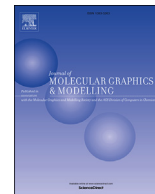




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Alignment independent 3D-QSAR studies and molecular dynamics simulations for the identification of potent and selective S1P<sub>1</sub> receptor agonists

Ali Akbar Alizadeh <sup>a, 1</sup>, Behzad Jafari <sup>a, b, 1</sup>, Siavoush Dastmalchi <sup>a, c, \*</sup>

<sup>a</sup> Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

<sup>b</sup> Department of Medicinal Chemistry, School of Pharmacy, Urmia University of Medical Sciences, Iran

<sup>c</sup> School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran

## ARTICLE INFO

### Article history:

Received 3 June 2019

Received in revised form

14 September 2019

Accepted 26 September 2019

Available online 29 September 2019

### Keywords:

Sphingosine 1-phosphate

Molecular docking

3D-QSAR

Molecular dynamics simulation

Multiple sclerosis

## ABSTRACT

Sphingosine 1-phosphate type 1 (S1P<sub>1</sub>) receptors are expressed on lymphocytes and regulate immune cells trafficking. Sphingosine 1-phosphate and its analogues cause internalization and degradation of S1P<sub>1</sub> receptors, preventing the auto reactivity of immune cells in the target tissues. It has been shown that S1P<sub>1</sub> receptor agonists such as fingolimod can be suitable candidates for treatment of autoimmune diseases. The current study aimed to generate GRIND-based 3D-QSAR predictive models for agonistic activities of 2-imino-thiazolidin-4-one derivatives on S1P<sub>1</sub> to be used in virtual screening of chemical libraries. The developed model for the S1P<sub>1</sub> receptor agonists showed appropriate power of predictivity in internal ( $r^2_{acc}$  0.93 and SDEC 0.18) and external ( $r^2$  0.75 and MAE (95% data), 0.28) validations. The generated model revealed the importance of variables DRY-N1 and DRY-O in the potency and selectivity of these compounds towards S1P<sub>1</sub> receptor. To propose potential chemical entities with S1P<sub>1</sub> agonistic activity, PubChem chemicals database was searched and the selected compounds were virtually tested for S1P<sub>1</sub> receptor agonistic activity using the generated models, which resulted in four potential compounds with high potency and selectivity towards S1P<sub>1</sub> receptor. Moreover, the affinities of the identified compounds towards S1P<sub>1</sub> receptor were evaluated using molecular dynamics simulations. The results indicated that the binding energies of the compounds were in the range of  $-39.31$  to  $-46.18$  and  $-3.20$  to  $-9.75$  kcal mol<sup>-1</sup>, calculated by MM-GBSA and MM-PBSA algorithms, respectively. The findings in the current work may be useful for the identification of potent and selective S1P<sub>1</sub> receptor agonists with potential use in diseases such as multiple sclerosis.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

Sphingosine 1-phosphate (S1P), a bioactive signaling molecule, is a lysophospholipid (LPL) that involves in different cellular actions such as vascular growth, development and regulation of immune cells, morphogenesis, and arrangement of cytoskeleton [1]. S1P is a lipid with a particular aminoalcohol structure, known as sphingosine which binds to S1P<sub>1-5</sub> receptors [2]. These receptors are G protein-coupled receptors (GPCRs), in which the N-terminal of the

proteins forms an amphipathic binding pocket by crafting a helical cap hindering ligands to access the binding pocket [3]. However, it is believed that LPLs attain their binding sites through the cell membrane instead of extracellular region [4].

S1P<sub>1</sub> receptors are expressed on lymphocytes, regulating immune cells (i.e.B- and T-cells) trafficking. The gradient concentration of S1P in blood, interstitial fluid, and lymph controls the egress of lymphocytes [5]. S1P and its analogues cause internalization and degradation of the receptors by sustain activation of S1P<sub>1</sub> receptor, preventing the auto reactivity of immune cells in the target tissues [5,6].

Diverse structural moieties have demonstrated S1P<sub>1</sub> receptor agonistic effects which can be potential candidates for being used in the treatment of autoimmune diseases [7,8]. Fingolimod (FTY720) is a S1P<sub>1</sub> receptor modulator that is approved for

\* Corresponding author. School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran.

E-mail addresses: [dastmalchi.s@tbzmed.ac.ir](mailto:dastmalchi.s@tbzmed.ac.ir), [siavoush11@yahoo.com](mailto:siavoush11@yahoo.com) (S. Dastmalchi).

<sup>1</sup> Equally contributed as the first authors.

relapsing forms of multiple sclerosis (MS), an autoimmune disorder affecting the central nervous system [9,10]. After conversion to the phosphate form, fingolimod activates S1P<sub>1</sub>, S1P<sub>3</sub>, S1P<sub>4</sub> and S1P<sub>5</sub> receptors. Although the exact roles for S1P<sub>4</sub> and S1P<sub>5</sub> receptors remain unclear, the adverse effects of fingolimod are attributed to the activation of S1P<sub>3</sub> receptor [11]. Therefore, in different studies several selective S1P<sub>1</sub> receptor agonists have been synthesized in order to retain the therapeutic properties of S1P<sub>1</sub> receptor agonists with the minimum side effects [5,12,13].

Nowadays, three-dimensional quantitative structure-activity relationship (3D-QSAR) methods based on molecular interaction fields (MIFs) approach are accepted as the standard tools in medicinal chemistry for drug design [14]. However, the most rate limiting step in 3D-QSAR studies is the obtaining of appropriate ligand alignment. The Grid-INdependent Descriptors (GRIND) offered an alternative approach in 3D-QSAR analyses in which the alignment of ligands is not required. Instead, the products of interaction energies for each pair of MIFs are used to build the models [15]. There are several examples of successful application of alignment-independent 3D-QSAR methods using GRIND descriptors where the results were comparable to GRID based alignment-dependent studies [16–19].

Bolli et al. have synthesized a series of 2-imino-thiazolidin-4-one derivatives and investigated their agonistic effects on S1P<sub>1</sub> and S1P<sub>3</sub> receptors [7]. Among all synthesized derivatives, compound 61 demonstrated the highest S1P<sub>1</sub> receptor agonistic activity with the minimum agonism on S1P<sub>3</sub>. Therefore, this compound, currently known as ponesomid, was selected for in vivo studies, in which it could significantly reduce circulating lymphocytes in rat. The current study aimed to generate GRIND-based 3D-QSAR predictive models for agonist activities of these 2-imino-thiazolidin-4-one derivatives both on S1P<sub>1</sub> and S1P<sub>3</sub> receptors in order to identify structural features relevant for the selective activation of S1P<sub>1</sub> over S1P<sub>3</sub> receptors. The identified structural features consisted in the developed 3D-QSAR model were used for searching PubChem small molecule database for selecting compounds with selective S1P<sub>1</sub> receptor agonistic activity.

## 2. Methods

### 2.1. Data set preparation

In the current work, a set of 62 S1P<sub>1</sub> receptor agonists, synthesized based on 2-imino-thiazolidin-4-one scaffold, was selected from literature for 3D-QSAR studies [7](Table 1). All activities expressed as EC<sub>50</sub> were assessed using membrane preparations of CHO cells expressing recombinant human S1P<sub>1</sub> and S1P<sub>3</sub> receptors. These values were converted to pEC<sub>50</sub> and used as the dependent variable in this study. HyperChem software (version 8.0.8) was employed for generation of the 3D structures of the ligands where MM<sup>+</sup> force field based on Polack-Ribiere algorithm was used for energy minimization [20]. The structures were further optimized by *semi-empirical* approaching AM1 algorithm [21].

### 2.2. Molecular docking study

The crystal structure of S1P<sub>1</sub> and the model structure of S1P<sub>3</sub> receptors were used for the docking calculations. The crystal structure of S1P<sub>1</sub> receptor in complex with an antagonist was retrieved from protein data bank (PDB code: 3V2Y) and the 3D-structural model of S1P<sub>3</sub> was generated using homology modeling in Swiss Model web server [22] based on S1P<sub>1</sub> receptor structure (PDB code: 3V2W.1.A) as the template. The quality and geometrical features of the obtained model was checked using PROCHECK and MolProbity programs [23–25]. Then, the 3D coordinates of the

antagonist in complex with S1P<sub>1</sub> receptor were used to dock the studied ligands into the S1P<sub>1</sub> receptor crystal structure using GOLD program (version 5.0, CCDC, Cambridge, UK) running under LINUX operating system [26,27]. To determine the ligand binding site on S1P<sub>3</sub>, the modeled S1P<sub>3</sub> receptor structure was aligned on S1P<sub>1</sub> receptor-antagonist complex and a new merged complex was built consisting S1P<sub>3</sub> receptor and the antagonist. Finally, the position of the antagonist in the new complex was used as the guide for determining the binding site in docking calculations. ChemPLP scoring function in GOLD program was used for the docking calculations. Docking of the experimentally co-crystallized antagonist into the binding site of S1P<sub>1</sub> receptor resulted in a solution with 0.51 Å RMSD to its experimentally observed position. Therefore, for all other dockings the same scoring function was used. The best active conformations for the most active compounds in data set (i.e. compounds 61 and 14 for S1P<sub>1</sub> and S1P<sub>3</sub> receptors, respectively) were selected based on their docking scores. The best conformers for the rest of compounds were selected based on the maximum shape similarity to the most active compound (i.e. compounds 61 and 14 for S1P<sub>1</sub> and S1P<sub>3</sub> receptors, respectively) using Tanimoto algorithm implemented in Shape-it program [28].

### 2.3. Generation of GRIND descriptors and 3D-QSAR models

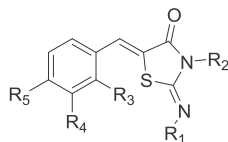
The obtained active conformations were introduced to Pentacle program for producing GRIND-based descriptors. Initially, the molecular interaction fields (MIFs) were generated using GRID based fields [29] by calculating the interaction energies at grid points called nodes between the compounds atoms and different probes including hydrophobic (DRY), hydrogen bond donor, HBD (O), hydrogen bond acceptor, HBA (N1) and shape (TIP) probes.

The process was followed by discarding the nodes with the energies below the default cut-off values. Then the most favorable regions were extracted from the produced MIFs using AMANDA algorithm based on field intensity at a node and the mutual node-node distances between the selected nodes [30]. Finally, the encoding of MIFs were processed by the Consistently Large Auto and Cross Correlation (CLACC) algorithm [31] which resulted in more consistent variables compared to Maximum Auto- And Cross-Correlation (MACC) method. The obtained values were then used for generating correlogram plots in which the product of node-node energies is represented against the distances between the nodes. The selected variables were used to build 3D-QSAR models using partial least square (PLS). The generated models were checked from the quality point of view by internal cross validation methods. Fractional factorial design (FFD) method was employed to extract the most relevant variables to the compounds activities. FFD selection was repeated several times on the models until no improvement in the statistical parameters ( $r^2$ ,  $q^2$  and SDEP values) was observed.

The models were further validated by randomly dividing the original data set into train (48 and 50 compounds for S1P<sub>1</sub> and S1P<sub>3</sub> receptors, respectively) and test sets (12 compounds for both of S1P<sub>1</sub> and S1P<sub>3</sub> receptors) based on the activities using SPSS (version 11.5). The train set compounds were used for construction of the predictive models in order to predict test compounds activities. To identify the outliers, the generated principle component analysis (PCA) for each compound in the 3D-QSAR models were used to define the applicability domain by the method of Roy and co-workers named AD using standardization approach (version 1.0) [32] where the PCA values were the X variables and the Y variable was the biological activities of the studied compounds. Validity of the final models was evaluated using both internal and external cross-validation methods. For internal validation, leave-one-out (LOO) and leave-two-out (LTO) methods were used. The

**Table 1**

Structures and biological activities of S1P<sub>1</sub> and S1P<sub>3</sub> agonists. The observed vs predicted pEC<sub>50</sub> activities for the studied compounds have been represented. The bold values are attributed to test set compounds.



No.	R1	R2	R3	R4	R5	S1P <sub>1</sub> pEC <sub>50</sub>		S1P <sub>3</sub> pEC <sub>50</sub>	
						exp.	pred.	exp.	pred.
1	Dimethyl amino	phenyl	H	Cl	OH	7.21	7.13	<b>6.86</b>	<b>7.33</b>
2	methyl	phenyl	H	Cl	OH	6.01	6.12	5.06	5.64
3	ethyl	phenyl	H	Cl	OH	<b>6.74</b>	<b>7.1</b>	5.92	5.81
4	n-propyl	phenyl	H	Cl	OH	7.18	7.52	6.73	6.51
5	n-butyl	phenyl	H	Cl	OH	6.96	7.08	<b>6.58</b>	<b>6.23</b>
6	isopropyl	phenyl	H	Cl	OH	7.33	7.13	6.93	6.56
7	tert-butyl	phenyl	H	Cl	OH	6.84	6.64	7.02	7.05
8	cyclopropyl	phenyl	H	Cl	OH	<b>6.99</b>	<b>7.17</b>	6.95	6.59
9	cyclobutyl	phenyl	H	Cl	OH	6.7	6.7	<b>6.9</b>	<b>6.89</b>
10	cyclopentyl	phenyl	H	Cl	OH	6.46	6.56	<b>6.52</b>	<b>6.57</b>
11	cyclohexyl	phenyl	H	Cl	OH	<b>6.24</b>	<b>6.35</b>	<b>6.51</b>	<b>6.51</b>
12	isopropyl	isopropyl	H	Cl	OH	7.24	7.26	6.8	6.52
13	isopropyl	n-hexyl	H	Cl	OH	6.55	6.29	6.93	6.79
14	isopropyl	cyclohexyl	H	Cl	OH	8	7.76	7.75	6.98
15	isopropyl	Ethoxy carbonylethyl	H	Cl	OH	6.32	6.4	6.91	6.82
16	isopropyl	allyl	H	Cl	OH	<b>7.26</b>	<b>7.85</b>	6.09	6.33
17	isopropyl	2-methyl-phenyl	H	Cl	OH	<b>7.47</b>	<b>7.16</b>	6.86	6.46
18	isopropyl	3-methyl-phenyl	H	Cl	OH	<b>6.96</b>	<b>7.37</b>	6.7	6.55
19	isopropyl	4-methyl-phenyl	H	Cl	OH	7.11	7.27	7.06	6.96
20	isopropyl	2,6-dimethyl-phenyl	H	Cl	OH	6.82	6.81	<b>6.52</b>	<b>6.35</b>
21	isopropyl	2,3-dimethyl-phenyl	H	Cl	OH	7.15	6.65	<b>6.46</b>	<b>6.33</b>
22	isopropyl	2,4-dimethyl-phenyl	H	Cl	OH	6.75	6.74	6.34	6.74
23	isopropyl	2-ethyl-phenyl	H	Cl	OH	6.91	6.86	6.48	6.57
24	isopropyl	2-chloro-phenyl	H	Cl	OH	7.27	7.27	6.38	6.38
25	isopropyl	3-chloro-phenyl	H	Cl	OH	7.46	7.54	6.89	6.49
26	isopropyl	3-chloro-2-methylphenyl	H	Cl	OH	7.51	7.6	6.61	6.57
27	isopropyl	3-chloro-4-methylphenyl	H	Cl	OH	7.33	7.26	6.85	6.69
28	isopropyl	2-methoxy-phenyl	H	Cl	OH	6.98	7.15	6.37	6.53
29	isopropyl	3-methoxy-phenyl	H	Cl	OH	6.59	6.63	6.29	6.56
30	isopropyl	4-methoxy-phenyl	H	Cl	OH	6.61	6.74	6.9	6.92
31	isopropyl	2,4-dimethoxy-phenyl	H	Cl	OH	<b>6.06</b>	<b>5.83</b>	<b>6.28</b>	<b>6.49</b>
32	isopropyl	3-pyridyl	H	Cl	OH	6.45	6.62	<b>6.65</b>	<b>6.75</b>
33	isopropyl	benzyl	H	Cl	OH	<b>6.21</b>	<b>6.68</b>	6.18	6.14
34	isopropyl	phenethyl	H	Cl	OH	6.04	5.76	<b>6.74</b>	<b>6.62</b>
35	isopropyl	4-phenyl-butyl	H	Cl	OH	5.75	5.84	6.68	6.66
36	isopropyl	phenyl	H	H	H	<b>5.91</b>	<b>6.05</b>	6.5	6.17
37	isopropyl	phenyl	H	H	OH	<b>6.92</b>	<b>7.23</b>	7.03	7.04
38	isopropyl	phenyl	H	F	OH	6.65	6.56	7.01	7.3
39	isopropyl	phenyl	H	CH <sub>3</sub>	OH	7.44	7.19	7.31	7.09
40	isopropyl	phenyl	H	OCH <sub>3</sub>	OH	6.7	6.98	6.48	6.27
41	isopropyl	phenyl	H	H	OCH <sub>3</sub>	<b>6.7</b>	<b>6.4</b>	6.98	6.76
42	isopropyl	phenyl	H	OCH <sub>3</sub>	OCH <sub>3</sub>	6.65	6.43	5.86	5.98
43	isopropyl	phenyl	H	H	NH(CH <sub>3</sub> ) <sub>2</sub>	6.83	6.56	<b>6.79</b>	<b>6.96</b>
44	isopropyl	phenyl	H	H	Br	5.99	6.13	6.87	7.14
45	isopropyl	phenyl	H	OCH <sub>3</sub>	H	6.74	6.63	7.46	7.16
46	isopropyl	phenyl	H	OH	H	6.13	6.09	6.99	6.9
47	isopropyl	phenyl	H	OH	OCH <sub>3</sub>	6.36	6.45	6.68	6.7
48	isopropyl	phenyl	CH <sub>3</sub>	H	H	5.2	5.43	5.24	5.74
49	isopropyl	phenyl	Cl	H	H	—	—	5.59	5.71
50	isopropyl	phenyl	OCH <sub>3</sub>	H	H	—	—	5.18	5.28
51	isopropyl	phenyl	H	H	CH <sub>2</sub> OH	6.39	6.71	7.07	7.3
52	isopropyl	phenyl	H	H	(CH <sub>2</sub> ) <sub>2</sub> OH	7	6.78	7.23	7.31
53	isopropyl	phenyl	H	H	(CH <sub>2</sub> ) <sub>3</sub> OH	<b>7.01</b>	<b>7.28</b>	7.27	7.63
54	isopropyl	phenyl	H	H	O(CH <sub>2</sub> ) <sub>2</sub> OH	7.03	7.02	6.95	7.44
55	isopropyl	phenyl	H	H	O(CH <sub>2</sub> ) <sub>3</sub> OH	7.24	7.34	7.17	7.48
56	isopropyl	phenyl	H	H	O(CH <sub>2</sub> ) <sub>2</sub> (OH) <sub>2</sub>	5.07	5.2	5.85	5.98
57	isopropyl	phenyl	H	O(CH <sub>2</sub> ) <sub>2</sub> OH	H	5.85	5.88	6.28	6.47
58	isopropyl	phenyl	H	H	O(CH <sub>2</sub> ) <sub>3</sub> N(CH <sub>3</sub> ) <sub>2</sub>	6.47	6.62	6.95	7.01
59	n-propyl	phenyl	H	Cl	OH	7.73	7.82	7.1	7
60	n-propyl	2-methyl-phenyl	H	Cl	O(CH <sub>2</sub> ) <sub>2</sub> OH	7.96	8.17	<b>6.91</b>	<b>7.04</b>
61	n-propyl	2-methyl-phenyl	H	Cl	(R)-OCH <sub>2</sub> CH(CH <sub>3</sub> )	8.05	8.04	6.92	6.91
62	n-propyl	2-methyl-phenyl	H	Cl	(S)-OCH <sub>2</sub> CH(CH <sub>3</sub> )	8.02	7.67	6.97	7.04

predictivity of the models was also assessed by calculating the standard deviation of the error of prediction (SDEP) and mean absolute error (MAE) based criteria using External Validation Plus (version 1.2) package.

#### 2.4. Database search for identification of S1P<sub>1</sub> receptor specific agonists

PubChem database of chemical molecules was searched based on the 80% structural similarity to the core structure (Fig. 1) derived from the studied molecules. The search resulted in 1796 compounds from which 90 compounds were discarded based on applying rule of five criteria implemented in PubChem web server. The 3D structures of remained 1706 compounds were retrieved from PubChem and after docking on S1P<sub>1</sub> and S1P<sub>3</sub> receptors, their activities were predicted using the proposed 3D-QSAR models. The physico-chemical properties of the molecules were calculated using HyperChem (version, 8.0.8). The profile-profile alignment was performed using ClustalW program implemented in ClustalX (version 2.0) [33].

#### 2.5. Molecular dynamic simulations

Energy minimization and binding free energy calculation were carried out using AMBER suite of programs with AMBER-ff99 force field (version 14) operating on a Linux-based (Centos 6.8) GPU work station. The best pose for each of the ligands (61, 400, 452, 798 and 799) obtained by docking calculations was used for MD simulations. Ligand input files for MD simulations were generated using Antechamber package which is designed to be used with the "general AMBER force field (GAFF) [34–36]. In order to prepare S1P<sub>1</sub> receptor-ligand complex in lipid membrane environment, CHARMM force field available at CHARMM-GUI web-based platform ([charmm-gui.org](http://charmm-gui.org)) [37] was employed where the complex was embedded in a hydrated, pre-equilibrated 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipid bilayer with 130 DOPC molecules per complex. The system was neutralized by addition of potassium and chloride ions to the final concentration of 150 mM and then converted to tleap and Lipid14 readable file using charmm lipid2amber.py script. Subsequently, the produced file was introduced to tleap program where Lipid14 [38] and Amber-ff99SB force fields implemented in AmberTools 14 were recruited to generate Amber topology and initial coordinates files.

The generated system was submitted to a short energy minimization process including 5000 steps of steepest descent and 5000 steps of conjugate gradient followed by a 100 ps heating step from 0 K to 100 K in a NVT and then from 100 K to 303 K in a NPT ensembles both with 10.0 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic restrains applied to the protein and to the lipids. Then, the system was equilibrated in the NPT ensemble at 303 K (controlled with Langevin thermostat) with 1 bar pressure for 2 ns followed by gradually removing the restraints. Only bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm.

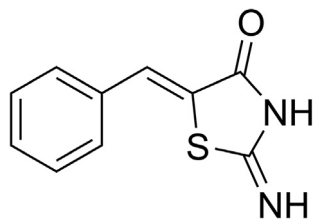


Fig. 1. The selected core structure from data set compounds for PubChem database search.

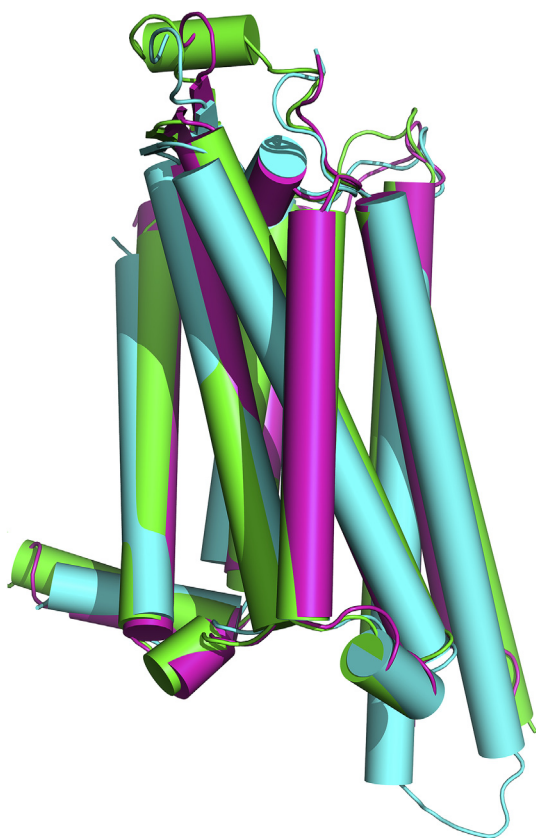
The final production of dynamic simulation was performed for 125 ns by applying the Particle Mesh Ewald (PME) method under periodic boundary condition where no constraint was applied to the protein, lipids and the ligand molecules. Binding energies were calculated for ligand–receptor complexes using molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) and the molecular mechanics generalized Born surface area (MM-GBSA) algorithms. The interaction energies were calculated excluding lipid, water molecules and counter ions and presented as the average value.

### 3. Results and discussion

The aim of this study is to derive 3D-QSAR models for a set of 62 compounds with 2-imino-thiazolidin-4-one scaffold (Table 1) in order to identify the structural features essential for the interaction of the compounds with S1P<sub>1</sub> and S1P<sub>3</sub> receptors. The obtained results could be used for identifying new ligands active on S1P<sub>1</sub> receptors in database search and drug design studies. In the current study, the crystal structure of S1P<sub>1</sub> receptor in complex with an antagonist was used for virtual screening to identify S1P<sub>1</sub> agonists. It has been elucidated that GPCRs adopt active or inactive structure upon binding to either an agonist or an antagonist, respectively. Up to now, the structure of S1P<sub>1</sub> with an agonist has not been solved experimentally, therefore, the antagonist-bound S1P<sub>1</sub> structure was used in the current study. To examine how different might be the active and inactive forms of S1P<sub>1</sub> receptor, the active and inactive structures of closely related GPCR was investigated. The basic local alignment tool (BLAST) was employed to search protein data bank for structurally known homologous GPCRs. The search identified adenosine A2A receptor bound to an antagonist (PDB code: 3PWH) as one of the closest homologues of S1P<sub>1</sub> receptor (28.67% identity, 85% coverage, and  $2 \times 10^{-23}$  E-value). As a matter of fact, the structure of adenosine A2A receptor could have been used as the template for homology modeling of S1P<sub>1</sub> receptor if there was no experimental structure for S1P<sub>1</sub> itself. There are two other structures for adenosine A2A receptor co-crystallized with agonist ligands (PDB IDs: 3QAK and 2YDO). Assuming agonist and antagonist bound structures of receptor representing the active and inactive forms, calculation of the C $\alpha$  RMSDs between the adenosine A2A active and inactive forms revealed that they are very similar (RMSD values of 3QAK and 2YDO active structures against 3PWH inactive structure are 1.37 and 1.33 Å), which is comparable to the difference between two active structures (RMSD value of 0.82 Å). The structures of adenosine A2A receptor in active and inactive forms are compared in Fig. 2. These observations for the most identical GPCR with known structure to the S1P<sub>1</sub> receptor indicate the viability of using antagonist-bound S1P<sub>1</sub> receptor structure for the purpose of current study, and it may not be feasible to distinguish fine structural differences between the active and inactive forms of the receptor by MD simulations. Moreover, in this study, semi-flexible docking procedure was utilized for the docking of ligands retrieved from the database structure similarity search.

#### 3.1. Homology modeling and docking studies

SWISS-MODEL automated homology modeling server was used to develop the 3D structure for S1P<sub>3</sub> receptor. The server was provided by the sequence of S1P<sub>3</sub> and the generated models as well as the corresponding quality measures were generated. The best model proposed by SWISS-MODEL was built based on S1P<sub>1</sub> receptor structure (PDB code: 3V2W.1.A) as the template with ~48% similarity to the target sequence. The RMSD value between the template and the model was 0.08, calculated by DeepView (version 3.7) program. The quality of the modeled structure was verified



**Fig. 2.** Comparison of conformations of adenosine A2A receptor in its active (PDB IDs: 2YDO and 3QAK) and inactive (PDB ID: 3PWH) states using cylinder representation. 2YDO and 3QAK are colored in green and purple, respectively, while 3PWH is shown as cyan.

using different methods as outlined below. PROCHECK and Mol-Probity programs showed that 100.0% (270/270) of all residues were in allowed (>99.8%) regions. The QMEAN z-score reported for the model by SWISS-MODEL server was  $-3.17$ . Although this value demonstrates a moderate quality for the built model, taking into account that S1P<sub>3</sub> belongs to GPCRs family, the model can be considered acceptable.

The model structure of S1P<sub>3</sub> and the experimental structure of S1P<sub>1</sub> receptors were used for the docking studies. The active conformation for each ligand was obtained by docking the compound into the active site of S1P<sub>1</sub> and S1P<sub>3</sub> receptors using GOLD program. To do this, the best active conformation for the most active compound in each data set (i.e. compounds 61 and 14 for S1P<sub>1</sub> and S1P<sub>3</sub> receptors, respectively) was selected based on just the docking scores (compounds 61 and 14 had docking scores of 64.56 and 60.18 upon docking into S1P<sub>1</sub> and S1P<sub>3</sub> receptors). Then, by applying Tanimoto similarity method implemented in Shape-it program (version 1.0.1), the active conformations of the remaining compounds in the data sets were selected based on the similarity of the generated docking solutions for each compound to the selected conformation of the most active compound (Tables 2 and 3). The compounds in the selected conformations were used in QSAR model building as outlined below.

### 3.2. Generation of 3D-QSAR model for activity prediction

After selection of the best conformations for the ligands, the compounds were divided into train and test sets randomly based on their biological activities using SPSS program (version 11.5) in an

effort to have similar range of biological activities for both sets. The train compounds were subjected to Pentacle program and the generated GRIND-based descriptors were used to build 3D-QSAR models. FFD variable selection was performed for several times on the generated models until no significant changes in statistical indices (squared correlation coefficient  $r^2$ , cross validated correlation coefficient  $q^2$  and SDEP) were observed. The internal predictivity of the models was evaluated by LOO and LTO methods and the obtained statistics are shown in Table 4. As seen in Table 4, the  $r^2_{acc}$  values and standard deviation of error in calculation (SDEC) for S1P<sub>1</sub> and S1P<sub>3</sub> models have been improved by increasing the number of latent variables and reached to the optimum values with five latent variables (5LVs). Although LOO and LTO  $q^2$  values (i.e., 0.39 and 0.42, respectively) for S1P<sub>1</sub> model with 5LVs are lower than that with 2, 3 and 4 LVs, the overall assessment of internal validation and prediction errors suggested the model with 5LVs as the best model for activity prediction. However, in the case of S1P<sub>3</sub> model, the model generated with 5LVs was validated as the best predictive model from both predictivity (i.e.,  $q^2$  LOO, 0.48 and  $q^2$  LTO 0.49) and level of errors (SDEC, 0.19) points of view.

Fig. 3 illustrates the distance dependent plot of PLS coefficients for the selected variables analyzed with 5LVs during 3D-QSAR model development. The most important variables for the 3D-QSAR models are summarized in Table 5. As it can be seen in Table 5 and Fig. 3, the S1P<sub>1</sub> and S1P<sub>3</sub> models share DRY-DRY (14–14.4 Å), N1–N1 (11.2–11.6 Å), TIP–TIP (17.6–18 Å) and N1–TIP (4.4–4.8 and 8–8.4 Å) variables with almost equal probe-probe distances; while, DRY–N1 (10.4–10.8 Å) and DRY–O (12.4–12.8 Å) variables were unique MIF pairs which were selected just in S1P<sub>1</sub> model with positive and negative impacts on biological activities, respectively. As seen in Fig. 4 DRY–N1 (10.4–10.8 Å) variable connects thiazolidine ring to hydroxyl moiety at the most end of the structure indicating that the presence of a hydrophobic region (for example thiazolidine ring in compound 61) at the distance of 10.4–10.8 Å from a hydrogen acceptor group (i.e., hydroxyl group in compound 61) positively influences the biological activities as well as the selectivity of the studied compounds towards S1P<sub>1</sub> receptor. On the other hand, analyzing the PLS coefficients revealed that the positive impact of DRY–N1 variable is higher for the compounds 02, 04, 21, 23, 25, 25, 26, 28, 29, 59 and 62 whose selectivity towards S1P<sub>1</sub> is higher compared to the compounds 44, 45, 46, 56 and 58 as the least selective compounds towards S1P<sub>1</sub> over S1P<sub>3</sub> receptor. Another selectivity inducing variable, DRY–O (12.4–12.8 Å), connects a hydrophobic region (i.e., thiazolidine ring in compound 61) to a hydrogen donor group (i.e., etheric oxygen in compound 61), far apart 10.4–10.8 Å from each other which adversely affects the S1P<sub>1</sub> receptor agonistic activity. The impact of this variable is lower for the S1P<sub>1</sub> selective compounds such as compounds 22, 25, 42, 60 and 62 which are the S1P<sub>1</sub> selective compounds rather than S1P<sub>3</sub> receptor. Taking into account all these explanations, it seems that DRY–N1 and DRY–O variables can be considered as the S1P<sub>1</sub> receptor selectivity-introducing factors for the studied compounds.

### 3.3. Predictivity assessment

For external prediction, based on the model statistics, five latent variables (LVs) were selected as the optimum number of PLS components for both S1P<sub>1</sub> and S1P<sub>3</sub> receptors activity prediction models. The  $r^2$  values for external prediction were 0.75 and 0.62 with SDEP values of 0.33 and 0.21 for S1P<sub>1</sub> and S1P<sub>3</sub> models, respectively, indicating the good predictive power of the developed models. Experimental vs. predicted values are presented in Table 1 and Fig. 5.

According to the Organization for Economic Co-operation and Development (OECD) [32,39], it is required to define the

**Table 2**  
Docking scores and Tanimoto index for the studied compounds upon docking the ligands into S1P<sub>1</sub> receptor. Shape-it program was used to select a conformation of each ligand structurally similar to the most active compound. GOLD scores and Tanimoto index values are included for comparison.

GOLD Docking			Shape-it selecton		
Compounds	Best rank GOLD score	Tanimoto index	Docking rank	GOLD score	Tanimoto index
1	61.97	0.483	1	61.97	0.483
2	56.03	0.396	5	50.67	0.406
3	57.86	0.392	9	56.43	0.469
4	57.85	0.393	1	52.94	0.51
5	57.72	0.355	1	56.57	0.545
6	54.56	0.494	2	52.61	0.551
7	51.49	0.358	4	49.93	0.462
8	63.15	0.411	7	56.83	0.548
9	58.83	0.456	3	58.83	0.456
10	62.03	0.41	9	57.03	0.473
11	59.22	0.401	4	58.74	0.509
12	58.32	0.372	3	57.96	0.442
13	61.32	0.337	2	60.16	0.373
14	61.74	0.4	8	45.63	0.471
15	64.68	0.383	7	61.4	0.412
16	58.16	0.391	7	54.89	0.469
17	66.74	0.371	1	56.31	0.489
18	56.26	0.465	10	56.26	0.465
19	58.35	0.434	10	56.36	0.462
20	54.09	0.393	3	49.53	0.515
21	57.01	0.338	7	50.63	0.467
22	58.23	0.483	2	58.23	0.483
23	69.35	0.442	3	45.06	0.455
24	57.28	0.426	10	51.47	0.504
25	56.5	0.461	9	53.84	0.48
26	60.43	0.377	10	51.47	0.496
27	58.54	0.425	6	46.17	0.486
28	57.62	0.406	4	49.34	0.535
29	55.36	0.453	3	50.39	0.479
30	58.1	0.458	4	58.54	0.458
31	49.38	0.387	3	46.62	0.437
32	61.13	0.381	3	60.47	0.485
33	64.3	0.391	2	51.39	0.42
34	64.35	0.392	6	57.59	0.424
35	71.89	0.392	3	65.65	0.426
36	55.5	0.437	10	53.89	0.495
37	58.83	0.445	5	51.34	0.491
38	61.44	0.4	1	52.33	0.528
39	57.67	0.481	2	55.5	0.481
40	58.9	0.383	10	46.98	0.523
41	57.67	0.429	4	50.52	0.574
42	55.03	0.469	5	52.83	0.52
43	61.29	0.511	9	54.18	0.569
44	56.9	0.422	10	50.43	0.637
45	56.44	0.396	1	53.42	0.505
46	55.05	0.461	9	51.25	0.53
47	58.6	0.459	2	54.11	0.503
48	53.91	0.407	1	49.27	0.503
51	58.72	0.553	9	49.98	0.49
52	59.22	0.585	9	59.22	0.585
53	63.71	0.528	7	61.38	0.534
54	61.47	0.507	3	59.53	0.518
55	62.49	0.419	7	61.75	0.555
56	63.17	0.409	5	56.68	0.517
57	61.04	0.369	3	59.77	0.544
58	67.4	0.457	7	62.72	0.535
59	58.76	0.502	9	54.94	0.55
60	67.01	0.53	10	54.3	0.583
61	64.56	1	1	64.56	1
62	66.3	0.55	3	58.79	0.609

applicability domain (AD) for the developed QSAR models. Therefore, the calculated PC values during the PLS analysis for the studied compounds and their biological activities were analyzed by "Applicability Domain using standardization approach" program to determine the possible outliers in train and test sets [32]. The results indicated none of the structures as the outlier. To further evaluate the robustness of the models, the errors of the predicted

values were calculated and the results are presented in Table 6. The calculated mean absolute errors of the predictions for 95% of the test sets were 0.28 and 0.13 for S1P<sub>1</sub> and S1P<sub>3</sub> models, respectively. These error values are smaller than 0.30 and 0.27 (i.e., 10% of the train set activity range for S1P<sub>1</sub> and S1P<sub>3</sub> models data sets, respectively) indicating adequate prediction ability of the models. Using more rigorous treatment of the predictions by MAE-based

**Table 3**

Docking scores and Tanimoto index values for the studied compounds upon docking the ligands into S1P<sub>3</sub> receptor. Shape-it program was used to select a conformation of each ligand structurally similar to the most active compound. GOLD scores and Tanimoto index values are included for comparison.

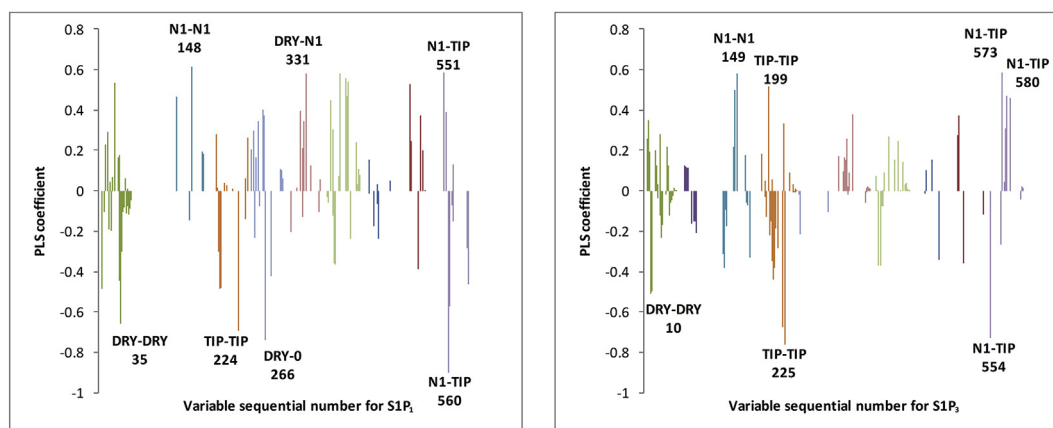
GOLD Docking			Shape-it selecton		
Compounds	Best rank GOLD score	Tanimoto index	Docking rank	GOLD score	Tanimoto index
1	59.4	0.546	7	53.67	0.56
2	60.43	0.474	9	60.43	0.474
3	61.58	0.443	4	60.59	0.525
4	60.05	0.459	7	51.49	0.642
5	58.58	0.546	9	58.44	0.599
6	51.66	0.517	4	48.13	0.528
7	65.97	0.551	8	50.82	0.572
8	62.63	0.499	7	51.14	0.567
9	55	0.509	9	53.08	0.617
10	63.5	0.512	2	51.53	0.515
11	54.93	0.547	9	53.74	0.603
12	60.6	0.493	10	49.85	0.649
13	62.56	0.573	8	61.38	0.708
14	60.18	10	10	60.18	1
15	58.88	0.421	7	56.87	0.728
16	60.89	0.439	3	52.11	0.597
17	55.46	0.534	6	48.96	0.561
18	53.53	0.511	2	53.53	0.511
19	57.27	0.453	3	52.15	0.545
20	50.04	0.485	6	49.59	0.594
21	53.86	0.445	6	53.4	0.581
22	52.83	0.497	8	51.99	0.529
23	55.91	0.548	2	51.31	0.602
24	49.87	0.522	5	46.9	0.584
25	54.42	0.499	1	48.06	0.554
26	53.74	0.45	3	50.15	0.568
27	54.3	0.417	4	52.02	0.539
28	53.01	0.517	7	50.44	0.659
29	55.54	0.51	7	55.54	0.51
30	52.18	0.447	10	49.48	0.546
31	54.24	0.462	6	53.19	0.498
32	54.32	0.458	1	49.04	0.593
33	58.68	0.561	10	54.48	0.645
34	63.99	0.519	8	60.92	0.599
35	71.84	0.494	9	71.84	0.494
36	55.37	0.487	7	47.03	0.609
37	51.75	0.462	5	48.6	0.532
38	56.71	0.482	2	51.62	0.551
39	56.07	0.464	6	52.84	0.558
40	56.72	0.467	5	49.36	0.509
41	55.3	0.476	7	44.95	0.675
42	58.43	0.455	8	55.43	0.553
43	57.58	0.502	4	55.31	0.521
44	57.24	0.505	7	48.69	0.563
45	58.88	0.527	1	51.76	0.544
46	57.64	0.48	3	49.67	0.533
47	59.52	0.448	1	53.38	0.591
48	60.02	0.54	5	44.22	0.637
49	50.76	0.54	5	48.06	0.566
50	58.65	0.623	9	46.68	0.645
51	57	0.543	4	47.55	0.547
52	61.28	0.512	4	60.44	0.545
53	61.27	0.462	2	56.68	0.507
54	61.71	0.453	9	53.73	0.545
55	66.66	0.516	8	56.32	0.53
56	67.54	0.449	1	53.97	0.512
57	63.4	0.441	2	52.27	0.532
58	69.98	0.453	5	56.93	0.48
59	56.96	0.477	1	55.73	0.628
60	66.76	0.552	1	63.99	0.563
61	63.76	0.545	7	63.76	0.545
62	62.47	0.367	1	60.35	0.497



**Table 4**  
The statistical data of the built PLS model for S1P<sub>1</sub> and S1P<sub>3</sub> agonists.

Models for s1p1							
No. LVs	$r^2_{acc}$	SDEC	$q^2$ LOO	SDEP	$q^2$ LTO	SDEP	
1	0.25	0.57	0.1	0.63	0.11	0.62	
2	0.69	0.36	0.46	0.49	0.45	0.49	
3	0.82	0.28	0.49	0.47	0.47	0.48	
4	0.87	0.24	0.46	0.49	0.42	0.50	
5	0.93	0.18	0.39	0.52	0.42	0.50	
Models for s1p3							
No. LVs	$r^2_{acc}$	SDEC	$q^2$ LOO	SDEP	$q^2$ LTO	SDEP	
1	0.56	0.38	0.36	0.45	0.33	0.46	
2	0.68	0.32	0.45	0.42	0.46	0.42	
3	0.79	0.26	0.43	0.43	0.46	0.42	
4	0.85	0.22	0.48	0.41	0.51	0.4	
5	0.89	0.19	0.48	0.41	0.49	0.41	

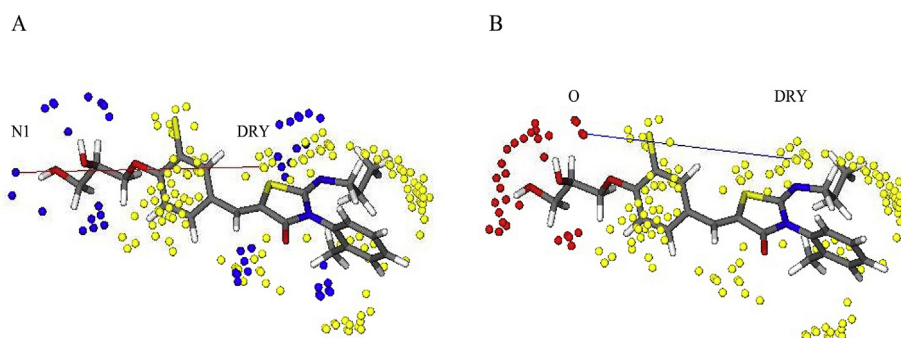
Abbreviations: SDEC, standard deviation of error in calculation; SDEP, standard deviation of error of prediction. The acc stands for accumulative value, Validation methods used for calculation of  $q^2$  are: leave one out (LOO), leave two out (LTO) and random five groups out (R5GO).



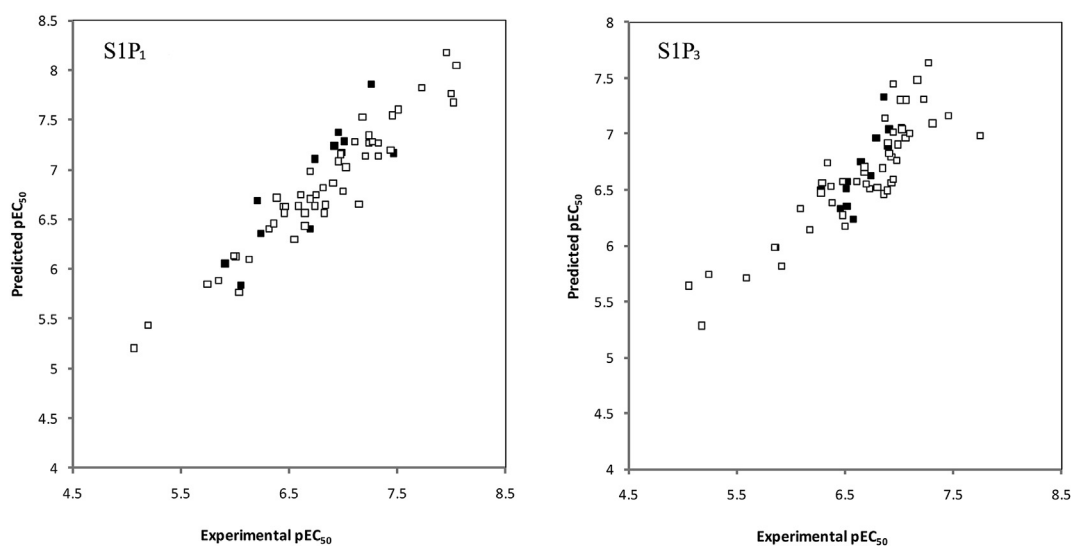
**Fig. 3.** 5LV PLS coefficient plots for the obtained models. The most intensive variables are labeled by sequential numbers. 5LV indicates 5 latent variables; PLS, partial least squares.

**Table 5**  
The most important structural variables in the 3D-QSAR model for S1P<sub>1</sub> and S1P<sub>3</sub> agonists.

S1P1			
Probe block	Variable No	Distance (Å)	Impact
DRY-DRY	35	14–14.4	–
N1–N1	148	11.2–11.6	+
TIP-TIP	224	17.6–18	–
DRY-O	266	10.4–10.8	–
DRY-N1	331	12.4–12.8	+
N1-TIP	551	4.4–4.8	+
N1-TIP	560	8–8.4	–
S1P3			
Probe block	Variable No	Distance (Å)	Impact
DRY-DRY	10	4–4.4	–
N1–N1	149	10.8–11.2	–
TIP-TIP	199	6.4–6.8	+
TIP-TIP	225	16.8–17.2	–
N1-TIP	554	2–2.4	–
N1-TIP	573	9.6–10	+
N1-TIP	580	12.4–12.8	+



**Fig. 4.** The unique structural elements associated with variables selected in S1P<sub>1</sub> model: (A) DRY-N1 (10.4–10.8 Å) and (B) DRY-O (12.4–12.8 Å).



**Fig. 5.** Experimental vs predicted EC<sub>50</sub> values for compounds. From data set, 48 and 50 compounds were used as train sets for S1P<sub>1</sub> and S1P<sub>3</sub> models, respectively while 12 compounds as test sets. Open and filled squares indicate train and test sets compounds, respectively.

**Table 6**

The calculated errors for the predicted pEC<sub>50</sub> values of external validation compounds.

	r2	SDEP	MAE (95% data)	MAE+3*SD (95% data)	training set range
S1P <sub>1</sub>	0.75	0.33	0.28	0.62	2.98
S1P <sub>3</sub>	0.62	0.21	0.13	0.43	2.69

Abbreviations: SDEP, standard deviation of error of prediction; MAE, mean absolute error.

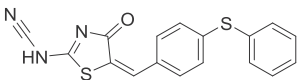
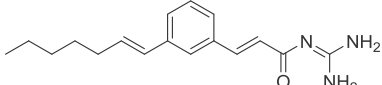
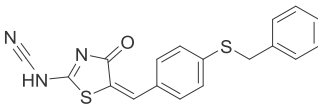
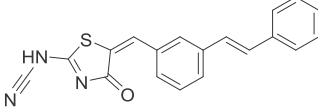
metrics of Roy et al., the prediction power of the S1P<sub>1</sub> and S1P<sub>3</sub> models are considered to be of moderate and good qualities, respectively.

#### 3.4. Identification of selective S1P<sub>1</sub> receptor agonist candidates

In order to propose potential S1P<sub>1</sub> receptor selective agonists, PubChem database of chemical molecules was searched based on 80% structural similarity to the core structure of the studied compounds (Fig. 1). The molecules identified by similarity search (1706 structures) were docked on the binding site of S1P<sub>1</sub> and S1P<sub>3</sub> receptors. The agonistic activities for the best docking poses of the identified molecules were predicted based on the developed 3D-QSAR models. The results demonstrated that compounds 400 (PubChem ID: 136647497), 452 (PubChem ID: 72785505), 798 (PubChem ID: 18758433) and 799 (PubChem ID: 18758427) could bind to S1P<sub>1</sub> receptor with predicted EC<sub>50</sub> values of 0.65, 3.63, 1.35

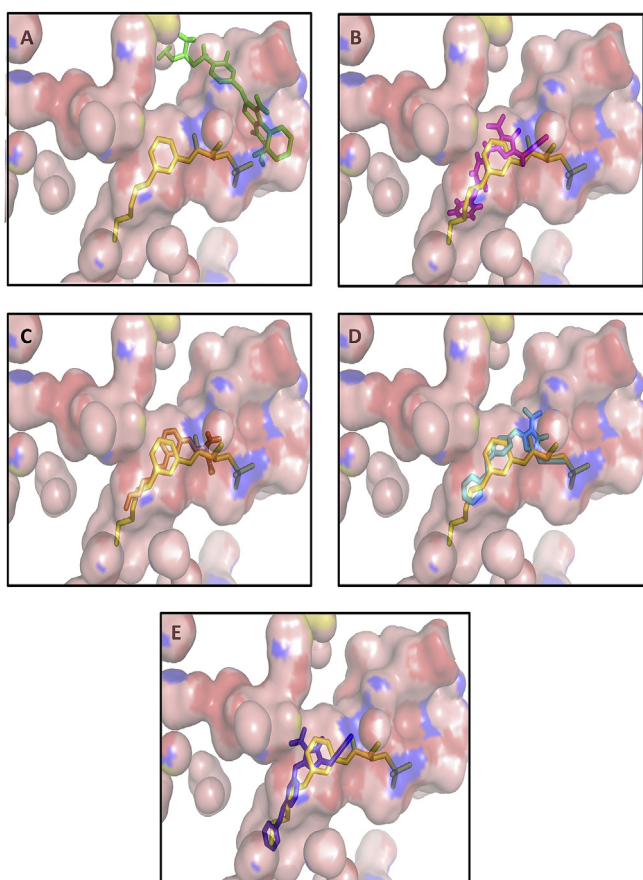
and 0.68 nM, respectively, while their predicted affinities towards S1P<sub>3</sub> receptor were estimated to be considerably low (Table 7). Comparing the predicted affinities towards S1P<sub>1</sub> and S1P<sub>3</sub> receptors showed excellent selectivity for compound 452 (S1P<sub>3</sub> EC<sub>50</sub>/S1P<sub>1</sub> EC<sub>50</sub>: 13489). This level of predicted selectivity is comparable with the selectivities reported for S1P<sub>1</sub> agonists in clinical trials such as ozanimod [19] and siponimod [40] and much greater than that of compound 61, indicating that compound 452 is noteworthy to be evaluated in experiments for development of a potent and selective S1P<sub>1</sub> agonist. Moreover, among the selected compounds, compound 799 showed high predicted EC<sub>50</sub> value (0.68 nM) with acceptable S1P<sub>1</sub> over S1P<sub>3</sub> selectivity (5888 times), demonstrating its feasibility to be explored as a potent and selective S1P<sub>1</sub> agonist. Overall, it seems that the selected compounds from PubChem search can be considered as promising selective and potent S1P<sub>1</sub> receptor agonists.

**Table 7**  
The structures and the predicted activities for compounds selected from PubChem search as the S1P<sub>1</sub> agonists. The database of chemical molecules in PubChem was searched based on 80% similarity to the core structure originated from the data set and the selected compounds activities were predicted by the generated 3D-QSAR models.

Compounds	structure	S1P <sub>1</sub> EC50 (nM)	S1P <sub>3</sub> EC50 (nM)	S1P <sub>3</sub> EC50/S1P <sub>1</sub> EC50
400		0.65	549.54	851.13
452		3.63	48977.88	13489.63
798		1.35	7413.10	5495.41
799		0.68	3981.07	5888.44

**Table 8**  
Physicochemical properties of PubChem selected compounds compared to S1P<sub>1</sub> agonists in clinical trials. The values have been calculated using QSAR properties module implemented in HyperChem software.

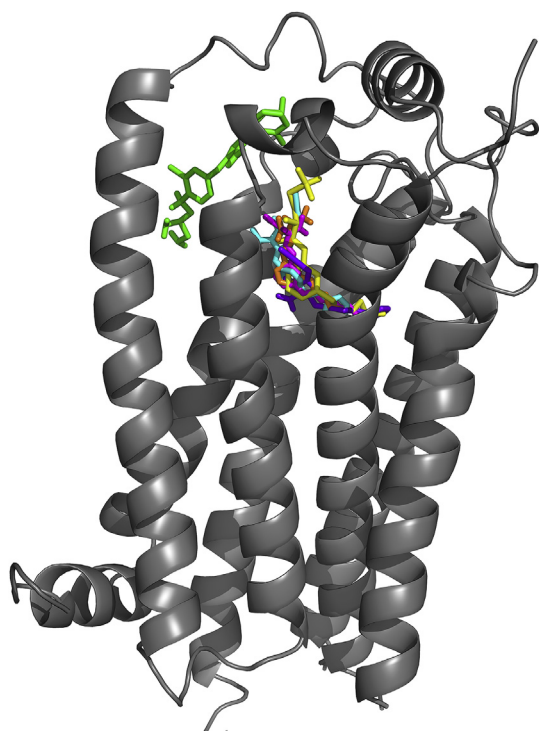
Compounds	LogP	Mass (amu)
400	0.63	339.43
452	3.09	285.39
798	2.06	351.44
799	2.84	330.38
61 (Ponesimod)	0.60	460.98
Sphingosine 1-phosphate	5.00	379.48
Ozanimod	1.30	404.47
Siponimod	5.04	516.60



**Fig. 6.** 3D-representation of the docked S1P<sub>1</sub> receptor agonists in comparison with S1P<sub>1</sub> receptor co-crystallized antagonist. S1P<sub>1</sub> receptor antagonist has been implemented in all figures and colored yellow. Panel A compound 61, B compound 400, C compound 452, D compound 798, and E compound 799. The compounds selected from PubChem search have been named by their appearance order in database search. PyMol program (version 1.7.0.0) was used to generate the images. The receptor and the ligands have been shown as surface and sticks, respectively. By making the surface 90% transparent the ligands inside of the protein were shown. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

### 3.5. Prediction of binding mode between S1P<sub>1</sub> receptor and its agonists

The 3D illustrations of the obtained best docking pose for compounds 61, 400, 452, 798 and 799 in complex with S1P<sub>1</sub> receptor are shown in Fig. 6 and compared with S1P<sub>1</sub> receptor co-crystallized antagonist. As seen in Fig. 6, compound 61 from the data set interacts differently with the receptor compared to S1P<sub>1</sub> receptor co-crystallized antagonist and the compounds identified from PubChem search. However, the latter compounds bind the same site as the antagonist, which may indicate their capability of S1P<sub>1</sub> receptor recognition. Analyzing the chemical structures of the selected compounds reveals some important details about the observed results in terms of predicted potency and selectivity of these structures. Table 8 summarizes the physicochemical properties of the selected compounds in comparison to compound 61, S1P, ozanimod, and siponimod. The analysis shows that those compounds more hydrophobic than compound 61 have higher predicted activities, which may be attributed to the easy penetration of more lipophilic compounds into the innermost binding site of S1P<sub>1</sub> receptor (Fig. 7). In an effort to identify the interaction modes of agonists with the receptor, the 2D illustrations of the bindings were generated using Discovery Studio Visualization (version 17.2.0) program (Fig. 8). The results revealed that the binding of compounds 400, 452, 798 and 799 with S1P<sub>1</sub> receptor are dominantly stabilized by hydrophobic interactions. Site directed mutational studies have revealed the importance of Arg<sup>120</sup> and Glu<sup>121</sup> residues located at the extracellular end of third transmembrane helix of S1P



**Fig. 7.** 3D-representation of S1P<sub>1</sub> receptor in complex with its ligands. As seen, compounds 61 (shown as green sticks) is interacting differently with S1P<sub>1</sub> receptor compared to S1P<sub>1</sub> receptor co-crystallized antagonist (yellow) and compounds 400 (pink), 452 (orange), 798 (cyan), and 799 (purple). PyMol program (version 1.7.0.0) was used to generate the images. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

receptors in signal transduction [41,42]. Gonzalez-Cabrera et al. showed that ligands not interacting with both Arg<sup>120</sup> and Glu<sup>121</sup> residues or interact just with one of them can exert agonistic activity with improved specificity [43]. 2D analysis of the identified ligands docked into the binding site of S1P<sub>1</sub> receptor (Fig. 8) revealed that they do not interact at a time with Arg<sup>120</sup> and Glu<sup>121</sup>. These observations are in agreement with the selective agonistic S1P<sub>1</sub> activities predicted for the identified compounds obtained from database search [43,44]. Moreover, the identified compounds also interact with Phe<sup>210</sup> and Trp<sup>269</sup> residues, which are known essential for agonistic activity as confirmed by the point mutagenesis studies [3]. Further, these ligands interact with Leu<sup>276</sup> and Met<sup>124</sup> at the binding site of S1P<sub>1</sub> which are S1P<sub>1</sub> selectivity-inducing residues over S1P<sub>3</sub> and S1P<sub>4</sub> receptors, respectively [3]. The identification of the selective S1P<sub>1</sub> receptor agonists demonstrated in the previous section outlines the potential for selectivity predictions based on the above mentioned qualitative descriptions of the binding modes to S1P receptors defined implicitly in the developed 3D-QSAR models. The interesting finding regarding the most selective compound (452), identified based on PubChem similarity search, was that it binds and inhibits E protein from Human coronavirus. This protein plays a central role in virus morphogenesis and assembly and acts as a viroporin and self-assembles in host membranes forming pentameric protein-lipid pores that allow ion transport. The experimental studies have shown that compound 452 binds to the transmembrane helical segments of E protein. Although, there is no apparent similarity between E protein and S1P<sub>1</sub> receptor sequences, both of them are helical transmembrane proteins. The docking studies conducted in this work identified the core of the helical bundle of S1P<sub>1</sub> receptor (a member of 7 transmembrane GPCR superfamily) as the binding

site for 452, which is in agreement to the experimentally identified binding site of this compound formed by transmembrane helical bundle of pentameric E protein. The alignment between profiles prepared for S1P<sub>1</sub> receptor and E protein using their closely related sequences and the known secondary structural information revealed that E protein can be aligned with the 5th transmembrane helix of S1P<sub>1</sub> receptor as demonstrated in Fig. 9. Although not very conclusive, but one may state that the binding sites available on S1P<sub>1</sub> receptor and E protein for compound 452 could share similar environment.

### 3.6. S1P<sub>1</sub> receptor binding energies of the ligands

In order to calculate the free binding energies of the compounds identified from PubChem search in complex with S1P<sub>1</sub> receptor, the MD simulations were performed using Lipid14 [38] and Amberff99SB force fields available in AMBER 14. To do this, S1P<sub>1</sub> receptor with docked ligands was embedded in lipid bilayer membrane (Fig. 10) and MD simulated for 125 ns. Table 9 represents the calculated binding energies for ligands-receptor interactions using MM-GBSA and MM-PBSA algorithms after 125 ns MD simulations. The results revealed that compound 400 binds S1P<sub>1</sub> receptor stronger than the rest of studied compounds with the binding  $\Delta G^\circ$  values of  $-46.18$  and  $-9.75$  kcal mol<sup>-1</sup>, calculated by MM-GBSA and MM-PBSA algorithms, respectively. In the MD simulations, compound 61 was used for comparison which showed the lowest affinity towards S1P<sub>1</sub> receptor with  $\Delta G^\circ$  binding energy values of  $-28.38$  and  $28.98$  kcal mol<sup>-1</sup>, calculated by MM-GBSA and MM-PBSA algorithms, respectively. Interestingly, the receptor binding energy ( $\Delta G^\circ$ ) values of the studied compounds (Table 9) correlate strongly with their predicted EC<sub>50</sub> values (Table 7) calculated by S1P<sub>1</sub> 3D-QSAR model, which in turn, further validates the developed model.

## 4. Conclusion

In the current work, we aimed to build 3D-QSAR predictive models for a series of S1P<sub>1</sub> and S1P<sub>3</sub> receptors agonists in order to identify structural features required to design S1P<sub>1</sub> receptor selective agonists. The validity of the developed models was evaluated by the internal and external cross validation methods, in which the generated models showed adequate statistical parameters for their predictive abilities. The results showed that interactions such as DRY-N1 and DRY-O promote the agonistic activity towards S1P<sub>1</sub> receptor. Furthermore, structural variables associated with hydrophobic moieties are the important features influencing the S1P<sub>1</sub> receptor selectivity of the studied compounds.

Structural similarity search of PubChem database using the core structure of the data set as the query structure identified 1706 compounds of which compounds 400, 452, 798, and 799 demonstrated excellent predicted potencies and selectivities towards S1P<sub>1</sub> receptor. Docking of the most potent compound (i.e., compound 61) in the data set and those identified by database similarity search showed that they interact differently with S1P<sub>1</sub> receptor which may be attributed to the higher lipophilicity of the identified compounds compared to compound 61 from the data set. The identified most selective compound from the similarity search bind to E protein from Human coronavirus which may share commonality with S1P<sub>1</sub> receptor in terms of binding site, providing further evidence in agreement with the results obtained from 3D-QSAR model. Moreover, S1P<sub>1</sub> receptor binding energies were calculated using molecular dynamics simulations for the agonists identified from PubChem database search. The results were indicative of their appropriate receptor affinities due to the negative binding energies, which are also in close agreement with the predicted EC<sub>50</sub> values.

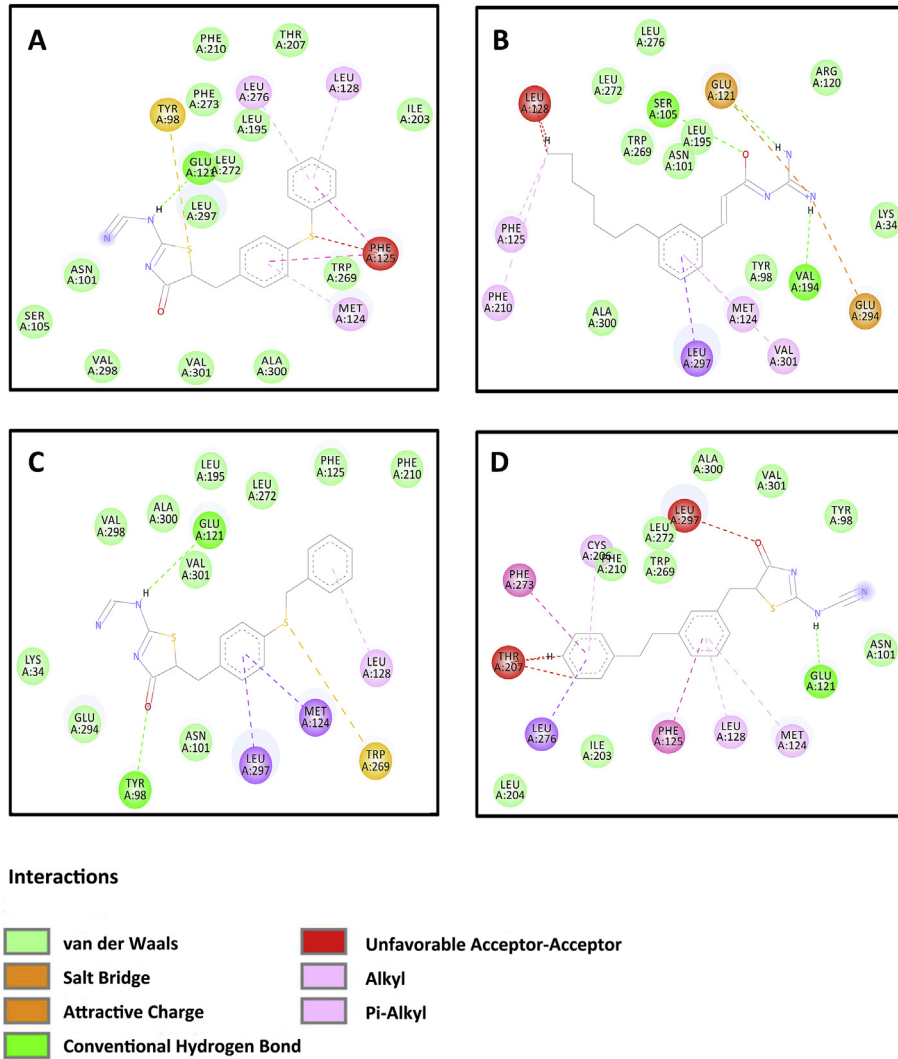


Fig. 8. 2D-representation of interactions of docked compounds (A) 400, (B) 452, (C) 798, and (D) 799 at the binding site of S1P<sub>1</sub> receptor. Discovery Studio Visualization (v 17.2.0) program was used for analysis.

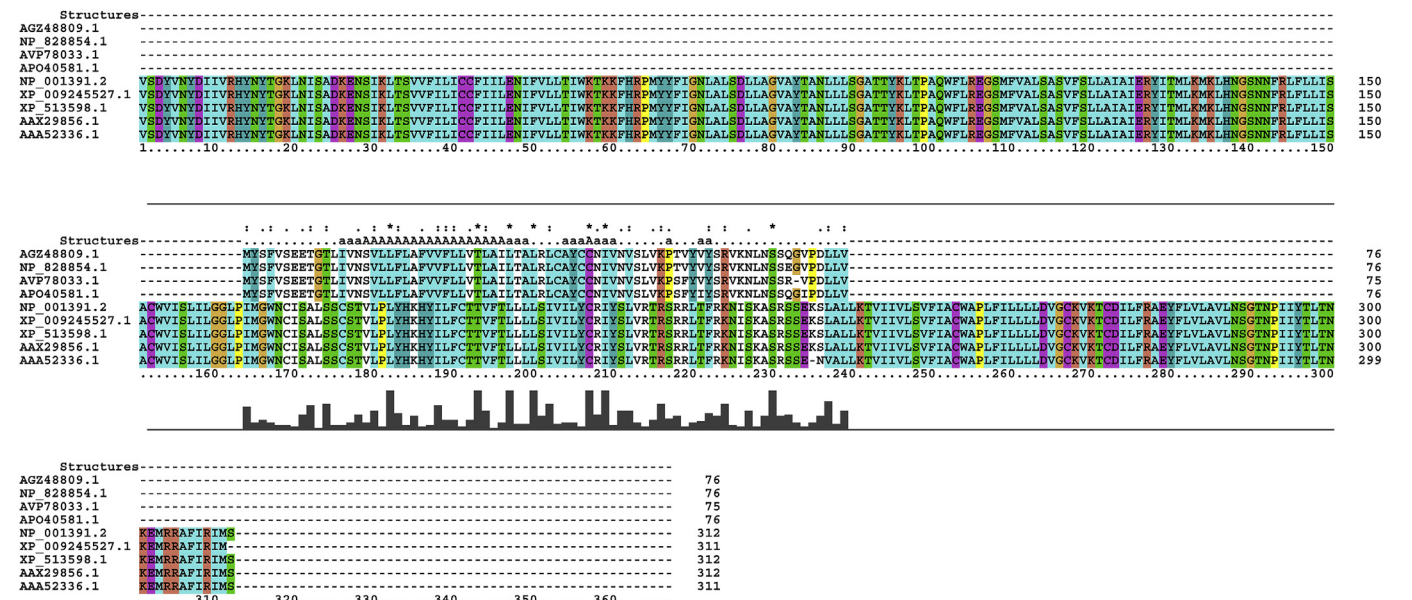
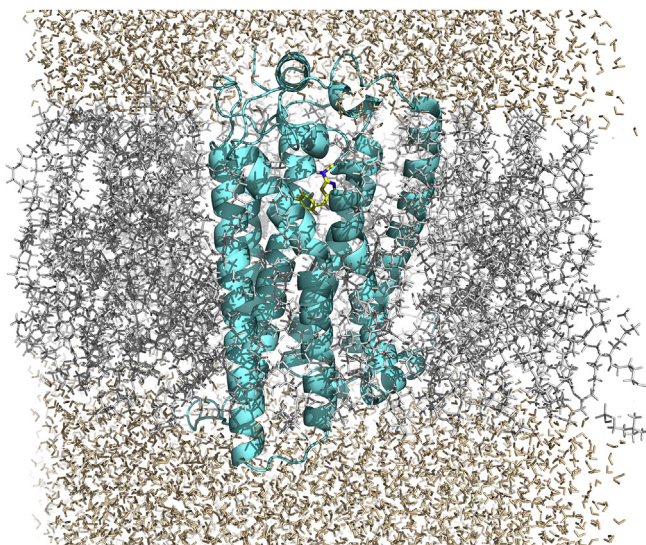


Fig. 9. The alignment between profiles prepared for S1P<sub>1</sub> receptor and E protein using the known secondary structural information. The profiles were generated using their closely related sequences extracted from NCBI. As seen, E protein can be aligned with the 5th transmembrane helix of S1P<sub>1</sub> receptor. The alignment was performed using ClustalW program implemented in ClustalX (version 2.0).



**Fig. 10.** S1P<sub>1</sub> receptor (cartoon representation in cyan) with docked compound 400 (stick representation) embedded in lipid bilayer membrane containing 130 DOPC molecules (gray) and 37H<sub>2</sub>O per each lipid molecule. KCl was added to neutralize the system. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 9**

S1P<sub>1</sub> receptor binding energies calculated for the ligands identified from PubChem search using MMGBSA/MMPBSA algorithms.

MMGBSA						
Ligand	EEL	VDW	$\Delta G_{\text{gas}}$	$\Delta G_{\text{sol}}$	$\Delta G_{\text{binding}}$	SEM $_{\Delta G_{\text{binding}}}$
400	-15.91	-50.22	-66.13	19.95	-46.18	0.04
452	-16.25	-44.54	-60.80	21.67	-39.31	0.04
61	4.65	-48.57	-43.91	15.53	-28.38	0.03
798	-22.21	-47.28	-69.29	26.68	-42.81	0.03
799	-19.55	-48.85	-68.40	22.97	-45.43	0.03
MMPBSA						
Ligand	EEL	VDW	$\Delta G_{\text{gas}}$	$\Delta G_{\text{sol}}$	$\Delta G_{\text{binding}}$	SEM $_{\Delta G_{\text{binding}}}$
400	-15.91	-50.23	-66.14	56.40	-9.75	0.07
452	-16.25	-44.54	-60.80	54.47	-6.33	0.07
61	4.66	-48.57	-43.9155	72.89	28.98	0.07
798	-22.21	-47.28	-69.49	61.00	-8.49	0.04
799	-19.55	-48.85	-68.40	65.20	-3.20	0.05

EEL; Electrostatic energy, VDW:Van der Waals interaction,  $\Delta G_{\text{gas}}$ :binding energy difference in gas state,  $\Delta G_{\text{sol}}$ : binding energy difference in solvated state, SEM $_{\Delta G_{\text{binding}}}$ : standard error of Mean of  $\Delta G_{\text{binding}}$ .

The results of this work can be useful for designing novel potent and selective S1P<sub>1</sub> receptor agonists.

### Declaration of competing interest

The authors declare no conflict of interest.

### Acknowledgements

This work was supported by Research Office of Tabriz University of Medical Sciences [grant number 62634]. The authors would like to thank Biotechnology Research Center for providing computational facilities.

### References

- [1] V. Brinkmann, Sphingosine 1-phosphate receptors in health and disease: mechanistic insights from gene deletion studies and reverse pharmacology, *Pharmacol. Ther.* 115 (2007) 84–105.

- [2] S. Crosignani, A. Bombrun, D. Covini, M. Maio, D. Marin, A. Quattropani, D. Swinnen, D. Simpson, W. Sauer, B. Francon, T. Martin, Y. Cambet, A. Nichols, I. Martinou, F. Burgat-Charvillon, D. Rivron, C. Donini, O. Schott, V. Elgert, L. Novo-Perez, P.A. Vitte, J.F. Arrighi, Discovery of a novel series of potent S1P<sub>1</sub> agonists, *Bioorg. Med. Chem. Lett* 20 (2010) 1516–1519.
- [3] M.A. Hanson, C.B. Roth, E. Jo, M.T. Griffith, F.L. Scott, G. Reinhart, H. Desale, B. Clemons, S.M. Cahalan, S.C. Schuerer, M.G. Sanna, G.W. Han, P. Kuhn, H. Rosen, R.C. Stevens, Crystal structure of a lipid G protein-coupled receptor, *Science* 335 (2012) 851–855.
- [4] H. Rosen, P.J. Gonzalez-Cabrera, M.G. Sanna, S. Brown, Sphingosine 1-phosphate receptor signaling, *Annu. Rev. Biochem.* 78 (2009) 743–768.
- [5] D. D'Ambrosio, M.S. Freedman, J. Prinz, Ponesimod, a selective S1P<sub>1</sub> receptor modulator: a potential treatment for multiple sclerosis and other immune-mediated diseases, *Ther. Adv. Chron. Dis.* 7 (2016) 18–33.
- [6] D.J. Buzard, T.O. Schrader, X. Zhu, J. Lehmann, B. Johnson, M. Kasem, S.H. Kim, A. Kawasaki, L. Lopez, J. Moody, S. Han, Y. Gao, J. Edwards, J. Barden, J. Thatte, J. Gatlin, R.M. Jones, Design and synthesis of new tricyclic indoles as potent modulators of the S1P<sub>1</sub> receptor, *Bioorg. Med. Chem. Lett* 25 (2015) 659–663.
- [7] M.H. Bolli, S. Abele, C. Binkert, R. Bravo, S. Buchmann, D. Bur, J. Gatfield, P. Hess, C. Kohl, C. Mangold, B. Mathys, K. Menyhart, C. Muller, O. Naylor, M. Scherz, G. Schmidt, V. Sippel, B. Steiner, D. Strasser, A. Treiber, T. Weller, 2-imino-thiazolidin-4-one derivatives as potent, orally active S1P<sub>1</sub> receptor agonists, *J. Med. Chem.* 53 (2010) 4198–4211.
- [8] D.J. Buzard, S. Han, L. Lopez, A. Kawasaki, J. Moody, L. Thoresen, B. Ullman, J. Lehmann, I. Calderon, X. Zhu, T. Gharbaoui, D. Sengupta, A. Krishnan, Y. Gao, J. Edwards, J. Barden, M. Morgan, K. Usmani, C. Chen, A. Sadeque, J. Thatte, M. Solomon, L. Fu, K. Whelan, L. Liu, H. Al-Shamma, J. Gatlin, M. Le, C. Xing, S. Espinola, R.M. Jones, Fused tricyclic indoles as S1P<sub>1</sub> agonists with robust efficacy in animal models of autoimmune disease, *Bioorg. Med. Chem. Lett* 22 (2012) 4404–4409.
- [9] R. Milo, E. Kahana, Multiple sclerosis: geoeidemiology, genetics and the environment, *Autoimmun. Rev.* 9 (2010) A387–A394.
- [10] M. Sanford, Fingolimod: a review of its use in relapsing-remitting multiple sclerosis, *Drugs* 74 (2014) 1411–1433.
- [11] V. Brinkmann, FTY720 (fingolimod) in Multiple Sclerosis: therapeutic effects in the immune and the central nervous system, *Br. J. Pharmacol.* 158 (2009) 1173–1182.
- [12] J.J. Clemens, M.D. Davis, K.R. Lynch, T.L. Macdonald, Synthesis of benzimidazole based analogues of sphingosine-1-phosphate: discovery of potent, subtype-selective S1P<sub>4</sub> receptor agonists, *Bioorg. Med. Chem. Lett* 14 (2004) 4903–4906.
- [13] J.J. Clemens, M.D. Davis, K.R. Lynch, T.L. Macdonald, Synthesis of 4(5)-phenylimidazole-based analogues of sphingosine-1-phosphate and FTY720: discovery of potent S1P<sub>1</sub> receptor agonists, *Bioorg. Med. Chem. Lett* 15 (2005) 3568–3572.
- [14] A. Artese, S. Cross, G. Costa, S. Distinto, L. Parrotta, S. Alcaro, F. Ortuso, G. Cruciani, Molecular interaction fields in drug discovery: recent advances and future perspectives, *Wiley, Interdiscipl. Rev. Comput. Mol. Sci.* 3 (2013) 594–613.
- [15] G. Ermondi, G. Caron, GRIND-based 3D-QSAR to predict inhibitory activity for similar enzymes, OSC and SHC, *Eur. J. Med. Chem.* 43 (2008) 1462–1468.
- [16] A.A. Alizadeh, M. Hamzeh-Mivehroud, B. Sokouti, S. Dastmalchi, An alignment-independent 3D-QSAR study on series of hydroxamic acid-based tumor necrosis factor- $\alpha$  converting enzyme inhibitors, *J. Chemom.* 30 (2016) 537–547.
- [17] G. Ermondi, G. Caron, I.G. Pintos, M. Gerbaldo, M. Perez, D.I. Perez, Z. Gandara, A. Martinez, G. Gomez, Y. Fall, An application of two MIFs-based tools (Volsurf+ and Pentacle) to binary QSAR: the case of a palinurin-related data set of non-ATP competitive glycogen synthase kinase 3beta (GSK-3beta) inhibitors, *Eur. J. Med. Chem.* 46 (2011) 860–869.
- [18] B. Jafari, M. Hamzeh-Mivehroud, A.A. Alizadeh, M. Sharifi, S. Dastmalchi, An alignment-independent 3D-QSAR study of FGFR2 Tyrosine kinase inhibitors, *Adv. Pharmaceut. Bull.* 7 (2017) 409–418.
- [19] F.L. Scott, B. Clemons, J. Brooks, E. Brahmachary, R. Powell, H. Dedman, H.G. Desale, G.A. Timony, E. Martinborough, H. Rosen, E. Roberts, M.F. Boehm, R.J. Peach, Ozanimod (RPC1063) is a potent sphingosine-1-phosphate receptor-1 (S1P<sub>1</sub>) and receptor-5 (S1P<sub>5</sub>) agonist with autoimmune disease-modifying activity, *Br. J. Pharmacol.* 173 (2016) 1778–1792.
- [20] N.L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms, *J. Am. Chem. Soc.* 99 (1977) 8127–8134.
- [21] J.J.P. Stewart, Optimization of parameters for semiempirical methods I. Method, *J. Comput. Chem.* 10 (1989) 209–220.
- [22] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Res.* 31 (2003) 3381–3385.
- [23] V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (2010) 12–21.
- [24] D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, *Methods Enzymol.* 277 (1997) 396–404.
- [25] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [26] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a

- genetic algorithm with a description of desolvation, *J. Mol. Biol.* 245 (1995) 43–53.
- [27] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* 267 (1997) 727–748.
- [28] J.A. GRANT, M.A. GALLARDO, B.T. PICKUP, A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape, *J. Comput. Chem.* 17 (1996) 1653–1666.
- [29] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, *J. Med. Chem.* 28 (1985) 849–857.
- [30] A. Duran, G.C. Martínez, M. Pastor, Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in Molecular Interaction Fields, *J. Chem. Inf. Model.* 48 (2008) 1813–1823.
- [31] M. Pastor, G. Cruciani, I. McLay, S. Pickett, S. Clementi, GRIND-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors, *J. Med. Chem.* 43 (2000) 3233–3243.
- [32] K. Roy, S. Kar, P. Ambure, On a simple approach for determining applicability domain of QSAR models, *Chemometr. Intell. Lab. Syst.* 145 (2015) 22–29.
- [33] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and clustal X version 2.0, *Bioinformatics* 23 (2007) 2947–2948.
- [34] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174.
- [35] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, *J. Comput. Chem.* 23 (2002) 1623–1641.
- [36] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem.* 97 (1993) 10269–10280.
- [37] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem.* 29 (2008) 1859–1865.
- [38] C.J. Dickson, B.D. Madej, Å.A. Skjevik, R.M. Betz, K. Teigen, I.R. Gould, R.C. Walker, Lipid14: the amber lipid force field, *J. Chem. Theory Comput.* 10 (2014) 865–879.
- [39] O. Nicolotti, D. Gadaleta, G.F. Felice Mangiatordi, M. Catto, A. Carotti, Applicability domain for QSAR models: where theory meets reality, *Int. J. Quant. Struct.-Prop. Relatsh.* 1 (2016) 45–63.
- [40] J. Fettel, B. Kuhn, N.A. Guillen, D. Surun, M. Peters, R. Bauer, et al., Sphingosine-1-phosphate (S1P) induces potent anti-inflammatory effects in vitro and in vivo by S1P receptor 4-mediated suppression of 5-lipoxygenase activity, *FASEB J.* 33 (2019) 1711–1726.
- [41] A.L. Parrill, D. Wang, D.L. Bautista, J.R. Van Brocklyn, Z. Lorincz, D.J. Fischer, D.L. Baker, K. Liliom, S. Spiegel, G. Tigyi, Identification of Edg1 receptor residues that recognize sphingosine 1-phosphate, *J. Biol. Chem.* 275 (2000) 39379–39384.
- [42] D.A. Wang, Z. Lorincz, D.L. Bautista, K. Liliom, G. Tigyi, A.L. Parrill, A single amino acid determines lysophospholipid specificity of the S1P1 (EDG1) and LPA1 (EDG2) phospholipid growth factor receptors, *J. Biol. Chem.* 276 (2001) 49213–49220.
- [43] P.J. Gonzalez-Cabrera, E. Jo, M.G. Sanna, S. Brown, N. Leaf, D. Marsolais, M.T. Schaeffer, J. Chapman, M. Cameron, M. Guerrero, E. Roberts, H. Rosen, Full pharmacological efficacy of a novel S1P1 agonist that does not require S1P-like headgroup interactions, *Mol. Pharmacol.* 74 (2008) 1308–1318.
- [44] S.C. Schurer, S.J. Brown, P.J. Gonzalez-Cabrera, M.T. Schaeffer, J. Chapman, E. Jo, P. Chase, T. Spicer, P. Hodder, H. Rosen, Ligand-binding pocket shape differences between sphingosine 1-phosphate (S1P) receptors S1P1 and S1P3 determine efficiency of chemical probe identification by ultrahigh-throughput screening, *ACS Chem. Biol.* 3 (2008) 486–498.