CA21145 Training school in Bioinformatics and Computational tools in antibacterial research

# Ligand-based *in silico* approaches for evaluation of ADME/Tox properties of biologically active compounds

*Petko Alov*,

Institute of Biophysics & Biomedical Engineering,

Bulgarian Academy of Sciences, Bulgaria

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

IBPhBME

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

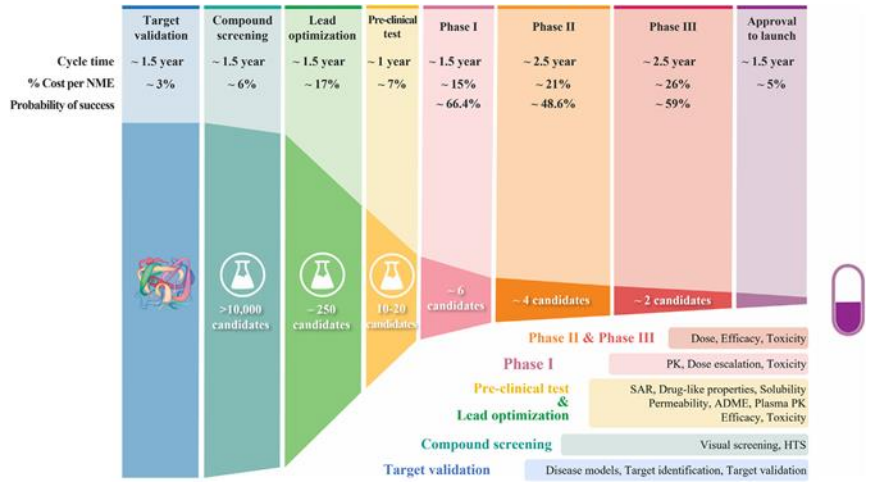
EURESTOP

Topic questions

- Questions to answer:
  - Why do we need *in silico*?
  - How has it been done?
  - How shall we do it correctly?

Why do we need *in silico*?

- Drug discovery and development takes over 10-15 years with an average cost of over $1-2 billion for each new drug to be approved (with a rate of success at and after regulatory approval <10%).
- Lead optimization and preclinical tests (where ADME/Tox evaluation is done) encounter, on average, for about 15% of the time and about 25% of the costs.




- Reasons attributed to the 90% of clinical failures in drug development: lack of clinical efficacy (40%–50%), unmanageable toxicity (30%), poor drug-like properties (10%–15%), and lack of commercial needs and poor strategic planning (10%).

doi:10.1016/j.apsb.2022.02.002

How has it been done?

- Main approaches used for *in silico* ADME/Tox properties prediction (excluding the structural ones):
  - Quantitave Structure-Activity Relationships (QSAR);
  - Structural alerts;
  - Physics-based approaches;
  - Other (not pure *in silico*):
    - Adverse outcome pathways (AOP);
    - Physiology-Based PharmacoKinetics (PBPK);
    - In Vitro to In Vivo extrapolation (IVIV);

How has it been done?

- Quantitave Structure-Activity Relationships (QSAR)

1964 – C. Hansch and T. Fujita relate the biological activity of a substance to the physico-chemical and electronic properties of the studied compounds:

$$\log\frac{1}{EC_{50}} = a + b\sigma + c\pi$$

(*in vitro* experiments)

or

$$\log\frac{1}{EC_{50}} = a + b\sigma + c\pi + d\pi^2$$

(*in vivo* experiments)

where σ is the Hammett substituent constant, and π is the Hansch hydrophobicity indicator. Hydrophobicity indicator, obviously, encounters for membrane permeability.

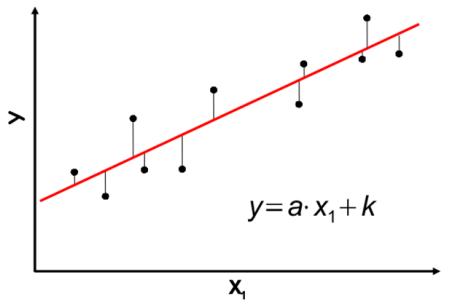
Thus, the first published QSAR in pharmacology deals with ADME!

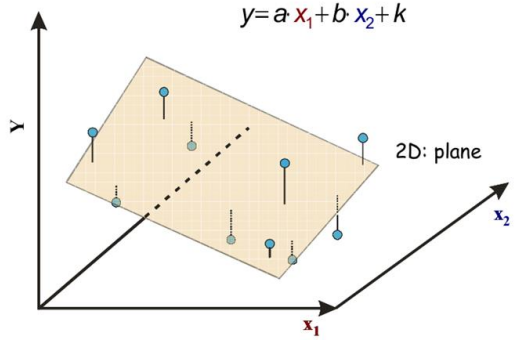
doi:10.1021/ja01062a035

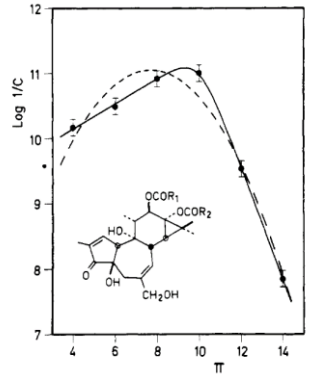
How has it been done?

- Quantitave Structure-Activity Relationships (QSAR)

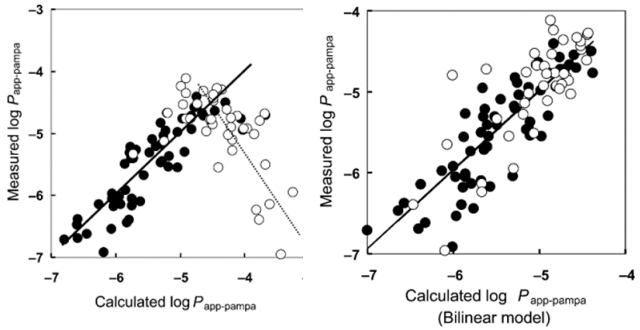
QSARs could be used for data regression or classification, and could be linear regression models (including bi- and multilinear), or entirely non-linear models (including SVMs, discriminant functions, NN models);

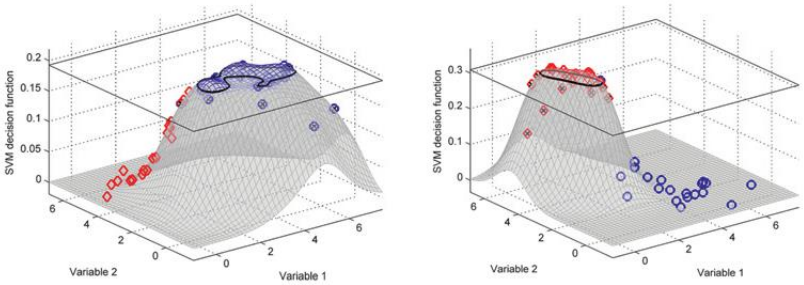


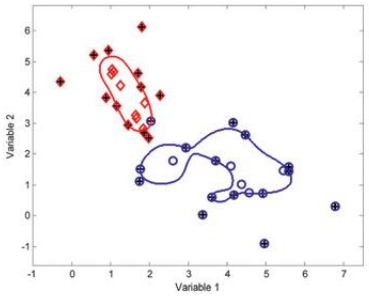




log 1/C = a log P – b log (βP + 1 ) + c




doi: 10.1021/jm00215a002 doi: 10.1002/cbdv.200900112 doi:10.1039/B918972F

- Quantitave Structure-Activity Relationships (QSAR)

  Most often QSARs are used for prediction of absorption and distribution, or compound properties directly related to trans-membrane permeability and water-lipid distribution;

  A drawback of QSAR models in relation to absorption and distribution predictions is that they hardly account for the active transport of some classes of compounds.

- Structural alerts

1995 – Lipinski's Rule of Five

"Poor absorption or permeation are more likely when:

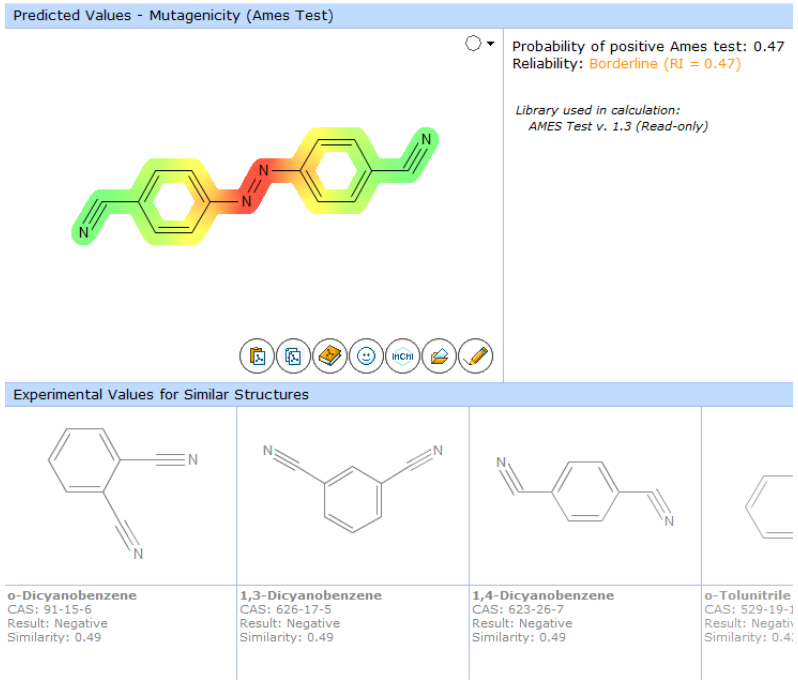
- There are more than 5 H-bond donors (expressed as the sum of OHs and NHs)
- The MWT is over 500;
- The Log P is over 5 (or MLogP is over 4.15);
- There are more than 10 H-bond acceptors (expressed as the sum of Ns and Os)
- Compound classes that are substrates for biological transporters are exceptions to the rule."

doi: 10.1016/S0169-409X(96)00423-1

How has it been done?

- Structural alerts

Structural alert approaches are based on the presence of specific structure segment in the compound's molecule and extrapolation (cross-reading) of specific toxicity patterns from known substances possessing same structural segments.




http://www.acdlabs.com/products/percepta-platform/tox-suite/

- Structural alerts

Structural alert approaches are, in general, knowledge-based SARs and include the most widely used methods used for prediction of ADME/Tox properties of putative drugs;

Most *in silico* methods for prediction of compounds' metabolism are based on structural alerts both for prediction of the metabolizing enzyme (e.g. specific CYP isoform) and for prediction of the products of metabolism ("site of metabolism" approach);

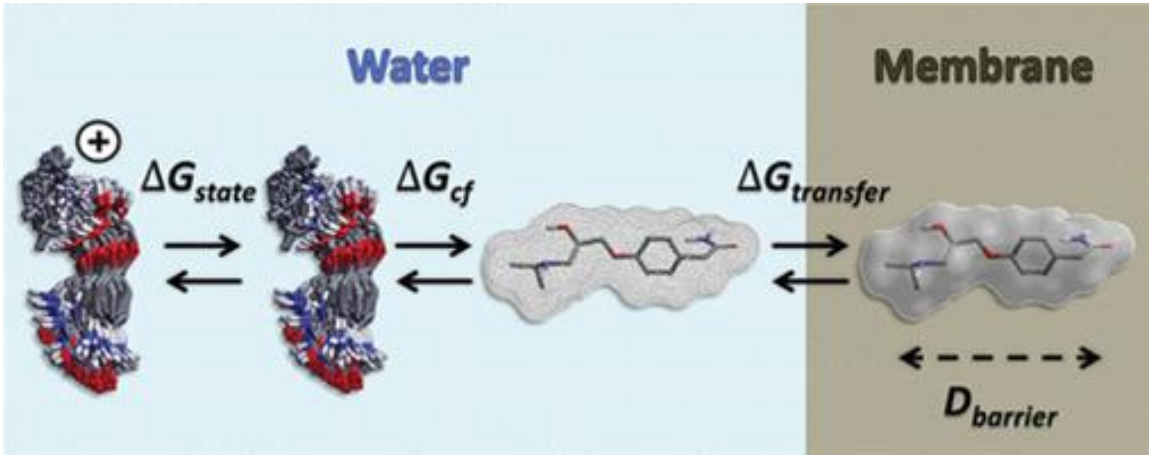
The toxicity prediction methods are also based on the presence of specific structural alerts in the compound's molecule and extrapolation (cross-reading) of specific toxicity patterns from known substances possessing same structural alerts.

A drawback of structural alert methods is that they rarely predict quantitave parameters, only classes and groupings.

- Physics-based methods

A recent trend, defined by continuous increase of computational power available, is used for prediction of passive membrane permeability. These methods try to avoid indirect capturing of the key physical aspects by statistical relationships.

Molecular-mechanics–based simulations in Schrodinger software platform calculates the total free energy penalty for the ligand to change state and enter the membrane using multiple generated conformations of the solute.

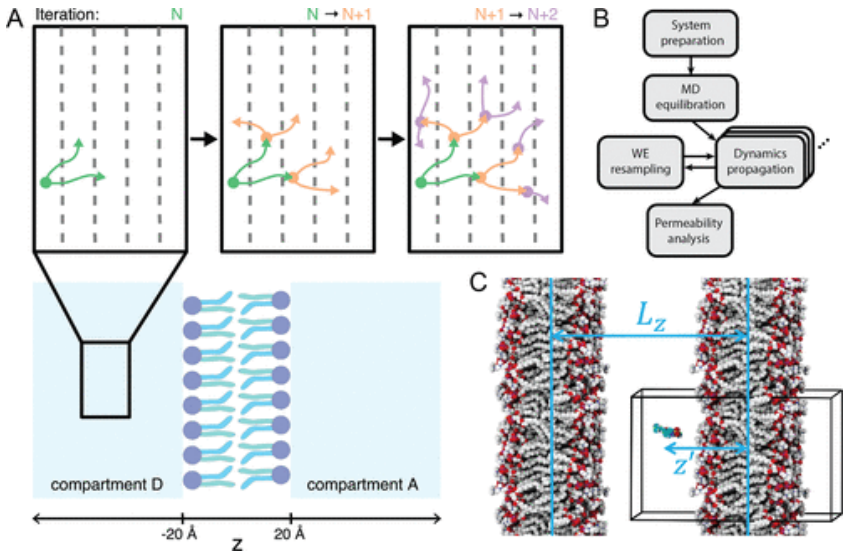



doi: 10.1021/ci200583t

How has it been done?

- Physics-based methods

Molecular-dynamics-based simulations in OpenEyes Orion Cloud platform uses weighted assemble approach to maximize number of membrane entry/exit events so to be able to calculate the free energy and enthalpy changes in solute permeation.



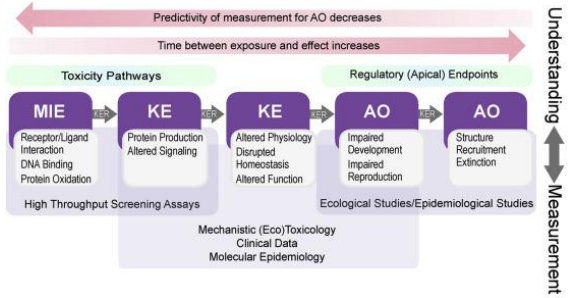

A drawback of physics-based is their high computational cost and very restricted area of application.

doi: 10.1021/acs.jcim.1c01540

How has it been done?

- Other:
  - Adverse outcome pathways (AOP)

AOPs could be used in toxicity predictions, based on the understanding that exposure to a xenobiotic may and, usually, have to invoke several steps in the organism to induce a toxic effect.



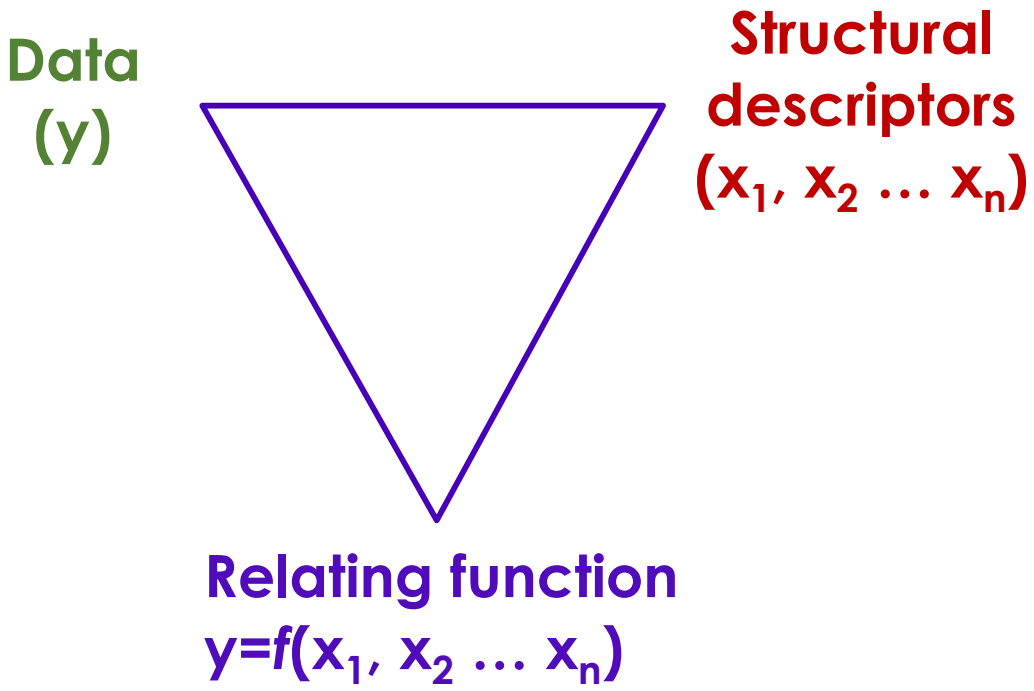

In practice, AOPs are a framework for integration of *in vivo*, *in vitro* and *in silico* results, greatly facilitating the understanding of compounds' toxic effects, but not always allowing for better prediction of specific dose/exposure regimens.

- Physiology-Based PharmacoKinetics (PBPK)
- In Vitro to In Vivo extrapolation (IVIV)

Despite their importance, these models operate on too many *in vitro* and *in vivo* experimental data to fill-in into the differential equation systems, so we hardly can classify them as pure *in silico* methods.

10.1016/j.cotox.2018.03.004

- QSAR modeling



- OECD and EC principles of building of reliable QSAR models

"In November 2004, the OECD Member Countries and the European Commission adopted five principles for the validation of (quantitative) structure-activity relationships ([Q]SARs) intended for use in the regulatory assessment of chemicals. International agreement on a set of validation principles was important, not only to provide regulatory bodies with a scientific basis for making decisions on the acceptability of data generated by (Q)SARs, but also to promote the mutual acceptance of (Q)SAR models by improving the transparency and consistency of (Q)SAR reporting."

The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN

- OECD and EC principles of building of reliable QSAR models

"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."

The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN

- A defined endpoint (deals with our **data**):

The intent of this principle is to ensure transparency in the endpoint being predicted by a given model, since a given endpoint could be determined by different experimental protocols and under different experimental conditions.

Ideally, (Q)SARs should be developed from homogeneous datasets in which the experimental data have been generated by a single protocol. However, this is rarely feasible in practice, and data produced by different protocols are often combined.

**Data** (membrane permeability):

- Flux: $F=dQ/dt$ (mol/sec)
- Effective permeability: $P_{app}=(dQ/dt) / (A \times C)$ (cm/sec)
- Membrane permeability: $P_m=(P_{app} \times P_{unstirred})/(P_{app} - P_{unstirred})$ (cm/sec)

The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN

10.1016/j.ejps.2003.10.009

- An unambiguous algorithm (deals with our **descriptors** and **relating function**):

The intent of this principle is to ensure transparency in the description of the model algorithm (not always the case with commercially-developed models).

**Molecular descriptors**:

- how they are obtained (more than 5000 molecular descriptors could be calculated by commercial and/or free software);
- how their preliminary filtering is done (low variability, collinearity);

**Relating function**:

- how the function (model) is selected;
- how the descriptors used in the model are selected (forward selection, backward elimination);

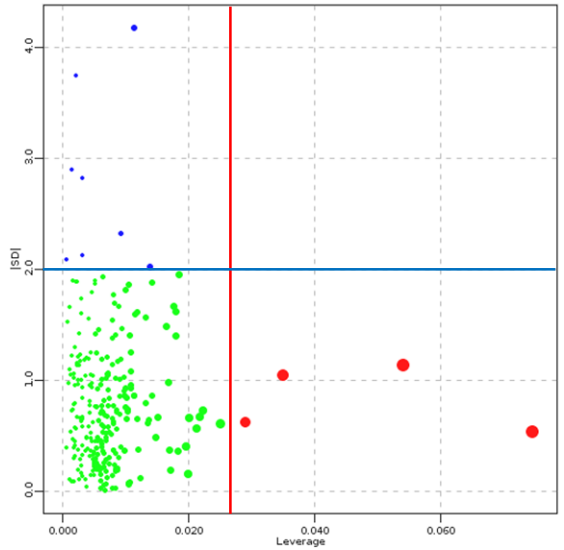
The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN
http://www.talete.mi.it/products/dragon_description.htm

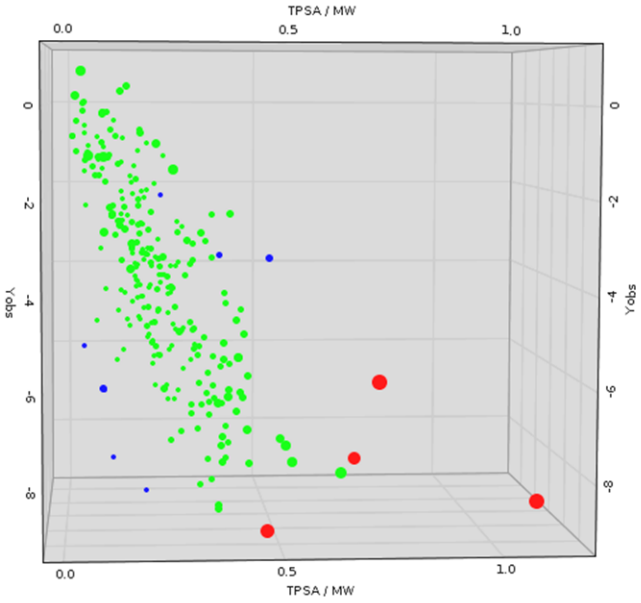
How shall we do it correctly?

- A defined domain of applicability (deals with our **data**, **descriptors** and **relating function**):

The need to define an applicability domain expresses the fact that (Q)SARs are reductionist models which are inevitably associated with limitations in terms of the types of chemical structures, physicochemical properties and mechanisms of action for which the models can generate reliable predictions.

Williams plot




The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN

10.1016/j.phymed.2018.09.001

How shall we do it correctly?

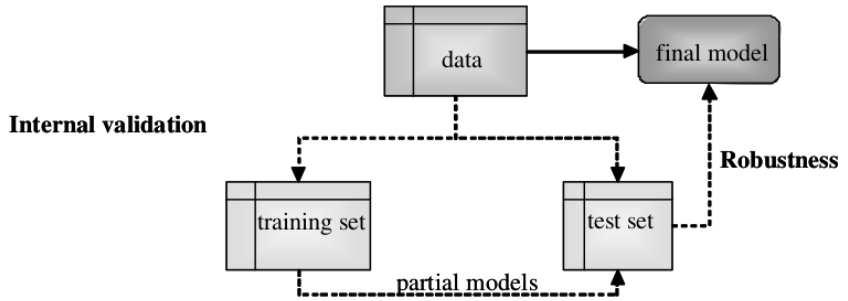
- Appropriate measures of goodness-of-fit, robustness and predictivity (deals mainly with our **relating function**):

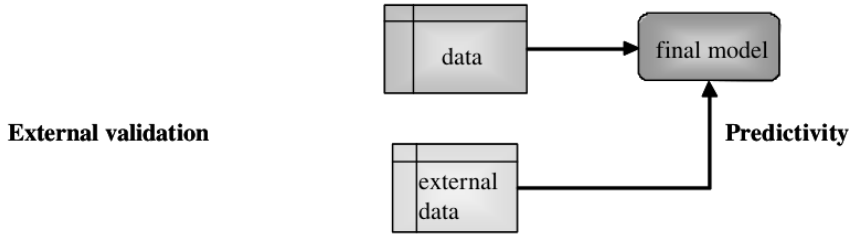
This principle expresses the need to provide two types of information: a) the internal performance of a model (as represented by coefficient of determination and robustness), determined by using a training set; and b) the predictivity of a model, determined by using an appropriate test set.

$$R^2 = 1 - \frac{\sum(y_{obs} - y_{calc})^2}{\sum(y_{obs} - y_{mean})^2}$$

$$R^2_{adj} = 1 - \frac{(n-1)\sum(y_{obs} - y_{calc})^2}{(n-k-1)\sum(y_{obs} - y_{mean})^2}$$

$$q^2 = 1 - \frac{\sum(y_{obs} - y_{pred})^2}{\sum(y_{obs} - y_{mean})^2}$$




The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN

- A mechanistic interpretation, if possible (deals mainly with our **descriptors** and **data**):

The intent of this principle is to ensure that there is an assessment of the mechanistic associations between the descriptors used in a model and the endpoint being predicted, and that any association is documented

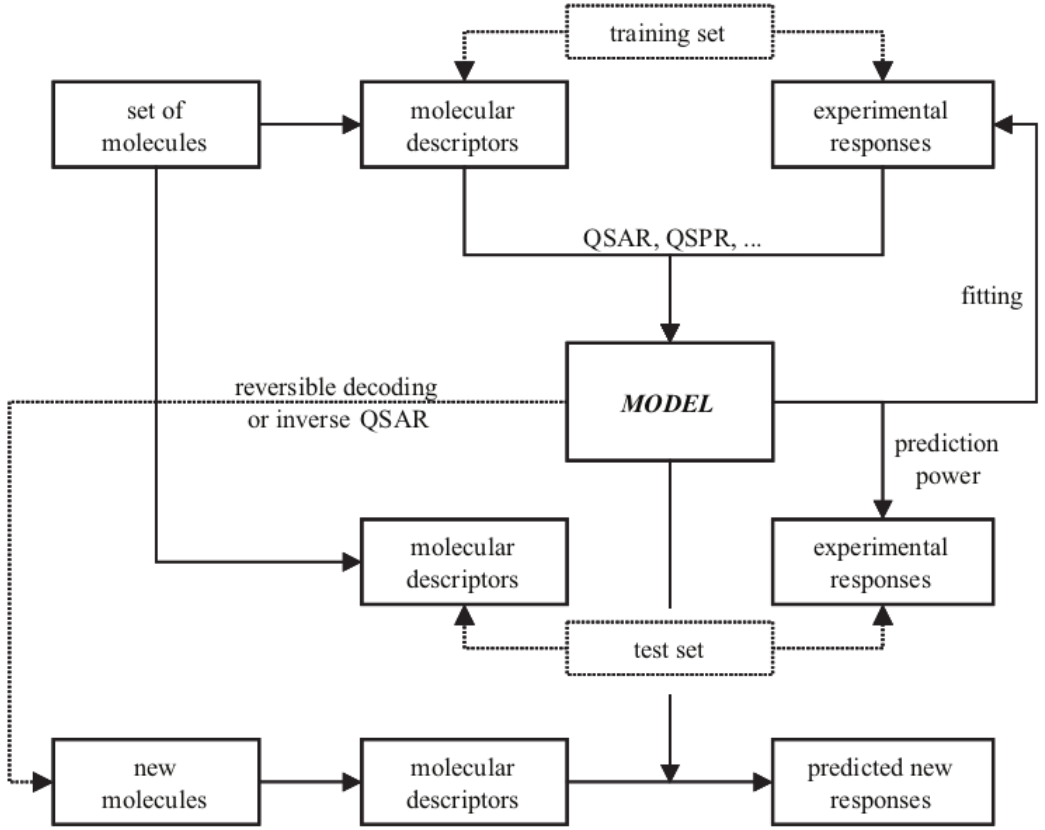
- Mechanism-based (Q)SARs are those (Q)SARs where a hypothesis is made as to the physicochemical properties or descriptors that are likely to be relevant. Statistical methods are then applied to seek out correlations existing between these descriptors and the endpoint of interest.

- In the case of purely empirical approaches, no assumptions are made as to the likely (biological) mechanism. A large number of physicochemical parameters or structural parameters are calculated and statistical approaches are applied to identify those features that correlate most closely with the biological activity.

The Characterisation of (Q)SARs: Preliminary Guidance. IHCP, ECB, 2005. EUR 21866 EN

General workflow

- QSAR modeling




ISBN 978-3-527-31852-0

- Figures-of-merit in QSAR modeling:

Coefficients of:

$$R^2 = 1 - \frac{\sum(y_{obs} - y_{calc})^2}{\sum(y_{obs} - y_{mean})^2}$$

determination ($R^2$)

$$q^2 = 1 - \frac{\sum(y_{obs} - y_{pred})^2}{\sum(y_{obs} - y_{mean})^2}$$

predictivity ($q^2$)

- Figures-of-merit in QSAR modeling:

$$q^2 = 1 - \frac{\sum(y_{obs} - y_{pred})^2}{\sum(y_{obs} - y_{mean})^2}$$

Silverman coefficient:

$$p^2 = 1 - \frac{\sum_{i,j}[(p_i - p_j) - (m_i - m_j)]^2}{\sum_{i,j}(m_i - m_j)^2}$$

$p_{i,j}$: predicted values $(y_{pred})$

$m_{i,j}$: observed values $(y_{obs})$

https://doi.org/10.1002/1521-3838(200006)19:3<237::AID-QSAR237>3.0.CO;2-A