3D QSAR
Why be square?

M Mackey, T Cheeseright, J Melville, R Scoffin, C Earnshaw

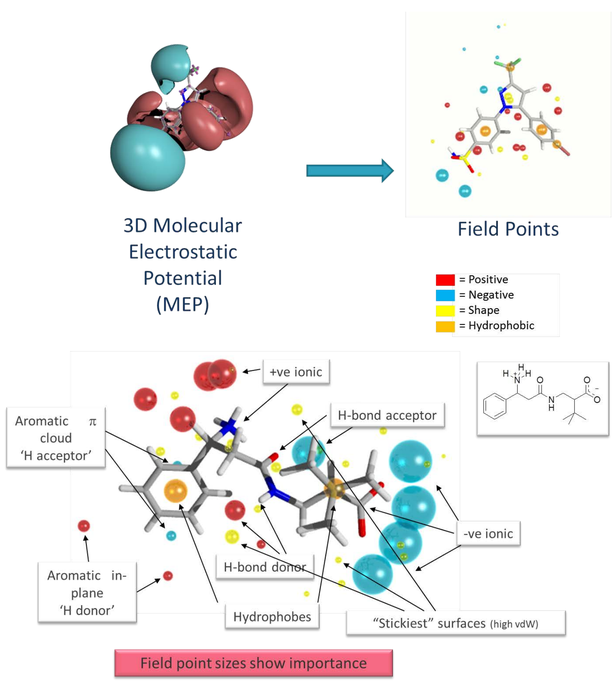
Cresset BioMolecular Discovery Ltd, Biopark, Broadwater Road, Welwyn Garden City, Herts, AL7 3AX, UK www.cresset-group.com

Introduction

3D QSAR based on molecular interaction potentials is a technique with a long pedigree. However, current techniques have a number of issues which make it difficult to reliably produce effective models. These include: (a) the alignment problem (how to align molecules in 3D as a pre-requisite to computing descriptors); (b) the sampling problem (sampling on a grid, as is widely done, produces edge effects and extra noise); and (c) the descriptor problem (which molecular properties should be calculated at the sample positions, and how accurate are those properties?).

Cresset presents **forge**V10, a new integrated approach to producing 3D QSAR models, utilizing our "field point" descriptors[1]. These descriptors are used first to align molecules, either using a free alignment or by first aligning common substructures. Once the molecules are aligned, the collective field points of the training set are used to derive a gauge-invariant set of sampling points, which reduces the number of descriptors that need to be considered while ensuring that all regions around the molecule which might contribute to activity are adequately sampled. Finally, sample values are calculated and a Partial Least Square (PLS) model built.

Fields

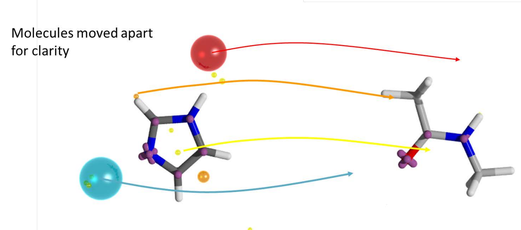


The field points of a molecule are generated by locating the points in space where the electrostatic and steric fields of a molecule are locally maximal. Each such point is labelled with the strength of the field at that point. We use four fields: positive and negative electrostatic, "shape" (van der Waals), and "hydrophobic" (a density function correlated with steric bulk and hydrophobicity).

The in-depth treatment of electrostatics provided by the XED force field is required to get good molecular electrostatic potentials.

Field Sampling

Given a set of aligned molecules, their collected field points define a set of sampling points in space. We run a sphere exclusion algorithm to filter this set down, and then calculate the field value for each molecule to generate a data matrix, on which we apply PLS. The technique is similar to CoMFA, but rather than sample at grid points around the molecules, we sample only at places where one or more molecules have field points, which are by construction the most important places. The advantages over CoMFA are that you get many fewer sample points than with grid-based sampling, there are no gauge variance problems, the sample positions are physically chosen rather than arbitrarily chosen, and there is a consistent framework for both alignment and QSAR.



Retrospective Testing

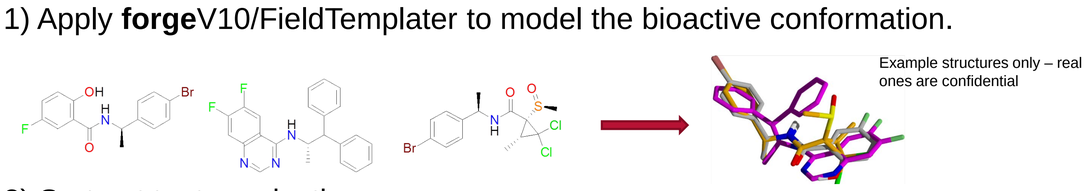
Tested against literature CoMFA datasets:

- 15 datasets with alignments available.
- CoMFA average cross-validated RMSE is 0.72.
- Field QSAR using CoMFA alignments is 0.74.
- **Simple model (volume indicator variable – 1 for "inside", 0 for "outside") is 0.83!**

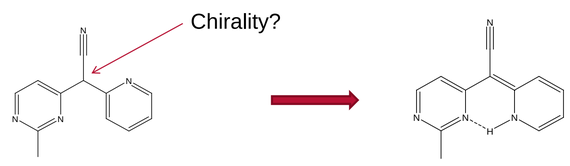
Over a wide variety of data sets we have found that volume indicator variables are uniformly superior to vdW potentials both in terms of model performance and in terms of interpretability. Addition of a hydrophobicity potential does not generally improve models.

Application to NK3

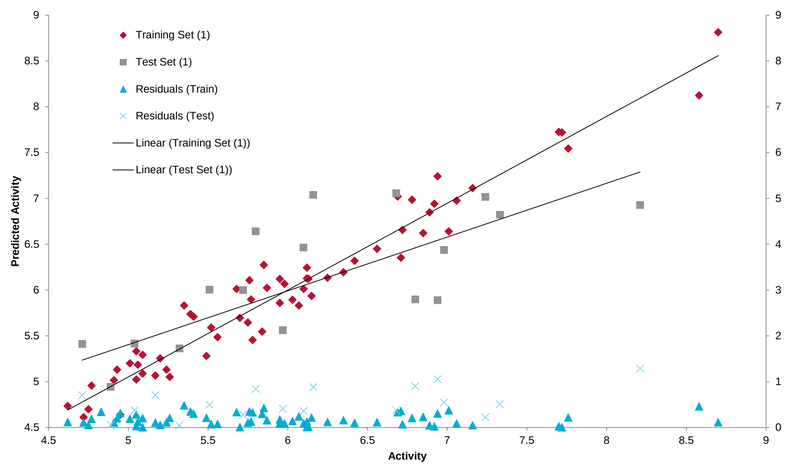
1) Apply **forge**V10/FieldTemplater to model the bioactive conformation.



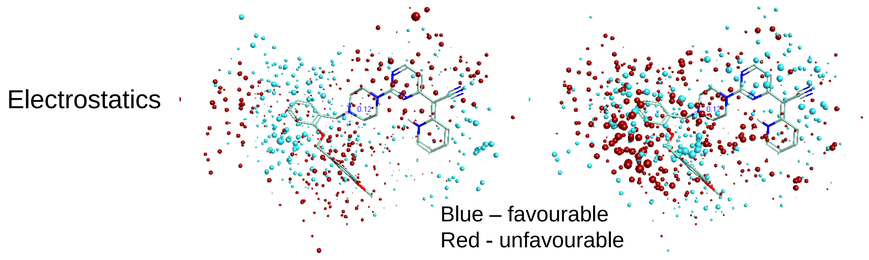
2) Sort out tautomerisation.



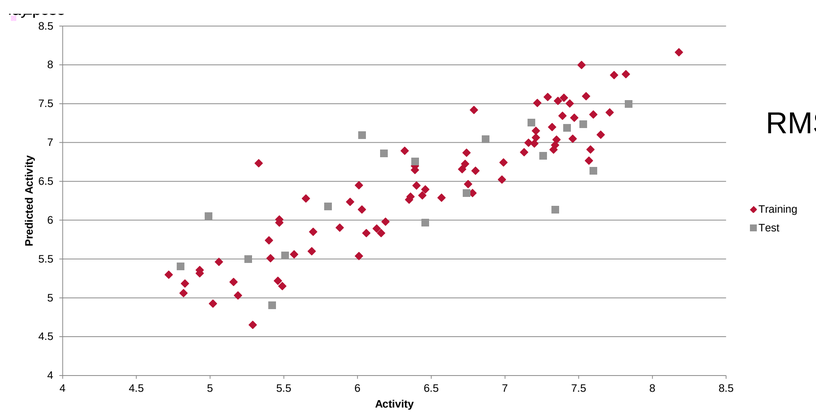
3) Align all molecules to the template and calculate descriptors.
4) Apply PLS.



RMSE 0.19, PRESS 0.51, $RMSE_{pred}$ 0.64



5) Repeat on different series.



$RMSE_{pred}$ 0.60

Conclusions

QSAR using molecular field sampling has important advantages over grid-based sampling methods. No parameterization of grid size or spacing needs to be performed, the method is gauge invariant and reproducible, and edge effect are reduced as the sample positions are placed at the most important regions of space by construction.

New software implementing this technique, '**forge**V10', is now available from Cresset.

References

[1] Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model*. **2006,** 46, 665-676.

Acknowledgements

Thanks to Euroscreen for the use of the NK3 data.