Journal of Computer-Aided Molecular Design **18:** 587–596, 2004.
DOI 10.1007/s10822-004-4068-0

3D-QSAR illusions

Arthur M. Doweyko
Department of Macromolecular Structure, CADD, Pharmaceutical Research Institute, Bristol-Myers Squibb, Princeton, NJ 08543, USA; Fax: +1-609-252-6030; E-mail: arthur.doweyko@bms.com

Received 30 April 2004; accepted in revised form 3 September 2004

Key words: CoMFA, HASL, QSAR

Summary

3D-QSAR is typically used to construct models (1) to predict activities, (2) to illustrate significant regions, and (3) to provide insight into possible interactions. To the contrary, examples are described herein which make it clear that the predictivity of such models remains elusive, that so-called significant regions are subject to the vagaries of alignment, and that the nature of possible interactions heavily depends on the eye of the beholder. Although great strides have been made in the imaginative use of 3D-descriptors, 3D-QSAR remains largely a retrospective analytical tool. The arbitrary nature of both the alignment paradigm and atom description lends itself to capricious models, which in turn can lead to distorted conclusions. Despite these illusionary pitfalls, predictions can be enhanced when the test set is bounded by the descriptor space represented in the training set. Interpretation of significant interaction regions becomes more meaningful when alignment is constrained by a binding site. Correlations obtained with a variety of atom descriptors suggest choosing useful ones, in particular, in guiding synthetic effort.

Introduction

3D-QSAR is an all-encompassing umbrella term which covers a wide variety of correlative methodologies, all of which are designed to go beyond the classical Hansch–Fujita [1] and Free–Wilson [2] analyses of 40 years ago to explore the brave new world of CPU-assisted drug design. With the advent of DYLOMMS [3] and PLS [4], CoMFA [5] emerged as the 3D-QSAR method most embraced by the scientific community today. In addition to CoMFA, there are a large number of other 3D-QSAR technologies, some of which rely on alignment (e.g., CoMSIA [6] and HASL [7]) and others which do not (e.g., COMMA [8], EVA [9] and VOLSURF [10]). Each 3D-QSAR method represents an attempt to correlate the observed activities for a series of molecules (e.g., enzyme inhibition or receptor binding) with the differences observed in their three-dimensional molecular structure. This is done with or without an alignment paradigm, utilizing a descriptor set based on molecular coordinates. Often, the objective of this exercise can include any of three generalized targets: (1) to predict activities of as-yet untested molecules, (2) to illustrate regions in or around a template molecule which may have significant effects on activity, and/or (3) to provide insights into interactions between the molecule and its protein target that may be significant.

Although the underlying theme of 3D-QSAR methods is prospective, i.e., it is desired that the model can predict the activities of new molecules, the single most obvious trait of such models is their retrospective nature. Attempts to gauge the 'predictivity' of models using popular techniques such as 'leave-one-out' (LOO) are likely to be misleading. Alignment-dependent 3D-QSAR approaches have very obvious issues regarding 'alignment rules' and related concerns which can lead to spurious models [11]. These are only part of the problem, in that different assumptions in how

Table 1. 3D-QSAR model statistics – literature values.

Method	Training set	r^2	q^2	Test set	r^2(pred)	Ref.
BCUTS	62	0.87	0.84	10	0.99	36
BCUTS	159	0.95	0.94	20	0.99	36
CoMFA	27	0.98	0.61	9	0.99	42
CoMSIA	27	0.97	0.53	9	0.99	42
HQSAR	27	0.95	0.72	9	0.99	42
HQSAR	157	0.86	0.72	20	0.89	22
CoMSIA	39	0.97	0.71	16	0.78	15
CoMFA	191	0.88	0.72	15	0.78	22
CoMSIA	21	0.94	0.67	10	0.76	16
CoMFA	22	0.92	0.56	11	0.76	41
CoMFA	44	0.97	0.80	9	0.76	45
CoMSIA	22	0.88	0.56	11	0.75	41
CoMFA	21	0.95	0.60	10	0.73	16
CoMSIA	32	0.92	0.75	6	0.72	39
EVA	21	0.96	0.80	10	0.69	24
CoMFA	40		0.57	16	0.69	37
HASL	40		0.93	16	0.66	33
EVA	44	0.79	0.46	9	0.66	45
EVA	44	0.95	0.54	9	0.66	45
EVA	135	0.96	0.68	76	0.65	35
GRID+	40		0.62	18	0.62	31
CATALYST	19	0.88		20	0.62	47
CATALYST	19		0.72	12	0.61	46
GRID/GOLPE	67	0.73	0.63	9	0.58	28
CoMFA	32	0.88	0.68	6	0.56	39
CoMFA	51	0.89	0.64	53	0.55	18
CoMSIA	28		0.57	4	0.55	40
CoMFA	52			14	0.52	46
CoMFA	26	0.97	0.52	28	0.51	17
CoMFA	59	0.98	0.78	36	0.49	13
CoMFA	59	0.99	0.61	36	0.47	13
CoMFA	59	0.98	0.64	36	0.45	13
HASL	54	0.75		13	0.44	33
EVA	44	0.98	0.58	9	0.43	45
CoMFA	59	0.99	0.65	36	0.40	13
CoMSIA	24	0.97	0.69	23	0.40	20
CoMFA+	49	0.84	0.82	11	0.40	21
CoMSIA/SEAL	21	0.94	0.67	10	0.40	30
CoMFA/SEAL	21	0.95	0.60	10	0.36	30
CoMFA	67	0.90	0.56	20	0.31	48
CATALYST	20	0.95		57	0.28	23
MS-WHIM	21			10	0.28	25
CoMFA	54	0.80	0.61	13	0.27	32
4D-QSAR	67	0.85	0.76	20	0.25	48
CoMFA	12	0.94	0.76	12	0.21	14
SOMFA	21	0.83		10	0.19	26
CoMFA	67	0.92	0.59	20	0.16	48
COMPASS	21			10	0.15	27
CoMFA	21			10	0.13	25
COMMA	21	0.94		10	0.13	44

Table 1. Continued.

Method	Training set	r^2	q^2	Test set	r^2(pred)	Ref.
CoMFA	59	0.99	0.66	36	0.11	13
CoMFA	25	0.99	0.74	4	0.11	19
CoMFA	24	0.97	0.42	23	0.11	20
CoMFA	171	0.98	0.84	24	0.10	38
CoMFA	33		0.93	17	0.10	12
CoMFA	23	0.99	0.43	6	0.09	43
4D-QSAR	29	0.93		10	0.04	29
CoMSIA	23	0.99	0.41	6	0.02	43
CoMFA	15	0.97	0.54	6	0.01	34
HASL	15	0.99	0.85	6	0.01	34
HASL	171	0.95	0.92	24	0.01	38
Averages	48	0.93	0.67	17	0.46	

alignments are obtained will lead to significantly different models. Lastly, with the advent of sophisticated protein homology modeling and available protein crystal structure information, it is often the hope that a 3D-QSAR model may provide insight into which intramolecular interactions are important and are consistent with a proposed binding mode. This hope represents the final and most alluring of the three illusions to be discussed.

Table 2. 17-Point model statistics.

SD	r^2(average, 10 trials)
0.7	0.59 (0.49–0.71)
0.8	0.54 (0.36–0.73)
0.9	0.44 (0.19–0.75)
1.0	0.48 (0.14–0.80)
1.1	0.41 (0.03–0.61)
1.2	0.35 (0.04–0.61)

Illusion 1. Prediction

In an attempt to get an idea of the predictive nature of 3D-QSAR models, 61 examples were culled from 37 papers published in the last decade [12–47]. These papers were selected in a near random manner, focusing on those models for which LOO and externalized test set data were listed. In this way it was possible to construct Table 1 wherein training set r^2/q^2 (LOO) and test set r^2(pred) are reported for these 61 examples of published 3D-QSAR analysis. Based on this listing, an average model might contain 48 training set members, which have excellent internal consistency ($r^2 = 0.93$) and appear to be reasonably predictive ($q^2 = 0.67$). Typically, the test set consisted of a smaller number of compounds (average = 17) with a r^2(pred) of this externalized set equal to 0.46. The question then is, ''is this a useful model?''.

In an effort to answer this question, a virtual data set was created which represented the members of the average 3D-QSAR test set. The data range was arbitrarily set to 3 orders of magnitude (activity = 1–4), reflecting a typical data range. A simple algorithm was written to endow the predicted activity values (which were exactly the same as the observed activity values) with a range of standard deviations (sd) in order to see what the effect would be on the resulting r^2. Table 2 lists the results of this analysis. As expected, r^2 was found to decrease with increasing sd. The data set having a standard deviation of 1.0 yielded an average r^2 of 0.48 (10 trials), which was similar to that observed for the average test set of Table 1 (0.46). Thus, the SD = 1.0 data set was plotted in Figure 1 as a representative test set result. One measure of this model's utility is to define a cutoff level of predicted activity of about 2.5 units (dotted line) as a means to choose which molecules to consider for synthesis or testing. It is clear that this model offers no predictive value, since the actual activities range from about 0.5 to 3.7, in effect spanning nearly the entire activity range. Thus, under the assumptions used in this analysis, the average 3D-QSAR model from the random sampling of this past decade offers no real predictive value.

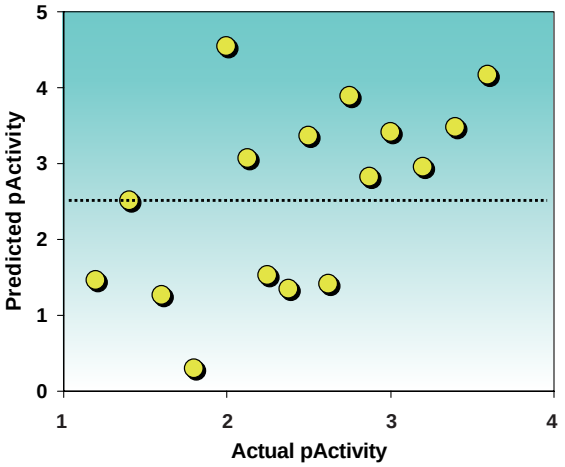

Figure 1. Typical test set prediction ($N = 17$, r^2(pred) = 0.48, SD = 1.0). Dotted line represents the cutoff level (ca. predicted activity = 2.5) for the identification of molecules of potential interest.

One of the most common measures of potential model predictivity in a training set is to conduct a LOO analysis (leave-one-out), or leave-some-number-out (typically done on larger training sets). The assumption is often made that a model with a good LOO r^2 (aka, q^2) has a better chance to predict an external test set. To test this hypothesis, the training set q^2 values obtained from the literature were plotted against their corresponding test set r^2(pred) values (Figure 2). As can be seen from the plot, there is no obvious correlation between the two values. Poor q^2 models as well as good q^2 models were found to be nicely spread between high and low test set r^2(pred), indicating that, in fact, there is no relationship between q^2 and the model's ability to predict an external test set. Such findings have been reported in the past [48–51]. This may be due to several reasons, two of which spring to mind: (1) a low q^2 value for a small training set may simply reflect the importance of each member of the training set to the model and have nothing to do with predictivity, and (2) a high q^2 may reflect redundancy in the training set, and once again, have nothing to do with predictivity. Instances where a high q^2 was associated with a high r^2(pred) may be attributed to the care taken to choose a test set that was well within the descriptor space utilized by the training set. The test set in these cases loses its independence and should actually be considered a special extension of the training set, since information from the training set was used in its selection.

Illusion 2. Significant regions

The alignment problem is well known in the 3D-QSAR field [52]. In order to illustrate the potential

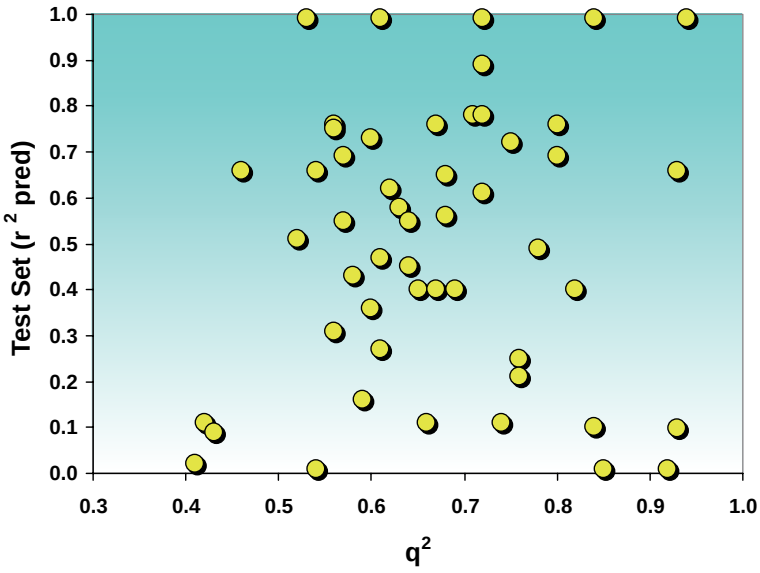

Figure 2. The relationship between q^2 and test set predictions (r^2(pred)). Data taken from Refs. [12–47].

effects of different alignment paradigms on alignment-dependent 3D-QSAR methods like CoMFA and HASL, several data sets were used. The DHFR (dihydrofolate reductase) set was based on two published crystal structures of protein-bound ligand, i.e., trimethoprim and methotrexate, as found in 1DYR and 4DFR, respectively (RCSB Protein Data Bank) [53], and 20 other structures loosely related to the 2,4-diaminopyrimidine chemotype extracted from published sources so that the range in K_i values for the complete 22-inhibitor set was 0.9×10^{-3}–1.3×10^{-11} M (pK_i 3.04–10.9) [32]. Two alignment paradigms were examined: template and MCP (Multiple Conformer Protocol) [54]. In the template procedure, the X-ray structures of TMP and MTX were used to form a static template upon which the 20 other inhibitor molecules were fitted (using the Match option within Flo [55], which forces similar atoms to overlay while accounting for internal ligand strain). The resulting overlayed structures were correlated to their DHFR inhibition (pK_i) using CoMFA [56] and HASL [57] (default settings) yielding relatively poor q^2 (LOO) values of 0.14 (1 component) and 0.33, respectively. The shape and location of the CoMFA electrostatic fields are shown in Figure 3. The same data set was subjected to MCP analysis (a HASL-based search for best conformers docked into the DHFR binding site) which yielded the binding site-derived overlays depicted in Figure 4. In this case, as expected, the resulting CoMFA and HASL models (q^2 0.31 and 0.61, respectively) were much improved and the location of the CoMFA electrostatic fields had changed significantly. These results pointed to the rather curious finding that the alignment typically considered the most conservative, i.e., template-based, surprisingly could not explain the data, while one based on binding site constraints did indeed offer a reasonable correlation.

To further examine the template and binding site alignment paradigms, a fictional set of 15 molecules was designed as possible p38 kinase inhibitors [32]. Their structures, shown in Figure 5, are based on a bicyclic core with simple changes within and around the core. Each structure was docked and minimized into a p38 binding site model using the Dockmin option within Flo (the major electrostatic interactions were set to hydrogen-bonding to Lys53 and Met109). The activity assigned to each molecule (range = 0.3–5.65) was loosely based on the calculated Flo scores. A template-based alignment (using the atoms of the six-membered ring) and a docked alignment (as provided by Dockmin) are illustrated in Figure 5 along with the resulting CoMFA electrostatic fields. Of most interest is the location of the positive electrostatic field (blue) pointing away from the Met109 interaction for the template-based series, while that for the dock-

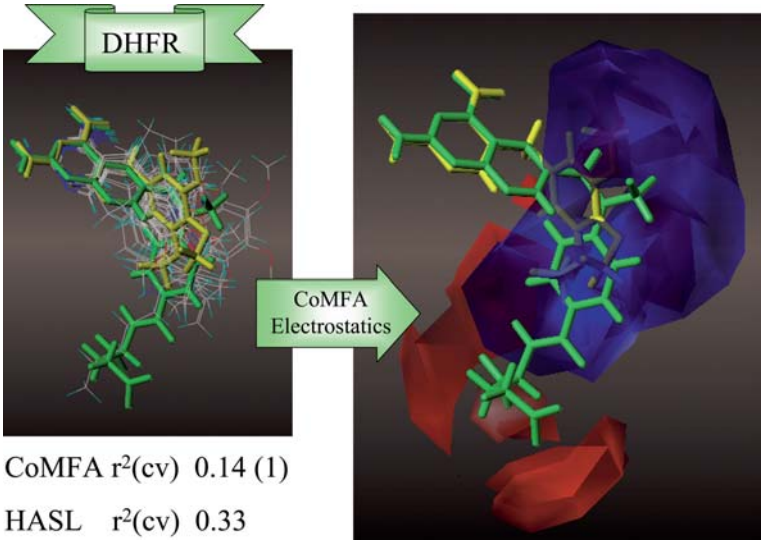

Figure 3. Template-based alignment of 22 dihydrofolate reductase inhibitors containing the 2,4-diaminopyrimidine motif, showing CoMFA electrostatic fields correlated to positive changes in activity (blue – positive charge; red – negative charge).

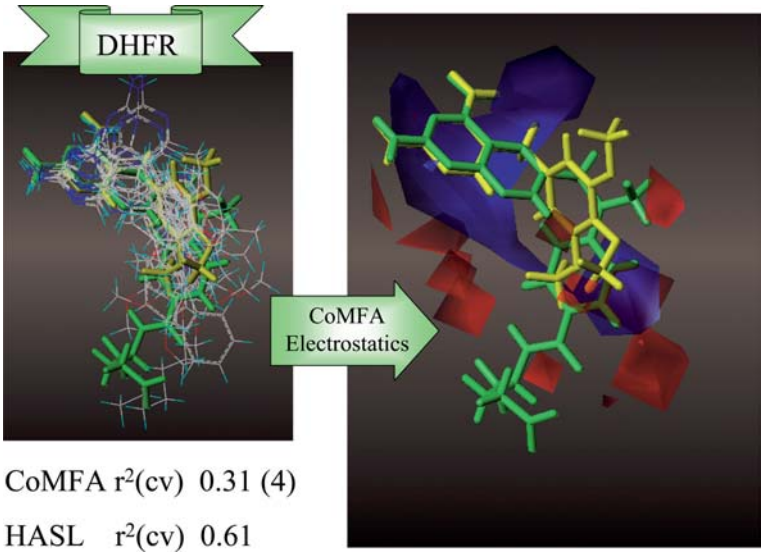

Figure 4. MCP-based alignment of 22 dihydrofolate reductase inhibitors containing the 2,4-diaminopyrimidine motif, showing CoMFA electrostatic fields correlated to positive changes in activity (blue – positive charge; red – negative charge).

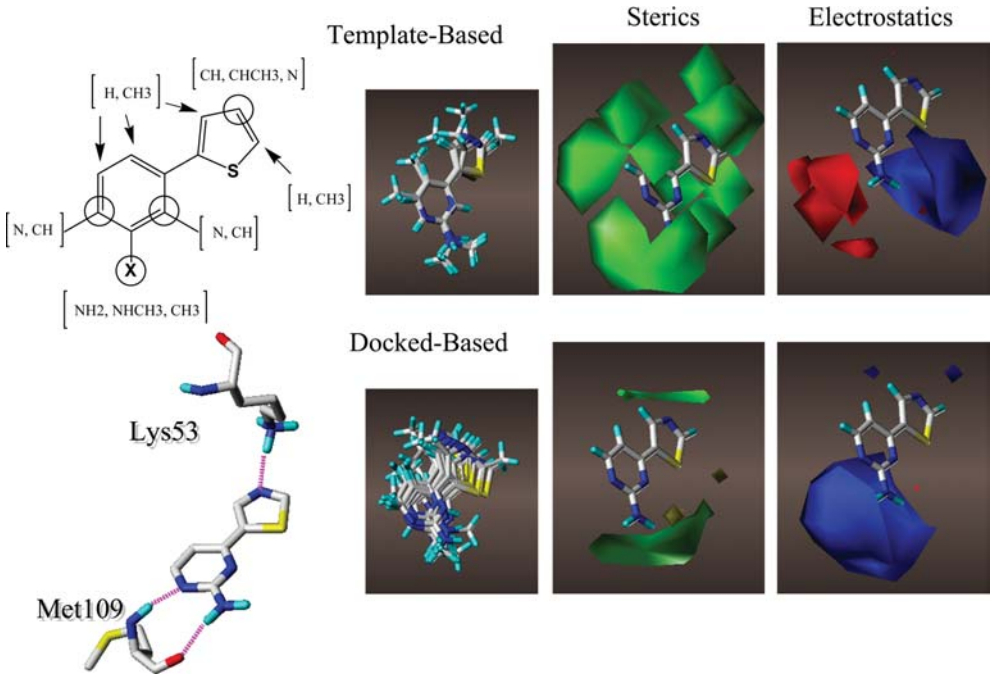

Figure 5. A comparison of template and docked based alignments for a series of 15 fictional p38 kinase inhibitors. CoMFA steric (green – positive; yellow – negative) and electrostatic (red – positive for negative charge; blue – positive for positive charge) fields are shown.

based series is pointed in the opposite direction, consistent with the expected direction of a critical ligand NH donor to Met109. Once again, depending on the alignment paradigm chosen, 3D-QSAR models can point to significantly different critical features, and the alignment based on a binding site constraint appeared to yield results consistent with the known structure.

Illusion 3. Insight

The huge number of descriptors available for QSAR analyses has made it very easy to both over-describe systems and to provide models with completely uninterpretable correlations, although that is not quite the case with all 3D-QSAR methods. The calculation of partial charges in a CoMFA analysis has often been cited as having potential consequence, with the intuitive feeling that the more accurately calculated charges would lead to better CoMFA models. As a test of this hypothesis, CoMFA models were created for five data sets using a number of different charge calculations (Table 3). The data sets were the CBG steroid set [5], an Elastase set [54], the DHFR set, a 5HT1A set [58] and a fictional set (I-Set) [59]. The values listed in Table 3 are all q^2 (LOO), with the number of components in the best model shown in parentheses. Models were created with four different partial charge calculations (Gasteiger/Hückel, Hückel, Pullman and MMFF94), as well as for one case using SAM1 [60]. It is clear from the table that the results do not vary significantly for any of the data sets, including the SAM1 version of the 5HT1A model. As a further examination of the CoMFA descriptors, models were also created using two extremes, i.e., using only a charge field (Gasteiger/Hückel) or only a steric field. The CBG and Elastase models were unaffected under these extremes, while some loss in predictivity (as measured by q^2) was observed for the other three models, suggesting that for these cases neither field alone was as good as a combination of the two.

The same data sets were subjected to HASL analyses in order to examine the effect of atom type definitions, as this methodology is based on the location of atom types in 3D space. The results of these analyses are also shown in Table 3. Using normal atom type definitions, i.e., three atom types (electron-rich, electron-poor and electron-neutral), HASL was able to provide reasonable models for all five data sets ($q^2 = 0.54$–0.83). Interestingly, when the atom type definitions were changed, models of similar predictivity were generated: zero = all atom types the same; H-bonding = O, N and polar H equivalent to each other and all others the same; random = scrambled definitions of three atom types. The observation that even scrambled definitions of the atom types resulted in statistically meaningful models was particularly of note, which suggested the following extreme. The three SLG atom types (solids, liquids and gases, at standard temperature and pressure, STP) were found to handle all five data sets reasonably well ($q^2 = 0.44$–0.77). An illustration of the predictivity of the 5HT1A SLG model is shown in Figure 6 and significant regions are compared to

Table 3. Behavior of 3D-QSAR models using different descriptors.

		Normal (electrostatics/sterics)					Electrostatics	Sterics
Dataset	*N*	Gast/Hück	Hückel	Pullman	MMFF94	SAM1	Gast/Hück	
CoMFA								
CBG[5]	21	0.74(2)	0.75(2)	0.68(2)	0.75(3)		0.74(2)	0.76(2)
Elastase[50]	32	0.64(3)	0.62(4)	0.62(4)	0.65(3)		0.65(3)	0.69(4)
DHFR[33]	22	0.32(4)	0.42(4)	0.34(4)	0.21(4)		−0.19(1)	0.33(3)
5-HT1A[54]	14	0.45(4)	0.38(4)	0.48(4)	0.52(2)	0.43(6)	0.39(4)	0.29(2)
I-Set[55]	20	0.59(3)	0.49(3)	0.56(3)	0.60(3)		0.14(4)	0.39(2)
Atom descriptors								
Dataset	*N*	Normal	Zero	H-bonding	Random			
HASL								
CBG	21	0.58	0.50	0.44	0.52			
Elastase	32	0.74	0.74	0.73	0.71			
DHFR	22	0.64	0.56	0.61	0.59			
5-HTIA	14	0.83	0.71	0.77	0.71			
CBG	21	0.54	0.58	0.41	0.54			

( ) = Number of components.

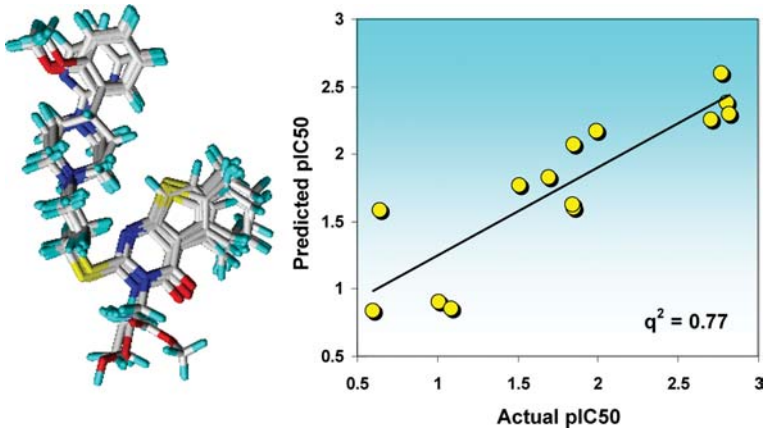

Figure 6. HASL cross-validated (LOO) analysis of the 5HT1A data set using SLG descriptors.

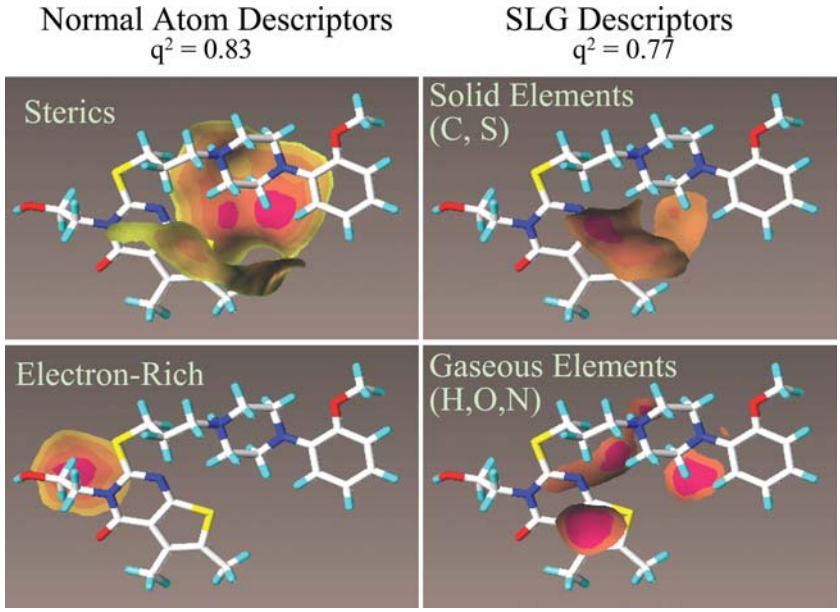

Figure 7. HASL models of the 5HT1A data set. Normal atom descriptor model shown on the left, SLG descriptor model shown on the right. Highlighted contours contribute positively to the predicted activity.

the Normal atom descriptors in Figure 7. As is readily apparent, since the SLG descriptor set does indeed meet reasonable statistical requirements (e.g., $q^2 = 0.77$), this 3D-QSAR model offers a perspective on interpretation that exceeds normal intuition or logic. For example, it is much easier to embrace a hypothesis which points to regions of higher electron density around a chemotype than it is to consider the ramifications of introducing atoms into a chemotype based solely on their physical state at STP. A careful examination of these 3D-QSAR models, in particular, when all atoms are considered the same (e.g., HASL, where all atoms are given an atom type of zero), would suggest that shape played a large role in the SAR and that further analyses which did not significantly increase model predictivity were unwarranted.

Conclusions

Although great strides have been made in the imaginative use of 3D-descriptors in correlation

efforts, 3D-QSAR remains largely a retrospective analytical tool. The random sampling of published 3D-QSAR models over the past decade supports this conclusion based on the rather poor test set predictions associated with such models, as well as the almost complete lack of correlation between q^2 and test set r^2. The fact remains that 3D-QSAR models are good at illustrating possible correlations existent in the data and thus provide a solid platform for retrospective hypotheses. Part of the problem may lie in the apparently arbitrary nature of the alignment paradigm and the descriptors themselves. The use of a molecular template appears to be the most logical choice when there is no target protein to act as a constraint, however, the several template-based models examined herein led to distorted conclusions when compared to those developed using a binding site constraint. Surprisingly, the choice of descriptors, whether they are field-based as in CoMFA or atom-based as in HASL, was found to be of little consequence in the creation of a viable 3D-QSAR model. Consequently, such a choice needs to be made with great care to insure that the resulting model can be interpreted in a useful manner.

In the end, 3D-QSAR methods need to be used responsibly, with care being taken to avoid undue extrapolation and the use of esoteric or uninterpretable descriptors. The good model will be one which is, at the least, a retrospective summary, highlighting features that are physically meaningful. Future efforts in 3D-QSAR should be aimed at overcoming the ambiguities and illusions described herein.

References

1. Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
2. Free Jr., S.M. and Wilson, J.W., J. Med. Chem., 7 (1964) 395.
3. Wise, M., Cramer, R.D., Smith, D. and Exman, I., In Dearden, J.C. (Ed.), Quantitative Approaches to Drug Design (Proceedings of the 4th European Symposium on Chemical Structure–Biological Activity: Quantitative Approaches), Elsevier, Amsterdam, The Netherlands, 1983, pp. 145–146.
4. Cramer III, R.D. and Wold, S.B., US Patent 5,025,388, June 18, 1991.
5. Cramer III, R.D., Patterson, D.E., and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
6. Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37 (1994) 4130.
7. Doweyko, A.M., J. Med. Chem., 31 (1988) 1396.
8. Silverman, B.D. and Platt, D.E., J. Med. Chem., 39 (1996) 2129.
9. Turner, D.B., Willet, P., Ferguson, A.M. and Heritage, T.W., J. Comput.-Aided Mol. Des., 11 (1997) 409.
10. Cruciani, G., Pastor, M. and Clementi, S., In Gundertofte, K. and Jorgensen, F.S. (Eds.), Modeling and Prediction of Activity. Kluwer, New York, 2000.
11. Thibaut, U., Folkers, G., Klebe, G., Kubinyi, H., Merz, A. and Rognan, D., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design, Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 711–716.
12. Oprea, T. and Garcia, A.E., J. Comput.-Aided Mol. Des., 10 (1996) 186.
13. Oprea, T., Waller, C.L. and Marshall, G.R., J. Med. Chem., 37 (1994) 2206.
14. Carroll, F.I., Mascarella, S.W., Kuzemko, M.A., Gao, Y., Abraham, P., Lewin, A.H., Boja, J.W. and Kuhar, M.J., J. Med. Chem., 37 (1994) 2865.
15. Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37(1994) 4130.
16. Kramer, O., Bohm, M., Schlitzer, M. and Klebe, G., In Höltje, H.-D. and Sippl, W. (Eds.), Rational Approaches to Drug Design (Proceedings of the 13th European Symposium on Quantitative Structure–Activity Relationships), Prous Science, Barcelona, Spain, 2001, pp. 359–363.
17. Baurin, N., Vangrevelinghe, E., Morin-Allory, L., Merour, J.-Y., Renard, P., Payard, M., Guillaumet, G. and Marot, C., J. Med. Chem., 43 (2000) 1109.
18. Zhang, S.-X., Feng, J., Kuo, S.-C., Brossi, A., Hamel, E., Tropsha, A. and Lee, K.-H., J. Med. Chem., 43 (2000) 167.
19. Lanig, H., Utz, W. and Gmeiner, P., J. Med. Chem., 44 (2001) 1151.
20. Buolamwini, J.K. and Assefa, H., J. Med. Chem., 45 (2002) 841.
21. Nair, A.C., Jayatilleke, P., Wang, X., Miertus, S. and Welsh, W.J., J. Med. Chem., 45 (2002) 973.
22. Avery, M.A., Alvim-Gaston, M., Rodrigues, C.R., Barreiro, E.J., Cohen, F.E., Sabris, Y.A. and Woolfrey, J.R., J. Med. Chem., 45 (2002) 292.
23. Debnath, A.K., Med. Chem., 45 (2002) 41.
24. Turner, D.B., Willett, P., Ferguson, A.M. and Heritage, T.W., J. Comput.-Aided Mol. Des., 13 (1999) 271.
25. Norinder, U., J. Chemometr., 10 (1996) 533.
26. Robinson, D.D., Winn, P.J., Lyne, P.D. and Richards, W.G., J. Med. Chem., 42 (1999) 573.
27. Lobato, M., Amat, Ll., Besalu, E. and Carbo-Dorca, R., QSAR, 16 (1997) 465.
28. Nielsen, S.F., Christensen, S.B., Cruciani, G., Kharazmi, A. and Liljefors, T., J. Med. Chem., 41 (1998) 4819.
29. Albuquerque, M.G., Hopfinger, A.J., Barreiro, E.J. and de Alencastro, R.B., J. Chem. Inf. Comput. Sci., 38 (1998) 925.
30. Kubinyi, H., Hamprecht, F.A. and Mietzner, T., J. Med. Chem., 41 (1998) 2553.
31. Nilsson, J., Homan, E.J., Smilde, A.K., Grol, C.J. and Wikstrom, H., J. Comput.-Aided Mol. Des., 12 (1998) 81.
32. McFarland, J.W., J. Med. Chem., 35 (1992) 2543.
33. Kaminiski, J.J. and Doweyko, A.M., J. Med. Chem., 40 (1997) 427.

34. Ferguson, A.M., Heritage, T., Jonathon, P., Park, S. E., Phillips, L., Rogan, J. and Snaith, P.J., J. Comput.-Aided Mol. Des., 11 (1997) 143.
35. Stanton, D.T., J. Chem. Inf. Comput. Sci., 39 (1999) 11.
36. Tafi, A., Anastassopoulou, J., Theophanides, T., Botta, M., Corelli, F., Massa, S., Artico, M., Costi, R., Di Santo, R. and Ragno, R., J. Med. Chem., 39 (1996) 1227.
37. Woolfrey, J.R., Avery, M.A. and Doweyko, A.M., J. Comput.-Aided Mol. Des., 12 (1998) 165.
38. Bostrom, J., Bohm, M., Gundertofte, K. and Klebe, G., J. Chem. Inf. Comput. Sci., 43 (2003) 1020.
39. Pearlstein, R.A., Vaz, R.J., Kang, J., Chen, X.-L., Preobrazhenskaya, M., Shchekotikhin, A.E., Korolev, A.M., Lysenkova, L.N., Miroshnikova, O.V., Hendrix, J. and Rampe, D., Bioorg. Med. Chem. Lett., 13 (2003) 1829.
40. Kharkar, P.S., Desai. B., Gaveria, H., Varu, B., Loriya, R., Naliapara, Y., Shah, A. and Kulkarni, V.M., J. Med. Chem., 45 (2002) 4858.
41. Huang, X., Xu, L., Luo, X., Fan, K., Ji, R., Pei, G., Chen, K. and Jiang, H., J. Med. Chem., 45 (2002) 333.
42. Chakraborti, A.K., Gopalakrishnan, B., Sobhia, M.E. and Malde, A., Bioorg. Med. Chem. Lett., 13 (2003) 1403.
43. Silverman, B.D., QSAR, 19 (2000) 237.
44. Turner, D.B. and Willet, P., J. Comput.-Aided Mol. Des., 14 (2000) 1.
45. Norinder, U., J. Comput.-Aided Mol. Des., 14 (2000) 545.
46. Tafi, A., Costi, R., Botta, M., Di Santo, R., Corelli, F., Federico, M., Silvio, C., Ciacci, A., Manetti, F. and Artico, M., J. Med. Chem., 45 (2002) 2720.
47. Ravi, M., Hopfinger, A.J., Hormann, R.E. and Dinan, L., J. Chem. Inf. Comput. Sci., 41 (2001) 1587.
48. Golbraikh, A. and Tropsha, A., J. Mol. Graphics, 20 (2002) 269.
49. Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y-D., Lee, K-H. and Tropsha, A., J. Comput.-Aided Mol. Des., 17 (2003) 241 (and ref. 21–24 within).
50. van Drie, J.H., Curr. Pharm. Des., 9 (2003) 1649.
51. van Drie, J.H., In Bultinck, P. (Ed.), Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, New York, 2004, pp. 437–460.
52. Tropsha, A. and Cho, S.J., Perspect. Drug Discov. Des., 12–14 (1998) 56.
53. RCSB (Research Collaboratory for Structural Bioinformatics), Rutgers University, NJ (www.rcsb.org/pdb).
54. Doweyko, A.M., In Holtje, H.-D. and Sippl, W. (Eds.), Rational Approaches to Drug Design (Proceedings of the 13th European Symposium on Quantitative Structure-Activity Relationships), Prous Science, Barcelona, Spain, 2001, pp. 307–315.
55. Flo(qxp), Colin McMartin, Thistlesoft, Colebrook, CT.
56. CoMFA module within Sybyl, Tripos, St. Louis, MO.
57. HASL, Hypothesis software, Long Valley, NJ.
58. Guccione, S., Doweyko, A.M., Chen, H., Barretta, G.U. and Balzano, F., J. Comput.-Aided Mol. Des., 14 (2000) 647.
59. I-Set: Twenty substituted naphthalenes (−log activity values 2.8–7.0) overlayed using a template paradigm. Activities were adjusted to optimize the CoMFA model using Gasteiger/Hückel charges.
60. SAM1 – Semiempirical quantum mechanical program – implemented within Sybyl (Tripos, St. Louis, MO).