

# Fragmental Methods in Accelerated Lead Optimization. Applications of Algorithm Builder

P. Japertas, L. Riauba, D. Zmuidinavicius, R. Didziapetris, and A. Petrauskas

Pharma Algorithms, Inc., [www.ap-algorithms.com](http://www.ap-algorithms.com)

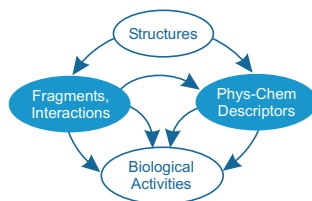
## Algorithm Builder

Algorithm Builder (AB) is a multitasking software system that:

- Integrates a variety of fragmental methods and calculates essential physico-chemical descriptors and statistics.
- Provides object-based, "click-drag-drop" development of new calculation algorithms and screening methods.
- Simplifies the design and selection of new compounds that have to be brought to the next stage of development

## Model Generation

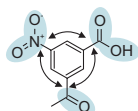
- Generate structural and phys-chem descriptors.
- Analyze phys-chem properties by fragmental QSPR
- Analyze biol. activities in terms of C-SAR, fragmental QSPR and descriptor-based QSAR.
- Integrate "structure-based design" and "property-based design" into a consistent approach.



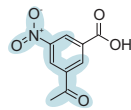
## Fragmental Methods

- Hansch-Leo (IC-based)**. Used to predict logP, logS (solubility) and other additive-constitutive properties in terms of the following equation:

$$\text{Log}X = \sum \text{Fragments} + \sum \text{Interactions}$$



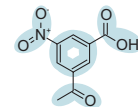
- Used in C-SAR analysis to identify structural factors ("biophores") that group compounds into classes according to biological activities.



- Pattern or Branched Chain** (similar to Klopman's linear chains). Suitable for identification of unknown biophores in C-SAR analysis of biological activities.

- Custom (Free-Wilson type)**. Suitable for analysis of congeneric series:

$$\text{Log}X = \sum \text{Skeletons} + \sum \text{Radicals}$$



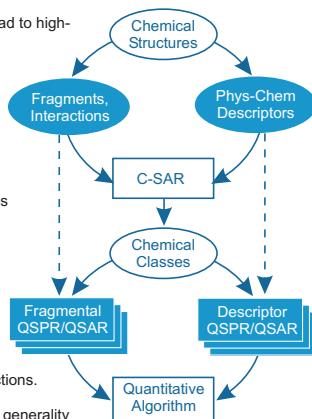
- Custom-defined skeletons can be used in other types of fragmentations to increase the efficiency and simplify interpretations of C-SAR and QSAR analyses.

## Combined Analysis

Combine qualitative and quantitative methods that lead to high-quality algorithms:

### C-SAR analysis

- Custom-defined skeletons can be used in other types of fragmentations to increase the efficiency and simplify interpretations of C-SAR and QSAR analyses.
- Identifies factors that group compounds into classes according to biological activities
- Leads to derivation of **rule-based filters**

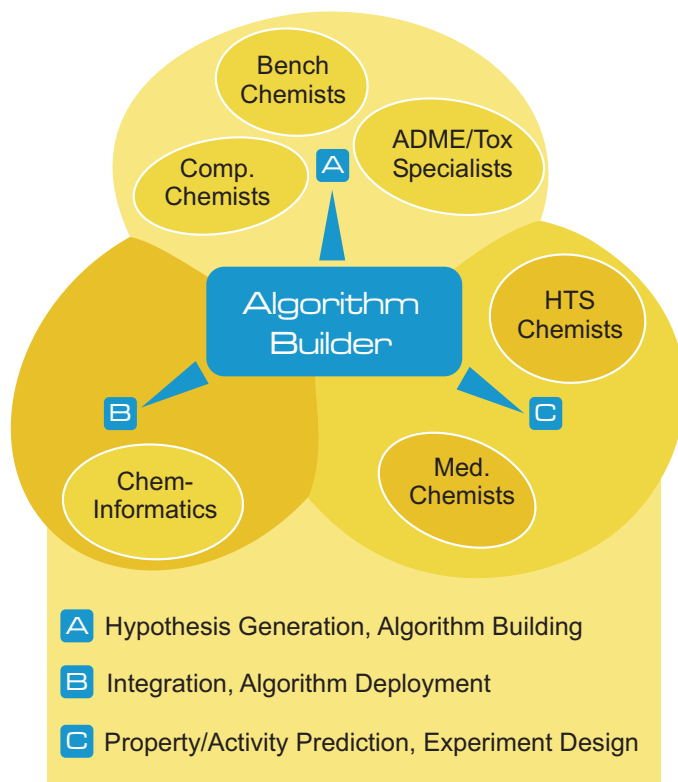


### QSAR/QSPR analysis

- Is always class-specific, as defined by the results of C-SAR analysis
- Fragmental QSPRs ensure high accuracy of predictions.

Alternatively, descriptor-based QSPRs ensure high generality of predictions.

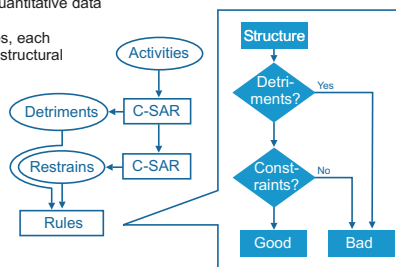
- Leads to derivation of quantitative algorithms



## Rule-Based Filters

Development of rule-based filters:

- Can be based on either qualitative or quantitative data
- Involves a sequence of C-SAR analyses, each considering many kinds of descriptors (structural and physico-chemical)
- Each analysis identifies certain rules that prohibit or enhance the biological effect of a new compound
- Replaces "trial-and-error" approaches based on similarity searching and/or data mining



### Example

Development of the **HIA Filter** for identifying compounds with poor intestinal permeability:

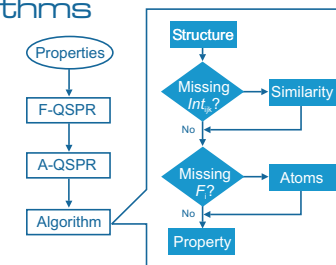
- Used experimental HIA values for >1,000 drugs. All data was converted into a binary format (0 if HIA<10%, 1 if HIA>10%).
- To distinguish paracellular transport, non-restricted and restricted membrane diffusion, all compounds were subdivided into three data sets according to MW.
- Within each data set, a series of C-SAR analyses was performed, leading to the development of the AB/HIA filter.

HIA	<10%	>10%
Exp.	98	879
True	83	852
False	15	27

## Quantitative Algorithms

Development of quantitative algorithms:

- F-QSPR denotes fragmental QSPR.
- A-QSPR denotes atom-based QSPR used to predict new fragmental increments ( $F_i$ ).
- If some interaction increments ( $Int_{ij}$ ) are missing, they are estimated by similarity-based analysis, whereas  $F_i$  are estimated from atomic increments.



### Example

Development of the **AB/LogP** calculation algorithm:

- Involved analysis of exp. data for 9,362 compounds.
- $F_i$  variables were analyzed in terms of atom-based QSPR.  $Int_{ij}$  variables were generalized by using similarity-based hierarchical clustering.
- After removing the single-point determinations and complex linear dependencies, the following statistics were obtained: N=8,760, SD=0.37, R<sup>2</sup>=0.96.

### Comparative Validation (N=180)

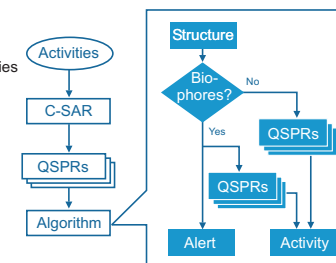
Program	AARS*		
	All	Drugs	Simple
AB/LogP	0.26	0.36	0.17
KOWWIN	0.28	0.44	0.13
CLOGP	0.30	0.51	0.10
SciLogP	0.31	0.44	0.18

\* AARS - Averaged abs. residual sums.  
All - all compounds,  
Drugs - 90 drugs,  
Simple - 90 simple organics.

## Hybrid Algorithms

Development of hybrid algorithms:

- Requires large databases of biological activities (e.g. screened generic/combinatorial libraries).
- C-SAR analysis identifies substructures ("biophores") and phys-chem factors that may cause changed biological mechanisms.
- Within each class of compounds, fragmental and descriptor-based QSPR analyses are performed.



### Example

**Tox Algorithm** for estimating LD50 values (mouse, intraperitoneal administration):

- Involved analysis of exp. LD50 values for >30,000 compounds.
- C-SAR analysis identified multiple toxicophores with particular types of toxic effects.
- After deriving class-specific QSPRs, the correlation between exp. and calc. values: N=30,242, SD=0.37, R=0.83

Toxic	Training		Validation	
	Yes	No	Yes	No
Exp.	1,599	28,643	175	3,164
True	1,296	26,391	130	2,929
False	303	2,252	44	235

## Lead Optimization

Multidisciplinary teams of scientists can benefit from using AB in the following ways:

- Analyze any type of data in order to generate and validate hypotheses.
- Replace manual "trial-and-error" approaches with automated C-SAR and fragmental QSAR methods.
- Develop new filters and algorithms for screening new compounds.
- Use existing AB filters to identify compounds with desirable ADME/Tox properties.
- Utilize existing company-wide databases to predict new types of ADME/Tox effects.
- Automate the process of selecting good and poor hits based on a cascade of standard and in-house filters that model real processes in the human organism.

