Biotechnological Journal of Environmental Microorganisms (BJEM)

Biotechnological Journal of Environmental Microorganisms(BJEM) 1(1) 2022 1-14

Molecular modeling of the *Toxoplasma gondii* adenosine kinase inhibitors

Taraneh Zandi Asadabadi*

Department of Microbiology, Faculty of Basic Sciences,Lahijan Branch, Islamic Azad University (IAU), Lahijan, Guilan, Iran.

Received:10 March 2022 / Revised:18 April 2022 /Accepted: 18 April 2022

Abstract

Computer-assisted approaches might be seen as a bridge to novel medication discoveries. In 2014, the World Health Organization declared antibiotic resistance in microorganisms to be a major global threat due to which simple diseases that were formerly manageable have now become deadly infections. Microbial resistance is a form of drug resistance in which a microorganism may live even when antibiotics are present in the environment. Toxoplasmosis is a major worldwide parasitic infection caused by Toxoplasma gondii. Since Toxoplasma gondii is not capable of purine synthesis, the protein adenosine kinase (EC.2.7.1.20) is an important enzyme in its life pathway. Therefore, Toxoplasma gondii adenosine kinase has recently been considered a target for developing anti-Toxoplasma agents. This study aimed to develop a 3D QSAR model to predict the activity of adenosine kinase inhibitors in Toxoplasma gondii and to find new potent inhibitors. Acceptable values of 0.98, 0.83, and 0.91 were observed for the goodness of fit (R_2), internal cross-validation (Q_2), and external cross-validation (R_2pred) indices, respectively. The robustness of the model was confirmed by applying the Y-scrambling analysis, and values of ~ 0.18 and ~ 0.0025 were observed for R_2intercept and Q_2intercept, respectively. This confirmed that indices calculated for the original model were not based on the chance correlation between independent and dependent variables. Following the structural virtual screening, new ligands were proposed using the SwissSimilarity web tool and the ZINC database. The SwissADME web tool was used to predict the pharmacokinetic properties of the new compounds, and a promising compound was suggested for further research.

Key words: Adenosine Kinase; Drug Design; Toxoplasma Gondii; QSAR modeling; ADME properties

*Corresponding Author: E-mail: taraneh.zed@gmail.com

1. Introduction

Toxoplasmosis, which is found all over the globe, is a serious parasitic infection caused by Toxoplasma gondii. It's worth noting that it's one of the most prevalent parasite infections of the central nervous system, as well as a serious public health issue affecting over a billion people worldwide (Babaie et al., 2019). Toxoplasma gondii is an obligate intracellular parasite that has the potential to cause infection in warm-blooded animals. Cats and feline cats are the only final hosts of this parasite that spread parasite oocysts through feces into the environment (Nicolle, 1908).

In order to detect toxoplasmosis, the most commonly used methods include molecular, biochemical, histological, immunohistochemical, direct spread, or a combination of these methods (Sterkers et al., 2011). Medical history, physical examination (Lin et al., 2000), and blood tests (Hill & Dubey, 2002) are all used to determine the presence of an infection. Anti-Toxoplasma antibodies (IgM and IgG antibodies) are routinely measured in the blood (Switaj et al., 2005). There are also polymerase chain reactions to detect infection in blood, amniotic fluid, and cerebral spinal fluid (Elmore et al., 2010). Stool testing for oocysts and serological testing are used to determine the presence of an infection in cats. Particularly important in terms of human health and infecting other intermediate host animals are kittens that are infected for the first time because of their lack of immunity with abundant oocyst excretion and environmental contamination (Zenner et al., 1999). Due to the importance of Toxoplasma gondii in medicine and veterinary medicine, the biological characteristics of this parasite can be predicted to some extent by using genotyping techniques, and the prevalence of infection in human and animal populations can be reduced by using control and prevention methods, including DNA vaccine design (Homan et al., 2000). Treatments for toxoplasma have had poor results and a wide range of undesirable side effects. In addition, there is presently no vaccination to treat the disease. Due to such shortcomings, the development of new, effective medications with fewer side effects is a critical and fundamental need for treating toxoplasmosis.

Rational drug design depends on physiological and biochemical differences between pathogen and host, and in the case of Toxoplasma gondii, the purine metabolism pathway is one of the most important goals in drug design studies against this pathogen (el Kouni, 2007). Toxoplasma gondii is not able to synthesize purines and depends on recovery pathways to meet its need for purines (Ngô et al., 2000). Adenosine kinase and adenosine monophosphate are the most important enzymes in the parasite's need for purine in recovery pathways (Reddy et al., 2008). Biochemical, metabolic, and molecular research on chemical performance and structure shows that Toxoplasma gondii adenosine kinase is an important and very suitable chemotherapy target for the treatment of toxoplasmosis (el Kouni, 2007). Toxoplasma gondii adenosine kinase is a monomeric protein with a length of 363 residues and a weight of 39.3 kDa (Fig. 1), which uses the ATP y-phosphate group as a phosphate donor by phosphorylation of adenosine to adenosine 5'-monophosphate (AMP) (Al Safarjalani et al., 2008), and because purine is essential for Toxoplasma gondii and other parasites, inhibition of the salvage pathway could stop the growth of this parasite. Benzyl adenosine analogues, also known as invasive substrates, are metabolized to the nucleotide level and become selectively toxic against this parasite but have no toxic effect on the host (el Kouni, 2007).

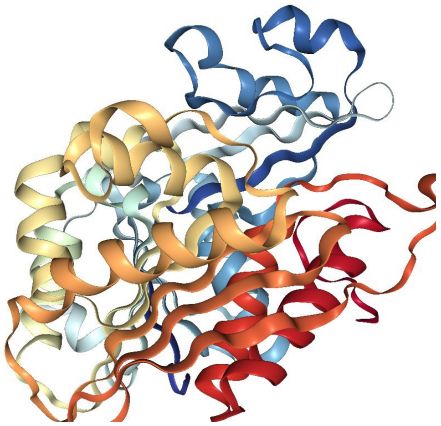


Figure 1. 3D structure of Toxoplasma gondii adenosine kinase (PDB ID: 2a9y)

Recently, 41 analogues of benzyl adenosine have been specified, which have been reported as invasive and, of course, potent substrates for Toxoplasma gondii adenosine kinase [el Kouni, 2007; Kim et al., 2008). The 7-Deaza adenosine is also an excellent ligand for Toxoplasma gondii adenosine kinase. Compared with 6-benzyl inosine and 7-Deaza inosine, the 7-Deaza-6-benzylthioinosine analogues are better ligands for Toxoplasma gondii adenosine kinase (Kim et al., 2008). All 7-Deaza-6-benzylthioinosine analogues showed a selective antitoxoplasmic effect against wild-type parasites. The proximity of the anti-toxoplasma and the toxic host with a distinct substitution on the aromatic ring enhances the proximity of Toxoplasma gondii adenosine kinase compared to when N6-Benzyladenosine is non-substituted. The difference in the position of these substitutes on the aromatic ring is in the different degrees of strength and ability of anti-toxoplasma factors (Al Safarjalani et al., 2008). These results indicate that Toxoplasma gondii adenosine kinase is an excellent target for chemotherapy and its compounds are anti-Toxoplasma factors.

Adenosine kinase inhibitors received much attention for drug production in the late 1990s and early 2000s because they prevent the secretion of adenosine metabolism in the body, and therefore reduce side effects (Cook et al., 2000). Therapeutic applications that include adenosine kinase-lowering performance modification include pharmacy (Singh et al., 1996), gene therapy, cell therapy, ketogenic diets, and transcriptional suppression (Cook et al., 2000). Prokaryotic and eukaryotic microorganisms usually have specific compatibility in their nucleotides. Nucleotide metabolism in microorganisms is different from that of the host. This difference can be used to produce antiparasitic drugs that are specific to parasites but do not affect the host. Mycobacterium tuberculosis is a species of pathogenic bacterium that causes tuberculosis whose adenosine kinase was the first bacterial adenosine kinase to be cloned. Currently, halogenated 3-deaza-adenosine analogues can be considered as anti-tuberculosis agents (Vodnala et al., 2008).

Traditional techniques of drug manufacture have long been a source of consternation and frustration for scientists due to their high production costs and the perception that they would take an eternity to scale up (Johns Hopkins Bloomberg School of Public Health, 2018). To this end, pharmaceutical companies have been looking for new and more accurate ways to spend money and speed up the drug production process. Due to important and ubiquitous computer science advances, initiatives to use this technology in the medical area in order to greatly minimize the challenges faced have been made. In point of fact, computer-assisted methods have the potential to be seen as a pathway leading to the discovery of innovative medications. These techniques center mostly on compound modeling and information analysis, both of which are carried out via the use of software calculations that are characterized by a high level of precision as well as an increased likelihood of accomplishment. It was not until the 1960s that computer-related methods for discovering and developing drugs were found to be potentially effective in making chemical compounds and inactivating inappropriate components and compounds.

QSAR is one of the most widely used methods in chemometrics because it seeks to find a meaningful relationship between structure and function. This is accomplished through the application of mathematical models such as multivariate linear regression, least squares regression, and other similar models (Brereton, 2003). The therapeutic and biological functions of a medicine are dependent not only on the bonding angles between the atoms but also on the capacity of the molecule to pass through the cell membrane and disperse its electrical charges, as well as to make hydrogen bonds with other molecules, and so on, to the degree that it can be claimed that QSAR methods can predict new compounds from the chemical structure and biological activity of compounds (Burger, 1970).

The biological activity of drugs depends on the binding of the receptor protein or enzyme, which leads to the formation of the drug-receptor complex. Structural properties determine the properties and activity of chemical compounds, and QSAR can show the relationship between the structural properties and the physical and

chemical properties of compounds well. In fact, QSAR can determine the relationship between the physico-chemical behavior of a molecule and its structural units. Thus, scientists can predict the behavior of molecules based on their structure. A variety of scientific disciplines are used in the process of designing drugs, as well as technologies such as QSAR, which play a significant part in the process of developing and perfecting compounds. The QSAR approaches may be broken down into a few different classes, each of which describes descriptors according to structure (Göller et al., 2006) Taking into consideration the information presented above, the purpose of this research was to investigate the most effective way to inhibit adenosine kinase in Toxoplasma gondi by using the three-dimensional QSAR approach.

2. Methods

2.1 Data Selection

The structures of inhibitory compounds along with their biological functions were obtained from the BindingDB (Chen et al., 2001; Liu et al., 2007). The BindingDB is one of the most popular databases used for drug design research. Data were screened before modeling, and bad data were removed from the dataset based on the following criteria:
1)Data belonging to organisms other than humans.
2)Performances reported on scales other than Ki or with inaccurate values (e.g. Ki <1000).
3)Compounds for which more than one amount of biological performance (functioning) has been reported.

After screening, the biological performances of the remaining 87 compounds were returned to the pKi (log Ki × 10-9).

2.2 Calculating the Descriptors of Chemical Compounds

GRIND descriptors are one of the most important molecular descriptors for studying the structure and interaction of ligands. These descriptions are 3D descriptors that represent the ability of the molecule to form appropriate interactions with independent pharmacophoric groups. The production of these descriptors includes the calculation of molecular interaction field maps. These descriptors are independent of the alignment of structures and are therefore a good way to describe inhibitory molecules with different structures. Drawing and optimizing the structure of ligands has been conducted through the use of the Sybil software (version 7.3). The partial charge of the atoms was calculated by the extended Hückel method. Then, to calculate Grid Independent Descriptors (GRINDs), the AMANDA algorithm (Duran et al., 2008; Wold et al., 1993) was applied. The operation steps of the AMANDA algorithm are as follows:
1)Calculation of the MIFs (molecular-interaction fields) and identification of favorable interaction energies called nodes,
2)Node filtration (to find a set of regions with the most favorable interaction energies) and
3)Encoding of the final remaining nodes into the GRINDs descriptors. After multiplication of the interaction energy pairs, the greatest product is kept for each internode distance.

Herein, the molecular interaction field maps for N1, O, and DRY probes have been calculated by the grid software. These probes represented hydrophobic interactions, hydrogen bond acceptor groups, and hydrogen bond donor groups. A 0.5 angstrom distance was obtained between the grid points. With six molecular interaction field maps and a smoothing window of 0.4 angstroms, 1200 descriptors were calculated for each chemical compound.

2.3 Variable Selection

Descriptors can determine the correctness or incorrectness of modeling. In other words, modeling will not be designed properly if the choice of descriptors is inappropriate. This expresses the importance and value of descriptors. Therefore, their selection must be done very carefully to depict the specific properties of the molecule. Given the problems raised, it is necessary to choose the appropriate variable to obtain useful and accurate information from the many descriptors that exist.

Genetic algorithm is a type of natural evolution in which variables (different descriptors) play the role of genes in a particular species (Baroni et al., 1993; Leardi, 2000). Through selection, mutation, and genetic crossover, better values than

the fitness function are sought. The population with the highest fitness function remains for the next generation. The genetic algorithm navigates the target space by randomly creating genetic mutations, combining variables in the process of genetic crossover, and combining variables in the process of genetic crossover. In fact, genetic algorithms are used in optimization, which means that they affect the variables of a problem if information is not available from the nature of the problem, and this feature is unique to the genetic algorithm and therefore distinguishes this method from other methods.

2.4 Multivariate Modeling

Using the Kennard–Stone method (Kennard et al., 1969), the whole dataset was partitioned into two groups—one for training, consisting of eighty percent of the data, and the other for testing, consisting of twenty percent. The ability to produce models¬ with high predictive power, even in the presence of perturbations, alignments, and redundant variables, has made the partial least squares (PLS) technique one of the most popular achievements in multivariate modeling. PLS is a more evolved approach than the principal component analysis technique. In PLS, the descriptors and biological activities are simultaneously reflected on PLS components, which are also known as latent variables. This is done in order to understand the linear relationships that exist between the descriptors and biological activities. The regression coefficients calculated by the output of the PLS show the direction and intensity of the X effect on Y. For systems such as a protein-ligand complex system, the PLS regression equation can be expressed as Equation 1:

$$y = \bar{y} + \sum_{l=1}^{L} coeff_l D_l \quad \text{(Equation 1)}$$

where $coeff_l$ are the regression coefficients and D_l are the descriptors of the ligand. To build the PLS model presented in this study, the PLS Toolbox 3.5 software available in MATLAB was used.

2.5 Predicting, Evaluating the Validity, and Interpreting the Model

Normally, parameters are required for a model to be scientifically valid. In the normal case of a model with fewer independent variables, the calculation of variables is naturally easier and more efficient than in other models. Also, a model whose standard error (mean square error) is low and whose correlation coefficient (R2) is closer to 1 will be a more appropriate model. Predictive power is measured by predictive groups that have no role in the model development process. The cross-validation method is used to make a broad prediction of the model. That is, all available molecules will be predicted. The cross-validation method uses another parameter called Q2 instead of the correlation coefficient (R2) (Gramatica, 2007). The validation parameter R2 indicates the goodness of fit of the model. In fact, this parameter shows how well a regression model fits into the training set. R2 is the percentage of the total response variance explained by the regression model and calculated by the following

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_{i-estimated})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \text{(Equation 2)}$$

$\bar{y}_{i\_estimated}$

where yi and are the observed and estimated responses of the model, respectively. The value of $\bar{y}$ is the average of the observed response variables. The closer R2 is to 1, the better the quality of the model and the smaller the estimation error. However, if the number of descriptors is relatively large compared to the number of observations, the existence of a chance correlation can lead to the development of simple models with good fit characteristics. In addition, the R2 parameter has nothing to do with the model's ability to perform well on the forthcoming datasets, and the model training error is a poor estimate of the test set error.

More validations are needed to avoid models containing only chance correlation. One of the most popular techniques for estimating model prediction errors is the cross-validation method. For the model, the cross-validation (Q2) represents the variance as specified in the prediction and is calculated by the following equation:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_{i-predicted})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \text{(Equation 3)}$$

where yi and $\hat{y}_{i\_estimated}$ the observed and the predicted responses during the cross-validation process. $\bar{y}$ is the average for the observed response variables. The venetian blind cross-validation technique was conducted on the training set (approximately 80% of the dataset), while the test set (approximately 20% of the dataset) was used to confirm the predictive power of the constructed model. The Y-scrambling technique was used to evaluate the coherence of the model (Gramatica, 2007; Tropsha et al., 2003). To perform the Y-scrambling technique, the observed data (Y) (biological performance) was randomly scrambled for 100 times while the X-matrix (descriptors) was preserved intact. Then the values of R2 and Q2 were calculated for 100 made models. Next, the graph of the values R2 and Q2 obtained for the reference model and the made models after Y-scrambling was plotted against the correlation coefficients between reference Ys and scrambled Ys. Y-intercept points for R2 and Q2 (Q2 intercept and R2 intercept) were calculated by crossing a regression line through the data points. Previous studies have shown that any regression model with $R^2_{intercept}<0.3$ and $Q^2_{intercept}<0.05$ can be considered as a coherent model (Eriksson et al., 2003).

2.6 Virtual Screening

Virtual screening is one of the methods that are widely used to identify powerful inhibitors. In this study, virtual screening was conducted by means of the SwissSimilarity web tool. SwissSimilarity's small molecule database is divided into the following four categories:

1) Drug molecules extracted from the Drug Bank (Wishart et al., 2018). This set of compounds has been classified as one of the subgroups related to approved drugs (1500 compounds), laboratory drugs (4800 compounds), drugs under investigation (500 compounds), discarded drugs (160 compounds), illegal compounds (170 compounds), and dietary supplements (78 compounds).

2) Small bioactive molecules. From this group, we can refer to a set of ligands that are in complex with macromolecular structures in the protein database (PDB) (Burley et al., 2019). These ligands were retrieved from databases such as LigandExpo (19,500 combinations) (LigandExpo, 2019), ChEMBL (177,000 combinations) (Gaulton et al., 2019), and ChEBI (28,000 combinations) (Hastings et al., 2016).

3) Commercially available compounds stored in the ZINC database (Irwin et al., 2012). This set of compounds is categorized in one of the subgroups related to drug-like molecules (10,600,000), lead-like molecules (4,300,000), fragment-like molecules (700,000) or compounds grouped by sellers (9,700,000).

4) A collection of 205 million virtual combinations that can be easily synthesized by existing commercial agents. These compounds can be synthesized through a single-step chemical reaction (Hartenfeller et al., 2012) and refined for chemical stability and non-toxicity.

2.7 Investigation of Pharmacokinetic Parameters and Drug-like Properties

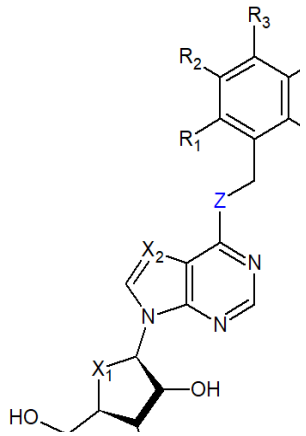
To evaluate the pharmacokinetic parameters and drug-like properties of compounds obtained from the virtual screening stage, indicators such as molecular weight, lipid solubility, drug similarity, polar surface area, and toxicological risk assessment (using the SwissADME web tool (http://www.swissadme.ch) were evaluated.

3. Results and Discussion

3.1 Data Selection

Since the growing phenomenon of antibiotic resistance is one of the most important concerns today and a serious threat to public health, extensive clinical efforts have been made to develop new antibacterial agents to effectively treat bacterial infections. Since the adenosine kinase enzyme plays a very important role in the life cycle of Toxoplasma gondii, it is recognized as an important target for the development of anti-toxoplasmosis compounds. Therefore, inhibition of adenosine kinase bioactivity by specific chemical inhibitors can be a powerful strategy to fight against the growing phenomenon of antibiotic resistance in related pathogens. The structures of inhibitory compounds along with the biological activities of these compounds, extracted from the BindingDB database, are shown in Table 1.

1- Table 1. Structure and Biological Activities of Inhibitor Compounds Extracted from the BindingDB Database



Comp.	X_1	X_2	Z	R_1	R_2	R_3	R_4	R_5	pK_i
M1*	O	N	S	H	H	H	H	H	5.62
M2	O	N	S	Cl	H	H	H	H	5.85
M3	O	N	S	CH_3	H	H	H	H	5.82
M4	O	N	S	H	NO_2	H	H	H	5.52
M5*	O	N	S	H	CH_3	H	H	H	5.88
M6	O	N	S	H	CF_3	H	H	H	5.54
M7	O	N	S	H	H	F	H	H	5.92
M8	O	N	S	H	H	Cl	H	H	5.68
M9	O	N	S	H	H	Br	H	H	5.12
M10	O	N	S	H	H	CN	H	H	6.04
M11	O	N	S	H	H	NO_2	H	H	5.96
M12	O	N	S	H	H	CH_3	H	H	5.48
M13	O	N	S	H	H	OCH_3	H	H	5.58
M14*	O	N	S	H	H	OCF_3	H	H	3.85
M15	O	N	S	H	H	t-Bu	H	H	3.95
M16*	O	N	S	H	H	cooc H3	H	H	4.02
M17	O	N	S	Cl	H	Cl	H	H	6.15
M18	O	N	S	H	Cl	Cl	H	H	5.60
M19	O	N	S	Cl	H	H	H	F	4.42
M20	O	N	S	CH_3	H	CH_3	H	CH_3	3.82
M21	O	N	NH	H	H	H	H	H	3.74
M22	O	N	NH	F	H	H	H	H	4.29
M23	O	N	NH	Cl	H	H	H	H	4.92
M24	O	N	NH	NO_2	H	H	H	H	4.43
M25	O	N	NH	OCH_3	H	H	H	H	4.95
M26*	O	N	NH	H	F	H	H	H	4.08
M27	O	N	NH	H	CH_3	H	H	H	4.60

M28*	O	N	NH	H	H	F	H	H	4.28
M29	O	N	NH	H	H	Cl	H	H	4.09
M30	O	N	NH	H	H	CN	H	H	4.74
M31	O	N	NH	H	H	CF_3	H	H	4.00
M32	O	N	NH	H	H	CH_3	H	H	4.57
M33	O	N	NH	H	H	i-Pr	H	H	4.33
M34	O	N	NH	H	H	cooc H3	H	H	4.05
M35	O	N	NH	H	H	OCH_3	H	H	5.00
M36	O	N	NH	F	H	F	H	H	4.91
M37	O	N	NH	Cl	H	Cl	H	H	4.69
M38	O	N	NH	H	Cl	Cl	H	H	4.45
M39	O	N	NH	F	H	NO_2	H	H	3.77
M40	O	N	NH	Cl	H	CN	H	H	4.27
M41	O	N	NH	F	H	OCH_3	H	H	4.59
M42	O	N	NH	OCH_3	H	OCH_3	H	H	4.56
M43*	O	CH	S	H	H	H	H	H	4.55
M44	O	CH	S	F	H	H	H	H	4.23
M45*	O	CH	S	Cl	H	H	H	H	4.44
M46	O	CH	S	Br	H	H	H	H	3.97
M47	O	CH	S	CH_3	H	H	H	H	3.64
M48*	O	CH	S	H	NO_2	H	H	H	4.34
M49	O	CH	S	H	CF_3	H	H	H	4.66
M50	O	CH	S	H	CH_3	H	H	H	5.07
M51	O	CH	S	H	H	F	H	H	4.42
M52*	O	CH	S	H	H	Cl	H	H	4.96
M53	O	CH	S	H	H	Br	H	H	3.60
M54	O	CH	S	H	H	NO_2	H	H	4.45
M55	O	CH	S	H	H	CN	H	H	5.28
M56	O	CH	S	H	H	CO_2CH_3	H	H	5.15
M57	O	CH	S	H	H	CH_3	H	H	4.72
M58	O	CH	S	H	H	t-Bu	H	H	4.62
M59	O	CH	S	H	H		H	H	3.74
M60	O	CH	S	H	H	OCF_3	H	H	4.45
M61	O	CH	S	H	H	OCH_3	H	H	5.24
M62	O	CH	S	F	H	F	H	H	4.79
M63	O	CH	S	Cl	H	Cl	H	H	4.58
M64	O	CH	S	H	Cl	Cl	H	H	4.55
M65	O	CH	S	F	H	H	H	Cl	5.14
M66	O	C-I	S	H	H	H	H	H	4.08
M67*	CH_2	N	S	H	H	H	H	H	3.98
M68*	CH_2	N	S	F	H	H	H	H	4.26
M69	CH_2	N	S	Cl	H	H	H	H	4.62

M70	CH_2	N	S	CH_3	H	H	H	H	4.60
M71*	CH_2	N	S	H	NO_2	H	H	H	4.57
M72	CH_2	N	S	H	CF_3	H	H	H	4.66
M73	CH_2	N	S	H	CH_3	H	H	H	4.57
M74	CH_2	N	S	H	H	F	H	H	4.49
M75*	CH_2	N	S	H	H	Cl	H	H	4.79
M76	CH_2	N	S	H	H	Br	H	H	4.68
M77	CH_2	N	S	H	H	NO_2	H	H	4.03
M78	CH_2	N	S	H	H	CN	H	H	3.89
M79	CH_2	N	S	H	H	CO_2CH_3	H	H	4.39
M80*	CH_2	N	S	H	H	CH_3	H	H	5.12
M81	CH_2	N	S	H	H	SO_2CH_3	H	H	4.13
M82	CH_2	N	S	H	H	OCF_3	H	H	4.18
M83	CH_2	N	S	H	H	OCH_3	H	H	4.21
M84*	CH_2	N	S	Cl	H	Cl	H	H	4.21
M85	CH_2	N	S	H	Cl	Cl	H	H	4.77
M86*	CH_2	N	S	F	H	H	H	H	4.51
M87	CH_2	N	S	CH_3	H	CH_3	H	H	4.20

*: test set

3.2 3D QSAR Modeling and Validation of the Model

Structure-dependent descriptors were described by the AMANDA algorithm and molecular interaction fields from DRY, N1, and O probes to calculate the GRIND descriptors. Prior to modeling, the feature selection process to find the most appropriate structure-dependent descriptors was performed. The relationships between the descriptors related to the selected structures and bioactivity (pKi) were investigated using PLS modeling. Figure 2 shows the diagram of the experimental pKi versus the pKi predicted by the 3D QSAR model. Acceptable values of 0.98, 0.83, and 0.91 were observed for the goodness of fit (R2), internal cross-validation (Q2), and external cross-validation (R2pred) indices, respectively.

The robustness of the model was also confirmed according to the results obtained from the scrambling analysis and the observed values for R2intercept and Q2intercept, which were ~0.18 and ~0025/0, respectively (Fig. 3). The results of R2intercept and Q2intercept showed that the optimal indices calculated for the original model were not based on the chance relationship between independent and dependent variables.

3.3 Virtual Screening

Virtual screening was performed using the SwissSimilarity web tool to search for chemical compounds that have the potential to inhibit the adenosine kinase enzyme. The ZINC database was selected for screening, and the search parameters were set to default. The most active chemical compound in the adenosine kinase inhibitor group (Table 1) was used as the input compound for the screening process. At the end of the screening process, 400 compounds that were most similar to the input compound in terms of structural parameters were extracted from the ZINC database.

In order to predict the inhibitory potential of database-derived compounds for adenosine kinase, the performance of these compounds was predicted in the test group by the constructed 3D QSAR model. Among the 400 compounds extracted from the ZINC database, those compounds that showed more activity than the input compound were nominated as compounds with inhibitory potential for adenosine kinase (Table 2). Then the pharmacokinetic parameters and drug-like properties of the candidate compounds were further studied.

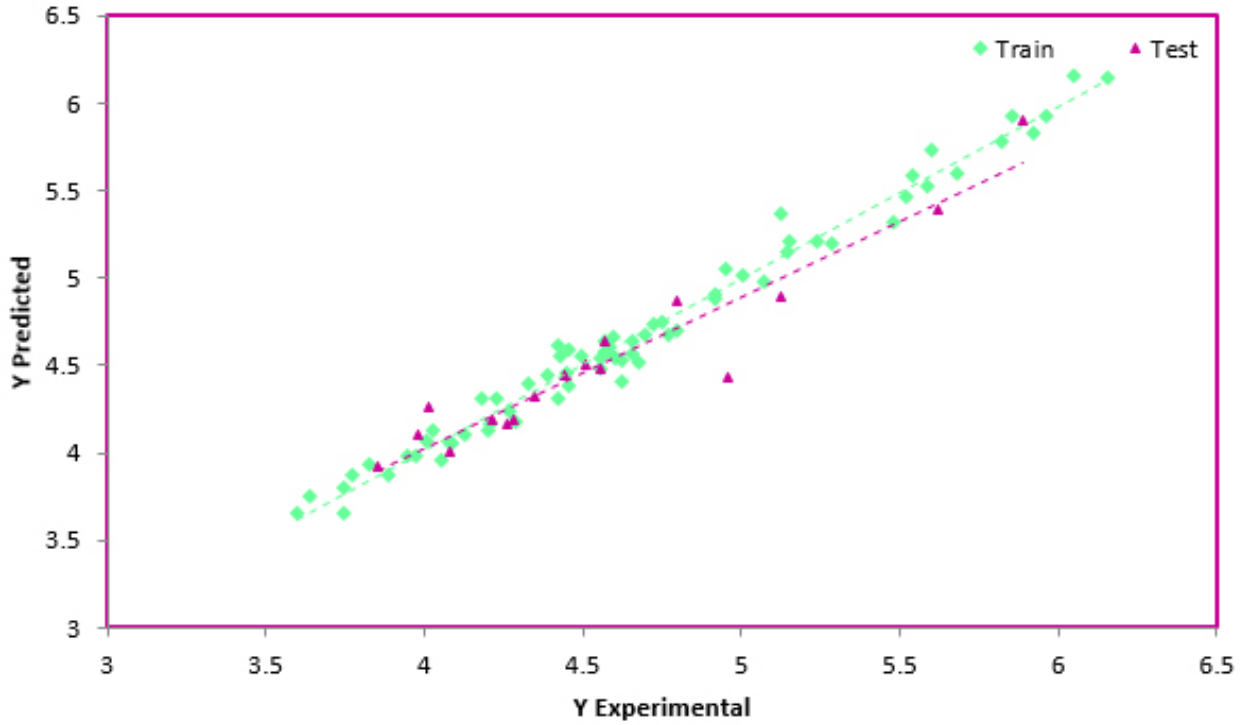


Figure 2. Plot of Experimental pKi Values Versus Predicted pKi Values for Training and Testing Datasets

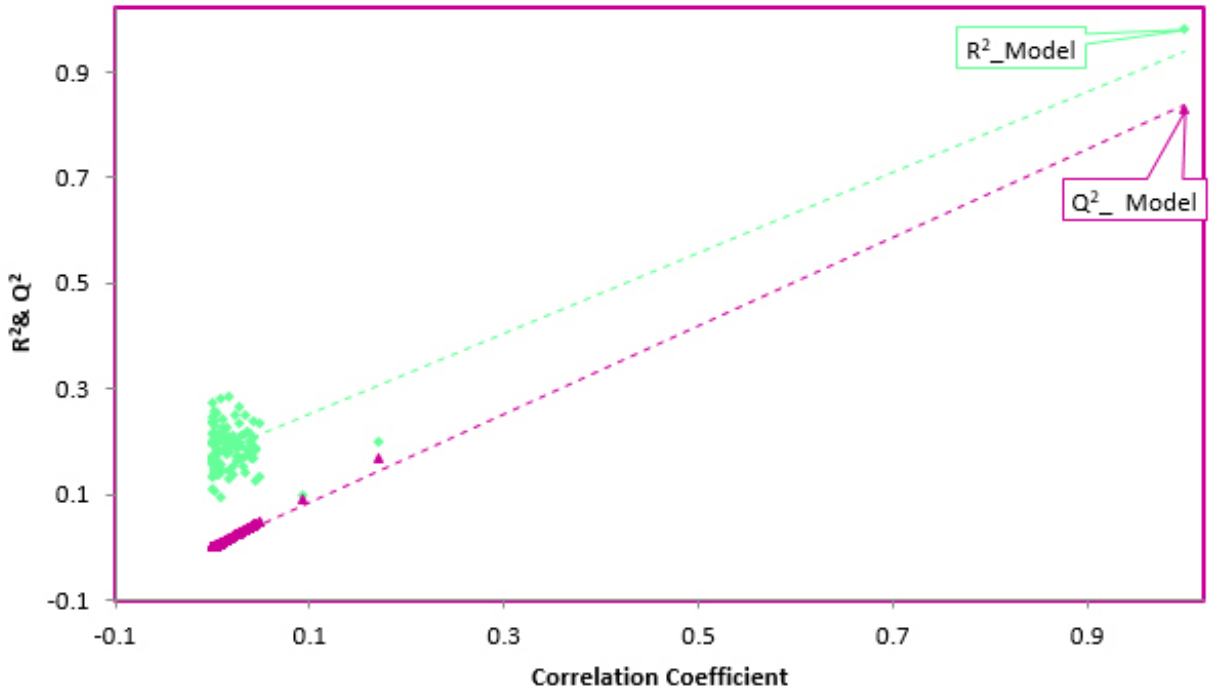


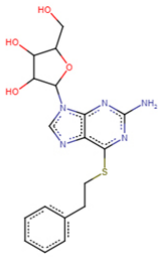
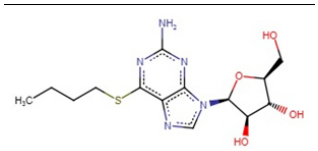
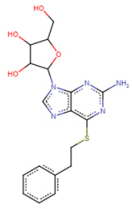
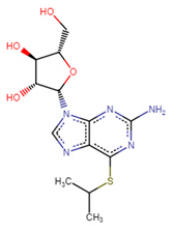
Figure 3. Y-scrambling plot of pKi for the QSAR model. The Y-axis represents R2 (green) and Q2 (violet) coefficients for the original model and 100 models built based on randomly scrambled response data. The X-axis represents the correlation coefficient between the original and permuted response data.

Table 2. Compounds with Inhibitory Potential for Adenosine Kinase

ZINC ID	Similarity%	pKi (calc.)
ZINC00057394	0.575	6.30
ZINC01668084	0.601	7.05
ZINC01669236	0.67	6.28
ZINC01675226	0.564	6.63
ZINC01675229	0.564	6.47
ZINC01689682	0.564	6.73
ZINC02044304	0.564	6.37
ZINC04311911	0.637	7.10
ZINC04311912	0.637	7.10
ZINC04945807	0.601	6.79
ZINC04945809	0.601	7.06
ZINC04964106	0.564	6.23
ZINC04964109	0.564	7.26
ZINC04964111	0.564	6.25
ZINC04964113	0.564	6.30
ZINC08715601	0.675	6.23
ZINC11616454	0.579	6.49
ZINC12958408	0.583	6.38
ZINC13454269	0.665	6.53
ZINC13547650	0.556	6.91
ZINC13597424	0.64	6.37
ZINC13829362	0.554	6.188
ZINC16382910	0.579	6.492
ZINC16923286	0.584	6.36
ZINC16951688	0.569	6.58
ZINC16951904	0.753	6.21
ZINC16952556	0.564	7.79
ZINC16953141	0.577	6.44
ZINC16953833	0.585	6.26
ZINC16970549	0.64	6.70
ZINC16990186	0.721	6.23
ZINC16990189	0.723	6.22
ZINC16990363	0.756	6.45
ZINC16990703	0.737	6.63
ZINC17020182	0.564	6.71
ZINC17021040	0.596	6.44
ZINC20150163	0.637	7.87
ZINC20150164	0.637	7.10
ZINC44069525	0.583	6.80
ZINC95921563	0.554	6.33
ZINC98091587	0.554	6.18
ZINC98091589	0.554	6.18

One of the essential issues that should be considered in the field of drug design is the analysis of pharmacokinetic properties such as absorption, distribution, metabolism, excretion, and toxicity of new drug-like compounds before entering the synthesis stage. Since this process is a multi-step and very time-consuming and costly process, and the existence of favorable interactions between the ligand and the target molecule is not a guarantee that the ligand is drug-like, so virtual evaluation and prediction of the pharmacokinetics feature for designed compounds can be extremely useful. Therefore, in this study, the total level of polarity and the Lipinski's rule of five (i.e., a molecule with a molecular mass < 500 Da, hydrogen bond donors < 5, hydrogen bond acceptors < 10, and a log P (octanol-water partition coefficient) < 5) were used as the basis for analyzing the pharmacokinetic properties of candidate drug-like compounds (Table 2). Of the 42 compounds studied, 5 are given in Table 3 as the most optimal pharmacokinetic conditions according to Lipinski's rule of five, calculated by the SwissADME web tool (http://www.swissadme.ch). From a pharmacokinetic point of view, since a total level of polarity greater than 140 is not desirable, among the 5 final candidates listed in Table 3, the first combination (ZINC ID: ZINC16951904) is introduced as the final candidate to enter the synthesis stage.

Table 3. Five Compounds with the Most Optimal Pharmacokinetic Conditions According to Lipinski's Rule of Five

ZINC ID		Molecular Weight	Hydrogen Bond Acceptors	Hydrogen Bond Donors	Total Level Of Polarity	Lipophilicity
ZINC16951904		298.32	7	3	138.82	0.09
ZINC16990186		355.41	7	4	164.84	0.76
ZINC16990189		403.46	7	4	164.84	1.47
ZINC16990363		341.39	7	4	164.84	0.31

4. Conclusion

The main goal of pharmaceutical chemists is to find new compounds showing desirable performance, selectivity, and pharmacokinetic properties for a specific target. The aim of this study was to propose a 3D QSAR model, constructed by alignment independent descriptors termed GRINDs, to predict the inhibitory activity of new compounds against the adenosine kinase of Toxoplasma gondii, and to find new potent inhibitors. Since GRINDs-based 3D QSAR models provide chemically interpretable data, they were applied for QSAR modeling. New ligands were then suggested based on virtual screening. SwissSimilarity and ZINC databases were used as web tools and data banks for virtual screening. The pharmacokinetic properties of drug-like molecules are major contributors to the design and development of new drugs. Therefore, compounds selected by virtual screening were subjected to the SwissADME web tool for pharmacokinetic properties calculation. Finally, based on pharmacokinetic properties, a leading compound was proposed as a potent inhibitor of adenosine kinase of Toxoplasma gondii.

References

Al Safarjalani, O. N., Rais, R. H., Kim, Y. A., Chu, C. K., Naguib, F. N., & el Kouni, M. H. (2008).
7-Deaza-6-benzylthioinosine analogues as subversive substrate of Toxoplasma gondii adenosine kinase: activities and selective toxicities. Biochemical Pharmacology, 76(8), 958-966.

Babaie, J., Sayyah, M., & Golkar, M. (2019). A systematic review and meta-analysis of the
relationship between toxoplasmosis and epilepsy. Iranian Journal of Physiology and Pharmacology, 2(4), 279-271.

Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., & Clementi, S. (1993).
Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. Quantitative Structure-Activity Relationships, 12(1), 9-20.

Brereton, R. G. (2003). Chemometrics: data analysis for the laboratory and chemical plant. John
Wiley & Sons.

Burger, A. (Ed.). (1970). Medicinal chemistry. Wiley-Interscience.

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., ... & Zardecki, C.
(2019). RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic acids research, 47(D1), D464-D474.

Chen, X., Liu, M., & Gilson, M. K. (2001). BindingDB: a web-accessible molecular recognition
database. Combinatorial Chemistry & High Throughput Screening, 4(8), 719-725.

Cook, W. J., Delucas, L. J., & Chattopadhyay, D. (2000). Crystal structure of adenosine kinase
from Toxoplasma gondii at 1.8 Å resolution. Protein Science, 9(4), 704-712.

Duran, A., Martínez, G. C., & Pastor, M. (2008). Development and validation of AMANDA, a
new algorithm for selecting highly relevant regions in molecular interaction fields. Journal of Chemical Information and Modeling, 48(9), 1813-1823.

el Kouni, M. H. (2007). Adenosine metabolism in Toxoplasma gondii: potential targets for
chemotherapy. Current Pharmaceutical Design, 13(6), 581-597.

Elmore, S. A., Jones, J. L., Conrad, P. A., Patton, S., Lindsay, D. S., & Dubey, J. P. (2010).
Toxoplasma gondii: epidemiology, feline clinical aspects, and prevention. Trends in Parasitology, 26(4), 190-196.

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., & Gramatica, P. (2003).
Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. Environmental Health Perspectives, 111(10), 1361-1375.

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., ... & Overington, J.
P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Research, 40(D1), D1100-D1107.

Göller, A. H., Hennemann, M., Keldenich, J., & Clark, T. (2006). In silico prediction of buffer
solubility based on quantum-mechanical and HQSAR-and topology-based descriptors. Journal of Chemical Information and Modeling, 46(2), 648-658.

Gramatica, P. (2007). Principles of QSAR models validation: internal and external. QSAR &
Combinatorial Science, 26(5), 694-701.

Hartenfeller, M., Eberle, M., Meier, P., Nieto-Oberhuber, C., Altmann, K. H., Schneider, G., ... &
Renner, S. (2012). Probing the bioactivity-relevant chemical space of robust reactions and common molecular building blocks. Journal of Chemical Information and Modeling, 52(5), 1167-1178.

Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., ... & Steinbeck, C.
(2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Research, 44(D1), D1214-D1219.

Hill, D., & Dubey, J. P. (2002). Toxoplasma gondii: transmission, diagnosis and prevention.
Clinical Microbiology and Infection, 8(10), 634-640.

Homan, W. L., Vercammen, M., De Braekeleer, J., & Verschueren, H. (2000). Identification of a

200-to 300-fold repetitive 529 bp DNA fragment in Toxoplasma gondii, and its use for diagnostic and quantitative PCR. International Journal for Parasitology, 30(1), 69-75.

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: a free tool to discover chemistry for biology. Journal of Chemical Information and Modeling, 52(7), 1757-1768.

Johns Hopkins Bloomberg School of Public Health. (2018, September 24). Cost of Clinical Trials For New Drug FDA Approval Are Fraction of Total Tab. Johns Hopkins Bloomberg School of Public Health. https://publichealth.jhu.edu/2018/cost-of-clinical-trials-for-new-drug-FDA-approval-are-fraction-of-total-tab.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. Technometrics, 11(1), 137-148.

Kim, Y. A., Sharon, A., Chu, C. K., Rais, R. H., Al Safarjalani, O. N., Naguib, F. N., & el Kouni, M. H. (2008). Structure− activity relationships of 7-deaza-6-benzylthioinosine analogues as ligands of Toxoplasma gondii adenosine kinase. Journal of Medicinal Chemistry, 51(13), 3934-3945.

Leardi, R. (2000). Application of genetic algorithm–PLS for feature selection in spectral data sets. Journal of Chemometrics, 14(5-6), 643-655.

LigandExpo. Available online: http://ligand-expo.rcsb.org (accessed on 15 August 2019).

Lin, M. H., Chen, T. C., Kuo, T. T., Tseng, C. C., & Tseng, C. P. (2000). Real-time PCR for quantitative detection of Toxoplasma gondii. Journal of Clinical Microbiology, 38(11), 4121-4125.

Liu, T., Lin, Y., Wen, X., Jorissen, R. N., & Gilson, M. K. (2007). BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Research, 35(suppl_1), D198-D201.

Ngô, H. M., Ngo, E. O., Bzik, D. J., & Joiner, K. A. (2000). Toxoplasma gondii: are host cell adenosine nucleotides a direct source for purine salvage?. Experimental parasitology, 95(2), 148-153.

Nicolle, C. (1908). Sur une infection a corps de Leishman (on organismes voisons) du gondi. CR Acad Sci, 147, 736.

Reddy, M. C., Kuppan, G., Shetty, N. D., Owen, J. L., Ioerger, T. R., & Sacchettini, J. C. (2008). Crystal structures of Mycobacterium tuberculosis S-adenosyl-L-homocysteine hydrolase in ternary complex with substrate and inhibitors. Protein Science, 17(12), 2134-2144.

Singh, B., Hao, W., Wu, Z. C., Eigl, B., & Gupta, R. S. (1996). Cloning and characterization of cDNA for adenosine kinase from mammalian (Chinese hamster, mouse, human and rat) species: high frequency mutants of Chinese hamster ovary cells involve structural alterations in the gene. European Journal of Biochemistry, 241(2), 564-571.

Sterkers, Y., Ribot, J., Albaba, S., Issert, E., Bastien, P., & Pratlong, F. (2011). Diagnosis of congenital toxoplasmosis by polymerase chain reaction on neonatal peripheral blood. Diagnostic Microbiology and Infectious Disease, 71(2), 174-176.

Switaj, K., Master, A., Skrzypczak, M., & Zaborowski, P. (2005). Recent trends in molecular diagnostics for Toxoplasma gondii infections. Clinical Microbiology and Infection, 11(3), 170-176.

Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR & Combinatorial Science, 22(1), 69-77.

Vodnala, M., Fijolek, A., Rofougaran, R., Mosimann, M., Maser, P., & Hofer, A. (2008). Adenosine kinase mediates high affinity adenosine salvage in Trypanosoma brucei. Journal of Biological Chemistry, 283(9), 5380-5388.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... & Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074-D1082.

Wold, S., Jonsson, J., Sjörström, M., Sandberg, M., & Rännar, S. (1993). DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. Analytica Chimica Acta, 277(2), 239-253.

Zenner, L., Estaquier, J., Darcy, F., Maes, P., Capron, A., & CESBRON-DELAUW, M. F. (1999). Protective immunity in the rat model of congenital toxoplasmosis and the potential of excreted-secreted antigens as vaccine components. Parasite Immunology, 21(5), 261-272.

CONFLICT OF INTEREST

No conflict of interest declared.