Scientia Pharmaceutica

ÖSTERREICHISCHE PHARMAZEUTISCHE GESELLSCHAFT

MDPI

Article

Assessment of Anti-Influenza Activity of Pyrimidin-4(3*H*)-one Derivatives Using Prediction Models

Yakov V. Gorokhov [1], Alexey D. Egorov [1,2], Nadezhda M. Andriyashina [1], Alexander N. Lobov [1], Irina S. Balashova [3], Aleksandrina S. Volobueva [4], Stanislav A. Grabovsky [1] and Sophia S. Borisevich [1,*]

[1] Ufa Institute of Chemistry, Ufa Federal Research Center of the Russian Academy of Sciences, 450054 Ufa, Russia; yakov.v.gorokhov@gmail.com (Y.V.G.); adegorov@mephi.ru (A.D.E.); andriyashina@anrb.ru (N.M.A.); lobovan@anrb.ru (A.N.L.); stas_g@anrb.ru (S.A.G.)
[2] Institute of Cyber Intelligence Systems, National Research Nuclear University MEPHI, 115409 Moscow, Russia
[3] Faculty of Biology, St. Petersburg State University, 199034 Saint Petersburg, Russia; wolshmir@gmail.com
[4] Saint-Petersburg Pasteur Institute, 197101 Saint Petersburg, Russia; sasha-khrupina@mail.ru
* Correspondence: monrel@mail.ru

Abstract

In this study, we used our own prediction models to assess the antiviral potential of pyrimidin-4(3*H*)-one derivatives against the A/H1N1 influenza virus strain. This assessment allows us to identify promising structures. The models are based on machine learning algorithms and molecular modeling results. In general, the prediction results are consistent with experimental data. The most promising compound, namely **12** (6-amino-2-(dimethylamino)pyrimidin-4(3*H*)-one), inhibits the reproduction of the A/Puerto Rico/8/34 (H1N1) influenza virus strain in vitro, likely by affecting the function of the endonuclease domain of the viral polymerase complex. Compound **12** can be used to create new PAN inhibitors by modifying its structure.

Keywords: pyrimidine-4(3*H*)-one derivatives; influenza virus A/H1N1; prediction models; machine learning; molecular modeling

Academic Editors: Osvaldo Andrade Santos-Filho and Arthur Eugen Kümmerle

Received: 30 May 2026
Revised: 29 June 2026
Accepted: 1 July 2026
Published: 14 July 2026

1. Introduction

The pyrimidine nitrogenous base of pyrimidin-4(3H)-one derivatives is one of the most well-known pharmacophores in medical chemistry [1] and is found in the structures of many biologically active drugs. Although uracil is most often associated with cancer treatments, there are many other uracil-based compounds that can be used to treat a variety of diseases, including infectious diseases caused by various pathogenic viruses [2,3], including the influenza virus [4].

In 1996, the authors of [5] reported a new antiviral agent, Flutimide **1**, which inhibits the replication of infectious influenza viruses of different strains (Figure 1). The mechanism of the antiviral action of compound **1** and its analogue **2** is the inhibition of the RNA-dependent RNA polymerase of the influenza virus [6]. Then, the authors of [7] described a structural analysis of the mechanism of inhibition of influenza virus endonuclease (PA) by small molecules, among which the antiviral agent **3** inhibits the replication of the infectious influenza virus of strain A/PR/8/34 (H1N1) at micromolar concentrations. It should be noted that RNA-dependent RNA polymerase is a heterotrimeric complex that includes the subunits of endonuclease PA and polymerases PB1 and PB2, which are associated with the 3′ and 5′ ends of each segment of the RNA genome [8]. Inhibition of the endonuclease activity of the PA viral protein affects the function of RNA-dependent

RNA polymerase and allows PA to be considered as an attractive target for the development of new antiviral drugs. The authors of [9] reported a series of phenyl-substituted 4-hydroxypyridazin-3(2*H*)-ones and 5-hydroxypyrimidin-4(3H)-ones: the inhibitors of influenza virus endonuclease, among which compounds 4 and 5 are the leaders. Derivatives of 1.6-bis[(benzyloxy)methyl]uracin [4] were initially considered as potential inhibitors of HIV-1 reverse transcriptase. However, antiviral agents, which include the lead compound **6**, also inhibit the replication of the influenza A/H1N1 virus (Figure 1). Although the authors report that, the mechanism of antiviral action of the studied compounds is unclear.

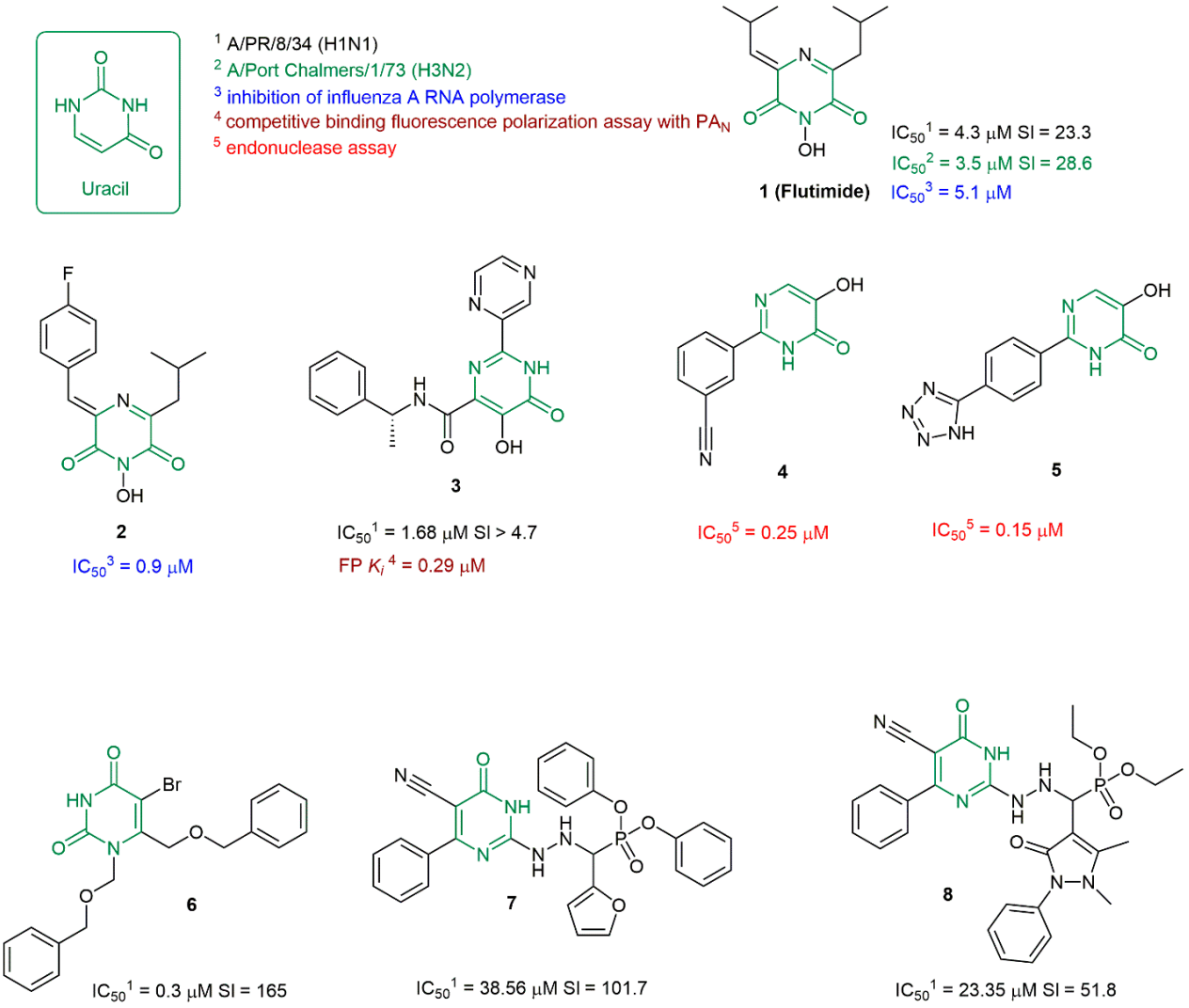


Figure 1. Compounds that contain a pyrimidine scaffold and inhibit viral proteins and/or exhibit antiviral activity against the infectious influenza virus. The key pyrimidine scaffold is shown in green. The colors of the antiviral activity values correspond to the color used to highlight the biological experiment method.

Finally, a recent publication [10] examines the mechanism of the antiviral activity of pyrimidinone-based α-aminophosphonates against influenza A/H1N1 viruses and HCoV-229E coronavirus. The structures of lead compounds **7** and **8**, whose selectivity indices (SI) exceed 50, are shown in Figure 1. Based on the molecular docking results, the antiviral agents under study can inhibit the function of influenza virus neuraminidase and/or RNA-dependent RNA polymerase. Thus, while analyzing even a small number of scientific publications, we can once again confirm the possibility of antiviral properties of compounds containing a pyrimidine scaffold. The following papers [4,6,7,9,10] share not only a common

fragment of the compounds under study, but also an approach to the search for new antiviral agents, which consists of the synthesis of new molecules followed by an assessment of their inhibitory activity through in vitro tests. Such an approach requires significant physical and financial costs and, unfortunately, very often leads to unsatisfactory results. Thus, in [4], among the thirteen synthesized compounds, only five structures can be noted as promising (with SI values > 8) in relation to HIV-1, and only two (among which is the leader compound **6**) in relation to the influenza A/H1N1 virus. A similar situation is observed in [10], where among the twelve synthesized compounds, only one is characterized by a selectivity index greater than 8 (SI (HCoV-229E) = 24.5) when assessing their antiviral activity against the HCoV-229E coronavirus, and two (compounds **7** and **8** in Figure 1) have such an index when assessing their antiviral activity against the influenza A/H1N1 virus. To be fair, it should be noted that it was found in [9] that most of the studied compounds inhibit influenza virus endonuclease at micromolar concentrations. However, the fact that a compound inhibits the function of a viral protein does not necessarily mean that it will suppress the replication of the infectious virus. The above examples further confirm the well-known fact that the search for promising biologically active compounds most often begins with experimental high-throughput screening, which requires significant physical and financial resources [11]. The limitations of traditional pharmacological methods can be overcome using the structure-based design of biologically active molecules, in which the development of theoretical models for predicting the biological potential of small molecules and/or de novo design have demonstrated their effectiveness [12]. The most interesting and desirable outcome of applying theoretical calculations to a range of biochemical problems is the creation of a prediction model that would allow for the assessment of the antiviral activity of a number of compounds prior to their direct synthesis and biological experimentation. At that, the prediction model can be constructed using machine learning (ML) methods [13,14], results of molecular modeling (docking and/or dynamics) [15], or a combination of both approaches [16,17].

In our recent publication, we have presented our own prediction model for assessing the antiviral potential of small molecules against the influenza virus of strain A/H1N1 [18]. The compiled database contained information on the small molecule structures in SMILES format, data on their cytotoxicity against the MDCK cell line, and IC_{50} values characterizing antiviral activity against different serotypes of the A/H1N1 influenza virus strain. The study utilized classical and non-classical machine learning methods to solve a binary classification problem to separate active and inactive molecules based on selectivity index values. Classification was performed at a threshold value SI = 8, since, according to the Guidelines for the Experimental (Preclinical) Study of New Pharmacological Substances [19], substances with SI values >8 can be considered promising. Analysis of the calculation results allowed us to identify a random forest model using Morgan fingerprints as a descriptor, characterizing the structure of a small molecule as a promising one. The accuracy of the classification problem solution was 75.94%. This prediction is ligand-oriented, i.e., it allows us to identify compounds active against the influenza virus, regardless of which viral protein may play the role of a biological target.

Alternatively, a prediction model can be constructed using molecular modeling methods. This approach is based on searching for a relationship between the pIC_{50} values (inverse logarithm of the half-maximal inhibitory concentration) and the energy of binding with the active site of a therapeutically important protein ΔG_{bind}. A similar approach that was described in [20] dealt with the search for histone demethylase inhibitors. The authors used various molecular docking algorithms, carried out a series of molecular dynamics simulations, and carefully analyzed the binding energy estimates. The resulting prediction models allowed them to classify inhibitors into active and inactive ones. The authors

of [20] believe that such models can be used as a primary filter for identifying promising compounds for subsequent synthesis and biological testing.

The relationship between experimental values of binding and docking energy is presented in [21]. The authors determined the cytotoxicity values of some compounds in relation to four cell line models, namely HepG2, LU-1, SW480, and HL-60; then, the CC_{50} values were set equal to Ki, and the experimental binding energy was calculated using formula $\Delta G_{exp} = RT \ln K_i$. For the molecular docking procedure, the authors considered the GSK-3β enzyme as a biological target. Theoretical and experimental data correlate with each other with correlation indices (R) from 0.90 to 0.93 depending on the cell line. For 10 ligand–protein complexes, the authors of [21] applied refinement calculations using the methodology of fast pulling of ligand (FPL) [22]. The correlation index of the calculated values of binding energies with experimental data is 0.8.

Although machine learning approaches are commonly applied to prioritize potentially active compounds, their ability to reliably identify chemical series with low antiviral potential has received considerably less attention. However, accurate negative predictions are equally important because they reduce unnecessary synthesis and biological testing, thereby improving the efficiency of medicinal chemistry campaigns.

The present study was designed to prospectively evaluate the predictive performance of computational models for pyrimidin-4(3*H*)-one derivatives in both directions: identifying compounds with promising antiviral activity and correctly recognizing compounds with low antiviral potential. To achieve this objective, we first refined our previously developed machine learning model [18] and applied it to predict the antiviral activity of a focused library of pyrimidin-4(3*H*)-one derivatives. In parallel, a complementary structure-based model was developed using the molecular docking of pyrimidine-containing ligands in the active site of the influenza A virus PA endonuclease (PAN) [9,23,24]. The predicted activities were subsequently validated by synthesis of the selected compounds and experimental evaluation of their antiviral activity against influenza A/Puerto Rico/8/34 (H1N1). The adequacy of both predictive approaches was additionally assessed using literature data for known PAN inhibitors and antiviral pyrimidine derivatives [4,7,9,10,23,24].

Thus, the aim of this study was to prospectively validate machine-learning- and docking-based prediction models for pyrimidin-4(3*H*)-one derivatives through synthesis and biological evaluation of a new compound library, with particular emphasis on their ability to prioritize compounds for antiviral screening by correctly identifying both promising and low-activity candidates.

2. Materials and Methods

2.1. Data Set, Descriptors, and Machine Learning Tools

The existing database was supplemented [18] with new records from scientific publications and the ChEMBL database [25]. The new database contains 2255 records of small-molecule structures and biological testing data (CC_{50} cytotoxicity against the MDCK cell line, IC_{50} antiviral activity against the A/H1N1 strain of different serotypes, and SI values, characterizing the CC_{50} to IC_{50} ratio). Among 2255 compounds, 910 are characterized by SI values > 8. The median SI value is 4.75. Class balancing was performed for SI = 8, i.e., compounds with SI > 8 were marked with an active label (1), and those with SI < 8 were marked with an inactive label (0). The structural chemical diversity was analyzed using the non-commercial software RDKit 2026.03.3 [26], and Bemis–Murko scaffolds were used for the evaluation [27]. The number of compounds containing a pyrimidine substructure was estimated. To solve the binary classification problem using machine learning, the most effective methods [18] were applied: the classical method combining molecular fingerprints (Morgan fingerprint) [28] and random forest [29], and the neural network method, combin-

ing MolFormer [30] as a method of feature extraction and a fully connected neural network. As was noted above, this combination of descriptors and the model showed the best result in predicting SI values.

The training of machine learning (fp-RF and MF-NN) models was carried out using algorithms that have a general logical structure.

In the case of the fp-RF model, the dataset stores Morgan fingerprint data and the selectivity index value for each molecule, line by line. It is divided into training and testing samples in a ratio of 1:3 and contains only unique lines. Due to the limited amount of data available, no further validation was carried out. Stratification was considered when dividing the data into training and test sets, so the balance of class labels in the two sets is equal. During training, no deferred sampling was used, so no data was leaked from the training set to the test set. The following basic hyperparameters were used to train the random forest: 200 trees; Ginny's criteria were used as the criterion for quality of partition, the minimum element in a leaf is 1, and bootstrap sampling was used.

A fully pre-trained MoLFormer (IBM/MoLformer-XL-Both-10Pct) was used to extract features from an existing dataset, line by line, and then the dataset was divided into two parts: a training set and a test set. The training set consisted of the extracted features and the selectivity index values, while the test set was used for validation. Only the training set was used in further training. A fully connected three-layer neural network with 256, 128, and 1 neuron was used as a classifier on the extracted features. ReLU was used as an activation function. To regularize after the first two layers, dropout with probability 30% was applied. Binary cross entropy was used as loss function, and the optimizer was Adam. The number of epochs was 100.

2.2. Molecular Modeling

Geometrical parameters of pandemic 2009 H1N1 influenza A endonuclease domain (PAN) were downloaded from Protein Data Bank [31]. Two PDB ID were used: 4M5U [23] and 4KIL [24]. Protein structure models were preprocessed with Schrodinger Protein prep-wizard tools: hydrogen atoms were added and refined, missing amino acid residues were added, bond orders were verified, solvent was removed, and protein structure minimization was performed with constraints in the OPLS4 force field [32]. The molecular docking procedure was performed on a group of small molecules. The synthesis of these molecules is described in [9,23,24]. Compounds inhibit the endonuclease activity of viral protein PAN. Structures are shown in Supplementary Materials (Figures S1–S3). Docking was restricted to reference position using the existing docked ligand. Restricted conditions (tolerance, maximum common position or SMART pattern) were dependent on the structures of ligands and the active site of proteins. When analyzing the docking results, attention was paid to the ligand location relative to metal cations (Mn^{2+}) and the correspondence of geometric parameters of ligands to geometric parameters of reference compounds located in active sites of ligand–protein complexes (PDB ID: 4M5U, 4KIL), deciphered using experimental methods. The MM/GBSA methodology [33] was used to estimate ligand-binding affinities of optimal docking positions. Calculation conditions: vacuum, OPLS4 [32] force field, flexible residues are defined by the distances from ligand 5A. Cations Mn^{2+} are included.

2.3. Synthesis and NMR Spectra

NMR spectra were recorded at 298 K on a Bruker Avance-III 500 MHz spectrometer (Bruker BioSpin GmbH, Karlsruhe, Germany) equipped with a PABBO X{^{1}H} direct detection probe, using 5 mm NMR tubes. The operating frequencies were 500.30 MHz for ^{1}H, 125.75 MHz for ^{13}C, and 50.58 MHz for ^{15}N. Chemical shifts in the ^{1}H and ^{13}C NMR spectra are reported in parts per million (ppm) relative to tetramethylsilane as an

external standard. Chemical shifts in the ^{15}N NMR spectra are given in ppm relative to external liquid ammonia at 25 °C. Gradient-selected {^{1}H,^{13}C}, {^{1}H,^{15}N} HSQC and {^{1}H,^{13}C}, {^{1}H,^{15}N} HMBC spectra were recorded using standard Bruker pulse sequences.

Elemental analysis was performed using a Perkin Elmer Ser. II CHNS/O 2400 elemental analyzer (PerkinElmer, Shelton, CT, USA). Melting points were determined with a Stanford Research Systems Optimelt MPA100 melting point apparatus (Stanford Research Systems, Sunnyvale, CA, USA) and have not been corrected.

Derivatives of pyrimidin-4(3*H*)-one were synthesized via the condensation of the corresponding β-ketoester with thiourea, guanidine hydrochloride, or acetamidine in ethanol using sodium ethoxide, or under solvent-free conditions with microwave irradiation [34,35]. In some cases, the latter method provided higher yields of the target compounds compared to the conventional sodium methoxide-based approach; however, its scalability remains challenging.

β-Ketoesters were obtained through the alkylation of Meldrum acid with corresponding acyl chloride in pyridine, followed by decomposition of the intermediate adduct under reflux in ethanol [36]. Alternatively, they were synthesized via the reaction of the magnesium complex of ethyl acetoacetate with acyl chloride or by reacting to the magnesium salt of a malonic ester with corresponding acyl chloride [37]. The choice of the method depended on the structure of β-ketoester: for esters with bulky or electron-withdrawing substituents (R = i-Pr, t-Bu, c-hexyl, n-hexyl), the first method resulted in low yields, whereas for non-branched alkyl substituents (R = Et, n-nonyl, n-pentadecyl, Ph), it was preferred.

Six uracil derivatives (R = Et, c-hexyl, t-Bu, n-nonyl, and n-pentadecyl) were synthesized from 2-thio derivatives via reaction with methyloxirane in an alkaline medium or with monochloroacetic acid in an aqueous medium [38]. The direct condensation of β-ketoesters with urea was ineffective in these cases, as it failed to provide the desired products in acceptable yields.

N-Alkylation of uracil derivatives was carried out using established methods with haloalkanes or dimethyl sulfate in an aqueous alkaline medium. The synthesis of dimethyl 6-pentadecyluracil was accomplished via a solvent-free methodology [39].

1,3-dimethyl-5-carboxyuracil (9): white solid; yield 44%; mp 187.6–188.2 °C (lit. 179–181 [40]); ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 3.207 (s, 3H, $N^{3}CH_3$), 3.444 (s, 3H, $N^{1}CH_3$), 8.709 (s, 1H, C^{6}-H), 12.780 (br. s, 1H, COOH). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 28.27 ($N^{3}CH_3$), 37.86 ($N^{1}CH_3$), 100.96 (C^{5}), 150.89 (O=C^{2}), 152.78 (C^{6}), 163.85, 163.95. ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 138.0 (N^{1}), 154.8 (N^{3}). Found, %: C, 45.60; H, 4.40; N, 15.23. $C_7H_8N_2O_4$. Calculated, %: C, 45.66; H, 4.38; N, 15.21.

6-amino-3-butyl-2-methylpyrimidin-4(3*H*)-one (10): white solid; yield 76%; mp 137–139 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.956 (t, 3H, *J* = 7.35 Hz, CH_3), 1.402 (td, 2H, *J* = 7.35, 7.48 Hz, $MeCH_2$), 1.631 (dd, 2H, *J* = 7.48, 7.90 Hz, $C\underline{H}_2CH_2N$), 2.442 (s, 3H, H_3CC^{2}), 3.910 (t, 2H, *J* = 7.90 Hz, H_2C-N), 4.881 (s, 2H, NH_2), 5.345 (s, 1H, HC^{5}), ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.74 ($CH_2\underline{C}H_3$), 20.20 ($\underline{C}H_2CH_3$), 22.51 (C^{2}-$\underline{C}H_3$), 30.82 ($N^{3}CH_2\underline{C}H_2$), 43.66 ($N^{3}\underline{C}H_2CH_2$), 85.96 ($C^{5}$), 158.97 ($MeC^{2}$), 161.25 ($C^{6}$), 163.19 ($C^{4}$). Found, %: C, 59.74; H, 8.36; N, 23.13. $C_9H_{15}N_3O$. Calculated, %: C, 59.64; H, 8.34; N, 23.19.

6-Pentadecyluracil (11): white solid; yield 56%; mp 165.0–166.5 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.854 (t, 3H, *J* = 6.7 Hz, CH_3), 1.15–1.33 (m, 24H, CH_2), 1.46–1.56 (m, 2H, $C\underline{H}_2CH_2C$), 2.264 (t, 2H, *J* = 7.6 Hz, $C^{6}CH_2$), 5.298 (s, 1H, HC^{5}), 10.754 (s, 1H, $N^{1}H$), 10.861 (s, 1H, $N^{3}H$). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.85 (CH_3), 22.01 (CH_2), 26,80 (CH_2), 28.22 (CH_2), 28.54 (CH_2), 28.63 (CH_2), 28.79 (CH_2), 28.85–29.05 (CH_2), 31.22 (CH_2), 31.63 ($C\underline{C}H_2$), 97.84 (HC^{5}), 151.62 (O=C^{2}), 156.49 (C^{6}),

164.09 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 135.5 (N^1), 156.3 (N^3). Found, %: C, 70.05; H, 10.71; N, 8.35. $C_{19}H_{34}N_2O_2$. Calculated, %: C, 70.76; H, 10.63; N, 8.69.

6-amino-2-(dimethylamino)pyrimidin-4(3*H*)-one (12): white solid; yield 63%; mp 290.6–291.2 °C (lit. 290.5–292.5 °C [41]); ^1H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 2.979 (s, 6H, CH_3), 4.540 (s, 1H, HC^5), 6.043 (s, 2H, NH_2), 10.586 (br. s, 1H, NH). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 36.77 (CH_3), 75,66 (C^5), 154.93, 164.43, 164.54. ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 64.1 (NH_2), 80.9 (NMe_2), 142.4 (N^3), 199.5 (N^1). Found, %: C, 46.23; H, 6.42; N, 36.21. $C_6H_{10}N_4O$. Calculated, %: C, 46.74; H, 6.54; N, 36.34.

1,3-Dimethyl-6-pentadecyluracil (13): white solid; yield 41%; mp 64.8–65.8 °C; ^1H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.856 (t, 3H, *J* = 6.85 Hz, CH_3), 1.20–1.40 (m, 24H, CH_2), 1.50–1.58 (m, 2H, CH_2), 2.509 (t, 2H, *J* = 7.65 Hz, C^6CH_2), 3.153 (s, 3H, N^3CH_3), 3.312 (s, 3H, N^1CH_3), 5.501 (s, 1H, HC^5). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.71 (CH_3), 21.99 (CH_2), 26.70 (CH_2), 27.27 (N^3CH_3), 28.44 (CH_2), 28.63 (CH_2), 28.85 (CH_2), 28.9–29.1 (CH_2), 31.07 (N^1CH_3), 31.26 (CH_2), 31.71 (CH_2), 98.91 (HC^5), 152.30 (C^2), 155.96 (C^6), 161.63 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 127.7 (N^1), 151.9 (N^3). Found, %: C, 70.87; H, 11.12; N, 7.83. $C_{21}H_{38}N_2O_2$. Calculated, %: C, 71.95; H, 10.93; N, 7.99.

5-Ethoxycarbonyluracil (14): white solid; yield 83%; mp 239.2–240.3 °C (lit. 241–242 [42]); ^1H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 1.237 (t, 3H, *J* = 7.0 Hz, CH_3), 4.163 (t, 2H, *J* = 7.0 Hz, CH_2), 8.130 (d, 1H, *J* = 6.4 Hz, C^6H), 11.303 (s, 1H, N^3H), 11.589 (d, 1H, *J* = 6.4 Hz, N^1H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 14.08 (CH_3), 59.86 (CH_2), 103.06 (C^5), 149.30 (C^6), 150.54 (C^4), 160.01 (C^2), 162.49 ($C^5C{=}O$). ^1H and ^{13}C NMR spectra was similar with [42]. ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 139.5 (N^1), 159.7 (N^3). Found, %: C, 45.11; H, 4.22; N, 15.34. $C_7H_8N_2O_4$. Calculated, %: C, 45.66; H, 4.38; N, 15.21.

6-Amino-2-(methylthio)pyrimidin-4(3*H*)-one (15): commercially available, was used after recrystallization from water, white solid; mp 267–268 °C; Found, %: C, 38.19; H, 4.42; N, 26.63; S, 20.42. $C_5H_7N_3OS$. Calculated, %: C, 38.21; H, 4.49; N, 26.73; S, 20.40.

1,3,6-Trimethyluracil (16): commercially available, was used after recrystallization from ethanol, white solid. Found, %: C, 54.01; H, 6.41; N, 18.32. $C_7H_{10}N_2O_2$. Calculated, %: C, 54.54; H, 6.54; N, 18.17.

6-Amino-2-thiouracil (17): commercially available, was used after recrystallization from water. Found, %: C, 33.39; H, 3.39; N, 29.24; S, 22.16. $C_4H_5N_3OS$. Calculated, %: C, 33.56; H, 3.52; N, 29.35; S, 22.39.

2-Amino-6-isobutylpyrimidin-4(3*H*)-one (18): white solid; yield 67%; mp 221.0–223.2 °C; ^1H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.869 (d, J = 6.6 Hz, 6H, C^6-$CH_2CH(C\underline{H}_3)_2$), 1.961 (sept, J = 6.8 Hz, 1H, C^6-$CH_2C\underline{H}(CH_3)_2$), 2.117 (d, J = 7.2 Hz, 2H, C^6-$C\underline{H}_2CH(CH_3)_2$), 5.376 (s, 1H, C^5-H), 6.673 (s, 2H, NH_2), 10.953 (s, 1H, N^1-H), 10.953 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 22.14 (C^6-$CH_2CH(\underline{C}H_3)_2$), 26.72 ($C^6$-$CH_2\underline{C}H(CH_3)_2$), 45.69 ($C^6$-$\underline{C}H_2CH(CH_3)_2$), 100.69 ($C^5$), 155.59 ($C^6$), 155.75 ($C^2$), 163.88 ($C^4$). Found, %: C, 57.01; H, 7.96; N, 24.82. $C_8H_{13}N_3O$. Calculated, %: C, 57.46; H, 7.84; N, 25.13.

6-Ethyluracil (19): white solid; yield 58%; mp 200.6–201.2 °C (lit. 202–203 °C [43]; ^1H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 1.103 (t, *J* = 7.5 Hz, 3H, CH_3), 2.306 (q, 2H, *J* = 7.5 Hz, CH_2), 5.327 (s, 1H, C^5H), 10.826 (s, 1H, N^1H), 10.913 (s, 1H, N^3H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 11.48 (CH_3), 24.96 (CH_2), 97.00 (C^5), 151.63 (C^2), 157.87 (C^6), 164.25 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 135.2 (N^1), 156.1 (N^3). Found, %: C, 51.27; H, 5.62; N, 19.81. $C_6H_8N_2O_2$. Calculated, %: C, 51.42; H, 5.75; N, 19.99.

2-Amino-6-nonylpyrimidin-4(3*H*)-one (20): white solid; yield 52%; mp 192.0–193.2 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.868 (t, 3H, *J* = 6.8 Hz, CH_3), 1.17–1.35 (m, 12H, CH_2), 1.45–1.58 (m, 2H, CH_2), 2.234 (t, 2H, J = 7.4 Hz, C^6-CH_2), 5.386 (s, 1H, C^5-H), 6.487 (s, 2H, NH_2), 10.669 (s, 1H, NH). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.93 (CH_3), 22.08 (CH_2), 27.48 (CH_2), 28.62 (CH_2), 28.66 (CH_2), 28.85 (CH_2), 28.92 (CH_2), 31.26 (CH_2), 37.02 (C^6CH_2), 99.63 (C^5), 155.47, 162.87, 169.84. Found, %: C, 65.68; H, 9.82; N,17.67. $C_{13}H_{23}N_3O$. Calculated, %: C, 65.79; H, 9.77; N, 17.70.

6-tert-Butyluracil (21): white solid; yield 43%; mp 226–228 °C (lit. 227–229 °C [44]); ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 1.193 (s, 9H, CH_3), 5.317 (s, 1H, HC^5), 10.579 (s, 1H, HN^1), 11.002 (s, 1H, HN^3). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 27.34 (CH_3), 34.36 ($\underline{C}Me_3$), 95.33 (C^5), 151.95 (C=O), 163.13 (C^6), 164.28 (C=O). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 133.2 (N^1), 154.2 (N^3). Found, %: C, 57.44; H, 7.26; N, 16.52. $C_8H_{12}N_2O_2$. Calculated, %: C, 57.13; H, 7.19; N, 16.66.

1,2-Dimethyl-6-ethyluracil (22): white solid; yield 46%; mp 61.0–62.2 °C; ^{1}H NMR spectrum (500.13 MHz, CDCl3), δ, ppm: 1.255 (t, 3H, *J* = 7.3 Hz, CH_3), 2.539 (q, 2H, *J* = 7.3 Hz, CH_2), 3.344 (s, 3H, N^3CH_3), 3.421 (s, 3H, N^1CH_3), 5.646 (s, 1H, HC^5). ^{13}C NMR spectrum ($CDCl_3$, 125.76 MHz), δ, ppm: 11.31 ($CH_2\underline{C}H_3$), 25.71 (CH_2), 27.93 (N^3CH_3), 31.09 (N^1CH_3), 99.16 (C^5), 152.69 (C^2), 156.36 (C^6), 162.79 (C^4). ^{15}N NMR ($CDCl_3$, 50.58 MHz), δ, ppm: 127.4 (N^1), 153.0 (N^3). Found, %: C, 57.28; H, 7.29; N, 16.41. $C_8H_{12}N_2O_2$. Calculated, %: C, 57.13; H, 7.19; N, 16.66.

6-Nonyluracil (23): white solid; yield 59%; mp 171.0–172.2 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.858 (t, 3H, *J* = 6.8 Hz, CH_3), 1.245–1.288 (m, 12H, CH_2), 1.521 (p, 2H, *J* = 6.8 Hz, C^6-CH_2-$C\underline{H}_2$), 2.276 (t, 2H, *J* = 7.6 Hz, C^6-CH_2), 5.315 (s, 1H, C^5-H), 10.801 (s, 1H, N^1-H), 10.905 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.91 (CH_3), 22.09 ($\underline{C}H_2CH_3$), 26.87 (C^6-$CH_2\underline{C}H_2$), 28.28 (CH_2), 28.63 (CH_2), 28.66 (CH_2), 28.84 (CH_2), 31.27 (CH_2), 31.69 (C^6-$\underline{C}H_2$), 97.90 (C^5), 151.70 (C^2), 156.56 (C^6), 164.19 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 135.7 (N^1), 156.5 (N^3). Found, %: C, 65.16; H, 9.35; N, 11.61. $C_{13}H_{22}N_2O_2$. Calculated, %: C, 65.52; H, 9.30; N, 11.75.

6-Phenyl-2-methylpyrimidin-4(3*H*)-one (24): white solid; yield 61%; mp 149–151 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 2.385 (s, 3H, CH_3), 6.743 (s, 1H, HC^5), 7.45–7.51 (m, 3H, Ph), 8.02–8.07 (m, 2H, Ph), 12.493 (s, 1H, NH). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 21.37 (CH_3), 106.63 (C^5), 126.77 (Ph), 128.55 (Ph), 130.28 (Ph), 136.14 (Ph), 159.03 (C^2), 160.41 (C^6), 162.82 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 176.5 (N^3), 234.5 (N^1). Found, %: C, 71.26; H, 5.54; N, 15.43. $C_{11}H_{10}N_2O$. Calculated, %: C, 70.95; H, 5.41; N, 15.04.

6-Cyclohexyluracil (25): white solid; yield 58%; mp 286–287 °C (lit. 286–287 °C [45]); ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 1.10–1.38 (m, 5H, CH_2), 1.60–1.68 (m, 1H, CH_2), 1.70–1.83 (m, 4H, CH_2), 2.18–2.27 (m, 1H, $C^{1'}H$), 5.284 (s, 1H, HC^5), 10.735 (s, 1H, N^1-H), 10.906 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 25.10 ($C^{4'}$), 25.53 ($C^{3'}$), 30.19 ($C^{2'}$), 40.43 ($C^{1'}$), 95.71 (C^5), 151.72 (C^2), 160.82 (C^6), 164.41 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 133.9 (N^1), 155.8 (N^3). Found, %: C, 61.56; H, 7.36; N, 14.98. $C_{10}H_{14}N_2O_2$. Calculated, %: C, 61.84; H, 7.27; N, 14.42.

6-Cyclohexyl-2-thiouracil (26): white solid; yield 72%; mp 286–288 °C (lit. 282–285 °C [46]); ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 1.11–1.38 (m, 5H, CH_2), 1.60–1.69 (m, 1H, CH_2), 1.70–1.87 (m, 4H, CH_2), 2.30–2.40 (m, 1H, $C^{1'}H$), 5.632 (s, 1H, HC^5), 12.150 (s, 1H, N-H), 12.323 (s, 1H, N-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 25.04 ($C^{4'}$), 25.52 ($C^{3'}$), 30.32 ($C^{2'}$), 39.82 ($C^{1'}$), 100.56 (C^5), 161.04 and 161.29 (C^4 and C^6), 175.97 (C^2). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 158.5 (N^1), 180.7 (N^3). Found, %: C, 57.92; H, 6.45; N, 13.52; S, 15.43. $C_{10}H_{14}N_2OS$. Calculated, %: C, 57.12; H, 6.71; N, 13.32; S, 15.25.

6-Isobutyluracil (27): white solid; yield 62%; mp 225.1–226.3 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.877 (d, *J* = 6.6 Hz, 6H, C^6-$CH_2CH(CH_3)_2$), 1.917 (sept, *J* = 6.8 Hz, 1H, C^6-$CH_2C\underline{H}(CH_3)_2$), 2.154 (d, *J* = 7.3 Hz, 2H, C^6-$C\underline{H_2}CH(CH_3)_2$), 5.301 (s, 1H, C^5-H), 10.771 (s, 1H, N^1-H), 10.913 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 21.76 (C^6-$CH_2CH(\underline{C}H_3)_2$), 26.53 ($C^6$-$CH_2\underline{C}H(CH_3)_2$), 40.63 ($C^6$-$\underline{C}H_2CH(CH_3)_2$), 98.81 ($C^5$), 151.64 ($C^2$), 155.36 ($C^6$), 164.02 ($C^4$). ^{15}N NMR (DMSO-$d_6$, 50.58 MHz), δ, ppm: 136.0 (N^1), 156.5 (N^3). Found, %: C, 58.02; H, 7.45; N, 16.92. $C_8H_{12}N_2O_2$. Calculated, %: C, 57.13; H, 7.19; N, 16.66.

2-((6-Amino-2-(butylthio)pyrimidin-4-yl)oxy)acetic acid (28): white solid; yield 39%; mp 264–265 °C (lit. 265 °C [47]) ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.893 (t, 3H, *J* = 7.35 Hz, CH_3), 1.380 (tq, 2H, *J* = 7.35, 7.38 Hz, CH_2Me), 1.580 (tt, 2H, *J* = 7.35, 7.38 Hz, $C\underline{H_2}CH_2S$), 2.950 (t, 2H, *J* = 7.35 Hz, CH_2S), 4.734 (s, 2H, CH_2OC^4), 5.483 (HC^5), 6.717 (s, 2H, NH_2), 12.846 (br. s, 1H, C(O)OH). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.59 (CH_3), 21.51 ($\underline{C}H_2Me$), 29.42 (SCH_2), 31.53 ($\underline{C}H_2CH_2S$), 61.82 ($\underline{C}H_2OC^4$), 81.93 (C^5), 165.24 (C^6), 167.73 (C^4), 169.10 (C^2), 170.08 (C(O)OH). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 82.6 (NH_2), 219.2, 230.8. Found, %: C, 46.02; H, 5.98; N, 16.52; S, 12.13. $C_{10}H_{15}N_3O_3S$. Calculated, %: C, 46.68; H, 5.88; N, 16.33; S, 12.46.

6-Nonyl-2-thiouracil (29): white solid; yield 73%; mp 147.5–148.6 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.858 (t, *J* = 6.9 Hz, 3H, CH_3), 1.246–1.289 (m, 12H, CH_2), 1.518 (p, J = 6.8 Hz, 2H, C^6-CH_2-$C\underline{H_2}$), 2.341 (t, *J* = 7.6 Hz, 2H, C^6-CH_2), 5.679 (s, 1H, C^5-H), 12.223 (s, 1H, N^1-H), 12.330 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.90 (CH_3), 22.07 ($\underline{C}H_2CH_3$), 27.13 (C^6-$CH_2\underline{C}H_2$), 28.27 (CH_2), 28.58 (CH_2), 28.64 (CH_2), 28.81 (CH_2), 31.24 (CH_2), 31.29 (C^6-$\underline{C}H_2$), 102.83 (C^5), 156.79 (C^6), 161.04 (C^4). 175.97 (C^2). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 159.5 (N^1), 180.0 (N^3). Found, %: C, 62.03; H, 8.78; N, 11.32; S, 12.33. $C_{13}H_{22}N_2OS$. Calculated, %: C, 61.38; H, 8.72; N, 11.01; S, 12.60.

6-(tert-Butyl)-2-methylpyrimidin-4(3*H*)-one (30): white solid; yield 59%; mp 127.1–129.5 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 1.178 (s, 9H, $C(CH_3)_3$), 2.279 (s, 3H, C^2CH_3), 6.022 (s, 1H, HC^5), 12.255 (s, 1H, NH). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 21.26 ($\underline{C}H_3C^2$), 28.38 ($C(\underline{C}H_3)_3$), 36.45 ($\underline{C}(CH_3)_3$), 105.68 (C^5), 157.82 (C^2), 162.88 (C^4), 174.36 (C^6). Found, %: C, 64.03; H, 8.61; N, 16.27. $C_9H_{14}N_2O$. Calculated, %: C, 65.03; H, 8.49; N, 16.85.

2-(Methylthio)-6-nonylpyrimidin-4(3*H*)-one (31): white solid; yield 52%; mp 113.7–114.9 °C; ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.828 (t, *J* = 6.9 Hz, 3H, CH_3), 1.15–1.30 (m, 12H, CH_2), 1.559 (p, J = 6.8 Hz, 2H, C^6-CH_2-$C\underline{H_2}$), 2.383 (t, *J* = 7.5 Hz, 2H, C^6-CH_2), 2.443 (s, 3H, SCH_3), 5.649 (s, 1H, C^5-H), 12.305 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 13.19 (CH_3), 14.41 (CH_3), 22.57, 27.72, 28.95, 29.14, 29.25, 29.35, 31.75, 31.80, 103.36 (C^5), 157.33, 161.55, 176.50. Found, %: C, 62.83; H, 9.35; N, 10.26; S, 12.21. $C_{14}H_{24}N_2OS$. Calculated, %: C, 62.65; H, 9.01; N, 10.44; S, 11.94.

6-Isobutyl-2-thiouracil (32): white solid; yield 73%; mp 217.2–218.2 °C (lit. 220.5–221.5 °C [47]); ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 0.879 (d, *J* = 6.6 Hz, 6H, C^6-$CH_2CH(C\underline{H_3})_2$), 1.903 (sept, *J* = 6.8 Hz, 1H, C^6-$CH_2C\underline{H}(CH_3)_2$), 2.224 (d, *J* = 7.3 Hz, 2H, C^6-$C\underline{H_2}CH(CH_3)_2$), 5.671 (s, 1H, C^5-H), 12.201 (s, 1H, N^1-H), 12.348 (s, 1H, N^3-H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 21.70 (C^6-$CH_2CH(\underline{C}H_3)_2$), 26.89 ($C^6$-$CH_2\underline{C}H(CH_3)_2$), 40.08 ($C^6$-$\underline{C}H_2CH(CH_3)_2$), 103.77 ($C^5$), 155.60 ($C^6$), 160.91 ($C^4$), 175.91 ($C^2$). ^{15}N NMR (DMSO-$d_6$, 50.58 MHz), δ, ppm: 160.2 (N^1), 180.6 (N^3). Found, %: C, 51.92; H, 6.71; N, 15.13; S, 17.28. $C_8H_{12}N_2OS$. Calculated, %: C, 52.15; H, 6.56; N, 15.20; S, 17.40.

2-Amino-6-phenylpyrimidin-4(3*H*)-one (33): white solid; yield 60%; mp 301–302 °C (lit. 303–304 °C [48]); ^{1}H NMR spectrum (500.13 MHz, DMSO-d_6), δ, ppm: 6.137 (s, 1H, HC^5), 6.698 (s, 2H, NH_2), 7.42–7.47 (m, 3H, Ph), 7.93–7.99 (m, 2H, Ph), 10.961 (s, 1H,

N^3H). ^{13}C NMR spectrum (DMSO-d_6, 125.76 MHz), δ, ppm: 97,53 (C^5), 126.56 (Ph), 128.30, (Ph), 129.86 (Ph), 137.19 (Ph), 155.66 (C^4), 162.55 (C^6), 163.38 (C^2). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 76.1 (NH_2), 148.8 (N^3), 191.7 (N^1). Found, %: C, 64.82; H, 4.77; N, 22.21. $C_{10}H_9N_3O$. Calculated, %: C, 64.16; H, 4.85; N, 22.45.

1,3-Dimethyl-6-isopropyluracil (34): colorless oil; yield 49%; ^1H NMR spectrum (500.13 MHz, $CDCl_3$), δ, ppm: 1.256 (d, 6H, *J* = 6.75 Hz, $(C\underline{H}_3)_2CH$), 2.902 (sept., 1H, *J* = 6.75 Hz, $C^6\underline{C}H$), 3.343 (s, 3H, N^3CH_3), 3.469 (s, 3H, N^1CH_3), 5.689 (s, 1H, HC^5). ^{13}C NMR ($CDCl_3$, 125.76 MHz) δ = 21.39 ($(\underline{C}H_3)_2$-CH), 27.93 (N^3CH_3), 29.64 ($C^6\underline{C}H$), 31.07 (N^1CH_3), 97.34 (C^5), 152.82 (C^2), 160.98 (C^6), 163.07 (C^4). ^{15}N NMR (DMSO-d_6, 50.58 MHz), δ, ppm: 127.1 (N^1), 153.3 (N^3). Found, %: C, 59.01; H, 7.86; N, 15.52. $C_9H_{14}N_2O_2$. Calculated, %: C, 59.32; H, 7.74; N, 15.37.

2.4. Biological Assays

Viruses and cells. Influenza virus A/Puerto Rico/8/34 (H1N1) was obtained from the collection of viruses of the St. Petersburg Pasteur Institute. Before the experiment, virus was incubated in Madin-Darby Canine Kidney (MDCK) cells (ATCC CCL-34) for 48 h at 36 °C. The infectious titer of the virus was determined in MDCK cells grown in 96-well plates in alpha-MEM serum-free medium by end-point dilution assay.

Cytotoxicity Assay. MDCK cells were seeded onto 96-well culture plates (104 cells per well) and incubated at 36 °C in 5% CO_2 until continuous monolayer formation. To assess the toxicity of compounds, a series of their 3-fold dilutions at concentrations of 300 to 3.7 μg/mL in Eagle's Minimal Essential Medium (MEM) were prepared. The dilutions were added to the wells of the plates. Cells were incubated for 48 h at 36 °C in a CO_2 incubator under 5% CO_2. Further, a microtetrazolium (MTT) assay was performed on 96-well plates. The cells were washed 2 times with saline (0.9% NaCl), and a 100 μL/well of MTT solution [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] at a concentration of 0.5 g/mL in MEM was added. The plates were incubated for 1 h at 36 °C, the liquid was removed, and dimethylsulfoxide (DMSO) (0.1 mL per well) was added. The optical density (OD) in wells was measured using a Thermo Multiskan FC spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) at a wavelength of 540 nm. Based on the data obtained, the CC_{50}—the concentration of the compound that destroys 50% of the cells in the culture—was calculated for each specimen.

CPE Reduction Assay. The compounds in appropriate concentrations were added to MDCK cells (0.1 mL per well). MDCK cells were further infected with the influenza virus (m.o.i. 0.01). Plates were incubated for 48 h at 36 °C at 5% CO_2. After that, cell viability was assessed by the MTT test, as described above. The cytoprotective activity of compounds was considered as their ability to increase the values of the OD compared to the control wells (with virus only; no drugs). Based on the results obtained, the IC_{50} values, i.e., the concentration of compounds that results in 50% cell protection, were calculated using GraphPad Prism 6.01 software. IC_{50} values in μg/mL were then calculated into micromoles. For each compound, the value of the selectivity index (SI) was calculated as a ratio of CC_{50} to IC_{50}.

Statistical analysis. All of the in vitro experiments were repeated three times. Calculation of CC_{50} and IC_{50} values was performed using the GraphPad Prism software package (v.6.01), using a 4-parameter logistic curve equation as the working model for analysis (menu items "Nonlinear regression"—"logarithm of inhibitor—response").

3. Results

3.1. Data Set Analysis

One of the important factors in the applicability of a machine-learning-based QSAR approach is determination of the model applicability domain [49]. This procedure improves the reliability of prediction results. In this study, we analyzed the molecules included into our database of small molecules for the presence of pyrimidine and pyrimidin-4(3H)-one fragments in their structures. The database contains 2255 compounds, whose structures are presented in SMILES format, and the results of their biological testing against the influenza A/H1N1 virus of different serotypes. The analysis revealed that the obtained substructures were found in 47 and 13 compounds. The structures of compounds are presented in the Supplementary Materials (Figures S4 and S5). In other words, the compounds under study are within the applicability domain of the resulting models. It is worth noting that among the molecules with a pyrimidin-4(3*H*)-one ring presented in the database, only two of the thirteen have an SI value > 8. These structures are shown in Figure 2.

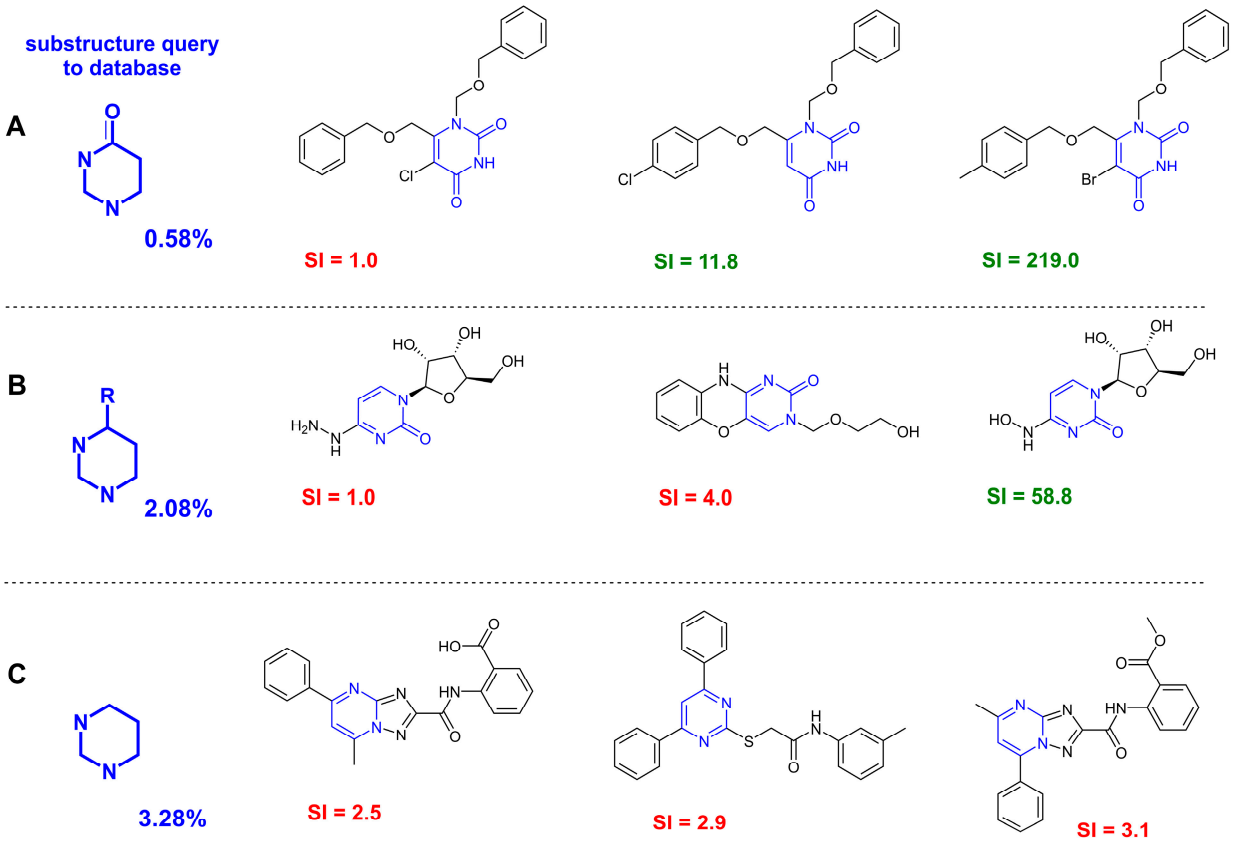


Figure 2. Results of a substructural search in the database. (**A**) A pyrimidin-4(3*H*)-one fragment in the structure. (**B**) A pyrimidine fragment with any substituent at position 4. (**C**) A pyrimidine fragment. The percentages indicate the number of small molecules with a given substructure in the database. The key substructure is highlighted in blue. SI values less than 8 are shown in red, and values greater than 8 are shown in green, respectively.

The database contains uracil derivatives tested against the influenza virus of strain A/PR/8/34 (H1N1) [4]. This fact should facilitate a more accurate classification of the studied small molecules as active or inactive, given that these compounds were present in the training set of the ML model.

3.2. Evaluation of Antiviral Activity

We have previously shown [18] that the classical Random Forest method [28], using Morgan fingerprints [29] (fp-RF) as descriptors, turned out to be the most accurate for solving the binary classification problem by means of the SI threshold > 8. The F-measure value was 0.74. The sensitivity value is 0.6, the specificity value is 0.78, and the balanced accuracy value is 0.69. Non-classical ML methods showed very mediocre results. The highest F-measure value corresponds to the MolFormer-Neural Network (MF-NN) model [30] and equals 0.67. The sensitivity value is 0.67, the specificity value is 0.73, and the balanced accuracy is 0.70. The uncertainty matrices of the rp-RF and MF-NN models are presented in Figure 3A,B. The ROC area values for fp-RF and MF-NN are 0.7594 and 0.7567, respectively (Figure 3C,D).

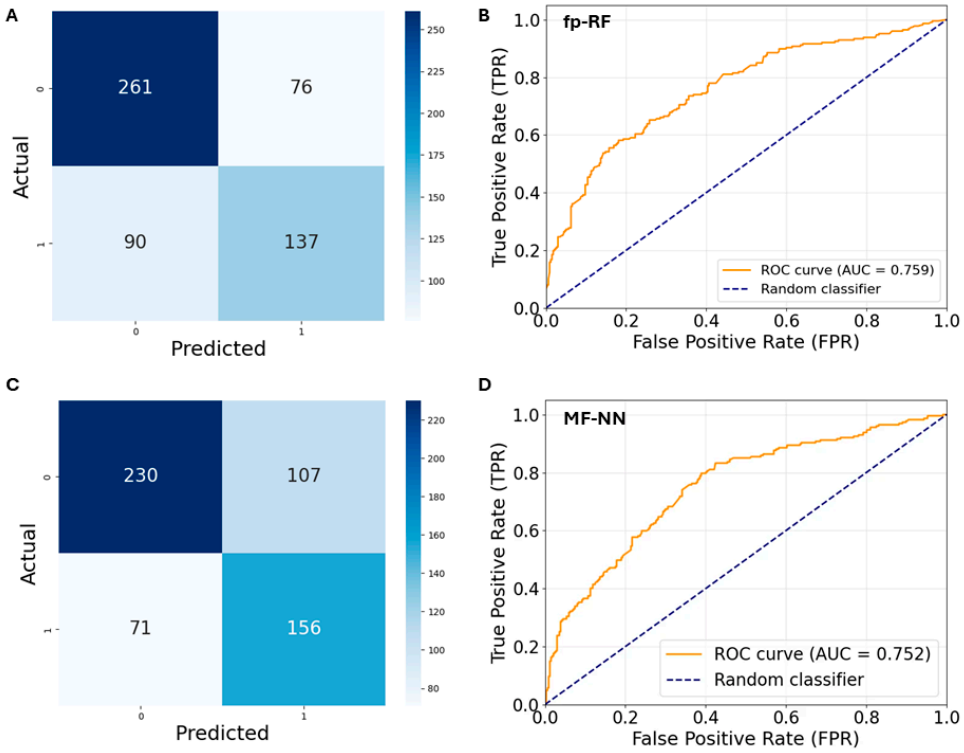


Figure 3. Final models after training: (**A**,**B**) uncertainty matrices; (**C**,**D**) ROC curves.

In this work, we used two ML methods (fp-RF and MF-NN) to separate pyrimidin-4(3*H*)-one derivatives (Figure 4) based on the selectivity index.

In total, 60 models were trained in the study (30 random forest models and 30 models based on neural networks), and each of them solved the binary classification problem with respect to a given threshold p: the value of selectivity index in the range [2; 8] with a step of 0.2. Each model allowed a prediction of whether the selectivity index value of a chemical compound is within the range [0, p] or (p, 300]. The threshold of 300 was chosen as the maximum possible SI value. Although the models failed to determine the exact value of the index, but on the basis of the entire set of predictions, taking into account the given range, it was concluded that the selectivity index values for each compound, if they are less than 8, are in the range from 2 to 4.8 inclusive.

Compounds (**9**–**34**) were synthesized, and the values of their cytotoxicity against the MDCK cell line and antiviral activity against the influenza virus of strain A/PR/8/34 (H1N1) were assessed.

The results of prediction by the classical ML method suggest that all pyrimidin-4(3*H*)-one derivatives under consideration are unpromising as anti-influenza agents. The

prediction result using the non-classical MF-NN ML method highlights compounds **16**, **22**, **30**, and **34**, for which the SI value may be greater than 8. The results of biological testing allow us to identify only two promising compounds, for which SI (**12**) = 59 and SI (**33**) = 10 (Table 1). In other words, the use of the fp-RF method led to erroneous conclusions only in two cases. The non-classical MF-NN method yields more erroneous conclusions.

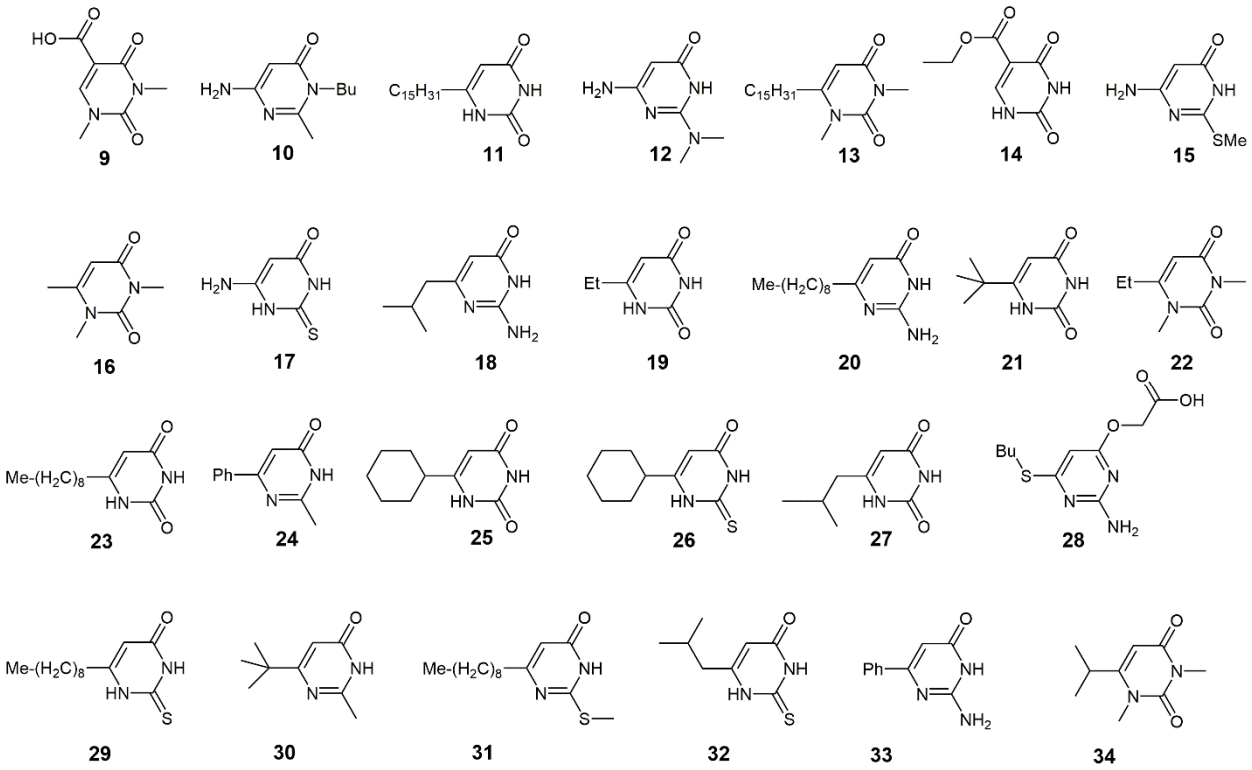


Figure 4. Derivatives of pyrimidin-4(3*H*)-ones.

Table 1. Results of classification of antiviral activity of pyrimidin-4(3*H*)-one derivatives. Green shading of the SI columns indicates a correct activity detection, while red shading indicates an incorrect one. The calculated CC_{50} and IC_{50} values were given as mean ± standard deviation for the three experimental replicates. ">" sign denotes that, across the investigated concentration range up to highest point, no evidence of toxicity or activity was observed. The green color shows compounds with a selectivity index (SI) greater than 8.

ID Compounds	CC_{50}, μM	IC_{50}, μM	SI	SI_{pred}(fp-RF)	SI_{pred}(MF-NN)
9	>1629.1	>1629.1	1	SI < 8	SI < 8
10	>1655.5	413.9 ± 47.5	4	SI < 8	SI < 8
11	491.5 ± 33.7	164.3 ± 20.6	3	SI < 8	SI < 8
12	**>1945.9**	**33.1 ± 5.3**	**59**	SI < 8	SI < 8
13	855.8 ± 79.6	716.0 ± 87.8	1	SI < 8	SI < 8
14	>1629.1	431.2 ± 55.8	4	SI < 8	SI < 8
15	>1908.5	>1908.5	1	SI < 8	SI < 8
16	>1947.2	668.5 ± 71.8	3	SI < 8	SI > 8
17	>2095.6	>2095.6	1	SI < 8	SI < 8
18	>1794.2	1501.1 ± 178.8	1	SI < 8	SI < 8
19	>2140.7	1791.1 ± 187.3	1	SI < 8	SI < 8
20	>1264.0	185.4 ± 19.3	7	SI < 8	SI < 8
21	>1783.6	>1783.6	1	SI < 8	SI < 8
22	>1783.6	>1783.6	1	SI < 8	SI > 8
23	141.4 ± 25.4	138.5 ± 23.7	1	SI < 8	SI < 8
24	>1611.1	>1611.1	1	SI < 8	SI < 8

Table 1. *Cont.*

ID Compounds	CC_{50}, μM	IC_{50}, μM	SI	SI_{pred}(fp-RF)	SI_{pred}(MF-NN)
25	>1544.6	816.0 ± 93.8	2	SI < 8	SI < 8
26	>1426.5	>1426.5	1	SI < 8	SI < 8
27	>1783.6	>1783.6	1	SI < 8	SI < 8
28	>1165.9	>1165.9	1	SI < 8	SI < 8
29	>1179.3	>1179.3	1	SI < 8	SI < 8
30	>1804.8	>1804.8	1	SI < 8	SI > 8
31	122.2 ± 13.6	29.8 ± 3.5	4	SI < 8	SI < 8
32	>1628.1	>1628.1	1	SI < 8	SI < 8
33	**>1602.6**	**164.0 ± 21.6**	**10**	SI < 8	SI < 8
34	>1646.4	>1646.4	1	SI < 8	SI > 8
Zanamivir	>903.6	3.3 ± 0.9	>274	No data	No data

3.3. Molecular Modeling Results

Molecular modeling results can be used to develop prediction models. This approach is based on searching for correlations between affinity, for example, the values of energy of binding small molecules (ΔG_{bind}) to the binding site of a biological target, and pIC_{50} values (the negative decimal logarithm of the half-maximal inhibitory concentration). This relationship is most often used to describe the mechanism of biological activity of a number of small molecules, as presented, for example, in [20,21]. We built prediction models based on the relationship between ΔG_{bind} and pIC_{50} values. To do this, we performed flexible molecular docking into the binding site of the PAN endonuclease domain of small molecules containing a pyrimidine fragment (Figures S1–S3). The synthesis of compounds **P1-18**, **SI1-16**, **SII1-11** and the results of their biological testing are described in [9,23,24]. A molecular docking protocol with variation in the values of ligand positional constraints and in terms of the ligand and amino acid residues located within a radius of 5 Å was selected for each specific group of compounds. The main criterion for the adequacy of the molecular docking results was the value of the root mean square deviation (RMSD) of atoms in the obtained docking positions relative to the geometric parameters of the ligand in ligand–protein complexes resolved by X-ray diffraction (PDB ID: 4M5U [23]) and 4KIL [24]). In addition, attention was paid to the arrangement of several ligand atoms at the binding site relative to the Mn^{2+} cations, the values of the binding energy parameters (Docking score, Emodel, etc.), and the presence of undesirable clash interactions between ligand atoms and atoms of amino acid side chains (Figure S6). This resulted in a set of docking solutions, from which the top docking positions, ranked by IFD score, and/or positions with the minimum RMSD value were selected. For the selected docking positions, the binding energy (ΔG_{bind}) was calculated (Figures S7–S9), and dependence diagrams of the pIC_{50} and ΔG_{bind} values were plotted. Thus, we obtained a set of binding energy values for each ligand under study. A natural question arises: which ΔG_{bind} value should be chosen: the minimum or the one corresponding to the minimum RMSD value at the selected docking position? In this study, several options were considered. First, the median ΔG_{bind} values corresponding to each of the docking positions under consideration were calculated. When using median ligand binding energy values, no correlation was observed between the data from biological experiments and the calculation results (Figure 5A). It should be noted here that to obtain the ligand location at the binding site, it was necessary to select a unique molecular docking protocol suitable only for a specific group of ligands. This is since, for a number of molecules, it was not possible to obtain docking solutions (for example, for compounds **P1**, **P2** (Figure S7), **SI-11-13**, **SI-15** (Figure S8), and **SII-2**, **SI-10** (Figure S9)).

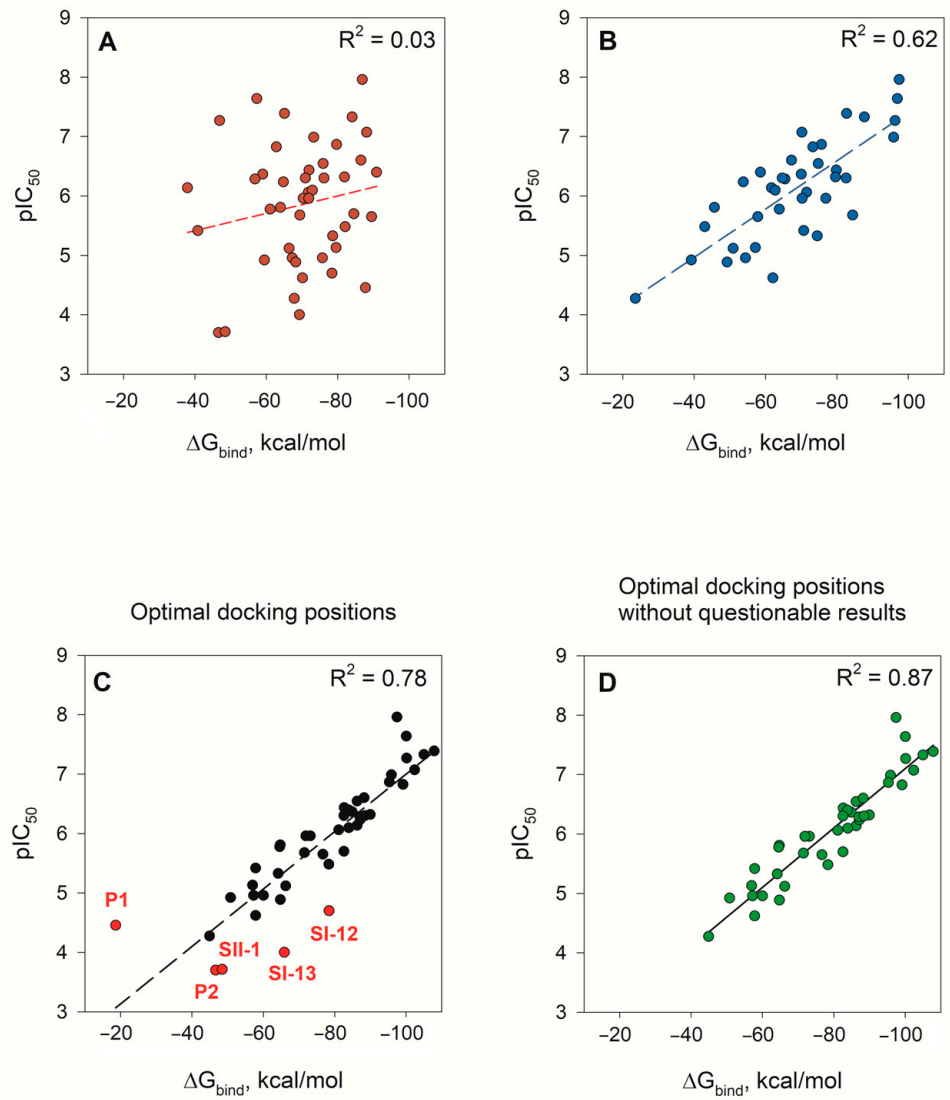


Figure 5. The relationship between the binding energies (ΔG_{bind}) and the negative logarithm of the half-maximal inhibitory concentration pIC_{50}: (**A**) median binding energies of all analyzed docking positions; (**B**) median binding energies obtained using one docking protocol (prediction model 1); (**C**) binding energies calculated for the selected optimal docking positions, outliers are highlighted in red; (**D**) outliers on graph C have been removed (prediction model 2).

It is interesting that unsuccessful calculation results were obtained in most cases for compounds with unreliable IC_{50} values (Figures S1–S3), i.e., the lower limit of the range to which the IC_{50} value belongs is presented.

Secondly, we calculated the median binding energies of the docking positions obtained using the following calculation protocol: PDB ID: 4KIL, SMART pattern: Occ(=O)[nH]c, tolerance = 1, Mn^{2+} (Figures S7–S9). This docking protocol was suitable for most of the ligands under consideration; however, in some cases, no possible ligand position at the binding site was found. As a result, the median ΔG_{bind} values correlate with pIC_{50} for which R^2 = 0.62 (Figure 5B). These results were used to construct a prediction model 1 (model 1).

The third approach involved the analysis of each individual docking solution, taking into account the ligand environment at the binding site. This method yielded a diagram of the relationship between biological assays and theoretical calculations with an accuracy of R^2 = 0.78 (Figure 5C). At that, there are not points characteristic of compounds whose biological assay results appear questionable in the diagram (Figure 5C, red dots). Ligand

P1 (Figure S1) is particularly noteworthy. According to experimental data, compound **P1** exhibits moderate inhibitory activity against PAN. However, the calculations yielded four docking positions, three of which suggest higher inhibitory activity due to their low binding energies, comparable to those characteristic of the leading compounds in this group, **P16-18**. Removing the red dots from the diagram in Figure 5C yielded a more accurate relationship between pCI_{50} and ΔG_{bind} with correlation index $R^2 = 0.87$ (Figure 5D), which was used as prediction model 2 (model 2).

To check the adequacy of the prediction models, a number of points were randomly removed from the diagrams in Figure 5B,D (Figures S10 and S11) and new dependence diagrams were plotted. In prediction model 1, the correlation index decreased significantly during the second validation (Figure S10, Table 2). In prediction model 2, the R^2 value in the validation diagram remained almost unchanged (Figure S11, Table 2). Then, using the real numbers of the straight-line equations y = b[0] + b[1]x, the pIC_{50} values were estimated (Table 2).

Table 2. Results of pIC_{50} value prediction based on molecular docking results.

ID Compound	ΔG_{bind}, kcal/mol	pIC_{50} (exp)	pIC_{50} (pred)	δ, %
Model 1 Validation 1 (Figure S10) $R^2 = 0.630$ b[0] = 3.220, b[1] = −0.043				
P3	−84.41	5.68	6.82	20%
P17	−96.30	7.27	7.33	1%
SII-2	−53.78	6.24	5.52	12%
SII-10	−73.27	6.82	6.35	7%
SI-3	−39.20	4.92	4.89	1%
Model 1 Validation 2 (Figure S10) $R^2 = 0.514$ b[0] = 3.376, b[1] = −0.040				
P5	−70.70	5.42	6.17	14%
P11	−87.70	7.33	6.85	7%
SII-7	−76.90	5.96	6.42	8%
SI-5	−50.93	5.12	5.39	5%
SI-9	−64.68	6.30	5.94	6%
Model 2 Validation 1 (Figure S11) $R^2 = 0.856$, b[0] = 1.986, b[1] = −0.051				
P5	−57.82	5.42	4.95	9%
P11	−105.02	7.33	7.38	1%
SII-7	−73.16	5.96	5.74	4%
SI-2	−44.92	4.28	4.29	0%
SI-9	−88.48	6.30	6.53	4%
Model 2 Validation 2 (Figure S11) $R^2 = 0.877$, b[0] = 1.797, b[1] = −0.054				
P3	−71.51	5.68	5.64	1%
P17	−100.17	7.27	7.18	1%
SII-10	−99.13	6.82	7.12	4%
SI-3	−50.81	4.92	4.53	8%
SI-15	−82.59	5.70	6.23	9%

The relative error in predicting the pIC_{50} value using the first prediction model (model 1) ranges from 1 to 20% (Table 2). The experimental and theoretical pIC_{50}values correlate with each other with a correlation index $R^2 = 0.52$ (Figure 6A). Using the second model (model 2), we obtained pIC_{50} (pred) values with relative errors of no more than 9%. The range of relative errors varies from 0 to 9% (Table 2). In this case, the relationship between the pIC_{50} (pred) and pIC_{50} (exp) values is pronounced, and the correlation index is 0.93 (Figure 6B).

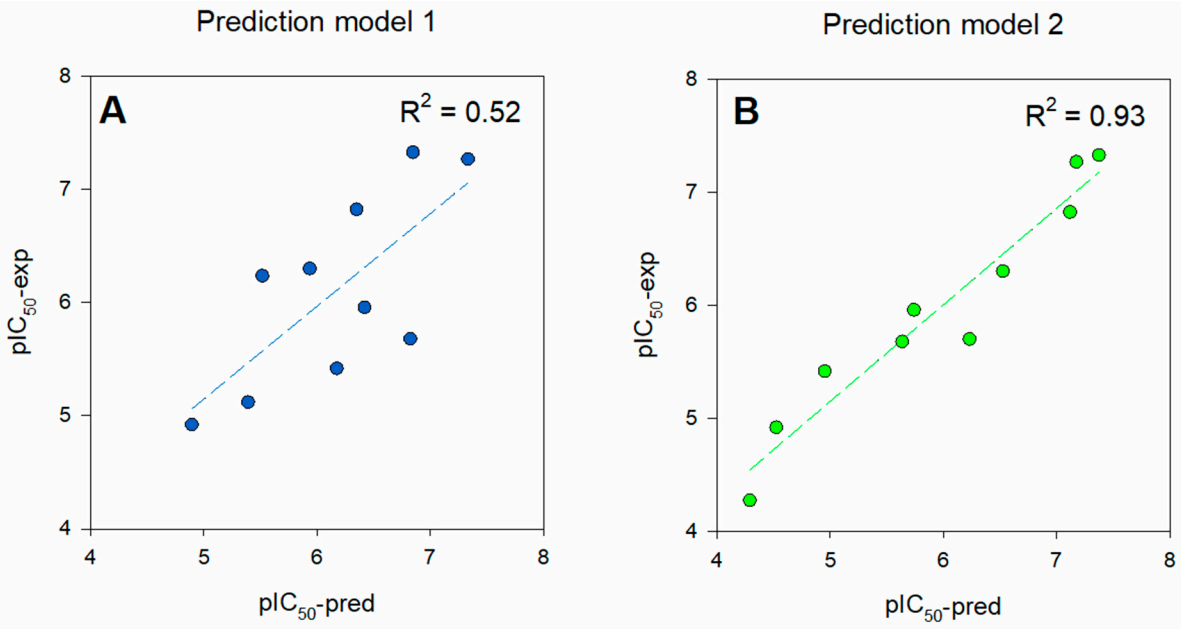


Figure 6. The relationship between predicted and experimental pIC_{50} values: (**A**) the median binding energy values for docking positions obtained using one protocol of molecular docking were used to calculate pIC_{50}; (**B**) the binding energy values obtained for the selected optimal docking solutions were used to calculate pIC_{50}.

It should be emphasized that optimal docking values were selected to construct the second prediction model. Their choice was based on a visual analysis of small-molecule binding, the absence of undesired clash interactions, and the minimal RMSD value, along with the minimal binding energy parameters. At that, docking protocols for different types of molecules may differ, since the primary goal is to find any probable ligand position within the binding site, or, in other words, to place the ligand within the site "at any cost". This method requires attention and time, unlike the second approach, which uses the median binding energy values obtained by means of one of the molecular docking protocols. However, the resulting prediction model (model 2) gives more accurate predictions.

3.4. Affinity of Pyrimidine-4(3H)-one Derivatives to PAN

The constructed prediction models were used to predict the inhibitory potential of pyrimidin-4(3*H*)-one derivatives against the endonuclease activity of the PA protein of influenza virus. To do this, flexible molecular docking procedures were performed for compounds **9**–**34** (Figure 4). Different molecular docking protocols were used, varying the parameters of the ligand positional constraints. However, docking positions were obtained only for four pyrimidin-4(3H)-one derivatives: **11**, **12**, **14**, and **33**. According to the results of biological testing, two of these compounds are characterized by selectivity indices greater than 8: SI (**12**) = 59 and SI (**33**) = 10; the other two compounds exhibit moderate anti-influenza activity (Table 1). The IC_{50} values for compounds **11** and **33** are comparable, but molecule **11** exhibited a greater cytotoxic effect. This result could perhaps be interpreted as follows: hit compounds **12** and **33** likely exhibit anti-influenza activity against the influenza virus by inhibiting the endonuclease activity of the viral protein PAN.

The location of pyrimidin-4(3*H*)-one derivatives in the PA binding site is shown in Figure 7 in comparison with the reference compounds **P16** and **SI-9**.

Compounds **11**, **12**, **14**, and **33** are positioned at the binding site in such a manner that one oxygen atom of the conjugated ring of the ligand coordinates to the Mn^{2+} ion. Using the protein corresponding to PDB ID 4M5U, the ligands interact with Y130 through a water-mediated hydrogen bridge (Figure 7A). The position of the side chain of K134

during binding of pyrimidin-4(3*H*)-one derivatives is noteworthy compared to the binding of the reference compounds. In the case of the reference compounds **P16** and **SI-9**, the side chain of K134 is located in close proximity to ligand oxygen (Figure 7A,B). In this case, a hydrogen bridge is formed between one of the oxygen atoms of **P16** and the NH2 group of the side chain of K134.

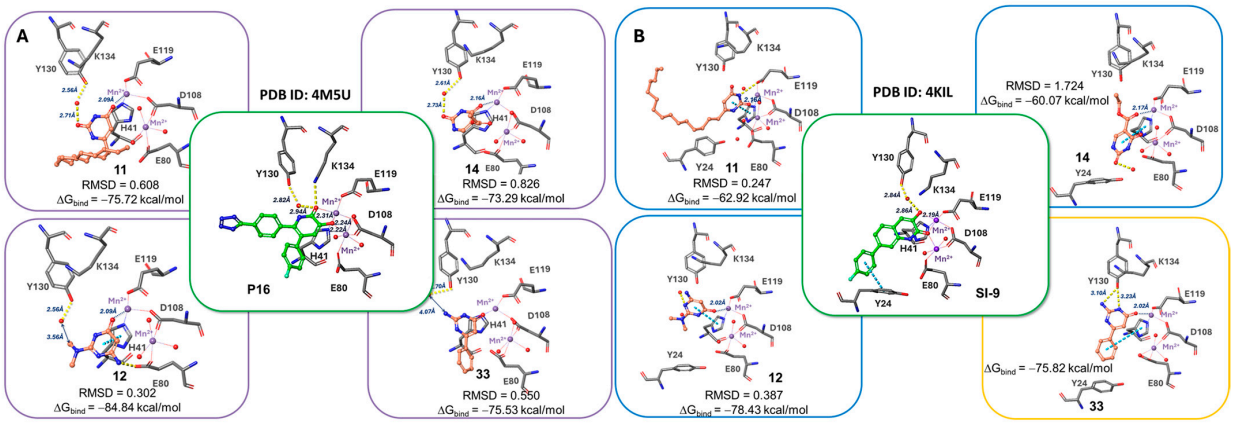


Figure 7. Results of molecular docking of pyrimidin-4(3*H*)-one derivatives at the binding site of the endonuclease domain of RNA polymerase: (**A**) docking to the PA binding site corresponding to PDB ID 4M5U [23]; (**B**) docking to the PA binding site corresponding to PDB ID 4KIL [24]. Purple beads—Mn^{2+} cation, red beads—water molecules; yellow dashed lines—hydrogen bonds, blue dashed lines—π-π stacking interactions. The color of the frames corresponds to the docking protocol (see Figures S7–S9).

The conjugated ring of pyrimidin-4(3*H*)-one derivatives is located in close proximity to the histidine at position 41. Therefore, upon ligand binding at the active site corresponding to the PDB code of 4KIL (Figure 7B), π–π stacking interactions are recorded between the derivative ring and H41. Molecular docking of compound **33** to the active site of PA (4KIL) was successful only in the absence of positional constraints on the ligand. Moreover, binding of ligand 33 is accompanied by patterns similar to those characteristic of binding of derivatives **11**, **12**, and **14**: a hydrogen bridge with Y130, an effect on the side chain of K134, and π–π stacking interaction with the conjugated ring H41.

The energy parameters of molecular docking are given in the Supplementary Materials (Table S1). The pIC_{50} values, estimated on the basis of prediction model 2, characterizing the inhibition of the viral protein PA are presented in Table 3. In general, the predicted pIC_{50} values (pred PAN) agree with the pIC_{50} values (A/H1N1), which characterizes the inhibition of influenza virus replication. At least the most active compound **12** and the least active compound **14** (according to the results of the biological assay) are characterized by high and low pIC_{50} (pred PAN) values, respectively (Table 3, Figure S12).

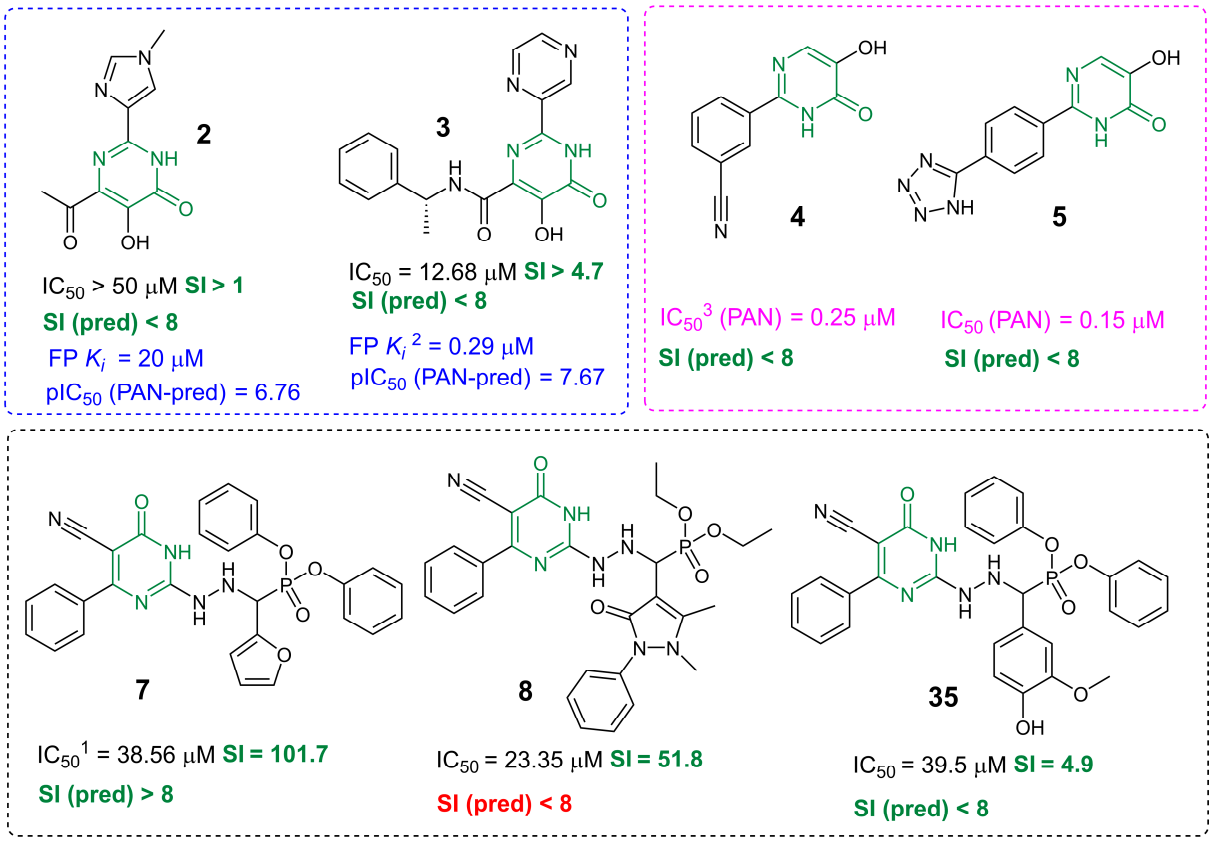
Table 3. Assessment of the predicted pIC_{50} value of PAN inhibition.

ID	pIC_{50} (A/H1N1)	Mediana Values		4M5U (min RMSD)		4KIL (min RMSD, Except 33)	
		ΔG_{bind}, kcal/mol	pIC_{50} (pred-PAN)	ΔG_{bind}, kcal/mol	pIC_{50} (pred-PAN)	ΔG_{bind}, kcal/mol	pIC_{50} (pred-PAN)
11	3.78	−67.43	5.47	−75.72	5.89	−64.92	5.35
12	4.48	−76.45	5.92	−84.84	6.34	−78.43	6.02
14	3.37	−66.68	5.44	−73.29	5.77	−60.07	5.11
33	3.79	−75.53	5.88	−75.53	5.88	−75.82	5.89

Thus, a combined analysis of biological assays and molecular docking results suggests that the mild antiviral activity of the studied pyrimidin-4(3*H*)-one derivatives against the influenza virus of strain A/PR/8/34 (H1N1) is most likely due to their effect on the function of the viral endonuclease. Compound **12**, in this case, can be considered a starting structure that needs to be modified to increase its affinity to the PAN active site.

3.5. Prediction Model Validations

The prediction models developed in this work, based on machine learning and molecular modeling methodology, have been validated on independent compounds; information on the synthesis and biological assay of these compounds is presented in [7,10,24]. The results of model validation are presented in Figure 8. All the structures under consideration include a pyrimidin-4(3*H*)-one fragment.



[1] A/PR/8/34 (H1N1)
[2] competitive binding fluorescence polarization assay with PA_N
[3] endonuclease assay

Figure 8. Results of prediction model validations. The key pyrimidine scaffold is shown in green. The colors of the antiviral activity values correspond to the color used to highlight the biological experiment method.

Compounds **2** and **3** in Figure 8 inhibit the replication of the infectious influenza virus and affect the function of the viral polymerase complex likely by binding to the endonuclease domain [7]. The ML-driven estimate attributes these structures to the unpromising ones, as in both cases the predicted SI value is less than 8, which is consistent with the results of biological assay. The low values of the selectivity index are associated with the high cytotoxicity of the compounds. The pIC_{50} (pred PAN) values estimated based on the molecular modeling results differ by an order of magnitude, which is also consistent with the results of the biological assay.

The PAN inhibitors **4** and **5** [24] (Figure 8) were used to construct prediction models based on molecular modeling. Unfortunately, no information on the inhibition of the replication of infectious influenza virus of any strain by these compounds was found in the scientific literature. Moreover, the prediction model based on machine learning classifies these structures as inactive.

Two antiviral agents **7** and **8** (Figure 8) are characterized by a selectivity index greater than 8, and for one agent this value is less than 8. According to [10], the compounds inhibit the replication of the influenza virus of strain A/PR/8/34 (H1N1), but the mechanism of their antiviral action is unknown. In such cases, the use of a prediction model based on ML methodology is most justified, since the target protein is not defined. Using the fp-RF method allows one to obtain the correct answer in two out of three cases (Figure 8).

It should be noted that our literature review did not yield a sufficiently large number of compounds containing a pyrimidine fragment and tested against the infectious influenza virus, which can be used to validate the prediction models independently. These compounds are not present in our database. In [7], the compounds under study inhibit the endonuclease function of influenza virus at micromolar concentrations; however, only one compound exhibits significant activity against the infectious A/PR/8/34 (H1N1) strain (Figure 1). In all other cases, as in the case shown in Figure 8, the IC_{50} value is presented as the lower limit of the range.

Papers [9,23,24] are devoted to the development of viral endonuclease inhibitors: they present the synthesis of small molecules, their description, biological assay in relation to PAN and even the geometric parameters of ligand-protein complexes (ligand-PAN) deciphered by the X-ray structural analysis. However, there are no data on the inhibition of replication of the infectious strain. In a relatively recent publication [10], out of twelve compounds tested against the influenza virus of strain A/PR/8/34 (H1N1), only three are characterized by an SI value greater than 8. Two of them, including one inactive one, we used to check our calculations. In general, the result of validation of prediction models can be considered satisfactory. In four cases out of five, the prediction result using the ML methodology is correct (Figure 8).

4. Discussion

In this study, we used two approaches to evaluate the antiviral potential of pyrimidin-4(3*H*)-one derivatives. The fp-RF machine learning method classifies all the compounds under study as inactive. Strictly speaking, in two cases an erroneous version was obtained. However, analysis of the results of biological testing of the studied compounds against the influenza virus of strain A/PR/8/34 (H1N1) hardly allows us to classify these molecules as antiviral agents. The SI values are significantly lower than those of zanamivir, which was used as a reference drug. Furthermore, for the well-known anti-influenza drugs oseltamivir and umifenovir, the selectivity index varies from 121 to 1000 depending on the virus strain [50–52]. Nevertheless, based on pharmacological principles [19], compound **12** can be considered as a starting molecule.

This involves molecular modeling methods and prediction models based on the relationship between the binding energy ΔG_{bind} and pIC_{50} values. According to molecular docking results, only four structures of the pyrimidin-4(3H)-one derivatives considered bind to the active site of the endonuclease domain. For two of the four compounds, the selectivity index exceeds 8. The predicted pIC_{50} (pred-PAN) values, characterizing PAN inhibition, correlate with pIC_{50} values obtained from biological testing and demonstrating the ability of compounds to suppress the replication of an infectious influenza virus strain in the MDCK cell line. The mechanism of action of the promising pyrimidin-4(3H)-one derivatives is likely related to their effect on the endonuclease function of the virus.

Compound **12** is located in the binding cavity, coordinating with the Mn^{2+} cation. Given the volume of known PAN inhibitors and the size of the active site, the molecule can be modified to increase its affinity for the target protein.

A combined analysis of calculations and biological assays allows us to conclude that machine learning methodology can be successfully used at the initial stage of selecting potentially active compounds. Molecular modeling methods can describe the mechanism of the antiviral action of small molecules, if the viral target is identified. We believe that both approaches should be used to reduce the physical and financial costs associated with the development of new low-toxicity antiviral agents.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/scipharm94030060/s1. SI-1: Figures S1–S3. PAN inhibitors; Figures S4 and S5. Results of a substructural search; Figure S6. Molecular redocking results; Figures S7–S9. Molecular docking results; Figures S10 and S11. Prediction models; Table S1. Molecular docking results; Figure S12. Prediction models validation. SI-2: NMR spectra.

Author Contributions: Conceptualization, S.S.B. and S.A.G.; methodology, Y.V.G. and A.D.E.; software, S.S.B. and Y.V.G.; validation (in vitro), I.S.B. and A.S.V.; formal analysis, Y.V.G. and A.D.E.; investigation, S.S.B.; synthesis, N.M.A., A.N.L. and S.A.G.; data curation, Y.V.G.; writing—original draft preparation, S.S.B.; writing—review and editing, S.S.B.; visualization, S.S.B.; supervision, S.S.B.; project administration, S.S.B. and S.A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been made in accordance with the research plan of Ufa Institute of Chemistry of the Ufa Federal Research Center of the Russian Academy of Sciences No. 125020601626-9.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Upon request from the corresponding author.

Acknowledgments: The authors are grateful to the theoretical group 'Quanta and Dynamics' (https://monrel.ru accessed on 10 February 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ramesh, D.; Vijayakumar, B.G.; Kannan, T. Therapeutic potential of uracil and its derivatives in countering pathogenic and physiological disorders. *Eur. J. Med. Chem.* **2020**, *207*, 112801. [CrossRef] [PubMed]
2. Schultz, D.C.; Johnson, R.M.; Ayyanathan, K.; Miller, J.; Whig, K.; Kamalia, B.; Dittmar, M.; Weston, S.; Hammond, H.L.; Dillen, C.; et al. Pyrimidine inhibitors synergize with nucleoside analogues to block SARS-CoV-2. *Nature* **2022**, *604*, 134–140. [CrossRef] [PubMed]
3. Ruiz, F.X.; Arnold, E. Evolving understanding of HIV-1 reverse transcriptase structure, function, inhibition, and resistance. *Curr. Opin. Struct. Biol.* **2020**, *61*, 113–123. [CrossRef] [PubMed]
4. Geisman, A.N.; Valuev-Elliston, V.T.; Ozerov, A.A.; Khandazhinskaya, A.L.; Chizhov, A.O.; Kochetkov, S.N.; Pannecouque, C.; Naesens, L.; Seley-Radtke, K.L.; Novikov, M.S. 1,6-Bis[(benzyloxy)methyl]uracil derivatives—Novel antivirals with activity against HIV-1 and influenza H1N1 virus. *Bioorg. Med. Chem.* **2016**, *24*, 2476–2485. [CrossRef] [PubMed]
5. Tomassini, J.E.; Davies, M.E.; Hastings, J.C.; Lingham, R.; Mojena, M.; Raghoobar, S.L.; Singh, S.B.; Tkacz, J.S.; Goetz, M.A. A novel antiviral agent which inhibits the endonuclease of influenza viruses. *Antimicrob. Agents Chemother.* **1996**, *40*, 1189–1193. [CrossRef] [PubMed]
6. Singh, S.B.; Tomassini, J.E. Synthesis of Natural Flutimide and Analogous Fully Substituted Pyrazine-2,6-diones, Endonuclease Inhibitors of Influenza Virus. *J. Org. Chem.* **2001**, *66*, 5504–5516. [CrossRef] [PubMed]
7. DuBois, R.M.; Slavish, P.J.; Baughman, B.M.; Yun, M.-K.; Bao, J.; Webby, R.J.; Webb, T.R.; White, S.W. Structural and Biochemical Basis for Development of Influenza Virus Inhibitors Targeting the PA Endonuclease. *PLoS Pathog.* **2012**, *8*, e1002830. [CrossRef] [PubMed]

8. Hagen, M.; Chung, T.D.; Butcher, J.A.; Krystal, M. Recombinant influenza virus polymerase: Requirement of both 5′ and 3′ viral ends for endonuclease activity. *J. Virol.* **1994**, *68*, 1509–1515. [CrossRef] [PubMed]
9. Sagong, H.Y.; Bauman, J.D.; Patel, D.; Das, K.; Arnold, E.; LaVoie, E.J. Phenyl Substituted 4-Hydroxypyridazin-3(2*H*)-ones and 5-Hydroxypyrimidin-4(3*H*)-ones: Inhibitors of Influenza A Endonuclease. *J. Med. Chem.* **2014**, *57*, 8086–8098. [CrossRef] [PubMed]
10. Hekal, H.A.; Hammad, O.M.; El-Brollosy, N.R.; Salem, M.M.; Allayeh, A.K. Design, synthesis, docking, and antiviral evaluation of some novel pyrimidinone-based α-aminophosphonates as potent H1N1 and HCoV-229E inhibitors. *Bioorg. Chem.* **2024**, *147*, 107353. [CrossRef] [PubMed]
11. Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int. J. Mol. Sci.* **2019**, *20*, 2783. [CrossRef] [PubMed]
12. Hartenfeller, M.; Schneider, G. De Novo Drug Design. *Methods Mol. Biol.* **2011**, *672*, 299–323. [PubMed]
13. Alves, V.M.; Bobrowski, T.; Melo-Filho, C.C.; Korn, D.; Auerbach, S.; Schmitt, C.; Muratov, E.N.; Tropsha, A. QSAR Modeling of SARS-CoV M pro Inhibitors Identifies Sufugolix, Cenicriviroc, Proglumetacin, and other Drugs as Candidates for Repurposing against SARS-CoV-2. *Mol. Inform.* **2021**, *40*, e2000113. [CrossRef] [PubMed]
14. Amiroch, S.; Irawan, M.I.; Mukhlash, I.; Al Faroby, M.H.Z.; Nidom, C.A. Machine Learning for the Prediction of Antiviral Compounds Targeting Avian Influenza A/H9N2 Viral Proteins. *Symmetry* **2022**, *14*, 1114. [CrossRef]
15. Sapozhnikova, T.; Borisevich, S.; Kireeva, D.; Gabdrakhmanova, S.; Khisamutdinova, R.; Makara, N.; Gibadullina, N.; Khursan, S.; Zarudii, F. Effects of novel hexahydropyrimidine derivatives as potential ligands of M1 muscarinic acetylcholine receptor on cognitive function, hypoxia-induced lethality, and oxidative stress in rodents. *Behav. Brain Res.* **2019**, *373*, 112109. [CrossRef] [PubMed]
16. Nguyen, T.H.; Thai, Q.M.; Pham, M.Q.; Minh, P.T.H.; Phung, H.T.T. Machine learning combines atomistic simulations to predict SARS-CoV-2 Mpro inhibitors from natural compounds. *Mol. Divers.* **2023**, *28*, 553–561. [CrossRef] [PubMed]
17. Ferdous, N.; Reza, M.N.; Hossain, M.U.; Mahmud, S.; Napis, S.; Chowdhury, K.; Mohiuddin, A.K.M. Mpropred: A machine learning (ML) driven Web-App for bioactivity prediction of SARS-CoV-2 main protease (Mpro) antagonists. *PLoS ONE* **2023**, *18*, e0287179. [CrossRef] [PubMed]
18. Egorov, A.D.; Gorohov, Y.V.; Kuznetsov, M.M.; Borisevich, S.S. Prediction of the small molecule selectivity index against influenza virus strain A/H1N1 using machine learning methods. *Russ. Chem. Bull.* **2025**, *74*, 851–864. [CrossRef]
19. Kharbriev, R.U. (Ed.) *Guide to Experimental (Preclinical) Studies of New Pharmacological Substances*; Scientific Research Publishing: Moscow, Russia, 2005.
20. Bülbül, E.F.; Melesina, J.; Ibrahim, H.S.; Abdelsalam, M.; Vecchio, A.; Robaa, D.; Zessin, M.; Schutkowski, M.; Sippl, W. Docking, Binding Free Energy Calculations and In Vitro Characterization of Pyrazine Linked 2-Aminobenzamides as Novel Class I Histone Deacetylase (HDAC) Inhibitors. *Molecules* **2022**, *27*, 2526. [CrossRef] [PubMed]
21. Dan, N.T.; Quang, H.D.; Van Truong, V.; Nghi, D.H.; Cuong, N.M.; Cuong, T.D.; Toan, T.Q.; Bach, L.G.; Anh, N.H.T.; Mai, N.T.; et al. Design, synthesis, structure, in vitro cytotoxic activity evaluation and docking studies on target enzyme GSK-3β of new indirubin-3′-oxime derivatives. *Sci. Rep.* **2020**, *10*, 11429. [CrossRef] [PubMed]
22. Ngo, S.T.; Hung, H.M.; Nguyen, M.T. Fast and accurate determination of the relative binding affinities of small compounds to HIV-1 protease using non-equilibrium work. *J. Comput. Chem.* **2016**, *37*, 2734–2742. [CrossRef] [PubMed]
23. Parhi, A.K.; Xiang, A.; Bauman, J.D.; Patel, D.; Vijayan, R.; Das, K.; Arnold, E.; LaVoie, E.J. Phenyl substituted 3-hydroxypyridin-2(1H)-ones: Inhibitors of influenza A endonuclease. *Bioorg. Med. Chem.* **2013**, *21*, 6435–6446. [CrossRef] [PubMed]
24. Sagong, H.Y.; Parhi, A.; Bauman, J.D.; Patel, D.; Vijayan, R.S.K.; Das, K.; Arnold, E.; LaVoie, E.J. 3-Hydroxyquinolin-2(1 H)-ones As Inhibitors of Influenza A Endonuclease. *ACS Med. Chem. Lett.* **2013**, *4*, 547–550. [CrossRef] [PubMed]
25. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef] [PubMed]
26. Lundrum, G.; Paolo, T.; Kelley, B.; Rodrigues, R.; Cosgrove, D.; Vianello, R.; Gareth, P.; Jones, G.; Nadine, D.; Kawashima, E.; et al. *RDKit: Open-Source Cheminformatics Software*; Zenodo: Geneva, Switzerland, 2026. [CrossRef]
27. Bemis, G.W.; Murcko, M.A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893. [CrossRef] [PubMed]
28. Breman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
29. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef] [PubMed]
30. Ross, J.; Belgodere, B.; Chenthamarakshan, V.; Padhi, I.; Mroueh, Y.; Das, P. Large-scale chemical language representations capture molecular structure and properties. *Nat. Mach. Intell.* **2022**, *4*, 1256–1264. [CrossRef]
31. Berman, H.M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
32. Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G.A.; Dahlgren, M.K.; Russell, E.; Von Bargen, C.D.; et al. OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300. [CrossRef] [PubMed]

33. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [CrossRef] [PubMed]
34. Orzeszko, B.; Kazimierczuk, Z.; Maurin, J.K.; Laudy, A.E.; Starościak, B.J.; Vilpo, J.; Vilpo, L.; Balzarini, J.; Orzeszko, A. Novel adamantylated pyrimidines and their preliminary biological evaluations. *Farmaco* **2004**, *59*, 929–937. [CrossRef] [PubMed]
35. Burgula, L.N.; Radhakrishnan, K.; Kundu, L.M. Synthesis of modified uracil and cytosine nucleobases using a microwave-assisted method. *Tetrahedron Lett.* **2012**, *53*, 2639–2642. [CrossRef]
36. Svenstrup, N.; Simonsen, K.B.; Thorup, N.; Brodersen, J.; Dehaen, W.; Becher, J. A Pyrazole to Furan Rearrangement. Thermolysis of 5-Azido-4-formylpyrazoles. *J. Org. Chem.* **1999**, *64*, 2814–2820. [CrossRef] [PubMed]
37. Vlassa, M.; Barabás, A. One-step Synthesis of 3-Oxo-4,4-dimethyl Pentanoic Esters from pivaloyl chloride and esters of malonic acid. *J. Für Prakt. Chem.* **1980**, *322*, 821–825. [CrossRef]
38. Novakov, I.A.; Orlinson, B.S.; Navrotskii, M.B. Desulfurization of 2-Thioxo-1,2,3,4-tetrahydropyrimidin-4-ones with Oxiranes and 2-Haloacetonitriles. *Russ. J. Org. Chem.* **2005**, *41*, 607–609. [CrossRef]
39. Bram, G.; Decodts, G.; Bensaïd, Y.; Farnoux, C.C.; Galons, H.; Miocque, M. N-Alkylation of Pyrimidine and Purine Derivatives (Uracils, Xanthines, Adenine) using Solid/Liquid Phase-Transfer Catalysis without Solvent. *Synthesis* **1985**, *1985*, 543–545. [CrossRef]
40. Deady, L.W. Substitutent Effects in Non-Aromatic Nitrogen Heterocycles: Alkaline Hydrolysis of Methyl N-Methyl (oxo)dihydropyridinecarboxylates and Diaza Analogues. *Aust. J. Chem.* **1985**, *38*, 637–641. [CrossRef]
41. Roth, B.; Smith, J.M., Jr.; Hultquist, M.E. Analogs of Pteroylglutamic Acid. VII. 2-Alkylamino Derivatives. *J. Am. Chem. Soc.* **1951**, *73*, 2864–2868. [CrossRef]
42. Shanmugam, P.; Perumal, P.T. An unusual oxidation–dealkylation of 3,4-dihydropyrimidin-2(1*H*)-ones mediated by $Co(NO_3)_2 \cdot 6H_2O/K_2S_2O_8$ in aqueous acetonitrile. *Tetrahedron* **2007**, *63*, 666–672. [CrossRef]
43. Schultz, O.E.; Warnecke, P. Analogs of nucleic acid bases as antimetabolites. II. *Arzneim.-Forsch. (Drug Res.)* **1967**, *17*, 1060–1064.
44. Helmkamp, G.K.; Kondo, N.S. Purine stacking: Effects of alkyl substituents. *Biochim. Biophys. Acta* **1968**, *157*, 242–257. [CrossRef] [PubMed]
45. Botta, M.; Cavalieri, M.; Ceci, D.; De Angelis, F.; Finizia, G.; Nicoletti, R. 6-alkyl and 5,6-dialkyl-2-methoxy-4(3H)- pyrimidinones in the transformations of pyrimidines—2. *Tetrahedron* **1984**, *40*, 3313–3320. [CrossRef]
46. Anderson, G.W.; Halverstadt, I.F.; Miller, W.H.; Roblin, R.O. Studies in Chemotherapy. X. Antithyroid Compounds. Synthesis of 5- and 6- Substituted 2-Thiouracils from β-Oxoesters and Thiourea. *J. Am. Chem. Soc.* **1945**, *67*, 2197–2200. [CrossRef] [PubMed]
47. Golubyatnikova, L.G.; Khisamutdinov, R.A.; Grabovskii, S.A.; Meshcheryakova, E.S.; Khalilov, L.M.; Kabalnova, N.N.; Murinov, Y.I. Synthesis and Structure of Chloro Complex of Palladium(II) with {[6-Amino-2-(butylsulfanyl)pyrimidin-4-yl]oxy}acetic Acid. *Russ. J. Gen. Chem.* **2019**, *89*, 1808–1815. [CrossRef]
48. Patel, A.; Lewis, W.; Searle, M.S.; Stevens, M.F.G.; Moody, C.J. Synthesis of 6-arylisocytosines and their potential for hydrogen bonding interactions. *Tetrahedron* **2015**, *71*, 7339–7343. [CrossRef]
49. Hanser, T.; Barber, C.; Marchaland, J.F.; Werner, S. Applicability domain: Towards a more formal definition. *SAR QSAR Environ. Res.* **2016**, *27*, 865–881. [CrossRef] [PubMed]
50. McAuley, J.L.; Gilbertson, B.P.; Trifkovic, S.; Brown, L.E.; McKimm-Breschkin, J.L. Influenza Virus Neuraminidase Structure and Functions. *Front. Microbiol.* **2019**, *10*, 39. [CrossRef] [PubMed]
51. Yang, Y.; He, Y.; Li, X.; Dinh, H.; Iyer, S.S. Bifunctional thiosialosides inhibit influenza virus. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 636–643. [CrossRef] [PubMed]
52. Boriskin, Y.; Leneva, I.; Pecheur, E.-I.; Polyak, S. Arbidol: A Broad-Spectrum Antiviral Compound that Blocks Viral Fusion. *Curr. Med. Chem.* **2008**, *15*, 997–1005. [CrossRef] [PubMed]