

# Large-Scale *in silico* Model Building

## Authors

Ulf Norinder<sup>1</sup>, Sarah Aaron<sup>2,3</sup>, David Buttar<sup>4</sup>, Lars Carlsson<sup>5</sup>, Hongming Chen<sup>6</sup>, Ryszard Czereminski<sup>7</sup>, Andy Davis<sup>8</sup>, Johan Ulander<sup>9</sup>, Dan Warner<sup>10</sup>, David Wood<sup>11</sup>

## Contact email

ulf.norinder@astrazeneca.com

Development of methods and procedures for industrialised large scale *in silico* model building is an important part of the 'Predictive Chemistry Programme' initiative within AstraZeneca.

Other important aspects considered in the program are areas such as: Accuracy, robustness and domain coverage of derived (ADMET) *in silico* models as well as the balance between complexity and transparency for the resulting models.

This poster displays two of the platforms developed at AstraZeneca R&D for enabling large scale *in silico* model building in an industrialised setting:

## AutoQSAR AZOrange

AZOrange and AutoQSAR are complementary tools in our predictive modeling platform.

AZOrange is a high-performance QSAR system designed for development and evaluation of new machine learning methods and descriptors.

AutoQSAR is an enterprise system to expose validated methods, e.g. by AZOrange, to design teams in AZ – the competitive workflow chooses methods and models to make the best predictions for the latest datasets.

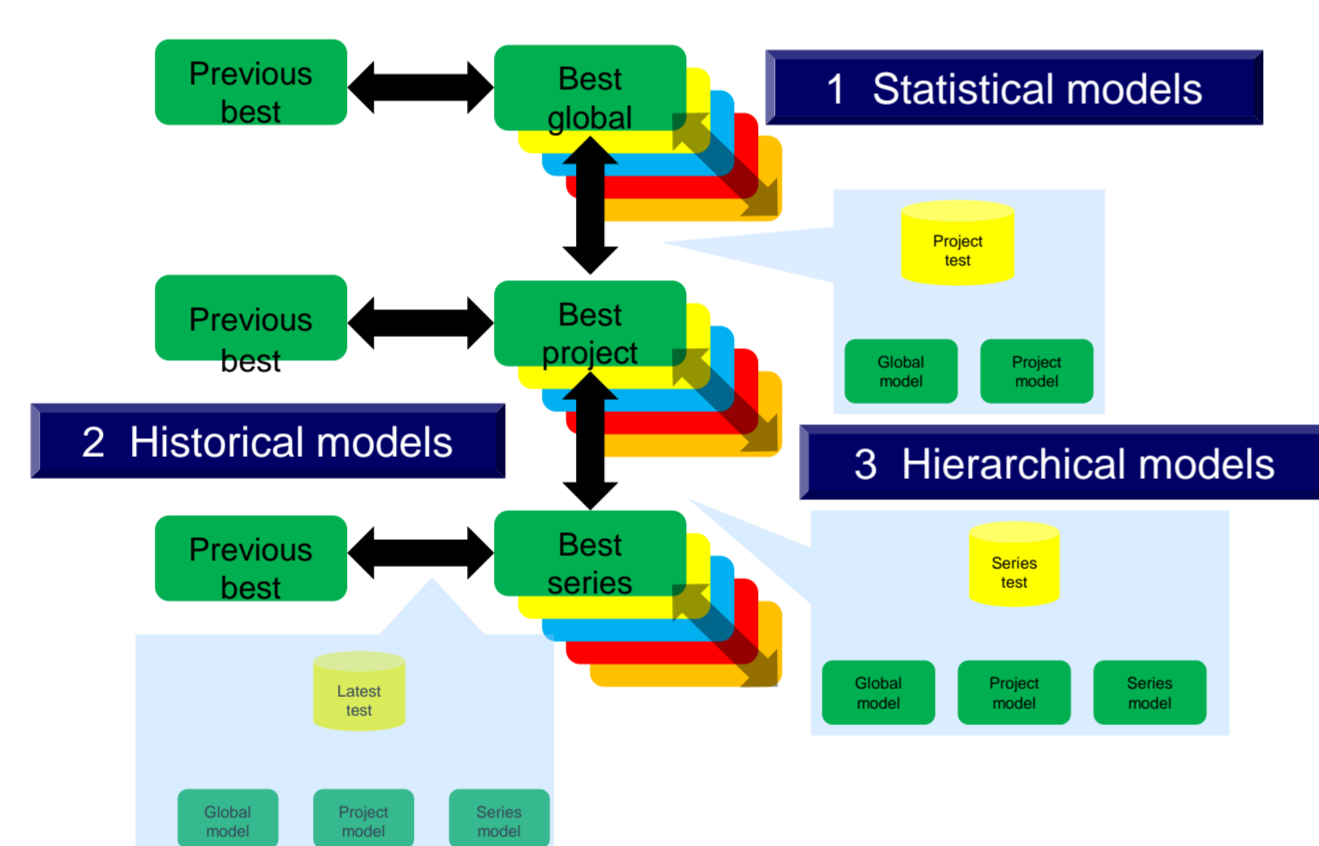
## AutoQSAR Platform

The AutoQSAR system is designed to automate much of the QSAR modelling process. Drug design scientists can register their datasets to the system, and AutoQSAR will build models and keep them updated so they reflect the most recent measurements.

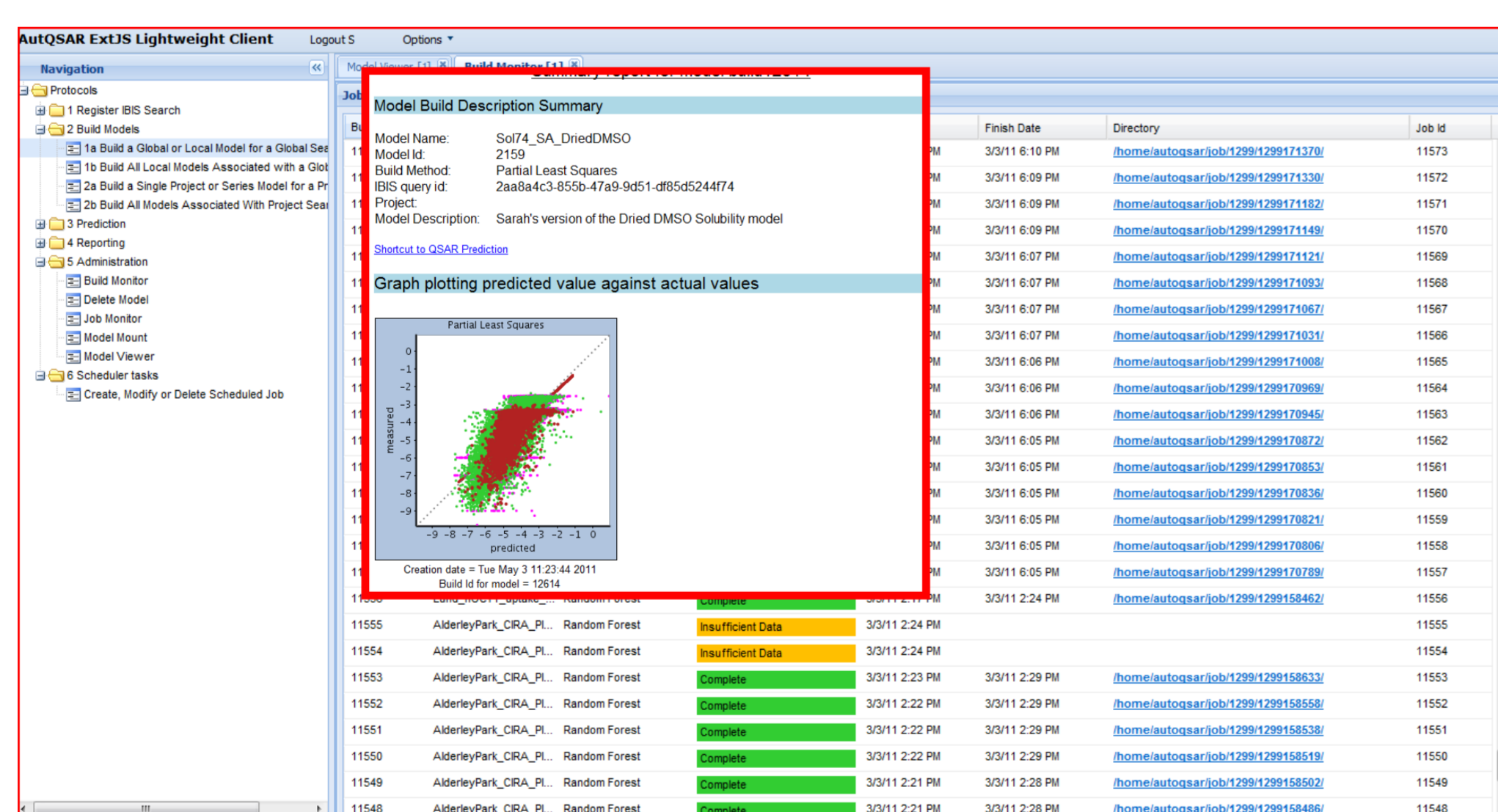
**Learning Algorithms:** The AstraZeneca AutoQSAR (AQ) platform enables automated model building with a range of statistical learning algorithms, including:

- Partial Least Squares
- Random Forest
- Bayesian Neural Network

**AZ Descriptors:** a compilation of 196 molecular descriptors of physico-chemical nature are used [1]. These descriptors contain 1D and 2D, as well as 3D, descriptors including properties such as various counts of atoms and bonds, charges, surfaces and lipophilicity..



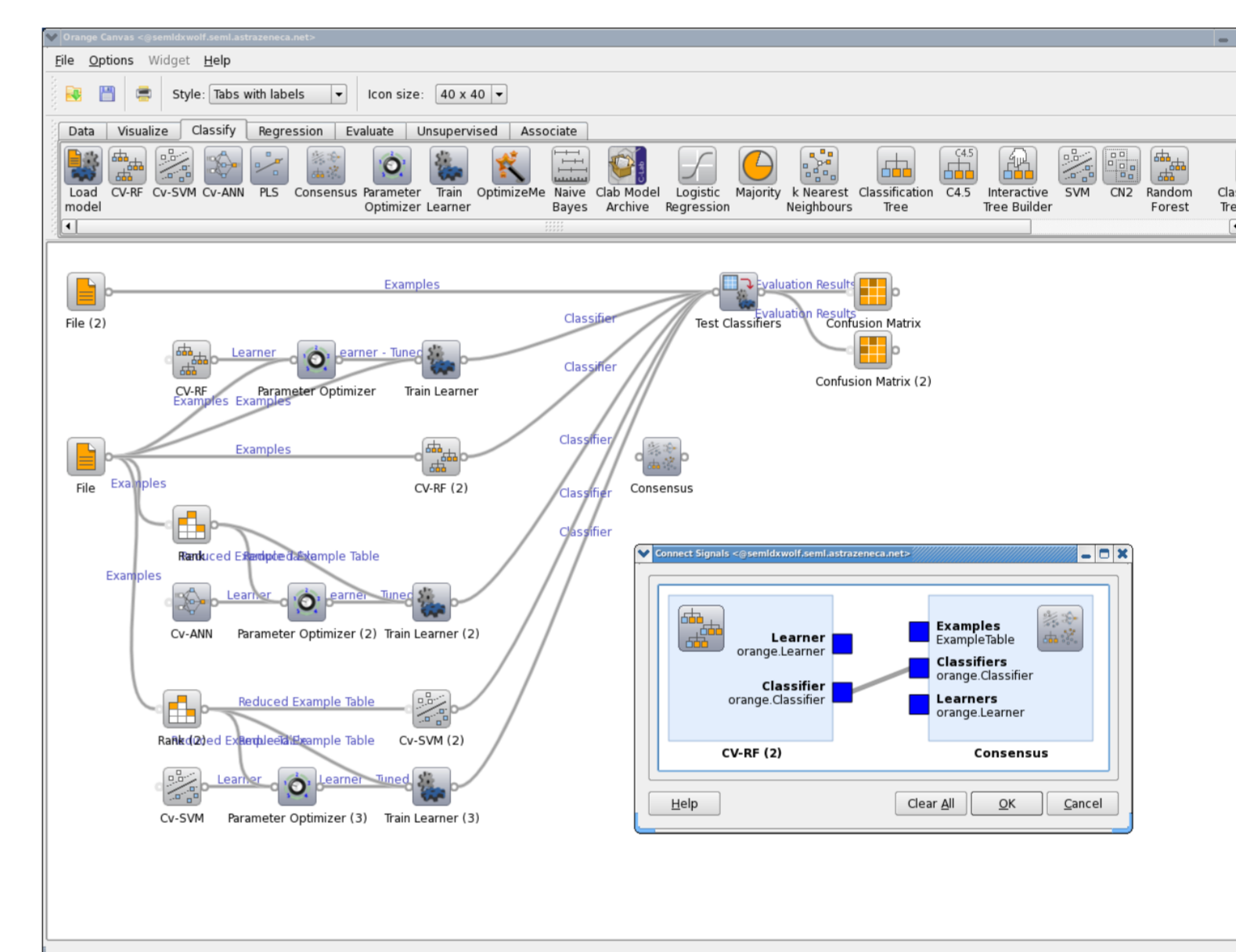
**Competitive Workflow:** The competitive workflow selects the most predictive model for a particular chemical domain from a pool of local and global models. Current and historical models built with the various learning algorithms are included in the model pool. The most predictive model is then made available to all users via our global model and descriptor platform, C-Lab.



AutoQSAR is currently being rolled out across AstraZeneca and is already starting to have an impact on projects in providing local focussed models as well as more generic global models. Since AutoQSAR automates the process of model building, it is now easy for all projects to have multiple predictive models generated using their own data.

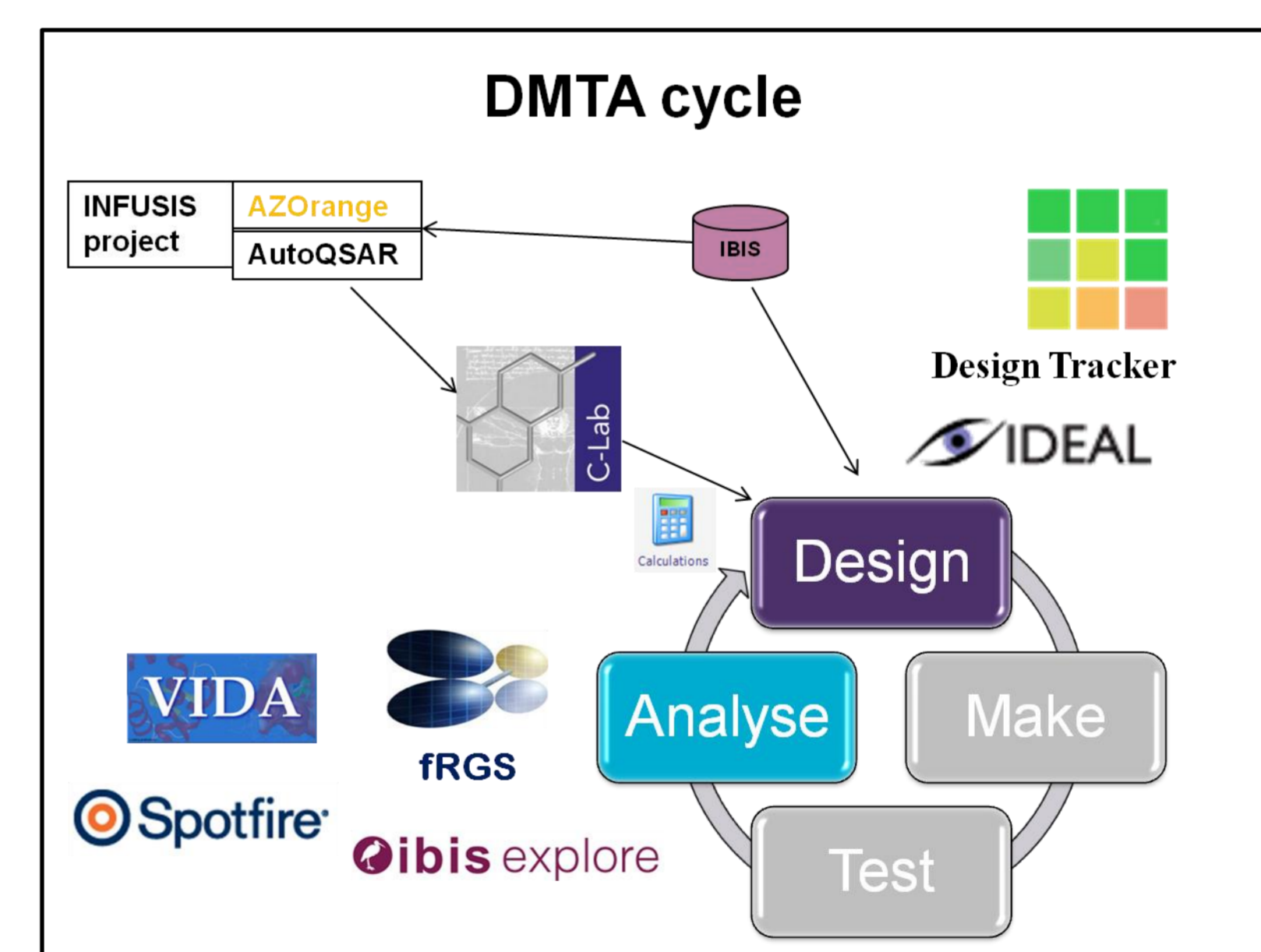
## AZOrange Platform

AZOrange [2] is a comprehensive platform for QSAR development with an integrated, highly flexible work flow built on Orange [3]. The package comprises methods for data retrieval, descriptor calculation, data pre-processing, model building and evaluation. Apart from the work flow, it also has extensive scripting capabilities. An automated model-parameter optimizer make batch development of QSAR models possible, testing multiple combinations of descriptors, parameters and machine-learning methods.



Example of Workflow in AZOrange.

This example shows a typical workflow in AZOrange to derive and validate QSAR/QSPR models



DMTA cycle @ AstraZeneca R&D

Figure shows the DMTA cycle employed at AstraZeneca R&D with various software components including the AutoQSAR and AZOrange platforms for automated *in silico* model building (mod. from presentation by John Cumming).

## References

1. Bruneau, P. Search for Predictive Generic Model of Aqueous Solubility Using Bayesian Neural Nets. *J. Chem. Inf. Comput. Sci.*, 41, 1605–1616 (2001)
2. <https://github.com/AZCompTox/AZOrange>
3. <http://orange.biolab.si/>