

AutoQSAR

Automation of the QSAR Modelling Process

UK-QSAR – Autumn Meeting, 2011

David Wood, Discovery Sciences

Andy Davis, Respiratory & Inflammation

Sarah Rodgers, Accelrys

Overview

- Predictive Chemistry Programme
- Introduction to AstraZeneca AutoQSAR platform
 - What is AutoQSAR and Why has it been Developed?
- Validation of the AutoQSAR Approach
 - *Automated QSAR with a Hierarchy of Global and Local Models*
- Impact of AutoQSAR at AstraZeneca
- What Next?
- Summary



Finding Better Compounds Faster

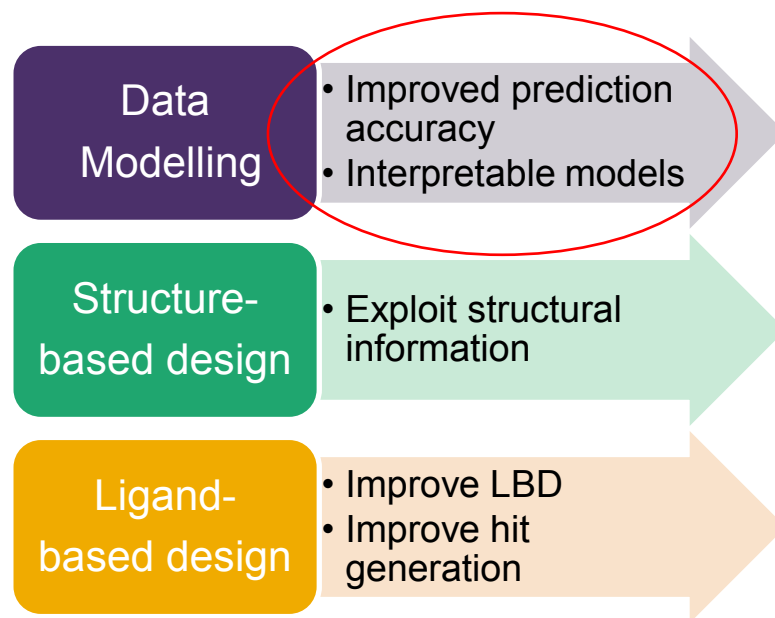
- **Predictive Chemistry Programme** @AstraZeneca

- External collaborations
- Internal projects
- Postdocs

- **Aims:**

- Enhance design quality and efficiency
- Improve prediction accuracy

- **Tools:**



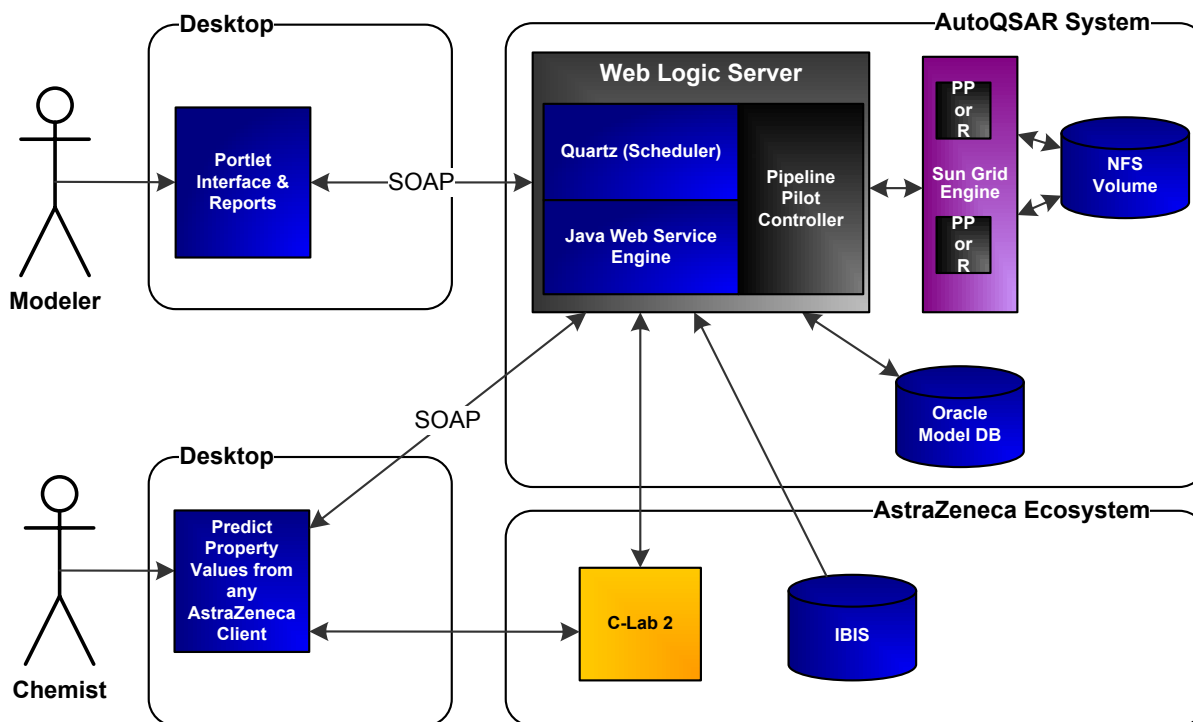
Predictive Chemistry Vision:

All compounds designed at AstraZeneca will have key properties calculated before synthesis



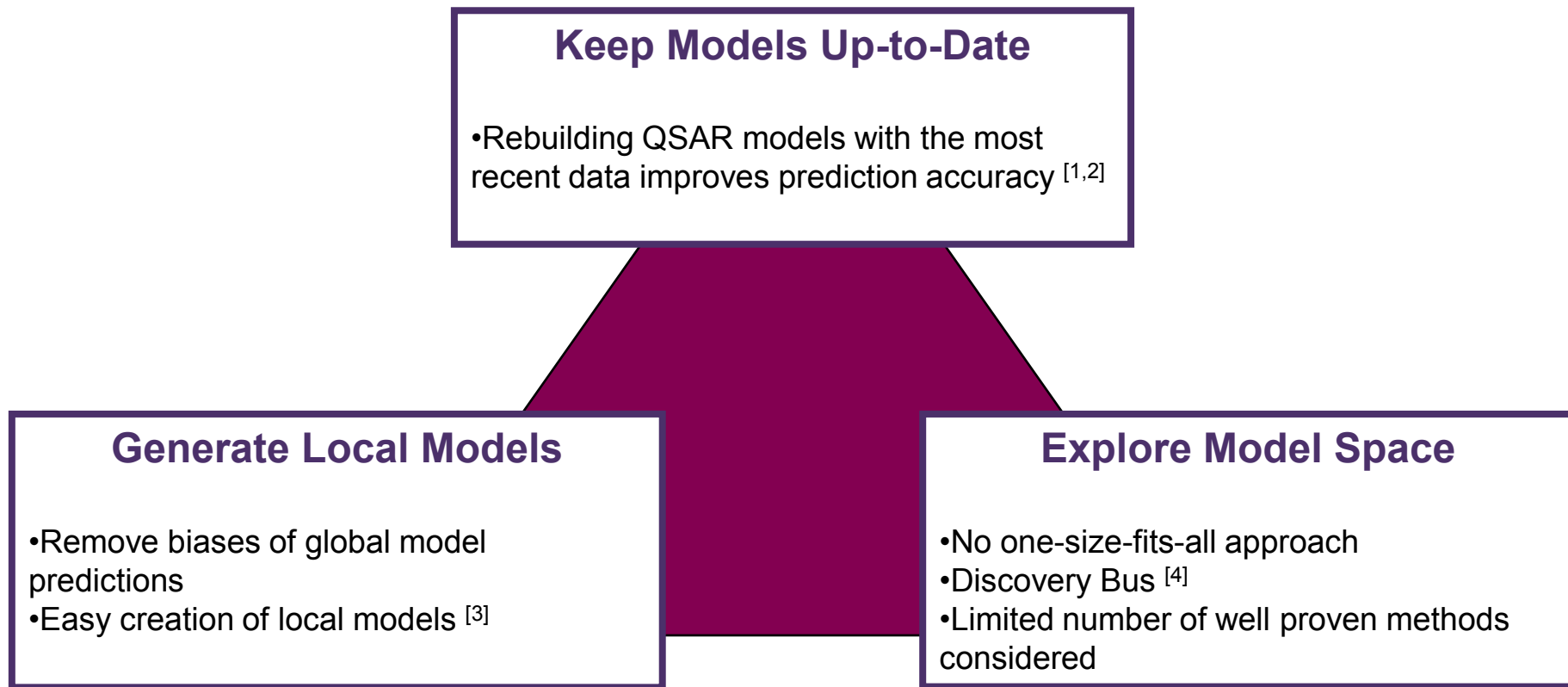
What is AutoQSAR?

A proprietary system developed at AstraZeneca in collaboration with Accelrys for the automatic creation, evaluation and maintenance of QSAR models



Why AutoQSAR? – Science Motivations

Improve Prediction Accuracy



1) Rodgers *et al.* *QSAR Comb. Sci.* **26**, 2007, p511

2) Rodgers *et al.* *JCIM*, **47**, 2007

3) Wood *et al.* *MolInf*, just accepted

4) Cartmell *et al.*, *JCAMD*, **19**, 2005, p821



Why AutoQSAR? – Business Motivations

Standardize AstraZeneca's Approach to QSAR

- Exploit all data generated within AstraZeneca drug discovery programs

Free Up Chemistry Resources

More time to...

- Research new methods
- Refine QSAR protocols
- Build bespoke models

Track Performance

- Demonstrate impact on compound design



Why AutoQSAR?

- Build global models
- Build project models
- Build chemical series models
- Objectively choose best to predict new compounds

- Scope
 - Global models - 20 models /week
 - 100 projects
 - Project/series models 20 endpoints = $20 \times 100 = 2000$
 - Explore 4 stats methods
 - 8000 models/week

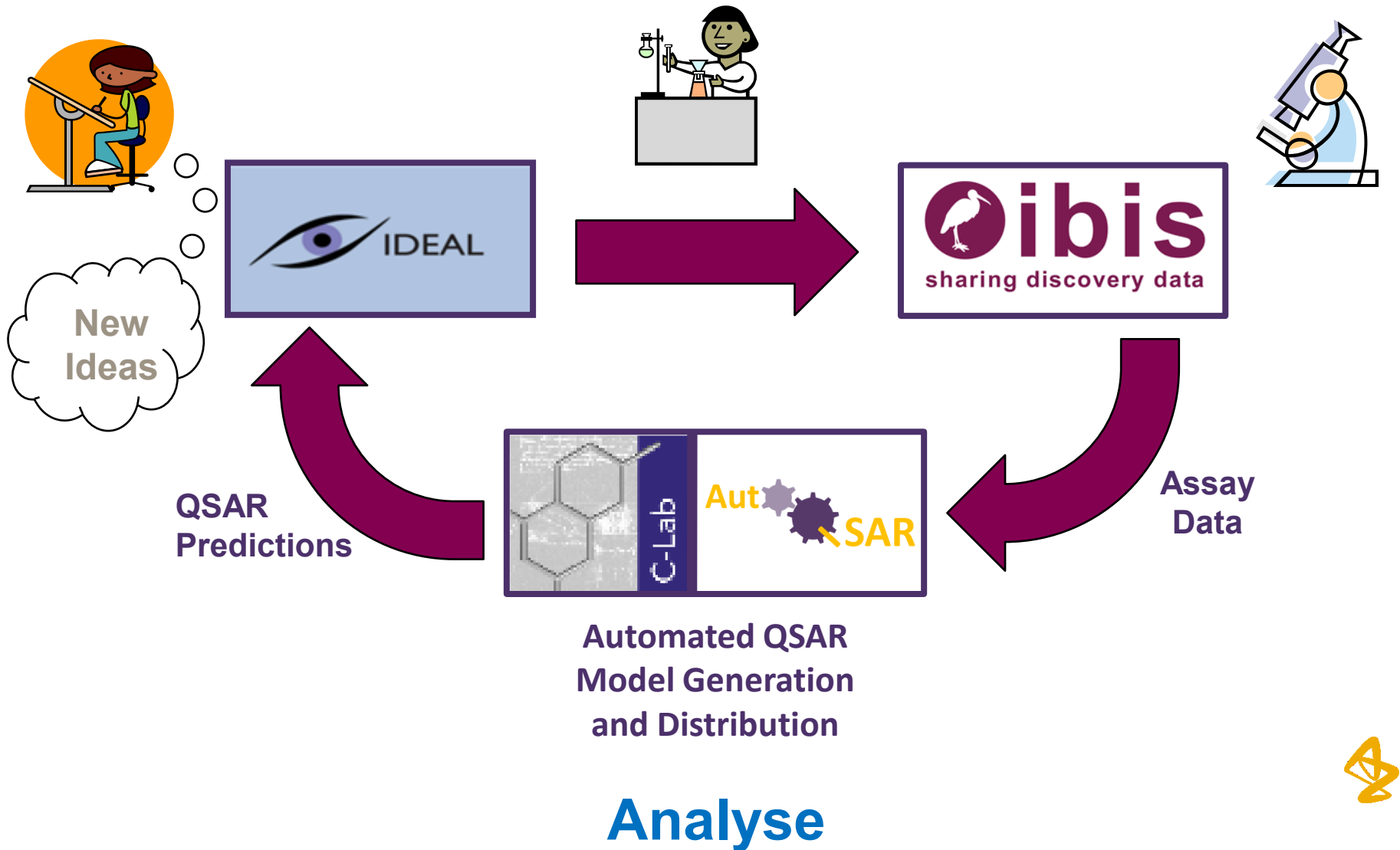


The Drug Design Process at AstraZeneca

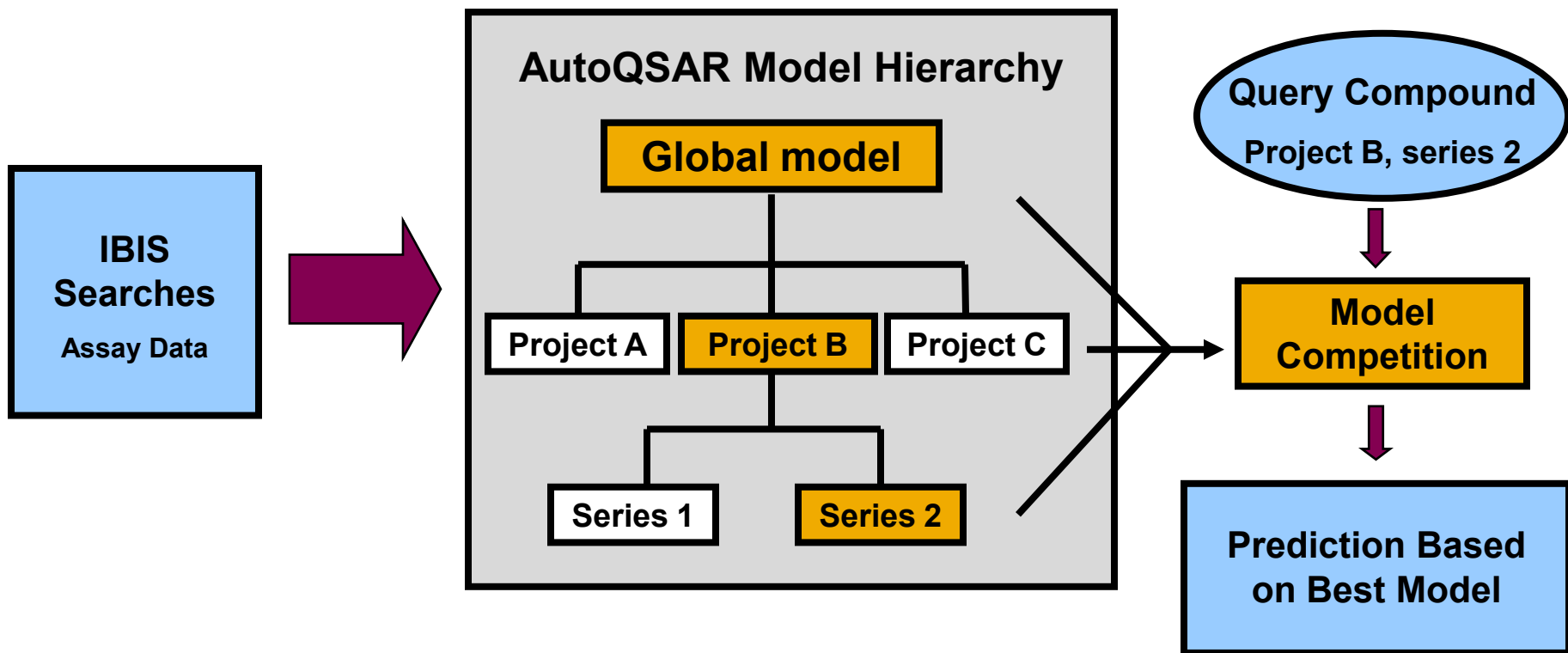
Design

Make

Test



AutoQSAR Under-the-Hood



- Random Forest and Partial Least Squares algorithms assessed for each sub-domain
- The model that is judged to make the best predictions by AutoQSAR is applied to new compounds



AutoQSAR Web Client - Windows Internet Explorer
http://autoqsar.rd.astrazeneca.net:9944/webapps/AutoQSAR/index.html

File Edit View Favorites Tools Help

AutoQSAR Web Client

AutQSAR ExtJS Lightweight Client Logout DW Options ▾

Navigation << 2 Register Project Search [1] ×

- Protocols
 - 1 Register IBIS Search
 - 1 Register Global Search
 - 2 Register Project Search
 - 2 Build Models
 - 1a Build a Global or Local Model for a Global Search
 - 1b Build All Local Models Associated with a Global Search
 - 2a Build a Single Project or Series Model for a Project
 - 2b Build All Models Associated With Project Search
 - 3 Prediction
 - QSAR Prediction
 - 4 Reporting
 - Competitive Analysis Report
 - Time Series Metrics
 - 5 Administration
 - Build Monitor
 - Delete Model
 - Job Monitor
 - Model Mount
 - Model Viewer

Register an IBIS Project Search

IBIS query:

Project (DIPP) Code:

IMED: CNSPain ▾

Sites: AlderleyPark ▾

Descriptors: Standard ▾

Test Set Selection Method: Temporal % ▾

Test Set Selection Value %: 10

Model Name Suffix:

Include OOR data in training set? yes no

Description:

Submit

AutoQSAR

[Link to Help](#)



Validation of AutoQSAR

molecular
informatics

Automated QSAR with a Hierarchy of Global and Local Models

David J. Wood^a *, David Buttar^a, John G. Cumming^a, Andrew M. Davis^b, Ulf Norinder^c and Sarah L. Rodgers^d

A retrospective study aimed at demonstrating the improvements to prediction accuracy that would have been achieved if AstraZeneca had a fully functional AutoQSAR system from 1st January 2009

9 AstraZeneca projects including 35 different chemical series

15 month time frame

3 molecular properties

LogD, Human Plasma Protein Binding, Solubility

3 QSAR strategies compared

1) Static Global

- A single global model built 1st January 2009

2) Global AutoQSAR

- A global model rebuilt with new data at the beginning of each month

3) Hierarchy (Global + Local AutoQSAR)

- Model selection from multiple chemical domains



Validation Study Methodology

Model Selection for the AutoQSAR Hierarchy Strategy

- **Date:** 1st January 2009
- **Endpoint:** LogD
- **Domain:** Project A, Series 1 (A1)
- **Future Compounds:** A1 compounds registered in the following month (Jan 2009)

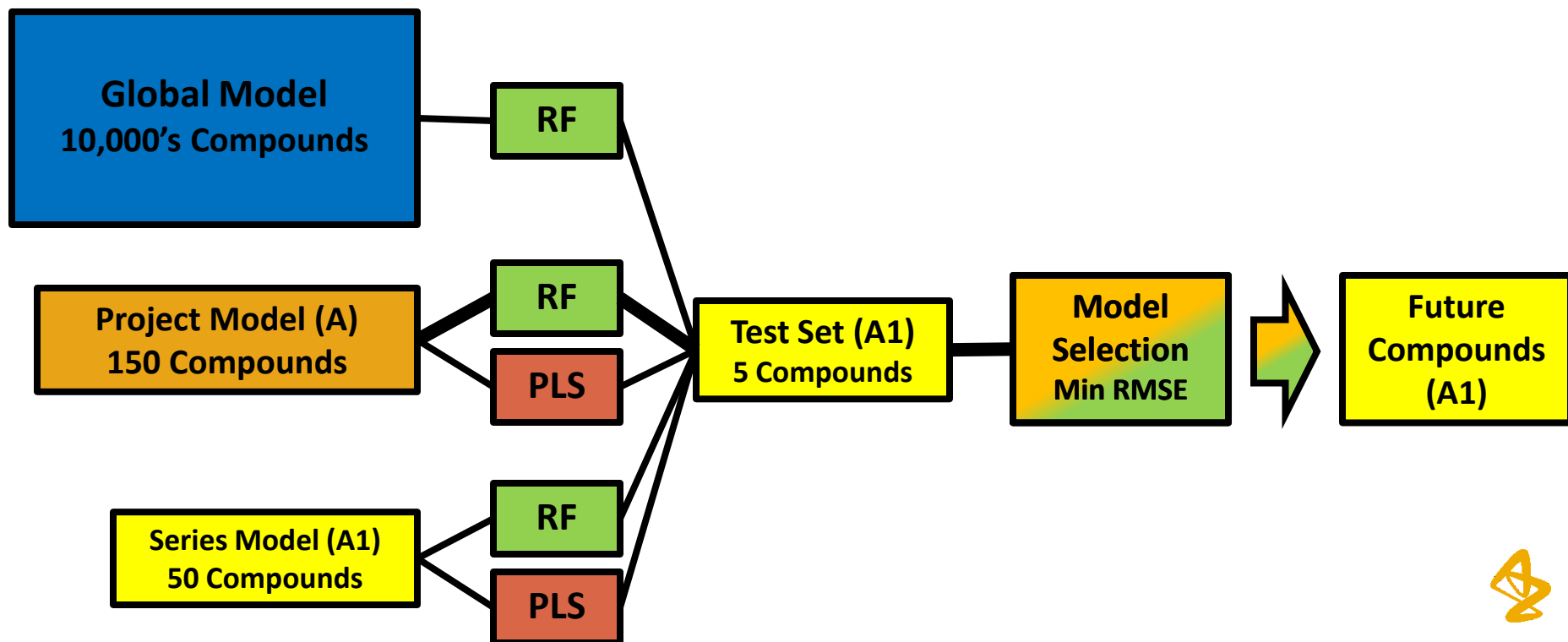
Training Sets

Algorithms

Test Set

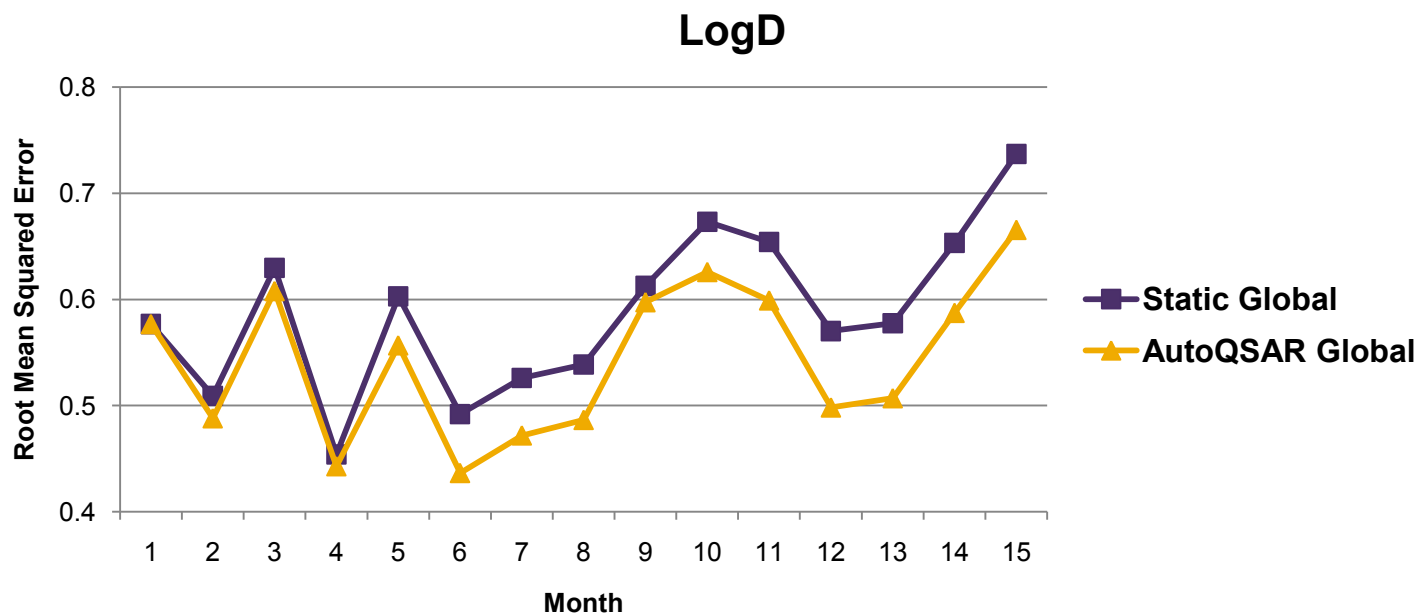
Model Selection

Future Predictions



Validation of AutoQSAR

Global QSAR Models Over Time



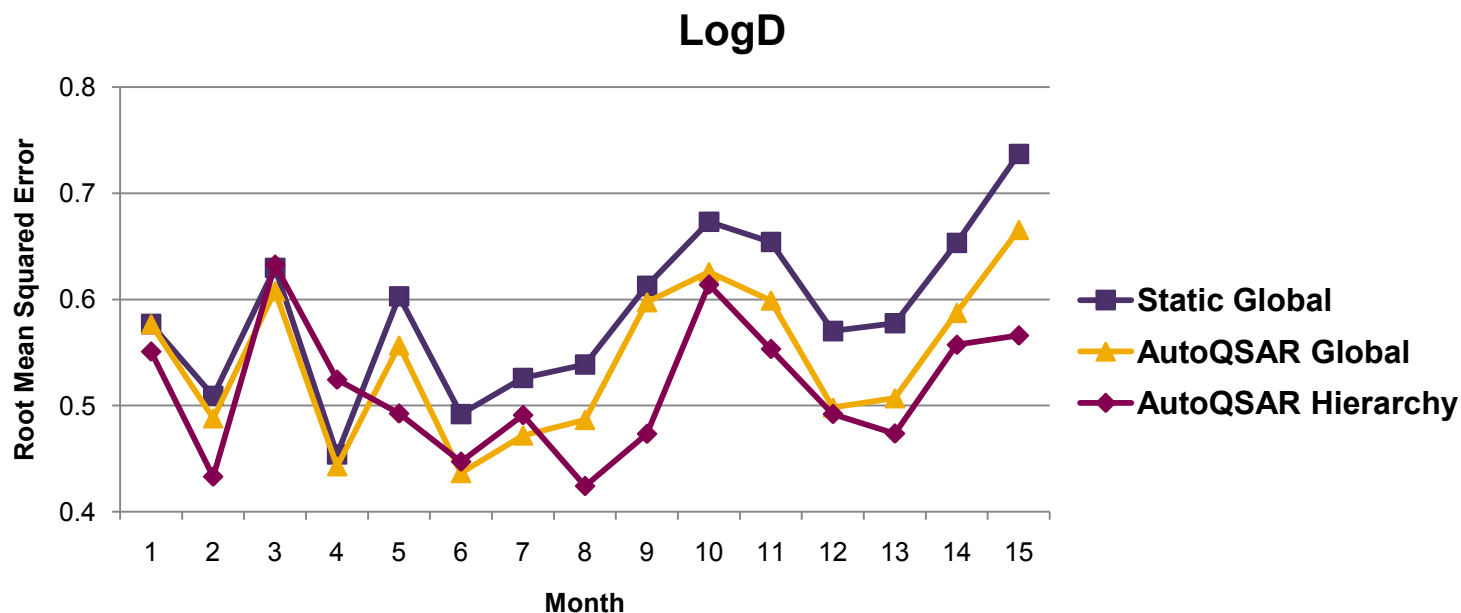
- **Static Global Model:** Statistically significant negative trend
- **AutoQSAR Global Model:** No statistically significant trend over time

Keeping models up-to-date, maintains, rather than improves, accuracy over time



Validation of AutoQSAR

Global QSAR Models Over Time



- **Static Global Model:** Statistically significant negative trend
- **AutoQSAR Global Model:** No statistically significant trend over time

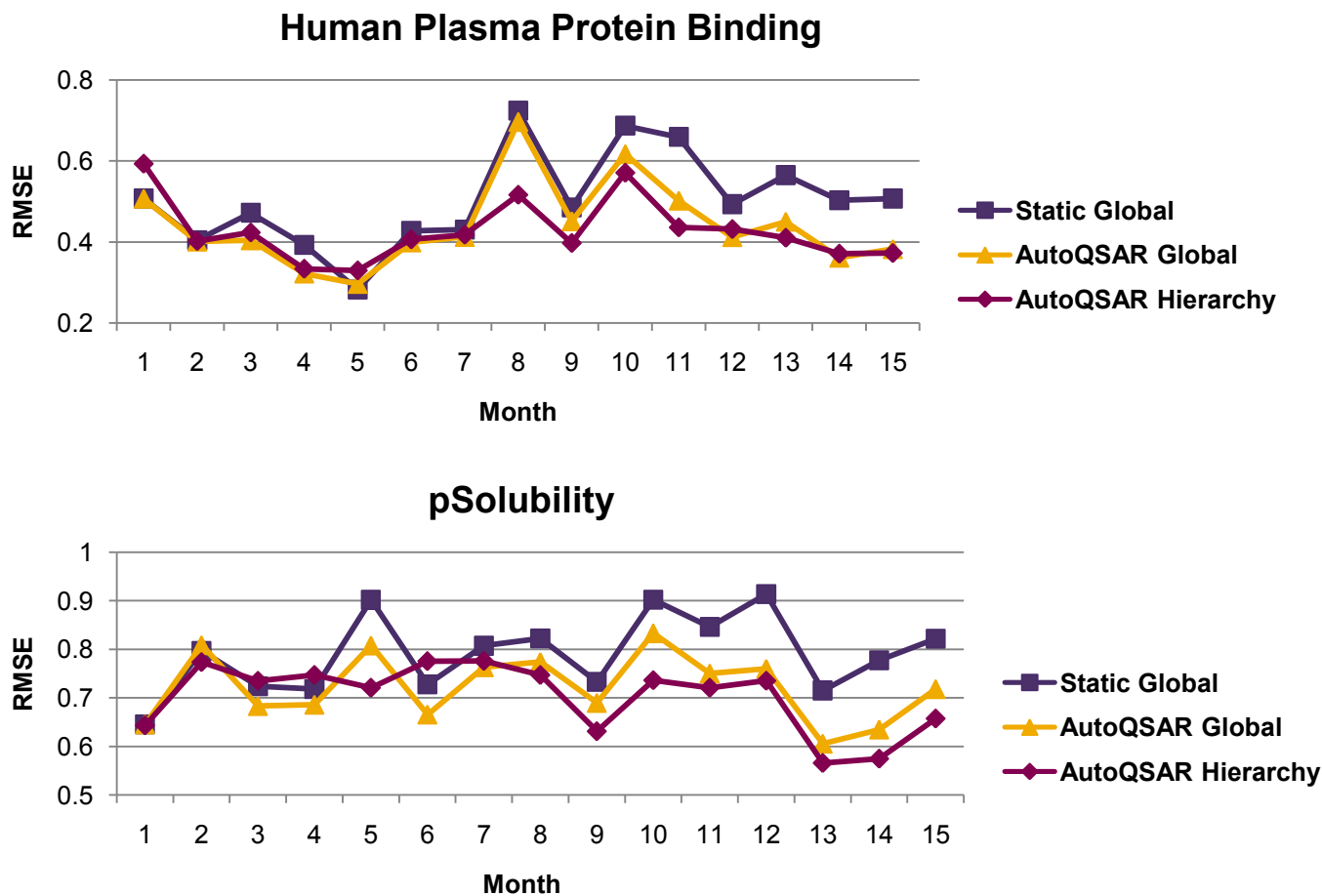
Keeping models up-to-date, maintains, rather than improves, accuracy over time

- **AutoQSAR Hierarchy Model:** No significant trend over time



Validation of AutoQSAR

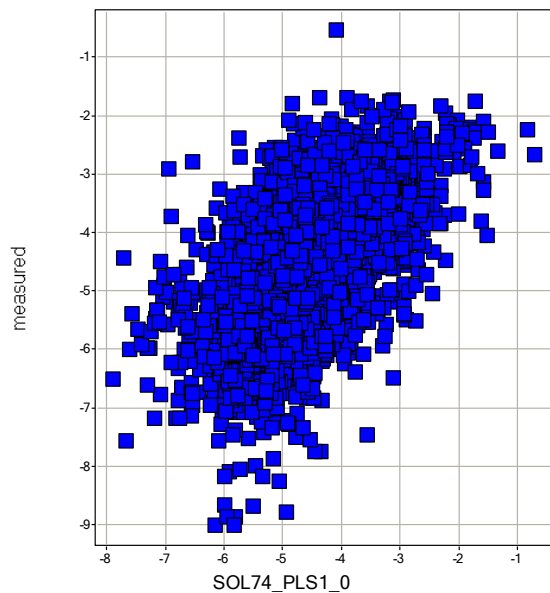
Global QSAR Models Over Time



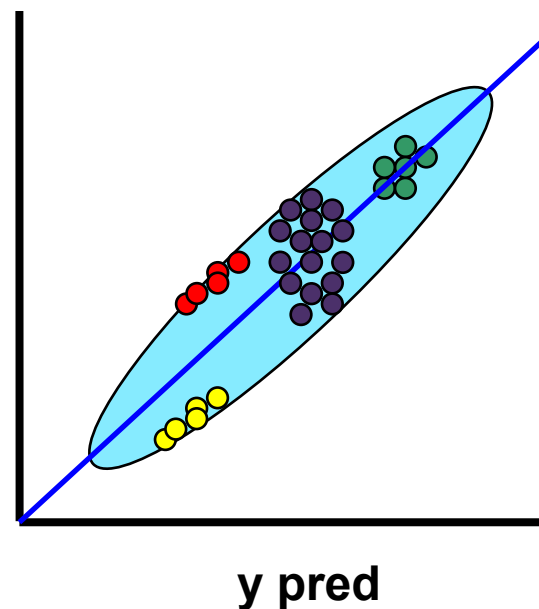
Validation of AutoQSAR

Inside “Averaged” Global Models

pH 7.4 Solubility



yobs

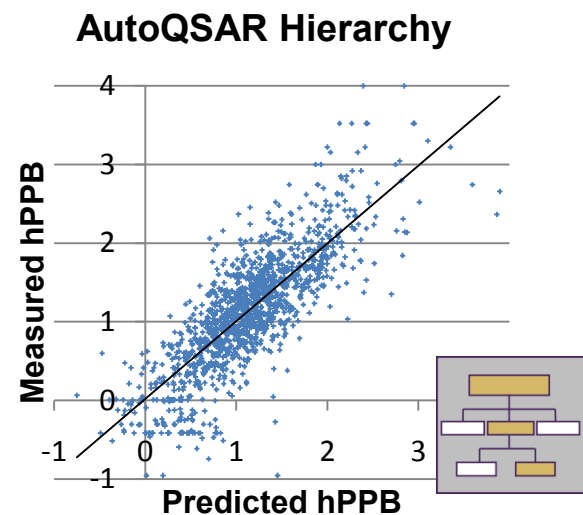
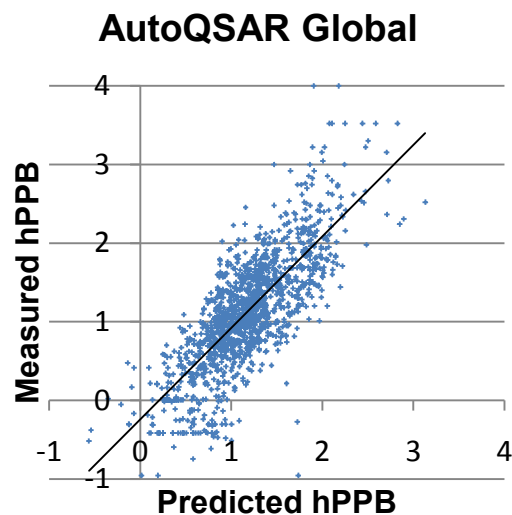
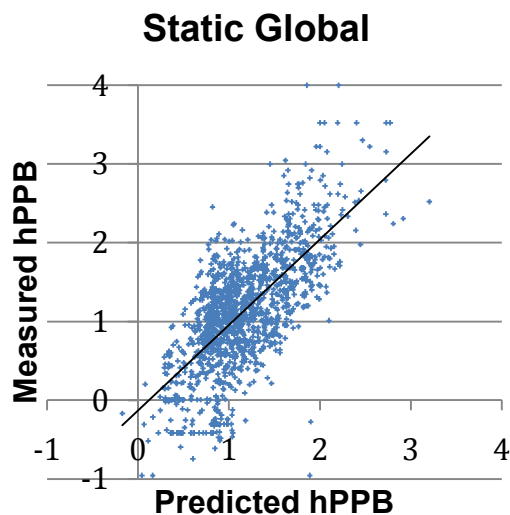


Rmse = 0.92



Validation of AutoQSAR

hPPB: All 'Future' Predictions



Measured LogD Statistics	
N	1763
Mean	2.18
S.D.	1.05

Modelling Strategy	RMSE	R ²	R ₀ ²
Static Global	0.58	0.73	0.69
AutoQSAR Global	0.54	0.76	0.74
AutoQSAR Hierarchy	0.51	0.76	0.76

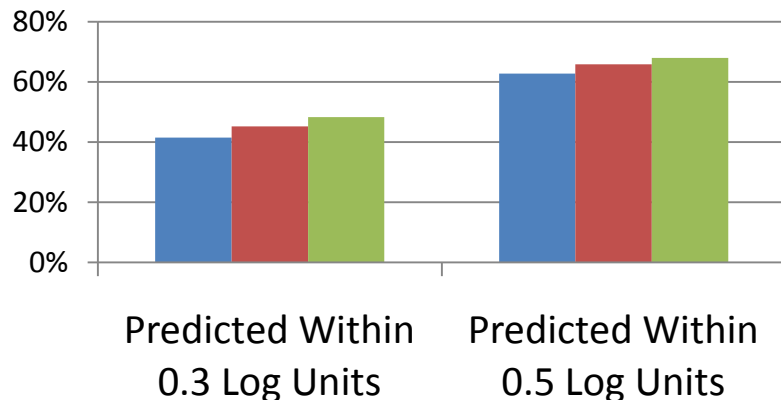
- ❑ Modelling strategy success: AutoQSAR Hierarchy > AutoQSAR Global > Static Global
- ❑ AutoQSAR Hierarchy strategy removes systematic biases



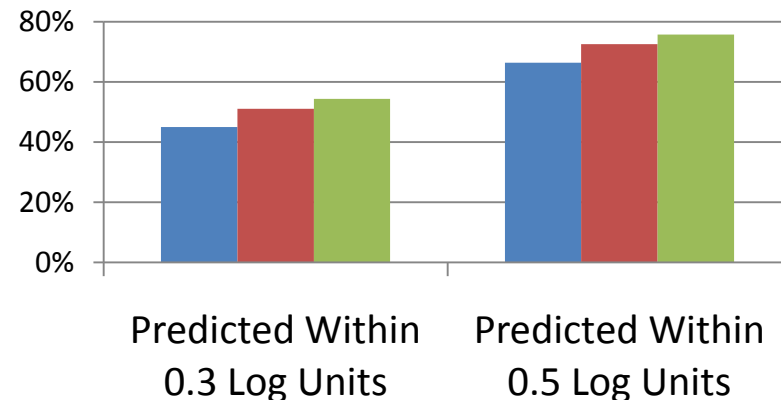
Validation of AutoQSAR

Future Prediction Accuracy

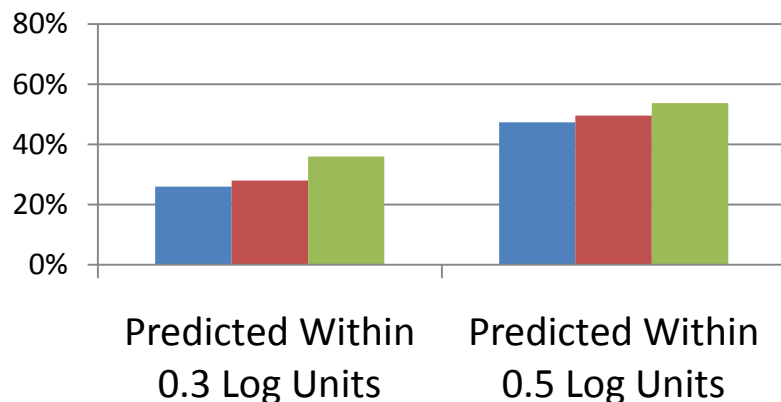
LogD



hPPB



pSol



■ **Static Global**
■ **AutoQSAR Global**
■ **AutoQSAR Hierarchy**

Typical chemical series:
mean values from 35 series



Validation of AutoQSAR

Model Selection for Project B

Test Set Model Selection

Period	Series 1	Series 2	Series 3
1	project-PLS	project-PLS	project-PLS
2	project-PLS	project-PLS	project-PLS
3	project-PLS	project-PLS	project-PLS
4	project-PLS	project-PLS	project-PLS
5	project-PLS	project-PLS	project-PLS
6	project-PLS	project-PLS	project-PLS
7	global-RF	global-RF	global-RF
8	project-RF	project-RF	global-RF
9	project-RF	project-RF	global-RF
10	project-RF	project-RF	global-RF
11	project-RF	project-RF	global-RF
12	project-RF	project-RF	project-PLS
13	project-RF	project-PLS	series-RF
14	project-PLS	project-PLS	project-PLS
15	project-PLS	project-PLS	project-PLS

Perfect Model Selection

Period	Series 1	Series 2	Series 3
1	project-RF		
2	project-PLS		
3	project-RF		
4	project-RF	project-RF	project-RF
5	project-RF		project-PLS
6	global-RF		
7	project-RF	global-RF	global-RF
8	project-RF	project-RF	project-RF
9	project-RF		
10	project-RF		global-RF
11	project-PLS		project-PLS
12		project-PLS	series-RF
13	project-RF		project-PLS
14	project-PLS	project-RF	project-PLS
15	project-RF		project-PLS

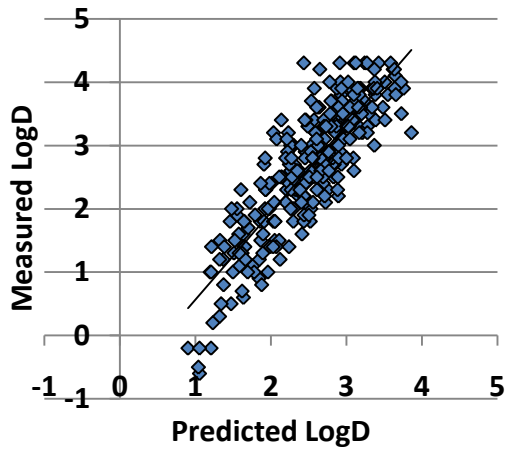
global-RF
project-RF
project-PLS
series-RF
series-PLS

- ❑ Best model selected 38% of the time (13/38)
- ❑ Global model best 11% of the time (4/38)

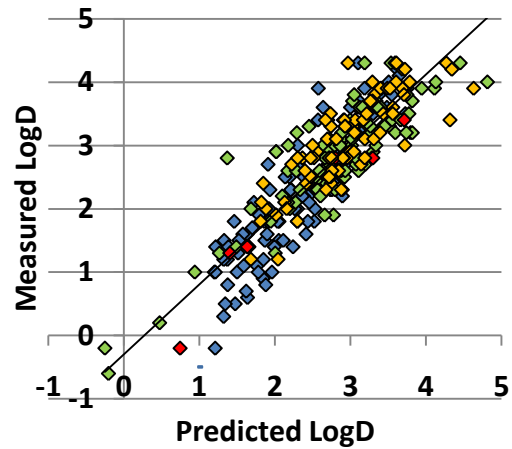


Project B LogD

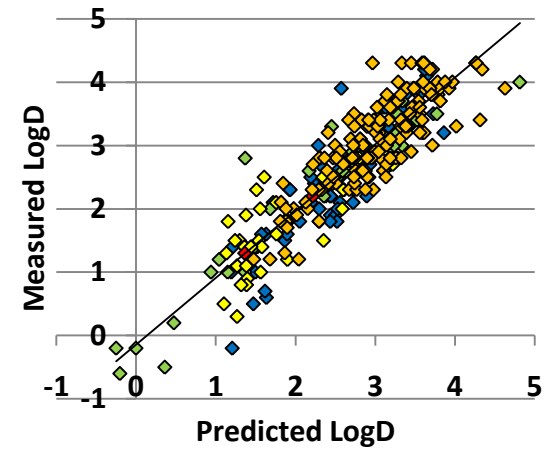
Updating



Hierarchy



Perfect Model Selection



global-RF
project-RF
project-PLS
series-RF
series-PLS

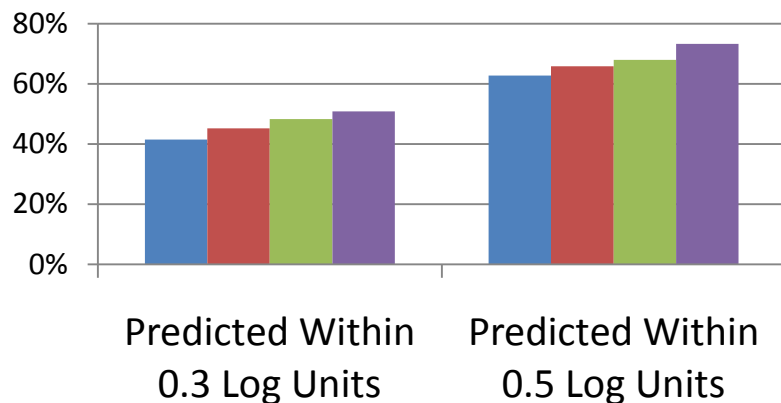
Model	RMSE	R^2	R_0^2
Updating	0.57	0.73	0.66
Hierarchy	0.47	0.78	0.77
Perfect	0.42	0.82	0.82



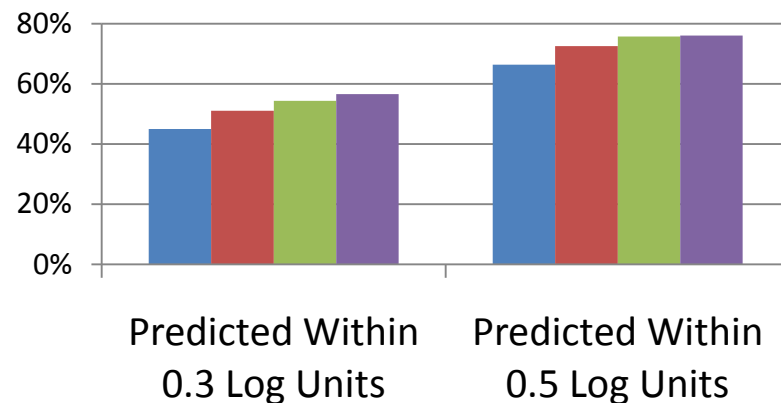
Validation of AutoQSAR

Temporal Test Sets vs. Cross Validation

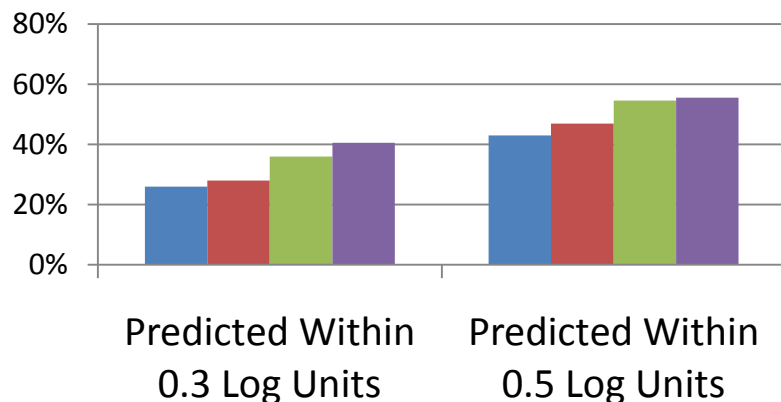
LogD



hPPB



pSol








- Static Global
- AutoQSAR Global
- AutoQSAR Hierarchy TS
- AutoQSAR Hierarchy CV

Typical chemical series:
mean values from 35 series



Questions to Answer

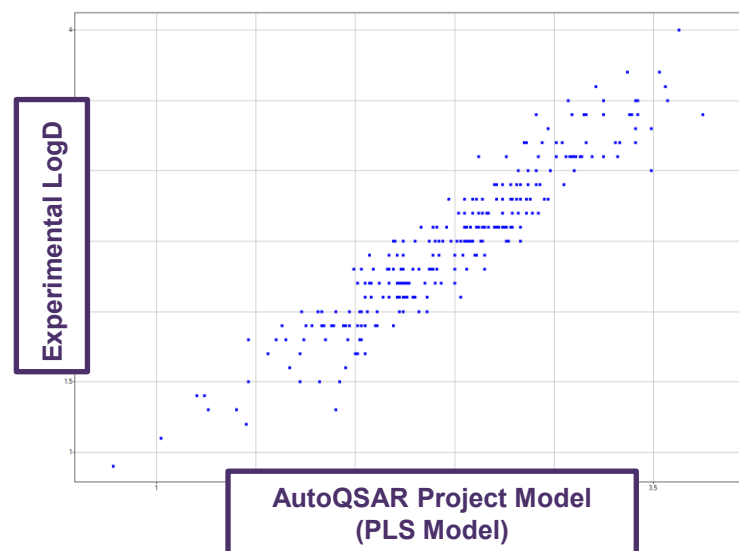
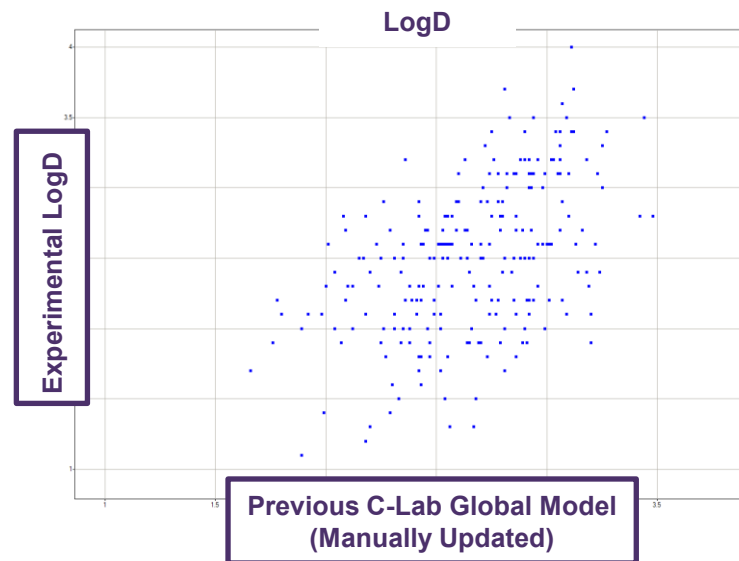
- Could a "system"
 - build models as good as a QSAR expert ? 
 - Be parameterised to work across all projects? 
 - Will my predictions change with time ? 
 - Could we define the project space well enough ? 
 - Would project teams use it ? 



AutoQSAR Impact on AZ Projects

Kinase Cancer Target

- Established project with many lipophilicity measurements
- Model allowed better selection of compounds for synthesis within a narrow lipophilicity range
- Other experimental endpoints (Caco2, hu. plasma protein binding) were also predicted more accurately
- Project team nominated AutoQSAR for an Innovation Award



What Next?

- New Methods and Descriptors
- Fragment-Based Models and Free-Wilson analysis
- Estimation of Prediction Confidence
- Classification Models



Summary

- Drug attrition is related to properties we can control
- Better predictions can improve efficiency and lead to better quality compounds
- AutoQSAR is a system for automated QSAR
 - Improves prediction accuracy by...
 - Keeping models up-to-date
 - Making local modelling easy
 - Automatically selecting the best model for each domain
- Since rollout this summer, it has made an impact on AstraZeneca projects



Acknowledgements

Andy Davis
Sarah Rodgers
John Cumming
David Buttar
Ulf Norinder

- Scott Boyer, Han van de Waterbeemd, Nick Lynch, Iain Comley, Alan Lamont

Previous Postdocs

- Kamaldeep Chohan, Paul Gleeson, Sarah Rodgers, Majeed Sheik, Dave Wood



AutoQSAR

Automation of the QSAR Modelling Process

UK-QSAR – Autumn Meeting, 2011

David Wood, Discovery Sciences

Andy Davis, Respiratory & Inflammation

Sarah Rodgers, Accelrys