Structural Bioinformatics and Molecular Dynamics Simulations Studies of Cathepsins as a Potential Target for Drug Discovery

Surapong Pinitglang*, Ratchanee Saiprajong*, Tossaporn Dussadee, Khanok Ratanakhanokchai

*Department of Food Science and Technology, School of Science and Technology, University of the Thai Chamber of Commerce 126/1 Vibhavadee Rangsit Road, Bangkok, 10400, Thailand.

Abstract

Prediction of three-dimensional structure of cathepsins, and molecular dynamics simulations of cathepsin S were studied by interaction with the drug molecule with virtual screening 681,158 compounds from ZINC database. The result of study showed top 1 ranked was obtained with drug molecule ZINC 23215439 reaction with cathepsin S. This demonstrates that the active site of cathepsin S Cys25, His164 and binding site Gln19 and Gly 20 are essential for interactions of cathepsin S-ZINC 23215439 inhibitor complex. Coulomb-SR and Lennard-Jones-SR interactions energy of amino acids and drug molecule ZINC code 23215439 which consisted in active site of cathepsin S have been evaluated.

© 2012 The Authors. Published by Elsevier B.V.
Selection and/or peer-review under responsibility of the Program Committee of CSBio 2012.

Keyword: cathepsin, homology modeling, virtual screening, cysteine proteinase

1. Introduction

Cathepsins are proteases that break apart other proteins, found in many types of cells including those in all animals. There are approximately a dozen members of this family, which are distinguished by their structure, catalytic mechanism, and proteins that are by them cleaved. Most of the members become activated at the low pH found in lysosomes [1]. Thus, the activity of this family lies almost entirely within those organelles. In this research purpose to study on three-dimensional structural investigation of cathepsins in cysteine proteinase C1

* Corresponding author. Tel.:+662-697-6525; fax: +662-277-7007.
E-mail address: surapong_p@utcc.ac.th.
papain family was carried out by Homology modeling method [2-3]. Cathepsin S is a lysosomal enzyme that belongs to the papain family of cysteine proteases as a potential target for anti-cancer cancer therapy [4]. The protein structure of cathepsin S (EC 3.4.22.27) is highly similar to several members of the family, with for example 57% sequence identity to cathepsin L and K and approximately 30% identity to cathepsin B [5]. Virtual screening of cathepsin S reactions with drug-like molecules from ZINC database were investigated and study on phenomenon a motion of protein structure in particular of active site with both free and complex of cathepsin S by using molecular docking studies. Virtual screening is a computational technique used in drug discovery research. It involves the rapid in silico assessment of large libraries of chemical structures in order to identify those structures that most likely to bind to a drug target, typically a protein receptor or enzyme [6]. Virtual screening has become an integral part of the drug discovery process [7-9]. Virtual screening is an emerging technology that is gaining an increased role in the drug discovery process. The technique involves analyzing large collections of compounds, leading to smaller subsets for biological testing [10]. It is now perceived as a complementary approach to experimental screening and, when coupled with structural biology, promised to enhance the probability of success in the lead identification stage of the drug discovery process. Structure-based virtual screening of cathepsin S for use as potential anticancer drug requires computational fitting of compounds into an active site of a receptor by use of sophisticated algorithms, followed by scoring and ranking of these compounds to identify potential leads [11-12].

2. Methods

2.1. Predictions of three-dimensional structure of cathepsins

*Homology modeling method*

No X-ray crystal structures of cathepsins 1, 2, 3, 6, Q, P, m, R, W and O. have been deposited in the protein Data Bank based on the amino acid sequence from GeneBank and UniProt Database [13]. X-ray crystallography structure of cathepsins in C1 cystein protienase family were obtained from the Brookhaven Protein Data Bank [14]. These models of three-dimensional structures were used as the template structure for prediction three-dimensional structure of cathepsins. The models of three-dimensional structure of cathepsins was generated using the homology module of Insight II [15] based on X-ray crystallographic structure of template which showed the highest identity percentage by using Geno3D. Geno3D is an automatic web server for protein molecular modeling [16]. The searching of the template structure was determined by starting with a query protein sequence, the server performs in three successive steps were identified homologous proteins with known three-dimensional structures by using PSI-BLAST, provide the user all potential templates through a very convenient user interface for target selection and perform the alignment of both query and subject sequence. The sequences of cathepsins and template were aligned by homology module to identify the blocks that are likely to contain structurally conserved region. All atomic coordinates of the residues in those blocks were transferred from template to build the modeled structure of cathepsins. However, for the mismatch residues, only the atomic coordinate of Cα were transferred from template while the residual atomic coordinate were generated by using library.

*Energy minimization*

All hydrogen atoms were added to the homology-modeled structure of cathepsin by a builder module and then were energy-minimized by discovery module of Insight II. The minimization was performed with a 5Å-water layer in dimensions of 60x60x60 Å in a water box. First, the steepest descent algorithm followed by the more efficient conjugate gradient was reached. After minimization, the layer water was removed. Validation of the model was carried out using Ramachandran plot calculations computed with the Procheck program [17].
Superimposing cathepsins structures

Superimposing cathepsins structures were generated by using Discovery Studio (DS) V2.5 [18]. The first step in comparing different cathepsin structures is to superimpose the structures based on either their c-alpha or backbone atoms. Since different cathepsins have different numbers and types of residues, the residues to be used must be known to map the coordinates from different cathepsin structures. DS provides a method for superimposing cathepsin structures specific matching residues between two cathepsins and superimposes them based on their sequence alignment, and then the matching residues for superimposition are automatically identified based on the aligned residues in the sequence alignment. The root-mean-square deviation (RMSD) was determined for the superimposition and the number of residues used to superimpose the cathepsins is used to measure the differences between values predicted by a model and the values actually observed.

2.2. Virtual screening study

The studies of three-dimensional structure of cathepsins papain C1 family involves the assignment of cathepsins have already been deposited and compared with no X-ray crystal structures of cathepsins have been deposited in public Protein Data Bank to account for protein structures. To understand molecular recognition in cathepsin-substrate system for structure-based drug design, however, it is necessary to consider the interdependence of binding interactions of all cathepsins papain C1 family. This dynamics aspect of molecular recognition of cathepsin S is one of least well understood aspect of molecular recognition. Structure-based virtual screening involves docking of candidate ligands into a protein target followed by applying a scoring function to estimate the ligand will bind to the protein with high affinity. The drug subset of the ZINC database was selected to study [19]. Cathepsin S, the cysteine proteinase has been used as target for virtual screening in this study. The coordinate the X-ray crystal structure of this enzyme (PDB 2R9M) was retrieved from the Protein Data Bank. Active compounds with molecular weights between on 200-600 daltons were selected as drug-like compounds from ZINC database. DOCK program (version 6.1) were selected for molecular docking studies [20]. The small molecules from drug-like compounds of ZINC database were initial screen based on method of Lipinski’s rule of five including ADME-Tox [21]. The small molecules which have heavy metal and toxic atom, aromatic ring more than 2 rings, the ring size more than 6 carbons and the charge more than +2 and less than -2 were removed from the database. The screening active compounds (681,158 molecules) were performed virtual screening with flexible docking and with grid energy scoring function. The final coordinates of each molecule were then stored in multi-mol2 files. The top-ranked 10 drug-like compounds were collected based on free energy binding and grid score for the last step. The energy scores and contact scores of DOCK were used in this work.

2.3. Molecular dynamics simulations studies

All simulations in this research were performed using the GROMACS (version 3.3.3) package [22]. The OPLS-AA/L all-atom force field was used for protein and SPC set as water model. The amino acids name Lys, Arg and Gln were set to +1 while Glu and Asp were set to -1. The minimum angle for hydrogen bonding was set to 135° and maximum distance of donor-acceptor was set to 30 Å. The box model was set to cubic by specified the distance between the solute and the box as 8 Å and also set the protein to the center of the system. The atom ion name Na⁺ and Cl⁻ were selected for this research by specified salt concentration as 0.02 M and set the option (-neutral) until the system became to neutralize. In order to performed molecular dynamic simulations, first step, the system was run energy minimization. In the second step, the output obtained from energy minimization was continued with pre-MD simulation until the solvent system became to equilibrium. In the last step, the output obtained from pre-MD simulation was continued with full molecular dynamic simulations from 0 until 20 ns.
3. Results

3.1. Prediction of three-dimensional structure of cathepsins by homology modeling method

The results of prediction three-dimensional structure of cathepsins in C1 family that have not been determined structures by X-ray crystallography were shown in Fig 1. The structure of cathepsins were generated by homology module using Insight II program based on the X-ray structure which showed the highest identity percentage, as shown in Table 1. The three-dimensional structures of cathepsin in C1 Family have been solved by X-ray crystallography available at Protein Data Bank such as cathepsins V, X, S, L, K, H, F and B. According to superimpositions of cathepsin in C1 family, we found that the structure of cathepsins have been identified into 3 groups; group 1 showed % identity less than 20 Å with model of three-dimensional structure of cathepsins 1, 2, 3, 6, F, H, L, K, O, P, Q, R, S, V and W. Group 2 has only cathepsin B as a member and group 3 has only cathepsin X as a member because they both have % identity more than 20 Å, when compared with the another as shown in Fig 2.

Fig. 1. Overall predictions models of three-dimensional structures of cathepsin 1, 2, 3, 6, M, O, P, Q, R and W by Homology Modeling method. The catalytic dyads (cysteine and histidine were shown under structure.
Table 1. Templates used for three-dimensional structure prediction of each sequences of cathepsins.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Enzyme Template calculated by Geno 3D server</th>
<th>% identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathepsin 1</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
<tr>
<td>Cathepsin 2</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
<tr>
<td>Cathepsin 3</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
<tr>
<td>Cathepsin 6</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
<tr>
<td>Cathepsin Q</td>
<td>3hha</td>
<td>Cathepsin L</td>
</tr>
<tr>
<td>Cathepsin P</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
<tr>
<td>Cathepsin M</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
<tr>
<td>Cathepsin R</td>
<td>3hha</td>
<td>Cathepsin L</td>
</tr>
<tr>
<td>Cathepsin W</td>
<td>1me3</td>
<td>Cruzain</td>
</tr>
<tr>
<td>Cathepsin O</td>
<td>1fh0</td>
<td>Cathepsin V</td>
</tr>
</tbody>
</table>

Fig. 2. Superimposition of all cathepsins three-dimensional structures in papain C1 family based on carbon backbone.

3.2. Virtual Screening

The results of virtual screening studies between cathepsin S and 681,158 active compounds as drug-like compounds from ZINC database, which are drug now subset and initial screen based on method of Lipinski’s rule of five including ADME-Tox was studied. Drug-like compounds are interacts with cathepsin S, that have lowest energy value top 10 ranked compounds as shown in Table 2. The binding site of cathepsin S for drug-like compounds interactions were shown in Fig 3. The result showed indicating that drug-like compounds bound in the active site within 2 parts. These parts are S and S' subsites.
### Table 2. Top 10 rank drug-like compound molecules interactions with cathepsin S.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Zinc code</th>
<th>IUPAC Name</th>
<th>Grid Score</th>
<th>Free energy binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ZINC23215439</td>
<td>(S)-2-[(2,5-dimethylphenyl)amino]-5-[[3-methoxy-4-(2-morpholino-2-oxo-ethoxy)phenyl]methylene]thiazol-4-one</td>
<td>-38.393898</td>
<td>-39.092602</td>
</tr>
<tr>
<td>2</td>
<td>ZINC25514259</td>
<td>N-(2,3-dihydro-1,4-benzodioxin-6-ylmethyl)-N-methyl-2-(2-oxo-1,3-dihydrobenzimidazol-5-yl)sulfonylamino]benzamide</td>
<td>-37.235909</td>
<td>-32.429466</td>
</tr>
<tr>
<td>3</td>
<td>ZINC09185764</td>
<td>5-(2-furyl)-4-(damanti-phenyl-methylene)-1-{2-(1H-indol-3-yl)ethyl}pyrroli dine-2,3-dione</td>
<td>-36.491997</td>
<td>-32.623199</td>
</tr>
<tr>
<td>4</td>
<td>ZINC22681343</td>
<td>3-(1,3-benzodioxol-5-yl)-N-[(2R)-2-(4-methyl-1-piperidyl)-2-(2-thienyl)ethyl]propanam id e</td>
<td>-36.359833</td>
<td>-34.095329</td>
</tr>
<tr>
<td>5</td>
<td>ZINC20057950</td>
<td>1-[[4-(1,3-benzothiazol-2-ylmethyl)piperazin-1-yl]methyl]-2-(phenoxymethyl)benzimidaz ole</td>
<td>-36.338936</td>
<td>-33.565563</td>
</tr>
<tr>
<td>6</td>
<td>ZINC22918093</td>
<td>1-[[4-(1,3-benzothiazol-2-ylmethyl)piperazin-1-yl]methyl]-2-(phenoxymethyl)benzimidaz ole</td>
<td>-36.169193</td>
<td>-35.388580</td>
</tr>
<tr>
<td>7</td>
<td>ZINC13123165</td>
<td>(4E,5R)-4-[damanti-(p-tolyl)methylene]-1-{{2-(1H-indol-3-yl)ethyl}-5-(3-pyridyl)pyrrolidine-2,3-dione</td>
<td>-36.140175</td>
<td>-34.131016</td>
</tr>
<tr>
<td>8</td>
<td>ZINC23640419</td>
<td>N-[4-[[5-chloro-2-thienyl]methyl]piperazin-1-yl]-4-oxobutyldamantine-1-carboxamide</td>
<td>-35.677483</td>
<td>-35.067104</td>
</tr>
<tr>
<td>9</td>
<td>ZINC20855900</td>
<td>N-[3-[(3R,5R)-3,5-dimethyl-1-piperidyl]propyl]-1-(1H-imidazol-4-yl)sulfonylimidazol-4-carboxamide</td>
<td>-35.623291</td>
<td>-31.000479</td>
</tr>
<tr>
<td>10</td>
<td>ZINC09124349</td>
<td>N-[(1-methyl-2-(2-morpholinoethyl)benzimidazol-5-yl)-2-(4-methyl-1-oxo-phthalazin-2-yl)-acetamide</td>
<td>-35.349407</td>
<td>-33.785519</td>
</tr>
</tbody>
</table>

Fig. 3. The result from virtual screening of cathepsin S with drug-like compounds. There are top 100 ranked superimposed from free energy binding. The structure of cathepsin S was shown in electrostatic surface by representing of drug-like compounds with stick. The binding mode was divided into 2 parts, S and S’ subsites. These parts were shown with yellow dashed circle.
Fig. 4. (A) Electrostatic surface representation of cathepsin S in complex with top-ranked 1, ZINC 23215439. The ZINC 23215439 was shown fit well in the part of S and S’ subsites of the active site of cathepsin S. (B) The molecular docking model of interactions in a cathepsin S-ZINC23215439 inhibitors adsorptive complex. Catalytic dyads of cathepsin S amino acid residues Cys25 and His164 are formed hydrogen bonds towards the O2 and O4 of ZINC23215439, respectively. The P1 subsite of ZINC23215439 inhibitor is formed hydrogen bond towards the Trp186 and Gln19.
Fig. 5. (A) Electrostatic surface representation of cathepsin S in complex with top-ranked 25, ZINC 06660322. The ZINC06660322 was shown fit well in the part of S and S' subsites of the active site of cathepsin S. (B) The molecular docking model of interactions in a cathepsinS-ZINC06660322 inhibitors adsorptive complex. The P1 and P'2 subsites of ZINC06660322 inhibitor are formed hydrogen bonds towards the Phe146, Trp186 and Gln19.
Fig. 6. Coulomb-SR interactions energy of amino acids and drug molecule ZINC 23215439 (blue), ZINC 25514259 (red) and ZINC 09185764 (green) which consisted in active site of cathepsin S were evaluated from molecular dynamics simulations during 20 ns.

Fig. 7. Lennard-Jones-SR interactions energy of amino acids and drug-like molecule ZINC 23215439 (blue), ZINC 25514259 (red) and ZINC 09185764 (green) which consisted in active site of cathepsin S were evaluated from molecular dynamics simulations during 20 ns.
The ribbon structure of cathepsin S complex with the best drug-like compound, ZINC 23215439 was shown in Fig 4A. The molecular docking model of interactions in cathepsin S-ZINC 23215439 inhibitor adsorptive complex was shown in Fig 4B. The binding interactions between of cathepsin S and ZINC 23215439 complex showed 7 hydrogen bonds within 3.5 Å. The amino acid residues of cathepsin S which form hydrogen bonds and interact with drug-like compound ZINC 23215439 are Cys 25…O2, Trp168…S1, Gln19…S1, Gln19…N1 and His164…O4. Comparison with the ribbon structure of cathepsin S complex with top-rank 25 drug-like compounds ZINC06660322 was shown in Fig 5 A and B. The result was shown that missing on hydrogen bonding interactions in between of active site amino acid residues of cathepbin S and ZINC06660322. Coulomb-SR interactions energy amino acids in active site of cathepsin S-ZINC 23215439, 25514259 and 09185764 complex were shown in Fig 6. The result of Lennard-Jones-SR interactions energy of amino acids in active site of cathepsin S–complex with ZINC code 23215439, 25514259 and 09185764 was shown in Fig 7. Comparison of Root mean square deviation (RMSD) of amino acids in active site of cathepsin S and drug-like molecules ZINC 23215439, 25514259 and 09185764 were evaluated from molecular dynamics simulations studies on during 20 nanosecond as shown in Fig 8.

4. Conclusions

The characteristics of active site and structure-based virtual screening of cathepsin S reactions with small molecules of drug-like compounds from the ZINC database Version 8 were investigated. The S and S’ subsites of cathepsin S were clearly classified based on molecular docking and virtual screening. Several of amino acid residues in the S subsite represented the hydrophobic pocket. More 60 % of the active small molecules based on virtual screening preferred S subsite rather than S’ subsite. The 10 top ranked drug-like compounds were classified to possess the lowest free energy of binding and grid score for cathepsin S. The result of top 1 ranked drug-like compounds reaction with cathepsin S was also studied by molecular docking. The results of these
studies are compared with those of top 2 and 3 ranked drug-like compounds from ZINC database in which that hydrogen bonds are essential for binding interactions between amino acid residues Cys25, His164, Gln19, Trp186 and Phe 146 of cathepsin S and inhibitors. Root Mean Square Deviation, Coulomb-SR and Lennard-Jones-SR interactions energy of amino acids and drug molecule ZINC 23215439 which consisted in active site of cathepsin S have been evaluated from molecular dynamics simulations during 20 nanoseconds. The values indicated that drug molecule ZINC 23215439 in which that hydrophobic interaction are essential for binding interactions between cathepsin S and inhibitors. The potentials of these cathepsin S inhibitors for anticancer therapy are now under intensive investigation in many academic institutions and pharmaceutical companies, and that attention is also being paid to the possibility of these inhibitors in controlling the inhibition and progression of immune disorders such as multiple sclerosis, rheumatoid arthritis, allergic asthma and pain.

Acknowledgements

The authors thanks University of the Thai Chamber of Commerce, King Mongkut’s University of Technology Thonburi and National Center for Genetic Engineering and Biotechnology, NSTDA, Thailand for the support and facilities.

References


