SAR and QSAR in Environmental Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gsar20

A QSAR study to predict the survival motor neuron promoter activity of candidate diaminoquinazoline derivatives for the potential treatment of spinal muscular atrophy

G. Sabuncu Gürses, S.S. Erdem & M.T. Saçan

To cite this article: G. Sabuncu Gürses, S.S. Erdem & M.T. Saçan (2023) A QSAR study to predict the survival motor neuron promoter activity of candidate diaminoquinazoline derivatives for the potential treatment of spinal muscular atrophy, SAR and QSAR in Environmental Research, 34:3, 247-266, DOI: 10.1080/1062936X.2023.2200975

To link to this article: https://doi.org/10.1080/1062936X.2023.2200975

SAR AND QSAR IN ENVIRONMENTAL RESEARCH
2023, VOL. 34, NO. 3, 247–266
https://doi.org/10.1080/1062936X.2023.2200975

Taylor & Francis
Taylor & Francis Group

A QSAR study to predict the survival motor neuron promoter activity of candidate diaminoquinazoline derivatives for the potential treatment of spinal muscular atrophy

G. Sabuncu Gürses[a], S.S. Erdem[a] and M.T. Saçan[b]

[a]Chemistry Department, Faculty of Science, Marmara University, Istanbul, Turkey; [b]Institute of Environmental Sciences, Bogaziçi University, Istanbul, Turkey

ABSTRACT

Spinal Muscular Atrophy is a genetic neuromuscular disease that leads to muscle weakness and atrophy and it is characterized by the loss of α-motor neurons in the spinal cord's anterior horn cells. The disease appears due to low levels of the survival motor neuron protein. There are continuing clinical trials for the treatment of Spinal Muscular Atrophy. Quinazoline–based compounds are promising since they were tested on fibroblasts derived from the patients and found to increase the survival motor neuron protein levels. In this study, using multiple linear regression, we generated robust and valid quantitative structure- activity relationship models to predict the survival motor neuron–2 promoter activity of the new candidate compounds using the experimental survival motor neuron–2 promoter activity values of 2,4–diaminoquinazoline derivatives taken from the literature. The novel compounds designed by combining the pyrido[1,2–α]pyrimidin–4–one moeity of the known drug Risdiplam with that of 2,4 – diaminoquinazoline scaffold were predicted to exhibit strong promoter activities.

ARTICLE HISTORY
Received 5 January 2023
Accepted 4 April 2023

KEYWORDS
SMA; SMN protein; QSAR; diaminoquinazoline; Risdiplam; MLR

Introduction

Spinal Muscular Atrophy (SMA) is an autosomal recessive neuromuscular disease that leads to muscle weakness and atrophy [1]. Its specific characterization is the loss of α-motor neurons in the spinal cord's anterior horn cells. SMA disease was the most frequent monogenic cause of death in infancy before pharmacological treatment became available. The disease appears due to low survival motor neuron SMN protein levels. SMN protein is involved in numerous vital roles, ranging from RNA splicing, neuronal survival, neurite outgrowth and neuromuscular junction formation [2]. SMN is expressed ubiquitously and molecularly and has an important role in pre-mRNA processing [3]. In humans, there are two *SMN* genes in chromosome 5q13 as *SMN1* and *SMN2*. The two genes are in the same genomic region which is a duplicated inverted region [4]. *SMN1* is the primary gene that encodes SMN protein. A low level of SMN protein is caused by deletion and/or mutation of

CONTACT S.S. Erdem msacan@boun.edu.tr
Supplemental data for this article can be accessed at: https://doi.org/10.1080/1062936X.2023.2200975

the *SMN1* gene. In this situation, the *SMN2* gene is not able to compensate for this loss of *SMN1* [5].

The *SMN* gene was duplicated prior to human-chimpanzee split 5–7 million years ago. Although almost all vertebrate species have SMN protein, both *SMN1* and *SMN2* genes are only specific to humans. The two genes produce different amounts of alternatively spliced transcripts, the full-length transcript being the major product of *SMN1* and a transcript lacking exon 7 being the major product of *SMN2* [6].

Although the two genes are nearly identical, in a healthy human without SMA disease, the *SMN1* gene is responsible for the production of 85% to 90% of the full-length protein, whereas the *SMN2* gene is responsible for only 10% to 15%. Therefore, *SMN2* can only partially compensate for *SMN1* loss. The number of copies of the *SMN2* gene varies in humans; the more *SMN2* gene copies a person has, the more SMN protein is produced [7]. This is especially important in the total absence of *SMN1* or in case it is mutated.

SMA clinical phenotype is heterogeneous and ranges from a severe to a mild phenotype. It is generally divided into three main subtypes: type I, typically presents itself in the first two years of children and they are unable to sit and most often die from respiratory failure, type II patients have an onset after 6 months of age and are usually diagnosed by 18 months. They achieve the ability to sit but not to walk, show moderate respiratory dysfunction, and experience scoliosis. Type III is usually diagnosed between 2 and 17 years of age. These type III patients are able to sit and walk but often become wheelchair-bound as the course of the disease progress. The adult form of the SMA (type IV) is characterized by age of onset at >30 years and shows very mild signs of muscle weakness [2,5].

SMA drugs are at the top list of the most expensive therapeutics worldwide. Therefore, SMA treatment is very costly and unreachable for most patients. Available treatment for SMA includes *SMN2* splicing modifiers and gene replacement therapy or upregulating muscle growth [8]. There are only three approved drugs. Nusinersen (Spinraza®), an antisense oligonucleotide and intrathecal splicing modifier, was approved by the Food and Drug Administration (FDA) in 2016 and by the European Medicines Agency (EMA) in 2017 for all subtypes of SMA. The adeno-associated virus vector-based gene therapy onasemnogene abeparvovec xioi (Zolgensma®) was approved by the FDA in 2019 for SMA in children <2 years and by the EMA in May 2020 for SMA in patients with 2 or 3 *SMN2* copies. The oral splicing modifier evrysdi (Risdiplam®) was approved by the FDA in 2020 and serves to treat patients two months of age and older with SMA. Several attempts at SMN independent therapies are currently underway [8].

An increase in SMN protein level might be achieved by (i) enhancing *SMN2* gene transcription or (ii) modifying defective splicing of the *SMN*2 mRNA, thereby increasing the number of the full-length transcript, or alternatively a combination of steps i and ii. Small molecule activators of the *SMN2* gene promoter, which enhance SMN expression, thus represent a promising strategy for the treatment of SMA [9]. In addition, stabilizing the functional SMN proteins, resulting in increased protein levels may lead to novel treatments for SMA [10,11]. There are continuing clinical trials for the treatment of SMA, ranging from repurposed drugs, gene therapy, antisense oligonucleotides, and novel small molecules [12]. Repurposed drugs offer both savings in time and cost since some drug development stages such as preclinical testing and safety evaluations have been already accomplished.

The identification of several small molecules was performed by Jarecki et al. [13] by a cell-based assay for *SMN2* promoter activation and ultra-high-throughput-screening (uHTS) campaign. As a result of this campaign, the molecules that have the quinazoline scaffold in their structure were among the most potent *SMN2* promoters. These molecules were tested on fibroblasts derived from SMA patients and found that they increased the relative abundance of the full-length SMN transcript and SMN protein levels. The quinazoline scaffold is the most attractive, as it is the most potent compound class identified from the screen [9].

In 2013, van Meerbeke et al. investigated one of the quinazoline derivatives (RG3039) as the lead clinical candidate due to its reduced off-target effects compared with other quinazoline derivatives [14]. Good clinical results were obtained as a Phase I clinical trial is completed with this molecule on healthy volunteers. RG3039 robustly inhibits DcpS (the RNA decapping enzyme which is strongly correlated with *SMN2* promoter activation) in central nervous system tissues [15]. SMA mice were treated with RG3039 and SMN transcripts levels showed an increase of 30–40% in neural tissues [14].

In spite of extensive research efforts [16], the exact cure for SMA is still not achievable. Available treatments are very few, extremely expensive, and have side effects and drawbacks. Therefore, any attempt to design and develop new cost-effective therapeutics is crucial and urgently necessary. Quantitative Structure-Activity Relationships (QSARs) models in accordance with Organization for Economic Co-operation and Development (OECD) principles can offer a faster and cheaper screening of the compound collections to identify new potent molecules in the early stage of ligand-based drug design and development process. However, except for a few machine/deep learning and molecular dynamics applications on similar systems [16–22], to the best of our knowledge, so far, we have not encountered a QSAR model development study in the literature related to SMN promoter activity end-point. The fact that the SMA disease is a progressive and fatal disorder with very limited treatment options and the lack of QSAR studies motivated us to develop validated QSAR models that can be used to predict the SMN promoter activity of the novel small molecules.

In order to achieve this goal, the following steps were taken in this study: (i) The measured *SMN2* promoter activities of 99 quinazoline derivatives reported by Thurmond et al. [9] were selected as data set. (ii) Three QSAR models were developed and validated by using internal and external validation tools (iii) The developed QSAR models were used to predict the unknown *SMN2* promoter activity of the external set molecules.

Materials and methods

Dataset compilation

A dataset containing 109 compounds with *SMN2* promoter activity values was taken from Thurmond et al. [9] (Supplementary Material (SM) Table S1). These *SMN2* promoter activity values were reported as 50% median effective concentration (EC_{50}) and obtained from the β–lactamase activity assay which was previously described by Jarecki et al. [13]. In this assay, the NSC-34 cell line, a hybrid between mouse spinal cord cells and a mouse neuroblastoma, was utilized because it exhibits characteristics of motor neurons. The unit of experimental EC_{50} values was given as μM and converted to – log EC_{50}. During the

compilation of the dataset, 10 racemic compounds were omitted. Molecular structures of the remaining 99 compounds were drawn and their conformations were scanned using the SPARTAN 16 software [23]. All conformations were initially optimized with the semiempirical PM6 method [24]. Then, density functional theory (DFT) with the M06-2X/6-31 G(d) method [25] was employed for the geometry optimization of the most stable conformation based on the performance of this method in our previous studies [26–30].

Descriptor calculation

The optimized geometries of the molecules were used to calculate the descriptors. A pool of 5273 descriptors for each molecule was calculated using Dragon 7.0.2 software [31] and uploaded to the QSARINS 2.2.1 software [32–34]. After the pre-filtration of descriptors (excluding semi-constant descriptors (>80%) and highly intercorrelated descriptors (>95%), a total of 1051 descriptors were considered before importing the data for model generation.

Training/Test set division: In order to create and confirm the prediction power of the generated QSAR models, the dataset was divided into training and test sets. As a consequence, the dataset was split into training (80% of compounds) and test sets (20% of compounds). Different division approaches have been employed, all of which are implemented in the QSARINS program. These strategies include (i) ordered response, (ii) grouping the molecules based on their structure, and (iii) random divisions.

Model generation

QSAR models were generated using the ordinary least square (OLS) method. All subsets and the genetic algorithm (GA) tools of the QSARINS 2.2.1 software [32–34] were used to select descriptors for the training set. The coefficient of determination (r^2), the determination coefficient adjusted for the number of variables (r^2_{adj}), the leave-one-out cross-validation (Q^2_{LOO}), and leave-many-out cross-validation (Q^2_{LMO}) were used to assess the statistical quality, internal predictivity, and robustness of the QSAR models created in the current study. In order to summarize the overall errors of the QSAR models, root mean square errors (RMSE) were evaluated for the training and test sets, $RMSE_{TR}$ and $RMSE_{TEST}$, respectively. Inter-correlation among descriptors was tested via the QUIK rule (Q Under the Influence of K) [35]. The QUICK rule was set to 0.05 to minimize inter-correlation between descriptors. The reliability of the models was tested by response randomization (Y-scrambling, 2000 iterations to check the fitting of the randomly reordered Y-data) procedure. After scrambling the response, the new models' noticeably low coefficient of determination (r^2_{Yscr} and Q^2_{Yscr}) suggest that the proposed model was not developed by chance.

External validation

The following metrics were used to evaluate the external validation of each model: Q^2 metrics such as, Q^2_{F1} [36], Q^2_{F2} [37], Q^2_{F3} [38,39], Concordance Correlation Coefficient (CCC) [40,41], for training and test set (CCC_{TR} and CCC_{TEST}) [42,43], r^2m, and Δr^2m [44]. When the three external Q^2 metrics are larger than 0.70 [42], the CCC is greater than 0.85 [42], the r^2m is greater than 0.5, the r^2m is less than 0.2, and the Δr^2m is greater than 0.5,

the predictivity of a QSAR model is deemed satisfactory. The model's reliability was also checked for additional validation parameters such as Golbraikh and Tropsha's Criteria [45].

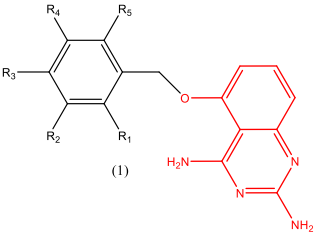
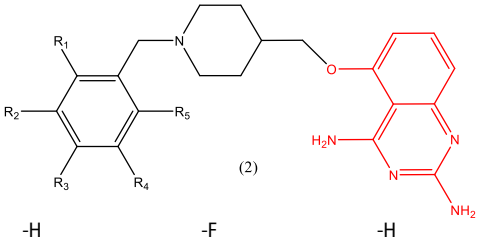
Applicability domain and model prediction reliability

As an important part of the OECD principles [46] for the validation and regulatory purpose of QSAR models, the applicability domain (AD) which is the structural and response boundary depending on the chemical nature of the data used for model development should be defined. The AD is an important tool for the reliable application of models. We can consider the compounds in the AD as reliable and outside the AD as less reliable. The leverage approach of AD was applied to all models in this study [47]. The leverage approach was used to visually assess the predictive performance of the model via a plot of standardized residuals vs. hat values (Williams graph). Additionally, we have assessed the prediction reliability of the created models using the prediction reliability indicator (PRI) tool, a free program available at http://dtclab.webs [48,49] in order to confirm their reliability. The PRI tool categorizes the quality of the test set or real external set predictions into three categories: good (with a composite score of 3), moderate (with a composite score of 2), and bad (with a composite score of 1). Furthermore, we used the external validation plus tool and provided the mean absolute error-based metrics (MAE_{TEST}) and mean absolute error for 95% of test set compounds data [49] in order to further assess the predictive power of the generated models. According to this approach, an error of 10% of the training set range should be acceptable while an error value of more than 20% of the training set range should be a very high error. Thus, the following criteria should be met for good predictions: MAE ≤ 0.1 x training set range and MAE $+3*\sigma \leq 0.2$ x training set range. Here, the σ value is the standard deviation of the absolute error values for the test set data. The predictions could be considered bad when: MAE > 0.15 x training set or MAE + 3 x σ > 0.25 x training set range. The predictions that do not comply with any of the two conditions may be considered moderate quality [49].

External set chemicals

The generated QSAR models were used to predict the *SMN2* promoter activity of 50 external chemicals. 31 of the 50 compounds were taken by Rietz et al. [50] (we labelled them with R where R refers to Rietz as shown in SM Table S2) having in vitro *SMN2* promoter activity measured by a different assay (luciferase reporter assay) and 19 of them were newly designed by us (labelled as D1-D19 where D refers to 'designed'). The structures of the designed compounds are given in Tables 1 and 2. The molecules labelled from D1 to D12 were constructed by attaching different substituents to the 5^{th} position of the 2,4–diaminoquinazoline scaffold which were not listed in the study of Thurmond et al. [9]. For D13-D19 labelled molecules, our design strategy was to combine the pharmacophore scaffolds pyrido[1,2-α]pyrimidin–4–one (blue) or 2,8–dimethylimidazo[1,2–β]pyridazine) (pink) of the known drug Risdiplam® (Figure 1) and of the quinazoline scaffold (red) from the study of Thurmond et al. [9]. The geometries of all the designed compounds and compounds used from the Rietz et al. [50] study were optimized with the same methodology as described for dataset compounds above.

Table 1. Chemical structures of external set compounds (designed D1-D12).

Compound No	*R_1*	*R_2*	*R_3*	*R_4*	*R_5*
					
	(1)				
D1	-F	-F	-H	-H	-H
D2	-H	-Cl	-H	-Cl	-H
D3	-H	-Cl	-Cl	-H	-H
D4	-H	-Cl	-H	-H	-Cl
D5	-Cl	-H	-Cl	-H	-H
D6	-Cl	-H	-H	-H	-Cl
D7	-Cl	-Cl	-H	-H	-H
D8	-H	$-OCH_3$	-H	$-OCH_3$	-H
D9	-H	$-OCH_3$	$-OCH_3$	-H	-H
D10	$-OCH_3$	-H	-H	$-OCH_3$	-H
					
	(2)				
D11	-H	-F	-H	-H	-H
D12	-H	-H	-F	-H	-H

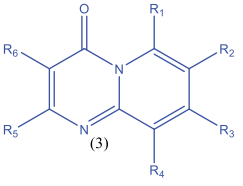
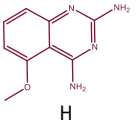
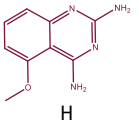
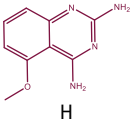
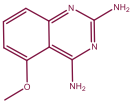
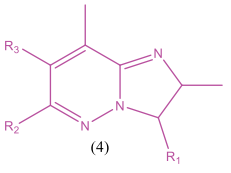
Results and discussion

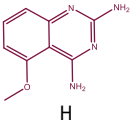
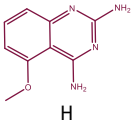
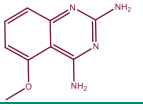
Supplementary Material Table S1 contains a complete list of chemicals in the dataset, together with *SMN2* promoter activity values [9] (pEC_{50} (μM)). Numerous attempts were made using alternative training/test set divisions with the random, response-based, and structure-based splitting setups in order to generate robust and validated QSAR models. Two 7-descriptor models and one 6-descriptor model were highlighted from random divisions.

Internal and external validation criteria of these models are given in Tables 3 and 4, respectively. In terms of fit, internal, and external validation parameters, all models meet the criteria. The r^2 (0.6161–0.6858) and Q^2_{LOO} (0.5341–0.6111) values reveal that all models are stable, while the higher r^2_{TEST} values (0.7564–0.7939) and low $RMSE_{TEST}$ (0.2335–0.2640) values show that these models have good predictive potential.

The structural coverage of all models for 50 external chemicals was found as 84%, 76% and 46%, for model 1, model 2 and model 3, respectively (Table 4). Model 1 (Tables 3 and 4) with Eq. (1) is the proposed 7-descriptor QSAR model for the prediction of pEC_{50} values and standard errors of coefficients, as it provides the highest structural coverage for external chemicals.

Table 2. Chemical structures of external set compounds (designed, D13-D19).

Compound No	*R_1*	*R_2*	*R_3*	*R_4*	*R_5*	*R_6*
	 (3)					
D13		H	H	H	H	H
D14	H		H	H	H	H
D18	H	H	H		H	H
D19	H	H	H	H		H
	 (4)					

Compound No	*R_1*	R_2	*R_3*
D15		H	H
D16	H		H
D17	H	H	

$$\begin{aligned} \mathrm{pEC}_{50} = {} & -2.9347\ (\pm 2.429) + 0.0448\ (\pm 0.0205)\ \mathrm{RDF020s}\ -0.2676\ (\pm 0.2027) \\ & \mathrm{Mor10e}\ -0.9965\ (\pm 0.5145)\ \mathrm{Mor27e}\ -0.0812\ (\pm 0.0596)\ \mathrm{CATS2D\_08\_AL} \\ & +0.3347\ (\pm 0.2334)\ \mathrm{B05[C-F]} - 0.2094\ (\pm 0.0982)\ \mathrm{F02[C-O]} \\ & +4.701\ (\pm 3.182)\ \mathrm{BIC2} \end{aligned} \quad (1)$$

The equations of models 2 and 3 are provided in SM as Eq. (S1) and Eq. (S2), respectively. The chemicals utilized in the training and test sets, as well as predicted pEC_{50} values, hat values, and the values of descriptors that appeared in Models 1, 2, and 3 are also given in

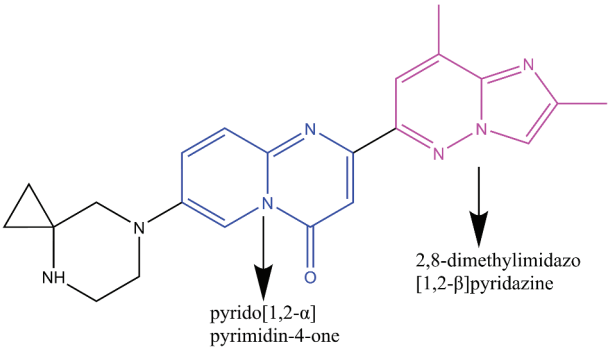


Figure 1. Risdiplam molecular structure and its pharmacophore groups used for the design of external set compounds D13-D19.

SM (through Tables S3-S5). Figure 2(a) shows a plot of predicted versus observed pEC_{50} values for model 1, whereas Figure 2(b) shows its Williams plot. The fit and Williams plots of Models 2 and 3 are shown in SM, Figure S1, and Figure S2, respectively. The predicted and observed *SMN2* promoter activity values are well distributed along the optimal line for all models and compatible with each other. Models 1 and 2 have no response and structural outliers (Figure 2(b) and Figure S2(a), respectively) whereas model 3 has only one structural outlier (Figure S2 (b). The hat value of the compound (5-(tetrahydropyran–2–ylmethoxy) quinazoline–2,4–diamine) is slightly greater than the critical hat value ($h^* = 0.284$), but its predicted value is very close to the experimental value, so it has been considered as "structural outlier" and a good leverage chemical (Figure S2(b)). The standardized residuals for all compounds in the training and test sets are less than three standard deviation units, indicating that the data set contains no response values greater than the response outlier limit (±3σ). These findings show that the QSAR models are able to successfully interpolate all of the pEC_{50} values of the dataset chemicals.

Xternal Validation Plus tool [48] (MAE-based criteria applied on 95% of data) was used to check the presence of systematic errors in models 1–3. The results of the Xternal Validation Plus tool for models 1, 2, and 3 are given in SM Table S6, S7, and S8, respectively. The prediction quality of each model was found to be 'Good'.

The prediction reliability index was also employed to ensure the prediction quality by using the Prediction Reliability Indicator tool by Roy et al. [51] and good prediction quality was observed for all test set compounds for all models (Table 4).

Model 1 has the highest structural coverage (84%) of the three models (Table 4), the Insubria graph for Model 1 is displayed in Figure 3 (for Models 2 and 3, please refer to SM Figure S3 (a and b, respectively)). Table 5 lists the definition of descriptors appearing in Model 1. The order of descriptors based on their importance for Model 1 is as follows:

$$\text{RDF020s} > \text{Mor27e} > \text{Mor10e} > \text{F02}[\text{C}-\text{O}] > \text{CATS2D\_08\_AL} > \text{BIC2} > \text{B05}[\text{C}-\text{F}]$$

The order of descriptors for Model 2 and Model 3 is given in SM. The descriptors appearing in Models 2 and 3 and their meanings are also given in SM (Table S9).

RDF020s (Radial Distribution Function-020 (weighted by I-state)) belongs to the RDF Descriptions class and has a positive impact on *SMN2* promoter activity. The BIC2 Bond

Table 3. Fit and internal validation parameters of developed models 1, 2 and 3.

Model No	n_{TR}/n_{TEST}	Number of variables	r^2	r^2adj	$RMSE_{tr}$	CCC_{TR}	F	CCC_{CV}	Q^2_{Loo}	$RMSE_{CV}$	Q^2_{LMO}	r^2Yscr	Q^2Yscr
1	83/14	7	0.6837	0.6542	0.3523	0.8122	23.1638	0.7697	0.6111	0.3907	0.5945	0.0858	–0.1226
2	75/22	7	0.6858	0.6530	0.3535	0.8136	20.8927	0.7682	0.6100	0.3938	0.5895	0.0952	–0.1371
3	74/23	6	0.6161	0.5818	0.3852	0.7625	17.9244	0.7139	0.5341	0.4244	0.5104	0.0809	–0.1260

Table 4. External validation parameters of developed Models 1,2 and 3.

Model No	r^2_{TEST}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	$r^2_m aver.$	Δr^2_m	$RMSE_{TEST}$	k'	k	PRI[a]	% Structural coverage	MAE_{TEST}
1	0.7633	0.7625	0.7376	0.8610	0.8719	0.6731	0.0426	0.2335	0.9735	0.9776	GOOD	84	0.4517
2	0.7564	0.7408	0.7403	0.8248	0.8697	0.6735	0.0020	0.2640	0.9576	0.9666	GOOD	76	0.5575
3	0.7939	0.8200	0.7745	0.8484	0.8903	0.7164	0.0590	0.2421	0.9709	0.9839	GOOD	46	0.5317

[a]PRI = Prediction reliability indicator.

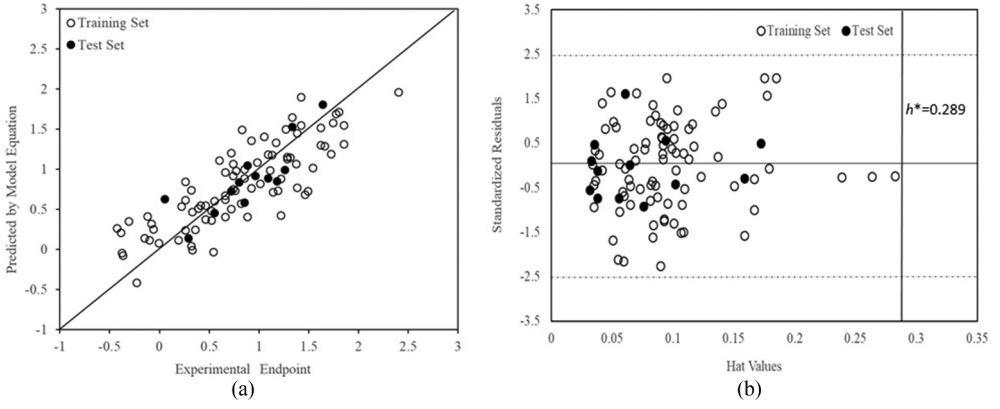


Figure 2. Scattered plot of observed versus predicted pEC_{50} values for Model 1 (a) and Williams plot of Model 1 (b).

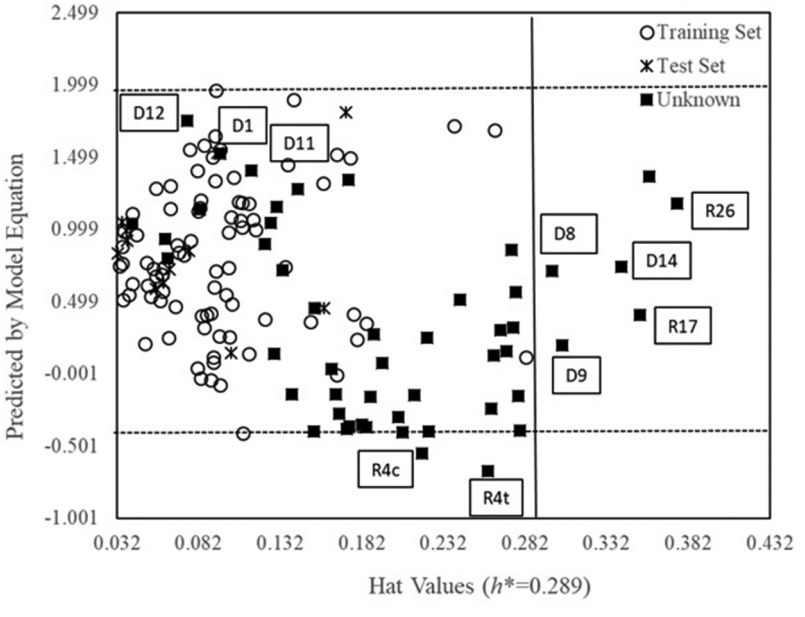


Figure 3. Insubria graph of Model 1.

Table 5. Molecular descriptors found in Model 1 with their contribution.

Symbol	*Contribution*	*Class*	*Definition*
BIC2	Positive	Information indices	Bond Information Content index (neighbourhood symmetry of 2-order)
CATS2D_08_AL	Negative	CATS2D	CATS2D Acceptor-Lipophilic at lag 08
5[C-F]	Positive	2D Atom Pairs	Presence/absence of C – F at topological distance 5
F02[C-O]	Negative	2D Atom Pairs	Frequency of C – O at topological distance 2
Mor10e	Negative	3D-MoRSE descriptors	Signal 10/weighted by Sanderson electronegativity
Mor27e	Negative	3D-MoRSE descriptors	Signal 27/weighted by Sanderson electronegativity
RDF020s	Positive	RDF Descriptions	Radial Distribution Function – 020/weighted by I-state

Information Content index (neighbourhood symmetry of 2-order) from the information indices group is another descriptor that appeared in this model. It has also a positive impact, although it contributes the least to the activity. B05[C-F] is a 2D Atom Pairs description that indicates "the presence or absence of C – F at topological distance 5" and is found in all three models, contributing positively to *SMN2* promoter activity. As the value of this descriptor increases, so does the predicted *SMN2* promoter activity value.

The descriptors from the 3D-MoRSE descriptors, Mor10e and Mor27e, which correspond to the mean "Signal 10 and Signal 27/weighted by Sanderson electronegativity", respectively, had a negative impact on the *SMN2* promoter activity. They reveal the molecules' three-dimensional arrangement of atoms. CATS2D_08_AL belongs to CATS2D descriptors, and it gives information about the topological distance between any pair of pharmacophore point types. These potential pharmacophore points are hydrogen-bond acceptor and lipophilic atoms reflected by CATS2D_08_AL. Likewise, the CATS2D descriptor was utilized in the MLR model of the antioxidant activity of coumarins [52]. F02[C-O] is from the 2D atom pairs descriptors indicating "Frequency of C – O at topological distance 2".

The predicted *SMN2* promoter activities of the external set compounds from Models 1, 2, and 3 are given together with the calculated descriptor values in Tables 6–8, respectively. Since there is no experimental activity value (*SMN2* promoter activity based on the β–lactamase activity assay) for these compounds, we considered all models in the assessment of their *SMN2* promoter activity. It should be noted that, although the *SMN2* promoter activities of compounds R4-R27 are reported by Rietz et al. [50], these measurements are based on a different end-point (Luciferase reporter assay standard system). This system can identify compounds based on the principle that three different mechanisms together increase SMN protein levels. These are respectively; the modulation of alternative splicing of *SMN2* exon 7; increasing transcription from the *SMN2* promoter; stabilizing the full-length SMN fusion protein or mRNA. In the dataset we used to develop the QSAR models, the measured EC_{50} (Thurmond et al. [9]) is based on a different principle and is defined as only *SMN2* promoter activity. Therefore, the predicted values for compounds R4-R27 are not directly relevant to the experimentally measured values by Rietz et al. [50].

Table 9 lists the pEC_{50} values of the external set compounds predicted from each model together with the average predicted pEC_{50} values. Note that only the predicted values within the AD are considered in calculating the average predicted pEC_{50}. As shown from the Insubria graph in Figure 3, 8 external set compounds are out of the structural space and response range of Model 1 (D8, D9, D14, R25, R26, R27 and R4c, R4t, respectively). Insubria graphs of Model 2 and Model 3 (Figures S3a, S3b) reveal that these two models have better coverage of designed compounds but weaker coverage of Rietz et al.'s compounds [50]. In Model 2, 12 compounds from Rietz et al. [50] and only D6 from the designed ones are out of AD. All designed compounds are covered by Model 3 but 27 compounds from Rietz et al. [50] are out of AD. The predicted pEC_{50} values of D12, R8c from Model 3, and R4f from Model 2 are slightly out of the response range, but they are within the range of experimental maximum and minimum values. Therefore, we included them in calculating the average of predicted pEC_{50} values in Table 9 while the compounds that are out of the structural and/or response ranges of the models were excluded (denoted as 'Not

Table 6. Predicted pEC_{50} *SMN2* promoter activity values[a] for external set compounds from Model 1 together with descriptor values.

Comp.	*RDF020s*	*Mor10e*	*Mor27e*	*CATS2D _08_AL*	*B05[C-F]*	*F02[C-O]*	*BIC2*	*Predicted pEC_{50} (μM)*
D1	17.783	−0.405	−0.385	2	1	3	0.771	**1.5221**
D2	17.72	−0.927	−0.087	0	0	3	0.76	**1.1382**
D3	17.686	−0.892	−0.111	2	0	3	0.771	**1.0404**
D4	17.667	−1.08	−0.114	4	0	3	0.771	0.9304
D5	17.725	−0.863	−0.074	6	0	3	0.781	0.7195
D6	17.657	−0.893	0.089	8	0	3	0.749	0.2492
D7	17.795	−1.048	0.017	4	0	3	0.771	0.797
D8	31.887	−1.053	−0.091	2	0	7	0.739	0.7116
D9	31.767	−0.196	−0.052	5	0	7	0.739	0.1943
D10	32.295	−0.689	0.062	4	0	7	0.736	0.3034
D11	25.746	−0.473	−0.269	3	0	3	0.78	**1.4077**
D12	25.824	−0.662	−0.258	3	1	3	0.773	**1.7526**
D13	21.474	−0.914	−0.367	4	0	4	0.823	**1.3436**
D14	21.907	−0.51	−0.031	7	0	4	0.837	0.7421
D15	21.492	−0.053	−0.081	4	0	3	0.793	0.8974
D16	18.252	−0.581	−0.314	5	0	3	0.793	**1.0445**
D17	21.061	−1.046	−0.235	5	0	3	0.806	**1.2771**
D18	22.313	−0.668	−0.007	6	0	4	0.837	0.8599
D19	19.156	−0.392	−0.261	2	0	4	0.823	**1.157**
R4	6.84	0.72	0.092	1	0	4	0.784	−0.1459
R4a	7.148	0.54	0.032	0	0	4	0.738	−0.1592
R4b	7.13	0.569	0.168	2	0	4	0.762	−0.3529
R4c	7.607	0.503	0.183	1	0	4	0.697	−0.5532
R4d	7.407	0.592	0.055	1	1	4	0.75	0.1254
R4e	7.147	0.591	0.07	1	1	4	0.762	0.1555
R4f	7.686	0.441	0.177	1	0	4	0.75	−0.2779
R4g	7.23	0.382	0.201	1	0	4	0.738	−0.3629
R4h	15.604	0.202	0.001	2	0	6	0.759	−0.1417
R4i	14.558	0.189	0.089	1	0	6	0.769	−0.1445
R4j	7.694	0.435	0.305	1	0	4	0.75	−0.4035
R4k	7.473	0.489	0.218	2	0	4	0.761	−0.3707
R4l	7.623	0.263	0.208	2	0	4	0.739	−0.3969
R4m	9.219	0.784	0.133	3	0	4	0.776	−0.2975
R4n	10.751	0.438	0.162	1	0	4	0.784	0.0349
R4p	6.412	0.11	0.271	2	0	4	0.807	−0.1533
R4r	12.186	0.64	0.097	6	0	4	0.762	−0.3998
R4s	10.93	−0.147	−0.076	8	0	4	0.761	−0.2402
R4t	7.131	0.494	0.149	6	0	4	0.754	−0.6765
R4v	5.741	1.298	0.011	1	0	4	0.758	−0.3913
R8c	6.819	0.522	−0.101	1	0	4	0.831	0.3194
R8d	7.507	0.56	0.003	1	0	4	0.796	0.0719
R8e	5.927	0.3	0.103	1	0	4	0.721	−0.3815
R8i	8.565	0.204	−0.373	3	0	4	0.734	0.1352
R10	7.893	0.284	−0.134	1	0	3	0.784	0.4525
R11	6.78	0.444	−0.015	0	0	3	0.773	0.2708
R16	6.85	0.1	−0.066	1	1	4	0.784	0.5126
R24	6.149	0.116	0.055	1	1	3	0.784	0.5657
R25	6.126	−0.14	−0.086	0	1	1	0.804	**1.3677**
R26	6.111	−0.012	−0.005	1	1	1	0.806	**1.1802**
R27	6.602	−0.061	0.046	3	1	4	0.813	0.4068

[a]High *SMN2* promoter activity predicted values $pEC_{50} > 1$ are highlighted in bold.

Table 7. Predicted pEC_{50} *SMN2* promoter activity values for external set compounds from Model 2 together with descriptor values.

Comp.	*O%*	*BIC3*	*Chi1_EA(dm)*	*C-026*	*CATS2D _08_AL*	*B05[C-F]*	*F06[C-N]*	*Predicted pEC_{50} (μM)*
D1	2.9	0.886	18.144	3	2	1	4	**1.3523**
D2	2.9	0.854	22.382	3	0	0	4	**1.0669**
D3	2.9	0.886	20.395	3	2	0	4	0.9651
D4	2.9	0.886	19.417	3	4	0	4	0.71
D5	2.9	0.897	19.087	3	6	0	4	0.5846
D6	2.9	0.843	17.389	3	8	0	4	−0.1169
D7	2.9	0.886	18.101	3	4	0	4	0.608
D8	7.1	0.812	31.598	3	2	0	4	0.7094
D9	7.1	0.829	27.765	3	5	0	4	0.2663
D10	7.1	0.826	27.607	3	4	0	4	0.322
D11	1.9	0.856	29.897	2	3	0	5	**1.6036**
D12	1.9	0.836	28.91	2	3	1	5	**1.944**
D13	5.1	0.898	26.643	1	4	0	8	0.8808
D14	5.1	0.898	29.394	1	7	0	6	**1.1302**
D15	2.4	0.867	20.214	1	4	0	11	0.0788
D16	2.4	0.867	22.929	1	5	0	9	0.5048
D17	2.4	0.867	20.214	1	5	0	8	0.4471
D18	5.1	0.898	29.196	1	6	0	8	0.8992
D19	5.1	0.898	26.429	1	2	0	8	**1.0435**
R4	6.5	0.833	15.82	3	1	0	4	−0.1875
R4a	6.5	0.784	14.295	2	0	0	4	−0.3489
R4b	6.5	0.833	15.54	2	2	0	4	−0.0774
R4c	6.5	0.768	13.633	1	1	0	4	−0.3841
R4d	6.5	0.821	11.51	2	1	1	4	0.1749
R4e	6.5	0.833	15.582	2	1	1	4	0.5772
R4f	6.5	0.821	11.464	2	1	0	4	−0.3903
R4g	6.5	0.784	14.274	2	1	0	4	−0.4402
R4h	8.6	0.82	15.697	2	2	0	4	−0.4535
R4i	8.6	0.831	19.989	2	1	0	4	0.0483
R4j	5.9	0.814	12.843	1	1	0	5	−0.1812
R4k	5.9	0.825	17.021	1	2	0	4	0.285
R4l	5.9	0.781	15.104	1	2	0	4	−0.1817
R4m	6.7	0.827	13.633	0	3	0	5	−0.0959
R4n	6.7	0.827	13.633	1	1	0	4	0.0146
R4p	7.4	0.807	13.633	0	2	0	5	−0.249
R4r	5.9	0.824	13.633	0	6	0	6	−0.4272
R4s	5.4	0.818	13.936	0	8	0	7	−0.709
R4t	6.1	0.818	13.633	0	6	0	5	−0.346
R4v	5.7	0.819	14.212	1	1	0	4	0.1417
R8c	7.1	0.887	14.35	3	1	0	3	0.1578
R8d	5.9	0.84	14.35	3	1	0	3	−0.014
R8e	5.9	0.764	18.762	3	1	0	5	−0.5272
R8i	5.7	0.86	23.173	3	3	0	5	0.3576
R10	3.2	0.833	15.593	3	1	0	4	0.2573
R11	3.4	0.826	24.546	3	0	0	1	**1.4199**
R16	6.5	0.833	16.472	2	1	1	2	0.9514
R24	6.5	0.833	16.472	2	1	1	2	0.9514
R25	3.2	0.833	16.472	2	0	1	2	**1.5034**
R26	3.1	0.866	16.472	2	1	1	3	**1.5138**
R27	6.1	0.871	23.266	2	3	1	3	**1.4767**

[a]High *SMN2* promoter activity predicted values $pEC_{50} > 1$ are highlighted in bold.

Table 8. Predicted pEC_{50} *SMN2* promoter activity values for external set compounds from Model 3 together with descriptor values.

Comp.	*SpMin5_ Bh(m)*	*Chi1_ EA(dm)*	*RDF020e*	*B05[C-F]*	*F02[C-O]*	*CATS3D _03_DL*	*Predicted pEC_{50} (μM)*
D1	1.276	18.144	8.118	1	3	0	**1.2624**
D2	1.275	22.382	8.289	0	3	0	**1.2193**
D3	1.274	20.395	8.143	0	3	0	**1.0581**
D4	1.277	19.417	8.052	0	3	0	0.9673
D5	1.275	19.087	8.109	0	3	0	0.9572
D6	1.277	17.389	8.009	0	3	1	0.6384
D7	1.276	18.101	8.087	0	3	0	0.8811
D8	1.53	31.598	13.274	0	7	1	0.6052
D9	1.531	27.765	13.196	0	7	1	0.3176
D10	1.53	27.607	13.595	0	7	1	0.3686
D11	1.526	29.897	12.901	0	3	1	**1.6547**
D12	1.525	28.91	12.985	1	3	1	**1.9729**
D13	1.45	26.643	9.889	0	4	0	**1.0306**
D14	1.382	29.394	9.95	0	4	1	**1.2247**
D15	1.503	20.214	9.764	0	3	1	0.5499
D16	1.524	22.929	8.535	0	3	0	0.6857
D17	1.502	20.214	9.586	0	3	0	0.7034
D18	1.441	29.196	10.376	0	4	0	**1.3078**
D19	1.295	26.429	8.943	0	4	1	**1.0751**
R4	1.196	15.82	3.364	0	4	2	−0.4536
R4a	1.198	14.295	3.521	0	4	3	−0.7215
R4b	1.227	15.54	3.481	0	4	2	−0.5319
R4c	1.231	13.633	3.711	0	4	3	−0.8211
R4d	1.229	11.51	3.585	1	4	3	−0.6129
R4e	1.227	15.582	3.528	1	4	2	−0.1483
R4f	1.228	11.464	3.637	0	4	3	−0.9796
R4g	1.191	14.274	3.539	0	4	3	−0.7032
R4h	1.417	15.697	6.551	0	6	3	−1.3118
R4i	1.417	19.989	6.147	0	6	2	−0.8884
R4j	1.417	12.843	3.796	0	4	3	−1.32
R4k	1.417	17.021	3.712	0	4	3	−1.0345
R4l	1.417	15.104	3.793	0	4	3	−1.1591
R4m	1.228	13.633	4.206	0	4	2	−0.5617
R4n	1.226	13.633	4.73	0	4	2	−0.4783
R4p	1.243	13.633	3.002	0	4	1	−0.6012
R4r	1.417	13.633	4.726	0	4	3	−1.1242
R4s	1.417	13.936	5.424	0	4	1	−0.6424
R4t	1.291	13.633	3.458	0	4	1	−0.6503
R4v	1.417	14.212	2.714	0	4	0	−0.8511
R8c	1.101	14.35	3.336	0	4	2	−0.3301
R8d	1.384	14.35	3.772	0	4	2	−0.9575
R8e	1.417	18.762	2.942	0	4	2	−0.8479
R8i	1.181	23.173	4.167	0	4	2	0.2281
R10	1.177	15.593	3.888	0	3	2	−0.0405
R11	1.137	24.546	3.486	0	3	1	0.8138
R16	1.203	16.472	3.546	1	4	1	0.1545
R24	1.208	16.472	3.285	1	3	1	0.4073
R25	1.14	16.472	3.614	1	1	2	**1.0538**
R26	1.095	16.472	3.353	1	1	1	**1.3026**
R27	1.094	23.266	3.845	1	4	1	0.9508

[a]High *SMN2* promoter activity predicted values $pEC_{50} > 1$ are highlighted in bold.

Table 9. Predicted pEC_{50} values (μM) from Models 1, 2, 3 and the mean predicted pEC_{50} values[a,b] with standard deviation (SD) of the external set compounds.

Comp.	*Predicted pEC_{50} Model 1*	*Predicted pEC_{50} Model 2*	*Predicted pEC_{50} Model 3*	*Mean pEC_{50} ± SD*	*Comp.*	*Predicted pEC_{50} Model 1*	*Predicted pEC_{50} Model 2*	*Predicted pEC_{50} Model 3*	*Mean pEC_{50} ± SD*
D1	1.5221	1.3523	1.2624	**1.3789 ± 0.1319**	R4f	−0.2779	−0.3903	NA	−0.3341 ± 0.0795
D2	1.1382	1.0669	1.2193	**1.1415 ± 0.0763**	R4g	−0.3629	NA	NA	c
D3	1.0404	0.9651	1.0581	**1.0212 ± 0.0494**	R4h	−0.1417	NA	NA	c
D4	0.9304	0.71	0.9673	0.8692 ± 1.1391	R4i	−0.1445	0.0483	NA	−0.0481 ± 0.1363
D5	0.7195	0.5846	0.9572	0.7538 ± 0.1886	R4j	−0.4035	−0.1812	NA	−0.29235 ± 0.1572
D6	0.2492	NA	0.6384	0.4438 ± 0.2752	R4k	−0.3707	0.285	NA	−0.04285 ± 0.4636
D7	0.797	0.608	0.8811	0.762 ± 0.1399	R4l	−0.3969	−0.1817	NA	−0.2893 ± 0.1522
D8	NA	0.7094	0.6052	0.6573 ± 0.0737	R4m	−0.2975	NA	NA	c
D9	NA	0.2663	0.3176	0.29195 ± 0.0363	R4n	0.0349	0.0146	NA	0.02475 ± 0.0144
D10	0.3034	0.322	0.3686	0.3313 ± 0.0336	R4p	−0.1533	NA	NA	c
D11	1.4077	1.6036	1.6547	**1.5553 ± 0.1304**	R4r	−0.3998	NA	NA	c
D12	1.7526	1.944	1.9729	**1.8898 ± 0.1197**	R4s	−0.2402	NA	NA	c
D13	1.3436	0.8808	1.0306	**1.085 ± 0.2361**	R4t	NA	NA	NA	NA
D14	NA	1.1302	1.2247	**1.1775 ± 0.0668**	R4v	−0.3913	0.1417	NA	−0.1248 ± 0.3769
D15	0.8974	0.0788	0.5499	0.5087 ± 0.4109	R8c	0.3194	0.1578	NA	0.2386 ± 0.1143
D16	1.0445	0.5048	0.6857	0.745 ± 0.2747	R8d	0.0719	−0.014	NA	0.02895 ± 0.0607
D17	1.2771	0.4471	0.7034	0.8092 ± 0.4250	R8e	−0.3815	NA	−0.8479	−0.6147 ± 0.3298
D18	0.8599	0.8992	1.3078	**1.0223 ± 0.2480**	R8i	0.1352	0.3576	NA	0.2464 ± 0.1573
D19	1.157	1.0435	1.0751	**1.0919 ± 0.0586**	R10	0.4525	0.2573	−0.0405	0.2231 ± 0.2483
R4	−0.1459	−0.1875	NA	−0.1667 ± 0.0294	R11	0.2708	NA	NA	c
R4a	−0.1592	−0.3489	NA	−0.25405 ± 0.1341	R16	0.5126	NA	NA	c
R4b	−0.3529	−0.0774	NA	−0.21515 ± 0.1948	R24	0.5657	0.9514	0.4073	0.6415 ± 0.2799
R4c	NA	NA	NA	NA	R25	NA	**1.5034**	NA	d
R4d	0.1254	NA	NA	c	R26	NA	**1.5138**	NA	d
R4e	0.1555	NA	NA	c	R27	NA	**1.4767**	NA	d

[a]The predicted values for the compounds that are out of the AD of the models are denoted as not applicable (NA). Only the predicted values within the AD are considered in calculating the average predicted pEC_{50}.
[b]High *SMN2* promoter activities mean $pEC_{50} > 1$ are highlighted in bold.
[c]Predicted pEC_{50} is only available from Model 1.
[d] Predicted pEC_{50} is only available from Model 2.

Applicable' in Table 9) to avoid ambiguity in their reliability. Therefore, the discussion below about predicted pEC_{50} values of external set compounds is based on the averages of two or three models.

All rationally designed compounds (D1-D19) exhibit predicted $pEC_{50} > 0$ (or $EC_{50} < 1$ μM), revealing their high potential as *SMN2* promoters (Table 9). Among them, compounds D1-D3, D11-D14, D18, D19 with $pEC_{50} > 1$ (or $EC_{50} < 0.1$ μM) are predicted as remarkably potent *SMN2* promoters (highlighted as bold in Table 9) whereas the activities of compounds D4, D5, D7, D8, D15-D17 are predicted to be slightly less ($0.5 < pEC_{50} < 1$). Compounds D13-D19 have special importance as they are designed as new substances inspired by the pharmacophore groups of the Risdiplam molecule. Larger pEC_{50} values predicted for D13, D14, D18, D19 than for D15-D17 reveals that pyrido[1,2-α] pyrimidin-4-one moiety increases the *SMN2* promoter activity more than the 2,8-dimethylimidazo[1,2 -β]pyridazine) moiety. On the other hand, almost half of the compounds from Rietz et al. [50] display $pEC_{50} < 0$ (or $EC_{50} > 1$ μM), indicating that, in general, these compounds with isoxazole scaffold have much weaker *SMN2* promoter activities based on the β–lactamase activity assay used as end-point in the development of the models in this study. Nevertheless, three of them (R25, R26, R27) consisting of 1,3-thiazole heterocyclic moiety (without isoxazole scaffold) are predicted to have very strong *SMN2* promoter activities with $pEC_{50} > 1$.

As a result, we propose D1-D3, D11-D14, D18, D19, and R25-R27 as compounds with strong promoter activity potential which should be studied further to measure their in vitro and in vivo activity. Among them, D11, D12, and R25-R27 were predicted to exhibit the highest *SMN2* promoter activities.

Conclusion

This study aimed to develop robust and predictive QSAR models using a dataset composed of 97 2,4–diaminoquinazoline derivatives and to predict the *SMN2* promoter activities of the compounds with no experimental values. Three validated MLR models with seven and six descriptors were generated. Their prediction quality was confirmed by PRI and Xternal Plus tool. Thus, these QSAR models were used to predict the *SMN2* promoter activities of the external set compounds, 19 of which were designed as new 2,4–diaminoquinazoline derivatives. Compounds D13, D14, D18, D19 are novel compounds designed by combining the pyrido[1,2-α]pyrimidin–4–one moiety of Risdiplam with that of 2,4–diaminoquinazoline scaffold. They were predicted to have notably high *SMN2* promoter activity. Hence, this study adds new candidate compounds to the literature that are predicted to have high *SMN2* promoter activity.

In conclusion, based on the average predicted pEC_{50} values, we propose D1-D3, D11-D14, D18, D19, and R25-R27 as promising lead compounds for SMA treatment, which need to be further investigated for their pharmacodynamic and pharmacokinetic properties.

Acknowledgements

We thank Prof. Paola Gramatica (University of Insubria, Italy) for providing the QSARINS software. SSE would like to acknowledge Marmara University Scientific Research Project Commission (BAPKO) for supporting the Computational Chemistry Laboratory.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

M.T. Saçan http://orcid.org/0000-0003-2902-4965

References

[1] G.E. Opera, S. Kröber, M.L. McWhorter, W. Rossoll, S. Müller, M. Krawczak, G.J. Bassell, C.E. Beattie, and B. Wirth, *Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy*, Science 320 (2008), pp. 524–527. doi:10.1126/science.1155085.

[2] R.S. Anderton and F.L. Mastaglia, *Advances and challenges in developing a therapy for spinal muscular atrophy*, Expert Rev. Neurother. 15 (2015), pp. 895–908. doi:10.1586/14737175.2015.1059757.

[3] C. Madocsai, S.R. Lim, T. Geib, B.J. Lam, and K.J. Hertel, *Correction of SMN2 pre-mRNA splicing by antisense U7 small nuclear RNAs*, Mol. Ther. 12 (2005), pp. 1013–1022. doi:10.1016/j.ymthe.2005.08.022.

[4] S. Lefebvre, L. Bürglen, S. Reboullet, O. Clermont, P. Burlet, L. Viollet, B. Benichou, C. Cruaud, P. Millasseau, M. Zeviani, D.L. Paslier, J. Frezal, D. Cohen, J. Weissenbach, A. Munnich, and J. Melki, *Identification and characterization of a spinal muscular atrophy-determining gene*, Cell 80 (1995), pp. 155–165. doi:10.1016/0092-8674(95)90460-3.

[5] M.D. Howell, N.N. Singh, and R.N. Singh, *Advances in therapeutic development for spinal muscular atrophy*, Future Med. Chem. 6 (2014), pp. 1081–1099. doi:10.4155/fmc.14.63.

[6] C.F. Rochette, N. Gilbert, and L.R. Simard, *SMN gene duplication and the emergence of the SMN2 gene occurred in distinct hominids: SMN2 is unique to* Homo sapiens, Hum. Genet. 108 (2001), pp. 255–266. doi:10.1007/s004390100473.

[7] M. Calucho, S. Bernal, L. Alías, F. March, A. Venceslá, F.J. Rodríguez-Álvarez, E. Aller, R.M. Fernández, S. Borrego, J.M. Millán, C. Hernández-Chico, I. Cuscó, P. Fuentes-Prior, and E.F. Tizzano, *Correlation between SMA type and SMN2 copy number revisited: An analysis of 625 unrelated Spanish patients and a compilation of 2,834 reported cases*, Neuromuscular Disord. 28 (2018), pp. 208–215. doi:10.1016/j.nmd.2018.01.003.

[8] D.C. Schorling, A. Pechmann, and J. Kirschner, *Advances in treatment of spinal muscular atrophy*, Neuromuscular Disord. 7 (2020), pp. 1–13. doi:10.3233/JND-190424.

[9] J. Thurmond, M.E.R. Butchbach, M. Palomo, B. Pease, M. Rao, L. Bedell, M. Pai, G. Keyvan, R. Mishra, M. Haraldsson, T. Andresson, G. Bragason, M. Thosteinsdottir, J.M. Bjornsson, D.D. Coovert, A.H.M. Burghes, M.E. Gurney, and J. Singh, *Synthesis and biological evaluation of novel 2,4-diaminoquinazoline derivatives as SMN2 promoter activators for the potential treatment of SMA*, Med. Chem. 51 (2008), pp. 449–469. doi:10.1021/jm061475p.

[10] F. Farooq, S. Balabanian, X. Liu, M. Holcik, and A. MacKenzie, *p38 mitogen-activated protein kinase stabilizes SMN mRNA through RNA binding protein HuR*, Hum. Mol. Genet. 18 (2009), pp. 4035–4045. doi:10.1093/hmg/ddp352.

[11] M.A. Waldrop and S.J. Kolb, *Current treatment options in neurology-SMA therapeutics*, Curr. Treat. Options Neurol. 21 (2019), pp. 1–11. doi:10.1007/s11940-019-0568-z.

[12] A.N. Calder, E.J. Androphy, and K.J. Hodgetts, *Small molecules in development for the treatment of spinal muscular atrophy*, Med. Chem. 59 (2016), pp. 10067–10083. doi:10.1021/acs.jmedchem.6b00670.

[13] J. Jarecki, X. Chen, A. Bernardino, D.D. Coovert, M. Whitney, A. Burghes, J. Stack, and B.A. Pollok, *Diverse small-molecule modulators of SMN expression found by high-throughput compound screening: Early leads towards a therapeutic for spinal muscular atrophy*, Hum. Mol. Genet. 14 (2005), pp. 2003–2018. doi:10.1093/hmg/ddi205.

[14] J.P. Van Meerbeke, R.M. Gibbs, H.L. Plasterer, W. Miao, Z. Feng, M.Y. Lin, and C.J. Sumner, *The DcpS inhibitor RG3039 improves motor function in SMA mice*, Hum. Mol. Genet. 22 (2013), pp. 4074–4083. doi:10.1093/hmg/ddt257.

[15] J. Singh, M. Salcius, S.W. Liu, B.L. Staker, R. Mishra, J. Thurmond, G. Michaud, D.R. Mattoon, J. Printen, and J. Christensen, *DcpS as a therapeutic target for spinal muscular atrophy*, ACS Chem. Biol. 3 (2008), pp. 711–722. doi:10.1021/cb800120t.

[16] L.C. Chong, G. Gandhi, J.M. Lee, W.W.Y. Yeo, and S.B. Choi, *Drug discovery of spinal muscular atrophy (SMA) from the computational perspective: A comprehensive review*, Int. J. Mol. Sci. 22 (2021), pp. 8962. doi:10.3390/ijms22168962.

[17] L. Zhang, J. Tan, D. Han, and H. Zhu, *From machine learning to deep learning: Progress in machine intelligence for rational drug discovery*, Drug Discov. Today 22 (2017), pp. 1680–1685. doi:10.1016/j.drudis.2017.08.010.

[18] W. Li, *How do SMA-linked mutations of SMN1 lead to structural/functional deficiency of the SMA protein?*, PLoS ONE 12 (2017), pp. e0178519. doi:10.1371/journal.pone.0178519.

[19] J.D. Zhang, L. Sach-Peltason, C. Kramer, K. Wang, and M. Ebeling, *Multiscale modelling of drug mechanism and safety*, Drug Discov. Today 25 (2020), pp. 519–534. doi:10.1016/j.drudis.2019.12.009.

[20] R. Duranay, R. Bashirov, and A. Şeytanoğlu, *Simulation-based identification of optimal combination of drug candidates for spinal muscular atrophy*, Procedia Comput. Sci. 120 (2017), pp. 253–259. doi:10.1016/j.procs.2017.11.236.

[21] A. Garcia-Lopez, F. Tessaro, H.R.A. Jonker, A. Wacker, C. Richter, A. Comte, N. Berntenis, R. Schmucki, K. Hatje, O. Petermann, G. Chiriano, R. Perozzo, D. Sciarra, P. Konieczny, I. Faustino, G. Fournet, M. Orozco, R. Artero, F. Metzger, M. Ebeling, P. Goekjian, B. Joseph, H. Schwalbe, and L. Scapozza, *Targeting RNA structure in SMN2 reverses spinal muscular atrophy molecular phenotypes*, Nat. Commun. 9 (2018), pp. 1–12. doi:10.1038/s41467-018-04110-1.

[22] J. Axford, M.J. Sung, J. Manchester, D. Chin, M. Jain, Y. Shin, I. Dix, L.G. Hamann, A.K. Cheung, R. Sivasankaran, K. Briner, N.A. Dales, and B. Hurley, *Use of intramolecular 1,5-sulfur–oxygen and 1,5-sulfur–halogen interactions in the design of n-methyl-5-aryl-n-(2,2 6,6-tetramethylpiperidin-4-yl)-1,3,4-thiadiazol-2-amine SMN2 splicing modulators*, Med. Chem. 64 (2021), pp. 4744–4761. doi:10.1021/acs.jmedchem.0c02173.

[23] SPARTAN 16, Wavefunction, Inc, Irvine, USA, 2016; software available at http://wavefun.com .

[24] J.J.P. Stewart, *Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements*, Mol. Model. 13 (2007), pp. 1173–1213. doi:10.1007/s00894-007-0233-4.

[25] Y. Zhao and D.G. Truhlar, *The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals*, Theor. Chem. Acc. 120 (2008), pp. 215–241.

[26] A.A. Kaya, E. Salamci, A. Menzek, S.S. Erdem, E. Şahin, and K. Ecer, *Reaction of 9-oxabicyclo [4.2.1]non-7-ene-1-ol with tetrazine: An unusually facile intramolecular rearrangement*, Tetrahedron 73 (2017), pp. 5381–5388. doi:10.1016/j.tet.2017.07.040.

[27] C. Dokuzparmak, F.O. Tuncay, S.B. Ozdemir, B. Kurnaz, I. Demir, A. Colak, S.S. Erdem, and N. Yildirim, *Newly synthesized piperazine derivatives as tyrosinase inhibitors: In vitro and in silico studies*, Iran. Chem. Soc. 19 (2022), pp. 2739–2748. doi:10.1007/s13738-021-02487-3.

[28] U. Cakmak, F.O. Tuncay, S.B. Ozdemir, E.A. Demir, İ. Demir, A. Colak, S.C. Uzuner, S.S. Erdem, and N. Yildirim, *Synthesis of hydrazine containing piperazine or benzimidazole derivatives and their potential as α-amylase inhibitors by molecular docking, inhibition kinetics and in vitro cytotoxicity activity studies*, Med. Chem. Res. 30 (2021), pp. 1886–1904. doi:10.1007/s00044-021-02785-8.

[29] H. Dulger, O. Sari, N. Demirel, and S.S. Erdem, *Computational insight into the enantioselectivity of homoboroproline catalyzed asymmetric aldol reaction*, ChemistrySelect 4 (2019), pp. 7959–7967. doi:10.1002/slct.201901737.

[30] Y. Du, O. Sari, S.S. Erdem, and A. Whiting, *A bifunctional B,N-Based asymmetric catalytic nitrostyrene-Michael addition acting through a 10-membered ring cyclic transition state*, Helv. Chim. Acta 104 (2021), pp. e202100199. doi:10.1002/hlca.202100199.

[31] Kode srl, Dragon (software for molecular descriptor calculation) version 7.0.2, 2016; software available at https://chm.kode-solutions.net.

[32] P. Gramatica, N. Chirico, E. Papa, S. Cassani, and S. Kovarich, *QSARINS: A new software for the development, analysis, and validation of QSAR MLR models*, Comput. Chem. Software news and updates 34 (2013), pp. 2121–2132. doi:10.1002/jcc.23361.

[33] P. Gramatica, S. Cassani, and N. Chirico, *QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS*, Comput. Chem. Software news and updates 35 (2014), pp. 1036–1044. doi:10.1002/jcc.23576.

[34] QSARINS 2.2.1., 2015; software available at www.qsar.it.

[35] R. Todeschini, V. Consonni, and A. Maiocchi, *The K correlation index: Theory development and its application in chemometrics*, Chemom. Int. Lab. Syst. 46 (1999), pp. 13–29. doi:10.1016/S0169-7439(98)00124-5.

[36] L.M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R.M. Blair, W.S. Branham, S.L. Dial, C.L. Moland, and D.M. Sheehan, *QSAR models using a large diverse set of estrogens*, Chem. Inf. Comput. Sci. 41 (2001), pp. 186–195. doi:10.1021/ci000066d.

[37] G. Schüürmann, R. Ebert, J. Chen, B. Wang, and R. Kühne, *External validation and prediction employing the predictive squared correlation coefficients test set activity mean vs training set activity mean*, Chem. Inf. Model. 48 (2008), pp. 2140–2145. doi:10.1021/ci800253u.

[38] V. Consonni, D. Ballabio, and R. Todeschini, *Comments on definition of Q^2 parameter for QSAR validation*, Chem. Inf. Model. 49 (2009), pp. 1669–1678. doi:10.1021/ci900115y.

[39] V. Consonni, D. Ballabio, and R. Todeschini, *Evaluation of model predictive ability by external validation techniques*, Chemom. 24 (2010), pp. 194–201. doi:10.1002/cem.1290.

[40] L.I. Lin, *A concordance correlation coefficient to evaluate reproducibility*, Biometrics 31 (1989), pp. 255–268. doi:10.2307/2532051.

[41] L.I. Lin, *Assay validation using the concordance correlation coefficient*, Biometrics 48 (1992), pp. 599–604. doi:10.2307/2532314.

[42] N. Chirico and P. Gramatica, *Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient*, Chem. Inf. Model. 51 (2011), pp. 2320–2335. doi:10.1021/ci200211n.

[43] N. Chirico and P. Gramatica, *Real external predictivity of QSAR models. Part 2. New inter-comparable thresholds for different validation criteria and the need for scatter plot inspection*, Chem. Inf. Model. 52 (2012), pp. 2044–2058. doi:10.1021/ci300084j.

[44] P.K. Ojha, I. Mitra, R.N. Das, and K. Roy, *Further exploring r^2m metrics for validation of QSPR models*, Chemom. Intell. Lab. Syst. 107 (2011), pp. 194–205. doi:10.1016/j.chemolab.2011.03.011.

[45] A. Golbraikh and A. Tropsha, *Beware of q^2!*, Mol. Graph. Model. 20 (2002), pp. 269–276. doi:10.1016/S1093-3263(01)00123-1.

[46] OECD, Guidance document on the validation of (quantitative) structure–activity relationship [(Q)SAR] models, ENV/JM/MONO (2007)2, OECD Environment Health and Safety Publications Series of Testing and Assessment No. 69, Organisation for Economic Co-operation and Development, Paris, 2007.

[47] P. Gramatica, S. Cassani, P.P. Roy, S. Kovarich, C.W. Yap, and E. Papa, *QSAR modeling is not "Push a button and find a correlation": A case study of toxicity of(Benzo-)triazoles on algae*, Mol. Inform. 31 (2012), pp. 817–835. doi:10.1002/minf.201200075.

[48] K. Roy, P. Ambure, and S. Kar, *On a simple approach for determining applicability domain of QSAR models*, Chemom. Int. Lab. Syst. 145 (2015), pp. 22–29. doi:10.1016/j.chemolab.2015.04.013.

[49] K. Roy, R.N. Das, P. Ambure, and R.B. Aher, *Be aware of error measures, further studies on validation of predictive QSAR models*, Chemom. Int. Lab. Syst. 152 (2016), pp. 18–33. doi:10.1016/j.chemolab.2016.01.008.

[50] A. Rietz, H. Li, K.M. Quist, J.J. Cherry, C.L. Lorson, B.G. Burnett, and K.J. Hodgetts, *Discovery of a small molecule probe that post-translationally stabilizes the survival motor neuron protein for the treatment of spinal muscular atrophy*, Med. Chem. 60 (2017), pp. 4594–4610. doi:10.1021/acs.jmedchem.6b01885.

[51] K. Roy, P. Ambure, and S. Kar, *How precise are our quantitative structure activity relationship derived predictions for new query chemicals?*, ACS Omega 3 (2018), pp. 11392–11406. doi:10.1021/acsomega.8b01647.

[52] P. Erzincan, M.T. Saçan, B.Y. Dursun, Ö. Danış, S. Demir, S.S. Erdem, and A. Ogan, *QSAR models for antioxidant activity of new coumarin derivatives*, SAR QSAR Environ. Res. 26 (2015), pp. 721–737. doi:10.1080/1062936X.2015.1088571.