Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. *90* (1) 39–52 (2025)
JSCS–5818
Original scientific paper
Published 1 February, 2025

Monte Carlo optimization-based QSAR modelling of *Staphylococcus aureus* inhibitory activity of coumarin-1,2,3-triazole hybrids

KRISHNA N. MISHRA, HARISH C. UPADHYAY* and POONAM VERMA

Laboratory of Chemistry, Department of Applied Sciences, Rajkiya Engineering College (Affiliated with Dr. A.P.J. Abdul Kalam Technical University, Lucknow), Churk, Sonbhadra-231206, India

(Received 30 March, revised 22 September, accepted 12 November 2024)

Abstract: In this study, 51 coumarin-1,2,3-triazole hybrids with known minimum inhibitory concentration (*MIC*) values against *Staphylococcus aureus* were used for the generation of a Monte Carlo based optimized QSAR model on correlations and logic (CORAL) software. The entire dataset was divided into four different sets, namely the training set (Tr), the invisible training set (iTr), the calibration set (C) and the validation set (V) of three random splits. For each split, five models were generated using various combinations of SMILES, graphs and hybrid optimal descriptors with various connectivity indices. Finally, fifteen models were obtained from three random, non-identical splits. For the best model from each split, the correlation coefficient (r^2) ranged from 0.9672 to 0.8693, while the cross-validated correlation coefficient (Q^2) ranged from 0.9478 to 0.8250. The mean absolute error (*MAE*) for the best models was less than 0.065. Additionally, favourable values of the index of ideality of correlation (*IIC*) and correlation intensity index (*CII*) were reported, justifying the robustness, reliability and predictive potential of the developed models. Further, good and bad fingerprints were estimated based on correlation weights for structural attributes.

Keywords: CORAL software; antibacterial; structural attributes; index of ideality of correlation; correlation intensity index.

INTRODUCTION

Natural products have played a crucial role in the discovery of drugs for almost all types of diseases.[1] The natural products, as such and their semisynthetic analogs, contribute to more than half of the approved drugs of today.[2,3] Because of safety and compatibility, natural products are and will be at the center

* Corresponding author. E-mail: harishcu@recsonbhadra.ac.in
https://doi.org/10.2298/JSC240330094M

of the drug discovery process.[4] Traditional drug discovery methods often rely on trial-and-error approaches, which are time-consuming and resource-intensive.[5] However, due to dependence on several factors, it is not possible to estimate exactly, but on average, the development of a single drug from a new chemical entity may cost up to $1 billion over twelve to fifteen years.[5,6] While natural products continue to inspire drug discovery efforts, there is a growing recognition of the need for innovative approaches that integrate advances in computational biology, synthetic chemistry and biotechnology to overcome the limitations of traditional drug discovery.[7,8] By leveraging interdisciplinary approaches and cutting-edge technologies, researchers can harness the potential of natural resources more effectively and develop novel therapeutics with improved efficacy, safety and sustainability.[9] Computer-aided drug design (CADD) has emerged as a powerful tool in the field of pharmaceutical research that utilizes computational methods to facilitate the discovery and development of new drugs.[10] The CADD approach integrates principles from various disciplines, such as chemistry, biology and computer science, to expedite the drug discovery process.[11,12] CADD-based rational drug design approaches, *viz*., virtual screening by docking simulations, quantitative structure-activity relationship (QSAR) studies for lead optimization and prediction of bioactivities and *in silico* prediction of physiochemical characteristics of molecules in terms of absorption, distribution, metabolism and toxicity, help the researchers prioritize the most promising candidates for further experimental validation, significantly reducing the time and cost associated with traditional screening methods.[10]

QSAR studies play a pivotal role in drug discovery and development by enabling the rational design, optimization, and selection of potential drug candidates. QSAR models provide valuable insights into the relationship between the chemical structure of compounds and their biological activities.[13] QSAR modelling utilizes computational techniques to analyse the structural features of molecules and predict their pharmacological properties, thus facilitating the identification of potential drug candidates.[10] The application of QSAR studies is poised to revolutionize the drug discovery process, driving innovation and accelerating the development of new therapeutic agents to address unmet medical needs.[10] The most popular classic approach to QSAR modeling utilizes descriptors of molecular structure by using a simplified molecular input line entry system (SMILES) to correlate with the biological activity reported in wet lab experiments.[14] These models are formulated as a training set (for defining the model) and a test set or validation set (for checking the model with compounds of an external validation) and are often associated with one or more drawbacks.[14] A developed model must be robust enough to have high predictive power.[14] To validate a QSAR as scientifically legitimate and so enable its regulatory approval, the Organization for Economic Cooperation and Development (OECD) has

proposed five guidelines.[15] For any QSAR model, "appropriate measures of goodness-of-fit, robustness, and predictivity" must be established, according to Principle 4 of OECD.[15] It highlights the necessity for both the QSAR model's external validation (predictivity) and internal validation (as demonstrated by robustness and goodness-of-fit). The key to determining the trustworthiness of predictions is to determine whether the built models can be used with any confidence on new sets of data by using validation procedures.[13,15] The correlations and logic (CORAL), a freeware software for QSAR modelling uses the standardized SMILES-based optimal descriptors.[16] CORAL generates random models based on the Monte Carlo approach, and from a probabilistic perspective, the algorithms employed in CORAL can provide some answers to the limitations of classic QSAR modeling issues.[16] If the statistical quality of the model can be replicated in a series of attempts to create the model for both the training and validation sets, then a random model may be a reasonable predictor for an endpoint.[16,17] Following this logic, CORAL allows random splits to build a robust QSAR model. One can utilize the balance of correlations that is provided by the CORAL in addition to the conventional scheme that was previously dealt with.[16,17] The division of the training set into a calibration set and a sub-training/invisible training set is the fundamental concept of the balancing of correlations.[16] The calibration set serves as the basis for the initial validation of the model and the preliminary check assists in preventing overtraining. An additional measure to enhance predictability has involved analysing the balance of correlations with optimal slopes.[16] If the values of the cluster's slopes on the internal training and calibration sets are as close to each other as possible, the plot of the experimental versus calculated endpoint values will be perfect.[16,18]

Coumarins (2*H*-1-benzopyran-2-one) are a varied class of naturally occurring pharmacophores with a wide range of bioactivities, *viz*., anti-inflammatory, antioxidant, antinociceptive, hepatoprotective, antithrombotic, antiviral, antimicrobial, antituberculosis, anticancer, antidepressant, antihyperlipidemic, anti-Alzheimer, anticholinesterase and antiviral activities.[19,20] On the other hand, triazoles are five-membered heterocyclic compounds containing three nitrogen atoms that exist in two isomeric forms: 1,2,3-triazoles and 1,2,4-triazoles.[21,22] Commonly found in both natural and synthetic leads, 1,2,3-triazoles have a wide range of biological actions, including antitubercular, anticancer, antibacterial and antifungal properties.[21,22] Being enriched with electronegative nitrogen atoms that can interact in a variety of ways with biological targets, the 1,2,3-triazole ring system has emerged as a versatile and intriguing scaffold in medicinal chemistry, offering a myriad of pharmacological effects that contribute to its significance in drug discovery.[21,22] In recent attempts to synthesize multitargeting "hybrid molecules" there are enough reports on different coumarin-1,2,3-triazole--based molecular hybrids synthesized by coupling the coumarin to 1,2,3-triazole

as pharmacologically active segments, either in the presence or absence of tethering agents (spacer/linker).[21,22] Additionally, enough SAR has been identified in numerous series of potentially active coumarin hybrids with antibacterial, anticancer, antitubercular, antifungal, anti-Alzheimer, antidiabetic, antidepressant and antithrombotic properties that have been synthesized.[21,22]

In the given frame of context, from our recently published review encompassing the antibacterial coumarin-1,2,3-triazole hybrids, we collected data on coumarin-1,2,3-triazole hybrids with known activity against *Staphylococcus aureus* in terms of *MIC* values.[21] The collected data were used for the development of Monte Carlo optimization-based predictive QSAR modeling of *S. aureus* inhibitory activity of coumarin-1,2,3-triazole hybrids.

EXPERIMENTAL

Data collection

A total of 51 coumarin-1,2,3-triazole derivatives showing minimum inhibitory concentration (*MIC*) values in the range of 0.4 to 75 μg mL^{-1} against *S. aureus* were retrieved from the earlier published reports.[21,23-30] The molecular structures were drawn using ChemDraw v22, saved in ".cdxml" format, and then converted to simplified molecular-input line-entry system (SMILES) by using optical structure recognition (OSRA) online tool (https://cactus.nci.nih.gov/cgi-bin/osra/index.cgi). The input files for building the QSAR models corresponding to SMILES of the derivatives along with their *MIC* values were saved as text (.txt, also see Table S-I of the Supplementary material to this paper).

Descriptor calculation

Three types of descriptors: *i*) SMILES-based, *ii*) graph-based and *iii*) hybrid descriptors, which are combinations of SMILES and graphs, were calculated using CORAL software (http://www.insilico.eu/coral).[17]

SMILES-based descriptors. The optimal SMILES based descriptor were calculated as follows:

$$\begin{aligned}\mathrm{SMILES}_{DCW(T,N_{\mathrm{epoch}})} = \alpha\sum CW(S_{\mathrm{k}}) + \beta\sum CW(SS_{\mathrm{k}}) + \gamma\sum CW(SSS_{\mathrm{k}}) + \\ + \delta CW(\mathrm{PAIR}) + xCW(\mathrm{BOND}) + yCW(\mathrm{NOSP}) + zCW(\mathrm{HALO}) + \\ + mCW(\mathrm{HARD})\end{aligned} \quad (1)$$

where T is the threshold used to classify molecular features into rare (noise) and active based on their frequency, and N_{epoch} is the number of iterations of the optimization process. *CW* represents the correlation weight for the above-mentioned eight types of structural attributes. S_{k}, SS_{k} and SSS_{k} are local SMILES attributes as representations of molecular fragments. PAIR, BOND, NOSP and HALO are global SMILES attributes, while the HARD index represents the association of BOND, NOSP and HALO attributes. The coefficients α, β, γ, δ, x, y, z and m are assigned binary values only, *i.e.*, 0 if the attribute is not used in building the model or 1 if the attribute is used in building the model. S_{k} represents the sole SMILES element *(e.g.*, Cl..., C..., N...), SS_{k} denotes the combination of two SMILES elements connected with each other, and they can be represented as C1... (..., =... O..., *etc*). SSS_{k} stands for the combination of the three SMILES elements, represented as C... C... (..., C... #... N..., *etc*). PAIR denotes the simultaneous presence of two SMILES atom compounds during descriptor calculation. The index BOND is related to the presence of three categories of chemical bonds

i.e., "=" double bond, "#" triple bond, and "@@" stereospecific chemical bond. NOSP indicates the presence or absence of nitrogen, oxygen, sulphur and phosphorous atoms. The index HALO is related to the presence or absence of only three halogen atoms, *i.e.*, fluorine, chlorine and bromine.

Graph-based descriptors. In CORAL, three types of graph-based descriptors, namely HSG (hydrogen-suppressed graph), HFG (hydrogen-filled graph), and GAO (graph of atomic orbitals), with many invariants, are available. The optimal graph-based descriptors are calculated as follows:

$$\mathrm{Graph}DCW(T, N_{\mathrm{epoch}}) = \sum CW(A_{\mathrm{k}}) + \alpha\sum CW({}^{0}EC_{\mathrm{k}}) + \beta\sum CW({}^{1}EC_{\mathrm{k}}) + \gamma\sum CW({}^{2}EC_{\mathrm{k}}) + \delta\sum CW({}^{3}EC_{\mathrm{k}}) \quad (2)$$

where CW is the correlation weight of the invariants of the molecular graph. The indexes A_{k} is the correlation weight of atomic elements, and ${}^{0}EC_{\mathrm{k}}$, ${}^{1}EC_{\mathrm{k}}$, ${}^{2}EC_{\mathrm{k}}$ and ${}^{3}EC_{\mathrm{k}}$ are the Morgan extended connectivity for each vertex in the molecular graph of the zero, first, second and third order, respectively. The coefficients α, β, γ and δ belong to invariants of molecular graphs, which can take up only binary values, *i.e.*, 0 if not used or 1 if used in model building.

Hybrid descriptors. Hybrid descriptors are the combination of SMILES-based and graph--based descriptors and are calculated as follows:

$$\mathrm{Hybrid}DCW(T, N_{\mathrm{epoch}}) = \mathrm{Graph}DCW(T, N_{\mathrm{epoch}}) + \mathrm{SMILES}DCW(T, N_{\mathrm{epoch}}) \quad (3)$$

Monte Carlo optimization

A three-step process was chosen for the generation of a Monte Carlo-based optimized model. As required in the input format of CORAL software, the very first step entails preparing SMILES attributes from the molecular structures in various splits. The optimization of T and N_{epoch} values for each model independently is the second step. Here, N_{epoch} is the number of iterations in the Monte Carlo optimization, and T is the number of thresholds for classifying the molecular characteristics taken from the SMILES. Herein, values 1–4 of T and 15–60 of N_{epoch} were considered the most preferable combination.[31] The third and final step belongs to the calculation of descriptors (SMILES, graph and hybrid) and optimization of correlation weights (CW) using the balance correlation scheme of the Monte Carlo algorithm.[32] The entire dataset was divided into four different sets, namely the training set (Tr), the invisible training set (iTr), the calibration set (C) and the validation set (V) of three random splits. The distribution of compounds in training and various sets of three random splits have been provided in Table S-II of the Supplementary material. For each split, five models were generated using various combinations of SMILES, graphs and hybrid optimal descriptors with various connectivity indices. Finally, fifteen models were obtained from three random non-identical splits.

Mechanistic interpretation

Correlation weights (CW) for structural attributes (SAKs), *i.e.*, descriptors for SMILES, hybrids and fragments of local symmetry, were obtained in several probes of the Monte Carlo optimization for the best models (M4, M10 and M11). Structural attributes were classified into four types: promoters of increase, promoters of decrease, undefined and blocked. Those SAKs whose CW are positive indicate that these attributes of the derivatives of coumarin are promoters of an increase in MIC. SAKs with negative CW are responsible for decreasing the MIC of the derivatives. SAKs having both positive and negative correlation weights fall in the

undefined category, and all the attributes, where *CW* shows null values, fall under the blocked category.

RESULTS AND DISCUSSION

Several earlier published reports on the antibacterial activity of coumarin--1,2,3-triazole hybrids, have been compiled in the review by our group.[21] These reports comprise a preliminary screening of the antibacterial activity of the synthesized hybrid molecules against various Gram-positive and Gram-negative bacterial strains. The results of most of the preliminary screening have been published in terms of zone of inhibition or in *MIC* values. We collected the coumarin-1,2,3-triazole hybrids with antibacterial activities tested against *S. aureus* presented in *MIC* for the current study.[21] The descriptors of 51 molecules with known *MIC*s were calculated using CORAL software, which calculates SMILES--based, graph-based and hybrid descriptors, which are combinations of SMILES and graph-based descriptors. The literature reports highlight the outstanding forecasting capability of QSAR models on SMILES-based optimal descriptors as calculated by CORAL.[16,33] Using the Monte Carlo optimization method, the CORAL software allows the user to build QSAR models as a mathematical function of descriptors (also known as correlation weights of fragments of quasi--SMILES) with biological activity.[16] The entire dataset was divided into four different sets, namely the training set (Tr), the invisible training set (iTr), the calibration set (C) and the validation set (V) of three random non-identical splits: Split 1 (Tr = 16 compounds, iTr = 14 compounds, C = 11 compounds and V = 10 compounds); Split 2 (Tr = 15 compounds, iTr = 14 compounds, C = 11 compounds and V = 11 compounds); Split 3 (Tr = 11 compounds, iTr = 15 compounds, C = 14 compounds and V = 11 compounds; also see Table S-II). Each set aimed to solve an individual task. The training set aims to find the maximum correlation coefficient between endpoint and descriptor for compounds present. The invisible training set aimed at finding out whether the correlation is satisfactory for similar substances that are not involved in the training set. The calibration set identifies the starting point of overtraining. For each split, five models were generated using various combinations of SMILES, graph and hybrid optimal descriptors with various connectivity indices. Finally, fifteen models were generated after the search for the best preferable threshold value and N_{epoch} value for Monte Carlo optimization. The statistical parameters, correlation coefficient (r^2) cross-validated correlation coefficient (Q^2), standard error of estimation (s), Fischer ratio (F), mean absolute error (MAE), index of ideality of correlation (IIC) and correlation intensity index (CII), were calculated to evaluate the predictive potential of each model (Table I).

A developed QSAR model must, as always, be sufficiently robust to be able to forecast new molecular characteristics in an impartial, reliable and precise way.[9,14] The so-called five OECD principles, which are internationally defined,

are: *i*) a defined endpoint; *ii*) an unambiguous algorithm; *iii*) a defined domain of applicability; *iv*) suitable measures of robustness, predictivity and goodness-of--fit; *v*) a mechanistic interpretation, if feasible.[15,34] Considering OECD principles, the best models with higher predictive potential for each split are presented in Fig. 1, and their equations are depicted below:

Split 1.

$$pMIC = -4.1880389(\pm 0.1069537) + 0.1795520(\pm 0.0063400) \times DCW(1,40) \quad \text{(M4)}$$

Split 2.

$$pMIC = -4.4055993(\pm 0.0701103) + 0.1119546(\pm 0.0022489) \times DCW(1,40) \quad \text{(M10)}$$

Split 3.

$$pMIC = -5.1755729(\pm 0.1209286) + 0.1698881(\pm 0.0045733) \times DCW(1,70) \quad \text{(M11)}$$

TABLE I. Various models for *S. aureus* inhibitory activity using Monte Carlo optimization; *T*: threshold, *N*: N_{epoch}/Number of iterations, *n*: number of compounds in set, r^2: correlation coefficient, *s*: standard error of estimation, *MAE*: mean absolute error, *F*: Fischer ration, Tr: training set, iTr: invisible training set, C: calibration set, V: validation set

No.	Split	Parameter	*T*, *N*	Set	*n*	r^2	Q^2	*IIC*	*CII*	*s*	*MAE*	*F*
M1	1	SMILES	1, 40	Tr	16	0.4147	0.1668	0.5008	0.7502	0.455	0.347	10
				iTr	14	0.9052	0.8791	0.4884	0.9264	0.510	0.343	115
				C	11	0.7681	0.6661	0.8762	0.8670	0.362	0.284	30
				V	10	0.8036	0.7118	0.1263	0.8281	0.535	0.355	33
M2	1	SMILES	1, 25	Tr	16	0.9062	0.8655	0.9519	0.9312	0.182	0.137	165
		and HSG		iTr	14	0.9009	0.8741	0.7171	0.9272	0.479	0.389	109
		with 0		C	11	0.7638	0.6257	0.8736	0.8526	0.309	0.246	29
		ECk		V	10	0.2453	-0.5510	0.1673	0.3575	0.897	0.493	3
M3	1	SMILES	1, 50	Tr	16	0.8467	0.7969	0.7157	0.8887	0.233	0.188	77
		and HSG		iTr	14	0.8470	0.8167	0.8623	0.8694	0.362	0.304	66
		with 1		C	11	0.8959	0.8539	0.9465	0.9194	0.231	0.178	77
		ECk		V	10	0.2455	-0.2148	0.1270	0.5582	0.969	0.610	3
M4	1	SMILES	1, 40	Tr	16	0.8738	0.8250	0.7270	0.9184	0.211	0.161	97
		and HFG		iTr	14	0.8819	0.8493	0.9371	0.9177	0.473	0.353	90
		with 0		C	11	0.8732	0.8114	0.9343	0.9066	0.221	0.188	62
		ECk		V	10	0.3808	0.0753	0.3603	0.5537	0.766	0.482	5
M5	1	SMILES	1, 45	Tr	16	0.8871	0.8462	0.7325	0.9097	0.200	0.155	110
		and HFG		iTr	14	0.8793	0.8515	0.6968	0.8947	0.414	0.354	87
		with 1		C	11	0.6791	0.4856	0.8239	0.8121	0.340	0.272	19
		ECk		V	10	0.4852	0.2614	0.1450	0.5698	0.580	0.360	8
M6	2		1, 30	Tr	15	0.8593	0.8258	0.6180	0.8861	0.288	0.226	79
		SMILES		iTr	14	0.8711	0.8339	0.3429	0.9105	0.286	0.223	81
				C	11	0.6806	0.4677	0.8248	0.7768	0.434	0.301	19
				V	11	0.6524	0.2616	0.4050	0.7163	0.419	0.249	17
M7	2	SMILES	1, 70	Tr	15	0.8447	0.8099	0.6127	0.8804	0.303	0.230	71
		and HSG		iTr	14	0.8773	0.8443	0.2638	0.9136	0.250	0.181	86
		with 0		C	11	0.7032	0.5015	0.8384	0.8000	0.426	0.279	21
		ECk		V	11	0.6121	0.4713	0.5489	0.7854	0.415	0.299	14

TABLE I. Continued

No.	Split	Parameter	*T*, *N*	Set	*n*	r^2	Q^2	*IIC*	*CII*	*s*	*MAE*	*F*
M8	2	SMILES	1, 40	Tr	15	0.8519	0.8210	0.8076	0.8772	0.296	0.236	75
		and HSG		iTr	14	0.8973	0.8661	0.5188	0.9334	0.305	0.254	105
		with 1		C	11	0.7634	0.6097	0.8735	0.8201	0.284	0.207	29
		ECk		V	11	0.4633	0.1582	0.0625	0.6474	0.771	0.534	8
M9	2	SMILES	2, 25	Tr	15	0.8416	0.8015	0.6116	0.8806	0.306	0.246	69
		and HFG		iTr	14	0.8852	0.8598	0.4471	0.9171	0.302	0.237	93
		with 0		C	11	0.6644	0.4191	0.8149	0.8271	0.438	0.334	18
		ECk		V	11	0.6085	0.4633	0.2208	0.7015	0.801	0.572	14
M10	2	SMILES	1, 40	Tr	15	0.8693	0.8399	0.6216	0.9011	0.278	0.222	86
		and HFG		iTr	14	0.8923	0.8602	0.0793	0.9248	0.294	0.219	99
		with 1		C	11	0.7627	0.5970	0.8733	0.8364	0.330	0.240	29
		ECk		V	11	0.6788	0.3623	0.5865	0.7518	0.511	0.381	19
M11	3		1, 70	Tr	11	0.9672	0.9478	0.8195	0.9687	0.087	0.065	265
		SMILES		iTr	15	0.8475	0.8045	0.5932	0.8851	0.624	0.489	72
				C	14	0.8889	0.8500	0.9428	0.9397	0.355	0.275	96
				V	11	0.2995	-0.0298	0.5029	0.6025	0.737	0.563	4
M12	3	SMILES	1, 33	Tr	11	0.9579	0.9282	0.8156	0.9632	0.099	0.079	205
		and HSG		iTr	15	0.8301	0.7703	0.5334	0.8923	0.576	0.456	64
		with 0		C	14	0.8099	0.7314	0.8998	0.9394	0.420	0.326	51
		ECk		V	11	0.3844	0.0716	0.5218	0.6268	0.891	0.750	6
M13	3	SMILES	3, 15	Tr	11	0.8739	0.7521	0.5341	0.8810	0.171	0.128	62
		and HSG		iTr	15	0.7665	0.6942	0.5871	0.8499	0.543	0.418	43
		with 1		C	14	0.7612	0.6833	0.8725	0.8876	0.500	0.415	38
		ECk		V	11	0.3552	-0.2766	0.3893	0.5673	0.688	0.474	5
M14	3	SMILES	1, 23	Tr	11	0.9302	0.8892	0.8037	0.9414	0.127	0.093	120
		and HFG		iTr	15	0.8668	0.8322	0.5049	0.8942	0.761	0.535	85
		with 0		C	14	0.8469	0.7822	0.9195	0.9543	0.440	0.347	66
		ECk		V	11	0.3105	-0.0192	0.2676	0.6277	0.706	0.545	4
M15	3	SMILES	3, 75	Tr	11	0.8928	0.7935	0.7874	0.9286	0.158	0.130	75
		and HFG		iTr	15	0.7843	0.7114	0.4777	0.8490	0.436	0.317	47
		with 1		C	14	0.6740	0.5633	0.8205	0.8872	0.567	0.483	25
		ECk		V	11	0.3723	0.0122	0.2434	0.5971	0.649	0.462	5

The problem of overtraining in prediction models can be calculated by using the newly introduced parameters, *IIC* and *CII*, which increase the predictive potential of the model.[18,35] The *IIC* is based on the correlation coefficient and mean absolute error, while the *CII* indicates the statistical quality of linear regression models with a unique ability since it is a measure that is sensitive both to the value of the correlation coefficient and to the value of the mean absolute error (*MAE*).[35] Both *IIC* and *CII* can expose the drawbacks of the predictive model usually seen with others. Favorable statistical parameters are obtained for all the models from each split.

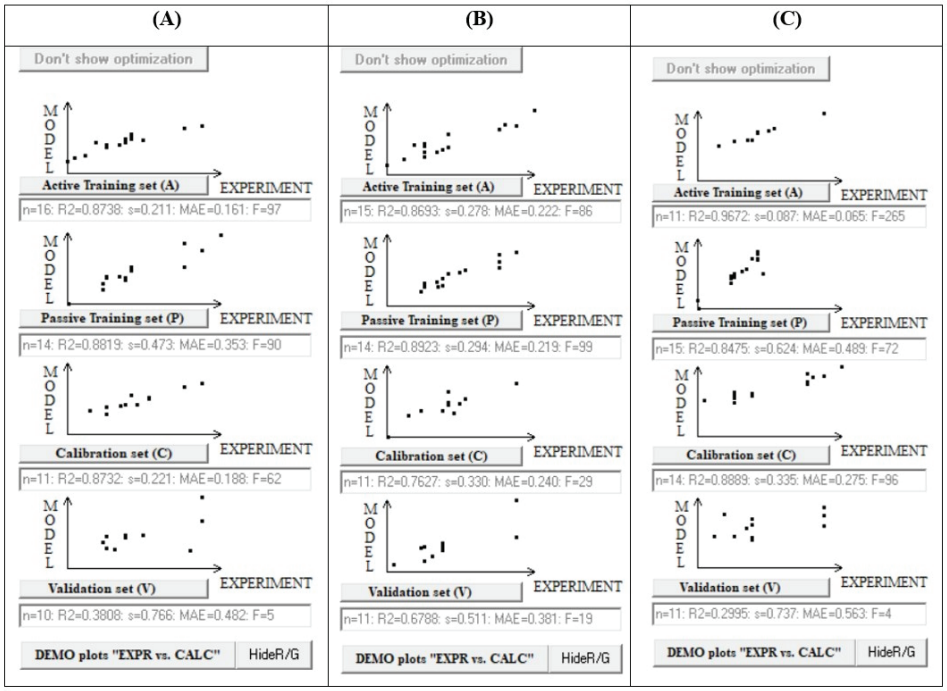

Fig. 1. Monte Carlo optimization based on best QSAR models from each split: M4 from Split 1 (A), M10 from Split 2 (B) and M11 from Split 3 (C).

Mechanistic interpretation

Correlation weights (*CW*) for structural attributes (SAKs) were estimated for both SMILES and hybrid descriptors using CORAL software. Structural attributes were classified into four types: promoters of increased activity, *i.e.*, good fingerprints (stable positive values), promoters of decreased activity, *i.e.*, bad fingerprints (stable negative values), undefined and blocked. SAKs having both positive and negative correlation weights fall in the undefined category, while all the attributes, where *CW* shows null values, fall under the blocked category. The data on *CW* for each structural attribute, which were obtained in several probes of the Monte Carlo optimization for the best models (M4, M10 and M11), are shown in supplemental Table S-III of the Supplementary material. The structural attributes of coumarin-1,2,3-triazole hybrids responsible for the increase in *MIC* value (Fig. 2) were c...3...c... (*i.e.*, presence of two aromatic cycles at three bond distances as can be seen in compound 42), c...2... (*i.e.*, presence of aromatic cycle followed by two bonds as can be seen in compounds 21 and 22), c...c...c... (*i.e.*, presence of three fused aromatic cycles as can be seen in compound 30), c...c...2... (*i.e.*, presence of branching on adjacent aromatic cycle as seen in compound 46), n...n... (*i.e.*, presence of adjacent nitrogen atoms as in the compound

4), O...=...(... (*i.e.*, presence of oxygen with double bonds as seen in the compound 7), c...(...Cl… (*i.e.*, presence of aromatic cycle and chlorine atom with branching as seen in compound 24) and c...(...N… (*i.e.*, presence of aromatic cycle and nitrogen atom with branching in between as can be seen in compound 31). On the other hand, among the hybrid descriptor correlation weights of Morgan's extended connectivity, the promoters responsible for the crucial increase in the inhibitory activity of the derivatives were found to be "EC0-O...1...", "EC1--C...6...", "EC1-C...9....", "EC1-N...5…", " EC1-Cl..6…", "EC1-O...4…", "EC1--N...2..." and "EC1-N...8..." of the model. "BOND10000000" constitutes the presence of a double bond is very important for compound's activity.

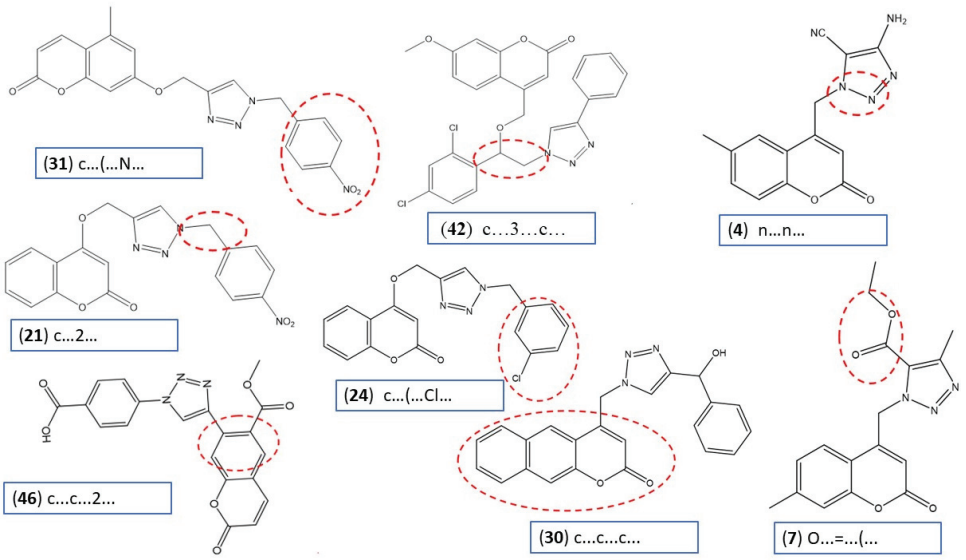

Fig. 2. Good fingerprints causing the increase in activity against *S. aureus*.

The structural attributes of coumarin-1,2,3-triazole hybrids responsible for the decrease in *MIC* value (bad fingerprints, Fig. 3) obtained from M4, M10, and M11 among SMILES descriptors were found to be c…(…=… (*i.e.*, presence of branching between aromatic carbon and double bonding as in compound 37), c...4...c... (*i.e.*, the presence of two aromatic cycles separated by four bonds as shown in compound 44) and c...(...C... (*i.e.*, the presence of branching between aromatic ring and aliphatic carbon as in compound 37). On the other hand, among hybrid descriptors correlation weights of Morgan's extended connectivity, the promoters responsible for the decrease of activity of the derivatives were found to be EC0-C...2..., EC1-C...10…

It can be seen that promoters of the decrease effect are not very much affected by the Morgan's extended connectivity in comparison to the promoters

of the increase. BOND11000000 (presence of a triple bond in the compound) and HALO00000000 (absence of a halogen atom) are detrimental to the inhibitory activity of bacteria.

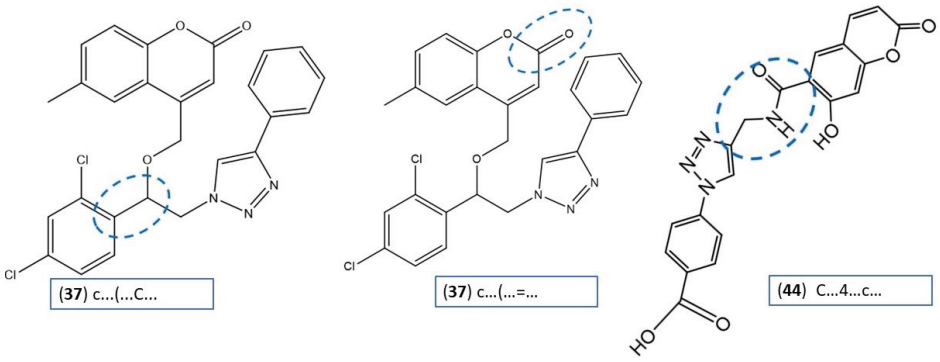

Fig. 3. Bad fingerprints causing the decrease in activity against *S. aureus*.

CONCLUSION

New 2D-QSAR models with good prediction ability were developed and validated for the prediction of the antibacterial activity of coumarin-1,2,3-triazole hybrids against *S. aureus* using the Monte Carlo optimization method. The consideration of additional statistical parameters such as the index of ideality of correlation and correlation intensity index gave clear insight into the improved predictability in all the QSAR models. Overall, among all the developed models, M11 may be considered best based on statistical parameters. The study provides a significant analysis of the important structural fingerprints required to be a potentially active substrate against *S. aureus*. Overall, the models will help design novel chemical entities with the coumarin-1,2,3-triazole moiety for the development of potent antibacterial drugs against *S. aureus*.

SUPPLEMENTARY MATERIAL

Additional data and information are available electronically at the pages of journal website: https://www.shd-pub.org.rs/index.php/JSCS/article/view/12871, or from the corresponding author on request.

Acknowledgements. The authors thank the Director of Rajkiya Engineering College, Sonbhadra, for his constant encouragement and for providing all types of facilities for this study. The financial support by the Council of Science and Technology Uttar Pradesh (Project No. 3027) is duly acknowledged.

И З В О Д

QSAR МОДЕЛОВАЊЕ ИНХИБИТОРНЕ АКТИВНОСТИ ХИБРИДА КУМАРИН-1,2,3--ТРИАЗОЛА НА *Staphylococcus aureus*, ЗАСНОВАНО НА МОНТЕ КАРЛО ОПТИМИЗАЦИЈИ

KRISHNA N. MISHRA, HARISH C. UPADHYAY и POONAM VERMA

Laboratory of Chemistry, Department of Applied Sciences, Rajkiya Engineering College (Affiliated with Dr. A.P.J. Abdul Kalam Technical University, Lucknow), Churk, Sonbhadra-231206, India

У овој студији су 51 кумаринski-1,2,3-триазолски хибриди са познатим минималним вредностима инхибирајућих концентрација (*MIC*) према *Staphylococcus aureus* искоришћени за генерисање, на Монте Карло заснованом, оптимизованог QSAR модела на *correlations and logic* (CORAL) софтверу. Целокупан скуп података је подељен на четири различита скупа података, наиме скуп за увежбавање (Tr), невидљиви скуп за увежбавање (iTr), скуп за калибрисање (C) и скуп за валидацију (V) из три случајне поделе. За сваку поделу, било је генерисано пет модела користећи различите комбинације SMILES, графова и хибридних оптималних дескриптора са различитим индексима повезаности, петнаест модела је добијено из три насумичне неидентичне поделе. За најбољи модел из сваке поделе, корелациони коефицијент (r^2) ишао је од 0,9672 до 0,8693, док унакрсно валидирани коефицијенти корелације (Q^2) ишли су од 0,9478 до 0,8250. Просечна апсолутна грешка (*MAE*) за најбоље моделе је била мања од 0,065. Додатно, саопштене су пожељне вредности индекса идеалности корелације (*IIC*) и индекс интензитета корелације (*CII*), потврђујући робусност, поузданост и потенцијал предвиђања развијених модела. Надаље, процењени су добри и лоши отисци прстију на основу корелационих тежина структурних атрибута.

(Примљено 30. марта, ревидирано 22. септембра, прихваћено 12. новембра 2024)

REFERENCES

1. G. M. Cragg, D. J. Newman, *Biochim. Biophys. Acta Gen. Subj.* **1830** (2013) 3670 (https://doi.org/10.1016/j.bbagen.2013.02.008)
2. D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **83** (2020) 770 (https://doi.org/10.1021/acs.jnatprod.9b01285)
3. H. C. Upadhyay, *Lett. Drug Des. Discov.* **20** (2023) 373 (https://doi.org/10.2174/1570180820042301131444O4)
4. H. C. Upadhyay, G. R. Dwivedi, M. P. Darokar, V. Chaturvedi, S. K. Srivastava, *Planta Med.* **78** (2012) 79 (https://doi.org/10.1055/s-0031-1280256)
5. M. Dickson, J. P. Gagnon, *Discov. Med.* **4** (2004) 172 (PMID: 20704981)
6. M. Dickson, J. P. Gagnon, *Nat. Rev. Drug. Discov.* **3** (2004) 417 (https://doi.org/10.1038/nrd1 382)
7. H. C. Upadhyay, G. R. Dwivedi, S. Roy, A. Sharma, M. P. Darokar, S. K. Srivastava, *ChemMedChem* **9** (2014) 1860 (https://doi.org/10.1002/cmdc.201402027)
8. H. C. Upadhyay, B. S. Sisodia, H. S. Cheema, J. Agrawal, A. Pal, M. P. Darokar, S. K. Srivastava, *Nat. Prod. Commun.* **8** (2013) 1591 (https://doi.org/10.1177/1934578x1300801123)
9. H. C. Upadhyay, M. Singh, O. Prakash, F. Khan, S. K. Srivastava, D. U. Bawankule, *SN Appl. Sci.* **2** (2020) 2069 (https://doi.org/10.1007/s42452-020-03798-5)
10. M. Xiang, Y. Cao, W. Fan, L. Chen, Y. Mo, *Comb. Chem. High Throughput Screen.* **15** (2012) 328 (https://doi.org/10.2174/138620712799361825)

11. S. Surabhi, B. Singh, *J. Drug Deliv. Therapeut.* **8** (2018) 504 (https://doi.org/10.22270/jddt.v8i5.1894)
12. G. R. Dwivedi, H. C. Upadhyay, D. K. Yadav, V. Singh, S. K. Srivastava, F. Khan, N. S. Darmwal, M. P. Darokar, *Chem. Biol. Drug. Des.* **83** (2014) 482 (https://doi.org/10.1111/cbdd.12263)
13. M. Rudrapal, D. Chetia, *J. Drug Deliv. Therapeut.* **10** (2020) 225 (https://doi.org/10.22270/jddt.v10i4.4218)
14. P. Gramatica, *QSAR Comb. Sci.* **26** (2007) 694 (https://doi.org/10.1002/qsar.200610151)
15. OECD, *OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationships models*, 2004
16. E. Benfenati, A. A. Toropov, A. P. Toropova, A. Manganaro, R. Gonella Diaza, *Chem. Biol. Drug. Des.* **77** (2011) 471 (https://doi.org/10.1111/j.1747-0285.2011.01117.x)
17. A. T. K. Baidya, K. Ghosh, S. A. Amin, N. Adhikari, J. Nirmal, T. Jha, S. Gayen, *New J. Chem.* **44** (2020) 4129 (https://doi.org/10.1039/c9nj05825g)
18. K. Bagri, A. Kumar, M. Nimbhal, P. Kumar, *Mol. Simul.* **46** (2020) 777 (https://doi.org/10.1080/08927022.2020.1770753)
19. T. G. Kraljević, A. Harej, M. Sedić, S. K. Pavelić, V. Stepanić, D. Drenjančević, J. Talapko, S. Raić-Malić, *Eur. J. Med. Chem.* **124** (2016) 794 (https://doi.org/10.1016/j.ejmech.2016.08.062)
20. Y. L. Fan, X. Ke, M. Liu, *J. Heterocycl. Chem.* **55** (2018) 791 (https://doi.org/10.1002/jhet.3112)
21. H. C. Upadhyay, *Curr. Top. Med. Chem.* **21** (2021) 737 (https://doi.org/10.2174/1568026621666210303145759)
22. K. N. Mishra, H. C. Upadhyay, *Frontiers Drug Discov.* **2** (2022) 1072448 (https://doi.org/10.3389/fddsv.2022.1072448)
23. J. M. Madar, L. A. Shastri, S. L. Shastri, R. Guda, M. Holiyachi, N. S. Naik, S. Dodamani, S. Jalapure, V. A. Sungar, *Chem. Data Collect.* **17**–**18** (2018) 219 (https://doi.org/10.1016/j.cdc.2018.09.005)
24. P. Yadav, B. Kumar, H. K. Gautam, S. K. Sharma, *J. Chem. Sci.* **129** (2017) 211 (https://doi.org/10.1007/s12039-016-1214-x)
25. S. Carmel Yesudass, P. Ranjan, H. P. Suresh, *J. Heterocycl. Chem.* **59** (2022) 309 (https://doi.org/10.1002/jhet.4385)
26. M. H. Shaikh, D. D. Subhedar, B. B. Shingate, F. A. Kalam Khan, J. N. Sangshetti, V. M. Khedkar, L. Nawale, D. Sarkar, G. R. Navale, S. S. Shinde, *Med. Chem. Res.* **25** (2016) 790 (https://doi.org/10.1007/s00044-016-1519-9)
27. S. M. Sutar, H. M. Savanur, C. Patil, G. M. Pawashe, G. Aridoss, K. M. Kim, R. G. Kalkhambkar, *Chem. Data Coll.* **28** (2020) 100480 (https://doi.org/10.1016/j.cdc.2020.100480)
28. A. V. Lipeeva, D. O. Zakharov, L. G. Burova, T. S. Frolova, D. S. Baev, I. V. Shirokikh, A. N. Evstropov, O. I. Sinitsyna, T. G. Tolsikova, E. E. Shults, *Molecules* **24** (2019) 2126 (https://doi.org/10.3390/molecules24112126)
29. M. N. Joy, Y. D. Bodke, S. Telkar, V. A. Bakulev, *J. Mex. Chem. Soc.* **64** (2020) 53 (https://doi.org/10.29356/jmcs.v64i1.1116)
30. X. M. Peng, K. V. Kumar, G. L. V. Damu, C. H. Zhou, *Sci. China Chem.* **59** (2016) 878 (https://doi.org/10.1007/s11426-015-0351-0)
31. P. Kumar, A. Kumar, J. Sindhu, *SAR QSAR Environ. Res.* **30** (2019) 63 (https://doi.org/10.1080/1062936X.2018.1564067)

32. A. A. Toropov, A. P. Toropova, *Toxicol. Mech. Methods* **29** (2019) 43 (https://doi.org/10.1080/15376516.2018.1506851)
33. A. A. Toropov, E. Benfenati, *Curr. Drug. Discov. Technol.* **4** (2007) 77 (https://doi.org/10.2174/157016307781483432)
34. A. Kumar, S. Chauhan, *Arch. Pharm. (Weinheim)* **350** (2017) e1600268 (https://doi.org/10.1002/ardp.201600268)
35. A. P. Toropova, A. A. Toropov, A. Roncaglioni, E. Benfenati, *Molecules* **28** (2023) 6587 (https://doi.org/10.3390/molecules28186587).