

CAS ONLINE®

Available on STN International®
The Scientific & Technical Information Network

SEARCHING COORDINATION COMPOUNDS

December 1986

Chemical Abstracts Service
A Division of the American Chemical Society
2540 Olentangy River Road
P.O. Box 3012
Columbus, OH 43210

Copyright © 1986 American Chemical Society

Quoting or copying of material from this publication
for educational purposes is encouraged, providing
acknowledgment is made of the source of such material.

SEARCHING COORDINATION COMPOUNDS

prepared by

Adrienne W. Kozlowski
Professor of Chemistry
Central Connecticut State University

while on sabbatical leave as a
Visiting Educator, Chemical Abstracts Service

Table of Contents

<u>Topic</u>	<u>Page</u>
PREFACE.....	i
CHAPTER 1: INTRODUCTION TO SEARCHING IN CAS ONLINE.....	1
What is Substructure Searching?.....	1
The Basic Commands.....	2
CHAPTER 2: INTRODUCTION TO COORDINATION COMPOUNDS.....	5
Definitions and Terminology.....	5
Ligand Characteristics.....	6
Metal Characteristics.....	8
CHAPTER 3: STRUCTURING AND REGISTRATION POLICIES FOR COORDINATION COMPOUNDS.....	11
Policies for Structuring Coordination Compounds.....	11
Ligands.....	11
Ligand Structures.....	11
Metal-Ligand Bonds.....	12
Pi-Donors.....	12
Hydrogens.....	12
Hydride Ions.....	13
Charges.....	14
Counter Ions.....	14
Ligands with Negative Charges.....	14
The Central Metal Atom.....	15
Structuring Assumptions.....	16
Metal-Metal Bonds.....	16
Carboxylates.....	16
Incomplete Structures.....	17
Policies for Assigning Registry Numbers.....	18
Counter Ions.....	19
Geometric Ions.....	19
Optical Isomers.....	20
Different Oxidation States.....	20
Isotopes.....	20
Ionized Ligands.....	21
Component Registration.....	21
Polymers.....	21

CHAPTER 4: BUILDING STRUCTURES - THE BASICS.....	23
The GRAPH Command.....	23
The NODE Command.....	23
The BOND Command.....	24
Examples Using the GRAPH, NODE, and BOND Commands.....	26
Creating a 6-Coordinate Structure.....	26
Creating Chelate Rings.....	28
Creating Fused Rings as Ligands.....	30
Creating Clusters.....	32
Creating Cyclopentadienyl Ligands.....	33
Creating Macrocyclic Ligands.....	34
The RECALL Command - Using Models to Build Structures.....	37
Recall a Registry Number.....	37
Recall and Modify a Structure from the Fragment File.....	38
Variability in Structures.....	41
Variable Donor Atoms and Variable Chelate Backbone.....	41
Defining a Central Metal.....	42
Defining Ligand Atoms.....	42
CHAPTER 5: ADDING ATTRIBUTES TO STRUCTURES.....	43
The HCOUNT Command.....	44
The RSPEC Command.....	44
The CONNECT Command.....	45
Using CONNECT to Retrieve Metallocenes.....	47
Using CONNECT in Ligand Structures.....	48
Metal Hydrides.....	49
The NSPEC Command.....	50
Effect of Bond Type on Node Specification.....	50
Effect of Bond Type and Node Specification.....	52
System Defaults.....	52
Ring or Chain Bonds.....	53
Chain Bonds and Ring or Chain Nodes.....	54
All Bonds in Rings.....	55
One Chelate Ring.....	56
One Tridentate Ligand.....	57
Ring or Chain Node.....	57
Bidentate or Tetradentate Ligands.....	58

CHAPTER 6: ADDING SCREENS TO REFINE YOUR SEARCH.....	61
Screens That Are Often Useful.....	64
Adding Metal Screens When Structure Contains the M Node Symbol	64
Adding Screens that Specify Minimum Numbers of Ring or Chain	
Bonds.....	65
Hydrogen Augmented Atom Screens for Ligands.....	69
Type of Ring Screens.....	71
Pi-Bonding Ligands, Organometallics.....	72
Screens Not to Use.....	73
BATCH Searching.....	74
Dividing a Search that Will Not Complete in One Search.....	74
CHAPTER 7: SUMMARY OF GENERAL SEARCH TECHNIQUES.....	75
Using Ring or Chain Bond Types.....	75
Using Single Exact Bond Values.....	75
Using Ring or Chain Node Specification.....	75
Ring Isolation.....	76
Using Gk Nodes.....	76
Tautomers.....	76
Using A (Any Atom) Nodes.....	76
Using the CONNECT Attribute.....	77
Determining the Coordination Number.....	77
Using Hydrogen Counts for Hydrides.....	77
Attributes and the Gk Node.....	77
Using Charges.....	77
Using Screens.....	78
Interpreting SAMPLE Searches.....	78

CASE STUDY 1: TAILORING THE COORDINATION SPHERE.....	79
All Octahedral Dichloro Complexes of Cobalt.....	79
Dichloro Cobalt Complexes of Lower Coordination Number.....	81
Two Different Halide Ligands.....	82
Bridging Halides.....	84
Simple Halides and Halo Complexes of Rhenium.....	88
CASE STUDY 2: ORGANOMETALLICS.....	93
Metallocenes.....	94
Substituted Cyclopentadienyl Ligands.....	96
Noncyclic Polyenes.....	97
Ethylene.....	98
Pentadiene.....	99
Partial Coordination of the Pi System.....	100
CASE STUDY 3: CARBOXYLIC ACIDS.....	103
Comprehensive Searching.....	104
All Substances with the Ligand Bonded to the Metal.....	107
Carboxylate Complexes as Salts - No Bond to Metal.....	109
Specific Structures - N,O Chelates.....	111
Structures with an Uncoordinated Carboxylate Group.....	113
Structures with an Uncoordinated Pyridine Group.....	115
CASE STUDY 4: UNSPECIFIED DERIVATIVES.....	117
Substances without Registry Numbers.....	117
Manually Registered Compounds.....	120

CASE STUDY 5: CHELATES AND MACROCYCLES.....	123
Tridentate Chelates.....	123
Tetradentate Macrocycles with Both Sulfur and Nitrogen Donors...	127
Macrocyclic Complexes with Varying Numbers of Donor Sites.....	129
CASE STUDY 6: HYDRIDES.....	133
CASE STUDY 7: METAL CLUSTERS.....	139
Gold Clusters with Phosphine Ligands.....	139
Iron Carbonyl Clusters.....	142
Terminal Carbonyl Groups.....	142
Bridging Carbonyls.....	147
Clusters of Specific Size.....	150
CASE STUDY 8: STEREOCHEMISTRY.....	155
Geometry.....	159
Ligand Stereochemistry.....	162
Cis or Trans Isomers.....	165
CASE STUDY 9: NAME AND FORMULA SEARCHING.....	171
Searchable Fields.....	172
Proximity Operators.....	173
Truncation Symbols.....	174
Searching Formulas in the Basic Index.....	175
Formula Searching in the Molecular Formula Index (/MF).....	176
Searching Element Counts.....	178
Searching with Compound Class Identifier Codes (/CI).....	181
Name Searching in the Basic Index.....	182

APPENDIX 1: STRUCTURES OF LIGANDS.....	187
APPENDIX 2: FRAGMENT FILE.....	193
APPENDIX 3: SCREENS.....	197
Element Count Screens.....	197
Selected Augmented Atom Screens.....	198
Selected Hydrogen Augmented Screens.....	198
Selected Type of Ring Screens.....	199
Ring Count Screens.....	199
Degree of Connectivity Screens.....	200
Atom Count Screens.....	201
Graph Modifier Screens.....	202
INDEX.....	205

PREFACE

Searching Coordination Compounds in CAS ONLINE is aimed at a dual audience - both chemical information specialists who do not have extensive experience with inorganic substances and inorganic chemists who wish to use online searching. For this reason background information has been included on both coordination chemistry and general search procedures for CAS ONLINE.

Approximately 10% of all compounds in the Registry File contain transition metals, which corresponds to over 800,000 substances. Coordination compounds play such major roles in many subject areas, including biochemistry, catalysis, nuclear medicine, organic synthesis, and energy conversion, that many chemists are interested in obtaining information on these substances. Substructure searching offers a powerful tool for retrieving information on coordination compounds.

This book is divided into two parts. The first part contains chapters describing general policies used by Chemical Abstracts Service in structuring and naming compounds when they are registered and some general techniques for searching for coordination compounds online. The second part is a series of nine case studies illustrating search strategies for specific types of coordination compounds and organometallic compounds.

Chapters 2, 3, 6, and 7 and the first four case studies should be useful to searchers who are unfamiliar with the area of coordination compounds. Chapters 5 and 6 give details of creating structures of coordination compounds. The case studies are intended to give coordination chemists insights into the variety of structural features that can be incorporated into substructure searches and tailored to their research interests. The case studies and examples have purposely been left as general as possible so they can be adapted to a number of more specific queries.

There are a number of search aids which have been developed by Chemical Abstracts Service to provide assistance in conducting substructure searches in the Registry File with as much precision as possible. You will probably find it useful to review the following publications. The material in this booklet draws heavily on the topics of adding screens, specifying bonds, and structuring conventions covered in the booklets Structuring Conventions for Functional Groups, Adding Screens in Structure Searching, and Specifying Bonds in Structures. These booklets are available from Chemical Abstracts Service.

CHAPTER 1: INTRODUCTION TO SEARCHING IN CAS ONLINE

CAS ONLINE offers several ways to find information on coordination compounds. In the Registry File, searching via substructure or by name and/or molecular formula will provide answers that consist of Registry Numbers with their associated names and structural information. In the CA File, subject searching using Registry Numbers, words used in titles of documents, and/or keywords and index terms as search terms will provide answers that consist of CA abstract numbers with their associated bibliographic information, abstracts, keywords, and index terms. The major emphasis in this booklet is on specialized techniques for substructure searching for coordination compounds in the Registry File.

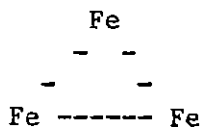
In this chapter we will give a brief overview of substructure searching and the basic commands needed to interact with the Registry File. More detail on the basic commands needed to create structures can be found in Chapters 4 and 5, with Chapter 4 describing the use of the structure building commands and Chapter 5 discussing the commands used to assign specific features to atoms. Still more information can be found in Case Study 9 which deals with searching for coordination compounds by formula or name in the dictionary portion of the Registry File. All chapters of this booklet are necessarily only brief introductions to the commands of CAS ONLINE. More detail is provided in Registry File workshops and full documentation is found in Volumes I-IV of Using CAS ONLINE: The Registry File.

What is Substructure Searching?

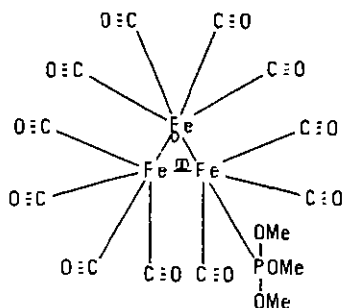
Substructure searching uses a structure diagram as the search term. Any structural fragment or functional group of interest may be used as a substructure. The answers you retrieve from a substructure search will contain your substructure somewhere in the structure of the retrieved answer. Unlimited substitution is permitted unless you restrict further substitution.

Introduction to Searching

SUBSTRUCTURE



ANSWER



One of the principal advantages of substructure searching is the ability to focus either on very generic structural features, such as all nitrogen donor ligands; or, on the other hand, on very specific combinations of features, like pi bonding through triple bonds of ligands in metal clusters. Such degrees of generality or specificity are usually not possible in bibliographic or dictionary searching. It is also possible in substructure searching to exclude unwanted features from specific portions of the molecule.

Through substructure searching you are able to avoid nomenclature complexities and changes and to bypass indexing policies.

The Basic Commands

The STRUCTURE command is used to create your substructure. Within the STRUCTURE command you use the commands GRAPH (GRA) to create the skeleton of the molecule (rings and chains), NODE (NOD) to assign an element other than carbon to a node, and BOND (BON) to specify the bond value (single, double, normalized, etc.) and/or the bond type (ring or chain).

Introduction to Searching

The SEARCH command is used to initiate the process of searching the substructure you have created against the more than 8 million substances in the Registry File. You may do substructure searches in which unlimited substitution is allowed or EXACT or FAMILY searches in which no further substitution is permitted.

The search itself is carried out in two stages. The first stage is a screen search based on the numerical codes (screen numbers) automatically generated from your query structure(s) plus any screen numbers you have manually added to enhance the search strategy. This initial search screens out, or eliminates, the structures that do not meet the requirements of your question. The second stage of the search is an atom-by-atom bond-by-bond comparison of the candidate structures that were not screened out by the screen search with the structure(s) you defined. The user is usually unaffected by the fact that the search is done in two stages. In Chapter 6 there is a more detailed discussion of screens.

The HELP command, or a ?, can be used to get online help in the use of commands. Type HELP at an arrow (=>) to see all the commands available. Or type HELP GRAPH or HELP NODE or HELP BOND at a colon (:) in the STRUCTURE command to get instructions on the use of the structure building commands.

Summary

This chapter has given a very brief overview of substructure searching in the CAS ONLINE Registry File. Some of the basic commands have been defined. The next chapter will discuss coordination compounds and the terms used by chemists to describe them.

CHAPTER 2: INTRODUCTION TO COORDINATION COMPOUNDS

This chapter will provide a very general overview of the field of coordination compounds and the terminology chemists use to describe them.

Coordination compounds have been recognized by chemists as a special type of compound since late in the eighteenth century. Throughout the nineteenth century, a wide variety of "complex inorganic compounds" were prepared by reacting metal salts with neutral organic or inorganic molecules, such as ethylene or ammonia, or with halide ions and other simple anions, such as cyanide and nitrite. The formation of new stable complex compounds from molecules and salts, each of which is capable of independent existence, required a dramatic new theory to explain how such complexes were able to form. In 1892 Alfred Werner, the great Swiss chemist, proposed the Coordination Theory. He spent the rest of his career expanding and refining it, culminating in his receiving the Nobel Prize in 1913.

Werner's Coordination Theory is one of the great landmarks of chemistry and has proved extremely fruitful in fostering a myriad of developments that have taken coordination chemistry into such diverse areas as the use of organometallic compounds in synthesis and catalysis, new insights into the role of metal ions in biochemical reactions, and investigations into metal-metal bonding and the formation of clusters, where a number of metal centers come together in one compound. Today, coordination compounds play key roles in medicine, agriculture, materials science, electrochemistry, catalysis, and industrial organic synthesis.

Definitions and Terminology

The following section provides a basic guide to the specialized terminology chemists use to describe coordination compounds.

A coordination compound is a molecule or ion in which a central atom (usually a metal) is bonded to a group of molecules and/or ions via coordinate bonds. Such bonding occurs until the metal has satisfied its coordination number, which is usually larger than its oxidation state. In some cases the coordination number may be smaller than the oxidation state, particularly if the central atom is a nonmetal.

Coordination Compounds

Coordination compounds may be uncharged or charged, either positive or negative. Charged coordination compounds are often referred to as complex ions. Both charged and uncharged coordination compounds may also be referred to as metal complexes.

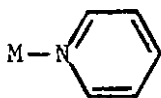
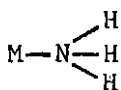
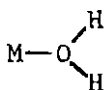
The terms coordination compounds, complex ions, and metal complexes are frequently used interchangeably.

The ions or molecules bonded to the central atom are called ligands, or electron pair donors, or Lewis bases. The atoms directly attached to the central metal make up the coordination sphere of the metal. The number of atoms in the coordination sphere is the coordination number of the metal atom. A coordinate covalent bond results when a ligand donates two electrons to an atom (metal, in our case) which accepts the electron pair.

Ligand Characteristics

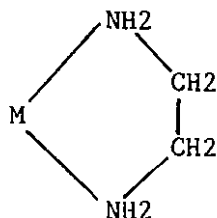
Both organic and inorganic molecules and ions can function as ligands in coordination compounds. All ligands attached to the central metal may be identical; or, a given complex may have several different ligands.

Ligands may be classified according to the structure of the complex formed. Ligands that form only one bond with the metal are called monodentate or unidentate, as illustrated below by the ligands water, ammonia, pyridine, chloride, and cyanide

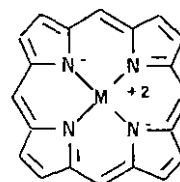


Coordination Compounds

Ligands with more than one electron pair donor site are called multidentate, or polydentate; specifically, bidentate, tridentate, tetradentate, etc., according to the number of attachments they make with the metal. If the multidentate ligand itself is a large ring, it is termed macrocyclic, and its complexes are macrocycles.



bidentate



tetradentate macrocycle

When a multidentate ligand becomes coordinated, chelate rings are formed as the ligand backbone wraps around to attach to the second and successive sites in the coordination sphere. Chelate ring formation is most likely when coordination to the metal results in either five or six membered rings, but examples are known with three, four, and seven membered or even larger rings.

Some monodentate ligands have two or more different donor sites, so they may coordinate in more than one way. Such ligands are called ambidentate ligands, illustrated below by the thiocyanate ion which can form complexes by donating the electron pairs from either the nitrogen atom or the sulfur atom.



thiocyanato-N



thiocyanato-S

Coordination Compounds

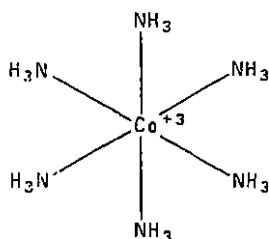
Organometallic compounds are a special category of coordination compounds in which one or more carbon atoms of an organic molecule are directly attached to the central metal atom. However, if the only carbon donors present are the inorganic donors carbon monoxide (CO), carbonyl sulfide (CS), or cyanide ion ($\text{CN}(\text{l-})$), the complex is not termed organometallic.

There is a category of ligands containing either double or triple bonds which coordinate through the π bond. These are called π donor ligands. Common examples include ethylene and cyclopentadienyl ion. Biscyclopentadienyl metal complexes are called metallocenes. Complexes of π bonding ligands are described as having hapto bonding and have the symbol .eta. (η) as part of their names.



Metal Characteristics

Every central metal atom in a coordination compound has a charge, also known as the oxidation state. The charge may be zero, positive, or occasionally even negative. The sum of the metal and ligand charges determines whether the coordination compound will be electrically neutral, a positive ion, or a negative ion.

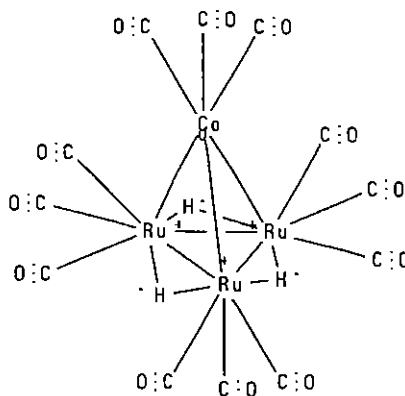


Coordination Compounds

Each central metal atom in a complex also has a coordination number, which indicates the number of donor atoms present. The size of the coordination number is a function of the particular metal involved, its oxidation state, and the particular ligands present. Chemically, the coordination number may range from two to twelve; coordination numbers four and six are especially common, although many examples are known of the other values. The coordination number in the previous structure is six.

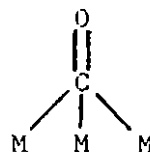
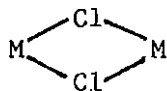
There are one or more geometries or stereochemistries corresponding to each coordination number. With coordination number four, for example, the complex may be either square planar or tetrahedral. With coordination number six, the complex usually exhibits octahedral geometry.

Certain coordination compounds have more than one central metal atom in their structure. They are called polynuclear complexes. Some polynuclear complexes possess direct metal-metal bonds; these are also called metal clusters. Clusters as large as 22 metal atoms are known.



Coordination Compounds

Polynuclear coordination compounds may also have bridging ligands, which simultaneously are coordinated to two or three metal centers. The presence of bridging ligands is indicated by .mu. (μ) before the name of the ligand which is attached to more than one metal atom.



μ -(thiocyanato-N:S)

di- μ -chloro

μ_3 -carbonyl

Summary

This chapter has given a brief overview of the nature of coordination compounds. The terms commonly used by chemists to describe these compounds have been defined. Chapter 3 will discuss the structuring conventions used for coordination compounds in the Registry File.

CHAPTER 3: STRUCTURING AND REGISTRATION POLICIES FOR COORDINATION COMPOUNDS

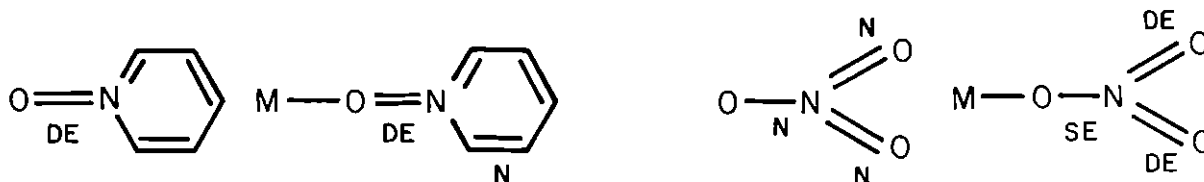
When a new substance enters the CAS Registry System, it is assigned a unique Registry Number and its structural information, e.g., its structural diagram, index name, synonyms, are entered into the Registry File. This chapter summarizes the CAS policies used in structuring and registering coordination compounds.

Policies for Structuring Coordination Compounds

When you create structures for searching in the CAS ONLINE Registry File, you will find it helpful to consider these structuring conventions that are followed when CAS registers coordination compounds.

Ligands

Ligand Structures. Ligands are structured, generally, in exactly the same manner as they would be if they were not bonded to the metal. Major exceptions to this generalization are those oxyanions and their thio analogs which require normalized bonds when they are not bonded to a metal. When such ligands coordinate, any normalized bonds on an atom coordinated to the metal become exact. Bonds between uncoordinated atoms may remain normalized or may become exact, depending on the particular oxyanions. In special cases, when certain ligands bridge two or more metal atoms, the bond values within the ligand may also change.

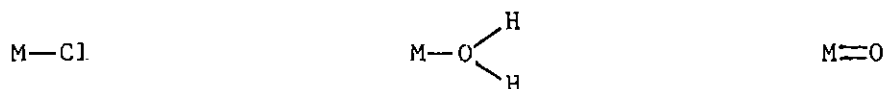


Appendix I shows how a number of ligands are structured in the Registry System.

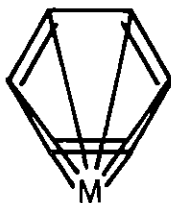
Structuring and Registration Policies

Ligands

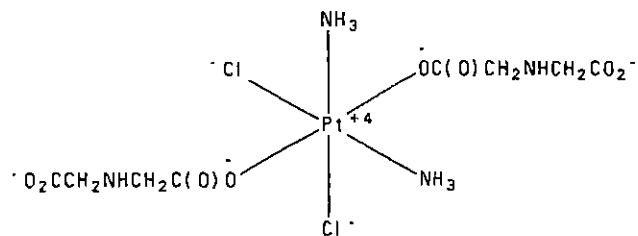
Metal-Ligand Bonds. Metal-ligand bonds, in general, are single exact bonds. Exceptions are the monatomic ions O(2-) and S(2-), which require double bonds, and N(3-) and P(3-), which require triple bonds, provided that each monatomic ion is attached to only one metal atom.



Pi-Donors. Pi-donors, such as cyclopentadiene, benzene, ethylene, or allyl groups, are structured with direct bonds from each atom involved in the pi-bonding to the metal. This convention creates one or more three-membered rings. Ring systems with normalized bonds prior to coordination remain normalized in the coordination compound.



Hydrogens. Hydrogens are removed from uncoordinated acid groups substituted on ligands and are included as a separate component, dot-disconnected from the ligand.



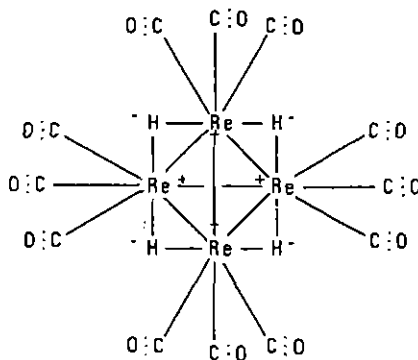
• 2H⁺

Structuring and Registration Policies

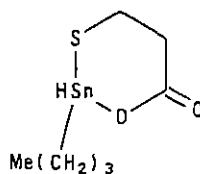
Ligands

Hydride Ions. The hydride ion, $H(1-)$, may function as a ligand and it may be attached to one, two, or three different metals. In the Registry System, a hydrogen atom bonded to a metal may be designated as a connection table hydrogen, i.e., an actual atom in the connection table for the substance, or as a nonconnection table hydrogen, i.e., an attribute assigned to the node. A hydrogen is a connection table atom if: it is a single atom fragment, it has two or more bonds, it has an abnormal valence, or it has a charge. Whenever the number of bonds to the metal is greater than the oxidation state of the metal, the hydrogen on the metal is a connection table hydrogen. However, if the number of bonds to the metal is the same as the oxidation state, the hydrogen connected to the metal is not a connection table hydrogen.

CONNECTION TABLE H



NONCONNECTION TABLE H

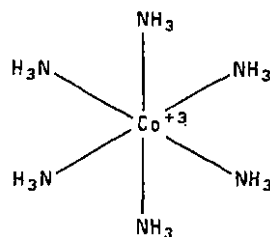


Structuring and Registration Policies

Charges

Charges

Counter Ions. Ionic coordination compounds may have counter ions present as separate components.



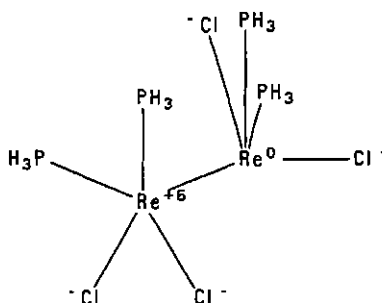
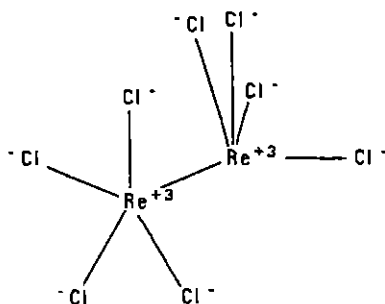
• 3Cl⁻

Ligands with Negative Charges. Cases where all of the ligands are negatively charged and where the number of bonds to the metal exactly equals the charge on the metal are not coded as coordination compounds in the Registry System. They do not have the coordination compound class identifier code. An example of such a compound is uranyl chloride. Note that no charges are placed on the metal or on the chlorines. This would not be coded as a coordination compound in the Registry File.



Structuring and Registration Policies Charges

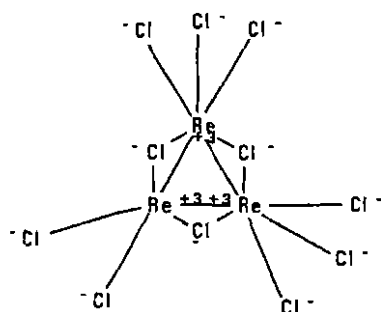
The Central Metal Atom. The central metal atom is assigned a charge, except in cases described in the previous example. The value of this charge is usually the same as the oxidation state of the metal, i.e., the algebraic sum of the metal charge and the ligands' charges must equal the net charge on the complex ion. Some coordination compounds contain more than one metal atom. In assigning charges to these, a series of arbitrary steps is followed. Metals with cyclopentadienyl ligands receive a charge of 1- for each Cp ring pi bonded to it. If possible, the remaining charge is distributed equally to the other metals so they each have the same charge. If it is not possible to give all metals the same charge, an attempt is made to assign the charge so that each metal can conform to the 18 electron rule. Finally, if none of these steps is possible, the charge is placed on the metal with the greatest atomic number, and the other metal atoms receive zero charge. This last step has been the indexing policy since 1982. Prior to 1982, the policy for this step was to register the substance by name and formula only. No structure was given.



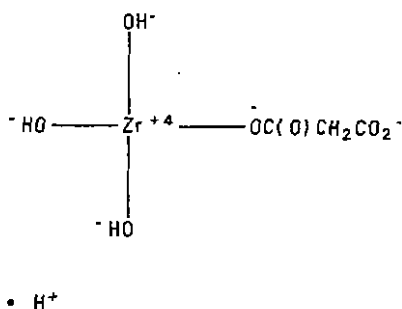
Structuring and Registration Policies Assumptions

Structuring Assumptions

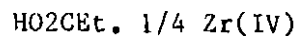
Metal-Metal Bonds. All metal-metal bonds are registered as single exact bonds, regardless of whether a multiple metal-metal bond is reported in the document.



Carboxylates. Carboxylates and their thio analogs are structured as coordinated if the mode of coordination is known or can reasonably be assumed.



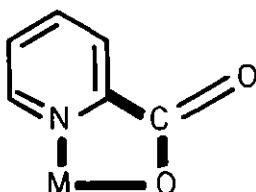
They are structured as two component metal salts if the bonding is not reported in the document and cannot be assumed.



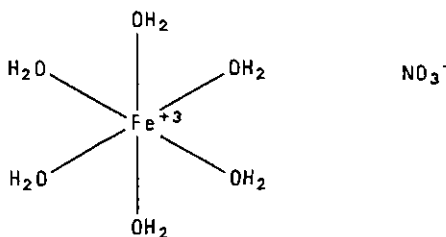
Structuring and Registration Policies Assumptions

Incomplete Structures. The structuring conventions discussed thus far assume that the bonding is completely known for each coordination compound, including the exact structure of the ligand, the points of attachment to the metal, and which ions are counter ions. If the complete structure is not known, certain assumptions about the bonding are made to permit the registration of the compound.

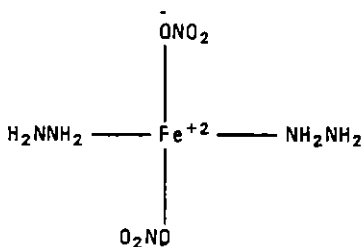
1. If a five or six membered chelate ring is possible, the ligand is assumed to be bidentate, up to the maximum coordination number of the central metal.



2. A chemically reasonable coordination number is assumed for the metal in its particular oxidation state.
3. Oxy ions, such as nitrate, perchlorate, and sulfate, are structured as counter ions, not ligands, unless the author specifically cites their coordination to the metal.



If coordinated, they are assumed to be monodentate unless evidence is presented for their bidentate attachment.

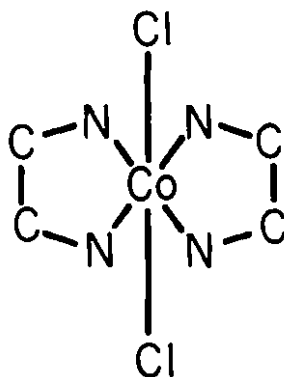


Structuring and Registration Policies Registry Number Assignment

Policies for Assigning Registry Numbers

After you perform a structure search, it may seem that a great many answers were retrieved, even when a specific structure had been created. Every unique chemical substance receives an individual Registry Number in the CAS Registry File, i.e., each stereoisomer, each compound with a different oxidation state for the metal (different charge), each isotopically labeled complex, and each salt of all the preceding cases are all unique chemical substances, and each receives its own Registry Number.

If we examine the answers retrieved during a substructure search for the fully defined coordination compound



we can see examples of these variations in complexes which have identical structures but different Registry Numbers. The following answers illustrate the various compounds that require unique Registry Numbers. We have not shown the complete record associated with these answers, only the Registry Number (RN), Index Name (IN), and, in some cases, synonyms (SY). We have underlined the term generating a unique Registry Number.

Structuring and Registration Policies
Registry Number Assignment

Counter Ions

Ionic coordination compounds may have oppositely charged counter ions present. Every one of these salts is registered individually.

RN 14040-33-6
IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, chloride

RN 14877-02-2
IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, nitrate

RN 28984-24-9
IN Cobalt(1+), dichlorobis(ethylenediamine)-, tetrakis(isothiocyanato)bis(pyridine)chromate(1-)

Geometric Isomers

The different stereochemistries are distinguished by different stereochemical descriptors. Every one of these isomers receives a different Registry Number.

RN 14875-15-1
IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22)-
SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, cis-

RN 14403-91-9
IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-12)-
SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, trans-

When the author does not indicate which isomer was used, CAS assigns a separate Registry Number for the complex with unspecified geometry.

RN 15275-04-4
IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-
SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion

Structuring and Registration Policies

Registry Number Assignment

Optical Isomers

Isomers can be both chiral and optically active. Each absolute configuration is assigned a unique Registry Number. The document must indicate which of these isomers is under consideration by quoting either the absolute configuration or the optical rotation in order for these Registry Numbers to be assigned.

RN 45837-15-8

IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22-.DELTA.)-

RN 18660-62-3

IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22-.LAMBDA.)-

SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, cis-(+)-

SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, D-cis-

SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, stereoisomer

Different Oxidation States

Each oxidation state for the metal receives a different Registry Number.

RN 21199-58-6

IN Cobalt, dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22)-

SY Cobalt, dichlorobis(ethylenediamine)-, cis-

RN 14875-15-1

IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22)-

SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, cis-

Isotopes

Isotopically labeled derivatives receive their own Registry Numbers.

RN 75829-80-0

IN Cobalt(1+)-57Co, dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22)-

Structuring and Registration Policies
Registry Number Assignment

Ionized Ligands

An ionized ligand, such as deprotonated ethylenediamine, generates a unique Registry Number.

RN 15318-86-2

IN Cobalt, dichloro(1,2-ethanediaminato-N,N')(1,2-ethanediamine-N,N')-(9Cl)

Component Registration

The free ion of a charged coordination compound, without any counter ion, receives a separate registration as a component. The Registry Number assigned to the free ion is quoted whenever that ion is part of a multicomponent system. For example, RN 14875-15-1, shown below, would also appear as part of the display of RN 14877-02-2, shown on page 19.

RN 14875-15-1

IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-22)-

SY cis-Dichlorobis(ethylenediamine)cobalt(1+) ion

Often, the component ion will not have any bibliographic references associated with it. The references may be associated with the multicomponent substance only.

Polymers

Polymers of a coordination compound are separately registered. At this time no polymers of dichlorobis(ethylenediamine)cobalt have been registered; but, the closely related substance shown below does form a homopolymer.

RN 75931-38-3

IN Cobalt(2+), chlorobis(1,2-ethanediamine-N,N')(4-ethenylpyridine)-, (OC-6-33)-, dichloride, homopolymer

Structuring and Registration Policies
Registry Number Assignment

Summary

In this chapter we have discussed some general guidelines for structuring coordination compounds. We have also listed the policies used for generating Registry Numbers for these compounds. Chapter 4 will discuss structure building commands used to create typical inorganic complexes.

CHAPTER 4: BUILDING STRUCTURES - THE BASICS

This chapter introduces the structure building commands and suggests techniques which are useful for creating structures of typical inorganic complexes. In many respects, creating structures for coordination compounds is easier than creating organic structures because the number of atoms in inorganic complexes is usually small.

This chapter provides an introduction to the basic structure building commands, GRAPH, NODE, and BOND. Full details of structure building are available in Using CAS ONLINE: The Registry File, Volume IIA.

The GRAPH Command

The GRAPH command is used to create the skeleton of the structure. It is used to create an atom, a chain of atoms, a ring, or a bond where nothing existed before. You may begin building a structure at any point. Most searchers find it easiest to start with a ring if one is present in the structure. After creating the major ring or chain in your structure, you can add chains to any node in the basic structure, or you can add fused or spiro ring systems. The system numbers the nodes in the structure in the order in which you create or add them. These node numbers have no relation to nomenclature.

The NODE Command

All atoms or nodes in a structure are carbons until you specify otherwise. The NODE command is used to assign any element to either a single position or to multiple positions in the structure. In addition to the standard element symbols, other symbols can be used to define a node.

Building Structures

GRAPH, NODE, and BOND

The variable node symbols are:

<u>Variable Node</u>	<u>Meaning</u>
M	Any metal*
X	Any halogen (F, Cl, Br, I, At)
A	Any element, including connection table hydrogens**
Q	Any element except carbon or non-connection table hydrogen**
G1 - G20	User defined multiple atoms or groups of atoms. These are defined using the VARIABLE and REPEATING commands.

* Metal is defined as all elements except Ar, As, At, B, Br, C, Cl, F, H, He, I, Kr, N, Ne, O, P, Rn, S, Se, Si, Te, Xe

** The most pertinent example of a connection table hydrogen is a hydride ion ligand coordinated to a transition metal. Some hydrogens attached to nontransition metals are not connection table hydrogens. See Chapter 3 for a more complete discussion.

The BOND Command

As you create a structure, all bonds are unspecified, i.e., they can have any value, until you define them otherwise. The BOND command is used to assign the bond value (single, double, triple, etc.), the bond type (ring or chain), or both the bond value and the bond type.

Table of Bond Values

Symbol	Definition	Text Display	Graphics Display
SE	Single Exact	C - C	C—C
DE	Double Exact	C = C	C=C
T	Triple	C # C	C≡C
N	Normalized	C + C	C---C
S	Single Exact or Normalized	C . C	C---C
D	Double Exact or Normalized	C : C	C=C
U	Unspecified	C ? C	C^C

For illustrative purposes only, the display symbols are shown connecting carbon (C) atoms. Elements other than carbon could have been chosen.

With the BOND command, bond values can be defined as either Unspecified (any bond value allowed), variable (Single or Double which encompasses either Exact or Normalized), or specific (Single Exact, Double Exact, Triple or Normalized). Most metal-ligand bonds can be defined as Single Exact. The bond values for a number of ligands are listed in Appendix I. If you are not certain of a bond value, or if you wish to allow for variation, leave the bonds in the structure Unspecified. However, in order to help insure that your search will run to completion, it is wise to fully specify the bond values for as many bonds as possible.

The BOND command can also be used to define the bond type. The bond type may be Ring, Chain, or variable, i.e., Ring or Chain. When you create a structure, the system assumes a default bond type. Bonds in rings are Ring bonds, and bonds not in rings are Chain bonds. You have the option of overriding the Chain bond default by specifying either a Ring bond or a Ring or Chain variable bond. You can reassign the bond type only for those bonds in a chain.

Building Structures

Six Coordinate Structures

For maximum retrieval of coordination compounds, specify the variable bond type of Ring or Chain.

The following Table summarizes the bond types in the Registry File.

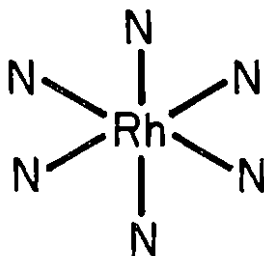
Table of Bond Types

Symbol	Definition
C	Chain
R	Ring
RC	Ring or Chain

Examples Using the GRAPH, NODE, and BOND Commands

Creating a 6-Coordinate Structure

The following steps are used to create the structure shown below.
 (The structures are illustrated as they appear on an ASCII (text) terminal.
 All commands are given at the colon (:) prompt within the STRUCTURE
 command. We have underlined what you would be typing.)



Building Structures Six Coordinate Structures

:GRA C3, DIS

```
C???????C???????C
1      2      3
```

:GRA 2 Cl, 2 Cl, 2 Cl, 2 Cl, DIS

```
      5
     C   C   C
    4 ?   ?   ? 6
      ?   ?   ?
      ???
1 C???????C???????C
      ?2      3
      ?
      ?
      C 7
```

:NOD 2 RH, 1 3 4 5 6 7 N, DIS

```
      5
     N   N   N
    4 ?   ?   ? 6
      ?   ?   ?
      ???
1 N???????Rh???????N
      ?2      3
      ?
      ?
      N 7
```

:BON ALL SE, DIS

```
      5
     N   N   N
    4 -   -   - 6
      -   -   -
      ---
1 N-----Rh-----N
      -2      3
      -
      -
      N 7
```

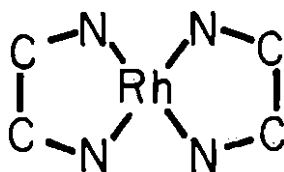
1. Graph a chain of 3 atoms.
All atoms are carbon.
All bonds are unspecified.
Each node is assigned a number.
2. Graph four chains of 1 atom each to node 2.
Atoms are numbered in the order you create them.
3. Specify Rhodium at node 2 and Nitrogen at nodes 1,3,4,5,6,7.
Specified elements display in the structure.
4. Specify all bonds as Single Exact.

Building Structures

Chelate Rings

Creating Chelate Rings (Spiro Ring Fusion)

Use the following steps to build the structure.



:GRA R5, DIS

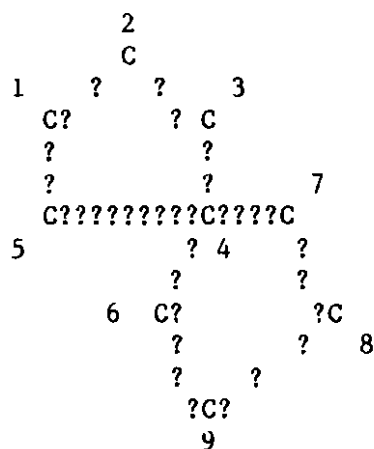
1. Graph a ring of 5 atoms.

```

      2
      C
    1   ?   ?   3
      C?       ? C
        ?       ?
        ?       ?
      C????????C
    5               4
  
```

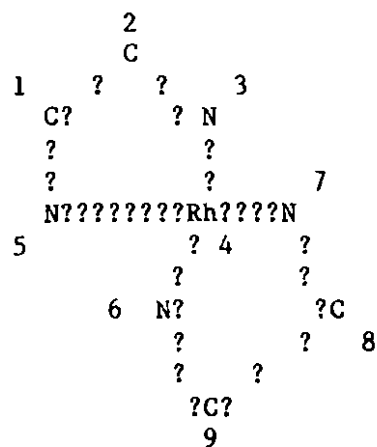
Building Structures Chelate Rings

:GRA 4 R5, DIS



2. Graph a second ring of 5 atoms at node 4. Atom 4 is counted as one of the five atoms in this ring.

:NOD 4 RH, 3 5 6 7 N, DIS

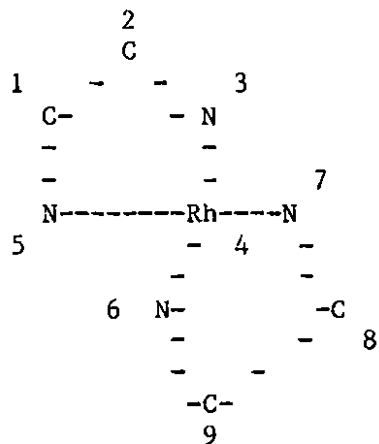


3. Specify Rhodium at node 4 and Nitrogen at nodes 3,5,6,7.

Building Structures
Fused Rings as Ligands

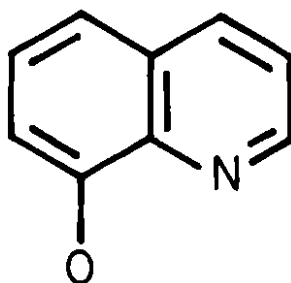
:BON ALL SE, DIS

4. Specify all bonds as
Single Exact.



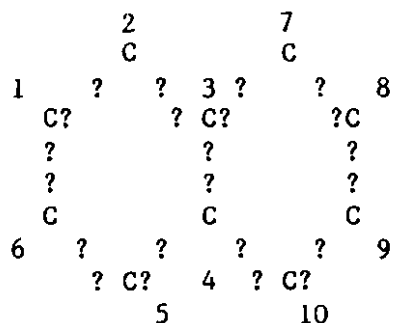
Creating Fused Rings as Ligands

The following steps are used to build the structure.



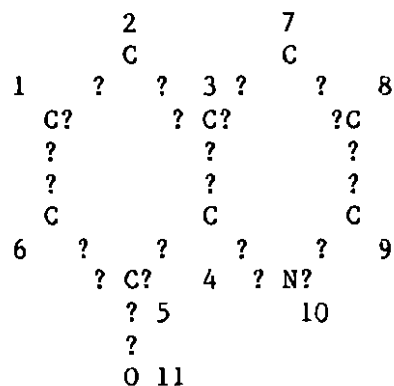
Building Structures Fused Rings as Ligands

:GRA R66, DIS



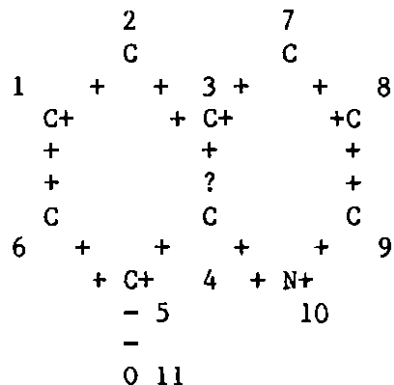
1. Create two fused six-membered rings.

:GRA 5 C1, NOD 10 N, 11 O, DIS



2. Add a chain of 1 atom at node 5.
Define node 10 as Nitrogen and node 11 as Oxygen.

:BON ALL N, 5-11 RCSE, DIS

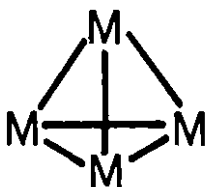


3. Specify all bonds as Normalized.
Respecify the bond between nodes 5 and 11 to Ring or Chain Single Exact to allow this bond to occur either as part of a chelate ring or as a monodentate ligand.

Building Structures Clusters

Creating Clusters

Use the following steps to build this four atom cluster.



:GRA R4, GRA 1 3, GRA 2 4, DIS

```

1              2
  C????????C
  ? ?      ? ?
  ?   ? ? ?
  ? ? ? ? ?
  ? ?      ? ?
  C????????C
4              3
```

1. Create a ring of 4 atoms
Create a bond between
atoms 1 and 3 and between
atoms 2 and 4.

:NOD ALL M, BON ALL SE, DIS

```

1              2
  M-----M
  - - - - -
  - - - - -
  - - - - -
  - - - - -
  M-----M
4              3
```

2. Define all nodes as
any Metal atom.
Specify all bonds as
Single Exact.

When the structure diagram (image) is confusing or garbled or when some of the atoms and/or bonds overlap, you can look at the Structure Connection Table (SCT) to ascertain the nodes and bonds. The SCT lists each node in the structure and the bond type and bond value between that node and all other nodes to which it is bonded. Displaying the Structure Connection Table is often the best way to be certain what nodes and bonds exist in the structure. The following two examples illustrate using the SCT to verify the bonding in the query structure.

Creating Cyclopentadienyl Ligands

Use the following steps to create the structure.



:GRA R5, C1, 6 1, 6 2, 6 3, 6 4, 6 5, DIS

```

      2
      C
    ? ? ?      3
1 C?  ?  ? C
  ?  ?  ??
  ?  ?  ??
5 C????????C
  ?  ?  ? ?  4
  ?  ?  ? ?
  ?  ?  ? ?
  ? ?  ? ?
  ?  ?  ??
  ? ?  ? ?
  ? ? ??
  ?? ??
  ?? ?
  ???
  C?

```

1. Create a ring of 5 atoms.
Create a separate chain of
of 1 atom.
Create a bond from atom
6 to each of the ring
atoms.

Building Structures Macrocyclic Ligands

:NOD 6 M, DIS SCT

2. Define node 6 as any metal atom.

:DIS SCT

3. Display Structure Connection Table to verify bonding.

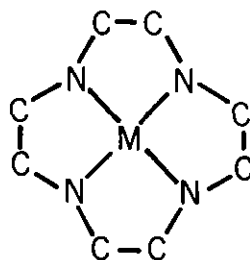
```

*****CONNECTIONS*****
NOD SYM      NOD/BON NOD/BON NOD/BON
  1 C        6 RU     2 RU     5 RU
  2 C        6 RU     3 RU     1 RU
  3 C        6 RU     4 RU     2 RU
  4 C        6 RU     5 RU     3 RU
  5 C        6 RU     1 RU     4 RU
  6 M        5 RU     4 RU     3 RU
              2 RU     1 RU

```

Creating Macrocyclic Ligands

Rings containing more than eight atoms, such as the one shown below, are created from chains.



Building Structures
Macrocyclic Ligands

:GRA C12, 1 12

1. Graph a chain of 12 atoms.
Connect the terminal nodes,
1 and 12.

:NOD 1 4 7 10 N,DIS

2. Define nodes 1,4,7,10 as
Nitrogen.
Structure display does not
show the bond between 1 and 12.

```
1  N????C????C????N????C????C????N????C????C????N????C????C 12
   2      3      4      5      6      7      8      9      10     11
```

:DIS SCT

3. Display the Structure
Connection Table to verify
that all bonds are Ring bonds
and that a bond exists
between atoms 1 and 12.

```
*****CONNECTIONS*****
NOD SYM      NOD/BON  NOD/BON  NOD/BON
  1 N          12 RU      2 RU
  2 C           3 RU      1 RU
  3 C           4 RU      2 RU
  4 N           5 RU      3 RU
  5 C           6 RU      4 RU
  6 C           7 RU      5 RU
  7 N           8 RU      6 RU
  8 C           9 RU      7 RU
  9 C          10 RU      8 RU
 10 N          11 RU      9 RU
 11 C          12 RU     10 RU
 12 C           1 RU     11 RU
```

Building Structures
Macrocyclic Ligands

:GRA C1, 13 1, 13 4, 13 7, 13 10

4. Create a separate chain of one atom (node 13).
Create a bond between 13 and each of the Nitrogen nodes, 1,4,7,10.

:NOD 13 M, DIS SCT

5. Specify node 13 as any Metal atom.
Display Structure Connection Table to verify that 13 is M and that it is bonded to 1,4,7, and 10.

```

*****CONNECTIONS*****
NOD SYM      NOD/BON NOD/BON NOD/BON
  1 N        13 RU   12 RU   2 RU
  2 C         3 RU    1 RU
  3 C         4 RU    2 RU
  4 N        13 RU    5 RU   3 RU
  5 C         6 RU    4 RU
  6 C         7 RU    5 RU
  7 N        13 RU    8 RU   6 RU
  8 C         9 RU    7 RU
  9 C        10 RU    8 RU
 10 N        13 RU   11 RU   9 RU
 11 C        12 RU   10 RU
 12 C         1 RU   11 RU
 13 M        10 RU    7 RU   4 RU
              1 RU

```

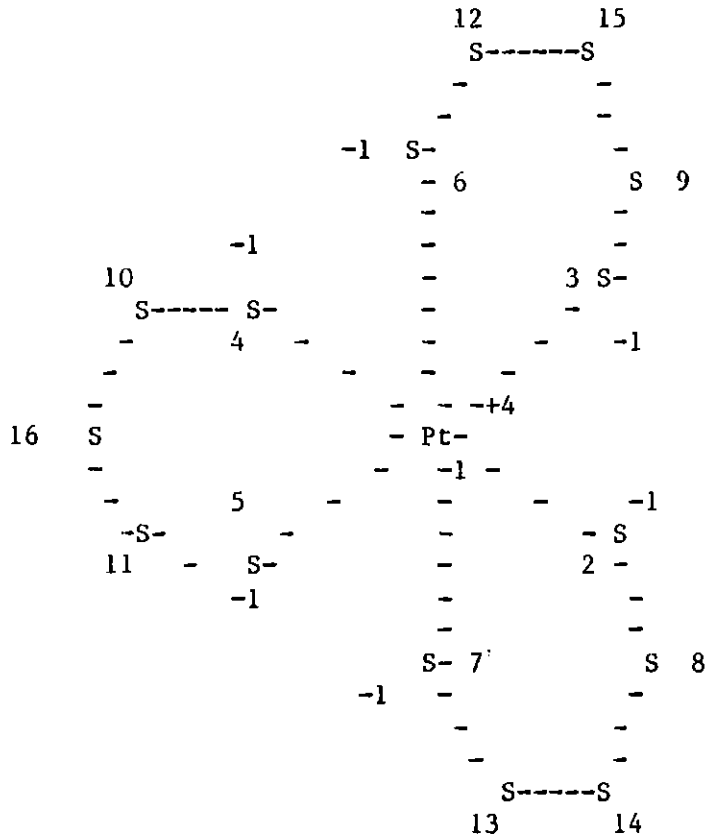
Building Structures Registry Numbers as Models

The RECALL Command - Using Models to Build Structures

The RECALL command is used to call up a copy of a structure to use as the starting point for a new structure. This may be the L-number of a structure you built earlier in the session, the Registry Number of a structure you want to use as a model, or the code name of a structure from the Fragment File. After displaying the recalled structure, you can use any of the structure building commands to modify it to meet your search requirements. A list of structures in the Fragment File useful for coordination compounds is contained in Appendix 3.

Recall a Registry Number

:REC 17611-65-3, DIS



Building Structures
Fragment File Models

Recall and Modify a Structure from the Fragment File

:REC PORPHN, DIS

1. Recall and display the
fragment.

```

      2      3      8
      C      C      C
    ? ? 1 ? ? 6 ? ?
    ? ? ? C? ? C? ? 11
5 C? ? ? ? ? C
    ? ? ? ? ?
    ? ? ? ? ?
    7 C????N???H 9 N????C? 12
    ? 4 25 ?
10 C ? ?C 14
    ? ?
    ? 16 21 ?
13 C????N 26H???N????C?
    ? ? ? ?
    ? ? ? 17 ?
15 C? 19 ? ? 24 C 20
    ? ? C? ? C? ?
    ? ? ? C ? ? ?
      C      C
    18      22      23

```

Building Structures Fragment File Models

:DEL 25 26, GRA C1, DIS

2. Delete the H nodes.

Create a chain of one atom
for the metal node.

```

      2      3      8
      C      C      C
    ? ? 1 ? ? 6 ? ?
    ? ? ? C? ? C? ? 11
5 C? ? ? ? ? C
  ? ? ? ? ? ?
  ? ? ? ? ? ?
  7 C?????N 9 N????C? 12
    ?      4      ?
10 C      ?C 14 C 25
  ?      ?
  ?      16 21 ?
  13 C?????N N????C?
    ? ? ? ? ?
    ? ? ? 17 ?
15 C? 19 ? ? 24 C 20
  ? ? ? C? ? C? ?
  ? ? ? C ? ? ?
    C      C
    18 22 23

```

Building Structures
Fragment File Models

:GRA 25 4, 25 9, 25 21, 25,16

3. Create bonds between node
25 and nodes 4,9,16,21.

:NOD 25 CO, DIS

4. Specify node 25 as Cobalt.

```

      2      3      8
      C      C      C
    ? ? 1 ? ? 6 ? ?
    ? ? ? C? ? C? ? 11
5 C? ? ? ? ? C
    ? ? ? ? ?
    ? ? ? 4 ? ?
    7 C?????N 9 N????C? 12? ? ? ?
    ? ? ? ? ? ? ?
10 C ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
    ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
    ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
    13 C?????N 16 21 ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
    ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
    ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ?
15 C? 19 ? ? ? 24 C ?20
    ? ? ? C? ? C? ? ?
    ? ? ? ? C ?? ? ?
      C      C
    18      22      23

```


Building Structures Variability in Structures

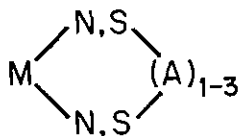
Variability in Structures

When you create structures for retrieving coordination compounds, you can define the nodes either generically or very specifically. You can define nodes as being specific elements, e.g., a node is a Nitrogen atom. You can use generic node symbols to introduce variability, e.g., a node is a Metal atom. Even more variability can be built into the structure using the variable symbols G1-G20, (collectively referred to as Gk). These represent from 2-20 optional definitions of your choice at a node. A Gk symbol may be used more than once in the structure. To specify multiple atoms or groups of atoms at a node, first use the NODE command to assign a G-group, e.g., G1, to the appropriate nodes in your structure. Then, use the VARIABLE or REPEATING commands to define the Gk group. The VARIABLE command (VAR) is used to

The VARIABLE command (VAR) is used to define the atoms or groups of atoms that are to be used in a Gk group. This command allows you to specify that up to 20 different values be present in a Gk group. These values can be elements, shortcut symbols, the system-defined variable atoms, or nodes in separate multiple-atom fragments built along with your structure.

If the variability at a node can be expressed in terms of a repeating group, the REPEATING command (REP) can be used to define the repeating fragment and specify the number of times it can repeat.

Allowing for Variable Donor Atoms and for Variable Chelate Backbone



:GRA R4, NOD 1 M, 2 4 G1, 3 G2, DIS

```

1           2
M?????G1
?         ?
?         ?
G1?????G2
3           4

```

1. Graph a ring of 4.
Define nodes 2,4 as G1.
Define node 3 as G2.

Building Structures Variability in Structures

:VAR G1=N/S, REP G2=(1-3) A

```

1           2
  M?????G1
  ?       ?
  ?       ?
  G1????G2
4           3

```

VAR G1=N/S
REP G2=(1-3) A

2. Define G1 as Nitrogen
or Sulfur.
Define G2 as Any atom
repeating 1 to 3 times.

Defining a Central Metal

The central metal may be specified as an individual metal, e.g., NOD 1 RH. It can be defined as multiple metals, e.g., NOD 1 G1, VAR G1=CR/MO/W. Or, with the broadest of definitions, the central metal can be defined as any metal, e.g., NOD 1 M.

Defining Ligand Atoms

Ligand atoms can be specified as being one element, e.g., NOD 2 N. They can be defined as having multiple values, e.g., NOD 2 G1, VAR G1=N/P/S. They can be allowed to be any atom, e.g., NOD 2 A. Or, they can be specified as being any atom but one, NOD 2 -S.

Summary

In this chapter we have illustrated the techniques used to create basic structures for coordination compounds. The next chapter will consider other ways of specifying nodes and bonds and adding attributes in order to retrieve more specific answers.

CHAPTER 5: ADDING ATTRIBUTES TO STRUCTURES

Once you have created a basic structure, you can tailor it to better meet your search requirements by adding attributes to selected nodes. Attribute commands are used to indicate special or unusual structural features that are not visible in the skeleton of the structure. The proper selection of attributes will insure complete retrieval of all the desired answers and/or will narrow the answer set to include exactly the type of complexes you want.

Attributes particularly useful for coordination compounds are those which indicate the presence of a certain number of hydrogen atoms on a node (HCOUNT), the number of substituents desired on a node (CONNECT), whether a node may be part of a ring or a chain (NSPEC), and whether all rings must be isolated or if they may be allowed to have other rings fused to them (RSPEC). Two of these, CONNECT and NSPEC, can also be combined with the BOND command to specify bond types that further refine search strategy.

COMMAND	DEFINITION
CONNECT (CON)	Specify the number and type of connections to a node. May be Exact (E), Minimum (M), or Maximum (X). May be Ring (R), Chain (C), or Ring or Chain (RC).
HCOUNT (HCO)	Specify the number of hydrogens attached to a node. May be Exact (E) or Minimum (M).
NSPEC (NSP)	Specify whether a node is in a ring, chain, or either. May be Ring (R), or Chain (C), or Ring or Chain (RC).
RSPEC (RSP)	Specify that a single ring, a ring system, or all rings in query must be isolated.

HCOUNT and RSPEC Uses

The HCOUNT (Hydrogen Count) Command

The HCOUNT command is used to specify the number of hydrogens, including hydrides, attached to a node. Either the Exact number of hydrogens (0 to 99) or the Minimum number (1 to 99) may be specified.

To use the HCOUNT command the format is:

HCO n ... Em or HCO n ... Mm

where n ... represents the node number(s) where the hydrogen count is to be assigned, Em is the Exact number of hydrogens present at that node, and Mm is the Minimum number of hydrogens allowed. Hydrogen counts may not be used on nodes defined as Gk groups. If no hydrogen count is specified, substances with all possible numbers of hydrogens, including none, will be retrieved.

Detailed examples using hydrogen counts to retrieve metal hydrides are given in Case Study 6 beginning on page 133.

The RSPEC (Ring Specification) Command

The RSPEC command allows you to specify that you will accept only substances in which the rings that you have included in your substructure are isolated, i.e., no other ring fusion is allowed. If you do not specify that the rings are isolated, it is assumed that you will allow them to be either isolated or embedded.

In coordination compounds, the existence of two or more chelate rings at a metal center constitutes ring fusion, with the metal atom as a spiro node:



You would not ordinarily use the ring specification attribute when searching for coordination compounds, unless you are certain that you wish to prohibit any further chelation on the metal.

The CONNECT Command

When you search a substructure, the answers you retrieve include all of the atoms you have included in the query. Unlimited additional connections may be present anywhere in the structure. The CONNECT command is used to restrict substitution at one or more nodes in the query structure.

The CONNECT command specifies the number of non-hydrogen connections (it ignores any hydride ligands) to a node and the type of bonds to those connections. With CONNECT you can assign the Minimum, Exact, or maximum number of connection table atoms at any node in the structure. You can also specify whether the atoms are connected by Ring bonds only, Chain bonds only, or the default option of either Ring or Chain bonds. The number of connections specified by the CONNECT command includes the connections already shown in the structure, as well as any additional ones required by the CONNECT command.

To use the CONNECT command, the format is:

CON n... Em bt

where n... represents the node number(s) where the connections are to take place, Em is the Exact number of attachments (0 to 16) allowed, and bt is the bond type (R, C, or RC) for these connections. If you do not specify a bond type, either Ring or Chain (RC) is assumed. Alternatively, you can specify the Minimum number or the maximum number of attachments (0 to 16).

User Input	Definition
En	Exactly n Ring or Chain attachments
Mn R	Minimum of n Ring attachments
Xn C	MaXimum of n Chain attachments

CONNECT

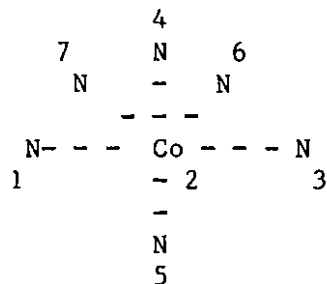
Retrieving Specific Coordination Numbers

Using CONNECT to Retrieve Specific Coordination Numbers

One or two coordination numbers dominate the chemistry of each transition metal. CONNECT gives you a way of finding the less common coordination numbers without getting many irrelevant answers. It can help make a search more efficient by reducing the number of substances that find their way into answer sets because the system was unable to match them with the query in the time allowed for iterating an individual structure.

It should be noted that CONNECT operates at the iteration step, not at the screen level. Using the CONNECT attribute has no effect on the number of iterations the search system will have to perform. When you use CONNECT to search for unusual coordination numbers on a metal, e.g. cobalt, that has a large numbers of six coordinate complexes, the substructure you create must be quite specific in order to get your search to run to completion because a potentially large number of the six-coordinate structures may have to be iterated. CONNECT E5 (Exactly 5) or E7 (Exactly 7) will exclude six coordinate complexes as answers since they have a CONNECT of E6 (Exactly 6). But, all of the six coordinate complexes will be passed through the screen step and will be eliminated when they do not match at the iteration step.

Cobalt is more likely to have more than six ligands. In seeking examples of higher coordination numbers for cobalt, a search was run using the structure below which includes a CONNECT value of a Minimum of 7 ring or chain attachments to the cobalt. The six attachments already present in the query structure are counted in the minimum of seven. Therefore, there must be six nitrogen donors, but the additional ligands may include any element as the donor atom.



:CON 2 M7, BON ALL RCSE

Using CONNECT to Retrieve Metallocenes

$$\begin{array}{ccccccc}
 & & 12 & & & & 2 \\
 & & \text{C} & & & & \text{C} \\
 11 & - & & - & 13 & & 1 & - & & - & 3 \\
 \text{C} - & & & & - \text{C} & & & & & & \text{C} \\
 = & & & & = & & & & & & = \\
 = & & & & = & & & & & & = \\
 \text{C} - - - - \text{C} - - - - \text{M} - - - - \text{C} - - - - \text{C} \\
 15 & & 14 & & & & 5 & & & & 4 \\
 & & & & - & & & & & & \\
 & & & & - & & & & & & \\
 & & & & - & & & & & & \\
 & & & & - & & & & & & \\
 & & & & - & & & & & & \\
 & & & & - & & & & & & \\
 & & & & - \text{C} - & & & & & & 7 \\
 & & 6 & \text{C} - & & - & \text{C} & 8 \\
 & & = & & & & = & \\
 & & = & & & & = & \\
 & & = & & & & = & \\
 & & \text{C} - - - - \text{C} & & & & & \\
 10 & & & & & & & 9
 \end{array}$$

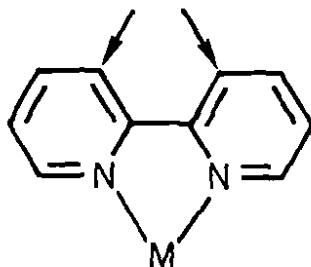
:BON ALL R, CON 16 M15 R

CONNECT Isolated Rings

Using CONNECT in Ligand Structures

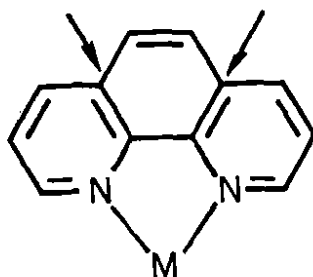
The CONNECT command can be used to isolate ligand ring system. For example, if you wish to search for complexes of bipyridine without retrieving phenanthroline complexes at the same time, you can block further substitution of the bipyridine with a CONNECT attribute. Bipyridine has exactly two ring bonds at the indicated carbon atoms,

CONNECT E2 R



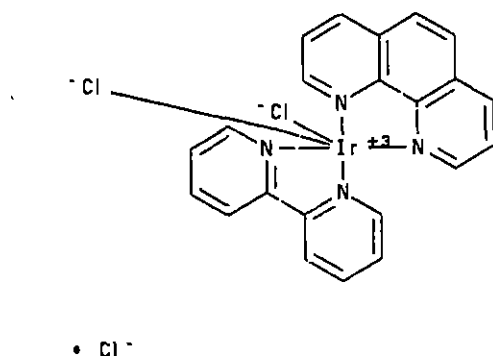
while phenanthroline has three ring bonds at that position.

CONNECT E3 R



CONNECT Metal Hydrides

RSPEC should not be used for this example because it could block further chelation on the metal. RSPEC would cause the loss of answers such as



Metal Hydrides

Whenever you use CONNECT on a metal node, you may get some irrelevant answers containing hydrides. Since CONNECT does not count hydrogens, the attribute is ignored whenever a complex contains a connection table hydrogen on the same node where the CONNECT Attribute was assigned. Some of the hydride answers may well have the designated coordination number, but others will not.

If you are searching for metal hydrides, do not use a CONNECT on the metal node. You may lose valid answers, especially those containing nontransition metals.

NSPEC and Bond Types Definitions

The NSPEC (Node Specification) Command

The NSPEC command allows you to specify whether a particular atom in your structure is in a ring, in a chain, or in either. Atoms that are in a ring are ring atoms and their node specification cannot be changed. On the other hand, atoms that are in a chain are chain atoms but can be redefined as either ring atoms or as ring or chain atoms by using the NSPEC command.

To use the NSPEC command, the format is:

NSP n... value

where n... represents the node number(s) of the nodes to be redefined, and value is R (Ring), C (Chain), or RC (Ring or Chain).

Effect of Bond Type on Node Specification

The bond type in the query structure determines the node specification for each node on either side of the bond. When a bond is assigned as Ring, the nodes on either side of the bond are automatically defined as Ring nodes. When the bond type Chain is assigned, the nodes on either side of the bond are automatically defined as Chain nodes. In the most general case, and the one that allows for maximum retrieval, the bond type is defined as Ring or Chain. This strategy not only allows either type of bond, but also allows the atoms on either side of the bond to be either ring or chain nodes.

NSPEC and Bond Types
Case 1 - System Defaults

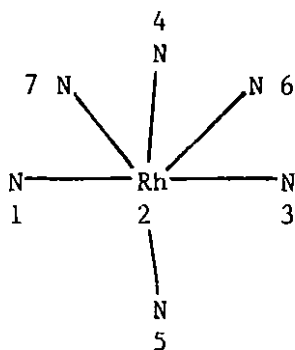
Table of Node Specification Defaults

Bond type	Node specification default for nodes on either side of bond
C	Chain----Chain
R	Ring-----Ring
RC	Chain----Chain or Chain----Ring or Ring-----Chain or Ring-----Ring

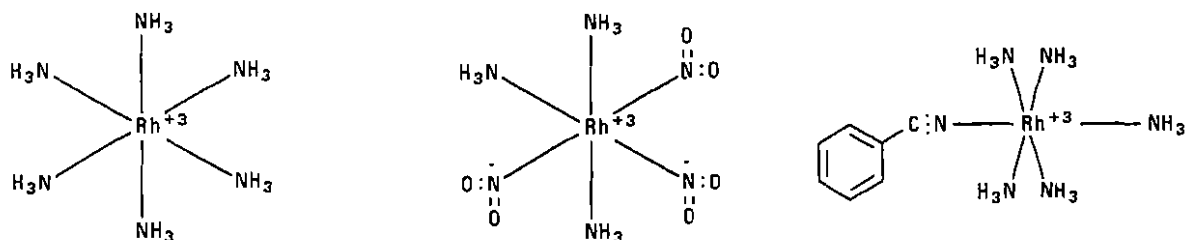
NSPEC and Bond Types
Case 1 - System Defaults

Effect of Bond Type and Node Specification on Coordination Compound
Retrieval

Specification of bond types and of node specifications is critically important when searching for coordination compounds. To illustrate the effects of different node and bond specifications, we will search for rhodium complexes containing six nitrogen donor atoms.



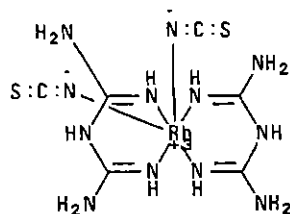
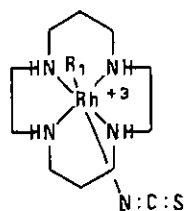
Case 1 (:BON ALL SE). This is the "what you see is what you get" default. The bonds and nodes shown above will automatically be assigned exactly as you have created them, i.e., all the nodes are chain nodes and all of the bonds are chain bonds. The bonds are specified with the bond value of Single Exact. This selection will retrieve complexes with only monodentate ligands which are not part of a cyclic system. Examples of complexes retrieved with this strategy include:



Note that it is possible to have two or more different types of ligand molecules coordinated to the same metal atom using this kind of bond and node specification.

NSPEC and Bond Types
Case 2 - Ring or Chain Bonds

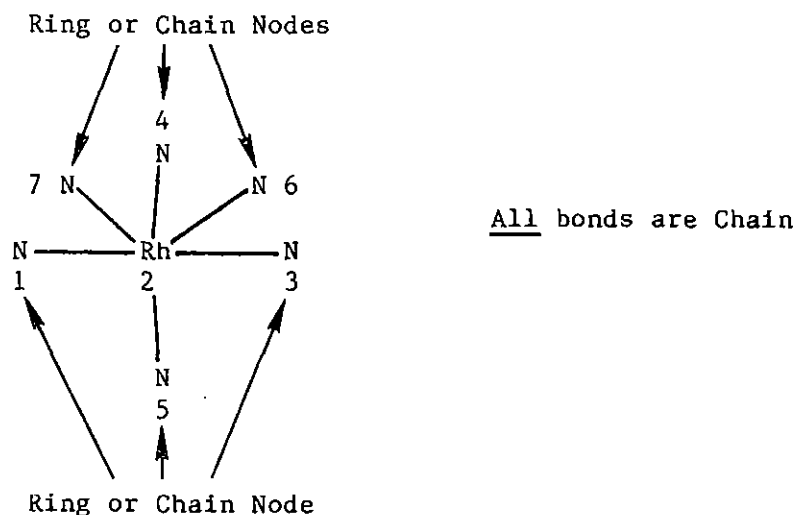
Case 2 (:BON ALL RCSE). This is the most general case and will retrieve the most answers because it allows all bonds in the structure shown on the previous page to be either Ring or Chain bonds and all nodes to be either in a Ring or in a Chain. It includes all examples of rhodium complexes containing six nitrogen donors, including all those retrieved in the first example. Examples of complexes retrieved with this strategy include:



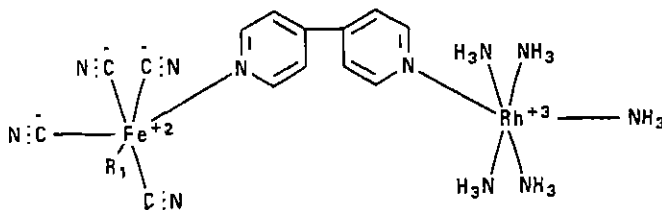
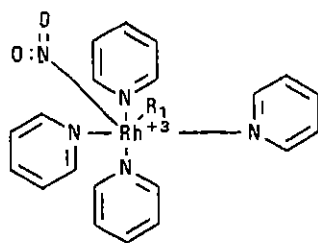
NSPEC and Bond Types

Case 3 - Bonds Chain, Nodes Ring or Chain

Case 3 (:BON ALL SE, NSP 1 3 4 5 6 7 RC). In this example, using the structure



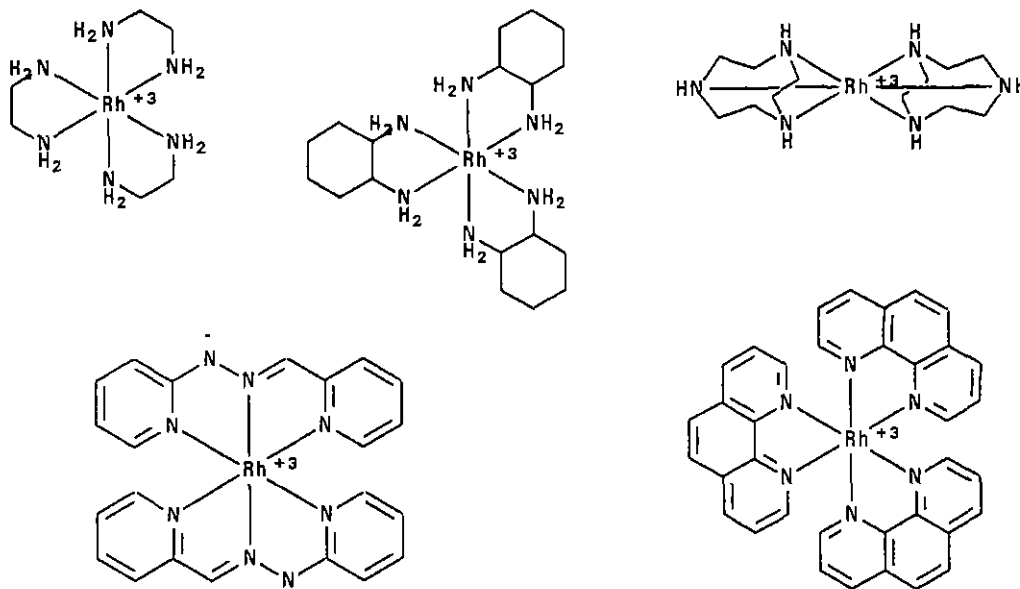
all the the bonds are chain bonds by default. We have also specified that all of the Nitrogen atoms may be either in a Ring or in a Chain. As in Case 1, only complexes with monodentate ligands will be part of the answer set. In addition to the complexes found in Case 1, this search will also retrieve complexes in which the nitrogen donor is part of a ring. Examples of complexes retrieved with this strategy include:



If you want to exclude ligands chelating through the N atom, these are the appropriate bond and node types to specify.

NSPEC and Bond Types
Case 4 - All Bonds are in Rings

Case 4 (:BON ALL RSE). In this case, we have specified all of the bonds in the structure shown on page 52 as being Ring bonds. All nodes are now automatically Ring nodes. No node specification is needed. Only complexes in which all ligands are part of chelate rings will be retrieved. Examples of complexes retrieved with this strategy include:



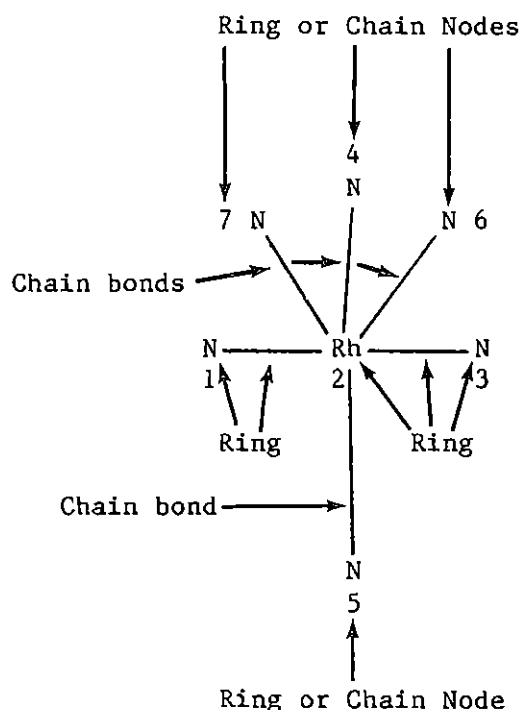
Note the ring bonds may be formed by three bidentate ligands, two tridentate ligands, one tetradentate and one bidentate ligands, or one hexadentate ligand. Also note that the chelating ligands may have either saturated or aromatic carbon rings for backbones, or saturated carbon chains, or there may be additional heteroatoms in the backbone.

NSPEC and Bond Types

Case 5 - One Chelate Ring

You can designate specific combinations of ring and chain bonds and node specifications to require that retrievals be just the type of complexes you are interested in. The next three cases show what is retrieved by overriding the defaults at selected bonds and nodes.

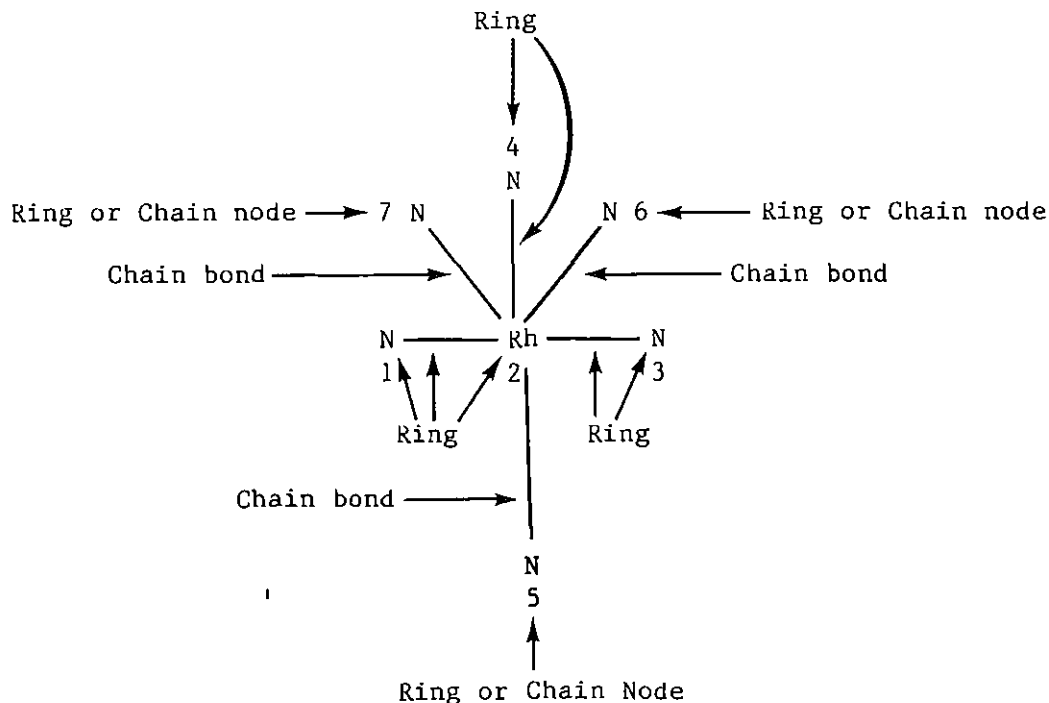
Case 5 (:BON ALL SE, 1-2 2-3 R, NSP 4 5 6 7 RC). For this case, we have used the following structure.



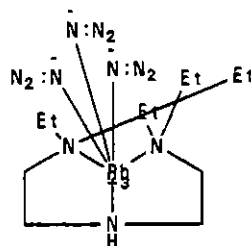
First, the system default of Chain is taken for all of the bonds in the structure. Then, two of the bonds are respecified as being in a Ring. This means that nodes 1, 2, and 3 are Ring nodes automatically. We also override the system default of Chain node for atoms 4, 5, 6, and 7 by specifying that they may occur either in a Ring or in a Chain. Remember, NSPEC does not effect bond types. Therefore, the bonds 2 to 4, 2 to 5, 2 to 6, and 2 to 7 are still Chain bonds. This strategy will retrieve rhodium-nitrogen complexes with exactly one chelate ring. To date, there are no examples of this type of complex in the Registry File.

NSPEC and Bond Types
Case 6 - One Tridentate Ligand

Case 6 (:BON ALL SE, 1-2 2-3 2-4 R, NSP 5 6 7 RC). In this case, for the structure



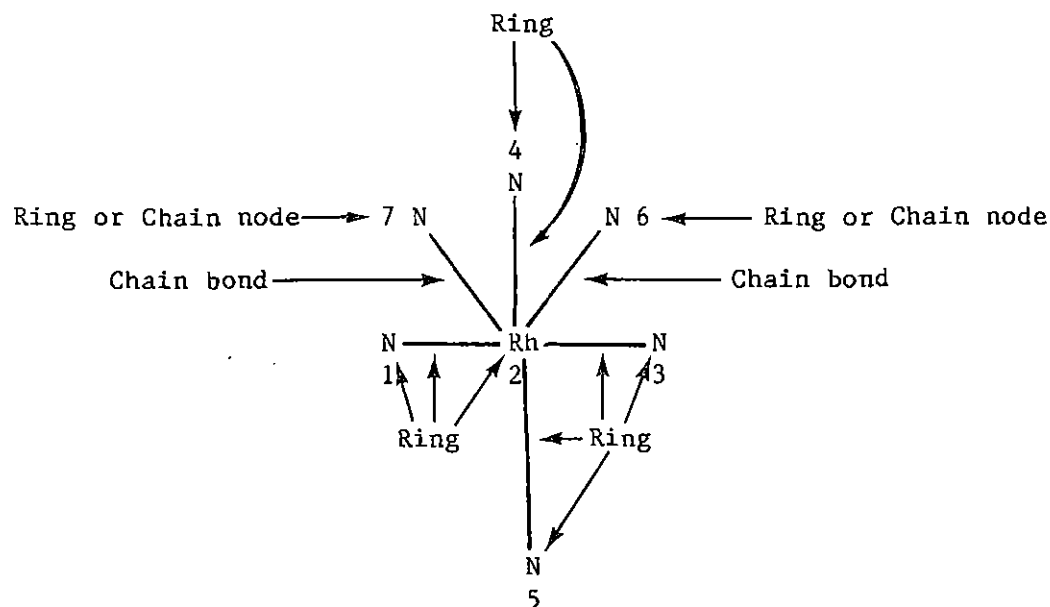
three of the bonds are specified to be in a Ring, i.e., 1 to 2, 2 to 3, and 2 to 4. This automatically requires that nodes 1, 2, 3, and 4 are nodes in Rings. Additionally, nodes 5, 6, and 7 are specified as being either in a Ring or in a Chain. Since the system default was taken for the remaining bonds, bonds 2 to 5, 2 to 6 and 2 to 7 are Chain bonds. This strategy will retrieve complexes with exactly one tridentate ligand. An example of the type of complex retrieved is



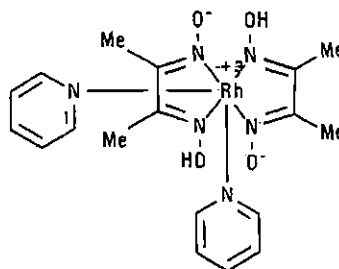
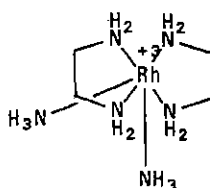
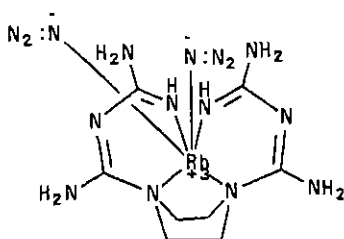
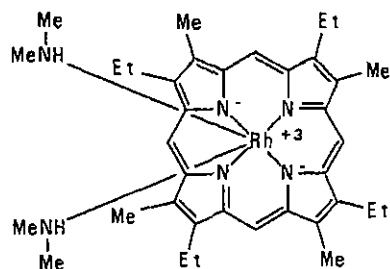
NSPEC and Bond Types

Case 7 - Bidentate or Tetradentate Ligands

Case 7 (:BON ALL SE, 1-2 2-3 2-4 2-5 R, NSP 6 7 RC). In this case, using the structure



the system default of bond type Chain is taken for the bonds between 2 and 6 and 2 and 7. All of the other bonds are respecified as being Ring bonds. This automatically makes the nodes 1, 2, 3, 4, and 5 Ring nodes. Nodes 6 and 7 are specified as Ring or Chain nodes using the NSPEC command. Remember that this does not change the bond type. This will retrieve Rh-N complexes with either two bidentate ligands or one tetradentate ligand, which may be either open chain or macrocyclic.



In Cases 5, 6, and 7, the bonds that were not specified as Ring bonds are chain bonds by default. Specifying them as ring or chain (RC) will retrieve not only answers which have a minimum of the requested chelating ligands, but will also get the answers that were included in Case 4 where all bonds were ring, i.e., you have lost the specificity of the search.

Summary

In this chapter, various attribute commands were described in terms of their usefulness in retrieving coordination compounds. These attributes are usually used in conjunction with specific bond types to specify the complexes desired. CONNECT and NSPEC are especially useful for refining coordination searches.

In the next chapter we will discuss the use of screens in coordination search strategy.

CHAPTER 6: ADDING SCREENS TO REFINE YOUR SEARCH

The previous chapters have considered how to create accurate and precise substructures to use as search queries to retrieve coordination compounds. This chapter deals with making the online search of the substructure you created more efficient by adding screens, fragment codes, to the search logic. Learning to use screens is one of the finer points of substructure searching; but, using them is often crucial to getting your coordination compound search to run to completion. This chapter will suggest which additional screens are most useful to add and how to select them. For more information on screens, see the search aid, Using CAS ONLINE: Adding Screens in Structure Searching.

After you have created your substructure and initiated the search, the structure is automatically described in terms of numeric codes, called screen numbers, for the various structural features you have specified. The computer search process operates in two stages, first the screen step and then the iteration process. The screen numbers are used to "screen out", or eliminate from further consideration, all substances in the Registry File which do not possess screen numbers for the structural features you have built into your query structure. The iteration process, an atom-by-atom bond-by-bond comparison of the database compounds to your query structure, is performed on the candidate substances which possess the appropriate screens. Those that match all of the required features of your substructure are the answers you retrieve.

This entire search process happens automatically after you enter the SEARCH command. For many searches you do not have to be concerned with which screens are generated by the system. In the interest of efficient use of computer resources, the search machine is programmed to run the screen search on the basis of choosing only the statistically most unusual screens and to limit the number of iterations performed and the time spent on them.

Screens

The Search Process

For many query structures a sufficient number of distinctive screens are automatically generated to enable the search to run to completion. With coordination compound structures, however, the number and variety of screens available is more limited because many of the screens available pertain only to organic ligands and bond sequences and number of connections that are independent of the element present. Relatively few screens describe metal centers.

In searching for coordination compounds you may find that a sample search indicates that the full-file search will not complete or is uncertain because the system limit for iterations will be exceeded, or that too many answers which do not meet the intent of the search are being retrieved. You can often overcome these difficulties by adding additional screens to the query.

Before adding screens, you should build your structure as specifically as possible, within the requirements of your question. Exact bond values and bond types should be specified for as many bonds as possible. The only way you can reduce the number of iterations is by creating a more specific substructure or by adding screens. Attributes such as hydrogen count, connect, charge, or valence operate only at the iteration level, after the substances have passed screening. Creating additional substructures to be NOTed out does not reduce the number of iterations either. These latter techniques help only to make the answers more specific.

Full file projections on coordination compound queries are likely to be less accurate than those made for many other kinds of searches because coordination compounds make up only a small percentage of the file. When they do occur, a whole group of similar compounds may have adjacent Registry Numbers since they are reported in one paper. You may find it useful to perform more than one sample search if the projection indicates the search will be just over or just under the system limits. This is especially true as you are trying to decide if the addition of one or more screens will be helpful in getting your search to complete. When you add multiple screens using one or more SCREEN commands, bear in mind that if you have used more than one substructure in the query, the entire screen expression must apply to all structures in the query.

Several factors commonly occur which require you to add screens to your coordination compound search.

- Some ligands, even when completely described by exact bonds and hydrogen counts, are small molecules containing common atom sequences.
- When searching for a class of ligands, it is possible to specify only the atoms in the immediate vicinity of the metal. This is a very generic search, particularly if you have not specified a specific metal atom.
- If bonds are left unspecified because the ligand may attach itself to the metal in a variety of ways, or you are uncertain of the ways the bonds ought to be specified, or you suspect that a ligand could have multiple structure representations, only very generic screens will be generated for the portion of your structure which contains the unspecified bonds.
- Your substructure has a feature which, while common in organic molecules, is very distinctive when part of a coordination compound. Since the search machine generally uses only the rarest of the screens it generates, i.e., those of less than 10% frequency, it may not include the screens for the feature of interest. If you wish to emphasize the presence of a feature which is described by a screen which occurs more frequently than 10%, add that screen manually.

To illustrate, Augmented Atom (AA) screens denoting either three ring bonds or three chain bonds to a central atom occur very frequently in organic compounds, so they are usually ignored by the search machine. Yet, requiring three ring ring bonds or three or more chain bonds in a coordination compound substructure can add a very precise piece of information to a coordination compound search. Furthermore, if these features occur at the same metal center, a very distinctive combination of high frequency screens is present. Adding those screen numbers to your query may make the search much more efficient.

- If the generic metal M node is used, the presence of a M node is not detected until the iteration step. The only screen available to describe the presence of a metal is routinely ignored. Manually adding screens for the metals you require improves the efficiency of the search.

Screens

Metal Screens for M Node Symbol

Screens That Are Often Useful

Adding Metal Screens when Structure Contains the Generic M Node Symbol

Generic Element Count (EC) screens for metals are available. They are based on location of the metal in the periodic table, both by horizontal row and by vertical group. A full listing of these screens is provided in Appendix 3. As many of these as are required to define the intent of your search may be combined into a screen set and then added to the search logic.

The most important of these screens to a coordination compound search are:

- 1964 Transition Series 1
- 1921 Transition Series 2
- 1931 Transition Series 3
- 1957 Lanthanide Series
- 1919 Actinide Series
- 1918 Metals (includes all elements except Ar As At B Br C Cl F H
He I Kr N Ne O P Rn S Se Si Te Xe)

Several of the metals which have very extensive coordination chemistries have their own Element Count screens:

- 1975 Iron
- 1963 Cobalt
- 2002 Nickel
- 2020 Platinum

Screens Screens for Ring or Chain Bonds

When you use the generic M node, it is recommended that you routinely add one or more EC screens to the search logic. This is not necessary if your substructure contains a node for a specific metal or if you have defined a variable Gk group containing several metals. Under these circumstances, the appropriate EC screens will be generated automatically.

Which EC screens to add to a query depends on how extensive you want the search to be. Screen 1918, the screen number for all metals, is the most comprehensive and includes non-transition as well as transition metals. Using it will get you the most retrievals, but in some searches you may find a lot of those answers are nothing more than sodium or other alkali or alkaline earth salts that are not really coordination compounds. For transition metal complexes, the combination of Screens 1964 or 1921 or 1931 is often very effective. You must note however that Zn, Cd, and Hg are not included in this screen set. If you want those metals, too, but not any others, include Screen 1956 for periodic group Ib. For most of the searches illustrated in this booklet, the three screens for the Transition Series have been selected, but you should add whichever screens best describe the metals you want to retrieve.

Adding Screens That Specify Minimum Numbers of Ring or Chain Bonds

Generic Augmented Atom (AA) Screens are descriptions of an atom, its immediate attachments, and the type of bonds to those attachments. These AA Screens are listed in Appendix 3. They describe any atom (A) in the compound. This means that they are not restricted to a description of the metal coordination site but serve well to describe these sites.

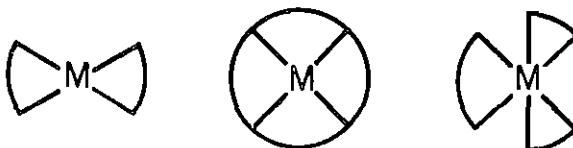
In the table of screens, the A represents Any atom, a star (*) indicates a ring bond and a dash (-) indicates a chain bond. The central atom is cited first, followed by the ring attachments (*) and then the chain attachments (-).

AA screens are generated automatically, but many of them occur too frequently to be used by the search machine. None of the AA screens are automatically generated for the metal center when you use a generic M atom in your structure. It can be helpful to add AA screens to your query because they are effective in reducing the number of iterations.

Screens

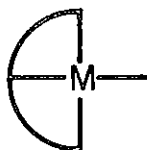
Screens for Ring or Chain Bonds

Two very useful screens of this type are Screens 945 and 947. Screen 945, A*A*A*A*A, describes an atom (A) with four or more ring bonds (*) to any other atom (A), as shown below.



SCR 945

Screen 947, A*A*A*A-A, describes an atom (A) with a minimum of three ring (*) bonds and one chain (-) bond to any atom (A), as shown below.



SCR 947

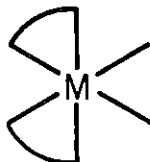
Since an AA screen only describes the minimum number of bonds required, two or more may be combined with AND logic in the same screen set to more completely describe a particular bonding arrangement. For example, in the figure below, Screen 940 describes three ring bonds and Screen 944 describes three chain bonds. Each occurs alone with high frequency, but together they are quite distinctive.



SCR 940 and 944

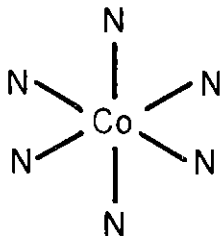
Screens
Screens for Ring or Chain Bonds

The combination of augmented atom definitions may overlap. In the figure below, Screen 945 covers the four ring bonds and Screen 947 specifies at least three ring and one chain bond.



SCR 945 and 947

A case illustrating the benefits of adding AA screens is a search for cobalt complexes of six or more nitrogen ligands. This is similar to the search for all rhodium-nitrogen complexes discussed in Chapter 4. The main difference is that the iteration limit is reached when searching for structures such as Structures A or B, because cobalt complexes are much more common than rhodium complexes.



BON ALL SE

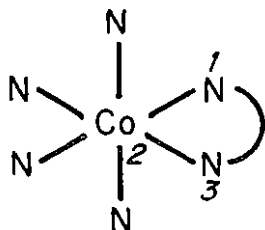
Structure A

If Screen 944, requiring the presence of three or more chain bonds to a central atom, is added to the search for Structure A, the projected iterations drop from 25,000 to 15,000 and the online search completes.

Screens

Screens for Ring or Chain Bonds

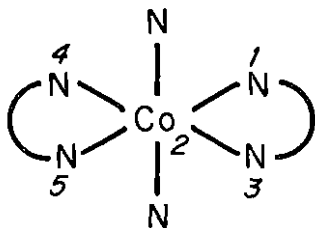
Combining Screen 941, requiring the presence of at least one ring bond and two chain bonds to a central atom, with Structure B reduces the projected 25,000 iterations for a search of Structure B alone to 8,000-10,000. Such improvement allows the online search to run to completion.



BON 1-2 2-3 RSE

Structure B

There were no problems searching Structure C. Screen 945, requiring four ring bonds to a central atom, was automatically generated and used by the system because it occurs with a low frequency.



BON 1-2 2-3 2-4 2-5 RSE

Structure C

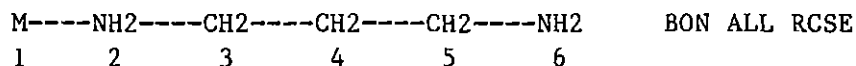
Screens
Screens for Hydrogens on Ligands

Hydrogen Augmented Atom Screens for Ligands

Hydrogen Augmented Atom (HA) screens are special types of augmented atom screens describing a central atom and its attachments, including the exact number of hydrogen atoms. Bond types and exact bond values must be specified. The central atom in an HA screen cannot have any attachments other than those shown in the particular HA definition. Therefore, HA screens can be very useful in describing a well defined ligand backbone; but, they cannot be used to describe a coordinated donor atom.

HA screens are not automatically generated, so adding them manually can help a search run to completion. A selected list of HA screens most useful to ligands is given in Appendix 3.

To illustrate the power of the EC and HA screens, we can follow the development of a search strategy for all transition metal complexes of 1,3-propanediamine, which includes examples of the ligand as a monodentate, bridging, and chelating group. To accomplish this, we create the structure using Single Exact Ring or Chain bonds. We soon discover that this is too general, even if we add screens for the three transition series, i.e., Screens 1964 or 1921 or 1931.



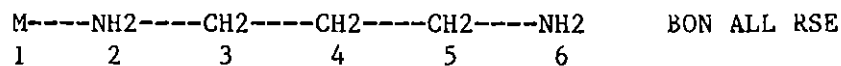
Structure D

<u>Query:</u>	<u>Iterations</u>
STR D	900,000
STR D and SCR (1964 or 1921 or 1931)	40,000

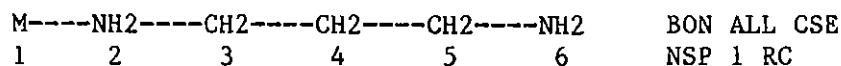
To lower the number of projected iterations we should completely define the bond types. We create two structures, one with Ring Single Exact bonds for the chelates and the other with Chain Single Exact bonds for the monodentates. The latter requires a node specification of Ring or Chain at the metal node. With the same set of transition metal screens, we find an improvement, but not enough to allow the search to complete online.

Screens

Screens for Hydrogens on Ligands



Structure E



Structure F

<u>Query:</u> STR E or STR F and SCR (1964 or 1921 or 1931)	<u>Iterations</u>
	25,000

We will use HA screens to describe the hydrogens attached to carbon atoms in the 1,3-propanediamine. Describing the carbon at node 4 is a pair of screens for a CH₂ group bonded to two carbons, one for the ring bonds in Structure E and the other for the chain bonds in Structure F:

HA 989	CH2*1C*1C
HA 1006	CH2-1C-1C

<u>Query:</u> STR E or STR F and SCR (1964 or 1921 or 1931)	<u>Iterations</u>
and SCR (989 or 1006)	20,000

Screens 989 or 1006 may not be quite specific enough, but there is also a second pair which describe two occurrences of a CH₂ group bonded to a nitrogen and to a carbon (nodes 3 and 5).

HA 1211	(2)	CH2*1C*1N
HA 1236	(2)	CH2-1C-1N

<u>Query:</u> STR E or STR F and SCR (1964 or 1921 or 1931)	<u>Iterations</u>
and SCR (1211 or 1236)	16,000

With the projected iterations well below the system limit, a full file search was run and over 1300 answers retrieved for transition metal complexes of 1,3-propanediamine. If complexes of Zn, Cd, and Hg or other post transition metals are desired, a different set of EC screens could be used.

Screens Screens for Type of Ring

Type of Ring Screens

Type of Ring (TR) screens enable you to specify the size of a ring as well as the points at which it is fused to other rings. TR screens are never automatically generated unless the rings in a substructure have been isolated by a ring specification attribute. Adding TR screens manually is risky, because it can cause lack of recall unless all possible TR screens are used. This danger is especially high for bridged and cage structures. Unfortunately, these are just the sorts of compounds where TR screens have the most potential value in coordination compound searching. If the user can accept retrieval of most of the relevant answers, while possibly missing a few answers, the risk is deemed tolerable. In this booklet we do not add TR screens to a search unless there is really no alternative but to abandon the search question.

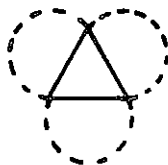
TR screens are very specific and imply that certain positions, called T, are the only places where fusion is required. When using TR Screens to describe a ring system in your substructures, you must include the screens for all of the new fused rings that could be formed upon substitution of your substructure.

A TR screen consists of a series of D's and T's:

- D represents any atom which is part of only one ring, i.e., no fusion occurs at this point
- T represents any atom which is part of two or more rings, i.e., fusion has occurred at this point

Each atom in the ring is represented in sequence by either D or T. TR screens apply to the smallest number of smallest rings.

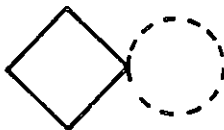
- TTT - Three-membered ring fused at all three atoms



Screens

Screens for Type of Ring

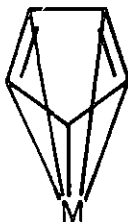
- DDDT - Four-membered ring with one point of fusion



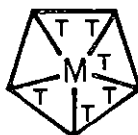
The following examples illustrate how to choose appropriate TR screens.

Pi-Bonding Ligands, Organometallics. Pi-donor ligands are structured in the Registry System by attaching the atoms at each end of the double bond to the metal. Three-membered rings are created as a result of this.

Cyclopentadienyl complexes are a very common example. In the cyclopentadienyl ligand the carbon without the double bond is negatively charged and is also part of the pi donor system. If you don't include substituents on the cp ring in your substructure, you can specify Single Exact and Double Exact bonds and the search will probably run without adding any TR Screens.



If, however, you wish to include a substituent in your substructure, the bonds in the cp ring must be left unspecified. In this case, TR Screens may be one of the only ways to get a complete search. A series of 5 three-membered rings fused at each corner are formed. They would be described by the TTT screen, TR 1851. Note that a five-membered ring is no longer used to describe the coordinated cyclopentadienyl group, because TR screens apply to the smallest number of smallest rings.



Screens
Screens NOT to Use

If you call your Search Assistance Desk, you can get advice about how to proceed if you wish to add TR screens.

Screens Not to Use

AA 1391 or AA 1392 or AA 1558. Do not use for an M atom bonded to a series of ligand atoms. These screens, while still listed in the Screen Dictionary, are not generated automatically and should never be added manually. You may loose answers that have cyclopentadienyl groups in addition to the ligands you have specified in your substructure.

GM (Graph Modifier) 2049. This is the class identifier screen for coordination compounds. This screen may not be present for certain uncharged compounds and is also used for a number of ions like PF₆(1-), which are common counter ions.

NOT Logic. Be aware that adding screens in NOT logic could cause a loss of retrievals by rejecting good answers that have the NOTed feature present elsewhere in the structure. This could happen when the NOTed feature is part of a second component, such as a second complex ion serving as a counter ion, or when the NOTed feature is in another part of the molecule remote from the substructure, such as a second metal center in a binuclear complex. It is better to ADD screens that describe the desired features rather than NOTing out undesired features.

Too Many Screens. Do not add more screens than you need to get the search to complete, because you could miss retrievals with structures you hadn't considered. Add the screens one by one, testing their effect by SAMPLE searches.

Screens

BATCH Searching

BATCH Searching

You should be aware that it is possible to request that structure queries be run for you offline. These searches are run each night and have higher iteration limits and answer set limits than online searches. Online, a search is limited to 20,000 or fewer iterations and 5,000 or fewer answers. The BATCH limits are 50,000 or fewer iterations and 10,000 or fewer answers. Refer to STN Technical Note 86/03 for complete details on this feature.

Dividing a Search that Will Not Complete in One Search

If these strategies are not sufficient, sometimes it becomes necessary to divide a search into several searches. If you need to do this, call the Search Assistance Desk for help in breaking the search into parts and obtaining a credit for the extra searches.

Depending on the nature of the search, the division can be made using screens for one transition row at a time, presence or absence of a fused (chelated) metal atom, ring count, atom count, or Registry Number range. The Search Assistance Desk will be able to devise the correct search strategy.

Summary

This chapter has discussed various screen numbers that can be added to a search profile to make the structure search run more efficiently. Many screen numbers are automatically generated by the system from the structure. However, the screens that describe many features common to coordination compounds occur with high frequency in the database and are not used by the system. They are, in fact, unique when combined with the coordination structure so the search proceeds more efficiently if you manually add them to the profile.

In the next chapter we will summarize some of the general techniques used when searching for coordination compounds.

CHAPTER 7: SUMMARY OF GENERAL SEARCH TECHNIQUES

This chapter summarizes the general guidelines to keep in mind when searching for coordination compounds in the Registry File. It is a synopsis of points that were discussed in more detail in Chapters 3 through 6 and that will be illustrated further in the case studies that make up the rest of this booklet.

Using Ring Or Chain Bond Types

Allow for metal-ligand bonds to be either Ring or Chain unless you know that you want to retrieve only complexes with monodentate ligands or only complexes with chelate rings. If the full file search projection is incomplete when you do this, build two structures, one with ring bonds and the other with chain bonds. Then search Structure 1 OR Structure 2.

Using Single Exact Bond Values

Specify the bonds between the metal and specific ligands as Single Exact (except for monodentate oxo, sulfido, nitrido, or phosphido). If you are searching for generic donor atoms, leave the metal-ligand bond values as Unspecified.

Using Ring or Chain Node Specification

Use node specifications of Ring or Chain on both the metal and donor atom nodes when you use a chain bond between the metal and an unsubstituted donor atom in a substructure describing only a portion of the coordination sphere. These node specifications will allow retrieval of cyclic ligands and of metals with chelate rings elsewhere in the structure.

Summary of General Search Techniques

Ring Isolation

Do not isolate the rings in a substructure unless you are sure you want to prohibit any further chelation on the metal. A metal with two or more chelate rings is an example of spiro ring fusion, not isolated rings.

Using Gk Nodes

Allow chelate ring backbones to vary both in size and in type of heteroatom unless you are searching for a specific type of chelate ring. Use repeating and variable Gk groups to do this.

Tautomers

When the ligand is a tautomer, any atoms that become coordinated no longer have Normalized bonds. The bonds between remaining atoms must be reexamined to see whether they remain Normalized or become Exact.

Using A (Any Atom) Nodes

When you have built a small substructure that only defines one or two ligands in the coordination sphere, add A groups to the metal to increase its coordination number to the minimum you will accept in answers.

Summary of General Search Techniques

Using the CONNECT Attribute

Use a CONNECT Attribute on the metal atom whenever you can specify the exact, minimum, or maximum coordination number you desire in your answers. Do this in addition to adding A groups to the metal. Do not use CONNECT if you want your answers to include pi bonded ligands.

Determining the Coordination Number

In determining the desired coordination number, remember that each atom in a pi bonding ligand that is attached to the metal (mostly organometallics) increases the coordination number by one, e.g., a cyclopentadienyl group requires five attachments to the metal. The apparent coordination numbers may well be higher than the coordination numbers based on the number of electron pairs donated.

Using Hydrogen Counts for Hydrides

Use hydrogen counts to search for metal complexes with hydride ligands.

Attributes and the Gk Node

If you plan to use attributes like connect and hydrogen count on a metal center, you cannot use a Gk group to specify more than one metal. Build separate structures instead.

Using Charges

Do not specify charges on any of the nodes in a coordination compound substructure.

Summary of General Search Techniques

Using Screens

Add screens for the appropriate periodic rows and/or groups when you have used the generic M group for the metal node.

Interpreting SAMPLE Searches

Sample searches for coordination compounds are often less reliable than for organic compounds. When a full file search projection is uncertain or incomplete (but close to the limits), it is often worthwhile to do several more sample searches to attempt to determine a more accurate "averaged" projection before revising your strategy.

Summary

Searching for coordination compounds requires special search techniques. This chapter has summarized some of the basics you should use when searching for this class of compounds. The remainder of this booklet will be Case Studies to supply more explicit details on particular aspects of searching in this field.

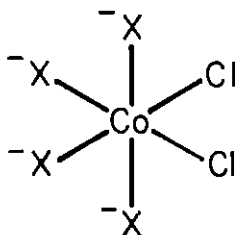
Tailoring the Coordination Sphere Octahedral Complexes

CASE STUDY 1: TAILORING THE COORDINATION SPHERE

This case study uses examples that illustrate some ways of using structures and screens to retrieve the kinds of coordination spheres you seek. Strategies for excluding unwanted features from the coordination sphere are equally as important to a successful search as specifying the features you require. Examples using halide ligands have been selected to illustrate these points, but the general principles can be applied to other ligands as well.

All Octahedral Dichloro Complexes of Cobalt

The essential feature desired in this search is the presence of exactly two chloride ligands in a six coordinate cobalt complex. More than two chlorides are not desired, nor are any other halide ligands. In the other four positions, any ligand other than halide may be present. To exclude a particular element from a node, assign a minus (-) element symbol to the node where the element may not be present. The generic halide symbol X is used in this example to exclude all halides from the other four coordination positions on the cobalt.



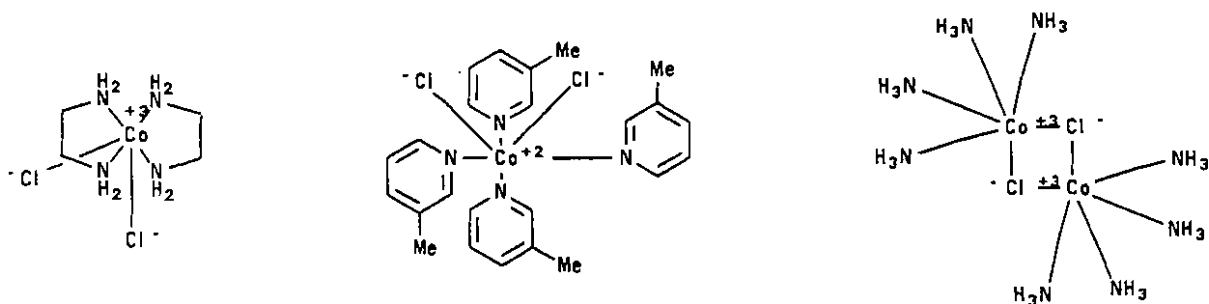
BON ALL RCSE

-X = any element except
halogen or hydrogen

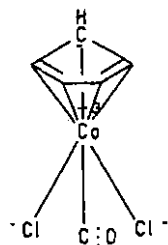
It was decided not to include a CONNECT attribute of Exactly 6 on the cobalt since all six ligands required for an octahedral complex are part of the structure and cobalt rarely shows coordination numbers greater than six. No screens were necessary to allow the search to run.

Tailoring the Coordination Sphere Octahedral Complexes

Examples of the types of answers retrieved by a search on this substructure are shown below.



Had a CONNECT of Exactly 6 been used for the cobalt, answers like the one below would have been lost. Although these organometallic derivatives are considered six coordinate from a chemical point of view, the CAS Registry System sees seven or more connections at the metal (each cp ligand contributes five connections).

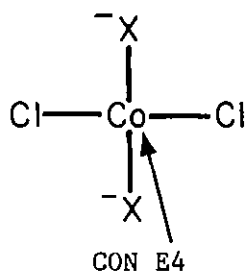


A strategy similar to this example works for most metals and can even be generalized to include all dihalo complexes by using the generic X group in place of the Cl's. A CONNECT attribute of Exactly 6 on the metal node might be appropriate if the metal in question is known to have many examples of coordination numbers greater than six.

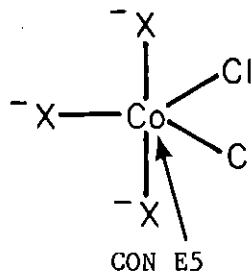
Tailoring the Coordination Sphere Complexes with Low Coordination Numbers

Dichloro Cobalt Complexes of Lower Coordination Number

Cobalt also exhibits both coordination number four and coordination number five. Either of these can be searched for individually. After creating the required structure, add the CONNECT attribute of Exactly 4 or Exactly 5 to the cobalt node. This will exclude from the answer set any complexes with coordination numbers greater than the assigned CONNECT value. It is important to add the CONNECT attribute in this case, because we know from the example above that there are many examples of cobalt with a coordination number of six or more. This strategy would not retrieve cyclopentadienyl complexes, but without the Exact CONNECT attribute there is no way to limit answers to the lower coordination numbers of cobalt.



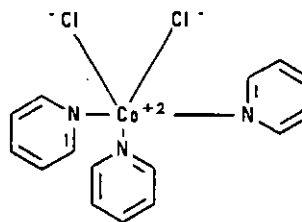
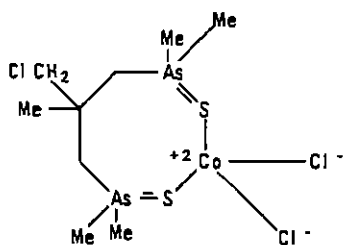
STRUCTURE A



STRUCTURE B

If you want answers with both coordination numbers four and five in the same answer set, you can search for them together with the query: STR A or STR B.

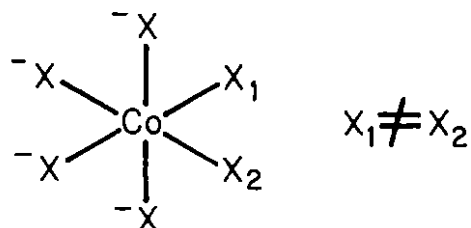
Examples of answers are shown below.



Tailoring the Coordination Sphere
Requiring Two Different Halide Ligands

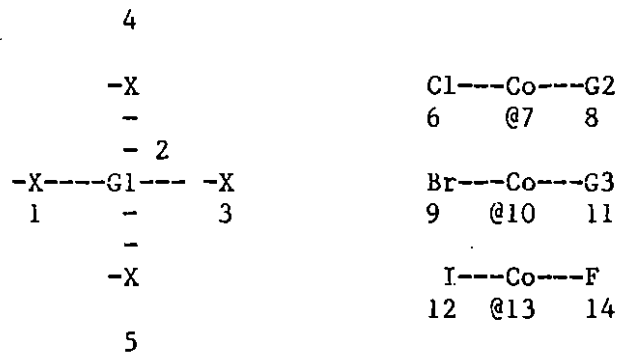
Two Different Halide Ligands

If we wish to require two different halide ions in a structure,



we can do this by using Gk groups.

There is a variety of ways to build the structure query. The most efficient search results when we use just one structure with three VARIABLE Gk groups. G1 is defined as any of three multiatom fragments containing a cobalt atom with two different halides. G2 allows for F, Br, or I in the presence of Cl. G3 allows for either F or I in the presence of Br, and the third fragment covers the final possibility of I and F together. We assigned a CONNECT of Exactly 6 to each of the cobalt atoms used in the G1 definitions. The query structure is shown below.



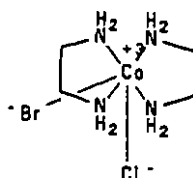
```

VAR G1=7/10/13
VAR G2=F/Br/I
VAR G3=F/I

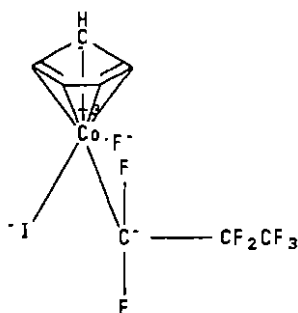
```

Tailoring the Coordination Sphere Requiring Two Different Halide Ligands

A search of this structure produces answers such as



Because we assigned a CONNECT value of E6 to the cobalt atoms, this strategy fails to retrieve such answers as



because the connectivity is greater than the coordination number.

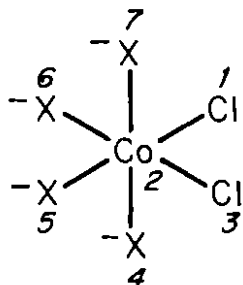
Running the same search without a CONNECT attribute on the cobalt retrieves "answers" that did not match the query because the iteration process for them was incomplete. The answer set is more complete at the expense of being considerably less precise.

Thus, it can be seen that one of the benefits of using a CONNECT attribute can be fewer "answers" that have not completed the iteration step. But, when you use the CONNECT attribute for coordination compounds, you trade off a possible loss of a few organometallic answers with connectivities higher than the formal coordination number for an answer set with a higher degree of precision. Sometimes you will willingly make this tradeoff in order to get a more manageable answer set. For example, the CONNECT attributes made it possible to get highly precise answer sets for four- and five-coordinate complexes without retrieving a very large number of answers with higher coordination number.

Tailoring the Coordination Sphere Bridging Halides

Bridging Halides

If the cobalt-chloride bonds in the structure are specified as Ring

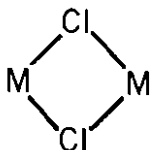


BOND 1-2 2-3 R,
2-4 2-5 2-6 2-7 RC

the answer set will have only those substances with a minimum of two bridging chlorides. Alternatively, you could exclude the chloro bridged species from the answer set by taking the Chain default for the bonds 1-2 and 2-3.

In the next example we will broaden the structure part of the query to encompass complexes where the chlorides bridge two metals. With this more general approach, we can explore the effect of both structures and screens on the number of iterations projected for the Full file search.

The essential part of the query structure is a four-membered ring containing the metals (M) and the bridging chlorides. Structure C, shown below, meets this requirement.



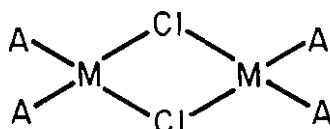
Structure C

Tailoring the Coordination Sphere Bridging Halides

Since the structure contains the generic M nodes, we will add a screen for the metals that are desired, e.g., Screens 1921 or 1931 or 1964 define the three transition series. If all metals are desired, Screen 1918 should be used. We choose to add the screens for the three transition series.

<u>Query</u>	<u>Iterations Projected</u>
STK C	over 300,000
STR C and SCR (1921 or 1931 or 1964)	over 60,000

From the high number of projected iterations, it is obvious that the query needs refinement. As a general rule, use specific structures in which the coordination sphere is as completely defined as possible within the scope of the query. Structure D assumes that both metals will be at least four coordinate



BON A-M RCSE

Structure D

A search of Structure D, in which coordination number four or greater is specified, illustrates the improvement possible when you add two generic A groups to each metal to increase the coordination number.

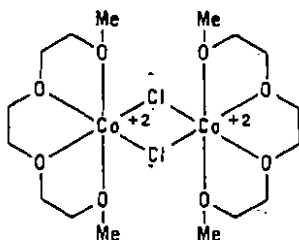
<u>Query</u>	<u>Iterations Projected</u>
STR D	over 80,000
STR D and SCR (1921 or 1931 or 1964)	over 32,000

Tailoring the Coordination Sphere Bridging Halides

Even though the search statistics are improved by the more specific Structure B, there are still too many iterations for a single online search. We also learn that the projected number of answers for this search is greater than the number of answers allowed for a single online search. The search could be run offline with BATCH or it could be run online by splitting the search into multiple searches. It is convenient to do this on the basis individual transition rows, so each of the transition metal screens is used separately.

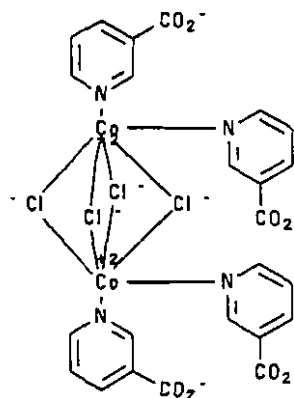
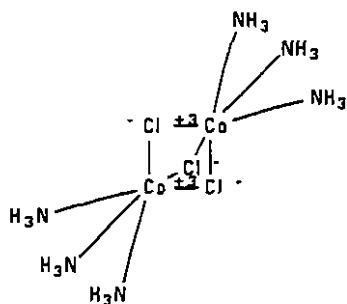
	<u>Iterations</u>	<u>Answers</u>
<u>Query:</u> STR D and SCR 1964	12887	1524
<u>Query:</u> STR D and SCR 1921	11554	3528
<u>Query:</u> STR D and SCR 1931	9348	1173

After combining the answer sets, the results of the search for the dihalo bridged complexes provide answers such as

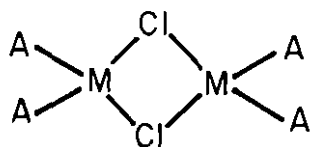


Tailoring the Coordination Sphere Bridging Halides

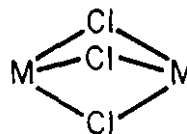
The search statement requires only the presence of at least two chloro bridges. We note that the answers include some examples of both tri- and tetra- bridged substances.



Although there are relatively few of these, you may wish to exclude them. You can limit the search to substances with only two halo bridges by building a second substructure with three bridges and adding it to the search query with NOT logic, as shown in the following query:



NOT



Tailoring the Coordination Sphere Halide and Halo Complexes of Rhenium

Simple Halides and Halo Complexes of Rhenium

All rhenium halides and halo complexes are sought. The rhenium should not have any other nonmetal atoms bonded to it.

Rhenium has a very rich halide chemistry. It exhibits a variety of oxidation states and undergoes dimer and trimer formation both with and without halide bridges. Thus, it is necessary to construct a very general structure to require the presence of at least one halide on the Re.

```
Re????X      BON 1-2 RC
 1      2
```

Structure E

To meet the objectives of this search, it is necessary to exclude elements other than halogen from the coordination sphere of the rhenium. Since there is no generic symbol to accomplish this in the manner that -X excludes all halogens, a more complicated strategy is necessary.

To eliminate any other elements bonded to the rhenium, a second structure is created bonding rhenium to other common ligands.

```
Re????G1      BON 1-2 RC
 1      2
```

```
VAR G1=N/O/S/P/As/C/H
```

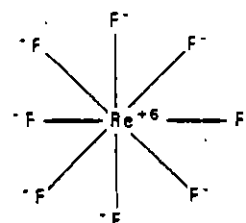
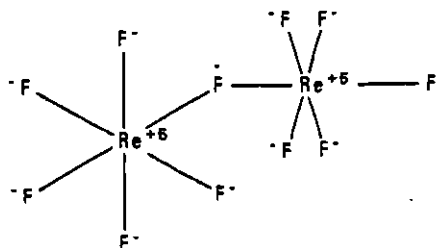
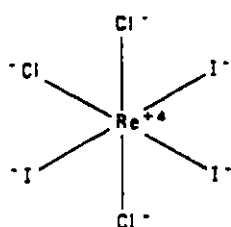
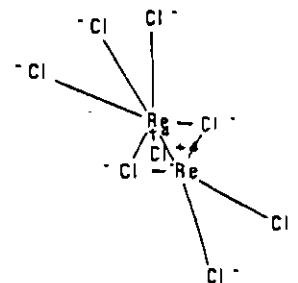
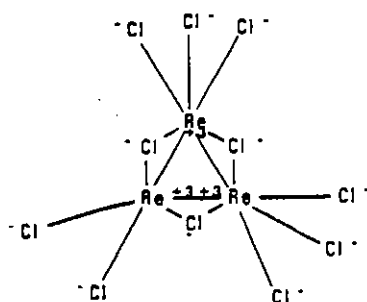
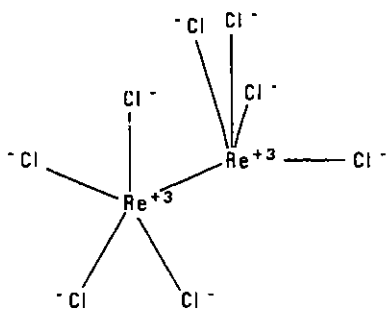
Structure F

Structure F is then used with NOT logic.

Query: STR E not STR F

Tailoring the Coordination Sphere Halide and Halo Complexes of Rhenium

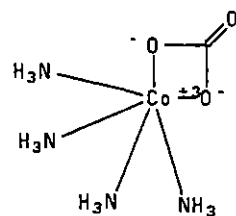
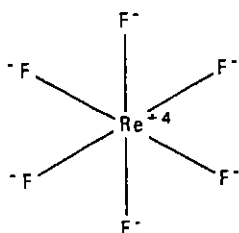
Answers include: ReI , ReCl_3 , and



Of the answers retrieved, a few are extraneous substances in which rhenium is bonded to elements other than halogen, such as selenium or other metals. These answers are examples of donor atoms we didn't think to exclude when we created the Variable G1 group in Structure F. Remember you can specify up to 20 different elements in a Gk group.

Tailoring the Coordination Sphere Halide and Halo Complexes of Rhenium

This strategy also allows us to retrieve two component answers such as

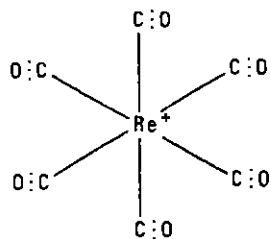
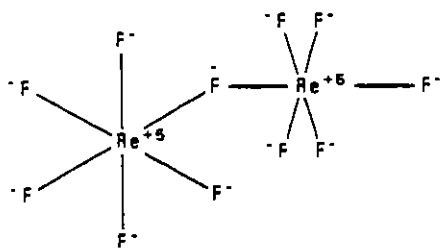


COMPONENT 1

COMPONENT 2

in which the counter ion has a different metal bonded to both nitrogen and oxygen ligands.

However, to get as precise an answer set as we did, we sacrificed answers in which a second rhenium complex is the counter ion. The second rhenium is indeed bonded to carbon, one of the donor atoms we excluded by NOT logic.



COMPONENT 1

COMPONENT 2

Tailoring the Coordination Sphere
Halide and Halo Complexes of Rhenium

Summary

Using halides as examples, Case Study 1 has illustrated some techniques used to specify the nature of the coordination sphere.

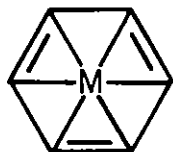
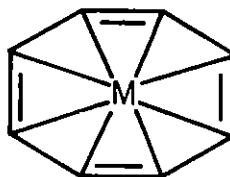
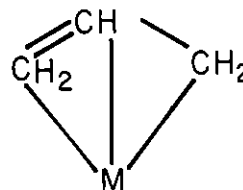
In Case Study 2, organometallics will be discussed.

CASE STUDY 2: ORGANOMETALLICS

Organometallic compounds have carbon atoms covalently bonded to a metal atom by either sigma or pi bonds. However, compounds with carbon monoxide, cyanate ion, or carbide ion as the sole carbon donor ligands are not considered to be organometallics.

The structuring conventions used by the Registry System for pi bonded systems differ considerably from those used by chemists for these compounds. In the Registry System, each carbon atom in the pi system is connected to the metal atom. This results in structures based on the connectivity of the metal atom rather than on typical coordination numbers derived from the electronic structure of the metal.

The bonds in cyclic pi donors containing an even number of carbon atoms joined by alternating double and single bonds are Normalized bonds. Cyclic pi donors with odd numbers of carbon atoms have Double Exact and Single Exact bonds. Linear pi bonded systems have alternating Double Exact and Single Exact bonds. In all cases, the bonds from the carbons to the metal are Single Exact bonds.

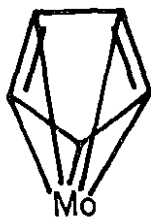
Exact
BondsNormalized
BondsNormalized
BondsExact
Bonds

All of these structures contain a series of three-membered fused rings. You may wish to combine TR screen 1851 for DTT or TTT rings with the structure queries for organometallics because it will markedly reduce the number of iterations required. Remember, however, that when you use TR screens, you risk losing some relevant answers.

Metallocenes (Cyclopentadienyl Complexes)

Metallocenes, i.e., cyclopentadienyl complexes, are exceedingly common, especially complexes with iron. It is not possible to search for cyclopentadienyl derivatives of iron without dividing the search into a number of smaller searches because of the large number of compounds reported. It is certainly not feasible to try to search for cyclopentadienyl derivatives of all metals using a generic M node. A substructure has to be explicit. However, if you specify Single Exact and Double Exact bonds and specific metals, other than iron, the search will complete.

A sample search for derivatives of cyclopentadienylmolybdenum



Structure A

projects more than 5,000 answers. This search would have to be run as a BATCH search or split into multiple online searches. Since there are five rings present with every cyclopentadienyl group, a Ring Count screen for ten or more rings, RC 1846, was chosen to divide the search into two groups.

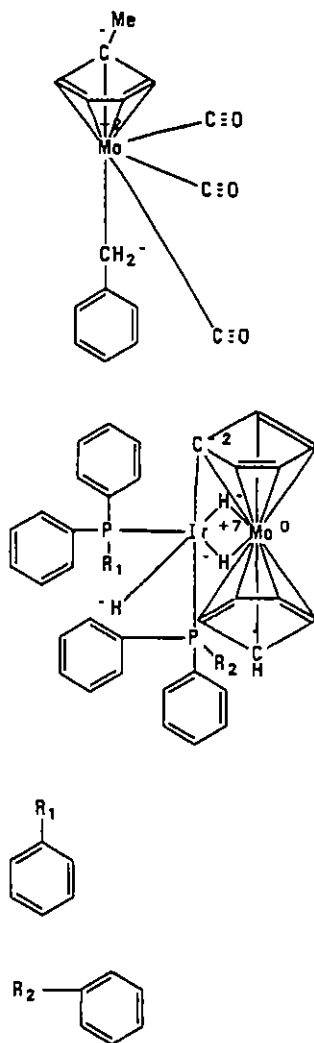
One query will retrieve substances with two or more cp groups.

Query: STR A and SCR 1846

The second query will retrieve substances with only one.

Query: STR A not SCR 1846

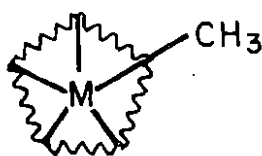
Examples of substances retrieved based on these queries are shown below.



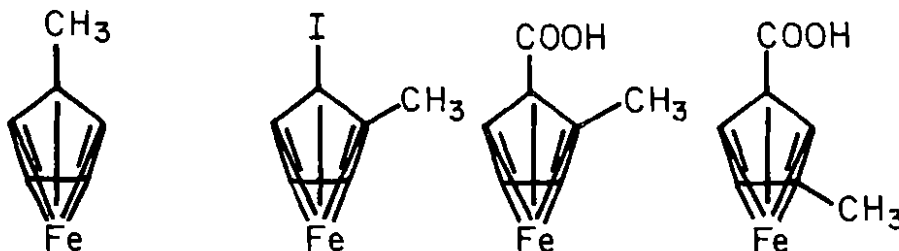
Organometallics
Substituted Cyclopentadienyl Ligands

Substituted Cyclopentadienyl Ligands

If you are allowing for substitution on the cyclopentadienyl ligand, you should leave the bonds Unspecified in the pi system of the query. The location of the single and double ring bonds may vary if more than one substituent is present. A search for methylcyclopentadienyls



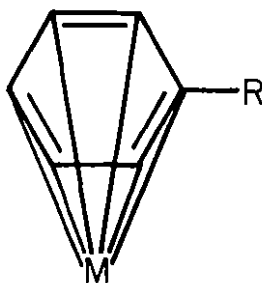
revealed three different types of structures. Examples of each are illustrated below.



However, a sample search of the query with Unspecified bonds shows that the full file search will not complete, even with the addition of EC screens 1921, 1931, and 1964 for the transition metals and HA screen 963 for a methyl group bonded to a carbon. Another approach is to use Exact bonds in multiple structures placing the substituent into each of the possible locations with respect to the exact bonds and then OR the structures together. With either strategy, the search will have to be divided into two portions because iteration limits are exceeded even with the TR screen 1851 included.

Organometallics
Noncyclic Polyenes

In cases where the pi donor ligand is uncharged and has normalized bonds, as in the example below, all positions of substitution are equivalent.



Therefore, there is no need to use Unspecified ring bonds.

Noncyclic Polyenes

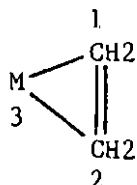
When searching for open chain polyene pi donors, a strategy is needed that will eliminate answers in which additional bond formation generates the closed cyclic ligand. There are several ways to prevent these answers, depending on the search requirements. A CONNECT attribute of Exactly 2 Ring bonds at the two terminal carbons is the most general way to insure the presence of the open diene. This allows for both an exact number of carbons in the pi system and the possibility of uncoordinated substituents anywhere in the ligand. If you wish to prevent substitution of any type on the terminal carbons, use a CONNECT attribute of Exactly 2 Ring or Chain at those nodes. A Hydrogen Count of Exactly 2 would also block any substitution on the end carbons.

Searching for this type of substance will often run into iteration limits since these attributes do not bring down the number of compounds to be iterated. They only reduce the number of answers. This problem of too many iterations will be especially true when the ligand is one of the smaller pi systems, such as allene and butadiene. In order to get searches for these substances to run, you will often have to divided the search into two searches based on ring counts.

Organometallics
Ethylene Ligand

Ethylene

Upon coordination to a metal, ethylene forms a three-membered ring which is not part of any larger fused ring system. It is necessary to prevent any further substitution on the carbon atoms. This is done by adding a CONNECT Attribute of Exactly 2 Ring bonds.



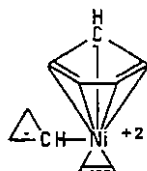
CON 1 2 E2 R

Structure B

Because this is such a small fragment, over 400,000 iterations are projected. This means this search is impractical without adding screens. You can effectively screen out larger pi donors by using the TR screen 1848, for a DDD ring system, together with TR screen 1850, for a DDT ring system, which allows for chelate formation on the metal. This lowers the iterations to 10,000, and, in spite of the potential loss of a few answers, allows us to perform the search.

Query: STR B and SCR (1848 or 1850)

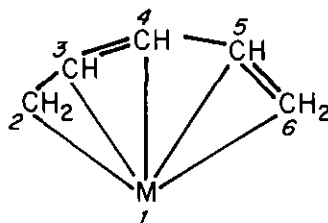
An examination of some of the answers indicates that larger pi donor ligands are often present together with the ethylene. An example is shown below.



Pentadiene

Any open chain pi donor larger than ethylene will probably present difficulties with respect to the projected number of iterations, since there is no way to exclude the cyclized derivatives of the ligand from the initial screen step. Even reasonably specific substructures have incomplete full file projections. The only way around this problem is to divide the search, basing the split on ring count.

We have chosen pentadienes to illustrate how the search can be divided. The structure for pentadiene is shown in Structure C.



CON 2 6 E2 R

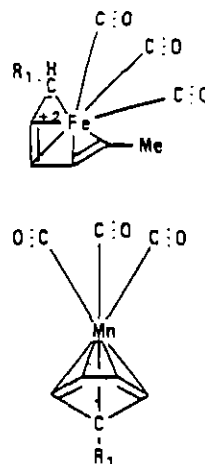
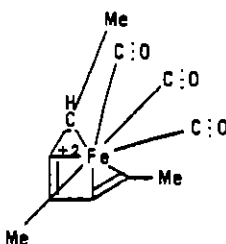
Structure C

The RC screen 1846, for a ring count of ten or more, was combined with this structure to define the two searches needed to allow the full file search to run to completion.

Query 1: STR C and SCR 1846

Query 2: STR C not SCR 1846

Examples of answers to these searches are shown below.

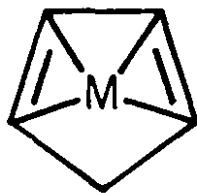


Organometallics
Partially Coordinated Pi System

Partial Coordination of the Pi System

There are a number of organometallic compounds where only some of the carbon atoms in the ring are bonded to the metal. This presents the familiar problem of excluding the fully coordinated ligand structure from flooding either the iteration step or the answer set. We will use the

-cyclopentadienyl ligand to illustrate strategies to get to the desired answer set. This ligand requires that one carbon in the five-membered ring remain unattached to the metal atom. Assigning a CONNECT Attribute of Exactly 2 Ring to that carbon will insure that the answers will not have a bond between carbon and metal at that point. Structure D represents the structure query.



CON E2 R

Structure D

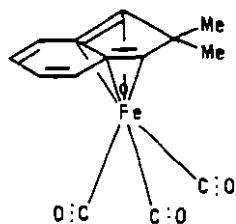
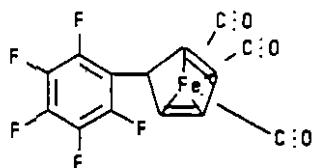
A sample search projects that the iterations are far above system limits. Without TR screens, this search is not practical. A unique pair of TR screens for both a four-membered ring and three-membered rings describes the the rings found in Structure D. There no longer is a set of 5 three-membered rings, but rather 3 three-membered TTT rings and 1 four-membered DTTT ring. The presence of both types of rings in the same substructure is a rare occurrence. The strategy is to combine these two screen with the structure.

Query: STR D and SCR (1851 and 1853)

This query will run well within system limits.

Organometallics
Partially Coordinated Pi System

Remember that there may be some loss of recall due to the addition of the TR screens. Examples of answers retrieved are shown below.



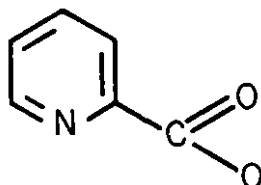
Summary

This case study has illustrated various strategies used when searching for organometallics. Case Study 3 will discuss complexes containing carboxylic acids in the ligands.

CASE STUDY 3: CARBOXYLIC ACIDS

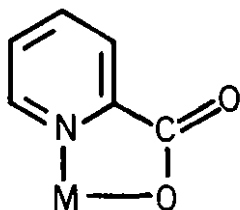
When searching for complexes of ligands that have several ways of coordinating to a metal, it is necessary to decide how general you want the search to be. Do you only want examples of one type of bonding or do you want all substances that contain that ligand and any metal? If it is necessary to retrieve all substances where the ligand and a metal occur, you need to build a very general structure and you may get a great many answers.

In this case study, the emphasis is on the completely defined ligand, picolinate ion, or 2-pyridinecarboxylate, its current Index Name,



and its interactions with metal ions. The general principles can, however, be applied to other carboxylates.

When a metal coordinates to picolinate, the most typical complex has a five-membered chelate ring.



Other possibilities could include monodentate coordination either through the pyridine nitrogen or through the carboxylate oxygen. You might also need to consider whether other functional groups substituted on the aromatic ring influence the mode of coordination. Finally, compounds that have the anionic form of the picolinate with the metal as the counter ion require yet another approach.

Comprehensive Searching

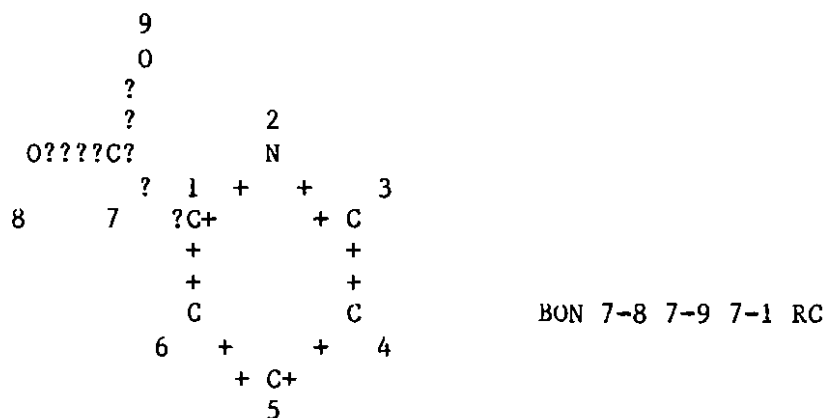
When you want complete retrieval of all substances with the carboxylate and the metal, you are searching with a very broad strategy and may retrieve a great number of answers. The general procedure for this most complete recall is given in the following steps.

- Build one structure containing only the metal node. Place a Node Specification of Ring or Chain on the metal.
- Build a separate structure for the carboxylate. Use Unspecified Ring or Chain bonds for the bonds in the ligand which might become part of a chelate ring.
- Add screens to define the metal.
- Search a query consisting the two structures and the screen set combined with AND logic.

The strategy of using two separate structures gives the most general search that can be run. It can retrieve all substances which have picolinate associated in any way with a metal. It will retrieve all of the answers with the ligand bonded to the metal in any way, as well as those where the metal is not bonded to the picolinate, such as the substances that are structured with the metal as a salt.

Carboxylic Acids
Comprehensive Search

Begin by building the structure of the picolinate, Structure A. The bonds in the carboxylate group must be Ring or Chain Unspecified, including the bond between the carboxylate carbon and the ring.



Structure A

A second structure for the metal, Structure B, is built. A Node Specification of Ring or Chain must be used on the metal.



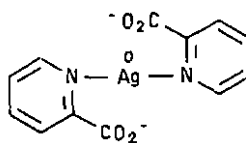
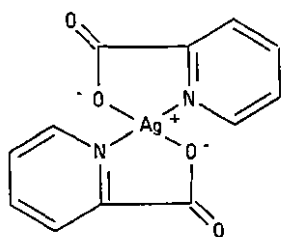
Structure B

Carboxylic Acids
Comprehensive Search

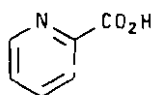
To complete the search strategy, appropriate screens for the metal should be included. Those for the transition metal groups were added.

Query: STR A and STR B and SCR (1921 or 1931 or 1964)

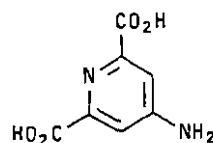
This is an extremely general strategy and will retrieve a great many answers, including all two component salts. Examples are shown below.



• 2H⁺



• 1/2 Mn(II)

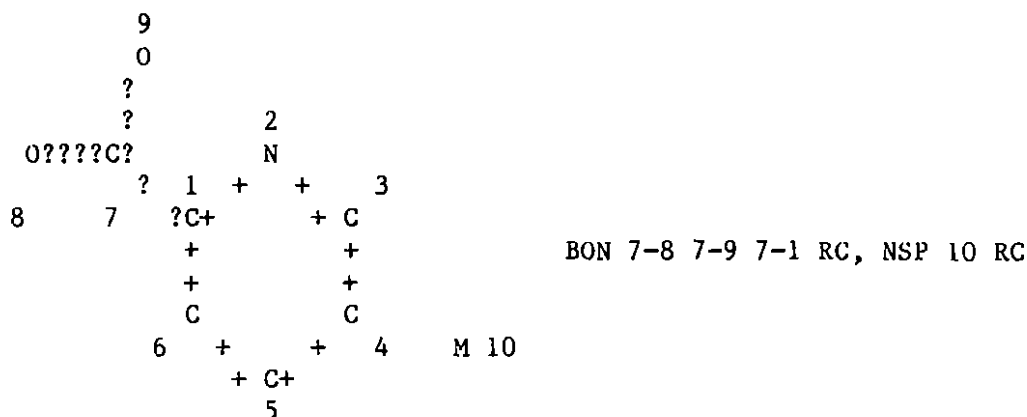


• 1/2 Fe(II)

Carboxylic Acids
Ligand - Metal Complexes

All Substances with the Ligand Bonded to the Metal

When you want to retrieve a comprehensive survey of ligand interactions with the answers containing all substances where a picolinate or substituted picolinate is bonded in any way to a metal, combine a separate metal node with the picolinate substructure in one query structure. To achieve complete retrieval, it is necessary to allow for all possible modes of coordination. To this end, the bonds in the carboxylate group must be Ring or Chain Unspecified, including the bond between the carboxylate carbon and the ring. In addition, the separate metal node must receive a Node Specification of Ring or Chain. Structure C illustrates this type of query structure.



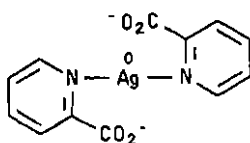
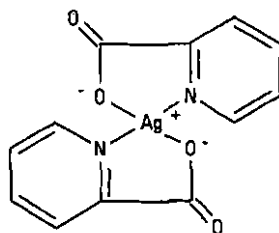
Structure C

Add whatever screens are needed to define the metal. We will use those for the transition metals, EC 1964, 1921, and 1931.

Query: STR C and SCR (1921 or 1931 or 1964)

Examples of the wide variety of chelated and monodentate structures are shown on the next page. There will be no answers with the metal as a separate component.

Carboxylic Acids
Ligand - Metal Complexes



• 2H⁺

The general procedure for retrieving structures that have a ligand bonded in any way to a transition metal is given in the following steps.

- Make separate structural fragments for the ligand and for the metal in one structure query.
- Put a Node Specification of Ring or Chain on the metal.
- Assign Ring or Chain bond types to any chain bonds in the ligand which could potentially become part of a chelate ring.
- Add screens to define the metal.
- Search a query consisting the structure and the screen set combined with AND logic.

Carboxylate anions can be structured as counter ions in metal salts, especially if the metal-ligand interaction is not clearly defined in the original document. Such salts could warrant consideration as metal complexes, even though this type of structure conveys no structural information about the bonding to the metal. These salts are two component systems with the carboxylate as the free acid in one component and the metal as a single atom fragment in the second component.

- Build one structure containing only the metal node.
- Build a separate structure for the carboxylate. Use Normalized bonds in the carboxylic acid portion of the structure.
- Add screens to define the metal.
- Add a screen to require the presence of two or more components.
- Search a query consisting the two structures and the screen sets combined with AND logic.



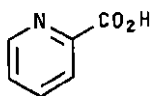
Carboxylic Acids
Metal Salts

To search for salts as multicomponent systems, the metal node and the carboxylate structure must be in separate substructures. Build two structures, one for the metal, Structure D and a second one for the picolinate, Structure E. Then, search on a query containing the two structures combined with AND logic.

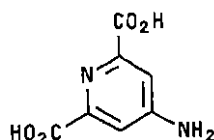
Appropriate screens for the metal should be included. Those for the transition metal groups were added. Also a screen for two or more components should be used to eliminate those structures where the metal is bonded to the ligand.

Query: STR D and STR E and SCR (1921 or 1931 or 1964) and SCR 2127

This strategy will retrieve answers such as those shown below.



• 1/2 Mn(II)



• 1/2 Fe(II)

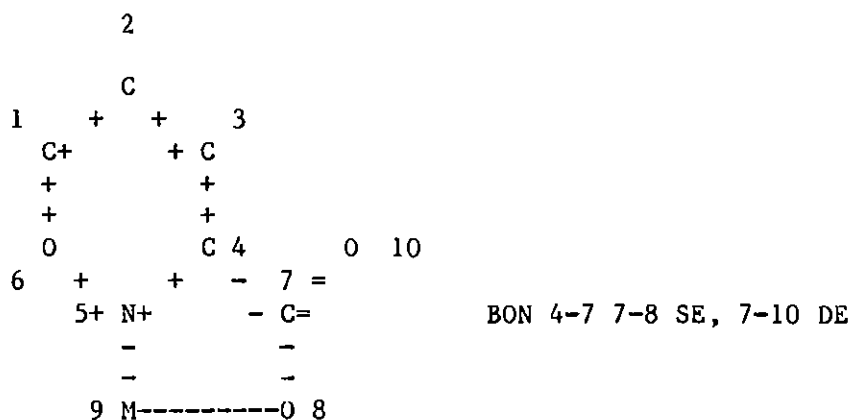
Carboxylic Acids
Nitrogen, Oxygen Chelates

Specific Structures - N,O Chelates

You can create structures with specific modes of coordination. Even if you want to retrieve complex picolines with a specific structure, it is important not to make the substructure too restrictive.

Picolinate ion most commonly forms a five membered chelate ring, bonded through the nitrogen of the pyridine ring and the single bonded oxygen of the carboxylate.

It is important to note that, when a carboxylate group is attached to a metal, the bonds become Exact. These bonds are Normalized bonds in the uncoordinated ligand.

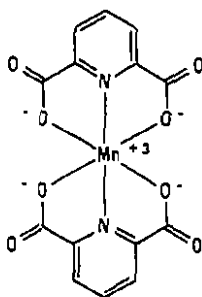
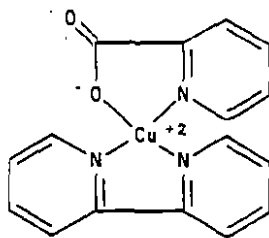


Structure F

Query: STR F and SCR (1921 or 1931 or 1964)

Carboxylic Acids
Nitrogen, Oxygen Chelates

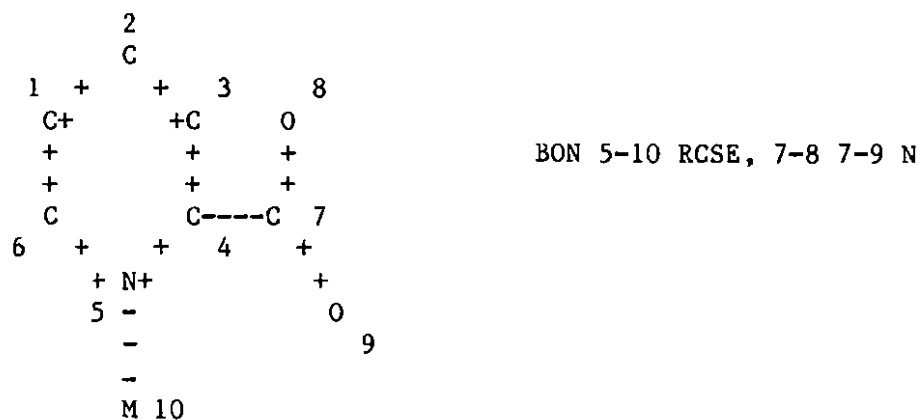
The search of this query give answers limited to transition metal complexes, because of the screen set chosen to describe the metal. Substances with this chelated structure comprise about three-quarters of all picolinate complexes discussed in this case study. Examples are shown below.



Carboxylic Acids
Uncoordinated Carboxylates

Structures with an Uncoordinated Carboxylate Group

The picolinate also could be coordinated through just one of the donor atoms, either as a monodentate ligand or as a chelate through another unspecified functional group. We can retrieve these structures by blocking coordination at either the carboxylate group or at the nitrogen in the pyridine ring. For example, Normalized bonds in the carboxylate group block any coordination to metals through the oxygen.

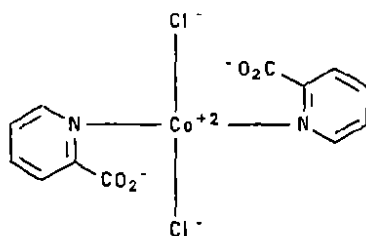


Structure G

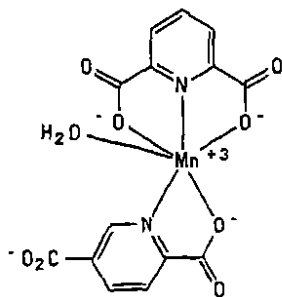
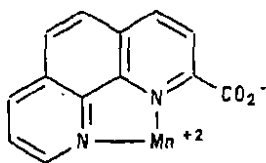
Query: STR G and SCR (1921 or 1931 or 1964)

Carboxylic Acids
Uncoordinated Carboxylates

Answers to this query all have a free carboxyl groups. Some have a monodentate pyridine ligand. In a number of others, the pyridine nitrogen forms a chelate ring with other functional groups which are part of a substituted ligand. Several of these chelated structures were also retrieved in the previous search, since some substances have both a coordinated and a free carboxylate group on the pyridine ring.



• 2H⁺



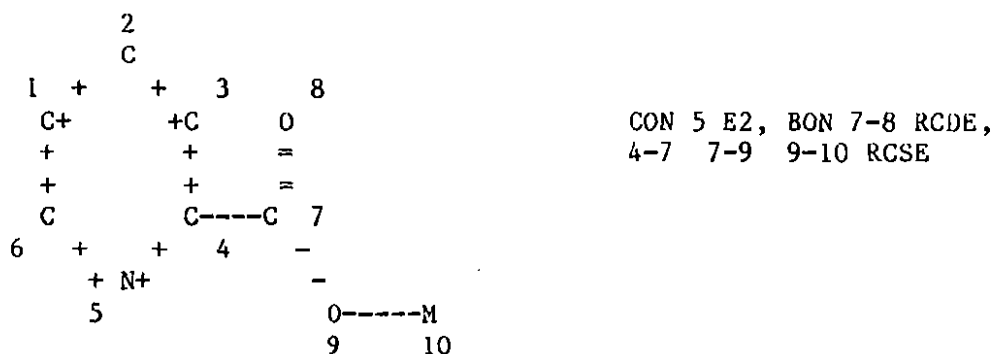
• H⁺

Carboxylic Acids
Uncoordinated Pyridine Group

Structures with an Uncoordinated Pyridine Group

To block metal coordination at the pyridine nitrogen, use a CONNECT attribute of Exactly 2 at the nitrogen node. This will block not only coordination to a metal, but also any other substitution, such as amine oxides.

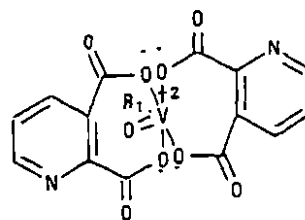
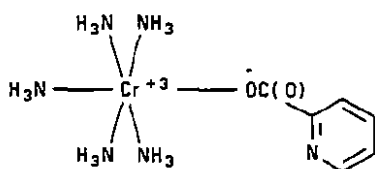
When a carboxylate coordinates to a metal, it may do so as a monodentate ligand through the single bonded oxygen or it may form a chelate ring either with the other oxygen or with some other substituent on the ligand. For this reason, the bond between the oxygen and the metal and those between the carbon and the oxygens in the carboxylate group should be specified as Ring or Chain. The bond between the pyridine ring and the carboxylate group will become part of a ring if another functional group participates in chelation. Therefore, that bond also should be specified as Ring or Chain.



Structure H

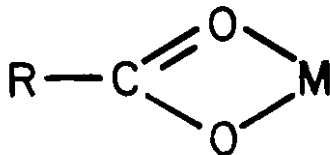
Query: STR H and SCR (1921 or 1931 or 1964)

This answer set contains both monodentate and chelating oxygens.



Carboxylic Acids
Uncoordinated Pyridine Group

This query would also retrieve answers where the carboxylate chelates through both oxygens.



To date there are no examples in the Registry File of picolates which exhibit this type of chelation. For picolates, five-membered chelate rings containing the pyridine nitrogen and one carboxylate oxygen predominate.

Summary

When searching for substances containing carboxylates, many strategies can be used. This study has shown various queries used to retrieve the type of answers your search requires. In the next case study, we will discuss some special strategies needed when unspecified derivatives are involved. The picolinate example will be used.

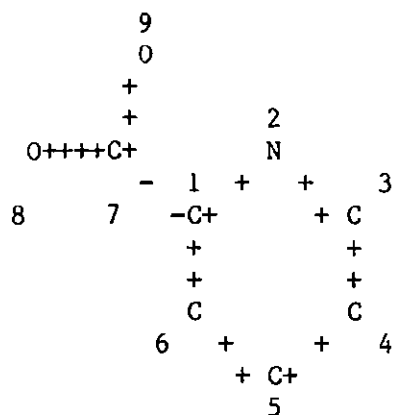
Unspecified Derivatives
Derivatives without Registry Numbers

CASE STUDY 4: UNSPECIFIED DERIVATIVES

In Case Study 3, we described a very general strategy to retrieve all structures that contained both a picolinate (2-pyridinecarboxylate) and a transition metal. There are yet other strategies that must be used in conjunction with those described in Case Study 3 in order to retrieve all substances and references which pertain to complex formation by picolines.

Substances without Registry Numbers

When a document mentions a metal complex whose exact nature is not specified, it is not possible to assign a molecular formula to the complex. This may occur in the cases of solution studies, an incompletely characterized reaction product of a picolinate and another complex, or part of a generalized series of related picoline complexes. For such studies, the complex reported will not be assigned Registry Numbers. Instead, the substance is indexed as an unspecified derivative of the ligand and of the metal or starting complex. You can retrieve such unspecified derivatives by first doing a substructure search on just the ligand structure, in this case picolinic acid, Structure A. Then, cross these answers over into the CA File.



Structure A

Unspecified Derivatives
Derivatives without Registry Numbers

A substructure search on Structure A places the answers in answer set L3.
This is the answer set used as the search term in the CA File.

=> FIL CA

.
.
.

=> S L3/D (L) COMPLEX?

.
.
.

Examples of answers are shown below. We have underlined the term causing the record to be retrieved. Note that the Registry Numbers for unspecified derivatives have a D appended.

IT 69-72-7D, copper-mixed ligand complexes 98-98-6D, copper-mixed
ligand complexes 7440-50-8D, mixed-ligand complexes with picolinic
and salicylic acids

IT 59-67-6DP, iron(III) complexes 89-00-9P 98-98-6DP, iron complexes
499-83-2P 7439-89-6DP, pyridinecarboxylic acid complexes

This strategy will retrieve references in the 10th and subsequent
Collective Indexes for complexes indexed as unspecified derivatives of
picolinic acids.

Unspecified Derivatives
Derivatives without Registry Numbers

Indexing for the 8th and 9th Collective Indexes did not include ligand Registry Numbers as part of unspecified derivative entries. Instead of the Registry Number, the complete CA Index Name for the ligand was used when unspecified derivative entries were made. The name included the word "complex." To retrieve the unspecified complexes that were indexed in the 8CI and 9CI, you could enter the 8CI and 9CI names for all the answers in L3 (L3 = substructure search for all picolinic acid derivatives) as search terms in the CA File. Then, these names would be linked to the word COMPLEX?. Or, short of this massive effort, you can search for complexes of ligands which are named as a picolinic acid (the 8CI Index Name for those ligands which are not part of fused ring systems) or a 2-pyridinecarboxylic acid (the 9CI Index Name). Complete recall cannot be guaranteed by this method, but at least some new references can be retrieved this way.

=> FIL CA

.
.
.

=> S PICOLINIC (W) ACID (L) COMPLEX? RANGE=(8CI)

.
.
.

=> S PYRIDINE (W) CARBOXYLIC (L) COMPLEX? RANGE=(9CI)

.
.
.

Examples of answer to these two searches are shown below. We have underlined the terms causing the retrieval.

IT Picolinic acid, iron complexes

TI Iron complexes with pyridinecarboxylic acids

Unspecified Derivatives Manually Registered Substances

Unspecified derivatives that do not have their own Registry Numbers should not be confused with the ligand heading index entries for completely structured and registered complexes. Entries at the ligand heading in the Chemical Substance Index are alternate names for the same complexes which you retrieve through a substructure search for metal complexes. These alternate names can be searched in the Basic Index or Complete Name index of the Registry File. For example, a completely defined tungsten complex with picolinate with the Registry Number 61880-51-1 has a synonym of 2-pyridinecarboxylic acid, 5-methyl-, tungsten complex. The alternate name at the ligand appears to be an unspecified derivative but is, in fact, a synonym for a fully defined complex.

References in the CA File for unspecified derivatives are in addition to those obtained via a substructure search for metal complexes. This crossover strategy to retrieve unspecified derivatives is very important when complete recall is needed. It is also an important technique for retrieving references to kinetic or thermodynamic properties, as complexes studied by these techniques frequently are indexed only as unspecified derivatives.

Manually Registered Compounds

Another type of substance that cannot be retrieved by substructure searching is one for which the formula is known, but not the structure. Substances of this sort do receive Registry Numbers, but they receive the designation Manual Registration because they have no structure connection table as part of the record in the Registry File. Manually registered substances can be searched in the dictionary portion of the Registry File by using name fragments in combination with the Class Identifier MAN and M>0 to require the presence of a metal. Again, it is possible to retrieve substances with name searching that you cannot get any other way, but complete retrieval is not possible.

Using the picolinic acid as an example, the strategy is:

=> S (2(W)PYRIDINE(W)CARBOXY? OR PICOLINIC)(L)COMPLEX AND MAN/CI AND M>0

Unspecified Derivatives
Manually Registered Substances

The following is an example of an answer that would be retrieved.

RN 60495-65-0
IN Iron, dihydroxytetrakis(2-pyridinecarboxylato)di- (9CI)
SY 2-Pyridinecarboxylic acid, iron complex (9CI)
MF C24 H18 Fe2 N4 O10
CI CCS, MAN

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***

Summary

In this case study we have discussed a variety of different search strategies needed to complement substructure searching in order to retrieve the maximum number of references to coordination compounds. Some substances are just not well enough characterized to have a structure in the Registry File. These principles can be applied to a variety of topics in coordination chemistry.

In Case Study 5 we will discuss chelates and macrocycles.

CASE STUDY 5: CHELATES AND MACROCYCLES

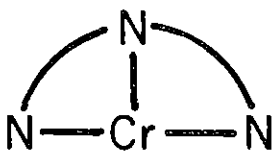
Chelate ring formation gives rise to a number of interesting features of coordination chemistry. Rather than focusing on the particular ligand donor atoms attached to the metal, you may wish to search for chelate rings of various sizes, or to allow for chelate rings with heteroatoms in the backbone, or to retrieve multidentate ligands with a particular configuration.

As we observed in Chapter 4, specifying ring bonds between the metal and donor atoms retrieves chelate rings. This is useful to allow for chelates of any ring size and any composition, but it is not very selective when searching for specific polydentate or macrocyclic rings.

The structure building tools of VARIABLE and REPEATING Gk groups are especially powerful for searching for specific types of chelate rings.

Tridentate Chelates

Chromium complexes of any tridentate nitrogen ligand are desired. The size of chelate ring may range from four to seven atoms and noncarbon atoms may be present in the ligand backbone.



Chelates and Macrocycles

Tridentate Chelates

Structure A, shown below, is used to describe the chelate ring. REpeating G1 Groups are used to allow for the variable size of the backbone. The A (Any atom) is used in the REPEAT definition to allow for any element to be part of the ring. The bonds are left unspecified because a wide variety of tridentate ligands is expected.

```

N?????G1
?      ?
?      ?
Cr????N      REP G1=(1-4) A
?      ?
?      ?
N?????G1

```

Structure A

The objective of this study is to retrieve tridentate ligands, but not tetradentate ligands. This can be accomplished by creating a second structure, Structure B, with three nitrogens and an additional donor group and adding this structure to the query using the NOT operator.

```

N?????G1
?      ?
?      ?
A?????Cr????N      REP G1=(1-4) A
?      ?      ?
?      ?      ?
G1????N????G1

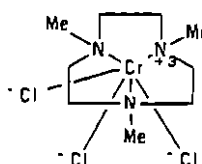
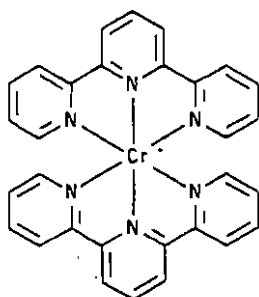
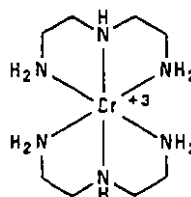
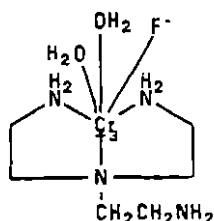
```

Structure B

Query: STR A not STR B

Chelates and Macrocycles
Tridentate Chelates

Included in the answer set are substances with either one or two open tridentate ligands and some with macrocyclic tridentates. Examples are shown below.



Whenever you use NOT logic to restrict an answer set, you lose any answers that have the desired feature in combination with the undesired feature. Any complex that has both a tridentate ligand and a larger polydentate will not be retrieved by this strategy. We have again traded loss of a few possible answers for the ability to exclude many unwanted ones.

Chelates and Macrocycles

Tridentate Chelates

It may have been tempting to use a CONNECT attribute to prevent closing the macrocycle by specifying Exactly 2 Ring bonds on the terminal nitrogens, but the answers displayed show that some ligands have the nitrogen embedded in an aromatic system, giving it three ring bonds.

If you specify a metal with an extensive nitrogen chemistry, such as nickel or cobalt, coordinated to the metal, the search may need to be divided with screens. For nickel, you can search for all six coordinate complexes in one answer set by adding a screen for a Degree of Connectivity of six or greater, DC 2126.

Query: STR A not STR B and SCR 2126

Four and five coordinate complexes can be obtained from a second search in which you use screens for Degree of Connectivity four, DC 2120, not six.

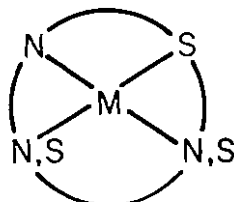
Query: STR A not STR B and SCR 2120 not SCR 2126

For cobalt, the search for six coordinate complexes has to be divided again using a Ring Count screen for four or more rings. Remember to call the Search Assistance Desk when a search must be divided.

Chelates and Macrocycles
Tetradentate Macrocycles

Tetradentate Macrocycles with Both Sulfur and Nitrogen Donors

In this example, a macrocycle such as



is desired. REpeating G1 groups will be used for the backbone of the macrocycle to allow the sizes of individual rings to vary from four to seven members. The structure, Structure C, is a closed circle, so macrocycles will be the only answers possible. Both nitrogen and sulfur donor atoms are required; but acceptable combinations are one nitrogen and three sulfurs, two nitrogens and two sulfurs, or three nitrogens and one sulfur. To meet these criteria, node 1 is required to be nitrogen, node 3 is required to be sulfur, and nodes 5 and 7 are defined as VARIABLE G2 groups which may be either nitrogen or sulfur.

```

      1      2
      ?N???G1
      ? -      ?
8 G1? -      ? S 3
      ? -      -?
      ? -      - ?
7 G2 -      - ?G1 4
      ? -      -?
      - ?G1????G2
      - -6      - 5
      --      --
      --      -
      --      -
      --      -
      M
9

```

REP G1=(1-4) A
VAR G2=N/S

Structure C

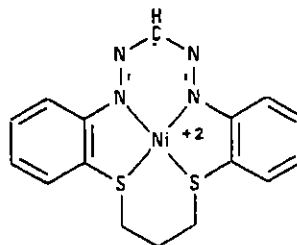
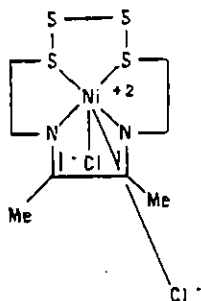
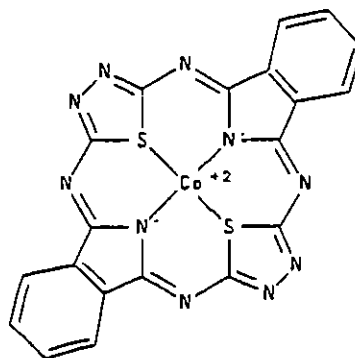
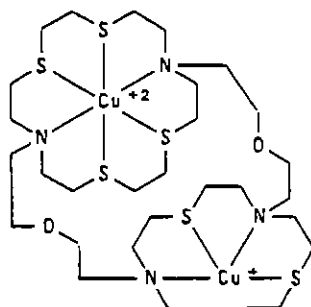
Chelates and Macrocycles
Tetradentate macrocycles

A number of screen sets must be manually added because this is a very generic query. As customary, the standard screens for transition metals, 1921 OR 1931 OR 1964, are used. AA Screen 945 for four ring bonds on the generic M node should be added, as well as the RC Screen 1841 for four or more rings. The search is then divided into two portions based on a Ring Count of eight or more, RC 1845.

Query 1: STR C and SCR (1964 or 1921 or 1931) and SCR (945 and 1841) and SCR 1845

Query 2: STR C and SCR (1964 or 1921 or 1931) and SCR (945 and 1841) not SCR 1845

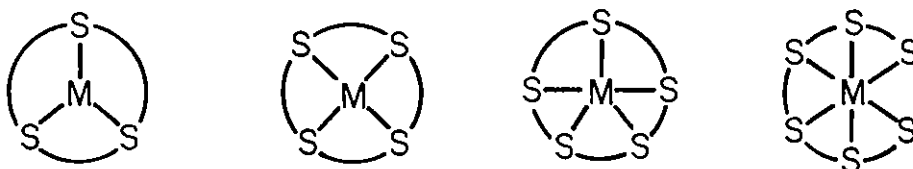
Looking at the selected answers below, we see that retrievals encompass both five- and six-membered rings, both nitrogen and sulfur as well as carbon in the backbone, and binuclear complexes. Coordination numbers of four, five, and six are represented.



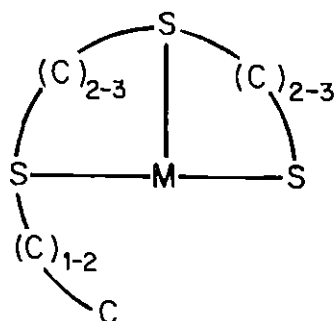
Chelates and Macrocycles
Varying Numbers of Donor Sites

Macrocyclic Complexes With Varying Numbers of Donor Sites

Rather than searching for tetradentate macrocycles as in the previous example, we now wish to search for a series of macrocycles with nothing but sulfur donors, where the number of sulfur atoms can vary. At least three sulfur atoms need to be present, and we wish to require that the backbone have either five or six-membered rings which include two or three carbon atoms. One way to do this would be to build a series of macrocycles with increasing numbers of sulfur atoms.



A strategy was devised for this problem which avoids having to create different structures to accommodate the varying number of sulfur atoms.



Chelates and Macrocycles

Varying Numbers of Donor Sites

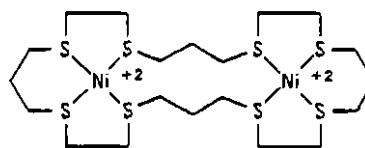
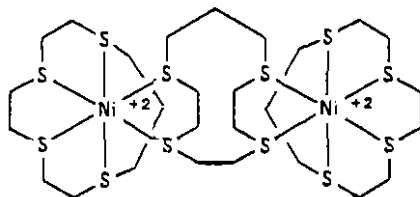
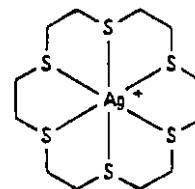
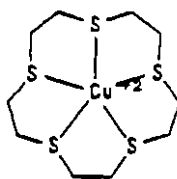
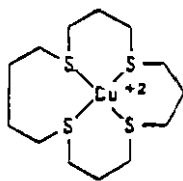
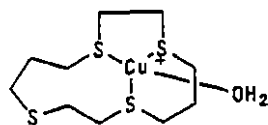
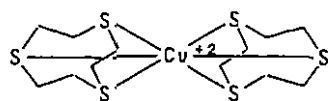
One structure, Structure D, is built in which one ring is left open in order that the larger macrocycles can be generated by substitution of more rings with sulfur at carbon 9 and sulfur 7. A CONNECT Attribute on sulfur 7 requires an additional ring bond to be present. We hope that, when these other rings form, a sulfur macrocycle will be formed, but other groups could be introduced as well. We suspect, however, that most macrocycles with at least three sulfur atoms will not have too many other kinds of donors.

2	3	4	
G1----	S-----	G1	BON ALL RSE
-	-	-	CON 7 M3 R
-	-	-	REP G1=(2-3) C
-	-	-	REP G2=(1-2) C
1 S-----	M-----	S	
-	6	7	
-			
-			
G2-----	C		
8	9		

Structure D

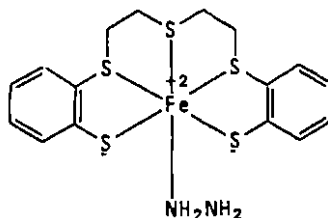
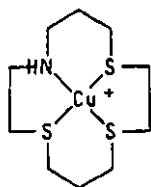
Chelates and Macrocycles
Varying Numbers of Donor Sites

The customary transition metal screens are added to the query. Tri-, tetra-, penta-, and hexadentate macrocycles are all retrieved, including



Chelates and Macrocycles Varying Numbers of Donor Sites

The open sites on the query structure did allow for substitution of any type of ring, so answers like those below do not fit the intent of the query.



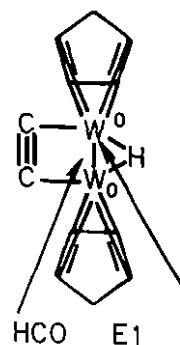
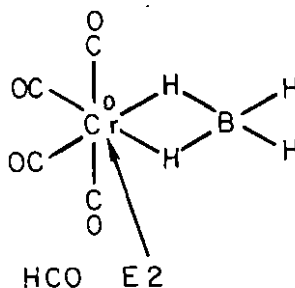
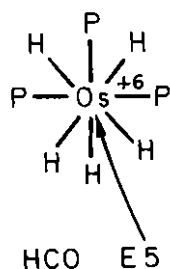
Summary

This case study illustrated ways of using repeating Gk groups to allow for retrieval of chelates and macrocycles.

Case Study 6 will discuss hydrides.

CASE STUDY 6: HYDRIDES

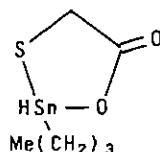
The hydride ion, $H(1-)$, functions as either a monodentate or a bridging ligand in many coordination compounds. Metal hydrides occur frequently in both organometallic and cluster compounds. The best way to search for metal hydrides is to specify a Hydrogen Count on the metal node. The following structures have the Hydrogen Counts on the metal nodes indicated below each structure.



Hydrogen bonded to a metal center can be either a connection table hydrogen or a nonconnection table hydrogen. Any hydrogen which has more than one bond, i.e., bridging hydrides, or any hydrogen which is attached to a metal center that has a number of bonds greater than its oxidation state is a connection table hydrogen. Protons and hydride ions, i.e., hydrogens bearing charges, are also connection table hydrogens. All of the hydrogens illustrated in the structures above are connection table hydrogens.

Hydrides

Any hydrogen connected to a metal center which has a number of bonds exactly equal to its oxidation state is a nonconnection table hydrogen. An example of a nonconnection table hydrogen is the one on the four coordinate Sn(4+) atom in the structure shown below.



All hydrogens connected to metals, whether connection or nonconnection table, may be retrieved by use of the Hydrogen Count (HCO) attribute. Hydrogen Counts may be assigned as either Minimum or Exact values. All Exact Hydrogen Counts are converted by the system to Minimum values when substances that have one or more connection table hydrogens are iterated.

Should you wish to designate an Exact or a Maximum Hydrogen Count for an input structure, you can do so by building a second structure consisting of a single metal atom node which you add to the query in NOT logic. Assign a Minimum Hydrogen Count to the metal atom one increment larger than the HCO in the first structure. You must also assign a Node Specification of Ring or Chain to the metal atom.

To search for all metal complexes with at least three hydride and three phosphine ligands, the preferred strategy is to create the structure below, Structure A, and add a Hydrogen Count of a Minimum of three to the metal node.

```

      4
      P
      ?
      ?
      ?
P???????M???????P
1         2         3

```

BON ALL RC
HCO 2 M3

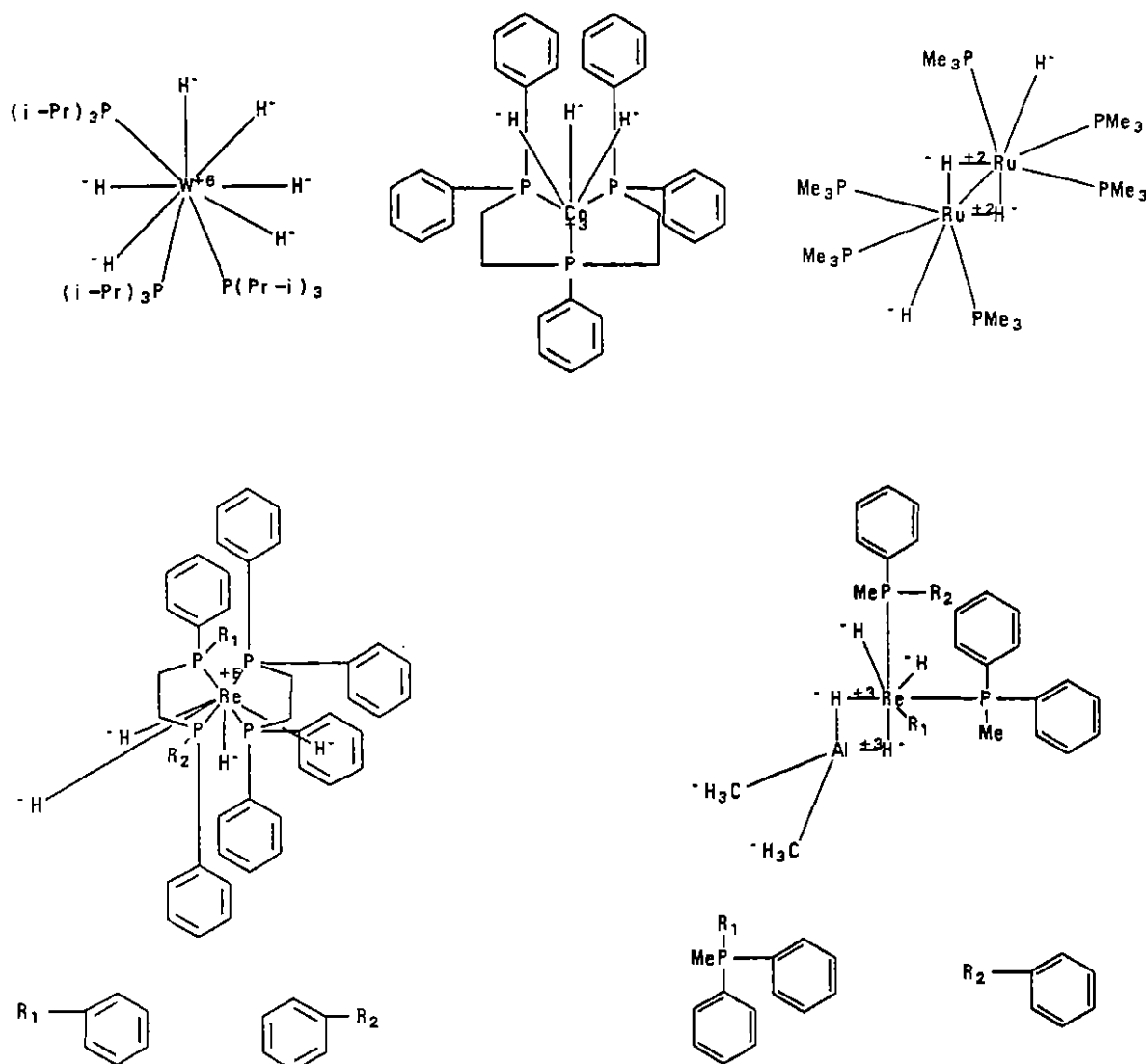
Structure A

Screens for whatever metals you require are added. In this case, the search was split using Screen 1964 for the first row transition metals, and Screens 1921 or 1931 for the second and third row transition metals.

Query 1: STR A and SCR 1964

Query 2: STR A and SCR (1921 or 1931)

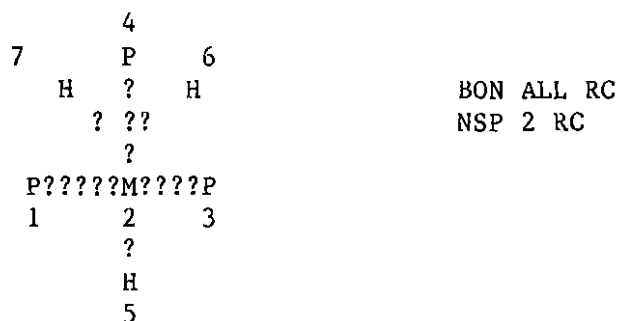
Among the answers retrieved are compounds with three to six hydrides, both monodentate and bridging, exhibiting coordination numbers six through nine.



Hydrides

Structures can also be input with a hydrogen directly bonded to the metal. In this case, the search system will treat those hydrogens as if a Minimum Hydrogen Count attribute had been assigned. No differentiation is made between connection and nonconnection table hydrogen nodes.

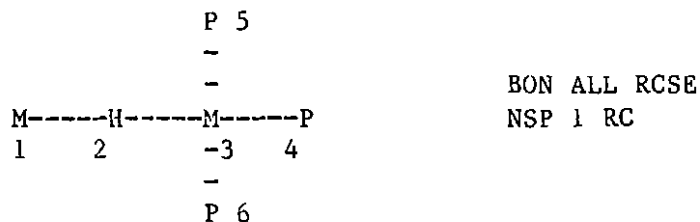
If you build a structure, Structure B, which has metal-hydrogen bonds, you should put a Node Specification of Ring or Chain on the metal node. In the case of hydrides, specifying a M--H bond as Ring or Chain does not force the node specification to follow because the system converts the hydride to a Hydrogen Count and does not see any M--H bond there. This results in the metal node reverting back to being a chain node. To allow for the possibility of it being in a ring, the Node Specification of Ring or Chain is required.



Structure B

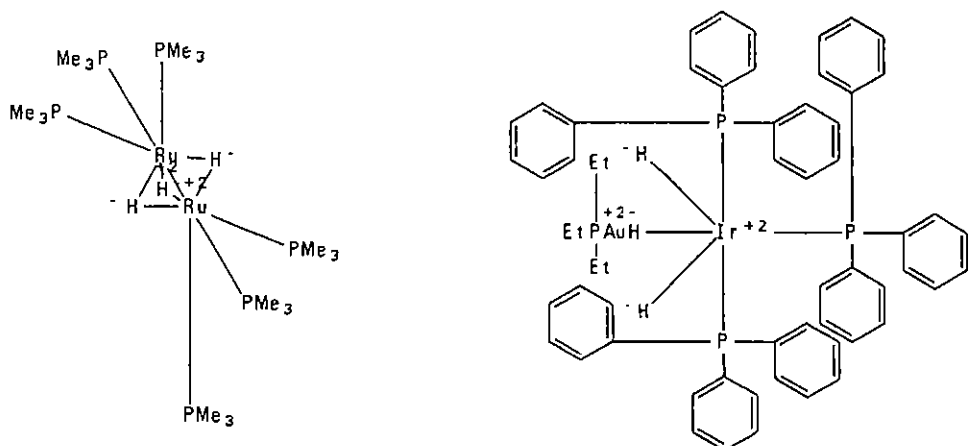
To the system, Structure A and Structure B are identical structure queries and give the same search results.

If you wish to retrieve only substances with bridging hydrides, you should create a node for the hydrogen atom. Add a Node Specification of Ring or Chain to any metal atom which has bonds only to hydrogen. Structure C illustrates this type of structure.



Structure C

Examples of answers retrieved by a search for Structure C are shown below.

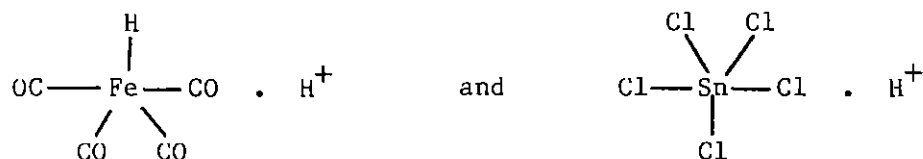


There is really no good way to limit retrievals to a specific coordination number for metals coordinated to hydride ligands. The CONNECT attribute is not effective on metal nodes which have Hydrogen Counts or H nodes, as the system ignores the CONNECT Attribute when it iterates an answer that has a connection table hydrogen. Thus, many of the answers you get may not have the number of hydrogens you are expecting.

Degree of Connectivity Screens include connection table hydrogens in the connectivity count, but using a DC Screen will cause you to loose answers for Sn, Ge, Bi, and Sb, or other metals with nonconnection table hydrogens.

It is also not advisable to use a Variable Gk group which includes a H as a substituent on any node to which you intend to assign CONNECT or an HCOUNT. If you do use a Variable Gk group, the answer set you get may contain complexes that do not have the desired number of hydrogens.

All hydrogens in the formula of a complex are not necessarily coordinated to the metal. Sometimes the hydrogen may be a proton functioning as a counter ion, H⁺. A proton is treated like any other counter ion and is structured as a separate component, as illustrated in the substances below.



Hydrides

There is a Graph Modifier screen, 2033, for hydrogen as a single atom fragment which can be used as part of the query if you want to retrieve H⁺ salts.

Summary

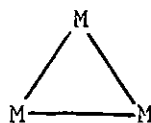
This case study has discussed ways of searching for hydrides. The next case study discusses metal clusters.

CASE STUDY 7: METAL CLUSTERS

Metal clusters and metal-metal bonding are areas of ever-increasing interest to inorganic chemists. They are topics where substructure searching can be particularly powerful. You can search for large or small clusters, clusters with or without bridges, or clusters with a particular geometry.

This case study will illustrate general techniques for building structures and selecting appropriate screens for searching clusters. As you will see from the answers displayed, metal clusters exhibit considerable variety in their structures.

Rings of three metal atoms predominate in metal cluster chemistry, but they are by no means universal.



There are also many dimers, a number of clusters with the metals in a linear arrangement, and some with rings of four or more metal atoms. When bridging groups are present, they usually form three- (occasionally four-) membered rings and commonly include a metal-metal bond.

Search examples discussed in this case study will be gold-phosphine clusters and iron carbonyls.

Gold Clusters with Phosphine Ligands

The objective of this example is to illustrate that very simple structures will suffice to retrieve metal clusters. A very general way to search for clusters is to create a structure with a metal-metal bond which has the desired ligand node attached to the metal. For example, this search requires clusters with gold-gold bonds and at least one phosphine ligand coordinated to a gold atom. All bonds should be specified as Ring

Metal Clusters
Gold-Phosphine Clusters

or Chain as we want to allow the possibility for both open chain and cyclic clusters. The phosphine ligand could bridge two metals and be part of a ring, or it could be in terminal positions with chain bonding. Structure A is used.

P????Au????Au
1 2 3

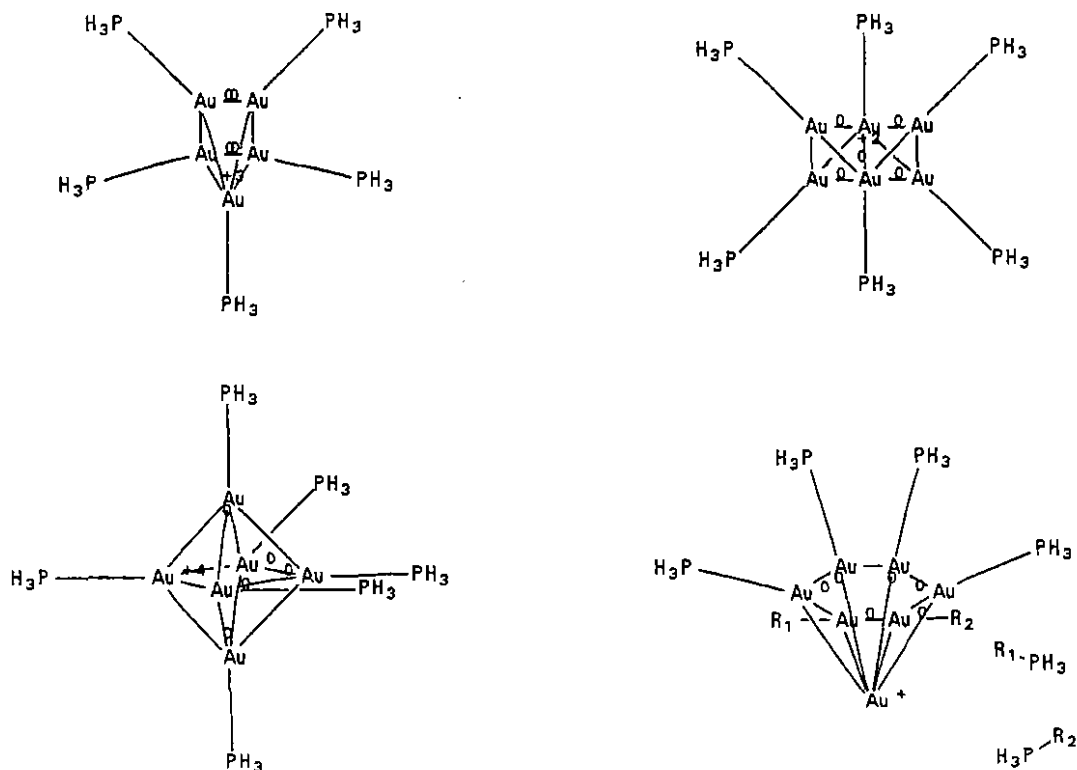
BON ALL RC

Structure A

The search objective was then refined to restrict the search to clusters with five or more gold atoms and five or more phosphorus atoms. There are two ways to do this. The first way makes use of the substructure answer set, L2, and refine it with formula searching in the dictionary portion of the Registry File by requiring that more than four gold atoms and more than four phosphorus atoms be present.

=> L2 and AU>4 and P>4

Some of the answers are shown below.



Metal Clusters
Gold-Phosphine Clusters

A second search strategy for clusters of this size is to create a structure, Structure B, which contains a minimum of five gold atoms. This can be done by using a REpeating G1 of between five and twenty gold atoms. (A maximum of twenty units may be included in a REpeating Group.) Bonds to the phosphines should be Ring or Chain. The system automatically assigns Ring or Chain bonds within the REpeating Group since the bonds leading to the REpeating Group are Ring or Chain.

```
P????G1????P      REP G1=(5-20) Au
 1      2      3
```

Structure B

For gold-phosphine clusters, Structures A and B give identical answer sets. However, this is not universally true, as the element count for metals in the dictionary part of the file will count all metals in the structure, whether they are part of the cluster or not.

The metal clusters which are structured in the Registry File are those for which the bonding is well elucidated. If the cluster has a definite formula, but the actual cluster geometry or location of substituents is not completely defined, the substance will have a Manual Registration and cannot be retrieved by a substructure search. A check of the dictionary portion of the Registry File using both the molecular formula portion and the class identifier code for manually registered substances will usually give additional substances. The search strategy for checking the dictionary portion of the file for this study is shown below.

=> S AU>4 and P>4 and MAN/CI

An example of the type of answers this search retrieves is shown below.

```
RN  67954-31-8
IN  Gold, triiodoheptakis[tris(4-methylphenyl)phosphine]undeca-,
    (31Au-Au) (9CI)
SY  Phosphine, tris(4-methylphenyl)-, gold complex (9CI)
MF  C147 H147 Au11 I3 P7
CI  CCS, MAN
*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
```

Metal Clusters

Iron Carbonyl Clusters with Terminal Carbonyls

Iron Carbonyl Clusters

A very large number of iron clusters are known. They range from simple dimers to large multiatom clusters with a variety of other metals and other ligands.

You can include only one Fe node in a structure and make the others M. This will include more mixed metal clusters. Or, you can restrict, but not exclude entirely, the number of mixed metal complexes by making all nodes Fe. Clusters which are larger than those required by the substructure can still have other metals present.

Terminal Carbonyl Groups

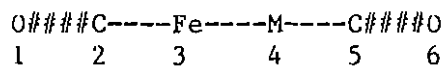
The most generic structure for an iron cluster with a terminal carbonyl is shown below.



This is far too general to search, however; and, most likely a chemist would have other substituents in mind which would make the structure more specific. We will consider another terminal carbonyl and will place it on the second metal atom. The cluster could be either open chain or cyclic. To allow for this, it is best to build two structures, Structure C and Structure D, and use exact bonding in each. (This requires fewer iterations than using a Ring or Chain bond.) Then, search the two structures using OR logic.

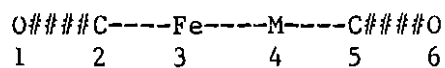
Metal Clusters

Iron Carbonyl Clusters with Terminal Carbonyls



BON 1-2 5-6 T,
2-3 3-4 4-5 SE

Structure C

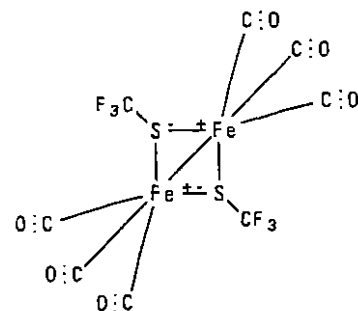
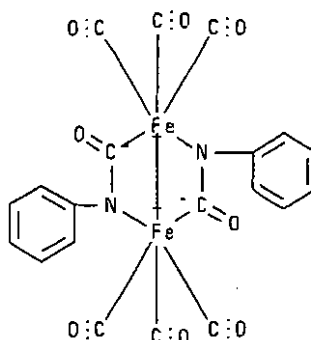
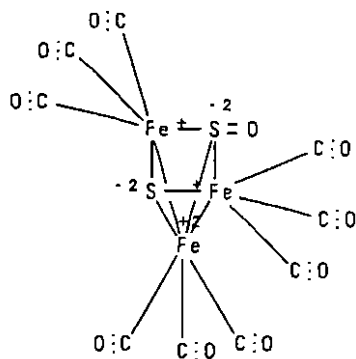
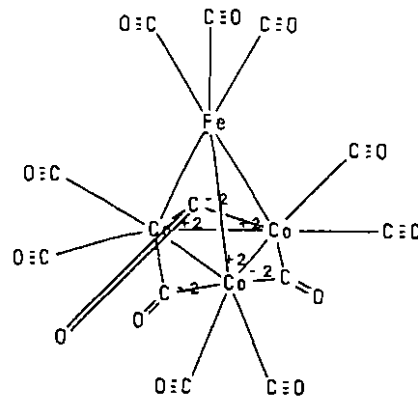
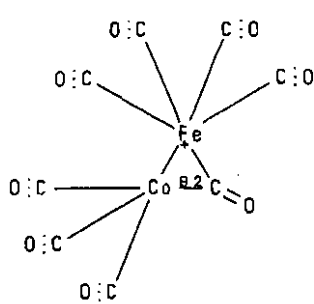
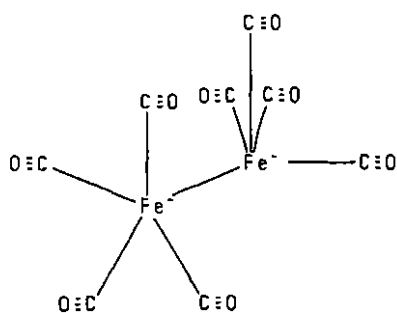


BON 3-4 RSE, 1-2
5-6 T, 2-3 4-5 SE

Structure D

Query: STR C OR STR D

A selection of answers that would be retrieved by this query are shown below.



Metal Clusters

Iron Carbonyl Clusters with Terminal Carbonyls

Using structures such as those that follow will give more specific answer sets. To get clusters with all, or mostly all, Fe atoms, create a structure with an Fe REpeating group, Structure E. The bonds within the REpeating G1 will be the same as those specified leading to the G group. Assign bonds 2-3 and 3-4 as Ring or Chain Single Exact to allow the cluster to also have Ring bonds.

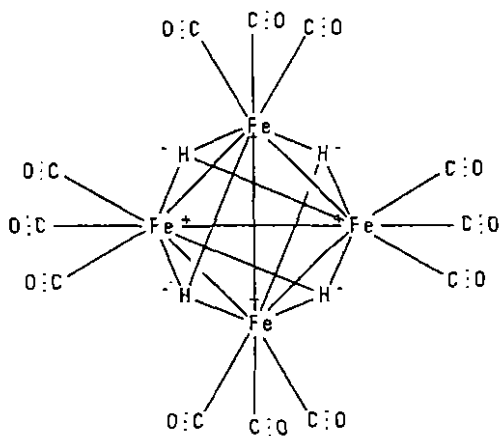
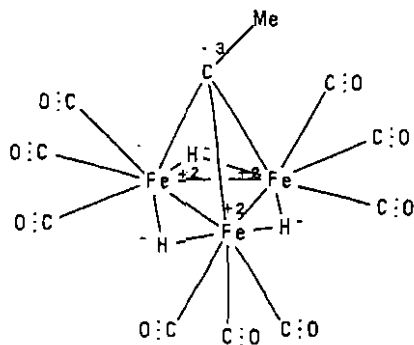
```

O####C----G1----C####O      BON 1-2 5-6 T,
1      2      3      4      5      2-3 3-4 KCSE,
REP G1=(1-12) Fe

```

Structure E

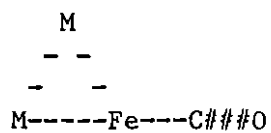
This will restrict the answer set somewhat to compounds such as those shown below.



Metal Clusters

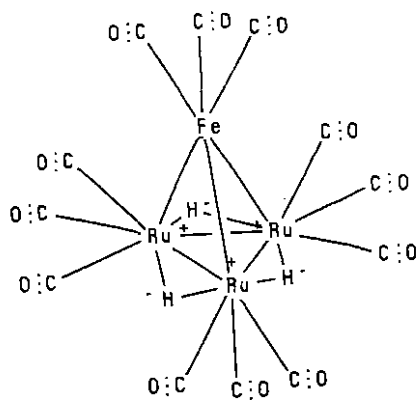
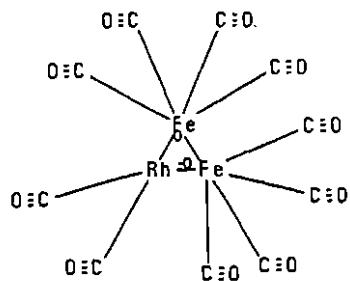
Iron Carbonyl Clusters with Terminal Carbonyls

You can use a ring structure, Structure F, as another possible query.



Structure F

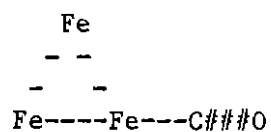
This query retrieves answers such as those shown below.



Metal Clusters

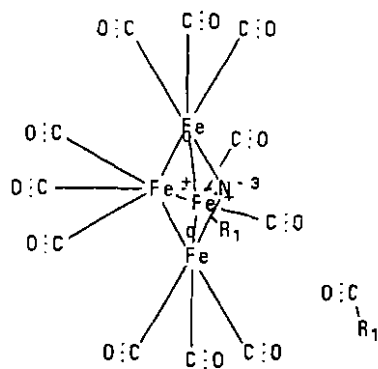
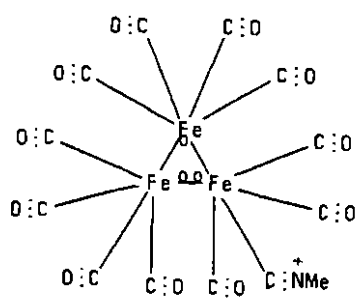
Iron Carbonyl Clusters with Terminal Carbonyls

You could use a ring structure with only Fe atoms in the ring,
Structure G.



Structure G

This is a more restrictive query and retrieves answers such as the ones shown below.

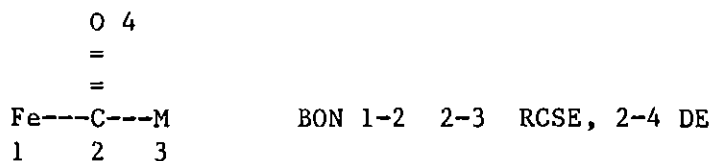


Metal Clusters
Iron Carbonyl Clusters with Bridging Carbonyls

Bridging Carbonyls

If you wish to require the presence of bridging carbonyls, the following structures offer several options. You also have the option of requiring both bridging and terminal carbonyl groups in the same structure. Note that the bonding in the CO group changes from Triple to Double Exact when it becomes a bridging group.

This structure, Structure H, is the most general. It requires only one iron with one carbonyl bridge. A metal-metal bond is not required, although it is allowed.



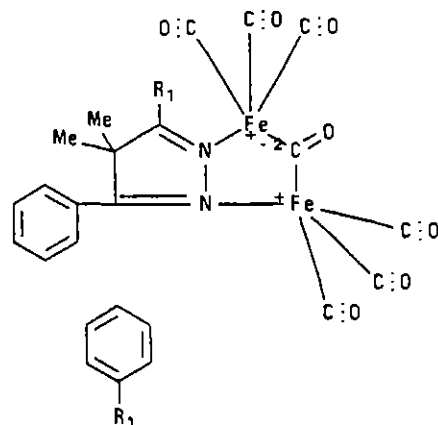
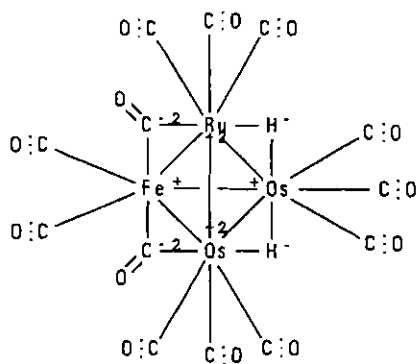
Structure H

A search for this structure will have too many iterations for a single online search. Many answers will have a number of three-membered rings, so the search can be divided on the basis of ring count. Screen 1842 for a ring count of five or more is satisfactory.

Query 1: STR H and SCR 1842

Query 2: STR H not SCR 1842

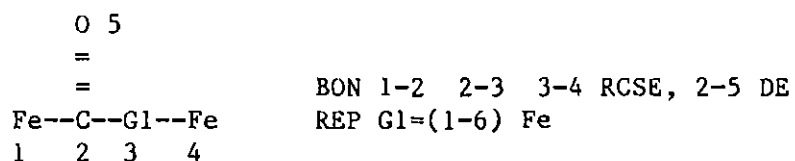
The following examples illustrate answers which can be retrieved only by this very general strategy.



Metal Clusters

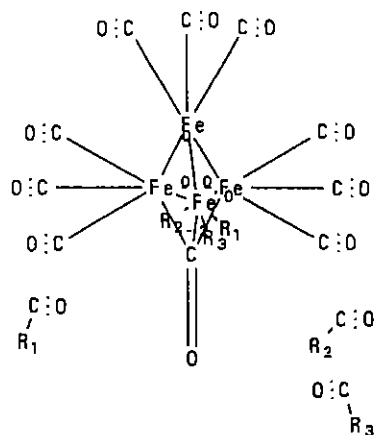
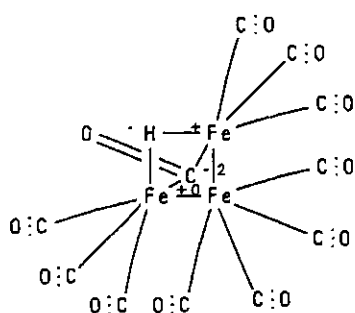
Iron Carbonyl Clusters with Bridging Carbonyls

You can require between three and eight irons in an answer with the Structure I which uses an Fe REpeating Group. All bonds involving Fe or Gl should be specified Ring or Chain. A search of this structure runs to completion unaided.



Structure I

Answers containing several iron atoms are shown below.



Metal Clusters

Iron Carbonyl Clusters with Bridging Carbonyls

If you wish to restrict answers to substances with several CO bridges between iron atoms, you can use an Fe-CO REpeating group, Structure J.

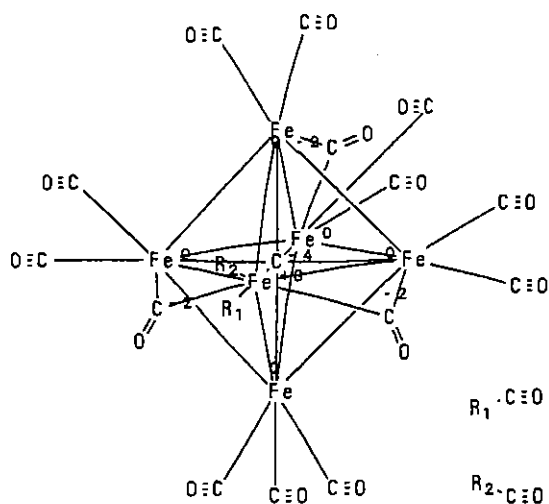
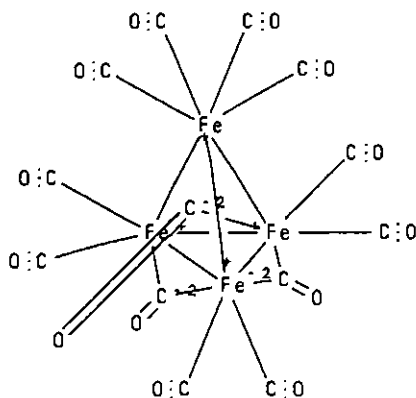
Fe---G1---Fe
1 2 3

Fe---C≡O
@4 @5 6

BON 1-2 2-3 4-5 RCSE, 5-6 DE
REP G1=(1-6) 5-1 4-3

Structure J

Examples of answers are shown below.



Metal Clusters

Iron Carbonyl Clusters of Specific Size

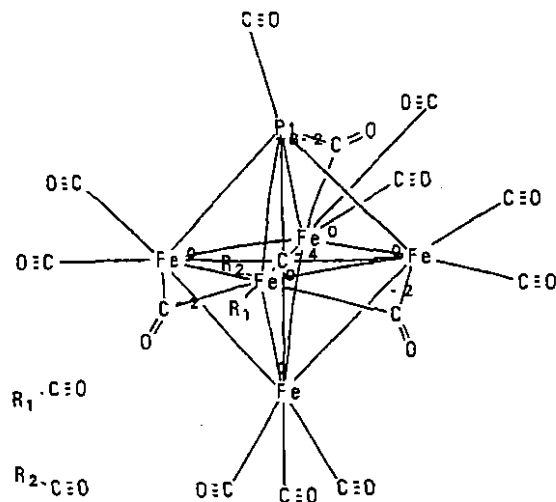
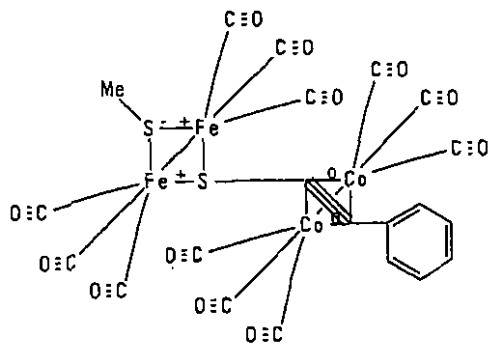
Clusters of Specific Size

To limit retrievals to clusters with a specific number of metal atoms, you can combine an answer set from a substructure search with an element count for M. This technique will give you the metal count over the entire cluster, not just within the ring system.

For example, you can restrict the answer set, L3, created from a search of the query shown on page 143, to answers that have more than three metal atoms in the following way.

=> S L3 AND M>3

This retrieves answers such as the ones shown below.

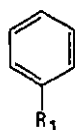
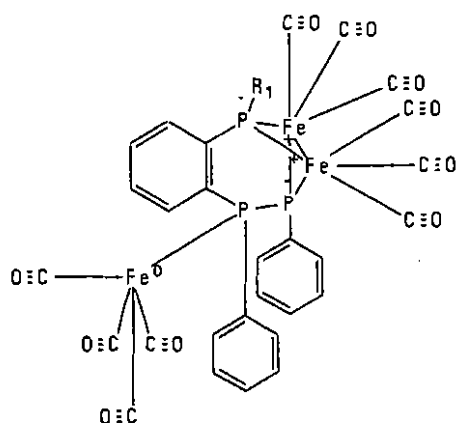
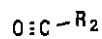
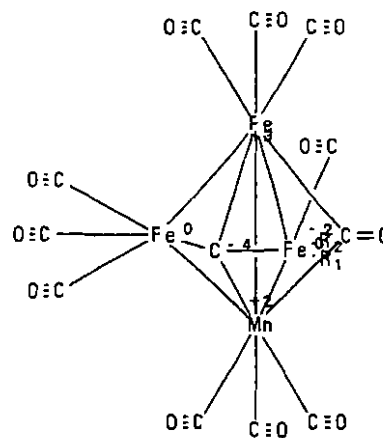
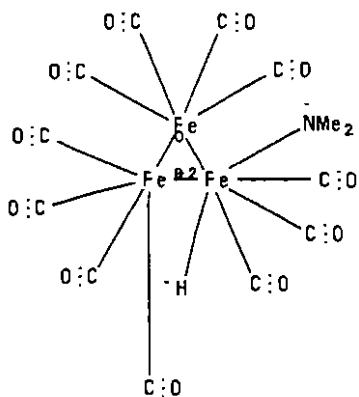


Metal Clusters Iron Carbonyl Clusters of Specific Size

Alternatively, you can require exactly three iron atoms in the following way.

=> S L3 AND 3/Fe

This retrieves answers such as the ones shown below.

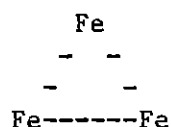


Metal Clusters

Iron Carbonyl Clusters of Specific Size

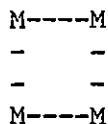
You can limit answers to clusters of a specific size, especially when you are looking for smaller clusters, by building two structures. One structure should be the precise cluster you seek, and the second should be one metal atom larger than you wish to retrieve. The larger structure is excluded from the answers by adding it to the query using the NOT operator.

If you wish to retrieve clusters with exactly three iron atoms, and no other metals in the cluster, build the three iron structure, Structure K.



Structure K

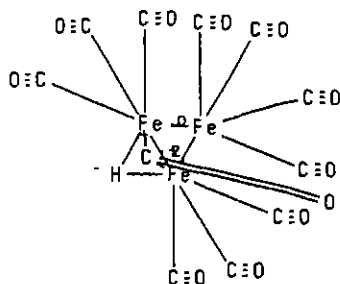
Then, build a second structure, Structure L, with a very generic four metal cluster. This structure will be used with NOT logic in the query.



Structure L

Query: STR K not STR L

Answers such as the one below are retrieved.



Metal Clusters
Iron Carbonyl Clusters of Specific Size

To limit answers to larger clusters, assign a large number of REpeating Groups in a structure, Structure M.

```
O####C-G1-C####O      BON 1-2 4-5 T, 2-3 3-4 RCSE
1      2 3 4      5      REP G1 = (5-6) Fe
```

Structure M

This will retrieve clusters with 5 or 6 or more Fe atoms.

Summary

This case study has discussed strategies from searching for metal clusters. Case Study 8 will discuss stereochemistry.

CASE STUDY 8: STEREOCHEMISTRY

The stereochemistry of coordination compounds is a topic which greatly interests inorganic chemists. The CAS Registry System, however, is limited to two dimensions, so stereochemistry is not directly searchable with structures. In other words, stereochemistry is not in the structure connection table.

When a document describes stereochemistry for a mononuclear compound, this fact is noted during registration by a Registry System stereochemical descriptor phrase. For coordination compounds, this notation was begun in 1972 for coordination numbers four, five, and six, and extended to coordination numbers seven, eight, and nine in 1977. The descriptors are being added to records of substances registered prior to 1972 or 1977 if they are cited again in the literature. For substances registered since the 9CI, the stereochemical text descriptor has also been included as part of the Index Name of the substance. During the 8CI, the Index Name includes the searchable terms, i.e., cis, trans, fac, mer, +, -, delta, lambda, D, or L, if such stereochemical information is reported for the compound. For a few substances that enter the Registry File as part of the pre-1965 registration project, stereochemical information present in the original document cannot be included in the Registry record because of limitations inherent in the indexes used for this pre-1965 registration.

The following examples show the various forms of stereochemical information available. We have underlined this information in the various online display fields.

EXAMPLE 1

IN Cobalt(1+), dichlorobis(1,2-ethanediamine-N,N')-, (OC-6-12)- (9CI)
SY Cobalt(1+), dichlorobis(ethylenediamine)-, ion, trans- (8CI)
ST 7:OC-6-12

Stereochemistry

EXAMPLE 2

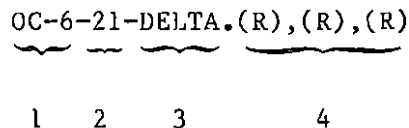
IN Cobalt(3+), tris(1,2-propanediamine-N,N')-,
[OC-6-21-DELTA.-(R),(R),(R)]- (9CI)
 SY 1,2-Propanediamine, cobalt complex, (R)- (9CI)
 SY trans-Tris(R-propylenediamine)cobalt(3+)
 SY Cobalt(3+), tris(1,2-propanediamine)-, ion, L-trans-(-)-(-)-(-)- (8CI)
 ST 7:OC-6-21-DELTA.(R),(R),(R)

EXAMPLE 3

IN Cobalt, [.mu.-(dimethylphosphino)][.mu.-[(dimethylphosphino)methyl-C:P]]
 tetrakis(trimethylphosphine)di-, (Co-Co), stereoisomer (9CI)
 ST *

The stereochemical notation system is composed of four parts:

1. the site symmetry term
2. the configuration number
3. the chirality label for the metal center
4. the ligand stereochemical label



If no chiral center is reported for the substance, only the first two parts denoting the geometry will be used.

The site symmetry term indicates the idealized geometry of the central atom by a letter code derived from the name of the coordination polygon. This is followed by the coordination number. These terms are summarized in the table on the next page.

SYMMETRY SITE TERMSFour-Coordinate Polyhedra

T-4	Tetrahedron
SP-4	Square Plane

Five-Coordinate Polyhedra

TB-5	Trigonal Bipyramid
SP-5	Square Pyramid

Six-Coordinate Polyhedra

OC-6	Octahedron
TP-6	Trigonal Prism

Seven-Coordinate Polyhedra

PB-7	Pentagonal Bipyramid
OCF-7	Octahedron Face Monocapped
TPS-7	Trigonal Prism Square Face Monocapped

Eight-Coordinate Polyhedra

CU-8	Cube
SA-8	Square Antiprism
DD-8	Dodecahedron
HB-8	Hexagonal Bipyramid
OCT-8	Octahedron Trans-Bicapped
TPT-8	Trigonal Prism Triangular Face Bicapped
TPS-8	Trigonal Prism Square Face Bicapped

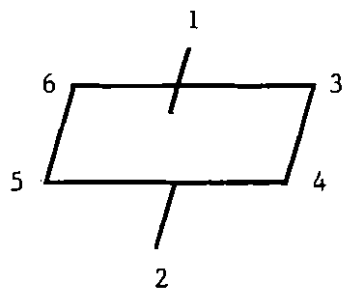
Nine-Coordinate Polyhedra

TPS-9	Trigonal Prism Square Face Tricapped
HB-9	Heptagonal Bipyramid

Stereochemistry

The second part, the configuration number, is one or more numbers used to identify atoms on symmetry elements of a structure. In this way, geometric isomers are distinguished. Configuration numbers are derived from priority numbers assigned to the donor atoms according to the Cahn - Ingold - Prelog sequence rules. The donor atoms are numbered in order of decreasing atomic number; the numbers derived this way are the priority numbers.

For octahedral complexes, atom number 1, the one with the greatest atomic number, is placed on the major axis. The first digit of the configuration number is the priority number of the atom trans to atom 1. The second digit of the configuration number is the priority number of the atom trans to the lowest numbered atom in the plane perpendicular to the major axis.



OC-6-25

A full description of the assignment of configuration numbers, including rules for other geometries and for complexes with unsymmetrical or multidentate ligands, is given in the references at the end of this case study.

The presence of a stereochemical text descriptor can be required during a substructure search through the use of Graph Modifier screens:

SCR 2065 - type 7: coordination compound stereochemical descriptor

SCR 2058 - type *: stereochemistry present, but cannot be expressed by CAS text descriptor conventions. This includes polynuclear complexes and complexes with pi bonding ligands.

Stereochemistry
Geometry

The text descriptor phrase of the Registry record is not searchable, but the fragments of any descriptor included in the Index Name are searchable in the dictionary portion of the file.

The assignment of stereochemical descriptors is complicated. They are not assigned for all types of coordination compounds. Furthermore, they are posted as part of the Index Name only for compounds which have 9CI names. For these reasons, the techniques in this case study are offered as a generalization which may be helpful in dividing an answer set into groups for further perusal. They may not give complete retrieval in some searches.

The following examples illustrate how some stereochemical information may be gleaned from the Registry File.

Geometry

The object of this search is to determine which nickel complexes with two or more phosphine ligands are tetrahedral, and which are square planar. This search question presupposes a four-coordinate complex, which is not indicated by the structure, Structure A. A CONNECT Attribute of Exactly 4 is assigned to the nickel node.

```
P???????Ni?????P      BON ALL RC, CON 2 E4
 1         2         3
```

Structure A

A screen for Degree of Connectivity of four or more, DC 2120, is added to to the query. Screens 2065 or 2058 for the presence of a stereochemical descriptor are also added.

Query: STR A and SCR 2120 and SCR (2065 or 2058)

Stereochemistry
Geometry

A search of this query retrieves substances that have stereochemistry described. These answers are stored in the answer set L4.

The tetrahedral descriptor, T-4, and the square planar descriptor, SP-4, are searched in the nomenclature portion of the file and used to sort the answers in answer set L4 into three sets.

=> S L4 AND T(W)4

All of the answers in this answer set, L5, are definitely tetrahedral. An example is shown below.

INDEX NAME = Nickel, tetrakis(diethylphenylphosphine)-, (T-4)- (9CI)

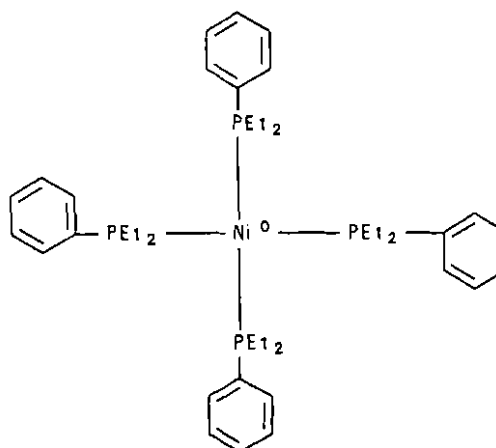
SYNONYM = Tetrakis(diethylphenylphosphine)nickel

SYNONYM = Phosphine, diethylphenyl-, nickel complex (9CI)

MOLECULAR FORMULA = $C_{40}H_{60}NiP_4$

SUBSTANCE CLASS = CCS

TEXT DESCRIPTOR = 7:T-4



=> S L4 AND SP(W)4

All of the answers in this answer set, L6, are definitely square planar. An example is shown below.

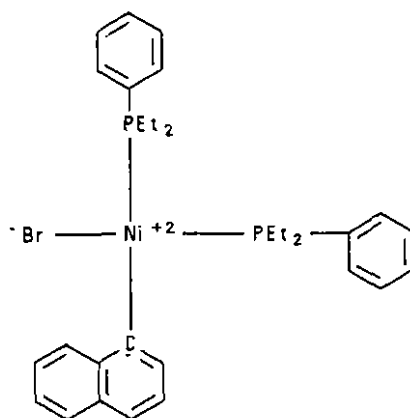
INDEX NAME = Nickel, bromobis(diethylphenylphosphine)-1-naphthalenyl-, (SP-4-3)- (9CI)

SYNONYM = Phosphine, diethylphenyl-, nickel complex (9CI)

MOLECULAR FORMULA = $C_{30}H_{37}BrNiP_2$

SUBSTANCE CLASS = CCS

TEXT DESCRIPTOR = 7:SP-4-3



Next we will sort out the answers that are not T-4 or SP-4.

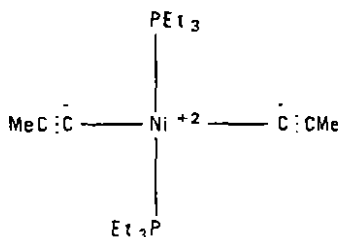
=> S L4 NOT L5 NOT L6

Stereochemistry

Ligand Stereochemistry

This answer set does contain some T-4 and SP-4 substances because the stereochemical descriptor is not part of the 8CI Name, as well as the other stereoisomers we were sorting. An example of an 8CI substance of this nature is shown below.

INDEX NAME = Nickel, di-1-propynylbis(triethylphosphine)-, *trans*- (8CI)
 MOLECULAR FORMULA = $C_{18}H_{36}NiP_2$
 SUBSTANCE CLASS = CCS
 TEXT DESCRIPTOR = 7:SP-4-1



Ligand Stereochemistry

The purpose of this study is to find substances with two bidentate ligands which contain optically active phosphines and/or arsines.

In this case, the emphasis is on the ligand stereochemistry. The text descriptor screens cannot single out substances for which the stereochemical notation includes ligand chirality, as opposed to those where there is merely metal atom geometry described. Screens 2065 and 2058 will not be very useful in this problem.

Stereochemistry
Ligand Stereochemistry

For the P or As to be optically active, it must have four different attachments, of which one may be a hydrogen. We will use a CONNECT Attribute of a Minimum of 3 on the phosphorus or arsenic to require at least one additional substituent. The structure, Structure B, is created to allow both five or six membered carbon containing chelate rings.

```

6  G2?????G1  5      As
   ?      ?      @8
   ?      ?
   ?      ? 1      2  P      CON 8 9 M3
7  G1?????M?????G1  @9      VAR G1=8/9
   ?      ?      REP G2=(2-3) C
   ?      ?
   ?      ?
   G1?????G2
      4          3

```

STRUCTURE B

There is no way in structure building to require that substituents be different, so we will NOT out a structures which have two of the most common substituents using Structure C.

```

6  G2?????G1  5      Me-P-Me
   ?      ?      @5
   ?      ?
   ?      ? 1      2  Et-P-Et
7  G1?????M?????G1  @6
   ?      ?      VAR G1=5/6/7/8/9/10
   ?      ?      REP G2=(2-3) C
   ?      ?
   G1?????G2
      4          3
      Me-As-Me
      @8
      Et-As-Et
      @9
      Ph-As-Ph
      @10

```

Structure C

Stereochemistry
Ligand Stereochemistry

In addition to Screens 1921, 1931, or 1964 for the transition metals, AA Screen 945 for four ring bonds to a central atom was added. (Screen 945 will not be used automatically when the generic M node is assigned.)

Query: STR B not STR C and SCR (1921 or 1931 or 1964) and SCR 945

Answers to the substructure search of this query are in L12. They were combined with a search in the dictionary portion of the file to select the complexes where the ligands had been designated as chiral or where a stereoisomer was indicated. R or S indicates an optically active atom where the absolute configuration is known; R* or S* indicates the relative configuration of a chiral atom.

=> S L12 AND (R OR S OR R* OR S* OR STEREOISOMER)

This strategy was designed to retrieve only substances which have a chiral center at phosphorus or arsenic. Substances with both a chiral P or As and a dimethyl-, diethyl-, or diphenyl- substituted P or As will not be in the answer set. An example of the type of answer retrieved is shown below.

INDEX NAME = Nickel(2+), bis[methyl[2-(methylphenylarsino)phenyl]phenylphosphine-As,P]-,
[SP-4-4-[R-(R*,R*)],[S-(R*,R*)]]- (9CI)

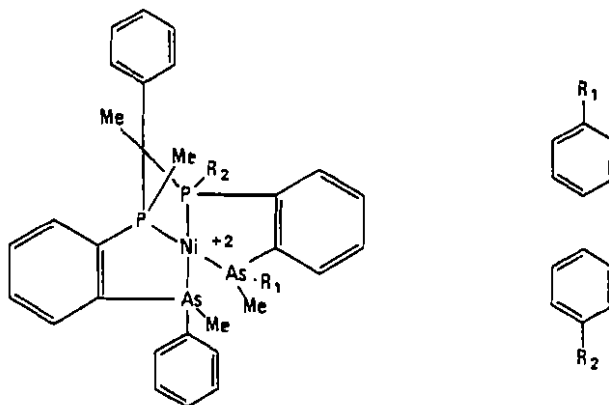
SYNONYM = Phosphine, methyl[2-(methylphenylarsino)phenyl]phenyl-, nickel complex, [R-(R*,R*)]- (9CI)

SYNONYM = Phosphine, methyl[2-(methylphenylarsino)phenyl]phenyl-, nickel complex, [S-(R*,R*)]- (9CI)

MOLECULAR FORMULA = C₄₀H₄₀As₂NiP₂

SUBSTANCE CLASS = CCS, COM

TEXT DESCRIPTOR = 7:SP-4-4.(1:R2:R*,R*),(1:S2:R*,R*)

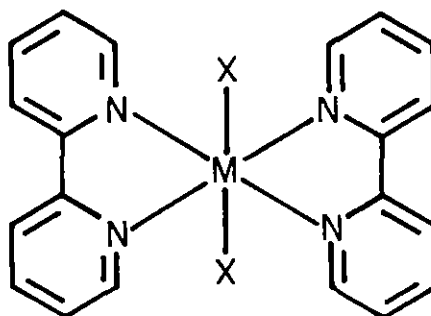


Stereochemistry
Cis or Trans Isomers

Cis or Trans Isomers

Of the compounds retrieved in the following search for dihalobisbipyridine or bisphenanthroline complexes, which are cis and which are trans?

This type of question is both the most difficult to answer of the three posed, and also gives the least precise results. Structure D, shown below, was combined with the screens for the transition metals and searched. The results of this search are in answer set L16.



Structure D

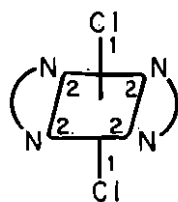
Query: STR D and SCR (1921 or 1931 or 1964)

The structures in the answer set are quite well defined, so we can predict the stereochemical descriptors for most of the answers in which the ligand is unsubstituted or symmetrically substituted.

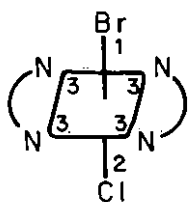
Stereochemistry
Cis or Trans Isomers

Illustrated below are the different isomers possible, together with their stereochemical descriptors. All the halide ions are heavier than the nitrogen donors, so the halides receive the highest priority.

TRANS

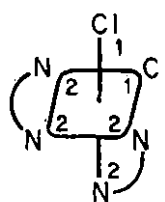


OC-6-12

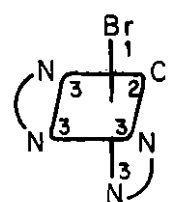


OC-6-23

CIS



OC-6-22



OC-6-33

The following search allows us to the trans isomers out of the answer set L16. The trans isomers are in answer set L17.

=> S L16 and (12 or 23 or TRANS)

Next we sort out the cis isomers. These isomers are in answer set L18.

=> S L16 and (22 or 33 or CIS)

Finally, we create an answer set that contains the isomers which are uncertain.

=> S L16 not L17 not L18

Stereochemistry
Cis or Trans Isomers

The last answer set will have to be examined by hand, but it will be a lot easier than sorting all answers in L16. These remaining answers comprise three categories of substances.

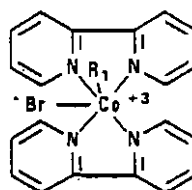
1. Stereochemical descriptor not included in the index name.

INDEX NAME = Cobalt(1+), bis(2,2'-bipyridine)dibromo-, bromide (8CI)

MOLECULAR FORMULA = $C_{20}H_{16}Br_2CoN_4.Br$

SUBSTANCE CLASS = CCS

TEXT DESCRIPTOR = 7:OC-6-22



Br^- R_1

Br^-

Stereochemistry
Cis or Trans Isomers

2. Unsymmetrically substituted bipyridines or phenanthrolines. These will have configuration numbers other than 12, 22, 23, or 33.

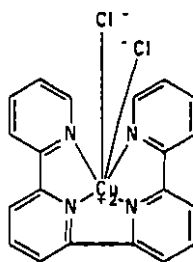
INDEX NAME = Copper, dichloro(2,2':6',2'':6'',2''':6''',2''''-quaterpyridine-*N,N',N'',N'''*)-, (OC-6-13)- (9CI)

SYNONYM = 2,2':6',2'':6'',2''':6''',2''''-Quaterpyridine, copper complex (9CI)

MOLECULAR FORMULA = $C_{20}H_{14}Cl_2CuN_4$

SUBSTANCE CLASS = CCS

TEXT DESCRIPTOR = 7:OC-6-13



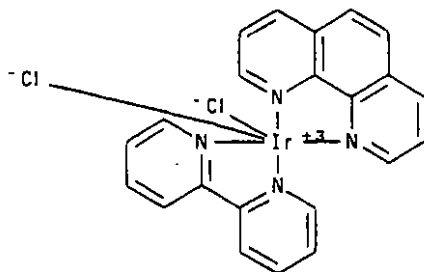
INDEX NAME = Iridium(1+), (2,2'-bipyridine-*N,N'*)dichloro(1,10-phenanthroline-*N1,N10*)-, chloride, (OC-6-32)- (9CI)

DELETED REGISTRY NUMBER(S) = 56928-52-0

MOLECULAR FORMULA = $C_{22}H_{16}Cl_2IrN_4$

SUBSTANCE CLASS = CCS

TEXT DESCRIPTOR = 7:OC-6-32



• Cl⁻

Stereochemistry
Cis or Trans Isomers

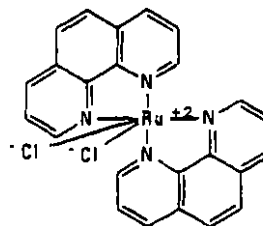
3. Those where no stereochemistry is described. This is the majority.

INDEX NAME = Ruthenium, dichlorobis(1,10-phenanthroline- Λ^1, Λ^{10})- (9CI)

SYNONYM = Dichlorobis(1,10-phenanthroline)ruthenium

MOLECULAR FORMULA = $C_{24}H_{16}Cl_2N_4Ru$

SUBSTANCE CLASS = CCS



Summary

This case study has discussed various ways of searching for specific stereoisomers. Case Study 9 will discuss name and formula searching.

References

1. T. E. Sloan, "Stereochemical Nomenclature and Notation in Inorganic Chemistry" in Topics in Inorganic and Organometallic Stereochemistry, ed. G. Geoffroy, John Wiley & Sons, Inc., New York, 1981, 1-36.
 2. T. E. Sloan and D. H. Busch, "Stereochemical Description and Notation for Coordination Systems" in Stereochemistry of Optically Active Transition Metal Compounds, ed. B. E. Douglas and Y. Saito, American Chemical Society, Washington, D.C., 1980, 397-419. (ACS Symposium Series No. 119.)
 3. M. F. Brown, B. R. Cook, and T. E. Sloan, "Stereochemical Notation in Coordination Chemistry. Mononuclear Complexes", Inorg. Chem. 1975, 14, 1273-8.
 4. M. F. Brown, B. R. Cook, and T. E. Sloan, "Stereochemical Notation in Coordination Chemistry. Mononuclear Complexes of Coordination Numbers Seven, Eight, and Nine", Inorg. Chem. 1978, 17, 1563-8.
-

CASE STUDY 9: NAME AND FORMULA SEARCHING

In the previous case studies we have emphasized how to use substructure searching to obtain Registry Numbers of coordination compounds. In several of the case studies, name and/or formula searching has been used to make answer sets more specific. If your search question pertains to a well defined structure, you can often obtain similar information less expensively by searching by chemical formula or by name in the dictionary portion of the Registry File. The types of information you can search in the dictionary portion of the Registry File are summarized in the next page.

In this case study we present an overview of the Basic Index and the specialized indexes of most interest to inorganic chemists. Full documentation for dictionary term searching is found in Using CAS ONLINE: The Registry File, Volume III. Search examples from previous case studies which lend themselves to dictionary searching have been selected for illustrations, so you may compare the results of name or formula searches with those obtained from substructure searches on the same topic.

Name and Formula Searching
Searchable Fields

SEARCHABLE FIELDS IN THE REGISTRY FILE

Field Code	Field Name	Examples
None (or /BI)	Basic Index (contains name fragments and molecular formulas)	CHLORO AND .MU. CoH18N6
/MF	Complete Molecular Formula in Hill System order (C > H > all other symbols in alphabetical order; if no C present, strictly alphabetic by element symbol)	C6H24CoN6.3Cl/MF ClOH10Fe/MF
/element symbol	Element Count (exact value, range of values, minimum or maximum values)	2-6/Fe Au > 4
/PG	Periodic Group codes for columns and rows of the Periodic Table	B3/PG LNTH/PG
/NC	Total Number of Components in the molecular formula, i.e., number of "dots" + 1 (exact value, range of values, minimum or maximum values)	3/NC 4-6/NC NC > 3 NC < 4
/FW	Formula Weight (exact range, range of values, minimum or maximum values)	195/FW 225-240/FW FW >1000
/CI	Substance Class Identifier (12 are available)	MAN/CI CCS/CI
/CN	Complete Name (CA Index Names as well as synonyms, e.g., trade names and commonly used names)	FERROCENE/CN FERRIC CHLORIDE/CN

Name and Formula Searching
Proximity Operators

Proximity Operators

In addition to the logical operators, AND, NOT, and OR, which merely specify that the search terms are in the record, there are three proximity operators which specify more precise relationships between search terms.

Proximity Operator	Definition	Example
(W)	name fragments must be adjacent and in the order specified	=> S TETRA(W)AMMINE
(nW)	name fragments must have n or fewer intervening terms; name fragments must be in the order specified	=> S BROMO(1W)CHLORO
(A)	name fragments must be adjacent but in either order	=> S PLATINUM(A)AMMINE
(nA)	name fragments must have n or fewer intervening terms; name fragments can be in either order	=> S CHLORO(1A)BROMO
(L)	name fragments must be in the same name (CA Index Name or synonym)	=> S BROMO(L)(PLATINUM OR PLATINATE)

Any one of the above search statements would have retrieved the following substance.

RN 23678-60-6
 IN Platinum(2+), tetraammine-, bromotrichloroplatinate(2-) (8CI)
 SY Platinatate(2-), bromotrichloro-, tetraammineplatinum(2+) (8CI)
 MF Br Cl3 Pt . H12 N4 Pt

Name and Formula Searching
Truncation Symbols

Truncation Symbols

Special characters are available to allow for alternative endings for a search term and for variation of a character within a term. These special characters may be used either with words or with chemical formulas. They are listed in the following table.

Symbol	Definition	Example	Retrieved
?	represents any number of characters, including none may be used only at the end of a term	=> S CHLOR?	chloro chlorine chloride chlorides etc.
#	represents zero or one character more than one # may be used to specify n or fewer characters may be used only at the end of a term	=> S PHOSPHIN# => S COBALT###	phosphine phosphino cobalt cobaltate etc.
!	represents exactly one character more than one ! may be used to specify exactly n characters may be used at the end of a term or within a term	=> S SULF!TE => S SULF!N!!	sulfate sulfite sulfonyl sulfonic sulfinic etc.

Name and Formula Searching
Formulas in the Basic Index

Searching Formulas in the Basic Index

The formulas in the Basic Index include the molecular formulas of all single component substances, e.g., ClOHI0Fe for ferrocene, and the molecular formulas for the individual components of multicomponent substances (such as salts with counter ions), e.g., C4H16Cl2CoN4 and ClO4 for dichlorobis(ethylenediamine)cobalt perchlorate.

In Chapter 3, a FAMILY substructure search was done to locate isomers, salts, and labeled substances containing the Co en2 Cl2 (2+) ion. To find Registry Numbers for this substance, both as a single component and as one of the components of a salt or other multi-component substance, you can search for the molecular formula in the Basic Index.

=> S C4H16CL2CON4

A comparison of the answers from the FAMILY search and the answers from the molecular formula search showed that the answer sets are not identical. The formula search also retrieved salts of two other complexes with the same formula.

RN 92276-05-6
IN Cobalt, dichlorobis(1,1-dimethylhydrazine)- (7Cl)
MF C4 H16 Cl2 Co N4

RN 52571-38-7
IN Cobalt(1+),
[N-(2-aminoethyl)-1,2-ethanediamine-N,N',N'']amminedichloro-,
(0C-6-14)-, perchlorate (9Cl)
MF C4 H16 Cl2 Co N4 . Cl O4

Retrieved by the substructure search, but not by the formula search, was a deprotonated complex which has a different molecular formula.

RN 15318-86-2
IN Cobalt, dichloro(1,2-ethanediaminato-N,N')(1,2-ethanediamine-N,N')-(9Cl)
MF C4 H15 Cl2 Co N4

Name and Formula Searching
Formulas in the Molecular Formula Index

Formula Searching in the Molecular Formula Index (/MF)

Complete molecular formulas may be searched in the /MF field. When you search on a molecular formula in the /MF field, you are restricting your results to just those substances in which the specified molecular formula is the complete molecular formula. No components except the one(s) you specified as part of the complete formula are allowed.

To retrieve a specific salt of an ionic coordination compound, such as [Co en₃] Cl₃, enter the entire formula with periods separating each of the component formulas. The carbon-containing components are listed first, according to decreasing number of carbon atoms. The components without carbon are listed after the carbon-containing components in alphabetical order by the first element in the Hill formula.

=> S C₆H₂₄CoN₆.3Cl/MF

Since molecular formulas are not unique, not all of the answers are for tris(ethylenediamine)cobalt chloride. An example of a false retrieval is shown below.

RN 15613-81-7
IN Cobalt(3+),
diammine[N,N'-bis(2-aminoethyl)-1,2-ethanediamine-N,N',N'',N''']-,
trichloride (9CI)
SY 1,2-Ethanediamine, N,N'-bis(2-aminoethyl)-, cobalt complex (9CI)
SY Diammine(triethylenetetramine)cobalt trichloride (7CI)
MF C₆ H₂₄ Co N₆ . 3Cl

Name and Formula Searching Formulas in the Molecular Formula Index

In Case Study 1, a search for all rhenium halides and halocomplexes was performed. If you needed only a more limited set of substances with more than one rhenium atom and were interested in seeing each ion only once (without any counter ions), you could perform the search in the Molecular Formula field. Character masking (!) is used to allow for variation in the coefficients. CL!RE! allows for the retrieval of CL2RE2 to CL9RE9 and CL1!RE! allows for the retrieval of CL10RE2 to CL19RE9.

=> S CL!RE!/MF OR CL1!RE!/MF OR BR!RE!/MF OR BR1!RE!/MF OR I!RE!/MF OR
I1!RE!/MF OR F!RE!/MF OR F1!RE!/MF OR BR!CL!RE!/MF OR BR1!CL!RE!/MF OR
BR!CL1!RE!/MF OR BR!F!RE!/MF OR BR1!F!RE!/MF OR BR!F1!RE!/MF OR BR!I!RE!/MF OR
BR1!I!RE!/MF OR BR!I1!RE!/MF OR CL!F!RE!/MF OR CL1!F!RE!/MF OR CL!F1!RE!/MF
OR CL!I!RE!/MF OR CL1!I!RE!/MF OR CL!I1!MF OR F!I!RE!/MF OR F1!I!RE!/MF OR
F!I1!RE!/MF

Examples of answers are shown below.

RN 67815-32-1
SY Rhenate(2-), octaiododi-, (Re-Re),
MF 18 Re2

RN 60260-27-7
SY Rhenate(1-), tri-.mu.-chlorohexachlorodi-,
bis(thiohypochlorito)nitrogen(1+) (9Cl)
MF C19 Re2

While it is tedious to type so many formulas, for a search of well defined substances this method is less expensive than a substructure search.

You can also use the technique of internal masking in the Basic Index to retrieve the salts of the dimers and trimers. You can also broaden this technique to retrieve all one-component rhenium chlorides with only a single rhenium atom by using the search term CL!RE.

Name and Formula Searching Element Counts

Searching Element Counts

Element Count is a count of the total number of atoms of each element in the molecular formula. For multi-component substances, element counts are calculated for each of the individual components. Total element counts are not calculated for the entire multi-component substance. The following examples illustrate the individual element counts that are generated from substance molecular formulas:

<u>Complete Formula</u>	<u>Element Counts Generated</u>
C4H16CoCl2N4	4/C, 16/H, 1/Co, 2/Cl, 4/N
C2H8Br2CoCl2N2. Cl4Co	component 1: 2/C, 8/H, 2/Br, 1/Co, 2/Cl, 2/N component 2: 1/Co, 4/Cl

Note that when the molecular formula contains a single occurrence of an element, an element count of 1 is generated for that element.

In addition to generating the element counts for each element in a molecular formula, the following special element counts are generated:

For a molecular formula with no carbon, the element count 0/C (zero carbon atoms) is generated.

For a molecular formula with no hydrogen, the element count 0/H (zero hydrogen atoms) is generated.

For a molecular formula with no oxygen, the element count 0/O (zero oxygen atoms) is generated.

For a molecular formula with no nitrogen, the element count 0/N (zero nitrogen atoms) is generated.

For a molecular formula containing halogens (F, Cl, Br, I, At), the number of halogens is summed and an element count of n/X is generated, where n is the total number of halogen atoms in the formula.

For a molecular formula containing one or more metal atoms, the total number of metal atoms is summed and an element count of n/M is generated, where n is the total number of metal atoms in the formula.

For a molecular formula containing D (deuterium) or T (tritium), a total H count that includes these elements is generated in addition to the individual element counts for each of the isotopes.

Name and Formula Searching
Element Counts

For the molecular formulas listed on the previous page, the following additional element counts are generated.

<u>Complete Formula</u>	<u>Additional Element Counts Generated</u>
C ₄ H ₁₆ CoCl ₂ N ₄	1/M, 2/X, 0/O
C ₂ H ₈ Br ₂ CoCl ₂ N ₂ . Cl ₄ Co	component 1: 1/M, 4/X, 0/O component 2: 1/M, 4/X, 0/H 0/O, 0/N, 0/C

In addition to specifying the exact number of atoms of a given element in a molecular formula, element count can be used to:

- specify an element count range,
- specify a minimum element count,
- specify the presence of an element, or
- specify a maximum element count.

The following table summarizes the search terms used for each of the cases listed above.

Element Count	Search Example
Exactly n atoms	3/Fe (formulas with exactly 3 iron atoms)
m-n atoms allowed	3-7/F (formulas with 3-7 fluorine atoms)
Greater than n atoms allowed *	S > 4 (formulas with more than 4 sulfur atoms)
The element must be present	Mn > 0 (manganese must be present in the formula)
The element must be absent	O < 1 (oxygen must not be present. Only valid for O, N, C, and H)
Fewer than n atoms allowed *	P < 3 (formulas with less than 3 phosphorus atoms)

* The combinations >= (greater than or equal to) and <= (less than or equal to) may also be used.

Name and Formula Searching
Element Counts

In Case Study 7, a search for gold-phosphine clusters was performed and then refined to include only those clusters with five or more gold atoms and five or more phosphorus atoms. This search could also be run entirely in the dictionary portion of the Registry File.

=> S AU>4 AND P>4

The substructure search for P????Au????Au limited to five or more Au and over four P retrieved fewer substances. The higher number of substances retrieved by formula searching includes those that are Manual Registrations and have no structure. An example of this type of answer is shown below.

RN 90857-52-6
IN Gold(3+), octakis[tris(4-methoxyphenyl)phosphine-P]nona-, (20Au-Au),
trinitrate (9CI)
SY Phosphine, tris(4-methoxyphenyl)-, gold complex (9CI)
MF C168 H168 Au9 O24 P8 . 3 N O3

CM 1

RN 82800-52-0
MF C168 H168 Au9 O24 P8
CI CCS, COM, MAN

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***

CM 2

RN 14797-55-8
MF N O3
CI COM

NO3-

Name and Formula Searching
Class Identifiers

Searching with Compound Class Identifier Codes (/CI)

Compound Class Identifier Codes are codes assigned by CAS to describe broad classes of substances. They are searched in the Class Identifier (/CI) field. The following table lists the 12 Compound Class Identifier Codes and their definitions:

Class Code	Definition	Example
AYS	Alloy	AYS/CI
CCS	Coordination Compound	CCS/CI
CTS	Registered Concept	CTS/CI
GRS	Generic Registration	GRS/CI
IDS	Incompletely Defined Substance	IDS/CI
MAN	Manually Registered Substance	MAN/CI
MNS	Mineral	MNS/CI
MXS	Mixture	MXS/CI
PMS	Polymer	PMS/CI
RIS	Radical Ion	RIS/CI
RPS	Ring Parent	RPS/CI
TSCA	Toxic Substance Control Act Inventory (Initial Inventory and Cumulative Supplement II)	TSCA/CI

Compound Class Identifier Codes are generally used in combination with other Registry File search terms, e.g., names, molecular formulas, structure search results, etc. Those codes most likely to be of use in inorganic searching are MAN/CI and AYS/CI. CCS/CI is not recommended for searching for the same reason that use of a screen for this same designation was not recommended: not all compounds which inorganic chemists consider as coordination compounds receive the identifier.

If you wish to locate coordination compounds which have both iron and cobalt in the same fragment, you would not wish to see the substances which are alloys. Alloys can be excluded by the use of AYS/CI, the class identifier code for alloys.

=> S FE>0 AND CO>0 NOT AYS/CI

Name and Formula Searching

Name Searching in the Basic Index

Name Searching in the Basic Index

In the Basic Index you can search with name fragments. This allows you to search for as much or as little of the name as you wish.

In order to give you the capability for searching for parts of chemical names, all 10 million names in the Registry File have been fragmented. All of these fragments are searchable in the Basic Index of the Registry File. The principles of name segmentation are:

- the name is first broken into segments at spaces and punctuation
- these segments are broken into smaller fragments by a computer program that identifies chemically significant fragments

These fragments are the searchable name fragments in the Basic Index of the Registry File. For example, the name

Cobalt(3+), tris(1,2-propanediamine)

is broken into the following searchable fragments:

cobalt	propane
3+	di
tris	amine
1	
2	

Note that partially recombined fragments such as diamine and 1,2 are not searchable name fragments. Only the smallest chemically significant fragments are searchable in the Basic Index. When deciding which name fragments to include in your search strategy, you should use the EXPAND command to verify your selections.

In Case Study 8, we searched for dihalophenanthroline or bipyridine complexes. To use name searching for these compounds, we must decide on which search terms to use. Expanding "phenanthroline", we find many postings for this word, so it is a valid name segment.

Name and Formula Searching
Name Searching in the Basic Index

=> E PHENANTHROLINE 5

E1	376	PHENANTHROLIN/BI	
E2	17	PHENANTHROLINATO/BI	
E3	12757	PHENANTHROLINE/BI	← many postings
E4	78	PHENANTHROLINEDIIUM/BI	
E5	1	PHENANTHROLINEDIOLATO/BI	

Expanding "bipyridine", however, shows no postings, so this word is further segmented into smaller fragments.

=> E BIPYRIDINE 5

E1	1	BIPYRIDINE/BI	
E2	1	BIPYRIDINIUM/BI	
E3	0	BIPYRIDINE/BI	← 0 postings
E4	2	BIPYRIMI/BI	
E5	37	BIPYRIMIDIN/BI	

"Bi" and "Pyridine" are the segment terms as we can verify with EXPAND.

=> E BI 5

E1	1	BHW38/BI	
E2	1	BHX/BI	
E3	315373	BI/BI	← many postings
E4	2	BI~/BI	
E5	2	BI1/BI	

=> E PYRIDINE 5

E1	7	PYRIDINATE/BI	
E2	186	PYRIDINATO/BI	
E3	166040	PYRIDINE/BI	← many postings
E4	1	PYRIDINEAOETONITRILE/BI	
E5	3	PYRIDINECARBOHYDRAZONAMIDE/BI	

Name and Formula Searching
Name Searching in the Basic Index

Name fragments may be combined in a search statement by the (L) operator, assuring that the terms will appear in the same CA name or synonym. Terms that you know were segmented from a larger segment, may be searched with the (W) operator, e.g., bi(w)pyridine. The name fragments of the ligand can be combined with a 1/M term to insure the presence of a metal atom in the complex.

=> S DI(W)(CHLORO OR BROMO OR IODO OR FLUORO)(L)BIS(L)PHENANTHROLINE AND 1/M

=> S DI(W)(CHLORO OR BROMO OR IODO OR FLOURO)(L)BIS(L)BI(W)PYRIDINE AND 1/M

There are many more answers than for the substructure search in the previous case study, and we didn't even search for any complexes with two different halides. The answer sets from name searching contain all metals, not just the transition metals required in the substructure search. We also retrieve many complexes of ligands with dihalo substitution.

RN 92543-37-8
IN Ruthenium(2+),
(4,7-dichloro-1,10-phenanthroline-N1,N10)bis(1,10-phenanthroline-N1,N10)-, (OC-6-22)- (9CI)
SY 1,10-Phenanthroline, 4,7-dichloro-, ruthenium complex (9CI)
MF C36 H22 Cl2 N6 Ru
CI CCS, COM

It is possible to use the (W) operator to keep the terms in the proper order. Ligands are named alphabetically in coordination compounds.

=> S DI(W)(CHLORO OR BROMO OR IODO OR FLUORO)(W)BIS(2W)PHENANTHROLINE AND 1/M

=> S BIS(2W)BI(W)PYRIDINE(2W)DI(W)(CHLORO OR FLUORO OR BROMO OR IODO) AND 1/M

Name and Formula Searching
Name Searching in the Basic Index

This strategy, however, is too restrictive; and, we are losing answers like:

Rhodium(1+), dichlorobis(4,7-diphenyl-1,10-phenanthroline-N1,N10)-

Iridium(1+), bis(2,2'-bipyridine-3,3',4,4',5,5',6,6'-d8-N,N')dichloro-

In addition to the drawbacks illustrated in the examples above, name searching is unable to retrieve any complexes which have ring fusion on the phenanthroline or bipyridine rings, as these ligands have different index names.

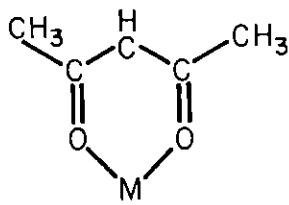
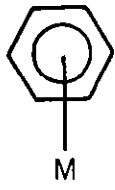

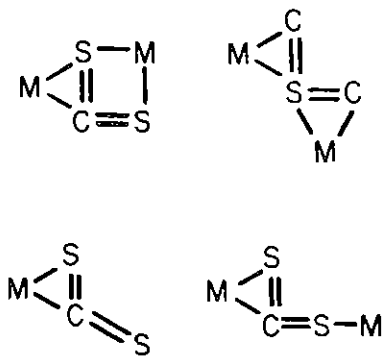
Summary

Searching by name or formula is rarely as precise as substructure searching. It can be used effectively, however, to make a substructure search more precise. Dictionary terms may not be searched directly with structures. You must first perform the substructure search and then you can combine the results of the substructure search with the dictionary terms.

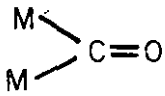
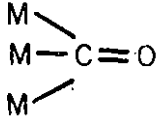
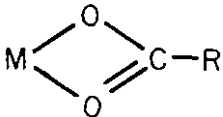


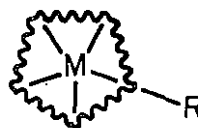
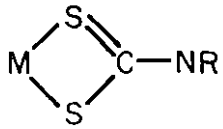
Appendixes

Appendix 1: Structures of Ligands

(All Bonds are Exact Unless Otherwise Noted)

CAS ONLINE Structure			
Name and	Chemist's	Monodentate	Chelated or
Structure			Bridging
Acetylacetone (acac) <div> $\begin{array}{c} \text{O} \quad \quad \text{O} \\ \quad \quad \\ \text{CH}_3\text{C}-\text{CH}-\text{C}-\text{CH}_3 \end{array}$ or $\begin{array}{c} \text{O} \quad \quad \text{O} \\ \quad \quad \\ \text{CH}_3\text{C}-\text{CH}=\text{C}-\text{CH}_3 \end{array}$ </div>			
Allyl - see propenyl			
Azide	N_3^-	$\text{M}-\text{N}=\text{N}=\text{N}$	
Benzene			 C---C normalized
Carbon Disulfide	CS_2	$\text{M}-\text{S}=\text{C}=\text{S}$	

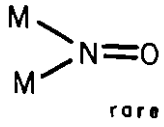
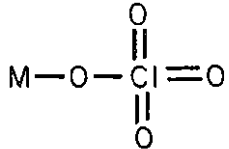
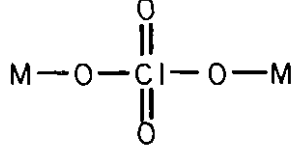
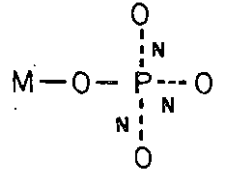
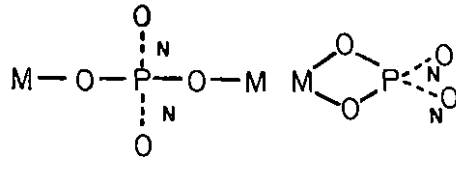
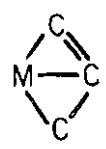
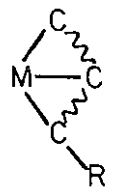

Appendix 1: Structures of Ligands

CAS ONLINE Structure			
Name and Chemist's Structure	Monodentate	Chelated or Bridging	
Carbon Monoxide (carbonyl) CO	$M-C \equiv O$		
Carboxylate COO ⁻	$M-O-C \equiv O$ R		
Cyanide (cyano) CN ⁻	$M-C \equiv N$		
Cyanate (cyanato-N) NCO ⁻	$M-N \equiv C \equiv O$		
Cyclopentadienyl (cp)		 unsubstituted	 substituted C~C unspecified
Dithiocarbamate RNCS ₂ ⁻			

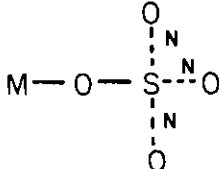
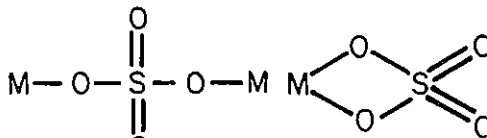
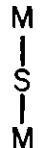
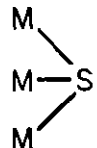
Appendix 1: Structures of Ligands

CAS ONLINE Structure			
Name and	Chemist's	Monodentate	Chelated or
Structure			Bridging
Dithiocarboxylato	RCSS^-	$\text{R}-\text{C} \begin{array}{l} \text{//} \text{S} \\ \text{\textbackslash} \text{S}-\text{M} \end{array}$	$\text{R}-\text{C} \begin{array}{l} \text{//} \text{S} \\ \text{\textbackslash} \text{S} \end{array} \text{M}$
Ethylene (ethene)	C_2H_4		$\text{M} \begin{array}{l} \text{C} \\ \text{//} \\ \text{C} \end{array}$
Nitrogen (dinitrogen)	N_2	$\text{M}-\text{N} \equiv \text{N}$	$\text{M}-\text{N} \equiv \text{N}-\text{M}$ $\text{M} \begin{array}{c} \text{N} \\ \text{ } \\ \text{N} \end{array}$ $\text{M} \begin{array}{c} \text{N} \\ \text{ } \\ \text{N} \end{array} \text{M}$
Nitrate (nitrato)	NO_3^-	$\text{M}-\text{O}-\text{N} \begin{array}{l} \text{=} \text{O} \\ \text{=} \text{O} \end{array}$ customary	$\text{M} \begin{array}{l} \text{O} \\ \text{O} \end{array} \text{N} = \text{O}$ only if author stresses
Nitrite (nitro = nitrito-N)	NO_2^-	$\text{M}-\text{N} \begin{array}{l} \text{=} \text{O} \\ \text{=} \text{O} \end{array}$	
Nitrite (nitrito-O)	ONO^-	$\text{M}-\text{O}-\text{N} = \text{O}$	

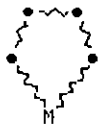
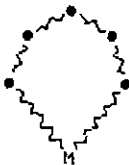
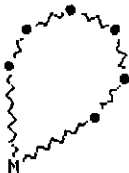
Appendix 1: Structures of Ligands

CAS ONLINE Structure			
Name and	Chemist's	Monodentate	Chelated or
Structure			Bridging
Nitrosyl	NO NO^+	$\text{M}-\text{N}\equiv\text{O}$	 rare
Oxide (oxo, oxy)	O^{-2}	$\text{M}=\text{O}$	$\text{M}-\text{O}-\text{M}$
Perchlorate	ClO_4^-		
Phosphate	PO_4^{-3}		
Propenyl (allyl)	C_3H_6^-	 Unsubstituted	 Substituted  Unspecified

Appendix 1: Structures of Ligands

		CAS ONLINE Structure	
Name and	Chemist's Structure	Monodentate	Chelated or Bridging
Sulfate	SO_4^{-2}		
Sulfide (sulfide, thio, thioxy)	S^{-2}	$\text{M}=\text{S}$	 
Thiocyanate (thiocyanato-N)	NCS^-	$\text{M}-\text{N}=\text{C}=\text{S}$	$\text{M}-\text{N}\equiv\text{C}-\text{S}-\text{M}$
Thiocyanate (thiocyanato-S = isothiocyanato)	SCN^-	$\text{M}-\text{S}-\text{C}\equiv\text{N}$	$\text{M}-\text{S}-\text{C}\equiv\text{N}-\text{M}$

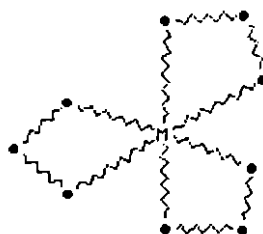
Appendix 2: Fragment File

<u>Code</u>	<u>STRUCTURE</u>
COORD5	
COORD6	
COORD7	

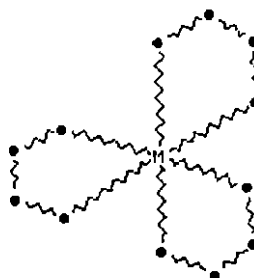
Appendix 2: Fragment File

CodeSTRUCTURE

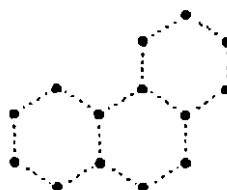
C06LG3



C06LG4



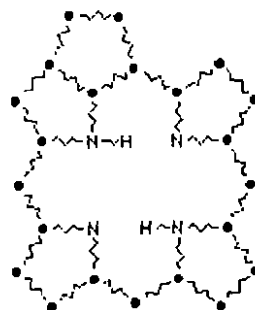
PHENAN



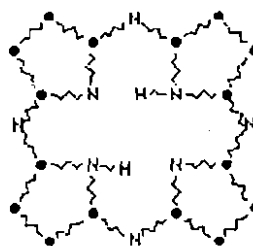
Appendix 2: Fragment File

CodeSTRUCTURE

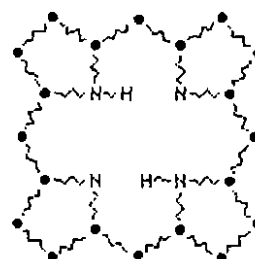
PHORBN



PORPHY



PHORPHN



Appendix 3: Screens
Element Count Screens

Element Count Screen for Metals

<u>Screen</u>	<u>Description</u>	<u>Elements</u>
1918	Metals	all except Ar As At B Br C Cl F H He I Kr N Ne O P Rn S Se Si Te Xe
1967	Group Ia	Cs Fr K Li Na Rb
1936	Group IIa	Ba Be Ca Mg Ra Sr
1988	Group IIIa	La Sc Y
1984	Group IVa	Hf Ti Zr
2001	Group Va	Nb Ta V
1966	Group VIa	Cr Mo W
1991	Group VIIa	Mn Re Tc
1965	Group VIII	Co Fe Ir Ni Os Pd Pt Ra Ru
1920	Group Ib	Ag Au Cu
1956	Group IIb	Cd Hg Zn
1923	Group IIIb	Al B Ga In Tl
1983	Group IVb	Ge Pb Si Sn
1925	Group Vb	As Bi P Sb
2003		O S
2019		Po Se Te
1929	Group VIIb	At Br Cl F I
1924	Group O	Ar He Kr Ne Rn Xe
1964	Transition Series I	Co Cr Cu Fe Mn Ni Ti V
1926		As Ga Ge Kr Rb Se Y Zn

Appendix 3: Screens

Augmented Atom and Hydrogen Augmented Atom Screens

Selected Augmented Atom Screens

Screen Number	Count	Fragment Definition					Freq. %
935		A	*2A				27.31
937		A	*3A				0.05
936		A	*4A				62.10
938		A	-2A				55.74
939		A	-4A				29.01
940		A	* A	* A	* A		37.58
941		A	* A	- A	- A		13.25
942	2	A	* A	- A	- A		3.03
943	3	A	* A	- A	- A		0.95
944		A	* A	* A	* A		68.16
945		A	* A	* A	* A	* A	7.01
946	2	A	* A	* A	* A	* A	1.23
947		A	* A	* A	* A	- A	8.87
948	2	A	* A	* A	* A	- A	4.16
949	3	A	* A	* A	* A	- A	1.17

Selected Hydrogen Augmented Atoms

Screen Number	Count	Fragment Definition		Freq. %
963		AlH2-1C		0.11
989		C H2*1C	*1C	24.89
990	2	C H2*1C	*1C	19.86
991	3	C H2*1C	*1C	14.10
996		C H *1C	*2C	11.82
1006		C H2-1C		27.86
1210		C H2*1C	*1N	11.29
1211	2	C H2*1C	*1N	7.54
1214		C H *1C	*2N	9.35
1214		C H *2C	*1N	9.35
1214		C H *4C	*4N	9.35
1235		C H2-1C	-1N	17.98
1236	2	C H2-1C	-1N	7.75
1398		C H3-1N		10.45
1399		C H3-1N		5.09

Appendix 3: Screens
Type of Ring and Ring Count Screens

Selected Type of Ring Screens

Screen Number	Count	Fragment Definition	Freq. %
1848	2	DDD	1.71
1849		DDD	0.17
1850		DDT	0.26
1851		DTT	2.18
1851		TTT	2.18
1852		DDDD	0.87
1853		DDDT	2.31
1853	2	DDTT	2.31
1854		DDTT	0.03
1853		DTDT	2.31
1853		DTTT	2.31
1853		TTTT	2.31

Ring Count Screens

Screen Number	Fragment Definition	Freq. %
1838	1	81.21
1839	2	62.14
1840	3	39.69
1841	4	23.04
1842	5	11.71
1843	6	6.53
1844	7	3.66
1845	8	2.65
1846	10	1.31
1847	15	0.19

Appendix 3: Screens

Degree of Connectivity Screens

Degree of Connectivity

Screen Number	Count	Fragment Definition	Freq. %
950		5	5.80
2111		3	91.76
2112	3	3	81.24
2113	5	3	62.98
2114	7	3	41.77
2115	9	3	25.18
2116	12	3	11.20
2117	14	3	6.46
2118	16	3	4.14
2119	20	3	2.08
2120		4	42.14
2121	2	4	16.30
2122	3	4	7.73
2123	4	4	3.77
2124	5	4	1.98
2125	6	4	1.15
2126		6	4.71

Appendix 3: Screens
Atom Count Screens

Atom Count		
Screen Number	Fragment Definition	Freq. %
1899	1	95.10
1900	6	91.51
1901	8	89.93
1902	10	87.11
1903	12	82.25
1904	14	75.63
1905	16	67.70
1906	18	58.70
1907	20	50.07
1908	22	41.73
1909	24	34.16
1910	26	27.62
1911	28	22.40
1912	30	17.95
1913	35	10.57
1914	40	6.76
1915	50	3.18
1916	70	1.09
1917	90	0.50

Appendix 3: Screens
Graph Modifier Screens

Graph Modifier Screens

2039	abnormal mass -- all isotopic specifications	1.36%
2045	deuterium isotope	0.71%
2046	tritium and higher H isotopes (e.g., ⁴ H, ⁵ H, etc.)	0.09%
2047	isotope at unknown location (IUL)	0.08%
2041	abnormal valence	31.40%

Multi-component substance screens.

2127	2 or more components	21.07%
2077	3 or more components	4.09%
2078	4 or more components	1.79%
2079	single atom fragment (SAF)	10.46%

Chemical substance class identifier screens. Note that these screens are not mutually exclusive -- a substance may belong to as many classes as are appropriate.

2050	alloy	2.11%
2049	coordination compound	8.97%
2048	incompletely defined (ID) substance	1.80%
2071	ID - unknown structure (ID molform)	0.21%
2072	ID - unknown point of attachment	1.29%
2073	ID - ester	0.16%
2074	ID - hydrogen (bond)	0.13%
2053	manual registration	3.44%
2052	mineral	0.10%
2051	mixture (substance named as "mixt. with" in <u>CA</u>)	0.26%
2043	polymer (general category)	4.30%

Appendix 3: Screens
Graph Modifier Screens

Text descriptor screens indicate the type of stereochemical data provided by a substance's text descriptor. (Note that familiarity with CAS stereochemistry practices is required to use these screens effectively.)

2058	type * -- stereochemistry cannot be expressed by the CAS text descriptor conventions	4.82%
2059	type 1 -- absolute configuration (using R/S to describe a single or reference center stereogenic atom)	2.42%
2060	type 2 -- relative configuration (using cis/trans, E/Z, R*/S*, etc., to describe multiple stereogenic atoms)	7.23%
2061	type 3 -- optical rotation (present only if absolute configuration is not specified)	1.70%
2062	type 4 -- stereoparent descriptor (based on 140+ common names of classes of natural products)	3.73%
2063	type 5 -- amino acid and carbohydrate descriptors	4.55%
2064	type 6 -- trivial name descriptor (based on 550+ common names of natural products having complex stereochemistry)	0.20%
2065	type 7 -- coordination compound descriptor (describes basic geometry and nuclear and ligand stereochemistry)	3.33%
2066	type 8 -- non-steric descriptor (used with incompletely-described substances, polymers, minerals, alloys and radical ions)	4.22%

Index

<u>Subject</u>	<u>Page</u>
A node symbol.....	24
(A) operator.....	173
Acids.....	12,16,103
bonds in.....	107
metal complexes.....	107
metal salts..	109
uncoordinated.....	113
Alloys.....	181
Allyl groups.....	12
Ambidentate ligands.....	7
Atoms, metal.....	15
Attribute commands.....	43
Attributes.....	77
Augmented Atom screens.....	63
Basic Index.....	172
formulas.....	175
names.....	182
BATCH searching.....	74
Benzene.....	12
Bidentate ligands.....	7
Bipyridine.....	48
BOND command.....	2,24
Bond types.....	25
node specification.....	50,52
ring or chain.....	75
Bond values.....	25
Bonds	
coordinate covalent.....	6
covalent.....	6
exact.	11
hapto.....	8
in carboxylates	104
metal-ligand.....	12,75
metal-metal.....	9,16,32,139
normalized.....	11,12,76,93,111
in organometallics.....	93
pi.....	8,12,93
in repeating groups.....	141
ring or chain.....	31,65,75
screens for.....	65
sigma.....	93
single exact.....	12,16,27,30,32,75,93
triple.....	12
Bridging carbonyls.....	147
Bridging halides.....	84
Bridging ligands.....	10

Index

CA File, searching for unspecified derivatives.....	117
CONNECT command.....	43,45,77,79,81,97,126
for isolating rings.....	48
for Metallocenes.....	47
for specific coordination numbers.....	46
ligands.....	48
metal hydrides.....	49
Carbide.....	93
Carbon Monoxide.....	93
Carbonyls	
bridging.....	147
terminal.....	142
Carboxylates.....	16,103
bidentate.....	116
bonds in.....	104
uncoordinated.....	113
Carboxylic Acids.....	103
bonds in.....	104,107
metal complexes.....	107
metal salts.....	109
Central Metal Atom.....	15
Charges.....	8,14,15,77
Chelate Rings.....	7,17,28,31,75,76,103,111
variable backbone.....	41,123
variable donors.....	41
Chelates.....	123
nitrogen.....	111
oxygen.....	111
tridentate.....	123
Chirality label.....	156
Class identifiers.....	181
Clusters.....	9,32
gold.....	139
iron.....	142
iron carbonyl.....	142
metal.....	139
specific size.....	150
Cobalt complexes.....	79
Commands.....	2
Complete Name Index.....	172
Complex ions.....	6
Complexes	
cobalt.....	79
cyclopentadienyl.....	94
low coordination numbers.....	81
octahedral.....	79
polynuclear.....	9
Components.....	12,14,16,21,109
Configuration number.....	156,158

Connection Table.....	32,34,35,36
Connection Table Hydrogens.....	13,24,133
Coordination Numbers.....	6,9,17,,46,76,77,79
Coordination Sphere.....	6,75,79
Coordination Theory.....	5
Counter Ions.....	14,17,19
Cyanate.....	93
Cyclopentadienyl.....	15,33,47,72,80,94
Cyclopentadienyl, substituted.....	12,96
Dictionary searching.....	171
to narrow results.....	140
for stereochemistry.....	160
Dividing searches.....	74,86
Donor atoms.....	6
nitrogen.....	127
sulfur.....	127,129
variable.....	41
Donor sites, varying numbers.....	129
Element Count Index.....	172
Element count searching.....	178
Elements counts to narrow results.....	140
Ethylene.....	98
Ethylene groups.....	12
Ethylenediamine.....	21
EXPAND command.....	183
Field codes.....	172
Field names.....	172
File crossover.....	117
Formula searching.....	171,175
Formula Weight Index.....	172
Fragment File, as models.....	38
Fragments.....	182
Fragments, recombined.....	182
Free ions.....	21
Full file projections.....	62
Full file searching.....	61,78
Fused rings.....	30
GRAPH command.....	2,23
Generic atoms.....	41
Generic registrations.....	181
Geometric isomers.....	19
Gk Nodes.....	24,41,76,77,82,124,141
in metal hydrides.....	137
Gold clusters.....	139

Index

HCOUNT command.....	43,44,134
HELP command.....	3
Halides	
bridging.....	84
cobalt complexes.....	79
different.....	82
rhenuim complexes.....	88
Halo complexes	
cobalt.....	79
rhenium.....	88
Halogens.....	24
Hapto bonding.....	8
Homopolymers.....	21
Hydride ions.....	13
Hydrides.....	13,44,49,77,133
bridging.....	136
Hydrogen augmented atom screens.....	69
Hydrogen count.....	43,44,77,97,133
Hydrogens.....	12,24,133
connection table.....	13
nonconnection table.....	13
number of.....	44
Incomplete search projection.....	61
Incomplete structures.....	17
Incompletely defined substances.....	181
Indexing policies.....	11,15
Ionized ligands.....	21
Ions.....	14
counter.....	17,19
free.....	21
hydride.....	13
monoatomic.....	12
nitrate.....	17
oxy.....	17
perchlorate.....	17
sulfate.....	17
Iron carbonyl clusters.....	142
Isomers	
geometric.....	19
optical.....	20
stereo.....	19,20
Isotopes.....	20
Iteration process.....	61
(L) operator.....	173
Labeled compounds.....	20
Lewis Base.....	6

Ligands.....	6
acids.....	103
allyl.....	12
ambidentate.....	7
atoms in.....	42
benzene.....	12
bidentate.....	7,17
bipyridine.....	48
blocked nitrogen coordination.....	115
bridging.....	10,11
carbide.....	93
carbon monoxide.....	93
carboxylates.....	103,107
chelates.....	123
chlorides.....	79
cyanate.....	93
cyclopentadienyl.....	12,15,33,47,72,80
deprotonated.....	21
ethylene.....	12,98
fused rings.....	30
halides.....	79,82
hydride.....	133
ionized.....	21
macrocyclic.....	7,34,123
monodentate.....	6,17,31,75
multidentate.....	7
negative charged.....	14
nitride.....	12,75
organometallics.....	72
oxide.....	12
oxo.....	75
pentadiene.....	99
phenanthroline.....	48
phosphide.....	12
phosphido.....	75
phosphine.....	134,139
pi-bonding.....	12,72
pi-donor.....	8,12,72
picolines.....	103,107,117
polydentate.....	7,123
porphyrins.....	38
1,3-propanediamine.....	69
stereochemistry.....	156,162
structuring conventions for.....	11
substituted cyclopentadienyl.....	96
sulfide.....	12,75
tetradentate.....	7,127
tridentate.....	7,123
unidentate.....	6
variable groups.....	41
M node symbol.....	24

Index

Macrocycles.....	34,123
tetradentate.....	127
Manual registration.....	120,141,181
Manual Registry Numbers.....	120,141,181
Manually registered substances.....	120,141,181
Metal, definition of.....	24
Metal atom.....	15,24,42
Metal clusters.....	9,32,139
Metal complexes.....	6
Metal hydrides.....	49,133
Metal-ligand bonds.....	75
Metal-metal bonds.....	16,32,139
Metal salts.....	16,109
Metal screens.....	64
Metallocenes.....	8,47,94
Minerals.....	181
Mixtures.....	181
Models, for structuring building.....	37
Molecular Formula Index.....	172,176
Molecular Formulas.....	175
Monatomic Ions.....	12
Monodentate ligand.....	6
Multidentate ligand.....	7
Name fragments.....	182
Name Searching.....	171
in the Basic Index.....	182
in the CA File.....	119
in the Registry File.....	120
Name segments.....	182
Nitrate ions.....	17
NODE command.....	2,23
Node Specification.....	43,50,75,107
and bond type.....	50,52
defaults.....	51
for hydrides.....	136
Nonconnection table hydrogens.....	13,133
Noncyclic polyenes.....	97
NSPEC command.....	43,50
Number of Components Index.....	172
Octrahedral complexes.....	79
Optical isomers.....	20
Organometallics.....	8,47,72,80,93
Oxidation states.....	8,20
Oxide.....	12
Oxy Ions.....	17
Oxyanions.....	11
Pentadiene.....	99
Perchlorate ions.....	17

Periodic Group Index.....	172
Phenanthroline.....	48
Phosphide.....	12
Phosphine.....	139
Pi-Bonds.....	12,93
Picolinates.....	103,117
chelates.....	111,113
monodentate.....	113
unspecified derivatives.....	117
Pi-donor ligands.....	72
Pi-donors.....	12,93,97
Pi-systems, partially coordinated.....	100
Polydentate ligands.....	7
Polydentates.....	123
Polyenes, noncyclic.....	97
Polymers.....	21,181
Polynuclear complexes.....	9
Porphyrins.....	38
1,3-Propanediamine.....	69
Pyridine, uncoordinated.....	115
2-Pyridinecarboxylate.....	103
Protons.....	12,133
Proximity operators.....	173
Q node symbol.....	24
RECALL command.....	37
Fragment File.....	38
Registry Number.....	37
REPEATING command.....	41,123,141
RSPEC command.....	43,44
Radical ions.....	181
Range searching.....	119
Registered concepts.....	181
Registration policies.....	11,15,18,119
Registry Numbers.....	18
as Models.....	37
manual.....	120
Ring parents.....	181
Ring specification.....	43,44

Index

Rings	
blocking fusion.....	43,44,48,76
building from Chains.....	35
chelates.....	28,41,75,123
fused.....	30
isolated.....	44,76
isolating.....	43,48
macrocyclic.....	123
polydentate.....	123
spiro.....	28
Salts.....	104,109
formula searching.....	176
metal.....	16
SAMPLE searches.....	62,78
SCREEN command.....	62
Screens.....	61,78
frequency of occurrence.....	63
for hydrogens on ligands.....	69
for metals.....	63,64
NOT to use.....	73
for ring or chain nonds.....	65
type of ring.....	71,93
SCT.....	34,35,36
SEARCH command.....	3
Search Process.....	61
Search Projections.....	61,62,78
Search strategy.....	75
Search strategy refinement.....	43,61
Searchable fields in Registry File.....	172
Segmentation.....	182
SET command.....	32
Spiro ring fusion.....	76
Spiro rings.....	28
Stereochemical descriptors.....	155,167
Stereochemistry.....	155
cis.....	165
geometry.....	159
ligand.....	162
screens.....	158
trans.....	165
Stereoisomers.....	19,20
Structure attributes.....	43
Structure building.....	23

STRUCTURE command.....	2,26
Structure connection table.....	34,35,36
Structure refinement.....	43
Structures	
6-Coordinate.....	26
needing screens.....	63
variability in.....	41
Structuring conventions.....	11
Substance Class Identifier Index.....	172
Substructure searching.....	1
Sulfate ions.....	17
Sulfide.....	12
Symmetry site terms.....	157
System limits.....	62
TSCA.....	74,181
Tautomers.....	76
Terminal Carbonyls.....	142
Tetradentates.....	127
Thiocarboxylates.....	16
Tridentate ligands.....	7
Tridentates.....	123
Truncation symbols.....	174
Type of ring screens.....	71
Uncertain search projection.....	61
Unidentate ligand.....	6
Unspecified derivatives.....	117
Using A atom.....	76
VARIABLE command.....	41,82,123
Variability in structures.....	41
Variable node symbols.....	24
(W) operator.....	173
Werner, Alfred.....	5
X node symbol.....	24