



Molecular features of thymidine analogues governing the activity of human thymidine kinase

Karolina Jagiello¹ · Samanta Makurat² · Sylwester Pereć¹ · Janusz Rak² · Tomasz Puzyn¹

Received: 10 January 2018 / Accepted: 1 May 2018 / Published online: 11 May 2018
© The Author(s) 2018

Abstract

Modified uridines seem to be able to work as hypoxic tumour cell radiosensitisers. Before they sensitise cells to ionising radiation, they have to be incorporated into the genomic DNA and the latter process has to be preceded by the phosphorylation of the modified uridine, which in human cells is executed by human thymidine kinase 1 (hTK1). In the current study, we present the quantitative structure–activity relationship (QSAR) model allowing to identify and understand the molecular features of nucleoside derivatives governing the hTK1 kinase activity. The developed model meets all requirements of a reliable QSAR model and is based on only two molecular properties: the shape of the nucleoside determined by atom substitutions and the ability of the molecule to intermolecular interactions with the enzyme. These results have important implications for the rational designing of new hTK1 substrates and should significantly reduce the time and cost of studies on new radiosensitisers.

Keywords Human thymidine kinase 1 · QSAR model · Nucleoside analogues · Deoxythymidine · Substrate specificity · Molecular descriptors

Introduction

The efficacy of radiotherapy, one of the most common modalities used to combat cancer [1], is impaired by the fact that solid tumour cells suffer from hypoxia, which make them radioresistant [2]. In order to overcome this therapeutically undesirable effect, radiosensitisation is necessary. One may mention several classes of cellular radiosensitisers that have been investigated so far [3], of which thymine analogues as 5-bromouracil (BrU)/5-iodouracil (IU) [4] and other

halogenated nucleobases [5, 6] seem to be especially promising since they work under hypoxia. Their radiosensitising action is related to the irreversible and swift elimination of halide anions (X^-) from the modified nucleobase anion [7, 8] that is formed due to solvated electron (e^-_{hyd}) attachment. The latter species, the second (beside hydroxyl radicals) most abundant product of water radiolysis [9], are amply stable under hypoxia, but native DNA is not sensitive to them [5]. This situation is different for the DNA with incorporated radiosensitiser, where the elimination of X^- leaves a reactive nucleoside radical in the biopolymer. As a consequence, secondary reactions comprising hydrogen atom transfer within the radical nucleoside [10, 11] or between the radical and adjacent nucleoside [12] lead ultimately to a serious DNA damage such as single-strand breaks [13], cyclopurine lesions [14] and DNA cross-links [15].

Despite the abovementioned advantages of halogenated nucleosides, they are not routinely employed in anticancer treatment. Even in 5-bromodeoxyuridine (BrdU)/5-iododeoxyuridine (IdU), the derivatives most thoroughly investigated in clinical trials were not introduced into practice due to only marginal therapeutic effects in patients [16, 17]. Probably one of the reasons that inhibit the development of this otherwise tempting approach to anticancer treatment is the lack of electrophilic

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11224-018-1124-2>) contains supplementary material, which is available to authorized users.

✉ Janusz Rak
janusz.rak@ug.edu.pl

✉ Tomasz Puzyn
t.puzyn@qsar.eu.org

¹ Department of Environmental Chemistry and Radiochemistry, Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 90-308 Gdańsk, Poland

² Department of Physical Chemistry, Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 90-308 Gdańsk, Poland

nucleosides with appropriately optimised features. Recently, we have undertaken attempts for rational designing of 5-substituted uracils prone to the dissociative electron attachment (DEA) process [18, 19]. As a result, we proposed a series of 5-modified uracils with the DEA characteristics much better than those of BrdU [18, 19] and the superior characteristics of some of them were confirmed in the photoelectron spectroscopy [18] and radiolytic studies [20, 21].

Here, it should be emphasised that in order to become a part of the cellular DNA, after translocation from the extracellular environment to the cytosol, a nucleoside has to be phosphorylated to its triphosphate. The first stage of phosphorylation, the conversion of the nucleoside to the nucleoside monophosphate, is usually the rate-limiting step [22]. Hence, the favourable DEA profile [18, 19] is necessary, but not a sufficient requirement for an efficient radiosensitising nucleoside. Taking into account that chemical synthesis is frequently not a trivial task and usually is a time-consuming and costly process, it is difficult to overestimate access to a computational model that could tell, yet before the actual synthesis, if the proposed derivative is a good substrate for nucleoside kinase. Basically, one can employ a variant of the quantum mechanics/molecular mechanics (QM/MM) approach to model the first, most demanding phosphorylation step of a studied nucleoside by hTK1 kinase. However, this is definitely not a simple task. Indeed, the experimental structures of the hTK1 kinase are not suitable for reaction mechanism studies since hTK1 changes its quaternary structure upon binding substrates [23]. On the other hand, the active site amino acids are flexible [24], which makes accurate diffraction measurement for the enzymatic pocket impossible. Out of three available crystal structures [24–26], only one monomer in one of those structures includes full amino acid sequence of the active site. Thus, in order to obtain the active hTK1 geometry, one has to complete missing amino acid sequence, dock proper substrates, and carry out a long molecular dynamics (MD) simulation that should lead to a ligand-hTK1 complex suitable for the QM/MM calculations. Only a representative structure from the MD simulation, with the proper choice of high-layer atoms and additional constraints, if necessary, would enable the QM/MM calculations to be completed. Moreover, a few enzymatic mechanisms for phosphate transfer have been proposed so far [27, 28]: dissociative, associative or concerted and they all should be modelled to decide which of the pathways is most probable. Although the computational procedure described above would bring us a direct insight into the enzymatic process under consideration as well as allow to select nucleosides suitable for phosphorylation, it is computationally very demanding and needs up to dozen or so months to be completed. Taking into account the fact that usually a significant number of possible derivatives have to

be scrutinised, a less time-consuming approach is desirable. Therefore, quantitative structure-activity relationship (QSAR) seems to be a method of choice for selecting the most promising nucleoside derivatives since to build a statistical model and choose appropriate compounds, one needs only a fraction of time compared to the described above QM/MM methodology. It is interesting to mention, that recently, the “intelligent” consensus modelling approach was proposed [29]. This approach integrates models that are developed by means of different combinations of descriptors and/or different modelling methods; in consequence, prediction is based on multiple individual models (each based on individual set of descriptors) instead of single regression equation. It offers improvements of the predictability of the model [29]. On the other hand, due to the fact that consensus model is based on large number of descriptors, its mechanistic interpretation is frequently more difficult (in comparison to simple QSAR model) [30]. In terms of the Organisation for Economic Cooperation and Development principles related to QSAR model development and validation [31], the benefits related to improvement of the model’s statistics are as valuable as the complexity and interpretability power of the model.

In the following, we will demonstrate the development of a simple correlation model that predicts with an acceptable accuracy the ease of various uridine derivatives to be phosphorylated by human thymidine kinase 1 (hTK1). The derived correlation equation depends on three descriptors only, which enables the observed activities to be rationalised in terms of the molecular features of the studied nucleosides.

Methodology

Data collection

The experimental data related to the hTK1 phosphorylation activity was collected from the available literature [32, 33] and was logarithmically transformed to decrease the range of data variation. The experimental data were available for 26 nucleoside analogues. Details related to the structures of analysed nucleosides one can find in Table S1 (Supplementary materials).

Molecular descriptors calculation

Geometries of nucleoside analogues in the trans-conformation were optimised with the Gaussian09 software [34], at the B3LYP/6-31++G(d,p) density functional theory level with the polarisable continuum model [35] to account for water environment. After the optimisation, the molecular descriptors were calculated with the use of the

DRAGON software [36]. We have calculated 957 descriptors for each nucleoside.

QSAR model development and validation

In the first step, the set of 26 nucleosides was sorted according to the increasing values of log A. Then, data was divided into training (T—1) and validation (V—2) sets with the usage of the “3:1 algorithm”, in which every the third compound is assigned to the validation set, whereas the remaining ones form the training set. The second and first before the last compounds were independently assigned to validation set.

QSAR model was developed with the application of the multiple linear regression (MLR) approach [37, 38]. We assumed that the modelled activity (log A) could be expressed as a function of molecular descriptors (x_1, x_2, x_3, \dots):

$$\log A = a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n + b \quad (1)$$

where $a_1, a_2, a_3, \dots, a_n$ are the regression coefficients and b is the intercept.

The optimal descriptors were selected in two steps: first, to reduce the descriptor-compound ratio, we selected descriptors significantly correlated with log A ($r > 0.60$), and then we applied genetic algorithm (GA) [39] implemented in the QSARINS software [40]. The purpose of the genetic algorithm application was to find the descriptors that allow to obtain model with the highest validation

and cross-validation parameters. The setup of GA was as follows: generation per size = 500 and mutation rate = 45%.

The model fitting, robustness and predictive abilities were evaluated based on the parameters summarised in Table 1 [41–46]. Model’s robustness was also verified by the application of the Y-scrambling procedure. The Williams plot technique was employed to determine the QSAR model applicability domain [47–50]. Additionally, in order to select an optimal model (based on optimal set of descriptors), the double cross-validation procedure was performed with the double cross-validation tool (version 2.0) developed by Roy and Ambure [51].

Results and discussion

The multiple linear regression technique was employed in order to develop the quantitative structure-activity relationship model (QSAR) [37, 38]. This model allows to identify the physicochemical features of nucleoside derivatives governing the hTK1 kinase activity, which should enable new hTK1 substrates to be designed rationally and tested before synthesis.

As it was mentioned in the “Methodology” section, the experimental data related to the hTK1 phosphorylation activity were collected from the available literature [32, 33] and were logarithmically transformed to decrease the range of data variation (Table 2).

Table 1 Quality measures for QSAR models [41–46]

| | Goodness of fit | Robustness | Predictive ability |
|-------------------------|---|---|---|
| Correlation coefficient | $R^2 = 1 - \frac{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{pred}})^2}{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{obs}})^2}$ | $Q_{CV}^2 = 1 - \frac{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{predcv}})^2}{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{obs}})^2}$ | $Q_{\text{Ext}}^2 = 1 - \frac{\sum_{j=1}^k (y_j^{\text{obs}} - y_j^{\text{pred}})^2}{\sum_{j=1}^k (y_j^{\text{obs}} - y_j^{\text{obs}})^2}$ |
| Root mean square error | $\text{RMSE}_C = \sqrt{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{pred}})^2}$ | $\text{RMSE}_{CV} = \sqrt{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{predcv}})^2}$ | $\text{RMSE}_P = \sqrt{\sum_{j=1}^k (y_j^{\text{obs}} - y_j^{\text{pred}})^2}$ |
| CCC | $\text{CCC} = \frac{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{x} - \bar{y})^2}$ | | |
| Modified r^2 | $r_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2} \right)$ | | |

y_i^{obs} means the experimental (observed) value of the property for the i th compound from training set, y_i^{pred} the predicted value of the property for the i th compound from training set, y_i^{predcv} the predicted value of the property for the temporarily included (cross-validated) i th compound from training set, y_j^{obs} the mean experimental value of the property in the training set, n the number of compounds in the training set, y_j^{obs} the experimental (observed) value of the property for the j th compound from validation set, y_j^{pred} the predicted value of the property for j th compound from validation set, y_j^{obs} the mean experimental value of the property in the validation set, k the number of compounds in the validation set, x and y the abscissa and ordinate values of the graph plotting the prediction experimental data values vs the ones calculated using the model, n the number of compounds and \bar{X} and \bar{Y} the averages of abscissa and ordinate values, respectively; squared correlation coefficient values between the observed and predicted values of the compounds with/without intercept (r^2/r_0^2)

Table 2 Experimental and predicted values of log A with data split into a training set and a validation set as well as other QSAR model details [32, 33]

| Substrate | P_VSA_LogP_6 | HATS4v | E3u | log A | T/V | log A _p | Leverages | Residuals |
|---|--------------|--------|-------|-------|-----|--------------------|-----------|-----------|
| 2'-Deoxythymidine (T) | 27.818 | 0.160 | 0.437 | 2 | 1 | 1.77 | 0.09 | 1.25 |
| 2'-Deoxyuridine (U) | 27.818 | 0.175 | 0.508 | 1.89 | 1 | 2.06 | 0.17 | -1.12 |
| 3'-Fluoro-5-ethynyl-2'-deoxyuridine (FEU) | 94.901 | 0.178 | 0.408 | 1 | 1 | 0.96 | 0.46 | 0.28 |
| 3'-Fluoro-methyl-2'-deoxythymidine (FMT) | 42.778 | 0.175 | 0.397 | 1.18 | 1 | 1.50 | 0.14 | -1.99 |
| 3'-Azido-methyl-2'-deoxythymidine (AZMT) | 55.520 | 0.151 | 0.354 | 1.18 | 1 | 1.14 | 0.32 | 0.26 |
| 3',5-Difluoro-2'-deoxyuridine (FFU) | 57.738 | 0.182 | 0.455 | 1.3 | 1 | 1.56 | 0.10 | -1.55 |
| 3-(2-Propynyl)-2'-deoxythymidine (3-PropT) | 79.941 | 0.125 | 0.482 | 1.32 | 1 | 1.24 | 0.43 | 0.69 |
| 3-(2,3-Dihydroxypropyl)-2'-deoxythymidine (3-DHPT) | 27.818 | 0.095 | 0.463 | 1.44 | 1 | 1.67 | 0.21 | -1.57 |
| 3'-Fluoro-5-cyclopropyl-2'-deoxyuridine (FCPU) | 42.778 | 0.142 | 0.452 | 1.52 | 1 | 1.60 | 0.07 | -0.50 |
| 2',3'-Dideoxythymidine (HHT) | 27.818 | 0.155 | 0.436 | 1.6 | 1 | 1.75 | 0.09 | -0.91 |
| 3-Methyl-2'-deoxythymidine (3-MeT) | 27.818 | 0.131 | 0.378 | 1.65 | 1 | 1.47 | 0.27 | 1.17 |
| 3'-Azido-2'-deoxythymidine (AZT) | 27.818 | 0.173 | 0.416 | 1.72 | 1 | 1.73 | 0.13 | -0.13 |
| 3-Ethyl-2'-deoxythymidine (3-EtT) | 27.818 | 0.111 | 0.471 | 1.79 | 1 | 1.74 | 0.16 | 0.22 |
| 3'-Azido-5-iodo-2'-deoxyuridine (AZIU) | 59.796 | 0.230 | 0.467 | 1.85 | 1 | 1.72 | 0.25 | 0.84 |
| 1-(2'-Deoxy-2'-fluoro-1-D-arabinofuranosyl)-5-iodouracil (FIAU) | 42.778 | 0.225 | 0.457 | 1.88 | 1 | 1.86 | 0.19 | 0.07 |
| 5-Bromo-2'-deoxyuridine (5-BrU) | 27.818 | 0.203 | 0.483 | 1.9 | 1 | 2.06 | 0.17 | -1.03 |
| 5-Ethyl-2'-deoxyuridine (5-EtU) | 27.818 | 0.144 | 0.432 | 1.9 | 1 | 1.70 | 0.10 | 1.06 |
| 3- <i>n</i> -Butyl-2'-deoxythymidine (3- <i>n</i> BuT) | 27.818 | 0.090 | 0.517 | 1.93 | 1 | 1.84 | 0.36 | 0.62 |
| 5-Fluoro-2'-deoxyuridine (5-FU) | 42.778 | 0.184 | 0.499 | 1.98 | 1 | 1.89 | 0.14 | 0.49 |
| 5-Chloro-2'-deoxyuridine (5-ClU) | 27.818 | 0.199 | 0.486 | 2.29 | 1 | 2.05 | 0.17 | 1.40 |
| 3-Isopropyl-2'-deoxythymidine (3- <i>Is</i> T) | 27.818 | 0.101 | 0.425 | 1.23 | 2 | 1.55 | 0.21 | -1.73 |
| 3'-Fluoro-2'-deoxythymidine (FLT) | 42.778 | 0.159 | 0.418 | 1.48 | 2 | 1.53 | 0.08 | -0.32 |
| 1-(2'-Deoxy-2'-fluoro-1-D-arabinofuranosyl)-thymine (FMAU) | 42.778 | 0.172 | 0.424 | 1.68 | 2 | 1.59 | 0.07 | 0.41 |
| 3'-Azido-2'-deoxyuridine (AZU) | 59.796 | 0.193 | 0.501 | 1.85 | 2 | 1.73 | 0.22 | 0.56 |
| 3-Benzyl-2'-deoxythymidine (3-BnT) | 27.818 | 0.114 | 0.475 | 1.92 | 2 | 1.77 | 0.15 | 0.73 |
| 5-Iodo-2'-deoxyuridine (5-IU) | 27.818 | 0.214 | 0.476 | 2.23 | 2 | 2.06 | 0.21 | 0.80 |

log A is the experimentally measured activity of hTK [29]; log A_p is the predicted activity of hTK

The QSAR model is described by Eq. (2):

$$\begin{aligned}
 \log A &= 0.073 - 0.011 P_VSA_LogP_6 + 2.92 HATS4v \\
 &+ 3.52 E3un \\
 &= 20, n_{val} = 6, \quad F = 16.53, \quad p < 10^{-4}, \quad R^2 \\
 &= 0.75 Q_{CV}^2 = 0.65 Q_{Ext}^2 = 0.71, RMSE_C \\
 &= 0.17, RMSE_{CV} = 0.19, RMSE_P = 0.17, r_m^2 \\
 &= 0.55, CCC = 0.86
 \end{aligned} \tag{2}$$

where n and n_{val} stand for the number of compounds in the training and validation set, respectively, while the remaining parameters are defined in Table 1.

The visual correlation between the experimentally measured and predicted values of log A for the training (T) and

validation (V) sets, presented in Fig. 1, confirms the high quality of the model. All parameters listed below Eq. (2) and employed to verified model's fitting, robustness and predictive abilities (calculated by means of quality measurement summarised in Table 1, [41–46]) meet the required criteria and, therefore, also proves the model's quality. The Y-scrambling procedure (Fig. 2) additionally confirms that obtained model was not a result of by chance correlation.

The applicability domain of the QSAR model was verified based on the plot of the standardised residuals (differences between the predicted and observed values of log A) versus the leverages, the so-called Williams plot (Fig. 3) [47–50]. The leverage value, in this case, expresses similarity of a given compound, for which the prediction is made to the set of training compounds. Thus, the Williams plot helps to assess the influence of structural similarity between the compounds on the predicting error. All deoxythymidine derivatives used in the training and validation sets were situated in the range of

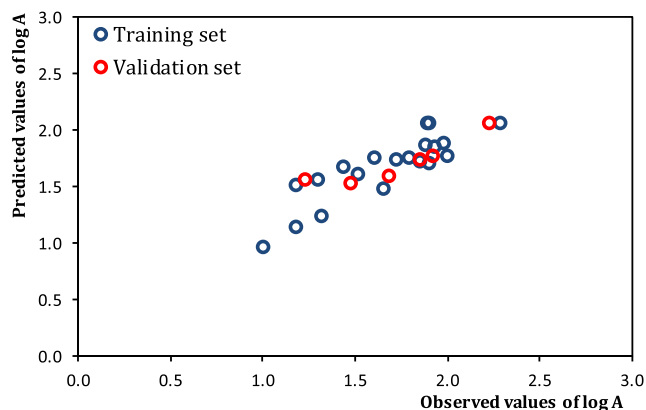


Fig. 1 Calculated vs. observed values of log A

residuals differing by ± 3 standard deviations from the mean value and do not exceed calculated leverage threshold for this model. Thus, there were no outlying predictions observed.

The QSAR model depicted by Eq. (2) is a linear combination of three descriptors (descriptors are not correlated to each other, Fig. S1 in the Supplementary materials), namely: P_VSA_logP_6, HATS4v and E3u. The selection of the optimal model (based on the optimal set of descriptors) was confirmed by the double cross-validation procedure [51] (Table S2 in the Supplementary materials). The conformation changes of a molecule as well as the total number and bond lengths influence the value of the HATS4v and E3u descriptors. Thus, both indirectly provide information related to the molecular shape. The P_VSA descriptor, on the other hand, expresses the van der Waals surface area (VSA) that is occupied by molecule atoms having given property in a certain range [52, 53]. Several properties could be taken into account in P_VSA descriptor calculations, such as atomic weight and ionisation potential. In the case of P_VSA_logP_6, lipophilicity is considered. Thus, this descriptor represents the size of VSA occupied by the atoms with lipophilicity within the specified range and thus encodes the availability of molecular fragments for intermolecular hydrophobic interactions [53].

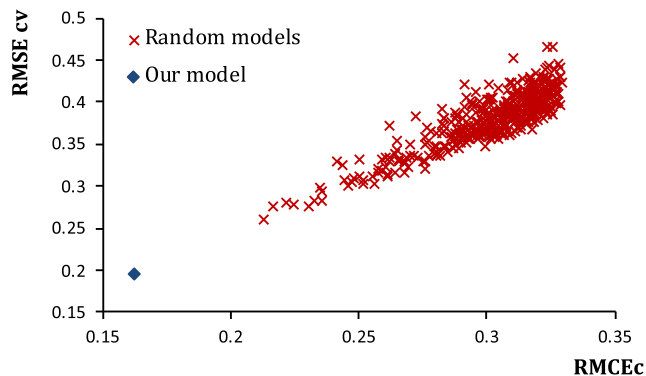


Fig. 2 Y-scrambling results: average values of the square errors of calibration and cross-validation of the real QSAR model and 400 random models

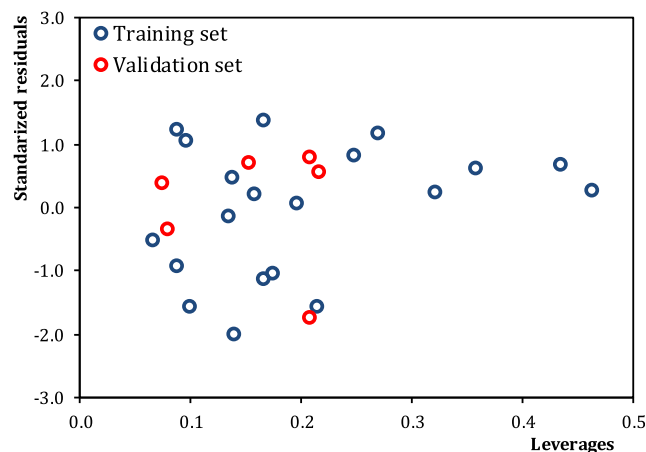


Fig. 3 Williams plot: standardised residuals versus leverages. Solid lines indicate ± 3 standard deviation units. Dash lines indicates the threshold value ($h^* = 0.60$)

The worked out QSAR model (Eq. (2)) indicates that both impact of atoms on the shape of the nucleoside analogues, coded in HATS4v and E3u descriptors, and the size of their VSA capable of intermolecular hydrophobic interactions, described by P_VSA_logP_6, are the key properties that influence the hTK1 phosphorylation activity against deoxythymidine derivatives. The sign of the equation coefficient related to the P_VSA_logP_6 descriptor is negative while those related to HATS4v and E3u are positive. This indicates that increasing the values of P_VSA_logP_6 results in decreasing the values of the hTK1 activity while increasing the values of HATS4v and E3u causes the increase in its activity (compounds with higher values of P_VSA_logP_6 and lower values of HATS4v and E3u are worse substrates for hTK1).

Among the nucleoside analogues employed for the development of the QSAR model, there are derivatives substituted at deoxyribose: 3'-OH (e.g. AZMT) and 2' analogues (e.g. FMAU) as well as molecules with the N3 (e.g. IsoT) and 5C modification of pyrimidine (e.g. 5-ClIdU; Table 2). The highest activities are observed for the compounds substituted with halogens at the 5C position. These derivatives exhibit the highest values of both the HATS4v and E3u descriptors (see Table 2). Modification of the N3 position with an alkyl residue results in the decreased values of both descriptors and in consequence decreases the hTK1 activity. It indicates that substitution at the N3 position with the alkyl group changes the shape of a molecule. It was proven that substitution in this position hinders the hydrogen bonding between the N3 nitrogen of pyrimidine base and the main chain carbonyl of residue 178 in the enzyme that is required to tight spacing of the lasso-like loop [32, 33, 54]. Substitution at deoxyribose is also not favourable. 3'-OH analogues as well as 2' derivatives exhibit lower activity than dT. The only exception is FIAU. However, this nucleoside is substituted not only at the 2' position but also at the 5 position of pyrimidine. The lower activity (higher values of P_VSA_logP_6 and lower values of HATS4v and E3u) of 3'-

OH derivatives is probably related to the lack of hydrogen bonds that cannot be formed in the ligand-enzyme complex between the substituted 3'-OH group and the amino group of Gly182 [32, 33, 54–56]. 2'-Fluoro derivative of deoxythymidine differs from dT mostly in the value of P_VSA_logP_6 while the HATS4v and E3u descriptors remain almost the same. This indicates that the shape of the FMAU is, in this case, not changed and that the addition of fluorine in 2' position affects intermolecular interactions. This agrees with the finding that interactions between the 2'-OH group and Tyr187 are hindered when the nucleoside is substituted at the 2' position [32]. Indeed, the Tyr187 nucleoside 2'-OH interactions are necessary to keep the lasso in place [32].

Considering that the most promising group of analogues constitutes 5C derivatives, we will finally focus on this class of nucleosides. It can be noticed that the changes in the activity of halogen derivatives follow the following pattern: 5-ClIdU > 5-IdU > 5-BrdU > 5-FdU and the worst substrate is the unsubstituted nucleoside, i.e. dT. When one compares the values of descriptors calculated for all these molecules, one finds out that the HATS4v and E3u descriptors which code the size and shape of the studied analogues have the smallest values, equal to 0.160 and 0.437, respectively, for dT (see Table 2). P_VSA_logP_6 for 5-halogenated derivatives are similar (except that for 5-FU), which means that they are similar in terms of intermolecular interactions. The impact of size of substituent at position 5 on the phosphorylation reaction was previously proven. The proper size of a substituent at the 5 position is required to fit the binding pocket in the hTK1 enzyme [32]. However, due to the fact that activities of the iodo and methyl analogues are different beside similar size of 5-substituents, the steric hindrance is not sufficient to explain the observed differences in activity. The other important issue is their ability to intermolecular interactions. The polarisability of iodine is much larger than that of the other analysed 5-substituents [57], which might be responsible for the stronger interaction between the nucleoside and enzyme, and in consequence makes 5-IdU the best (among the analysed compound) substrate for hTK1.

Conclusion

Due to the limited availability of structural data concerning hTK1, the significant flexibility of its enzymatic pocket, complex phosphorylation mechanism and high computational cost of QM/MM simulations, showing that a given nucleoside is a suitable substrate for hTK1 with calculations at the atomic level, seem to be more than difficult. Therefore, we have developed a QSAR model that allows to identify the molecular features of thymidine analogues governing their activity against hTK1. Two properties turned out to be the key features responsible for the phosphorylation process: the shape of the

nucleoside determined by atom substitution coded in the HATS4v and E3u descriptors and the ability of molecules to hydrophobic interactions coded by the P_VSA_logP_6 descriptor. The model meets all requirements related to QSAR model's development and therefore can be applied to predict the activity of new nucleoside before its synthesis. The obtained results also indicate that the most promising analogues are those substituted at the 5C position. Designing a new substrate for hTK1 one should focus on the modification of uridine at this position. A valuable substituent needs to meet two requirements: be at least as large as methyl group and be able to interact strongly with the enzyme-binding pocket.

Funding information This work was supported by the Polish National Science Centre [Grant No. UMO-2014/14/A/ST4/00405] (J.R.). K.J. and T.P. thank the Polish Ministry of Science and Higher Education [Grant No. DS 530-8637-D510-17]. This research was supported in part by PL-Grid Infrastructure.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA Cancer J Clin* 61(2):69–90. <https://doi.org/10.3322/caac.20107>
2. Oronsky BT, Knox SJ, Scicinski J (2011) Six degrees of separation: the oxygen effect in the development of radiosensitizers. *Transl Oncol* 4(4):256–265. <https://doi.org/10.1593/tlo.11166>
3. Wardman P (2007) Chemical radiosensitizers for use in radiotherapy. *Clin Oncol-UK* 19(6):397–417. <https://doi.org/10.1016/j.clon.2007.03.010>
4. Shewach DS (2006) Nucleoside radiosensitizers. In: Godefridus JP (ed) *Deoxynucleoside analogs in cancer therapy*. Humana Press Inc., New York
5. Westphal K, Wiczak J, Miloch J, Kciuk G, Bobrowski K, Rak J (2015) Irreversible electron attachment—a key to DNA damage by solvated electrons in aqueous solution. *Org Biomol Chem* 13(41):10362–10369. <https://doi.org/10.1039/c5ob01542a>
6. Westphal K, Skotnicki K, Bobrowski K, Rak J (2016) Radiation damage to single stranded oligonucleotide trimers labelled with 5-iodopyrimidines. *Org Biomol Chem* 14(39):9331–9337. <https://doi.org/10.1039/c6ob01713d>
7. Chomicz L, Rak J, Storoniak P (2012) Electron-induced elimination of the bromide anion from brominated nucleobases. A computational study. *J Phys Chem B* 116(19):5612–5619. <https://doi.org/10.1021/jp3008738>

8. Wieczor M, Wityk P, Czub J, Chomicz L, Rak J (2014) A first-principles study of electron attachment to the fully hydrated bromonucleobases. *Chem Phys Lett* 595:133–137. <https://doi.org/10.1016/j.cplett.2014.01.052>
9. Lehnert S (2008) Biomolecular action of ionizing radiation. Series in medical physics and biomedical engineering. Taylor & Francis, Boca Raton
10. Chomicz L, Leszczynski J, Rak J (2013) Electron-induced degradation of 8-Bromo-2'-deoxyadenosine 3',5'-diphosphate, a DNA radiosensitizing nucleotide. *J Phys Chem B* 117(29):8681–8688. <https://doi.org/10.1021/jp4022689>
11. Chomicz L, Furmanchuk A, Leszczynski J, Rak J (2014) Electron induced single strand break and cyclization: a DFT study on the radiosensitization mechanism of the nucleotide of 8-bromoguanine. *Phys Chem Chem Phys* 16(14):6568–6574. <https://doi.org/10.1039/c3cp55411b>
12. Schyman P, Zhang RB, Eriksson LA, Laaksonen A (2007) Hydrogen abstraction from deoxyribose by a neighbouring uracil-5-yl radical. *Phys Chem Chem Phys* 9(45):5975–5979. <https://doi.org/10.1039/b711083a>
13. Resendiz MJE, Pottiboyina V, Sevilla MD, Greenberg MM (2012) Direct strand scission in double stranded RNA via a C5-pyrimidine radical. *J Am Chem Soc* 134(8):3917–3924. <https://doi.org/10.1021/ja300044e>
14. Chatgililoglu C, Ferreri C, Terzidis MA (2011) Purine 5',8-cyclonucleoside lesions: chemistry and biology. *Chem Soc Rev* 40(3):1368–1382. <https://doi.org/10.1039/c0cs00061b>
15. Dextraze ME, Cecchini S, Bergeron F, Girouard S, Turcotte K, Wagner JR, Hunting DJ (2009) Reaching for the other side: generating sequence-dependent Interstrand cross-links with 5-bromodeoxyuridine and gamma-rays. *Biochemistry-Us* 48(9):2005–2011. <https://doi.org/10.1021/bi801684t>
16. Prados MD, Scott CB, Rotman M, Rubin P, Murray K, Sause W, Asbell S, Comis R, Curran W, Nelson J, Davis RL, Levin VA, Lamborn K, Phillips TL (1998) Influence of bromodeoxyuridine radiosensitization on malignant glioma patient survival: a retrospective comparison of survival data from the Northern California Oncology Group (NCOG) and Radiation Therapy Oncology Group trials (RTOG) for glioblastoma, multiforme and anaplastic astrocytoma. *Int J Radiat Oncol* 40(3):653–659. [https://doi.org/10.1016/S0360-3016\(97\)00770-0](https://doi.org/10.1016/S0360-3016(97)00770-0)
17. Prados MD, Seiferheld W, Sandler HM, Buckner JC, Phillips T, Schultz C, Urtasun R, Davis R, Gutin P, Cascino TL, Greenberg HS, Curran WJ (2004) Phase III randomized study of radiotherapy plus procarbazine, lomustine, and vincristine with or without BUdR for treatment of anaplastic astrocytoma: final report of RTOG 9404. *Int J Radiat Oncol* 58(4):1147–1152. <https://doi.org/10.1016/j.ijrobp.2003.08.024>
18. Chomicz L, Zdrowowicz M, Kasprzykowski F, Rak J, Buonaugurio A, Wang Y, Bowen KH (2013) How to find out whether a 5-substituted uracil could be a potential DNA radiosensitizer. *J Phys Chem Lett* 4(17):2853–2857. <https://doi.org/10.1021/jz401358w>
19. Makurat S, Chomicz-Manka L, Rak J (2016) Electrophilic 5-substituted uracils as potential radiosensitizers: a density functional theory study. *ChemPhysChem* 17(16):2572–2578. <https://doi.org/10.1002/cphc.201600240>
20. Zdrowowicz M, Chomicz L, Zyndul M, Wityk P, Rak J, Wiegand TJ, Hanson CG, Adhikary A, Sevilla MD (2015) 5-Thiocyanato-2'-deoxyuridine as a possible radiosensitizer: electron-induced formation of uracil-C5-thiyl radical and its dimerization. *Phys Chem Chem Phys* 17(26):16907–16916. <https://doi.org/10.1039/c5cp02081f>
21. Sosnowska M, Makurat S, Zdrowowicz M, Rak J (2017) 5-Selenocyanatouracil: a potential hypoxic radiosensitizer. Electron attachment induced formation of selenium centered radical. *J Phys Chem B* 121(25):6139–6147. <https://doi.org/10.1021/acs.jpcc.7b03633>
22. Rose JD, Parker WB, Someya H, Shaddix SC, Montgomery JA, Secrist JA (2002) Enhancement of nucleoside cytotoxicity through nucleotide prodrugs. *J Med Chem* 45(20):4505–4512. <https://doi.org/10.1021/jm020107s>
23. Segura-Pena D, Lichter J, Trani M, Konrad M, Lavie A, Lutz S (2007) Quaternary structure change as a mechanism for the regulation of thymidine kinase 1-like enzymes. *Structure* 15(12):1555–1566. <https://doi.org/10.1016/j.str.2007.09.025>
24. Welin M, Kosinska U, Mikkelsen NE, Camrot C, Zhu CY, Wang LY, Eriksson S, Munch-Petersen B, Eklund H (2004) Structures of thymidine kinase 1 of human and mycoplasmic origin. *Proc Natl Acad Sci USA* 101(52):17970–17975. <https://doi.org/10.1073/pnas.0406332102>
25. Birringer MS, Claus MT, Folkers G, Kloer DP, Schulz GE, Scapozza L (2005) Structure of a type II thymidine kinase with bound dTTP. *FEBS Lett* 579(6):1376–1382. <https://doi.org/10.1016/j.febslet.2005.01.034>
26. Segura-Pena D, Lutz S, Monnerjahn C, Konrad M, Lavie A (2007) Binding of ATP to TK1-like enzymes is associated with a conformational change in the quaternary structure. *J Mol Biol* 369(1):129–141. <https://doi.org/10.1016/j.jmb.2007.02.104>
27. Klahn M, Rosta E, Warshel A (2006) On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins. *J Am Chem Soc* 128(47):15310–15323. <https://doi.org/10.1021/ja065470t>
28. Alberts IL, Wang Y, Schlick T (2007) DNA polymerase beta catalysis: are different mechanisms possible? *J Am Chem Soc* 129(36):11100–11110. <https://doi.org/10.1021/ja071533b>
29. Roy K, Ambure P, Kar S, Ojha PK (2018) Is it possible to improve the quality of predictions form an “intelligent” use of multiple QSAR/QSPR/QSTR models? *J Chemom* 194:1–18
30. Hewitt M, Cronin MTD, Madden JC, Rowe PH, Johnson C, Obi A, Enoch J (2007) Consensus QSAR models: do the benefits outweigh the complexity? *J Chem Inf Model* 47:1460–1468
31. The Report from the Expert Group on (QuantitatiVe) Structure ActiVity Relationship ([Q]SARs) on the Principles for the Validation of (Q)SARs; Organisation for economic cooperation and development: Paris, France, 2004
32. Kosinska U, Camrot C, Eriksson S, Wang L, Eklund H (2005) Structure of the substrate complex of thymidine kinase from *Ureaplasma urealyticum* and investigations of possible drug targets for the enzyme. *FEBS J* 272(24):6365–6372. <https://doi.org/10.1111/j.1742-4658.2005.05030.x>
33. Bandyopadhyaya AK, Johnsamuel J, Al-Madhoun AS, Eriksson S, Tjarks W (2005) Comparative molecular field analysis and comparative molecular similarity analysis of human thymidine kinase 1 substrates. *Bioorg Med Chem* 13:1681–1690
34. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson BA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd J, Brothers EN, Kudin KN, Staroverov VN, Kobayashi RN, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, revision D.01. Gaussian, Inc., Wallingford
35. Miertus S, Scrocco E, Tomasi J (1981) Electrostatic interaction of a solute with a continuum—a direct utilization of abinitio molecular

- potentials for the prevision of solvent effects. *Chem Phys* 55(1): 117–129. [https://doi.org/10.1016/0301-0104\(81\)85090-2](https://doi.org/10.1016/0301-0104(81)85090-2)
36. DRAGON T (2016) Software for molecular descriptor calculation. Milano, Italy
 37. Schultz TW, Cronin MTD, Netzeva TI (2003) The present status of QSAR in toxicology. *J Mol Struct-Theochem* 622(1–2):23–38. [https://doi.org/10.1016/S0166-1280\(02\)00615-2](https://doi.org/10.1016/S0166-1280(02)00615-2)
 38. Cronin MTD, Schultz TW (2003) Pitfalls in QSAR. *J Mol Struct-Theochem* 622(1–2):39–51. [https://doi.org/10.1016/S0166-1280\(02\)00616-4](https://doi.org/10.1016/S0166-1280(02)00616-4)
 39. Cho SJ, Hermsmeier MA (2002) Genetic algorithm guided selection: variable selection and subset selection. *J Chem Inf Comp Sci* 42(4):927–936. <https://doi.org/10.1021/ci010247v>
 40. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S (2013) QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. *J Comput Chem* 34(24):2121–2132. <https://doi.org/10.1002/jcc.23361>
 41. Gramatica P (2013) On the development and validation of QSAR models. *Methods Mol Biol* 930:499–526. https://doi.org/10.1007/978-1-62703-059-5_21
 42. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 29(6–7):476–488. <https://doi.org/10.1002/minf.201000061>
 43. Chirico N, Gramatica P (2011) Real external Predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J Chem Inf Model* 51(9):2320–2335. <https://doi.org/10.1021/ci200211n>
 44. Chirico N, Gramatica P (2012) Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J Chem Inf Model* 52(8):2044–2058. <https://doi.org/10.1021/ci300084j>
 45. Roy K, Mitra I (2011) On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb Chem High Throughput Screen* 14(6):450–474. <https://doi.org/10.2174/138620711795767893>
 46. Ojha PK, Mitra I, Das RN, Roy K (2011) Further exploring r(m)(2) metrics for validation of QSPR models. *Chemom Intell Lab Syst* 107(1):194–205. <https://doi.org/10.1016/j.chemolab.2011.03.011>
 47. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *Atla-Altern Lab Anim* 33(5):445–459
 48. Jagiello K, Sosnowska A, Walker S, Haranczyk M, Gajewicz A, Kawai T, Suzuki N, Leszczynski J, Puzyn T (2014) Direct QSPR: the most efficient way of predicting organic 435 carbon/water partition coefficient (log KOC) for polyhalogenated POPs. *Struct Chem* 25:997–436 1004. <https://doi.org/10.1007/s11224-014-0419-1>
 49. Bielińska-Wąż D, Wąż P, Jagiełło K, Puzyn T (2014) Spectral density distribution moments 391 as novel descriptors for QSAR/QSPR. *Struct Chem* 25:29–35. <https://doi.org/10.1007/s11224-013-0229-x>
 50. Jagiello K, Grzonkowska M, Swirog M, Ahmed L, Rasulev B, Avramopoulos A, Papadopoulos MG, Leszczynski J, Puzyn T (2016) Advantages and limitations of classic and 431 3D QSAR approaches in nano-QSAR studies based on biological activity of fullerene 432 derivatives. *J Nanopart Res* 18:256. <https://doi.org/10.1007/s11051-016-3564-1>
 51. Roy K, Ambure P (2016) The “double cross-validation” software tool for MLR QSAR model development. *Chemom Intell Lab Syst* 159:108–126
 52. Labute P (2000) A widely applicable set of descriptors. *J Mol Graph Model* 18(4–5):464–477. [https://doi.org/10.1016/S1093-3263\(00\)00068-1](https://doi.org/10.1016/S1093-3263(00)00068-1)
 53. Todeschini R, Consonni V (2006) Molecular descriptors for chemoinformatics. Wiley-VCH
 54. Al-Madhoun AS, Tjarks W, Eriksson S (2004) The role of thymidine kinases in the activation of pyrimidine nucleoside analogues. *Mini-Rev Med Chem* 4(4):341–350
 55. Johansson NG, Eriksson S (1996) Structure-activity relationships for phosphorylation of nucleoside analogs to monophosphates by nucleoside kinases. *Acta Biochim Pol* 43(1):143–160
 56. Carnrot C, Wehelie R, Eriksson S, Bolske G, Wang LY (2003) Molecular characterization of thymidine kinase from *Ureaplasma urealyticum*: nucleoside analogues as potent inhibitors of mycoplasma growth. *Mol Microbiol* 50(3):771–780. <https://doi.org/10.1046/j.1365-2958.2003.03717.x>
 57. Auffinger P, Hays FA, Westhof E, Ho PS (2004) Halogen bonds in biological molecules. *Proc Natl Acad Sci USA* 101(48):16789–16794. <https://doi.org/10.1073/pnas.0407607101>