International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Homology Modeling Studies, Compounds Targeting the CXCR4 Protein of Using Molecular Docking

Kuncoro B[1], Mumpuni E[2], Ratih D[3], Kumala S[4]

[1, 2, 3, 4] Faculty of pharmacy Universitas Pancasila, Jakarta , Indonesia

Abstract: ***The cancer profile in the Indian state of Uttarakhand reveals that the breast cancer is the most prevalent type of cancers in females followed by cervical and ovarian type. Literature survey shows that the CXCR4 protein of is responsible for causing several forms of cancer in human. Therefore, it is of interest to screen CXCR4 target protein with known natural compounds using computer aided molecular modeling and docking tools. The complete structure of CXCR4 is unknown. Hence, the CXCR4 structure model was constructed using different online servers followed by molecular docking five known compounds of4-aminosalicylic acid, eugenol, gallic acid, salicylic acid,gentisic acid, and ifosfamideas reference ligandwith best CXCR4 protein model predicted by Swissmodel Server. The screening exercise shows that eugenol (with reranke score -64,68), a natural compound has the top binding properties.Thus, it is of interest to consider the compound for further validation.***

Instrumentation: Computational chemistry calculations using a Personal Computer with an Intel (R) processor type Core Intel Core i5-6400 @ 2.70 GHz with a 1000 GB Hard disk SSD 512 GB and 16 GB RAM. Programs used include Molegro Virtual Docker 6.0 (MolegroApS), Discovery Studio 2019.

Keywords: CXCR4 Protein, Eugenol, Molecular Docking, Homology Modeling

1. Introduction

Cervical cancer is one of the leading causes of deaths worldwide. The cancer profile of the Indian state of Uttarakhand reveals that the Breast cancer was most prevalent in female followed by cervical and ovarian cancer (1). In our study, five different compounds were collected 4-Aminosalicylic acid, Eugenol, Gallic acid, Salicylic acid, Gentisic Acid. Eugenol, now finds its application as potential anti-cancer compound. It has been observed to induce apoptosis of various cancer cells and cause the modulation of the cell cycle pathway. However the exact mechanism is yet to be studied. In order to understand the potential role of compounds as anticancer molecules, there is a need of computational drug designing tools that can identify and analyze protein-ligand interactions with respect to their binding affinity for investigation of novel drug molecule against CXCR4.

2. Material and Methods

Methods

Sequence retrieval and phylogenetic analysis

Amino Acid sequences of the CXCR4 protein was retrieved in FASTA format *https://www.uniprot.org/*databaseP61073|CXCR4_HUMAN C-X-C chemokine receptor type 4 OS=Homo sapiens OX=9606 GN=CXCR4 PE=1 SV=1 for multiple sequence alignment (2).Uniprot was taken for protein 3D model construction by Swissmodel (3).

Protein structure prediction and validation

The 352 amino acids residue long CXCR4 protein was subjected to BLASTp (http: //blast. ncbi.nlm. nih. gov/Blast.cgi) analysis against PDB database (http://www.rcsb.org /) (4) to identify suitable template for comparative protein modeling.Comparative homology modeling depends on a sequence alignment between target sequence and the template sequence whose 3D structure has been determined by experimental method and protein 3D model construction by Swissmodel (2).

The best model was selected on the basis of Ramachandran plot and protein stability analysis by ProCheck https://servicesn.mbi.ucla.edu/PROCHECK/ (5) and Discovery Studio 2019 visualizer (http:// accelrys.com/products /discovery-studio/) was used for the visualization of modelled protein structure (6). Figure 1.

The three-dimensional crystal structure were loaded in the Molegro virtual docker (MVD) with the removal of all water molecules. The standard Molegro algorithm was utilized for rendering the missing charges, protonation states, and assigning of polar hydrogen to the receptor.

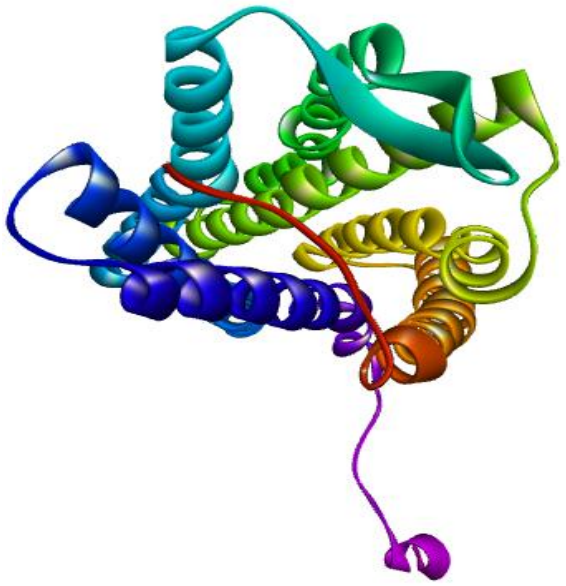
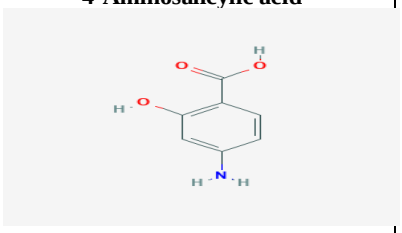
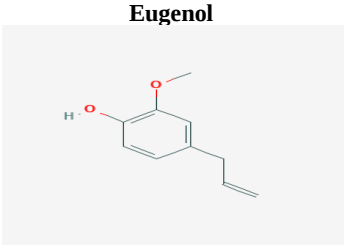
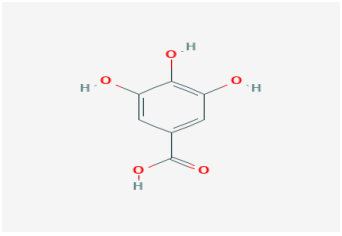
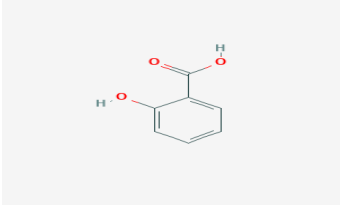
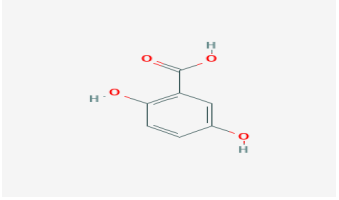
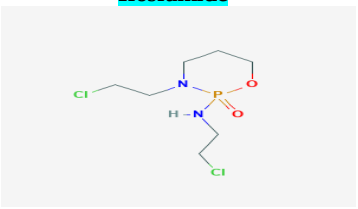


Figure 1: Homology model of CXCR4 by Swissmodel

***Ligand preparation*:**
The Structure of compounds 4-aminosalicylic acid (CID 4649), eugenol (CID 3314), gallic acid (CID370), salicylic acid (CID 338), gentisic acid(CID 3469), and ifosfamide (CID 3690) were retrieved from PubChem database at (http://www.ncbi.nlm.nih.gov/). The mol.2 files of ligands were obtained from PubChem database. Structures of ligands were drawn using marvin sketch and energy minimization was done using MMFF94 force field. Energy minimization is done to help the docking programme for identifying the bioactive conformer from the local minima. One major advantage of MVD is that it helps in assigning the missing bond orders, charges, bonds, and hybridization states of the imported ligands. The 2D structures of 6 ligands are illustrated in Table 1.Compound Structure of test

Table 1: Compound Structure of test

No.	Structure Compound
1	**4-Aminosalicylic acid** 
2	**Eugenol** 
3	**Gallic acid** 
4	**Salicylic acid** 
5	**Gentisic Acid** 
6	**Ifosfamide** 

***Molecular docking*:**
Molecular docking studies were performed using Molegro Virtual Docker 6. Preparation of required input files for Molegro Virtual Docker 6 (7). Preparation of files through Molegro Virtual Docker involved addition of polar hydrogen atoms and gasteiger charges. On the basis of pilot docking studies, the MolDock rerank scores were selected for ranking the inhibitor poses, and for all the docking CXCR4 performed here, the poses selected as the best were taken as those with the best re-rank score. The size of constrain was kept as user defined. It is one of the most important highly cited molecular docking tools for the prediction of proteinligand interaction. It requires the three dimensional structure of both ligand and protein. The results with best conformation and energetic were selected for analysis Discovery Studio 2019 visualizer (http://accelrys.com/products /discovery-studio/) was used for visualization and analysis of protein-ligand complex.

3. Results and Discussion

Molegro Virtual Docker : Molegro Virtual Docker (MVD) was used to perform docking. MVD is an integrated platform for predicting protein - ligand interactions. It handles all aspects of the docking process from preparation of the molecules to determination of the potential binding sites of the target protein, and prediction of the binding modes of the ligand. It provides the user with high-quality docking based on a novel optimization technique combined with a user interface experience focusing on productivity and usability. MVD has been shown to yieldhigher docking accuracy than other state-of-the-art docking products (MVD: 87%, Glide: 82%, Surflex: 75%, FlexX: 58%).

RMSD (Root Mean Square Deviation): The Root Mean Square Deviation (RMSD) is the measure of the average distance between the backbones of superimposed proteins. A widely used way to compare the structures of biomolecules or solid bodies is to translate and rotate one structure with respect to the other to minimize the RMSD. Coutsias, *et al.* presented a simple derivation, based on quaternion's, for the optimal solid body transformation [rotation-translation] that minimizes the RMSD between two sets of vectors.
The equation:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\delta_i^2}$$

Where δ is the distance between N pairs of equivalent atoms [usually *Cα* and sometimes *C,N,O,Cβ*].

Normally a rigid superposition which minimizes the RMSD is performed, and this minimum is returned. Given two sets of *n* points W and V the RMSD is defined as follows:

$$\mathrm{RMSD}(\mathbf{v},\mathbf{w}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|v_i - w_i\|^2}$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$$

An RMSD value is expressed in length units. The most commonly used unit in structural biology is the Ångström (Å) which is equal to10^{-10}m.

Table 2: Validation and Analysis of Docked Receptor-Ligand Complex Structures

RMSD (Angstrom)	MolDock Optimizer	MolDock SE	Iterated Simplex
MolDock Score	1,51	1,02	2,49
MolDock (Grid) Score	1,02	1,59	4,93
PLANTS Score	3,85	3,79	3,69
PLANTS Score (Grid)	0,82	3,79	0,77

The first method validation is re-docking. This procedure is done by trying all combinations of placement scoring and alogirotma available in the Molegro Virtual Docker docking module. A total of 100 docking poses were generated for each combination and an evaluation of the 10 docking poses with the lowest scores. The parameter evaluated is the RMSD value for each docking pose and its mean. The validation results show that the combination of PLANTS Score (Grid) and Iterated Simplex scoring function produces the lowest RMSD (0,77 Å) having an RMSD value <2.0 Å.

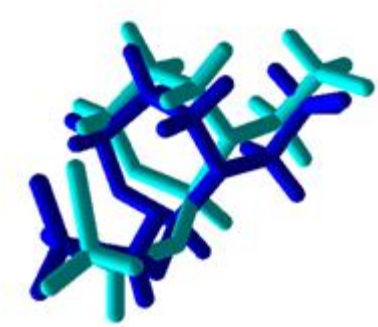


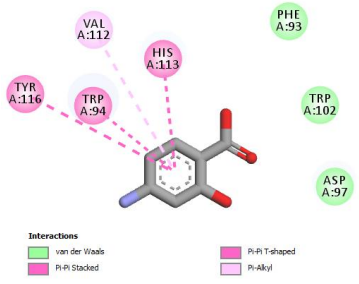
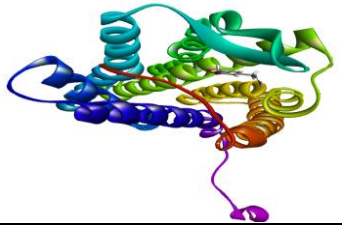
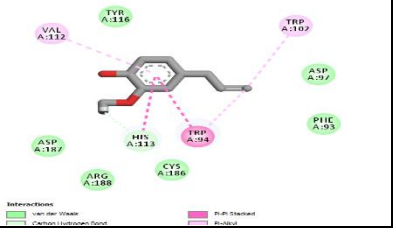
Figure 2: Superimpose native ligand and pose after docking

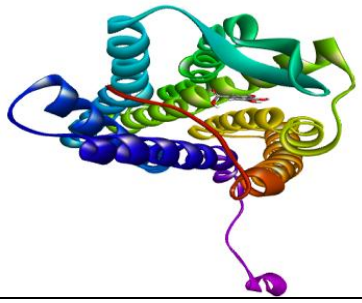
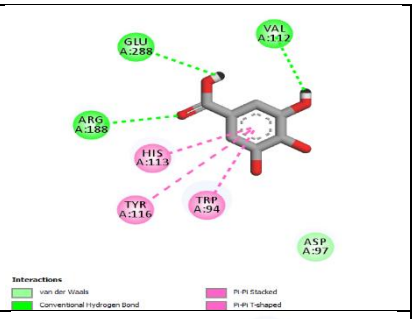
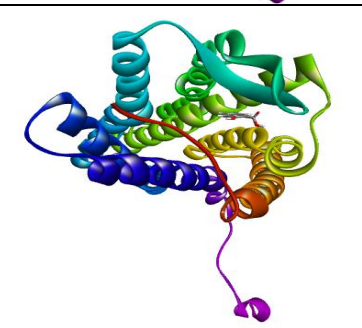
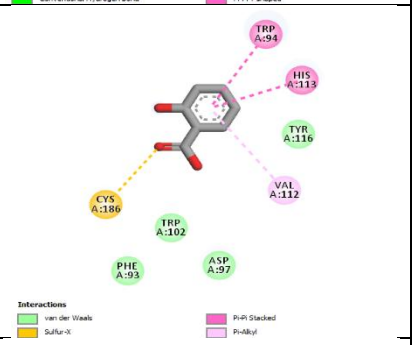
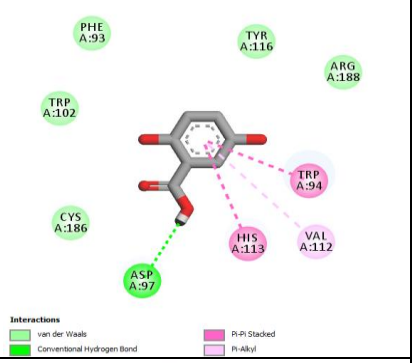
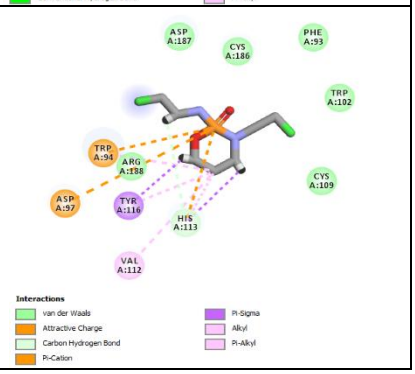
Table 3: Molecular Docking Value of CXCR4 Protein and the tested compounds

No.	Structure Compound	*Rerank score*
1	4-Aminosalicylic acid	-61,99
2	Eugenol	-64,68
3	Gallic acid	-57,86
4	Salicylic acid	-58,01
5	Gentisic Acid	-59,81
6	Ifosfamide	-69.87

The results, concluded which conformation produced the lowest energy state when bound to the target protein, were shown as Rerank Score. In this study, we also examined eugenol, since this compound was also potential to be developed as an anticancer agent.

Table 4: 3 and 2-dimensional amino acid interactions

No.	Compound Structure	3-dimensional amino acid interactions	2-Dimensional amino acid interaction
1	4-Aminosalicylic acid		
2	Eugenol		

3	Gallic acid		
4	Salicylic acid		
5	Gentisic Acid		
6	Ifosfamide		

4. Acknowledgement

Thank you to Prof. Dr. Siswandono, MS., Apt., For granting permission use of a license from Molegro Virtual Docking (MVD) and their guidance so far.

The authors gratefully acknowledge the efforts of Ms. Okta Nursanti who helped a lot for the literature search and the understanding of homology modelling and the other webserver.

5. Conflict of Interest

The authors declare no conflict of interest.

References

[1] Bag A, Rawat S, Pant NK, Jyala NS, Singh A, Pandey KC. Cancer patterns in Nainital and adjoining districts of Uttarakhand: A one year survey. *J Nat Sci Biol Med.* 2012;3(2):186–188. doi:10.4103/0976-9668.101911

[2] The UniProt Consortium **UniProt: a worldwide hub of protein knowledge** Nucleic Acids Res. 47: D506-515 (2019)

[3] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303 (2018).

[4] Bolton EE, Chen J, Kim S, Han L, He S, Shi W, Simonyan V, Sun Y, Thiessen PA, Wang J, Yu B,

Zhang J, Bryant SH. PubChem3D: A new resource for scientists. J Cheminform. 2011 Sep 20;3(1):32. [PubMed PMID: 21933373] doi: 10.1186/1758-2946-3-32.

[5] Laskowski R A, MacArthur M W, Thornton J M (2001). PROCHECK: validation of protein structure coordinates, in *International Tables of Crystallography, Volume F. Crystallography of Biological Macromolecules*, eds. Rossmann M G & Arnold E, Dordrecht, Kluwer Academic Publishers, The Netherlands, pp. 722-725.

[6] *Dassault Systèmes BIOVIA, Discovery studio 2019San Diego: Dassault Systèmes,* (2019).