DOI 10.1007/s11094-018-1710-z
Pharmaceutical Chemistry Journal, Vol. 51, No. 10, January, 2018 (Russian Original Vol. 51, No. 10, October, 2017)

QSAR MODELLING OF THYMIDYLATE SYNTHASE INHIBITORS IN A SERIES OF QUINAZOLINE DERIVATIVES

V. R. Khairullina,[1] A. Ya. Gerchikov,[1] A. A. Lagunin,[2, 3] and F. S. Zarudii[4]

Translated from Khimiko-Farmatsevticheskii Zhurnal, Vol. 51, No. 10, pp. 33 – 37, October, 2017.

Original article submitted January 15, 2016.

Thymidylate synthase (ThS) is a target for antimetabolite antitumor drugs. Such drugs have been used in the clinic although they cause several severe side effects and accumulate in tissues. Therefore, new less toxic ThS inhibitors must be sought and created. The GUSAR 2013 program was used to study the quantitative structure – activity relationship (QSAR) of a series of antifolate ThS inhibitors in the IC_{50} range 0.52 – 24,800.00 nM. Statistically significant QSAR models were constructed using MNA- and QNA-descriptors and self-consistent regression. They typically predicted highly accurately the structures of the training and test sets (R^2_{train}: 0.855 – 0.922; R^3_{train}: 0.810 – 0.895; R^2_{test1}: 0.734 – 0.790; R^2_{test2}: 0.800 – 0.835).

Keywords: antifolate thymidylate synthase inhibitors, QSAR, GUSAR 2013, QNA- and MNA-descriptors, structure—activity relationship analysis.

Thymidylate synthase (ThS, EC 2.1.1.45) is a bisubstrate enzyme in which deoxythymidine monophosphate, a nucleotide required for DNA synthesis, is synthesized [1 – 4]. ThS activity was elevated in tumor cells because of their high growth and development rates [3, 4]. Therefore, this enzyme is a target for antimetabolite antitumor drugs. However, antitumor drugs that are currently used in medical practice reduce ThS activity directly or indirectly, e.g., methotrexate and raltitrexed, cause several severe side effects. Furthermore, these drugs and their analogs accumulate in tissues, which further enhances their toxic properties [5 – 7]. Therefore, the search for biologically active compounds that can inhibit ThS activity and; therefore, slow DNA biosynthesis in tumor cells, is an important practical problem for medicinal chemistry aimed at the development of efficacious drugs. Solving the problem using exclusively empirical analysis of biological data without invoking computational chemistry methods is a difficult task that requires significant time and material expenses [8]. Also, virtual screening methods based on analysis of (quantitative) structure—activity relationships [(Q)SARs] [9 – 12] can be employed in preclinical tests to select lead compounds with a given activity profile from libraries and databases for *in vivo* biological tests [11, 12]. This approach could shorten considerably the time and material expenses for seeking and developing potential antifolate-type ThS inhibitors.

Thus, the goal of the present work was to construct and validate QSAR models for selective ThS inhibitors among quinazoline derivatives with general structural formulas **I** – **VI** (Fig. 1) based on two-dimensional representations of their structural formulas. These compounds have structures that are highly similar to ThS coenzyme, 5,10-methylenetetrahydrofolate, so that they are promising for development of ThS inhibitors based on them.

EXPERIMENTAL PART

The QSAR analysis for the ThS inhibitors used the General Unrestricted Structure Activity Relationships (GUSAR 2013) computer program [9 – 11]. QSAR models were constructed in several steps using the method given below and described in detail before [13, 14].

[1] Department of Chemistry, Bashkir State University, 32 Zaki Validi St., Ufa, Bashkortostan, 450076, Russia; fax: +7 (347) 229-9707; e-mail: gerchikov@inbox.ru, Veronika1979@yandex.ru
[2] N. I. Pirogov Russian National Research Medical University, 1 Ostrovityanova St., Moscow, 117997, Russia; fax: +7 (495) 434-1422.
[3] V. N. Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, 10/8 Pogodinskaya St., Moscow, 119121, Russia; fax: +7 (499) 245-0857; e-mail: alexey.lagunin@ibmc.msk.ru
[4] Bashkir State Medical University, 3 Lenina St., Ufa, Bashkortostan, 450077, Russia; fax: +7 (347) 272-3751; e-mail: zarudii.f@yandex.ru

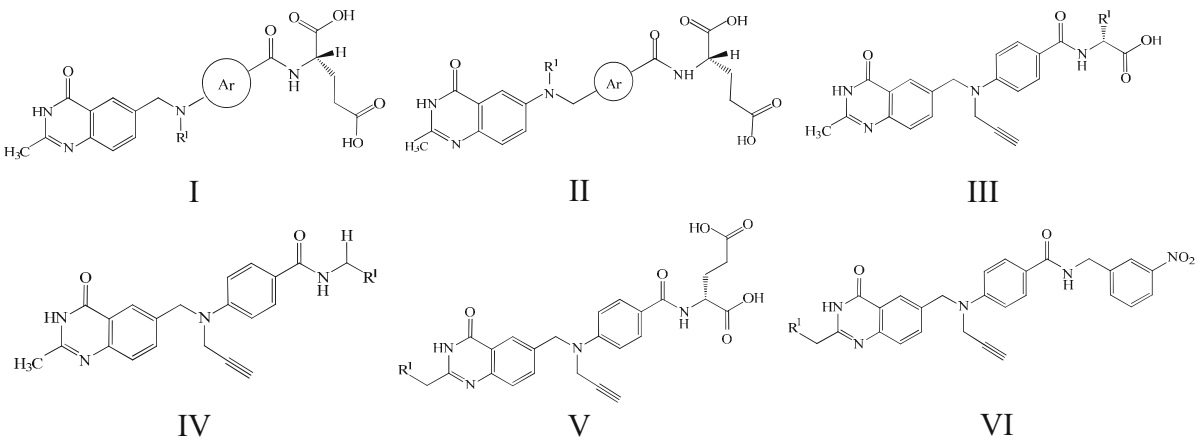


Fig. 1. General structural formulas of modeled antifolate ThS inhibitors.

Training (TrS1-TrS3) and test sets (TeS1-TeS2) for constructing the QSAR models (M1-M9) and validating them were formulated based on literature data [15 – 20] presented in sets S1-S3 according to the diagram shown in Fig. 2.

Training set TrS1 was intended for constructing QSAR models M1-M3, was formulated based on set S1, and included 196 antifolate-type ThS inhibitor structures. The inhibitory activities IC_{50} of these compounds were evaluated experimentally before [15 – 20]. Sets S2 and S3 were produced by breaking down set S1 in a 2:1 ratio after preliminary ranking by increasing IC_{50} values. Then, training set TrS2 and test set TeS1 were formulated based on sets S2 and S3. Set S3 contained 66 compounds covering a broad activity range. This enabled it to be used as training set TrS3 for formulating models M7 – M9. In this instance, structures of set S2 were selected as test set TeS2 for validating models M7 – M9. Also, the predictive abilities of models M1 – M3 were evaluated using structures of the external test set TeS3, which contained 16 quinazoline derivatives with structures similar to those of sets S1 – S3. The inhibitory activity of compounds in TeS3 were previously studied experimentally under the same conditions as for compounds in TrS1 – TrS3 [21]. QSAR models M1-M9 were constructed by transforming IC_{50} data in sets S1 – S3 (in mol/L) into pIC_{50} values using the formula:

$$pIC_{50} = -\log10(IC_{50}).$$

The GUSAR 2013 program was used to create the QSAR models so that they could be used further for quantitative predictions of the inhibitory activity of the quinazoline derivatives for ThS. The ideology for constructing QSAR models using this program was discussed before in detail [11 – 14]. Two types of atom-centered descriptors, i.e., substructural multilevel neighborhoods of atoms (MNA) and electro-topological quantitative neighborhoods of atoms (QNA) were used to describe the compound structures and construct the QSAR models [9 – 14]. These types of descriptors were computed automatically from chemical structural formulas represented in 2D-format considering the valence and partial charges on their atoms. Specifics of the bond types with respect to stereochemistry were not taken into account. Three

TABLE 1. Statistical Characteristics and Evaluation of Accuracy of Predicted pIC_{50} Values for ThS Inhibitors by Consensus Models M1-M9

Training set	Model	N	R^2_{TrS}	Q^2_{TrS}	F	S. D.	V	R^2_{TeSi}
QSAR models based on QNA descriptors								
TrS1	M1	196	0.895	0.869	65.386	0.363	21	-
TrS2	M4	130	0.857	0.821	42.917	0.420	14	0.800[a]
TrS3	M7	66	0.867	0.818	30.872	0.420	10	0.734[b]
QSAR models based on MNA descriptors								
TrS1	M2	196	0.896	0.872	54.243	0.364	24	-
TrS2	M5	130	0.867	0.832	37.718	0.408	16	0.812[a]
TrS3	M8	66	0.855	0.810	26.404	0.449	9	0.755[b]
QSAR models based on QNA and MNA descriptors								
TrS1	M3	196	0.917	0.895	65.352	0.325	25	-
TrS2	M6	130	0.895	0.868	47.100	0.364	17	0.835[a]
TrS3	M9	66	0.922	0.891	47.501	0.329	11	0.790[b]

Note. N, number of structures in training set; R^2_{TrS}, determination coefficient calculated for training set compounds; R^2_{TeS}, determination coefficient calculated for test set compounds; Q^2_{TrS}, correlation coefficient calculated for the training set with leave-one-out cross validation; *F*, Fisher criterion; S. D., standard deviation; *V*, number of variables in the final regression equation; hyphen (-), no data; [a] predicted pIC_{50} values for ThS inhibitors in TeS1; [b] predicted pIC_{50} values for ThS inhibitors in TeS2.

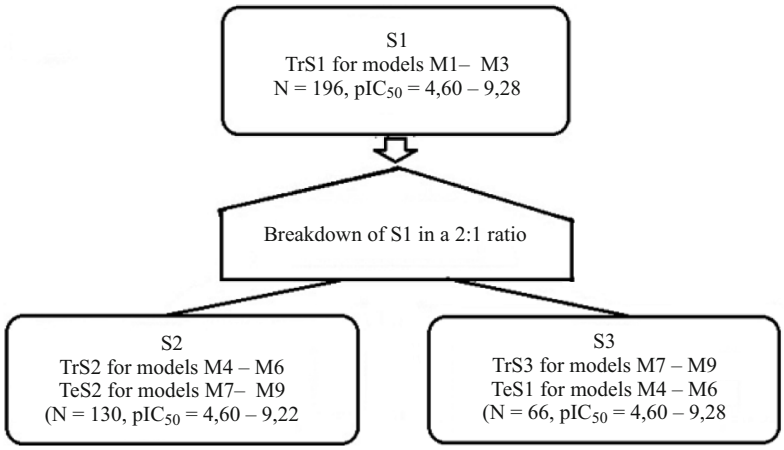


Fig. 2. Diagram of training and test sets for constructing QSAR models M1 – M9, where TrS is a training set; TeS, test set; N, number of compounds in sets S1 – S3, training sets TrS1 – TrS3, and test sets TeS1 – TeS2.

versions for constructing QSAR consensus models were used in the present investigation, i.e., 1) a combination of the whole set of regression equations constructed from QNA descriptors; 2) a combination of the whole set of regression equations constructed from MNA descriptors; and 3) a combination of all regression equations constructed from both QNA and NMA descriptors. Self-consistent regression was used as the mathematical method for determining the optimum set of descriptors and constructing the QSAR models [9-11]. The program allowed QSAR models to be created automatically from single regression equations and sets of regression equations combined into a single consensus model. All 360 partial regression equations (180 each for each descriptor type) were considered for constructing the consensus models. The accuracy of the constructed QSAR models was evaluated from predicted IC_{50} values for structures of TrS1-TrS3 and TeS1-TeS3. Leave-20%-out cross validation of the training set was used for internal validation. The final predicted activity for an actual compound was found by averaging the predicted IC_{50} values of separate QSAR models in the consensus model. This reduced the variability of the results. Also, consensus model M3 with 196 structures of quinazoline derivatives was used to evaluate the contributions of atoms to the activity of the antifolate-type ThS inhibitors. This procedure was performed automatically in the GUSAR 2013 program during construction of the QSAR models from QNA descriptors [13, 14].

RESULTS AND DISCUSSION

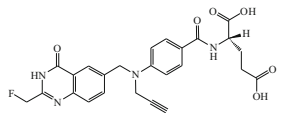
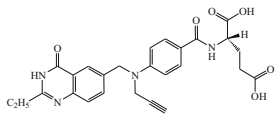
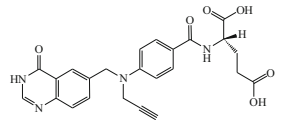
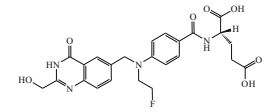
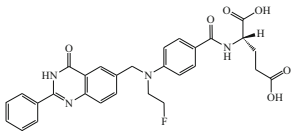
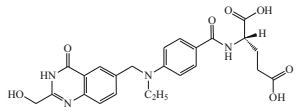
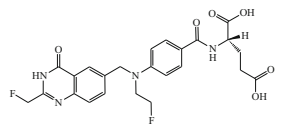
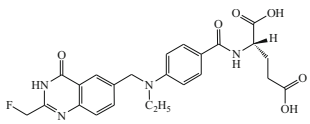
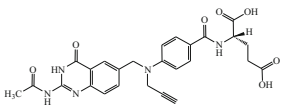
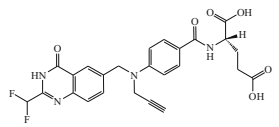
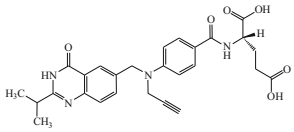
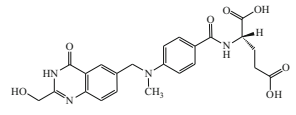
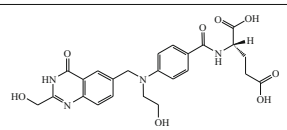
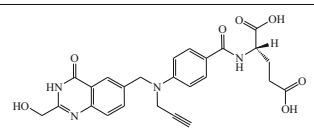
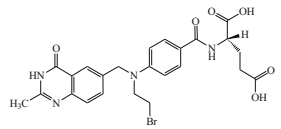
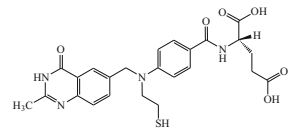
The QSAR of quinazoline derivatives with general structural formulas **I** – **VI** as inhibitors of male white-mouse ThS was modeled using the consensus approach embedded in the GUSAR 2013 program (Fig. 1). Three consensus models were obtained for each training set depending on the type of descriptors (MNA and QNA) used in the calculations. Table 1 lists the statistical parameters of these models and the accuracy of the predicted pIC_{50} values for the ThS inhibitors included in the training and test sets.

The results in Table 1 led to the conclusion that all constructed QSAR consensus models had acceptable predictive ability for structures of training sets TrS1-TrS3 ($R^2 > 0.8$, $Q^2 > 0.7$) and internal test sets TeS1 and TeS2 ($R^2_{TeS} > 0.7$) in the range $\Delta pIC_{50} = 4.67$ units. QSAR models M6 and M9 had the greatest predictive ability of QSAR models M4-M9 constructed from training sets TrS2-TrS3. This was explained by the fact that the ideology of combining consensus models constructed using either QNA or MNA descriptors into a single QSAR model in several instances, e.g., for models M3, M6, and M9, reduced the variability of the predictions of the separate models. Our previous research and that of others were consistent with this conclusion [9, 11, 22]. Furthermore, the reliable prediction of the target property for compounds of test set TeS2 with 130 compounds using models M7-M9 constructed from training set TeS3 with 66 ThS inhibitor structures provided obvious proof that the GUSAR 2013 program could be used correctly to model the QSAR for the ThS inhibitory activity of the quinazoline derivatives.

Next, QSAR models M1 – M3 constructed from the maximum set of ThS inhibitors were used to predict the pIC_{50} values for structures of the external test set TeS3 with structural analogs of quinazoline derivatives with general structural formulas **I** – **VI** (Fig. 1). Table 2 shows that QSAR consensus models M1 – M3 showed rather high predictive ability for structures of external test set TeS3. Discrepancies between the experimental IC_{50} values and those predicted by these models were less than an order of magnitude (RMSE < 0.55). These models could be used for virtual screening of virtual libraries and databases to search for new antifolate-type ThS inhibitors based on quinazoline derivatives.

Structural analysis of quinazoline derivatives with general structural formulas **I** – **VI** was also performed (Fig. 1). The effects of including various structural fragments on the ThS inhibitory activity were analyzed by comparing research results and data obtained using the GUSAR 2013 program. It was found that the results obtained using the GUSAR 2013 program for the SAR of compounds with general structural formulas **I** – **VI** agreed satisfactorily with the experimental data described in detail before [15 – 20]. The exceptions were N-, F-, and Cl-containing compounds and ThS inhibitors with bulky heterocyclic fragments. We supposed that the discrepancy between the structural analyses performed using GUSAR 2013 for halogen- and N-containing ThS inhibitors and the experimental results was due to several factors. For example, structural and steric factors had significant effects on the ThS inhibitor activity in addition to physicochemical factors used in the program to calculate the QNA descriptors and evaluate their contributions to the target property. In particular, the orientation of the aromatic fragments at the enzyme active center and the nature of the acyclic linker, aromatic fragments Ar, and terminal substituents R^1, including those bound to the asymmetric C atom, affected considerably the ThS inhibitor activity (Fig. 1) [15 – 20]. The ability of substituents in the aromatic fragment and other terminal groups to interact at the enzyme active center with nearest amino-acid residues was just as important for the ThS inhibi-

TABLE 2. Predicted pIC_{50} Values for External Test Set TeS3 by QSAR Consensus Models M1 – M3

Structural formula	pIC_{50exp}	pIC_{50pred}		Structural formula	pIC_{50exp}	pIC_{50pred}
	7.00	6.84[a] 6.91[b] 7.08[c]			6.85	7.29[a] 7.29[b] 7.17[c]
	6.80	6.90[a] 6.84[b] 6.72[c]			6.65	6.48[a] 6.67[b] 6.63[c]
	6.65	7.01[a] 6.82[b] 6.78[c]			6.59	6.51[a] 6.57[b] 6.52[c]
	6.47	6.28[a] 6.55[b] 6.62[c]			6.43	6.42[a] 6.69[b] 6.57[c]
	6.32	7.00[a] 7.16[b] 7.07[c]			6.24	7.24[a] 6.83[b] 7.07[c]
	6.21	7.25[a] 7.09[b] 7.00[c]			6.19	6.38[a] 6.29[b] 6.25[c]
	6.11	6.32[a] 6.35[b] 6.29[c]			6.04	6.37[a] 6.36[b] 6.28[c]
	5.89	6.69[a] 6.32[b] 6.60[c]			5.75	6.51[a] 6.29[b] 6.52[c]

[a] Predicted pIC_{50} values by QSAR consensus model M1;
[b] predicted pIC_{50} values by QSAR consensus model M2;
[c] predicted pIC_{50} values by QSAR consensus model M3.

tor activity. Distortion of the planarity of the aromatic fragment by adding bulky functional groups to it was shown to contribute to a reduction of the target property [15 – 20]. All these factors and the optical activity of the modeled molecules were not completely considered in the GUSAR 2013 program. We supposed that these factors had a considerable influence on the activity of the quinazoline ThS inhibitors. However, an important advantage of the GUSAR 2013 program was its ability to consider the influence of the nature of the structural constituents and their bonding mode to each other. Thus, QSAR consensus models constructed from MNA structural descriptors were characterized by rather high accuracy. In turn, this was obvious proof that the nature of the structural fragments contributed significantly to the activity.

The employed approach modeled with high reliability the activity of ThS inhibitors based on quinazoline derivatives in order to develop new antifolate-type inhibitors of this enzyme. Model M3 was preferred for virtual screening because it was constructed from the greatest number of quinazoline derivatives (196) with pronounced inhibitory activity for ThS. However, the rather accurate prediction of IC_{50} values for sets S2 and S3 using models M4-M9 indicated that the target property was modeled well using the approach embedded in the GUSAR 2013 program.

ACKNOWLEDGMENTS

The work was supported financially under a state task and the Basic Research Program of the State Academies of Sciences for 2013 – 2020 (A. A. Lagunin).

REFERENCES

1. J. Liu, J. C. Schmitz, X. Lin, et al., *Biochim. Biophys. Acta*, **1587**(2 – 3), 174 – 182 (2002).
2. P. R. Subbarayan, K. Lee, B. Ardalan, *Anticancer Res.*, **30**(4), 1157 – 1162 (2010).
3. J. A. van der Zee, C. H. J. van Eijck, H. van Dekken, et al., *Eur. J. Surg. Oncol. (EJSO)*, **38**(11), 1058 – 1064 (2012).
4. O. M. H. Salo-Ahen, A. Tochowicz, C. Pozzi, D. Cardinale, et al., *J. Med. Chem.*, **58**(8), 3572 – 3581 (2015).
5. J. Walling, *Invest. New Drugs*, **24**, 37 – 77 (2006).
6. E. Chu, *J. Biol. Chem.*, **265**(15), 8470 – 8478 (1990).
7. V. J. Chen, *Br. J. Cancer*, **78**(3), 27 – 34 (1998).
8. D. A. Filimonov and V. V. Poroikov, *Ross. Khim. Zh.*, **50**(2), 66 – 75 (2006).
9. A. V. Zakharov, A. A. Lagunin, D. A. Filimonov, et al., *Chem. Res. Toxicol.*, **25**(11), 2378 – 2385 (2012).
10. I. A. Taipov, V. R. Khairullina, A. Ya. Gerchikov, et al., *Vestn. Bashkir. Univ.*, **17**(2), 886 – 891 (2012).
11. D. A. Filimonov, A. V. Zakharov, A. A. Lagunin, et al., *SAR QSAR Environ. Res.*, **20**(7 – 8), 679 – 709 (2009).
12. A. V. Zakharov., E. V. Varlamova, A. A. Lagunin, et al., *Mol. Pharm.*, **13**(2), 545 – 556 (2016).
13. V. R. Khairullina, A. Ya. Gerchikov, A. A. Lagunin, et al., *Biokhimiya*, **80**(1), 96 – 110 (2015).
14. V. R. Khairullina, A. Ya. Gerchikov, F. S. Zarudii, et al., *Vestn. Bashk. Univ.*, **19**(2), 417 – 422 (2014).
15. P. J. Marsham, A. L. Jackman, J. Oldfield, et al., *J. Med. Chem.*, **33**(11), 3072 – 3078 (1990).
16. P. J. Marsham, L. R. Hughes, A. L. Jackman, et al., *J. Med. Chem.*, **34**(5), 1594 – 1605 (1991).
17. P. J. Marsham, A. L. Jackman, A. J. Hayter, et al., *J. Med. Chem.*, **34**(7), 2209 – 2218 (1991).
18. P. J. Marsham, A. L. Jackman, A. J. Barker, et al., *J. Med. Chem.*, **38**(6), 994 – 1004 (1995).
19. P. J. Marsham, J. M. Wardleworth, F. T. Boyle, et al., *J. Med. Chem.*, **42**(19), 3809 – 3820 (1999).
20. L. F. Hennequin, F. T. Boyle, J. M. Wardleworth, et al., *J. Med. Chem.*, **39**, 695 – 704 (1996).
21. L. R. Hughes, A. L. Jackman, J. Oldfield, et al., *J. Med. Chem.*, **33**(11), 3060 – 3067 (1990).
22. J. C. Dearden, M. T. Cronin, K. L. Kaiser, *SAR QSAR Environ. Res.*, **20**(3 – 4), 241 – 266 (2009).