

Interpretable Activity Models: exploring the limits of pharmacophores and 3D QSAR methods

D.A. Evans¹

D.A. Thorner¹ & M.J. Bodkin¹

¹ *Eli Lilly & Co. Ltd., Lilly Research Centre, Windlesham, UK*

QSAR modeling in *pharma* is often focused on the rapid production and systematic updating of models based upon large numbers of easy to calculate molecular descriptors or fingerprints whose interpretation can be difficult. Such methods typically output either a predicted activity class or a numerical value with a significant uncertainty. This makes them most useful in the early stage of a drug discovery project to screen existing compounds, prioritize library design options or check for potential off-target activities. In the later lead optimization phase significant activity data on a particular compound series already exist and medicinal chemists tend to be searching for a systematic understanding of the SAR. Activity models which are interpretable by a medicinal chemist are more useful in deciding ‘what to make next?’

Recent validation studies on QSAR methods have demonstrated that equivalent or superior performance can be obtained with more interpretable 3D QSAR methods such as CoMFA compared to selected 2D methods.¹ The atom based grid method Phase was also demonstrated to be generally superior in performance to the pharmacophore alignment method Catalyst Hypogen.² The work to be reported aims to establish the viability of using pharmacophore-based alignment to produce 3D QSAR models from a project SAR of potentially thousands of compounds in an automated fashion. We compare the performance of Phase and CoMFA and the quality of QSAR models generated from manual and pharmacophore based alignments and also rank these against the now commonly used fingerprint-based support vector machine (SVM) models. Interestingly, we present results on the variation of predictive QSAR performance with training set size and investigate the observed performance plateau.

The use of automated pharmacophore overlays become computationally demanding when hundreds of molecules need to be considered simultaneously. We address this problem by pre-clustering the data set using 2D methods before building pharmacophore models within and across each cluster and combining the derived pharmacophores into a minimal set. This approach aims to answer the common question of whether compounds of two chemical series have the same binding mode and hence transferable SAR, by testing both whether they have the same pharmacophore and whether the combined series can produce a good quality QSAR model.³ We will discuss the delivery of the approach as a desktop tool for medicinal chemists

1. Sutherland, J.J.; O'Brien, L.A.; Weaver, D.F. A Comparison of Methods for Modeling Quantitative Structure-Activity Relationships. *J. Med. Chem.*, **2004**, 47, 5541 -5554,
2. Evans, D.A.; Doman, T.N.; Thorner, D.A.; Bodkin, M.J. 3D QSAR methods: Phase and Catalyst Compared. *J. Chem. Inf. Model.*, **2007**, 47, 1248 -1257.
3. Iyer, M.; Hopfinger, A. J. Treating Chemical Diversity in QSAR Analysis: Modeling Diverse HIV-1 Integrase Inhibitors Using 4D Fingerprints. *J. Chem. Inf. Model.*, **2007**, 47, 1945 -1960.