Journal of Molecular Graphics and Modelling 85 (2018) 198–211

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Quantitative structure–activity relationship of the thymidylate synthase inhibitors of *Mus musculus* in the series of quinazolin-4-one and quinazolin-4-imine derivatives☆

Veronika R. Khairullina [a, *], Alfiya R. Gimadieva [b], Anatoly Ya Gerchikov [a], Akhat G. Mustafin [a, b], Felix S. Zarudii [c]

[a] Bashkir State University, 32 Z. Validi str., Ufa, 450076, Russian Federation
[b] Ufa Institute of Chemistry - Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, 71 prospect Oktyabrya, Ufa, 450054, Russian Federation
[c] Bashkir State Medical University, 3 Lenina str, Ufa, 450000, Russian Federation

ARTICLE INFO

Article history:
Received 3 April 2018
Received in revised form
19 July 2018
Accepted 2 September 2018
Available online 10 September 2018

Keywords:
Antifolate thymidylate synthase inhibitors
QSAR models
GUSAR 2013
QNA and MNA descriptors
Structure–activity relationships

ABSTRACT

A quantitative structure–activity relationship analysis of the 2-methylquinazolin-4-one and quinazolin-4-imine derivatives, well-known antifolate thymidylate synthase (TYMS) inhibitors, has been performed in the range $IC_{50} = 0.4 \div 380000.0$ nmoL/L using the GUSAR 2013 program. Based on the MNA and QNA descriptors using the self-consistent regression, 6 statistically significant consensus models for predicting the IC_{50} numerical values have been constructed. These models demonstrate high and moderate prognostic accuracies for the training and external validation test sets, respectively. The molecular fragments of TYMS inhibitors regulating their antitumor activity are identified. The obtained data open opportunities for developing novel promising inhibitors of TYMS.

1. Introduction

Thymidylate synthase (TYMS, EC2.1.1.45) is a bi-substrate enzyme that takes part in the synthesis of deoxythymidine monophosphate (dTMP), an important nucleotide for the DNA synthesis [1–4]. The mechanism of dTMP biosynthesis from dUMP is now thoroughly studied and described in a number of works [1–7]. Synthesis of deoxythymidine monophosphate occurs in this enzyme by transferring the methyl group from 5,10-methylenetetrafolate (FH4, substrate I) to C5 deoxyuridylate (dUMP, substrate II). In this process, tetrahydrofolate FH4 is oxidized to dihydrofolate FH2, (Fig. 1S in Supplementary Material). Subsequently, dTMP is metabolized to deoxythymidine triphosphate, which is a structural element of DNA molecules. The reduction of FH2 to FH4 occurs with the participation of dihydrofolate reductase (DHFR) although this enzyme is not the only source of biosynthesis of substrate I (Fig. 1).

For several decades, biochemists have been actively studying inhibitors of TYMS. As is known, the increased TYMS activity is typical for tumor cells due to their high growth rate [8–15]. Therefore, this enzyme is a drug target of antitumor drugs classified as antimetabolites. The TYMS activity can be inhibited using the three strategies:

1) searching for DHFR inhibitors that indirectly inhibit TYMS;
2) searching for indirect TYMS inhibitors among the analogues of deoxyuridylate;
3) searching for analogues of 5,10-methylenetetrafolic acid [16,17].

Inhibition of DHFR reduces the synthesis of 5,10-methylenetetrafolic acid from folic and dihydrofolic acids, *i.e.*, it affects the source of the growth and development of tumor cells. Methotrexate is a well known inhibitor of DHFR and efficient cytostatic drug (Fig. 1). This substance is attributed to the group of folic acid antagonists and indirectly affects the TYMS activity. The active component of this drug irreversibly binds to dihydrofolate

☆ Veronika R. Khairullina, Alfiya R. Gimadieva, Anatoly Ya. Gerchikov, Akhat G. Mustafin, Felix S. Zarudii are contributed equally to this work.

* Corresponding author.
E-mail address: KhVerRad@gmail.com (V.R. Khairullina).

https://doi.org/10.1016/j.jmgm.2018.09.002
1093-3263/

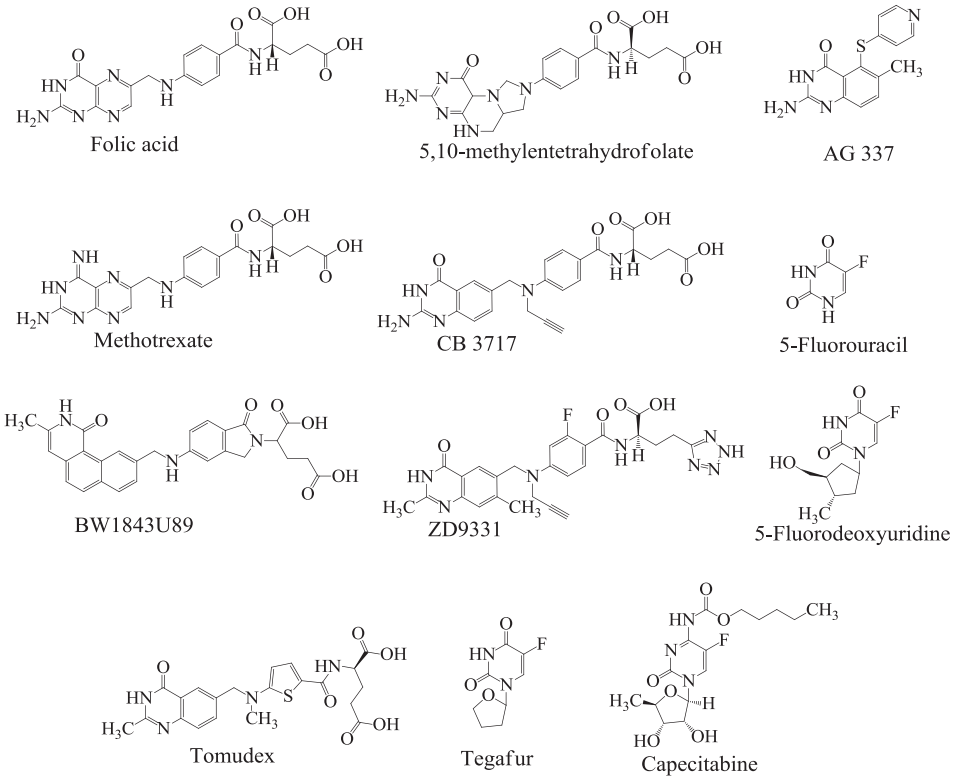

Fig. 1. Leader compounds and drugs inhibiting the DHFR and TYMS activities.

reductase and thereby inhibits the biosynthesis of thymidylate in tumor cells. In particular, methotrexate has an immunosuppressive effect even in relatively low doses, so it is very hard to tolerate by patients with tumors of various etiologies due to a pronounced side effects on the digestive and urinary systems. These pronounced side effects are followed from the fact that the DHFR inhibitors, in addition to the dTMP synthesis, indirectly inhibit purine biosynthesis. Furthermore, it is known that the point mutations in DHFR significantly increase the resistance of tumor cells to the DHFR inhibitors without reducing the ability of this enzyme to suppress purine biosynthesis [16]. Thus, the following two strategies aimed at searching for selective TYMS inhibitors are promising.

In particular, the search for structural analogues of deoxyuridylate among diverse pyrimidine derivatives is performed. 5-Fluorouracil (5-FU) is one of such analogs used to treat cancer for more than 50 years (Fig. 2). It is known that 5-FU not itself inhibits the biosynthesis of dTMP from dUMP but its metabolite, *viz.*5-fluorodeoxyuridine (5-FdUMP) (Fig. 1). It forms a covalent bond with 5,10-methylenetetrahydrofolate that leads to the irreversible inhibition of TYMS. Although 5-FdUMP alone or in combination with other agents is efficient against various human tumor cells (*e.g.*, in combination with uracil in preparate Tegafur), the low therapeutic stability and efficacy (due to its low selectivity toward TYMS) are the main limitation of use of this drug in chemotherapy [17,18].

The third strategy for the dTMP biosynthesis inhibitors seems more promising. It is focused on the search for selective TYMS inhibitors of antifolate type. These compounds, in contrast to pyrimidine derivatives, reveal a broader spectrum of *in vivo* activity and higher therapeutic efficacy. In this aspect, the 2-amino-quinazolin-4-one and 2-methyl-quinazolin-4-one derivatives are intensively studied. These compounds are structural analogs of folic

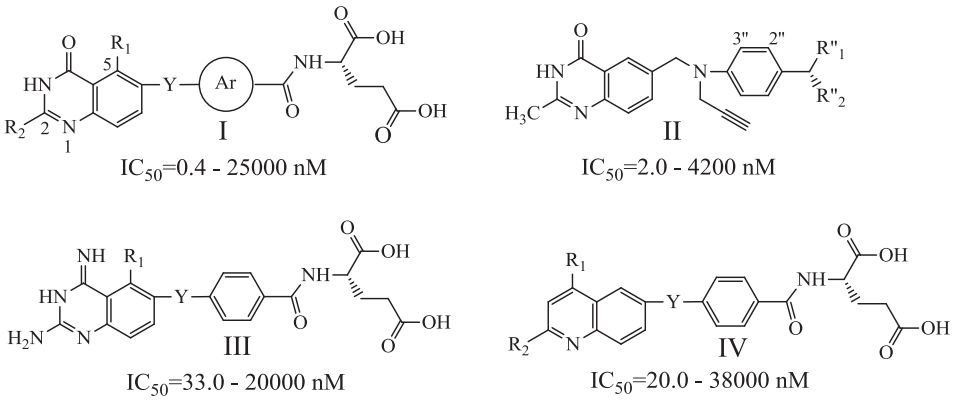

Fig. 2. General formulae of the antifolate-type TYMS inhibitors under study with ranges of their activity.

acid. In contrast to folic acid, the quinazolin-4-one derivatives contain a quinazoline ring as the core fragment (instead of the pteridine ring). Currently, several hundred quinazolin-4-one derivatives have been studied as the TYMS inhibitors with different degrees of selectivity toward this enzyme. In this group, five compounds, CB 3717, ZD 9331, BW 1843U89, AG 337, and Tomudex, are the most known (Fig. 1) and only one of them, Tomudex, is used in Europe as the inhibitor of the growth of tumor cells of different genesis. Leader compound CB 3717 has been rejected from clinical trials due to its unacceptably high hepatic and renal toxicity, which are associated with the poor solubility. Leader compounds ZD 9331, AG 337, BW 1843U89 (Fig. 1) are currently undergoing various phases of clinical trials as potential antitumor drugs. Despite its high antitumor efficacy, Tomudex has a high inhibitory effect on DHFR and, therefore, inhibits the purine biosynthesis. This induces a high gastrotoxicity of this drug. In addition, the joint use of Tomudex with leucovorin can significantly reduce the activity of the first drug. Thus, the search for biologically active substances that are able to selectively inhibit TYMS, and therefore slow down the process of DNA biosynthesis in tumor cells, is an important task of medical chemistry. However, the empirical analysis of biological data for developing new potential chemotherapeutic drugs without involving computational techniques *in silico* is a difficult task requiring excessive needs of time and resources [16,17].

Modern medical and bioorganic chemistry focuses on the rational synthesis of low-toxic organic compounds with a desirable set of biological and physicochemical properties. However, the experimental search for compounds satisfying these and other additional requirements to potential pharmaceuticals (Lipinski's rule of five, etc.) without virtual screening is inexpedient both in terms of time and material costs. In this regard, QSAR/QSPR methods are highly relevant at the stage of search for hit compounds with the purpose of their further proceeding as lead compounds at the preclinical stage. The use of these methods, either alone or in combination with other virtual screening methodologies (pharmacophore search or molecular docking) allows selecting the hit compounds for further testing. As is known, QSAR/QSPR methods are reliable for the search for effective agonists and antagonists of various receptors and selective inhibitors of enzymes [1–23].

We have previously reported on the results of modeling some TYMS inhibitors based on the quinazolin-4-one derivatives [24]. In the present study, using the GUSAR 2013 program, we have studied the quantitative structure–activity relationship of 196 inhibitors of thymidylate synthase of the antifolate type having the IC_{50} values from the range 0.52÷24800.00 nmoL/L. We have designed the statistically meaningful QSAR models for numerical prediction of the IC_{50}values. These models have a high accuracy of the prediction for the structures of training and test sets ($R^2_{TRi} = 0.855–0.922$; $Q^2_{TRi} = 0.810–0.895$; $R^2_{TS1} = 0.734–0.790$; $R^2_{TS2} = 0.800–0.835$). However, the obtained QSAR models are directed mainly to the IC_{50} prediction for TYMS-inhibiting 2-methylquinazolin-4-one derivatives and reveal a low prognostic ability for antifolates containing hydrophilic fragments in the quinazoline fragment. To construct statistically significant QSAR models useful for virtual screening of virtual libraries, it is necessary to supplement the training set with antifolates containing various substituents in the quinazoline ring and other aromatic moieties. This should expand the range of applicability of the QSAR consensus models.

Therefore, in this work, we have designed and validated the QSAR consensus models for the search for selective TYMS inhibitors in a series of quinazolin-4-one and quinazolin-4-imine derivatives with the general structural formulas **I**–**IV** shown in Fig. 2. The structures under study differ in the nature of the substituents in the quinazoline moiety and aromatic rings attached to quinazoline by the aliphatic linkers.

2. Computational details

A quantitative analysis of the structure–activity relationships was carried out only for the antifolate TYMS inhibitors. In total, 6 QSAR models M1–M6 were built. For this purpose, computer program GUSAR 2013 (General Unrestricted Structure Activity Relationships) was used [25–36]. A brief description of the capabilities of this program and the algorithms of constructing the quantitative structure–activity relationships is provided in Supplementary Material.

2.1. Constructing of the training and test sets

The construction of QSAR models M1–M6 was performed in the GUSAR 2013 program in several stages based on training sets TR1 and TR2. For the validation of these models, external and internal test sets of TS1 and TS2 were used. Training sets TR1, TR2, external and internal test sets TS1 and TS2 have been formed based on set S1 according to the scheme shown in Fig. 3.

Set S1 contains the IC_{50} data for 294 direct antifolate TYMS inhibitors of *Mus musculus*. The experimental IC_{50} estimates of these compounds have been obtained [37–47] under identical experimental conditions by the binding method in a model system containing the purified TYMS isolated from leukemia cells of mice L1210. The structures of the simulated compounds of set S1 differ in the nature of the linker, aromatic fragments Ar and end the marginal fragments R_1, R_2 (Fig. 2). Training set TR1 for constructing QSAR models M1–M3 includes 245 structures of antifolate TYMS inhibitors. It is formed based on set S2. The external test set TS1 is designed to test the validity of QSAR models M1–M6. It is generated from set S3. In turn, sets S2 and S3 are obtained as a result of splitting the set S1 in the ratio 5:1 by the following procedure: every sixth compound was transferred from S1 to S3.

Set S2 is made up by the structures of TYMS inhibitors not included in set S3. Previously, all structures of set S1 were ranked by increasing IC_{50}. Sets S4 and S5 are obtained as a result of splitting set S2 in a 4:1 ratio as every fifth compound was transferred from S2 to S5. Subsequently, training set TR2 and test set TS2 were formed based on sets S4 and S5, respectively. A detailed description of training TR1–TR2 and test sets TS1–TS2 is presented in Tables 1 and 2, respectively. A comparison of the data of these tables indicates that the distributions of compounds by activity in training and test sets are almost identical. Consequently, the mean values of the $\overline{pIC_{50}}$parameter of the TYMS inhibitors from training and test sets TR1, TR2, TS1, and TS2 almost coincide.

The structures of the compounds of training and test sets TR1–TR2, TS1–TS2 were generated in the MarvinSketch 17.22.0 program [48] and was converted into SDF-format using the Discovery Studio Visualizer program [49]. To construct QSAR models M1–M6, we used the IC_{50} values in mol/L, which were then expressed as pIC_{50}:

$$pIC_{50} = -\lg(IC_{50})$$

As follows from Table 1, the spread of IC_{50} values for training set ΔpIC_{50} exceeds 1.5, which is a necessary condition for constructing an adequate QSAR model [29,30].

2.2. Calculation of structural descriptors

To describe the structures of compounds within the program, two types of atom-centered descriptors were used, *viz.* substructural MNA (Multilevel Neighborhoods of Atoms), electrotopological

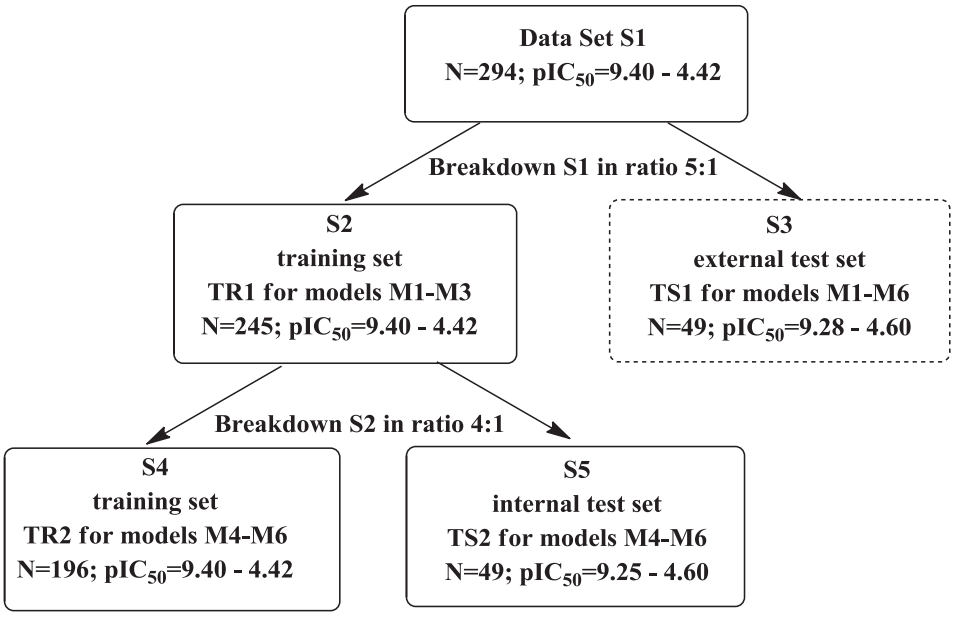

Fig. 3. Scheme of constructing the training and test sets and designing the M1–M6 QSAR consensus models (TR and TS are training and test sets, respectively, N is a number of compounds included to the corresponding sets and arrays). Designations: 1) S1 is all data set; 2) S2 is the training set TR1 for the M1–M3 models; 3) S3 is the external test set TR1 for the M1–M6 models; 4) partition S2 in the ratio 4:1; S2 is the training set TR1 for the M1–M3 models; 5) S4 is the training set TR1 for the M4–M6 models; 3) S5 is the internal test set TR1 for the M4–M6 models.

Table 1
Parameters of the training sets.

Comments	Parameters of training sets	Code of the training set	
		TR1	TR2
Number of training compounds	N_{TRi}	245	196
Mean pIC_{50}value for training set	$\overline{pIC_{50\,(TRi)}}$	6.9090	
Range of training set in pIC_{50}	$\Delta pIC_{50(TRi)}$	4.9777	
Distribution of observed response values of training sets TRi around training mean (in %)	$\overline{pIC_{50\,(TRi)}} \pm 0.5$, %	30.6120	34.1837
	$\overline{pIC_{50\,(TRi)}} \pm 1.0$, %	58.3670	58.6735
	$\overline{pIC_{50\,TRi}} \pm 1.5$, %	75.510	78.0612
	$\overline{pIC_{50\,(TRi)}} \pm 2.0$, %	93.469	90.8163
	$0.10 \times \Delta pIC_{50(TRi)}$	0.4978	
	$0.15 \times \Delta pIC_{50(TRi)}$	0.7467	
	$0.20 \times \Delta pIC_{50(TRi)}$	0.9955	
	$0.25 \times \Delta pIC_{50(TRi)}$	1.2444	

Table 2
Parameters of the test sets.

Comments	Characteristics parameters TSi	Code of test sets	
		TS1	TS2
Number of test compounds	N_{TSi}	49	49
Mean of test set value in pIC_{50}	$\overline{pIC_{50\,(TSi)}}$	6.9621	6.9288
Range of test set in pIC_{50}	$\Delta pIC_{50(TSi)}$	4.6438	4.6459
Distribution of observed response values of test sets TSi around test mean (in %)	$\overline{pIC_{50\,(TSi)}} \pm 0.5$, %	32.6531	34.6939
	$\overline{pIC_{50\,(TSi)}} \pm 1.0$, %	59.1837	59.1837
	$\overline{pIC_{50\,(TSi)}} \pm 1.5$, %	77.5510	77.5510
	$\overline{pIC_{50\,(TSi)}} \pm 2.0$, %	89.7959	89.7959
Distribution of observed response values of test sets TSi around train mean (in %)	$\overline{pIC_{50\,(TRi)}} \pm 0.5$, %	28.5714	30.6122
	$\overline{pIC_{50\,(TRi)}} \pm 1.0$, %	59.1837	59.1837
	$\overline{pIC_{50\,TRi}} \pm 1.5$, %	75.5102	75.5102
	$\overline{pIC_{50\,(TRi)}} \pm 2.0$, %	93.8776	93.8776

QNA (Quantitative Neighborhoods of Atoms) [26,27], and, additionally, three descriptors of the whole molecule (topological length, topological volume, and lipophilicity). QNA-descriptors were calculated automatically from the matrices of molecular

connectivity, standard ionization potentials (IP) and electron affinities (EA) of each atom. Thus, they depend on the structure of the molecule as a whole [28] and they are the basic information for regression coefficients. Thus, the regression equations constructed in the GUSAR 2013 program based on QNA-descriptors take into account both the specificity and physicochemical properties of each atom entering the training set [28,32–34].

MNA-descriptors are generated automatically based on the structural formulae of chemical compounds without using any precompiled list of structural fragments [22,26–28]. They are computed using the PASS algorithm (Prediction of Activity Spectra for Substances) [26,27]. The ideology of calculating QNA- and MNA-descriptors is expounded in previous works [32,33] and Supplementary material.

However, it noteworthy that the features of the QNA and MNA calculations retain these descriptors without unambiguous physical interpretation. For this reason, in the commercial and academic versions of the GUSAR 2013 program for broad use, the regression equations are not displayed.

2.3. Selection of the optimal number of descriptors

Self-consistent regression was used as a mathematical algorithm [27–33]. Previously, it has been shown [50] that self-consistent regression (SCR) can be successfully used to generate models from a large number of descriptors under different noise levels in the data. This method is correctly applied to modeling compounds with a rather high degree of similarity. The mathematical apparatus of this selection methodology is presented in works [32,33] and Supplementary material.

2.4. Constructing of the QSAR models

The GUSAR 2013 program allows constructing both partial regression dependencies and consensus models based on them. In this study, we use the consensus approach to construct the QSAR models. This allows reducing the variability of the predictions. Consensus models were designed in GUSAR 2013 automatically based on the principle of common similarity of particular regression dependencies [28–34].

Note that each of these partial models involved by the consensus model was made independently based on either QNA or MNA descriptors. As a result, 6 consensus QSAR models were designed. These models included 720 partial models. However, not all of them had acceptable statistical parameters. To select the most predictive models, a 20-fold crosscheck was performed for each model. Thus, 240 models were chosen from initial 720. These models have the R^2 values exceed 0.6 (from the cross-validation procedure after the randomized rejection of 20% of the training set). Each of the final consensus models M1–M2, M4–M5 is made up with 10 particular regression dependencies. Consensus models M3 and M6 include 100 regression equations. However, as the QNA and MNA descriptors have no direct physical meaning, the regression equations constructed on their basis are not explicitly displayed in the GUSAR 2013 program. Only the QSAR models satisfying the abovementioned condition have been further used for numerical predicting pIC_{50} for the compounds of the external training set.

It should be noted that the program is able to construct QSAR models both relying solely on one of these types of descriptors, and on their combination in terms of the consensus approach [29,30]. At the same time, based on the consensus approach methodology, models for quantitative prediction of biological activity for these descriptors are calculated independently of each other. The examples of the sample QSAR GUSAR models for predicting the toxic effects of chemical compounds are available free *via* the link http://www.way2drug.com/GUSAR.

2.5. Assessment of the range of the applicability

To assess the applicability of models, GUSAR 2013 provides three different approaches based on similarity, leverage, and accuracy previously described in detail [32,33].

Similarity. Using the Pearson correlation coefficients for each compound, program GUSAR 2013 calculated the distances toward its three nearest neighbors in the training set in the space of independent variables obtained after SCR. The compound is considered in the range of the model's applicability if the average value of these three distances is lower or equal to 0.7.

Leverage. The compound is considered out of the applicability range if its leverage is larger than 99% in the distribution of the leverage values of the training set.

Accuracy degree (AD). Here, the prediction of the applicability range for each compound is calculated based on the prediction error for the three most similar compounds in the test set relative to the training set as a whole [26,27]:

$$AD_{value} = RMSE_{3NN}/RMSE_{train}$$

In the present study, a threshold value of 1 was used for AD.

2.6. Assessment of the quality and predictive power of the QSAR models

2.6.1. Methodology of calculation of the pIC_{50} values using the consensus approach in the GUSAR 2013 program

The predictive efficiency of the constructed QSAR consensus models M1–M6 was estimated based on the IC_{50} parameter predictions for the training sets TR1-TR2 and test sets TS1–TS2. The final predicted pIC_{50} value for a particular compound using any consensus model is the weighted average predicted value of pIC_{50} estimated for all regression relationships included in this consensus model. In other words, first, the degree of similarity of the tested compound to the three structurally similar compounds of the training set is established. Then, based on the degree of this similarity, the prognosis of the numerical value of the activity for the test compound for each regression equation entering into the consensus model is made. After this, the calculated pIC_{50} values are averaged resulting in the final pIC_{50} of the test compound. As internal validation, the sliding control was used with a random 20-fold exclusion of 20% structures from the training set.

2.6.2. Statistical parameters characterizing the predictive power of the QSAR models

The predictive power of the QSAR models evaluated using external and internal test sets TS1 and TS2 was characterized using two categories of metrics:

1) metrics based on the determination coefficients R^2 (R^2_{TSi}, $R^2_{0(TSi)}$, $Q^2_{F1(TSi)}$, $Q^2_{F2(TSi)}$, $\overline{R^2_{m(TSi)}}$, CCC_{TSi});
2) metrics that estimate the errors for predicted pIC_{50} values (root mean square errors of prediction $RMSEP_{TSi}$, mean absolute errors MAE_{TSi} and standard deviations $S.D._{TSi}$) [51–55]. These statistical parameters are calculated by the program Xternal Validation Plus 1.2 using formulas (1–15) presented in Supplementary Material [56,57]. In addition, based on the results of the prediction of the pIC_{50} values for test sets TS1 and TS2, a systematic error of the constructed consensus models M1–M6 was evaluated using the same program.

2.7. Estimation of the atomic contributions to the target activity

Additionally, the GUSAR program allows visualizing the contribution of each atom into the predicted value [26–36]. This capability is implemented in the QSAR models based on the QNA descriptors and, accordingly, in the consensus combination of the QSAR models designed in different modes. It opens opportunities to identify "strong" and "weak" points in the biologically active molecules and, consequently, to rationalize the conclusions about the replacement of certain fragments upon molecular design directed to enhancing/weakening the target property.

The contribution of atoms to the activity of the antifolate TYMS inhibitors was estimated by the consensus models M3 and M6 containing 245 and 196 structures of the inhibitors, respectively. As reported [26–36], this procedure is implemented in the GUSAR 2013 program automatically when constructing the QSAR models based on QNA descriptors and consensus models.

3. Results and discussion

Using the consensus approach implemented in the GUSAR 2013 program, a quantitative relationship between the structure and inhibition efficiency of the catalytic activity of male white mice by the quinazoline-4-one and quinazolin-4-imine derivatives **I**–**IV** (Fig. 2) has been simulated with the general structural formulas. Herewith, depending on the type of the descriptors (MNA or QNA), three consensus QSAR models were obtained for each of the training sets. In total, 6 QSAR consensus models for the prediction of pIC_{50} values of the TYMS inhibitors were constructed including 240 partial regression equations. However, due to the fact that QNA and MNA descriptors do not explicitly reflect the physicochemical parameters of the substances compounds, the regression equations in the GUSAR 2013 program are not displayed. Therefore, it is impossible to estimate the contributions of each descriptors into the simulated activity [27–36]. However, this was not among the aims of this study. Indeed, we were solving two main problems: 1) to demonstrate the applicability of the GUSAR 2013 program to simulating the molecules with more than 10 mobile bonds, and 2) to construct the QSAR models that can be used in virtual screening of the TYMS inhibitors.

All models are characterized with a rather high descriptive power as follows from the determination coefficients calculated for the complete (100% data) and reduced (95%) data sets TrS1–TrS2 ($R^2_{TRi} > 0.85$, see Tables 3 and 4).

Designations. N_{TRi} is the number of structures in the training set; N_{PM} is the number of the unique regression equations used for the design of the consensus model; $\overline{R^2_{TRi}}$ are the averaged determination coefficients calculated for the compounds of the training set; $\overline{Q^2_{TRi}}$ are the averaged determination coefficients calculated for the training set with a sliding control with exception of one; $\overline{F_{TRi}}$ is the averaged Fisher criterion; $\overline{S.D._{TRi}}$ is the averaged standard deviation; V is the number of variables in the regression equation.

In general, a comparative analysis of the statistical parameters presented in Tables 3 and 4 allows concluding that the stable regression curves with acceptable statistical characteristics ($R^2_{TRi} > 0.6$, $Q^2_{TRi} > 0.5$) can be constructed in the GUSAR 2013 program both on one specific type and on both types of descriptors (QNA or MNA).

According to Table 4, the numerical values of different criteria R^2_{TRi} of the descriptive abilities of models M1–M6 almost coincide approaching to unit. The MAE_{TRi} values do not exceed 17% of training sets TR1–TR2, which indicates a good simulation of the target property using the GUSAR 2013 program.

where R^2_{TRi}, $R^2_{0(TRi)}$, R'^2_{0TRi}, $\overline{R^2_{m(TRi)}}$ are determination coefficients calculated for training sets TR1 and TR2 taking into account the average pIC_{50} values of these training sets; CCC_{TRi} is the concordance correlation coefficients; $RMSE_{TRi}$ is the root mean square error for training sets; MAE_{TRi} is mean absolute error; $S.D._{TRi}$ is the standard deviation; ωN_{TRi} is the part of training sets TRi (TR1 or TR2) having the prognostic error not exceeding the interval proportional to 0.1, 0.15, 0.20, and 0.25 of ΔpIC_{50} for training sets TR1 (a) and TR2 (b), respectively.

For the internal validation of QSAR models M1–M6, we have applied the sliding control procedure with a 20-fold randomized exception of 20% of the compounds of training sets TR1–TR2. The relevant numerical data are also presented in Table 3. The numerical values of the R^2_{TRi} and Q^2_{TRi} parameters for the compounds of both training sets are rather close. The difference between these parameters is less than 0.1, which indicates the stability of the constructed regression equations. High values of the coefficients and CCC_{TRi} is also observed.

The discrepancies in the numerical values of the determination coefficients R^2_{TRi} obtained for consensus models M1–M6 in the GUSAR 2013 and Xternal Validation Plus 1.2 programs (Tables 3 and 4) are explained with the different methodologies for calculating these parameters. Indeed, in GUSAR 2013, the activity prediction (pIC_{50} in our case) is performed according the degree of similarity of the tested compound toward the three structurally similar training set. Herewith, the numerical value of the activity for the test compound within any consensus model is the weighted average predicted pIC_{50} value estimated for all regression relationships included in this consensus model. In other words, all pIC_{50} values predicted by means of the partial regression models included in one consensus model are averaged and result in the final output pIC_{50}. Accordingly, the estimation of statistical parameters is also performed separately for each particular regression

Table 3
Statistical parameters and estimation of the prediction accuracy of the pIC_{50} values for the TYMS inhibitors according to the M1–M6 consensus models; $\Delta pIC_{50(TR1)} = \Delta pIC_{50(TR2)} = 4.9777$.

Training set	Consensus model	N_{TRi}	N_{PM}	$\overline{R^2_{TRi}}$	$\overline{F_{TRi}}$	$\overline{S.D._{TRi}}$	$\overline{Q^2_{TRi}}$	V
QSAR models based on the QNA descriptors								
TR1	M1	245	10	0.867	42.873	0.416	0.834	28
TR2	M4	196	10	0.812	34.001	0.493	0.771	20
QSAR models based on the MNA descriptors								
TR1	M2	245	10	0.858	39.392	0.432	0.826	28
TR2	M5	196	10	0.836	31.793	0.465	0.795	23
QSAR models based on both QNA and MNA descriptors								
TR1	M3	245	100	0.877	45.981	0.403	0.847	28
TR2	M6	196	100	0.841	34.681	0.458	0.803	21

Table 4
The validation parameters of the QSAR models estimated in the Xternal Validation Plus 1.2 program based on the experimental and predicted pIC_{50} for TYMS inhibitors of training sets TR1 (M1–M3) and TR2 (M4–M6); $\Delta pIC_{50(TR1)} = \Delta pIC_{50(TR2)} = 4.9777$.

Comments	Code of the model Prediction parameters	QSAR model used for the pIC_{50} prognosis for sets TR1 and TR2					
		based on TR1			based on TR2		
		M1	M2	M3	M4	M5	M6
Classical Metrics (100% data)	R^2_{TRi}(100% data)	0.9452	0.9401	0.9517	0.9260	0.9332	0.9404
	$R^2_{0(TRi)}$ (100% data)	0.9337	0.9327	0.9418	0.9138	0.9244	0.9297
	$R'^2_{0(TRi)}$ (100% data)	0.9119	0.9143	0.9245	0.8819	0.9009	0.9066
	$\overline{R^2_{m(TRi)}}$(100% data)	0.8147	0.8307	0.8308	0.7855	0.8127	0.8126
	$\Delta R^2_{m(TRi)}$ (100% data)	0.0637	0.0630	0.0558	0.0824	0.0712	0.0670
	CCC_{TRi} (100% data)	0.9622	0.9623	0.9671	0.9502	0.9571	0.9599
Classical Metrics (after removing 5% data with high residuals)	R^2_{TRi} (95% data)	0.9494	0.9476	0.9568	0.9327	0.9394	0.9456
	$R^2_{0(TRi)}$ (95% data)	0.9427	0.9437	0.9510	0.9259	0.9342	0.9392
	$R'^2_{0(TRi)}$ (95% data)	0.8117	0.8310	0.8327	0.7786	0.8032	0.8062
	$\overline{R^2_{m(TRi)}}$(95% data)	0.8497	0.8721	0.8682	0.8294	0.8499	0.8482
	$\Delta R^2_{m(TRi)}$ (95% data)	0.0530	0.0504	0.0445	0.0672	0.0594	0.0553
	CCC_{TRi} (95% data)	0.9681	0.9689	0.9729	0.9582	0.9632	0.9661
Mean absolute error and standard deviation for training set (100% data)	$RMSE_{TRi}$ (100% data)	0.2907	0.2928	0.2725	0.3313	0.3104	0.2993
	MAE_{TRi} (100% data)	0.2367	0.2358	0.2226	0.2739	0.2584	0.2482
	$S.D._{TRi}$ (100% data)	0.1691	0.1740	0.1575	0.1870	0.1724	0.1677
	$MAE_{TRi}+3 \times S.D._{TRi}$ (100% data)	0.7440	0.7578	0.6951	0.8349	0.7756	0.8024
Mean absolute error and standard deviation for training set (after removing 5% data with high residuals)	$RMSE_{TRi}$ (95% data)	0.2580	0.2604	0.2425	0.2969	0.2823	0.2695
	MAE_{TRi} (95% data)	0.2145	0.2134	0.2022	0.2503	0.2382	0.2276
	$S.D._{TRi}$ (95% data)	0.1437	0.1497	0.1342	0.1601	0.1519	0.1448
	$MAE_{TRi}+3 \times S.D._{TRi}$ (95% data)	0.6456	0.6623	0.6048	0.7306	0.6939	0.6619
Distribution of prediction errors (in %)	ωN_{TRi} in range $0.10 \times \Delta pIC_{50(TRi)}$	7.755[a]	9.796[a]	4.898[a]	14.796[b]	10.204[b]	7.653[b]
	ωN_{TRi} in range $0.15 \times \Delta pIC_{50(TRi)}$	0.000[a]	0.000[a]	0.000[a]	1.531[b]	0.510[b]	0.510[b]
	ωN_{TRi} in range $0.20 \times \Delta pIC_{50(TRi)}$	0.000[a]	0.000[a]	0.000[a]	0.000[b]	0.000[b]	0.000[b]
	ωN_{TRi} in range $0.25 \times \Delta pIC_{50(TRi)}$	0.000[a]	0.000[a]	0.000[a]	0.000[b]	0.000[b]	0.000[b]
Prediction quality	n/a	Good					
Systematic error presence	n/a	Absent					

model included in the consensus model used. Particularly, a set of 10 predicted pIC_{50pred} values and 10 sets of different internal validation criteria (R^2_{TRi}, Q^2_{TRi}, F_{TRi}, $S.D._{TRi}$.) is obtained for each individual set, *e.g.*, TR1, using the consensus model M1. Furthermore, in GUSAR 2013, the results of the activity prediction are averaged taking into account all 10 particular QSAR models. Similarly, the statistical parameters are averaged too.

At the same time, in the Xternal Validation Plus 1.2 program, calculating the statistical parameters of the external validation of the QSAR model is based on a comparison of the experimental pIC_{50} values with the averaged $pIC_{50\ pred}$, which are predicted using GUSAR 2013. This procedure is performed twice: 1) for a complete test set, and 2) for the same test set containing 95% of the original data. However, the results of statistical analysis in this program are not averaged [56].

Further, QSAR models M1–M6 based on TYMS inhibitors were used to predict the numerical values of pIC_{50} for external test set TS1 and internal test set TS2 containing the structural analogues of the quinazoline derivatives **I**–**IV** (Fig. 2). The inhibitory activity of sets TS1–TS2 has been determined under the same conditions as for training and test set. It should be noted that the TS1–TS2 sets were characterized by a uniform distribution of the pIC_{50} values relative to the range of training sets TR1–TR2. The centers of these sets also almost coincide with the centers of TR1–TR2. The results of the prediction of the pIC_{50} values for TS1 and TS2 sets by models M1–M6 are presented in Tables 2S and 3S in Supplementary Material.

The prognostic potential of QSAR models M1–M6 applied to on test sets TS1–TS2 was estimated using the Xternal Validation Plus 1.2 program. To characterize the predictive power of the model, two categories of metrics were used: 1) metrics based on the determination coefficients R^2 (R^2_{TSi}, $R^2_{0(TSi)}$, Q^2_{F1}, Q^2_{F2}, $\overline{R^2_{m(TSi)}}$, CCC_{TSi}); and 2) metrics estimating the prediction errors for pIC_{50} values ($RMSEP_{TSi}$, MAE_{TSi}, and $S.D._{TSi}$). Xternal Validation Plus 1.2 also allows assessing a systematic error occurring under prediction of the pIC_{50} values for structures of external and internal test sets. To avoid the false predictions related to unreliable experimental data, the program automatically removes 5% of the compound with high residual values. This step is also justified by the fact that most statistical tests are usually performed at a probability level of 5%.

The statistical parameters characterizing the predictive power of QSAR models M1–M6 based on various external validation criteria obtained from a comparison of the predicted and experimental pIC_{50} values for sets TS1 and TS2 are presented in Tables 5 and 6, respectively.

where R^2_{TS1}, $R^2_{0(TS1)}$, $R'^2_{0(TSi)}$ are the determination coefficients calculated, respectively, with or without taking into accout the origin of coordinates; $\overline{R^2_{m(TS1)}}$ is the averaged determination coefficient of the regression function calculated using the determination coefficients along with the ordinate axis (R^2_m) or abscissa (R'^2_m), respectively; $\Delta R^2_{m(TS1)}$ is the difference between $R^2_{m\ (TS1)}$ and $R'^2_{m(TS1i)}$; $Q^2_{F1(TS1)}$, $Q^2_{F2(TS2)}$ are the determination coefficients calculated for test set TS1 taking into account the average pIC_{50} value of the test set; CCC_{TS1} is the concordance correlation coefficient; MAE_{TS1} is the mean absolute error; $S.D._{TS1}$ is the standart deviation; ωN_{TS1} is a percentage of the compounds of test set TS1, for which the error of prediction does not exceed the interval proportional to 0.1, 0.15, 0.20, and 0.25 of ΔpIC_{50} for training sets TR1 (a) and TR2 (b), respectively.

where R^2_{TS2}, $R^2_{0(TS2)}$, $R'^2_{0(TS2)}$ are the determination coefficients calculated, respectively, with or without taking into accout the origin of coordinates; $\overline{R^2_{m(TS2)}}$ is the averaged determination coefficient of the regression function calculated using the determination coefficients along with the ordinate axis ($R^2_{m(TS2)}$) or abscissa

Table 5
The validation parameters of the QSAR models estimated using the Xternal Validation Plus 1.2 program based on the experimental and predicted pIC_{50} values of TYMS inhibitors of external test set TS1; $\Delta pIC_{50(TR1)} = \Delta pIC_{50(TR2)} = 4.9777$; $\Delta pIC_{50(TS1)} = 4.6438$.

Comments	Code of the model	Prediction parameters	QSAR model used for the pIC_{50} prognosis for set TS1					
			M1	M2	M3	M4	M5	M6
Classical Metrics (100% data)	R^2_{TS1} (100% data)		0.7856	0.8182	0.8212	0.7676	0.7826	0.8085
	$R^2_{0(TS1)}$ (100% data)		0.7772	0.8136	0.8151	0.7639	0.7807	0.8037
	$R'^2_{0(TS1)}$ (100% data)		0.6500	0.7355	0.7324	0.6476	0.6922	0.7168
	$Q^2_{F1(TS1)}$ (100% data)		0.7815	0.8209	0.8172	0.7689	0.7888	0.8058
	$Q^2_{F2(TS1)}$ (100% data)		0.7703	0.8116	0.8077	0.7569	0.7779	0.7958
	$\overline{R^2_{m(TS1)}}$(100% data)		0.5984	0.6736	0.6565	0.6043	0.6472	0.6472
	$\Delta R^2_{m(TS1)}$ (100% data)		0.2003	0.1621	0.1672	0.2014	0.1819	0.1745
	CCC_{TS1}(100% data)		0.8616	0.8901	0.8880	0.8564	0.8712	0.8812
Classical Metrics (after removing 5% data with high residuals)	R^2_{TS1} (95% data)		0.8660	0.8868	0.8851	0.8371	0.8285	0.8649
	$R^2_{0(TS1)}$ (95% data)		0.8624	0.8859	0.8837	0.8365	0.8285	0.8642
	$R'^2_{0(TS1)}$ (95% data)		0.6793	0.7504	0.7401	0.6654	0.6774	0.7136
	$Q^2_{F1(TS1)}$ (95% data)		0.8649	0.8872	0.8855	0.8394	0.8315	0.8661
	$Q^2_{F2(TS1)}$ (95% data)		0.8624	0.8851	0.8834	0.8364	0.8284	0.8636
	$\overline{R^2_{m(TS1)}}$(95% data)		0.7559	0.8220	0.7991	0.7485	0.7610	0.7795
	$\Delta R^2_{m(TS1)}$ (95% data)		0.1190	0.0952	0.1000	0.1342	0.1244	0.1144
	CCC_{TS1} (95% data)		0.9070	0.9028	0.9360	0.9090	0.9063	0.9252
Mean absolute error and standard deviation for test set TS1 (100% data)	$RMSEP_{TS1}$ (100% data)		0.5427	0.4914	0.4965	0.5582	0.5336	0.5117
	MAE_{TS1} (100% data)		0.3934	0.3715	0.3471	0.4259	0.4072	0.3808
	$S.D._{TS1}$ (100% data)		0.3777	0.3250	0.3586	0.3646	0.3484	0.3453
	$MAE_{TS1}+3 \times S.D._{TS1}$ (100% data)		1.5265	1.3465	1.4229	1.5197	1.4272	1.4167
Mean absolute error and standard deviation for test set TS1 (after removing 5% data with high residuals)	$RMSEP_{TS1}$ (95% data)		0.3992	0.3647	0.3675	0.4352	0.4458	0.3974
	MAE_{TS1} (95% data)		0.3195	0.3065	0.2789	0.3576	0.3513	0.3170
	$S.D._{TS1}$ (95% data)		0.2419	0.1999	0.2419	0.2508	0.2774	0.2422
	$MAE_{TS1}+3 \times S.D._{TS1}$ (95% data)		1.0453	0.9061	1.0046	1.1100	1.1835	1.0437
Distribution of prediction errors (in %)	ωN_{TS1} in range $0.10 \times \Delta pIC_{50(TRi)}$		30.612[a]	24.490[a]	22.449[a]	34.693[b]	32.653[b]	24.490[b]
	ωN_{TS1} in range $0.15 \times \Delta pIC_{50(TRi)}$		12.245[a]	10.204[a]	10.204[a]	12.245[b]	22.449[b]	12.245[b]
	ωN_{TS1} in range $0.20 \times \Delta pIC_{50(TRi)}$		6.122[a]	6.122[a]	8.163[a]	8.163[b]	6.122[b]	8.163[b]
	ωN_{TS1} in range $0.25 \times \Delta pIC_{50(TRi)}$		6.122[a]	4.082[a]	6.122[a]	6.122[b]	4.082[b]	4.082[b]
Prediction quality	n/a		Moderate					
Systematic error presence	n/a		Absent					

($R'^2_{m(TS2)}$), respectively; $\Delta R^2_{m(TS2)}$ is the difference between $R^2_{m(TS2)}$ and $R'^2_{m(TS2)}$; $Q^2_{F1(TS2)}$, $Q^2_{F2(TS2)}$ are the determination coefficients calculated for test set TS2 taking into account the average pIC_{50} value of the test set; CCC_{TS2} is the concordance correlation coefficient; MAE is the mean absolute error; $S.D._{TS2}$ is the standart deviation; ωN_{TS2} is a percentage of the compounds of test set TS2, for which the error of prediction does not exceed the interval proportional to 0.1, 0.15, 0.20, and 0.25 of ΔpIC_{50} for training sets TR2.

The criteria of the external validation of QSAR models M1–M6 (Tables 5 and 6) shows that the use of the full sets of the compounds from TS1 and TS2 allows achieving very close determination coefficients R^2_{TSi} and $R^2_{0(TSi)}$ being in the narrow ranges 0.7229–0.8212 and 0.7229–0.8156, respectively. Determination coefficients $Q^2_{F1(TSi)}$ and $Q^2_{F2(TSi)}$ lie in the range 0.7209–0.8209 and are close to determination coefficients, R^2_{TSi} и $R^2_{0(TSi)}$. However, the $Q^2_{F1(TSi)}$ value is usually higher than $Q^2_{F2(TSi)}$, R^2_{TSi} and R^2_{0TSi}, *i.e.*, $Q^2_{F1(TSi)} > Q^2_{F2(TSi)}$ (in all models), $Q^2_{F1(TSi)} > R^2_{TSi}$ (in models M2, M4-M5), $Q^2_{F1(TSi)} > R^2_{0(TSi)}$ (in all models) for 100% of TS1 and TS2 sets.

At the same time, criterion $\overline{R^2_{m(TSi)}}$ proposed by Roy et al. [52] for a more objective and rigorous assessment of the predictive power of QSAR/QSPR models is rather low and lies in the range 0.5882–0.6736 for external validation of consensus models M1–M6 using test sets TS1–TS2. Criterion $\Delta R^2_{m(TSi)}$, proposed by the same authors as an additional parameter for assessing the quality of prediction for external model validation, in all cases does not exceed 0.21. In all cases, for 100% of the data of TS1 and TS2 sets, the CCC_{TSi} coefficient obtains a satisfactory value from the range 0.8390–0.8901. All the data above indicate a moderate prognostic ability of QSAR consensus models M1–M6 with external validation.

The removal of 5% of compounds with high residues from both test sets increases all determination coefficients used for the external validation of models M1–M6 (Tables 5 and 6). Determination coefficients R^2_{TSi} of sets TS1 and TS2 in the case of 95% of the compounds lie in the range 0.8285–0.8868 and 0.8251–0.8483, respectively. Similar conclusions can be made about changing the $R^2_{0(TSi)}$ values, which increase to 0.8285–0.8859 and 0.8198–0.8364 for TS1 and TS2, respectively (in the case of 95% data sets). After removing 5% of compounds with high residues, determination coefficients $Q^2_{F1(TSi)}$ and $Q^2_{F2(TSi)}$ also increased to 0.8284–0.8872 and 0.8195–0.8431 for TS1 and TS2, respectively. Herewith, in the external validation of models M1–M6 with 95% of the original data, as in the case of the complete data set in TS1 and TS2, there was a contradictory insignificant excess of the Q^2_{F1}, parameter over $Q^2_{F2(TSi)}$, R^2_{TSi}, and $R^2_{0(TSi)}$ (Tables 5 and 6).

Removing 5% of the compounds from the TS1 and TS2 sets also positively affects the CCC parameters and allows reducing $\Delta R^2_{m(TSi)}$.

In general, based on the analysis of the determination coefficients, CCC, and $\Delta R^2_{m(TSi)}$ criteria (Tables 5 and 6), we conclude that all QSAR consensus models show satisfactory predictive abilities with external validation on the TS1 and TS2 structures, regardless of the volume of the sets. However, $Q^2_{F1(TSi)} > Q^2_{F2(TSi)}$, $Q^2_{F1(TSi)} > R^2_{TSi}$, $Q^2_{F1(TSi)} > R^2_{0(TSi)}$ in most cases. In some cases, the $Q^2_{F2(TSi)}$ and $R^2_{0(TSi)}$ values coincide up to the third decimal place. In other words, consensus models M1–M6 better predict the activity for the compounds of test sets TS1 and TS2 in comparison with the training set, although the quality of the activity prediction for the external test sets is usually lower than for the training samples. These facts were noted previously by other authors [51–54]. They suggest that the use only of the metrics based on R^2 and Q^2 to evaluate the prognostic ability of QSAR models is obviously insufficient.

Table 6
The validation parameters of the QSAR models estimated using the Xternal Validation Plus 1.2 program based on the experimental and predicted pIC_{50} values of TYMS inhibitors of external test set TS2; $\Delta pIC_{50(TR1)} = \Delta pIC_{50(TR2)} = 4.9777$; $\Delta pIC_{50(TS2)} = 4.6459$.

Comments	Code of the model	Prediction parameters	QSAR model used for the pIC_{50} prognosis for set TS2		
			M4	M5	M6
Classical Metrics (100% data)	R^2_{TS2} (100% data)		0.7777	0.7229	0.7945
	$R^2_{0(TS2)}$ (100% data)		0.7660	0.7229	0.7844
	$R'^2_{0(TS2)}$ (100% data)		0.6127	0.6183	0.6583
	$Q^2_{F1(TS2)}$ (100% data)		0.7785	0.7364	0.7936
	$Q^2_{F2(TS2)}$ (100% data)		0.7654	0.7209	0.7814
	$\overline{R^2_{m(TS2)}}$(100% data)		0.5882	0.6076	0.6089
	$\Delta R^2_{m(TS2)}$ (100% data)		0.2068	0.2096	0.1941
	CCC_{TS2} (100% data)		0.8540	0.8390	0.8670
Classical Metrics (after removing 5% data with high residuals)	R^2_{TS2} (95% data)		0.8407	0.8251	0.8483
	$R^2_{0(TS2)}$ (95% data)		0.8280	0.8198	0.8364
	$R'^2_{0(TS2)}$ (95% data)		0.5688	0.5901	0.5871
	$Q^2_{F1(TS2)}$ (95% data)		0.8317	0.8291	0.8431
	$Q^2_{F2(TS2)}$ (95% data)		0.8260	0.8195	0.8355
	$\overline{R^2_{m(TS2)}}$(95% data)		0.6840	0.6858	0.6767
	$\Delta R^2_{m(TS2)}$ (95% data)		0.1521	0.1560	0.1513
	CCC_{TS2} (95% data)		0.8939	0.8942	0.9015
Mean absolute error and standard deviation for test set TS2 (100% data)	$RMSEP_{TS2}$ (100% data)		0.5476	0.5973	0.5286
	MAE_{TS2} (100% data)		0.4498	0.4599	0.4231
	$S.D._{TS2}$ (100% data)		0.3155	0.3850	0.3200
	$MAE_{TS2}+3 \times S.D._{TS2}$ (100% data)		1.3963	1.6149	1.3831
Mean absolute error and standard deviation for test set TS2 (after removing 5% data with high residuals)	$RMSEP_{TS2}$ (95% data)		0.4703	0.4819	0.4608
	MAE_{TS2} (95% data)		0.3998	0.3925	0.3781
	$S.D._{TS2}$ (95% data)		0.2504	0.2827	0.2664
	$MAE_{TS2}+3 \times S.D._{TS2}$ (95% data)		1.1510	1.2407	1.1772
Distribution of prediction errors (in %)	ωN_{TS2} in range $0.10 \times \Delta pIC_{50(TR2)}$		38.7755	40.8163	34.6939
	ωN_{TS2} in range $0.15 \times \Delta pIC_{50(TR2)}$		18.3673	18.3673	20.4082
	ωN_{TS2} in range $0.20 \times \Delta pIC_{50(TR2)}$		6.1224	6.1224	2.0408
	ωN_{TS2} in range $0.25 \times \Delta pIC_{50(TR2)}$		2.0408	6.1224	2.0408
Prediction quality	n/a		Moderate		
Systematic error presence	n/a		Absent		

According to the results of Roy et al. [54], the predictive power of QSAR models can be correctly estimated using two criteria deduced from MAE: 1) directly the MAE value itself; and 2) the range of the spread of the predicted activities, taking into account MAE in the interval mσ (or mS.D.): $MAE_{TSi} + 3 \times S.D_{TSi}$. Conventionally, the predictive ability of a model is high if the MAE value when predicting pIC_{50} for test structures is 10% of the range of the pIC_{50} values of the training sets, on which the QSAR model is based. Here, the range of the pIC_{50} values of the training sets is designated as $\Delta pIC_{50(TRi)}$. The following relationship must be fulfilled: $MAE_{TSi} + 3 \times S.D._{TSi} \leq 0.2 \times \Delta pIC_{50\ (TRi)}$, where ΔpIC_{50} is the range of pIC_{50} values, in which the structures of training set TRi are located. At the same time, the predictive ability of the model is low if MAE_{TSi} of the pIC_{50} prediction for the test set is higher than 15% of the training set range used for designing the QSAR model. In this case, another relation should be satisfied: $MAE_{TSi}+3 \times S.D._{TSi} > 0.25 \times \Delta pIC_{50(TRi)}$ [54,56]. It is predicted that do not meet any of the above conditions are considered moderate.

The analysis of the predicted pIC_{50} values of the of external test sets TS1–TS2 (Tables 5 and 6), performed according to the above-described criteria, indicates the moderate predictive power of QSAR models M1–M6 after removing 5% of compounds with high residues in each of them. At the same time, using 100% of these test sets, there are discrepancies in the estimation of predictive abilities of models M1–M6. Particularly, in all cases, the magnitude of MAE does not exceed the parameter $0.1 \times \Delta pIC_{50(TR1)} = 0.1 \times \Delta pIC_{50(TR2)} = 0.4978$, which indicates a rather high predictive ability of M1–M6. However, the criterion $MAE_{TSi} + 3 \times S.D._{TSi} = 1.2444$ in the case of 100% of test sets TS1 and TS2 significantly exceeds the allowable threshold value $0.25 \times \Delta pIC_{50(TRi)}$, *i.e.*, the predictions in this case are considered unsatisfactory. This discrepancy may have two possible reasons: 1) the insufficient number of descriptors in the regression equation and the need to add new independent variables; 2) the presence of experimental errors in the experimental data. As follows from Table 3, each consensus model M1–M6 contains more than 20 QNA and MNA descriptors. Consequently, in our case, the discrepancy in assessing the predictive parameters of M1–M6, depending on the percentage of the data used (100 or 95%) is probably due to the experimental errors. This conclusion is supported by the absence of a systematic error in the prediction, which also allows rejecting the necessity of retraining the M1-M6 models.

As follows from Tables 3–6, all QSAR models reveal moderate prognostic ability on the structures of external and internal test sets TS1–TS2. These models are applicable to the virtual screening of virtual libraries and databases to search for new antifolate TYMS inhibitors based on the quinazoline derivatives.

As a rational finalization of this step of our study, we have performed virtual screening of the ChEMBL database using consensus model M3 to select potential TYMS inhibitors among the leader compounds and active components of known drugs. This model has been chosen because it is constructed using different types of descriptors and contains maximal number of structures. Additionally, the predictions errors of this model is acceptable and in most cases is close to the minimal value if we compare this parameter for TS1 and TS2. The facts above favor reliability and accuracy of the forecast results.

The virtual screening involved 200 quinazoline derivatives with the pronounced antitumor and antibacterial properties and no inhibitory activity toward thymidylate synthase. However, only 98 leading compounds and known pharmaceuticals are fitted into the

Table 7
Potential effective thymidylate synthase inhibitors selected from the ChEMBL database using virtual screening with QSAR model M3.

№	Code of compound[a]	Applicability domain (AD) of the model	Predicted IC_{50} value, nM
1	CHEMBL150607	in AD	740
2	CHEMBL36323	in AD	871
3	CHEMBL1738741	in AD	802
3	CHEMBL3228300	in AD	580
5	CHEMBL326511	in AD	971
6	CHEMBL146917	in AD	888
7	CHEMBL127972	in AD	734
8	CHEMBL331165	in AD	404
9	CHEMBL3228304	in AD	486
10	CHEMBL149218	in AD	908
11	CHEMBL475332/Chlorasquin	in AD	440
12	CHEMBL453872/Denopterin	in AD	777
13	CHEMBL459050/Diopterin	in AD	120
14	CHEMBL38937	in AD	742
15	CHEMBL1783014	in AD	49
16	CHEMBL3244856	in AD	675
17	CHEMBL162414	in AD	612
18	CHEMBL67297	in AD	888
19	CHEMBL40385	in AD	897
20	CHEMBL3094439	in AD	639
21	CHEMBL126579	in AD	946
22	CHEMBL37936	in AD	564
23	CHEMBL3244859	in AD	484
24	CHEMBL38313	in AD	439
25	CHEMBL3228305	in AD	549
26	CHEMBL22708	in AD	572
27	CHEMBL118230	in AD	592
28	CHEMBL3244853	in AD	456
29	CHEMBL80133	in AD	946
30	CHEMBL75914	in AD	457
31	CHEMBL141997	in AD	455
32	CHEMBL77257	in AD	649
33	CHEMBL118927	in AD	480
34	CHEMBL3706582	in AD	822
35	CHEMBL3228303	in AD	684
36	CHEMBL435217	in AD	969
37	CHEMBL2153708	in AD	920
38	CHEMBL476400	in AD	920
39	CHEMBL586489	in AD	19

[a] Compound codes in the ChEMBL database (https://www.ebi.ac.uk/chembl/).

range of the applicability of consensus-model M3. For 39 structures from them, the predicted IC_{50} values are less than 1 μmoL/L. These compounds are presented in Table 7. We assume that in living systems these compounds should behave as multi-target drugs. They are promising for further detailizing studies. A complete list of the structures of potential TYMS inhibitors predicted by consensus model M3 is presented in the Table 4S in Supplementary Material.

Thus, the approach used in the GUSAR 2013 program allows highly reliable modeling the inhibitory activity of the quinazoline-4-on derivatives toward TYMS and finding new antifolate inhibitors of this enzyme.

Additionally, the structures of the quinazolin-4-one, quinazolin-4-imine and quinolin derivatives **I**–**IV** have been analyzed. For most of the compounds studied, the structure–activity relationship analysis (SAR) was previously performed [37–47] but those SAR data have mostly a disparate nature. In the present work, we have performed a systematic comparative study of 294 known antifolate TYMS inhibitors and detected the functional groups permitting modulation of their activity. The analysis of the contribution of the structural fragments to the resulting inhibition activity has been based on the results of comparative studies and data obtained using the program GUSAR 2013. These have been deduced only from the comparative reasoning.

A detailed analysis of the results is presented in Tables 5S–15S in Supplementary Material. On the basis of numerous experimental data, the following regularities can be revealed.

Effect of inhibition of TYMS activity by the quinazoline and quinaline derivatives **I**–**IV** significantly depends on the nature of the substituents R_1 and R_2 in positions 2, 4 and 5 of the in bicyclic rings, acyclic linkers Y, aromatic fragments Ar and terminal R''_1 and R''_2 substituents bound to the asymmetric carbon atom (Figs. 4–6).

The analysis of the structures of TYMS inhibitors I (Fig. 5) reveals, that the derivatives quinazolin-4-one with an unsubstituted methylene linker at position Y, except for the compound 2 (IC_{50} = 24.8 μM), exhibit moderate inhibitory activity in the range of IC_{50} values = 3.78–9.34 μM. Substitution of a hydrogen atom in the secondary amino group in the linker by nonpolar and weakly polar methyl, propyl, propylene, fluoroethyl, hydroxyethyl, hydroxybutyl, cyanomethyl and propargyl group contributes in most cases increasing the activity containing compounds almost an order of magnitude, irrespective of the nature of the aromatic ring Ar (Fig. 4). In the majority of cases, with the exception of compounds containing a 3 ″-fluoroaromatic fragment, the maximum positive effect in structures with a substituted methyleneamine linker is provided by the propargyl group. In contrast, the reduction of the activity of TYMS inhibitors with the general structural formula I results in the replacement of the hydrogen atom in the secondary amino group of the methyleneamine linker with methoxyethyl, methoxypropyl and acetylmethylene moieties.

The replacement of the unsubstituted methyleneamine linker Y with aminomethylene, methyleneoxy, methylenethio and thio fragments also significantly reduces the activity of TYMS inhibitors

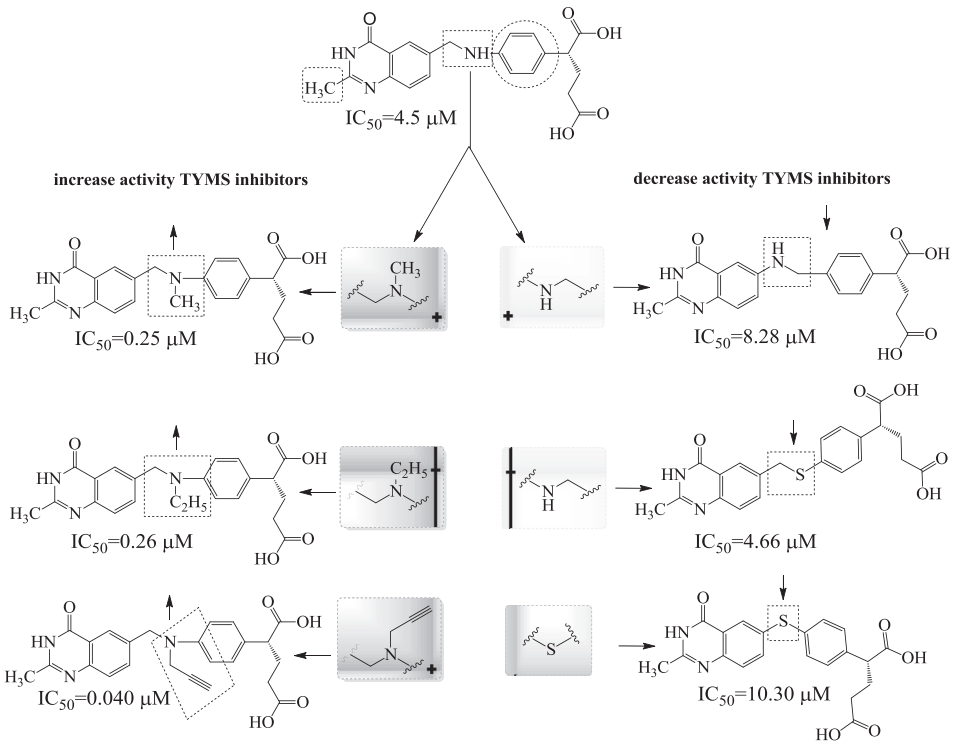

Fig. 4. The effect of the nature of acyclic linkers on the activity of the TYMS inhibitors general formulae **I**. The analyzed structural fragments are indicated by dashed lines. The arrows pointing up and down correspond to the positive and negative effects of the isolated fragment on the inhibitory activity toward TYMS.

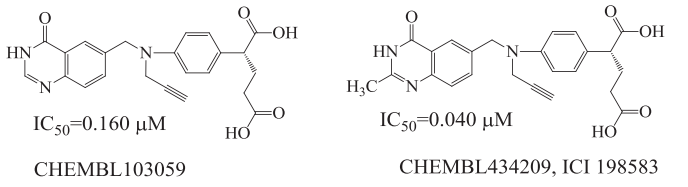

Fig. 5. Comparison compounds in the analysis of the effect of substituents in the position of R_2 in inhibitors of TYMS with the general structural formula I.

of the general structural formula I. It further reduces the activity of this group of TYMS inhibitors by replacing the hydrogen atom in the secondary amino group in the aminomethylene linker Y by on the propargyl group (see Table 5S in Support Material). At the same time, various modifications of the aminomethylene linker contribute to an increase in the target property. In general, the decrease in the activity of the considered TYMS inhibitors **I** with the aminomethylene linker in comparison with their structural analogs with the methyleneamine linker is explained by weakening the delocalization of the electron density of unpaired nitrogen atoms through the aromatic ring Ar.

As follows from the data of Fig. 6, replacing the 1,4-benzene ring with 2,5-substituted thiophene leads to a sharp decrease in the inhibition activity, regardless of the nature of substituent in the methyleneamine linker. The same effect is caused by the replacement of 1,4-benzene with 2,5-substituted thiazols with a nitrogen atom in position 2" and 3″ or 2,5-disubstituted 1,3,4-thiadiazole (see Fig. 4, Table 6S in Supplementary Material). Isosteric replacement of 1,4-benzene with 1,4-pyridine, in which the nitrogen atom is in position 2" in most cases does not affect. The replacement of the hydrogen atom in positions 2" and 3" of the benzene ring has an ambiguous effect on the inhibitory activity. In some cases, the effect of such a modification is determined by nature of the substituent in the methylenamine linker Y ((see Table 6S in Supplementary Material). For example, if the hydrogen atom of the secondary amino group in the linker Y is replaced by a ethyl moiety, the introduction of the chlorine atom into position 2" of the benzene ring increases the target property. Replacing the hydrogen atom by propargyl group in the secondary amino group in the same linker Y leads to the opposite effect (see Supplementary Material). It should be noted that the conclusions about the effect of the nature of the aromatic ring on the activity of TYMS inhibitors, made by us in this work, are in some cases limited to small sets of compounds available in the literature that contain methyl, ethyl and propargyl groups in the linker. However, these data make it possible to trace the following trend: compounds containing five-membered aromatic rings as an Ar substituent have slightly lower inhibitory activity than TYMS inhibitors containing six-membered aromatic fragments in the same position. These observations can be explained in terms of the differences in the location of five- and six-membered rings in comparison with the six-membered rings of *p*-aminobenzoic isosters in the active center of the enzyme. The reason for the decrease in inhibitory activity shown by the quinazoline derivatives in the series with five-membered aromatic fragments is not known precisely and may be related either to a decrease in specific interactions of the inhibitors with the enzyme or a decrease in the total surface of these molecules, which further leads to reducing the contacting surfaces of the binding sites.

Replacing of the hydrogen atom in the R_1 position of the quinazoline ring with a methyl group and a chlorine atom results in a decrease in the activity of TYMS inhibitors of the general structural formula I (see Table 7S in Supplementary Material).

When analyzing the effect of substituents in the R_2 position of the quinazoline ring, the following explanation should be made. According to generally accepted standards, a compound containing an unsubstituted hydrogen atom in the R_2 position should be taken as the standard. However, for the predominant majority of

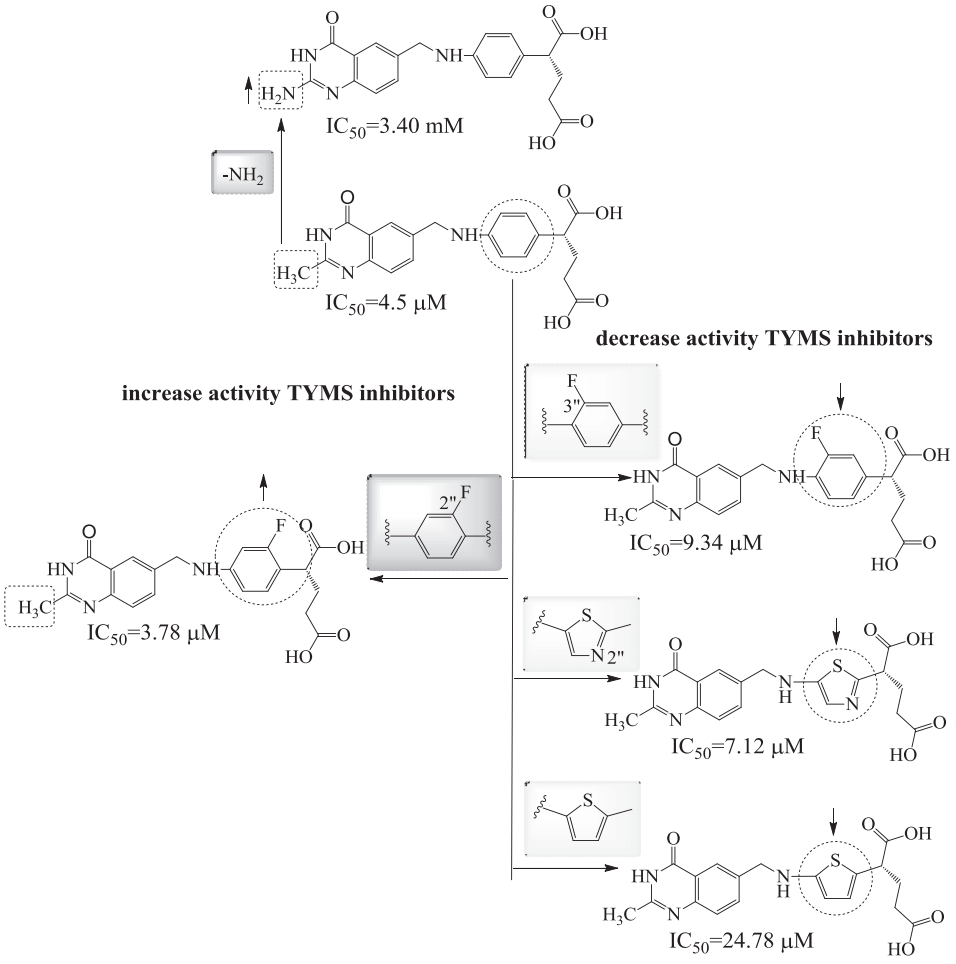

Fig. 6. The effect of the nature of aromatic fragments and substituents in the R_1 position on the activity of the TYMS inhibitors general formulae **I**. The analyzed structural fragments are indicated by dashed lines. The arrows pointing up and down correspond to the positive and negative effects of the isolated fragment on the inhibitory activity toward TYMS.

compounds with the general structural formula I, such information is not available in the literature. It is presented only for compounds containing a propargyl group in the linker. In this regard, for objectivity in the structural analysis of propargyl-containing TYMS inhibitors, we introduced a system of double standards. It consists in simultaneously taking into account two comparison compounds: a compound with the code CHEMBL103059 (contains the hydrogen atom in the position R_2) and a compound with the code CHEMBL434209, in the literature known as ICI 198583 (contains a methyl group in the R_2 position). The final conclusions about the effect of substituents in this position on the activity of propargyl-containing TYMS inhibitors were made based on a comparison of their activity with the compound with the code CHEMBL434209 (Fig. 5). In all other cases, for the reasons mentioned above, the compounds containing the methyl group in the R_2 position were chosen as reference substances.

It was determined that the replacement of the hydrogen atom at the R_2 position of the same ring with methyl and amino groups in compounds with the general structural formula I results in an increase in the inhibitory activity. An analogous replacement of the hydrogen atom by more voluminous isopropyl, fluoromethyl, trifluoroethyl and aromatic fragments leads to a decrease in the activity of TYMS inhibitors of the group under discussion. (see Fig. 6 and Table 8S in Supplementary Material).

The activity of TYMS inhibitors **II** is significantly influenced by the nature of the substituents at the positions R''_1 and R''_2 and their stereoisomerism. Any replacement of the hydrogen atoms in these positions by acyclic or heterocyclic aromatic fragments increases the target property (see Tables 9S, 10S and 11S in Supplementary Material). But in compounds of this series, which contain as a terminal fragment a benzene moiety with a nitro group in the meta-position, any replacement of the methyl group at position R_2 by more bulky acyclic and cyclic fragments leads to a decrease in inhibitory activity against TYMS. The most pronounced negative effect is observed when a methyleneamine group is introduced into this position (see Table 12S in Supplementary Material).

Modification of the quinazoline fragment in position R_1 by methyl group and fluorine atomin in compounds **III** has a positive effect on the target activity. The effect of the nature of the cyclic substituents in the R_1 position of in the compounds of this group has not been experimentally studied. Due to the insufficient experimental data, the clear effect of the substituent polarity on the activity of the TYMS inhibitors **III** could not be identified (see Table 13S in Supplementary Material).

In the quinoline derivatives with the general structural formula **IV** the modification of the position of R_1 ($R_2 = CH_3$) by small acyclic electron-donating and electron-withdrawing substituents, with the exception of the thiomethylene group, increases their inhibitory activity against TYMS. For this series of compounds, the influence of bulky cyclic substituents has not been experimentally studied. But a similar modification of the R_2 position by electron-donor and electron-withdrawing substituents, with the exception of the trifluoromethyl and amino groups, as well as the chlorine atom, decreases the target activity. The greatest negative effect is

observed when a carboxyl group is introduced into the R_2 position (see Tables 14S and 15S in Supplementary Material).

4. Conclusion

Using the QSAR methodology implemented in the GUSAR 2013 program, the structure–inhibitory activity toward thymidylate synthase (TYMS) of 303 methylquinazolin-4-one and quinazolin-4-imine derivatives was quantified.

The inhibitory activity of the simulated compounds lies in the range of $IC_{50} = 0.4 \div 380000.0$ nmoL/l. Based on the MNA and QNA descriptors using the self-consistent regression method, six statistically significant QSAR consensus models have been designed, These models are highly accurate in predicting the IC_{50} values of the compounds from the training and test sets: $R^2_{TRi} > 0.6$; $Q^2_{TRi} > 0.5$; and $R^2_{TSi} > 0.5$. All of them can be used for virtual screening to find new selective antifolate TYMS inhibitors among the quinazolin-4-one derivatives.

We have shown that the algorithms underlying the construction of the QSAR models in the GUSAR 2013 program allow simulating biological activity of condensed heterocyclic organic compounds with 10 and a large number of free-rotating single bonds although the program does not utilize the ideology of 3D-QSAR. It is known that not all the molecular docking algorithms allow obtaining objective and reliable results for bulk molecules with a large number of single bonds.

Thus, the approach used in the GUSAR 2013 program allows modeling the inhibitory activity of the quinazoline-4-on and quinazolin-4-imine derivatives toward TYMS with a high reliability to search for new antifolate inhibitors of this enzyme. Based on the structural analysis, which is the basis of this program, we have identified the structural descriptors modulating the activity of TYMS inhibitors. The information on the structural contribution of various substituents to the target property can be taken into account in the development of new TYMS inhibitors. High predicted activity of a number of compounds with known antitumor activity suggests the mechanism of their antitumor effect.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jmgm.2018.09.002.

References

[1] N. Touroutoglou, R. Pazdur, Thymidylate synthase inhibitors, Clin. Canc. Res. 2 (1996) 227–243. PMID: 9816165.

[2] D.A. Matthews, J.E. Villafranca, C.A. Janson, W.W. Smith, K. Welsh, S. Freer, Stereochemical mechanism of action for thymidylate synthase based on the X-ray structure of the covalent inhibitory ternary complex with 5-fluoro-2'-deoxyuridylate and 5,10-methylenetetrahydrofolate, J. Mol. Biol. 214 (1990) 937–948. https://doi.org/10.1016/0022-2836(90)90347-O.

[3] A. Kamb, J.S. Finer-Moore, R. Stroud, Cofactor triggers the conformational change in thymidylate synthase: implications for an ordered binding mechanism, Biochemistry 31 (1992) 12876–12884. PMID: 1281428.

[4] A.L.J. Pogolotti, P.V. Danenberg, D.V. Santi, Kinetics and mechanism of interaction of 10-propargyl-5,8-dideazafolate with thymidylate synthase, J. Med. Chem. 29 (1986) 478–482, https://doi.org/10.1021/jm00154a009.

[5] D.V. Santi, C.S. McHenry, R.T. Raines, K.M. Ivanetich, Kinetics and thermodynamics of the interaction of 5-fluoro-2'-deoxyuridylate with thymidylate synthase, Biochemistry 26 (1987) 8606–8613. PMID: 3126806.

[6] J. Liu, J.C. Schmitz, X. Lin, N. Tai, W. Yan, M. Farrell, M. Bailly, T.-M. Chen, E. Chu, Thymidylate synthase as a translational regulator of cellular gene expression, Biochim. Biophys. Acta 1587 (2002) 174–182. https://doi.org/10.1016/S0925-4439(02)00080-7.

[7] C.W. Carreras, D.V. Santi, The catalytic mechanism and structure of thymidylate synthase, Annu. Rev. Biochem. 64 (1995) 721–762, https://doi.org/10.1146/annurev.bi.64.070195.003445.

[8] M.P. Costi, S. Ferrari, A. Venturelli, S. Calò, D. Tondi, D. Barlocco, Thymidylate synthase structure, function and implication in drug discovery, Curr. Med. Chem. 12 (2005) 2241–2258. PMID: 16178783.

[9] L. Rahman, D. Voeller, M. Rahman, S. Lipkowitz, C. Allegra, J.C. Barrett, F.J. Kaye, M. Zajac-Kaye, Thymidylate synthase as an oncogene: a novel role for an essential DNA synthesis enzyme, Canc. Cell 5 (2004) 341–351. PMID: 15093541.

[10] D. Edler, M. Hallström, P.G. Johnston, I. Magnusson, P. Ragnhammar, H. Blomgren, Thymidylate synthase expression: an independent prognostic factor for local recurrence, distant metastasis, disease-free and overall survival in rectal cancer, Clin. cancer res. 6 (2000) 1378–1384. PMID: 10778966.

[11] K. Kubota, S. Niho, S. Enatsu, Y. Nambu, Y. Nishiwaki, N. Saijo, M. Fukuoka, Efficacy differences of pemetrexed by histology in pretreated patients with stage IIIB/IV non-small cell lung cancer: review of results from an open-label randomized phase II study, J. Thorac. Oncol. 4 (2009) 1530–1536, https://doi.org/10.1097/JTO.0b013e3181b9e608.

[12] S.L. Showalter, T.N. Showalter, A. Witkiewicz, R. Havens, E.P. Kennedy, T. Hucl, S.E. Kern, C.J. Yeo, J.R. Brody, Evaluating the drug-target relationship between thymidylate synthase expression and tumor response to 5-fluorouracil. Is it time to move forward, Canc. Biol. Ther. 7 (2008) 986–994. PMID: 18443433.

[13] K. Takezawa, I. Okamoto, S. Tsukioka, J. Uchida, M. Kiniwa, M. Fukuoka, K. Nakagawa, Identification of thymidylate synthase as a potential therapeutic target for lung cancer, Br. J. Canc. 103 (2010) 354–361, https://doi.org/10.1038/sj.bjc.6605793.

[14] H. Ozasa, T. Oguri, T. Uemura, M. Miyazaki, K. Maeno, S. Sato, R. Ueda, Significance of thymidylate synthase for resistance to pemetrexed in lung cancer, Canc. Sci. 10 (2009) 161–166, https://doi.org/10.1111/j.1349-7006.2009.01358.x.

[15] T. Oguri, H. Achiwa, Y. Bessho, H. Muramatsu, H. Maeda, T. Niimi, S. Sato, R. Ueda, The role of thymidylate synthase and dihydropyrimidine dehydrogenase in resistance to 5-fluorouracil in human lung cancer cells, Lung Canc. 49 (2005) 345–351, https://doi.org/10.1016/j.lungcan.2005.05.003.

[16] J.J. McGuire, Anticancer antifolates: current status and future directions, Curr. Pharmaceut. Des. 9 (2003) 2593–2613. PMID:14529544.

[17] N.L. Lehman, Future potential of thymidylate synthase inhibitors in cancer therapy, Expet Opin. Invest. Drugs 11 (2002) 1775–1787, https://doi.org/10.1517/13543784.11.12.1775.

[18] N. Tzioumaki, S. Manta, E. Tsoukala, J.V. Voorde, S. Liekens, D. Komiotis, J. Balzarini, Synthesis and biological evaluation of unsaturated keto and exomethylene D- arabinopyranonucleoside analogs: novel 5-fluorouracil analogs that target thymidylate synthase, Eur. J. Med. Chem. 46 (2011) 993–1005, https://doi.org/10.1016/j.ejmech.2011.01.005.

[19] S. Guariento, M. Tonelli, S. Espinoza, A.S. Gerasimov, R.R. Gainetdinov, E. Cichero, Rational design, chemical synthesis and biological evaluation of novel biguanides exploring species-specificity responsiveness of TAAR1 agonists, Eur. J. Med. Chem. 146 (2018) 171–184, https://doi.org/10.1016/j.ejmech.2018.01.059.

[20] N. Liessi, E. Cichero, E. Pesce, M. Arkel, A. Salis, V. Tomati, M. Paccagnella, G. Damonte, B. Tasso, L.J.V. Galietta, N. Pedemonte, P. Fossa, E. Millo, Synthesis and biological evaluation of novel thiazole- VX-809 hybrid derivatives as F508del correctors by QSAR-based filtering tools, Eur. J. Med. Chem. 144 (2018) 179–200, https://doi.org/10.1016/j.ejmech.2017.12.030.

[21] E. Cichero, S. Cesarini, A. Spallarossa, L. Mosti, P. Fossa, Acylthiocarbamates as non-nucleoside HIV-1 reverse transcriptase inhibitors: docking studies and ligand-based CoMFA and CoMSIA analyses, J. Mol. Model. 15 (2009) 871–884, https://doi.org/10.1007/s00894-008-0441-6.

[22] S. Guariento, O. Bruno, P. Fossa, E. Cichero, New insights into PDE4B inhibitor selectivity: CoMFA analyses and molecular docking studies, Mol. Divers. 20 (2016) 77–92, https://doi.org/10.1007/s11030-015-9631-1.

[23] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, Ch Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR Modeling: where have you been? Where are you going to? J. Med. Chem. 57 (2014) 4977–5010, https://doi.org/10.1021/jm4004285.

[24] V.R. Khairullina, A.Ja Gerchikov, A.A. Lagunin, F.S. Zarudii, QSAR modeling of thymidylate synthase inhibitors in a series of quinazoline derivatives, Pharm. Chem. J. 51 (2018) 884–888, https://doi.org/10.1007/s11094-018-1710-z.

[25] L. Guasch, A.V. Zakharov, O.A. Tarasova, V.V. Poroikov, Ch Liao, M.C. Nicklaus, Novel HIV-1 integrase inhibitor development by virtual screening based on QSAR models, Curr. Top. Med. Chem. 16 (2016) 441–448, https://doi.org/10.2174/1568026615666150813150433.

[26] V.H. Masand, D.T. Mahajan, K.N. Patil, T.B. Hadda, M.H. Youssoufi, R.D. Jawarkar, I.G. Shibi, Optimization of antimalarial activity of synthetic prodiginines: QSAR, GUSAR, and CoMFA analyses, Chem. Biol. Drug Des. 81 (2013) 527–536, https://doi.org/10.1111/cbdd.12099.

[27] A.V. Zakharov, A.A. Lagunin, D.A. Filimonov, V.V. Poroikov, Quantitative prediction of antitarget interaction profiles for chemical compounds, Chem. Res. Toxicol. 25 (2012) 2378–2385, https://doi.org/10.1021/tx300247r.

[28] D.A. Filimonov, A.V. Zakharov, A.A. Lagunin, V.V. Poroikov, QNA-based "star track" QSAR approach, SAR QSAR Environ. Res. 20 (2009) 679–709. https://doi.org/10.1080/10629360903438370.

[29] V.R. Khayrullina, A.Ya Gerchikov, A.A. Lagunin, F.S. Zarudii, Quantitative analysis of structure-activity relationships of tetrahydro-2H-isoindole cyclooxygenase-2 inhibitors, Biochemistry 80 (2015) 74–86, https://doi.org/10.1134/S0006297915010095.

[30] V.R. Khairullina, Y.Z. Akbasheva, A.R. Gimadieva, A.G. Mustafin, Analysis of the relationship «structure-activity» in the series of certain 5-ethyluridine derivatives with pronounced anti-herpetic activity, Vestn. Bashk. Univ. 22

(2017) 960–963.
[31] A.V. Zakharov, E.V. Varlamova, A.A. Lagunin, A.V. Dmitriev, E.N. Muratov, D. Fourches, V.E. Kuz'min, V.V. Poroikov, A. Tropsha, M.C. Nicklaus, QSAR modeling and prediction of drug-drug interactions, Mol. Pharm. 13 (2016) 545–556, https://doi.org/10.1021/acs.molpharmaceut.5b00762.
[32] A.V. Zakharov, M.L. Peach, M. Sitzmann, M.C. Nicklaus, A new approach to radial basis function approximation and its application to QSAR, J. Chem. Inf. Model. 54 (2014) 713–719, https://doi.org/10.1021/ci400704f.
[33] A.V. Zakharov, M.L. Peach, M. Sitzmannand, M.C. Nicklaus, QSAR modeling of imbalanced high-throughput screening data in PubChem, J. Chem. Inf. Model. 54 (2014) 705–712, https://doi.org/10.1021/ci400737s.
[34] A. Lagunin, A. Zakharov, D. Filimonov, V. Poroikov, QSAR modelling of rat acute toxicity on the basis of PASS prediction, Mol. Interv. 30 (2011) 241–250, https://doi.org/10.1002/minf.201000151.
[35] G.V. Kokurkina, M.D. Dutov, S.A. Shevelev, S.V. Popkov, A.V. Zakharov, V.V. Poroikov, Synthesis, antifungal activity and QSAR study of 2-arylhydroxynitroindoles, Eur. J. Med. Chem. 46 (2011) 4374–4382, https://doi.org/10.1016/j.ejmech.2011.07.008.
[36] A.V. Zakharov, M.L. Peach, M. Sitzmann, I.V. Filippov, H.J. McCartney, L.H. Smith, A. Pugliese, M.C. Nicklaus, Computational tools and resources for metabolism-related property predictions. 2. Application to prediction of half-life time in human liver microsomes, Future Med. Chem. 4 (2012) 1933–1944, https://doi.org/10.4155/fmc.12.152.
[37] P.R. Marsham, A.L. Jackman, J. Oldfield, L.R. Hughes, T.J. Thornton, G.M. Bisset, B.M. O'Connor, J.A. Bishop, A.H. Calvert, Quinazoline antifolate thymidylate synthase inhibitors: benzoyl ring modifications in the C2-methyl series, J. Med. Chem. 33 (1990) 3072–3078, https://doi.org/10.1021/jm00173a026. PMID: 2231608.
[38] P.R. Marsham, L.R. Hughes, A.L. Jackman, A.J. Hayter, J. Oldfield, J.M. Wardleworth, J.A. Bishop, B.M. O'Connor, A.H. Calvert, Quinazoline antifolate thymidylate synthase inhibitors: heterocyclic benzoyl ring modifications, J. Med. Chem. 34 (1991) 1594–1605, https://doi.org/10.1021/jm00109a011.
[39] P.R. Marsham, A.L. Jackman, A.J. Hayter, M.R. Daw, J.L. Snowden, B.M. O'Connor, J.A. Bishop, A.H. Calvert, L.R. Hughes, Quinazoline antifolate thymidylate synthase inhibitors: bridge modifications and conformationally restricted analogues in the C2-methyl series, J. Med. Chem. 34 (1991) 2209–2218, https://doi.org/10.1021/jm00111a042.
[40] P.R. Marsham, A.L. Jackman, A.J. Barker, F.T. Boyle, S.J. Pegg, J.M. Wardleworth, R. Kimbell, B.M. O'Connor, A.H. Calvert, L.R. Hughes, Quinazoline antifolate thymidylate synthase inhibitors: replacement of glutamic acid in the C2-methyl series, J. Med. Chem. 38 (1995) 994–1004, https://doi.org/10.1021/jm00006a019.
[41] P.R. Marsham, J.M. Wardleworth, F.T. Boyle, L.F. Hennequin R. Kimbell, M. Brown, A.L. Jackman, Design and synthesis of potent non-polyglutamatable quinazoline antifolate thymidylate synthase inhibitors, J. Med. Chem. 42 (1999) 3809–3820. PMID: 10508430.
[42] L.F. Hennequin, F.T. Boyle, J.M. Wardleworth, P.R. Marsham, R. Kimbell, A.L. Jackman, Quinazoline antifolates thymidylate synthase inhibitors: lipophilic analogues with modification to the C2-methyl substituent, J. Med. Chem. 39 (1996) 695–704, https://doi.org/10.1021/jm950645n.
[43] L.R. Hughes, A.L. Jackman, J. Oldfield, R.C. Smith, K.D. Burrows, P.R. Marsham, J.A. Bishop, T.R. Jones, B.M. O'Connor, A.H. Calvert, Quinazoline antifolate thymidylate synthase inhibitors: alkyl, substituted alkyl, and aryl substituents in the C2 position, J. Med. Chem. 33 (1990) 3060–3067, https://doi.org/10.1021/jm00173a024.
[44] J.B. Hynes, S.A. Patil, A. Tomazic, A. Kumar, A. Pathak, X. Tan, L. Xianqiang, M. Ratnam, T.J. Delcamp, J.H. Freisheim, Inhibition of murine thymidylate synthase and human dihydrofolate reductase by 5,8-dideaza analogues of folic acid and aminopterin, J. Med. Chem. 31 (1988) 449–454, https://doi.org/10.1021/jm00397a031.
[45] V. Bavestias, A.L. Jackman, J.H. Marriott, R. Kimbell, W. Gibson, F.T. Boyle, G.M. Bisset, Folate-based inhibitors of thymidylate synthase: synthesis and antitumor activity of γ-linked sterically hindered dipeptide analogues of 2-desamino-2-methyl-N^{10}-propargyl-5,8-dideazafolic acid (ICI 198583), J. Med. Chem. 40 (1997) 1495–1510, https://doi.org/10.1021/jm960878u.
[46] P. Warner, A.J. Barker, A.L. Jackman, K.D. Burrows, N. Roberts, J.A. Bishop, B.M. O'Connor, L.R. Hughes, Quinazoline antifolate thymidylate synthase inhibitors: variation of the C2- and C4-substituents, J. Med. Chem. 35 (1992) 2761–2768, https://doi.org/10.1021/jm00093a007.
[47] G.M. Bisset, V. Bavetsias, T.J. Thornton, K. Pawelczak, A.H. Calvert, L.R. Hughes, A.L. Jackman, The synthesis and thymidylate synthase inhibitory activity of L-γ-linked dipeptide and L-γ-amide analogues of 2-desamino-2-methyl-N^{10}-propargyl-5,8-dideazafolic acid (ICI 198583), J. Med. Chem. 37 (1994) 3294–3302, https://doi.org/10.1021/jm00046a014.
[48] MarvinSketch, version 17.22.0, ChemAxon, Ltd., Budapest, Hungary, 2017. https://chemaxon.com/download/marvin-suite.
[49] Discovery Studio Visualiser, Accelrys GmbH, Cologne, Germany, 2017. http://accelrys.com/products/collaborative-science/biovia-discovery-studio/visualization-download.php.
[50] D.A. Filimonov, D.V. Akimov, V.V. Poroikov, The method of self-consistent regression for the quantitative analysis of relationships between structure and properties of chemicals, Pharm. Chem. J. 38 (2004) 21–24. https://doi.org/10.1023/B:PHAC.0000027639.17115.5d.
[51] P. Gramatica, A. Sangion, A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology, J. Chem. Inf. Model. 56 (2016) 1127–1131, https://doi.org/10.1021/acs.jcim.6b00088.
[52] V. Consonni, D. Ballabio, R. Todeschini, Evaluation of model predictive ability by external validation techniques, J. Chemometr. 24 (2010) 194–201. https://doi.org/10.1002/cem.1290.
[53] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance Correlation coefficient, J. Chem. Inf. Model. 51 (2011) 2320–2335, https://doi.org/10.1021/ci200211n.
[54] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative studies on some metrics for external validation of QSPR models, J. Chem. Inf. Model. 52 (2012) 396–408, https://doi.org/10.1021/ci200520g.
[55] P.P. Roy, S. Paul, I. Mitra, K. Roy, On two novel parameters for validation of predictive QSAR models, Molecules 14 (2009) 1660–1701, https://doi.org/10.3390/molecules14051660.
[56] K. Roy, R.N. Das, P. Ambure, R.B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, Chemometr. Intell. Lab. Syst. 152 (2016) 18–33. https://doi.org/10.10.16/j.chemolab.2016.01.008.
[57] Xternal Validation Plus, version 1.2, 2016, DTC Lab., Kolkata, India. URL: https://sites.google.com/site/dtclabxvplus/.