



optibrium

Combining Quantum and QSAR Methods for the Prediction of Acid Dissociation Constants

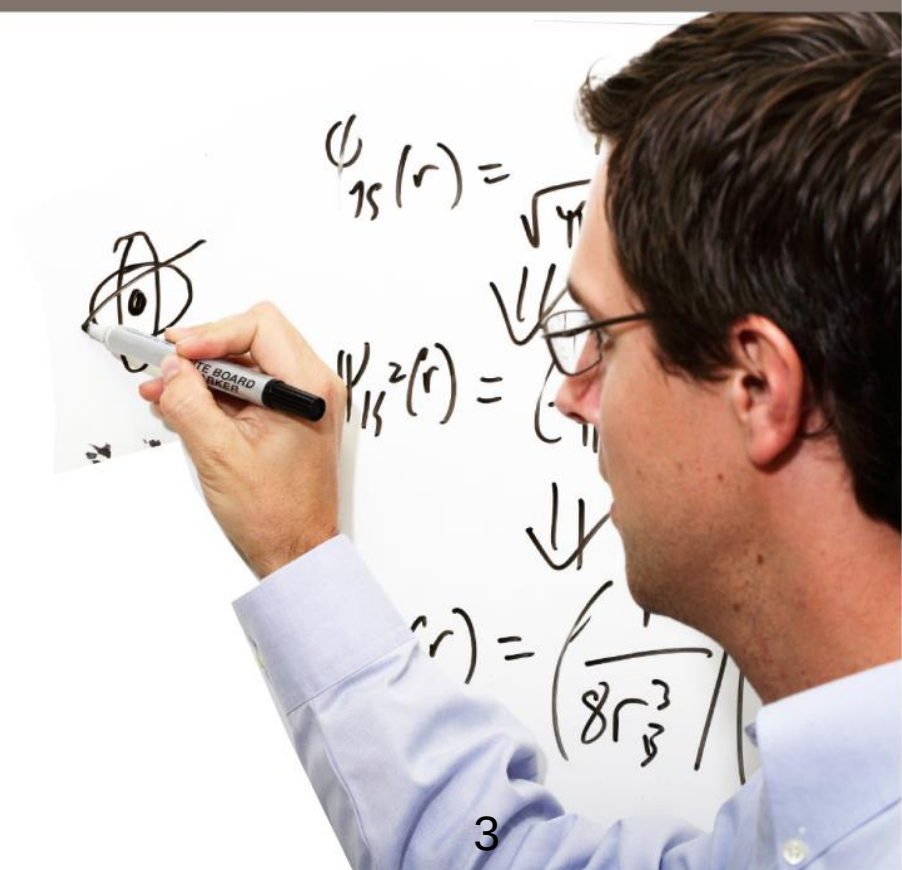
Layla Hosseini-Gerami*, Rasmus Leth, Peter Hunt, Matthew Segall

*layla@optibrium.com

Overview

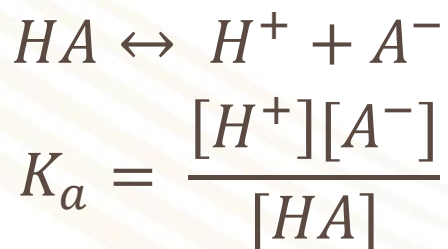
- The value of a good pK_a prediction
 - Context and past research
- DFT approach
- Semi-empirical approach
 - Predictions for separate compound classes
 - Unified model
- Summary and future steps

The Value of a Good pK_a Prediction



The Negative Logarithmic Acid Dissociation Constant: pK_a

- Quantitative measure of the strength of an acid in solution
- K_a is the equilibrium constant for an acid dissociation reaction

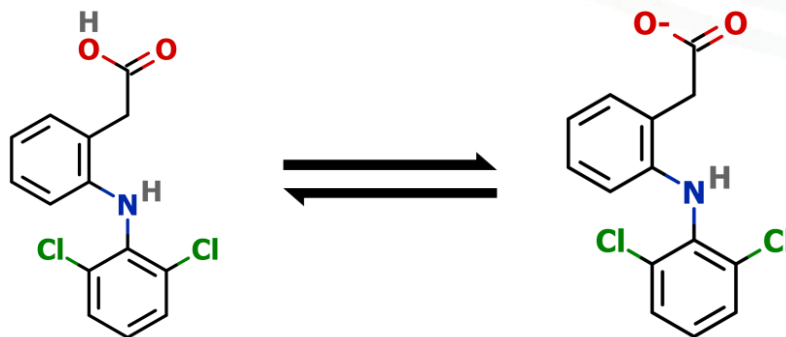


- pK_a is the negative logarithmic constant for the acid dissociation reaction – the smaller the pK_a , the stronger the acid

$$pK_a = -\log K_a$$

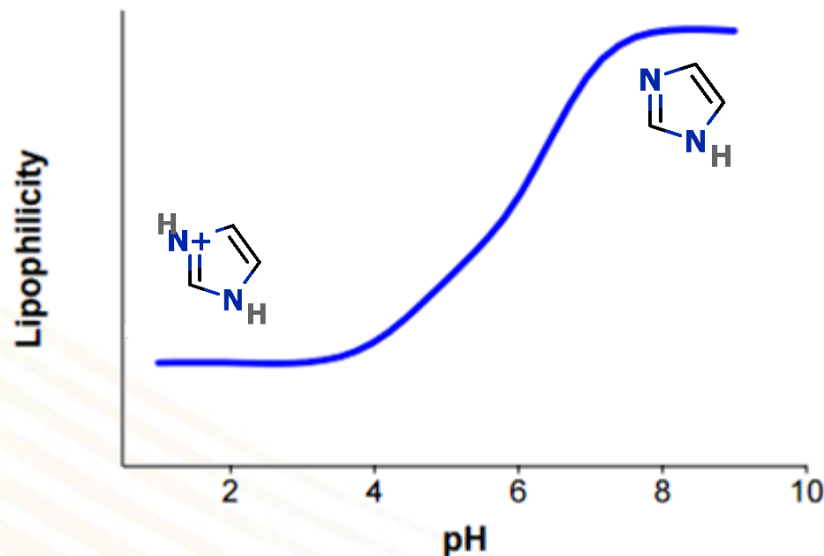
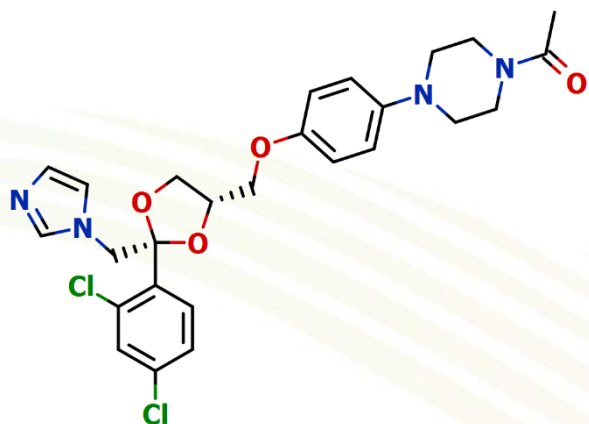
The Value of a Good pK_a Prediction

- Improved *in silico* tools are useful to medicinal chemists in the research stage
 - Manage time, cost, resources
 - Ensure that good quality compounds make it to clinical trial
- Knowledge of pK_a is valuable to chemists' decision making
 - Influence on physicochemical and ADME properties
 - o Absorption, Distribution, Metabolism, Excretion



The Value of a Good pK_a Prediction

Ketoconazole – pK_a 6.43, 3.64



pH	LogD [1]	Aqueous solubility (mg/mL) [2]	% Ionised
3	0.35	0.43	Highest
5	2.37	0.1	↓
7.4	3.46	0.006	Lowest

[1] Mackenzie, H. (2013). *The Central Role of pK_a in Drug Discovery*.

[2] Ghazal, H., Dyas, A., Ford, J. and Hutcheon, G. (2014). The impact of food components on the intrinsic dissolution rate of ketoconazole. *Drug Development and Industrial Pharmacy*, 41(10), pp.1647-1654.

Project Context

Current and past research

- Direct pK_a calculation
 - *Ab initio* and DFT studies with implicit solvation model [1]
- QSAR: Correlating pK_a with *ab initio* and DFT calculated descriptors
 - Partial atomic charges [2]
 - Bond lengths [3]
- Correlation with semi-empirical FMO descriptors calculated with computed eigenvectors and eigenvalues [4]

[1] da Silva, G., Kennedy, E. and Dlugogorski, B. (2006). *Ab Initio* Procedure for Aqueous-Phase pK_a Calculation: The Acidity of Nitrous Acid. *The Journal of Physical Chemistry A*, 110(39), pp.11371-11376.

[2] Svobodová Vařeková et al. (2011). Predicting pK_a Values of Substituted Phenols from Atomic Charges: Comparison of Different Quantum Mechanical Methods and Charge Distribution Schemes. *Journal of Chemical Information and Modeling*, 51(8), pp.1795-1806.

[3] Harding, A. and Popelier, P. (2011). pK_a Prediction from an *ab initio* bond length. *Physical Chemistry Chemical Physics*, 13(23), p.11264.

[4] Tehan et al. (2002). Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids. *Quantitative Structure-Activity Relationships*, 21(5), pp.457-472.

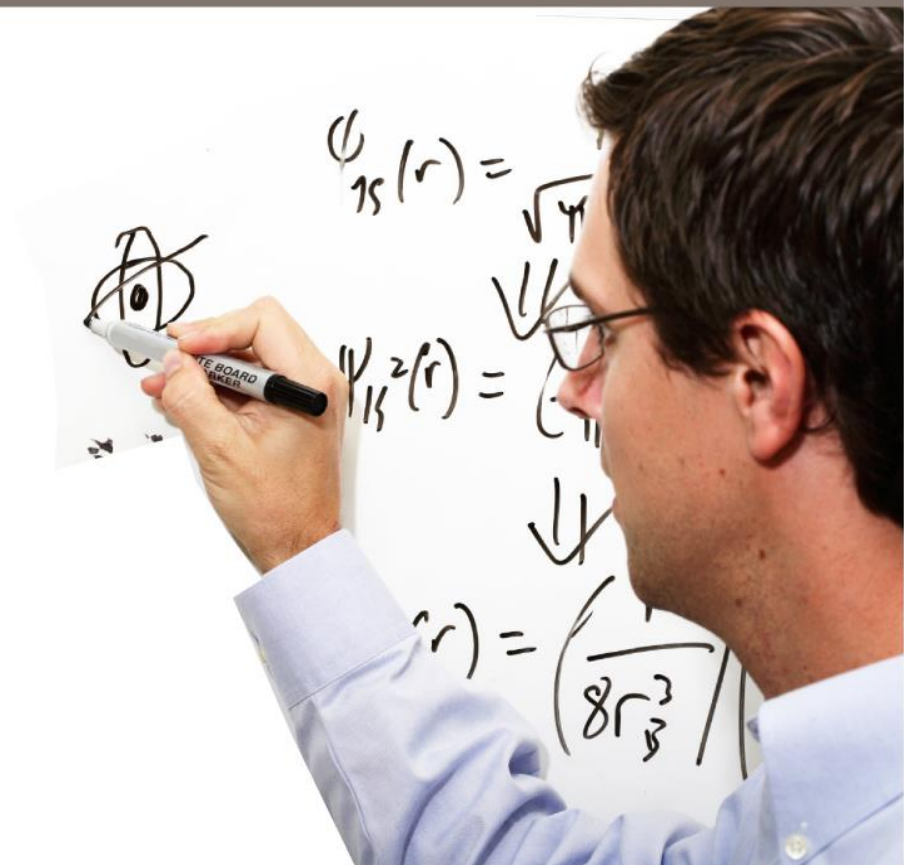
Drawbacks

- DFT and *ab initio* are computationally expensive
- FMO descriptors alone were not enough to produce a good correlation
- Separate models for separate compound classes – not universal across any ionisable centre.

Goals for Improvements to Current Methods

- Multi-faceted QSAR model **combining** partial atomic charge and bond length descriptors with FMO and energy descriptors
 - All calculated with AM1 level of theory
 - Much faster than calculating descriptors using *ab initio* and DFT methods
- Unified model: One model accurately predicting pK_a for multiple compound classes
 - Previous research involved one model for each compound class resulting in the production of a number of different models

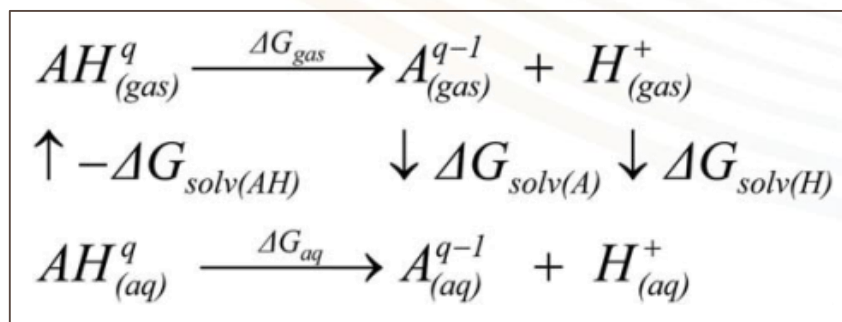
DFT Approach



DFT Approach

Computational details

- First looked at accurate DFT direct calculation method
 - Using a thermodynamic cycle and Gibbs energies
- B3LYP/6-31++G** level of theory and COSMO solvent model
 - Continuum solvation model
- Calculated using NWCHEM software

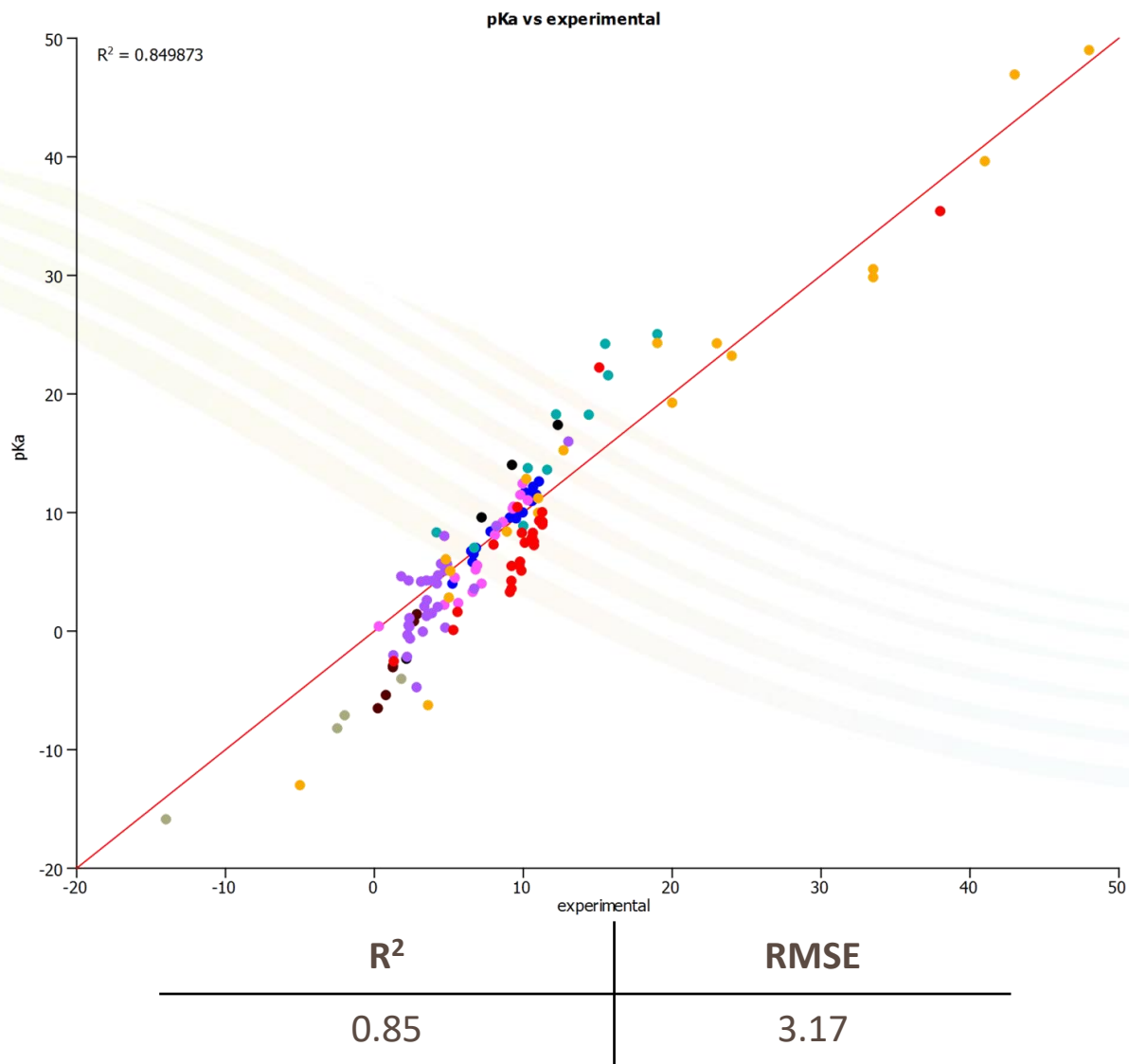


$$pK_a = \Delta G_{(aq)} / 2.303 RT$$

Image credit: Casasnovas, R., Ortega-Castro, J., Frau, J., Donoso, J. and Muñoz, F. (2014). Theoretical pK_a calculations with continuum model solvents, alternative protocols to thermodynamic cycles. *Int. J. Quantum Chem.*, 114(20), pp.1350-1363.

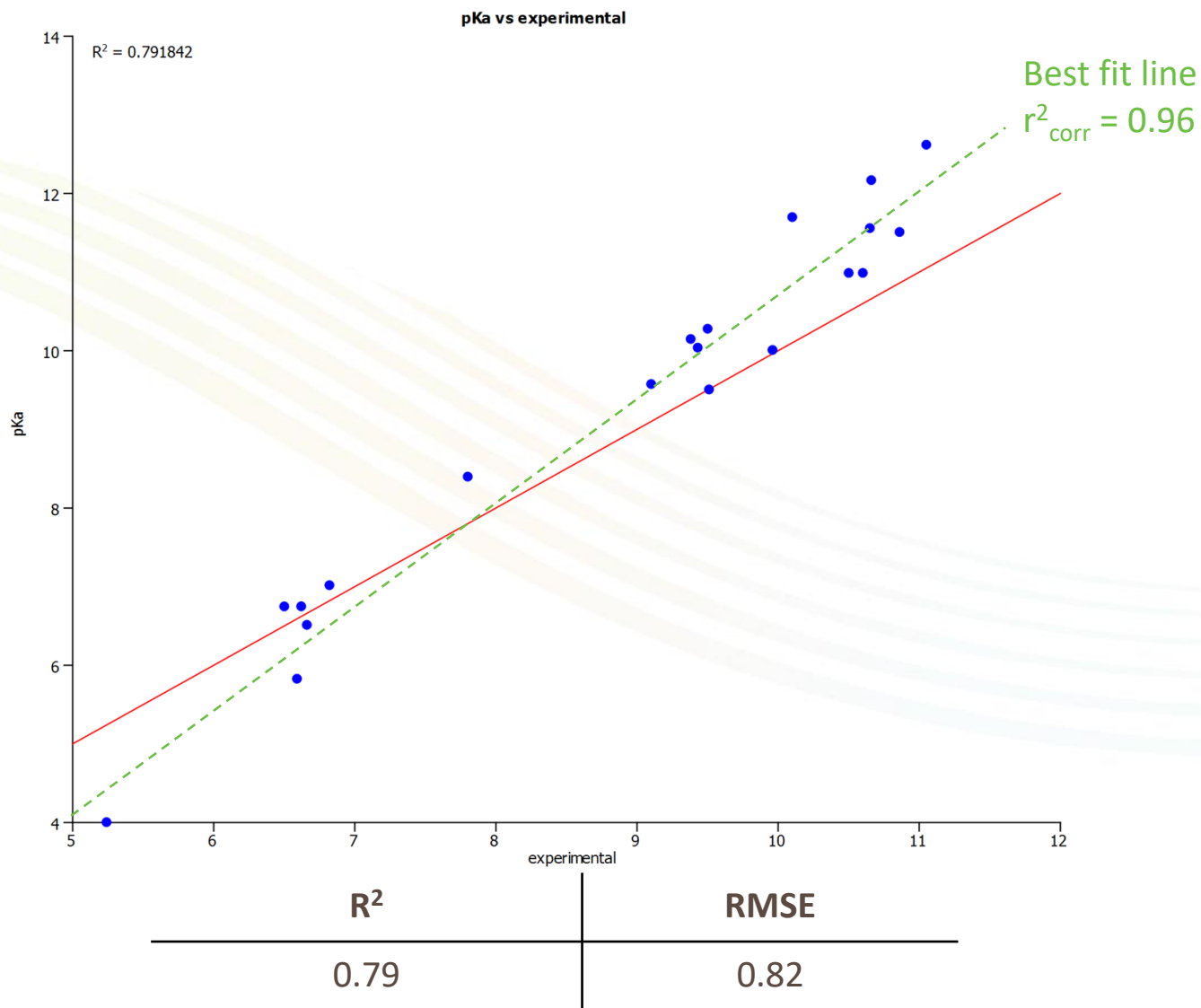
DFT Results

140 compounds



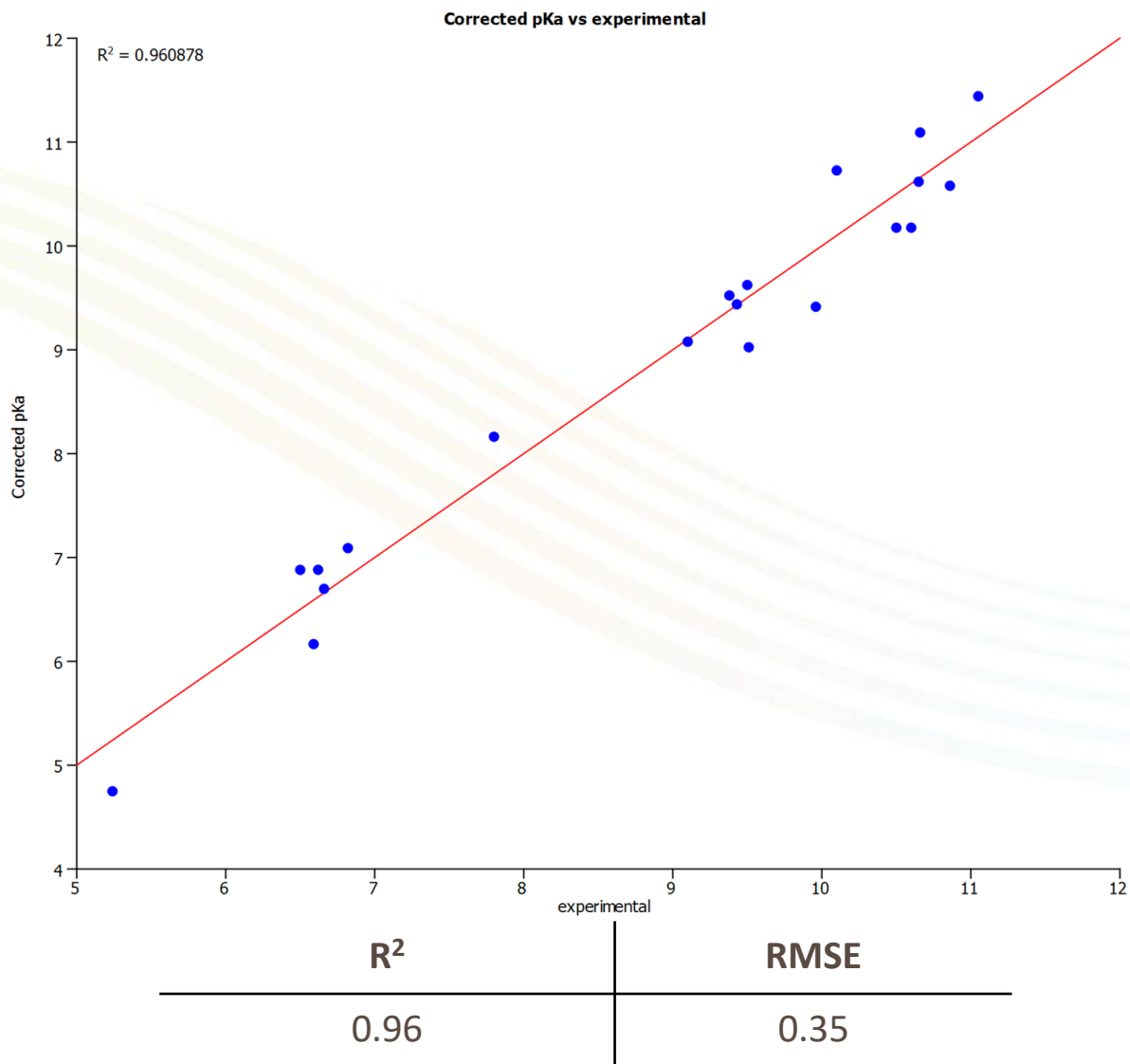
DFT Results

Thiols



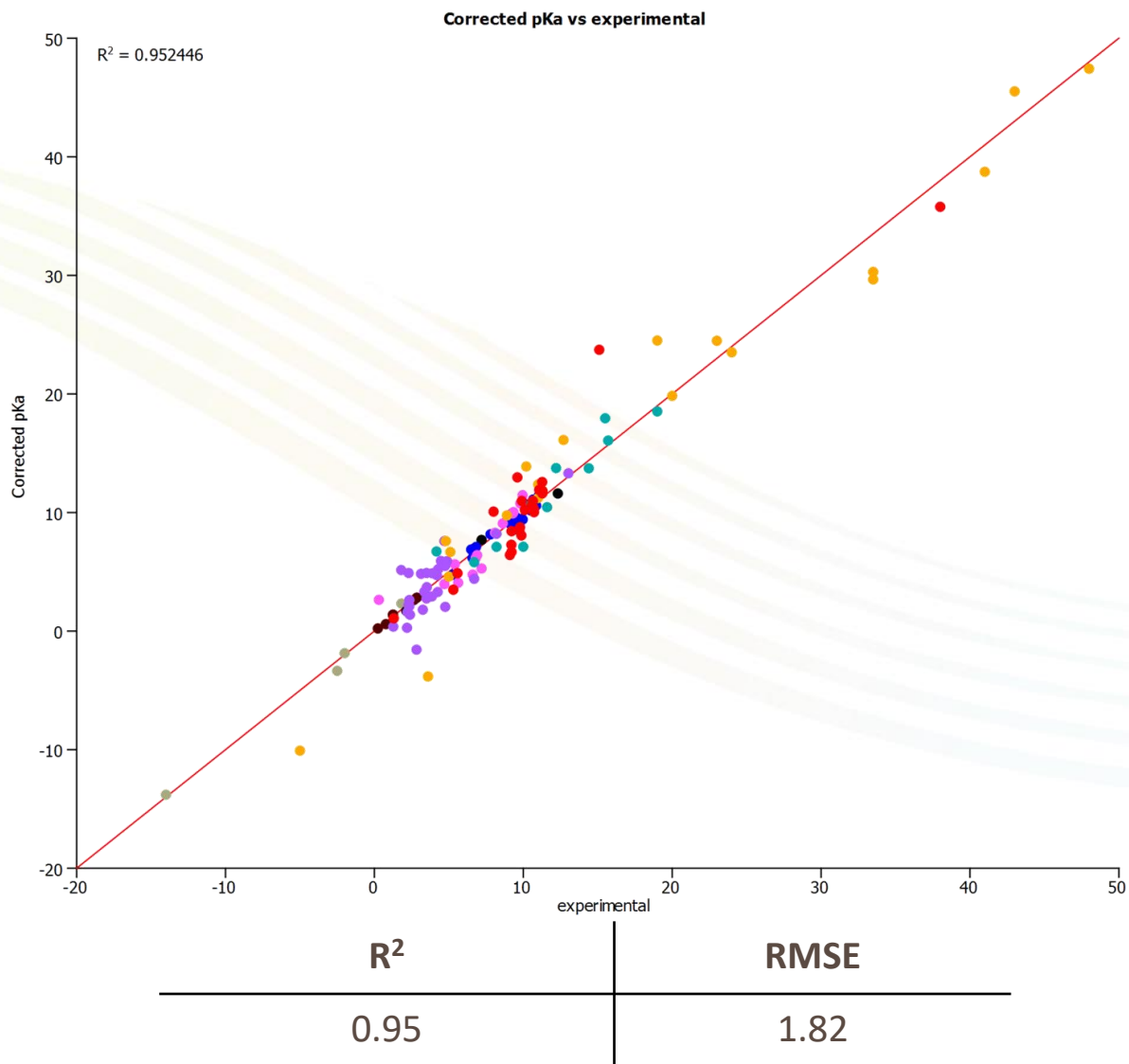
DFT Corrected Results

Thiols



DFT Corrected Results

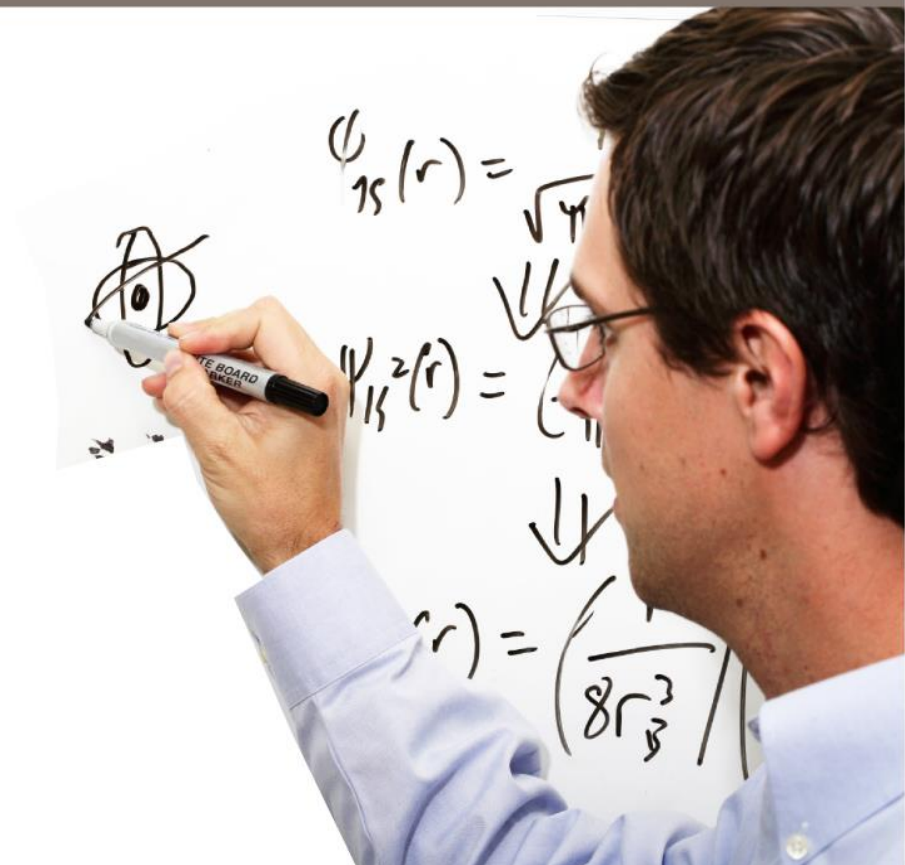
Corrections to all compound classes



DFT Drawbacks

- Although an accurate method, computational time is slow
 - Calculations took days, or weeks for large drug-like compounds
- End-user will want accurate **and** fast results
 - Smaller basis sets still computationally expensive
- Ultimately, a compromise is needed between accuracy and speed
 - Can semi-empirical methods produce the same results?

Semi-empirical Approach



Direct Calculation of pK_a with AM1

- MOPAC calculations of energy terms
 - AM1 semi-empirical method with COSMO solvation model
 - Method based on the Neglect of Differential Diatomic Overlap (NDDO) integral approximation
- Direct calculation of pK_a failed
- No correlation between AM1 and DFT energies
- How to use AM1 to get a meaningful prediction?

QSAR

An effective way to use AM1

- pK_a of a compound depends on its structure
 - Thus, it may be possible to find a Quantitative Structure Activity Relationship (QSAR)
 - Is AM1 able to calculate descriptors to build a QSAR model?
 - Including the *ab initio* and DFT calculated descriptors which showed a correlation in previous research
- Calculation of various descriptors for ~600 compounds shows positive results

Computational Details

- AM1 in MOPAC
- All calculations carried out in gas phase
- Descriptors calculated:
 - Bond lengths
 - Partial charges
 - HOMO/LUMO energies
 - Heats of formation
 - FMO descriptors: electrophilic superdelocalisability (SE) and nucleophilic frontier electron density (FN) [1]

$$SE(p) = 2 * \sum_{j=1,m} \sum_{\alpha=1,q} (c_{\alpha j}^2 / \lambda_j) \quad FN(p) = \sum_{\alpha=1,q} c_{\alpha(m+1)}^2$$

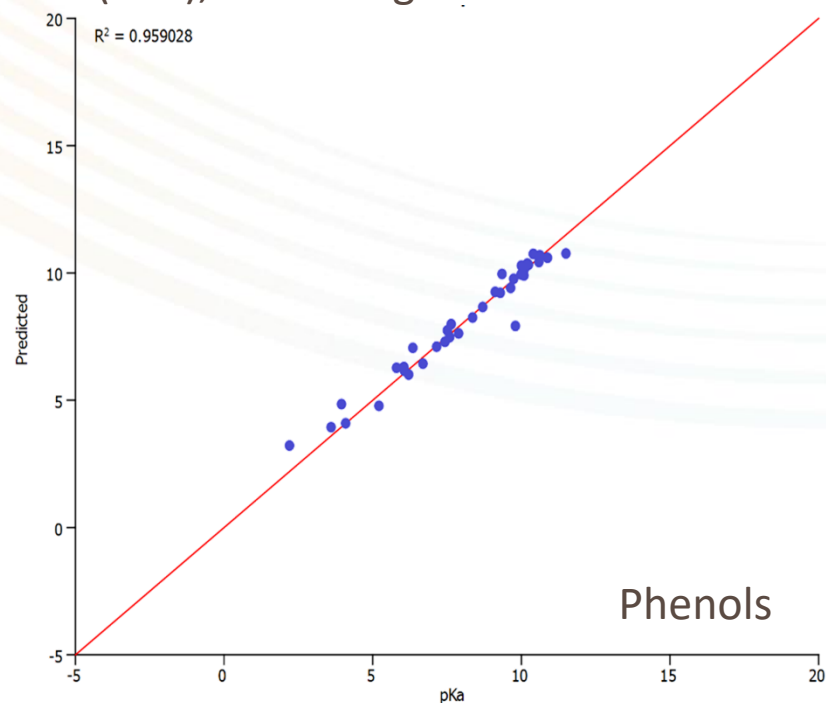
[1] Fukui, K., Yonezawa, T. and Nagata, C. (1954). Theory of Substitution in Conjugated Molecules. *Bulletin of the Chemical Society of Japan*, 27(7), pp.423-427.

Results

One model per compound class

- Auto-Modeller in StarDrop™ was used to build QSARs for individual compound classes
- Tested various modelling methods such as Random Forests, PLS and Gaussian Processes
- The most successful, Radial Basis Function (RBF), modelling method used for all compound classes

Compound class	R ² (test set)	RMSE
Carboxylic acids	0.89	0.37
Heterocycles	0.89	0.64
Amines	0.96	0.77
Phenols	0.96	0.47
Thiols	0.72	0.97
Oxygen acids	0.88	0.84

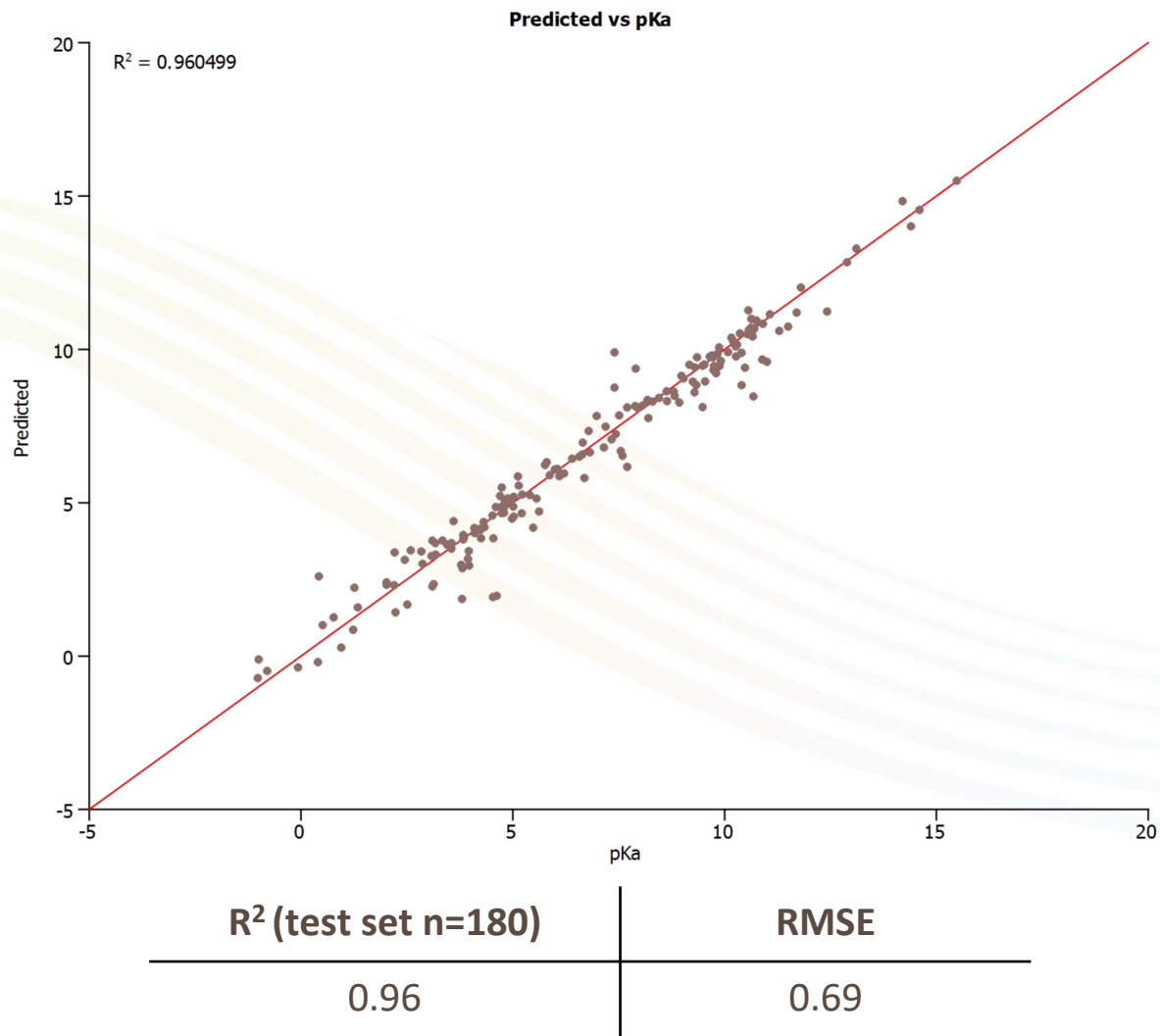


Unifying the Model

- Attempted to build a unified model
 - One model which will be used for all compound classes as opposed to separate models for each compound class
- Encompasses all descriptors used in each model and, additionally, a binary indicator variable to classify the site of deprotonation
- RBF model produced in Auto-Modeller
 - Keeping consistency in the modelling method from the individual models to the unified model

Results

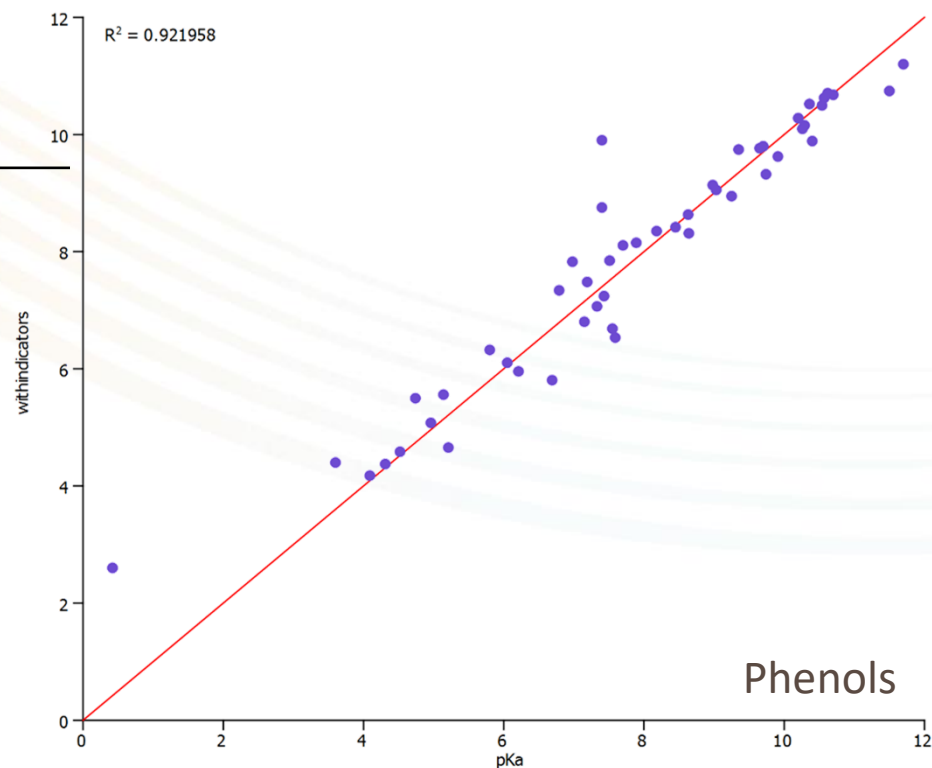
External test set



Performance of Model on Different Classes

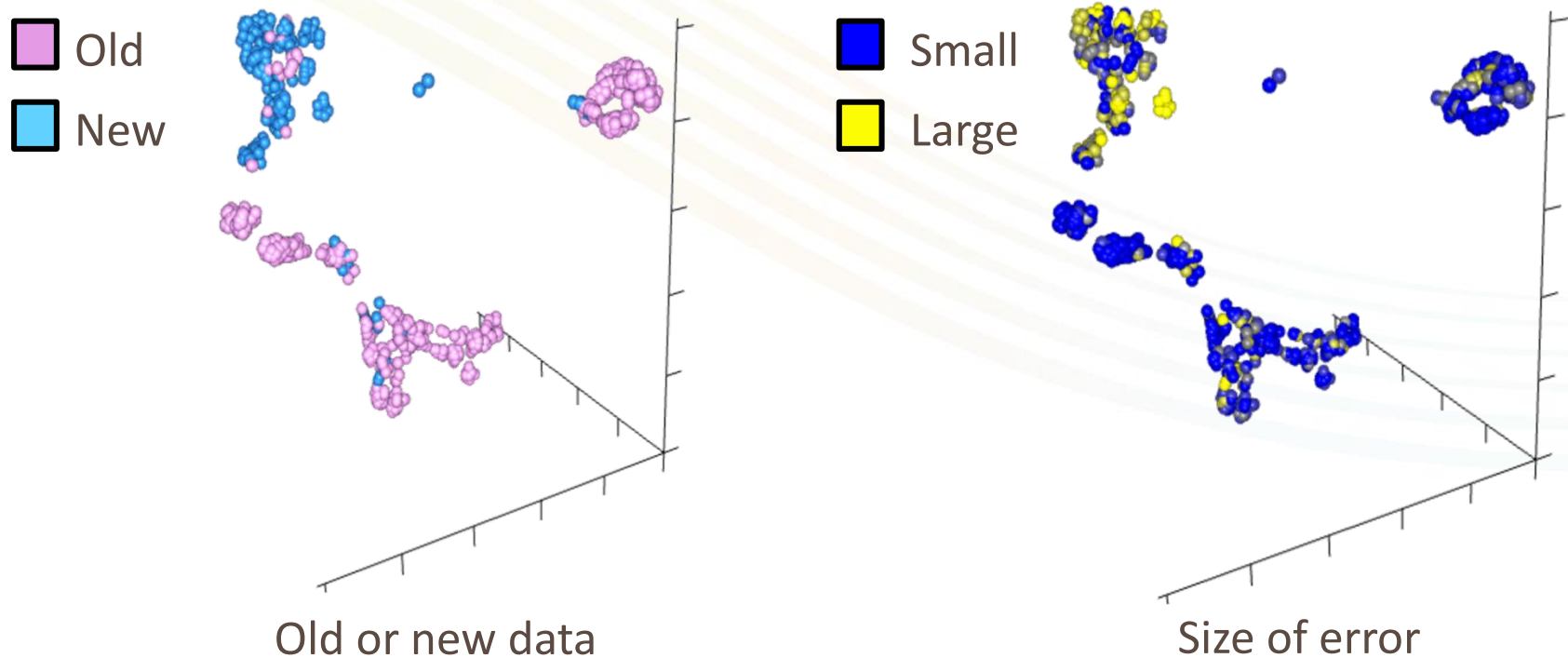
- Tested unified model on separated compound classes
 - To evaluate if different classes are predicted better than others

Compound class	R ²	RMSE
Carboxylic acids	0.82	0.47
Heterocycles	0.92	0.58
Amines	0.94	0.99
Phenols	0.92	0.66
Thiols	0.70	0.74
Oxygen acids	0.94	0.57



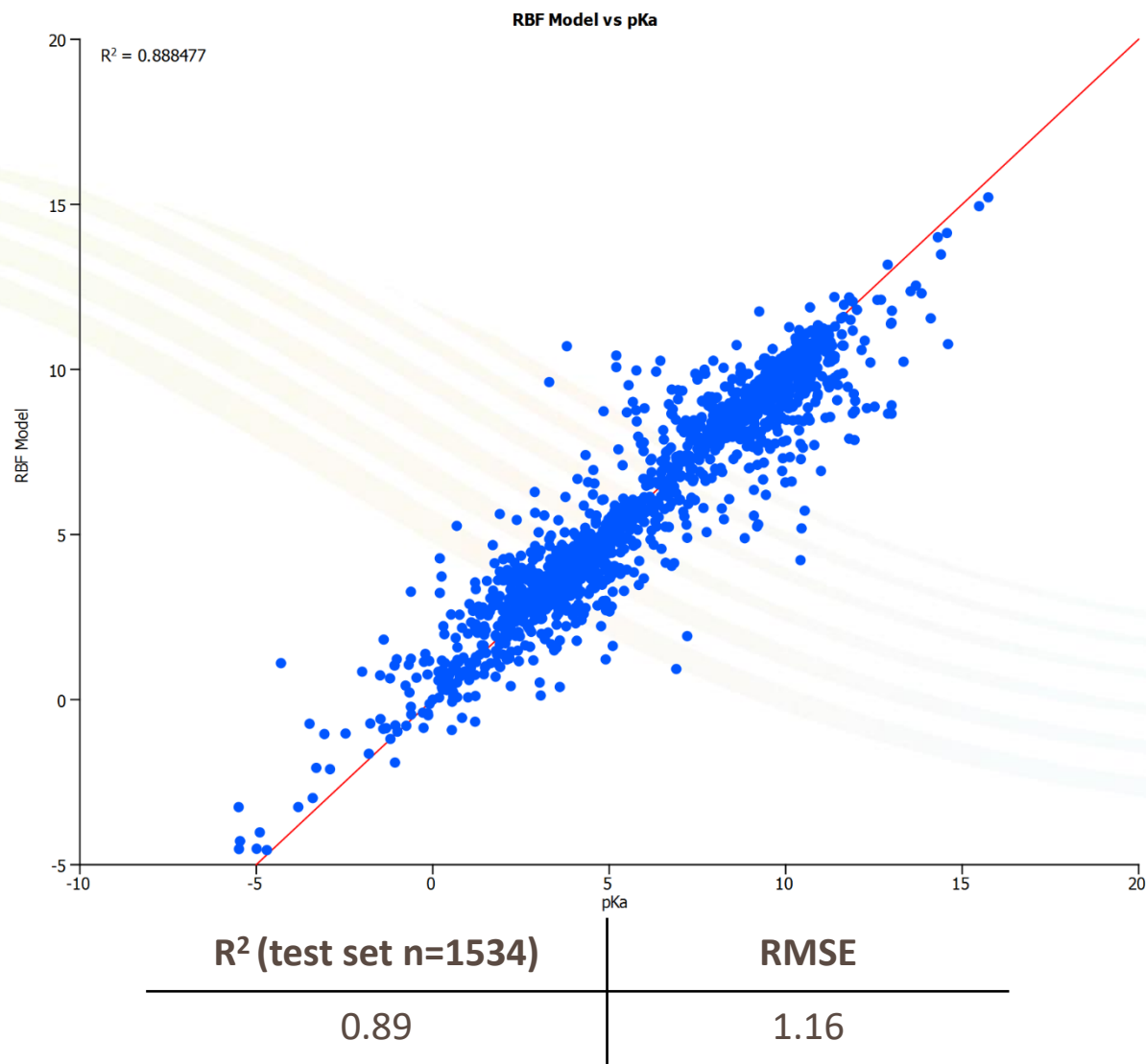
Applying the Model to a Larger Dataset

- Data provided by Lhasa Limited, <https://www.lhasalimited.org/>
- New data outside of domain of applicability

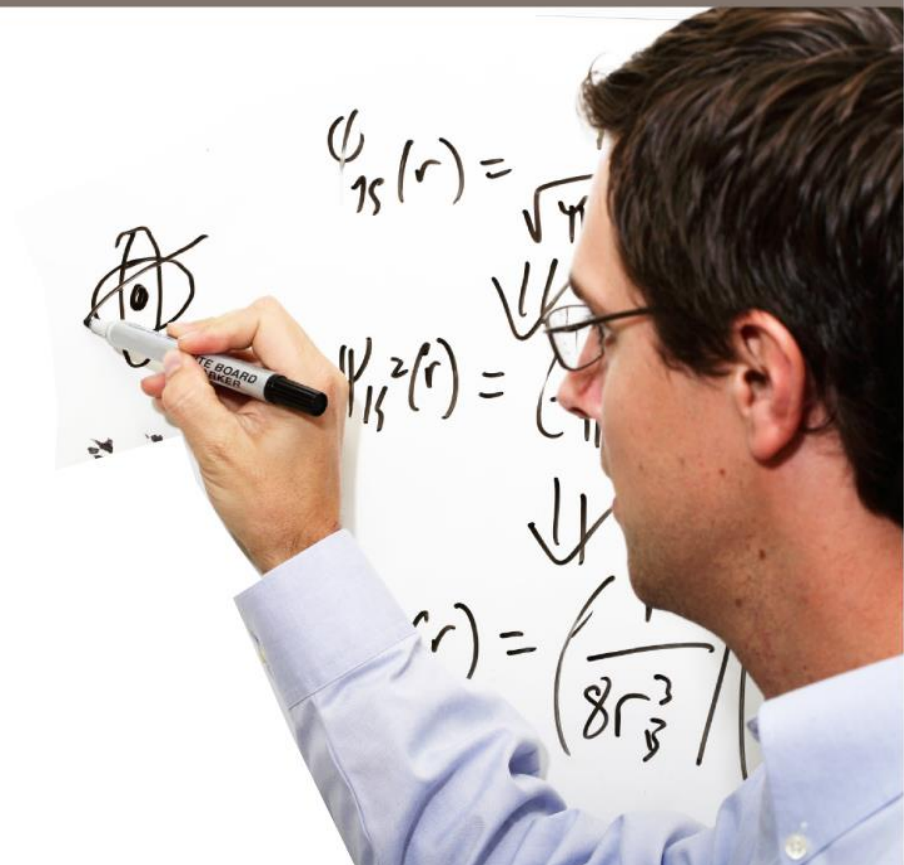


Retraining the Model on a Larger Dataset

Independent test set results



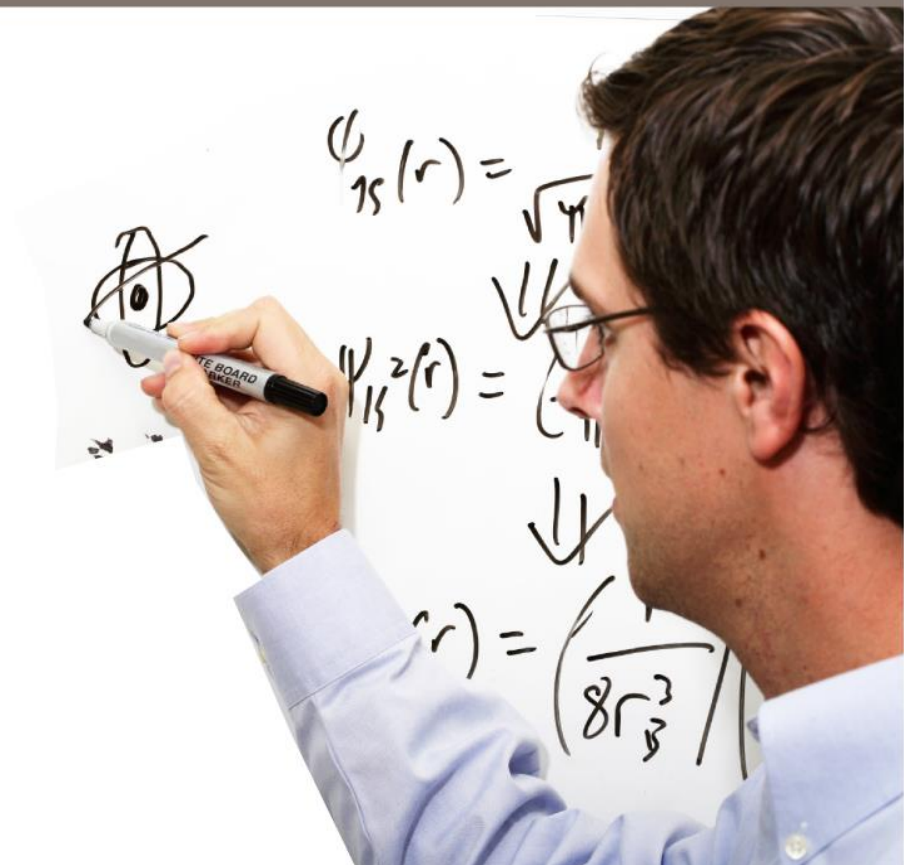
Summary and Further Steps



Summary and Further Steps

- Semi-empirical quantum calculations used in QSAR produced an accurate, computationally inexpensive pK_a model
- Attempts to create a unified model proved successful with the ability to make excellent predictions
- Future considerations:
 - Inclusion of consecutive deprotonation of multiple ionisation states
 - More accurate semi-empirical method (e.g. PM6)
 - Solvation effects taken into account

Thank You for Listening



Acknowledgements

- Optibrium
 - Continued guidance, advice and support from Peter Hunt, Matt Segall and Rasmus Leth
- Data provided by Lhasa Limited
- Funding from the Royal Society of Chemistry
- University of Leeds
 - Organisation of the placement program and support from Profs. Dmitry Shalashilin and Paul Seakins

