

REACH Seminar – 26th March 2009

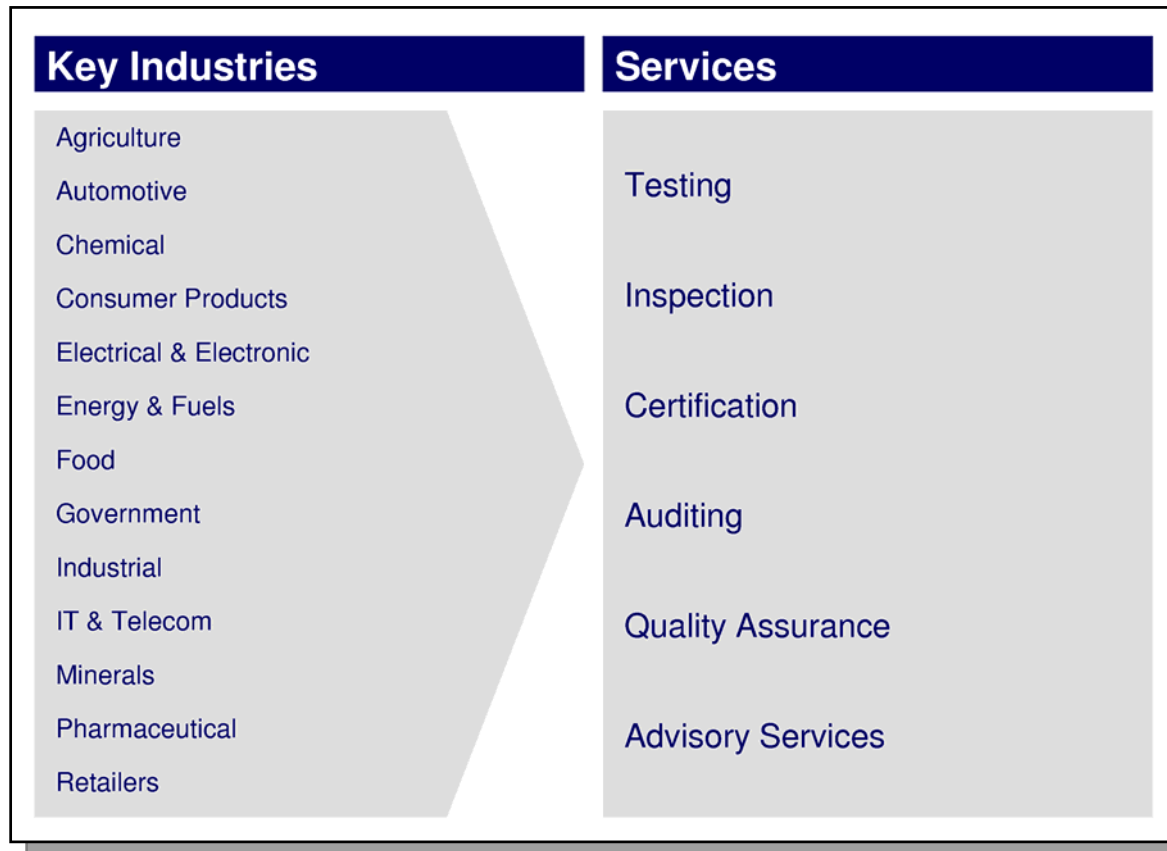
After the GAP Analysis:
Completing Physico-Chemical Data Gaps Using QSAR
and Laboratory Testing

Dr Daniel Tackley

Intertek ASG

daniel.tackley@intertek.com

Background: About Intertek



Background: Intertek Facts & Figures

- UK Company with a Global Reach
- 110 Countries
- >1000 Laboratories & Offices
- >23,000 Employees
- FTSE 250 (FTSE 100), Support Services Sector
- Market Capitalisation ~ £1.5bn



Background: Intertek ASG

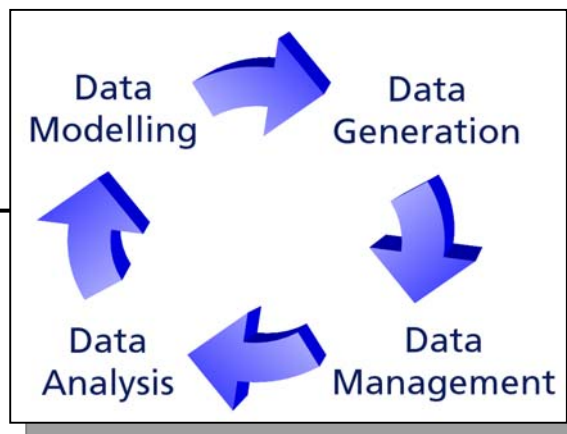
- Located in Blackley, North Manchester
- Ex-Corporate R&D Facility (ICI, Zeneca, Avecia)
- Transferred to Intertek in 2004
- Skills in:
 - Advanced Analytical Chemistry
 - Computational Science
- Inspected by:
 - MHRA (UK) for GLP and GMP compliance
 - FDA (US) for GMP compliance

Introduction: Dr Daniel Tackley

- Senior Scientist, Intertek ASG
- Informatics and Computational Chemistry
- BSc (York) – Chemistry
- MSc (York) – Ultrafast Laser Spectroscopy
- PhD (Strathclyde) – Vibrational Spectroscopy and Computational Chemistry

Informatics & Computational Chemistry

- Chemometrics
 - Prediction of test properties from information rich sources, especially spectra
- QSAR / QSPR
 - Quantitative Structure Activity / Property Relationships
 - Library enumeration
 - Similarity / diversity analysis
 - Intelligent molecular design
 - *In silico* screening



- Computational Chemistry / Molecular Modelling
 - Molecular structure and interactions
 - Molecular property calculations
- Physical Property Prediction
 - pK_a , logP, logD, aq. sol., boiling point, vapour pressure
 - Single molecules or compound libraries

- Data Visualisation
 - Large and / or multi-dimensional datasets
 - Interactive interrogation of data
- Statistics
 - Variability analysis
 - Trending
- Experimental Design

- Database Design and Implementation
- Excel Customisation
- Data Cleanup
- Automation
 - Macros, Workflow Management (Pipelining)

About REACH: Objectives

- Protection of Human Health and Environment
- Competitiveness of Chemical Industry
- Single European Market
- Transparency
- International Consistency
- Non-animal Testing
- WTO Obligations

About REACH: Aims

- REACH aims to:
 - Get manufacturers and importers to:
 - Generate information on hazards
 - Health, Safety and Environment
 - Assess risks
 - Manage risks
 - Ensure information is available to:
 - Users
 - Regulators
 - General public

Data Requirements for REACH

- Technical Dossier
 - Identification, Manufacture and Use, Classification, Guidance on Safety, Information, Proposals
- Chemical Safety Report
- Risk Reduction
 - Human Health
 - Physico-Chemical
 - Environmental
 - PBT and vPvB
- Risk Assessment

Obtaining Data for REACH

- Laboratory Testing
 - *In vivo* tests (animal tests)
 - *In vitro* tests (cell cultures)
 - Analytical measurements
 - Other tests
- Estimation Methods
 - Qualitative
 - Structure-Activity Relationships (SAR)
 - Read-Across
 - Categorisation
 - Quantitative
 - Quantitative Structure-Activity Relationships (QSAR)

Physico-Chemical Testing: Regulatory Requirements

- Need for Physico-Chemical data pre-dates REACH
- Long standing regulatory requirement for placing new chemical on the market in the EU
 - NONS (Notification of New Substances)
- Physico-Chemical data usually generated as part of a broader Tox / Eco-Tox study
 - Performed to GLP standards
- Test materials should be characterised for identity, purity and impurities prior to testing
 - Sameness

Importance of Physico-Chemical Testing

- Physico-Chemical data are pivotal to the whole of the notification process
 - Designing complete test packages
 - Risk assessment
- The order in which tests are carried out should be considered
 - Results of one test can influence how others are run
 - Decomposition during a melting point study would preclude the need for a boiling point study

Importance of Physico-Chemical Testing

- Physico-Chemical tests for hazardous properties used mainly for safe handling and hazard communication
 - Classification and Labelling
 - Properties include:
 - Flash point
 - Flammability
 - Explosive properties
 - Oxidising properties
- Result of other Physico-Chemical tests are vital in designing appropriate toxicological and eco-toxicological test packages and for risk assessment calculations

Common Physico-Chemical Tests

- Melting temperature
 - Boiling temperature
 - Relative density
 - Vapour pressure
 - Surface tension
 - Solubility in water
 - Octanol:water partition coefficient
 - Flash point
 - Flammability
 - Pyrophoric properties
 - Flammability on contact with water
 - Explosive properties
 - Self-ignition temperature
 - Oxidising properties
 - Particle size distribution
-
- Additionally, two of the environmental fate tests:
 - Hydrolysis as a function of pH
 - Adsorption / desorption coefficient

Impact of Tests

Test	Impact on other physico-chemical tests	Examples of Impact on toxicology	Examples of use in risk assessments
Melting / freezing point	Boiling point study not require if decomposition observed		
Boiling point	Relates to vapour pressure. Can affect flammability classification.		Need to consider temperatures used in chemical / manufacturing processes.
Relative density	Used in the determination of viscosity.		Fire-fighting measures: H2O extinguishers may not be suitable if $D_{204} < 1$.
Vapour pressure	Related to boiling point. May influence the ability to measure an accurate log Kow.	Dermal or inhalation exposure route for acute toxicity test; Route of exposure for sub-acute toxicity test; Exposure and excretion routes for toxicokinetic assessment.	Calculation of vapour exposure for human health risk assessment. Calculation of Predicted Environmental Concentrations (PECs) for environmental risk assessment- vapour pressure is a key parameter in determining environmental fate and behaviour. Determination of atmospheric behaviour as for exposure of man via the environment calculations.
Surface tension	May influence the approach for testing log Kow. Can occasionally interfere with measurement of water solubility.	Surface active substances have a higher local irritant or corrosive effect. As a consequence of local corrosion, the dermal uptake of a substance can be enhanced.	Environmental fate.

Test	Impact on other physico-chemical tests	Examples of Impact on toxicology	Examples of use in risk assessments
Water solubility	<p>Water solubility < 1 mg / l at 20 °C. Need to prepare 90% saturated solution (up to a maximum of 1 g / l) for surface tension test. Time to achieve saturation can be relevant to solution preparation for surface tension test. Water solubility affects concentration used in hydrolysis test. May influence the ability to measure log Kow accurately</p>	Impact on toxico-kinetic behaviour.	Environmental classification and labelling. Water solubility is a key parameter in determining environmental fate and behaviour
Partition coefficient n-octanol/water (Kow)	<p>Generally, substances with a high log Kow will be hydrophobic and have low water solubilities. Substances with negative log Kow will be hydrophilic and have high water solubilities.</p>	<p>Toxico-kinetic behaviour: Kow indicates the potential for absorption across biological membranes and for passive diffusion (e. g. prediction of dermal absorption). It provides information on the potential for accumulation in the body. Choice of suitable vehicle for toxicity studies.</p>	Environmental classification and labelling. Log Kow is a key parameter in determining environmental fate and behaviour and is used as a surrogate for bioaccumulation potential in the absence of bioaccumulation tests.
Dissociation constant	<p>The process of dissociation (depending on pH) in solution can have a huge effect on water solubility, Kow, and volatilisation of a substance.</p>	<p>Exposure to hydrolysis products <i>in vivo</i>: the dissociation constant indicates the potential for absorption from the gastrointestinal tract, because ionised compounds are thought not to cross biological membranes.</p>	The dissociation constant is a key parameter in determining environmental fate and behaviour.
Viscosity	Choice of methods for the determination of density.	Parameter for aspiration hazard	Assessment of spreadability of liquids.

Non-Testing Methods

- Saving Animals
 - Regulatory use of (Q)SARs is encouraged
 - Tests on vertebrate animals shall only be conducted as a last resort
- Saving Money
 - Huge cost-saving potential[†]
 - Average QSAR (based on current models)
 - Estimated €700 million savings
 - Max QSAR (requiring further development and validation)
 - Estimated €940 million savings

[†]EC Joint Research Centre Data, 2003

REACH ANNEX XI

- 1.3 (Q)SARs
 - Qualitative or Quantitative Structure-Activity relationship
 - Results for (Q)SARs may be used instead of testing when the following conditions are met:
 - Results are derived from a (Q)SAR model whose scientific validity has been established
 - The substance falls within the applicability domain of the (Q)SAR model
 - Results are adequate for the purpose of classification and labelling and/or risk assessment
 - Adequate and reliable documentation for the applied method is provided

REACH ANNEX XI

- 1.5 Grouping and Read-Across
 - Substances whose physico-chemical, toxicological and ecotoxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity may be considered as a group, or “category” of substances
 - ... physico-chemical properties, human health effect and environmental effects or environmental fate may be predicted from data for reference substance(s) within the group by interpolation to other substances in the group
 - Similarities based on:
 - Common functional group
 - Common precursors and/or breakdown products
 - Constant pattern in the changing of potency of properties

Physico-Chemical (Q)SARs

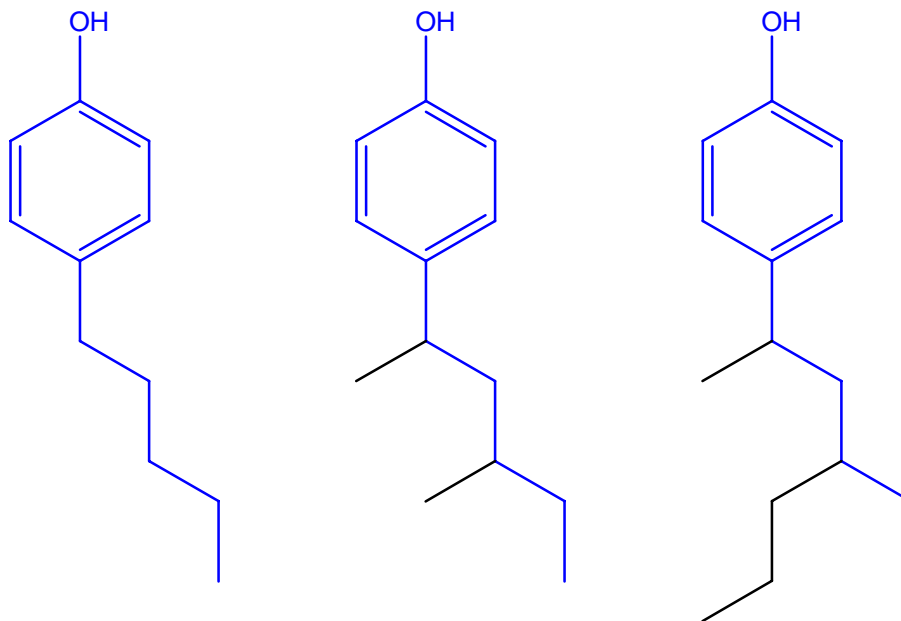
- Physico-chemical properties accessible via (Q)SARs:
 - pK_a
 - Octanol:water partition coefficient
 - Aqueous Solubility
 - Melting Point
 - Boiling Point
 - Vapour Pressure
 - Henry's Law Constant
 - Relative Liquid Density
 - Pure Liquid Viscosity
 - Liquid Surface Tension
 - Flash Point
 - Auto-Ignition Temperature
 - Soil Sorption

Practical Considerations

- Need to consider the domain of applicability
 - Global model
 - One model fits all
 - Simpler to generate
 - Very hard (if not impossible) to get good accuracy across the whole of molecular space
 - Local model
 - Individual model for a small group of molecules
 - Need to generate multiple models for each group
 - How to identify groups of molecules
 - Need sufficient data to build and test the model
 - Should give better accuracy across the chosen area of molecular space

Molecular Similarity (Grouping)

- By inspection
- Common substructure



Molecular Similarity (Grouping)

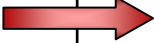
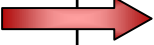
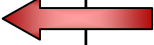
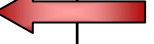

- Clustering based on molecular fingerprints
 - Binary yes/no representation of molecular features
 - Linear path
 - Subgraph path
 - MACCS (Molecular Access System) substructure keys
 - 160 key “public” set
 - 960 key “private” set
 - NEMA substructure keys
 - Newly Enhanced Morgan Algorithm
 - Hierarchical clustering
 - Similarity scoring

Molecular Similarity (Grouping)

- Molecular descriptors
 - Mathematical representation of molecular structure
 - Constitutional, topological, walks and path counts, connectivity, geometrical, functional group, molecular properties
 - 1500+ commonly used descriptors
 - Descriptor clean-up
 - Remove consistent descriptors
 - Remove highly-correlated descriptors
 - Data reduction procedure
 - Principal Component Analysis (PCA)
 - Transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components
 - PC1 accounts for as much of the variability in the data as possible
 - Similarity scoring on principal components

Read-Across and SAR

- Structurally similar molecules are used to predict for the molecule of interest
 - Read-Across for categorical data (Yes/No)
 - Interpolation or extrapolation for numerical data

Molecule 1	Molecule 2	Molecule 3	Molecule 4	SAR		
✓		x	✓	x	Read-Across	
✓		x		✓	✓	Interpolation
x		✓	✓		x	Extrapolation

Quantitative Structure-Activity Relationships

- What:
 - Relating activity to chemical structure
- How:
 - Structure generation and identification of domain of applicability
 - Molecular descriptor calculation
 - Model generation relating descriptors to activity
 - Training data
 - Model validation
 - Test data
 - Application of the model predictively for molecules whose properties are unknown

Pre-Built QSAR Models

- Various software packages available which use QSAR models to generate physico-chemical properties
 - Freely available
 - EPISUITE (US EPA), SPARC, VCCLAB, etc.
 - Commercial
 - ACD/Labs, Pallas, TOPKAT, etc.
- Each property from each package will have its own level of accuracy
 - Some accuracy comparisons are available
- Need to be used intelligently rather than “Black Box”

“In Silico Prediction for Physicochemical Properties”,

John Dearden and Andrew Worth (JRC Scientific and Technical Reports)

Consensus Modelling

- Look at the statistics for the error in the QSAR model
- Compare with the error in an experimental measurement
- Run predictions using multiple software packages
 - If practical, obtain 3+ predictions
 - If one is very different, can it be discarded?
 - Use the mean of the predictions rather than a single value

Application of Literature Models

- Find a method that is appropriate to the property you want to predict
 - Check the training set to see whether similar molecules to that of interest are included
 - Check the ranges of descriptors used in the training set
 - Check the statistics of the QSAR model
 - $R^2 > 0.9$
 - Does the model fit the data
 - $Q^2 > 0.6$
 - Can the model predict for molecules outside its training set
- Calculate values for the descriptors and apply the model
 - Validate the descriptor values with literature data as different software may produce different values

QSAR Model Development

- Identify the domain of applicability
- Obtain data to generate and build a model
 - Experimental or literature
 - Ideally enough for a training + test set
 - 20+ molecules total
 - Test set ~ 20% of training set
- Calculate and select descriptors
 - Systematic approach
 - Sequentially add or delete descriptors and evaluate model
 - Stochastic approach
 - Simulated annealing
 - Genetic algorithm



Data
Generation

Data
Management

Data
Analysis

QSAR Model Development

- Build the model based on the training data
 - Multiple regression
 - Neural networks
- Validate the model based on the test data
 - Q^2 (cross-validated R^2)
- Apply the model predictively to an unknown structure
 - Within the domain of applicability
 - Structure
 - Descriptor values



Useful Tools

- Data Gathering
 - ChemSpider [<http://www.chemspider.com>]
- Property Calculation
 - Virtual Computational Chemistry Lab [<http://www.vcclab.org>]
- SAR and Read-Across
 - OECD QSAR Toolbox [<http://www.oecd.org>]

Other QSARs for REACH

- Talk has focussed on physico-chemical properties
- QSARs are also available and can be applied to
 - Eco-Toxicity
 - Bioaccumulation, aquatic toxicity etc.
 - Human Toxicity
 - Irritation / corrosion
 - Sensitisation
 - Carcinogenicity
 - Neurotoxicity
 - Immunotoxicity
- Intertek can offer QSAR services covering all three areas

Conclusions

- Generation of physico-chemical data is not a new requirement for REACH
- Laboratories generating this data need the appropriate regulatory qualifications (GLP, ISO 17025 or equivalent)
- Where they exist, QSARs can offer large cost savings over testing methods
- A model applied outside its domain of applicability is useless
- No clear guidance yet on who will generate and validate QSAR models for REACH

Intertek REACH Services

- REACH Product assessment
- Third Party & Only Representation
- Late Pre-registration (if qualifications are met)
- SIEF & Consortia Participation
- Data Gap Analysis
- Dossier Creation & Submission
- Auditing & Testing
- SVHC (Substance of Very High Concern) Assessment, Screening & Testing
- REACH Compliance Certificate
- SDS/MSDS Creation and Update
- GHS Labelling & Classification
- Supply Chain Management (Auditing & Data Gathering)
- Strategy, Legal, Lobbying and Regulatory Consulting

www.intertek.com/reach

