Bioorganic & Medicinal Chemistry 20 (2012) 851–858

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

An integrated computational workflow for efficient and quantitative modeling of renin inhibitors ☆

Govindan Subramanian [a,*], Shashidhar N. Rao [b]

[a] Molecular Innovative Therapeutics, Sanofi U.S., PO Box 6800, 1041, U.S. Route 202-206, Bridgewater, NJ 08807, USA
[b] Tripos, A Certara Company, 1699 South Hanley Road, St. Louis, MO 63144, USA

ARTICLE INFO

Article history:
Received 12 October 2011
Revised 22 November 2011
Accepted 28 November 2011
Available online 7 December 2011

Keywords:
CoMFA®
CoMSIA
Topomer CoMFA®
Renin inhibitors
3D-QSAR

ABSTRACT

A new integrated computational workflow that couples the strength of the molecular overlay methods to achieve rapid and automated alignments along with 3D-QSAR techniques like CoMFA® and CoMSIA for quantitative binding affinity prediction is presented. The results obtained from such techniques are compared with rule-based Topomer CoMFA® method, where possible. The developed 3D-QSAR models were prospectively used to predict the affinities of new compounds designed through R-group deconvolution starting from the core chemical scaffold and subsequent virtual combinatorial library enumeration. The general applicability of the seamless in silico modeling workflow is demonstrated using several datasets reported for small molecule inhibitors of renin.

1. Introduction

Affinity enhancement of small molecules through subtle structural and physicochemical property changes on a given chemical scaffold is a major hallmark of medicinal chemistry design activities in lead optimization (LO) efforts. While structure-guided efforts like docking offer insights into the favored steric and electrostatic environments, they are often unsuccessful in consistently providing quantitative affinity improvement estimates.[1] Although a similar scenario is exhibited by several ligand-based modeling approaches, molecular field based 3D-QSAR techniques like CoMFA®,[2] CoMSIA,[3] Topomer CoMFA®,[4] and others present a unique path forward to quantitatively model affinity/activity and offer rational insights using a 'pseudo binding site' hypothesis[5] with several success stories in scientific literature.

One of the major hurdles in developing a reasonable structure–affinity relationship (SAR) model using the above ligand-based techniques rests in the molecular alignment of the small molecules. While rigid molecules can be treated with ease, increasing molecular flexibility results in several low energy conformational alternatives that complicate the alignment and model building and hence biases the outcome. More often the small molecule flexibility issue is addressed by using the 'bound conformation' reported in a published X-ray structure, if available. However, until recently, the alignment of a large set of molecules to this kind of reference ligand structure was a rate limiting factor. With the advent of new algorithms and tools like MOE-FlexAlign™,[6] ROCS,[7] Phase Shape,[8] Surflex-Sim,[9] etc., rapid shape/feature based overlays of diverse chemical scaffolds and even large sets of analog structures are now within reach and can be used to develop quantitative 3D-QSAR models and prospectively apply them for predicting the biological property of interest.

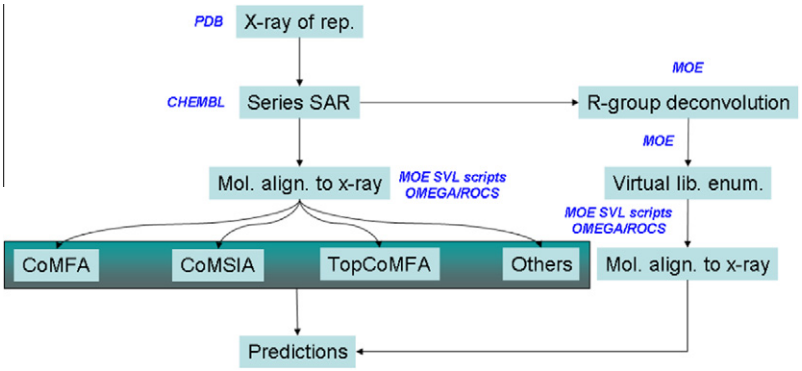
In this article, we illustrate the power of combining such techniques to achieve unique molecular alignments automatically with minimal user interference to the reference template molecule and subsequently utilize the generated alignments to rapidly and efficiently evaluate affinity in silico employing CoMFA® and CoMSIA (Scheme 1). Where permissible, the recently popular rule-based Topomer CoMFA® is also employed. We also highlight how the choice/size of the dataset and method influences the modeling outcome and how care should be taken in interpreting the results for rationalizing the predictive capability/uncertainty using published literature datasets for renin.

Renin is a clinically and therapeutically validated ~335 amino acid long aspartyl protease target for the development of antihypertensive drugs like aliskiren (trade names: Tekturna and Rasilez) in the renin-angiotensin-aldosterone system (RAAS) regulatory network.[10] Renin has long been viewed as a desirable therapeutic target as the aspartyl protease performs the initial and rate-limiting step in the RAAS. The low oral bioavailability of aliskiren (launched by Novartis) prompted other pharmaceutical companies

☆ Dedicated to Professor Eluvathingal D. Jemmis on the occasion of his 60th birthday.

* Corresponding author. Tel.: +1 908 231 3618; fax: +1 908 231 2492.
E-mail address: govisubra66@gmail.com (G. Subramanian).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.11.063



Scheme 1. Generic modeling workflow.

to develop direct renin inhibitors. Currently, only two follow-up oral drugs for renin from Takeda (TAK-272) and Vitae Pharmaceuticals (VTP-27999), respectively, seem to be under active clinical evaluation. In spite of the wealth of the knowledge reported in literature for the target as well as the small molecules and peptides, identifying a low molecular weight direct renin inhibitor that demonstrates oral efficacy is still a daunting challenge.

The ~50 X-ray structures reported for human renin bound with various types of inhibitors unequivocally provide a detailed and comprehensive view on the nature of interactions in the catalytic active site. Accordingly, the binding site for small molecules, peptides, and peptidomimetics lie in a deep cleft between the N- and C-terminal domains with a significant portion of the site solvent exposed (Fig. 1). The large size of the active site as well as the flexibility of several reported small molecule renin inhibitors (Fig. 2) poses a significant challenge for structure-based docking approaches to predict the appropriate binding mode and subsequently enable quantitative correlations of the docking scores to the observed binding affinities.[11] However, the diverse chemical scaffolds (Fig. 2) and X-ray structures reported in literature facilitate the application of ligand-based approaches to rapidly develop quantitative binding affinity models for the rational optimization and evaluation of new and potent small molecule inhibitors of renin.

By combining the structure-guided information (X-ray structures in this instance) and ligand-based methods (molecular alignments), the current article circumvents the limitations of the individual approaches and builds upon the collective strength to achieve quantitative predictions for newly designed renin inhibitors.

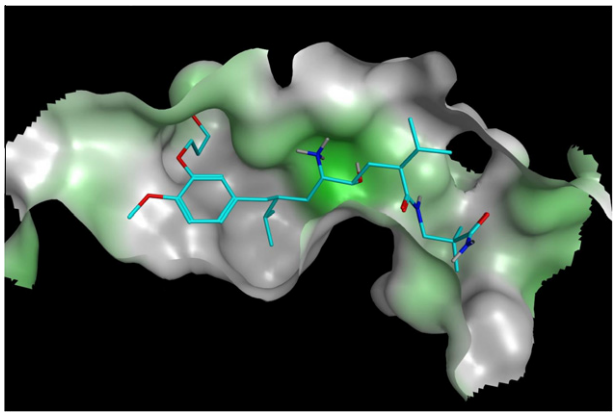


Figure 1. X-ray structure pose of Aliskiren bound to human renin (PDB: 2V0Z). Hydrophobic and hydrophilic regions are colored in grey and green, respectively.

2. Datasets and methods

Literature searches for small molecule direct renin inhibitors identified 5 chemically/structurally different datasets published by Sanofi,[12] Pfizer,[13] Merck,[14] Roche/Actelion,[15] and Vitae[16] with varying degree of small molecule inhibitor flexibility (Fig. 2) and representing 2-different protein conformational states. The small molecule structures and reported activities for these datasets were extracted from the manually curated ChEMBL chemical database (v09).[17] The recently published renin inhibitors that are still not part of this database were drawn using MDL ISIS™ Draw (v2.5 SP 2) and exported as SD file. All the 2D-structures were converted to an initial 3D-structure using MOE (v2010.10) and conformational databases for the 3D-structures were generated using OMEGA (v2.4.3) from OpenEye Scientific Software. The X-ray structure of bound renin-Aliskiren (Fig. 1; PDB code: 2V0Z) was selected for the reference coordinate frame. Representative X-ray structures reported in the literature for each of the dataset (Fig. 2) was identified and a protein structural alignment was carried out using the *structalign* script distributed with the Maestro (v9.2) software suite from Schrödinger. The inhibitor coordinates alone were extracted out from the structurally aligned PDB files and 'ligand-prepared' subsequently using MOE. The closest inhibitor structure (PDB code: 3G70/3OAD) was used to hypothesize the 'bound conformation' for the Merck datasets[14] as no X-ray structure is available in the public domain. Where possible, a ROCS (v3.1.1) search was performed using the OMEGA generated conformational databases and aligned to the X-ray coordinate of the reference inhibitor structure for that chemical series. Further refinements to the alignment were carried out using flexalign_manual.svl script obtained from Chemical Computing Group (CCG) and used within MOE. In cases where the combined OMEGA/ROCS approach did not succeed, the fragment_superpose.svl script was used to achieve a tentative alignment of the inhibitor datasets with subsequent refinements using the flexalign_manual.svl script available from CCG.[18]

The combined OMEGA/ROCS approach, as used herein, utilizes a precomputed conformational database of the small molecules considered in this study. While the approach provided satisfactory results for fairly rigid molecules, it did not succeed in retrieving the complete *bioactive/bound* conformation for flexible molecules (**2**, **4**–**9** in Fig. 2). Since the bound conformation of flexible ligands in the reported X-ray structures often adjusts and accommodates to the shape of the binding site, conformation generation engines like OMEGA and others[19] might miss to pick-up such orientations and hence will not result in the identification and capture of the maximally overlaid molecular alignment in the subsequent ROCS search. This limitation was overcome by using the *svl* scripts mentioned above that utilizes substructure/SMARTS pattern to align the molecular core rigidly to the conformation of the bound inhib-

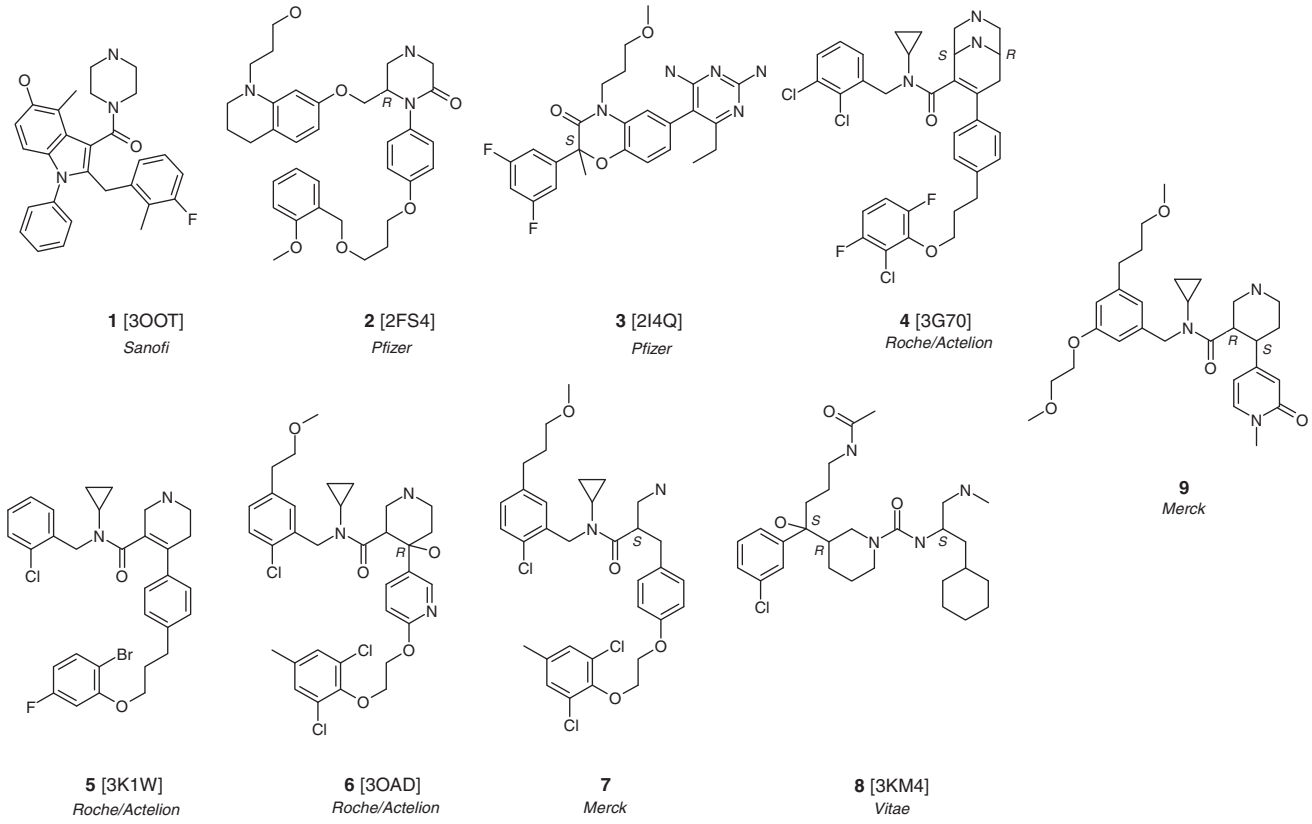


Figure 2. Representative small molecule renin inhibitors reported by different pharmaceutical/biotech companies in literature. Molecular alignments for the structural analogs in Table 1 were achieved using the bound inhibitor coordinates of the reported X-ray structures. The corresponding PDB codes are given in parenthesis. All tertiary nitrogens were considered in its protonated state.

itor chosen. The remaining parts of the molecule being aligned were subsequently minimized (part of the *svl* script) to obtain a maximal overlap with the reference structure.

Prospective predictions are provided for new compounds and libraries designed through virtual combinatorial library enumeration for the Roche/Actelion[15] and Vitae[16] dataset. For instance, R-group deconvolution was performed on the 59 Vitae training set compounds using combidecompose.svl script[18] and a virtual combinatorial library of ~2300 compounds enumerated via 3-point diversity using the *combigen* module within MOE. All the enumerated compounds were aligned to the template inhibitor structure using the *svl* scripts described above and affinities predicted using the models developed from 3D-QSAR techniques mentioned below. A similar exercise was performed for the Roche/Actelion dataset while a scaffold replacement and virtual library enumeration was carried out by morphing the core from Pfizer and R-groups from Roche/Actelion.

The molecular structures and alignments along with the experimental and predicted affinities for all the datasets are included as tables (Tables S1–S10) and SDF format files as part of the Supplementary data (may also be obtained from authors upon request) and are only referenced in the main article.

CoMFA®, CoMSIA, and Topomer CoMFA® as implemented within Sybyl-X (v1.3) were used for 3D-QSAR model building. In a standard CoMFA® approach, the biological property of interest is correlated through the steric and electrostatic properties of the molecules computed via the Lennard–Jones and Coulombic potentials, respectively. The complimentary CoMSIA technique includes the hydrophobic and hydrogen bond donor/acceptor features to supplement the steric and electrostatic indices. In contrast to the whole molecule alignments used in CoMFA® and CoMSIA field calculations, Topomer CoMFA®[4] employs structural fragments generated by splicing of molecules at user-specified acyclic bonds, to generate electrostatic and steric fields. The 3D structures of fragments are generated by a set of 'Topomer rules' which are consistently applied to all the molecules in the training and prediction sets. In a congeneric series, this allows for focusing of differences in activities between molecules to the corresponding differences in the fields brought about by different substitution patterns. Artifacts due to conformational sensitivity of the alignments are avoided in this approach.

Where possible, training and prediction sets were constructed in such a way that inhibitors representing several regions of the experimental affinity range (~4-log units) are considered, instead of assembling compounds based on structural similarity or diversity. Except for the Sanofi dataset, an 80/20 split of the literature data was used to group the training and prediction set compounds. The Sanofi dataset was treated such that data from one publication[12a] was used to develop the 3D-QSAR models with the intent of predicting and comparing the experimental outcome from another publication (akin to a prospective prediction).[12b,20] Standard physicochemical descriptors to probe any potential correlation with affinity were generated using MOE.

3. Results and discussion

Before embarking on sophisticated molecular modeling efforts, it is often advisable to verify if simple physicochemical descriptors like lipophilicity, polarity, etc. can capture the affinity trends and guide the affinity optimization efforts. Similarly, the impact of molecular size on affinity can also be assessed through the ligand efficiency indices.[21] In the current datasets, linear correlation be-

tween log *P* or polar surface area and affinity was not observed (see Supplementary data), thereby warranting the use of more sophisticated modeling techniques. Initially, the results for each dataset are discussed separately since the experimental biochemical protocol for determining renin affinity did not report a positive control that could be used to assess the inter laboratory assay variability. Additionally, the experimental uncertainties in the affinity measurements have not been published for several inhibitors and hence the prediction errors need to be considered within that scope. It should also be noted that the affinities of most inhibitors are in the (sub)nanomolar range and hence a log unit error may still be within acceptable deviations; 1 versus 10 nM is still a log unit difference. Similarly, several compounds were screened as enantiomeric mixtures, while only one enantiomer was used for model building/prediction (Tables S3–S5, S7). Such bias in selecting the so-called 'active' enantiomer was based on the comparison of the experimental observation of the mixture and the individual enantiomer activities that were reported for representatives in each chemical series. In several instances, the enantiomeric mixture activities were comparable to that of the more active individual enantiomer and hence the mixture values were used for the individual enantiomers chosen. Finally, compounds were not excluded either from the training or prediction sets (by calling them outliers) to reflect a better statistics. Table 1 summarizes the statistical outcome for the various datasets considered in this study and additional statistical measures are included in the Supplementary data (Table S1).

3.1. Sanofi dataset

Among the SAR datasets considered, the inhibitors from this group represent the most rigid molecular framework and hence easily amenable for molecular overlay techniques (**1**, Fig. 2). Thirty-four compounds from this dataset[12a] were used to train and build the model and 41 additional molecules[12b,20] that included representatives from the new 5- or 7-azaindole chemotype were used as an external prediction set (Table S2). In spite of its simple molecular architecture a reasonable model could only be achieved using CoMSIA with CoMFA® and Topomer CoMFA® yielding non-productive models (cv-r^2 values in Table 1). We partially attribute the failure of the latter technique to the limited R-group (or compound split) representation that does not capture enough or consistent field variations (activity cliffs). This reasoning was further ascertained by the improved cv-r^2 achieved when a subset of 24 inhibitors from this dataset was modeled (results not shown) with 2-points of diversity (split at indole 1- and 2-positions; **1**,

Table 1
Statistical results for the different datasets considered in CoMSIA, CoMFA® and Topomer CoMFA® approaches

Dataset[a]	N[b]	CoMSIA				CoMFA®				Topomer CoMFA®			
		r^{2c}	rmse[d]	MAE[e]	cv-r^{2f}	r^{2c}	rmse[d]	MAE[e]	cv-r^{2f}	r^{2c}	rmse[d]	MAE[e]	cv-r^{2f}
Sanofi													
Training	34	0.796	0.439	0.298	0.382	0.186	0.879	0.707	NM[g]	0.462			
Prediction	41	0.173	1.130	0.878		0.118	0.857	0.694					
Pfizer_KP[h]													
Training	42	0.924	0.284	0.221	0.564	0.467	0.747	0.579	0.284	0.908[i]	0.332[i]	0.250	0.616
Prediction	14	0.569	0.772	0.495		0.680	0.740	0.499		0.204[i]	0.993[i]	0.737	
Pfizer_DAP[j]													
Training	36	0.908	0.294	0.200	0.406	0.722	0.539	0.426	0.420				
Prediction	12	0.610	0.731	0.645		0.651	0.489	0.397					
Roche/Actelion													
Training	108	0.928	0.320	0.254	0.749	0.911	0.390	0.290	0.739	0.923[k]	0.291[k]	0.326	0.799[k]
										0.872[i]	0.363[i]	0.375	0.771[i]
Prediction	37	0.691	0.709	0.483		0.642	0.749	0.508		0.865[k]	0.488[k]	0.383	
										0.792[i]	0.554[i]	0.470	
Prediction[l]	17	0.225	1.088	0.844		0.488	1.295	1.025		0.415[k]	1.711[k]	1.622	
										0.093[i]	1.421[i]	0.948	
Vitae													
Training	59	0.664	0.586	0.439	0.306	0.732	0.526	0.392	0.337	0.918	0.370	0.298	0.527
Merck													
Training	57	0.631	0.656	0.532	0.516	0.651	0.635	0.523	0.471	0.683[i]	0.637[i]	0.530	0.521
Prediction	19	0.394	0.862	0.576		0.476	0.802	0.545		0.477[i]	0.831[i]	0.510	
Global													
Training	356	0.853	0.534	0.406	0.639	0.737	0.734	0.580	0.626				
Prediction	118	0.691	0.797	0.581		0.717	0.757	0.572					
Training[m]	334	0.813	0.588	0.456	0.811	0.784	0.669	0.528	0.625				
Prediction[n]	140	0.633	0.882	0.661		0.649	0.858	0.647					

[a] The training and prediction set compounds are denoted by *Training* and *Prediction* labels.
[b] Number of molecules used in the model development or validation.
[c] Statistical fit obtained through linear regression.
[d] Root mean square error.
[e] Mean absolute error.
[f] Cross-validated r^2 obtained using leave-one-out. The predictive capabilities of the statistical model should be interpreted and used appropriately.
[g] Statistical results are not meaningful.
[h] Ketopiperazine series.
[i] Statistical results obtained by fragmenting at the piperidine 3- or ketopiperazine 2-position.
[j] Diaminopyrimidine series.
[k] Statistical results obtained by fragmenting at the piperidine 4- or ketopiperazine 1-position.
[l] Statistical results obtained for the dataset reported in Ref. 22.
[m] Inhibitors from *Training* (108), *Training* (57), *Training* (42), *Training* (36), *Training* (34), and *Training* (59) datasets. Values of '*N*' in parenthesis.
[n] Inhibitors from *Prediction* (37), *Prediction* (17), *Prediction* (19), *Prediction* (14), *Prediction* (12), and *Prediction* (41) datasets. Values of '*N*' in parenthesis.

Fig. 2). Although quantitative predictions based on the CoMSIA model for the 41 external inhibitors demonstrate a larger rmse and weaker r^2, the model does indeed capture the affinity trends appropriately (Table S2). The current results for the prediction set do indeed suggest that the effects of atomic mutations (indole to azaindole) commonly performed in LO stages towards affinity manipulation can be captured reasonably well even though the training set inhibitors did not include such representatives. This also offers a plausible reasoning for the larger deviations in predictions when compared to experimental affinities for the prediction set, since the effect of such variations were not considered while training the model.

3.2. Pfizer dataset

The SAR for the two scaffolds reported by this group (**2**, **3**; Fig. 2) were not combined for modeling purposes as the 'ketopiperazine' framework[13b–d] (**2**) binds to a 'flap-open' protein conformation while the 'diaminopyrimidine' core scaffold[13a] (**3**) binds to a 'flap-closed' renin conformation. Moreover, the X-ray structures of representatives from both series clearly reveal that the inhibitors occupy different sub-sites in the binding pocket and hence interact differently with the binding site residues.

All the three 3D-QSAR approaches fared well (Table 1) with CoMFA® reflecting a weaker statistics for both the datasets. In spite of the small size of the training set compounds, the predictive capability of the models allow the use of the same for hypothesizing potential 'affinity zones' of new analogs. Two-points of SAR variation are reported for the 'ketopiperazine' core and hence a Topomer CoMFA® could be developed using a fragmentation scheme at the 2-position of the core. On the other hand, the SAR reported for the 'diaminopyrimidine' series presents only a single-region of variation outside the core and hence a split or fragmentation scheme was not pursued.

3.3. Roche/Actelion dataset

The SAR evolution for this dataset started with the modification to the core diazabicyclononene to finally land in the piperidine (**4**–**6**; Fig. 2) framework while maintaining similar R-groups for the non-core parts.[15] By far, inhibitors reported in this set represent the largest collection (Table S4) among the datasets considered in this study. Consequently, the results are superior and consistent across the three 3D-QSAR approaches for the training set compounds (Table 1). More strikingly, the Topomer CoMFA® seems to capture the R-group field contributions appropriately and is relatively insensitive to the type of the fragmentation scheme applied. This is also echoed in the statistical outcome for the 37 prediction set compounds with better responses than displayed by CoMFA® and CoMSIA.

Akin to Pfizer's ketopiperazine analogs,[13b–d] the renin inhibitors from Roche/Actelion also bind to the 'flap-open' conformation of the aspartyl protease and display analogous interactions with the binding site residues. Utilizing this background information, researchers at Merck contemplated an acyclic amine warhead (**7**, Fig. 2) by replacing the piperidine-like core from Roche/Actelion (**4**–**6**) and still capture the interactions with the catalytic aspartate. The SAR for the 17 compounds reported recently[22] using this approach was passed-on to the 3D-QSAR models developed using the Roche/Actelion dataset. The predictions, however, reveal weaker r^2 and higher rmse (Prediction (17) in Table 1; Table S5). Part of this statistical disparity is attributed to the 'unknown' inter laboratory experimental variation and the fact that the biochemical assay results for most compounds were for racemic mixtures and not for the single enantiomer as considered in this study. In fact, enantiospecific, **7**, is ~threefold more active compared to its racemic counterpart and the predicted pIC_{50} using this active *S*-enantiomer is close to the experimental findings (Table S5).

The promising outcome observed for this dataset motivated the design of new analogs based on the R-group deconvolution approach and using the 3-amido piperidine like core in **4**–**6**. In silico virtual combinatorial enumeration resulted in ~2900 compounds that were subsequently used for predicting renin affinity using the Topomer CoMFA® model developed for this dataset. The prospective predictions are provided as part of the Supplementary data for potential comparison with experimental finding as and when they become publicly available.

3.4. Vitae dataset

Like the Merck strategy briefed above,[22] lead finding and SAR progression for this dataset used the structural cue's extrapolated from the reported X-ray structures of a small molecule inhibitor (PDB code: 1RNE).[23] With at-least 3 diversity sites covered through the ~60 compounds reported in literature (**8**, Fig. 2), the complete dataset was used for the model building exercise (Table S6). Through the statistics reported in Table 1, it appears that the Topomer CoMFA® outperforms the traditional CoMFA® and CoMSIA approaches.

Validation of the 3D-QSAR models performance and the design of new analogs that retain or improve affinity can be achieved by deconvoluting the 59 inhibitors into a central core (with and without –OH functionality) and 3-points of SAR variation. The predictions for the ~2250 member virtual library thus designed are provided as part of the Supplementary data for comparison with the experimental affinities of inhibitors that may also have been realized synthetically, but unavailable in the public domain.

3.5. Merck dataset

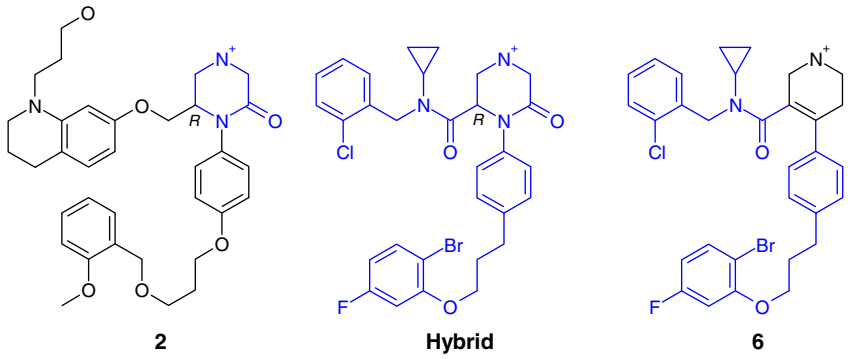
Prudent incorporation of polar and small substitutions to the Roche/Actelion piperidine scaffold[15a,b] (**6**) allowed this group to trim the 'floppy-and-greasy' region while maintaining the renin affinity. As exemplified in Figure 2 (**9**) and the recent publications,[14] the SAR for this series involved variations at the piperidine 3- and 4-positions (Table S7). Unlike the other datasets considered in this study, all the 3D-QSAR approaches pursued here performed equally well (Table 1) for the training and validation sets.

3.6. Hybrid scaffold

The recent and successful demonstration of morphing and identifying new war-heads[22] by judiciously retaining and optimizing the appropriate R-groups using known literature SAR information for renin prompted us to pursue the same path. Through a subtle analogy, the ketopiperazine warhead[13b–d] (**2**) was morphed into the Roche/Actelion scaffold[15a,b] (**6**) as described in Scheme 2. Using the 2-point R-group diversity at the 3- and 4-piperidine positions in the Roche/Actelion dataset and the ketopiperazine warhead, a new virtual combinatorial library was enumerated resulting in ~330 compounds (Table S8).

The predictions obtained using the Roche/Actelion 3D-QSAR models reveal a potential affinity spectrum (Fig. 3) clearly suggesting that only certain R-group combinations are more favored. While the subtle variations in the predicted affinity between the various 3D-QSAR models is not presented and discussed here, conventional consensus or 'weighted' approach could be pursued to identify the optimal subset of R-group combination for library synthesis efforts.

The above results provided impetus for novelty assessments and synthetic feasibility resulting in the identification of a patent filing by Roche/Actelion,[24] with the exemplification of compounds



Scheme 2. Morphing **2** and **6** to design new hybrid series.

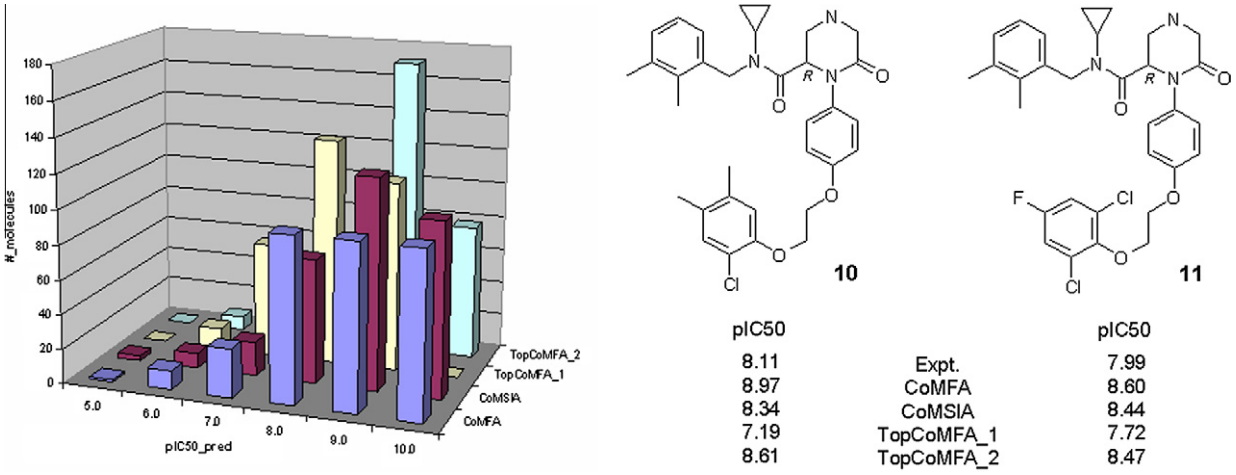


Figure 3. Predicted pIC50 affinity distribution obtained from various 3D-QSAR models for ~330 member ketopiperazine library. TopCoMFA_1 and TopCoMFA_2 are the results obtained by fragmenting the molecules at the 1-, 2-ketopiperazine positions, respectively. Experimental pIC50 for **10** and **11** were obtained from Ref. 24.

containing substitutions that were not considered in this library design (e.g., **10**, **11**; Fig. 3). A comparison of the experimental and predicted renin affinity values for **10** and **11** offer enough incentive to explore this new warhead and R-group decorations that takes advantage of the residues in the $S3^{sp}$ pocket of renin.

3.7. Global model

Although the 3D-QSAR local models developed above for structurally diverse renin inhibitor would seem to be more appropriate, the reported dataset represent very limited SAR. This prevents a plausible widespread application of the models and limits the model scope to systems that fall within similar overall shape. Consequently, a global model using all the reported chemotypes (Fig. 2) investigated here was pursued. While the inter-laboratory variation in the measurements are being forgiven, the model permits future investigation and prioritization of new chemical scaffolds similar to a docking approach, as the potential contribution of several sub-site residues to the overall binding may be captured (Fig. 4) in a more balanced manner.

In spite of using a mixed dataset that does not exhibit a maximal overlap of the inhibitors commonly hypothesized in many molecular superposition/alignment methods,[25] the statistical results from both CoMFA® and CoMSIA are extremely good (Table 1 and Table S9) and comparable between these two 3D-QSAR methods. The cross-validated r^2, rmse and the MAE are reasonable

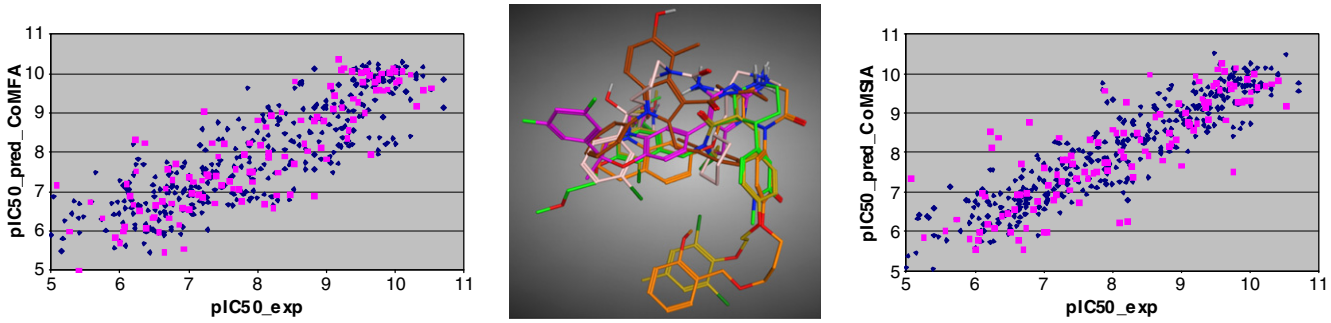


Figure 4. Scatter plot for experimental versus predicted pIC50 affinity by CoMFA® (left) and CoMSIA (right) for 356 training set (blue) and 118 prediction set (magenta) renin inhibitors. The structural diversity of the chemotypes considered are shown through the bound conformation of representatives (middle).

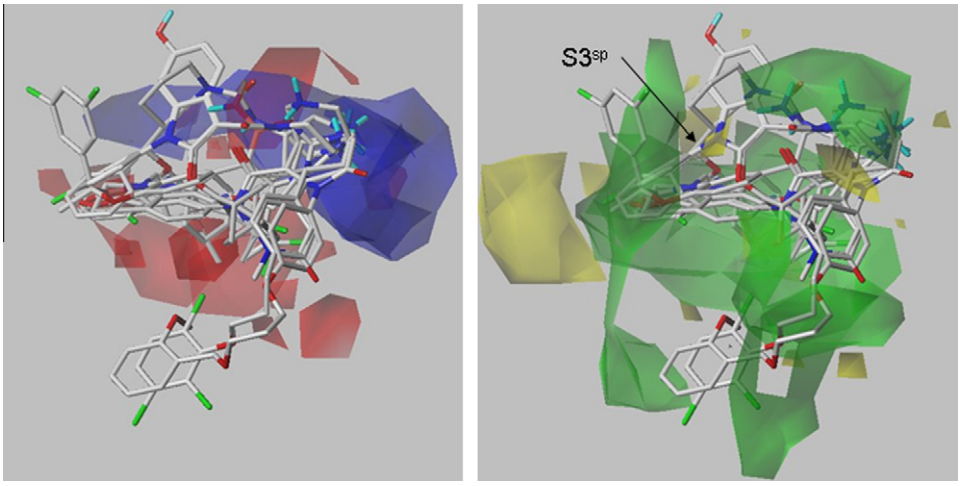


Figure 5. CoMFA® 3D-contour maps around representative actives in the global model. The standard electrostatic (left) and steric (right) fields that rationalize the variation in renin inhibition are shown.

and within the expectations of the model's predictive capability.[26] Figure 5 illustrates the CoMFA® contour plots with similar features exhibited by the CoMSIA model as well (not shown).

It is interesting to note that the performance of the models does not diminish by changing the membership of the training and prediction set inhibitors. This additional model developed by combining all the training set compounds from the local models (and in a similar vein, the prediction set molecules) resulted in a quantitative model well suited for predictive purposes (Table 1; Table S10).

The steric and electrostatic fields captured by the 3D-QSAR model are in accordance with what would be anticipated from the perspective of interactions with protein active site residues. For instance, the narrow $S3^{sp}$ sub-pocket that would not allow branched substitutions is represented by the unfavorable steric fields (yellow region with pointed arrow in Figure 5).

4. Conclusion and outlook

Several pertinent activities pursued during a typical LO effort have been addressed and modeled quantitatively by leveraging application tools from disparate vendors and streamlining them to enhance the efficiency and accuracy of the model development. In one instance, the results emanating from traditional medicinal chemistry COST (changing one substitution at a time) approach have been used to design virtual combinatorial libraries and identify/predict affinity pockets within the permissible SAR landscape achievable synthetically by permuting various R-group combinations. The results for the Vitae and Roche/Acetlion (see Supplementary data) dataset demonstrate the power of the approach. To our knowledge, this is the first time where a CoMFA® or CoMSIA approach has been reported on such a large dataset.

The extrapolation of the developed quantitative models to suggest new substitution patterns has also been exemplified through the Sanofi prediction set as the approach is field-based and that the structural patterns are not memorized. Our effort here was to predict and propose modifications that would retain the affinity, but also provide avenues that would minimize other 'off-target' liabilities frequently encountered in the LO stage and not discussed here.

By morphing the core-scaffold and R-groups from different chemical series, new novelty avenues could be achieved and affinities predicted prior to synthetic efforts. This has been show-cased through the hybridization of the Roche/Actelion and Pfizer structures. A similar scheme could be applied to toggle part of the Merck (**9**) and Pfizer (**2**) structures, but is beyond the scope of this article. Such approaches seem to have been pursued in literature,[24] but the current study provides a much broader scope and quantification for prioritizing valuable chemical resources.

As would be expected from any QSAR modeling approach, the predictive capability for **10** and **11** demonstrates that the model interpolates reasonably well for closely related scaffolds. However, the extent of extrapolation or the model applicability domain to new chemical scaffolds and unforeseen substitution patterns needs further investigation before the model could be generalized. For instance, most of the peptide and peptide mimic inhibitors including Aliskiren occupy the P′-sub-sites of the renin active site. This region was not captured by the dataset considered in this study and hence predictions for compounds that occupy these sub-sites should be treated with caution.

In general, all the 3D-QSAR approaches considered here perform equally well and no distinct statistical advantage was observed for any individual method. However, as mentioned above, the size of the dataset used to train the model as well as the type of substitution patterns seen in the training and prediction sets seem to contribute to the quantitative activity prediction errors, but, none-theless are not completely mispredicted if a classification scheme were to be applied. This is clearly evident for the global models shown in Figure 4 (also see Supplementary data for the local model results).

Finally, the intent of this work was not to perform in-depth analysis of the various models and approaches as it pertain to renin, but to present an alternate workflow that resolves the bottleneck associated with the molecular alignment commonly cited as a barrier to the extended use of the CoMFA® and CoMSIA approaches. Since these two approaches consider the compound property as a whole with surprisingly good results for the global model, new and alternate scaffolds identified through shape-based or docking approaches can be evaluated quantitatively instead of relying on the qualitative scoring schemes. This facilitates a new priority criterion or filter that could be set to minimize potentially top ranking false-positives obtained from shape-based and docking approaches in the quest for scaffold-hoping based on literature or in-house SAR information. Efforts along this direction are on-going and the results will be reported in a future publication.

Acknowledgments

G.S. thanks the help and application support provided by Alain Deschenes from Chemical Computing Group, James from OpenEye Scientific Software, Steven Dixon from Schrödinger, Richard Cra-

mer and Bernd Wendt from Tripos. The authors also thank Suresh Singh from Vitae Pharmaceuticals for validating the predictions with their unpublished experimental results and the reviewers for the constructive suggestions to improve the manuscript quality.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.11.063.

References and notes

1. Sotriffer, C.; Matter, H. In Sotriffer, C., Ed.; Virtual Screening: Principles, Challenges, and Practical Guidelines; Wiley-VCH Verlag GmbH: Weinheim, 2011; Vol. 48, pp 177–221.
2. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
3. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
4. Cramer, R. D. *J. Med. Chem.* **2003**, *46*, 374.
5. Verma, J.; Khedkar, V. M.; Coutinho, E. C. *Curr. Top. Med. Chem.* **2010**, *10*, 95.
6. http://www.chemcomp.com/.
7. Grant, J. A.; Gallardo, M. A.; Pickup, B. T. *J. Comput. Chem.* **1996**, *14*, 1653.
8. Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. *J. Comput. Aided Mol. Des.* **2006**, *20*, 647.
9. Jain, A. N. *J. Med. Chem.* **2004**, *47*, 947.
10. Webb, R. L.; Schiering, N.; Sedrani, R.; Maibaum, J. *J. Med. Chem.* **2010**, *53*, 7490.
11. Politi, A.; Durdagi, S.; Moutevelis-Minakakis, P.; Kokotos, G.; Mavromoustakos, T. *J. Mol. Graphics Modell.* **2010**, *29*, 425.
12. (a) Scheiper, B.; Matter, H.; Steinhagen, H.; Stilz, U.; Böcskei, Z.; Fleury, V.; McCort, G. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6268; (b) Matter, H.; Scheiper, B.; Steinhagen, H.; Böcskei, Z.; Fleury, V.; McCort, G. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5487.
13. (a) Powell, N. A.; Ciske, F. L.; Cai, C.; Holsworth, D. D.; Mennen, K.; Van Huis, C. A.; Jalaie, M.; Day, J.; Mastronardi, M.; McConnell, P.; Mochalkin, I.; Zhang, E.; Ryan, M. J.; Bryant, J.; Collard, W.; Ferreira, S.; Gu, C.; Collins, R.; Edmunds, J. J. *Bioorg. Med. Chem.* **2007**, *15*, 5912; (b) Holsworth, D. D.; Cai, C.; Chen, X.-M.; Cody, W. L.; Downing, D. M.; Erasga, N.; Lee, C.; Powell, N. A.; Edmunds, J. J.; Stier, M.; Jalaie, M.; Zhang, E.; McConnell, P.; Ryan, M. J.; Bryant, J.; Li, T.; Kasani, A.; Hall, E.; Subedi, R.; Rahim, M.; Maiti, S. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2500; (c) Powell, N. A.; Clay, E. H.; Holsworth, D. D.; Bryant, J. W.; Ryan, M. J.; Jalaie, M.; Edmunds, J. J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4713; (d) Powell, N. A.; Clay, E. H.; Holsworth, D. D.; Bryant, J. W.; Ryan, M. J.; Jalaie, M.; Zhang, E.; Edmunds, J. J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2371.
14. (a) Chen, A.; Cauchon, E.; Chefson, A.; Dolman, S.; Ducharme, Y.; Dubé, D.; Falgueyret, J.-P.; Fournier, P.-A.; Gagné, S.; Gallant, M.; Grimm, E.; Han, Y.; Houle, R.; Huang, J.-Q.; Hughes, G.; Jûteau, H.; Lacombe, P.; Lauzon, S.; Lévesque, J.-F.; Liu, S.; MacDonald, D.; Mackay, B.; McKay, D.; Percival, M. D.; St-Jacques, R.; Toulmond, S. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3976; (b) Chen, A.; Campeau, L.-C.; Cauchon, E.; Chefson, A.; Ducharme, Y.; Dubé, D.; Falgueyret, J.-P.; Fournier, P.-A.; Gagné, S.; Grimm, E.; Han, Y.; Houle, R.; Huang, J.-Q.; Lacombe, P.; Laliberté, S.; Lévesque, J.-F.; Liu, S.; MacDonald, D.; Mackay, B.; McKay, D.; Percival, M. D.; Regan, C.; Regan, H.; St-Jacques, R.; Toulmond, S. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3970.
15. (a) Corminboeuf, O.; Bezencon, O.; Remeň, L.; Grisostomi, C.; Richard-Bildstein, S.; Bur, D.; Prade, L.; Strickner, P.; Hess, P.; Fischli, W.; Steiner, B.; Treiber, A. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6291; (b) Corminboeuf, O.; Bezencon, O.; Grisostomi, C.; Remeň, L.; Richard-Bildstein, S.; Bur, D.; Prade, L.; Hess, P.; Strickner, P.; Fischli, W.; Steiner, B.; Treiber, A. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6286; (c) Remeň, L.; Bezencon, O.; Richard-Bildstein, S.; Bur, D.; Prade, L.; Corminboeuf, O.; Boss, C.; Grisostomi, C.; Sifferlen, T.; Strickner, P.; Hess, P.; Delahaye, S.; Treiber, A.; Weller, T.; Binkert, C.; Steiner, B.; Fischli, W. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6762; (d) Bezencon, O.; Bur, D.; Weller, T.; Richard-Bildstein, S.; Remeň, L.; Sifferlen, T.; Corminboeuf, O.; Grisostomi, C.; Boss, C.; Prade, L.; Delahaye, S.; Treiber, A.; Strickner, P.; Binkert, C.; Hess, P.; Steiner, B.; Fischli, W. *J. Med. Chem.* **2009**, *52*, 3689; (e) Güller, R.; Binggeli, A.; Breu, V.; Bur, D.; Fischli, W.; Hirth, G.; Jenny, C.; Kansy, M.; Montavon, F.; Müller, M.; Oefner, C.; Stadler, H.; Vieira, E.; Wilhelm, M.; Wostl, W.; Märki, H. P. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1403; (f) Vieira, E.; Binggeli, A.; Breu, V.; Bur, D.; Fischli, W.; Güller, R.; Hirth, G.; Märki, H. P.; Müller, M.; Oefner, C.; Scalone, M.; Stadler, H.; Wilhelm, M.; Wostl, W. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1397.
16. (a) Xu, Z.; Cacatian, S.; Yuan, J.; Simpson, R. D.; Jia, L.; Zhao, W.; Tice, C. M.; Flaherty, P. T.; Guo, J.; Ishchenko, A.; Singh, S. B.; Wu, Z.; McKeever, B. M.; Scott, B. B.; Bukhtiyarov, Y.; Berbaum, J.; Mason, J.; Panemangalore, R.; Cappiello, M. G.; Bentley, R.; Doe, C. P.; Harrison, R. K.; McGeehan, G. M.; Dillard, L. W.; Baldwin, J. J.; Claremon, D. A. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 694; (b) Tice, C. M.; Xu, Z.; Yuan, J.; Simpson, R. D.; Cacatian, S.; Flaherty, P. T.; Zhao, W.; Guo, J.; Ishchenko, A.; Singh, S. B.; Wu, Z.; Scott, B. B.; Bukhtiyarov, Y.; Berbaum, J.; Mason, J.; Panemangalore, R.; Cappiello, M. G.; Müller, D.; Harrison, R. K.; McGeehan, G. M.; Dillard, L. W.; Baldwin, J. J.; Claremon, D. A. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3541.
17. https://www.ebi.ac.uk/chembldb/index.php.
18. http://www.svl.chemcomp.com/.
19. (a) Boström, J.; Greenwood, J. R.; Gottfries, J. *J. Mol. Graphics Modell.* **2003**, *21*, 449; (b) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. *J. Chem. Inf. Model.* **2010**, *50*, 572; (c) Chen, I.-J.; Foloppe, N. *J. Chem. Inf. Model.* **2010**, *50*, 822; (d) Musafia, B.; Senderowitz, H. *Expert Opin. Drug Discov.* **2010**, *5*, 943.
20. Steinhagen, H.; Scheiper, B.; Matter, H.; Stilz, H. U.; McCort, G. WO 2009/095163.
21. Abad-Zapatero, C.; Perišić, O.; Wass, J.; Bento, P.; Overington, J.; Al-Lazikani, B.; Johnson, M. E. *Drug Discovery Today* **2010**, *15*, 804.
22. Chen, A.; Bayly, C.; Bezencon, O.; Richard-Bildstein, S.; Dubé, D.; Dubé, L.; Gagné, S.; Gallant, M.; Gaudreault, M.; Grimm, E.; Houle, R.; Lacombe, P.; Laliberté, S.; Lévesque, J.-F.; Liu, S.; MacDonald, D.; MacKay, B.; Martin, D.; McKay, D.; Powell, D.; Remeň, L.; Soisson, S.; Toulmond, S. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2204.
23. Rahuel, J.; Priestle, J. P.; Grütter, M. G. *J. Struct. Biol.* **1991**, *107*, 227.
24. Bezencon, O.; Boss, C.; Bur, D.; Cormin-Boeuf, O.; Grisostomi, C.; Remen, L; Richard-Bildstein, S.; Weller, T. WO 2007/034445.
25. Nicholls, A.; McGaughey, G. B.; Sheridan, R. P.; Good, A. C.; Warren, G.; Mathieu, M.; Muchmore, S. W.; Brown, S. P.; Grant, J. A.; Haigh, J. A.; Nevins, N.; Jain, A. N.; Kelley, B. *J. Med. Chem.* **2010**, *53*, 3862.
26. Cramer, R. D. *J. Comput. Aided Mol. Des.* **2011**, *25*, 197.