JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

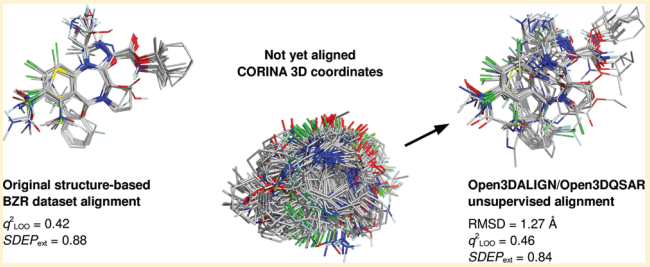
A 3D-QSAR-Driven Approach to Binding Mode and Affinity Prediction

Paolo Tosco*,† and Thomas Balle‡

†Department of Drug Science and Technology, University of Turin, Via Pietro Giuria 9, 10125 Torino, Italy

‡Department of Medicinal Chemistry, The Faculty of Pharmaceutical Sciences, University of Copenhagen, 2 Universitetsparken, 2100 Copenhagen, Denmark

ABSTRACT: A method for predicting the binding mode of a series of ligands is proposed. The procedure relies on three-dimensional quantitative structure−activity relationships (3D-QSAR) and does not require structural knowledge of the binding site. Candidate alignments are automatically built and ranked according to a consensus scoring function. 3D-QSAR analysis based on the selected binding mode enables affinity prediction of new drug candidates having less than 10 rotatable bonds.



INTRODUCTION

Three-dimensional quantitative structure−activity relationships (3D-QSAR) are aimed at building quantitative models by relating biological activities of a series of ligands to their 3D properties.[1] For this purpose, most methods, including the well-established comparative molecular field analysis/comparative molecular similarity indices analysis (CoMFA/CoMSIA)[2,3] and the techniques implemented in the programs GRID and GOLPE,[4] rely on partial least-squares (PLS) statistical treatment of computed molecular interaction fields (MIFs). In general, prior to computing MIFs and carrying out PLS analysis, 3D-QSAR require the determination, or an educated guess, of the bioactive conformation of a template molecule, followed by alignment of the whole data set onto the latter. The resulting PLS model may have the power to predict the activity of new molecules before they are synthesized and tested. Several other regression methods, including neural networks,[5] support vector machines and random forest,[6] may yield similar or even better predictions, but PLS techniques are unique in the sense that regression pseudocoefficients may be mapped in the 3D space of the molecules, thus allowing visualization of the regions which contribute most to explaining the differences in biological activity across a data set. These 3D maps have turned out to be a powerful tool that can be used in the design of new ligands with improved activity/selectivity and are among the reasons why 3D-QSAR techniques, after more than 20 years since they were first introduced, are still very popular, though more in the academic community than in pharmaceutical industry.[7]

However, the strength of 3D-QSAR may as well be regarded as its achilles heel, since a reasonable guess of the bioactive conformation is required in order to obtain good quality models and interpretable results. Identifying the bioactive conformation of a template molecule is not a trivial task, and even when it is known, usually from an experimentally determined structure of the ligand−target complex or from carefully designed and synthesized rigid analogues, the alignment procedure itself is a difficult and time-consuming operation, especially in the presence of flexible or structurally heterogeneous ligands. When the structure or even the identity of the target is not known, it becomes difficult to hypothesize a univocal and reliable alignment, thus making it hard to apply 3D-QSAR. Unfortunately, the lack of knowledge of the structure of the target is also the situation where a ligand-based approach would be most desirable, being basically the only option for computer-aided drug design.

We have recently realized a 3D-QSAR project where we aimed at overcoming the template selection bottleneck by using virtually all conformers within an energetically accessible window as possible templates, in order not to be biased toward arbitrarily selecting a single conformer. Using this procedure on a series of nicotinic $\alpha_4\beta_2$ receptor agonists and partial agonists, we showed that, among all evaluated alignments, one compatible with pharmacophore models, site-directed mutagenesis studies, and X-ray complexes of acetylcholine binding proteins could be identified. For this purpose, a 3D-QSAR model was built on each individual alignment, and its predictive performance was used as a scoring function.[8] The idea of reversing the approach to 3D-QSAR models, namely using them as a tool to select the best among many possible alignments, was first pioneered by Jain and co-workers in COMPASS, a method based on nonlinear modeling of molecular surface properties encoded by steric and hydrogen-bonding descriptors.[9] COMPASS requires an initial guess of the alignment, based on a user-identified pharmacophore common to all molecules; the initial alignment, which needs to be good enough to yield a predictive model from the beginning, is then iteratively optimized until for each molecule the conformer yielding the best performance in the associated

Special Issue: 2011 Noordwijkerhout Cheminformatics

Received: September 1, 2011
Published: November 16, 2011

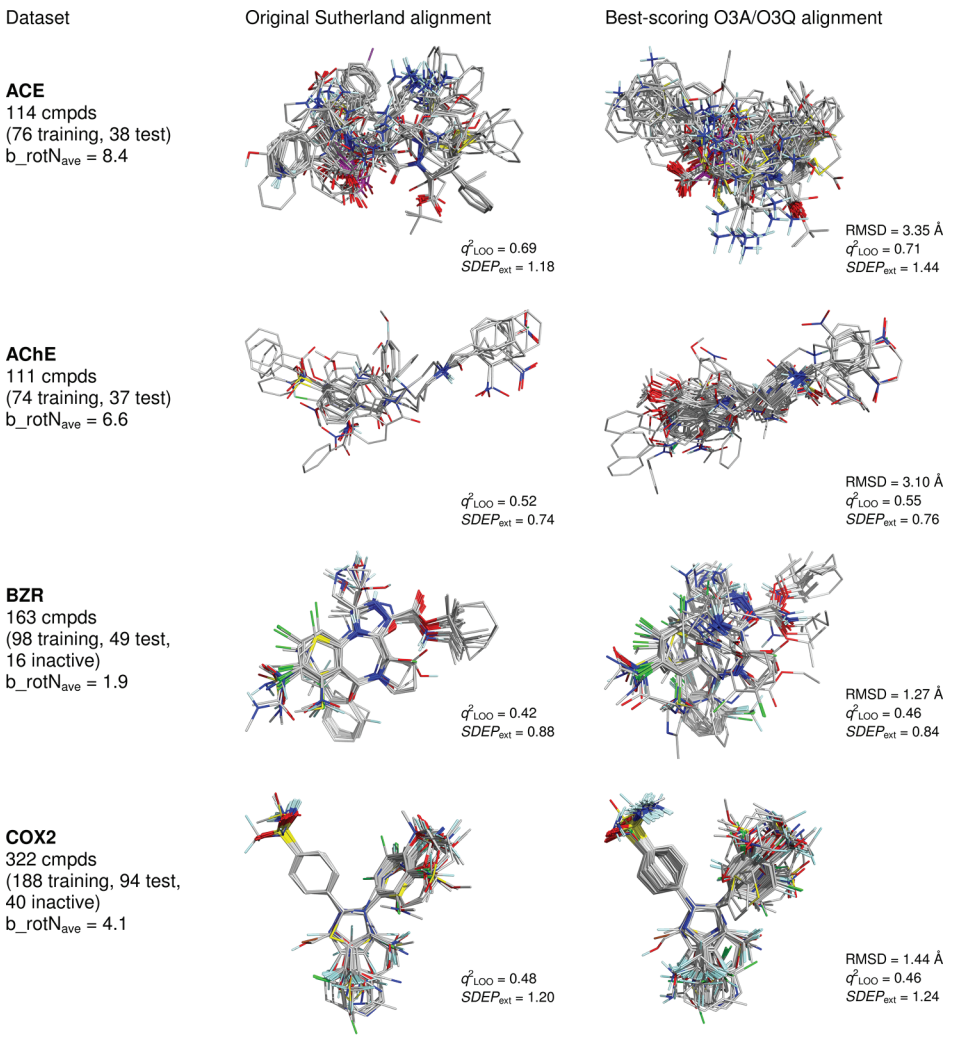


Figure 1. Original alignment compared to the best-scoring alignment as found by the unsupervised Open3DALIGN/Open3DQSAR strategy for ACE, AChE, BZR, and COX2 data sets. Average numbers of rotatable bonds across the eight data sets (b_rotN$_{ave}$) were computed with MOE.[21] Molecules were rendered with PyMOL.[22]

model is found. Our approach differs from COMPASS in two major features: (1) it uses linear PLS statistics on any combination of grid-based MIFs and (2) rather than optimizing a single, user-generated binding mode hypothesis, it challenges a large number of highly diverse, machine-generated binding modes by assessing via a consensus scoring function their degree of internal consistency along with their ability to yield a predictive 3D-QSAR model. On the way toward an automated and unbiased binding mode explorer, we have realized two open-source software projects: (1) Open3DALIGN for unsupervised generation and scoring of alignments[10] and (2) Open3DQSAR for MIF computation, PLS model building, validation, and refinement through variable selection.[11] Herein we describe how Open3DALIGN and Open3DQSAR can be combined to generate and score hypotheses on the bioactive conformation of a series of ligands.

METHODS

To determine the feasibility as well as the domain of applicability of our approach, we chose as test bench the eight data sets gathered from literature by Sutherland and co-workers.[12] These data sets include inhibitors of the angiotensin converting enzyme (ACE), acetylcholinesterase (AChE), benzodiazepine receptor (BZR), cyclooxygenase-2 (COX2), dihydrofolate reductase (DHFR), glycogen phosphorylase b (GPB), thermolysin (THERM), and thrombin (THR). They are quite large (from 66 to 397 compounds) and characterized by a wide range of size, flexibility, and stereoelectronic properties; biological activities cover at least 4 orders of magnitude, thus constituting the ideal playground for 3D-QSAR; moreover, CoMFA model statistics are available for comparison purposes. For each data set the following automated workflow was adopted:

- Not yet aligned 3D coordinates of the compounds generated by Sutherland and co-workers with CORINA[13] were imported into Open3DALIGN in the SD file format.
- Conformational sampling by quenched molecular dynamics (QMD)[14] was carried out within Open3DALIGN on each structure (MMFF94 force-field, GB/SA implicit solvent model, 1000 5 ps molecular dynamics runs at 1000 K followed by energy minimization), keeping the most stable conformations in a 8 kcal mol^{-1} range from the global minimum; this energy strain threshold was previously shown to include a percentage of experimentally found bioactive

dx.doi.org/10.1021/ci200411s |J. Chem. Inf. Model. 2012, 52, 302–307

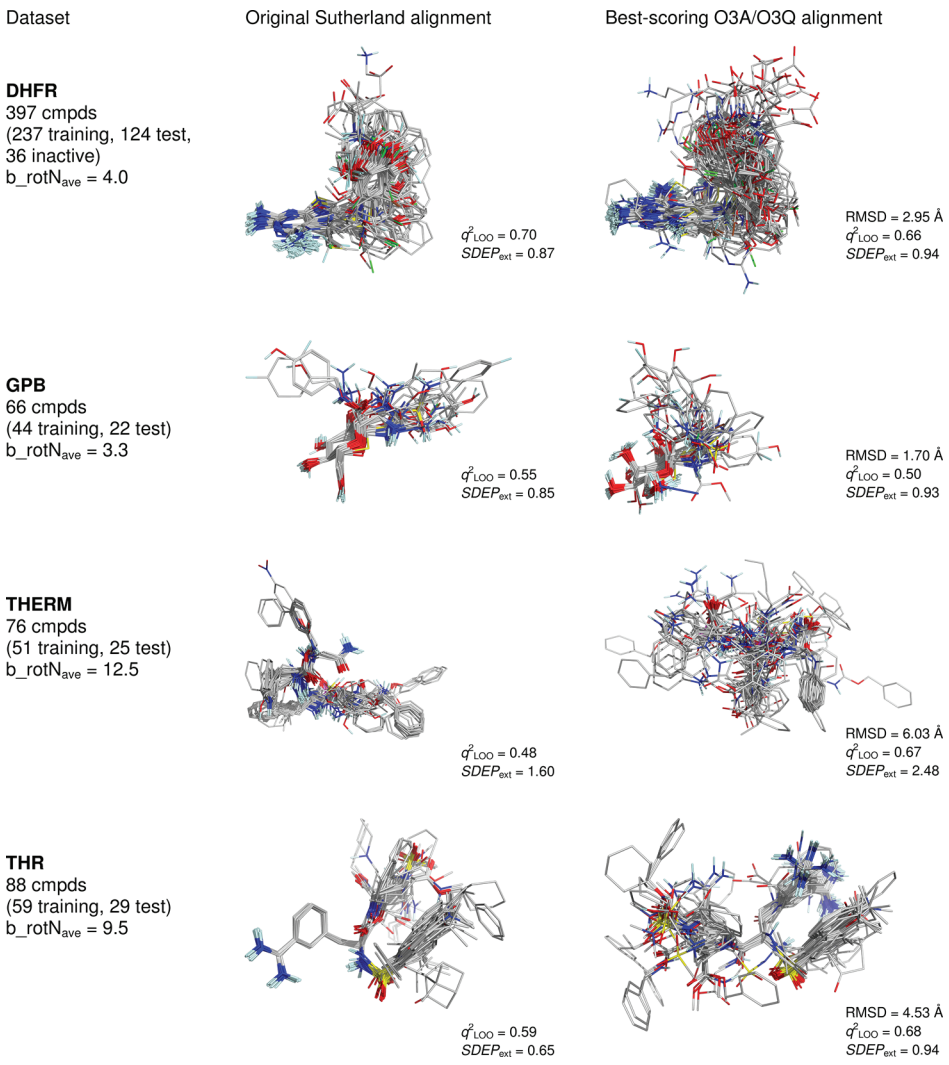


Figure 2. Original alignment compared to the best-scoring alignment as found by the unsupervised Open3DALIGN/Open3DQSAR strategy for DHFR, GPB, THERM, and THR data sets.

conformations ranging from 60 to 90%, depending on their flexibility.[15] Pairs of conformers whose heavy atom root-mean-square deviation (rmsd) was below 0.2 Å were considered duplicate, and the higher energy one was discarded.

- The data set was split into a training set (2/3 of the compounds) and a test set (1/3 of the compounds) according to the original composition chosen by Sutherland et al.,[12] in order to ensure comparability of results.
- According to the paradigm mentioned above, each conformer of training set compounds could act as a potential template to align the rest of the data set. To reduce the number of possible templates, a two-step filtering procedure recently implemented in Open3DALIGN was adopted. Initially, pharmacophores were extracted from potential template conformations with Pharao.[16] Subsequently, pairwise comparisons were made between pharmacophores, assessing the value of two similarity indexes computed by Pharao, namely Tanimoto and Tversky scores, both ranging from 0 to 1.[16] In the first filtering step, conformers of the same molecule whose pharmacophores had a Tanimoto similarity score higher than a threshold value (here, 0.7) were considered as duplicates (intradatabase filtering). In the second step, all conformers retained from the individual databases were merged together and sorted according to decreasing number of pharmacophoric features, then submitted to pairwise comparisons (interdatabase filtering). Again, a threshold value was chosen, depending on the degrees of torsional freedom, to assess similarity between the pharmacophores under comparison (here, 0.7 for BZR, COX2, DHFR, and GPB and 0.6 for ACE, AChE, THERM, and THR). All conformer pairs whose Tanimoto score exceeded the threshold were treated as duplicates, while the conformers whose Tversky score was above the threshold were considered as subsets/supersets of each other; in both cases, the conformers whose pharmacophore had the highest number of features was retained. The overall filtering procedure allowed using as templates only the structures which encoded the most diverse and informative arrangements of pharmacophoric features and were therefore representative of a truly distinct binding mode.
- The retained templates were used to align both training and test sets, picking for each compound the best-fitting conformer from the QMD pool. Alignments were carried out in an atom-based fashion using the "mixed" algorithm as implemented in Open3DALIGN.[10]

dx.doi.org/10.1021/ci200411s |*J. Chem. Inf. Model.* 2012, 52, 302–307

Table 1. Statistics of the 3D-QSAR Models Obtained by Different Strategies from the Eight Benchmark Data Sets

		ACE	AChE	BZR	COX2	DHFR	GPB	THERM	THR
n (training set)		76	74	98	188	237	44	51	59
n (test set)		38	37	49	94	124	22	25	29
optimal PCs	CoMFA[a]	3	5	3	5	5	4	4	4
	O3Q[b]	5	5	3	5	5	5	4	4
	O3A/O3Q[c]	4	5	3	5	5	5	5	5
r^2	CoMFA[a]	0.80	0.88	0.61	0.70	0.79	0.84	0.85	0.86
(training set)	O3Q[b]	0.86	0.87	0.61	0.73	0.80	0.94	0.78	0.83
	O3A/O3Q[c]	0.85	0.89	0.69	0.73	0.80	0.95	0.95	0.90
SDEC	CoMFA[a]	1.04	0.41	0.41	0.56	0.59	0.43	0.73	0.36
(training set)	O3Q[b]	0.86	0.43	0.41	0.53	0.56	0.27	0.89	0.39
	O3A/O3Q[c]	0.92	0.41	0.37	0.53	0.56	0.25	0.41	0.31
q^2_{LOO}	CoMFA[a]	0.68	0.52	0.32	0.49	0.65	0.42	0.52	0.59
(training set)	O3Q[b]	0.69	0.52	0.42	0.48	0.70	0.55	0.48	0.59
	O3A/O3Q[c]	0.71	0.55	0.46	0.46	0.66	0.50	0.67	0.68
$q^2_{L10\%O}$	CoMFA[a]	0.69	0.52	0.32	0.48	0.65	0.47	0.49	0.50
(training set)	O3Q[b]	0.68	0.49	0.41	0.47	0.69	0.52	0.46	0.53
	O3A/O3Q[c]	0.70	0.53	0.45	0.44	0.65	0.46	0.66	0.63
$q^2_{L20\%O}$	CoMFA[a]	0.67	0.49	0.31	0.46	0.63	0.37	0.45	0.47
(training set)	O3Q[b]	0.67	0.45	0.40	0.45	0.69	0.50	0.43	0.49
	O3A/O3Q[c]	0.69	0.51	0.44	0.42	0.65	0.41	0.63	0.59
$q^2_{L33\%O}$	CoMFA[a]	0.65	0.42	0.28	0.43	0.62	0.33	0.42	0.40
(training set)	O3Q[b]	0.66	0.39	0.39	0.43	0.67	0.44	0.40	0.41
	O3A/O3Q[c]	0.67	0.46	0.42	0.39	0.63	0.34	0.59	0.51
r^2_{pred}	CoMFA[a]	0.49	0.47	0.00	0.29	0.59	0.42	0.54	0.63
(test set)	O3Q[b]	0.69	0.67	0.17	0.32	0.60	0.50	0.51	0.67
	O3A/O3Q[c]	0.54	0.65	0.24	0.28	0.53	0.41	−0.18	0.30
SDEP	CoMFA[a]	1.54	0.95	0.97	1.24	0.89	0.94	1.59	0.70
(test set)	O3Q[b]	1.18	0.74	0.88	1.20	0.87	0.85	1.60	0.65
	O3A/O3Q[c]	1.44	0.76	0.84	1.24	0.94	0.93	2.48	0.94

[a] Statistics of the original CoMFA model (ref 12). [b] Statistics of the model computed by Open3DQSAR using the built-in MMFF94 van der Waals parameters and charges on the same alignment. [c] Statistics of the best-ranked model obtained by the combined Open3DALIGN/Open3DQSAR strategy using the consensus scoring function.

- The aligned ligand ensembles were enclosed in a grid box exceeding the largest molecule by 5 Å in each direction; a 1 Å step size was preferred to the original coarser 2 Å mesh to reduce the dependency from grid-to-molecule reciprocal orientations. Steric and electrostatic MIFs were computed with Open3DQSAR using MMFF94 van der Waals parameters and charges. The probe was constituted by a sp^3 carbon atom bearing a unit positive charge.
- Training set MIF data were prefiltered as in the Sutherland paper;[12] namely, an energy cutoff was set at ±30 kcal/mol; variables having a standard deviation below 2.0 were discarded; block unscaled weighting[17] was applied to steric and electrostatic fields to give them the same importance in the PLS model.
- pIC_{50} values corresponding to the individual molecules were correlated with MIF data for each alignment using PLS regression; the optimal number of principal components (PC) to be extracted was chosen with the same criterion adopted by Sutherland et al.,[12] namely the one giving rise to the best leave-one-out cross-validation performance, expressed as q^2_{LOO}. Leave-10%-out ($q^2_{L10\%O}$), leave-20%-out ($q^2_{L20\%O}$), and leave-33%-out ($q^2_{L33\%O}$) cross-validation procedures were also carried out, to enable comparison with previously reported results.[12] Finally, the predictive power of each PLS model was evaluated against the external test set and expressed both as r^2_{pred} and as standard deviation of the error of prediction ($SDEP_{ext}$).[18]
- The quality of each alignment/model was ranked by a consensus scoring function which combines three different ranking criteria, namely the Open3DALIGN alignment score ($score_{O3A}$; the higher, the better), the solvent accessible surface area (SAS; the lower, the better), and q^2_{LOO} (the higher, the better). Since no significant differences in ranking were noticed when replacing q^2_{LOO} with more robust leave-many-out q^2 indexes, the former was preferred since its determination does not rely on randomness. While $scoreO_{3A}$ assesses the goodness-of-fit using rmsd and the degree of chemical match between atoms, SAS proved very useful to discriminate more consistent, "cavity-shaped" alignments (lower SAS) from inconsistent and spread ones (higher SAS). All scores where normalized, rounded to one decimal and classified on an 11 level scale to reduce the impact of the uncertainties inherent in each metric. The rank indexes obtained with each criterion were then summed up according to the following consensus function: $\text{overall rank} = \text{rank}(\text{score}_{O3A}) + \text{rank}(\text{SAS}) + 10 \times \text{rank}(q^2_{LOO})$

dx.doi.org/10.1021/ci200411s |J. Chem. Inf. Model. 2012, 52, 302–307

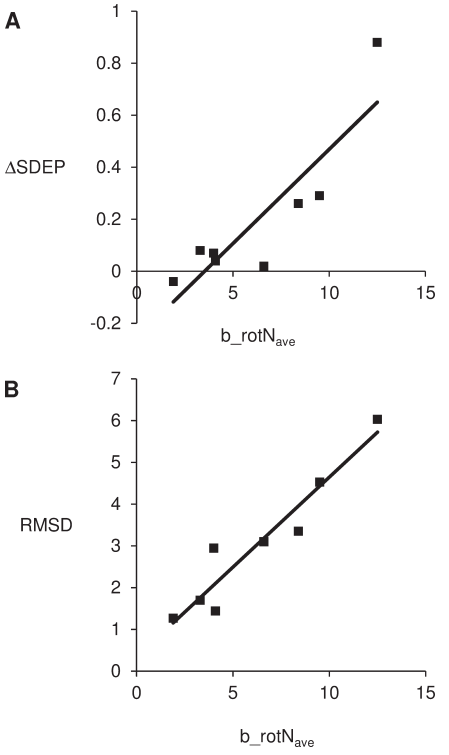
This function penalizes binding modes for which the associated 3D-QSAR model does not have good internal predictive power, while using alignment quality metrics to refine ranking.

RESULTS AND DISCUSSION

The best-scoring alignment for each data set is depicted in Figures 1 and 2, along with the original alignment as a reference. Complete 3D-QSAR model statistics are collected in Table 1; the latter reports the results obtained on the original alignment by Sutherland with CoMFA using modified neglect of differential overlap (MNDO) charges ("CoMFA" rows), those obtained with Open3DQSAR on the same alignments using MMFF94 van der Waals parameters and electrostatic charges ("O3Q" rows), and finally those obtained with Open3DQSAR on the best-scoring Open3DALIGN alignment selected using the consensus scoring function ("O3A/O3Q" rows). Using the original Sutherland alignments, models obtained with Open3DQSAR have equal or better statistics than CoMFA; the improvement is amenable both to the higher grid resolution and to the different force field parameters.

Moving to the O3A/O3Q models obtained from the multiconformational, unsupervised alignments, they all have a performance close to the O3Q model obtained on the Sutherland alignment as far as internal predictivity indexes are considered. However, when the ability to predict affinities of an external test set which has never come into contact with the model previously is evaluated, only five models (namely, AChE, BZR, COX2, DHFR, and GPB) retain a predictive power within 0.1 standard deviation errors of prediction (SDEP) units with respect to those built on the original alignment. There is a clear trend correlating the average number of rotatable bonds across the data set to the ΔSDEP between the original O3Q and the corresponding O3A/O3Q models (Chart 1A, $r^2 = 0.78$); a similar trend exists with rms deviations between the original Sutherland alignments and those associated with the O3A/O3Q models (Chart 1B, $r^2 = 0.91$). While O3Q/O3A-generated alignments for BZR, COX2, and GPB data sets are largely reminiscent of the Sutherland ones (2−4 rotatable bonds; rmsd < 1.7 Å), more flexible data sets display larger deviations from the original (ACE, AChE, and DHFR; 4−6.6 rotatable bonds; rmsd ≈ 3 Å), even if external predictivity is still almost as good. A closer examination of the alignments shows that certain parts of the scaffold are aligned just as in the Sutherland reference, e.g., the diaminoheteroaryl moiety in DHFR or the arylalkylpiperidino moiety in AChE, while other parts, being consistently aligned with respect to each other, assume different orientations. This is largely unavoidable, since the scoring function does enforce internal consistency via the $score_{O3A}$ and SAS terms but is not aware of the shape of the cavity; wherever compounds can bend due to intrinsic flexibility, the scoring function cannot warrant that they bend according to the cavity constraints, but only that they bend consistently. This phenomenon becomes even more evident in the THR data set (9.5 rotatable bonds; rmsd = 4.5 Å) where the amidinophenyl and the amide moiety consistently exchange places compared to the original alignment; while this exchange obviously has a large impact on the rmsd, external predictivity is not affected as much (ΔSDEP = 0.29). Instead, when molecules become very flexible (THERM; 12.5 rotatable bonds; rmsd = 6.0 Å) and pharmacophoric features more heterogeneous across the data set, the methodology clearly fails to obtain a consistent alignment, and

Chart 1. Correlation between Model Reliability and Average Number of Rotatable Bonds Across Each Data Set[a]



[a] b_rotN_{ave}, average number of rotatable bonds across the data set; ΔSDEP, difference between $SDEP_{ext}$ of the 3D-QSAR model based on the Sutherland alignment and the respective $SDEP_{ext}$ of the best scoring O3A/O3Q 3D-QSAR model; and rmsd, root-mean-square deviation in Å between the reference Sutherland alignment and the best scoring O3A/O3Q alignment.

the associated model is basically devoid of external predictivity; the failure is not amenable to the scoring function, since none of the models has QSAR statistics comparable to the Sutherland reference. These observations appear to limit the domain of safe applicability of this multiple-alignment 3D-QSAR strategy to data sets having on average less than six rotatable bonds. While the method still has validity on compounds characterized by 6−10 rotatable bonds, in this case less credit should be given to the hypothesized binding mode, while the activity predictions for new drug candidates remain quite reliable. However, while individual moieties in the ligands may assume absolute orientations which turn out to be incompatible with the shape of the binding site, their relative orientations are still rather consistent. Above 10 rotatable bonds, the method fails and should not be applied, since fair cross-validated q^2 indices might eventually induce confidence in models which have no real predictive power and are very far from yielding a sound binding mode hypothesis. This flexibility threshold is also in accordance with the energy strain range imposed to the conformational search, so that it would not be reasonable to expect any success beyond this limit. Our findings confirm once again that in the absence of the ability to correctly predict activities of a truly external test set, a high cross-validated q^2 value may simply indicate redundancy in the training set and does not certify per se the quality of a 3D-QSAR model.[19,20] However, combining the relatively weak q^2 indicator with alignment-based metrics in the above-described consensus

scoring function allows a much more robust ranking of binding mode hypotheses.

This work shows that it is possible to reverse the usual 3D-QSAR paradigm, where a model is built on the basis of prior knowledge about ligands' binding mode, obtained mostly by crystallographic complexes. We show that, in the absence of any information about the target, meaningful 3D-QSAR models can be generated and selected by a purely ligand-based approach and used in combination with other metrics to assess the trustworthiness of the underlying alignment. On the Sutherland validation suite, in five cases out of eight the models identified by our unsupervised approach had an external predictive power very close to structure-derived models and could therefore be used to predict the activity of new drug candidates.

The whole Open3DALIGN/Open3DQSAR procedure (QMD conformational search, template filtering, alignment, 3D-QSAR model building) for a 300 compound data set requires about one day on a 2.4 GHz 8-core Intel Xeon E5620 CPU.

Open3DALIGN and Open3DQSAR are available for download free of charge under the terms of the GNU GPLv3 at the URLs http://open3dalign.org/ and http://open3dqsar.org/, respectively.

AUTHOR INFORMATION

Corresponding Author

*E-mail: paolo.tosco@unito.it. Telephone: +39 011 670 7680.

ACKNOWLEDGMENT

We are grateful to the developers of OpenBabel, Pharao, and TINKER on which our tools partly depend to do their job. We acknowledge the support of Chemical Computing Group. Part of this work was carried out by P.T. at the University of Copenhagen under a visiting scientist grant supported by the Drug Research Academy (DRA). T.B. was supported by grants from the Lundbeck Foundation and from DRA.

REFERENCES

(1) Akamatsu, M. Current state and perspectives of 3D-QSAR. *Curr. Top. Med. Chem.* **2002**, *2*, 1381–1394.

(2) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(3) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

(4) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.

(5) Tetko, I. V.; Kovalishyn, V. V.; Livingstone, D. J. Volume learning algorithm artificial neural networks for 3D QSAR studies. *J. Med. Chem.* **2001**, *44*, 2411–2420.

(6) Manchester, J.; Czerminski, R. SAMFA: simplifying molecular description for 3D-QSAR. *J. Chem. Inf. Model.* **2008**, *48*, 1167–1173.

(7) Cross, S.; Cruciani, G. Molecular fields in drug discovery: getting old or reaching maturity? *Drug Discovery Today* **2010**, *15*, 23–32.

(8) Tosco, P.; Ahring, P. K.; Dyhring, T.; Peters, D.; Harpsøe, K.; Liljefors, T.; Balle, T. Complementary three-dimensional quantitative structure-activity relationship modeling of binding affinity and functional potency: a study on $\alpha_4\beta_2$ nicotinic ligands. *J. Med. Chem.* **2009**, *52*, 2311–2316.

(9) (a) Jain, A. N.; Koile, K.; Chapman, D. Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J. Med. Chem.* **1994**, *37*, 2315–2327. (b) Jain, A. N.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow, R. E., Jr.; Bauer, B. E.; Webster, T. A.; Lozano-Perez, T. Compass: a shape-based machine learning tool for drug design. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 635–652.

(10) Tosco, P.; Balle, T.; Shiri, F. Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 777–783.

(11) Tosco, P.; Balle, T. Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J. Mol. Model.* **2011**, *17*, 201–208.

(12) 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.

(13) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547.

(14) O'Connor, S. D.; Smith, P. E.; al-Obeidi, F.; Pettitt, B. M. Quenched molecular dynamics simulations of tuftsin and proposed cyclic analogs. *J. Med. Chem.* **1992**, *35*, 2870–2881.

(15) (a) Boström, J.; Norrby, P.-O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383–396. (b) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499–2510.

(16) Taminau, J.; Thijs, G.; De Winter, H. Pharao: pharmacophore alignment and optimization. *J. Mol. Graphics Modell.* **2008**, *27*, 161–169.

(17) Kastenholz, M. A.; Pastor, M.; Cruciani, G.; Haaksma, E. E. J.; Fox, T. GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43*, 3033–3044.

(18) Cruciani, G.; Baroni, M.; Clementi, S.; Costantino, G.; Riganelli, D.; Skagerberg, B. Predictive ability of regression models. Part I: standard deviation of prediction errors (SDEP). *J. Chemometr.* **1992**, *6*, 335–346.

(19) Golbraikh, A.; Tropsha, A. Beware of q^2!. *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.

(20) Doweyko, A. M. 3D-QSAR illusions. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 587–596.

(21) *MOE*, version 2010.10; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2010.

(22) *The PyMOL Molecular Graphics System*, version 1.4.1; Schrödinger LLC: New York, 2011.