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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
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 Aoudate e, Salah Belaaid f,g, Mohammed Moutaabbid a, Said Belaaoud a, Mohammed Bouachrine b,h, Tahar Lakhli b

a Laboratory of Physical Chemistry of Materials, Faculty of Sciences Ben M’Sil, 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, Chouaib 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, Algeria
h High School of Technology of Khenifra, Sultane Slimane University, B.P. 591, Khenifra, Morocco

ARTICLE INFO

Keywords:
Quantitative structure activity relationship (QSAR)
Density functional theory (DFT)
Coronavirus
SARS-CoV
Disulfide
ADMET

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 IC50 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 R2, R2test and Q2cv (R2 = 0.838, R2test = 0.735, Q2cv = 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 (Ehomo), energy of molecular orbital below HOMO energy (EHOMO-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. Branoswka 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 thiacarboxylic 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 di-sulfide 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-Tri-fluoroacetyl aminothiol derivatives 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, ajone and dysosoxysulfone 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 activity of some known drugs to inhibit SARS-CoV-2 [21]. Other studies were reported by Masand et al. 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 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^2]_{adj}), Mean of Squared Errors of model (MSE), Fischer’s value (F_{cal}), Variance Inflation Factor (VIF), Coefficient of determination of Leave-One-Out Cross Validation ([R^2]_{CV}), Coefficient of determination of external test ([R^2]_{ext}) and Y-randomization parameters ([R^2]_{rand} and Q^2_CV (Rand)) [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.

| No | Structure | IC50 | No | Structure | IC50 | No | Structure | IC50 |
|----|-----------|------|----|-----------|------|----|-----------|------|
| 1  | ![Structure 1](image1.png) | 1.871 | 8  | ![Structure 8](image2.png) | 1.762 | 13 | ![Structure 13](image3.png) | 1.651 |
| 2  | ![Structure 2](image4.png) | 2.863 | 14 | ![Structure 14](image5.png) | 2.075 | 15 | ![Structure 15](image6.png) | 5.954 |
| 3  | ![Structure 3](image7.png) | 3.675 | 16 | ![Structure 16](image8.png) | 3.957 | 17 | ![Structure 17](image9.png) | 4.126 |
| 4  | ![Structure 4](image10.png) | 3.130 | 18 | ![Structure 18](image11.png) | 2.565 | 19 | ![Structure 19](image12.png) | 1.947 |
| 5  | ![Structure 5](image13.png) | 1.506 | 20 | ![Structure 20](image14.png) | 2.029 | 21 | ![Structure 21](image15.png) | 1.250 |
| 6  | ![Structure 6](image16.png) | 4.344 | 22 | ![Structure 22](image17.png) | 2.211 | 23 | ![Structure 23](image18.png) | 3.321 |
| 7  | ![Structure 7](image19.png) | 1.871 | 24 | ![Structure 24](image20.png) | 2.555 | 25 | ![Structure 25](image21.png) | 2.452 |
|    |           |      | 26 | ![Structure 26](image22.png) | 1.679 | 27 | ![Structure 27](image23.png) | 1.557 |
|    |           |      |    |           |      |    |           |      |
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 ≤ 500 g/mol, Log P ≤ 5, NR ≤ 10, HBA ≤ 10, HBD ≤ 5, PSA ≤ 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 ≤ −6, moderately soluble if −6 < log S ≤ −4, soluble if −4 < log S) [39].

Finally, different ADME 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, hERG1 and hERG2 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 other descriptors (Table S3).

3.2. Principal component analysis (PCA)

The 45 descriptors are computed 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 ≤ 0.15) with respect to the dependent variable (IC50). 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 dissides. 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 IC50 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 R2, R2_adjust, Q2, and R2_pred and low values of MSE point out that all these models are statistically satisfactory.

| Model equations | R2 | R2_adjust | MSE | R2_pred | Q2 |
|-----------------|----|-----------|-----|---------|----|
| IC50 = -85.468 + 1.0164 + 36.289 S1S2 + 1.081 log S - 0.860 HLC + 0.042 BP | 0.801 | 0.763 | 0.418 | 0.655 | 0.722 |
| IC50 = -98.914 - 1.813 F_EHOMO - 3.652 F_EXOMO + 44.737 S1S2 - 100.408 S2Bnz + 5.382 10^{-16} BI | 0.789 | 0.749 | 0.562 | 0.907 | 0.675 |
| IC50 = -87.944 + 2.948 E_REMDA + 34.295 S1S2 - 76.116 S2Bnz + 5.487 10^{-16} BI - 0.60600C | 0.763 | 0.718 | 0.632 | 0.907 | 0.641 |
| IC50 = -85.852 - 1.272 F_EHOMO + 3.355 F_EXOMO + 40.557 S1S2 - 87.281 S2Bnz + 5.560 10^{-16} BI | 0.789 | 0.749 | 0.566 | 0.617 | 0.639 |
| IC50 = -63.514 + 1.828 E_REMDA + 0.927 E_REMDA + 42.783 S1S2 - 77.343 S2Bnz + 5.680 10^{-16} BI | 0.752 | 0.704 | 0.522 | 0.617 | 0.602 |
| IC50 = -0.265 - 0.616 E_REMDA + 1.906 E_REMDA + 0.852 log P + 3.537 10^{-16} BI + 0.1490% | 0.747 | 0.698 | 0.664 | 0.862 | 0.580 |
| IC50 = -72.252 - 1.866 F_EHOMO + 3.590 F_EXOMO + 44.919 S1S2 - 85.554 S2Bnz + 5.202 10^{-16} BI | 0.821 | 0.787 | 0.488 | 0.655 | 0.722 |
| IC50 = -199.399 - 1.573 F_EHOMO + 3.848 F_EXOMO + 44.839 S1S2 - 110.400 S2Bnz + 5.872 10^{-16} BI | 0.742 | 0.692 | 0.572 | 0.694 | 0.605 |
| IC50 = -167.793 + 1.726 E_REMDA + 58.021 S1S2 + 26.914 S2H1Br + 5.454 10^{-16} BI - 0.1290 HLC | 0.852 | 0.824 | 0.336 | 0.637 | 0.796 |
| IC50 = -174.066 + 3.601 E_REMDA + 1.977 E_REMDA + 95.138 S1S2 - 0.330 0% + 0.1960 PSA | 0.772 | 0.728 | 0.453 | 0.735 | 0.640 |
| IC50 = -105.697 + 2.507 E_REMDA + 46.181 S1S2 + 349.936 S2Bnz + 0.039 GFE - 0.8854% | 0.741 | 0.691 | 0.545 | 0.862 | 0.552 |
| IC50 = -126.683 + 3.235 E_REMDA + 67.333 S1S2 + 0.656 NHBA + 0.011 BP - 0.1760% | 0.743 | 0.694 | 0.508 | 0.953 | 0.584 |
| IC50 = -120.893 - 2.510 E_REMDA + 4.209 E_REMDA + 45.514 S1S2 - 113.825 S2Bnz + 5.821 10^{-16} BI | 0.768 | 0.723 | 0.556 | 0.776 | 0.638 |
| IC50 = -97.222 - 0.800 E_REMDA + 3.663 E_REMDA + 42.570 S1S2 - 95.527 S2Bnz + 4.692 10^{-16} BI | 0.800 | 0.761 | 0.545 | 0.655 | 0.680 |
| IC50 = -64.336 + 2.234 E_REMDA + 0.574 E_REMDA + 39.641 S1S2 - 72.988 S2Bnz + 5.597 10^{-16} BI | 0.800 | 0.761 | 0.526 | 0.694 | 0.666 |
| IC50 = -74.931 + 2.538 E_REMDA + 37.952 S1S2 - 76.311 S2Bnz + 0.634 NHBD - 5.498 10^{-16} BI | 0.800 | 0.761 | 0.526 | 0.694 | 0.666 |
| IC50 = -83.018 + 2.112 E_REMDA + 42.762 S1S2 - 87.785 S2Bnz + 6.665 10^{-16} BI - 0.0130 PSA | 0.762 | 0.716 | 0.505 | 0.735 | 0.582 |
| IC50 = -56.425 + 2.848 E_REMDA + 41.4501 S1S2 - 67.856 S2Bnz + 4.983 10^{-16} BI - 0.0390% | 0.750 | 0.702 | 0.495 | 0.655 | 0.593 |
| IC50 = -57.755 + 0.708 + 0.507 + 0.776 + 0.566 | (continued on next page) |
IC₅₀ = 98.414 - 1.539 \text{E}_{\text{HOMO}} - 1 + 3.391 \text{E}_{\text{LUMO}} + 40.628 S1S2 - 95.249 S2Bnz + 5.368 10^{-6} \text{BI}

20 \text{ IC₅₀ = 106.474 - 1.800 E}_{\text{HOMO}} - 1 + 3.416 \text{E}_{\text{LUMO}} + 40.309 S1S2 - 100.234 S2Bnz + 5.211 10^{-6} \text{BI}

21 \text{ IC₅₀ = 30.182 + 3.025 E}_{\text{HOMO}} + 52.323 S1S2 - 66.486 S2Bnz + 0.040 GFE - 9.3964 \text{BI}

22 \text{ IC₅₀ = 126.976 - 2.222 E}_{\text{HOMO}} + 3.939 \text{E}_{\text{LUMO}} + 49.656 S1S2 - 122.318 S2Bnz + 4.955 10^{-6} \text{BI}

23 \text{ IC₅₀ = 48.294 + 1.842 E}_{\text{HOMO}} + 0.825 \text{E}_{\text{LUMO}} + 43.447 S1S2 - 69.770 S2Bnz + 5.907 10^{-6} \text{BI}

24 \text{ IC₅₀ = 80.883 + 2.481 E}_{\text{HOMO}} + 28.980 S1S2 - 72.516 S2Bnz + 5.432 10^{-6} \text{BI} - 0.0456 \text{Bi}

25 \text{ IC₅₀ = 97.940 - 1.322 E}_{\text{HOMO}} - 1 + 3.469 \text{E}_{\text{LUMO}} + 44.105 S1S2 - 97.956 S2Bnz + 5.330 10^{-6} \text{BI}

26 \text{ IC₅₀ = -177.107 + 3.565 E}_{\text{HOMO}} + 2.450 \text{E}_{\text{LUMO}} + 96.952 S1S2 - 0.271 094 + 0.090 \text{PSA}

27 \text{ IC₅₀ = 57.171 + 1.561 E}_{\text{HOMO}} + 40.099 S1S2 - 72.972 S2Bnz + 0.016 GFE + 5.090 10^{-6} \text{Bi}

28 \text{ IC₅₀ = 115.209 - 1.762 E}_{\text{HOMO}} + 3.428 \text{E}_{\text{LUMO}} + 43.393 S1S2 - 108.561 S2Bnz + 5.575 10^{-6} \text{Bi}

29 \text{ IC₅₀ = 30.718 + 2.616 E}_{\text{HOMO}} + 43.359 S1S2 - 57.653 S2Bnz + 0.589 NHBD + 4.666 10^{-6} \text{Bi}

30 \text{ IC₅₀ = 111.463 - 1.448 E}_{\text{HOMO}} + 3.393 \text{E}_{\text{LUMO}} + 45.060 S1S2 - 107.296 S2Bnz + 5.226 10^{-6} \text{Bi}

31 \text{ IC₅₀ = 103.548 - 2.125 E}_{\text{HOMO}} + 4.035 \text{E}_{\text{LUMO}} + 49.253 S1S2 - 108.085 S2Bnz + 4.767 10^{-6} \text{Bi}

32 \text{ IC₅₀ = 109.178 - 2.098 E}_{\text{HOMO}} + 3.915 \text{E}_{\text{LUMO}} + 46.178 S1S2 - 107.953 S2Bnz + 5.459 10^{-6} \text{Bi}

33 \text{ IC₅₀ = 121.803 - 1.602 E}_{\text{HOMO}} + 3.553 \text{E}_{\text{LUMO}} + 48.317 S1S2 - 117.039 S2Bnz + 6.338 10^{-6} \text{Bi}

34 \text{ IC₅₀ = 128.780 - 2.590 E}_{\text{HOMO}} + 4.855 \text{E}_{\text{LUMO}} + 51.701 S1S2 - 123.760 S2Bnz + 5.682 10^{-6} \text{Bi}

35 \text{ IC₅₀ = 220.048 - 0.734 E}_{\text{HOMO}} + 0.938 \text{E}_{\text{LUMO}} - 119.872 S2Bnz - 0.912 NHBD + 0.930 NR

36 \text{ IC₅₀ = 114.995 - 2.281 E}_{\text{HOMO}} + 4.025 \text{E}_{\text{LUMO}} + 41.273 S1S2 - 105.794 S2Bnz + 5.359 10^{-6} \text{Bi}

37 \text{ IC₅₀ = 6.687 - 0.936 \text{E}_{\text{HOMO}} + 1.909 \text{E}_{\text{LUMO}} + 29.959 S1S2 - 40.520 S2Bnz + 7.103 10^{-6} \text{BI}

Robust and also possess good external predictive ability.

For all developed models, values of R^2 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 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 ≥ 0.800 and R^2 adj ≥ 0.760). However, looking at the others statistical parameters (MSE, R^2_in and Q^2_in) we can suggest models 28, 30 and 34 as the most desirable three models. The three pentavariable MLR equations are able to predict IC50 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 $p$IC$_{50}$ data must be considered reliable only for those fall within this AD on which the model was constructed. Compounds with $h^* > h^*$ or with standardized residual greater than $y = \pm 3$ 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 $\pm 3$ 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](http://teqip.jdvu.ac.in/QSAR_Tools), were employed in the y-randomization.

The average value of $R^{2}_{\text{Rand}}$, $R^{2}_{\text{Rand}}$, and $Q^{2}_{\text{cv}}$($R^{2}_{\text{Rand}}$) are 0.413, 0.183 and $-0.272$ respectively, the $R^{2}_{\text{cv}}$ value equals a 0.847, and all the new QSAR models having significantly low $R^{2}_{\text{Rand}}$ and $Q^{2}_{\text{cv}}$($R^{2}_{\text{Rand}}$) 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 (EHOMO, E$\text{HOMO}$-1, BI, S$_2$Bnz and S$_1$S$_2$) 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 $R^{2}$ ($R^{2}_{\text{adj}}$) 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 ($Q^{2}_{\text{cv}}$) by LOO approach was 0.757 which showed a good internal predictive ability of the model. The low $R^{2}$ and $Q^{2}_{\text{cv}}$ 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 $R^{2}_{\text{cv}}$ as shown in the developed model ($R^{2}_{\text{cv}} = 0.735$) 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$_{\text{HOMO}}$, E$_{\text{HOMO}-1}$, BI, S$_2$Bnz and S$_1$S$_2$, 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$_{\text{HOMO}}$) and bond length between the two sulfur atoms (S$_1$S$_2$) promote activity, while molecular orbital energy below HOMO energy (E$_{\text{HOMO}-1}$) and bond length between sulfur atom and the benzene ring (S$_2$Bnz) 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, −2.955, 8.162, −5.080 and 5.425 for E$_{\text{HOMO}}$, E$_{\text{HOMO}-1}$, BI, S$_2$Bnz and S$_1$S$_2$, 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,
and by electronic characteristics (29% of the variance in IC50) related
with the EHOMO and EHOMO-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 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 IC50, which suggests that increased activity
(decreased IC50) can be achieved by decreasing the rami-

- 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 IC50, suggesting that increased activity can be achieved by sub-
stitute 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 sub-
stituents 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 IC50, 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

| N1 | Observed IC50 | Predicted IC50 | Error | N2 | Observed IC50 | Predicted IC50 | 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         | −0.976|
| 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.)
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 $E_{\text{HOMO}}$ in describing anti-SARS-CoV activity may be attributed to the interaction of disulfide derivatives with nucleophilic amino acid residue of microorganisms. $E_{\text{HOMO}}$ has a positive sign in the model and variation in BI accounts for 19% of the variance in IC50, which suggests that the higher of $E_{\text{HOMO}}$, 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 IC50, we will decrease $E_{\text{HOMO}}$ for which we must substitute the disulfide derivatives for a weaker donating electron ability group that removes electron density (don’t donate density) from the conjugated $\pi$ system via mesomerism effect, making it less reactive.

- $E_{\text{HOMO}}$ 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 IC50.

In the conclusion, these results illustrate that to increase the anti-SARS-CoV, decrease IC50, we will substitute the disulfide derivatives with smaller size electron withdrawing groups such as Nitro (NO2), Sulfinic acid (SOH), Cyano (CN), Trifluoromethyl (CF3), Haloformal (COX), Carbonyl (CO$\equiv$), alkoxycarbonyl (CO$\equiv$R), Acril (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 IC50 based on the built model and was validated by the proposed model equation:

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

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 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 (IC50 = 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 IC50 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 $\leq 10$, octanol-water partition coefficient $< 5$, hydrogen bond donor $< 5$, molecular weight $< 500$Da 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 4
Comparison of the statistical parameters of model 34 and Golbraikh and Tropsha’s criteria.

| Parameter | Equation | Model score | Threshold | Comment |
|-----------|----------|-------------|-----------|---------|
| $R^2$     | $R^2 = 1 - \frac{\sum(Y_{\text{obs}} - Y_{\text{calc}})^2}{\sum(Y_{\text{obs}} - \bar{Y}_{\text{obs}})^2}$ | 0.832 | > 0.600 | Passed |
| $R^2_{\text{adj}}$ | $R^2_{\text{adj}} = (N - 1)R^2 - p$ | 0.802 | > 0.600 | Passed |
| $R^2_{\text{certain}}$ | $R^2_{\text{certain}} = 1 - \frac{\sum(Y_{\text{calc}}(\text{test}) - Y_{\text{calc}}(\text{train}))^2}{\sum(Y_{\text{obs}} - \bar{Y}_{\text{obs}})^2}$ | 0.737 | > 0.600 | Passed |
| $Q^2$ | $Q^2 = 1 - \frac{\sum(Y_{\text{calc}} - Y_{\text{obs}})^2}{\sum(Y_{\text{obs}} - \bar{Y}_{\text{obs}})^2}$ | 0.740 | > 0.500 | Passed |
| MSE | MSE = $\frac{\sum(Y_{\text{calc}} - Y_{\text{obs}})^2}{N}$ | 0.483 | A low value | Passed |
| $R^2_{\text{pred}}$ | $R^2_{\text{pred}} = \frac{\sum(Y_{\text{calc}} - Y_{\text{obs}})^2}{N - p - 1}$ | 27.654 | a high value | Passed |
| $Q^2_{\text{LOO}}(\text{Pred})$ | $Q^2_{\text{LOO}}(\text{Pred}) = 1 - \frac{\sum(Y_{\text{calc}} - Y_{\text{obs}})^2}{\sum(Y_{\text{obs}} - \bar{Y}_{\text{obs}})^2}$ | 0.142 | $< R^2$ | Passed |
| $Q^2_{\text{LOO}}(\text{Pred})$ | $Q^2_{\text{LOO}}(\text{Pred}) = 1 - \frac{\sum(Y_{\text{calc}} - Y_{\text{obs}})^2}{\sum(Y_{\text{obs}} - \bar{Y}_{\text{obs}})^2}$ | -0.270 | $< Q^2$ | Passed |
| $cR^2$ | $cR^2 = R^2 - \sqrt{R^2 - (\text{Average } R^2)_{\text{CV}}^2}$ | 0.764 | > 0.500 | Passed |

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

$\bar{Y}_{\text{obs}}$ and $\bar{Y}_{\text{calc}}$, 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.
3.8. ADMET properties

Absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of designed sulfi de 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 5

### Table 5

| Compound | Lipinski’s parameters | Descriptors | calculated anti-SARS-CoV activity and leverage | h |
|----------|-----------------------|-------------|---------------------------------------------|----|
| Xi       | MW: 289.725           | 4           | 7                                           | 108.293 | 0 | Moderately |
| X2       | MW: 269.738           | 3           | 6                                           | 104.398 | 3 | Moderately |
| X4       | MW: 272.738           | 4           | 6                                           | 104.166 | 0 | Moderately |
| X5       | MW: 272.738           | 3           | 6                                           | 104.166 | 0 | Moderately |
| X7       | MW: 262.718           | 3           | 5                                           | 97.806   | 0 | Soluble |
| X8       | MW: 279.173           | 4           | 5                                           | 103.943  | 0 | Moderately |
| X9       | MW: 323.624           | 4           | 5                                           | 107.508  | 0 | Soluble |
| X10      | MW: 280.708           | 3           | 5                                           | 101.971  | 0 | Soluble |
| Y1       | MW: 279.777           | 3           | 5                                           | 111.635  | 0 | Moderately |
| Y2       | MW: 299.764           | 4           | 6                                           | 115.530  | 0 | Moderately |
| Y3       | MW: 298.776           | 3           | 5                                           | 116.198  | 0 | Moderately |
| Y4       | MW: 272.757           | 4           | 5                                           | 105.043  | 0 | Soluble |
| Y5       | MW: 289.212           | 3           | 4                                           | 111.181  | 0 | Moderately |
| Y6       | MW: 333.663           | 4           | 5                                           | 114.745  | 0 | Moderately |
| Y7       | MW: 317.222           | 4           | 5                                           | 117.769  | 0 | Moderately |
| Y8       | MW: 296.804           | 4           | 5                                           | 117.769  | 0 | Moderately |
| Y9       | MW: 312.803           | 4           | 6                                           | 122.882  | 0 | Moderately |
| Y10      | MW: 296.804           | 4           | 5                                           | 117.769  | 0 | Moderately |

### Table 6

| Compound | Lipinski’s parameters | Descriptors | calculated anti-SARS-CoV activity and leverage | h |
|----------|-----------------------|-------------|---------------------------------------------|----|
| X1       | MW: 289.725           | 4           | 7                                           | 108.293 | 0 | Moderately |
| X2       | MW: 269.738           | 3           | 6                                           | 104.398 | 3 | Moderately |
| X4       | MW: 272.738           | 4           | 6                                           | 104.166 | 0 | Moderately |
| X5       | MW: 272.738           | 3           | 6                                           | 104.166 | 0 | Moderately |
| X7       | MW: 262.718           | 3           | 5                                           | 97.806   | 0 | Soluble |
| X8       | MW: 279.173           | 4           | 5                                           | 103.943  | 0 | Moderately |
| X9       | MW: 323.624           | 4           | 5                                           | 107.508  | 0 | Soluble |
| X10      | MW: 280.708           | 3           | 5                                           | 101.971  | 0 | Soluble |
| Y1       | MW: 279.777           | 3           | 5                                           | 111.635  | 0 | Moderately |
| Y2       | MW: 299.764           | 4           | 6                                           | 115.530  | 0 | Moderately |
| Y3       | MW: 298.776           | 3           | 5                                           | 116.198  | 0 | Moderately |
| Y4       | MW: 272.757           | 4           | 5                                           | 105.043  | 0 | Soluble |
| Y5       | MW: 289.212           | 3           | 4                                           | 111.181  | 0 | Moderately |
| Y6       | MW: 333.663           | 4           | 5                                           | 114.745  | 0 | Moderately |
| Y7       | MW: 317.222           | 4           | 5                                           | 117.769  | 0 | Moderately |
| Y8       | MW: 296.804           | 4           | 5                                           | 117.769  | 0 | Moderately |
| Y9       | MW: 312.803           | 4           | 6                                           | 122.882  | 0 | Moderately |
| Y10      | MW: 296.804           | 4           | 5                                           | 117.769  | 0 | Moderately |

### 3.8. ADMET properties

Absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of designed sulfi de derivatives were predicted using pkCSM (Table 7).
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 further 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 uptake of drugs resulting in therapeutic failure because the drug concentration would be lower than expected [46,48].

The skin permeability, expressed as the skin permeability constant log (Kp), (A compound is considered to have relatively low skin permeability if it has log Kp(cm²/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 IC50) related, firstly, with the size and volume of the substituents 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 IC50) 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 should 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 perspectve oral drugs.

Dehydrogenative coupling of thiols, Organic Chemistry Frontiers 6 (2019) 2222–2225.

Dehydration of allylic alcohols via silylation, Tetrahedron Lett. 51 (2010) 3369–3372.

Dehydrogenative coupling of thiols, Organic Chemistry Frontiers 6 (2019) 2222–2225.

Dehydration of allylic alcohols via silylation, Tetrahedron Lett. 51 (2010) 3369–3372.
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) 476–486.

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

[37] Mahmud Tareq Hassan Khan, Predictions of the ADME 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] S. Chtita, Modélisation de moléules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR - Recherche de nouveaux médicaments, Faculty of Sciences Meknes - Moulay Ismail 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, R. Hnamouchi, M. Bouachrine, T. Lakkhlifi, 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. Struct. Prop. Rel. 2 (2) (2017) 90–115.

[45] S. Chtita, M. Ghamali, R. Hnamouchi, B. Elidrisi, M. Bourass, M. Lari, M. Bouachrine, T. Lakkhlifi, 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.