Food and Drug Administration

FDA GLOBAL SUBSTANCE REGISTRATION SYSTEM (FDA GSRS)

RELEASE NUMBER: 2.1

ENTERPRISE PERFORMANCE LIFECYCLE

USER MANUAL

Document Version Number: 2.1

Document Version Date: 4/2/2018

VERSION HISTORY

Version Number	Implemented By	Revision Date	Approved By	Approval Date	Description of Change
1.0	Frank Switzer	11/27/2016			Initial Draft
1.1	Sarah Stemann	10/20/2017			Updated for v2
2.1	Sarah Stemann/Frank Switzer	4/2/2018			Updated for v2.1 and registration (chemical)

Table of Contents

1.

PURPOSE AND SCOPE

The purpose of the Global Substance Registration System (GSRS) is to unambiguously and consistently identify all substances which may be present in regulated products or substances produced metabolically from these substances or interacting with these substances in biological systems. To accomplish this, each substance entered into the GSRS is assigned a strong, non-proprietary identifier known as a Unique Ingredient Identifier (UNII) that is permanently associated with the defined substance. GSRS provides UNIIs freely for general use and specifically for electronic ingredient listing activities. This data is available to FDA users and systems via a web client and REST and JAVA APIs. The public portion of this data is available for search and download at https://fdasis.nlm.nih.gov/srs/srs.jsp

The scope of regulated products covered by GSRS is as follows:

Foods
: Specific foods or components of food, regardless of whether the food is in conventional food form or a dietary supplement, such as vitamins, minerals, herbs or other nutritional substances.

Drugs
: Both active and inactive ingredients used in drug products, including those for veterinary purposes.

Biologics
: Both active and inactive ingredients used in biologics, such as blood products, therapeutic products, vaccines, cellular and gene therapy products, allergenic products and tissues.

Devices
: Components of devices including, for example, silicon for implants and chemical reagents for glucose test kits.

Cosmetics
: Components of cosmetic products, such as flavors, fragrances, colorants, vitamins, plant and animal derived ingredients and polymers.

Tobacco
: Components of tobacco products, such as flavors, fragrances, colorants, vitamins, plant and animal derived ingredients and polymers.

2. CONSIDERATIONS

The process of registering substances in the GSRS is captured by detailed business rules developed by the SRS Project Management Team located in FDA's Office of Health Information with the Office of the Chief Scientist and overseen by the SRS Program Review and Technical Boards. These boards work closely with multiple global stakeholders in the development of these rules.

To use effectively use GSRS, users should understand how GSRS:

Categorizes substances (principles)

Defines substances (identities)

Associates data with substances (metadata)

Links substance information from other FDA systems (interoperability)

2.1. *Principles*

A substance is any matter (has mass and occupies space) that has discrete existence. Substances are defined based on what they are and not how they are made or used. They are defined independent of grade or level of purity.

Broadly, substances can be divided into three major categories. Classifying a substance by category determines which attributes are necessary to fully identify it in GSRS.

- Monodisperse: single molecular connectivity including:
  - Chemicals aka small molecules.
  - Proteins: defined sequence chains of amino acid residues connected by peptide bonds.
  - Nucleic acids: defined sequence chains of nucleotide residues connected by phosphate esters.
- Polydisperse: collection of multiple but related molecular connectivities including:
  - Polymers: chemical entities that are related due to being based on one or more structural repeating units.
  - Structurally-diverse: entities that are related due to originating from a single well-defined biological or mineral source.

Together, monodisperse and polydisperse substances account for all single substances.

- Mixture: limited (<50) set of synthetic or naturally-occurring single substances that are synthesized and/or isolated together (does not include substances synthesized separately and then recombined).

2.2. *Identities*

The guiding principle of the GSRS is to limit ambiguity, which implies uniqueness and internal consistency of representations.

In the legacy SRS system, a substance was identified using two fields: Structure and an XML Identifying description. In GSRS, a JSON schema is used but the data management principles are the same.

The primary identifying attribute of a chemical substance is a two-dimensional molfile (.mol) format array. This non-proprietary data format is compatible with chemical search engines, may be faithfully rendered as an image showing molecular connectivity and stereochemistry or may be converted to other common chemical data formats.

GSRS uses precise stereochemistry when identifying chemical substances distinguishing ABSOLUTE, RACEMIC, ACHIRAL, UNKNOWN, EPIMERIC, MIXED and other less common isomeric types.

Many substances cannot be described completely or at all as a single molfile structure or sets of molfile structures so other forms of representation are also necessary. Fielded controlled vocabulary is used for substances that are not easily defined using molfile structures. Data captured and organized using the JSON schema is used either to add additional information about a substance or to describe substances for which a structure is not drawn. Different types of substances, such as polymers and proteins have specific attributes in the schema used to describe them.

3. USER MANUAL

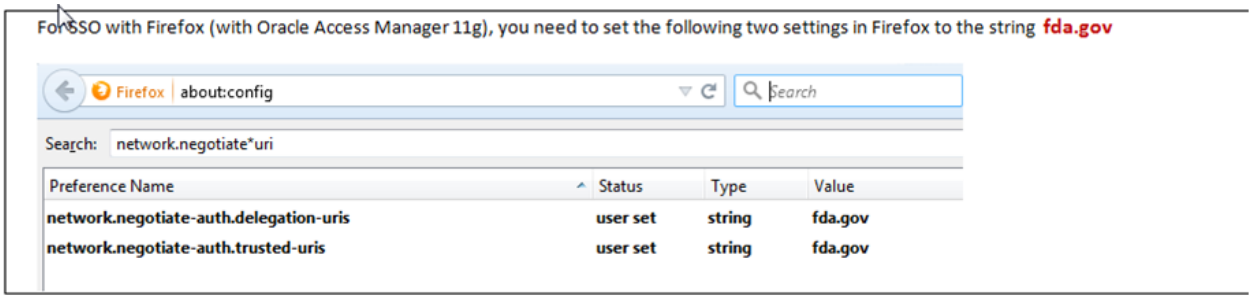
3.1. *Access*

To access FDA GSRS use an FDA supported browser to navigate to:

Pre-prod (testing and UAT): http://gsrs.preprod.fda.gov/ginas/app

Prod: http://gsrs.fda.gov/ginas/app

Google Chrome is the preferred browser for use with GRSS. Firefox also works well with the changes below:

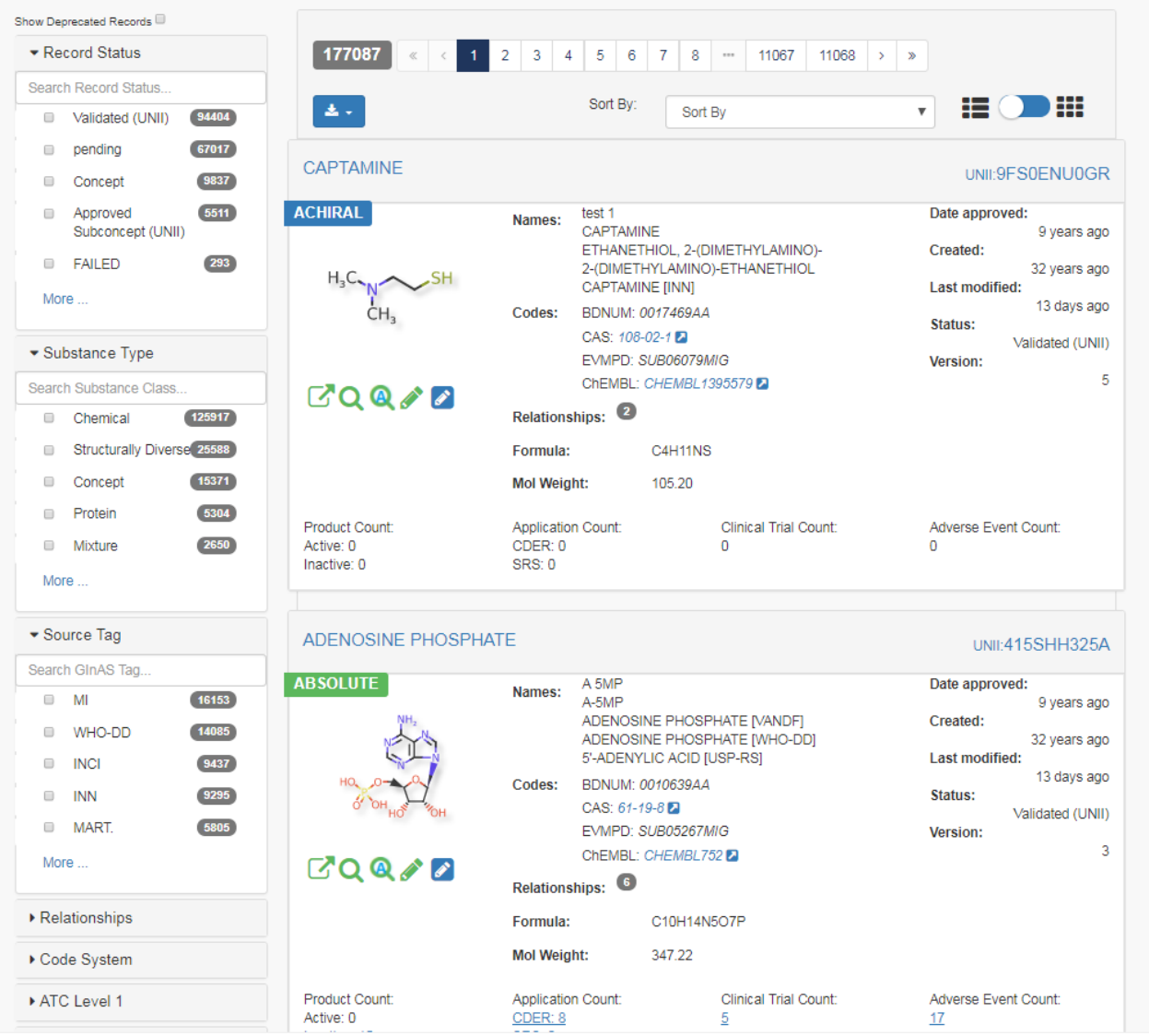


3.2. *Browse*

FDA GSRS supports a browsing interface that provides a high-level overview of all the records contained within the system while still allowing a user to filter and drill down to specific records as desired.

To navigate to the browser interface, click on the Browse Substances tab on the navigation pane at the top of any page on the FDA GSRS website.

Browsing substances in the FDA GSRS is conducted via a dashboard interface. The content of the Browse Substances page looks like this:



The dashboard can be divided into three major components:

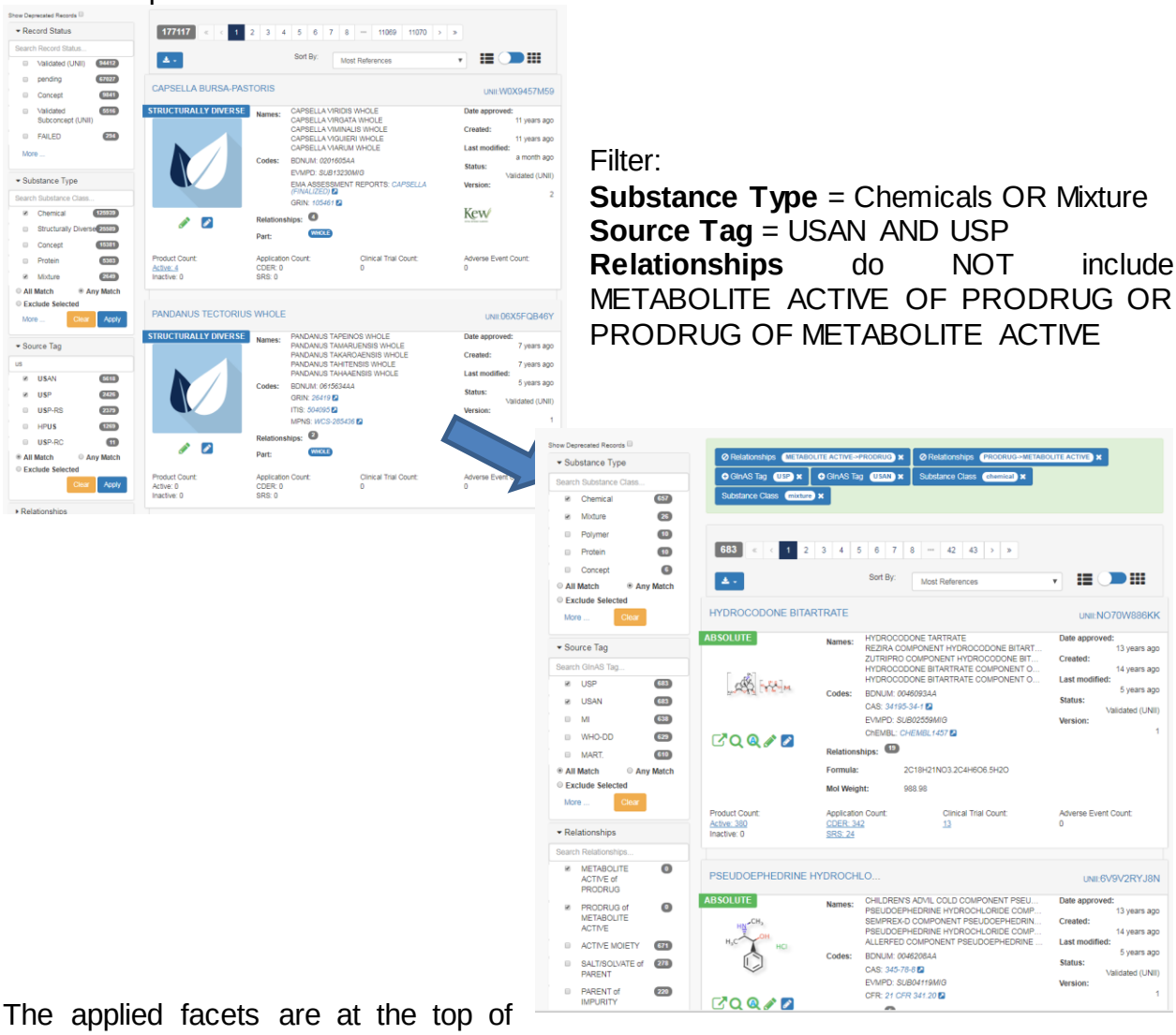
- Filter/Facets pane
- Display (header/footer) pane
- Results pane

3.2.1. Filter Pane

The filter pane is located on the far left of the dashboard. Here, the user can see the number of substance records stored in the FDA GSRS database based on a select number of facets. This count data is displayed in the pill-shaped field found to the right of each category name.

Under each facet title is the " Search [Facet name]…" box. Enter the facet name (full or partial) and search for the facet category within the facet. To the left of each facet category name is a checkbox. At the bottom of the displayed list is a link to "More…". This will expand the list to display the top 20 options. Use the search to filter for options that are not explicitly displayed.

After making your facet selections Boolean options are displayed. Exclude, All Match or Any Match may be selected. By default, the filters are applied as “Any Match”. To display facets that fulfill all criteria, select “All Match”. Click “Apply” to filter based on the selected options (for one or more filters at once). The filtered records are displayed in the results pane located at the center of the dashboard.



Filter:
Substance Type = Chemicals OR Mixture
Source Tag = USAN AND USP
Relationships do NOT include METABOLITE ACTIVE OF PRODRUG OR PRODRUG OF METABOLITE ACTIVE

The applied facets are at the top of the display pane, and can be removed which will update the data set. Facets with the icon indicate selections that are excluded from the results. Facets with the icon indicate selections where one or more criteria must be met by the substance.

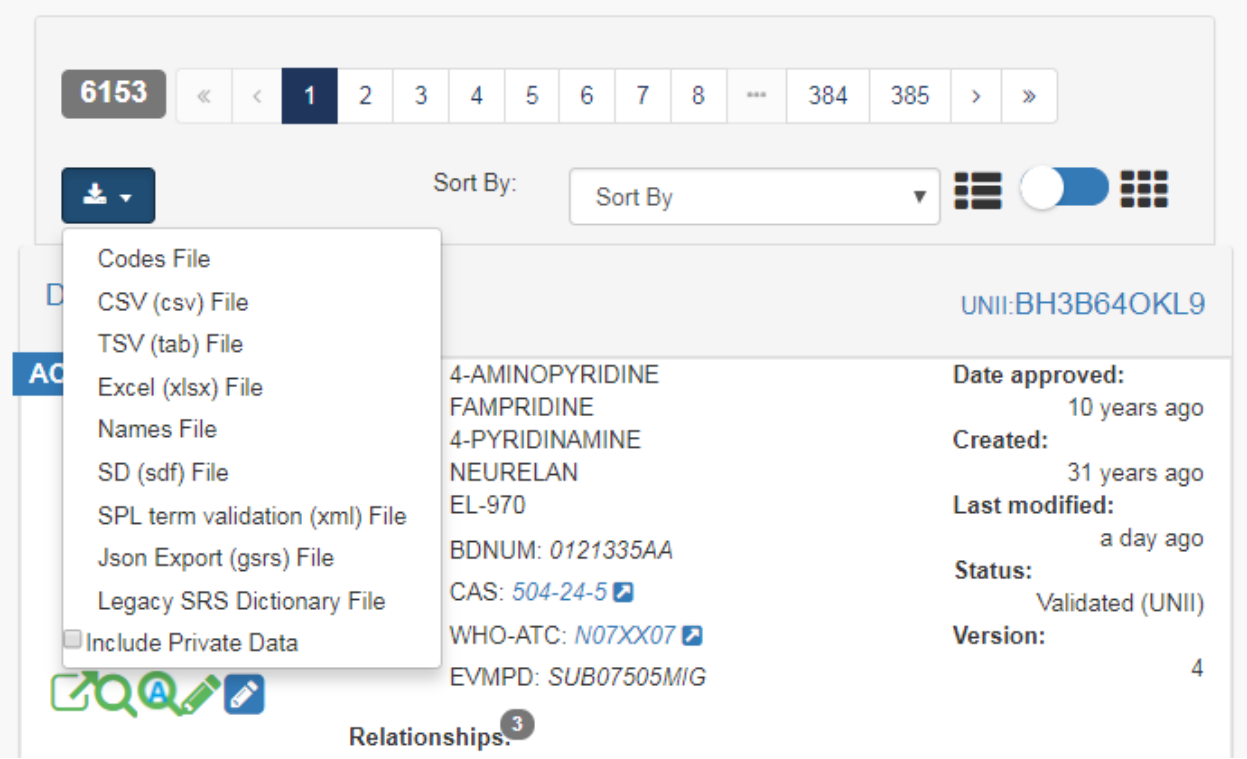
The facets are organized into sections, such as stereochemistry, molecular weight, relationships, etc. The first three facets are fully expanded by default, the remaining facets are collapsed and can be expanded as needed.

3.2.2. Display (Header & Footer) Pane

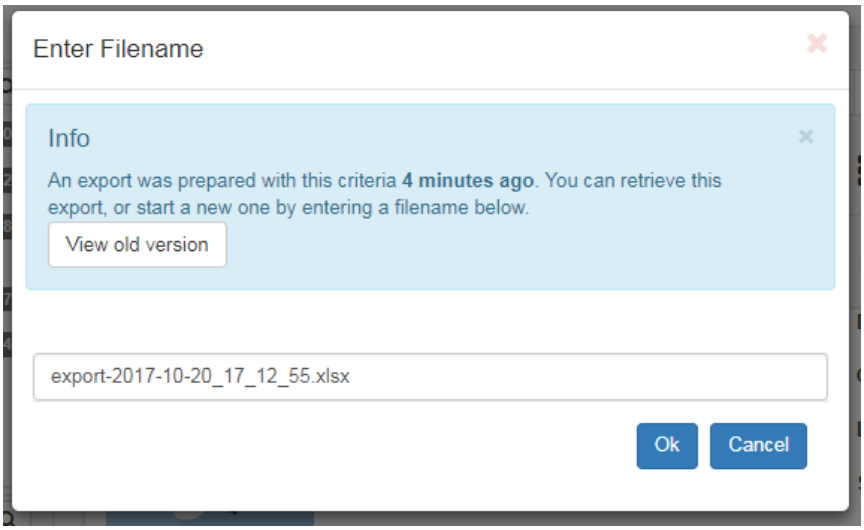
The display pane has components located at the top and bottom of the dashboard. In both places, the display pane shows the user the total number of records returned by the FDA GSRS database based on any (defaults to none) search criteria specified in the filter pane and includes pagination for the user to navigate to a specific page of results.

At the top portion of the dashboard only, there is a download icon that allows the user to export a summary of each of the records displayed in the results pane to a separate file. The following export formats are supported:

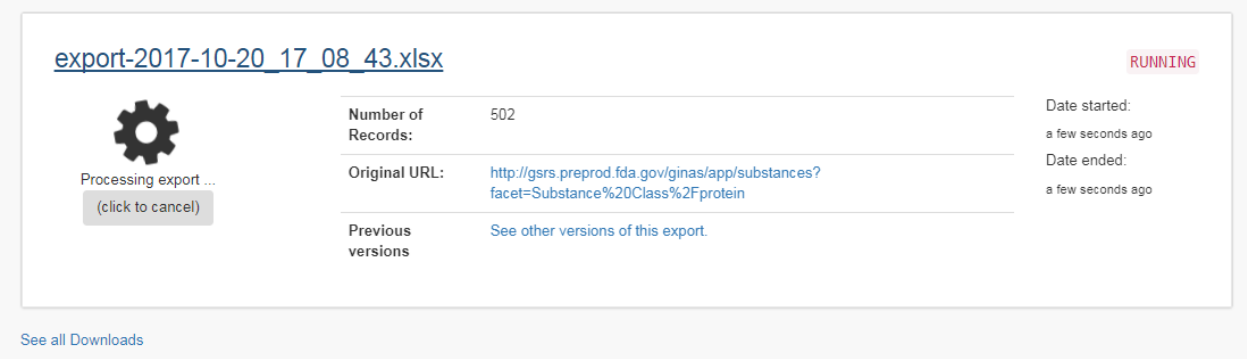
- Codes File
- JSON
- TSV
- CSV
- SDF
- XLSX
- Names File
- SPL Term Validation (xml)
- Legacy SRS Dictionary



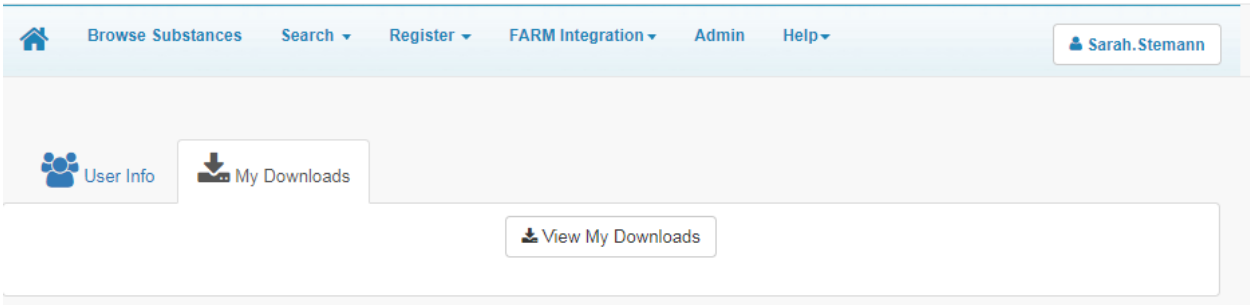
After initiating a download, enter a filename (or accept the default date/time) and retrieve the file from My Downloads



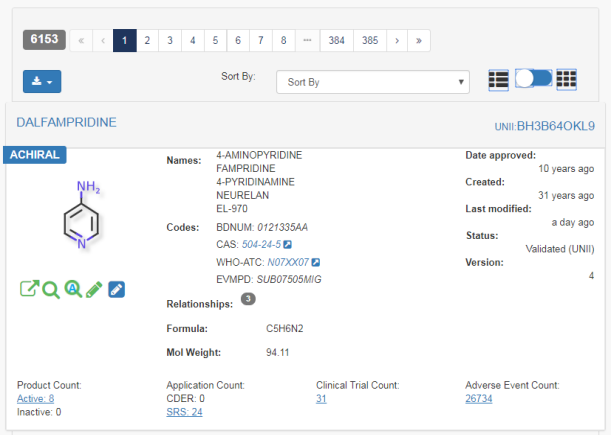
While the download is generated, continue using GSRS for your research.



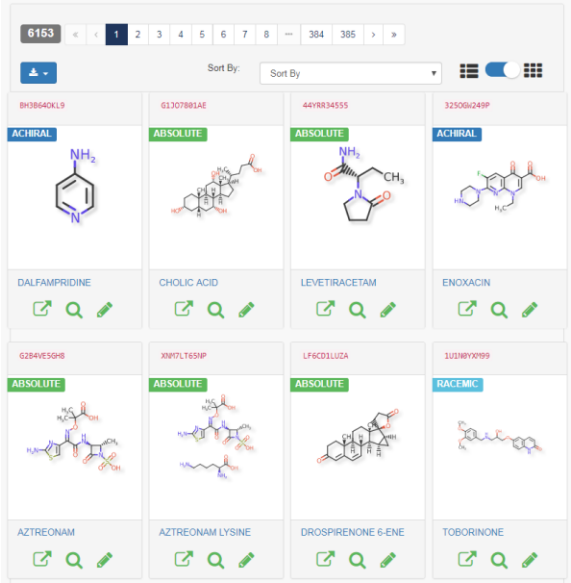
To access your download(s), click your user name in the upper right corner, click My Downloads, and View My Downloads



In addition to the download icon, there is a toggle that allows the user to change the display format of records in the results pane between a summary list and grid format.



List Format (default)



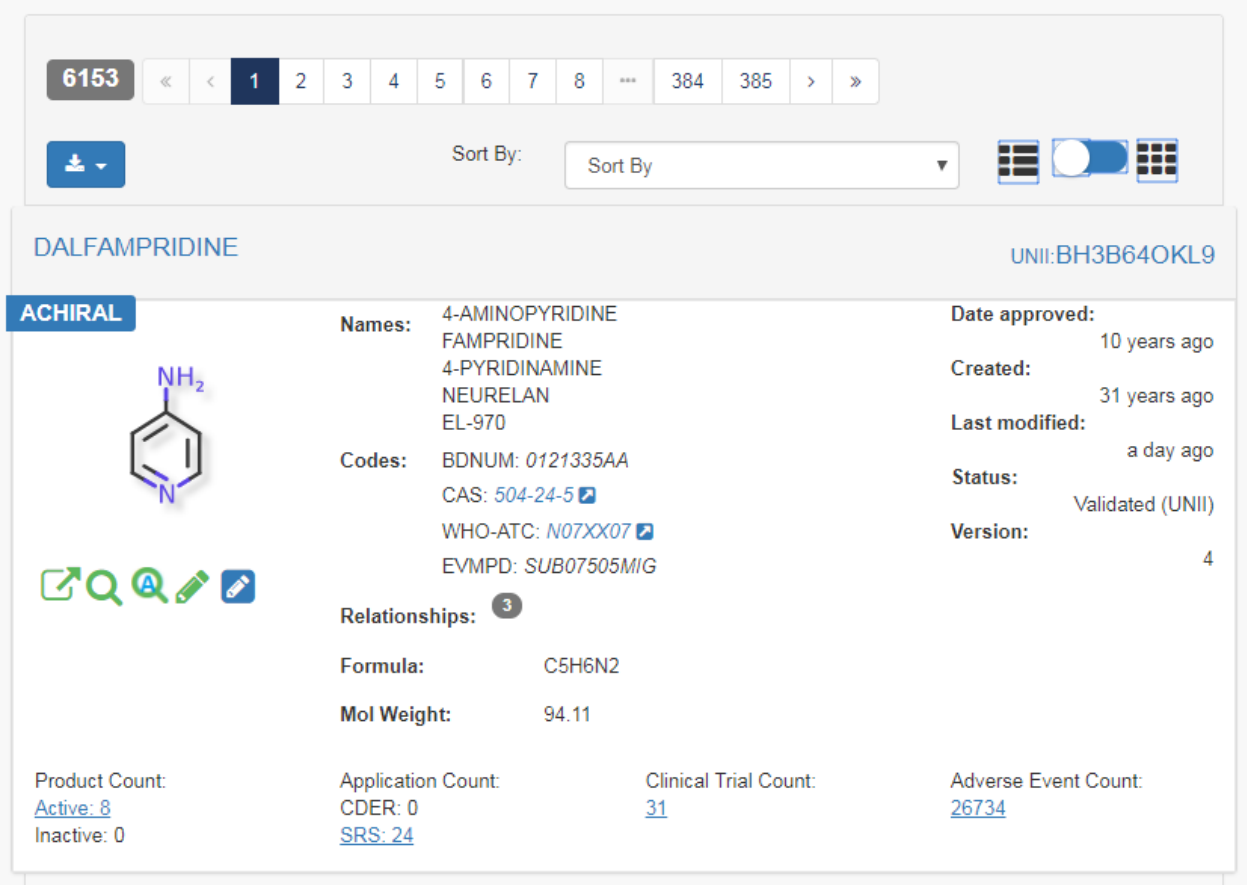
Grid Format

3.2.3. Results Pane

The results pane is located at the center of the dashboard. Here, the substance answers are displayed in either list or grid format depending on which setting is toggled from the display pane at the top.

The list format provides a summary view for each record returned by the FDA GSRS database. Basic information captured in a summary view includes the substance UNII code (if any), names (up to six simultaneously), codes (up to six simultaneously), number of relationships to other substance records, date approved, and time last modified.

Additional information depending on substance type is also presented; for instance, substance records include the chemical formula, number of moieties, the stereochemistry category and a two-dimensional chemical structure. An example summary view for a chemical substance record is shown below:

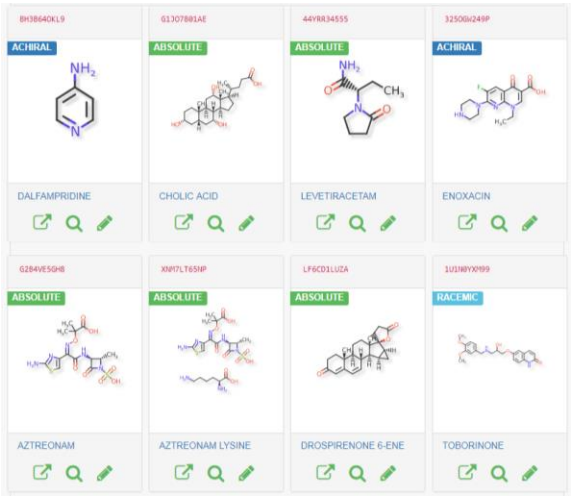


The additional information incorporated into the summary view for each substance type is included below:

- Substance
  - Formula
  - Number of moieties
  - Stereochemistry
  - 2D structure
- Structurally diverse
  - Part
- Mixture
  - Number of components
- Polymer
  - Display structure
- Protein
  - Number of subunits
- Nucleic Acid
  - Number of subunits
- Actions available from the summary view include:

- Export – options for individual substance export provided in next window
- Structure (or Sequence) Search – if applicable
- Structure Search in Advanced Search – if applicable
- Edit – power users
- New Form – power users

The grid format solely presents the UNII code (if any), the preferred name of the substance, and the structure (if the substance is either a chemical or a polymer). An example grid view displaying several chemical substance records is shown below:



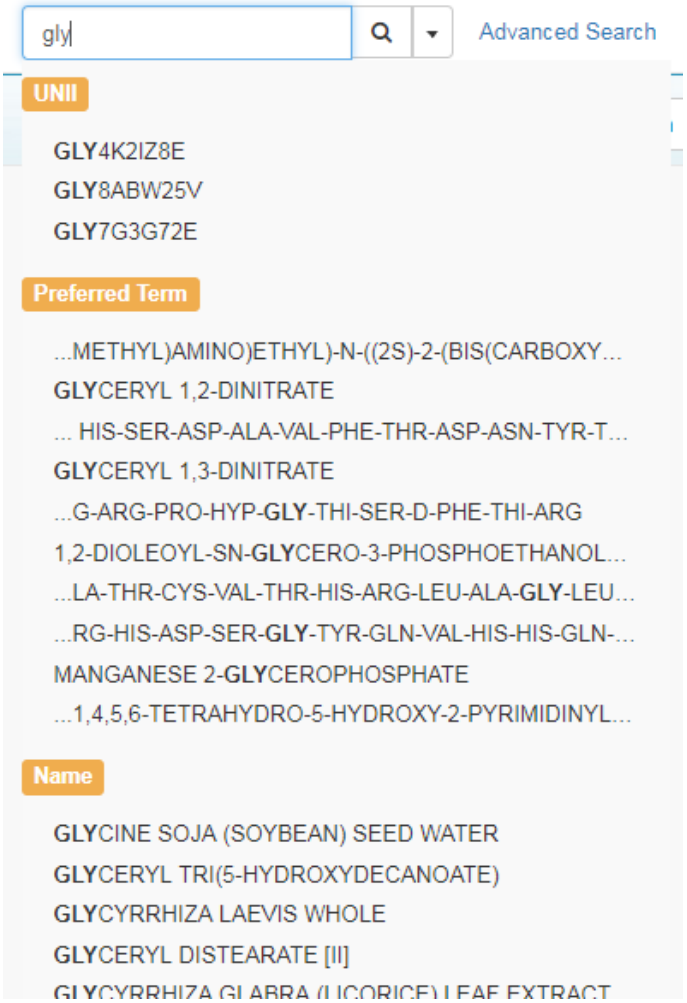
For chemical and polymer records in either view, the user can export a chemical table file or SMILES string of the substance record by clicking on the export icon below the structure. The user can also transfer the structure to the Structure Search interface to perform an additional search by clicking on the search icon.

3.3. *Search*

FDA GSRS supports several means for searching substances based on user-specified criteria. These are Global Search, Advanced (and free-text) Search, Structure Search, and Sequence Search

3.3.1. Global Search

Global search is a free-text search box in the upper right of every page. Type the text you are looking for and the type-ahead feature will display the top matchs for UNIIs and names separated by preferred names and all other name types. You can select from one of the matches, or use the magnifying glass for a wider search. Results are presented in the display pane.

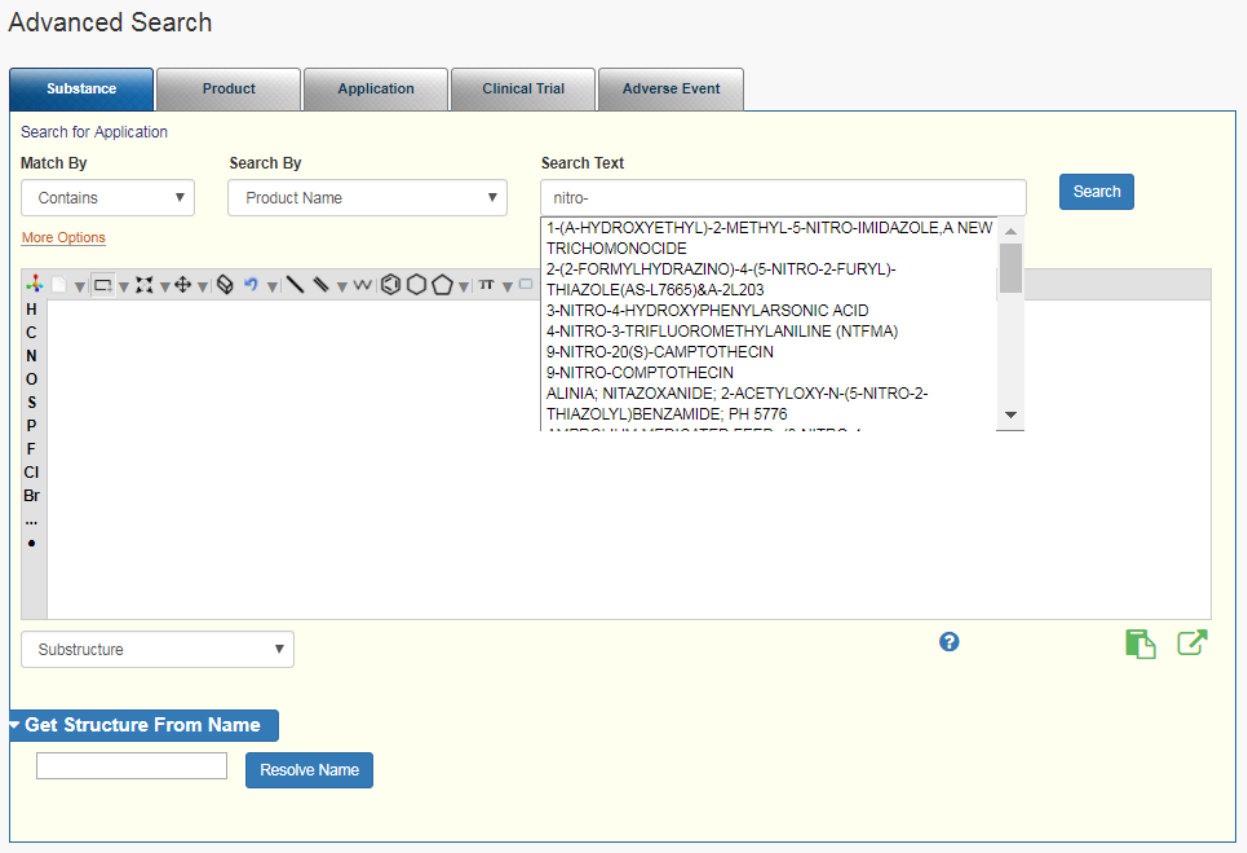

gly
Advanced Search
UNII
GLY4K2IZ8E
GLY8ABW25V
GLY7G3G72E
Preferred Term
...METHYL)AMINO)ETHYL)-N-((2S)-2-(BIS(CARBOXY...
GLYCERYL 1,2-DINITRATE
... HIS-SER-ASP-ALA-VAL-PHE-THR-ASP-ASN-TYR-T...
GLYCERYL 1,3-DINITRATE
...G-ARG-PRO-HYP-GLY-THI-SER-D-PHE-THI-ARG
1,2-DIOLEOYL-SN-GLYCERO-3-PHOSPHOETHANOL...
...LA-THR-CYS-VAL-THR-HIS-ARG-LEU-ALA-GLY-LEU...
...RG-HIS-ASP-SER-GLY-TYR-GLN-VAL-HIS-HIS-GLN-...
MANGANESE 2-GLYCEROPHOSPHATE
...1,4,5,6-TETRAHYDRO-5-HYDROXY-2-PYRIMIDINYL...
Name
GLYCINE SOJA (SOYBEAN) SEED WATER
GLYCERYL TRI(5-HYDROXYDECANOATE)
GLYCYRRHIZA LAEVIS WHOLE
GLYCERYL DISTEARATE [II]
GLYCYRRHIZA GLABRA (LICORICE) LEAF EXTRACT

3.3.2. Advanced Search

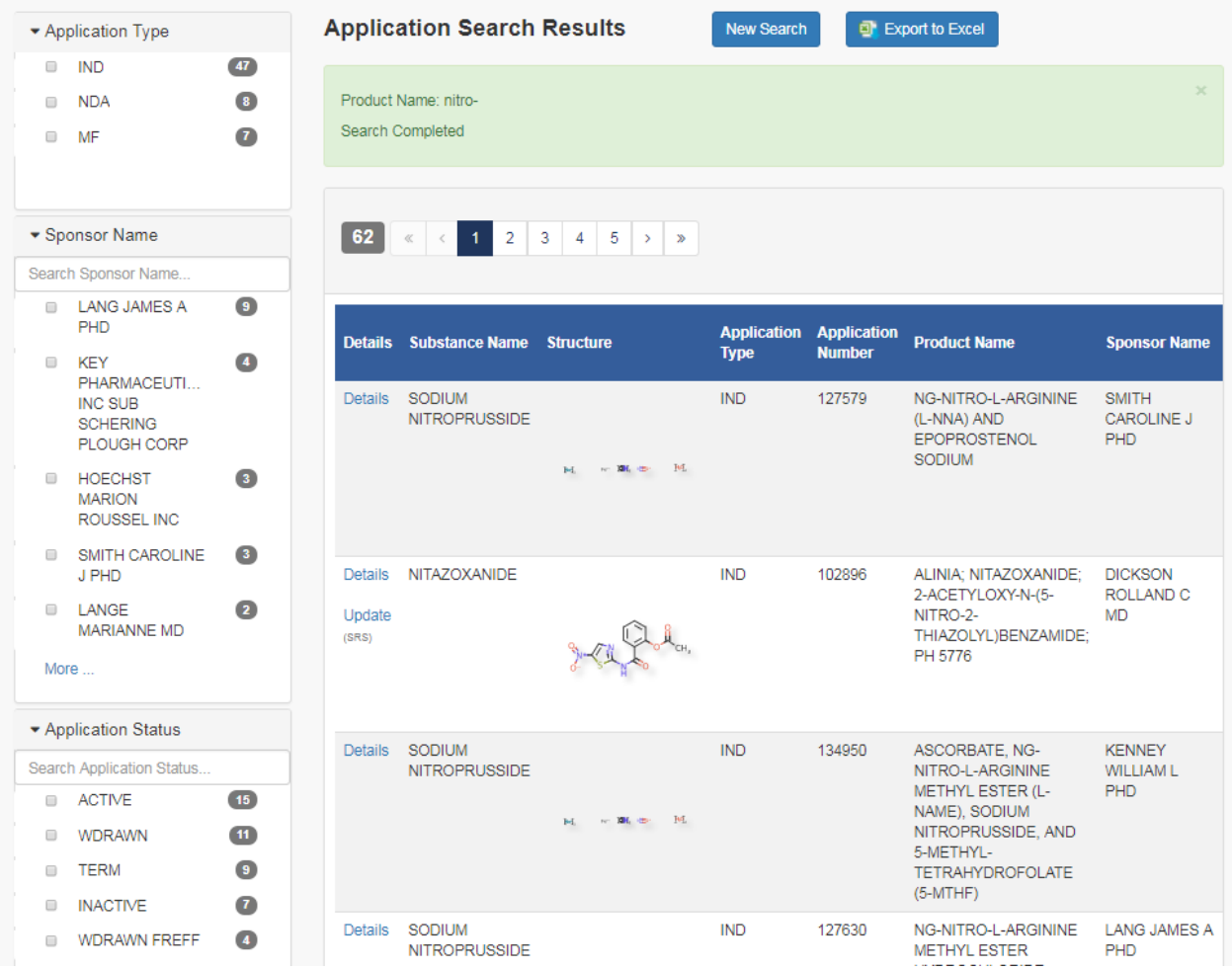
Not included in the current public-facing site. The advanced search can be reached via the navigation pane, global search results or by selecting the A magnifying glass from a specific substance. The structure search pane within Advanced Search can be used similar to Structure Search described in section 3.4.3.

The Advanced Search criteria is organized by five tabs: Substance, Product, Application, Clinical Trial, and Adverse Events. Within each of the tabs, a structure can be searched (similar to structure search), or criteria specific to each tab can be searched. For the criteria there are three selections:

- Match By: Contains or Exact
- Search By: field name
- Search Text: free text, with type-ahead matches displayed as optional selections

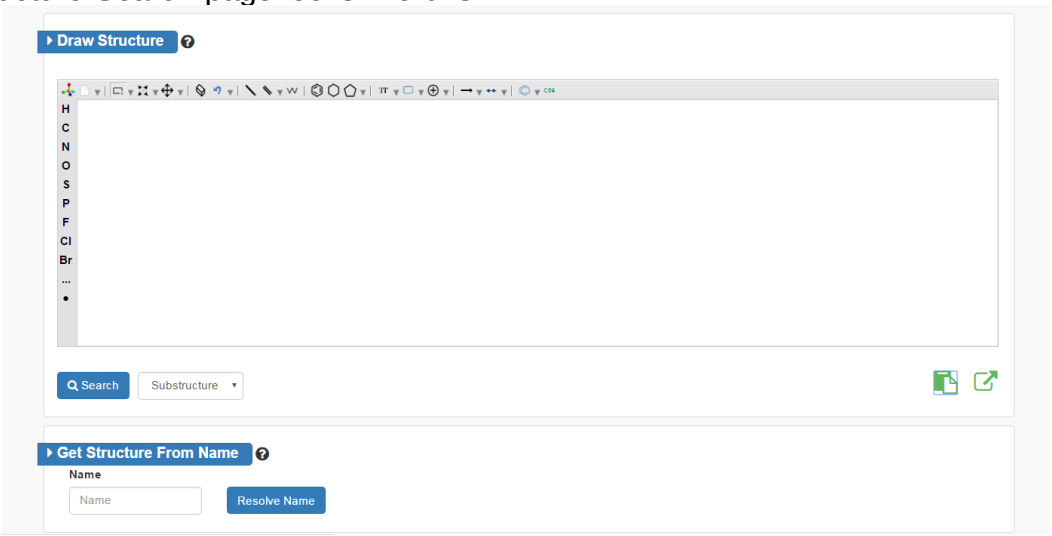


After entering the criteria and clicking search, the results are presented in a grid, with details specific to the advanced search type. Results can be exported to Excel or filtered similar to Browse Substances.



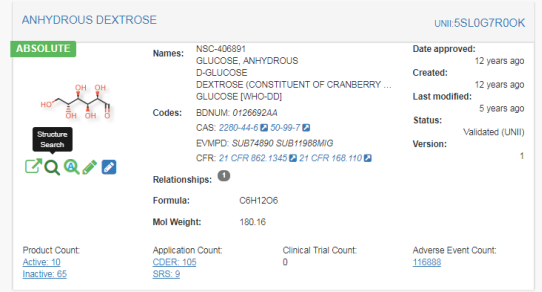
3.3.3. Structure Search

FDA GSRS Structure Search is a means for searching for specific chemical and polymer records based on their two-dimensional chemical structure. The interface on the Structure Search page looks like this:

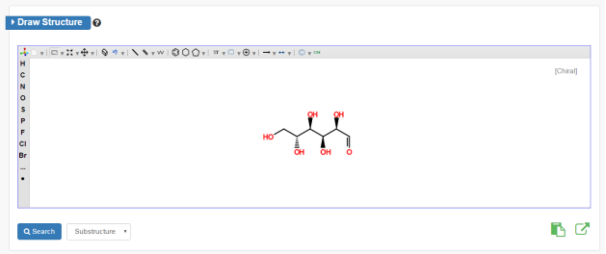


There are several ways that a structure can be prepared to perform a search.

- Transfer a structure found during a search or browse session using the results pane search icon.

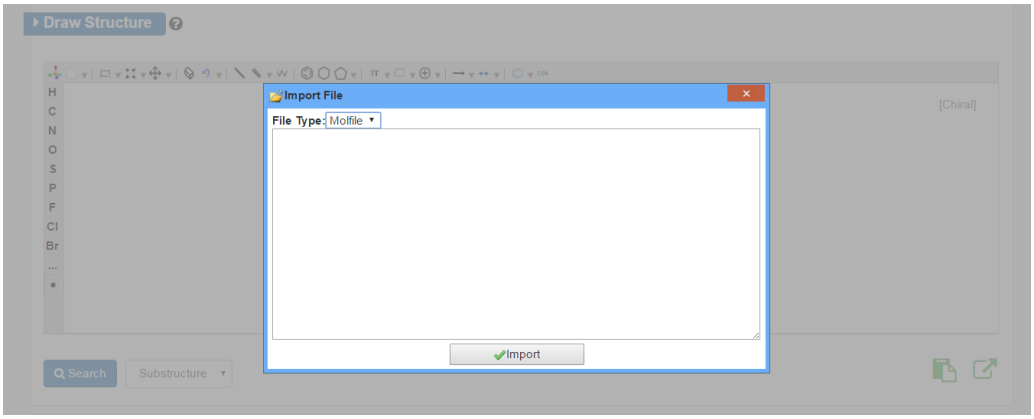


Clicking on the search (magnifying glass) icon below the structure…

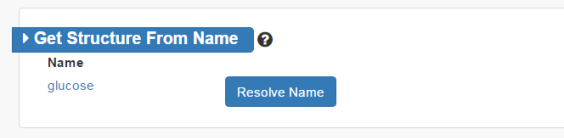


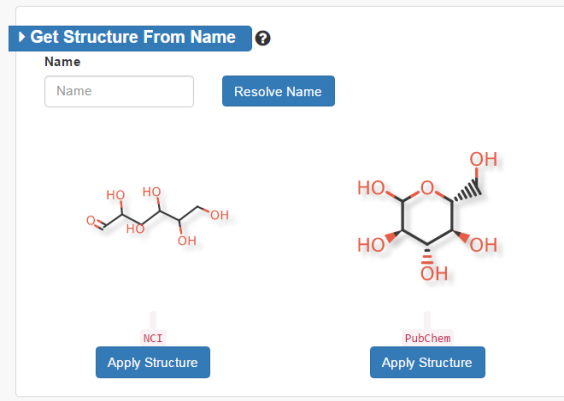
Transfers the structure to the Structure Search window.

- Import an external file (supported formats: Molfile, Rxnfile, Html).

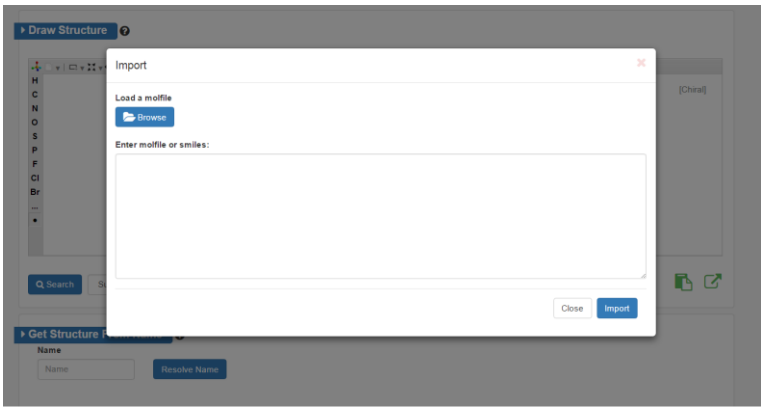


- Draw a structure by using the included drawing tool. You may use either the default JSDraw editor or more standalone chemical editor. SymyxDraw is the preferred standalone because the molfile is its native format.
- Get structure from its substance name





- Get structure from its SMILES string representation

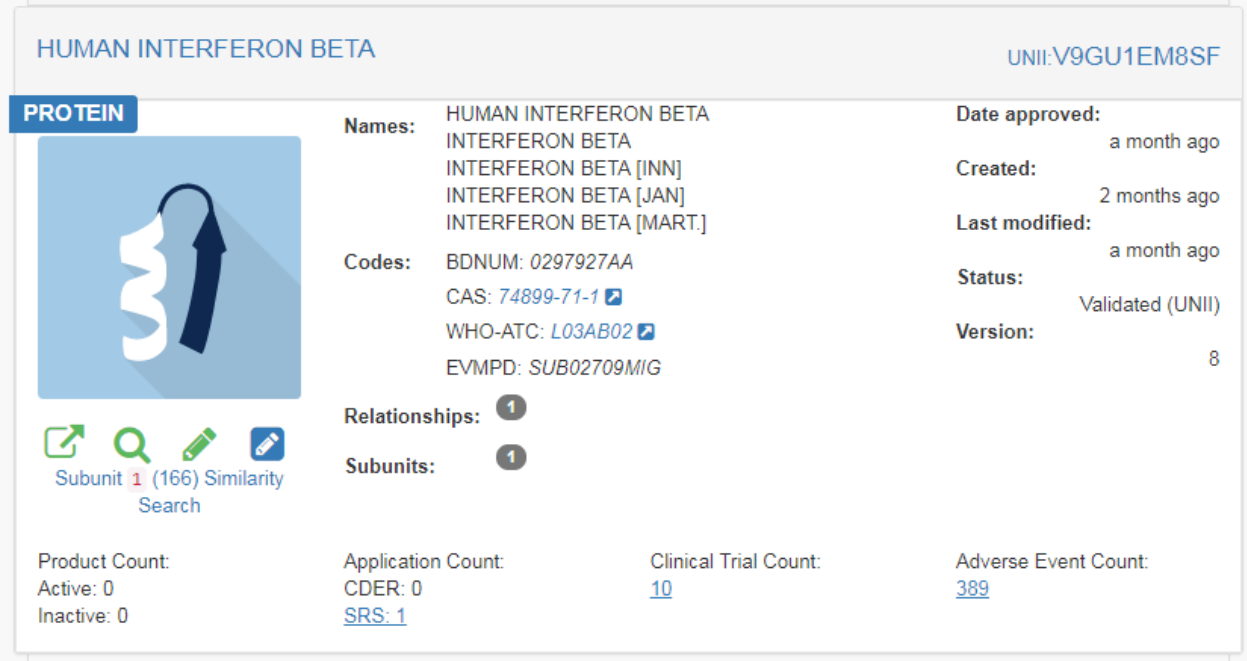


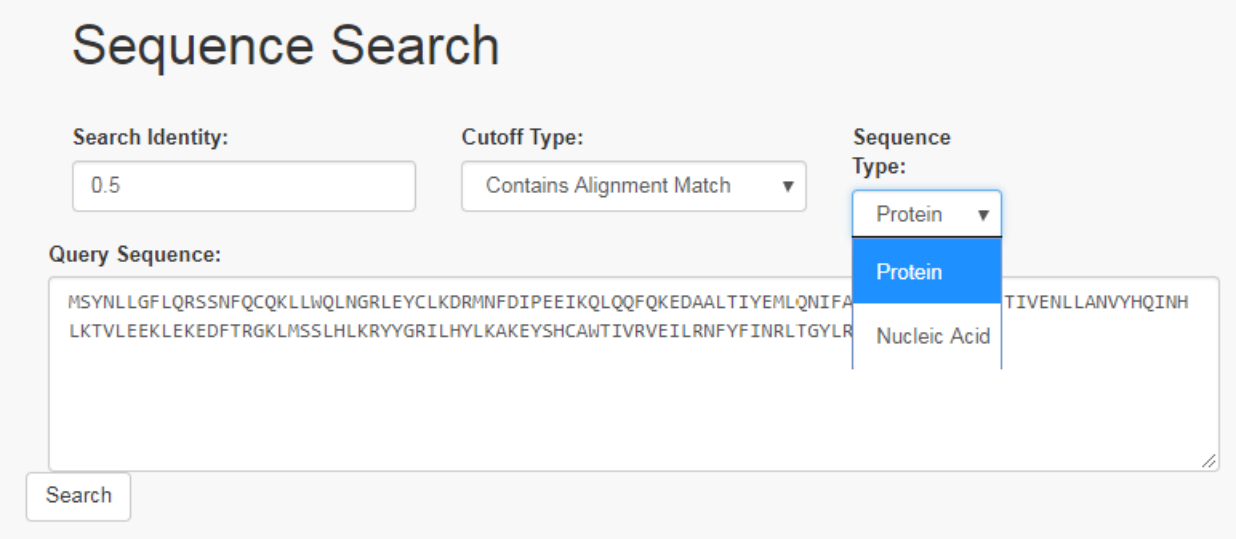
With the structure ready, select the search option.

- Substructure: find other compounds that the query structure is embedded in. This is useful for finding a set of compounds that share a common "substructure."
- Similarity: find other compounds with structural features that are similar to those of the query structure. Similarity percentages between 50% and 100% are returned. Compounds that are related and are relatively close in overall size can be found using this method.
- Exact: find the query structure as a complete entity, where all the structure's atoms and bonds are identical in the retrieved compound.
- Flex: Find the query structure, with all bonds identical in the retrieved compound including stereochemical and tautomeric bonds. Flex also finds salts, mixtures, hydrates, and polymers of the parent.

3.3.4. Sequence Search

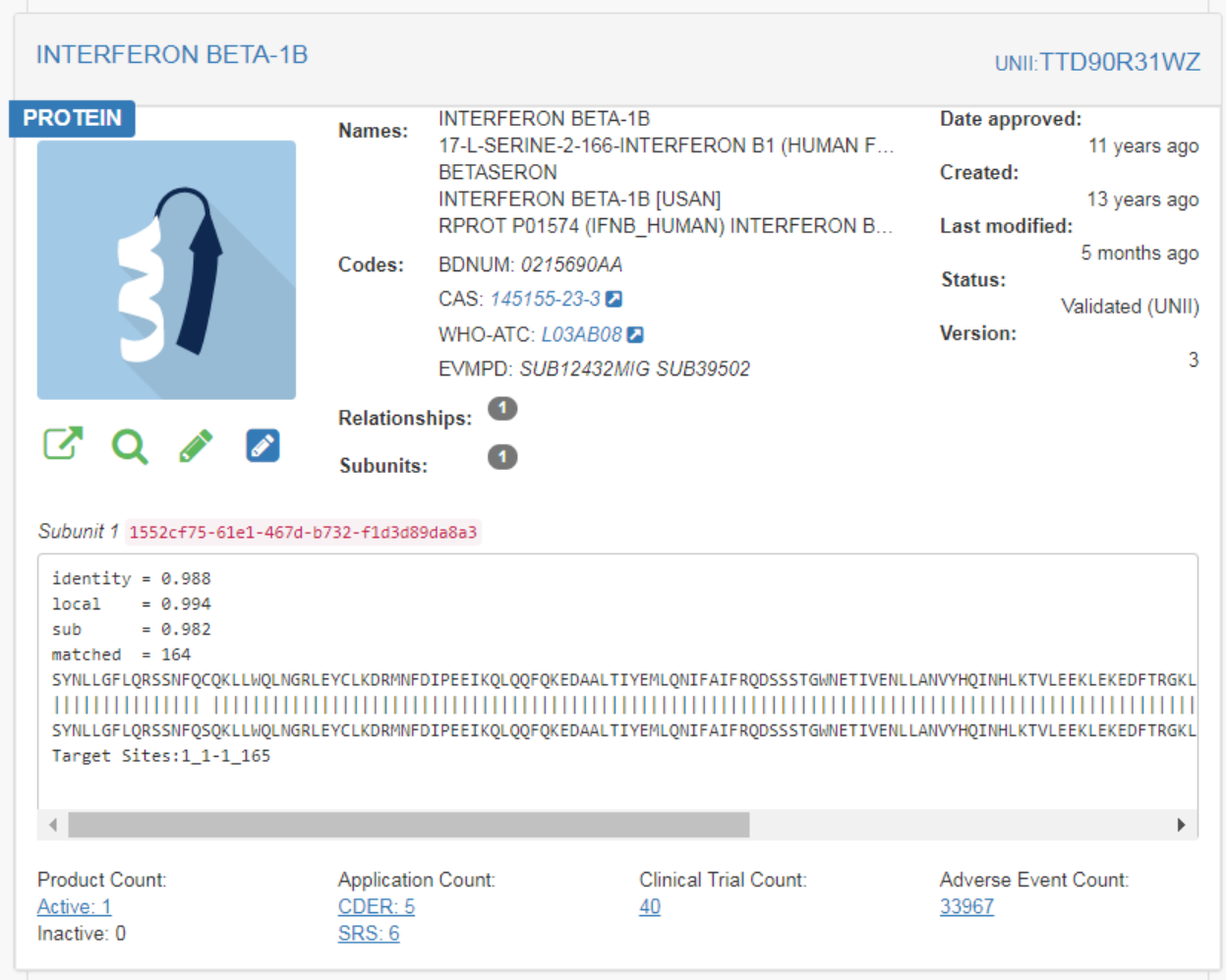
FDA GSRS also supports searching on sequences to locate protein and nucleic acid substance records. Sequence searching can be accessed from the Search menu or from an existing substance, using the magnifying glass which now links to the sequence page.





FDA GSRS has three alignment methods with which to search sequence candidates: contains, global, and local alignment matches. There are also two sequence types.

The results display the statistics for the match:

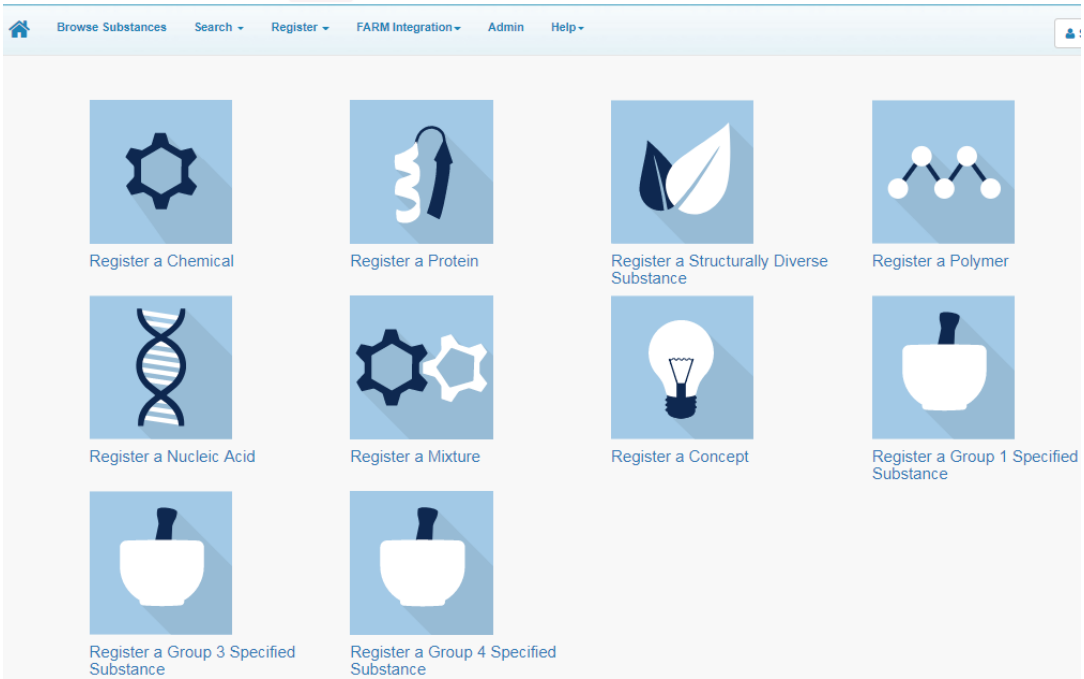

INTERFERON BETA-1B
UNII:TTD90R31WZ
PROTEIN
Names:
INTERFERON BETA-1B
17-L-SERINE-2-166-INTERFERON B1 (HUMAN F...
BETASERON
INTERFERON BETA-1B [USAN]
RPROT P01574 (IFNB_HUMAN) INTERFERON B...
Codes:
BDNUM: 0215690AA
CAS: 145155-23-3
WHO-ATC: L03AB08
EVMPD: SUB12432MIG SUB39502
Relationships: 1
Subunits: 1
Date approved:
11 years ago
Created:
13 years ago
Last modified:
5 months ago
Status:
Validated (UNII)
Version:
3
Subunit 1 1552cf75-61e1-467d-b732-f1d3d89da8a3
identity = 0.988
local = 0.994
sub = 0.982
matched = 164
SYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIYEMLQNIFAIFRQDSSSTGWNETIVENLLANVYHQINHLKTVLEEKLEKEDFTRGKL
SYNLLGFLQRSSNFQSQKLLWQLNGRLEYCLKDRMNFDIPEEIKQLQQFQKEDAALTIYEMLQNIFAIFRQDSSSTGWNETIVENLLANVYHQINHLKTVLEEKLEKEDFTRGKL
Target Sites:1_1-1_165
Product Count:
Active: 1
Inactive: 0
Application Count:
CDER: 5
SRS: 6
Clinical Trial Count:
40
Adverse Event Count:
33967

3.4. *Registration*

3.4.1. Chemical

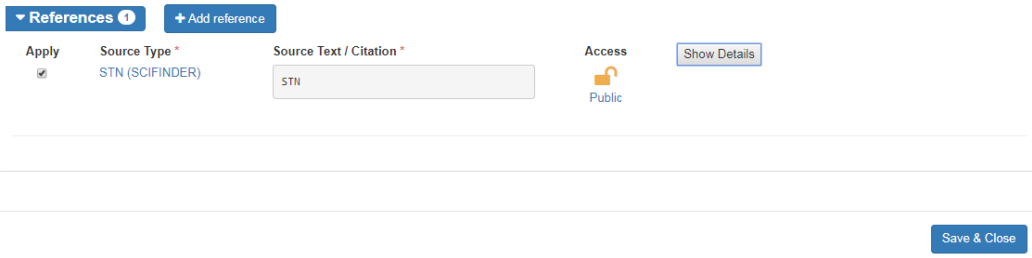
1. From the navigation pane, select Register > Register Substance

2. For this example, from the list of substance types, select “Register a Chemical”



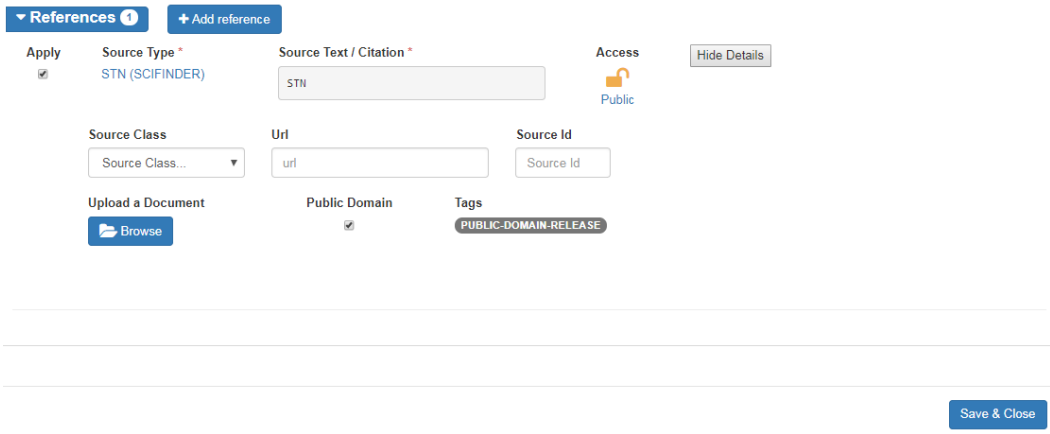
3. Definition Level should be set to Complete. Incomplete and Representative definition levels should be avoided. Also avoid generating records with unknown stoichiometry, stereochemistry, or substitution.

4. Add a Definition Reference by clicking the  icon
   a. Click Apply to ensure the reference is selected after Save
   b. Update Source Type and Source Text



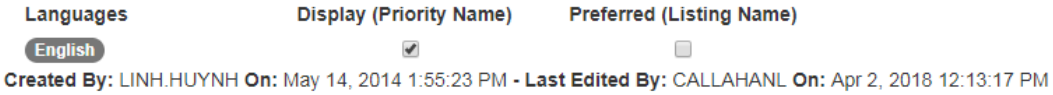
   c. Click Show Details
   d. Select Public Domain checkbox and add the Public Domain Release tag

e. Click Save & Close

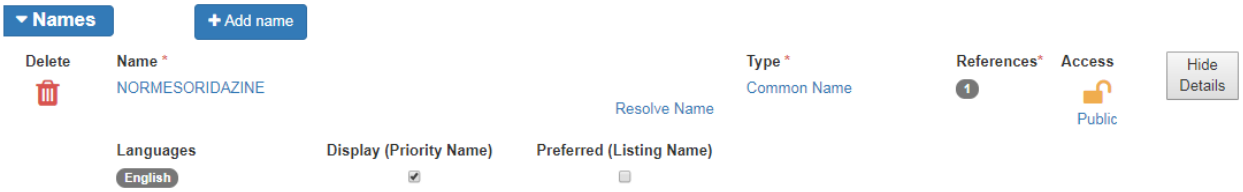


f. Ensure the Reference is attached to the Definition, the  is replaced with  number will correspond to the number of references added

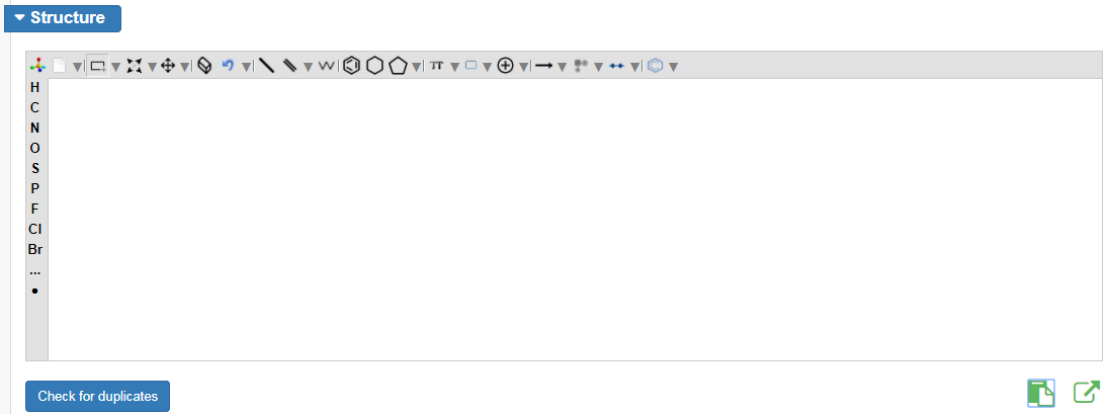
5. Add a Name by clicking the Add Name button
   a. Add a Name and Reference, click Show Details. Names should be standardized to ASCII UPPER CASE.Names should be globally unique and most relevant to the specific record.
   b. Update Display/Priority Name checkbox as necessary. Only one Display name is allowed per record per software instance. Display names should use natural word order EXCEPT for stereochemical attributes which should be preceded by a comma and followed by a dash at the end of the name; i.e.display names should be formatted like the following example: 1,2-DIMETHYLCYCLOHEXANE, (1R,2R)- . **Only ADMIN users should select Preferred/Listing names**.



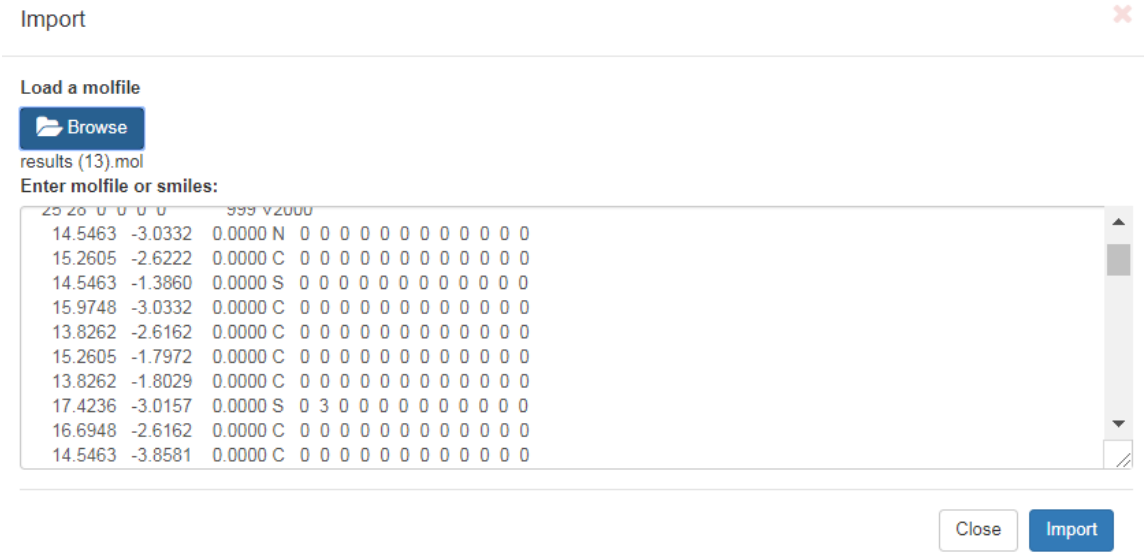
   c. Click Add Name as many times as needed



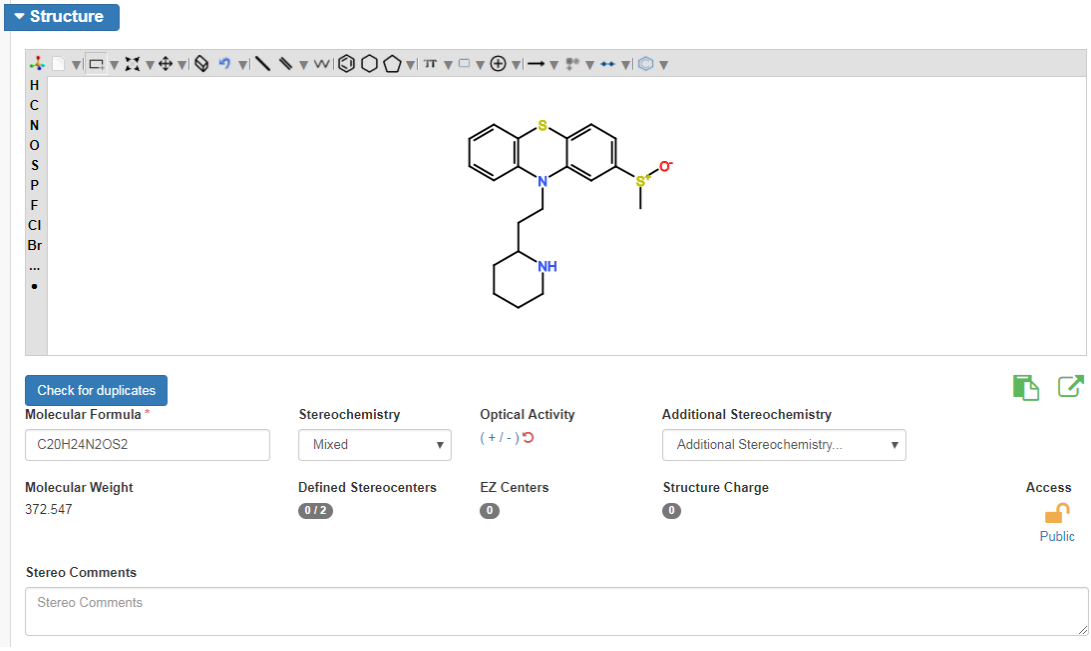
6. Add a Structure, as created in ChemDraw or your preferred tool and saved in molfile format, using the Import button  .



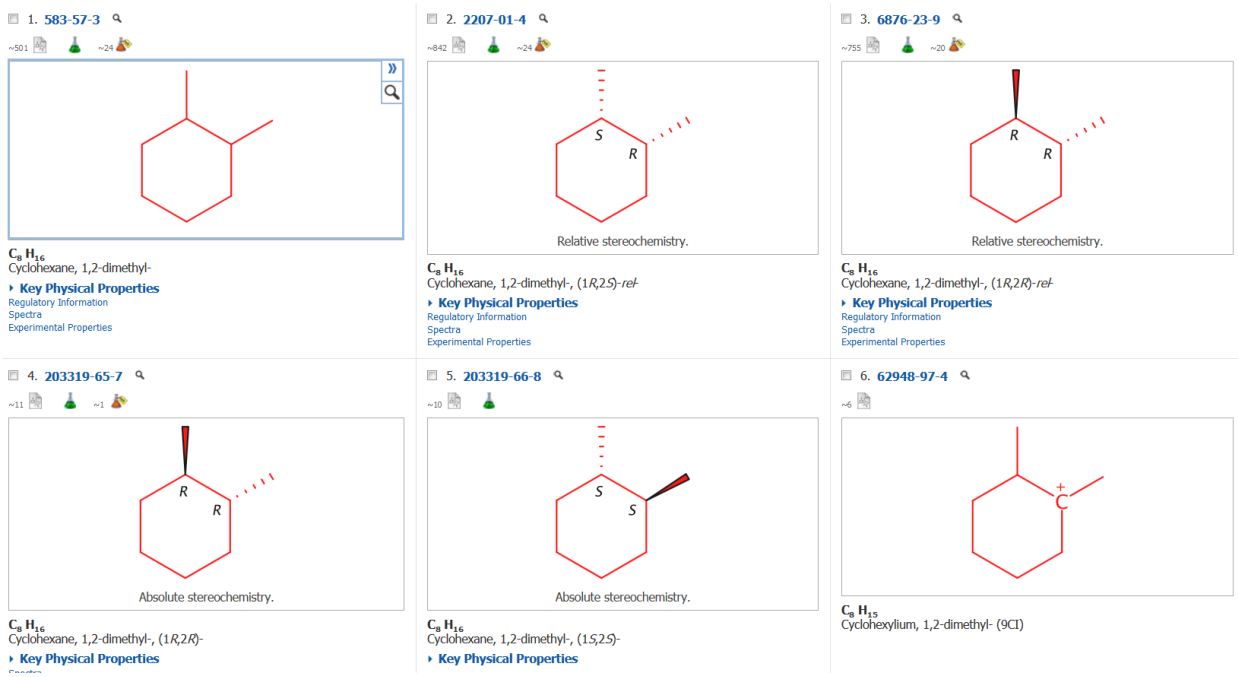
a. Browse for your file and click Import



b. The structure and it's details are available for editing via JSDraw

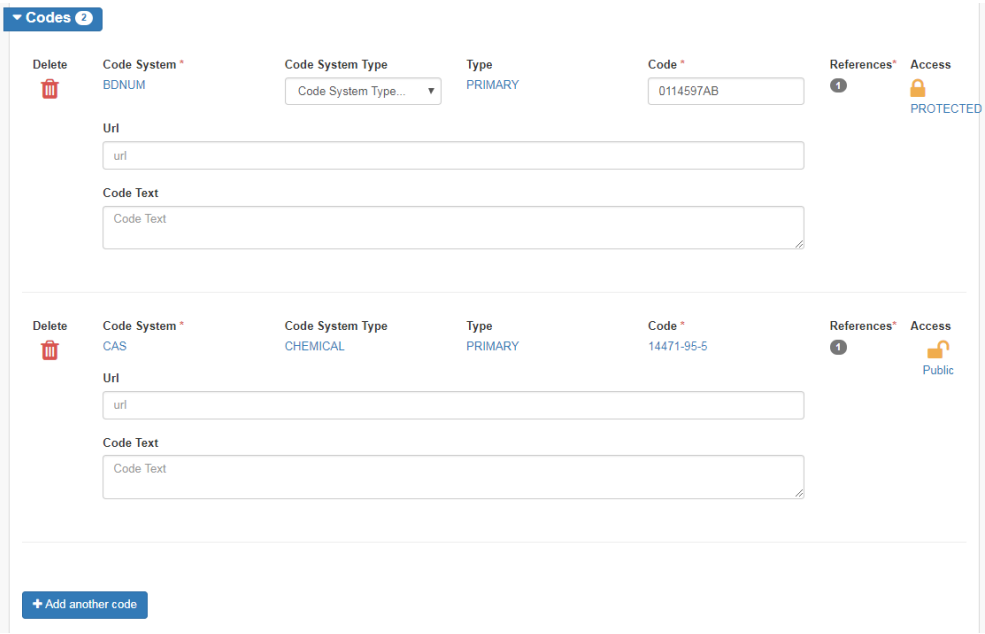


7. Each unique stereochemical option should have its own GSRS record. For the exampe of 1,2-DIMETHYLCYCLOHEXANE, 5 GSRS records should be created: (1R,2R), (1S,2S), (+/-)-TRANS, CIS (meso), and mixed isomers (with two component CIS/TRANS mixture substance as alternative definition). These are the first five CAS Registry numbers below.

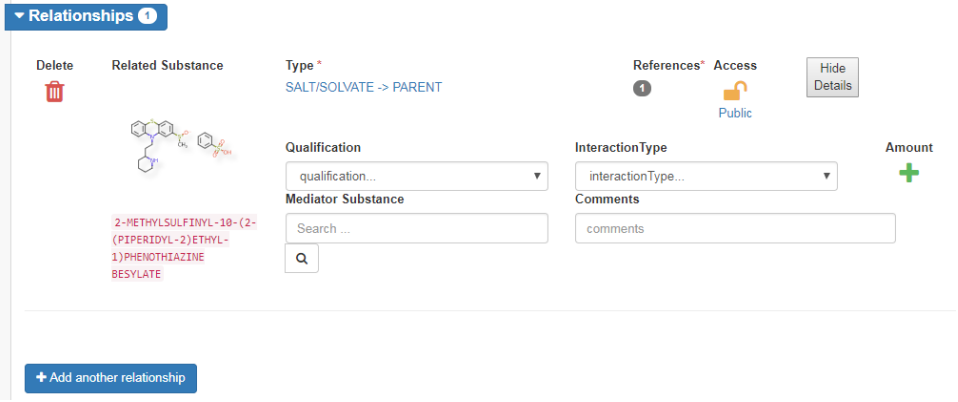


Exact CAS Registry structure searches are equivalent to Flex searches in GSRS.

8. Add Codes as needed, update Access if the code is protected. URLs will be automatically generated for the most common code systems. The first code added for a particular code system should have the PRIMARY Type attribute. Additional codes using the same code system require selection of the most appropriate other attribute. CAS Registry numbers listed by CAS as DELETED are referred to as SUPERCEDED in GSRS.

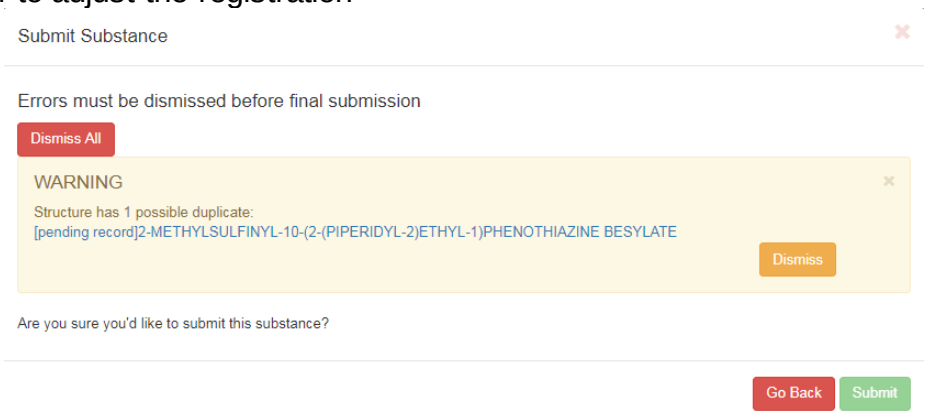


9. Add Relationships, select the substance after searching. Salts and solvates containing multiple discrete molecular entities should always include reference to the organic parent GSRS record.



10. Skip Properties for Chemicals

11. To avoid repeating work, click Submit frequently during the registration process after a structure, a name, and a reference have been added to the record.
    a. Any Errors (in red) will need to be resolved
    b. Warnings (in orange) can be dismissed after research has determined there is no need to adjust the registration



12. Once a Chemical is submitted, it can be Viewed, Edited, and Edited in New Forms – it's status is Pending

13. A privleged user account will review the registration and approve
    a. At this point that status will change from Pending to Approved
    b. A UNII will be assigned

4. Metadata

The attributes included in the schema are briefly summarized below:

Contains GSRS identifier information for the substance such as the BDNUM and the UNII.

Substance Type Specific:

This attribute is shared across all substance types.

This attribute is included in all records.

GSRS prefaces the detail attribute name with the type of substance described (e.g. chemical detail, polymer detail, etc.).

---

i. Structure

Contains the two-dimensional chemical structure as well as additional metadata such as stereochemistry, optical activity, the molecular formula and weight, number of defined stereocenters and EZ centers, and the charge.

Substance Type Specific:

This attribute is chemical specific.

---

ii. Moieties

Contains structure information (see Structure) and stoichiometry of moieties contained within the substance.
A moiety is defined as an entity within a substance that has a complete and continuous molecular structure. This includes single substances, ions and solvate molecules.

Substance Type Specific:
This attribute is chemical specific.

---

iii. Active Moieties

Contains links to the active moieties of the substance.

An active moiety is defined in 21CFR316.3 as the “molecule or ion, excluding those appended portions of the molecule that cause the drug to be an ester, salt (including a salt with hydrogen or coordination bonds), or other noncovalent derivative (such as a

complex, chelate, or clathrate) of the molecule, responsible for the physiological or pharmacological action of the drug substance."

Substance Type Specific:

This attribute is shared across all substance types.

---

iv. Names

Contains Chemical Names, compiled from authoritative and other public sources, associated with the substance.

Chemical Name Types:

- Brand: trade name.
- Code: code name.
- Common: trivial name.
- Official: name given by an official registration authority.
- Systematic: any name constructed using a nomenclature that describes the structure of the substance.

Substance Type Specific:

This attribute is shared across all substance types.

This attribute is included in all records.

---

v. Codes

Contains numerical identifiers, compiled from authoritative and other public sources, associated with the substance.

Substance Type Specific:

This attribute is shared across all substance types.

This attribute is included in all records.

---

vi. Relationships

Contains links to substance records related to the substance.

Substance Type Specific:

This attribute is shared across all substance types.

---

vii. Notes

Contains free-text comments related to the substance and its data.

Substance Type Specific:

This attribute is shared across all substance types.

---

viii. References

Contains source text/citations for all data presented in the schema.

Substance Type Specific:

This attribute is shared across all substance types.

This attribute is included in all records.

---

ix. Audit Information

Contains timestamps and identifies the Data Steward associated with approval, creation, and last edit of the substance record.

Substance Type Specific:

This attribute is shared across all substance types.

This attribute is included in all records.

---

x. Metabolites

Contains links to the metabolites of the substance.

Substance Type Specific:

This attribute is shared by chemical, mixture, protein, concept, and structurally diverse substance types.

---

xi. Definition

Contains a schema that is used to unambiguously identify a substance that cannot be adequately defined by its structure, sequence, or components.

Substance Type Specific:

This attribute is shared by concept and structurally diverse substance types.

---

xii. Subunits

Contains sequence information for each subunit in a substance.

Substance Type Specific:

This attribute is shared by protein and nucleic acid substance types.

---

xiii. Disulfide Bridges

Identifies which pairs of residues are connected by disulfide bridges.

Substance Type Specific:

This attribute is protein specific.

---

xiv. Modifications

Contains information on structural, physical, and agent modifications to a substance.

Substance Type: Specific:

This attribute is shared by protein, structurally diverse, polymer, and nucleic acid substance types.

---

xv. Properties

Contains properties information on a substance.

Substance Type Specific:

This attribute is shared by protein, polymer, and nucleic acid substance types.

---

xvi. Components

Contains links to substance records that are contained within a mixture substance definition and identifies the mixture type.

Mixture Types:

- All of: single substances that are synthesized or isolated together
- One of: used for ambiguous substance classes

Substance Type: Specific:

This attribute is mixture specific.

---

xvii. Applications, Products, Clinical Trials, Adverse Events

Contains a data table that lists the product, application, clinical trial, and adverse event details associated with the substance and links to the individual records.

Substance Type Specific:

This attribute is shared across all substance types.

---

xviii. Display Structure

Contains the two-dimensional chemical structure with repeating units shown in brackets.

Substance Type Specific:

This attribute is polymer specific.

---

xix. Monomers

Contains links to substance records from which repeated units are derived.

Substance Type Specific:

This attribute is polymer specific.

---

xx. Structural Units

Contains the two-dimensional chemical structure of repeating units.

Substance Type Specific:

This attribute is polymer specific.

---

xxi. Sugars

Lists the nucleotide sugars contained in a nucleic acid.

Substance Type Specific:

This attribute is nucleic acid specific.

---

xxii. Linkages

Lists the type of phosphate ester bonds joining nucleotides in a nucleic acid.

Substance Type Specific:

This attribute is nucleic acid specific.

APPENDIX A: USER MANUAL APPROVAL

The undersigned acknowledge that they have reviewed the FDA GSRS **User Manual** and agree with the information presented within this document. Changes to this **User Manual** will be coordinated with, and approved by, the undersigned, or their designated representatives.

Signature: _____ Date: _____

Print Name: _____

Title: _____

Role: Project Manager

APPENDIX B: REFERENCES

The following table summarizes the documents referenced in this document.

Document Name	Description	Location

APPENDIX C: KEY TERMS

The following table provides definitions and explanations for terms and acronyms relevant to the content presented within this document.

Term	Definition
CSV	Comma Separated Values
GSRS	Global Substance Registration System
JSON	JavaScript Object Notation
SDF	Structure-data File
SMILES	Simplified Molecular Input Line Entry System
SRS	Substance Registration System
TSV	Tab Separated Values
UNII	Unique Ingredient Identifier
XML	eXtensible Markup Language