Introduction

- QSAR
  - Quantitative Structure-Activity Relationships
- QSPR
  - Quantitative Structure-Property Relationships

What is QSAR?

- QSAR is a mathematical relationship between a biological activity of a molecular system and its geometric and chemical characteristics.
- This involves mathematical relationships linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds.

Quantitative Structure-Activity Relationships (QSAR)

and

Quantitative Structure-Property Relationships (QSPR)

The structure of a chemical influences its properties and biological activity

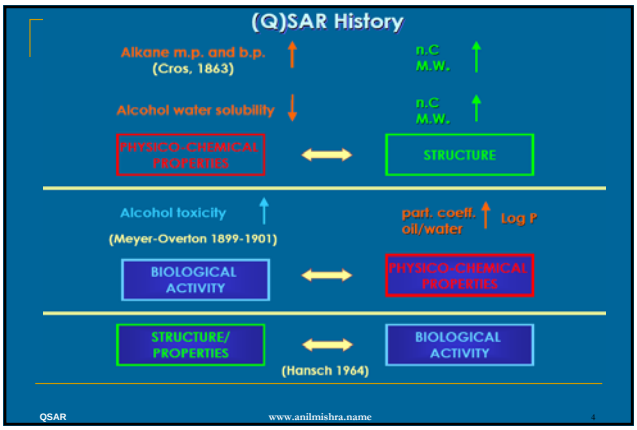
"Similar compounds behave similarly"

(Hansch 1964)

Activity or Property = f (Structure)

It is possible to find a relationship (f) between Structure and behavior (Activity or Property) of a chemical

PREDICTED DATA



QSAR

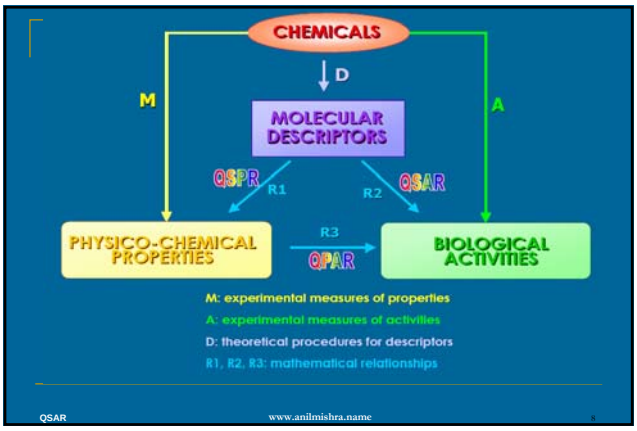
- QSAR attempts to find consistent relationship between biological activity and molecular properties, so that these "rules" can be used to evaluate the activity of new compounds.
- Methods which can be used in QSAR include various regression and pattern recognition techniques.

QSAR

- These physicochemical descriptors, which include parameters to account for
  - Hydrophobicity
  - Topology
  - Electronic properties
  - Steric effects
- These are determined empirically or, more recently, by computational methods.
- Activities used in QSAR include chemical measurements and biological assays.
- QSAR currently are being applied in many disciplines, with many pertaining to drug design and environmental risk assessment.

QSAR

- Biological activity can be expressed quantitatively as in the concentration of a substance required to give a certain biological response.
- Additionally, when physiochemical properties or structures are expressed by numbers, one can form a mathematical relationship, or quantitative structure-activity relationship, between the two.
- The mathematical expression can then be used to predict the biological response of other chemical structures.



Basic Requirements in QSAR Studies

- All analogs belong to a congeneric series
- All analogs exert the same mechanism of action
- All analogs bind in a comparable manner
- The effects of isosteric replacement can be predicted
- Binding affinity is correlated to interaction energies
- Biological activities are correlated to binding affinity

THE 3 NECESSITIES:

GOOD INPUT DATA

High-quality experimental data as input data to find the Structure-Activity Relation

MEANINGFUL STRUCTURAL INFORMATION

Good representation of the chemical structure: molecular descriptors

PREDICTIVE MODELS

Quantitative models with validated predictive performances

Experimental data set

There is a need for a "limited" number of HIGH-QUALITY experimental data on which to develop QSAR models!

The models will only be as good as the data used to develop them!
"Garbage in, garbage out"

NEEDS FOR EXPERIMENTAL DATA:

- AS NUMEROUS AS POSSIBLE
- CORRECT
- REPRESENTATIVE
- HOMOGENEOUS

(ideally, same lab, same method)

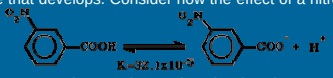
Linear Free Energy Relationships

- Little additional development of QSAR occurred until the work of Louis Hammett (1894-1987), who correlated electronic properties of organic acids and bases with their equilibrium constants and reactivity.
- Consider the dissociation of benzoic acid

Linear Free Energy Relationships

- Hammett observed that adding substituents to the aromatic ring of benzoic acid had an orderly and quantitative effect on the dissociation constant. For example,

- a nitro group in the meta position increases the dissociation constant, because the nitro group is electron-withdrawing, thereby stabilizing the negative charge that develops. Consider now the effect of a nitro group in the para position:



- The equilibrium constant is even larger than for the nitro group in the meta position, indicating even greater electron-withdrawal.

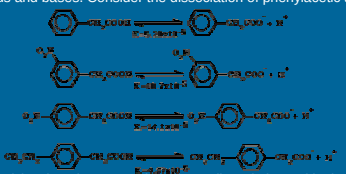
Linear Free Energy Relationships

- Now consider the case in which an ethyl group is in the para position:

- In this case, the dissociation constant is lower than for the unsubstituted compound, indicating that the ethyl group is electron-donating, thereby destabilizing the negative charge that arises upon dissociation.

Linear Free Energy Relationships

- Hammett also observed that substituents have a similar effect on the dissociation of other organic acids and bases. Consider the dissociation of phenylacetic acids:

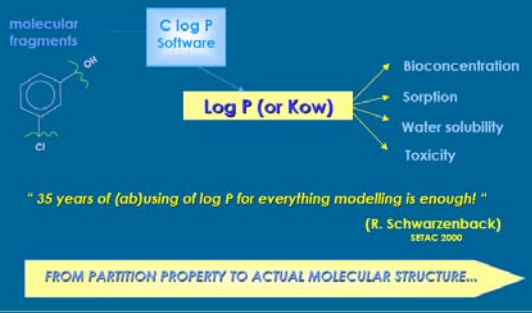


- Electron-withdrawal by the nitro group increases dissociation, with the effect being less for the meta than for the para substituent, just as was observed with benzoic acid.
- The electron-donating ethyl group decreases the equilibrium constant, as would be expected.

Linear Free Energy Relationships

- Hammett constant takes into account both **resonance and inductive** effects; thus, the value depends on whether the substituent is *para or meta* substituted
- --*ortho* not measured due to steric effects
- In some positions only inductive effects effect & some both resonance & inductive effects play a part
- Aliphatic electronic substituent constants are also available

The "magic" molecular descriptor



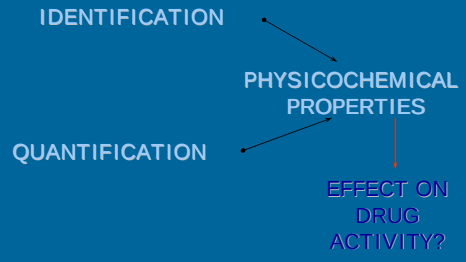
" 35 years of (ab)using of log P for everything modelling is enough! "

(R. Schwarzenback)
SETAC 2000

FROM PARTITION PROPERTY TO ACTUAL MOLECULAR STRUCTURE...

Role of QSAR

- To optimise lead identification by:



QSAR

- Quantitative structure-activity relationships are used when there is little or no receptor information, but there are measured activities of (many) compounds
- Quantitative structure-activity relationships correlate chemical/biological activities with structural features or atomic, group or molecular properties within a range of structurally similar compounds

QSAR

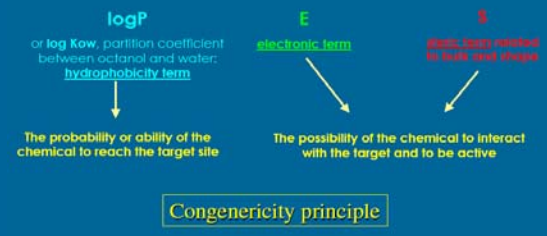
- Attempts to identify and quantitate physicochemical properties of a drug in relation to its biological activity or binding
- Studies hydrophobic, electronic, and steric properties--either whole molecule or pieces
- *Medicinal chemist draws up an equation that quantifies the relationship & allows one to predict (to some extent) the biological activity*

First Approaches: The Early Days

- Free- Wilson Analysis
  - Free-Wilson analysis is a regression technique using the presence or absence of substituents or groups as the only molecule descriptors in correlations with biological activity
- Hansch Analysis
  - Hansch analysis is the investigation of the quantitative relationship between the biological activity of a series of compounds and their physicochemical substituent or global parameters representing hydrophobic, electronic, steric and other effects using multiple regression correlation methodology

Classical Hansch equation:

"Toxicity" = a + b logP + c E + d S



Congenericity principle

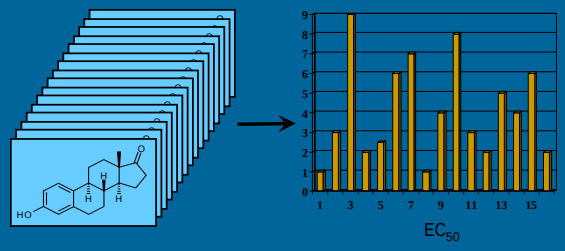
QSAR Approach

- Equations are produced to predict relationship properties have on the mechanism/distribution of drug
- Used as a prediction of biological activity
- targets leads with improved activity, minimises number of analogues made.
- If analogue does not 'fit' the equation, identifies some other feature of the compound as important
  - review development

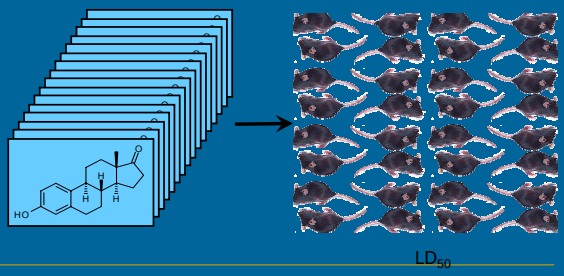
General Procedure of QSAR

- Select a set of molecules interacting with the same receptor with known activities.
- Calculate features (e.g. physical, chemical properties, etc., 2D, 3D)
- Divide the set to two subgroups: one for training and one for testing.
- Build a model: find the relations between the activities and properties (regression problem, statistic methods, machine learning approaches, etc).
- Test the model on the testing dataset.
- You can also develop new descriptors, new methodologies, algorithms, etc.

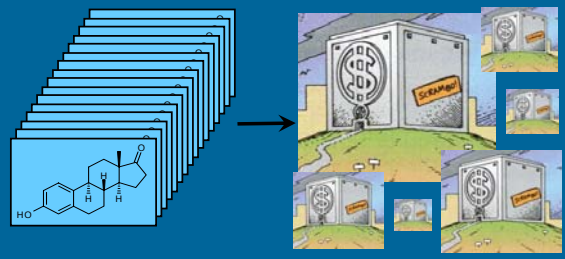
From Structure to Property



From Structure to Property



From Structure to Property

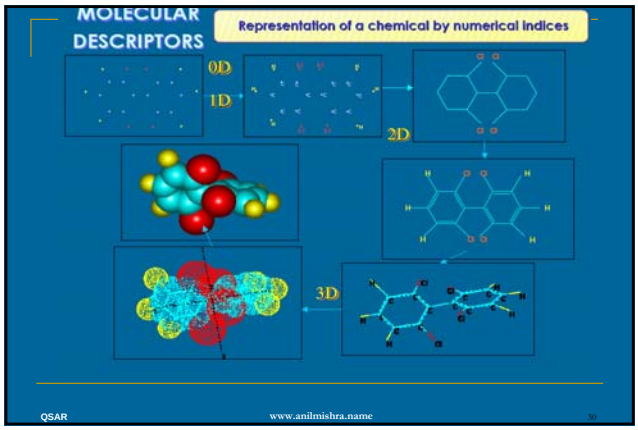


Dimensions in QSAR

- 2D-QSAR models are based on descriptors derived from a two dimensional graph representation of a molecule
- 3D-QSAR involves the analysis of the quantitative relationship between the biological activity of a set of compounds and their three-dimensional properties using statistical correlation methods.

Dimensions in QSAR

- 1D - molecular formula
- 2D - molecular connectivity / topology
- 3D - molecular geometry / stereochemistry
- 4D/5D/… - conformational ensembles



Advantages of QSAR:

- Quantifying the relationship between structure and activity provides **an understanding of the effect of structure on activity**.
- It is also **possible to make predictions** leading to the synthesis of novel analogues.
- The results can be used to help **understand interactions** between functional groups in the molecules of greatest activity, with those of their target.

Disadvantages of QSAR

- **False correlations** may arise because biological data that are subject to considerable experimental error (noisy data).
- If training dataset is not large enough, the data collected **may not reflect the complete property space**. Consequently, many QSAR results cannot be used to confidently predict the most likely compounds of best activity.
- Features may not be reliable as well. This is particularly serious for 3D features because 3D structures of ligands binding to receptor may not be available. Common approach is to use minimized structure, but that may not represent the reality well.
- **Bottom-line**: there are many successful applications but do not expect QSAR works all time. Also be aware of overfitting problem!!

SAR vs. QSAR

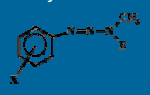
- SAR is supposed to be not quantitative concept
- SAR is based on the notion of "similarity" :
  - *"Similar compounds have similar activity"*
  - *"Dissimilar compounds have dissimilar activity"*
- QSAR aims to derive a quantitative model of the activity

Application

- Researchers have attempted for many years to develop drugs based on QSAR. Easy access to computational resources was not available when these efforts began, so attempts consisted primarily of statistical correlations of structural descriptors with biological activities.

Application

- An early example of QSAR in drug design involves a series of 1-(X-phenyl)-3,3-dialkyl triazenes



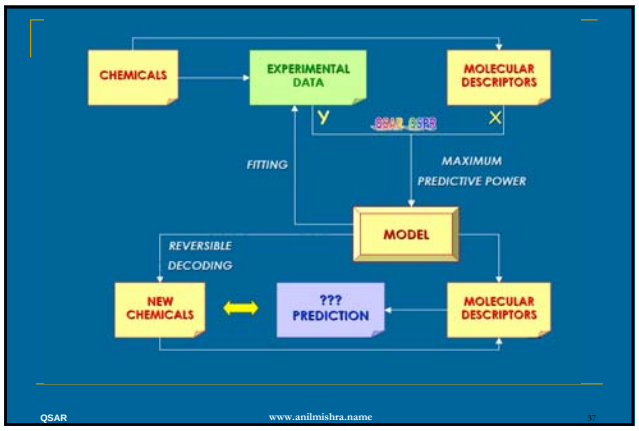
- These compounds were of interest for their anti-tumor activity, but they also were mutagenic. QSAR was applied to understand how the structure might be modified to reduce the mutagenicity without significantly decreasing the anti-tumor activity.

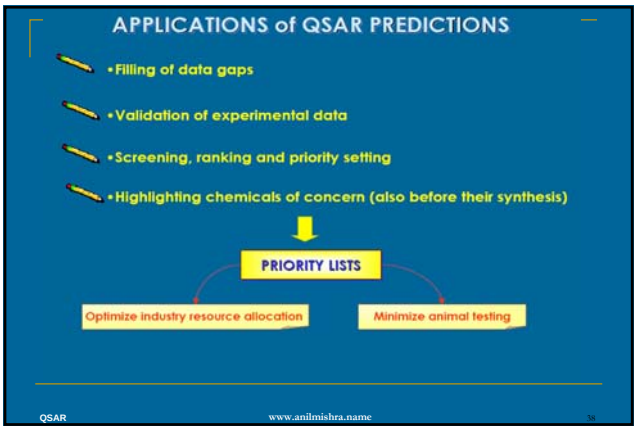
Application

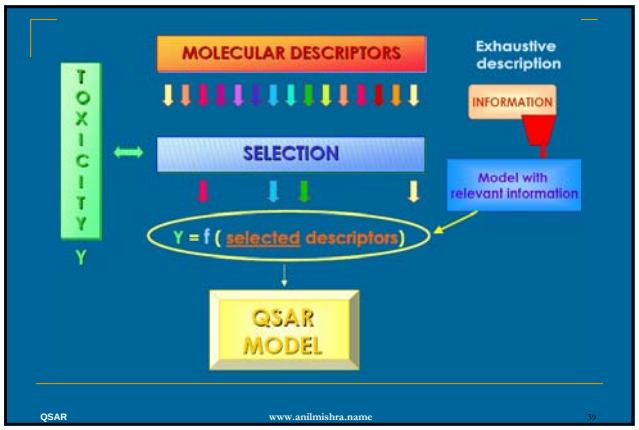
- Based on equations it was observed that mutagenicity is more sensitive than anti-tumor activity to the electronic effects of the substituents. Thus, electron-withdrawing substituents were examined

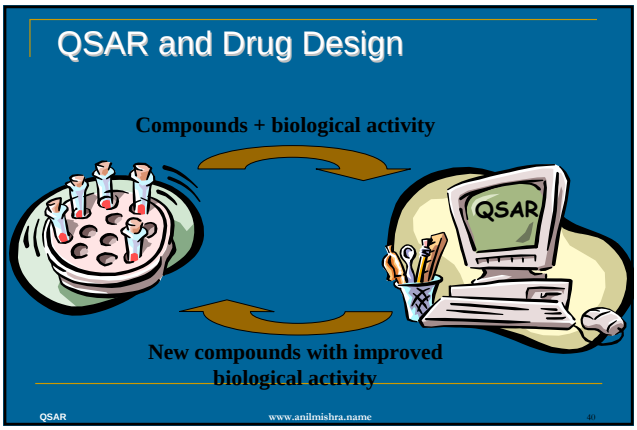
4-X substituent	log(1/C) mutagenicity	antitumorigenicity
-H	5.75	3.58
$-SO_2NH_2$	3.15	3.48

- By substituting a sulfonamide group at the para position, the anti-tumor activity was reduced 1.2-fold, whereas the mutagenicity was reduced by about 400-fold.


CHEMICALS
EXPERIMENTAL DATA
MOLECULAR DESCRIPTORS
Y
QSAR, QSPR
X
FITTING
MAXIMUM PREDICTIVE POWER
MODEL
REVERSIBLE DECODING
NEW CHEMICALS
??? PREDICTION
MOLECULAR DESCRIPTORS
QSAR
www.anilmishra.name


APPLICATIONS of QSAR PREDICTIONS
•Filling of data gaps
•Validation of experimental data
•Screening, ranking and priority setting
•Highlighting chemicals of concern (also before their synthesis)
PRIORITY LISTS
Optimize industry resource allocation
Minimize animal testing
QSAR
www.anilmishra.name


MOLECULAR DESCRIPTORS
Exhaustive description
INFORMATION
TOXICITY
SELECTION
Model with relevant information
Y = f (selected descriptors)
Y
QSAR MODEL
QSAR
www.anilmishra.name


QSAR and Drug Design
Compounds + biological activity
QSAR
New compounds with improved biological activity
QSAR
www.anilmishra.name