Journal of Molecular Graphics and Modelling 29 (2011) 843–864

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Discovery of new renin inhibitory leads via sequential pharmacophore modeling, QSAR analysis, *in silico* screening and *in vitro* evaluation

Afaf H. Al-Nadaf[a], Mutasem O. Taha[b,*]

[a] *Department of Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Science University, Amman, Jordan*
[b] *Drug Discovery Unit, Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan*

ARTICLE INFO

Article history:
Received 11 July 2010
Received in revised form 31 January 2011
Accepted 3 February 2011
Available online 13 February 2011

Keywords:
Renin
Pharmacophore modeling
QSAR
In silico screening

ABSTRACT

The renin–angiotensin–aldosterone system is a major target for the clinical management of hypertension. Development of renin inhibitors has proven to be problematic due to poor bioavailability and complex synthesis. In this study, we combined pharmacophore modeling and quantitative structure–activity relationship (QSAR) analysis to explore the structural requirements for potent renin inhibitors employing 119 known renin ligands. Genetic algorithm and multiple linear regression analysis were employed to select an optimal combination of pharmacophoric models and physicochemical descriptors to yield self-consistent and predictive QSAR. Two binding pharmacophore models emerged in the optimal QSAR equation ($r^2_{96} = 0.746$, *F*-statistic = **43.552**, $r^2_{LOO} = 0.697$, r^2_{PRESS} against 23 test inhibitors = **0.527**). The successful pharmacophores were complemented with exclusion spheres to optimize their receiver operating characteristic curve (ROC) profiles. The QSAR equations and their associated pharmacophore models were validated by the identification and experimental evaluation of new promising renin inhibitory leads retrieved from the National Cancer Institute (NCI) structural database. The most potent hits illustrated IC_{50} value of 2.6 μM. Successful pharmacophore models were found to be comparable with crystallographically resolved renin binding pocket.

1. Introduction

The renin–angiotensin system (RAS) is a major regulator of blood pressure; it has a renoprotective role and an important role in the vascular response to injury [1]. Renin belongs to the aspartic proteases, which comprise one of the four primary classes of peptide cleaving enzymes [2]. It is secreted by the kidneys in response to decrease in circulating volume and blood pressure. Renin cleaves the substrate angiotensinogen to form the inactive angiotensin I. Angiotensin I is converted to the pro-hypertensive agent angiotensin II by angiotensin converting enzyme (ACE) [3]. Therefore, renin is a key player in the renin–angiotensin system, and its manipulation provides a means for the therapeutic treatment of hypertension and heart failure [4,5].

Renin inhibition is an attractive target for drug intervention due to its remarkable specificity for its substrate, which should reduce unwanted interactions and side effects [4]. In contrast, ACE is implicated in several pathways, and therefore, its inhibition results in side effects [2,3,6].

Earlier renin inhibitors were mainly peptidomemtic [7,8]. However, the unfavorable pharmacokinetic behavior of peptidomemtic inhibitors prompted continuous efforts towards developing non-peptidic renin inhibitors. These culminated in the development of aliskiren, Fig. 1, a nonpeptidic sub-nanomolar renin modulator of acceptable oral bioavailbility [2,9].

Nevertheless, development of renin inhibitors has faced many problems: the high cost of synthesis, low and variable bioavailability and lack of appropriate animal models [10,11]. Development efforts relied mainly on classical rational drug design [10,12,13] and structure-based concepts [9,11,14] with only one ligand-based exception that relied on CoMFA and CoMSIA [15]. A recent study combined docking and QSAR to attempt developing predictive model for rennin inhibitors [73]. The first renin crystallographic structure was determined by Sialecki et al. at a resolution of 2.5 Å [15]. Subsequent crystallographic studies achieved better resolutions [16–22].

The continued interest in designing new renin inhibitors combined with drawbacks of structure-based design [23–27], e.g., limitations in dealing with the induced fit flexibility of renin [11,19], and inappropriateness of CoMFA and CoMSIA models as search queries to mine virtual three-dimensional (3D) structural databases [15,28], prompted us to explore the possibility of developing ligand-based 3D pharmacophore(s) integrated within

* Corresponding author. Tel.: +962 6 5355000x23305; fax: +962 6 5339649.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.02.001

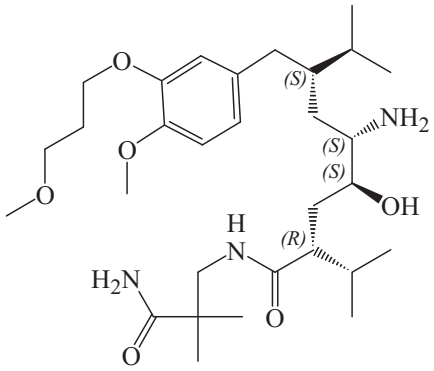


Fig. 1. Aliskiren.

self-consistent QSAR model. The pharmacophore model(s) can be used as 3D search query(ies) to mine 3D libraries for new renin inhibitors, while the associated QSAR model can be used to predict the bioactivities of captured hits and therefore prioritize them for *in vitro* evaluation. We previously reported the successful use of this combination to probe the induced fit flexibilities of activated factor X (fXa) [29] and towards the discovery of new inhibitory leads against glycogen synthase kinase 3β (GSK-3β) [30], hormone sensitive lipase (HSL) [31], bacterial MurF [32], protein tyrosine phosphatase 1B (PTP 1B) [33] and influenza neuraminidase [34], beta-secretase [35] and cholesteryl ester transfer protein [36].

We employed the HYPOGEN module from the CATALYST software package to construct numerous plausible binding hypotheses for renin inhibitors [20–22,37–40]. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analysis were employed to search for optimal QSAR that combines high-quality binding pharmacophores with other molecular descriptors and capable of explaining bioactivity variation across a collection of diverse renin inhibitors.

QSAR-selected pharmacophores were further validated by comparing them with crystallographic structures of renin bound to known inhibitors, and evaluating their ability to successfully classify a list of compounds as actives or inactives (i.e., by assessing their receiver operating characteristic (ROC) curves). Subsequently, the optimal pharmacophores were complemented with exclusion spheres to enhance their ROC profiles. Thereafter, the resulting exclusion spheres-complemented models were used as 3D search queries to screen the national cancer institute (NCI) virtual molecular database for new renin inhibitors.

2. Materials and methods

2.1. Molecular modeling

CATALYST models drug–receptor interaction using information derived only from the drug structure [28,41–45]. HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen-bond donors and acceptors (HBDs and HBAs, respectively), aliphatic and aromatic hydrophobes (Hbic features), positive and negative charges, positive and negative ionizable groups and aromatic planes. CATALYST pharmacophores have been used as 3D queries for database searching and in 3D-QSAR studies [29–32,46].

2.1.1. Software and hardware

The following software packages were utilized in the present research.

- CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA.
- CERIUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA.
- CS ChemDraw Ultra 6.0, Cambridge Soft Corp. (http://www.cambridgesoft.Com), USA.
- Pharmacophore and QSAR modeling studies were performed using CATALYST (HYPOGEN module) and CERIUS2 software suites from Accelrys Inc. (San Diego, California, www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a dual 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system. Structure drawing was performed employing ChemDraw Ultra 6.0 which was installed on a Pentium 4 PC.

2.1.2. Data set

The structures of 119 renin inhibitors (Fig. 2 and Table A under Supplementary Materials) were collected from articles published by a single research group [20–22,37–40], which strongly supports the notion that their *in vitro* bioactivities were determined by a single assay procedure. The bioactivities were expressed as the concentration of the test compound that inhibited the activity of renin by 50% (IC_{50}). Table A in Supplementary Information and Fig. 2 show the structures and IC_{50} values of the considered inhibitors. The logarithm of measured IC_{50} (nM) values were used in pharmacophore modeling and QSAR analysis, thus correlating the data linear to the free energy change.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra, installed on a PC, and saved in MDL-mol file format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for conformational analysis.

2.1.3. Conformational analysis

The molecular flexibilities of the collected compounds were taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor (**1–119**, Fig. 2 and Table A under Supplementary Materials) was explored adopting the "best conformer generation" option within CATALYST. Default parameters were employed in the conformation generation procedure, i.e., a conformational ensemble was generated with an energy threshold of 20 kcal/mol from the local minimized structure which has the lowest energy level and a maximum limit of 250 conformers per molecule [41].

2.1.4. Pharmacophoric hypotheses generation

All 119 molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an "Uncertainty" value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor [43,45,47]. CATALYST requires the uncertainty parameter for two purposes: (i) as means to classify the training compounds into most-active, moderate and inactive categories, which is essential for the three modeling phases of CATALYST (see SM-1 under Supplementary Materials), and (ii) as means to define the tolerance sizes of the binding features' spheres in the resulting pharmacophore. The default value of this parameter is 3. Typically,

CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least three and a half logarithmic cycles [43,45,47]. Four structurally diverse training subsets (Table 3) were carefully selected from the collected compounds for pharmacophore modeling.

Each training subset was utilized to conduct four modeling runs to explore the pharmacophoric space of renin inhibitors. Different hypotheses were generated by altering the interfeature spacing and the number of allowed features in the resulting pharmacophores (see Table B under Supplementary Materials).

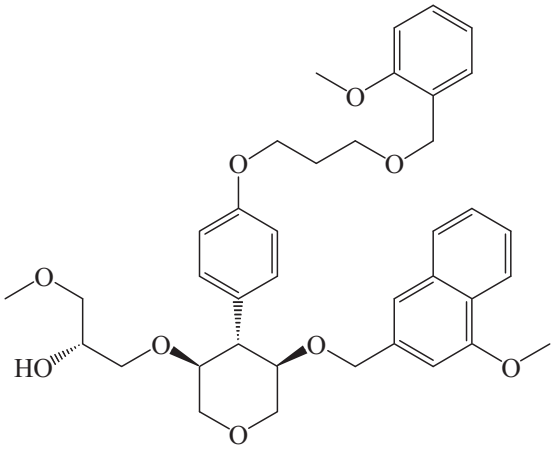
Eventually, our pharmacophore exploration efforts (16 automatic runs, Table 3 and Table B in Supplementary Materials) culminated in 160 pharmacophore models of variable qualities (see SM-1 under Supplementary Materials for details about CATALYST pharmacophore generation algorithm [43,45,47]).

2.1.5. Assessment of the generated hypotheses

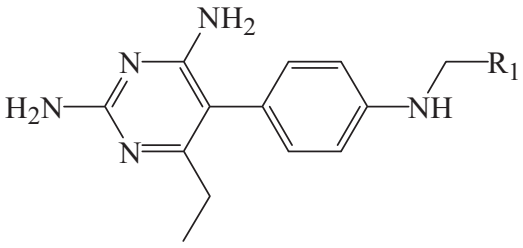
When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost and configuration cost (see SM-2 pharmacophore assessment in CATALYST under Supplementary Materials [41,45,48]). In a successful automatic modeling run, CATALYST ranks the generated models according to their total costs [41].

CATALYST also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the cost of the null hypothesis, the more likely that the hypothesis does not reflect a chance correlation.

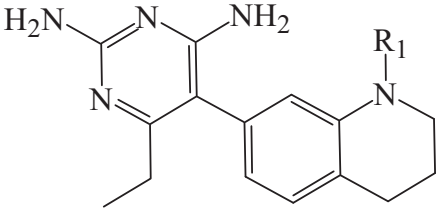
CATALYST-HYPOGEN implements additional approach to assess the quality of resulting pharmacophores, namely, the Cat-Scramble



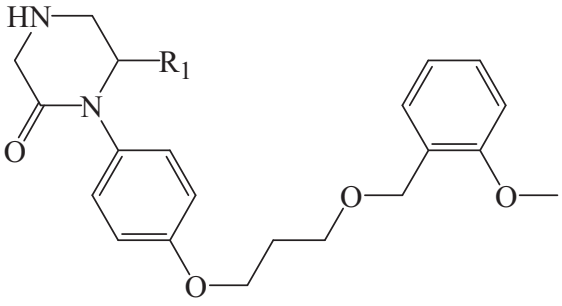
1



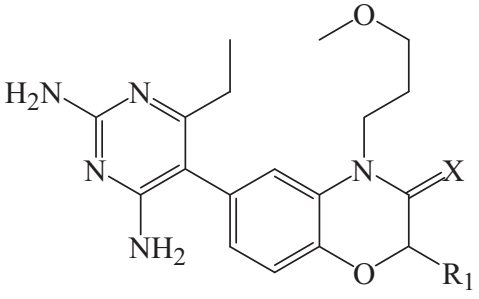
2-3



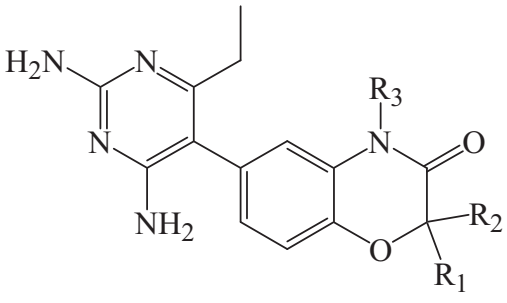
4-8



9-20

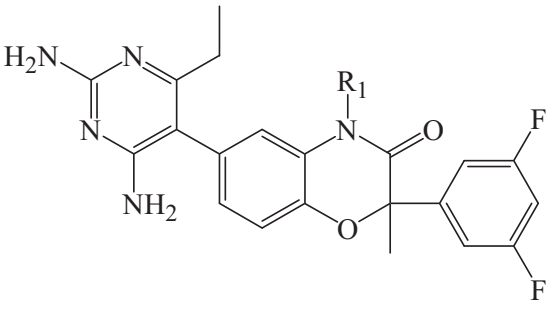


21-34

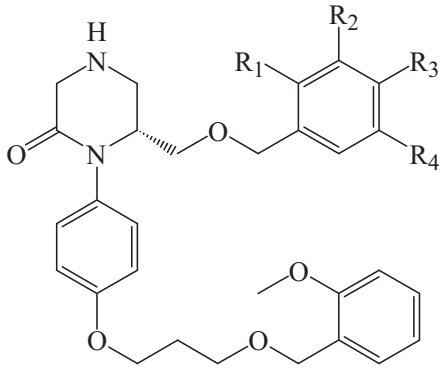


35-50

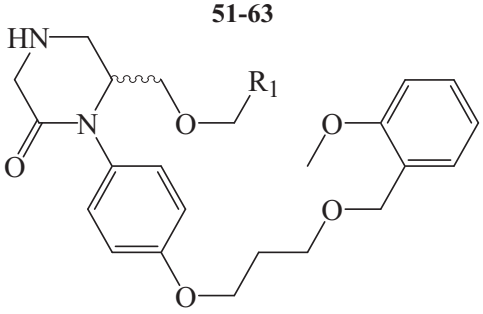
Fig. 2. The chemical scaffolds of training compounds, the corresponding structures and bioactivities are as in Table A in Supplementary Materials.



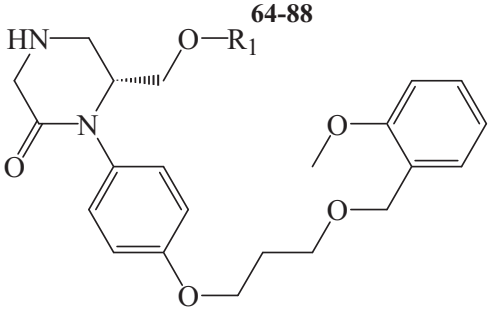
51-63



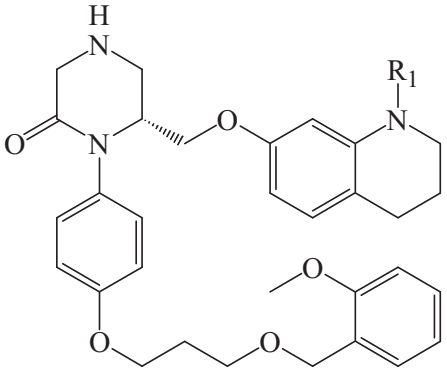
64-88



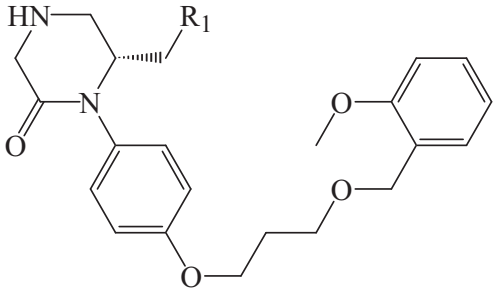
89-94



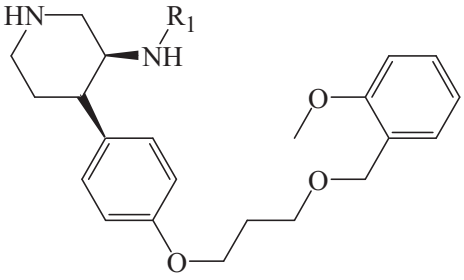
95-98



99-107



108-111



112-119

Fig. 2. (*Continued*).

approach. This validation procedure is based on Fisher's randomization test [47]. In this test, a 95% confidence level was selected, which instructs CATALYST to scramble the bioactivities of training compounds to generate 19 random spreadsheets. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating pharmacophore models from unscrambled bioactivity data. Success in generating pharmacophores of comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training

compounds and their respective pharmacophore models [41,59]. Only 62 pharmacophores, out of 160 generated models, were found to possess Fisher confidence values ≥85%. Table C in Supplementary Materials shows the success criteria of representative pharmacophores from each run.

2.1.6. Clustering of the generated pharmacophore hypotheses

The successful models (62) were clustered into 12 groups utilizing the hierarchical average linkage method available in CATALYST. Therefore, closely related pharmacophores were grouped in five-membered clusters. Subsequently, the highest-ranking representatives, as judged based on their fit-to-bioactivity correlation r^2-values (calculated against collected compounds **1–119**), were selected to represent their corresponding clusters in subsequent QSAR modeling (Table C in Supplementary Materials).

2.1.7. QSAR modeling

A subset of 96 compounds from the total list of inhibitors (**1–119**) was utilized as a training set for QSAR modeling. However, since it is essential to assess the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 23 molecules (*ca*. 20% of the dataset) were employed as an external test subset for validating the QSAR models. The selected test inhibitors are: **10**, **11**, **21**, **33**, **36**, **45**, **46**, **48**, **61**, **65**, **72**, **76**, **84**, **88**, **101**, **102**, **106**, **107**, **108**, **111**, **114**, **115** and **116** (numbers are as in Table A in Supplementary Materials and Fig. 2).

The test molecules were selected as follows: the collected inhibitors (**1–119**, Table A in Supplementary Materials and Fig. 2) were ranked according to their IC_{50} values, and then every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set.

The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors were 128 properties (see section SM-4 under Supplementary Materials) that included various simple and valence connectivity indices, electro-topological state indices and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, energies of the lowest and highest occupied molecular orbitals, etc.) [47]. Furthermore, the training compounds were fitted (using the Best-fit option in CATALYST) against the representative pharmacophores (12 models, Table C in Supplementary Materials), and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically via equation (D) in Supplementary Materials [41].

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF) [47]. However, to avoid overwhelming GFA-MLR with large number of poor descriptors; we removed 50% of those showing lowest variance prior to QSAR analysis.

We were obliged to normalize the potencies of the training compounds via division by their corresponding molecular weights, i.e., ligand efficiency ($Log(1/IC_{50})$/Mwt), to achieve reasonable self-consistent QSAR models [35,49,50].

Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear, quadratic and spline equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30,000 and lack-of-fit (LOF) smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 4 to 20 explanatory terms.

All QSAR models were validated employing leave one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}), leave 25%-out cross-validation ($r^2_{L25\%O}$) and predictive r^2 (r^2_{PRESS}) calculated from the randomly selected external test subset (see selection criteria mentioned earlier).

In $r^2_{L25\%O}$ procedure the training set is divided into two subsets: fit and test subsets. The test subset is randomly selected to represent 25% of the training compounds. This procedure is repeated over four cycles; accordingly, four test subsets with their complementary fit subsets were selected for the particular QSAR model. The four test subsets should cover 100% of the training compounds by avoiding selecting the same compound in more than one test subset. The fit sets are then utilized to generate four QSAR sub-models using the same descriptors. The resulting sub-models are then utilized to predict the bioactivities of the corresponding testing subsets. Finally, the predicted values of all four test subsets are correlated with their experimental counterparts to determine the corresponding $r^2_{L25\%O}$.

On the other hand, predictive r^2_{PRESS} is defined as:

$$r^2_{PRESS} = \frac{SD - PRESS}{SD} \quad (1)$$

Where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

2.1.8. Receiver operating characteristic (ROC) curve analysis

Successful QSAR-selected pharmacophore models (i.e., Hypo1/5 and Hypo1/7) were validated by assessing their abilities to selectively capture diverse renin inhibitors from a large list of decoys employing ROC analysis.

Therefore, it was necessary to prepare valid evaluation structural database (testing set) that contains an appropriate list of decoy compounds in combination with diverse list of known active compounds. The decoy list was prepared as described by Verdonk and co-workers [51,52]. Briefly, the decoy compounds were selected based on three basic one-dimensional (1D) properties that allow the assessment of distance (D) between two molecules (e.g., i and j), namely: (1) the number of hydrogen-bond donors (NumHBD); (2) number of hydrogen-bond acceptors (NumHBA) and (3) count of nonpolar atoms (NP, defined as the summation of Cl, F, Br, I, S and C atoms in a particular molecule). For each active compound in the testing set, the distance to the nearest other active compound is assessed using their Euclidean distance (equation (2)):

$$D(i,j) = \sqrt{(\mathrm{NumHBD}_i - \mathrm{NumHBD}_j)^2 + (\mathrm{NumHBA}_i - \mathrm{NumHBA}_j)^2 + (\mathrm{NP}_i - \mathrm{NP}_j)^2} \quad (2)$$

The minimum distances are then averaged over all active compounds (Dmin). Subsequently, for each active compound in the testing set an average of 20 decoys were randomly chosen from the ZINC database [53]. The decoys were selected in such a way that they did not exceed Dmin distance from their corresponding active compound.

Moreover, to further diversify the actives members, i.e., to avoid close similarity among actives in the testing set, any active compound having zero distance ($D(i,j)$) from other active compound(s) in the testing set were excluded. Active testing compounds were defined as those possessing renin affinities (IC_{50} values) ranging

from 0.067 nM to 100 nM. The testing set included 12 active compounds and 238 ZINC compounds.

The testing set (250 compounds) was screened by each pharmacophore for ROC analysis employing the "Best flexible search" option implemented in CATALYST, while the conformational spaces of the compounds were generated employing the "Fast conformation generation option" implemented in CATALYST. Compounds missing one or more features were discarded from hit lists. The *in silico* hits were scored employing their fit values (best-fit values) as calculated by equation (D) in Supplementary Materials. Subsequently, hit lists were used to construct ROC curves for corresponding pharmacophores (see section SM-3 ROC analysis in Supplmenetary Materials [51,54–56]).

2.1.9. Addition of exclusion volumes

To account for the steric constraints of the binding pocket and to optimize the ROC curves of our QSAR-selected pharmacophores, it was decided to add exclusion volumes to Hypo1/5 and Hypo1/7 employing the HIPHOP-REFINE module of CATALYST. HIPHOP-REFINE uses inactive training compounds to add exclusion spheres to resemble the steric constraints of the binding pocket. It identifies spaces occupied by the conformations of inactive compounds and free from active ones. These regions are then filled with excluded volumes [29–33,46].

In HIPHOP-REFINE the user defines how many molecules must map the selected pharmacophore hypothesis completely or partially through controling the Principal and Maximum Omitted Features (MaxOmitFeat) parameters. Active compounds are normally assigned a MaxOmitFeat parameter of zero and Principal value of two to instruct the software to consider all their chemical moieties to fit them against all the pharmacophoric features of the particular hypothesis. On the other hand, inactive compounds are allowed to miss one (or more) features by assigning them a MaxOmitFeat of one (or two) and Principal value of zero. Moderately active compounds are normally assigned a principal value of one and a MaxOmitFeat of zero or one to encode their intermediate status.

A subset of training compounds was carefully selected for HIPHOP-REFINE modeling. It was decided to consider IC_{50} of 10 nM as an arbitrary activity/inactivity threshold. Accordingly, inhibitors of IC_{50} values $\leq$10 nM were regarded as "actives" and were assigned principal and MaxOmitFeat values of two and zero, respectively, while less active inhibitors were assigned principal values of one or zero [29–33,46] and were carefully evaluated to assess whether their lower potencies are attributable to missing one or more pharmacophoric features, i.e., compared to active compounds (MaxOmitFeat = 1 or 2), or related to possible steric clashes within the binding pocket (MaxOmitFeat = 0). HIPHOP-REFINE was configured to allow a maximum of 150 exclusion spheres to be added to the generated pharmacophoric hypotheses.

2.2. Fluorometric quantification of renin activity

SensoLyte™ 520 Renin Assay Kit was used for the assay using a Mc-Ala/Dnp FRET peptide [57]. The sequence of this peptide is derived from the cleavage site of renin. In this peptide, the fluorescence of Mc-Ala is quenched by Dnp. Upon cleavage into two separate fragments by renin, the fluorescence of Mc-Ala is recovered, and can be monitored at excitation/emission λs of 490/520 nm.

Test compounds and renin solutions were added into the microplate wells and incubated at 37 °C for 30 min. Subsequently, 50 μL renin substrate solution were added into each well. The reagents were subsequently mixed thoroughly by shaking the plate gently for 30 s. The fluorescence intensity was immediately measured continuously and recorded every 5 min for 15 min at 37 °C. Appropriate positive and negative controls were prepared. The reaction rates were determined from the slopes of absorbance versus time plots constructed from 4 time points: 0, 5 10, and 15 min. Blank and standard inhibitor (Ac-His-Pro-Phe-Val-Sta-Leu-Phe-NH2) [57] were used as negative and positive controls, respectively.

3. Results and discussion

CATALYST enables automatic pharmacophore construction by using a collection of molecules with activities ranging over a number of orders of magnitude. CATALYST pharmacophores (hypotheses) explain the variability of bioactivity with respect to the geometric localization of the chemical features present in the molecules used to build it. Different hypotheses were generated for a series of renin inhibitors. A total of 119 compounds were used in this study (Fig. 2 and Table A in Supplementary Materials). Four training subsets were selected from the collection (Table 3). Each subset consisted of inhibitors of wide structural diversity. The biological activity in the training subsets spanned from 3.5 to 4.0 orders of magnitude. Genetic algorithm and multiple linear regression statistical analysis were subsequently employed to select an optimal combination of complementary pharmacophores capable of explaining bioactivity variations among all inhibitors.

3.1. Data mining and conformational coverage

The literature was surveyed to collect as many structurally diverse renin inhibitors as possible. However, the fact that pharmacophore and QSAR modeling necessitates that the training compounds should have been assayed by a single bioassay procedure restricted us to certain published inhibitors (**1–119**, see Table A in Supplementary Materials and Fig. 2) [20–22,37–40]. Nevertheless, in order to assess the structural diversity of the collected compounds, and hence their aptness for pharmacophore and QSAR modeling, we calculated several diversity-related parameters for the collected list and compared them with closely related list of compounds extracted from the ZINC database (329 compounds) [53], which we used as decoys for ROC analysis (see Section 2.1.8). Table 1 summarizes the calculated parameters of the two lists together with the implemented calculation methodologies. The comparison in Table 1 suggests that our training set is much more diverse than most QSAR training sets, which are normally very limited in their structural modifications.

Furthermore, clustering through maximal dissimilarity partitioning using several physicochemical descriptors and connectivity fingerprints classified the collected compounds into 12 different clusters (Table 2), which further points the chemical diversity of the collected list. Molecular diversity is essential for efficient pharmacophore and QSAR modeling, and to unveil different binding modes assumed by diverse binding ligands.

The 2D structures of the inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations. The resulting 3D structures were used as starting points for conformational analysis and in the determination of various molecular descriptors for QSAR modeling. The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within the CONFIRM module of CATALYST [58] and via the "Best" module to ensure extensive sampling of conformational space. Efficient conformational coverage guarantees minimum conformation-related noise during pharmacophore generation and validation stages [58].

Table 1
Molecular diversity among training compounds (**1–119**) compared to diversity within the decoy molecules in the ROC list (see Section 2.1.8).

Diversity parameters		Collected compounds[e]	Decoys in the ROC list[f]
Normalized number of assemblies[a]		0.26	0.68
Normalized number of fingerprint features[b]		9.85	17.47
Fingerprint distances[c]	Minimum	0	0
	Maximum	0.95	0.98
	Average	0.66	0.87
Property distances[d]	Minimum	0	0
	Maximum	1.03	1.19
	Average	0.46	0.47

[a] Defined as the total number of Murcko assemblies [71] divided by the number of library molecules [72].
[b] Calculates the specified fingerprint for each input molecule and counts the total number of unique fingerprint features collected over all the molecules, then it averages the number of unique fingerprint features per molecule. In this case we implemented extended connectivity molecular fingerprints of a maximum length of six bonds (ECFP_6) [72].
[c] Calculates the distance (defined as 1 − similarity) between every pair of input molecules based on Tanimoto similarity coefficient calculated using ECFP_6 fingerprints. Minimum distance: the minimum distance observed between any pair of molecules in the set. Average distance: the average distance for all pairs of molecules in the set. Maximum distance: the maximum distance observed between any pair of molecules in the set [72].
[d] Defined as the Euclidean distance for every pair of molecules based on the following numerical properties: Log *P*, molecular weight, number of hydrogen-bond donors, number of hydrogen-bond acceptors, number of rotatable bonds, number of rings, number of aromatic rings, molecular fractional polar surface area [72].
[e] Training compounds **1–119** (Fig. 2 and Table A in Supplementary Material).
[f] The inactive decoys employed in ROC analysis. See Section 2.1.8.

Table 2
Clustering of training compounds **1–119** (Fig. 2 and Table A in Supplementary Materials) based on connectivity fingerprints and calculated physicochemical properties[a,b].

Cluster			Members[c]
Number	Centre[c]	Size[d]	
1	**82**	31	**64**, **65**, **66**, **67**, **68**, **69**, **70**, **71**, **72**, **73**, **74**, **75**, **76**, **77**, **78**, **79**, **80**, **81**, **82**, **83**, **84**, **85**, **86**, **87**, **88**, **89**, **90**, **91**, **92**, **93**, **94**
2	**28**	8	**22**, **26**, **27**, **28**, **29**, **35**, **38**, **39**
3	**111**	3	**109**, **110**, **111**
4	**2**	2	**2**, **3**
5	**1**	1	**1**
6	**20**	8	**10**, **20**, **43**, **59**, **60**, **61**, **62**, **63**
7	**113**	8	**112**, **119**, **113**, **114**, **115**, **116**, **117**, **118**
8	**23**	9	**4**, **5**, **6**, **7**, **8**, **21**, **23**, **24**, **25**
9	**17**	19	**13**, **14**, **15**, **16**, **17**, **18**, **37**, **48**, **98**, **99**, **100**, **101**, **102**, **103**, **104**, **105**, **106**, **107**, **108**
10	**31**	5	**30**, **31**, **32**, **33**, **34**
11	**42**	18	**11**, **19**, **36**, **40**, **41**, **42**, **45**, **46**, **47**, **49**, **51**, **52**, **53**, **54**, **55**, **56**, **57**, **58**
12	**44**	7	**9**, **12**, **44**, **50**, **95**, **96**, **97**

[a] Clustering was performed based on ECFP_6 (extended connectivity fingerprints of a maximum of six bonds) combined with several physiochemical descriptors: Log *P*, molecular weight, number of hydrogen-bond donors, number of hydrogen-bond acceptors, number of rotatable bonds, number of rings, number of aromatic rings, molecular fractional polar surface area.
[b] The clustering was done by a relocation method based on maximal dissimilarity partitioning implementing a combination of root-mean-square (RMS) difference for physicochemical descriptors and Tanimoto distance for connectivity fingerprints [72].
[c] Compounds' numbers as in Fig. 2 and Table A (in Supplementary Materials).
[d] Number of compounds in corresponding cluster.

3.2. Exploration of renin pharmacophoric space

CATALYST-HYPOGEN enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over 3.5 orders of magnitude [41–45,48]. The fact that we have a list of 119 renin inhibitors of evenly spread bioactivities over more than 3.5 orders of magnitude prompted us to employ HYPOGEN algorithm to identify as many binding modes assumed by renin inhibitors as possible.

HYPOGEN implements an optimization algorithm that evaluates large number of potential binding pharmacophores through perturbations to hypotheses that survived the constructive and subtractive phases of the modeling algorithm [45] (see CATALYST pharmacophore generation algorithm under Supplementary Materials). The number of evaluated pharmacophoric models is reflected by the configuration (Config.) cost calculated for each modeling run. It is generally recommended that the Config. cost of any HYPOGEN run not to exceed 17 (corresponding to 2^{17} hypotheses to be assessed by CATALYST) to guarantee thorough analysis of all models [44].

The size of the investigated pharmacophoric space is a function of the structral diversity of training compounds, selected chemical features and other control parameters. Restricting the extent of explored pharmacophoric space should improve the efficiency of optimization via allowing effective evaluation of limited number of pharmacophoric models. On the other hand, extensive restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypotheses, as they might occur outside the "boundaries" of the pharmacophoric space.

Therefore, we decided to explore the pharmacophoric space of renin inhibitors within plausible "boundaries" through sixteen HYPOGEN automatic runs performed on four carefully selected training subsets (i.e., from the collected compounds): subsets A, B, C, and D in Table 3. The training compounds in these subsets were selected in such away to guarantee maximal 3D diversity and continuous bioactivity spread over more than 3.5 logarithmic cycles. To ensure sufficient molecular diversity within training subsets, member compounds were selected in such a way that each structural cluster of the collected compounds (Table 2) was sampled at least once in each training subset, as in Table 3.

Furthermore, the training subsets were selected in such a way that differences in their anti-renin bioactivities are primarily attributable to the presence or absence of pharmacophoric features (e.g., HBA or HBD or Hbic or RingArom) rather than steric shielding and/or bioactivity-enhancing or -reducing auxiliary groups (e.g., electron donating or withdrawing groups). We gave special emphasis to the 3D diversity of the most-active compounds in each training subset (Table 3) because of their significant influence on the extent of the evaluated pharmacophoric space during the constructive phase of HYPOGEN algorithm.

Guided by our reasonably restricted pharmacophoric exploration concept, we instructed HYPOGEN to explore only 4- and 5-featured pharmacophores, i.e., ignore models of lesser number of features (as shown in Table B in Supplementary Materials). The later restriction has the advantage of narrowing the investigated pharmacophoric space and representing the feature-rich nature of renin ligands.

The resulting binding hypotheses from each HYPOGEN run were automatically ranked according to their corresponding "total cost" value, which is defined as the sum of error cost, weight cost and configuration cost [41–45,48]. Eventually, 160 pharmacophore models emerged from 16 automatic HYPOGEN runs, out of which only 62 models illustrated confidence levels ≥85% (Fisher scrambling criteria, see Section 2.1.5 Assessment of the generated hypotheses under Experimental) [41–45,48]. These successful models were clustered and their best 12 representatives were used in subsequent QSAR

Table 3
The training subsets employed in exploring the pharmacophoric space of renin inhibitors, numbers correspond to compounds in Fig. 2 and Table A in Supplementary Materials.

Training sets	Most-active[a]	Moderate active	Least active[b]	Structural clusters[c]
A	**98**, **99**, **103**, **109**	**9**, **18**, **20**, **75**, **97**, **100**, **106**, **112**, **115**	**22**, **64**, **118**, **92**, **94**	1, 2, 3, 6, 7, 9, 12
B	**41**, **42**, **49**, **55**	**4**, **5**, **7**, **6**, **26**, **30**, **32**, **34**, **36**, **52**, **53**, **56**	**3**, **21**	2, 4, 8, 10, 11
C	**98**, **103**, **108**, **109**	**9**, **14**, **16**, **20**, **74**, **75**, **77**, **91**, **97**	**3**, **64**, **69**, **70**, **90**, **95**	1, 3, 4, 6, 9, 12
D	**42**, **49**, **58**, **98**, **103**, **108**	**4**, **9**, **16**, **30**, **34**, **41**, **55**, **69**, **75**, **90**	**3**, **21**, **64**	1, 4, 8, 9, 10, 11, 12

[a] Potency categories as defined by Eqs. A and B under Supplementary Materials.
[b] Potency categories as defined by Eqs. A and B under Supplementary Materials.
[c] Cluster numbers as in Table 2. These represent the clusters from which the corresponding training subsets were selected.

modeling. (Table C in Supplementary Materials shows the statistical criteria of the best representatives). Clearly from this table, representative models shared comparable features and acceptable statistical success criteria.

Emergence of several statistically comparable pharmacophore models suggests the ability of renin ligands to assume multiple pharmacophoric binding modes within the binding pocket. Therefore, it is quite challenging to select any particular pharmacophore hypothesis as a sole representative of the binding process.

3.3. QSAR modeling

Pharmacophoric hypotheses are important tools in drug design and discovery as they provide excellent insights into ligand-macro molecule recognition. However, their predictive value as 3D-QSAR models is limited by steric clashes and bioactivity-enhancing or reducing auxiliary groups [42]. This point combined with the fact that pharmacophore modeling of renin inhibitors furnished numerous binding hypotheses of comparable success criteria (Table C in Supplementary Materials) prompted us to employ classical QSAR analysis to search for the best combination of pharmacophore(s) and 2D descriptors capable of explaining bioactivity variation across the whole list of collected inhibitors (**1–119**, Table A in Supplementary Materials and Fig. 2).

Furthermore, QSAR modeling was implemented in the current case as competition ground to select the best pharmacophore(s) that can explain bioactivity variation across the whole training list. The fact that pharmacophore modeling requires limited number of carefully selected training compounds (from 16 to 45 compounds only [41]) that exhibit bioactivity variations attributable solely to the presence or absence of pharmacophoric features (i.e., not due to steric or electronic factors) makes it impossible to explore the pharmacophore space of large training sets in one shot, partly because CATALYST cannot handle large number of compounds and partly because pharmacophore modeling is generally confused by electronic and steric bioactivity modifying factors commonly encountered in large SAR list. This dilemma prompted us to break the collected compounds into smaller training subsets (as mentioned earlier) compatible with pharmacophore modeling, i.e., of bioactivity variations attributable solely to the presence or absence of pharmacophoric features. Nevertheless, the basic problem in this approach is it generates large number of pharmacophores, and therefore, it becomes quite hard to identify a particular training set capable of representing the whole list of collected compounds. This problem can be very significant in cases of large SAR lists, as in this case. We found that the best way to solve this problem is by exploring the pharmacophoric space of several carefully selected training subsets, i.e., from the whole list of collected compounds, followed by allowing the resulting pharmacophores to compete within the context of GFA-QSAR analysis carried out on the complete list of compounds, such that the best pharmacophore(s) that are capable of explaining bioactivity variations across the whole list of collected compounds is selected. However, since pharmacophore models fail in explaining electronic and steric bioactivity-modulating effects, the GFA-QSAR process should be allowed to select other 2D physicochemical descriptors to complement the selected pharmacophore(s). We implemented this basic principle against a variety of biological targets [29–36,46].

We employed genetic function approximation and multiple linear regression QSAR (GFA-MLR-QSAR) analysis to search for optimal QSAR equation(s) [41].

The fit values obtained by mapping 12 representative hypotheses (see Section 2.1.6 Clustering of the generated pharmacophore

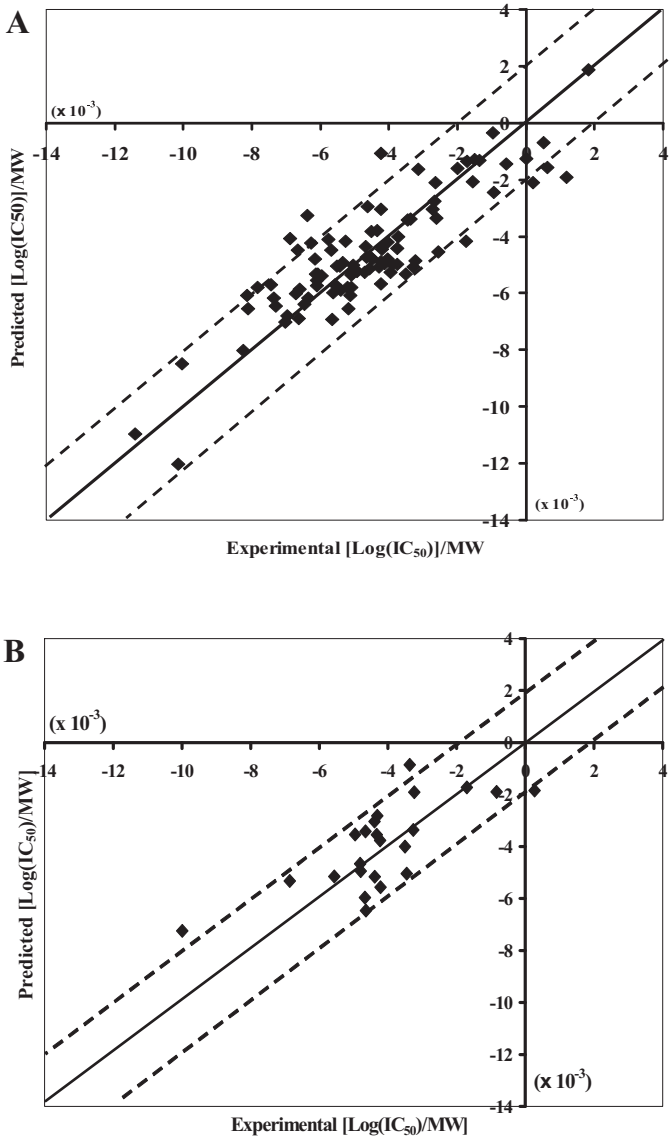


Fig. 3. Experimental versus fitted (A): 96 compounds and predicted (B): 23 compounds bioactivities calculated from the best QSAR model equation (3). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the error margins.

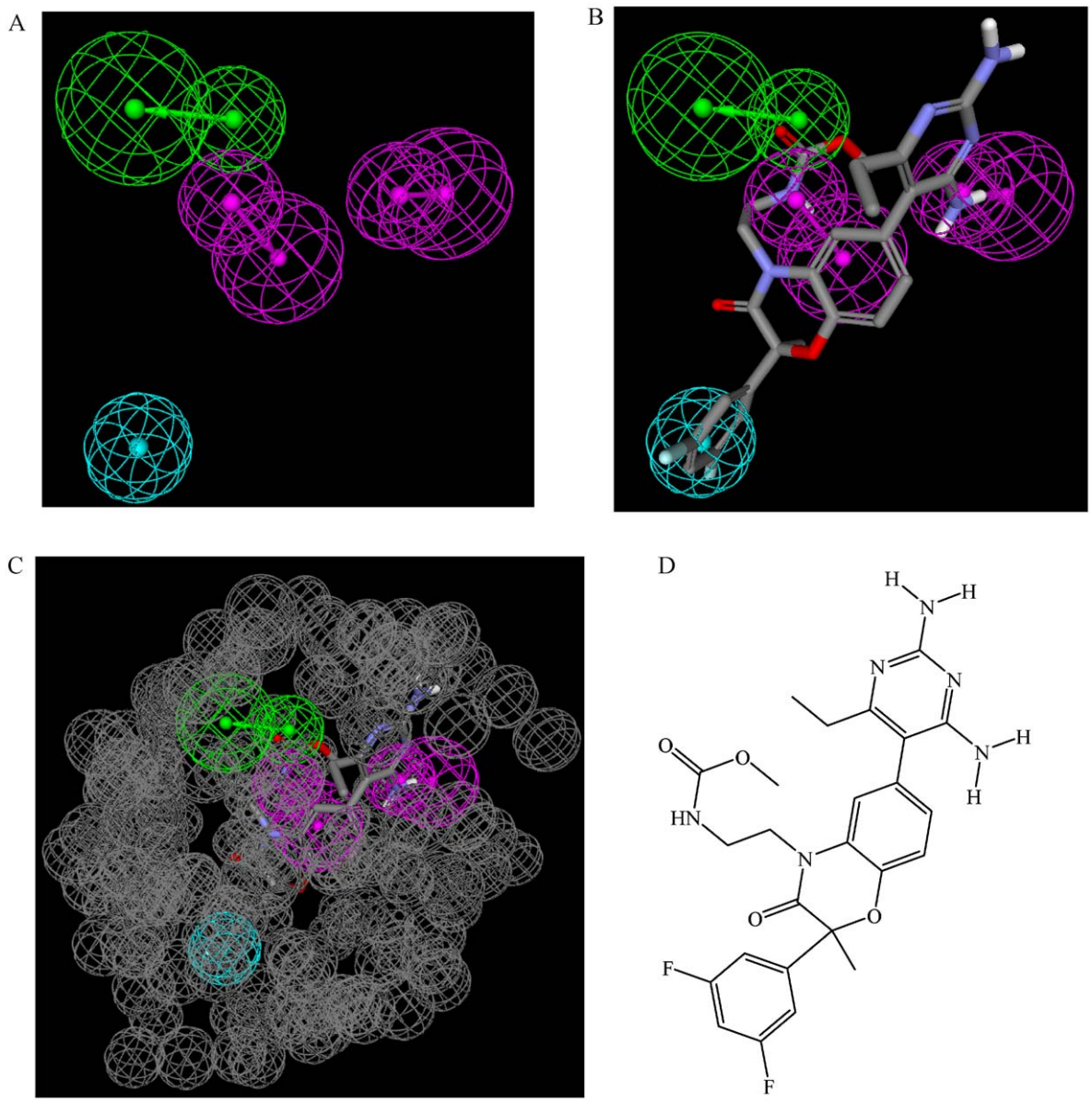


Fig. 4. Hypo1/5. (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres and Hbic as blue spheres; (B) Hypo1/5 fitted against **58** (Table A in Supplementary Materials and Fig. 2, IC_{50} = 2 nM); (C) Hypo1/5 with exclusion volumes (gray spheres); (D) The chemical structures of compound **58**.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 4
Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of **Hypo1/5** and **Hypo1/7**.

Model	Definitions		Chemical features						
			HBA		HBD		HBD		Hbic
Hypo1/5[a]	Weights		2.05054		2.05054		2.05054		2.05054
	Tolerances		1.60	2.20	1.60	2.20	1.60	2.20	1.60
	Coordinates	*X*	1.21	−1.30	2.55	5.44	5.89	8.15	4.51
		Y	−3.39	−1.80	−1.70	−1.02	−5.46	−6.27	4.94
		Z	3.91	3.45	2.56	3.15	1.79	3.58	−1.23
			HBA		HBA		Hbic		Hbic
Hypo1/7[b]	Weights		1.84924		1.84924		1.84924		1.84924
	Tolerances		1.60	2.20	1.60	2.20	1.60		1.60
	Coordinates	*X*	−3.58	−5.53	2.37	4.99	−2.82		0.40
		Y	−1.32	−2.45	−2.62	−3.07	0.32		1.54
		Z	0.07	−1.96	−3.17	−4.64	2.74		1.74

[a] Hypo1/5: the 1st pharmacophore hypothesis generated in the 5th HYPOGEN run (as in Tables B and C in Supplementary Materials).
[b] Hypo1/7: the 1st pharmacophore hypothesis generated in the 7th HYPOGEN run (as in Tables B and C in Supplementary Materials).

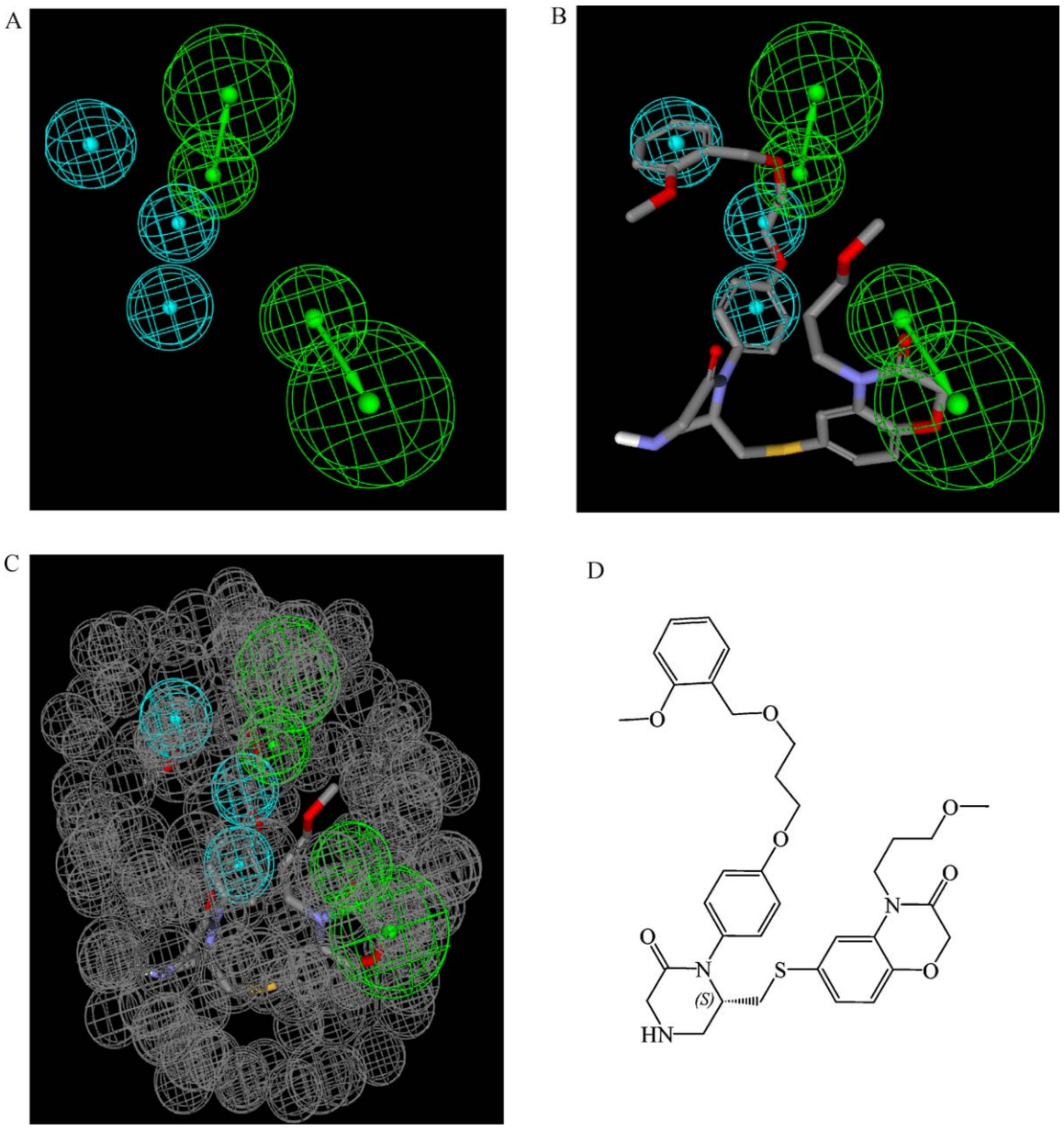


Fig. 5. Hypo1/7. (A) Pharmacophoric features of the binding model: HBA as green vectored spheres and Hbic as blue spheres, (B) Hypo1/7 with exclusion volumes (gray spheres); (C) Hypo1/7 fitted against **109** (Table A in Supplementary Materials and Fig. 1, IC_{50} = 0.18 nM); (D) The chemical structures of compound **109**.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

hypotheses under Experimental) against collected inhibitors (**1–119**, Fig. 2 and Table A under Supplementary Information) were enrolled together with a selection of 2D descriptors as independent variables in GFA-MLR-QSAR analysis.

Unfortunately, all our attempts to achieve self-consistent and predictive QSAR models were futile prompting us to evaluate an alternative modeling strategy, namely, to employ ligand efficiency [log(IC_{50})]/MW [49] as the response variable instead of activ-

Table 5
ROC[a] performances of QSAR-selected pharmacophores and their sterically refined versions as 3D search queries.

Pharmacophore model	ROC[a] – AUC[b]	ACC[c]	SPC[d]	TPR[e]	FNR[f]
Hypo1/5	0.60	0.95	0.97	0.67	0.034
Hypo1/7	0.63	0.95	0.95	0.92	0.046
Sterically refined Hypo1/5	0.76	0.95	0.98	0.50	0.025
Sterically refined Hypo1/7	0.83	0.95	0.96	0.75	0.038

[a] ROC: receiver operating characteristic.
[b] AUC: area under the curve.
[c] ACC: overall accuracy.
[d] SPC: overall specificity.
[e] TPR: overall true positive rate.
[f] FNR: overall false negative rate.

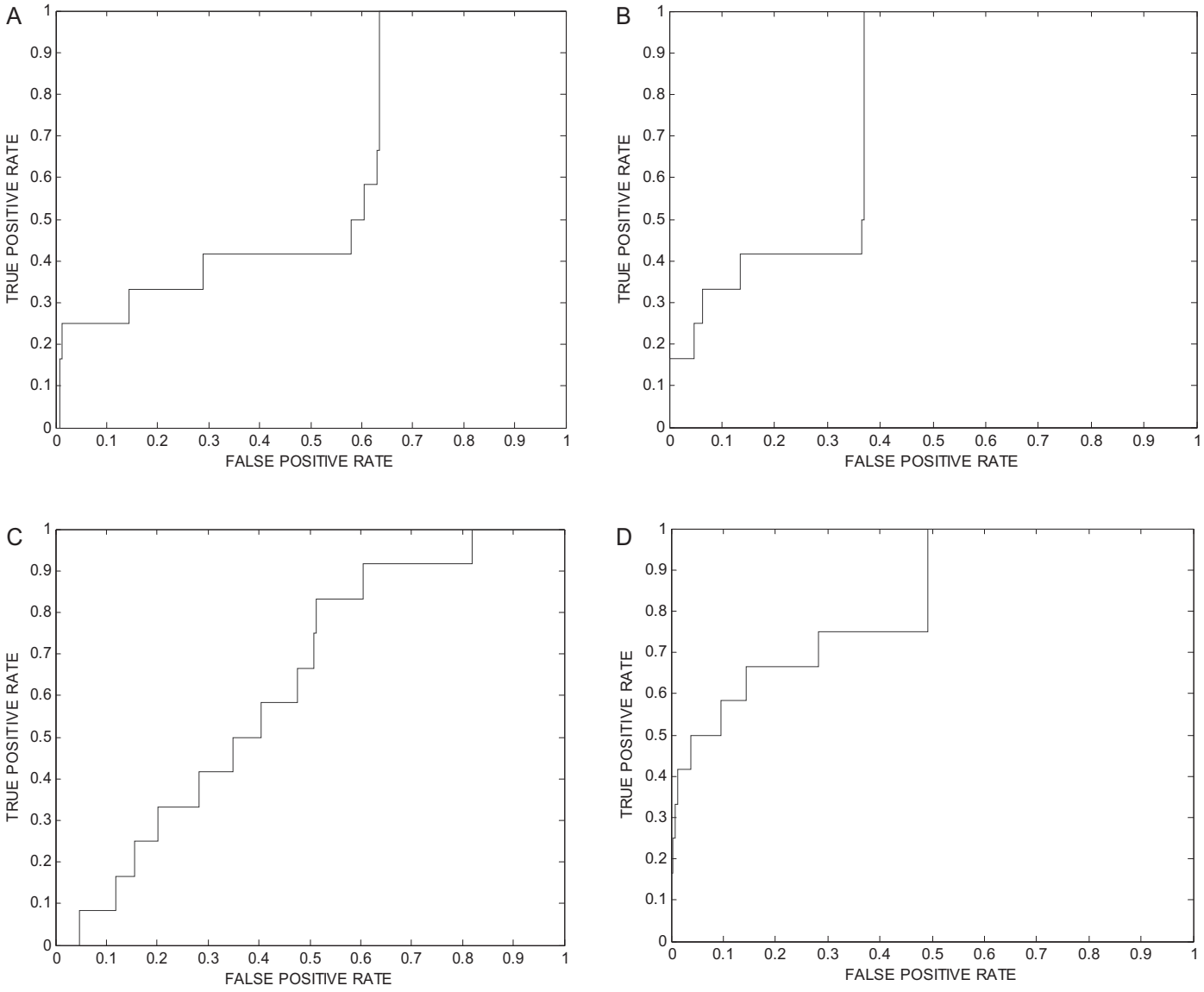


Fig. 6. Receiver operating characteristic (ROC) curves of QSAR-selected pharmacophores: (A) Hypo1/5, (B) sterically refined Hypo1/5, (C) Hypo1/7 and (D) sterically refined Hypo1/7.

ity ($\log(IC_{50})$). We felt that normalizing the bioactivities based on molecular weights can achieve significant QSAR correlations because most reported (in particular potent) renin inhibitors are heavily functionalized and have large molecular weights, which seem to create structure-bioactivity noise that confuse attempts to correlate structural descriptors with bioactivity. The use of ligand efficiency values should neutralize molecular weight-related noise. In fact we previously reported the use of ligand efficiency as response variable in QSAR modeling of β-secreatase inhibitors [35]. Fortunately, this strategy proved successful in this case.

To access the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected 23 molecules (marked with asterisks in Table A in Supplementary Materials) and employed them as external testing set for validating QSAR models. Moreover, all QSAR models were cross-validated automatically using the leave-one-out cross-validation implemented in CERIUS2 [47,59].

Eq. (3) shows the optimal QSAR model. Fig. 3 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors.

$$\begin{aligned}\frac{\log(1/IC_{50})}{MW} &= +3.15\times10^{-4}(\text{Hypo1/5}) + 2.68\times10^{-4}(\text{Hypo1/7}) \\ &\quad -1.39\times10^{-3}(\text{ShadowNU}) - 2.327\times10^{-3}({}^{1}\chi^{v}) \\ &\quad + 1.94\times10^{-3}({}^{0}\chi^{v}) - 4.45\times10^{-4} \\ &\quad (\text{ADME\_Absorption\_Level\_2D}) - 1.52\times10^{-2}n \\ &= 96,\ r^2_{96} = 0.75,\ r^2_{BS} = 0.746,\ r^2_{LOO} = 0.697,\ r^2_{L25\%O} \\ &= 0.717,\ F = 43.52,\ r^2_{PRESS} = 0.527\end{aligned} \tag{3}$$

where n is the number of training compounds, r^2_{96} is the correlation coefficient against 96 training compounds, r^2_{LOO} is the leave-one-out cross-validation correlation coefficient, $r^2_{L25\%O}$ is the leave 25% out cross-validation correlation coefficient (see experimental Section 2.1.7 for more details [47]), r^2_{BS} is the bootstrapping regression coefficient and r^2_{PRESS} is the predictive r^2 determined for 23 randomly selected test compounds (see how they were selected in experimental Section 2.1.7) [47,59]. Hypo1/5 and Hypo1/7 represent the fit values of the training compounds (as calculated from equation D under Supplementary Materials) against the 1st pharmacophore

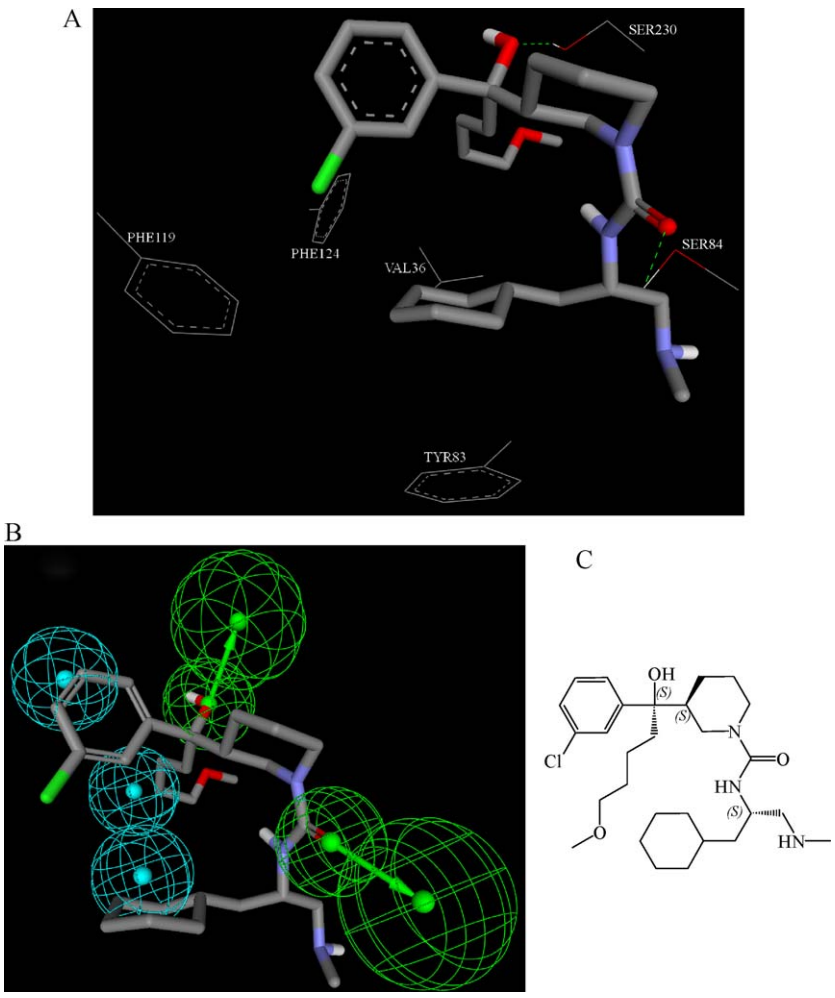


Fig. 7. Co-crystallized ligand of renin X-ray structure 3GW5 (resolution = 2.0 Å): (A) binding pocket of 3GW5 complexed with co-crystallized ligand 72X. (B) Mapping Hypo1/7 against 72X (rigid mapping) and (C) the chemical structure 72X.

models in 5th and 7th automatic runs, as in Tables B and C under Supplementary Information.

ShadowNU is one of the shadow descriptors related to surface area projections of a particular molecule in the *X*, *Y* and *Z* dimensions and depend on the conformation and orientation of the molecule. ShadowNU represents the ratio between the largest to smallest *X*, *Y* and *Z* dimensions of the molecule [47,60]. ${}^{0}\chi^{v}$ and ${}^{1}\chi^{v}$ are the zero and first order chi valence-modified connectivity indices [47,61]. ADME_Absorption_Level_2D is a categorical descriptor related to the distance of a particular molecule from the centre of the chemical space defined by well-absorbed compounds on the scatter plot constructed between Log *P* (Atom-based Log *P*) versus polar surface area (PSA). Higher ADME_Absorption_Level_2D values indicate longer distances from the optimal chemical space centre and infer inferior oral bioavailabilities [29,30,32,47,61,62].

The statistical significance of QSAR equation (3) is supported by r^2_{LOO}, $r^2_{L25\%O}$, and r^2_{PRESS} against 23 randomly selected external list of compounds. Moreover, in order to further assess the significance of the QSAR model we implemented a vigorous validation procedure based on Fisher's randomization test [59]. We selected a 99% confidence level, which scrambles the ligand efficiencies of training compounds 98 unique times (i.e. to generate 99 scrambled lists). Subsequently, GFA-MLR is challenged to use these randomized data to generate QSAR models. Success in generating QSAR equations of better (or comparable) r^2 values to those produced by the original unscrambled data reduces the confidence in the training compounds and their respective QSAR model [47,59]. Table D in Supplementary Materials summarizes the results of the randomization test. The fact that all 98 trials failed to achieve higher r^2 than the original nonrandom QSAR further indicates the statistical significance of the QSAR model.

Emergence of two binding models in the optimal QSAR equation (3) suggests the existence of at least two binding modes assumed by inhibitors within the binding pocket of renin. Figs. 4 and 5 show Hypo1/5 and Hypo1/7 and how they map training compound **58** and **109** (IC_{50} = 2.0 nm and 0.18 nM), respectively, while Table 4 shows the *X*, *Y*, and *Z* coordinates of the two pharmacophores.

Emergence of shadow and connectivity descriptors in equation (3) illustrates certain role played by the ligands' topology in the binding process. However, despite their predictive significance, the information content of such topological descriptors is quite obscure.

On the other hand, emergence of ADME_Absorption_Level_2D in equation (3) associated with negative slope suggests that deviations from oral absorption-related optimal log *P*/PSA chemical space (reflected in higher values for this descriptor) lead to inferior ligand–renin affinities. This trend can be explained based on the fact that most reported potent renin inhibitors (including those in our training list) tend to have large molecular sizes

Table 6
The training compounds used for adding excluded spheres for Hypo1/5 and Hypo1/7 using HIPHOP-REFINE module of CATALYST.

Compd[a]	IC_{50} (nM)	Principal value	MaxOmitFeat[b]	Compd[a]	IC_{50} (nM)	Principal value	MaxOmitFeat[b]
2	27000	0	1	**51**	890	0	1
3	4000	0	1	**52**	1080	0	1
5	198	1	1	**53**	43	1	0
6	120	0	1	**54**	3	2	0
8	178	0	1	**55**	6	2	0
9	54	1	1	**56**	200	0	1
10	66	1	1	**57**	630	0	0
13	23	1	0	**58**	2	2	0
22	3900	0	1	**59**	62	1	0
23	235	0	1	**60**	182	1	0
24	325	0	1	**61**	188	1	0
25	310	0	1	**62**	141	1	1
26	520	0	1	**63**	48	1	1
27	125	0	0	**66**	3900	0	1
28	1040	0	1	**67**	1600	0	1
29	430	0	0	**68**	1200	0	1
30	220	0	0	**69**	900	0	1
31	206	0	0	**70**	600	0	1
32	325	0	0	**71**	450	0	1
33	95	1	1	**74**	370	0	1
34	72	1	1	**75**	340	0	1
35	90	1	1	**76**	330	0	1
38	52	1	0	**77**	300	0	1
39	245	0	0	**79**	270	1	1
40	100	1	0	**83**	230	0	1
41	7	2	0	**89**	1000	0	1
42	1	2	0	**90**	1000	0	1
43	175	0	0	**94**	10000	0	1
44	66	1	1	**95**	1700	0	1
45	60	1	0	**97**	440	0	1
46	40	1	0	**114**	175	1	0
47	7	2	0	**115**	255	0	0
48	7	2	0	**117**	393	0	0
49	0.8	2	0	**118**	2280	0	0

[a] Compounds' numbers are as in Table A (in Suplementary Information) and Fig. 2.
[b] MaxOmitFeat: maximum omitted features.

(of MW > 600) [16,18,7,19,63–65] and therefore require well-balanced combination of hydrophobic and hydrophilic groups. Lack of such balance for such large molecules can lead to excessive hydration at one end, or poor water-solubility at the other, which in both cases seem to undermine their affinities to renin.

3.4. Receiver operating characteristic (ROC) curve analysis

To further validate the resulting models (both QSARs and pharmacophores), we subjected our QSAR-selected pharmacophores to receiver operating curve (ROC) analysis to assess their abilities to correctly classify a list of compounds as actives or inactives. In this case, the validity of a particular pharmacophore is indicated by the area under the curve (AUC) of the corresponding ROC curve, as well as the overall accuracy, specificity, true positive rate and false negative rate of the pharmacophore (see Section 2.1.8 under Experimental). Table 5 and Fig. 6 show the ROC performances of our QSAR-selected pharmacophores.

Clearly from table and figure, both pharmacophores gave mediocre ROC performances. The fact that active and decoy members of the testing set exhibit close physicochemical properties (see Section 2.1.8 Receiver operating characteristic (ROC) curve analysis under Experimental) [52] combined with their large molecular sizes and heavy functionalization, complicates the task of active/inactive classification and explains the observed mediocre performances of both models. Large molecular sizes and heavy functionalization increase the probability of ligand-pharmacophore mapping regardless to bioactivity.

3.5. Addition of exclusion volumes

Although ligand-based pharmacophores serve as excellent tools to probe ligand/macromolecule recognition and as useful 3D QSAR models and search queries, they lack steric constrains necessary to define the size of the binding pocket. This liability renders pharmacophoric models rather promiscuous, as can be concluded from the ROC profiles of Hypo1/5 and Hypo1/7 (Table 5 and Fig. 6). Furthermore, this problem is particularly significant in this case because renin seems to have large binding pocket (as can be concluded from the sizes of crystallographically resolved complexes of potent inhibitors, e.g., 3GW5 and 3D91) while potent renin inhibitors exhibit numerous potentially pharmacophoric functional groups. Both factors tend to reduce the selectivity of renin pharmacophore models as mentioned earlier (see Section 3.4 Receiver operating characteristic (ROC) curve analysis).

Therefore, we decided to complement our QSAR-selected pharmacophores with exclusion spheres employing HipHop-REFINE module of CATALYST to resemble sterically inaccessible regions within the binding site that correspond to each binding mode [41,66]. Understandably, the induced fit motion accompanying each binding mode is reflected on both the arrangement of binding features, i.e., emergence of multiple pharmacophoric binding modes in the QSAR, and on the steric size of the binding pocket accompanying each binding mode. Accordingly, it is logical to attempt characterizing the steric size of the binding pocket during different binding modes (pharmacophoric models).

A structurally diverse training subset was selected for HipHop-REFINE modeling, as in Table 6. The training compounds were

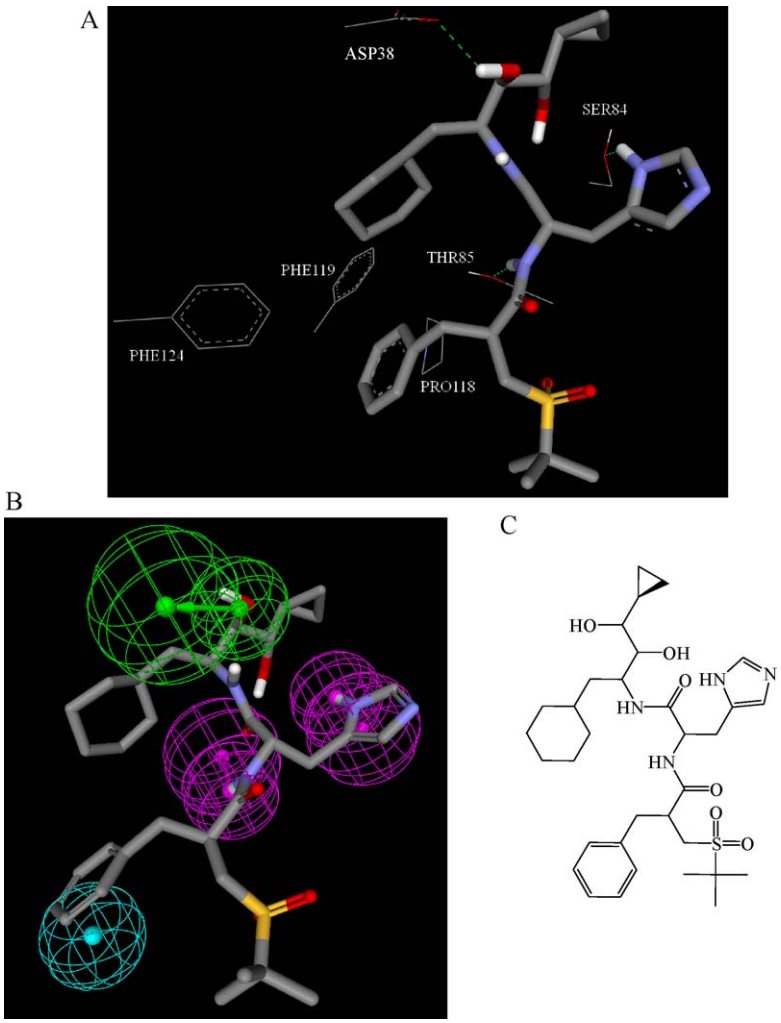


Fig. 8. Co-crystallized ligand of renin X-ray structure 3D91 (resolution = 2.2 Å): (A) binding pocket of 3D91 complexed with corresponding co-crystallized ligand REM, (B) mapping Hypo1/5 against REM (rigid mapping) and (C) the chemical structure of REM.

selected in such a way that the bioactivities of weakly active members are mainly explainable by steric clashes within the binding pocket (see Section 2.1.9 Addition of exclusion volumes under Experimental for more details).

Figs. 4b and 5b show sterically refined versions of Hypo1/5 (150 added exclusion volumes) and Hypo1/7 (147 added exclusion volumes), respectively, while Table 5 and Fig. 6 illustrate their ROC profiles. Clealry form Table 5 and Fig. 6, the sterically refined versions of Hypo1/5 and Hypo1/7 outperformed their unrefined counterparts, which point to improvements in their classification power due to addition of exclusion spheres.

3.6. Comparing pharmacophore models with crystallographic complexes

To further emphasize the validity of our pharmacophore/QSAR modeling approach, we compared the crystallographic structures of two renin/ligand complexes (PDB codes: 3GW5 and 3D91 at resolutions of 2.0 Å and 2.2 Å, respectively) [64,65] with Hypo1/7 and Hypo1/5. Figs. 7 and 8 show the chemical structures of the ligands and compare their renin complexes with the ways they map Hypo1/7 and Hypo1/5 employing rigid mapping, i.e., fitting the ligands' bound states against corresponding pharmacophores without conformational adjustments.

Fitting the hydroxyl fragment of the co-crystallized ligand in 3GW5 complex (ligand PDB code: 72X) [64] against HBA feature in Hypo1/7 (Fig. 7a) corresponds to hydrogen-bonding interactions connecting this fragment with the hydroxy side chain of SER230. Similarly, mapping the ligand's amidic carbonyl against an HBA feature in Hypo1/7 corresponds to hydrogen-bonding interaction connecting this group with the hydroxyl side chain of SER84 (Fig. 7a and b). Finally, mapping the chlorobenzene; cyclohexyl and methoxybutyl moieties of 72X against three Hbic features in Hypo1/7 (Fig. 7b) correlates with fitting these fragments in a hydrophobic pocket composed of the side chains of PHE119, PHE124, VAL36 and TYR83, as in Fig. 7a.

A similar pharmacophore-to-binding pocket analogy is concluded by comparing Hypo1/5 with the co-crystallized ligand in 3D91 complex (ligand PDB code: REM or remikiren) [65]. Mapping one of the hydroxy groups of REM against HBA in Hypo1/5 (Fig. 8b) corresponds to hydrogen-bonding interaction connecting this group with the carboxylic acid of ASP38 (Fig. 8a), while mapping the imidazole and amidic NH groups in REM against two HBD features in Hypo1/5 (Fig. 8b) correspond to hydrogen-bonding interactions connecting these moieties with the hydroxyl side chains of SER84 and THR85, respectively, as in Fig. 8a and b.

Table 7
The captured hit molecules with their fit values, their corresponding QSAR estimates from equation (3) and their *in vitro* bioactivities.

Fit values against[b]			QSAR-based estimates		*In vitro* anti-renin activity[c]	
Tested Hits[a]	Hypo1/5	Hypo1/7	Log(1/IC_{50})	IC_{50} (nM)	% Inhibition at 10.0 μM	IC_{50} (μM)
120	7.6	6.5	−1.9	74.8	95	2.62 (99.8%)[f]
121	6.3	5.6	−1.9	85.9	90	3.16 (100%)[f]
122	5.4	7.1	−0.7	4.9	86	3.0 (99.9%)[f]
123	6.3	8.0	−1.5	34.5	57	7.53 (99.9%)[f]
124	5.9	7.0	−1.2	17.1	47	ND[g]
125	7.1	7.1	−2.0	108.0	0	ND
126	6.7	4.8	−1.7	50.6	17	ND
127	6.7	7.3	−1.3	22.4	74	43.8 (97.5%)[f]
128	6.3	6.2	−1.9	79.9	50	10
129	7.1	7.1	−0.9	8.7	0	ND
130	7.2	4.2	−1.9	80.3	31	ND
131	5.2	6.1	−1.7	47.3	32	ND
132	7.3	5.9	−1.3	20.4	64	61.9 (99.4%)[f]
133	7.3	7.1	−2.1	118.1	40	ND
134	5.6	6.5	−1.8	63.9	31	ND
135	5.4	5.3	−1.9	83.2	15	ND
136	5.3	7.6	−1.5	32.9	29	ND
137	8.0	4.9	−1.6	37.8	19	ND
138	5.8	4.6	−1.8	63.1	13	ND
139	7.4	6.7	−1.6	41.8	5	ND
140	7.4	6.9	−1.9	82.5	0	ND
141	7.6	8.3	−1.5	32.4	0	ND
142	7.8	5.9	−1.9	86.3	0	ND
143	5.6	6.1	−1.9	72.1	0	ND
144	6.8	6.3	−1.0	10.5	0	ND
145	5.8	6.7	−1.8	67.4	0	ND
146	4.9	6.3	−1.7	55.7	0	ND
147	7.5	6.8	−2.1	134.8	0	ND
148	7.5	7.5	−2.0	105.5	0	ND
149	8.0	6.7	−1.5	28.7	0	ND
150	5.1	7.3	−2.0	93.5	0	ND
151	7.0	7.1	−1.3	20.9	0	ND
152	5.2	6.6	−1.8	64.5	0	ND
153	6.1	5.7	−1.9	76.6	0	ND
154	7.1	5.4	−2.0	102.1	0	ND
155	4.7	7.4	−1.7	53.0	0	ND
156	7.3	6.7	−1.7	51.0	0	ND
157	7.4	7.7	−1.6	43.6	0	ND
Standard inhibitor[d]				100[e]		

[a] Compound numbers as in Fig. 7.
[b] Best-fit values against each binding hypothesis calculated by equation (D under Supplementary Materials).
[c] Bioactivity values are the average of at least duplicate measurements.
[d] Standard renin inhibitor Ac-His-Pro-Phe-Val-Sta-Leu-Phe-NH2 provided in the assay kit [57].
[e] Percent inhibition at 1.0 μM [57].
[f] Values between brackets represent the correlation coefficients of the corresponding dose–response curves at three concentrations (0.1 μM, 1 μM, and 10 μM). Data fitting was performed using GraphPad Prism (version 5.04) via fitting against sigmoidal dose-inhibition model.
[g] ND: not determined.

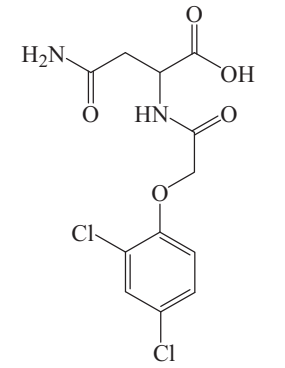
Finally, mapping the phenyl ring of co-crystallized REM against Hbic feature in Hypo1/5 nicely correlates with stacking this group against the side chains of PHE119, PHE124 and PRO118.

Clearly from the above discussion, Hypo1/7 and Hypo1/5 represent two valid binding modes assumed by ligands within renin binding pocket. However, pharmacophore models point to limited number of critical interactions required for high ligand–renin affinity in each of the binding modes. In contrast, crystallographic complexes reveal many bonding interactions without highlighting critical ones. Incidentally, Figs. 7a and 8a only show interactions corresponding to pharmacophoric features while other binding interactions were hidden for clarity.

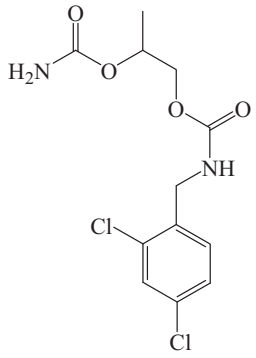
Table 8
Comparison between the molecular properties of training compounds (**1**–**119**, Fig. 2 and Table A under Supplementary Materials) versus active NCI hits (**120**–**123**, **127**, **128** and **132** in Table 7 and Fig. 9).

		Log *P*[a]	MWt	Rotatable bonds	Rings	Aromatic rings	HBA	HBD	Ligand efficiency
Training compounds	Mean	3.45	507.72	10.69	4.16	2.95	6.93	1.8	-4.42×10^{-3}
	St.D.	1.21	78.73	3.81	0.72	0.53	0.99	0.77	2.37×10^{-3}
	Min	−0.64	341.45	5	3	2	5	1	-1.14×10^{-2}
	Max	6.07	645.79	18	6	4	10	4	1.82×10^{-3}
Active hits	Mean	2.59	372.94	5.29	2.57	1.71	5.43	2.71	-9.62×10^{-3}
	St.D.	1.36	55.66	1.58	1.05	1.03	1.18	1.28	1.62×10^{-3}
	Min	0.92	321.16	3	1	1	4	1	-1.22×10^{-2}
	Max	4.31	483.3	7	4	4	7	5	-7.00×10^{-3}

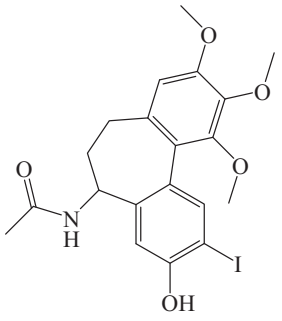
[a] Log *P*: calculated as implemented in Discovery Studio 2.5.



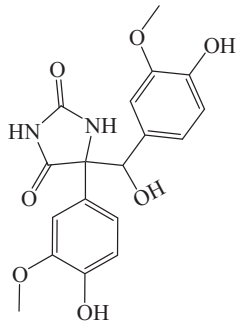
120 (NSC 9200)



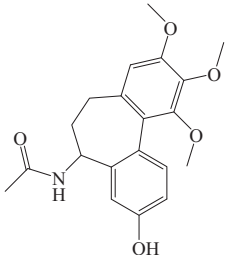
121 (NSC 29145)



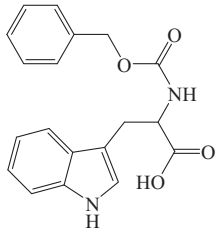
122 (NSC 33408)



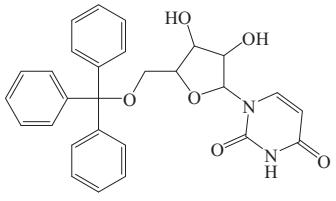
123 (NSC 50120)



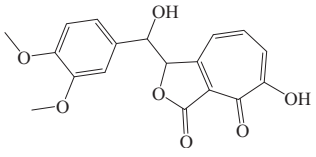
124 (NSC 51045)



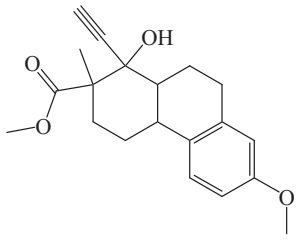
125 (NSC 88478)



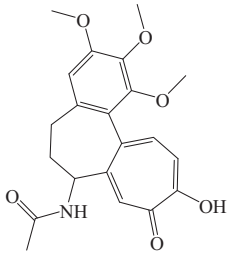
126 (NSC 97038)



127 (NSC 101035)

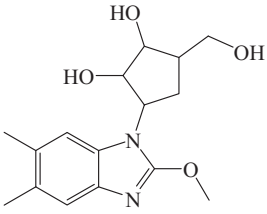


128 (NSC 119035)

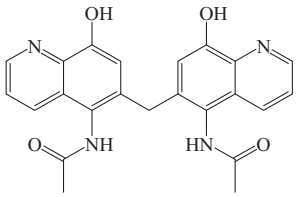


129 (NSC 123396)

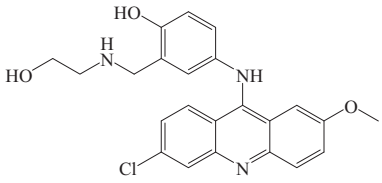
Fig. 9. The chemical structures of the tested highest-ranking hits predicted by the QSAR model (equation (3)) and associated pharmacophores.



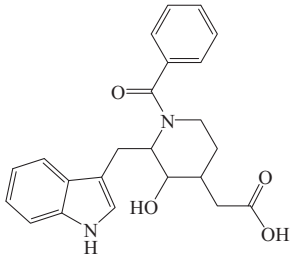
130 (NSC 124157)



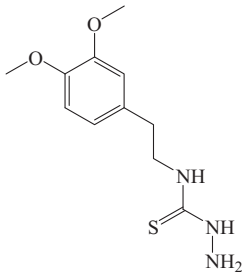
131 (NSC 130811)



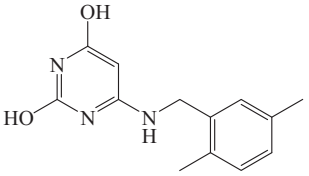
132 (NSC 130815)



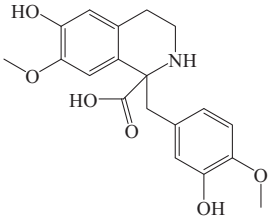
133 (NSC 152548)



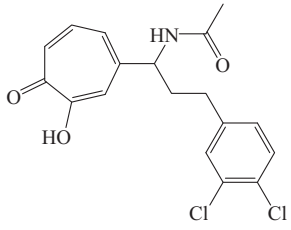
134 (NSC 176360)



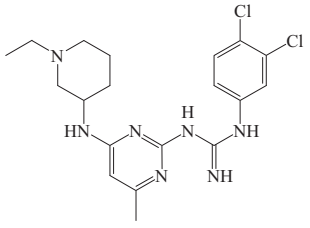
135 (NSC 211350)



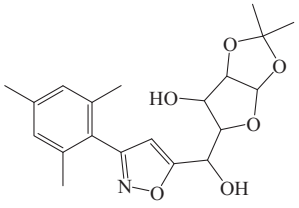
136 (NSC 267402)



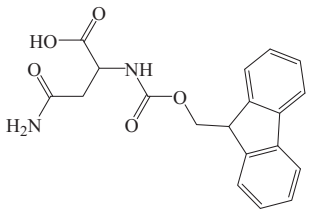
137 (NSC 282885)



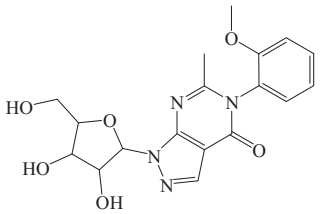
138 (NSC 305761)



139 (NSC 331782)

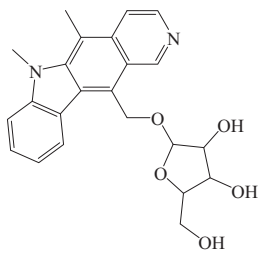


140 (NSC 334297)

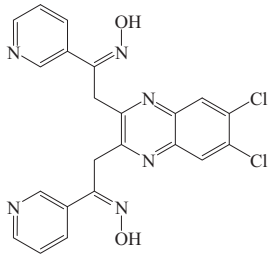


141 (NSC 357750)

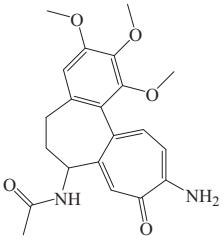
Fig. 9. (*Continued*).



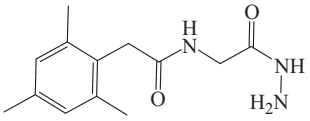
142 (NSC 360319)



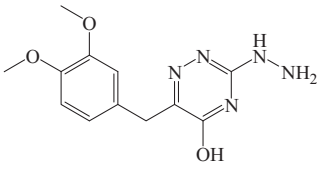
143 (NSC 373267)



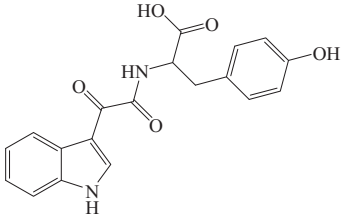
144 (NSC 403144)



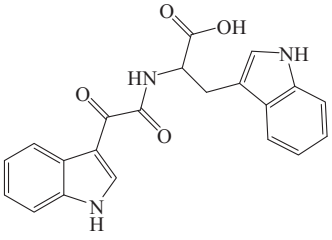
145 (NSC 408435)



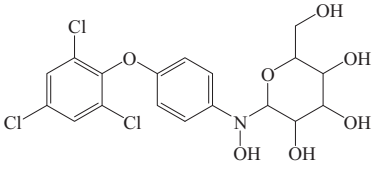
146 (NSC 602112)



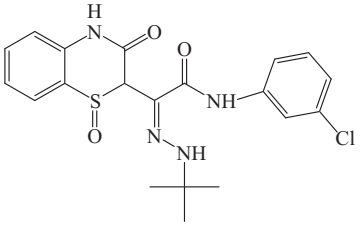
147 (NSC 608048)



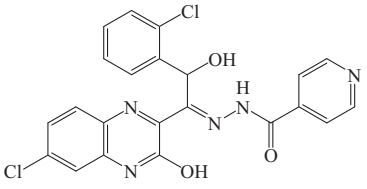
148 (NSC 608049)



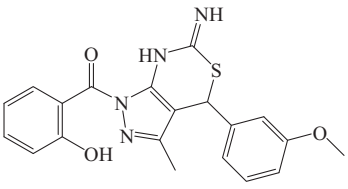
149 (NSC 609597)



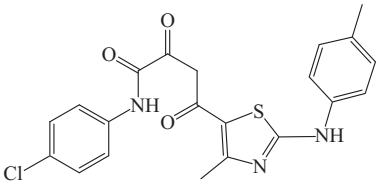
150 (NSC 646811)



151 (NSC 647082)



152 (NSC 657994)



153 (NSC 658278)

Fig. 9. (*Continued*).

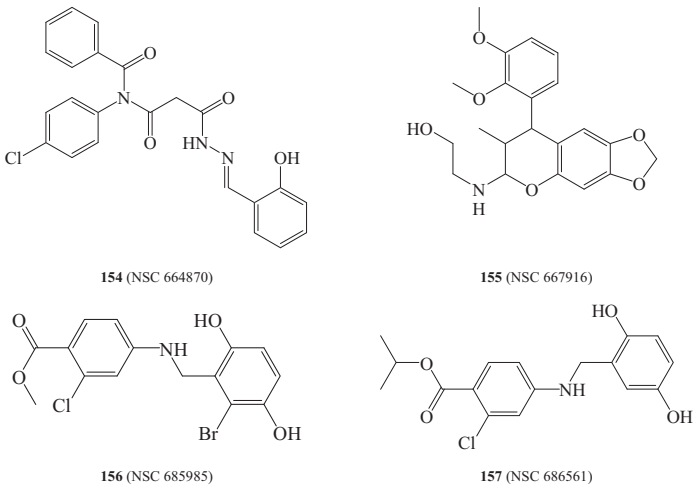


Fig. 9. (*Continued*).

3.7. *In silico screening and subsequent in vitro evaluation*

Due to the fact that steric refinement enhanced the ROC profiles of Hypo1/5 and Hypo1/7, we decided to use their sterically refined versions to screen the national cancer institute list of compounds (NCI, includes 238,819 compounds) [41] for new renin inhibitors. However, we were prompted by the moderate ROC performances (AUC range from 0.76 to 0.83) of the models to screen the NCI against the two pharmacophores sequentially, i.e., hits captured by screening the NCI against Hypo1/7 (33175 hits) were filtered by Hypo1/5. Surviving hits (3668) were subsequently filtered by Lipinski's [67] and Veber's criteria [68] to yield 1282 compounds.

Surviving hits were fitted against Hypo1/7 and Hypo1/5 (fit values determined by Eq. (D) under Supplementary Information) and their fit values were substituted in QSAR Eq. (3) to determine their predicted bioactivities. However, in order to minimize the impact of any possible extrapolatory QSAR prediction errors on decisions regarding which hits merit subsequent *in vitro* testing [69,70], $\text{Log}(1/\text{IC}_{50})$ predictions were merely employed to rank the corresponding hits and prioritize subsequent *in vitro* testing [29–34]. This approach should act as additional filter to enrich the hit list with more active compounds. Table 7 and Fig. 9 show the high-

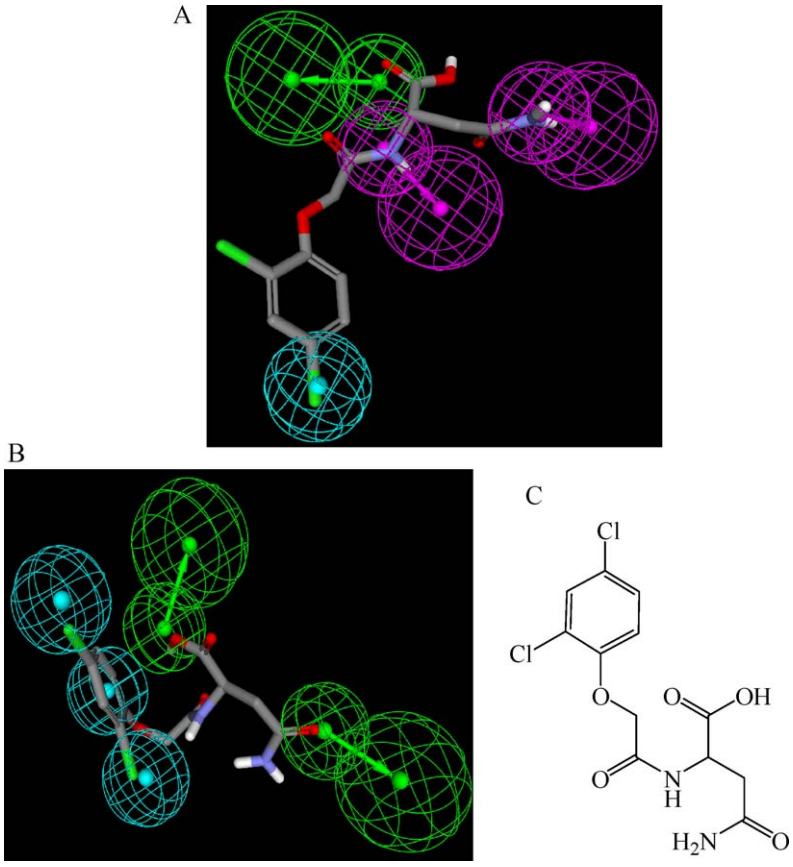


Fig. 10. Hit **120** fitted against (A) Hypo1/5, (B) Hypo1/7 and (C) chemical structure of **120**.

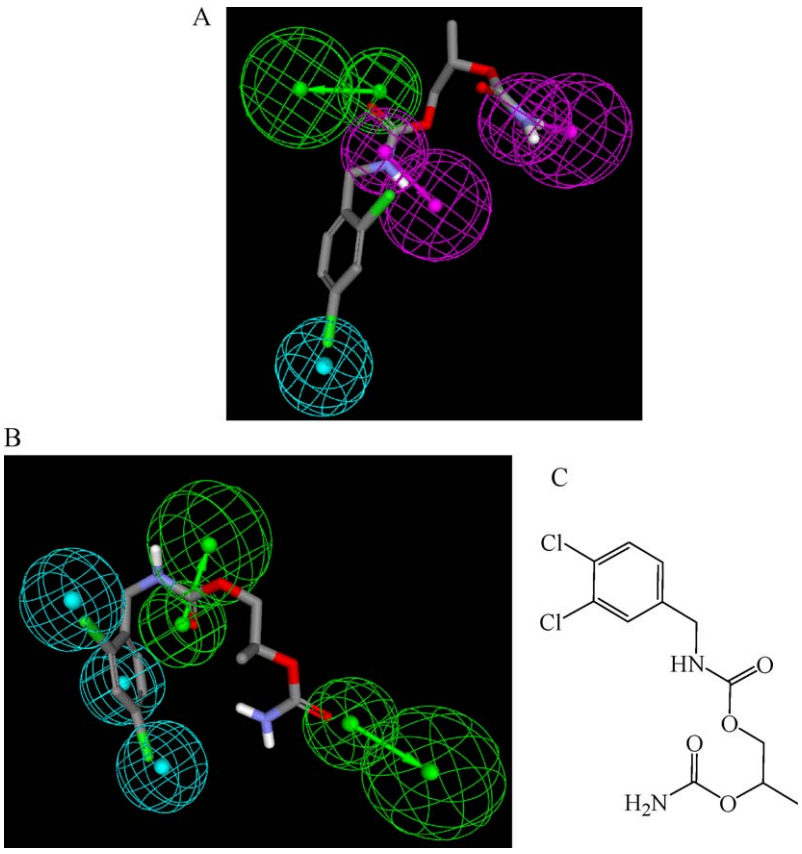


Fig. 11. Hit **121** fitted against: (A) Hypo1/5, (B) Hypo1/7 and (C) chemical structure of **121**.

est predicted hits, their QSAR-based predictions, as well as their experimental *in vitro* bioactivities.

Out of the 38 highest-ranking hits (selected based on availability from the NCI and cost of screening) acquired for experimental validation, 18 were found to possess inhibitory activities against renin ranging from 5.0% to 95% at 10 μM. Seven hits, namely, **120** (NSC 9200), **121** (NSC 29145), **122** (NSC 33408), **123** (NSC 50120), **127** (NSC 101035), **128** (NSC 119035) and **132** (NSC 130815), exhibited ≥50% renin inhibition at 10 μM prompting us to evaluate their IC_{50} values. Fig. 10 shows how Hypo1/5 and Hypo1/7 fit **120** (IC_{50} = 2.62 μM), while Fig. 11 shows Hypo1/5 and Hypo1/7 fitted against **121** (IC_{50} = 3.16 μM).

Table 8 compares the properties of active hits and training compounds. Clearly from the table that although our active *in silico* hits exhibit somewhat inferior ligand efficiencies, i.e., compared to training compounds; they are significantly smaller (i.e., in molecular weight), furthermore, they tend to have lesser number of rotable bonds, hydrogen-bond acceptors and donors, which should promote our active hits as excellent leads for subsequent optimization.

Apparently from the Table 7, QSAR-based predictions deviated from experimental values by at least two logarthmic cycles. We believe this deviation is related to the significant structural differences between training compounds and captured hits, particularly their molecular weight distributions. Training compounds range from 341 to 645*D*, while tested hits range from 245 to 500*D*. To evaluate this hypothesis we compared the means, standard deviations, minimal and maximal values of several physicochemical descriptors between active hits and training compounds. Table 8 shows the results of the assessment. Clearly, captured hits deviate significantly from collected training compounds. This difference is particularly evident from the values of molecular weights, log *P*, and rotable bonds.

However, despite inaccuracies in bioactivity predictions, our overall strategy of combining pharmacophore modeling, QSAR analysis and *in silico* screening succeeded in capturing seven novel scaffolds of low micromolar IC_{50} values against renin.

4. Conclusion

Renin inhibitors are currently considered as potential treatments for hypertension. The pharmacophoric space of renin inhibitors was explored via six diverse sets of inhibitors and using CATALYST-HYPOGEN to identify high-quality binding model(s). Subsequently, genetic algorithm and multiple linear regression analysis were employed to access optimal QSAR model capable of explaining anti-renin bioactivity variation across 119 collected renin inhibitors ($r^2_{96} = 0.746$, F − statistic = **43.552**, $r^2_{LOO} = 0.697$, r^2_{PRESS} against 23 test inhibitors = 0.527). Two pharmacophoric models emerged in the QSAR equation suggesting the existence of two complementary binding modes accessible to ligands within renin binding pocket. The validation of the QSAR equations and the associated pharmacophoric models was experimentally established by the identification of several potent renin inhibitors retrieved from NCI structural database. NCI hits **120** and **121** were most potent with anti-renin IC_{50} values of 1.0 μM.

Acknowledgments

The authors thank the Deanship of Scientific Research and Hamdi-Mango Centre for Scientific Research at the University of Jordan for their generous funds.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.02.001.

References

[1] T. Berl, Renal protection by inhibition of the renin angiotensin aldosterone system, JRAAS 10 (2009) 1–9.
[2] K.B. Lindsay, T. Skrydstrup, Formal total synthesis of the potent renin inhibitor aliskiren: application of a SmI2-promoted acyl-like radical coupling, J. Org. Chem. 71 (2006) 4766–4777.
[3] G.T. Andreas, P.G. Ioannis, N.T. Anastassios, On the structural basis of the hypertensive properties of angiotensin II: a solved mystery or a controversial issue, Curr. Top. Med. Chem. 4 (2004) 431–444.
[4] K.H. Norman, Therapeutic opportunities for renin inhibitors, JRAAS 1 (2005) 107–109.
[5] M.J. Brown, Direct renin inhibition – a new way of targeting the renin system, JRAAS 7 (Suppl. 2) (2006) S7–S11.
[6] S.B. Jeffrey, Angiotensin-converting enzyme inhibition: a landmark advance in treatment for cardiovascular diseases, Eur. Heart J. Suppl. 9 (2007) E2–E9.
[7] O. Bezenc, D. Bur, T. Weller, S. Richard-Bildstein, L. Remen, T. Sifferlen, O. Corminboeuf, C. Grisostomi, C. Boss, L. Prade, S. Delahaye, A. Treiber, P. Strickner, C. Binkert, P. Hess, B. Steiner, W. Fischli, Design and preparation of potent, nonpeptidic, bioavailable renin inhibitors, J. Med. Chem. 52 (2009) 3689–3702.
[8] N.D.L. Fisher, N.K. Hollenberg, Renin inhibition: what are the therapeutic opportunities? J. Am. Soc. Nephrol. 16 (2005) 592–599.
[9] W.B. Wendy, D. Ah jan, Prorenin and the (pro)renin receptor: binding kinetics, signalling and interaction with aliskiren, JRAAS 12 (2008) 181–184.
[10] P.L Beaulieu, J. Gillard, M. Bailey, C. Beaulieu, J. Duceppe, P. Lavallée, D. Wernic, Practical synthesis of BILA 2157 BS, a potent and orally active renin inhibitor: use of an enzyme-catalyzed hydrolysis for the preparation of homochiral succinic acid1 derivatives, J. Org. Chem. 64 (1999) 6622–6634.
[11] E. Vieira, A. Binggeli, V. Breu, D. Bur, W. Fischli, R. Gtiller, G. Hirth, H.P. Marki, M. Mtiller, C. Oefner, M. Scalone, H. Stadler, M. Wilhelm, W. Wostl, Substituted piperidines, highly potent renin inhibitors due to induced fit adaptation of the active site, Bioorg. Med. Chem. Lett. 9 (1999) 1397–1402.
[12] R. Gtiller, A. Binggeli, V. Breu, D. Bur, W. Fischli, G. Hirth, C. Jenny, M. Kansy, F. Montavon, M. Mtiller, C. Oefner, H. Stadler, E. Vieira, M. Wilhelm, W. Wostl, H.P. Marki, Piperidine-renin inhibitors compounds with improved physicochemical properties, Bioorg. Med. Chem. Lett. 9 (1999) 1403–1408.
[13] J. Maibaum, S. Stutz, R. Go¨schke, P. Rigollier, Y. Yamaguchi, F. Cumin, J. Rahuel, H. Baum, N. Cohen, C.R. Schnell, W. Fuhrer, M.G. Gruetter, W. Schilling, J.M. Wood, Structural modification of the P2′ position of 2, 7-dialkyl-substituted 5(S)-Amino-4(S)-hydroxy-8-phenyl-octanecarboxamides: the discovery of aliskiren, a potent nonpeptide human renin inhibitor active after once daily dosing in marmosets, J. Med. Chem. 50 (2007) 4832–4844.
[14] J. Rahuel, V. Rasetti, J. Maibaum, H. Rüeger, R. Göschke, N.-C. Cohen, S. Stutz, F. Cumin, W. Fuhrer, J.M. Wood, M.G. Grütter, Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin, Chem. Biol. 7 (2000) 493–504.
[15] A.R. Sialecki, K. Hayakawa, M. Fujinaga, M. Murphy, M. Fraser, A. Muir, C. Carilli, J.A. Lewicki, J.D. Baxter, M.N.G. James, Science 243 (1989) 1346–1351.
[16] L. Remen, O. Bezencon, S. Richard-Bildstein, D. Bur, L. Prade, O. Corminboeuf, C. Boss, C. Grisostomi, T. Sifferlen, P. Strickner, P. Hess, S. Delahaye, A. Treiber, T. Weller, C. Binkert, B. Steiner, W. Fischli, New classes of potent and bioavailable human renin inhibitors, Bioorg. Med. Chem. Lett. 19 (2009) 6762–6765.
[17] P. Källblad, P.M. Dean, Efficient conformational sampling of local side-chain flexibility, J. Mol. Biol. 326 (2003) 1651–1665.
[18] Z. Xu, S. Cacatian, J. Yuan, R.D. Simpson, L. Jia, W. Zhao, C.M. Tice, P.T. Flaherty, J. Guo, A. Ishchenko, S.B. Singh, Z. Wu, B.M. McKeever, B.B. Scott, Y. Bukhtiyarov, J. Berbaum, J. Mason, R. Panemangalore, M.G. Cappiello, R. Bentley, C.P. Doe, R.K. Harrison, G.M. McGeehan, L.W. Dillard, J.J. Baldwin, D.A. Claremon, Optimization of orally bioavailable alkyl amine renin inhibitors, Bioorg. Med. Chem. Lett. 2 (2010) 694–699.
[19] R.W. Sarver, J. Peevers, W.L. Cody, F.L. Ciske, J. Dyer, S.D. Emerson, J.C. Hagadorn, D.D. Holsworth, M. Jalaie, M. Kaufman, M. Mastronardi, P. McConnell, N.A. Powell, J. Quin, C.A. Van Huis, E. Zhang, I. Mochalkin, Binding thermodynamics of substituted diaminopyrimidine renin inhibitors, Anal. Biochem. 360 (2007) 30–40.
[20] N.A. Powell, F.L. Ciske, C. Cai, D.D. Holsworth, K. Mennen, C.A. Van Huis, M. Jalaie, J. Day, M. Mastronardi, P. McConnell, I. Mochalkin, E. Zhang, M.J. Ryan, J. Bryant, W. Collard, S. Ferreira, C. Gu, R. Collins, J.J. Edmunds, Rational design of 6-(24 diaminopyrimidinyl)-1,4-benzoxazin-3-ones as small molecule renin inhibitors, Bioorg. Med. Chem. 15 (2007) 5912–5949.
[21] D.D. Holsworth, C. Cai, X. Cheng, W.L. Cody, D.M. Downing, N. Erasga, C. Lee, N.A. Powell, J.J. Edmunds, M. Stier, M. Jalaie, E. Zhang, P. McConnell, M.J. Ryan, J. Bryant, T. Li, A. Kasani, E. Hall, R. Subedi, M. Rahimc, S. Maiti, Ketopiperazine-based renin inhibitors: optimization of the "C" ring, Bioorg. Med. Chem. Lett. 16 (2006) 2500–2504.
[22] N.A. Powell, E.H. Clay, D.D. Holsworth, J.W. Bryant, M.J. Ryan, M. Jalaie, E. Zhang, J.J. Edmunds, Equipotent activity in both enantiomers of a series of ketopiperazine-based renin inhibitors, Bioorg. Med. Chem. Lett. 15 (2005) 2371–2374.
[23] A.M. Davis, S.A. St-Gallay1, G.J. Kleywegt, Limitations and lessons in the use of X-ray structural information in drug design, Drug Discov. Today 20 (2008) 831–841.
[24] G. Klebe, Virtual ligand screening: strategies perspectives and limitations, Drug Discov. Today 11 (2006) 580–594.
[25] H. Steuber, M. Zentgraf, C. Gerlach, C.A. Sotriffer, A. Heine, G.J. Klebe, Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions, Mol. Biol. 363 (2006) 174–187.
[26] M.T. Stubbs, S. Reyda, F. Dullweber, M. Moller, G. Klebe, D. Dorsch, W. Mederski, H. Wurziger, pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design, Chem. Biol. Chem. 3 (2002) 246–249.
[27] M.A. DePristo, P.I.W. de Bakker, T.L. Blundell, Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography, Structure 12 (2004) 831–838.
[28] Proceedings of the 9th European CATALYST User Group Meeting Advanced Seminars in CATALYST Frankfurt, Germany, Accelrys Inc., San Diego, CA, 2006.
[29] M.O. Taha, A. Qandil, D. Zaki, M. AlDamen, Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling, Eur. J. Med. Chem. 40 (2005) 701–727.
[30] M.O. Taha, Y. Bustanji, M.A.S. Al-Ghussein, M. Mohammad, H. Zalloum, I.M. Al-Masri, N. Atallah, Pharmacophore modeling, quantitative structure-activity relationship analysis and in silico screening reveal potent glycogen synthase kinase-3β inhibitory activities for cimetidine, hydroxychloroquine and gemifloxacin, J. Med. Chem. 51 (2008) 2062–2077.
[31] M.O. Taha, L.A. Dahabiyeh, Y. Bustanji, H. Zalloum, S. Saleh, Combining ligand-based pharmacophore modeling, quantitative structure-activity relationship analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors, J. Med. Chem. 51 (2008) 6478–6494.
[32] M.O. Taha, N. Atallah, A.G. Al-Bakri, C. Paradis-Bleau, H. Zalloum, K. Younis, R.C. Levesque, Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening, Bioorg. Med. Chem. 16 (2008) 1218–1235.
[33] M.O. Taha, Y. Bustanji, A.G. Al-Bakri, M. Yousef, W.A. Zalloum, I.M. Al-Masri, N. Atallah, Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening, J. Mol. Graph. Modell. 25 (2007) 870–884.
[34] A.M. Abu Hammad, M.O. Taha, Pharmacophore modeling, quantitative structure-activity relationship analysis, and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors, J. Chem. Inf. Model. 49 (2009) 978–996.
[35] A. Al-Nadaf, G. Abu Sheikha, M.O. Taha, Elaborate ligand-based pharmacophore exploration and QSAR analysis guide the synthesis of novel pyridinium-based potent beta-secretase inhibitory leads, Bioorg. Med. Chem. 9 (2010) 3088–3115.
[36] R. Abu Khalaf, G. Abu Sheikha, Y. Bustanji, M.O. Taha, Discovery of new cholesteryl ester transfer protein inhibitors via ligand-based pharmacophore modeling and QSAR analysis followed by synthetic exploration, Eur. J. Med. Chem. 4 (2010) 1598–1617.
[37] D.D. Holsworth, M. Jalaie, T. Belliotti, C. Cai, W. Collard, S. Ferreira, N.A. Powell, M. Stier, E. Zhang, P. McConnell, I. Mochalkin, M.J. Ryan, J. Bryant, T. Li, A. Kasani, R. Subedi, S.N. Maitib, J.J. Edmunds, Discovery of 6-ethyl-2, 4 diaminopyrimidine-based small molecule renin inhibitors, Bioorg. Med. Chem. Lett. 17 (2007) 3575–3580.
[38] N.A. Powell, E.H. Clay, D.D. Holsworth, J.W. Bryant, M.J. Ryan, M. Jalaie, J.J. Edmunds, Benzyl ether structure–activity relationships in a series of ketopiperazine-based renin inhibitors, Bioorg. Med. Chem. Lett. 15 (2005) 4713–4716.
[39] D.D. Holsworth, N.A. Powell, D.M. Downing, C. Cai, W.L. Cody, J. Michael Ryan, R. Ostroski, M. Jalaie, J.W. Bryantb, J.J. Edmunds, Discovery of novel non-peptidic ketopiperazine-based renin inhibitors, Bioorg. Med. Chem. 13 (2005) 2657–2664.
[40] W.L. Cody, D.D. Holsworth, N.A. Powell, M. Jalaie, E. Zhang, W. Wang, B. Samas, J. Bryant, R. Ostroski, M.J. Ryand, J.J. Edmunds, The discovery and preparation of disubstituted novel amino-aryl-piperidine-based renin inhibitors, Bioorg. Med. Chem. 13 (2005) 59–68.
[41] Catalyst 4.11 User Guide, Accelrys Software Inc., San Diego, CA, 2005.
[42] I.B. Bersuker, S. Bahçeci, J.E. Boggs, Pharmacophore Perception, Development and Use in Drug Design, International University Line, California, 2000, pp. 457–473.
[43] Y. Kurogi, O.F. Guner, Pharmacophore modeling and three dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.
[44] J. Sutter, O.F. Güner, R. Hoffmann, H. Li, M. Waldman, Pharmacophore Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 501–511.
[45] H. Sutter, J. Li, R. Hoffmann, O.F. Güner, in: O.F. Güner (Ed.), Pharmacophore Perception, Development, and Use in Drug Design, International University Line, California, 2000, pp. 173–189.
[46] M. Taha, A. Al-Bakri, W. Zalloum, Discovery of potent inhibitors of pseudomonal quorum sensing via pharmacophore modeling and In silico screening, Bioorg. Med. Chem. Lett. 16 (2006) 5902–5906.
[47] CERIUS2, QSAR Users' Manual, Version 4.10, Accelrys Inc.: San Diego, CA, 2005, pp. 43–88, 221–235.

[48] K. Poptodorov, T. Luu, T. Langer, R. Hoffmann, Methods and Principles in Medicinal Chemistry, Pharmacophores and Pharmacophores Searches, 2, WILEY–VCH, Weinheim, 2006, pp. 17–47.
[49] E. Perola, An analysis of the binding efficiencies of drugs and their leads in successful drug discovery programs, J. Med. Chem. 53 (2010) 2986–2997.
[50] L.H. Andrew, R.G. Colin, A. Alexander, Ligand efficiency: a useful metric for lead selection, Drug Discov. Today 9 (2004) 430–431.
[51] J. Kirchmair, P. Markt, S. Distinto, G. Wolber, T. Langer, Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—what can we learn from earlier mistakes? J. Comput. Aid. Mol. Des. 22 (2008) 213–228.
[52] M.L. Verdonk, L. Marcel, V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor, P. Watson, Virtual screening using protein-ligand docking: avoiding artificial enrichment, J. Chem. Inf. Comp. Sci. 44 (2004) 793–806.
[53] J.J. Irwin, B.K. Shoichet, ZINC – a free database of commercially available compounds for virtual screening, J. Chem. Inf. Comp. Sci. 45 (2005) 177–182.
[54] N. Triballeau, F. Acher, I. Brabet, J.-P. Pin, H.-O. Bertrand, Virtual screening workflow development guided by the "Receiver Operating Characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4, J. Med. Chem. 48 (2005) 2534–2547.
[55] H. Gao, C. Williams, P. Labute, J. Bajorath, Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands, J. Chem. Inf. Comp. Sci. 39 (1999) 164–168.
[56] M. Jacobsson, P. Liden, E. Stjernschantz, H. Bostroem, U. Norinder, Improving structure-based virtual screening by multivariate analysis of scoring data, J. Med. Chem. 46 (2003) 5781–5789.
[57] SensoLyte® 520 Renin Assay Kit Fluorimetric.
[58] R.P. Sheridan, S.K. Kearsley, Why do we need so many chemical similarity search methods? Drug Discov. Today 7 (2002) 903–911.
[59] R. Fischer, The Principle of Experimentation Illustrated by a Psycho-Physical ExpeHafner Publishing Co., 8th ed., Hafner Publishing, 1966.
[60] W.M. Suhre, S. Ekins, C. Chang, P.W. Swaan, S.H. Wright, Molecular determinants of substrate/inhibitor binding to the human and rabbit renal organic cation transporters hOCT2 and rbOCT2, Mol. Pharmacol. 67 (2005) 1067–1077.
[61] L.B. Kier, L.H. Hall, in: J. Devillers, A.T. Balaban (Eds.), Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach, London, 1999, pp. 491–562.
[62] A. Puratchikodya, M. Dobleb, Antinociceptive and antiinflammatory activities and QSAR studies on 2-substituted-4,5-diphenyl-1H-imidazoles, Bioorg. Med. Chem. 15 (2007) 1083–1090.
[63] J. Rahuel, V. Rasetti, J. Maibaum, H. Rueger, R. Goschke, N.C. Cohen, S. Stutz, F. Cumin, W. Fuhrer, J.M. Wood, M.G. Grutter, Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin, Chem. Biol. 7 (2000) 493–504.
[64] C.M. Tice, Z. Xu, J. Yuan, R.D. Simpson, S.T. Cacatian, P.T. Flaherty, W. Zhao, J. Guo, A. Ishchenko, S.B. Singh, Z. Wu, B.B. Scott, Y. Bukhtiyarov, J. Berbaum, J. Mason, R. Panemangalore, M.G. Cappiello, D. Muller, R.K. Harrison, G.M. McGeehan, L.W. Dillard, J.J. Baldwin, D.A. Claremon, Design and optimization of renin inhibitors: orally bioavailable alkyl amines, Bioorg. Med. Chem. Lett. 19 (2009) 3541–3545.
[65] L. Prade, O. Bezencon, D. Bur, T. Weller, W. Fischli, L. Remen, Human renin in complex with remikiren, in press, (www.PDB.org), doi:10.2210/pdb3d91/pdb.
[66] O.O. Clement, A.T. Mehl, O.F. Guner, Pharmacophore Perception, Development, and Use in Drug Design, in: O.F. Guner (Ed.), IUL Biotechnology Series, International University Line, La Jolla, CA, 2000, pp. 71–84.
[67] C. Lipinski, F. Lombardo, B. Dominy, P. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, J. Adv. Drug Deliv. Rev. 46 (2001) 3–26.
[68] D. Veber, S. Johnson, H. Cheng, B. Smith, K. Ward, K. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, J. Med. Chem. 45 (2002) 2615–2623.
[69] M.T.D. Cronin, T.W. Schultz, Pitfalls in QSAR, J. Mol. Struct. (Theochem.) 622 (2003) 39–51.
[70] M.F. Paola, N.M. Beatriz, O.N. Beatriz, C.E. Alberto, Methodologies QSAR/QSPR/QSTR: current state and perspectives, Int. J. Chem. Model. 1 (2009) 221–243.
[71] G.W. Bemis, M.A. Murcko, The properties of known drugs. 1. Molecular frameworks, J. Med. Chem. 39 (1996) 2887–2893.
[72] Discovery Studio Users' Manual, Version 2.5.5, Accelrys Inc., San Diego, CA, 2010.
[73] A. Politi, S. Durdagi, P. Moutevelis-Minakakis, G. Kokotos, T. Mavromoustakos, Development of accurate binding affinity predictions of novel renin inhibitors through molecular docking studies, J. Mol. Graph. Modell. 29 (2010) 425–435.