DERWENT

WORLD PATENTS INDEX

CPI Chemical Indexing User Guide

Prepared by Paul Bryan in association with Derwent Information

Revised Edition 1
ISBN: 0 901157 73 2

THOMSON

Published by Thomson Scientific
14 Great Queen Street, London WC2B 5DF,
United Kingdom

Visit the Thomson Scientific web site at www.thomson.com/scientific

First edition published March 1998
Revised first edition published November 2000

ISBN: 0 901156 60 9 (Edition 1)
ISBN: 1 903836 37 4 (Revised Edition 1)

Contents

1 Introduction to the User Guide

1 Introduction to the user guide

About This User Guide

The **Chemical Indexing User Guide** describes the Chemical Fragmentation Coding System, a structural indexing language used for indexing and retrieving chemical patents in Derwent World Patents Index (DWPI). The indexing and retrieval of chemical patents is a challenge because chemical patents often cover **far too many chemicals to index individually**. Such broad patent coverage is achieved by patentees through the use of special **Markush** chemical structures with many variable fragments. Derwent's answer to this problem was the development of a chemical coding system, a system that enables **comprehensive indexing and retrieval** of the broad range of structures disclosed in chemical patents. The use of Chemical Codes has made DWPI one of the most comprehensively indexed chemical patent databases in the world since its inception in 1963.

Although understanding the meanings and uses of the Chemical Codes requires study and practice, searching with the Chemical Codes does not have to be complicated. The **Markush TOPFRAG** software automatically converts a variable structure drawn by the user into a correct Chemical Code search strategy in seconds. However, coding expertise enables the user to modify Markush TOPFRAG searches so that they search more precisely for the information desired.

Before reading this user guide, it is suggested that the new user become acquainted with the information in the Derwent publication *Introduction to Derwent Chemical Indexing*, which discusses the fundamental concepts of Chemical Code searching in detail. An introductory class is also recommended. The section entitled "Help Tools and Customer Assistance", which starts on page 7, lists these and other means that Derwent provides for helping users develop expertise in chemical patent searching.

Introduction

Derwent World Patents Index – Historical Background

The following brief history of Derwent World Patents Index will give the reader an appreciation of the Chemical Codes in their historical context.

Derwent's first major product was **Farmdoc**, a PHARMaceutical patent DOCumentation service that commenced publication in 1963. (The early product planning took place in Italy, so "Farmdoc" derived its name from the Italian word FARMaceutii rather than the English word PHARMaceutical.) The **Chemical Fragmentation Codes**, designed especially for indexing and retrieving chemical patents, were introduced with Farmdoc.

Agdoc, an agricultural patent documentation service that is indexed and searched with the same chemical coding system as Farmdoc, was initiated in 1965.

Derwent's polymer documentation service, **Plasdoc**, began in 1966. ICI's polymer indexing language was adapted for use in the new service, because the codes used for Farmdoc and Agdoc were not suitable for indexing polymers.

In 1970, the existing chemical patent indexes were organized together with several new subject indexes into the **Chemical Patents Index** (CPI). CPI covers a broad range of chemical technology, as indicated by the 12 CPI Section topics listed below.

CPI Section	Topic
A	Polymers and Plastics
B	Pharmaceuticals
C	Agricultural Chemicals
D	Food, Detergents, Water Treatment and Biotechnology
E	General Chemicals
F	Textiles and Paper – Making
G	Printing, Coating, Photography
H	Petroleum
J	Chemical Engineering
K	Nucleonics, Explosives and Protection
L	Refractories; Ceramics; Cement and Electro (in)organics
M	Metallurgy

The Chemical Codes are used to search for patents in CPI Sections B, C, and E. They are sometimes referred to as the "BCE Chemical Codes", to distinguish them from Section A (Plasdoc) codes. In 1970, some new codes were added to the existing BCE Codes to enable the formulation of more specific Chemical Code search strategies.

In 1972, **new BCE codes** were introduced, as were **Ring Index Numbers** (RIN's). RIN's are based on the *Patterson Ring Index*, published by the American Chemical Society. RIN's represent individual ring systems that are not uniquely identified by Chemical Codes.

Patent coverage was extended to **all areas of technology** in 1974 with the introduction of World Patents Index (DWPI). In 1976, DWPI was made available for interactive online searching on the ORBIT search system.

In Derwent Week 27 of 1981 the original "punch codes", i.e. three-digit codes derived from hole positions on IBM data cards, were replaced by the current four-character alphanumeric codes. At the same time, many new, more specific codes were added to the existing codes.

In 1987, the **TOPFRAG** software was introduced. TOPFRAG translated a (TOPological) chemical structure drawn by the user into a FRAGmentation code search strategy. The current version of TOPFRAG, called Markush TOPFRAG, allows the inclusion of free sites and Markush variations in the structure being searched.

In 1988, a completely new type of patent database and search system were introduced jointly by Derwent, Institut Nationale de la Propriete Intellectuelle, and Telesystemes Questel. The name of the new database is **Derwent World Patents Index Markush** (DWPIM), and the search system used to find information in DWPIM is called **Markush DARC**. DWPIM is searched by means of variable chemical structures drawn by the user, rather than by use of the BCE Chemical Codes; therefore, DWPIM is not discussed in this manual. Because the earliest patents in DWPIM were published in 1987, DWPI will continue to have active patents not found in DWPIM at least until the year 2005, which means that searching with the Chemical Codes will remain important for at least as many years.

Today, DWPI contains over 14 million patent documents organized into over 8 million patent families. This user guide describes in detail one method of finding relevant chemical patents in DWPI - the BCE Chemical Codes. A general discussion of the DWPI database and the wide variety of methods for finding information in it are found in Derwent's *DWPI Online User Guide*.

Derwent World Patents Index Timeline

1963 **Farmdoc** and the Chemical Fragmentation Codes introduced

1965 **Agdoc** introduced, which also utilizes the Chemical Codes

1966 **Plasdoc** and a polymer indexing language introduced

1970 **Chemical Patents Index** (CPI) introduced

1972 **Ring Index Numbers** introduced

1974 **Derwent World Patents Index** (DWPI) introduced, covering all areas of technology

1976 DWPI available **online** on ORBIT (and later on other vendors)

1981 BCE chemical coding system undergoes **major changes**

1987 **TOPFRAG** introduced

Sample Patent Record from DWPI

```
009636110
WPI Acc No: 93-329659/199342
XRAM Acc No: C93-145711
XRPX Acc No: N93-254523
  Flexible optical pulse sensor used partic. as blood oximeter - where
  circuitry on flexible substrate is covered by insulating tape to prevent
  shorting
Patent Assignee: GOULD INC (GOUN  ); GOULD ELECTRONICS INC (GOUN  ); GOULD
  ELECTRONICS INC (GOUL-N)
Inventor: CENTA J A; HALAZ L; HALASZ L
Number of Countries: 003  Number of Patents: 005
Patent Family:
Patent No  Kind  Date   Applicat No Kind  Date   Main IPC      Week
EP 566354   A1 19931020 EP 93302843  A  19930413 A61B-005/00   199342 B
US 5263244  A  19931123 US 92870247  A  19920417 H05K-003/30   199348
JP 6094516  A  19940405 JP 9366344   A  19930325 G01J-001/02   199418
US 5390670  A  19950221 US 92870247  A  19920417 A61B-005/00   199513
                        US 93139564  A  19931020
CN 1078374  A  19931117 CN 93103230  A  19930407 A61B-006/00   199710

Priority Applications (No Type Date): US 92870247 A 19920417; US 93139564 A
  19931020
Cited Patents: DE 3809084; EP 357249; EP 481612; EP 529412; US 4865038; US
  4964408
Patent Details:
Patent   Kind Lan Pg Filing Notes     Application  Patent
EP 566354   A1 E  12
US 5263244  A      9
JP 6094516  A      8
US 5390670  A      9 Div ex           US 92870247
                     Div ex                       US 5263244

Abstract (Basic): EP 566354 A
        A sensor has a flexible printed circuit mounting at least one LED
    and at least one photoelectric detector, connected to diff. circuit
    patterns and with insulating tape (26) secured over circuits and
    components. The sensor can be folded so that the detector can be spaced
    over the LED.
```

The circuit substrate is pref. polyester or polyamide sheet 0.0015-0.010 inc thick, the circuitry is of copper and the tape is of PTFE 0.0005-0.003 inch thick. There are pref. two LEDs (24, 25) mounted across a set of terminals, one emitting red light and the other IR, and only one photodetector (18) mounted across a second set of terminals.

USE/ADVANTAGE - Partic. for mounting in a finger pouch for use as a pulse oximeter, is easy and inexpensive to mfr. and does not require an opening to be formed in the substrate for components.

Dwg.3/5

Abstract (Equivalent): US 5390670 A

Flexible sensor assembly for detecting optical pulses comprises a flexible sheet carrying a first pair of conductive strips (6,7) with a photodetector (18) mounted across its terminals (8,9) at one end.

A second set of strips (12,13) has LED's (24, 25) mounted across its top terminals (14, 15). An insulating tape is secured over the photodetector, LED's and circuitry at least up to the bottom sets of terminals on both conductive strips (10, 11, 16, 17). The assembly can be folded so that the photodetector can be spaced from but disposed over the LED.

The flexible sheet is pref. made of polyester or polyamide and the insulating tape is pref. made of PTFE.

USE - In pulse oximeters for measuring the 02 content of blood.

Dwg.2/5

US 5263244 A

Flexible sensors for detecting optical pulses are produced by forming a circuit pattern on a Cu-coated flexible sheet, each circuit having two pairs of spaced strips with terminals, etching off exposed copper, and securing a photodetector (18) across a set of terminals (8,9) attached directly to one strip and attached to the other by a conductive clip for physical and electrial securement.

A LED (24,25) is secured across a set of terminals on the second pair of strips, a layer of non-conductive adhesive tape is applied across the circuits and components at least up to the remaining terminals, and each circuit with its components is cut from the sheet. The sheet is pref. of polyester or polyamide, the tape is PTFE, and the components are attached with conductive epoxy resin which is then heat-cured.

USE/ADVANTAGE - E.g., for use in a pulse oximeter for detecting blood oxygen level, does not require an opening formed in the substrate for the components and has a non-conductive covering to prevent shorting in use.

Dwg.3/56

```
Title Terms: FLEXIBLE; OPTICAL; PULSE; SENSE; BLOOD; OXIMETER; CIRCUIT;
  FLEXIBLE; SUBSTRATE; COVER; INSULATE; TAPE; PREVENT; SHORT
Derwent Class: A96; B04; P31; S03; S05; V04
International Patent Class (Main): A61B-005/00; A61B-006/00; G01J-001/02;
  H05K-003/30
International Patent Class (Additional): A61B-005/024; H01L-033/00;
  H05K-003/00; H05K-003/02
File Segment: CPI; EPI; EngPI
Manual Codes (CPI/A-N): A04-E08; A12-E13; A12-V03D; B04-C03B; B04-C03D;
  B11-C08B; B12-K04
Manual Codes (EPI/S-X): S03-E04A4; S05-D01G; V04-R05D
Plasdoc Codes (KS): 0207 0210 0231 0947 0968 1282 1283 1288 2020 2198 2493
  2522 2532 2551 2628 2654 2682 2706 2740 2743 2768 2815
Polymer Fragment Codes (PF):
  *001* 017 04- 141 143 502 551 560 566 575 596 623 627 628 643 645 722 726
  *002* 017 04- 062 064 087 090 487 506 509 575 596 623 627 628 643 645 668
        688 722 726
  *003* 017 04- 226 231 359 473 506 509 609 623 627 628 643 645 722 726
Chemical Fragment Codes (M1):
  *01* M423 M424 M740 M760 M903 N102 V600 V615
  *03* G013 G019 G100 H4 H401 H481 H8 J0 J014 J1 J131 J2 J232 M280 M312
       M323 M332 M342 M383 M393 M423 M424 M430 M510 M520 M533 M540 M740
       M782 M903 M904 M910 N102 P831 V0 V743 R02038-D R02038-M
  *04* H1 H100 H181 J0 J014 J1 J171 J3 J373 M280 M314 M315 M316 M323 M332
       M342 M381 M382 M383 M393 M423 M424 M430 M510 M520 M530 M540 M620
       M740 M782 M903 M904 M910 N102 P831 V0 V743 R02035-D R02035-M
  *06* H6 H601 H607 H609 H684 H689 H7 H721 M280 M312 M321 M332 M344 M363
       M391 M423 M424 M430 M510 M520 M530 M540 M740 M782 M903 M904 M910
       N102 P831 V0 V743 R00975-D R00975-M R00975-Q
Chemical Fragment Codes (M2):
  *02* C108 C550 C810 M411 M424 M740 M750 M903 M904 M910 N102 R01779-A
  *05* A429 C810 M411 M424 M430 M740 M782 M903 M904 N102 P831 R05099-D
       R05099-M
Chemical Fragment Codes (M6):
  *07* M903 P831 R150 R514 R515 R528 R533 R611 R627 R639
```

Help Tools and Customer Assistance

The resources listed below are provided to help customers learn how to effectively search for chemical patents with the BCE Chemical Codes.

User Aids

- **Introduction to Derwent Chemical Indexing** helps new and prospective users of DWPI become acquainted with the history, logical construction, and use of the Chemical Codes.
- The **Chemical Indexing User Guide** (this manual) is a reference work that describes coding rules, code definitions, and the construction of code search strategies.
- The **Online User Guides** discuss the Derwent World Patents Index database and the many methods that can be used to access its contents.
- The **Chemical Code Dictionary** is an alphabetical subject listing that refers the user from concepts and nomenclature to the corresponding Chemical Codes.
- The **Chemical Coding Sheet** is a concise summary of the BCE Chemical Codes. It gives an abbreviated description of each code, provides an overview of the logical sequence of the codes, and indicates the year that each code became available for searching. Copies of the Chemical Coding Sheet are available from your local Derwent office.
- The **Markush TOPFRAG** is a software program that converts specific and variable chemical structures drawn by the user into Chemical Code search strategies.

Training

Specialist staff run regular training programmes to help you to search our online databases effectively and comprehensively. Training classes are held in major towns and cities in Europe, North America and Japan. A wide range of classes are available introducing Derwent files to both beginners and experienced searchers and subject specialists. These classes include the following specialist training courses covering the Chemical Codes:

- Searching Derwent World Patents Index Online
- Introduction to Chemical Code Searching
- Advanced Chemical Code Searching
- Searching Chemical Structures in DWPI using Markush TOPFRAG/STN Express

The Derwent Help Desk

Expert advice and support is available via our Derwent Help Desk staff, to provide a fast and efficient response to all your enquiries. The experienced Help Desk staff have an in-depth knowledge of all Derwent's products and services and are familiar with the command languages of the various online hosts.

In addition to your online and other product queries, the staff will also assist with delivery and invoice enquiries, take details of amendments to contact and address details and action correction of any reported errors. In fact, from general customer queries through to technical questions on chemical or polymer indexing or how to set up your internet browser, the Help Desk is there to help you.

Contact your local Help Desk by phone, fax or email or visit the Customer Services area on the Derwent World Wide Web Site.

Universal Freephone

The European Help Desk has recently launched a Universal Freephone Number +800 33 44 2999.

At present this is available in the following countries:

Belgium
Denmark
France
Germany
Norway
Sweden
Switzerland
The Netherlands
UK*

Additional countries will be made available as they join the Universal Freephone system.

*UK users should note that it is necessary to dial the international prefix '00', ie 00800 33 44 2999

Europe & Rest of World

Tel +44 (0)20 7344 2999
Fax +44 (0)20 7344 2900
Email custserv@derwent.co.uk

Universal Freephone

UK 00800 33 44 2999
International +800 33 44 2999

North & South America

Tel +1 703 706 4220
Toll Free +1 800 451 3551
Fax +1 703 838 0450
Email search@derwentus.com

Japan

Tel +81 3 5218 6500
Fax +81 3 5218 7840

Derwent World Wide Web Site

http://www.derwent.com/
http://www.derwent.co.jp/

2 Chemical Code Overview

2 Chemical code overview

Structure and Arrangement of the Chemical Codes

The BCE Chemical Codes are a set of alphanumeric symbols, each of which represents a chemical structure fragment, a biological activity, a formulation, a use, or a property disclosed in a patent document. The code **F431**, for example, represents a **pyridine ring**. As indicated by their name, the BCE Chemical Codes are used to index patents classified in Derwent Sections B (pharmaceutical, medical, and veterinary), C (agricultural), and E (general chemical). The Chemical Codes are divided into 21 categories, Parts A: through W:.

Elements Present	Part A:	Metals Present
	Part B:	Less Common Non-metals Present
	Part C:	Common Non-metals Present
Ring Systems	Part D:	Fused Ring Heterocycles
	Part E:	More Fused Ring Heterocycles
	Part F:	Mononuclear Heterocycles
	Part G:	Carbocycles
Functional Groups	Part H:	Common Functional Groups without >C=O or >C=S
	Part J:	Common Functional Groups with >C=O or >C=S
	Part K:	Bonds Between Heteroatoms in Organic Compounds
	Part L:	Other Less Common Groups
Miscellaneous	Part M:	Miscellaneous Descriptors (Ring to ring linkages, carbon chains, type of patent, etc.)
Non-Structural Codes	Part N:	Chemical Reactions, Bonds Broken and Formed
	Part P:	Activities, Properties, Uses - Sections B and C
	Part Q:	Activities, Properties, Uses - Section E
	Part R:	Formulations, Galenicals
	Part V:	Natural Products, Polymers
	Part W:	Dyes
Steroids	Part S:	Steroid Descriptors I
	Part T:	Steroid Descriptors II
	Part U:	Steroid Descriptors III

Note The fact that CPI Sections and Chemical Code Parts are both represented by letters of the alphabet may cause some confusion for new users. All chemical patents received by Derwent are classified by subject matter into one or more CPI Sections (listed on page 2). Most chemical structures in patents classified in Sections B, C, and E are indexed with the Chemical Codes. The Chemical Codes are divided into 21 Parts, or categories, as listed above.

The principal four-character Chemical Codes are hierarchical: the first character represents the code **Part**, the first two characters the code **Set**, the first three characters the code **Subset**, and all four characters represent the code. This hierarchy allows reference to be made to larger or smaller groups of related codes as shown in the example below, in which the definitions in the right column become increasingly more specific as one moves from the **Part** level of the hierarchy to the **code** level.

Part	H:	Common Functional Groups Without >C=O or >C=S
Set	H7:	Olefin, Acetylene Bonds
Subset	H73:	C≡C Present
Code	H7301	C≡C Present in a Ring

The code hierarchy is easily observable on the coding sheet and in the portions of this user guide that list code definitions. The colon at the end of the Part, Set, and Subset characters is a truncation symbol that can represent any further character(s). An entire Set or Subset of codes can be searched by truncating a code after the second or third character, respectively.

This user guide is organized similarly to the coding hierarchy. Correct use of the Chemical Codes requires an awareness of the rules, scope of applicability, and restrictions presented in the text at each level of the coding hierarchy. The applicability of codes is discussed more at length in the section entitled "Selecting Applicable Chemical Codes", beginning on page 16.

Special-Purpose Codes

In addition to the four-character codes discussed above, two-character codes were created for negative searching, i.e. for the removal of irrelevant references. ***Negation codes***, or **essential codes** as they are also called, are primarily used to indicate the functional groups that are present or absent in a structure. They are therefore discussed in detail in the chapter on functional groups, beginning on page 101.

A ***Poly*** code is a standard four-character code found at the end of many Sets and Subsets. A **Poly** code indicates that one of the group of codes to which it refers is applicable more than once to a structure. **Poly** does not mean that more than one code from a group of codes is applicable to a structure. No **Poly** code is provided in Sets or Subsets with codes that indicate a specific number of occurences of a structural feature. For example, no **Poly** code is needed in Set H4:, because the codes themselves specify exactly how many hydroxy groups are present.

The ***control codes*** in Set M9: direct the search strategy to examine only a designated portion of the DWPI database. The control codes are defined on page 188. **Time range** control codes warrant a more detailed discussion, which begins on page 22.

A ***precursor code***, also known as a **discontinued generic code**, is a code that was replaced by a series of more specific codes, but which can still be searched in the earlier references of the database. Precursor codes are discussed in the following section.

Code Modifications in 1981 Derwent Week 27

The Chemical Codes underwent many changes in 1981. The older three-digit **punch codes**, which were derived from hole positions on IBM data cards (or punch cards), were replaced by the current four-character codes. Many of the older codes were simply converted to the newer codes on a one-to-one basis, so that searching a newer code seamlessly retrieves references that were originally indexed with the older code. In some cases, however, the transition to the newer codes was not a simple code-for-code conversion. A number of coding modifications were made that improved retrieval for more recent references (post-1981), but which require special attention in the search strategy or in the analysis of search results. Two such modifications are discussed in the following paragraphs.

A **#** sign next to a code on the coding sheet indicates that several newer codes have replaced an older code that was less specific. The less specific code, known as a **"precursor code"** or **"discontinued generic code"**, can still be searched in the time ranges preceding the introduction of the current code. (Precursor codes are listed and defined in this user guide, but they do not appear on the coding sheet. *Time ranges* are discussed on page 22). Precursor codes are created by replacing the final digit in the current codes with the digit "0". The discontinued code and the applicable current code are 'OR'ed together as one search term in the search statement appropriate for the precursor code, as in the search strategy on page 24, where M370 is the precursor code and M373 is the newer, more specific code. Thus a precursor code introduced in 1970 is 'OR'ed with a more specific code in the search statement used for codes introduced in 1970 (i.e. the red codes on the coding sheet). The OR logic ensures that either of the terms can produce a hit, the precursor code in earlier references and the newer code in more recent references. If precursor codes are omitted completely from the search strategy, the result is less-specific retrieval of references published prior to 1981. Precursor codes are not listed in indexing examples; the newer, more specific codes are listed in the examples, and it is assumed that their precursor codes would be included in a search strategy.

In a few cases, a single older code was **divided into two** new, more specific codes. The two new codes allow more precise searching from 1981 forward, because the recent coding distinguishes between the two meanings represented by the codes. However, references dated prior to 1981 that were indexed with the older code are retrieved equally by either of the newer codes. The older code didn't distinguish between the two meanings of the newer codes, and thus searches on the older documents tend to result in more irrelevant references than searches on the newer documents.

3 Searching With The Chemical Codes

3 Searching with the chemical codes

Introduction

This section describes the selection of applicable Chemical Codes for structure searches, and the construction of Chemical Code search strategies. Before the selection of codes begins, the user should verify that the role that the searched compound is likely to play in patents of interest is one that Derwent considers significant enough to warrant chemical coding. (See "Selecting the Compounds to be Coded" on page 27.)

Markush TOPFRAG – Automatic Search Strategy Construction

The simplest way for a new user to get started with Chemical Code searching is to use the Markush TOPFRAG software. Markush TOPFRAG:

- analyses chemical structures drawn by the user;
- converts them into the corresponding fragmentation codes; and
- constructs Chemical Code search strategies that take into account all relevant rules.

The substructures that are input may include free sites and the general chemical terms that are found in patents, e.g. *alkyl*, *heterocycle*, and *halogen*. With Markush TOPFRAG, a user can learn to use the Chemical Codes at a convenient pace, and can formulate effective Chemical Code searches even during the early stages of the learning process. The search strategies created by Markush TOPFRAG can be simple or very complex, depending on how many variable units are included in the structure being searched.

Although Markush TOPFRAG is a very helpful tool for automatically constructing Chemical Code search strategies, the strategies it generates are not always exactly what the user intends. For example, some search strategies created by Markush TOPFRAG are too precise; codes may need to be deleted to make the search more general and comprehensive. Using Markush TOPFRAG is in some respects similar to using an expensive automatic camera. The results from the automatic camera are almost always quite good, and inexperienced users could not do better by setting the exposure themselves. Even experienced camera users could use the automatic features to some advantage. However, there inevitably arise situations when an experienced user wants to have more control over the outcome, to produce special effects for example, and would thus adjust the settings to produce a result consistent with the user's better judgment. In the same way, experience with the Chemical Codes enables users to make adjustments to Markush TOPFRAG's search strategies, yielding results that are more relevant to the search question.

It is good practice to look up the definitions of all codes generated by Markush TOPFRAG in this user guide, checking to see if the codes are appropriate for the specific search being conducted.

Selecting Applicable Chemical Codes

Structure Analysis

The first step in developing a Chemical Code search strategy is analysis of the structure being searched. Once the structure is divided into "codeable" fragments, potentially relevant codes that correspond to those fragments can be selected for further consideration. The following tips mention the most important types of chemical elements and fragments that are coded, and point the user to many of the applicable codes.

- **Inorganic compounds, ions, and ligands** are indexed in Part A:, B:, and/or C:.
- **Elements other than H, C, N, O, S** are indexed in Part A:, Part B:, or Part C:.
- **Ring systems** are indexed in Part D:, Part E:, Part F:, or Part G:.
- The **number of each type of ring system** present is indexed in Set M5:.
- **Functional groups**, i.e. non-ring structural fragments that contain at least one heteroatom, may be indexed in Parts B:, H:, J:, K:, and/or L:, depending on the atoms involved.
- **Carbon chains, ring to ring linkages**, and other **miscellaneous** features are indexed in Part M:.

The most common elements, i.e. **H, C, N, O, and S**, are only indexed as individual elements if they are present in inorganic compounds, ions, or ligands. Otherwise, they are indexed in terms of the rings and functional groups in which they are present. For more information about coding elements, see Chapter 6, "Elements Present".

Some ring codes refer ambiguously to more than one ring system. Searches that utilize ambiguous ring codes can be made more precise by the addition of **Ring Index Numbers** (RIN's) to the search strategy. An asterisk (*) next to a ring code in this user guide or on the coding sheet indicates that RIN's have been assigned to various structures covered by the code. RIN's are discussed in more detail on page 74.

Rules and Restrictions of Code Usage

The application of Chemical Codes to structures is governed by coding rules and conventions. Parts, Sets, and Subsets are in most cases accompanied by an Introduction and notes that specify coding rules applicable to use of the codes they contain. Rules that only apply to individual codes follow the code definitions. The notes and rules given at all levels of the hierarchy of a code (i.e. rules given at the Part level, rules given at the Set level, etc.) must be understood before using the code in a search. Otherwise, the search may not perform as intended.

While the definition of a code may match the structure being searched, if the overall *scope of applicability* of the code does not include the type of fragment or compound in question, the code cannot be used in the search strategy for that structure. For example, looking at the coding sheet, a user might decide that a trivalent N atom present in a pyridine ring should be coded C107. However, reading the introduction to Part C: reveals that Part C: codes are only applicable to inorganic compounds, with a few exceptions. Instead, the N atom in the ring is correctly specified by the ring code F431 ("Pyridine").

Related to the scope of applicability of a code are **restrictions** placed on the types of compounds that can be assigned a certain code. For example, each code can only be used with certain **subheadings**, which are based on compound type. (Subheadings are discussed on page 23.) Some codes can only be used for patents classified in specific **Derwent Sections**; for example, most codes in Part R: are only assigned to Section B (Pharmaceutical, Veterinary, Medical) patents.

The use of some codes is governed by **priority rules** that determine which code is used when there are several apparently "correct" alternatives. For example, the priority rule for finding the correct functional group codes states that Part K: should be consulted first, followed by Part L:, Part H:, and finally Part J:. Although an apparently correct code may be found in Part J:, it can be **superseded by a code of higher priority** in Part L:, in which case the Part J: code is not assigned. As another example, the priority rule for ethers states that only the highest priority ether attachment is coded, in the order **heterocyclic > aromatic > alicyclic > aliphatic**. Therefore, if an ether oxygen is bonded to one alicyclic ring and one heterocyclic ring, "Ether bonded to heterocyclic carbon" is coded, but "Ether bonded to alicyclic carbon" is not coded.

The **priority rule** of many Sets and Subsets states that only the **first correct code** in that Set or Subset is assigned; other equally correct codes in that Set or Subset are ignored. One such Set of codes, Set M4:, describes the type of compound patented:

M411	Inorganic/Miscellaneous
M412	Fused Ring Heterocycle
M413	Monocyclic Heterocycle
M414	Benzene or benzene fused to any carbocycle
M415	Non-aromatic carbocycle
M416	Aliphatic
M417	Incomplete structure

There may be aliphatic, heterocyclic, aromatic, and inorganic fragments in the structure, but if "inorganic" is the first term applicable to the structure, it is the only one whose code is added to the indexing record. Other similar priority rules are listed throughout this user guide.

Some coding rules pertain to the **preferred form** of a structure. Rules have been developed that specify which of several possible structural forms should be the one that is coded. For example, when keto-enol **tautomerism** is possible, the structure is coded in the keto form, unless the -OH group of the enol form is bonded to a fully conjugated carbocyclic ring, e.g. benzene. The same rule is assigned to analogous thioxo-thiol compounds. Tautomer rules with structural examples are presented in the section "Tautomerism", which begins on page 106. Preferred forms of dyestuffs are listed on page 270.

Because the **coding sheet** gives only a summary definition for each code, it should be used as a memory device or a means of finding potentially applicable Sets of codes rather than as a coding guide.

Coding Rules for Polymers and Halogens

Although coding rules for specific types of elements and compounds are listed throughout the user guide where applicable, rules for coding **polymers** and **halogens** are discussed in this section because the codes used to describe them are not found in any one chapter.

Polymers are most specifically indexed with special polymer codes. (Contact your local Derwent office for information about Derwent's extensive polymer indexing and searching resources.) However, some Chemical Codes are also assigned to polymers. Both natural and synthetic polymers of patents classified in Section B (Pharmaceutical, Veterinary, Medical) are assigned codes from Part V:. Section B synthetic polymers are assigned additional codes as follows:

- The hypothetical monomers of >C=C< addition polymers are assigned all applicable Chemical Codes. For example, patents on polyvinyl chloride are assigned applicable Chemical Codes for vinyl chloride.

- Prior to 1981, monomers or condensants of polymers other than addition polymers were structure-coded, except for groups that reacted during the condensation or polymerisation. Linking groups were coded as present more than once using applicable **Poly** codes. (**Poly** codes, which did not derive their name from the term *polymer*, are discussed on page 12.)

- From 1981 forward, polymers other than addition polymers have been coded as if they were single compounds, meaning that monomers or condensants have not been coded individually. For example, all groups in polyethylene terephthalate except the end groups are coded as present more than once using applicable Poly codes.

- Also from 1981 forward, negation codes and codes from Sets M1: ("Ring to Ring Linkages"), M2: ("0- or 1-valent Carbon Chains"), and M3: ("Polyvalent Carbon Chains") have been assigned to polymers.

The code Part(s) used for coding a **halogen-containing group** depend on the atom(s) to which the halogen is bonded. The halogen atom itself is identified either in Set C0: or Subset H60:, again depending on the atom(s) to which it is bonded. The functional group formed with the halogen may receive further codes. The following list of halogen fragments and codes directs the user to potentially applicable codes; the code Parts and Sets listed should be consulted for specific guidelines.

Functional Group Description	Codes Potentially Applicable to Halogen
1. Halogen bonded to metal	C0:, A940, (see Part C: for additional codes)
2. Halogen anion	C0:, C720, (see Part C: for additional codes)
3. Halogen in inorganic compound, ion, ligand	C0:
4. Monovalent halogen bonded to C, but not to C=U (U is O, S, Se, Te, N)	H6:
5. Halogen bonded to O, S, Se, Te, N, halogen	K1:, K222, K352, K410, K531, K620, K741, K751, K810, K910, L310, L320, L351, L410, L431, L520, L541

Note The halogen atom in these codes is represented by X or Y. The particular halogen represented by X or Y has been identified since 1978 by a code in Set C0:

6. Valency of Halogen	L760, C100, C200, C300
7. Halogen as ring member	C600, C0:
8. Halogen bonded to C=U (U = O, S, Se, Te, N)	L330, L353, L511, L512

Note The halogen atom in these codes is represented by X. The particular halogen represented by X in L330 and L353 has been identified since 1963 by a code in Subset H60:. The particular halogen represented by X in L511 and L512 has been identified since 1978 by a code in Set C0:.

Variable or Unknown Substituents

Variable substituents can be included in Chemical Code searches by 'OR'ing together the codes applicable to the possible substituents. The varying substituents may affect related codes as well, which can make the development of a variable-structure search strategy complicated. All variable possibilities must be taken into account, and if related codes change according to the variable substituents, the different possibilities for those codes are also 'OR'ed together. The user must judge whether the increased complexity of the search strategy is warranted by the increase of relevance gained by inserting the variable substituents codes. One alternative is to delete the highly variable portions of the structure being searched, thus gaining simplicity at the cost of a less-specific search. Another alternative is to use Markush TOPFRAG, which accepts variable substituents and automatically creates a search strategy that takes the variability into account.

If there are **unknown substituents**, all *counting codes*, i.e. codes that indicate the number of times a certain group is present in a structure, must be modified or deleted to take the unknown substituents into account. For example, if it is known that there is one hydroxy group bonded to a ring that also has an unknown substituent, then the codes for all possible numbers of hydroxy substituents should be 'OR'ed together. **Negation codes** are not used when there are unknown substituents in the search structure.

Several examples of coding variable and unknown substituents are given in Chapter 8, beginning on page 101.

Comprehensiveness vs. Relevance

As codes are being chosen for inclusion in a search strategy, the user must decide how comprehensive the search is to be. In effect this is a decision on which applicable codes, if any, should be **omitted from the search strategy**. Suppose one fragment of a structure being searched is an ethyl group attached to a phenyl ring. The code M212 is the correct code to describe the ethyl group. However, if patents claiming compounds with a methyl or propyl group attached to a phenyl ring would be of interest to the user as well, then the M212 code should either be deleted from the search strategy or the strategy should be **broadened with "OR" logic** to include the other acceptable possibilities. The **less restrictive** the search strategy is, the **more comprehensive** the results will be. Unfortunately, comprehensive search strategies often retrieve many irrelevant answers. This is a tradeoff which the user must always be aware of when selecting Chemical Codes.

Once a Chemical Code search has been conducted, **other kinds of indexing terms** can be used to restrict the search to the most relevant answers, e.g. keywords, Manual Codes, company names, etc. The increased relevance may be offset by a decrease in the comprehensiveness of the results. That is, some pertinent references may be eliminated when the other indexing terms are used to narrow down the answer set.

Four Elements of the Chemical Code Search Strategy

Once applicable codes have been selected, a search strategy can be formulated. The simplest search strategy, one that is not recommended, would be to search all of the selected Chemical Codes in one search statement using the AND operator. The weaknesses of this search strategy, and the solutions for overcoming those weaknesses, are discussed in the following paragraphs.

Card Records and the LINK Operator

The **Markush structures** of chemical patents disclose many compounds without naming them individually. Instead, there are lists of possible ring systems, substituents, elements, etc., which together constitute a highly variable structure. Markush structures can represent thousands or even millions of possible compounds, making it **impossible to index each one individually**. The Chemical Codes were developed as a solution to this problem, as discussed in the Derwent manual *Introduction to Derwent Chemical Indexing*. Indexers describe each variable chemical fragment of a chemical invention with one or more Chemical Codes. The codes corresponding to the core patented structure are listed together with codes for all of the variable parts. Formerly this was done on a single data entry card, which was called a **card record**. In database terminology, the codes for the many different possible compounds described by a single Markush structure are listed in the same **database subfield**. Different chemical inventions within each patent were in earlier years coded on different card records, and are currently listed in different database subfields. For example, the patent record on page 4 has seven code subfields, numbered *01*, *02*, *03* … *07*. (The term "code subfield" is used in the remainder of this manual, rather than the older term "card record".)

The **LINK** operator is used for Chemical Code searching instead of the **AND** operator, thus requiring that all matching codes be located within the same subfield to be retrieved by the code search. This ensures that only codes related to a single chemical invention are examined for a match with the search strategy, rather than allowing codes found anywhere within the same patent record to cause the record to be retrieved.

There are limits to the variety of compounds that can be coded in one card record, or database subfield. These limits are called "overcoding rules", and are listed on page 30.

Negation codes

Although the user can LINK the codes of the search strategy so that they must all be present in the same code subfield, the codes causing the hit may nevertheless have been derived from different individual structures represented by one Markush formula. Thus grouping the codes for many related structures into a single code subfield, which is referred to as **overcoding,** can result in **false drops** (irrelevant answers). Some of these false drops can be removed from the search results using **negation codes**. Negation codes, which are discussed in detail on page 108, are codes that were created to eliminate patents in which every structure disclosed has a particular fragment not wanted by the user.

Time Ranging

A number of improvements have been made to the Chemical Codes since their introduction in 1963. New codes were created and some existing codes were sub-divided in order to enable the creation of more specific search strategies.

The currently valid Chemical Codes are shown on the coding sheet (see page 7), grouped according to code Part. Groups of codes introduced at the same time are printed in the same color:

- Codes introduced in **1963** are **BLACK**
- Codes introduced in **1970** are **RED**
- Codes introduced in **1972** are **BLUE**
- Codes introduced in **1981 Derwent Week 27** are **GREEN**

The four distinct time periods delineated by the introduction of new codes are referred to as **time ranges**. Every patent in DWPI is assigned a time range Chemical Code based on when it was introduced.

M900	was assigned to all documents coded between **1963-1969**
M901	was assigned to all documents coded between **1970-1971**
M902	was assigned to all documents coded between **1972-1981 Derwent Week 26**
M903	has been assigned to all documents coded from **1981 Derwent Week 27 to the present**

Most code concepts introduced since the beginning of the file are still valid for searching. Therefore, a different search is needed for each time range, with each search utilizing the **most precise codes available** during the corresponding time period. Although it takes some study to understand how it functions, the Standard Search Strategy uses a **time-ranging** search statement to integrate the four necessary searches into a single search strategy. If instead the codes selected for searching were simply 'AND'ed together in one search statement, it is likely that only references from 1981 and later would be retrieved, because only those references would include the most recently introduced codes. The logic behind the time-ranging search statement is discussed further in the section on the Chemical Code search strategy on page 24.

Subheadings

Examining the entire DWPI database for every code search is an inefficient process, because the vast majority of patents in DWPI are not relevant to any one search structure or topic. The chemical portion of DWPI, Chemical Patents Index (CPI), has been divided into a number of subsets to allow the user to avoid searching for codes in parts of the database that are definitely irrelevant to the search structure. Patents may be included in more than one subset if their subject matter warrants it. The subsets are designated by the **subheadings** M0 through M6, and are both year- and subject-oriented. Patent records in CPI have a separate Chemical Code field for every subset in which the patent is classified, and each code field is labelled with the appropriate subheading as defined in the following table.

M0	All Pre-1970 Agricultural, Pharmaceutical Non-Steroids	1963 - 1969
M1	Agricultural, Pharmaceutical - Natural Products and Polymers	1970 to present
M2	General Chemical - Agricultural, Pharmaceutical	1970 to present
M3	General Chemical - Other	1970 to present
M4	Dyes	1970 to present
M5	Steroids	1963 to present
M6	Galenicals (pharm. apparatus, formulation, diagnostics)	1976 to present

The subheading M0, for example, was assigned before 1970 to non-steroidal compounds with pharmaceutical or agricultural uses. Because M0 was only assigned prior to 1970, it is only used to search for codes that are black on the coding sheet. The phrase "General Chemical" in the definitions of M2 and M3 means that these subheadings are not used for the other specifically mentioned compound types, i.e. natural products, polymers, dyes, steroids, or galenicals.

The use of subheadings as field labels in the search strategy is explained in the following section. The user includes the subheadings of interest and the rest are ignored when the search is conducted. If the user does not know which subheadings have been used for the structures of interest, all subheadings can be included in the search strategy . The only negative effect this will have is a lack of efficiency in the database searching process. It is preferable, however, to use as few subheadings as possible.

The subheadings used are determined by the type of structure being searched, and the years relevant to the search. For example, if the structure being searched has a cyclopenta(a)phenanthrene ring system, then M5 is the only subheading needed in the search strategy. Conversely, if it is clear that compounds of interest do not have a cyclopenta(a)phenanthrene ring system, the subheading M5 can be omitted from the search strategy. The same inclusion/exclusion logic applies to the subheading M6, which is used to designate codes referring to pharmaceutical apparatus, formulations, or medical diagnostics.

If the compound being searched is a polymer assigned to Section B or a natural product, the subheading M0 is used to search for black codes (as represented on the coding sheet), and the subheading M1 is used for all other codes. In most other searches, the subheading M0 is used to search for black codes, and the codes M2, M3, and M4 are used to search for the remaining codes. These three subheadings are not usually searched individually, because the categories they represent are not as distinct as those represented by the other subheadings, and users don't want to miss a relevant structure just because it's category is other than expected.

Putting It All Together – The Chemical Code Search Strategy

A typical Chemical Code search strategy automatically generated by Markush TOPFRAG is shown below.

```
S M0,M2,M3,M4=(F433(S)G100(S)H121(S)H401(S)H481(S)H602(S)M413(S)M531)
S S1(S)M2,M3,M4=(M123(S)M132(S)M521)
S S2(S)M2,M3,M4=(M280(S)M312(S)M321(S)M332(S)M343(S)M391(S)(M370+M373))
S S3 and RR=03412
S S4(S)M2,M3,M4=(F012(S)F013(S)F016(S)G010(S)H100(S)H621)
S (S1(S)M0=M900)+(S2(S)M2,M3,M4=M901)+(S4(S)M2,M3,M4=M902)+S5
S S5(NOT S)M2,M3,M4=(H2+H3+H5+H7+H9+J0+J1+J2+J3+J4+J5+J6+J9+K0)
```

This search strategy looks quite different from the usual strategies used in online searching. The two main differences (other than the "+" abbreviation for "OR") are:

- The search strategy above is designed to only search for codes in relevant **subsets** of the database.
- Chemical Codes have changed over the years, and this strategy is designed to search the most specific codes available during each distinct time period.

The **components of the search strategy** were discussed in previous sections. They are:

Subheadings	Field labels that limit the code search to certain subsets of the database e.g. M0, M2, M3, M4
Chemical Codes	Search terms described in this manual e.g. F433, G100, H121, H401, H481
Time range codes	Codes which limit search statements to certain time periods M900, M901, M902, M903
LINK operator	Search operator that requires all matching codes to be located within the same subfield
Negation codes	Codes used to remove irrelevant patents from the search results e.g. H2, H3, H5, H7
Ring Index Number	Number denoting the presence of a specific ring system e.g. 03412

The codes in the search strategy shown above are grouped according to the time of their introduction, which is indicated by their color on the coding sheet:

Search Statement 1	All **BLACK** codes
Search Statement 2	S1 (LINK) all **RED** codes
Search Statement 3	S2 (LINK) all **BLUE** codes
Search Statement 4	**RIN** 'AND'ed with S3
Search Statement 5	S4 (LINK) all **GREEN** codes
Search Statement 6	**Time range** linking
Search Statement 7	**Negation codes**

Answer Set 1 (S1) consists of records that can have any time range code, i.e. **M900**, **M901**, **M902** or **M903,** because the **BLACK** codes were assigned to patents during all time periods. As newer codes are LINKed to the first Set of codes in later search statements, the older records are eliminated from the answer set.

S2 consists of records with **M901**, **M902** or **M903** only. The **RED** codes were introduced in **1970**, so there will be no records in S2 with the time-range code M900.

S3 consists of records with **M902** or **M903**.

S4 consists of records with **M902** or **M903** and the **Ring Index Number** (RIN) 03412.

S5 consists of records with **M903** only.

The records in S5 are the most specifically selected records from the database, because the records in S5 have all of the codes from the search strategy in a single code subfield. However, S5 only contains records added to DWPI since 1981, because codes used before 1981 were 'LINK'ed in the search strategy to the GREEN codes, and no reference prior to 1981 Derwent Week 27 contains GREEN codes. A method is therefore needed to selectively retrieve documents prior to 1981. This method, which is performed in Search Statement 6 (SS6) above, is called **time ranging**.

Using the time-range codes, SS6 accepts patent records retrieved only by BLACK codes (i.e. by SS1) IF the records were indexed before 1970. If they were indexed after 1970, they must have the RED codes searched in SS2 to be considered relevant. The records found by SS2 are accepted by SS6 IF they were indexed between 1970 and 1971, i.e. if they have the time-range code M902. Records indexed later than 1971 should have been retrieved by SS3 if they are to be considered relevant. Records found by SS3 are accepted IF they were indexed between 1972 and 1981 (M903), and if they have the RIN 03412. Records indexed after 1981 should be found by SS5 if they are really relevant, and should include all of the codes and RIN's used in the search strategy.

SS7 removes records found by the search strategy that have any of the negation codes listed.

Note This section described the most commonly used Chemical Code search strategy, but variations are sometimes necessary, depending on the structure being searched.

Searching With Other Kinds of Codes

Derwent Registry Numbers

In 1981, a list of about 2100 compounds and elements commonly mentioned in claims and examples was compiled, and each compound was assigned a Registry number. The purpose of the Registry is to improve retrieval of compounds that occur often in patents. A patent is indexed with a Registry number when it mentions a Registry compound as a **significant compound or non-metallic element** in the claims or examples. (An example of an **insignificant** compound is a solvent mentioned in a patent when any solvent may be used.)

The CPI Registry Compounds User Guide contains a list of these compounds in both alphabetical orders and in number order.

The alphabetical listing contains the standard compound name, synonyms if appropriate, molecular formula and the Registry number.

n-Octanoic acid (CAPRYLIC ACID)
$C_8H_{16}O_2$ **1061**

Manual Codes

Manual Codes are alphanumeric codes that appear similar to Chemical Codes. Each Manual Code represents a term in a hierarchical controlled term vocabulary. While the Chemical Codes are designed to **determine if certain chemical structures are present** among the thousands or millions of compounds claimed or disclosed in a chemical patent, Manual Codes are most useful for **general classification** of patented compounds, reactions, and uses. Manual Codes can be 'AND'ed with results from a Chemical Code search to increase the relevance of the results to the search question. Manual Codes are discussed in detail in Derwent's CPI Manual Codes Manual.

4 How Derwent Indexes Patents

4 How Derwent indexes patents

Introduction

Today, most Chemical Code indexing is done automatically by sophisticated graphics programs similar to Markush TOPFRAG. The results are then verified by code experts. Understanding the guidelines and processes used in the coding of patents can help the user create better search strategies and more accurately interpret search results.

Selecting the Compounds to be Coded

Derwent indexes the widest chemical disclosure of patents, that is, the greatest range of compounds each patent can exclude others from making, using, or selling. **The possible chemical permutations** covered in each structural invention of the patent are separated into chemical fragments, and these fragments are then translated into Chemical Codes. The following types of compounds and uses are indexed with Chemical Codes:

- **all compounds and reaction intermediates stated to be new**
- **products of new processes**
- **new uses of known materials**
- **materials detected** and **detecting agents**
- **detection media** since 1970 (Derwent Sections B and C only)
- **materials recovered or purified in new ways**
- **materials removed** and **removing agents** (only since 1977, unless they were the only codeable chemical in the invention)
- **activities, properties, and uses**
- **components of compositions**, if essential to the invention
- **chemical formulations** and **apparatuses**
- **novel catalysts** since 1970
- **novel starting materials**

Non-essential components of compositions in patents classified in Derwent Sections B and/or C may be structure-coded if they are the reason for the patent being classified in one or both of these sections. Known compounds that serve a standard function in a composition are not structure-coded; for example, a known propellant used in a flame retardant composition is not structure-coded. However, compounds that are not chemical-coded may be searchable with Derwent Registry numbers, as explained on page 26.

The coding of **catalysts** is dicussed in detail on page 191.

Assigning Index Terms Other Than Chemical Codes

Before a patent is assigned Chemical Codes, it is entered into the DWPI database with indexing terms that can be assigned more quickly than the Chemical Codes. When patent records are first entered in DWPI they can be found using uncontrolled keywords, International Patent Classification numbers (IPC's), bibliographic data, Manual Codes, and words in the abstract. Searches on these terms can be used instead of a Chemical Code search when non-structural information is sufficient for retrieval, or they can be used after a Chemical Code search to reduce the number of irrelevant hits. The ***Derwent Online User Guides*** discuss the many methods that can be used to access the contents of DWPI.

Dividing the Patent Into Code Subfields

As explained in the previous section, Chemical Codes assigned to different chemical aspects of a single patent are often listed in different database subfields, sometimes referred to as ***card records***. (The terminology *card record* is borrowed from a former era when Chemical Codes were represented by holes punched into different positions on a series of data cards.)

The figure below shows Chemical Codes from one of the more complex DWPI records.

```
Chemical Fragment Codes (M1):
     *01* V902 D601 D602 F211 F431 F521 F522 G221 G299 G100 G040 M150 M532
     M531 G530 K442 L140 L250 H121 H181 H182 H183 J171 J172 J173 J371 J321
     J341 H401 H441 H481 H498 J581 H321 H341 H581 H598 H582 H583 H584 H589
     H599 H601 H608 H609 H685 H602 H603 H604 M620 H721 H711 H722 H723 M240
     M232 M233 M331 M333 M640 M650 M510 M511 M520 P445 P447 M521 M522 P442
     M530 M540 P517 P446 M541 M710 R002 M423 M903
     *02* V902 D601 D602 F211 F431 F521 F522 G221 G299 G100 G040 M150 M532
     M531 G530 K442 L140 L250 H121 H181 H182 H183 ...etc.
Chemical Fragment Codes (M2):
     *03* J6 J5 M121 M111 M122 M112 M123 M113 M126 M116 M129 M131 M132 M147
     M282 M283 M210 M220 M231 M232 M233 M240 M270 M281 M311 M312 M313 M314
     M315...
     *04* A546 C810 M411 M416 M730 M903 Q421
```

This patent record has four different code subfields, delineated by the numbers *01* through *04*. Each of the first three subfields represents a distinct chemical invention disclosed or claimed in the patent; the codes listed correspond to different structural features specified by a variable Markush structure. Subfield *04* refers to a catalyst (Q421 is defined as "Catalyst"). Because the patent corresponding to this record was published after 1970 and involves a natural product and a synthetic pharmaceutical compound, the subheadings M1 and M2 are assigned. When search terms are combined in a search strategy with the LINK operator, the search must find all matching codes in a single **subfield**, i.e. *01*, *02*, *03*, or *04*, to generate a "hit".

The rules that determine which compounds should be coded in the same subfield, i.e. which compounds should be **overcoded in the same code subfield**, and which ones should be separated, have developed over time. The development of these rules is decribed in the following paragraphs.

From 1963 Forward

1. **Steroids** and **non-steroids** are not overcoded. (However, all steroids were coded together until 1981. From 1981 forward, individual components of steroidal compositions have been coded in separate code subfields, as have steroid starting materials, intermediates, and end products.)

From 1970 Forward

1. Natural products whose structures are known, and their derivatives of natural or synthetic origin, are coded once in a subfield under the subheading M1, and in a second subfield under the subheading M2.

2. Dye precursors (see Part W:) such as coupling components, colour couplers, oxidation bases, leuco bases, photo-, thermo-, piezo- or halo-chromic compounds are coded once in a subfield under the subheading M3, and in a second subfield under the subheading M4.

3. Different essential ingredients in compositions are coded in separate code subfields.

4. Compounds comprising different facets of an invention (e.g. purifying agent and compound purified) are coded in separate code subfields.

From 1979 forward

Specific **overcoding rules** were first introduced in 1979, with the aim of reducing the number of irrelevant hits in search results caused by the overcoding of large numbers of alternative compounds from Markush formulae in a single code subfield. Although the application of these rules increases the relevance of search results, they do not affect the search logic used for retrieval.

1. Compounds coded under different subheadings are never overcoded. (Subheadings are discussed on page 23.)

2. Compounds with "polyvalent chain absent" (M320) should not be overcoded with compounds having "polyvalent chain present" (M321, M322 or M323).

3. Compounds that have no functional groups coded in Parts H: to L: should not be coded in the same code subfield as compounds that do have functional groups coded in Parts H: to L:. In addition, compounds that have functional groups coded in Parts H: to L: should always be assigned at least one negation code.

4. Compounds with different Basic Group codes (M411 through M416) should not be overcoded.

5. Since 1981, catalysts have been coded in a separate subfield from other compounds in the same patent, and they are always coded Q421 ("Catalyst"). The coding of catalysts is discussed on page 191.

Assigning Chemical Codes

The following topics are briefly discussed from the coding point of view. They are discussed in more detail from the searching point of view in the section entitled "Searching with the Chemical Codes", on page 15.

Subheadings

Chemical Codes are listed under one or more subheadings, according to the subject matter of the coded compounds.

Time Range Code

The time range code assigned to each code subfield indicates the time range during which the patent was added to DWPI.

Chemical Codes

Chemical Codes are assigned to patents according to the same rules used for searching. Indexers tend to include extra codes, making it more likely that relevant patents will be retrieved. Because of the nature of patent searches, the choice is often made to err on the side of comprehensiveness rather than on the side of greater relevance of search results.

Negation Codes

Negation codes are assigned to patent records in a manner opposite to their use in search strategies. Negation codes are called **essential codes** when viewed from the perspective of the indexer, because the indexer includes a particular negation/essential code whenever all compounds coded in a single coding subfield have the corresponding functional group, ring linkage, or natural product designated by that code. Groups that are essential to the invention, but which are undesirable from the perspective of the user, can be negated.

Markush Coding

Derwent World Patents Index Markush is a new type of patent database created for use with Markush DARC software. Although chemical indexing and Markush coding are being performed concurrently at the time of this writing, chemical coding will probably be discontinued in the future. However, because many patent documents are valid for 20 years (longer with extensions), and because Markush indexing didn't begin until 1987, the Chemical Codes will remain essential to comprehensive patent searches for many years to come.

5 Terms and Symbols Used in this Manual

5 Terms and symbols used in this user guide

The following notes explain terms and symbols used throughout the user guide. Although an understanding of the conventions listed below are essential to a correct interpretation of the coding notes and definitions, the new user may want to skip this section and come back to it after becoming familiar with the Chemical Codes and this user guide.

1. One-character **element symbols** are used as a chemical shorthand for groups of atoms that are coded similarly. This shorthand makes code definitions much shorter than they would be if all variations of a coded fragment were specifically mentioned. The symbols that remain consistent throughout the manual are the following:

T	=	O, S, Se, Te
U	=	O, S, Se, Te or N
V	=	Heteroatom other than O, S, Se, Te or N
W	=	O, S, Se, Te, N, C or halogen in Parts: H: through L:
	=	Any heteroatom in Parts other than H: through L:
X	=	Halogen (F, Cl, Br, I)
Y	=	O, S, Se, Te, N or halogen
Z	=	B, Si, P or As

Some chapters define additional element symbols or variations of the symbols given above. The notes in each chapter should be consulted to determine the meaning of the element symbols used in that chapter.

2. Some codes and symbols appear to be the same or similar, but have completely different meanings. For example, ***M1*** is a subheading used for natural products and polymers. ***M1*** is also a negation code, stating that every structure coded in a particular code subfield has a ring to ring linkage. ***M1:*** is a truncated code used to represent the Set of codes from M111 to M150. The meaning of these codes and symbols can be easily determined from the written context in which they are used, but the user should be aware that the possibility for confusing such codes exists.

3. The phrases “additionally coded” and “also coded” mean that two different code Sets or Subsets are used, the one where this phrase appears and the one to which the user is referred.

4. The phrases “as applicable” and “if applicable” mean that the codes to which the user is referred can potentially be used, but the definitions and rules for those codes must be consulted to determine if they actually fit the structure.

5. The punctuation marks “;” and “,” are used to separate equally applicable code definitions. For example, the code definition “P448 Tranquiliser; anxiolytic” means that the code P448 is equally applicable to a compound that is a tranquiliser, a compound that is an anxiolytic, or a compound that is both a tranquiliser and an anxiolytic.

6. Major changes occured in the chemical coding system in **1981/Derwent Week 27**. For the sake of brevity, that point in time is referred to throughout this manual as **1981**. Thus, a coding note that begins "From 1981 forward, ..." should be more specifically interpreted as "From 1981/Derwent Week 27 forward, ...".

7. The term ***other*** in a code definition covers all relevant, specific possibilities other than those that have been defined by the immediately preceding codes. For example, in the Set of codes shown below, the code V570 is assigned to patent documents that disclose specific types of microorganisms that are not covered by the codes V510 through B560.

V500	Microorganism - general
V510	Gibberellin
V520	Growth stimulant for microorganisms; sideramines
V530	Microbial growth stimulant for animals (e.g. lactobacillus)
V540	Bacteria
V550	Fungi; algae (e.g. Streptomyces; yeast)
V560	Virus
V570	Other specific microorganism

Thus, codes should be read in the order that they are listed for correct interpretation of their definitions.

6 Elements Present

6 Elements present

Introduction

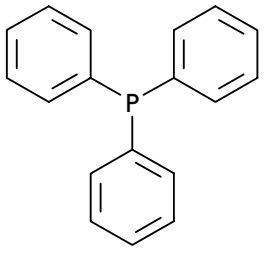
Chemical codes in Parts A:, B:, and C: are used to identify and describe individual elements present in specific types of chemical compounds. The types of compounds which are described by these codes include:

- inorganic compounds
- inorganic ions and ligands that are bonded to organic and/or inorganic groups
- organic compounds containing phosphorus, arsenic, boron, silicon, selenium and tellurium
- organometallic compounds
- free elements

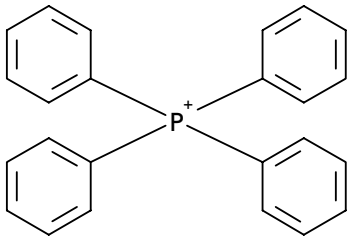
Organic compounds other than the types listed above are not usually assigned codes in Parts A:, B:, or C:. Instead, most elements in organic compounds are instead coded in terms of the rings, functional groups, and carbon chains in which they are found. In a few specific cases, however, Part C: codes are used to further identify functional group atoms coded in Parts K: and L:. These special uses of Part C: codes are listed in the notes preceding the Part C: code definitions.

Notes on Codes in Parts A:, B:, and C:

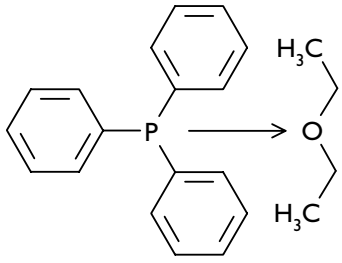
1. In the chemical coding system, the ***valency*** of an atom is defined as the number of bonds in which it participates. Therefore, the valency coded for an atom may be non-standard. When determining valencies, a double bond counts as two bonds, a coordination bond also counts as two bonds, and a triple bond counts as three bonds.



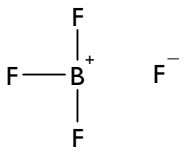
P is 3-valent (B813)



P is 4-valent (B814)



P is 5-valent (B815)



B is 4-valent (B809)

2. Ions or ligands coded in Parts A:, B:, and/or C: may be bonded to organic groups that are coded in Parts other than A:, B:, and C:, e.g. when an inorganic cation is bonded to an organic anion. The organic groups of such compounds are coded independently from each other and from the inorganic groups(s) to which they're bonded, using all applicable code Parts. (The following section, "Ions and Ligands", describes the coding of ionic and coordinated compounds.)

3. Carbon-containing compounds that are considered inorganic for coding purposes include: CO, CO_2, COS, CS_2, HCN, HCNO, HCNS, COX_2, and CSX_2, where X are the same or different halogens; metal carbonyls, cyanogen, thiocyanogen, cyanamide, cyanogen halides, thiocyanogen halides and fulminates; and metal salts of carbonic acid, thiocarbonic acid, carbamic acid, and thiocarbamic acid. Carbon oxides not listed above are coded as organic compounds. Cyanamide and metal salts of carbamic acid and thiocarbamic acid have in some cases been coded as organic compounds. Urea is always organic.

4. **Hydrates** are coded in their anhydrous form. If the water of hydration is an important part of the invention, it is coded as C101, C108, and C550 and if the material is organic all the codes for the elements absent in water are also coded (C80: codes). On the other hand, if the water of hydration is not an important part of the invention, it is not coded. For this reason, a search on the chemical codes for water will not usually retrieve all hydrates of inorganic compounds.

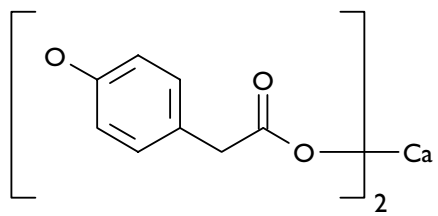
Ions and Ligands

When selecting codes for an **ionic or coordinated compound**, the compound is first divided into its component ions or ligands. Each ion or ligand is considered separately when codes are being selected, even though the codes for all ions or ligands have usually been **overcoded in a single code subfield**. (Code subfields are discussed on page 28.) The ions or ligands are searched as if only one formula unit of each is present, even when the molecular formula indicates that multiple units of that ion or ligand are present in the compound. This rule affects the codes used for counting the various elements, functional groups, and ring systems present in the compound, as illustrated in the first example below.

Although the ions or ligands of a compound are coded individually, some codes are assigned to the compound as a whole. Thus, to receive one or more of the **Elements Absent** codes from Subset C80:, the element(s) must be absent from all **inorganic** ions or ligands in the compound. Negation codes in Parts H: through L:, the carbon chain codes M280 and M320, and Basic Group codes in Subsets M41: and M42: also apply collectively to all organic ions and ligands of the compound.

Inorganic ions and ligands bonded to **organic** groups are usually only coded when important to the invention. Prior to 1981, metal salts of organic acids received the Basic Group code of the free acid from Subset M4: when the metal was unimportant, but M411 when the metal was important. From 1981 forward, however, the metal in such compounds is coded in Part A: whether it's significant or not, and the Basic Group is always M411. An **exception** to this rule is that alkali and alkaline earth metals, because they are so common, are still only coded if they are a limiting feature of the invention; otherwise they are not coded and the compound receives the Basic Group code of the free acid. **All** metal salts of organic acids have been assigned the codes A960 ("Metal bonded to O, S, Se, Te, or N of an organic ion or ligand") and C710 ("Inorganic cation in an organic compound") since 1963.

Examples

a 

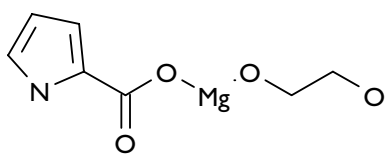
Codes include:

H401, H441	**One** -OH group
J011, J171	**One** -COOH group
M630	Metal salt
M321, M371	**One** carbon chain
M531	**One** aromatic ring
A960	Metal to O,S of organic ligand
C710	Inorganic cation

b $Al(OH)(NO_3)(CH_3COO)$

Search as:

Al^{3+}
OH^-
NO_3^-
CH_3COOH

c 

Search for 2-carboxypyrrole and ethylene glycol, in each case as if the other ligand were absent, e.g. code M620 (saturated aliphatic compound) can be used for the glycol.

Complexes

For comprehensive retrieval of complexes it is necessary to perform two searches: one search for the complex as a single compound and a second search for the component compounds (i.e. as if the complex were a simple mixture). For example, a patent concerning the treatment of a mixture of MgO and FeO would probably be coded as a mixture of the two oxides, even though the product may actually be magnesium ferrite. However, if magnesium ferrite were mentioned in the patent, it would be coded as a single compound instead of as MgO and FeO.

To retrieve both types of patent, MgO and FeO are first searched as individual compounds, and the results are 'AND'ed together. The AND operator is used in this case instead of LINK, because the codes for the individual compounds of the mixture may be listed in separate code subfields. (Code subfields are explained on page 28). Next, a search is conducted on magnesium ferrite. The results of the 'AND'ed searches are 'OR'ed with the results of the latter search, so that the final results contain patents on MgO and FeO mixtures as well as patents on magnesium ferrite.

Part A: Metals

Main Headings

A1: Alkali metals
A2: Alkaline earth metals
A3: Metals in Periodic Groups IIIA to VA
A4: First Transition Series
A5: Second Transition Series
A6: Third Transition Series, La
A7: Lanthanides
A8: Actinides, Po, At, Fr, Ra, Ac
A9: Bonds to metal atoms

Introduction

Part A: codes cover:

- all metals except As (As is coded in Part B:)
- Ge, Sb
- elements with an atomic number greater than 54, except for Rn

The first digit of most Part A: codes indicates the metal group to which the element belongs, and the last two digits (up to 99) are the atomic number.

Notes on Part A: Codes

1. **Generic codes**, i.e. those ending in "00", are assigned when patents disclose a generic group of metals, rather than specifying the particular metal(s) involved in the invention. Generic codes alone are not sufficient to retrieve all references to a certain group of metals. To retrieve patents referring to all metals within a particular Set, the codes for all of the individual metals are 'OR'ed together with the generic "00" code at the beginning of the Set. The same results can also be accomplished by searching on the applicable Set code truncated after the second character. For example, to retrieve all references to metals in the First Transition Series, whether disclosed individually or as a group, include the truncated search term "A4:".

2. Additional descriptors for compounds receiving Part A: codes are found in Part C:. For example, C710 ("Inorganic cation in an organic compound") is assigned to metal salts of organic acids. The use of Part C: codes for metal-containing compounds is described in the notes accompanying Part C: codes, beginning on page 59.

Guidelines for Coding Compounds Containing Metals

1. First divide the compound into its component ions or ligands. All bonds to the metal atom(s) are deleted **except** those from **organic carbon** atoms. (Note 3 on page 36 lists carbon compounds that are NOT considered organic.)

2. Each metal in the compound receives the Part A: code that identifies it (except for arsenic, which is coded in Part B:). Note that there are no **Poly** codes in Part A: to indicate the presence of more than one formula unit of a particular metal in a compound.

3. Codes in Subsets A91: and A92: and the code A930 are used to describe various types of **bonds between metal atoms and organic C atoms**.

4. The rules for coding a **metal bonded to two C atoms in a ring** are discussed in the section "Ring Heteroatoms Other Than O, S, and N", beginning on page 72. The metal receives the Part A: code that identifies it, as well as the code A922 for "Metal in a ring".

5. A bond between a metal atom and an atom other than organic C is assumed to be either ionic or coordinate; the group bonded to the metal is considered a separate ion or ligand for coding purposes. Ionic and coordinate bonds, including bonds between metal atoms and inorganic C atoms, are assigned one of the following codes:

A940	Metal bonded to an inorganic anion
A950	Metal bonded to an inorganic neutral ligand
A960	Metal bonded to O, S, or N of an organic group
A970	Metal bonded to an atom other than O, S, or N of an organic group
A980	Metal is part of an inorganic anion

6. **Metal carbides** are only considered organic when the C is bonded to one or more H atoms.

7. **Ions** or **ligands** bonded to metal atoms are coded individually in the code Parts that apply to them, with protons replacing metal atom(s) in the case of ions. The coding of ions and ligands is discussed on page 37 in the section entitled "Ions and Ligands." The metal and all ions or ligands are overcoded in a single code subfield. (Code subfields are discussed on page 28.) However, two or more complex organic fragments in a single compound, e.g. chromium complexed with two different azo dyes, may be coded in separate code subfields.

8. **Functional groups** in organic ions or ligands that are bonded to a metal atom via C are searched in Parts H: through L: in one of two ways: as if the metal were replaced by H when the metal is bonded to only one C atom, or as if the metal were replaced by C when the metal is bonded to more than one C atom.

9. If a compound can exist as either a metal-C or a metal-heteroatom **tautomer**, only the latter is coded. The metal-tautomer bond is assigned the appropriate code from among those listed in Note 5 above (i.e. A940 through A980).

10. Examples illustrating the coding of metal compounds are given on pages 43-45.

Subheadings for Part A: codes

Unless specified, codes are applicable to M0 from 1963; M1 to M4 from 1970.

A1: ALKALI METALS

A100	General disclosure of alkali metals(1963-1969, 1981)	
A103	Lithium	
A111	Sodium	
A119	Potassium	
A137	Rubidium	
A155	Cesium	

Notes on A1: Codes

1. The code A100 is only valid from 1963 through 1969, and from 1981 to the present.
2. For complete retrieval of potassium from 1963 through 1969, search (**A119 OR A100**).

A2: ALKALINE EARTH METALS

A200	General disclosure of alkaline earth metals	(1981)
A204	Beryllium	
A212	Magnesium	
A220	Calcium	
A238	Strontium	
A256	Barium	

A3: METALS IN PERIODIC GROUPS III-A TO V-A (EXCEPT ARSENIC)

A300	General disclosure of Group III-A, IV-A,or V-A metals	(1981)
A313	Aluminium	
A331	Gallium	
A332	Germanium	
A349	Indium	
A350	Tin	
A351	Antimony	
A381	Thallium	
A382	Lead	
A383	Bismuth	

A4: METALS OF THE FIRST TRANSITION SERIES

A400	General disclosure of First Transition Series metals	
A421	Scandium	(1981)
A422	Titanium	(1970)
A423	Vanadium	(1970)
A424	Chromium	(1970)
A425	Manganese	(1970)
A426	Iron	
A427	Cobalt	
A428	Nickel	(1970)
A429	Copper	
A430	Zinc	(1970)

Notes on A4 Codes:

1 A421 only applies from 1981; prior to 1981 search A400 for scandium.

2 A422, A423, A424, A425, A428, A430 only apply from 1970; prior to 1970 search A400 for these metals.

A5: METALS OF THE SECOND TRANSITION SERIES

A500	General disclosure of Second Transition Series metals	(1981)
A539	Yttrium	(1981)
A540	Zirconium	(1981)
A541	Niobium	(1981)
A542	Molybdenum	(1981)
A543	Technetium	(1981)
A544	Ruthenium	(1981)
A545	Rhodium	(1981)
A546	Palladium	(1981)
A547	Silver	(1981)
A548	Cadmium	(1981)

Notes on A5: codes

1. A539 to A545 are only applicable from 1981; prior to 1981, search A500 for Y, Zr, Nb, Mo, Tc, Ru, Rh.

2. A546, A547, and A548 are only applicable from 1970; prior to 1970, search A500 for palladium, silver, or cadmium.

A6: METALS OF THE THIRD TRANSITION SERIES; LA

A600	General disclosure of the Third Transition Series metals	
A657	Lanthanum	(1981)
A672	Hafnium	(1981)
A673	Tantalum	(1981)
A674	Tungsten	(1981)
A675	Rhenium	(1981)
A676	Osmium	(1981)
A677	Iridium	(1981)
A678	Platinum	(1970)
A679	Gold	(1981)
A680	Mercury	(1981)

Notes on A6: codes

1. The codes A657 through A677 and the code A679 are applicable from 1981 forward; prior to 1981, search A600.
2. A678 is applicable from 1970 forward; prior to 1970, search A600 for platinum.

A7: LANTHANIDES

A700	General disclosure of lanthanides	(1981)
A758	Cerium	(1981)
A759	Praseodymium	(1981)
A760	Neodymium	(1981)
A761	Promethium	(1981)
A762	Samarium	(1981)
A763	Europium	(1981)
A764	Gadolinium	(1981)
A765	Terbium	(1981)
A766	Dysprosium	(1981)
A769	Thulium	(1981)
A770	Ytterbium	(1981)
A771	Lutetium	(1981)

Note The codes A758 to A771 are applicable from 1981 forward; prior to 1981 search A700 for the metals identified by these codes.

A8: ACTINIDES, PO, AT, FR, RA, AC (INCLUDING TRANS-URANICS)

A800	General disclosure of Po, At, Fr, Ra, Ac & the actinides	(1970)
A884	Polonium	(1981)
A885	Astatine	(1981)
A887	Francium	(1981)
A888	Radium	(1981)
A889	Actinium	(1981)
A890	Thorium	(1981)
A891	Protoactinium	(1981)
A892	Uranium	(1981)
A893	Neptunium	(1981)
A894	Plutonium	(1981)
A895	Americium	(1981)
A896	Curium	(1981)
A897	Berkelium	(1981)
A898	Californium	(1981)
A899	Trans-californium	(1981)

Notes on A8: codes

1. A800 is applicable from 1970 forward; prior to 1970, search C811 ("radioactive element") in place of A800.
2. The codes A884 through A899 are applicable from 1981 forward. The elements identified by these codes are searched using the code C811 ("radioactive element") prior to 1970, and using the code A800 from 1970 to 1981.

A9: BONDS TO METAL ATOMS

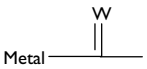
A910 One bond from metal to organic C

A92: Two or more bonds from metal to organic C

A921 Each metal atom is bonded to one C atom only

A922 Metal bonded to two C atoms in the same ring (1970)

A923 Other cases of two or more bonds from metal to organic C

A930 

(W is any heteroatom) (1981)

Note Structures coded A930 are additionally coded A910 or in Subset A92:, depending on the number and type of C-metal bonds present. In addition, the -C(=W)- fragment receives all applicable codes from Parts H: through L:.

A940: Metal bonded to an inorganic anion (1970)

Note If a metal is bonded to B, Si, P, or As, search B770 in addition to any applicable codes in Set A94: or Set A95:.

A950: Metal bonded to an inorganic neutral ligand (1970)

Notes on A950 code:

1 Used for coordinated neutral inorganic ligands such as CO, NO, PH_3, NH_3

2 If a metal is bonded to B, Si, P, or As, search B770 in addition to any applicable codes in Set A94: or Set A95:.

A960 Metal bonded to O, S, Se, Te, or N of an organic ion or ligand (1970)

A970 Metal bonded to an organic ion or ligand, excluding the cases covered by Subsets A91: to A96:. (1970)

A980 Metal is a member of an inorganic anion (1970)

A990 Metal bonded to metal (1981)

Examples Using Part A: Codes

Note Only the codes relevant to the coding concepts presented in Part A: are included in the following examples.

Ions and ligands bonded to metals:

a $NaCl$
A111, A940

b $Pt(PH_3)_4$
A678, A950, B770

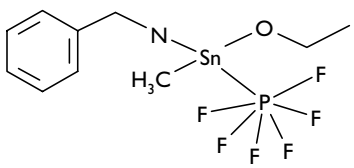
c $SnCl_4$
A350, A940

d $K_4Fe(CN)_6$
A119, A426, A940, A980

e $Fe(CO)_4Cl_2$
A426, A940, A950

f $Pt(PPh_3)_4$
A678, A970, B770

g $K_2Cr_2O_7$
A119, A424, A940, A980

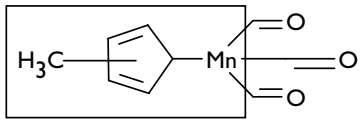
h 

A350, A910, A940, A960, B770
Search also $PhCH_2NH_2$, CH_3CH_2OH, PF_6^-, CH_3Sn

Metal bonded to C:

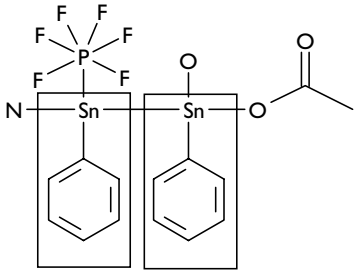
Note The organometallic fragments in the following examples are outlined with broken lines. All codes given for a single compound are overcoded in a single code subfield, even where different lists are given for different fragments of the structure.

a



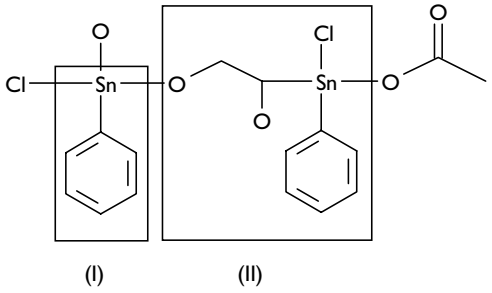
A425, A910, A950

b



(I.) A350, A910, A940
(II.) A350, A910, A940, A960, A990

c

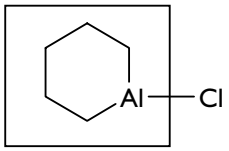


(I.) A350, A910, A940, A960
(II.) A350, A923, A940, A960

d $(C_2H_5OCO)_2CH^-\ Na^+$

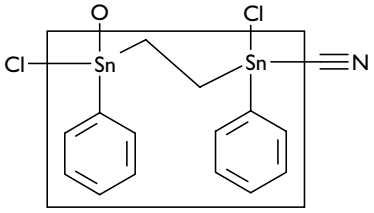
A111, A960 (but not A910)

e



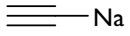
A313, A922, A940

f



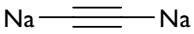
A350, A923, A940

g



A111, A910

h

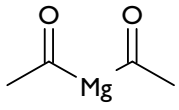


A111, A940 (but not A921) (i.e. Na acetylide treated as inorganic salt as it contains only Na ions and H-free acetylene ions)

i $(CH_3)_2Ca$

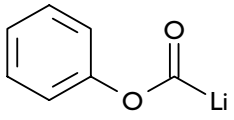
A220, A923

j



A212, A923, A930, J582

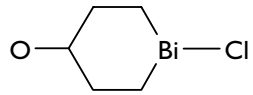
k



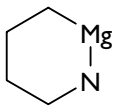
A103, A910, A930, J241

Metal as ring heteroatom

Note The selection of codes for the rings in the following examples is discussed in Part A: and in the section "Ring Heteroatoms Other Than O, S, and N", on page 72.

a 

A212, A910, A960, M630
Searched as mono-substituted heterocyclic, hydroxy cyclohexane,

b 

A383, A922, A940, G563, H401, H421
Searched as Mg $CH_2CH_2CH_2NH_2$

Part B: Noble Gases, B, Si, P, As, Se, Te

Main Headings

B0: Identification of noble gas(es) present
Bl: Noble gas, B, Si, P, As, Se, Te as free element or in an inorganic compound, ion, or ligand
B2: Noble gas, B, Si, P, As, Se, Te in an inorganic anion of an organic compound
B3: Noble gas, B, Si, P, As, Se, Te in an inorganic cation of an organic compound
B4: Noble gas, B, Si, P, As, Se, Te in an aliphatic compound
B5: Noble gas, B, Si, P, As, Se, Te in a carbocyclic compound
B6: Noble gas, B, Si, P, As, Se, Te in a heterocyclic compound
B7: Bonds to B, Si, P, As
B8: Valency and number of B, Si, P, and As atoms; B, Si, P and As in a ring

Introduction

Part B: covers the noble gases, B, Si, P, As, Se, and Te, both as free elements and in compounds.

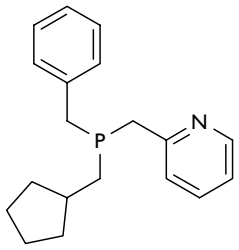
Notes on Codes in Part B:

1. The phrase "Part B: elements" is used to denote the noble gases, B, Si, P, As, Se, and Te. The elements B, Si, P, and As are referred to collectively as "Z" atoms.

2. The particular code used to identify a Part B: element depends on the type of compound in which the element is found. Only one code is used to identify a Part B: element, chosen according to the following priority order:

   B1 = B2 = B3 > B6 > B5 > B4

   Thus, the P atom in the compound shown below receives the code B615, but not the codes B515 or B415.



3. In addition to codes that identify Part B: elements, Part B: codes indicate the type of compound in which a Part B: element is found, elements bonded to Z atoms, functional groups bonded to Z atoms, valency of Z atoms, total number of Z atoms, and Z atoms as ring members.

4. The last two digits of codes in Sets B0: through B6: correspond to the atomic number of the element described. Codes ending in "00" are used in Sets B1: through B6: to indicate the noble gases.

5. **Compounds with Se or Te** receive the applicable Se or Te element code from Part B:, plus any codes from Parts A: through L: that would apply to the compound if the Se and Te atoms were replaced by S. Exceptions to this rule are the codes in Set C5: and the code C801, which are not assigned to structures containing Part B: elements.

6. Rules for coding **Part B: elements in rings** are given in the section "Ring Heteroatoms Other Than O, S, and N", on page 72.

Symbols used for chemical elements in this section:

T	=	O, S, Se, Te	Y	=	O, S, Se, Te, N, or halogen
U	=	O, S, Se, Te, N	Z	=	B, Si, P, As
X	=	halogen			

Guidelines for Coding Functional Groups Bonded to B, Si, P, or As

Codes in Subset B79: are assigned to functional groups bonded to B, Si, P, and As (i.e. "Z" atoms). As explained in the guidelines given below, some of the functional groups bonded to Z atoms are additionally coded in Parts H: through L:. Examples illustrating the use of B79: codes are presented later in this chapter, beginning on page 53.

1. Codes in Subset B79: are assigned to every structural fragment that matches their code definitions. However, codes in Parts H: through L: that appear applicable are in many cases not assigned, being superseded by the codes in Subset B79:. An O, S, N, or halogen atom bonded to one or more Z atoms is only coded as a functional group in Parts H: through L: if the O, S, N, or halogen atom is additionally bonded to U, C=U, C≡N or halogen.

Exceptions:
- a -Z-N=C- is coded both B731 and L355
- b acetals, ketals, and their sulphur analogues that are bonded to Z atoms are coded as acetals, etc., rather than as ethers

2. Z atoms bonded to halogens are searched in Subset B75:. From 1978 forward, the specific halogen bonded to the Z atom has been identified by a halogen code from Set C0:.

3. The functional groups -Z-C(=O)-OH, -Z-C(=O)-X, -Z-C(=O)-O-, and their sulphur analogues are searched from 1963 forward with the code B791. The functional group -Z-O-C(=O)- and its sulphur analogues are searched before 1981 with the code B790. From 1978 forward, the halogen in the functional group Z-C(=T)-X can be identified with a code from Set C0:.

Note The functional groups discussed in this note are NOT additionally searched as carboxylic acids (Set J1:), esters (Set J2:), or acyl halides (L512).

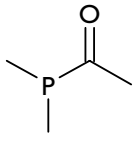
---

4. If a Z atom is bonded to a carboxylic amide group, the Z atom is ignored and the amide is searched according to the other groups linked to it. If the amide is bonded to three Z atoms, the amide is searched in Set J3: as if the Z atoms were replaced by C atoms. The amide fragment of the group Z-C(=O)-NH2 is searched as an aliphatic amide.

5. The functional groups listed below are assigned all applicable codes from Subset B79:. They are additionally coded in Part J: or Part L: as indicated, the choice of code depending on the types of groups to which the functional group is bonded. By convention, the Part J: or Part L: code for each of the groups is chosen as if the Z atom(s) were replaced by C.

Z-C(=T)-C-	coded as oxo in Set J5:
Z-C(=O)-H, Z-C(=S)-H	coded in Set J4: as aldehyde or thioaldehyde, respectively
Z-C≡N	coded as -C≡N in Subset L14:
$(Z)_2$-N-C(=T)-Z	coded as amide in Set J3: (Z atoms may be the same or different)

For example, the following structure is coded B792 and J581.



(Other applicable Part B: codes include: B415, B743, B720, B813, B831.)

If a Z atom in one of the functional groups shown above is a ring member, the functional group is considered to be bonded to a heterocyclic ring, even if there are no other heteroatoms in the ring with Z.

Subheadings Applicable to Part B: Codes

Unless specified, codes are applicable to M0 from 1963; M1 to M6 from 1970

B0: IDENTIFICATION OF NOBLE GAS(ES) PRESENT

B000	Non-specific disclosure of noble gases; generic code for searching all noble gases before 1981	
B002	Helium	(1981)
B010	Neon	(1981)
B018	Argon	(1981)
B036	Krypton	(1981)
B054	Xenon	(1981)
B086	Radon	

B1: NOBLE GAS, B, SI, P, AS, SE, TE AS FREE ELEMENT, OR IN AN INORGANIC COMPOUND, ION, OR LIGAND

B100	Noble gas
B105	Boron
B114	Silicon
B115	Phosphorus
B133	Arsenic
B134	Selenium
B152	Tellurium

B2: NOBLE GAS, B, SI, P, AS, SE, TE IN AN INORGANIC ANION OF AN ORGANIC COMPOUND

B200	Noble gas
B205	Boron
B214	Silicon
B215	Phosphorus
B233	Arsenic
B234	Selenium
B252	Tellurium

B3: NOBLE GAS, B, SI, P, AS, SE, TE IN AN INORGANIC CATION OF AN ORGANIC COMPOUND

B300	Noble gas
B305	Boron
B314	Silicon
B315	Phosphorus
B333	Arsenic
B334	Selenium
B352	Tellurium

B4: NOBLE GAS, B, SI, P, AS, SE, TE IN AN ALIPHATIC COMPOUND

B400	Noble gas
B405	Boron
B414	Silicon
B415	Phosphorus
B433	Arsenic
B434	Selenium
B452	Tellurium

B5: NOBLE GAS, B, SI, P, AS, SE, TE IN A CARBOCYCLIC COMPOUND

B500	Noble gas
B505	Boron
B514	Silicon
B515	Phosphorus
B533	Arsenic
B534	Selenium
B552	Tellurium

B6: NOBLE GAS, B, SI, P, AS, SE, TE IN A HETEROCYCLIC COMPOUND

B600	Noble gas
B605	Boron
B614	Silicon
B615	Phosphorus
B633	Arsenic
B634	Selenium
B652	Tellurium

Note Heterocyclic inorganic compounds are searched in Set B1:, not in Set B6:.

B7: BONDS TO B, SI, P, AS ("Z" ATOMS)

Notes on B7: codes

1. In Set B7:, the number of each type of bond is totalled for all Z atoms present in the compound, ion, or ligand.
2. Acids are coded in their non-ionized form; therefore HBF_4 and H_2SiF_6 are coded B760.
3. Unlike acids, metal salts are coded in their ionized forms; therefore, B770 is not applicable to $NaBF_4$ or Na_2SiF_6. B770 is coded for metal phosphides and for Z atoms coordinated in a neutral molecule to a metal, e.g. $(Ph_3P)_3Pt$.

B70: Total number of Z=O and Z=S bonds

B701	One
B702	Two or more

B71: Total number of Z-O and Z-S bonds

B711	One
B712	Two
B713	Three or more

B72: Total number of Z-S and Z=S bonds

B720	None	(1972)
B721	One	(1972)
B722	Two	(1972)
B723	Three or more	(1972)

Note B720 is only searched if there is a Z atom present in the structure. B72: codes not used for natural products (subs M1) before 1981.

B73: Total number of Z to N bonds

B730	One or more Z=N or Z≡N bonds
B731	One Z-N bond
B732	Two or more Z-N bonds

B74: Total number of Z to C bonds

B740	One or more Z=C or Z≡C bonds
B741	One Z-C bond
B742	Two Z-C bonds
B743	Three Z-C bonds
B744	Four or more Z-C bonds

B75: Total number of Z-halogen bonds

B751	One
B752	Two or more

Note From 1978 forward, the specific halogen bonded to a Z atom has been coded in Set C0:.

B760: One or more of Z-H bonds

B770: One or more Z to metal bonds

B780: One or more Z to Z bonds

B79: Non-ring functional groups containing Z atoms (organic compounds only)		
B790	Anhydrides of Z-atom acids (Pre-1981 precursor code for Subset B79:)	
B791	Z-C(=U)-U; Z-C(=U)-X	(1981)
B792	Z-C(=U)-; Z-C≡N	(1981)
B793	Z-U-C=Y; Z-U=C-Y; Z-U=C=Y	(1981)
B794	Z-U-U; Z-U=U; Z=U-U	(1981)
B795	P-O-P-O-P	(1981)
B796	Z-U-Z-U-Z (excluding structures covered by B795)	(1981)
B797	P-O-P (excluding structures covered by B795 and B796)	(1981)
B798	Z-U-Z (excluding structures covered by B795, B796, B797)	(1981)
B799	Poly (for Subset B79: only)	

Note Rules for using B79: codes in conjunction with codes from Parts H: through L: are given in the section entitled "Guidelines for Coding Functional Groups Bonded to B, Si, P, and As ("Z" Atoms) on page 48.

B8: VALENCY AND NUMBER OF B, SI, P, AND AS ATOMS

Notes on B8: Codes

1. The ***valency*** of an atom is defined as the number of bonds in which it participates; therefore, the valency coded for an atom may be non-standard. When coding valencies, a double bond counts as two bonds, a coordination bond also counts as two bonds, and a triple bond counts as three bonds.

2. Below are shown valency codes for Z atoms in different kinds of compounds.

$HBCl_4$; $NaBF_4$	B809 (treated as $^-BF_4$)
HPF_6; $NaPF_6$	B819 (treated as $^-PF_6$)
$Et_2O.BBr_3$	B805 if coded as a single compound; B803 if coded as a mixture
$Ph_3P.NEt_2$	B815 if coded as a single compound; B813 if coded as a mixture
$Ph_4P^+Cl^-$	B814 (the anion is only coded if important to the invention)

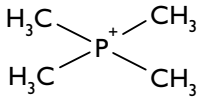
   Rules for coding complexes are given in the section entitled "Complexes", on page 38.

3. To ensure the recall of all references to P+ bonded to H, search (B813 OR B814) instead of B814 alone. Similarly, to ensure the recall of all references to As+ bonded to H, search (B823 OR B824) instead of B824 alone. However, the small number of additional valid references retrieved by the supplemental codes B813 and B823 may be outweighed by the increased number of irrelevant references.

B80: Valency of B		
B803	Trivalent	(1970)
B805	Pentavalent	(1970)
B809	Other valency	(1970)

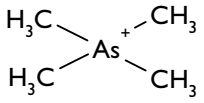
Note B809 includes electron-deficient boron hydrides.

B81: Valency of P		
B813	Trivalent	(1970)
B814	Tetravalent	
B815	Pentavalent	(1970)
B819	Other valency	(1970)



B814

B82: Valency of As		
B823	Trivalent	(1970)
B824	Tetravalent	
B825	Pentavalent	(1970)
B829	Other valency	(1970)



B824

B83: Total number of Z atoms present in the compound, ion, or ligand		
B831	One	(1970)
B832	Two of the same Z atoms	
B833	Three or more of the same Z atoms	
B834	Two or more different Z atoms	

Note In the case of ionic and coordinated compounds, the number of Z atoms is totalled for each ion or ligand, not for the compound as a whole.

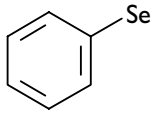
B840: Z as ring heteroatom in an organic compound (1970 /M2-M4; 1981 /M1)

Examples Using Part B: Codes

The following examples illustrate the use of Part B: codes. Codes that are applicable to the structures but unrelated to Part B: coding rules are not included.

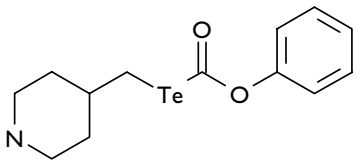
Organic Se and Te compounds

a



B534, H494 ("aromatic -SH") (H494 used for Se as well as S; see Note 5 on page 48)

b



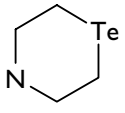
B652, L471 ("S-C(=O)-O")

c

CH_3-Se-CH_2CH_3

B434, M620 ("aliphatic thio ether")

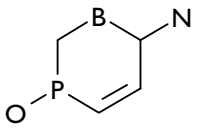
d



B652, F740 ("thiazine")

Note The Ring Index Number (RIN) for 1,4-tellurazine (i.e. the ring shown above) is assigned, not the RIN of 1,4-thiazine (i.e. the ring coded).

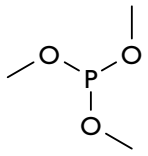
Z atom as a ring member

a 

B605, B615, B711, B720, B744, B760, B803, B813, B834, B840, G562 ("cyclohexene"), H100, H121 ("amino group bonded to heterocyclic ring")

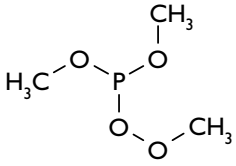
Note No -OH codes are assigned to this ring, because of the rule given in Note 1 of "Guidelines for Coding Functional Groups Bonded to B, Si, P, or As" on page 48.

Z atom in functional groups

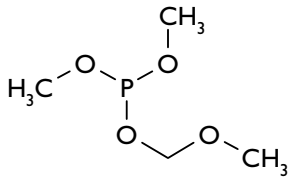
a 

B415	Phosphorus in an aliphatic compound
B713	Three or more Z-O bonds
B720	No Z to S bonds
B813	Trivalent P
B831	Total of one Z atom in the compound

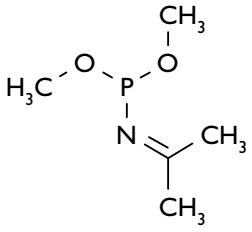
Note No codes from Parts H: through L: are used.

b 

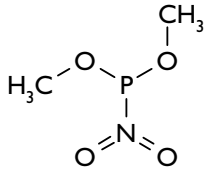
B415	Phosphorus in an aliphatic compound
B713	Three or more Z-O bonds
B720	No Z to S bonds
B794	Z-U-U
B813	Trivalent P
B831	Total of one Z atom in the compound
K930	Peroxide in an organic compound

c 

B415	Phosphorus in an aliphatic compound
B713	Three or more Z-O bonds
B720	No Z to S bonds
B813	Trivalent P
B831	Total of one Z atom in the compound
L660	O-C-O (ketal) in an organic compound

d 

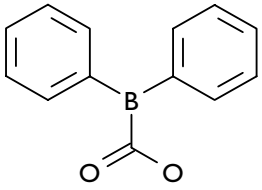
B415	P in an aliphatic compound
B712	Two Z-O bonds
B720	No Z to S bonds
B731	One Z-N bond
B813	Trivalent P
B831	Total of one Z atom in the compound
L355	-C=N-

e 

B415	Phosphorus in an aliphatic compound
B712	Two Z-O bonds
B720	No Z to S bonds
B731	One Z-N bond
B794	Z-U=U

B813	Trivalent P
B831	Total of one Z atom in the compound
H381	One aliphatic nitro group

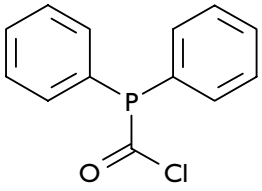
f



B505	B in a carbocyclic compound
B720	No Z to S bonds
B743	3 Z to C bonds
B791	Z-C(=U)-U
B803	Trivalent B
B831	Total of one Z atom in the compound

Note Set J1: codes are not assigned.

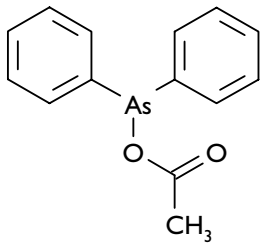
g



B515	P in a carbocyclic compound
B720	No Z to S bonds
B743	3 Z to C bonds
B791	Z-C(=U)-U
B813	Trivalent P
B831	Total of one Z atom in the compound
C017	Used to identify Cl in -C(=O)-Cl from 1978 forward (treated as parent acid prior to 1978)

Note L512 is not assigned.

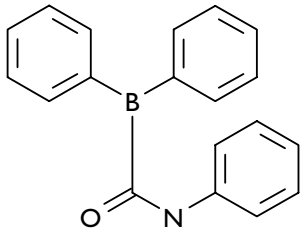
h



B533	As in a carbocyclic compound
B711	One Z to O bond
B720	No Z to S bonds
B742	Two Z to C bonds
B793	Z-U-C=U
B823	Trivalent As
B831	One Z atom

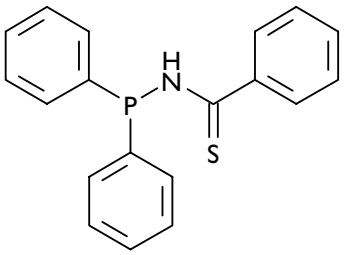
Note Set J2: codes are not assigned.

i



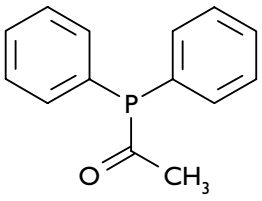
B505	B in a carbocyclic compound
B720	No Z to S bonds
B743	Three or more Z to C bonds
B791	Z-C(=U)-U
B803	Trivalent B
B831	One Z atom
J341	One amide, aromatic ring bonded to N

j



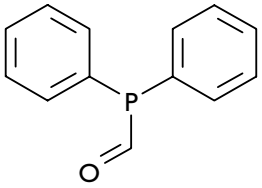
B515	P in a carbocyclic compound
B720	No Z to S bonds
B731	One Z-N bond
B742	Two Z to C bonds
B793	Z-U-C=U
B813	Trivalent P
B831	One Z atom
J331	One amide, aromatic ring bonded to -C(=S)-
J390	Thioamide

k



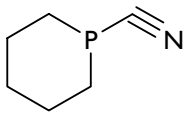
B720	No Z to S bonds
B743	Three Z to C bonds
B792	Z-C(=U)-
B813	Trivalent P
B831	One Z atom
J581	One aliphatic oxo group

l



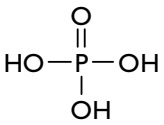
B515	P in a carbocyclic compound
B720	No Z to S bonds
B743	Three Z to C bonds
B792	Z-C(=U)-
B813	Trivalent P
B831	One Z atom
J471	One aliphatic -C(=O)-H

m



B615	P in a heterocyclic compound
B720	No Z to S bonds
B743	Three Z to C bonds
B792	Z-C≡N
B813	Trivalent P
B831	One Z atom
B840	Z as ring heteroatom in an organic compound
L142	-C≡N bonded to heterocycle
00281	RIN for phosphorin

Inorganic compounds, ions or ligands; free elements

a 

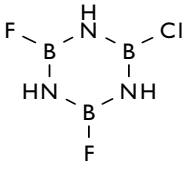
B115 P in an inorganic compound
B701 One Z=O bond
B713 Three or more Z-O bonds
B720 No Z to S bonds
B831 One Z atom
B815 Pentavalent P

b Se=Si=Se

B114 Si in an inorganic compound
B134 Se in an inorganic compound
B702 Two or more Z=S bonds
B722 Two Z=S bonds
B831 One Z atom

c He

B000 Code used for all noble gases before 1981/week 27
B002 Helium (valid from 1981/week 27 forward)
B100 Noble gas as a free element
C810 Free element in an inorganic compound

d 

B105 B in an inorganic compound (NOT heterocyclic)
B720 No Z to S bonds
B732 Two or more Z-N bonds
B752 Two or more Z-halogen bonds
B803 Trivalent B
B833 Three or more of the same Z atoms
C009 F atom present
C017 Cl atom present

Note B840 ("Z as ring heteroatom in an organic compound") is not coded.

e $(Me_4N^+)_2SeO_4^{2-}$

B234 Se in an inorganic anion

f $(Me_4N)^+ (AsH_4^+)_2 \ BO_3^{3-} AsH_4^+$

i. AsH_4^+

B333 As in an inorganic cation
B720 No Z to S bonds
B760 One or more Z-H bonds
B824 Tetravalent As^+
B831 One Z atom

Note To retrieve all references to As^+ bonded to H, search (B823 OR B824).

ii. BO_3^{3-}

B205 B in an inorganic anion
B713 Three or more Z-O bonds
B720 No Z to S bonds
B803 Trivalent B
B831 One Z atom

Part C: Halogen, H, C, N, O, and S in Inorganic Compounds; Miscellaneous Descriptors for Elements

Main Headings

C0: Identification of halogens present
C1: Presence of lower-valency halogen, H, C, N, O, or S in inorganic compounds, ions,or ligands
C2: Presence of intermediate-valency halogen or S in inorganic compounds, ions, or ligands
C3: Presence of high-valency halogen, N, or S in inorganic compounds, ions, or ligands
C4: -O-O- or -S-S- in inorganic compounds, ions, or ligands
C5: Inorganic compounds and ions consisting solely of H, C, N, O, and/or S atoms
C6: Halogen as a ring heteroatom
C7: Descriptors for inorganic compounds and ions that do not contain Part B: elements
C8: Elements not present in inorganic compounds, ions, or ligands; miscellaneous descriptors for elements

Introduction

Part C: codes cover:

- Presence and valency of halogen, H, C, N, O, and S in **inorganic** compounds, ions, and ligands
- Identification of **organic** halogens in Part K: codes, L511, L512,and L760; valid from 1978 forward
- Valency of S or N in **organic** compounds, when the S or N code from Part K: or Part L: can represent more than one valency; valid from 1970 forward
- -O-O- or -S-S- in **inorganic** compounds, ions, or ligands
- Halogen as a ring member
- Descriptors for **inorganic** ions and compounds that do NOT contain noble gases, B, Si, P, As, Se, or Te
- Elements that are NOT present in **inorganic** compounds, ions, or ligands
- Free elements, radioactive elements, and non-radioactive isotopes in **organic** and **inorganic** compounds, ions, or ligands

1. Part C: codes are not assigned to **organic** compounds, ions, or ligands, other than the **special cases** listed above.

2. Assigning Part C: codes to **halogens** is discussed in the "Notes on C0: Codes". A general discussion on coding halogens is found in the section "Coding Rules for Polymers and Halogens", on pages 18-19.

3. Using Part C: codes to specify the valencies of S and N in **organic** compounds is discussed in Note 7 on page 141, in the section "Notes on Codes in Parts K: and L:".

4. Examples using Part C: codes are given at the end of this chapter.

Subheadings Applicable to Part C: Codes

Unless otherwise specified, codes are valid from 1963, although sets C1: (excluding C116), C2:, C3:, C4:, C6: and C8 (excluding C811 and C812) are valid from 1970.

Codes in sets C1:, C2:, C3:, C4: and C6: are not valid for natural products (M1) before 1981.

C0: IDENTIFICATION OF HALOGENS PRESENT

C000	Unspecified halogen present	(1981)
C009	Fluorine	
C017	Chlorine	
C035	Bromine	
C053	Iodine	

Notes on C0: Codes

1. Set C0: codes are used to identify halogens in the following cases:

   a Halogens in **inorganic** compounds, ions, and ligands

   Exceptions:

   - only assigned after 1978 for halogens in compounds with Z atoms
   - not used for halogens that are part of ions containing a metal atom e.g. $PtCl_6$

   b Halogens bonded to heteroatoms, to -C(=O)-, or to -C(=S)- in **organic** compounds. The overall functional group is coded L511, L512, or in Set K1: as applicable; the "X" atom in these code definitions has, from 1978 forward, been identified by a Set C0: code.

   c Halogen atoms as **ring heteroatoms**

   d **Radioactive** halogens

2. **Astatine** is coded A885, not in Set C0:.

Examples Using C0: Codes

a NaCl
C017, C100

b CaFCl
C009, C017, C100

c BF_3
C009 (only from 1978 forward)

C1: PRESENCE OF LOWER-VALENCY HALOGEN, H, C, N, O, OR S IN INORGANIC COMPOUNDS, IONS, OR LIGANDS

C100	Halogen with valency of I or II	(1970)
C101	H (**excluding** NH_3 and NH_4^+)	(1970)
C106	C	(1970)
C107	N with valency of II, III, or IV (**excluding** NH_3 and NH_4^+)	(1970)
C108	O (**excluding** -O-O-)	(1970)
C116	S with valency of II or III	

Note Although many codes defined for S are also assigned to Se and Te, C116 is not used for elemental Se or Te.

C2: PRESENCE OF INTERMEDIATE-VALENCY HALOGEN OR S IN INORGANIC COMPOUNDS, IONS, OR LIGANDS

C200	Halogen with valency of III or IV	(1970)
C216	S with valency of IV or V	(1970)

C3: PRESENCE OF HIGH-VALENCY HALOGEN, N, OR S IN INORGANIC COMPOUNDS, IONS, OR LIGANDS

C300	Halogen with valency of V or more	(1970)
C307	N with valency of V (**excluding** NH_3 and NH_4^+)	(1970)
C316	S with valency of VI or more	(1970)

C4: -O-O- OR -S-S- IN INORGANIC COMPOUNDS, IONS, OR LIGANDS

C408	-O-O-	(1970)
C416	-S-S-, -S=S-	(1970)

Notes on C4: Codes

1. C408 includes O_3 but does not include O_2. Ions or non-ionic compounds are additionally coded C108 when C408 is used, unless a non-peroxy O is present in a separate ion (see e.g. 'h' below).
2. C416 does not include elemental S.

Examples Using Codes from C1: through C4:

a H_2SO_4
C101, C108

b H_3PO_4
C101, C108, C316

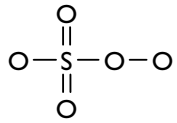
c $KClO_4$
C108, C300

d H_2O_2
C101, C408

e $NaHCO_3$
C101, C106, C108

f $(CH_3)_4N^+ Cl^-$
C100

g $Na_2S_2O_4$
C108, C216, C416

h 
C101, C316, C408 (not C108)

C5: IDENTIFICATION OF INORGANIC COMPOUNDS, IONS, AND LIGANDS CONSISTING SOLELY OF H, C, N, O, AND/OR S ATOMS

C500	NH_3 or NH_4^+
C510	HNO_3 or NO_3^-
C520	Other N-containing compounds, ions, and ligands that consist solely of H, C, N, O, and/or S
C530	Carbonic acid, thiocarbonic acid, and derivatives (including CO_2)
C540	Sulphur acids and derivatives (including SO_2)
C550	Other compounds, ions, and ligands consisting solely of H, C, O, and/or S

Notes on C5: Codes

1. A priority system is followed when using C5: codes: only the **first applicable code** is assigned to a compound or ion.
2. Compounds coded in Set C5: may consist of **elements other than H, C, N, O, and S,** but only when these other elements are present in ions not coded in Set C5:.
3. Elemental N is coded C520. Elemental H and elemental O are both assigned the code C550. Elemental C and elemental S are not searched in Set C5:; elemental C is searched using the codes C106 and C810, and elemental S is searched using the codes C116 and C810.

Examples Using C5: Codes

a NH_2SO_3H
C520 (but not C540)

b HNO_3
C510

c NH_4NO_3
C500, C510

d H_2O_2
C550

e $PbCO_3$
C530

f K_2SO_4
C540

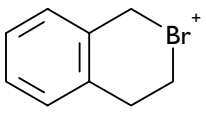
g Al (OH) Cl_2
C017, C550

h KIO_3
C053 (but not C550)

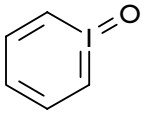
C600: Halogen as a ring heteroatom (1970)

Note The particular ring halogen present is identified by a code from Set C0:, and its valency is specified by C100, C200, C300, and in some instances L760 as applicable.

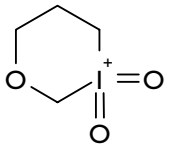
Examples Using C600

a 

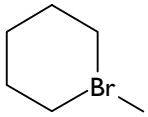
C035, C100, C600, L730 (hal$^+$, etc) ("halogen valency III-VII")

b 

C053, C300, C600, K140, ("halogen-O bond"), L760 ("halogen III, V, VII")

c 

C053, C300, C600, K140, L760

d 

C035, C200, C600, C760

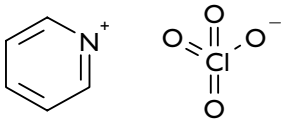
C7: DESCRIPTORS FOR INORGANIC COMPOUND AND IONS THAT DO NOT CONTAIN PART B: ELEMENTS

C710	Inorganic cation in an organic compound
C720	Inorganic anion in an organic compound
C730	Entirely inorganic compound

Note C7: codes are not used for free elements or for ions or compounds containing a noble gas, B, Si, P, As, Se, or Te.

Examples Using C7: Codes

a $FeCl_3$
C730

b 

C720

c CH_3COO^- NH_4^+
C710

d NH_4 (Me_4N) SO_4
C710, C720

C8: ELEMENTS NOT PRESENT IN INORGANIC COMPOUNDS, IONS, OR LIGANDS; MISCELLANEOUS DESCRIPTORS

C80: Elements not present in inorganic compounds, ions, or ligands

C800	Metals, other Part A: elements, and NH_4^+ are not present
C801	Noble gases, B, Si, P, As, Se, and Te are not present
C802	Halogens are not present
C803	H is not present
C804	C is not present
C805	N is not present
C806	O is not present
C807	S is not present

Notes on C80: Codes

1. Subset C80: is not applicable to:

- free elements
- mixtures during the period 1970-1971
- compounds for which the only inorganic part of the molecule is a metal cation or NH_4^+

2. Subset C80: is applicable to NH_3.
3. Elements present in **organic** group(s) of an ionic or coordinated compound are considered absent if not present in any of the **inorganic** groups.

Examples Using C80: Codes

a Na_2CO_3
C801, C802, C803, C805, C807

b $Cl_2P\text{-}C{\equiv}N$
C800, C803, C806, C807

c SO_2Cl_2
C800, C801, C803, C804, C805

d $(NH_4)_2\ SO_4$
C801, C802, C804

e $(CH_3)_4N^+\ ClO_4^-$
C801, C803, C804, C805, C807

f H_2O_2
C800, C801, C802, C804, C805, C807

C81: Miscellaneous descriptors for elements in organic or inorganic compounds, ions, or ligands

C810	Free element	(1970)
C811	Radioactive element	
C812	Specific isotope, non-radioactive	

Notes on C81: Codes

1. An element coded in Subset C81: is additionally assigned a code to identify it, even if the element is part of an organic group. For example, $^{14}CH_3CHO$ is coded C811 and C106, even though C in CH3- is not usually assigned the code C106.
2. C810 (together with C106) is coded for elemental C including graphite or diamond and for elemental S (together with C116).

Examples Using C81: Codes

a D_2
C101, C810, C812

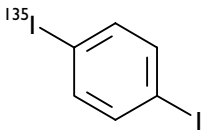
b H_2
C101, C810

c $^{14}CO_2$
C106, C811

d $^{14}CH_3CHO$
C106, C811, J471 (“aliphatic aldehyde”)

e CH_3OT
C101, C811, H481 (“aliphatic -OH”)

f



C053, C811, H604, H608 ("bis iodine"), H642 ("2 aromatic halogens")

7 Ring Systems

7 Ring systems

Introduction

Rings are usually the largest, most easily identifiable fragments in any organic chemical structure, and are thus a good starting point for selecting chemical codes for searching.

Definitions of Ring Types Discussed in Parts D:, E:, F:, and G:

Alicyclic ring system consisting only of carbon atoms, with no aromatic rings in system

Aromatic benzene or a quinonoid derivative of benzene, unfused or fused to a carbocyclic ring. The term ***aromatic*** is defined in more detail on page 90.

Bridged rings that have three or more adjacent atoms in common; or ring system formed by a chain of one or more atoms bonded at either end to two non-adjacent atoms of the same ring

Carbocyclic ring system consisting only of carbon atoms; can be alicyclic or aromatic

Fused rings that have two adjacent atoms in common

Heterocyclic ring system with at least one ring atom that is not carbon

Mononuclear an unfused ring

Ring System either an unfused ring, rings that are fused together, or bridged rings

Spiro ring systems that have one atom in common

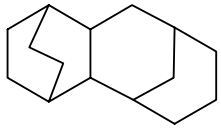
Notes on Ring System Codes

1. Only **one code** from Parts D: through G: is assigned to a mononuclear, fused, or bridged ring system of an organic compound. Inorganic compounds, even if regarded as cyclic, are not assigned D: through G: codes.

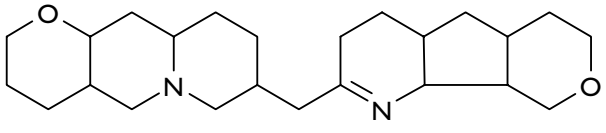
2. Component rings of **fused and bridged ring systems** cannot be searched individually; they are searched together as a **single system**.

3. Some ring codes describe more than one possible ring system; for example, D230 is used for any 3-ring fused system consisting of two O atoms and any number of C atoms. Searches using codes that refer ambiguously to more than one ring system are made more specific by the addition of Ring Index Numbers (RIN's) to the search strategy. An asterisk (*) next to a ring code indicates that Ring Index Numbers (RIN's) may have been assigned to various structures covered by the code. RIN's are discussed in detail on page 74.

4. Ring systems are named according to *The Ring Index* by Patterson, Capell and Walker, 2nd Edition (American Chemical Society). The number of rings in the system can be determined by counting the minimum number of bonds which must be broken to make it an open chain structure. The ring sizes used to describe the system are by convention the smallest numbers which cover the entire ring system. For example, the structure below is considered a 6:6:6:6 system rather than a 6:6:6:8 system.



5. The ***Poly*** code at the end of most Sets of ring codes indicates that a single code in that set has been used more than once to describe a particular structure, although not necessarily for identical ring systems. ***Poly*** does not mean that two or more different codes in a Set have been used. The compound shown below is coded **E520** and **E599**.



6. The total number of each type of ring system (fused-ring heterocyclic, mononuclear heterocyclic, aromatic and alicyclic) is searchable in Set M5:, the description is on page 182.

7. Many ring code definitions start with the term ***other***. Consider as an example the following ring codes:

G4: Aromatic Systems With 4 or More Rings

G400	4-ring system with at least one 3-, 4- or 5-membered ring
G410	Chrysene
G420	Naphthacene
G430	***Other*** 6:6:6:6 systems
G440	***Other*** 4-ring systems

The first code, G400, covers all 4-ring systems that have at least one aromatic ring and at least one ring that has less than six members. The codes G410 and G420 cover specific aromatic compounds having four rings of six members each. The code G430 covers all 4-ring aromatic systems with six members in each ring **except** those ring systems covered by G410 and G420. G440 covers all 4-ring aromatic systems **except** those covered previously by G400, G410, G420, and G430; thus G440 covers 4-ring systems that have at least one aromatic ring, at least one ring with more than six members, and no rings with less than six members.

Degree of Unsaturation in Rings

Some chemical codes for ring systems differentiate between various hydrogenation states of the component rings, while other codes cover all possible hydrogenation states. In general, each code in Sets D1: to G4: covers all hydrogenated (i.e. more saturated) derivatives of the ring systems specified, unless a separate specific code for the hydrogenated form is provided. Conversely, each code in Sets G5: to G8: covers all less-saturated derivatives of the ring systems specified, unless a separate specific code for the less-saturated derivative is provided. Examples are given below:

- F610 is defined as oxazole, but it also covers di- and tetrahydrooxazoles.
- G650 is defined as azulene, but it also covers optionally unsaturated derivatives.
- Quinoline is searched as D621; di- and polyhydroquinolines are searched as D622.
- Benzene is searched with an appropriate G1: code, cyclohexadiene with G561, cyclohexene with G562, and cyclohexane with G563.

Ring Substituents

A ***ring substituent*** is any atom or group of atoms other than H that is bonded to a ring atom. As a rule, indexers attempt to index all specific substitution patterns disclosed in a patent for the most important ring system in a particular structure. Substituents do not normally affect the choice of a ring code, unless tautomerism of the substituent changes the hydrogenation state of the ring; in such cases tautomerism may change the ring code used. (Tautomerism is discussed on page 102.) A spiro ring is considered as one substituent; points of fusion of fused ring systems are not considered substituents.

The numbers specified for substituents in Sets D0:, F0: and G0: refer to the ring positions bearing substituents, not to the total number of substituents. In particular, the substituent number in each code of Sets D0: and G0: refers to the number of ring atoms bearing substituents. The substituent number in each Set F0: code refers to the IUPAC position number of each ring atom bearing substituent(s). These rules are presented in more detail in the discussions preceding each ring system code Part.

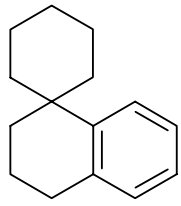
The codes D010, D020, F010, G001, G002, and G003 (defined as "Non-specific substitution pattern disclosed") are applied when very general patterns of substitution are disclosed in a patent, e.g. "optionally substituted benzene". All specifically exemplified or claimed substitution patterns in such patents are given more specific substitution codes.

Spiro Structures

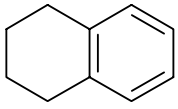
Each ring system of a spiro structure is coded separately. Additionally, each spiro system receives one or more of the following codes, depending on whether the rings spiro to one another are alike or different:

D041	A ring system coded in Part D: or E: is spiro to another ring
F022	A ring system coded in Part F: is spiro to another ring
G041	A ring system coded in G1: to G4: is spiro to another ring
G052	A ring system coded in G5: to G7: is spiro to another ring

Spiro Example 1

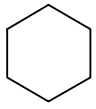


search as



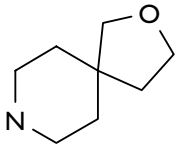
G223, G041, G031

and

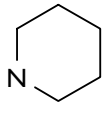


G563, G052, G030

Spiro Example 2

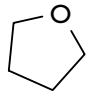


search as



F433, F022, F014

and



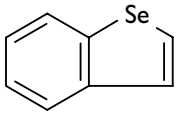
F113, F022, F013

Even though the code F022 occurs twice in the second example above, a **Poly** code is not used because there is only one spiro linkage in the structure. The entire spiro system receives a Ring Index Number (RIN). (RIN's are discussed on page 74.)

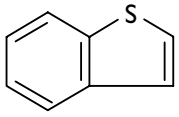
Ring Heteroatoms Other Than O, S, and N

The ring codes in Parts D:, E:, F:, and G: are specifically defined for rings consisting of C, O, S, and N. The rules for adapting these codes for use with rings composed of other atoms are listed below.

1. **Se or Te** – Use the code for the ring system obtained by replacing all Se and Te atoms with S atoms. Include the code for Se (B634) and/or Te (B652) as applicable.

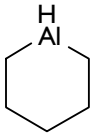


search as

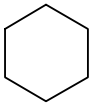


D300, B634

2. **Metals** – If a metal atom is bonded to two C atoms in a ring, the ring is coded as if the metal atom were replaced with a C atom. However, even when a carbocyclic ring code is used for rings containing one or more metal atoms, substituents are coded as if bonded to a heterocyclic ring. The code A922 ("metal in ring") is also applicable, as is the Part A: code for the specific metal in the ring. A Ring Index Number (RIN) may also apply. (RIN's are discussed on page 74).



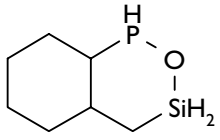
search as



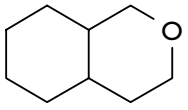
G563, A922, A313

If the metal is adjacent to a ring atom other than C, consult the rules for metal-heteroatom bonds in Chapter 6, beginning on page 35.

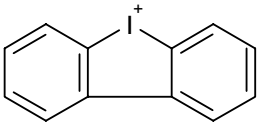
3. **Other Elements** – Use the code for the ring system obtained by replacing the element(s) with C atom(s). Although a carbocyclic ring code is in some cases applicable to such rings, substituents are coded as if bonded to a heterocyclic ring. The code in Part B: or Part C: for each specific element present in the ring (i.e. the elements "other" than C, O, S, N, Se, Te, or metal) is assigned. B840 (B, Si, P or As as ring heteroatom) or C600 (halogen as ring heteroatom) may also apply.



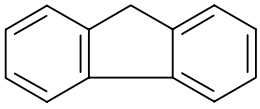
search as



D130, B840, B614, B615



search as



G310, C600

Note The "Basic Group" code for the three types of ring systems described above is M411 (**except** in the case of pre-1970 multicomponent compositions, which are discussed on page 181).

Ring Atoms That Are Coded In Non-Ring Code Parts

As explained in the previous section, rings containing elements other than C, O, S, and N receive both a ring code from Parts D: through G:, and one or more element codes from Parts A: through C:. In addition, when certain combinations of atoms and bonds are found together in a ring they receive a functional group code from Set L9:, as explained on page 165.

Steroids

Compounds that contain the cyclopenta(a)phenanthrene ring system, i.e. the steroid nucleus, are never assigned codes in Parts D: through G:, nor are they assigned the steroid Ring Index Number (RIN) 04781. (RIN's are discussed on page 70.) Instead, steroids are structure-coded in Parts S:, T:, and U:, as discussed in Chapter 13 beginning on page 233. However, when the cyclopenta(a)phenanthrene system is spiro or fused to another ring, then the RIN for the complete system is assigned. The RIN for the ring system spiro to the steroid nucleus is also assigned if it exists, but the RIN for the steroid nucleus itself is never assigned.

General Ring Codes

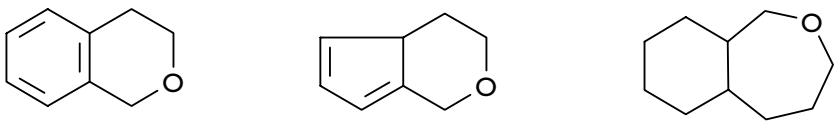
Patents sometimes mention general types of rings rather than specifying the exact ring system involved in an invention - e.g. “aryl” or “aromatic heterocyclic ring system”. Chemical codes exist for indexing such patents, but they are only recommended for use in the most exhaustive searches. Such “general codes” are not applied to patent documents in which all possible specific ring systems are disclosed.

D040 Fused ring heterocycle - general
F020 Aromatic mononuclear heterocycle - general
F021 Non-aromatic mononuclear heterocycle - general
G040 Aromatic carbocycle - general
G050 Monocyclic alicycle - general
G051 Polycyclic alicycle - general

Suppose all patent records related to pyridine are being sought. **F431** retrieves all patent records in which pyridine has been specifically mentioned. **F020**, which indicates the general disclosure of aromatic mononuclear heterocycles, retrieves patent records in which terms such as “heterocyclic aryl” were used, but in which pyridine itself was not specified. Thus the strategy (**F431 OR F020**) retrieves all patents related to pyridine. In practice, using **F431** alone will probably satisfy most patent search requirements.

Ring Index Numbers (RIN’s)

Some rings are not uniquely described by a single code. For example, **D130** is assigned to all of the following ring systems:



Furthermore, each code not only covers the ring(s) for which it is defined, but also equivalent rings with heteroatoms other than O, S, and N, as explained earlier in this chapter on page 68. To enable more specific searches on ring systems, Derwent began assigning ring numbers from *The Ring Index* (by Patterson, Capell, and Walker, 2nd Edition, American Chemical Society) to patent indexing records in 1972. These numbers, known as Ring Index Numbers (RIN’s), are five digit numbers that appear in the RR field of the DWPI database. Applicable RIN’s are ‘AND’ed to the search statement containing codes introduced in 1972 (the blue codes on the coding sheet). Since 1986 RINs are also linked to the appropriate codes in the chemical code fields, to enable more specific retrieval.

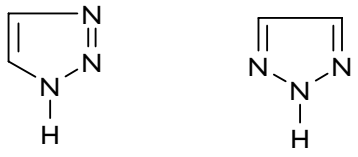
RIN's are assigned in the following cases:

- ring systems whose chemical codes additionally cover other ring systems (codes covering more than one ring system are marked with an *)
- rings containing elements other than C, N, O, or S
- spiro rings

RIN's are NOT used for:

- ring systems uniquely identified by a single code
- tautomers or different states of hydrogenation
- metals in rings, unless the metal is bonded to two ring C atoms

RIN's indicate the skeleton of ring atoms present, not taking into account substituents, tautomeric forms, or degrees of unsaturation. When *The Ring Index* assigns different numbers to different tautomers, Derwent applies the lowest assigned ring number to all tautomeric forms. For example, even though the two rings shown below each have their own ring number in *The Ring Index*, Derwent indexers assign the lowest ring number, 00094, to both rings.

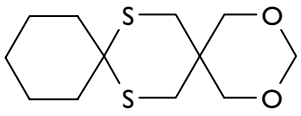


00094 00095

When indexers encountered patented ring systems that were not included in *The Ring Index*, Derwent created its own system of RIN's. The rings in *The Ring Index* are numbered from 00001 to 25000; Derwent's ring numbers start at 40000. Derwent-generated RIN's are available on cards, CD-ROM, and via the Markush TOPFRAG software. Markush TOPFRAG automatically inserts applicable RIN's from *The Ring Index* and Derwent-generated RIN's in the search strategies it constructs.

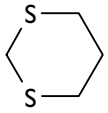
For historical reasons, a maximum of 117 RIN's are assigned to a single patent. For spiro ring systems, the RIN of the overall spiro structure can be searched, as well as the RIN's for the individual rings systems (if RIN's exist for those particular ring systems). For example, when searching for patents on structure (a) below, the RIN's for structures (b) and (c) can also be searched, but not the RIN's for structures (d) and (e).

a



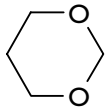
03423

b



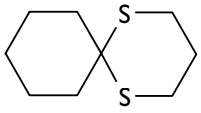
00266

c

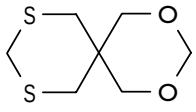


00262

d



e



Parts D: and E: Fused Ring Heterocycles – I & II

Main Headings

D0: General descriptors
D1: Ring systems consisting solely of O and C; 2 rings
D2: Ring systems consisting solely of O and C; 3 or more rings
D3: Ring systems consisting solely of S and C; 2 rings
D4: Ring systems consisting solely of S and C; 3 or more rings
D5: Ring systems consisting solely of O, S, and C; 2 or more rings
D6: Ring systems having 1 N as the sole heteroatom; 2 rings
D7: Ring systems having 2 N atoms as the sole heteroatoms; 2 rings
D8: Ring systems having 3 N atoms as the sole heteroatoms; 2 rings
D9: Ring systems having 4 or more N atoms as the sole heteroatoms; 2 rings

E1: Ring systems having 1 N as the sole heteroatom; 3 rings
E2: Ring systems having 2 or more N atoms as the sole heteroatoms; 3 rings
E3: Ring systems having any number of N atoms as the sole heteroatom(s); 4 or more rings
E4: Ring systems consisting solely of O, N, and C; 2 rings
E5: Ring systems consisting solely of O, N, and C; 3 or more rings
E6: Ring systems consisting solely of S, N, and C; 2 rings
E7: Other 2-ring systems consisting solely of S, N, and C
E8: Ring systems consisting solely of S, N and C; 3 or more rings
E9: Ring systems with O, S, and N as heteroatoms; 2 or more rings

Introduction

Parts D: and E: refer to fused ring heterocyclic systems; Part E: is a continuation of Part D:. Fused ring heterocyclic systems consist of **two or more fused rings, at least one of which is heterocyclic**. The codes in Parts D: and E: are hierarchical. In all Sets except D0:, the first digit indicates the heteroatom(s) present and the number of rings; the second digit differentiates between various ring sizes and ring positions of heteroatoms; and the third digit indicates the state(s) of hydrogenation covered by the code. In most cases, a final "0" indicates that the code covers all hydrogenation states; "1" means that the ring system is fully unsaturated; and "2" indicates a hydrogenated form.

In addition to the rings specified in the definitions of codes in Parts D: and E:, each code also covers rings in which one or more C atoms have been replaced by heteroatoms other than O, S, N, Se, or Te, and rings in which one or more S atoms have been replaced by Se or Te. These substituted elements in the ring code definitions are discussed on page 72 in the section entitled "Ring Heteroatoms Other than O, S, and N".

Heterocyclic Fused Ring Code Finder

Heteroatoms present (specific # if given)	# or Rings	Set Used
O	2	D1:
O	3 or more	D2:
S	2	D3:
S	3 or more	D4:
N (1)	2	D6:
N (1)	3	E1:
N (2)	2	D7:
N (3)	2	D8:
N (2 or more)	3	E2:
N (4 or more)	2	D9:
N (any number)	4	E3:
O, S	2 or more	D5:
O, N	2	E4:
O, N	3 or more	E5:
S, N (S+N=2)	2	E6:
S, N (S+N=3 or more)	2	E7:
S, N	3 or more	E8:
O, S, and N	2 or more	E9:

Notes on Coding Ring Substituents in Parts D: and E:

1. The component rings are considered separately when determining ring substitution codes. Each heterocyclic ring in the system that bears substituents receives a code from Subset D01:; each carbocyclic ring in the system that bears substituents receives a code from Subset D02:.

2. The number of ring atoms bearing substituents is searched, not the relative positions of the substituents. For example, D015 means that there are four or more atoms in a heterocyclic ring of a fused ring heterocyclic system that bear substituents. D023 means that two atoms in a carbocyclic ring of a fused ring heterocyclic system bear substituents.

3. If there is only one substituent, the codes D011 and D012 specify adjacent (i.e. alpha) or not adjacent (i.e. ≥beta) to ring fusion, respectively.

4. D016 indicates that two or more substituents are on a single atom. D030 indicates substitution on a ring fusion atom.

5. A "Poly" code is used when one or more of the substitution codes is valid for more than one ring.

6. Spiro rings are considered as one substituent. A ring spiro to a fused heterocyclic system receives the spiro code D041.

Subheadings Applicable To Codes in Parts D: and E:

Subsets D00: to D03: are applicable to subheadings M1 to M4 from 1981 forward. All other codes apply to M0 from 1963 and to M1, M2, M3 and M4 from 1970 forward.

D0: GENERAL DESCRIPTORS

D000: Unsubstituted system (1981)

D01: Substituents on a heterocyclic ring of a fused-ring heterocyclic system

D010	Non-specific heterocyclic substitution pattern disclosed	(1981)
D011	1 atom of fused heterocyclic ring bears substituent(s); atom is adjacent (alpha) to ring fusion	(1981)
D012	1 atom of fused heterocyclic ring bears substituent(s); atom is not at or adjacent (atom is ≥beta) to ring fusion	(1981)
D013	2 atoms of a fused heterocyclic ring bear substituents	(1981)
D014	3 atoms of a fused heterocyclic ring bear substituents	(1981)
D015	4 or more atoms of a fused heterocyclic ring bear substituents	(1981)
D016	2 or more substituents are on the same atom	(1981)
D019	Poly	(1981)

D02: Substituents on a carbocyclic ring of a fused-ring heterocyclic system

D020	Non-specific carbocyclic substitution pattern disclosed	(1981)
D021	1 C atom of fused carbocyclic ring bears substituent(s); C atom is adjacent (alpha) to ring fusion	(1981)
D022	1 C atom of fused carbocyclic ring bears substituent(s); C atom is not at or adjacent (C atom is ≥beta) to ring fusion	(1981)
D023	2 C atoms of a fused carbocyclic ring bear substituents	(1981)
D024	3 C atoms of a fused carbocyclic ring bear substituents	(1981)
D025	4 or more C atoms of a fused carbocyclic ring bear substituents	(1981)
D026	2 substituents are on the same C atom	(1981)
D029	Poly	(1981)

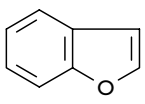
D030: One or more substituents at ring fusion (1981)

D04: General types of ring system

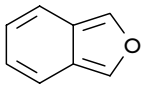
D040	Non-specific disclosure of a fused-ring heterocyclic system
D041	A ring is spiro to a fused-ring heterocyclic system
D049	Poly

D1: RING SYSTEMS CONSISTING SOLELY OF O AND C; 2 RINGS

D100 Benzo(b)furan

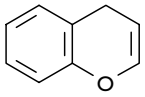


D111 Benzo(c)furan, 1,3-dihydrobenzo(c)furan



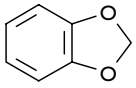
D112 Di- and polyhydrobenzo(c)furan (excluding 1,3-dihydro form covered by D111)

D120 Benzo(b)pyran



D130 Other 2-ring systems with 1 O atom as the sole heteroatom

D140 Benzodioxole



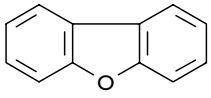
D150 Benzodioxins

D160 Other 2-ring systems with 2 or more O atoms as the sole heteroatoms

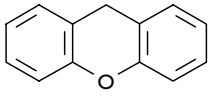
D199 Poly

D2: RING SYSTEMS CONSISTING SOLELY OF O AND C; 3 OR MORE RINGS

D200 Dibenzofuran



D210 Xanthene



D220 Other ring systems with 1 O atom as the sole heteroatom; 3 or more rings

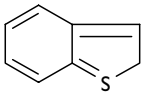
D230 3-ring systems with 2 O atoms as the sole heteroatoms

D240 Other ring systems with O as the sole heteroatom; 3 or more rings

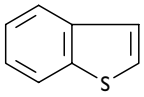
D299 Poly

D3: RING SYSTEMS CONSISTING SOLELY OF S AND C; 2 RINGS

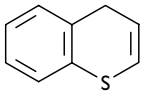
D300 1-Benzothiophene



D310 2-Benzothiophene



D320 1-Benzothiopyran



D330 Other 2-ring systems with 1 S atom as the sole heteroatom

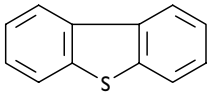
D340 All 2-ring systems with 2 S atoms as the sole heteroatoms

D350 All 2-ring systems with 3 or more S atoms as the sole heteroatoms

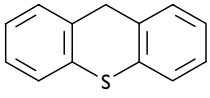
D399 Poly

D4: RING SYSTEMS CONSISTING SOLELY OF S AND C; 3 OR MORE RINGS

D400 Dibenzothiophene



D410 Thioxanthene



D420 Other ring systems with 1 S atom as the sole heteroatom; 3 or more rings

D430 Other ring systems with 2 or more S atoms as the sole heteroatoms; 3 or more rings

D499 Poly

D5: RING SYSTEMS CONSISTING SOLELY OF O, S, AND C; 2 OR MORE RINGS

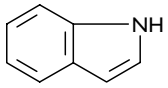
D500 Ring systems with O and S as the sole ring heteroatoms; 2 rings

D510 Ring systems with O and S as the sole ring heteroatoms; 3 or more rings

D599 Poly

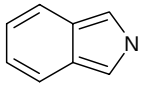
D6: RING SYSTEMS WITH 1 N AS THE SOLE HETEROATOM; 2 RINGS

D601 Indole



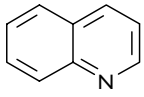
D602 Di- and polyhydroindoles

D611 Isoindole, 1,3-dihydroindole



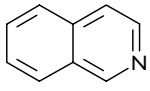
D612 Di- and polyhydroisoindoles (excluding 1,3-dihydro form covered by D611)

D621 Quinoline



D622 Di- and polyhydroquinolines

D631 Isoquinoline

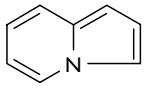


D632 Di- and polyhydroisoquinolines

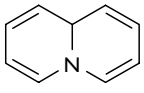
D640 Benzazepines

D650 Benzazocines

D660 Indolizine, quinolizine



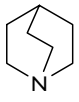
01276



01686

D670 Nortropane and optionally unsaturated derivatives

D680 Quinuclidine and optionally unsaturated derivatives

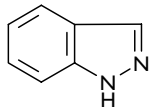


D690 Other 2-ring systems with 1 N as the sole heteroatom

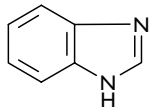
D699 Poly

D7: RING SYSTEMS WITH 2 N ATOMS AS THE SOLE HETEROATOMS; 2 RINGS

D700 Indazole



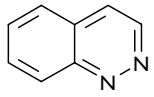
D711 Benzimidazole



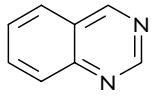
D712 Di- and polyhydrobenzimidazoles

D720 Pyrrolo-pyridines

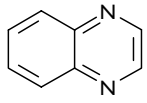
D730 Cinnoline



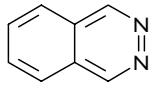
D740 Quinazoline



D750 Quinoxaline



D760 Phthalazine



D770 Pyrido-pyridines

D780 Benzodiazepines, benzodiazocines

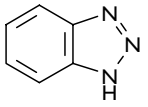
D790 Other 2-ring systems with 2 N atoms as the sole heteroatoms

D799 Poly

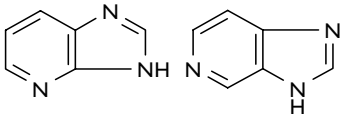
D8: RING SYSTEMS WITH 3 N ATOMS AS THE SOLE HETEROATOMS; 2 RINGS

D800 Both are 5-membered rings

D810 Benzotriazole



D820 Imidazo-pyridines



01192 01194

D830 Pyrazolo-pyridines

D840 Pyrrolo-pyrimidines

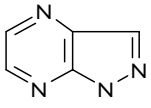
D850 Pyrido-pyrimidines

D860 Other 2-ring systems with 3 N atoms as the sole heteroatoms

D899 Poly

D9: RING SYSTEMS WITH 4 OR MORE N ATOMS AS THE SOLE HETEROATOMS; 2 RINGS

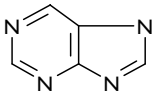
D900 Pyrazolo-pyrazine



D910 Pyrazolo-pyridazines

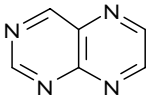
D920 Pyrazolo-pyrimidines

D931 Purine

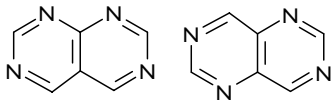


D932 1,2,3,6-Tetrahydropurine, hexahydropurine

D940 Pteridine

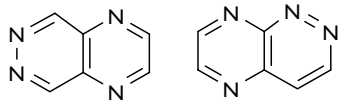


D950 Pyrimido-pyrimidines



01589 01586

D960 Pyrazino-pyridazines



01591 08099

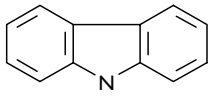
D970 Other 2-ring systems with 4 N atoms as the sole heteroatoms

D980 2-ring systems with 5 or more N atoms as the sole heteroatoms

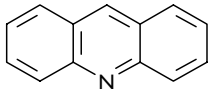
D999 Poly

E1: RING SYSTEMS WITH 1 N AS THE SOLE HETEROATOM; 3 RINGS

E100 Carbazole

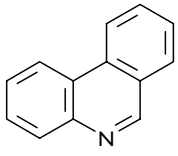


E111 Acridine

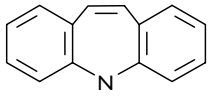


E112 Di- and polyhydroacridines

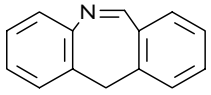
E120 Phenanthridine



E130 Dibenzo(b,f)-azepine



E140 Morphanthridine



E150 N-atom angular (at ring fusion)

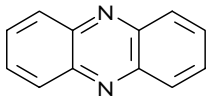
E160 Other 3-ring systems with 1 N as the sole heteroatom

E199 Poly

E2: RING SYSTEMS WITH 2 OR MORE N ATOMS AS THE SOLE HETEROATOMS; 3 RINGS

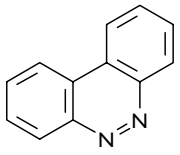
E200 Carboline

E210 Phenazine



E220 Phenanthrolines

E230 Dibenzopyridazine

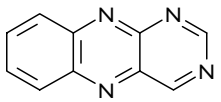


E240 Dibenzodiazepines, dibenzodiazocines

E250 Other 3-ring systems with 2 N atoms as the sole heteroatoms

E260 3-ring systems with 3 N atoms as the sole heteroatoms

E270 Benzo(g)pteridine



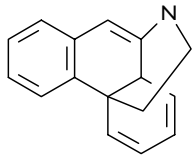
E280 Other 3-ring systems with 4 N atoms as the sole heteroatoms

E290 3-ring systems with 5 or more N atoms as the sole heteroatoms

E299 Poly

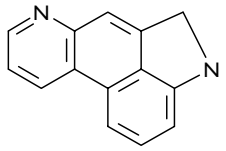
E3: SYSTEMS WITH 4 OR MORE RINGS AND N AS THE SOLE HETEROATOM

E300 Morphinan



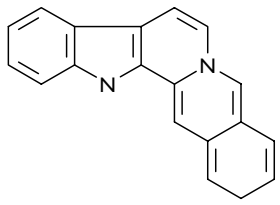
E310 Other 4-ring systems with 1 N atom as the sole heteroatom

E320 Ergoline



E330 Other 4-ring systems with 2 or more N atoms as the sole heteroatoms

E340 Yohimbane

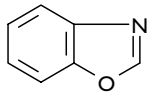


E350 Other systems with 5 or more rings and N as the sole heteroatom(s)

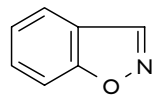
E399 Poly

E4: RING SYSTEMS CONSISTING SOLELY OF O, N, AND C; 2 RINGS

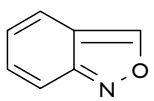
E400 Benzoxazole



E410 Benzisoxazoles

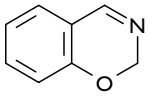


01123

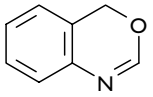


01124

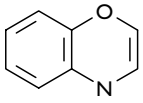
E420 1,3-Benzoxazine



E430 3,1-Benzoxazine



E440 1,4-Benzoxazine



E450 Benzoxazepines, benzoxazocines

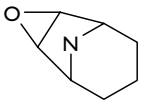
E460 Other 2-ring systems with 1 O atom, 1 N atom, and any number of C atoms

E470 Other 2-ring systems consisting solely of O, N, and C

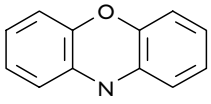
E499 Poly

E5: RING SYSTEMS CONSISTING SOLELY OF O, N, AND C; 3 OR MORE RINGS

E500 Scopolamine nucleus, optionally unsaturated



E510 Phenoxazine

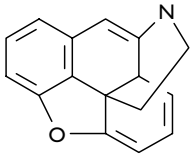


E520 Other 3-ring systems with 1 O atom, 1 N atom, and any number of C atoms

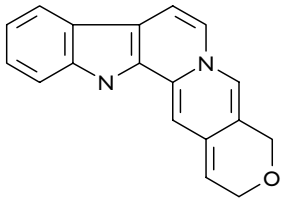
E530 Other 3-ring systems consisting solely of O, N, and C

E540 4-ring systems consisting solely of O, N, and C

E550 Morphine nucleus



E560 Reserpiline nucleus

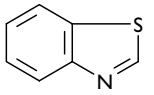


E570 Other ring systems consisting solely of O, N, and C; 5 or more rings

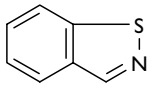
E599 Poly

E6: RING SYSTEMS CONSISTING SOLELY OF S, N, AND C; 2 RINGS

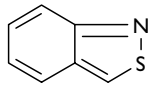
E600 Benzothiazole



E610 Benzisothiazoles

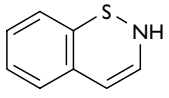


01150

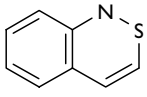


01151

E620 1,2-Benzothiazine; 2,1-benzothiazine

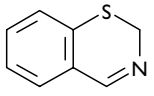


01577

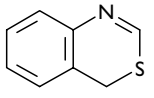


10041

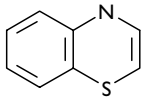
E630 1,3-Benzothiazine



E640 3,1-Benzothiazine

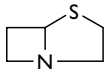


E650 1,4-Benzothiazine

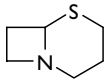


E660 Benzothiazepines, benzothiazocines

E670 Penicillin nucleus, optionally unsaturated



E680 Cephalosporin nucleus, optionally unsaturated



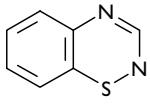
E690 Other 2-ring systems with 1 S atom, 1 N atom, and any number of C atoms

E699 Poly

E7: OTHER 2-RING SYSTEMS CONSISTING SOLELY OF S, N, AND C (EXCLUDING RING SYSTEMS COVERED BY CODES IN SET E6:.)

E700 2-ring systems with 2 S atoms, 1 N atom, and any number of C atoms

E710 1,2,4-Benzothiadiazine



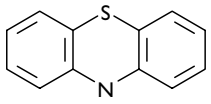
E720 Other 2-ring systems with 1 S atom, 2 N atoms, and any number of C atoms

E730 2-ring systems consisting solely of S, N, and C; (number of S atoms) + (number of N atoms) = 4 or more

E799 Poly

E8: RING SYSTEMS CONSISTING SOLELY OF S, N AND C; 3 OR MORE RINGS

E800 Phenothiazine



E810 3-ring system with 1 S atom, 1 N atom, and any number of C atoms; the S and N atoms are in the same ring

E820 3-ring system with 1 S atom, 1 N atom, and any number of C atoms; the S and N atoms are in different rings

E830 3-ring system with 2 S atoms, 1 N atom, and any number of C atoms

E840 Azaphenothiazines

E850 Other 3-ring systems with 1 S atom, 2 N atoms, and any number of C atoms

E860 Other 3-ring systems consisting solely of S, N, and C

E870 Ring systems consisting solely of S, N, and C; 4 rings

E880 Ring systems consisting solely of S, N, and C; 5 or more rings

E899 Poly

E9: RING SYSTEMS WITH O, S, AND N AS HETEROATOMS; 2 OR MORE RINGS

E900 Ring systems with O, S, and N as heteroatoms; 2 rings

E910 Ring systems with O, S, and N as heteroatoms; 3 rings

E920 Ring systems with O, S, and N as heteroatoms; 4 or more rings

E999 Poly

Part F: Mononuclear Heterocycles

Main Headings

F0: General Descriptors
F01: Positions substituted
F02: Other general descriptors

F1: Unfused Rings Consisting Solely of O and C
F2: Unfused Rings Consisting Solely of S and C
F3: Unfused Rings Consisting Solely of O, S, and C
F4: Unfused Rings With 1 N as the Sole Heteroatom
F5: Unfused Rings With 2 or More N atoms as the Sole Heteroatoms
F6: Unfused Rings Consisting Solely of O, N, and C
F7: Unfused Rings Consisting Solely of S, N, and C and Optionally O

Introduction

The codes in Part F: are hierarchical. In all Sets except F0:, the first digit indicates the heteroatom(s) present; the second digit identifies the specific type of ring system present; and the third digit indicates the state of hydrogenation. In most codes a final "0" means that the code covers all hydrogenation states; "1" means fully unsaturated; "2" means intermediate saturation; and "3" means fully saturated.

In addition to the rings specified in the code definitions for Part F:, each code also covers rings in which 1 or more C atoms have been replaced by heteroatoms other than O, S, N, Se, or Te, and rings in which one or more S atoms have been replaced by Se or Te. These substituted elements in the ring code definitions are discussed on page 72 in the section entitled "Ring Heteroatoms Other than O, S, and N".

Notes on Coding Ring Substituents in Part F:

1. The **ring positions** of atoms bearing substituents are searched (not the total number of ring atoms bearing substituents as in Parts D: and G:).

2. The numbering of heterocyclic rings follows **IUPAC** conventions. Numbering starts with the highest priority heteroatom, and then the rest of the ring is numbered in the direction yielding the lowest possible number(s) for the remaining heteroatom(s). The heteroatom priority is: O > S > N (descending group order).

3. **F017** indicates that two substituents are on the same C atom in the ring. **F018** indicates that two or more substituents are on the same heteroatom.

4. **Poly** is used when different rings have the same positions substituted, not when there are multiple substituents on the same ring.

5. **Spiro rings** are considered one substituent; a ring spiro to a mononuclear heterocycle receives the spiro code F022.

Subheadings Applicable to Codes in Part F:

Codes F000 to F019 are applicable to subheadings M1 to M4 from 1981 forward. All other codes apply to M0 from 1963 and to M1, M2, M3 and M4 from 1970 forward.

F0: GENERAL DESCRIPTORS

F000	Unsubstituted mononuclear heterocycles	(1981)

F01: Positions substituted

F010	Non-specific substitution pattern disclosed	
F011	Position 1 substituted	(1981)
F012	Position 2 substituted	(1981)
F013	Position 3 substituted	(1981)
F014	Position 4 substituted	(1981)
F015	Position 5 substituted	(1981)
F016	Position 6 or higher substituted	(1981)
F017	2 substituents on the same C atom	(1981)
F018	2 or more substituents on the same heteroatom	(1981)
F019	Poly	(1981)

F02: Other general descriptors

F020	Non-specific disclosure of aromatic heterocyclic ring
F021	Non-specific disclosure of non-aromatic heterocyclic ring
F022	A ring is spiro to a mononuclear heterocyclic ring
F029	Poly

Note An aromatic heterocyclic ring is one containing the maximum possible number of double bonds. Examples of aromatic heterocyclic rings are pyrrole, furan, and thiophene.

F1: UNFUSED RING CONSISTING SOLELY OF O AND C

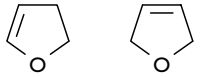
F100 Oxirene; oxete

00012

00044

F111 Furan

F112 Dihydrofuran



F113 Tetrahydrofuran

F121 Pyran, dihydropyran

F123 Tetrahydropyran

F130 7- or more membered ring consisting of 1 O atom and 6 or more C atoms

F140 1,3-Dioxole

F150 Other 3-, 4- or 5-membered ring consisting of 2 or more O atoms as the sole heteroatoms

F161 Dioxins

F163 Dioxans

F170 Ring of 6 or more members consisting of 2 or more O atoms as the sole heteroatoms

F199 Poly

F2: UNFUSED RING CONSISTING SOLELY OF S AND C

F200 Thiirene, thiete

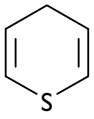
00013 00045

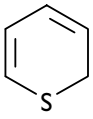
F211 Thiophene

F212 Dihydrothiophene

F213 Tetrahydrothiophene

F220 Thiopyran





F230 Ring consisting of 1 S atom and 6 or more C atoms

F240 3-, 4- or 5-membered ring consisting of 2 or more S atoms as the sole heteroatoms

F250 Ring of 6 or more members consisting of 2 or more S atoms as the sole heteroatoms

F299 Poly

F3: UNFUSED RING CONSISTING SOLELY OF O, S, AND C

F300 3- or 4-membered ring consisting solely of O, S, and C

F310 5-membered ring consisting solely of O, S, and C

F320 6-membered ring consisting solely of O, S, and C

F330 Ring of 7 or more members consisting solely of O, S, and C

F399 Poly

F4: UNFUSED RING WITH 1 N AS THE SOLE HETEROATOM

F400 Azirine

F410 Azete

F421 Pyrrole

F422 Dihydropyrrole

F423 Pyrrolidine

F431 Pyridine

F432 Dihydropyridine, tetrahydropyridine

F433 Piperidine

F443 Hexamethyleneimine (unsaturated forms of hexamethyleneimine are coded F450 with RIN 00355)

00355

F450 Other unfused ring with 1 N as the sole heteroatom

F499 Poly

F5: UNFUSED RING WITH 2 OR MORE N ATOMS AS THE SOLE HETEROATOMS

F500 3- or 4-membered ring consisting of 2 or more N atoms and at least 1 C atom

F511 Pyrazole

F512 Dihydropyrazole, pyrazoline

F513 Pyrazolidine

F521 Imidazole

F522 Dihydroimidazole, imidazoline

F523 Imidazolidine

F530 Pyridazine

F541 Pyrimidine

F542 Dihydropyrimidine, tetrahydropyrimidine

F543 Hexahydropyrimidine

F551 Pyrazine

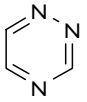
F552 Dihydropyrazine, tetrahydropyrazine

F553 Piperazine

F560 7-or more-membered unfused ring consisting of 2 N atoms and 5 or more C atoms

F570 5-membered ring consisting of 3 or more N atoms

F580 Triazines



00212 00210 00211

F590 Other ring of 6 or more members consisting of 3 or more N atoms as the sole heteroatoms and at least 1 C atom

F599 Poly

F6: UNFUSED RING CONSISTING SOLELY OF O, N, AND C

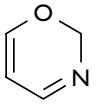
F600 3- or 4-membered ring consisting solely of O, N, and C

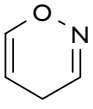
F610 Oxazole

F620 Isoxazole

F630 Other 5-membered ring consisting solely of O, N, and C

F640 1,2-Oxazine; 1,3-oxazine





00237 00234

F651 1,4-Oxazine; 1,4-dihydrooxazine

F653 Morpholine

F660 Other ring of 6 or more members consisting solely of O, N, and C

F699 Poly

F7: UNFUSED RING CONSISTING OF S, N, C AND OPTIONALLY O

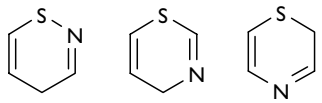
F700 3- or 4-membered ring consisting solely of S, N, and C

F710 Thiazole

F720 Isothiazole

F730 Other 5-membered ring consisting solely of S, N, and C

F740 Thiazines



00241 00243 00245

F750 Other ring of 6 or more members consisting solely of S, N, and C

F760 Ring consisting solely of O, S, N, and C

F799 Poly

Part G: Carbocyclic Ring Systems

Topics Covered in Part G:

G0: General descriptors
- G00: Unsubstituted or non-specific substitution pattern disclosed
- G01: Number of C atoms substituted in an unfused aromatic ring
- G02: Number of C atoms substituted in an aromatic ring fused to another carbocycle
- G03: Substituents on an alicyclic ring that is either unfused or that is fused to a carbocyclic ring

G1: Unfused aromatic ring(s)

G2: 2-ring carbocyclic systems with at least 1 aromatic ring

G3: 3-ring carbocyclic systems with at least 1 aromatic ring

G4: Carbocyclic systems with 4 or more ring, at least 1 of which is aromatic

G5: Unfused alicyclic ring

G6: 2-ring alicyclic systems

G7: 3-ring alicyclic systems

G8: Alicyclic systems with 4 or more rings

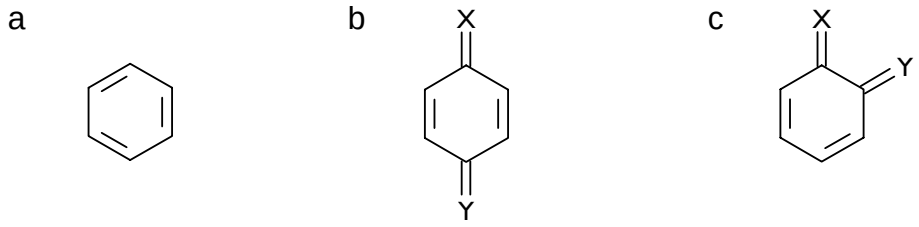
Introduction

Part G: codes refer to carbocyclic ring systems, i.e. ring systems consisting solely of C atoms. The codes in Part G: are hierarchical. In most cases the first digit indicates whether the ring is aromatic or alicyclic, as well as the number of rings in the system; the second digit indicates the type of ring system present; and the third digit indicates the state of hydrogenation. A final "0" means that the code covers all hydrogenation states; "1" means fully unsaturated; "2" means intermediate saturation; and "3" means fully saturated.

In addition to the rings specified in the code definitions for Part G:, each code also covers rings in which one or more C atoms have been replaced by heteroatoms other than O, S, N, Se, or Te. These substituted elements in the ring code definitions are discussed on page 72 in the section entitled "Ring Heteroatoms Other than O, S, and N".

Rings Coded As Aromatic

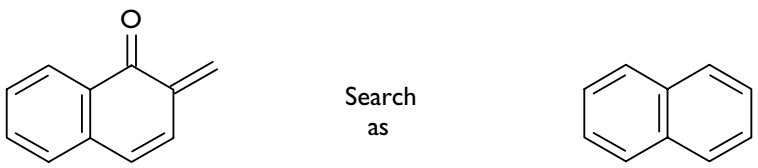
Ring systems are considered ***aromatic*** if they consist only of C atoms and if one or more of the following rings is present in a fused or unfused state:



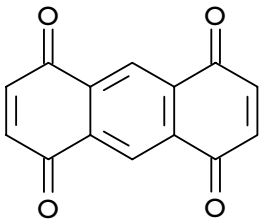
X and Y in the structures above may be any elements, and they may be bonded to additional atoms. Quinonoid derivatives (b) and (c) above are searched with the ring code for the "equivalent" parent aromatic ring (a). Thus the term **aromatic** in this manual is nearly synonymous with fused or unfused benzene, except that it includes the fused or unfused quinonoid derivatives of benzene shown above. Aromatic rings are searched using codes in Sets G1: through G4:.

Some example ring systems are shown below, each accompanied by its aromatic "equivalent", i.e. the ring system whose ring code it receives, as well as some of the additional chemical codes that are applicable. Note that the quinonoid substituents are counted and coded like any other aromatic substituents.

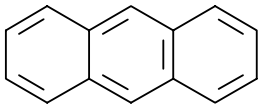
G100 (also search G013, L951)



G221 (also search G022, H720, L952)

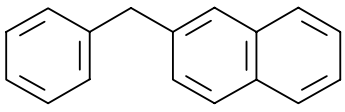


Search as

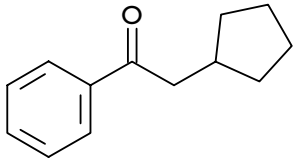


G331 (also search G022, G029, L951, L999)

The code G100 is only used for structures with one or more unfused aromatic rings, and only when no other carbocyclic ring systems are present. Thus G100 retrieves simple aromatic ring derivatives or structures that consist of one or more unfused aromatic rings and one or more heterocyclic ring systems. If unfused aromatic rings are present in a structure with other carbocyclic ring systems, one of the codes G111, G112 or G113 is applicable instead of G100. For example, a search for either of the two structures shown below is conducted using the code G111, rather than the code G100.



G111 (also G221)

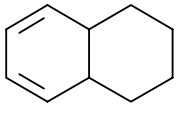


G111 (also G553)

Unless specifically excluded, codes also cover hydrogenated derivatives, provided that at least one ring in the system is aromatic, e.g. G310 includes fluorene, 1,2-dihydro-, and 1,2,3,4-tetrahydrofluorene.

Alicyclic Ring Systems

All carbocyclic ring systems that are not aromatic are searched using a code from Set G5: to G8:. For example, the structure shown below is searched using the code G670.



Notes on Coding Substituents of Carbocyclic Rings

1. Each ring in a fused carbocyclic system is considered individually when selecting codes for substituents.

2. The substituent number in codes from Subsets G01: to G03: refers to the number of C atoms in each ring that bear substituents. The second set of numbers (in parentheses on the coding sheet) refers to the relative positions in the ring of the C atoms bearing substituents.

3. If only one atom of a fused benzene ring in a carbocyclic system bears substituent(s), the position of the substituent(s) can be specified as adjacent (alpha) or not adjacent (beta) to ring fusion using the codes G020 and G021, respectively.

4. If two or three C atoms in an unfused aromatic ring bear substituents, the relative positions of the substituents can be specified. For example, G015 indicates that positions 1, 2, and 4 are substituted. Note that the numbers "1, 2 and 4" refer to relative positions in the ring, not atom numbers from a conventional numbering scheme. Also note that quinonoid substituents are counted and coded like any other aromatic substituents.

5. If only one atom in an unfused alicyclic ring bears substituents, the code G030 is used. If only one atom of a fused alicyclic ring in a carbocyclic system bears substituents, the position of the substituent can be specified as adjacent (alpha) or not adjacent (≥ beta) to ring fusion using the codes G031 and G032, respectively.

6. If there are two atoms in a fused or unfused alicyclic ring that bear substituents, the relative positions of the substituted C atoms can be specified. For example, **G034** means that there are substituents on two adjacent alicyclic C atoms. The lowest possible numbers are always assigned to the substituents.

7. Spiro groups are considered one substituent. A ring spiro to a carbocycle is additionally assigned one of the carbocyclic spiro codes **G041** (aromatic) or **G052** (alicyclic), depending on the type of carbocyle to which the ring is spiro.

Subheadings Applicable to Codes in Part G:

Subsets G00:, G01:, G02:, G03:, G06:, and G11: are applicable to subheadings M1 to M4 from 1981 forward. All other codes apply to M0 from 1963 and to M1, M2, M3 and M4 from 1970 forward.

G0: GENERAL DESCRIPTORS

G00: Unsubstituted or non-specific substitution pattern disclosed

G000	Unsubstituted carbocyclic system	(1981)
G001	Non-specific substitution pattern disclosed on benzene	(1981)
G002	Non-specific substitution pattern disclosed on fused aromatic ring	(1981)
G003	Non-specific substitution pattern disclosed on fused or unfused alicyclic ring	(1981)

G01: Number of substituted C atoms in an unfused aromatic ring

G010	1	(1981)
G011	2; relative positions in ring = 1 and 2	(1981)
G012	2; relative positions in ring = 1 and 3	(1981)
G013	2; relative positions in ring = 1 and 4	(1981)
G014	3; relative positions in ring = 1, 2, and 3	(1981)
G015	3, relative positions in ring = 1, 2, and 4	(1981)
G016	3, relative positions in ring = 1, 3, and 5	(1981)
G017	4	(1981)
G018	5 or 6	(1981)
G019	Poly	(1981)

G02: Number of substituted C atoms in an aromatic ring fused to another carbocycle

G020	1; C is adjacent (alpha) to ring fusion	(1981)
G021	1; C is not at or adjacent (beta) to ring fusion	(1981)
G022	2	(1981)
G023	3	(1981)
G024	4	(1981)
G029	Poly	(1981)

G03: Number of substituted C atoms in an alicyclic ring that is either unfused OR is fused to a carbocyclic ring

G030	1 C atom of unfused alicyclic ring bears substituent(s)	(1981)
G031	1 C atom of fused alicyclic ring bears substituent(s); C is adjacent (alpha) to ring fusion	(1981)
G032	1 C atom of fused alicyclic ring bears substituent(s); C is not at or adjacent (C is ≥beta) to ring fusion	(1981)
G033	2; relative positions in ring = 1 and 2	(1981)
G034	2; relative positions in ring = 1 and 3	(1981)
G035	2; relative positions in ring = 1 and 4	(1981)
G036	3	(1981)
G037	4 or more	(1981)
G038	2 substituents are on the same alicyclic C	(1981)
G039	Poly	(1981)

G04: Other descriptors for aromatic rings

G040	Non-specific disclosure of a carbocyclic aromatic ring
G041	A ring is spiro to an aromatic ring system

Note G041 is assigned when a ring is spiro to an alicyclic ring that is fused to benzene.

G05: Other descriptors for alicyclic rings

G050 Non-specific disclosure of unfused alicyclic ring

G051 Non-specific disclosure of polycyclic alicyclic ring system

G052 A ring is spiro to an alicyclic ring system

Note G041 is assigned when a ring is spiro to an alicyclic ring that is fused to benzene; G052 is not assigned in such cases, except if the first ring is alicyclic, when both codes would be used.

G060 Substituent(s) at the point of carbocyclic ring fusion (1981)

G1: UNFUSED AROMATIC RING(S)

G100: Unfused aromatic ring(s) present, no other carbocyclic ring systems are present

G11: Number of unfused aromatic rings in a structure with other carbocyclic system(s)

G111 1 benzene ring (1981)

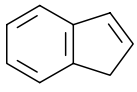
G112 2 benzene rings (1981)

G113 3 or more benzene rings (1981)

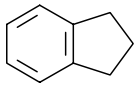
G2: CARBOCYCLIC SYSTEMS WITH 2 RINGS, AT LEAST ONE OF WHICH IS AROMATIC

G200 3- or 4-membered alicyclic ring fused to aromatic ring

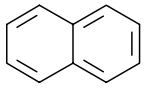
G211 Indene



G212 Indane

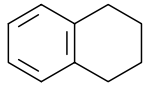


G221 Naphthalene

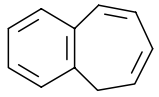


G222 Dihydronaphthalene (excluding 4a and 8a)

G223 Tetralin



G230 Benzocycloheptatriene



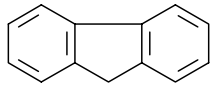
G240 Ring with 8 or more members fused to aromatic ring

G299 Poly

G3: CARBOCYCLIC SYSTEMS WITH 3 RINGS, AT LEAST ONE OF WHICH IS AROMATIC

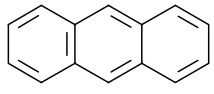
G300 System that includes 3- or 4-membered ring(s), or a 5:5:6 system

G310 Fluorene



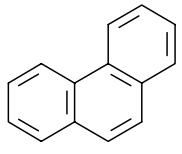
G320 Other 3-ring systems with one 5-membered ring, one 6-membered ring, and one ring with 6 or more members

G331 Anthracene (excluding rings covered by G332)



G332 Di- and polyhydroanthracenes

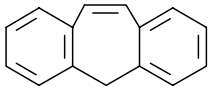
G341 Phenanthrene (excluding rings covered by G342)



G342 Dihydrophenanthrene, polyhydrophenanthrene

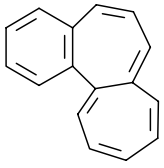
G350 Other 6:6:6 systems

G360 Dibenzo(a,d)cycloheptatriene



G370 Other 3-ring systems with two 6-membered rings and one ring with 7 or more members

G380 Benzo(a)heptalene



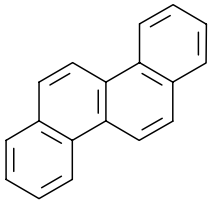
G390 Other 3-ring systems with one 6-membered ring and two rings with 7 or more members each

G399 Poly

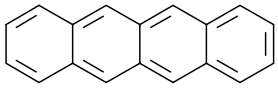
G4: CARBOXYLIC SYSTEMS WITH ≥4 RINGS, AT LEAST ONE OF WHICH IS AROMATIC

G400 4-ring system with at least one 3-, 4- or 5-membered ring

G410 Chrysene



G420 Naphthacene



G430 Other 6:6:6:6 systems

G440 Other 4-ring systems

G450 6:6:6:6:6 systems

G460 Other 5-ring systems

G470 6 or more rings, all with 6 members

G480 Other systems with 6 or more rings

G499 Poly

G5: UNFUSED ALICYCLIC RINGS

G530 Cyclopropane, cyclopropene

G541 Cyclobutadiene

G542 Cyclobutene

G543 Cyclobutane

G551 Cyclopentadiene

G552 Cyclopentene

G553 Cyclopentane

G561 Cyclohexadiene

G562 Cyclohexene

G563 Cyclohexane

G571 Cycloheptatriene

G572 Cycloheptene, cycloheptadiene

G573 Cycloheptane

G581 Alicyclic ring with 8 or more members and 2 or more multiple bonds

G582 Alicyclic ring with 8 or more members and 1 multiple bond

G583 Alicyclic ring with 8 or more members and no multiple bonds

G599 Poly

G6: ALICYCLIC SYSTEMS WITH 2 RINGS

G600 1 ring with 3 members and another ring with 3 or more members

G610 1 ring with 4 members and another ring with 4 or more members

G621 Norbornadiene

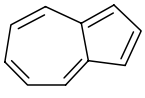
G623 Norbornene

G623 Norbornane

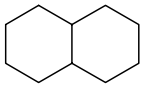
G630 Pentalene

G640 1 ring with 5 members and another ring with 6 members

G650 Azulene, optionally unsaturated



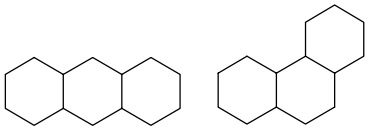
G670 Decalin, non-aromatic unsaturated derivatives



G680	Other systems with two 6-membered alicyclic rings
G690	1 ring with 6 members and another ring with 7 or more members
G695	2 rings, each with 7 or more members
G699	Poly

G7: ALICYCLIC SYSTEMS WITH 3 RINGS

G700	3-ring systems with at least one 3- or 4-membered ring
G710	2 rings with 5 members each and 1 ring with 5 or more members
G720	1 ring with 5 members, and 2 rings with 6 or more members
G730	Anthracene, phenanthrene (non-aromatic)



03618 03619

G740	Other 3-ring systems with 6 members in each ring
G750	2 rings with 6 or more members and 1 ring with 7 or more members
G799	Poly

G8: ALICYCLIC SYSTEMS WITH 4 OR MORE RINGS

G800	Four 6-membered rings
G810	Other 4-ring systems
G820	5-ring systems
G830	Systems with 6 or more rings
G899	Poly

Note G810 does not include cyclopenta(a) phenanthrene, i.e. the steroid nucleus, which is coded in Parts S:, T:, and U:.

8 Functional Groups

8 Functional groups

Introduction

The codes in Parts H:, J:, K:, and L: are used to describe ***functional groups*** (i.e. predefined patterns of atoms and bonds) in organic compounds. The common characteristics of all functional group codes are discussed first, followed by the common characteristics of Parts H: and J:, the common characteristics of Parts K: and L:, and finally the individual code Parts themselves.

Notes and Guidelines Pertaining to All Functional Groups

General Notes on Functional Group Codes

1. Three challenges of selecting functional group codes are:

   a choosing between several codes that appear to be equally applicable to a particular functional group;

   b finding codes that match complex arrangements of functional group atoms; and

   c knowing when enough codes have been selected for a complex functional group

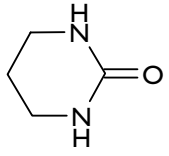
   Many codes in Parts H: through L: may appear to accurately describe a particular functional group in a structure, and yet are not used when coding that structure because of the associated coding rules. For example, a carbonyl group (>C=O) is found in many code definitions, but all of these codes are not used each time a structure with a carbonyl group is to be coded. Instead, coders use the **least possible number of codes** to describe functional groups. Rules for selecting functional group codes are given in the following section, "Choosing Applicable Functional Group Codes In Parts H:, J:, K:, and L:".

2. One-character **symbols** are used as a chemical shorthand for groups of atoms that are coded similarly. These symbols are defined as follows:

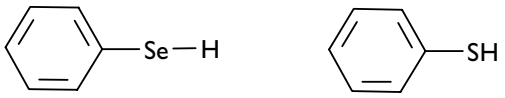
   T = O, S, Se, Te
   U = O, S, Se, Te or N
   V = Heteroatom other than O, S, Se, Te or N
   W = O, S, Se, Te, N, C or halogen in Parts: H: through L:
   = Any heteroatom in Parts other than H: through L:
   X = Halogen (F, Cl, Br, I)
   Y = O, S, Se, Te, N or halogen
   Z = B, Si, P or As

3. Functional group codes from Parts H:, J:, K:, and L: are usually not assigned to groups having **more than one atom in the same ring system**.

   Exceptions:
   the code L710, certain designated codes in Set H7: ("Olefinic/acetylinic groups"), Set K1: ("X-Y and X=Y bonds"), Set L8: ("Sugars and derivatives"), and Set L9: ("Functional groups in rings"). Thus the compound shown below is searched as J521 ("Oxo bonded to C in heterocyclic ring") and L921 ("Urea in a ring"), but not L432 ("Urea").



4. A compound with an **Se or Te** atom receives **all** codes from Parts H: through L: that would be applicable to the compound if Se or Te were replaced by S. In addition, the Se and Te atoms are themselves assigned codes from Part B:. For example, the code H494 ("-SH group bonded to an aromatic ring") is assigned to both of the structures shown below.



The structure containing the Se atom is additionally assigned the code B534 ("Selenium in a carbocyclic compound").

5. **Functional groups in organic ions or ligands** that are bonded to a metal atom via C are searched in Parts H: through L: in one of two ways:

   a as if the metal were replaced by H when the metal is bonded to only one C atom, or

   b as if the metal were replaced by C when the metal is bonded to more than one C atom.

6. **Functional groups bonded to B, Si, P, and As** (i.e. "Z" atoms) are coded in Subset B79:. Some of these functional groups are additionally coded in Parts H: through L:. Guidelines for searching for functional groups bonded to Z atoms are found on page 48.

7. Each ion or ligand bonded to a **metal** atom is searched as if only one formula unit of the ion or ligand is present. Thus the compound shown below is searched as H181 ("One aliphatic amino") and H481 ("One aliphatic -OH group"), i.e. as the Mg salt of $HOCH_2CH_2NH_2$.

$$H_2NCH_2CH_2O\text{-}Mg\text{-}OCH_2CH_2NH_2$$

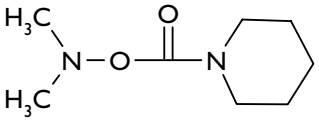
More complete guidelines for coding compounds containing both metals and functional groups are discussed in Part A: in the section entitled "Guidelines for Coding Compounds Containing Metals", beginning on page 40.

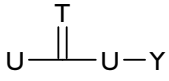
Choosing Applicable Functional Group Codes In Parts H:, J:, K:, and L:

The following guidelines can be used for choosing applicable functional group codes for a structure.

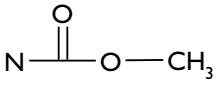
1. Locate all non-ring O, S, Se, Te, N, and halogen atoms in organic compounds, and all C atoms that are not ring members and not members of carbon chains. (The term "carbon chain" is defined on page 174). These atoms are likely to form functional groups coded in Parts H: through L:. If these atoms are bonded to N, S, Se, or Te in a ring, the ring atom may also form part of the functional group, unless the applicable code definition states that no atom in the functional group can be a ring member. Furthermore, the code L710 or some of the codes from Sets H7:, K1:, L8: or L9: may be applicable to functional groups with more than one atom in the same ring system.

2. If a functional group contains a bond between heteroatoms, first look for matching codes in Part K:. If no bonds between heteroatoms are present in the structure, begin looking for applicable codes in Part L:. If the atoms and bonds of the functional group are not completely described by the first matching code, then look for further matching codes **in the order that the Sets appear**, first in Part K:, then in Part L:. By convention, the **minimum number of codes** are assigned which between them identify every atom of the group and all bonds between the atoms. At each stage, the largest possible matching group is chosen, even if this means that a given atom of a large fragment is covered by more than one code. For example, consider the following structure and the functional group codes that describe it.

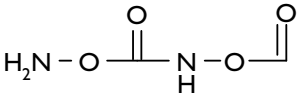




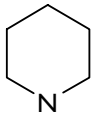
L410



L463



K820

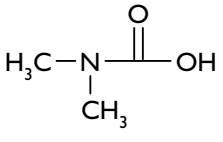


H211 and of course F433 and F011

3. Codes occuring later in a Part or Set are often less specific, or cover fewer atoms, than the first applicable code(s). Even though these less-specific codes may also appear to be applicable to the functional group, they are not used if they add no new information about the structure. For example, a structure coded Y-C(=T)-X (L511) would not be further coded as -C(=T)-X (L512), nor would a compound coded -O-S-N (K340) be further coded as Y-S-N (K352).

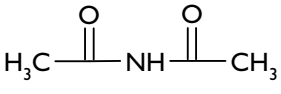
4. If the definition of an applicable code includes **variable element symbols** (i.e. T, U, V, W, X, Y and/or Z), additional codes are used to identify the elements represented by the symbols, unless the first code states that this should not be done. In the structure shown on the previous page, for example, the entire functional group was actually represented by the single code L410, but the codes K820 and L463 were needed to identify the elements represented by the symbols T, U, and Y in the definition of L410. On the other hand, note 5 on page 162 states that no additional codes should be used to identify the "T" atom in functional groups searched with the code L620.

5. If all functional groups in a structure have not been fully defined by codes from Parts K: and L:, Parts H: and J: can be scanned for relevant codes. Functional groups searched in Parts H: and J: include: amine, hydroxy, mercapto, ether, thioether, halogen, C=C, C≡C, carboxylic acid, thiocarboxylic acid, ester, thioester, amide, thioamide, aldehyde, thioaldehyde, oxo, and thioxo groups. Although these functional groups are easily identifiable, it is often the case that they are coded in Part K: and/or Part L: instead of in Part H: or Part J: because of the atoms to which they are bonded. The structures shown below, for example, are assigned Part L: codes instead of similar codes in Part H: or Part J:.



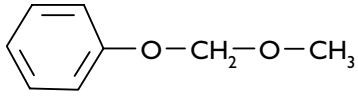
L463

not amine, amide carboxyl acid or ester



L532

not amide



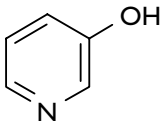
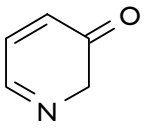
L660

not ether

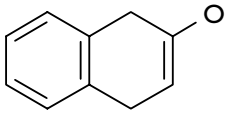
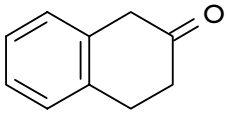
Tautomerism

The following rules pertain to the coding of certain tautomeric structures, the presence of which can significantly affect the coding of functional groups.

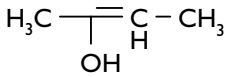
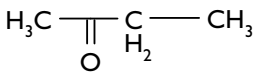
1. When keto-enol (thioxo-thiol) tautomerism is possible, the structure is searched in the keto (thioxo) form, unless the -OH (-SH) group of the enol (thiol) form is bonded to a *fully conjugated carbocyclic ring*, e.g. benzene.

a  search as 

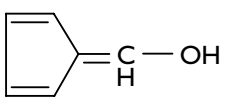
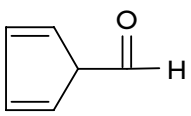
J521 ("Oxo bonded to heterocyclic C"), NOT H421 ("-OH bonded to heterocyclic C") and F432 ("Di or tetra hydro pyridine"), NOT F431 ("pyridine")

b  search as 

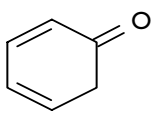
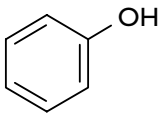
J561 ("Oxo bonded to alicyclic C"), NOT H461 ("-OH bonded to alicyclic C") and G223 ("Tetralin"), NOT G222 ("Dihydronaphthalene")

c  search as 

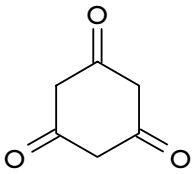
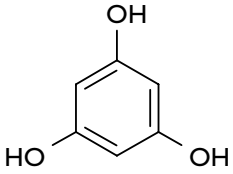
J581 ("Oxo bonded to aliphatic C"), NOT H481 ("-OH bonded to aliphatic C")

d  search as 

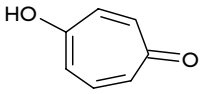
J451 ("Aldehyde bonded to alicyclic C"), NOT H720 ("Exocyclic C=C").

e  search as 

H443 (“1 aromatic -OH”), NOT J561 (“exo bonded to alicyclic C”)

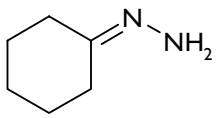
f  search as 

H443 (“3 aromatic -OH groups”), NOT J563 (“≥3 oxo bonded to alicyclic C”).

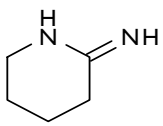
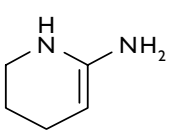
Note:  is treated as a fully conjugated carbocyclic ring, so the hydroxy group is NOT tautomerised and is coded as H401, H461 and J561.

---

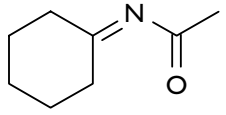
2. An **imine** group (>C=N-) in which the **C atom is part of a ring** is tautomerized and searched in the =C-NH- form if such tautomerization causes the N atom to form an amine or amide group. If the tautomerization leads to an N group other than amine or amide, the imine form is coded. (The term “amine”, as used in the chemical coding system, is defined on page 119.)



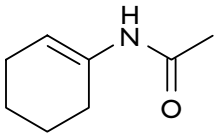
K630 (“N-N=C”) NOT K640 (“N-N”) or K534 (N=N)

 search as 

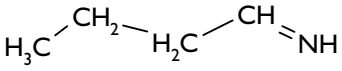
H121 (“1 amine bonded to a heterocycle”), NOT L355 (“C=N”)



search as



J361 ("1 amide group bonded to an alicycle via N") NOT L355



L355 ("C=N")

Negation Codes

The two-character codes found on the coding sheet at the beginning of each Set in Parts H:, J:, K:, and L: are called ***negation codes,*** or alternatively ***essential codes***. (To avoid confusion, the term **negation code** will be used throughout this section to indicate **negation/essential codes,** as this better describes their use in searching. The term **essential code** is more descriptive when discussing the coding of a patent document.) Negation codes are searched negatively by users to eliminate some of the "false drops", i.e. irrelevant answers, that are retrieved by chemical code searches. Whereas the four-character chemical codes are searched with the **LINK** operator, negation codes are negated using the operator **LINK NOT**.

A negation code is assigned by coders when the functional group represented by that code is present in ALL variations of a disclosed compound in a patent document. Conversely, users negate a negation code when it is known with certainty that the corresponding functional group is not present in any variation of the structure(s) being searched, thereby eliminating irrelevant records from the search results. If there is any possibility that a particular functional group may be present in variable or undefined **R** groups, the negation code for that functional group should not be negated.

The four-character chemical codes, e.g. H441 ("One aromatic hydroxy"), should not be negated. This stems from the fact that the Chemical Codes listed together in a single code subfield usually represent many different possible structures disclosed in a patent document. Thus, negating the Chemical Code of an unwanted functional group is likely to eliminate references with code subfields that include codes for compounds of interest and codes for compounds that are not of interest.

Negation codes, on the other hand, are designed for removing false drops while not getting rid of valid answers. Suppose, for example, that it is known with certainty that a structure to be searched has **no hydroxy substituents**. It is desirable, therefore, to eliminate all patents that **only** disclose compounds **with** hydroxy groups. It is not desirable, however, to eliminate all patents that disclose compounds with hydroxy groups, because these patents may additionally disclose the compounds being searched. Since the only records that receive an H4 negation code are those in which **every** compound coded in that record has a hydroxy substituent, negating all records that contain the code H4 will not eliminate any records that refer to compounds of interest.

If a search structure contains none of the functional groups defined in Parts K: and L:, the negation code **K0** can be negated in place of all of the individual K: and L: negation codes. Outside of Parts H:, J:, K:, and L:, the only negation codes available are **M1**, a negation code used for **groups linking ring systems**, and **V0**, a negation code for **natural products**. Although negation codes are not linked to particular time ranges, they are usually negated at the end of the search strategy.

Negation codes applicable Parts H and J:

H1	Amine	1970
H2	Ring tert.nitrogen	1970
H3	Nitro	1970
H4	Hydroxy or mercapto	1970
H5	(Thio) ether	1970
H6	Halogen	1970
H7	Olefinic/acetylenic	1970
H8	Ether or hydroxy	M3 = 1970 others = 1980
H9	Thio-group in Part H:	1981

J0	Carboxylic acid, ester, or amide	M3=1970 others = 1980
J1	(Thio) carboxylic acid	1970
J2	(Thio) carboxylic ester	1970
J3	(Thio) carboxylic amide	1981
J4	(Thio) aldehyde	1981
J5	Oxo; thioxo	1970
J6	(Thio) aldehyde or (thio) amide	1970-1981
J9	Thio-group in Part J:	1981

Negation codes applicable Parts K and L:

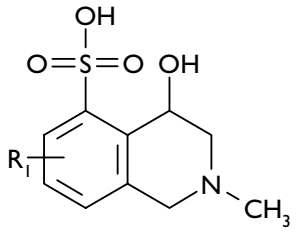
K0	K or L group (see below)	1970
K1-L9	Relevant K1-L9 group	1981

Other Negation Codes

M1	Ring Linkage	1981
V0	Natural Products	1981

Example 1

Suppose the following compound were disclosed in a patent document:



where R_1 = halogen, NO_2 , or OCH_3

The following groups are "essential" to this structure, and thus their two-character negation codes are assigned by coders:

H1 amine
H2 tertiary ring N
H4 hydroxy, mercapto
H8 hydroxy or ether
K0 a group from Part K: or Part L: is essential
K4 S=O or S-O

Note that the variable groups in this structure - halogen, nitro, and methoxy - are not essential (i.e. not present in all structures disclosed in the patent document), and are thus not assigned negation codes.

A search strategy formulated to find structures like the one shown above would include all negation codes (including M1 and V0) EXCEPT the following:

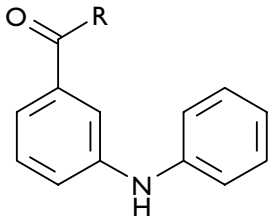
H1 amine
H2 tertiary ring N
H3 nitro
H4 hydroxy, mercapto
H5 ether, thioether
H6 halogen
H8 hydroxy or ether
K0 a group from Part K: or Part L: is essential
K4 S=O or S-O

Thus, the user does not negate any functional group that could conceivably be present in the structure being searched, so that no relevant references are lost.

Note If the R group in the molecule above was undefined in the search query, no negation codes could be used, because it would not be known which functional groups could potentially be present.

Example 2

Suppose the following compound were disclosed in a patent document:



where R_1 = OH, OCH_3 , or NH_2

The following groups are "essential" to this structure, and thus their two-character negation codes would be assigned by coders:

H1 amine
J0 carboxylic acid or derivative
M1 ring linkage

A search strategy formulated to find structures like the one shown above would include all negation codes EXCEPT the following:

H1 amine
J0 carboxylic acid or derivative
J1 carboxylic acid or thiocarboxylic acid
J2 ester or thioester
J3 amide (thioamide)
M1 ring linkage

K0 can be negated when searching for this compound, which is equivalent to negating all negation codes in Parts K: and L:.

Notes and Guidelines Pertaining to Common Functional Groups

Introduction

This section lists coding guidelines that Parts H: and J: have in common. Functional groups coded in Part H: do not have a carbonyl or thionyl group: i.e.amine, hydroxy, mercapto, ether, thioether, halogen, C=C, and C≡C. Functional groups coded in Part J: do have a carbonyl or thionyl group: i.e. carboxylic acid, thiocarboxylic acid, ester, thioester, amide, thioamide, aldehyde, thioaldehyde, oxo, and thioxo groups.

Notes and guidelines applicable to all functional groups were discussed earlier in this chapter (pages 102-103), and those applicable only to the individual Parts H: and J: are discussed later in their respective sections of this chapter.

Notes on Codes in Parts H: and J:

1. Most codes in Parts H: and J: are hierarchical: the first digit indicates the type of functional group represented; the second digit indicates the type of group to which the functional group is bonded (i.e. the "attachment group"); and the third digit indicates the number of times a particular functional group/attachment group combination is present in a structure.

2. Thiocarboxylic acids, thioxo groups, and thioethers have their own codes; they do not receive the codes of their oxygen analogues. For example, a compound with one thiocarboxylic acid group bonded to a heterocyclic ring is coded J191 ("Thiocarboxylic acid group bonded to a heterocyclic ring"). Other **thio functional groups** receive the codes of their corresponding oxygen analogues, as well as a general code indicating that these codes are being used for the thio analogue. For example, a compound that has a single thiocarboxylic ester group bonded via thionyl to a heterocyclic ring is coded J211 ("One ester or thioester group bonded via carbonyl or thionyl to a heterocyclic ring") and J290 ("Thioester").

3. Some functional groups that normally receive codes from Part H: or Part J: do not receive those codes when they are bonded to **B, Si, P, or As** ("Z" atoms). Rules for coding functional groups bonded to Z atoms are given in the Notes pertaining to Part B: codes, beginning on page 47.

Choosing Applicable Functional Group Codes in Parts H: and J:

1. Most functional group codes in Parts H: and J: indicate both the type of group to which the functional group is attached, and the total number of occurences of that particular functional group/attachment group combination. There are four types of **attachment groups** that are coded:

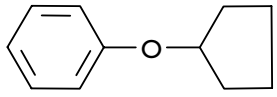
   a Heterocyclic ring
   b Aromatic ring
   c Alicyclic ring
   d Aliphatic chain

   In most cases, the **atom of attachment** in these groups must be C. However, nitro groups are searched using codes in Set H3: when bonded to either a C atom or a N atom of the attachment group.

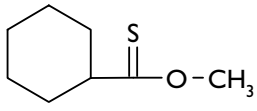
2. Codes of **multivalent functional groups** - i.e. amines, ethers, esters, amides, and their sulphur analogues - are chosen according to the attachment group with the highest priority. The priority order used is:

   Heterocyclic ring > Aromatic ring > Alicyclic ring > Aliphatic chain

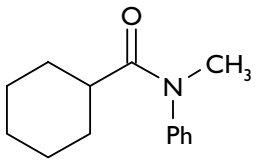
   Thus, the compound shown below is coded H541 ("One aromatic ether"), but not H561 ("1 alicyclic ether").



   Non-symmetrical, multivalent functional groups - i.e. esters, amides, and their sulphur analogues - are additionally coded according to the **point of attachment** of the highest priority group. Thus, a compound with one amide group bonded to an aliphatic C atom on the carbonyl side and to a heterocyclic C atom on the N side would be coded J321 ("One amide group bonded to a heterocyclic ring via N"). If the attachment groups on both sides of the functional group are of equal priority, the code for the carbonyl (or thionyl) attachment has a higher priority and is assigned.

a 

   J251 ("1 ester group bonded to an alicyclic ring via carbonyl)
   J290 ("Thioester present")

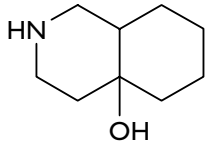
b 

J011 ("1 COOH group or derivative")
J341 ("1 amide group bonded to an aromatic ring via N")

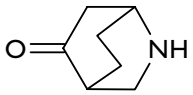
3. The codes in many Sets in Parts H: and J: refer to monovalent functional groups bonded to heterocyclic, aromatic, and alicyclic rings. Only the ring to which the functional group is bonded is taken into consideration when choosing these codes; further rings fused to the substituted ring have no effect on the choice of an applicable code. For example, if a structure contains a single amine bonded to an aromatic ring that is fused to a heterocyclic ring, the amine is searched as H141 ("One amine bonded to an aromatic C atom"). This differs from the coding of ring substituents in Parts D: through G:, where, for example, a substituent on benzene could be searched in Subset D02: ("Carbocyclic ring substituent in a fused ring heterocyclic system"), Subset G01: ("Substituent on unfused benzene"), or Subset G02: ("Substituent on benzene fused to another carbocycle"), depending on the presence of other rings in the same ring system as the substituted benzene ring.

4. A compound with an **Se or Te** atom receives **all** codes from Parts H: and J: that would be applicable to the compound if Se or Te were replaced by S. In addition, Se and Te atoms are themselves coded in Part B:.

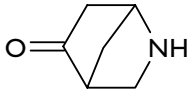
5. If the C atom to which a functional group is bonded is a member of more than one ring, the highest priority ring (according to the order of priority given above in Note 2) determines which functional group code is assigned. **Exception:** if the atom bonded to the functional group is a non-angular C atom in a bridged ring system, and if the C atom could be seen as being a member of different sized rings, the smallest ring size is chosen, even if it is of lower priority than the larger ring, (see example below).

a 

H421 ("1 OH bonded to a heterocyclic ring"), but not H461 ("1 OH bonded to an alicyclic ring")

b 

J521 (“1 oxo group bonded to a heterocyclic ring”), but not J561 (“1 oxo group bonded to an alicyclic ring”)

c 

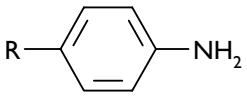
J561 (“1 oxo group bonded to an alicyclic ring”), not J521 (“1 oxo group bonded to a heterocyclic ring”) as the alicyclic ring is smaller than the heterocyclic.

Searching Structures with Variable or Unknown Substituents

It was pointed out in the previous section that most functional group codes in Parts H: and J: indicate both the type of group to which the functional group is bonded, and the total number of occurences of that particular functional group/attachment group combination. If known functional groups are present in a structure that also has unknown structural fragments (***R*** groups), each known functional group is considered separately, and the codes that would be applicable to that functional group for all definitions of **R** are 'OR'ed together. The resulting search strategy may retrieve some unwanted references, but not as many as a search strategy that completely omits codes for the known functional groups.

The following examples illustrate the practice of broadening the search strategy to compensate for unknown structural fragments. The search strategies in this series of examples are not formulated to describe all functional groups that might conceivably be present in the example structures. If that were the case, all functional group codes would be 'OR'ed together in the search strategy, because **R** and **R'** are unknown. Instead, the search logic in each example is constructed to include all codes that might possibly be applicable to the functional group(s) shown, i.e. those known to be present.

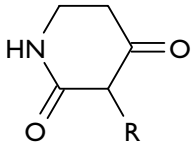
Example 1



R = anything

The code for the single amino group shown in this structure is H141 ("One aromatic amino group"). If **R** is optionally a second aromatic amino group, the code H142 must be searched as well; thus, H141 would be replaced by (H141 OR H142) in the search logic. If **R** could be anything, any number of aromatic amino groups might be present in this structure, and thus (H141 OR H142 OR H143) would replace H141 in the search logic.

Example 2



R = anything

If R is anything, any number of additional heterocyclic oxo groups may be present, and thus the search logic used to code the oxo groups shown in this structure would be: (J522 OR J523).

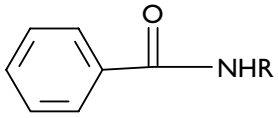
Note In the following examples, **R** and **R'** are groups bonded to the functional group in question by C atoms that are not doubly bonded to O, N, S, Se, or Te.

Example 3

$H_3C—CH_2—NRR'$

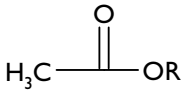
If **R** and **R'** do not consist of more amino groups, the search logic for the amino group shown would be: (H121 OR H141 OR H161 OR H181). This search logic covers all possible attachment groups that might be bonded to the N atom. If more amino groups may be present in **R** or **R'**, the search logic should be formulated with truncated Subset codes as follows: (H12: OR H14: OR H16: OR H18:). This search strategy indicates that any number of amino groups may be present, each bonded to any kind of attachment group.

Example 4



The search logic for the amide group shown would be: (J321 OR J322 OR J331 OR J332). Fewer possibilities exist for coding this amide group than for coding the amino group in Example 3, because the only codes that could possibly be assigned to the amide group shown are "Amide group bonded to aromatic ring via carbonyl", and "Amide group bonded to heterocyclic ring via N". All other possible definitions of **R** would result in amide attachments of lower priority than the one already known to be present, which means they would not be coded.

Example 5

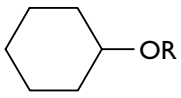


The search logic for the ester group in this structure would be formulated with truncated Subset codes as follows: (J22: OR J24: OR J26: OR J27:).

J22 covers the case of R being heterocyclic
J24 covers the case of R being aromatic
J26 covers the case of R being alicyclic
J27 covers the case of R being aliphatic, in which case the carbonyl-attached aliphatic group has a higher priority and is therefore the one that is coded

All of the codes in this example are truncated to allow for the presence of other similar groups in the undefined portion of the structure.

Example 6



The search strategy for the ether group would be formulated with truncated Subset codes as follows: (H52: OR H54: OR H56:). Codes for alicyclic and aliphatic definitions of **R** are not needed in the search strategy, because such groups would have a priority equal to or less than the alicyclic group already known to be bonded to the ether -O- shown. The codes are truncated because it is unknown how many other such ether groups are present in the undefined portion of this structure. In addition of R can be H, the codes H401 and H461 must be 'OR'ed to the H5 codes to allow for the alcohol.

Part H: Common Functional Groups Without >C=O, >S=O

Main Headings

H1: Amine
H2: Ring Tertiary Nitrogen
H3: Nitro
H4: Hydroxy, Mercapto
H5: Ether, thioether
H6: Halogen (F, Cl, Br, I)
H7: Olefin, Acetylenic

Introduction

Part H: codes refer to common functional groups that don't have a carbonyl or thionyl group. Some of the guidelines for using Part H: codes were given earlier in this chapter, both in the context of all functional group codes (Parts H: through L:, starting on page 102) and in the context of common functional group codes (Parts H: and J:, starting on page 112). Those general guidelines are not repeated in this section.

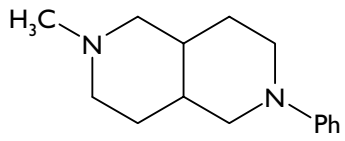
Amines and Non-amines

An ***amine*** is defined in the chemical coding system as a trivalent N atom having the following characteristics:

- N is **single-bonded** to three atoms, all of which are **C** or **H**, at least one of which is C
- N is not bonded to any C atom that is multiply bonded to a heteroatom, unless the C atom and the heteroatom are part of the same ring system
- if N is a ring member, **all three atoms of attachment are C**, and the three C atoms are not members of the same ring system

Aliphatic amines are coded in Set H1: and amines in which N is a ring member are coded in Set H2:. N atoms that do not meet the above criteria are called "non-amines", and are coded in code Sets other than H1:.

A *ring tertiary nitrogen* is a trivalent N atom in a ring, bonded to an atom other than H that is outside the N atom's ring system. Ring tertiary nitrogens may be either **amines or non-amines**, as defined above, and are coded in Set H2:. **Ring tertiary amines** additionally receive a code from Set H1:, the choice of which depends on the C group (i.e. heterocyclic, aromatic, etc.) bonded to the N atom from outside the N atom's ring system.

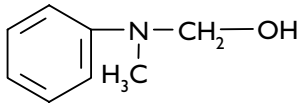


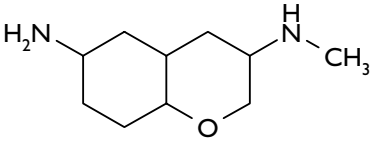
H202, H141, H181

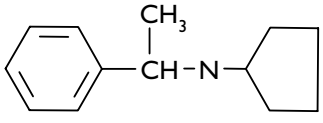
The presence of imine **tautomers** can significantly affect the coding of N atoms. An imine group (>C=N-) in which the C atom is part of a ring is tautomerized and searched in the =C-NH- form when such tautomerization causes the N atom to form an amine or amide group. If the tautomerization leads to an N group other than amine or amide, the imine form is searched.

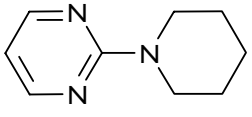
Examples of Amines and Non-amines

The following structures are coded as amines:

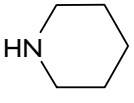




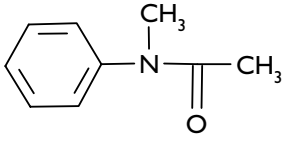




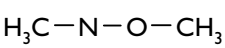
The following structures are NOT coded as amines:



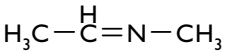
(no H: to L: code)



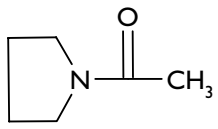
(J341)



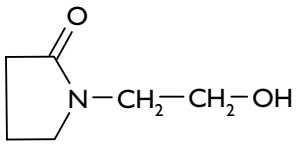
(K850)



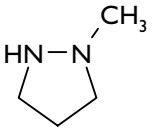
(L355)



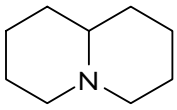
(J371, H211)



(H211, H481, J521, L941)



(H211)



(No H: to L: Code)

Halogens

The code Part used for coding a halogen atom depends on the atom(s) to which the halogen is bonded. Set H6: codes, defined later in this section, are assigned to every monovalent halogen atom bonded to a C atom that is not doubly bonded to O or S. The codes in Subset H60: are additionally used to identify X atoms in functional groups that are coded L330 and L353, and are searchable from 1963 forward (i.e. in the same search statement as the black codes from the coding sheet). The coding of halogens is discussed more completely in the section "Coding Rules for Polymers and Halogens", on pages 18-19.

Subheadings Applicable to Part H: Codes

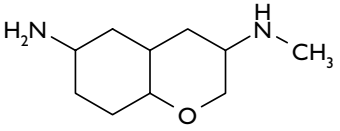
Unless otherwise specified, codes are valid in Subs M0 to M4 from 1963.

H1: AMINE

Note Amine codes are selected according to the rules given for multivalent functional groups in Note 2 on page 113.

H10: Type of amine

H100	One primary amine (M4 - 1970 M1-M3 - 1980)
H101	Two or more primary amines (M4 - 1970 M1-M3 - 1980)
H102	One or more secondary, non-ring amines (M4 - 1970 M1-M3 - 1980)
H103	One or more tertiary, non-ring amines (M4 - 1970 M1-M3 - 1980)



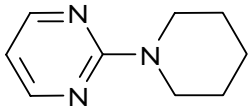
H100, H102, H121, H161

Notes on H10: codes

1. **Primary** amines are amines in which the N atom is bonded to two H atoms and one C atom. **Secondary** amines are amines in which the N atom is bonded to two C atoms and one H atom. **Tertiary** amines are amines in which the N atom is bonded to three C atoms and no H atoms.
2. The coding of **quaternary ammonium salts** is discussed on page 154. They are not assigned the code H103.

H12: Amine bonded to heterocyclic C

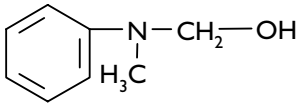
H121	One
H122	Two
H123	Three or more



H121, H201

H14: Amine bonded to aromatic C

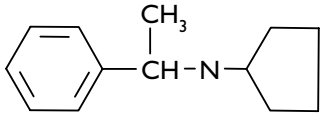
H141	One
H142	Two
H143	Three or more



H103, H141, H481 (+L640)

H16: Amine bonded to alicyclic C

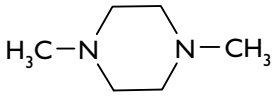
H161	One
H162	Two
H163	Three or more



H102, H161

H18: Amine bonded to aliphatic C

H181	One
H182	Two
H183	Three or more



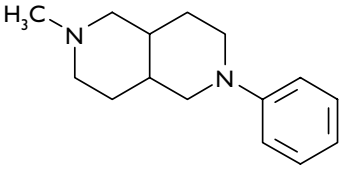
H182, H202

H2: RING TERTIARY NITROGEN

Note **Ring tertiary nitrogen** is defined on page 120 in the section entitled "Amines and Non-amines".

H20: Ring tertiary amine

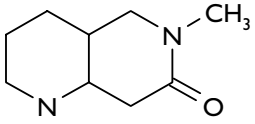
H201	One
H202	Two
H203	Three or more



H141, H181, H202

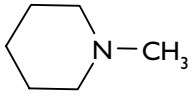
H21: Ring tertiary N, non-amine

H211	One
H212	Two
H213	Three or more

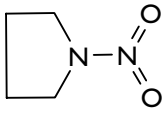


H211, J521, L941

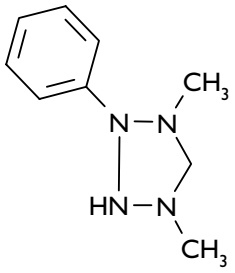
Examples of Ring Tertiary N Atoms

a 

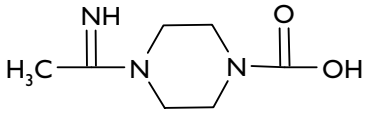
H211, H321, K610

b 

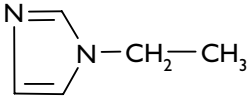
H181, H201

c 

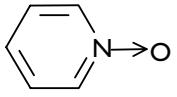
H213

d 

H212, L340, L463

e 

H181, H201

f 

K742 (but not H211)

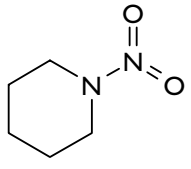
H3: NITRO

Notes on H3: codes

1. Thionitro (i.e. nitro groups with one or both O atoms replaced by S atoms), -S-NO2, and O-NO2 are coded in Part K: ("Bonds Between Heteroatoms"), not in Set H3:.
2. Nitramines are coded (K610 OR K600) (">N-NO2"), as well as in Subset H32: or Subset H38: as applicable.
3. Groups of formula R-C(=O)-NO2 or R-C(=S)-NO2 (R ≠ O, S, Se, Te, N or halogen) are assigned the following codes: a code in Subset H38:; a code in Subset J01: ("Number of -COOH derivatives)"; a code in Set J3: ("Amide (thioamide))"; and the code L520 ("-C(=O)-N=O or -C(=S)-N=O)". Such functional groups are specifically excluded, however, from the definition of the code K741.

H32: Nitro group bonded to heterocyclic C or N

H321	One
H322	Two
H323	Three or more



H321, H211, K610

H34: Nitro group bonded to aromatic C

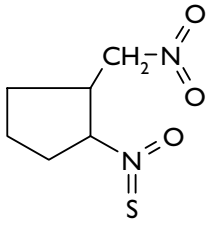
H341	One
H342	Two
H343	Three or more

H36: Nitro group bonded to alicyclic C

H361	One
H362	Two
H363	Three or more

H38: Nitro group bonded to aliphatic C or N

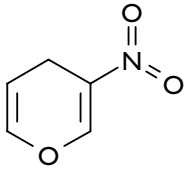
H381	One
H382	Two
H383	Three or more



H381, K320, K741

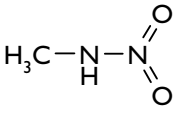
Further Examples Using H38: Codes

a



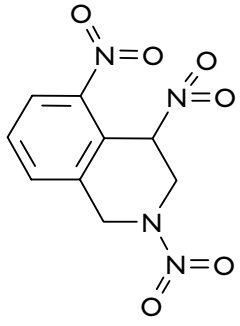
H321

b



H381, K610

c



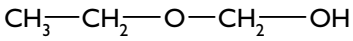
H211, H322, H341, K610

H4: HYDROXY AND MERCAPTO GROUPS

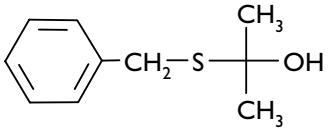
Notes on H4: Codes

1. The presence of keto-enol **tautomers** can significantly affect the coding of hydroxy groups. When keto-enol (thioxo-thiol) tautomerism is possible, the structure is coded in the keto (thioxo) form, unless the -OH (-SH) group of the enol (thiol) form is bonded to a fully conjugated carbocyclic ring, e.g. benzene. Examples of selecting codes for -OH tautomers are given in the section "Tautomerism", on page 106.

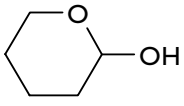
2. Compounds not searchable in Set H4: include: hemi-acetals, hemi-thioacetals, non-cyclic hemi-ketals and hemi-thioketals, and hydrates of aldehydes and thioaldehydes. Hemi(thio) acetals or ketals are coded in L6:, whilst hydrates of (thio)aldehydes or ketones are coded with H_20 removed as the parent aldehyde or ketone.

a 

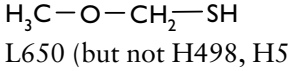
L660 (but not H401, H481, H581)

b 

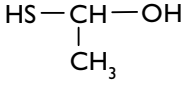
L650 (but not H401, H481, H598)

c 

H401, H421
(coded in H4: because it's cyclic)

d 

L650 (but not H498, H581)

e 

J471, J490 (but not L650) - code with H_20 removed, i.e. aldehyde

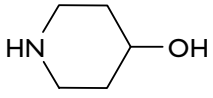
H40: Total number of -OH groups

H401	One	
H402	Two	(M3 - 1972; rest - 1981)
H403	Three	(M3 - 1972; rest - 1981)
H404	Four	(M3 - 1972; rest - 1981)
H405	Five or more	(M3 - 1972; rest - 1981)

Note: In previous manuals it was wrongly reported that H402: H405 were valid from 1970 in M3.

H42: -OH group bonded to heterocyclic C

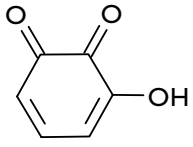
H421 One
H422 Two
H423 Three



H401, H421

H44: -OH group bonded to aromatic C

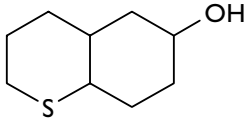
H441 One
H442 Two
H443 Three
H444 Four or more



H401, H441, L951

H46: -OH group bonded to alicyclic C

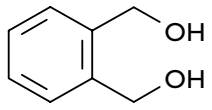
H461 One
H462 Two
H463 Three
H464 Four or more



H401, H461

H48: -OH group bonded to aliphatic C

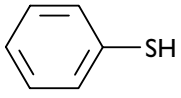
H481	One
H482	Two
H483	Three
H484	Four or more



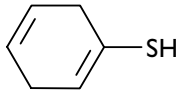
H402, H482

H49: -SH group

H492	One or more -SH groups bonded to heterocyclic C atom(s)
H494	One or more -SH groups bonded to aromatic C atom(s)
H496	One or more -SH groups bonded to alicyclic C atom(s)
H498	One or more -SH groups bonded to aliphatic C atom(s)

a 

H494

b 

J596 (but not H496)

H5: ETHER AND THIOETHER

Note Ether codes are selected according to the rules given for **multivalent functional groups** in Note 2 on page 113.

H52: Ether with -O- bonded to heterocyclic C

H521	One
H522	Two
H523	Three or more

H54: Ether with -O- bonded to aromatic C

H541	One
H542	Two
H543	Three or more

H56: Ether with -O- bonded to alicyclic C

H561	One
H562	Two
H563	Three or more

H58: Ether with -O- bonded to aliphatic C

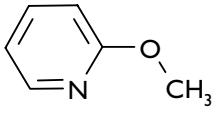
H581	One
H582	Two
H583	Three
H584	Four to eight
H589	Nine or more

H59: Thioether

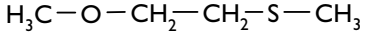
H592	One or more thioether group with -S- bonded to heterocyclic C
H594	One or more thioether group with -S- bonded to aromatic C
H596	One or more thioether group with -S- bonded to alicyclic C
H598	One or more thioether group with -S- bonded to aliphatic C
H599	More than one thioether group is present

Note Thioethers coded H599 are also assigned one or more of the codes H592 through H598 as applicable.

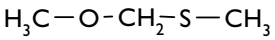
Examples Using H5: Codes

a 

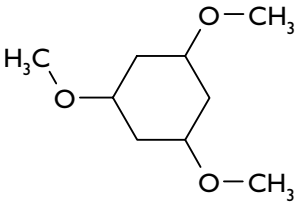
H521 (L941 may also have been coded see Note 2 on page 165)

b 

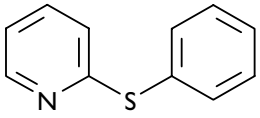
H581, H598

c 

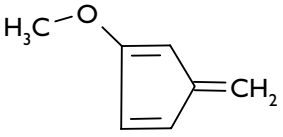
L650 (but not H481, H598)

d 

H563

e 

H592, (L943 may also have been coded - see Note 2 page 165)

f 

H561, H7200

H6: HALOGEN

Notes on H6: Codes

1. The codes in Set H6: are assigned to every monovalent halogen atom bonded to a C atom that is not doubly bonded to O or S. In addition, codes in Subset H60: have been used since 1963 to identify the halogen atom represented by **X** in groups coded L330 and L353. A complete discussion on coding halogens is found on page 14.
2. The group -CF3 is coded H685, not H601 ("Fluorine") and H699 ("Poly").
3. To retrieve all disclosures of specific halogens (in the type of halogen compounds searched with Set H6: codes), use the following search terms in Search Statement 1:

   (H601 OR H602 OR H603 OR H604 OR H685)

   In order to retrieve all specific and non-specific disclosures of halogens (in the type of halogen compounds searched with Set H6: codes), add the code H600 to the above list of 'OR'ed terms.

H60: Halogen(s) present

H600	Non-specific disclosure of halogens	
H601	Fluorine (other than -CF3, which is coded H685)	
H602	Chlorine	
H603	Bromine	
H604	Iodine	(1981)
H607	Per-halo compound	
H608	Bis-halo (only assigned to compounds coded H600 through H604)	
H609	Poly (only assigned to compounds coded H600 through H604)	

Notes on H60: Codes

1. H600 is coded when a patent document uses the term "halogen" without specifying every individual halogen intended. Halogens mentioned individually in such patents are also coded. In an exhaustive search for a specific halogen, the code for the specific halogen should be 'OR'ed with H600. However, the code H600 should never be searched without one or more of the codes H601, H602, H603, H604, or H685.
2. The code H607 is assigned to compounds that: (a) receive one or more of the codes H600 through H604; and (b) do not have any C-H bonds other than the group -C(=W)-H. (W = any heteroatom.)

3. If one of the codes H601 through H604 is applicable to a compound twice, the code H608 is also assigned. If one of the codes H601 through H604 is applicable to a compound more than twice, the code H609 is also assigned.

H62: Halogen bonded to heterocyclic C		
H621	One	(1981)
H622	Two	(1981)
H623	Three or more	(1981)

H64: Halogen bonded to aromatic C		
H641	One	(1981)
H642	Two	(1981)
H643	Three or more	(1981)

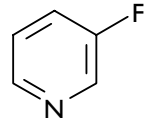
H66: Halogen bonded to alicyclic C		
H661	One	(1981)
H662	Two	(1981)
H663	Three or more	(1981)

H68: Halogen bonded to aliphatic C		
H681	-CH2-X	(1981)
H682	>CH-X	(1981)
H683		(1981)
H684	>CX2	(1981)
H685	-CF3 (including X-CF3)	(1981)
H686	-CX3, CX4 (excluding -CF3 and CF4)	(1981)
H689	Poly (for compounds coded H681 through H686)	(1981)

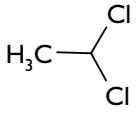
Notes on H68: Codes

1. X is F, Cl, Br, or I, and may vary independently in the code definitions of H684 and H686.
2. The codes H682, H683, and H684 are not assigned if the C atom to which the halogen is bonded is multiply bonded to O, S, Se, Te, or N, in which case the codes in Set L3: or Set L5: are applicable.
3. The **Poly** code, H689, indicates that at least one code from Subset H68: is applicable more than once to a given structure. **Poly** does not mean that more than one code from Subset H68: is applicable.

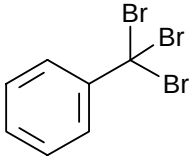
Examples Using H6: Codes

a 

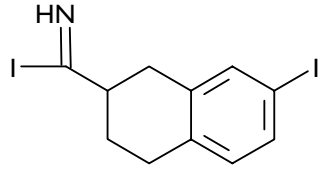
H601, H621

b 

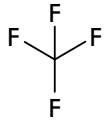
H602, H608, H684

c 

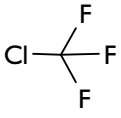
H603, H609, H686

d 

H604, H608, H641, L353 (but not H683)

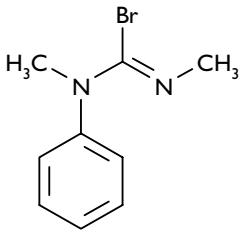
e 

H685, H689, H607 (but not H601 or H609)

f 

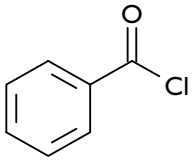
H602, H685, H686, H607 (but not H601, H609)

g



H603, L330

h



L512, C017 (but not H602)

H7: OLEFINIC/ACETYLINIC GROUPS

Notes on H7: Codes

1. A functional group may be searched using codes from both Subset H71: and Subset H72: when applicable.
2. Unlike most functional group codes, the codes H711, H712, and H724 can be assigned to functional groups with more than one atom in the same ring.

H71: Identification of C=C groups present

H711	>C=C=C<	(1981)
H712	>C=C=W	(M3 - 1972; rest - 1981)
H713	CH_2=CH-W	(M3 - 1972; rest - 1981)
H714	CH_2=CH-C=W or CH_2=CH-C≡W	(M3 - 1972; rest - 1981)
H715	CH_2=CH-(ring C)	(M3 - 1972; rest - 1981)
H716	CH_2=CH-CH_2-W	(M3 - 1972; rest - 1981)

where W is any heteroatom.

Note: In previous manuals it was wrongly reported that H713: H716 were valid from 1970 in M3.

H72: Non-cyclic >C=C< group present

H720	Exocyclic C=C
H721	One >C=C< group present, not exocyclic
H722	Two unconjugated >C=C< groups present
H723	Three or more unconjugated >C=C< groups present
H724	Two conjugated >C=C< groups present
H725	Three or more conjugated >C=C< groups present

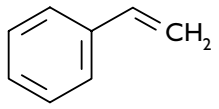
Notes on H72: Codes

1. The code H720 is assigned to >C=C< groups in which one of the C atoms is a ring member, as well as to >C=C< groups in which the two C atoms are members of different rings.
2. Conjugation of double bonds does not include double bonds in rings. However, conjugation does continue through one single bond in a ring. Thus, two exocyclic >C=C< groups on adjacent ring C atoms are coded as two conjugated double bonds, and H720 and H724 are assigned.
3. If conjugated and unconjugated double bonds are present in the same structure, they are coded separately. For example, if a carbon chain has three double bonds, two of which are conjugated, it is coded H721 and H724.
4. Exocyclic >C=C< groups are not counted when applying H721, H722, or H723.

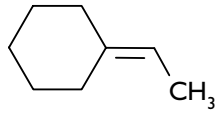
H73: Acetylinic bonds present

H730	One or more -C≡C- groups present in a ring
H731	One non-cyclic -C≡C- group present
H732	Two or more non-cyclic -C≡C- groups present

Examples Using H7: Codes

a 

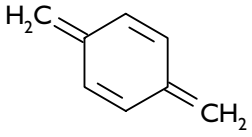
H715, H721

b 

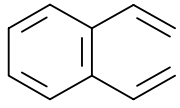
H720

c CH2=C=CH2

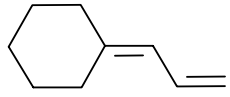
H711, H722

d 

H720, L952

e 

H715, H722 (but not H724)

f 

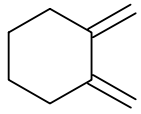
H720, H724

g CH_2=CH-CH=CH_2

H724

h CH_2=CH-CH_2CH_2CH=CH_2

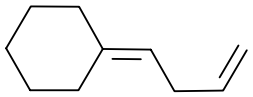
H722

i 

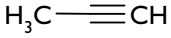
H720, H724

j CH_2=CH-CH=CH-CH_2-CH=CH_2

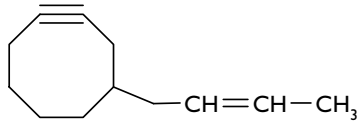
H721, H724 (but not H722)

k 

H720, H721 (but not H722)

l 

H731

m 

H721, H730

n O=C=C=C=O

H711, H712, H722, J582

Part J: Common Functional Groups With >C=O or >C=S

Main Headings

J0: Number of Carboxylic Acid, Ester, and Amide Groups Present
J1: Carboxylic Acid, Thiocarboxylic Acid
J2: Ester (Thioester)
J3: Amide (Thioamide)
J4: Aldehyde (Thioaldehyde)
J5: Oxo, Thioxo

Introduction

Part J: codes refer to common functional groups with a carbonyl or thionyl group. Some of the guidelines for using Part J: codes were given earlier in this chapter, both in the context of all functional group codes (Parts H: through L:, starting on page 102) and in the context of common functional group codes (Parts H: and J:, starting on page 112). Those general guidelines are not repeated in this section.

Notes on Part J: Codes

1. Code definitions and notes that apply equally to O-containing functional groups and their S-containing analogues include the name of the sulphur analogue in parentheses. For example, "**Set J2: Ester (thioester)**".

2. The **symbols** used in Part J: code definitions, which can vary independently when appearing more than once in a fragment, are defined as follows:

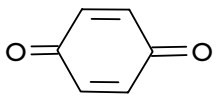
   T = O, S, Se, or Te, with a valency of II
   T' = S, Se, or Te, with a valency of IV or more
   U = O, S, Se, Te, or N
   R = C or H
   X = Halogen (F, Cl, Br, I)
   Y = O, S, Se, Te, N, or halogen

   Most of these symbols are standard throughout this user guide, but the element symbols **T'** and **R** are defined specifically for this section.

3. A number of carbonyl- and thionyl-containing functional groups are coded in Parts K: and L:; some of these functional groups are additionally coded in Part J:, and some are not. Conversely, some carbonyl- and thionyl-containing functional groups are coded in Part J: that are not coded in Parts K: or L:. To be coded in Part J:, the carbonyl (or thionyl) side of the functional group must be directly bonded to H or C.

Exception:
oxo and thioxo groups bonded to ring C atoms are coded in Set J5: even if the resulting carbonyl or thionyl group is further bonded solely to one or two of the following atoms: O, S, Se, Te, N, or halogen. (Choosing the correct code Part(s) in which to code a functional group is discussed in the section "Guidelines For Choosing Applicable Functional Group Codes" on page 104.)

The following groups are NOT coded in Part J:

Group	Code Used	Example
Non-ring U-C(=T)-U	From Set L4:	N-C(=O)-N
-C(=T)-U-C(=T)-	From Set L5:	-C(=O)-N-C(=O)-
-C(=T)-T'-	L550	-C(=S)-S^{IV}(=O)-
-R-C(=T')-R	L570	-CH2-C(=S^{IV}=O)-H
-R-C(=T)-S-Y	L520	-C-C(=O)-S-O-
-R-C(=T)-O-S	L520	H-C(=O)-O-S
-R-C(=T)-O-X	L520	H-C(=O)-O-Cl
Quinones	L951	

The following groups ARE coded in Part J:

Group	Coded in Part J: as:	Example
R-C(=T')-U-	Same as S^{II} analogue	-C-C(=S^{IV}=O)-S-
R-C(=T)-NO2	Amide (thioamide)	CH3-C(=S)-NO2
R-C(=T)-O-O-R	Same as R-C(=T)-O-R in Set J1: or Set J2:; also coded K910	-C-C(=O)-O-O-H
R-C(=T)-T-C=N, R-C(=T)-T-C≡N	Ester (both also indexed L351; -T-C≡N is also indexed as either L120 or L130)	CH3-C(=O)-O-C≡N

J0: NUMBER OF CARBOXYLIC ACID, ESTER, AND AMIDE GROUPS PRESENT

J011	One carboxylic acid, ester, or amide group present (1981)
J012	Two carboxylic acid and/or ester and/or amide groups present (M3 - 1972 (rest - 1981)
J013	Three carboxylic acid and/or ester and/or amide groups present (M3 - 1972 (rest - 1981)
J014	Four or more carboxylic acid and/or ester and/or amide groups present (M3 - 1972 (rest - 1981)

Notes on J0: Codes

1. Thiocarboxylic acid, thioester, and thioamide groups are not counted when applying Set J0: codes.
2. Prior to 1981, the codes J012, J013, and J014 were additionally assigned to the groups C(=O)-X, -C(=O)-N=U, C(=O)-U-Y, C(=O)-N°N, C(=O)-N-C(=O), and C(=O)-O-C(=O).

J1: CARBOXYLIC ACID, THIOCARBOXYLIC ACID

Notes on J1: Codes

1. To be coded in Set J1:, a carboxylic acid group or thiocarboxylic acid group must be bonded to H or C.
2. Salts of carboxylic acids and thiocarboxylic acids are assigned the same code from Set J1: as the parent acid.
3. Peroxyacids, i.e. compounds of formula R-C(=O)-O-OH, are assigned the same code as the corresponding carboxylic acid, as well as the code K910.

J11: Carboxylic acid group bonded to heterocyclic C

J111	One
J112	Two
J113	Three or more

J13: Carboxylic acid group bonded to aromatic C

J131	One
J132	Two
J133	Three or more

J15: Carboxylic acid group bonded to alicyclic C

J151	One
J152	Two
J153	Three or more

J17: Carboxylic acid group bonded to H or aliphatic C

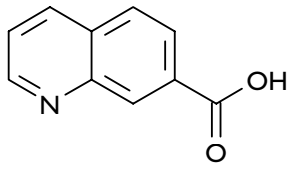
J171	One
J172	Two
J173	Three or more

J19: Thiocarboxylic acid group

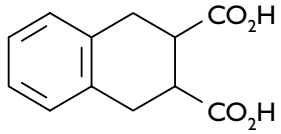
J191	One or more thiocarboxylic acid groups bonded to heterocyclic C
J193	One or more thiocarboxylic acid groups bonded to aromatic C
J195	One or more thiocarboxylic acid groups bonded to alicyclic C
J197	One or more thiocarboxylic acid groups bonded to H and/or aliphatic C

Note Thiocarboxylic acid groups are not coded in Sets J11: through J17:.

Examples Using J1: Codes

a 

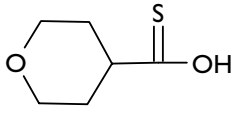
J011, J131

b 

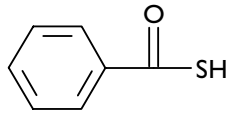
J012, J152

c HCOOH

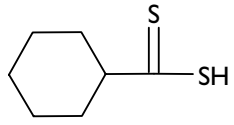
J011, J171

d 

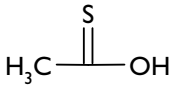
J191

e 

J193

f 

J195

g 

J197

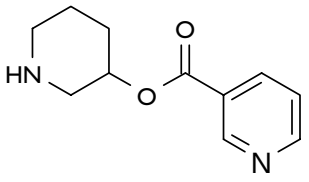
J2: ESTER (THIOESTER)

Notes on J2: Codes

1. **Thioesters** are ester groups with either one or both carboxylic O atoms substituted by S. Thioesters receive the same code as their oxygen analogues, plus the code J290 ("Thioester present").
2. Esters (thioesters) are coded according to the guidelines for **non-symmetrical functional groups**, which are listed in Note 2 on page 113.
3. To be coded in Set J2:, the carbonyl (thionyl) group of the ester (thioester) must be **directly bonded to either H or C**.
4. Esters of formula R-C(=O)-O-Het N are assigned a code based on the group R, not based on the heterocyclic N. However, prior to 1981 a few references of this type were incorrectly coded in Subset J22:. (See Note 8 for thioesters)
5. Peroxyesters, i.e. compounds of formula R-C(=O)-O-O-, are assigned the same code as the corresponding ester of formula R-C(=O)-O- (i.e. with one -O- removed), and are additionally assigned the code K910 ("C(=U)-O-O"). (See note 8 for thio derivatives).
6. Codes in Set J2: are applicable to esters of hydroxylamines, but not to thioesters of hydroxylamines. (See Note 8).
7. Codes in Set J2: are applicable to groups of the type R-C(=T)-T-C=N- and R-C(=T)-T-C≡N, as long as the -T- atom does not have a valency of IV or VI. (T = O, S, Se, or Te)
8. The groups R-C(=T)-S-Y (T = O, S, Se, or Te; Y = O, S, Se, Te, N, or halogen) and R-C(=T)-O-Y' (Y' = S, Se, Te or halogen) are coded L520, not in Set J2:.

J21: Ester (thioester) group bonded to heterocyclic C via >C=O (>C=S)

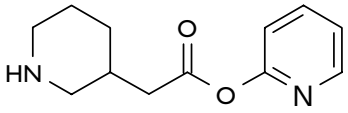
J211 One
J212 Two or more



J211

J22: Ester (thioester) group bonded to heterocyclic C via O (S)

J221 One
J222 Two or more



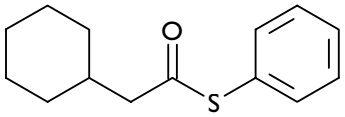
J221

J23: Ester (thioester) group bonded to aromatic C via >C=O (>C=S)

J231 One
J232 Two or more

J24: Ester (thioester) group bonded to aromatic C via O (S)

J241 One
J242 Two or more



J241, J290

J25: Ester (thioester) group bonded to alicyclic C via >C=O (>C=S)

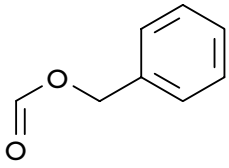
J251 One
J252 Two or more

J26: Ester (thioester) group bonded to alicyclic C via O (S)

J261 One
J262 Two or more

J27: Ester (thioester) group bonded to H or aliphatic C via >C=O (>C=S)

J271 One
J272 Two
J273 Three or more



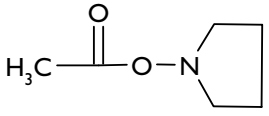
J271

J290 Thioester group present

Note Thioesters are also assigned an applicable code from Subsets J21: through J27:.

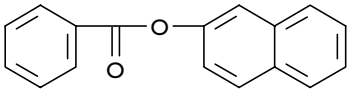
Examples Using J2: Codes

a



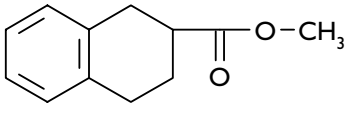
J011, J271, H211, K820

b

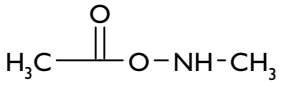


J011, J231

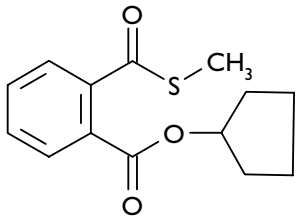
c



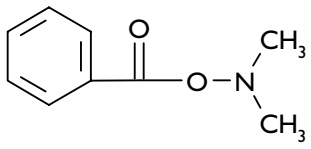
J011, J251

d 

J011, J271, K820

e 

J011, J232, J290

f 

J011, J231, K820

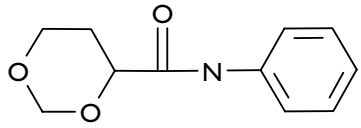
J3: AMIDE (THIOAMIDE)

Notes on J3: Codes

1. **Thioamides** receive the same code as their oxygen analogues, plus the code J390 ("Thioamide present").
2. Amides (thioamides) are coded according to the guidelines for **non-symmetrical functional groups**, listed in Note 2 on page 113.
3. Codes in Set J3: are assigned to all groups of the type R-C(=T)-N (where R = H or C bonded to any other atoms, and T = O, S, Se, or Te), except for R-C(=T)-N-C(=T) groups which are coded in Set L53:.
4. Amides of formula R-C(=O)-N-(Heterocyclic N) are assigned a code based on the R-C(=O) group, not based on the heterocyclic N. However, prior to 1981 a few references of this type were incorrectly coded in Subset J32:.

J31: Amide (thioamide) group bonded to heterocyclic C via >C=O (>C=S)

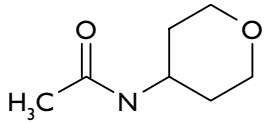
J311 One
J312 Two or more



J311

J32: Amide (thioamide) group bonded to heterocyclic C via N

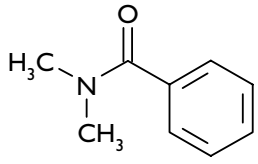
J321 One
J322 Two or more



J321

J33: Amide (thioamide) group bonded to aromatic C via >C=O (>C=S)

J331 One
J332 Two or more



J331

J34: Amide (thioamide) group bonded to aromatic C via N

J341 One
J342 Two or more

J35: Amide (thioamide) group bonded to alicyclic C via >C=O (>C=S)

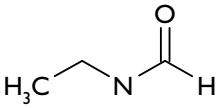
J351 One
J352 Two or more

J36: Amide (thioamide) group bonded to alicyclic C via N

J361	One
J362	Two or more

J37: Amide (thioamide) group bonded to H or aliphatic C via >C=O (>C=S)

J371	One
J372	Two
J373	Three or more



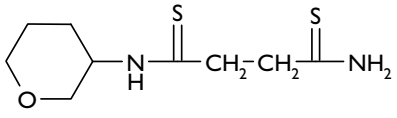
J371

J390 Thioamide group present

Note Thioamides are also assigned the applicable code from Subsets J31: through J37:.

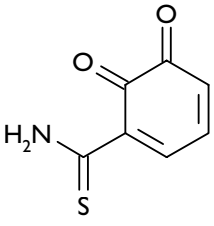
Examples Using J3: Codes

a



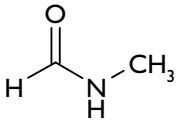
J321, J371, J390

b



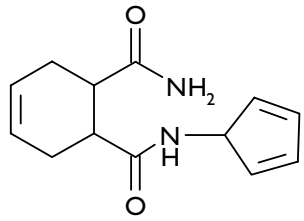
J331, J390, L951

c



J011, J371

d



J012, J352

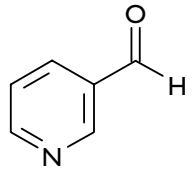
J4: ALDEHYDE (THIOALDEHYDE)

Notes on J4: Codes

1. To be coded in Set J4:, the -C(=O)-H group of aldehydes or the -C(=S)-H group of thioaldehydes can only be bonded to H or C. Thioaldehydes receive the same code as their oxygen analogues, plus the code J490 ("Thioaldehyde present").
2. Thioaldehydes in which the thionyl S has a valency of IV or more are coded L570, not in Set J4:.
3. Hydrates are assigned the same codes as the corresponding aldehydes or thioaldehydes; they are not coded L650 ("S-C-S or S-C-O") or L660 ("O-C-O").

J41: Aldehyde (thioaldehyde) group bonded to heterocyclic C

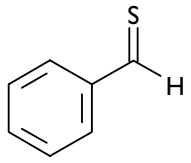
J411	One
J412	Two or more



J411

J43: Aldehyde (thioaldehyde) group bonded to aromatic C

J431 One
J432 Two or more



J431, J490

J45: Aldehyde (thioaldehyde) group bonded to alicyclic C

J451 One
J452 Two or more

J47: Aldehyde (thioaldehyde) group bonded to H or aliphatic C

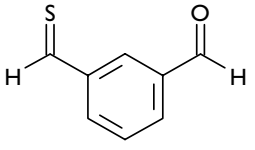
J471 One
J472 Two or more

J490 Thioaldehyde group present

Note Thioaldehyde groups are also coded in Sets J41: through J47:.

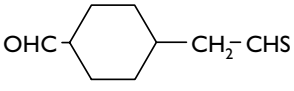
Examples Using J4: Codes

a



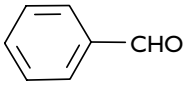
J432, J490

b



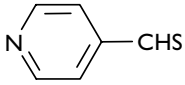
J451, J471, J490

c



J431

d

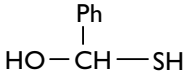


J411, J490

e

J471 (but not L660 or H482)

f



J431, J490 (but not H481, H498, or L650)

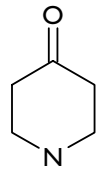
J5: OXO (=O), THIOXO (=S)

Notes on J5: Codes

1. Thioxo groups are coded in Subset J59:, not as the oxo analogue in Subsets J51: through J58:.
2. The aliphatic oxo and thioxo codes (i.e. J581, J582, J583, and J598) are not used for carbonyl or thionyl groups bonded to O, S, Se, Te, N, or halogen. However, oxo and thioxo groups bonded to ring C atoms are coded in Set J5: even if the resulting carbonyl or thionyl group is further bonded only to O, S, Se, Te, N, or halogen instead of C.
3. Keto-enol (thioxo-thiol) tautomerism can affect the coding of oxo (thioxo) groups. When keto-enol (thioxo-thiol) tautomerism is possible, the structure is coded in the keto (thioxo) form, unless the -OH (-SH) group of the enol (thiol) form is bonded to a fully conjugated carbocyclic ring, e.g. benzene.
4. Thioketones in which S has a valency greater than II are coded L570, not in Set J5:.
5. The oxo groups of quinones are coded in Subset L95:, not in Set J5:.

J52: Oxo group bonded to heterocyclic C

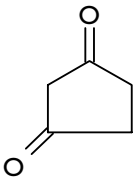
J521	One
J522	Two
J523	Three or more



J521

J56: Oxo group bonded to alicyclic C

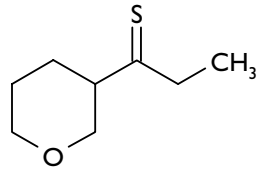
J561	One
J562	Two
J563	Three or more



J562

J58: Oxo group bonded to aliphatic C

J581	One
J582	Two
J583	Three or more



J581

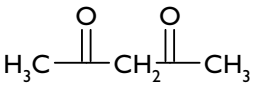
J59: Thioxo group

J592	One or more thioxo groups bonded to heterocyclic C
J596	One or more thioxo groups bonded to alicyclic C
J598	One or more thioxo groups bonded to aliphatic C
J599	Poly, for Subset J59: (1981)

Note Thioxo groups are not coded in Subsets J52: through J58:.

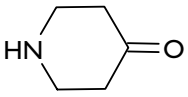
Examples Using J5: Codes

a



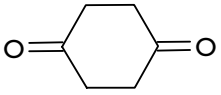
J582

b



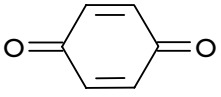
J521

c



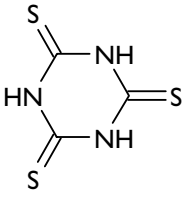
J562

d



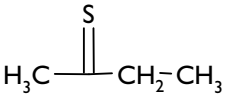
L951 (but not J562)

e



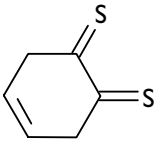
J592, J599, L910, L999

f



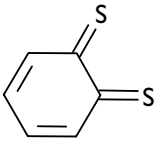
J598

g



H494, L960 (but not J596, J599)

h



L951 (but not J596, J599)

Notes and Guidelines Pertaining to Less-Common Functional Groups

Introduction

This section lists coding guidelines that Parts K: and L: have in common. Functional groups coded in Parts K: and L: are predefined arrangements of atoms and bonds, that, for the most part, do not have common names like "alcohol", "amine", "nitro", etc. Part K: codes represent bonds between heteroatoms, and Part L: codes represent a wide variety of element and bond combinations.

Notes and guidelines applicable to all functional groups were discussed earlier in this chapter (pages 98-99), and those applicable only to the individual Parts K: and L: are discussed later in their respective sections of this chapter.

Notes on Codes in Parts K: and L:

1. Codes in Parts K: and L: differ from codes in Parts H: and J: in a number of ways:

   - Many of the K: and L: codes are **green** on the coding sheet, which means they can only be searched from 1981/Derwent Week 27 to the present;
   - Some groups of codes have a corresponding **precursor code** (explained in Note 6 below) that is searched prior to 1981 in place of the current, more specific codes;
   - **Poly** codes are used to indicate that a particular group occurs more than once, rather than using codes defined for a precise number of occurences of a group; and
   - Codes in Parts K: and L: are not selected according to **the type of atom** to which a functional group is bonded except for cyano and carbamate groups.

2. Many codes in Parts K: and L: use one-character **symbols** as a chemical shorthand for groups of atoms that are coded similarly. These symbols are defined as follows:

   T = O, S, Se, Te
   U = O, S, Se, Te or N
   V = Heteroatom other than O, S, Se, Te or N
   W = O, S, Se, Te, N, C or halogen, in Parts: H: through L:
   = Any heteroatom, in Parts other than H: through L:
   X = Halogen (F, Cl, Br, I)
   Y = O, S, Se, Te, N, or halogen
   Z = B, Si, P or As

3. More than one code is usually required to fully describe **large structural fragments containing heteroatoms**. As stated earlier in the section entitled "Choosing Applicable Functional Group Codes In Parts H:, J:, K:, and L:", the minimum number of functional group codes are assigned that between them specify every atom of the group and all linkages between the atoms. The largest possible fragment is searched at each successive stage of the process, even if that means that some atoms and bonds are covered by more than one code.

4. If the definition of a code includes one or more variable element symbols (i.e. T, U, V, W, X, Y or Z), **additional functional group codes** are used to identify the elements represented by the symbols, unless the definition of the first code specifically states that this should not be done. For example, Note 5 for the code L620 on page 162 states that no additional code should be searched to identify the T atom in the code definition ("S-C(-T)-U"). In most cases, however, one or more codes from Part K: or Part L: are used to fully specify all elements present in the functional group, even if that means that some atoms are covered by more than one code. For example, -S-S-N< is searched as -S-S-Y (K222) and also as -Y-S-N< (K352). Used together, these two codes leave no ambiguity as to the identity of the atoms being searched. The group >N-N-C≡N is searched as N-N-C≡U (K620) and N-C≡N (L110).

5. The group -C≡U is used several times in Part K: code definitions; for example, K910 is defined as -O-O-C≡U. Although the -C≡U group can only be -C≡N, the symbol U is included in the definition so that the user realizes that the -C≡U group requires further identification, e.g. with a code from Set L1: ("-C≡N, C<four bond>N-").

6. A code with a **#** sign next to it on the coding sheet indicates that several current codes have replaced a discontinued code that was less specific. The less specific code, known as a **"discontinued generic code"** or **"precursor code"**, is searchable in the time ranges preceding the introduction of the current code. Although the coding examples in this chapter only list the current codes, it is assumed that any available precursor codes would be included in the search strategy. (Precursor codes and coding time ranges are discussed in the Introduction to this manual.)

7. Codes in Parts K: and L: that do not specify a valency for S, N, or halogen can be used for any valency of those atoms. For example, the code K351 illustrated above can be used for any valency of S and N, because no valencies are specified in the code definition. A valency code from Set C1:, C2:, or C3: is assigned in addition to the functional group code(s) from Part K: or Part L: in the following cases:

   - halogen with a valency of II or more
   - N with a valency of V or more
   - all valencies of S
   - a Part K: or Part L: code that is defined for more than one specific valency

The Part C: valency codes, searchable from 1970 forward, are defined as follows:

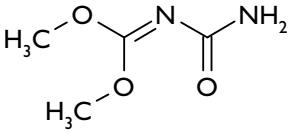
Lower Valency	Intermediate Valency	Higher Valency
C100 X^{I}, X^{II+}	**C200** X^{III}, X^{IV+}	**C300** X^{V}, X^{VI+}, X^{VII}
C107 N^{II-}, N^{III}, N^{IV+}	**C216** S^{IV}	**C307** N^{V}
C116 S^{II}, S^{+}		**C316** S^{VI}

8. As indicated in Note 2 of this section, the symbols **X** and **Y** may be used to represent **halogens** in Part K: and Part L: codes. The halogens that these symbols represent in a particular structure are identified using halogen codes from Subset H60: or Set C0:. **Halogen codes from Subset H60:** have been used to identify X atoms in groups that are coded L330 or L353 from **1963** forward; thus the H60: codes are included in the same search statement as the black codes from the coding sheet. **Halogen codes from Set C0:** codes have been used to identify all other **X** and **Y** atoms coded in Parts K: and L: since **1978**; thus C0: codes used to identify halogens in organic compounds are included in the same search statement as the green codes from the coding sheet. For example, the group S-N-Cl is coded K352 ("S-N-Y") and C017 ("Cl"). The coding of halogens is discussed more completely in the section "Coding Rules for Polymers and Halogens", beginning on page 18.

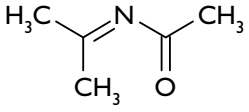
9. **The presence of tautomers** can significantly affect the coding of functional groups in Parts K: and L:. Tautomers are discussed in the section entitled "Tautomerism", on page 106.

10. Heteroatoms bonded to metals are discussed in Part A: in the section entitled "Guidelines for Coding Compounds Containing Metals", which begins on page 40. Heteroatoms bonded to B, Si, P and As are discussed in Part B: in the section entitled "Guidelines for Coding Functional Groups Bonded to B, Si, P, or As", which begins on page 48.

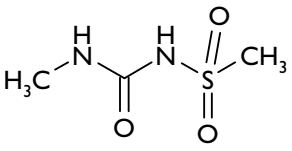
Examples Illustrating Coding Rules for Parts K: and L:

a 

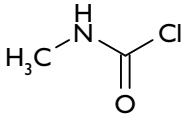
L290, L432 (not L352 or L660)

b 

L355 , J371

c 

K351, L431, C316

d 

L463, L511, C017

Part K: Bonds Between Heteroatoms in Organic Compounds

Main Headings

K1: X=Y and X-Y bonds (X = F, Cl, Br, I; Y = O, S, Se, Te, N, F, Cl, Br, or I)
K2: S=S and S-S bonds
K3: S=N and S-N bonds
K4: S=O and S-O bonds
K5: N≡N and N=N bonds
K6: N-N bonds
K7: N=O bonds
K8: N-O bonds
K9: O-O bonds

Introduction

Part K: codes refer to bonds between any two of the elements O, S, Se, Te, N, and halogen. Some of the guidelines for using Part K: codes were given earlier in this chapter, both in the context of all functional group codes (Parts H: through L:, starting on page 102) and in the context of less-common functional group codes (Parts K: and L:, starting on page 140). Those general guidelines are not repeated in this section.

Notes on Part K: Codes

1. Heteroatoms bonded to metals are discussed in the section entitled "Guidelines for Coding Compounds Containing Metals", which begins on page 40.

2. Heteroatoms bonded to B, Si, P or As ("Z" atoms) are discussed in the section entitled "Guidelines for Coding Functional Groups Bonded to B, Si, P, or As", which begins on page 48.

3. The symbols used in Part K: code definitions, which can vary independently when appearing more than once in a fragment, are defined as follows:

   T = O, S, Se, or Te
   U = O, S, Se, Te, or N
   X = Halogen (F, Cl, Br, I)
   Y = O, S, Se, Te, N, or halogen
   W = Any heteroatom

Subheadings Applicable to Part K: Codes

Unless otherwise specified, codes are valid in Subs M0-M4 from 1963

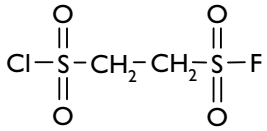
K1: X-Y BONDS (X = F, CL, BR, I; Y = O, S, SE, TE, N, OR HALOGEN)	
K110	X-X, X=X
K121	X-S^{VI}, X=S^{VI}
K122	X-S^{II}, X=S^{II}, X-S^{IV}, X=S^{IV}
K130	X-N<, X=N-
K140	X-O-, X=O
K141	Poly, for Set K1:

Notes on K1: Codes

1. From 1978 forward, the X atom of groups coded in Set K1: has been identified by a halogen code from Set C0:. C0: codes used in this way should be included in the search statement with the codes that are green on the coding sheet, i.e. those introduced in 1981.
2. The **valency of halogen** in groups coded K140 is always coded C100 ("Halogen with valency of I or II), C200 ("Halogen with valency of III or IV"), or C300 ("Halogen with valency of V or higher"). The valency of halogen in groups assigned other codes in Set K1: is only coded if greater than I. (Valency codes, which are searchable from 1970 forward, are explained in Note 7 on page 141).
3. The valency of the S atom in K122 is further specified by C116 (S^{II}) or C216 (S^{IV}).
4. Unlike most functional groups coded in Parts H: through L:, both atoms represented by K1: codes may be in the same ring.

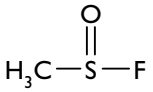
Examples Using K1: Codes

a



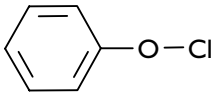
K121, K199, C009, C017

b



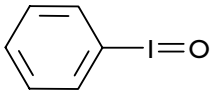
K122, C009, C216

c



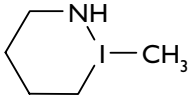
K140, C017, C100

d



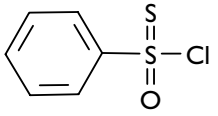
K140, C053, C200

e



K130, C053, C200, C600

f



K121, K210, C017

K2: S=S AND S-S BONDS

K210	S=S	(1970)
K220	S-S (Pre-1981 precursor code for K221 through K224)	(1965-1981)
K221	S-S-$(S)_n$ (n≥1)	(1981)
K222	S-S$(=U)_n$-; S-S-Y; or S-S$(=U)_n$-Y (n = 1 or 2; excludes groups coded K221)	(1981)
K223	S-S-C(=U)- or S-S-C(≡U)	(1981)
K224	Other S-S- groups	(1981)
K299	Poly, for Set K2:	

Notes on K2: Codes

1. Prior to 1970, the code K220 was not assigned to -S-S- groups containing S^{IV} or S^{VI}.

2. The valency of S in Set K2: codes is coded in Part C: for S^{IV} (C216) and S^{VI} (C316), unless either K422 ("O-S^{VI}-O") or K432 ("Other S^{VI}-O") is also assigned to the structure. These latter codes specify the valency of S as VI, and therefore require no valency code.

3. Prior to 1981, groups containing S^{IV} or S^{VI} bonded solely to S atoms, or bonded solely to S and C atoms, were additionally coded in Set K4: as if one S-S or S=S bond were replaced by S-O or S=O, respectively. Since 1981, such groups have not been coded in Set K4:.

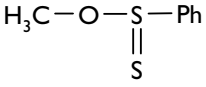
4. The fragments -S-S-S$(=U)_n$- and -S-S$(=U)_n$-S- (n = 1 or 2) are coded K221, not K222 and K299. This is in keeping with the rule that the first applicable K: code is chosen over subsequent (and usually less precise) codes. The S=U group is additionally assigned an applicable code from Part K:.

5. Groups coded K223 are additionally coded in Part L: from 1981. Prior to 1981/week 27, such groups were also coded in Set J2: as thioesters, but this practice has been discontinued.

Examples Using K2: Codes

a CH_3-$(S)_4$-CH_3

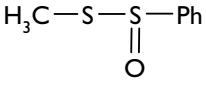
K221 (but not K299)

b 

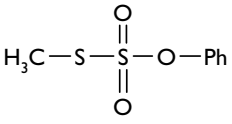
K210, K433, C216

c CH_3-$(S)_4$-CH_2-$(S)_4$-CH_3

K221, K299, L650

d 

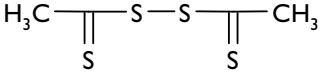
K222, K433, C216

e 

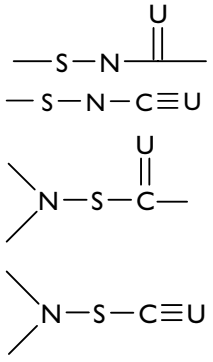
K222, K410, K422

f CH_3-S-S-CN

K223, L120

g 

K223, K299, L520, L599
(Prior to 1981/Week 27, search J272, J290 instead of L520, L599)

K3: S=N and S-N bonds		
K310	-C≡N=S	
K320	$S{=}N^{V}$ (excluding groups coded K310)	(1981)
K330	$S{=}N^{III}$	
K340	-O-S-N<	
K350	Other -S-N< groups (Pre-1981 precursor code for K35: codes)	(1963-1981)
K351		(1981)
K352	-S-N-Y, Y-S-N (excluding groups coded K340), $-S-N(=U)_n$ (n = 1 or 2)	(1981)
K353	Other -S-N groups	(1981)
K399	Poly, for Set K3:	

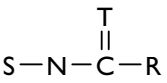
Notes on K3: Codes

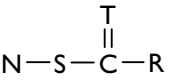
1. Groups coded in Set K3: are additionally coded C116 (S^{II}), C216 (S^{IV}), or C316 (S^{VI}), depending on the valency of S in the structure being coded. **Exception:** -C≡N=S^{II} is coded K310 with no additional valency code.

2. Prior to 1981, amine sulphides were coded K742; from 1981 forward, they have been coded K320.

3. Before 1981, groups with S=N were assigned the code K330, as well as any other K3: codes that would be applicable to the fragment if S=N were changed to S-N. An **exception** was that the code K350 was only used with K330 when the S atom was multiply bonded to one or more O atoms. Thus -N=S=O was coded K330, C216, and K350, but -N=S was coded K330 and C116. Since 1981, these rules have not been assigned, and groups are only coded according to the definitions.

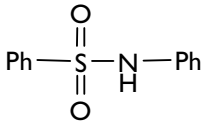
4. Prior to 1981, thionitrates were coded K710 and thionitrites were coded K720. These groups were additionally assigned the code C116 (S^{II} or S^{III}) from 1970 forward. Group (a) shown below is additionally coded in Set J3:. Group (b) shown below is additionally coded L520, but not in Set J2:.

a 

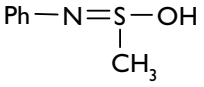
b 

R is not O, S, Se, Te, N, or halogen

Examples Using K3: Codes

a 

K353, C316

b 

K330, K433, C216
Prior to 1981, coded K340 instead of K433

c CH_3-S-N=O

K352, K751, C116
Prior to 1981, coded K720 instead of K751 and K352

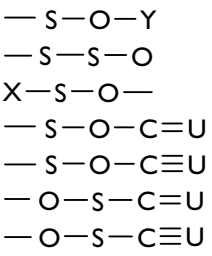
d CH_3-S-N=S

K330, K352, C116
Prior to 1981, coded K720 instead of K330 and K352

e CH_3-O-N=S

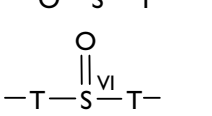
K330, K810, C116
Prior to 1981, coded K720 instead of K330 and K810

K4: S=O and S-O bonds

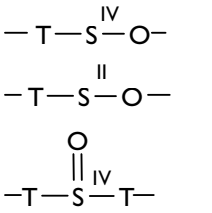
K410 

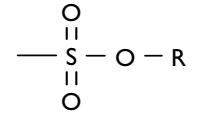
(1981)

K421 -O-SO_2-O-

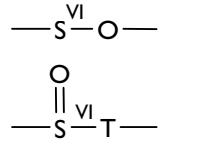
K422 

(Excluding compounds coded K421)

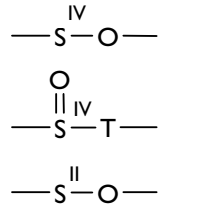
K423 

K431 

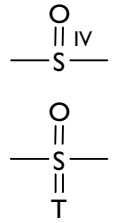
(R = H or a cation; excludes compounds coded K421, K422, or K423)

K432 

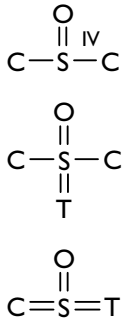
(Not used for groups coded more specifically with other K4: codes.)

K433 

(Not used for groups coded more specifically with other K4: codes.)

K441 

S is a ring member

K442 

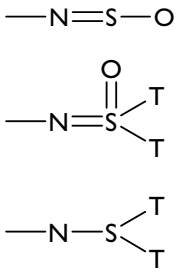
C-S-C is in a chain

K499 Poly for Set K4:

Notes on K4: Codes

1. Groups with a S-X bond are additionally coded in Set K4: if they contain the fragment X-S(=T)$_n$-T- (n = 0, 1, or 2; at at least one T atom is O).

2. Groups with a N-S or N=S bond are not coded in Set K4: unless they contain at least one S-O or S=O bond, and one of the following fragments:

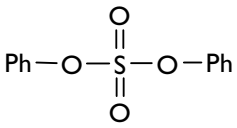


3. Groups coded K410 are additionally assigned applicable codes from K421 through K433. The code K410 is searchable for anhydrides of S acids from 1963. Mixed anhydrides may also have applicable codes in Part L:, e.g. L520.

4. Groups coded K423, K433, K441, or K442 are additionally coded C116 (S^{II}), C216 (S^{IV}), or C316 (S^{VI}), depending on the valency of S in the structure being coded.
5. Groups coded K442 that contain one of the fragments C=S^{IV} or C=S^{VI} are additionally coded L570.
6. The code K431 is not assigned if the S atom is bonded to O, S, Se, Te, N, or halogen.
7. K499 is only assigned when two or more S atoms are present, all of them are bonded to O, and they require the same K4: code to be assigned more than once.

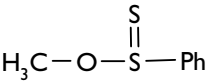
Examples Using K4: Codes

a



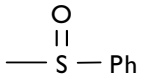
K421

b



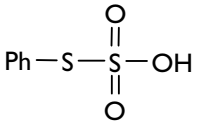
K210, K433, C216

c



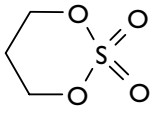
K442, C216

d



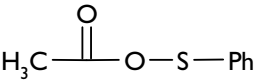
K222, K410, K422

e



K441, L970, C316

f

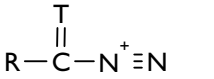


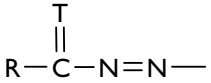
K410, K433, L520, C116

K5: N≡N and N=N bonds		
K510	-N=N^+=N-, N≡N=N-, N≡N^+-N<	
K520	-N=N-N<	
K530	-N=N-, -N^+≡N, other -N=N^+= groups (Pre-1981 precursor code for K53: codes)	(1963-1980)
K531	N≡N^+-T, -N=N(=T)-, N=N-Y (Y is not N), N≡N^+-C(=U)-, -N=N^+=T	(1981)
K532	-N=N-C(=U)-, -N=N-C≡U	(1981)
K533	-N^+≡N, N=N^+=, other =N≡N groups	(1981)
K534	Other -N=N- groups	(1981)
K599	Poly, for Set K5: codes	

Notes on K5: Codes

1. Azo Dyes and their precursors are additionally coded in Part W:.
2. From 1978 forward, N=N^+= and -N^+≡N are additionally coded in Subset L72:.
3. Groups of the types shown below in diagrams (a) and (b) are coded K531 and K532 respectively, and also in Set J3:. Prior to 1981, a few of these references were incorrectly coded L520 instead of in Set J3:.

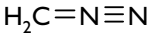




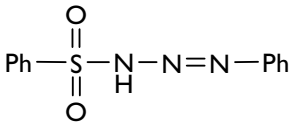
R is any atom **except** O, S, Se, Te, N, or halogen.

4. The groups C=N^+=N and C=N≡N are coded K533, but not L355 ("C=N-").

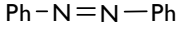
Examples Using K5: codes

a 

K533

b 

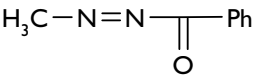
K520, K352, C316

c 

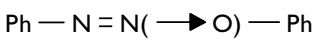
K534

d CH_3-N=N-S-CH_3

K352, K531, C116

e 

K532, J331

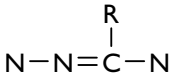
f 

K531, K741

	K6: N-N bonds	
K600	>N-N< (Pre-1981 precursor code for K610, K620, K640)	(1963-1980)
K610	>N-N^{III}=O, >N-N^{V}(=O)$_n$ (n = 1 or 2)	(1981)
K620	>N-N-Y or >N-N-C(=U) or >N-N-C≡U	(1981)
K630	>N-N=C<	
K640	Other >N-N< groups	(1981)
K699	Poly, for K6: codes	

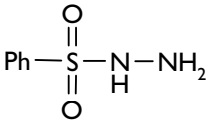
Notes on K6: Codes

1. If K510 or K520 is applicable to a N-N group, K6: codes are not also assigned.
2. Nitrosamines are searched using the codes K610 and K751. Nitramines are searched using the code K610 and the applicable code from H3: ("Nitro groups"); they are not coded in Set K7:.
3. The group N-N=S is coded K620, even though the N-S bond in the code definition is a single bond. The group N-N=N, however, is coded K520, not K620.
4. The group R-C(=T)-N-N (T is O, S, Se, or Te; R is not O, S, Se, Te, N, or halogen) is assigned the code K620 and a code from Set J3: ("Amide, Thioamide").
5. Prior to 1981, the group shown below should be searched as K600 instead of K630. The fragment N=C(-R)-N is not covered by K600, and therefore must be searched with additional applicable codes, e.g. from Sets L2: or L3:.

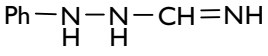


R is anything

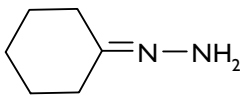
Examples Using K6: Codes

a 

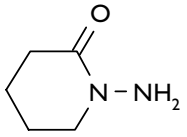
K620, K352, C316

b 

K620, L320

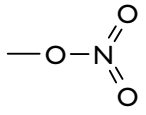
c 

K630

d 

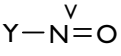
K630
K640, H211, L941 (but not K620)

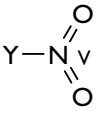
K7: N=O bonds

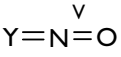
K710 

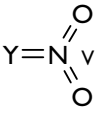
K720 -O-N^{III}=O

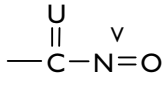
K730 -C≡N=O

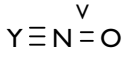
K741 











(Not used for O-NO$_2$, N-NO$_2$, or C-NO$_2$)

K742 Other structures with -N^{V}=O

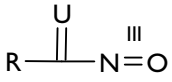
K751 Y-N^{III}=O; U=C-N^{III}=O
(Not used for structures coded K720)

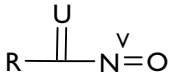
K752 -N^{III}=O; Other =N^{+}=O groups

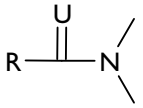
K799 Poly for Set K7:

Notes on K7: Codes

1. Prior to 1981, thionitrates were coded K710 (not K741) and thionitrites were coded K720 (not K751).
2. The group =N^{+}=O is assigned the code K752 and a code from Subset L72: (for N^{+}).
3. The group (a) shown below is coded K751 and the group (b) shown below is coded K741 (R is anything). The fragment (c) from these groups, which is not covered by the Set K7: codes, is further coded using applicable codes.

a 

b 

c 

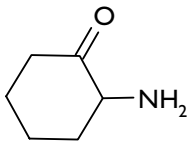
In the structures shown above, R is anything.

4. Nitro groups bonded to C are coded in Set H3:, not in Set K7:.

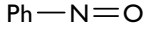
Examples Using K7: Codes

a CH$_3$-S-N=O

K751, K352, C116
Prior to 1981, K720 was assigned instead of K751

b 

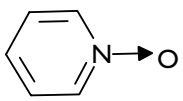
K730

c 

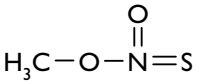
K752

d CH$_3$-N=N^{+}=O

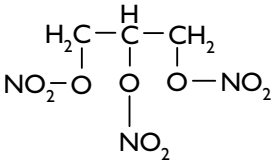
K752, K531, L722

e 

K742

f 

K741, K320, C116
Prior to 1981, K710 was assigned instead of K741

g 

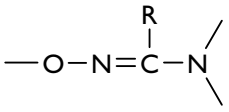
K710, K799

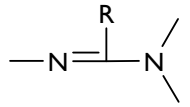
	K8: N-O bonds	
K800	>N-O- (Pre-1981 precursor code for K810, K820, K830, and K850)	(1963-1980)
K810	N-O-Y; Y-N-O-; -O-N=Y; -O-N(=Y)$_2$	(1981)
K820	>N-O-C=U; >N-O-C≡U	(1981)
K830	U=C-N-O-; U≡C-N-O-	(1981)
K840	>C=N-O-	
K850	Other >N-O- groups	(1981)
K899	Poly for Set K8:	

Notes on K8: Codes

1. If K710 or K720 is applicable to a N-O group, K8: codes are not also assigned.

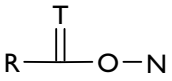
2. Prior to 1981, the group shown in (a) below was coded K800 instead of K840; the fragment from this group shown in (b) below, which is not covered by K800, was further coded using applicable codes.

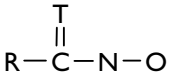
a 

b 

R in the above groups can be anything.

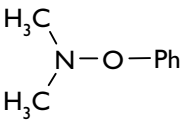
3. Groups of the type shown in diagram (a) below are assigned the code K820 and a code from Set J2: ("Ester, Thioester"). Groups of the type shown in diagram (b) below are assigned the code K830 and a code from Set J3: ("Amide, Thioamide").

a 

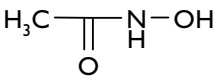
b 

R in the above groups can be anything.

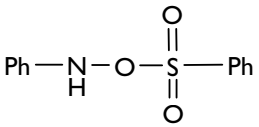
Examples Using K8: Codes

a 

K850

b 

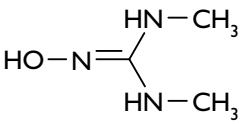
K830, J371

c 

K810, K410, K432

d $(CH_3)_2C$=N-OH

K840

e 

K840, L250

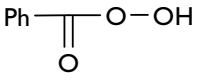
	K9: O-O bonds
K910	-O-O-Y, -O-O-C=U, -O-O-C≡U
K920	-O-OH (excluding groups coded K910)
K930	Other -O-O- groups
K999	Poly, for K9: codes

Notes on K9: Codes

1. Compounds coded K910 are additionally assigned a code for the group resulting from the removal of one O atom, unless there is a specific code that takes both O atoms into consideration. For example, perbenzoic acid is coded K910 and J131 ("One carboxylic acid group bonded to an aromatic ring"), but persulphonic acid is coded K910, K410 ("S-O-Y"), and K432 ("S^{VI}-O").

2. Prior to 1981, the code K999 was only assigned to compounds consisting of two or more -O-O- groups. Since 1981, K999 is assigned when any code from Set K9: is applicable more than once to a compound, even if there is only one -O-O- group present in the structure.

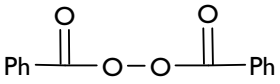
Examples Using K9: Codes

a 

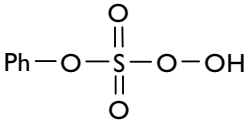
K910, J131

b CH_3-O-OH

K920

c 

K910, K999, L543

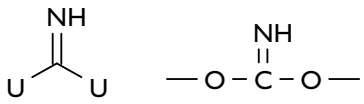
d 

K410, K421, K910 (but not K920)

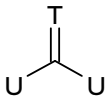
Part L: Complex Functional Groups

Main Headings

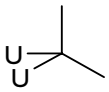
L1: 

L2: 

L3: Other >C=N- Groups

L4: 

L5: Other C=T Groups

L6: 

L7: Miscellaneous element descriptors - onium, free radicals, ions, valencies

L8: Sugars and Derivatives

L9: Groups in Rings

Introduction

Codes in Part L: cover various combinations of the elements O, S, Se, Te, N, and halogen bonded to C atoms; onium and free radical compounds; sugars and sugar residues; and functional groups that have more than one atom in the same ring. Some of the guidelines for using Part L: codes were given earlier in this chapter, both in the context of all functional group codes (Parts H: through L:, starting on page 102) and in the context of less-common functional group codes (Parts K: and L:, starting on page 140). Those general guidelines are not repeated in this section.

Notes on Part L: Codes

1. Heteroatoms bonded to metals are discussed in the section entitled "Guidelines for Coding Compounds Containing Metals", which begins on page 40.

2. Heteroatoms bonded to B, Si, P and As are discussed in the section entitled "Guidelines for Coding Functional Groups Bonded to B, Si, P, or As", beginning on page 48.

3. The element symbols used in Part L: code definitions, which can vary independently when appearing more than once in a fragment, are defined as follows:

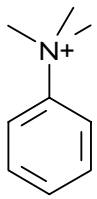
   T = O, S, Se, or Te
   U = O, S, Se, Te, or N
   X = Halogen (F, Cl, Br, I)
   Y = O, S, Se, Te, N, or halogen
   W = Any heteroatom

4. Except for L710 and codes in L8: ("Sugars") and L9: ("Groups in Ring"), Part L: codes are not applied if two or more adjacent atoms of the functional group are ring members. The code for the appropriate group with only 1 atom (or none) in the ring is applied, plus any appropriate L9: code.

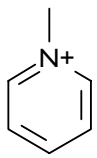
Quaternary Ammonium Compounds

The N^+ of quaternary ammonium compounds, i.e. N^+ bonded to four atoms other than H, are coded in Subset L72:. The ***parent amine*** is the amine that remains when the simplest substituent on N^+ has been removed. (The order of increasing simplicity is: heterocyclic, aromatic, alicyclic, aliphatic.) The parent amine receives all applicable codes from Parts H: through L:, e.g. an amine code from Set H1:. Thus, negating H1 eliminates many documents referring to ammonium compounds. **Exceptions:** (1) the code for tertiary amine (H103) is not assigned to ammonium compounds; and (2) if N and all substituents are members of the same ring system, the parent amine is not coded.

Examples of Coding Ammonium Compounds

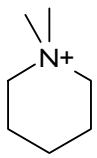
a 

L722, H141

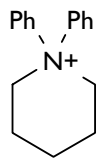
b 

L721

(no H1: code, because parent compound is not an amine)

c 

L721, H181

d 

L721, H141

Subheadings Applicable to Part L: Codes

Unless otherwise specified, codes are valid from 1963 in Subs M0-M4.

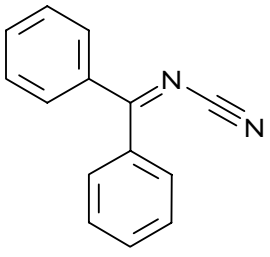
	L1: -C≡N -N≡C	
L110	>N-C≡N ; -N=C=NH	
L120	-S-C≡N	
L130	-O-C≡N	
L140	-C≡N -N≡C (Pre-1981 precursor code for Subset L14:)	(1963-1980)
L141	U=C-C≡N	(1981)
L142	-C≡N bonded to heterocyclic C	(1981)
L143	-C≡N bonded to aromatic C	(1981)
L144	-C≡N bonded to alicyclic C	(1981)
L145	-C≡N bonded to aliphatic C (not assigned to structures coded L141)	(1981)
L146	-N≡C	(1981)
L199	Poly for L1:	

Notes on L1: Codes

1. All groups containing -C≡N are assigned the first applicable code in Set L1:. If the group is not fully described by this code, additional codes from Parts K: or L: may be used, e.g. those with C≡U.

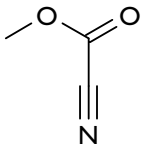
Examples Using L1: Codes

a



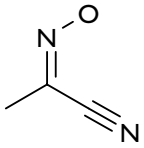
L110, L355

b



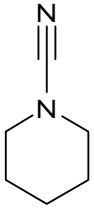
L141, J271

c



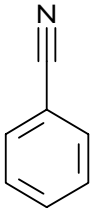
L141, K840

d



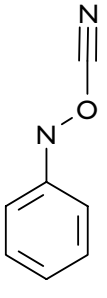
L110, H211

e

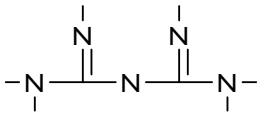
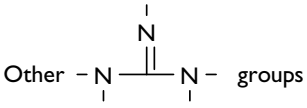
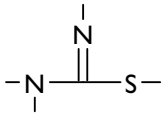
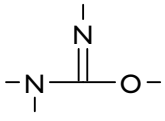
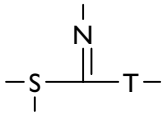
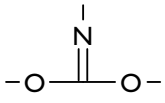


L143

f



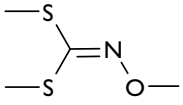
L130, K820

L2: U=C=N-, U-C(=N-)-U	
L210	-N=C=N- (excluding compounds coded L110)
L220	-N=C=S
L230	-N=C=O
L240	
L250	
L260	
L270	
L280	
L290	
L299	Poly for Set L2:

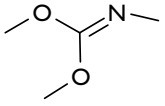
Notes on L2: Codes

1. If a group that matches a code definition in Set L2: is bonded to R-C(=T)- (R is not O, S, Se, Te, N, or halogen), the group is assigned a code from Set L2: and a code from either Set J2: ("Ester, Thioester") or Set J3: ("Amide, Thioamide") as applicable. The group is not assigned a code from L3:, as these codes are used for groups other than those that are coded in L2:.
2. If one of the terminal N-atoms of a group coded in Set L2: is double-bonded to C, the group is assigned the additional code L355 (>C=N-).
3. Groups coded L280 or L290 do not receive additional codes to identify the particular atom that "T" represents. Furthermore, groups coded L280 or L290 are not additionally assigned L650 or L660, which are less specific than L280 and L290.

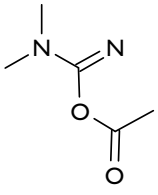
Examples Using L2: Codes

a 

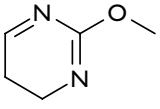
L280, K840
(but not L650, H598, or H599)

b 

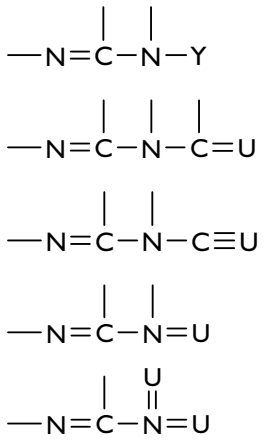
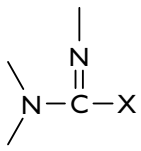
L290
(but not L352, L660, or H582)

c 

L270, J271 (but not L351)

d 

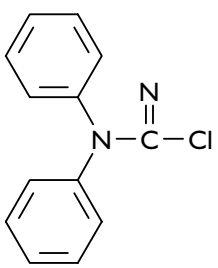
H521, L921 (but not L270)

L3: Other C=N groups		
L310	>N-C=N-C=U; >N-C=N-C≡U ; >N-C=N-Y	
L320		
L330		
	(Also search "X" in Subset H60:)	
L340	Other >N-C=N- groups	
L350	>C=N- (Pre-1981 precursor code for Subset L35:)	(1963-1980)
L351	-N=C-T-Y ; -N=C-T-C=U ; -N=C-T-C≡U	(1981)
L352	Other -N=C-T- groups	(1981)
L353	-N=C-X (Also search "X" in Subset H60:)	(1981)
L354	-N=C-C=U	(1981)
L355	Other -N=C< groups	(1981)
L399	Poly for Set L3:	(1981)

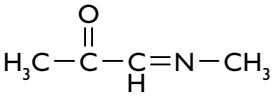
Notes on L3: Codes

1. Groups containing >C=N⁺< receive a code from Set L3: for the >C=N group, and a code from Subset L72: for N⁺.
2. The groups >C=N-N and >C=N-O are fully defined by the codes K630 and K840, respectively, and are therefore not additionally coded in Set L3:.
3. Certain limitations exist on the element(s) that may be bonded to the **leftmost C** in the Set L3: code definitions (i.e. the C in the group C=N). When the "prohibited" elements are bonded to the leftmost C atom, they either result in a tautomeric form that is not covered by L3: codes, or they change the functional group to the extent that a code outside of Set L3: more precisely matches the group than an L3: code. The limitations are as follows:
   a To be coded L310, L320, L351, or L352, C may be bonded to a halogen atom, but not to O, S, Se, Te, or N.
   b To be coded L340 or L355, C may not be bonded to O, S, Se, Te, N, or halogen.
   c To be coded L350 or L353, C may be bonded to O, S, Se, Te, or halogen, but not to N.
4. The C atoms of groups coded L354 may be bonded to any atoms. Additional functional group codes that apply to the smaller functional groups formed by -C(=N-)- and -C(=U)-, **except** L355, are also assigned.
5. Iminodisulphide is coded L351 and K223. Iminoether and iminothioether are coded in Set L3: and also in Set H5:.

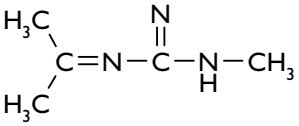
Examples Using L3: Codes

a 

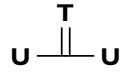
L330, H602

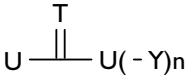
b 

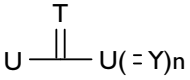
L354, J581

c 

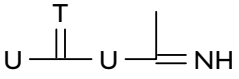
L355, L250

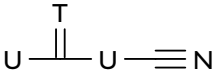
L4: 

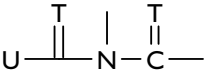
L410 



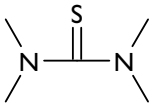
n = 1 or 2

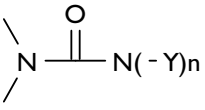


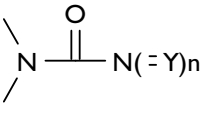




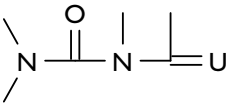
(Excludes structures that are coded L431.)

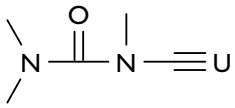
L420 

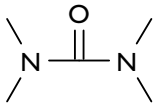
L431 

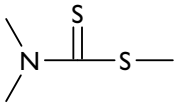


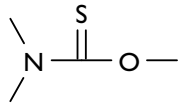
n = 1 or 2

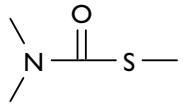


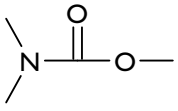


L432 Other  groups

L440 

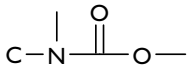
L450 



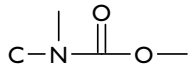
L460 

(Pre-1981 precursor code for Subset L46:) (1963-1980)

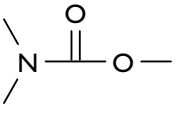
L461 Heterocyclic

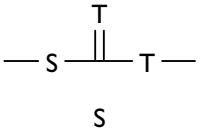
 (1981)

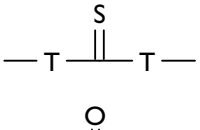
L462 Aromatic

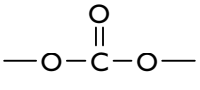


L463 Other

 groups (1981)

L471 



L472 

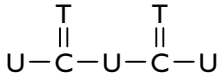
L499 Poly for Set L4:

Notes on L4: Codes

1. Groups coded L410 are additionally assigned all other applicable code(s) from Set L4:, **except** L431 and L432. If L431 or L432 are applicable, they are coded **instead of** L410.

   In example (a) below, L410 is used for the N-C(=S)-N-C(=O)- fragment; whilst L431 covers the -C(=S)-N-C(=O)-N part. L420 is applied to further define the thiourea.

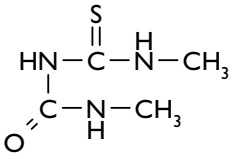
2. Groups with the structural formula shown below are assigned two or more codes from L4:, but no codes from L5:.



3. Groups coded L471 do not receive additional codes to identify the atom that "T" represents. In particular, the codes L650 and L660 are not assigned to groups coded L471.

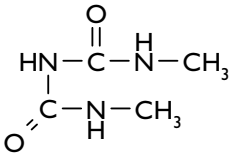
Examples Using L4: Codes

a



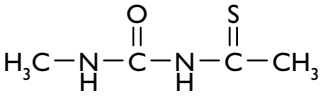
L410, L420, L431 (but not L531)

b



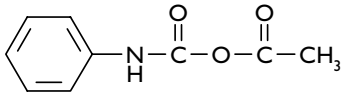
L431, L499 (but not L532)

c



L431, L531

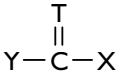
d



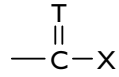
L462, L541

L5: Other groups with >C=O or >C=S

L511

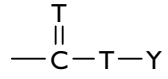


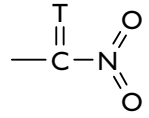
L512



(Excludes structures coded L511.)

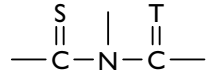
L520



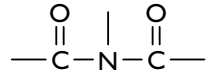


(If T is O in these definitions, then Y may not be O or N. Use L550 also if applicable.)
See note 4 for more details.

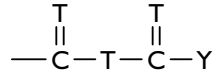
L531



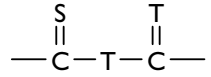
L532

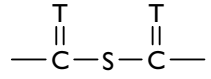


L541



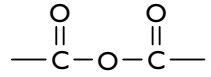
L542





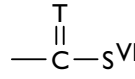
(Excludes structures coded L541.)

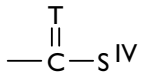
L543



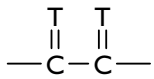
(Excludes structures coded L541.)

L550

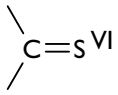


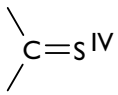


(Use L520 also if applicable.)

L560 

See Note 7 for more details.

L570 



See Note 8 for more details

L599 Poly for Set L5:

Notes on L5: Codes

1. The definitions for the codes L512 through L543 have one or more C atoms with unspecified attachments. These codes are not used when **all** of the unspecified atoms bonded to C are O, S, Se, Te, N, and/or halogen. The codes L550, L560, and L570 are applicable even when all atoms bonded to C are O, S, Se, Te, N, and/or halogen.

2. From 1978 forward, the X atom of groups coded L511 or L512 has been identified by a halogen code from Set C0: The parent acid of compounds coded L511, i.e. the corresponding acid of formula Y-C(=T)-OH, is additionally assigned all applicable functional group codes, including the code L410 if applicable. (Y is O, S, Se, Te, N, or halogen; T is O, S, Se, or Te).

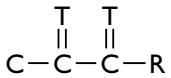
3. Groups coded L512, L531, L541, or L542 do not receive additional codes to identify the particular atom that "T" represents.

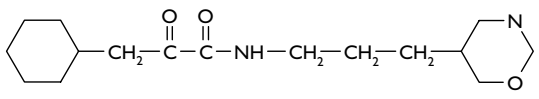
4. Groups coded L520 receive additional codes that describe the T-Y portion (coded as a bond between heteroatoms in Part K:), the -C(=T)-N< portion (coded as an amide or thioamide in Set J3:), and -NO2 (coded as nitro in Set H3:). The group C(=T)-T does not receive additional codes. Please note that prior to 1981 this code was not reliably applied.

5. Groups coded L541 receive additional applicable codes from L440 through L472, as well as the code L511 if applicable.

6. Groups coded L550 receive additional codes for the functional group formed by S, but such groups are not additionally coded as thioesters.

7. The code L560, although green on the coding sheet, is searchable prior to 1981 for the group shown below in (a). The negation code for this group (inlcuding K0), however, has only been assigned since 1981. The functional groups formed by the two C=T portions of this group are also coded individually as C-C(=T)-C (oxo or thioxo in Set J5:) and C-C(=T)-R (which could be an ester, amide, aldehyde, etc., depending on the definition of R). For example, the structure shown in (b) below is coded J011 ("1 COOH deriv."), J371 ("1 amide bonded to an aliphatic group via C=O"), J581 ("1 aliphatic oxo"), and L560.

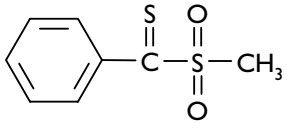
a 

b 

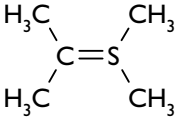
where R is H, O, S, Se, Te, N, or halogen)

8. Although black on the coding sheet, the code L570 is only searchable prior to 1981 for groups in which the S atom is bonded solely to C atoms. L570 is coded even if the >C=S is part of a larger functional group that receives its own code. For example, the group >C=S=O is coded L570, K442 (">C=S=O") and C216 ("S with a valency of IV"). The group -C(=S)-Y (S is S^{IV} or S^{VI}; Y is O, S, Se, Te, N, or halogen) receives the code L570 and a code from J1: ("Carboxylic acid, thiocarboxylic acid"), J2: ("Ester, Thioester"), or J3: ("Amide, thioamide"). The valency of S in groups coded L570 is always additionally coded.

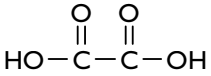
Examples Using L5: Codes

a 

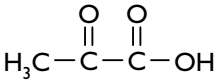
L550, K442, C316

b 

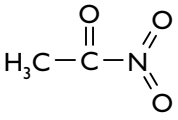
L570, C216

c 

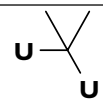
L560, J012, J172

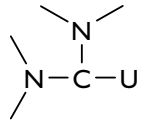
d 

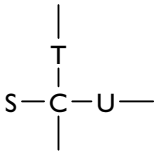
L560, J011, J171, J581

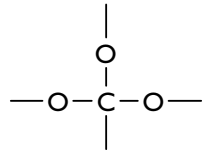
e 

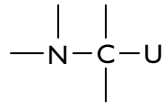
L520, H381, J371

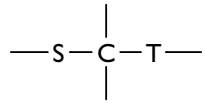
L6: 

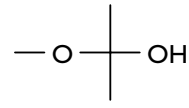
L610  (1981)

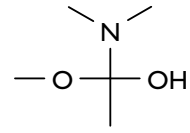
L620 

L630 

L640  (1981)

L650 

L660  including



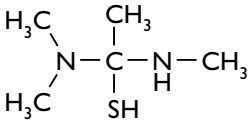
Notes on L6: Codes

1. Groups coded L610, L620, or L630 are not additionally assigned the less-specific codes L640, L650, or L660.

2. The C atom in the definitions of L640, L650, and L660 may be a ring member, but the adjacent heteroatoms may not be members of the same ring. None of these three codes are assigned

if the unspecified attachments of C in their code definitions constitute a multiple bond to a heteroatom.

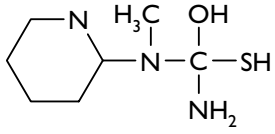
3. The smaller functional groups formed by N and by U in fragments coded L610 and L640 are additionally assigned any applicable codes. For example, the structure shown below is assigned the codes H102 ("One or more secondary, non-ring amines"), H103 ("One or more tertiary, non-ring amines"), H182 ("Two amines bonded to aliphatic C atoms"), H498 ("One or more -SH groups bonded to aliphatic C atoms"), and L610.



4. If the functional group S-C(-T)-N< is present, the overall group is coded L620 and the smaller functional group formed by the N atom is also assigned any applicable codes. See example (d) below.

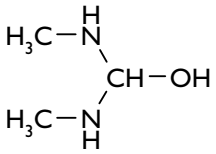
5. Groups coded L620 or L650 do not receive an additional code to identify the particular atom that "T" represents.

6. Hydrates of formula HO-C-OH and HO-C-SH are not coded L650 or L660, respectively. Instead, they are assigned the codes for the corresponding >C=O or >C=S group, i.e. water is removed. For example, the structure shown below is assigned the code L430, not L610 or amino codes.



Examples Using L6: Codes

a



L610, H102, H182, H401, H481

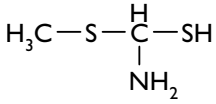
b $CH_3O\text{-}CH_2\text{-}OH$

L660

c $HO\text{-}CH_2\text{-}OH$

J471 (but not L660)

d



L620, H100, H181

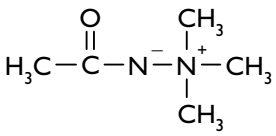
L7: MISCELLANEOUS ELEMENT DESCRIPTORS - ONIUM, FREE RADICALS, IONS, VALENCIES

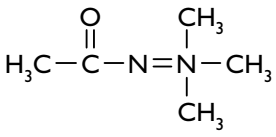
L710	a $S^{IV}R_4$ (R is H, C, O, S, Se, Te, N, or halogen, and may vary independently)
	b $S^{VI}R_6$
	c $R{=}S^{VI}R_4$
	d Any group in which S^{IV} or S^{VI} is bonded solely to C and/or H See Note 1 for more details.
L721	One N^+, in ring
L722	One N^+, in chain
L723	More than one N^+, one or more of which is in a ring
L724	More than one N^+, none of which are in rings
L730	C^+, O^+, S^+, Se^+, Te^+, F^+, Cl^+, Br^+, I^+
L740	C^-, N^-

L750	C^-, O^-, S^-, Se^-, Te^-, N^-, F^-, Cl^-, Br^-, I^- (Includes all organic free radicals)
L760	N^V bonded only to C atoms; X^{III} ; X^V ; X^{VII}
L799	Poly, for codes L710, L730, L740, L750, L760

Notes on L7: Codes

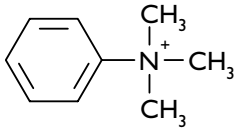
1. Although L710 is black on the coding sheet, it is only searchable prior to 1981 if the S atom is bonded solely to C and/or H atoms. Such groups are also coded C216 ("S^{IV} or S^V") or C316 ("S^{VI}"). S-Y or S=Y bonds present in compounds coded L710 are additionally coded in Part K: (Y is O, S, Se, Te, N, or halogen). L710 does not cover onium salts.
2. The coding of quaternary ammonium compounds as amines is discussed in the section entitled "Quaternary Ammonium Compounds" on page 154.
3. The central element in the definitions of L730, L740, L750, and L760 also receives an element code from Part C:.
4. Charge separation formulae, i.e. structures having opposing charges on adjacent atoms, are coded with a multiple bond instead of the charges. For example, the structure shown in (a) below is indexed as the structure shown in (b); thus, the structure in (a) is indexed J371 ("One amide or thioamide group bonded to H or aliphatic C via C=O or C=S") and K532 ("-N=N-C(=U)-"; U is O, S, Se, Te, N, or halogen) rather than in Set L7:.

a 

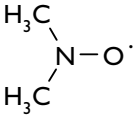
b 

5. The code L740 is not used for metal salts. Guidelines for coding metal salts are found in the Notes accompanying Part A: codes, which begin on page 39.
6. Valency (L710, L760) is defined by the number and type of bonds an atom participates in, not by change, so N^+R_4 is 4-valent and has not got the L760 code.

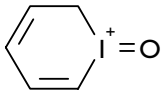
Examples Using L7: Codes

a 

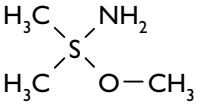
L722, H141

b 

L750, C108, K850

c 

L730, C053, C200, C600, K140

d 

L710, K340, C216
(not L760 as L730 takes presidence)

L8: SUGARS AND DERIVATIVES

Notes on L8: Codes

1. Codes in Set L8: are used to describe both cyclic and acyclic sugars and derivatives that have two or more asymmetric centres.
2. From 1981 forward, the codes in Subsets L81: and L83: have been assigned to all polysaccharides to which the code V735 ("Polysaccharide") is applicable. Polysaccharides of unknown structure are coded L810. Polysaccharides do not receive any codes from Subset L82:.
3. Functional groups present in a sugar residue, e.g. -OH or -CHO, receive applicable functional group codes in addition to sugar codes.
4. A sugar containing a free aldehyde or ketone group is always searched in the open chain form (as illustrated in example b on the following page).
5. Glyceraldehyde, pentaerythritol, cyclitols, and other similar compounds are not coded in Set L8:.

L81: Type of configuration (sugar or residue)

L810	Unspecified configuration	
L811	Arabinose	(1981)
L812	Ribose	(1981)
L813	2-Deoxyribose	(1981)
L814	Glucose	(1981)
L815	Galactose	(1981)
L816	Mannose	(1981)
L817	Monodeoxyhexose, dideoxyhexose	(1981)
L818	Other configurations	(1981)
L819	Poly, for Subset L81: codes (from 1981/week 27 forward)	(1981)
L899	Poly, for Subset L81: codes (prior to 1981/ week 27)	(1963-1981)

Notes on L81: Codes

1. Subset L81: is used to indicate the type of configuration of a sugar or sugar residue. These codes also cover all derivatives that have the same configuration, other than higher sugars.
2. Deoxy-sugars, i.e. sugars in which C-OH is replaced by C-H or >C=O is replaced by >CH2, are not assigned the code of the corresponding oxysugar. For example, deoxymannose is coded L817, not L816.
3. The code L819, searchable since 1981/week 27, indicates that one of the sugar residues defined in codes L811 through L818 is present in a structure more than once. The code L899, searchable prior to 1981/week 27, indicates the presence of two separate sugar residues in a structure.

L82: Number of linked sugar units

L821	Monosaccharide	(1981)
L822	Disaccharide	(1981)
L823	Trisaccharide	(1981)
L824	Oligosaccharide (Four, five, or six residues in sequence)	(1981)

L83: Type of compound or residue

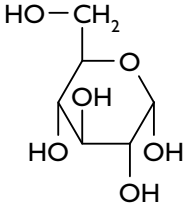
L831	Sugar (excluding those covered by L832 through L835)
L832	Oxidised sugar
L833	Reduced sugar
L834	C atom of sugar bonded to N
L835	C atom of sugar bonded to an atom other than H, N, or O

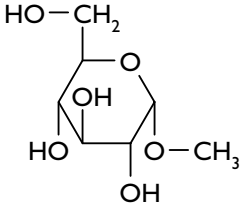
Notes on L83: Codes

1. L831 covers all sugar derivatives for which none of the other codes from Subset L83: are applicable, including esters, ethers, acetals, etc.

2. L832 covers acids, ketones, unsaturated sugars, and their derivatives, e.g. esters and lactones.
3. L833 covers sugar alcohols and their derivatives, e.g. esters, ethers, cyclic ketals, etc.
4. L834 covers compounds such as sugar amines, amides of sugar acids (which are also coded L832), nucleosides, nucleotides, and osazones.
5. L835 covers compounds such as thiosugars, halides of sugar acids (which are also coded L832), and sugars in which the central chain of C atoms is either branched or contains more than six C atoms. The C atoms in the sugars with a branched central chain or with a central chain of more than six C atoms may be ring members.

Examples Using L8: Codes

a  glucose

b 

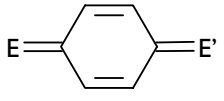
L814, L821, L831, H484,
L814, L821, L831, H404,
H405, J471, M620, etc.
H423, H481, H521, F123, etc.

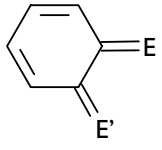
L9: GROUPS IN RINGS

L910	-U-C(=U)-U-C(=U)-	(1981)
L921	>N-C(=O)-N<	(1981)
L922	U-C(=U)-U (Excludes structures coded L921)	(1981)
L930	-C(=U)-U-C(=U)	(1981)
L941	-C(=O)-N<	(1981)
L942	-C(=O)-O-	(1981)
L943	-C(=U)-U (Excludes structures coded L941 or L942)	(1981)
L951	Quinone or thioquinone	
L952	Derivative of quinone or thioquinone	
L960	-C(=U)-C(=U)- (Excludes structures coded in Subset L95:)	(1981)
L970	-S(=U)$_n$-U ; U-S(=U)$_n$-U (n = 1 or 2)	(1981)
L980	-O-O-	(1981)
L999	Poly for Set L9:	(1981)

Notes on L9: Codes

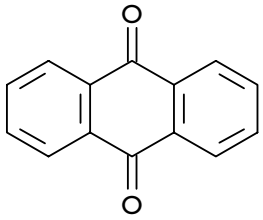
1. For a Set L9: code to be applicable, the atoms underlined must be members of the same ring, and the atoms that are not underlined must not be members of the same ring system as the underlined atoms.
2. Except for L951, L952, and L960, the codes in Set L9: include optionally substituted tautomeric forms. These have not been reliably applied, however, for substituted tautomers, and should be used with care in such cases.
3. The code L951 covers the quinone/thioquinone structures A and B shown below.

a 

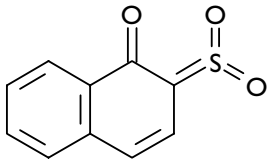
b 

E and E', which may be the same or different, are any of the following atoms: O, S, Se, or Te. E and E' may be bonded to further atoms. The valency of S, Se, and Te in quinones and thioquinones are additionally coded C116 ("S^{II}"), C216 ("S^{IV}"), or

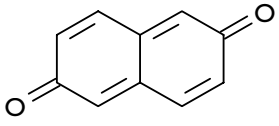
C316 ("S^{VI}"). Extended quinones and thioquinones in which **E** and **E'** are substituents on different rings of a fused ring system are not searchable as L951. Three examples of these rules are shown below.

a 

L951

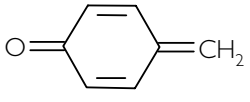
b 

L952, C316, K442, L570

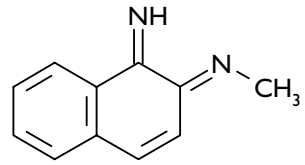
c 

J562

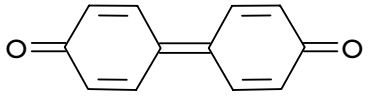
4. The code L952 covers the structures A and B shown in Note 3 above when at least one of the atoms **E** or **E'** is an atom other than O, S, Se, or Te. **E** and **E'** may be further bonded to other atoms. Functional groups formed by **E** and/or **E'** receive all applicable codes. Extended quinones and thioquinones in which **E** and **E'** are substituents on different rings of a fused ring system are not searchable as L952. Four examples of these rules are shown below.

a 

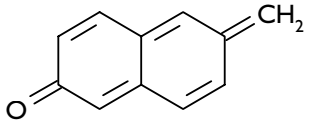
L952, H720

b 

L952, L355, L399

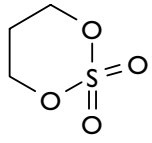
c 

L952, L999, H720

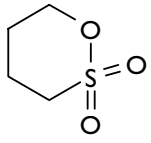
d 

J561, H720 (but not L952)

5. The following structures are searched with the same codes, including L970, K441, C316, and F320, but can be distinguished by their Ring Index Numbers (RIN's).

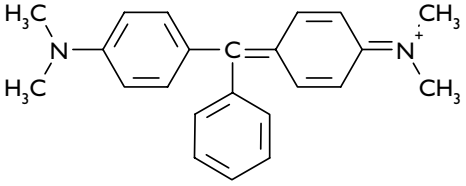


RIN is 00219

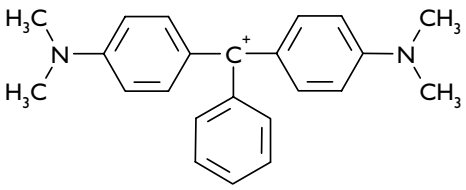


RIN is 00253

6. Prior to 1981, diphenyl and triphenyl methane dyes were coded as quinone derivatives in CPI Sections B and C, but as carbonium salts in Section E.



Section B (Farmdoc) and C (Agdoc)

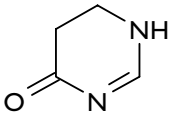


Section E (Chemdoc)

From 1981 forward, such structures are coded as carbonium salts in all three CPI sections.

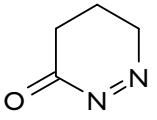
Examples Using L9: Codes

a



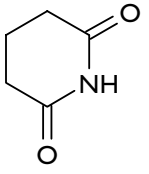
L941, J521 (but not L943)

b



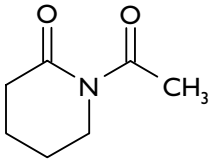
J521, L941

c



L930, J522 (but not L941)

d

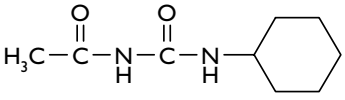


L941, J371, J521, H211 (but not L930)

Examples Using Codes from Parts K: and L:

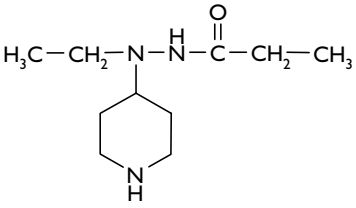
Note All applicable codes from Parts C:, H:, J:, K:, and L: are listed in the following examples.

a



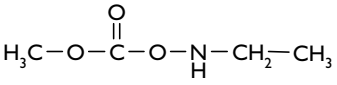
L431, L532

b



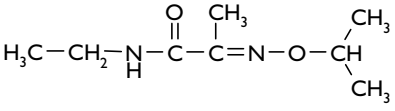
K620, J011, J371

c



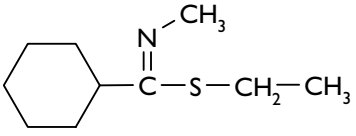
L410, L472, K820

d



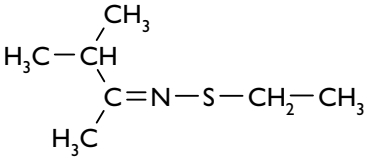
K840, L354, J371, J011

e



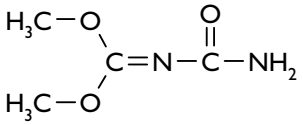
L352, H598

f



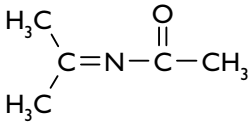
L355, K353, C116

g



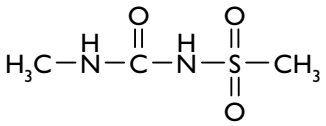
L290, L432 (but not L352 or L660)

h



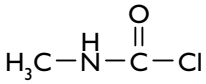
L355, J011, J371

i



K351, L431, L570, C316

j



L463, L511, C017

9 Part M: Miscellaneous Descriptors

9 Part M: Miscellaneous descriptors

Main Headings

M1: Groups Linking Ring Systems
M2: Carbon Chains (0-valent or Monovalent)
M3: Carbon chains (Polyvalent)
M4: Basic Groups
M5: Ring Systems Present
M6: Miscellaneous Structure Codes
M7: Patent Type
M8: Stereochemistry
M9: Control Codes

Introduction

Part M: codes are called "Miscellaneous Descriptors" because they cover a wider variety of concepts than other code Parts. Part M: covers bonds and linking groups between rings, carbon chains, type of structure, number and type of rings present, the role of the compound(s), stereochemistry, time ranging and codes for retrieving various subsets of the DWPI database.

Notes on Part M: Codes

1. The symbols used in Part M:, whch can vary independently when appearing more than once in a fragment, are defined as follows:

   U = O, S, Se, Te or N
   V = atom other than C, H, O, S, Se, Te or N
   V' = atom other than C, H, O, S, Se, Te, N, or halogen
   W,W',W" = heteroatoms

2. The **precursor codes** in Part M:, i.e. M260, M270, M350, M360, M370, and M380 are only searchable between the years 1972 and 1981. Thus the codes marked with a **#** on the coding sheet are searched by 'OR'ing them with the corresponding precursor code in the search statement for codes introduced in 1972, i.e. with the codes that are blue on the coding sheet. (Precursor codes are discussed in the Chemical Code Overview, on page 12.)

M1: GROUPS LINKING RING SYSTEMS

Notes on M1: Codes

1. Codes in Set M1: are assigned to the linking bonds or groups between **C atoms** of two separate ring systems, as well as the types of rings linked. Codes from Set M1: are only assigned when all of the following conditions are met:

- The entire linking group must either be a bond or be fully described by one of the codes in Subset M13: or Subset M14:;
- The linking group must be bonded to C atoms in both rings;
- No atom in the straight chain between the two rings may itself be a ring member (except in the case of M150 - consult its definition for details); and
- The two rings to which the linking group is bonded cannot form part of the same ring system.

2. Definitions of terms used in Set M1:

   Benzene – an unfused benzene ring or benzoquinone
   Aryl – a benzene ring fused to one or more carbocyclic or heterocyclic rings
   Other – a ring other than fused or unfused benzene

3. In most cases the code for a linking group depends only on the chain of atoms directly connecting the two rings together. Other atoms bonded to the linking atoms and the type of bonds between the linking atoms (single or multiple) do not affect the code unless a specific code is provided for the arrangement. For example, M131 specifies that a heteroatom atom is bonded to the linking group, and M133 specifies double or triple bonds in the linking group.

4. The codes in Subsets M13: and M14: are not used for describing only a part of the group linking two rings; the entire direct line of atoms between the two rings must be fully described by a code for that code to be assigned.

5. If no code in Set M1: is used, the code M1 can be negated. (See page 108 for information about negation codes.)

Subheadings Applicable to M1: Codes

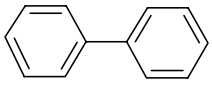
Nearly all codes in Set M1: are applicable to subheadings M2, M3 and M4 from 1970 and to M1 from 1981. Exceptions are M150 (applicable to subheading M0 from 1963 and to subheadings M1 to M4 from 1970) and the negation code M1 (applicable to subheadings M1 to M4 from 1981).

M11: Rings linked by a single or double bond	
M111	***Benzene*** - bond - ***benzene***
M112	***Benzene*** - bond - ***aryl***
M113	***Benzene*** - bond - ***other***
M114	***Aryl*** - bond - ***aryl***
M115	***Aryl*** - bond - ***other***
M116	***Other*** - bond - ***other***
M119	Poly, for Subset M11:

Note Definitions for ***benzene***, ***aryl*** and ***other*** are given in the Introduction to Part M:.

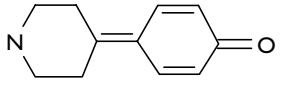
Examples Using M11: Codes

a



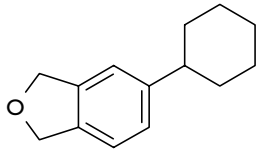
M111

b



M113 (because ring at right is coded as benzene)

c



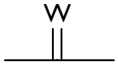
M115

M12: Rings linked by linking groups

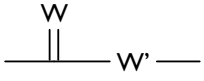
M121	***Benzene*** - to - ***benzene***
M122	***Benzene*** - to - ***aryl***
M123	***Benzene*** - to - ***other***
M124	***Aryl*** - to - ***aryl***
M125	***Aryl*** - to - ***other***
M126	***Other*** - to - ***other***
M129	Poly, for Subset M12:

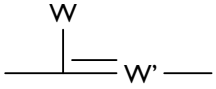
Note Definitions for ***benzene***, ***aryl*** and ***other*** are given in the Introduction to Part M:.

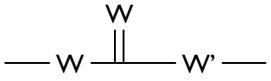
M13: Linking groups containing C atom(s)

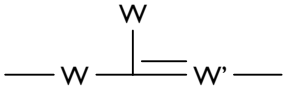
M131 

M132	1 C (excluding structures coded M131)
M133	-C=C -; -C≡C-
M134	Polymethine; azamethine; polyazamethine
M135	2 or more C atoms (excluding structures covered by previous codes)

M136 



M137 



M139 Poly, for Subset M13:

Notes On M13: Codes

1. The free bonds at either end of the group indicate the points of attachment to the rings.
2. M134 is not assigned to linking groups coded M133, M136, M145, or M146. The double bonds in groups coded M134 must be conjugated throughout the bridge (i.e. the direct line of atoms linking the two rings), and the bridge may be bonded to the ring(s) via double bonds.
3. The code M139 is used to indicate that one or more of the codes from M131 through M137 applies to more than one group of atoms linking different rings; it is not used to indicate the same group of atoms linking more than one pair of rings. (See example n on page 173).

M14: Linking groups containing no C atom(s)

M141	- O -
M142	-S- or -Se- or -Te-
M143	-N- or =N-
M144	Single heteroatom other than O, S, Se, Te, N
M145	-N=N-
M146	$-(W)_n-$ (n > 1, excluding -N=N- covered in M145)
M147	- W - W' -
M148	- W -W -W'' -
M149	Poly, for Subset M14:

Notes on M14: Codes

1. The free bonds at either end of the above groups indicate the points of attachment to the rings.
2. All **W** atoms in the definition of M146 must be the same. For M147 and M148 the **W** atoms must not all be the same.
3. The code M149 is used to indicate that one or more of the codes from M141 through M148 applies to more than one set of atoms linking different rings; it is not used to indicate the same group of atoms linking more than one pair of rings.

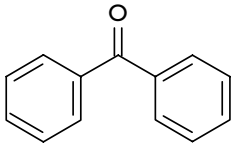
M150:	Geminal aromatic linking group R-C-R'	1963

Notes on M150

1. **R, R'** = aryl ring systems bonded via benzene ring or quinonoid derivative of benzene to the C atom shown above. The linking group C atom may form part of a ring, be multiply bonded to a heteroatom, or be substituted by additional aryl group(s).
2. **R** and **R'** cannot be part of the same ring system.

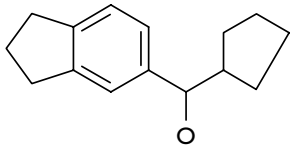
Examples Using Codes M12:, M13:, M14:, and M150

a



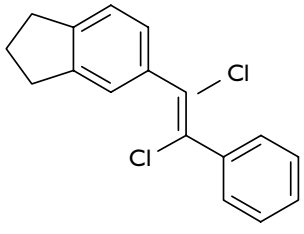
M121, M131, M150

b



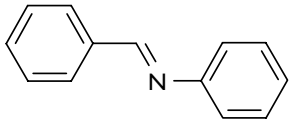
M125, M132

c



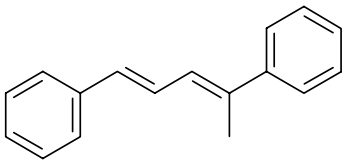
M122, M133

d



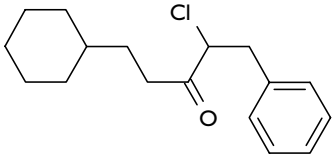
M121, M134

e



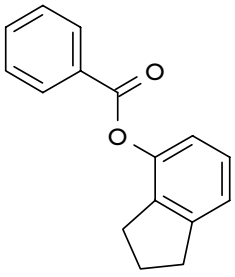
M121, M134

f



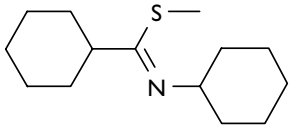
M123, M135

g



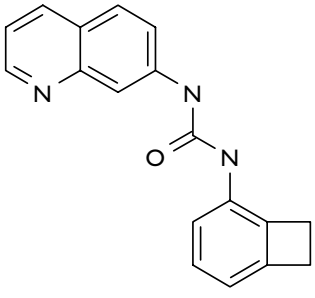
M122, M136

h



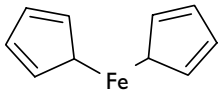
M126, M136

i



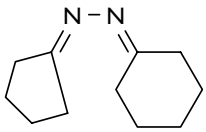
M124, M137

j



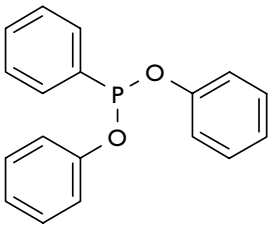
M126, M144

k



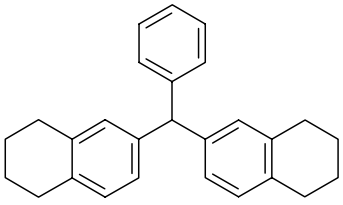
M126, M146

l



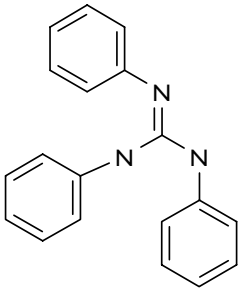
M121, M129, M147, M148, M149

m



M122, M124, M129, M132, M150 (but not M139)

n



M121, M129, M137 (but not M139)

Sets M2: and M3: Carbon Chains

Introduction

A ***carbon chain*** is defined in the chemical coding system as an optionally branched, optionally unsaturated aliphatic hydrocarbon group. The carbon chain ends with, but includes, the C atom(s) bonded to any of the following ***terminating atoms***:

- Ring C;
- Chain C multiply bonded to a heteroatom; or
- Any heteroatom (i.e. any atom other than C or H).

The ***valency*** of a carbon chain is defined as the total number of bonds linking the carbon chain to **terminating atoms**. The number of C atoms in the chain includes those in branched parts of the chain. Thus a tert-butyl substituent that is not multiply bonded to a heteroatom would be assigned the code M214 ("4 C atoms in the carbon chain"). C atoms that are terminating atoms are NOT considered as part of the chain when counting the number of C-atoms, so CH_3-CO_2H is a 1-C chain.

Selecting Codes for Carbon Chains

Sets M2: and M3: consist of codes used to describe carbon chains. Caution should be used when devising search strategies for carbon chains, both because of the way chemical patents are written and because of the way patents are indexed in the chemical coding system. Many patents give only general information about the carbon chains in the structures they cover; others give a wide range of possible lengths of the chains. Thus, creating a chemical code search that is too precise may eliminate relevant references.

Additionally, carbon chains with a valency of 0 or 1 are coded in such a way that comprehensive retrieval is best achieved by searching the general length code(s), i.e. M210 for carbon chains of 1 to 6 C atoms and M220 for carbon chains of 7 to 10 C atoms. The reason for this is explained in the section "Notes on M2: Codes".

With these points in mind, one rule of thumb for using the carbon chain Chemical Codes is: "When in doubt, leave it out."

M2: 0-VALENT, MONOVALENT CARBON CHAINS

Notes on M2: Codes

1. Set M2: codes are used to describe isolated carbon chains or carbon chains bonded to one terminating atom.

2. One of the general codes M210 or M220 is always added to the code record when a specific code from Subset M21: or Subset M22: is assigned. For example, a propyl chain that is not multiply bonded to a heteroatom receives the codes M210 and M213. Also, prior to 1981 the specific codes in Subsets M21: and M22: were only assigned if there were no polyvalent chains present; from 1981 forward, the specific codes have only been assigned if the chain length is specifically claimed or exemplified. Therefore, **comprehensive retrieval** is best achieved by searching M210 instead of the more specific chain lengths.

3. If a structure contains both a 0- or 1-valent carbon chain and a polyvalent chain, and if the user chooses to include a code for the specific number of C atoms present (e.g. when searching for Et- at a position where previous references have all used Methyl), that code is placed in the search statement with the codes from the time range 1981/Week 27 to the present (i.e. with the codes that are green on the coding sheet). The applicable general code, however, i.e. M210 or M220, is placed in the search statement with the codes from the time range 1972 to 1981/Week 26 (i.e. with the codes that are blue on the coding sheet).

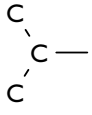
Subheadings Applicable to M2: Codes

Set M2: codes are applicable to subheadings M2, M3 and M4 from 1972 and to subheading Ml from 1981. Exceptions are M232, M233 and M240, which are applicable to subheading M0 from 1963 and to subheadings Ml to M4 from 1970.

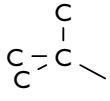
M21: Number of C atoms in chains 1 to 6 C	
M210	1-C to 6-C (general)
M211	1-C
M212	2-C
M213	3-C
M214	4-C
M215	5-C
M216	6-C

M22: Number of C atoms in chains 7 C	
M220	7 C to 10 C (general)
M221	7- C
M222	8- C
M223	9- C
M224	10- C
M225	11- to 18- C
M226	≥19- C

M23: Carbon chain branching	
M231	Straight chain
M232	Carbon chain containing secondary branching



M233	Carbon chain containing tertiary branching



(1963)

Notes on M23: Codes

1. M231 has not been used for methyl or ethyl chains since 1981.
2. Carbon chains coded M232 or M233 are not additionally coded M231.
3. In the case of codes M232 and M233, none of the C atoms shown may be multiply bonded to a heteroatom, nor may they be ring members (either of which would exclude them from being considered part of the carbon chain). The free bond must be attached to an atom other than hydrogen; for example, tert-butyl alcohol and neopentyl benzene are both coded M233.
4. Carbon chains containing C=C or C≡C are assigned the same code in Subset M23: as the corresponding hydrogenated chains. For example, CH_2=C(-CH_3)-CH_3 and CH_3-CH(-CH_3)-CH_3 are both coded M232.
5. Care should be taken when searching M232 and M233, since the previously used "punch codes" were inconsistently assigned before 1972. (Punch codes are discussed briefly on pages 3 and 13.)

M240	Chain bonded to ring C (1963)

M250	Chain bonded to V, C=V, or C≡V

Note: V = atom other than C, H, O, S, Se, Te, or N

M26: Chain bonded to C=U or C≡U		
M260	Chain bonded to C=U or C≡U (Pre-1981 precursor code Subset M21:)	(1972-1981)
M261	Carbon chain bonded to C=S, C=Se or C=Te	(1981)
M262	Carbon chain bonded to C=O	(1981)
M263	Carbon chain bonded to C=N or C≡N	(1981)

Note: In the case of tautomeric forms, the first applicable code is assigned; thus, CH3-C(=O)-SH is coded as its tautomer CH3-C(=S)-OH (M261).

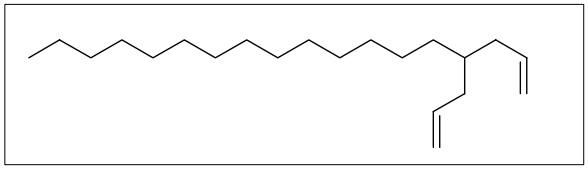
M27: Chain bonded to U		
M270	Chain bonded to U (Pre-1981 precursor code for Subset M27:)	(1972-1981)
M271	Chain bonded to S, Se or Te	(1981)
M272	Chain bonded to O	(1981)
M273	Chain bonded to N	(1981)

M28: Multipliers for Subsets M24: through M27:	
M280	No 0-valent or monovalent carbon chains present
M281	One or more of codes M240 through M273 applicable once
M282	One or more of codes M240 through M273 applicable twice
M283	One or more of codes M240 through M273 applicable 3 times

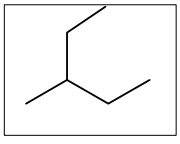
Examples Using M2: Codes

The chain is shown in the boxes

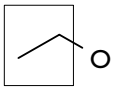
a



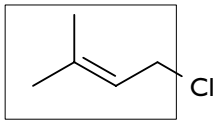
M226, M232, M320, M610

b 

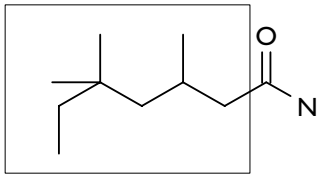
M210, M216, M232, M320, M610, M620

c 

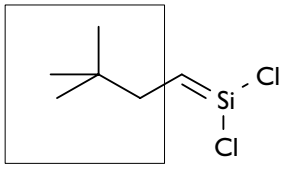
M210, M212, M272, M281, M320, M620

d 

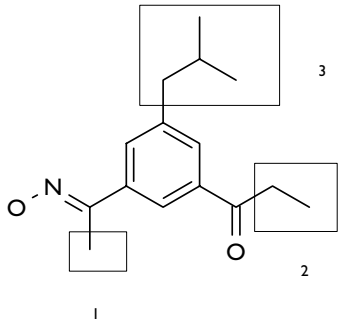
M210, M215, M232, M250, M281, M320

e 

M220, M223, M232, M233, M262, M281, M320, M620

f 

M210, M215, M233, M250, M281, M320, M620

g 

1 M210, M211, M263
2 M210, M212, M262
3 M210, M214, M232, M240

Note Chains (1) and (2) may be searched with the precursor code M260 for references added to the database between 1972 and 1981. Although all other carbon chain codes for this structure are used only once each and are therefore accompanied by the multiplier code M281 ("One or more of codes M240 through M273 applicable once"), the code M260 may be used twice for references between 1972 and 1981, and was therefore indexed with the additional code M282 ("At least one of codes M240 through M273 applicable twice") during those years. (Precursor codes are discussed in the section "Code Modifications in 1981 Week 27", beginning on page 13.)

M3: POLYVALENT CARBON CHAINS (VALENCY 2)

Notes on M3: Codes

1. ***Polyvalent*** carbon chains are carbon chains bonded to more than one terminating atom. See introduction to Sets M2: and M3: on page 170 for definitions of ***carbon chain***, ***terminating atom*** and ***valency***.
2. Sets M35: through M38: use the notation "**And** (**U and/or C=U and/or C≡U**)" to mean that at least one of the options in parentheses must be present, but any two or all three could also be present. The expression "**And optionally** (**U and/or C=U and/or C≡U**)" means that any, all, or none of the options in parentheses may be present in a structure to be a match for the code in question.

Subheadings Applicable to M3: Codes

Set M3: codes are applicable to subheadings M2, M3 and M4 from 1972, and to subheading Ml from 1981. Exceptions are M331, M333 and M341 (applicable to subheading M0 from 1963 and to subheadings Ml to M4 from 1970).

M31: Number Of C atoms in polyvalent chain

M311	1- C
M312	2- C
M313	3- C
M314	4- C
M315	5- to 8- C
M316	≥9- C

M32: Multipliers for Subset M31:

M320	No polyvalent chains present
M321	One or more of M31: codes used once
M322	One or more M31: code is used twice
M323	One or more M31: code is used three or more times

Notes on M32: Codes

1. Search strategies for pre-1981 references must be modified with regard to the multiplier code used with M313 and M314, since these two codes were represented by a single punch code (050) prior to 1981. (Punch codes are discussed in the Introduction to the manual on page 3.) Consider a search strategy in which M313 and M314 are each used once. After 1981 the multiplier code would be M321, because M313 and M314 were each used once. Prior to 1981, the punch code 050 would have been applicable twice, once to cover the 3-C chain, once to cover the 4-C chain; therefore, the code M322 would have also been assigned. To find references like this, the search term (**M321 OR M322**) should be included in the search statement with codes that were introduced in 1972, i.e. the codes that are blue on the coding sheet.

M33: Straight or branched carbon chains

M331	Straight carbon chain with -CH3, -C=CH2, and/or -C≡CH	(1963)
M332	Straight carbon chain with none of the groups -CH3, -C=CH2, or -C≡CH	
M333	Branched carbon chain with -CH3, -C=CH2, and/or -C≡CH	(1963)
M334	Branched carbon chain with none of the groups -CH3, -C=CH2, or -C≡CH	

Notes On M33: Codes

1. From 1981 forward, M332 has not been assigned to chains with 1 C atom.
2. Chains of the type Q-$(CH_2)_n$-CH=(ring C) or Q-$(CH_2)_n$-CH=C=W are coded M332 and not M331

Note Q is -CH3, -C=CH2, or -C≡CH; W = any heteroatom; n=0; $(CH_2)_n$ may contain C=C and/or C≡C unsaturation).

3. M333 and M334 are only used for carbon chains containing one of the fragments shown below .

None of these C-atoms may be a ring member or be multiply bonded to a heteroatom.

4. Care should be taken when searching the codes M331 and M333, since the punch code 536, (which was used to indicate Q-containing chains prior to the creation of codes M331 and M333) may not have been consistently assigned prior to 1981. (Punch codes are discussed in the Introduction to the manual on pages 3 and 13.)

M34: Valency specification of polyvalent carbon chains

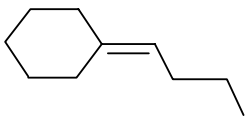
M340	All valencies are to the same C atom in a chain of 2 or more C atoms (excludes structures coded M341)	
M341	Aliphatic hydrocarbon group bonded to ring C via double bond	(1963)
M342	2-valent	
M343	3-valent	
M344	4-valent	
M349	C^1(-W)-C^2(=W') or C^1(-W)-C^2(≡W')	(1981)

Notes on M34: Codes

1. In the definition of M349, C^1 is a member of the carbon chain being coded and C^2 is not a member of a carbon chain, nor is it a ring member. W and W' may be the same or different heteroatoms. M349 is used in addition to other chain codes.
2. A double bond to a carbon chain counts as two valencies, a triple bond as three.

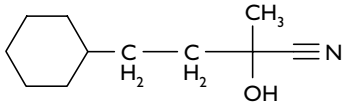
Examples Using M34: Codes

a



M341
(also M280, M314, M321, M331)

b



M343, M349
(also, M280, M314, M321, M331, M371, M391)

M35: Carbon chain bonded to ring C and (V and/or C=V and/or C=V) and optionally (U and/or C=U and/or C=U)

M350	Pre-1981 precursor code for Subset M35:	(1972-1980)
M351	Bonded to ring-C *and* (V' and/or C=V' and/or C=V') and optionally halogen	(1981)
M352	Bonded to ring-C *and* halogen *and* (U and/or C=U and/or C=U) only	(1981)
M353	Bonded to ring-C and halogen only	(1981)

Note V = atom other than C, H, O, S, Se, Te or N; V' = atom other than C, H, O, S, Se, Te, N, or halogen; U = O, S, Se, Te or N

M36: Carbon chain bonded to (V and/or C=V and/or C=V) and optionally (U and/or C=U and/or C≡U), *but not* to a ring C

M360	Pre-1981 precursor code for Subset M36:	(1972-1980)
M361	Bonded to (V' and/or C=V' and/or C=V') and optionally halogen, U, C=U, and/or C≡U	(1981)
M362	Bonded to halogen *and* (U and/or C=U and/or C≡U) only	(1981)
M363	Bonded to halogen only	(1981)

Note V = atom other than C, H, O, S, Se, Te or N; V' = atom other than C, H, O, S, Se, Te, N, or halogen; U = O, S, Se, Te or N

M37: Carbon chain bonded to a ring C *and* (U and/or C=U and/or C≡U), *but not* to V, C=V, or C≡V

M370	Pre-1981 precursor code for Subset M37:	(1972-1981)
M371	Bonded to ring-C *and* U *and* (C=U and/or C≡U)	(1981)
M372	Bonded to ring -C *and* (C=U and/or C≡U) only	(1981)
M373	Bonded to ring C *and* U only	(1981)

Notes on M37: Codes

1. U = O, S, Se, Te or N
2. Codes for chains bonded to ring -C only are in Subsets M13: and M34:.

M38: Carbon chain bonded to (U and/or C=U and/or C≡U), *but not* to ring C, V, C=V or C≡V

M380	Pre-1981 precursor code for Subset M38:	(1972-1981)
M381	Bonded to U *and* (C=U and/or C≡U)	(1981)
M382	Bonded to C=U and/or C≡U only	(1981)
M383	Bonded to U only	(1981)

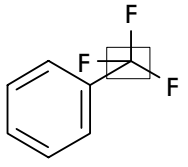
Notes on M38: Codes

1. For more information about precursor codes, see page 13.
2. U = O, S, Se, Te or N

M39: Multipliers for codes M350 to M383 (polyvalent carbon chain attachments)	
M391	One or more of codes used once
M392	One or more of codes used twice
M393	One or more of codes used 3 times

Note For structures with no polyvalent chains, search M320.

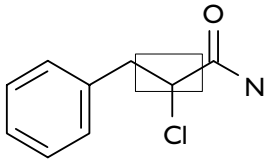
Examples Using M3: Codes

a 

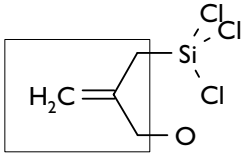
M311, M321, M344, M353, M391, M280

b 

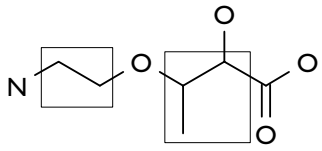
M312, M321, M332, M342, M381, M391, M280, M620

c 

M312, M321, M332, M349, M343, M352, M391, M280

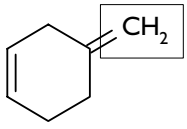
d 

M314, M321, M333, M342, M361, M391, M280

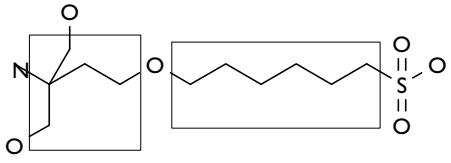
e 

1 2

1 M312, M332, M342, M383
2 M313, M331, M343, M381, M349
Overall structure: M321, M391, M280, M620.
Prior to 1981: M380, M392

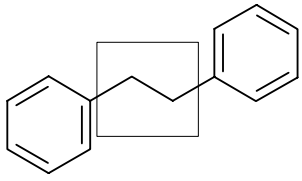
f 

M311, M321, M341, M280, M610, H720 ("Exocyclic C=C")

g 

1 2

1 M315, M334, M344, M383; Prior to 1981: M380
2 M315, M332, M342, M383; Prior to 1981: M380
Overall structure: M322, M392, M280, M620

h 

M312, M321, M332, M342, M280, M121, M135, M610

M4: BASIC GROUPS

Notes on M4: Codes

1. Basic Group codes allow the user to indicate the general category of the compound, composition or apparatus being patented.
2. Prior to 1970, components of multicomponent compositions were coded in Subset M43:, not in Subsets M41: or M42:. For example, before 1970 a fused ring heterocycle that was a component of a composition with no active ingredients was coded M430, not M412. Therefore, if multicomponent compositions of a compound are of interest, the applicable code from Subset M41: or M42: for the structure in question should be 'OR'ed with M430 ("Composition with no active ingredients") or M431 ("Composition with one or more active ingredients") as applicable.

Subheadings Applicable to M4: Codes

Set M4: codes are applicable to subheading M0 from 1963 and to subheadings Ml to M4 from 1970. In addition, the codes M417, M430 and M431 are applicable to subheading M5 from 1981.

M41: Basic group category, based on chemical structure	
M411	Miscellaneous; inorganic
M412	Fused ring heterocyclic
M413	Mononuclear heterocyclic
M414	Carbocyclic aromatic
M415	Alicyclic
M416	Aliphatic
M417	Incomplete structure disclosed

Notes on M41: Codes

1. M411 covers all inorganic compounds; free elements; organic compounds containing elements other than C, H, N, O, S and halogen; and organic compounds with halogen as a ring member.
2. Only one of the M41: codes are assigned to a single compound. The Basic Group codes are listed in the order of their priority, so that the first one which is applicable to a compound is selected. For example, a structure having both a fused ring heterocyclic system and an aliphatic portion, but which does not fit into "Inorganic or Miscellaneous", is coded as a fused ring heterocycle (M412), not as an aliphatic compound (M416).
3. Inorganic salts of organic acids or bases are not always coded M411. When searching for these salts, M411 should be 'OR'ed with the Subset M41: code for the corresponding free acid or base.
4. The Basic Group code for polymers is M423, not a code in Subset M41:.
5. The term ***aromatic*** in the definition of M414 means that the compound contains a ring system coded in one of the Sets Gl: through G4:. The term ***aromatic*** is defined on page 94.
6. If M417 is assigned, the Basic Group codes corresponding to all other compounds claimed or exemplified are also assigned.

Examples Using M411

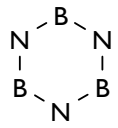
a O_2

M411

b $Ni(CO)_4$

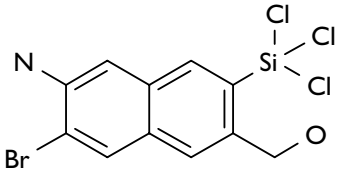
M411

c



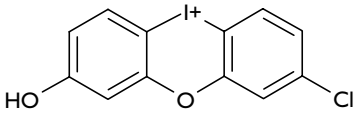
M411

d



M411, M510, M520, M531, M540

e



M411, M511, M520, M530, M540

Note More examples using Subset M41: codes are found on page 181.

M42: Natural materials, polymers, apparatus	
M421	Antibiotic or vaccine for which no structure is indexed
M422	Vitamin for which no structure is indexed
M423	Other natural products; polymers; polypeptides
M424	Apparatus

Notes on M42: Codes

1. The codes in this Subset are used for all polymers, apparatus, and natural materials of unknown structure (even when a partial structure is known).
2. The code M421 is only used with subheadings M0 and Ml. M423 is used primarily with subheadings M0 and M1, but may also be used with subheading M4 when searching for natural and polymeric dyes.
3. M424 is assigned when the apparatus is the only codeable concept in the patent, e.g. a patent on a syringe.

M43: Multicomponent compositions	
M430	Composition with no active ingredients
M431	Composition with one or more active ingredients

Notes on M43: Codes

1. Subset M43: is only used for pharmaceutical (Section B - Farmdoc) and agricultural (Section C - Agdoc) patents for which two or more ingredients of a mixture have been indexed. From 1970 forward, the applicable codes in M41: and M42: have also been assigned to each ingredient of the composition that is coded.
2. "Active" is used in a pharmaceutical, agricultural, or veterinary sense.

M5: RING SYSTEMS PRESENT

Notes on M5: Codes

1. Set M5: codes are used to indicate the number of each kind of ring system present in a structure. Codes describing the particular ring systems present in the structure are found in Chapter 7, starting on page 67.
2. From 1981 forward, M510, M520, M530, and M540 have not been assigned to aliphatic compounds; the code M416 implies these codes. However, if the M411 basic group is used rather than M416, because of the presence of Metals or other "less-common" elements (e.g. P), these codes *are* applied.

Subheadings Applicable to M5: Codes

Set M5: codes are are used with subheadings Ml through M4 from 1970 forward; in addition, M531, M532 and M533 are used with subheading M0 from 1963.

M51: Fused ring heterocyclic systems	
M510	None
M511	One
M512	Two
M513	Three or more

M52: Mononuclear heterocycles	
M520	None
M521	One
M522	Two
M523	Three or more

M53: Carbocyclic systems with at least one aromatic ring		
M530	None	
M531	One	(1963)
M532	Two	(1963)
M533	Three or more	(1963)

Note The ring systems referred to in Subset M53: codes are those coded in Sets Gl: through G4:. The term ***aromatic*** is defined on page 94.

M54: Alicyclic Systems	
M540	None
M541	One
M542	Two
M543	Three or more

M6: MISCELLANEOUS STRUCTURE CODES

Note Like Basic Group codes, the codes in Set M6: allow general classification of the compound being searched.

Subheadings Applicable to M6: Codes

M610 is applicable to subheadings M2, M3, and M4 from 1970 and to subheading Ml from 1981. M620 to M650 are applicable to subheading M0 from 1963 and to subheadings Ml to M4 from 1970.

M610	Hydrocarbon

M620	Saturated aliphatic compound

Note M620 is used for all compounds which contain no rings and no C to C unsaturation, but which may contain heteroatoms.

M630	Metal or amine salt of an organic acid

Note The complete structure of the acid is always indexed, but the metal or amine is only indexed if significant to the invention. (The amine, if indexed, would normally be indexed on a separate record, unless the structure is multi ligand complex.

M640	Inorganic acid salt of an organic base

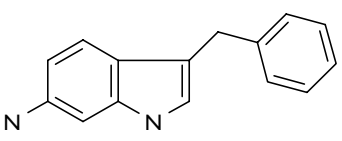
Note The complete structure of the base is always indexed, but the acid is only indexed if significant to the invention.

M650	Organic acid salt of an organic base

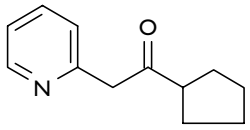
Note The complete structure of the base is always indexed, but the acid is only indexed if significant to the invention. If both organic acid AND organic base are indexed, they are indexed in separate records, unless the structure is a multi ligand complex.

Examples Using M4:, M5:, and M6: Codes

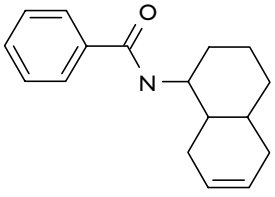
Note Codes in Sets M4:, M5:, and M6: are closely related. Because of the priority system utilized in Set M4:, searching codes from Sets M4: and M6: may make it unnecessary to search Set M5: codes. Redundant codes in the examples below are underlined, and can be omitted from the search strategy.

a 

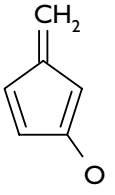
M412, M511, M520, M531, M540
(redundant because the code G100 always implies M540)

b 

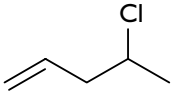
M413, M510, M521, M530, M541
(M413 implies mo fixed heterocyclic)

c 

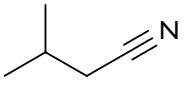
M414, M510, M520, M531, M541

d 

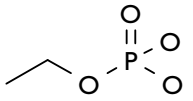
M415, M510, M520, M530, M541

e 

M416

f 

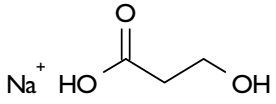
M416, M620

g 

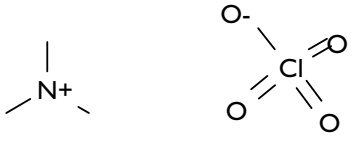
M411, M510, M520, M530, M540, M620

h 

M416, M610, M620

i 

M416, M620, M630
(Na not coded unless important to invention)

j 

M416, M620, M640
(Perchlorate not coded unless important to the invention)

k 

M416, M620, M650 (If acetate and amine are both important to the invention, include M630, M771)

Note See page 181 for examples using M411.

M7: PATENT TYPE

Notes on M7: Codes

1 Set M7: codes are used to specify the role of a compound in a patent, the involvement of a compound in a complex, or the use of a compound as a major feature of an invention.

Searching For Patents On Detection, Analysis, Or Diagnostics

Three different **code subfields** may be used when indexing a detection, analysis, or diagnostic patent: one for the material used in the detection, analysis or diagnosis (Material A); another for the material being detected, analysed or used diagnostically (Material B); and a third for the medium used (Material C). A suggested formula for searching on all three types of materials is shown below; however, all three materials are not always structure-coded. This depends on the information given in the patent specification.

Material A (M781 OR M782); P831 for disease diagnosis, P832 for other types of diagnosis; and Q505 for the reagent itself

Material B M750

Material C M760

If a major feature of a detection, analysis or diagnostic patent is a **process** or **apparatus**, the code N102 is also assigned (searched as N100 before 1981). If Material B is coded, then N102 (searched as N100 before 1981) is also coded. Prior to 1981, either P831 or P832 was added to all three code subfields, not only to the code subfield for Material A. This is useful for searching because the pre-1981 codes used for M750 and M760 before 1981 were much broader than M750 and M760.

Note Code subfields are discussed in the Introduction to the manual, beginning on page 28.

Searching For Patents On Removal Of A Compound

As in the previous section, three different **code subfields** may be used when indexing a patent on removal of a compound or compounds: one for the reagent used (Material X); a second for the material removed (Material Y); and a third for the material treated (Material Z). A suggested formula for searching on all three types of materials is shown below; as before, all three materials are not always structure-coded.

Material X (M781 OR M782); Q508 OR Q500

Material Y M750; and M720 if the material is recovered

Material Z M720

If a major feature of a removal patent is a process or apparatus, a code from Subset N16: is added to all three code subfields, along with the code Q431 (which is searched in subheadings M1 and M2 from 1981, subheadings M3 and M4 from 1970). A code from Subset N16: and the code Q431 are always coded when the material removed (Material Y) or the material treated (Material Z) is coded. If the patent involves water treatment apart from sewage and industrial waste, then the code Q231 (subheadings M1 and M2 from 1981, subheadings M3 and M4 from 1970) is added to all three code subfields. If the patent involves pollution control, then one or more of the codes Q436 through Q439 are added to all three code subfields.

Note Code subfields are discussed in the Introduction to the manual, beginning on page 28.

Subheadings Applicable to M7: Codes

Set M7: codes are applicable to subheadings Ml to M4 from 1970 and to subheading M5 from 1981. An exception is M770 applicable to subheadings Ml to M5 from 1981). In addition, M771 and M772 are applicable to subheading M0 from 1963.

M710: through M760: - role of compound in patent	
M710	New compound or intermediate
M720	Known compound is produced
M730	Compound is used in a synthetic process
M740	Apparatus
M750	Compound is detected or removed
M760	Compound is a medium for a substance being analysed, detected or used in a diagnosed

Notes on Codes M710 Through M760

1. New compounds, intermediates, and microorganisms are normally coded M710 only, and not also elsewhere in M7:. However, from 1979 forward, compounds disclosed but not claimed to be new receive the code M710 as well as other applicable codes from Set M7:. Novel ingredients of compositions receive the code M710 as well as other applicable codes in Subset M78:.
2. Because of the different rules and practices of patent-granting and patent-receiving organizations, it is suggested that M710 not be relied upon for comprehensive chemical novelty searches.
3. The code M720 includes purification, extraction, or separation of the compound(s) coded, including separation of optical isomers.
4. The code M730 was rarely used prior to 1981 because before that time starting materials, catalysts and reagents were only coded if novel. From 1981 forward, the codes M730 and Q421 have been used to identify the separate code subfields for catalysts. (The complete coding of catalysts is discussed in Part N:, beginning on page 189.) M730 is also used when the end-product of a reaction is unknown.
5. The use of M740 implies that the inventive feature of a patent is an apparatus, i.e. the chemical features are less important. The code M740 has not been assigned to compounds used in "other" processes since 1981; however, such compounds continue to be coded in Subset M78:.
6. The code M750, which has only been consistently assigned since 1979, indicates that a compound is detected, analysed, used in diagnostics, or is removed during a purification process. M750 is also assigned to patents involving monoclonal antibodies detected in diagnostic tests.
7. The code M760, which has only been consistently applied since 1979, is assigned to detection, analysis, or diagnostic media. M760 is also used for solvents when they are an important feature of an invention.

M77: Complexes		
M770	Molecular complex	(1963)
M771	Acid addition salt	
M772	Onium salt	

Notes on M77: Codes

1. M770 is assigned when two or more uncharged compounds are complexed, e.g. clathrates and other interstitial compounds, charge transfer complexes, etc. Examples of compounds coded M770 are $BF_3.Et_2O$ and picric acid.mesitylene. M770 is not used to indicate water of hydration, metal coordinate bonding to an uncharged particle, or pi-bonding to an uncharged particle; therefore, M770 is not used to code metal carbonyls, ferrocene, or other similar compounds.
2. M771 and M772 indicate that the structure codes assigned refer to one ion of a salt for which both ions are coded, i.e. a salt in which both anion and cation are important, limiting factors of the invention. Specifically excluded are simple metal salts. If the

salt is a new compound with known component ions, then M710 is also assigned. If the salt is known, then M782 is assigned instead of M710. Normally both ions of compounds that are indexed M771 or M772 are organic. Only one of the two codes, M771 or M772, is assigned to all code subfields related the salt. In addition, the code M630 is assigned to the code subfield of the acid, and M650 is assigned to the code subfield of the base.

M78 Use of a compound

M781	Use of one compound
M782	Use of two or more compounds

Notes on M78: Codes

1. M78: codes are only assigned to ingredients of compositions that are significant to the patent. Because it may be difficult for the user to know for certain how many ingredients have been coded, it is usually best to search (M781 OR M782) for the use of a single compound.

M8: STEREOCHEMISTRY

M800 Stereochemistry is an important feature of the invention (1981)

M9: CONTROL CODES

The codes in Set M9: allow the user to select a subset of the online database in which to conduct a search.

M901 - M903 time range codes

M900	Pre-1970 records
M901	1970 - 1971 records
M902	1972 - 1981/Derwent Week 26 records
M903	1981/Derwent Week 27 to the present records

Note Time-range codes are discussed in the section "Time Ranging", beginning on page 22.

M904 Both Markush DARC indexing and Chemical Codes have been assigned to a structure

Note M904 can be used to remove records that are retrievable by Markush DARC software.

M910 Some or all structure codes in this subfield were generated from a Derwent Registry Number

Note The Derwent Registry is discussed in the section "Searching With Other Kinds of Codes", beginning on page 26.

M911 Patent has a very wide disclosure

Note M911 indicates that a patent has such a wide disclosure that it is often retrieved by searches to which it has no relevance. This code is not systematically assigned, but can be negated to eliminate such references from searches which retrieve large numbers of answers. However, this could lose relevant references, so use with care.

10 Part N: Chemical Reactions, Processes and Apparatus

10 Part N: Chemical reactions, processes, and apparatus

Main Headings

N0: General Chemical Reaction Descriptors (Pre-1981)
N1: Reaction Processes
N2: Bonds Broken
N3: Bonds Formed
N4: Auxiliary Compound Descriptors
N5: Reaction Conditions

Introduction

Part N: codes refer to chemical reactions, processes, and apparatuses claimed to be novel. Topics include: reaction processes; bonds broken and formed; auxiliary compounds such as solvents and catalysts; and reaction conditions.

Notes on Part N: Codes

1. To be coded in Part N:, a chemical process must be claimed and must be an important feature of the invention.

2. Set N0: codes are very general descriptors for chemical processes; they have not been used since 1981, when the more specific codes in Sets N1: through N5: were introduced.

3. Most Part N: codes are only searchable from 1981 forward. The codes N100, N130, N160, N180, N310, N330, and N340 are **precursor codes** (discussed on page 9), each of which were replaced by more specific codes marked with a # on the coding sheet. Precursor codes introduced in 1963, i.e. those in Set N1:, are 'OR'ed with a more specific code in Search Statement 1. Precursor codes introduced in 1970, i.e. those in Set N3:, are 'OR'ed with a more specific code in the search statement line used for codes introduced in 1970 (usually Search Statement 2).

4. The scope of chemical reactions covered by Part N: codes expanded twice between 1979 and 1981. Prior to 1979, Part N: codes were only used for processes involving the production of known compounds, i.e. they were not assigned to processes in which **all** products were **disclosed or claimed** to be novel. [From 1979, patents where the process was claimed were given reaction codes even if all compounds were disclosed (but not all claimed) as new. Therefore, the only occasions where N: codes were *not* applied to process patents after 1979 was when all the compounds were **claimed** to be new]. From 1981 forward, Part N: codes have been additionally assigned to all processes having novel features, e.g. those which give unexpected results compared to prior art reactions, even if all of the end products are new. The novelty of a process is determined from the wording of the whole patent, not from its presence or absence in

the claims. Also from 1981 forward, Part N: codes have been assigned to production of new natural materials of unknown structure.

5. All relevant process codes are assigned to a reaction. In a **multi-stage reaction**, all steps in the process which lead to isolated products are coded, **except** tautomerism of starting materials, intermediates, or end-products.

6. The compounds involved in reactions receiving Part N: codes are themselves structure-coded in Parts A: through M: as follows:

   - **Starting materials** - if stated to be novel
   - **Catalysts** and other **auxiliary compounds** - if stated to be novel
   - **Intermediate products** - if stated to be novel
   - **Final products** - whether novel or not

7. Derwent Registry numbers have been assigned to the following types of compounds since 1981, provided the compounds are specifically mentioned in the claims or examples, and provided that Registry numbers exist for the compounds:

   - all **starting materials** and isolatable **intermediates**
   - **auxiliaries** that provide at least one atom to an organic product
   - **auxiliaries** that are a limiting feature of the invention
   - all **catalysts**

   Derwent Registry numbers are discussed in the section "Searching With Other Kinds of Codes", beginning on page 26.

8. The following element symbols are used in Part N: code definitions:

W	=	Any heteroatom
W'	=	Heteroatom other than N, O, S, P, F, Cl, Br, or I
X	=	F, Cl, Br, or I

   W', which is denoted by "rem" on the coding sheet, includes the atoms **Se** or **Te**. Thus, unlike the convention in most code Parts, compounds with Se or Te do not receive all codes defined for S in Part N:.

9. Dyeing processes are coded in Set W5:. Galenical processes (coating, bioanalysis, diagnostics, etc.) are coded in Sets R5: and R6:.

Starting Materials and Auxiliaries

Definitions

A ***starting material for an organic product*** is an organic compound which contributes at least one C atom to the product. ***A starting material for an inorganic product*** is a compound which contributes at least one atom to the product. Thus, not all reactants are considered starting materials. The term ***product*** is used in this chapter to indicate the target compound(s) of the reaction, not all reaction products. Any substances present in a reaction which are not starting materials or products are called ***auxiliaries***, e.g. reagents, solvents, catalysts, acid scavengers, etc. Set N4: codes are used to describe auxiliaries.

The structures of starting materials and auxiliaries stated to be novel have been fragmention-coded in Parts A: through M: since 1963, excluding catalysts – see below.

Catalysts

From 1970 forward, all catalysts stated to be novel or produced by a novel process have been assigned applicable fragmentation codes from Parts A: through M:. From 1977 forward, all catalysts mentioned in Derwent's Documentation Abstract of a patent are searchable using CPI Section N Manual Codes. (Manual Codes are discussed on page 26.)

From 1981 forward, all catalysts that appear in the claims or examples of patent documents have been assigned to a separate code subfield coded with the code Q421 ("Catalyst"). ("Code subfields" are discussed on page 28.) For catalysts claimed to be novel or produced by a novel process, the applicable fragmentation codes from Parts A: through M: are applied as normal. If the catalyst is not stated to be novel and is not produced by a novel process, then only the following Chemical Codes are assigned:

- applicable codes from Parts A: and B: (**except** Sets B7: and B8:)
- applicable codes from Sets C0: through C5: for inorganic H, C, N, O, S, and halogens, **if these elements are significant** to the catalyst. For example, cobalt oxide in a catalyst receives the code C108 ("Oxygen present"); but if a catalyst can contain any cobalt salt, anions mentioned with cobalt in the examples would not be coded.
- applicable codes from Sets C6: and C7:, and from Subset C81:
- the Basic Group code from Set M4:
- Q421 and M730 ("Known compound used in a reaction")
- applicable codes from Set C8: ("Elements absent"), but only if **every element** in the catalyst is assigned a fragmentation code

---

Exception Alumina or aluminosilicate catalysts (including zeolites) are not coded in Parts A: through M:, because they have specific CPI Section N Manual Codes.

Selecting Codes for Bonds Broken and Bonds Formed in Sets N2: and N3:

Introduction

Sets N2: ("Bonds broken") and N3: ("Bonds formed") were derived from the Chemical Reaction Documentation Service (CRDS) coding system, with several minor differences. Codes from Sets N2: and N3: are only assigned to reactions whose products are considered significant enough to the patent to be structure-coded. Codes for the bonds broken and formed to produce the product are listed in the same code subfield as the structural indexing of the product.

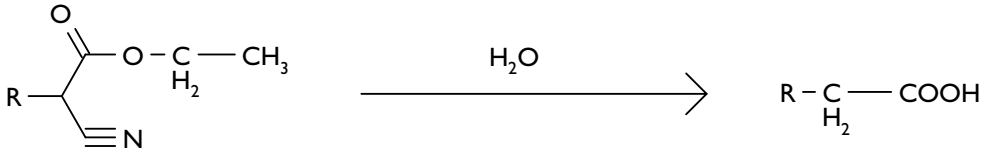
Bonds broken are bonds present in a **starting material** that are changed to yield new bonds in the **product**. NOT coded as broken bonds are the following:

- bonds broken in auxiliary compounds
- bonds broken in the decomposition of the eliminated parts of a molecule
- bonds that are temporarily modified but which are present in the final product

Bonds formed are bonds that are present in a **product** that were not present in the starting materials or auxiliaries. All bonds formed in the product are coded, irrespective of the reactants that contributed the bonded atoms.

The bonds broken and formed in **multistage reactions** are coded for each reaction step between one isolable structure and the next in the reaction sequence. These are overcoded in one Code Subfield (i.e. the product) unless the intermediate(s) is (are) also coded. Fermentations are excluded from this rule, unless the intermediate stages are specified. The entire Krebs cycle, for example, is not coded for the sugar --> alcohol process. **Intermediate products** are only structure-coded in Parts A: through M: if they are stated to be novel.

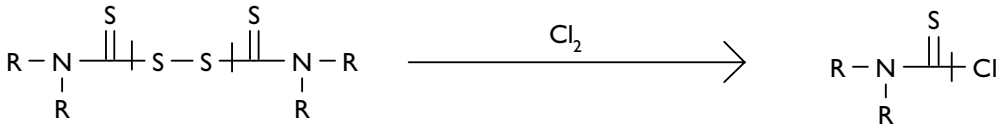
When different functional groups in the same starting material react to give a single product, the two reactions are overcoded. For example, the reaction shown below is coded "C-C-heteroatom broken" (N211), "O=C-O broken" (N241), "C≡N broken" (N235), "O=C-O formed" (N341) and "C-H formed" (N321).



Examples of Indexing Bonds Broken and Formed

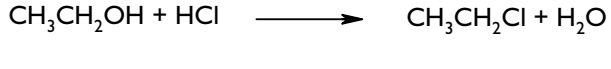
In the examples below, ≠ or ⫨ in the reactants indicates bonds broken, and ≠ or ⫨ in the products indicates bonds formed.

Example 1



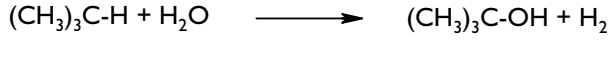
The above reaction is indexed N252 ("C-S broken"), N209 ("Poly"), and N322 ("C-Cl formed"). It is not indexed N283 ("S-S broken") because the S-S group is completely eliminated.

Example 2



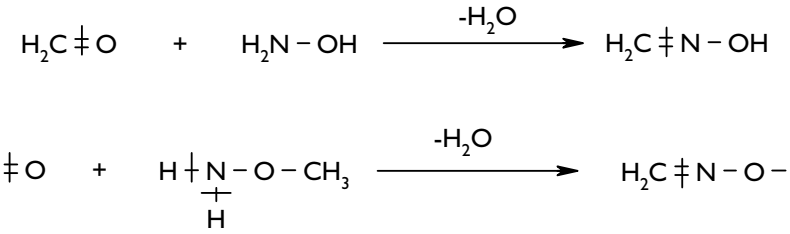
In the reaction shown above, A is the **starting material** and B is the **product**. In this reaction, a C-O bond is broken (N242) and a C-halogen bond is formed (N322). The H-Cl bond broken and the H-OH bond formed are not indexed because they are auxiliary compounds.

Example 3



In the reaction shown above, a C-H bond is broken (N224) and a C-O bond is formed (N342). The H-O bond present in compound B was already present in H_2O, so N362 ("H-O formed") is not indexed.

Example 4

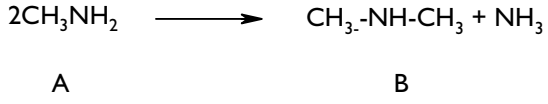


Both of the above reactions are indexed "C=O broken" (N243) and "C=N formed" (N334). The bottom reaction is additionally indexed "N-H broken" (N261) and "Poly, bonds broken" (N209), but the top reaction does not receive these codes. The difference is that NH_2OCH_3 is a **starting material** (it supplies a C atom to the product), and NH_2OH is only an **auxiliary**.

Example 5

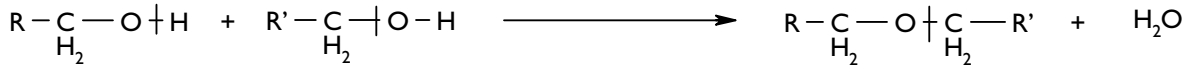
This reaction is indexed "O=C-O broken" (N241) "O-H broken" (N262) and "O=C-O formed" (N341). These codes take the reaction mechanism into account, i.e. cleavage of the C-O bond of the acid and formation of the C-O bond of the ester. The apparent replacement of H by Ph is not indexed. If the reaction mechanism were not clear, all possibilities would be indexed.

Example 6



The reaction shown above is indexed "C-N broken" (N233), "H-N broken" (N261) and "C-N formed" (N333).

Example 7

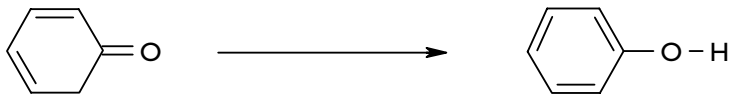


This reaction is indexed "C-O broken" (N242), "H-O broken" (N262) and "C-O formed" (N342). No **Poly** code is assigned.

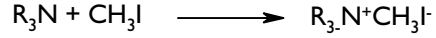
Structural Conventions

The following conventions are assigned when selecting codes for bonds broken and formed in reactions:

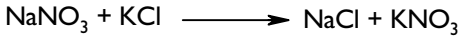
a The starting materials and products are considered in the form in which they would be indexed in Parts A: through M:. For example, only the preferred tautomeric form is indexed; alternative tautomers of the starting materials and products are not considered. {Rules for coding tautomers are found on page 106.}



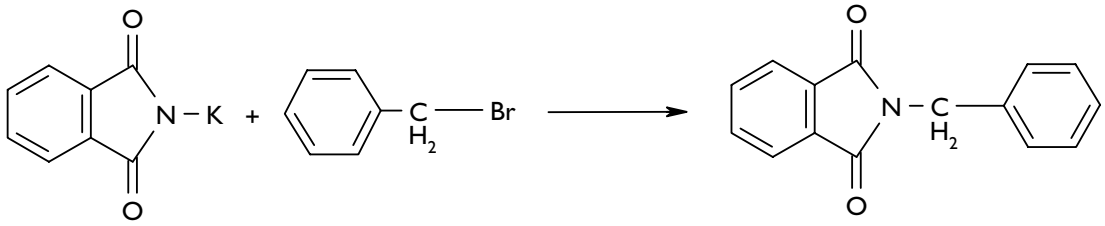
b Ionic bonds are ignored in both organic and inorganic reactions. For example, the reaction shown below is indexed "C-halogen broken" (N225) and "C-N formed" (N330), but not "N-halogen formed" (N366).



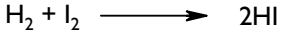
The following reaction only involves the breaking and forming of ionic bonds, so no codes in N2: or N3: are assigned.



c Acid addition and base salts of organic bases and acids are treated as the parent base or acid, unless salt formation or decomposition is the only novel feature of the invention. The reaction shown below is indexed "C-halogen broken" (N225), **"H-N broken"** (N261), and "C-N formed" (N333), but not "N-W' broken" (N275). Thus, unless the N-K bond is a significant feature of the patent, the N-H form is indexed instead.

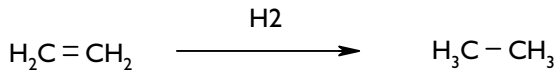


d Bonds broken and formed in elements are ignored. In the reaction shown below, only "H-halogen formed" (N363) is indexed.

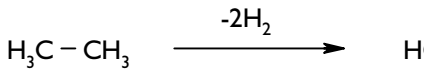


Multiple Bonds Broken or Formed

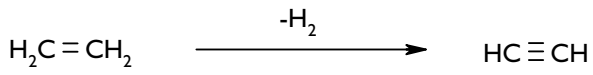
When a single bond is converted to a double or triple bond, breaking the single bond is not indexed, but formation of the multiple bond is indexed. When a double or triple bond is converted to a single bond, breaking the multiple bond is indexed, but formation of the single bond is not indexed. However, for conversion of a double bond to a triple bond or vice versa, both the bond broken and the bond formed are indexed.



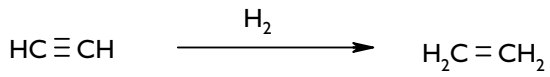
N213, N321, N309, but not N311



N314, N224, N209, but not N212



N213, N221, N209, N314



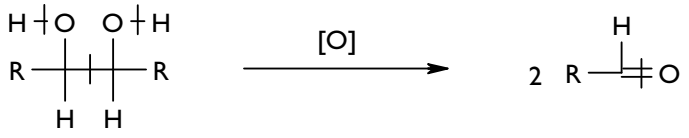
N214, N312, N321, N309

When no specific code exists for breaking or forming a particular multiple bond, the code for breaking or forming the corresponding single bond is indexed in combination with one of the generic codes N201 (“Double bond opened”), N202 (“Triple bond opened”), N301 (“Double bond formed”) or N302 (“Triple bond formed”). For example, C-S bond formation is indexed N352 (“C-S bond formed”), whereas C=S bond formation is indexed N352 and N301.

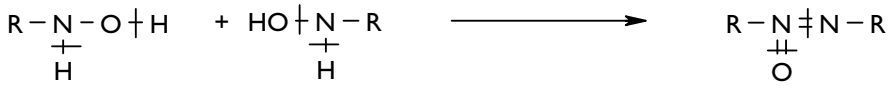
Poly Codes

Either N209 ("Poly, bonds broken") or N309 ("Poly, bonds formed") is used to indicate that one of the codes in Sets N2: or N3:, respectively, is applicable more than once to a particular reaction. These codes are assigned from the point of view of one unit of product. Thus, when reactions result in two molecules of a product, the bonds formed are not indexed **Poly**. However, if two molecules of a starting material react together, and if a particular N2: code is applicable more than once to the reaction, "Poly, bonds broken" (N209) is indexed.

The following reaction is indexed "C-C broken" (N211), "H-O broken" (N262), "Poly, bonds broken" (N209), and "C = O formed" (N343), but not "Poly, bonds formed" (N309).



The reaction shown below is indexed "H-N broken" (N261), "Poly, bonds broken" (N209), "H-O broken" (N262), "N-O broken" (N273), "N=N formed" (N372) and "N=O formed" (N373 for N-O formed and N301 for a double bond formed).



Isotopes

Different isotopes of an element receive the same code as the element itself would receive, and when an element is replaced by one of its isotopes, the bond broken is considered the same as the bond formed. For example, when a C-H bond is broken and a C-deuterium bond is formed, the reaction is indexed N224 ("C-H bond broken") and N321 ("C-H bond formed").

Subheadings Applicable to Part N: Codes

Codes are applicable from 1963 to Subheadings M0-M4 unless otherwise specified. From 1981 codes apply to all subheadings, unless otherwise specified.

N0: GENERAL CHEMICAL REACTION DESCRIPTORS (PRE-1981)

Note Codes in Set N0: were only assigned prior to 1981/Week 26. In 1981/Week 27, the more specific codes in Sets N1: through N5: were introduced.

N000	Process not specified elsewhere in Set N0: or Set N1:
N010	Addition reaction (non-steroids only)
N020	Elimination reaction (non-steroids only)

N03: Steroid fermentation processes

N030	Non-specific	(M5 only)
N031	Oxidation	(M5 only)
N032	Reduction	(M5 only)
N033	Processes producing unsaturation	(M5 only)

N040: Hydrolysis; alcoholysis (non-steroids only) (1970 M3: M4 only)

N050: Oxidation process (non-steroids only)

N060: Reduction process (non-steroids only)

Notes on N0: Codes

1. N000 covers processes that cannot be more specifically indexed in Subsets N01: through N06:, or in Subsets N13: through N18:. For example, N000 covers all condensation reactions.
2. The code N010 covers reactions in which two or more ingredients combine to form a single product. N010 is not assigned to oxidation (N050) or reduction (N060) reactions.
3. N020 covers reactions in which a single compound is split into more than one product, including pyrolysis.
4. Alcoholysis is only indexed N040 in pre-1981 general chemical patents (Chemdoc - Section E). Reactions indexed N010 are not also indexed N040.
5. Ozonation and dehydrogenation are indexed as oxidation processes (N050). Hydrogenation is indexed as a reduction process (N060).

N1: REACTION PROCESSES

N10: Apparatus; physical process

N100	Apparatus; physical process(Pre-1981 precursor code for Subset N10:)	
N101	Container; packaging	(1963 M0-M2, M5; 1981 M3, M4)
N102	Detection, analytical, or diagnostic (apparatus or process)	(1981)
N103	Means of administration or application of a compound or composition	(1981)
N104	Means for producing a compound or composition	(1981)
N105	Other apparatus or physical process (including reaction apparatus)	(1981)

Notes on N10: Codes

1. The codes in Subset N10: are assigned to apparatuses and physical processes that are not more specifically covered elsewhere in Set N1:.

2. Galenical (formulation) apparatuses and processes are indexed in Set R5:, not in Subset N10:.

3. N102 is usually searched in combination with other codes. The indexing of detection, analytical, and diagnostic patents is discussed on page 185.

N11: Ring expansion/contraction, chain expansion/contraction

N111	Ring expansion	(1970 M3, M4; 1981 otherwise)
N112	Ring contraction	(1970 M3, M4; 1981 otherwise)
N113	Carbon chain expansion	(1970 M3, M4; 1981 otherwise)
N114	Carbon chain contraction	(1970 M3, M4; 1981 otherwise)

Notes on N11: Codes

1. Either N111 or N112 is assigned if: (a) the total number of atoms in a ring is changed; (b) at least two ring members from the previous ring remain in the ring; and (c) there is no change in the overall number of rings in the ring system, e.g. production of naphthalene from benzene would NOT be coded N111. The codes N111 and N112 are always used in conjunction with a "ring opening" code (N204 or N205) and a "ring formation by cyclisation" code (N303, N304, N305, or N306).

2. To be indexed N113 or N114, the number of acyclic C atoms in a straight or branched carbon chain must be altered, which excludes carbon chain rearrangements where the number of carbons in the chain does not change. The carbon chain must be present in the starting material and product, which means that alkylation of an aromatic ring would not be considered a chain expansion.

Note "Carbon chain" in codes N113 and N114 refers to any straight or branched chain of C atoms, including C atoms that are multiply-bonded to heteroatoms. This is different from the definition of carbon chain in Part M:, which excludes C atoms multiply bonded to heteroatoms.

N120: Electrolytic process; electric discharge (1970 M3, M4; 1980 otherwise)

Note From 1981 forward, electric discharges have been indexed N142 in addition to N120.

N13: Fermentation; culture

N130	Fermentation; culture (Pre-1981 precursor code for Subset N13:)	(1981)
N131	Cultivation of bacteria	(1981)
N132	Cultivation of fungi (or algae), including yeast	(1981)
N133	Cultivation of virus	(1981)
N134	Enzymatic process	(1981)
N135	Genetic engineering	(1981)
N136	Tissue or cell culture from plant or animal	(1981)
N137	Other (including whole plant culture)	(1981)

Notes on N13: Codes

1. Subset N13: includes attenuation and use of enzymes, but not culture resolution.

2. Reactions indexed in Subset N03: are not also indexed N130.

N14: Irradiation

N141	Irradiation by light
N142	Other irradiation

Notes on N14: Codes

1. From 1981 forward, electric discharge processes have been indexed N142 in addition to N120.
2. Infrared radiation and ultraviolet radiation are indexed N141.
3. Examples of irradiation processes indexed N142 include: gamma-rays, X-rays, magnetisation, demagnetisation, sonic and ultrasonic waves.

N15: Polymer processes

N151	Dimerisation
N152	Oligomerisation; polymerisation; telomerisation
N153	Polymer modification; depolymerisation (1970 M3, M4; 1981 otherwise)

N16: Purification; extraction; separation

N160	Purification; extraction; separation (Pre-1981 precursor code for Subset N16:)
N161	Extraction from natural materials (1963 M5; 1981 otherwise)
N162	Isomer and isotope separation (1981)
N163	Purification by chemical reaction (and optional regeneration) (1981)
N164	Other purification, extraction, separation processes (1981)

Notes on N16: Codes

1. N160 does not cover steroids. Prior to 1981, N161 only covered steroids.
2. Separation of stereoisomers is indexed in Subset N18:, not in Subset N16:.

N17: Racemisation; rearrangemenet

N171	Racemisation
N172	Rearrangement; isomerisation
N173	Conversion of one solid form to another by physical change (1981)

Notes on N17: Codes

1. N172 covers a change in position of C atoms and/or heteroatoms between two atoms or groups of atoms present in the starting material. Excluded from N172 are H-migrations, D↔L and cis↔trans conversions, and the migration of a group between two radicals originating from different starting materials.
2. Prior to 1981, N172 was only assigned when the starting material and product contained exactly the same atoms, e.g. isomerisation. Therefore N172 would have been used for production of t-butyl alcohol from iso-butanol before and after 1981. However, dehydration and rearrangement of 2 molecules of isobutanol to produce di-t-butyl ether would not have been coded N172 prior to 1981.

N18: Resolution; production of stereoisomers

N180	Resolution; production of stereoisomers (Pre-1981 precursor code for Subset N18:) (1981)
N181	D → L, L → D, cis → trans, trans → cis (1981)
N182	DL separation by fermentation or biochemical process (1981)
N183	DL separation by other methods (1981)
N184	Production of one stereoisomer by fermentation or biochemical process (1981)
N185	Production of one stereoisomer by other methods (1981)

Notes on N18: Codes

1. N182 and N184 include enzymatic processes.
2. N184 and N185 include stereoselective processes, but exclude reactions in which the configuration is maintained without racemisation.

N2: BONDS BROKEN

Note The indexing of multiple bonds broken is discussed on page 196. If no specific code exists for the multiple bond broken, use the code for breaking the corresponding single bond and either N201 or N202.

N20: General descriptors

N200	No bonds broken	(1981)
N201	Double bond broken or modified	(1981)
N202	Triple bond broken or modified	(1981)
N203	Reaction at a benzene ring	(1981)
N204	Carbocyclic ring opened	(1981)
N205	Heterocyclic ring opened	(1981)
N209	Poly, for all Subsets in Set N2:	(1981)

Notes on N20: Codes

1. N200 is used for reactions at an atom that do not involve broken bonds at that atom or the adjacent atom(s), and for reactions in which only the stereochemistry has changed.
2. Either N201 or N202 is assigned when no specific code exists for the particular multiple bond that is broken. The code for breaking the corresponding single bond is also assigned.
3. N203 indicates that an exocyclic bond from a benzene C atom is modified, but the benzene ring itself is retained. The benzene ring, which may be isolated or fused with another ring, must be present in both the starting material and the product. Specifically excluded from the scope of N203 are ring closure reactions in which the modified bond forms part of a ring condensed to the original benzene ring, as in the reaction shown below.

4. A ring opening that is followed by cyclisation to form a new ring of the same size is assigned a "ring opening" code (N204 or N205) and a "ring formation by cyclisation" code (N303, N304, N305, or N306).
5. N209 indicates that a code in Set N2:, other than N201 or N202, is applicable to a reaction more than once. N209 does not indicate that more than one bond in a structure is broken in a reaction.

N21: C-C, C=C, or C≡C bond broken

N211	C-|-C-W, C-|-C=W, or C-|-C≡W	(1981)
N212	Other C-C broken	(1981)
N213	C=C	(1981)
N214	C≡C	(1981)
N215	A single C atom is eliminated	(1981)

Notes on N21: Codes

1. In reactions indexed N211, one or both C atoms can be substituted by one or more heteroatoms.
2. N215 indicates that the breaking of a single or multiple bond between C atoms leads to the elimination of a single C atom, which may rebond with the original C-structure via a heteroatom, e.g. $CH_3C(=O)$-|-$CH_3 \rightarrow$ $CH_3C(=O)$-|-O-|-CH_3

N22: C-H or C-X bond broken		
N221	C=C-|-H, C≡C-|-H	(1981)
N222	C=C-C-|-H, C≡C-C-|-H	(1981)
N223	W=C-C-|-H, W≡C-C-|-H, W-C(W)-C-|-H	(1981)
N224	Other C-H broken	(1981)
N225	C-X broken	(1981)

N23: C-N, C=N, or C≡N broken		
N231	O=C-|-N	(1981)
N232	N=C-|-N, N-C=|=N	(1981)
N233	Other C-N broken	(1981)
N234	Other C=N broken	(1981)
N235	C≡N broken	(1981)

Note N209 ("Poly, bonds broken") is not assigned when N=C and N-C are each broken once.

N24: C-O or C=O broken		
N241	O=C-|-O, O-C=|=O	(1981)
N242	Other C-O broken	(1981)
N243	Other C=O broken	(1981)

Note N209 ("Poly, bonds broken") is not assigned when O=C and O-C are each broken once.

Note for N25: → N28 codes:
W' is a heteroatom other than N, O, S, P, F, Cl, Br, or I. X is F, Cl, Br, or I.

N25: Other bonds to C broken		
N251	C-P	(1981)
N252	C-S	(1981)
N253	C-W'	(1981)

N26: Bonds to H or X broken (Other than C-H or C-X)		
N261	H-N	(1981)
N262	H-O	(1981)
N263	H-P, H-S, H-W', H-X	(1981)
N264	X-P	(1981)
N265	X-S	(1981)
N266	X-N, X-O, X-W', X-X	(1981)

N27: Other bonds to N broken		
N271	N-N	(1981)
N272	N=N, N≡N	(1981)
N273	N-O	(1981)
N274	N-S	(1981)
N275	N-P, N-W'	(1981)

Notes on N27: Codes

1. To indicate N≡N broken, search N272 and N202 ("Triple bond broken"). N=N broken is indicated by N272 alone. Other multiple bonds are specified using N201, N202 together with the N27: code as normal.

N28: Other bonds broken		
N281	O-P	(1981)
N282	O-O, O-S, O-W'	(1981)
N283	P-P, P-S, P-W', S-S, S-W'	(1981)
N284	W'-W'	(1981)

N3: BONDS FORMED

Note The indexing of multiple bonds formed is discussed on page 196. If no specific code exists for the multiple bond formed, use the code for forming the corresponding single bond in combination with either N301 or N302.

N30: General descriptors

N301	Double bond formed	(1981)
N302	Triple bond formed	(1981)
N303	Carbocyclic ring formed by cyclisation – not fused or spiro	(1970 M3, M4; 1981 otherwise)
N304	Carbocyclic ring formed by cyclisation – fused or spiro	(1970 M3, M4; (1981 otherwise)
N305	Heterocyclic ring formed by cyclisation – not fused or spiro	(1970 M3, M4; (1981 otherwise)
H306	Heterocyclic ring formed by cyclisation – fused or spiro	(1970 M3, M4; (1981 otherwise)
N309	Poly, for Set N3:	(1981)

Notes on N30: Codes

1. Either N301 or N302 is assigned when no specific code exists in Set N3: for the particular multiple bonds that are formed. The code for forming the corresponding single bond is also assigned.
2. N303 through N306 are not used for cleavages of polycyclic structures, e.g. norbornane → methylcyclohexane.
3. Prior to 1981, cyclisation reactions were additionally indexed N111 or N112 (ring expansion or contraction).
4. A ring opening that is followed by cyclisation to form a new ring of the same size is assigned a "ring opening" code (N204 or N205) and a "ring formation by cyclisation" code (N303, N304, N305, or N306).
5. N309 indicates that one or more codes in Set N3: is applicable to a reaction more than once. N309 does not indicate that more than one bond in a structure has been formed in a reaction.

N31: C-C, C=C, or C≡C bond formed

N310	C-C, C=C, or C≡C bond formed (1970-1981 M3, M4 only)	
N311	C-C formed, with neither C bonded to a heteroatom	(1981)
N312	C=C formed from C-C or C≡C	(1981)
N313	Other C=C formed	(1981)
N314	C≡C formed	(1981)
N315	C-|-C-W	(1981)
N316	C-|-C=W, C-|-C≡W	(1981)
N317	W-C-|-C-W, W=C-|-C=W, W-C-|-C≡W	(1981)
N318	(Ring C)-C	(1981)
N319	Single C atom becomes bonded to C	(1981)

Notes on N31: Codes

1. N315 indicates that **one** of the two C atoms is bonded to one heteroatom via a single bond. If the bond formed is C=C-W or C≡C-W, search N315 and either N312, N313, or N314 as applicable.
2. N316 indicates that **one** of the two C atoms is multiply bonded to one or more heteroatoms. If the bond formed is C=C=W, search N316 and either N312 or N313 as applicable.

3. N317 indicates that **both** C atoms are **singly or multiply** bonded to heteroatoms. If the bond formed is W-C=C-W or W-C≡C-W, search N317 and either N312, N313, or N314 as applicable.
4. N318 indicates that one of the C atoms in the bond formed is acyclic and the other is part of a ring that was present in a starting material.
5. N319 indicates that a single C atom is condensed, which bears only H and/or seteroatoms. It is not used for rearrangments within the same C-C structure.

N32: C-H or C-X formed		
N321	C-H	(1981)
N322	C-X	(1970 M3, M4; 1981 otherwise)

N33: C-N, C=N, or C≡N formed		
N330	C-N, C=N, or C≡N formed (1970-1981, M3, M4 only)	
N331	O=C-|-N	(1981)
N332	N=C-|-N, N-C=|=N	(1981)
N333	Other C-N formed	(1981)
N334	Other C=N formed	(1981)
N335	C≡N formed	(1981)

Note N309 ("Poly, bonds formed") is not assigned when N=C and N-C are each formed once.

N34: C-O OR C=O FORMED		
N340	C-O or C=O formed (1970-1981 M3, M4 only)	
N341	O=C-|-O, O-C=|=O	(1980)
N342	Other C-O formed	(1980)
N343	Other C=O formed	(1980)

Note N309 ("Poly, bonds formed") is not assigned when O=C and O-C are each formed once.

Note for codes N35: → N38:
W' is a heteroatom other than N, O, S, P, F, Cl, Br, or I. X is F, Cl, Br, or I.

N35: Other bonds to C formed		
N351	C-P	(1980)
N352	C-S	(1970 M3, M4; 1981 otherwise)
N353	C-W'	(1981)

N36: Bonds to H or X formed (Other than C-H or C-X)		
N361	H-N	(1981)
N362	H-O	(1981)
N363	H-P, H-S, H-W', H-X	(1981)
N364	X-P	(1981)
N365	X-S	(1981)
N366	X-N, N-O, X-W', X-X	(1981)

N37: Other bonds to N formed		
N371	N-N	(1981)
N372	N=N, N≡N	(1981)
N373	N-O	(1981)
N374	N-S	(1981)
N375	N-P, N-W'	(1981)

Note To indicate N≡N formed, code N372 and N302 ("Triple bond formed"). N=N formed is only indexed N372. Other multiple bonds are specified using N301, N302 together with the N37: codes as normal.

N38: Other bonds formed		
N381	O-P	(1981)
N382	O-O, O-S, O-W'	(1981)
N383	P-P, P-S, P-W', S-S, S-W'	(1981)
N384	W'-W'	(1981)

N4: AUXILIARY COMPOUND DESCRIPTORS

N41: Redox reagents

N411	Oxidising agent	(1981)
N412	Reducing agent	(1981)

Note These two codes only refer to O_2, H_2, compounds giving off O or H, and compounds taking up O or H.

N42: Type of medium

N421	Acid medium (pH < 6)	(1981)
N422	Basic medium (pH > 7)	(1981)
N423	Fluidised bed	(1981)
N424	Melt	(1981)
N425	pH specific (e.g. buffer)	(1981)
N426	Water free (anhydrous)	(1981)

N430: Condensing agent (1981)

Notes on N430

1. N430 covers compounds that react with one or more organic molecules, causing the elimination of a simple molecule such as H_2O, HCl.
2. Condensing agents do not contribute atoms to the product. In particular, N430 is not assigned to Mg in Grignard reactions.

N44: Catalytic reactions

N441	Heterogeneous	(1981)
N442	Homogeneous	(1981)

Note Undefined catalytic reactions are indexed both N441 and N442.

N450: Inhibitors (1981)

Note N450 covers antioxidants, polymerisation inhibitors, catalytic poisons, etc.

N460: Biological auxiliaries (excluding reactions indexed in Subset N13:) (1981)

Note N460 covers enzymes, extracts with enzymic activity, other proteins, polysaccharides, microorganisms, etc.

N470: Polymers as auxiliaries (excluding auxiliaries indexed N460) (1981)

N480: Reactions in the absence of oxygen (1981)

Note N480 is not used when the gas reacts with the reaction medium, e.g. CO_2 in neutralisations.

N5: REACTION CONDITIONS

N51: Temperature (T)

N511	$T < 10\ ^oC$	(1981)
N512	$10\ ^oC \leq T \leq 30\ ^oC$	(1981)
N513	$30\ ^oC < T \leq 200\ ^oC$	(1981)
N514	$200\ ^oC < T \leq 500\ ^oC$	(1981)
N515	$T > 500\ ^oC$	(1981)

N52: Pressure (P)

N520	P < 1 atm
N521	1 atm ≤ P < 2 atm
N522	2 atm ≤ P ≤ 20 atm
N523	20 atm < P ≤ 100 atm
N524	100 atm < P ≤ 1000 atm
N525	P > 1000 atm

Notes on N52: Codes

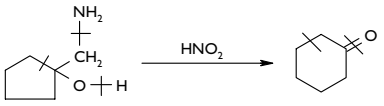
1. N520 includes reactions under vacuum, but not vacuum distillation after a reaction.
2. N521 is only indexed if the process disclosed is more commonly carried out at reduced or elevated pressures, i.e. if the pressures is the novel feature of the invention.

Examples of Part N: Codes

The following examples are arranged in order of the particular concepts (underlined codes) being illustrated.

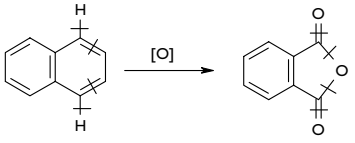
N11: Ring/chain expansion/contraction

a



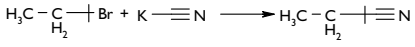
1981 to present: <u>N111</u>, N204, N211, N233, N262, N303, N311, N343, N430
Pre-1981: N000, N110, N303, N310

b



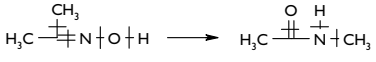
1981 to present: <u>N112</u>, N204, N209, N213, N221, N306, N309, N341, N411
Pre-1981: N050, N340

c



1981 to present: <u>N113</u>, N225, N316, N319
Pre-1981: N000, N110, N310

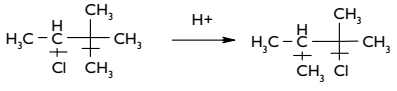
d



1981 to present: <u>N114</u>, N172, N211, N215,N234, N262, N273, N333, N343, N361
Pre-1981: N110, N172, N330, N340

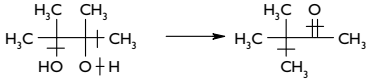
N172: Rearrangement

e



1981 to present: <u>N172</u>, N212, N225, N311, N322, N421
Pre-1981: N172, N310, N322

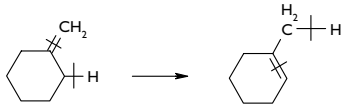
f



1981 to present: <u>N172</u>, N211, N242, N262, N311, N343
Pre-1981: N020, N310, N340 but **not** N172

Note Prior to 1981, N172 was only used if product and starting material contained exactly the same atoms

g

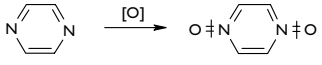


1981 to present: N213, N222, N312, N321
Pre-1981: <u>N172</u>, N310

Note H migrations have not been indexed N172 since 1981.

N20: Generic descriptors for bonds broken

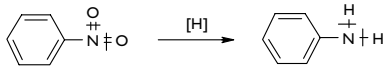
h



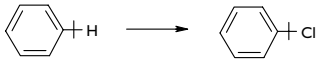
1981 to present: <u>N200</u>, N209, N301, N309, N373, N411
Pre-1981: N050

Note N209 is applied as no bonds were broken at two reaction sites

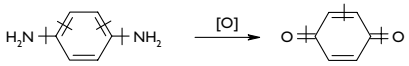
i



1981 to present: <u>N201</u>, N273, N209, N361, N309, N412
Pre-1981: N060

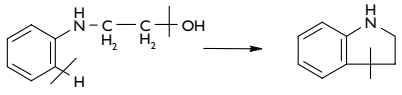
j 

1981 to present: N203, N221, N322
Pre-1981: N000, N322

k 

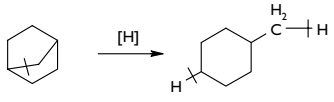
1981 to present: N213, N209, N233, N312, N343, N309, N411 **not** N203
Pre-1981: N050, N310, N340

Note Although the ring in the product is indexed as an aromatic ring, the reaction is not indexed N203 because the position of the double bond has changed.

l 

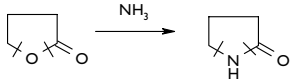
1981 to present: N221, N242, N306, N311 **not** N203
Pre-1981: N020, N306, N310

Note This reaction is not indexed N203 because the modified bond forms part of a ring condensed to the original benzene ring.

m 

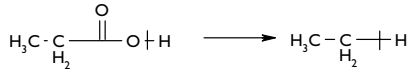
1981 to present: N204, N212, N321, N309, N412 **not** N203 or N209
Pre-1981: N060

Note This reaction is not indexed N113 because the number of rings in the ring system has changed. The reaction is not indexed N209 because there is only one ring cleavage.

n 

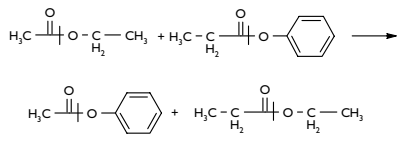
1981 to present: N205, N241, N242, N305, N331, N333
Pre-1981: N000, N305, N330

N215: A single C atom is eliminated

o 

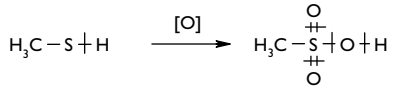
1981 to present: N215, N114, N211, N262, N321
Pre-1981: N020, N110

N241, N341: O=C-O broken, formed

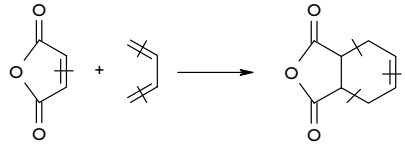
p 

1981 to present: N241, N209, N341, N309
Pre-1981: N000

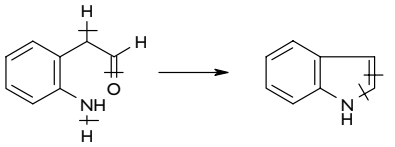
N30: Generic descriptors for bonds formed

q 

1981 to present: N301, N263, N362, N382, N309, N411
Pre-1981: N050

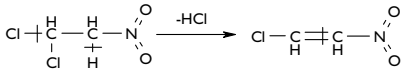
r 

1981 to present: N304, N213, N209, N311, N309, N312
Pre-1981: N010, N304, N310

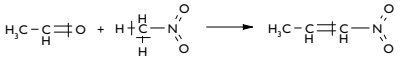
s 

1981 to present: N306, N223, N243, N261, N312, N333
Pre-1981: N020, N310, N330, N306

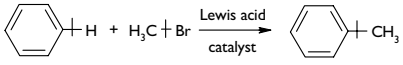
N31: C-C, C=C, C=C bond formed

t 

1981 to present: N312, N317, N223, N225 (+N430 if a specific condensing agent is used)
Pre-1981: N020, N310

u 

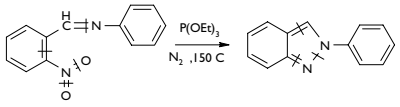
1981 to present: *N313, N315, N319*, N113, N224, N209, N243
Pre-1981: N000, N113, N310

v 

1981 to present: N311, N318, N319, N203, N221, N225, N442
Pre-1981: N000, N310

Note When the code N319 is assigned, either N113 or N318 is also assigned.

N4: Auxiliary Compound Descriptors

w 

1981 to present: N412, N430, N480, N513, N213, N234, N273, N201, N209, N306, N312, N315, N334, N371
Pre-1981: N000, N060, N306, N310, N330

11 Activities, Properties, Uses

11 Activities, properties, uses

Introduction

The codes in Parts P: and Q: are used to search for activities, properties, and uses of chemical compounds and compositions.

Notes on Codes in Parts P: and Q:

1. The code for a particular activity, property, or use is most easily located by consulting the *Chemical Code Dictionary*.

2. Codes in Parts P: and Q: are only assigned when a patent specifically states that a compound or composition has a particular activity, property or use - i.e. they are not inferred. For example, a process for the production of acetylsalicylic acid (aspirin) would not be indexed as "analgesic" (P411) unless this activity is specifically stated. **Exception:** Process patents whose end products are certain polymer condensants and monomers are indexed Q110 ("Polymer condensant, monomer"), whether this use is stated in the patent or not. (See the note after Q110 on page 220 for a list of these monomers and condensants).

3. When a patent discloses a generic use rather than a specific use, e.g. describing a known "insecticide" (P341) as simply a "pesticide" (P340), only the generic use is indexed. Thus, for comprehensive retrieval it is sometimes necessary to OR a specific use of a compound or composition with its more generic use, e.g. (P340 OR P341).

4. When a patent discloses new uses of a compound that has known uses, only the new uses are indexed. For example, if a known deodorant (Q604) is found to have flame-proofing activity (Q621), only the flame-proofing activity is indexed.

5. Because of the coding limitations discussed in Notes 2, 3, and 4 above, Chemical Codes in Parts P: and Q: do not give retrieval of all references to compounds having a certain activity, property, or use. The codes in Parts P: and Q: are most useful for finding out if a known compound has a particular activity, property, or use, and for limiting the number of hits from structure searches. **Manual Codes** (discussed on page 26) can often be used to search more specifically and comprehensively for activities, properties, and uses than the corresponding Chemical Codes, particularly in the area of general chemicals (Chemdoc - Section E).

6. Part P: codes are not assigned to inactive intermediates of active products. For example, a novel intermediate of an insect sterilant is not assigned the code P342 ("Insect sterilant"). Part Q: codes, on the other hand, often indicate a broad field of use rather than a specific activity, and are thus assigned in some cases to intermediates.

7. Detection, analysis, and removal patents may be assigned one or more of the following codes: P831, P832, Q231, a code from Subset Q43:, Q505, or Q508. The indexing of detection, analysis, diagnostic, and removal patents is discussed in detail on page 185.

8. Part P: codes are primarily applicable to CPI Sections B (Farmdoc) and C (Agdoc), and Part Q: codes to Section E (Chemdoc). However, from 1981 forward, Part Q: codes may also be assigned to patents that are only classified in Sections B and C, and Part P: codes may be assigned to patents that are only classified in Section E.

Part P: Pharmaceutical (Farmdoc) and Agricultural (Agdoc) Descriptors For Activities, Properties, and Uses

Main Headings

P0: General Activity Descriptors
P1: Plant Growth Media, Regulant
P2: Antimicrobial; Disinfection
P3: Other Biocides
P4: Analgesic; Anaesthetic; Antiinflammatory; Immunological; CNS
P5: Autonomic Nervous System; Cardiovascular System
P6: Metabolic
P7: Anabolic; Gastrointestinal System
P8: Blood; Respiratory System; Miscellaneous Descriptors
P9: Skin; Dental; Ear, Nose, and Throat; Mouth; Eye; Hair Treatment

Subheadings Applicable to Part P:

Unless otherwise specified, codes are valid from 1993 for subs M0, M1, M2, and M6 and from 1981 for subs M3 and M4

P0: GENERAL ACTIVITY DESCRIPTORS

P001	Human or veterinary activity	(1965)
P002	Agricultural activity	(1965)
P003	Contact	(1965)
P004	Systemic or stomach	(1965)

Notes on P0: Codes

1. Codes in Set P0: are only used for agricultural (Section C - Agdoc) patents, in combination with P200, P210, P220, P241, P242, P330, P331, P340, P341, or P361.
2. The code P003 is not used for human or veterinary activities, and is therefore not used together with the code P001.

P1: PLANT GROWTH MEDIA, REGULANT

P100: Plant or soil use of steroids (M5: 1965-1981)

P11: Fertiliser

P111	Trace element(s)	(1965)
P112	Other inorganic fertiliser	(1965)
P113	Other organic fertiliser	(1965)

P12: Soil treatment; growth media

P121	Erosion inhibition	(1965; M3:, M4:, 1970)
P122	Frost protection	(1965)
P123	Moisture conservation; mulch	(1965)
P124	Nitrification inhibitor	(1965)
P125	Soil fumigant	(1965)
P126	Soil texture improvement	(1965)
P127	Synthetic growth media	(1965)

P13: Plant growth regulants

P130	Plant growth regulants - general	(1965)
P131	Defoliant; dessicant	(1965)
P132	Fruit drop or set; blossom stimulant or retardant	(1965)
P133	Mutation inducer	(1965)
P134	Phytohormone, growth stimulant	(1965)
P135	Plant growth inhibitor	(1965)
P136	Rooting compound, rhizogene	(1965)

P14: Herbicide

P140	Herbicide - general	(1965)
P141	Aquatic	(1981)
P142	Pre-emergence	(1981)
P143	Post-emergence	(1981)
P144	Selective	(1965)
P145	Total	(1965)

Notes on P14: Codes

1. Prior to 1981, the codes P141, P142, and P143 were not in use. Thus aquatic, pre-emergence, and post-emergence herbicides were assigned either P140, P144, or P145 as applicable.
2. Antimoss and antilichen substances are assigned the code P243.

P2: ANTIMICROBIAL; DISINFECTION

P200: Generic disclosure of antimicrobial or disinfection substance (1981; M3:, M4:, M5: 1970)

Note: Prior to 1981 other subs used P220 for general antimicrobials

P21: Antiviral

P210	Antiviral - general	
P211	Interferon-inducer	(1981)

Note Prior to 1981, P210 was assigned to all antivirals. From 1981 forward, P210 has been assigned to all antivirals that are not interferon-inducers.

P220: Antibacterial

P23: Antimycobacterial; antivenereal

P231	Antileprotic
P232	Antitubercular
P233	Antivenereal

P24: Antifungal; antialgal; antilichen, antimoss

P241	Antifungal
P242	Antialgal
P243	Antilichen, antimoss

P3: OTHER BIOCIDES

Note Biocides indexed as general chemicals (Chemdoc - Section E) before 1981 and steroidal biocides are searched using the code P200, rather than a code from Set P3:.

P300: Antifouling

P31: Antiprotozoal

P310	Antiprotozoal – general or unspecified
P311	Amoebicide
P312	Antimalarial
P313	Coccidiostat
P314	Histomonacide; trichomonicide
P315	Trypanocide
P316	Other specified antiprotozoal (e.g. babesicide, leishmaniacide)

P32: Anthelmintic

P320	Anthelmintic – general or unspecified
P321	Schistosomicide
P322	Fasciolicide, distomicide

Note Nematocides are assigned the code P345, not a code in Subset P32:.

P33: Antiparasitic

P330	Antiparasitic – general or unspecified
P331	Acaricide (agricultural)
P332	Tickicide (veterinary)

Note Codes in Subset P33: are only used for substances that can not be classified in previous Subsets.

P34: Other pesticide

P340	Other pesticide – general or unspecified	(1965)
P341	Insecticide	(1965)
P342	Insect sterilant	(1965)
P343	Molluscicide; crustacicide	(1965)
P344	Rodenticide; piscicide; avicide	(1965)
P345	Nematocide (soil)	(1965)

P35: Lure; bait

P350	Lure; bait	(1965)
P351	Pheromone	(1981)

Note Prior to 1981, pheromones were assigned the code P350.

P36: Repellant

P361	Insect repellant	
P362	Rodent, fish, or bird repellant	(1965)

P4: ANALGESIC; ANAESTHETIC; ANTIINFLAMMATORY; IMMUNOLOGICAL; CNS

P41: Analgesic; local anaesthetic

P411	Analgesic (1963 including M5)
P412	Local anaesthetic (1963 including M5)

P42: Antiinflammatory

P420	Antiinflammatory – general or unspecified (1963 including M5)
P421	Antiarthritic
P422	Antipyretic
P423	Antirheumatic

P43: Immunological

P430	Immunological – general (pre-1981 steroids) (M5 1963-1981)
P431	Antiallergic
P432	Antihistamine
P433	Autoimmune disease treatment; immunosuppressant, complement inhibitor
P434	Immunostimulant

P44: CNS depressant, tranquilisers

P440	CNS depressant – general (1963 including M5)
P441	Anaesthetic – general (local anaesthetic is searched as P412) (1963 including M5)
P442	Anticonvulsant (1963 including M5)
P443	Antiemetic
P444	Antiparkinson
P445	Hypnotic
P446	Neuroleptic; antipsychotic
P447	Sedative
P448	Tranquiliser; anxiolytic

P45: CNS stimulant

P450	CNS stimulant – general (1963 including M5)
P451	Antidepressant
P452	Analeptic
P453	Convulsant (1963 including M5)

P5: AUTONOMIC NERVOUS SYSTEM; CARDIOVASCULAR SYSTEM

P51: Autonomic nervous system

P510	Autonomic nervous system – general (1963 including M5)
P511	Alpha sympathetic blocker
P512	Beta sympathetic blocker
P513	Alpha sympathetic stimulant
P514	Beta sympathetic stimulant
P515	Anticholinergic
P516	Cholinergic
P517	Muscle relaxant
P518	Mydriatic, myopic
P519	Uterus active

P52: Cardiovascular system

P520	Circulatory active – general
P521	Antiarrythmic
P522	Cardioactive – general
P523	Coronary dilator
P524	Ganglion blocker
P525	Hypertensive
P526	Hypotensive
P527	Vasoconstrictor
P528	Vasodilator
P529	Veins (varicose); haemorrhoid treatment

Notes on P52: Codes

1. P520 or P522 are only assigned to a compound when there is not a more specific term for the disclosed activity.
2. General circulatory disease treatment is also indexed in Subset P51:.

P6: METABOLIC

P61: Antihormone, Antimetabolite

P610	Antimetabolite – general	(1963 including M5)
P611	Antiandrogenic	(1981)
P612	Antioestrogenic	(1981)
P613	Antiprogestational	(1981)
P614	Other antihormone	(1981)
P615	Antivitamin	(1981)
P616	Non-polymeric enzyme inhibitor of known structure	(1981)
P617	Other antimetabolite	

Note Compounds assigned the code P616, including synthetic enzyme inhibitors, additionally receive a code from Subset V81: to indicate the type of enzyme. Enzyme inhibitors of polymeric or unknown structure are indexed in Subset V80:.

P62: Hormone active

P620	Hormone active – general (Pre-1981 precursor code for P621, P622, P623, P625 including M5)	
P621	Androgenic	(1981)
P622	Oestrogenic	(1981)
P623	Progestational	(1981)
P624	Thyroid active	(1963 including M5)
P625	Hormone active - other	(1981)

Note Contraceptives are assigned additional codes from Subset P84:.

P63: Cancer treatment; cytostat

P631	Inhibitor of cell division	(1963 including M5)
P632	Leukaemia treatment	(1963 including M5)
P633	Other antitumour agent	(1963 including M5)
P634	Tumour inducing agent	(1963 including M5)

P64: Antidote

P640	Antidote – general (Pre-1981 precursor code for Subset P64: including M5)	
P641	Alcoholism treatment	(1981)
P642	Antismoking	(1981)
P643	Antidote for heavy metal poisoning	(1981)
P644	Antidote for poisoning by pesticide or herbicide	(1981)
P645	Antidote protecting plants against poisons	(1981)
P646	Other antidote	(1981)

P7: ANABOLIC; KIDNEY OR LIVER ACTIVE; GASTROINTESTINAL

P71: Anabolic

P710	Anabolic – general (Pre-1981 precursor code for Subset P71: including M5)	
P711	Appetite stimulant	(1981)
P712	Gastrointestinal flora	(1981)
P713	Animal growth promoter	(1981)
P714	Other anabolic agent	(1981)

Notes on P71: Codes

1. Food, animal feeds, and additives that cannot be indexed in Subset P71: are indexed in Subsets Q21: and Q22:.
2. Prior to 1981/Week 27, animal growth promoters were assigned the code R002. Human growth promoters were assigned the code P710.

P72: Kidney, liver active

P720	Steroid with kidney or liver activity (Pre-1981 only)	
P721	Choleretic	
P722	Diuretic	
P723	Other renal activity	

P73: Gastrointestinal

P730	Steroid with gastrointestinal activity (Pre-1981 only)
P731	Anorectic (excluding compounds coded P732 or P733)
P732	Gastric secretion inhibitor (excluding compounds coded P733)
P733	H2 secretion inhibitor
P734	Antacid
P735	Antidiarrhoeal
P736	Emetic
P737	Laxative
P738	Ulcer treatment

Note Gastrointestinal flora is coded P712, not in Subset P73:.

P8: BLOOD; RESPIRATORY SYSTEM; MISCELLANEOUS DESCRIPTORS

P81: Blood

P810	Steroid with blood activity (Pre-1981 only)
P811	Iron deficient anaemia
P812	Other anaemia
P813	Anticoagulant
P814	Antilipaemic, anticholesterol
P815	Coagulant
P816	Hypoglycaemic
P817	Blood or plasma substitute

P82: Respiratory system

P820	Respiratory activity, general or unspecified; Pre-1981 Steroid with respiratory activity (Pre-1981 only)
P821	Antitussive
P822	Bronchodilator
P823	Expectorant

Note Pre-1981 steroids were coded P820. Therefore search P820 OR specific P821-3 codes to retrieve steroids from 1963.

P83: Diagnostic; analytical material

P831	Diagnostic; analysis of biological material
P832	Other analytical material (1963 M5; M3, M4 1970)

Notes on P83: Codes

1. The indexing of detection, analytical, and diagnostic patents is discussed on page 181.
2. Prior to 1981, P831 is only searchable for pharmaceutical (Farmdoc - Section B) and agricultural (Agdoc - Section C) patents.

P84: Contraceptive

P840	Contraceptive general (Pre-1981 only) (1963-1981 including M5)
P841	Abrotifacient, anti-implantation (1981)
P842	Antiovulatory (1981)
P843	Other contraceptive, including devices (1981)

Notes on P84: Codes

1. Prior to 1981, P610 was assigned instead of P841 or P842.
2. Insect sterilants are coded P342, not in Subset P84:.

P850: X-ray contrast agent

Note Prior to 1981, P850 is only searchable for pharmaceutical (Farmdoc - Section B) and agricultural (Agdoc - Section C) patents including steroids.

P86: Synergist

P861	Pharmaceutical synergist (Farmdoc - Section B) (1963 including M5)
P862	Steroidal agricultural synergist (Agdoc - Section C) (1963 including M5)
P863	General chemical steroidal synergist (Chemdoc - Section E) (1970 M3: M4)

P9: SKIN; DENTAL; EAR, NOSE, AND THROAT; MOUTH; EYE; HAIR TREATMENT

Notes on P9: Codes

1. Cosmetics, perfumes, and non-medical hair treatments are coded in Subset Q25:.
2. Skin protection is coded in Subset Q26:.

P91: Dental agents

P910	Dental agents - general (Pre-1981 precursor code for Subset P91:) (1963-1981 M0-M6)
P911	Toothpaste or tooth powder (1981)
P912	Anticaries agent, antiplaque agent (1981)
P913	Dentistry, including dentures and dental fillings (1981)

P92: Ear, nose, and throat; mouth; eye

P920	Ear, nose, and throat; mouth; eye - general (Pre-1981 precursor code for Subset P92:) (1963-1981 M0-M6)
P921	Ear preparation (1981)
P922	Eye preparation, including composition for contact lenses (1981)
P923	Mouth (including gums) preparation (1981)
P924	Nasal preparation (1981)

P930: Treatment of hair or scalp disease; chemical hair restoration (1963 including M5)

Note Non-medical hair treatment and non-chemical hair restoration are coded Q252.

P94: Skin treatment

P940	Skin treatment - general (Pre-1981 precursor code for Subset P94:) (1963-1981 including M5)
P941	Burn treatment (1981)
P942	Other wound treatment (1981)
P943	Other skin treatment (1981)

Notes on P94: Codes

1. Bandages, dressings, surgical sponges, and tampons are coded in Subset R04:.
2. Skin protection, including sunscreen agents, are coded in Subset Q26:.

Part Q: Other Chemical Descriptors

Main Headings

Q0: General Use Descriptors
Ql: Descriptors for compounds indexed in CPI Section A
Q2: Descriptors for compounds indexed in CPI Section D
Q3: Descriptors for compounds indexed in CPI Sections E, F, and G
Q4: Descriptors for compounds indexed in CPI Sections H through M
Q5: Chemical Properties
Q6: Other Properties and Uses

Introduction

The codes in Part Q: are used to describe the activities, properties, and uses of a broad range of compounds and compositions, other than those coded in Part P: with pharmaceutical (Farmdoc – Section B) or agricultural (Agdoc – Section C) applications.

Derwent's CPI Sections, which form the basis of Set divisions in Part Q:, are listed below.

CPI Section	Topic
A	Polymers (Plasdoc)
B	Pharmaceutical (Farmdoc)
C	Agricultural (Agdoc)
D	Food, Cosmetics, Detergents
E	General Chemical (Chemdoc)
F	Textiles, Paper, Cellulose
G	Printing, Coating, Photography
H	Petroleum, Fuel
J	Chemical Engineering
K	Nucleonics, Explosives
L	Refractories; Cement; Electroorganic or Electroinorganic Substance
M	Metallurgy

Subheadings Applicable to Part Q: Codes

Codes in Part Q are generally applicable to M3 from 1970 and subs M1 to M6 from 1981. Some codes are applicable to M4 from 1970. Unless otherwise specified, subheadings are only valid from 1981.

Q0: GENERAL USE DESCRIPTORS

Note Set Q0: codes are only assigned when the use is not covered by a more specific code in Sets Q1: through Q6:.

Q010	Electrical; instrumentation	(M3, M4: 1970
Q020	Engineering; transport	(M3, M4: 1970
Q030	Household; personal	(M3, M4: 1970)

Q1: POLYMER DESCRIPTORS (CPI SECTION A)

Q110 Polymer condensant or monomer

Note In most cases, activities, properties, and uses are only coded if specifically disclosed in a patent document. However, when a patent concerned with the production of one or more of the following polymer condensants and monomers are coded in Section E, the code Q110 is assigned, even if use as a polymer monomer or condensant is not stated.

Acrolein
Acrylic acid
Acrylonitrile
Adipic acid
Bisphenol A
Butadiene
Caprolactam
Chloroprene
Diethylterephthalate
Dimethylterephthalate
Ethylacrylate
Ethylene
Ethylmethacrylate
Formaldehyde
Hexamethylenediamine (1,6-hexanediamine)
Isobutene (isobutylene)
Isoprene
Maleic anhydride
Melamine
Methacrylic acid
Methyl(meth)acrylate
Methylstyrene (vinyl toluene)
2,6-Naphthalenedicarboxylic acid (2,6-naphthalic acid)
Phenol
Phthalic anhydride
Propylene
Sebacic acid
Styrene
Terephthalic acid
Tetrafluoroethylene
Urea
Vinyl acetate
Vinyl chloride
Vinylidene chloride

Q12: Polymer production (excluding polymer condensants and monomers)

Q120	Polymer production - general	(M3: 1970)
Q121	Polymerisation catalyst or initiator	(M3: 1970)
Q122	Polymerisation inhibitor	(M3: 1970)

Note Polymer production compounds are only coded Q120 if neither Q121 nor Q122 is applicable.

Q13: Polymer additives

Q130	Polymer additive - general	(M3: 1970)
Q131	Blowing agent	(M3: 1970)
Q132	Crosslinking agent or accelerator	(M3: 1970)
Q133	Lubricant or mould release	(M3: 1970)
Q134	Depolymerising agent	(M3: 1970)

Notes on Q13: Codes

1. Additives are coded Q130 if no other code in Subset Q13: is applicable. Additives coded Q130 are further defined elsewhere in Parts P: and Q: if possible. For example, flame retardants are coded Q130 and Q621.
2. Additives to the surface of a polymer are coded in Subset Q13:, not Q140. However, prior to 1981, such compounds were coded inconsistently, so for that time period Q140 should be 'OR'ed with the applicable code from Subset Q13:.

Q140 Polymer processing, testing, etc. (M3: 1970)

Notes on Q140

1. Polymer modification and surface treatment of polymers are coded Q140, unless compounds coded Q132 or Q133 are used in the process.
2. Additives to the surface of a polymer are coded in Subset Q13:, not Q140. However, prior to 1981, such compounds were coded inconsistently, so for that time period Q140 should be 'OR'ed with the applicable code from Subset Q13:.
3. Depolymerisation processes and subsequent monomer recovery are coded Q110 and N153, not Q140.

Q2: FOOD, COSMETICS, AND DETERGENT DESCRIPTORS (CPI SECTION D)

Q21: Human or animal foods

Q210	Human or animal food (Pre-1981 precursor code for Subset Q21:) (M3, M4: 1970-1981)	
Q211	Human food	(1981)
Q212	Food for other mammals	(1981)
Q213	Avian food	(1981)
Q214	Food for other animals	(1981)

Notes on Q21: Codes

1. Food additives are coded in Subset Q22:, not in Subset Q21:.
2. Prior to 1981, human food compositions were coded Q210 or P710, dependent on CPI section. (Search Q210/M3 or P710/M0, M1, M2, M5).
3. Prior to 1981, animal food compositions were coded Q210 or R002, dependent on CPI section. (Search Q210 or R002/M0, M1, M2, M5).
4. Human or animal foods containing appetite stimulants, gastrointestinal flora, or other anabolic drugs are coded in Subset P71:, not in Subset Q21:.

Q22: Food additives; food preservation

Q220	Food additive or food preservation not covered elsewhere in Subset Q22: (M3, M4: 1970)	
Q221	Condiment; flavour; aroma-producer	(M3: 1970)
Q222	Sweetener	(M3: 1970)
Q223	Vitamin	(M3: 1970)
Q224	Food preservation, other than a food preservative	(M3: 1970)
Q225	Food preservative	(M3: 1970)

Notes on Q22: Codes

1. Prior to 1981, human food additives were additionally coded P710.
2. Prior to 1981, animal food additives were additionally coded R002.
3. Human or animal food additives such as appetite stimulants, gastrointestinal flora, or other anabolic drugs are coded in Subset P71:, not in Subset Q22:.
4. Q224 and Q225 became separate codes in 1981. Prior to 1981, either code retrieves references to both food preservatives and other types of food preservation.

5. Animal food additives are additionally coded Q212, Q213, or Q214 as applicable.
6. Prior to 1981, dyes within food compositions were not coded Q221, Q222, Q223, Q224, or Q225.

Q23: Water treatment; brewing; microbiology

Q231	Water treatment other than sewage, wastewater, and pollution control	(M3, M4: 1970)
Q232	Brewing; fermentation producing ethanol	(M3, M4: 1970)
Q233	Microbiology, laboratory culture, mutant formation	(M3, M4: 1970)

Notes on Q23: Codes

1. Sewage, wastewater, and pollution control are coded Q437, not Q231.
2. Fermentations that produce ethanol are assigned the code Q232 and a code from Subset N13:. Fermentations that produce compounds other than ethanol are coded in Subset N13:, not in Subset Q23:.

Q24: Sugar, starch; leather; animal fur, skin, or hide; tobacco

Q241	Sugar and starch industry	(M3: 1970)
Q242	Leather; animal fur, skin, or hide	(M3: 1970)
Q243	Tobacco; tobacco requisite	(M3: 1970)

Q25: Cosmetics; toiletries; hair treatment; perfume

Q250	Cosmetics; toiletries; hair treatment (Pre-1981 precursor code for Q251, Q252, and Q254)	(M3, M4: 1970-1981)
Q251	Antiperspirant; human deodorant	(M3, M4: 1981)
Q252	Non-medical hair treatment; non-chemical hair restoration	(M3, M4: 1963)
Q253	Perfume	(M3, M4: 1963)
Q254	Depilatory; other cosmetics	(M3, M4: 1963)

Q26: Disinfection or sterilisation; skin protection

Q261	Disinfection or sterilisation	(M3, M4: 1970)
Q262	Skin protection other than sunscreening agent	(M3, M4: 1970)
Q263	Sunscreening agent for skin	(M3, M4: 1970)

Notes on Q26: Codes

1. Bandages, dressings, surgical sponges, and tampons are coded in Subset R04:.
2. For disinfection and sterilisation in Sections B or C prior to 1981, P200 and/or P220 were used.
3. Prior to 1981, pharmaceutical and veterinary skin treatments are coded P940, not in Subset Q26:.

Q27: Oil; fat; wax; soap; detergent

Q271	Oil; fat; animal or vegetable wax	(M3, M4: 1970
Q272	Soap composition	(M3, M4: 1970)
Q273	Non-soap detergent composition	(M3, M4: 1970)

Q3: DESCRIPTORS FOR COMPOUNDS INDEXED IN CPI SECTIONS E, F, AND G

Q31: Dye-related descriptors

Q311	Colour coupler	
Q312	Organic dye (including mordant)	(M3, M4: 1970)
Q313	Dye assistant (excluding compounds coded Q314)	(M3: 1970)
Q314	Resist, discharge agent	(M3: 1970)
Q315	Other dye additive	(M3: 1970)
Q316	Dye intermediate	(1981)
Q317	Oxidative dye precursor	(M3, M4: 1970)
Q318	Non-oxidative dye precursor	(M3, M4: 1970)

Notes on Q31: Codes

1. Dyes are also coded in Part W:. For example, couplers for azo dyes and for oxidation bases receive a code from Subset W12: in addition to the code Q311.
2. Examples of compounds coded Q318 include: leuco dyes from triaryl methane; vat, sulphur, and phthalein dyes; cryochromic, halochromic, photochromic, piezochromic, and thermochromic compounds; and diazonium compounds.

Q32: Textiles; paper; cellulose; wood (CPI Section F)

Q321	Colouring of textiles	(M3: 1970)
Q322	Other chemical treatment of textiles	(M3: 1970)
Q323	Textile production or treatment (excluding processes coded Q321 or Q322)	(M3: 1970)
Q324	Paper; cellulose; wood	(M3: 1970)

Q33: Printing; Coating (CPI Section G)

Q331	Adhesive	(M3, M4: 1970)
Q332	Ink; paint; polish	(M3, M4: 1970)
Q333	Inorganic pigment	(M3: 1970)
Q334	Liquid crystal - electrical	(M3, M4: 1970)
Q335	Other liquid crystal	(M3, M4: 1970)
Q336	Cleaning composition (excluding compositions coded in Subset Q27:)	(1981)
Q337	Other composition	(M3, M4: 1970)
Q338	Printing materials and processes (excluding those coded Q339 or in Subset Q34:)	(M3, M4: 1970)
Q339	Thermography, magnetography, etc.	(M3, M4: 1970)

Notes on Q33: Codes

1. Codes in Subset Q33: do not cover compounds coded in Subset Q34: ("Photography, electrophotography, electrography"), but they do cover non-optical image production.
2. Steroidal liquid crystals prior to 1981 were only indexed using subs M5 codes.

Q34: Photography; electrophotography; electrography

Q341	Antihalation; fog inhibitor	(M3, M4: 1970)
Q342	Developer; fixer	(M3, M4: 1970)
Q343	Photosensitiser	(M3, M4: 1970)
Q344	Photosensitive compound	(1981)
Q345	Photosensitive composition used in photography	(M3, M4: 1970)

Q346	Photosensitive composition used in electrography or electrophotography	(M3, M4: 1970)
Q347	Other photographic material	(M3, M4: 1970)
Q348	Other electrographic or electrophotographic material	(M3, M4: 1970)
Q349	Photographic production of printing surface	(M3, M4: 1970)

Notes on Q34: Codes

1. Prior to 1981, photosensitive compounds were coded Q345 or Q346 as applicable. From 1981 forward, Q344 has been assigned to the photosensitive compound the remaining material when the photosensitive composition has been indexed with either of codes Q345 or Q346, as appropriate.
2. Q347, Q348, and Q349 are only assigned to substances that are not coded elsewhere in Subset Q34:.

Q4: DESCRIPTORS FOR COMPOUNDS INDEXED IN CPI SECTIONS H THROUGH M

Q41: Petroleum, Fuels (CPI Section H)

Q411	Carbon black; coke; production of carbon black or coke	(M3, M4: 1970)
Q412	Crude oil production, transport, storage	(M3, M4: 1970)
Q413	Gaseous fuel; gaseous fuel additive	(M3, M4: 1970)
Q414	Liquid fuel; liquid fuel additive	(M3, M4: 1970)
Q415	Solid fuel of petroleum origin; additive to solid fuel of petroleum origin	(M3, M4: 1970)
Q416	Lubricant; lubrication; lubricant additive	(M3, M4: 1970)
Q417	Other petroleum product; other petroleum production	(M3, M4: 1970)
Q418	Coal; coal production	(M3, M4: 1970)
Q419	Other non-petroleum fuel product; production of other non-petroleum fuel	(M3, M4: 1970)

Note Rocket fuel is coded Q443, not in Subset Q41:.

Q42: Catalysis (CPI Section J)

Q421	Catalyst, including initiator	(M3, M4: 1970)
Q422	Catalyst accelerator	(M3, M4: 1970)
Q423	Catalyst carrier	(M3, M4: 1970)

Note The complete indexing of catalysts is discussed on page 191.

Q43: Other Chemical Engineering Descriptors (CPI Section J)

Q431	Separation of solids; mixing; filtration	(M3, M4: 1970)
Q432	Boiling; refrigeration; cooling (excluding the refrigerant itself)	(M3, M4: 1970)
Q433	Refrigerant	(M3, M4: 1970)
Q434	Drying; heat exchange	(M3, M4: 1970)
Q435	General process; laboratory apparatus	(M3, M4: 1970)
Q436	Air pollution control or prevention	(M3, M4: 1970)
Q437	Water pollution control or prevention	(M3, M4: 1970)
Q438	Land pollution control or prevention	(M3, M4: 1970)

Q439	Civil or industrial waste treatment, at source or after collection (M3, M4: 1970)

Notes on Q43: Codes

1. Q433 is coded for compounds disclosed as refrigerants. Other compounds involved in processes other than their own synthesis, where refrigeration is important, are coded Q432.
2. Q435 was assigned to pre-1981 references to general chemical (Chemdoc - Section E) **analysis**, **testing**, and **processes**. Q435 is no longer used for general chemical **processes** because Part N: codes introduced in 1981 describe such processes more specifically. For example, chemical analysis procedures and apparatuses are coded N102, not Q435.
3. Q437 includes sewage and wastewater treatment.
4. Q439 can be used in combination with Q436, Q437, or Q438 if applicable.

Q44: Nucleonics, Explosives (CPI Section K)

Q441	Fire-fighting composition, including breathing apparatus, protective materials, etc. (M3, M4: 1970)
Q442	Explosive charge, blasting (M3, M4: 1970)
Q443	Explosives; matches; rocket fuel (M3, M4: 1970)
Q444	Nucleonics (M3, M4: 1970)

Q45: Refractive Materials; Cement; Electroorganic or Electroinorganic Substance – (CPI Section L)

Q450	General Section L descriptor, pre-1981 Dyes only (M4: 1970-1981)
Q451	Abrasive (M3: 1970)
Q452	Glass; vitreous enamel (M3: 1970)
Q453	Cement; ceramic; refractory compound or composition (M3: 1970)
Q454	Electroorganic or electroinorganic substance (M3: 1970)

Notes on Q45: Codes

1. Liquid crystals are coded Q334, not in Subset Q45:.

Q46: Metallurgy (CPI Section M)

Q461	Metal cleaning (M3, M4: 1970)
Q462	Corrosion inhibition (M3, M4: 1970)
Q463	Electroplating; electroforming; electrolytic surface treatment of metal or with metal (M3, M4: 1970)
Q464	Non-electrolytic coating, e.g. diffusion coating (M3, M4: 1970)
Q465	Other metal surface treatment (M3, M4: 1970)
Q466	Shaping and casting of metal (M3, M4: 1970)
Q467	Soldering; welding (M3, M4: 1970)
Q468	Flux - general (M3, M4: 1970)
Q469	Alloy and metal production (M3, M4: 1970)

Q5: CHEMICAL PROPERTIES

Q500	Other uses of dye, excluding detecting agents and indicators (Pre-1981 Dyes only) (M4: 1970-1981)
Q501	Acid acceptor (M3: 1970)
Q502	Base acceptor (M3: 1970)
Q503	Buffer (M3: 1970)
Q504	Chelating agent (M3: 1970)
Q505	Detecting agent, indicator (M3, M4: 1970)
Q506	Ion exchange agent (M3: 1970)
Q507	Oxidising agent (M3: 1970)

Q508	Purifying agent	(M3: 1970)
Q509	Reducing agent	(M3: 1970)

Note Most codes for dyes are found in Part W:.

Q6: OTHER PROPERTIES AND USES

Notes on Q6: Codes

1. Stabilisers are indexed in Subset Q62:, not Q616. General chemical synergists are assigned the code P863, not Q616.

2. Personal deodorants are assigned the codes Q604 and Q251.

3. Galenical additives are indexed in Set R3:, not in Set Q6:.

4. Prior to 1981, thermal insulators and sound insulators were assigned the codes Q450 (dyes) or Q453 (non-dyes). From 1981, Q453 is used.

5. U.V. absorbers used in skin protection are assigned the code Q263, not Q623.

Q60: and Q61: Miscellaneous properties and uses

Q600	Other uses, excluding those coded Q613 or Q618 (Pre-1981 dyes only)	(M4: 1970-1981)
Q601	Anticaking, antisticking	(M3: 1970)
Q602	Antifoam	(M3: 1970)
Q603	Antistatic	(M3: 1970)
Q604	Deodorant	(M3: 1970)
Q605	Drier; solvent remover	(M3: 1970)
Q606	Filler; reinforcing agent	(M3: 1970)
Q607	Flocculent; antiflocculent; antidusting	(1981)
Q608	Foaming agent; foam stabiliser	(M3: 1970)
Q609	Functional fluid (e.g. antifreeze)	(M3: 1970)
Q610	Electrical insulation; dielectric	(M3: 1970)
Q611	Magnetic material	(M3: 1970)
Q612	Odorant (excluding perfume)	(1981)
Q613	Optical bleach; fluorescent	(1963)
Q614	Plasticiser	(M3: 1970)
Q615	Solvent; swelling agent; solubiliser	(M3: 1970)
Q616	Surfactant, emulsifier	(1963)
Q617	Thermal material (e.g. insulation)	(1981)
Q618	Trace material; tracer	(M3, M4: 1970)
Q619	Viscosity modifier; thixotropic	(M3: 1970)

Q62: Stabilisers

Q620	Stabiliser - general	(1963)
Q621	Flame-proofing; smoke-reducing	(M3, M4: 1970)
Q622	Stabiliser for heat	(M3, M4: 1970)
Q623	Stabiliser for radiation; U.V. absorber	(M3, M4: 1970)
Q624	Stabiliser for oxidation	(M3, M4: 1970)

12 Part R: Pharmaceutical Formulations; Medical Devices and Analysis; Packaging

12 Part R: - Pharmaceutical formulations; medical devices and analysis; packaging

Main Headings

R0: Formulation Types1
R1: General Terms for Galenicals
R2: Routes of Pharmaceutical Administration
R3: Additives in Formulations
R4: Pharmaceutical Administration Devices
R5: Formulation Processes, Apparatuses, and Machines
R6: Diagnostic and Bionalytical Methods
R7: Packaging

Introduction

Part R: codes cover pharmaceutical formulations, routes of administration, medical devices and diagnostics, additives, and packaging.

Notes on Codes in Part R:

1. Set R0: codes are used with a variety of subheadings, but codes in Sets R1: through R7: are only used with the **subheading M6**. M6 was introduced in 1976, and it is used for novel pharmaceutical (Farmdoc - Section B) formulations; administration devices; production of dosage forms; formulation machines, devices and physical processes; and packaging.

2. At least one of the codes in Set R0: is applied whenever M781 ("Use of one compound") or M782 ("Use of two or more compounds") is coded, i.e. for **patents on compositions**. Otherwise, Part R: codes are only assigned if they describe an important feature of the invention.

Note: This applies to any claimed use/composition patent, not only pharmaceuticals.

---

3. Concepts describing shape or form searched with Part R0: codes are listed in the index at the end of this manual and in the *Chemical Code Dictionary*. The code that is applied usually corresponds to the shape or form in which the compound is commercially used. For example, an emulsion paint is coded R022 ("Emulsion"), not R043 ("Film or sheet").

Subheadings Applicable to Part R: Codes

Unless otherwise specified set R0: is applicable to subheadings M0-M5 from 1963, M6 from 1976. Sets R1: to R7: apply to subheading M6 (from 1976 onwards unless otherwise specified).

R0: FORMULATION TYPES

R00: General Use Descriptors (Pre-1981 only)

R000	Pharmaceutical (1963, M0-M2, M5)
R001	For human use (1976, M6)
R002	For veterinary use (1963, M0-M2, M5)
R003	For agricultural use (1965, M0-M2, M5)
R004	For general chemical use (1970, M3, M4)

Note R000 and R004 were not applied if the composition was in a specific form; for example, a tablet was coded R038, not R000.

R01: Gaseous forms

R010	Gaseous form - general (Pre-1981 precursor code for Subset R01:) (1963-1981, M0-M5)
R011	Aerosol (1976, M6, 1981 otherwise)
R012	Inhalant (1976, M6, 1981 otherwise)
R013	Other gaseous form (i.e. gas, smoke, etc.) (1976, M6, 1981 otherwise)

R02: Liquid or semi-solid forms

R021	Cream; paste
R022	Emulsion
R023	Solution; syrup; liquid
R024	Suspension or dispersion of solids

Note Liquid crystals are coded in Q33: together with R023 and R032. The combination of R023 and R032 was also used to describe liquid crystals prior to introduction of specific liquid crystal (Q33:) codes.

R03: Solid forms; moulded articles

R031	Capsule
R032	Grain; granule; crystal (see also R023 for liquid crystals)
R033	Microcapsule
R034	Pellet
R035	Dusting powder
R036	Other kind of powder
R037	Suppository
R038	Tablet; other moulded articles

R04: Foams; other solid forms

R041	Bandage; wound dressing
R042	Filament; suture; wire
R043	Film; sheet (1963 M0-M2, M5; 1970 M3, M4; 1981, M6)
R044	Foam; porous or expanded solid form (1970 M3, M4; 1976, M6; 1981 others)
R045	Laminate
R046	Prosthesis (1976, M6; 1981 otherwise)
R047	Surgical sponge; tampon (1976, M6; 1981 otherwise)

R05: Timed release formulations

R051	Delayed release
R052	Sustained release

Note R051 and R052 are not searchable for general-use chemicals (Chemdoc - Section E) prior to 1981.

R1: GENERAL TERMS FOR GALENICALS

R111	Novel dosage form
R112	Novel production of dosage form
R120	Coated form
R130	Disposable
R140	Ink; label
R150	Multicomponent
R160	Reference standard
R170	Special shape

Note The codes R111 and R112 are used in combination with an R0: code for the particular dosage form.

R2: ROUTES OF PHARMACEUTICAL ADMINISTRATION

R200	Buccal; sublingual
R210	External, topical
R220	Implant
R231	Injection
R232	Infusion
R241	Intraarterial
R242	Intravenous
R251	Intraaural
R252	Intranasal; intratracheal
R253	Intraocular
R261	Intramuscular
R262	Subcutaneous
R263	Transdermal (Searchable since 1983/Week 37)
R271	Intrauterine
R272	Intravaginal
R280	Oral
R290	Rectal

R3: ADDITIVES IN FORMULATIONS

R301	Absorbent; adsorbent
R302	Accelerator
R303	Adhesive; binder
R304	Anticaking, disintegrating agent
R305	Buffer
R306	Chelating agent
R307	Coating
R308	Diluent; filler
R309	Dye; indicator
R310	Flavour
R311	Gelling agent
R312	Lubricant
R313	Plasticiser
R314	Propellant
R315	Stabiliser, general
R316	Stabiliser, to heat
R317	Stabiliser, to oxidation
R318	Stabiliser, to radiation
R319	Surfactant
R320	Suspending agent

Note Additives are coded in M0-M5 and their uses are coded in Part Q: where the additive itself is important. For example, a particular stabiliser is used in a pharmaceutical formulation, the additive would be indexed in M2 with appropriate Q: codes, whilst the formulation would be covered in M6 with appropriate R: codes.

R4: PHARMACEUTICAL ADMINISTRATION DEVICES

R410	Catheter
R420	Dosage into orifice
R430	Implanting device
R440	Inhalation device
R450	Injection gun

R46: Syringes

R460	Syringe - general
R461	Hypodermic
R462	Remote control
R463	Syringe needle

R5: PROCESSES; APPARATUSES; MACHINES

R50: General

R501	Physical process or apparatus
R502	Chemical process or apparatus

Note R501 and R502 are only used for processes and apparatuses that cannot be more specifically coded elsewhere in Set R5:.

R51: Analytical; diagnostic; testing

R510	Analytical; diagnostic; testing - general (Pre-1981 precursor code for Subset R51:)	
R511	Automated analysis, diagnosis, or test	(1981)
R512	Chromatographic	(1981)
R513	Isotope detection	(1981)
R514	Spectroscopic, colorimetric	(1981)
R515	Other processes or apparatuses for analysis, diagnosis, or testing	(1981)

R52:, R53: Other processes and apparatuses

R520	Agglutination; precipitation
R521	Biological; fermentation
R522	Coating
R523	Compression; vacuum (see also R531)
R524	Cooling; freezing
R525	Cutting; grinding
R526	Distilling; heating
R527	Drying; lyophilisation
R528	Electrical; magnetic
R529	Filling
R530	Flow control; transporting (excluding aerosol cans or valves, which is coded R700)
R531	Forming; moulding (see also R523)
R532	Impregnating
R533	Microscopy; optical
R534	Mixing
R535	Purification; separation (excluding drying and lypholisation, which are coded R527)
R536	Radiation
R537	Sampling
R538	Spraying

R6: DIAGNOSTIC AND ANALYTICAL METHODS

R61: Material tested

R611	Blood or blood plasma	(1981)
R612	Urine	(1981)
R613	Other body fluid	(1981)
R614	Other material	(1981)

R62: Reagent used

R621	Antibody	(1981)
R622	Antigen	(1981)
R623	Colour indicator	(1981)
R624	Enzyme	(1981)
R625	Fluorescent-labelled material	(1981)
R626	Isotopically or radioactively labelled material	(1981)
R627	Other reagent used	(1981)

R63: Substance detected

R630	Antibody	(1981)
R631	Antigen	(1981)
R632	Enzyme	(1981)
R633	Drug	(1981)
R634	Fat	(1981)
R635	Microorganism	(1981)
R636	Pesticide or other agrochemical	(1981)
R637	Protein	(1981)
R638	Steroid	(1981)
R639	Other substance detected	(1981)

Note A substance detected is only coded R633 if no other code in Subset R63: is applicable.

R7: Packaging	
R700	Aerosol can and valve
R710	Ampoule
R720	Bottle
R730	Box; carton
R740	Bag; blister/bubble pack
R750	Cap; closure
R760	Dispenser
R770	Safety container

Note Prior to the introduction of Set R7: codes in 1976, N101 was assigned to pharmaceutical (Farmdoc – Part B:) packaging materials.

13 Steroid Descriptors

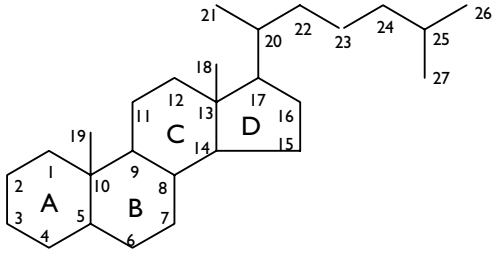
13 Steroid descriptors

Introduction

Codes in Parts S:, T:, and U: refer to steroids, i.e. compounds containing a complete cyclopenta(a)phenanthrene ring system. Steroid descriptors in Parts S:, T:, and U: are based on a coding system used previously at the U.S. Patent and Trademark Office (USPTO). An understanding of the coding rules listed in the USPTO document is essential for proper use of the steroid codes presented in this chapter. The rules of this document are included as notes in the appropriate code sections.

The Selection of Codes for Steroids

To be indexed as a steroid, a structure must contain a complete cyclopenta(a)phenanthrene ring system. The ring system may have additional rings fused to it, but it cannot be broken, e.g. as in secosteroids; Carbon atoms in the ring cannot be replaced by heteroatoms, e.g. as in oxa-steroids; and the rings cannot vary in size from the basic ring system, e.g. as in ring nor-steroids and homo-steroids. Steroid codes are defined in terms of ring letters A through D and position numbers 1 through 23, as illustrated below.



A steroid is coded in Part S:, T:, or U: if it is a:

- starting material;
- intermediate in a reaction that produces a steroid;
- end product; or
- component in a composition.

Prior to 1981, all steroid codes assigned to a patent were listed in a single database field, i.e. all steroids from a patent were overcoded on one **code subfield**. (Code subfields are discussed on page 28.) From 1981 forward, codes assigned to individual steroids in a steroid composition are listed in separate code subfields, as are codes for steroid starting materials, intermediates, and end-products, to reduce the false combinations caused by listing codes of unrelated structures together.

Format of Steroid Codes

Each Set in Parts S:, T:, and U: covers a single feature of the steroid: unsaturation, carboxylic acid substituent, epoxy subsituent, etc. Within each Set, codes known as ***2A-terms*** indicate the position on the steroid ring system where the feature described by that Set is found. The last two digits of each code (01 to 23) correspond to a numbered position on the steroid nucleus. For example, a hydroxy group at position 3 is coded S503 (Set S5: pertains to hydroxy substituents). Listed next to each 2A-term is the position number on the steroid nucleus of the feature being described. Codes ending in "00" indicate that the feature is ***exo***, i.e. not directly bonded to the **steroid nucleus**. (The nucleus includes the group bonded to position 17.) Codes ending in "23" indicate position 23 or higher.

In some cases, the 2A-terms in a Set do not completely describe a feature. For example, Set S7: refers to acyloxy groups, but an additional code is required to specify the particular acyl group present. Such additional codes are called ***2B-terms***, and they are listed after the 2A-terms to which they pertain. The last two digits of 2B-terms are 30 or higher.

A code with an **indented definition** is only assigned if the closest preceding code with an unindented definition is also assigned. For example, U521 ("Position 21 unsubstituted") is only coded if U520 ("2 C atoms in chain bonded to position 17") is also coded. **Hyphens** indicate positions where a particular feature cannot occur.

Codes Applicable To Steroids

Listed below are the codes within each code Part that are searchable with subheading M5, the subheading reserved for steroids. (Subheadings are discussed on page 23.)

1 Parts A:, B:, and C: - Elements Present

From 1963 forward: C811 ("Radioactive element"),
C812 ("Specific isotope")

2 Parts D:, E:, F:, and G: - Ring Systems

The cyclopenta(a)phenanthrene ring system itself is never assigned codes in Parts D: through G:, nor is the steroid Ring Index Number (RIN = 04781) ever assigned. However, when the cyclopenta(a)phenanthrene ring system is spiro or fused to another ring or ring system, the RIN for the complete system is assigned. The RIN's of rings spiro to the steroid nucleus may also be assigned, but the RIN for the steroid nucleus itself is not assigned. (The application of Ring Index Numbers is discussed on page 74.)

3 Parts H:, J:, K:, L: - Functional Groups

Functional groups bonded to the steroid nucleus are coded in Parts S:, T: and U:, not in Parts H:, J:, K:, and L:.

4 Part M: - Miscellaneous Descriptors

From 1981 forward: "Basic Group" codes
M417 ("Incomplete structure coded"),
M430 ("Composition with no active ingredients"), and
M431 ("Composition with one or more active ingredients");
Set M7: ("Patent type")

5 Part N: - Chemical Reactions, Processes, and Apparatuses

From 1963 forward: Subset N03: ("Fermentation processes");
N101 ("Containers, packaging");
N161 ("Extraction from natural materials")

From 1981 forward: all other codes in Part N:.

6 Parts P: and Q: - Activities, Properties, and Uses

From 1963 forward: Only general codes such as P420 ("Antiinflammatory - general or unspecified")

From 1981 forward: all codes in Parts P: and Q:.

Note More specific retrieval of pre-1981 references to activities or uses can usually be obtained by searching with Manual Codes. (Manual Codes are discussed on page 26.)

7 Part R: - Pharmaceutical Formulations; Medical Devices and Analysis; Packaging

From 1963 forward: Set R0: ("Formulation Types").

8 Parts S:, T:, and U: - Steroid Descriptors

From 1963 forward: All codes in Parts S:, T:, and U:.

Note Codes in Parts S:, T:, and U: are only used with subheading M5.

9 Part V: - Natural Products; Polymers

From 1963 forward: V340 ("Vitamin D derivative or intermediate")

From 1981 forward: V402 ("Steroid glycoside")

Note A steroid linked to a natural product or polymer that is coded in Part V: is assigned all applicable Part V: codes.

10 Part W: - Dye Descriptors

From 1981 forward: W314 ("Steroid dye")

Part S: - Steroid Descriptors I

Main Headings

S0: Carbon to Carbon Double Bonds
S1: Ring Saturation and Substitution
S2: Alpha Configuration
S3: Beta Configuration
S4: Miscellaneous Chemical Groups
S5: Hydroxy
S6: Keto
S7: Acyloxy
S8: O-R
S9: Epoxy

Note The diagram on page 233 is labelled with the position numbers and ring letters referred to in the code definitions below.

S0: CARBON TO CARBON DOUBLE BONDS

Code	Position	Code	Position
S000	Exo	S012	12
S001	1	S013	13
S002	2	S014	14
S003	3	S015	15
S004	4	S016	16
S005	5	S017	17
S006	6	S018	18
S007	7	S019	19
S008	8	S020	20
S009	9	S021	21
-	10	S022	22
S011	11	S023	23 or higher

Positions joined by double bond

S030	1 - 2	S036	9 - 10
S031	1 - 10	S037	9 - 11
S032	5 - 6	S038	13 - 14
S033	5 - 10	S039	13 - 17
S034	8 - 9	S040	20 - 21
S035	8 - 14	S041	20 - 22

S050 Ring "A" is an aromatic ring

Note 2-A terms – in coding the positional location of a double bond, the lower number of the pair of position numbers in recorded. Thus for $\Delta^{4,5}$, position 4 is coded and for 17(20), code 17. The 2-A field punches will not be definitive for positions 1, 5, 8, 9, 13 and 20; distinction between the two variants in each of these positions is provided in the 2-B field (S030-S041. The unsaturated linkage of methylene groups connected to the steroid nucleus is coded as "exo" double bond only.

An aromatic A ring is coded 1, 3, 5 in the 2-A field, and in addition is coded 1(2), 5(10) and aromatic A in the 2-B field.

Ring saturation is also indicated in the 2-B terms. Double bonds that are common to two rings are treated as follows:

5(10)	-	A ring
8(9)	-	C ring
13(14)	-	C ring

S1: RING SATURATION AND SUBSTITUTION

Atoms bonded to H

S110	H at position 10
S113	H at position 13
S120	H at position 20

Note These codes are applied only when the hydrogen atom is attached to the 10, 13 or 20 carbon atom to indicate the absence of a methyl group in the 18, 19 or 21 position. It is not used for a 17-methyl group.

Saturated rings

S131	Ring "A" saturated
S132	Ring "B" saturated
S133	Ring "C" saturated
S134	Ring "D" saturated

Unsubstituted rings

S141	Ring "A" unsubstituted
S142	Ring "B" unsubstituted
S143	Ring "C" unsubstituted
S144	Ring "D" unsubstituted

Note -H

Ring A – no substituents at any of the positions 1, 2, 3, 4, 5 and 10 in any one of the compounds of the document. (10-methyl is not a substituent)

Ring B – positions 6 and 7

Ring C – positions 8, 9, 11, 12, 13 and 14 (13-methyl is not a substituent)

Ring D – positions 15, 16 and 17 (the 17-side chain in pregnanes, sterols, etc., is considered a substituent)

S2: ALPHA (ALLO) CONFIGURATION

Code	Position	Code	Position
-	Exo	S212	12
S201	1	S213	13
S202	2	S214	14
S203	3	S215	15
S204	4	S216	16
S205	5	S217	17
S206	6	S218	18
S207	7	S219	19
S208	8	S220	20
S209	9	S221	21
S210	10	S222	22
S211	11	S223	23 or higher

Note α or allo – this term represents a particular orientation at an asymmetric steroidal carbon.

The α descriptor is implied and coded wherever the configuration is known from the trivial name of a steroid, e.g. "cholic acid" is coded 3α, 7α, 12α and "cortisone" receives a 17α code.

When a steroidal carbon is substituted such that one of the groups must be α-oriented (no hydrogens present on the carbon) the α descriptor is automatically coded e.g. a 20-cyanohydrin pregnane or any 17-disubstituted steroid.

The 9α and 14α codes are assumed and not coded unless the substituent is other than H.

Note that i compounds have a specific 2-B punch and α or allo is not used to identify i compounds. The i compounds are located in the 2-B under "General".

However, a word of caution in using the α descriptor. Since, in the absence of a specific statement in the document, the coder may not have known whether a particular substituent was α or not, the descriptor may have been omitted in some cases. It should not be used when a complete search is required.

S3: BETA CONFIGURATION

Code	Position	Code	Position
-	Exo	S312	12
S301	1	S313	13
S302	2	S314	14
S303	3	S315	15
S304	4	S316	16
S305	5	S317	17
S306	6	S318	18
S307	7	S319	19
S308	8	S320	20
S309	9	S321	21
S310	10	S322	22
S311	11	S323	23 or higher

Note β-(Beta) – this code is only used when the document *specifically states* the β configuration. It is never implied and therefore may not be relied on for complete searches.

The 5-B code is not used unless the substituent is other than hydrogen.

Since coding the α and β descriptors is sometimes arbitrary, they should not be used if a complete search is desired.

S4: MISCELLANEOUS CHEMICAL GROUPS

Code	Position	Code	Position
-	Exo	S412	12
S401	1	S413	13
S402	2	S414	14
S403	3	S415	15
S404	4	S416	16
S405	5	S417	17
S406	6	S418	18
S407	7	S419	19
S408	8	S420	20
S409	9	S421	21
S410	10	S422	22
S411	11	S423	23 or higher

Note Miscellaneous – this designation is used for any group not provided for by any other 2-A term.

Examples of groups in this category are: -O-Na, azides (consider  as a unit), and Grignard intermediates.

Rules of Superiority. The Miscellaneous code is used only if no other designation is applicable. (In cases of doubt between C-sub and Misc., the C-sub code takes precedence.)

S5: HYDROXY

Code	Position	Code	Position
S500	Exo	S512	12
S501	1	S513	13
S502	2	S514	14
S503	3	S515	15
S504	4	S516	16
S505	5	S517	17
S506	6	S518	18
S507	7	S519	19
S508	8	S520	20
S509	9	S521	21
S510	10	S522	22
S511	11	S523	23 or higher

S6: KETO

Code	Position	Code	Position
S600	Exo	S612	12
S601	1	-	13
S602	2	-	14
S603	3	S615	15
S604	4	S616	16
-	5	S617	17
S606	6	S618	18
S607	7	S619	19
-	8	S620	20
-	9	S621	21
-	10	S622	22
S611	11	S623	23 or higher

S7: ACYLOXY

Code	Position	Code	Position
S700	Exo	S712	12
S701	1	S713	13
S702	2	S714	14
S703	3	S715	15
S704	4	S716	16
S705	5	S717	17
S706	6	S718	18
S707	7	S719	19
S708	8	S720	20
S709	9	S721	21
S710	10	S722	22
S711	11	S723	23 or higher

Type of O-acyl group present

S730	Carboxylic
S731	poly
S732	unsaturated
S733	substituted
S734	aromatic
S735	aliphatic
S736	straight chain
S737	cycloaliphatic
S738	branched
S740	Heterocyclic
S750	Inorganic
S751	phosphorous acid
S752	SO_4 (all S)
S753	osmium
S754	boron
S755	aliphatic
S756	aromatic
S760	N-containing
S761	S-containing
S762	O-containing
S763	Halo-containing
S770	Miscellaneous

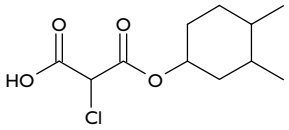
Note -O-Acyl – this designation refers to an ester group attached to the steroid throught the -O- atom of the group. It is further defined in the 2-B terms as follows:

Carboxylic – a carboxylic acid radical

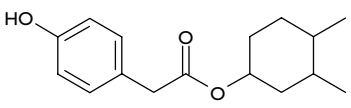
Poly – a polycarboxylic acid radical (exo-COOR is not coded when poly is recorded)

Unsat – an unsaturated acid radical. The double bond is not coded exo in 2-A. Aromatic unsaturation is excluded (e.g. benzoyl group is not coded as an unsaturated acid radical).

Subst – a hydrocarbon carboxylic acid radical having a non hydrogen susbtituent. This term is applicable to poly, unstaturated, aromatic, and aliphatic radicals. The carboxyl group of a polycarboxylic acyl is excluded from this category. Examples are:



coded as carboxylic, poly, subst, halo cont., "exo" halo, and "Cl" in 2B. (S730, S731, S733, S763, T200, T231)



coded as carboxylic, aromatic, subst, O-cont., and "exo" OH. (S730, S733, S734, S762, S500).

Aromatic – a carboxylic acid radical containing an aromatic hydrocarbon ring

Aliphatic – an aliphatic, non-aromatic carboxylic acid radical.

St. Chain – a straight chain aliphatic acid radical wiht no non-hydrogen susbtituents (formate and acetate included).

Cycloalkyl – an aliphatic hydrocarbon carboxylic acid containing a cycloalkyl group (e.g. cyclopentylpropionate)

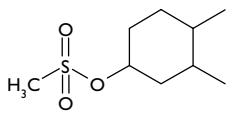
Branched – an aliphatic branched hydrocarbon carboxylic acid radical

Heterocyclic – a heterocyclic containing carboxylic or inorganic acid radical (e.g. nicotinate)

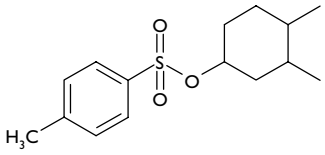
Inorganic Acyl – includes the acyl group of any inorganic acid with the exception of the halogen acids. (This term includes the acyl radicals derived from carbonic acid, its alkyl esters, and its inorganic derivatives (e.g. phosgene), but does *not* include acyl radicals derived from carbamic or xanthic acids – these are coded as substituted aliphatic carboxylic acid radicals).

Phosphorus A – an inorganic phosphorous - containing acyl radical.

$S0_4$ (all S) – this term includes all inorganic sulfur-containing acyl radicals. Examples are the mesyl and tosyl radicals.



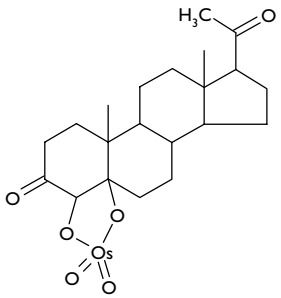
coded inorganic, SO_4, aliphatic, S-cont. and O-cont. (S750, S752, S755, S761, S762)



coded inorganic, SO_4, aromatic, S-cont., and O-cont. (S750, S752, S756, S761, S762).

Osmium – an inorganic osmium-containing acyl radical.

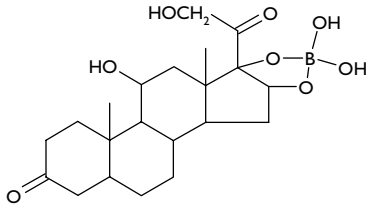
The 4(5) osmate of progesterone exemplifies this term:



The osmate radical is coded as 4, 5-O-acyl in 2-A and as inorganic, osmium, and O-containing in 2-B (S704, S705, S750, S753, S762).

Boron – an inorganic boron-containing acyl radical.

The16, 17-cycloborate of 16α hydroxyhydrocortisone exemplifies this term.



The cycloborate radical is coded as exo-OH and 16, 17 O-acyl in 2-A and as inorganic, boron, and O-containing in 2-B. (S-500, S716, S717, S750, S754, S762).

Aliphatic – an acyclic inorganic acyl radical.

Aromatic – an inorganic acyl containing an aromatic ring.

N, S, O and Halo containing are coded when an -O-Acyl radical, either organic or inorganic, contains any one of these elements, excluding the oxygen contained in the carboxyl groups of organic acyl radicals. The O-containing descriptor is coded in the case of SO_4, PO_4 and osmates. It is not coded for polycarboxylic acids.

Miscellaneous – any O-Acyl not provided for above.

(As a result of the multiple coding principle, all applicable terms are applied to a particular compound).

S8: O-R

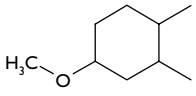
Code	Position	Code	Position
S800	Exo	S812	12
S801	1	S813	13
S802	2	S814	14
S803	3	S815	15
S804	4	S816	16
S805	5	S817	17
S806	6	S818	18
S807	7	S819	19
S808	8	S820	20
S809	9	S821	21
S810	10	S822	22
S811	11	S823	23 or higher

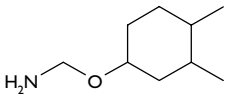
Type of R group in O-R

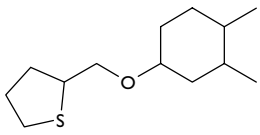
S830	Hydrocarbon
S831	N-containing
S832	S-containing
S833	O-containing
S834	Halo-containing
S835	Containing other element

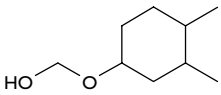
Note -O-R – this defines the ether linkage, i.e. R-O-R, where one R is a carbon which is part of the steroid nucleus (except for "Exo") and the other R is a substituent (aliphatic, aromatic, cycloaliphatic or

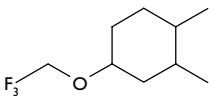
heterocyclic) which is attached to the oxygenation independently through a carbon atom and which is further defined in 2-B by the following terms:

R= hydrocarbon, e.g. 

R= N group, e.g. 

R= S group, e.g. 

R= O group, e.g. 

R= halo group, e.g. 

R= other

(Because they are specifically provided for elsewhere acetals and ketals are not coded under O-R.)

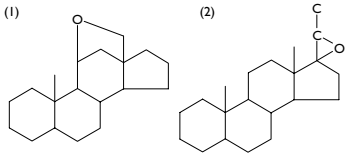
S9: EPOXY

Code	Position	Code	Position
S900	Exo	S912	12
S901	1	S913	13
S902	2	S914	14
S903	3	S915	15
S904	4	S916	16
S905	5	S917	17
S906	6	S918	18
S907	7	S919	19
S908	8	S920	20
S909	9	S921	21
S910	10	S922	22
S911	11	S923	23 or higher

Note Epoxy – this description refers to an epoxy group attached to two nuclear carbon atoms. The two positions to which the epoxy is attached are recorded.

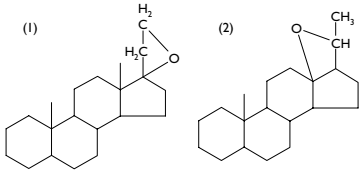
Since this term is closely related to the O-hetero category the following rules are applied:

I all carbons linked by a single oxygen atom are coded as epoxy unless the oxygen is a member of a lactone group in which case "epoxy" is not used.

E.g. 

are coded (1) 11, 18 epoxy and (2) 17, 20 epoxy.

II epoxides of non-adjacent carbons involving one or two of the 20, 21, 22 and/or 23 + carbons are also coded as O-hetero. See the O-hetero section for the proper coding procedure in such case, e.g.



are also coded (1) 17, 20 epoxy; 17-O-hetero, spiro, misc and (2) 18, 20 epoxy; 13, 17-O-hetero.

Part T: Steroid Descriptors II

Main Headings

T0: Ketal
T1: Heterocyclic Ring with O
T2: Halogen
T3: -S-R, -Se-R, or -Te-R
T4: Heterocyclic Ring with S
T5: N-R-R
T6: Heterocyclic Ring with N
T7: Keto Reagent
T8: Methyl Group
T9: -C≡C-

Note The diagram on page 233 is labelled with the position numbers referred to in the code definitions below.

T0: KETAL

Code	Position	Code	Position
T000	Exo	T012	12
T001	1	-	13
T002	2	-	14
T003	3	T015	15
T004	4	T016	16
-	5	T017	17
T006	6	T018	18
T007	7	T019	19
-	8	T020	20
-	9	T021	21
-	10	T022	22
T011	11	T023	23 or higher

Note Ketal – this designation refers to the reaction product of a keto or aldehyde group with an alcohol to give either a cyclic or non-cyclic ketal or acetal.

The term ketal includes thioketal, semithioketal, acetal, thioacetal and hemithioacetal.

Non-cyclic ketals (acetals) should also be coded under the 2-B term "Bis-substituents".

Cyclic oxygen ketals are not coded in the O-hetero category in 2-A or 2-B. Cyclic thioketals, including hemithioketals are not coded under S-hetero in 2-A but are coded as thioketal in the 2-B S-hetero column.

Non-cyclic ketals (acetals) are not coded under O-R or S-R in 2-A or 2-B.

T1: HETEROCYCLIC RING WITH O

Code	Position	Code	Position
T100	Exo	T112	12
T101	1	T113	13
T102	2	T114	14
T103	3	T115	15
T104	4	T116	16
T105	5	T117	17
T106	6	T118	18
T107	7	T119	19
T108	8	T120	20
T109	9	T121	21
T110	10	T122	22
T111	11	T123	23 or higher

Specific rings present

T130	Morpholine
T131	Furan
T132	Lactone
T133	Spirostane
T134	spirostane substituted in O-spiro
T135	Pseudosapogenin
T136	Acetonide
T137	Peroxide
T138	Pyran
T139	Other heterocyclic ring with O

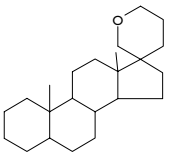
Type of attachment

T140	Spiro
T141	Fused
T142	Independent

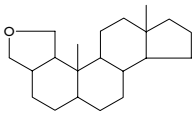
Note O-Hetero – this designation refers to an oxygen-containing heterocyclic group. The heterocycle can be attached to the steroid nucleus through any atom of the heterocycle.

The heterocycle may be fused, independent or spiro. When it is fused to the nucleus, the two nuclear positions (which need not be adjacent) to which it is fused are recorded. When it is spiro or independent the one position through which it is attached is recorded.

Spiro

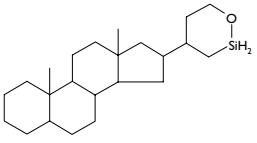


Fused



Code: 1,2-O-hetero; furan; fused

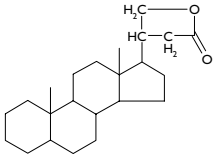
Independent



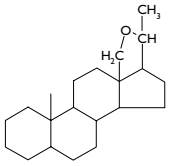
Code: 16-O-hetero; independent; misc

Acid and anhydride adducts (e.g. maleic anhydride adducts), epoxides, cyclic oxygen containing ketals and steroidal sapogenins are specifically excluded from the 2-A term O-hetero.

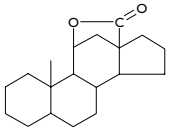
When the 18-23+ side chain carbons are members of the hetero ring, the point of attachment to the steroid *ring* is recorded unless the hetero *ring* is independently connected ro a side chain carbon, e.g.



Code: 17-O-hetero; independent; lactone; furan



Code: 13, 17-O-hetero; fused; furan (Note: Also code 18, 20-epoxy)



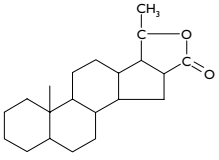
Code: 11, 13-O-hetero; fused; lactone; furan

The O-hetero is further defined in 2-A as follows:

Morpholine – this is also coded as N-hetero

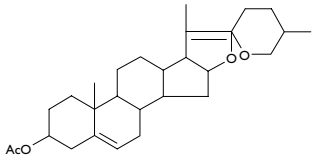
Furan – includes saturated and unsaturated forms and also 5-member lactone

Lactone – e.g.



is coded 16, 17-O-hetero, furan, lactone, fused.

Spirostane – includes both normal and iso, e.g.



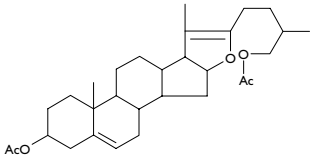
(diosgenin)

The above formula is coded as "spirostane" - (no other O-hetero descriptors in 2-A or 2-B are used)

Sub in O-Spiro Ring – spirostanes substituted in the oxygen rings by a non-hydrocarbon substituent. The substituent is also coded as "exo".

(The 2-A designations 16, 17-O-hetero and 2-B fused are not recorded for sapogenins or for pseudo-sapogenins.)

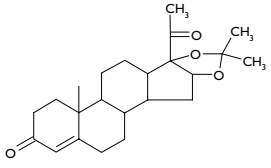
Pseudosapo – the pseudosapogenins are derived from steroidal sapogenins by treatment with acid anhydrides. The free hydroxy analog is included in this descriptor. The side chain hydroxy or acetate is not coded "exo". The double bond in the fused O-hetero ring is not coded as "exo" - it is included in the term "pseudosapogenin", e.g.



(pseudodiosgenin acetate)

is coded as "pseudosapogenin".

Acetomide – cyclic acetal or ketal which is the reaction product of an aldehyde or ketone with 2-hydroxy groups attached to the steroid nucleus. It is also coded as "fused", e.g.

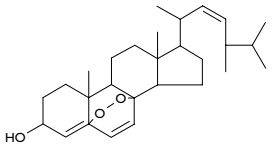


(16α, 17α isopropylidene dioxyprogesterone)

is coded 16, 17 O-hetero, acetonide, fused.

(The methyl groups may be replaced by hydrogen or any other substituent (hydrocarbon or substituted hydrocarbon, heterocyclic, halogen, etc.) or by substituents which themselves form a ring.

Peroxide – the C-O-O-C linkage. This is limited to peroxide in ring configuration only. The C-O-O-H group is coded "misc" in 2-A. The peroxide descriptor does not include "ozonide" which is coded as O-hetero-fused-misc. E.g.: the peroxide group of



(ergosterol peroxide)

is coded 5, 8 O-hetero, peroxide, fused.

Pyranyl – includes saturated and unsaturated forms.

Spiro, Fused and Independent – these terms refer to the manner in which the heterocycle ring is attached to a carbon atom or atoms of the steroid nucleus.

Miscellaneous – O-hetero susbtituents not specifically provided for the above but specifically excluding epoxides, ketals, and acid and anhydride adducts.

T2: HALOGEN

Code	Position	Code	Position
T200	Exo	T212	12
T201	1	T213	13
T202	2	T214	14
T203	3	T215	15
T204	4	T216	16
T205	5	T217	17
T206	6	T218	18
T207	7	T219	19
T208	8	T220	20
T209	9	T221	21
T210	10	T222	22
T211	11	T223	23 or higher

Specific halogens present

T230	Fluorine
T231	Chlorine
T232	Bromine
T233	Iodine

T3: -S-R, -SE-R, OR -TE-R

Code	Position	Code	Position
T300	Exo	T312	12
T301	1	T313	13
T302	2	T314	14
T303	3	T315	15
T304	4	T316	16
T305	5	T317	17
T306	6	T318	18
T307	7	T319	19
T308	8	T320	20
T309	9	T321	21
T310	10	T322	22
T311	11	T323	23 or higher

Type of group present

T330	-Se-R
T331	-Te-R
T332	=S, =Se, or =Te
T333	R is H
T334	R is not H

Note: -S(Se, Te)-R – any substituent joined to the steroid nucleus by attachment through a sulfur, selenium or tellurium atom. S-containing heterocyclics are excluded.

It is further defined in 2-B as follows:

-Se-R – The substituent is joined to the steroid nucleus through a selenium atom

-Te-R – The substituent is joined to the steroid nucleus through a tellurium atom

=S(Se, Te) – =S, =Se, or =Te is the substituent, and is attached by a double bond

R=H – used for -S-H group

R=other – this includes all -S-R groups where R is not H, including $-SO_3H$, $-SO_2H$, etc.

In using the exo 2-A term =S(Se, Te) is specifically excluded where the =S(Se, Te) is part of an acyl group. Also specifically excluded are NCS and SCN which are provided for elsewhere.

T4: HETEROCYCLIC RING WITH S

Code	Position	Code	Position
T400	Exo	T412	12
T401	1	T413	13
T402	2	T414	14
T403	3	T415	15
T404	4	T416	16
T405	5	T417	17
T406	6	T418	18
T407	7	T419	19
T408	8	T420	20
T409	9	T421	21
T410	10	T422	22
T411	11	T423	23 or higher

Specific rings present

T430	Thiophene
T431	Thiazole (see also T636)
T432	Thioketal
T433	Other heterocyclic ring with S

Type of attachment

T440	Spiro
T441	Fused
T442	Independent

Note S-het Ring – this designation refers to any sulfur-containing heterocyclic group attached to any of the 1 to 23+ postions of the steroid nucleus through any of the atoms of the heterocycle.

The same general rules for coding rings are followed that were recited under O-hetero.

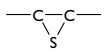
This category is further defined in 2B as follows:

Thiophene – the S-hetero in thiophene configuration, saturated or unsaturated.

Thiazole – the S-hetero in thiazole configuration, saturated or unsaturated. (Also coded as N-hetero).

Thioketal – includes both cyclic monothioketals and cyclic dithioketals. The 2-A S-hetero descriptor is not used for "thioketal". The appropriate position is coded under "ketal".

Spiro, Fused and Independent – have been defined above (see O-hetero).

Miscellaneous – S-hetero substituents not specifically provided for above. The thiirane ring  is coded here.

T5: N-R-R

Code	Position	Code	Position
T500	Exo	T512	12
T501	1	T513	13
T502	2	T514	14
T503	3	T515	15
T504	4	T516	16
T505	5	T517	17
T506	6	T518	18
T507	7	T519	19
T508	8	T520	20
T509	9	T521	21
T510	10	T522	22
T511	11	T523	23 or higher

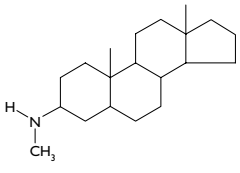
Specific groups present

T530	Primary amine
T531	Secondary amine
T532	Tertiary amine
T533	Quaternary ammonium
T534	NO_2, NO, and miscellaneous
T535	R is acyl
T536	Imino

Note: N-R-R – this defines a nitrogen containing group connected through its nitrogen atom to the nucleus. The term is not applicable when the nitrogen atom is part of a hetercyclic ring.

The 2-B descriptors for N-R-R are:

Primary, Secondary, Tertiary and Quaternary Amines – are self explantory, e.g.



is coded 3- NRR in 2-A and "secondary" amine in 2-B.

NO_2, NO and Misc. – this term includes NO_2, nitroso and all NRR groups not covered by the other definitions in 2-B (e.g. nitrone, nitrate).

R=Acyl – converts the amine group to amides.

Imino – defines an imino group.

(=N – ketone reagents are specifically excluded from N-R-R).

T6: HETEROCYCLIC RING WITH N

Code	Position	Code	Position
T600	Exo	T612	12
T601	1	T613	13
T602	2	T614	14
T603	3	T615	15
T604	4	T616	16
T605	5	T617	17
T606	6	T618	18
T607	7	T619	19
T608	8	T620	20
T609	9	T621	21
T610	10	T622	22
T611	11	T623	23 or higher

Specific rings present

T630	Morpholine
T631	Piperidine
T632	Pyridine
T633	Pyrimidine
T634	Azole
T635	Pyrrolidine
T636	Thiazole (see also T431)
T637	Piperazine
T638	Other heterocyclic ring with N

Type of attachment

T640	Spiro
T641	Fused
T642	Independent

Note N-hetero Ring – this designation refers to a nitrogen-containing hetercyclic group attached through any of the atoms of the heterocyclic group.

The heterocycle can be fused, independent, or spiro. The rules are set forth under O-hetero are followed.

The 2-B terms further define as follows:

Morpholine – (also coded as O-hetero)

Piperidine –

Pyridine – (including the dihydro and tetrahydro forms)

Pyrimidine – (including all saturated forms)

Azole – a five membered saturated or unsaturated ring containing at least one nitrogen. This is a generic term and all azoles are coded here.

Pyrrolidine – (Pyrrole) including all saturated and unsaturated forms. (Azole is also coded whenever this descriptor is used.)

Thiazole – including the saturated and unsaturated forms. (Azole is also coded whenever this descriptor is used).

Piperazine –

Spiro, Fused and Independent – have been defined above (see O-hetero.)

Miscellaneous – N-hetero substituents not specifically provided for by the terms above.

T7: KETO REAGENT

Code	Position	Code	Position
T700	Exo	T712	12
T701	1	-	13
T702	2	-	14
T703	3	T715	15
T704	4	T716	16
-	5	T717	17
T706	6	T718	18
T707	7	T719	19
-	8	T720	20
-	9	T721	21
-	10	T722	22
T711	11	T723	23 or higher

Specific groups present

T730	Hydrazone
T731	Oxime
T732	Semicarbazone
T733	Girard reagent

Note Keto Reagent – this designation refers to a reaction product of a ketone or aldehyde group attached to any of the positions 1 to 23+ with well-known keto reagents.

It is further defined in 2-B as follows:

Hydrazone, Oxime and Semi-Carbazone – are self explanatory. Thiosemicarbazone is coded as semicarbazone and "exo" S. When uncommon exo groups are present they are recorded (e.g. in dinitrophenyl hydrazone the nitro groups would *not* be coded; but in di-cyano phenylhydrazone the cyano groups would be coded). The above should be considered when attempting to find substituted keto reagents.

Girard Reagent – includes all acyl hydrazones substituted on the acyl moiety by a quaternary ammonium radical. Girard Reagents are not coded as hydrazones and are not split into "exo" terms, except for halogens.

(Substituents coded here are not coded under NRR).

T8: METHYL GROUPS

Code	Position	Code	Position
-	Exo	T812	12
T801	1	-	13
T802	2	T814	14
T803	3	T815	15
T804	4	T816	16
T805	5	T817	17
T806	6	T818	18
T807	7	T819	19
T808	8	-	20
T809	9	T821	21
-	10	-	22
T811	11	-	23 or higher

Note CH_3 – this represents a methyl group. This code is not used at positions 10, 13, 20, 22 or 23+ because methyl groups in these positions are considered integral parts of the steroid nucleus. When a methyl group is coded the 2A term "hydrocarbon chain" and the 2B term "lower alkyl (1-7)" are also coded. The "exo" CH_3 descriptor is not used.

T9: -C≡C-

Code	Position	Code	Position
T900	Exo	T912	12
T901	1	T913	13
T902	2	T914	14
T903	3	T915	15
T904	4	T916	16
T905	5	T917	17
T906	6	T918	18
T907	7	T919	19
T908	8	T920	20
T909	9	T921	21
T910	10	T922	22
T911	11	T923	23 or higher

Note When this ethynyl code is used, "hydrocarbon chain" (U001-U023) and "alkynyl" (U035) are also coded.

Part U: Steroid Descriptors III

Main Headings

U0: Hydrocarbon Chain
U1: Carbocyclic Ring
U2: -CN
U3: -COOR
U4: -C-R
U5: Carbon Chain Bonded to Position 17; General Descriptors

Note The diagram on page 233 is labelled with the position numbers and ring letters referred to in the code definitions below.

U0: Hydrocarbon Chain

Code	Position	Code	Position
U000	Exo	U012	12
U001	1	U013	13
U002	2	U014	14
U003	3	U015	15
U004	4	U016	16
U005	5	U017	17
U006	6	U018	18
U007	7	U019	19
U008	8	U020	20
U009	9	U021	21
U010	10	U022	22
U011	11	U023	23 or higher

Type of chain

U030	Lower alkyl (1 to 7 C atoms)
U031	Higher alkyl (More than 7 C atoms)
U032	$=CH_2$
U033	Alkenyl
U034	With aromatic ring
U035	Alkynyl

Note HC (hydrocarbon) chain – this term represents a hydrocarbon chain. Methyl groups at the 10 and 13 positions are not coded (see the $-CH_3$ descriptor). Also, in pregnanes, the chain of 2 carbons attached to the 17 position is not coded as a hydrocarbon unless it is substituted (i.e. $-CH_2$, CH_3). In cholesterols and other sterols, the hydrocarbon chain in the 17, 20 and 22 positions is not coded as a hydrocarbon.

Hydrocarbon chain includes aralkyl or any other cyclic hydrocarbon ring attached to the steroid nucleus through an aliphatic carbon chain but the carbons of the ring are not counted as chain members.

This descriptor is not used at positions 20, 22 or 23+ (except in the case of 20-, 22- or 23+- ethinyl) since carbon attached to these positions are integral parts of the steroid nucleus.

"Exo"-HC Chain is not coded except when used in conjunction with the "exo" -C≡C- descriptor.

The 2-B definitions are:

L. Alkyl (1-7) – a chain of 1 to 7 carbons

Hi. Alkyl (8+) – a chain of 8 or more carbons

$=CH_2$ – this includes both H and unsubstituted hydrocarbon substituents attached to =C.

Alkenyl – if the HC chain contains a double bond other than at the point of attachment to the steroid nucleus this code is used.

With Aromatic – an aliphatic side chain substituted by an aromatic hydrocarbon ring.

Alkinyl – if the HC chain contains a triple bond this descriptor is used.

U1: CARBOCYCLIC RING

Code	Position	Code	Position
-	Exo	U112	12
U101	1	U113	13
U102	2	U114	14
U103	3	U115	15
U104	4	U116	16
U105	5	U117	17
U106	6	U118	18
U107	7	U119	19
U108	8	U120	20
U109	9	U121	21
U110	10	U122	22
U111	11	U123	23 or higher

Ring size

U133	3-membered ring
U134	4-membered ring
U135	5-membered ring
U136	6-membered ring
U137	7-membered or larger ring

Type of attachment

U140	Spiro
U141	Fused
U142	Independent

Type of ring

U150	Saturated
U151	Unsaturated (non-aromatic)
U152	Aromatic

Note Hydrocarbon ring – this term includes any hydrocarbon ring including those having non-hydrocarbon substituents, attached to the nucleus in positions 1-23+. In the case where the ring is fused to the nucleus the positions on the ring to which it is attached are recorded. The "exo"-HC ring descriptor is not used.

The 2-B definitions are:

3M, 4M, 5M, 6M and 7M – all denote the number of carbon atoms in the ring (M-members).

Sat – used if the ring is saturated

Unsat (N arom.) – used if the ring is unsaturated but not for aromatic rings.

Aromatic –

Spiro, Fused and Independent – are described under O-hetero.

U2: -CN

Code	Position	Code	Position
U200	Exo	U212	12
U201	1	U213	13
U202	2	U214	14
U203	3	U215	15
U204	4	U216	16
U205	5	U217	17
U206	6	U218	18
U207	7	U219	19
U208	8	U220	20
U209	9	U221	21
U210	10	U222	22
U211	11	U223	23 or higher

U3: -COOR

Code	Position	Code	Position
U300	Exo	U312	12
U301	1	U313	13
U302	2	U314	14
U303	3	U315	15
U304	4	U316	16
U305	5	U317	17
U306	6	U318	18
U307	7	U319	19
U308	8	U320	20
U309	9	U321	21
U310	10	U322	22
U311	11	U323	23 or higher

Type of group

U330	R is H
U331	R is a metal or ammonium salt
U332	R is an alkyl group
U333	-C(=O)-W (W is a heteroatom)
U334	-C(=W)-O- (W is a heteroatom)

Note: COOR – this terms represents carboxylic acid radicals and their salts, esters, and amides. It also includes thioesters, thioamides, and halides. COOR is considered as a unit and the -OH, =O, -NH_2, etc, portions are not coded elsewhere. The 2-B terms are:

R=H – the carboxyl group

R=salts – metal and amine salts of the carboxyl group

R=alkyl – esters of the carboxyl group with any alcohol, including aromatic alcohols.

C(O)-X, X=hetero – X includes any hetero atom, the most common of which are nitrogen (amides) or the halides.

C(X)-O, X=hetero – most commonly X will be sulfur. (In the case of dithio acids, both of the latter codes are recorded.)

U4: -C-R

Code	Position	Code	Position
-	Exo	U412	12
U401	1	-	13
U402	2	U414	14
U403	3	U415	15
U404	4	U416	16
U405	5	-	17
U406	6	U418	18
U407	7	U419	19
U408	8	-	20
U409	9	U421	21
-	10	-	22
U411	11	-	23 or higher

Type of group

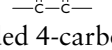
U430	R is a N group
U431	R is a S group
U432	R is an O group
U433	R is a halogen group
U434	R is not one of the groups listed above and is not oxalyl
U435	Oxalyl

Note -C-R (C-subs) – this symbol represents a non-hydrocarbon substituent linked through a carbon atom to the steroid nucleus and not specifically provided for by any other 2-A term.

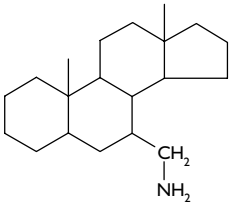
In cases of doubt the C-sub is superior to "Miscellaneous" and if a group can be coded in C-sub, it is coded there rather than in "Miscellaneous".

Substituted cycloalkyls attached to the nucleus are coded HC-ring and not as C-sub. The substituting groups are coded as "exo".

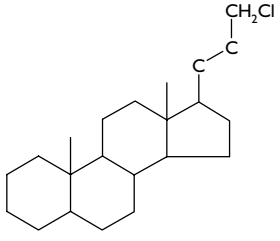
The substituted carbon group is further defined in 2-B

R=N	containing group
R=S	containing group
R=O	containing group
R=halogen	containing group
R=other	containing group
R=oxalyl	this defines the  group. (21-oxalyl is also coded 4-carbons at 17; the sodium enolate is coded 21-double bond and 21-C-sub; R=O containing group).

The exo C-sub descriptor is not used. In these cases the radical that is substituted on the carbon chain is also coded as exo. Also C-sub is not employed at the 10, 13, 17, 20, 21, 22 and 23+ positions. For example:



Code 7-C-sub in 2-A and N-cont. group in 2-B (as well as "exo" N-R-R and primary amine)



Code 21-C-sub in 2-A and "halogen group" in 2-B. (Also code "exo" halo, chlorine, and 3 carbons at 17).

U5: CARBON CHAIN BONDED TO POSITION 17; GENERAL DESCRIPTORS

Notes on U5: Codes

1. The final digits of codes in Set U5: do not necessarily refer to position numbers on the steroid nucleus.
2. Steroids that are vitamin D derivatives or intermediates for vitamin D are coded V340.

Carbon chain bonded to position 17

U500	0 or 1 C atom (androstane)
U501	0 C atoms
U502	1 C atom
U520	2 C atoms (pregnane)
U521	position 21 is unsubstituted
U530	3 C atoms
U540	4 C atoms
U550	5 C atoms
U560	6 or more C atoms
U561	6 C atoms
U562	7 C atoms
U563	8 C atoms
U564	9 C atoms
U565	10 or more C atoms
U566	no 3-OR
U570	Bile acids (3-5)
U571	Bile (non-acids) (3-5)
U572	Isopregnane α C-C

Bis substituents (same)

U580	At C (17)
U581	At same C (not 17)

Note Bis subst. (Same) – this is used when a steroidal carbon is attached to two identical groups. It is not used to record bismethyl substitutions in the side chain as in the case of the 26 and 27 methyl groups of cholesterol nor is it used for hydrogen atoms or for symmetrical spiro rings (e.g. ketals). It is further divided into two specific descriptors.

At C(17) *and* *At same C (not 17)*

E.g. *lanosterol* is coded bis "at same C (not 17)" for the 4,4 bismethyl group.

Other general terms

U590	Addition
U591	Maleic adduct
U592	CNO, NCO, NCS, SCN
U593	21 diazo
U594	i compound

Note M General (Misc.)

Addition – this term designates addition compounds such as bisulfate addition procedures, hydrates, Grignard addition compounds, etc. (Also code under 2-A Miscellaneous when possible).

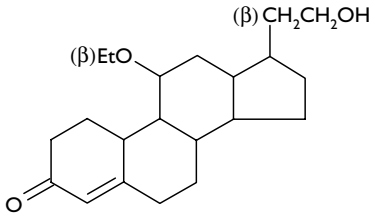
Maleic adducts – steroid reaction products of the type of maleic acid anhydride or ester adducts. Note that maleic adducts take precedence over all other codes and are only coded as such except for nuclear substituents. No other codes are used for the adduct.

CNO, NCO, NCS, SCN – self explanatory (also code in 2-A under "Miscellaneous")

21-Diazo – self explanatory (not coded 21-N-R-R or 21-Misc).

i-compounds – 3,5-cyclo, not considered as a fused compound.

Steroid Example

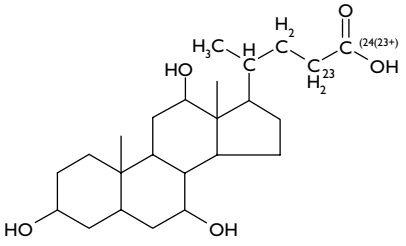


Position Number	Group	Code(s)
3	C=O	S603
4	C=C	S004
11	-OEt (β)	S311, S811
17	Side chain (β)	S317
21	-OH	S521

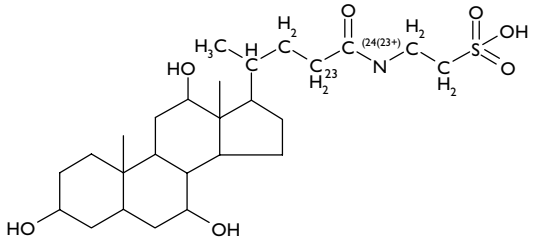
S132	Ring B is saturated
S133	Ring C is saturated
S134	Ring D is saturated
S142	Ring B is unsubstituted
U520	Carbon chain bonded to position 17 contains 2 C atoms
S830	Ether group bonded to position 11 is a hydrocarbon

Notes on the use of Exo

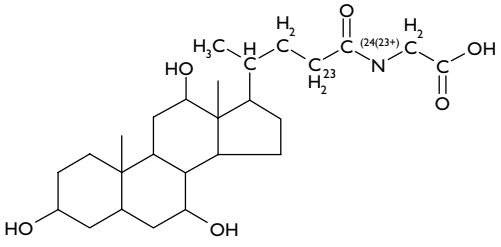
The designation of "Ex" represents substituents or pairs of double-bonded carbon atoms which are not directly connected through carbon-to-carbon linkages to the 1-23+ carbon atoms of the steroid nucleus. For example, the carboxyl substituent in cholic acid.



is directly connected, and is coded as 23+ COOR. The sulfo radical is taurocholic acid



and the carboxyl radical is glycholic acid



are not directly connected, and are coded as Ex-S-R and Ex-COOR respectively.

Note that in each of these compounds, the 2-A term 23+ COOR and the 2B term U333 -C(O), X=hetero are also coded to represent the carboxylic acid derivative substituents at the 23 positions.

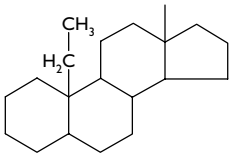
The designation "Ex" is *not coded* in the following 2A columns:

H ; α ; β ; Misc ;
CH_3 ; HC Ring ; HC Chain* ;
-C-Sub.

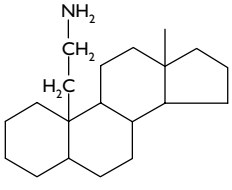
(*But note: alkinyl -C≡C- is coded as an "Ex" term, and when it is, HC Chain is also coded.)

Coding at 10, 13, 18 and 19

When the methyl groups at positions 10 and 13 are substituted the substituents are coded from positions 18 and 19 unless these positions are part of a ring in which case they are coded from positions 10 and 13.



is coded 19-methyl (not 10-HC chain)



is coded 19-C-sub (not 10-C-sub).

14 Part V: Natural Products; Polymers

14 Part V: - Natural products; polymers

Main Headings

V0: Antibiotics "A" through "I"
V1: Antibiotics "J" through "S"
V2: Antibiotics "T" through "Z"; Vaccines
V3: Vitamins
V4: Plant Extracts
V5: Microorganisms
V6: Animal Extracts
V7: Polymers; Proteins; Fats; Oils and Waxes; Miscellaneous Natural Products
V8: Enzymes; Coenzymes; Enzyme-related Terms
V9: Polypeptides

Introduction

Part V: codes refer to:

- natural products of unknown structure
- polymers
- polypeptides
- the following natural products, whether their structure is known or unknown: ajmalicine, antibiotics, atropine, coenzymes, gibberellins, lysergic acid, morphine, nucleosides, nucleotides, papaverine, prostaglandins, pyrethrins, chrysanthemum acids, quinidine, quinine, rotenone, tripeptides, vitamins, xanthines, and yohimbanes

Notes on Codes in Part V:

1. The natural products listed above in the Introduction are **structure-coded** in Parts A: through M: as with any other compound whose chemical structure is known. The construction of a search strategy that combines stucture codes and Part V: codes is discussed below in the section "Part V: Codes in the Standard Search Strategy". It may be best NOT to search for natural products of known chemical structure with Part V: codes, however, because the patentee may not note that the compound synthesized is a natural product. Such references will not be retrieved if Part V: codes are used in the search strategy. Instead, the Part V: codes could be used to narrow down the results of a structure-code search.

2. Part V: codes are applicable to compounds specifically mentioned in the code definitions and to all derivatives, unless a code definition specifically states that derivatives are not covered. ***Derivatives*** of natural products, polymers, and polypeptides are compounds that have the same basic structure as the parent compound, e.g. the same ring system. In the case of substances of unknown structure, derivatives include extracts and chemically modified forms, e.g. an acylated form.

3. Most codes in Part V: cover both the specified natural product and **similar synthetic compounds**. For example, V161 is assigned to both penicillin and ampicillin.

4. The indexing of **polymers** is discussed in the section "Coding Rules for Polymers and Halogens", which begins on page 18. Although Set V7: contains some polymer descriptors for patents classified in Derwent's Section B (Pharmaceutical), polymers have their own indexing system apart from the BCE Chemical Codes discussed in this manual. For information about Derwent's polymer indexing, contact your local Derwent office.

5. A **steroid** is only coded in Part V: if no structure is disclosed, or if the steroid is in a multi-component composition with at least one non-steroidal natural product.

6. From 1981 forward, **processes** by which new natural products of unknown structure are produced have been coded in Part N:.

Indexing the Source of Natural Products

Some natural products, e.g. proteins and polysaccharides, can be more precisely described by identifying their source of production. From 1981 forward, **codes marked with an** * may be assigned in combination with one or more of the following codes to indicate the source:

Plant source – V400, V403, V406

Microbial source – V500, V540, V550, V560, V570

Animal source – V600, V623, Subset V61:, Subset V63:, Subset V64:

The codes listed above are only assigned as supplementary source codes if the source is mentioned in the specification, or if it can be readily determined from reference books. If any of the codes listed above are assigned without a code marked with an *, the applicable generic code (V400, V500, or V600) is also assigned.

Part V: Codes in the Standard Search Strategy

The natural products listed in the Introduction are structure-coded when their chemical structure is known. The **structure** codes are searched with subheading M0 prior to 1970, and with subheading M2 from 1970 forward. The corresponding Part V: codes for these compounds are searched with subheading M0 prior to 1970, with subheading M1 from 1970 through 1980, and with subheading M2 from 1981 forward. Structure codes and Part V: codes can be searched together using the following search logic, where A represents all applicable structure codes separated by the appropriate LINK operators, and V represents all applicable Part V: codes separated by LINK operators:

Dialog (LINK = (S))

S M0= A LINK V	(Pre-1970)
S M2= A AND M1= V AND NOT M2=M903	(1970 - 1980)
S M2=(A LINK V LINK M903)	(1981 to present)

Orbit/Questel (LINK = L)

/M0 A LINK V
/M2 A AND V/M1 AND NOT M903
/M2 A LINK V LINK M903

STN (LINK = (P))

(A LINK V)/M0
A/M2 AND V/M1 AND NOT M903/M2
(A LINK V LINK M903)/M2

V0 – The Part V: Negation Code

Chemical structure searches may retrieve references to patents on natural products and their derivatives in cases when such references are not wanted. From 1981 forward, the **negation code V0** has been added to code subfields whose structural codes **refer only to natural products**. (Negation codes are discussed on page 108. Code subfields are discussed on page 28.) Negating V0 therefore eliminates records that contain the desired structure codes, but which pertain only to natural products.

The negation code V0 is NOT assigned by indexers if:

- the only Part V: codes assigned are from Subset V81:;
- there is any doubt that the structure being coded is a natural product or a closely related structure;
- the structure is very simple and could be overlooked by the user as being one of the natural products listed in the Introduction (e.g. some phosphonomycins and carotenes); or
- the code subfield also contains codes for structures that are not natural products.

Subheadings Applicable to Codes in Part V:

Unless otherwise specified, codes are applicable to M0 from 1963, M1 from 1970 and M2 from 1981.

V0:, V1:, V2: ANTIBIOTICS AND VACCINES

Notes on Selecting Codes for Antibiotics

1. An antibiotic is indexed according to the first letter of its generic name. For example, an antibiotic whose generic name begins with the letter **K** is assigned the code V110 ("K" antibiotics). However, some of the more important types of antibiotics are assigned their own code; for example, neomycin and derivatives are assigned the code V141 ("Neomycin"), not V140 ("N" antibiotics). Ampicillin is assigned the code V161 ("Penicillin") rather than V010 ("A" antibiotic), because ampicillin is a derivative of penicillin.
2. V001 is assigned to antibiotics that either do not have a name or that are represented by numbers and/or letters (e.g. "antibiotic 123B").

V000	Antibiotic - general disclosure
V001	Antibiotic with no generic name
V010	"A" antibiotic
V020	"B" antibiotic (excluding compounds coded V021)
V021	Bleomycin
V030	"C" antibiotic (excluding compounds coded V031, V032, V033, or V034)
V031	Cephalosporin
V032	Chloramphenicol
V033	Clavulanic acid
V034	Cycloheximide
V040	"D" antibiotic (excluding compounds coded V041)
V041	Daunomycin; daunorubicin
V050	"E" antibiotics (excluding compounds coded V051)
V051	Erythromycin
V060	"F" antibiotics
V070	"G" antibiotics (excluding compounds coded V071 or V072)
V071	Gentamycin
V072	Griseofulvin
V080	"H" antibiotics
V090	"I" antibiotics
V100	"J" antibiotics
V110	"K" antibiotics (excluding compounds coded V111)
V111	Kanamycin
V120	"L" antibiotics (excluding compounds coded V121 or V122)
V121	Leucomycin
V122	Lincomycin
V130	"M" antibiotics (excluding compounds coded V131)
V131	Mitomycin
V140	"N" antibiotics (excluding compounds coded V141, V142, or V143)
V141	Neomycin
V142	Novobiocin
V143	Nystatin
V150	"O" antibiotics
V160	"P" antibiotics (excluding compounds coded V161 or V162)
V161	Penicillin; ampicillin
V162	Polymyxin
V170	"Q" antibiotics
V180	"R" antibiotics (excluding compounds coded V181)
V181	Rifamycin
V190	"S" antibiotics (excluding compounds coded V191 or V192)
V191	Spiramycin
V192	Streptomycin
V200	"T" antibiotics (excluding compounds coded V201 or V202)

V201	Tetracycline
V202	Thienamycin
V210	"U" antibiotics
V220	"V" antibiotics
V230	"W" antibiotics
V240	"X" antibiotics
V250	"Y" antibiotics
V260	"Z" antibiotics

V27: Vaccines - antiviral

V270	Antiviral vaccines - general or unspecified (also for pre-1981 specifics)	
V271	Arbo	(1981)
V272	Entero	(1981)
V273	Hepatitis	(1981)
V274	Herpes	(1981)
V275	Interferon	(1981)
V276	Myxo	(1981)
V277	Pox	(1981)
V278	Rabies	(1981)
V279	Other specific antiviral vaccine	(1981)

V28: Vaccines - other than antiviral

V280	Vaccines - general or unspecified (also pre-1981 precursor for V282-V288)	
V281	Tuberculosis, leprosy	
V282	Typhoid, paratyphoid	(1981)
V283	Cholera	(1981)
V284	Plague	(1981)
V285	Tetanus	(1981)
V286	Diphtheria	(1981)
V287	Pertussis	(1981)
V288	Other specific vaccine that is not anti-viral	(1981)

V299: Mixture of antibiotics and/or vaccines

Notes on V299:

1. Prior to 1981, V299 was assigned to mixtures of two or more antibiotics that had the same "punch code", e.g. rosamicin and rondomycin. From 1981 forward, V299 has been assigned to all mixtures of two or more antibiotics and/or vaccines.
2. Antibiotic complexes, e.g. gentamycin, are usually treated as a single antibiotic rather than as a mixture.

V3: VITAMINS

Notes on V3:

1. Set V3: codes are assigned to compounds with vitamin activity and to compounds whose structure is derived from known vitamins.
2. Not considered as vitamins in Set V3: are the following: nicotinic acid, pantothenic acid, folic acid, choline, inositol, biotin, p-aminobenzoic acid, linoleic acid, and other unsaturated aliphatic acids.
3. Prior to 1981, V399 was assigned to mixtures of two or more vitamins that had the same "punch code", e.g. a mixture of Vitamin A and a carotene. From 1981 forward, V399 has been assigned to all mixtures containing two or more vitamins.

V300	Vitamin - general or unspecified
V310	Vitamin A, carotenoids
V321	Vitamin B1 (e.g. thiamin, aneurin)
V322	Vitamin B2 (e.g. riboflavin)
V323	Vitamin B6 (e.g. pyridoxine)
V324	Vitamin B12, cobalamin
V330	Vitamin C (e.g. ascorbic acid)
V340	Vitamin D (e.g. calciferol including provitamin D)
V350	Vitamin E, tocopherol
V360	Vitamin K
V370	Vitamin P; other specific vitamin
V399	Mixture of vitamins

V4: PLANT EXTRACTS

Note Alkaloids of known structure that cannot be classified in one of the Subsets V41: through V46:, e.g. berberine or strychnine, are coded by structure only, not in Part V:.

V40: Non-alkaloid plant extracts

V400	Non-alkaloid plant extract - general
V401	Derris, rotenone
V402	Glycoside, saponin
V403	Mushroom, toadstool
V404	Peat; straw; whole plant; humic acid; silage; humus; compost; moss; sawdust; tobacco; seaweed; lignite; wood; bark; seed
V405	Pyrethrins; allethrins; chrysanthemum acids
V406	Other specific, non-alkaloid plant extract

Note V402 covers compounds of unknown structure and, from 1981 forward, steroid glycosides.

V41: Belladonna alkaloids

V410	Belladonna alkaloids, excluding those coded V411
V411	Tropane, scopolamine

V42: Cinchona alkaloids

V420	Cinchona alkaloids, excluding those coded V421
V421	Quinine; quinidine

V43: Ergot alkaloids

V430	Ergot alkaloids, excluding those coded V431
V431	Lysergic acid

V44: Opium alkaloids

V440	Opium alkaloids (e.g. papaverine), excluding those coded V441
V441	Morphine, codeine, thebaine

V45: Rauwolfia alkaloids

V450	Rauwolfia alkaloids (e.g. ajmalicine, ajmaline, and reserpiline), excluding those coded V451
V451	Yohimbanes (e.g. reserpine and rescinnamine)

V460: Xanthines, including caffeine and theobromine

V470: Other alkaloids, structure unknown

V5: MICROORGANISMS

Note Each code in Set V5: covers the extracts and metabolites of the microorganisms for which they are defined, unless the extracts and/or metabolites can be more specifically described elsewhere. For example, an antibiotic produced by a microorganism is coded in Set V2: ("Antibiotics and Vaccines") rather than in Set V5:.

V500	Microorganism - general or unspecified
V510	Gibberellins
V520	Growth stimulant for microorganisms; sideramines
V530	Microbial growth stimulant for animals (e.g. lactobacillus)
V540	Bacteria
V550	Fungi; algae (e.g. Streptomyces; yeast)
V560	Virus
V570	Other specific microorganism

V6: ANIMAL EXTRACTS

V600: Animal extracts - general

V61: Blood and derivatives

V610	Blood and derivatives - general (Pre-1981 precursor code for V612 through V616)	(1963-1981)
V611	Antibody (excluding vaccines, which are coded in Subset V27: or V28:)	
V612	Blood cell (i.e. erythrocyte, etc.)	(1981)
V613	Blood factor	(1981)
V614	Serum; plasma	(1981)
V615	Whole blood	(1981)
V616	Other blood derivative (including haemoglobin)	(1981)

Note Cells other than blood cells are coded V754.

V62: Hormones, prostaglandin, gland extracts

V621	Insulin or derivative	(1973)
V622	Prostaglandin	
V623	Non-hormonal gland extract	
V624	Other hormone of unknown structure, including endocrine gland extract	

Note V622 has only been assigned to prostaglandins from 1973 forward.

V63: Other body secretions or products

V631	Milk, including synthetic milk
V632	Urine
V633	Faeces
V634	Other body secretion or product, excluding those covered in Subset V62:.

V64: Other animal extracts

V641	Body factors (excluding blood factors, which are coded V613)
V642	Bone; hair; nail; etc.
V643	Heart, kidney, liver, etc.
V644	Other mammalian extract
V645	Other non-mammalian extract

V7: POLYMERS; OTHER NATURAL PRODUCTS

Notes on V7: Codes

1. Set V7: codes cover all natural polymers; polymers involved in pharmaceutical, veterinary, or medical inventions; and all natural products that cannot be classified elsewhere in Part V:.
2. Polymeric dyes are coded in Subset W31:.

V71: Cellulose and derivatives

V711	Cellulose, unmodified
V712	Cellulose ester
V713	Cellulose ether
V714	Other cellulose derivative

Note The modifying group in a cellulose derivative coded V712, V713, or V714 is additionally structure-coded in Parts A: through M:.

V72: Starch; dextran

V721	Dextran or derivative
V722	Starch, modified
V723	Starch, unmodified

Note Modifying groups on dextran or starch are additionally structure-coded in Parts A: through M:.

V73: Other polysaccharides

V731	Chondroitin or derivative
V732	Heparin or derivative
V733	Alginic acid or derivative
V734	Algar or derivative
V735*	Other polysaccharides

Notes on V73: Codes

1. Polysaccharides of known structure that are coded V735 also receive codes from Subsets L81: ("Type of configuration of sugar or residue") and L83: ("Type of sugar compound or residue"), and the negation code L8 ("Sugar is essentially present").
2. Polysaccharides of unknown structure that are coded V735 are additionally coded L810 ("Unspecified sugar or residue configuration").

V74: Other polymers

V741*	Other natural polymers (e.g. lignin)
V742	Acidic or basic synthetic polymer
V743	Non-ionic synthetic polymer (e.g. PVC)

Note Synthetic polymers coded V742 or V743 are assigned additional chemical codes in Parts A: through M:.

V75: Proteins; nucleic acid; cells

V751	Gelatin
V752*	Other protein
V753*	Nucleic acid
V754*	Cells

Note Blood cells are coded V612, not V754.

V76: Nucleoside; nucleotide

V761*	Nucleoside
V762*	Nucleotide

V77: Phospholipid; fat; lipid

V771*	Phospholipid
V772*	Fat; lipids (excluding phospholipids)

V780*: Oil; wax (excluding fats and lipids coded in Subset V77:)

V79: Other natural products

V791*	Non-microbial antigen	(1981)
V792*	Dye of unknown structure	
V793	Mineral	
V794	Monosaccharide, disaccharide	
V795*	Steroid of unknown structure	
V796*	Terpene of unknown structure	
V797	Other natural product	

Notes on V79: Codes

1. Prior to 1981, non-microbial antigens were only coded according to their sources. For example, antigenic pollen was coded V406 ("Other non-alkaloid plant extracts").
2. V792 is assigned to dyes of unknown structure and to natural dyes of known structure that are indexed as general chemicals (Chemdoc - Section E).
3. V794 is assigned to sugars of unknown structure and to sugars that are ingredients in multicomponent mixtures.
4. Prior to 1970, V795 was additionally assigned to steroids of known structure in multicomponent mixtures with one or more non-steroid compounds. From 1970 forward, at least one of the non-steroid components in the mixture had to be a natural product for V795 to be assigned.

V8: ENZYMES; COENZYMES; ENZYME-RELATED TERMS

V80: General enzyme descriptors

V800	Enzyme; coenzyme; proenzyme; zymogen (Pre-1981 precursor code for V801, V802, V804, and V805)	(1963-1981)
V801	Coenzyme (e.g. ubiquinones)	(1981)
V802*	Enzyme	(1981)
V803*	Enzyme inhibitor (polymeric or unknown structure)	(1981)
V804	Proenzyme	(1981)
V805	Zymogen	(1981)

Notes on V80: Codes

1. The codes V802, V803, and P616 ("Enzyme inhibitor") are used in combination with codes from Subset V81: ("Type of enzyme or enzyme inhibitor").
2. V803 is only used for polymers and compounds of unknown structure. Other enzyme inhibitors are coded P616 ("Non-polymeric enzyme inhibitor of known structure").

V81: Type of enzyme or enzyme inhibitor

V810	General or unspecified type of enzyme or enzyme inhibitor	(1981)
V811	Oxidoreductase	(1981)
V812	Transferase	(1981)
V813	Esterase	(1981)
V814	Peptide hydrolase	(1981)
V815	Hydrolase (other than those coded V813 or V814)	(1981)
V816	Lyase	(1981)
V817	Isomerase	(1981)
V818	Synthetase	(1981)

Note Codes in Subset V81: are used in combination with the codes V802 ("Enzyme"), V803 ("Enzyme inhibitor of polymeric or unknown structure"), and P616 ("Non-polymeric enzyme inhibitor of known structure").

V9: POLYPEPTIDES

Notes on V9: Codes

1. Set V9: codes are not used for dipeptides.
2. If an amino acid contains 1-3 alpha amino acid units, it is indexed as a non-polymeric compound, and is structure coded in Parts D: through M:. Furthermore, from 1981 forward a 3-unit amino acid chain receives applicable codes from Part V: in a code subfield under the subheading M2. (Subheadings are discussed on page 23). Prior to 1981 tripeptides were coded in subheadings M1 and M2 (i.e. as a standard chemical (M2) and a peptide (M1))
3. If an amino acid chain consists of 4 or more alpha amino acids, the structure is indexed as a polypeptide, and structure codes from Parts D: through M: for the individual amino acid units are listed in a single code subfield under the subheading M1. The amide linkages of the polypeptides, however, are not coded.
4. Codes in Sets M1:, M2:, and M3: have only been assigned to polypeptides since 1981.
5. The "Basic Group" code for polypeptides is M423 ("Other natural product or polymer").
6. From 1981 forward, negation codes from Parts H: through L: have been assigned to polypeptides. (Negation codes are discussed on page 108). These are applied to the polypeptide as a whole – excluding the amide linking groups.

V90: Polypeptide source

V901*	Natural
V902	Synthetic or semisynthetic

Note A patent that discloses a synthesis of a polypeptide is coded V902, not V901, even if the polypeptide is obtainable from a natural source.

V91: Number of amino acid units

V911	3 - 5	(1981)
V912	6 - 10	(1981)
V913	11 - 15	(1981)
V914	16 - 20	(1981)
V915	21 - 25	(1981)
V916	26 - 30	(1981)
V917	31 or more	(1981)

V92: Polypeptide structural type

V921	Linear	(1981)
V922	Cyclic with S-S bond	(1981)
V923	Cyclic with other bond	(1981)
V924	α-Amino acid unit replaced by a non α-amino acid unit	(1981)
V925	Polypeptide with a D-α-amino acid unit	(1981)

15 Part W: Dyes; Dyed Materials: Dyeing Processes

15 Part W: - dyes; dyed materials; dyeing processes

Main Headings

W0: General Dye Descriptors
W1: Azo Dyes
W2: Reactive Dyes
W3: Dye Types
W4: Pigment or Dye Treatment
W5: Dyeing Processes and Materials

Introduction

Codes in Part W: are assigned to organic compounds used as colouring matter, including dyes, pigments, immediate precursors of dyes and pigments, fluoroescent brighteners, and other similar compounds. Part W: codes also cover dyeing processes and materials treated.

Notes on Codes in Part W:

1. In addition to the special dye codes of Part W:, dyes are assigned structure codes from Parts A: through M:, and non-structure codes from other code Parts. All codes for dyes are searched with the subheading M4; the different code Parts must be consulted to determine which codes are applicable to M4, and hence, to dyes.

2. Immediate **precursors of dyes or pigments** receive dye or pigment codes from Part W: in combination with either Q317 ("Oxidative dye precursor") or Q318 ("Non-oxidative dye precursor") as applicable. Examples of precursors include: oxidation bases (Q317); leuco dyes from triarylmethane (Q318); vat, sulphur, and phthalein dyes (Q318); cryochromic, halochromic, photochromic, piezochromic, and thermochromic compounds (Q318); and diazonium compounds (Q318).

3. **Couplers** for azo dyes and couplers for oxidation bases are assigned the code Q311 and an applicable code from Subset W12:.

4. **Natural dyes of unknown structure** are additionally coded V792 ("Dye of unknown structure") and M423 ("Other natural product").

5. Fluorescent-coloured compounds are coded in Part W: as dyes or pigments, and additionally Q613 ("Optical bleach; fluorescent"). Indicators, liquid crystals, fluorescent compounds other than brighteners, pearlescent pigments, white pigments, photoconductors, scintillators, and other similar compounds are **not coded in Part W:**.

6. Conventional, non-essential counter ions in **salts** of ionic dyes or diazonium compounds are not coded, whether organic or inorganic.

7. Examples illustrating the use of Part W: codes are presented at the end of this chapter.

Conventions for Selecting Dye Codes

1. The dyes and dye precursors listed in the first column below are coded as the **preferred form** listed in the second column.

These dyes or precursors:	Are coded in this form:
Keto-hydrazone dyes	Azo-phenol
Quinone-diazide precursors	Diazonium-phenoxide
Tri-aryl methane dyes	Carbonium
Xanthene, oxazine dyes	Oxonium
Thiazine dyes	Sulphonium
Phenazine, acridine dyes	Form having an ionic charge on the ring-N

Note Prior to 1981, tri-aryl methane dyes in pharmaceutical (Farmdoc - Section B) or agricultural (Agdoc – Section C) patents were coded as quinone derivatives

2. Phthalein and xanthene **dyes** are coded in the "ring-open" form, but phthalein, xanthene, spiro-pyran, and related dye **precursors** are coded in the "ring-closed" form.

3. Ionic cyanine dyes with one heterocyclic ring bonded to each end of a conjugated chain are coded first with the charge on one ring and then with the charge on the other ring. If there is only one heterocyclic ring, then it is assumed that the charge is on that ring. The ionic charge alters the hydrogenation state of the heterocyclic ring on which it is assumed to exist, which may in turn affect the structure-coding of that ring.

4. Codes for metal-complex dyes are divided into different code subfields based on the structure of the dye. (Code subfields are discussed on page 24.) Compounds with the structure (**dye1**) - **metal** - (**dye 2**) are coded with (**dye1**)-**metal** codes in one field and **metal**-(**dye2**) codes in another field. Similarly, compounds with the structure (**dye**) - **metal** - (**non-dye**) are coded with (**dye**)-**metal** codes in one field and **metal**-(**non-dye**) codes in another field.

Subheadings Applicable to Part W codes

Unless otherwise specified, codes are applicable to Subs M4 from the beginning of the service, 1970.

W0: GENERAL DYE DESCRIPTORS

W00: Generic terms for dyes and pigments

W001	Dye (Novel dye or dye production)
W002	Metallised dye or pigment (metal complex) (1981)
W003	Non-metallised dye or pigment
W004	Dye residue - general
W005	Reactive group - general

Notes on W00: Codes

1. The following types of compounds are not assigned the code W001: pigments (W334), fluorescent brighteners (Q613), couplers (Q311), dye precursors (Q317 or Q318), and dye intermediates (Q316).
2. W002 is assigned to all dyes containing **metal(s) complexed in a chelate ring**, but not to simple metal salts. W003 is assigned to all dyes except those containing metal(s) complexed in a chelate ring, including **simple metal salts**.
3. W004 and W005 are only assigned when there is **insufficient information** for detailed structure indexing, or when a wide range of dyes are disclosed. In the latter case, all specific structures disclosed are also indexed. W004 and W005 are primarily useful for exhaustive searches on a specific dye.

W01: Substituents on anthraquinone

W011	0 or 1
W012	2
W013	3
W014	4
W015	5
W016	6 or more

Notes on W01: Codes

1. Subset W01: codes are assigned separately to each unfused anthraquinone ring system present in a structure.
2. Substituents coded in Subset W01: may be any atom or group except H, and may be bonded to any of the anthraquinone position(s) 1 through 8. The substituents may not be linked to one another outside of the anthraquinone ring.

W020 Extended quinone

Note Extended quinones are similar to the quinonoid structures shown on page 165, with E and E' always bonded to different rings within a linear series of 6-membered, fully conjugated rings. E and E' may be the same or different atoms, and may be bonded to further atoms. See the example on page 276.

W03: Sulphonic acid groups present

W030	None
W031	1
W032	2
W033	3
W034	4 or more

W04: Diazonium groups present

W041	One
W042	Two or more

Note Compounds coded in Subset W04: include stabilised diazonium groups, e.g. -N=N-ONa or -N=N-SO_2ONa.

W1: AZO DYES

W11: Azo type

W111	Mono
W112	Dis (A→M→E)
W113	Dis (E←D→E)
W114	Dis (A→Z←A')
W115	Dis, other configuration
W116	Tris or poly azo

Notes on W11: Codes

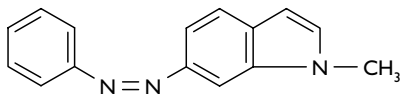
1. A and A' are diazo components (amino), M is middle component (amino and coupling), E is coupling (end) component, D is tetraazo component (diamino), Z is a "double" coupling component, and "→" means "diazotised and coupled to".
2. If insufficient information is given about a dis-azo dye to assign W112, W113, or W114, then W115 is assigned. Also coded W115 are condensed dis-azo dyes. When constructing an exhaustive search on a specific dis-azo dye type, W115 should be 'OR'ed in the search strategy with any other code(s) in Subset W11: that may be applicable.

W12: Coupling components in azo dyes

W120	Coupling component - unspecified
W121	Amino
W122	Phenol
W123	Aminophenol
W124	Other heterocyclic coupling component (including reactive CH_2)
W125	Other specified coupling component (including aliphatic reactive CH_2)

Notes on W12: Codes

1. Only substituents on rings directly bonded to azo or diazo groups are considered in this Subset; substituents on other rings in the same ring system are not considered when selecting an applicable code.
2. The amino group in the code definitions of W121 and W123 may be a ring tertiary amine. In such cases, the ring with the amine is fused to the ring that is bonded to an azo or diazo group, as in the structure shown below.



3. W123 is assigned to two types of azo structures: those with amino and phenol substituents on the same ring; and those with amino and phenol substituents on different rings in the same ring system, each ring being bonded to a different azo or diazo group.
4. Several examples of structures that are assigned codes in Subset W12: are given at the end of this chapter.

W13: Diazo components in azo dyes

W130	Diazo component - unspecified
W131	Carbocyclic ring
W132	Other specified diazo component (including heterocyclic ring)

Notes on W13: Codes

1. Only the ring(s) directly bonded to azo or diazo groups are considered in this Subset; other rings in the same ring system do not affect the selection of applicable codes.
2. Several examples of structures that are assigned codes in Subset W13: are given at the end of this chapter.

W2: REACTIVE DYE

Note Each reactive centre is coded separately in Set W2:. Reactive groups present at different reactive centres are not added together when selecting codes in this Set.

W21: Reactive centres present

W211	One
W212	Two or more

W22: W23: - Heterocyclic or activated aromatic system

W22: Number of reactive halogen groups

W221	One
W222	Two or more

W23: Number of other types of reactive groups

W231	One
W232	Two or more

W24:, W25:, W26: Systems other than heterocyclic or activated aromatic

W24: Number of reactive halogen groups

W241	One
W242	Two or more

W251: One or more reactive C=C or C≡C groups

W26: Number of reactive groups other than halogen, C=C, or C≡C

W261	One
W262	Two or more

W3: DYE TYPES

W31: Polymeric; polymerisable; natural; steroidal

W311	Polymeric dye
W312	Polymerisable dye that bonds with substrate
W313	Polymerisable dye that does not bond with substrate
W314	Natural dye of known structure; steroidal dye

Notes on W31: Codes

1. W314 covers natural dyes of known structure, compounds identical to natural dyes that are synthesized, and their derivatives. From 1981 forward, steroidal dyes have also been coded W314 in subheading M5.
2. Dyes produced from natural sources are additionally coded N161 ("Extraction from natural materials").

W32: Type of dye - chemical descriptors

W321	Anionic	(1981)
W322	Azoic (coupled on fibre)	
W323	Basic (cationic)	
W324	Mordant; ingrain (excluding compounds coded W322, W328, or W329)	(1981)
W325	Nitro; nitroso	
W326	Phthalocyanine or other similar macrocyclic compound	
W327	Reactive	
W328	Sulphur	
W329	Vat	

Notes on W32: Codes

1. W325 is only assigned to dyes whose main chromophore is nitro or nitroso.
2. Chlorophyll, haemin, corrin, and their derivatives are examples of compounds coded W326.

W33: Type of dye - other descriptors; pigment

W331	Fugitive or sighting dye
W332	Liquid concentrate
W333	Soluble in organic solvent (1981)
W334	Pigment
W335	Water insoluble (disperse)
W336	Water soluble
W339	Other type of dye (1981)

Notes on W33: Codes

1. Oxidation dyes are coded in Subset Q31:.
2. To be coded W333, a dye's solubility in organic solvents must be specifically mentioned.
3. Water insoluble salts of water soluble acidic or basic dyes are coded W334.

W4: PIGMENT OR DYE TREATMENT

W410	Pigment treatment
W420	Dye treatment

W5: DYEING PROCESSES AND MATERIALS

W51: Substrate treatment

W511	Treatment before dyeing
W512	Treatment during dyeing
W513	Treatment after dyeing

W52: Natural material (or derivative) treated

W521	Cellulose
W522	Cellulose acetate
W523	Other cellulose derivative
W524	Paper; wood
W525	Leather
W526	Wool; silk; hair; fur
W529	Other natural material

W53: Synthetic material treated

W530	Unspecified synthetic material (1981)
W531	Polyacrylonitrile, other similar compounds
W532	Polyamide; polurethane
W533	Polyester (including polycarbonate)
W534	Polyolefin; polyhydrocarbon
W535	Substituted polyolefin (excluding compounds coded W531)
W539	Other specified synthetic material

Note W531 includes all polymers prepared from -C≡N containing monomers, e.g. poly(vinylidene cyanide). All other acrylics are coded W535.

W54: Colouring processes – general descriptors

W541	Dyeing
W542	Mass colouring
W543	Printing

Notes on W54: Codes

1. The codes in Subset W54: can be used in combination with any applicable codes in Subset W55:.
2. W542 is used for mass colouration with compounds that are either soluble or insoluble in the substrate.

W55: Colouring processes – specific descriptors	
W551	Assisted
W552	Cross-linking and colouring
W553	Hot transfer (dyeing/printing)
W554	Resin and colour
W555	Resist; special printing paste
W556	Non-aqueous solvent
W557	Union (one)
W558	Union (two or more)
W559	Other colouring process

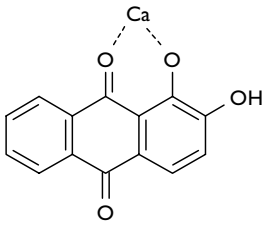
Notes on W55: Codes

1. W557 is assigned if the use of a dye for union dyeing is stated, but only the name of one fibre type is mentioned.
2. W559 is not used for sulphur dyeing processes (coded W328) or vat dyeing processes (coded W329).

Examples Using Part W: Codes

W002 – Metallised

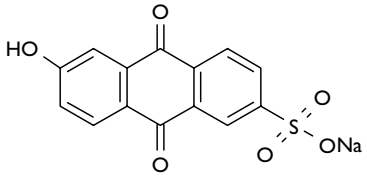
a



W002 (but not W003)

W003 – Non-metallised

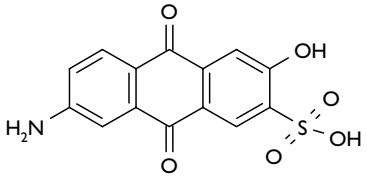
a



W003

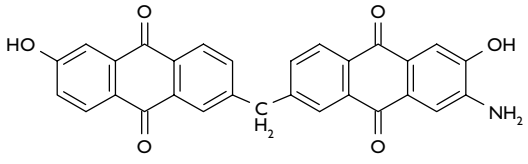
W01 – Substituents on anthraquinones

a



W013

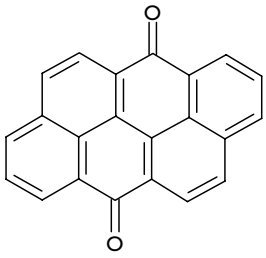
b



W012, W013 (but not W015)

W02 – Extended quinone

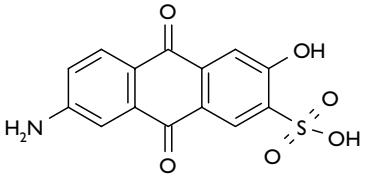
a



W020

W03 – Sulphonic acid groups present

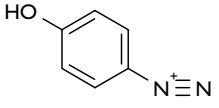
a



W032

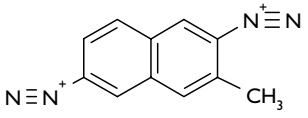
W04 – Diazonium groups present

a



W041

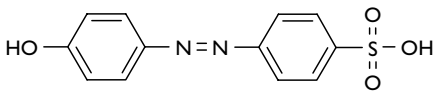
b



W042

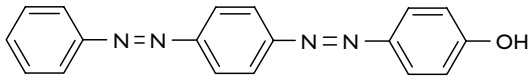
W11 – Azo type

a



W111

b

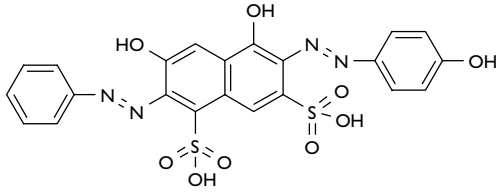


W112

c

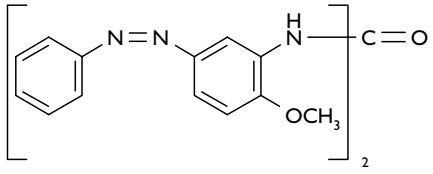
W113

d

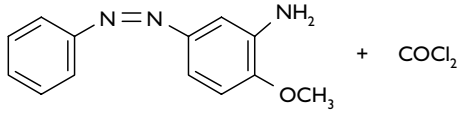


W114

e

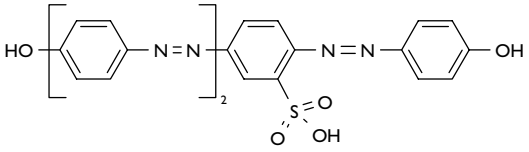


prepared from



W115

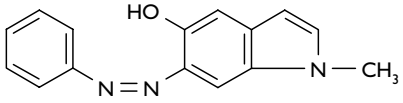
f



W116

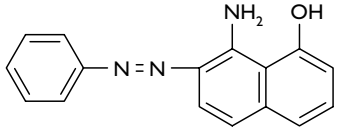
W12: and W13: Coupling and diazo components in azo dyes

a



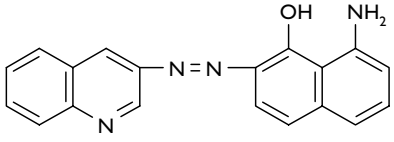
W131, W123

b



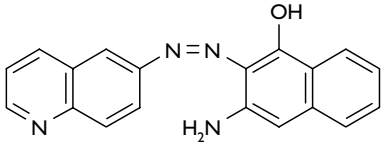
W131, W121 (but not W123)

c



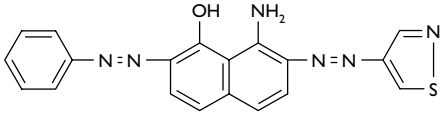
W132, W122 (but not W123)

d



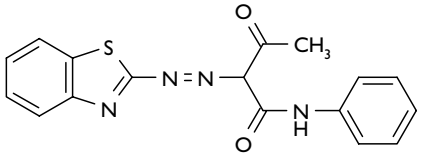
W131, W123

e

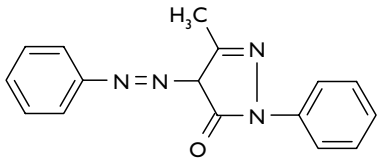


W131, W123, W132 (but not W121 or W122)

f

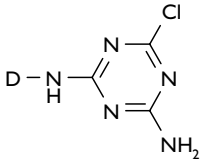


W132, W125

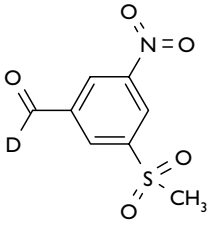
g 

W131, W124

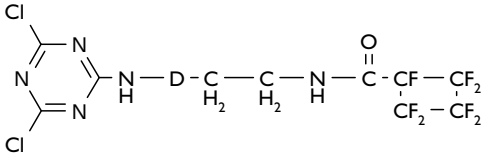
W2 – Reactive dyes

h 

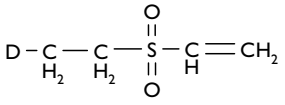
W211, W221

i 

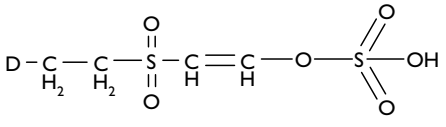
W211, W231

j 

W222, W212, W242

k 

or



W211, W251