EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

Original article:

COMPUTER-GUIDED DESIGN OF NOVEL NITROGEN-BASED HETEROCYCLIC SPHINGOSINE-1-PHOSPHATE (S1P) ACTIVATORS AS OSTEOANABOLIC AGENTS

Rattanawan Tangporncharoen[1], Chuleeporn Phanus-Umporn[2], Supaluk Prachayasittikul[3], Chanin Nantasenamat[4], Veda Prachayasittikul[3*], Aungkura Supokawej[1*]

[1] Department of Clinical Microscopy, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[2] Department of Community Medical Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[3] Center for Research Innovation and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[4] Streamlit Inc., San Francisco, CA 94121, USA

* **Corresponding authors:** Aungkura Supokawej, Department of Clinical Microscopy, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand. Phone: +66(0)2 441 4371-5 Ext. 2723; Fax: +66(0)2 441 4380; E-mail: aungkura.jer@mahidol.ac.th
Veda Prachayasittikul, Center for Research Innovation and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand. Phone: +66(0)2 441 4371-5 Ext. 2726; Fax: +66(0)2 441 4380; E-mail: veda.pra@mahidol.ac.th

https://dx.doi.org/10.17179/excli2024-7214

ABSTRACT

Osteoanabolic agents, or drugs that promote bone formation, have gained considerable attention for osteoporosis management due to their curative and preventive potentials. Sphingosine-1-phosphate receptor 2 (S1PR2) is an attractive drug target, in which its activation leads to osteogenesis-promoting effect. Nitrogen-containing heterocyclic scaffolds (i.e., quinoxaline and indole) are promising pharmacophores possessing diverse bioactivities and were reported as S1PR2 activators. Quantitative structure-activity relationship (QSAR) modeling is a computational approach well-known as a fundamental tool for facilitating successful drug development. This study demonstrates the discovery of new S1PR2 activators using computational-driven rational design. Herein, an original dataset of nitrogen-containing S1PR2 activators was collected from ChEMBL database. The retrieved dataset was separated into two datasets according to their core scaffolds (i.e., quinoxaline and indole). QSAR modeling was performed using multiple linear regression (MLR) algorithm to successfully obtain two models with good predictive performance. The constructed models also revealed key properties playing essential roles for potent S1PR2 activation, such as Van der Waals volume (R2v+ and E3v), mass (MATS5m and Km), electronegativity (H3e), and number of 5-membered rings (nR05). Subsequently, the constructed models were further employed to guide rational design and predict S1PR2 activating effects of an additional set of 752 structurally modified compounds. Most of the modified compounds were predicted to have higher potency than their parents, and a set of promising potent newly designed compounds was highlighted. Additionally, drug-likeness prediction was performed to reveal that most of the newly designed compounds are druggable compounds with possibility for further development.

Keywords: Sphingosine-1-phosphate activators, Quantitative Structure-Activity Relationship, computer-aided drug design, osteoanabolic, quinoxalines, indoles

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

INTRODUCTION

Aging drives many degenerative conditions and is the most concerning health issue globally (United Nations Department of Economic Social Affairs, 2021). Osteoporosis is a degenerative condition typically found among elders. It is initiated by an imbalanced bone remodeling leading to the loss of bone density and porous bone appearance with increased risk of fragility fractures (Akkawi and Zmerly, 2018; Ensrud and Crandall, 2017). Pharmacotherapeutics for osteoporosis are mainly aimed to increase bone mineral density either by preventing bone resorption (antiresorptive agents) or facilitating osteogenesis (osteoanabolic agents). The osteoanabolic agents have gained greater interest due to their healing-promoting effects (Kostenuik et al., 2023). Several classes of anti-osteoporotic drugs are clinically available (i.e., bisphosphonates, selective estrogen receptor modulators, sclerostin inhibitors, and receptor activator of nuclear factor kappa-β ligand inhibitors (Akkawi and Zmerly, 2018; Ensrud and Crandall, 2017; Tu et al., 2018). However, their adverse effects and complications are still concerning (Black et al., 2020; Gilbert et al., 2022; Wotton et al., 2019). Accordingly, the discovery of novel osteoanabolic agents for osteoporosis management is of great interest.

Sphingosine-1-phosphate (S1P) plays crucial roles in regulating many biological systems including angiogenesis, inflammation, neurogenesis (Mendelson et al., 2014; Takuwa et al., 2012), osteogenesis (Cao et al., 2019; Sartawi et al., 2017), and bone remodeling (Grewe et al., 2022). Therefore, their receptors are attractive drug targets for several clinical indications (Baumruker et al., 2007; Yazdi et al., 2020). Among other subtypes, S1P receptor 2 (S1PR2) was noted for its role in bone remodeling. An activation of S1PR2 was reported to induce the expression of a key transcriptional factor (RUNX2) playing a role in osteoblast differentiation (Higashi et al., 2016; Sato et al., 2012), increase osteoblast differentiation markers at mRNA level (Matsuzaki et al., 2022), and improve bone structural parameters in several *in vivo* models (Matsuzaki et al., 2022; Weske et al., 2018, 2019). Accordingly, targeting the activation of S1PR2 is a promising strategy for managing bone diseases.

Nitrogen-based heterocyclic compounds (i.e., quinoxaline and indole) are promising pharmacophores for drug discovery due to their various bioactivities. Quinoxalines were reported as antimicrobial, anticancer, anti-inflammatory and antidepressant agents (Pereira et al., 2015). Similarly, indoles were noted for their anti-inflammatory, anticancer, antimicrobial, antidiabetic, and antioxidant activities (Sravanthi and Manju, 2016). Notably, these nitrogen-containing compounds were reported to take parts in bone remodeling via modulating the differentiations and functions of both osteoclast and osteoblast (Ha et al., 2021; Yu et al., 2015; Zhou et al., 2021). Therefore, these pharmacophores are attractive scaffolds for the discovery of new anti-osteoporotic drugs.

Computational (*in silico*) approaches have been employed as facilitating tools in drug development for decades (Prachayasittikul et al., 2015) due to their benefits in reducing late-stage failures, increasing success rate, and saving time (Vemula et al., 2023). Quantitative Structure-Activity Relationship (QSAR) modeling is a computational method widely used to elucidate the relationships between chemical properties of the compounds and their biological activities (Nantasenamat et al., 2009; Roy et al., 2015b). Among others, the multiple linear regression (MLR) algorithm is well-known for its interpretable nature that suitably reveals informative knowledge for guiding the effective design of new derivatives. Successful QSAR-driven rational drug designs have been reported for diverse activities by our group (Phanus-umporn et al., 2020; Pingaew et al., 2022; Prachayasittikul et al., 2017; Pratiwi et al., 2019; Worachartcheewan et al., 2020). Drug-likeness is one of the key properties determining successful drug development, therefore, *in silico* drug-likeness prediction is widely included in the initial phases of development to improve the success rate (Ekins et al., 2023).

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

To date, QSAR studies regarding the S1PR modulating compounds are still limited. Herein, previously reported quinoxaline-based and indole-based S1PR2 activators (Pubchem, 2004a, b, c) (Figure 1) were retrieved as input datasets for QSAR modeling using MLR algorithms. The constructed models were further applied to guide the design of new structurally modified compounds. Subsequently, the newly designed compounds were predicted for their S1PR2 activating effects as well as drug-likeness. SAR analysis was also performed to reveal key beneficial knowledge. Finally, a set of promising novel compounds was highlighted for further potential development.

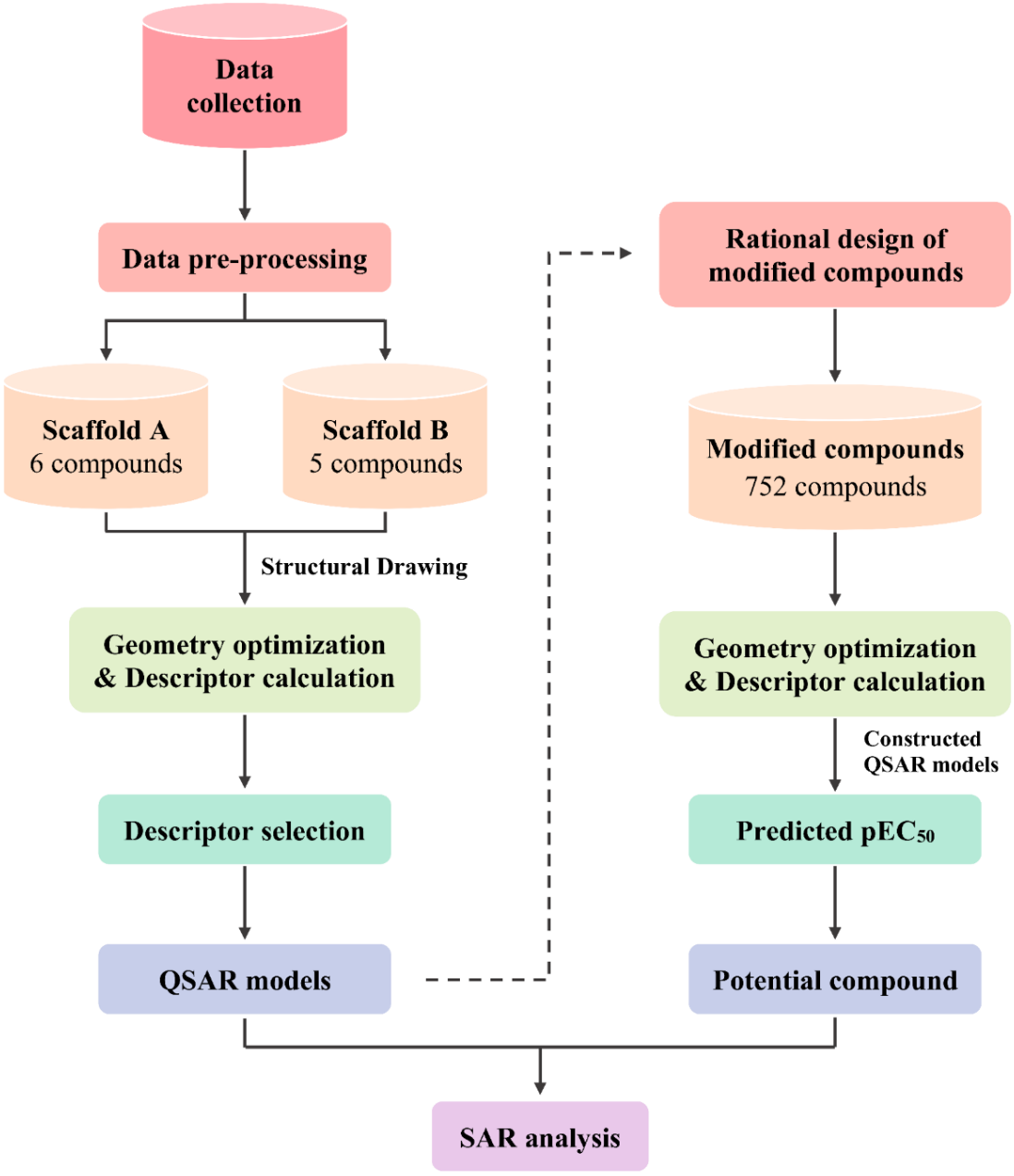

Figure 1: Overview workflow of the study

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

MATERIALS AND METHODS

Data collection

An original set of quinoxaline- and indole-based S1PR2 activators was obtained from the ChEMBL database (Mendez et al., 2018) and was curated (Fourches et al., 2010) by the following steps (i) removal of inorganic compounds (i.e., salts and mixtures), (ii) structural validation and cleaning, (iii) normalization of specific chemotypes, (iv) deletion of duplicates, and (v) final checking to obtain the final curated data set containing 11 active compounds along with their bioactivity values, represented by the half maximal effective concentration (EC_{50}) (Neubig et al., 2003). EC_{50} is a concentration that increases the S1P activity by 50 %. The EC_{50} values were converted into pEC_{50} values by taking the negative logarithm based 10 to normalize the data points. According to the compound's core structure, the dataset was separately preprocessed into two final datasets (i.e., scaffold A = 6 quinoxaline-based compounds and scaffold B = 5 indole-based compounds), Figure 2.

Geometry optimization

Chemical structures of compounds in SMILES format were converted into MOL format using MarvinSketch (ChemAxon Ltd., 2018). All structures were geometrically optimized to obtain the lowest energy conformers by density functional theory (DFT) computation with Becke's three-parameter Lee–Yang–Parr hybrid functional (B3LYP) and 6-31g(d) basis set using Gaussian 09 software (Frisch et al., 2009). The optimized structures were subsequently used as input files for descriptor calculation.

Molecular descriptors calculation

Molecular descriptors are numerical values used to represent characteristics of the compounds in terms of connectivity, constitution, and physicochemical properties (Nantasenamat et al., 2009). A set of 13 quantum chemical descriptors was calculated using an in-house developed script (i.e., mean absolute atomic charge (Q_m), molecular dipole moment (μ), electronegativity (χ), total energy (E_T), electron affinity (EA), ionization potential (IP), electron ionization (EI), highest occupied molecular orbital energy (HOMO), lowest unoccupied molecular orbital energy (LUMO), energy difference of HOMO and LUMO ($HOMO\text{-}LUMO_{gap}$), absolute hardness (η), softness (S), and electrophilicity (ω). Additionally, the optimized structures were used as input files for calculation of molecular descriptor using Dragon 5.5 software (Mauri et al., 2006) to obtain a set of 3,224 molecular descriptors, comprising 48 constitutional descriptors, 119 topological descriptors, 47 walk and path counts, 33 connectivity indices, 47 information indices, 96 2D autocorrelations, 107 edge adjacency indices, 64 Burden eigenvalues, 21 topological charge indices, 44 eigenvalue-based indices, 41 Randic molecular profiles, 74 geometrical descriptors, 150 RDF descriptors, 160 3D-MoRSE descriptors, 99 WHIM descriptors, 197 GETAWAY descriptors, 154 functional group counts, 120 atom-centered fragments, 14 charge descriptors, 29 molecular properties, 780 2D binary fingerprints, and 780 2D frequency fingerprints.

Descriptor selection

Descriptor selection is a process to select a set of informative descriptors from a large set of calculated descriptors to be used as final predictors. Correlation-based feature selection was initially performed to filter a set of descriptor variables that are highly correlated with bioactivity of the compounds. Pearson's correlation coefficient (r) values for each pair of descriptor value and bioactivity value (pEC_{50}) were calculated. Descriptors with $|r| \geq 0.8$ were considered as highly correlated predictors and were selected for further selection process using stepwise MLR in PASW Statistics 18 software (SPSS Inc, 2009) and M5 method in WEKA software (Frank et al., 2016) to obtain a final set of informative descriptors. Finally, the values of selected descriptors along with the pEC_{50} values were used to prepare final datasets for model construction.

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

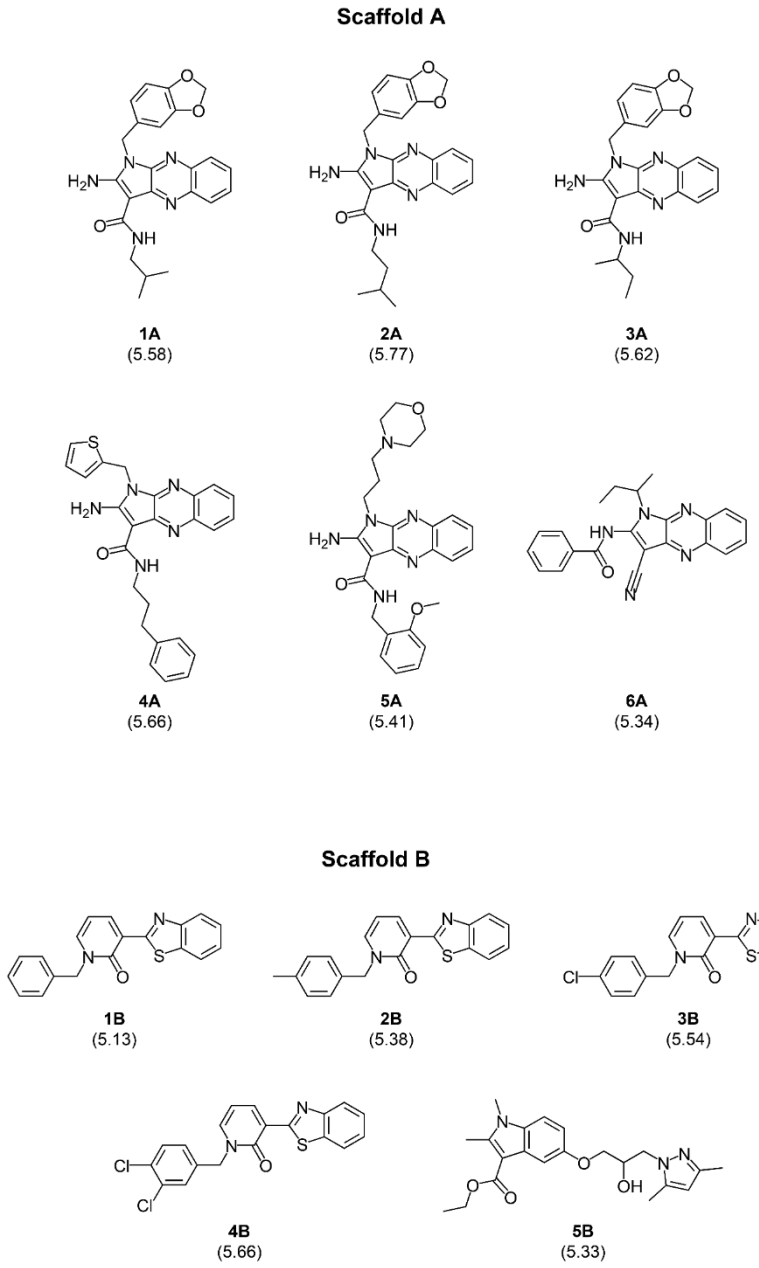

Figure 2: Chemical structures of quinoxaline-based (**Scaffold A**: 6 compounds) and indole-based (**Scaffold B**: 5 compounds) activators. Experimental pEC_{50} values of the compounds are presented in parentheses.

Model construction

Two final datasets were used as input files for construction of QSAR models by WEKA software 3.8.4 (Frank et al., 2016) using MLR algorithm. The MLR model represents a linear relationship between multiple descriptors (X variables) and bioactivity (Y variable) as shown in equation (1).

$$Y = B_0 + \sum B_n X_n \quad (1)$$

where Y is the pEC_{50} of compounds, B_0 is the Y-intercept and B_n are the regression coefficients of descriptors (X_n).

Data sampling and model validation

The dataset was divided into training set and testing set by leave-one-out cross validation (LOO-CV). Since the datasets contain small samples with less than 50 compounds, the LOO-CV sampling method is reliable for model validation (Roy et al., 2015a). The training set is used to build the model whereas the testing set was used to validate the predictive performance of the built model. For each round, one sample was excluded from the whole dataset (N) to be used as a

testing set (in which its activity was predicted using the trained model) whereas the remaining samples (N-1) were used as a training set to train and construct the model. The same process was repeated until every compound in the dataset was selected to be used as a testing set (Sammut and Webb, 2011).

Assessment of model performance

The constructed models were assessed for their predictive performance using two statistical parameters i.e., correlation coefficient (R) and root mean square error (RMSE). The first parameter reflects predictive correlation of the model whereas the second one represents predictive error (Rácz et al., 2015).

Application of the constructed models for guiding the design of modified compounds

To increase structural diversity, a set of 752 structurally modified compounds was rationally designed based on key descriptors presented in the constructed QSAR models. All newly designed compounds were undergone the same processes as their parent compounds (e.g., drawing, geometrical optimization, and descriptor calculation) to obtain their informative descriptor values, and their S1P activities (pEC_{50} values) were subsequently predicted using the constructed models.

In silico drug-likeness prediction

The newly designed compounds were predicted for their drug-likeness using *in silico* web-based tool SwissADME (Daina et al., 2017). Chemical structures in SMILES format were uploaded for prediction. The drug-likeness of the compounds was assessed based on Lipinski's rule of five (LRo5).

RESULTS AND DISCUSSION

QSAR modeling

Two final datasets of 6 quinoxaline-based (scaffold A) and 5 indole-based (scaffold B) S1PR2 activators were prepared for construction of two QSAR models, Figure 2. Compounds were preprocessed to obtain their descriptor values. Bioactivity values were normalized into pEC_{50} values. Descriptor selection was performed to select a final set of 6 informative descriptors for construction of scaffold A model (4 descriptors: R2v+, MATS5m, nR05, and Km) and scaffold B model (2 descriptors: H3e and E3v), Table 1. Dataset and predicted activities of scaffolds A and scaffold B models are provided in Tables 2 and 3.

Two QSAR models were successfully constructed using MLR (i.e., scaffold A and scaffold B models). Two built models provided acceptable predictive performance as shown by their validated statistical values for both training and testing sets (R_{tr} = 0.9667-0.9997, $RMSE_{tr}$ = 0.0046-0.0375, Q_{cv} = 0.7902-0.9989, and $RMSE_{cv}$ = 0.0093-0.1057, Table 4). Comparative plots of experimental pEC_{50} versus predicted pEC_{50} values are provided in Figure 3.

From scaffold A, four descriptors play roles as predictors for the activity of quinoxaline-based activators. According to regression coefficient, the most influential descriptor was noted to be a van der Waals volume descriptor (R2v+) followed mass descriptors (MATS5m and Km), and number of five-membered ring (nR05), respectively (Table 4). The high positive values of R2v+ and MATS5m together with low positive or high negative values of the nR05 and Km are required to provide potent activity (high pEC_{50} value). For scaffold B, the van der Waals volume descriptor (E3v) also played major influence on S1PR activating effect, followed by the electronegativity (H3e) descriptor (Table 4). Herein, both predictors possessed positive regression coefficient (R) values suggesting their positive effects on the pEC_{50} values.

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

Table 1: Definitions of informative descriptors used for QSAR modeling

	Descriptor	Definition	Type
Scaffold A	*R2v+*	R maximal autocorrelation of lag 2 / weighted by van der Waals volume	GETAWAY descriptors
	MATS5m	Moran autocorrelation of lag 5 weighted by mass	2D autocorrelations
	nR05	number of 5-membered rings	Ring descriptors
	Km	global shape index / weighted by mass	WHIM descriptors
Scaffold B	*H3e*	H autocorrelation of lag 3 / weighted by Sanderson electronegativity	GETAWAY descriptors
	E3v	3rd component accessibility directional WHIM index / weighted by van der Waals volume	WHIM descriptors

Table 2: Dataset and predicted activities of scaffold A compounds (**1A** – **6A**)

		Scaffold A					
		1A	**2A**	**3A**	**4A**	**5A**	**6A**
Descriptor	R2v+	0.034	0.035	0.032	0.035	0.028	0.027
	MATS5m	-0.026	0.007	0.008	0.008	-0.158	-0.241
	nR05	2	2	2	2	1	1
	Km	0.494	0.436	0.486	0.518	0.579	0.692
pEC_{50} value	Experimental	5.58	5.77	5.62	5.66	5.41	5.34
	Predicted	5.65	5.72	5.61	5.66	5.44	5.31

Table 3: Dataset and predicted activities of scaffold B compounds (**1B – 5B**)

		Scaffold B				
		1B	**2B**	**3B**	**4B**	**5B**
Descriptor	H3e	2.613	2.697	2.74	2.832	2.723
	E3v	0.131	0.145	0.159	0.155	0.134
pEC_{50} value	Experimental	5.13	5.38	5.54	5.66	5.33
	Predicted	5.13	5.37	5.54	5.66	5.33

Table 4: Summary of predictive performance of the constructed QSAR models

	QSAR model	**Training set**		**LOO-CV**	
		R	**RMSE**	**Q**	**RMSE**
Scaffold A	pEC_{50} = 25.0339 (*R2v+*) + 0.1487 (*MATS5m*) - 0.0246 (*nR05*) - 0.7551 (*Km*) + 5.2207	0.9667	0.0375	0.7902	0.1057
Scaffold B	pEC_{50} = 1.5949 (*H3e*) + 7.4202 (*E3v*) - 0.0061	0.9997	0.0046	0.9989	0.0093

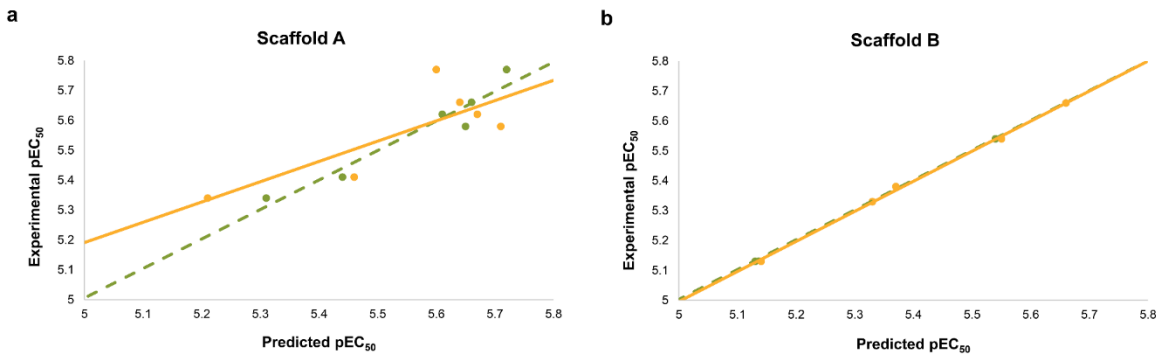

Figure 3: Comparative plots of experimental pEC_{50} and predicted pEC_{50} values from two QSAR models. **a**: Scaffold A (quinoxaline-based compounds), **b**: Scaffold B (indole-based compounds). Plots of training set are presented as green circles and dotted regression lines.

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

Application of the QSAR models for rational design of modified compound series

Key descriptors of the constructed models were further used to guide the design of new derivatives. Modification strategies are conceptually depicted in supplementary information section 1.2 and 2.2 (Tables S2 and S5). Finally, two sets of structurally modified compounds were designed (i.e., scaffold A = 635 compounds, scaffold B = 117 compounds). Chemical structures of the newly designed compounds and their predicted pEC_{50} values are provided in supplementary data and supplementary information, Figures S1-S11. In overview, most of the modified compounds exhibit improved predicted activity when compared to their parents.

Structure-Activity Relationship (SAR) analysis

Understanding structure-activity relationship (SAR) is essential for successful drug development (Macalino et al., 2015; Roy et al., 2015b). Herein, SAR analysis was performed to reveal key features influencing S1PR activating effect of both compound series. For each series, a modification template is illustrated to facilitate effective discussion (as shown in the main panels of Figures 4 and 6). Insight analysis of each sub-modified series was provided to suggest a potential strategy for rational design and development of new derivatives with improved properties. Additionally, the most potent compounds of each subseries are summarized in Figures 4, 5, and 6.

From the original quinoxaline-based activators, compound **2A** (pEC_{50} = 5.77) is the most potent compound followed by **4A**, **3A**, **1A**, **5A**, and **6A** (pEC_{50} = 5.66, 5.62, 5.58, 5.41, and 5.34, respectively, Table 2. An increased length of the alkyl chain amide linker at A site on ring D (Figure 4a) could improve activity via increasing MATS5m values (**4A** and **3A**, Table 2). The presence of long-length alkyl chain at A site is required to give high R2V+ and MATS5m values, thereby, high pEC_{50} values of the compounds (as observed for compounds **2A** and **4A**). The aromatic ring (i.e., benzene, and thiophene) nearly attached at B site is also essential for potent activity. Two compounds without (**6A**) or with aromatic ring attached at this position in further distance (**5A**) provided lower nR05 value (nR05 =1) leading to their lesser potency when compared to other compounds (**1A-4A** with nR05 =2), Figure 4a and Table 2. Summarized SAR analysis of the original compounds **1A-6A** is provided in supplementary information, Table S1.

An additional set of 635 modified quinoxaline-based activators were designed. The modified compounds were grouped by their main core structures into three main templates (Figure 4b-d) and their modification strategies are provided in supplementary information Table S2. Summarized SAR analysis of the modified compounds **1A-6A** is provided in supplementary information, Table S3. In overview, the modified series **6A** provided the most potent predicted activities (predicted pEC_{50} = 5.262-5.772, supplementary data, Scaffold A). Notably, the replacement of nitrogen-containing ring D from the five-membered (pyrrole) to six-membered ring (piperidine or pyridine) gave various effects depending on the parent compounds. Improved activities were observed for the six-membered ring modified series **5A-6A**, whereas decreased activities were found for those of the series **1A-4A**. This phenomenon could be due to the decreased nR05 values (supplementary data, Scaffold A).

For the best modified series **6A**, modifications were performed on A, B, and C sites to obtain 287 modified compounds. Most of them provided better activities (predicted pEC_{50} = 5.262-5.772, supplementary data, Scaffold A) than the parent **6A** (experimental pEC_{50} = 5.340, Table 2). Compounds **6A-116** and **6A-115** were the two most potent compounds (predicted pEC_{50}: **6A-116** = 5.772 and **6A-115** = 5.769). The improved activity (higher predicted pEC_{50}) was suggested to be via decreasing the nR05 and Km values while increasing the R2v+ value (supplementary information, Table S3). A set of promising compounds in modified **6A** series is summarized in Figure 5. Some essential key modifications

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

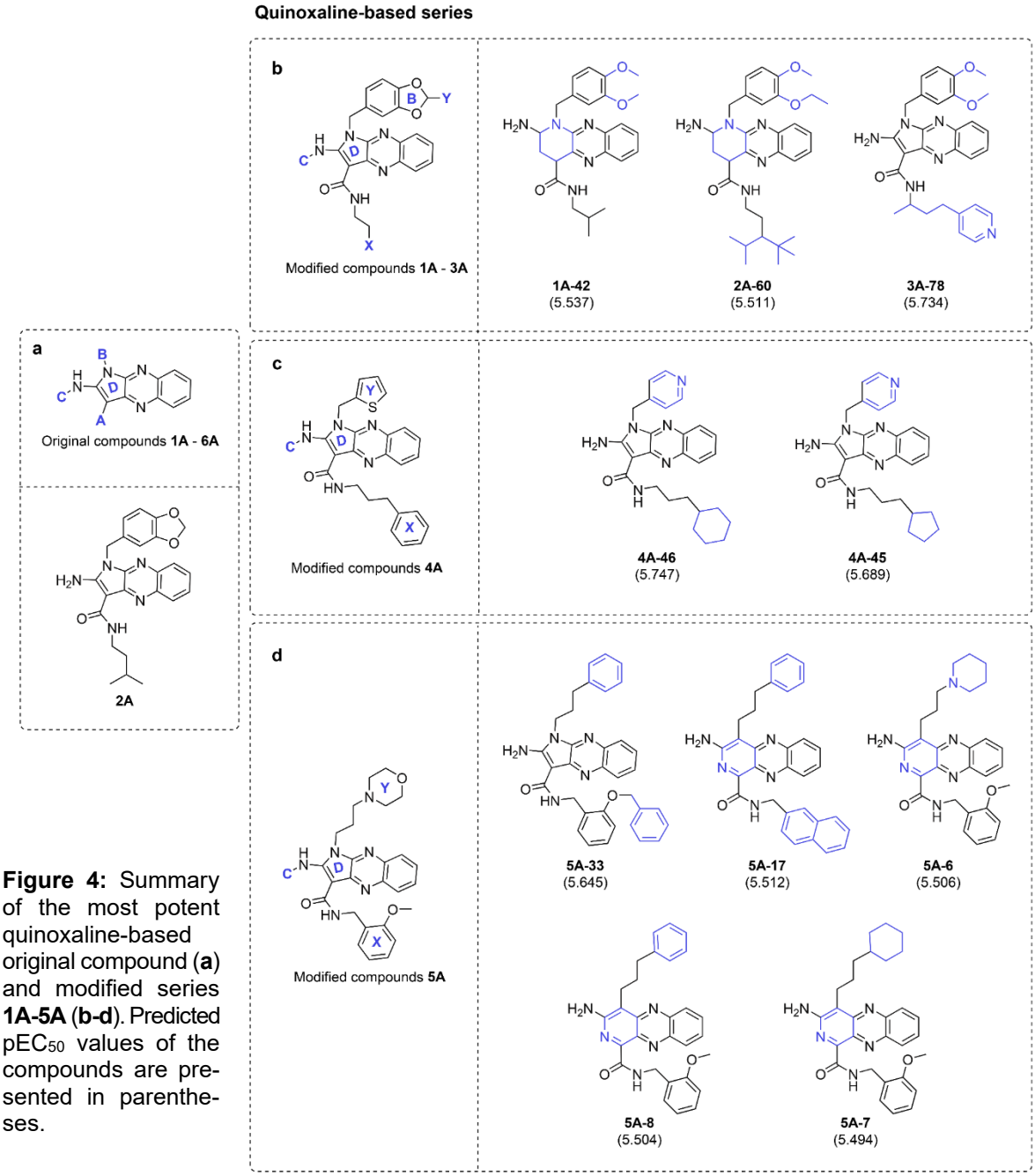

Figure 4: Summary of the most potent quinoxaline-based original compound (**a**) and modified series **1A-5A** (**b-d**). Predicted pEC_{50} values of the compounds are presented in parentheses.

to achieve preferable activity of series **6A** were revealed such as 1) a replacement of ring D with six-membered ring, 2) modification on A site by an insertion of a moiety containing terminal benzene ring, 3) a modification on B site by replacing the branched alkyl chain with -COH and -CN groups as well as 4) a modification on C site by substitutions of $-COCH_3$ or $-COCF_3$ gave compounds with better activities than those substituted with benzoyl group. The better activities were provided when the longer chain was substituted at A position (**6A-116** > **6A-44** and **6A-115** > **6A-43**, Figure 5). Substitution at B position with -COH group provided better activity than with the CN group (**6A-116** > **6A-115** and **6A-44** > **6A-43**, Figure 5). The replacement of COH group on B position of compound **6A-44** (predicted pEC_{50} = 5.767) by the $COCH_3$ group gave compound **6A-45** (predicted pEC_{50} = 5.755) with the lesser potency. Similarly, the bulky ring substitutions on A and B positions impaired activity of compound **6A-140** (predicted pEC_{50} = 5.741) when compared to **6A-116** (predicted pEC_{50} = 5.772). Additionally, the substitution at C position

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

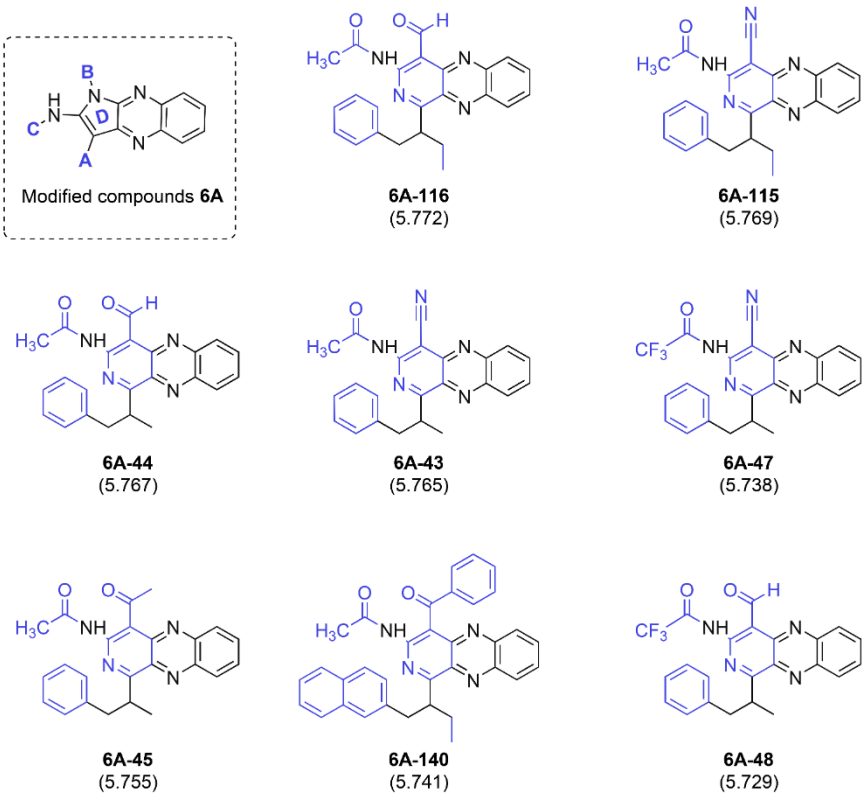

Figure 5: Summary of the most potent compounds from modified series **6A**. Predicted pEC_{50} values of the compounds are presented in parentheses.

with acetyl group ($COCH_3$) gave the more potent compounds rather than the $COCF_3$ group (**6A-43** > **6A-47** and **6A-44** > **6A-48**). The lesser potency of these $COCF_3$ derivatives was suggested to be via a shift of MATS5m value to positive values and the increased Km value (supplementary data, Scaffold A).

Activity of the original indole-based compounds (scaffold B) was ranked as **4B** > **3B** > **2B** > **5B** > **1B** (pEC_{50} = 5.66, 5.54, 5.38, 5.33, and 5.13, respectively, Table 3). The two most potent compounds (**4B** and **3B**) are chloro-containing derivatives indicating that substitutions on ring A with chlorine (Cl) atom provides better improved activity than substitutions with alkyl (**2B**) or hydrogen (**1B**) groups. This could be via the increasing values of both H3e and E3v descriptors (Figure 2 and Table 3). Moreover, an insertion of additional chloro group on the A ring of **3B** gave the **4B** compound with increased activity via increasing the value of van der Waals volume (**4B** > **3B**: H3e value: **4B** = 2.832, **3B** = 2.74, Table 3). Summary SAR analysis of original compounds **1B-5B** is provided in supplementary information, Table S4.

The QSAR scaffold B model was used to guide a design of 117 modified compounds (supplementary information, Table S4) and are subdivided into 5 subseries (**1B-5B**). The most potent compounds of each subseries are summarized in Figure 6b. In overview, the modifications provided compounds with improved activities in every subseries. The substitution on Xa position of ring A with the ring-containing moiety (aliphatic ring: **1B-3**, **2B-3**, and **3B-4** or aromatic ring: **4B-48**) provided the best activities when compared to other types of moieties (i.e., -OH, $-OCH_3$, $-OCF_3$, and alkyl chain). Notably, improvement was observed when the $-CH_3$ group of the parent **2B** (experimental pEC_{50} = 5.38) is replaced by a cyclohexane ring to give compound **2B-3** (predicted pEC_{50} = 8.025). The same improvement was noticed when comparing parent mono-chloro containing compound **3B** (experimental pEC_{50} = 5.54)

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

with the modified **3B-4** (predicted pEC_{50} = 8.269). The improved activity of the **3B-4** is also influenced by a substituted fluorine (F) atom on the Yb moiety of ring B. The type of halogen atom substituted on this Yb position also affected the activity of the compounds as observed for modified subseries **3B**. The best activity is achieved by substitution with fluorine (F: **3B-4**: pEC_{50} = 8.269) followed by chlorine (Cl: **3B-5**: pEC_{50} = 8.223) > bromine (Br: **3B-6**: pEC_{50} = 8.148) > iodine (I:**3B-7**: pEC_{50} = 8.061), supplementary information, Table S5 and supplementary data, Scaffold B). For modified **4B** series, the compounds with $COCF_3$ substitution on Xa position (**4B-48** to **4B-55**: predicted pEC_{50} = 6.275-6.795, supplementary information, Figure S10) provided the best improvement among others. This could be due to their high E3v value (all compounds possess E3v value > 3, greater than those of other series).

For compound **5B** possessing different core structure, the modifications effectively improved activity (as shown by high predicted pEC_{50} = 5.794-8.784, supplementary information, Table S5 and supplementary data, Scaffold B). The best improvements were noted when the X moieties are substituted by alkyl chain (**5B-9**) or aromatic ring (**5B-7** and **5B-11**) moieties, while the Y moieties are replaced with long (**5B-7**) or branched alkyl (**5B-9** and **5B-11**) chain, Figure 6b.

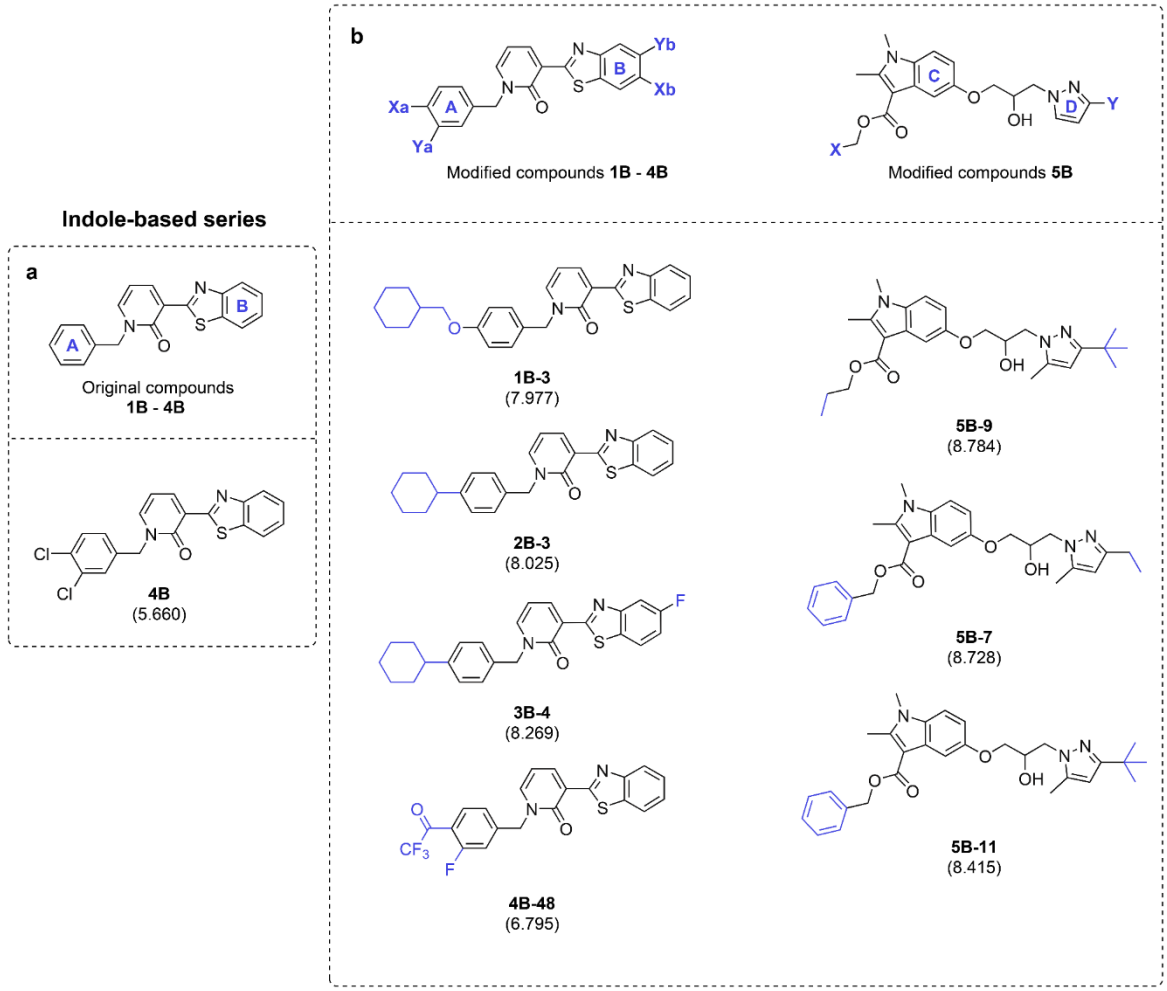

Figure 6: Summary of the most potent indole-based original compounds (**a**), and modified series (**b**). Predicted pEC_{50} values of the compounds are presented in parentheses.

EXCLI Journal 2024;23:818-832 – ISSN 1611-2156
Received: March 26, 2024, accepted: April 17, 2024, published: May 27, 2024

Predicted drug-likeness of the newly designed compounds

Regarding the Lipinski's rule, newly designed compounds were predicted for their drug-like parameters including molecular weight (MW), Ghose-Crippen-Viswanadhan octanol-water partition coefficient (AlogP), number of hydrogen bond donors and acceptors (nHBDon and nHBAcc) (Lipinski et al., 2001). The parameter values were visualized as distribution plots (Figure 7). It was shown that all compounds are accepted by the rule of five, suggesting that they are drug-like molecules with the possibility for further development.

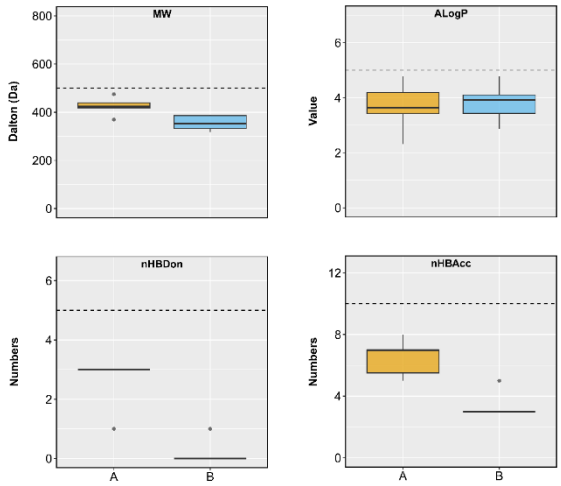

Figure 7: The distribution of Lipinski's descriptors of scaffold A (yellow) and scaffold B (blue). Abbreviation: MW, molecular weight; AlogP, Ghose-Crippen-Viswanadhan octanol-water partition coefficient; nHBDon, number of hydrogen bond donors; nHBAcc, number of hydrogen bond acceptors

CONCLUSION

This study demonstrates the utilization of QSAR modeling for facilitating the effective design of novel nitrogen-based heterocyclic S1PR2 activators. Two QSAR models were successfully constructed and provided good predictive performance (R_{tr} = 0.9667-0.9997, $RMSE_{tr}$ = 0.0046-0.0375, Q_{cv} = 0.7902-0.9989, and $RMSE_{cv}$ = 0.0093-0.1057). The constructed models were further applied to guide the design and activity prediction of an additional set of 752 modified compounds. The QSAR-driven modification on scaffold B outperformed those on scaffold A in improving the activities (predicted pEC_{50} values: modified scaffold B > 6.00, modified scaffold A ≤ 5.80). All newly designed compounds were predicted to be drug-like molecules. Additionally, the models revealed key properties required for potent S1PR2 activities including van der Waals volume (R2v+ and E3v), mass (MATS5m and Km), number of 5-membered rings (nR05), and electronegativity (H3e), which would be beneficial for future design, screening, and development of the related compounds. Finally, a set of 25 newly designed compounds with outstanding predicted activities (scaffold A = 18 and scaffold B = 7 compounds) were highlighted for further development. This work demonstrates an initial step in the discovery journey of the novel osteoanabolic agents, and further studies regarding the synthesis and biological investigations (i.e., *in vitro*, *in vivo*, clinical trials) are essentially recommended.

Supplementary information

Supplementary information and supplementary data are available on EXCLI Journal's website.

Conflict of interest

There are no conflicts to declare.

Acknowledgments

Rattanawan Tangporncharoen is supported by the Royal Golden Jubilee (RGJ) Ph.D. Programme (grant no. PHD/0066/2561), through the National Research Council of Thailand (NRCT), Thailand Research Fund (TRF) and Thailand Science Research and Innovation (TSRI). Veda Prachayasittikul was supported by Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation, Research Grant for New Scholar (grant no. RGNS 64-167), and by Mahidol University (Fundamental Funds: fiscal year 2023 by National Science Research and Innovation Fund (NSRF)). Chuleeporn Phanus-Umporn was supported by Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation (grant no. RGNS 65-145).

REFERENCES

Akkawi I, Zmerly H. Osteoporosis: current concepts. Joints. 2018;6:122-7. doi: 10.1055/s-0038-1660790.

Baumruker T, Billich A, Brinkmann V. FTY720, an immunomodulatory sphingolipid mimetic: translation of a novel mechanism into clinical benefit in multiple sclerosis. Expert Opin Investig Drugs. 2007;16:283-9. doi: 10.1517/13543784.16.3.283.

Black DM, Geiger EJ, Eastell R, Vittinghoff E, Li BH, Ryan DS, et al. Atypical femur fracture risk versus fragility fracture prevention with bisphosphonates. N Engl J Med. 2020;383:743-53. doi: 10.1056/NEJMoa1916525.

Cao Y, Xiao L, Cao Y, Nanda A, Xu C, Ye Q. 3D printed β-TCP scaffold with sphingosine 1-phosphate coating promotes osteogenesis and inhibits inflammation. Biochem Biophys Res Commun. 2019;512: 889-95. doi: 10.1016/j.bbrc.2019.03.132.

ChemAxon Ltd. MarvinSketch [Computer software]. Version 18.10. Budapest: ChemAxon Ltd., 2018. http://www.chemaxon.com.

Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7(1):42717. doi: 10.1038/srep42717.

Ekins S, Lane TR, Urbina F, Puhl AC. *In silico* ADME/tox comes of age: twenty years later. Xenobiotica. 2023:epub ahead of print. doi: 10.1080/00498254.2023.2245049.

Ensrud KE, Crandall CJ. Osteoporosis. Ann Intern Med. 2017;167(3):ITC17-32. doi: 10.7326/AITC201708010.

Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J Chem Inf Model. 2010;50:1189-204. doi: 10.1021/ci100176x.

Frank E, Hall MA, Witten IH. Appendix B - The WEKA workbench. In: Witten IH, Frank E, Hall MA, Pal CJ (eds). Data mining: practical machine learning tools and techniques. 4th ed (pp 553-71). Burlington, MA: Morgan Kaufmann, 2016.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09 [Computer software]. Version A.02. Wallingford, CT: Gaussian Inc, 2009. http://www.gaussian.com.

Gilbert ZA, Muller A, Leibowitz JA, Kesselman MM. Osteoporosis prevention and treatment: the risk of comorbid cardiovascular events in postmenopausal women. Cureus. 2022;14(4):e24117. doi: 10.7759/cureus.24117.

Grewe JM, Knapstein P-R, Donat A, Jiang S, Smit DJ, Xie W, et al. The role of sphingosine-1-phosphate in bone remodeling and osteoporosis. Bone Res. 2022;10 (1):34. doi: 10.1038/s41413-022-00205-0.

Ha MT, Tran PT, Tran HNK, Kim O, Kim JA, Lee J-H, et al. Anti-osteoclastogenic effects of indole alkaloids isolated from Barley (*Hordeum vulgare* Var. *Hexastichon*) Grass. J Agric Food Chem. 2021;69: 12994-3005. doi: 10.1021/acs.jafc.1c04337.

Higashi K, Matsuzaki E, Hashimoto Y, Takahashi-Yanaga F, Takano A, Anan H, et al. Sphingosine-1-phosphate/S1PR2-mediated signaling triggers Smad1/5/8 phosphorylation and thereby induces Runx2 expression in osteoblasts. Bone. 2016;93:1-11. doi: 10.1016/j.bone.2016.09.003.

Kostenuik PJ, Binkley N, Anderson PA. Advances in osteoporosis therapy: focus on osteoanabolic agents, secondary fracture prevention, and perioperative bone health. Curr Osteoporos Rep. 2023;21:386-400. doi: 10.1007/s11914-023-00793-8.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Del Rev. 2001; 46(1):3-26. doi: 10.1016/S0169-409X(00)00129-0.

Macalino SJY, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. Arch Pharm Res. 2015;38:1686-701. doi: 10.1007/s12272-015-0640-5.

Matsuzaki E, Hirose H, Fujimasa S, Yoshimoto S, Yanagi T, Matsumoto K, et al. Sphingosine-1-phosphate receptor 2 agonist induces bone formation in rat apicoectomy and alveolar bone defect model. J Dent Sci. 2022;17:787-94. doi: 10.1016/j.jds.2021.10.004.

Mauri A, Consonni V, Pavan M, Todeschini R. DRAGON software: An easy approach to molecular descriptor calculations. MATCH Commun Math Comput Chem. 2006;56:237-48.

Mendelson K, Evans T, Hla T. Sphingosine 1-phosphate signalling. Development. 2014;141(1):5-9. doi: 10.1242/dev.094805.

Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL:towards direct deposition of bioassay data. Nucleic Acids Res. 2018;47(D1): D930-40. doi: 10.1093/nar/gky1075.

Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. A practical overview of quantitative structure-activity relationship. EXCLI J. 2009;8: 74-88. doi: 10.17877/DE290R-690.

Neubig RR, Spedding M, Kenakin T, Christopoulos A. International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. Pharmacol Rev. 2003;55:597-606. doi: 10.1124/pr.55.4.4.

Pereira JA, Pessoa AM, Cordeiro MNDS, Fernandes R, Prudêncio C, Noronha JP, et al. Quinoxaline, its derivatives and applications: a state of the art review. Eur J Med Chem. 2015;97:664-72. doi: 10.1016/j.ejmech.2014.06.058.

Phanus-umporn C, Prachayasittikul V, Nantasenamat C, Prachayasittikul S, Prachayasittikul V. QSAR-driven rational design of novel DNA methyltransferase 1 inhibitors. EXCLI J. 2020;19:458-75. doi: 10.17179/excli2020-1096.

Pingaew R, Prachayasittikul V, Worachartcheewan A, Thongnum A, Prachayasittikul S, Ruchirawat S, et al. Anticancer activity and QSAR study of sulfur-containing thiourea and sulfonamide derivatives. Heliyon. 2022;8(8):e10067. doi: 10.1016/j.heliyon.2022.e10067.

Prachayasittikul V, Worachartcheewan A, Shoombuatong W, Songtawee N, Simeon S, Prachayasittikul V, et al. Computer-aided drug design of bioactive natural products. Curr Top Med Chem. 2015;15:1780-800. doi: 10.2174/1568026615666150506151101.

Prachayasittikul V, Pingaew R, Worachartcheewan A, Sitthimonchai S, Nantasenamat C, Prachayasittikul S, et al. Aromatase inhibitory activity of 1,4-naphthoquinone derivatives and QSAR study. EXCLI J. 2017;16:714-26. doi: 10.17179/excli2017-309.

Pratiwi R, Prachayasittikul V, Prachayasittikul S, Nantasenamat C. Rational design of novel sirtuin 1 activators via structure-activity insights from application of QSAR modeling. EXCLI J. 2019;18:207-22. doi: 10.17179/excli2019-1274.

Pubchem [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004a. PubChem Bioassay Record for AID 729, Source: The Scripps Research Institute Molecular Screening Center; [cited 2019 June 21]. Available from: https://pubchem.ncbi.nlm.nih.gov/bioassay/729.

Pubchem [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004b. PubChem Bioassay Record for AID 736, Source: The Scripps Research Institute Molecular Screening Center; [cited 2019 June 21]. Available from: https://pubchem.ncbi.nlm.nih.gov/bioassay/736.

Pubchem [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004c. PubChem Bioassay Record for AID 843, Source: The Scripps Research Institute Molecular Screening Center; [cited 2019 June 21]. Available from: https://pubchem.ncbi.nlm.nih.gov/bioassay/843.

Rácz A, Bajusz D, Héberger K. Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters. SAR QSAR Environ Res. 2015;26:683-700. doi: 10.1080/1062936X.2015.1084647.

Roy K, Kar S, Das RN. Chapter 7 - Validation of QSAR models. In: Roy K, Kar S, Das RN (eds). Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment (pp 231-89). Boston, MA: Academic Press, 2015a.

Roy K, Kar S, Das RN. Chapter 11 - SAR and QSAR in drug discovery and chemical design—some examples. In: Roy K, Kar S, Das RN (eds). Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment (pp 427-53). Boston, MA: Academic Press, 2015b.

Sammut C, Webb GI. Leave-one-out cross-validation. In: Sammut C, Webb GI (eds). Encyclopedia of machine learning (pp 600-1). Boston, MA: Springer, 2011.

Sartawi Z, Schipani E, Ryan KB, Waeber C. Sphingosine 1-phosphate (S1P) signalling: Role in bone biology and potential therapeutic target for bone repair. Pharmacol Res. 2017;125:232-45. doi: 10.1016/j.phrs.2017.08.013.

Sato C, Iwasaki T, Kitano S, Tsunemi S, Sano H. Sphingosine 1-phosphate receptor activation enhances BMP-2-induced osteoblast differentiation. Biochem Biophys Res Commun. 2012;423:200-5. doi: 10.1016/j.bbrc.2012.05.130.

SPSS Inc. PASW Statistics for Windows [Computer software]. Version 18.0. Chicago: SPSS Inc, 2009.

Sravanthi TV, Manju SL. Indoles — A promising scaffold for drug development. Eur J Pharm Sci. 2016;91: 1-10. doi: 10.1016/j.ejps.2016.05.025.

Takuwa Y, Okamoto Y, Yoshioka K, Takuwa N. Sphingosine-1-phosphate signaling in physiology and diseases. Biofactors. 2012;38:329-37. doi: 10.1002/biof.1030.

Tu KN, Lie JD, Wan CKV, Cameron MA, Austel AG, Nguyen J, et al. Osteoporosis: a review of treatment options. P T. 2018;43(2):92-104.

United Nations Department of Economic Social Affairs. World population ageing 2020: Highlights. New York: United Nations, 2021.

Vemula D, Jayasurya P, Sushmitha V, Kumar YN, Bhandari V. CADD, AI and ML in drug discovery: A comprehensive review. Eur J Pharm Sci. 2023;181: 106324. doi: 10.1016/j.ejps.2022.106324.

Weske S, Vaidya M, Reese A, von Wnuck Lipinski K, Keul P, Bayer JK, et al. Targeting sphingosine-1-phosphate lyase as an anabolic therapy for bone loss. Nat Med. 2018;24:667-78. doi: 10.1038/s41591-018-0005-y.

Weske S, Vaidya M, von Wnuck Lipinski K, Keul P, Manthe K, Burkhart C, et al. Agonist-induced activation of the S1P receptor 2 constitutes a novel osteoanabolic therapy for the treatment of osteoporosis in mice. Bone. 2019;125:1-7. doi: 10.1016/j.bone.2019.04.015.

Worachartcheewan A, Prachayasittikul V, Prachayasittikul S, Tantivit V, Yeeyahya C, Prachayasittikul V. Rational design of novel coumarins: A potential trend for antioxidants in cosmetics. EXCLI J. 2020;19:209-26. doi: 10.17179/excli2019-1903.

Wotton CJ, Green J, Brown A, Armstrong MEG, Floud S, Beral V, et al. Use of oral bisphosphonates and risk of hospital admission with osteonecrosis of the jaw: Large prospective cohort study in UK women. Bone. 2019;124:69-74. doi: 10.1016/j.bone.2019.04.003.

Yazdi A, Ghasemi-Kasman M, Javan M. Possible regenerative effects of fingolimod (FTY720) in multiple sclerosis disease: An overview on remyelination process. J Neurosci Res. 2020;98:524-36. doi: 10.1002/jnr.24509.

Yu T-y, Pang W-j, Yang G-s. 3,3′-Diindolylmethane increases bone mass by suppressing osteoclastic bone resorption in mice. J Pharmacol Sci. 2015;127(1):75-82. doi: 10.1016/j.jphs.2014.11.006.

Zhou Y, Xue X, Guo Y, Liu H, Hou Z, Chen Z, et al. A quinoxaline-based compound ameliorates bone loss in ovariectomized mice. Exp Biol Med. 2021;246: 2502-10. doi: 10.1177/15353702211032133