MDPI

Article

Quantitative Structure–Activity Relationship in the Series of 5-Ethyluridine, N2-Guanine, and 6-Oxopurine Derivatives with Pronounced Anti-Herpetic Activity

Veronika Khairullina * and Yuliya Martynova

Institute of Chemistry and Defence in Emergency Situations, Ufa University of Science and Technology, 50076 Ufa, Russia; martynovayuz@uust.ru
* Correspondence: khajrullinavr@uust.ru; Tel.: +7-963-906-6567

Abstract: A quantitative analysis of the relationship between the structure and inhibitory activity against the herpes simplex virus thymidine kinase (HSV-TK) was performed for the series of 5-ethyluridine, N2-guanine, and 6-oxopurines derivatives with pronounced anti-herpetic activity (IC_{50} = 0.09 ÷ 160,000 µmol/L) using the GUSAR 2019 software. On the basis of the MNA and QNA descriptors and whole-molecule descriptors using the self-consistent regression, 12 statistically significant consensus models for predicting numerical pIC_{50} values were constructed. These models demonstrated high predictive accuracy for the training and test sets. Molecular fragments of HSV-1 and HSV-2 TK inhibitors that enhance or diminish the anti-herpetic activity are considered. Virtual screening of the ChEMBL database using the developed QSAR models revealed 42 new effective HSV-1 and HSV-2 TK inhibitors. These compounds are promising for further research. The obtained data open up new opportunities for developing novel effective inhibitors of TK.

Keywords: inhibitors of herpes simplex virus thymidine kinase; HSV-1; HSV-2; QSAR models; GUSAR 2019 program; QNA descriptors; MNA descriptors; structure–activity relationships

1. Introduction

Herpes virus infections induced by viruses of the Herpesviridae family are among the most widespread human diseases. Antibodies to various herpes viruses are identified in about 95% of the world's population. Eight types of Herpesviridae viruses cause herpesvirus infections [1–3]. Herpes simplex viruses of the first and second type (HSV-1 and HSV-2) are the most common [4]. HSV-1 usually affects the upper body (mouth, eyes, and brain), whereas HSV-2 relates to genital infections [5]. These viruses are usually latent. However, when immunity is reduced, they are activated, which, in turn, provokes diseases, such as oral herpes, genital herpes, keratitis, conjunctivitis, herpes zoster, etc. It is reported that herpes virus infections induced by HSV-1 and HSV-2 (and other types) increase the possibility to be infected with the human immunodeficiency virus (HIV) and are almost always diagnosed in patients with an HIV infection, which complicates the course of this disease. There is evidence that HSV-1 can participate in the development of multiple sclerosis [6] and lead to male infertility [7].

Currently, there are three classes of drugs in active medical practice for the treatment of infectious diseases caused by different types of herpes viruses, including HSV-1 and HSV-2: (1) acyclic guanosine analogues; (2) acyclic nucleotide analogues; and (3) pyrophosphate analogues (foscarnet) [8].

Acyclovir is known to be an effective inhibitor of viral thymidine kinase (TK). This drug is the gold standard for the prevention and treatment of infections caused by HSV-1 and HSV-2, as this drug combines a pronounced clinical effect and low toxicity [9]. However, a significant disadvantage of acyclovir is its low oral bioavailability, poor solubility, and short blood circulation time. An increase in the therapeutic dose of this drug is undesirable

with long-term use, as it leads to an increase in its toxicity. Another disadvantage of acyclovir is related to drug resistance development. This problem is not significant for patients with good immunity, since the incidence of acyclovir-resistant herpes simplex virus strains among them is ~0.5% of cases. However, among patients with immunodeficiency conditions, it exceeds 30% of cases. In 95% of cases, acyclovir resistance is due to mutations in the thymidine kinase and DNA polymerase genes, which are related to the mechanism of action of this drug [10–15]. In addition to acyclovir, the phenomenon of drug resistance was observed for pentacyclovir and its analogues used as HSV-1- and HSV-2-replication inhibitors. The analogues of adenine, adefovir, and tenofovir have found the greatest application in clinical practice among the phosphonate derivatives of guanosine. These drugs are included in therapy to suppress HSV-1 strains resistant to acyclovir (and its analogues) and some similar cases with the deficiency of viral thymidine kinase. The same drugs are used in the treatment of hepatitis B and HIV. However, these drugs have a pronounced nephro- and hepatotoxic effect. Foscarnet, a covalent inhibitor of DNA polymerase, has not been widely used in clinical practice due to the rather high toxicity and the lack of selectivity. The above-mentioned issues demonstrate the necessity for the search for new anti-herpetic drugs [8].

Additional promising strategies against the HSV-induced herpes infections deal with the development of inhibitors of other enzymes (helicase-primase or ribonucleotide reductase) and inhibitors of the adhesion/penetration of the virus into the cell. Currently, inhibitors of the mentioned enzymes are at various stages of preclinical and clinical trials and out of medical practice.

Thymidine kinase inhibitors occupy a special place in the development of new-generation antiviral drugs. It should be noted that this enzyme plays a key role in the thymidine metabolism both in healthy and virus-infected cells. In healthy cells, this intracellular enzyme catalyzes the conversion of thymidine to thymidine monophosphate (TMP) in the presence of adenosine triphosphate (ATP). Viral thymidine kinase differs from the thymidine kinase of the host cell in its much greater substrate specificity and it is able to catalyze the phosphorylation of thymidine, pyrimidines, and purines. Subsequently, in both healthy and virus-infected cells, the resulting monophosphates of pyrimidines and purines are converted into the corresponding bi- and triphosphates. Triphosphates are then incorporated into deoxyribonucleic acid. Thus, viral thymidine kinase inhibitors cannot be simultaneously used with preparations containing acyclovir and its analogues as an active component [5,8].

DNA polymerase inhibitors prevent the replication of the virus after reactivation. In contrast, thymidine kinase inhibitors are aimed at preventing reactivation by lengthening the latent period of the virus. However, when developing antiviral drugs based on thymidine kinase, one should bear in mind that viral thymidine kinase is not a key target in the replication of herpes virus in rapidly dividing cells where the amount of thymidine triphosphate is sufficient for the synthesis of viral DNA due to cellular metabolism. However, this enzyme plays a key role in non-proliferating (non-dividing) nerve cells, in which the synthesis of cellular DNA occurs at a low level (if at all). In this case, the inhibition of viral thymidine kinase leads to the growth of damaged and, therefore, non-viable cells in primary neuronal cultures.

In addition to a potential target in the fight against viral infections, thymidine kinase (TK) is considered a tumor marker, which is used to diagnose and monitor the increased proliferation of tumor cells. It is known that tumor cells have an increased concentration of TK due to their high-intensity division and growth.

Thus, the search for efficient TK inhibitors, including TK of the human herpes viruses HSV-1 and HSV-2, can be considered one of the promising medical treatment options for herpetic infections and cancer diseases of various origins [16,17]. However, the rational search for new drugs without involving virtual screening methods is impractical both from an economic point of view and because of the high time costs [18–22].

In this regard, scientists use various approaches for computer-aided drug design (CADD) to search for hit compounds at the initial stage of development of new potential drugs [23]. There are known drugs that have been developed using this approach, such as tirofiban [24], zanamivir [25], boceprevir [26], saquinavir [27], captopril [28,29], and aliskiren [30]. CADD approaches are classified into structure-based and ligand-based methods. In the first category of methods, computational drug design is carried out by studying the interactions between ligands and target molecules. Accordingly, the aim of this variant of CADD is to optimize the binding structure of the ligand under study and the corresponding receptor in a three-dimensional form. Virtual protein–ligand complexes are modeled using pharmacophore search, molecular docking, and molecular dynamics methods. The most obvious disadvantage of structure-based approaches in CADD is the requirement of correct information on the receptor structure and the high time and computational costs [31].

In the alternative category of ligand-based CADD approaches, the leading factors contributing to biological activity are the physicochemical, electronic, and conformational features of the ligands. The key advantage of the latter strategy over the former one is mainly that in the latter case, knowledge of the spatial structure and amino acid composition of the target is not required for the design of potential drugs [23,31].

Quantitative structure–activity relationship (QSAR) is a valuable method in CADD which aims to build statistically significant mathematical models for predicting different biological activity parameters (pIC_{50}, pLD_{50}, pK_i, etc.) based on different physicochemical, electronic, and structural characteristics of organic compounds [32,33]. In terms of dimensionality, the type of QSAR models depends on the descriptors used, ranging from 0D-QSAR to 7D-QSAR [34]. Several descriptors (e.g., atomic properties, number of fragments, and topological descriptors) make up the 0D to 2D-QSAR components. Modeling using 3D-QSAR methods requires the inclusion of 3D descriptors giving an additional dimension in spatial coordinates [35,36]. Additional aspects of 3D-QSAR models require the use of multidimensional molecular descriptors based on conformational flexibility, induced fit, solvation function, and target-based receptor models. These supplements generate multidimensional QSAR (i.e., 4D to 7D-QSARs) [32]. A factor complicating the practical use of 3D-7D QSAR methods is the required knowledge of the bioactive conformation of the ligands that are structural analogues of the compounds being modeled [37–39]. Taking into account of all of the above factors in terms of time and computational cost can, in some cases, be far superior to the first category of structure-based CADD methods. In this regard, today there is a growing interest in the use of 2D-QSAR models against the background of a relatively smaller number of studies using multivariate QSAR approaches despite the high predictive power, logical validity, and objectivity of the latter.

The GUSAR program, developed at the V.N. Orekhovich Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, is modern software for the construction of quantitative and classification models (QSAR and SAR models) and the prediction of various types of biological activity, as well as other properties of organic compounds based on a 2D approach (structural formulae of organic compounds) [20]. In this software, the chemical structure is described in terms of descriptors called quantitative neighborhoods of atoms (QNAs) and multilevel neighborhoods of atoms (MNAs) developed at the same institute [40–42]. The functioning algorithm of this software is based on the method of self-consistent regression previously developed by the same team with the inclusion of additional estimates of the quality of prediction of the target property (based on the method of nearest neighbors and artificial neural network with a radial basis function) and construction of a consensus of the set of models [20]. It is reported that GUSAR is not inferior to other methods (CoMFA, CoMSIA, HQSAR, etc.) used to build QSAR/QSPR models in terms of accuracy and predictive ability [43,44]. As a result, the software can be successfully applied to a variety of QSAR/QSPR tasks [45–58]. In particular, the GUSAR software has been used for more than a decade to model various types of biological activity and toxicity of organic compounds [40–49]. In addition, the successful application of this software has

been demonstrated for the QSPR modeling of several physicochemical properties of organic compounds, including the n-octanol–water partition coefficient (logP) [45], boiling and melting points, density, thermal conductivity, viscosity, surface tension, water solubility, and gas pressure [40].

Additionally, our earlier publications demonstrated the successful application of this software for QSPR modeling of antioxidants under conditions of the liquid-phase radical-chain oxidation of organic substrates [59–63].

This software has been used for more than a dozen years for modeling different types of biological activity. It was shown by the developers and other researchers, including our research team, that GUSAR software can be successfully applied to multiple QSAR/QSPR problems [50–63].

In this work, we used GUSAR 2019 software to study the quantitative structure–activity relationship for inhibitors of HSV-1 and HSV-2 viral thymidine kinase using the series of 5′-amino-2′,5′-dideoxy-5-ethyluridine (I–III), N2-phenylguanine (IV), and 2-phenylamino-6-oxopurine carboxamide derivatives (V–VI, Figure 1) and developed coupled and statistically significant QSAR models for screening virtual libraries and databases.
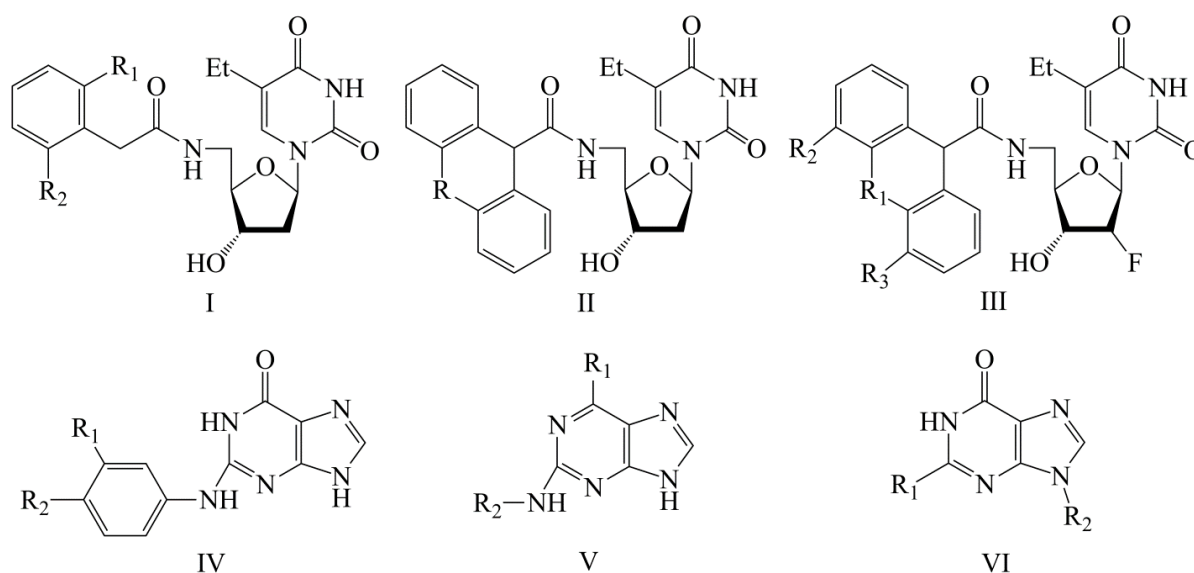

Figure 1. General structural formulas of simulated inhibitors of HSV-1 and HSV-2 thymidine kinases based on a series of 5′-amino-2′,5′-dideoxy-5-ethyluridine (I–III), N2-phenylguanine (IV), and 2-phenylamino-6-oxopurine derivatives (V,VI).

2. Results

Using the consensus approach implemented in the GUSAR 2019 program, we have studied the quantitative relationship between the structure and the efficiency of inhibition of HSV-1 and HSV-2 TK with 5-ethyluridine, N2-guanine, and 6-oxopurine derivatives with general structural formulas I–VI (Figure 1). These compounds made up the training sets TrS1–TrS4. Depending on the type of descriptors used in the calculations (MNA and/or QNA), three QSAR consensus models have been obtained for each of the training sets. In total, we have built 12 QSAR consensus models for predicting pIC_{50} values for HSV-1 and HSV-2 TK inhibitors that included from 20 to 360 partial regression models. The pIC_{50} values for inhibitors included in TrS1–TrS4 derived from these QSAR consensus models M1–M12 were compared with the experimental values of pIC_{50} (see Tables S2–S5 in Supplementary Materials).

The regression models were not explicitly displayed, as a clear physical interpretation of the descriptors was absent. Hence, we could not determine the descriptors making the largest/the smallest contributions to the simulated activity [64,65]. However, this was beyond the scope of this study. Our goal was to solve two problems:

(1) to show that the ideology of descriptor formation and selection implemented in the GUSAR 2019 software is applicable for modeling potential inhibitors of HSV-1 and HSV-2 TK enzymes in the series of 5-ethyluridine, N2-guanine, and 6-oxopurine derivatives;

(2) to develop statistically significant QSAR models suitable for the virtual screening of HSV TK inhibitors.

For the internal validation of the QSAR models M1–M12 over the TrS1–TrS4 structures, we used a cross validation procedure with a 20-fold randomized exclusion of 20% of the compounds. Here, the averaged values of determination coefficients $\overline{R^2}$ and $\overline{Q^2}$ for the inhibitors of all training sets were similar (Table 1); the difference between these two indicators ($\Delta = \overline{R^2} - \overline{Q^2}$) did not exceed 0.1. This assessment indicates the stability of the constructed consensus models.

Table 1. Statistical parameters and accuracy of the predicted pIC_{50} values of the compounds included in the training sets TrS1–TrS4 within the M1–M12 consensus models. $\Delta pIC_{50\,TrS1} = \Delta pIC_{50\,TrS3} = 5.867$, $\Delta pIC_{50\,TrS2} = \Delta pIC_{50\,TrS4} = 6.250$ [1].

Training Set	Model	N	N_{PM}	$\overline{R^2}$	$\overline{F}$	$\overline{SD}$	$\overline{Q^2}$	V
				QSAR models based on the QNA descriptors				
TrS1	M1	73	20	0.878	67.101	0.569	0.848	7
TrS2	M4	74	20	0.891	84.683	0.593	0.869	6
TrS3	M7	61	20	0.875	50.879	0.579	0.837	7
TrS4	M10	62	20	0.891	65.152	0.598	0.863	6
				QSAR models based on the MNA descriptors				
TrS1	M2	73	20	0.878	63.594	0.568	0.854	7
TrS2	M5	74	20	0.906	79.140	0.552	0.887	8
TrS3	M8	61	20	0.882	51.831	0.565	0.853	7
TrS4	M11	62	20	0.894	70.947	0.589	0.872	6
				QSAR models based on both QNA and MNA descriptors				
TrS1	M3	73	320	0.891	57.523	0.542	0.862	8
TrS2	M6	74	320	0.905	70.945	0.559	0.882	8
TrS3	M9	61	320	0.881	45.955	0.570	0.846	7
TrS4	M12	62	320	0.899	63.865	0.578	0.873	7

[1] N is the number of structures in the training set; N_{PM} is the number of regression equations used for the consensus model; $\overline{R^2}$ is the coefficient of determination calculated for the compounds of TrSi; $\overline{Q^2}$ is the correlation coefficient calculated for the training set based on cross-validation with the exception of one; $\overline{F}$ is Fisher's criterion; $\overline{SD}$ is the standard deviation; V is the number of variables in the final regression equation.

This is exemplified in Tables 1–3, which present the numerical values of the statistical criteria estimated by comparing the experimental and predicted pIC_{50} values calculated using models M1–M12 with 95% of the data included in the corresponding training set. Full information about all of these criteria using the twelve developed QSAR models, which enables an objective evaluation of the descriptive and predictive ability of the models, taking into account 95% and 100% of the data included in the training and test sets, respectively, is given in the Supplementary Materials (Tables S2–S5).

The data of Tables 1–3 provide the conclusion that all constructed QSAR models had high descriptive ability. However, the data presented in Tables 1–3 clearly demonstrate the discrepancy between the numerical values of determination coefficients (R^2) found while evaluating the descriptive ability of models M1–M12 in the GUSAR 2019 and XternalValidationPlus 1.2 software, due to different ideologies underlying the calculations.

It should be taken into account that in the GUSAR 2019 software, the target parameter (in our case, pIC_{50}) for each chemical structure included in the training or test set is predicted as a result of averaging the numerical values of this parameter calculated using each of the particular models included in a single consensus model. The final statistical parameters are calculated in a similar way.

Table 2. 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 the HSV-1 TK inhibitors from training sets TrS1 (M1–M3) and TrS3 (M7–M9). $\Delta pIC_{50\,TrS1} = \Delta pIC_{50\,TrS3} = 5.867$ [1].

Comments	Prediction Parameters	QSAR Model Used for Predicting pIC_{50}					
		TrS1			TrS2		
		M1	**M2**	**M3**	**M7**	**M8**	**M9**
Classical metrics (after removing 5% of the data with high residuals)	R^2	0.9609	0.9594	0.9653	0.9591	0.9611	0.9654
	R^2_0	0.9555	0.9579	0.9614	0.9556	0.9587	0.9593
	$R^{2'}_0$	0.8443	0.8804	0.8661	0.8568	0.8725	0.8483
	$\overline{R^2_m}$	0.8776	0.9052	0.8952	0.8883	0.8971	0.8819
	ΔR^2_m	0.0379	0.0355	0.0326	0.0379	0.0352	0.0342
	CCC	0.9755	0.9777	0.9790	0.9759	0.9779	0.9775
Mean absolute error and standard deviation for the test set (after removing 5% of the data with high residuals)	RMSE	0.3368	0.3331	0.3193	0.3323	0.3327	0.3331
	MAE	0.2914	0.2784	0.2673	0.2872	0.2768	0.2830
	SD	0.1701	0.1844	0.1758	0.1687	0.1861	0.1773
	MAE + 3·SD	0.8016	0.8314	0.7948	0.7933	0.8351	0.8149
Prediction quality	-	Good					
Presence of systematic errors	-	Absent					

[1] R^2, R^2_0, and $R^{2'}_0$ are the determination coefficients calculated with and without taking into account the origin; average $\overline{R^2_m}$ is the averaged determination coefficient of the regression function calculated using the determination coefficients on the ordinate axis (R^2_m) and on the abscissa axis ($R^{2'}_m$), respectively; $\Delta\overline{R^2_m}$ is the difference between R^2_m and $R^{2'}_m$; CCC is the concordance correlation coefficient; MAE is the mean absolute error; SD is the standard deviation.

Table 3. 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 the HSV-2 TK inhibitors from training sets TrS2 (M4–M6) and TrS4 (M10–M12). $\Delta pIC_{50\,TrS2} = \Delta pIC_{50\,TrS4} = 6.250$ [1].

Comments	Prediction Parameters	QSAR Model Used for Predicting pIC_{50}					
		TrS2			TrS4		
		M4	**M5**	**M6**	**M10**	**M11**	**M12**
Classical metrics (after removing 5% of the data with high residuals)	R^2	0.9714	0.9712	0.9719	0.9708	0.9676	0.9743
	R^2_0	0.9687	0.9701	0.9694	0.9681	0.9664	0.9710
	$R^{2'}_0$	0.8890	0.9086	0.8927	0.8889	0.9009	0.8891
	$\overline{R^2_m}$	0.9137	0.9267	0.9142	0.9148	0.9216	0.9109
	ΔR^2_m	0.0270	0.0260	0.0267	0.0273	0.0290	0.0252
	CCC	0.9830	0.9843	0.9836	0.9827	0.9823	0.9844
Mean absolute error and standard deviation for the test set (after removing 5% of the data with high residuals)	RMSE	0.3278	0.3121	0.3146	0.3333	0.3328	0.3164
	MAE	0.2712	0.2590	0.2624	0.2739	0.2822	0.2676
	SD	0.1856	0.1753	0.1748	0.1915	0.1781	0.1703
	MAE + 3·SD	0.8279	0.7850	0.7868	0.8484	0.8164	0.7785
Prediction quality	-	Good					
Presence of systematic errors	-	Absent					

[1] R^2, R^2_0, and $R^{2'}_0$ are the determination coefficients calculated with and without taking into account the origin; average $\overline{R^2_m}$ is the averaged determination coefficient of the regression function calculated using the determination coefficients on the ordinate axis (R^2_m) and on the abscissa axis ($R^{2'}_m$), respectively; $\Delta\overline{R^2_m}$ is the difference between R^2_m and $R^{2'}_m$; CCC is the concordance correlation coefficient; MAE is the mean absolute error; SD is the standard deviation.

For example, when predicting pIC_{50} values for any compound from the training set TrS1 using the consensus model M1, we get a set of 20 predicted $pIC_{50\,pred}$ values and

20 sets of different internal validation criteria: R^2, Q^2, F, and SD. Further, all the same data are averaged, which is displayed as the final results.

Meanwhile, in the XternalValidationPlus 1.2 program, the calculation of statistical parameters for assessing the descriptive and predictive ability of QSAR models is based on comparing the experimental pIC_{50} data with the average values previously predicted using the GUSAR 2019 software. This procedure is performed twice without averaging the final results [66]:

(1)   for the full dataset in each training and test set (100% of data);
(2)   for 95% of the data in each training and test set (95% of the data).

In general, a comparison of the data given in Tables 1–3 demonstrates that the SCR method of GUSAR 2019 for selecting significant descriptors produces stable regression dependences with acceptable statistical characteristics ($R^2_{TrS} > 0.6$ and $Q^2_{TrS} > 0.5$) for simulated HSV-1 and HSV-2 TK inhibitors, regardless of the selected types of descriptors.

The different determination criteria of the descriptive ability of models M1–M12 are similar irrespective of the amount of data in the sets (95 or 100%) and tend to be 1 (Tables 2 and 3). The MAE error values do not exceed 15% of the ΔpIC_{50} range of the inhibitory activity of the TrS1–TrS4 structures. The parameter ΔR^2_m is in all cases is much lower than 0.2 and does not exceed 0.048. All of these data indicate the rather high simulability of the target properties using the selected algorithms for a calculation of descriptors and construction of regression equations [67] implemented in the GUSAR 2019 software.

An external validation of the M1–M3 and M7–M9 QSAR models was performed by predicting the pIC_{50} for HSV-1 TK inhibitors using test sets TS1 and TS3. The validity of the models M4–M6 and M10–M12, meant for the prediction of the pIC_5 for HSV-2 TK inhibitors, was evaluated in relation to test sets TS2 and TS4. All estimates of the predictive ability of the M1–M12 models were based on three criteria:

(1)   numerical values of various coefficients of determination based on R^2 (R^2, R^2_0, Q^2_{F1}, Q^2_{F2}, CCC);
(2)   numerical values of the MAE prediction error;
(3)   the scatter range of activity prediction data taking into account MAE in the mσ (or mSD) range: MAE + 3·SD. All of these parameters were computed using the XternalValidationPlus 1.2 program. In addition, this program was used to trace the systematic error that can arise in QSAR modeling.

Figures 2–5 show the distribution of different determination coefficients and prediction errors for pIC_{50} values for 95% of the HSV inhibitors from test sets TS1–TS4 calculated using the XternalValidationPlus 1.2 program. The complete set of all statistical parameters obtained from a comparison of experimental and predicted pIC_{50} values for the TS1–TS4 structures determined based on models M1–M12 is given in Tables S2–S5 (Supplementary Materials).

The more stringent criterion $\overline{R^2_m}$ is relatively high for the external validation of models M1–M12 using the full size of TS1–TS4, being in the range of 0.8273–0.8859 and 0.7587–0.8683 for HSV-1 and HSV-2 inhibitors, respectively. After removing 5% of outliers from TS1–TS4, the ranges of $\overline{R^2_m}$ become 0.8207–0.9294 and 0.8664–0.9294 for HSV-1 and HSV-2 inhibitors, respectively (Figures 2–5, Tables S2–S5 in Supplementary Materials). The $\Delta \overline{R^2_m}$ criterion, proposed by the same authors as an additional parameter for assessing the predictive ability for the external validation of regression models, did not exceed 0.09 in any of the cases. This also indicates the rather high predictive ability of QSAR models M1–M12 (Figures 2–5, Tables S6–S13 in Supplementary Materials).
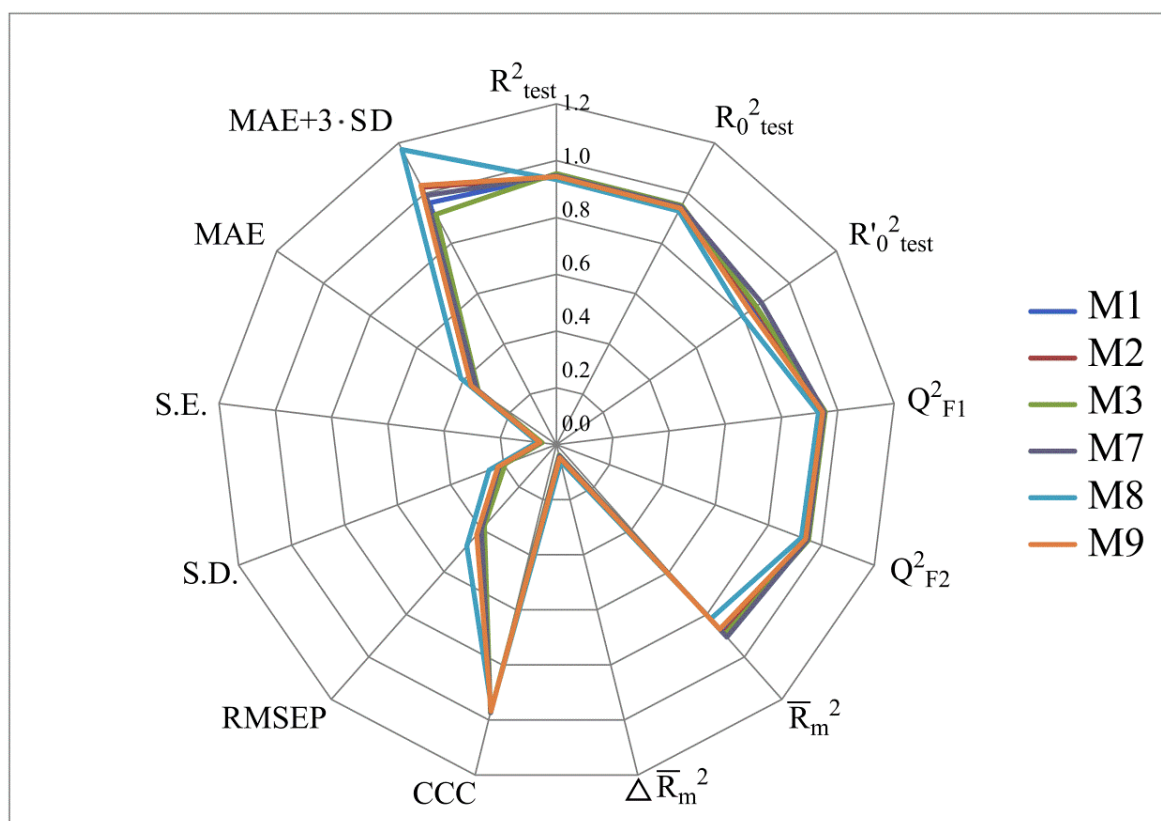
Figure 2. Distribution of statistical characteristics of QSAR models derived from predicted pIC_{50} values for the structures of the external test set TS1.
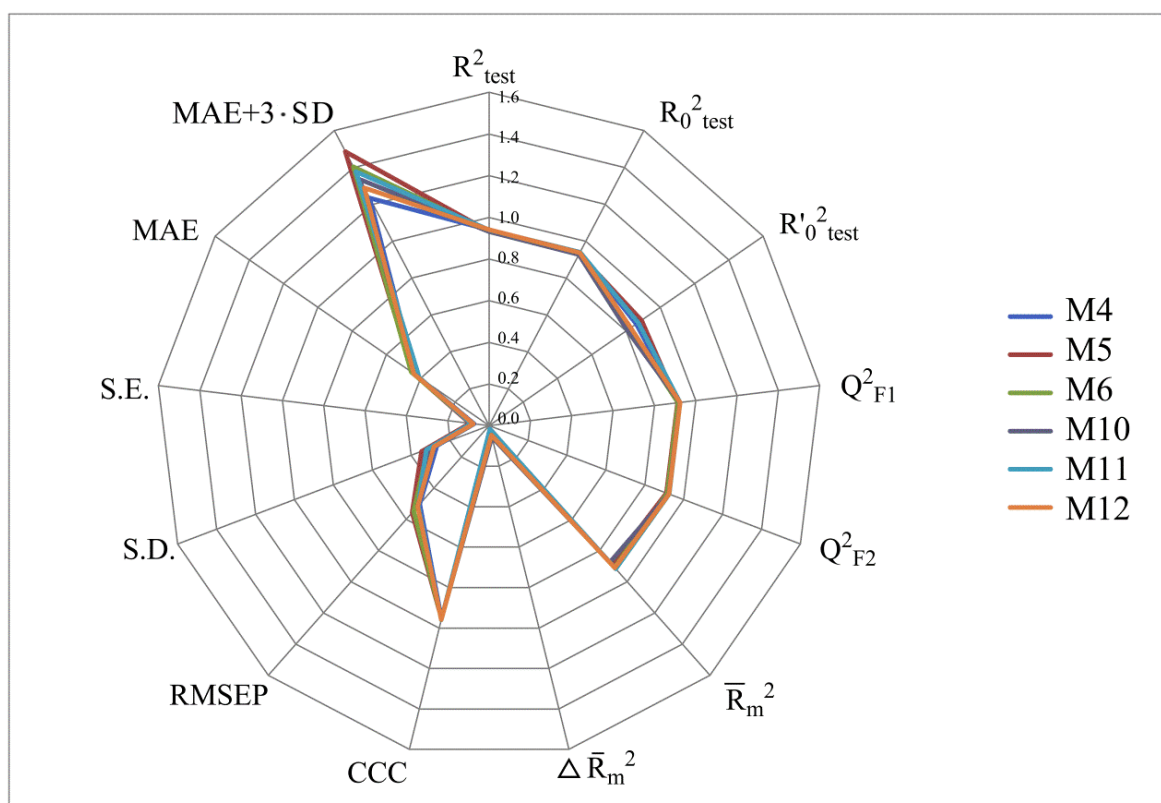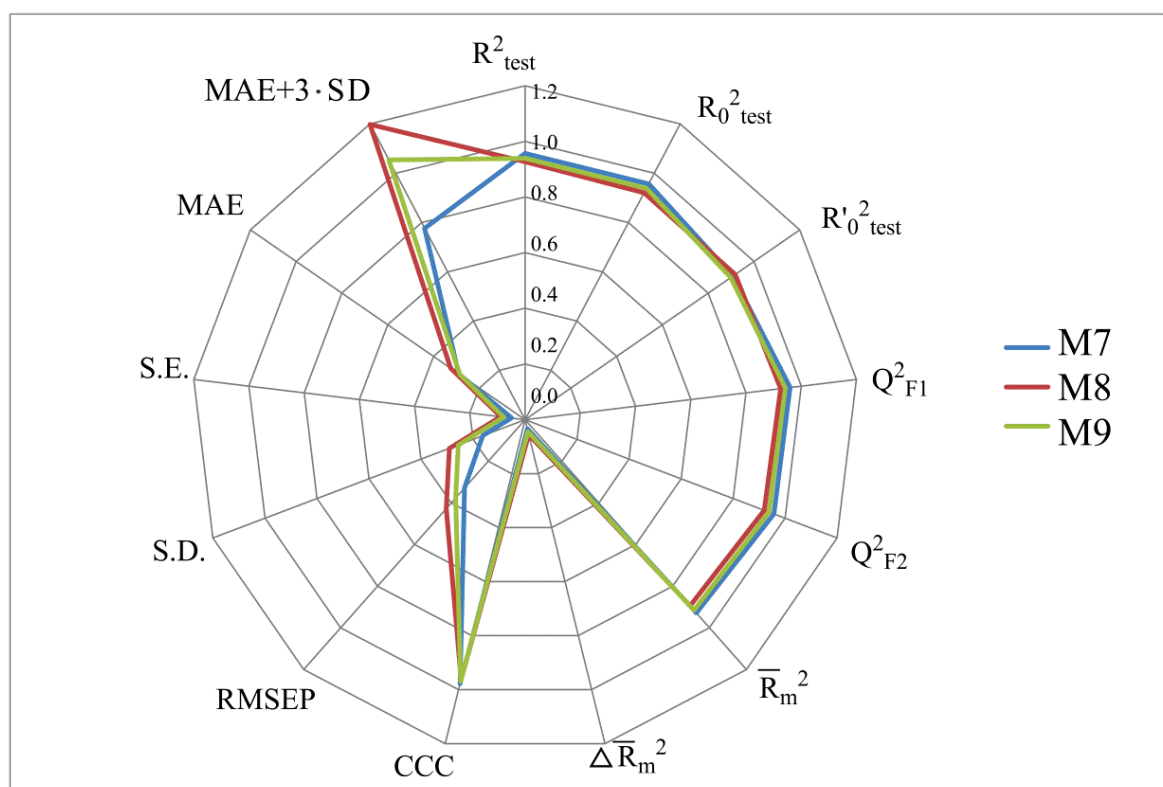

Figure 3. Distribution of statistical characteristics of QSAR models derived from the predicted pIC_{50} values for the structures of the external test set TS2.
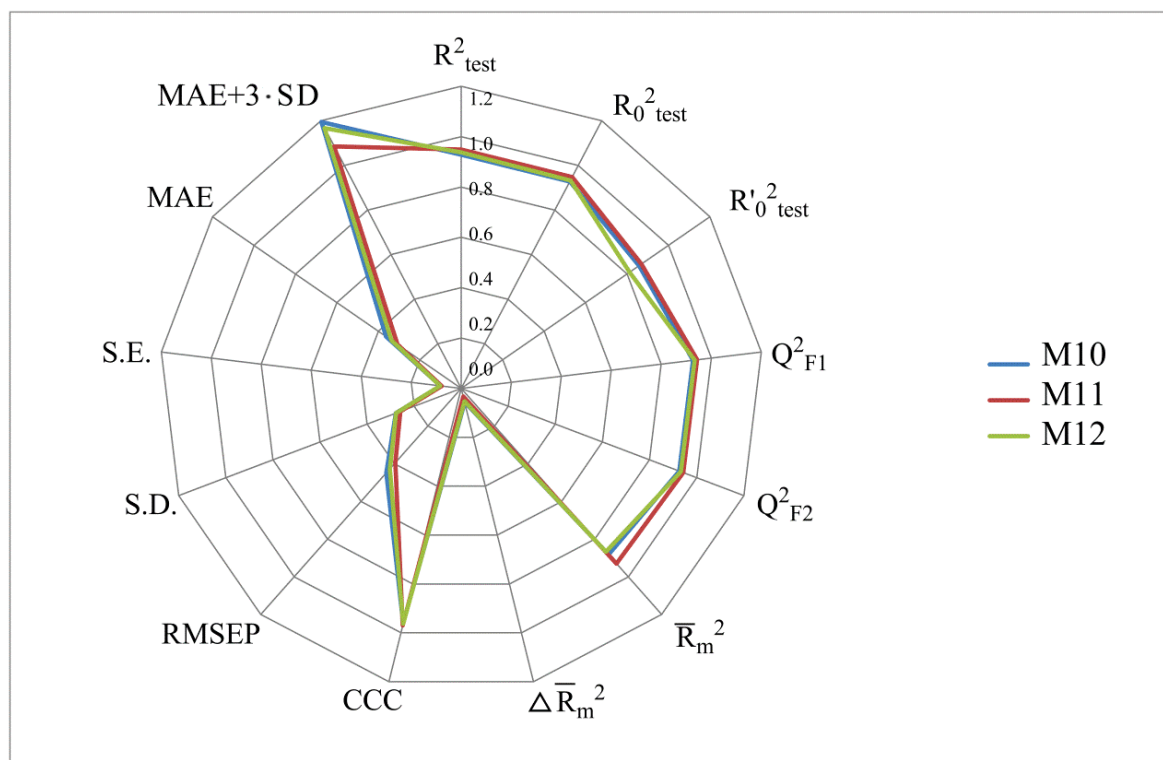
Figure 4. Distribution of statistical characteristics of QSAR models derived from the predicted pIC_{50} values for the structures of the external test set TS3.

Figure 5. Distribution of statistical characteristics of QSAR models derived from the predicted pIC_{50} values for the structures of the external test set TS4.

Based on a comparison of different determination coefficients obtained during the external validation of models M1–M12, we have found that the parameter pIC_{50} for 5-ethyluridine, N2-guanine, and 6-oxopurine derivatives with respect to HSV-1 is modeled with higher accuracy than that for the same compounds against HSV-2.

As we noted above, an analysis of different types of determination coefficients is faced with the following contradictory situation: the R^2 and R^2_0 values for the activity of 5-ethyluridine, N2-guanine, and 6-oxopurine derivatives against HSV-1 are equal to or less than Q^2_{F1} and Q^2_{F2}. This means that the constructed models M1–M12 predict the activities of TS1–TS4 compounds better than the activities of the training set structures. Note that in practice, the situation is usually opposite. This fact was repeatedly noted by other researchers [68–71]. Thus, the use of the metrics based on R^2 and Q^2 alone for assessing the predictive ability of QSAR models seems to be insufficient.

According to these two criteria, the predictive ability of models M1–M3, M7, and M9 has been classified as high for both test sets TS1 and TS3. Since the MAE + 3·SD criterion has been at the boundary of the allowable threshold value equal to 1.1735, the predictive ability of model M8 for 95% of the data of test set TS1 is moderate. The MAE and MAE + 3·SD values in the case of sets TS2 and TS4 do not exceed 0.6250 and 1.5625, respectively. As a result, the predictive ability of model M4 in the sets TS2 and TS4 has also been rated as high. At the same time, considering these two threshold values, the predictive abilities of M5–M12 can be estimated as satisfactory for set TS2 and as high for set TS4.

A comparative analysis of the statistical characteristics and prediction errors of pIC_{50} indicate that all constructed models have rather high descriptive and predictive ability. However, to solve the problem of searching for new potential inhibitors of HSV-1 and HSV-2 TK enzymes among the title compounds, it is most preferable to use the consensus models M3 and M6 because they include 100 particular regression models and each of them is based on the maximum set of structures and descriptors.

In this regard, we have applied the consensus models M3 and M6 to virtual screening through the CHEMBL database for new HSV TK inhibitors among various lead compounds and active drug components of different pharmacological profiles. Unlike traditional methods of QSAR modeling (multilinear regression (MLR), partial least squares (PLS) method, etc.), the GUSAR software does not specify clear threshold criteria regarding the Tanimoto coefficient, which would limit the search for new potential biologically active substances in virtual databases. However, adhering to concepts of the classical school, we limited the scope of the search for new potential inhibitors of HSV-1 TK and HSV-2 TK in the ChEMBL database by the degree of similarity of at least 70% with respect to the reference compounds.

The virtual screening involved 400 5-ethyluridine, N2-guanine, and 6-oxopurine derivatives with pronounced antitumor and antibacterial properties and no antiviral properties. However, only 192 lead compounds and known pharmaceuticals fitted in the range of applicability of consensus models M3 and M6. For 155 structures of these, the predicted IC_{50} values were <1 μmol/L. The most promising hit compounds are presented in Table 4. The complete list of the structures of the potential HSV TK inhibitors predicted using consensus models M3 and M6 is given in Table S14 in the Supplementary Materials. We assume that in living systems, these compounds should behave as multi-target drugs. They are promising for further detailed studies.

Additionally, using the GUSAR 2019 program, we carried out a structural analysis of TK inhibitors. Since for 42 compounds presented in Table 4, there were no experimental data on the inhibitory activity against human herpes viruses HSV-1 and HSV-2, these compounds were not included in the structural analysis. We used the consensus model M3, as it provides more objective and accurate results due to the maximum number of modeled structures and involvement of all types of descriptors implemented in GUSAR 2019 [20–22]. However, it should be noted that these compounds have been extensively studied for their inhibitory activity against HPV in previous biological experiments [17,72,73]. Therefore, here, we will briefly discuss this issue. Figures 6–8 show the analysis of the contribution of different functional groups

to the activity of inhibitors of HSV-1 and HSV-2 thymidine kinase with general structural formulas I–VI. For compounds with the general structural formula I, it was experimentally shown that the replacement of a hydrogen atom in the R_1 position of the benzene ring (**1**) increases the inhibitory activity, irrespective of the nature of the acyclic substituent. The results of a structural analysis of the same compounds obtained using the GUSAR 2019 program lead to a similar conclusion. This enhancement is manifested for compounds **2–7** containing fluoro (**2**), chloro (**3**), methyl (**4**), and trifluoromethyl (**5**) substituents in the ortho-positions (Figure 6a).

Table 4. Potential effective HSV-1 and HSV-2 TK inhibitors selected from the ChEMBL database using virtual screening with QSAR models M3 and M6.
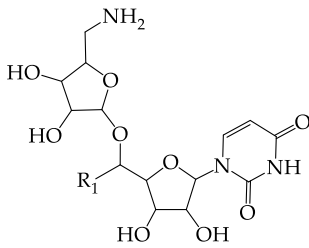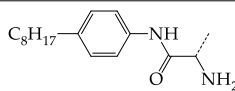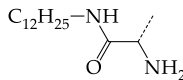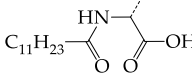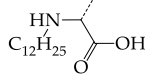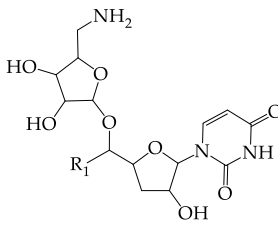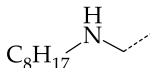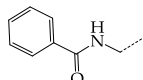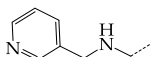
No.	Name in ChEBIL	Structure	pIC_{50pred}		Selectivity $\left(\text{Selectivity}=\frac{IC_{50,HSV-1}}{IC_{50,HSV-2}}\right)$
			HSV-1	HSV-2	
		R_1			
1	CHEMBL1199108		15.29	2.87	5.3359
2	CHEMBL1199070		32.52	13.98	2.3267
3	CHEMBL1199059		27.75	21.38	1.2980
4	CHEMBL1780207		30.42	21.46	1.4176
		R_1			
5	CHEMBL20028		35.85	27.30	1.3131
6	CHEMBL1178256		31.91	5.91	5.4029
7	CHEMBL19326		14.87	3.82	3.8897
8	CHEMBL1178302		13.77	3.27	4.2105

Table 4. *Cont.*

No.	Name in ChEBIL	Structure	pIC$_{50pred}$		Selectivity $\left(\text{Selectivity}=\frac{IC_{50,HSV-1}}{IC_{50,HSV-2}}\right)$
			HSV-1	HSV-2	
9	CHEMBL19510		9.73	1.37	7.0878
10	CHEMBL1178307 *		13.97	2.63	5.3210
11	CHEMBL19608		6.88	0.83	8.3308
12	CHEMBL19725		10.33	2.06	5.0177
13	CHEMBL19782		8.01	1.41	5.6706
14	CHEMBL1178314		8.42	1.52	5.5286
15	CHEMBL1178315		9.27	1.73	5.3491
16	CHEMBL277025		12.04	1.51	7.9804
17	CHEMBL1183046		11.01	0.99	11.0940
18	CHEMBL277844		5.76	0.70	8.2058
19	CHEMBL1183063		12.58	2.05	6.1317
20	CHEMBL278626		8.87	0.89	9.9477
21	CHEMBL1183081		12.39	2.53	4.9020
22	CHEMBL1183082		34.36	7.19	4.7770
23	CHEMBL1183089		10.95	1.20	9.1477
24	CHEMBL1183095		8.78	0.71	12.4135

Table 4. *Cont.*
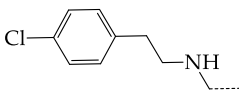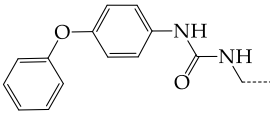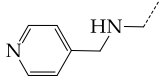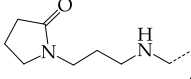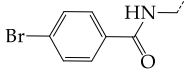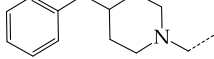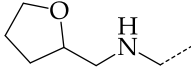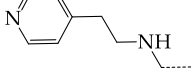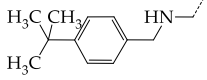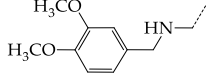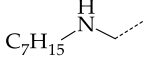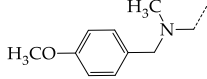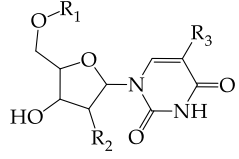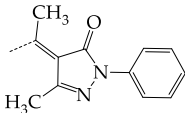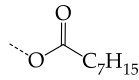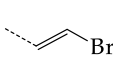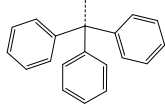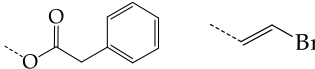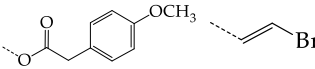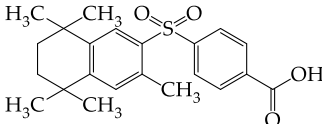
No.	Name in ChEBIL	Structure	pIC$_{50pred}$		Selectivity $\left(\text{Selectivity}=\frac{\text{IC}_{50,\text{HSV}-1}}{\text{IC}_{50,\text{HSV}-2}}\right)$
			HSV-1	HSV-2	
25	CHEMBL1183096		7.31	1.04	7.0456
26	CHEMBL279892		8.78	0.74	11.7868
27	CHEMBL1183107		14.50	4.72	3.0716
28	CHEMBL1183108		13.71	3.16	4.3415
29	CHEMBL280909		5.38	1.07	5.0082
30	CHEMBL1183123		8.20	0.83	9.8336
31	CHEMBL1183154		15.30	4.26	3.5958
32	CHEMBL1183178		11.23	2.77	4.0530
33	CHEMBL1183185		8.32	1.57	5.2872
34	CHEMBL1185346		8.28	1.06	7.7791
35	CHEMBL1185463		32.63	6.28	5.1918
36	CHEMBL1185716		8.81	0.93	9.4314

No.	Name in ChEBIL	R$_1$	R$_2$	R$_3$			
37	CHEMBL217675 *	-H	-H		62.83	26.96	2.3306
38	CHEMBL238635	-H			36.62	42.98	0.8520

Table 4. *Cont.*

No.	Name in ChEBIL	Structure		pIC$_{50pred}$		Selectivity $\left(\text{Selectivity}=\dfrac{\text{IC}_{50,\text{HSV}-1}}{\text{IC}_{50,\text{HSV}-2}}\right)$
				HSV-1	HSV-2	
39	CHEMBL2403290 *	(triphenylmethyl)	-H	-CH$_3$	26.44	40.28	0.6564
40	CHEMBL241407 *	-H	(phenylacetyl ester)	=Br	14.48	22.16	0.6535
41	CHEMBL241408 *	-H	(4-OCH$_3$-phenylacetyl ester)	=Br	10.05	6.47	1.5544
42	CHEMBL1183075 *	(tetramethyl-dihydronaphthalene sulfonyl benzoic acid structure)			15.18	3.11	4.8817

The asterisk marks structures that obey Lipinski's rule of five.

Figure 6. Effect of acyclic substituents on the activity of herpes virus inhibitors with general formulas I and II with the chemical group contributions to the activity; superscripts 1 and 2 refer to the activities against HSV-1 TK and HSV-2 TK, respectively. Dotted lines highlight the substituents. The up and down arrows indicate the positive or negative effect of the selected group. A and B denote fragments that remained unchanged during structural analysis.

In compounds with the general formula II, replacement of the dihydroxanthene moiety (**8**) with a xanthene (**9**) or thioxanthene dioxide (**12**) moiety somewhat increases the activity of the TK inhibitors of HSV-1 and HSV-2. At the same time, replacement by dibenzosuberene, anthracene, or NMe-acridine (**10**) has an adverse effect on both target properties. Note that the first two of these groups induce a pronounced decrease in the inhibitory activity, while the third replacement has only a moderate effect. The replacement of the dihydroxanthene moiety in **8** with a thioxanthene moiety (**11**) decreases the inhibitory activity against HSV-1 TK by a factor of 1.5 and has almost no effect on the inhibitory activity against HSV-2 TK (Figure 6b).
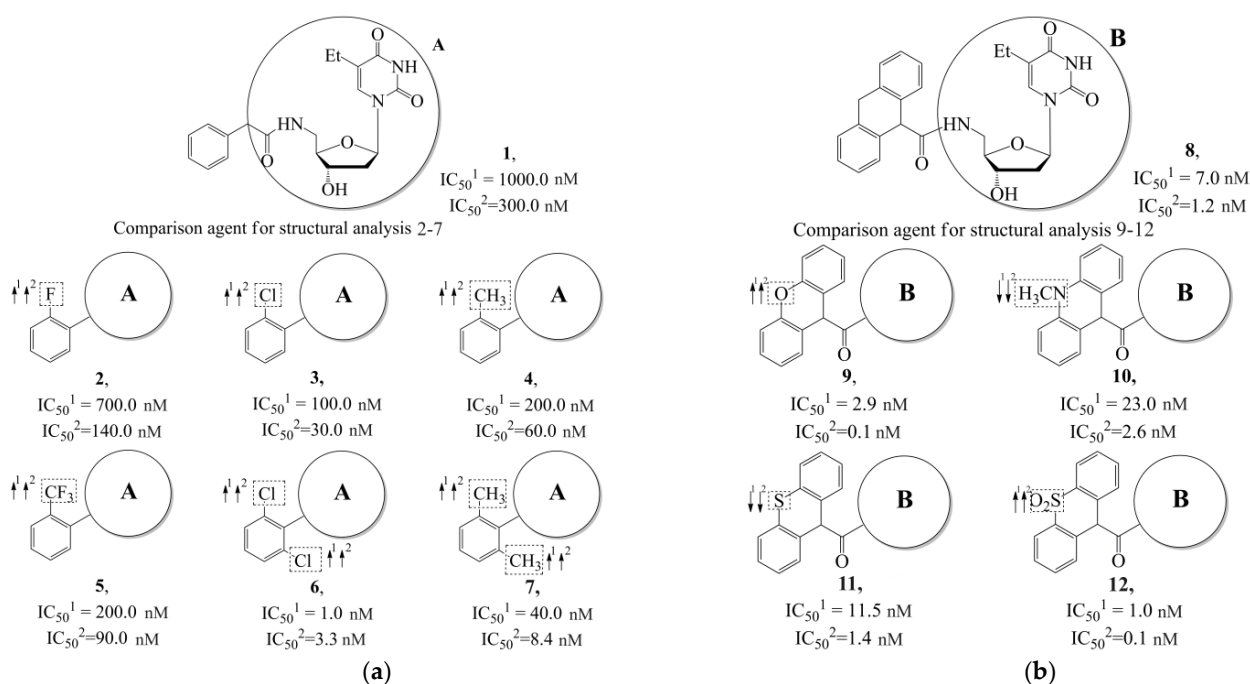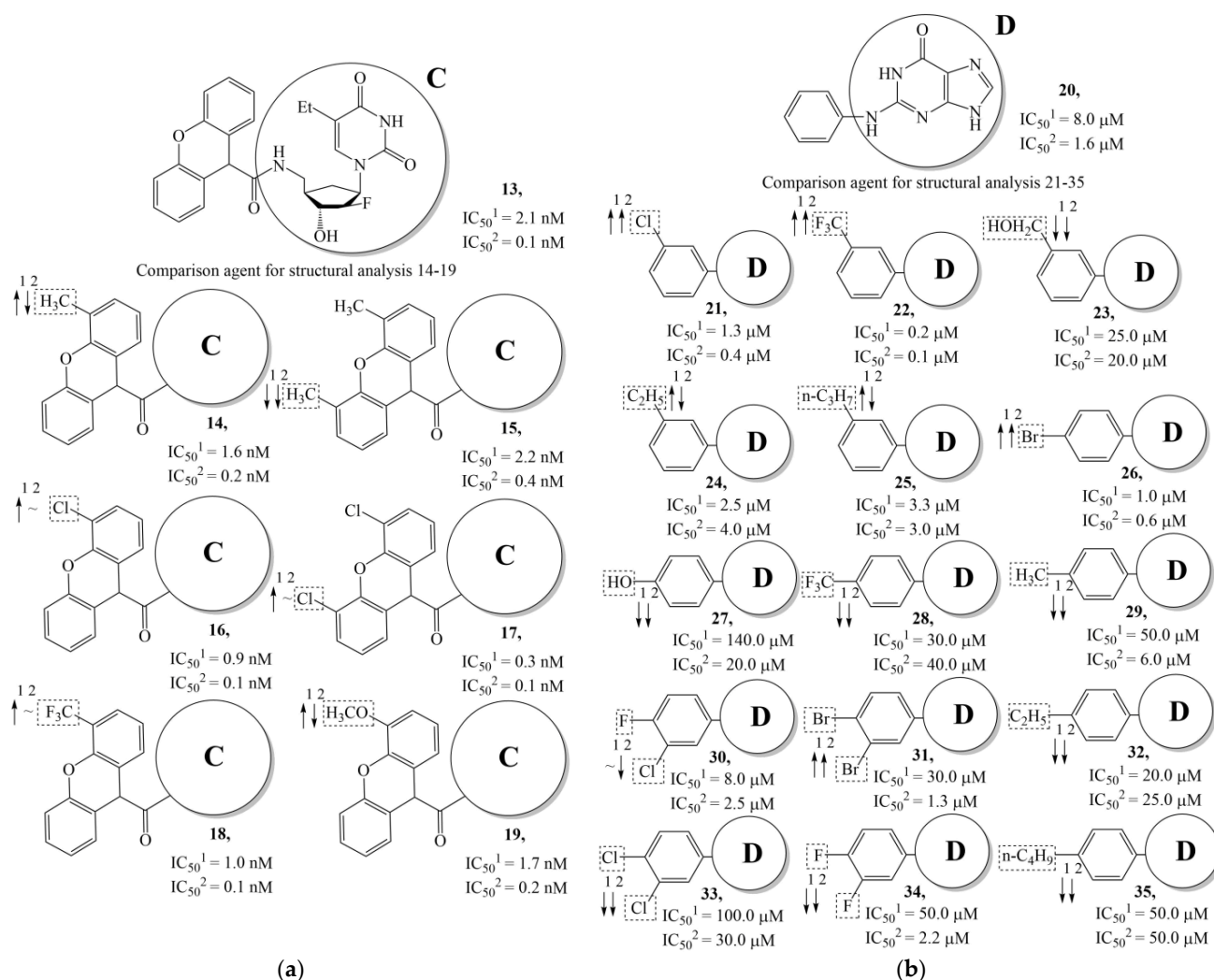
Figure 7. Effect of acyclic substituents on the activity of herpes virus inhibitors with general formulas III–IV with the chemical group contributions to the activity; the superscripts 1 and 2 refer to activities against HSV-1 TK and HSV-2 TK, respectively. The dotted lines highlight the substituents. The up and down arrows indicate the positive or negative effect of the selected group. C and D denote fragments that remained unchanged during structural analysis.

In compounds with the general structural formula III containing an oxygen atom in position R_1 (i.e., xanthene ring, **13**) (Figure 7a), the replacement of the hydrogen atom in position R_2 with a methyl group (**14**) increases the inhibitory activity against HSV-1 TK and impairs the activity of TK inhibitors against HSV-2. However, the effect is not clearly pronounced in both cases. The introduction of a second methyl group into position R_3 (**15**) of the xanthene ring decreases both target properties. The alternative replacement of the hydrogen atom at position R_2 by a chlorine atom (**16**) increases the activity of TK inhibitors for HSV-1 almost 2-fold, but barely affects the inhibitory activity against HSV-2 TK. The additional incorporation of a second chlorine atom at position R_3 (**17**) is favorable for the activity against HSV-1 TK and almost does not influence the activity against HSV-2 TK. The replacement of the hydrogen atom in position R_2 with a trifluoromethyl group (**18**) and unsubstituted phenyl increases the TK inhibitory activity against HSV-1 and has almost no effect on this activity against HSV-2. Meanwhile, the modification of position R_2 by introducing a methoxy group (**19**) increases the activity of HSV-1 TK inhibitors and decreases the activity of HSV-2 TK inhibitors. However, the changes caused by a hydrogen atom replacement with the above substituents are moderate.
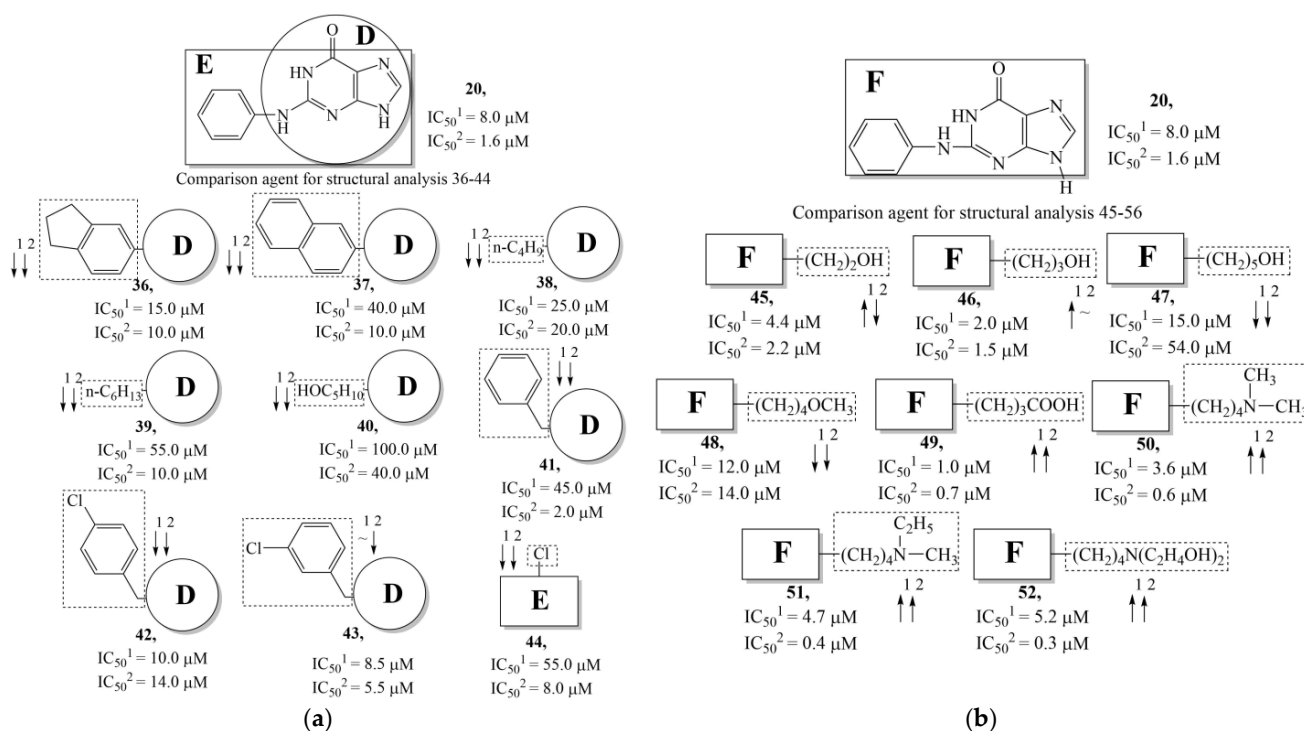
Figure 8. Effect of acyclic substituents on the activity of herpes virus inhibitors with general formulas V–VI with the chemical group contributions to the activity, where superscripts 1 and 2 refer to activities against HSV-1 TK and HSV-2 TK, respectively. Dotted lines highlight substituents. The up and down arrows indicate the positive or negative effect of the selected group. D, E and F denote fragments that remained unchanged during structural analysis.

In compounds with the general structural formula IV (Figure 7b), the replacement of the hydrogen atom (**20**) in the meta-position by chlorine (**21**) or a trifluoromethyl group (**22**) increases the activity of both TK isoforms quite significantly. Modification of the meta-position in the benzene ring with a hydroxymethyl group (**23**) negatively affects both target properties, and the adverse effect is high. At the same time, the alternative replacement of the hydrogen atom with ethyl (**24**) or n-propyl (**25**) increases the activity of TK inhibition of HSV-1 and decreases the activity of TK inhibition of HSV-2.

The replacement of hydrogen in the para-position with a bromine atom (**26**) favorably affects both target properties. In contrast, the alternative replacement of hydrogen with methyl (**29**), ethyl (**32**), n-butyl (**35**), trifluoromethyl (**28**), or hydroxyl (**27**) markedly decreases the inhibitory activity of compounds with the general structural formula IV against both TK isoforms (Figure 7b).

The simultaneous substitution of hydrogen atoms in the meta- and para-positions of the benzene ring by a bromine atom (**31**) considerably increases the efficiency of inhibitors of HSV-1 TK and almost does not affect the efficiency against HSV-2 TK. However, if we consider this substitution as sequential, the introduction of the second bromine atom in the meta-position of the benzene ring decreases the activity of both TKs compared to the modification of only the para-position by this substituent. The inclusion of fluorine and chlorine atoms in the para- and meta-positions (**30**) of the benzene ring, respectively, does not affect the inhibitory efficiency against HSV-1 and markedly decreases that against HSV-2. Similar modifications of para- and meta-positions based on the inclusion of two chlorine (**33**) or fluorine atoms (**34**) significantly decrease both target properties (Figure 7b).

The replacement of benzene (**20**) with a 2,3-dihydro-1H-indene (**36**) or naphthalene (**37**) ring and with a number of acyclic groups, including n-butyl (**38**), n-hexyl (**39**), and 1-hydroxypentyl (**40**), in compounds with general structural formula V has the same effect (Figure 8a).

In addition, in compounds with the general structural formula V, replacement of the benzene ring (**20**) in position R_2 with a benzyl moiety (**41**) markedly reduces the efficiency of inhibition of HSV-1 TK and has almost no effect on the activity of HSV-2 TK. At the same time, structural analogues of benzyl containing a chlorine atom in the meta- (**43**) or para-position (**42**) have the opposite effect, which is also markedly pronounced. The replacement of the oxo group (hydroxyl group, if we consider the alternative resonance structure) with a chlorine atom (**44**) and a hydroxyl group significantly reduces the inhibitory activity against both TK isoforms (Figure 8a).

In compounds of general formula VI, the introduction of hydroxyalkyl, aminoalkyl, or carboxyalkyl substituents in position 9 (position R_2) of the purine ring, except for 2-hydroxyethyl (**45**) and 3-hydroxypentyl (**46**), increases both target properties. The introduction of 4-(piperidinyl)butyl and its derivatives containing a benzene moiety and acyclic substituents at positions 2, 3, and 4 of the pyridine ring has a similar effect. The only exceptions in the latter case are the two oxopurine derivatives with the general structural formula VI containing 4-(4-hydroxypyridyl)butyl and 4-(1,4'-bipyridine)butyl at position R_1. However, these two moieties have a negative effect only for the inhibition of the TK activity of HSV-1. The activities of HSV-2 TK inhibitors are not affected by these two modifications. The modification of the R_2 position in the oxopurine ring by replacing the hydrogen atom with 4-(decahydroquinolyl)butyl or 4-(1,2,3,4-tetrahydroquinolyl)butyl makes a positive contribution to both target properties (Figure 8b).

In oxopurine derivatives with the general formula VI containing a 4-hydroxyl group at position R_2, the replacement of the phenylamine moiety at position R_1 with a primary amino group or with a methylamine moiety significantly decreases the inhibitory activity against both TKs. Meanwhile, the introduction of a 2-phenoxyl or 2-phenylthiol moiety instead of 2-phenylaminyl moiety promotes the activity of inhibitors of HSV-1 TK, but negatively affects the inhibition efficiency of HSV-2 TK.

Overall, the comparison of experimental and calculated data indicates that the results of the structural analysis performed in GUSAR-2019 were 80% consistent with the results of previous biological studies.

The discrepancies in predicted estimates of the influence of structural descriptors on the target activities are observed only for the simulated structures containing bulky cyclic moieties, such as dibenzosuberene, NMe-acridine, thioxanthene dioxide, and their structural analogues. The mismatch is explained by the fact that the structural analysis in the GUSAR 2019 program is based on the 2D approach and, therefore, does not take into account steric features of the receptor that the activity of which the simulated compounds are intended to inhibit.

3. Discussion

In the present work, using the GUSAR 2019 program, we have modeled the quantitative structure–activity relationship for 89 TK inhibitors for HSV-1 and HSV-2 in the series of some carboxamide derivatives of 5'-amino-2',5'-dideoxy-5-ethyluridine, N2-phenylguanine, and 2-phenylamino-6-oxopurine with general structural formulas I–VI. The modeled TK inhibitors differed quite significantly in structure and belonged to different classes of organic compounds. In particular, compounds with general structural formulas I–III had a rather high degree of similarity to thymidine in the backbone structure. Compounds with general structural formulas IV–VI were more diverse and were actually structural analogues of guanine.

The modeling resulted in the construction of 12 valid QSAR consensus models focused on predicting target properties in the form of pIC_{50}. Each of these consensus models contains 20 to 100 partial regression relationships, which differ from each other by a set of descriptors. The validity of the use of structurally diverse TK inhibitors for modeling is confirmed based on the rather high numerical values of statistical criteria of the internal and external validation of QSAR models M1–M12. In particular, the high descriptive ability of the consensus models M1–M12 was confirmed based on the reliable prediction of activities

performed for compound structures of four training sets using two categories of metrics: (1) metrics based on R_2 coefficients of determination (R^2, R^2_0, $\overline{R^2_m}$, CCC); and (2) metrics based on errors in predicting pIC_{50} values (root mean square error (RMSEP), mean absolute error (MAE), standard deviation (SD)).

The predictive ability of QSAR models M1–M12 was evaluated using similar statistical criteria and prediction errors. Additionally, the criteria Q^2_{F1} and Q^2_{F2}, which are also used in the scientific literature to evaluate the predictive ability of QSAR/QSPR models, were determined. All models demonstrated rather high predictive ability in predicting target properties for both internal and external test set structures regardless of their size (95 and 100% of data).

This result is not an exception to the general rule, although it may be rather cautiously perceived by followers of the methodology of Gunch, Hammett, Taft, etc. In this context, note that the GUSAR program has been used for more than ten years since the release of its first version to build (Q)SAR/QSPR models focused on the detection and quantitative prediction of different types of biological activity. The developers of this program have repeatedly demonstrated in their publications that an important and undeniable advantage of their software product is the correct modeling of organic compounds that differ significantly in the structure and type of experimental studies. This important benefit of the GUSAR software is once again confirmed by the results of the present studies and is related to the unique algorithms used to calculate descriptors, as well as methods used to select the most significant descriptors for building the final QSAR models. In particular, the calculation of descriptors, the ideology of which is described in detail in the Section 4.1 and in the Supplementary Materials, is performed in the GUSAR program not only on the basis of whole molecules, but also on the basis of their individual structural parts, including individual atoms, as well as their various combinations. The calculation of descriptors based on the nature and properties of all atoms included in the modeled structures and their local environments is dominant in the descriptor-generation methodology, unlike the calculation of properties using whole molecular structures. This approach to the calculation of descriptors allows for common elements to be found among various organic compounds differing in the nature of cyclic and acyclic moieties, and, accordingly, expands the possibilities of QSAR modeling in general.

The ideology of the consensus approach, which actually takes into account the predictions of all partial regression relationships with a focus on their statistical weights, also significantly increases the reliability of adequate prediction of quantitative indicators of biological activity.

In addition, GUSAR program developers have repeatedly reported that QSAR models based on a diverse range of compound structures have a broader applicability in virtual screening than models based on a narrow set of multiple data. Such models allow for the identification and quantification of target properties for a broader class of organic compounds.

4. Computational Details

4.1. Computational Methodology

The QSAR analysis of HSV-1 and HSV-2 TK inhibitors was performed using the GUSAR 2019 (general unrestricted structure activity relationships) program [40–49]. In total, 12 QSAR models (M1–M12) were built.

The construction of QSAR models was performed using GUSAR 2019 software in several stages based on the training sets TrS1–TrS4. To validate these models, we used the external and internal test sets (TS1–TS2 and TS3–TS4, respectively).

4.2. Formation of Training and Test Sets

The training sets TrS1–TrS4 and external and internal test sets TS1–TS4 were formed from sets S1 and S2 according to the chart shown in Figure 9.
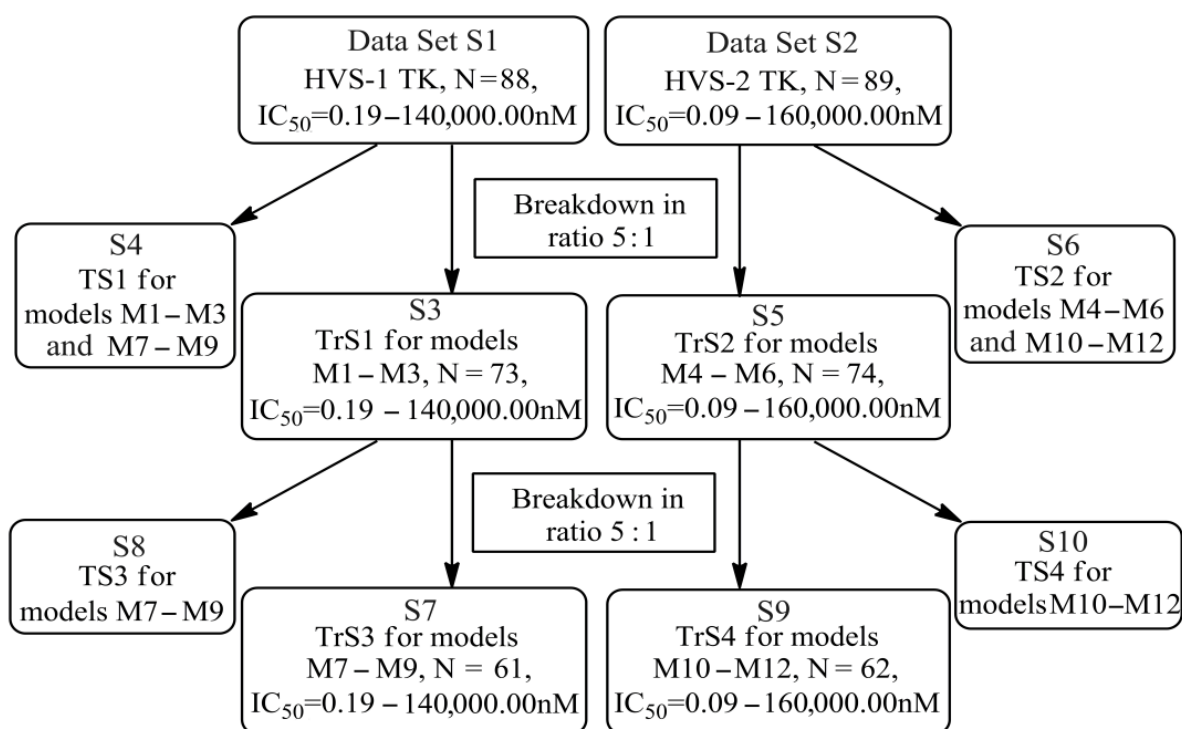
Figure 9. Chart of construction of the training and test sets and design of the QSAR consensus models M1–M12 (S is set, TrS and TS are training and test sets, respectively, N is the number of compounds included to the corresponding sets and arrays). Designations: (1) S1 and S2 are all datasets; (2) S3 is the training set TrS1 for models M1–M3; (3) S4 is the external test set TS1 for models M1–M3 and M7–M9; (4) S5 is the training set TrS2 for models M4–M6; (5) S6 is the external test set TS2 for models M4–M6 and M10–M12; (6) S7 is the training set TrS3 for models M7–M9; (7) S8 is the internal test set TS3 for models M7–M9; (8) S9 is the training set TrS4 for models M10–M12; (9) S10 is the internal test set TS4 for models M10–M12.

The datasets S1 and S2 comprised the same chemical structures (general structural formulas I–VI, Figure 1), the inhibitory activity of which against HSV-1 and HSV-2 TKs was quantitatively expressed as the IC_{50}. The IC_{50} values for these compounds were determined in earlier experimental studies [17,72,73]. The minor difference between the numbers of compounds in these sets is due to elimination of one compound with an inaccurately measured IC_{50} value from the set S1 ($IC_{50} > 10$ µmol/L).

The training set TrS1 was designed to build QSAR models M1–M3 and included 73 HSV-1 TK inhibitors. To assess the predictive power of M1–M3, we used the external test set TS1. Both of these sets were obtained by splitting the original data set S1 in a 5:1 ratio by moving every sixth chemical compound from S1 to TS1. Previously, all structures of S1 were ranked by an increasing IC_{50} (Figure 9).

The training set TrS2 contained 74 HSV-2 TK inhibitors. It was designed to build QSAR models M4–M6. To assess the validity of these models, we used external test set TS2. Both sets were obtained from the original S2 set in the same way as TrS1 and TS1.

The training sets TrS3 and TrS4 and internal test sets TS3 and TS4 were obtained by splitting TrS1 and TrS2 in a 5:1 ratio (the chemical structures of the sets were ranked by increasing IC_{50} values). A detailed description of training sets TrS1–TrS4 and test sets TS1–TS4 is presented in Tables 5 and 6, respectively. A comparison of the data in these tables indicated that the activity distribution of compounds in all training and test sets was almost identical. As a result, the average $\overline{pIC_{50}}$ values for HSV-1 and HSV-2 TK inhibitors were almost equal for TrS1–TrS4 and TS1–TS4.

Table 5. Statistical characteristics of the training sets TrS1–TrS4.

| Designation of TrS_i | Code of the Training Set | | | |
| | HSV-1 | | HSV-2 | |
	TrS1	TrS3	TrS2	TrS4
N	73	61	74	62
$\overline{pIC_{50}}$	6.788		6.921	
ΔpIC_{50}	5.867		6.250	
Thresholds used to evaluate the model's forecast				
$0.10 \times \Delta pIC_{50}$	0.587		0.625	
$0.15 \times \Delta pIC_{50}$	0.880		0.938	
$0.20 \times \Delta pIC_{50}$	1.174		1.250	
$0.25 \times \Delta pIC_{50}$	1.467		1.563	

Table 6. Statistical characteristics of the test sets TS1–TS4.

| Designation of TS_i | Code of the Test Set | | | |
| | HSV-1 | | HSV-2 | |
	TS1	TS3	TS2	TS4
N	15	12	15	12
$\overline{pIC_{50}}$	6.788		6.921	
ΔpIC_{50}	5.867		6.250	
Distribution of the observed response values of test sets TSi around the test mean				
$\overline{pIC_{50}} \pm 0.5$, %	26.667	16.667	20.000	25.000
$\overline{pIC_{50}} \pm 1.0$, %	40.000	41.667	40.000	41.667
$\overline{pIC_{50}} \pm 1.5$, %	60.000	58.333	46.667	50.000
$\overline{pIC_{50}} \pm 2.0$, %	73.333	83.333	66.667	66.667
Distribution of the observed response values of test sets TSi around the training mean				
$\overline{pIC_{50}} \pm 0.5$, %	13.333	8.333	26.667	16.667
$\overline{pIC_{50}} \pm 1.0$, %	33.333	25.000	33.333	41.667
$\overline{pIC_{50}} \pm 1.5$, %	46.667	50.000	46.667	50.000
$\overline{pIC_{50}} \pm 2.0$, %	66.667	75.000	66.667	75.000

The chemical structures of the compounds of TrS1–TrS4 and TS1–TS4 were created with the Marvin Sketch 17.22.0 program [74] and converted to the SDF format using the Discovery Studio Visualizer program [75]. To build QSAR models M1–M12, we used the IC_{50} values in mol/L, which were then converted to the pIC_{50} values:

$$pIC_{50} = -\log_{10}(IC_{50}) \tag{1}$$

Tables 5 and 6 show that the scatter of IC_{50} values $\Delta pIC_{50} > 5$ in the training sets is an important condition for constructing reliable QSAR models [76].

4.3. Building QSAR Models

QSAR models were developed using the GUSAR 2019 software. Chemical structures were described using three types of descriptors, the calculation of which is incorporated in this software: whole-molecule descriptors, 35 electro-topological descriptors (quantitative neighborhoods of atoms, QNAs), and 30 substructural descriptors (multilevel neighborhoods of atoms, MNAs). The whole-molecule descriptors used in GUSAR include the topological length, topological volume, and lipophilicity [40–49,55–63]. We will briefly explain the ideology of QNA and MNA descriptor formation, as it is rather new and unconventional in terms of classical QSAR approaches. A detailed description of the ideology of these computations is presented in the Supplementary Materials.

Formally, QNA descriptors represent the structure of a molecule using only two descriptors (P and Q). The P and Q values are calculated on the basis of the connectivity matrix (C) and atomic characteristics, such as the standard ionization potential (IP) and electron affinity (EA). The values for P and Q for each atom i are calculated as follows:

$$P_i = B_i \sum_k \left(\exp\left(-\frac{1}{2}C \right) \right)_{ik} B_k \tag{2}$$

$$Q_i = B_i \sum_k \left(\exp\left(-\frac{1}{2}C \right) \right)_{ik} B_k A_k \tag{3}$$

$$A_k = \frac{1}{2}(IP_k + EA_k) \tag{4}$$

$$B_k = (IP_k - EA_k)^{-1/2} \tag{5}$$

where k stands for all the other atoms in the molecule, IP is the first ionization potential, EA is the electron affinity for each atom (in eV), and C is the connectivity matrix for the molecule as a whole [42]. The standard IP and EA values of atoms in a molecule were taken from the literature.

Bivariate Chebyshev polynomials are used for the further approximation of P and Q functions over the entire junction structure. The regression equations use the averaged values of specific bivariate Chebyshev polynomials as independent variables. The averaging of these functions takes into account all atoms that are directly bonded to at least two other neighboring atoms. A detailed account of the calculation of QNA descriptors is presented in the Supplementary Materials.

In addition, GUSAR allows for the creation of QSAR models based on different biological activity profiles that are predicted for the compounds included in the training sets. All theoretically acceptable biological activities for the compounds used to build a QSAR model can be predicted with the PASS algorithm. The current version of PASS predicts more than 6000 types of biological activity with an average prediction accuracy of about 95%. This list includes pharmacotherapeutic effects, mechanisms of action, side and toxic effects, metabolic conditions, sensitivity to transporter proteins, and gene expression related activity. Adequate performance of the PASS algorithm is realized under conditions where each structure is represented as a list of MNA descriptors. Accordingly, before running the PASS algorithm, a set of MNA descriptors is preliminarily automatically calculated for each compound [40–42]. Thus, it is fair to say that in this case, regression models are based on MNA descriptors. The results of the PASS procedure for each type of biological activity are outputted in the form of a list of Pa-Pi parameter values, representing the difference between the probabilities that a compound is active (Pa) or inactive (Pi), respectively. Subsequently, the selection of independent variables necessary for the construction of regression relationships is performed automatically from this list at random. A detailed account of the calculation of MNA descriptors is also presented in the Supplementary Materials.

The self-consistent regression method (SCR) was applied to select the optimal number of descriptors for the QSAR models [20–22]. As previously reported by the developers of the GUSAR 2019 program, this makes it possible to remove the variables that poorly describe the target value. Additionally, this method is resistant to the noise in the data [55–58,60].

The SCR method of descriptor selection is a regularized least squares method based on the use of the mathematical apparatus of Bayesian statistics for the optimal estimation of regularization parameters and descriptor selection for the subsequent construction of regression relationships. The essence of the SCR method is the iterative selection of regularization coefficients v_i, first, to find the optimal number of descriptors in the regression equation, and second, to find the maximum values for regression coefficients a_i, and thus to obtain the maximum and, hence, reliable values of the dependent variable y for the training set compounds. As a result, the minimum error of quantitative prediction of the target

property is achieved. Unlike the multiple linear regression method, which is traditionally used to solve such problems, the SCR method, based on its ideology, does not impose restrictions on the number of regressors in the final regression equation or on the absence of a correlation (or the presence of a weak correlation) between them. Thus, the advantages of the SCR descriptor selection method over classical multiple linear regression are obvious. Unlike various heuristic approaches that solve multiple linear regression problems, descriptor selection in the SCR method is mathematically sound. Because of this, the SCR method can be successfully applied to remove variables that poorly describe the modeled activity value, while retaining a set of variables that correctly represent the existing relationships. The detailed mathematical apparatus on which the SCR method is based is presented in previous publications [47,55–58,60] and in the Supplementary Materials.

The GUSAR 2019 program allows for the building of both partial regression dependences and consensus models based on them. In this study, to reduce the variability of the final results, we used the consensus approach to build QSAR models. The final values of statistical criteria (coefficient of determination, Fisher's test, etc.) and the predicted pIC_{50} values for each molecule were estimated, in accordance with the consensus approach incorporated in the GUSAR 2019 software, as the weighted averages of these values derived from a set of partial QSAR models (for predictions that were within the respective areas of applicability). Meanwhile, each of these private models included in the consensus model was built independently based on a combination of QNA and/or MNA descriptors with the above three types of whole-molecule descriptors. This algorithm allowed us to combine the results of QSAR modeling based on different types of descriptors that are provided in the GUSAR 2019 software. As a result, we built 12 QSAR consensus models, which included 20 to 320 partial models. Since we used QSAR consensus models derived from dozens or even hundreds of single QSAR models, it is not possible to provide a general equation describing all selected variables. For this reason, the created QSAR consensus models could not provide information about positively and negatively influencing descriptors. Instead, GUSAR 2019 shows the positive and negative impact of each atom of the molecule on the predicted value [61]. An analysis of the effect of atoms on predicted pIC_{50} values and the search for general relationships between the structures of active compounds interacting with targets is described in this publication in Section 2.

4.4. Assessment of Applicability

GUSAR 2019 uses three different approaches to assess the applicability of the models: similarity, leverage, and accuracy assessment, which were described in detail previously [20–22,42]. A detailed description of these characteristics is given in the Supplementary Materials.

5. Evaluation of the Quality and Predictive Ability of QSAR Models

5.1. Calculating the pIC_{50} Values Using the Consensus Approach in the GUSAR 2019 Program

The descriptive and predictive ability of the M1–M12 consensus models was evaluated using the results of predicting pIC_{50} values for the structures included in the training sets TrS1–TrS4 and test sets TS1–TS4, respectively. For internal validation, a cross-validation control was used with a random twenty-fold exclusion of 20% of the structures from each training set.

5.2. Statistical Parameters Characterizing the Predictive Ability of QSAR Models

The predictive ability of QSAR models was estimated by predicting the pIC_{50} values for HSV-1 and HSV-2 TK inhibitors included in the external and internal test sets using two types of metrics:

(1) based on coefficients of determination R^2 (R^2, R^2_0, Q^2_{F1}, Q^2_{F2}, $\overline{R^2_m}$, CCC);
(2) estimating the prognostic errors of pIC_{50}: standard error (RMSEP), mean absolute error (MAE), and standard deviation (SD) [68–71,77].

Statistical parameters were calculated using the XternalValidationPlus 1.2 program [66,78]. The relevant formulas are presented in the Supplementary Materials. The same program

was used to identify systematic errors in the constructed consensus models. To avoid false predictions associated with outliers in experimental data, XternalValidationPlus 1.2 automatically removes 5% of compounds with high residuals.

In this work, when assessing the descriptive and predictive ability of the QSAR models, we mainly focused on the recommendations of Roy et al. [71]. We rated the descriptive and predictive power of each of the M_i QSAR models we developed as high if the following four conditions were met simultaneously:

(1) different coefficients of determination, calculated by comparing the experimental data with the calculated pIC_{50} data contained in each of the training and test sets, respectively, were numerically similar and tended to be 1;

(2) MAE values for predicted pIC_{50} of compounds of the training or test set, respectively, did not exceed 10% of the range of variation of the experimental pIC_{50} values for this set;

(3) the following relation held: $MAE + 3 \cdot SD_{TrS} \leq 0.2 \cdot pIC_{50\ TrS}$, where ΔpIC_{50} is the range of variation of pIC_{50} values for the TrS structures (this criterion refers to the assessment of the descriptive ability of the model);

(4) the following relation held: $MAE + 3 \cdot SD_{TrS} \leq 0.2 \cdot pIC_{50\ TrS}$, where ΔpIC_{50} is the range of variation of pIC_{50} values for the TrS structures (the criterion refers to the assessment of the predictive ability of the model).

We rated the descriptive and predictive ability of each of the M_i QSAR models we developed as low if the following conditions were met simultaneously:

(1) the numerical values of different coefficients of determination, calculated by comparing the experimental data with calculated pIC_{50}, did not exceed 0.6;

(2) MAE values estimated from the results of comparing the experimental and predicted pIC_{50} values of compounds of the training or test set, respectively, did not exceed 20% of the range of variation of the experimental pIC_{50} values in the training set used to build the M_i model;

(3) the following relation held: $MAE + 3 \cdot SD_{TrS} \geq 0.25 \cdot pIC_{50\ TrS}$, where ΔpIC_{50} is the range of variation of pIC_{50} values for the TrS structures (the criterion refers to the assessment of the descriptive ability of the model);

(4) the following relation held: $MAE + 3 \cdot SD_{TS} \geq 0.25 \cdot pIC_{50\ TrS}$, where ΔpIC_{50} is the range of variation of pIC_{50} values for the TrS structures (the criterion refers to the assessment of the predictive ability of the model).

The predictions not meeting any of the above conditions are considered moderate. If the QSAR model contained a systematic error, then it was excluded from consideration.

5.3. Evaluation of the Contribution of Atoms to the Target Activity

QSAR consensus models M3 and M6, containing 73 and 74 HSV-1 and HSV-2 TK inhibitors, respectively, were further used for assessing the contribution of atoms and functional groups to the simulated activity. It should be noted that this procedure is automatically implemented in the GUSAR 2019 program when calculating QNA descriptors and constructing QSAR models based on them.

6. Conclusions

Based on the QSAR methodology using the GUSAR 2019 program, a quantitative relationship between the structure and inhibitory activity against thymidine kinase of the herpes viruses HSV-1 and HSV-2 have been found in a series of 89 derivatives of 5-ethyluridine, N2-guanine, and 6-oxopurine. The inhibitory activities of the simulated compounds were in the range of $IC_{50} = 0.09$–160,000.00 nmol/L. Based on the MNA and QNA descriptors and whole-molecule descriptors using the self-consistent regression, we have constructed 12 statistically significant QSAR consensus models characterized by high accuracy of the prediction of pIC_{50} values for inhibitors of thymidine kinase of the herpes viruses HSV-1 and HSV-2 ($R^2_{TrS} > 0.6$; $Q^2_{TrS} > 0.5$; $R^2_{TS} > 0.5$). All of them can be used

for virtual screening of new TK inhibitors in a series of 5-ethyluridine, N2-guanine, and 6-oxopurine derivatives.

Thus, the approach implemented in the GUSAR 2019 program makes it possible to model, with a high degree of reliability, the inhibitory activity of derivatives of 5-ethyluridine, N2-guanine, and 6-oxopurine with respect to the TK of human herpes viruses in order to develop new inhibitors of this enzyme.

The approach implemented in the GUSAR 2019 program allows for the reliable simulation of the inhibitory activity of derivatives of 5-ethyluridine, N2-guanine, and 6-oxopurine against TK of human herpes viruses to develop new inhibitors of this enzyme. In addition, the correctness of the computational protocols used and the construction of regression models in the GUSAR 2019 program are confirmed by the absence of a systematic error in the calculations.

References

1. Sachs, S.L.; Straub, S.E.; Griffiths, P.D.; Whitley, R.J. *Clinical Management of Herpes Viruses*; Sachs, S.L., Straub, S.E., Griffiths, P.D., Whitley, R.J., Eds.; IOS: Washington, DC, USA, 1995; p. 398.
2. Tenser, R.B. Role of herpes simplex virus thymidine kinase expression in viral pathogenesis and latency. *J. Intervirol.* **1991**, *32*, 76–92. [CrossRef]
3. Jamieson, A.T.; Gentry, G.A.; Subak-Sharp, J.H. Induction of both thymidine and deoxycytidine kinase activity by herpes viruses. *J. Gen. Virol.* **1974**, *24*, 465–480. [CrossRef]

4. Kukhanova, M.K.; Korovina, A.N.; Kochetkov, S.N. Virus prostogo gerpesa cheloveka: Zhiznennyy tsikl i poisk ingibitorov. *J. Uspekhi Biol. Him.* **2014**, *54*, 457–494.

5. Richtin, T.; Black, M.; Mao, F.; Lewis, M.; Drake, R. Purification and Photoaffinity Labeling of Herpes Simplex Virus Type-1 Thymidine Kinase. *J. Biol. Chem.* **1995**, *270*, 7055–7060. [CrossRef]

6. Bello-Morales, R.; Crespillo, A.; Fraile-Ramos, A.; Tabares, E.; Alcina, A.; Lopez-Guerrero, J. Role of the small GTPase Rab27a during Herpes simplex virus infection of oligodendrocytic cells. *BMC Microbiol.* **2012**, *12*, 265–278. [CrossRef]

7. Schuppe, H.C.; Meinhardt, A.; Allam, J.P.; Bergmann, M.; Weidner, W.; Haidl, G. Chronic orchitis: A neglected cause of male infertility? *J. Androl.* **2008**, *40*, 84–91. [CrossRef]

8. Jiang, Y.-C.; Feng, H.; Lin, Y.-C.; Guo, X.-R. New strategies against drug resistance to herpes simplex virus. *J. Oral Sci.* **2016**, *8*, 1–6. [CrossRef]

9. Klein, R.; Czelusniak, S. Effect of a thymidine kinase inhibitor (L-653,180) on antiviral treatment of experimental herpes simplex virus infection in mice. *J. Antivir. Res.* **1990**, *14*, 207–214. [CrossRef]

10. Kim, C.U.; Luh, B.Y.; Misco, P.F.; Bisacchi, G.; Terry, B.; Mansuri, M.M. (2R, 4S, 5S)-1-(tetrahydro-4-hydroxy-5-methoxy-2-furanyl)thymine: A potent selective inhibitor of herpes simplex thymidine kinase. *J. Bioorg. Med. Chem. Lett.* **1993**, *3*, 1571–7576. [CrossRef]

11. Cheng, Y.C. A rational approach to the development of antiviral chemotherapy: Alternative substrates of herpes simplex virus Type 1 (HSV-1) and Type 2 (HSV-2) thymidine kinase (TK). *J. Ann. N. Y. Acad. Sci.* **1977**, *284*, 594–598. [CrossRef]

12. Watkins, A.M.; Dunford, P.J.; Moffatt, A.M.; Wong Kai-In, P.; Holland, M.J.; Pole, D.S.; Thomas, G.J.; Martin, J.A.; Roberts, N.A.; Mulqueen, M.J. Inhibition of virus-encoded thymidine kinase suppresses herpes simplex virus replication in vitro and in vivo. *J. Antivir. Chem. Chemother.* **1998**, *9*, 9–18.

13. Focher, F.; Hildebrand, C.; Freese, S.; Ciarrocchi, G.; Noonan, T.; Sangalli, S.; Brown, N.; Spadari, S.; Wright, G. N2-phenyldeoxyguanosine: A novel selective inhibitor of herpes simplex thymidine kinase. *J. Med. Chem.* **1988**, *31*, 1496–1500. [CrossRef]

14. Manikowski, A.; Lossani, A.; Savi, L.; Maioli, A.; Gambino, J.; Focher, F.; Spadari, S.; Wright, G.E. N2-Phenyl-9-(hydroxyalkyl)guanines and related compounds are substrates for Herpes simplex virus thymidine kinases. *J. Mol. Biochem.* **2012**, *1*, 21–25.

15. Nutter, L.M.; Grill, S.P.; Dutschman, G.E.; Sharma, R.A.; Bobek, M.; Cheng, Y.C. Demonstration of viral thymidine kinase inhibitor and its effect on deoxynucleotide metabolismin cells infected with herpes simplex virus. *J. Antimicrob. Agents Chemother.* **1987**, *31*, 368–374. [CrossRef]

16. Martin, J.A.; Thomas, G.J.; Merrett, J.H.; Lambert, R.W.; Bushnell, D.J.; Dunsdon, S.J.; Freeman, A.C.; Hopkins, R.A.; Johns, I.R.; Keech, E.; et al. The design, synthesis and properties of highly potent and selective inhibitors of herpes simplex virus types 1 and 2 thymidine kinase. *J. Antivir. Chem. Chemother.* **1998**, *9*, 1–8.

17. Martin, J.A.; Lambert, R.W.; Thomas, G.J.; Duncan, I.D.; Hall, M.J.; Merrett, J.H. Nucleoside analogues as highly potent and selective inhibitors of herpes simplex virus thymidine kinase. *J. Bioorg. Med. Chem. Lett.* **2001**, *11*, 1655–1658. [CrossRef]

18. Ferreira, M.M.C. Multivariate QSAR. *J. Braz. Chem. Soc.* **2002**, *13*, 742–753. [CrossRef]

19. Aremenko, N.V.; Baskin, I.I.; Palyulin, V.A.; Zefirov, N.S. Prediction of Physical Properties of Organic Compounds Using Artificial Neural Networks within the Substructure Approach. *J. Dokl. Chem.* **2001**, *381*, 317–320. [CrossRef]

20. Poroikov, V.V. Computer-aided drug design: From discovery of novel pharmaceutical agents to systems pharmacology. *J. Biochem. Mosc. Suppl. Ser. B Biomed. Chem.* **2020**, *14*, 216–227. [CrossRef]

21. Lagunin, A.A.; Rudik, A.V.; Pogodin, P.V.; Savosina, P.I.; Tarasova, O.A.; Dmitriev, A.V.; Ivanov, S.M.; Biziukova, N.Y.; Druzhilovskiy, D.S.; Filimonov, D.A.; et al. CLC-Pred 2.0: A Freely Available Web Application for In Silico Prediction of Human Cell Line Cyto-toxicity and Molecular Mechanisms of Action for Druglike Compounds. *Int. J. Mol. Sci.* **2023**, *24*, 1689. [CrossRef]

22. Muratov, E.N.; Bajorath, J.; Sheridan, R.P.; Tetko, I.V.; Filimonov, D.; Poroikov, V.; Oprea, T.I.; Baskin, I.I.; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *J. Chem. Soc. Rev.* **2020**, *49*, 3525–3564. [CrossRef]

23. Schaduangrat, N.; Lampa, S.; Simeon, S.; Gleeson, M.P.; Spjuth, O.; Nantasenamat, C. Towards reproducible computational drug discovery. *J. Cheminform.* **2020**, *2*, 4–30. [CrossRef]

24. Hartman, G.D.; Egbertson, M.S.; Halczenko, W.; Laswell, W.L.; Duggan, M.E.; Smith, R.L. Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. *J. Med. Chem.* **1992**, *35*, 4640–4642. [CrossRef]

25. Kim, C.U.; Lew, W.; Williams, M.A.; Liu, H.; Zhang, L.; Swaminathan, S. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: Design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **1997**, *119*, 681–690. [CrossRef]

26. Njoroge, F.G.; Chen, K.X.; Shih, N.Y.; Piwinski, J.J. Challenges in modern drug discovery: A case study of boceprevir, an HCV protease inhibitor for the treatment of hepatitis C virus infection. *Acc. Chem. Res.* **2008**, *41*, 50–59. [CrossRef]

27. McQuade, T.J.; Tomasselli, A.G.; Liu, L.; Karacostas, V.; Moss, B.; Sawyer, T.K. A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science* **1990**, *247*, 454–456. [CrossRef]

28. Ondetti, M.A.; Rubin, B.; Cushman, D.W. Design of specific inhibitors of angiotensin-converting enzyme: New class of orally active antihypertensive agents. *Science* **1977**, *196*, 441–444. [CrossRef]

29. Cushman, D.W.; Cheung, H.S.; Sabo, E.F.; Ondetti, M.A. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* **1977**, *16*, 5484–5491. [CrossRef]

30. Cohen, N.C. Structure-based drug design and the discovery of aliskiren (Tekturna): Perseverance and creativity to overcome a R&D pipeline challenge. *Chem. Biol. Drug Des.* **2007**, *70*, 557–565. [CrossRef]

31. Sokouti, B.; Hamzeh-Mivehroud, M. 6D-QSAR for predicting biological activity of human aldose reductase inhibitors using quasar receptor surface modeling. *BMC Chem.* **2023**, *17*, 1–9. [CrossRef]

32. Damale, M.G.; Harke, S.N.; Kalam Khan, F.A.; Shinde, D.B.; Sangshetti, J.N. Recent advances in multidimensional QSAR (4D-6D): A critical review. *Mini Rev. Med. Chem.* **2014**, *14*, 35–55. [CrossRef]

33. Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524. [CrossRef]

34. Giordano, D.; Biancaniello, C.; Argenio, M.; Facchiano, A. Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals* **2022**, *15*, 646. [CrossRef]

35. Ab, A.; Bhatt, H. 3D-QSAR (CoMFA, CoMFA-RG, CoMSIA) and molecular docking study of thienopyrimidine and thienopyridine derivatives to explore structural requirements for aurora-B kinase inhibition. *Eur. J. Pharm. Sci.* **2015**, *79*, 1–12. [CrossRef]

36. Ankitkumar, P.; Hardik, B.; Bhumika, P. Structural insights on 2-phenylquinazolin-4-one derivatives as tankyrase inhibitors through CoMFA, CoMSIA, topomer CoMFA and HQSAR studies. *J. Molec. Struct.* **2022**, *1249*, 131636. [CrossRef]

37. Duraiswami, C.; Madhav, P.J.; Hopfinger, A.J. Application of 4D-QSAR Analysis to a Set of Prostaglandin, PGF2α, Analogs. In *Molecular Modeling and Prediction of Bioactivity*; Springer: Boston, MA, USA, 2000; pp. 323–324. [CrossRef]

38. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, *23*, 2139–2149. [CrossRef]

39. Vedani, A.; Dobler, M.; Lill, M.A. Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor. *J. Med. Chem.* **2005**, *48*, 3700–3703. [CrossRef]

40. Zakharov, A.V.; Peach, M.L.; Sitzmann, M.; Nicklaus, M.C. A New Approach to Radial basis function approximation and Its application to QSAR. *J. Chem. Inf. Model.* **2014**, *54*, 713–719. [CrossRef]

41. Zakharov, A.V.; Peach, M.L.; Sitzmann, M.; Nicklaus, M.C. QSAR modeling of imbalanced high-throughput screening data in PubChem. *J. Chem. Inf. Model.* **2014**, *54*, 705–712. [CrossRef]

42. Lagunin, A.; Zakharov, A.; Filimonov, D.; Poroikov, V. QSAR Modelling of Rat Acute Toxicity on the Basis of PASS Prediction. *J. Mol. Inform.* **2011**, *30*, 241–250. [CrossRef]

43. Filimonov, D.A.; Zakharov, A.V.; Lagunin, A.A.; Poroikov, V.V. QNA based "Star Track" QSAR approach. SAR QSAR En-viron. *J. Resolut.* **2009**, *20*, 679–709. [CrossRef]

44. Zakharov, A.V.; Lagunin, A.A.; Filimonov, D.A.; Poroikov, V.V. Quantitative structure—Activity relationships of cyclin-dependent kinase 1 inhibitors. *J. Biomed. Chem.* **2006**, *52*, 3–18. [CrossRef]

45. Filimonov, D.A.; Akimov, D.V.; Poroikov, V.V. The Method of Self-Consistent Regression for the Quantitative Analysis of Relationships Between Structure and Properties of Chemicals. *Pharm. Chem. J.* **2004**, *38*, 21–24. [CrossRef]

46. Ivanov, S.M.; Lagunin, A.A.; Filimonov, D.A.; Poroikov, V.V. Relationships between the structure and severe drug-induced liver injury for low, medium, and high doses of drugs. *J. Chem. Res. Texicol.* **2022**, *35*, 402–411. [CrossRef]

47. Lagunin, A.A.; Zakharov, A.V.; Filimonov, D.A.; Poroikov, V.V. A new approach to QSAR modelling of acute toxicity. *J. SAR QSAR Environ. Res.* **2007**, *18*, 285–298. [CrossRef]

48. Lagunin, A.A.; Geronikaki, A.; Eleftheriou, P.; Pogodin, P.V.; Zakharov, A.V. Rational Use of Heterogeneous Data in Quantitative Structure–Activity Relationship (QSAR) Modeling of Cyclooxygen-ase/Lipoxygenase Inhibitors. *J. Chem. Inf. Mod.* **2019**, *59*, 713–730. [CrossRef]

49. Zakharov, A.V.; Varlamova, E.V.; Lagunin, A.A.; Dmitriev, A.V.; Muratov, E.N.; Fourches, D.; Kuz'min, V.E.; Poroikov, V.V.; Tropsha, A.; Nicklaus, M.C. QSAR Modeling and Prediction of Drug–Drug Interactions. *J. Mol. Pharm.* **2016**, *13*, 545–556. [CrossRef]

50. Tarasova, O.A.; Urusova, A.F.; Filimonov, D.A.; Nicklaus, M.C.; Zakharov, A.V.; Poroikov, V.V. QSAR Modeling Using Large-Scale Databases: Case Study for HIV-1 Reverse Transcriptase Inhibitors. *J. Chem. Inf. Mod.* **2015**, *55*, 1388–1399. [CrossRef]

51. Tarasova, O.A.; Rudik, A.V.; Ivanov, S.M.; Lagunin, A.A.; Poroikov, V.V.; Filimonov, D.A. Machine Learning Methods in Antiviral Drug Discovery. In *Topics in Medicinal Chemistry*; Tarasova, O.A., Rudik, A.V., Ivanov, S.M., Lagunin, A.A., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 37, pp. 245–279. [CrossRef]

52. Kokurkina, G.V.; Dutov, M.D.; Shevelev, S.A.; Popkov, S.V.; Zakharov, A.V.; Poroikov, V.V. Synthesis, antifungal activity and QSAR study of 2-arylhydroxynitroindoles. *Eur. J. Med. Chem.* **2011**, *46*, 4374–4382. [CrossRef]

53. Masand, V.H.; Mahajan, D.T.; Patil, K.N.; Dawale, N.E.; Hadda, T.B.; Alafeefy, A.A.; Chinchkhede, K.D. General Unrestricted Structure Activity Relationships based evaluation of quinoxaline derivatives as potential influenza NS1A protein inhibitors. *Der Pharma Chem.* **2011**, *3*, 517–525.

54. Masand, V.H.; Devidas, T.; Mahajan, D.T.; Patil, K.N.; Hadda, T.B.; Youssoufi, M.H.; Jawarkar, R.D.; Shibi, I.G. Optimization of Antimalarial Activity of Synthetic Prodiginines: QSAR, GUSAR, and CoMFA analyses. *J. Chem. Biol. Drug Des.* **2013**, *81*, 527–536. [CrossRef]

55. Khairullina, V.R.; Gerchikov, A.Y.; Lagunin, A.A.; Zarudii, F.S. QSAR modeling of thymidilate synthase inhibitors in a series of quinazoline derivatives. *J. Pharm. Chem.* **2018**, *51*, 884–888. [CrossRef]

56. Khairullina, V.R.; Gerchikov, A.Y.; Zarudii, F.S. Analysis of the relationship "structure cyclooxygenase-2 inhibitory activity" in the series of di-tret-butylphenol, oxazolone and thiazolone. *J. Vestn. Bashk. Univ.* **2014**, *19*, 417–422.

57. Khayrullina, V.R.; Gerchikov, A.Y.; Lagunin, A.A.; Zarudii, F.S. Quantitative Analysis of Structure−Activity Relationships of Tetrahydro-2H-isoindole Cyclooxygenase-2 Inhibitors. *J. Biokhimiya* **2015**, *80*, 74–86. [CrossRef]

58. Khairullina, V.R.; Akbasheva, Y.Z.; Gimadieva, A.R.; Mustafin, A.G. Analysis of the relationship «structure-activity» in theseries of certain 5-ethyluridine derivatives with pronounced anti-herpetic activity. *J. Vestn. Bashk. Univ.* **2017**, *22*, 960–965.

59. Martynova, Y.Z.; Khairullina, V.R.; Nasretdinova, R.N.; Garifullina, G.G.; Mitsukova, D.S.; Gerchikov, A.Y.; Mustafin, A.G. Determination of the chain termination rate constants of the radical chain oxidation of organic compounds on antioxidant molecules by the QSPR method. *J. Russ. Chem. Bull.* **2020**, *69*, 1679–1691. [CrossRef]

60. Khairullina, V.; Safarova, I.; Sharipova, G.; Martynova, Y.; Gerchikov, A. QSAR Assessing the Efficiency of Antioxidants in the Termination of Radical-Chain Oxidation Processes of Organic Compounds. *J. Mol.* **2021**, *26*, 421. [CrossRef]

61. Khairullina, V.; Martynova, Y.; Safarova, I.; Sharipova, G.; Gerchikov, A.; Limantseva, R.; Savchenko, R. QSPR Modeling and Experimental Determination of the Antioxidant Activity of Some Polycyclic Compounds in the Radical-Chain Oxidation Reaction of Organic Substrates. *J. Mol.* **2022**, *27*, 6511. [CrossRef]

62. Martynova, Y.Z.; Khairullina, V.R.; Garifullina, G.G.; Mitsukova, D.S.; Zarudiy, F.S.; Mustafin, A.G. QSAR-modeling of the relationship "structure—Antioxidative activity" in a series of some benzopirane and benzofurane derivatives. *J. Vestn. Bashk. Univ.* **2019**, *24*, 573–580. [CrossRef]

63. Martynova, Y.Z.; Khairullina, V.R.; Gerchikov, A.Y.; Zarudiy, F.S.; Mustafin, A.G. QSPR-modeling of antioxidant activity of potential and industrial used stabilizers from the class of substituted alkylphenols. *J. Vestn. Bashk. Univ.* **2020**, *25*, 723–730. [CrossRef]

64. Oguri, T.; Achiwa, H.; Bessho, Y.; Muramatsu, H.; Maeda, H.; Niimi, T.; Sato, S.; Ueda, R. The role of thymidylate synthase and dihydropyrimidine dehydrogenase in resistance to 5-fluorouracil in human lung cancer cells. *J. LungCan.* **2005**, *49*, 345–351. [CrossRef]

65. McGuire, J.J. Anticancer Antifolates: Current Status and Future Directions. *J. Cur. Pharm. Des.* **2003**, *9*, 2593–2613. [CrossRef]

66. Roy, K.; Das, R.N.; Ambure, P.; Aher, R.B. Be aware of error measures. Further studies on validation of predictive QSAR models. *J. Chemom. Intell. Lab. Syst.* **2016**, *152*, 18–33. [CrossRef]

67. Ivanov, A.S.; Veselovsky, A.V.; Dubanov, A.V.; Skvortsov, V.S.; Archakov, A.I. The integral platform "From gene to drug prototype" in silico and in vitro. *J. Ross. Khim. Zh.* **2006**, *1*, 18–35.

68. Gramatica, P.; Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. *J. Chem. Inform. Model.* **2016**, *56*, 1127–1131. [CrossRef]

69. Consonni, V.; Ballabio, D.; Todeschini, R. Evaluation of model predictive ability by external validation techniques. *J. Chemom.* **2010**, *24*, 194–201. [CrossRef]

70. Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models: How to Evaluate It? Comparisonof Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. *J. Chem. Inform. Model.* **2011**, *51*, 2320–2335. [CrossRef]

71. Roy, K.; Mitra, I.; Kar, S.; Ojha, P.K.; Das, R.N.; Kabir, H. Comparative Studies on Some Metrics for External Validation of QSPR Models. *J. Chem. Inform. Model.* **2012**, *52*, 396–408. [CrossRef]

72. Hildebrand, C.; Sandoli, D.; Focher, F.; Gambino, J.; Ciarrocchi, G.; Spadari, S.; Wright, G. Structure-activity relationships of N2-substituted guanines as inhibitors of HSV1 and HSV2 thymidine kinases. *J. Med. Chem.* **1990**, *33*, 203–206. [CrossRef]

73. Manikowski, A.; Lossani, A.; Verri, A.; Gebhardt, B.-M.; Gambino, J.; Focher, F.; Spadari, S.; Wright, G.E. Inhibition of Herpes Simplex Virus Thymidine Kinases by 2-Phenylamino-6-oxopurines and Related Compounds: Structure-Activity Relationships and Antiherpetic Activity in Vivo. *J. Mol. Biochem.* **2006**, *48*, 3919–3929. [CrossRef]

74. MarvinSketch. Available online: https://chemaxon.com/download/marvin-suite (accessed on 31 August 2023).

75. DiscoveryStudioVisualiser. Available online: https://www.3ds.com (accessed on 31 August 2023).

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

77. Roy, P.P.; Paul, S.; Mitra, I.; Roy, K. On Two Novel Parameters for Validation of Predictive QSAR Models. *J. Mol.* **2009**, *14*, 1660–1701. [CrossRef]

78. Xternal Validation Plus. Available online: https://sites.google.com/site/dtclabxvplus (accessed on 31 August 2023).