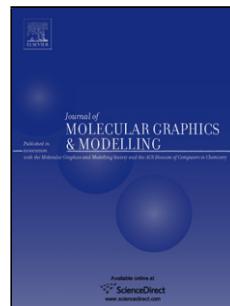


# Accepted Manuscript

Title: Constructing an atomic-resolution model of human P2X<sub>7</sub> receptor followed by pharmacophore modeling to identify potential inhibitors

Author: Mehdi Ahmadi Amin Nowroozi Mohsen Shahlaei



PII: S1093-3263(15)30038-3

DOI: <http://dx.doi.org/doi:10.1016/j.jmgm.2015.08.005>

Reference: JMG 6587

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 3-4-2015

Revised date: 22-6-2015

Accepted date: 10-8-2015

Please cite this article as: Mehdi Ahmadi, Amin Nowroozi, Mohsen Shahlaei, Constructing an atomic-resolution model of human P2X<sub>7</sub> receptor followed by pharmacophore modeling to identify potential inhibitors, *Journal of Molecular Graphics and Modelling* <http://dx.doi.org/10.1016/j.jmgm.2015.08.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Constructing an atomic-resolution model of human P2X<sub>7</sub> receptor followed by pharmacophore modeling to identify potential inhibitors**

**Mehdi Ahmadi<sup>a</sup>, Amin Nowroozi<sup>b</sup>, Mohsen Shahlaei<sup>c\*</sup>**

<sup>a</sup>*Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran;* <sup>b</sup>*Pharmaceutical sciences Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran;* <sup>c</sup>*Medical Biology Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran*

**\*Corresponding author.** Mohsen Shahlaei, PhD

Department of Medicinal Chemistry, Faculty of Pharmacy, Kermanshah University of Medical Sciences PO: 67346-67149, Kermanshah, Iran. Tel: +98-831-4276489; Fax: +98-831-4276493; Email: mohsenshahlaei@yahoo.com AND mshahlaei@kums.ac.ir

**Highlights**

□ A reliable 3D model for P2X<sub>7</sub> receptor is developed by using homology modelling and MD simulation. ▶ A pharmacophore model was constructed to discover small-molecule inhibitors of P2X<sub>7</sub> receptor ▶ Two different databases were screened by means of validated pharmacophore model. ▶ Docking analysis shows important residues in binding of lead compounds.

Graphical abstract **Homology modeling and molecular dynamics simulation of human P2X<sub>7</sub> receptor followed by pharmacophore modeling and virtual screening to identify potential inhibitors**

**Mehdi Ahmadi<sup>a</sup>, Amin Nowroozi<sup>b</sup>, Mohsen Shahlaei<sup>c\*</sup>**

<sup>a</sup>Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran; <sup>b</sup>Novel Drug Delivery Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran; <sup>c</sup>Medical Biology Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran

A computational pipeline is used to construct a reliable model of P2X<sub>7</sub> receptor and then a ligand-based quantitative pharmacophore modeling methodology was used to introduce some hit compounds.

### Abstract

The P2X purinoceptor 7 (P2X<sub>7</sub>R) is a trimeric ATP-activated ion channel gated by extracellular ATP. P2X<sub>7</sub>R has important role in numerous diseases including pain, neurodegeneration, and inflammatory diseases such as rheumatoid arthritis and osteoarthritis. In this prospective, the discovery of small-molecule inhibitors for P2X<sub>7</sub>R as a novel therapeutic target has received considerable attention in recent years. At first, 3D structure of P2X<sub>7</sub>R was built by using homology modeling (HM) and a 50 ns molecular dynamics simulation (MDS).

Ligand-based quantitative pharmacophore modeling methodology of P2X<sub>7</sub>R antagonists were developed based on training set of 49 compounds. The best four-feature pharmacophore model, includes two hydrophobic aromatic, one hydrophobic and one aromatic ring features, has the highest correlation coefficient (0.874), cost difference (368.677), low *RMSD* (2.876), as well as it shows a high goodness of fit and enrichment factor.

Consequently, some hit compounds were introduced as final candidates by employing virtual screening and molecular docking procedure simultaneously. Among these compounds, six potential molecule were identified as potential virtual leads which, as such or upon further optimization, can be used to design novel P2X<sub>7</sub>R inhibitors.

**Keywords:** P2X<sub>7</sub> receptor antagonists; Pharmacophore model; Virtual screening; Homology modeling; Molecular dynamics simulation; Molecular docking

## Introduction

There are two major families of purine and pyrimidine receptors: the P1 adenosine receptors (A1, A2a, A2b, and A3), and the P2 family, which is divided into the ionotropic (P2X) and metabotropic receptors (P2Y). The ionotropic P2X receptor subfamily comprises seven subtypes of ATP-gated ion channels, called P2X<sub>1-7</sub> receptors [1-3]. P2X receptors are widely distributed in mammalian tissues [4]. For all P2X receptor subtypes, the preferred agonist is ATP [5]. The discovery of potent and receptor-selective P2X<sub>7</sub>R antagonists has significantly expanded the field of P2X receptor pharmacology [6]. Because of the importance of P2X<sub>7</sub>R and its antagonists, there is a lot of attention paid to design new antagonists with high potency [7]. Over the past few years there have been many interesting developments in the study of P2X<sub>7</sub>R antagonists, including the identification of diverse new chemical series, increased understanding of receptor pharmacology, and reports of in vivo studies both in preclinical animal models of disease and, critically, in early clinical trials [7]. With regard to P2X receptors, there is no medicine (agonists or antagonists) in clinical use. However, pre-clinical studies indicate that P2X<sub>7</sub>R antagonists might be used to treat inflammatory or partially inflammatory diseases in any tissue, such as chronic pulmonary conditions, glomerulonephritis, rheumatoid arthritis, inflammatory bowel diseases, stroke, brain trauma, amyotrophic lateral sclerosis, multiple sclerosis, and pain disorders [3]. Therefore, P2X<sub>7</sub>R has become a pharmacological target for several pharmaceutical companies, with several clinical trials in phase I and II to evaluate the efficacy of P2X<sub>7</sub>R antagonists on many diseases [3]. Although published data are lacking, a P2X<sub>7</sub>R antagonist from Astra-Zeneca, AZ9056 (structure not disclosed), is reportedly under clinical development for the treatment of rheumatoid arthritis and inflammatory bowel disease [6]. There are other reports for starting

or completing clinical trials of P2X<sub>7</sub>R antagonist [7]. It is evident that high-throughput screening has provided a rich source of leads for this target [7]. The P2X<sub>7</sub>R, first cloned from rat brain and soon after from human monocytes, is the most disparate of the P2X subtypes, in terms of both structure and function [8,9]. Structurally, it differs from the other P2X subtypes in having a long intracellular C-terminal chain. Functionally, upon brief stimulation of the P2X<sub>7</sub>R, a nonselective cation channel is opened [7]. The minimum stoichiometric conformation of the P2X<sub>7</sub>R channel appears to be a trimeric channel with three binding sites for ATP [10,11]. It can be mentioned that the P2X<sub>7</sub>R has a 595 amino acid sequence with an average of 40% homology to other members of purinergic P2X receptors [9]. However, at present, this approach is hampered by the lack of a good P2X<sub>7</sub>R model [7,12].

Some of the major health challenges, such as many types of cancers, inflammatory and infectious diseases, and diabetes and neurodegenerative diseases are in desperate need of innovative medicines. Despite of this challenge, fast and affordable drug development is a vision that contrasts sharply with the current state of drug discovery. It takes an average of 12 to 15 years and (depending on the therapeutic area) as much as 1 billion USD to bring a single drug into market for using by patients [13].

Pharmacoinformatics approaches including structure activity relationship (SAR), pharmacophore, virtual screening and molecular docking are widely used in the field of drug discovery as it saves time, investment and animal sacrifice [14].

Pharmacophore approaches have been used extensively in virtual screening, Denovo design and other applications such as lead optimization and multitarget drug design [15]. The advantage of pharmacophore-based virtual screening in the drug discovery process is that most of the compounds with low probability to be active can be excluded from further studies already in a very early stage of a project. Thereby, a lot of resources in the further drug

discovery process, especially in in vitro experiments, can be saved [16]. The concept of pharmacophore was first introduced in 1909 by Paul Ehrlich [17], who defined the pharmacophore as ‘a molecular framework that carries (phoros) the essential features responsible for a drug’s (pharmacon) biological activity’. After a century’s development, the basic pharmacophore concept still remains unchanged, but its intentional meaning and application range have been expanded considerably. According to the very recent definition by IUPAC [18], a pharmacophore model is ‘an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response’ [15].

Computer-aided drug design methods are usually divided into two categories: ligand-based and structure-based methods, which depend on the available information about the 3D structure of the macromolecule target of interest. If this is known, either from experimental (i.e., X-ray crystallography, NMR studies) or computational (building of 3D structures using methods such as HM technique) sources, structure-based pharmacophores can be developed. Otherwise, ligand-based pharmacophoric models can be generated from a set of molecules known to act with a common mechanism of action [19,20]. Once generated, a pharmacophoric hypothesis represents a versatile tool for drug design. In fact, it can be used as a query to screen databases of commercially available compounds or, alternatively, to guide chemists in the synthesis of new compounds during the hit-to-lead optimization process. Moreover, it can also be used to align molecules based on the 3D arrangement of chemical features or to develop predictive 3D QSAR models [20,21].

Recently, as software and hardware advance together, much successful combination of structure- and ligand-based computational methods has been effectively employed in many drug design projects [22,23]. It has been postulated that using both techniques simultaneously on the same biological system is advantageous as it takes into account all possible

information [24,25]. The combination of structure- and ligand-based techniques can occur either in a sequential, parallel or hybrid type. It is possible to implement all of approaches in most current molecular modeling routines [22]. In the sequential method, different structure- and ligand-based methods are applied in a virtual screening routine to sequentially filter the number of hits retrieved until the number is small enough for extensive biological testing. Regularly, computationally low-cost techniques like pharmacophore screening are applied in the beginning of the multi-step screening process (prefiltering). As the number of hits decreases, computationally more expensive techniques, particularly docking and molecular dynamics simulation (MDS), can be used to further filter the retrieved ligands. In recent years, several successful uses of sequential ligand- and structure-based type's methods have been reported. In many of them, hits retrieved by screening with single or multiple pharmacophores [26-30] are further filtered using druglikeness or absorption, distribution, metabolism, excretion and toxicity (ADMET) filters and evaluated using docking into the active pocket of protein of interest. Docking technique commonly employ an energy-based scoring function to estimate the most favorable ligand position, orientation and conformation, necessary for binding in the active pocket. As seen in many scientific reports, virtual screening based on pharmacophore modeling procedure has been found to be a successful method, especially when combined with docking simulation [31,32]. It should be mentioned that in the parallel methods, several approaches are run independently and the top hits of each approach are selected for biological testing. The methods used should be complementary and can include pharmacophore models, ligand similarity methods with the application of two- and three dimensional descriptors, as well as docking. On the other hand, there is a "protein-ligand pharmacophores" approach (hybrid approach) which represents a true combination of structural and ligand information into a standalone approach. This method have been developed and used successfully. "Protein-ligand pharmacophores" approach represents a

combination of ligand- and protein-information as they are developed based on experimental structures of protein–ligand complexes. In this manner, the observed protein–ligand interactions are directly translated into pharmacophore features [22]. A comparison of different combination approaches disclosed that most reported studies apply a sequential combination of virtual screening techniques [22]. The remaining challenge in drug development is not to find hits, but to advance them into lead compounds by predicting their metabolism and adverse effects. By combining structure- and ligand-based methods, modelers are hoping to address this challenge and to improve accuracy, performance and development of molecular modeling methods [22]. If an experimental obtained structure of a biological target is not available, it may be likely to create a computational 3D structure using techniques such as homology modeling (HM) (also known as comparative modeling). HM procedure can be described as a procedure to build an atomic-resolution 3D structure of a "target" protein. This 3D structure is built using its amino acid sequence and the experimental structure of a related protein (called "template"). Employing developed structure of the biological target, hit compounds with high affinity and selectivity to the target may be designed.

In this study, an attempt was made to combine ligand- and structure-based drug design methods using the sequential approach including pharmacophore models, HM, MD simulation, docking and virtual screening of the MayBridge and CoCoCo-SC Asinex compound database. The ultimate goal is recognizing novel and future P2X<sub>7</sub>R inhibitors with specificity for their possible use in pain, neurodegeneration, and inflammatory diseases such as rheumatoid arthritis and osteoarthritis.

At first part of study an HM based model of P2X<sub>7</sub>R is constructed and then the developed model was inserted in a pre-equilibrated lipid bilayer. One of the most important uses of molecular dynamics simulations (MDS) of proteins is actually the refinement of HM

structures. The coupling of HM techniques with MDS is helpful in that it tackles the sampling deficiency of conformational space of dynamics simulations by providing good quality initial guesses for the proteins [33-35].

Subsequently, here in this research, various structural from MayBridge and CoCoCo-SC Asinex compound databases were filtered by using the drug-like ADMET properties, such as Lipinski's rule of five [36]. Then, a qualitative pharmacophore model was developed based on P2X<sub>7</sub>R antagonists that were acquired from the recently published research [37,38] and was successfully used in the further screening of the database compounds. The hit molecules were chosen based on their best fit values and then subjected to docking analyses to refine the list of retrieved hits. This study has resulted to the set of lead ligands as possible candidates for the designing of potential P2X<sub>7</sub>R antagonists.

## 2. Materials and methods

### 2.1. Template searching and homology modeling

The primary sequences of human P2X<sub>7</sub>R were obtained from Swiss-Prot database (Accession number Q99572). In the HM step, an experimentally determined structure is searched for with high sequence identity with a P2X<sub>7</sub>R. Blastp (Basic Local Alignment Search Tool) at the NCBI was used to find the homologous proteins with known structures to be employed as the template in the in the HM process [39]. The crystal structure of the P2X<sub>4</sub> receptor (PDB ID: 4WD0) at 2.90 Å resolution was selected.

MODELER 9v2 [40] was used to align the sequence and build homology models of P2X<sub>7</sub>R. From the alignment, 3D models containing all non-hydrogen atoms were obtained automatically using the method implemented in MODELER. An ensemble of 10 model structures of P2X<sub>7</sub>R was generated, among, the one corresponding to the lowest value of the

probability density function (*pdf*) and the fewest restraints violations was used for further analysis. From the generated ensemble, a single structure was selected for further modeling and as a starting structure for the MD simulations. The overall stereochemical quality of the final selected model for P2X<sub>7</sub>R was accessed by the program PROCHECK [41]. Environment profile of the final developed model was checked using verify-3D (Structure Evaluation Server) [42,43].

## 2.2. Molecular dynamics simulation

The best model from HM step was inserted at the center of the POPC bilayer with its long axis normal to the membrane–water interface. The  $\alpha$ -helical domain of the receptor was placed at the same level as the lipid bilayer. Overlapping lipid and water molecules were discarded to remove overlaps between atoms of protein and lipids.

All MD simulations were performed by the GROMACS 4.5.3 package [44] . All of the MD simulations were performed in an explicit 1-palmitoyl-2-oleoylphosphatidylcholine (POPC). A biomembrane was modeled by a lipid bilayer consisting of 408 POPC lipids (204 per leaflet). An initial configuration of a POPC bilayer was taken from Tieleman website ([www.ucalgary.ca/~tieleman/download.html](http://www.ucalgary.ca/~tieleman/download.html) ) [45]. A 25 residue segment in C-terminus of the P2X<sub>7</sub>R was ignored in the MD simulation to keep high similarity to template.

The GROMOS 53a6 force field [46] was employed for the MD simulation of P2X<sub>7</sub>R while the lipid bilayer were described using a previously developed topology file [47].

Water molecules were represented applying a simple point charge (SPC216) model. Solvent (i.e. water) ions, lipid, protein and ligand were coupled separately to a temperature bath.

Final protein-lipid membrane system was placed in a box with the dimensions of 99×94 × 140 (all in Å) with a total of 10834 atoms.

The coupling scheme of Berendsen was employed in both of NVT and NPT ensembles. The particle mesh Ewald (PME) method interaction was used [48]. A 12 Å cutoff for long-range and the Lincs algorithm for covalent bond constraints were applied [49]. All the molecular images and animations were produced using VMD [50].

### 2.3. Dataset

An adequately set of molecules with their P2X<sub>7</sub>R inhibitory activity data is necessary in a typical process of pharmacophore model development. Bioactivities of 49 P2X<sub>7</sub>R antagonists were taken from the literature [37,38] and are presented in Table 1. The IC<sub>50</sub> value is defined as the concentration of a given molecule required to inhibit 50% of the P2X<sub>7</sub>R activity. The 2D structure of each compound was drawn by Chem Draw [51] and subsequently converted into 3D structure by Chem3D, and then followed by optimization using PM3 method. The generated 3D structure of each molecule was manually checked to ensure that the chirality of the chiral agent is correctly prepared and no structure of molecules was duplicated. The molecular structures were optimized using the Polak-Ribiere algorithm until the root mean square gradient was 0.01 kcal mol<sup>-1</sup>. Molecules were further separated into the training (29 compounds), and test (20 compounds) sets.

The selection of an appropriate training set is one of the most main steps in pharmacophore modeling procedure. This step is responsible for the quality of the generated pharmacophore hypothesis. The test set, which is not employed in model building step, but in the pharmacophore validation procedure, has equal weight. The data set was split into a training set and a prediction set using Kenard and Stone algorithm. In running of Kennard and Stone algorithm the calculated features matrix was used as input [52]. The IC<sub>50</sub> values of these 49 molecules spanned across an extensive range from 24 nM to 2752 nM. In addition, 51 molecules were retrieved from the literature and used in the validation step (Table S1). The

activity values of the data set were classified into four categories, active ( $IC_{50} \leq 50$  nM, ++++), moderately active ( $50 \leq IC_{50} \leq 200$  nM, +++) , less active ( $200 \leq IC_{50} \leq 1000$  nM, ++), and inactive ( $IC_{50} > 1000$  nM, +), to simplify the results of pharmacophore model building and validation.

<Table 1 near here>

#### **2.4 Molecular docking enabled pharmacophore modeling**

Every pharmacophore modeling study that employs the “3D QSAR Pharmacophore Generation” or “Common Feature Pharmacophore Generation” protocols of Accelrys Discovery Studio (ADS) conventionally starts with the diverse conformation generation step. The conformations of the selected molecules were generated using the “best conformer generation” method with a cutoff value of 20 kcal/mol from the local energy minimum conformation [53-55].

Pharmacophore model generation was performed using generated conformations of each training set molecule. "Feature Mapping" protocol was applied to identify the chemical features present in all training set molecules. Pharmacophore model generation was performed by choosing chemical features defined in the ADS, such as hydrogen bond acceptor, hydrogen bond donor, hydrophobic, and aromatic ring.

Each of the molecules in training set was submitted to the “Common Feature Pharmacophore Generation” procedure. The “Common Feature Pharmacophore Generation” protocol implemented in ADS was used to identify and overlay common features shared by a training set.

Hypothesis generation run develops 10 possible pharmacophore hypotheses having a different arrangement of above mentioned features and sorts them according to the ranking scores.

The Uncertainty value was changed to 1.2 from the default value of 3 as the training set molecules scarcely spanned the required range of bioactivity (i.e., four orders of magnitude) [56].

Docking as an efficient in silico tool is playing an ever increasing role in structure-based drug design [57,58]. The program AutoDock 4 was used in this docking study, in which Lamarckian genetic algorithm was used to search for the optimum binding site [59]. During docking process, a maximum of 200 conformers was considered for each studied compound. The conformer with the lowest binding free energy was used for further analysis. The 2D structure of studied ligands were constructed by ChemOffice 2004. The optimized structures by a semi empirical method (AM1), implemented in HyperChem, was used as input of Auto Dock Tools and the partial charges of atoms were calculated using Gasteiger– Marsili procedure [60]. After determining Kollman united atom charges [61], non-polar hydrogens were merged to their corresponding carbons and as a final processing step in protein preparation, desolvation parameters were assigned to each atom. Using auto Grid tool, the grid maps were constructed adequately large to include the active site of protein as well as significant regions of the surrounding surface. In all the cases, a grid map of 60 points in each Cartesian direction besides grid-point spacing of 0.375 was generated. Based on Lamarckian Genetic Algorithm [59], using the pseudo-Solis and Wets local search method [62], Auto Dock Tools was employed to produce both grid and docking parameter files. Applying 0.5 Å clustering tolerance to construct clusters of the closest compounds, the initial coordinates of the ligand were used as the reference structure.

## 2.5 Pharmacophore validation

Ordinarily, developed pharmacophore models are employed as 3D queries to search chemical databases to discover new and highly potent drug lead compounds. These developed pharmacophore models should be statistically significant and able to predict the IC<sub>50</sub>s of new molecules and retrieve active molecules from the database. The best pharmacophore model was validated employing four techniques, cost analysis, test set prediction, Fischer randomization test and Enrichment factor calculation (E). The “Common Feature Pharmacophore Generation” protocol ranks the 10 generated hypothesis pharmacophore models on their cost values. The weight, error and configuration costs are three components that build the overall cost of a hypothesis. The value of the weight cost increases in a Gaussian form as this function weighs a model's deviation from the ideal value of two. The error cost value shows the root mean square (*RMS*) difference between the observed and predicted IC<sub>50</sub>s of the training set molecules. The configuration cost denotes the complexity or the entropy of the conformational space being optimized and is constant for a given data set.

The “Common Feature Pharmacophore Generation” also calculates two additional costs for each hypothesis, fixed cost and null cost, and a cost for every hypothesis namely total cost. The fixed cost is the lowest possible cost denoting a simplest hypothetical model that fits all data completely and the null cost denotes the maximum cost of a pharmacophore with no features and estimates the biological activity to be the average activity of the molecules in training set and the total cost for every hypothesis. A larger difference between the fixed and null costs than that between the fixed and total costs signifies the quality of a pharmacophore model. All of these cost values are reported in bits and a difference of 40–60 bits between the total and null costs suggests a 75–90% chance of representing a true correlation in the data. Twenty molecules were used as the test set to validate the developed hypothesis. Fischer

randomization (Cat-Scramble) is another technique for pharmacophore model validation. The 95% confidence level was selected in this validation study and 19 random spreadsheets were built. This validation technique checks the correlation between the chemical structures and IC<sub>50</sub>. This technique produces model employing the same parameters as those employed to generate the original pharmacophore model by shuffling the IC<sub>50</sub>s of the training set molecules. The fourth way for validating the developed pharmacophore model is based on the E value, which is estimated employing a database containing active and inactive molecules.

## 2.6 Enrichment factor calculation

The GH scoring method or Güner-Henry scoring method [63,64], was employed following external test set validation to evaluate the quality of the developed pharmacophore model. The GH score has been successfully applied to quantify model selectivity precision of hits and the recall of actives from a 1676 (D) molecule dataset consisting of known actives and in-actives.

Of these molecules, 44 compounds (A) are known inhibitors of P2X<sub>7</sub>R antagonists that were selected from literature [65], while the other 1632 molecules were from the already prepared chemical dataset. The Güner-Henry scoring method consists of calculating the following parameters: the percent yield of actives in a database (%Y, recall), the percent ratio of actives in the hit list (%A, precision), and the enrichment factor E, and the GH score. The following formula was used to calculate the different parameters,

$$E = (Ha \times D) / (Ht \times A).$$

The GH score ranges from 0 to 1, where a value of 1 signifies the ideal model.

$$\%A = \frac{Ha}{A} \times 100$$

$$\%Y = \frac{Ha}{Ht} \times 100$$

$$E = \frac{Ha/Ht}{A/D}$$

$$GH = \left( \frac{Ha(3A + Ht)}{4HtA} \right) \left( 1 - \frac{Ht - Ha}{D - A} \right)$$

## 2.7. Virtual screening

Virtual screening, an in silico technique for drug design and discovery, has been extensively employed for lead identification in drug discovery process. Virtual screening techniques are generally divided into ligand-based screening and structure-based virtual screening.

As mentioned above, we used a sequential approach which combines structure- and ligand-based methods. A well-validated pharmacophore model includes the chemical features responsible for bioactivities of drug candidates; consequently, it can be applied to carry out a database search.

The best pharmacophore model developed was employed as a 3D query in database searching. This virtual screening was carried out to find novel and diverse virtual leads. Leads introduced are appropriate for further drug design and discovery. One of the main advantages of applications of database searching is that the retrieved molecules are typically more easily available for testing than those based on Denovo design techniques [66]. A molecule must be able to map the entire features of the developed hypothesis to be listed as a hit. All screening experiments were carried out employing the “Ligand Pharmacophore Mapping” protocol with the Best Flexible Search option as available in ADS. Hit molecules from the database searching with less than 4 nM predicted IC<sub>50</sub> values were retained. In addition, hit compounds with good estimated activity were predicted for the drug-likeness using Lipinski's rule of five [67]. A "Lipinski rule of five" said a drug candidate molecule has

(i) a molecular weight less than 500; (ii) less than 10 hydrogen bond acceptor groups; (iii) less than 5 hydrogen bond donor groups and (iv) an octanol/water partition co-efficient ( $\log P$ ) value less than 5.

## 2.8. Docking protocol

Using the equilibrated structure of P2X<sub>7</sub>R resulted from MD simulation, compounds that were predicted to be positive in Lipinski drug-likeness screening were subjected to molecular docking studies. The structure of P2X<sub>7</sub>R built by HM in previous step was used. Then, after determining Kollman united atom charges [61] and merging non-polar hydrogens, rotatable bonds were assigned. Using auto Grid tool, the grid maps (one for each atom type in the ligand, and one for electrostatic interactions) were constructed adequately large to include the active site of protein as well as significant regions of the surrounding surface. In all the cases, a grid map of 70 points in each Cartesian direction besides a grid-point spacing of 0.375 Å (a quarter of the carbon-carbon single bond) were generated. By the ligand location in the complex, the maps were centered on the ligand's binding site, searching favorable interactions with the functional groups. Based on "Lamarckian Genetic Algorithm" [59], using the pseudo-Solis and Wets local search method [62], Auto Dock Tools was employed to produce both grid and docking parameter files i.e. *.gpf* and *.dpf*. Applying 2.0 Å clustering tolerance to construct clusters of the closest compounds, the initial coordinates of the ligand were used as the reference structure.

## 3. Results and discussion

### 3.1 Sequence alignment and homology modeling

The building of protein models by HM normally proceeds along a series of well-defined and commonly accepted steps: (1) sequence alignment between the target and the template; (2) building an initial model; (3) refining the model; (4) evaluating the quality of the model [68,69].

The crystal structure of A-chain of ATP-gated P2X<sub>4</sub> ion channel receptor was the structurally homologous protein that was found by Blastp analysis and hence chosen as template for developing the P2X<sub>7</sub>R model.

An optimal sequence alignment is necessary to the success of a typical HM procedure. The sequence identity between P2X<sub>4</sub>R and P2X<sub>7</sub>R is ~48.16% in the primary structure (Figure S1), making this element of the alignment relatively straightforward. To optimize the alignment, the initial alignment was adjusted so as to maximize overlap between the predicted locations of the elements of secondary structure such as helices in P2X<sub>7</sub>R and their locations in the experimentally resolved structure of P2X<sub>4</sub>R. This provided a first-pass refinement of the alignment of the proteins. Just about 400 residues of P2X<sub>4</sub>R have been used for HM.

According to report [70] if two proteins have more than 30% “identity”, then their  $C_{\alpha}$  atoms are within 3.5 Å of their correct position.

To study the orientation of the secondary structure elements of P2X<sub>7</sub>R such as helices, we overlaid the regions of the generated model with the template. From the obtained *RMSD* value (1.2 nm) we can deduce that there is a good agreement between the proteins.

Further checked by Procheck, the final P2X<sub>7</sub>R model indicates that more than 99 % of residue  $\phi$ - $\psi$  angles are in the favored or additional allowed regions of Ramachandran plot (Figure S2), showing again that the final obtained 3D model of P2X<sub>7</sub>R is satisfactory. With

respect to Ramachandran plot, it is observed only three residues are in disallowed region. Residues located in the unfavorable regions are far from the substrate-binding domain, indicating that these residues may not affect the ligand–protein binding.

A parameter indicating the quality of covalent and bond angle distance is the G-factors. The G-factor was -0.08° for dihedrals, -0.23° for covalent and overall -0.13°. The overall main-chain and side-chain parameters, as evaluated by Procheck, are all very favorable. No clash between residues of the model has also been identified in the viewer. The final step of testing is the packing quality of each residue as evaluated by the Verify-3D method represents the profile attained with respect to the residues. The compatibility score above zero in the Verify-3D graph is corresponding to acceptable side-chain environments (Figure S3). This suggests that the model has overall self-consistency in terms of sequence-structure compatibility.

### **3.2 Geometry of structure**

The P2X<sub>7</sub>R receptor adopts a chalice-shaped, homotrimetric geometry consisting of a large hydrophilic and glycosylated extracellular domain, a transmembrane domain composed of 6 α-helices and short intracellular amino and carboxy termini (Figure 1).

Each subunit resembles the shape of a dolphin, with the transmembrane helices and the extracellular region akin to the tails and the body, respectively (Figure 4C).

MD simulations of the developed 3D structure of P2X<sub>7</sub>R embedded in a lipid bilayer were carried out using trimeric assemblies. Each of assemble was build using three subunits resulted from HM. Subunits were put together manually.

<Figure 1 near here>

MD simulations of the developed 3D structure of P2X<sub>7</sub>R embedded in a lipid bilayer were carried out for two reasons: (i) to “refine” the developed homology model via simulations in an explicit biomembrane plus water environment, by helping recognition of the less

conformationally stable regions and therefore (presumably) less well-modeled parts of the structure; and (ii) to obtain reliable structure P2X<sub>7</sub>R allowing investigate interactions with ligands and introduced hits in docking step. Accordingly, the P2X<sub>7</sub>R model was inserted into a POPC lipid bilayer that contained a hole of dimensions sufficient to accommodate P2X<sub>7</sub>R. The resultant system was solvated, and counterions were added by replacing water molecules randomly, to give an overall electroneutral system. To optimize the conformations of the lipids after protein insertion, a 50000 ps simulation was run, during which the protein atoms were restrained to their initial positions while the lipids and water were free to move. MD simulation was carried out to determine the stability of the predicted 3D structure of P2X<sub>7</sub>R. The trajectories were stable during the whole production part of the 50000 ps MD simulation run. The trajectory stability was monitored and was confirmed by the analysis of  $C_a$  RMSD (Figure 2) and the total energy (Figure 3) as a function of time. On the other words, to illustrate the detailed events occurred by the MD simulation in studied system, Figure 2 shows the time history of RMSD for protein structure immersed in lipid bilayer relative to the starting structure (the output of HM process). As can been seen  $C_a$  RMSD was about 1 Å after 50 ns of simulation. The RMSD value implies that this protein structure has been affected by its environment. Therefore, the used MD simulation was essential to specify native geometry of P2X<sub>7</sub>R embedded in lipid bilayer. In Figure 3 Variation in total energy versus time in 50 ns of MD is illustrated, small fluctuation was observed in the last 45 ns of MD.

<Figure 2 near here>

<Figure 3 near here>

Furthermore, average temperature of 50 ns of simulation at 310 K for the investigated system was equal to 310±1.4 K (Figure S4). Therefore, the extracted equilibrium structure at 310 K

belonging to the P2X<sub>7</sub>R was obtained under stable temperature conditions. These facts show that energy conservation satisfies in MD simulation.

How is the developed 3D model of P2X<sub>7</sub>R affected due to the bilayer environment? The mobility of the generated structure of 3D model of P2X<sub>7</sub>R in the membrane was studied by computing the rms fluctuations (*RMSF*) of the C<sub>α</sub> backbone structure versus the average structure during the simulation (Figure 4A)

The *RMSF* is measure of mobility. It can be seen from figure that the helical domains of P2X<sub>7</sub>R exhibit relatively small fluctuations, whereas the extramembranous domains, show higher mobility. A similar pattern emerges from analysis of the secondary structures of the 3D model of P2X<sub>7</sub>R as functions of time (Figure 4B).

It was found that throughout the dynamic simulations very few fluctuations exceeded 0.5 nm and even less fluctuations over passed 0.6 nm for total protein. The residues (49-60, 105-108, 131-133, 277) with fluctuations close to 0.6 nm observed in the dynamic plots were located at the loops of body region (Figure 4A, B).

<Figure 4 near here>

After structural refinement of P2X<sub>7</sub>R model by MD, the geometric quality of the backbone conformation, namely all tests performed in HM step, was carried out again and the quality of model was confirmed (data not shown).

In brief, the quality of the geometry of developed model, the residue interaction, the residue contact, and the dynamic stability of the structure are in a good level and within the limits established for reliable structures. Passing all checks by developed model proposes that a desired model for P2X<sub>7</sub>R is a good estimation to describe various protein-ligand interactions and to study the relation between the structure and function.

### 3.3. Pharmacophore modeling

Before beginning of pharmacophore modeling procedure, a total of 49 P2X<sub>7</sub>R antagonists were gathered from published resources. Of these molecules, 29 were selected to form a training set based on Kennard and Stone algorithm.

The top ten hypotheses were composed of hydrogen bond acceptor (HBA), hydrogen bond acceptor lipid (HBA-L), hydrophobic (HY), hydrophobic aromatic (HY-A), ring aromatic (RA) features (Table 2) which reports the statistics of the generated pharmacophore hypotheses. The values of the ten hypotheses such as pharmacophore features, root-mean-square deviations (*RMSD*), correlation, and cost values showed statistical significance (Table 2).

<Table 2 near here>

All of developed hypothesis contained at least three chemical features. A significant pharmacophore model should have a large difference between its total and null cost values. In this work, the best hypothesis, Hypo1, as indicated in Figure 5 and reported in Table 2 is characterized by the lowest total cost value (186.5), the highest cost difference (368.677) and the lowest *RMSD* (2.87). Developed pharmacophore showed the highest correlation coefficient value of 0.87, highlighting its strong predictive ability. The fixed cost and null cost are 66.403 and 555.177 bits, respectively. The total cost is low and close to the fixed cost, as well as being less and differs greatly from the null cost. The entire evidences indicate that the developed model has good predictive ability. Consequently, Hypo1 was chosen as the best pharmacophore model for further analyses and application.

<Figure 5 near here>

### **3.4. Application of developed Pharmacophore model on training set**

As mentioned above, all of the molecules in the studied dataset set were categorized into four different groups based on their experimental activity ( $IC_{50}$ ) values active ( $IC_{50} \leq 50$  nM, ++++), moderately active ( $50 \leq IC_{50} \leq 200$  nM, +++) , less active ( $200 \leq IC_{50} \leq 1000$  nM, ++), and inactive ( $IC_{50} > 1000$  nM, +) (Table 3). The activity of each training set molecule was predicted by a fitting procedure based on the best pharmacophore model, Hypo1, and the results are displayed in Table 3. As it can be seen, six of training set molecules were predicted as having different  $IC_{50}$ s than their experimental values. Noticeably, all active molecules in the training set were predicted as active P2X<sub>7</sub>R antagonists with predicted activity values, which are very close to their experimental  $IC_{50}$  values except molecule 22. Two moderately active compounds were underestimated as less active compounds (molecules 14 and 16) and two compounds were overestimated as active compounds (molecules 20 and 40). One of inactive molecule was overestimated and predicted to be less active (molecule 1). Error values depict the ratio between the experimental and estimated activity values. Positive error values are calculated when the predicted activity value is higher than the observed value and a negative value shows the opposite. All of the active compounds listed in Table 3 contained all of the four chemical features in Hypo1, whereas all of the other compounds mapped four or less pharmacophoric features of hypothesis. Figure 6A and B depicts the mapping of the least and most active molecules (molecule 1 and 36 respectively) of the training set on Hypo1, respectively. The fluorine atom of difluorobenzene ring and nitrogen atom of pyridine ring present in the most active molecule (molecule 36) mapped upon the two present HY-A features, while the imidazole ring of purine mapped over the HY feature. The fit values of these least and most active compounds are 6.08 and 7.21, respectively.

<Table 3 near here>

<Figure 6 near here>

### 3.5. Validation methods applied on developed model

As well as the training set prediction by Hypo1, the predictive ability of the best developed pharmacophore model was tested using additional methods such as cost analysis, prediction of biological activity of test set, Fischer randomization, and E value calculation. Cost analysis is based on the statistical cost values generated during pharmacophore model building phase. A diverse test set was employed to verify if the pharmacophore model predicts the biological IC<sub>50</sub> of the molecules that are structurally distinct to the training set. The Fischer randomization test also was used to verify that the chosen pharmacophore model was not generated as a result of chance correlation. The E value calculation was built to verify the selectivity of the developed pharmacophore model towards actives molecules rather than inactives.

#### 3.5.1. Cost analysis

The “Common Feature Pharmacophore Generation” algorithm generated three cost values during pharmacophore building step to evaluate the quality and reliability of the pharmacophore hypothesis. As described in Method section, the first cost value is the fixed cost value (also called ideal cost) denotes the simplest model that fits the data completely. The second one is the null cost value (no correlation cost) denotes the highest cost of a pharmacophore with no features estimating the activity to be the average IC<sub>50</sub>s of the training set molecules. A statistical significant and predictive pharmacophore hypothesis should have a large difference between these two cost values. Hypo1 was generated with a fixed cost value of 66.403 and a null cost value of 555.177, thus with a difference of 488.774. The third cost is the total cost value estimated for every pharmacophore hypothesis and should be close to the fixed cost value. A large difference between the total and null costs shows a more predictive and statistical meaningful pharmacophore model. Hypo1 scored a total cost value

of 186.5, which is closer to the fixed cost, for a cost difference of 368.677 (shown in Table 2).

### **3.5.2. Test set prediction**

A set of 20 molecules with structures quite similar to training set and range of IC<sub>50</sub> values was employed to assess the best developed pharmacophore model, Hypo1. The chemical structures of the test set compounds are provided in Table 1. The “Ligand Pharmacophore Mapping” protocol implemented in ADS with the Best Flexible Search option was applied to map all of the molecules in test data (Table 4). Using this protocol the activity values were calculated for each molecule in test data. In particular, no compound in the test set was predicted with an error value more than 8 except molecule 47, not exhibiting more than one order of magnitude between experimental and estimated activities (Table 4). Noticeably, 40% (8 molecules) of the test set molecules were predicted within their IC<sub>50</sub> scales while the remaining 60% (12 molecules) were estimated in different activity scales. From these 12 molecules, 1 active molecule was underestimated as moderately active; 3 moderately active molecules were overestimated as active molecules; 6 less active molecules were overestimated as moderately active molecules and 2 poor active molecules were overestimated as moderately active molecules. All of the less active and inactive compounds were predicted within their activity scales.

<Table 4 near here>

### **3.5.3. Fischer randomization test**

Furthermore, Fischer randomization test technique was applied to evaluate the statistical robustness of developed pharmacophore model (Hypo1). The third method to validate the

robustness of the developed model is based on Fischer's randomization technique. The observed biological activities of the training set were shuffled randomly and the resulting training set was used in "Common Feature Pharmacophore Generation" protocol with the parameters selected for the original model building step. Thereby, a set of 19 random tables was generated to reach a 95% confidence level that the best pharmacophore Hypo1 was not developed by chance. Figure 7 indicates that none of the randomly developed pharmacophore models were generated with better statistical values than Hypo1. The results of Fischer randomization test technique clearly demonstrate that the original hypothesis is far more superior to those of the 19 randomization produced hypotheses, which give confidence on developed pharmacophore model.

<Figure 7 near here>

### **3.5.4 Estimation of Enrichment factor**

Generated pharmacophore model was also validated employing a special approach. This approach determines whether the best hypothesis can choose active molecules during the virtual screening procedure from a database of 1676 molecules consisting of 44 experimentally determined P2X<sub>7</sub>R antagonists retrieved from recently published studies.

Statistics used in this section includes calculation of false positives, false negatives, enrichment factor, and goodness of hit to determine the robustness of the generated hypotheses (reported in Table 5). Not only should the pharmacophore model generated predict the biological activity of the molecules applied for model building, but it should also be skilled for predicting the biological activities of other molecules as active or inactive. Using the best developed pharmacophore model, Hypo1, 45 molecules ( $H_t$ ) were retrieved as hits from the database screening.

Among these hits, 36 ( $H_a$ ) molecules were from the A list of known antagonists. Therefore, the enrichment factor was calculated to be 30.47, indicating that model is 30.47 times more probable to pick an active compound from the database than an inactive one. This value of enrichment factor and a GH score of 0.86 indicated the quality of the model and high efficiency of the screening test.

As it can be seen from this table, selected pharmacophore model is successful in retrieving 80% of the active molecules, 9 inactive molecules as active (false positives), and predicted 8 active molecules as inactive (false negatives). Overall, a strong correlation was observed between the selected pharmacophore model calculated biological activity and the observed P2X<sub>7</sub>R inhibitory activity (IC<sub>50</sub>) of the training and test set molecules.

<Table 5 near here>

### **3.6 Virtual screening and drug-likeness prediction**

In drug design and discovery procedure virtual screening (database searching) is an efficient alternative way to high throughput screening (HTS). The best pharmacophore model, Hypo1, was used as a 3D query to search a chemical databases, CoCoCo-SC Asinex (514,957 compounds) and Maybridge (54,318) for a total of 569275 compounds. The “Ligand Pharmacophore Mapping protocol” with the Best Flexible Search option was employed to search these databases. Inhibitory activity values were estimated for the compounds obtained from the database screening. A total of 136099 (81954 and 54145 molecules from CoCoCo-SC Asinex and Maybridge, respectively) molecules were mapped upon all of the pharmacophoric features present in Hypo1. A total of 148 compounds mapped in previous step scored an estimated activity value less than 4 nM and were considered for further studies.

### 3.7 Lipinski's rule of five evaluation

Drug-likeness properties are one of the key indicators for selecting the molecules for in vitro studies, which includes molecular or physicochemical properties that contribute to favorable Lipinski's rule of five. The parameters described in the Lipinski's rule of five including  $\log P$  (the logarithm of octanol/water partition coefficient), number of hydrogen bond donor groups, number of hydrogen bond acceptor groups and molecular weight. They have been proved to have a correlate with drug absorption. These properties describe the 'drug-likeness' and predict a poor oral absorption or permeation when the investigated molecules have more than five H-bond donors (HBD), 10 H-bond acceptors (HBA), a molecular weight ( $M_w$ ) greater than 500 Da and calculated  $\log P$  (cLogP) higher than 5. Compounds breaching more than one of the conditions may have small oral bioavailability. However, among 148 considered compounds, the 22 compounds that are listed in Table 6 did not breach any parameter of Lipinski's proposed rule, and thus are supposed to have high bioavailability.

So, finally 22 molecules were further selected for docking studies.

<Table 6 near here>

### 3.8 Docking

The proposed approach was further validated by docking a series of retrieved inhibitors (The 22 hit compounds that were chosen from the pharmacophore filtering studies) reported in Table 6 in the binding site. Docking studies were carried out to gain insight into the binding conformation of pharmacophore models derived from studied dataset. Docking studies on binding modes are very informative to clarify key structural characteristics and interactions to provide helpful data for suggesting effective P2X<sub>7</sub>R inhibitors.

Two subunits, B and C, were used in this study for docking. As previously reported, most of the conserved residues within or adjacent to a pocket formed between the body region of two subunits (Figure 8) [4].

<Figure 8 near here>

To take a snapshot of the activities and  $\Delta G$  of the selected compounds, the predicted  $\Delta G$  (estimated free energy change of binding) values for each compound are presented in Table 7.

<Table 7 near here>

With respect to the obtained results, compounds Maybridge 15178, Maybridge 15177, Asinex 93637, Asinex 251362, Asinex 260860 and Asinex 469708 were selected for further evaluation (Figure 9). As reported in Table 7, these compounds have more negative estimated free energy of binding value than -6.5. It is obvious from Table 7 that the order of predicted bioactivities of these compounds is:

Maybridge 15178> Asinex 93637> Asinex 260860> Maybridge 15177> Asinex 469708>  
Asinex 251362

<Figure 9 near here>

On the basis of  $\Delta G$  binding, the order is: Maybridge 15178> Maybridge 15177> Asinex 93637> Asinex 251362> Asinex 260860> Asinex 469708, that is, compound Maybridge 15178 interacts more strongly with P2X<sub>7</sub>R binding site than the other compounds. The binding modes and molecular interactions between compound Maybridge 15178 (with higher value of  $\Delta G$  binding) and the active site components are discussed below.

Inspection of docking results derived from P2X<sub>7</sub>R- Maybridge 15178 complex immediately showed a prominent feature consistent with the molecular composition of an ATP molecule. Previous experimental studies show that there are three equivalent ATP-binding sites at each of the three pairs of subunit interfaces in the trimeric receptor. It must be noted that binding sites located about 40 Å from the extracellular boundary of the transmembrane domain. Ligand is recognized by the upper (Chain B) and lower (Chain C) body domains through extensive hydrophilic interactions (Figure 10).

<Figure 10 near here>

Compound Maybridge 15178 binds to the P2X<sub>7</sub>R catalytic ATP site. Docking results revealed that the amino acids Pro96 (Chain C), Tyr299 (Chain C), Arg294 (Chain C), Ala296 (Chain B), Tyr295 (Chain B), Tyr298 (Chain B), Phe95 (Chain C), Leu97 (Chain C), Val312 (Chain C) and Met105 (Chain C) (Figure 10) played vital role to bind the introduced lead compound. The pyrimidine ring of compound Maybridge 15178 is situated in the pocket with high degree of hydrophobic property that binds the adenine of ATP. This pocket includes the side chains of residues Val312 (Chain C) and Leu97 (Chain C).

The thiazole ring of Maybridge 15178 and Tyr299 (Chain C) lay with their rings in parallel (distance between the rings <3.5 Å), and forming a van der Waal contact. Since the structure of compound Maybridge 15178 at physiological pH accept a proton on N of thiazole ring, are very probable, the more likely explanation of our results is that, when the thiazole ring becomes protonated, a new attractive interaction is established between the aromatic ring of the Tyr299 (Chain C) residue and the charge of the protonated thiazole ring. This is a typical example of a catione-aromatic interaction where the cation is a protonated aromatic ring.

Our docking results and binding data on P2X<sub>7</sub>R ligands were comparable with the results of an earlier experimental study of P2X receptors [4,71].

Experimental studies including site directed mutations [72-76] led to an important result that the ATP binding site is generally conserved within the P2X receptor family, and positively charged amino acid residues coordinate the negatively charged phosphate oxygens of ATP. In particular, amino acids such as Lys298(Chain B), Lys300(Chain B) and Arg294(Chain C) (correspond to Lys68, Lys70, Arg292 and Lys 309 in P2X<sub>1</sub>R) were shown to be of importance for ATP potency (Figure 10) [76-80]. Docking results showed residues Phe95 (Chain C) contribute to agonist action.

Binding mode of the lead compound correlated well with the pharmacophore overlay (Figure 11). These findings well corroborates with the best pharmacophore hypothesis where the importance of hydrophobic functionality at the active site has been described by HY feature.

<Figure 11 near here>

#### **4. Conclusion**

In this study, In order to identify new P2X<sub>7</sub>R inhibitors, a sequential combination of pharmacophore modeling, virtual screening, HM, MD simulation and docking methodologies was used.

In the first part of study, a detailed 3D structure of the P2X<sub>7</sub>R is constructed using the HM procedure. The resulting model, consisting of the transmembrane, extra and intracellular segments, is validated by MD simulation. It can be noted that the general structure of the TM helices remained stable during the MD simulation. The generated model is used in pharmacophore modeling.

Three-dimensional distances among pharmacophoric features were used as the criteria in the screening process. Since the top 10 pharmacophore models contained the same feature (i.e., HY, HY-A, RA), first hypothesis (Hypo1) seemed to be the most predictive model with

highest rank score. The selected hypothesis was made of two aromatic hydrophobic, one hydrophobic and one aromatic ring (Figure 5) features with a high correlation value of 0.874. This hypothesis was validated using 20 molecules assigned as test set compounds. In addition, Fischer randomization and E value calculation were also used in pharmacophore validation. This 3D pharmacophore model was then further assessed by using it to search 3D databases. The model was further used in database screening to find novel and diverse virtual leads for P2X<sub>7</sub>R inhibitors. The retrieved compounds from database searching were further examined at the active site of P2X<sub>7</sub>R protein where the docking study well collaborates with the pharmacophore model. All results suggested that this model has accurate prediction for P2X<sub>7</sub>R antagonists. Moreover, the pharmacophore features in Hypo1 and docking conformation were correctly matched onto P2X<sub>7</sub>R protein binding pocket. Therefore, the developed pharmacophore hypothesis may be useful in finding new scaffolds that may aid in design and develop New Chemical Entities as potent P2X<sub>7</sub>R inhibitors.

Finally, it can be concluded that such a rational method enhances the efficiency in recognizing novel lead compounds. This approach can also help to decrease time consuming and cost-intensive steps in the drug design and development procedure.

### Acknowledgment

The authors gratefully acknowledge the Vice Chancellor for Research and Technology, Kermanshah University of Medical Sciences for financial support.

### References

1. Burnstock G, Kennedy C (1985) Is there a basis for distinguishing two types of P<sub>2</sub>-purinoceptor? General Pharmacology: The Vascular System 16 (5):433-440

2. North RA (2002) Molecular physiology of P2X receptors. *Physiological reviews* 82 (4):1013-1067
3. Alves LA, Bezerra RJS, Faria RX, Ferreira LGB, da Silva Frutuoso V (2013) Physiological roles and potential therapeutic applications of the P2X7 receptor in inflammation and pain. *Molecules* 18 (9):10953-10972
4. Browne LE, Jiang L-H, North RA (2010) New structure enlivens interest in P2X receptors. *Trends in pharmacological sciences* 31 (5):229-237
5. Di Virgilio F (2007) Liaisons dangereuses: P2X<sub>7</sub> and the inflammasome. *Trends in pharmacological sciences* 28 (9):465-472
6. Carroll WA, Donnelly-Roberts D, Jarvis MF (2009) Selective P2X7 receptor antagonists for chronic inflammation and pain. *Purinergic signalling* 5 (1):63-73
7. Guile SD, Alcaraz L, Birkinshaw TN, Bowers KC, Ebden MR, Furber M, Stocks MJ (2009) Antagonists of the P2X7 receptor. From lead identification to drug development. *Journal of medicinal chemistry* 52 (10):3123-3141
8. Surprenant A, Rassendren F, Kawashima E, North R, Buell G (1996) The cytolytic P2Z receptor for extracellular ATP identified as a P2X receptor (P2X7). *Science* 272 (5262):735-738
9. Rassendren F, Buell GN, Virginio C, Collo G, North RA, Surprenant A (1997) The permeabilizing ATP receptor, P2X7 Cloning and expression of a human cDNA. *Journal of Biological Chemistry* 272 (9):5482-5486
10. North RA, Surprenant A (2000) Pharmacology of cloned P2X receptors. *Annual Review of Pharmacology and Toxicology* 40 (1):563-580
11. Skaper SD, Debetto P, Giusti P (2009) P2 X 7 Receptors in Neurological and Cardiovascular Disorders. *Cardiovascular psychiatry and neurology* 2009

12. Mager PP, Illes P, Walther H (2006) Prediction of the conformation of the human P2X7 receptor. *Letters in Drug Design & Discovery* 3 (10):675-682
13. Csermely P, Korcsmáros T, Kiss HJ, London G, Nussinov R (2013) Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacology & therapeutics* 138 (3):333-408
14. Islam MA, Pillay TS (2014) Exploration of the structural requirements of HIV-protease inhibitors using pharmacophore, virtual screening and molecular docking approaches for lead identification. *Journal of Molecular Graphics and Modelling*
15. Yang S-Y (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* 15 (11):444-450
16. Vuorinen A, Schuster D (2014) Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*
17. Wulff G (1995) Molecular imprinting in cross-linked materials with the aid of molecular templates—a way towards artificial antibodies. *Angewandte Chemie International Edition in English* 34 (17):1812-1832
18. Wermuth C, Ganellin C, Lindberg P, Mitscher L (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry* 70 (5):1129-1143
19. Guner OF (2005) The impact of pharmacophore modeling in drug design. *IDrugs: the investigational drugs journal* 8 (7):567-572
20. Tintori C, Corradi V, Magnani M, Manetti F, Botta M (2008) Targets looking for drugs: a multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *Journal of chemical information and modeling* 48 (11):2166-2179

21. Khedkar SA, Malde AK, Coutinho EC, Srivastava S (2007) Pharmacophore modeling in drug discovery and development: an overview. *Medicinal Chemistry* 3 (2):187-197
22. Drwal MN, Griffith R (2013) Combination of ligand-and structure-based methods in virtual screening. *Drug Discovery Today: Technologies* 10 (3):e395-e401
23. Drwal MN, Marinello J, Manzo SG, Wakelin LP, Capranico G, Griffith R (2014) Novel DNA Topoisomerase II $\alpha$  Inhibitors from Combined Ligand-and Structure-Based Virtual Screening. *PloS one* 9 (12):e114904
24. Wilson GL, Lill MA (2011) Integrating structure-based and ligand-based approaches for computational drug design. *Future medicinal chemistry* 3 (6):735-750
25. Hein M, Zilian D, Sotriffer CA (2011) Docking compared to 3D-pharmacophores: the scoring function challenge. *Drug Discovery Today: Technologies* 7 (4):e229-e236
26. Khan KM, Wadood A, Ali M, Ul-Haq Z, Lodhi MA, Khan M, Perveen S, Choudhary MI (2010) Identification of potent urease inhibitors< i> via</i> ligand-and structure-based virtual screening and< i> in vitro</i> assays. *Journal of Molecular Graphics and Modelling* 28 (8):792-798
27. Weidlich IE, Dexheimer T, Marchand C, Antony S, Pommier Y, Nicklaus MC (2010) Inhibitors of human tyrosyl-DNA phosphodiesterase (hTdp1) developed by virtual screening using ligand-based pharmacophores. *Bioorganic & medicinal chemistry* 18 (1):182-189
28. Banoglu E, Çalışkan B, Luderer S, Eren G, Özkan Y, Altenhofen W, Weinigel C, Barz D, Gerstmeier J, Pergola C (2012) Identification of novel benzimidazole derivatives as inhibitors of leukotriene biosynthesis by virtual screening targeting 5-lipoxygenase-activating protein (FLAP). *Bioorganic & medicinal chemistry* 20 (12):3728-3741
29. Drwal MN, Agama K, Wakelin LP, Pommier Y, Griffith R (2011) Exploring DNA topoisomerase I ligand space in search of novel anticancer agents. *PloS one* 6 (9):e25150

30. Smith JR, Evans KJ, Wright A, Willows RD, Jamie JF, Griffith R (2012) Novel indoleamine 2, 3-dioxygenase-1 inhibitors from a multistep in silico screen. *Bioorganic & medicinal chemistry* 20 (3):1354-1363
31. Saxena S, Chaudhaery S, Varshney K, Saxena A (2010) Pharmacophore-based virtual screening and docking studies on Hsp90 inhibitors. *SAR and QSAR in Environmental Research* 21 (5-6):445-462
32. Sakkiah S, Thangapandian S, John S, Lee KW (2011) Pharmacophore based virtual screening, molecular docking studies to design potent heat shock protein 90 inhibitors. *European journal of medicinal chemistry* 46 (7):2937-2947
33. Shahlaei M, Mousavi A (2015) A Conformational Analysis Study on the Melanocortin 4 receptor Using Multiple Molecular Dynamics Simulations. *Chemical biology & drug design*
34. Shahlaei M, Madadkar-Sobhani A, Mahnam K, Fassihi A, Saghai L, Mansourian M (2011) Homology modeling of human CCR5 and analysis of its binding properties through molecular docking and molecular dynamics simulation. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1808 (3):802-817
35. Shahlaei M, Mousavib A (2014) A 3D Model for Human Melanocortin 4 Receptor Refined with Molecular Dynamics Simulation. *Journal of Reports in Pharmaceutical Sciences (J Rep Pharm Sci)* 3 (1):42-53
36. Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. *Journal of pharmacological and toxicological methods* 44 (1):235-249
37. Matasi JJ, Brumfield S, Tulshian D, Czarnecki M, Greenlee W, Garlisi CG, Qiu H, Devito K, Chen S-C, Sun Y (2011) Synthesis and SAR development of novel P2X<sub>7</sub> receptor antagonists for the treatment of pain: Part 1. *Bioorganic & medicinal chemistry letters* 21 (12):3805-3808

38. Brumfield S, Matasi JJ, Tulshian D, Czarniecki M, Greenlee W, Garlisi C, Qiu H, Devito K, Chen S-C, Sun Y (2011) Synthesis and SAR development of novel P2X<sub>7</sub> receptor antagonists for the treatment of pain: Part 2. Bioorganic & medicinal chemistry letters 21 (24):7287-7290
39. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25 (17):3389-3402
40. Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology 234 (3):779-815
41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography 26 (2):283-291
42. Liithy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356 (6364):83-85
43. Shahlaei M, Madadkar-Sobhani A, Fassihi A, Saghai L (2011) Exploring a model of a chemokine receptor/ligand complex in an explicit membrane environment by molecular dynamics simulation: the human CCR1 receptor. Journal of chemical information and modeling 51 (10):2717-2730
44. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. Journal of computational chemistry 26 (16):1701-1718
45. Tieleman D, Berendsen H (1996) Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. The Journal of chemical physics 105 (11):4871-4880
46. van Gunsteren WF, Billeter S, Eising A, Hünenberger PH, Krüger P, Mark AE, Scott W, Tironi IG (1996) Biomolecular simulation: The {GROMOS96} manual and user guide.

47. Tieleman DP, MacCallum JL, Ash WL, Kandt C, Xu Z, Monticelli L (2006) Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid–protein interactions, side chain transfer free energies and model proteins. *Journal of Physics: Condensed Matter* 18 (28):S1221
48. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *The Journal of chemical physics* 98 (12):10089-10092
49. Hess B, Bekker H, Berendsen HJ, Fraaije JG (1997) LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* 18 (12):1463-1472
50. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14 (1):33-38
51. Ultra C (2001) 6.0 and Chem3D Ultra. Cambridge Soft Corporation, Cambridge, USA
52. Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11 (1):137-148
53. Schuster D, Laggner C, Steindl TM, Paluszak A, Hartmann RW, Langer T (2006) Pharmacophore modeling and in silico screening for new P450 19 (aromatase) inhibitors. *Journal of chemical information and modeling* 46 (3):1301-1311
54. Bharatham N, Bharatham K, Lee KW (2007) Pharmacophore identification and virtual screening for methionyl-tRNA synthetase inhibitors. *Journal of Molecular Graphics and Modelling* 25 (6):813-823

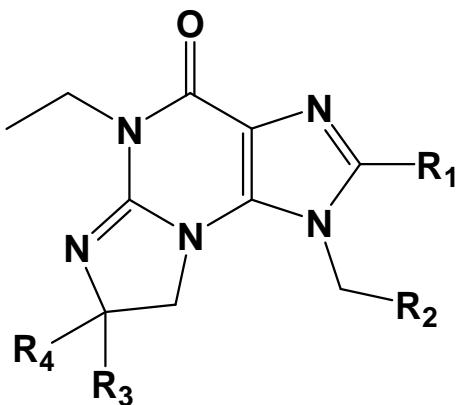
55. Neves MA, Dinis TC, Colombo G, Sá e Melo ML (2009) An efficient steroid pharmacophore-based strategy to identify new aromatase inhibitors. European journal of medicinal chemistry 44 (10):4121-4127
56. Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural networks 2 (5):359-366
57. Shahlaei M, Fassihi A, Saghie L, Arkan E, Madadkar-Sobhani A, Pourhossein A (2013) Computational evaluation of some indenopyrazole derivatives as anticancer compounds; application of QSAR and docking methodologies. Journal of enzyme inhibition and medicinal chemistry 28 (1):16-32
58. Ayatollahi AM, Ghanadian M, Afsharypuor S, Mesaik MA, Abdalla OM, Shahlaei M, Farzandi G, Mostafavi H (2011) Cycloartanes from Euphorbia aellenii Rech. f. and their Antiproliferative Activity. Iranian journal of pharmaceutical research: IJPR 10 (1):105
59. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of computational chemistry 19 (14):1639-1662
60. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36 (22):3219-3228
61. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. Journal of the American Chemical Society 106 (3):765-784
62. Solis FJ, Wets RJ-B (1981) Minimization by random search techniques. Mathematics of operations research 6 (1):19-30
63. Güner O, Henry DR (2000) Metric for analyzing hit lists and pharmacophores. Pharmacophore Perception, Development, and Use in Drug Design:191-211

64. Clement OO, Freeman CM, Hartmann RW, Handratta VD, Vasaitis TS, Brodie AM, Njar VC (2003) Three dimensional pharmacophore modeling of human CYP17 inhibitors. Potential agents for prostate cancer therapy. *Journal of medicinal chemistry* 46 (12):2345-2351
65. Perez-Medrano A, Donnelly-Roberts DL, Florjancic AS, Nelson DW, Li T, Namovic MT, Peddi S, Faltynek CR, Jarvis MF, Carroll WA (2011) Synthesis and in vitro activity of N-benzyl-1-(2, 3-dichlorophenyl)-1H-tetrazol-5-amine P2X 7 antagonists. *Bioorganic & medicinal chemistry letters* 21 (11):3297-3300
66. Langer T, Wolber G (2004) Pharmacophore definition and 3D searches. *Drug Discovery Today: Technologies* 1 (3):203-207
67. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* 64:4-17
68. Forrest LR, Tang CL, Honig B (2006) On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophysical journal* 91 (2):508-517
69. Ginalski K (2006) Comparative modeling for protein structure prediction. *Current opinion in structural biology* 16 (2):172-177
70. Sánchez R, Šali A (1999) Comparative protein structure modeling in genomics. *Journal of Computational Physics* 151 (1):388-401
71. Hattori M, Gouaux E (2012) Molecular mechanism of ATP binding and ion channel activation in P2X receptors. *Nature* 485 (7397):207-212
72. Evans R (2009) Orthosteric and allosteric binding sites of P2X receptors. *European Biophysics Journal* 38 (3):319-327

73. Bodnar M, Wang H, Riedel T, Hintze S, Kato E, Fallah G, Gröger-Arndt H, Giniatullin R, Grohmann M, Hausmann R (2011) Amino acid residues constituting the agonist binding site of the human P2X3 receptor. *Journal of Biological Chemistry* 286 (4):2739-2749
74. Fischer W, Zadori Z, Kullnick Y, Gröger-Arndt H, Franke H, Wirkner K, Illes P, Mager PP (2007) Conserved lysin and arginin residues in the extracellular loop of P2X 3 receptors are involved in agonist binding. *European journal of pharmacology* 576 (1):7-17
75. Rassendren F, Buell G, Newbolt A, North RA, Surprenant A (1997) Identification of amino acid residues contributing to the pore of a P2X receptor. *The EMBO journal* 16 (12):3446-3454
76. Roberts JA, Digby HR, Kara M, El Ajouz S, Sutcliffe MJ, Evans RJ (2008) Cysteine substitution mutagenesis and the effects of methanethiosulfonate reagents at P2X2 and P2X4 receptors support a core common mode of ATP action at P2X receptors. *Journal of Biological Chemistry* 283 (29):20126-20136
77. Jiang L-H, Rassendren F, Surprenant A, North RA (2000) Identification of amino acid residues contributing to the ATP-binding site of a purinergic P2X receptor. *Journal of Biological Chemistry* 275 (44):34190-34196
78. Yan Z, Liang Z, Tomić M, Obsil T, Stojilkovic SS (2005) Molecular determinants of the agonist binding domain of a P2X receptor channel. *Molecular pharmacology* 67 (4):1078-1088
79. Zemkova H, Yan Z, Liang Z, Jelinkova I, Tomic M, Stojilkovic SS (2007) Role of aromatic and charged ectodomain residues in the P2X4 receptor functions. *Journal of neurochemistry* 102 (4):1139-1150
80. Ennion S, Hagan S, Evans RJ (2000) The role of positively charged amino acids in ATP recognition by human P2X1 receptors. *Journal of Biological Chemistry* 275 (38):29361-29367

**Table 1.** Details of compounds used in this study: (A) Main structure, (B) Training Set, (C) Test Set.

(A)

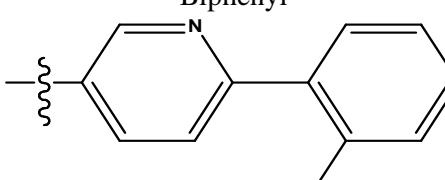
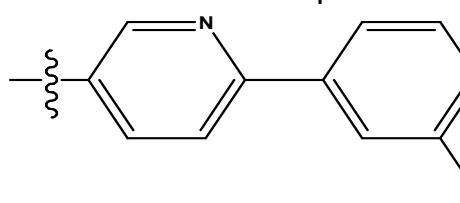
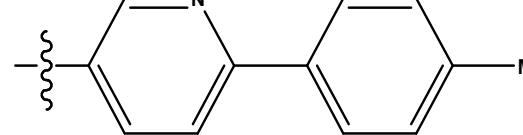
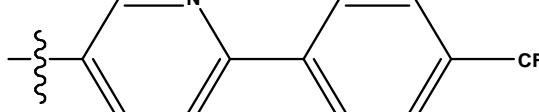
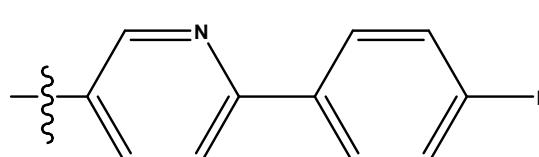
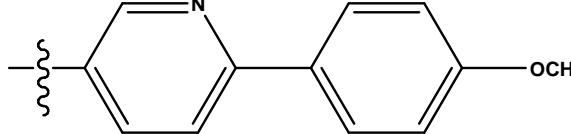
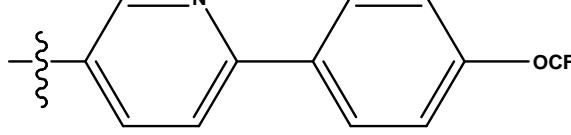


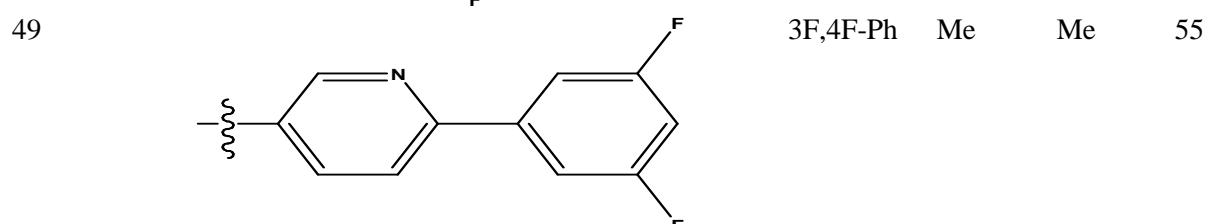
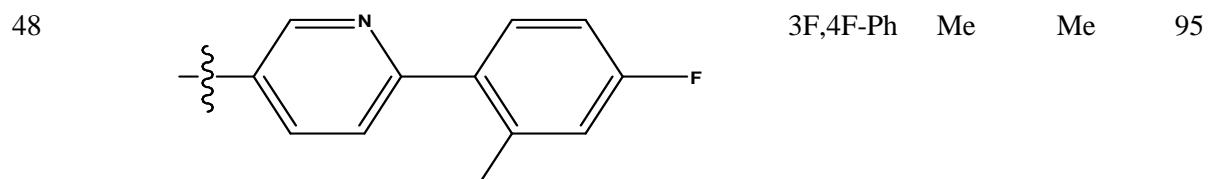
(B)

Compound	Substituent's				IC <sub>50</sub> (nM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	
1	4F-Ph	Ph	H	Isopropyl	2752
2	4F-Ph	3F,4F-Ph	H	Isopropyl	295
3	4F-Ph	2F,4F-Ph	H	Isopropyl	893
4	4F-Ph	4F-Ph	H	Isopropyl	614
5	4F-Ph	3F,5F-Ph	H	Isopropyl	454
6	4F-Ph	3,4,5-TriF-Ph	H	Isopropyl	278
7	4F-Ph	3F,4F-Ph	H	tButyl	614
8	4F-Ph	3F,4F-Ph	H		248
9	4F-Ph	3F,4F-Ph	H		334
10	4F-Ph	3F,4F-Ph	Me	Me	368
11	4-CF <sub>3</sub> OPh	3F,4f-Ph	H	Isopropyl	223
12	3-CF <sub>3</sub> OPh	3F,4F-Ph	H	Isopropyl	279
13	4-CN-Ph	3F,4F-Ph	H	Isopropyl	223
14	4-CN,3F-Ph	3F,4F-Ph	H	Isopropyl	163
15		3F,4F-Ph	H	Isopropyl	241
16		3F,4F-Ph	H	Isopropyl	147
17		3F,4F-Ph	H	Isopropyl	82

18		3F,4F-Ph	H	Isopropyl	114
19		3F,4F-Ph	H	Isopropyl	92
20		3F,4F-Ph	H	Isopropyl	56
21		3F,4F-Ph	H	Isopropyl	53
22		3F,4F-Ph	H	Isopropyl	38
23 24 36	Biphenyl Biphenyl 	2F-Ph 3F-Ph 3F,4F-Ph	H H Me	Isopropyl Isopropyl Me	937 269 24
38		3F,4F-Ph	Me	Me	101
39		3F,4F-Ph	Me	Me	25
40		3F,4F-Ph	Me	Me	104
42		3F,4F-Ph	Me	Me	110

(C)

Compound	Substituent's				IC <sub>50</sub> (nM)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	
25	Biphenyl	2F-Ph	H	Isopropyl	237
26	Biphenyl	2F,3F-Ph	H	Isopropyl	870
27	Biphenyl	2F,4F-Ph	H	Isopropyl	845
28	Biphenyl	2F,5F-Ph	H	Isopropyl	1036
29	Biphenyl	3F,5F-Ph	H	Isopropyl	107
30	Biphenyl	3F,4F,Ph	H	Isopropyl	178
31	Biphenyl	3F,4F-Ph	Me	Me	85
32	Biphenyl	3F,4F-Ph	-	Cyclohexyl	388
33	Biphenyl	4F-Ph	Me	Me	276
34	Biphenyl	Biphenyl	Me	Me	700
35	Biphenyl	Biphenyl	Me	Me	91
37	Biphenyl	3F,4F-Ph	Me	Me	52
					
41		3F,4F-Ph	Me	Me	28
43		3F,4F-Ph	Me	Me	39
44		3F,4F-Ph	Me	Me	137
45		3F,4F-Ph	Me	Me	45
46		3F,4F-Ph	Me	Me	50
47		3F,4F-Ph	Me	Me	1944



**Table 2.** Results of the top 10 pharmacophore hypotheses generated by the *HypoGen* algorithm.

Hypothesis	Total cost	Cost difference <sup>a</sup>	RMSD	Error cost	Correlation	Features <sup>b</sup>
1	186.500	368.677	2.876	165.451	0.874	HY-A, HY-A, HY, RA
2	190.975	364.202	2.930	170.023	0.869	HBA, HBA-L, HY-A
3	218.890	336.287	3.236	197.310	0.837	HBA-L, HY-A, HY, RA
4	219.702	335.475	3.220	195.837	0.839	HBA-L, HBA-L, HBA-L, HY
5	219.832	335.345	3.220	195.814	0.839	HBA, HY-A, HY, RA
6	219.930	335.247	3.213	195.216	0.840	HBA, HBA-L, HBA-L, HY
7	221.596	333.581	3.247	198.381	0.836	HBA, HBA-L, HBA-L, HY-A
8	221.910	333.267	3.168	191.001	0.845	HBA, HBA-L, HY, RA
9	222.003	333.174	3.268	200.328	0.834	HBA-L, HY, RA
10	223.224	331.953	3.235	197.294	0.837	HBA, HBA, HY, RA

Null cost = 555.177; fixed cost = 66.403; configuration cost = 19.818.

<sup>a</sup> Cost difference = null cost - total cost.

<sup>b</sup> Abbreviations used for features: HBA, hydrogen bond acceptor; HBA-L, hydrogen bond acceptor lipid; HY, hydrophobic; HY-A, hydrophobic aromatic; RA, ring aromatic.

**Table 3.** Experimental and predicted IC<sub>50</sub> values of the training set compounds against Hypo1.

Compound	IC <sub>50</sub> (nM)		Error <sup>a</sup>	Fit value <sup>b</sup>	Activity Scale <sup>c</sup>	
	Experimental	Predicted			Experimental	Predicted
1	2752	652	-4.21	6.08	+	++
2	295	395	1.33	6.30	++	++
3	893	599	-1.49	6.11	++	++
4	614	396	-1.54	6.29	++	++
5	454	472	1.04	6.22	++	++
6	278	320	1.15	6.39	++	++
7	614	479	-1.27	6.21	++	++
8	248	575	2.32	6.13	++	++
9	334	313	-1.06	6.40	++	++
10	368	246	-1.49	6.50	++	++
11	223	201	-1.10	6.59	++	++
12	279	292	1.04	6.43	++	++
13	223	205	-1.08	6.58	++	++
14	163	240	1.47	6.51	+++	++
15	241	632	2.62	6.09	++	++
16	147	258	1.75	6.48	+++	++
17	82	108	1.32	6.86	+++	+++
18	114	120	1.05	6.81	+++	+++
19	92	115	1.25	6.83	+++	+++
20	56	36	-1.55	7.34	+++	++++
21	53	51	-1.02	7.80	+++	+++
22	38	70	1.84	7.05	++++	+++
23	937	316	-2.95	6.39	++	++
24	269	328	1.22	6.38	++	++
36	24	47	1.98	7.21	++++	++++
38	101	106	1.05	6.86	+++	+++
39	25	21	-1.19	7.57	++++	++++
40	104	50	-2.06	7.19	+++	++++
42	110	96	-1.14	6.91	+++	+++

<sup>a</sup>Positive value indicates that the predicted IC<sub>50</sub> is higher than the experimental IC<sub>50</sub>; negative value indicates that the predicted IC<sub>50</sub> is lower than the experimental IC<sub>50</sub>.

<sup>b</sup> Fit value indicates how well the features in the pharmacophore map the chemical features in the compound.

<sup>c</sup> Activity scale: active, +++; moderately active, ++ (50 < IC<sub>50</sub> ≤ 200 nM); less active, + (200 < IC<sub>50</sub> ≤ 1000 nM); poor active, + (IC<sub>50</sub> > 1000 nM).

**Table 4.** Experimental and predicted IC<sub>50</sub> data values of the test set compounds against Hypo1.

Compound	IC <sub>50</sub> (nM)		Error	Fit value	Activity Scale	
	Experimental	Predicted			Experimental	Predicted
25	237	134	-1.76	6.76	++	+++
26	870	167	-5.20	6.67	++	+++
27	845	106	-7.97	6.86	++	+++
28	1036	192	-5.39	6.61	+	+++
29	107	147	1.37	6.72	+++	+++
30	178	141	-1.26	6.74	+++	+++
31	85	148	1.74	6.72	+++	+++
32	388	138	-2.81	6.75	++	+++
33	276	154	-1.79	7.70	++	+++
34	700	121	-5.78	6.81	++	+++
35	91	169	1.85	6.66	+++	+++
37	52	41	-1.26	7.27	+++	++++
41	28	42	1.50	7.26	++++	++++
43	39	48	1.23	7.21	++++	++++
44	137	56	-2.44	7.14	+++	+++
45	45	47	1.04	7.22	++++	++++
46	50	55	1.10	7.15	++++	+++
47	1944	53	-36.67	7.17	+	+++
48	95	30	-3.16	7.35	+++	++++
49	55	39	-1.41	7.29	+++	++++

**Table 5.** Statistical parameters of GH score validation for Hypo1.

Serial No.	Parameters	Results
1	Total number of molecules in database (D)	1676
2	Total number of actives in database (A)	44
3	Total number of hit molecules from the hit database (Ht)	45
4	Total number of active molecules in hit list (Ha)	36
5	% Yield of actives (Ha/Ht) 100	80
6	% Ratio of actives (Ha/A) 100	81.8
7	Enrichment Factor (EF)	30.47
8	FALSE negative [A-Ha]	8
9	FALSE positives [Ht-Ha]	9
10	Goodness of hit (GH)*	0.86

$$*[(\text{Ha}/4\text{HtA})(3\text{A} + \text{Ht}) \times (1 - (\text{Ht} - \text{Ha})/(\text{D} - \text{A}))]$$

**Table 6.** Results for the calculated Lipinski's rule of five.

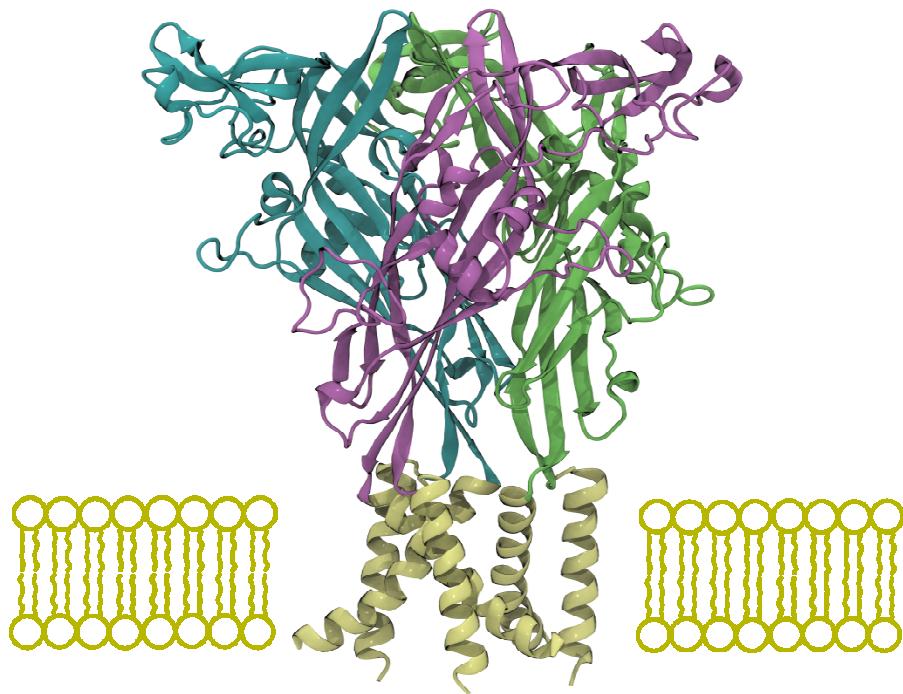
Molecule code	logP	Mw	Number of hydrogen bond acceptors	Number of hydrogen bond donors
Asinex 326689	3.39	491.582	9	3
Asinex 326670	2.56	447.483	8	2
Asinex 243006	4.39	428.480	6	0
Asinex 157767	4.05	463.526	8	3
Asinex 93637	4.89	484.609	5	1
Asinex 470117	4.08	389.469	5	0
Asinex 46912	3.61	375.467	6	0
Asinex 251362	4.61	486.585	7	1
Asinex 157847	4.38	489.520	9	3
Asinex 260860	3.47	473.547	8	1
Asinex 469708	4.32	404.480	5	0
Asinex 467064	2.37	388.462	6	0
Asinex 276866	3.39	457.521	7	2
Asinex 141160	4.59	470.583	5	1
Asinex 444679	1.82	393.462	7	1
Asinex428356	4.18	437.511	7	1
Maybridge 14936	3.66	351.382	7	1
Maybridge 15177	3.23	493.548	6	1
Maybridge 14858	4.37	385.827	7	1
Maybridge 15178	4.32	459.995	6	1
Maybridge 46837	3.24	481.572	8	1
Maybridge 15238	4.78	467.630	6	1

**Table 7.** Predicted activity is theoretical inhibitory activity of each retrieved compound calculated using Hypo1. Estimated Free Energy of Binding is  $\Delta G$  of each retrieved compound calculated by autodock.

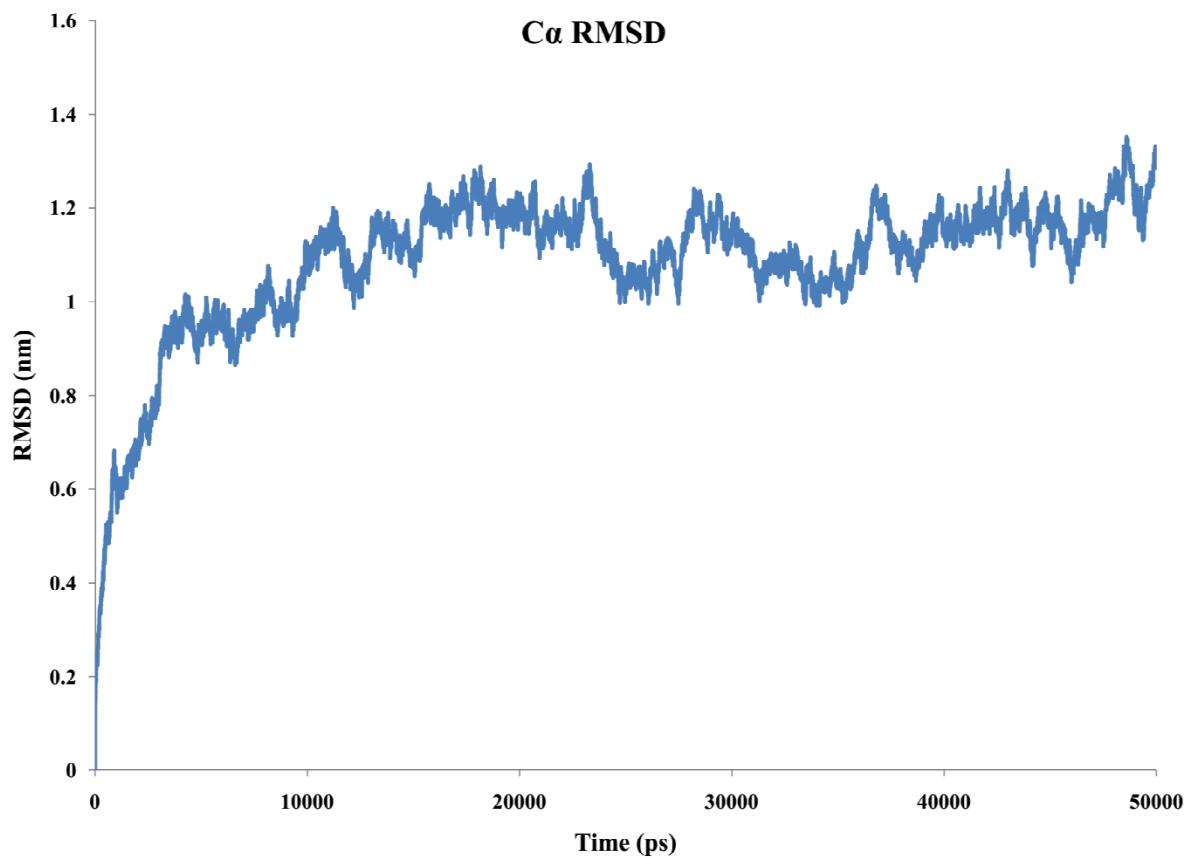
Molecule code	Free Binding Energy (kcal/mol)	Predicted Activity (nM)
Maybridge 15178	-7.58	3.314
Maybridge 15177	-7.01	3.660

Asinex 93637	-6.97	3.390
Asinex 251362	-6.83	3.921
Asinex 260860	-6.82	3.623
Asinex 469708	-6.73	3.709
Maybridge 14858	-6.37	3.777
Asinex 467064	-6.01	3.672
Maybridge 14936	-5.92	3.204
Asinex 469912	-5.85	3.300
Asinex 470117	-5.78	3.866
Asinex 326670	-5.70	3.328
Asinex 428356	-5.69	3.860
Maybridge 46837	-5.59	2.511
Asinex 141160	-5.10	3.720
Asinex 157847	-5.04	3.340
Asinex 444679	-5.00	3.302
Asinex 157767	-4.79	3.102
Maybridge 15238	-4.79	2.969
Asinex 276866	-4.26	3.300
Asinex 243006	-4.24	3.147
Asinex 326689	-3.18	3.204

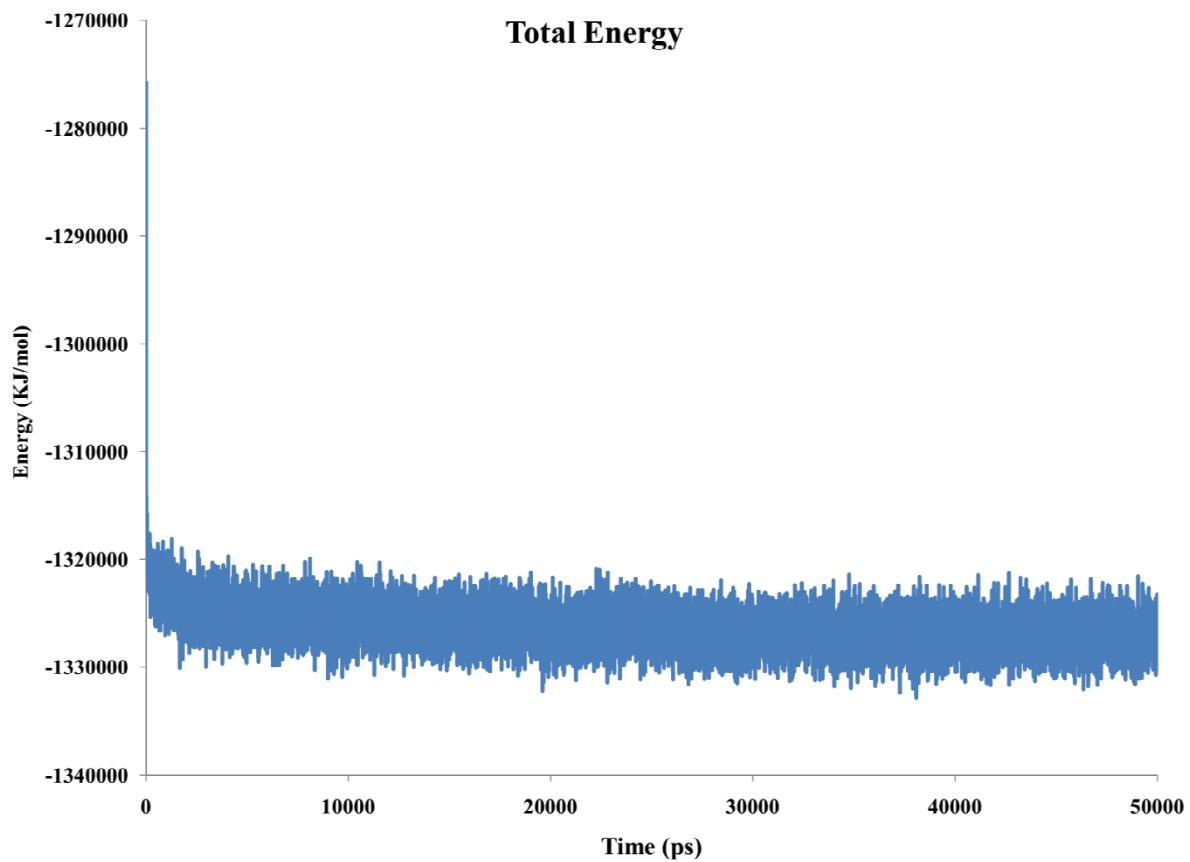
---



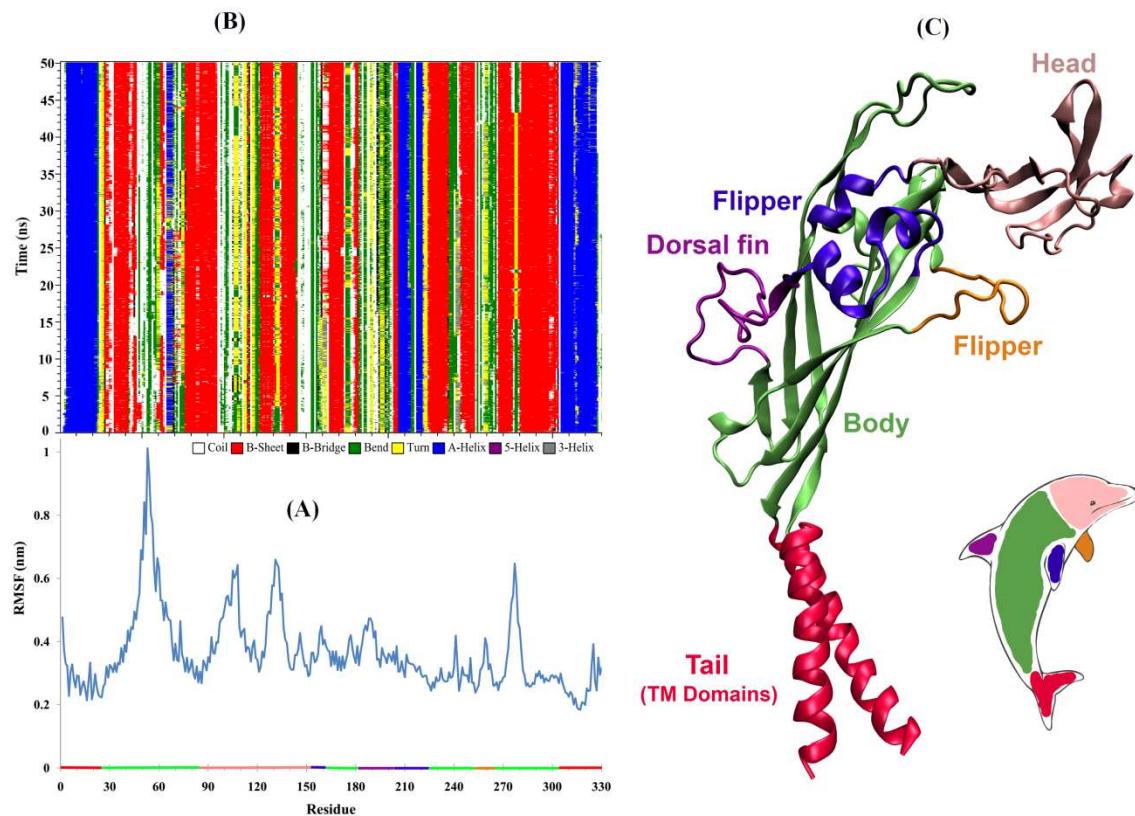
**Figure 1** The architectures of P2X<sub>7</sub>R; Trimer structures viewed parallel to the membrane. Each subunit is shown in a different colour.



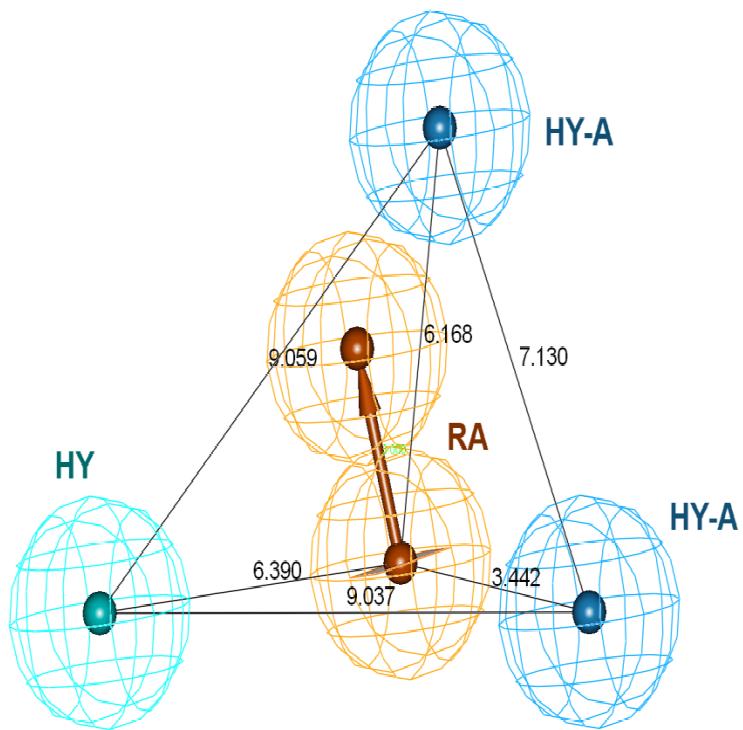
**Figure 2** Graphical representation of  $RMSD$  of  $C_\alpha$  atoms from starting structure of P2X<sub>7</sub>R model as a function of time.



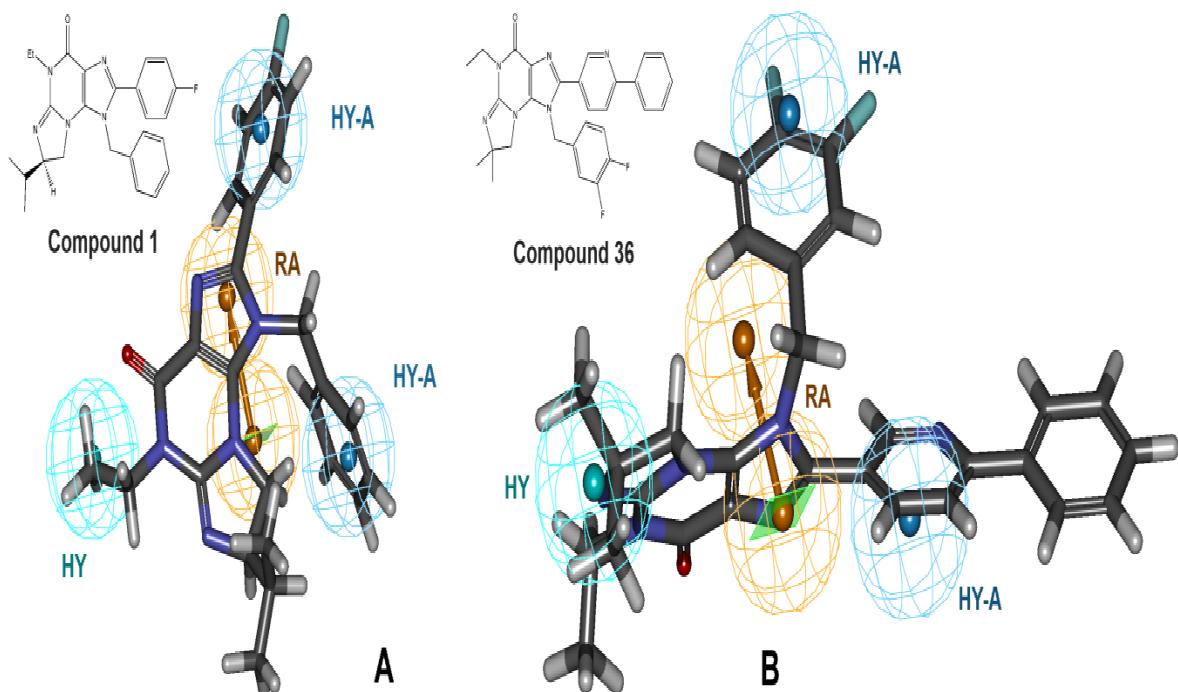
**Figure 3** Graphical representation of energy *vs.* simulation time from starting structure of P2X<sub>7</sub>R model as a function of time



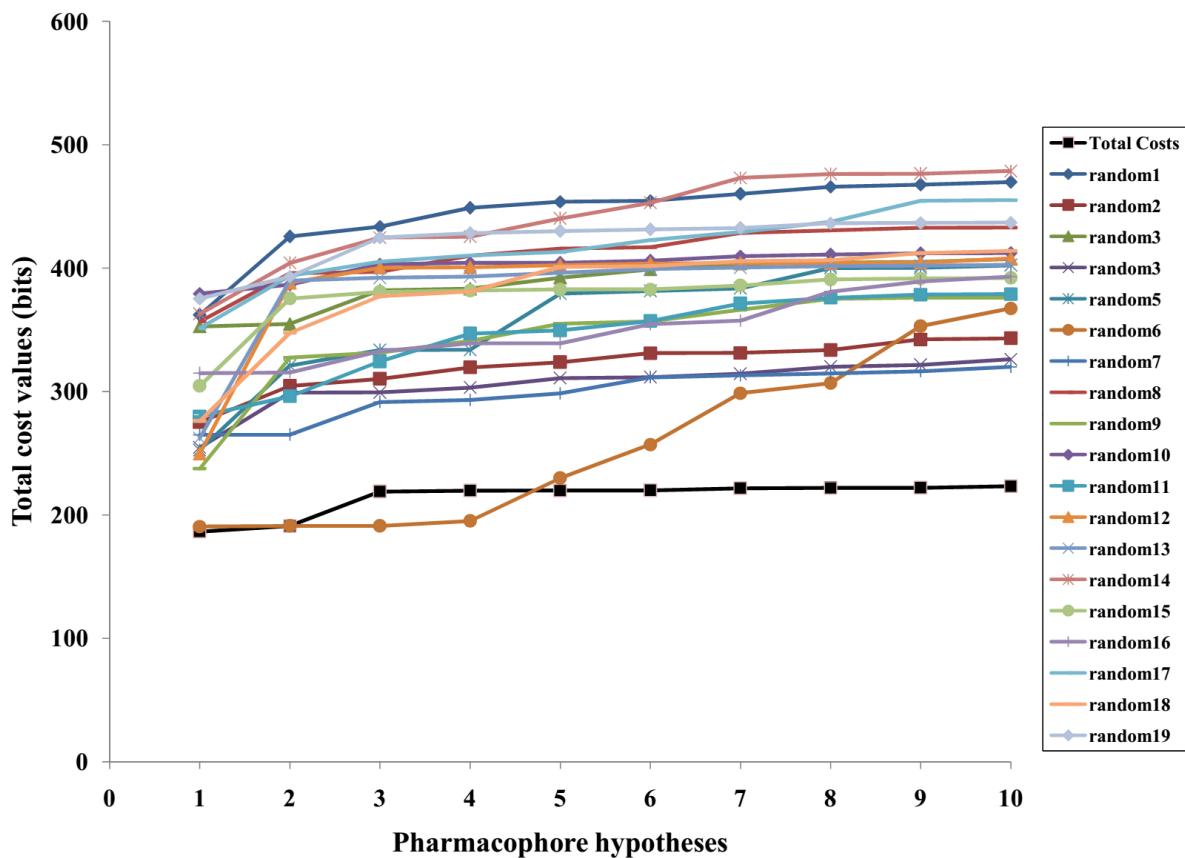
**Figure 4** (A) rms fluctuation (*RMSF*) for a single P2X<sub>7</sub>R subunit during MD simulation and (B) secondary structure per residue along the trajectories of molecular dynamics simulation. (C) A single P2X<sub>7</sub>R subunit, viewed parallel to the membrane plane with the outline dolphin.



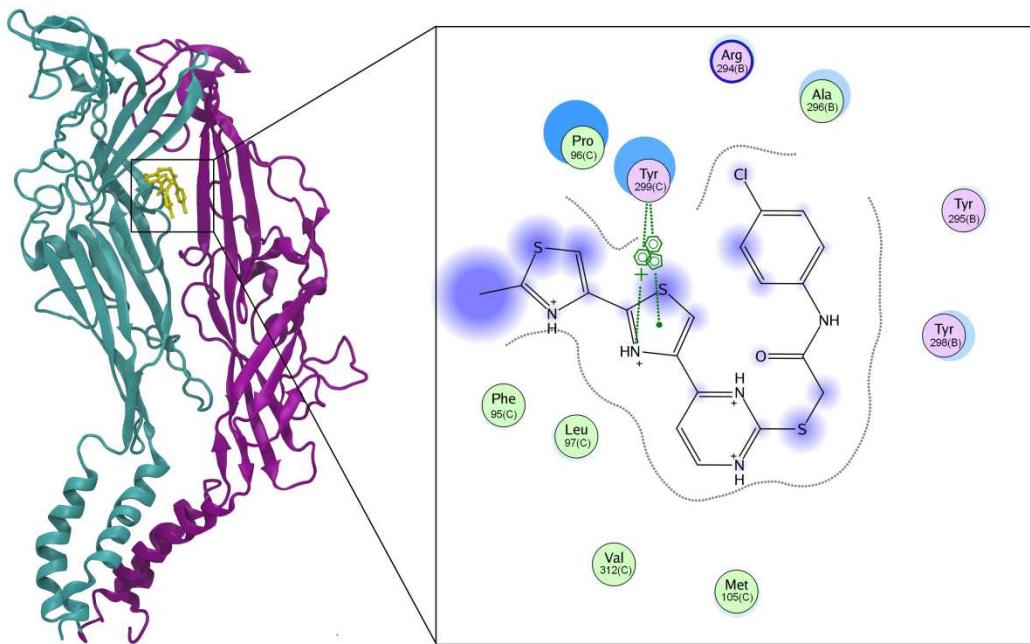
**Figure 5** The best developed pharmacophore hypothesis, Hypo1, showed with distance constraints. Pharmacophore features abbreviated as follows: Hydrophobic (HY), Hydrophobic aromatic (HY-A), Ring aromatic (RA).



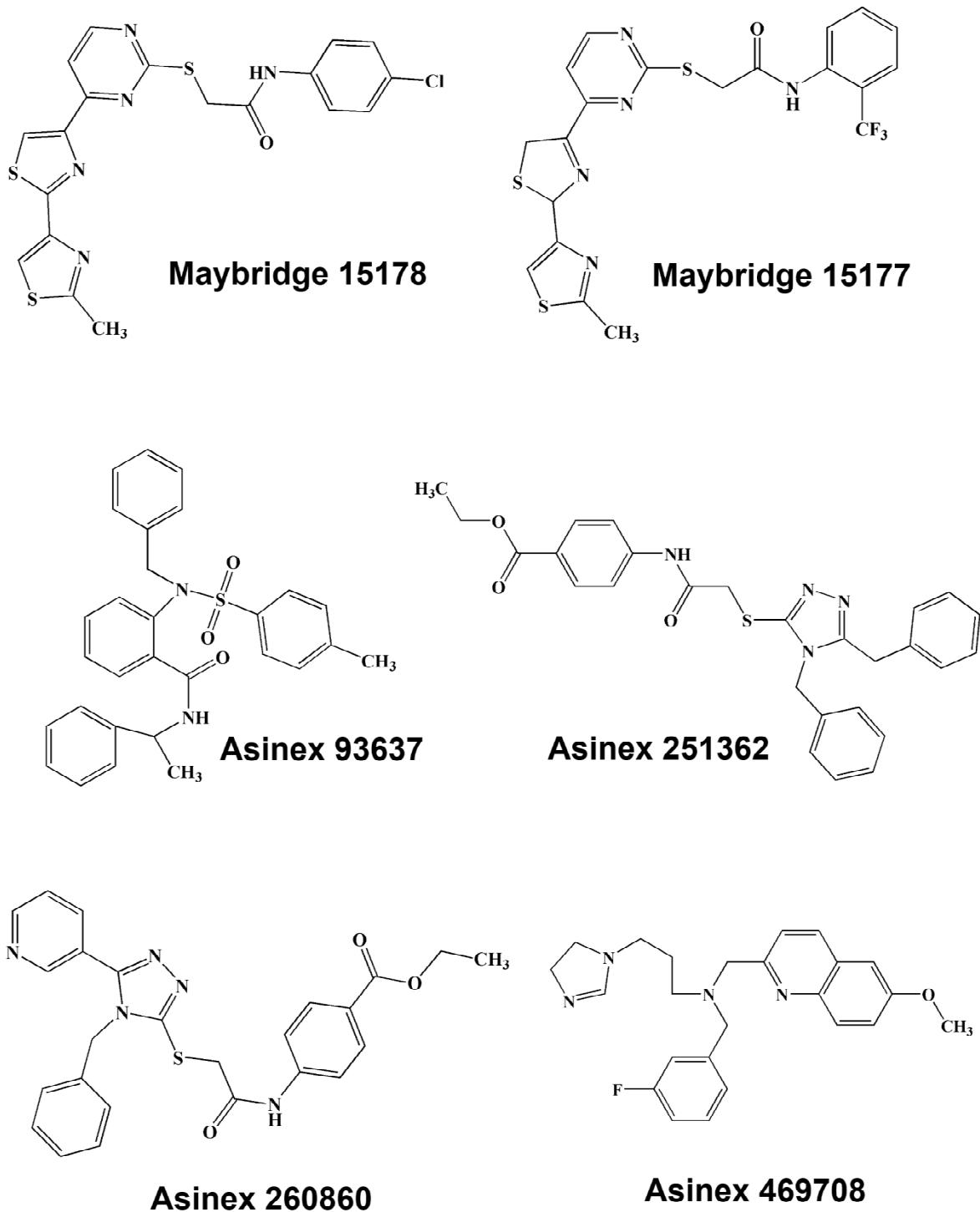
**Figure 6** Overlay of least active (A) and most active (B) molecules in the training set upon the best pharmacophore model Hypo1.



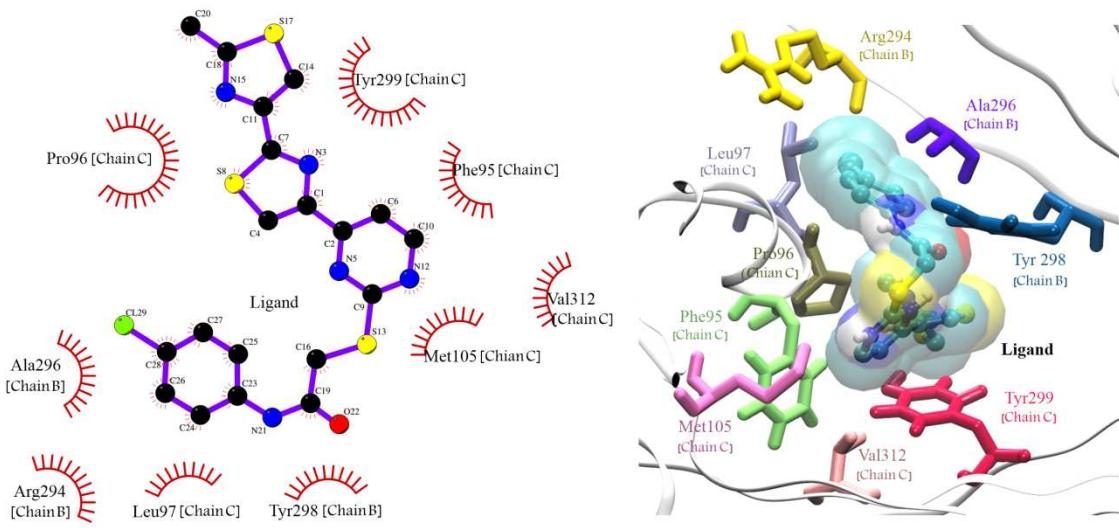
**Figure 7** The difference between cost values of developed pharmacophore model and the scrambled models.



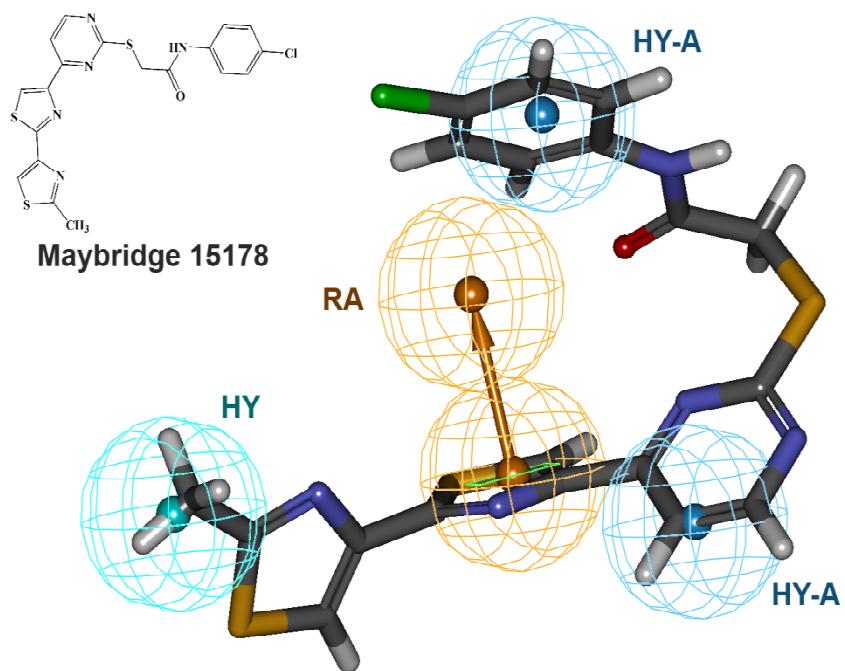
**Figure 8.** P2X<sub>7</sub>R structure and pose of ligand binding. Only two subunits are shown (B is cyan and C is purple) in the left. 2D close-up view of the ligand binding site and important interactions is shown in the right.



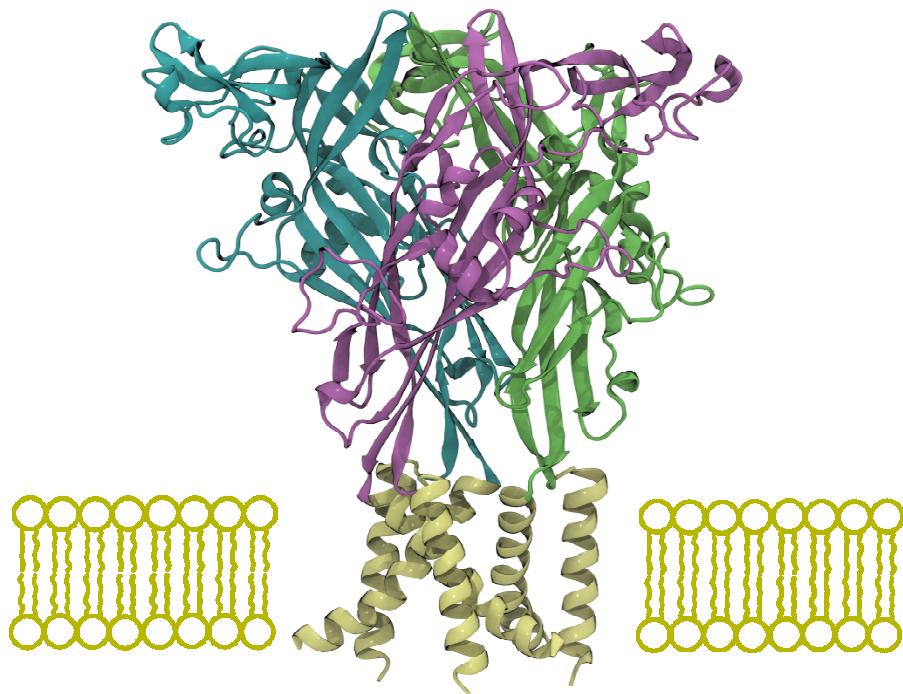
**Figure 9.** Lead molecules retrieved from the database searching as potent P2X<sub>7</sub>R inhibitors.



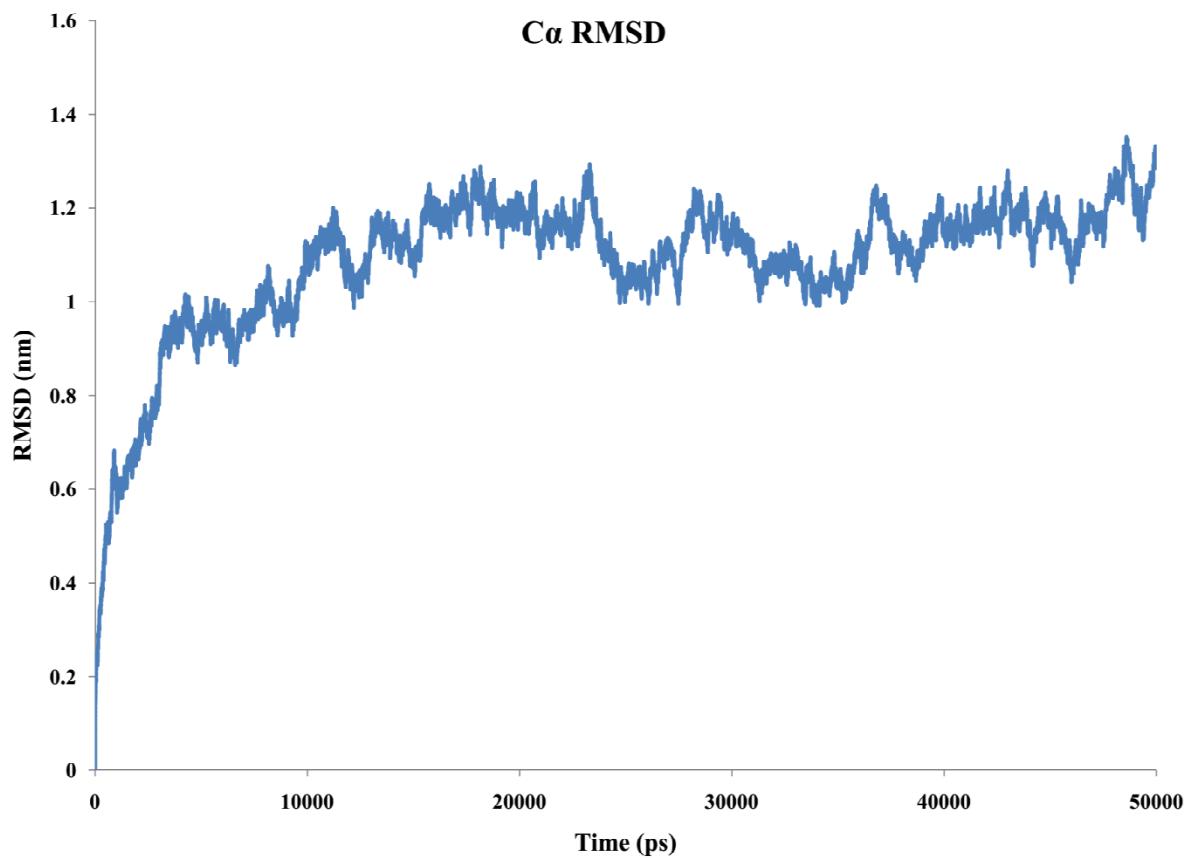
**Figure 10.** Molecular docking results. Docked orientations of compound Maybridge 15178. All active site residues are shown ball and stick form.



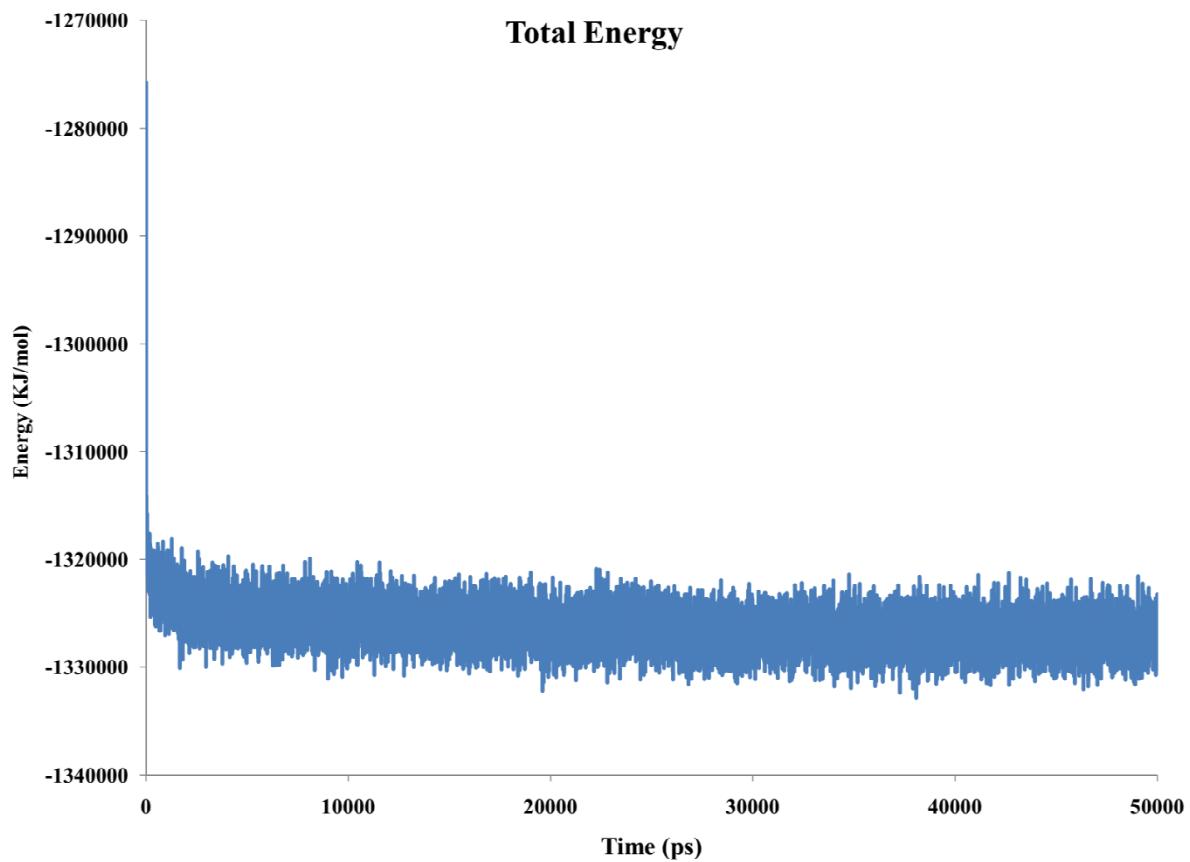
**Figure 11.** The pharmacophore overlay of the compound Maybridge 15178 on the Hypo1.



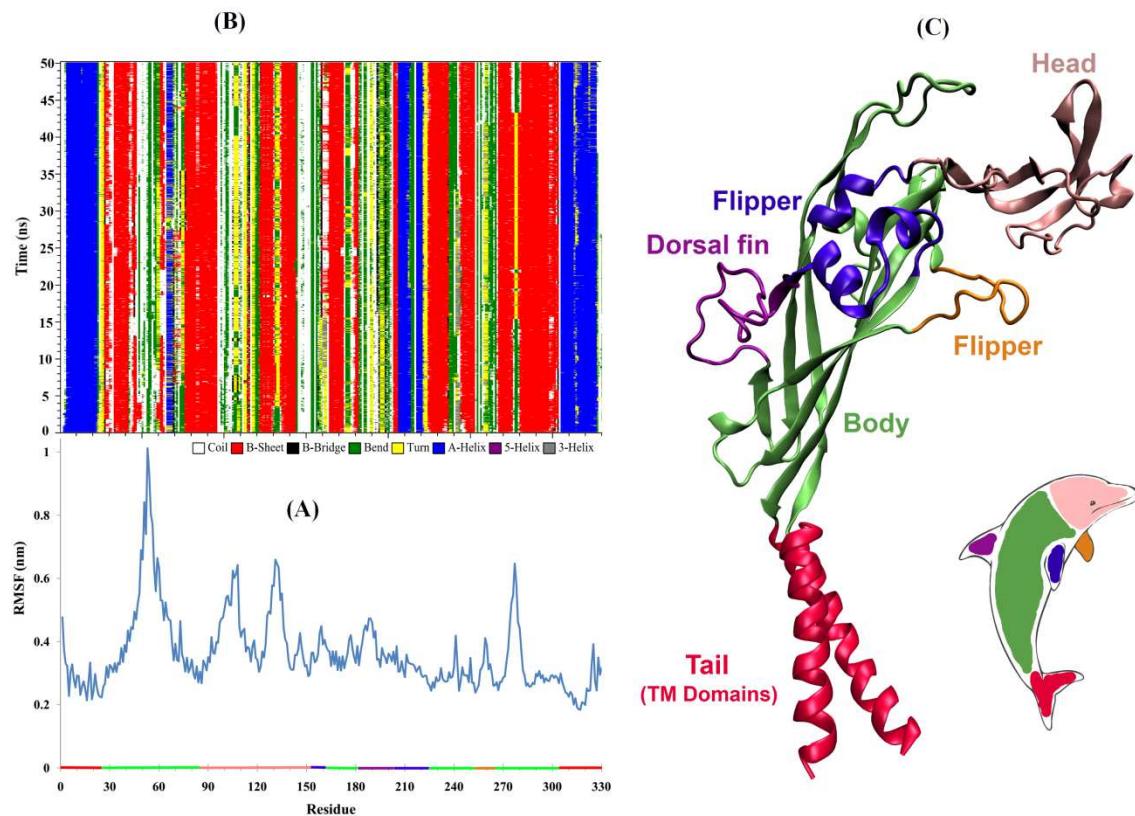
**Figure 1** The architectures of P2X<sub>7</sub>R; Trimer structures viewed parallel to the membrane. Each subunit is shown in a different colour.



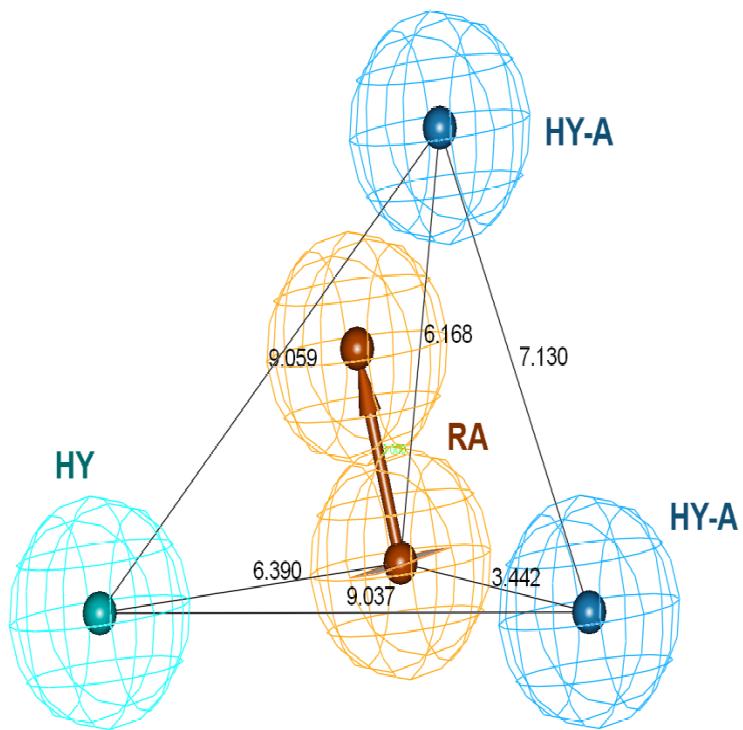
**Figure 2** Graphical representation of *RMSD* of  $C_{\alpha}$  atoms from starting structure of P2X<sub>7</sub>R model as a function of time.



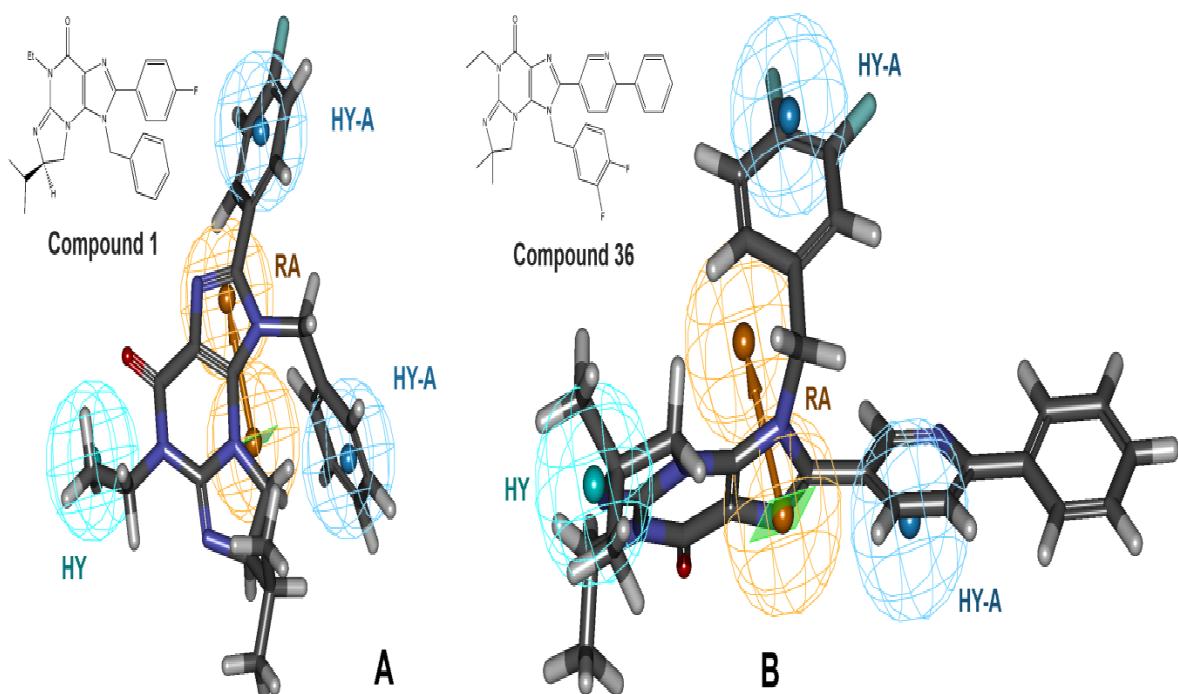
**Figure 3** Graphical representation of energy *vs.* simulation time from starting structure of P2X<sub>7</sub>R model as a function of time



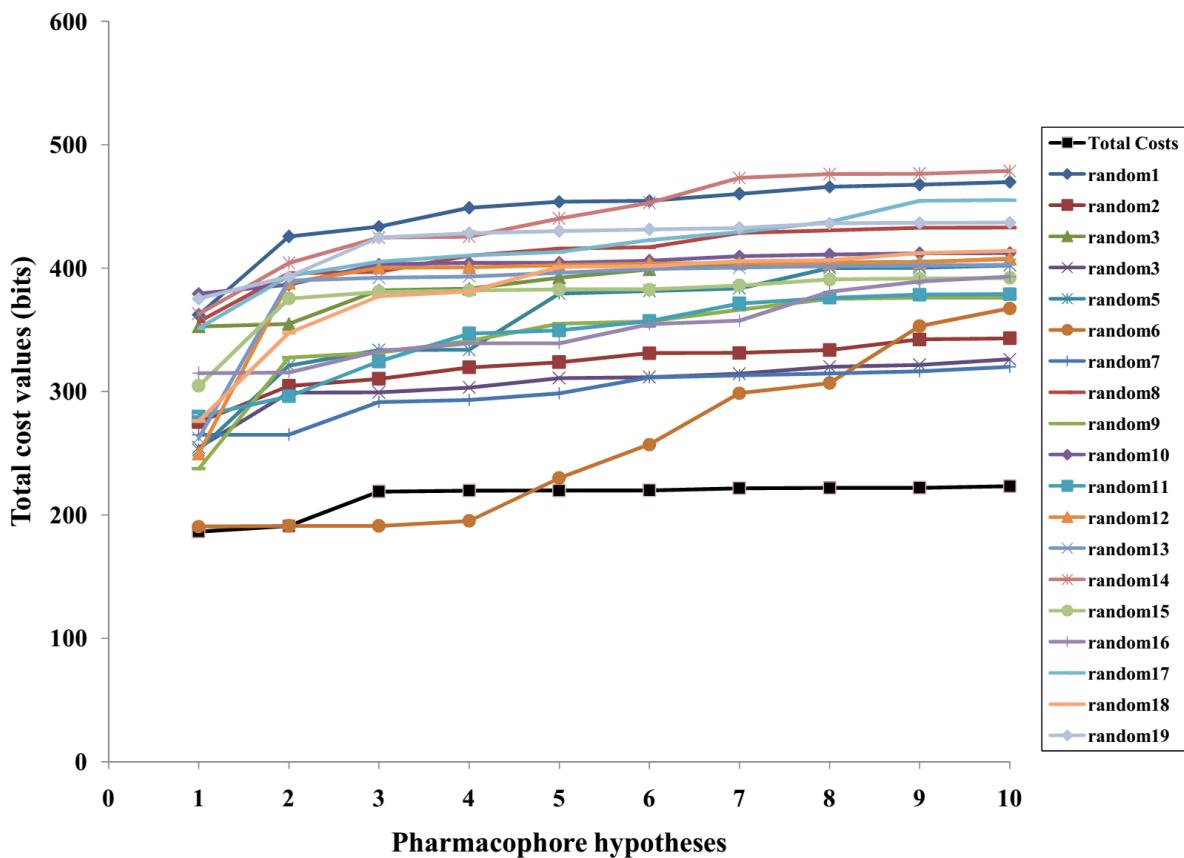
**Figure 4** (A) rms fluctuation (*RMSF*) for a single P2X<sub>7</sub>R subunit during MD simulation and (B) secondary structure per residue along the trajectories of molecular dynamics simulation. (C) A single P2X<sub>7</sub>R subunit, viewed parallel to the membrane plane with the outline dolphin.



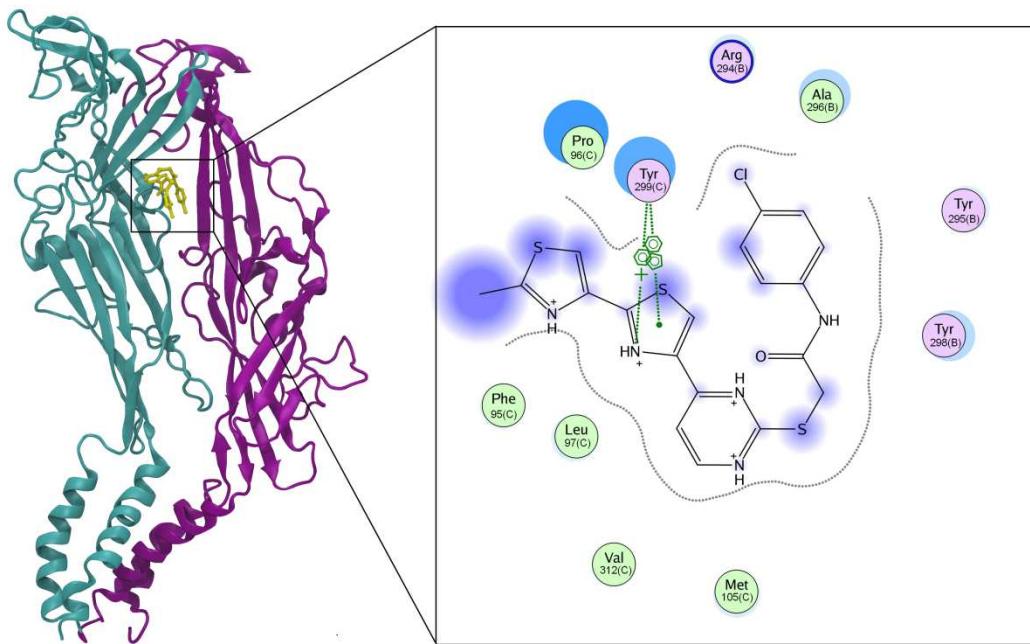
**Figure 5** The best developed pharmacophore hypothesis, Hypo1, showed with distance constraints. Pharmacophore features abbreviated as follows: Hydrophobic (HY), Hydrophobic aromatic (HY-A), Ring aromatic (RA).



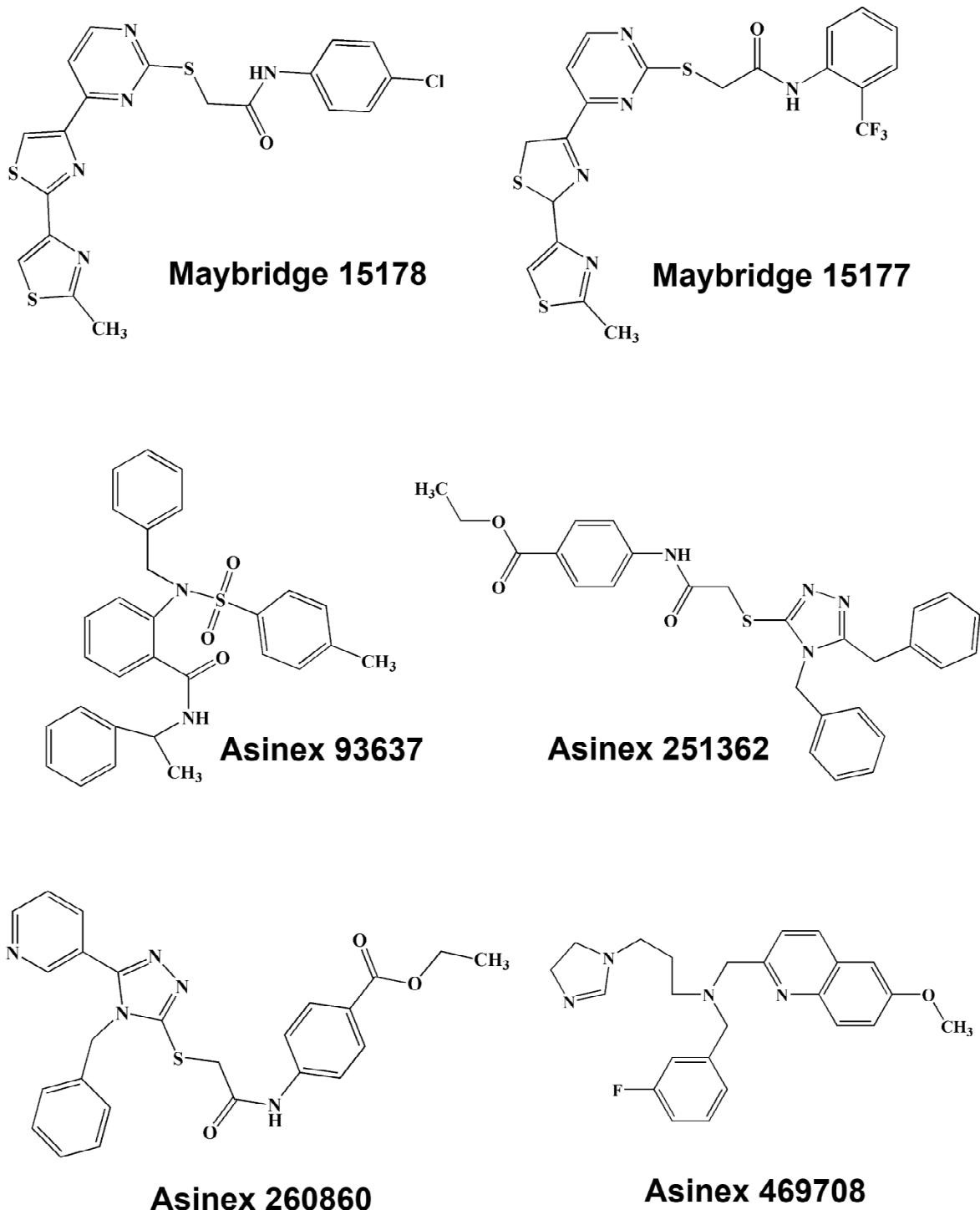
**Figure 6** Overlay of least active (A) and most active (B) molecules in the training set upon the best pharmacophore model Hypo1.



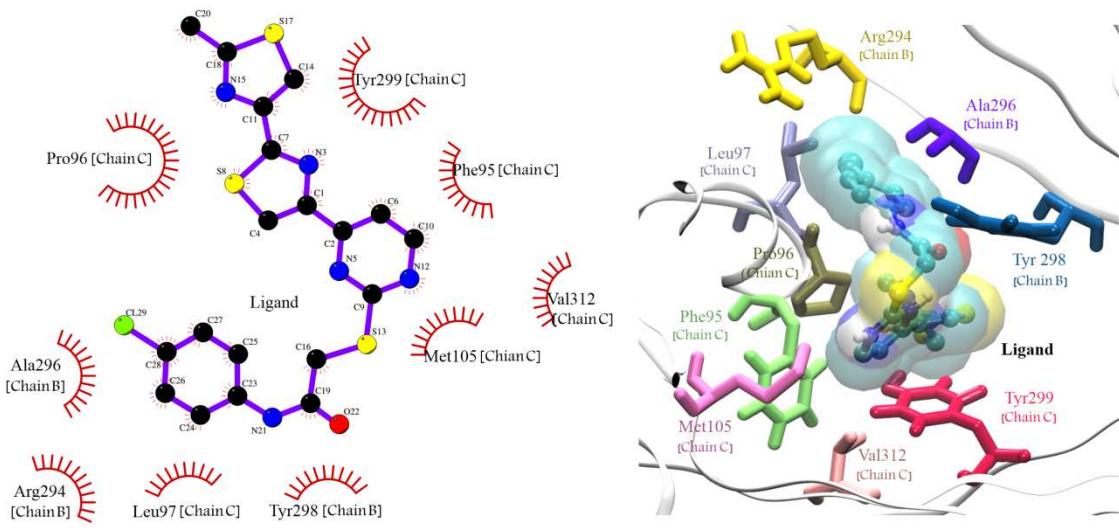
**Figure 7** The difference between cost values of developed pharmacophore model and the scrambled models.



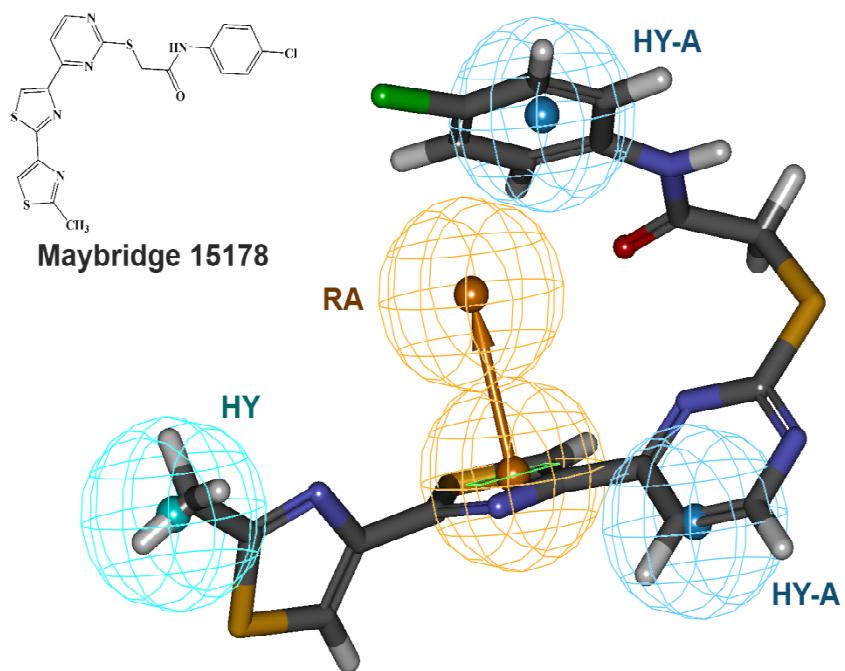
**Figure 8.** P2X<sub>7</sub>R structure and pose of ligand binding. Only two subunits are shown (B is cyan and C is purple) in the left. 2D close-up view of the ligand binding site and important interactions is shown in the right.



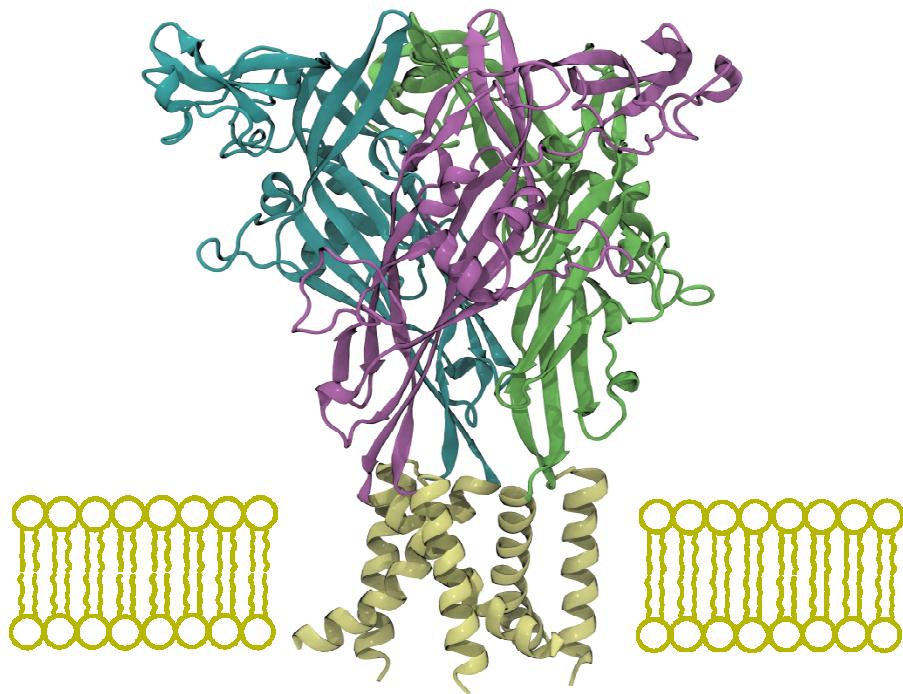
**Figure 9.** Lead molecules retrieved from the database searching as potent P2X<sub>7</sub>R inhibitors.



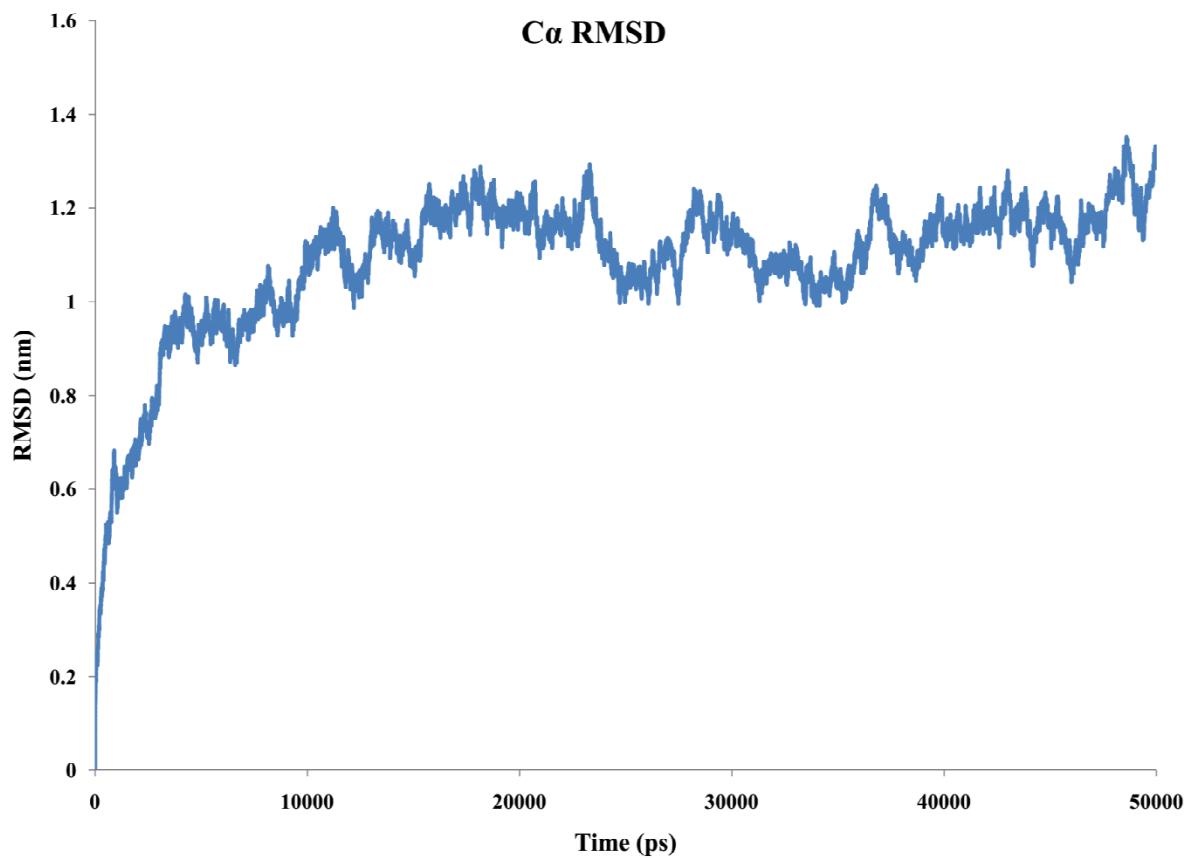
**Figure 10.** Molecular docking results. Docked orientations of compound Maybridge 15178. All active site residues are shown ball and stick form.



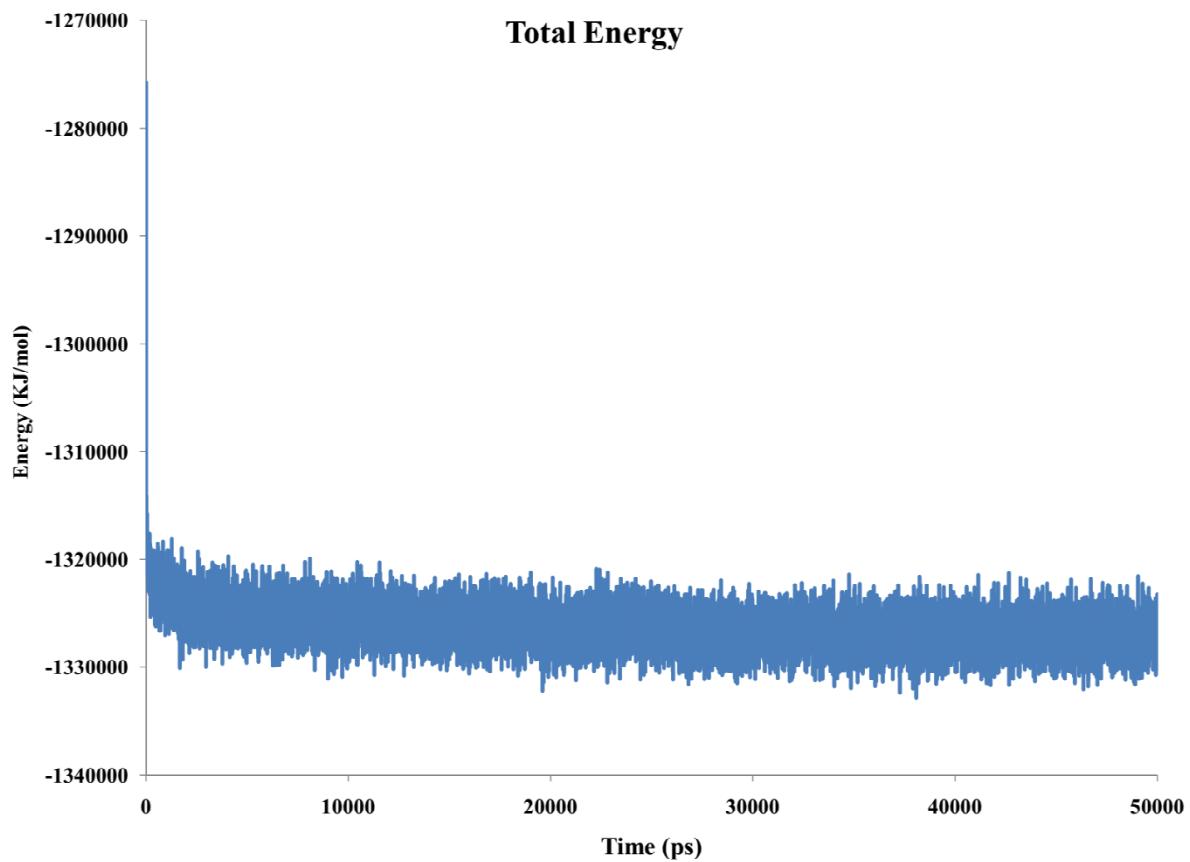
**Figure 11.** The pharmacophore overlay of the compound Maybridge 15178 on the Hypo1.



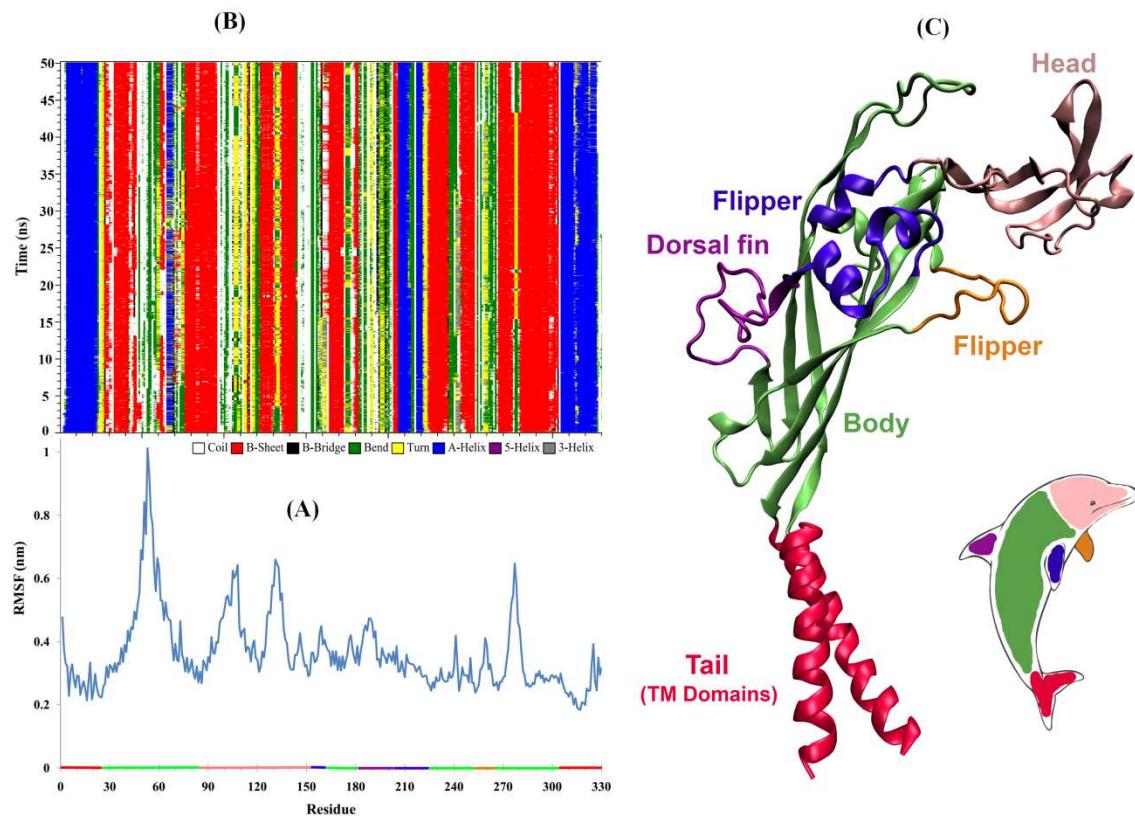
**Figure 1** The architectures of P2X<sub>7</sub>R; Trimer structures viewed parallel to the membrane. Each subunit is shown in a different colour.



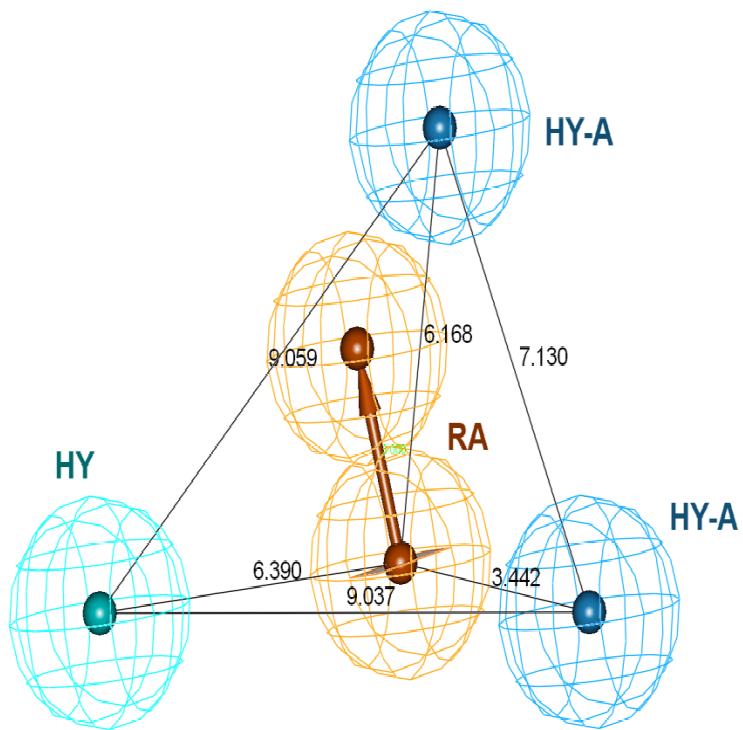
**Figure 2** Graphical representation of  $RMSD$  of  $C_\alpha$  atoms from starting structure of P2X<sub>7</sub>R model as a function of time.



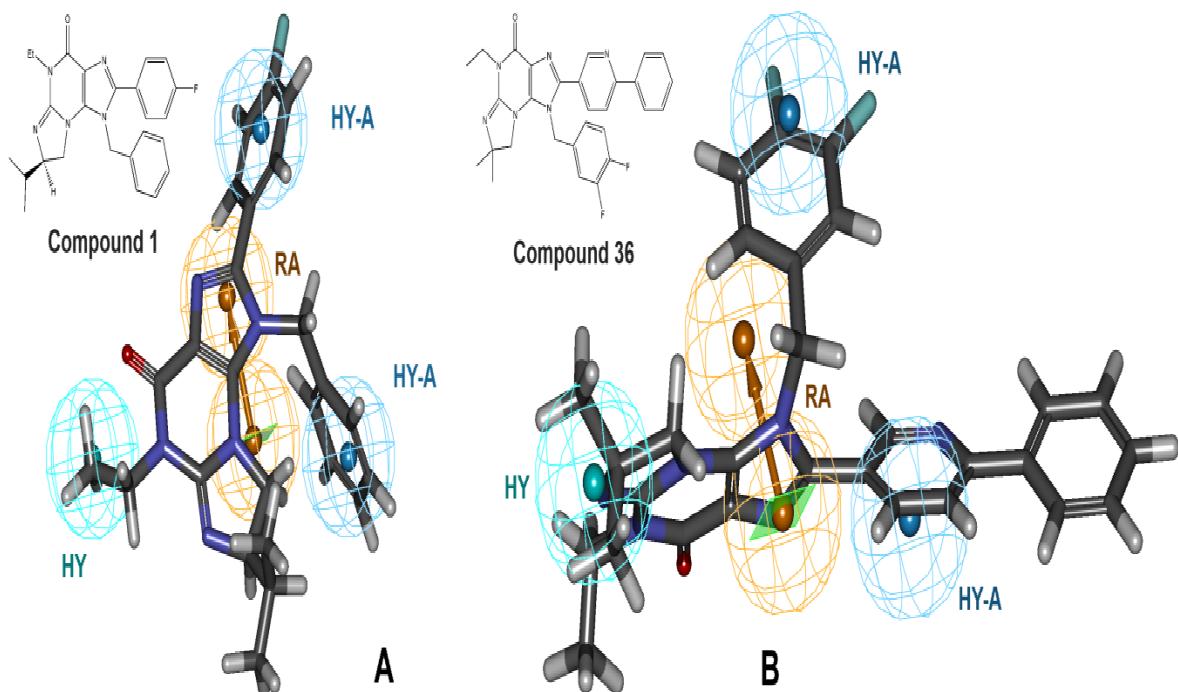
**Figure 3** Graphical representation of energy *vs.* simulation time from starting structure of P2X<sub>7</sub>R model as a function of time



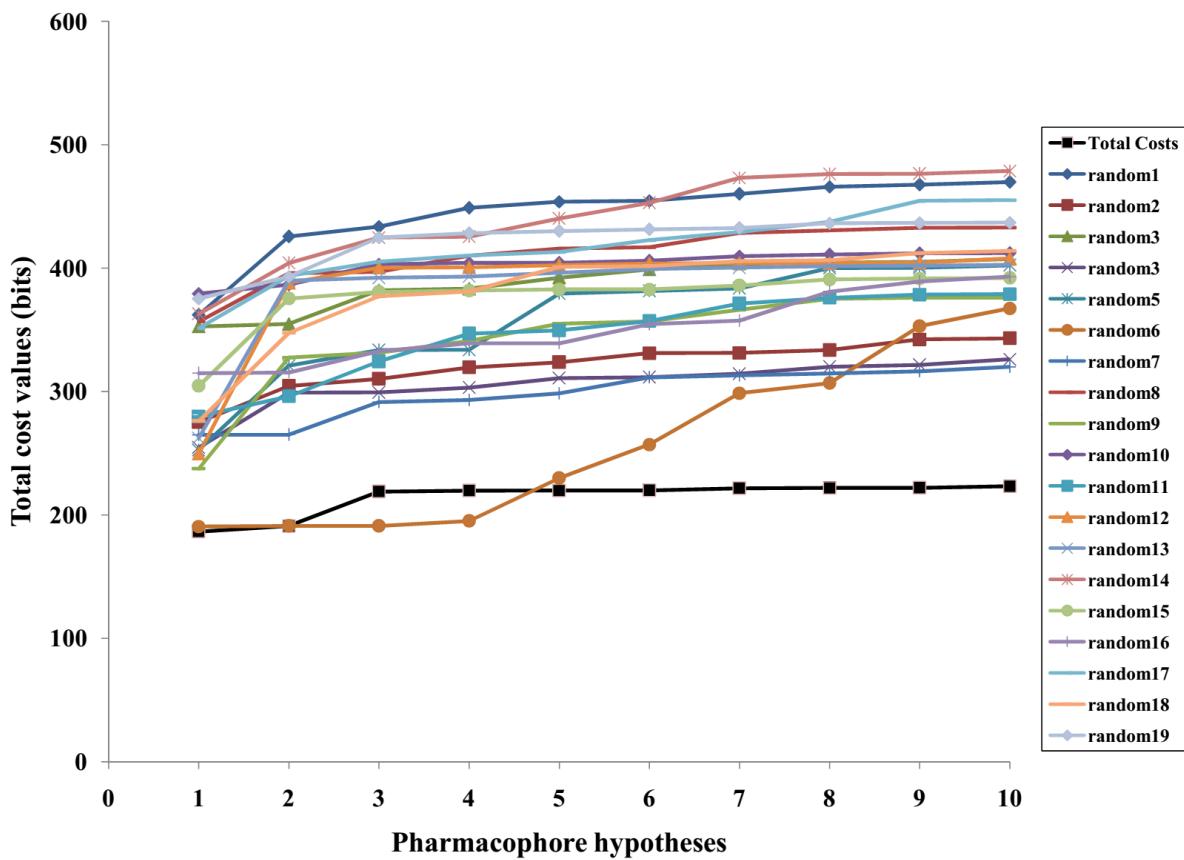
**Figure 4** (A) rms fluctuation (*RMSF*) for a single P2X<sub>7</sub>R subunit during MD simulation and (B) secondary structure per residue along the trajectories of molecular dynamics simulation. (C) A single P2X<sub>7</sub>R subunit, viewed parallel to the membrane plane with the outline dolphin.



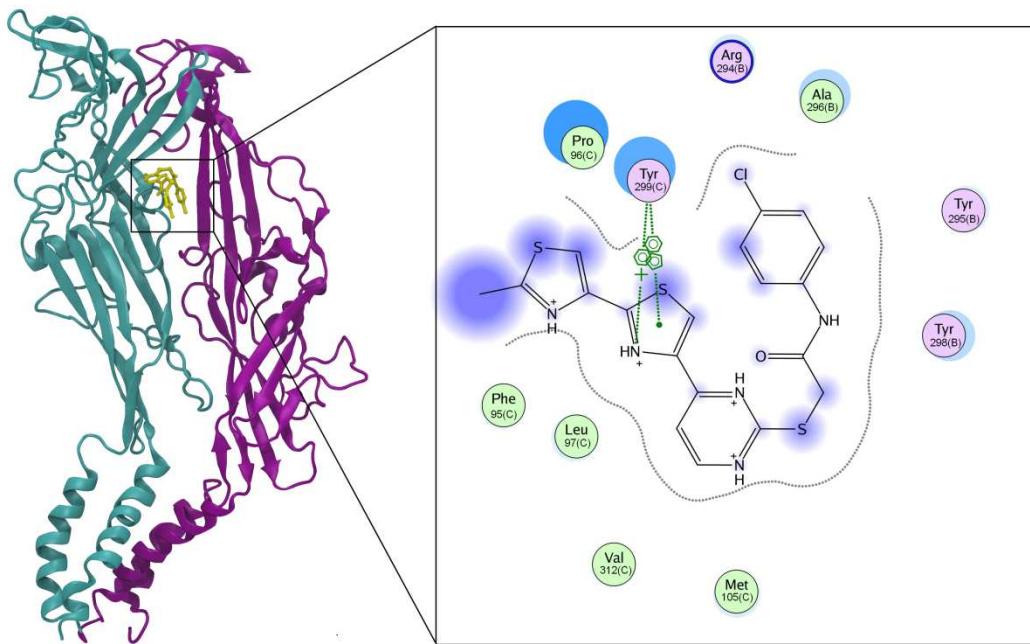
**Figure 5** The best developed pharmacophore hypothesis, Hypo1, showed with distance constraints. Pharmacophore features abbreviated as follows: Hydrophobic (HY), Hydrophobic aromatic (HY-A), Ring aromatic (RA).



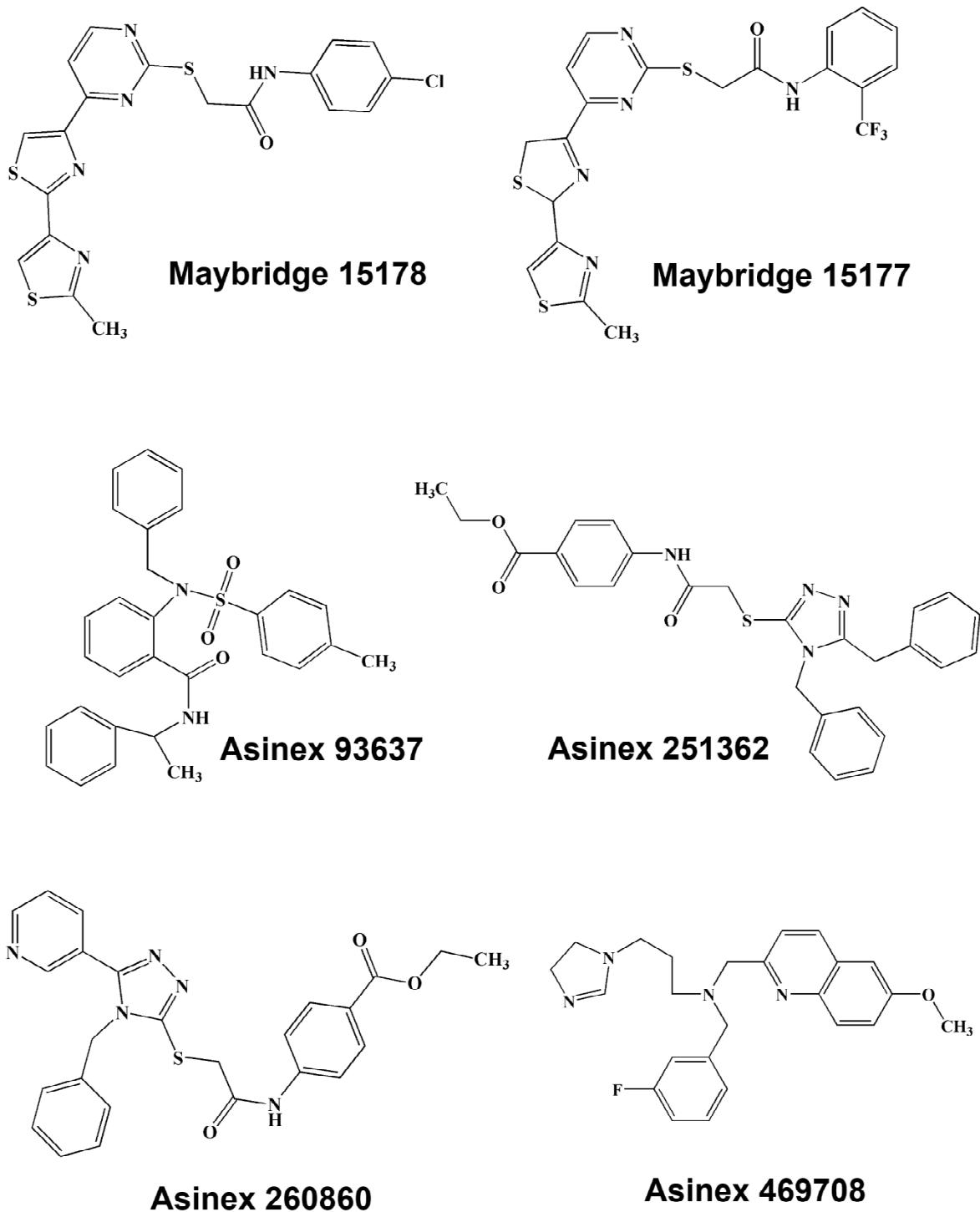
**Figure 6** Overlay of least active (A) and most active (B) molecules in the training set upon the best pharmacophore model Hypo1.



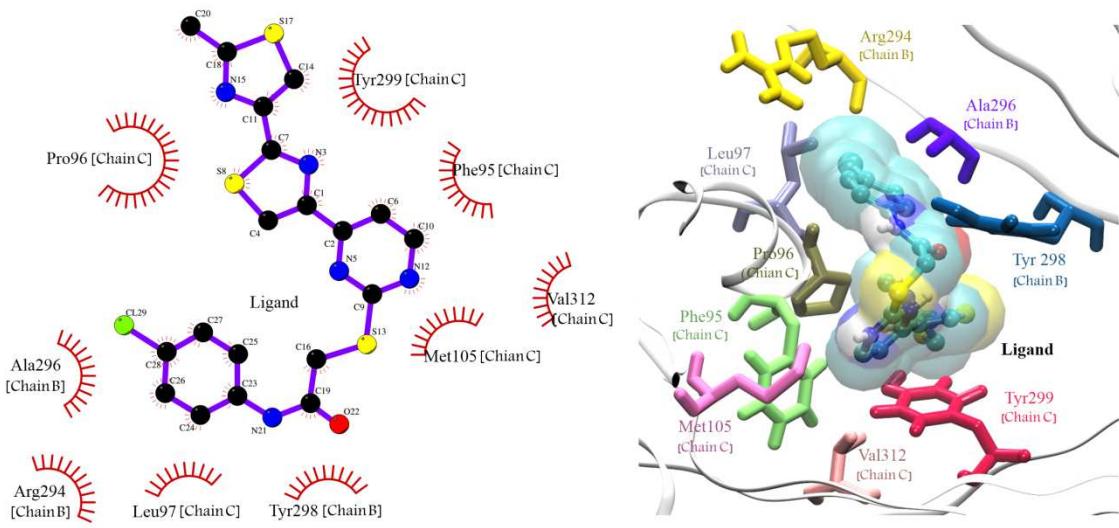
**Figure 7** The difference between cost values of developed pharmacophore model and the scrambled models.



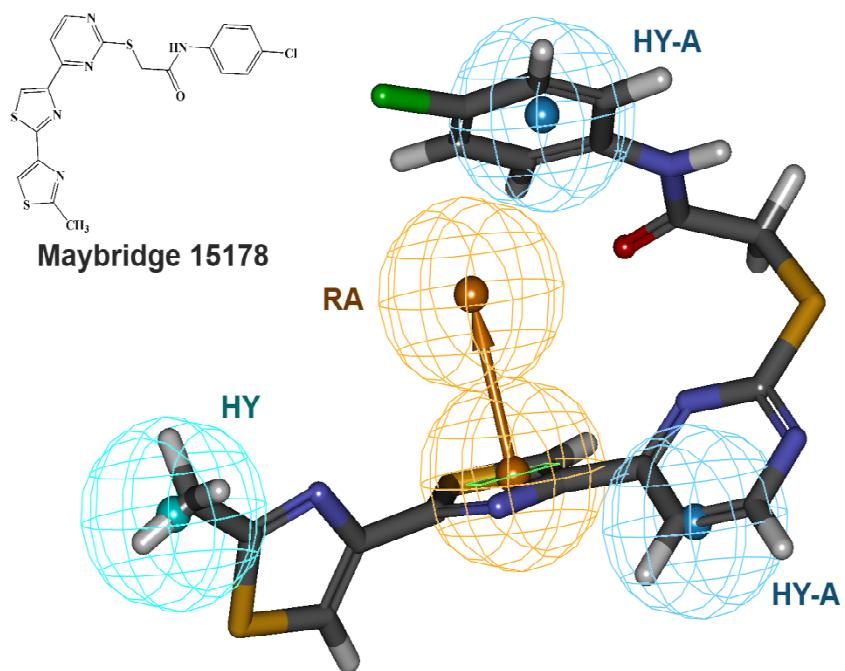
**Figure 8.** P2X<sub>7</sub>R structure and pose of ligand binding. Only two subunits are shown (B is cyan and C is purple) in the left. 2D close-up view of the ligand binding site and important interactions is shown in the right.



**Figure 9.** Lead molecules retrieved from the database searching as potent P2X<sub>7</sub>R inhibitors.



**Figure 10.** Molecular docking results. Docked orientations of compound Maybridge 15178. All active site residues are shown ball and stick form.



**Figure 11.** The pharmacophore overlay of the compound Maybridge 15178 on the Hypo1.