A Pharmaceutical Perspective on QSAR and Expert System Tools

Pesticide Program Dialogue Committee (PPDC)
Crystal City, VA

January 6, 2009

Edwin J. Matthews, Ph.D.

Informatics and Computational Safety Analysis Staff (ICSAS)
Office of Pharmaceutical Science
Center for Drug Evaluation and Research
US Food and Drug Administration

The views expressed during this talk are those of the author; this talk is not an official US FDA guidance or policy statement.

Outline

- ICSAS
- *In silico* testing paradigm at the FDA
- QSAR and expert system tools
- Applications
- FDA/EPA Pesticide database project
- Critical endpoints
- Consensus prediction strategy
- Unmet needs

WARNING!

The QSAR strategy and methodologies described in this presentation are a work in progress!

ICSAS

- Computational Toxicology Consulting Service
  - Internal SAR & QSAR Consults
- Computational Toxicology Program
  - (Q)SAR Research & Development

Computational Toxicology Program

- An applied regulatory research unit
- Create toxicological and clinical databases
- Develop rules for quantifying toxicological and clinical endpoints
- Evaluate data mining and (Q)SAR software
- Develop toxicological and clinical effect prediction programs through collaborations with software companies

Primary Goal

To be able to predict accurately chemical toxicities with *in silico* software for all toxicological and clinical effect endpoints of interest to the US FDA

Potential Benefits

Substantially reduce, replace and refine the need for animal toxicological testing in establishing the safety of chemical substances

What FDA / CDER Comtox Program Can Do Now

Provide rapid and reliable decision support information to support regulatory decisions without additional testing

Toxicology and Clinical Predictions Used to Support Regulatory Decisions

- Prevent additional review cycles by using *in silico* data information on chemically similar substances and predicting toxicological and clinical effects
- Use *in silico* data when you must make a regulatory decision in the absence of all the safety information you would like (degradents, contaminants, metabolites)
- When a safety study is equivocal, investigate *in silico* data for toxicologically related endpoints

New *In Silico* Testing Paradigm at the FDA

ICSAS is Facilitating an Orderly Transition to a New *In Silico* Testing Paradigm

- Introduction, education, and consensus
  - FDA and CRADA partner presentations, 2008 - 2009
- FDA MaPPs (SOPs) and Guidances
  - Public comment, 2009
- Once in place, *in silico* methods will be performed as a means of reduction, replacement, refinement for longer, more expensive testing, not as an additional burden

The ICSAS *in silico* testing paradigm is being articulated in parallel to the current OECD and EU QSAR efforts,

but there are substantial strategic differences in these approaches

ICSAS *In Silico* Paradigm

- Employs a multiple software platform strategy
  - Different software use the same training data sets
- Employs commercial software products
  - Freeware only used for special applications
- Emphasizes read-only tools to produce consistent results between institutions
  - All QSARs are prepared and validated by the FDA
- Emphasizes global QSARs and expert systems
  - Universe of non-congeneric chemicals in commerce
- Objective: facilitate a competitive advancement of (Q)SAR science and technology

ICSAS *In Silico* Paradigm

- Emphasizes human health effect endpoints
- Includes data sets from both preclinical (animal) and clinical (human) endpoints
- Includes studies from public domain and knowledge from archival proprietary studies
- CBI / proprietary data is not made transparent
- Emphasizes data regulatory submissions using standard test protocols
- Commits resources to quality review of data
- Commits resources to the maintenance and enhancement of the (Q)SAR knowledge base

What are the software platforms currently included in the new multiple platform strategy?

In Silico Toolbox for Toxicity Screening*

- *Derek for Windows & Meteor*
- *Leadscope FDA Model Applier, Predictive Data Miner*
- *MC4PC & META*
- *BioEpisteme & Integrity*
- *QSARIS* (Scimatics, *MDL-QSAR*)

**No endorsement is implied*

Current Status of Validated Software

- 3 Software platforms are validated
- 2 Additional platforms are being validated
- Additional platforms are being added to diversify the *in silico* battery capabilities
- New functionalites are rapidly being developed and added to all of the platforms
- Testing paradigm is flexible and suitable for different regulatory applications

Comparison of Software Platforms

	Leadscope FDA Model Applier	QSARIS (MDL-QSAR)	MC4PC	BioEpisteme	Derek for Windows
(Q)SAR Algorithm	Partial Logistic Regression / Expert Rules	Discriminant Analysis	Recursive partitioning Statistics	Genetic Algorithm/ Statistics	Human Expert Rules
Molecular Structure Interpretation	Fingerprint Molecular Features / Scaffolds	Connectivity Indices (2D Descriptors)	2-10 Atom Molecular Fragments	None	Structural Alert (Molecular Fragment)
Molecular Descriptors (2D / 3D)	Limited 2D (n~10)	2D (n~200, Kier and Hall)	Limited 2D (n~6)	2D (n~126, volume & shape descriptors; 3D is a future functionality)	Limited 2D (n~4)
Training Data Sets	FDA/ICSAS	FDA/ICSAS	FDA/ICSAS	FDA/ICSAS and PIBR	Private Industry, Government, Literature, and FDA/ICSAS
Coverage Measure	Presence in Molecular Feature Domain	Descriptor-based Membership in Class	Presence of 2-3 Atom Unknown Fragments	None (Future Affinity Constant Functionality)	None
Operating System	Windows Desktop	Windows Desktop	Windows Desktop	Windows Desktop (client server work station)	Windows Desktop

Where are the software platforms currently being used?

What are the current applications?

- **FDA**
  - Center for Drug Evaluation and Research
    - Toxicities of drug contaminants, metabolites
    - Drug adverse effects and off-target MOAs
    - Prospective study with OND / DCR drug products
  - Center for Food Safety and Applied Nutrition
    - Food Contact Notification Program (120 day clock)
  - Center for Veterinary Medicine (soon)
- **EPA**
  - Office of Pollution Prevention and Toxics (soon)
    - Pre-manufacture Notice Program (90 day clock)
  - Office of Pesticide Programs (soon)
    - Toxicities of pesticide contaminants and metabolites
  - Office of Water (soon)
    - Prioritization of chemicals in municipal water systems based upon human QSARs

- **NIH**
  - **National Institute on Drug Abuse**
    - **Selection of lead chemicals for substance abuse**
  - **National Cancer Institute**
    - **Screening of cancer chemotherapeutics / chemopreventatives for potential adverse effects**
- **Pharma**
  - **All major Pharma use one or more programs**
    - **Lead selection and discovery applications**
- **EU, REACH, 7th Amendment**
  - ***InSilico First* (soon)**
    - **Battery of FDA QSARs on multiple platforms**

Related Partnership

www.insilicofirst.com/

*No endorsement is implied

What toxicological and clinical endpoints are currently available?

Non-Clinical QSAR Suites

Prediction Suite	Models	Chems	Records
• Carcinogenicity	7	1,584	24,708
• Genetic toxicity	20	8,200	> 27,498
• Reproductive toxicity	9	686	
• Developmental toxicity	27	2,115	51,724
• Behavioral toxicity	3	503	
• Phospholipidosis	1	583	227
• Quantitative MTD	8	1,266	3,925
• Organ specific toxicities	R&D		
• Regulatory dose conc.	R&D		

Human Clinical QSAR Suites

Prediction Suite	Models	Chems	Records
• Hepatobiliary	5	1,660	120,419
• Renal / Bladder	6	1,660	214,563
• Cardiological	13	1,632	396,985
• Immunological	26	1,586	823,954
• Pulmonary	25	1,579	242,344
• Quantitative MRDD	2	1,246	4,500
• Other organ systems	R&D		
• Human metabolism	R&D		
• Human bioavailability	Suppl. Programs		

What Efforts are being done to expand coverage of FDA QSARs for pesticides and other non-pharmaceutical molecules?

FDA/CDER Data Sharing Initiatives

- **EPA**
  - OPPT: CBI PMN industrial chemical studies
  - OPP: CBI pesticide studies
- **NIH**
  - NCI: CBI chemotherapeutics
  - NCI: public domain oncolytics (IAG)
- **FDA**
  - CFSAN: food additives, food contact substances (Leadscope)
  - CVM: veterinary drug products

FDA/CDER Data Sharing Initiatives

- **EU**
  - UK/DEFRA: CBI pesticide studies
- **Industry, Institutions, Academia**
  - PharmaPendium™ (FDA/CDER and Elsevier MTA, FOI records)
  - Drug-Drug Interaction Database (U.Washington)
  - Zenith Project (Lhasa Limited)
  - BioPrint®: Drug MOA / Clinical AE database (Cerep?)

Pesticide Database Project

- **Data Sources**
  - Public Domain: EPA/OPP, PAN, CA/EPA/OEHA, HC, WHO (~2500 documents)
  - CBI: EPA/OPP, DEFRA
- **Linguistic Software tools**
  - *I2E* (Linguimatics Ltd)
  - *SARF* (FDA/CDRH) software
- **QSAR Model / Expert System Enhancement**
  - *MC4PC, LFMA, BioEpisteme*
  - *DfW* and *Vitic* (FY 2010)

Pesticide Database Project

Database Objectives:

Compile a chemical structure database

- ~6500 substances (~3500 simple organics)

- Construct a comprehensive, relational database of pesticide toxicology data
- Publicly available (*Vitic* software, Lhasa Limited)
- Enhance FDA QSAR database

Pesticide Database Project

QSAR and Expert System Objectives:

- Expand coverage for pesticide-like chemicals in current FDA global QSARs and expert systems
- Construct new QSARs for EPA safety assessment endpoints (NOEL, neurotoxicity)
- Predict toxicities of pesticide metabolites, contaminants, reagents
- Provide data to support EU REACH
- Provide data to facilitate EPA/OPP tier-testing

Consensus Guidance and MaPPs

Decision Support Information for:

- Specific FDA actions, tasks and applications
- Other Agencies governed by different legislation
- Providing dossiers of *in silico* data for data gaps
- Defining acceptable experimental design criteria
- Defining acceptable model validation criteria
- Defining acceptable criteria for evaluation of data from multiple platforms

Why does the FDA use more than one QSAR software program?

FDA CRADA Programs

- None of the programs have all the necessary functionalities
- None of the programs have 100% coverage, sensitivity, and specificity
- All of the programs are complementary and can be used for consensus prediction strategies
- FDA cannot endorse a single (Q)SAR program
- FDA receives CRADA contributions which can only support CRADA activities
  - Funds are used to harvest data from FDA archives
- All of the programs must protect CBI data

What are the critical components for the FDA multiple platform strategy?

1. PREDICTION VS. EXPERIMENTAL VALUE

How do you know whether your test chemical is included in the (Q)SAR model training data set?

- Create a *MC4PC** test chemical .sdf
- Run the test chemical .sdf with each FDA QSAR model
- Examine the ICSAS.xls file

A	experimentally active
M	marginal experimental activity
I	experimentally inactive
-,+,?	test chemical not in the training data set

No endorsement is implied

2. BIOAVAILABILITY

Is the test chemical likely to be bioavailable?

If it is, how will it be processed by the host?

In Silico Toolbox for Bioavailability*

- *MC4PC*
  - Lipinski Alert
  - Human Intestinal Absorption Coefficient
  - Water Solubility
  - LogP (Log Octanol / Water Partition Coefficient)
- *BioEpisteme (PIBR)*
- *ADMET Predictor*, *Gastro Plus* (Simulations Plus)
- *QSARIS* (Scimatics)

MultiCASE Incorporated

PROUS INSTITUTE FOR BIOMEDICAL RESEARCH

*No endorsement is implied

3. STRUCTURAL ANALOGUES

Structural analogues may have data relevant to the test chemical!

How do you identify structural analogues of a test chemical?

Resources Utilized by ICSAS

Some Sources of Chemical / Drug Information

- **Derwent World Drug Index**
- **Discovery Gate (MDL/Symyx)**
- ***Integrity* (PIBR)**
- ***Leadscope Client***
- **National Library of Medicine *ChemID & PubChem***
- ***Pharmaprojects***
- ***PharmaPendium* (Elsevier)**
- ***Physicians' Desk Reference***
- **The *Merck Index***
- **Thomson *MicroMedex***
- **US Pharmacopeia**

Analogue Evaluation Criteria

- Molecular feature clustering
  - ISIS keys ≥ 85%
  - Tanamoto coefficient ≥ 85%
- Molecular property clustering
  - Compounds that share a pharmacological activity

4. COVERAGE

Predictions for test chemicals outside the domain of applicability should be discarded

How do you know whether a test chemical is covered by a FDA QSAR model?

Coverage (Domain of Applicability)

- Coverage is dependent upon QSAR prediction paradigm
- Coverage must be determined for each QSAR model
- Coverage is independent of chemical toxicity
- Coverage does not depend upon the product use; it is good for drugs and other products

Coverage ADRI

- Select set of congeneric test chemicals
  - Analogues and 1st pass metabolites, n≥10
- Create a *Leadscope FDA software* .sd-file
  - Currently the most sensitive molecular feature tool
- Calculate the applicability domain representation index (ADRI) for test chemicals in each QSAR model

ADRI = (# covered test chemicals) / (total # test chemicals)

ADRI = 1.00 for total coverage

ADRI = 0.00 for no coverage

***No endorsement is implied**

Suggested Coverage Evaluation Criteria

- QSAR Models with acceptable coverage
  - ≥ 75% of congeneric test chemicals are covered
- QSAR Models with poor coverage
  - < 75% of congeneric test chemicals are covered

5. METABOLISM

Toxicity can be related to the metabolites of a chemical

How can metabolites of a test chemical be predicted?

Test Chemical Metabolites

In Silico Prediction Systems

- *METEOR* (Lhasa Limited)
- *META* (MultiCASE)
- *MetaDrug* (GeneGo)
- Metabolic Simulator
  (P. Schmieder/DUL/USEPA)

Data Mining Sources

- *PharmaPendium* (Elsevier)
- *Metabolite* (Discovery Gate, Symyx)

No endorsement is implied

6. WEIGHT OF EVIDENCE PREDICTIONS

How does ICSAS recommend that predictions from multiple QSAR programs be combined?

General Recommendations

- Use FDA validated software platforms
- Use FDA validated QSAR models
- Use identical training data sets in different QSAR platforms
- Use training data sets containing both non-proprietary data and knowledge from proprietary studies
- Use data derived from standard protocol studies

Weight of Evidence Predictions*

Test Chemicals with Coverage Issues

- QSAR Model Poor Coverage (PC)
  - LFMA software: coverage <75% for a set of structurally related test chemicals
  - *All predictions for the QSAR model are discarded*
- Not in the Applicability Domain (ND)
  - *LFMA* software: individual test chemical not covered
  - *All predictions for the test chemical are discarded*

**** WOE is a work in progress***

Weight of Evidence Predictions*

- Test chemicals with no coverage issues

- Consensus Inactive (-)
  - Predicted negative by two or more QSAR programs
  - *Risk Assessment & Risk Management*

- Marginal Activity (?):
  - Predicted positive by only one QSAR program
  - *Risk Management*

**** WOE is a work in progress***

Weight of Evidence Predictions*

- Active (+)
  - Predicted positive by one QSAR program and by QSAR model(s) for two or more toxicologically related endpoints
  - *Risk Management*
- Consensus Active (C+)
  - Predicted positive by two or more programs for a single toxicological endpoint
  - *Risk Management & Risk Assessment*
- Consensus Strong Active (C++)
  - Predicted positive by QSAR model(s) for two or more QSAR programs and two or more toxicologically related endpoints
  - *Risk Assessment & Hazard Identification*

**** WOE is a work in progress***

7. MECHANISM OF ACTION

How can a QSAR prediction be related to a plausible test chemical mechanism of action?

Test Chemical MOAs

QSAR & Expert System Predictions:

- *Derek for Windows*:
  - Expert and prototype structural alerts
  - Plausible MOAs
- *BioEpisteme* QSAR Models:
  - Pharmacological MOA
  - Off-target MOAs
  - Drug-drug interaction*
- *MetaDrug*:
  - Xenobiotic metabolism by humans

**** QSAR model under development***
No endorsement is implied

Concluding Remarks on Unmet Needs

Unmet Database Needs

- Pharmaceutical off-target activities
- Pharmaceutical IND/CBI data, untouched
- Integration of FDA and EPA regulatory archival data
- Compilation of regulatory dose concentration endpoint data sets
- Validation of advanced linguistic software to extract archival data

Unmet QSAR and Expert System Needs

- Integrated fragment & descriptor paradigms
- Validation of 3D descriptor paradigms
- QSARs based upon PAI plus metabolites
- QSARs for drug-drug interactions
- QSARs for animal organ toxicities
- QSARs for regulatory dose concentration endpoints (LOEL, NOEL, etc.)
- Expert system rules for toxicities of non-QSARable substances (biologics, etc.)

General References

- Combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows software to achieve high performance, high confidence, mode of action-based predictions of chemical carcinogenesis in rodents. Matthews, et al. *Toxicol. Mechan. Methods* 18:189-206, 2008
- In silico approaches to explore toxicity endpoints: issues and concerns for estimating human health effects. Matthews and Contrera. *Expert Opin. Drug Metab. Toxicol.* 3:125-134, 2007
- Use of toxicological information in drug design. Matthews, et al. *J. Mol. Graphics Mod.* 18:605-615, 2000

http://www.fda.gov/cder/Offices/OPS_IO/default.htm