

Original Research Article

In silico design of small molecules targeting cathepsin-k for the treatment of osteoarthritis

Shenghong Zhang, Yao Chi Wu*

Department of Acupuncture, Tuina and Traumatology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, 200233, China

*For correspondence: **Email:** webberLepati@yahoo.com; **Tel:** 0086-021-24058567

Sent for review: 20 April 2018

Revised accepted: 22 July 2018

Abstract

Purpose: To investigate cathepsin-K as an alternative drug target for the development of a new therapy for osteoarthritis (OA) in humans.

Method: *In silico* molecular docking simulation-based virtual screening was used to identify probable lead molecules with significant binding affinities for the target receptor, and acceptable pharmacokinetic profiling. Lamarckian genetic algorithm was used for molecular docking simulation, while various physicochemical parameters such as molecular weight, calculated partition coefficient, topological polar surface area, and hydrogen bond acceptor and donor counts were evaluated for their pharmacokinetic profile.

Results: The compound, ZINC05386901, was found to be a potent inhibitor of human cathepsin-K protein with excellent pharmacokinetic properties, and had no toxic effects.

Conclusion: The designed inhibitor molecule for cathepsin-K protein is a promising lead molecule for further structure-based discovery of novel drugs for the treatment of OA.

Keywords: Osteoarthritis, Cathepsin-K, Drug design, Ligand

This is an Open Access article that uses a funding model which does not charge readers or their institutions for access and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>) and the Budapest Open Access Initiative (<http://www.budapestopenaccessinitiative.org/read>), which permit unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

Tropical Journal of Pharmaceutical Research is indexed by Science Citation Index (SciSearch), Scopus, International Pharmaceutical Abstract, Chemical Abstracts, Embase, Index Copernicus, EBSCO, African Index Medicus, JournalSeek, Journal Citation Reports/Science Edition, Directory of Open Access Journals (DOAJ), African Journal Online, Bioline International, Open-J-Gate and Pharmacy Abstracts

INTRODUCTION

Osteoarthritis (OA) is a human pathological and cartilaginous disorder caused by biological and mechanical events occurring simultaneously to interfere with the normal coupling between the synthesis and degradation of the tissues involved in the disease [1]. The initiation of the osteoarthritic disease state is governed by genetic, metabolic, developmental and traumatic factors. The diagnostic features of OA are based on biochemical, morphological, molecular and bio-mechanical changes in cells and cellular

matrix which result in loss of articular cartilage, ulceration, softening, sclerosis, and fibrillation. Other features include eburnation of osteophytes, subchondral bone, and subchondral cysts. The characteristic features associated with OA are crepitus, tenderness, painful joints, limited movement, intermittent effusion, and variable degrees of inflammation without systemic effects [2]. One common feature of OA is pain in the joints leading to decreased mobility and function. Thus, the primary goal of the OA therapy is to provide relief to the patient from the painful condition associated with the disease.

Thus, the medical care of OA patients includes the initiation of non-pharmacological therapy as a foundation for pain relief, followed by concurrent pharmacological therapy for improvement of the diseased condition [3]. On the basis of safety, efficacy and cost considerations, a regularly scheduled dose of acetaminophen (up to 4 g/day) is prescribed to the OA patients for symptomatic relief from pain. When this fails, drugs like NSAIDs, capsaicin, opioids, glucosamine and chondroitin are prescribed for topical applications, in combination with intra-articular injections of corticosteroids or hyaluronic acid. The major problem associated with the use of these drugs is their toxicity in the gastrointestinal tract and cardiovascular system, as well as hepatic, renal, and central nervous system toxicities. Due to these limitations, there is a need to explore new drug targets for development of safer therapies for OA [1-3]. Cathepsins are enzymes involved in the expression of osteoclasts by cleaving the telopeptide region of triple helical collagens. Human cathepsin K represents critical bone resorbing proteases having strong collagenase activity on the type-I and type-II collagens present in the main connective tissue. The expression of cathepsin K enzyme occurs predominantly in the osteoclasts and some other multinucleated cells like Langhans cells and giant foreign body cells [1-3].

EXPERIMENTAL

Selection and preparation of cathepsin-K protein

Human cathepsin-K protein with complexed ligand 3XT (pdb id-4X6H) was obtained from RCSB protein data bank. The macromolecular receptor protein was prepared for molecular docking simulations by removal of ligand from its active site, removal of water molecules, manual addition of hydrogen atoms, addition of charge, assigning autodock tools (ADT) atom type, and finally saving it in *pdbqt format [4].

Preparation of ligand for molecular docking

The substrate (ligand) 3XT i.e 4-amino-3-fluoro-N-(1-((2Z)-2-iminoethyl)carbonyl)cyclohexyl)benzamide was prepared for molecular docking simulation by providing the rotatable, non-rotatable and un-rotatable bonds present in the ligand.

Identification of binding site

The ligand-binding site present in human cathepsin-K protein was identified using protein

visualization software like DS Visualizer, Pymol and Chimera. The ligand molecule 3XT binds to the active site of the receptor.

Grid-box formation

The grid parameter points of the grid box used in the present study for performing the molecular docking simulation of human cathepsin-K protein receptor are shown in Table I. These grid parameters were utilized for all docking runs. The grid-box was placed by centering the ligand so that all the interacting binding residues associated with the binding of the ligand are well covered within the grid box to ensure that all the extended conformations of ligand fit within the grid-box.

Preparation of grid maps

The map files of various atom types present within the receptor and the ligands viz: A,HD, C, NA, N, SA, F and OA were prepared by running Autogrid utility of the AutoDock suite. These map files prepared by *Autogrid* were further utilized by the *AutoDock* program for performing molecular docking simulations.

Docking parameters

Lamarckian genetic algorithm (LGA) is the primary conformational search approach in AutoDock [19]. A trail population was created for various possible conformations, and mutations and exchange of conformational parameters took place in a similar approach to that of biological evolution in successive generations, for final selection of individuals with minimum binding energy. Semi-empirical free energy forcefield was used to identify free binding energies for the binding of small ligand molecules to macromolecular drug targets. The forcefield was centred on an inclusive thermodynamic model base that permitted intramolecular energy assimilation within the predicted free binding energy by assessing energies individually for bound and unbound states. The docking parameter file (*.dpf) was composed separately for each of the ligands by compiling parameters like GA runs to 150, maximum number of evaluations to 250000, maximum number of generations to 27000, and rate of gene mutation to 0.02.

Validation of docking method

The position and orientations of the ligand obtained after the molecular docking study represent potential modes involved in the binding of inhibitors. The various parameters considered

in the molecular docking methods were further validated by independently re-docking the crystallized form of ligand. The molecular docking simulation technique was validated using the following parameters:

Binding energy

To validate the molecular docking method, the binding energy of the docked ligand should be within the pre-defined range i.e. between -5 and -15 kcal/mol.

Overlay methods

The molecular docking method was deemed validated when the docked conformation of the ligand was absolutely overlaid with the crystallized bioactive conformation of the ligand present in the downloaded protein.

Chemical resemblance

The molecular docking method was deemed validated when the interactions of the docked ligand with the residues of macromolecule were similar to interactions present in the downloaded crystallized macromolecule.

Selection of chemical libraries

A total of 1880 diverse molecules present in the "NCI Diversity Set II" molecular library were virtually screened to identify possible lead compounds. All the ligands used in virtual screening followed Lipinsky's rule of five for good pharmacokinetic properties. According to Lippinski's rule of five, the ligands should have the following physicochemical properties in the given range: calculated partition coefficient (ClogP) up to 5, hydrogen bond donor up to 5, hydrogen bond acceptor up to 10, and molecular weight up to 500.

Virtual screening process

The essential files required for performing the virtual screening process were procured with the help of Raccoon. Raccoon is a graphical user interface for splitting molecular ligand library file having multiple number of individual ligand molecules. The individual ligand molecules are filtered on the basis of Lipinski's rule of five, fragment like rule of 3, and druglikeness value. These filtered ligand molecules are further converted and saved in the pdbqt format required by the *AutoDock*. The input filename, coherent format of grid maps, existence of nonstandard atom types, and confirming that these

parameters were evaluated at every step for the validation of the process [19].

Analysis of results of docking simulation

Virtual screening results were summarized and sorted for identification of best hits by using `summarize_results.py` python script from the Scripps Research Institute. The docking results were evaluated on the basis of hydrophilic and hydrophobic interactions obtained among ligand and the binding residues. The empirical range of the free binding energy should be in the predefined range of -5 to -15 kcal/mol. Eq 1 was used for the calculation of binding affinity [20]

$$K_i = e^{[(\Delta G)/(RT)]} \dots\dots\dots (1)$$

where ΔG = Gibbs free energy of binding, T = absolute temperature, and R = gas constant.

Prediction of ADME and toxicity of lead compounds

The toxic effects of the selected lead molecules were predicted by using the online program, OSIRIS-molecular property explorer. The occurrence of major toxic effects such as mutagenicity, tumorigenicity, irritant effect and reproductive effect within the lead molecules were identified on the basis of functional groups present in their chemical structures. Druglikeness scores and drug scores of the selected lead molecules were also predicted by this program, based on their physicochemical properties.

RESULTS

Selection and preparation of cathepsin-K protein

The 4X6H protein complex downloaded from RSCB protein data bank consisted of a single polypeptide chain of 215 amino acids and a bound ligand 4-amino-3-fluoro-N-(1-[[[(2Z)-2-iminoethyl]carbonyl]cyclohexyl]benzamide. All the 290 water molecules present in the receptor molecule were removed using autodock tools and the receptor saved in *.pdbqt format [4].

Preparation of ligand for molecular docking

No rotatable and non-rotatable bonds were present within the ligand molecule adenine. All the bonds of the ligand molecule were unrotatable in nature. The prepared ligand was saved in *.pdbqt format [5,6].

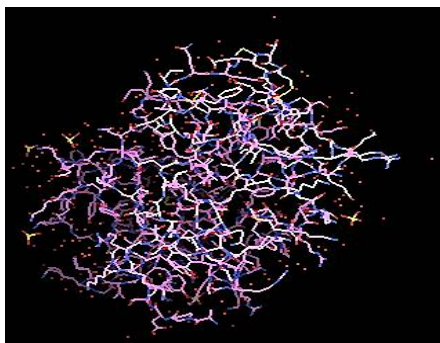


Figure 1: Crystal structure of cathepsin K protein of *Homosapiens* acquired from RCSB protein data bank (PDB ID-4X6H)

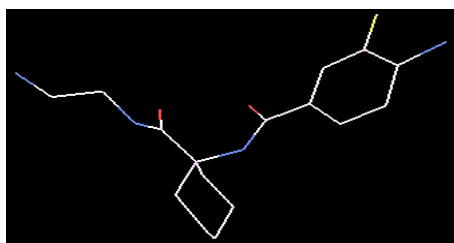


Figure 2: Bound ligand 3XT separated from human cathepsin-K protein using software chimera

Identification of binding site

Amino acids ASP61, ASN161, GLN19 and GLY66 of protein 4X6H were identified as the binding residues present within the ligand binding site in the human cathepsin-K protein [7,8].

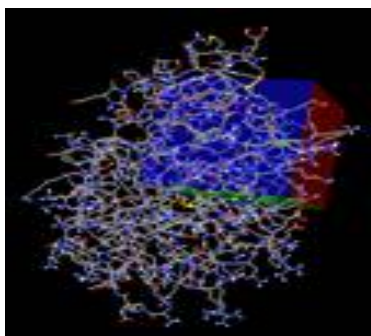


Figure 3: Three dimensional grid box covering the binding site present in the receptor molecule and all the binding residues involved in ligand binding

Grid-Box

The coordinates used for the preparation of the grid box are indicated in Table 1.

Molecular docking simulation

The results obtained from molecular docking of the complexed ligand 4-amino-3-fluoro-N-(1-{{(2Z)-2-iminoethyl]carbamoyl}cyclohexyl)benzamide (3XT) with huma cathepsin-K protein are presented in Table 2 [5,6].

Validation of molecular docking

The molecular docking process was successfully validated, because the binding energy obtained for the complex of ligand 3XT with cathepsin-K protein was within the pre-defined range of -5 to -15 kcal/mol. Moreover, the docked conformation of the ligand was exactly superimposed over its bioactive conformation. The overlaid conformation of the docked ligand relative to the bioactive conformation of ligand is shown in Figure 4.

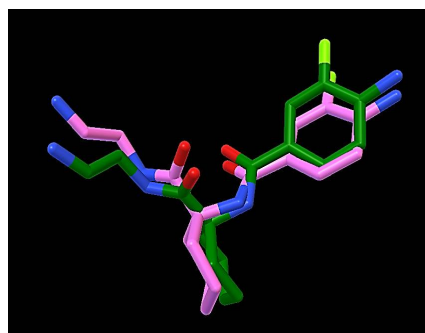


Figure 4: Superimposition of the docked conformation of the ligand, relative to the bioactive conformation of the ligand obtained from the crystal structure of the downloaded protein

The molecular docking method was further validated based on the observation that the chemical interactions of the docked ligand with the macromolecular binding residues was similar to that present in the downloaded crystallized macromolecular complex. The interactions

Table 1: The grid coordinates used for the preparation of the grid box for cathepsin-K protein

Proteins	x-D	y-D	z-D	Spacing (Å)	x center	y center	z center
4X6H	40	40	40	0.458	16.559	8.022	-18.988

Table 2: Results from molecular docking of ligand 3XT with cathepsin-K protein (4X6H)

Protein	Interacting residues	Internal validation rmsd value	Binding energy (kcal/mol)	Binding affinity (µM)
4X6H	ASP61, ASN161, GLN19 and GLY66	0.84	-5.02	132.17

present in the crystal structure are shown in Figure 5 A, while the interactions present in the docked structure are shown in Figure 5 B.

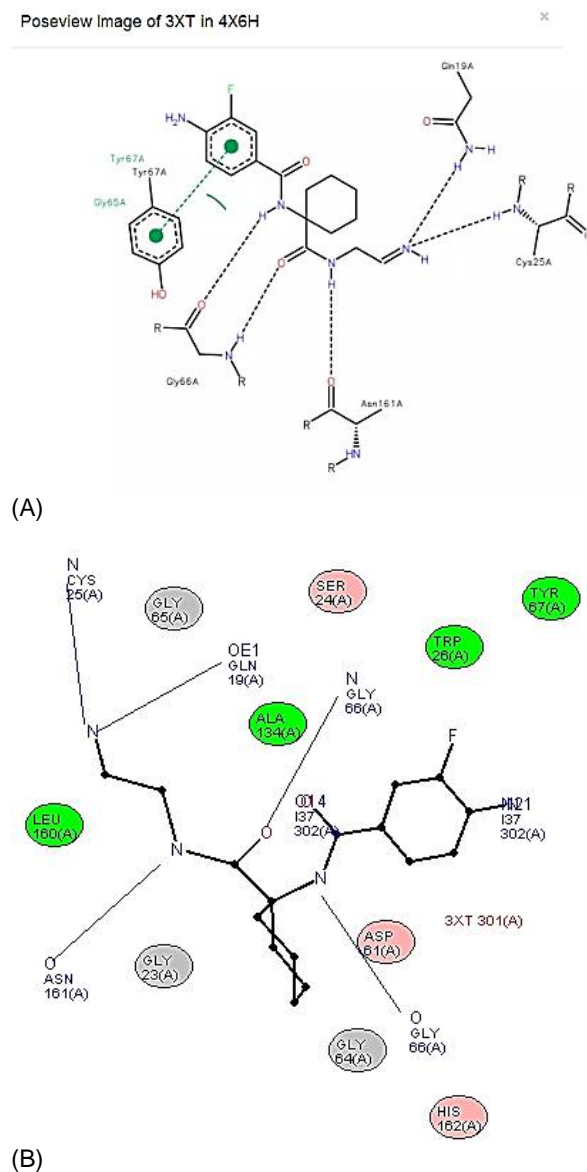


Figure 5: Binding mode and chemical interactions of the bound ligand 3XT within the active ligand binding site of human cathepsin-K protein. A: Bioactive interactions, B: Docking interactions

Virtual screening

Potential ligands were selected by analyzing the ligand-protein interactions for top ranking pose of each ligand, and interactions of docked compound were visualized. The molecular docking results of five top molecules selected in *Autodock*-based virtual screening of the human cathepsin-K protein with *NCI Diversity set-II*

ligand library having 1880 small ligand molecules are shown in Table 3.

Pharmacokinetic profiles of potential lead molecules

The five potent lead molecules obtained after performing molecular docking simulation-based virtual screening of *NCI Diversity set-II* ligand library with 1880 diverse ligand molecules were evaluated for their pharmacokinetic profiles. The important physicochemical properties of the ligand such as molecular weight, calculated partition coefficient (cLogP), two dimensional polar surface area (2D PSA), hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA) counts were studied to predict their pharmacokinetic profiles. *Osiris Molecular Property Explorer*, and *Marvin Sketch* software were used to calculate the physicochemical properties of the selected ligands. The physicochemical properties of the five best lead molecules for cathepsin-K protein of human are shown in Table 4 [12-14].

ADME and toxicity profiling of lead compounds

Pharmacokinetic parameters such as absorption, distribution, metabolism, excretion and toxicity were evaluated using online server *Osiris Molecular Property Explorer*. The best five virtually screened lead molecules were tested for their pharmacokinetic parameters, presence of major toxic effects, druglikeness, and drug score values. On the basis of pre-defined range of the explored physicochemical properties according to the Lipinski's rule, it was observed that two out of the five selected lead molecules, i.e. ZINC05386901 and ZINC18157167 showed very good pharmacokinetic profiles without toxic effects. It was also revealed that they had very good drug-likeness and drug score values. On the other hand, the other lead molecules showed poor pharmacokinetic profiles, in addition to some major toxic effects such as mutagenicity, tumorigenicity and reproductive effects. The ADME and toxicity results of the best lead molecules obtained after the virtual screening are shown in Figure 5 to Figure 9.

DISCUSSION

The separated ligands and receptor protein were prepared for the *Autodock*-based molecular docking simulation by removal of non-interacting water molecules, addition of polar hydrogens and

Table 3: Binding energies and affinities of five top hits from *NCI Diversity Set II* after virtual screening against cathepsin-K protein

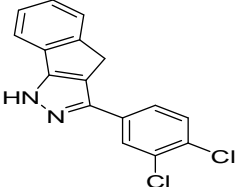
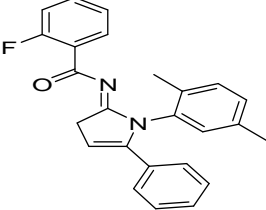
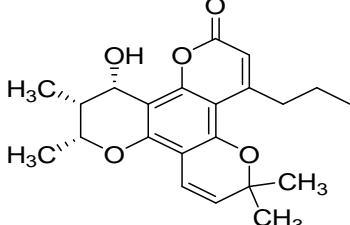
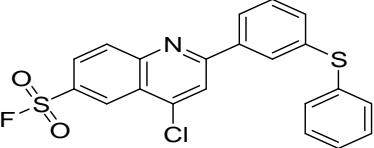
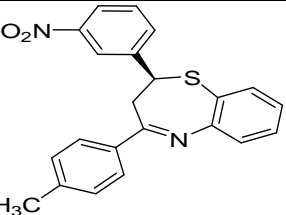
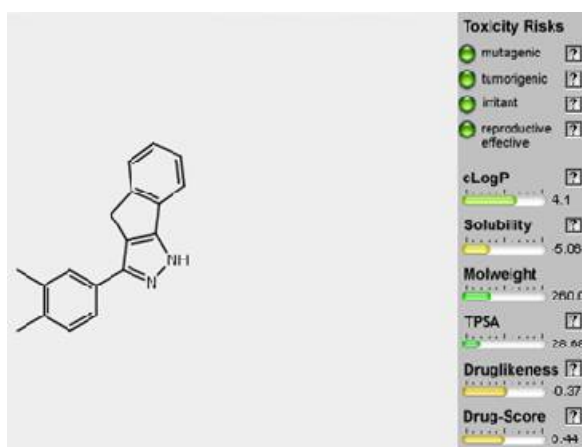
S/no.	Compound ID	Chemical structure	Binding energy (kcal/mol)
1	ZINC05386901		-9.66
2	ZINC18157167		-9.45
3	ZINC05536814		-9.41
4	ZINC17465958		-9.13
5	ZINC04376856		-9.12

Table 4: Lipinski's "rule of five" for the hits on the cathepsin-K protein

Compound ID	Mol. Wt.	c log P	2D PSA (Å ²)	HBA	HBD
ZINC05386901	260	4.1	28.68	1	1
ZINC18157167	383	6.06	20.31	3	0
ZINC05536814	370	3.95	64.99	4	1
ZINC17465958	429	6.23	80.71	3	0
ZINC04376856	374	5.25	65.73	4	2

**Figure 5:** ADME profiling and toxicity prediction of ZINC05386901 molecule using online *Molecular Properties Prediction Server- Osiris Property Explorer*

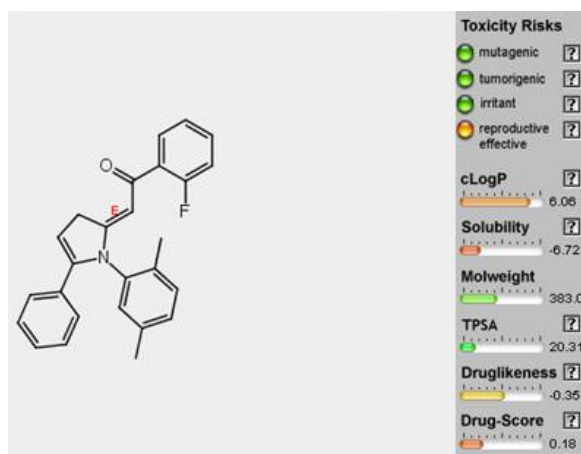


Figure 6: ADME profiling and toxicity prediction of ZINC18157167 molecule using online *Molecular Properties Prediction Server- Osiris Property Explorer*

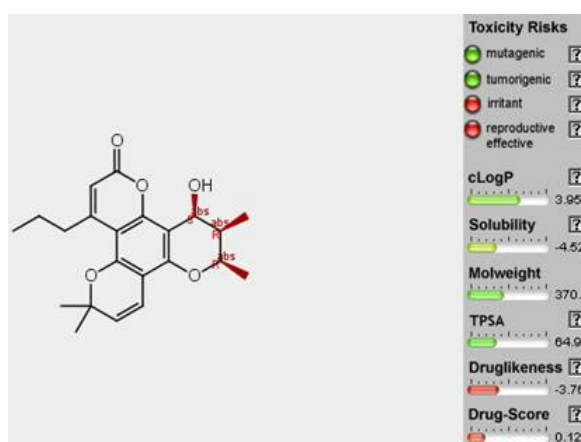


Figure 7: ADME profiling and toxicity prediction of ZINC05536814 molecule using online *Molecular Properties Prediction Server- Osiris Property Explorer*

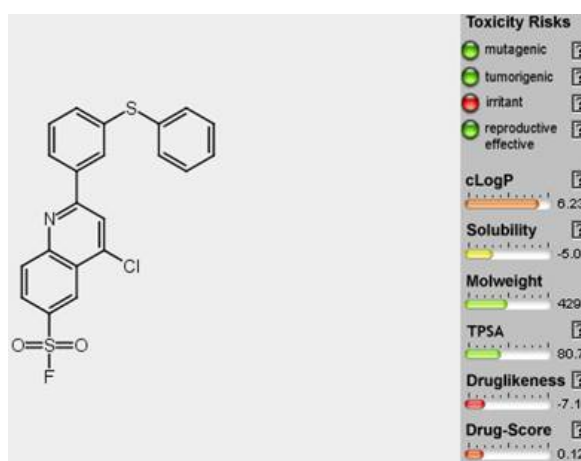


Figure 8: ADME profiling and toxicity prediction of ZINC17465958 molecule by using online *Molecular Properties Prediction Server- Osiris Property Explorer*

assigning of Gastgeiger charge. A grid box was prepared by covering the different conformations of ligands as well as the binding cavity of the macromolecule which consisted of ASP61, ASN161, GLN19 and GLY66 as active binding residues. The grid information was utilized by the

Autogrid utility of the Autodock suite for the generation of various map files which were further utilized for performing molecular docking simulations.

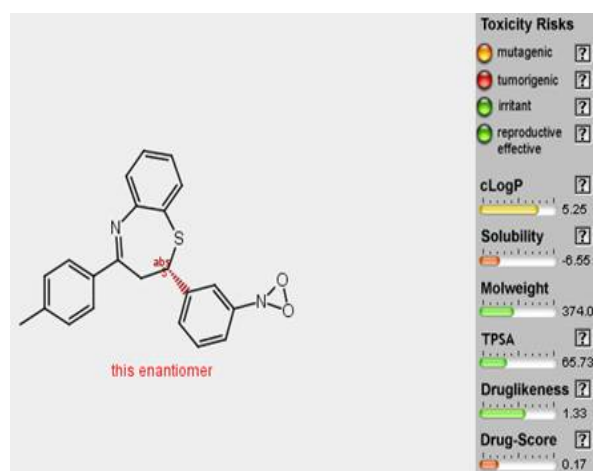


Figure 9: ADME profiling and toxicity prediction of ZINC04376856 molecule using online *Molecular Properties Prediction Server- Osiris Property Explorer*

The bound ligand 3XT was docked with cathepsin-K protein with acceptable binding energy of -5.02 kcal/mol. The molecular docking process was further validated by perfect overlay of the docked conformation of ligand with respect to its bioactive conformation. The chemical interaction after molecular docking was similar to that present in the bioactive protein complex obtained from the protein data bank. Raccoon software was used for performing virtual screening of cathepsin-k protein with *NCI Diversity Set-II* having 1880 diverse ligand molecules. Out of 1880 ligands molecules of the ligand library, five best molecules i.e. ZINC05386901, ZINC18157167, ZINC05536814, ZINC17465958, ZINC04376856 were selected based on their minimum binding energies (-5 to -15 kcal/mol). These selected molecules were further evaluated for their pharmacokinetic profiling by calculating various physicochemical properties such as molecular weight, calculated partition coefficient (cLogP), two dimensional polar surface area (2D PSA), hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA) counts.

Furthermore, these molecules were tested for the presence of major toxic effects like tumorigenicity, mutagenicity, irritant and reproductive effects. They were also tested for druglikeness and drug score values. After performing intense evaluation of the selected molecules, two molecules ZINC05386901 and ZINC18157167 were proposed as potent inhibitor molecules for human cathepsin-K protein, with potential for the treatment of osteoarthritis.

CONCLUSION

Molecular docking simulation-based *in silico* virtual screening using *Autodock4* is very useful for shortlisting potential lead molecules from libraries containing a large number of ligand molecules. The compound, ZINC05386901, is a potent inhibitor of human cathepsin-K protein; it has excellent pharmacokinetic properties, and is not associated with any major toxic effects. Moreover, it has excellent druglikeness and drug score. Thus, this molecule can serve as a promising lead compound for further structure-based discovery of novel drugs for the treatment of osteoarthritis.

DECLARATIONS

Acknowledgement

The authors are highly thankful to the Plan of Action of Traditional Chinese Medicine of Shanghai and Project of Shanghai Commission of Science and Technology for providing financial support for this study.

Conflict of interest

The authors declare no conflict of interest is associated with this study.

Contribution of authors

We declare that this work was done by the authors named in this article and all liabilities pertaining to claims relating to the content of this article will be borne by the authors. Both authors contributed equally to this work. Yao Chi Wu designed the frame work for performing this research. Shenghong Zhang performed all the molecular docking simulation as well as virtual screening process. The pharmacokinetic profiling as well as toxicity evaluation was performed by Yao Chi Wu.

REFERENCES

- Brömme D, Lecaillon F. Cathepsin K inhibitors for osteoporosis and potential off-target effects. *Expert Opin Investig Drugs*. 2009; 18(5): 585-600.
- Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, Kington RS, Lane NE, Nevitt MC, Zhang Y, Sowers M. Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med*. 2000; 133(8): 635-646.
- Dipiro JT, Talbert RL, Yee GC, Matzke GR, Wells BG, Posey LM. *Pharmacotherapy- A Pathophysiologic Approach*. 7th Edition, Chapter 77, pp 1205-1242.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000; 28(1): 235-242.
- Mujwar S, Pardasani KR. Prediction of Riboswitch as a Potential Drug Target for Infectious Diseases: An *In Silico* Case Study of Anthrax. *J Med Imaging Health Inform* 2015; 5(1): 1-10.
- Morris G. *Autodock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. *J Comput Chem* 2009; 30(16): 2785-2791.
- DeLano WL. *PyMOL*. DeLano Scientific, San Carlos, CA. 2002; 700.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25(13): 1605-1612.
- Mujwar S, Pardasani KR. Prediction of Riboswitch as A Potential Drug Target and Design of Its Optimal Inhibitors for *Mycobacterium Tuberculosis*. *Int J Comput Biol Drug Des*. 2015; 8(4): 326-347.
- Forli S, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. Raccoon Autodock Preparation Tool; The Scripps Research Institute. *J Comput Chem*. 2009; 30(16): 2785-2791.
- Tuccinardi T. Structure-based virtual screening: identification of novel CB2 receptor ligands. *Bioorg Med Chem Lett*. 2007; 4(1): 15-19.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997; 23(1-3): 3-25.
- Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods*. 2000; 44(1): 235-249.
- Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol*. 2004; 1(4): 337-341.
- Thomas Sander, Actelion Pharmaceuticals Ltd., Gewerbestrasse 16, 4123 Allschwil, Switzerland, Email: thomas.sander@actelion.com.
- Patani A, George LaVoie, J Edmond. Biososterism: A rational approach in drug design. *Chem Rev*. 1996; 96: 3147-3176.
- Kaur A, Mujwar S, Adlakha N. *In-silico* analysis of riboswitch of *Nocardia farcinica* for design of its inhibitors and pharmacophores. *Int J Comput Biol Drug Des*. 2016; 9(3): 261-276.
- Pradhan P, Soni NK, Chaudhary L, Mujwar S, Pardasani KR. *In-Silico* Prediction of Riboswitches and Design of their Potent Inhibitors for H1N1, H2N2 and H3N2 Strains of Influenza Virus. *Biosci Biotechnol Res Asia*. 2015; 12(3): 2173-2186.
- Sander T, Freyss J, von Korff M, Reich JR, Rufener C. OSIRIS, an entirely in-house developed drug discovery informatics system. *J Chem Inf Model*. 2009; 49(2): 232-246.

20. Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit.* 1996; 9(1): 1-5.
21. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules.* 2015; 20(7): 13384-13421.
22. Gregory S, Kothiwale SK, Meiler J, Lowe EW. Computational methods in drug discovery. *Pharmacol Rev.* 2014; 66(1): 334-395.
23. Rang HP, Dale MM, Ritter JM, Flower RJ, Henderson G. Chapter 3: How drugs act: molecular aspects. *Rang and Dale's Pharmacology.* Edinburgh; New York: Elsevier/Churchill Livingstone. 2003; pp 20–48.
24. Mujwar S, Pardasani KR. Prediction of Riboswitch as A Potential Drug Target in Streptococcus. *Online J Bioinfo.* 2012; 13(2): 285-313.
25. Soni N, Pardasani KR, Mujwar S. Insilico analysis of dietary agents as anticancer inhibitors of insulin like growth factor 1 receptor (IGF1R). *J Pharm Pharm Sci.* 2015; 7(9): 191-196.
26. Khan MR, Soni LK. Molecular docking simulation of small diverse chemical molecules based virtual screening for treatment of tuberculosis. *Asian J Chem.* 2016; 28(12): 2617-2621.