Improving (Q)SAR Review with Structure-Data Files (SD Files)

Naomi L. Kruhlak, Ph.D.

Scientific Lead, Computational Toxicology Consultation Service

Division of Applied Regulatory Science
Office of Clinical Pharmacology
Office of Translational Sciences
Center for Drug Evaluation and Research
US Food and Drug Administration

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

FDA Disclaimer

The findings and conclusions in this presentation reflect the views of the author and should not be construed to represent FDA's views or policies.

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.

Outline

- (Q)SAR Modeling Basics
  - How (Q)SAR works
  - Chemical structure formats
  - Benefits of structure-data (SD) file format
- ICH M7(R1) Guideline
  - (Q)SAR recommendations
  - Importance of structural accuracy
  - Structure-linked databasing

U.S. FOOD & DRUG ADMINISTRATION

(Q)SAR Basics

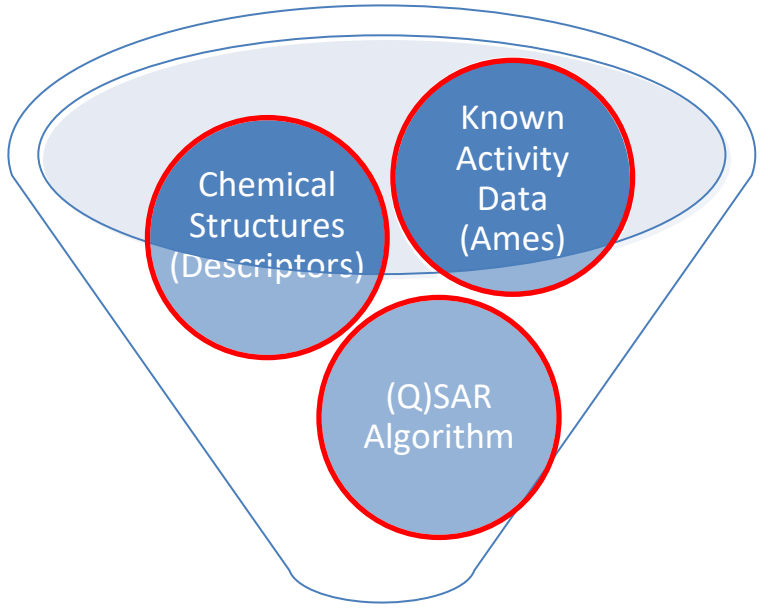
(Q)SAR Modeling: What is it?

- (Q)SAR = (Quantitative) Structure-Activity Relationship
  - Modeling identifies associations between attributes of chemical structures and biological activity (e.g., mutagenicity)
  - General assumption: Similar molecules exhibit similar chemical and biological properties

  ⇨ Toxicity can be explained by chemical structure
- Model learns from the results of actual laboratory testing
  - Use a computer to examine "pieces" of chemical structures to find those associated with activity ➔ structural alerts
  - Can also identify attributes that mitigate activity
- Model can be used to make a prediction of a new chemical's toxicity based on its structure
  - Fill data gaps when empirical data are unavailable or inadequate

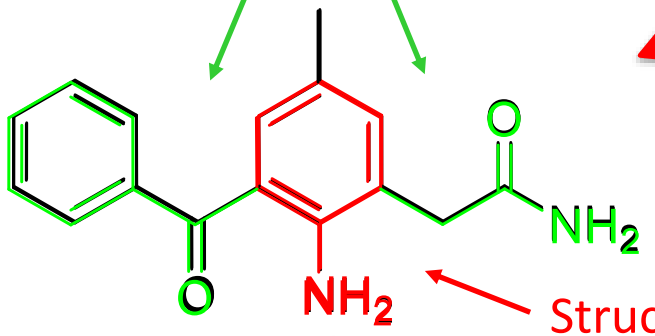
QSAR – quantitative – statistical-based model
SAR – qualitative – expert rule-based model
} (Q)SAR

Building and Applying a (Q)SAR Model



(Q)SAR Model

Mitigating Features



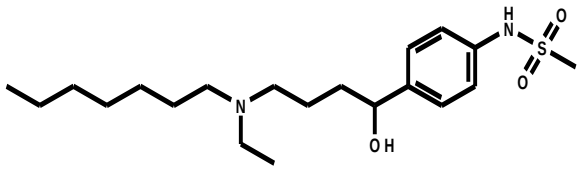
Structural Alert

Computer-readable structure!

Activity Prediction (Negative)

Chemical Structure Formats

- Structural representation



2D

3D

Ibutilide

- Structure formats
  - SMILES
    - ➔ CCCCCCCN(CC)CCCC(C1=CC=C(C=C1)NS(=O)(=O)C)O
  - InChI
    - ➔ InChI=1S/C20H36N2O3S/c1-4-6-7-8-9-16-22(5-2)17-10-11-20(23)18-12-14-19(15-13-18)21-26(3,24)25/h12-15,20-21,23H
  - Molfile/SD File
    - ➔ 2858.mol or Drugs.sdf

Single Structure (Molfile)

Multiple Structures (SD File)

Most Common Format

SD File Benefits

- SD File:
  - Enables batch processing of 10s to 100s of structures
  - Open standard – does not require a commercial software to open or create files
  - Small file size (ASCII text) – makes data transfer facile
- Format works well for regulatory submissions of drug impurity structures under ICH M7(R1) with associated data fields:
  - Name, Application Number, ID, Role, UNII, CAS, Notes*

```
>  <Name>
p-Chloroaniline

>  <CAS Number>
106-47-8

$$$$



  6  5  0  0  0  0  0  0  0  0999 V2000
    4.4736    2.5260    0.0000 Cl  0  0  0  0  0  0  0  0  0  0  0  0
    5.1880    2.9385    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.9025    2.5260    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.1880    3.7635    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    6.6170    2.9385    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.9025    1.7010    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  1  0  0  0  0
  2  4  2  0  0  0  0
  3  5  1  0  0  0  0
  3  6  1  0  0  0  0
M  END
```

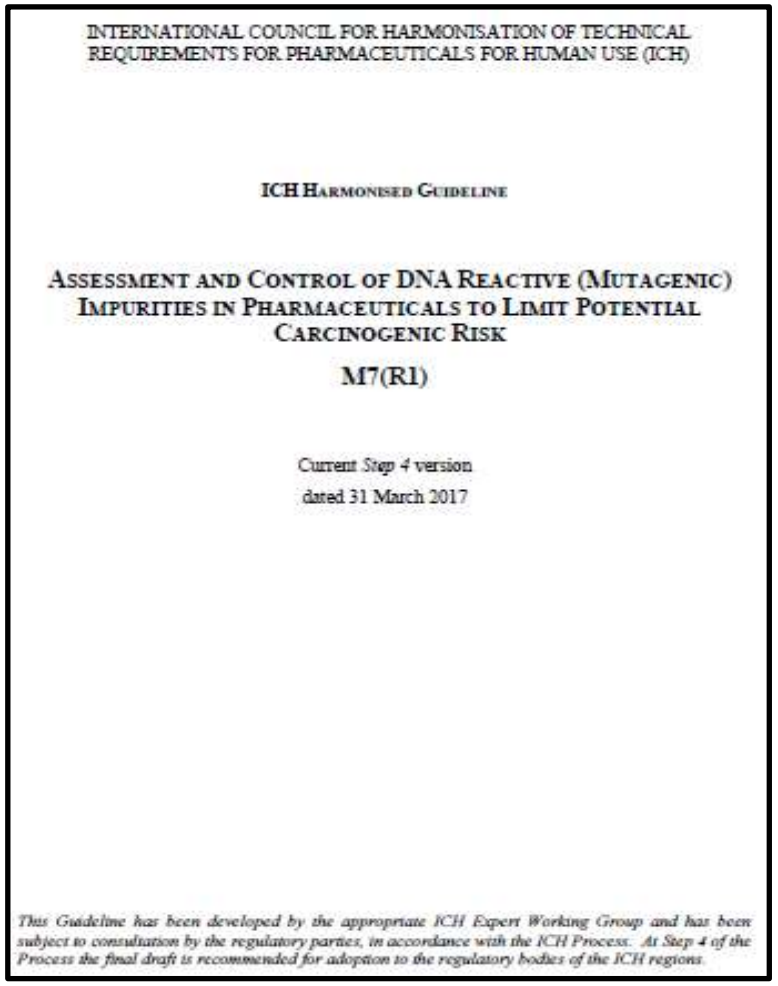

*For more details, see presentation by Marlene Kim, entitled, "Quick Guide to Creating an SD File for eCTD Submissions"

U.S. FOOD & DRUG ADMINISTRATION

ICH M7(R1) (Q)SAR Assessment of Drug Impurities

ICH M7 Pharmaceutical Regulatory Guideline

- Published in June 2014, revised (to version R1) in March 2017, currently under revision to R2
- Title: ASSESSMENT AND CONTROL OF DNA REACTIVE (MUTAGENIC) IMPURITIES IN PHARMACEUTICALS TO LIMIT POTENTIAL CARCINOGENIC RISK
- Describes how a hazard assessment should be conducted on a pharmaceutical impurity to classify it as mutagenic or non-mutagenic based on experimental data and/or (Q)SAR predictions
  - (Q)SAR models should predict the outcome of Ames assay
  - Use two complementary modeling methodologies: statistical-based and expert rule-based
  - Models should be consistent with OECD validation principles
  - If no structural alerts, no further genetox testing is required



INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE (ICH)

ICH HARMONISED GUIDELINE

ASSESSMENT AND CONTROL OF DNA REACTIVE (MUTAGENIC) IMPURITIES IN PHARMACEUTICALS TO LIMIT POTENTIAL CARCINOGENIC RISK

M7(R1)

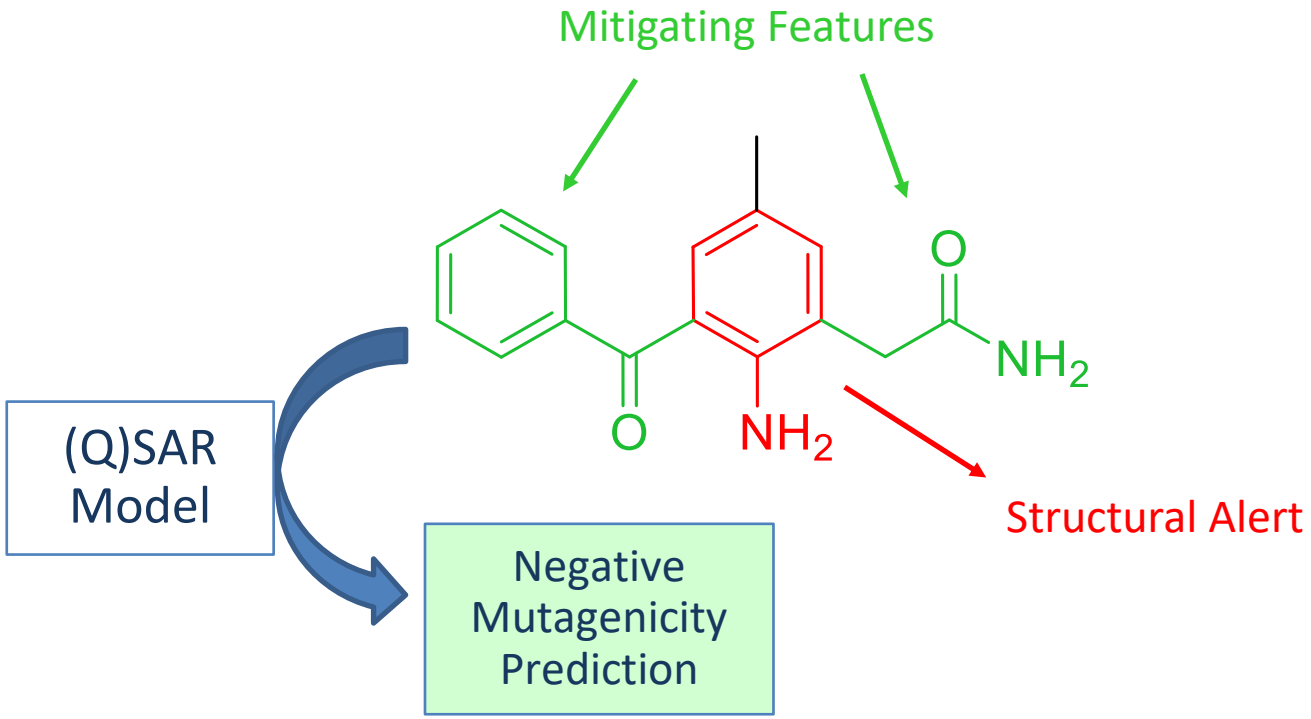
Current *Step 4* version
dated 31 March 2017

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the ICH regions.

International Council for Harmonization (ICH) M7(R1), 2017. https://database.ich.org/sites/default/files/M7_R1_Guideline.pdf; Organisation for Economic Co-operation and Development (OECD), 2007. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2

Application of Expert Knowledge

Model output "… can be reviewed with the use of expert knowledge in order to provide additional supportive evidence on relevance of any positive, negative, conflicting or inconclusive prediction and provide a rationale to support the final conclusion."



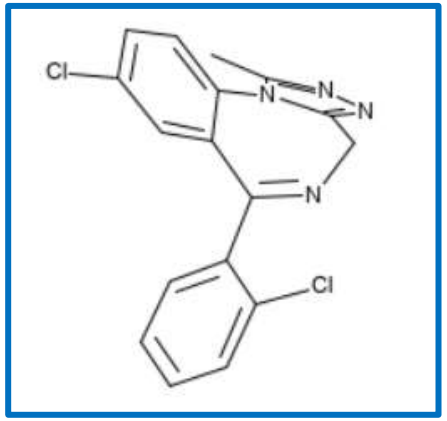
- Identify and interpret alerting portion of the molecule
- Consider mechanism of reactivity, where possible
- Assess training set structures used to derive alerts and mitigating features [review model output]
- Consider data from structurally similar compounds (analogs) not used by the model [search supplemental databases]

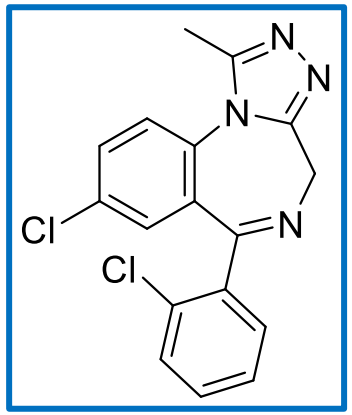
Expert knowledge is applied to all (Q)SAR analyses conducted in-house by FDA/CDER

(Q)SAR Review

- Industry (Q)SAR predictions for impurities in DMFs are triaged by review staff
- Any questionable predictions/conclusions or inadequately documented analyses (including application of expert knowledge), are sent forward for an internal (Q)SAR consultation
  - Electronically-readable structure needed for analysis
- In the absence of an electronic structure, impurity and corresponding API need to be redrawn from submitted image
  - Quality of submitted images varies – takes time to decipher them
  - For example:

Original Submission





Final Structure

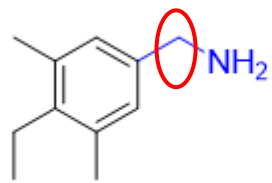
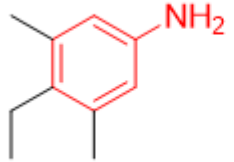
But redrawing takes extra time and can introduce errors!

Structural Accuracy Matters

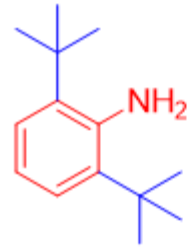
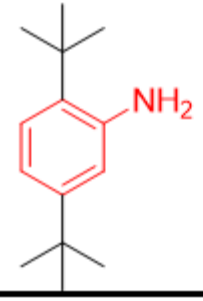
Prediction is based on chemical structure

- Incorrect structure = invalid prediction!
- Slight changes in structure can lead to different predictions

Addition of a methylene group

Structure	Predictions		
	-	-	-
	+	Equ	+

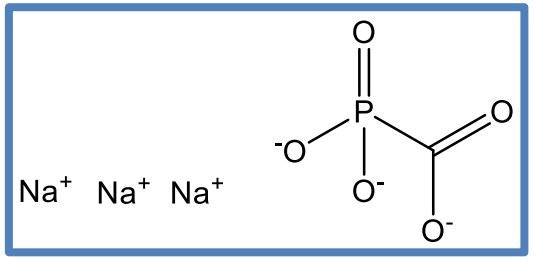
Position of substituents

Structure	Predictions		
	-	-	-
	+	-	-

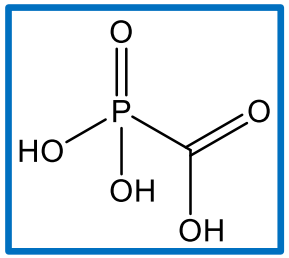
Electronic submission of structures in SD File improves accuracy and efficiency!

Chemical Structure (Q)SAR Pre-Processing

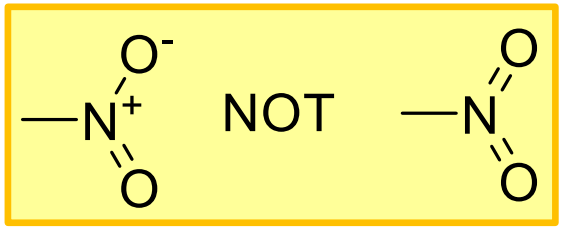
- If structures are submitted as an SD File, only minimal pre-processing is needed by the agency for (Q)SAR analysis
- For (Q)SAR analysis, structural formatting may be needed:
  - Salts and counter-ions stripped, and some charges may need to be neutralized (e.g., $-COO^-$, $-NH_2^+$)
  - Some functional groups may need to be standardized, e.g., nitro groups



Original Submission

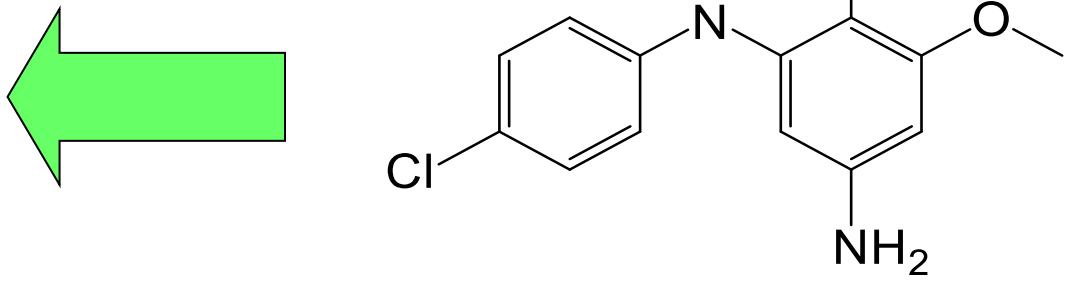


Final Structure



Structure-Linked Databasing Supports Review

Yes



No

- Has FDA/CDER reviewed this chemical before?
- Are experimental data available?
- Have we previously performed a (Q)SAR analysis?
- Are there data for structurally related compounds (for expert review)?

Structure-based searching facilitated by electronically-readable structures in SD File format

Databasing is Essential for (Q)SAR Review!

- The Computational Toxicology Consultation Service (CTCS) team maintains an internal database of ~35K chemical structures, published toxicological data and past Agency (Q)SAR analysis reports
  - Structures can be imported directly from SD File format
- Reviewers have the ability to search the database for past Agency (Q)SAR analyses
  - Eliminates duplicate consult requests for previously evaluated compounds
- CTCS reviewers use the database of toxicology data to support their (Q)SAR analyses
  - Facilitates application of expert knowledge to (Q)SAR predictions--rapid retrieval of structural analogs using exact, substructure and global similarity queries
  - Rapid retrieval and review of past regulatory decisions for related compounds—ensures consistency

Summary

- (Q)SAR models make predictions of toxicity based on chemical structure
  - Require structures in computer-readable format
- SD File format is commonly-used in cheminformatics
  - Encodes multiple chemical structures and associated data in computer-readable format
  - Suitable for regulatory submissions of drug impurity structures under ICH M7(R1)
- In a (Q)SAR review workflow, SD Files:
  - Eliminate the need for redrawing of structures ➔ reduces introduction of structural errors
  - Facilitate application of expert knowledge to (Q)SAR predictions by enabling structure-based searching
  - Enable direct databasing of impurity structures for tracking purposes

Drug impurity structures submitted in SD files improve efficiency and accuracy of (Q)SAR analysis review

Acknowledgements

The CTCS Team

Patra Volarath

Jahan Cooper

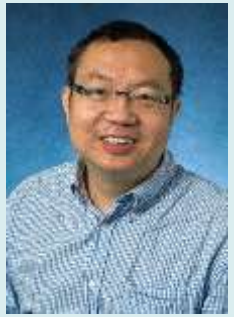
Franny Sweder Gold

Govindaraj Kumaran

Benon Mugabe

Barbara Scott

Diandian Shen



Jian Yang

Andy Zych

FDA Collaborators

David Green
Marlene Kim
Tyler Peryea
Andre Raw
Kirk Roy
David Skanchy

External Collaborators

Leadscope Inc.
Lhasa Limited
MultiCASE Inc.


FDA
U.S. FOOD & DRUG
ADMINISTRATION