same number of internal third-degree carbon atoms (N_{Ic}) and disconnections (d_s) of the internal bonds, respectively. Graph theoretical concepts of vertex degree (degree of carbon atom), edges (σ C–C bonds), and trees (formed by the internal edges of PAH) have been uniquely employed in this description of the structure-formula related properties of PAH's.

Graph-theoretically related formulas have been derived or empirically noted for the first time for PAH's. From the formula of a PAH, the number of rings ($r = (1/2)[N_c + 2 - N_H]$) and σ C–C bonds ($q = (1/2)[3N_c - N_H]$) can be determined. Also, the range of possible peripheral and internal carbon atoms (N_{Pc} and N_{Ic}), peripheral and internal σ bonds (q_p and q_I), and the number of five-membered rings (r_5) that are associated with the various related structures of a specific PAH formula can be directly determined.

Finally, a method for determining whether a specific C/H composition can be represented by a PAH6 structure has been presented. A systematic approach for the computer enumeration of all the structures of PAH6 has been proposed. Although, eq 2, 4, and 5 are presented without proofs (in the

mathematical sense), it is believed that they will have general application in mathematical graph theory, and additional results emanating from their utility will be forthcoming.

REFERENCES AND NOTES

(1) Balaban, A. T. *Rev. Roum. Chim.* **1966,** *11,* 1097.
(2) Balaban, A. T; Harary, F. *Tetrahedron* **1968,** *24,* 2505.
(3) Balaban, A. T. "Chemical Application of Graph Theory"; Academic Press: New York, 1976; Chapter 5.
(4) Trinajstić, N., work in progress.
(5) (a) Harary, F. "A Seminar on Graph Theory"; Holt, Rinehart, and Winston: San Francisco, CA, 1967. (b) "Graph Theory"; Addison-Wesley Publishing Co.: Reading, MA, 1969.
(6) Jenny, W.; Peter, R. *Angew. Chem.* **1965,** *77,* 1027.
(7) Flammang-Barbieux, M.; Nasielski, J.; Martin, R. H. *Tetrahedron Lett.* **1967,** 743.

The Chemical Abstracts Service Chemical Registry System. 9. Input Structure Conventions

ANNE W. RYAN* and ROBERT E. STOBAUGH

Chemical Abstracts Service, Columbus, Ohio 43210

The Chemical Abstracts Service (CAS) Chemical Registry System is a computer-based system that uniquely identifies chemical substances on the basis of composition and structure. The structure conventions for substances registered in the Chemical Registry System have been described. The types of substances are fully defined organic and inorganic compounds, ions, polymers, coordination compounds, alloys, mixtures, and certain partially defined substances.

INTRODUCTION

Chemical Abstracts Service, in its mission of covering the world's primary chemical and chemical engineering literature, has from the beginning been involved with documenting and indexing information about chemical substances. A particular problem has been the identification of reoccurrences of substances in scientific literature. Such identification is necessary to assure consistency in the indexes to *Chemical Abstracts.* These indexes are based on a systematic chemical nomenclature, and the technique used for many years to assure consistency was to name each indexed substance every time it was cited and to file the name alphabetically in an index with its citations. This resulted in a large amount of redundant effort in repeated renamings. To remedy this situation, the CAS Chemical Registry System was developed in the early 1960s. The main objective was the establishment of a computer-based system to identify chemical substances uniquely on the basis of molecular structure. The chemical substances involved comprise the whole of chemistry in that all classes are handled—organic and inorganic, completely and partly described. The design, content, function, and some special features of the Chemical Registry System, including both machine and manual registration processes, have been described in detail in previous papers.[1-8] In the development of the Registry System, the structural diagram rather than nomenclature was chosen as the means for recording substances because in actual usage names vary greatly in consistency, standardization, systematization and the language involved. On the other hand, structural representation is standardized to such an extent that structural diagrams are virtually the same on a worldwide basis, even when different languages and

alphabets are involved. The techniques and mechanics of typed structure input have been described,[5] and in previous Registry papers[1-8] structure conventions have been alluded to and illustrated. However, these conventions have not been treated specifically, and it is the purpose of this paper to do so.

Since the Chemical Structure File has become available as the basis for the new service CAS ONLINE, the knowledge of how structures are represented for input will enable users to phrase queries in a way to take maximum advantage of particular structural characteristics. While the majority of use of CAS ONLINE can be effective without knowledge of many of the details described here, more retrieval selectivity is possible with more intellectual specificity. Also, in order for users to take full advantage of the CAS Private Registry Service, structural information should be represented in conformance with conventions established for the Chemical Registry System. General structuring conventions will be presented first, followed by those for specialized classes.

GENERAL STRUCTURING CONVENTIONS

The conventions used in preparing structural diagrams, especially of organic compounds, are based on those exemplified in chemical literature. The primary aim has been to allow as much flexibility as possible, keeping the number of arbitrary conventions to a minimum. In cases where more than one method of representation is in current use, a single method has been chosen, for example, for salts, polymers, coordination compounds, and incompletely defined substances. Where it has been possible to recognize certain variations in representation within the registration programs, it has been done, especially for alternating single–double bond and tautomeric
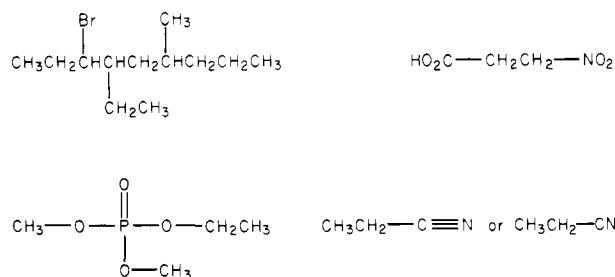
CAS CHEMICAL REGISTRY SYSTEM

J. Chem. Inf. Comput. Sci., Vol. 22, No. 1, 1982 **23**

CH₃CH₂CHCHCH₂CHCH₂CH₂CH₃ (with Br, CH₃, CH₂CH₃ substituents)

$HO_2C-CH_2CH_2-NO_2$

$CH_3-O-P-O-CH_2CH_3$ (with O double bond and O—CH₃)

$CH_3CH_2-C\equiv N$ or CH_3CH_2-CN

Figure 1. Acyclic structures.

Figure 2. Cyclic structures.

Figure 3. Input conventions for charges and isotopes.
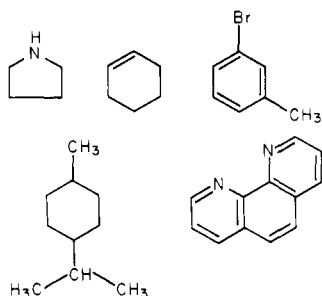
situations.[7] Atoms and bonds are generally drawn explicitly, but in some specific situations, methods are used to simplify the structuring technique, such as group symbols or implied bonds between carbon and hydrogen. Information about stereochemical configuration,[3] isotopic content, and molecular formula[1] is also provided where applicable.

For acyclic structures each atom is shown explicitly. While single bonds may or may not be shown, carbon–carbon double and triple bonds are always represented. Multiple bonds between carbon and a noncarbon atom are shown explicitly in some cases and not in others, which are represented in short-cut notation, as will be discussed. The point of attachment for substituents is shown as clearly as possible. The machine representation of these structures is, as previously described,[1] a detailed and specific atom-bond inventory or connection table. Examples of acyclic structures are shown in Figure 1.

For cyclic structures, the carbon atoms in the ring and their attached hydrogen atoms are implied by the angled bond lines, but all noncarbon atoms and any attached hydrogen atoms are explicitly shown. All multiple bonds are shown, and benzenoid rings are drawn with alternating single and double bonds. Acyclic substituents are represented in accordance with the above conventions for acyclic compounds. The machine representation[1] is that of an atom-bond connection table. Examples of cyclic structures are shown in Figure 2.

A charge is indicated by a superscript "+" (plus) or "−" (minus) to the right, if possible, of the atom involved. A numeral used as a multiplier is placed to the left of the plus or minus superscript, but the numerical prefix "one" is not expressed. The charge and the coefficient are associated with the specific atom involved. Examples are shown in Figure 3.

Isotopes of elements other than hydrogen are represented in structural diagrams by placing the number representing the isotope mass as a superscript at the left of the element symbol involved. This numerical value is associated with the specific atom in the machine representation. The hydrogen isotope deuterium (^2H) is represented by the symbol D and the isotope tritium (^3H) by the symbol T. Examples are shown in Figure 3.

STRUCTURING SHORTCUTS

Structuring shortcuts are used extensively. These are groups or symbols which contain implicit bonding that does not have to be drawn or input since the shortcuts are expanded by
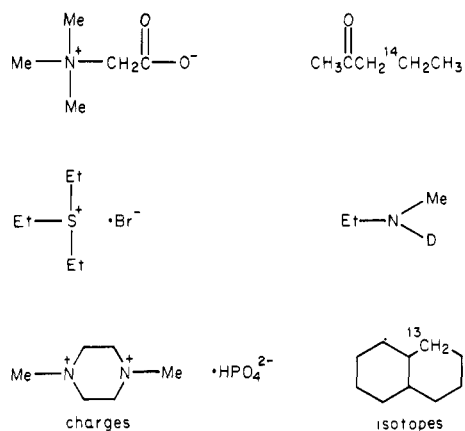
program to the detailed atom-bond record. That is, the same computer record results whether the shortcut or the expanded version is used. The approximately 200 shortcuts represent groups and also complete compounds. Alkyl groups, functional groups, coordinating ligands, mono- and disubstituted phenyl groups, amino acid groups, inorganic acid ions, certain complex boranes, and some common salt- and complex-forming compounds are included. Some examples are shown in Figure 4.

Compounds that contain sequentially repeating groups may be represented by the notation $-(X)_n-$, where X is the repeated unit composed of any number of atoms and bonds. Restrictions are as follows: the bonds at each end of the repeating group must have the same value; multiple repeating groups cannot be sequential; repeating groups cannot be nested; in a ring structure, a nonrepeating group must occur before and after the repeating group; repeating groups must have at least two bonds. The machine representation is the same whether the shortened notation is used for repeating groups or the more detailed, expanded form. Examples are shown in Figure 5.

The molecular formula always accompanies the structural diagram and is input to the Registry System. The editing/checking role of the molecular formula in the Registry System has been described.[1,5] The molecular formula is written in the order C, H, followed by other elemental symbols in alphabetical order. If the structure does not contain carbon, all symbols are given in alphabetical order. Only in the case of deuterium and tritium are isotopic elements shown separately in the molecular formula. Charges that are shown in the diagram do not appear in the molecular formula.

The steric descriptor of the substance as a whole is provided by means of terms called descriptors recorded in a separate segment of the connection table. These stereochemical descriptors are of seven types and are machine validated. The detailed handling of stereochemistry has been described[3] and will not be covered further in this paper. The CAS Chemical Registry System handles certain chemically equivalent representations of the same substance by algorithmically recognizing tautomeric and alternating single–double bond, aromatic or resonant type structures, replacing the explicit single and double bonds with special normalized bonds, and associating the migrating tautomeric hydrogen with groups of atoms rather than just single atoms. Detailed handling of these situations has been described[7] and will not be further described in this paper.

METAL SALTS

In considering input conventions of metal salts, the term "salt" refers to the metal replacement of hydrogen in the acidic function of acids, the amino group of amines and amides, and the hydroxyl group of phenols and alcohols and their thio, seleno, and telluro analogues.
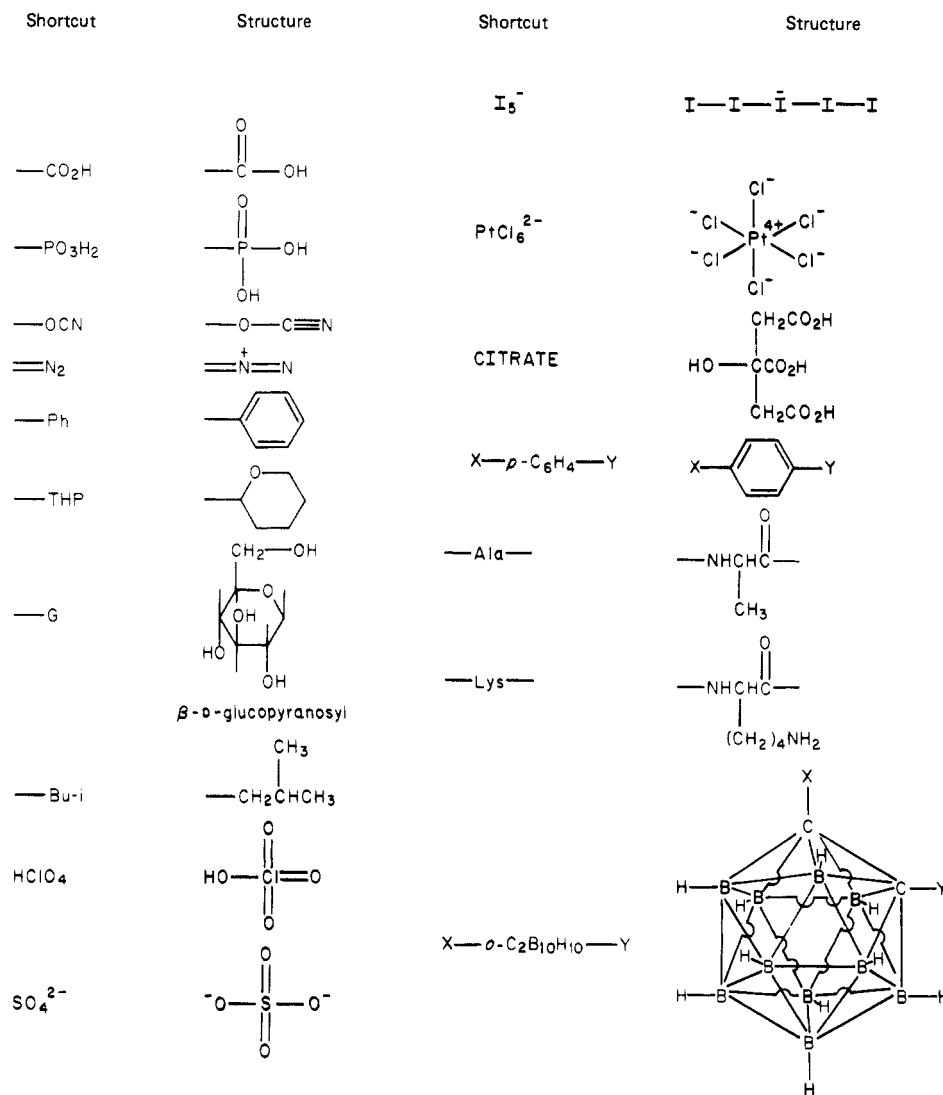
Shortcut   Structure   Shortcut   Structure

I_5^- I—I—$\bar{I}$—I—I

—CO_2H

—PO_3H_2 $PtCl_6^{2-}$

—OCN

=N_2 CITRATE

—Ph

—THP   X—p-C_6H_4—Y

—G   —Ala—

β-D-glucopyranosyl

—Lys—

—Bu-i

HClO$_4$

X—o-$C_2B_{10}H_{10}$—Y

SO_4^{2-}

Figure 4. Input shortcut conventions.

F—$(CF_2)_7$—F

Ph—N=$(CH—CH)_3$=N—Ph

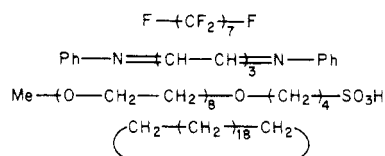Me—$(O—CH_2—CH_2)_8$—O—$(CH_2)_4$—SO_3H

$(CH_2—(CH_2)_{18}—CH_2)$

Figure 5. Repeating units.

In view of the existence of different ways of representing such metal salts in the literature, one method has been selected for consistency, using the parent substance as the basis. The salts are structured in a "disconnected" form, consisting of the parent compound with no hydrogens replaced, and the metallic element symbol, as illustrated in Figure 6. The correct ratio of acid to metal is expressed by a numerical prefix associated with the metal symbol, both in the structural diagram and the machine representation. This prefix may be an integer ("1" is not expressed) or a common fraction. Where the ratio is not known, the prefix "x" is used. For multivalent metals, the oxidation state is given by appending the appropriate Roman numeral in parentheses to the metal symbol, an x being used for an undetermined value. The appropriate numerical value or x is associated with the metal atom in the machine representation. The molecular formula follows the same format, except that the oxidation state of the metal is not used. The elements listed in Table I are considered metals. Their salts are structured according to the "disconnected" convention.
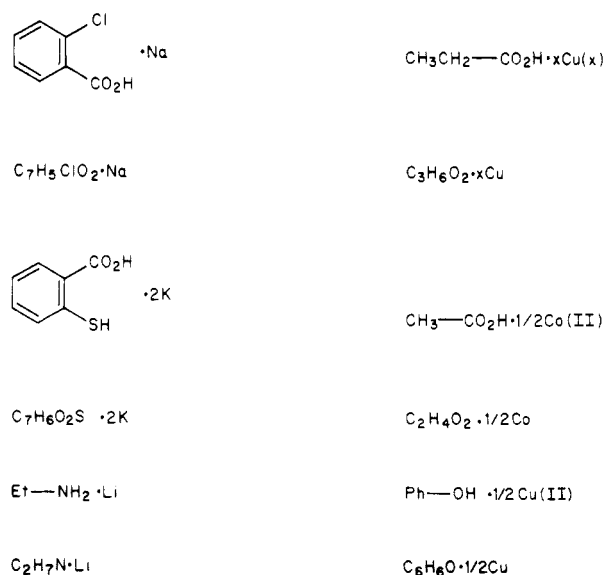
$C_7H_5ClO_2$·Na CH_3CH_2—CO_2H·xCu(x)

$C_3H_6O_2$·xCu

$C_7H_6O_2S$ ·2K CH_3—CO_2H·1/2Co(II)

$C_2H_4O_2$ ·1/2Co

Et—NH_2 ·Li   Ph—OH ·1/2Cu(II)

C_2H_7N·Li C_6H_6O·1/2Cu

Figure 6. Input conventions for metal salts.

ADDITION COMPOUNDS AND ONIUM COMPOUNDS

Addition compounds and acid salts of organic nitrogen and other bases are structured by using the disconnected format

Table I. Metallic Elements

Name	Symbol	Name	Symbol
*Actinium	Ac	*Molybdenum	Mo
Aluminum	Al	*Neodymium	Nd
*Americium	Am	*Neptunium	Np
*Antimony	Sb	*Nickel	Ni
Barium	Ba	*Niobium	Nb
*Berkelium	Bk	*Nobelium	No
Beryllium	Be	*Osmium	Os
*Bismuth	Bi	*Palladium	Pd
Cadmium	Cd	*Platinum	Pt
Calcium	Ca	*Plutonium	Pu
*Californium	Cf	*Polonium	Po
*Cerium	Ce	Potassium	K
Cesium	Cs	*Praseodymium	Pr
*Chromium	Cr	*Promethium	Pm
*Cobalt	Co	*Protactinium	Pa
*Copper	Cu	Radium	Ra
*Curium	Cm	*Rhenium	Re
*Dysprosium	Dy	*Rhodium	Rh
*Einsteinium	Es	Rubidium	Rb
*Erbium	Er	*Ruthenium	Ru
*Europium	Eu	*Samarium	Sm
*Fermium	Fm	*Scandium	Sc
Francium	Fr	*Silver	Ag
*Gadolinium	Gd	Sodium	Na
Gallium	Ga	Strontium	Sr
*Germanium	Ge	*Tantalum	Ta
*Gold	Au	*Technetium	Tc
*Hafnium	Hf	*Terbium	Tb
*Holmium	Ho	*Thallium	Tl
*Indium	In	*Thorium	Th
*Iridium	Ir	*Thulium	Tm
*Iron	Fe	*Tin	Sn
*Lanthanum	La	*Titanium	Ti
Lawrencium	Lr	*Tungsten	W
*Lead	Pb	*Uranium	U
Lithium	Li	*Vanadium	V
*Lutetium	Lu	*Ytterbium	Yb
Magnesium	Mg	*Yttrium	Y
*Manganese	Mn	Zinc	Zn
*Mendelevium	Md	*Zirconium	Zr
*Mercury	Hg		

* -- variable valence

$C_6Br_4O_2 \cdot xC_{14}H_{10}$
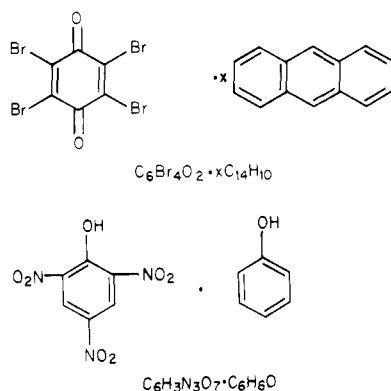
$C_6H_3N_3O_7 \cdot C_6H_6O$

Figure 7. Addition compounds.

described in the previous section. Each component is structured separately, the ratio of components being expressed by integers or common fractions. The letter x is used when the ratio is not known. The numerical value, or x, is associated with the appropriate component in the machine representation. The molecular formula follows the same format. Examples of addition compounds are shown in Figure 7.

Hydrogen acid salts of organic nitrogen bases are represented in this disconnected format. Compounds of primary and secondary nitrogen and other bases with alkyl halides, sulfates, etc., are structured as secondary and tertiary amines with the hydrogen acid salt separated. Compounds of tertiary nitrogen bases with alkyl halides, sulfates, etc., are also structured in a disconnected, as opposed to onium, form when the site of specific attachment of the alkyl is not specified. Illustrations are shown in Figure 8.
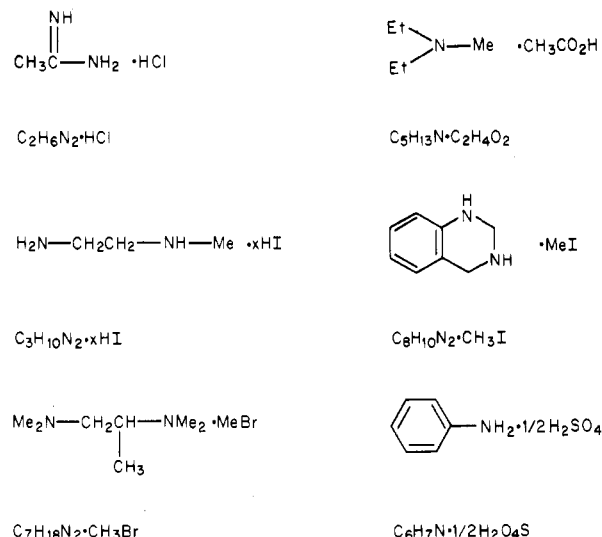

$C_2H_6N_2 \cdot HCl$

$C_5H_{13}N \cdot C_2H_4O_2$

$H_2N-CH_2CH_2-NH-Me \cdot xHI$

$C_3H_{10}N_2 \cdot xHI$

$C_8H_{10}N_2 \cdot CH_3I$

$Me_2N-CH_2CH-NMe_2 \cdot MeBr$

$C_7H_{18}N_2 \cdot CH_3Br$

$C_6H_7N \cdot 1/2 H_2SO_4$

Figure 8. Salts of organic nitrogen bases.

$C_2H_5AsCl_3 \cdot Cl$

$C_{12}H_{28}N \cdot ClO_4$

$C_{14}H_{12}IO_2 \cdot I$

$C_6H_{13}S \cdot Br$

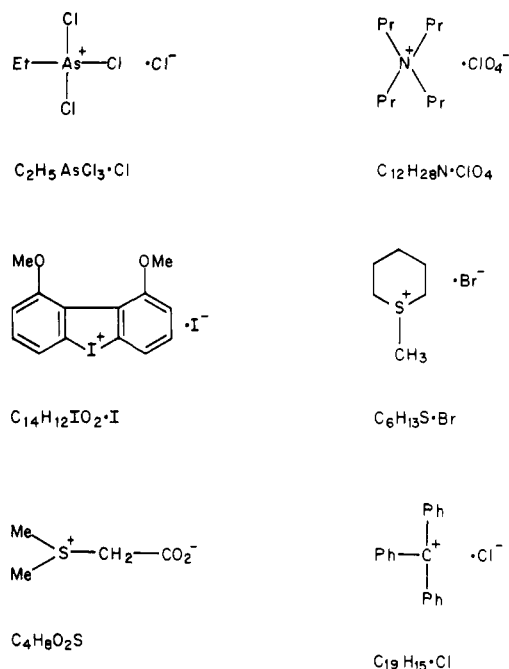$C_4H_8O_2S$

$C_{19}H_{15} \cdot Cl$

Figure 9. Onium compounds.

The term "onium compounds" refers to compounds containing organic cations in which the charged center is completely substituted, i.e., all hydrogen atoms have been replaced. Those that are not completely substituted are represented as addition compounds, described in the preceding paragraphs. Such onium compounds are ammonium, diazonium, oxonium, carbonium, sulfonium (with four bonds to carbon atoms), arsonium, selenonium, etc. These compounds are represented in the disconnected format, with the anion structured separately, unless an inner salt or zwitterion is involved. Positive and negative charges and numerical coefficients for ratios are used as appropriate and are associated with the specific atom or component as appropriate in the machine representation. Molecular formulas follow the same format, without the use of charge indications. Examples are shown in Figure 9.

COORDINATION COMPOUNDS

Structural representation of coordination compounds has probably been richer in variety, over many decades, than any other class of compounds. This is still true, since in today's usage, the manner of representation varies depending on the

particular viewpoint involved in a given publication. For the sake of consistency in structuring, and to provide for detailed machine editing procedures, one method of representation has been selected.

A coordination compound is defined as a compound consisting of a central atom (usually metallic) to which are bonded neutral or ionic molecules via coordinate bonds and/or anions or cations bonded via coordinate covalent bonds. In these coordination compounds, the central atom is not bonded in its oxidation state; that is, the number of bonds from the central atom to other groups or atoms is either smaller or (usually) larger than its oxidation state. For example, the cation $Co(HOEt)_6^{2+}$ consists of a central atom, Co, whose oxidation state is 2+, to which six neutral ethanol molecules are attached via coordinate bonds.

The coordination structuring conventions involve the following terms:

Central atom or coordination center—the atom, usually a metal, which is surrounded by ligands.

Ligand—a molecule or ion that is attached to a central atom through coordinate or covalent bonds or both.

Ligating atoms or donor atoms—those atoms in a ligand which are directly involed in bonding to the central atom.

The central atom is represented in the structure by its elemental symbol. The ligands are placed around the central atom in the structure, with the ligating atoms oriented toward the central atom whenever possible. The central atom and the ligating atoms are connected by lines which represent the coordinate and/or covalent bonds in the compound. Some ligands are monodentate and contain only one ligating atom; others are bi-, tri-, tetra-, etc., dentate, indicating that they have two, three, four, etc., ligating atoms. In general, only one line or direct bond is shown between the central atom and each ligating atom, although double and triple bonds are used in some cases. No distinction is made between coordinate and covalent bonds in the machine representation.

The ligands are structured, generally, in exactly the same manner as they would be if they were not bonded to the coordination center. Neutral organic ligands are structured according to the guidelines previously described; since they form only coordinate bonds, they remain formally neutral in the coordination compound and are not assigned charges. Ionic organic ligands are also structured as described in an earlier section except that ionic delocalized ligands are structured with localized charges and bonds.

The central atom is always assigned a charge, which is indicated to the upper right of its elemental symbol in the structure. The value of this charge is that which results in the entire coordination molecule having the appropriate charge, be it neutral, cationic, or anionic. The charge assigned to the central atom takes into account the charges on ionic ligands and usually is the same as the actual oxidation state of the central atom. This charge may be positive, zero (shown as Ø), or negative. If the coordination molecule is ionic, there will be counterions present if the entire compound is neutral. These counterions are structured according to the appropriate organic or inorganic conventions. Illustrations of coordination compounds are shown in Figure 10.

Some coordination compounds contain organic unsaturated molecules (neutral or ionic) as ligands which bond to the central atom through π bonds or both σ and π bonds. These compounds are structured with direct bonds from the central atom to each atom involved in π bonding or σ bonding. All hydrogens connected to atoms involved in bonding are shown in the structure. Metallocenes are compounds containing two 2,4-cyclopentadien-1-yl ligands which are π bonded to a central metal atom, to which no other ligands are bonded. Charges and double bonds of ionic π-bonded ligands are localized in
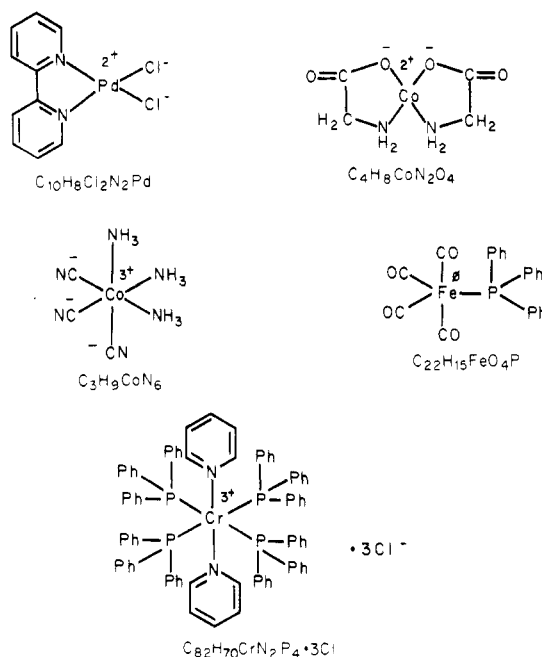

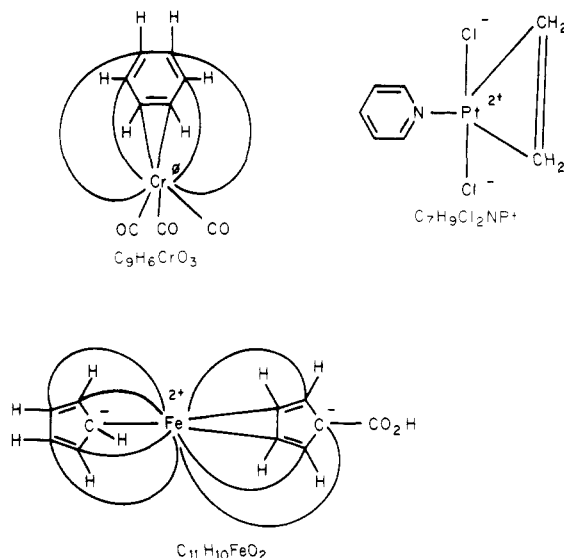**Figure 10.** Coordination compounds.

Figure 11. π-bonded coordination compounds.

the structure and are assigned according to nomenclature policies whenever localization would otherwise be ambiguous. Examples of π-bonded coordination compounds are shown in Figure 11.

INCOMPLETELY DEFINED SUBSTANCES

Incompletely defined (ID) substances fall into four general classes and any one substance may be wholly of one class or a combination of two or more. Besides the structural diagram itself and molecular formula, an additional item, or descriptor, is required to augment the atom-bond description. These descriptors are similar in format to the stereodescriptor described previously,[3] and have been referred to in a general way in an earlier publication.[1]

The following are descriptions of the four classes of incompletely defined substances.

Class I, substances with unknown structural components, i.e., those which have only a portion of the structure represented by a molecular formula segment.

Class II, substances with unknown location of unsaturation—those for which the location of a partic-
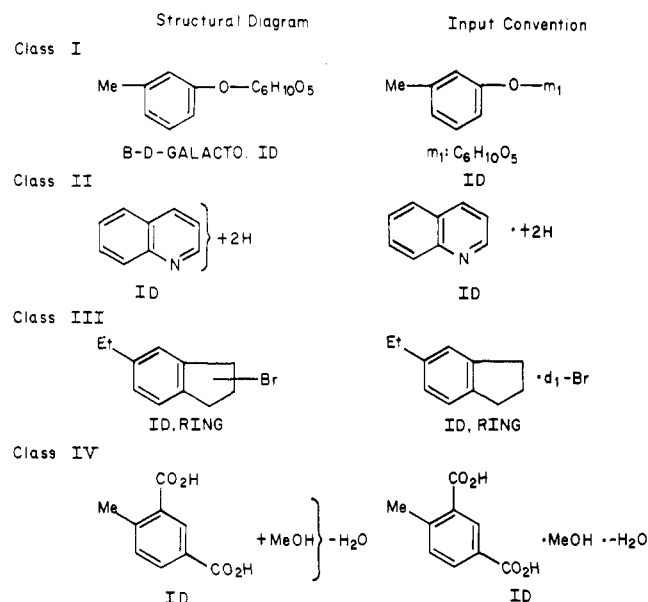
CAS CHEMICAL REGISTRY SYSTEM

J. Chem. Inf. Comput. Sci., Vol. 22, No. 1, 1982 **27**

Figure 12. Incompletely defined structures.

Figure 13. Input conventions for polymers as component monomers.

Figure 14. Input conventions for polymers in SRU terms.
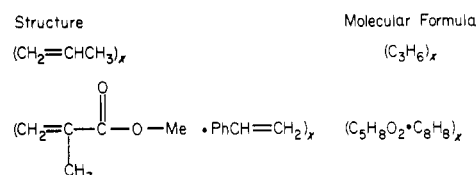

Figure 15. Input conventions for SRUs without end groups.

Figure 16. Input convention for a mixture.

ular multiple bond is unspecified and which can be represented only by the addition or subtraction of the appropriate number of hydrogen atoms.

Class III, substances with unknown position of attachment of structural component(s).

Class IV, substances with unknown position for functional derivative formation—those that have more than one functional group available for derivative formation, but the group which has actually been involved is unknown.

Examples are shown in Figure 12 together with the corresponding machine record. Each has a descriptor, consisting of the term "ID" as an indicator of the class plus any additional terms for specificity.

POLYMERS

Polymers are structured either on the basis of the monomers from which they are prepared or on the basis of the structure of the polymer. The former is the most frequently used, since it is in terms of monomers that most polymers are described in chemical literature.

Polymers in terms of monomers are structured and machine recorded very much like addition compounds. If more than one monomer is involved, each is structured separately and the complete expression enclosed in parentheses with an "x" subscript and an indication that the substance is a polymer. Monomer ratios are not included in the record. Examples of polymers represented in terms of the component monomer(s) are shown in Figure 13, with corresponding molecular formulas and descriptors.

Polymers are also described in terms of the structural repeating unit (SRU), the smallest group of atoms which on sequential repetition represents the backbone of the polymer. The SRU may or may not have specific end groups. The structure is enclosed in parentheses with an "n" subscript, the end bonds being left open or connected to end groups as ap-
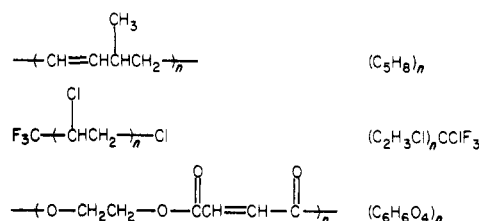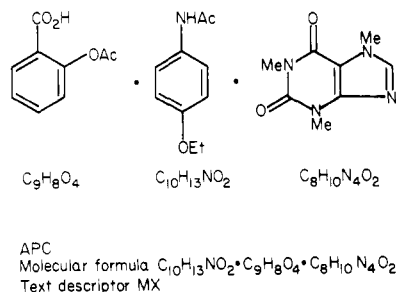
propriate. The molecular formula of the SRU is also enclosed in parentheses with an "n" subscript. If end groups are present the molecular formula is recorded as described, followed by the formula of the combined end groups. Examples of polymers in terms of SRU are shown in Figure 14.
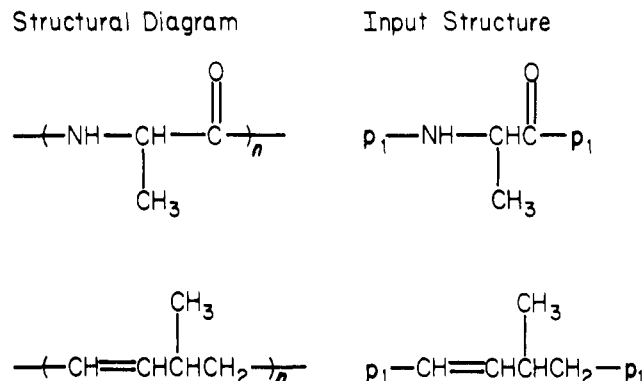
In the case of polymer SRUs without end groups, the open bonds are attached to a dummy node p_1 (p_2 if a double bond) in the machine representation, as shown in Figure 15. In the input conventions for SRUs with end groups, the atoms and connections inside and outside of the SRU are identified separately to distinguish polymers from nonpolymeric substances.

MIXTURES

Mixtures are expressed in terms of the components. Each component is structured separately, with the appropriate molecular formula and steric and/or nonsteric descriptor. The ratio among components is not included in the structure record. The molecular formula for the mixture is a composite made up of the formulas of each component. The descriptor MX is associated with the mixture itself. Figure 16 illustrates the input convention for the pharmaceutical APC (aspirin–phenacetin–caffeine).

ALLOYS

Alloys are structured for input in a tabular format, consisting of the element symbols with their weight percent composition, if present; however, composition is not required data. Components equal to or greater than 0.1% are listed

```
Alnico V

Molecular Formula  Fe.Co.Ni.Al.Cu.Ti.Si

Text Descriptor  AY,ALNICO V

Structure    Fe      49

             Co      24

             Ni      15

             Al       8

           . Cu       3

             Ti      0.5

             Si      0.1
```

Figure 17. Input convention for an alloy.

Structural Diagram	Molecular Formula
Cl—Mg—Cl	Cl_2Mg
Cl—P—Cl (with O double bond above, Cl below)	Cl_3OP
H—Be—H	BeH_2
HO—Fe═O	$FeHO_2$
S═Mn═S	MnS_2

Figure 18. Inorganic compounds.

in order of decreasing concentration. The text descriptor "AY" and the alloy name or code, if any, is applied to the alloy as an entity. If present, the percent composition, which is a basis for differentiation during registration, may be specific or may be given in ranges for some or all of the components. The molecular formula for the alloy is cited in a disconnected format, with the components given in the same order as listed in the tabular structure, and without composition data. The input convention for an alloy is shown in Figure 17.

INORGANIC COMPOUNDS

Inorganic compounds, that is, compounds that contain only elements other than carbon, follow in general the basic structuring conventions for organic compounds. Those compounds for which the number of groups attached to a given atom is equal to the valence of that atom are structured as shown in the examples in Figure 18. Metal salts of oxo acids and their thio, seleno, and telluro analogues are represented in the disconnected convention previously described. Some examples are shown in Figure 19.

Stereochemistry specified for inorganic compounds is expressed by text descriptors in accordance with conventions described previously.[3]

Structural Diagram	Molecular Formula
HO—Cr—OH .Ca (with O double bonds above and below)	$CrH_2O_4 \cdot Ca$
HO—P—OH .3/2 Mg (with S double bond above, OH below)	$H_3O_3PS \cdot 3/2Mg$
HO—N (with two O double bonds) .1/3 Fe(III)	$HNO_3 \cdot 1/3Fe$

Figure 19. Metal salts of inorganic oxo acids.

SUMMARY

Input structuring conventions for substances registered in the CAS Chemical Registry System have been described. These substances include fully defined organic (carbon-containing) compounds, inorganic compounds, ions, polymers, coordination compounds, alloys, mixtures, and certain partially defined substances. Knowledge of these conventions is very useful for those using various Registry services, such as the Private Registry Service and CAS ONLINE, for registration and for searching involving specific substances or sets of substances sharing structural features.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Dittmar, P. G.; Stobaugh, R. E.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. I. General Design". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111–121.
(2) Freeland, R. G.; Funk, S. A.; O'Korn, L. J.; Wilson, G. A. "The Chemical Abstracts Service Chemical Registry System. II. Augmented Connectivity Molecular Formula". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 94–97.
(3) Blackwood, J. E.; Elliott, P. S.; Stobaugh, R. E.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. III. Stereochemistry." *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 3–8.
(4) Vander Stouw, G. G.; Gustafson, C. R.; Rule, J. D.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. IV. Use of the Registry System to Support the Preparation of Index Nomenclature". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 213–218.
(5) Zamora, A.; Dayton, D. L. "The Chemical Abstracts Service Chemical Registry System. V. Structure Input and Editing". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 219–222.
(6) Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. VI. Substance-Related Statistics". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 76–82.
(7) Mockus, J.; Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. VII. Tautomerism and Alternating Bonds". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 18–22.
(8) Moosemiller, J. P.; Ryan, A. W.; Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. VIII. Manual Registration". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 83–88.