

System Number	Compound Name
1---	1-Acenaphthenone, 3-*sec*-butyl-
	1-ACENAPHTHYLENOL, 3-*sec*-BUTYL-
2---	Acetaldehyde
	Acetophenone
	ACRYLONITRILE, 2-HYDROXY-
	Benzonitrile, 3-acetyl-5-nitro-
	BENZONITRILE, 3-(1-HYDROXYVINYL)-5-NITRO-
	Benzyl alcohol, α-methylene-
	Cyclohexanone, 2-acetyl-2-vinyl-
	CYCLOHEXANONE, 2-(1-HYDROXYVINYL)-2-VINYL-
	1-CYCLOHEXEN-1-OL, 6-(1-HYDROXYVINYL)-6-VINYL-
	1,2-Ethenediol
	Fumaric acid, hydroxy-
	Glycolaldehyde
	KETONE, 2-HYDROXY-1-VINYL-2-CYCLOHEXEN-1-YL METHYL
	Maleic acid, hydroxy-
	Oxalacetic acid
	Pyruvonitrile
	Vinyl alcohol

Table II. (*Continued*)

System Number	Compound Name
3---	Acetamide, *N*-(3-formylpropyl)-
	ACETAMIDE, *N*-(4-HYDROXY-3-BUTENYL)-
	BENZYL ALCOHOL, α-PROPYLIDENE-
	1-Butanone, 4-chloro-1-(2-thienyl)-
	3-BUTENOIC ACID, 4-CYCLOPENTADIENYL-4-HYDROXY-2-METHYL
	1-BUTEN-1-OL
	Butyraldehyde
	Butyrophenone
	Cyclopentadienebutyric acid, α-methyl-γ-oxo-
	Hexanedioic acid, 2-oxo-4-phenyl-
	2-HEXENEDIOIC ACID, 2-HYDROXY-4-PHENYL-
	2-THIOPHENEMETHANOL, α-(3-CHLOROPROPYLIDENE)-
4---	Acetoacetic acid
	Acetone
	Crotonic acid, 3-hydroxy-
	2-Norbornanecarbonitrile, 2-acetonyl-3,3-dimethyl-
	2-NORBORNANECARBONITRILE, 2-(2-HYDROXYALLYL)-3,3-DIMETHYL-
	2-NORBORNANECARBONITRILE, 2-(2-HYDROXYPROPENYL)-3,3-DIMETHYL-
	1-Propen-2-ol
5---	9-Anthrol
	Anthrone
6---	1,3-Butadiene-1,4-diol, 2-*tert*-butyl-
	3-BUTENAL, 2-*tert*-BUTYL-4-HYDROXY-
	3-BUTENAL, 3-*tert*-BUTYL-4-HYDROXY-
	SUCCINALDEHYDE, 2-*tert*-BUTYL-
7---	Cyclobutaneacetic acid, 2,2-dimethyl-3-oxo-
	Cyclobutanone
	2-CYCLOBUTENE-1-ACETIC ACID, 3-HYDROXY-4,4-DIMETHYL-
	1-CYCLOBUTEN-1-OL
8---	Benzoic acid, *m*-hydroxy-
	2,4-Cyclohexadien-1-one
	2,4-CYCLOHEXADIEN-1-ONE, 2,3-DIMETHYL-
	2,4-CYCLOHEXADIEN-1-ONE, 5,6-DIMETHYL-
	Phenol
	Salicylic acid
	2,3-Xylenol
9---	Cyclohexanone, 2-acetyl-2-vinyl-
	CYCLOHEXANONE, 2-(1-HYDROXYVINYL)-2-VINYL-
	1-CYCLOHEXEN-1-OL, 6-(1-HYDROXYVINYL)-6-VINYL-
	KETONE, 2-HYDROXY-1-VINYL-2-CYCLOHEXEN-1-YL METHYL
	PHTHALIMIDE, *N*-(2-HYDROXY-1-CYCLOHEXEN-1-YL)-
	PHTHALIMIDE, *N*-(2-HYDROXY-2-CYCLOHEXEN-1-YL)-
	Phthalimide, *N*-(2-oxocyclohexyl)-
10---	Cyclopentaneheptanoic acid, 3-hydroxy-2-(3-hydroxy-1-octenyl)-5-oxo-
	Cyclopentanone
	1-CYCLOPENTENE-1-HEPTANOIC ACID, 2,4-DIHYDROXY-5-(3-HYDROXY-1-OCTENYL)-
	2-CYCLOPENTENE-1-HEPTANOIC ACID, 2,4-DIHYDROXY-5-(3-HYDROXY-1-OCTENYL)-
	1-Cyclopenten-1-ol
11---	Decanal, 10-hydroxy-
	1-DECEN-1,10-DIOL
	1-DECEN-1-OL
12---	1,4-Dioxaspiro[4.5]decan-8-one
	1,4-DIOXASPIRO[4.5]DEC-7-EN-8-OL
13---	2*H*-1-BENZOPYRAN-4-OL, 2-PHENYL-
	Flavanone
14---	3-FURANOL, 5-METHYL-
	3(2*H*)-Furanone, 5-methyl-
15---	1,6-Heptadien-4-one
	4-Heptanone
	1,3,6-HEPTATRIEN-4-OL
	3-HEPTEN-4-OL
	Octanoic acid, 8-(4-methoxy-*m*-tolyl)-5-oxo-
	4-OCTENOIC ACID, 5-HYDROXY-8-(4-METHOXY-*m*-TOLYL)-
	5-OCTENOIC ACID, 5-HYDROXY-8-(4-METHOXY-*m*-TOLYL)-
16---	BENZYL ALCOHOL, α-BUTYLIDENE-
	Hexanoic acid, 2-oxo-
	2-HEXENOIC ACID, 2-HYDROXY-
	1-PENTEN-1-OL
	Valeraldehyde
	Valerophenone
17---	3-BUTENOIC ACID, 4-HYDROXY-
	Butyric acid, 2-oxo-
	CINNAMIC ACID, α-HYDROXY-β-METHYL-*o*-NITRO-
	CROTONIC ACID, 2-HYDROXY-
	Hydrocinnamic acid, β-methyl-*o*-nitro-α-oxo-
	1-Propen-1-ol
	Propionaldehyde
	Succinaldehydic acid
18---	1,3-Indandione, 2-phenyl-
	INDONE, 3-HYDROXY-2-PHENYL-
19---	1-DODECEN-1-OL
	Lauraldehyde
20---	2-Butanone
	1-BUTEN-2-OL
	2-BUTEN-2-OL
	Levulinic acid
	3-Pentenoic acid, 4-hydroxy-
	4-PENTENOIC ACID, 4-HYDROXY-
21---	Malonic acid, (3-hydroxy-2-heptenylidene)-
	MALONIC ACID, (3-HYDROXY-3-HEPTENYLIDENE)-
	MALONIC ACID, (3-OXOHEPTYLIDENE)-
	1,3-OCTADIEN-4-OL
	1,4-OCTADIEN-4-OL
	4-OCTANONE
	3-OCTEN-4-OL
	4-OCTEN-4-OL
	1-OCTEN-4-ONE
22---	3-Hexanone
	2-HEXEN-3-OL
	3-HEXEN-3-OL
	HYDROCINNAMIC ACID, β-[2-HYDROXY-4-(1-NAPHTHYL)-1-BUTENYL]-
	HYDROCINNAMIC ACID, β-[2-HYDROXY-4-(1-NAPHTHYL)-2-BUTENYL]-
	1-Naphthaleneheptanoic acid, δ-oxo-β-phenyl-
23---	1-Pentalenol, octahydro-
	1(2*H*)-Pentalenone, hexahydro-
24---	Spiro[2.4]heptan-4-one
	SPIRO[2.4]HEPT-4-EN-4-OL
25---	2-Undecanone
	1-UNDECEN-2-OL
	2-UNDECEN-2-OL

Table III.

System Number	Primary Tautomeric System
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

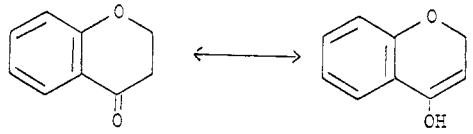
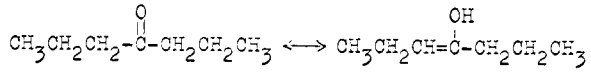
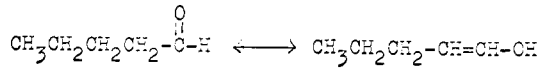
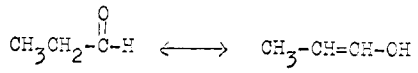
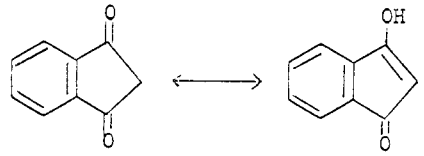
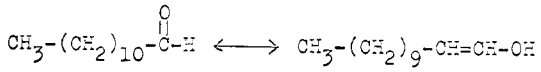
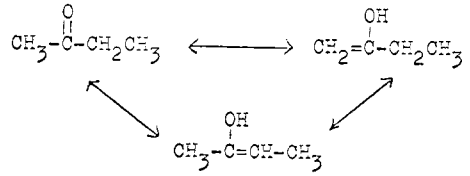
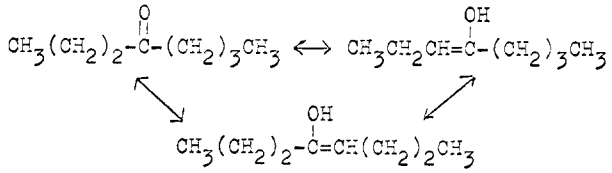
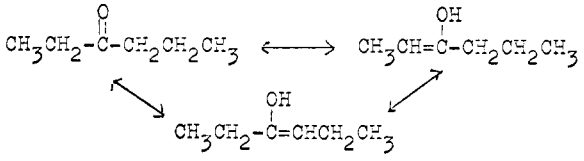
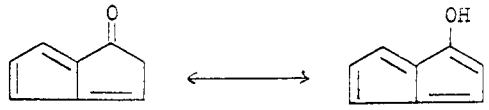
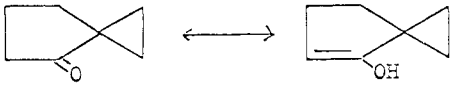
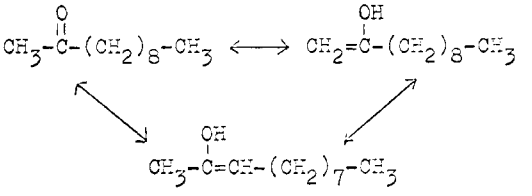
System Number	Primary Tautomeric System
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	

Table III. (*Continued*)

System Number	Primary Tautomeric System
24	
25	

has been scattered. In a larger file, which would also normally contain many compounds other than keto-enol tautomers, this scattering would be even worse.

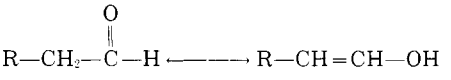
It is not possible to determine the percentage of compounds which are keto-enol tautomers in a given subject index without examining all of the compounds which have been indexed, and since there are thousands of compounds in an index such as the Volume 57 Subject Index of *Chemical Abstracts*, this would clearly be a monumental task. The percentage of these compounds is sufficiently high, however, to warrant the investigation of a system which improves the retrieval of tautomeric information, and in any case it would be necessary from the standpoint of thoroughness.

For some compounds, this system results in what can be called a nontraditional association with the aldehyde structure. In Table II under system number 2, compounds are listed which contain the acetaldehyde. . . vinyl alcohol tautomeric system. "Acetophenone" and "benzyl alcohol, α-methylene-" are names designating structures which fall into this category, and in fact, they represent different tautomeric forms of the same compound. The nontraditional association in this case is between acetophenone and acetaldehyde, because acetophenone is a ketone. This is really an example of a pseudo-problem, since the two compounds are clearly related structurally regardless of the generic classes into which they fall from the standpoint of nomenclature.

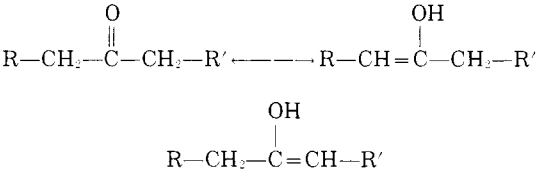
Another consideration is the case in which one compound contains more than one tautomeric system. The compound "cyclohexanone, 2-acetyl-2-vinyl-" is an example of such a case. Figure 1 shows the name and the structure of this compound as they would appear on an edge-notched card. Holes two and nine have been notched to correspond to the primary tautomeric systems present in this structure. If a more complicated code is used, which requires more than one hole per tautomeric system, then it would probably be more advisable to make duplicate cards which could each be coded for a different tautomeric system present in that particular compound.

The foregoing considerations are important from the standpoint of coding the information being put into the file, but they do not in any way hamper the actual operation of the file, which in itself is extremely simple. The cards should be arranged alphabetically as shown in Table I so that filing remains a straightforward clerical operation. Whenever information associated with a particular tautomeric system is desired, a single needle-sort will provide it, thus obviating the necessity of searching through the entire file.

Ultimately, a code based solely on structural considerations will probably provide a more satisfactory system for relating tautomers. One possibility would be to adopt a more generic approach than has been suggested in this paper. For example,



could be used to include system numbers 2, 3, 11, 16, 17, and 19;



could be used to include system numbers 4, 15, 20, 21, 22, and 25; and ring structures could also be generalized by using an appropriate system number or descriptor. It should be kept in mind, however, that a more generic approach is not necessarily better and that the needs of the users should always be considered when formulating any such system.

Recently, the Chemical Abstracts Service has developed a technique for generating a unique machine description for chemical structures (5). This technique is being used to establish the computer-based file of chemical structures known as the Chemical Abstracts Registry System. Since it is the structure which is being used for input rather than the compound, scattering of information associated with keto-enol tautomers can also occur in this system unless an algorithm is written to tie it together. It is suggested that the method of coding described in this paper could be used as the basis for such an algorithm.

LITERATURE CITED

(1) Chemical Abstracts Service, "The Naming and Indexing of Chemical Compounds," p. 33N, § 195, American Chemical Society, Easton, Pa., 1962.
(2) *Ibid.*, p. 12N, § 83.
(3) Bourne, Charles P., "Methods of Information Handling," p. 100, John Wiley & Sons, New York, N. Y., 1963.
(4) *Ibid.*, p. 85.
(5) Morgan, H. L., *J. Chem. Doc.* **5**, 107 (1965).