Identification of small molecule inhibitors of CXCR4 – an important drug target in renal fibrosis

Senthil Kumar
SASTRA Deemed University

Elakkiya E
Pondicherry University

Krishna Kant Gupta (krishnakant@scbt.sastra.edu)
SASTRA Deemed University

Research Article

Keywords:

Posted Date: January 9th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2442430/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

The final stage of almost all chronic kidney diseases is renal fibrosis (CKD). Following tissue inflammation, the healing process leads to fibrosis. Simple wounds or persistent inflammation can cause tissue inflammation, which, in the case of the kidney, results in scarring. Vascular sclerosis, tubulointerstitial fibrosis, and glomerular fibrosis are all types of kidney fibrosis. Tubular atrophy, tubular dilatation, interstitial leukocyte infiltration, fibroblast accumulation, vascular rarefaction, and persistent matrix protein deposition make up the tubulointerstitial fibrosis histological appearance. Renal damage will therefore be exacerbated and fibrosis will be encouraged by persistently elevated Cxcr4 expression (on tubules or immune cells like macrophages). Since various effector cells, including tubular and infiltrating lymphoid cells, are involved in fibrosis, blocking this pathway should reduce it. This study aimed to identify possible pharmacological agents which could bind to and inhibit isoform I of CXCR4 and determine their strength of interactions. The I-TASSER, Phyre and Robetta were used to predict and refine the structure of the CXCR4 protein. ModBase was used to improve the loops, and then the quality was evaluated using the ERRAT value. The improved 3D structure was subjected to small molecule database docking using Maestro (from Schrodinger) and the glide module. GROMACS was used to simulate molecules with the lowest glide scores and the best ADME properties. For docking studies, we employed the CXCR4 protein 3D rendered structure, which had an ERRAT score of 92.15%. The maximum glide score was achieved by the ligand 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE, which was followed by 1-CYCLOHEXYL-4-[(2-NITROPHENYL)METHYL]PIPERAZINE. GROMACS simulation simulations revealed that 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE and II-TASSER interacted in a more stable manner.

1. Introduction

The GPCR superfamily’s chemokine receptors and ligands work together to create chemokine networks that control a variety of cellular activities, immunological responses, and physiological processes. Chemokine receptors are categorised similarly to their ligands, and they can be further separated into CR, CCR, CXCR, and CX3CR receptors depending on the chemokine subfamily to which they bind [1]. Particular residues have been found to be crucial for receptor signalling, particularly for the 352 total residues of CXCR4 (comprising 81 mostly extracellular residues) (4). These receptors are crucial for a number of physiological and pathological processes that promote cell migration because of their tight links to cell mobility. The CXCR4 has been reported to promote tumor growth [2]. The GPCR superfamily of proteins, the biggest class of integral membrane proteins encoded in the human genome, includes the CXCR4-chemokine receptor 4. (CXCR4).CXCR4 controls the development of the haematological and nervous systems as well as stress tolerance and cellular processes include cell migration, chemotaxis, differentiation, growth, activation, proliferation, and apoptosis [3].

The G protein-coupled chemokine (C-X-C motif) receptor (CXCR4) is considerably changed by kidney injury, and ongoing activation of CXCR4 expression worsens the fibrotic response [4]. According to studies, CXCR4 gene and protein expression in nephron tubular cells was considerably downregulated by unilateral ureteral obstruction (UUO) (5). CXCR4/CXCR7/SDF-1 has recently emerged as a key regulator of both renal developments [5]. Targeting the Ang II/AT1 receptor/CXCR4 pathogenic pathways in severe glomerular proliferative disorders that typically cause rapid loss of renal function needing dialysis or transplantation may be beneficial. The inhibition of CXCR4 protein stops the formation of calcium oxalate crystal in Kidneys [6].

Compared to original tumours and lymph node metastatic lesions, cancer cells present in bone metastatic lesions show higher expression of CXCR4 in prostate cancer [7]. Rapid morbidity and death are linked to acute kidney injury (AKI), which is defined by a quick loss of renal function. The AKI pathophysiology has been linked to both the tumour suppressor p53 and the chemokine receptor CXCR4. The CXCR4 antagonist polymeric CXCR4 antagonist (PCX) /siRNA nanoparticles (polyplexes) defend renal injury [8]. CXCR4 is thus a potential target in nephrolithiasis patients to prevent renal fibrosis. CXCR4 expression is elevated in the renal tissue of hyperoxaluric mice and people with nephrolithiasis [9].

CXCR4, one of the chemokine receptors that has drawn the most attention, is crucial in the pathogenesis of liver disease in addition to helping to attract immune cells. Treatment strategies for liver illness nowadays typically focus on regulating certain liver cells’ important functions, where the CXCR4 pathway plays a crucial role. There is strong evidence that the liver diseases cirrhosis, liver fibrosis, liver injury, and hepatitis are all impacted by CXCR4 and its ligand. Accordingly, CXCL12 carefully controls signal transduction by activating CXCR4, which subsequently aids in the onset of liver disease, depending on the internal and external conditions of cells as well as the pathophysiology of the body [10]. CXCR4 and CXCL12 have been related to angiogenesis and the interaction between cancer cells and the microenvironment of tumours promotes spread of tumours. CXCR4 inhibition is crucial in a therapeutic setting, as is being done. For cancer cells to connect with their environment, tumour and tumor-supporting cells must express CXCR4 and CXCL12 [11].

The objectives of this study were to identify potential pharmacological compounds that could bind to and inhibit the CXCR4 protein, and assess the potency of interactions between these compounds. Phyre2 and 3D refine were used to predict and refine the structure of the CIITA-I isoform. ModBase was used to improve the loops, and then the quality was evaluated using the ERRAT value. The improved 3D structure was subjected to small molecule database docking using Maestro (from Schrodinger) and the glide module. GROMACS was used to simulate molecules with the lowest glide scores and the best ADME properties. For docking studies, we employed the CXCR4 protein 3D refined structure, which had an ERRAT
score of 92.15. The ADME following ligands 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE achieved the highest glide score − 11.5, which was followed by 1-CYCLOHEXYL-4-[(2-NITROPHENYL)METHYL]PIPERAZINE (-8.3). The docked molecules were used as input for MM-GBSA calculation. The interactions between 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE and CXCR4 protein were more stable, according to GROMACS simulation simulations. The top simulated complexes were considered for MM-PBSA calculation and residue per decomposition graphs were plotted. Based on our findings, we propose 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE and 1-CYCLOHEXYL-4-[(2-NITROPHENYL)METHYL]PIPERAZINE as potential modulators for CXCR4 protein.

2. Materials And Methods

2.1. Structure prediction

CXCR4's amino acid sequence was obtained from Uniprot (https://www.uniprot.org/). Its 3D structure was predicted using the Galaxy Web Server - TBM (Template Based Modelling) [12], I-TASSER (Iterative Threading ASSEmbly Reninement)[13], and Robetta Server [14]. Each server provides five distinct types of models PDB. For structure validation, the ERRAT quality score (assuming > 80% as an acceptable score) [15] and the Ramachandran plot server (https://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/ramachandran/ramachandran.html) were used. The structure was refined further using the GalaxyWEB refine server. Based on hydrogen bonding and atomic level energy minimization, the server refines the predicted structure. The SAVES server v 6.0 (http://nihserver.mbi.ucla.edu/SAVES) was used to evaluate the protein structure's quality.

2.2. Identification of ligand binding pockets

On the CASTp (Computer Atlas of Surface Topography of Proteins, v 3.0) server, the refined 3D structure of CXCR4 (.pdb format) was used to identify ligand binding pockets [16]. CASTp uses computational geometry to characterise the interior voids and surface pockets of proteins. Interproscan was used to identify domains furthermore [17].

2.3. Small molecule databases

Small molecule databases are an essential resource for virtual lead molecule screening. High throughput screening was done using a molecular database library from Chem Div (.sdf file contain 23,839 molecules; http://www.chemdiv.com/catalog/focused-and-targeted-libraries/anti-inflammatory-library).

2.4. Molecular docking

Schrodinger Maestro v 12's glide module was used to dock small molecules with CXCR4. The receptor grid generation module was used to create a grid for the CXCR4 amino acids D97, D187, and E288. With the help of Glide [18], millions of compounds can be effectively screened using high-throughput virtual screening (HTVS) [19]. In this case, we used HTVS for about 23,839 molecules. The grid's size was kept constant at 20 Å. A 2D interaction diagram in Schrodinger was used to find the interactive residues.

2.5. ADME property

Knowing the properties of potential drug molecules' absorption, distribution, metabolism, and elimination (ADME) would be crucial to determining their biocompatibility. In order to determine the reference values for the candidate ligands, ADME properties were examined using ADMETlab 2.0 server [20].

2.6 Simulation of native protein

To better understand molecular stability, GROMACS 2020.3 was used to simulate the CXCR4 model's three-dimensional structure while employing the OPLS all-atom force field. GROMACS is a powerful programme for computing the dynamics of biomolecules. Studies on protein-ligand interactions and small molecule solvation are highly accurate as a result of the OPLS4 implementation [21]. Here, a periodic body with a dimension extending to 1.50 nm was used, and the protein was enclosed in a cube. With the help of sodium ions, the water molecules were neutralised. The system underwent 50,000 steps of energy minimization using the steepest descent algorithm, then 50,000 steps of minimization using the conjugate gradient method. For positional restraint, the minimised systems were then equilibrated for 100 picoseconds (ps) at NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature). The LINCS algorithm [22] was used to constrain all bond angles, and the SETTLE algorithm [23] was used to constrain the geometry of water molecules. The system's temperature (at 310 K) was controlled by a V-rescale, in which velocity is rescaled by a carefully selected random factor. Performance was excellent regardless of thermostat and property dynamics. The pressure (at 1 atm) was set using the Parrinello-Rahman method [24], which is the pressure-induced structural transition that elaborates the metadynamics of pressure-induced changes. V-rescale is a weak coupling method. The equilibrated systems were then prepared for a simulation run with a time step of 170 nanoseconds (ns) and a resolution of 2 femtoseconds (fs). The GROMACS package's appropriate tools were used to save and analyse structural coordinates every two ps. The CXCR4 structure with the lowest energy was located.

2.7. Simulation of best three docked complexes
In order to comprehend the molecular stability of simulated three-dimensional complex structures, molecular dynamics simulations using the all-atom force field were conducted using GROMACS 2020.3. The topology of the protein and ligand was first built. Next, ligand topology was generated using the LigParGen server [25]. The protein-ligand complexes were submerged in a 1.50 nm-tall cubic box of solution. All directions used a periodic boundary system. By substituting sodium ions for the water molecules in the solvated systems, the systems were made neutral. The remaining steps were identical to those in section 2.6.1. The structural coordinates for every 2 ps were saved and analysed using the GROMACS package's appropriate tools. The RMSD, RMSF, H-bond, Radius of gyration, clustering, binding energy, Free Energy Landscape (FEL) analysis, Principal Component Analysis (PCA), Solvent Accessible Area (SASA), and Molecular Mechanics Poisson-Boltzmann Surface Area were used to compare the minimum energy complex structures (MMPBSA).

3. Results And Discussion

3.1. Structure prediction

CXCR4 has a remarkable capacity to recognise a variety of proteins, peptides, and tiny compounds that are unrelated to one another. Due to receptor conformational plasticity involving changes to the receptor side-chain and backbone, the ligands occupy various areas of the binding pocket while still binding to a conserved set of binding determinants. Due to its adaptability, the receptor may accept ligands from several classes, including allosteric inhibitors and chemokines of the CC and CXC types. The increasing diversity of ligands for chemokine receptors creates opportunities for the rational design of ligands with better inhibitory profiles and mechanisms of action [26].

The sequence of CXCR4 protein with Uniprot ID P61073 has 352 amino acids. A 38-residue, sulftosyrine-containing peptide generated from the CXCR4 N-terminus was complexed with CXCL12 to form its NMR structure (PDB ID: 2K05) [27] but the full-length experimental structure is not available. Therefore, we have modelled the protein. The tertiary structure of CXCR4 obtained from GalaxyWEB-TBM (Fig. 1), indicating a low confidence level. The Ramachandran plot for these structures showed more disallowed regions than the raw structure. Since these scores are not considered suitable for docking.

The specific portions of the protein were further refined using the GalaxyWEB-refine was obtained. The optimised structure was chosen based on the highly preferred observations green crosses: 316 (98.442%) score in the ramachandran plot server (Supplementary Fig. 1). Superimposing the raw and refined structure revealed a root mean square deviation (RMSD) of 0.3. Positions of alpha helix differed a bit between the two structures (Fig. 2).

3.2. Druggable site identification

The drug targeting site can be an active or allosteric site that influences protein function. Putative pockets for ligand interaction were identified by CASTp analysis, which identified forty-six pockets. Pocket 1 was chosen for docking as it possessed the largest surface area and volume (Fig. 3). The three residues namely D97, D187 and E288 plays an important role in CXCR4 signalling [28]. The identification of inhibitors binding site in the crystal structure of CXCR4 (PDB ID: 3OE8) has been reported [29]. Interproscan revealed that a portion (residue 6–37) contains the chemokine receptor 4 N-terminal domain (IPR022726) and (residue 55–302) GPCR, rhodopsin-like, 7TM (IPR017452) domain. Interestingly, this domain has been implicated in programmed cell death [30].

3.3. Screening of anti-inflammatory molecules

CXCR4 is engaged in the persistent inflammation of the arterial wall, which is characterised by a chemokine-mediated influx of leukocytes, in the aetiology of atherosclerosis [31]. Additionally, CXCR4 has been found as a significant factor in the development of aneurysms, atherosclerotic plaque instability, and vascular remodelling following injury. Additionally, persistent inflammation and localised immune cell infiltration with CXCR4 expression substantially encourage the development of esophageal cancer [32]. CXCR4 overexpression has been discovered to significantly contribute to renal injury and neurodegenerative illnesses in addition to its involvement in a number of inflammation-related processes [33].

Therefore, we have focussed on anti-inflammatory molecules for screening. Libraries containing anti-inflammatory molecules (23,839) were selected for docking studies due to their relevance for CXCR4 function. Molecules with top 20 glide scores are shown in Table 1. The ADME score for all Top 20 compounds is given in Table 2. The ADME following top three compounds are given in Table 3. In table, the table heading, green colour indicates safe; orange colour indicates partially safe; red colour indicates risk. The 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE, showed maximal glide score ~ 11.5 followed by 1-CYCLOHEXYL-4-[(2-NITROPHENYL)METHYL]PIPERAZINE which showed a glide score of -8.2 and 1-BENZYL-7-OXO-N-[2-(PYRROLIDIN-1-YL)ETHYL]AZEPANE-2-CARBOXAMIDE, showed a glide score of -7.0. The 1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE interacts with CXCR4 via TWO hydrogen bonds at GLU288 and TRP94 (Fig. 4a; 2D and 3D diagram), 1-CYCLOHEXYL-4-[(2-
NITROPHENYL)METHYL]PIPERAZINE interacts with CXCR4 via same two hydrogen bonds at GLU288 and TRP94 (Fig. 4b), and 1-BENZYL-7-OXO-N-[2-(PYRROLIDIN-1-YL)ETHYL]AZEPANE-2-CARBOXAMIDE interacts with CXCR4 via three hydrogen bonds at GLU288, TRP94 and ASN101 (Fig. 4c). The Piperazine compounds also exhibit anti-tubercular and anti-depressant activity [34, 35].

Table 1
Top scoring ligands

<table>
<thead>
<tr>
<th>Entry ID</th>
<th>PubChem CID</th>
<th>Name</th>
<th>Glide Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>4990</td>
<td>740944</td>
<td>1-[(4-ETHYLPHENYL)METHYL]-4-[(3-NITROPHENYL)METHYL]PIPERAZINE</td>
<td>-11.5</td>
</tr>
<tr>
<td>1597</td>
<td>46073602</td>
<td>1-[4-(2,4-DIFLUOROPHENOX)-2-(4-METHYLPIPERAZIN-1-YL)-5H,6H,7H,8H-PYRIDO[4,3-D]PYRIMIDIN-6-YL]-2-(4-METHOXYPHENYL)ETHAN-1-ONE</td>
<td>-10.9</td>
</tr>
<tr>
<td>20574</td>
<td>136665827</td>
<td>7-[(3-CHLOROPHENYL)METHYL]-2-(4-METHYLPIPERAZIN-1-YL)-3H,4H,5H,6H,7H,8H,9H-PYRIMIDO[4,5-D]AZEPIN-4-ONE</td>
<td>-10.6</td>
</tr>
<tr>
<td>18401</td>
<td>3240862</td>
<td>1-[5-(3-CHLOROPHENYL)-1,2-OXAZOLE-3-CARBONYL]-4-METHYLPIPERAZINE</td>
<td>-9.2</td>
</tr>
<tr>
<td>22857</td>
<td>50801419</td>
<td>3-[2-(1-METHYL-1H-INDOL-3-YL)-2-(PYRIDIN-3-YL)ETHYL]-1-PHENYLUREA</td>
<td>-8.7</td>
</tr>
<tr>
<td>23537</td>
<td>50801367</td>
<td>3-(3-BROMOPHENYL)-1-[2-(1-METHYL-1H-INDOL-3-YL)-2-(PYRIDIN-3-YL)ETHYL]UREA</td>
<td>-8.6</td>
</tr>
<tr>
<td>2938</td>
<td>16764992</td>
<td>2-[[2,1,3-BENZOTHIADIAZOL-5-YL]METHYL]SULFANYL]-4H-1,3-BENZOTHIAZIN-4-ONE</td>
<td>-8.2</td>
</tr>
<tr>
<td>4993</td>
<td>880196</td>
<td>1-CYCLOHEXYL-4-[(2-NITROPHENYL)METHYL]PIPERAZINE</td>
<td>-8.2</td>
</tr>
<tr>
<td>4661</td>
<td>695088</td>
<td>1-[3-FLUOROPHENYL]METHYL]-4-[4-NITROPHENYL]METHYL]PIPERAZINE</td>
<td>-7.8</td>
</tr>
<tr>
<td>21889</td>
<td>124063622</td>
<td>1-BENZYL-7-OXO-N-[2-(PYRROLIDIN-1-YL)ETHYL]AZEPANE-2-CARBOXAMIDE</td>
<td>-7.0</td>
</tr>
<tr>
<td>4239</td>
<td>71686729</td>
<td>1-[2-(METHYLPHENYL)METHYL]-2-[(4-METHYLPIPERAZIN-1-CARBONYL]-1-PHENYLUREA</td>
<td>-6.6</td>
</tr>
<tr>
<td>1702</td>
<td>53137502</td>
<td>4-CYANO-N-[4-(MORPHOLIN-4-YL)QUINAZOLIN-7-YL]BENZAMIDE</td>
<td>-5.5</td>
</tr>
<tr>
<td>22407</td>
<td>87051447</td>
<td>6-METHYL-N-[(4-METHYLGLYCOLAMIDO)PHENYL]METHYL]IMIDAZO[1,2-A]PYRIDINE-2-CARBOXAMIDE</td>
<td>-5.0</td>
</tr>
<tr>
<td>5746</td>
<td>135511588</td>
<td>2-AMINO-5-(5-BROMO-2-FLUOROPHENYL)-3H,4H,5H,6H,7H,8H-PYRIDO[2,3-D]PYRIMIDINE-4,7-DIONE</td>
<td>-4.4</td>
</tr>
<tr>
<td>22192</td>
<td>135511596</td>
<td>2-AMINO-5-(2-ETHOXYPHENYL)-3H,4H,5H,6H,7H,8H-PYRIDO[2,3-D]PYRIMIDINE-4,7-DIONE</td>
<td>-4.4</td>
</tr>
<tr>
<td>5747</td>
<td>135511592</td>
<td>2-AMINO-5-(2-METHOXYPHENYL)-3H,4H,5H,6H,7H,8H-PYRIDO[2,3-D]PYRIMIDINE-4,7-DIONE</td>
<td>-4.3</td>
</tr>
<tr>
<td>18449</td>
<td>2738592</td>
<td>4-CHLORO-N-[4-(METHYLPIPERAZIN-1-YL)PHENYL]METHYL]BENZENESULFONAMIDE</td>
<td>-3.9</td>
</tr>
<tr>
<td>7055</td>
<td>46257553</td>
<td>N-[4-(ACETAMIDOPHENYL)-1-(3-METHYL-2-OXO-2,3-DIHYDRO-1,3-BENZOXAZOL-6-YL)SULFONYL]PIPERIDINE-4-CARBOXAMIDE</td>
<td>-3.2</td>
</tr>
</tbody>
</table>
Table 2
ADME study of top 20 ligands (Based on Glide Score) Green colour indicates safe; orange colour indicates partially safe; red colour indicates risk

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Ligands</th>
<th>PubChem CID</th>
<th>Green</th>
<th>Orange</th>
<th>Red</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1597</td>
<td>46073602</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>1702</td>
<td>53137502</td>
<td>8</td>
<td>5</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>2938</td>
<td>16764992</td>
<td>10</td>
<td>1</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>4239</td>
<td>71686729</td>
<td>17</td>
<td>2</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>4661</td>
<td>695088</td>
<td>16</td>
<td>3</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>4990</td>
<td>740944</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>4993</td>
<td>880196</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>5746</td>
<td>135511588</td>
<td>15</td>
<td>1</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>5747</td>
<td>135511592</td>
<td>14</td>
<td>3</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>7055</td>
<td>46257553</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>18401</td>
<td>3240862</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>12</td>
<td>18449</td>
<td>2738592</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>18456</td>
<td>2162624</td>
<td>12</td>
<td>5</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>14</td>
<td>20574</td>
<td>13666827</td>
<td>19</td>
<td>0</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>15</td>
<td>21889</td>
<td>124063622</td>
<td>19</td>
<td>4</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>16</td>
<td>22191</td>
<td>135511596</td>
<td>14</td>
<td>2</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>17</td>
<td>22407</td>
<td>87051447</td>
<td>19</td>
<td>1</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>18</td>
<td>22857</td>
<td>50801419</td>
<td>16</td>
<td>2</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>19</td>
<td>23357</td>
<td>50801367</td>
<td>17</td>
<td>1</td>
<td>5</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 3
Final top 3 ligands Based on ADME and docking study

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Ligands</th>
<th>PubChem CID</th>
<th>Green</th>
<th>Orange</th>
<th>Red</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4990</td>
<td>740944</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>4993</td>
<td>880196</td>
<td>17</td>
<td>3</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>21889</td>
<td>124063622</td>
<td>19</td>
<td>4</td>
<td>0</td>
<td>23</td>
</tr>
</tbody>
</table>

These interacting residue GLU288, plays very important role for fusion elicited by HIV-1 envelope glycoprotein [36]. All the three ligands interacting with these residues which exert their anti-HIV activities also. In all complexes, the ligands were interacting with the active site residue GLU288. In our study, we focused on the prevention of kidney injury. The inhibition of CXCR4 prevents kidney injury due to modulation of leukocyte infiltration and expression of proinflammatory chemokines/cytokines, rather than a HSC-mediated effect [37].

3.4. Simulation of native CXCR4 and best three complexes

Based on docking and molecular dynamics, EPI-X4 peptide was proposed as a promising inhibitor of CXCR4. The peptide was able to bind to D97 residue of CXCR4 [38]. The ID of 1-[(4-ETHYLPHENYL)METHYL]-4-[3-NITROPHENYL]METHYL]PIPERAZINE was 4990. Similarly, the ID of 1-CYCLOHEXYL-4-[2-NITROPHENYL]METHYL]PIPERAZINE was 4993. The ID of 1-BENZYL-7-OXO-N-[2-(PYRROLIDIN-1-YL)ETHYL]AZEPANE-2-CARBOXAMIDE was 21889. The RMSD plot (Fig. 5a) indicates the good stability of the compound 21889 as compared with the other two compounds. The RMSD of all conformations with step size of 10 ns were calculated and compared with 10 ns conformation (Supplementary Fig. 1). The average RMSD for the compounds 21889, 4993 and 4990 were found to be 0.7, 6.5 and 11.7 (Supplementary Table 1). This data strongly suggests the good stability of 21889 with CXCR4. The RMSF plot clearly suggest overall stability of all complexes except CXCR4-4993. The active site residues D97, D187 and E288 was not fluctuating in CXCR4-4990 and CXCR4-21889. The folding of CXCR4-4990 and CXCR4-
4993 was well maintained with time. The solvent accessible surface area got reduced with time. The highest number of H-bond was present in complex CXCR4_4990 followed by CXCR4_4993. The average potential energy of CXCR4 and all the three complexes is as follows: CXCR4: -1.98163e+06, CXCR4_4990: -1.83261e+06, CXCR4_4993: -1.85430e+06 and CXCR4_21889: -1.847295e+06. All these analyses strongly suggest the promising potential of 4990, 4993 and 21889 as inhibitor to CXCR4. People have targeted CXCR4 for COVID-19 therapies, they have reported few inhibitors but safety and stability of these molecules haven’t been investigated [39].

3.4. Principle Component Analysis (PCA) CXCR4 and three complexes

PCA helps us to find functionally relevant motions. During simulation, protein and protein-ligand complexes appear to be moving. It is difficult to discern between local fluctuations and collective motions because they both happen at the same time. In these situations, a principle components analysis can be useful because it can separate local, quick motions from global, collective (often sluggish) motions [40]. The trajectory (md1 backbone.xtc) and the structure (ref.pdb) file were used for PCA analysis.

In Fig. 6 the projection on eigenvector 1 and projection on eigenvector 2 represents the largest amplitude collective motions. It shows PCA for native protein and complexes. The global motion in native protein was less as compared to complexes. For complexes CXCR4-4990 and CXCR4-4993, the projection on eigenvector 1 and projection on eigenvector 2 is mostly positive which clearly indicates more motions after binding. The complex CXCR4-21889 attained positive and negative projection equally in both vectors. It shows motion more than native protein but less than other two complexes.

3.5 Per residue decomposition analysis of three complexes

The energy function in molecular mechanics is a pair-wise additive function with per-residue components. This is important for determining how much each protein residue contributes to ligand binding [41]. From this analysis, we reported that in Total Decomposition contribution (Fig. 7); TRP94, GLU288, TYR45 and ASP97 are the important residues contributing to 4990, 4993 and 21889 binding. Similarly, in Side Chain Decomposition contribution (Fig. 7) same residues plays an important role in ligand binding.

3.6 MM-GBSA of three complexes

The MMGBSA of the three complexes were done to find the ligand binding affinities [42]. In Table 4, The ligand binding affinity of 4990 was better than other two ligands.

<table>
<thead>
<tr>
<th>Complexes</th>
<th>r_psp_MMGBSA_dG_Bind</th>
<th>r_psp_MMGBSA_dG_Bind_Coulomb</th>
<th>r_psp_MMGBSA_dG_Bind_Cовалент</th>
<th>r_psp_MMGBSA_dG_Bind_Hbond</th>
<th>r_psp_MMGBSA_ligand_efficiency_Ln</th>
</tr>
</thead>
<tbody>
<tr>
<td>ligprep_1.maegz:4990</td>
<td>-36.1190422432</td>
<td>-78.0801184995</td>
<td>3.0621877673</td>
<td>-0.4301900774</td>
<td>-8.561295412</td>
</tr>
<tr>
<td>ligprep_1.maegz:4993</td>
<td>-34.5109171155</td>
<td>-78.897123741</td>
<td>2.6722354822</td>
<td>-0.9667112506</td>
<td>-8.4357269593</td>
</tr>
<tr>
<td>ligprep_1.maegz:21889</td>
<td>-17.9352366805</td>
<td>11.5013438204</td>
<td>-0.1972390855</td>
<td>-0.0282814165</td>
<td>-4.2511885689</td>
</tr>
</tbody>
</table>

4. Conclusion

We employed in silico methods in this case to predict the structure of CXCR4. Through molecular docking and molecular dynamics research, we reported promising CXCR4 small molecule inhibitors. Three compounds were found by our investigation to have the ability to target CXCR4 and reduce kidney damage: 4990, 4993, and 21889. Additionally, these compounds demonstrated pharmacological characteristics appropriate for clinical use. To ascertain the safety and effectiveness of these suggested compounds for avoiding renal injury, more experimental research would be beneficial.

Declarations

Acknowledgement

The authors thankfully acknowledge SASTRA Deemed University, Thanjavur, India, for providing infrastructural support and the Schrodinger platform. We thank Bioinformatics Resources & Applications Facility (BRAF) at Centre for Development of Advanced Computing (CDAC), Pune, India, for providing access to GROMACS software.
Disclosure statement

The all authors declare no competing interest.

Funding

The research was not supported by any funding agency.

Authors’ contribution

Krishna Kant Gupta, Elakkiya Elumalai and Senthil Kumar G designed, analysed and interpreted data. Krishna Kant Gupta and Senthil Kumar G wrote the manuscript. Elakkiya Elumalai cross checked the manuscript.

References


**Figures**

![Modelled structure of CXCR4 protein](image-url)
Figure 2

The red colour and green colour refers refined model and un refined model respectively.
Figure 3

Pocket 1 in CXCR4 protein predicted from CASTp server
Figure 4

Figure 5

a) RMSD of CXCR4 (black colour), CXCR4-4990 (red colour), CXCR4-4993 (green colour) and CXCR4-21889. The same colour coding was followed in other plots also. B) RMSF plot c) Radius of gyration d) SASA plot
Figure 6

A) Hydrogen bonds in all the three complexes  B) Potential energy of three complexes
Figure 7

Figure 6: PCA of CXCR4 and three complexes
Figure 7: Total decomposition and Side chain decomposition analysis of three complexes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfigure1.pdf
- Supplementarytable1.xlsx