Quick Guide to Creating an SD File for eCTD Submissions

Marlene Kim, PhD

Chemist/Computational Toxicologist
US FDA
Office of the Commissioner/ODT/ODAR/Health Informatics Staff

SBIA DMF Workshop: GDUFA III Enhancements and Structured Data Submissions
November 30, 2022

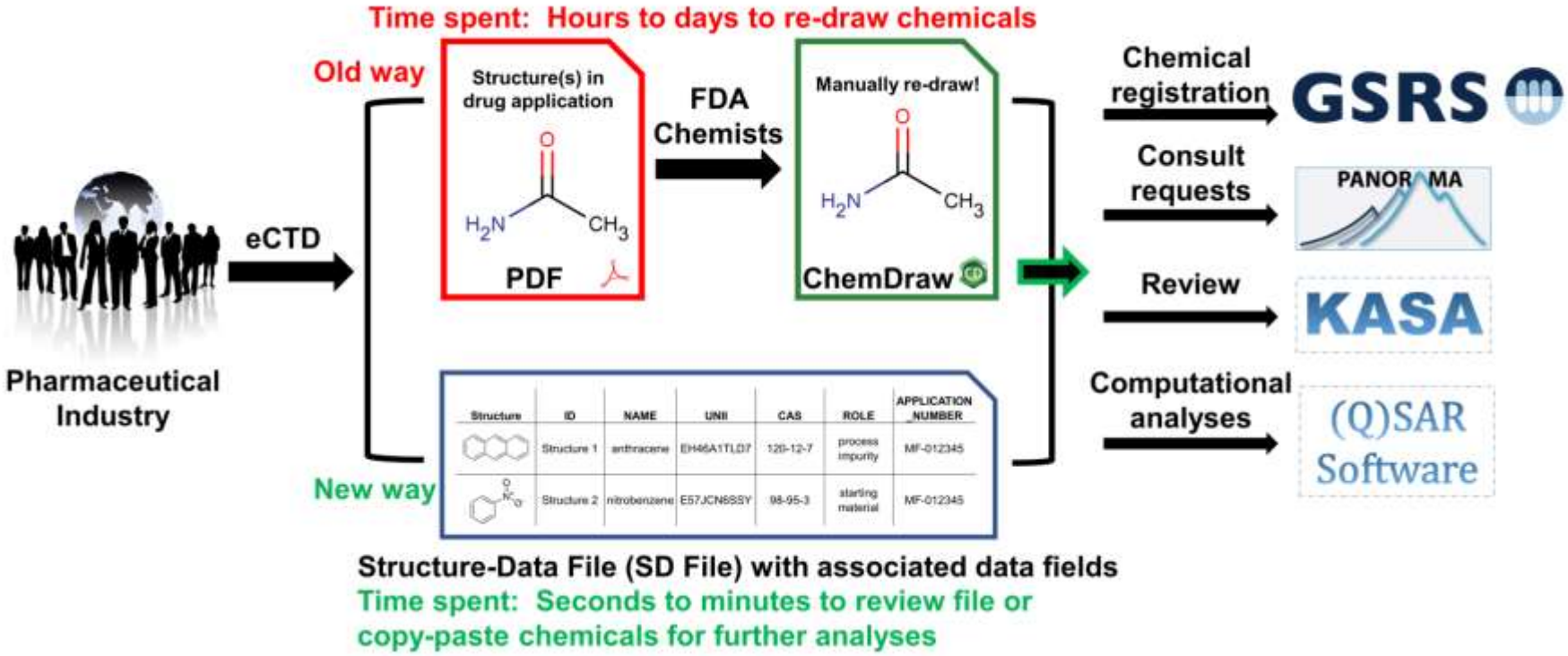
Disclaimer

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services. This presentation reflects the views of the author and should not be construed to represent the Food & Drug Administration views or policies.

Creating and submitting a structure-data file (SD File)

- Define an SD File
- Format
- File locations in electronic common technical document (eCTD)
- Available software
- Special chemical structure scenarios
- Additional Support

What happens to the chemical structure at FDA?



Structure-Data File from Sponsors' Submissions Promotes Rapid Chemical Registration and Computational Analyses, Scientific Computing Days 2021, https://www.fda.gov/media/151913/download

Moving forward

- In 2021, SD Files became an acceptable file format in the eCTD.
- As of October 21, 2022, FDA received 65 SD files with an average of 30 structures per file.

What is an SD File?

- SD File – chemical-data file format that can associate data with a chemical structure
- The chemical structure uses the MOLfile[1] format, which describes the atoms, bonds, connectivity, coordinates, and properties of the structure using strings of text
- File extension is “.sdf”
- Also known as SDF

[1]Chemical table file – Molfile. https://en.wikipedia.org/wiki/Chemical_table_file#Molfile

The *computer* interprets the file and translates the string/text into a graphical structure and table.

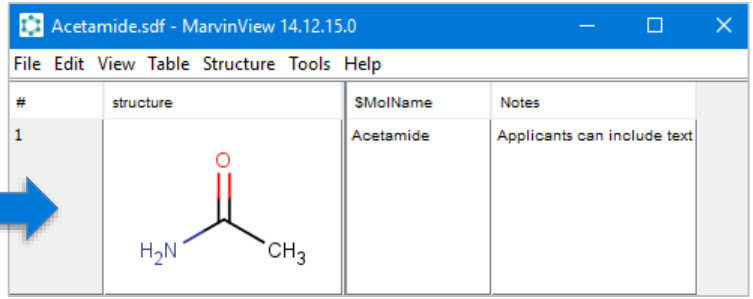
SD File text contents

```
Acetamide


  4  3  0  0  0  0  0  0  0  0999 V2000
    0.0000   -0.2062    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.6188    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.7145   -0.6188    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    0.7145   -0.6188    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0  0
  1  3  1  0  0  0  0
  1  4  1  0  0  0  0
M  END
>  <Notes> (Acetamide)
Sponsors can include text

$$$$
```

Graphical chemical structure and data table



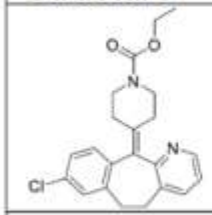
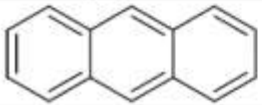
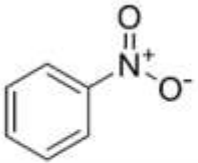
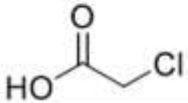
Suggested file type and name

- Please use V2000 format
- One SD file containing more than one chemical structure will be expected
- File name - SDF_DMF_012345_Loratidine.sdf
  - Application type and number (i.e., DMF 012345, ANDA 012345)
  - Name of the Active Pharmaceutical Ingredient (API)

Columns/data items linked to each chemical structure

- **NAME** - Name(s) of the chemical as referenced in the application
- **APPLICATION_NUMBER** - Application type prefix (e.g., DMF, ANDA, NDA, IND, or BLA) before the number (e.g., DMF 012345 or DMF 12345)
- **ID** - Unique identifier for cross-references
- **ROLE** – e.g., active ingredient, process impurity, intermediate, degradant, metabolite, and/or starting material
- **UNII** - Unique Ingredient Identifier, if available from https://precision.fda.gov/uniisearch and https://gsrs.ncats.nih.gov/ginas/app/beta/
- **CAS** - If available, from SciFinder - https://scifinder.cas.org
- **NOTES** – if needed. Notes specify/qualify something about the substance that is not easily discernable from the structure alone

Sample SD File with columns/data items and associated data

Structure	ID	NAME	UNII	CAS	ROLE	APPLICATION_NUMBER
	Structure 4	loratadine	7AJO3BO7QN	79794-75-5	active ingredient	DMF-012345
	Structure 1	anthracene	EH46A1TLD7	120-12-7	process impurity	DMF-012345
	Structure 2	nitrobenzene	E57JCN6SSY	98-95-3	starting material	DMF-012345
	Structure 3	2-chloroacetic acid	5GD84Y125G	79-11-8	starting material	DMF-012345

Note – Helpful if the active ingredient is the first record

Types of chemicals to include

- Impurities of drug substance and drug product
- Leachables exceeding the analytical evaluation threshold (AET) (calculated from the safety concern threshold (SCT))
- Impurities evaluated by (Q)SAR to support selection of a surrogate molecule or predict potential mutagenicity

Place SD File(s) in appropriate section of eCTD

- Module 3 has several sections which may contain information relating to drug substance (3.2.S) or drug product (3.2.P)
- Impurities evaluated by (Q)SAR
  - 3.2.S.3.2 for drug substance
  - 3.2.P.5.5 for drug product

Place SD File(s) in appropriate section of eCTD

(continued)

- 3.2.R Supportive files placed here for any of the sections above should be properly hyperlinked. One SD File may be suitable when substances pertain to multiple sections
- Nonclinical study reports submitted in Module 4 may include hyperlink(s) to SD Files submitted in Module 3 to assist cross-references
- May also submit to other sections, if needed, e.g., 3.2.S.6 and 3.2.P.7 Container Closure System

Software available

Commercial

- ChemAxon Marvin
- Microsoft Excel with ChemDraw Add-in from ChemOffice
- Instem Leadscope SDF Editor
- Molecular Operating Environment

Publicly available and free

- Open Babel
- DataWarrior
- KNIME SDF Writer
- RDKit
- CDK

Special chemical structure scenarios

- Salts, mixtures, 3-D configurations, isotopes, or stereochemistry, please use the SD File Guide

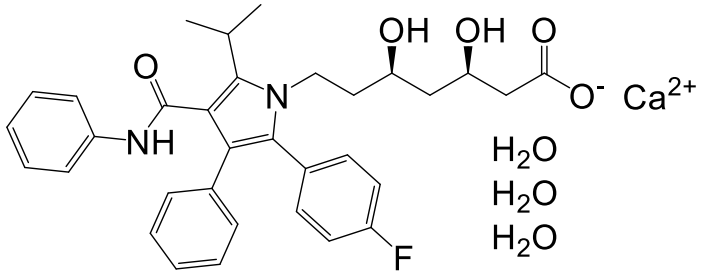
- For complex substances other than small molecules (e.g., proteins, nucleic acids) please contact FDA-SRS@fda.hhs.gov for a UNII

Include stoichiometry explicitly

Magnesium chloride ($MgCl_2$)

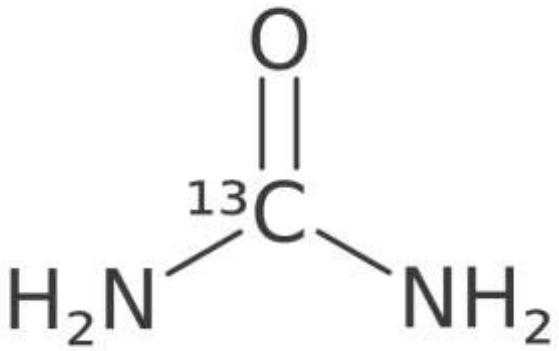
Cl^- Mg^{2+} Cl^-

Atorvastatin calcium trihydrate

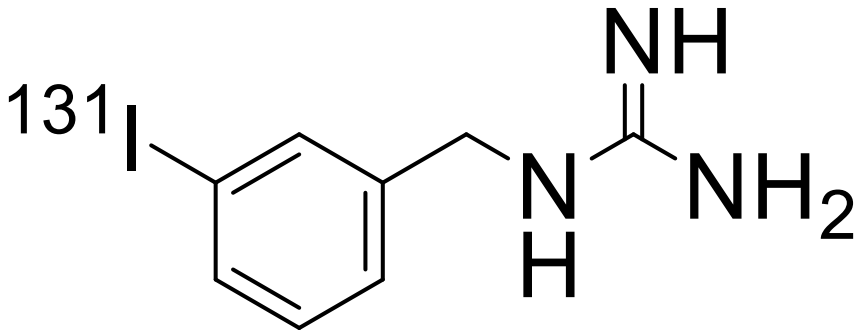


Specify isotopes

Urea C^{13}

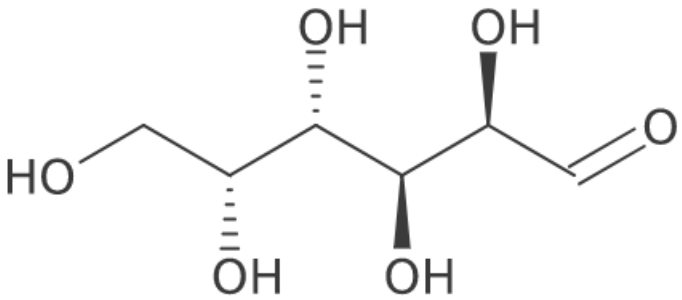


Iobenguane I-131

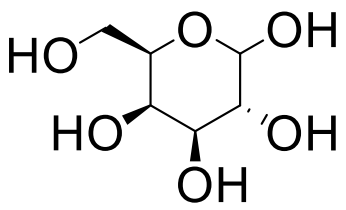


Use planar drawings for cyclic structures and ring systems

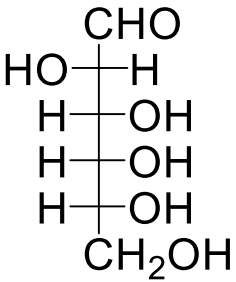
D-Galactose



Preferred drawing

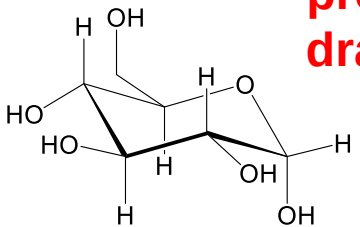


Haworth



Fischer

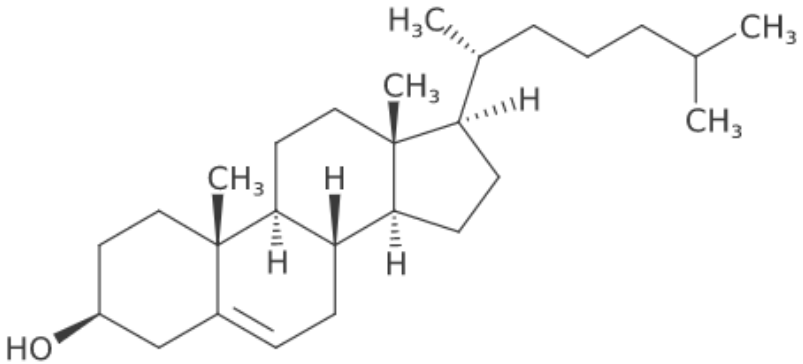
NOT preferred drawings



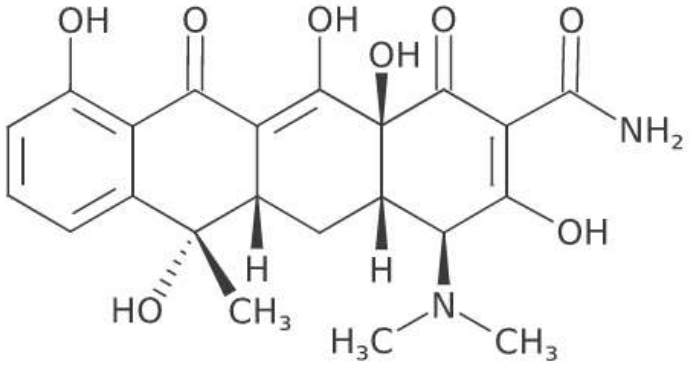
Chair

Use wedge and dash bonds to indicate stereochemistry

Cholesterol



Tetracycline



Additional support

- FDA-SRS@fda.hhs.gov -Requesting UNII, formatting SD Files, drawing complex structures
- DMFOGD@fda.hhs.gov - DMF submissions
- esub@fda.hhs.gov - CDER related to eCTD submissions
- esubprep@fda.hhs.gov - CBER related to eCTD submissions

We've come a long way

- Standardized data fields meets FDA's cheminformatics needs
- Errors associated with redrawing structures are eliminated
- Process increases accuracy, quality review efficiency, and eliminates redundancy
- Impacts drug substance and drug product reviews, (Q)SAR analyses, and GSRS registration efforts

Submit a comprehensive SD File with your submission!

Resources

- Dalby A, et al. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. J. Chem. Inf. Comput. Sci. 1992, 32, 3, 244–255.
- SD File Quick Guides
  - DMF submissions
  - Other types of submissions (ANDA, NDA, BLA)
- Electronic Common Technical Document (eCTD) | FDA
- Study Data Standards Resources | FDA

Thank you!

SD File Implementation Team

Tyler Peryea

Barbara Scott

David Green

Naomi Kruhlak

GSRS Staff and Collaborators

Frank Switzer

Larry Callahan

Elaine Johanson

Dammika Amugoda

Marian Nkeng

David Skanchy

Charles Robertson

Andre Raw

Geoffrey Wu

Extractables and Leachables Subcommittee

Data Standards Board

Jonathan Resnick

Heather Crandall

Helena Sviglin

Norman Schmuff

Jeff Florian

Mary Anne Slack

Ray Wang