



Subtractive Proteomics Supported with Rational Drug Design Approach Revealed ZINC23121280 as a Potent Lead Inhibitory Molecule for Multi-drug Resistant *Francisella tularensis*

Naima Javed[†], Sajjad Ahmad[†], Saad Raza, and Syed Sikander Azam^{*}

Computational Biology Lab, National Center for Bioinformatics,
Quaid-i-Azam University, Islamabad-45320, Pakistan

Abstract: *Francisella tularensis* is a Gram-negative bacterium and is the etiological agent of tularemia. The prolonged use of antibiotics is the reason for pathogen resistance to antibiotics such as beta-lactams and macrolides. This leads to the search to explore novel drug targets for *F. tularensis* to inhibit its growth. Subtractive proteomics revealed Glucose-1-phosphate thymidyltransferase (G1PTT) as the most promising protein as a drug target. A pharmacophore model was generated for virtual screening of a druglike library comprised of 1,000,000 drug molecules. Based on a pharmacophore-based search, a set of 152 compounds was predicted as the most potent inhibitors against this enzyme. The screened hits were docked with the target enzyme; which unveiled ZINC23121280 as the best-docked inhibitor having Autodock Vina binding energy of -7.2 kcal/mol and the GOLD score of 64.06. Moreover, the time-dependent dynamic behavior of the complex was analyzed using Molecular Dynamics (MD) simulation studies that revealed a stable system with a Root Mean Square Deviation (RMSD) average value of 2.25 Å and Root Mean Square Fluctuations (RMSF) of 1.16 Å. Radial Distribution Function (RDF) predicted strong hydrogen interactions between the ligand and Trp221 from the enzyme active pocket. The higher affinity of the antagonist for the enzyme was further supported by Molecular Mechanics Energies combined with the Poisson-Boltzmann and Surface Area (MMPBSA) and or Generalized Born and Surface Area (MMGBSA) with the estimated binding free energy of -1.07 kcal/mol and -29.59 kcal/mol, respectively. Findings from this present computational framework may provide the foundation for future drug discovery against *F. tularensis*.

Keywords: *Francisella tularensis*; Subtractive proteomics; Glucose-1-phosphate thymidyltransferase; Pharmacophore; Molecular docking; Molecular Dynamic simulation; MMPB/GBSA.

1. INTRODUCTION

Francisella tularensis is a pleomorphic Gram-negative coccobacilli and is the etiologic agent of a zoonotic disease of the northern hemisphere, tularemia [1, 2]. *F. tularensis* is highly virulent to a wide range of animals and humans [3]. The pathogen may cause epizootics or epidemics [4]. No human-to-human transmission is observed whereas transmission to humans occurs by four main routes: (i) through direct contact with the infected animals, infectious animal fluids or tissues, (ii) through arthropod bites, (iii) by inhaling infective aerosols, and (iv) by ingesting contaminated food or water [5, 6]. Tularemia is

highly prevalent in Sweden, Finland, and Turkey [7]. According to the pre-antibiotic era, the reported mortality rate of tularemia was 30%–60% [8]. The mortality rate associated with respiratory tularemia is as high as 5 – 30% [9]. According to Center for Disease Control and Prevention (CDC) statistics, *F. tularensis* is common in south-central America, and Massachusetts, and cases of Tularemia have been reported from all states of the USA except Hawaii. The number of cases from 1950 declined from 927 to 229 in 2018 (<https://www.cdc.gov/tularemia/statistics/index.html>). This bacterium is considered an aerosol biological weapon by several countries in the past [10]. Various subspecies include tularensis, holarctica, mediasiatica, and

Received: March 2020; Accepted: March 2021

*Corresponding Author: Syed Sikander Azam <ssazam@qau.edu.pk; syedazam2008@gmail.com>

[†]Both authors contributed equally to this work

novicida, where tularensis type A and holarctica type B is the most significant clinical subspecies of *F. tularensis* [11, 12]. Strains of the *F. tularensis* subspecies are common in North America where they cause rapidly progressive diseases [13] leading to prominent lymph node enlargement and flu-like symptoms [14]. The pathogen is also capable of infecting many types of eukaryotic cells and tissue macrophages [15]. With the increasing trend of antibiotic resistance to beta-lactams and macrolides in *F. tularensis* and the absence of a licensed vaccine for boosting the immune response to infections, identification of novel drug targets for designing novel antibiotics is an imperative need of time [4, 16].

The first phase of the drug designing process is the identification of potential drug targets against bacterial pathogens [17]. In traditional drug discovery, this process is time and resource-consuming and often results in failures [18]. On the other hand, using computational power and available genomic and proteomic data of bacterial pathogens is now common to discover new antibiotics, optimization, and development [19, 20]. In this context, subtractive proteomics is a widely used approach that in a step-wise process, filters proteins of high pathogen specificity [21]. The use of such in silico methodologies, not only saves extensive labor cost and time but also expedites the process of characterization of bacterial host's non-homologous and essential proteins to eradicate the disease with fewer side effects [22]. In the present study, a subtractive proteomics approach was employed in combination with the applications of computer-aided drug designing (CAAD) for the identification of potential drug candidates for the potential druggable protein against *F. tularensis* reference strain SCHU4 [23, 24]. The pathogen is investigated for host non-homologous proteins followed by essential proteins mapping using the Database of Essential Genes (DEG) [25]. The Glucose-1-phosphate thymidyltransferase (G1PTT) is a target of choice for novel antibacterial discovery. The G1PTT is the first enzyme in the dTDP-L-rhamnose biosynthesis pathway that acts as an L-rhamnose precursor. L-rhamnose is an important component of bacterial surface antigens such as the O-lipopolysaccharide [26]. In addition, it aids in mediating pathogen adhesion to host tissues and virulence [26]. The best-docked

complex was simulated to unveil the enzyme dynamics in the presence of ligand [27]. To further explore the ligand affinity towards the enzyme active site, binding free energies were estimated using Molecular Mechanics Energies combined with the Poisson-Boltzmann or Generalized Born and Surface Area Continuum Solvation (MMPB/GBSA) [28].

2. MATERIALS AND METHODS

The workflow for characterizing potential drug targets in *F. tularensis* proteome and subsequent steps of pharmacophore generation, molecular docking, MD simulations, and binding free energies is illustrated in Fig. 1.

2.1 Subtractive Proteomics

The Uniprot database [29] was used to retrieve the complete proteome of the reference strain of *F. tularensis* labeled as SCHU S4 [30]. The proteome was subjected to the subtractive proteomics where proteins relevant to antibiotics design were extracted in a step-wise manner [31]. In the first step, the CD-HIT suite [32] was applied to eliminate the redundant proteins sharing the identity of 80%. Redundant proteins are paralogous proteins that arise because of duplication during evolution and are not conserved across bacterial species. Non-redundant proteins, in contrast, are orthologous and are well conserved across bacterial species and strains; thus can be targeted for the design of broad-spectrum inhibitors [23, 33]. The BLASTp search of the National Center for Biotechnological Information (NCBI) was performed against reference human proteome (*Homo sapiens*: Tax id. 9606) to obtain proteins specific to the pathogen with a percentage identity threshold $\geq 30\%$ and the Expectation value (E-value) of 10^{-5} . The host non-homologous proteins were then used in BLASTp of DEG [25] with the threshold E-value of 10^{-10} , sequence identity of $\geq 30\%$, and bit score of 100 to predict pathogen essential proteins. The identified essential proteins were then allowed to enter the metabolic pathway mapping stage, where the proteins were mapped to the metabolic pathways of the pathogen [34]. In order to predict protein sequences involved in different metabolic pathways of the organism, the KEGG Automatic Annotation server (KAAS) [35] was used. Further in the framework, virulence

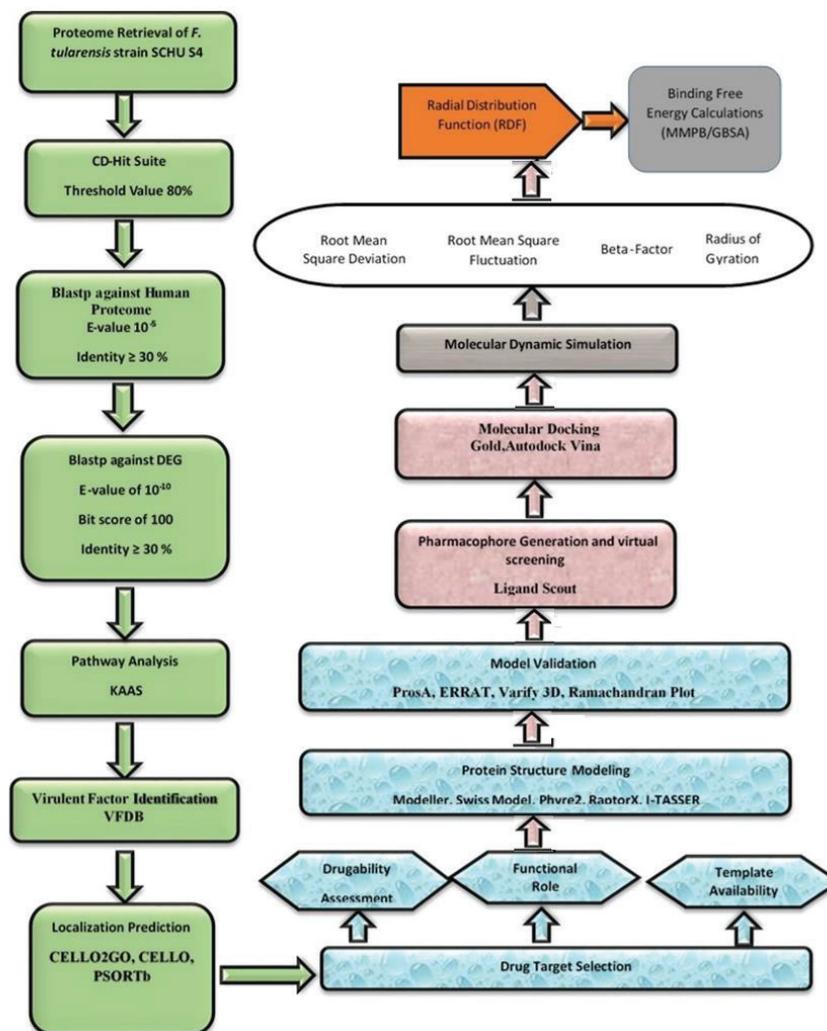


Fig. 1. Workflow of the methodology used in the present study

proteins of the bacteria were screened as they aid in bacteria adherence, colonization, invasion, evasion of host defense, and disease etiology as such are attractive targets to deactivate the pathogen [36]. For virulent protein identification, the Virulent Factor Database (VFDB) was accessed [37]. The unique proteins from *F. tularensis* metabolic pathways were subjected to BLASTp of VFDB to screen proteins having a threshold bit score ≥ 100 and identity \geq of 50%. In the concluding step of the subtractive proteomics, cytoplasmic proteins were identified using a comparative subcellular localization prediction approach. In this approach, the cellular localization of selected proteins was determined using three online servers: PSORTb [38] CELLO [39], and CELLO2GO [40]. The cytoplasmic proteins are presumed to be attractive targets because of the higher availability of drugs [41].

2.2 Physicochemical Characterization of Cytoplasmic Proteins

The physicochemical properties of the cytoplasmic proteins were unraveled to evaluate several vital parameters of the targets important from an experimental validations point of view [42]. The characterization was done based on molecular weight, theoretical pI, atomic composition, instability index, and grand average of hydropathicity (GRAVY). For this, an online server of ProtParam [43] was used.

2.3 Drug Target Selection

The selection of a drug target was done based on their subcellular localization, virulence, and physicochemical properties of the protein [16, 17]. Another parameter for the target selection was

the non-availability of experimentally determined structure. The Protein Data Bank (PDB) was explored for the availability of protein experimental structures. Enzymes in the shortlisted targets were especially targeted because of the following reasons: (i) Enzymes are essential to life, (ii) dysregulated enzymes lead to disease states, (iii) Enzymes are highly amenable to inhibition by small druglike molecules, (iv) Majority of the pharmaceutical companies (50% - 75%) around the globe focused on enzyme as primary target [44].

2.4 Comparative Structure Modeling and Validation

The availability of 3D structures of the targeted protein was checked using an online BLASTp tool of NCBI against PDB. The 3D structure is a prerequisite of the subsequent molecular docking study, MD simulation, and binding free energy calculations. In the absence of 3D structure, the sequence of proteins was used in a comparative structure modeling approach. First, in this approach, the amino acid sequence of the selected proteins was BLASTp against PDB for the identification of the appropriate template structure. Once template structure was identified, an automated protein modeling program, Modeller 9.14, was run to predict the most optimal 3D structure of proteins. Parallel to Modeller, several online servers: ReceptorX [45], Phyre2 [46], SWISS-MODEL [47] and I-TASSER [48] were used. To check the thermodynamic stability of the generated models, online structure quality assessment tools: ERRAT, Verify3D [49], ProSA [50], and Ramachandran Plot [51] were utilized. The protein was then energetically minimized using UCSF Chimera [52] to improve the quality of the structure by removing steric clashes. Minimization was performed for 1500 steps which can be split into 750 conjugate gradient steps and 750 steepest descent steps under Tripos Force Field [53].

2.5 Pharmacophore Model Generation and Virtual Screening

For ligand-based pharmacophore model generation, 15 compounds against the target protein were retrieved from the extensive literature reviews and binding database as illustrated in Table S1 [38]. The pharmacophore model was generated using

LigandScout 4.5. Pharmacophoric sites: aromatic ring, hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), positive and negative ionizable groups, and hydrophobic sites were characterized carefully. To incorporate associated features of the utilized compounds, merge feature model generation and atom overlapping scoring function of the LigandScout was employed. The pharmacophore model with the best score was selected and used in virtual screening of Zinc database druglike libraries containing 1 million compounds. Once the screening was done, Lipinski's rule of five filters was additionally applied to the shortlisted inhibitors for filtering drug molecules with properties: HBA < 10, HBD < 5, molecular weight < 500 Da, and the value of log P < 5 [54].

2.6 Molecular Docking

The binding site in the selected enzyme for the screened set of inhibitors was predicted using a combined approach of online binding site prediction tools and Multiple Sequence Alignment (MSA) [55]. Meta pocket [56] was used first to predict the enzyme active cavity, followed by aligning the orthologues of GIPTT enzyme through ClustalW [57] to look for the most conserved active site pocket residue. Genetic Optimization for Ligand Docking (GOLD) [58] and AutoDock Vina (AD-Vina) [59] were used for docking of the compounds to enzyme active pockets. The coordinates of the oxygen atom from Gly09 were set as the point of inhibitors binding. The GOLD docking was accomplished with a genetic run for each compound was set to 10. In AD-Vina, the same active site coordinates used in GOLD were used with the grid box size set to 15 Å along with the X, Y, and Z-axis. The docking results were visualized using LIGPLOT [60], UCSF Chimera [60], Visual Molecular Dynamics (VMD) [61], and Discovery Studio (DS).

2.7 MD Simulation

To determine the dynamic behavior of the enzyme in complex with the ligand, MD simulation for 100-ns was carried out [62]. Assisted model building with the energy refinement 14 (AMBER 14) [63] was used to design and perform simulation protocol. Initial libraries of the complex were generated using the Antechamber program. The docked complex was integrated into a TIP3P water

box with a padding distance size of 12 Å between the protein and water box boundary conditions. This was accomplished using ff14SB force field using Leap program [64]. The addition of 12 Na⁺ ions was involved in neutralizing the hydrated complex. During minimization, hydrogen atoms, water box, carbon alpha atoms of the complex were minimized for 1000 cycles whereas non-heavy atoms were relaxed for 300 steps [65]. After that, the system was subjected to a heating step, where a temperature of 300 K for 20-ps with the restraint of 5 kcal/mol. on alpha-carbon atoms are used. SHAKE algorithm [66] was used to heat the system while Langevin dynamics was used for maintaining system temperature. The system equilibrium was achieved for 100-ps with a time scale of 2-fs [67]. The NPT ensemble was employed for 50-ps to maintain system pressure. The system was then equilibrated for 1-ns. In production phase, the Berendsen algorithm combined with NVT ensemble was used with cut-off value of non-bonded interactions set to 0.8 Å. The production run was carried out for 100-ns. For simulation trajectories analysis, CPPTRAJ program of AMBER was used [68].

2.8 Binding Free Energy Calculations

The MM-PBSA and MM-GBSA of AMBER14 were used to estimate the binding free energies of the system [62]. A total of 500 frames were extracted from the simulation trajectories. The prmtop files of receptor, ligand, and complex were generated using the anti-MMPBSA.py module of AMBER whereas for estimating binding free energy MMPBSA.py module was used.

3. RESULTS AND DISCUSSION

3.1 Subtractive Proteomics

The emergence of resistance in bacteria is rapid and occurring worldwide thus endangering the efficacy of life-saving antibiotics [69]. This resistance to antibiotics has been attributed to misuse and overuse of these medications, in addition, to the lack of new antibiotic development by pharma industries due to challenging regulatory requirements and lesser economic incentives [70]. Because of many limitations of conventional drug target identification and drug discovery, computational identification of potentially druggable proteins and subsequent

drug designing played a major role in providing therapeutically important molecules against several medical complications [71]. The drug target identification is the first step in drug discovery and can be applied to a range of biological entities that may include protein, DNA, and RNA [72] associated with the disease. Subtractive proteomics is now a widely used approach for the identification and validation of bacterial proteins involved in regulating essential biological processes [73]. This is a step-wise approach that gradually reduces the number of proteins involved in the host's non-homologous, essential and unique pathways of the bacterial pathogen [31]. Using this approach, the complete proteome of *F. tularensis* strain SCHU S4 that contains 1556 proteins were thoroughly screened first for non-redundant proteome. The non-redundant proteins are attractive targets for antibiotics because of their broad-spectrum conservation across bacterial species and strains [31]. On the other hand, redundant proteins are paralogous that are less conserved and not preferred as drug targets [31]. The CD-HIT analysis revealed 28 duplicated proteins and thus excluded them from *F. tularensis* proteome. The 1528 orthologous proteins were forwarded to the homology check. At this check, homologous proteins between the host (*Homo sapiens*) and bacteria were compared using an online BLASTp tool of NCBI. The homologous proteins sharing 30% of identity were discarded whereas those having identity of <30% were passed to the next step of DEG analysis. A homology check revealed 1218 proteins as host non-homologous proteins as a potential target for drug discovery. Screening host non-homologs is vital as targeting host homologous proteins could lead to autoimmune reactions and adverse side effects [73]. The essential proteins are the foundation to bacterial life without such proteins the bacterium is unable to survive, and as such are attractive targets for designing novel antibiotics [73]. The essentiality analysis unraveled 732 proteins in number while the remaining 486 proteins are non-essential hence excluded from further analysis [73]. Mapping essential proteins to organism metabolic pathways are vital as it provides an array of opportunities to block pathogen survival. The KAAS mapped a total of 262 proteins to metabolic pathways including 11 unique and 251 common proteins. The output of KASS was investigated first enzymatic and non-enzymatic proteins. The enzymatic proteins

were recognized through its Enzyme Classification (EC) number and were 71% of the total proteins compared to the non-enzymatic proteins that were 29%. The metabolic proteins were categorized into the following: 1) cellular process, 2) environment information processing pathways, 3) metabolism 4) human diseases and drug development and 5) organismal systems. Cellular processes can be divided into two main systems including peroxisome cell-cycle –caulobacter. Genetic information processing includes proteins that take part in folding, translation, transcription, sorting and degradation, repair, and replication. Environmental information processing include transport systems including phosphotransferase system (PTS), ABC transporters, two-component system and bacterial secretion system. The virulent proteins were identified through VFDB. Six proteins were found virulent including *wcaJ*, *oppF*, *qseC*, *phnA*, *wbtL* and *tolQ* (Table 1). The *WcaJ* is considered as initiating enzyme in the synthesis of colonic acid (CA) [74]. *OppF* plays part in transporting oligonucleotides. *QseC* is a member of two-component regulatory system (*QseB/QseC*) and functions by activating the flagella regulon of *FlhDC* [75]. The *PhnA* is involved in hydrolysis of phosphonoacetate [76]. The *WbtL* drives formation of dTDP-glucose from glucose 1-phosphate and dTTP [26]. The *TolQ* is a part of *Tol-Pal* system which plays a role in outer membrane invagination during cell division and is important for maintaining outer membrane integrity

[77]. Knowledge of subcellular localization is significant in identification of therapeutic targets. Cytoplasmic proteins are preferred as drug targets compared to membrane proteins because of the following reasons: (i) membrane proteins have low permeation rate thus can block the access of drugs to the biological target, (ii) the presence of energy driven efflux systems may use the drug as effluxing substrate for broad specificity. The comparative subcellular localization for the virulent proteins is illustrated in Fig. 2. According to PSORTb, majority of the proteins were found in the cytoplasmic membrane (73%), followed by cytoplasmic

Table 1. Virulent proteins screened in the study.

Gene	Protein Name	Bit score	Identity
<i>wcaJ</i>	Hydrogen peroxide-inducible genes activator	226	65
<i>oppF</i>	Oligopeptide transport ATP-	157	56
<i>qseC</i>	Sensor histidine Protein PhnA	107	31
<i>phnA</i>	Protein PhnA	32	3
<i>wbtL</i>	Glucose-1-phosphate thymidyltransferase	381	64
<i>tolQ</i>	Biopolymer transport protein TolQ	86	43

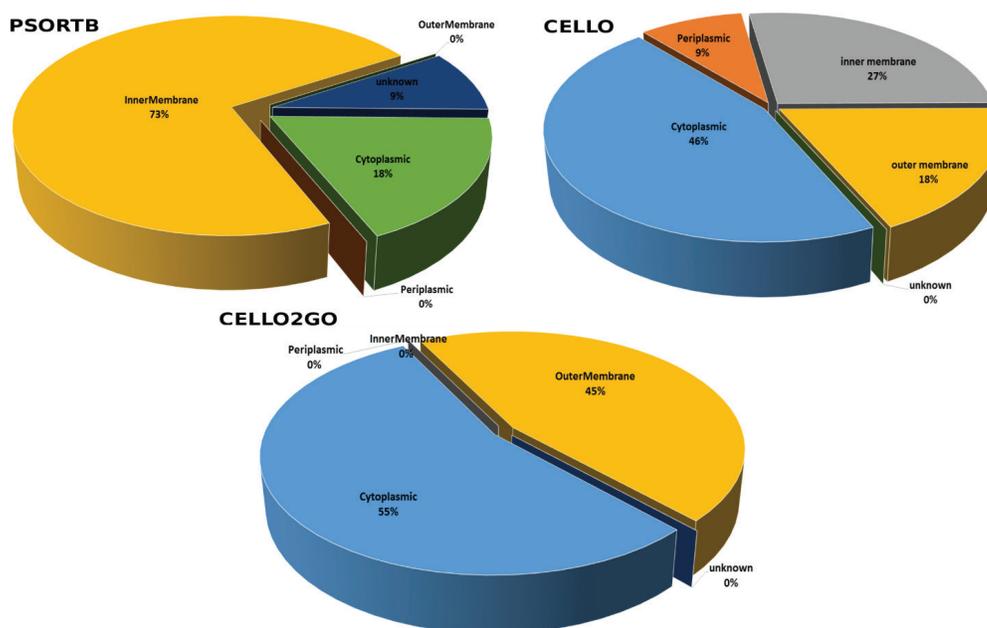


Fig. 2. Sub-Cellular localization of *F. tularensis* virulent proteins.

proteins (18%), 9% of unknown proteins, 0% of outer membrane, inner membrane and periplasmic proteins. Cello demonstrated majority of the proteins as cytoplasmic (46%), inner membrane (27%), outer membrane (18%), periplasmic (9%), cytoplasmic membrane and unknown (0%). Lastly, Cello2Go revealed 55% cytoplasmic, 45% outer membrane, 0% of cytoplasmic membrane, periplasmic, and unknown proteins. It was found that around 5 proteins were cytoplasmic, 1 was periplasmic, 3 in inner membrane and 2 were in outer membrane. By comparative analysis, only 3 proteins: *wcaJ*, *qseC*, and *wbtL* were selected as cytoplasmic proteins and forwarded along the framework.

3.2 Physicochemical Characterization

The physicochemical characterization of shortlisted proteins was an important consideration for shedding light on the suitability of selected proteins for a wet lab analysis [42]. The major parameter during this analysis was to compute the molecular weight of proteins. The proteins having <110 kDa molecular weight are preferred as drug and vaccine targets because of their easy purification [42, 53]. It was estimated that all four proteins have a molecular weight less than the threshold and can be used in the development of the novel drug. The stability of the protein is the next most important characteristic of protein. The proteins having an estimated <40 stability index are considered stable while those proteins with >40 are regarded as unstable. For all three proteins, stability were calculated <40

while no protein stability value was > 40. The GRAVY negative score of all the stable proteins indicates the hydrophilic nature of the proteins. The theoretical pI value of 2 proteins (*wcaJ*, *qseC*) were greater than 7 represented the basic nature of the protein and the 1 protein (*wbtL*) with less than 7 pI value indicating the acidic nature of drug-protein. The aliphatic index value for the proteins ranged from 117.9 - 95 show the high thermal stability of proteins. The physicochemical parameters for the shortlisted 3 proteins are tabulated in Table 2.

3.3 Selection of Drug Target

The G1PTT enzyme was selected as the potential drug target against the pathogen based on its cytoplasmic location, involvement in virulent pathways, suitable molecular weight, theoretical PI, instability index, and GRAVY index. Another parameter for the target selection was the non-availability of experimental structure and availability of a suitable template. The G1PTT is involved in the biosynthesis of different antibiotics, polyketide sugar unit biosynthesis, and acarbose and validamycin biosynthesis pathways. As the target is an enzyme, involved pathogen-specific and selective pathways, structure modeling, molecular docking, and dynamics simulation can provide an excellent platform for designing antibiotics against this target enzyme.

3.4 Comparative Structure Modelling

The 3D structure of the selected protein was not

Table 2. Physicochemical characterization of selected cytoplasmic proteins.

Uniprot ID	Gene	Protein Name	Residues length	molecular weight (KDa)	Theoretical pI	Instability Index	Aliphatic index	GRAVY
Q5NEZ2	<i>wcaJ</i>	Hydrogen peroxide-inducible genes activator	205	23.3	9.4	33.1	95.5	-0.073
Q5NIH6	<i>qseC</i>	Sensor histidine kinase QseC	4745	54.7	7.1	31.8	105.0	-0.22
Q5NF04	<i>wbtL</i>	Glucose-1-phosphate thymidyltransferase (G1PTT)	294	32.4	5.5	39.1	104.1	-0.052

present in the PDB, therefore, a comparative structure prediction approach was used. The X-ray crystallographic structure of a template ‘‘PDB id: 1H5T’’ was selected for the model building process with 98% query coverage and 64% sequence identity. For the selection of the best model, a detailed comparison of stereochemical properties was performed as tabulated in Table 3. Based on quality assessment measurements, Phyre2 predicted structure was selected having 97.9% residues in the favored region of the Ramachandran plot. The number of residues in the allowed and outlier region was 1.7% and 0.3%, respectively. ERRAT, Verify3D, and Z-score were 93.116, 93.10%, and -8.41, respectively thus further confirmed the reliability of the optimal model selected for antibiotics screening. Moreover, the superimposed structure of the template and the target unraveled Root Mean Square Deviation (RMSD) of 0.001 Å is in a highly acceptable range and shown in Fig. 3. The tertiary structure of G1PTT can be found in Fig. 4.

3.5 Pharmacophore Model Generation and Virtual Screening

A pharmacophore model was generated to shortlist druglike compounds from Zinc database druglike libraries that share molecular features necessary for recognizing a ligand by druggable biological macromolecules. Pharmacophore model-based virtual screening of 1,000,000 drug molecules was then performed to shortlist the best possible drug molecules. The screening filtered 152 compounds. Structures of these compounds are tabulated in Table S2. These compounds were energetically minimized using the MMFF94 force field and

further utilized in molecular docking studies.

3.6 Molecular Docking

The sequence 8GGSGTR13 was found conserved in all orthologues of the G1PTT enzyme. The coordinates oxygen atom from Gly9 was selected for molecular docking studies. Comparative docking performed using two different tools: GOLD and AD-Vina. In molecular docking, structure-based virtual screening of 152 drug-like compounds extracted based on pharmacophore-based virtual screening. All the inhibitors were docked into the enzyme active site using GOLD, and AD-Vina. The top ten best inhibitors based on descending order of GOLD fitness score together with AD-Vina binding energy and drug-likeness are shown in Table 4. The correlation coefficients between GOLD fitness score and AD-Vina binding energy of the compounds can be found in Fig. 5. Compound ZINC23121280 ((3R)-N-(6-amino-1-benzyl-2,4-dioxo-pyrimidin-5-yl)-1-cyclopentyl-5-oxo-N-propyl-pyrrolidine-3-carbox) was selected as the best-docked inhibitor with GOLD fitness score and AD-Vina binding energy of 64.02 and -7.2 kcal/mol, respectively. The complex was selected based on several parameters including strong interactions between ligand and target protein, drug-likeness of the compound, and its pharmacokinetics. In both tools, the inhibitor was investigated to dock in the same position and interacts with almost the same residues of the active site (Fig. 6). Visual inspection of complexes from both GOLD and AD-Vina revealed inhibitor binding with active residues: Leu6, Ala7, Gly8, Gly9, Ser10, Gly11, Arg13, Lys23, Gln24, Pro83, Gly85, Leu86, Leu106, Gly107, Asp108, Glu194, and Gly225.

Table 3. Stereo-chemical properties of comparative homology modeled structure.

Structure Resources	Favored region	Allowed region	outlier region	Errat	Z-score	Verify-3D
Modeller 1	279 (96.5%)	9 (3.1%)	1 (0.3%)	83.039	-8.41	88.32%
Modeller 2	279 (96.5%)	6 (2.1%)	4 (1.4%)	84.099	-8.41	90.38%
Modeller 3	280 (96.9%)	8 (2.8%)	1 (0.3%)	83.746	-8.73	96.22%
Modeller 4	279 (96.5%)	8 (2.8%)	2 (0.7%)	80.565	-8.32	94.50%
Modeller 5	279 (96.5%)	9 (3.1%)	1 (0.3%)	84.452	-8.47	92.44%
Phyre 2	282 (97.9%)	5 (1.7%)	1 (0.3%)	93.116	-8.41	93.10%
Swiss-Model 1	1073 (93.1%)	55 (4.8%)	24 (2.1%)	96.791	-8.39	93.10%
Swiss-Model 2	220 (93.6%)	12 (5.1%)	3 (1.3%)	83.772	-8.04	99.16%
I-TASSER	271 (92.8%)	14 (4.8%)	7 (2.4%)	97.895	-8.75	98.64%
RaptorX	280 (95.9%)	10 (3.4%)	2 (0.7%)	80.42	-8.4	98.80%

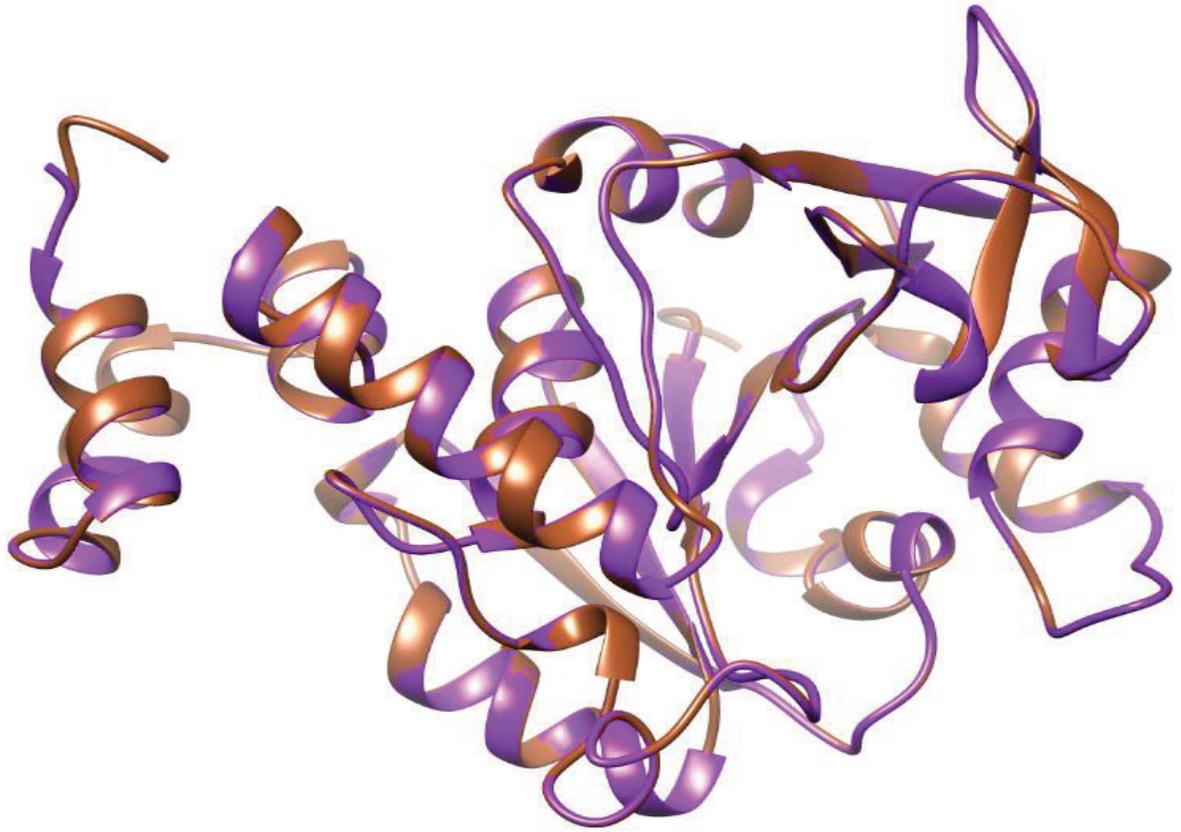


Fig. 3. Superimposition of selected optimum model (purple) over the template (sienna).

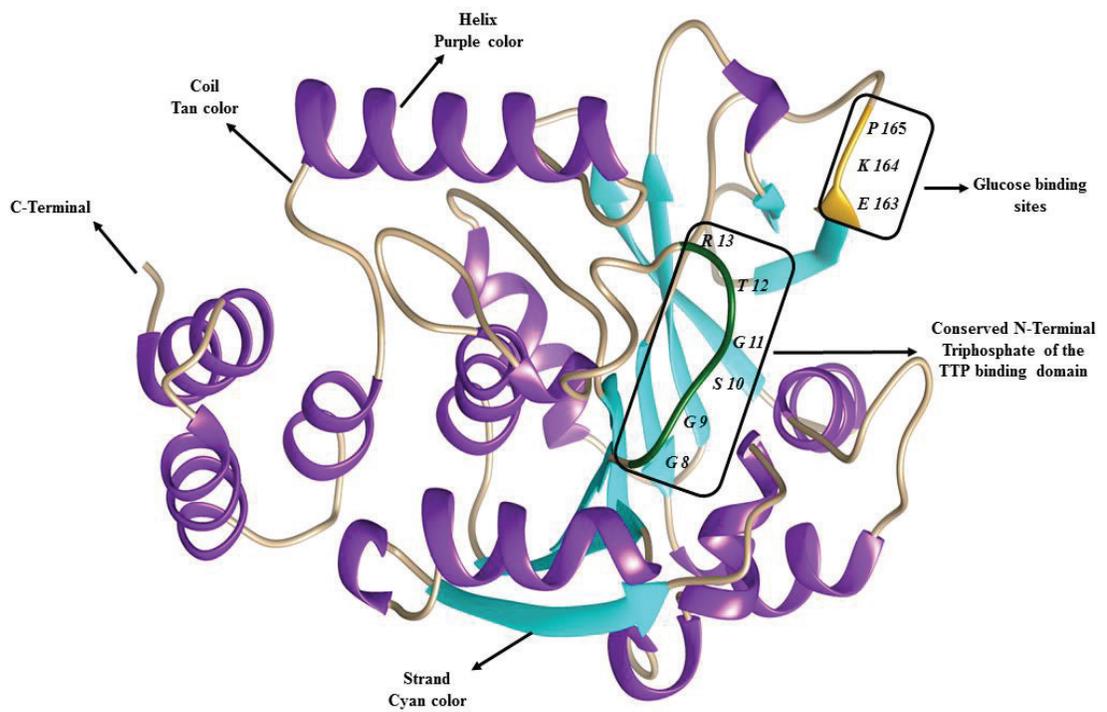


Fig. 4. Tertiary structure of GIPTT enzyme.

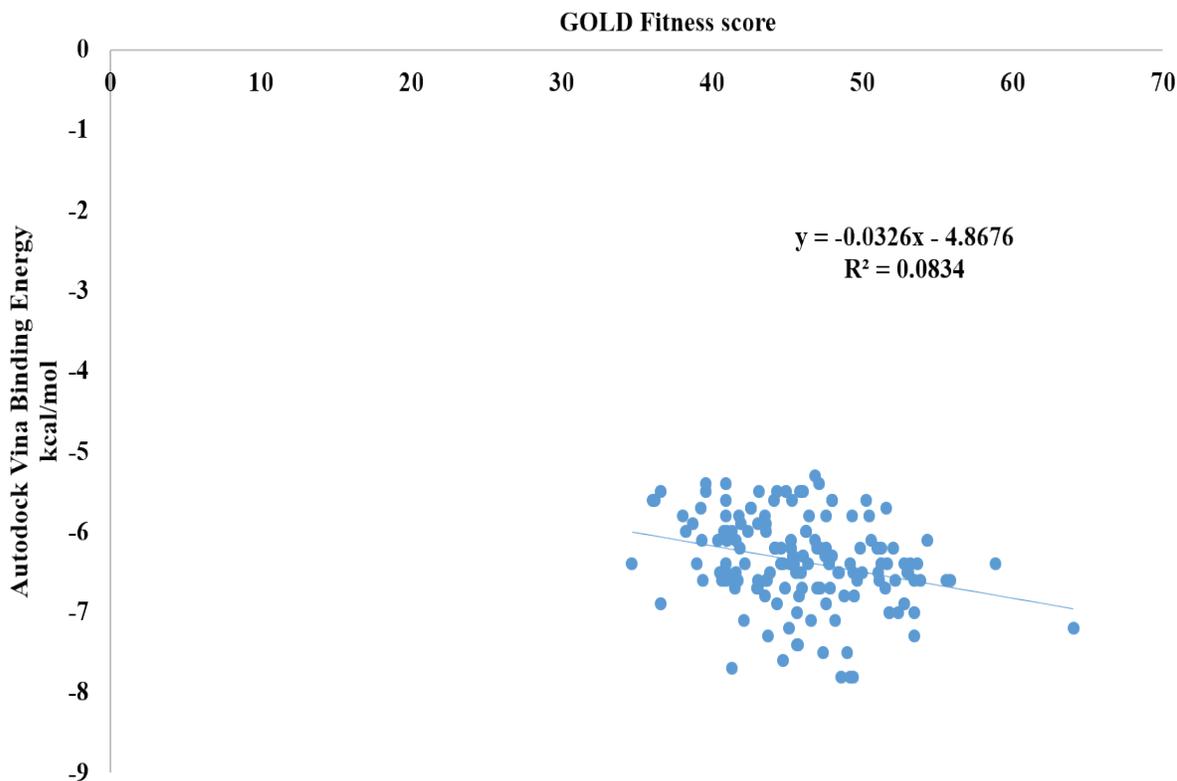


Fig. 5. Correlation coefficient between GOLD scores and binding free energies of 152 inhibitors.

The binding mode of the compound in the G1PTT pocket in GOLD was positioned as such to allow deep binding of 6-amino-1-(cyclohexylmethyl)-5-(ethyl(methyl)amino)-hexahydropyrimidine-2,4-diol ring and covered the major portion of the active site (Fig. 7). The oxygen atom of the ring was observed in hydrogen bond interaction with Gly85 and Pro83. The ring, (4R)-4-(hydroxymethyl) pyrrolidin-2-ol, bind in the cavity of the active site and the nitrogen atom of the ring was observed to interact hydrophilically with Gly09. It also forms interactions with Ser10, Leu106, Asp108, and Lys23.

At the drug design stage, unveiling drug-likeness and pharmacokinetic behavior of drugs are important as it reduces the number of unsuccessful hits in clinical trials [78]. In addition, it also enables chemists to select the most appropriate compounds for lead optimization and novel drug development. The compound completely follows Lipinski's rule of five: molecular weight (453.53 g/mol), number of H-bonds acceptors (4), number of H-bonds donors (2), topological polar surface area (TPSA) value (130.78 Å²), and LogP value (1.53). The

number of heavy atoms in the compound is 33, while aromatic heavy atoms and rotatable bonds are 12 and 8, respectively. The molar refractivity of the compound is 127.95. An important consideration of drugs is their lipophilicity, which describes the compound's ability to be dissolved in lipophilic solutions (non-aqueous). Lipophilicity determines the compound's ability to permeate across different biological membranes [79]. Higher LogP, higher the capacity of drugs to cross biological membranes and hence can access targets for inhibition. The efficient delivery of a drug to the target site depends on the tendency of drugs to retain in blood for an extended period [78]. This was disclosed using the plasma protein binding (PPB) feature of the compound (described as LogK). The different ADMET properties of the best 10 inhibitors screened in the study can be found in Table 5.

3.7 Molecular Dynamics Simulation

The MD Simulations for 100-ns of the enzyme complex were carried out to investigate system stability. There are different examples in which time-dependent MD simulations have been applied

Table 4. Docking scores of top ten docked inhibitors.

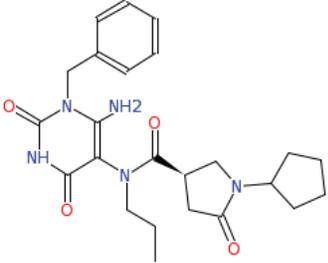
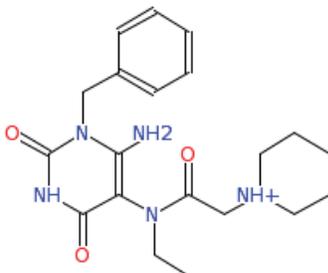
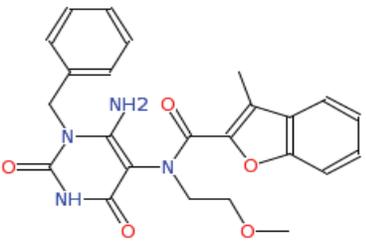
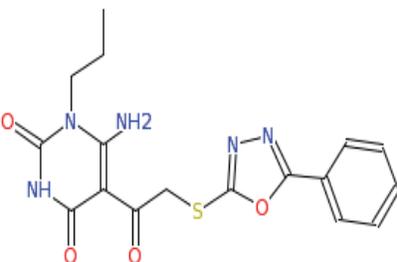
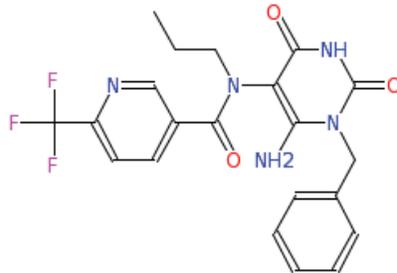
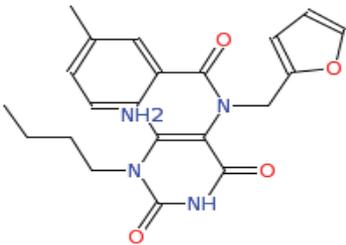
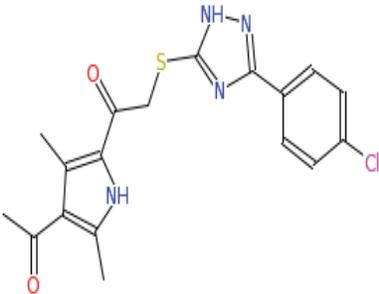
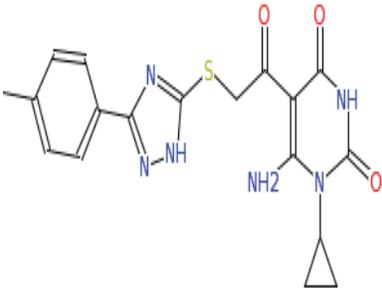
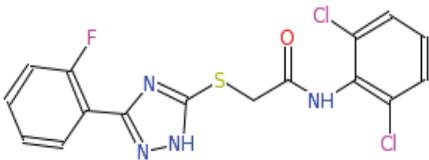
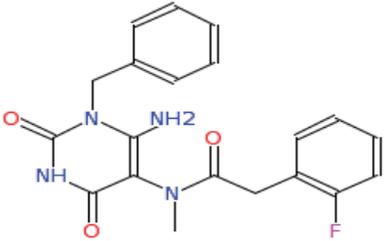
Compounds	GOLD Fitness Score	AD-Vina binding energy (kcal/mol)	Druglikenss rule violations
 <p>ZINC23121280</p>	64.06	-7.2	No violations
 <p>ZINC02629047</p>	58.94	-6.4	No violations
 <p>ZINC14511277</p>	55.86	-6.6	Egan rule (1 violations: TPSA>131.6)
 <p>ZINC03348170</p>	55.62	-6.6	Veber rule (1 violation: TPSA>140), Egan rule (1 violation: TPSA>131.6)
 <p>ZINC12763362</p>	54.31	-6.1	No violations

Table 4. Docking scores of top ten docked inhibitors.

Compounds	GOLD Fitness Score	AD-Vina binding energy (kcal/mol)	Druglikens rule violations
 <p>ZINC06221653</p>	53.85	-6.6	No violations
 <p>ZINC28807288</p>	53.69	-6.4	No violations
 <p>ZINC08343860</p>	53.46	-6.6	Veber rule (1 violations: TPSA>140), Egan rule (1 violations: TPSA>131.6), Muegge filter (1 violation: TPSA> 150)
 <p>ZINC05678255</p>	53.44	-7.3	No violations
 <p>ZINC13135410</p>	53.44	-7	No violations

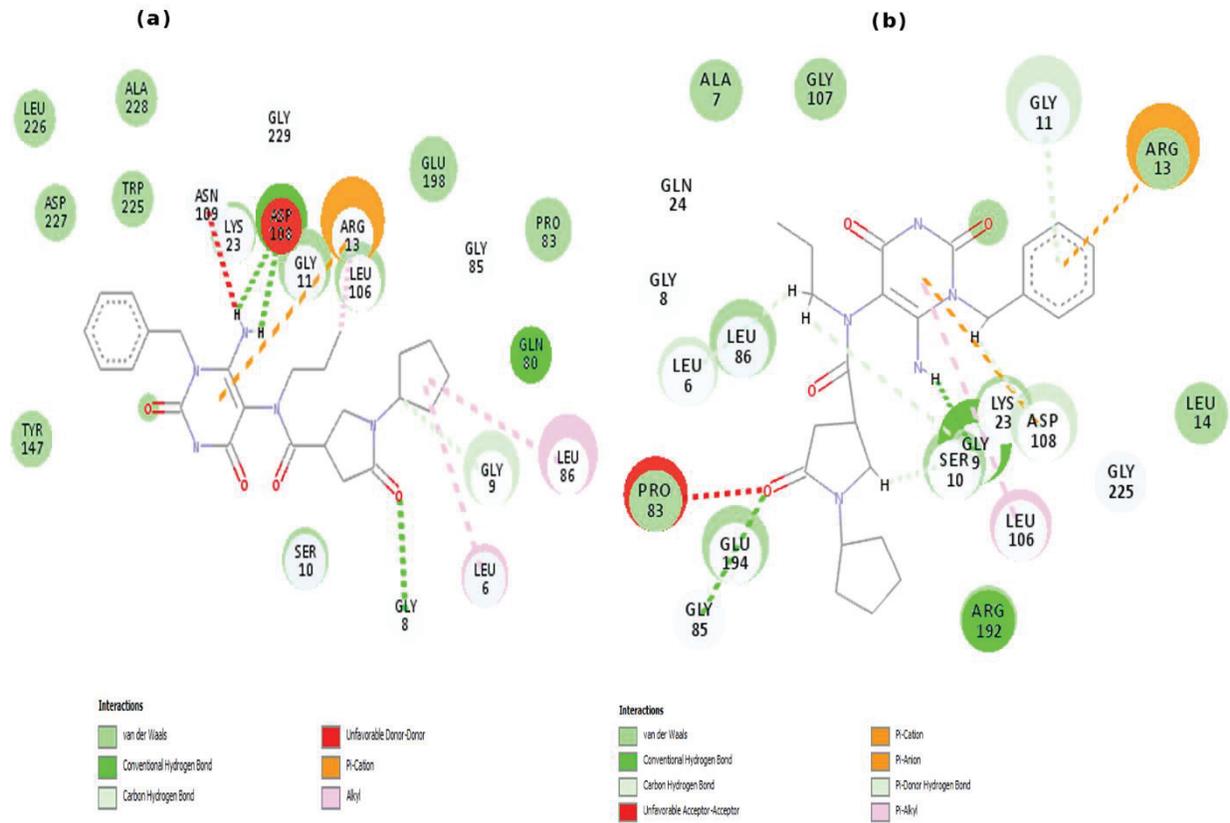


Fig. 6. Interacting residues of the enzyme with the ligand in AD-Vina (a) and GOLD (b).

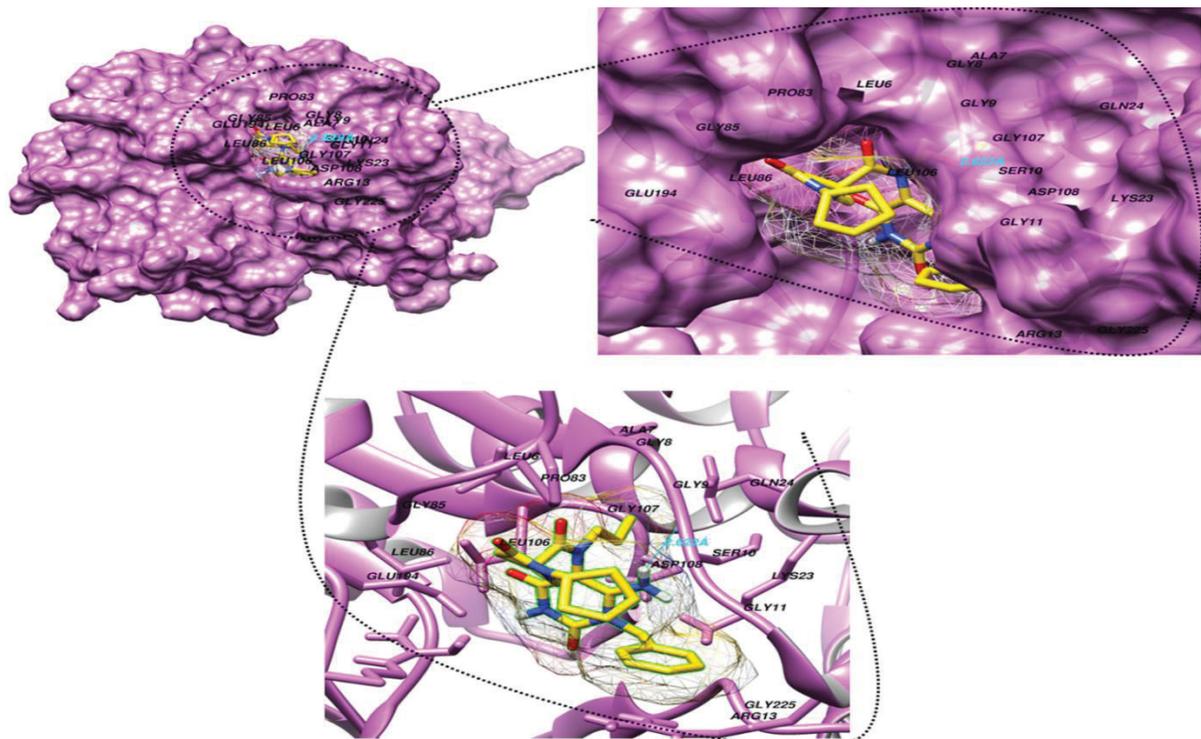


Fig. 7. Binding mode and interactions of ZINC23121280 in the binding pocket of the G1PTT enzyme.

Table 5. AdmetSAR properties of selected Compounds

Selected compounds	Blood-brain barrier (BBB) (probability)	Human intestinal absorption (HIA) (probability)	Caco2 permeability (probability)	CYP450 2D6 inhibitor (probability)	Carcinogens (probability)	Acute oral toxicity (probability)	Aqueous solubility (LogS)	Rat acute toxicity (LD50, mol/kg)	fish toxicity (pLC50, mg/L)	Ames toxicity (probability)	Honey bee toxicity (HBT) (probability)
ZINC 23121280	BBB-0.7	HIA+0.9	Caco2-0.7	Non-substrate 0.8301	Non-carcinogens 0.8	0.6	-2.6	2.5	1.9	Non-AMES toxic 0.7	Low HBT 0.8
ZINC 02629047	BBB-0.7	HIA+0.9	Caco2-0.6	Non-substrate 0.8090	Non-carcinogens 0.8	0.6	-2.4	2.5	2.2	Non-AMES toxic 0.6	Low HBT 0.8
ZINC 14511277	BBB-0.5	HIA+0.8	Caco2-0.7	Non-substrate 0.8273	Non-carcinogens 0.9	0.5	-2.5	2.6	1.6	Non-AMES toxic 0.6	Low HBT 0.6
ZINC 03348170	BBB+0.7	HIA+0.9	Caco2-0.6	Non-substrate 0.8322	Non-carcinogens 0.8	0.6	-3.1	2.6	1.3	Non-AMES toxic 0.5	Low HBT 0.8
ZINC 06221653	BBB+0.5	HIA+0.9	Caco2-0.6	Non-substrate 0.8	Non-carcinogens 0.9	0.6	-3.0	2.6	1.3	Non-AMES toxic 0.6	Low HBT 0.8
ZINC 28807288	BBB+0.9	HIA+1.0	Caco2-0.5	Non-substrate 0.8	Non-carcinogens 0.6	0.6	-3.7	2.4	1.2	Non-AMES toxic 0.5	Low HBT 0.7
ZINC 08343860	BBB+0.9	HIA+1.0	Caco2-0.5	Non-substrate 0.8	Non-carcinogens 0.7	0.6	-3.1	2.4	1.4	Non-AMES toxic 0.5	Low HBT 0.8
ZINC 05678255	BBB+0.9	HIA+1.0	Caco2-0.5	Non-substrate 0.8	Non-carcinogens 0.6	0.5	-3.6	2.7	1.2	Non-AMES toxic 0.6	Low HBT 0.8

Table 5. AdmetSAR properties of selected Compounds

Selected compounds	Blood-brain barrier (BBB) (probability)	Human intestinal absorption (HIA) (probability)	Caco2 permeability (probability)	CYP450 2D6 inhibitor (probability)	Carcinogens (probability)	Acute oral toxicity (probability)	Aqueous solubility (LogS)	Rat acute toxicity (LD50, mol/kg)	Fish toxicity (pLC50, mg/L)	Ames toxicity (probability)	Honey bee toxicity (HBT) (probability)
ZINC 13135410	BBB +	HIA+ 0.9	Caco2- 0.5	Non-substrate 0.8	Non-carcinogens 0.9	0.6	-3.0	2.4	1.6	Non-AME S toxic 0.6	Low HBT 0.9

on docked complexes to explore the protein-ligand interactions, conformational fluctuations, structural, architectural changes, and dynamical shifts of the proteins [80]. The MD simulations aid in understanding the dynamic behavior of the complex and also highlight the important residues playing a vital role in identifying and binding the ligand [81]. To shed light on biomolecular movements within a solvated environment, RMSD, RMSF, B-factor, and radius of gyration were plotted as a function of time (Fig. 8). Investigation of the enzyme in ligand-bounded form led to the assessment of structural minor structural variations and atomic level transitions [27]. The deviation of the backbone C α atoms was observed first for the complete production run. The average RMSD value calculated for the complex is 2.25 Å, with a maximum of 3.10 Å at 70th ns. No substantial structural movements were reported that elucidates complex stability. The average RMSF value for the complex was 1.16 Å. The regions illustrating higher fluctuation were loops: that involve the regular conversion of sheets into helix and helix into the sheet. The graph indicates that most of the residues of the active site have remained stable. The highest peak of the graph indicates the region in the loop region. The β -factor is a thermal disorderness calibrating function which stipulates the structural stability at the atomic position in term of RMSF. Therefore, its value depends on the level of atomic fluctuations which collectively contribute to the global vibrational movements of the protein and its thermal stability. The pattern of β -factor for protein is consistent with the RMSF trend. The β -factor average value calculated for the complex was 47 Å. To evaluate the structural

compactness, radius of gyration was calculated as a time function. The average value of 25.3 Å for the docked protein denotes the stability of the protein structure. Snapshots at different ns of the docked enzyme complex over simulation period of 100-ns is presented in Fig. 9.

3.8 RDF Analysis

The RDF is a key tool to describe the probability of the distance 'r' between two particles in a system [82]. The distribution of atoms, molecules, and species around a specific residue of targeted protein can be described by RDF. For this, the first vital residues of the enzyme involved in hydrogen bonding with the inhibitor toward the end of the simulation were identified using an in-house script in VMD. It was found that TRP221 is the main residue from the enzyme active pocket that contributes significantly to ligand binding. The RDF graphs were generated for all the three hydrogen interactions of TRP221 atoms: HE1 and HH2 as illustrated in Fig. 10. The highest distribution was observed between HE1 atom of TRP221 and ligand N atom, at a distance of 4.50 Å with a g(r) value of 0.16. The highest distribution of TRP221: HE1 and N1 atom of ligand was observed at 3.89 Å having a g(r) value of 0.20. The third plot describes the highest distribution at 3.39 Å with a g(r) value of 0.26 between the HH2 atom of TRP221 and the N4 atom of the ligand.

3.9 Binding Free Energy Calculations

MM_PBSA/GBSA methods of the AMBER14 were used to describe the binding free energy of the

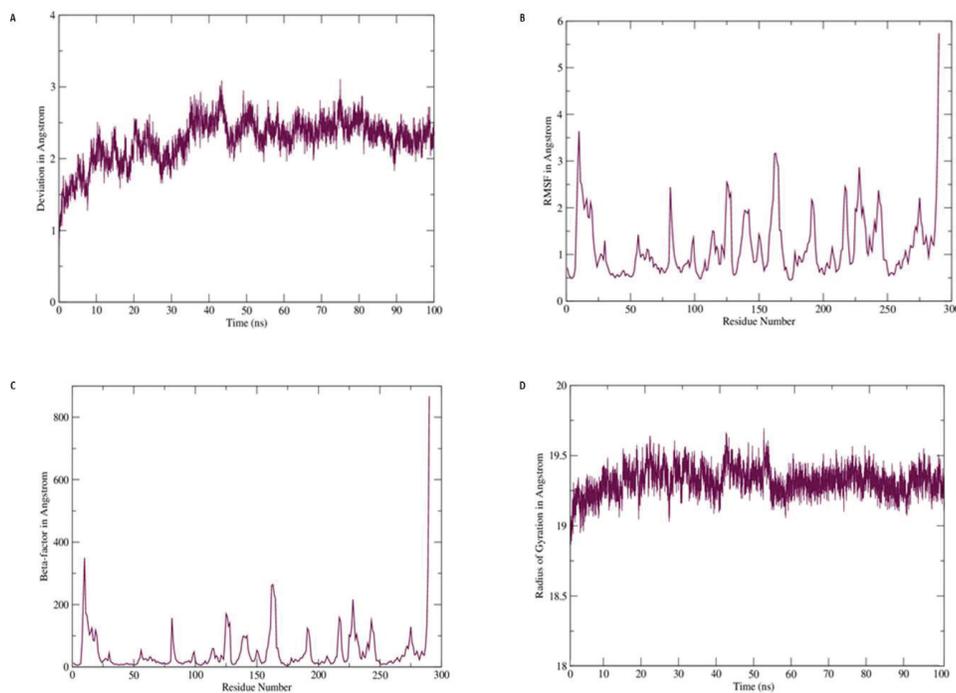


Fig. 8. Trajectories analysis.

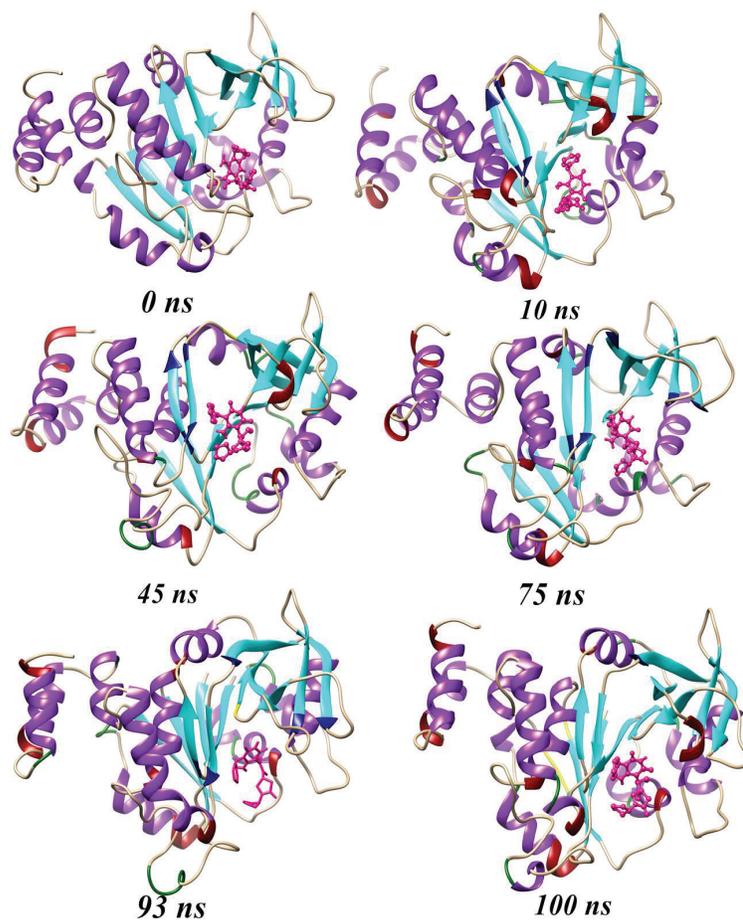


Fig. 9. Snapshots of structural variations of the docked complex at different ns. Dark red color indicates loop converted into the helix, the navy blue color indicates loop converted into the sheet, dark green indicates helix converted into a loop, and yellow color indicates the sheets convert into loops

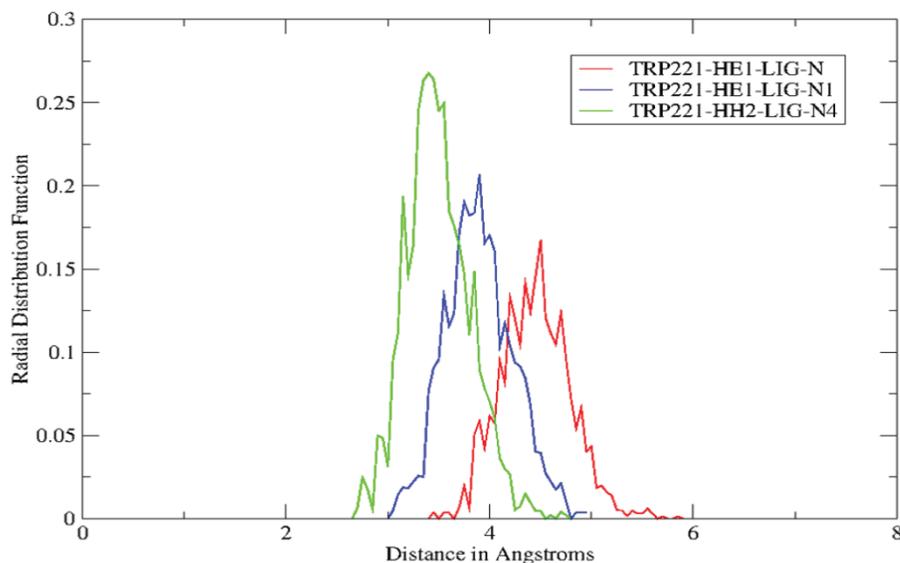


Fig. 10. RDF plots for G1PTT enzyme TRP221 atoms: HE1 and HH2.

system as well as molecular interactions between the protein and ligand. The MM_PBSA/GBSA technique combines the molecular mechanical energies with the continuum solvent approaches. The values of binding free energy were explained in Table 6. The entropy term is eliminated because of convergence problems in some cases, and it cannot be calculated.

The formation of the complex leads to highly favorable columbic interactions (-30.15 kcal/mol) as opposed to non-favorable contributions from the polar part of solvation free energy (44.79 kcal/mol in case of PB and 40.93 kcal/mol in

GB). The total electrostatic contribution is 10.41 kcal/mol in GB and 15.71 in PB calculations, respectively. The binding energy value for van der waal interactions is -35.93 kcal/mol, which depicts system stability. The total binding free energies in PB and GB were determined as -1.07 and -29.59 kcal/mol, respectively. The difference in values due to the solvation energy which was 66.11 kcal/mol in MMPBSA compared to 36.85 kcal/mol from MMGBSA. The binding energy of active site residues are as follow: Leu6 (-0.2 kcal/mol), Ala7 (-0.0 kcal/mol), Gly8 (-0.03 kcal/mol), Gly9 (-0.50 kcal/mol), Ser10 (-0.55 kcal/mol), Gly11 (-0.09 kcal/mol), Arg13 (-0.36 kcal/mol), Lys23 (-0.13 kcal/mol), Gln24 (-0.01 kcal/mol), Pro83 (-0.04 kcal/mol), Gly85 (-0.0 kcal/mol), Leu86 (-0.58 kcal/mol), Leu106 (-1.22 kcal/mol), Gly107 (-0.02 kcal/mol), Asp108 (-0.56 kcal/mol), Glu194 (0.24 kcal/mol) and Gly225 (-0.17 kcal/mol). These values indicate that the overall system was stable as described in RMSD and RMSF earlier. The fluctuations pattern observed in RMSD and MM(PB/GB)SA analyses were almost identical and indicates the system stability.

Table 6. Binding energies values.

Contribution	Energy Values (kcal.mol ⁻¹)
Van der Waals energy	-35.93
Columbic energy	-30.15
Gas phase energy	-66.44
Polar solvation free energy (PB)	44.79
Solvation free energy (PB)	66.11
Total binding free energy (PB)	-1.07
Polar solvation free energy (GB)	40.93
Solvation free energy (GB)	36.85
Electrostatic energy (PB)	15.71
Electrostatic energy (GB)	10.41
Total binding free energy (GB)	-29.59

4. CONCLUSION

The current study was based upon a combinatorial approach highlighting the G1PTT enzyme of *F. tularensis* as a potential drug target. The findings revealed ZINC23121280 ((3R)-N-(6-amino-1-benzyl-2,4-dioxo-pyrimidin-5-yl)-1-cyclopentyl-

5-oxo-N-propyl-pyrrolidine-3-carbox) as potential inhibitor of the enzyme. Although the inhibitor seems to show good binding efficacy for the enzyme the still these predictions required experimental *in vivo* and *in vitro* validation. These findings can be used to design and develop more specific, efficient, and potent drugs against *F. tularensis*.

5. ACKNOWLEDGEMENTS

All the authors are indebted to Pak-US Science and Technology Cooperation Program under the grant No. Pak-US2017/360 for providing financial assistance to make this possible.

6. CONFLICT OF INTEREST

The authors declare no conflict of interest.

7. REFERENCES

1. M.E. Evans., D.W. Gregory., W. Schaffner, and Z.A. McGee. Tularemia: a 30-year experience with 88 cases. *Medicine* 64:251–269 (1985).
2. L.D. Thomas, and W. Schaffner. Tularemia pneumonia. *Infection Disease Clinics* 24:43–55 (2010).
3. I. Golovliov., V. Baranov., Z. Krocova., H. Kovarova, and A. Sjostedt. An attenuated strain of the facultative intracellular bacterium *Francisella tularensis* can escape the phagosome of monocytic cells. *Infection Immunity* 71:5940–5950 (2003).
4. World Health Organization, WHO guidelines on Tularaemia: epidemic and pandemic alert and response (2007).
5. A. Johansson., A. Larkeryd., M. Widerstrom., S. Mortberg., K. Myrtannas., C. Ohrman., D. Birdsell., P. Keim., D.M. Wagner., M. Forsman, and P. Larsson. An Outbreak of Respiratory Tularemia Caused by Diverse Clones of *Francisella tularensis*. *Clinics Infection Diseases* 59:1546–1553 (2014).
6. A.M. Hauri., I. Hofstetter., E. Seibold., P. Kaysser., J. Eckert., H. Neubauer, and W.D. Splettstoesser. Investigating an airborne tularemia outbreak, Germany. *Emerging Infectious Diseases* 16:238–243 (2010).
7. A. Desvars., M. Furberg., M. Hjertqvist., L. Vidman., A. Sjostedt., P. Ryden, and A. Johansson. Epidemiology and ecology of tularemia in Sweden, 1984–2012. *Emerging Infectious Diseases* 21:32–39 (2015).
8. A. Sjostedt. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. *Annals of the New York Academy of Sciences* 1105:1–29 (2007).
9. J. Ellis., P.C.F. Oyston., M. Green, and R.W. Titball. Tularemia. *Clinical Microbiology Review* 15:631–646 (2002).
10. J. Su., J. Yang., D. Zhao., T.H. Kawula., J.A. Banas, and J.R. Zhang. Genome-Wide Identification of *Francisella tularensis* virulence determinants. *Infection Immunity* 75:3089–3101 (2007).
11. J.E. Staples., K.A. Kubota., L.G. Chalcraft., P.S. Mead, and J.M. Petersen. Epidemiologic and molecular analysis of human tularemia, United States, 1964–2004. *Emerging Infectious Diseases* 12:1113–1118 (2006).
12. K.J. Kugeler., P.S. Mead., A.M. Janusz., J.E. Staples., K.A. Kubota., L.G. Chalcraft, and J.M. Petersen. Molecular epidemiology of *Francisella tularensis* in the United States. *Clinical Infectious Diseases* 48:863–870 (2009).
13. J. Celli, and T.C. Zahrt. Mechanisms of *Francisella tularensis* intracellular pathogenesis. *Cold Spring Harbor Perspective Medicine* 3:a010314 (2013).
14. W.J. Dai., W. Bartens., G. Kohler., M. Hufnagel., M. Kopf, and F. Brombacher. Impaired macrophage listericidal and cytokine activities are responsible for the rapid death of *Listeria monocytogenes*-infected IFN-gamma receptor-deficient mice. *Journal of Immunology* 158:5297–5304 (1997).
15. A. Tarnvik. Nature of protective immunity to *Francisella tularensis*. *Reviews of Infectious Diseases* 11:440–451 (1989).
16. S. Ahmad., S. Raza., S.W. Abbasi, and S.S. Azam. Identification of natural inhibitors against *Acinetobacter baumannii* D-alanine-D-alanine ligase enzyme: A multi-spectrum in silico approach. *Journal of Molecular Liquids* 262:460–475 (2018).
17. S. Ahmad., S. Raza., R. Uddin, and S.S. Azam. Comparative subtractive proteomics based ranking for antibiotic targets against the dirtiest superbug: *Acinetobacter baumannii*. *Journal of Molecular Graphics and Modelling* 82:74–92 (2018).
18. G.R. Marshall. Computer-Aided Drug Design. *Annual Reviews in Pharmacology Toxicology* 27:193–213 (1987).
19. S. Ahmad., S. Raza., A. Abro., K.R. Liedl, and S.S. Azam. Toward novel inhibitors against KdsB: a highly specific and selective broad-spectrum bacterial enzyme. *Journal of Biomolecular Structure and Dynamics* 0:1–20 (2018).
20. R. Gupta., D. Pradhan., A.K. Jain, and C.S. Rai.

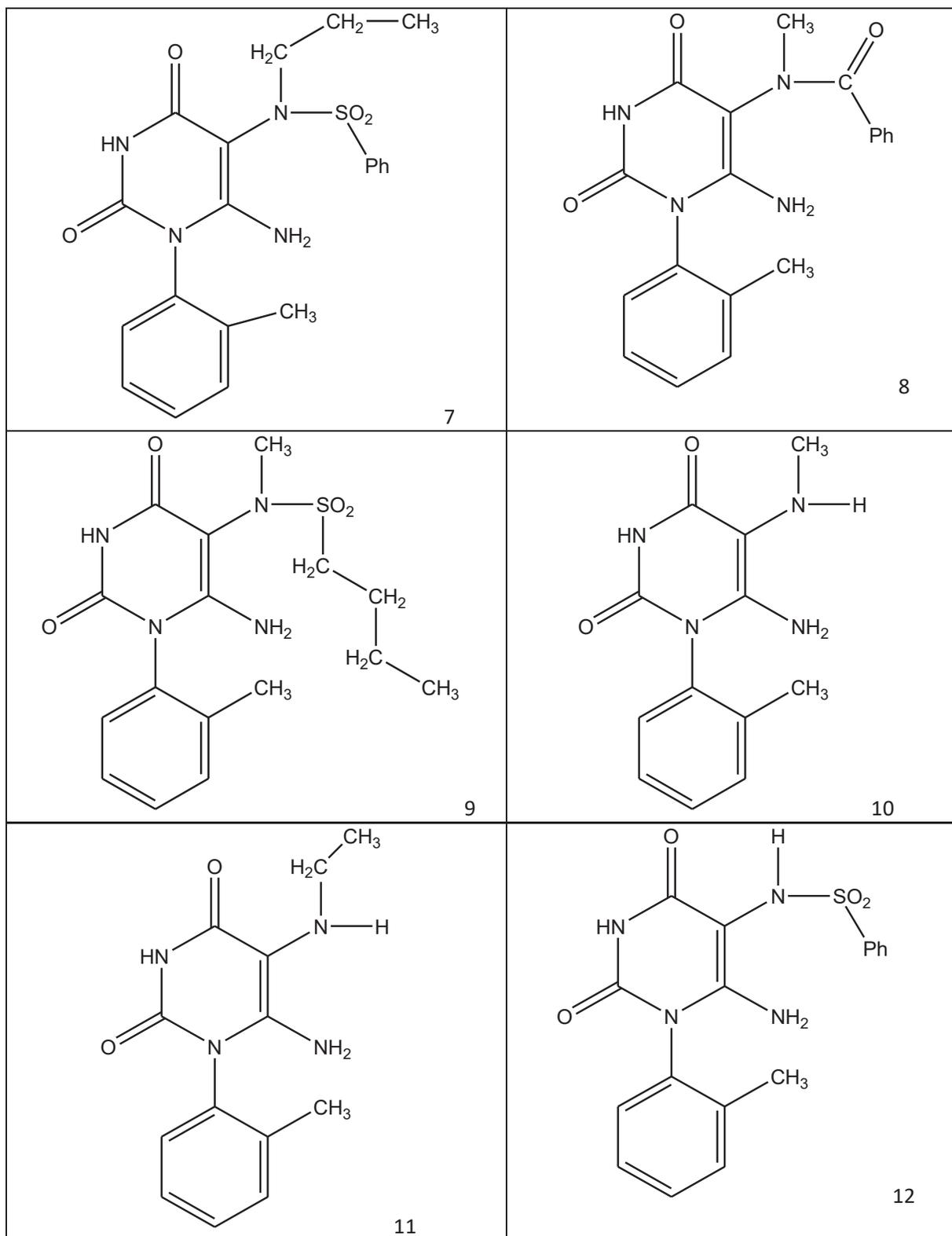
- TiD: Standalone software for mining putative drug targets from bacterial proteome. *Genomics* 109:51–57 (2017).
21. Z. Khalid., S. Ahmad., S. Raza, and S.S. Azam. Subtractive proteomics revealed plausible drug candidates in the proteome of multi-drug resistant *Corynebacterium diphtheriae*. *Meta Gene* 17: 34–42 (2018).
 22. A.L. Harvey. Natural products in drug discovery. *Drug Discovery Today* 13:894–901(2008).
 23. G. Sanober., S. Ahmad, and S.S. Azam. Identification of plausible drug targets by investigating the druggable genome of MDR *Staphylococcus epidermidis*. *Gene Reports* 7:147–153 (2017).
 24. S. Ahmad., S. Raza., R. Uddin, and S.S. Azam. Comparative subtractive proteomics based ranking for antibiotic targets against the dirtiest superbug: *Acinetobacter baumannii*. *Journal of Molecular Graphics and Modelling* 82:74–92 (2018).
 25. R. Zhang, and Y. Lin. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Research* 37:D455–D458 (2009).
 26. S. Zuccotti., D. Zanardi., C. Rosano., L. Sturla., M. Tonetti, and M. Bolognesi. Kinetic and crystallographic analyses support a sequential-ordered bi-bi catalytic mechanism for *Escherichia coli* glucose-1-phosphate thymidyltransferase. *Journal of Molecular Biology* 313:831–843 (2001).
 27. S. Ahmad., S. Raza., R. Uddin, and S.S. Azam. Binding mode analysis, dynamic simulation and binding free energy calculations of the MurF ligase from *Acinetobacter baumannii*. *Journal of Molecular Graphics and Modelling* 77:72–85 (2017).
 28. C. Wang., P.H. Nguyen., K. Pham., D. Huynh., T.B.N. Le., H. Wang, P. Ren, and R. Luo. Calculating protein–ligand binding affinities with MMPBSA: method and error analysis. *Journal of Computational Chemistry* 37:2436–2446 (2016).
 29. UniProt. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 46:2699 (2018).
 30. S. Twine., M. Bystrom., W. Chen., M. Forsman., I. Golovliov., A. Johansson., J. Kelly., H. Lindgren., K. Svensson., C. Zingmark., W. Conlan, and A. Sjostedt. a mutant of *Francisella tularensis* strain Schu S4 lacking the ability to express a 58-kilodalton protein is attenuated for virulence and is an effective live vaccine. *Infection Immunity* 73:8345–8352 (2005).
 31. G. Sanober., S. Ahmad, and S.S. Azam. Identification of plausible drug targets by investigating the druggable genome of MDR *Staphylococcus epidermidis*. *Gene Reports* 7:147–153 (2017).
 32. Y. Huang., B. Niu., Y. Gao., L. Fu, and W. Li. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682 (2010).
 33. AR. Ortiz., P. Gomez-Puertas., A. Leo-Macias., P. Lopez-Romero, and E. Lopez-Vinas. Computational Approaches to Model Ligand Selectivity in Drug Design. *Current Topics in Medicine Chemistry* 6:41–55 (2006).
 34. R. Cloete., E. Oppon., E. Murungi., W.D. Schubert, and A. Christoffels. Resistance related metabolic pathways for drug target identification in *Mycobacterium tuberculosis*. *BMC Bioinformatics* 17: 75 (2016).
 35. Y. Moriya., M. Itoh., S. Okuda., A.C. Yoshizawa, and M. Kanehisa. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35:182–185 (2007).
 36. A. Naz., F.M. Awan., A. Obaid., S.A. Muhammad., R.Z. Paracha., J. Ahmad, and A. Ali. Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: A reverse vaccinology based approach. *Infection Genetics Evolution* 32:280–291 (2015).
 37. L. Chen., J. Yang., J. Yu., Z. Yao., L. Sun., Y. Shen, and Q. Jin. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* 33:325–328 (2005).
 38. N.Y. Yu., J.R. Wagner., M.R. Laird., G. Melli., S. Rey., R. Lo., P. Dao., S.C. Sahinalp., M. Ester., L.J. Foster, and F.S.L. Brinkman. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615 (2010).
 39. C.S. Yu., Y.C. Chen., C.H. Lu, and J.K. Hwang. Prediction of protein subcellular localization. *Proteins Structure Function Bioinformatics* 64:643–651 (2006).
 40. C.S. Yu., C.W. Cheng., W.C. Su., K.C. Chang., S.W. Huang., J.K. Hwang, and C.H. Lu. CELLO2GO: a web server for protein subcellular localization prediction with functional gene ontology annotation. *PLOS ONE* 9:e99368 (2014).
 41. S.S. Azam, and A. Shamim. An insight into the exploration of druggable genome of *Streptococcus gordonii* for the identification of novel therapeutic candidates. *Genomics* 104:203–214 (2014).

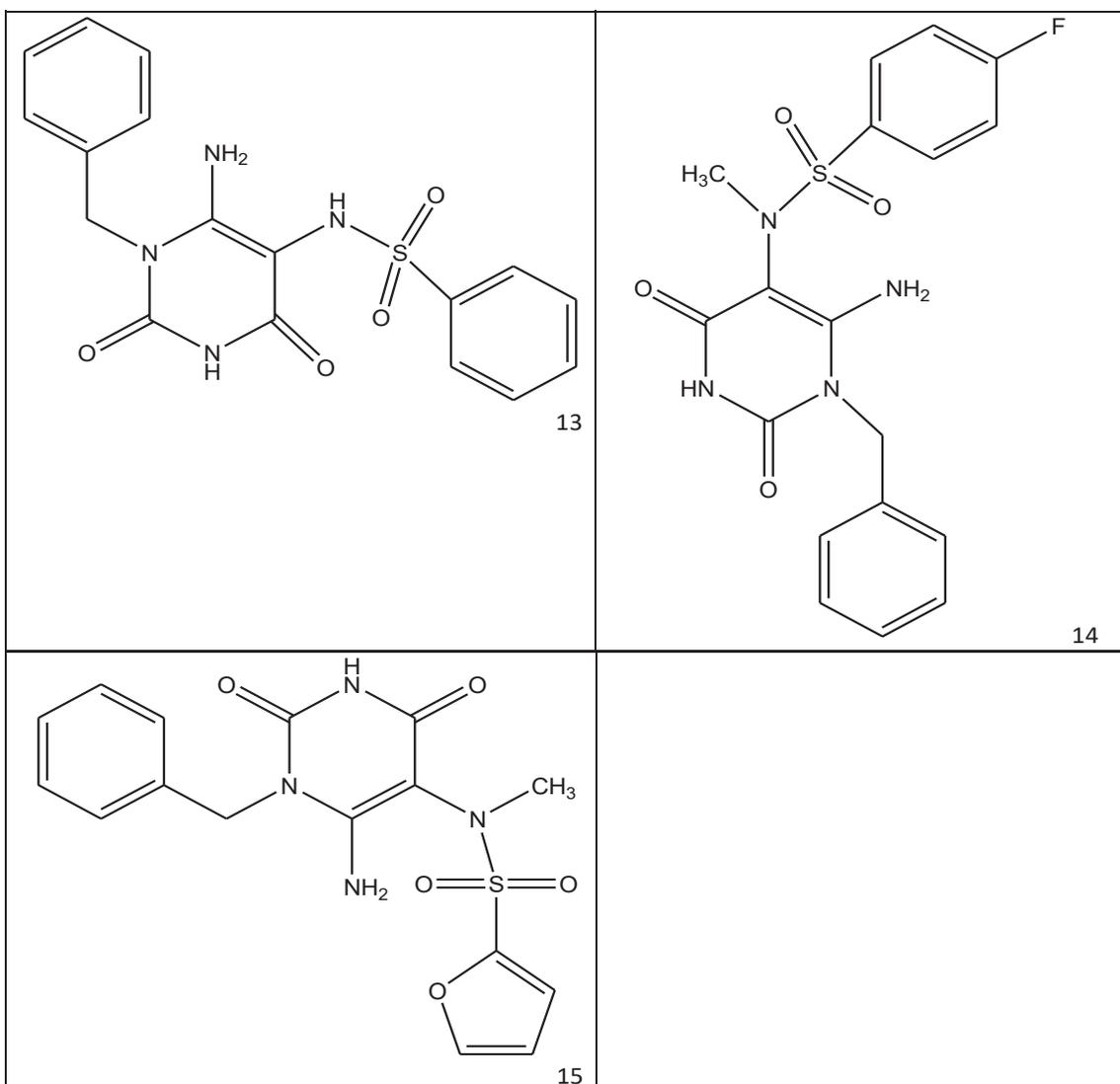
42. Y.Asad., S. Ahmad., T. Rungrotmongkol., K.E. Ranaghan, and S.S. Azam. Immuno-informatics driven proteome-wide investigation revealed novel peptide-based vaccine targets against emerging multiple drug resistant *Providencia stuartii*. *Journal of Molecular Graphics and Modelling* 80:238–250 (2018).
43. E.Gasteiger., C. Hoogland., A. Gattiker., S. Duvaud., M.R. Wilkins., R.D. Appel, and S. Bairoch, A. Protein identification and analysis tools on the expasy server. In: the proteomics protocols handbook. Humana Press, pp 571–607 (2005).
44. B. Hesp, and A.K. Willard. Enzymes as targets for drug design. In: Enzyme Chemistry. Springer, Dordrecht, pp 119–161 (1984).
45. M. Kallberg., H. Wang., S. Wang., J. Peng., Z. Wang., H. Lu, and J. Xu. Template-based protein structure modeling using the RaptorX web server. *Nature Protocol* 7:1511–1522 (2012).
46. Kelley, L.A. and Sternberg, M.J.E. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocol* 4:363–371 (2009).
47. K.Arnold., L. Bordoli., J. Kopp, and T. Schwede. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201 (2006).
48. A. Roy., A. Kucukural, and Y. Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocol* 5:725–738 (2010).
49. D. Eisenberg., R. Luthy, and J.U. Bowie. VERIFY3D: Assessment of protein models with three-dimensional profiles. In: methods in enzymology. Academic Press, pp 396–404 (1997).
50. M.J. Wiederstein, and M. Sippl. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* 35:407–10 (2007).
51. R.A. Laskowski., J.A. Rullmannn., M.W. MacArthur., R. Kaptein, and J.M. Thornton. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* 8:477–486 (1996).
52. E.F. Pettersen., T.D. Goddard., C.C. Huang., G.S. Couch., D.M. Greenblatt., E.C. Meng, and T.E. Ferrin. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25:1605–1612 (2004).
53. S. Baseer., S. Ahmad., K.E. Ranaghan, and S.S. Azam. Towards a peptide-based vaccine against *Shigella sonnei*: A subtractive reverse vaccinology based approach. *Biologicals* 50:87–99 (2017).
54. C.A. Lipinski., F. Lombardo., B.W. Dominy, and P.J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3–26.
55. J. Lange., L.S. Wyrwicz, and G. Vriend. (2016) KMAD: knowledge-based multiple sequence alignment for intrinsically disordered proteins. *Bioinformatics* 32:932–936 (2001).
56. B.Huang. MetaPocket: A meta approach to improve protein ligand binding site prediction. *OMICS: A Journal of Integrative Biology* 13:325–330 (2009).
57. J.D. Thompson., J. Gibson, and D.G. Higgins. Multiple sequence alignment using clustalw and clustalx. *Current Protocols in Bioinformatics* 1:2-3 (2002).
58. M.L. Verdonk., J.C. Cole., M.J. Hartshorn., C.W. Murray, and R.D. Taylor. Improved protein–ligand docking using GOLD. *Proteins Structure Function Bioinformatics* 52:609–623 (2003).
59. O. Trott, and A.J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:455–461 (2010).
60. E.F. Pettersen., T.D. Goddard., C.C. Huang., G.S. Couch., D.M. Greenblatt., E.C. Meng, and T.E.L. Ferrin. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25:1605–1612 (2004).
61. W. Humphrey., A. Dalke, and K.Schulten. VMD: visual molecular dynamics. *Journal of Molecular Graphics* 14:33–38 (1996).
62. A.Abro, and S.S. Azam. Binding free energy based analysis of arsenic (+3 oxidation state) methyltransferase with S-adenosylmethionine. *Journal of Molecular Liquids* 220:375–382 (2016).
63. D.A. Case., V. Babin., J. Berryman., R.M. Betz., Q. Cai., D.S. Cerutti., I.T.E. Cheatham., T.A. Darden., R.E. Duke., H. Gohlke, and A.W. Goetz. Amber 14.
64. P.K. Weiner, and P.A. Kollman. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry* 2:287–303 (1981).
65. S. Andleeb., I.U. Din., M.K. Rauf., S.S. Azam., A. Badshah., H. Sadaf., A. Raheel., M.N. Tahir, and S. Raza. A one-pot multicomponent facile synthesis of dihydropyrimidin-2(1H)-thione derivatives using triphenylgermane as a catalyst and its binding

- pattern validation. *RSC Advances* 6:79651–79661 (2016).
66. V. Krautler., G.W.F. Van, and P.H. Hunenberger. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry* 22:501–508 (2001).
 67. H.C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics* 72:2384–2393 (1980).
 68. D.R. Roe, and T.E. Cheatham. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* 9:3084–3095 (2013).
 69. World Health Organization. WHO publishes list of bacteria for which new antibiotics are urgently needed. WHO Geneva Switzerland (2017).
 70. C.L. Ventola. The Antibiotic Resistance Crisis. *Pharmacy and Therapeutics* 40:277–283 (2015)
 71. B.M. Hassan., K. Ahmad., S. Roy., A.J. Mohammad., M. Adil., S.S. Haris., S. Khan., K.M. Amjad., I. Provazník, and I. Choi. Computer aided drug design: success and limitations. *Current pharmaceutical design* 22: 572-581 (2016).
 72. J.Drews. Drug Discovery: A Historical Perspective. *Science* 287:1960–1964 (2000).
 73. T.Hossain., M. Kamruzzaman., T.Z. Choudhury., H.N. Mahmood., A.H.M.N Nabi, and M.I. Hosen. Application of the subtractive genomics and molecular docking analysis for the identification of novel putative drug targets against *Salmonella enterica* subsp. *enterica* serovar Poona. *BioMed Research International* 2017: (2017)
 74. K.B. Patel., E. Toh., X.B. Fernandez., A. Hanuszkiewicz., G.G. Hardy., Y.V. Brun., M.A. Bernards, and M.A. Valvano. Functional Characterization of UDP-Glucose:Undecaprenyl-Phosphate Glucose-1-Phosphate Transferases of *Escherichia coli* and *Caulobacter crescentus*. *Journal of Bacteriology* 194:2646–2657 (2012).
 75. V. Sperandio., A.G. Torres, and J.B. Kaper. Quorum sensing *Escherichia coli* regulators B and C (QseBC): a novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in *E. coli*. *Molecular Microbiology* 43:809–821 (2002).
 76. G. McMullan, and J.P. Quinn. In vitro characterization of a phosphate starvation-independent carbon-phosphorus bond cleavage activity in *Pseudomonas fluorescens* 23F. *Journal of Bacteriology* 176:320–324 (1994).
 77. R.E. Webster. The *tol* gene products and the import of macromolecules into *Escherichia coli*. *Molecular Microbiology* 5:1005–1011 (1991).
 78. A. Wadood., M. Ghufra., S.F. Hassan., H. Khan., S.S. Azam, and U. Rashid. *In silico* identification of promiscuous scaffolds as potential inhibitors of 1-deoxy-d-xylulose 5-phosphate reductoisomerase for treatment of *Falciparum malaria*. *Pharmaceutical Biology* 55:19–32 (2017).
 79. E. Rutkowska., K. Pajak, and K. Jozwiak. Lipophilicity-- methods of determination and its role in medicinal chemistry. *Acta Poloniae Pharmaceutica -Drug Research* 70:3–18 (2013).
 80. S.A. Adcock, and J.A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews* 106:1589–1615 (2006).
 81. S. Abbasi., S. Raza., S.S. Azam., K.R. Liedl, and J.E. Fuchs. Interaction mechanisms of a melatonergic inhibitor in the melatonin synthesis pathway. *Journal of Molecular Liquids* 221:507–517 (2016).
 82. J. Donohue. Radial distribution functions of some structures of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America* 40:377–381 (1954).

Table 1S. Supplementary Table 1

<p>1</p>	<p>2</p>
<p>3</p>	<p>4</p>
<p>5</p>	<p>6</p>

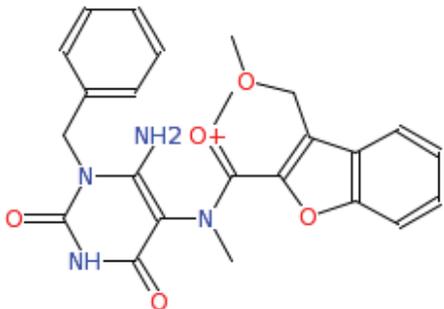
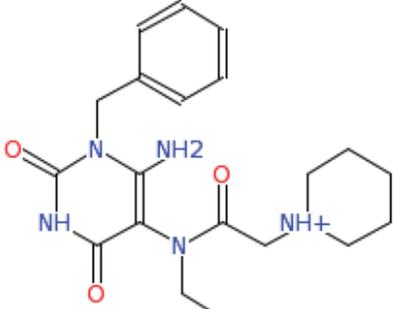
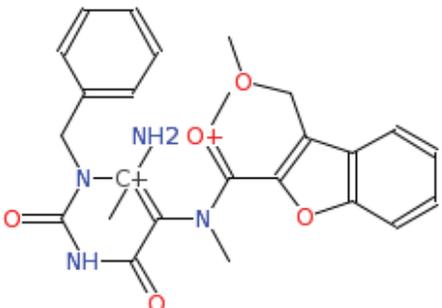
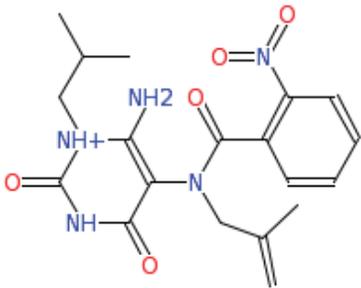
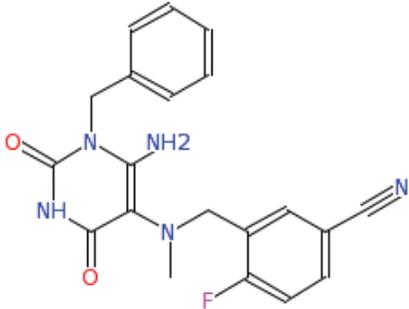
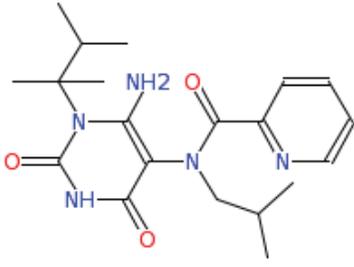
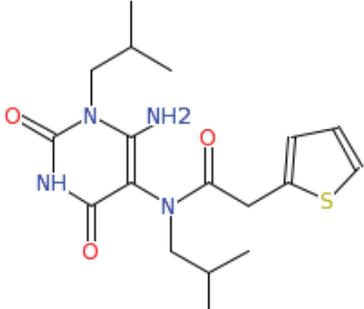
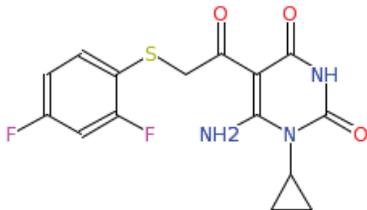


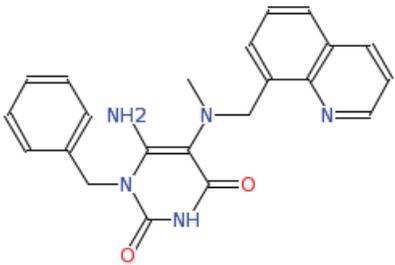
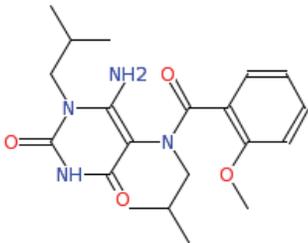
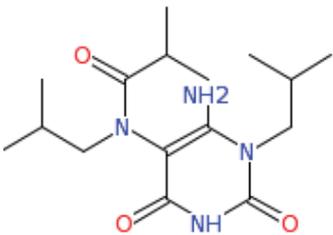
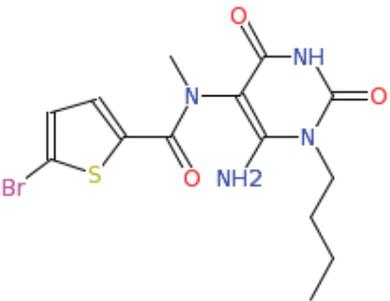
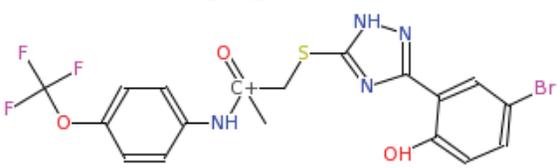
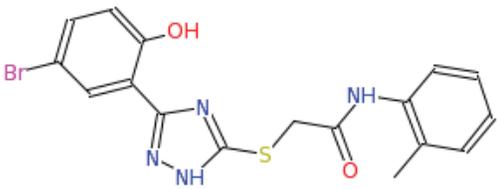
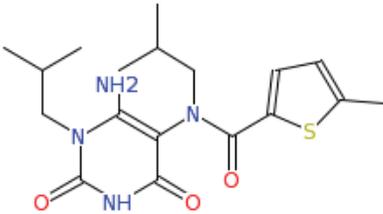


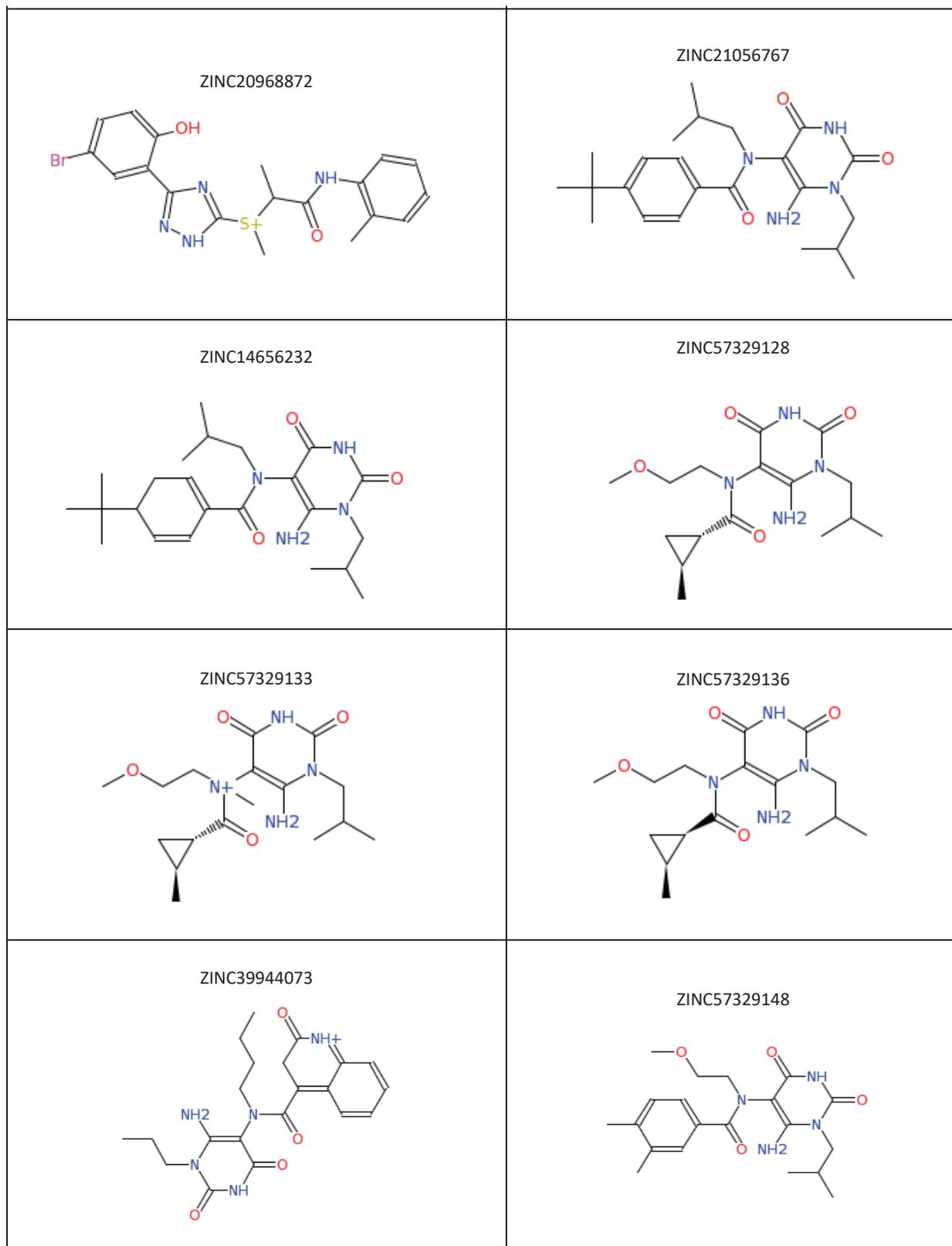
Name of Compounds

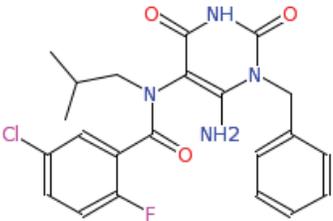
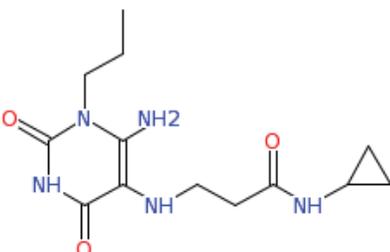
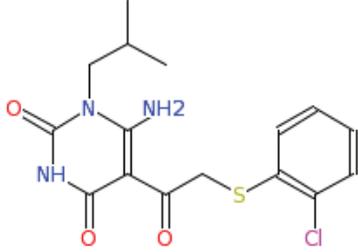
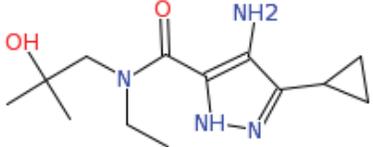
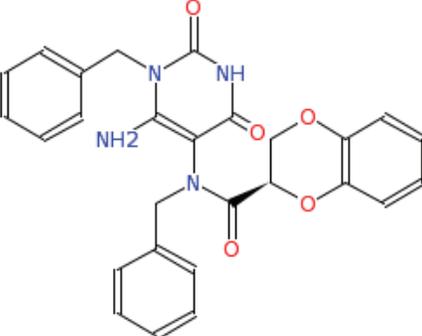
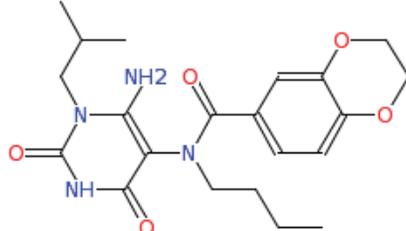
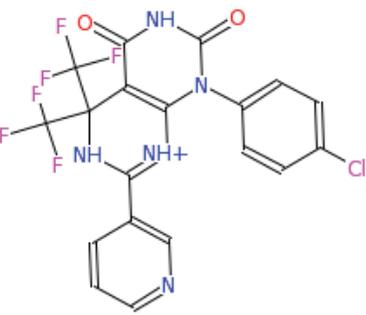
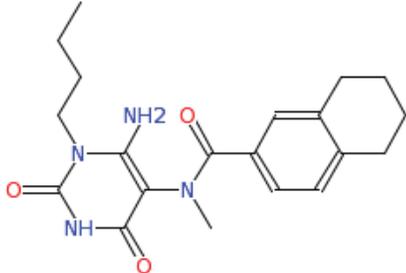
- N-(6-amino-1-butyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)-N-methylbenzenesulfonamide
- 5-amino-6-hydroxy-1-o-tolylpyrimidine-2,4(1H,3H)-dione
- N-(6-amino-1-butyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)-N-methylbenzamide
- N/A
- N-(6-amino-2,4-dioxo-1-o-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)-N-methylbenzenesulfonamide
- N-(6-amino-2,4-dioxo-1-o-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)-N-ethylbenzenesulfonamide
- N-(6-amino-2,4-dioxo-1-o-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)-N-propylbenzenesulfonamide
- N-(6-amino-2,4-dioxo-1-o-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)-N-methylbenzamide
- N-(6-amino-2,4-dioxo-1-o-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)-N-methylbutane-1-sulfonamide
- 6-amino-5-(methylamino)-1-o-tolylpyrimidine-2,4(1H,3H)-dione
- 6-amino-5-(ethylamino)-1-o-tolylpyrimidine-2,4(1H,3H)-dione
- N-(6-amino-2,4-dioxo-1-o-tolyl-1,2,3,4-tetrahydropyrimidin-5-yl)benzenesulfonamide
- N-(6-amino-1-benzyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)benzenesulfonamide
- N-(6-amino-1-benzyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)-4-fluoro-N-methylbenzenesulfonamide
- N-(6-amino-1-benzyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)-N-methylfuran-2-sulfonamide

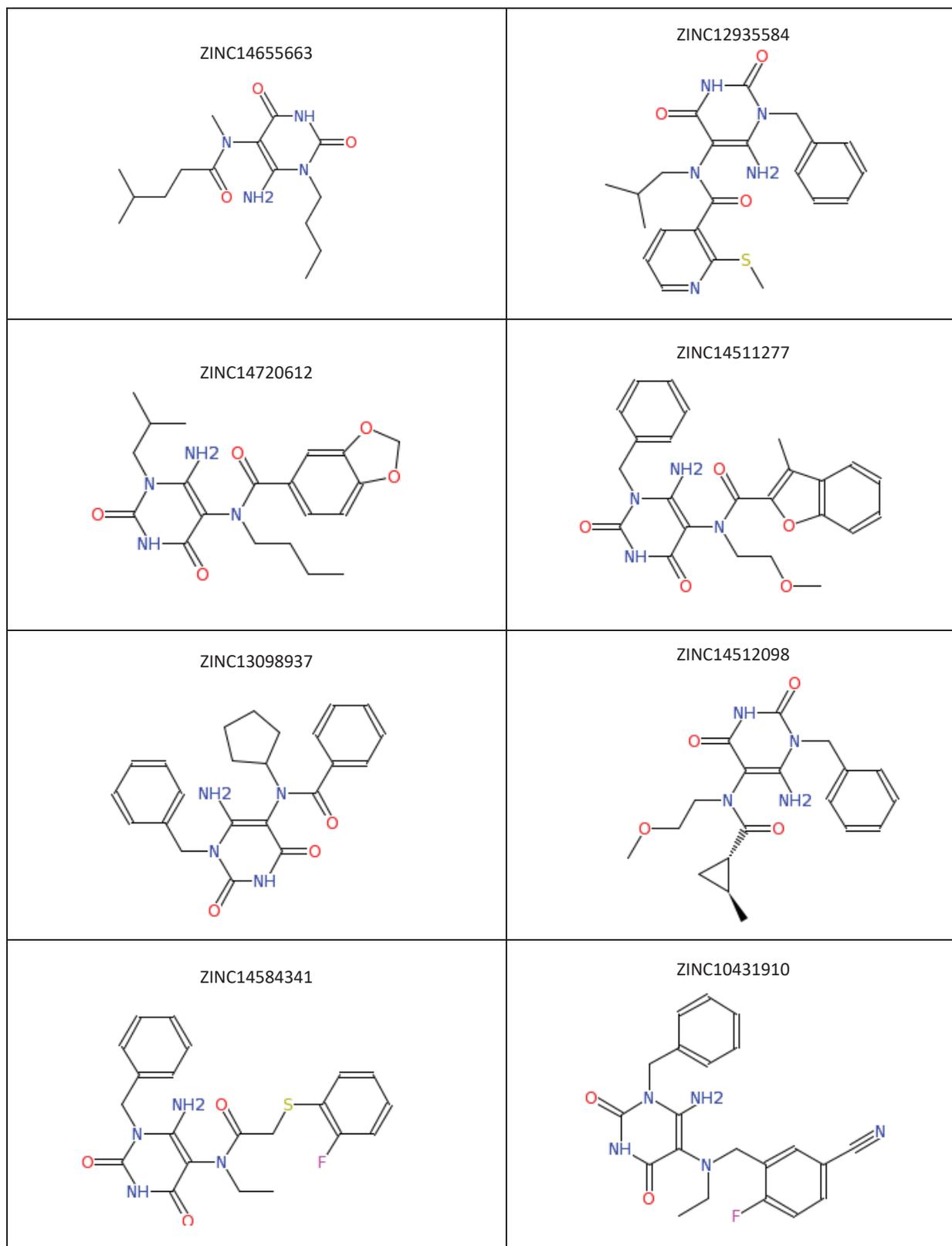
Table 2S. Supplementary Table 2

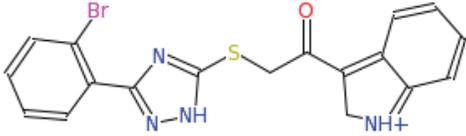
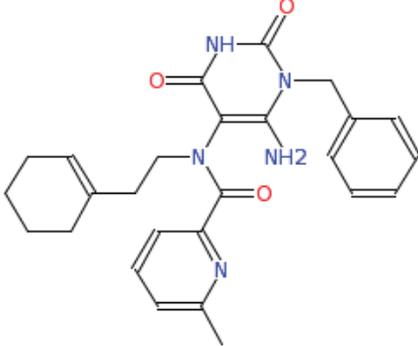
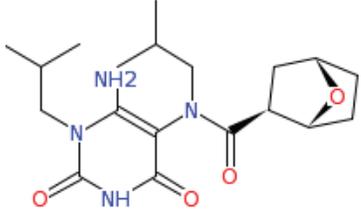
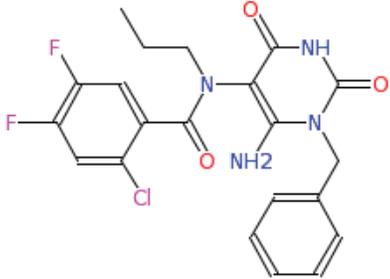
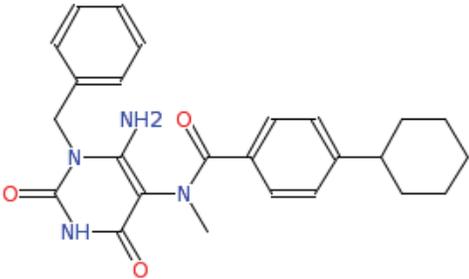
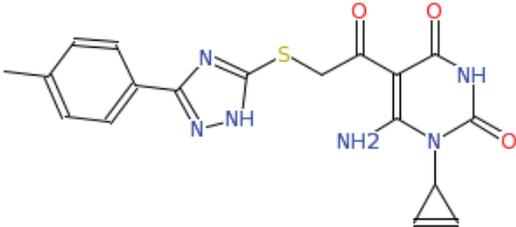
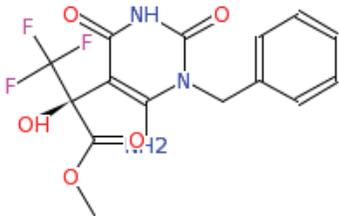
<p>ZINC09255350</p> 	<p>ZINC02629047</p> 
<p>ZINC21083072</p> 	<p>ZINC20968846</p> 
<p>ZINC12688361</p> 	<p>ZINC20968850</p> 
	<p>ZINC09503746</p> 

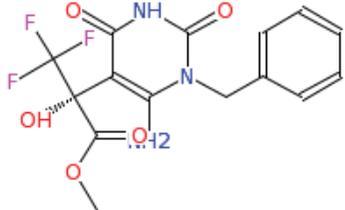
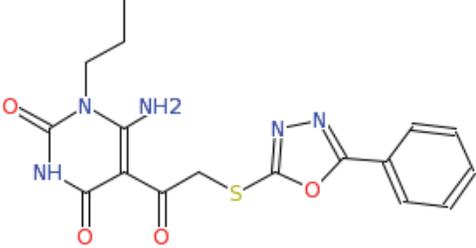
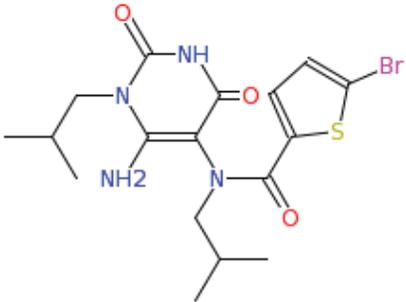
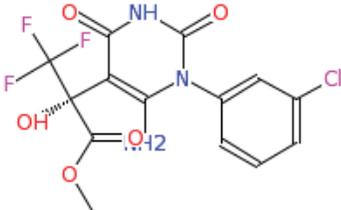
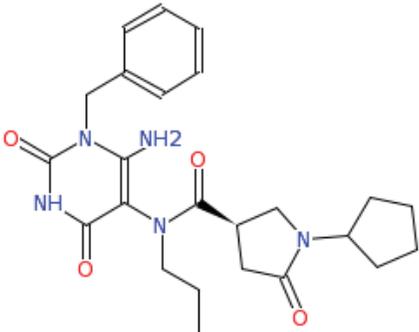
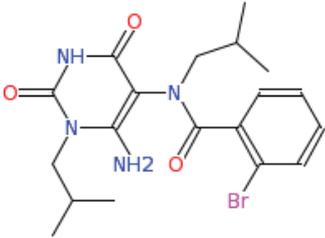
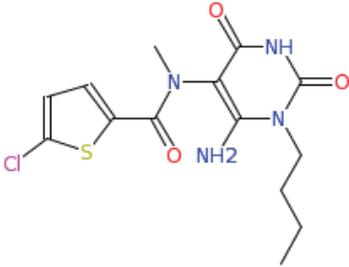
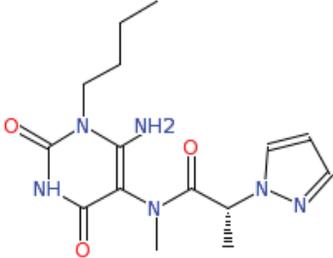
<p>ZINC12688372</p> 	<p>ZINC21056754</p> 
<p>ZINC20968860</p> 	<p>ZINC14656215</p> 
<p>ZINC20968864</p> 	<p>ZINC06973908</p> 
<p>ZINC20968868</p> 	<p>ZINC14670047</p> 

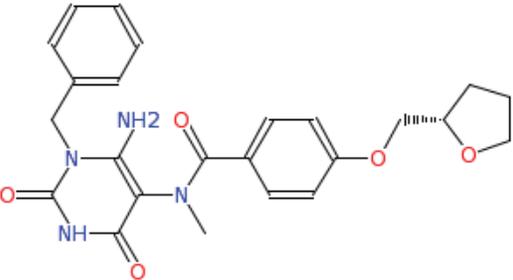
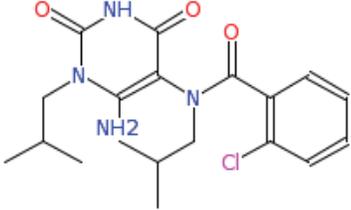
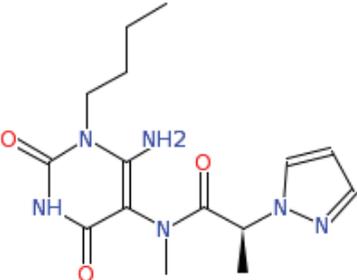
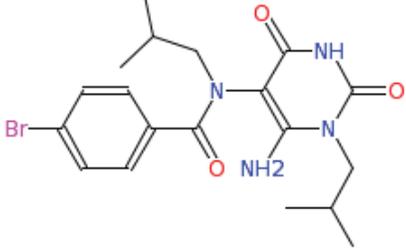
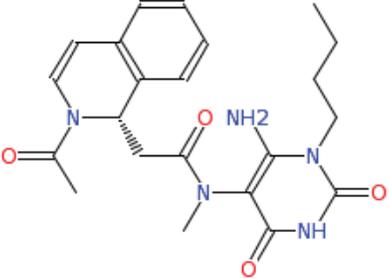
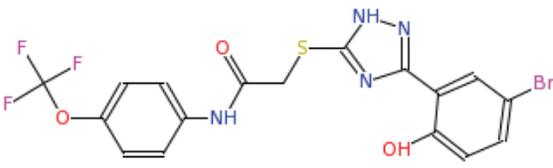
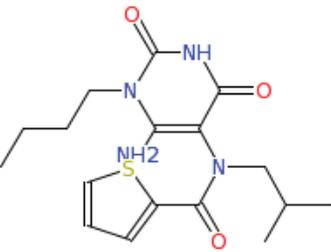
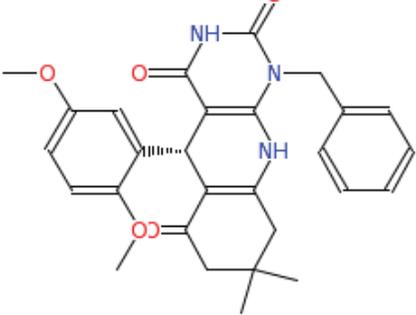


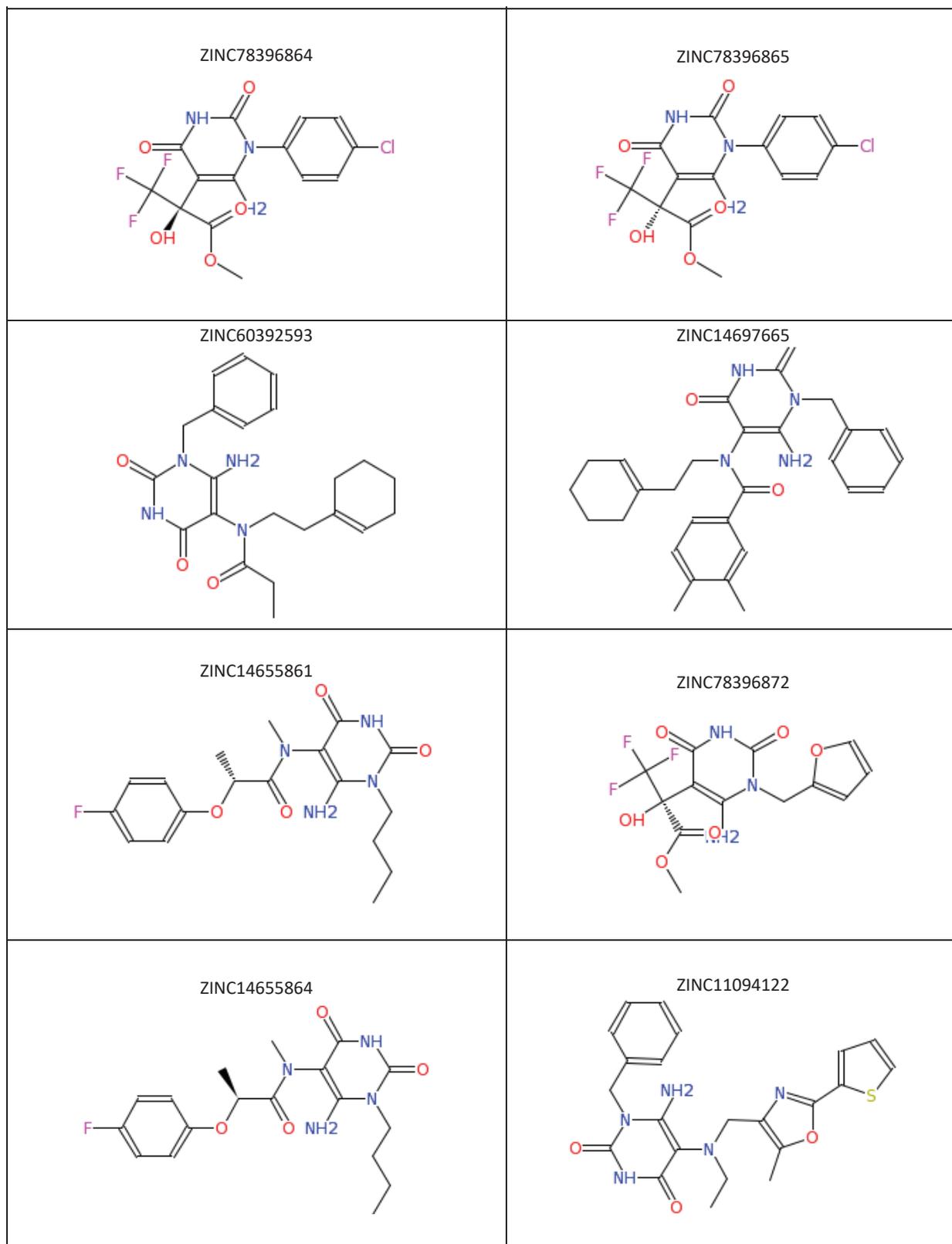
<p>ZINC13006236</p> 	<p>ZINC92996529</p> 
<p>ZINC22066775</p> 	<p>ZINC92958387</p> 
<p>ZINC12568770</p> 	<p>ZINC14720788</p> 
<p>ZINC19848781</p> 	<p>ZINC15480656</p> 

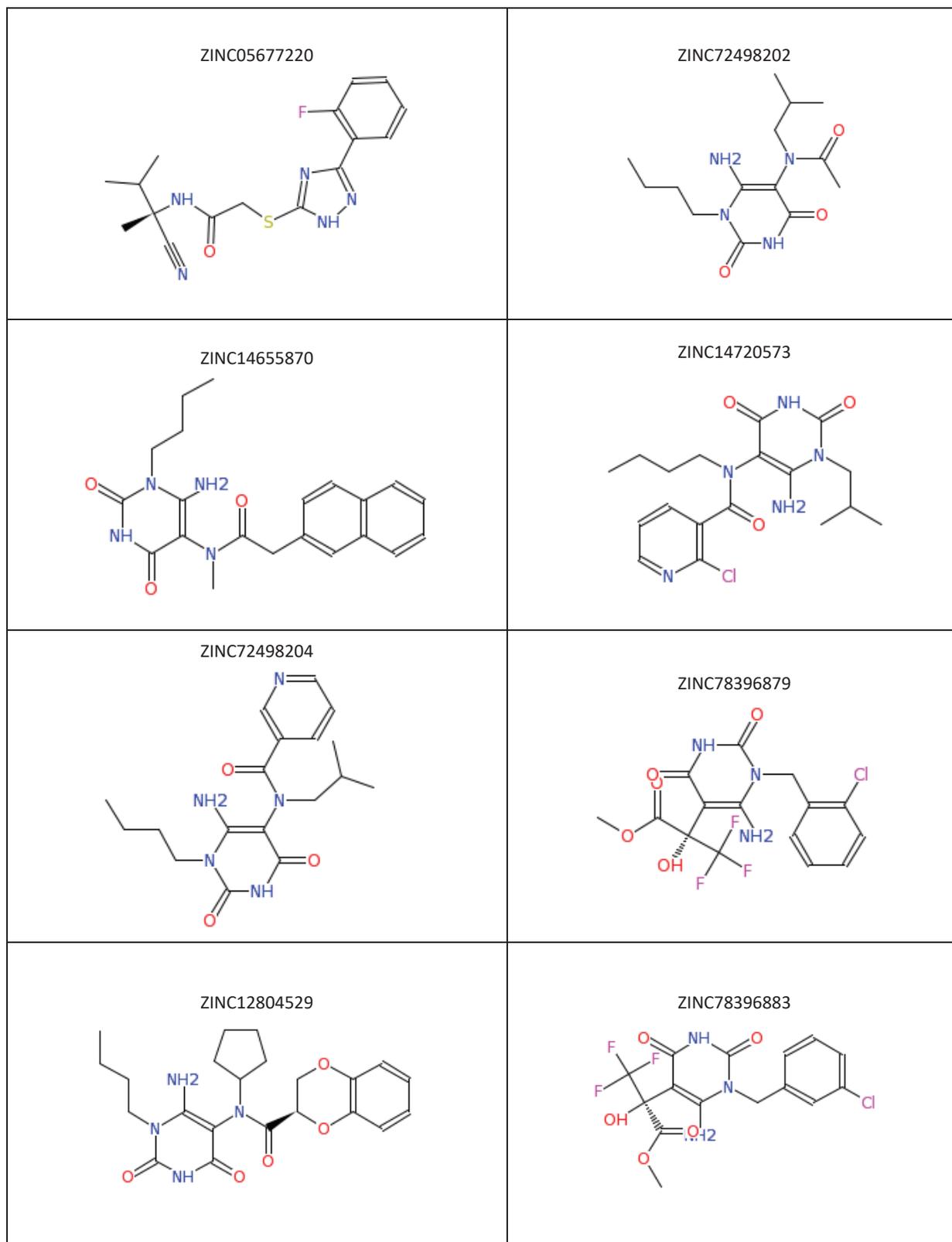


<p>ZINC40061682</p> 	<p>ZINC24616811</p> 
<p>ZINC72645280</p> 	<p>ZINC12821917</p> 
<p>ZINC09255029</p> 	<p>ZINC08343860</p> 
<p>ZINC20989147</p> 	<p>ZINC12417381</p> 

<p>ZINC12417384</p> 	<p>ZINC03348170</p> 
<p>ZINC20989150</p> 	<p>ZINC78396862</p> 
<p>ZINC23121280</p> 	<p>ZINC21056728</p> 
<p>ZINC14656171</p> 	<p>ZINC65433956</p> 

<p>ZINC09255338</p> 	<p>ZINC21056731</p> 
<p>ZINC65433958</p> 	<p>ZINC21056736</p> 
<p>ZINC15605260</p> 	<p>ZINC09235457</p> 
<p>ZINC14670114</p> 	<p>ZINC44974019</p> 

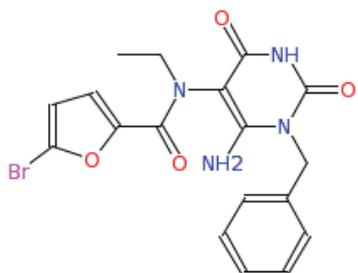




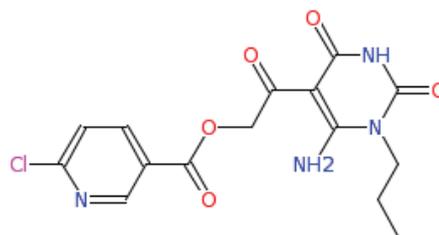
ZINC09467411

ZINC09452073

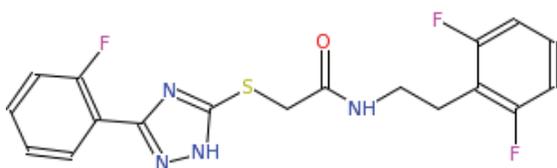
ZINC09467411



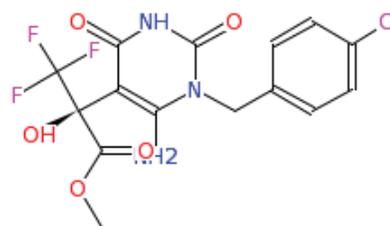
ZINC09452073



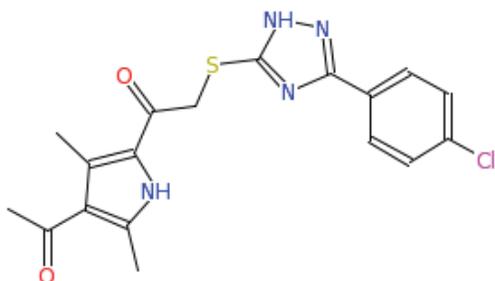
ZINC53264670



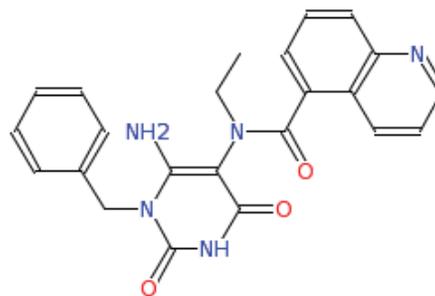
ZINC78396886



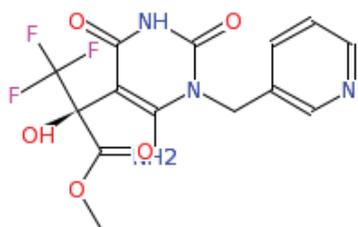
ZINC28807288



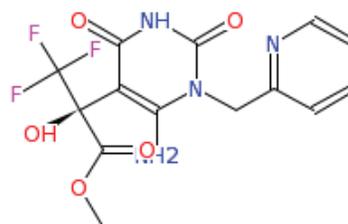
ZINC12270700

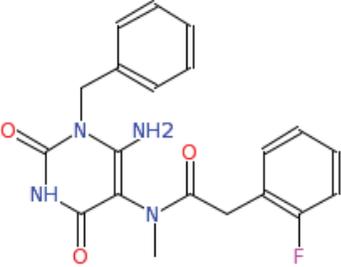
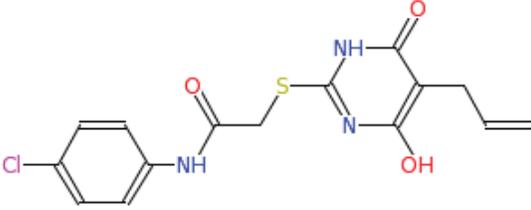
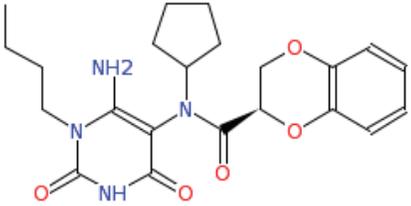
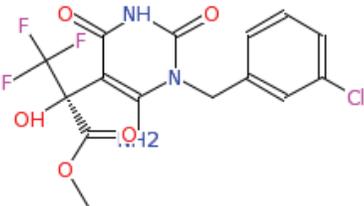
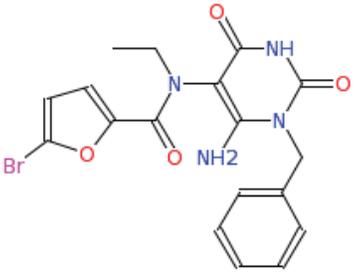
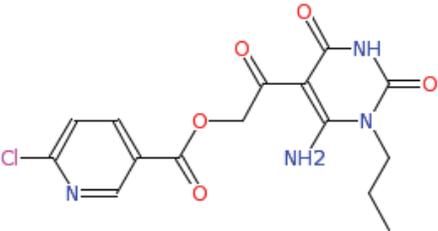
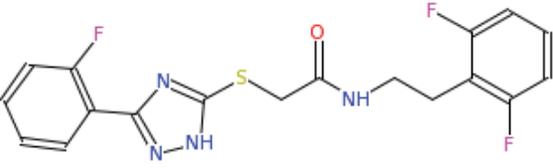
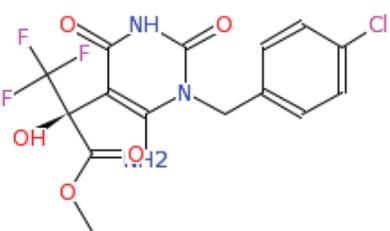


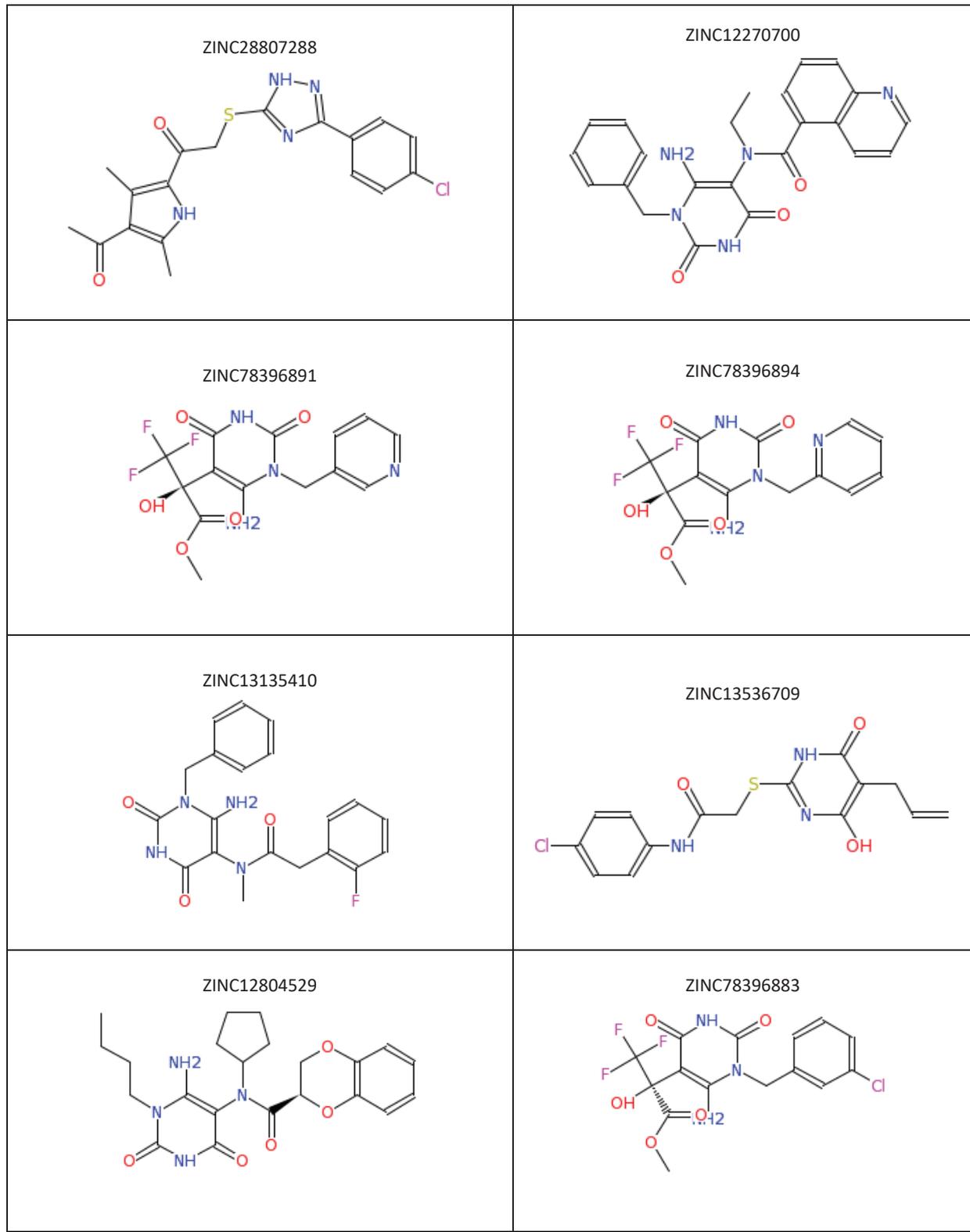
ZINC78396891



ZINC78396894

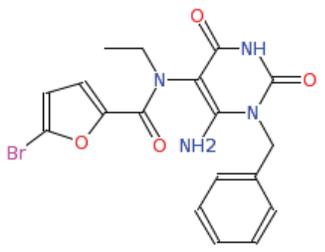
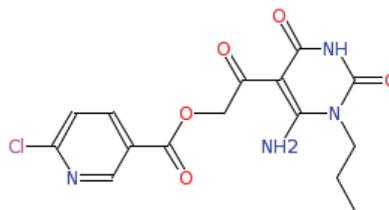
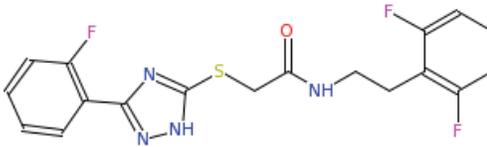
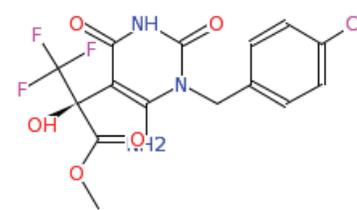
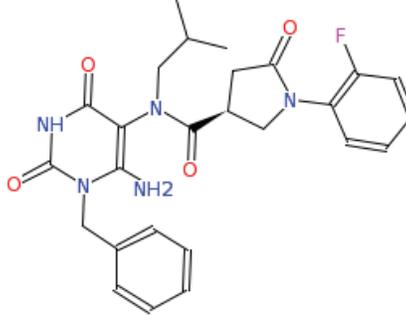
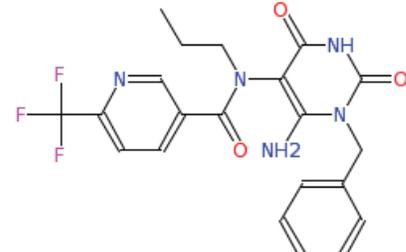
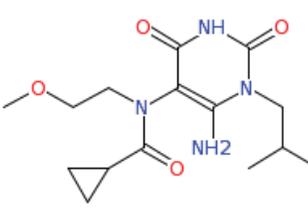
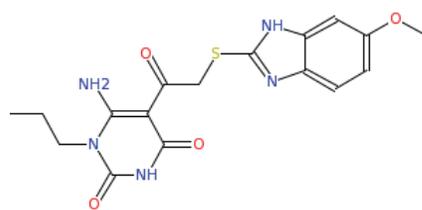
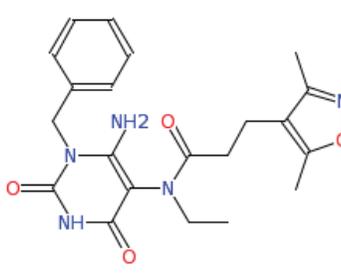


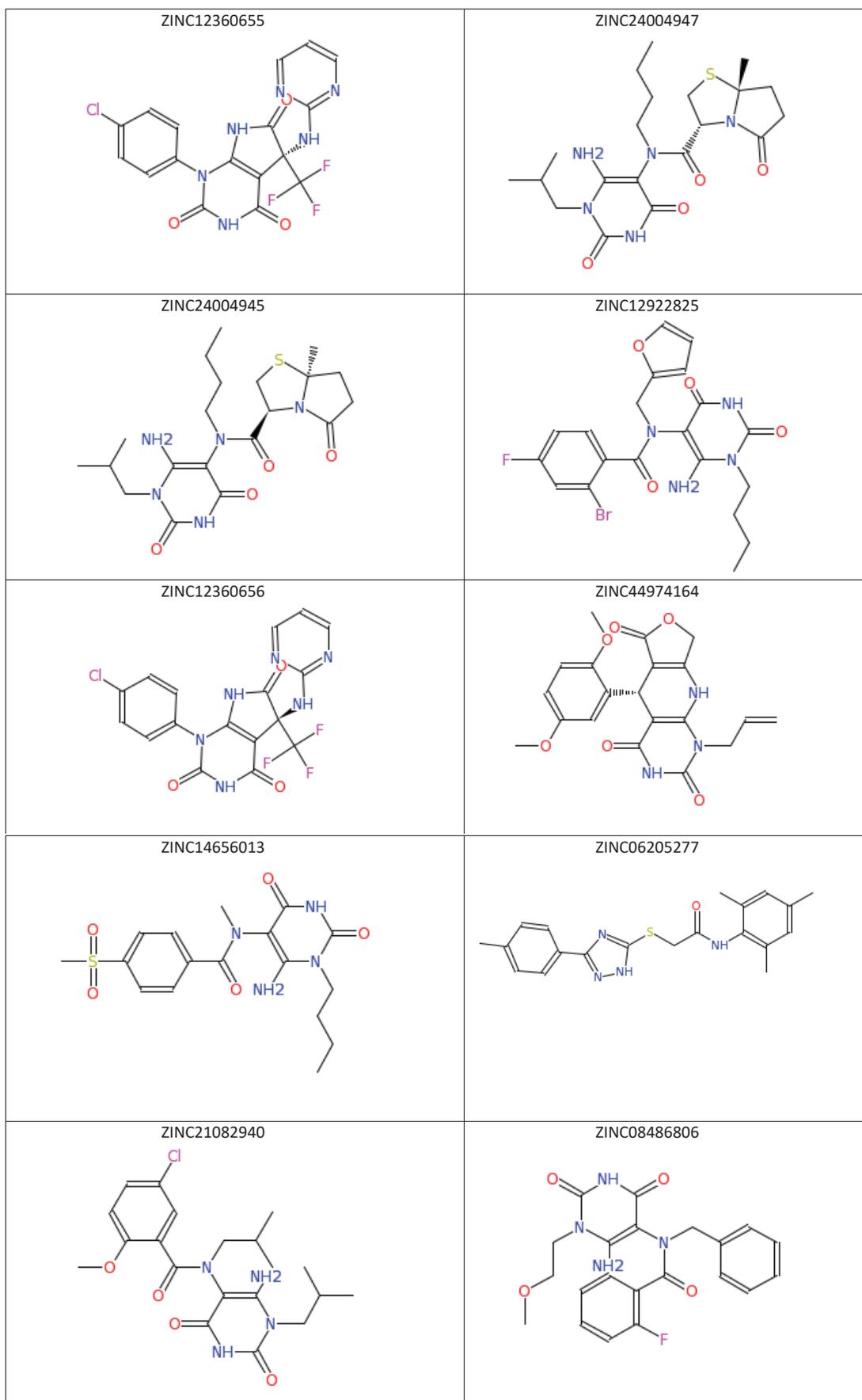
<p>ZINC13135410</p> 	<p>ZINC13536709</p> 
<p>ZINC12804529</p> 	<p>ZINC78396883</p> 
<p>ZINC09467411</p> 	<p>ZINC09452073</p> 
<p>ZINC53264670</p> 	<p>ZINC78396886</p> 

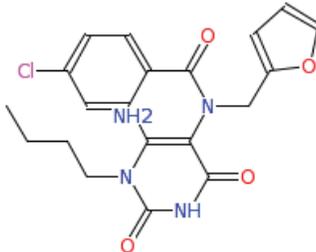
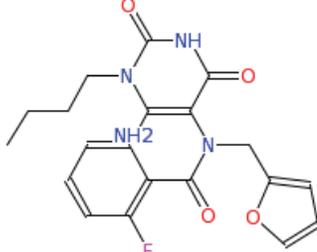
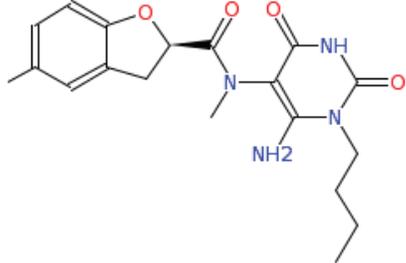
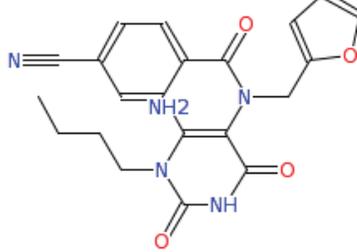
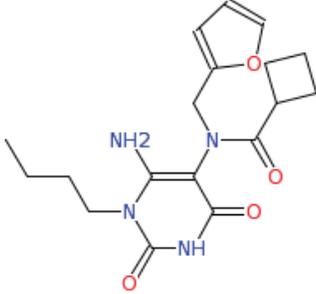
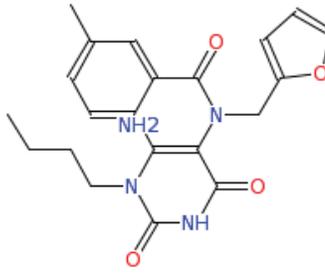
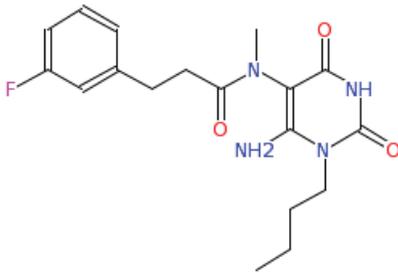
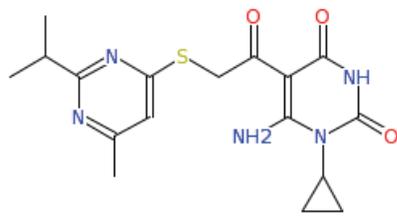
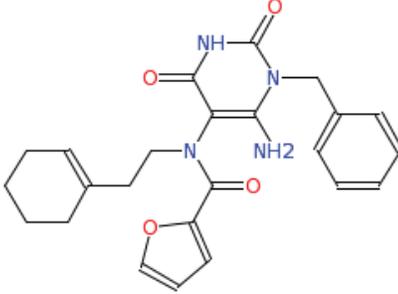
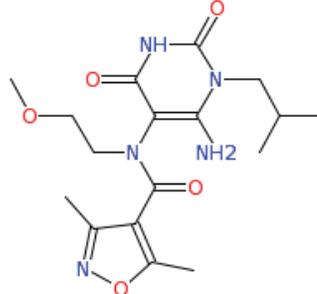


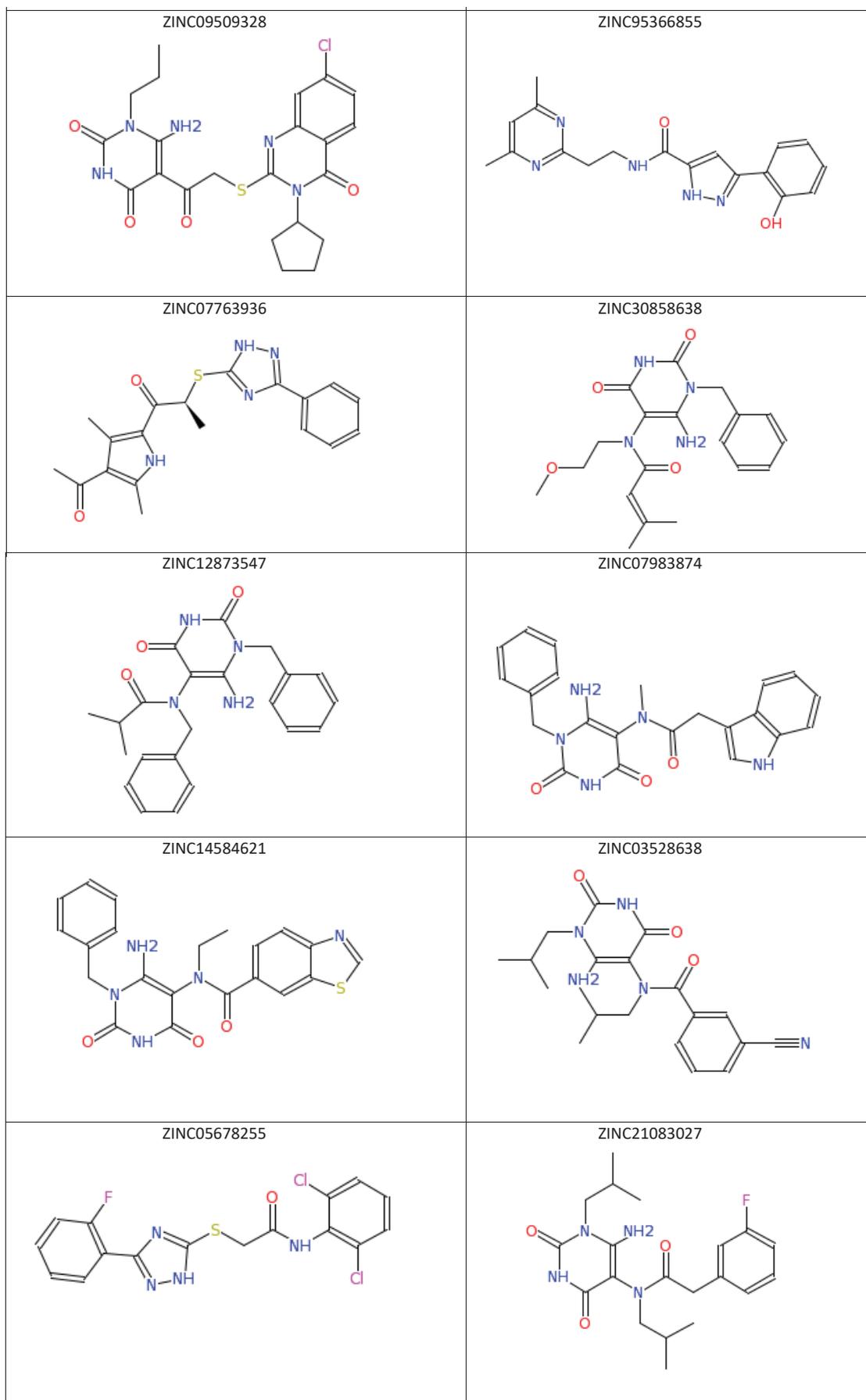
ZINC09467411

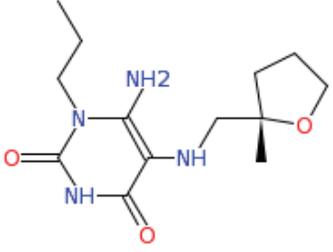
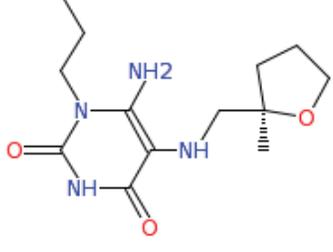
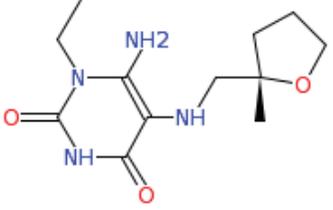
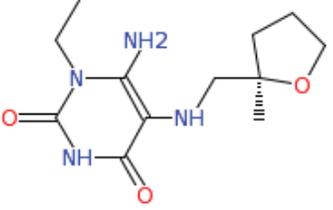
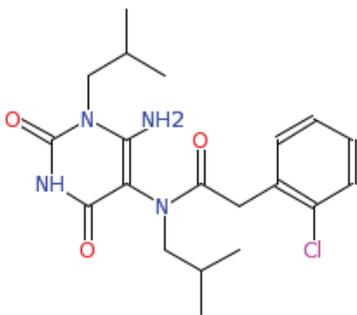
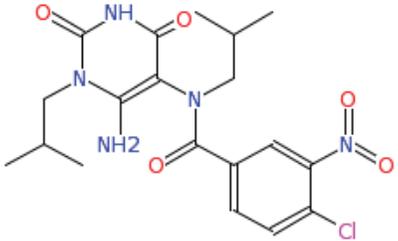
ZINC09452073

<p>ZINC09467411</p>  <chem>CCN(C1=NC(=O)NC(=O)C1=NC2=CC=CC=C2)C(=O)c3cc(Br)oc3</chem>	<p>ZINC09452073</p>  <chem>CCCCN1C(=O)NC(=O)C1=NC(=O)C2=CC=C(Cl)C=C2CO</chem>
<p>ZINC53264670</p>  <chem>COc1cnc2c(=O)[nH]c(=O)c2n1CSC3=NC=CC(F)=C3F</chem>	<p>ZINC78396886</p>  <chem>CCCCN1C(=O)NC(=O)C1=NC(=O)C2=CC=C(Cl)C=C2OC(O)C</chem>
<p>ZINC24833725</p>  <chem>CC1(C)C(=O)N(C1)C2=CC=C(F)C=C2N(C1)C(=O)NC(=O)C1=NC(=O)C2=CC=CC=C2N1C</chem>	<p>ZINC12763362</p>  <chem>CCCCN1C(=O)NC(=O)C1=NC(=O)C2=CC=C(F)C=C2FCC3=CC=CC=C3</chem>
<p>ZINC68713803</p>  <chem>CC(C)N1C(=O)NC(=O)C1=NC(=O)C2=CC=CC=C2N(C1)C(=O)C3CCOC3</chem>	<p>ZINC04273915</p>  <chem>CC(C)N1C(=O)NC(=O)C1=NC(=O)C2=CC=C(OC)C=C2SC3=NC=CC=C3N</chem>
<p>ZINC23121423</p>  <chem>CCN(C1=NC(=O)NC(=O)C1=NC2=CC=CC=C2)C(=O)C3=CC=C(C)N=C3</chem>	<p>ZINC44974155</p>  <chem>CC(C)N1C(=O)NC(=O)C1=NC(=O)C2=CC=C(OC)C=C2C3=CC=C(C)N=C3C=C</chem>



<p>ZINC06221643</p> 	<p>ZINC06221644</p> 
<p>ZINC14119574</p> 	<p>ZINC06221650</p> 
<p>ZINC06221651</p> 	<p>ZINC06221653</p> 
<p>ZINC10442496</p> 	<p>ZINC08548182</p> 
<p>ZINC09244451</p> 	<p>ZINC14721420</p> 



<p>ZINC94804951</p> 	<p>ZINC94804952</p> 
<p>ZINC94804953</p> 	<p>ZINC94804954</p> 
<p>ZINC21083040</p> 	<p>ZINC21056718</p> 
<p>ZINC57329501</p> 