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

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

QSAR study of unsymmetrical aromatic disulfides as potent avian SARS-CoV main protease inhibitors using quantum chemical descriptors and statistical methods

Samir Chtita [a,*], Assia Belhassan [b], Mohamed Bakhouch [c], Abdelali Idrissi Taourati [d], Adnane Aouidate [e], Salah Belaidi [f,g], Mohammed Moutaabbid [a], Said Belaaouad [a], Mohammed Bouachrine [b,h], Tahar Lakhlifi [b]

[a] Laboratory of Physical Chemistry of Materials, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, B.P. 7955, Sidi Othmane, Casablanca, Morocco
[b] Molecular Chemistry and Natural Substances Laboratory, Department of Chemistry, Faculty of Sciences, University Moulay Ismail, Meknes, Morocco
[c] Laboratory of Bioorganic Chemistry, Department of Chemistry, Faculty of Sciences, Chouaïb Doukkali University, 24000, El Jadida, Morocco
[d] Laboratory of Biological Chemistry Applied to the Environment, Department of Chemistry, Faculty of Sciences, University Moulay Ismail, Meknes, Morocco
[e] Institut de Chimie Organique et Analytique (ICOA), UMR7311 Université d'Orléans-CNRS, Rue de Chartres, BP 6759, 45067, Orléans, Cedex 2, France
[f] Laboratory of Molecular Chemistry and Environment, Group of Computational and Pharmaceutical Chemistry, University of Biskra, BP145, 07000, Biskra, Algeria
[g] Centre de Recherche en Sciences Pharmaceutiques (CRSP) la nouvelle ville Ali Mendjeli, Constantine, Alegria
[h] High School of Technology of Khenifra, Sultane Slimane University, B.P. 591, Khenifra, Morocco

ARTICLE INFO

ABSTRACT

In silico research was executed on forty unsymmetrical aromatic disulfide derivatives as inhibitors of the SARS Coronavirus (SARS-CoV-1). Density functional theory (DFT) calculation with B3LYP functional employing 6-311 + G(d,p) basis set was used to calculate quantum chemical descriptors. Topological, physicochemical and thermodynamic parameters were calculated using ChemOffice software. The dataset was divided randomly into training and test sets consisting of 32 and 8 compounds, respectively. In attempt to explore the structural requirements for bioactives molecules with significant anti-SARS-CoV activity, we have built valid and robust statistics models using QSAR approach. Hundred linear pentavariate and quadrivariate models were established by changing training set compounds and further applied in test set to calculate predicted IC_{50} values of compounds. Both built models were individually validated internally as well as externally along with Y-Randomization according to the OECD principles for the validation of QSAR model and the model acceptance criteria of Golbraikh and Tropsha's. Model 34 is chosen with higher values of R^2, R^2_{test} and Q^2cv ($R^2 = 0.838$, $R^2_{test} = 0.735$, $Q^2_{cv} = 0.757$).

It is very important to notice that anti-SARS-CoV main protease of these compounds appear to be mainly governed by five descriptors, i.e. highest occupied molecular orbital energy (E_{HOMO}), energy of molecular orbital below HOMO energy (E_{HOMO-1}), Balaban index (BI), bond length between the two sulfur atoms (S1S2) and bond length between sulfur atom and benzene ring (S2Bnz). Here the possible action mechanism of these compounds was analyzed and discussed, in particular, important structural requirements for great SARS-CoV main protease inhibitor will be by substituting disulfides with smaller size electron withdrawing groups. Based on the best proposed QSAR model, some new compounds with higher SARS-CoV inhibitors activities have been designed. Further, in silico prediction studies on ADMET pharmacokinetics properties were conducted.

1. Introduction

Since its first appearance in Southern China in November 2002, the SARS coronavirus has been recognized as a global threat [1,2]. It's an epidemic caused by severe acute respiratory syndrome SARS-CoV-1 and affected more than 8500 cases in 32 countries [3]. Symptoms are influenza-like and include high fever, malaise, myalgia, headache, non-productive cough, diarrhea, and shivering [4]. No individual

symptom or cluster of symptoms has proved to be specific for a diagnosis of SARS. Although fever is the most frequently reported symptom, it is sometimes absent on initial measurement, especially in elderly and immunosuppressed patients [5].

The SARS was successfully controlled in July 2003, however, the potential reemergence of pandemic SARS-CoV is still posing a risk. In fact, the new strain of SARS (SARS-CoV-2) is potentially more virulent than the strain of 2003 outbreak [6]. SARS-CoV encodes a main protease which plays a pivotal role in processing viral polyproteins and controlling replicas complex activity. Main protease is an enzyme indispensable for viral replication and infection processes, making it an ideal target for the design of antiviral therapies [7].

In order to understand the chemical–biological interactions and predict their activities toward SARS-CoV-1 and to open a new way in SARS inhibitors drug research, in the current work, a series of 40 unsymmetrical aromatic disulfides derivatives against SARS-CoV-1 were collected and constructed with QSAR models.

Perusal of the literature reveals a variety of methods for synthesizing disulfides and a great number of disulfide analogues had been designed and synthesized. For example, Xu Qiu et al. [8] demonstrated a novel carbonate salts catalyzed aerobic oxidative heterocoupling of thiols for the efficient synthesis of unsymmetrical disulfides; D. Branowska et al. [9] had described a series of new 1,2,4-triazine unsymmetrical disulfane analogues that were prepared and evaluated as anticancer activity compounds against MCF-7 human breast cancer cells with some of them acting as low micromolar; J. K. Vandavasi et al. [10] have developed an efficient 'one pot' method for the synthesis of unsymmetrical dithio compounds directly from corresponding thiols and thiocarboxylic acids in the presence of DDQ (2,3-Dichloro-5,6-dicyano-1,4-benzoquinone). In addition, F. Yang et al. [11] have also developed one-pot synthesis of aromatic-aromatic and aromatic-aliphatic disulfide unsymmetrical disulfide using TCCA (Trichloroisocyanuric Acid). N. Stellenboom et al. [12,13] prepared unsymmetrical glycosyl disulfides derived from sugar, alkyl/aryl or thiols. M. Bao et al. [14] have developed the *N*-Trifluoroacetyl arenesulfenamides effective precursors for the synthesis of unsymmetrical disulfides.

Disulfides exist in many synthetic and natural products and have been applied extensively in organic transformation and medicinal chemistry. As example, ajoene and dysoxysulfone are found in garlic, onions and mahogany trees and have shown promising antifungal [15,16], antibacterial [17], antitumor [9,18], antimalarial [19] and analgesic properties [20].

On the other hand, a literature survey reveals that several published papers describe the molecular modeling towards the main protease of SARS-CoV-1 and SARS-CoV-2 viruses. Thus, Alves et al. have performed QSAR studies to evaluate the ability of some known drugs to inhibit SARS-CoV-2 [21]. Other studies were reported by Masand et al. which describe the development of QSAR model from a dataset of peptide-type compounds as SARS-CoV inhibitors [22,23].

The significance and novelty of findings presented in this work are reflected from the fact that we have used quantum chemistry descriptors which describe electron proprieties of the studied molecules. The use of density functional theory (DFT) is justified for the reason that some our previously QSAR studies have shown that the descriptors calculated using the DFT method can improve the accuracy of the results and lead to more reliable QSAR models [24–26].

2. Material and methods

2.1. Selection of dataset and generation of molecular descriptors

Dataset of the inhibitor activities toward SARS Coronavirus (SARS-CoV) main protease of 40 unsymmetrical aromatic disulfides derivatives was collected from the literature [27]. Structures of the studied molecules with their activity IC_{50} (µM) values are presented in Table 1. The inhibitory activity factor IC_{50} biochemical assays spectacles the required concentration of an inhibitor to achieve 50% inhibition of replication of

SARS-CoV main protease.

To predict the correlation between the anti-SARS-CoV activity with various quantum, topological, thermodynamic and physicochemical parameters, and to develop linear models, all the three-dimensional structures were drawn and built by GaussView 06 program [28], quantum parameters were calculated by DFT approach performed with Gaussian 09 program package [29] using the hybrid functional B3LYP combining the Becke's three-parameter and the Lee-Yang-Parr exchange-correlation functional employing the 6-31G+(d,p) basis set in gas phase and all others parameters were calculated using Chem3D software [30]. The geometry of the compounds was determined by optimizing all geometrical variables with no symmetry constraints (Table S1).

2.2. Principal component analysis (PCA)

The pre-processing of the dataset is to eliminate the irrelevant descriptors in order to avoid the phenomenon of over-fitting. Therefore, we must reduce the variables (descriptors) that do not have or have little influence on the studied activity. With the XLSTAT software [31], we have used PCA to overview the examined compounds for similarities and dissimilarities in order to eliminate descriptors that are highly correlated and to select those that show a high correlation with the response activity; this one gives extra weight because it will be more effective at prediction. The most important result obtained by PCA is the correlation matrix, a diagonal matrix which represents the correlation between the activity and the descriptors retained. Descriptor with highest correlation is taken and compared to other descriptors in the correlation matrix.

2.3. Data splits and model development

Dataset was randomly split into several training set and test set before descriptors selection. It was recommended that analysis of the models should be obtained from various splits into training set (80%) and test set (20%). Then, all-subset regression for the whole dataset was obtained from the training sets and was performed using multiple linear regression (MLR) method with XLSTAT software.

We have used the stepwise MLR analysis based on the elimination of aberrant descriptors one by one, which takes the following form: $Y = a_0 + \sum_{i=1}^{n} a_i x_i$. Where: Y: the studied activity, which is, the dependent variable; a_0 QUOTE a0 a_0: the intercept of the equation; x_i: the molecular descriptors; a_i: the coefficients of those descriptors.

This method is one of the most popular methods of QSAR due to its simplicity in operation, reproducibility and ability to allow easy interpretation of the features used. The important advantage of the linear regression analysis is its transparent nature, therefore, the algorithm is accessible and predictions can be made easily [32].

2.4. Model validation

Statistical parameters for modeling, internal and external validation metrics were adopted to evaluate the fit, stability and predicative power of the QSAR model.

Quality validation parameters include Coefficient of determination (R^2), Adjusted coefficient of determination (R_{adj}^2), Mean of Squared Errors of model (MSE), Fischer's value (F_{test}), Variance Inflation Factor (VIF), Coefficient of determination of Leave-One-Out Cross Validation (Q_{CV}^2), Coefficient of determination of external test (R_{test}^2) and Y-randomization parameters (R_{Rand}^2 and $Q_{cv\ (Rand)}^2$) [33]. A model is valid only within its training domain and new molecules must be considered as belonging to the applicability domain (AD) before the model is applied (OECD Principle 3 [34]). (Supporting information).

2.4.1. Drug-likeness and ADMET properties

In drug discovery, the prediction of ADMET properties is an important study to escape the failure of drugs in the clinical phases [35].

Table 1
Structures of 40 unsymmetrical aromatic Disulfides and their activities anti-SARS-CoV MPro.

N°	Structure	IC50	N°	Structure	IC50	N°	Structure	IC50
1		1.871	8		1.762	13		1.651
2		2.803				14		2.075
3		3.675	9		5.654	15		5.954
4		3.130				16		3.957
5		1.506	10		4.511	17		4.126
6		4.344				18		2.565
7		1.871	11		5.794	19		1.947
			12		2.626	20		2.029

N°	Structure	IC50	N°	Structure	IC50	N°	Structure	IC50
21		1.250	28		1.713	35		1.991
22		2.211	29		1.118	36		1.495
23		3.321	30		1.264	37		0.883
24		2.555	31		0.516	38		0.684
25		2.452	32		0.921	39		0.697
26		1.679	33		1.437	40		1.522
27		1.557	34		1.121			

Pharmacokinetic and bioavailability predictions are an essential tool in drug discovery process and should be considered to develop a new drug. Based on the pkCSM online tool [36], the physicochemical properties of the active components were predicted, including molecular weight (MW), Partition coefficient (log P), rotatable bonds count (RB), H-bond acceptors and donors count (HBA and HBD) and polar surface area (PSA).

Lipinski's rule (with MW $\leq$ 500 g/mol, Log P $\leq$ 5, NR $\leq$ 10, HBA $\leq$ 10, HBD $\leq$ 5, PSA $\leq$ 140) has been applied to evaluate the molecules drug likeness [37]. Candidate violating no more than one of these criteria is likely to be developed as a prospective oral drug [38].

Log S was also calculated to evaluate the water solubility of the proposed compounds (compound is insoluble or poorly soluble if log S $\leq$ -6, moderately soluble if $-6 <$ log S ≤ -4, soluble if $-4 <$ log S) [39].

Finally, different ADMET properties were predicted including, Absorption (Caco-2 cell permeability, *P*-glycoprotein (*P*-gp) and Human Intestinal Absorption (HIA)), Distribution (blood-brain barrier (BBB)), Metabolism (interaction of molecules with cytochrome enzyme system P450 CYP2D6 and CYP3A4), Excretion (total clearance TC)) and Toxicity (AMES toxicity, hERG I and hERG II inhibitors). These *in silico* pharmacokinetics parameters were evaluated to prevent the failure of those compounds during clinical studies and enhance their chances to reach the stage of being drug-candidates against the SARS-CoV-1.

3. Results and discussions

3.1. Molecular descriptors

From the results of DFT(B3LYP/6-31G(d,p)) calculations, 11 quantum chemistry descriptors values were computed (Table S2). ChemOffice 3D software was used to calculate 34 others descriptors (Table S3).

3.2. Principal component analysis (PCA)

The 45 descriptors are competed for the 40 studied molecules; these descriptors were subjected to a principal component analysis. The results of this analysis are used to select the input data of multiple linear regression studies. Thus, at the beginning, we excluded all descriptors having a low correlation coefficient value (r $\leq$ 0.15) with respect to the dependent variable (IC_{50}). Instead, the descriptors with a correlation coefficient value greater than 0.95 are omitted to reduce the uncertainty present in our data matrix. The 25 descriptors presented in Tables S2 and S3 are selected by the PCA analysis and used in MLR models development.

3.3. Data splits and models development

QSAR analysis was performed using calculated molecular descriptors and experimental values of anti-SARS-CoV activity for the forty disulfides. Therefore, the whole dataset was randomly split into training and test sets by a good number of pentavariate and quadrivariate MLR models with nearly similar statistical performance but encompassing different descriptors (One hundred splits, 1–100) for the same size of training and test sets. Of the chemicals in the dataset, 32 compounds were selected for training set and remaining 8 compounds were considered as test set. The models that do not satisfy OECD principles [34] and Golbraikh and Tropsha's criteria [33] were summarily excluded. Fifty MLR models with highest coefficients of determination, explained variance in "leave one-out" cross validation prediction and with good ability to predict IC_{50} values of test set compounds were selected for the whole dataset from all splits. The splits into training and test sets results and the performances of MLR models are shown in Tables 2 and S4.

All equations models presented in Table 2 with usual meaning of the statistical symbols are statistically sound and predictive with adequate values of statistical parameters used to judge for internal and external validation of QSAR models. High values of R^2, R^2_{adj}, Q^2_{cv} and R^2_{test} and low values of MSE point out that all these models are statistically satisfactory,

Table 2

Statistical parameters and model equations for the fifty splits of training and test sets.

Model equations		R^2	R^2_{adj}	MSE	R^2_{test}	Q^2_{cv}
1	$IC_{50} = -85.468 + 1.0164$ $E_{HOMO}+ 36.289$ S1S2 + 1.081 Log S - 0.860 HLC +0.042 BP	**0.801**	0.763	**0.418**	0.655	**0.722**
2	$IC_{50} = 98.914 - 1.813$ $E_{HOMO-1} + 3.652$ E_{HOMO} + 44.737 S1S2 - 100.408 S2Bnz +5.382 10^{-06} BI	0.789	0.749	0.562	0.907	0.675
3	$IC_{50} = 87.944 + 2.948$ E_{HOMO} + 34.295 S1S2 - 76.116 S2Bnz +5.487 10^{-06} BI -0.060C%	0.761	0.715	0.564	0.819	0.627
4	$IC_{50} = 85.852 - 1.272$ $E_{HOMO-1} + 3.355$ E_{HOMO} + 40.557 S1S2 - 87.281 S2Bnz +5.363 10^{-06} BI	0.763	0.718	0.632	0.907	0.641
5	$IC_{50} = 63.514 + 1.828$ E_{HOMO} + 0.927 E_{LUMO+1} + 42.783 S1S2 - 77.343 S2Bnz +5.680 10^{-06} BI	0.789	0.749	0.566	0.617	0.639
6	$IC_{50} = -0.265 - 0.616$ E_{LUMO} + 1.906 E_{LUMO+1} + 0.852 log P + 3.537 10^{-06} BI +0.149 O%	0.752	0.704	0.522	0.617	0.602
7	$IC_{50} = 72.252 - 1.686$ $E_{HOMO-1} + 3.590$ E_{HOMO} + 44.919 S1S2 - 85.554 S2Bnz +5.202 10^{-06} BI	0.747	0.698	0.664	0.862	0.580
8	$IC_{50} = 119.399 - 1.573$ $E_{HOMO-1} + 3.848$ E_{HOMO} + 44.839 S1S2 - 110.400 S2Bnz +5.872 10^{-06} BI	**0.821**	0.787	**0.488**	0.655	**0.722**
9	$IC_{50} = -167.793 + 1.726$ E_{LUMO+1} + 58.021 S1S2 + 26.914 S1Htr +5.454 10^{-06} BI +0.139 HLC	0.742	0.692	0.572	0.694	0.605
10	$IC_{50} = -174.066 + 3.601$ E_{HOMO} + 1.977 E_{LUMO+1} + 95.158 S1S2 - 0.330 O% + 0.106 PSA	**0.852**	**0.824**	**0.336**	0.617	**0.796**
11	$IC_{50} = 105.697 + 2.507$ E_{HOMO} + 46.181 S1S2 - 103.286 S2Bnz +0.039 GFE - 0.885H%	0.772	0.728	0.453	0.735	0.640
12	$IC_{50} = -126.683 + 3.235$ E_{HOMO} + 67.333 S1S2 + 0.656 NHBA +0.011 BP - 0.176 O%	0.741	0.691	0.545	0.862	0.552
13	$IC_{50} = 120.893 - 2.510$ $E_{HOMO-1} + 4.299$ E_{HOMO} + 45.514 S1S2 - 113.825 S2Bnz +4.821 10^{-06} BI	0.743	0.694	0.508	0.953	0.584
14	$IC_{50} = 97.222 - 1.458$ $E_{HOMO-1} + 3.663$ E_{HOMO} + 42.570 S1S2 - 95.527 S2Bnz +4.692 10^{-06} BI	0.768	0.723	0.556	0.776	0.638
15	$IC_{50} = 64.336 + 2.234$ E_{HOMO} + 0.574 E_{LUMO+1} +39.641 S1S2 - 72.988 S2Bnz +5.597 10^{-06} BI	**0.800**	0.761	0.545	0.655	0.680
16	$IC_{50} = 74.931 + 2.538$ E_{HOMO} + 37.952 S1S2 - 76.311 S2Bnz +0.634 NHBD +5.498 10^{-06} BI	**0.800**	0.761	0.526	0.694	0.666
17	$IC_{50} = 83.018 + 2.112$ E_{HOMO} + 42.762 S1S2 - 87.785 S2Bnz +6.665 10^{-06} BI - 0.013 PSA	0.762	0.716	0.505	0.735	0.582
18	$IC_{50} = 56.425 + 2.848$ E_{HOMO} + 41.4501 S1S2 - 67.856 S2Bnz +4.983 10^{-06} BI - 0.039C%	0.750	0.702	0.459	0.655	0.593
19		0.755	0.708	0.507	0.776	0.566

Table 2 (*continued*)

	Model equations	R^2	R^2_{adj}	MSE	R^2_{test}	Q^2_{cv}
	$IC_{50} = 98.414 – 1.539\ E_{HOMO-1} + 3.391\ E_{HOMO} + 40.628\ S1S2 - 95.249\ S2Bnz + 5.368\ 10^{-06}\ BI$					
20	$IC_{50} = 106.474 – 1.800\ E_{HOMO-1} + 3.416\ E_{HOMO} + 40.309\ S1S2 - 100.234\ S2Bnz + 5.211\ 10^{-06}\ BI$	0.788	0.748	0.502	0.694	0.671
21	$IC_{50} = 30.182 + 3.025\ E_{HOMO} + 52.323\ S1S2 - 66.486\ S2Bnz + 0.040\ GFE - 0.930 H\%$	0.799	0.760	0.452	0.819	0.689
22	$IC_{50} = 126.976 – 2.222\ E_{HOMO-1} + 3.939\ E_{HOMO} + 49.656\ S1S2 - 122.318\ S2Bnz + 4.992\ 10^{-06}\ BI$	0.766	0.721	0.421	0.617	0.617
23	$IC_{50} = 48.294 + 1.842\ E_{HOMO} + 0.825\ E_{LUMO+1} + 43.447\ S1S2 - 69.770\ S2Bnz + 5.907\ 10^{-06}\ BI$	0.765	0.719	0.633	0.862	0.599
24	$IC_{50} = 88.883 + 2.481\ E_{HOMO} + 28.980\ S1S2 - 72.516\ S2Bnz + 5.432\ 10^{-06}\ BI - 0.045 C\%$	0.777	0.734	0.478	0.862	0.631
25	$IC_{50} = 97.940 – 1.322\ E_{HOMO-1} + 3.469\ E_{HOMO} + 44.105\ S1S2 - 97.956\ S2Bnz + 5.330\ 10^{-06}\ BI$	0.727	0.675	0.553	0.776	0.572
26	$IC_{50} = -177.107 + 3.565\ E_{HOMO} + 2.450\ E_{LUMO+1} + 96.952\ S1S2 - 0.271\ O\% + 0.090\ PSA$	0.706	0.649	0.535	0.907	0.519
27	$IC_{50} = 57.171 + 1.561\ E_{HOMO} + 40.099\ S1S2 - 72.972\ S2Bnz + 0.010\ GFE + 4.090\ 10^{-06}\ BI$	0.773	0.729	0.512	0.735	0.645
28	$IC_{50} = 115.209 – 1.762\ E_{HOMO-1} + 3.428\ E_{HOMO} + 43.393\ S1S2 - 108.561\ S2Bnz + 5.575\ 10^{-06}\ BI$	**0.820**	0.785	**0.459**	**0.735**	**0.727**
29	$IC_{50} = 30.718 + 2.616\ E_{HOMO} + 43.359\ S1S2 - 57.653\ S2Bnz + 0.589\ NHBD + 4.666\ 10^{-06}\ BI$	0.768	0.723	0.553	0.617	0.630
30	$IC_{50} = 111.463 – 1.448\ E_{HOMO-1} + 3.393\ E_{HOMO} + 45.060\ S1S2 - 107.296\ S2Bnz + 5.226\ 10^{-06}\ BI$	**0.839**	**0.808**	**0.389**	**0.776**	**0.753**
31	$IC_{50} = 103.548 – 2.125\ E_{HOMO-1} + 4.035\ E_{HOMO} + 49.253\ S1S2 - 108.085\ S2Bnz + 4.767\ 10^{-06}\ BI$	0.748	0.700	0.528	0.819	0.603
32	$IC_{50} = 109.178 – 2.098\ E_{HOMO-1} + 3.915\ E_{HOMO} + 46.178\ S1S2 - 107.950\ S2Bnz + 5.459\ 10^{-06}\ BI$	0.798	0.759	0.559	0.776	0.682
33	$IC_{50} = 121.803 – 1.602\ E_{HOMO-1} + 3.553\ E_{HOMO} + 48.317\ S1S2 - 117.039\ S2Bnz + 6.338\ 10^{-06}\ BI$	**0.848**	**0.819**	**0.408**	0.655	**0.748**
34	$IC_{50} = 128.780 – 2.590\ E_{HOMO-1} + 4.855\ E_{HOMO} + 51.701\ S1S2 - 123.760\ S2Bnz + 5.682\ 10^{-06}\ BI$	**0.838**	**0.807**	**0.453**	**0.735**	**0.757**
35	$IC_{50} = 220.048 + 0.734\ E_{HOMO} + 0.938\ E_{LUMO+1} – 119.872\ S2Bnz - 0.912\ NHBD + 0.930\ NRB$	0.775	0.732	0.511	0.819	0.605
36	$IC_{50} = 114.995 – 2.281\ E_{HOMO-1} + 4.025\ E_{HOMO} + 41.273\ S1S2 - 105.704\ S2Bnz + 5.359\ 10^{-06}\ BI$	0.763	0.717	0.545	0.862	0.619
37	$IC_{50} = 6.687 – 0.936\ E_{HOMO-1} + 1.909\ E_{LUMO+1} + 29.959\ S1S2 - 40.320\ S2Bnz + 7.103\ 10^{-06}\ BI$	0.729	0.677	0.545	0.617	0.598
38	$IC_{50} = 13.618 + 1.525\ E_{LUMO+1} + 30.697\ S1S2 - 41.729\ S2Bnz + 6.660\ 10^{-06}\ BI$	0.716	0.674	0.550	0.617	0.595
39	$IC_{50} = 68.873 + 2.072\ E_{HOMO} + 41.201\ S1S2 - 78.416\ S2Bnz + 5.146\ 10^{-06}\ BI$	0.756	0.720	0.589	0.694	0.642
40	$IC_{50} = 66.390 + 1.954\ E_{HOMO} + 39.051\ S1S2 - 74.950\ S2Bnz + 5.672\ 10^{-06}\ BI$	0.762	0.727	0.515	0.735	0.640
41	$IC_{50} = 101.177 + 2.167\ E_{HOMO} + 40.756\ S1S2 - 95.484\ S2Bnz + 5.528\ 10^{-06}\ BI$	0.794	0.764	0.504	0.735	0.699
42	$IC_{50} = 55.547 + 2.600\ E_{HOMO} + 47.887\ S1S2 - 76.796\ S2Bnz + 4.858\ 10^{-06}\ BI$	0.743	0.705	0.590	0.735	0.608
43	$IC_{50} = 62.052 + 2.510\ E_{HOMO} + 37.292\ S1S2 - 68.345\ S2Bnz + 5.065\ 10^{-06}\ BI$	0.776	0.743	0.583	0.776	0.682
44	$IC_{50} = 101.404 + 2.313\ E_{HOMO} + 42.408\ S1S2 - 96.970\ S2Bnz + 5.231\ 10^{-06}\ BI$	**0.820**	0.793	**0.420**	0.694	**0.734**
45	$IC_{50} = 107.565 + 2.328\ E_{HOMO} + 46.599\ S1S2 - 105.417\ S2Bnz + 6.280\ 10^{-06}\ BI$	**0.827**	**0.801**	**0.447**	0.617	**0.728**
46	$IC_{50} = 111.430 + 2.796\ E_{HOMO} + 48.125\ S1S2 - 107.524\ S2Bnz + 5.663\ 10^{-06}\ BI$	0.796	0.766	0.549	0.819	0.705
47	$IC_{50} = 70.676 + 2.205\ E_{HOMO} + 40.127\ S1S2 - 77.609\ S2Bnz + 5.724\ 10^{-06}\ BI$	0.791	0.760	0.521	0.694	0.696
48	$IC_{50} = 211.172 + 1.308\ E_{LUMO+1} + 0.990\ NRB - 117.448\ S2Bnz - 1.075\ NHBD$	0.762	0.727	0.520	0.735	0.656
49	$IC_{50} = 100.017 + 2.555\ E_{HOMO} + 39.880\ S1S2 - 92.409\ S2Bnz + 5.305\ 10^{-06}\ BI$	0.707	0.664	0.541	0.694	0.585
50	$IC_{50} = 72.115 + 2.348\ E_{HOMO} + 41.353\ S1S2 - 79.341\ S2Bnz + 5.187\ 10^{-06}\ BI$	0.751	0.714	0.685	0.819	0.645

robust and also possess good external predictive ability.

For all developed models, values of R^2_{adj} are quite close to R^2 suggesting that number of descriptors in the models is not too high, thereby, indicating that the models are free from over-fitting [40]. This is further supported by the low MSE values. Values of Cross Validation parameter Q^2_{cv}, are high, thereby, indicating good statistical robustness of models. High values of R^2_{test} indicate that models possess high external predictive ability. In short, the developed models satisfy the recommended interrelations and threshold values for various statistical parameters suggested by different researchers.

According to the R^2 and R^2_{adj} values for the fifty proposed models in Table 2, it's clear that models 10, 33, 30, 34, 45, 8, 44, 28, 1, 15 and 16 are, in this order, the first-class MLR models (we chose models with $R^2 \geq 0.800$ and $R^2_{adj} \geq 0.760$). However, looking at the others statistical parameters (MSE, R^2_{test} and Q^2_{cv}) we can suggest models 28, 30 and 34 as the most desirable three models. The three pentavariate MLR equations are able to predict IC_{50} values for the disulfide derivatives.

In addition, evaluation of applicability domains of these top three

models shows that only model 34 that have no responses outside or outlier in *Williams* plots (Fig. 1). Applicability domains were evaluated by leverage analysis expressed as *Williams* plot, in which standardized residuals and the leverage threshold values h* = 0.563 (h* = 3*(k+1)/n); k = 5, n = 32) were plotted. Any new value of predicted pIC_{50} data must be considered reliable only for those compounds that fall within this AD on which the model was constructed. Compounds with hi>h* or with standardized residual greater than y = ±3 can be considered as chemically different from the data set compounds and, thus, outside or outline the AD. From Fig. 1, it is obvious that all compounds in training and test sets satisfy outlier/ outside criteria for model 34. There is no response outlier in training set and no response outside in test set; only one compound (N° 14) has a residual out of the ±3 times standard deviation interval.

3.4. Y-randomization test for model 34

In this step, all calculations were repeated with randomized activities of the training set compounds as well to evaluate model robustness (y-randomization test). In the present case, 100 random trials were run for the MLR model. None of the random trials could match the original model (Table S5). The standalone QSAR-tools, available online at http://teqip.jdvu.ac.in/QSAR_Tools, were employed in the y-randomization.

The average value of R_{Rand}, R^2_{Rand} and $Q^2_{cv\ (Rand)}$ are 0.413, 0.183 and −0.272 respectively, the cR^2_p value equal a 0.847, and all the new QSAR models having significantly low R^2_{Rand} and $Q^2_{cv\ (Rand)}$ values for the 100 trials, which confirm that the developed QSAR models are robust.

The *p-value* is lower than 0.0001, it means that we would be taking a lower than 0.01% risk in assuming that the null hypothesis is wrong. The high correlation coefficient R (0.915) indicates the susceptibility of descriptors (E_{HOMO}, E_{HOMO-1}, BI, S2Bnz and S1S2) to form the above model and do bring a significant amount of information.

Further, the generated model has achieved high activity-descriptor relationship efficiency of 84% as shown by the regression-coefficient (R^2 = 0.838). The large adjusted regression-coefficient R^2 (R^2_{adj}) value presented in the generated MLR model and its closeness to the value of regression-coefficient (R^2) indicates that the developed model has perfect descriptive ability to descriptors in it and it further illustrates the true impact of used descriptors on the IC_{50}. Cross-validated square correlation coefficient (Q^2_{cv}) by LOO approach was 0.757 which showed a good internal predictive ability of the model. The low R^2 and Q^2_{cv} values obtained

for all the random models by shown in Table 2 indicate that there is no chance of correlation or structural dependency in the proposed model. The high R^2_{test} as shown in the developed model (R^2_{test} = 0.735) explains that the generated model can provide a good and valid prediction for the new compounds. Consequently, we can conclude with confidence that model 34 can be considered as a perfect model with both high statistical significance and excellent predictive ability and thus, can be used as a reliable tool for discovering anti-SARS-CoV with novel disulfides.

The activity values and the correlation diagram with calculated IC_{50} versus experimental IC_{50} of the best model (model 34) of training and test sets are shown in Table 3 and Fig. 2. VIF values of the five descriptors are smaller than 5.0 (4.785, 3.794, 1.217, 1.266 and 1.492 for E_{HOMO}, E_{HOMO-1}, BI, S2Bnz and S1S2, respectively) indicating that there is no multicollinearity among selected descriptors and resulting model has good stability [41].

3.5. Golbraikh and Tropsha's criteria

The results of model 34 were compared with threshold values of the Golbraikh and Tropsha's acceptable limit. The results in Table 4 reflected the reliability and acceptability of our proposed model.

3.6. Design of new compounds

In the equation of model 34, Balaban index (BI), highest occupied molecular orbital energy (E_{HOMO}) and bond length between the two sulfur atoms (S1S2) promote activity, while molecular orbital energy below HOMO energy (E_{HOMO-1}) and bond length between sulfur atom and the benzene ring (S2Bnz) increases activity.

Comparing the significance of each descriptor on IC_{50} activity, one must know the standardized coefficient or *t-test* values in the model equation. The bigger absolute value of *t-test* value is, the greater influence of descriptor is. *T-test* values for our model descriptors are 5.031, −2.595, 8.162, −5.080 and 5.425 for E_{HOMO}, E_{HOMO-1}, BI, S2Bnz and S1S2, respectively.

Our best MLR model clearly show that the most relevant factors to the anti-SARS-CoV activity of disulfide derivatives are steric characteristics (71% of the variance in IC_{50}) related, on one hand, with the size and volume of the substituent described by Balaban index and, on the other hand, with the distances parameter described by the bond length between the two sulfur atoms and between sulfur atom and benzene ring,
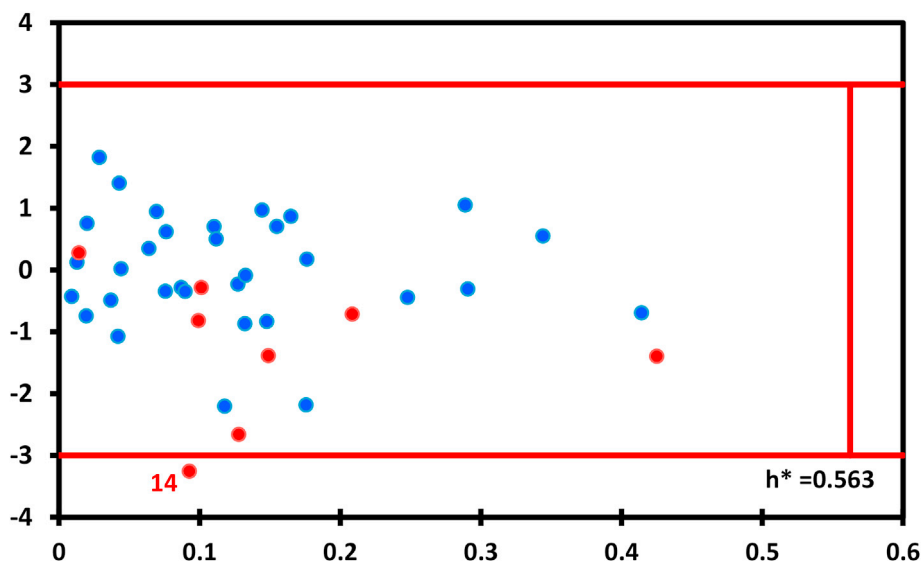

Fig. 1. Williams plot of standardized residual versus leverage for the best MLR model (model 34) (
train samples in black color and test samples in red color). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Observed and predicted activities by model 34.

N°	Observed IC$_{50}$	Predicted IC$_{50}$	Error	N°	Observed IC$_{50}$	Predicted IC$_{50}$	Error
1	1.871	2.163	−0.292	21*	1.250	2.648	−1.398
2	2.803	2.675	0.128	22	2.211	2.203	0.008
3	3.675	3.660	0.015	23	3.321	2.285	1.036
4	3.130	1.997	1.133	24	2.555	2.263	0.292
5*	1.506	1.837	−0.331	25	2.452	2.365	0.087
6	4.344	3.617	0.727	26*	1.679	1.776	−0.097
7	4.100	5.465	−1.365	27	1.557	1.999	−0.442
8*	1.762	3.258	−1.496	28	1.713	1.338	0.375
9*	5.654	4.685	0.969	29*	1.118	1.217	−0.099
10	4.511	4.475	0.036	30	1.264	1.907	−0.643
11	5.794	5.547	0.247	31	0.516	1.139	−0.623
12	2.626	2.176	0.450	32	0.921	1.696	−0.775
13	1.651	2.211	−0.560	33	1.437	1.529	−0.092
14*	2.075	3.905	−1.830	34	1.121	1.657	−0.536
15	5.954	4.786	1.168	35	1.991	1.322	0.669
16	3.957	4.395	−0.438	36	1.495	1.725	−0.230
17*	4.126	3.437	0.689	37	0.883	1.154	−0.271
18	2.565	2.372	0.193	38	0.684	0.657	0.027
19	1.947	2.448	−0.501	39	0.697	0.518	0.179
20	2.029	2.273	−0.244	40	1.522	1.283	0.239

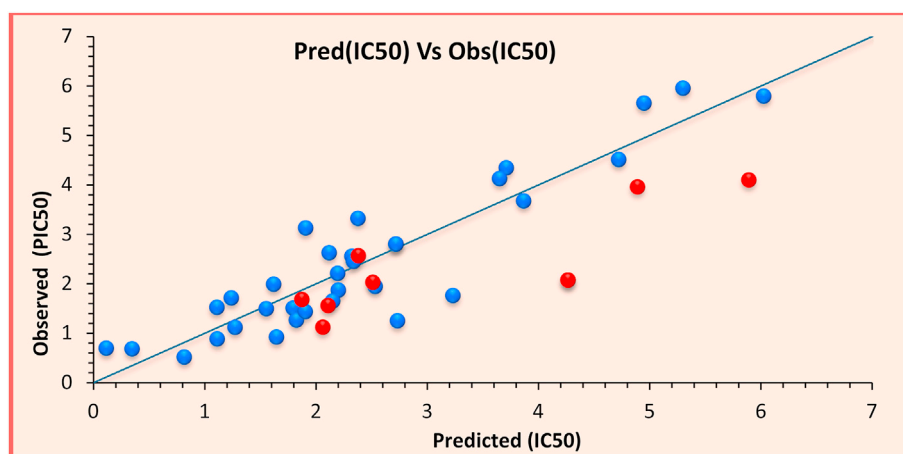* refer to test set compounds.

Fig. 2. Correlations of observed and predicted activities values calculated using model 34
(training set in blue and test set in red). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and by electronic characteristics (29% of the variance in IC$_{50}$) related with the E$_{HOMO}$ and E$_{HOMO-1}$.

By interpreting the descriptors contained in QSAR model, it is possible to gain some insights into factors, which are related to anti-SARS-CoV activity. For this reason, an acceptable interpretation of the selected descriptors is provided below:

- Balaban index (BI) of a molecular graph calculates the average distance sum connectivity index. It is describes very well the degree of ramification of non-cyclic molecules [42]. In the model equation, BI mean effect has a positive sign in the model and variation in BI accounts for 31% of the variance in IC$_{50}$, which suggests that increased activity (decreased IC$_{50}$) can be achieved by decreasing the ramification of molecular skeleton.

- The bond length between the two sulfur atoms (S1S2) has a positive sign in the model and variation in S1S2 accounts for 21% of the variance in IC$_{50}$, which suggests that increased activity can be achieved by substitute the molecular skeleton with stronger electron withdrawing ability group decrease S–S bond lengths. A relatively neutral or electron-withdrawing group in only one ortho position of phenyl (or any substituents at any more distant positions) allows the S–S bond to be short [43].

- The bond length between sulfur atom and benzene ring (S2Bnz) has

a negative sign in the model and variation in S2Bnz accounts for 19% of the variance in IC$_{50}$, suggesting that increased activity can be achieved by substitute the molecular skeleton with weaker donating electron ability group that can decrease the S2Bnz bond length. The bigger the bond length between sulfur atom and benzene ring is, the weaker conjugated π system via mesomerism or inductive effects, and higher the activity is.
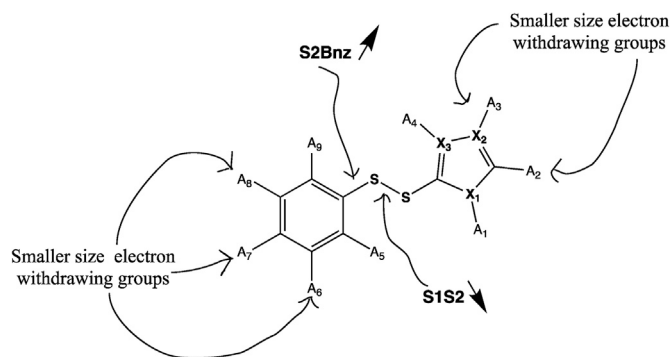
- The energy of HOMO is directly related to the ionization potential and characterizes the susceptibility of the molecule toward attack by

Table 4
Comparison of the statistical parameters of model 34 and Golbraikh and Tropsha's criteria.

Parameter	Equation	Model score	Threshold	Comment
R^2	$R^2 = 1 - \dfrac{\sum (Y_{obs} - Y_{calc})^2}{\sum (Y_{obs} - \overline{Y_{obs}})^2}$	0.832	> 0.600	**Passed**
R^2_{adj}	$R^2_{adj} = \dfrac{(N-1)R^2 - p}{N - p - 1}$	0.802	> 0.600	**Passed**
R^2_{test}	$R^2_{test} = 1 - \dfrac{\sum (Y_{calc}\,(test) - Y_{obs}(test))^2}{\sum (Y_{obs}\,(test) - \overline{Y_{obs}}(train))^2}$	0.737	> 0.600	**Passed**
Q^2_{cv}	$Q^2_{cv} = 1 - \dfrac{\sum (Y_{Calc} - Y_{Obs})^2}{\sum (Y_{Obs} - \overline{Y_{obs}})^2}$	0.740	> 0.500	**Passed**
MSE	$MSE = \dfrac{\sum (Y_{Obs} - Y_{Calc})}{N}$	0.483	A low value	**Passed**
F_{test}	$F_{test} = \dfrac{\sum (Y_{calc} - \overline{Y_{Calc}})^2}{\sum (Y_{Obs} - Y_{Calc})^2} \dfrac{N - p - 1}{p}$	27.654	a high value	**Passed**
R^2_{Rand}	Average of the 100 R^2_{Rand} (i)	0.142	$< R^2$	**Passed**
$Q^2_{cv\ LOO\ (Rand)}$	Average of the 100 $Q^2_{cv\ LOO\ (Rand)}$ (i)	-0.270	$< Q^2_{cv}$	**Passed**
cR^2_p	$cR^2_p = R*\sqrt{(R^2 - (Average\ R_{rand})^2)}$	0.764	> 0.500	**Passed**

Y_{obs} and Y_{calc}: refer to the observed and calculated/predicted response values.

$\overline{Y_{obs}}$ and $\overline{Y_{calc}}$: refer to the mean of the observed and calculated/predicted response values.

N and p refer to the number of data points (compounds) and descriptors.

electrophiles. Hard nucleophiles have a low-energy HOMO, soft nucleophiles have a high energy HOMO. Hence, molecule with high energy HOMO will give up electrons more easily because it does not cost much to donate these electrons toward making a new bond [32,44]. The contribution of E_{HOMO} in describing anti-SARS-CoV activity may be attributed to the interaction of disulfide derivatives with nucleophilic amino acid residue of microorganisms. E_{HOMO} has a positive sign in the model and variation in BI accounts for 19% of the variance in IC_{50}, which suggests that the higher of E_{HOMO}, the weaker donating electron ability, is showing the fact that the nucleophilic reaction occurs more easily and the activity of the compound is higher [45]. Consequently, if we want to decrease the value of IC_{50}, we will decrease E_{HOMO} for which we must substitute the disulfide derivatives for a weaker donating electron ability group that removes electron density (don't donates density) from the conjugated π system via mesomerism effect, making it less reactive.

- E_{HOMO-1} has a negative sign in the model. This sign suggests that the anti-SARS-CoV activity is inversely related to this descriptor. Whereas, the significance of this descriptor in the activity when its compared to the other descriptors is very weak and account for only 10% of the variance in IC_{50}.

In the conclusion, these results illustrates that to increase the anti-SARS-CoV, decrease IC_{50}, we will substitute the disulfide derivatives with smaller size electron withdrawing groups such as Nitro (NO_2), Sulfonic acid (SO_3H), Cyano (CN), Trifluoromethyl (CX_3), Haloformyl (COX), Carboxyl (CO_2H), alkoxycarbonyl (CO_2R), Acyl (COR), Formyl (CHO), halogens (X) …

The results obtained by the best MLR model (model 34) are very sufficient to conclude the performance of the models. Consequently, we can design new compounds with improved values of activity than the studied compounds using this model. The *in-silico* screening method was achieved by deletion, insertion, and substitution of various substitutes at different positions on the original templates of molecules and the results of the structural adjustments on the biological activity were studied. Therefore, the *in-silico* screening was employed to design novel compounds with good IC_{50} based on the built model and was validated by the proposed model equation:

$$IC_{50} = 128.780 - 2.590\ E_{HOMO-1} + 4.855\ E_{HOMO} + 51.701\ S1S2 - 123.760\ S2Bnz + 5.682\ 10^{-06}\ BI$$

Therefore, this suggested model will reduce the time and the cost of synthesis as well as the determination of the anti-SARS-CoV activity for the unsymmetrical aromatic disulfide derivatives.

The proposed model using 2D-QSAR suggests that the studied activity

study is highly affected by steric and electrostatic. These outcomes were supported by those obtained by L. Wang et al. [27] using CoMFA analysis.

According to the above discussions, our proposed model could be applied to other unsymmetrical aromatic disulfide derivatives accordingly to Table 1 and could add further knowledge in the improvement of new way in anti-SARS-CoV drug research. If we develop a new compound with better values than the existing ones, it may give rise to the development of more active compounds than those currently in use.

For this purpose, compounds 31 and 38 was selected as templates because they had relatively highest anti-SARS-CoV activity ($IC_{50} = 0.516$ and 0.684, respectively). The molecules were adjusted in such a way that their synthesis was experimentally achievable. Next, *in-silico* screen was employed by replacing various groups in R1 to R4 sites of the benzene ring; which lead to compounds with improved predicted anti-SARS-CoV activity values as shown in Table 5.

From the predicted activities, it has been observed that all the designed compounds (X1 to X12, and Y1 to Y10) have good IC_{50} values compared to the 40 studied compounds in Table 1.

Compounds X3 and X6 are defined as outliers and consequently they are not be considered, because they have higher leverage which is greater than h* = 0.563; we suggest all other twenty compounds for a drug-likeness and an ADMET studies.

3.7. Drug-likeness

The eminent Rule of Five by Lipinski helps to evaluate the drug-likeness of a chemical compound or determine if a compound has the properties that would make it a potential orally active drug for humans [46]. As reported by Lipinski, an orally active drug should not breach more than one of the following rules: hydrogen bond acceptor ≤10, octanol-water partition coefficient <5, hydrogen bond donor ≤5, molecular weight <500Da and topological polar surface area <140. The results of the Lipinski's calculations using pkCSM online software are depicted in Table 6.

These results suggest that all proposed compounds show good result and are in agreement with this rule.

Hence, it suggests that all proposed compounds present acceptable bioavailability of oral medications. In addition, all these compounds show moderate to good water solubility, the log S value being between −6 and −2 and thus could facilitate good oral adsorption.

Table 5

Values of descriptors, calculated anti-SARS-CoV activity and leverages (h) for the new designed unsymmetrical aromatic disulfide derivatives.

	R_2	R_3	R_5	R_6	BI	S1S2	S2Bnz	E_{HOMO}	E_{HOMO-1}	IC_{50}	h_i
31	H	H	H	H	47752	2.130	1.791	−7.433	−7.484	0.815	0.148
X1	H	NO₂	H	H	114215	2.127	1.793	−7.684	−7.684	0.069	0.496
X2	CN	H	H	H	87155	2.131	1,791	−7.717	−7.759	0.413	0.500
X3	H	CN	H	H	87155	2.128	1,793	−7.718	−7.899	0.322	0.575
X4	CHO	H	H	H	84981	2.131	1,800	−7.560	−7.737	0.049	0.376
X5	H	CHO	H	H	87155	2.128	1,794	−7.542	−7.725	0.593	0.374
X6	COOH	H	H	H	110547	2.108	1,814	−6.962	−7.733	0.150	0.707
X7	H	F	H	H	64575	2.128	1,793	−7.516	−7.603	0.543	0.319
X8	H	Cl	H	H	64575	2.129	1,792	−7.535	−7.568	0.379	0.331
X9	H	Br	H	H	64575	2.128	1,792	−7.464	−7.568	0.707	0.278
X10	H	F	F	H	85224	2.127	1,793	−7.685	−7.778	0.119	0.500
X11	H	Cl	Cl	H	85224	2.127	1,794	−7.653	−7.710	0.069	0.457
X12	H	Br	Br	H	85224	2.127	1,794	−7.581	−7.626	0.194	0.384
	R_2	R_3	R_5	R_6	BI	S1S2	S2Bnz	EHOMO	E_{HOMO-1}	IC_{50}	h_i
38	H	H	H	H	66628	2.101	1.796	−6.843	−6.973	0.346	0.112
Y1	H	CN	H	H	117275	2.100	1.797	−7.089	−7.316	0.140	0.128
Y2	H	NO2	H	H	151406	2.099	1.796	−7.135	−7.259	0.043	0.093
Y3	H	H	COOH	H	151406	2.099	1.796	−6.976	−7.021	0.222	0.149
Y4	H	H	F	H	88474	2.100	1.796	−6.946	−7.146	0.301	0.014
Y5	H	Cl	H	H	88474	2.100	1.796	−6.933	−7.104	0.274	0.209
Y6	H	Br	H	H	88474	2.100	1.796	−6.951	−7.037	0.004	0.101
Y7	H	H	COCl	H	151406	2.099	1.796	−7.090	−7.256	0.205	0.099
Y8	H	H	COCH3	H	151406	2.099	1.797	−6.955	−6.994	0.149	0.425
Y9	H	H	COOCH3	H	195234	2.100	1.796	−6.911	−7.019	0.787	0.128
Y10	H	H	H	COCH3	146622	2.102	1.799	−7.023	−7.141	0.017	0.093

Table 6

Prediction of molecular properties of descriptors for the new designed compounds.

Compound	Lipinski's parameters						Number of violations	Water solubility	
	MW	Log P	RB	HBA	HBD	PSA		Log S	Class
X1	289.725	3.431	4	7	0	108.293	0	−4.255	Moderately
X2	269.738	3.394	3	6	0	104.398	0	−4.248	Moderately
X4	272.738	3.335	4	6	0	104.166	0	−3.979	Soluble
X5	272.738	3.335	4	6	0	104.166	0	−3.991	Soluble
X7	262.718	3.662	3	5	0	97.806	0	−3.798	Soluble
X8	279.173	4.176	3	5	0	103.943	0	−4.509	Moderately
X9	323.624	4.285	3	5	0	107.508	0	−4.652	Moderately
X10	280.708	3.801	3	5	0	101.971	0	−3.987	Soluble
X11	313.618	4.829	3	5	0	114.247	0	−5.340	Moderately
X12	402.520	5.047	3	5	0	121.375	1	−5.613	Moderately
Y1	279.777	3.801	3	5	0	111.635	0	−4.633	Moderately
Y2	299.764	3.838	4	6	0	115.530	0	−4.743	Moderately
Y3	298.776	3.628	4	5	1	116.198	0	−4.154	Moderately
Y4	272.757	4.069	3	4	0	105.043	0	−4.281	Moderately
Y5	289.212	4.583	3	4	0	111.181	0	−4.979	Moderately
Y6	333.663	4.692	3	4	0	114.745	0	−5.121	Moderately
Y7	317.222	4.308	4	5	0	121.707	0	−5.170	Moderately
Y8	296.804	4.132	4	5	0	117.769	0	−4.469	Moderately
Y9	312.803	3.716	4	6	0	122.882	0	−4.481	Moderately
Y10	296.804	4.132	4	5	0	117.769	0	−4.474	Moderately
Threshold	MW ≤ 500	Log P ≤ 5	RB ≤ 10	HBA ≤10	HBD ≤5	PSA ≤140	N. Viol ≤1	Log S ≥ −6	

3.8. ADMET properties

Absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of designed sulfide derivatives were predicted using pkCSM (Table 7).

The blood-brain barrier (BBB) permeation is a prominent property in the pharmaceutical field, it helps to determine whether or not a compound can or not cross the BBB and thus exert its therapeutic effect on the brain [47]. Based on BBB report, it is clear that all proposed compounds, except X1, are capable of crossing the BBB through by passive diffusion,

Table 7
Prediction of ADMET properties for the new designed compounds.

	Absorption and Distribution						Metabolism				Excretion and Toxicity		
	BBB	Caco-2	HIA	Skin log (Kp)	P-gp Substrate	P-gp Inhibitor	CYP2D6 Substrate	CYP3A4 Substrate	CYP2D6 Inhibitor	CYP3A4 Inhibitor	TC	AMES Tox	hERG I/II
X1	−**1.108**	0.810	90.416	−2.583	No	No	No	Yes	No	No	0.088	Yes	No
X2	0.098	0.940	93.261	−2.683	No	No	No	No	No	No	0.015	No	No
X4	−0.102	1.393	93.773	−2.725	No	No	No	No	No	No	−0.011	No	No
X5	−0.100	1.392	92.855	−2.683	No	No	No	No	No	No	0.054	No	No
X7	0.772	1.965	91.218	−2.244	No	No	No	No	No	No	0.045	No	No
X8	0.466	1.836	90.623	−2.172	No	No	No	No	No	No	0.122	No	No
X9	0.465	1.835	90.556	−2.175	No	No	No	No	No	No	−0.120	No	No
X10	0.406	2.070	90.571	−2.421	No	No	No	No	No	No	0.002	No	No
X11	0.466	1.846	88.962	−2.264	No	No	No	Yes	No	No	0.227	No	No
X12	0.463	1.844	88.828	−2.277	No	No	No	Yes	No	No	−0.261	No	No
Y1	0.321	1.450	94.802	−2.335	No	No	No	No	No	No	−0.083	No	No
Y2	−0.940	0.824	91.269	−2.572	No	Yes	No	Yes	No	No	−0.077	Yes	No
Y3	0.077	1.171	95.056	−2.73	No	No	No	No	No	No	−0.054	No	No
Y4	0.099	2.014	92.117	−2.178	No	No	No	No	No	No	−0.120	No	No
Y5	0.011	1.890	91.522	−2.105	No	No	No	Yes	No	No	−0.043	No	No
Y6	−0.006	1.889	91.455	−2.112	No	No	No	Yes	No	No	−0.285	No	No
Y7	0.238	1.645	93.740	−2.613	No	No	No	No	No	No	−0.189	No	No
Y8	0.332	1.992	93.037	−2.354	No	No	No	Yes	No	No	−0.165	No	No
Y9	0.159	1.433	94.910	−2.734	No	No	No	Yes	No	No	0.064	No	No
Y10	0.332	1.976	93.148	−2.361	No	No	No	Yes	No	No	−0.233	No	No

without upsetting the normal central nervous system (CNS) functions.

P-glycoprotein (*P*-gp) is a *trans*-membrane efflux pump that transport drugs away from the cytoplasm and cell membrane causing compounds to undergo farther metabolism and clearance, thereby limiting cellular uptake of drugs resulting in therapeutic failure because the drug concentration would be lower than expected [46,48].

The study showed that only compound Y2 can be an inhibitor for *P*-glycoprotein, responsible for drug effluxes and various compounds to undergo further metabolism and clearance.

The intestine is normally the primary site of a drug being absorbed from an orally administered solution. This method is constructed to predict the proportion of compounds that have been absorbed through the small intestine of humans. It estimates the percentage for a given compound that will be consumed in the human intestine. A molecule with less than 30% absorbance is considered poorly absorbed [48]. Based on the predicted values of HIA, all the proposed compounds can be absorbed through human intestines.

The skin permeability, expressed as the skin permeability constant log (Kp), (A compound is considered to have relatively low skin permeability if it has kog Kp(cm/h)) is also an important parameter in the pharmaceutical industry to determine the risk of compounds in case there is direct contact with skin. The more negative the log (Kp) value, the less skin permeate is the molecule [49]. Hence, all proposed compounds are found to be poorly permeable to skin and accidental contact will not have any effect on the skin.

The cell line Caco-2 is composed of cells of human epithelial adenocarcinoma. The cell monolayer Caco-2 is commonly used to predict the absorption of orally administered drugs through an *in vitro* model of the human intestinal mucosa [48]. A compound is considered to be extremely permeable to Caco-2 should translate into expected values>0.90. It is obvious from the Caco-2 values in Table 7 that all proposed compounds, except for X1 and Y2, can be considered to be highly permeable to Caco-2.

Drug clearance is measured by the proportionality constant CLtot (Low value of total clearance (logCLtot) means high drug half lifetime), and occurs primarily as a combination of hepatic clearance (metabolism in the liver and biliary clearance) and renal clearance (excretion via the kidneys). It is related to bioavailability, and is important for determining dosing rates to achieve steady-state concentrations. All compounds have a low value of total clearance which means high drug half lifetime of these compounds.

The Ames toxicity test is a tool commonly used to determine mutagenic ability of a compound using bacteria. A positive test indicates the compound is mutagenic, and can therefore act as a carcinogen. Most proposed new compounds, except for X1 and Y2, are likely to be AMES-negative and thus non-mutagenic.

hERG of the potassium channels encoded by hERG (Human ether-a-go-go gene) are the principal causes for the development of squire long QT syndrome - leading to fatal ventricular arrhythmia. Inhibition of hERG channels has resulted in the withdrawal of many substances from the pharmaceutical market. All proposed compounds are likely to be non-hERG I/II inhibitor as shown in Table 7.

In conclusion, based on the Drug-likeness and ADMET studies, we suggest thirteen compounds, including X2, X3, X4, X5, X6, X7, X8, X9, X10, Y1, Y3, Y4 and Y7, which present good absorption, distribution and metabolism properties, and they present low total clearance property and show no AMES mutagenicity or hERG inhibition properties, as promising inhibitors candidates for the main protease of SARS-CoV-1 that will be synthesized and evaluated as SARS-CoV inhibitory drugs.

4. Conclusion

In this study, we have used multi-MLR approaches as linear feature QSAR method to interpret the relationship between SARS-CoV inhibitory activity for forty unsymmetrical aromatic Disulfide derivatives and their chemical structural descriptors.

The above QSARs study describing the anti-SARS-CoV activity of disulfides revealed that the most relevant factors to the anti-SARS-CoV activity of disulfide derivatives are steric characteristics (71% of the variance in IC_{50}) related, firstly, with the size and volume of the substituent described by Balaban index and, secondly, with the distances parameter described by the bond length between the two sulfur atoms and between sulfur atom and benzene ring, and finally by electronic characteristics (29% of the variance in IC_{50}) related with the E_{HOMO} and E_{HOMO-1}.

The results suggest that derivatives of unsymmetrical aromatic Disulfide with the following structural feature may exhibit great anti-SARS-CoV activity by substituting disulfides with smaller size electron withdrawing groups. According to the developed model, the most important findings of this research are that we have designed and suggest some new compounds with possible great activities. Consequently, the proposed models can be used in anti-SARS-CoV drug research for the unsymmetrical aromatic Disulfide derivatives.

ADMET evaluation shows that 13 compounds passed the stringent

lead-like criteria and *in silico* drug-likeness test, which are excellent candidates for drug discovery and are expected to be developed as prospective oral drugs.

These results encourage the collaboration with pharmacologists, academic or industrial, because the last ones many times are groping new drugs.

CRediT authorship contribution statement

Samir Chtita: Resources, Conceptualization, Methodology, Formal analysis, Writing - review & editing. **Assia Belhassan:** Writing - review & editing, Data curation. **Mohamed Bakhouch:** Writing - review & editing, Data curation. **Abdelali Idrissi Taourati:** Writing - review & editing, Data curation. **Adnane Aouidate:** Writing - review & editing, Data curation. **Salah Belaidi:** Visualization, Writing - review & editing. **Mohammed Moutaabbid:** Visualization, Writing - review & editing. **Said Belaaouad:** Visualization, Writing - review & editing. **Mohammed Bouachrine:** Project administration, Supervision. **Tahar Lakhlifi:** Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2021.104266.

References

[1] C. Drosten, S. Gunther, W. Preiser, S. VenderWerf, H.R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolensnikova, R.A.M. Fouchier, A. Berger, A.M. Burguiere, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J. Manuguerra, S. Muller, V. Rickerts, M. Sturmer, S. Vieth, H.D. Klenk, A.D.M.E. Osterhaus, H. Schmitz, H.W. Doerr, N. Engl. J. Med. 348 (2003) 1967e1976.
[2] N. Lee, D. Hui, A. Wu, P. Chan, P. Cameron, F.M. Joynt, A. Ahuja, M.Y. Yung, C.B. Leung, K.F. To, M.D. Leu, C.C. Szeto, S. Chung, J.J.Y. Sung, N. Engl. J. Med. 348 (2003) 1986e1994.
[3] Sho Konno, Pillaiyar Thanigaimalai, Takehito Yamamoto, Kiyohiko Nakada, Rie Kakiuchi, Kentaro Takayama, Yuri Yamazaki, Fumika Yakushiji, Kenichi Akaji, Yoshiaki Kiso, Yuko Kawasaki, Shen-En Chen, Ernesto Freire, Yoshio Hayashi, Design and synthesis of new tripeptide-type SARS-CoV 3CL protease inhibitors containing an electrophilic arylketone moiety, Bioorg. Med. Chem. 21 (2013) 412–424.
[4] N. Lee, D. Hui, A. Wu, P. Chan, P. Cameron, G.M. Joynt, A. Ahuja, M.Y. Yung, C.B. Leung, K.F. To, S.F. Lui, C.C. Szeto, S. Chung, J.J.Y.N. Sung, Engl. J. Med. 348 (2003) 1986.
[5] WHO, SARS (Severe acute respiratory syndrome) - disease information. https://www.who.int/ith/diseases/sars/en/, 2003.
[6] Y. Yang, F. Peng, R. Wang, K. Guan, T. Jiang, G. Xu, J. Sun, C. Chang, The deaordly conaviruses: the 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China, J. Autoimmun. 109 (2020) 102434, https://doi.org/10.1016/j.jaut.2020.102434.
[7] V. Thiel, K.A. Ivanov, A.A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weiabrich, E.J. Snijder, H. Rabenau, H.W. Doerr, J. Ziebuhr, J. Genet. 84 (2003) 2305–2315.
[8] Xu Qiu, Xiaoxue Yang, Yiqun Zhang, Song Song, Jiao Ning, Efficient and practical synthesis of unsymmetrical disulfides via base-catalyzed aerobic oxidative dehydrogenative coupling of thiols, Organic Chemistry Frontiers 6 (2019) 2220–2225.
[9] D. Branowska, J. Ławecka, M. Sobiczewski, Z. Karczmarzyk, W. Wysocki, E. Wolinska1, E. Olender, B. Mirosław, A. Perzyna, A. Bielawska, K. Bielawski, Synthesis of unsymmetrical disulfanes bearing 1,2,4-triazine scaffold and their in vitro screening towards anti-breast cancer activity, Monatshefte für Chemie - Chemical Monthly 149 (2018) 1409–1420.
[10] K. Vandavasi, W.P. Hu, C.Y. Chen, J.J. Wang, Efficient synthesis of unsymmetrical disulfides, Tetrahedron 67 (2011) 8895–8901.
[11] F. Yang, W. Wang, K. Li, W. Zhao, X. Dong, Efficient one-pot construction of unsymmetrical disulfide bonds with TCCA, Tetrahedron Lett. 58 (3) (2017) 218–222.
[12] N. Stellenboom, R. Hunter, M.R. Caira, L. Szilágyi, A high-yielding, one-pot preparation of unsymmetrical glycosyl disulfides using 1-chlorobenzotriazole as an in situ trapping/oxidizing agent, Tetrahedron Lett. 51 (2010) 5309–5312.
[13] N. Stellenboom, R. Hunter, M.R. Caira, One-pot synthesis of unsymmetrical disulfides using 1-chlorobenzotriazole as oxidant: interception of the sulfenyl chloride intermediate, Tetrahedron Lett. 51 (2010) 5309–5312.
[14] M. Bao, M. Shimizu, N-Trifluoroacetyl arenesulfenamides, effective precursors for synthesis of unsymmetrical disulfides and sulfonamides, Tetrahedron 59 (2003) 9655–9659.
[15] M. Koketsu, K. Tanaka, Y. Takenaka, C. Dkwong, H. Ishihara, Eur. J. Pharmaceut. Sci. 15 (2002) 307.
[16] B. Tozkoparan, G. Aktay, E. Yolanda, Farmaco 57 (2002) 145.
[17] J.G. Sheppard, Keely R. Frazier, P. Saralkar, M.F. Hossain, W.J. Geldenhuys, T.E. Long, Disulfiram-based disulfides as narrow-spectrum antibacterial agents, Bioorg. Med. Chem. Lett 28 (8) (2018) 1298–1302.
[18] S. André, Z.C. Pei, H.C. Siebert, O. Ramström, H. Gabius, J. Bioorg. Med. Chem. 14 (2006) 6314–6326.
[19] J.L. Venerstrom, M.T. Makler, C.K. Angerhofer, J.A. Williams, Antimicrob. Agents Chemother. 39 (1995) 2671.
[20] W. Malinka, M. Kaczmarz, B. Filipek, J. Sapa, D.B. Glo, Farmaco 57 (2002) 737.
[21] Vijay H. Masand, Vesna Rastija, Meghshyam K. Patil, Ajaykumar Gandhi, A. Chapolikar, Extending the identification of structural features responsible for anti-SARS-CoV activity of peptide-type compounds using QSAR modeling, SAR QSAR Environ. Res. 31 (9) (2020) 643–654.
[22] M. Vinicius, Alves tesia bobrowski cleber C. Melo-Filho daniel korn scott auerbach charles schmitt eugene N. Muratov alexander Tropsha, QSAR modeling of SARS-CoV mpro inhibitors identifies sufugolix, cenicriviroc, proglumetacin, and other drugs as candidates for repurposing against SARS-CoV-2, Molecular Informatics 40 (2021) 2000113.
[23] Vijay H. Masand, Siddhartha Akasapu, Ajaykumar Gandhi, Vesna Rastija, Meghshyam K. Patil, Structure features of peptide-type SARS-CoV main protease inhibitors: quantitative structure activity relationship study, Chemometr. Intell. Lab. Syst. 206 (2020) 104172.
[24] Samir Chtita, Majdouline Larif, Mounir Ghamali, Mohammed Bouachrine, Tahar Lakhlifi, Quantitative structure–activity relationship studies of dibenzo[a,d] cycloalkenimine derivatives for non-competitive antagonists of N-methyl-d-aspartate based on density functional theory with electronic and topological descriptors, Journal of Taibah University for Science 9 (2) (2015) 143–155.
[25] Samir Chtita, Rachid Hmamouchi, Majdouline Larif, Mounir Ghamali, Mohammed Bouachrine, Tahar Lakhlifi, QSPR studies of 9-aniliioacridine derivatives for their DNA drug binding properties based on density functional theory using statistical methods: model, validation and influencing factors, Journal of Taibah University for Science 10 (6) (2016) 868–876.
[26] Samir Chtita, Mounir Ghamali, Rachid Hmamouchi, Bouhya Elidrissi, Bourass Mohamed, Majdouline Larif, Mohammed Bouachrine, Tahar Lakhlifi, Investigation of antileishmanial activities of acridines derivatives against promastigotes and amastigotes form of parasites using quantitative structure activity relationship analysis, Advances in Physical Chemistry (2016) 1–16.
[27] Li Wang, Bo-Bo Bao, Guo-Qing Song, Cheng Chen, Xu-Meng Zhang, Wei Lu, Zefang Wang, Yan Cai, Sheng Fu Shuang, Fu-Hang Song, Haitao Yang, Jian-Guo Wang, Discovery of unsymmetrical aromatic disulfides as novel inhibitors of SARS-CoV main protease: chemical synthesis, biological evaluation, molecular docking and 3D-QSAR study, Eur. J. Med. Chem. 137 (2017) 450–461.
[28] Version 6 GaussView, Dennington Roy, Todd A. Keith, John M. Millam, Semichem Inc., Shawnee Mission, KS, 2016.
[29] Gaussian 09, Revision B.01, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian, Inc., Wallingford CT, 2010.
[30] ChemOffice, PerkinElmer Informatics. http://www.cambridgesoft.com, 2016.
[31] XLSTAT, Software, XLSTAT company. www.xlstat.com, 2013. (Accessed 17 April 2020). Accessed.
[32] S. Chtita, M. Ghamali, A. Ousaa, A. Aouidate, A. Belhassan, A. Idrissi Taourati, V.H. Masand, M. Bouachrine, T. Lakhlifi, QSAR study of anti-Human African Trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and

Lipinski's descriptors, Heliyon 5 (3) (2019), e01304, https://doi.org/10.1016/j.heliyon.2019.e01304.

[33] a Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276; b A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inf. 29 (6–7) (2010) 476–488.

[34] OECD Guidance Document on the Validation of QSAR Models, Organization for Economic Co-operation & Development, Paris, 2007.

[35] Mahmud Tareq Hassan Khan, Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches, Curr. Drug Metabol. 11 (4) (2010) 285–295, https://doi.org/10.2174/138920010791514306.

[36] D.E. Pires, T.L. Blundell, D.B. Ascher, pkCSM, Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures, J. Med. Chem. 58 (9) (2015) 4066–4072.

[37] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 46 (1997) 3–26.

[38] A. Aouidate, A. Ghaleb, M. Ghamali, et al., Investigation of indirubin derivatives: a combination of 3D-QSAR, molecular docking, and ADMET towards the design of new DRAK2 inhibitors, Struct. Chem. 29 (2018) 1609–1622, https://doi.org/10.1007/s11224-018-1134-0.

[39] T.J. Hou, K. Xia, W. Zhang, X.J. Xu, ADME evaluation in drug discovery. Prediction of aqueous solubility based on atom contribution approach, J. Chem. Inf. Comput. Sci. 44 (2004) 266–275.

[40] J.C. Dearden, M.T. Cronin, K.L. Kaiser, How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR), SAR QSAR, Environ. Res. 20 (2009) 241–266.

[41] J.F. Hair Jr., R.E. Anderson, R.L. Tatham, W.C. Black, Multivariate Data Analysis, third ed., Macmillan, New York, 1995.

[42] S. Chtita, Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR - Recherche de nouveaux medicaments, Faculty of Sciences Meknes - Moulay Ismaïl University, Morocco, 2017. Thesis.

[43] L.S. Higashi, M. Lundeen, K. Sef, Empirical relations between disulfide bond lengths, (N or C)-C-S-S torsion angles, and substituents in aromatic disulfides. Crystal and molecular structure of 3,3'-Dihydroxydi-2-pyridyl disulfide, American Chemical Society 100 (26) (1978) 8101–8106.

[44] S. Chtita, M. Ghamali, M. Larif, R. Hmamouchi, M. Bouachrine, T. Lakhlifi, Quantitative structure–activity relationship studies of anticancer activity for Isatin (1H-indole-2,3-dione) derivatives based on density functional theory with electronic and topological descriptors, Int. J. Quan. Struc. Prop. Rel. 2 (2) (2017) 90–115.

[45] S. Chtita, M. Ghamali, R. Hmamouchi, B. Elidrissi, M. Bourass, M. Larif, M. Bouachrine, T. Lakhlifi, Investigation of antileishmanial activities of acridines derivatives against promastigotes and amastigotes form of parasites using quantitative structure activity relationship analysis, Adv. Phys. Chem. (2016) 1–16.

[46] M.L. Amin, P-glycoprotein inhibition for optimal drug delivery, Drug Target Insights 7 (2013) 27–34.

[47] G.M. Levin, P-glycoprotein: why this drug transporter may be clinically important, Cur Psychiatry 11 (2012) 38–40.

[48] E. V.Pires Douglas, Tom L. Blundell, David B. Ascher, pkCSM: predicting small-molecule pharmacokinetic properties using graph-based signatures, J. Med. Chem. 58 (9) (2015 May 14) 4066–4072, https://doi.org/10.1021/acs.jmedchem.5b00104.

[49] A.H. Ahmed, Y.I. Alkali, In silico pharmacokinetics and molecular docking studies of lead compounds derived from Diospyros mespiliformis, Pharma 7 (2019) 31–37.