US Food and Drug Administration

Registering chemicals or proteins in FDA's Global Substance Registration System within precisionFDA

Last updated March 18, 2024

Table of Contents

Introduction

FDA's Global Substance Registration System (GSRS) makes it possible to exchange information on substances efficiently and accurately through their UNique Ingredient Identifiers (UNIIs), which can be generated before submissions and at any time in the regulatory life cycle. This guide will only cover the minimal requirements for registering new chemicals and proteins in GSRS using precisionFDA.

A chemical is a small molecule or substance that is defined by atoms and bonds. The chemical definition should contain as much known information as possible about stereochemistry, configurations, optical activity, and molar ratio. A protein is a biomolecule made of more than 40 proteinogenic amino acids. Disulfide bonds and glycosylation sites are usually included. This is to ensure that you are accurately describing the sbstances in your submission. All work and data generated in precisionFDA is private until you share it.

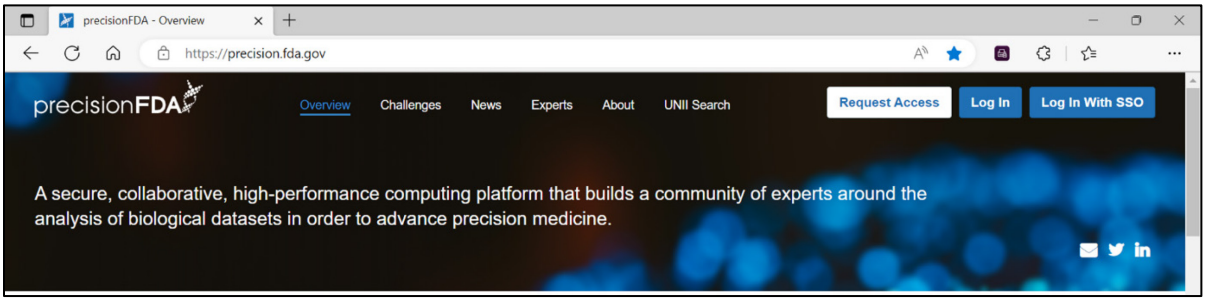
For specific questions about information not mentioned here or for troubleshooting, contact FDA-SRS@fda.hhs.gov. FDA-SRS staff will review your request.

Important information before starting

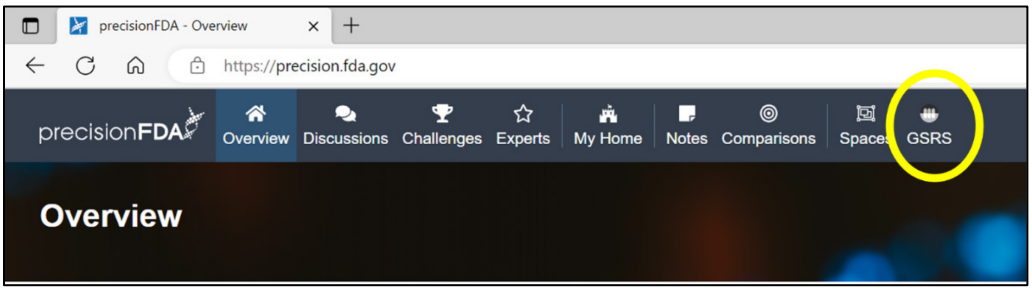
- Request access to precisionFDA by completing the form in this link: https://precision.fda.gov/request_access.
- Request access to a privately shared space to exchange chemicals and UNIIs with FDA-SRS by contacting FDA-SRS@fda.hhs.gov
- Many tools are available to help you draw or import the structure. Please familiarize yourself with Appendix I. Brief description of drawing application tools and features.
- The minimum requirements to complete a substance registration form are:
    1. One Definitional Reference. See Appendix II. Examples of Definitional References.
    2. One Name. Appendix III. Examples of Names and Type.
    3. Defined molecular composition or taxonomy
- Save registration form often. The precisionFDA browser times out after 15 minutes of being idle.

Search for existing record

1. Use an internet browser and go to https://precision.fda.gov/.



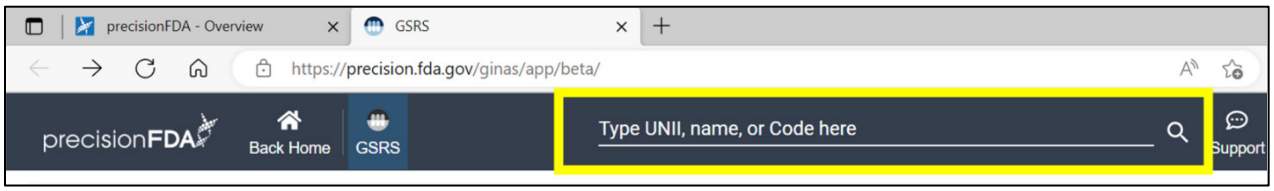
2. Click on the Log In (users outside of FDA) or Log In With SSO (badged FDA users) icon on the top menu bar.
3. Click on the GSRS icon on top menu bar.



4. Determine whether the record exists in GSRS by either:

Searching by UNII/Name/Code

Click on the Search text field next to the magnifying glass on the top menu bar. Type the UNII, name, or code. Click on the magnifying glass or press "enter" to start the search.



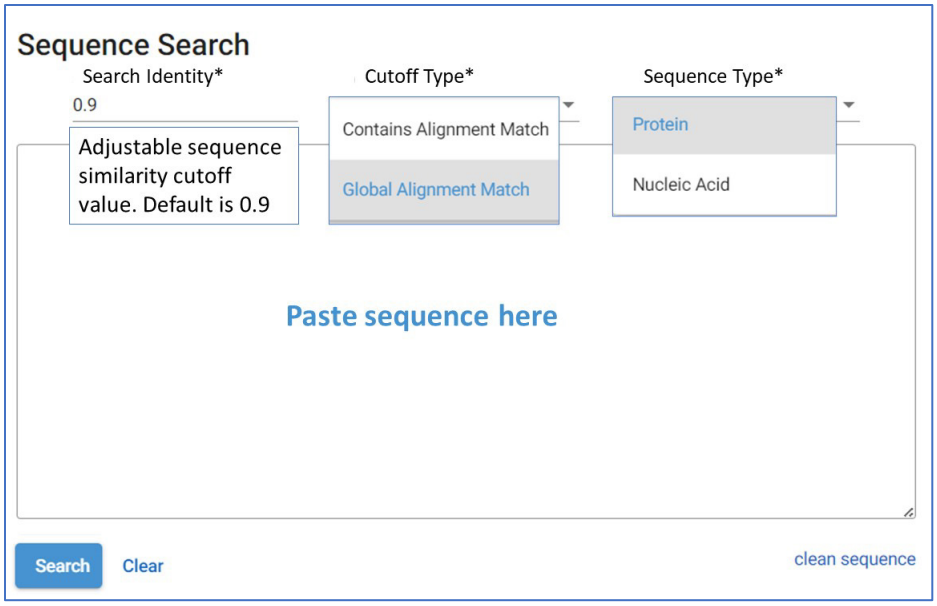
Searching by chemical structure

Click on the Structure Search icon. This will open a new window with a drawing application. Draw or import the structure. Click on the drop-down arrow of the "Search type" option and select a type of search. Click "Search" to submit query. *Note: Many tools are available to help you draw or import the structure. See* Appendix I. Brief description of drawing application tools and features.

Searching by sequence (protein or nucleic acid)

Copy the sequence you want to find. Note: For proteins, ONLY USE ONE-LETTER CODES. (e.g., The one-letter code of Alanine-Cysteine-Aspartic acid is A-C-D.) Paste the sequence into the text box. Click on clean sequence button. This will remove spaces, dashes, numbers, and delimiters from the text box. (e.g., A-C-D will convert to ACD in the window.) There are three parameters that need to be set (description below). Click Search.

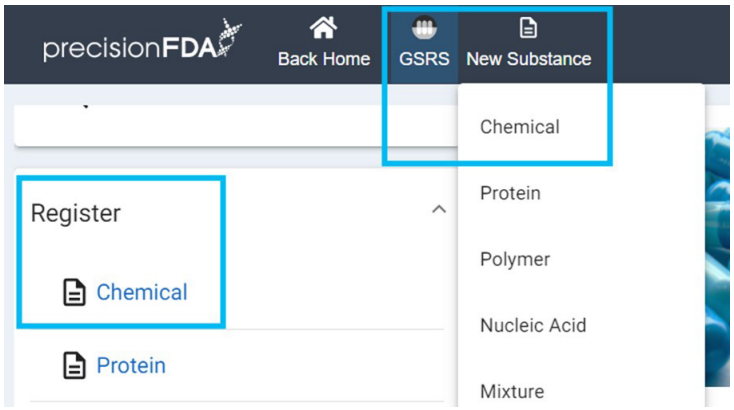
- Search identity cutoff - The default Search Identity cutoff is 0.9. The recommended cutoff is 0.98.
- Cutoff types
    - Contains Alignment Match - searches for specific sequences (proteins, RNA, DNA) within chains. This kind of search is useful when searching for a relatively short sequence motif which may be found in a much larger sequence with a few small edits.
    - Global Alignment Match - searches for sequences (proteins, RNA, DNA) similar to the complete query sequence. This kind of search is useful when searching for a complete sequence (e.g., a protein subunit, peptide, oligonucleotide, etc.) where the goal is to find a substance with a nearly exact complete sequence match.
- Sequence type - Select protein for amino acid sequences or nucleic acid for nucleotides.



5. If no records return, the substance needs to be registered and submitted to GSRS. A GSRS registrar will review the substance. Someone will contact you if they have questions. A UNII will be assigned after the substance is approved. Go to the next sections for instructions on registering a chemical or protein.

Registering chemicals

1. Go back to the GSRS homepage. On the left pane under Register or on the top navigation page under New Substance, click on Chemical. The chemical registration form will open.



2. Minimum requirements to complete chemical registration form are one reference, one name, and a new chemical structure:

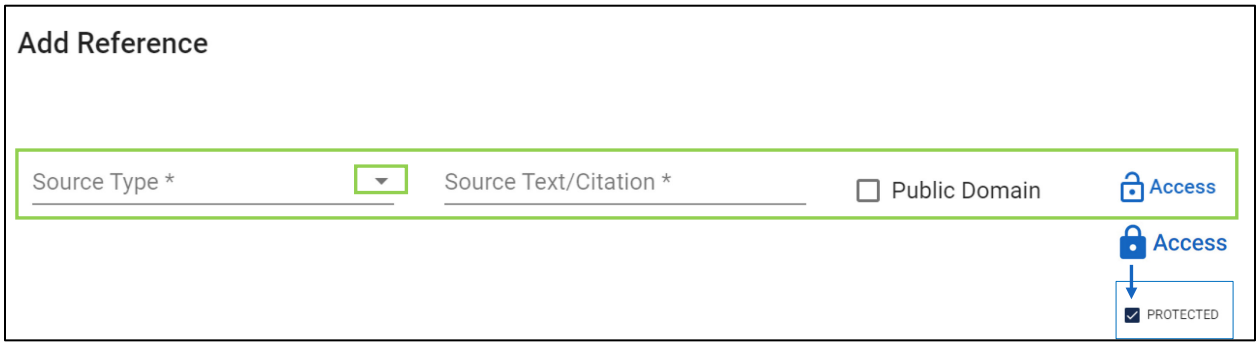
   a. One reference

   See Appendix II. Examples of Definitional References.

      1. Go to the Overview section
      2. Find Definitional References
      3. Click on Create new (+) button

      4. Click on the drop-down arrow and select a Source Type (required).
      5. Enter the Source Text/Citation (required).
      6. For proprietary references or to keep the reference confidential, click on the unlocked padlock Access icon, and click on the PROTECTED checkbox. The padlock icon will become a locked icon.

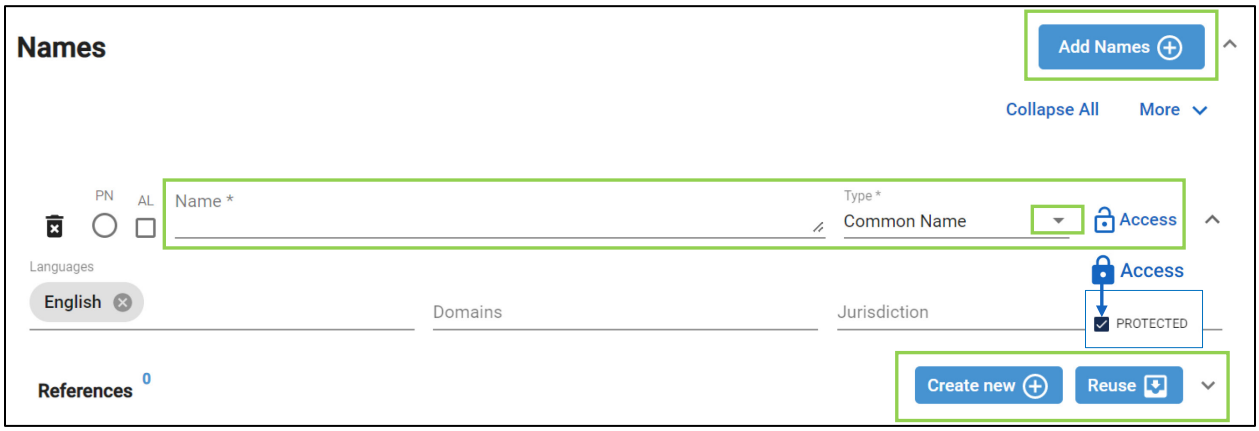
7. For public references, click on the Public Domain checkbox.



b. One name

See Appendix III. Examples of Names and Type.

1. Scroll down to the Names section.
2. Click on the Add Names (+) button.
3. Enter the name (required).
4. Click on the drop-down arrow and select a Type (required).
5. Add a reference. Click on the Create new (+) button or the Reuse (+) button to enter the reference of the name.
6. For proprietary names or to keep the name confidential, click on the padlock Access icon, and click on the PROTECTED checkbox. The padlock icon will become a locked icon.



c. New chemical structure

See Appendix I. Brief description of drawing application tools and features.

1. Scroll down to the Structure section.
2. Draw or import the structure. Resolve errors, if any.

Note. The molecular formula, stereochemistry, optical activity, and additional stereochemistry fields will automatically populate. You may adjust these to provide a more complete description of the chemical structure.

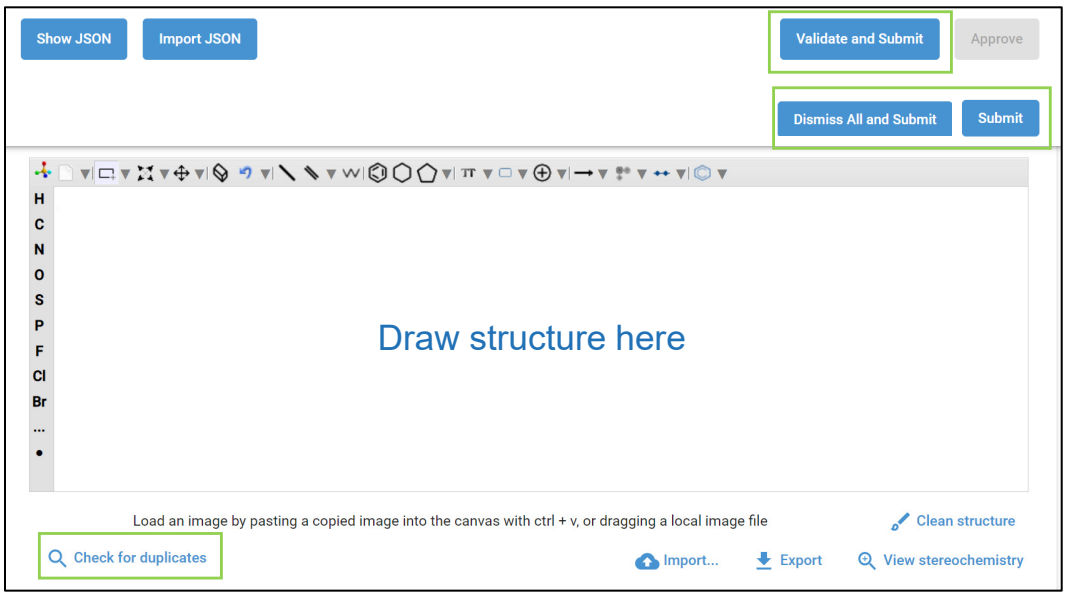
3. Click on the Validate and Submit button. Resolve error and warning messages, if any.

- 

Errors must be resolved before submitting.
- 

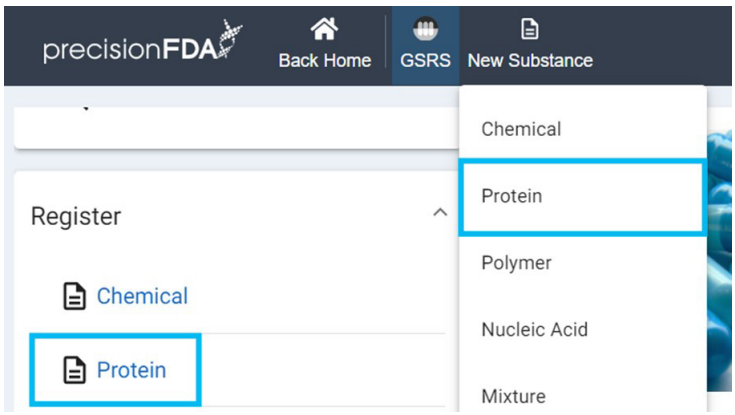
Warning messages may appear for various reasons (e.g., salt form of the substance is a record in GSRS; substance *xyz* appears to be a full duplicate). Please review each one. If they are intentional, click on the Dismiss All and Submit button.

4. Click on the Submit button.



Registering proteins

1. Go back to the GSRS homepage. On the left pane under Register or on the top navigation page under New Substance, click on Proteins. The protein registration form will open.



2. Similar to chemical registration, the minimum requirements to complete the protein registration form are one reference, one name, and a new amino acid sequence:

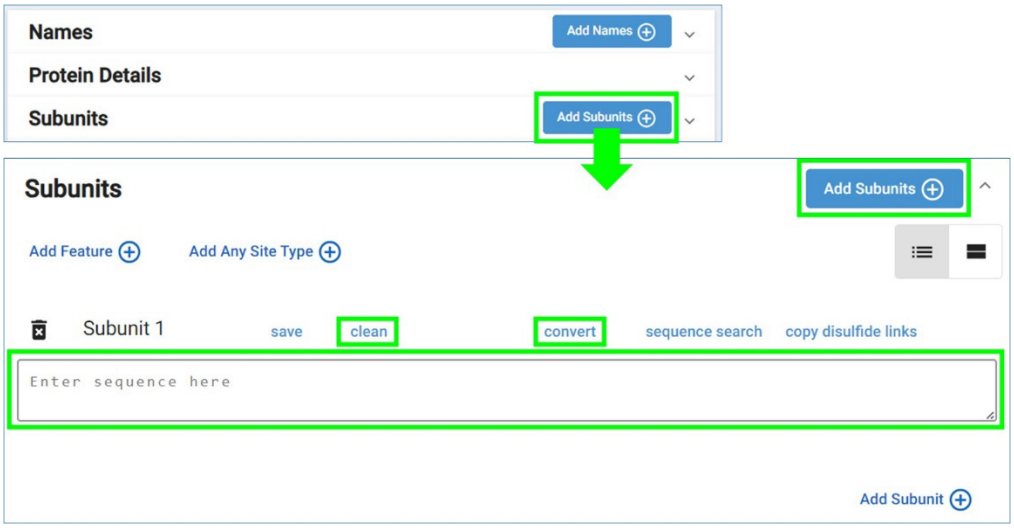
   a. One reference
   (Go to registering chemicals – one reference section for full instructions)
   See Appendix II. Examples of Definitional References.

   b. One name
   (Go to registering chemicals – one name section for full instructions)
   See Appendix III. Examples of Names and Type.

   c. New protein sequence*
   *Disulfide bonds and glycosylation sites are usually included.

      1. Scroll down to the Subunits section
      2. Click on Add Subunits (+)
      3. Compare the subunits. Add all subunits in order of decreasing length (e.g., heavy-heavy, light-light)
      4. Copy the sequence of the protein (one-letter code recommended. See Appendix IV. Twenty amino acid code conversion table)

5. Paste one-letter code sequence into text field. If you use three letter code, click Convert.
6. Click Clean
7. Repeat steps 2-6 for additional subunits

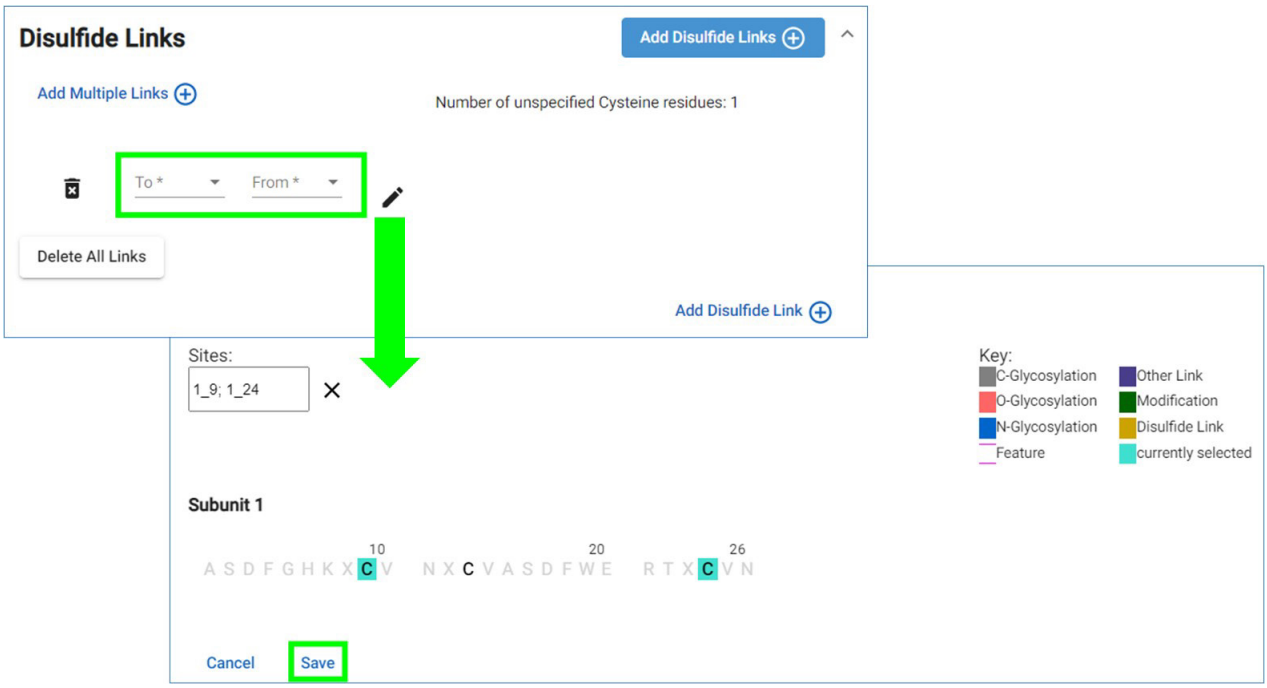


8. Click on the Validate and Submit button. Resolve error and warning messages, if any. Errors must be resolved before submitting. Warning messages may appear for various reasons Please review each one. If data entries were intentional, click on the Dismiss All and Submit button.
9. Click on the Submit button

*Disulfide Links

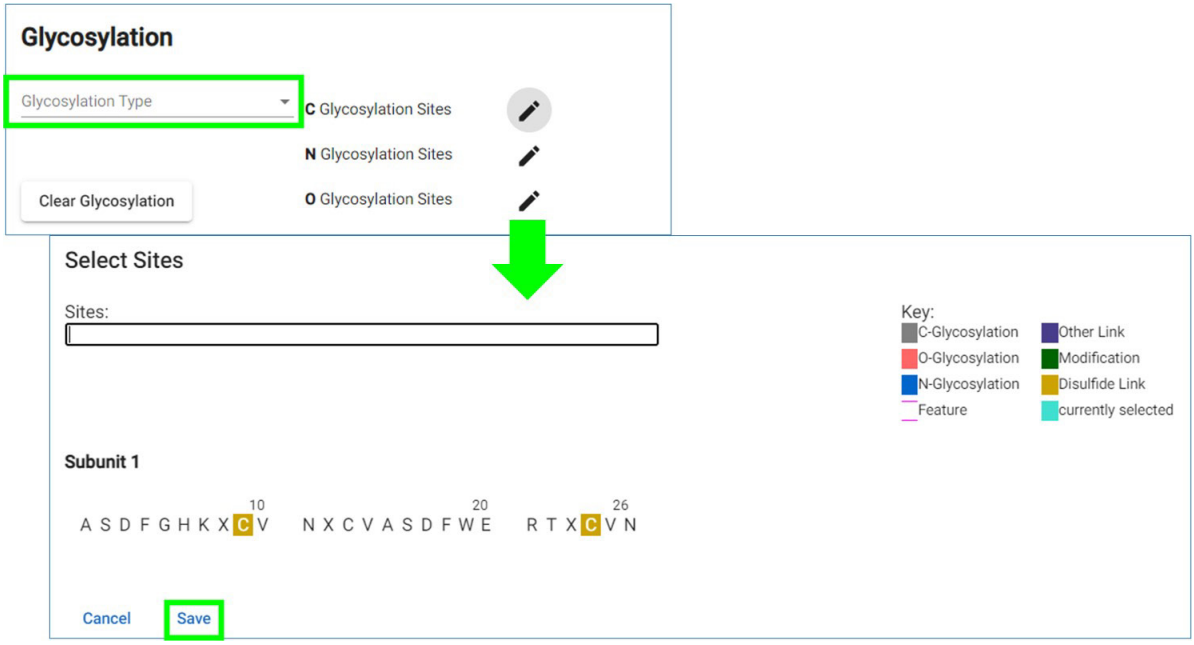
Add disulfide links one at a time with the pre-populated sites for pairing

1. Press the Add Disulfide Links (+) button
    a. Click on To* and From* drop-down arrows to select sites.
2. Repeat for each pair
3. At any time, if you need to start over, click the trashcan icon or Delete All Links Delete All Links button



*Glycosylation

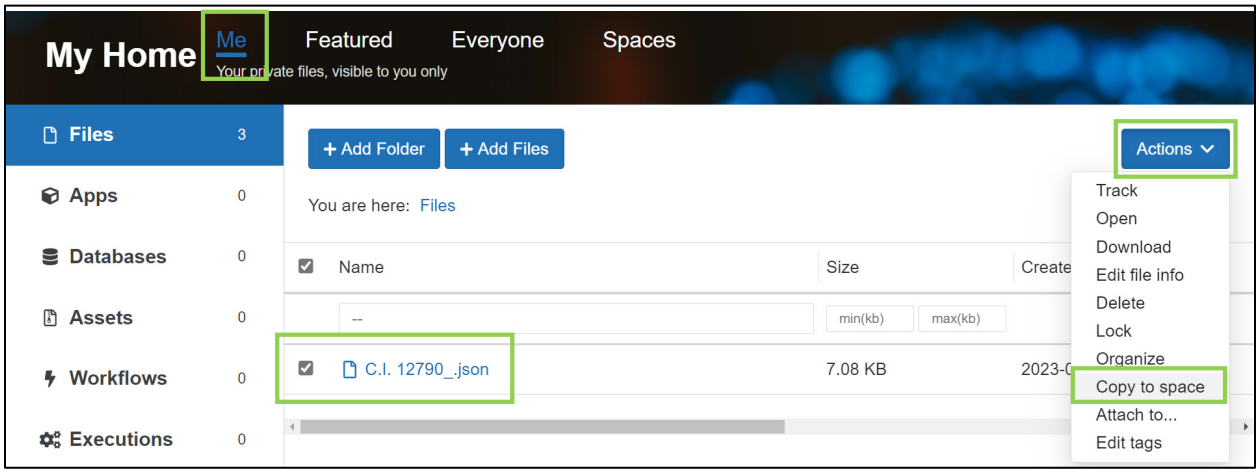
1. Click on Glycosylation section to expand the card
2. Optional. Select the Glycosylation Type. The choices are based on the controlled vocabulary of your instance.
3. For C, N, and O Glycosylation Sites, click pencil icon and the current sequence selector will be displayed
4. Sites should be in the form of subunit_residue, use hyphen "-" to indicate a range. This is automatically applied if you select the sites within the pop-up window.
5. Click Save

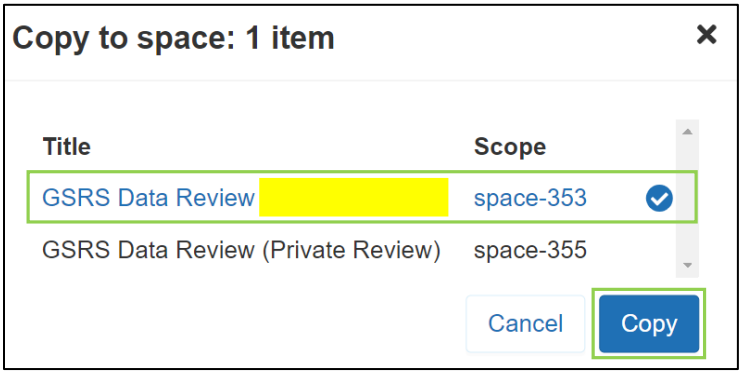


Sharing substance with GSRS

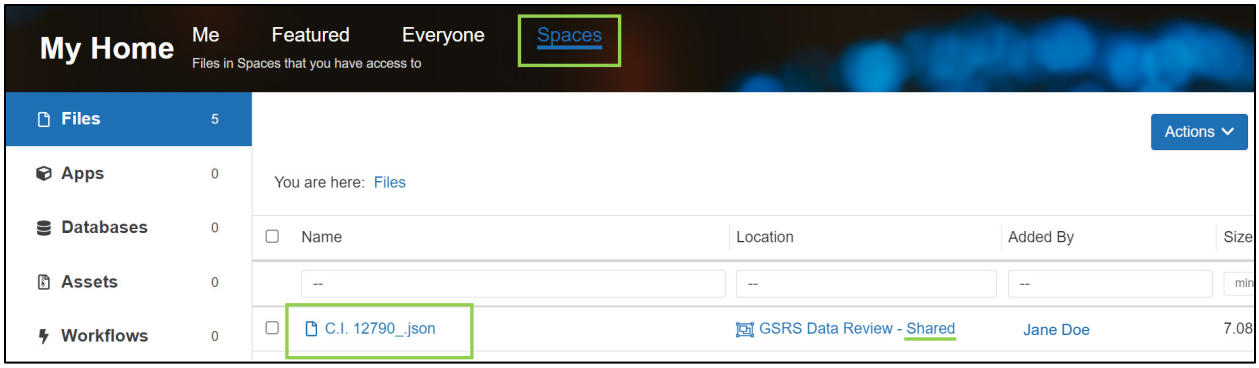
1. Click on the My Home (castle) icon 

to navigate to shared spaces. Your space is the Me tab. This space contains your private files and is visible only to you.
2. In the Me side pane, Click on Files.
3. Click the checkmark box of the files you want to share.
4. Click on the Actions button. Select the "Copy to space" option. A window will open. Select the Shared space assigned to you (email FDA-SRS@fda.hhs.gov if you do not have one). The text will turn blue and have a checkmark on the right. Click Copy.





5. Click on the Spaces tab and confirm the file was shared. *Note: An automated message is generated to notify FDA-SRS that a file was shared.*



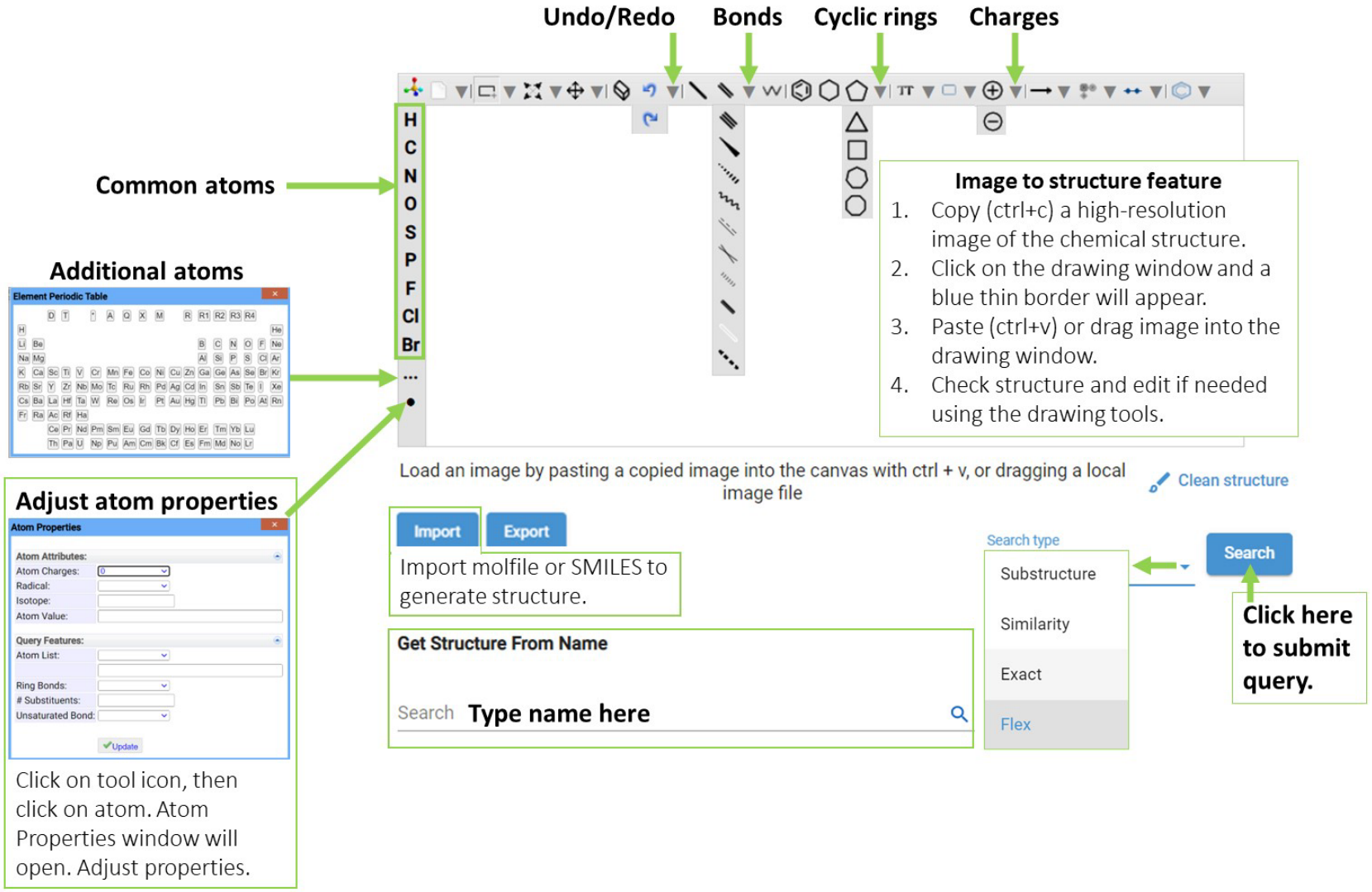
6. FDA-SRS will directly respond with UNIIs if there are no issues with your submission.

Additional support

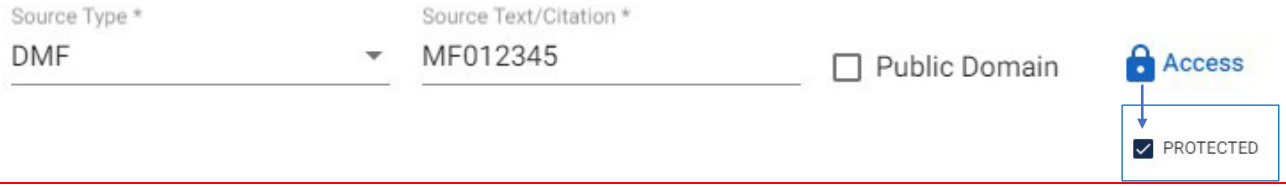
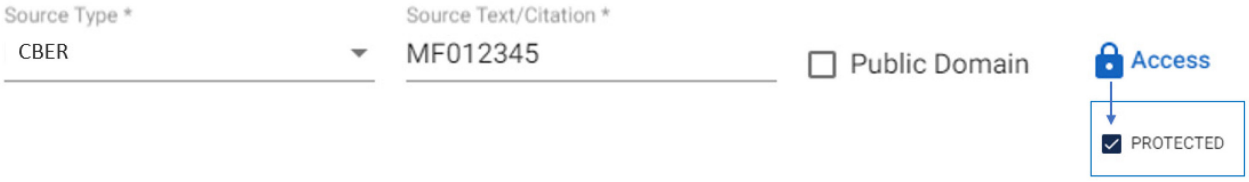
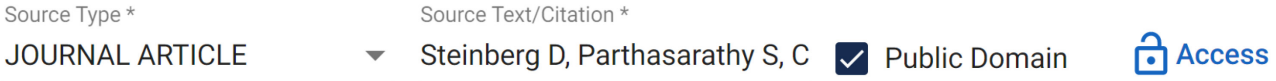
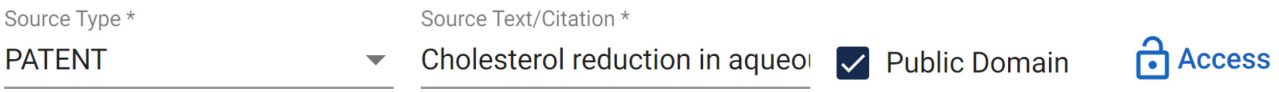
- For questions about obtaining a UNII for a substance, please contact FDA-SRS@fda.hhs.gov.
- For complex substances that require additional data elements to describe or are difficult to represent with a single chemical structure, please contact FDA-SRS@fda.hhs.gov for assistance.

Appendices

Appendix I. Brief description of drawing application tools and features



Appendix II. Examples of Definitional References

Source Type (required)	Source Text/Citation (required)	Screenshots
DMF, NDA, ANDA, or IND	MF012345 NDA010203 • No dashes or spaces • Include leading zero(s)	Set Access to PROTECTED for proprietary or confidential chemicals/references! 
***CBER	MF012345 NDA010203 • No dashes or spaces • Include leading zero(s)	CBER APPLICATIONS ONLY  *****Note: CBER and CDER application numbers are not unique. (e.g., There are applications in CBER and CDER with the same exact number!) For CBER applications only, use "CBER" as Source Type.**
JOURNAL ARTICLE	Steinberg D, Parthasarathy S, Carew TE, Khoo JC, Witztum JL. Beyond cholesterol. *New England Journal of Medicine.* 1989 Apr 6;320(14):915-24.	
PATENT	Cholesterol reduction in aqueous emulsions with cyclodextrins. By: Oakenfull, David George; Sidhu, Gurcharn Singh; Rooney, Michael Laurence. World Intellectual Property Organization, WO9111114 A1 1991-08-08.	

Appendix III. Examples of Names and Type

Type	Name	Format
Systematic Name	N-(4-Hydroxyphenyl)acetamide	Sentence case
Brand Name	TYLENOL	UPPERCASE
Common Name	Acetaminophen	Sentence case
Official Name	Acetaminophen	Proper Case
Official Name	Paracetamol	Proper Case
Code Name	NSC-109028	Standardize to UPPERCASE letters, a hyphen in between, then numbers.

Appendix IV. Twenty amino acid code conversion table

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic acid	Asp	D
Glutamic acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y