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

Abstract

Purpose: To investigate cathespin-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. Lamarkian 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

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]. Cathespins 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 cathespin-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]carbamoyl)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 Lipinski’s rule of five for good pharmacokinetic properties. According to Lipinski’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]

\[ K_i = e^{\frac{\Delta G}{RT}} \] ........................ (1)

where \( \Delta G \) = 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-((2Z)-2-iminoethyl)carbamoyl)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 un-rotatable in nature. The prepared ligand was saved in *.pdbqt format [5,6].
Figure 1: Crystal structure of cathespin 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-((2Z)-2-iminoethyl)carbamoylcyclohexyl)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.

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

<table>
<thead>
<tr>
<th>Proteins</th>
<th>x-D</th>
<th>y-D</th>
<th>z-D</th>
<th>Spacing (Å)</th>
<th>x center</th>
<th>y center</th>
<th>z center</th>
</tr>
</thead>
<tbody>
<tr>
<td>4X6H</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>0.458</td>
<td>16.559</td>
<td>8.022</td>
<td>-18.988</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Protein</th>
<th>Interacting residues</th>
<th>Internal validation rmsd value</th>
<th>Binding energy (kcal/mol)</th>
<th>Binding affinity (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4X6H</td>
<td>ASP61, ASN161, GLN19 and GLY66</td>
<td>0.84</td>
<td>-5.02</td>
<td>132.17</td>
</tr>
</tbody>
</table>

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
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, drug likeness, 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-likeliness 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, tumorogenicity 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

<table>
<thead>
<tr>
<th>S/no.</th>
<th>Compound ID</th>
<th>Chemical structure</th>
<th>Binding energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ZINC05386901</td>
<td><img src="image1" alt="Chemical structure" /></td>
<td>-9.66</td>
</tr>
<tr>
<td>2</td>
<td>ZINC18157167</td>
<td><img src="image2" alt="Chemical structure" /></td>
<td>-9.45</td>
</tr>
<tr>
<td>3</td>
<td>ZINC05536814</td>
<td><img src="image3" alt="Chemical structure" /></td>
<td>-9.41</td>
</tr>
<tr>
<td>4</td>
<td>ZINC17465958</td>
<td><img src="image4" alt="Chemical structure" /></td>
<td>-9.13</td>
</tr>
<tr>
<td>5</td>
<td>ZINC04376856</td>
<td><img src="image5" alt="Chemical structure" /></td>
<td>-9.12</td>
</tr>
</tbody>
</table>

Table 4: Lipinski’s “rule of five” for the hits on the cathepsin-K protein

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Mol. Wt.</th>
<th>c</th>
<th>log P</th>
<th>2D PSA (Å²)</th>
<th>HBA</th>
<th>HBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC05386901</td>
<td>260</td>
<td>4.1</td>
<td>28.68</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ZINC18157167</td>
<td>383</td>
<td>6.06</td>
<td>20.31</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ZINC05536814</td>
<td>370</td>
<td>3.95</td>
<td>64.99</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ZINC17465958</td>
<td>429</td>
<td>6.23</td>
<td>80.71</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ZINC04376856</td>
<td>374</td>
<td>5.25</td>
<td>65.73</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: ADME profiling and toxicity prediction of ZINC05386901 molecule using online Molecular Properties Prediction Server- Osiris Property Explorer.
Assigning of Gasteiger 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 Autodock utility of the Autodock suite for the generation of various map files which were further utilized for performing molecular docking simulations.

The bound ligand 3XT was docked with cathespin-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 cathespin-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 cathespin-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 cathespin-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

15. Thomas Sander, Actelion Pharmaceuticals Ltd., Gewerbestrasse 16, 4123 Allschwil, Switzerland, Email: thomas.sander@actelion.com.


