

Why you should read Dr. Cramer's perspective

Yvonne Connolly Martin

Received: 24 January 2011 / Accepted: 25 January 2011 / Published online: 6 February 2011
© Springer Science+Business Media B.V. 2011

This issue contains a Perspective article by Dr. Richard Cramer, who is known by all as the inventor of the popular 3D QSAR method CoMFA. A December 2010 search of Google Scholar identified approximately 3,000 articles that discuss or use CoMFA. Hence, it is of great interest to read Dr. Cramer's perspective of QSAR as he reflects on the results from the 22 years since the original publication of the method and his more than 30 years experience with QSAR.

Although the Perspective focuses on propriety software, the results have implications beyond that software, indeed beyond QSAR and CoMFA. The most important lesson is the realization that there is a distinction between a predictive QSAR model and molecular modeling to investigate physical reality. Any QSAR model must emphasize the differences in the ligand structure that lead to differences in biological potency. If these differences are correctly encoded, the QSAR will be predictive. Because traditional QSAR meets the criterion of focusing on differences in ligand structures, this may explain its continued success. Its problem can be to generate accurate descriptions of the molecules.

Dr. Cramer shows that considering the 3D properties of the ligand structures is a two-edged sword. Traditional CoMFA and related methodologies base their fits and predictions on the 3D locations of hydrogen bonding and steric occupancy. This improvement over traditional 2D QSAR is complicated by the necessity to choose a conformation for each molecule and how to align the molecules. Whereas these 3D properties may be more realistic, one may be tempted to generate a realistic model of the

binding. It often happens that there are subtle differences in the low-energy 3D structures of molecules that are similar in 2D. Because these differences are the direct result of changes in the structures of the ligands, using these changes to suggest the potency of untested molecules adds noise to the relationship between changes in the ligand structures and changes in their bioactivity.

The Perspective starts with a summary of unpublished results: specifically, the accuracy of Topomer CoMFA potency forecasts. In the four known cases in which the method was applied, the potency of the 140 newly synthesized molecules was predicted within 0.5 log units. In every series the accuracy of the predictions was greater than that from leave-one-out cross-validation. To understand the significance of these results, it is important to recognize that Topomer CoMFA discards the idea of generating molecular descriptors from low energy conformations of whole molecules as superimposed: Instead, the molecules are decomposed into molecular fragments, one fragment per position on the common "core", and aligned on the bond that attaches the fragment to the core. Importantly, the 3D conformation of each fragment is not chosen to be a low-energy one, but rather to follow a pre-defined heuristic. The QSAR analysis uses these fragments in the traditional CoMFA paradigm of calculating electrostatic and steric fields and analyzing the results with PLS.

Dr. Cramer emphasizes the relentless focus of Topomer CoMFA on changes in the structure of the small molecule as the sole reason for their change in biological potency: it is changes in the structure of the small molecule that is responsible for changes in the observed potency, whether or not the changes in the 2D structure of the small molecule leads to changes in the 3D structure of the target biomolecule or even in the small molecule as compared to other

Y. C. Martin (✉)
Waukegan, IL, USA
e-mail: yvonnecmartin@comcast.net

analogues. The accuracy of the potency forecasts from Topomer CoMFA may be partly due to the fact that the method allows no subtlety in the alignment; small differences in location of a key hydrogen bonding group disappear and the steric boundaries are more clearly defined. Such a focus is missing in traditional CoMFA.

The predictivity of the Topomer CoMFA models prompts a re-examination of the fundamental assumptions about how a 2D or 3D QSAR should be generated. Specifically, the results challenge the notion that CADD

modeling should include considerations of the binding conformation, because the conformations used in Topomer CoMFA are often very unrealistic.

Although it is possible that equal predictivity may result from other strategies that include a strict focus on the variable portions of the molecules, this remains to be demonstrated. In particular, it seems worthwhile to investigate the predictivity of 2D QSARs in which the substituents at different positions of the core are treated independently.