Application of (Q)SAR and Expert Knowledge for ICH M7 Impurity Classification

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 and Drug Substance Workshop - March 3, 2021

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
  - How it works
  - Published alerts vs. (Q)SAR models
  - (Q)SAR software
  - Applicability domain
  - Structural limitations
- ICH M7(R1) Guideline
  - (Q)SAR recommendations
  - Application of expert knowledge
- (Q)SAR Reporting
  - Report components
  - Example format
  - Special considerations

U.S. FOOD & DRUG ADMINISTRATION

(Q)SAR Basics

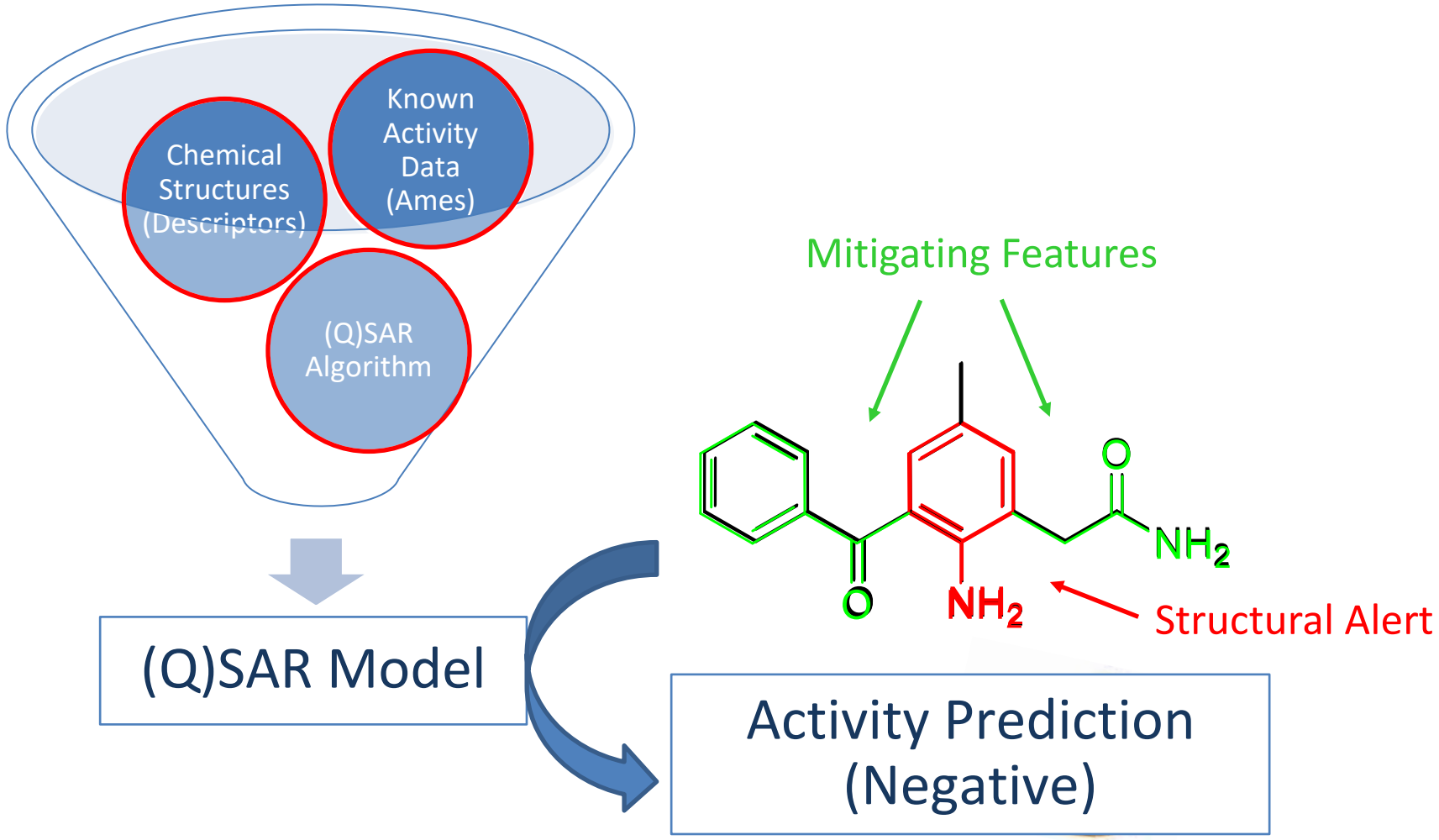
(Q)SAR Modeling: What is it?

- (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 a (Q)SAR Model



(Q)SAR Methodologies

- Statistical-based models
  - Constructed through machine learning
  - Use a large, classic training set of known examples
  - Identify statistical relationships between chemical features and activity
  - Algorithmically extract structural alerts and mitigating patterns
- Expert rule-based models
  - Derived from human expert knowledge
  - Generated through manual extraction of alert and mitigating patterns from known examples
  - Often capture mechanistic considerations or knowledge from proprietary sources
  - Structural alerts and mitigating patterns are encoded into software for consistent application (predictions)

Examples of (Q)SAR Software

Name	Type	Source
CASE Ultra	Both	Commercial
Derek Nexus	Expert Rule-Based	Commercial
Leadscope Model Applier	Both	Commercial
Sarah Nexus	Statistical-Based	Commercial
EPA T.E.S.T.	Statistical-Based	Commercial
TOPKAT	Statistical-Based	Commercial
vLife	Both	Commercial
CAESAR/VEGA	Statistical-Based	Free
OECD Toolbox	Expert Rule-Based	Free
Toxtree	Expert Rule-Based	Free

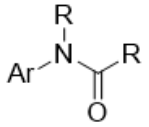
Software highlighted in red are used by FDA/CDER under Research Collaboration Agreements (RCAs)

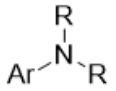
Applicability Domain

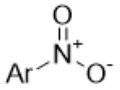
- Applicability Domain: Region of chemical space within which a model makes predictions with a given reliability
- Chemical space defined by structural attributes/properties of training set molecules
- Impurities outside of a model's applicability domain ("out-of-domain," or "OOD") should not be predicted
- Overall, different models have different coverage (applicability domain measurement)
  - Can be used to our advantage to obtain a valid prediction
  - However, when multiple models yield OODs, then extra attention needed

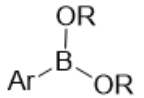
Structural Classes Associated with Mutagenicity

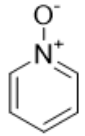
Group 1: Aromatic

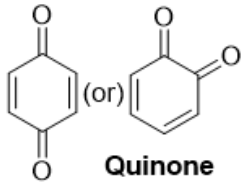

Aromatic amide

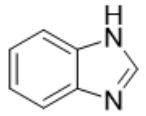

Aromatic amine

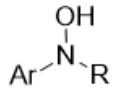

Aromatic nitro


Aryl boronic acids/esters


Aryl N-oxide


Quinone

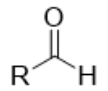

Benzimidazole ring

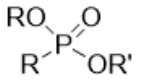

N-Hydroxyaryl

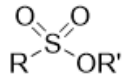

Polycyclic aromatic system (n≥3)

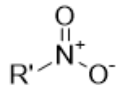
Ar = aromatic ring
R = alkyl, aryl, or H
R' = alkyl only
R'' = alkyl or aryl only
X = Cl, Br, I

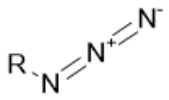
Group 2: Alkyl

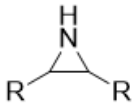

Aldehyde

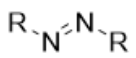

Alkyl ester of phosphonate

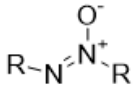

Alkyl ester of sulfonate

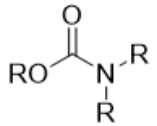

Alkyl nitro compound

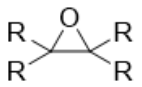

Azide

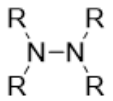

Aziridine

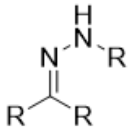

Azo compound

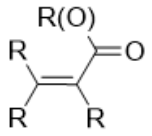

Azoxy

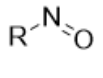

Carbamate (Urethane)

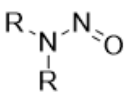

Epoxide

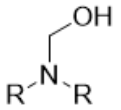

Hydrazine

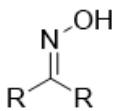

Hydrazone

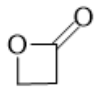

Michael-reactive acceptor


Nitroso

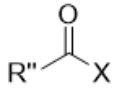

Nitrosamine

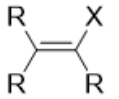

N-Methylol

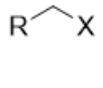

Oxime

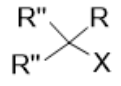

Propiolactone

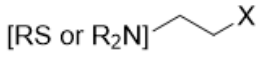
Group 3: Halide

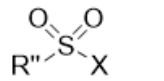

Acyl halide


Halo-alkene


Primary alkyl halide


Non-primary alkyl halide


N or S mustard

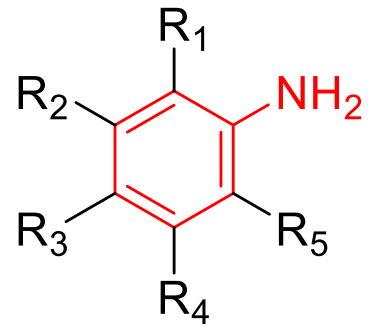

Sulfonic acid halide

Compiled from 1) Benigni and Bossa, 2011. Chem. Rev. 111, 2507-2536; 2) Müller et al., 2006. Regul. Toxicol. Pharmacol. 44, 198-211; 3) Enoch and Cronin, 2012. Mut. Res. 743, 10-19; 4) O'Donovan et al., 2011. Mut. Res. 724, 1-6.

Why use a computer?

Why not simply use visual inspection?

- Highly complex associations can be captured by a model
  - Published alerts are quite general—do not consider mitigating features or cumulative effect of multiple substituents
- E.g., primary aromatic amines
  - Stabilization of corresponding nitrenium ion increases mutagenicity (electronic)
  - Steric bulk near amine reduces formation of DNA adducts
- Underscores the value of (Q)SAR as a more refined approach to predicting activity based on all aspects of chemical structure



Functional group	ortho	meta	para	Functional group	ortho	meta	para
amino(NH2)-	SA	SA	SA	sulfonate-	SD	SD	SD
methyl-	SA	SA	SA	sulfonyl-	SD	SD	SD
methoxy-	SA	WA	SA	sulfonamide-	WD	SD	SD
alkylthio-	SA	?	SA	aminocarbonyl-	SD	WD	SD
trifluoromethyl-	SD	SD	SA	aminomethyl-	SD	SD	WD
phenyl-	WA	-	SA	trifluoromethyl-	SD	SD	SA
heteroaryl-	?	-	SA	carboxylate-	SD	-	SD

Ahlberg, et al. Regul Tox Pharmacol. 2016, 77, 1-12.

Structural Limitations

- Can predict:
  - Organic molecules
- Cannot predict:
  - Polymers (~>1000 Da)
    - Plastics, proteins, polysaccharides, *etc.*
  - Inorganics
    - Simple inorganic salts
    - Coordination compounds
  - Organic chemicals that are:
    - Mixtures
    - Poor coverage (unknown molecular features)
- Cannot differentiate:
  - Stereochemical or geometric isomer pairs

ICH M7

How to Apply (Q)SAR Under ICH M7

Section 6:

"A computational toxicology assessment should be performed using (Q)SAR methodologies that predict the outcome of a bacterial mutagenicity assay (Ref. 6). Two (Q)SAR prediction methodologies that complement each other should be applied. One methodology should be expert rule-based and the second methodology should be statistical-based. (Q)SAR models utilizing these prediction methodologies should follow the general validation principles set forth by the Organisation for Economic Co-operation and Development (OECD)."

"The absence of structural alerts from two complementary (Q)SAR methodologies (expert rule-based and statistical) is sufficient to conclude that the impurity is of no mutagenic concern, and no further testing is recommended (Class 5 in Table 1)."

Ref 6: Sutter et al., 2013. Regul Toxicol Pharmacol. 67:39-52.

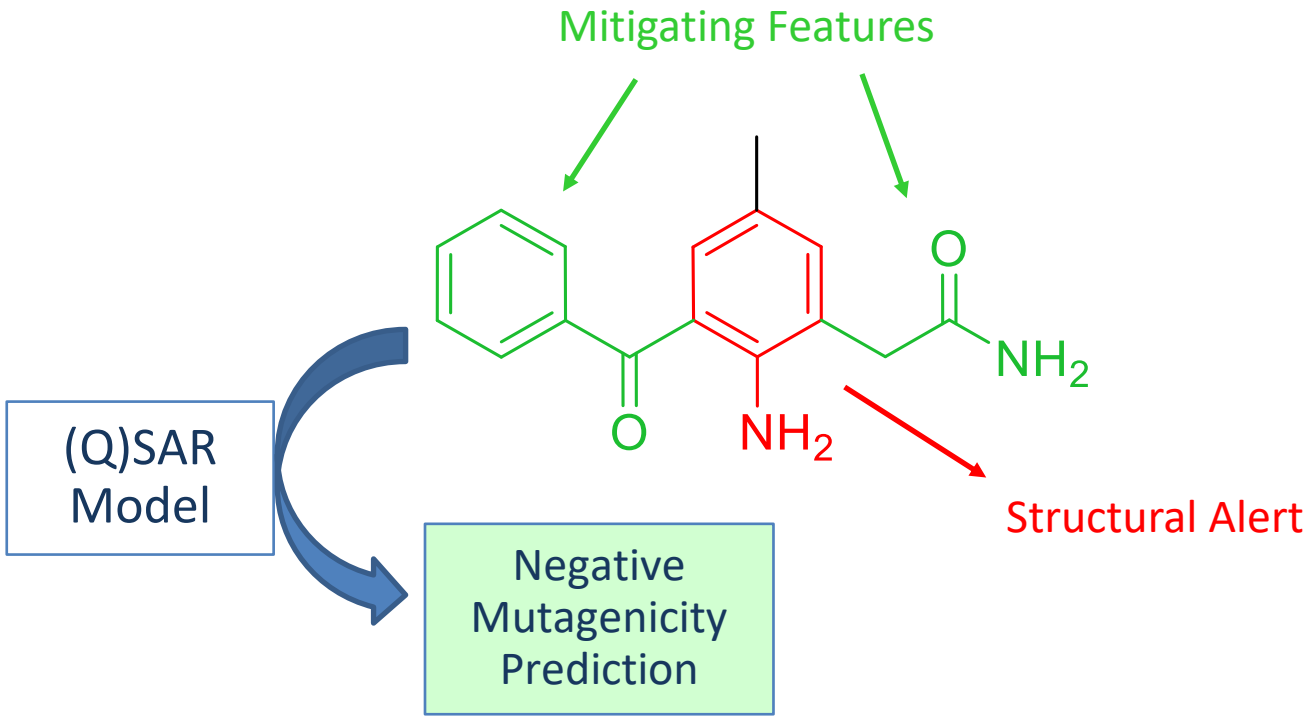
OECD Validation Principles

- To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:
  1) a defined endpoint
  2) an unambiguous algorithm
  3) a defined domain of applicability
  4) appropriate measures of goodness-of-fit, robustness and predictivity
  5) a mechanistic interpretation, if possible

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

U.S. FOOD & DRUG ADMINISTRATION

(Q)SAR Reporting

(Q)SAR Software Acceptability

- Under the ICH M7 guideline, applicants may submit (Q)SAR analyses performed using models that are fit-for-purpose
  - Commercially available
  - Freely available
  - Constructed in-house
- CDER has prior knowledge of several commercial and freely available (Q)SAR software
- For software that CDER has no prior knowledge of, supporting documentation demonstrating that a model is fit-for-purpose is required (e.g., in a (Q)SAR Model Reporting Format, or QMRF)

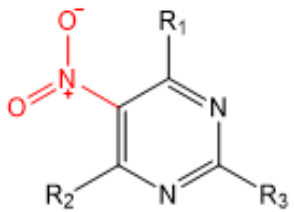
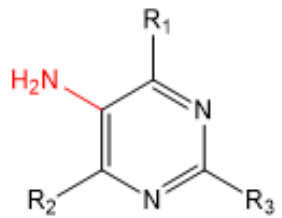
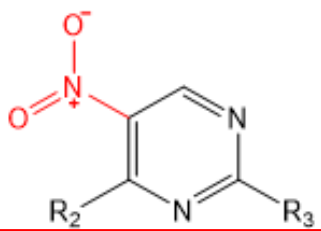
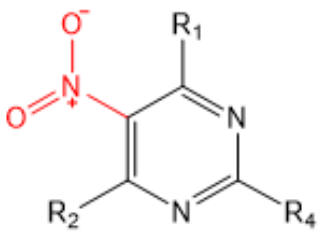
For QMRF see 1) OECD, 2007. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2
2) OECD, 2017. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC107491/kjna28713enn.pdf

(Q)SAR Analysis Reporting

Components of a well-documented (Q)SAR analysis:

- Materials and Methods, including software names/version
- Individual model predictions and overall classification in tabular format
- Detailed explanation of expert knowledge applied, particularly if model predictions are overturned
- Ideally, an appendix containing model output files, plus pivotal experimental data supporting overall classification

(Q)SAR Results Table

Chemical Name	Structure (Alert(s) highlighted in red)	Expt. Data	Bacterial Mutagenicity Predictions		ICH M7 Class	Comments
			Expert Rule-Based	*Statistical-Based*		
Chemical 1		**No data**	**+**	**Eqv**	**4**	Aromatic nitro alert in Chemical 1 is present in the same chemical environment in Chemical 4, which is Ames negative (see Appendix for data). Additionally, disubstitution in the *ortho* position has been shown to deactivate the nitro group.
Chemical 2		**No data**	**-**	**-**	**5**	Corresponding aromatic nitro analog is predicted negative. Disubstitution in the *ortho* position is known to deactivate aromatic amines.
Chemical 3		**No data**	**+**	**OOD**	**3**	Alert: Aromatic nitro
Chemical 4		**Ames Neg**	**+**	**Eqv**	**5**	Ames negative (see Appendix for data).

(Q)SAR Results: Special Considerations

Software Versions

- Predictions with the most recent software version are preferred
  - Old predictions may be acceptable unless the conclusions are questionable (e.g., a negative prediction for a chemical with an alert)
  - We consider a (Q)SAR predictions to have 2-year shelf-life as an unofficial rule-of-thumb

Out-of-Domain Results

- An OOD result is not a prediction and does not contribute to a negative (Class 5) classification
- Application of expert knowledge can be used to address OOD results
- FDA/CDER uses a 3rd model to resolve most OODs in internal analyses

Concluding Remarks

- (Q)SAR models provide a state-of-the-art approach for assessing mutagenicity
- Prediction is based solely on chemical structure
- Replaces visual inspection to identify structural alerts for ICH M7 impurity classification
- Published structural alerts can provide an preliminary screen for potential mutagens but (Q)SAR is needed for a more refined prediction
- Models should be consistent with OECD Validation Principles
- Comprehensive reporting of impurity (Q)SAR analyses can avoid multiple review cycles
- FDA/CDER applies expert knowledge to all internal (Q)SAR assessments

Acknowledgements

CDER Computational Toxicology Consultation Service Team


Marlene Kim


Neil Hartman
(In memoriam)


Suresh Jayasekara


Govindaraj Kumaran


Zheng (Jenny) Li


Diandian Shen


Jian Yang


Andy Zych

Internal Collaborators

Aisar Atrakchi
Megan Boals
Bob Dorsam
David Green
Rick Houghtling
Sruthi King
Tim McGovern
Kirk Roy
Barbara Scott
David Skanchy
Lidiya Stavitskaya

External Collaborators

MultiCASE Inc.
Lhasa Limited
Leadscope Inc.

Thank You!

- Send questions regarding this presentation by March 19, 2021, to DMFWorkshop2021@fda.hhs.gov for inclusion in the follow-on webinar on April 9, 2021.
- Please refer to the following presentation on March 3rd for additional information:
  - "ICH M7(R1) – Chemistry and manufacturing control (CMC) Perspective on Impurity Hazard Assessment" by Barbara O. Scott


FDA
U.S. FOOD & DRUG
ADMINISTRATION