Abstract

The primary goal of this article is to assist researchers and librarians to accurately and completely search Chemical Abstracts Service's (CAS) SciFinder web-based system for coordination compounds including organometallics. CAS indexing policies and conventions are described. Appropriate search strategies are explained. Although this article is of primary value for SciFinder researchers, the analysis of how CAS handles this class of compounds will be of use to searchers regardless of the platform used to access Chemical Abstracts information.

Introduction & Literature Review

This article focuses on searching coordination compounds including organometallics in the Chemical Abstracts REGISTRY substance database via SciFinder, a user-friendly web-based platform, as of September 2011. This paper follows up an article on searching SciFinder for all inorganic substances other than coordination compounds (Wagner 2011). Coordination compounds play key roles in catalysis, biochemical reactions, and the synthesis of organic chemicals including pharmaceuticals and agricultural chemicals, to name just a few highlights of their utility. Chemical Abstracts Service (CAS) has identified and registered over 2.15 million unique coordination compounds as of this writing.

A basic familiarity with the SciFinder system is assumed. Chemical Abstracts Service offers numerous e-seminars, interactive tutorials, and quick reference documents via its CAS Learning Solutions center (free registration required for access). Other good overviews of SciFinder are available (Haldeman et al. 2005; Ridley 2009; Wagner 2006).

Little has been written about searching SciFinder specifically for coordination compounds. REGISTRY database system conventions for inorganics are briefly described in a few articles, but either predate SciFinder or say little about how to actually search for these compounds (Cooke and Ridley 2004; Moulton 1993; Ryan and Stobaugh 1982). Only the recent book by
Ridley (2009) provides both an excellent overview of the entire system and also discusses how to search for all types of inorganic compounds in Appendix 4 of the book including metal complexes in Appendix 4.3.

Also of special interest is an old but extensive CAS training manual on searching coordination compounds still available on the web (Kozlowski 1986). Although the examples go all the way back to the command line-based structure drawing and non-graphical display of the original CAS Online System, Chapter 2 of this manual provides an excellent introduction to coordination compound terminology and characteristics while Chapter 3 details CAS structure conventions and registration policies for this class of compounds. The most complete documentation of the Registry System occurs in old print STN International REGISTRY File documentation that is now hard to find (STN International 1990; 1991; 1993).

Some of this discussion also will be applicable to searchers still using other platforms such as STN on the Web and STN Easy. For STN users, detailed discussion of dictionary (non-structural) and structure searching can be found in various REGISTRY File online and print documentation (Chemical Abstracts Service 2011). Since the SciFinder substance file and REGISTRY are based on the same underlying data, the explanations and strategies described in this article will be of value to the STN searcher.

Note: In this article, CAS Registry Numbers, the unique number assigned by CAS to every substance entered into their database, are enclosed in brackets, e.g. [28966-86-1].

**Definitions**

The following definitions indicate some of the terms used to describe various subclasses of coordination compounds. Given this variety of terminology, patrons may not use the term "coordination compound" when making their request for search assistance.

**Coordination Compounds** are molecules or ions in which a central atom has atoms or molecules (ligands) attached to it, and the number of bonds to the central atom (its coordination number) is not equal to the valence. Coordination compounds may be charged or uncharged. If charged, they are often referred to as complex ions.

The central atom may be any element, but it is usually a metal atom. Every central metal atom has a charge, also known as the oxidation state. The charge is usually zero or positive, but it can be negative. Coordination numbers can range from 2 to 12, and they usually exceed its oxidation state. Coordination compounds are often called metal complexes or simply complexes (Kozlowski 1986 p.5-6, 8).

**Organometallics** are a special category of coordination compounds in which one or more carbon atoms of an organic molecule or atom are directly attached to the central metal atom; i.e., there is a carbon-metal direct bond. However, if the only carbon-containing species present are carbon monoxide, carbonyl sulfide, or cyanide ions (inorganic species), the complex is not classified as an organometallic (Kozlowski 1986 p.8-10).

**Polynuclear Complexes** are coordination compounds with multiple central metal atoms. If a polynuclear complex contains any direct metal-metal bonds, they may also be called metal clusters. Clusters containing up to at least 22 metal atoms are known. This term should not be confused with the term homogeneous metal clusters, which contain only a single metallic element and hence are not coordination compounds (Kozlowski 1986 p.9).
SciFinder has four main options for retrieving substances under the *Explore Substances* main screen (Figure 1):

- **Chemical Structure** permits drawing of structures with exact, substructure, and similarity search options.
- **Markush** structure searching is beyond the scope of this article.
- **Molecular Formula** searches molecular formulas, usually requiring an exact match.
- **Substance Identifier** searches CAS Index Names, synonyms including trade names, and CAS Registry Numbers.

![Explore Substances query screen](image)

Because coordination compound names and molecular formulas are typically complex, this article focuses on structure searching. However, should a searcher have a reasonably simple name or a CAS Registry Number from another source such as a journal article, chemical supply catalog, or website such as Common Chemistry, this information can readily be input into the **Substance Identifier** query screen to rapidly retrieve the desired compound.

Likewise, one can also search molecular formula (MF). Be aware that some coordination compounds are considered by CAS to be multicomponent substances. Hence, they are assigned a separate MF segment for each component that is then strung together into an overall MF formally known as a **dot disconnect formula**. Periods are used between each component. For example, tris(2,2'-bipyridine)iron(2+) bis(tetrafluoroborate) [28966-86-1] is registered as a two-component substance; the tetrafluoroborate ions [BF(1-)] being the second component. Hence, it is assigned a molecular formula of C30H24FeN6.2BF4.

This compound illustrates the pitfalls of name and molecular searching. Since a **Substance Identifier** search generally requires an exact match, it would be a challenge to type in the long chemical name letter perfect. Dot disconnect formulas are discussed in detail in the predecessor

http://www.istl.org/11-fall/tips.html
Searching Coordination and Organometallic Compounds in SciFinder

article (Wagner 2011). A multicomponent coordination compound will be shown in examples that follow. Unless the searcher has a CAS Registry Number or simple name in hand, generally a structure search is the safest and most efficient option.

**Structure Drawing Basics**

The *Explore Substances* query screen contains a small drawing pane with the annotation "Click to Edit." Doing so brings up the Structure Editor (Figure 2). A detailed tutorial in drawing structures is beyond the scope of this paper and best done via web training. For anyone not familiar with the structure drawing features, the CAS Learning Solutions tutorials will provide the necessary assistance.

![Structure Editor](http://www.istl.org/11-fall/tips.html[12/6/2012 7:41:34 PM])

**Figure 2:** Screen shot of *Structure Editor* drawing window.

Note that along the bottom of the pane are options for selecting atoms and bonds. The icons on the left-hand side are various drawing tools including shortcuts for functional groups, charge assignments, and variable atoms. Helpfully, when the mouse cursor stays over an icon, a brief text note appears that identifies the function of the icon. For example, if one hovers over the 3rd icon down in the right-hand column (=$R$), the text line "Define R-groups" pops up.

It is important to understand CAS conventions when drawing coordination compound structures. Again, Kozlowski provides a helpful, detailed description of these conventions.
summarized in the bullet points below (Kozlowski 1986 p. 11-22). However, because all rules are subject to interpretation (or misinterpretation!), a searcher is strongly encouraged to first search for a simple, well-known analog; i.e., a similar type of material, to the substances one is looking for. Then note how this type of substance is drawn and test queries to make sure one can retrieve the simple analog before conducting a search for the desired structures. When in doubt about the exact value of bonds (single, double, normalized, triple), it seldom hurts to choose the "unspecified" bond value. If one retrieves too many hits, one can browse the results, determine the correct bond value, and modify the query structure.

- Metal-to-metal bonds are always drawn as single exact bonds.

- Although ligands are usually drawn in exactly the same way they would be were they not bonded to the central metal atom(s), in certain cases, normalized bonds in ligands may be changed to an exact bond.

- Metal-ligand bonds are usually single exact bonds. The exceptions are the following monoatomic ions, provided they are bonded to only a single metal atom:
  - Double bonds are used for O(2-) & S(2-).
  - Triple bonds are used for N(3-) & P(3-).

- Pi-donors, such as cyclopentadiene or benzene, are drawn with direct bonds from each carbon atom in the ring to the central metal atom.

- Hydrogen atoms should be included in the query structure drawing if the hydrogen is bonded to more than one atom, has an abnormal valence, or has a charge, e.g., the hydride ion (H1-).

- Simple non-coordinated ionic species accompanying the complex (counterions) are treated as separate components from the metal complex structure. Oxy ions such as nitrate, perchlorate, and sulfate are structured as counterions unless the author of the paper being indexed explicitly cites their coordination to the metal atom. Note that in a given structure, the same ionic species may be present both as a counterion and as a coordinated ion bonded to the central metal ion.

- Assignment of a charge to the central metal ion follows a set of somewhat complex rules. Unless one is familiar with how CAS registers the types of coordination compounds of interest, it is generally better to leave the metal atom(s) uncharged. That way the results will contain both uncharged and all charged species.

Note that CAS assigns a separate Registry Number to virtually every possible variation of a molecule. Differences in charges, number and types of counterions, stereochemistry, oxidation states, and isotopes are all assigned separate Registry Numbers (Kozlowski 1986 p.18-22). In general, it is good practice to run searches using a base structure without specifying stereochemistry, precise bonding, counterions, and charges. Then one can browse all the variants in a single set of results and determine exactly how CAS has treated this type of compound. If the retrieval set is too large or contains substances of no interest, then either more specification can be added to the query structure or various Analyze/Refine options can be used to limit retrieval.

What happens from this point forward is best shown by the examples in the next section.

**Structure Searching**

When one is done drawing the structure, one chooses one of three search options:
• **Exact search** -- no additional atoms are allowed except for hydrogen atoms to fill valences.

• **Substructure search** -- any sites not specifically blocked may have additional attachments.

• **Similarity search** -- uses the Tanimoto algorithm ([Willett et al. 1998](https://www.istl.org/11-fall/tips.html)) to compare all substances in the database with your query structure, and then determines which are the most similar. Different, but related, atoms and groups may appear in place of the specific atoms and groups drawn in the original structure. For example, if the query substructure contains a chlorine atom at a given position, substances with other halogens (F, Br, I, etc.) would be retrieved and ranked highly via the similarity algorithm.

Then click on the **OK** button to get back to the main *Explore Substances* screen to choose additional search options. One will likely receive a notice that the structure "Exceeds standard valency," since by definition, coordination compounds have a central atom where its valency does not match standard oxidation state. Simply click the **OK** button.

Figure 3 shows this screen with a structure that has already been drawn in and options normally selected for a coordination compound search.

![Abridged Explore Substances Screen: Ready to perform search](https://www.istl.org/11-fall/tips.html)

Figure 3: Abridged *Explore Substances* Screen: Ready to perform search

A 12-membered ring with four evenly spaced nitrogen atoms each connected to the central iron (Fe) atom has been drawn in the *Structure Drawing* window. Note that the Fe-N bonds are dotted, denoting an unspecified bond value. This assures retrieval of substances that conform to a different CAS bonding convention than the searcher expects.
Before hitting the Search button, make sure the desired Search type radio button is selected. In general, it is highly recommended that the Show precision analysis box is always checked. This will permit the searcher to choose the exact level of specificity in the match between the query structure drawing and the results, as we will see in a moment. Naturally, the Coordination Compounds box under Class(es) grouping should be checked whenever only coordination compounds are desired. Unless one is absolutely certain that one wants only single component answers (no counterions or any other associated species not directly bonded to the structure drawn), the single component box should not be checked.

After clicking the Search button, if the Show precision analysis box has been checked, a Precision Candidates pop-up window will appear (Figure 4).

![Precision Candidates Pop-up Window](image)

Figure 4: Abridged Precision Candidates Pop-up Window.

Generally one should select only the first option: Conventional Substructure if a substructure search is being performed or, alternately, Conventional Exact if an exact search is being performed. Choosing any of the other options will produce results that can be quite different from the drawn structure. In particular, bonds drawn in the query structure between metal atoms and heteroatoms like oxygen, sulfur, and nitrogen may not exist in the answer set structures.

At the time this particular substructure search was run, 85 Conventional Substructure results were retrieved. One of the results (Figure 5) illustrates the point that a given ionic species, in this case the chloride ion, can be both a counterion (separate component) and directly bonded in the metal complex structure.
Figure 5: Sample Result from the Search Query Shown in Figure 3.

The CAS Registry Number is the number with two hyphens directly below the text "Substance Detail." The Component Registry Number given after the main CAS Registry Number is the number for the iron-nitrogen structure without the chloride counterion.

Great care must be taken when drawing structures and choosing search options to account for the great diversity of coordination compounds and proper CAS conventions so that relevant answers will not be missed. Novice searchers often have a tendency to overspecify when drawing structures to be used in a substructure search. Draw only the essential aspects of the structure. If one is uncertain about or wants to see all possibilities for a given feature, then leave it unspecified whether the feature is a charge, stereochemistry, bond value, or attachments at a particular position.

As demonstrated in this example, once a proper structure has been drawn, the search process is quite straightforward other than remembering to check the Show Precision Analysis box. The importance of this check box and the meaning of the choices in the resulting Precision Candidates pop-up window are not apparent to the novice searcher. An additional explanation of this feature is provided below.

Our second example is based on ferrocene, a pi-bonded coordination compound with a central iron atom bonded to two 5-membered carbon rings. Although the ferrocene structure can be drawn from scratch, it is far quicker to retrieve the substance record by typing in "ferrocene" into the Substance Identifier query screen. Then simply hover the mouse over the retrieved structure, then click on the double chevron [>>] that appears in the upper right-hand corner of the structure (See highlighted structure in Figure 8). Choose the "Explore by Structure: Chemical Structure" option. This will insert the ferrocene structure in the Explore Substances
query screen (Figure 6).

One can then modify this structure by opening the Structure Drawing window. For this example, we have added a carbonyl group (-C=O) to one of the negatively charged carbons in the ring. We used the Lock Ring fusion or formation icon (shown as bold in the structure in Figure 7) to assure that the carbonyl group is not part of a ring.
At the time this search was run, it retrieved 9,849 compounds using the Conventional Substructure option in the Precision Candidates window. The next section will discuss refinement and use of this answer set.

**Working with Answer Sets: The Basics**

Although the emphasis of this article is on searching coordination compounds, a brief overview of the many things one can do with an answer set will be reviewed. The 9,849 ferrocene answer set will be used as our sample set.

Results can be sorted by CAS Registry Number, number of references, molecular weight, or molecular formula. Sorting by number of references shows that the most common compound is formylferrocene (Figure 8). Clicking on the Substance Detail link displays the full record including experimental and calculated properties, though most coordination compounds do not have calculated properties.
Figure 8: Formylferrocene -- Brief Substance Display showing Analysis & Refine tabs (Refine tab is active).

The set can be narrowed by using the Refine tab in the upper right-hand corner. Options are to limit the set to substances containing isotopes or metals, commercially available, any properties available, specific property values including range searching, or having at least one literature reference or having no references. Two other options are especially powerful:

- **Chemical Structure** -- a second structure can be drawn to further limit the set to answers containing that additional substructure.
- **Atom Attachment** -- any atom (position) in the original structure drawing can be specified and a pick list of all atoms at that attachment point is produced. Figure 9 shows the top 6 atoms attached to the carbon atom marked with an arrow in Figure 7 ((formylferrocene structure).
The *Analysis* Tab analyzes any set of substances by the following characteristics and generates a pick list:

- **Substance Role** (in literature references) -- indicates availability of references dealing with major aspects of the substance such as preparation, properties, uses, and biological studies.
- **Commercial Availability** -- shows how many substances are available commercially. Only 50 out of the 9,849 formylferrocene compounds are commercially available.
- **Elements** -- shows how many substances have a given element anywhere in the structure. This is an especially powerful research tool. Analyzing the formylferrocene results shows that the most common element (other than C, Fe, O, and H inherent in the query structure) is nitrogen, which is contained in 5,025 compounds out of the 9,849 substances retrieved.
- **Reaction Availability** -- indicates which substances are referenced in a reaction.

Finally, one can retrieve literature references (*Get References* button), reactions (*Get Reactions* button), or suppliers (*Tools: Commercial Sources*) for any individual compound, selected set (using check boxes) or the entire retrieved set. Regulatory information (link underneath the structure in Figure 8) can be displayed one compound at a time.

### Searching for Multicomponent Coordination Compounds

There may be times when one wishes to search for coordination compounds that have multiple components; i.e., having more than one structure drawing that are not connected to each other and a corresponding dot disconnect formula. As noted in Section IV, counterions are always treated as separate components. Assuming the *Single component* box is not checked on the *Explore Substances* query screen, any substructure search will automatically retrieve multicomponent substances that have additional structural components separate from the target structure.

The key point is that, if one wishes to search for counterions or other structural features that are considered to be separate components by CAS conventions, they must be drawn as separate,
isolated structures (fragments) in the drawing pane. Unfortunately, there is no way to limit SciFinder structure/substructure searches to only multicomponent substances. Hence, answer sets will always contain compounds where the specified structural fragments are either in different places within the same component or are in separate components entirely, assuming such compounds are known. Only via molecular formula searching can one search explicitly for a multicomponent substance, but of course, that approach eliminates the power of substructure searching.

More Details: Show Precision Analysis

If the searcher leaves the Coordination Compounds class option unchecked while checking the Show Precision Analysis box when performing a substructure search, the Closely Associated Tautomers and Zwitterions set may well contain organic salts of the metal ion. Following CAS conventions for ionic salts, the metal ion is considered a separate component. For example, if one performs a substructure search not limited to coordination compounds on a zinc benzoate [553-72-0] structure and bonds the zinc to the oxygen (Zn-O-C(=O)-Phenyl), the Precision Candidates window (shown in Figure 4) gave the following results at the time this paper was written:

- **Conventional Substructure**: 3,200 substances with 3,082 of them being coordination compounds. The remaining were substances where zinc was directly bonded to the oxygen atom, but did not meet the definition of a coordination compound.
- **Closely Associated Tautomers and Zwitterions**: 6,975 substances with 6,082 of them being coordination compounds. The remaining included salts like zinc benzoate where CAS conventions treat the zinc ion as a separate component.
- **The Loosely Associated Tautomers and Zwitterions** (741 compounds) and **Other** (7 compounds): Mostly coordination compounds containing variations in bonding and attachments compared to the query structure to a greater degree than those compounds in the Closely Associated set.

One of the advantages of a flat-rate system like SciFinder is that the searcher can always experiment by examining each option one at a time to determine exactly what types of structures are being retrieved. If a particular query does not produce any results for the first category, it would be useful to examine results in the other categories. However, usually that first "Conventional" option gives the best results with answers matching the query (sub)structure exactly as drawn. If the searcher is uncertain that only coordination compounds are desired in the results, the Coordination Compounds box should be left unchecked and the various Precision Candidates sets should be reviewed to assure comprehensive retrieval.

Conclusion

SciFinder provides name, molecular formula, and structure searching for over 2.15 million coordination compounds as of September 2011. This article has reviewed the structure drawing conventions and search options that assure a comprehensive search of coordination compounds including organometallics in the CAS substance database via SciFinder. Once results have been retrieved, one is a click or two away from literature references, reactions, properties, spectra, supplier, and regulatory information.

Literature/References

Chemical Abstracts Service. 2011. STN user documentation [Internet]. Columbus (OH): The Service; [cited 2011 Sep 7]. Available from:


