Virtual Screening, Molecular Docking and Dynamic Simulation of Shikimate kinase from *Mycobacterium tuberculosis* using in silico Approach

Mustafa Alhaji Isa

Department of Microbiology, Faculty of Sciences, University of Maiduguri, P.M.B. 1069, Nigeria
Corresponding author email: mustafaisa@unimaid.edu.ng

Abstract

Shikimate kinase (SK) is an enzyme that catalyzes the fifth steps in the shikimate pathway. The enzyme facilitate the transfer of phosphoryl from ATP to shikimate, to produce ADP and shikimate-3-phosphate from *Mycobacterium tuberculosis* (MTB). The 3D structure of SK bound ligands (4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid (EPE)), ADP and metals (Mg2+, Cl- and Pt+) obtained from PDB (PDB ID: 1L4U and resolution 1.8Å). The structural analysis of the SK revealed that it has a substrate or shikimate binding site (Asp34, Arg58, and Lys136) and substrate binding via amide nitrogen (Gly80). It also possessed nucleotide binding region (Gly12─Thr17), the ATP binding site (Arg117 and Arg153) and metallic ion (Mg2+) binding site (Ser16 and Asp32). All these residues mentioned above play an essential role in the catalytic activity of the SK. Therefore inhibition any of these residues serve as a stumbling block for the normal function of the enzyme. A total of eleven thousand three hundred and twenty-three (11323) compounds obtained from two public databases (Zinc Database and PubChem) capable of binding to SK with good binding affinities. These compounds further filtered for Lipinski's rule of five, drug-likeness, molecular docking analysis, and ADME and toxicity analysis. Three compounds with minimum binding energies─ PubChem15478 (─11.75 kcal/mol), ZINC02838601 (─11.52 kcal/mol), and ZINC11790367 (─9.88 kcal/mol) ─were selected and used for the MD simulation analysis. Also, MD simulation of the SK bound to EPE, ADP, and Mg2+ were carried out to compare their stabilities with the selected protein-ligand complexes. The result showed that the two compounds (ZINC11790367 and PubChem15478) formed stable and rigid complexes comparable to the bound ligand and the cofactors during the 50ns MD simulation. Therefore, it concluded that the above mentioned two compounds capable of inhibiting SK considered as prospective drugs for MTB after successful experimental validation.

Key Word: SK, Docking, Pharmacokinetic, PDB, and MD simulation

Introduction

Shikimate pathway plays an essential role in the biosynthesis of aromatic amino acid and metabolites in bacteria, fungi, algae, plant, and apicomplexan parasites, but it is absent in mammals. This pathway is a seven-step reaction, which yields chorismate. Chorismate is used as a prerequisite for the synthesis of amino acids and other aromatic compounds including mycobactin and ubiquinone in *Mycobacterium tuberculosis* (Zhang *et al.*, 2005). Thus, the pathway offers an excellent target for the development of new antimycobacterial agent. Shikimate kinase (SK) is the fifth enzyme in the shikimate pathway that catalyzes the transfer of phosphoryl from ATP to shikimate, to produce ADP and shikimate-3-phosphate from *Mycobacterium tuberculosis* (MTB). The protein is encoded by aroK, which plays a vital role in the survival of MTB (Parish and Stoker, 2002). The aroK is made up of 531 base pair (bp) nucleotide, which gives rise to 18.58 kDa protein with 176 residues (Oliveira *et al.*, 2001). Several studies reported the structure and function of the SK in MTB, and gene knockout has shown to plays an essential role for the survival of the pathogen (Parish and Stoker, 2002; Filgueira *et al.*, 2002). That why SK is regarded as an attractive target for drugs especially those inhibitors capable of binding to the substrate (shikimate) and the ATP-binding site, which probably interfere with shikimate pathway (Ducati *et al.*, 2007). The enzyme belonged to the family of nucleoside monophosphate and composed of three domains. The CORE domain (which provides the binding site of the nucleotide), LID domain, and nucleoside (NMP) binding domain. NMP binding domain also known as shikimate binding domain due to its ability to recognize and binds to shikimate. The LID domain consisted of residues between Gly112─Asp124, while NMP or shikimate binding domain had residues between Thr33─Glu61. In addition to these domains, the enzyme has P-loop or Walker A-motif, which consisted of the residues between Gly9─Ser16. The P-loop forms the binding pocket of ATP and ADP, due to their interactions with β-phosphate and γ-phosphate respectively (Dhaliwal *et al.*, 2004: Krell *et al.*, 2001). Also, ATP/ADP interacts with an adenine-binding loop (Val148─Pro155). Lastly, it has Walker-B motif (Val75─Gly80), which lies close to the interface between ATP-binding sites and shikimate (Pereira *et al.*, 2004). Similar to other types of nucleoside binding domain, Shikimate kinase from MTB presented significant conformational changes when binds to the substrate or shikimate binding site with the wide shifts from the two domains (shikimate binding and LID domains). The NMP or substrate binding domain composed of α-helices α2, α3, and N-terminal α4 regions next to β2 strand. The Arg58 and Arg136 interact with the carboxyl group of shikimate. Also, Glu34 and Gly80 interact with a 3-hydroxyl group of the shikimate, while a 2-hydroxyl group of shikimate interacts with Glu34 (Dhaliwal *et al.*, 2004). ADP and Mg^{2+} interact with the amino acids

located in the P-loop, LID domain, and adenine-binding loop. Mg^{2+} binds explicitly to β-phosphate, oxygen, ADP, Ser16 of the P-loop and four water molecules.

Methods

Protein Preparation and Structure Refinement

The crystal structure of shikimate kinase (SK) (PDB ID: 1L4U, resolution 1.8 Å) complexed with 4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid, ADP and Metals ions (Mg^{2+}, Cl^{-} and Pt^{+}), retrieved from PDB (Berman *et al.*, 2000). The 3D structure of the enzyme prepared to provides maximum reliability. The bound 4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid, ADP, Mg^{2+}, Cl^{-}, and Pt^{+} were removed to cleaned the strcuture. All the missing , atoms, hydrogen, loops were identified and fixed via structure optimization and energy minimization using programs implemented in Chimera (Pettersen *et al.*, 2004) SwissPDViewer (Johansson *et al.*, 2012) and Chiron energy minimization and refinement tool (Ramachandran *et al*., 2011).

Selection of the Ligands

The ligands with desirable therapeutic properties (antibacterial activity) against SK was identified through virtual screening using Zinc and PubChem databases. The SK bound to 4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid submitted to RASPD tool. A total of eleven thousand three hundred and twenty-three compounds (11323) with good binding energy was obtained and further validated using PyRx tool. The compounds with better binding energies after validation checked for physicochemical properties (Lipinski rule of five) which include: Molecular weight, partition coefficient (logP), hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA) via DataWarrior tool program, to determine the one with desirable antibacterial properties. The selected compounds used for the molecular docking studies.

Docking Studies

Docking studies were performed to ascertain the free binding energy of the protein-ligand complex via an AutoDock4.2 tool (Morris *et al.*, 1998). All the ligands with desirable physicochemical properties were selected and subjected to docking analysis. During the analysis, both the protein and the ligand converted to PDBQT. PDBQT is a coordinate file which contained detailed information concerning partial charges of the atom, the torsional degree of freedom and atom types of both the protein and the ligand. The ligands docked to the site of the previously bound ligands (4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid). Also, the 4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid, ADP and Metals ions (Mg2+,

Cl- and Pt+) were docked to the SK to compare their binding energies with the selected ligands. The binding energy of the protein-ligand complex was analyzed using the below formula.

$$\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elect} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}$$

Where ΔG_{bind} = estimated free binding energy

ΔG_{vdw} = sum of van der Waals energy

ΔG_{hbond} = sum of hydrogen bond and desolvation energy

ΔG_{elect} = sum of electrostatic energy

$\Delta G_{conform}$ = sum of final total internal energy

ΔG_{tor} = sum of torsional free energy

ΔG_{sol} = sum unbound system energy (Hariono *et al.*, 2016)

Protein-Ligands Complexes Analysis

The protein-ligand complex was analyzed and visualized via Pymol Molecular Graphics System (DeLano, 2002) and Ligplot+ v.1.4.5 program (Laskowski and Swindells, 2011; Wallace *et al.*, 1996). All the hydrogen bonds, hydrophobic, polar and non-polar interactions analyzed with tools mentioned above.

Pharmacokinetic Analysis

The ligands with good docking scores were selected and subjected pharmacokinetic analysis (Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET using AdmetSAR program (Cheng *et al.*, 2012), DataWarrior program and ADME/TOX tool (Lipinski *et al.*, 2001; Veber *et al.*, 2002). The pharmacokinetic properties include Blood-Brain Barrier (BBB) penetration, Human Intestinal Absorption (HIA), cytochrome P450 (CYP450 2D6) inhibitor, aqueous solubility, and plasma protein binding (PPB), mutagenicity, tumorigenicity, irritability, and reproducibility. All compounds with desirable pharmacokinetic properties subjected to molecular dynamics simulation.

Molecular Dynamic Simulation Analysis

The ligands which possessed desirable pharmacokinetic properties were subjected to the MD simulation to determine their stability with the SK using an AMBERTOOLS10 package (Case *et al.*, 2010). The Antechamber and the protonate 3D used for the addition of missing parameters and explicit hydrogen respectively. The coordinate and topology file of the

complex built via Tleap. The GAFF and ff12SB force field of the protein and the ligand respectively were assigned using the Tleap. The entire system complex neutralized with chloride ion contained in a buffer solution of an octahedral box of TIP3P. The structural artifact of the system complex removed via minimization. Then the system was heated with an initial and final temperature of 0 and 300k respectively for 0.1ns via Langevin dynamics temperature regulation. Finally, the production of the simulation was carried out at constant temperature and pressure of 300k and 1atm respectively. The analysis of the MD simulation (root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration) performed using the PTRAJ component of the AMBERTOOLS10 package. Besides, the MM-GBSA (Molecular Mechanics Generalized Born and Surface Area) analysis were performed to ascertain the binding affinity of the SK-ligand complex using the MD simulation trajectory of the last 5ns

Results and Discussion

Molecular Docking Studies

The 3D structure of SK btained from PDB (PDB ID: 1L4U and resolution 1.8Å). The structure was prepared to ensure high-quality 3D structure. The bound ligands (4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (EPE)), ADP and metals (Mg2+, Cl- and Pt+) removed. Based on the structural analysis of the SK, revealed that it has a substrate or shikimate binding site (Asp34, Arg58, and Lys136) and substrate binding via amide nitrogen (Gly80). It also possessed nucleotide binding region (Gly12─Thr17), the ATP binding site (Arg117 and Arg153) and metallic ion (Mg^{2+}) binding site (Ser16 and Asp32). The submission of the SK into the Ligand Contact Tool (LCT) revealed the residues interacted with ADP are Leu10, Pro11, Gly12, Ser13, Gly14, Lys15, Ser16, Thr17, Arg110, Arg117, Arg153, Asn154, and Pro155. It also unravels the residues interacted with five molecules of Cl^{-} (Met133, Arg66, Ala90, Ala91, Leu10, Pro11, Lys15, and Gly80), Pt^{+} (Met133) and EPE (Arg130). All these residues mentioned above play an essential role in the catalytic activity of the SK. Therefore inhibition any of these residues serve as a stumbling block for the normal function of the enzyme. The prepared structure of SK was used for virtual screening. A total of 11323 compounds was obtained from two public databases (Zinc Database and PubChem) capable of binding to SK with good binding affinities. These compounds were further filtered for physicochemical properties or Lipinski's rule of five (molecular weight ≤ 500 Da, number of hydrogen-bond acceptors ≤ 10, number of hydrogen bond donors ≤ 5, and logP ≤ 5), to select

compounds with desirable properties (Table 1). The selected compounds were subjected to molecular docking analysis to calculate the free binding energies based on the final intermolecular energy, hydrogen and desolvation energies, van der Waals, electrostatic energy and torsional free energy. The compounds were docked to the binding site of the previous ligand (EPE). Nineteen (19) compounds presented good binding affinities ranged between −11.75 and −8.49kcal/mol, less than the binding free energies of EPE (−8.02kcal/mol) and Mg2+ (−0.98kcal/mol), although it's greater than the binding energies of ADP (−12.44kcal/mol) (Fig. 1). The amino acids residues involved in hydrogen bonds with the different ligands include Gly12, Gly14, Lys15, Ser16, Thr17, Asp32, Asp34, Arg41, Arg58, Gly80, Gly81, Arg110, Arg117 and Arg136 (Table 5.11). PubChem15478 had the best binding energy of −11.75kcal/mol. The compound formed five hydrogen bonds through accepting electrons from hydroxyl group of Ser16 (distance = 3.25Å), a carboxylic group of Asp32 (distance = 2.74Å), an amino group of Arg58 (distance = 2.82Å and 2.99Å) and Arg117 (distance = 3.21Å) (Table 2). It also presented hydrophobic interactions with a carboxylic group of Glu61 and Asp34, and the hydrogen atom of Gly80 (Fig. 2a). Ser16 is among the residues formed the nucleotide binding site in P-loop region or Walker A-motif, Arg117 involved in the ATP binding site, Arg58 and Asp34 formed the shikimate binding site and Gly80 involved in the substrate or shikimate binding site via amide nitrogen. Interactions with all these residues by the ligands would block the catalytic activity of the enzyme. Arg117 which involved in the ATP binding site of SK, exhibited hydrogen bonds with 15 compounds (PubChem15478, ZINC8442189, ZINC633842, ZINC8442094, ZINC8442075, ZINC634006, ZINC633992, PubChem221071, ZINC16192643, ZINC09191993, ZINC633978, ZINC11881196, PubChem24360, PubChem13844301and PubChem9997719). Interactions of these compounds with Arg117 would block the transfer of phosphoryl from ATP to shikimate, which requires the formation of shikimate-3-phosphate. This interaction would inhibit the catalytic activity of the enzyme. Similarly, eight compounds (ZINC8442189, ZINC11790367, ZINC8442094, ZINC634006, ZINC633992, ZINC09191993, ZINC633978, and ZINC11881196) interacted and formed hydrogen bonds with Asp34. This residue involved in the substrate or shikimate binding site and binding of these ligands to the residue impede the catalytic role of the enzyme. Lys15 interacted and formed hydrogen bonds with eight compounds (ZINC633842, ZINC8442094, ZINC02843658, ZINC634006, ZINC633992, PubChem221071, ZINC11881196 and PubChem13844301). This residue involved in the nucleotide binding site of P-loop region or Walker A-motif. Also, Thr17 presented hydrogen bonds with seven compounds (ZINC11790367, ZINC633842, PubChem265237,

ZINC02843658, ZINC8442075, ZINC634006 and ZINC633992) via its hydroxyl group. Thr17 involved in nucleotide binding domain and play an important role in the catalytic activity of the SK. Similarly, seven ligands (ZINC8442189, ZINC02843658, PubChem221071, ZINC16192643, ZINC633978, PubChem24360 and PubChem13844301) formed hydrogen bonds by donating electrons to the hydroxyl group of Arg136, which plays an important role in the substrate or shikimate binding site. Six compounds (ZINC8442189, ZINC8442094, ZINC634006, ZINC16192643, PubChem13844301, and PubChem9997719) interacted and formed hydrogen bonds with hydrogen group of Gly12, which involved in the nucleotide binding site of P-loop region. Five compounds (ZINC633842, ZINC634006, ZINC633992, ZINC16192643, and ZINC11881196) presented hydrogen bonds with hydrogen of Gly80, which involve in the substrate binding site via amide nitrogen. Ser16 and Arg58 exhibited hydrogen bonds with 3 (PubChem15478, ZINC02843658, and ZINC8442075) and 2 (PubChem15478 and PubChem24360) different compounds respectively. Ser16 involved in nucleotide binding site of P-loop region, while Arg58 form part of substrate binding site. All these residues are necessary for the catalytic activity of the enzyme. Therefore, the interaction of these ligands with the residues above impedes the catalytic activity of the enzyme (Figure aa-s).

Table 1: Molecular Properties and Drug-likeness of the ligands interacted with SK

S/No.	Zinc Code/ PubChem ID	Molecular Weight (≤500)	cLogP (≤5)	H-bond Acceptors (≤10)	H-bond Donors (≤5)	Drug-likeness
1	ZINC02838601	496.57	3.53	10	1	0.02
2	ZINC8442189	454.48	-0.46	10	2	4.33
3	ZINC11790367	454.48	-0.46	10	2	4.33
4	ZINC633842	494.37	3.74	7	2	0.42
5	ZINC8442094	486.64	3.53	8	2	6.15
6	ZINC02843658	499.64	3.67	9	2	1.48
7	ZINC8442075	470.57	2.40	9	2	-3.43
8	ZINC634006	437.56	3.28	7	2	4.59
9	ZINC633992	421.93	2.38	7	2	6.12
10	ZINC16192643	477.58	2.87	8	2	1.96
11	ZINC09191993	494.61	2.40	9	1	9.17
12	ZINC633978	469.52	2.04	10	2	1.86
13	ZINC11881196	497.66	1.62	6	2	4.19
14	PubChem15478	390.51	2.07	5	3	-0.33
15	PubChem265237	470.60	2.49	6	2	1.68
16	PubChem221071	246.30	1.84	3	0	1.81
17	PubChem24360	348.35	1.18	6	1	5.32
18	PubChem13844301	342.39	3.81	5	1	-2.31
19	PubChem9997719	346.33	2.29	7	2	-0.07

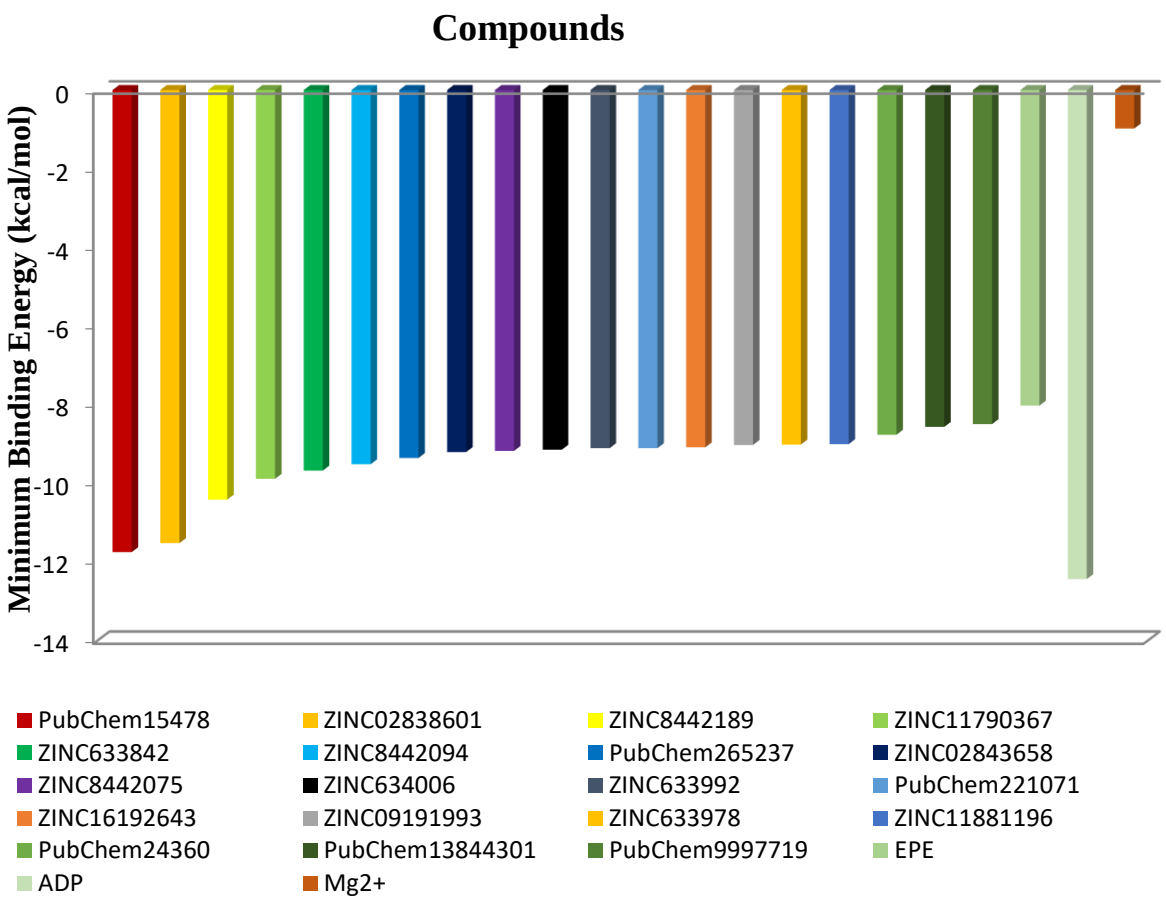
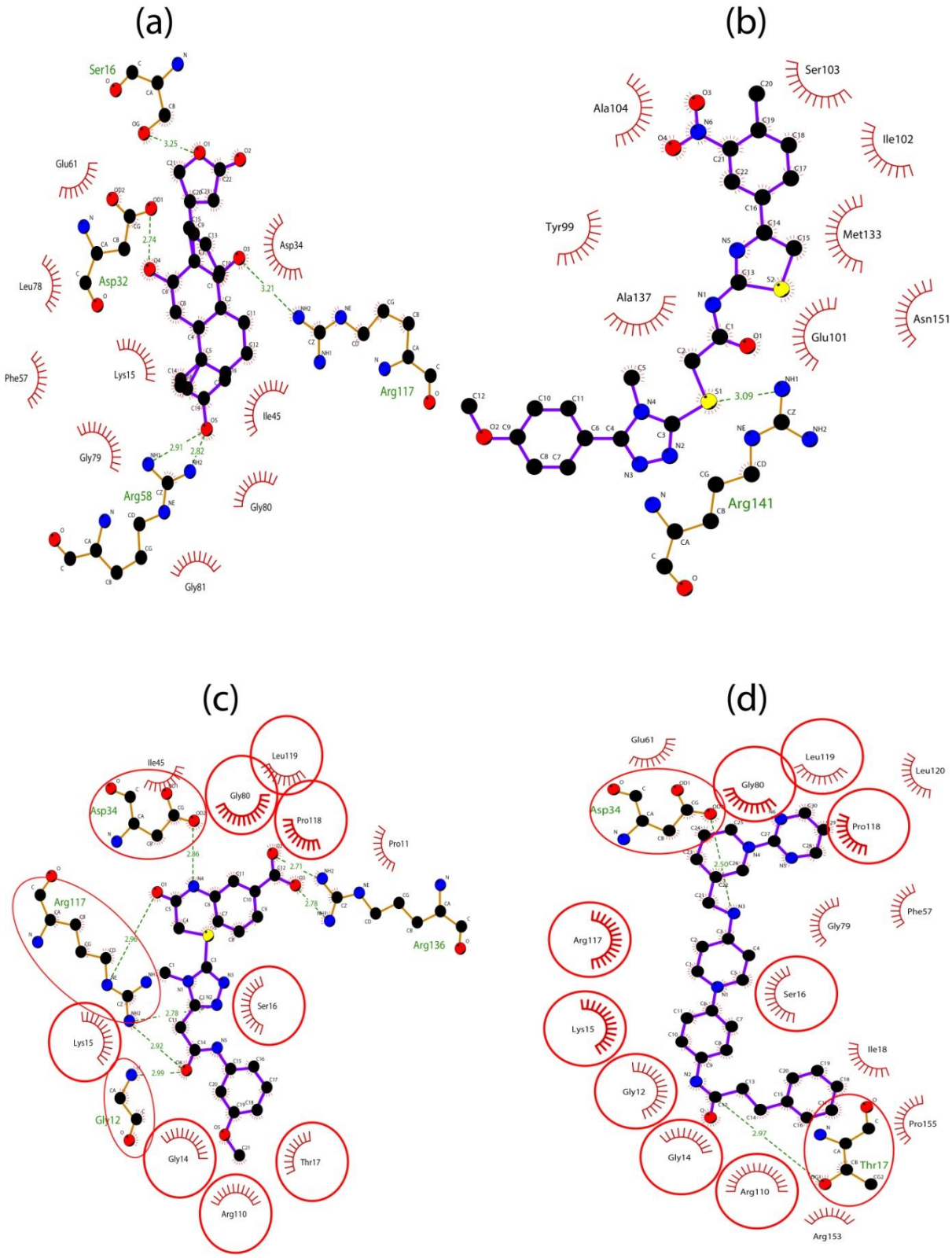


Figure 1: Docking scores of ligands with good binding energies interacted with SK

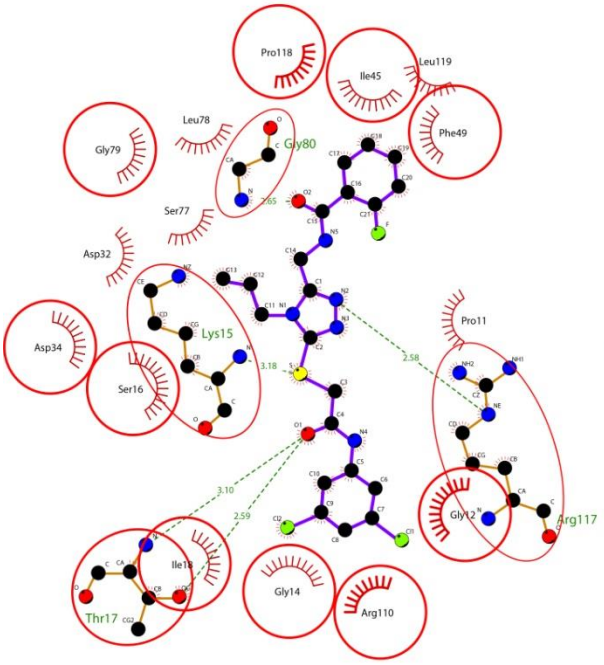
Table 2: Binding energies and hydrogen bonds interaction of SK with selected ligands

S/No.	Zinc Code/ PubChem ID	Minimum Free Energy of Binding (kcal/mol)	Hydrogen bond Interaction	Hydrogen bond distance (Å)
1	PubChem15478	−11.75	Ser16 Asp32 Arg58 Arg58 Arg117	3.25 2.74 2.82 2.91 3.21
2	ZINC02838601	−11.52	Arg41	3.09
3	ZINC8442189	−10.41	Gly12 Asp34 Arg117 Arg117 Arg117 Arg136 Arg136	2.99 2.86 3.10 2.78 2.92 2.78 2.71
4	ZINC11790367	−9.88	Thr17 Asp34	2.97 2.50
5	ZINC633842	−9.67	Lys15 Thr17 Thr17 Gly80 Arg117	3.18 2.59 3.10 2.65 2.58
6	ZINC8442094	−9.51	Thr17 Lys15 Lys15 Asp34 Gly14 Gly12 Gly12 Arg117	3.01 3.22 2.85 3.29 2.80 3.15 3.19 3.15
7	PubChem265237	−9.35	Gly12 Gly14 Lys15 Thr17	2.96 2.69 3.21 2.54
8	ZINC02843658	−9.20	Lys15 Ser16 Thr17 Arg136 Arg136 Arg136	2.84 3.23 3.30 3.14 2.98 3.14
9	ZINC8442075	−9.17	Ser16 Th17 Th17 Arg110 Arg117 Arg117	2.88 3.02 2.65 3.08 3.01 3.27
10	ZINC634006	−9.14	Gly12 Lys15 Lys15	3.35 2.99 2.45

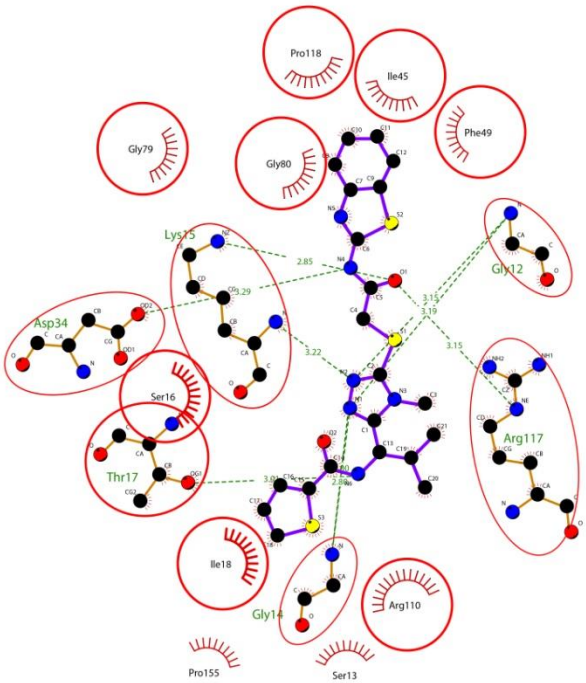
			Thr17 Thr17 Asp34 Gly80 Arg117 Arg117	2.64 2.87 2.66 2.85 3.24 3.28
11	ZINC633992	−9.10	Lys15 Thr17 Thr17 Asp34 Gly80 Arg117	3.32 3.08 2.62 2.88 2.72 2.98
12	PubChem221071	−9.10	Lys15 Arg117 Arg117 Arg136	3.21 3.07 3.14 2.67
13	ZINC16192643	−9.08	Gly12 Gly14 Gly80 Gly80 Arg117 Arg136	3.26 2.59 3.16 3.34 3.17 3.16
14	ZINC09191993	−9.02	Thr17 Asp34 Arg117 Arg117 Arg117	3.04 2.90 3.05 3.26 2.96
15	ZINC633978	−9.01	Asp34 Arg117 Arg136 Arg136	2.64 3.19 2.93 2.62
16	ZINC11881196	-9.00	Lys15 Asp34 Gly80 Arg117	2.93 3.03 3.02 3.27
17	PubChem24360	−8.76	Arg58 Arg58 Arg117 Arg136 Arg136	3.19 2.90 3.05 3.24 2.56
18	PubChem13844301	−8.56	Gly12 Lys15 Arg117 Arg136	3.00 3.22 2.91 3.27
19	PubChem9997719	−8.49	Gly12 Asp32 Gly81 Arg117	3.00 2.76 3.18 3.00


(a)
Ser16
Glu61
Asp32
Leu78
Phe57
Lys15
Asp34
Ile45
Gly79
Arg58
Gly80
Gly81
Arg117
(b)
Ser103
Ala104
Ile102
Tyr99
Met133
Ala137
Asn151
Glu101
Arg141
(c)
Ile45
Leu119
Gly80
Asp34
Pro118
Pro11
Arg117
Arg136
Ser16
Lys15
Gly12
Gly14
Thr17
Arg110
(d)
Glu61
Leu119
Leu120
Gly80
Asp34
Pro118
Phe57
Gly79
Arg117
Ser16
Lys15
Ile18
Gly12
Pro155
Gly14
Thr17
Arg110
Arg153

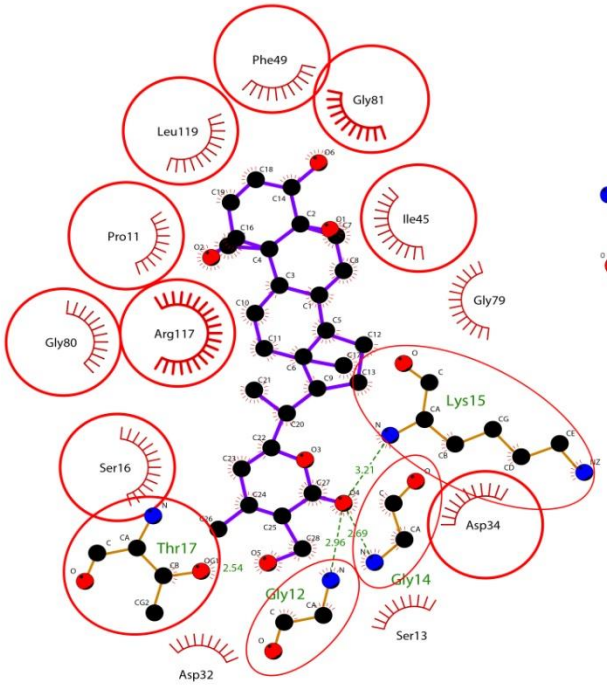
(e)


Pro118
Ile45
Leu119
Phe49
Gly79
Leu78
Gly80
Ser77
Asp32
Lys15
Asp34
Ser16
Pro11
Gly12
Arg117
Ile18
Thr17
Gly14
Arg110

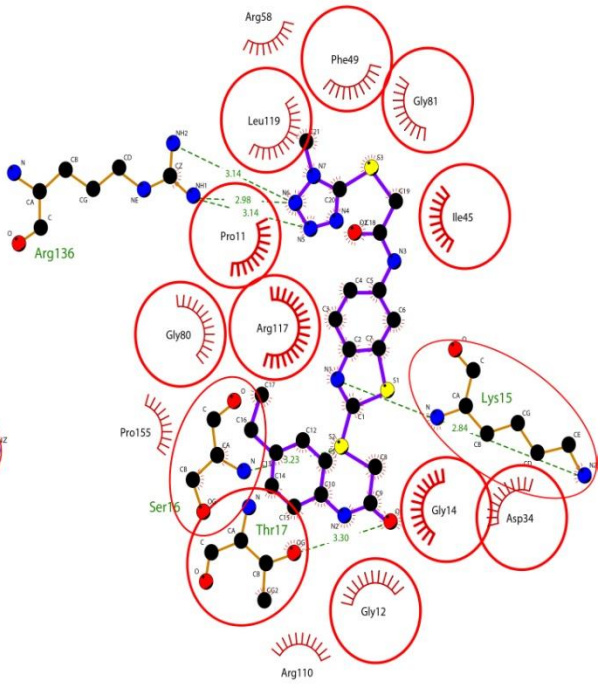
(f)

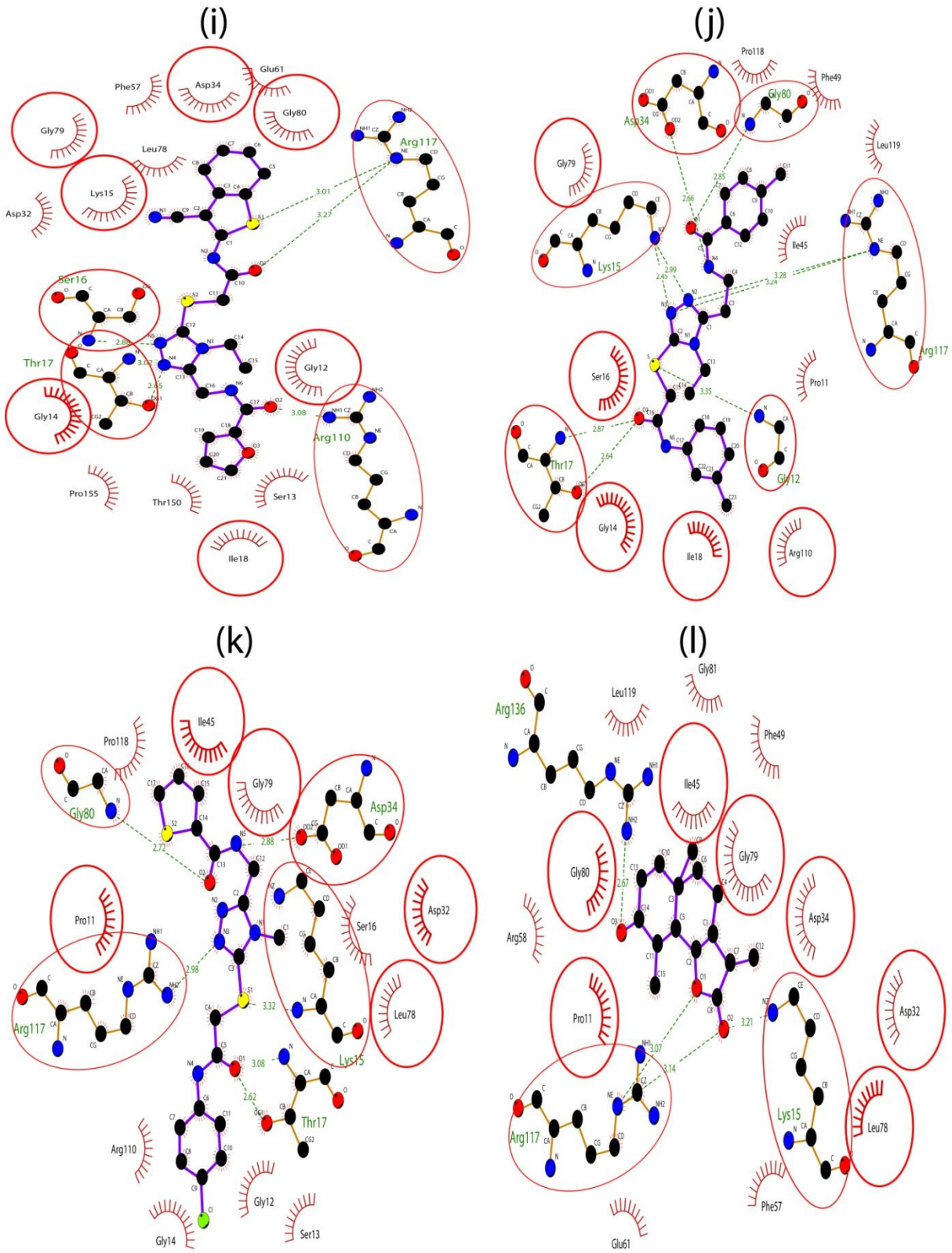

Pro118
Ile45
Phe49
Gly79
Gly80
Lys15
Gly12
Asp34
Ser16
Arg117
Thr17
Ile18
Gly14
Arg110
Pro155
Ser13

(g)

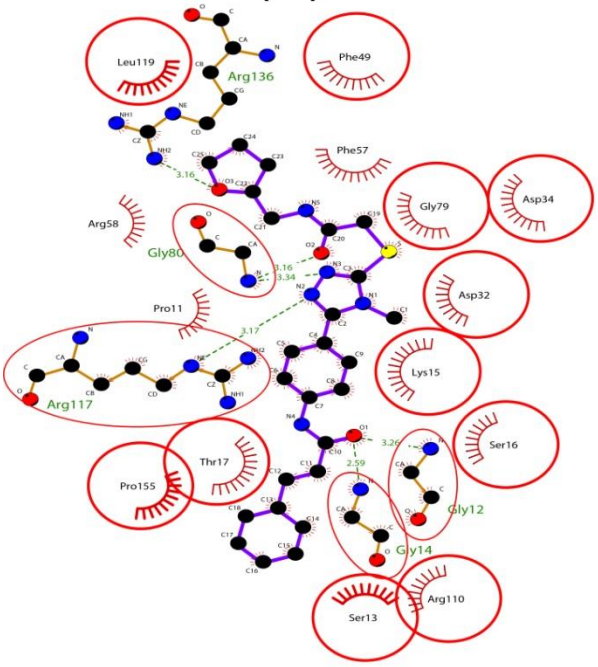

Phe49
Gly81
Leu119
Ile45
Pro11
Gly79
Gly80
Arg117
Lys15
Ser16
Asp34
Thr17
Gly14
Gly12
Ser13
Asp32

(h)

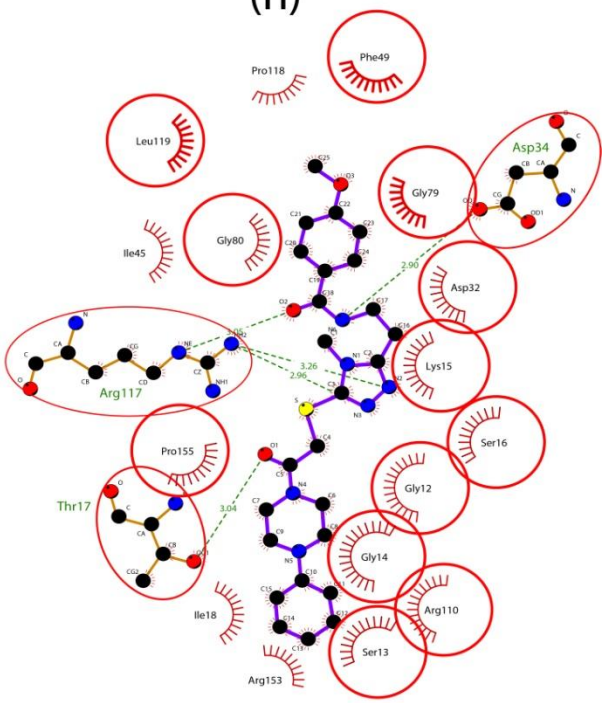

Arg58
Phe49
Gly81
Leu119
Ile45
Arg136
Pro11
Gly80
Arg117
Lys15
Pro155
Ser16
Thr17
Gly14
Asp34
Gly12
Arg110


(i)
Phe57
Asp34
Glu61
Gly80
Gly79
Leu78
Lys15
Asp32
Arg117
Ser16
Thr17
Gly14
Gly12
Arg110
Pro155
Thr150
Ser13
Ile18
(j)
Pro118
Phe49
Gly80
Asp34
Gly79
Leu119
Lys15
Ile45
Arg117
Ser16
Pro11
Thr17
Gly12
Gly14
Ile18
Arg110
(k)
Ile45
Pro118
Gly79
Gly80
Asp34
Pro11
Ser16
Asp32
Arg117
Leu78
Lys15
Thr17
Arg110
Gly12
Gly14
Ser13
(l)
Gly81
Leu119
Arg136
Phe49
Ile45
Gly80
Gly79
Asp34
Arg58
Pro11
Asp32
Arg117
Lys15
Leu78
Phe57
Glu61

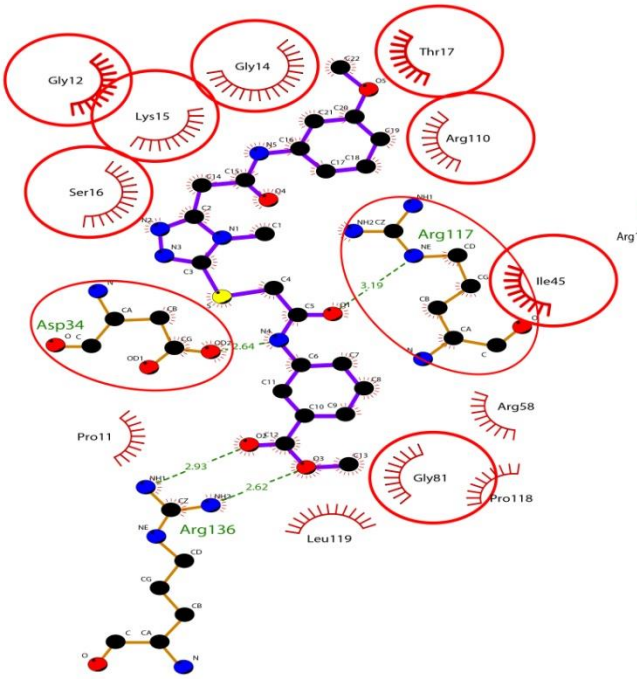
(m)


Leu119
Arg136
Phe49
Phe57
Gly79
Asp34
Arg58
Gly80
Asp32
Pro11
Lys15
Arg117
Ser16
Thr17
Pro155
Gly12
Gly14
Arg110
Ser13

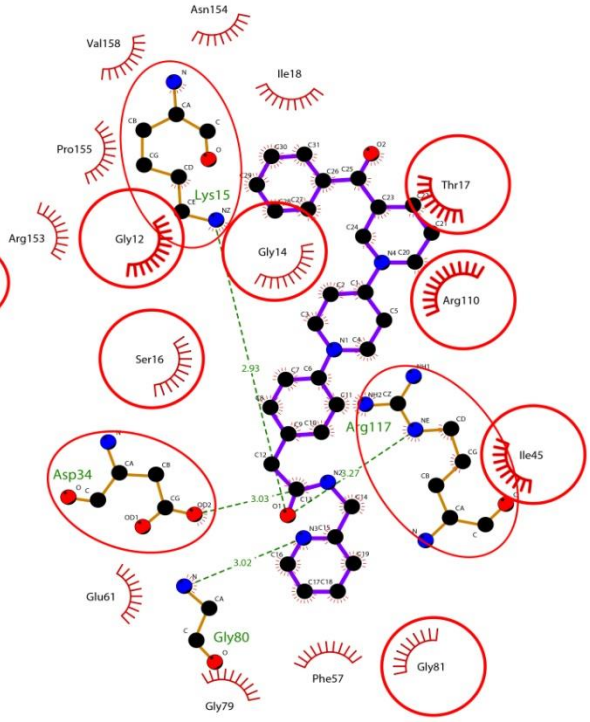
(n)


Pro118
Phe49
Leu119
Asp34
Gly79
Ile45
Gly80
Asp32
Lys15
Arg117
Ser16
Pro155
Gly12
Thr17
Gly14
Arg110
Ile18
Ser13
Arg153

(o)


Gly12
Gly14
Thr17
Lys15
Arg110
Ser16
Arg117
Ile45
Asp34
Arg58
Pro11
Gly81
Pro118
Arg136
Leu119

(p)


Asn154
Val158
Ile18
Pro155
Lys15
Thr17
Arg153
Gly12
Gly14
Arg110
Ser16
Arg117
Ile45
Asp34
Glu61
Gly80
Phe57
Gly81
Gly79

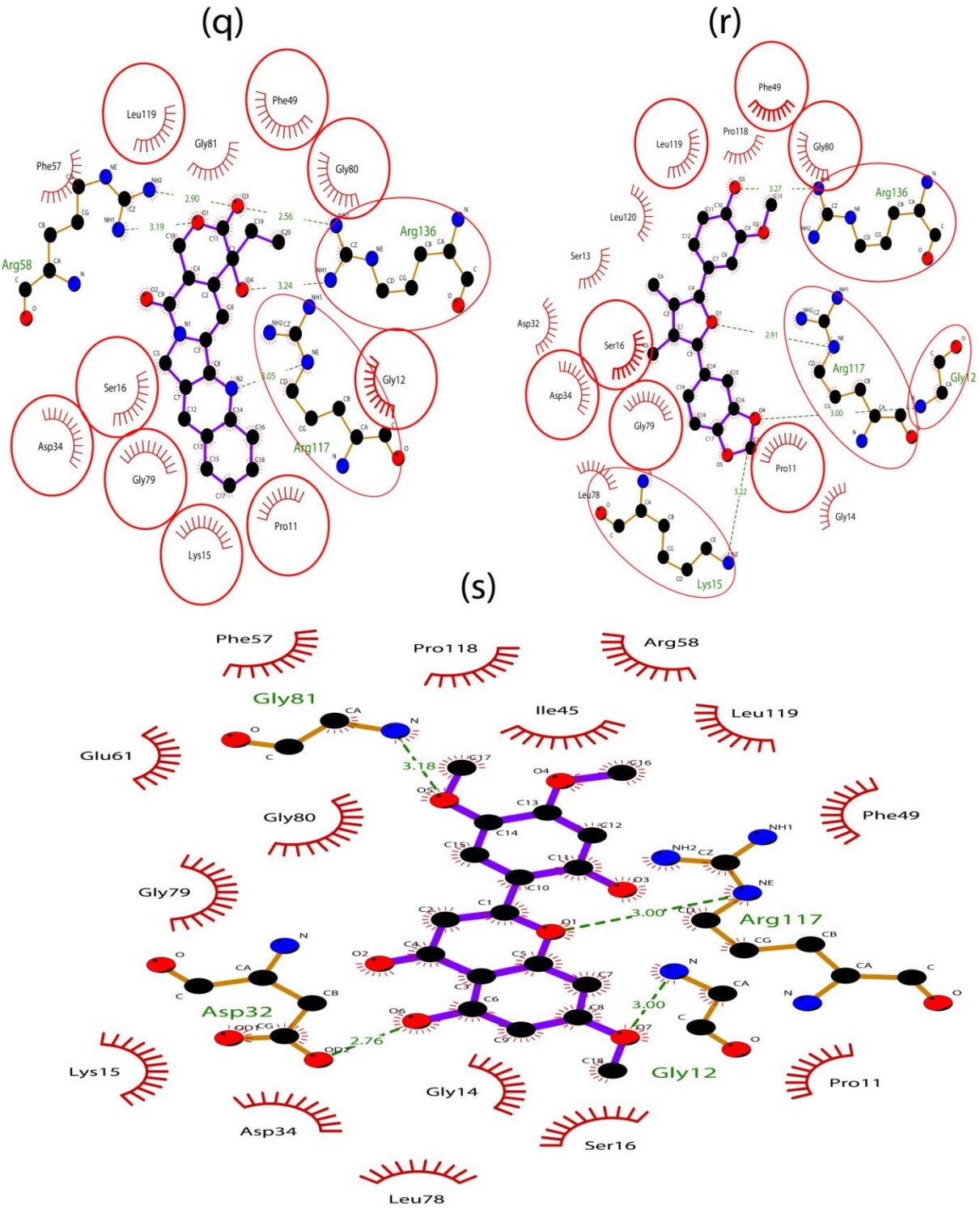


Figure 2: Hydrogen bonds and hydrophobic interactions between the SK and the selected ligands (a) PubChem15478 (b) ZINC02838601 (c) ZINC8442189 (d) ZINC11790367 (e) ZINC633842 (f) ZINC8442094 (g) PubChem265237 (h) ZINC02843658 (i) ZINC8442075 (j) ZINC634006 (k) ZINC633992 (l) PubChem221071 (m) ZINC16192643 (n) ZINC09191993 (o) ZINC633978 (p) ZINC11881196 (q) PubChem24360 (r) PubChem13844301(s) PubChem9997719

Pharmacokinetic Analysis

The molecular docking studies of the selected ligands against the SK revealed that 19 compounds had good binding energies and could serve as potential inhibitors of SK. These compounds were further screened for pharmacokinetic properties (absorption, distribution, metabolism, excretion, and toxicity analyses) to select compounds with desirable properties. The majority of ligands fail to reach clinical trial due to undesirable pharmacokinetics properties. The predicted properties include human intestinal absorption (HIA), cytochrome P450 (CYP450 2D6) inhibition, aqueous solubility, plasma protein binding, Ames test results, carcinogenicity, mutagenicity, tumorigenicity, reproducibility, and irritability (Table 3). The human intestinal absorption of all the ligands was positive except ZINC8442189, where it has less than 30% rate of absorption. Similarly, the blood-brain barrier of all compounds was positive except ZINC11881196, PubChem24360, PubChem13844301, and PubChem9997719, due to their inability to cross the brain barrier. The predicted CYP450 2D6 inhibitory activities of all the compounds revealed that five compounds (ZINC8442075, ZINC634006, ZINC16192643, ZINC09191993, and PubChem13844301) were inhibitors of cytochrome. ZINC8442075, ZINC634006, ZINC16192643, and ZINC09191993 had AC50 values of >57 μM based on the model of Cheng *et al.* (2012), whereas PubChem13844301 had a PubChem activity score of >40, made them inhibitors of cytochrome. The toxicity analysis revealed that two compounds (ZINC8442094 and PubChem265237) were toxics. ZINC8442094 predicted to undergoes mutation, although at a slow rate, while PubChem265237 was found to be carcinogenic and probably can cause irritation. Therefore, out of the 19 compounds screened for the pharmacokinetics properties, 11 compounds (ZINC8442189, ZINC11881196, PubChem24360, PubChem13844301, PubChem9997719, ZINC8442094, PubChem265237, ZINC8442075, ZINC634006, ZINC16192643, and ZINC09191993) had undesirable pharmacokinetics properties. The remaining 8 compounds (ZINC02838601, ZINC11790367, ZINC633842, ZINC8442094, ZINC633992, ZINC633978, PubChem15478, and PubChem221071) possessed all the desirable features and considered as novel inhibitors of SK (Table 3).

Table 3: ADME and Toxicity Analyses

S/No.	Zinc/PubChem Code	HIA	BBB	CYP450 2D6 Inhibitor	PPB (%)	Aqueous Solubility	Carcinogens	Mutagenic	Tumorigenic	Reproducibility	Irritant
1	ZINC02838601	+	+	Non-inhibitor	98.0	-5.34	Non-carcinogens	none	none	none	none
2	ZINC8442189	-	+	Non-inhibitor	88.0	-3.26	Non-carcinogens	none	none	none	none
3	ZINC11790367	+	+	Non-inhibitor	69.0	-9.93	Non-carcinogens	none	none	none	none
4	ZINC633842	+	+	Non-inhibitor	97.0	-5.14	Non-carcinogens	none	none	none	none
5	ZINC8442094	+	+	Non-inhibitor	93.0	-5.25	Non-carcinogens	none	none	none	none
6	ZINC02843658	+	+	Non-inhibitor	100	-5.38	Non-carcinogens	low	none	none	none
7	ZINC8442075	+	+	Inhibitor	87.0	-4.45	Non-carcinogens	none	none	none	none
8	ZINC634006	+	+	Inhibitor	98.0	-4.42	Non-carcinogens	none	none	none	none
9	ZINC633992	+	+	Non-inhibitor	99.0	-3.95	Non-carcinogens	none	none	none	none
10	ZINC16192643	+	+	Inhibitor	97.0	-4.30	Non-carcinogens	none	none	none	none
11	ZINC09191993	+	+	Inhibitor	99.0	-4.20	Non-carcinogens	none	none	none	none
12	ZINC633978	+	+	Non-inhibitor	91.0	-3.48	Non-carcinogens	none	none	none	none
13	ZINC11881196	+	-	Non-inhibitor	61.0	-5.31	Non-carcinogens	none	none	none	none
14	PubChem15478	+	+	Non-inhibitor	85.0	-2.89	Non-carcinogens	none	none	none	none
15	PubChem265237	+	+	Non-inhibitor	98.0	-4.97	Carcinogens	none	none	none	low
16	PubChem221071	+	+	Non-inhibitor	87.0	-2.81	Non-carcinogens	none	none	none	none
17	PubChem24360	+	-	Non-inhibitor	79.0	-3.49	Non-carcinogens	none	none	none	none
18	PubChem13844301	+	-	Inhibitor	98.0	-4.60	Non-carcinogens	none	none	none	none
19	PubChem9997719	+	-	Non-inhibitor	88.0	-3.91	Non-carcinogens	none	none	none	none

BBB = Blood-Brain Barrier, HIA = Human Intestinal Absorption, PPB = Plasma Protein Binding, Aqueous Solubility = Insoluble < -10 < Poorly soluble < -6 < Moderately soluble < -4 < Soluble < -2 < Very soluble < 0 < Highly soluble

Molecular Dynamic Simulation Analysis

The result of pharmacokinetics analysis revealed that 8 compounds (ZINC02838601, ZINC11790367, ZINC633842, ZINC8442094, ZINC633992, ZINC633978, PubChem15478 and PubChem221071) possessed all the desirable properties and thereby satisfied all the ADMET criteria. Three compounds with minimum binding energies─ PubChem15478 (─11.75 kcal/mol), ZINC02838601 (─11.52 kcal/mol), and ZINC11790367 (─9.88 kcal/mol) ─were selected and used for the MD simulation analysis, to determine their dynamic behaviors inside solvent. Also, MD simulation of the SK bound to EPE, ADP, and Mg^{2+} were carried out to compare their stabilities with the selected protein-ligand complexes. The six complexes (SK─ PubChem15478, SK─ZINC02838601, SK─ZINC11790367, SK─EPE, SK─ADP and SK─Mg^{2+}) were simulated differently. The stabilities of all the protein-ligand complexes were determined by examining the RMSD during the 50ns MD simulation (Fig. 3). RMSF was used to determine the deviation of the ligands around the binding pocket of the protein as well as the movement of residues around their mean position in the complexes. Also, the radius of gyration was checked to determine the compactness of each complex based on how folded or unfolded their structures were. The RMSD of the SK─PubChem15478 complex equilibrated and oscillated at 10ns and remained steady throughout the 50ns MD simulation, with a mean value of 3.2699±0.00778Å, a maximum value of 4.02Å, and a minimum value of 1.13Å. The means value is greater than the mean RMSD values of SK─EPE (3.0271±0.01529Å), SK─ADP (2.866±0.01349Å) and SK─Mg^{2+} (1.7066±0.00670Å). The equilibration of SK─PubChem15478 complex at 3Å during the 50ns MD simulation indicated that the complex is stable and rigid, which probably due to the interaction of the ligand with the flexible loop region of SK. Similarly, SK─ZINC11790367 complex equilibrated and oscillated with a mean value of 3.3243±0.1184Å, a maximum value of 3.96Å, and a minimum value 1.17Å. This mean value is greater than the mean RMSD values of SK─EPE (3.0271±0.01529Å), SK─ADP (2.866±0.01349Å) and SK─Mg^{2+} (1.7066±0.00670Å). The equilibration of the SK─ZINC11790367 complex at 3Å suggests that the complex is stable and rigid, probably due to the interaction of the bound ligand with the flexible loop region of the SK, leading to the stability and rigidity of the complex (Fig. 3). In contrast, SK─ZINC02838601 complex partially equilibrated between the RMSD values of 4─5Å throughout the 50ns MD simulation, with a mean value of 4.5985±0.1527Å, a maximum value of 6.08Å, and a minimum value of 1.43Å. This mean value is greater than the mean RMSD values of SK─EPE (3.0271±0.01529Å), SK─ADP (2.866±0.01349Å) and SK─Mg^{2+} (1.7066±0.00670Å), with a

highly statistically significant difference (P<0.05). This high RMSD value of SK─ZINC02838601 complex is probably due to the inability of the ligand to interact with the flexible region of the SK, leading to high fluctuation in the complex (Fig. 3). Figure 4 indicated the RMSF values of all the six complexes (SK─ PubChem15478, SK─ ZINC02838601, SK─ZINC11790367, SK─EPE, SK─ADP and SK─Mg^{2+}), which represents the deviation of the ligands in the binding pocket of SK as well as the movement of each residue from their initial position. In the SK─PubChem15478 complex all residues located in the N-terminus, ω-loop, C-terminus, CORE domain, LID domain and NMP domain had RMSF values <10Å, except Thr76 (Walker B─motif) which had value ≥10Å. This RMSF value of SK─PubChem15478 complex is similar to the corresponding to the RMSF values of SK─EPE, SK─ADP and SK─Mg^{2+} complexes, where they had RMSF values of <10Å in all their regions except SK─Mg^{2+} complex which had RMSF value of >10Å between residues Phe49─Asp52. Similarly, SK─ZINC11790367 complex had all it residues located in the N-terminus, ω-loop, C-terminus, CORE domain, LID domain and NMP domain with RMSF value of <10Å. The low RMSF values in both SK─PubChem15478 and SK─ZINC11790367 complexes were probably due to hydrogen and hydrophobic interactions between the ligands and the substrate binding site of the SK, leading to high stability and rigidity. This RMSF values also agree with the RMSD plot of the two complexes, where they stabilized and equilibrated with low RMSD value. In contrast to SK─ZINC02838601 complex, where all its residues located in the N-terminus, ω-loop, C-terminus, CORE domain, LID domain and NMP domain had RMSF values of ≥10Å, except residues between Leu10─Lys15, Glu101─Gly112, and Thr150─Asn151, which had RMSF values <10Å. This high RMSF values in all the regions of the complex were probably due to the failure of the ligand to form hydrogen bonds with residues involved in the substrate binding site, nucleotide binding site and ATP binding site of the SK, leading to high fluctuation of the residues in all the regions (Fig. 4). However, the main secondary structural elements of the SK and the ligands in all the complexes (SK─PubChem15478, SK─ZINC02838601, SK─ZINC11790367, SK─EPE, SK─ADP and SK─Mg^{2+}) partially deviated from their initial structures after the 50ns MD simulation, leading to low RMSF values (1─2Å). As clearly indicated in Figure 6a─f, where the final structures (after MD simulation) superposed to their respective initial structures (Before MD simulation) and showed that all the ligands bound to the position closed to the initial binding site. The radius of gyration was also checked during the 50ns MD simulation for each complex to determine the compactness based on how stably folded or unfolded the structures were. SK─PubChem15478 complex presented relatively constant values of radius of gyration

throughout the 50ns, with a mean value of 16.0080±0.00424Å, a maximum value of 16.81Å, and a minimum value of 15.70Å. These values are close to the mean radius of gyration values of SK─EPE (16.1391±0.00345Å), SK─ADP (17.6936±0.03145Å) and SK─Mg^{2+} (16.8466±0.2200Å). Similarly, SK─ZINC02838601 and SK─ZINC11790367 complexes move at relatively constant values throughout the 50ns with mean values of 15.7191±0.00195Å and 16.0116±0.00339Å, maximum values of 16.26Å and 16.37Å, and minimum values of 15.49Å and 15.73Å respectively. Therefore, all the selected ligands (PubChem15478, ZINC02838601 and ZINC11790367) formed relatively stable folded polypeptide structures (Fig. 5).

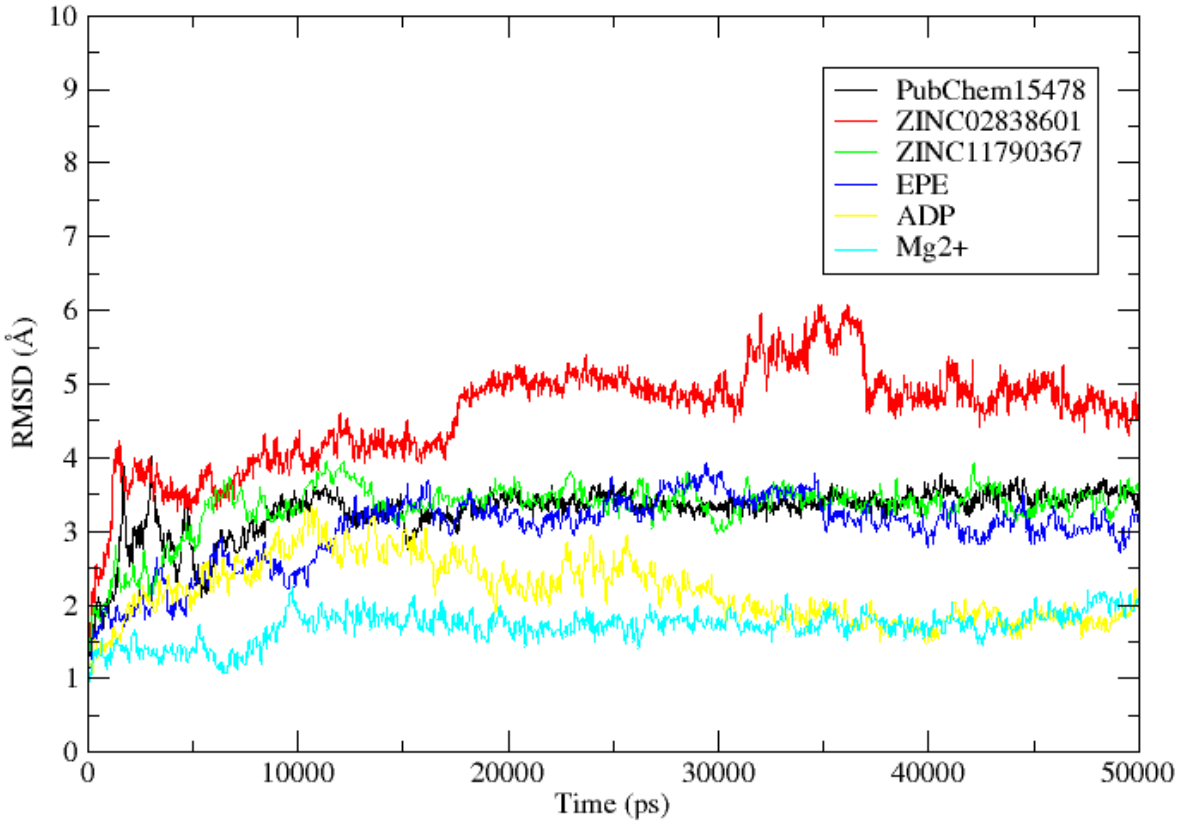


Figure 3: RMSD results for the SK─ PubChem15478, SK─ ZINC02838601, SK─ ZINC11790367, SK─EPE, SK─ADP and SK─Mg^{2+} complexes, based on data from 50-ns MD simulations

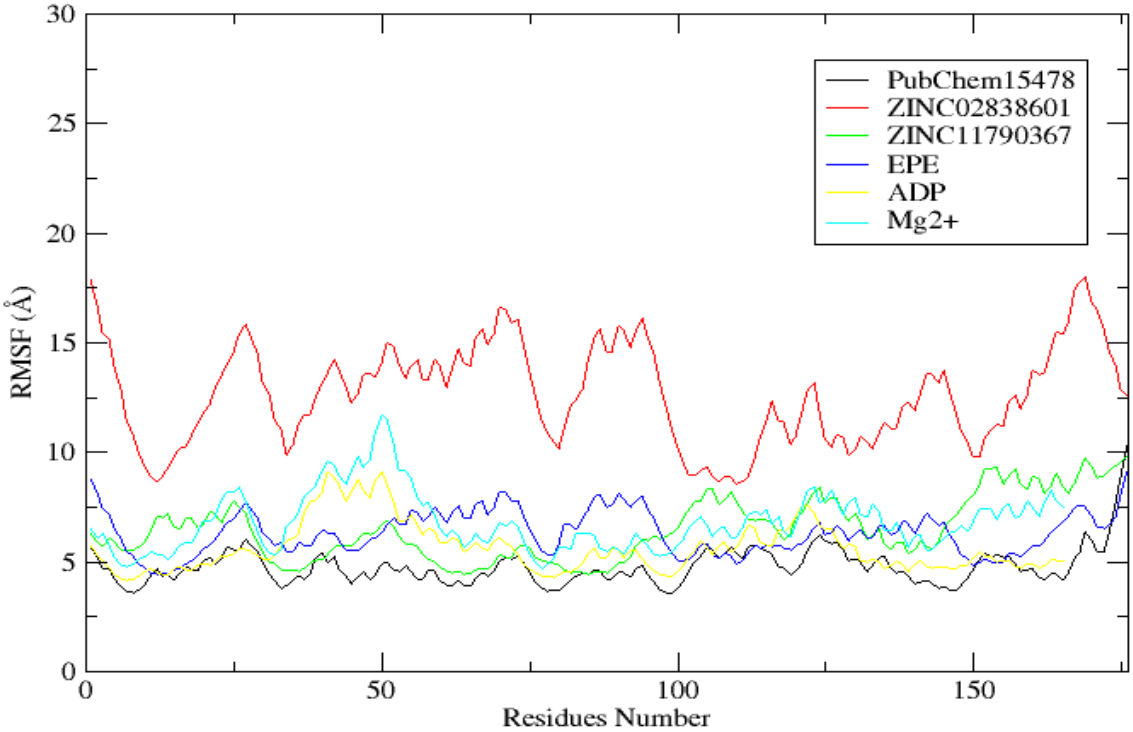


Figure 4: RMSF results for the SK–PubChem15478, SK–ZINC02838601, SK–ZINC11790367, SK–EPE, SK–ADP and SK–Mg^{2+} complexes, based on data from 50-ns MD simulations

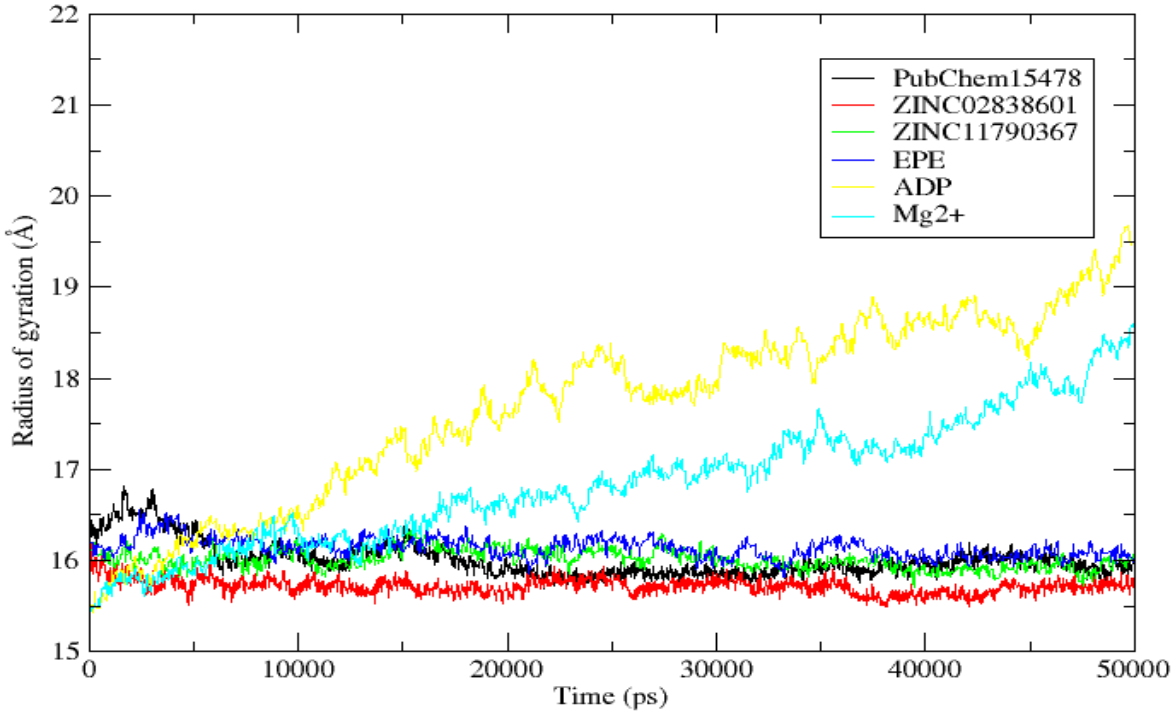


Figure 5: Radius of gyration results for the SK–PubChem15478, SK–ZINC02838601, SK–ZINC11790367, SK–EPE, SK–ADP and SK–Mg^{2+} complexes, based on data from 50-ns MD simulation

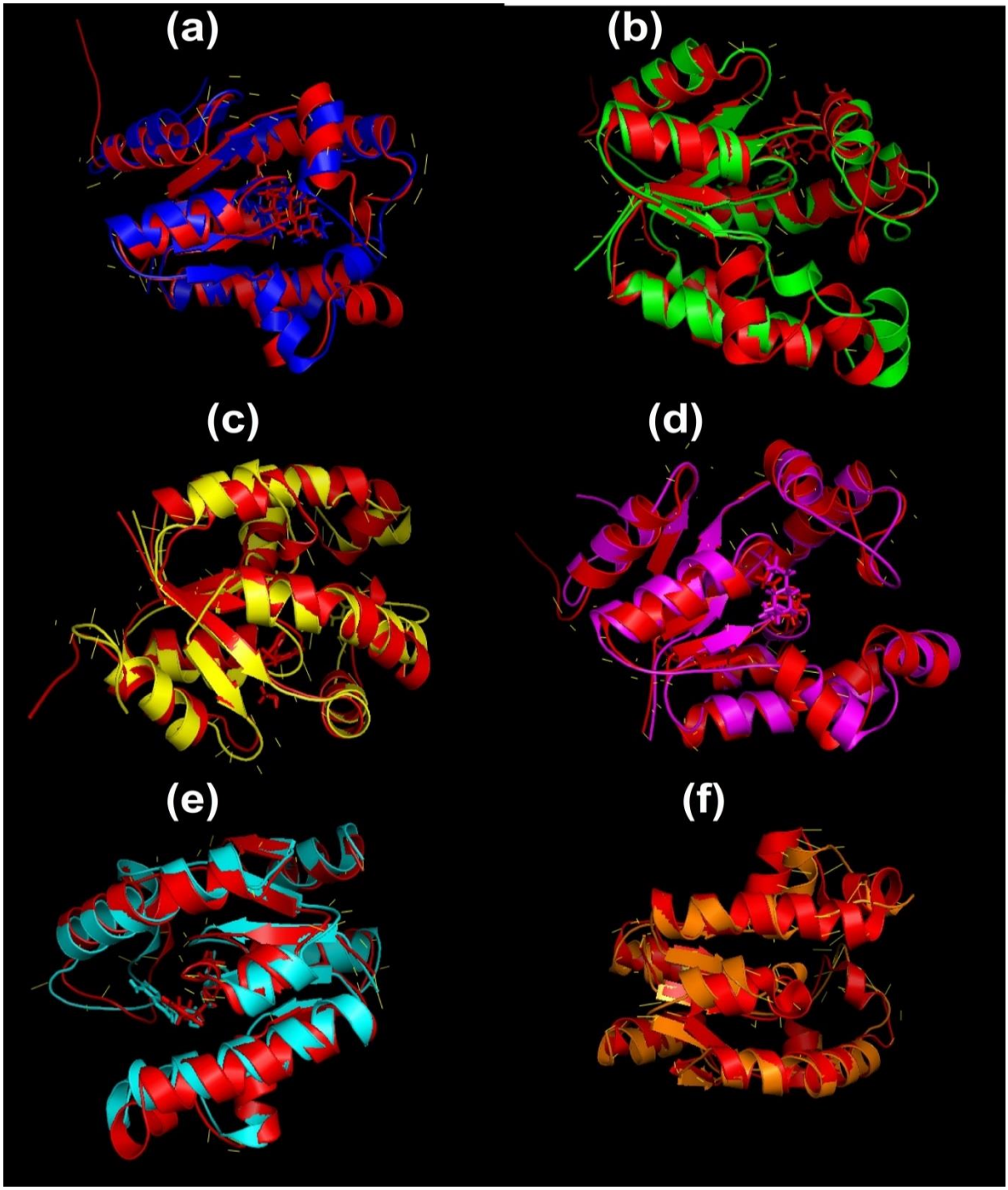


Figure 6a–f: Superposition of initial complex structures (before the 50-ns MD simulation) and final complex structures (after the MD simulation) obtained from MD simulation analysis. (a) SK─PubChem15478 complex. The final structure is shown in blue and the initial structure is in red (RMSF = 1.509 Å). (b) SK─ZINC02838601 complex. The final structure is shown in green and the initial structure is in red (RMSF = 1.594 Å). (c) SK─ ZINC11790367 complex. The final structure is shown in yellow and the initial structure is in red (RMSF = 1.921 Å). (d) SK–EPE complex. The final structure is shown in purple and the initial structure is in red (RMSF = 1.324 Å). (e) SK–ADP complex. The final structure is shown in greenish-blue and the initial structure is in red (RMSF = 1.278 Å). (f) SK–Mg^{2+} complex. The final structure is shown in orange and the initial structure is in red (RMSF = 1.983 Å)

Conclusion

Shikimate kinase is an essential enzyme in the shikimate pathway and considered as a drug target in MTB. The analysis of the 3D structure of SK obtained from PDB (PDB ID: 1L4U and resolution 1.8Å) revealed that it has a substrate or shikimate binding site (Asp34, Arg58, and Lys136) and substrate binding via amide nitrogen (Gly80). It also possessed nucleotide binding region (Gly12─Thr17), the ATP binding site (Arg117 and Arg153) and metallic ion (Mg2+) binding site (Ser16 and Asp32). These critical binding sites were evaluated using docking and MD simulation analyses with the ligand obtained from both Zinc and PubChem databases. The result of the analyses indicated that the two compounds (ZINC11790367 and PubChem15478) formed stable and rigid complexes comparable to the bound ligand (4-(2-Hydroxyethyl)-1-Piperazine Ethanesulfonic Acid (EPE)) and cofactors (ADP and Mg2+) during the 50ns MD simulation. Therefore, these compounds considered as prospective inhibitors of MTB after successful experimental validation.

Conflict of Interest

We have no conflict of interest

Acknowledgments

The author of this paper is very much grateful to **Prof. Pawan Dhar** (Jawaharlal Nehru University), **Prof. B. Jayaram** (Indian Institute of Technology Delhi), **Dr. Kalaiarasan P.** (Jawaharlal Nehru University), and **Mr. Shashank Shekhar** (Indian Institute of Technology Delhi) for providing facilities.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The Protein Data Bank Nucleic Acids Research, 28, 235-242. *URL: www. rcsb. org Citation*.

Case, D. A., Darden, T. A., Cheatham III, T. E., Simmerling, C. L., Wang, J., Duke, R. E., ... & Roberts, B. (2010). AMBER 12; University of California: San Francisco, 2012.

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., ... & Tang, Y. (2012). AdmetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties.

DeLano, W. L. (2002). The PyMOL user's manual. *DeLano Scientific, San Carlos, CA, 452*.

Dhaliwal, B., Nichols, C. E., Ren, J., Lockyer, M., Charles, I., Hawkins, A. R., & Stammers, D. K. (2004). Crystallographic studies of shikimate binding and induced conformational changes in MTB shikimate kinase. *FEBS letters*, *574*(1-3), 49-54

Ducati, R. G., Basso, L. A., & Santos, D. S. (2007). Mycobacterial shikimate pathway enzymes as targets for drug design. *Current drug targets*, *8*(3), 423-435.

Filgueira de Azevedo Jr W, Canduri F, Simões de Oliveira J, Basso LA, Palma MS, Pereira JH, Santos DS (2002). Molecular model of shikimate kinase from MTB. Biochem Biophys Res Commun. 295:142e8.

Hariono, M., Abdullah, N., Damodaran, K. V., Kamarulzaman, E. E., Mohamed, N., Hassan, S. S., ... & Wahab, H. A. (2016). Potential New H1N1 Neuraminidase Inhibitors from Ferulic Acid and Vanillin: Molecular Modelling, Synthesis and in Vitro Assay. *Scientific reports*, *6*, 38692.

Johansson, M. U., Zoete, V., Michielin, O., & Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics*, *13*(1), 173.

Krell, T., Maclean, J., Boam, D. J., Cooper, A., Resmini, M., Brocklehurst, K., ... & Coggins, J. R. (2001). Biochemical and X-ray crystallographic studies on shikimate kinase: The important structural role of the P-loop lysine. *Protein Science*, *10*(6), 1137-1149

Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery.

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1. *Advanced drug delivery reviews*, *46*(1-3), 3-26.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, *19*(14), 1639-1662.

Oliveira, J. S., Pinto, C. A., Basso, L. A., & Santos, D. S. (2001). Cloning and overexpression in the soluble form of functional shikimate kinase and 5-enolpyruvylshikimate 3-phosphate synthase enzymes from MTB. *Protein expression and purification*, *22*(3), 430-435.

Parish, T., & Stoker, N. G. (2002). The common aromatic amino acid biosynthesis pathway is essential in MTB. *Microbiology*, *148*(10), 3069-3077.

Pereira, J. H., de Oliveira, J. S., Canduri, F., Dias, M. V., Palma, M. S., Basso, L. A., & de Azevedo, W. F. (2004). Structure of shikimate kinase from MTB reveals the binding of shikimic acid. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12), 2310-2319.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Ramachandran, S., Kota, P., Ding, F., & Dokholyan, N. V. (2011). Automated minimization of steric clashes in protein structures. *Proteins: Structure, Function, and Bioinformatics*, *79*(1), 261-270.

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, *45*(12), 2615-2623.

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1996). Derivation of 3D coordinate templates for searching structural databases: Application to Ser-His-Asp catalytic triads in the serine proteinases and lipases. *Protein Science*, *5*(6), 1001-1013.

Zhang, X., Zhang, S., Hao, F., Lai, X., Yu, H., Huang, Y., & Wang, H. (2005). Expression, purification and properties of shikimate dehydrogenase from MTB. *Journal of biochemistry and molecular biology*, *38*(5), 624.