Translating Methods from Pharma to Flavours & Fragrances

CINF 27: ACS National Meeting, New Orleans, LA - 18th March 2018

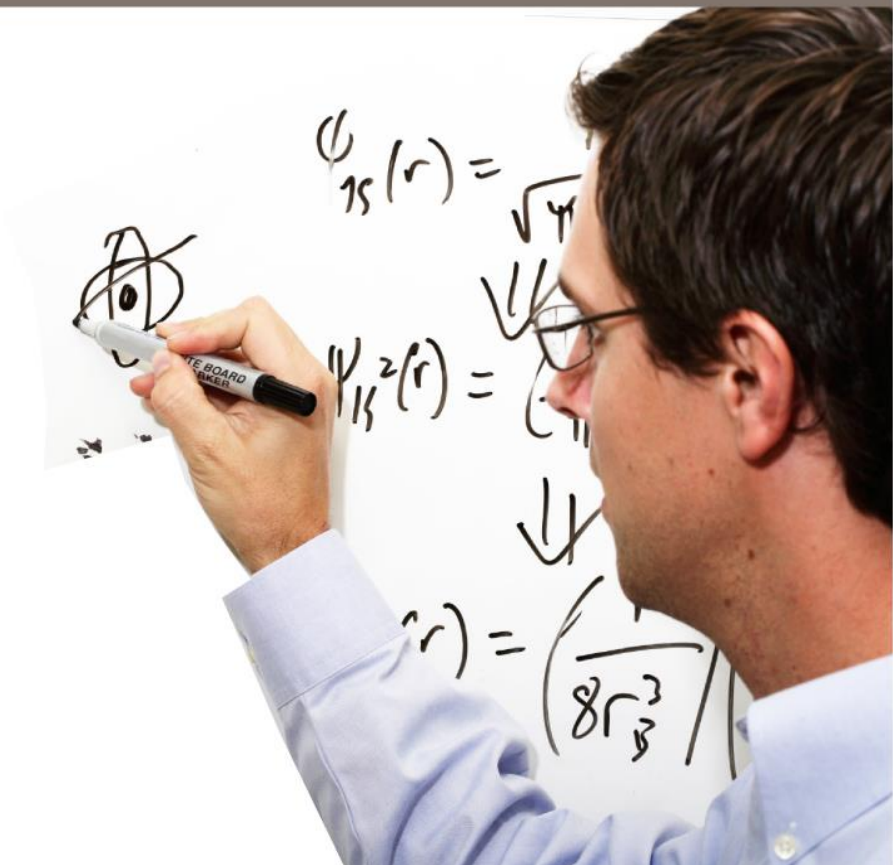
Peter Hunt, Edmund Champness, Nicholas Foster, Tamsin Mansley & Matthew Segall

Overview

- Datasets for pharma, flavours and fragrances
- Chemical Space
  - Similarity & diversity
- Multi-parameter optimisation
  - Using predictive models
- QSAR model building with Auto-Modeller™
  - Case Studies
- Conclusions

Introducing the Data Sets

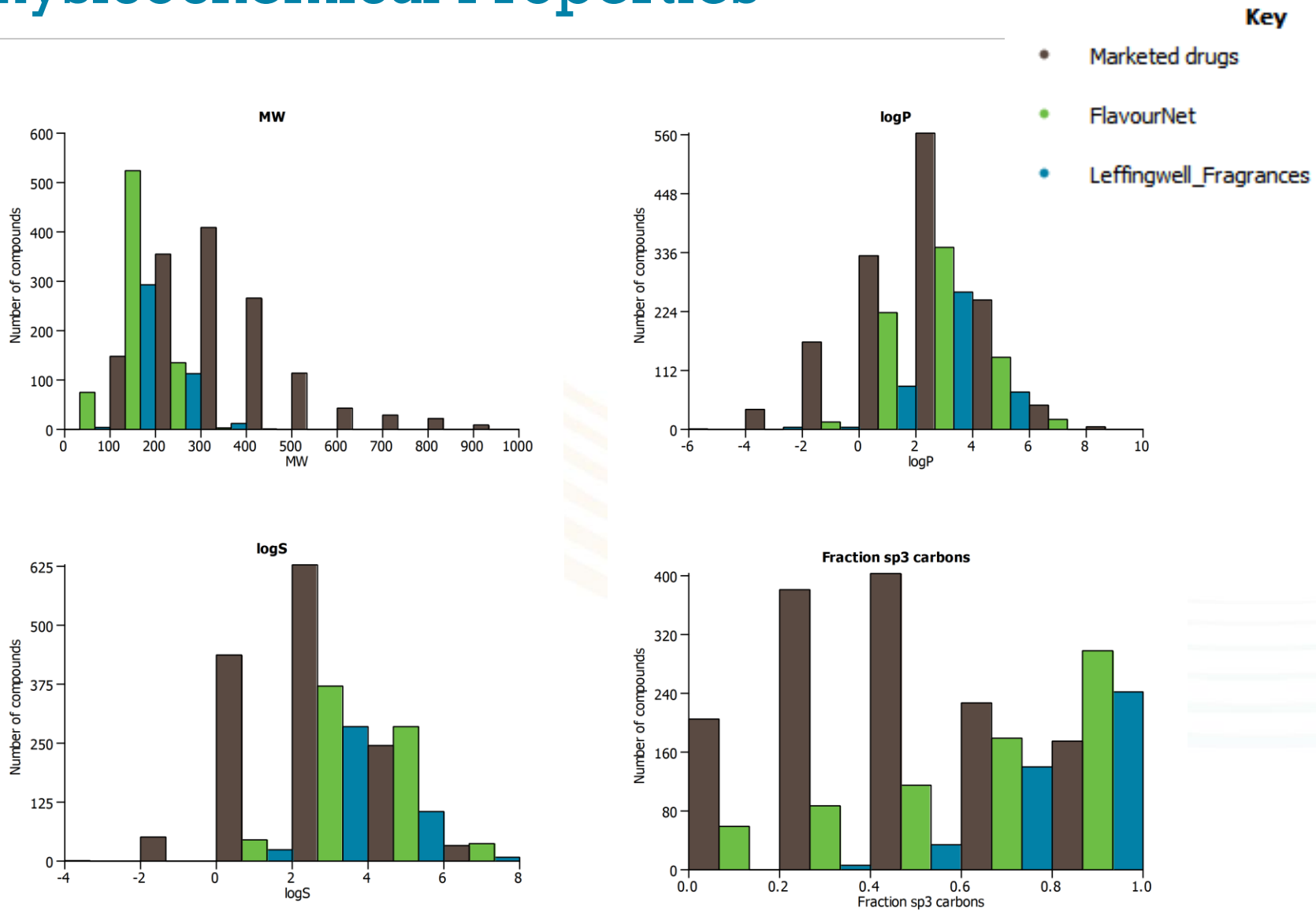
optibrium



Data Sets for Pharma, Flavours & Fragrances

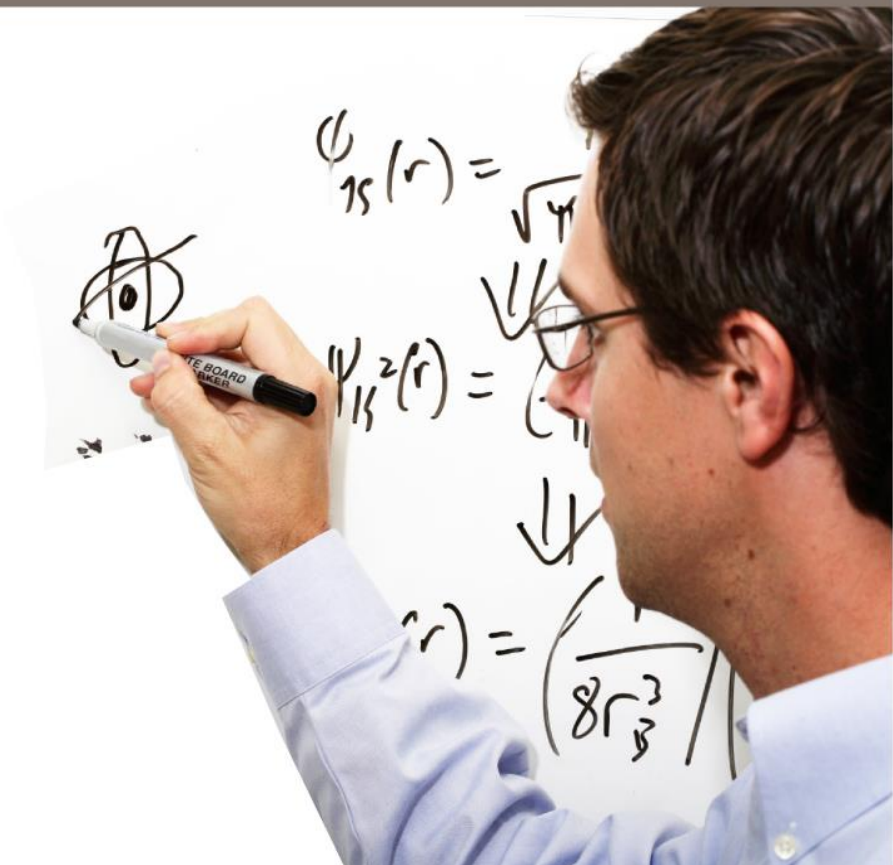
- Marketed Drugs
  - 1396 diverse marketed small-molecule drug compounds
  - Internal Optibrium dataset
- FlavourNet Database
  - 738 flavours compounds
  - http://www.flavornet.org/flavornet.html
- Leffingwell Odour Data Set
  - 422 fragrance compounds
  - http://www.leffingwell.com/chirality/chirality.htm

Physicochemical Properties

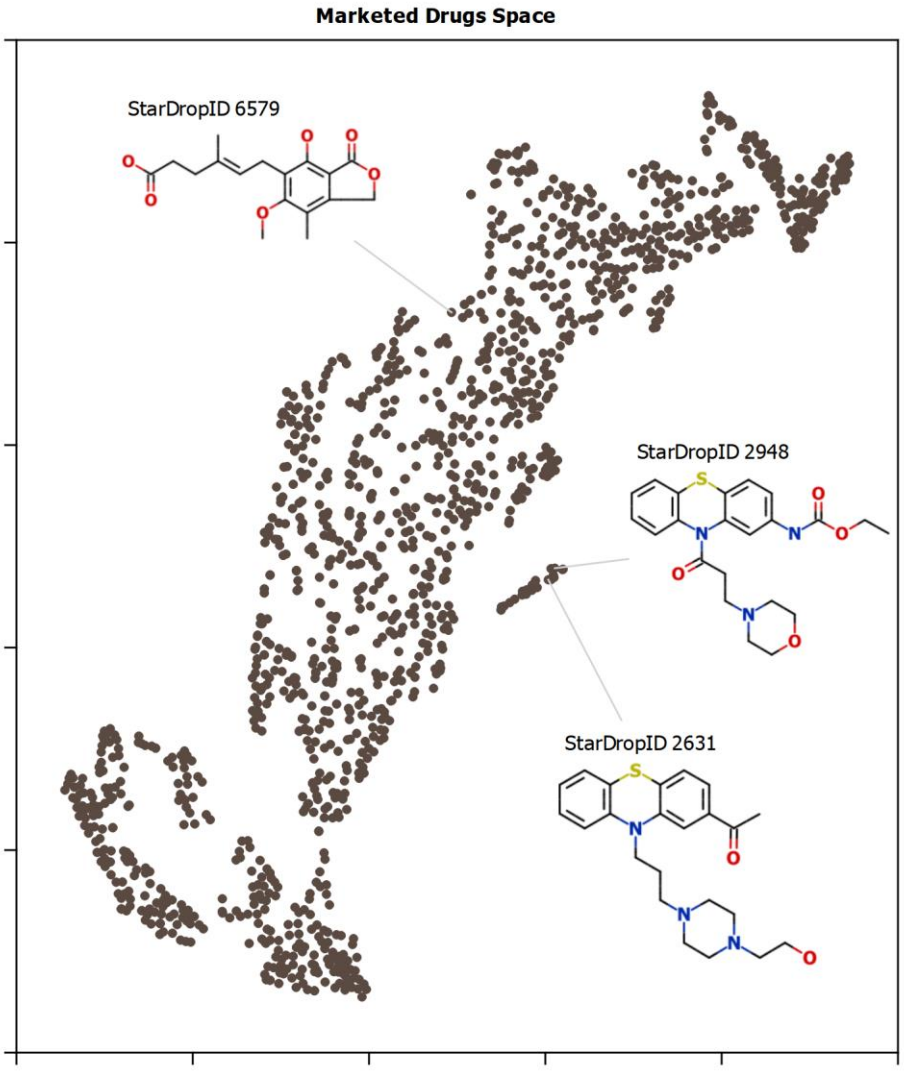


Chemical Space
Dataset Similarity & Diversity

optibrium

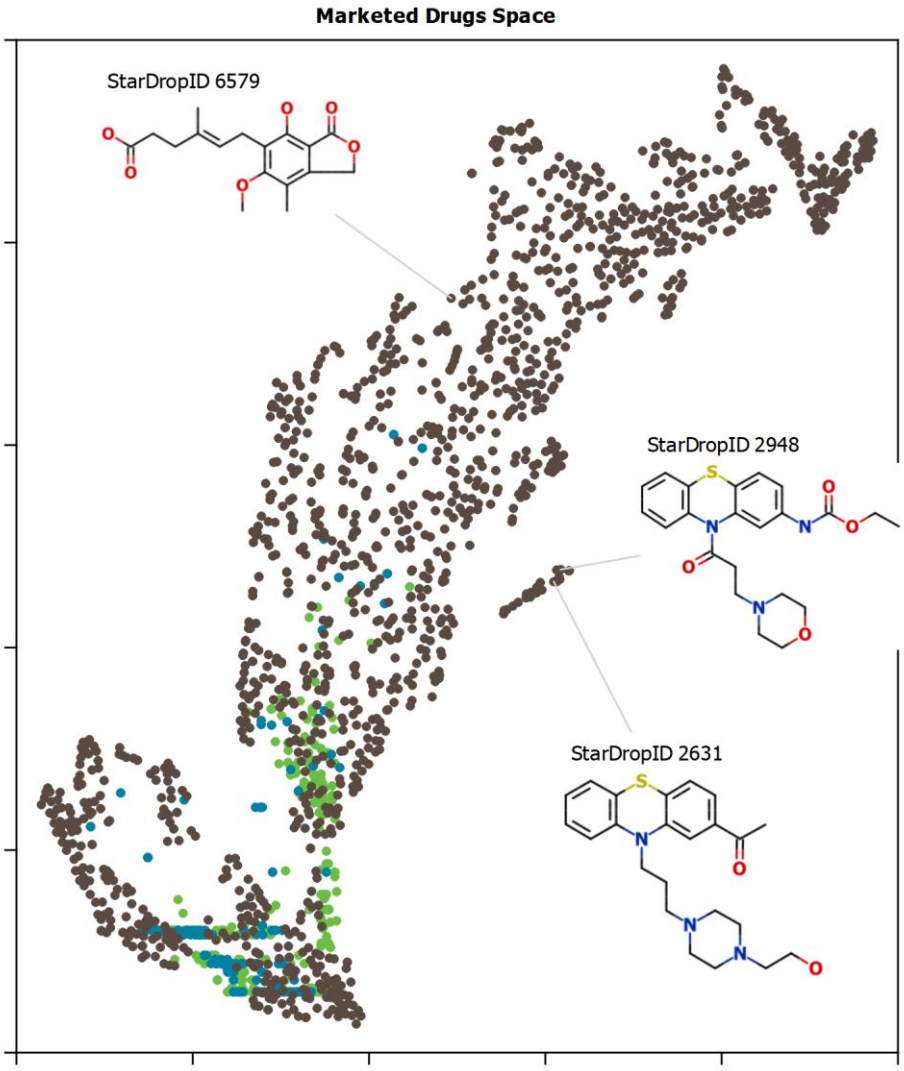


Chemical Space



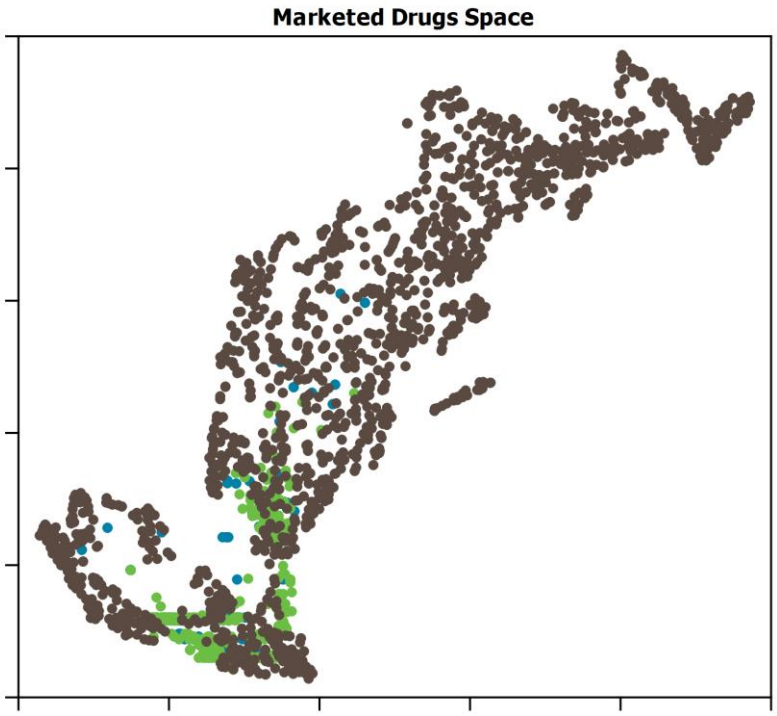
- A chemical space allows you to visualise trends across your data set
- Each point represents one compound
- The closer two points are the greater their similarity
  - Structure
  - Properties
- A space is defined by a single data set…
- …but other data sets can be plotted in that space at the same time

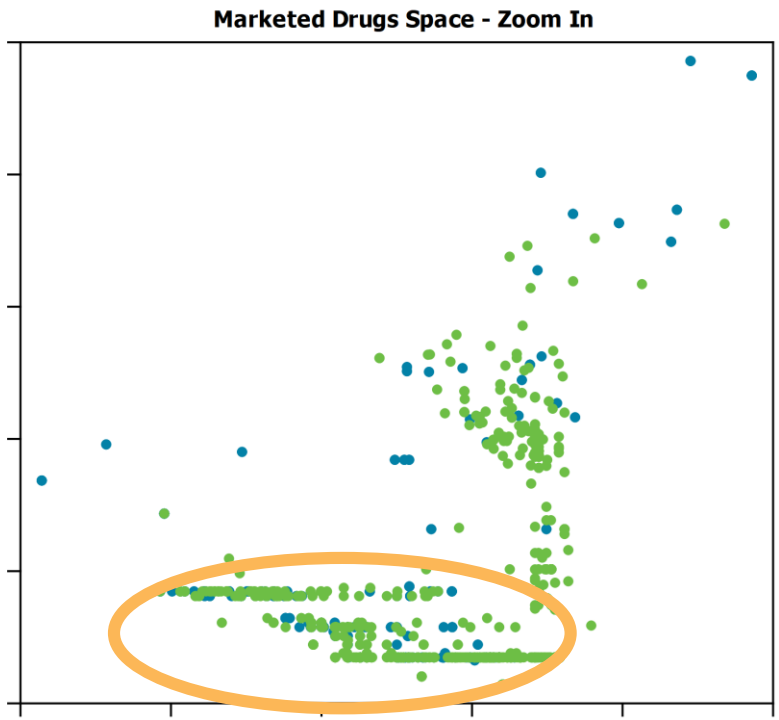
Chemical Space



- A chemical space allows you to visualise trends across your data set
- Each point represents one compound
- The closer two points are the greater their similarity
  - Structure
  - Properties
- A space is defined by a single data set…
- …but other data sets can be plotted in that space at the same time

Marketed Drug Space

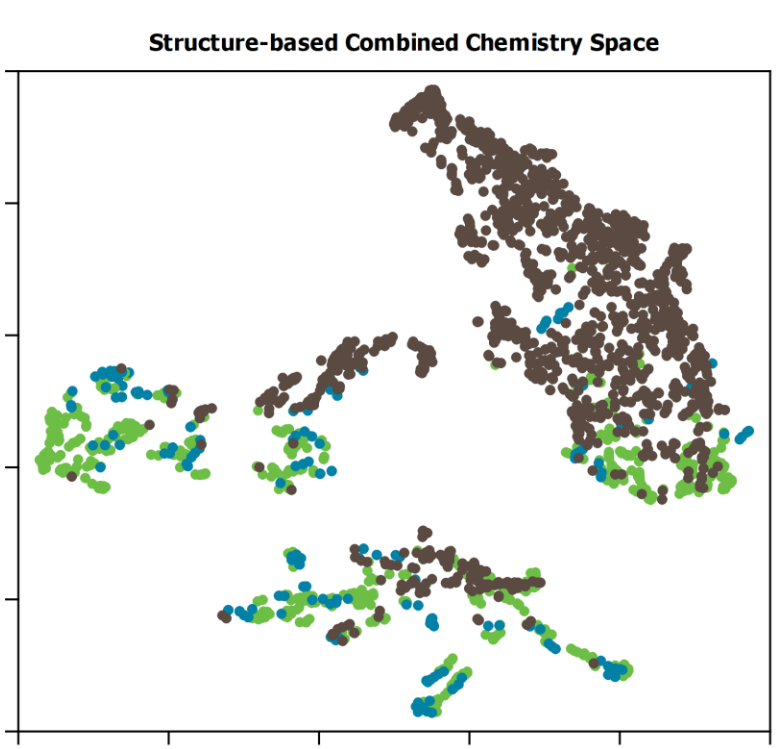


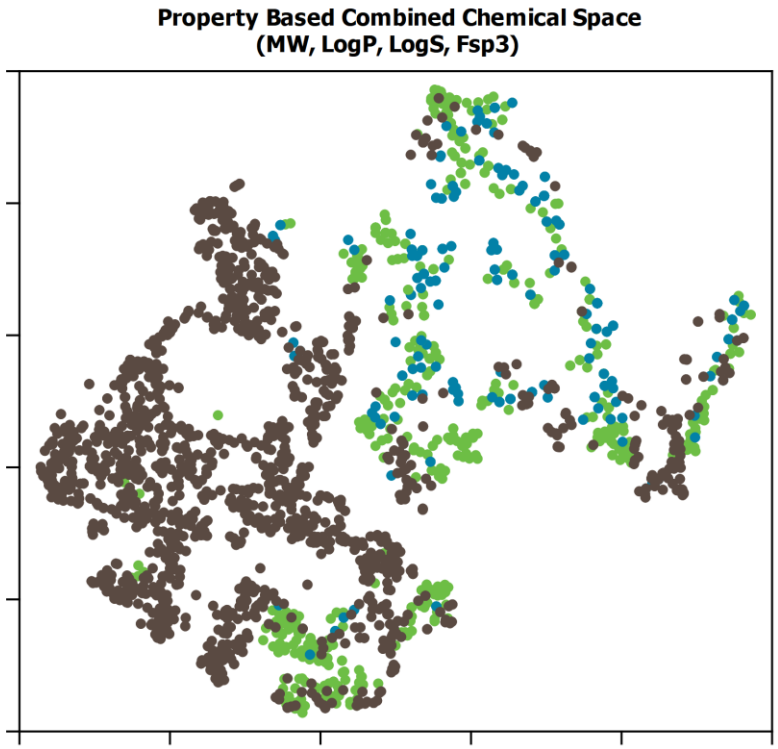


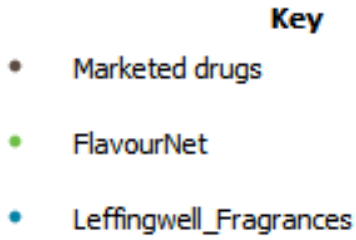
These compounds are not well defined by the Chemical Space

Combined Chemical Spaces

- New areas explored by flavours & fragrances

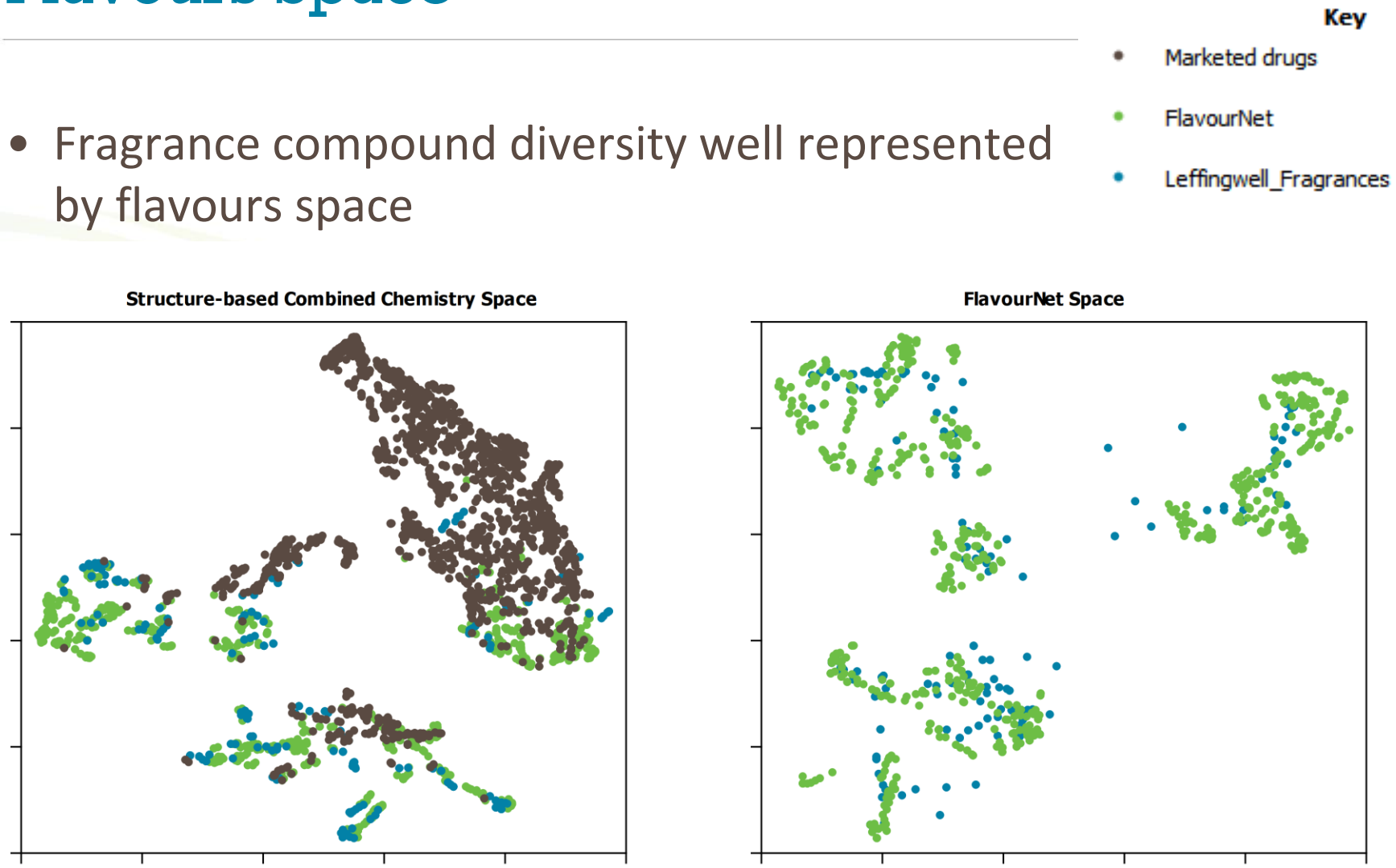






Flavours Space

- Fragrance compound diversity well represented by flavours space



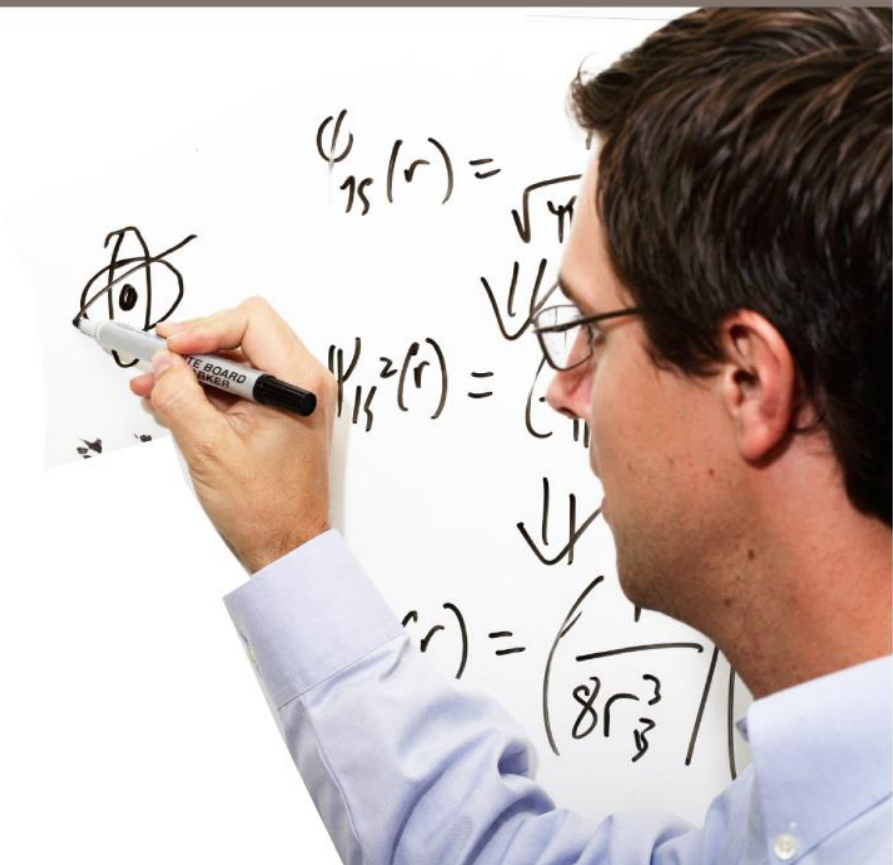
Conclusions – Chemical Spaces

- Chemical Spaces for pharma cannot be applied directly to flavours & fragrances
  - Molecules are typically smaller with a higher Fsp^3
- Approaches
  - Build 'Global' Chemical Space in which flavour and fragrance compounds are well represented
  - Build specific Chemical Spaces for flavours and fragrances datasets
- May be possible to share Chemical Spaces between flavours and fragrances with more success

Multi-parameter Optimisation

Prioritising Compounds with a Balance of Properties, Predicting Properties

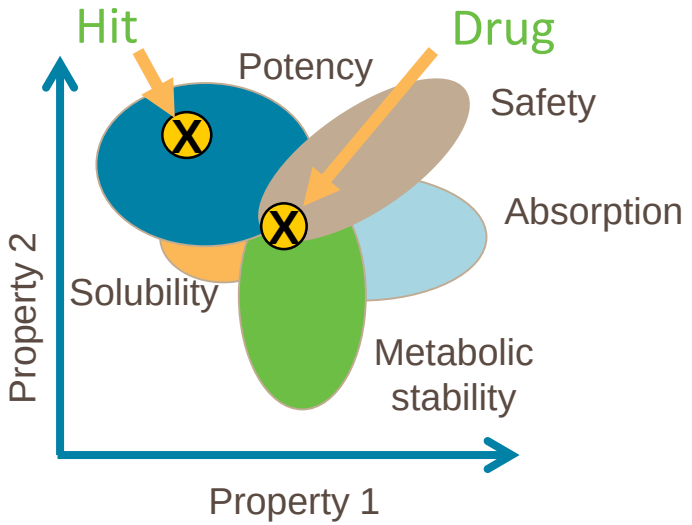
optibrium

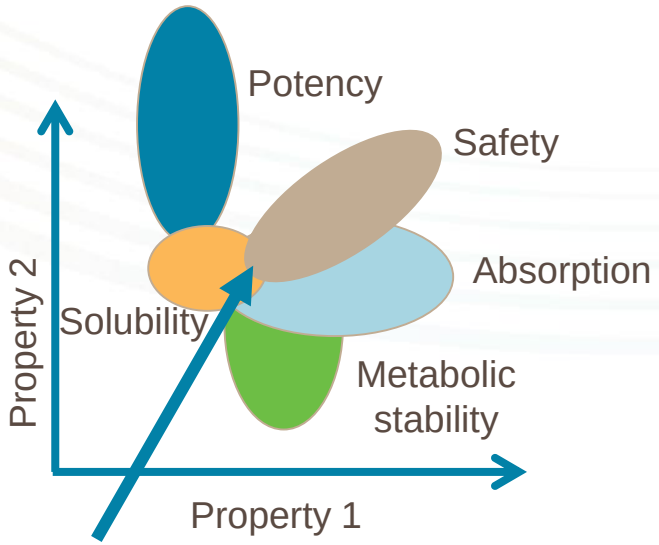


Multi-parameter Optimisation
Drug Discovery

- Identify chemistries with an optimal **balance** of properties

- Quickly identify situations when such a balance is not possible
  - –Fail fast, fail cheap
  - –Only when **confident**
  - –Avoid **missed opportunities**



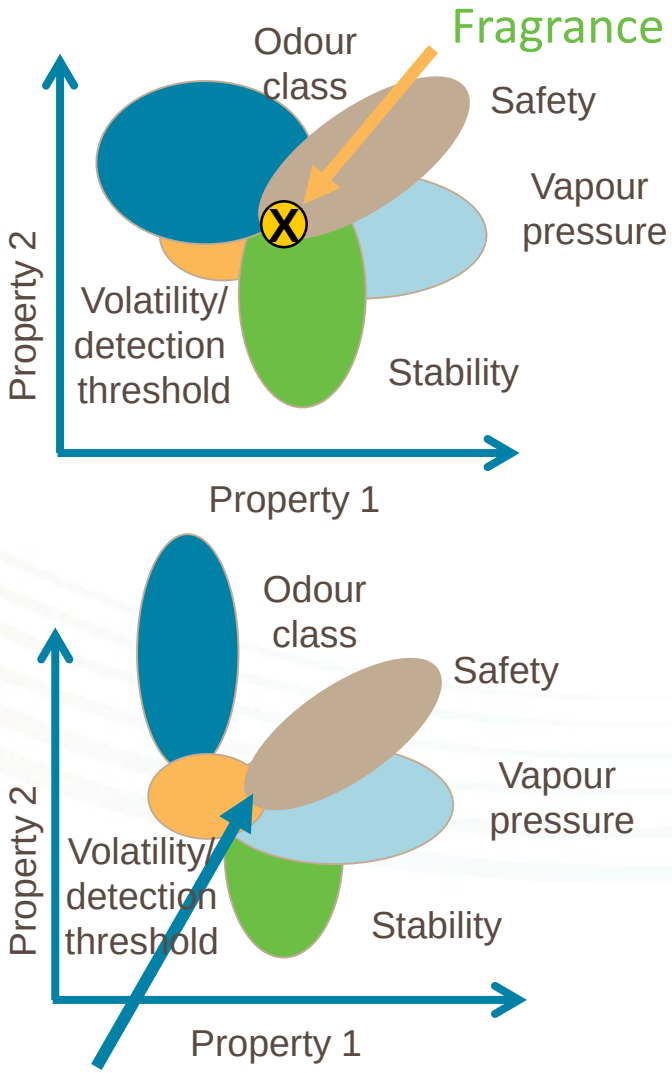


No good drug

Multi-parameter Optimisation
Flavours & Fragrances

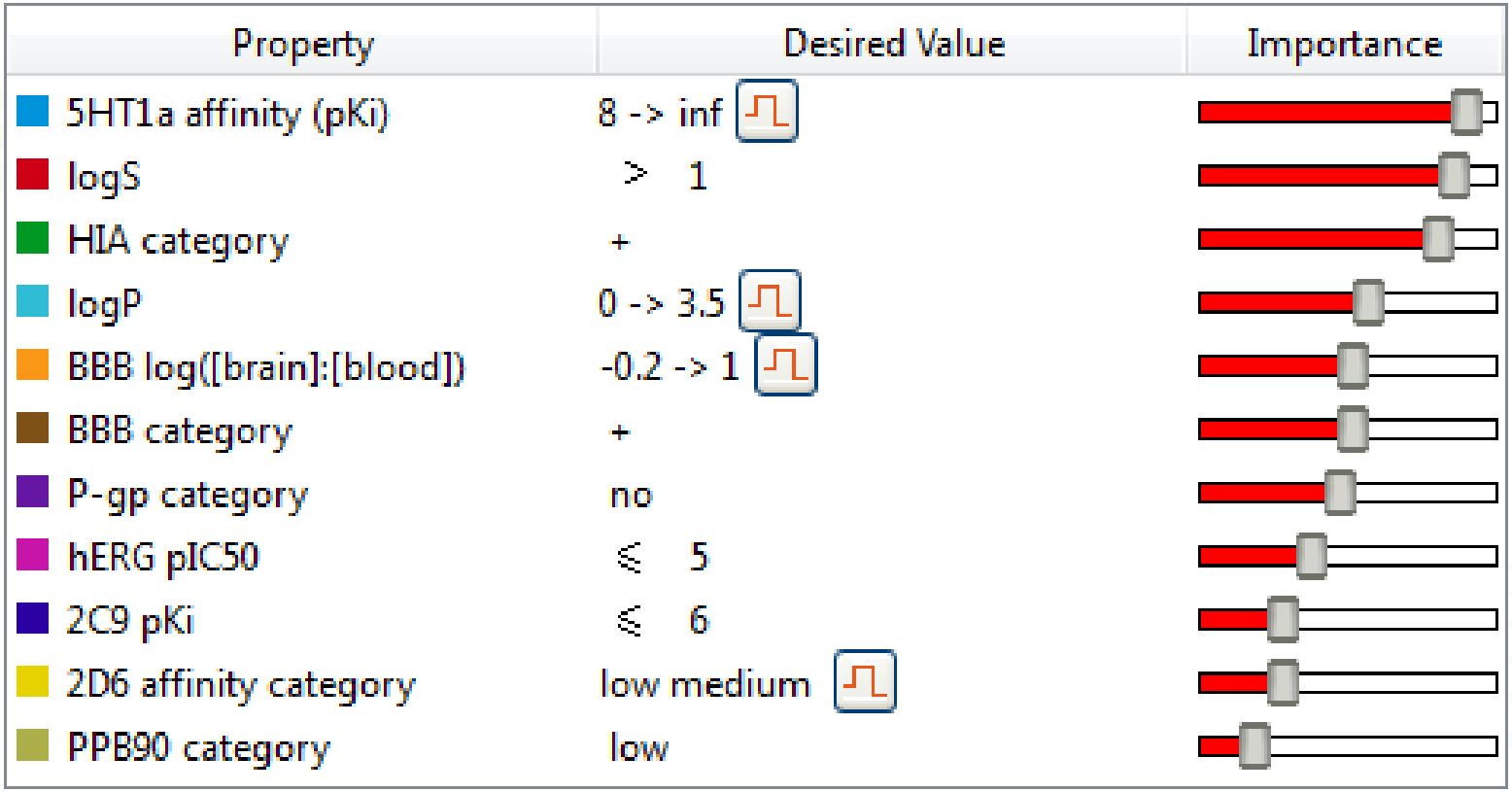
- Identify chemistries with an optimal **balance** of properties

- Quickly identify situations when such a balance is not possible
  - –Fail fast, fail cheap
  - –Only when **confident**
  - –Avoid **missed opportunities**



 M.D. Segall Curr. Pharm. Des. 18(9) pp. 1292-1310 (2012)

Probabilistic Scoring
Scoring Profile



Segall *et al.* (2009) Chem. & Biodiv. **6** p. 2144

Target Product Profiles

- Pharma
  - Potency
  - Selectivity
  - Physicochemical properties
    - LogP, LogS, MW
  - Off-target effects
    - hERG affinity, CYP inhibition
  - Distribution
    - Human intestinal absorption, Plasma-protein binding, BBB penetration

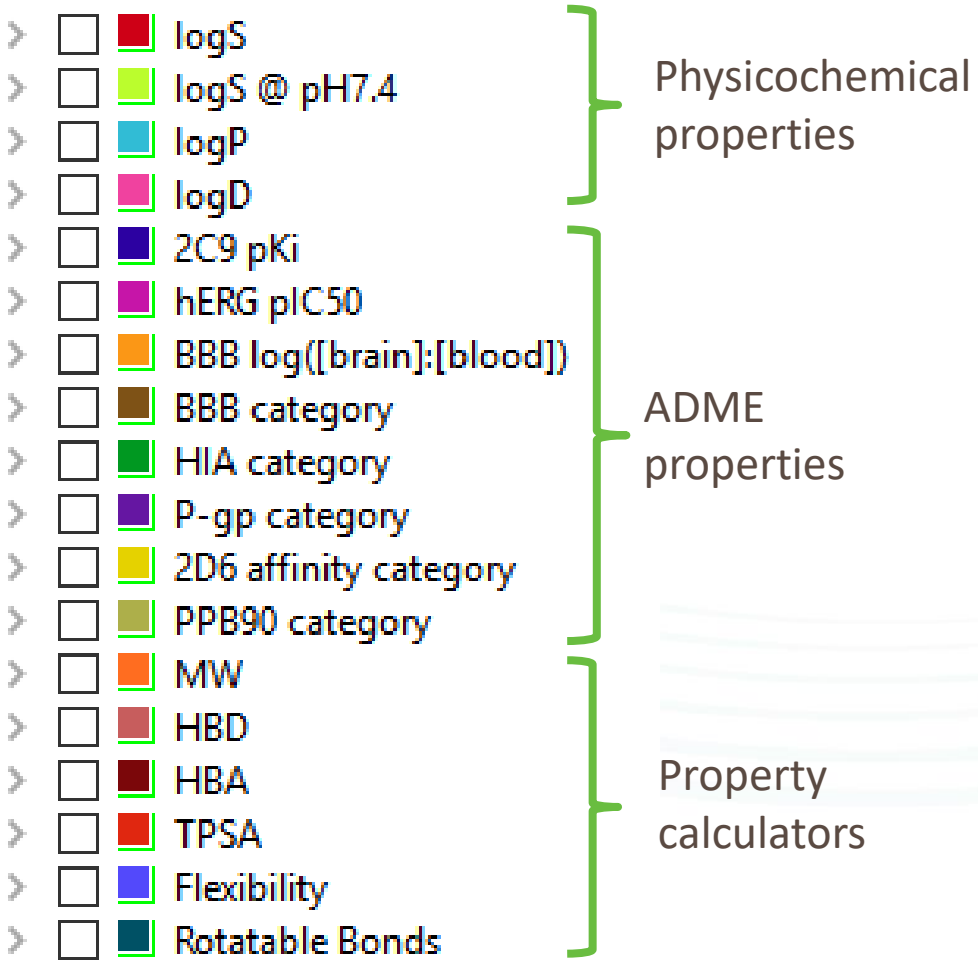
- Flavours
  - Potency
    - Flavour
    - Taste class (sweet, sour, bitter, umami, salt)
    - Taste threshold
  - Physicochemical properties
    - LogP, LogS, MW

- Fragrances
  - Potency
    - Odour class
    - Detection level
  - Physicochemical properties
    - MW, Vapour pressure

StarDrop™ Predictive Models

- High quality predictive models for key ADME properties
- Global models rigorously validated against independent test sets
  - Defined ‘domains of applicability’

Can we apply these models for flavours & fragrances datasets?



Assessing Predictive Ability
Defining the Domain of Applicability

- The diversity of the training set defines the **chemical space** of the model
- The position of a new compound relative to chemical space is reflected in the reported confidence in the prediction

LogP (octanol/water)

- Predicts the logarithm of the octanol/water partition coefficient for neutral compounds
- Dataset of 9000 experimental octanol/water partition coefficient values obtained from the Medchem* database
- Model statistics on Test Set
  - N = 2950, R2 = 0.92, $RMSE_{IN}$ = 0.44 log units, $RMSE_{OUT}$ = 0.63 log units

Dataset	% Predictions Within Chemical Space of Model	% Predictions Outside Chemical Space of Model
StarDrop Marketed Drugs	91.8	8.2
FlavourNet	98.8	1.2
Leffingwell Fragrances	100	0

*Leo, A., Chem. Reviews, 93(4), 1281-1306 (1993)

Intrinsic Aqueous Solubility (logS)

- Predicts the logarithm of the intrinsic aqueous solubility, S in µM, for neutral compounds
- Dataset of >3,300 aqueous solubility data points for intrinsic water solubility, S in µM, defined as the thermodynamic solubility of uncharged compound in water between 20-30°C. The data comes from the Syracuse* database
- Model statistics on Test Set
  - N = 663, R2 = 0.82, $RMSE_{IN}$ = 0.70 log units, $RMSE_{OUT}$ = 1.03 log units

Dataset	% Predictions Within Chemical Space of Model	% Predictions Outside Chemical Space of Model
StarDrop Marketed Drugs	34.4	65.6
FlavourNet	87.7	12.3
Leffingwell Fragrances	86.7	13.3

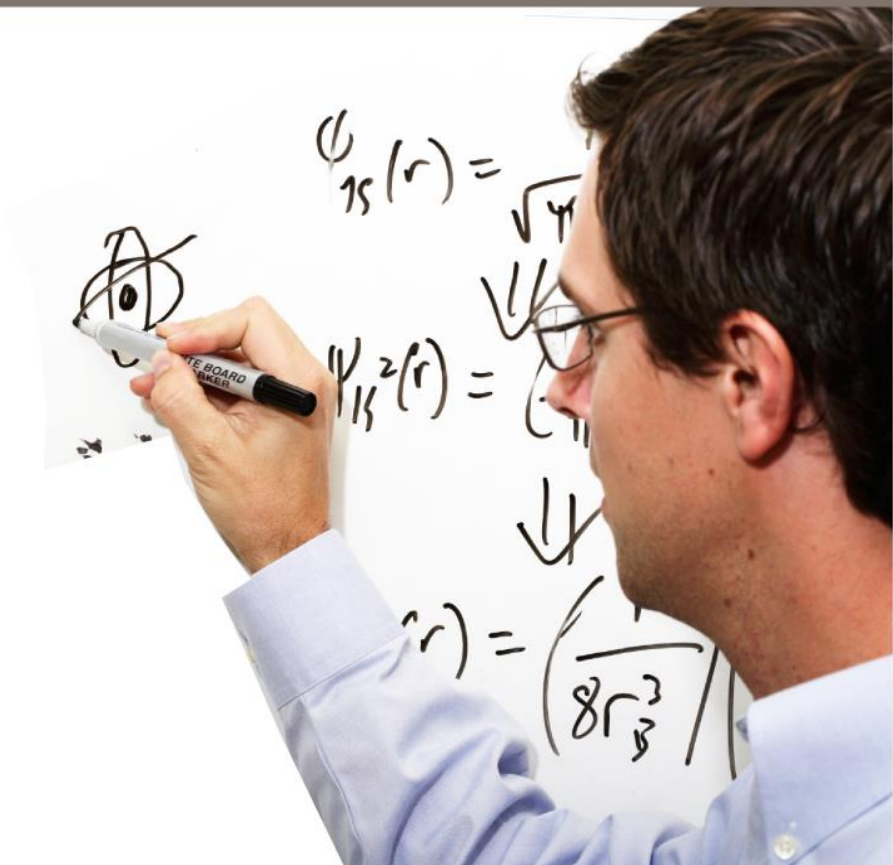
*Butina, D., & Gola, J. M., J. Chem. Inf. Comput. Sci., 43(3), 837-841.(2003,05)

Conclusions – Predictive Models

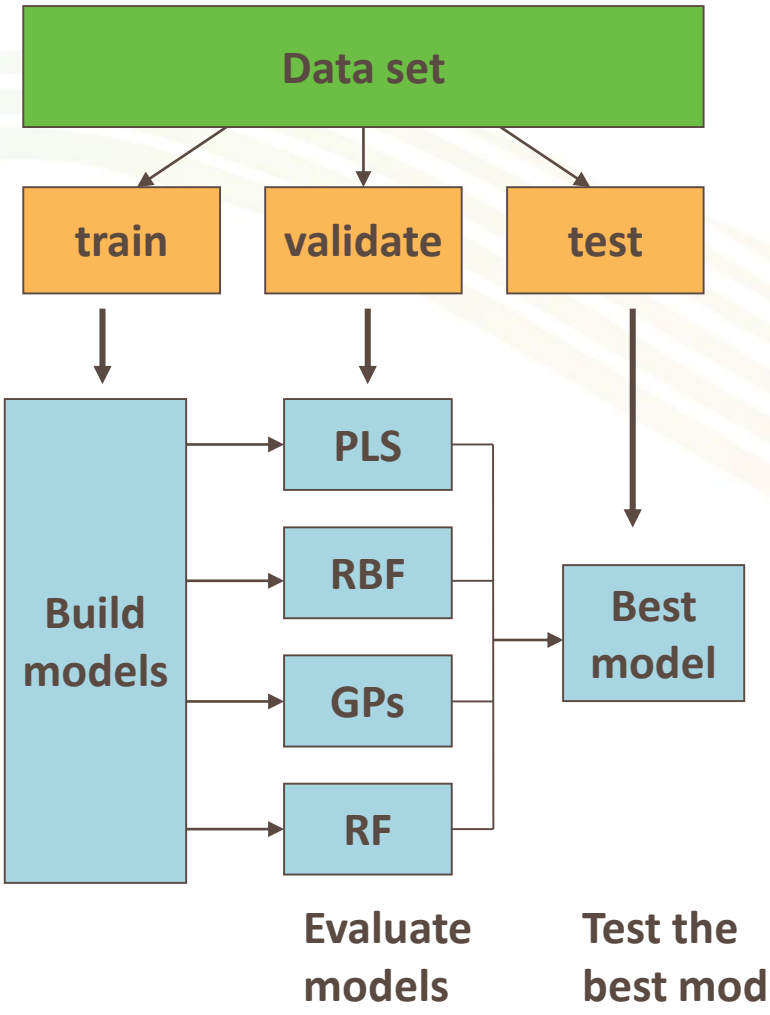
- Some properties are important across pharma, flavours and fragrances
- Where pharma models exist it may be possible to apply these to flavours and fragrances
  - Important to consider Chemical Space of the model (training set)
  - Assess uncertainty in predictions
- Where models are not predictive, or no model exists, we can consider building tailored QSAR models

Building QSAR Models
Auto-Modeller in StarDrop

optibrium



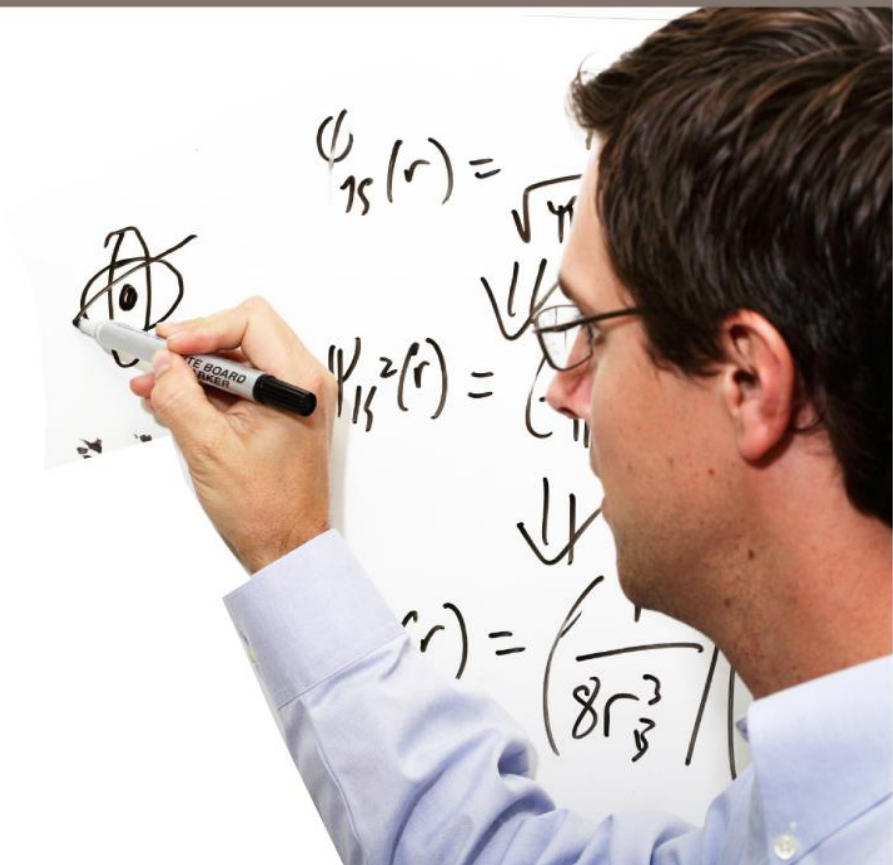
Automatic Model Generation



- Split data set
- Calculate descriptors (2D SMARTS, logP, TPSA, MW, charge etc.)
- Multiple modelling techniques
- Select the best model by performance on the validation set
- Test with an **independent** set

Case Study
FlavourNet Kovats Indices Models

optibrium

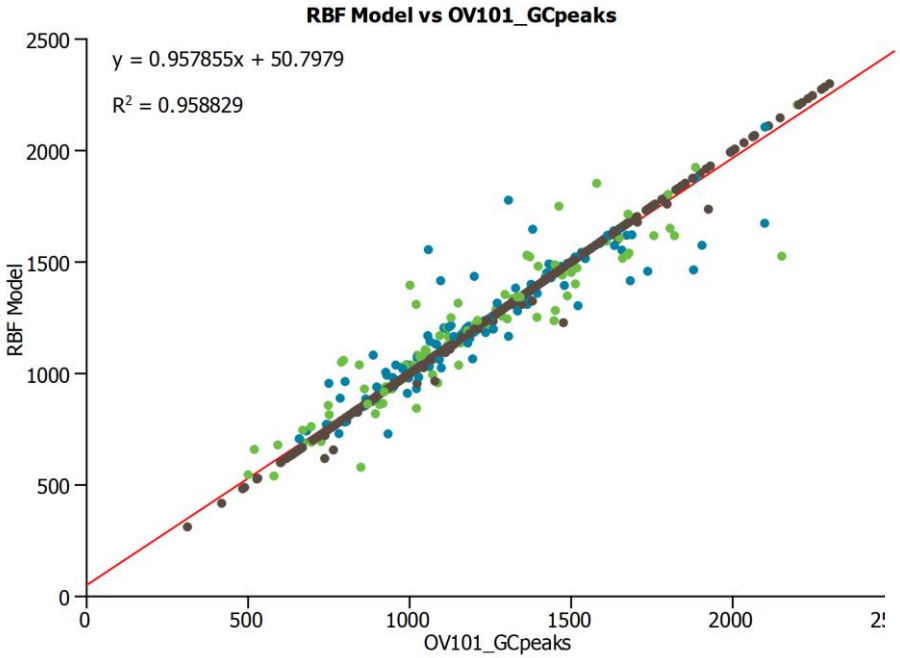


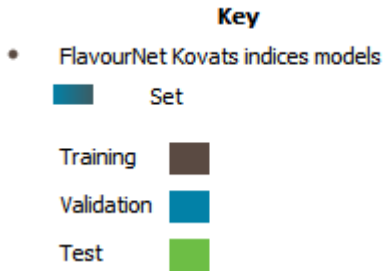
Kovats Indices

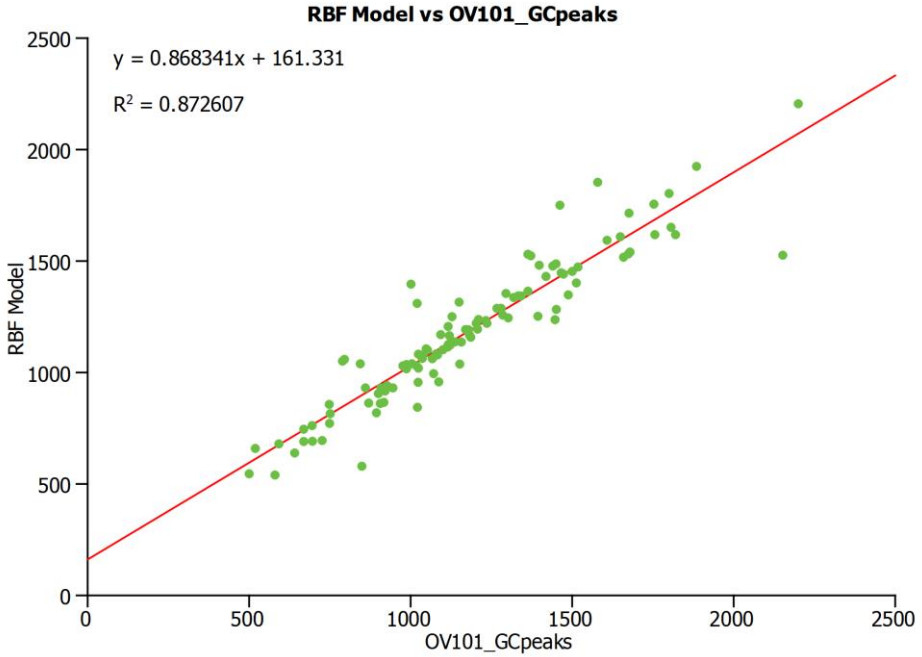
- FlavourNet
  - 738 fragrance compounds
  - Odour type is important for flavours
  - Kovats Indices (Gas Chromatography peaks) used in compound identification
  - Compounds with similar volatility may have similar odour profiles

	Training Set		Validation Set		Test Set	
Model	**Rsqr**	**RMSE**	**Rsqr**	**RMSE**	**Rsqr**	**RMSE**
PLS Model	0.6839	194.3	0.719	164	0.7667	163.9
RBF Model	0.9976	17.08	0.8351	125.7	0.8723	121.3
Random Forest Regression Model	0.9472	79.45	0.827	128.7	0.839	136.1
GPFixed	0.8778	120.8	0.8291	127.9	0.8794	117.8
GP2DSearch	0.8787	120.4	0.8294	127.8	0.8789	118

Kovats Indices

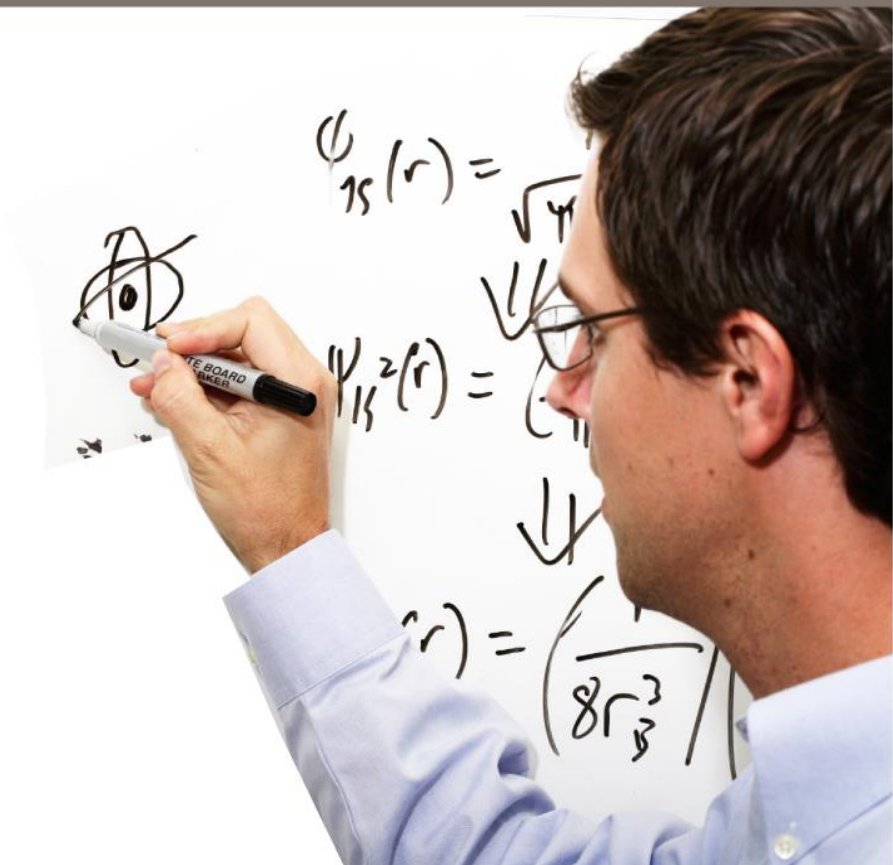






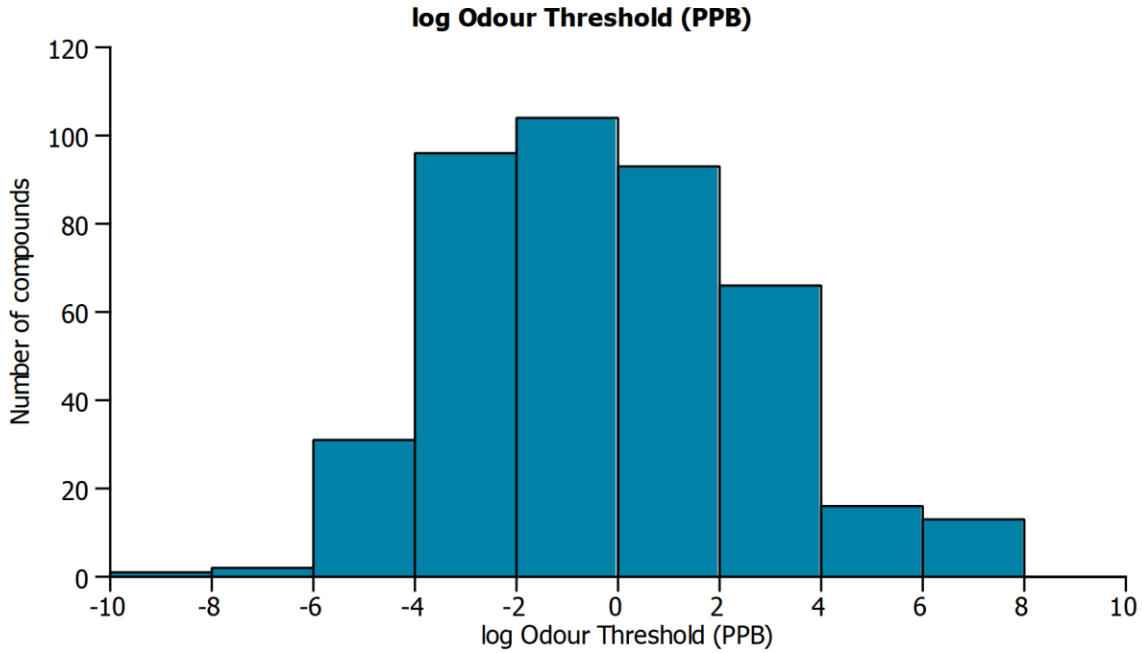
Case Study
Leffingwell Odour Threshold Models

optibrium



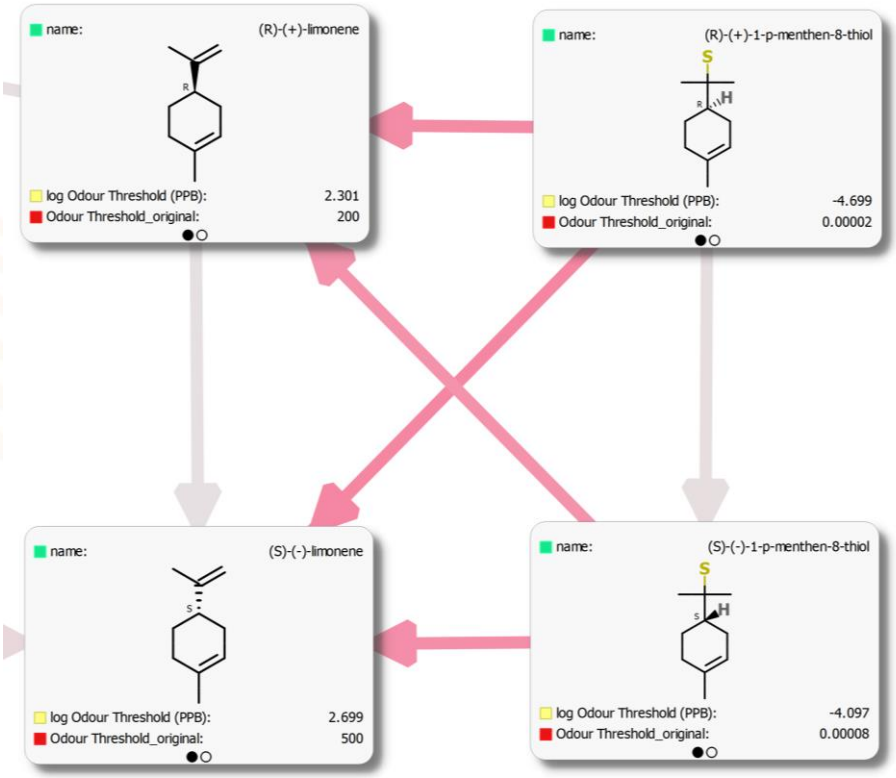
Leffingwell Odour Threshold

- Leffingwell
  - 422 compounds with odour descriptions and measured detection threshold



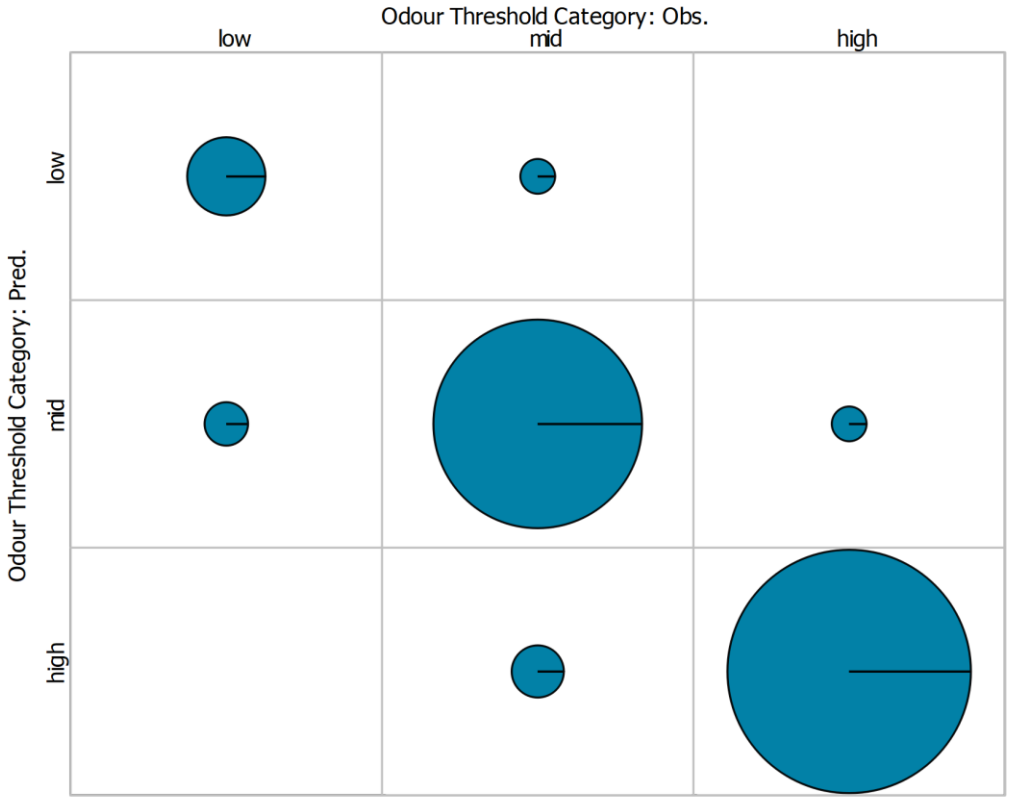
Enantiomer matched pairs

- Enantiomer pairs can have very different odour
  - Limonenes and related thiols
- All have citrus-like odours
  - Thiols are grapefruit
  - Limonenes are orange (+ or R) or harsh lemon/turpentine like (- or S)
- Thiols are more odourous (6 log units)



Odour Threshold Model

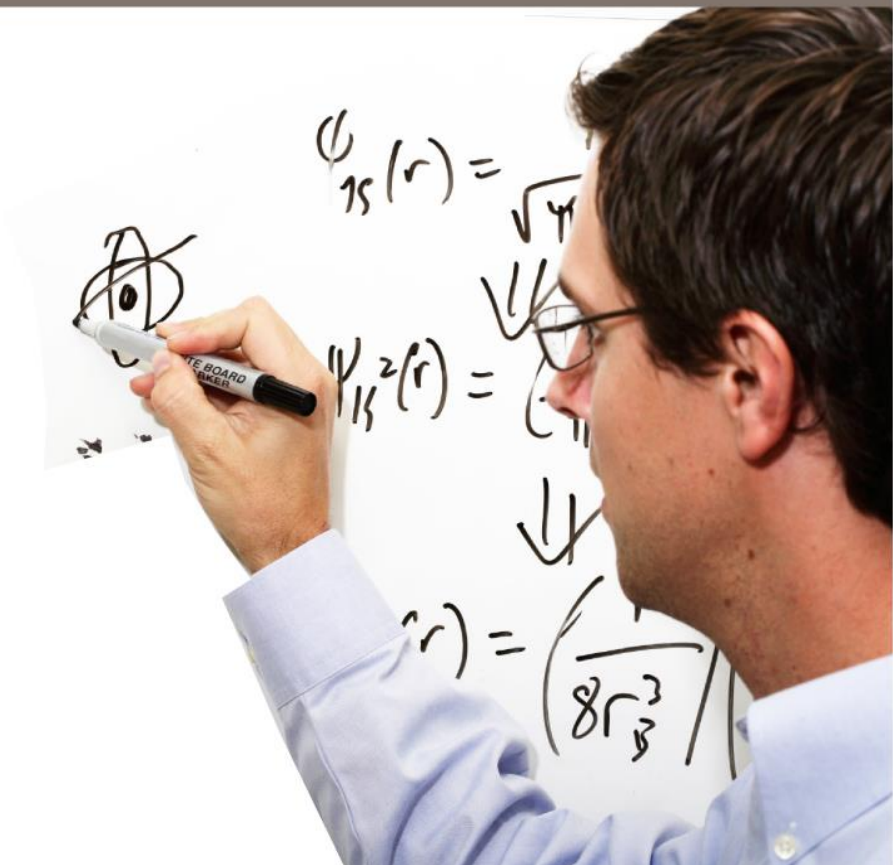
- Decision Trees Categorical Model
  - Log Odour Detection (PPB)
    - Low: ≤ -3
    - Mid: > -3 to ≤ 1
    - High: > 1
  - Training: Validation: Test 80:1:19 Y-based split
  - Test set
    - Kappa 0.62
    - Accuracy 0.76



Confusion Matrix: Test Set

Case Study
Skin Sensitisation Model

optibrium



Food and Chemical Toxicology 74 (2014) 164–176

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

An *in silico* skin absorption model for fragrance materials

Jie Shen [a], Lambros Kromidas [a,*], Terry Schultz [b], Sneha Bhatia [a]

[a] Research Institute for Fragrance Materials, Inc., 50 Tice Blvd., Woodcliff Lake, NJ 07677, USA
[b] College of Veterinary Medicine, The University of Tennessee, 2407 River Dr., Knoxville, TN 37996, USA

ARTICLE INFO

Article history:
Received 6 August 2014
Accepted 23 September 2014
Available online 5 October 2014

Keywords:
Toxicology
QSAR
Computational
Topical
Safety
Dermal

ABSTRACT

Fragrance materials are widely used in cosmetics and other consumer products. The Research Institute for Fragrance Materials (RIFM) evaluates the safety of these ingredients and skin absorption is an important parameter in refining systemic exposure. Currently, RIFM's safety assessment process assumes 100% skin absorption when experimental data are lacking. This 100% absorption default is not supportable and alternate default values were proposed. This study aims to develop and validate a practical skin absorption model (SAM) specific for fragrance material. It estimates skin absorption based on the methodology proposed by Kroes et al. SAM uses three default absorption values based on the maximum flux (J_{max}) – namely, 10%, 40%, and 80%. J_{max} may be calculated by using QSAR models that determine octanol/water partition coefficient (K_{ow}), water solubility (S) and permeability coefficient (K_p). Each of these QSAR models was refined and a semi-quantitative mechanistic model workflow is presented. SAM was validated with a large fragrance-focused data set containing 131 materials. All resulted in predicted values fitting the three-tiered absorption scenario based on J_{max} ranges. This conservative SAM may be applied when fragrance material lack skin absorption data.

Shen Skin Absorption Model

- Skin absorption is important for assessing systematic exposure to fragrances in cosmetics
- When no experimental data, 100% absorption is assumed
- Categorise absorption using J_{max}
  - Formulation independent, theoretically achieved dose based on Fick's first law of diffusion
  - J_{max} = Permeability coefficient (K_p) * Saturated water solubility (C_{water})
  - K_p is proportional to MW and log K_{ow}
- Shen's 3-category *in-silico* semi-quantitative model for J_{max}
  - **'% Abs <10%' (Low):** J_{max} <= 0.1 µg/cm^2/h
  - **'% Abs 10 - ≤40%' (Medium):** 0.1 µg/cm^2/h < J_{max} <= 10 µg/cm^2/h
  - **'% Abs >40 - ≤80%' (High):** J_{max} > 10 µg/cm^2/h

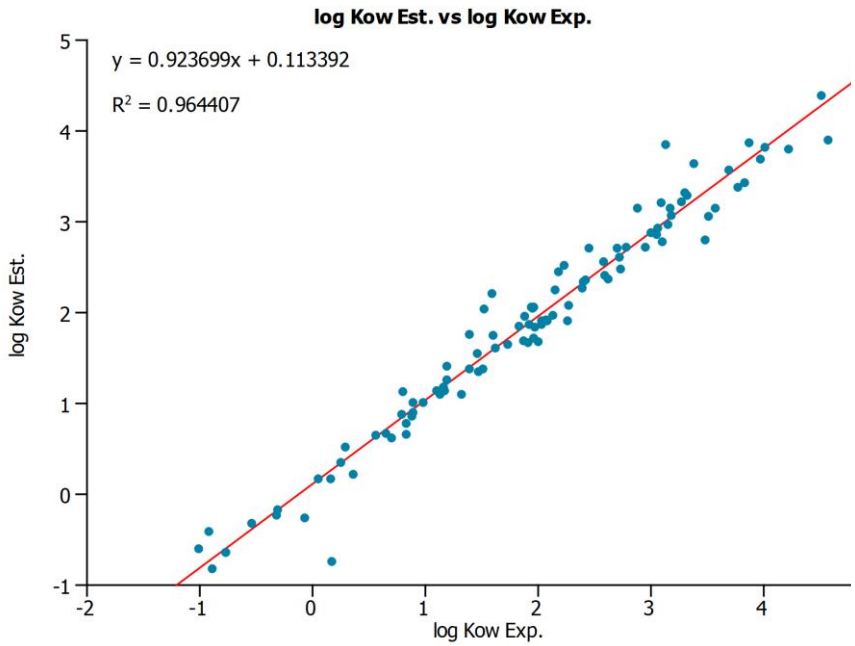
Shen Data Sets

- Fragrance and fragrance-like molecules
- Data set 1
  - 105 compounds
  - Experimental and calculated values for log K_{ow} and log K_p
- Data set 2
  - 155 compounds
    - 131 compounds
    - 24 additional compounds derived by hydrolysis of some of the 131
  - Either experimental or calculated values for the key parameters:

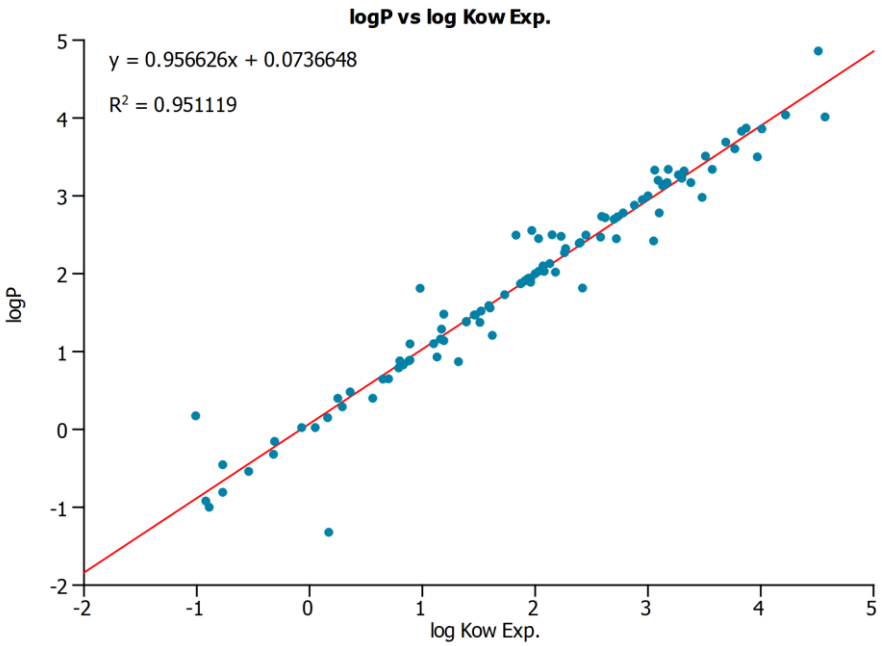
    logK_{ow}, log K_p, log K_p corrected, C_{water}, J_{max}, categorical %abs estimated and %abs experimental

logP can estimate log K_{ow}

7-model LogP Consensus

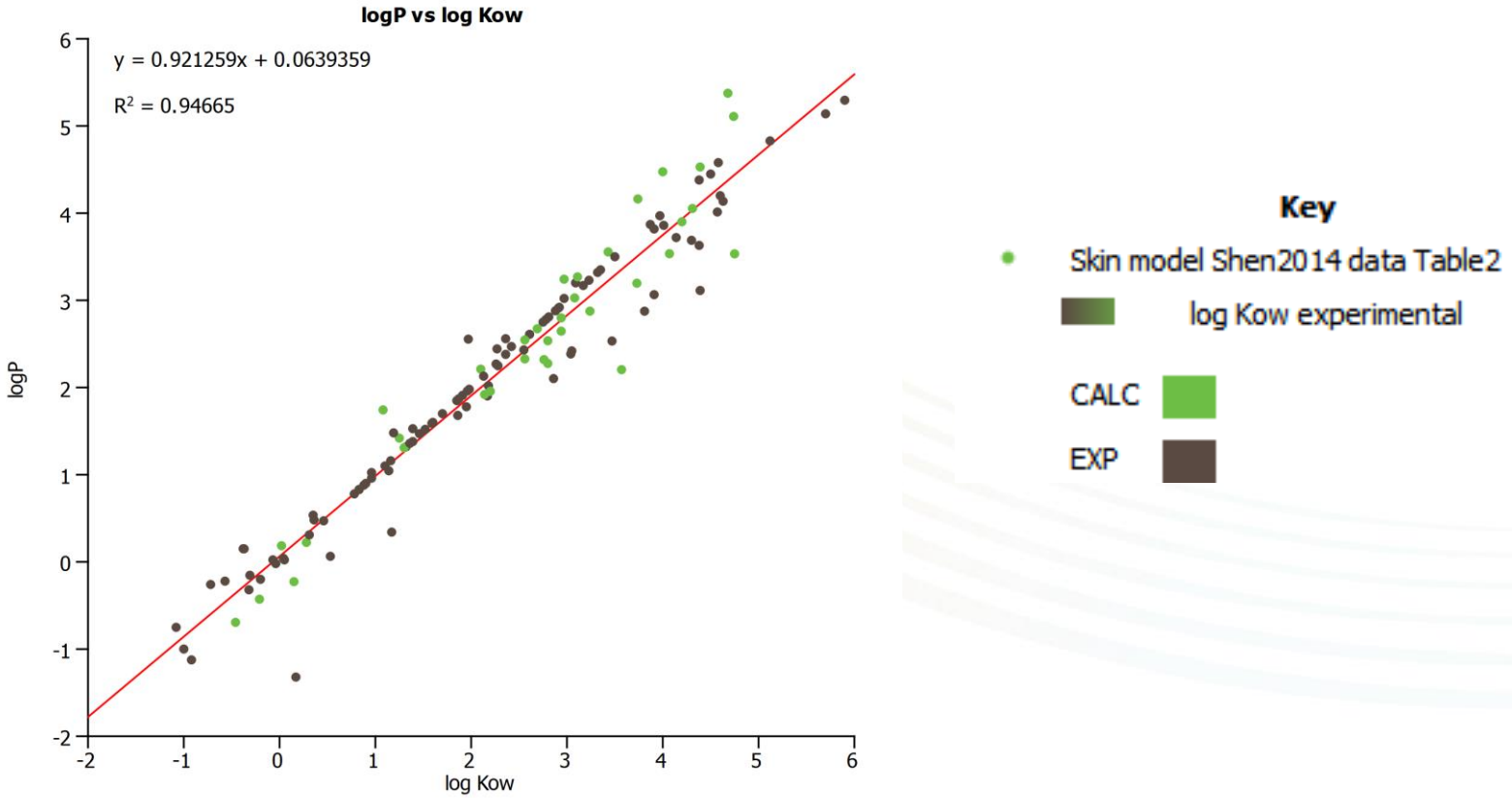


StarDrop™ LogP



logP can estimate log K_{ow}

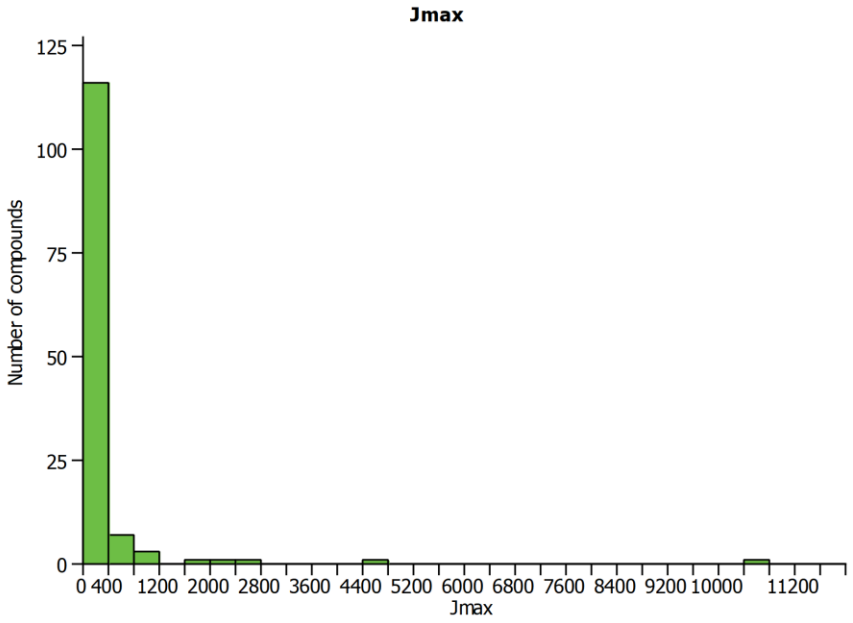
- Data set 2 (155 compounds)

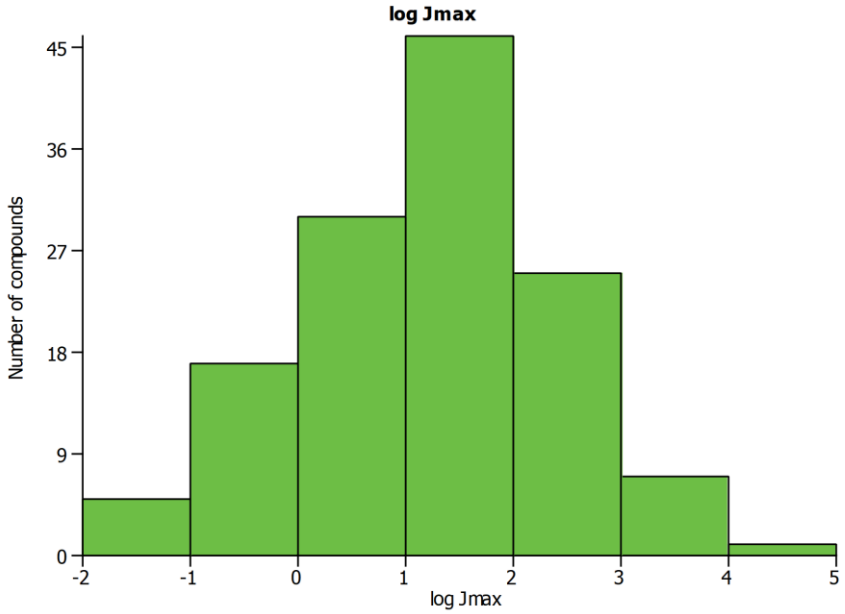


- No significant bias towards experimental (brown) vs calculated (green) data

Modelling J_{max} directly

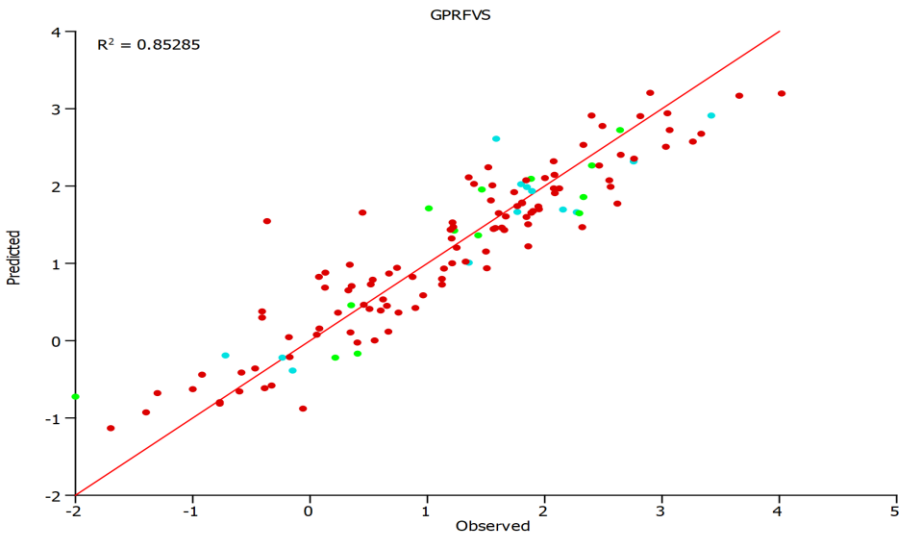
- J_{max} data is skewed; therefore model normalized log J_{max}





Log J_{max} Model

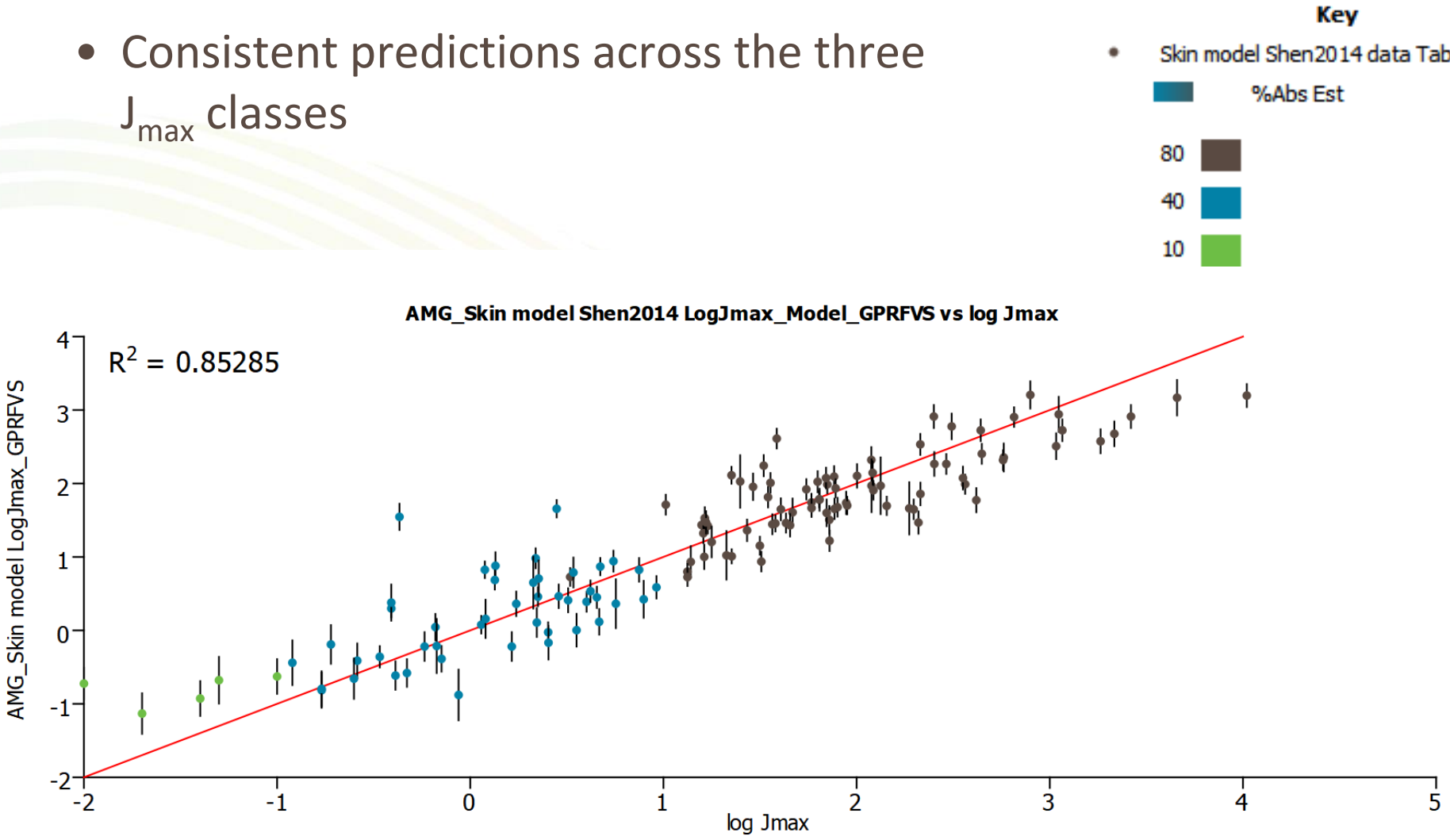
- Data set 2 (131 compounds)
  - training:validation:test sets in an 80:10:10 ratio
- Best model is GP-RVFS model using 11 descriptors



	Training set		Validation set		Test set	
Model	R^2	RMSE	R^2	RMSE	R^2	RMSE
GPRFVS	0.857	0.449	0.81	0.529	0.849	0.45

Log J_{max} Model

- Consistent predictions across the three J_{max} classes



Conclusions

- Pharma, flavours and fragrances all require a similar MPO approach to identify compounds with a balance of desirable properties
- Most flavours and fragrances molecules are outside the pharma Chemical Space
  - But flavours and fragrances more similar to each other
- Pharma models can be a good starting point for flavours and fragrances
  - Confirm predictions are within the model's Domain of Applicability
- Automated QSAR model building can produce robust models for flavours and fragrances properties

Acknowledgements

- Optibrium
  - Peter Hunt
  - Matt Segall
  - Scott Lyon
- International Flavours and Fragrances
  - Jack Bikker