

TRPV1 antagonism by piperazinyl-aryl compounds: A Topomer-CoMFA study and its use in virtual screening for identification of novel antagonists



Rajendra Kristam ^{a,b,*}, Shashidhar N. Rao ^{c,1}, Anne Sudha D'Cruz ^a, Vijayalakshmi Mahadevan ^b, Vellarkad N. Viswanadhan ^{a,2}

^a Department of Computational Chemistry, Jubilant Biosys Ltd. #96, Industrial Suburb, 2nd Stage, Yeshwanthpur, Bangalore 560 064, India

^b School of Chemical & Biotechnology, Shamugha Arts, Science, Technology, and Research Academy (SASTRA University), Thanjavur, Tamil Nadu 613 402, India

^c Tripos International, A Certara Company, St. Louis, MO 63101, USA

ARTICLE INFO

Article history:

Received 21 August 2016

Received in revised form 8 December 2016

Accepted 6 January 2017

Available online 7 January 2017

Keywords:

Topomer-CoMFA

3D-QSAR

TRPV1 antagonists

Piperazinyl-aryl

Pain

ABSTRACT

Transient Receptor Potential Vanilloid, member 1 (TRPV1), is a non-selective cation channel belonging to the transient receptor potential (TRP) family of ion channels. It occurs in the peripheral and central nervous system, activated by a variety of exogenous and endogenous stimuli, thus playing a key role in transmission of pain. This has been a target for chronic pain since more than a decade and a number of antagonists that progressed into clinical trials have failed due to the unexpected side effect of core body temperature rise, thus halting progress in this field. Of late, there has been an upsurge in research on this target, with the rat TRPV1 structure being determined, many new antagonists discovered that are temperature-neutral and many new therapeutic avenues being discovered for TRPV1, including diseases of respiratory and digestive systems, skin and bladder. Towards identifying diverse compounds to decipher the role of this target in various indications, here we report a 3D-QSAR model built using the new topomer-CoMFA methodology on a series of piperazinyl-aryl TRPV1 antagonists and the use of this model, along with a pharmacophore model and the shape of one of the potent compounds of this series, to virtually screen a subset of the ZINC database to find novel and diverse hits. These can serve as starting points to develop modality-selective antagonists for chronic pain and to elucidate the critical role of TRPV1 in the various new therapeutic areas.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

TRPV1 (transient receptor potential cation channel, subfamily V, member 1) is a non-selective cation channel activated by a variety of exogenous and endogenous stimuli. Cloning of TRPV1 and demonstration of its therapeutic value have led to intense research in understanding the molecular mechanisms encompassing the responses of sensory neurons to stimuli such as heat, protons, some endogenous activators such as anandamide and exogenous activa-

tors such as capsaicin [1–3]. TRPV1, also known as VR1 (Vanilloid Receptor 1), is primarily expressed on unmyelinated pain-sensing nerve fibers (C-fibers) and small A δ fibers in the dorsal root and trigeminal ganglia [4]. Activation of the channel leads to an influx of calcium and sodium ions into the cell, causing depolarization, and that in turn results in the excitation of primary sensory neurons and ultimately in the perception of pain, though a reduction of the activation thresholds of the channel to other stimuli indicate that agonists of the channel can cause desensitization with therapeutic application in the management of pain. However, such desensitization leads to the side effects of burning sensation, irritation, and neurotoxicity, resulting from continuous influx of calcium ions into the cells, thus limiting the clinical use of agonists [5]. Blocking the TRPV1-mediated pain signaling pathways with receptor antagonists is an alternative promising strategy for the development of novel analgesic drugs with potentially fewer side effects [6,7]. The chemistry and pharmacology of several classes of competitive

* Corresponding author at: Jubilant Biosys Ltd. #96, Industrial Suburb, 2nd Stage, Yeshwanthpur, Bangalore, Karnataka 560022, India.

E-mail address: Rajendra.kristam@jubilantbiosys.com (R. Kristam).

¹ Current address: Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA.

² Current address: Sahyadri Study Center, Sahyadri School, PO Tiwai Hill, Pune District, Maharashtra 410513, India.

TRPV1 antagonists, based on chemotypes containing thiourea, urea, and amide groups, have been described and reviewed [8–12]. They have been classified into four profiles based on their ability to differentially modulate TRPV1 activation by different modes such as capsaicin, pH 5 and heat [13,14]. Over the years, many antagonists that have entered and progressed into various clinical phases have failed due to the development of hyperthermia as an undesired on-target side-effect (Table S1, supplementary information).

With increased structural, mechanistic and biophysical knowledge on TRPV1, there is an unrecognized need to find more and diverse chemotypes for TRPV1 antagonists. New chemotypes are needed to evolve functionally-selective or modality-selective antagonists that can probe and prove if hyperthermia can be dissociated from the other functions of this ion channel [15–23]. The need for new chemotypes is encouraged by the fact that there numerous new indications attributed to TRPV1 for which small molecule chemical probes are required for validation [12,24–27]. Apart from validation for newer indications, probes are required to decipher the potential consequences of targeting a single indication. In this regard, a generally applicable TRPV1 antagonist pharmacophore that can enable discovery of novel chemotypes is critical. For this reason, we had recently developed a pharmacophore model for a series of piperazinyl-aryl compounds and then used the pharmacophore model to develop a 3D-QSAR model using the Schrodinger/Phase methodology [28–30]. This 3D-QSAR model was able to clearly explain and confirm the known SAR as well as outline some more insights that were not apparent in the original publications associated with the piperazinyl-aryl series. One more salient aspect of this work was that the pharmacophore model, representative of the piperazinyl-aryl series, was used to align other known antagonists, including several clinical candidates and compounds in pre-clinical and discovery stages, to clearly show the common features across the diverse chemotypes they represent. In this current work, a 3D-QSAR model of the same piperazinyl-aryl series of TRPV1 antagonists [29,30] is built using the Topomer-CoMFA methodology [31,32] and then used it, along with the pharmacophore model developed in our earlier work, to screen a subset of the ZINC database [33,34] to find new core fragments that can potentially give rise to new chemotypes.

There have been some efforts to use ligand-based methodologies to build pharmacophore and/or QSAR models for TRPV1 antagonists. Some of them used molecular field-based or similar methodologies such as CoMFA & CoMSIA and Phase QSAR [35–38], some used descriptor-based algorithms to find the descriptors most correlating with antagonistic activity [39,40]. In two studies analyses of 3D alignments of the antagonists were performed using compounds in the context of homology models [35,41]. In a recent effort, Goldmann *et al.*, describe pharmacophore modeling utilizing the publicly-available data on TRPV1 antagonists [42]. This work involved extensive validation of the pharmacophore models and use of those models for virtually screening the LifeChem database of 305841 compounds to yield 12 hits. The hits showed promising activity and diversity compared to the reference antagonist and other active compounds. The authors further speculate that this kind of approach of utilizing pharmacophore models to mine public data can help in revealing pharmacophoric ensembles that can distinguish compounds that are safe from those that have undesirable profiles. Feng et al., reported building the hTRPV1 homology models using the recently released rTRPV1 structure, their validation using a set of known agonists and antagonists, prediction of binding modes of some of the well known antagonists and a virtual screening exercise using the putative binding site of the antagonist [43].

Topomer-CoMFA is an alignment-independent approach that retains the CoMFA philosophy and methodology, but makes the process of building CoMFA models less cumbersome and more

automated [31,32]. One of the most attractive features of this methodology is that the dependence of the quality of the CoMFA models on alignment of the dataset is done away with, as this critical step is automated and is invariable. In terms of utility, Topomer-CoMFA method has many useful attributes that other methods lack, such as: (1) Core-hopping (Fig. 1B), (2) expansion of SAR (Fig. 1C), (3) generation of novel chemotypes (Fig. 1D) and (4) R-Group-Hopping (Fig. 1E). SAR expansion and R-group-hopping are very similar operationally, but with different objectives. The former entails searching for new R-groups that can primarily enhance potency, while the latter deals with improving properties other than potency. This aims at retaining potency as much as possible, while searching for R-groups that might improve other physico-chemical properties that in turn might affect some ADME outcomes.

Topomer-CoMFA has been mostly used to study and understand the SAR of a dataset or determine the critical structural requirements for a modulator of a target [S1–S18, table S2 of supplementary information]. While some studies report design of new compounds based on the developed models, some studies report using this methodology for R-group searches and a few studies report unique applications [S19–S37, table S2 of supplementary information].

In addition to the topomer-CoMFA model that is built using the piperazinyl-aryl series in this work, we also employ the pharmacophore model reported earlier [28], and conduct a shape query to virtually screen the subset of the ZINC database to find novel chemotypes. To our knowledge, there have been a few reports of predictive and quantitative 3D-QSAR and pharmacophore models, but most of them for understanding the structure-activity relationship of the respective datasets. Except for the work by Goldmann *et al.* and Feng *et al.*, no other effort proceeded further to use the developed models for virtual screening. In the current work, application of two different QSAR methodologies is carried out followed by virtual screening thereby illustrating the composition, quality and diversity of the top ranked hits from multiple screens. These efforts have yielded novel chemotypes that can be potentially used to develop modality-specific chemotypes or to probe the role of TRPV1 for new indications. We have listed and analyzed outliers in the 3D-QSAR models to understand the reasons for their outcome. This analysis informs us about the limitations of the models and how they compare to the models in our earlier work [28]. We also analyze a set of top-ranked hits from each screen in the context of various aspects such as their diversity, distribution over ranges of physico-chemical properties and predicted activities. This analysis of virtual screening results informs us of the quality and diversity of hits and strengths and weaknesses of each screening method. Finally, the hits from each screen are clustered and a representative set of hits from each screen is illustrated to show their novelty & diversity.

2. Methodology

2.1. Dataset preparation

The structures of 98 VR1 antagonists [29,30] were sketched in SymXDraw and saved in the form of a multi-molecule MDL-SDF formatted file, which was used as input for Topomer-CoMFA calculations [31,32] using the Graphical User Interface in Sybyl-X 1.3 [44]. Using the SELECTOR module in Sybyl-X, the complete dataset of 98 molecules (Table S3 in Supplementary information) was split into training and validation sets. The SELECTOR module uses a protocol, in which diverse subsets of ligands are selected in a space defined by three parameters, namely, UNITY fingerprints, Atom pairs and biological activity measured as pIC_{50} of VR1 antagonism, and identified to form the validation sets. The three parameters

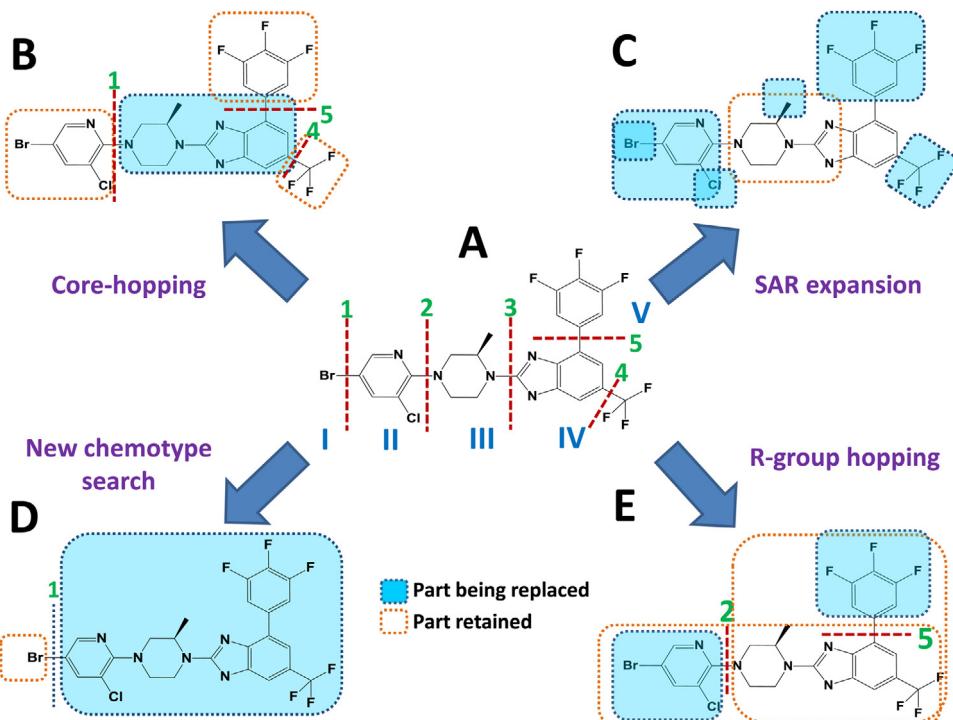


Fig. 1. Schematic illustration of the varied uses of topomer (R-group) search with **JM-45** as an example compound; A. Examples of acyclic bonds that can be cut during fragmentation; others also can be cut. B. Core-hopping; once acyclic bonds 1,4 & 5 are cut, the core (blue-colored dashed box) can used as a query. C. Virtual screening; replacements for the whole of fragment in blue-colored dashed box. D. SAR expansion; each of the fragments in blue-colored boxes can be topomer search queries to get a list of groups that can potentially enhance potency. E. R-group hopping; examples of two groups in blue-colored dashed boxes that can be replaced by hits from topomer searches towards improving any physico-chemical property.

can be assigned different weights to bias the selection. For this dataset, the biological activity was given a weight of 5 while the other two were given a weight of 1, to ensure a balanced split of active and inactive compounds in the selected subsets. The remaining compounds corresponding to each of the subsets form the corresponding training sets. This process was repeated four times with different numbers of members in the training sets (as well as corresponding test sets) to yield four sets of training and validation sets (Tables S4 through to S7 in Supplementary Information). One of the 98 compounds is **BCTC**, (*N*-(4-tertiarybutylphenyl)-4-(3-chlorophyridin-2-yl) tetrahydropyrazine-1(2*H*)-carboxamide; compound ID: **JM-4**), that is based on piperazine-1-carboxamide and is structurally slightly different from the piperazinyl-aryl series that is used in this work. Initially, including it in the dataset for building Topomer-CoMFA models proved problematic as its alignment to all other compounds was not good enough after splitting and hence was excluded from the dataset. It was included only in one of the datasets used for building Topomer-CoMFA and in all the datasets used for building CoMFA models. Table S4 represents a 58–39 split into training and test sets TRG1 and TEST1, respectively. Tables S5 and S6 represent 60–37 and 67–30 splits corresponding to two separate selections (TRG2 and TEST2, TRG3 and TEST3). Finally, Table S7 represents a 79–19 split (TRG4 and TEST4).

2.2. Topomer-CoMFA models

Topomer-CoMFA models were developed for each of the training sets using the “Split into 2” splitting scheme. In this splitting scheme, the two fragments R1 and R2 are created by splitting the molecules on either side of the bond connecting the imidazole (blue) and the piperazine ring (red) as illustrated in Fig. 2A using **BM-2**.

As previously described, model building in Topomer-CoMFA involves generation of CoMFA models for the training set fragments overlaid on top of one another using Topomer rules. In this study, CoMFA models would thus be generated for the R1 and R2 fragments resulting from the “Split into 2” fragmentation scheme. These models are then collectively regressed to the biological activity using the PLS (partial least squares) methods. Various statistical parameters are computed and reported in the Topomer-CoMFA GUI, while the predictions are reported in Sybyl-X molecular spreadsheets. The models were subsequently used to predict the activities of the corresponding test set compounds. Two-dimensional correlation plots were generated between the experimental and predicted pIC_{50} values for VR1 antagonism, using Microsoft Excel. Trend lines characterizing these correlations are also shown. We have illustrated the Topomer-CoMFA models graphically in which favorable steric contributions to activity are represented by green-colored regions, while unfavorable steric contributions are shown by yellow-colored regions. Similarly, favorable electrostatic contributions are represented by red- and blue-colored regions for partial negative and partial positive charges respectively. By extension, these regions are detrimental for activity when occupied by partial positive and partial negative charge carrying atoms, respectively.

2.3. CoMFA models

In addition to Topomer-CoMFA models, we have also generated CoMFA models of the VR1 antagonists using the QSAR project manager functionality in Sybyl-X 1.3. To obtain the alignment of all the ligands, the following workflow was employed: (a) the most active molecule in the series (**JM-37a** with a pIC_{50} value of 9.53; Fig. 2B) was energy minimized using the MMFF94s force field [45–50] in Sybyl-X, (b) the resultant structure was used to generate a

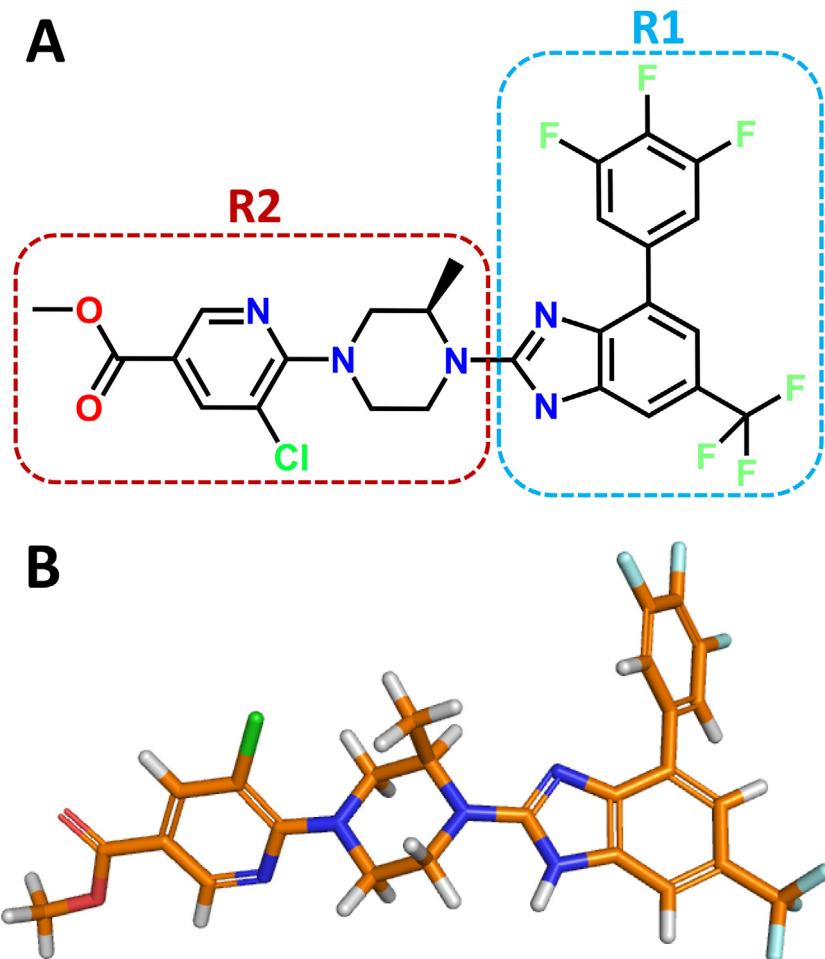


Fig. 2. (A) Splitting scheme used to define the two fragments (R1 – blue and R2 – red) for the purpose of developing Topomer-CoMFA models of VR1 antagonists. As an example, compound **BM-2** is used to illustrate the splitting scheme which is automatically applied to all the members of the training and test sets based on 2-D topology similarity around the split bond. (B) Low energy minimized 3-D structure of **JM-37a** used as a reference in the Surflex-SIM calculations with atoms colored by atom types – C (Orange), N (Blue), H (White), O (red), F (Light-blue), Cl (Green). The top-ranked alignments of training set and test set ligands to this reference molecule were used in the generation of 3-D QSAR (CoMFA) models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conformational ensemble of **JM-37a** using the random conformation generation method in Sybyl-X, while continuing to employ the MMFF94s force field, (c) the lowest energy structure resulting from such a search was taken to represent the template of **JM-37a** against which all the other molecules in the data set of 98 were overlaid. Incidentally, this lowest energy conformation of **JM-37a** is very close to the conformation of the reference compound of the pharmacophore model that we had developed earlier [28] using the Schrodinger/Phase methodology, with an energy difference of 0.8 Kcal/mol. Since the conformation of the reference compound of the pharmacophore model can be deemed to be very close to the putative bioactive conformation, the lowest energy conformation of **JM-37a** can itself be considered representative of the putative bioactive conformation. In the work reported by Feng et al. [43], the predicted binding mode of AMG2674 – a representative of the series of piperazinyl-aryl compounds used in this work – in the binding site of capsaicin is visually very similar to the lowest energy conformation of **JM-37a**. All the molecules of the dataset were aligned using the Surflex-SIM methodology [51] using the graphical user interface in Sybyl-X with default options. The pose with the highest similarity to the template structure of **JM-37a** was retained for each compound to obtain the collective alignment.

The total set of 98 molecules was then divided into three sets of training and test sets (listed in Tables S8, S9 and S10) using the diversity-based method OptiSim inside the QSAR project man-

ager of Sybyl-X. The OptiSim method involves extracting a subset of size k (compounds) from the whole dataset randomly. From each subset, the compound that is most different from compounds selected from earlier subsets is added to the selection set. This method ensures that compounds that are structurally representative of compounds in the whole dataset, but diverse from those in the selection set are incorporated into the test set [52]. The similarity of the compounds is determined by Tanimoto similarity based on UNITY fingerprints. Compounds having similarity greater than 0.8 to any other compound in the test set is excluded from further selection in subsets. This procedure was repeated three times with an arbitrary $k=6$ and $n=98$, each time using a different random number as seed. The 3-D CoMFA fields [53] were generated using essentially default parameters. For electrostatic fields, Gasteiger-Huckel partial atomic charges were employed for all the molecules. The models generated using the partial least squares (PLS) method were employed to predict the activities of the corresponding test sets and correlation plots were plotted using the Microsoft Excel program. As in the case of Topomer-CoMFA models, here too the graphical illustrations are characterized by green and yellow zones corresponding to sterically favorable and unfavorable regions, respectively. Similarly, electrostatically favorable regions are indicated by blue (partial positive charge, e.g. donor hydrogen atom) and red (partial negative charge, e.g. acceptor atom) regions.

Table 1

Results of Topomer-CoMFA and CoMFA models.

Model	# Comp	# Cpd	Intercept	LOO Q ²	R ²	R ² Stderr	F	Test set Pearson R	R ² (test set)	RMSE
Topomer-CoMFA models										
TRG1	2	58 (39)	5.71	0.53	0.73	0.56	65.7	0.78	0.61	0.78
TRG2	2	60 (37)	5.87	0.48	0.66	0.61	34.8	0.83	0.67	0.74
TRG3	3	67 (30)	5.66	0.54	0.70	0.55	34.4	0.85	0.63	0.85
TRG4	2	79 (19)	6.21	0.48	0.63	0.73	41.5	0.79	0.62	0.66
CoMFA models										
TRG5	5	73 (24)	4.21	0.52	0.82	0.43	67.7	0.79	0.58	0.90
TRG6	3	68 (29)	5.23	0.56	0.79	0.46	55.3	0.69	0.47	1.01
TRG7	4	63 (34)	4.76	0.55	0.73	0.36	29.5	0.79	0.60	0.86

Note: # Comp: number of components; # Cpd: number of compounds indicated as 'training set (test set)'; LOO: Leave-One-Out; R² Stderr: standard error in R²; RMSE: root mean squared error.

2.4. Virtual screening

One of the main aims of this study is to find novel chemotypes that can probably yield modality-specific antagonists of TRPV1 channel [16,17,20–22]. Literature evidence indicates that TRPV1 antagonists that block Capsaicin-induced activation of the channel but not proton- or acid-induced activation of the channel appear to be temperature neutral or in some cases lead to slight hypothermia [13,16–20]. The differential pharmacology appears to be chemotype-independent. Several mutagenesis studies indicate different binding sites for compounds that block capsaicin-induced activation and those that block proton- or acid-induced activation [54–57]. In this situation, starting with diverse chemotypes may be a viable option to find specific chemotypes that differentially block only the Capsaicin-induced activation and not acid-induced activation of the TRPV1 channel. Towards this end, we undertook virtual screening of the ZINC database of commercially-available compounds using three different methods to find several hits that can be further developed into TRPV1 antagonists: (1) Topomer-CoMFA model for finding hits based on R1 and R2 topomers of the model, (2) a Schrodinger/Phase pharmacophore model that we had build and reported in our earlier work [28] and (3) Schrodinger/Phase Shape screening using one of the potent compounds (**JM-45**) from the piperazinyl-aryl dataset as the query compound. We chose to use **JM-45** as the shape query instead of **JM-37a**, the most active compound of the dataset, as it was the reference compound of the Schrodinger/Phase pharmacophore model developed earlier.

Of all the CoMFA models developed in this work, the Topomer-CoMFA model developed using the training and test sets TRG2 and TEST2, respectively, gave better statistics and this model was used to screen a subset of the ZINC database to find potential replacements for the R1 & R2 groups. The Phase shape screening was done twice, once using pharmacophore features and once using Macro-Model atomtypes for scoring the hits. For all screening runs, we have used the respective default parameters.

In order to make sure that the hits from various screening approaches are indeed promising, we took a subset of hits from each screening approach and visually inspected them. This analysis of hits revealed a lot of information about the quality and diversity of hits. As clustering based selection of compounds to maintain structural diversity was prescribed in literature, we clustered the hits from each screening approach to select diverse compounds for further work.

3. Results and discussions

Table 1 summarizes the statistics associated with the Topomer-CoMFA models developed using the four training sets. The standard formulae used to calculate these values have been described in the manuals associated with the software modules used for building QSAR models [58] and hence not recollected here. The R² val-

ues associated with the training set compounds range from 0.6 to 0.7, the leave-one-out (LOO) cross-validation Q² values are around 0.5, while the R² values associated with the test set compounds are around 0.6. Thus, these models are deemed to be statistically significant in terms of their ability to make predictions. The standard errors associated with these models are comparable to typical experimental standard errors of about 0.5–0.8. The following sections describe the models obtained from each of the training sets.

The Topomer-CoMFA models for the two R groups (R1 and R2) obtained from the training set TRG1 compounds (Table S4) are shown separately in Figs. 3a and b.

The Topomer-CoMFA model for R1 is characterized by a large one-sided region of favorable steric interactions around one of the aromatic substituents on the imidazole ring which is connected to the piperazine. The positive electrostatic fields are symmetrically distributed around the imidazole ring while regions of favorable negative electrostatic fields are somewhat sparse by comparison. In the case of R2, the Topomer-CoMFA model seems to favor the presence of an electronegative functional group (e.g. aromatic nitrogen of the pyridine ring) or a hydrogen bond acceptor in the *ortho* position of the ring system attached to the piperazine. The substituent in the corresponding *para* position needs to have hydrogen bond donor functionality to ensure activity. On the other hand, the *para* position seems to be much more tolerant of steric bulk compared to the *ortho* or *meta* position substituents on the aromatic ring attached to the piperazine in R1. The values listed in the last two columns in Tables S4–S7 for the test sets correspond to the contributions of fragments R1 and R2 to the total activity. The predicted activity can be obtained by adding up these contributions to the intercept for the model, which is 5.7092 for TEST1. The model with the best value of cross-validated Q² (0.5) is characterized by two PLS components with an R² value of 0.7. The correlation plot for the training and test sets is shown in Fig. 4A. The Topomer-CoMFA models for TRG2 (Table S5) are qualitatively and quantitatively similar to that obtained with TRG1. As seen from Table 1, the statistics for training and test set predictions are very similar to those of TRG1. The correlation plot for the training and test sets is shown in Fig. 4B.

Larger training sets (TRG3 and TRG4 listed in Tables S6 and S7, respectively) do not necessarily lead to significantly different Topomer-CoMFA models and the corresponding prediction set statistics do not differ from those for TRG1 and TRG2. The correlation plots between experimental and predicted activities derived from Topomer-CoMFA models for TRG3 and TRG4 and their corresponding test sets are shown in Figs. 4C and D respectively.

Three more 3D-QSAR models using the CoMFA methodology were built primarily to check if the models provide any more information on the SAR of the dataset and also as a comparison to the models built using the newer CoMFA methodology, the Topomer-CoMFA. These were based on the alignment of the training set compounds obtained through Surflex-SIM studies. The CoMFA models were obtained with this alignment using the QSAR

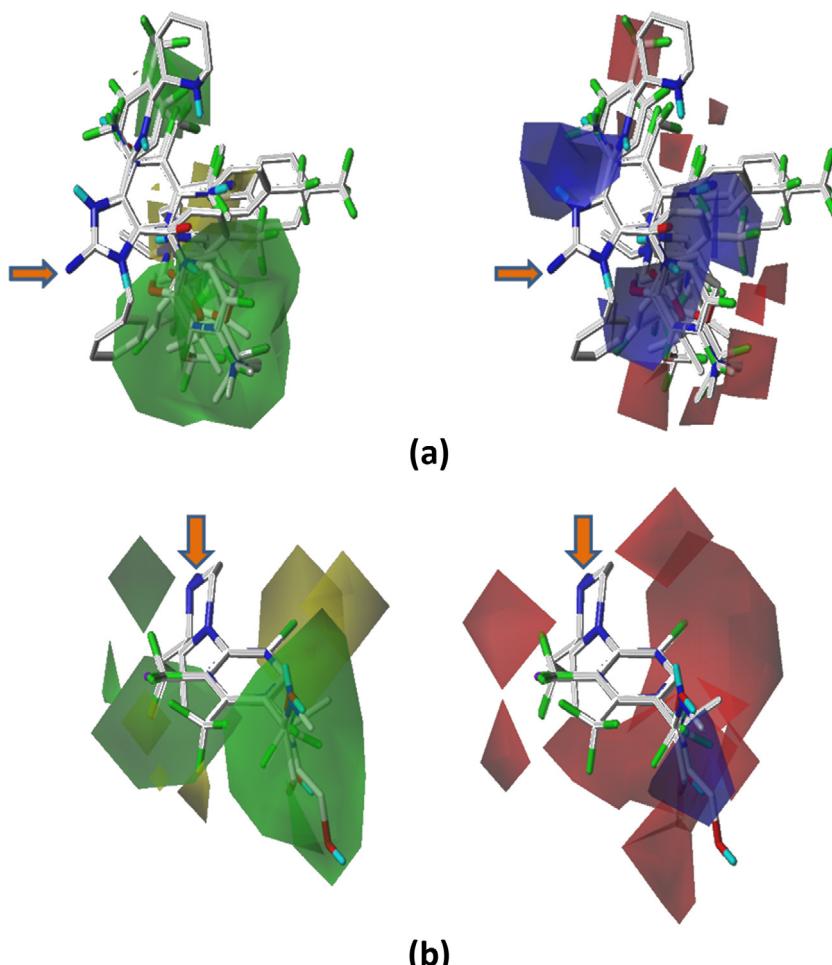


Fig. 3. Topomer-CoMFA models for R1 (a) and R2 (b) fragments obtained with the training set TRG1. The orange colored block arrow shows the split point in the generation of R1 and R2 fragments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

project manager (QPM) in Sybyl-X (version 1.3) employing primarily default options. For all the CoMFA models, the **BTC** compound (**JM-4**) was part of the test set and its activity was quite underpredicted, with the residual ranging from 3.1 (TRG5/TEST5) to 3.8 (TRG7/TEST7). Hence it was eliminated and the statistics for the test sets were recalculated. The statistics associated with the three models developed corresponding to three different training sets (and validated by corresponding three different test sets) are shown in Table 1. It is noted that the three models have very comparable R^2 (0.7–0.8) and Q^2 (0.5) values for the training sets (Fig. 5A to C). However, the test set predictions are appreciably better for the models obtained with TRG5 and TRG7 compared to the corresponding test set predictions from the model obtained with TRG6. The experimental pIC₅₀ values, the predicted pIC₅₀ values and residuals for all these three datasets (TRG5/TEST5, TRG6/TEST6 and TRG7/TEST7) are given in the tables S8–S10 in the supplementary information.

In light of the similarity of the statistics for the three training sets only the CoMFA model corresponding to TRG5 is illustrated in Fig. 6. It is interesting to note that sterically favorable regions (green-colored regions) in this map are also favored by the occurrence of donor atom types or atoms carrying partial positive charges as indicated by the blue regions of the CoMFA map. By comparison, the red regions are very small. The overall inference from the CoMFA models is depicted in Fig. 7. H-bond donors or electro-positive groups are tolerated and favored (1) at the 1st position of imidazole ring, (2) towards the 4th position of imidazole or benzimidazole system and (3) at the *ortho* and *para* positions of the

terminal phenyl ring. Hydrophobic groups or steric bulk is favored (1) towards the 4th position of imidazole or benzimidazole system and (2) at the *ortho* and *para* positions of the terminal phenyl ring. Steric bulk is also favored in the middle, near the piperazine ring and towards 6th position of the benzimidazole ring or 5th position of the imidazole ring. The model contours from the CoMFA models and the regression effects visualized in the Schrodinger/Phase model in our earlier work are very similar.

All the correlation plots show a line denoting the ideal correlation, flanked by two red-colored dashed lines on either side. These two lines indicate one log unit on either side of the ideal correlation line and help identify outliers in each case. We arbitrarily define an outlier as a compound whose predicted activity is more than one log unit from its experimental activity. As can be observed from the plots, all models have outliers. The number of outliers and the specific compounds whose activities were either over-predicted or under-predicted are given in Tables 2 and S11. Table S11 shows that the number of outliers among the test sets appears to be dependent on the number compounds in the training sets, with less number of outliers in the test set if more number of compounds are present in the training set. This could either be due to better training due to more number of compounds in the training set or due to less number of compounds in the test set. Topomer-CoMFA set 4 unusually has more number of outliers in the training set than in other training sets and the least number of outliers in its test set than in other test sets. Again this could be due to more number of compounds in the training set and less number of compounds in the test set.

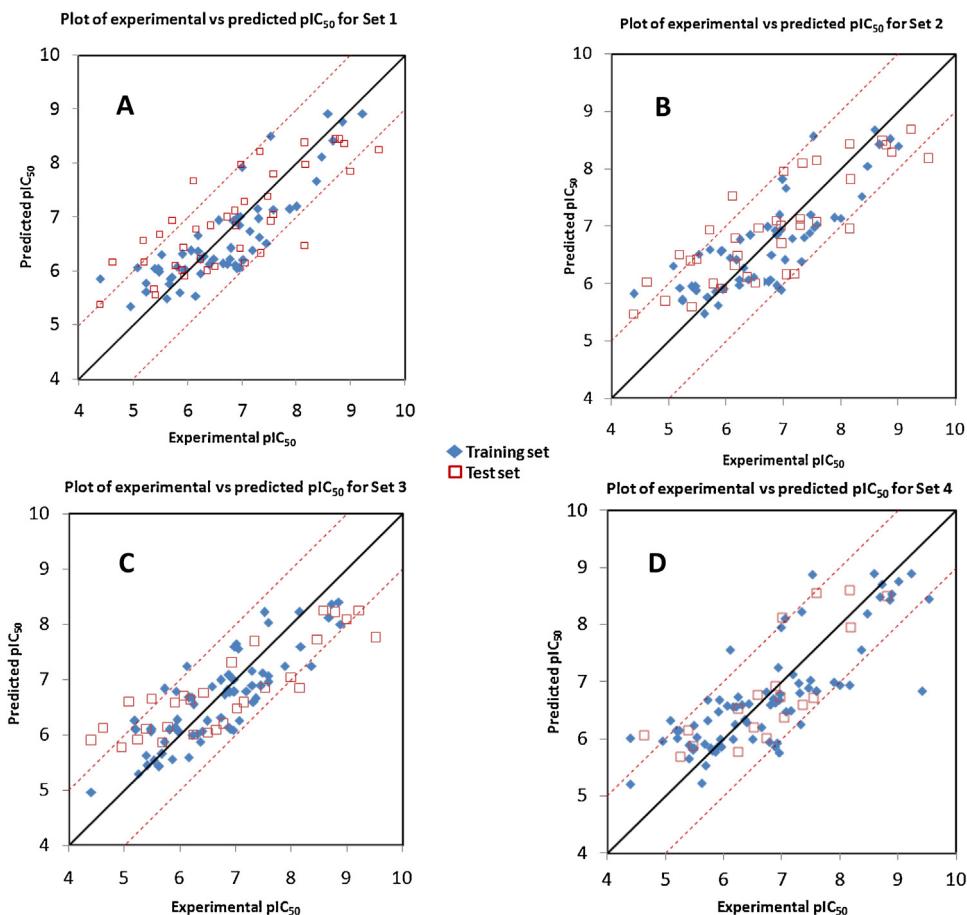


Fig. 4. A-D Correlation plot of experimental (X-axis) versus predicted (Y-axis) pIC_{50} of VR1 antagonism for the training sets TRG1–TRG4 (blue, diamond-shape) and corresponding test sets TEST1–TEST4 (squares with red-colored outline). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

The number of outliers in each dataset.

Set	# in Training	# in Test	Total	% of Training	% of Test	Test set R^2	Training set/Test set	Ratio
1	1	9	10	1.7	23	0.61	58/39	1.49/1
2	4	9	13	6.7	24.3	0.67	60/37	1.62/1
3	4	6	10	6	20	0.63	67/30	2.23/1
4	13	2	15	16.4	10.5	0.62	79/19	4.16/1
5	3	8	11	4.1	33	0.58	73/24	3.04/1
6	2	10	12	2.9	34.5	0.47	68/29	2.34/1
7	5	9	14	7.9	26.5	0.60	63/34	1.85/1

The Topomer-CoMFA models developed using the training and test sets TRG1 and TEST1 and TRG3 and TEST3 appear to give the best outcome in terms of number outliers, size of the training set and the test set R^2 . If one considers the sizes of training and test sets and the number of outliers in each, there does not seem to be any particular trend, though the CoMFA model 7 appears to have slightly higher number of outliers.

Table S11 shows that compounds **BM-23**, **JM-37a**, **JM-4**, **JM-46b**, **JM-46z**, **JM-50a**, **JM-51d**, **JM-52i** and **JM-53e** are outliers in most of the models. Of these, **JM-37a**, **JM-51d**, **JM-52i** and **JM-53e** have been consistently under-predicted (**JM-37a**) or over-predicted (**JM-51d**, **JM-52i** and **JM-53e**) by all the models. **JM-37a** is the only compound in the dataset that has a methyl ester group in the *para* position on the terminal pyridine ring. It is to be noted that although **JM-37a** (the most active compound of the dataset) was used to align all the compounds of the dataset, it does not influence the training of the 3D-QSAR model or its performance. It just acts as a means to get the alignment of all

compounds of the dataset and does not influence the model to have its activity predicted more accurately. While alignment is based on pharmacophoric features or the maximum common substructure (pyridine-cyclohexane-benzimidazole), prediction of its activity would be based on structural features and their molecular fields alone. Similarly, **JM-52i** is the only compound in the dataset that has an ethyl ester group in the *ortho* position on the terminal pyridine ring. **JM-53e** is the only compound in the dataset that has 2,6-dichloro substitution on the terminal ring which is a phenyl. **JM-53d** also has 2,6-dichloro substitution, but it is on the terminal pyridine ring, while **JM-53a** has a terminal phenyl ring, but it has trifluoro-methyl substitution in the *ortho* position. **JM-51d** is one of two compounds in the whole dataset to have a methyl substitution on the 3rd position of the piperazine ring, the other compound being **JM-51c**. Surprisingly, only **JM-51d** appears as an outlier in six out of seven datasets, while **JM-51c** appears as an outlier only in Topomer-CoMFA model 3. This appears to be due to the fact that the *R*-isomer (**JM-51c**) has the methyl projected

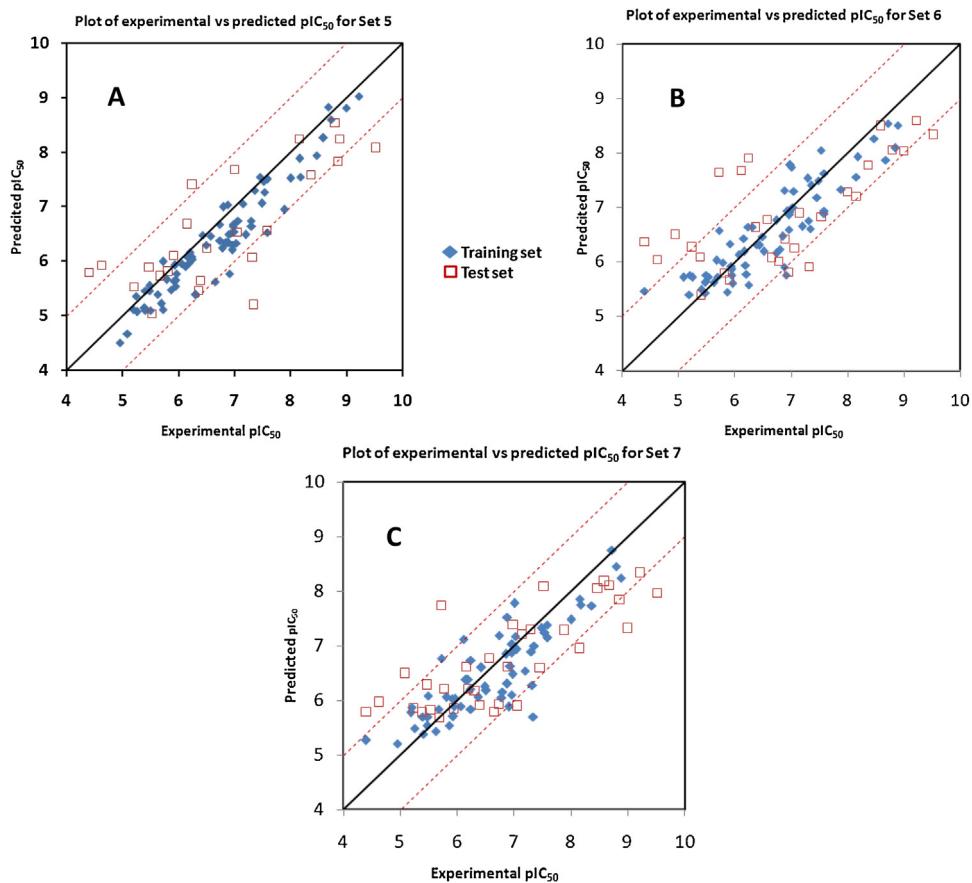


Fig. 5. A-C Correlation plot of experimental (X-axis) versus predicted (Y-axis) pIC_{50} of VR1 antagonism for the training sets TRG5–TRG7 (blue, diamond-shape) and corresponding test sets TEST5–TEST7 (squares with red-colored outline). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

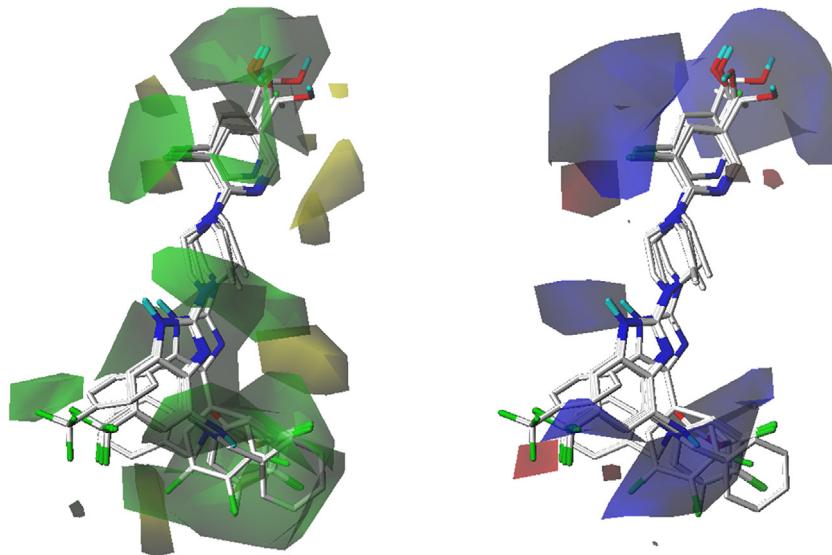


Fig. 6. CoMFA model derived from TRG5 training set. The five most active molecules in the training set are also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

towards the same region as the *ortho*-substituted hydrophobic groups, mostly halogens, on the terminal aromatic ring, while the S-isomer (**JM-51d**) has its methyl projecting away from the plane having the *ortho*-substituted hydrophobic groups on the terminal aromatic ring. Hence these hydrophobic groups appear to represent the hydrophobic region on the scaffold into which **JM-51c** also has

its methyl projected and hence mostly predicted well, while **JM-51d** has its methyl projected into an area mostly unrepresented by any groups. **JM-37a** appears in the test set in all models except for Topomer-CoMFA model 4. **JM-52i** appears in test set for all models, while **JM-53e** appears in the test set for Topomer-CoMFA model 3 and CoMFA models 5 and 7. **BCTC** compound (**JM-4**), being

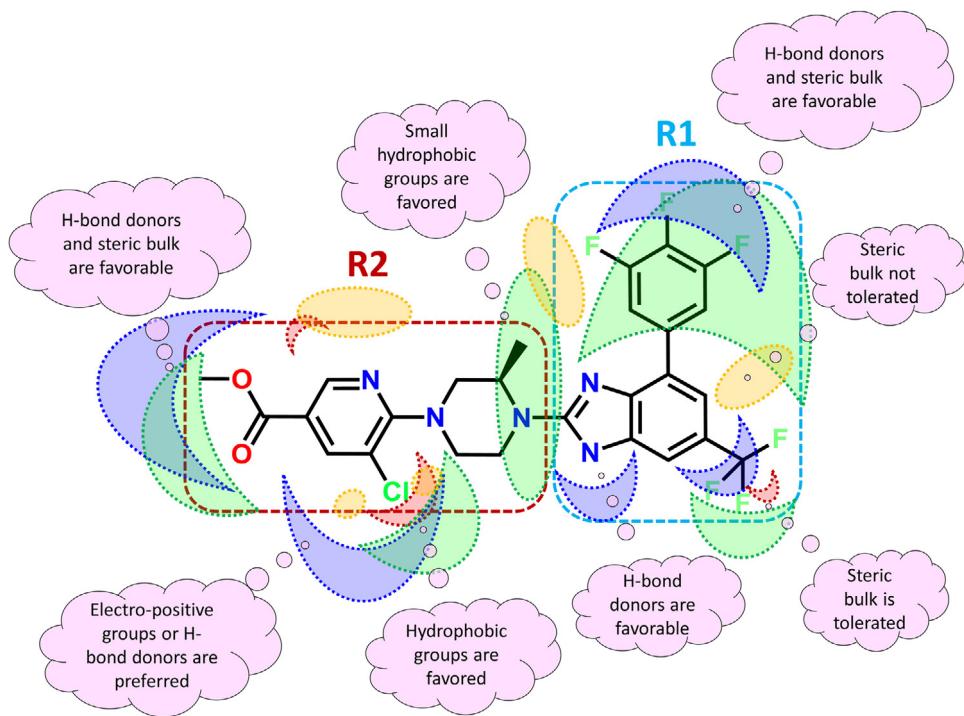


Fig. 7. Summary of the SAR of piperazinyl-aryl series of TRPV1 antagonists according to the CoMFA model contours. R1 and R2 represent the two fragments after splitting for the purpose of developing Topomer-CoMFA models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

unlike other compounds of the dataset in having a phenyl ring in R1, is not predicted well in all the three CoMFA models as well as the Topomer-CoMFA model 4. In general, compounds appear to become outliers if any of their structural features is not represented in the training set.

The various metrics for all models having **BCTC** (JM-4) in their datasets appear to suffer due to the occurrence of **BCTC** compound as an outlier. In our earlier work, the 3D-QSAR model built using Schrodinger/Phase did not over- or under-predict the activity for **BCTC** and the pharmacophore was applicable for other chemotypes. In contrast, the Topomer-CoMFA and the CoMFA models built in this work appear to be sensitive to other chemotypes. Once the **BCTC** compound was removed from the datasets (for all CoMFA models), the re-calculated metrics appear much improved. This analysis of outliers was done to understand why those compounds are outliers, due to either under- or over-prediction and also about how sensitive the models are to structural changes in some compounds of the dataset. Most of the models appear to be less robust if the dataset has any outliers that are structurally different from other dataset compounds.

4. Virtual screening

Virtual screening helped us to mine the databases to find as many diverse hits as possible and also to understand the strengths and weaknesses of each approach. The results of the various screens are given in the Table 3 below.

The top-ranking 15 R1 and 15 R2 topomers/fragments, found by screening the ZINC database using the Topomer-CoMFA model, were selected for further analysis. Combining R1 fragments with R2 fragments would give 225 unique compounds. To compare the hits from the different screens, 225 top ranked hits were taken for analysis from each screen. Although the number of hits from each screen differs, analysis using the top-ranked 225 compounds from each would give an idea of the quality and kind of hits from each screen.

Fig. 8 shows four pairs of virtual screening hits, each pair consisting of the hits with highest and the lowest value of the parameters, fitness (**Fig. 8A**, “Fitness”), Shape similarity calculation with volume scoring involving MacroModel atomtype overlap (**Fig. 8B**, “Shape similarity-MacroModel atomtypes”), Shape similarity calculation with pharmacophore feature overlap for volume scoring (**Fig. 8C**, “Shape similarity-Pharmacophore features”) and dataset similarity (**Fig. 8D**, “Dataset similarity”). The various parameters in the Table 3 have been described in the following sections.

4.1. Topomer-CoMFA screen

A typical Topomer-CoMFA search against a conventional database involves generation of fragments from each database compound by breaking apart each compound around each acyclic single bond. In a R-group search, each such fragment will be matched to the query fragment in terms of shape similarity and pharmacophoric composition. For example if each query structure is composed of two query fragments, then in a R-group search, the resulting hit fragments for each query fragment can come from different original database compound. This is illustrated in **Fig. 9**. In order to preserve the geometry at the point of breaking of an acyclic bond, the resultant fragments tend to have same or very similar group at that point and diversity in hit fragments tends to increase on moving away from the bond that is broken. This is very typical in a conventional lead optimization program wherein, during the course of designing new compounds for optimizing potency or any other property, the core is preserved as much as possible while the outward lying R-groups are changed. In the R-group search done in the current effort, the query compound has been broken roughly into two equal-sized fragments and the ZINC database searched for each fragment. The resulting hit fragments retain the essential core structure but have different R-groups. The new fragments represent good groups to triage in a typical lead optimization stage. This kind of R-group search also represents a unique way of

Table 3
Results of the three screens.

Attribute		Pharmacophore screening	Shape screen – MM	Shape screen – Pharm	Topomer-CoMFA	Pharmacophore screening: all compounds
Fitness	Average	2.10	0.29	0.37	0.54	1.24
	High	2.26	0.54	0.68	0.89	2.26
	Low	2.05	-0.16	-0.20	0.16	-0.39
Shape sim-Pharm	Average	0.44	0.35	0.52	0.35	0.34
	High	0.57	0.52	0.59	0.47	0.60
	Low	0.30	0.21	0.5	0.14	0.16
Shape sim-MacroModel	Average	0.48	0.71	0.48	0.60	0.44
	High	0.62	0.81	0.65	0.72	0.70
	Low	0.31	0.68	0.31	0.38	0.17
Dataset similarity	Average	0.40	0.31	0.29	0.36	
	High	0.51	0.40	0.42	0.43	ND
	Low	0.17	0.20	0.13	0.21	
Similarity with template	Average	0.15	0.21	0.17	0.26	
	High	0.21	0.34	0.26	0.42	ND
	Low	0.09	0.11	0.09	0.14	
Scaffold similarity with template	Average	0.35	0.48	0.32	0.69	
	High	0.57	0.87	0.50	0.93	ND
	Low	0.19	0.23	0.10	0.32	
Scaffold similarity within dataset	Average	0.57	0.40	0.32	0.63	
	High	0.69	0.49	0.44	0.72	ND
	Low	0.28	0.28	0.11	0.37	
Predicted activity*	# compounds	225	225	225	225	36,513
	≥7	0	0	1	55	6
	7–6	34	23	36	141	2893
	6–5	188	165	178	29	29082
	5–4	3	37	10	0	4532
Molecular properties						
Mol.Wt	Average	376.5	430.14	412.72	533.94	413.84
	High	517.7	573.3	554.7	628.5	599.30
	low	272.4	315.4	271.2	438.6	201.10
AlogP	Average	3.72	5.13	4.66	6.50	3.94
	High	8.50	7.20	7.7	8.40	9.80
	low	0.00	2.60	2.1	3.40	-2.00
PSA	Average	82.07	54.20	75.53	78.38	83.86
	High	146.2	126.90	119	190.90	231.20
	low	37.9	18.80	25.8	35.20	12.00
AMR	Average	103.51	123.41	111.40	153.42	114.28
	High	149.30	143.90	156.2	179.60	176.00
	low	76.00	98.20	65.5	130.70	56.40
Strain (Kcal/mol)	Average	2.37	3.74	2.15	2.77	349.34
	High	10.40	17.26	17.83	13.14	1063110.21
	low	0.0174	-4.00E-06	-5.00E-06	-1.50E-05	-2.10E-05

Notes: ND = Not Determined. * pIC50 values; Shape screen-MM: shape-based screening in which each structure is treated as a collection of atoms, whose volumes are defined by their van der Waals radii. Scoring is done by computing overlapping volumes between atoms of the same type. Shape screen-Pharm: Each structure as a collection of pharmacophore sites and the scoring is by computing the overlapping volumes between sites of the same feature type. Shape-sim Pharm: Shape similarity calculation with pharmacophore feature overlap for volume scoring. Shape-sim MacroModel: Shape similarity calculation with volume scoring involving MacroModel atomtype overlap. PSA: Polar Surface Area. AMR: Atomic Molar Refractivity.

breaking compounds and recombining the fragments to form new, novel and diverse compounds.

4.2. Pharmacophore screen

Out of the 225 top-ranked hits from the pharmacophore model-based screening, 199 compounds had a similar linker, the acetamide-thioether linker, though they had diverse moieties corresponding to termini of the dataset compounds (piperazinyl-aryl core). Of the 26 compounds that had a different linker, 25 compounds either have a methoxy-ethyl-methyl-thioether linker or just a thioether linker and the remaining compound has a piperazine linker. In most of the compounds, the acetamide-thioether linker or the methoxy-ethyl-methyl-thioether linker or the thioether linker appears to be loosely mapped to the hydropho-

bic feature in the center of the pharmacophore model. Among the various moieties corresponding to R1 and R2 Topomers, relatively diverse rings represent R1, while R2 is mostly a phenyl ring with varied substitutions. To check if any diversity can be found in the linker by clustering the hits from the screen and checking representative compounds from each cluster, the top-ranked 10,000 hits from the pharmacophore screen were clustered using k-means clustering method [59] into 225 clusters and the cluster centers of the 225 clusters were extracted. Analysis of these cluster centers revealed that although the moieties representing R1 and R2 were very diverse, the linkers tend to be very similar: linear amide- or ether- or thioether-containing linkers that linked ring systems representing R1 & R2 on their either sides. This could be due to the fact that a linear linker enables a better mapping of all pharmacophore features due to its flexibility and thus improve fitness than

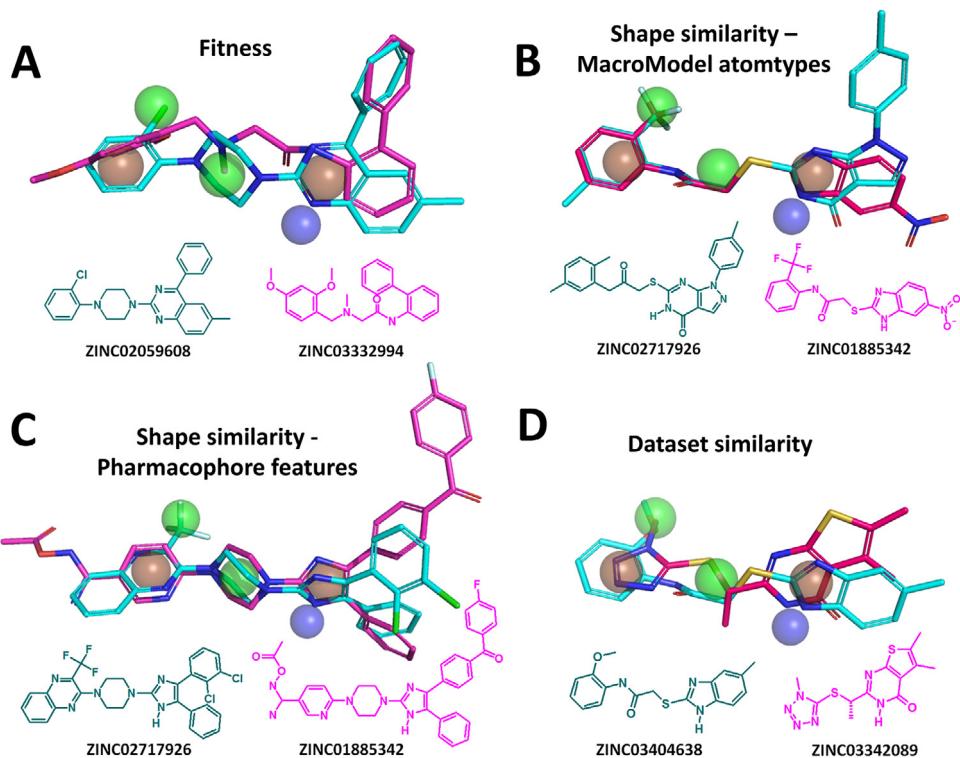


Fig. 8. Four pairs of virtual screening hits, each pair consisting of the hits with highest (cyan-colored; left) and the lowest (magenta-colored; right) value of the parameters: (A) Fitness (“Fitness”), (B) Shape similarity calculation with volume scoring involving MacroModel atomtype overlap (“Shape similarity-MacroModel atomtypes”), (C) Shape similarity calculation with pharmacophore feature overlap for volume scoring (“Shape similarity-Pharmacophore features”) and (D) Dataset similarity (“Dataset similarity”). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

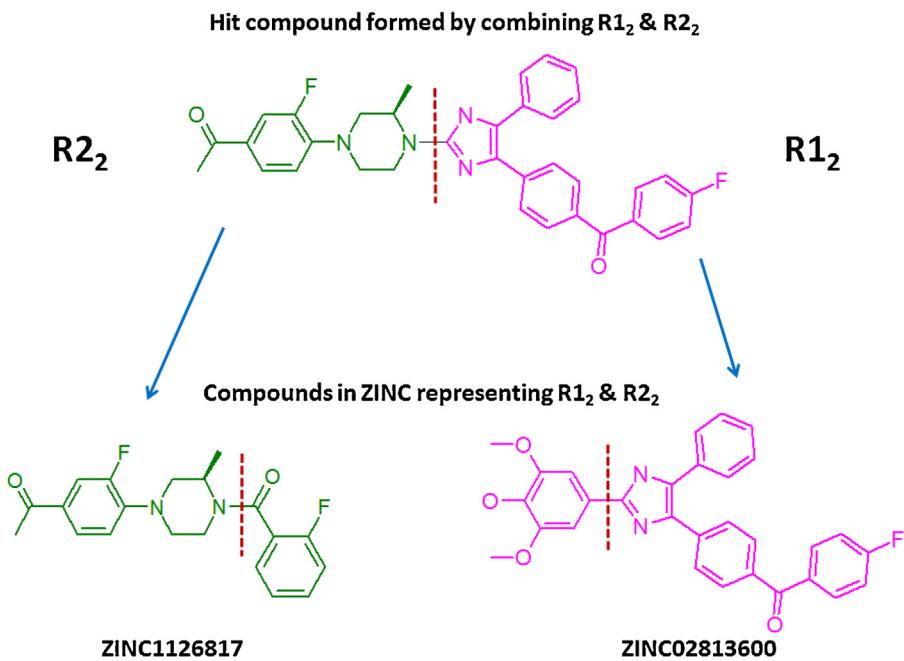


Fig. 9. An example of a hit (R1 + R2) and its parent compounds from Topomer-CoMFA screen.

a non-linear linker. The fact that a hydrophobic pharmacophoric feature is a non-vector group feature also helps a linear flexible moiety in the ligands to match/map this feature. The hits do reveal that mapping of pharmacophore features other than the middle hydrophobic features is much better and the middle hydrophobic feature is very loosely mapped, mostly by sulphur-containing linear linkers. The occurrence of sulphur atom among the linear

linkers is very high among all the hits in this screen. The scant occurrence of sulphur atom-containing linkers or linear linkers among other screens, especially the shape screens indicates that there is no dearth of diverse linkers in the ZINC dataset. Another observation is that most compounds in the set after clustering appear to be strained (average strain for k-means set is 4.67 Kcal/mol compared to 2.37 Kcal/mol for the original set of 225 compounds). **Fig. 10**

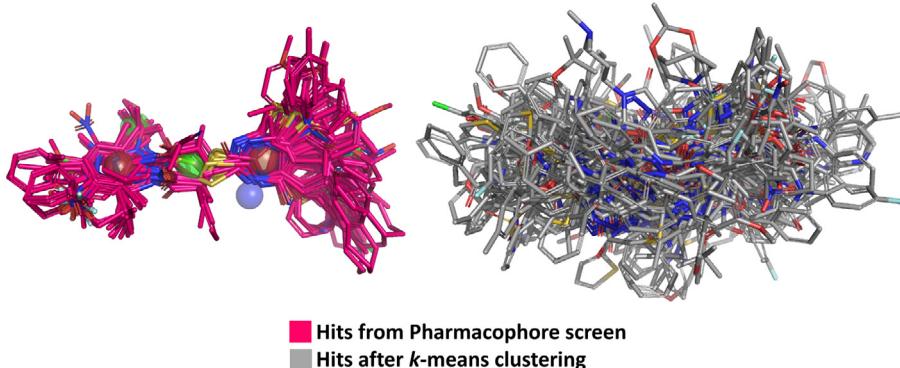


Fig. 10. Top 225 hits of the pharmacophore screen and the 225 hits from clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

showing all the hits in the original set and in the *k*-means set clearly indicates that the hits after clustering tend to be more diverse, and less aligned spatially.

4.3. Shape-screen with MacroModel atomtypes-based volume scoring

Of all the top-ranked 225 hits of the shape search with MacroModel atomtype-based volume scoring, only four compounds had a hydrogen-bond donor feature that maps to the corresponding feature of the pharmacophore model. Most of the compounds have piperazine linker and halogen-substituted phenyl or naphthyl as the R2 moiety. Quinazoline appears as a commonly occurring R1 moiety. Hence this dataset has less diversity in the middle and one of the terminal parts and most of the compounds lack a hydrogen bond donor feature, which is very essential for potency [29]. To improve the hits in terms of having a hydrogen-bond donor feature, the screen was repeated with increased weight for the nitrogen and its proton that together represent the hydrogen-bond donor feature in the piperazinyl-aryl dataset compounds. The repeat screen gave much improved results. A lot of diverse moieties corresponding to R1 and R2 and slightly diverse linkers were observed among the hits, though piperazine still is the commonly occurring linker. All the hit compounds have the hydrogen-bond donor feature represented by more diverse moieties compared to the original hits. Hence, much improved hits resulted by applying appropriate constraints in this screen.

4.4. Shape-screen with pharmacophore features-based volume scoring

Most of the hits in this screen do not match all the five features of the pharmacophore model. Among the 225 hits, only 86 compounds have a R1 moiety other than pyrimidinone as part of the core. Among the other R1 moieties, 30 have benzimidazole as part of the core. As in the case of pharmacophore model based screening, most of the hits have either an acetamide-thioether or a thioether as a linker linking the rings of R1 & R2. Only 23 compounds have a different linker. So, this screen gave hits that are less diverse in terms of R1 moieties and the linker. As in the case of hits from the pharmacophore screen, linear linkers allow better mapping of pharmacophore features and hence their common occurrence in screens that require matching of pharmacophore features. This may also be true in screens using pharmacophore models where a hydrophobic feature represents a linker.

All the hits from each screen were then analyzed in terms of their fitness to the pharmacophore, shape similarity with pharmacophore features or MacroModel atomtypes used for scoring,

similarity of the compounds compared to the query compound, similarity of compounds within their respective datasets (intra-dataset similarity), distribution of compounds in various intervals of predicted activity when the activity of compounds was predicted using the 3D-QSAR model, built and reported by us in an earlier report [28], distribution of hits over ranges of molecular weight, calculated log P (AlogP) [60], and polar surface area (PSA). Additionally, in order to discount the similarity due to side-chains on the scaffold and to get an idea of diversity of scaffolds within the hits, the compounds were fragmented following the rules proposed by Murcko & Bemis [61] and the resulting scaffolds were compared: (1) similarity of the scaffolds compared to the scaffold of the query compound and (2) similarity of scaffolds within their respective datasets (intra-dataset similarity). Each of the aspects analyzed are briefly described below.

4.5. Fitness

Fitness of the hit compounds from each screen was calculated based on their alignment with the pharmacophore. As expected we found that hits from the pharmacophore screen showed the highest fitness, followed by those from the screen using Topomer-CoMFA model. The compounds from the shape screen were less fit to the pharmacophore, again as expected. Pharmacophore-based screening using a pharmacophore model would screen based on matching pharmacophore features in database compounds and the deviations of matching features from their respective features in the query pharmacophore is checked and restricted by the default tolerances. Hence their fitness to the query pharmacophore is expected to be high. Similar logic can be ascribed to Topomer-CoMFA model based screening also. On the other hand, shape-based screening with volume scoring based on criteria such as pharmacophore features or MacroModel atomtypes would not consider matching pharmacophore features or atom types as the primary criterion, but would use feature overlap or atomtype overlap only for scoring purposes. The scoring would help in choosing the best alignment to the query compound shape from among the possible alignments using all the available conformations for each compound. Hence the hits from shape-based screening are expected to be less fit to the pharmacophore. In fact, the alignment of a hit compound to the query compound and to the pharmacophore model need not be same as illustrated by an example: ZINC03019475 is the only common compound, common in pharmacophore screening and shape screening with pharmacophore feature-based volume scoring, found among the top-ranked 225 compounds from each screen. As Fig. 11A below shows, although the conformations are very similar, mapping to the pharmacophore features differs, with better alignment observed for conformation

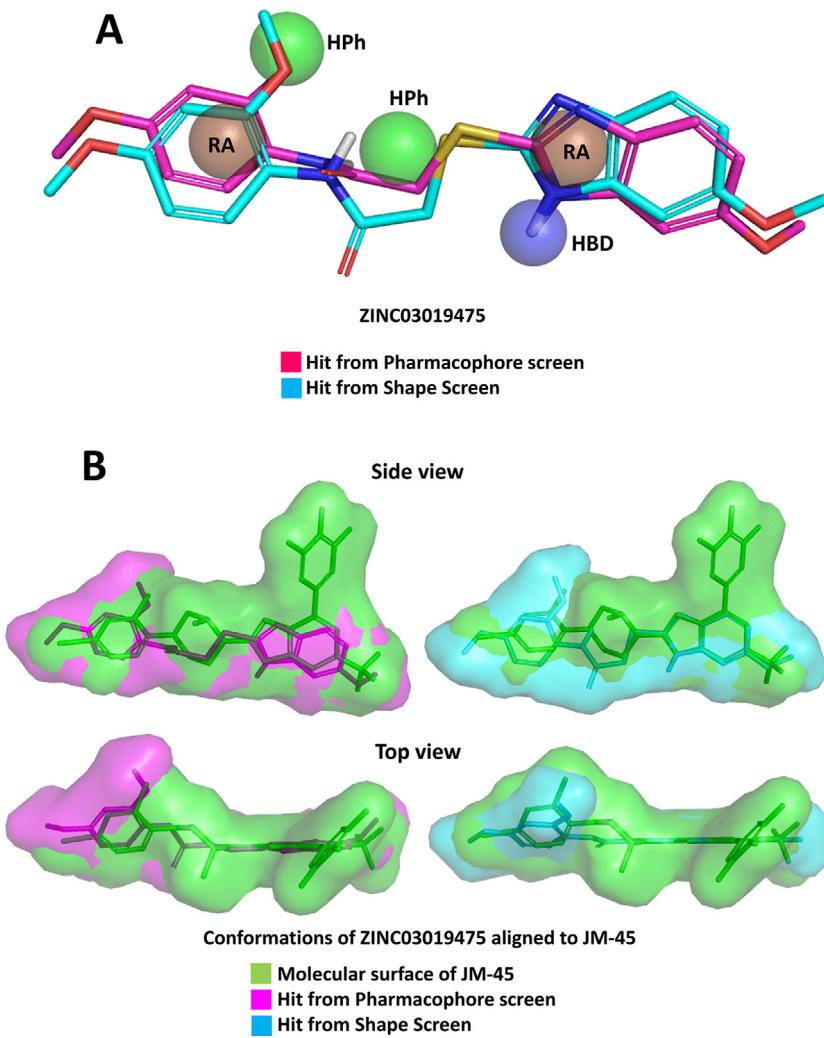


Fig. 11. (A) The compound, ZINC03019475, from the pharmacophore screen and shape screen in the context of the pharmacophore model. (B) Two views of the same compound represented as a molecular surface compared to the molecular surface of JM-45, the query compound. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

from the pharmacophore screen. The fitness value for the conformation from pharmacophore screen is 2.14, while it is 0.64 for the conformation from the shape screen. The calculated strain for the pharmacophore screen conformation is 5.94 Kcal/mol, while it is 0.24 Kcal/mol for the conformation from the shape screen. So an optimal alignment of shape is observed for the shape screen conformation at the cost of alignment of its corresponding pharmacophore features, while the reverse holds true for the conformation from the pharmacophore screen. As Fig. 11B (top and side views) shows, the conformation from the shape screen differs with respect to matching of corresponding pharmacophore features, though it aligns well with respect to shape. The volume overlap of the conformation from shape screen is better than that of the conformation from pharmacophore based screen. The strain values indicate that the pharmacophore screen conformation is slightly strained to obtain an optimal pharmacophore alignment. The comparison of hits from pharmacophore screen and shape screen has been made here to drive home the fact that although pharmacophore screen ensures hits that are very similar pharmacophorically, often they are higher energy conformations and hence may not end up as potent as expected when tested in biological assays. On the other hand, the shape screen ensures diversity of structure and lower energy of conformations, but compromises on

both the composition of pharmacophore features as well as their alignment to those of the query compound.

4.6. Shape similarity

Shape similarity calculation with pharmacophore feature overlap for volume scoring for all the hits showed that hits from shape search using pharmacophore feature-based volume scoring were more similar than hits from other screens. Similar conclusion can be drawn for shape similarity calculation with volume scoring involving atomtype overlap: hits from shape screen with volume scoring based on atomtype overlap were more similar than hits from other screens.

4.7. Dataset similarity

Average pair-wise similarity of compounds in each screen using atom pair similarity method was similar – around 0.30 – with the hits from the shape screening using pharmacophore feature-based volume scoring slightly diverse than those from other screens, while the hits from the pharmacophore screening are slightly more similar than hits from other screens. The average atom pair similarity with the query compound (JM-45) for hits in all

screens was also similar – around 0.20 – with hits from the pharmacophore-based screening and from shape screening using pharmacophore feature overlap for volume scoring being slightly less similar (more diverse). It was surprising to find that the hits from pharmacophore-based screening were less similar to the query compound as their average fitness was the highest when fit to the pharmacophore. This is due to the fact that although the hits had all the pharmacophore features and they do match the corresponding features of the pharmacophore model, the moieties representing the individual features vary significantly and hence the compounds appear diverse compared to the query compound.

The average dataset similarity as well as average similarity to the template compound, **JM-45**, is only slightly higher for hits from Topomer-CoMFA model, while at the same time, their fitness to the pharmacophore model is unexpectedly less, even though all the hits retain the same chemotype as the piperazinyl-aryl compounds used to build the model. A higher similarity and fitness was expected for this set. So, to check if the side-chains on the chemotype add to the diversity of the set, the chemotype of the compounds was extracted by fragmenting the compounds following the rules proposed by Murcko & Bemis [61] and the resulting chemotypes were compared. This chemotype extraction was done for all datasets of hits. Table 3 clearly shows that the average chemotype similarity for the Topomer-CoMFA set is much higher than the same for others, indicating that the side-chains appear to add significantly to the diversity and the compounds of the dataset are not much diverse in terms of their chemotype. The same holds for the average similarity of the chemotypes of hits from the Topomer-CoMFA screen to the chemotype of the template compound. Indeed when the dataset was analyzed, it was found that all the compounds had the same piperidinyl-imidazole core and diversity was observed only in moieties representing R2.

4.8. Distribution of predicted activities

All the hits from all the screens were aligned to the pharmacophore model and their activity was predicted using the Phase 3D-QSAR model built earlier [28]. Distribution of predicted activities of hits from each screen shows that more hits from the Topomer-CoMFA screen are predicted to be active than hits from other screens. In fact, 55 compounds from this screen are predicted to be less than 100 nM in potency, while among hits from all other screens, only one hit from the shape screen with pharmacophore feature-based volume scoring is predicted to be less than 100 nM. The distribution of hits from pharmacophore and shape screens is similar with slightly more number of hits showing predicted sub-micromolar potency from the shape screen with MacroModel atomtype-based volume scoring. It is interesting that compounds from the Topomer-CoMFA screen are predicted to be more potent in terms of potency value and number of compounds compared to compounds from the pharmacophore screen. It appears to be due to the fact that while the pharmacophore screen collects the hits that match just the corresponding pharmacophore features of the pharmacophore model (measured as fitness), the Topomer-CoMFA screen collects the compounds that are potentially more potent, predicted according to the Topomer-CoMFA model. In the context of the Phase 3D-QSAR model, the hits from the Topomer-CoMFA screen have many side-chains or larger moieties apart from and attached to the moieties representing the pharmacophore features and since many of these project into QSAR fields (contours) that are favorable for potency, they are predicted to be more active. In contrast, the hits from the pharmacophore screen have minimal moieties representing and aligning well with the pharmacophore features of the pharmacophore model. The Fig. 12 below shows a couple of the large sized hits – large in terms of molecular volume – one each from

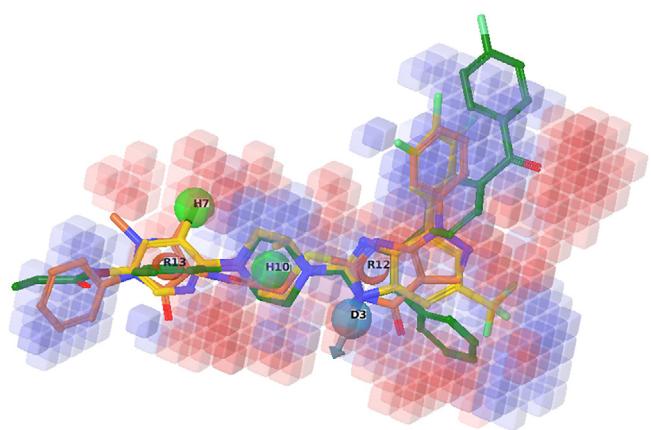


Fig. 12. Large sized hits, one each from pharmacophore screen (brown-colored) and Topomer-CoMFA screen (green-colored) compared to **JM-45**, the query compound for shape screening (yellow-colored). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Topomer-CoMFA and pharmacophore screens compared to the pharmacophore model and **JM-45**. A compound from the Topomer-CoMFA screen (**ZINC02813600 + ZINC00814718**) with a molecular surface area of 535.9 Å² and a hit from the pharmacophore screen (**ZINC02625228**) with a molecular surface area of 428.2 Å² compares to **JM-45** that has a molecular surface area of 430.5 Å². Clearly, the hit from the Topomer-CoMFA screen is larger in size with a larger surface area that can translate into more interactions and hence better potency if projected into an appropriate area around the pharmacophore model. It can be seen from the figure that the Topomer-CoMFA hit has a *para*-chloro-benzophenone moiety (from the 4th position of imidazole) that extends beyond the 3,4,5-trifluoro-phenyl moiety of **JM-45**, a phenyl moiety (from the 5th position of imidazole) that projects into regions that are favorable for hydrophobic moieties for a possible increase of potency and this is confirmed by the predicted potency by the Topomer-CoMFA model (predicted pIC50: 8.05; potency predicted by Phase model: 7.4). In contrast, the hit from the pharmacophore screen, having a similar surface area as **JM-45** of the dataset, has only the terminal phenyl ring extending beyond the pyridine ring of **JM-45**, resulting in a predicted potency of 6.87, the highest among the top-ranked 225 compounds of the pharmacophore screen. This compound does not have the fluoro substitutions on the phenyl, as in **JM-45** (4th position of benzimidazole) or the CF₃ group as in **JM-45** (6th position of benzimidazole) and hence is predicted to be less potent than **JM-45**. It is to be noted that neither the Phase 3D-QSAR model nor the Topomer-CoMFA model predicts any compound having potency greater than the most potent compound of the dataset.

4.9. Distribution of molecular/physico-chemical properties

The hits from Topomer-CoMFA have a higher average molecular weight as well as average AlogP than hits from other screens. Of significance is the observation that the pharmacophore screen hits have the lowest average AlogP and the highest average PSA than hits of other screens. The hits from the shape screens have property averages between those of the Topomer-CoMFA screen and the pharmacophore screen. So, the hits from the pharmacophore screen have properties nearer to those typical of lead compounds and hence can be considered as seed compounds or starting points. The hits from the shape screen are more like the dataset compounds and can be considered for scaffold-hopping by extracting the core

Table 4

Salient observations for each screen, highlighting the advantages and disadvantages.

Query mode	Disadvantages	Advantages
Pharmacophore model	Not much diversity in linker connecting R1 & R2; R2 less diverse	R1 rings are diverse
Shape/MacroModel atomtype	Very few compounds have a hydrogen-bond donor feature; R1 less diverse	R2 rings are diverse compared to the same in other screens
Shape/pharmacophore feature	Not much diversity in linker connecting R1 & R2; most of the compounds do not match all five pharmacophore features	A few R1 rings may be useful
Topomer-CoMFA	Diversity among fragments but not in core, due to fragmentation pattern and R-group search.	Best useful for SAR expansion. Activity can be predicted using the model

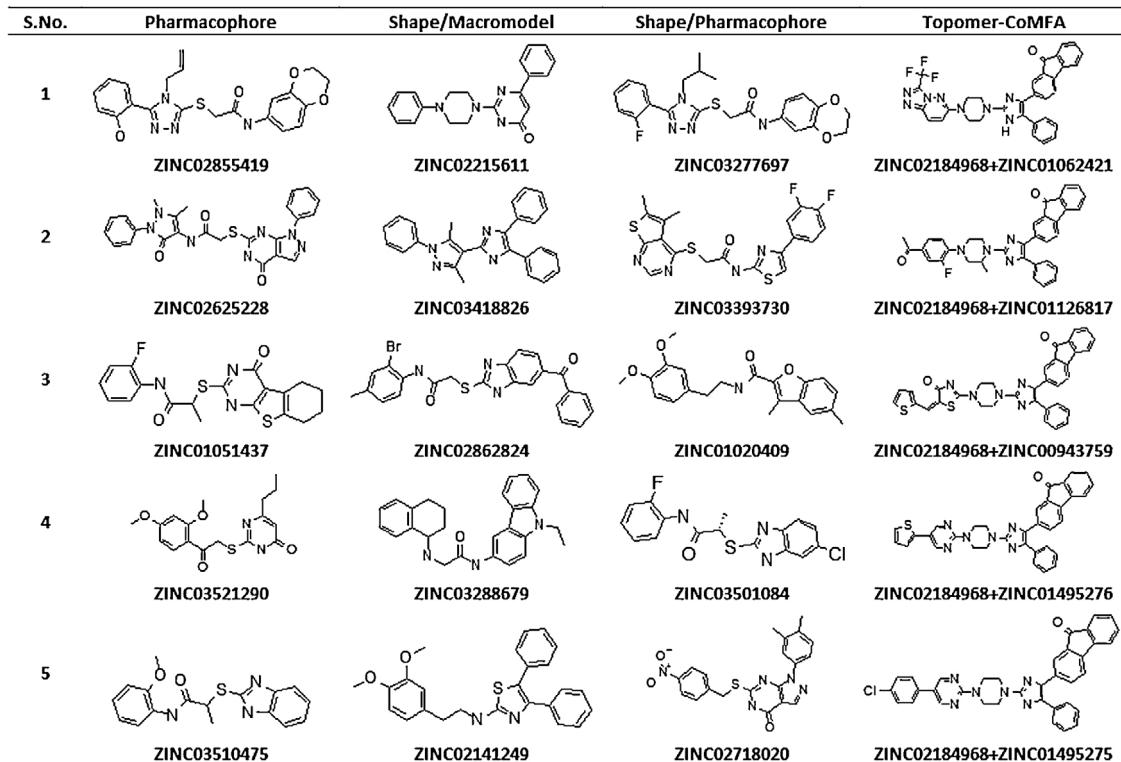


Fig. 13. Five compounds from each screen. These compounds have been picked after clustering the hits from each screen using k -means method and ranking the clusters according to their size. The five compounds from each screen are cluster centers from the top five largest clusters.

scaffolds from the hits compounds. The hits from the Topomer-CoMFA screen can be considered as extensions to the SAR of the dataset as all of them retain most of the core scaffold.

4.10. Distribution of strain values

Strain or internal energy of ligands plays an important role when binding to the receptor, with less strained conformations closer to the putative bioactive conformation [62]. Hence, consideration of strain is an important factor in the selection of hits. Strain of the hits was calculated using the strain rescore script of Schrödinger/Maestro. The distribution of strain values of the hit compounds of the various screens throws up interesting points. The average strain of hits of the pharmacophore screen, the shape screen using the pharmacophore features-based volume scoring and the Topomer-CoMFA screen is similar, with that of the Topomer-CoMFA screen slightly higher at 2.77 Kcal/mol and that of the shape screen slightly lower at 2.15 Kcal/mol. Although the hits from shape screen using pharmacophore features based volume scoring show the lowest average strain, they also show the highest strain for any compound in all the screens. The average strain of hits of the shape screen using the MacroModel atomtypes-

based volume scoring at 3.74 Kcal/mol is higher than all other screens. Though the average strain of the hits of the pharmacophore screen is a moderate 2.37 Kcal/mol, the lowest strain value recorded among the top-ranked 225 compounds is 0.017 Kcal/mol, while that for all other screens is much lower. Another interesting observation is that the highest strain value for any compound of each screen is lower for the pharmacophore screen, while that for the shape screen it is much higher. Taken together, all these observations indicate a difference in variance of the strain values across each screen. When calculated, the variance in strain values for each screen does confirm this: 2.37 for the pharmacophore screen (SD is 1.54), 16.17 for the shape screen with MacroModel atomtypes-based volume scoring (SD is 4.02), 11.14 for the shape screen with pharmacophore features-based volume scoring (SD is 3.34) and 4.92 for the topomer-CoMFA screen (SD is 2.22). All these values indicate that the pharmacophore screen and topomer-CoMFA screen give more compounds that are better in terms of ligand strain, while the shape screens tend to give many highly strained compounds that may need to be filtered out.

Analysis of the top-ranked 225 hits from each screen lead to some interesting observations and overall the results indicate moderate diversity. The salient observations for each screen are

captured in the Table 4 below. Although we analyzed only the top-ranked 225 hits, we believe that most of the top-ranking hits below these might also show the same trends that were observed. The last column of Table 3 shows some of the properties for all the compounds (36000 compounds) of the pharmacophore screen for comparison. In a recent work by Zhu et al. [63], clustering or scaffold decomposition of the compounds from virtual screens was proposed to maintain structural diversity. We did observe improvement in the diversity of compounds when the hits were clustered and representative compounds were picked from each cluster, but it was limited to some parts of the chemotype. Visual inspection of hit structures appears to be necessary for final selection of compounds for biological testing. Fig. 13 shows five diverse compounds picked from each screen after clustering and picking the cluster centers from the top-ranked clusters, ranked by size of clusters. These compounds can potentially serve as starting points for designing antagonists for TRPV1.

It is necessary to do a comprehensive analysis of hits of virtual screens to eliminate compounds based on rational bases and for the right reasons rather than take a few top-ranked (top scoring or top in terms of predicted potency) diverse compounds. Consideration of strain also has to be incorporated in selecting diverse compounds to maximize actual hit rate. Although methods such as clustering can be employed to maximize diversity while selecting compounds for biological testing, employing multiple ways of virtual screening is also very important. A better workflow can be built by incorporating multiple screens each with a different technique along with diversity sampling using methods like clustering.

A cryo-Electron Microscopy based crystal structure of rat TRPV1 was released giving atomic details of the channel and some details of binding of RTX and capsaicin [64]. This structure is expected to spawn many more studies of this target including studies that involve using structure-based virtual screening protocols to supplement this work. We plan to utilize this homology model to validate the binding site of the piperazinyl-aryl compounds that were used in this work, decipher more details of the SAR of this series as well as undertake a structure-based virtual screening to arm the research on this target with more antagonist starting points. Similar to the ligand-based virtual screening that we performed in this work, the structure-based virtual screening using the homology model would greatly aid in finding novel chemotypes that move away from the amides and ureas that are present in most of the known antagonists.

5. Conclusions

We have developed and validated predictive 3D-QSAR models for a collection of TRPV1 receptor antagonists using the methodologies of CoMFA and Topomer-CoMFA with the aim of finding a set of diverse compounds that can potentially lead to modality-specific TRPV1 antagonists. The Topomer-CoMFA models allow for potential screening of databases for alternative fragments which can be used to replace either piperazine aryl moiety (R2 fragment) or the disubstituted imidazolyl moiety (R1 fragment), using the R-group virtual screening protocol. One of the Topomer-CoMFA models, together with a Schrodinger/Phase pharmacophore model built by us earlier, as well as the most potent compound of the dataset, was used to screen the ZINC database. An in-depth analysis of the results of each screen leads to some interesting observations that enabled us to refine the searches and find a set of diverse hits that can potentially serve as multiple starting points towards modality-specific TRPV1 antagonists. These being hits from a virtual screening effort, it cannot be understated that these need to be validated by additional biological testing for potency confirmation, type of modulation (agonism/antagonism) and whether these

are suitable as starting points for expansion as series demonstrating a tractable SAR. These hits can be diverse starting points and potentially aid in validating TRPV1 in disease indications other than pain as well as aid in finding antagonists without side effect of hyperthermia.

Appendix A. Supplementary data

Electronic Supplementary Content

Tables showing the status of TRPV1 antagonists in the Clinical-Trials.gov website (as of 26-April-2016), table of list of references of topomer-CoMFA studies reported in literature, table with compound structures, activity information and details of dataset – training, internal test external test and not included – membership, tables of datasets used in the experiments showing compounds in training and test sets and table of outliers in each dataset.

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2017.01.010>.

References

- [1] M.J. Caterina, M.A. Schumacher, M. Tominaga, T.A. Rosen, J.D. Levine, D. Julius, The capsaicin receptor: a heat-activated ion channel in the pain pathway, *Nature* 389 (6653) (1997) 816–824.
- [2] M.J. Caterina, T.A. Rosen, M. Tominaga, A.J. Brake, D. Julius, Capsaicin-receptor homologue with a high threshold for noxious heat, *Nature* 398 (6726) (1999) 436–441.
- [3] M. Tominaga, M.J. Caterina, A.B. Malmberg, T.A. Rosen, H. Gilbert, K. Skinner, B.E. Raumann, A.I. Basbaum, D. Julius, The cloned capsaicin receptor integrates multiple pain-producing stimuli, *Neuron* 21 (3) (1998) 531–543.
- [4] M.M. Moran, M.A. McAlexander, T. Bíró, A. Szallasi, Transient receptor potential channels as therapeutic targets, *Nat. Rev. Drug Discov.* 10 (8) (2011) 601–620.
- [5] A. Szallasi, P.M. Blumberg, Vanilloid (Capsaicin) receptors and mechanisms, *Pharmacol. Rev.* 51 (2) (1999) 159–212.
- [6] L.A. Roberts, M. Connor, TRPV1 antagonists as a potential treatment for hyperalgesia, *Recent Pat. CNS Drug Discov.* 1 (1) (2006) 65–76.
- [7] A. Szallasi, D.N. Cortright, C.A. Blum, S.R. Eid, The vanilloid receptor TRPV1: 10 years from channel cloning to antagonist proof-of-concept, *Nat. Rev. Drug Discov.* 6 (5) (2007) 357–372.
- [8] J. Lázár, L. Gharat, N.K. Joshi, P.M. Blumberg, A. Szallasi, Screening TRPV1 antagonists for the treatment of pain: lessons learned over a decade, *Expert Opin. Drug Discov.* 4 (2) (2009) 159–180.
- [9] C.C. Correll, A. Palani, Advances in the development of TRPV1 antagonists, *Expert Opin. Ther.* 16 (6) (2006) 783–795.
- [10] D.J. Kyle, L. Tafesse, TRPV1 antagonists: a survey of the patent literature, *Expert Opin. Ther.* 16 (7) (2006) 977–996.
- [11] Y. Lee, S. Hong, M. Cui, P.K. Sharma, J. Lee, S. Choi, Transient receptor potential vanilloid type 1 antagonists: a patent review (2011–2014), *Expert Opin. Ther. Pat.* 25 (3) (2015) 291–318.
- [12] E.A. Voight, M.E. Kort, Transient receptor potential vanilloid-1 antagonists: a survey of recent patent literature, *Expert Opin. Ther. Pat.* 20 (9) (2010) 1107–1122.
- [13] S.G. Lehto, R. Tamir, H. Deng, L. Klionsky, R. Kuang, A. Le, D. Lee, J.C. Louis, E. Magal, B.H. Manning, J. Rubino, S. Surapaneni, N. Tamayo, T. Wang, J. Wang, J. Wang, W. Wang, B. Youngblood, M. Zhang, D. Zhu, M.H. Norman, N.R. Gavva, Antihyperalgesic effects of (R,E)-N-(2-hydroxy-2,3-dihydro-1H-inden-4-yl)-3-(2-piperidin-1-yl)-4-(trifluoromethyl)phenyl)-acrylamide (AMG8562), a novel transient receptor potential vanilloid type 1 modulator that does not cause hyperthermia in rats, *J. Pharmacol. Exp. Ther.* 326 (1) (2008) 218–229.
- [14] G.Y. Wong, N.R. Gavva, Therapeutic potential of vanilloid receptor TRPV1 agonists and antagonists as analgesics: recent advances and setbacks, *Brain Res. Rev.* 60 (1) (2009) 267–277.
- [15] N.R. Gavva, A.W. Bannon, S. Surapaneni, D.N. Hovland Jr., S.G. Lehto, A. Gore, T. Juan, H. Deng, B. Han, L. Klionsky, R. Kuang, A. Le, R. Tamir, J. Wang, B. Youngblood, D. Zhu, M.H. Norman, E. Magal, J.J.S. Treanor, J.C. Louis, The vanilloid receptor TRPV1 is tonically activated in vivo and involved in body temperature regulation, *J. Neurosci.* 27 (13) (2007) 3366–3374.
- [16] T. Watabiki, T. Kiso, T. Kuramochi, K. Yonezawa, N. Tsuji, A. Kohara, S. Kakimoto, T. Aoki, N. Matsuoka, Amelioration of neuropathic pain by novel transient receptor potential vanilloid 1 antagonist AS1928370 in rats without hyperthermic effect, *J. Pharmacol. Exp. Ther.* 336 (3) (2011) 743–750.
- [17] P. Holzer, The pharmacological challenge to tame the transient receptor potential vanilloid-1 (TRPV1) nocisensor, *Br. J. Pharmacol.* 155 (8) (2008) 1145–1162.
- [18] R.M. Reilly, H.A. McDonald, P.S. Puttfarcken, S.K. Joshi, L. Lewis, M. Pai, P.H. Franklin, J.A. Segreti, T.R. Neelands, P. Han, J. Chen, P.W. Mantyh, J.R. Ghilardi, T.M. Turner, E.A. Voight, J.F. Daanen, R.G. Schmidt, A. Gomtsyan, M.E. Kort, C.R.

- Faltynek, P.R. Kym, Pharmacology of modality-specific transient receptor potential vanilloid-1 antagonists that do not alter body temperature, *J. Pharmacol. Exp. Ther.* 342 (2) (2012) 416–428.
- [19] A. Garami, Y.P. Shimansky, E. Pakai, D.L. Oliveira, N.R. Gavva, A.A. Romanovsky, Contributions of different modes of TRPV1 activation to TRPV1 antagonist-induced hyperthermia, *J. Neurosci.* 30 (4) (2010) 1435–1440.
- [20] A.A. Romanovsky, M.C. Almeida, A. Garami, A.A. Steiner, M.H. Norman, S.F. Morrison, K. Nakamura, J.J. Burmeister, T.B. Nucci, The transient receptor potential vanilloid-1 channel in thermoregulation: a thermosensor it is not, *Pharmacol. Rev.* 61 (3) (2009) 228–261.
- [21] A. Fernández-Carvajal, G. Fernández-Ballester, I. Devesa, J.M. González-Ros, A. Ferrer-Montiel, New strategies to develop novel pain therapies: addressing thermoreceptors from different points of view, *Pharmaceuticals* 5 (1) (2012) 16–48.
- [22] K. Kaszas, J.M. Keller, C. Coddou, S.K. Mishra, M.A. Hoon, S. Stojilkovic, K.A. Jacobson, M.J. Iadarola, Small molecule positive allosteric modulation of TRPV1 activation by vanilloids and acidic pH, *J. Pharmacol. Exp. Ther.* 340 (1) (2012) 152–160.
- [23] M. Papakosta, C. Dalle, A. Haythornthwaite, L. Cao, E.B. Stevens, G. Burgess, R. Russell, P.J. Cox, S.C. Phillips, C. Grimm, The chimeric approach reveals that differences in the TRPV1 pore domain determine species-specific sensitivity to block of heat activation, *J. Biol. Chem.* 286 (45) (2011) 39663–39672.
- [24] B. Nilius, A. Szallasi, Transient receptor potential channels as drug targets: from the science of basic research to the art of medicine, *Pharmacol. Rev.* 66 (3) (2014) 768–814.
- [25] M. Cui, V. Gosu, S. Basith, S. Hong, S. Choi, Polymodal transient receptor potential vanilloid type 1 nocisensor: structure, modulators, and therapeutic applications, *Adv. Protein Chem. Struct. Biol.* 104 (2016) 81–125.
- [26] M.J. Gunthorpe, A. Szallasi, Peripheral TRPV1 receptors as targets for drug development: new molecules and mechanisms, *Curr. Pharm. Design* 14 (1) (2008) 32–41.
- [27] K. Starowicz, L. Cristina, V. Di Marzo, TRPV1 receptors in the central nervous system: potential for previously unforeseen therapeutic applications, *Curr. Pharm. Design* 14 (1) (2008) 42–54.
- [28] R. Kristam, V. Parmar, V.N. Viswanadhan, 3D-QSAR analysis of TRPV1 inhibitors reveals a pharmacophore applicable to diverse scaffolds and clinical candidates, *J. Mol. Graph. Model.* 45 (2013) 157–172.
- [29] V.I. Ognyanov, C. Balan, A.W. Bannon, Y. Bo, C. Dominguez, C. Fotsch, V.K. Gore, L. Klionsky, V.V. Ma, Y.X. Qian, R. Tamir, X. Wang, N. Xi, S. Xu, D. Zhu, N.R. Gavva, J.J. Treanor, M.H. Norman, Design of potent, orally available antagonists of the transient receptor potential vanilloid 1 Structure-activity relationships of 2-piperazin-1-yl-1H-benzimidazoles, *J. Med. Chem.* 49 (12) (2006) 3719–3742.
- [30] V.K. Gore, V.V. Ma, R. Tamir, N.R. Gavva, J.J. Treanor, M.H. Norman, Structure–activity relationship (SAR) investigations of substituted imidazole analogs as TRPV1 antagonists, *Bioorg. Med. Chem. Lett.* 17 (21) (2007) 5825–5830.
- [31] R.D. Cramer, Topomer-CoMFA: a design methodology for rapid lead optimization, *J. Med. Chem.* 46 (3) (2003) 374–388.
- [32] R.D. Cramer, P. Cruz, G. Stahl, W.C. Curtiss, B. Campbell, B.B. Masek, F. Soltanshahi, Virtual screening for R-groups, including predicted pIC₅₀ contributions, within large structural databases, using Topomer-CoMFA, *J. Chem. Inf. Model.* 48 (11) (2008) 2180–2195.
- [33] J.J. Irwin, B.K. Shoichet, ZINC – a free database of commercially available compounds for virtual screening, *J. Chem. Inf. Model.* 45 (1) (2005) 177–182.
- [34] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, ZINC: a free tool to discover chemistry for biology, *J. Chem. Inf. Model.* 52 (7) (2012) 1757–1768.
- [35] V.N. Viswanadhan, Y. Sun, M.H. Norman, Three-dimensional quantitative structure–activity relationships and activity predictions of human TRPV1 channel antagonists: comparative molecular field analysis and comparative molecular similarity index analysis of cinnamides, *J. Med. Chem.* 50 (23) (2007) 5608–5619.
- [36] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.* 110 (18) (1988) 5959–5967.
- [37] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *J. Med. Chem.* 37 (24) (1994) 4130–4146.
- [38] K.H. Kim, 3-D-QSAR analysis of N-(3-acyloxy-2-benzylpropyl)-N'-dihydroxytetrahydrobenzazepine and tetrahydroisoquinoline and N-(3-acyloxy-2-benzylpropyl)-N'-(4-hydroxy-3-methoxybenzyl) thioureas analogues as potent vanilloid receptor ligands, *Bioorg. Med. Chem.* 10 (5) (2002) 1367–1372.
- [39] W.S. Cheung, R.R. Calvo, B.A. Toung, S.P. Zhang, D.R. Stone, M.R. Brandt, T. Hutchinson, C.M. Flores, M.R. Player, Discovery of piperidine carboxamide TRPV1 antagonists, *Bioorg. Med. Chem. Lett.* 18 (16) (2008) 4569–4572.
- [40] T. Taskin, F. Sevin, QSAR approach to correlate TRPV1 antagonist activity for a series of heteroaromatic urea, *QSAR Comb. Sci.* 28 (10) (2009) 1098–1111.
- [41] P.R. Kym, M.E. Kort, C.W. Hutchins, Analgesic potential of TRPV1 antagonists, *Biochem. Pharmacol.* 78 (3) (2009) 211–216.
- [42] D. Goldmann, P. Pakfeifer, S. Hering, G.F. Ecker, Novel scaffolds for modulation of TRPV1 identified with pharmacophore modeling and virtual screening, *Future Med. Chem.* 7 (3) (2015) 243–256.
- [43] Z. Feng, L.V. Pearce, X. Xu, X. Yang, P. Yang, P.M. Blumberg, X.Q. Xie, Structural insight into tetrameric hTRPV1 from homology modeling, molecular docking, molecular dynamics simulation, virtual screening, and bioassay validations, *J. Chem. Inf. Model.* 55 (3) (2015) 572–588.
- [44] SYBYL-X (version 1.3), Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144 USA, 2011.
- [45] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94, *J. Comput. Chem.* 17 (5–6) (1996) 490–519.
- [46] T.A. Halgren, Merck molecular force field. II. MMFF94 van der waals and electrostatic parameters for intermolecular interaction, *J. Comput. Chem.* 17 (5–6) (1996) 520–552.
- [47] T.A. Halgren, Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94, *J. Comput. Chem.* 17 (5–6) (1996) 553–586.
- [48] T.A. Halgren, R.B. Nachbar, Merck molecular force field. IV. Conformational energies and geometries for MMFF94, *J. Comput. Chem.* 17 (5–6) (1996) 587–615.
- [49] T.A. Halgren, Merck Molecular Force Field. V. Extension of MMFF94 using experimental data, additional computational data and empirical rules, *J. Comput. Chem.* 17 (5–6) (1996) 616–641.
- [50] B.L. Bush, C.I. Bayly, T.A. Halgren, Consensus bond-charge increments fitted to electrostatic potential or field of many compounds: application to MMFF94 training set, *J. Comput. Chem.* 20 (14) (1999) 1495–1516.
- [51] A.N. Jain, Ligand-based structural hypotheses for virtual screening, *J. Med. Chem.* 47 (4) (2004) 947–961.
- [52] R.D. Clark, OptiSim: an extended dissimilarity selection method for finding diverse representative subsets, *J. Chem. Inf. Comput. Sci.* 37 (6) (1997) 1181–1188.
- [53] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.* 110 (18) (1988) 5959–5967.
- [54] Z. Winter, A. Buhala, F. Ötvös, K. Jósavay, C. Vizler, G. Dombi, G. Szakonyi, Z. Oláh, Functionally important amino acid residues in the transient receptor potential vanilloid 1 (TRPV1) ion channel – an overview of the current mutational data, *Mol. Pain* 9 (2013) 30.
- [55] N.R. Gavva, L. Klionsky, Y. Qu, L. Shi, R. Tamir, S. Edenson, T.J. Zhang, V.N. Viswanadhan, A. Toth, L.V. Pearce, T.W. Vanderah, F. Porreca, P.M. Blumberg, J. Lile, Y. Sun, K. Wild, J.C. Louis, J.J. Treanor, Molecular determinants of vanilloid sensitivity in TRPV1, *J. Biol. Chem.* 279 (19) (2004) 20283–20295.
- [56] S.E. Jordt, D. Julius, Molecular basis for species-specific sensitivity to hot chili peppers, *Cell* 108 (3) (2002) 421–430.
- [57] S.E. Jordt, M. Tominaga, D. Julius, Acid potentiation of the capsaicin receptor determined by a key extracellular site, *Proc. Natl. Acad. Sci. U. S. A.* 97 (14) (2000) 8134–8139.
- [58] Phase, version 4.2, Schrodinger LLC, New York, NY, 2015.
- [59] J. Ghosh, A. Liu, K-Means, in: X. Wu, V. Kumar (Eds.), *The Top Ten Algorithms in Data Mining (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)*, Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL, 2009.
- [60] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods, *J. Phys. Chem. A* 102 (1998) 3762–3772.
- [61] G.W. Bemis, M.A. Murcko, The properties of known drugs. 1. Molecular frameworks, *J. Med. Chem.* 39 (15) (1996) 2887–2893.
- [62] E. Perola, P.S. Charifson, Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding, *J. Med. Chem.* 47 (10) (2004) 2499–2510.
- [63] T. Zhu, S. Cao, P.C. Su, R. Patel, D. Shah, H.B. Chokshi, R. Szukala, M.E. Johnson, K.E. Hevener, Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis, *J. Med. Chem.* 56 (17) (2013) 6560–6572.
- [64] M. Liao, E. Cao, D. Julius, Y. Cheng, Structure of the TRPV1 ion channel determined by electron cryo-microscopy, *Nature* 504 (7478) (2013) 107–112.