

Virtual Screening for R-Groups, including Predicted pIC₅₀ Contributions, within Large Structural Databases, Using Topomer CoMFA

Richard D. Cramer,* Phillip Cruz, Gunther Stahl, William C. Curtiss, Brian Campbell,
Brian B. Masek, and Farhad Soltanshahi

Tripos International, 1699 South Hanley Road, St. Louis, Missouri 63144

Received May 5, 2008

Multiple R-groups (monovalent fragments) are implicitly accessible within most of the molecular structures that populate large structural databases. R-group searching would desirably consider pIC₅₀ contribution forecasts as well as ligand similarities or docking scores. However, R-group searching, with or without pIC₅₀ forecasts, is currently not practical. The most prevalent and reliable source of pIC₅₀ predictions, existing 3D-QSAR approaches, is also difficult and somewhat subjective. Yet in 25 of 25 trials on data sets on which a field-based 3D-QSAR treatment had already succeeded, substitution of objective (canonically generated) topomer poses for the original structure-guided manual alignments produced acceptable 3D-QSAR models, on average having almost equivalent statistical quality to the published models, and with negligible effort. Their overall pIC₅₀ prediction error is 0.805, calculated as the average over these 25 topomer CoMFA models in the standard deviations of pIC₅₀ predictions, derived from the 1109 possible “leave-out-one-R-group” (LOORG) pIC₅₀ contributions. (This novel LOORG protocol provides a more realistic and stringent test of prediction accuracy than the customary “leave-out-one-compound” LOO approach.) The associated average predictive r^2 of 0.495 indicates a pIC₅₀ prediction accuracy roughly halfway between perfect and useless. To assess the ability of topomer-CoMFA based virtual screening to identify “highly active” R-groups, a Receiver Operating Curve (ROC) approach was adopted. Using, as the binary criterion for a “highly active” R-group, a predicted pIC₅₀ greater than the top 25% of the observed pIC₅₀ range, the ROC area averaged across the 25 topomer CoMFA models is 0.729. Conventionally interpreted, the odds that a “highly active” R-group will indeed confer such a high pIC₅₀ are 0.729/(1–0.729) or almost 3 to 1. To confirm that virtual screening within large collections of realized structures would provide a useful quantity and variety of R-group suggestions, combining shape similarity with the “highly active” pIC₅₀, the 50 searches provided by these 25 models were applied to 2.2 million structurally distinct R-group candidates among 2.0 million structures within a ZINC database, identifying an average of 5705 R-groups per search, with the highest predicted pIC₅₀ combination averaging 1.6 log units greater than the highest reported pIC₅₀s.

Lead optimization is usually the most challenging medicinal chemistry activity. Not only are its structural candidates usually more demanding to synthesize but also their high structural similarity makes them challenging to select among, especially considering the overall goal of improving upon some still unsatisfactory properties while retaining all the satisfactory ones. Sooner or later, only one or two R-group positions (substitution sites) may seem amenable to further modifications. In this situation any CADD methodology that would provide a wide range of alternative R-group suggestions, all promising on the basis of current SAR knowledge from all relevant bioassays as well as shape similarity to established R-groups, and all synthesizable, seems particularly desirable.

Today large repositories of molecular structures are universally available, most offered for commercial purchase (at least in principle) and therefore presumably synthesizable. One publicly available example is the ZINC database,¹ containing around two million structures. Such databases are

routinely used as sources of screening candidates, selections being based on receptor docking and/or various types of ligand similarity calculations. Less obviously, these structures also contain a large population of R-group suggestions for lead optimization. Cleavage of any acyclic bond within any of these molecular structures generates two such R-group candidates (even though many, for example simply substituted phenyl groups, become duplicates).

To benefit lead optimization, such a pool of R-groups also requires project-relevant means of candidate selection.² It is often unappreciated that the two most important ligand-based selection approaches, structural similarity and pIC₅₀ prediction, conflict with one another. According to similarity searching, any and every structural change is undesirable, with the highest possible similarity score being “indistinguishable from the query”. However, the expectation that indeed motivates every discovery project is instead that structural change can be found that will have desirable rather than undesirable effects, with (Q)SAR reasoning often used, explicitly or not, to detect and extrapolate underlying trends, thereby allowing promising R-group alternatives to be identified.

* Corresponding author phone: (505)982-0406 or (314)647-1099; fax: (314)647-9241; e-mail: cramer@tripos.com.

Thus pIC50 predictions are potentially the most productive criteria for candidate selection and within as many relevant bioassays as possible. However the average error of any such prediction should be less than the standard deviation of experimental pIC50s among the tested structures. (Otherwise a random selection would be at least as effective.) While the actual variation of potencies within a particular series depends greatly on the particular structures and assays, a preliminary tabulation of the SAR publications that often accompany successes in lead optimization suggests a typical standard deviation of primary assay pIC50s to be, very roughly, 1.0 to 1.3 in units of pIC50.³

Most published predictions of pIC50 values, especially those also reporting errors in those predictions, are provided by various 3D-QSAR methods. Two summaries^{4,5} reference a total of over fifty studies that reported the outcomes of "true" pIC50 predictions (involving "test sets" disjoint from "training sets"). The errors of pIC50 prediction as stated by two of these summaries are 0.55 (average over 15 studies of SDEP values either reported or directly calculated) and 0.85 (imputed from the reported average predictive- r^2 values over 61 studies). These moderate errors support this customary use of 3D-QSAR for ranking candidate structures. However the costs and uncertainties associated with application of any current 3D-QSAR methodology are also well-known. Taking CoMFA⁶ as an example, each included structure, whether used for model derivation or pIC50 prediction, must be individually posed,⁷ i.e., converted into a 3D representation that is somehow aligned with all the other 3D representations. Such pose generation⁸ remains time-consuming, frustrating, and somewhat subjective. During model derivation, at least an aura of objectivity is conferred by the q^2 criterion, but when then using that model to predict, the only justification for any pose choice is some sort of similarity to a training set ligand pose, with the resulting pIC50 value depending greatly on that pose choice. Thus there exists a reasonable concern that the apparent accuracy of pIC50 predictions from 3D-QSAR is biased by analyst subjectivity. And the labor needed to select a pose also discourages pIC50 predictions for more than a few ligand candidates.

However these 3D-QSAR weaknesses seem to be overcome by the new topomer CoMFA methodology.⁴ A topomer is a pose of a *fragment*, generated canonically, that is, by considering only its "2D" structure, not by any sort of relationship to either a receptor site or to some other ligand.⁹ The intent of the rules that generate a topomer pose¹⁰ is to produce a "bioisosteric shape", such that similarity in that topomer shape is a strong predictor of biological similarity, an outcome that has been validated in both retrospective and prospective trials.¹⁰⁻¹³ When performing topomer CoMFA, these topomers simply become the 3D-QSAR poses, with each training set ligand (as a "2D" structure) being cut apart at one or more user-selected acyclic bonds and topomers then being generated automatically from the resulting R-groups (monovalent fragments). Rather than the familiar single CoMFA column, a completed topomer CoMFA table contains multiple "CoMFA columns", each column containing a different set of presumptively self-comparable ligand R-groups.

A first assessment of topomer poses in CoMFA involved reanalysis of fifteen 3D-QSAR-amenable data sets from

eleven randomly encountered publications.⁴ Statistically significant 3D-QSAR relationships emerged in every one of these fifteen topomer CoMFA trials, a very unlikely outcome if success in 3D-QSAR instead requires the customary user-generated poses of individual training set structures. At least as importantly, the averaged errors in 133 pIC50 true predictions were, indistinguishably, 0.55 from the various pose methods of the original workers and 0.62 (or 0.565 if an unusual prediction is omitted) from topomer CoMFA poses. Such small averaged errors in pIC50 prediction would surely provide useful guidance in lead optimization, whether for R-group searching or simply ranking known candidates. Furthermore, the topomer CoMFA pIC50 predictions result from an almost entirely objective process, with only the fragmentation bonds being user-controllable. Finally the time and effort of performing a topomer CoMFA analysis are negligible, and the only action then needed to proceed with R-group virtual screening¹⁴ is selection of the (conventional 2D) database to be searched. It cannot be much of an exaggeration to suggest that the advent of topomer CoMFA promises to transform 3D-QSAR from being one of the more demanding CADD approaches to becoming the easiest.

On the other hand, so far all these apparent benefits of topomer CoMFA are supported only by inference from a single application. Therefore we now report additional supporting investigations, in four areas:

1) Whenever a result as unexpected as the consistent effectiveness of topomer poses in 3D-QSAR is reported, it is simple prudence to confirm that result. Therefore another one of us applied a different implementation of topomer CoMFA to ten more data sets included within an additional seven randomly encountered 3D-QSAR publications.¹⁵⁻²¹

2) The accuracy of R-group potency contributions based on topomer CoMFA has been further investigated. A novel and more stringent "leave-out-one-R-group" (LOORG) validation protocol was used to generate 1350 true pIC50 predictions (for every one of the training set structures underlying the 25 topomer CoMFA models). The accuracies of these predictions were assessed by comparison to their experimental values.

3) Usually a discovery project's actual goal is identification of the R-groups most likely to confer superior pIC50s, as contrasted with accuracy in predictions of pIC50s for R-groups expected to be similar to those already in hand. Therefore we have also investigated the probability that topomer CoMFA-based virtual screening can discriminate the most active R-groups from other similarly shaped R-groups, by applying a Receiver Operator Characteristic (ROC) approach to all of the 1103 R-group training set variations underlying the 25 topomer CoMFA models.

4) How likely is topomer CoMFA searching within typical compound collections to encounter a usefully large number and variety of apparently superior R-group alternatives? (Superiority is defined as (1) sufficient shape similarity to be a promising lead-hopping candidate and (2) a pIC50 contribution predicted to be higher than all but the most potent training set R-groups.) We performed searches within the ZINC database,¹ using the topomer CoMFA models for both R-groups from each of the 25 topomer CoMFA models, a total of 50 such searches because each of the models defines two R-groups. We report the structures of the most promising

Table 1. Data and Methodologies for the Ten 3D QSAR Literature Studies Repeated with Topomer CoMFA

Dataset Name	Biological activity	Structural class	Literature CoMFA methods			Topomer CoMFA	
			Conformer	Orient	PLS ^a	Frag Case ^b	Common Core or Fragmentation Bond
Kchnl	activation of K _{ATP} channel openers	pyrido-thiazines	SYBYL minimization of SYBYL-built models	super-position of bicyclic; field fit	A	2	
MDR	inhibition of P-glyco-protein activity	anthranil-amides	minimization of a lead-inspired conformation	super-position of aromatic, O, and N	A	2	
CB1	binding to cannabinoid receptors	cannabinoid side chain modifications	similarity to the NMR/modeling inferred conformation of a lead	RMS fit of selected common atoms	A	2	
CB2						2	
comt	inhibition of rat brain soluble COMT	3-nitro-catechol derivatives	docking into 1VID		A	2	
MAOa2	inhibition of MAO-A	coumarin derivatives	previously published protocol; includes inspection of docked conformation		B	2	
MAOb2	inhibition of MAO-B						
MAOdiff	MAO-B selectivity over MAO-A						
rvtrans2	not stated	TIBO derivatives	docking into HIV-1 RT	superposition of TIBO ring system	A	2	
EV71	anti-EV71 (caspid protein) activity	aryl imidazolones	docking into EV71 VP1		A	1	

^a A denotes the standard CoMFA method within SYBYL. B indicates use of GOLPE. ^b See description of Classes in the Introduction to ref 4.

individual R-group found, the associated pIC₅₀ predictions, and the overall retrieval rates.

METHODS

Topomer CoMFA of Ten Additional Data Sets. The software used was an alpha version of SYBYL 8.0, with the underlying methods being the same as those detailed previously,⁴ though now mostly automated, except as follows. Complete rather than R-group structures are now entered, with selection of their fragmentation bonds being the user's responsibility, albeit with considerable support from a GUI that tries to generalize the user's apparent intention, as inferred from the 2D structure surrounding a user-selected fragmentation bond. The topomer generation process enforces the standardization of ionizable groups (to the neutral protomer if possible), tautomers, and stereoisomers. Finally, to agree with "standard CoMFA", the criterion for the

optimal number of PLS components has been restored to the first maximum in q^2 .

Also as before, the data sets were simply the first ten encountered by scanning random recent journals and including every 3D-QSAR training set that was encountered. From left to right, Table 1 provides for each model a short "data set name" as its subsequent identifier, the literature reference footnote, the biological end point, the general kinds of structures, the general method originally used to generate the 3D poses, and the PLS variation used to derive the published 3D-QSAR. Whenever an article described multiple 3D-QSAR models for the same structures and biological end-point, the individual model chosen for comparison was that whose derivation conditions seemed most similar to those of "standard CoMFA".

Table 1 also provides examples for correcting an assumption that is occasionally mistakenly made about topomer

CoMFA, that it is applicable only to sets of structures sharing a “common core”. Its last column indicates, with one or two arrow(s), the one or bond(s) of fragmentation for the structures in each of these training sets, the only variable in topomer CoMFA beyond the training set composition itself. In all but one of these eleven analyses, the training set fragmentations did not postulate a common core, the “R-groups” being created instead by simply cutting a single acyclic bond. For training sets that do share a core, it might be mentioned that the topomer CoMFA software currently supports up to nine sites per core for R-group substitution.

“Leave Out One R-Group” (LOORG) Validation. The “leave-X-out” procedure (X typically being one) usual for QSAR validation²² actually ignores the architecture of typical lead series, which frequently combine many trials of a few different R-groups at one position with a few (often single) trials of many different R-groups at a second position. Moreover, the emphasis that topomer CoMFA places on prediction and searching for R-groups implies that the focus of its validation should also be on R-groups.

The “leave out one R-group” procedure takes as input a completed topomer CoMFA table, with its underlying R-group fragments representing each individual compound variation tested, and iterates through every R-group position. The first action at an R-group position is to assemble a list of every distinct structural variation that the data set includes for that R-group position. That list is then traversed, with the following steps applied to each of the R-groups in turn.

- 1) all the tested compounds having that R-group at that position are omitted;
- 2) a topomer CoMFA is derived from the resulting SAR table;
- 3) the resulting topomer CoMFA equation is used to predict the activities of the omitted compounds; and
- 4) those activity predictions are recorded, becoming the fundamental LOORG results.

When all R-group positions have thus been processed, the procedure ends. The number of activity predictions associated with each individual compound will then be the same as the number of R-group positions, or two predictions of the effect on pIC50 for every compound within each of these 25 data sets.

After its implementation within SYBYL using SPL, the above algorithm was applied to generate an additional molecular spreadsheet for each of the 25 data sets, containing in particular the pIC50 predictions of the two R-group effects and the experimentally observed pIC50 for each tested compound. Analogously to the usual method of calculating a predictive- r^2 from “standard CoMFA”, for each of the 25 series, a predictive r^2 value along with a single pIC50 prediction for each tested compound was then obtained from a regression equation expressing the experimental pIC50 as a linear function of the two R-group predictions. The predictive- r^2 for each data set is then calculated from these pIC50 predictions using the formula

$$r_{pred}^2 = 1 - \frac{\sum_{test} (y_{obs} - y_{pred})^2}{\sum_{test} (y_{obs} - \bar{y}_{pred})^2}$$

Identifying Superior R-Groups. To facilitate comparisons with other virtual screening methods such as receptor

docking, we adopted its widely used ROC (Receiver Operating Characteristics) criterion²³ to assess topomer CoMFA’s performance, for simply identifying R-groups that might confer superior pIC50. The ROC approach compares a ranked list of experimental measurements with a ranked list of independently generated predictions, here becoming a comparison between the ranked experimental pIC50s and each of its two LOORG predictions based on its two R-groups, for each structure underlying one of the 25 topomer CoMFA models. A ROC analysis requires a binary experimental result, with the goal being a successful forecast distinguishing its positive and negative outcomes. Therefore in this study, the threshold of a pIC50 value (measured or predicted) greater than 75% of the observed pIC50 range for that data set was taken as the indication of a positive experimental outcome, lower pIC50 values of course then being taken as negative. For example, if the observed pIC50 range was between 3.0 and 7.4, an experimental pIC50 greater than 6.1 would represent a positive outcome, and any other (lower) pIC50 would be a negative outcome.

The starting point for identifying superior R-groups was the molecular spreadsheet introduced above in the leave-out-one-R-group methods section. Two LOORG pIC50 predictions were made for each structure (corresponding to its R1 and R2 fragments), as the LOORG-generated prediction for an R-group added to the full topomer CoMFA model prediction with that R-group contribution omitted. Both predicted values were written into one of two files, depending on whether the compound’s observed pIC50 represented a positive or a negative outcome. These two files were the input data required by a general ROC-area calculator program, kindly supplied to us by Professor A. J. Jain (UCSF).

Topomer Searching for R-Groups Based on Topomer CoMFA. From a user’s perspective, this process simply annotates each of the hit structures from a “standard” R-group topomer shape similarity search with a prediction of its expected effect on pIC50, obtained by applying the topomer CoMFA QSAR to each sufficiently shape similar R-group (of course also requiring the calculations of atomic charges and electrostatic fields for that R-group). However there remain some important questions.

Which R-group structure(s) (within the training set used to build the topomer CoMFA) should become the shape similarity query(s) for Topomer Search,²⁴ and what should the topomer distance cutoff then be? As a matter of policy, the default parameters for the initial release of this technology were used throughout these studies, so that a pIC50 prediction was done for every R-group whose topomer distance to any one of the training set R-groups was less than 185. However, it seems likely that even more novel R-groups may be retrieved by using a larger topomer distance while selecting only the more potent training set R-groups as the topomer shape similarity queries. The inherent tension between the goals of similarity searching and the goals of any SAR study (as discussed in the second introductory paragraph) expresses itself here, in the specific context of topomer CoMFA R-group virtual screening, as the following issues: (1) whenever a structure with lower activity is the one most similar to a candidate R-group, the structural changes then needed to produce high activity are larger and therefore less likely to pass the topomer similarity distance cutoff and (2) a topomer similarity distance cutoff that is too small will limit the extrapolation

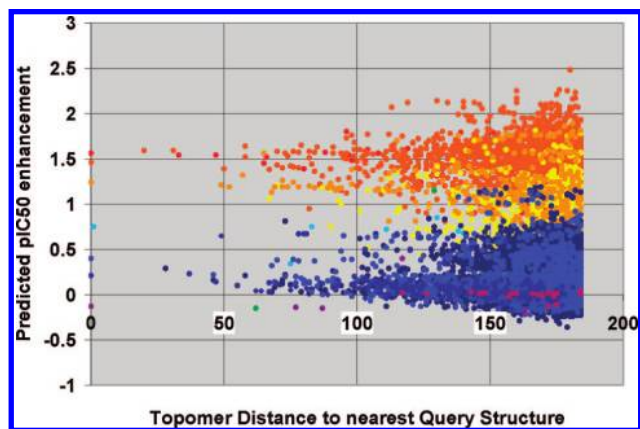


Figure 1. Dependence upon topomer distance to the closest query structure of the predicted pIC50 enhancements, for the replacement R1-groups retrieved from ZINC by the FLAV topomer CoMFA model. The redder an R1-group data point, the more active the closest query structure.

of any positive shape change effects that topomer CoMFA potentially reveals, since the R-groups with the greatest extrapolated potential for improving pIC50s may be too shape-different to be considered. On the other hand, of course, the greater the change in ligand shape, the greater the risk that activity will instead be lost altogether. Different project situations will suggest different risk/reward tradeoffs.

These theoretical expectations are empirically confirmed by our findings, where the R-groups with the highest predicted pIC50s are indeed those with the greatest shape similarity to the more active R-groups within the training set. Figure 1 illustrates this finding for an arbitrary R-group search, the R1 from the FLAV data set. It shows a plot of the predicted pIC50 values, for all the R-groups retrieved from ZINC, as a function of the smallest topomer distance of that R-group to any training set R-group. Furthermore each point is color coded by the experimental pIC50 contribution of that most similar training set R-group, in a red-to-blue scheme with red indicating the highest pIC50 contribution for that R-group. Two trends are clearly evident.

1) The larger the smallest topomer distances between a retrieved R-group and its nearest R-group within the training set, the larger the spread in the predicted pIC50 values.

2) The higher the pIC50 of that nearest training set R-group (the more red the color of a retrieved R-group data point), the higher the predicted pIC50 of the retrieved R-group.

Which candidate fragmentations are relevant in an R-group search? When the search objective is to identify R-groups rather than entire structures, the remaining fragment of the candidate structure is irrelevant. With this application even C–H bonds are reasonable sites for fragmentation, although their much greater abundance and diversity greatly extends search times and the resulting R-groups may lack synthetic accessibility. The default behavior used herein in R-group searching, whether based on shape similarity only or also including a topomer CoMFA model as in this study, is to break only those acyclic bonds and pairs of acyclic bonds that yield fragments containing at least three heavy atoms.

There is another tradeoff between speed and accuracy to consider. Conversion of a 2D structure into a 3D structure via Concord (which establishes the bond lengths, valence angles, hence ring geometries of a topomer) is the most time-

consuming step in topomer generation. When topomer searching for shape similarity only, the approximation of performing this conversion only once per entire candidate structure seemed reasonable, assuming that the breaking of the various acyclic bonds would have a negligible effect on the localized valence geometries of the resulting fragments. However for topomer CoMFA searching this approximation proved completely inadequate, with differences of several log units between the predicted pIC50s for the “same” topomer, either directly from topomer CoMFA derivation (where the topomer is always generated from a fragment) or instead encountered during R-group virtual screening. (The usual cause was a dependence of the root atom’s valence geometry on the nature of the detached atoms, particularly whenever the root atom is nitrogen.) To achieve internal self-consistency in predicted pIC50 contributions, during R-group virtual screening, whenever an R-group pIC50 is to be generated, Concord is reapplied to that individual R-group.

For all of these reasons, topomer R-group searches that include pIC50 prediction are currently about an order of magnitude slower than similarity-only queries. Most of these 50 searches took between 15 and 20 h on assorted individual “Linux box” processors, the lowest and highest timings being 13 and 36 h. (The remarkably fast search times experienced with other topomer technologies^{25,26} depend on topomer databases that are instead constructed in advance. However, the design tradeoff for the topomer search approach, intended for the database sizes in this study, instead emphasizes the ease-in-use of “no database preparation required”.

RESULTS

As a reader convenience, validation results from both this and the previous topomer CoMFA validation study are collected in Table 2. These results fall into the two general categories indicated by the major column block headings. The left block facilitates comparison of the pose results appearing in the original literature with the results obtained for the same input data with topomer CoMFA. Specifically, its paired “lit” and “top” columns compare the customary 3D-QSAR quality metrics: q^2 ; SDEP, the underlying number of PLS components; and the conventional r^2 , for all the CoMFA models as built from each of the 25 training sets. Also shown are the rms of the errors in pIC50 for every possible test set prediction. The right block provides the results most relevant to virtual screening. For the leave-out-one-R-group (LOORG) calculations, the “predictive r^2 ” (calculated as shown in the Experimental Section) and the RMS of errors in pIC50 prediction are shown. For the recognition of particularly promising R-groups, the “ROC areas”, from identifying R-groups within compounds whose measured pIC50 is in the highest 25% of the observed pIC50 range, appear in the penultimate column. The rightmost column shows the highest pIC50 predictions for the R1 and R2 replacements identified from within the ZINC database, as the sum of their predicted contributions over the best R1 and R2 pIC50 contributions generated from the corresponding training set by its topomer CoMFA model.

Table 3 provides individual results for each of the 50 R-group searches (“virtual screens”), mainly as triplets of fragment structures (each with the open valence being the horizontal “bond” at the left of any fragment structure). The

Table 2. Results from All Topomer CoMFA Validation Studies

comparison of literature alignment results with topomer CoMFA results																	
CoMFA model construction (training set)											test set predictions			LOORG and search results			
data set		no. of cpds	x-validated q ²		x-val SDEP		opt no. Cp		final r ²		rms error		pIC50 predictions		pred pIC50 > top 25% obsd range		
			lit	top	lit ^b	top	lit	top	lit	top	no. of cpds	lit ^a	top	pred r ²	RMS error	ROC areas	max incr
1	ICEc	36	0.630	0.362	0.816	1.002	6	5	0.970	0.883	9	0.568	0.740	0.494	0.886	0.801	1.15
2	ICEb	38	0.630	0.433	0.816	0.951	6	3	0.970	0.806	10	0.553	0.595	0.502	0.844	0.761	0.74
3	thrombin	72	0.687	0.533	0.594	0.726	4	4	0.881	0.838	16	0.673	0.619	0.296	0.878	0.720	0.65
4	trypsin	72	0.629	0.657	0.556	0.531	5	4	0.916	0.886	16	0.524	0.523	0.644	0.533	0.779	1.28
5	factorXa	72	0.374	0.186	0.515	0.591	3	4	0.680	0.747	16	0.278	0.340	0.163	0.593	0.449	1.67
6	MAOa	71	0.440	0.544	1.025	0.939	2	2	0.680	0.681				0.442	1.039	0.832	1.82
7	MAOb	71	0.430	0.520	1.253	1.199	2	3	0.880	0.782				0.296	1.441	0.750	3.29
8	hiv	25	0.680	0.389	0.571	0.845	3	3	0.950	0.878	7	0.823	0.449	0.120	0.991	0.509	0.39
9	a2a	78	0.541	0.226	0.563	0.742	4	3	0.817	0.555	23	0.668	0.761	0.304	0.699	0.712	3.82
10	d4	29	0.739	0.636	0.734	0.802	7	5	0.996	0.957				0.626	0.779	0.636	1.45
11	flav	38	0.752	0.763	0.475	0.495	4	5	0.969	0.952	4	0.337	1.314	0.743	0.494	0.934	3.94
12	cannab	61	0.592	0.423	0.570	0.696	4	3	0.905	0.777	6	0.452	0.540	0.374	0.720	0.741	1.44
13	ACEest	41	0.937	0.746	0.346	0.726	4	3	0.990	0.916	7	0.413	0.478	0.848	0.553	0.920	1.37
14	5ht3	61	0.645	0.295	1.193	1.804	5	2	0.913	0.519				0.239	1.875	0.678	0.74
15	rvtrans	82	0.837	0.830	0.567	0.587	4	4	0.936	0.916	19	0.791	0.608	0.830	0.578	0.920	−0.23
		847	0.636	0.503	0.706	0.842	4.2	3.5	0.897	0.806	133	0.553	0.633	0.461	0.860	0.743	1.568
16	Kchnl	131	0.682	0.718	0.469	0.442	6	12	0.842	0.869				0.574	0.522	0.880	2.02
17	MDR	49	0.564	0.422	0.445	0.450	4	2	0.874	0.608	13	0.190	0.429	0.599	0.375	0.755	−0.02
18	CB1	30	0.784	0.707	0.570	0.664	6	5	0.981	0.919				0.614	0.719	0.561	1.73
19	CB2	29	0.572	0.357	0.674	0.826	6	2	0.972	0.658				0.555	0.687	0.596	0.40
20	comt	92	0.594	0.677	0.538	0.475	5	3	0.903	0.834				0.646	0.495	0.614	1.88
21	MAOa2	34	0.830	0.586	0.561	0.630	2	5	0.899	0.909				0.320	0.771	0.568	0.97
22	MAOb2	34	0.751	0.607	0.441	0.942	2	4	0.859	0.877				0.641	0.874	0.843	0.94
23	MAOdifff	34	0.883	0.765	0.575	0.903	2	4	0.932	0.922				0.571	1.184	0.876	0.72
24	rvtrans2	50	0.704	0.725	0.326	0.867	4	4	0.972	0.894				0.566	1.067	0.702	7.48
25	EV71	20	0.813	0.563	0.298	0.456	5	4	0.991	0.929				0.366	0.516	0.680	0.81
		503	0.718	0.613	0.490	0.666	4.2	4.5	0.923	0.842	13	0.190	0.429	0.545	0.721	0.707	1.693
		1350	0.669	0.547	0.620	0.772	4.2	3.9	0.907	0.820	146	0.523	0.616	0.495	0.805	0.729	1.618

^a For ICEc, ICEb, hiv, and a2a, the individual prediction values were read from the graphs in Figure 3, Figure 3, Figure 6, and Figure 3, respectively, of the original publications. Others were taken directly from tables. ^b For MAOa, MAOb, hiv, a2a, flav, cannab, ACEest, Kchannel, CB1, CB2, and EV71 the sdep was calculated from the original variance in biological activity and the reported q^2 . Other values were taken directly from the tables.

leftmost structure in each triplet represents the R-group replacement suggestion, among those whose minimum topomer distance to any training set R-group was no greater than 185, that promised the highest predicted pIC50 contribution according to that topomer CoMFA. (It should be noted that these R-group searches did not consider either the therapeutic or the synthetic appropriateness of a suggested structure. These important concerns would of course be addressed in subsequent "2D" searches.) The middle structure is the most potent R-group in the training set used to derive that topomer CoMFA model. Between these two R-group structures is shown the difference in pIC50 contribution, between the value predicted for the leftmost structure and the "experimental value" (as generated from the training set by its topomer CoMFA model) for that most potent middle structure. The rightmost structure is the training set R-group that is topomerically most similar to the leftmost structure, with the corresponding topomeric distance to its immediate left. The total number of "highly promising" R-groups found in ZINC, those combining sufficient topomer similarity with a predicted pIC50 contribution greater than the top 25% of the experimental range in pIC50, appears in the rightmost column of Table 3.

Returning to Table 2, those rows having ID values 1–15 contain the validation results for the data sets employed in the original topomer CoMFA publication,⁷ and the remaining 16–25 are for the data sets introduced here. Two of the other

three unlabelled rows summarize these two groups and the unlabelled bottom summarizes all 25 data sets. Their values are sums for the two columns labeled "no. of cpds" and averages for the other columns. The results from applying topomer poses as CoMFA alignments to the ten additional 3D-QSAR data sets evidently have an overall statistical quality essentially equivalent to the original results, reproduced above them. Every one of the ten data sets provided a statistically significant topomer CoMFA model, making the cumulative success rate 25 statistically significant topomer CoMFA in 25 trials. The averages of the statistical parameters, in the summary lines immediately below both of the blocks, are somewhat better for the second trials, although the literature 3D-QSAR poses also provided higher statistical averages than in the first trials. These consistently strong results by a different investigator using the first released (a different) version of the topomer CoMFA software completely confirm the surprising original result of 15 CoMFA successes in 15 trials using the topomer poses.

Among these seven first-encountered additional 3D-QSAR publications there was unfortunately only a single test set prediction, for MDR. Its rms error of pIC50 prediction was 0.429, which although inferior to a remarkably small rms error of 0.19 reported in the original satisfactory is quite satisfactory in absolute terms.

The bottom line in Table 2 accumulates outcomes for all 25 data sets. The CoMFA models from the literature poses

Table 3. Results from R-Group Searching within the ZINC Data Base

R1 Searches	Most potent training set R-group		Most similar training set R-group		# of >25% obsvd range R-groups found
	Most potent R-group found	plC50 effect Structure	Tpmr dist Structure		
1 ICEc		0.67 	183 		288
2 ICEb		0.81 	183 		242
3 thrombin		0.47 	121 		8553
4 trypsin		0.56 	121 		4638
5 factorXa		0.43 	121 		2394
6 MAOa		1.80 	162 		377
7 MAOb		2.83 	164 		6808
8 hiv		0.17 	145 		3286
9 a2a		0.04 	115 		1182
10 d4		0.29 	155 		1655
11 flav		0.92 	180 		2284
12 cannab		0.08 	138 		19
13 ACEest		0.92 	184 		2218

Table 3. Continued

	Most potent R-group found	Most potent training set R-group		Most similar training set R-group		# of >25% obsvd range R-groups found
		pIC50 effect	Structure	Tpmr dist	Structure	
R2 searches						
14 5ht3		0.58		138		907
15 rvtrans		0.90		150		6350
16 Kchannel		1.70		124		565
17 MDR		-0.05		154		16
18 CB1		1.60		179		2606
19 CB2		0.36		166		344
20 comt		0.74		176		37392
21 MAOa2		0.79		111		3409
22 MAOb2		0.79		160		
23 MAOdif		0.36		111		
24 rvtrans2		8.51		169		
25 EV71		0.26		138		

Table 3. Continued

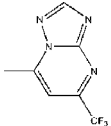
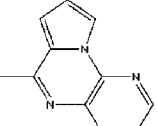
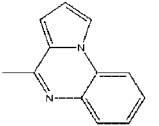
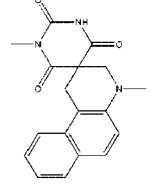
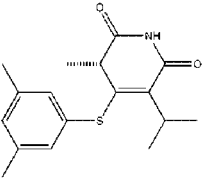
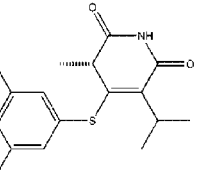
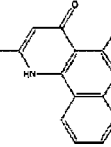
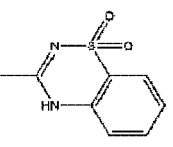
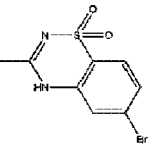
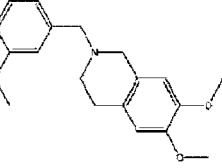
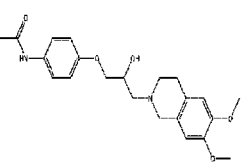
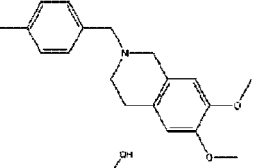
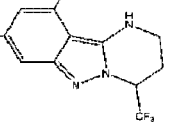
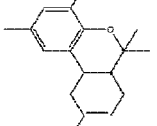
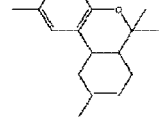
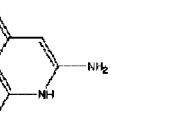
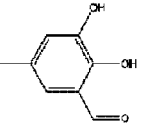
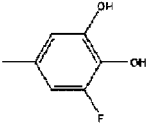
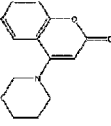
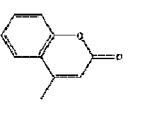
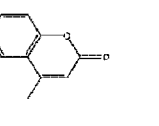
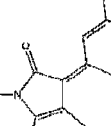
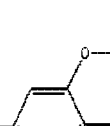
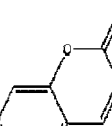
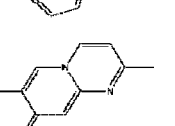
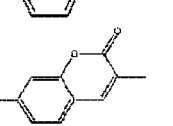
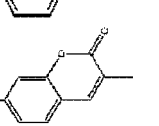
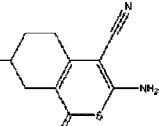
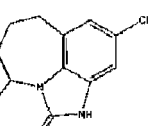
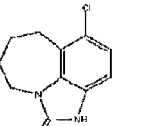
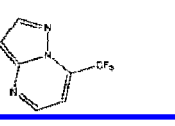
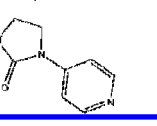
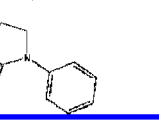



R1 Searches	Most potent R-group found	Most potent training set R-group		Most similar training set R-group		# of >25% obsvd range R-groups found
		pIC50 effect	Structure	Tpmr dist	Structure	
14 5ht3		0.16		172		706
15 rvtrans		-1.13		181		1
16 Kchnl		0.32		171		10
17 MDR		0.03		74		226
18 CB1		0.13		184		26
19 CB2		0.04				26
20 comt		1.14		168		350
21 MAOa2		0.18		121		1071
22 MAOb2		0.15		182		1899
23 MAOdiff		0.36		161		1349
24 rvtrans2		-1.03		162		0
25 EV71		0.55		161		3279

Table 3. Continued

	Most potent R-group found	Most potent training set R-group		Most similar training set R-group		# of >25% obsvd range R-groups found
		pIC50 effect	Structure	Tpmr dist	Structure	
R2 searches						
1 ICEc			0.48		144	
2 ICEb		-0.07			180	
3 thrombin		0.18			184	
4 trypsin		0.72			171	
5 factorXa		1.24			177	
6 MAOa		0.02		173		17537
7 MAOb		0.46		147		18887
8 hiv		0.22		164		2728
9 a2a		3.78		157		31960
10 d4		1.16		126		1261
11 flav		3.02		176		4030
12 cannab		1.36		55		21108
13 ACEest		0.45		101		157

do have a moderate overall superiority over those from the topomer poses, with a higher averaged q^2 , 0.669 compared to 0.543 and a lower averaged standard error in pIC50 prediction within test sets, 0.523 compared to 0.616. However these superiorities are far too small to have statistical significance. Furthermore the literature results should be better, simply as the outcome of some sort of context-relevant and q^2 -driven pose optimization, whereas topomer CoMFA results are based on a single trial of context-ignorant rule-generated poses. Yet a quarter of the individual topomer CoMFA models are superior to their manually aligned counterparts (7/25 for q^2 and 3/12 for rms test set pIC50 prediction). And of course the critical aspect of all these model-building successes is the consistent statistical significance of their outcomes rather than their comparative statistics. The topomer poses have now provided 25 satisfactory CoMFA results in 25 trials, even if so far only for data sets with which some 3D-QSAR method has already succeeded.

The average predictive r^2 of 0.495 indicates that these predictions are reducing uncertainty in pIC50 by about 50% on average, compared to the "non-prediction" that the pIC50 for any unknown structure will be the average of all values so far measured. Expressed complementarily, the average error in SDEp, the conventional leave-one-out pIC50 prediction, for the 1101 distinct R-groups, across the 17 chemical series and 23 biological assays, is 0.805. That value is less than the approximate (maximum) usefulness threshold³ of 1.0 to 1.3 proposed for the standard deviation of pIC50s within most chemical series that undergo lead optimization. Turning attention to the individual data sets, the three data sets whose predictive r^2 was less than 0.25 should not have been surprising, because the q^2 when deriving the factorXa and 5ht3 models were also very poor, while the hiv data set because of internal symmetries in many of the training structures is difficult to fragment self-consistently (this was not noticed during its original topomer CoMFA analysis, which was based on R-group structures already tabulated within the publication rather than fragmentation of complete structures). If, as would be likely in practice, no predictions are attempted for these three suspect data sets or for 5ht3 (with its q^2 of 0.295), the average predictive r^2 for the other 21 sets improves to 0.550 and the average error of pIC50 prediction to 0.760.

Because the rationale and metric of the "ROC-areas" column may be unfamiliar to many QSAR practitioners, here is a brief explanation. R-group virtual screening will be particularly helpful if its selections are more likely than others to impart superior potency. Because such a goal is already at least implicit in the validation approaches for other virtual screening methodologies, we have also applied the most widely accepted virtual screening validation approach, receiver-operator-characteristic (ROC) analysis, to the leave-out-one-R-group results. A frequent interpretation of the resulting "ROC areas" (the values tabulated in the "Superior R-group" column of Table 2) is the following: if structure A has a higher predicted value (here pIC50) than structure B, then the ROC-area may be interpreted as the probability that, when measured, the experimental value for structure A will also exceed the experimental value for structure B.²⁷ So the average ROC-area of 0.729 reported here suggests that the odds that a prediction of such a pIC50 superiority

will be confirmed experimentally are 0.729/(1-0.729), or almost three to one. Only one of these data sets, factorXa, generated a ROC area less than 0.5 (the value indicating a useless "worse than random" result), and of course the q^2 of 0.186 for its CoMFA derivation would have discouraged its actual use for prediction. Omitting all four of the suspect data sets increases the average ROC-area to 0.756.

The rightmost "max incr" column provides the overall pIC50 enhancements inferred from a combination of the most potent R1 and R2 group suggestions found among the ZINC structures. In 14 of the 25 possibilities, R-group combinations are suggested that might improve pIC50 by at least an order of magnitude, and in only 2 of those 25 possibilities was the best R-group pairing found within ZINC less potent than the most active training set structure. If the suspiciously high potency enhancement for the rvtrans2 data set is omitted, the 1.67 average of the suggested improvements in pIC50 becomes 1.38.

The individual structures shown in Table 3 demonstrate that the topomer CoMFA models are indeed effecting a significant discrimination, in that the most potent retrieved R-group is usually more shape (topomerically) similar to some other training set structure than to the most active one. Also encouraging are the large numbers of promising R-group alternatives usually reported in the rightmost "no. of.. found" column. Of course a few of the retrieved R-groups shown are far too reactive and/or toxic to be considered in a therapeutic project. However most corporate collections will not be as accommodating as ZINC to such nondruglike substructures, which can also be eliminated by substructural searches. The retrieved R-groups also seem rather likely to confer additional structural novelty.

Primarily to allow readers to assess the interpretability of the topomer CoMFA contours, a few rather arbitrarily chosen examples appear in Figures 2-4. Although only R1 groups are shown, it should be understood that separate contours are produced for each distinct set of R-groups. Each figure shows an orthonormal view of a superposition of the topomer conformations of the most active R-group proposed by a ZINC search (capped stick structure) onto the most active known R-group at the same position (wire structure), embedded within the standard CoMFA contours. The corresponding 2D structures appear in Table 3, as the left and middle of the three columns containing structures.

DISCUSSION

In summary, the potential value of R-group virtual screening, based on topomer CoMFA, seems strongly supported by the results shown in Tables 2 and 3:

- the remarkably consistent effectiveness of topomer poses in generating satisfactory CoMFA models from training sets;
- the errors of pIC50 prediction from most topomer CoMFA models being lower than an estimate of the standard deviation of experimental pIC50s within a lead series;
- the strong tendency for topomer CoMFA models to correctly identify R-groups that actually confer superior potency; and
- the many varied ideas for apparently new and more potent R-groups that emerged from most of 50 virtual screening trials with topomer CoMFA models.

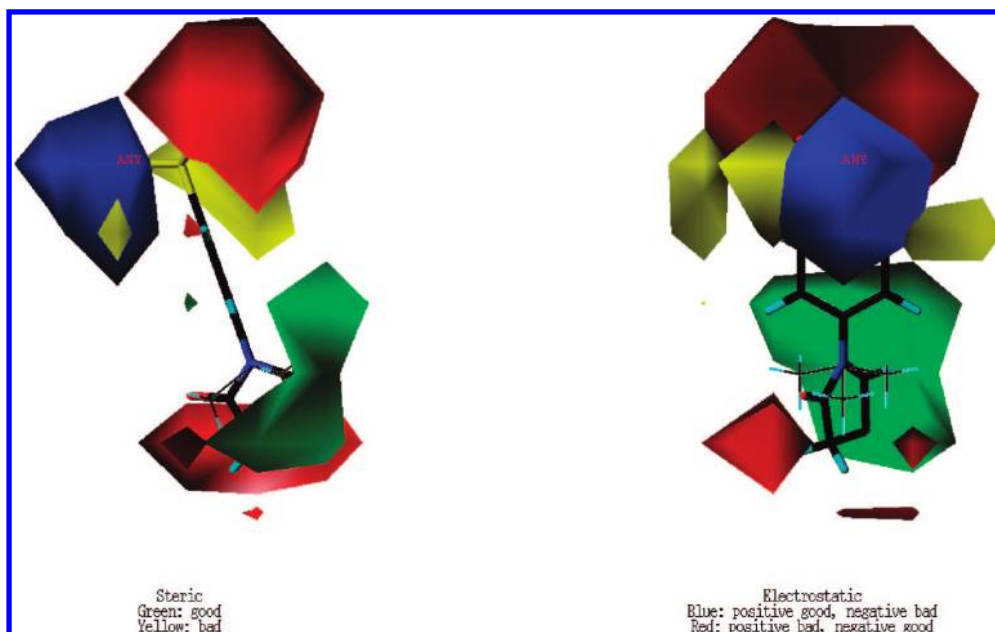


Figure 2. For the R1 group of the thrombin data set, the overlay onto the topomer CoMFA stdev x coeff contour plot of the topomers of the R-groups with the highest predicted pIC50 enhancements, the wire representation from among the training set, and the capped stick representation from the virtual screening results.

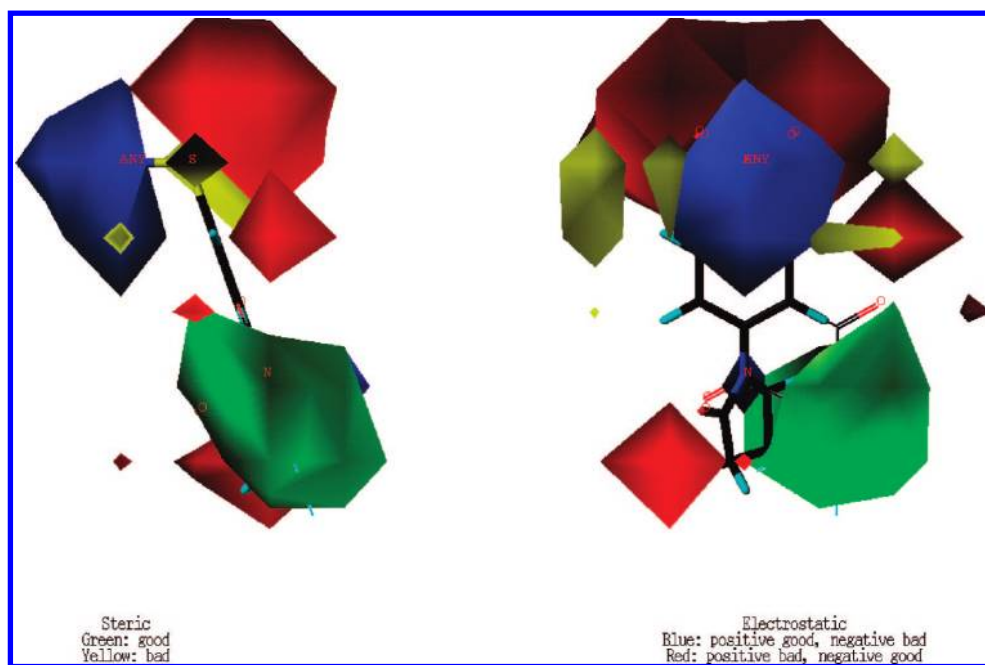


Figure 3. For the R1 group of the trypsin data set, the overlay onto the topomer CoMFA stdev x coeff contour plot of the topomers of the R-groups with the highest predicted pIC50 enhancements, the wire representation from among the training set, and the capped stick representation from the virtual screening results.

But what are the limitations on these results that also need to be considered?

Probably the most obvious limitation is that the uniform successes of topomer CoMFA poses are still limited to data sets on which a 3D-QSAR has already performed acceptably. It would also be interesting to learn how frequently topomer CoMFA works on a “typical” data set. However a discovery project’s real concern is how well topomer CoMFA will work with its particular lead series, and that question may better be answered quickly and directly, by the q^2 value that results from a topomer CoMFA trial, any value greater than 0.2 being statistically significant if probably not yet practically useful. Admittedly the challenge of data set definition must

also first be met, *i.e.*, which more-or-less related structures to include or exclude within the topomer CoMFA training set. And fragmentation bond selections must be made for all the structures in that training set. Yet encouraging progress is being made in using topomer similarity to automatically propose resolutions of these issues as well.^{28,29}

A more important limitation is concern about the actual accuracy of the pIC50 predictions that a statistically satisfactory topomer CoMFA claims to provide. This entirely reasonable concern might be factored among several somewhat independent issues, as follow:

- No matter how encouraging the statistical parameters resulting from any particular QSAR derivation might be, the

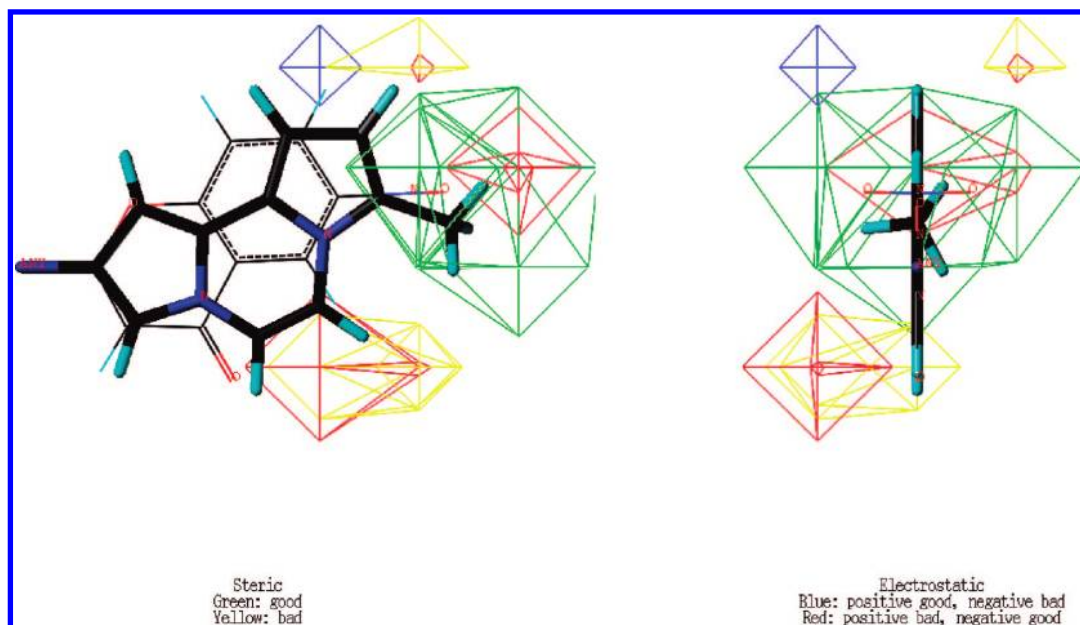


Figure 4. For the R1 group of the flav data set, the overlay onto the topomer CoMFA stdev \times coeff contour plot of the topomers of the R-groups with the highest predicted pIC50 enhancements, the wire representation from among the training set, and the capped stick representation from the virtual screening results.

dominantly disappointing outcomes from decades of QSAR predictions encourage skepticism.^{30,31}

- The interpretation of such statistical parameters^{32–34} depends at least somewhat on the appropriateness of the validation methodologies used to generate them.

- Because topomer poses are innocent of receptor-binding constraints, the appearance of the contour plots produced by topomer CoMFA is likely to disagree with the apparent structural implications of alignments based on an experimental X-ray structure. Perhaps the predictions are then less trustworthy.

- Inherent to topomer CoMFA is the perhaps questionable approximation that structural fragments can make independent and additive contributions to pIC50.

- The topomer CoMFA statistical parameters must provide predictions that are sufficiently accurate to identify promising novel R-groups.

The goal in the following discussion is to address each of these concerns as carefully and impartially as possible.

The predictive performances aggregated over all nominal “QSAR” studies have indeed not been very good. But innovative analysts have devised a very large variety of ways to computationally process an SAR table with the goal of obtaining a predictive relationship, and the merits of these vary greatly. Distinguishing field-based 3D-QSAR from QSAR approaches generally does seem to be justified by the thousands of publications reporting some degree of satisfaction with its results.³⁵ Furthermore, field-based 3D-QSAR has clear superiorities over many other QSAR approaches with respect to two long-recognized cautions, to trust only physically interpretable QSAR³⁶ (noting that the physical interpretability of 3D-QSAR’s field contours probably contributes greatly to its popularity) and to defend vigorously against the risk of chance correlation³² (noting PLS’s inherent *loss* of sensitivity as random descriptors are introduced,³⁷ and the introduction that the advent of 3D-QSAR also provided to such then-novel statistical tests as cross-validation, bootstrapping, and data scrambling²²).

Of course field-based QSAR is not immune from a general QSAR weakness, one most vividly captured by a well-known metaphor to the contrasting landscapes of Kansas and Bryce Canyon.³⁸ Every QSAR method seeks to predict future pIC50s by extrapolating continuous trends discovered among existing pIC50s. Sooner or later any such extrapolation is almost certain to encounter some serious discontinuity and provide misleading predictions (consider only the physico-chemical complexity obvious in the display of any receptor-bound ligand structure). Or put differently, at best QSAR can only reveal the “knowable knowns” within a data set. Yet is it not prudent for a discovery project to seek maximal advantage from any such locally valid “knowable known”, particularly if the costs and risks of doing so are minimal? If one does sincerely doubt the generalizations that similar structures do tend to have similar biological properties, or that reasonably well-behaved SARs are often encountered within lead series, why attempt lead optimization at all? (If such generalizations were completely futile, testing random structures would be just as (un)likely to succeed!)

Turning attention to the statistical parameters, the LOORG cross-validation approach introduced herein seems somewhat more appropriate for QSAR applications generally than conventional LOO cross-validation, at least to the first approximation that each R-group effect is independent and roughly constant. If one considers either the decision-making process that underlies a typical QSAR training set, or else simply inspects the resulting training sets, it will soon become apparent that the distribution of structural changes over the accessible sites of variation is usually not at all “independent” statistically. At one variation site, a project will often settle upon a particular R1-group after trying only two or three possibilities, and that presumably superior R1-group will then recur as dozens of variations at a second site are explored. Whenever conventional LOO cross-validation of each individual structure is applied to such a distribution of R-groups over variation sites, the influence of an oft-used R1-group on the individual models that perform each individual LOO

prediction is not omitted (its average influence is still being estimated by other training set structures containing that R1-group). The resulting bias in directly applying conventional LOO cross-validation to most QSAR data sets will be favorable—the QSAR methodology will seem more effective in predicting “unknowns” since actually the LOO-omitted nominal “unknowns” are still partially known. Such a bias will be eliminated by using the LOORG cross-validation approach. The LOORG approach will also have an appropriately higher difficulty in predicting R-group effects at those sites where one structural variation predominates, as the fewer other variations are less likely to yield a trend predictive for the presumably significant superiority of the numerically dominant R-group variation.

At least so far there is no indication that any visual discrepancy between topomer CoMFA contours and a receptor/ligand X-ray structure, or for a ligand between its topomer pose and its lowest energy conformation, affects the accuracy of predictions. Table 2 provides the most direct evidence, as it is highly unlikely that the topomer poses would so consistently yield satisfactory models if the 3D-QSAR trends that generate those LOO or LOORG predictions contradicted the trends found by the original researchers. The goal of the topomer poses, to produce the most consistently comparable shapes for ligand fragment, is further supported by the direct comparisons that other researchers have occasionally made between ligand-based and receptor-based alignments, with the ligand-based alignment usually producing the better 3D-QSAR.³⁹ To help appreciate how these “wrong” ligand poses can nevertheless produce satisfactory predictions, several additional points might be mentioned: the high likelihood that similar topomers will also be able to adopt similar conformations when receptor constrained; the stasis of the conformations in an X-ray structure, at most a snapshot of the dynamic behavior of molecules within living cells; and the general difficulty in converting such observed X-ray structures into accurate pIC50 predictions.

The interpretability of topomer CoMFA contour plots can also be assessed by reviewing Figures 2–4. Figure 2 portrays the R1 position of the thrombin data set, with the simple replacement of a *t*-butyl group by a lactam promising a 3x potency enhancement. Its contour plots suggest the cause to be a somewhat better occupancy of a large green region by the added $-\text{CH}_2\text{CH}_2-$ cycle. The same R1 position for the trypsin data set, shown in Figure 3, provides an interesting comparison to Figure 2. The most active known R-group is an anthraquinone, a very different structure from the 4-*t*-butyl-phenyl that thrombin most favors. Yet both the thrombin and trypsin models propose the same lactam as the most potent R-group found within ZINC. There do not appear to be significant differences in how well the lactam and distal anthraquinone rings occupy the large green region which both the thrombin and trypsin models generate. Perhaps the 3x potency superiority of the lactam ring results from its lack of an oxygen that the anthraquinone ring imposes between two red (negative charge disfavoring regions).

The suggested R1 replacement for the flav data set, depicted in Figure 4, represents a relatively larger departure from the training set, in both 2D structures and in forecast pIC50 contributions. The substantial potency effect of this R1 replacement is easily attributed to its methyl group

burrowing more deeply into a green region than does any portion of the most potent known inhibitors.

The assumption in topomer CoMFA, that R-groups do to a useful approximation exert independent and additive influences, also seems implicit and necessary in almost any of the qualitative SAR considerations that medicinal chemists customarily apply to lead optimization. Quantitatively this assumption is an explicit motivator of Free-Wilson analysis and an implicit assumption in Hansch analysis. Exceptions seem to be most likely whenever the properties, particular the size, of an R-group change dramatically, yet topomer CoMFA virtual screening tends to disfavor large R-group changes because of its concurrent requirement for shape similarity.

Although the emphasis in this work has been on the new application of R-group virtual screening, the pIC50 predictions obtained also seem accurate enough to make topomer CoMFA helpful in prioritizing or ranking R-group ideas already on hand (the familiar application of a 3D-QSAR). Indeed the ease of constructing a model seems to suit topomer CoMFA particularly well for the “multiple parameter optimization” goal in current lead optimization practice. One simply generates an additional topomer CoMFA model for each experimental end-point of ongoing project interest and applies all such models as additional candidate selection filters. A possible objection would be that the average value of rms error of LOORG predictions shown there, 0.805, is arguably not that much lower than our very rough estimate for the average standard deviation of experimental pIC50s for a biological activity that is actively being optimized, 1.0 to 1.3. However, the “pIC50 predictions” in Table 2 do provide strong counters to this objection, with the average predictive r^2 value of 0.495 and the average ROC area of 0.729 agreeing that actually these LOORG-predicted pIC50s are on average about halfway in accuracy between “useless, as no better than random” and “perfectly accurate”.

Once the effectiveness of searching very large structural spaces for partial (or complete) structures satisfying a variety of similarity, dissimilarity (e.g., off-target effects), and pIC50 constraints is appreciated, the question arises of how such very large structural spaces might otherwise be created. Two alternatives to ZINC-like compendia of existing structures can be mentioned:

- De novo*: random incremental modifications of input structures performed at run time.⁴⁰ The structural output is constrained primarily by some means of assessing biological promise.

- Synthesis engine: advance construction of the most synthetically accessible structural space, by iteratively applying a few reliable reactions to a few available building blocks.²⁶ The larger such a database, the more critical becomes an efficient and effective means of recognizing biological promise.

In general, these three approaches to the generation of structural space seem complementary. A large conventional structural database has the advantage of convenience and presumptive synthetic accessibility, but lacks novelty. *De novo* approaches emphasize novelty, constrained only by the “biological promise” filters, with synthetic accessibility limited to the uncertainties of a scoring function. A synthesis engine offers explicit synthetic accessibility, along with rapid and convenient access to the potential novelty of much the

most structurally diverse space, but that structural space entails higher maintenance costs. Topomer-based searches, including topomer CoMFA queries, are of course applicable to any of these three structural spaces. So are many other CADD approaches in principle, yet in practice only with topomers can the largest spaces be efficiently and effectively explored. A CADD “platform” that facilitates refinement of the still-numerous candidates that topomer CoMFA initially proposes, by applying other selection methodologies, should benefit and accelerate daily decision making in almost any drug discovery project.

CONCLUSIONS

After 25 trials the topomer poses have yet to fail in yielding satisfactory alignments for data sets to which 3D-QSAR has been successfully applied. And according to every established statistical criterion as well as some novel ones, the pIC50 predictions that these 3D-QSAR provide are about halfway between useless and perfect in their accuracies. The resulting ability to virtually screen within large compound collections for novel yet shape similar R-groups promising to confer superior potency is as far as we know entirely unprecedented. Finally, as a practical matter, topomer CoMFA followed by topomer searching seems at least as simple and convenient and rapid to carry out as any other CADD methodology. Considering also the lack of established alternatives for predicting pIC50s, it would seem very desirable for a lead optimization project to try using this new methodology to seek the maximum benefit from any “known-able knowns” buried within its accumulated SAR information.

ACKNOWLEDGMENT

We particularly thank Bob Clark and certain of the referees for helpful comments on the manuscript.

REFERENCES AND NOTES

- (1) Irwin, J. J.; Shoichet, B. K. ZINC—A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
- (2) Hillebrecht, A.; Klebe, G. Use of 3D QSAR Models for Data Screening: A Feasibility Study. *J. Chem. Inf. Model.* **2008**, *48*, 384–396. This article seems to be the first publication that explicitly explores the possibility of including a potency prediction in ligand based virtual screening, concurrently with but of course antedating this publication, and also introducing the ROC methodology to help validate the results.
- (3) In order to provide at least a very rough and preliminary estimate of this important baseline value, we calculated the standard deviation of pIC50s for the primary biological assay reported within 23 recent *Journal of Medicinal Chemistry* articles, limited to those from commercial organizations that mentioned a potential clinical candidate. The average value of these 23 pIC50 standard deviations was 0.89, with a standard deviation of 0.43. Not surprisingly, it was also apparent that in the majority of these publications the lead optimization goal in preparing and testing the most reported structures was no longer to improve potency in the primary assay but to maintain that potency while improving secondary properties. However such an activity will tend to depress the cumulative pIC50 standard deviation for the primary assay. So on balance we postulate a value somewhere between 1.0 and 1.3 for the standard deviation of any pIC50 that a project is actively seeking to improve.
- (4) Cramer, R. D. Topomer CoMFA: A Design Methodology for Rapid Lead Optimization. *J. Med. Chem.* **2003**, *46*, 374–389.
- (5) Doweyko, A. 3D-QSAR illusions. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 587–596. Actually the author proposes that 3D-QSAR predictions are not sufficiently accurate to be useful. However, the data that have been assembled and the results that are presented from its further analysis induce us to prefer the opposite conclusion. First, the average predictive r^2 over 61 data sets assembled from the literature is 0.46, which, as we agree, implies a pIC50 prediction error slightly under 1.0. Thus 0.46, though perhaps not what might be desired, is a long way from the 0.00 value which truly indicates nonpredictivity. Furthermore, Figure 1 in this publication summarizes the result of a further analysis, as a “typical” plot of predicted vs actual pIC50. Necessarily taking this plot at face value, and in contradiction of the original conclusion, the utility of the pIC50 predictions as shown seems very strong. If one seeks an actual pIC50 greater than 2.5 (structures to the right of the middle of the graph) and therefore accepts only candidates whose predicted pIC50 is greater than 2.5 (above the dotted line shown), the figure indicates that 7 of 8 (or 9) of the candidates would succeed, while only 2 (or 3) of 8 (or 9) of the failed candidates would have succeeded if pursued. We therefore fail to understand how it is “lear that this model offers no predictive value”.
- (6) 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*, 5939–5947.
- (7) To minimize semantic confusion, we concur with a current nomenclatural trend to use “pose” as a term denoting both the generation of a conformer and the spatial positioning or “alignment” of that conformer, i.e., the preparation activities required for any structure involved in 3D-QSAR.
- (8) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *24*, 215–232.
- (9) Cramer, R. D.; Clark, R. D.; Patterson, D. E.; Ferguson, A. M. Bioisosterism as a molecular diversity descriptor: steric fields of single topomeric conformers. *J. Med. Chem.* **1996**, *39*, 3060–3069.
- (10) Jilek, R. J.; Cramer, R. D. Topomers: A Validated Protocol for their Self-Consistent Generation. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1221–1227.
- (11) Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B.; Clark, R. D. “Lead-Hopping”. Validation of Topomer Similarity as a Superior Predictor of Similar Biological Activities. *J. Med. Chem.* **2004**, *47*, 6777–6791.
- (12) Cramer, R. D.; Poss, M. A.; Hermsmeier, M. A.; Caulfield, T. J.; Kowala, M. C.; Valentine, M. T. Prospective Identification of Biologically Active Structures by Topomer Shape Similarity Searching. *J. Med. Chem.* **1999**, *42*, 3919–3933.
- (13) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: a useful concept for validation of molecular diversity descriptors. *J. Med. Chem.* **1996**, *39*, 3049–60.
- (14) Cramer, R. D.; Jilek, R. J.; rews, K. M. dbtop: Topomer Similarity Searching of Conventional Databases. *J. Mol. Graphics Modell.* **2002**, *20*, 447–462.
- (15) De Tullio, P.; Dupont, L.; Francotte, P.; Counerotte, S.; Lebrun, P.; Pirotte, B. Three-Dimensional Quantitative Structure-Activity Relationships of ATP-Sensitive Potassium (KATP) Channel Openers Belonging to the 3-Alkylamino-4H-1,2,4-benzo- and 3-Alkylamino-4H-1,2,4-pyridothiazine 1,1-Dioxide Families. *J. Med. Chem.* **2006**, *49*, 6779–6788.
- (16) Labrie, P.; Maddaford, S. P.; Fortin, S.; Rakhit, S.; Kotra, L. P.; Gaudreault, R. C. A Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Analysis (CoMSIA) of Anthranilamide Derivatives That Are Multidrug Resistant. *J. Med. Chem.* **2006**, *49*, 7646–7660.
- (17) Durdagi, S.; Kapou, A.; Kourouli, T.; reou, T.; Nikas, S. P.; Nahmias, V. R.; Papahatjis, D. P.; Papadopoulos, M. G.; Mavronmoustakos, M. The Application of 3D-QSAR Studies for Novel Cannabinoid Ligands Substituted at the C1 Position of the Alkyl Side Chain on the Structural Requirements for Binding to Cannabinoid Receptors CB1 and CB2. *J. Med. Chem.* **2007**, *50*, 2875–2885.
- (18) Tervo, A. J.; Nyronen, T. H.; Ronkko, T.; Poso, A. A structure-activity relationship study of catechol-O-methyltransferase inhibitors combining molecular docking and 3D QSAR methods. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 797–810.
- (19) Catto, M.; Nicolotti, O.; Leonetti, F.; Carotti, A.; Favia, A. D.; Soto-Otero, R.; Mendez-Alvarez, E.; Carotti, A. Structural Insights into Monoamine Oxidase Inhibitory Potency and Selectivity of 7-Substituted Coumarins from Ligand- and Target-Based Approaches. *J. Med. Chem.* **2006**, *49*, 4912–4925.
- (20) Zhou, Z.; Madura, J. D. CoMFA 3D-QSAR Analysis of HIV-1 RT Nonnucleoside Inhibitors, TIBO Derivatives Based on Docking Conformation and Alignment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2167–2178.
- (21) Ke, Y.-Y.; Lin, T.-H. Modeling the Ligand-Receptor Interaction for a Series of Inhibitors of the Capsid Protein of Entovirus 71 Using Several Three-Dimensional Quantitative Structure-Activity Relationship Techniques. *J. Med. Chem.* **2006**, *49*, 4517–4525.
- (22) Cramer, R. D.; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.

- (23) Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, 27, 861–874.
- (24) Please note that this topomer similarity operation significantly differs from the shape averaging approach described in the original topomer CoMFA publication.
- (25) Andrews, K. M.; Cramer, R. D. Toward General Methods of Targeted Library Design: Topomer Shape Similarity Searching with Diverse Structures as Queries. *J. Med. Chem.* **2000**, 43, 1723–1740.
- (26) Cramer, R. D.; Soltanshahi, F.; Jilek, R.; Campbell, B. AllChem: Generating and Searching 10^{20} Synthetically Accessible Structures. *J. Comput.-Aided Mol. Des.* **2007**, 21, 341–350.
- (27) Jain, A. N. Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2008**, in press.
- (28) Cramer, R. D.; Wendt, B. Pushing the Boundaries of 3D-QSAR. *J. Comput.-Aided Drug Des.* **2007**, 21, 23–32.
- (29) Wendt, B.; Cramer, R. D. Quantitative Series Enrichment Analysis (QSEA): a novel procedure for 3D-QSAR analysis. *J. Comput.-Aided Mol. Des.* **2008**, 22, in press.
- (30) Doweyko, A. QSAR: dead or alive? *J. Comput.-Aided Mol. Des.* **2008**, 22, in press.
- (31) Johnson, S. R. The Trouble with QSAR. *J. Chem. Inf. Model.* **2008**, 48, 25–26.
- (32) Topliss, J. G.; Edwards, R. P. Chance Factors in Studies of Quantitative Structure-Activity Relationships. *J. Med. Chem.* **1979**, 22, 1238–1244.
- (33) Golbraith, A.; Tropsha, A. Beware of q^2 . *J. Mol. Graphics Modell.* **2002**, 20, 269–276.
- (34) Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSAR) from SEAL similarity matrices. *J. Med. Chem.* **1998**, 41, 2553–2564.
- (35) According to a no longer recent publication, “In the papers cited in SCI from 1989 to Dec. 7, 2000, the total number of papers with the keyword “CoMFA” are more than 5000 . Zhu, L. L.; Hou, T. J.; Chen, L. R.; Xu, X. J. 3D QSAR Analyses of Novel Tyrosine Kinase Inhibitors Based on Pharmacophore Alignment. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1032–1040.
- (36) Unger, S. H.; Hansch, C. On model-building in structure-activity relationships. A reexamination of adrenergic blocking activity of β -halo- β -arylamines. *J. Med. Chem.* **1973**, 16, 745–749.
- (37) Clark, M.; Cramer, R. D. The Probability of Chance Correlation Using Partial Least Squares. *Quant. Struct.-Act. Relat.* **1993**, 12, 137–145.
- (38) Maggiora, G. M. On Outliers and Activity Cliffs—Why QSAR disappoints. *J. Chem. Inf. Model.* **2006**, 46, 1535.
- (39) Clark, R. D. A ligand’s-eye view of protein binding. *J. Comput.-Aided Mol. Des.* **2008**, 22, 507–521.
- (40) Schneider, G.; Fechner, U. Computer-based *de novo* design of drug-like molecules. *Nature Rev. Drug Discovery* **2005**, 4, 649–660.

CI8001556