3D-QSAR CONTINUED

Limitations, Challenges, Opportunities for the Future Application of 3D QSAR

1. Choosing the bioactive conformation and alignment

Many of the 3D QSAR methods discussed require that the chosen conformations of the molecules be aligned before the software develops the quantitative model; other methods select a conformation and an alignment as part of the development of the model. Usually, one assumes that the conformation used should be the best assessment of the bioactive conformation and, furthermore, that the alignment represents how the different molecules bind to the target macromolecule. In fact, a 3D QSAR model simply provides a summary of how changes in the structure of the ligand affect its affinity for a target molecule. Furthermore, in many cases, either multiple binding modes of the same compound or closely related compounds have been observed crystallographically and could be expected for many of the series studied by 3D QSAR. Consider a 3D QSAR model that suggests that increased affinity results from added steric bulk (or electronegative group) at a certain position with respect to the groups used for the alignment. A simple explanation would be a hydrophobic (or electropositive) pocket accessible in the given alignment, whereas the true one might be that this steric bulk (or electronegative group) leads to favoured binding in an alternative orientation. Although one would expect that alignment of ligands based on minimizing the structures of the corresponding ligand–macromolecule complexes would produce the most robust 3D QSAR models, several groups have found this not to be the case.

This is probably a reflection of the uncertainties in the structure minimization programs. However, as noted above, the structure of the macromolecular binding site does provide a starting point for choosing the bioactive conformation and alignment. If one has no structure of the macromolecular target but yet has decided to use a method that needs at least a starting orientation and conformation of every molecule, then either manual molecular modelling or automated pharmacophore mapping tools will be needed; along with advances in 3D QSAR, recent years have produced advances in these techniques as well. However, no computer program can substitute for good structure–activity data. A pharmacophore mapping exercise can be expected to be successful if there is one relatively rigid active compound or several somewhat rigid compounds that collectively restrict the common distances between key recognition atoms or site points. A truly complete study would involve synthesis and testing of such molecules before a pharmacophore and a 3D QSAR study was undertaken. There have been a number of interesting suggestions of ways to improve the alignment of molecules. Usually these are applied once one has chosen the bioactive conformation or a preliminary model. The downside of these strategies that modify alignment or conformation to improve fit or predicted activity is that one must become increasingly alert to the possibility of deriving a chance model. With the receptor surface strategy, it is suggested to optimize the structures of the less potent compounds within the model receptor surface generated from the three or four most potent compounds. This could lead to very distorted structures of molecules that in a CoMFA analysis penetrate into negative steric regions. Investigating alternative alignment strategies should certainly be an area of active research; hopefully, more analysis of the reliability of the forecasts that result from different strategies will provide definitive guidelines for future work. CoMMA, EVA or the WHIM descriptors

promise an advantage because they provide 3D descriptors that are independent of the orientation of the molecules in space; they do not have to be aligned. However,we need to remember that the CoMMA inertial, dipole, and quadrapole moments are sensitive to conformation, as are most of the WHIM descriptors. The best way to find corresponding conformations in a set of molecules is to align them with each other, so one does not totally escape the alignment problem. However, the CoMMA and WHIM descriptors are less sensitive to exact conformation than are lattice-based energy values used in CoMFA and related methods. The EVA descriptors appear to be even less sensitive to conformation. This is somewhat adjustable within a run; sometimes the lack of sensitivity to conformation occurs at the expense of statistical quality of the model. A philosophical issue arises: if a method is insensitive to the 3D structure, the conformation, of a molecule, is it really a 3D QSAR method? Clearly, there are opportunities to continue to explore the role these and other alignment-free methods will play in QSAR analyses.

2. Choosing the type of descriptors

Many investigations are being carried out for an alternative molecular descriptor for 3D QSAR. For lattice-based methods, there is now evidence that hydrophobic fields do not generally increase the statistical quality of the model, that steric fields can profitably be replaced with somewhat softer functions and that electrostatic fields based on semiempirical electrostatic potentials are superior to empirical schemes. The CoMSIA descriptors appear to contain the same information as those of traditional CoMFA but produce contour plots that are easier to transform mentally into molecules to synthesize. Several groups have proposed 3D QSAR methods that are not based on properties calculated at a lattice. The GERM, COMPASS and receptor surface methods rely on properties calculated at discrete locations in the space at or near the union surface of the active molecules, presumably a model of the macromolecular binding site. If all molecules of the set do bind in a manner that doesn't distort the binding site too much, this can be a reasonable strategy as evidenced by the fact that these methods have led to the development of reasonable models. However, in series for which there is a large positive contribution of steric energy at certain points, as in the case of our D1 dopaminergic agonists, this type of descriptor might not be able to detect that the *absence* of steric bulk at a certain point leads to a decrease in potency. Both of these methods base their 3D QSAR on interaction energies with the hypothetical receptor and, hence, are subject to all the limitations of such interaction energies, even when the structure of the target macromolecule is known. The positive feature of these two methods is that the model is presented as a 3D display of properties of the receptor in space. The EVA, CoMMA and WHIM descriptors differ from the lattice- or surface-based descriptors, in that they do not consider properties at locations in space, but rather 3D properties of the molecules themselves. Hence, it is not possible to provide a 3D display of the resulting models.

3. Designing the series and choosing the training set

Within the CoMFA paradigm, some attention has been paid to the design of series for 3D QSAR analysis. For example, one might generate a number of principal components from the steric and electrostatic fields of the aligned molecules and cluster the molecules based on these descriptors. Alternatively, one might choose to use steric field descriptors suited to substituents. However, today most models are derived from datasets that were not designed

for 3D QSAR analysis. A particular concern is that, in poorly designed series, electrostatic and steric properties are not varied independently, nor are they varied continuously. Although good statistical models may result, their predictivity may be low if the new compounds break the correlations in the training set. The use of 3D QSAR or related descriptors in series planning represents an opportunity to help the medicinal chemist synthesize fewer and better distributed compounds for the derivation of the first QSAR model, or to select substituents for combinatorial libraries. Sometimes it happens that there are too few active compounds to derive a CoMFA model, even one based on active versus inactive sets. In that case, simply designing compounds that are similar to the active ones but different from the known inactives in one or more dimensions might lead to the identification of more active compounds. There is also evidence that one can derive 3D QSAR models of equivalent or better quality by considering a carefully selected subset of the compounds in the dataset and that such models are more robust and provide more accurate forecasts of affinity. Accordingly, for retrospective analyses, it appears advantageous to select a training subset of all compounds tested and to use the remaining compounds as a biased test set.

4. Selecting variables for the model

CoMFA requires that one considers thousands of 3D descriptors rather than the small number used in traditional QSAR. Even after discarding descriptors that do not vary significantly in the data set, there are often thousands remaining. Additionally, there is the conflict between using many lattice points to produce more accurate energy values (smaller lattice spacing) and the notion of keeping the number of variables low (larger lattice spacing) to reduce the noise in the models. Since PLS is very sensitive to noise in the descriptors, more predictive models should result if we could eliminate unnecessary descriptors. Experiences with HASL and genetic PLS suggest that for typical CoMFA models the energy at only a very few points explains most of the variance in biological potency. Models derived with the steroid dataset using different approaches reinforces this point since several of the methods use very few descriptors to provide the same level of statistical quality. Similarly, traditional QSAR provides equations in very few variables. However, in spite of the promise of cross-validated R2-guided region selection and GOLPE-guided region selection, it is too early to tell if variable reduction based on preliminary QSARs lead to models with better ability to forecast the potency of new compounds. The same problem might apply to genetic selection based on cross-validation. Again, it is to be expected that variable selection for 3D QSAR will continue to be an area of active research just as it is currently in traditional QSAR and other lower-dimensional problems.

5. Deriving the model

For those methods that use only a few descriptors or that calculate a single interaction energy to be correlated with biological potency, multiple linear regression is a suitable method. However, if several variables are considered for possible inclusion in the model, it is all too easy to overfit a regression equation, suggesting a preference for partial least squares, PLS, modelling instead. Although the simplicity of LS is a positive attribute, its modelling power decreases when noise is mixed with the relevant descriptors. Additionally, a PLS model is linear in the descriptors, although quadratic PLS identifies certain nonlinear relationships. Hence, there is considerable interest in finding new methods to establish the relationship between (selected) 3D descriptors and biological potency. However, one should

be aware that the deficiencies of PLS may be more noticed only because so much more attention has been devoted to PLS, and that alternative methods may suffer from the same problems. Nonlinear relationships can be detected by the PLS analysis of a transformation of the original data matrix into a matrix of the distances between each pair of observations as measured in the original property space. A problem with using this approach with CoMFA fields is that there is no obvious way to display the nonlinear relationship on the CoMFA lattice. Another problem is that including irrelevant descriptors in the distance calculation can weaken the nonlinear signal. Several chapters in this volume report modelling with neural networks. This is another area that deserves more attention to establish the conditions for reliable 3D QSAR model development.

6. Validating the model

The primary test of any model is how well it forecasts the potency of compounds not used in its derivation, typically a test set reserved for this purpose. Less common, but to be recommended, is to repeat the model derivation on different subsets of the data to test for the consistency of the models produced. Despite all the caution one uses, it is all too easy to overfit the training set data Hence, it is becoming common to scramble the biological data, often many times, and repeat the variable selection and model generation procedure. This randomization procedure preserves the correlations between the predictor variables and the distribution of the potency while breaking any true relationship between them. It is becoming clear that the cross-validated $R2$ is not a good measure of the quality of a 3D QSAR method, particularly if variable- or alignment-selection strategies have been used. A further complication with this statistic is that it is sensitive to the composition of the dataset: if there are many near-duplicates, then the cross-validation will indicate a robust model, whereas it will indicate no or a poor model if the dataset has been consciously designed to include no similar compounds. Larger datasets, usually preferred by QSAR modelers, have a larger chance of containing many near-duplicates If the 3D structures of the target macromolecule become available after the QSAR determination, then one can compare it with the 3D QSAR model.

7. Forecasting potency

Most forecasts of potency from 3D QSAR models are simply a value with no estimate of reliability, except the cross-validated root mean square error. However, it is important to know if the test compound is very different from every molecule in the training set and, hence, that its potency forecast is much less accurate than one for which a very similar molecule is in the training set. The use of molecular similarity to align molecules for potency forecasts suggests that all 3D QSAR forecasts should also include how similar the test molecule is to one in the dataset. The similarity should be calculated over all the properties considered for the model, rather than for those properties that were found important for the model, since if a new compound changes a property that was not previously changed, then no QSAR model can be expected to give reliable forecasts. There is no perfect way to summarize the accuracy of potency forecasts, because each method depends on the distribution of potency in the test set. Consider two QSAR methods: the first predicts only fairly accurately but consistently under-predicts potent compounds and over-predicts less active ones, whereas the second method predicts each compound more closely but has no

such bias. For datasets that contain most compounds at the extremes of activity, the former will have a higher *R*2pred, even though the slope between observed and forecast is not 1.0. On the other hand, for datasets in which all compounds have potency near the mean, the mean unsigned error of prediction would favour the latter method. The common use of plots of observed versus forecast affinities, on the same figure or at least the same scale as a similar figure for the training set, provides a more detailed picture of the quality of the forecasts.

8. Comparing 3D QSAR methods

A serious problem in comparing methods is that often the only information provided concerns the relative precision of models derived from the same dataset with different methods, whereas what one wants to know is how well the different methods forecast the affinity of new compounds. In particular, the comparison of methods must deal with the perception that at least some variable-selection methods provide optimistic cross-validation estimates of model accuracy and that feedback neural networks may overfit a model. Compounds to consider for true potency forecasting may be hard to find, and it is tempting to include all known molecules in the development of a model or when statistically selecting those to include and those to predict. Although most new methods provide a result on a reference set of compounds, errors of many sorts can confound these comparisons. Furthermore, it is possible that some methods are unintentionally tuned to the test datasets and will perform less well with other data. Until benchmark studies are done, how does one choose which method to use? Frequently, the choice depends on the software available. However, if no satisfactory quantitative relationship is found, one must decide if another method will be successful.

Role of 3D QSAR in Combinatorial Chemistry and High-throughput Screening

1. Generating 3D QSARs and forecasts quickly

The modern pharmaceutical industry has embraced two strategies that were just emerging a decade ago, when CoMFA was devised: mass or high-throughput screening hundreds of thousands of compounds in a particular assay and synthesis and testing of mixtures of compounds. In view of its success in small sets of compounds, it would be an important contribution if 3D QSAR could contribute to the success of these ventures. In industry today, computational chemists often participate in the design of targeted combinatorial libraries that can include any of millions of compounds. A QSAR method that could efficiently forecast the potency of so many compounds would be very attractive, even if it were less accurate than more time-consuming methods. Yet another challenge is to develop QSAR models based on high-throughput screening of thousands of compounds with associated errors in structure. The first challenge to basing a 3D QSAR model on high-throughput screening or screening of combinatorial libraries will be to establish the validity of the structures actually tested. Typically, the success of the chemistry to produce combinatorial libraries is measured only in rehearsal runs and on compounds identified as active. Similarly, the identity of the structures of the compounds in collections is often assessed only when activity has been identified. In both cases, the modeler cannot be assured that certain compounds are not active because there is a small chance that they have not been tested. This ambiguity suggests that methods that tolerate ambiguity might find application in this context. The second challenge to developing a QSAR based on high-throughput screening is that often the biological

activities are simple active versus inactive. Hence, the PLS variant of discriminant analysis or a neural network method might be useful. Since there are usually 10–1000 times more inactive compounds than active ones, a clever strategy to select only a subset of the inactive compounds for model development will conserve considerable time. A third challenge is for the computer to be fast enough to complement high throughput screening methods or SAR by NMR for the identification of novel existing compounds to fit a target of known 3D structure. A final challenge is that the QSAR modelling must be done quickly. Often, not only must a QSAR be derived, but new compounds for combinatorial synthesis must be designed within a matter of a week or two. This challenge means that any QSAR method used must be robust without human valuation of the results. The positive aspect is that the QSAR need not be especially reliable since any enrichment of active compounds in a second library will improve the efficiency of the search for new compounds. It is an open question whether a traditional or 3D QSAR approach will be more useful in this context.

2. Designing, diverse combinatorial libraries

The success of 3D QSAR in predicting the affinity of new compounds suggests that this type of descriptor has relevance to biological properties of molecules. Accordingly, some have based their selection of substituents for combinatorial libraries on 3D fields. A positive aspect of combinatorial library synthesis is that often there are more potential compounds that can be made than will actually be made. The result is that the computational chemist can influence the decision of which compounds to make and design a set that should lead to an interpretable QSAR.

Conclusion

All evidence suggests that 3D QSAR techniques will continue to make a valuable contribution to the computer-assisted analysis of structure–bioactivity relationships. The search for new descriptors of 3D properties of ligands and innovative strategies to investigate the relationships between these properties and bioactivity continues to be a fruitful research enterprise. Increasing information from structural biology will provide valuable feedback to the hypotheses that form the basis of 3D QSAR methods. 3D QSAR methods complement traditional QSAR based on physical properties. They offer the advantage that it is easy to calculate descriptors for most molecules, and the disadvantage that one must select a conformation and usually a superposition rule as part of the analysis. Because of their speed and accuracy, 3D QSAR methods complement calculations based on the structure of the ligand–macromolecular complex. Whereas the structure of at least one complex aids in the selection of the bioactive conformation and the alignment of the molecules for 3D QSAR, a QSAR model can be derived much more quickly than calculations based on the complex. Frequently, it is just as predictive. Knowledge of the structure of the complex can also prevent unwarranted extrapolation from a QSAR model. It is expected that concepts from 3D QSAR will continue to impact the analysis of high-throughput screening structure–activity data and the diversity of compound collections and combinatorial libraries.

References

1. Kim, K.H., Greco, G. and Novellino, E., *A critical review of recent CoMFA applications*, In Kubinyi, H., Folkers, G., and Martin, Y.C., (Eds.) 3D QSAR in drug design: Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 257–316.
2. Dunn III, W.J. and Hopfinger, A.J., *3D QSAR of flexible molecules using tensor representation*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 167–182.

3. Hahn, M. and Rogers, D., *Receptor surface models*, in Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 117–134.
4. Heritage, T.W., Ferguson, A.M., Turner, D.B. and Willett, P., *EVA — a novel theoretical descriptor for QSAR studies*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 381–398.
5. Klebe, G., *Comparative molecular similarity indices analysis — CoMSIA*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 87–104.
6. Walters, D.E., *Genetically evolved receptor models (GERM) as a 3D QSAR tool*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 159–166.
7. Wade, R.C., Ortiz, A.R. and Gago, F., *Comparative binding energy analysis*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 19–34.
8. Holloway, M.K., *A priori predicition of ligand affinity by energy minimization*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 63–84.
9. Todeschini, R. and Gramatica, P., *New 3D molecular descriptors: The WHIM theory and QSAR applications*, In Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 355–380.
10. Silverman, B.D., Platt, D.E., Pitman, M. and Rigoutsos, I., *Comparative molecular moment analysis (COMMA)*, in Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in drug design: Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998, pp. 183–196.
11. Jain, A.N., Koile, K. and Chapman, D., *Compass: Predicting biological activities from molecular surface properties — performance comparisons on a steroid benchmark*, J. Med. Chem., 37 (1994) 2315–2327.
12. Martin, Y.C., Kim, K.-H. and Lin, C.T., *Comparative molecular field analysis: CoMFA*, In Charton, M. (Ed.) Advances in quantitative structure property relationships, JAI Press, Greenwich, CT, 1996, pp. 1–52.
13. Greco, G., Novellino, E. And Martin, Y.C., *Approaches to 3D-QSAR*, In Martin, Y.C. and Willett, P. (Eds.) Designing bioactive molecules: Three-dimensional techniques and applications, America Chemical Society, Washington, DC, 1997 (in press).
14. Ajay and Murcko, M.A., *Computational methods to predict binding free-energy in ligand–receptor complexes*, J. Med. Chem., 38 (1995) 4953–4967.
15. Kollman, P.A., *Advances and continuing challenges in achieving realistic and predictive*

simulations of the properties of organic and biological molecules, Acc. Chem. Res., 29 (1996) 461–469.
16. Bush, B.L. and Nachbar Jr., R.B., *Sample-distance partial least-squares — PLS optimized for many variables, with application to CoMFA*, J. Comput.-Aided Mol. Design, 7 (1993) 587–619.
17. Burger, A., *Medical chemistry — the first century*, Med. Chem. Res., 4 (1994) 3–15.
18. Willett, P., *Similarity and clustering techniques in chemical information systems*, Research Studies Press, Letchworth, 1987.
19. Hodgkin, E.E. and Richards, W.G., *Molecular similarity based on electrostatic potential and electric field*, Int. J. Quantum Chem., 14 (1987) 105–110.
20. Kier, L.B., *Molecular orbital theory in drug research*, Academic Press, New York, 1971, p. 258.