TÜBİTAK

Turkish Journal of Chemistry

http://journals.tubitak.gov.tr/chem/

Research Article

Turk J Chem
(2019) 43: 251 – 265
© TÜBİTAK
doi:10.3906/kim-1807-61

# Investigation of the structural and physicochemical requirements of quinoline-arylamidine hybrids for the growth inhibition of K562 and Raji leukemia cells

**Vesna RASTIJA[1,*], Marijana JUKIĆ[2], Teuta OPAČAK-BERNARDI[2], Luka KRSTULOVIĆ[3], Ivana STOLIĆ[3], Ljubica GLAVAŠ-OBROVAC[2], Miroslav BAJIĆ[3]**

[1]Department of Agroecology and Environmental Protection, Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

[2]Department of Medicinal Chemistry, Biochemistry, and Laboratory Medicine, Faculty of Medicine, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

[3]Department of Chemistry and Biochemistry, Faculty of Veterinary Medicine, University of Zagreb, Zagreb, Croatia

**Received:** 14.07.2018 • **Accepted/Published Online:** 26.11.2018 • **Final Version:** 05.02.2019

**Abstract:** Quantitative structure–activity relationship (QSAR) analysis of 28 quinoline-arylamidine (CQArA) hybrids against two leukemia cells, K562 and Raji, was performed. Multiple linear regression (MLR) models were obtained by genetic algorithm. The best models involved the following descriptors: radial distribution function (RDF) descriptors, GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptor, bond information content index, and dipole moment. The best MLR models for K562 and Raji cells demonstrated satisfactory stability in internal and external validation. Since the QSAR model for Raji cells has better predictive ability, two new highly potent CQArA analogues were proposed based on it. The QSAR models revealed important physicochemical and structural requirements for the antitumor activity: enhanced 3D molecular distribution of mass calculated at radius 11 Å from the center of molecule, a higher number of terminal electronegative atoms, extension of the molecules' central linker between quinoline and arylamidine, higher ratio of single bonds and total number of atoms, and symmetric charge distribution. Molecular docking study was applied to ensure the anticancer activity affinity to the binding site of the tyrosine-protein kinase (c-SRC). It was confirmed that the most active compound binds on the pocket between the small and large lobes of c-SRC, mostly throughout the hydrogen bonds and van der Waals interactions.

**Key words:** Structure–activity, molecular docking, hybrid anticancer drugs, leukemia

## 1. Introduction

A major problem in chemotherapeutic treatment of cancer is drug resistance. In order to overcome this problem, multiple cytotoxic drugs with different chemical structures and mechanisms of action have been employed.[1] Hybrid anticancer drugs are designed and synthesized from two or more different bioactive moieties conjugated into a single drug. Such multitarget anticancer agents have great therapeutic potential since they trigger two or more cytocidal pharmacological mechanisms of action, acting in synergy to inhibit the growth of cancer cells.[2,3] Considering their contribution in the discovery of a new anticancer therapy, the research on hybrid molecules is rapidly evolving.

Chloroquine (CQ), a 7-chloro-4-aminoquinoline, has been the mainstay of malaria chemotherapy for years.

*Correspondence: vrastija@fazos.hr

A number of other existing antimalarial agents, as well as novel derivatives of CQ, have been tested as possible cancer therapeutics.[4–9] Hydroxychloroquine (HCQ), a CQ derivative, has recently been investigated in clinical trials for cancer therapy.[10] Arylamidines exhibited broad pharmacological activities, such as antiparasitic,[11] antibacterial,[12,13] and anticancer activity.[14]

Recently, we synthesized and evaluated the antitumor activity of a new class of hybrid molecules joining these two pharmacophores, 7-chloroquinoline and arylamidine.[15] The four series of molecular hybrids contain a quinoline moiety, whereas the linkers between quinoline and arylamidine, the aryl core, and terminal amidine moiety were modified. The investigated compounds showed higher efficacy against a human chronic myelogenous leukemia cell line (K562) and lymphoblastoid cell line (Raji) than carcinoma cells. Recent advances have implicated the role of tyrosine kinases in the pathophysiology of cancer. Their activity may acquire transforming functions due to mutation(s), overexpression, and autocrine paracrine stimulation, leading to malignancy. Selective tyrosine kinase inhibitors can block the oncogenic activation in cancer cells and can be applied as a new mode of cancer therapy.[16] Nonreceptor protein-tyrosine kinase (c-Src) is a protooncogene that plays key roles in cell morphology, motility, proliferation, and survival.[17] Numerous studies have shown evidence of an association between c-Src kinases and leukemia, as Li[18] summarized previously. Active site residues of Src kinase are derived from both the small and large lobes of the kinase, which are moving relative to each other and open or close the cleft. Any substances that block the interconversion of the open and closed forms of the cleft could be inhibitors.[17] Imatinib, sold under the brand name Gleevec (Novartis STI571), is used for chronic myelogenous leukemia treatment. Series of inhibitors (DSA compounds), which are based on the central chemical scaffold of imatinib, have been designed to bind to c-Src with equally high potency in the Asp-Phe-Gly (DFG)-flipped conformation.[19]

In order to signify the importance of structural and chemical attributes for the anticancer activity of 7-chloroquinoline-arylamidine (CQArA) hybrids and to improve further research on the mechanisms of action of these compounds against leukemia cells and also to propose new analogues with improved activity, quantitative structure–activity relationship (QSAR) analysis was performed. The interaction of CQArA hybrids and protooncogene tyrosine-protein kinase Src in silico was evaluated by molecular docking based on the binding mode of the DSA inhibitor.

## 2. Results and discussion

### 2.1. QSAR models

The best model obtained for anticancer activity against the K562 leukemia cell line is:

$$\log GI_{50} = 2.48 - 0.17RDF120m + 0.08RDF150p - 9.32R2e^{+} \quad (1)$$

Compounds in the test set, **9**, **13**, **18**, **20**, and **25**, were selected by ranking method.

For antitumor activity against the Raji leukemia cell line, the best model is:

$$\log GI_{50} = 5.69 - 6.94BIC1 - 0.11RDF110m - 0.2DIPX \quad (2)$$

Compounds in the test set, **1**, **2**, **5**, **11**, and **20**, were selected by random method.

Structures of analyzed compounds are presented in the Scheme. Experimental and calculated $\log GI_{50}$ by Eqs. (1) and (2) are shown in Table 1. The values of descriptors included in the models of Eqs. (1) and (2) are

**Table 1**. Experimentally obtained antitumor activities ($\log GI_{50exp}$)[15] and calculated values ($\log GI_{50calc}$) by QSAR models (1) and (2) for 28 CQArA hybrids against leukemia cell lines K562 and Raji, respectively.

| | K562 | | Raji | |
|---|---|---|---|---|
| Compound | $\log GI_{50exp}$ | $\log GI_{50calc}$ | $\log GI_{50\,exp}$ | $\log GI_{50calc}$ |
| **1** | 1.94 | 1.89 | 1.96 | 2.07 |
| **2** | 1.51 | 1.71 | 1.72 | 1.86 |
| **3** | 1.92 | 1.86 | 1.71 | 1.88 |
| **4** | 1.23 | 1.67 | 1.41 | 1.55 |
| **5** | 1.27 | 1.19 | 1.29 | 1.24 |
| **6** | 1.69 | 1.81 | 1.61 | 1.83 |
| **7** | 1.79 | 1.50 | 1.71 | 1.79 |
| **8** | 2.00 | 1.64 | 2.00 | 1.68 |
| **9** | 2.00 | 1.77 | 2.05 | 2.06 |
| **10** | 1.84 | 1.86 | 1.68 | 1.59 |
| **11** | 1.78 | 1.96 | 1.67 | 1.68 |
| **12** | 1.74 | 1.75 | 1.62 | 1.70 |
| **13** | 1.44 | 1.61 | 1.61 | 1.69 |
| **14** | 1.33 | 1.27 | 1.70 | 1.51 |
| **15** | 0.86 | 1.19 | 0.79 | 0.99 |
| **16** | 1.69 | 1.48 | 1.77 | 1.59 |
| **17** | 2.00 | 1.78 | 2.00 | 1.93 |
| **18** | 1.89 | 1.83 | 1.94 | 1.80 |
| **19** | 1.89 | 1.73 | 1.99 | 1.73 |
| **20** | 1.72 | 1.60 | 1.71 | 1.68 |
| **21** | 2.00 | 2.04 | 2.00 | 1.86 |
| **22** | 0.89 | 1.13 | 0.72 | 0.57 |
| **23** | 1.54 | 1.27 | 1.47 | 1.33 |
| **24** | 1.24 | 1.03 | 1.29 | 1.42 |
| **25** | 1.18 | 1.30 | 1.45 | 1.50 |
| **26** | 1.06 | 1.21 | 1.23 | 1.40 |
| **27** | 0.93 | 0.96 | 1.44 | 1.50 |
| **28** | 1.12 | 1.31 | 1.28 | 1.59 |

given in the Supplemental Information, Table S1. The statistical results for the obtained models are presented in Table 2.

Both models satisfy fitting criteria ($R^2 \geq 0.60$). In order to exclude the possibility that the models are overfitted, the collinearity of descriptors included in the models of Eqs. (1) and (2) was evaluated by correlation matrix (Tables 3 and 4). Low collinearity was confirmed by the values of correlation coefficient, $R \leq 0.7$, and verified by the low values of $Kxx$ and $\Delta K \geq 0.05$ (Table 2).

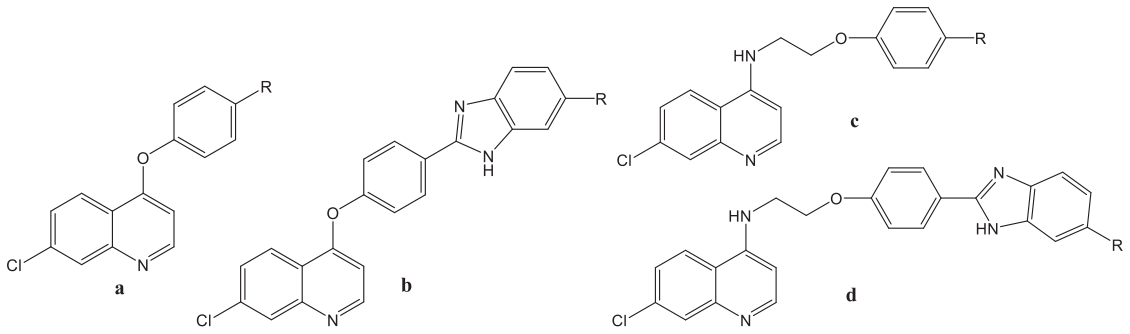
The models also demonstrate satisfactory stability in internal and external validation: $R^2_{ext} \geq 0.60$; $RMSE$ and $MAE$ close to zero; and $RMSE_{tr} < RMSE_{cv}$ and small difference between $CCC_{tr}$ and $CCC_{ext}$.[20]

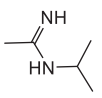
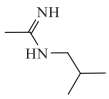
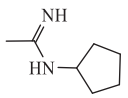
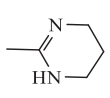
**Table 2**. The statistical results of QSAR models (1) and (2).

| | Model (1) | Model (2) |
|---|---|---|
| $N_{tr}$ | 23 | 23 |
| $N_{ex}$ | 5 | 5 |
| $R^2$ | 0.71 | 0.78 |
| $R^2_{adj}$ | 0.66 | 0.75 |
| $s$ | 0.23 | 0.18 |
| $F$ | 15.35 | 22.46 |
| $Kxx$ | 0.30 | 0.29 |
| $\Delta K$ | 0.06 | 0.05 |
| $RMSE_{tr}$ | 0.21 | 0.17 |
| $MAE_{tr}$ | 0.17 | 0.15 |
| $CCC_{tr}$ | 0.83 | 0.88 |
| $Q^2_{LOO}$ | 0.58 | 0.62 |
| $RMSE_{cv}$ | 0.25 | 0.22 |
| $MAE_{cv}$ | 0.21 | 0.19 |
| $PRESS_{cv}$ | 1.40 | 1.09 |
| $CCC_{cv}$ | 0.75 | 0.80 |
| $R^2_Y scr$ | –0.14 | 0.13 |
| $Q^2_Y scr$ | –0.27 | –0.30 |
| $RMSE_{ext}$ | 0.15 | 0.08 |
| $MAE_{ext}$ | 0.14 | 0.07 |
| $PRESS_{ext}$ | 0.12 | 0.03 |
| $R^2_{ext}$ | 0.84 | 0.96 |
| $Q^2_{F1}$ | 0.78 | 0.87 |
| $Q^2_{F2}$ | 0.74 | 0.85 |
| $Q^2_{F3}$ | 0.84 | 0.95 |
| $CCC_{ext}$ | 0.81 | 0.94 |
| $\bar{r}^2_m$ | 0.43 | 0.75 |
| $r^2_m$ | 0.32 | 0.08 |
| Applicability domain | | |
| $N$ compounds outlier | - | - |
| $N$ compounds out of app. dom. | - | 1 (**22**) |

**Table 3**. Correlation matrix for the descriptors included in model (1).

| | *RDF120m* | *RDF150p* | *R2e*$^+$ |
|---|---|---|---|
| *RDF120m* | 1 | | |
| *RDF150p* | 0.53 | 1 | |
| *R2e*$^+$ | –0.18 | 0.05 | 1 |




| R | —CN |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| a | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| b | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
| c | **15** | **16** | **17** | **18** | **19** | **20** | **21** |
| d | **22** | **23** | **24** | **25** | **26** | **27** | **28** |

**Scheme.** Structures of the analyzed compounds.

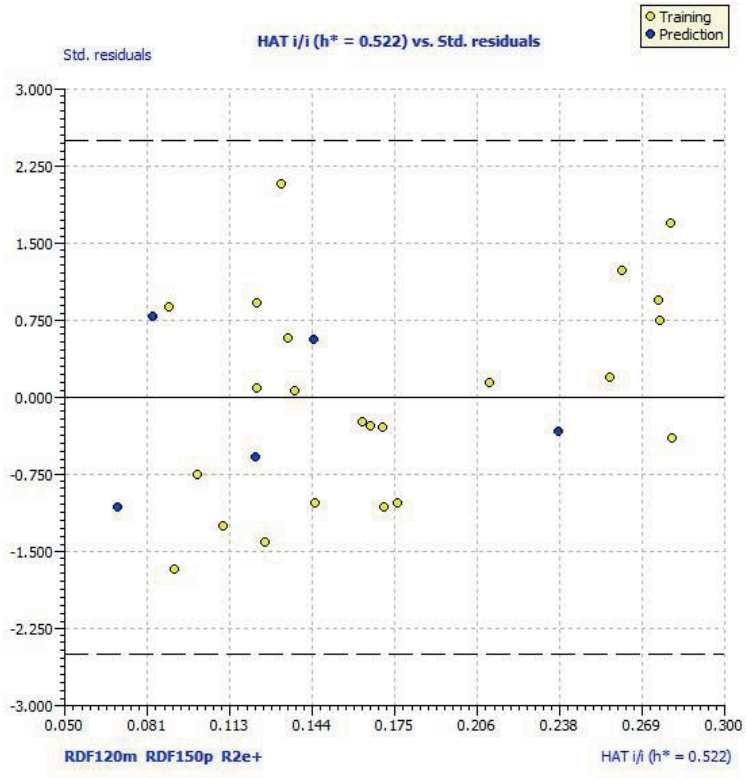
**Table 4**. Correlation matrix for the descriptors included in model (2).

| | *BIC1* | *RDF110m* | *DIPX* |
|---|---|---|---|
| *BIC1* | 1 | | |
| *RDF110m* | 0.17 | 1 | |
| *DIPX* | –0.45 | –0.51 | 1 |

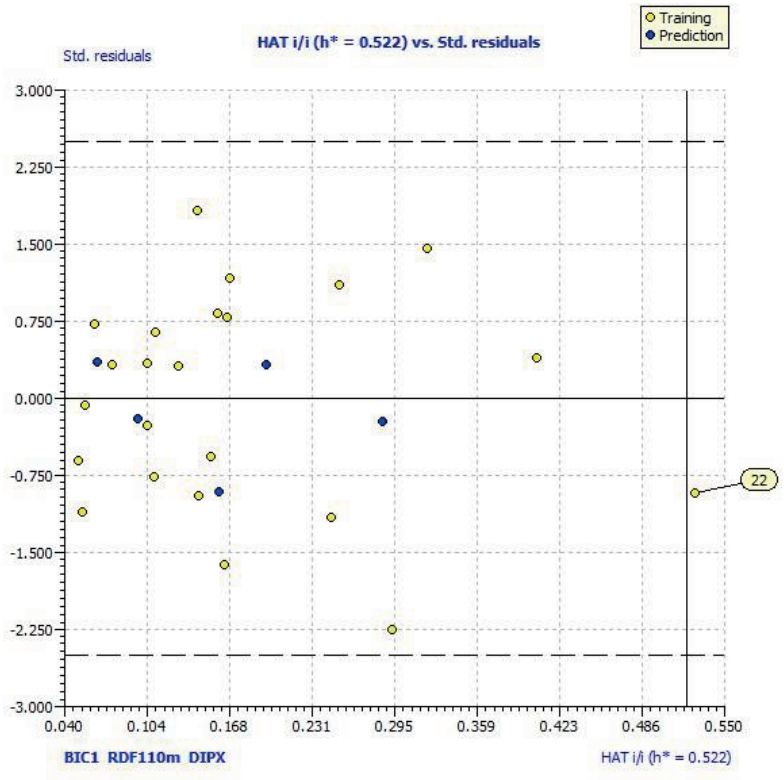
The lower values of $CCC_{ext}$, $Q^2_{F1}$,$Q^2_{F2}$,$Q^2_{F3}$, and $\bar{r}^2_m$ for model (1) indicate weaker external predictivity of this model compared to model (2). In addition, *CCC* values for model (1) are lower than an arbitrary cut-off value of 0.85, but it is not very restrictive.[21] The leave-one-out (LOO) validation highlights that the model is stable, robust, and not obtained by chance, since $Q^2_{LOO} > 0.5$, $R^2_y$scr, and $Q^2_y$scr $< 0.2$, as $R^2_y$scr $> Q^2_y$scr.

Williams plots (Figure 1) were used in order to define the chemical domain of applicability for which given QSAR models make reliable predictions.[22] Inspection of the Williams plot for model (1) (Figure 1a) revealed no outliers and no compounds outside of the hat value of leverage or warning leverage ($h^* = 0.52$), which means that all the predicted data for compounds belong to the chemical domain. Model (2) also has no outliers, except compound **22**, which slightly exceeds the critical hat value of leverage. This can be explained by the fact that the applicability domain represents the "distance" of the compound from the experimental model, and compound **22** has shown exceptionally high anticancer activity against the Raji leukemia line. For that compound, prediction is not reliable and predicted values of these compounds must be used with reserve (Figure 1b).

Despite the difference in the predictive potential of the proposed models, the included molecular descriptors may aid in the elucidation of important physicochemical and structural requirements for the antitumor




a)




b)

**Figure 1**. Applicability domain of the QSAR models for antitumor activity: a) against K562 leukemia cell lines calculated by model (1) and b) against Raji leukemia cell lines calculated by model (2).

activity of CQArA hybrids. Two radial distribution function (RDF) descriptors are present in model (1). The radial distribution function can be interpreted as the probability distribution of finding an atom in a spherical volume of a certain radius.[23] Therefore, the descriptor *RDF120m* included model (1) presents 3D molecular distribution of mass in a spherical volume of radius 12 Å. The molecules with the largest molecular radius and highest molecular weight have the highest values of descriptor *RDF120m* have (Table S1). According to the negative sign of *RDF120m* in model (1), higher values of those descriptors are favorable for the exhibition of antitumor activity. Similarly, the negative coefficient of descriptor *RDF110m* in model (2) implies that enhanced 3D molecular distribution of mass calculated at radius 11 Å from the specific geometrical centers of each molecule increases the antitumor potential. Figure 2 presents the 3D distribution of the mass at 11 Å from the center of the most active compound, **22**.

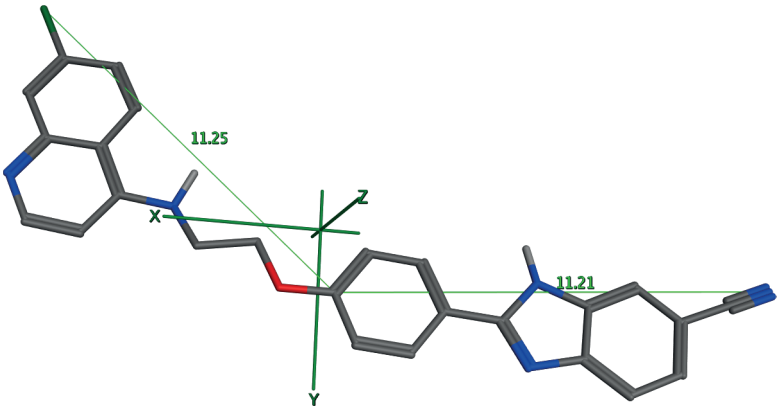



**Figure 2**. Structure of the most active compound (**22**) against Raji cells with representation of distance between the external atoms (nitrogen and chlorine) from the geometrical center of molecule.

The positive regression coefficient of *RDF150p* in model (1) suggests that enhanced atomic polarizability at the spherical volume of radius 15 Å negatively influences the antitumor activity of CQArA compounds. The value of that descriptor is strongly influenced by the number of carbon atoms in the external part of a molecule. Difference in values of that descriptor is mostly visible between the isopropyl and isobutyl-substituted **10** and **11** and **17** and **18**, respectively (Scheme 1). Descriptor $R2e^+$ is a descriptor of R maximal autocorrelation of lag 2 weighted by Sanderson electronegativity, which belongs to the GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors. The given descriptor is derived from the representation of molecular structure called the influence/distance matrix, **R**, where the elements of the molecular influence matrix are combined with geometric interatomic distances in the molecule.[24] It takes into account local aspects of the molecule, with atomic electronegativity at topological distance 2. The most external atoms at the small interatomic space have the highest influences on the value of R-GETAWAY descriptors. Therefore, molecules with a higher number of terminal electronegative atoms (nitrogen) have higher values of $R2e^+$ (Table S1) and consequently lower values of $\log GI_{50}$, according to the model (1). Thus, the highest values of $R2e^+$ are seen for the two most active compounds, **15** and **22**. Those molecules possess a strongly polarized nitrile group with an electrophilic carbon atom. Nitriles have a nonbonding electron pair localized on the nitrogen atom in *sp* hybrid orbitals, which makes them much less basic than amines and better hydrogen bond acceptors. Furthermore, involving a 2-aminoethanol linker in compounds **15** and **22**, compared to compounds **1** and **8**, which also have a terminal nitrile group, but shorter linker to 7-CQ, increased the value of $R2e^+$ (Table S1) and enhanced antitumor

activities against the growth of K562 cells. As we pointed out previously, extension of the molecules' central linker can improve the antiproliferative effect since it enhances the flexibility and adaptability of the molecular conformation when binding to DNA.[15]

Descriptor *BIC1* is the bond information content index (neighborhood symmetry of 1-order) that belongs to the information indices. The *BIC* descriptors take into account the number of bonds and bond orders. More saturated molecules with lower total number of atoms have greater values of *BIC1*. According Eq. (2), these structural properties enhanced the activity of molecules. Dipole moment along the $x$ axis (*DIPX*) is a third variable in Eq. (2). It is an important structural feature of the analyzed CQArA compounds since the most highly polar bond, C-Cl, is placed near the $x$axis (Figure 2). The negative sign of *DIPX* in model (2) implies that an enhanced value of *DIPX* is favorable for the antitumor activity of CQArA compounds. For example, the two most active compounds, **15** and **22**, have *DIPX* close to zero, compared to inactive compounds **17** and **18** with *DIPX* values over –5 (Figure 3). These observations suggest that the inclusion of terminal polar groups, such as a nitrile group, makes the molecule less polar due to symmetric charge distribution.

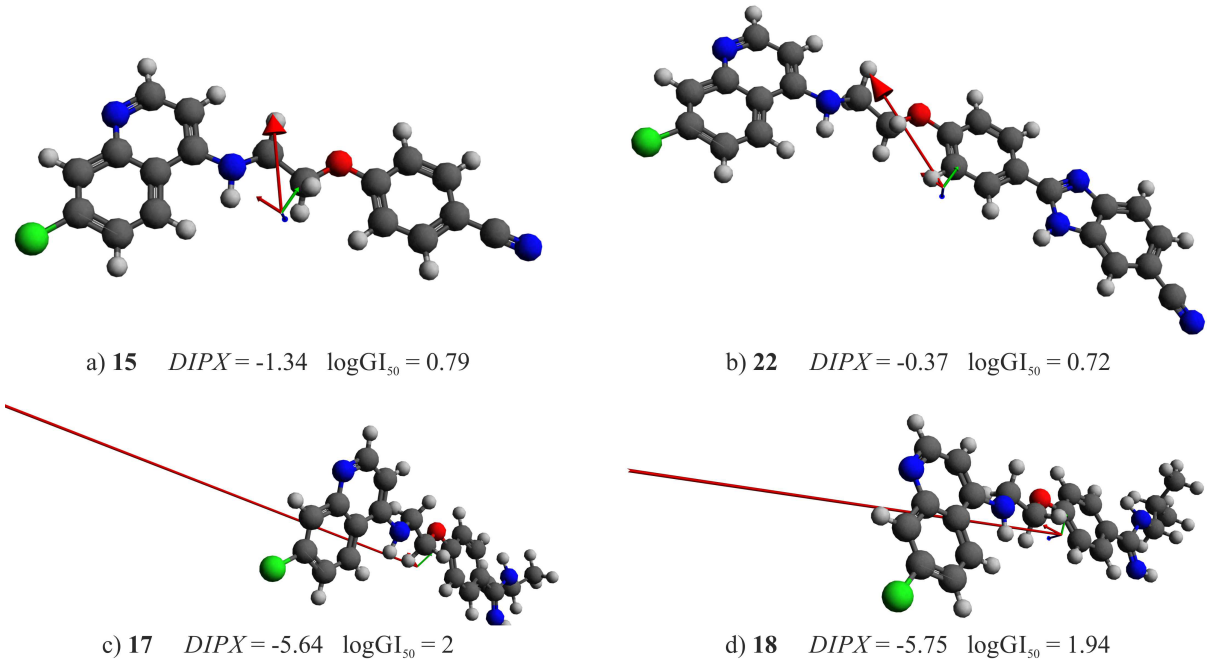



**Figure 3**. Dipole moments of the two most active compounds (**15**, **22**) and the two least active compounds (**17**, **18**).

The results of QSAR analysis could help to discover new compounds with possibly great antitumor potential. We propose two molecules as new potential anticancer agents. Activity against Raji leukemia cells has been predicted by means of model (2) (Figure 4). Antitumor activities of the proposed compounds are $\log GI_{50calc} = 0.02$ and –0.29 for **29** and **30**, respectively. Compared to compound **15**, proposed molecule **29** has an extended central linker between 7-CQ and arylamidine by one -$CH_2$- group, which resulted in enhancement of the calculated activity. Involving the second nitrile group on the benzimidazole, the polar charge is distributed more symmetrically, which makes a molecule less polar and potentially extremely powerful.

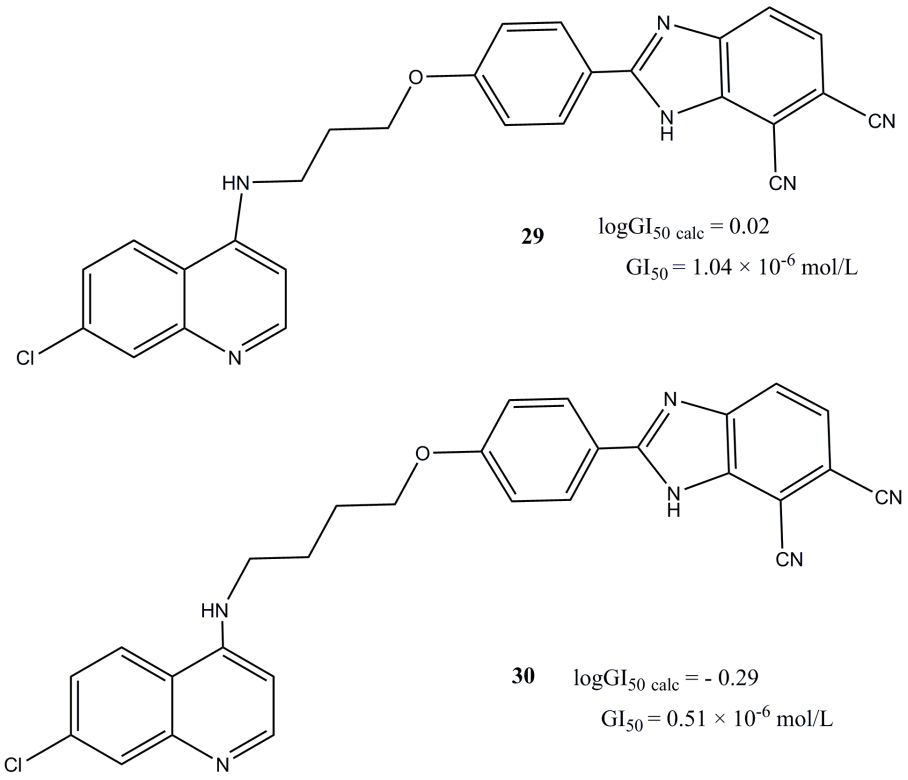



**Figure 4**. Structures of the proposed molecules with promising antiproliferative effects against leukemia Raji cell line. Activities have been predicted according to model (2).

### 2.2. Molecular docking study

Molecular docking studies provide virtual screening of different molecules as potential ligands to predict their ability to interact with given target candidates. Potential ligand molecules are ranked using various scoring functions.[25] For the present study we have chosen protooncogene tyrosine-protein kinase c-Src (PDB ID: 3G6H) as a target molecule. Compounds were ranked by total energy of predicted pose in the binding site. Docking results for the seven best ranking CQArA compounds are tabulated in Table 5. Results were compared with docking results of bonded ligand pyridinyl triazine (DSA1, PDB ID: G6H).[19]

**Table 5**. Ranking of the screened compounds by docking study.

| Compound (pose numb.) | Total energy/kcal mol$^{-1}$ | vdW | HBond | Elec |
|---|---|---|---|---|
| **22** (1) | –137.64 | –124.22 | –13.424 | 0 |
| **29** (0) | –135.57 | –125.29 | –10.28 | 0 |
| **28** (0) | –129.18 | –118.68 | –10.5 | 0 |
| **30** (1) | –127.79 | –118.29 | –9.5 | 0 |
| **27** (0) | –125.93 | –118.93 | –7 | 0 |
| **8** (0) | –121.05 | –113.09 | –7.9542 | 0 |
| **12** (0) | –120.21 | –105.67 | –14.54 | 0 |
| **G6H** | –118.94 | –102.72 | –16.223 | 0 |

As shown in Table 5, compound **22**, 2-{4-[2-(7-chloroquinolin-4-ylamino)ethoxy]phenyl}-1Hbenzo[d]imidazole-6-carbonitrile, has the biggest negative total energy (–137.64 kcal/mol), which is consistent with the fact that **22** is also the most active compound (Table 1). After compound **22**, compound **29** appears in the ranking, followed by compound **30**, both proposed by the QSAR study. Their high energy-based scoring functions confirm their possible high antitumor activity predicted by QSAR.

The binding site of tyrosine kinase was defined according to the inhibitor DSA1.[19] The structure of the Src kinase domain consists of a small amino-terminal (N) lobe (residues 267–337) and large carboxyl-terminal (C) lobe (residues 341–520) connected by a "hinge". The N lobe contains antiparallel $\beta$-sheet structures and an $\alpha$-helix ($\alpha$C-helix). The larger C lobe is predominantly $\alpha$-helical and contains a flexible activation loop, which starts with a DFG motif. The ATP binding pocket of Src lies deep in the cleft between the two lobes. The two lobes move relative to each other and can open or close the cleft.[17] A "gatekeeper" is a single residue in the ATP binding pocket located in the hinge region between the N and C lobes of the kinase that separates the adenine binding site from an adjacent hydrophobic pocket.[26] Molecular docking confirmed that, similar to DSA1, compound **22** is located in the cavity between two lobes of c-Src (Figures 5 and 6). Interactions of compound **22** with c-Src in the binding site are presented in Figure 7. Compound **22** forms hydrogen bonds with Tyr-340 via nitrogen from the nitrile group and $\pi$-alkyl hydrophobic interactions of the benzimidazole ring with Leu-273 and Leu-393 in the hinge region. For comparison, DSA1 in the same region forms the hydrogen bond of the exocyclic nitrogen from the trimethoxyaniline ring and nitrogens from the triazine ring.[19] Key interactions are H bonds between amino groups from the 2-aminoethanol linker and Asp-404 and the quinolone nitrogen atom and Tyr-382 in the DFG motif. Hydrogen bond interactions with the hinge region indicate that compound **22** is a potent tyrosine kinase inhibitor.

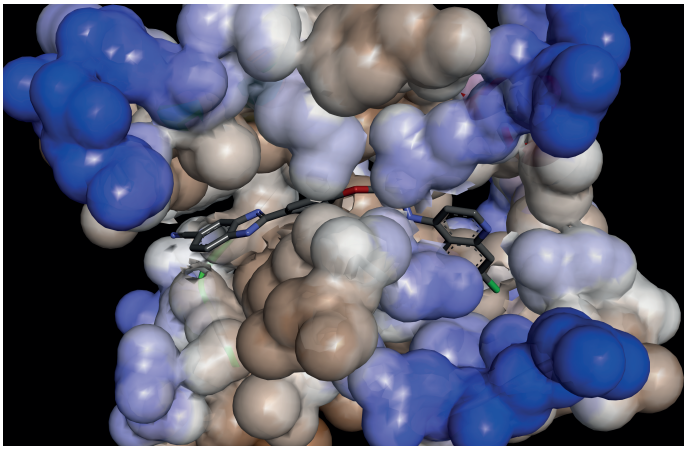


**Figure 5**. vdW surface representation of a segment of c-Src active site cleft with docked compound **22**.

In addition, other interactions are mediated through van der Waal (vdW) and hydrophobic contacts with Val-402, Glu-310, Asp-404, Met-314, and Ala-403 (see Figure 6). Similar to DSA1, compound **22** is connected with a glutamate residue in the $\alpha$C helix (Glu-310) and Asp-404 and Ala-403 in the DFG motif. Asp-404 is the first residue of the activation loop found in the large lobe of c-SRC. The chloroquinoline group of **22** occupied the specificity pocket by hydrophobic interactions with Val-377, Leu-322, Leu-317, and His-384. Likewise, the 3-trifluoromethylphenyl group of DSA1 is anchored into the hydrophobic pocket composed of the side chain Val-313, Leu-317, Leu 322, Val-377, Tyr-382, and Val-402.[19]

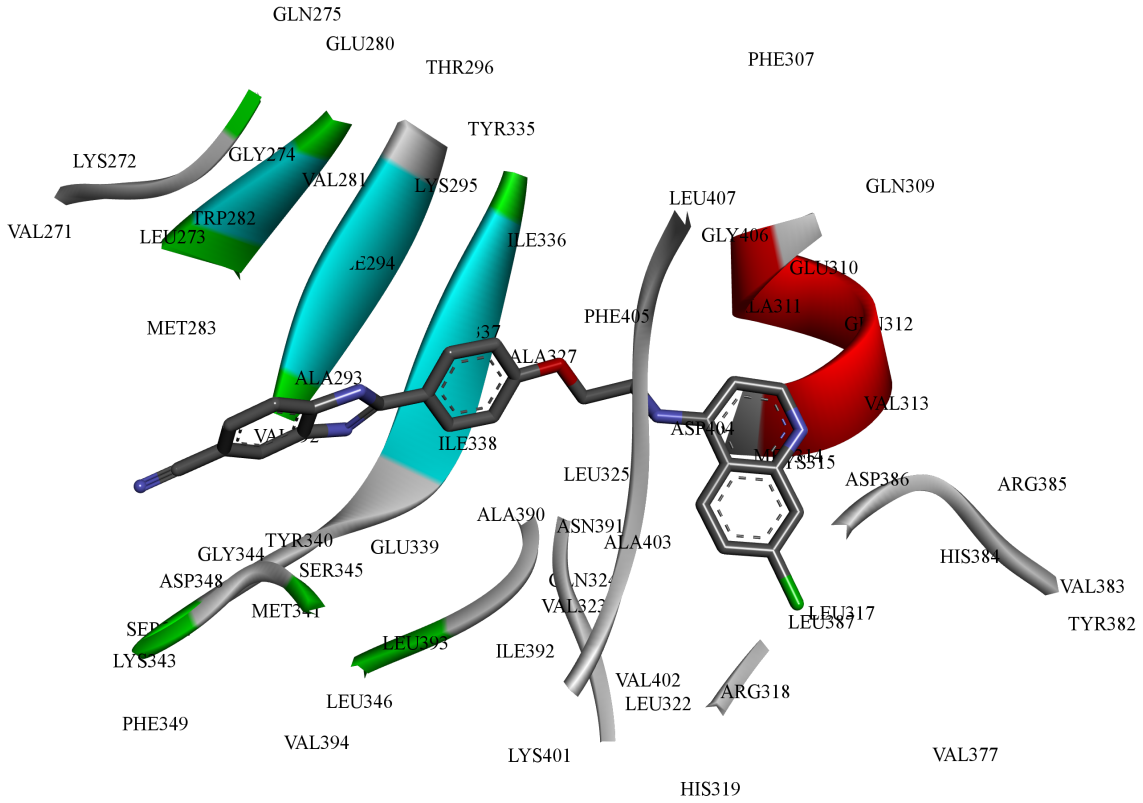



**Figure 6**. Cavity of c-Src with docked compound **22**: P-loop (turquoise); helix $\alpha$C (red). Figure 7 shows the interactions of compound **22** with c-Src in binding site.

Our results are consistent with the previous results of a docking study of azaacridine derivatives,[27] which revealed that the amido moiety formed two hydrogen bonds with Glu-310 and Asp-404 in the DFG out-pocket of c-SRC. This is in agreement with previously QSAR findings that the 2-aminoethanol linker (compounds **15**–**28**) between quinoline and arylamidine enhanced anticancer activity compared to the other compounds (**1**–**14**) without an amino group in the central linker.

### 2.3. Conclusions

In the present work, the QSAR analysis of anticancer activities of 7-chloroquinoline-arylamidine hybrids against two leukemia cell lines, Raji and K562, has been performed. The best obtained QSAR models revealed that important physicochemical and structural requirements for the anticancer activity of CQArA hybrids are enhanced 3D molecular distribution of mass calculated at radius 11 Å from the geometrical center of the molecule; the higher number of terminal electronegative atoms; extension of the molecules' central linker between the quinoline and arylamidine parts, which should contain an amino group; a higher ratio of single bonds and total number of atoms; and symmetric charge distribution that leads to the less polar molecule. According to the predictive model for anticancer activity against Raji cells, two new potent quinolone-arylamidine analogues were proposed.

Molecular docking has confirmed the importance of the terminal nitrile group and amino groups from the 2-aminoethanol linker for enhanced binding affinity toward tyrosine kinase. Our study suggests that CQArA hybrids could be developed as novel tyrosine kinase inhibitors in the treatment of leukemia.

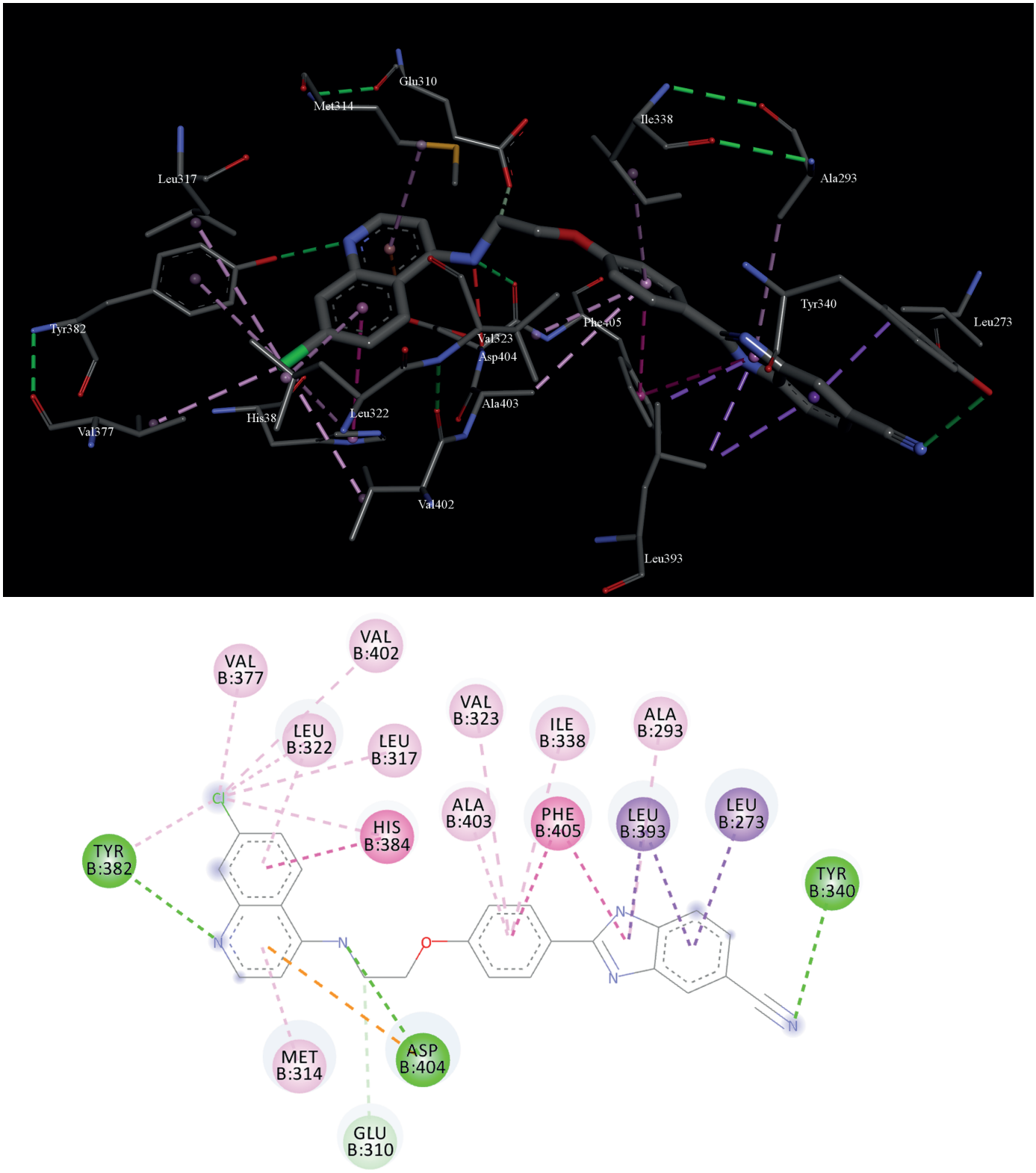



**Figure 7**. Interactions of compound 22 with c-Src in binding site: a) 3D representation; b) 2D representation (green = hydrogen bond; purple = $\pi$-$\sigma$ interactions; light purple = $\pi$- $\pi$ interactions; pink = alkyl and $\pi$-alkyl interactions).

## 3. Experimental

### 3.1. Dataset

The synthesis and evaluation of anticancer activity derivatives were described in our previous study.[15] Structures of the analyzed compounds are presented in the Scheme. Anticancer activities, expressed as the concentration achieving 50% cell growth inhibition ($GI_{50}/\mu M$) against leukemia cell lines (Raji and K562), were converted in the form of the logarithm ($\log GI_{50}$) and are presented in Table 1. For the inactive compounds whose $GI_{50}$ values were estimated as 100, $\log GI_{50}$ was set to 2.

### 3.2. QSAR

#### 3.2.1. Descriptor calculation

The 3D structures of 28 molecules were optimized applying Avogadro 1.2.0 (University of Pittsburgh, Pittsburgh, PA, USA) using the molecular mechanics force field (MM+).[28] Subsequently, all the structures were submitted to geometry optimization using the semiempirical PM3 method.[29] Two sets of descriptors were generated using Parameter Client (Virtual Computational Chemistry Laboratory, http://146.107.217.178/lab/pclient/), an electronic remote version of the Dragon program, and ADMEWORKS ModelBuilder 7.9.1.0 (Fujitsu Kyushu Systems Limited). In order to reduce the huge number of calculated descriptors (1728), zero value descriptors were first excluded from the initial pool. Further exclusion was performed using QSARINS-Chem 2.2.1 (University of Insubria, Varese, Italy): constant and semiconstant descriptors, i.e. those having chemical compounds with a constant value for more than 80%, and descriptors that were too intercorrelated ($> 85\%$) were rejected.[30] After reducing the number of calculated descriptors, the number of remaining descriptors in the pool was 492.

#### 3.2.2. Regression analysis and validation of models

Two algorithms were used for creating the training and test sets (20%), a random method and a ranking method, using QSARINS. The best QSAR models were obtained with a genetic algorithm (GA) using QSARINS. The number of descriptors ($I$) in the multiple regression equation was limited to three. The models were assessed by fitting criteria and internal cross-validation using the LOO method. Robustness of QSAR models was tested by Y-randomization test: the coefficient of determination ($R^2_Y\ scr$) and cross-validated correlation coefficient ($Q^2_Y\ scr$) should be significantly lower than the original model.

Fitting criteria included the coefficient of multiple determination ($R^2$), adjusted coefficient ($R^2_{adj}$), cross-validated correlation coefficient using LOO method ($Q^2_{LOO}$), global correlation among descriptors ($Kxx$), difference in correlation between the block of X variables plus response Y ($Kxy$) and $Kxx$ in models ($\Delta K$), standard deviation of regression ($s$), and Fisher ratio ($F$). Internal validations also included the following parameters: root-mean-square error of the training set ($RMSE_{tr}$), root-mean-square error of the training set determined through cross-validated LOO method ($RMSE_{cv}$), concordance correlation coefficient of the training set ($CCC_{tr}$), test set using LOO cross-validation ($CCC_{cv}$), mean absolute error of the training set ($MAE_{tr}$), mean absolute error of the internal validation set ($MAE_{cv}$), and predictive residual sum of squares determined through cross-validated LOO method ($PRESS_{cv}$) in the training set.[31–33]

Investigation of the applicability domain[22] of prediction models was performed by leverage plot (Williams plots) using QSARINS.

### 3.2.3. Molecular docking

The molecular docking of the compounds (**1**–**30**) was performed with *i*GEMDOCK (BioXGEM, Taiwan). Crystal coordinates of the c-Src (PDB ID: 3G6H) in the complex with DSA compound (PDB ID: G6H) were downloaded from the Protein Data Bank (PDB, https://www.rcsb.org/). In the first step, the structure of c-Src was prepared, including the removal of water molecules and optimized protein structure using BIOVIA Discovery Studio 4.5 (Dassault Systèmes, France). The binding site of the protein was defined according the bonded ligand (PDB ID: G6H). After preparing the protein target and set of optimized structures of 28 compounds as ligands, genetic parameters were set (population size: 200; generations: 70; number of solution or poses: 2). Each compound in the library was docked into the binding site and protein–compound interaction profiles of electrostatic (Elec), hydrogen-bonding (Hbond), and vdW interactions were generated. Finally, compounds were ranked by combining the pharmacological interactions and energy-based scoring function. Energy-based scoring function or total energy (E) is: E = vdW + Hbond + Elec.[34]

### Acknowledgment

The authors gratefully acknowledge the financial support from the J. J. Strossmayer University of Osijek supporting grant to Lj. Glavas-Obrovac (IZIP-2016) and Faculty of Medicine intramural grant VIF2017-MEFOS-8.

## References

1. Lage, H. *Cell. Mol Life Sci.* **2008**, *65*, 3145-3167.
2. Fortin, S.; Berube, G. *Expert Opin. Drug Discov.* **2013**, *8*, 1029-1047.
3. Kucuksayan, E.; Ozben, T. *Curr. Top. Med. Chem.* **2017**, *17*, 907-918.
4. Kimura T.; Takabatake Y.; Takahashi A.; Isaka Y. *Cancer Res.* **2013**,*73*, 3-7.
5. Solomon, V. R.; Lee, H. *Eur. J. Pharmacol.* **2009**, *625*, 220-233.
6. Mei, L.; Chen, Y.; Wang, Z.; Wang, J.; Wan, J.; Yu, C.; Liu, X.; Li, W. *Br. J. Pharmacol.* **2015**, *172*, 223-224.
7. Goodall, M.; Wang, T.; Martin, K. R.; Kortus, M. G.; Kauffman, A. L.; Trent, J. M.; Gately, S.; MacKeigan, J. P. *Autophagy* **2014**, *10*, 1120-1136.
8. Qiao, S.; Tao, S.; de la Vega; M. R.; Park; S. L.; Vonderfecht, A. A.; Jacobs, S. L.; Zhang, D. D.; Wondrak, G. T. *Autophagy* **2013**, *9*, 2087-2102.
9. McAfee, Q.; Zhang, Z.; Samanta, A.; Levi, S. M.; Ma, X. H.; Piao, S.; Lynch, J. P.; Uehara, T.; Sepulveda, A. R.; Davis, L. E. et al. *P. Natl. Acad. Sci. USA* **2012**, *19*, 8253-8258.
10. Rebecca, V. W.; Amaravadi, R. K. *Oncogene* **2016**, *7*, 1-11.
11. Soeiro, M. N. C.; Werbovetz, K.; Boykin, D. W.; Wilson, W. D.; Wang, M. Z.; Hemphill, A. *Parasitology* **2013**, *140*, 929-951.
12. Zhu, W.; Wang, Y.; Li, K.; Gao, J.; Huang, C. H.; Chen, C. C.; Ko, T. P.; Zhang, Y.; Guo, R. T.; Oldfield, E. *J. Med. Chem.*, **2015**, *58*, 1215-1227.
13. Stolić, I.; Čipčić Paljetak, H.; Perić, M.; Matijašić, M.; Stepanić, V.; Verbanac, D.; Bajić, M. *Eur. J. Med. Chem.* **2015**, *90*, 68-81.
14. Bistrović, A.; Krstulović, L.; Harej, A.; Grbčić, P.; Sedić, M.; Koštrun, S.; Kraljević Pavelić, S.; Bajić, M.; Raić-Malić, S. *Eur. J. Med. Chem.* **2018**, *143*, 1616-1634.

15. Krstulović, L.; Stolić, I.; Jukić, M.; Opačak-Bernardi, T.; Starčević, K.; Bajić, M.; Glavaš-Obrovac, Lj. *Eur. J. Med. Chem.* **2017**, *137*, 196-210.

16. Paul, M. K.; Mukhopadhyay, A. K. *Int. J. Med. Sci.* **2004**, *1*, 101-115.

17. Roskoski, R. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 1155-1164.

18. Li, S. *Int. J. Biochem. Cell Biol.* **2007**, *39*, 1483-1488.

19. Seeliger, M. A.; Ranjitkar, P.; Kasap, C.; Shan, Y.; Shaw, D. E.; Shah, N. P.; Kuriyan, J.; Maly, D. J. *Cancer Res.*, **2009**, *69*, 2384-2392.

20. Masand, V. H.; Mahajan, D. T.; Nazeruddin, G. M.; Hadda, T. B.; Rastija, V.; Alfeedy, A. M. *Med. Chem. Res.* **2015**, *24*, 1241-1264.

21. Chirico, N.; Gramatica P. *J. Chem. Inf. Model.* **2011**, *51*, 2320-2335.

22. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. *Environ. Health Perspect.* **2003**, *111*, 1361-1375.

23. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley: Weinheim, Germany, 2008.

24. Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Model.* **2002**, *42,* 682-692.

25. Agarwal, S.; Mehrotra, R. *JSM Chem.* **2016**, *4*, 1024.

26. Vijayan, R. S. K.; Peng, H.; Modi, V.; Duong-Ly, K. C.; Ma, H.; Peterson, J. R.; Dunbrack, R. L.; Levy, R. M. *J. Med. Chem.* **2015**, *58*, 466-479.

27. Cui, Z.; Chen, S.; Wang, Y.; Gao, C.; Chen, Y.; Tan, C.; Jiang, Y. *Eur. J. Med. Chem.* **2016**, *136*, 372-381.

28. Hocquet, A.; Langgård, M. *J. Mol. Model.* **1998**, *4*, 94-112.

29. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209-220.

30. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich S. *J. Comput. Chem.* **2013**, *34*, 2121-2132.

31. Gramatica, P. *QSAR Comb. Sci.* **2007**, *26*, 694-701.

32. Todeschini, R.; Consonni, V.; Maiocchi A. *Chemometr. Intell. Lab. Syst.* **1999**, *46*, 13-29.

33. Tropsha, A. *Mol. Inform.* **2010**, *29*, 476-488.

34. Hsu, K. C.; Chen, Y. F.; Lin, S. R.; Yang, J. M. *BMC Bioinformatics* **2011**, *12*, (Suppl. 1): S33.

**Table S1.** Values of the most relevant descriptors in QSAR models (1) and (2) with predicted values of values ($logGI_{50calc}$) calculated by QSAR models (1) and (2) for 30 CQArA hybrids against leukemia cell lines K562 and Raji, respectively.

| *No.* | *RDF120m* | *RDF150p* | *R2e+* | logGI 50 (K562 ) | *BIC1* | *RDF110m* | *DIPX* | logGI 50 (Raji ) |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.00 | 0.06 | 1.94 | 0.55 | 0.07 | -1.11 | 1.96 |
| **2** | 0.00 | 0.00 | 0.08 | 1.51 | 0.61 | 1.52 | -2.77 | 1.72 |
| **3** | 0.39 | 0.00 | 0.06 | 1.92 | 0.62 | 0.91 | -3.02 | 1.71 |
| **4** | 1.36 | 0.02 | 0.06 | 1.23 | 0.62 | 3.98 | -3.00 | 1.41 |
| **5** | 4.48 | 0.00 | 0.06 | 1.27 | 0.60 | 7.58 | -2.99 | 1.29 |
| **6** | 0.03 | 0.00 | 0.07 | 1.69 | 0.63 | 0.83 | -2.81 | 1.61 |
| **7** | 2.14 | 0.00 | 0.07 | 1.79 | 0.63 | 1.76 | -3.12 | 1.71 |
| **8** | 1.49 | 0.65 | 0.07 | 2.00 | 0.56 | 2.48 | -0.76 | 2.00 |
| **9** | 1.65 | 2.67 | 0.07 | 2.00 | 0.56 | 2.27 | -2.67 | 2.05 |
| **10** | 2.44 | 4.89 | 0.06 | 1.84 | 0.60 | 4.60 | -2.96 | 1.68 |
| **11** | 3.00 | 7.72 | 0.06 | 1.78 | 0.60 | 3.89 | -2.97 | 1.67 |
| **12** | 3.57 | 5.79 | 0.06 | 1.74 | 0.59 | 4.41 | -3.02 | 1.62 |
| **13** | 3.64 | 5.08 | 0.07 | 1.44 | 0.60 | 3.41 | -2.70 | 1.61 |
| **14** | 6.12 | 6.07 | 0.07 | 1.33 | 0.60 | 5.26 | -2.93 | 1.70 |
| **15** | 1.28 | 0.58 | 0.12 | 0.86 | 0.64 | 4.98 | -1.34 | 0.79 |
| **16** | 0.92 | 1.86 | 0.11 | 1.69 | 0.66 | 5.00 | -5.15 | 1.77 |
| **17** | 1.35 | 5.29 | 0.09 | 2.00 | 0.63 | 4.95 | -5.64 | 2.00 |
| **18** | 1.52 | 6.38 | 0.10 | 1.89 | 0.61 | 7.01 | -5.75 | 1.94 |
| **19** | 1.16 | 3.72 | 0.09 | 1.89 | 0.64 | 5.32 | -5.48 | 1.99 |
| **20** | 1.34 | 3.65 | 0.10 | 1.72 | 0.63 | 6.15 | -5.14 | 1.71 |
| **21** | 1.16 | 6.40 | 0.08 | 2.00 | 0.62 | 5.91 | -5.70 | 2.00 |
| **22** | 2.51 | 2.73 | 0.12 | 0.89 | 0.61 | 8.35 | -0.37 | 0.72 |
| **23** | 2.80 | 2.81 | 0.10 | 1.54 | 0.62 | 8.41 | -4.53 | 1.47 |
| **24** | 5.57 | 4.89 | 0.09 | 1.24 | 0.61 | 9.43 | -5.03 | 1.29 |
| **25** | 4.30 | 5.86 | 0.10 | 1.18 | 0.60 | 9.36 | -5.15 | 1.45 |
| **26** | 7.08 | 5.57 | 0.05 | 1.06 | 0.60 | 10.38 | -5.07 | 1.23 |
| **27** | 6.41 | 4.88 | 0.09 | 0.93 | 0.60 | 7.92 | -4.49 | 1.44 |
| **28** | 4.78 | 4.24 | 0.07 | 1.12 | 0.60 | 8.20 | -5.07 | 1.28 |
| **29** | 4.05 | 3.86 | 0.07 | 1.44 | 0.61 | 11.60 | 0.42 | 0.02 |
| **30** | 4.91 | 2.43 | 0.07 | 1.05 | 0.62 | 12.96 | 0.93 | -0.29 |