# INTERNATIONAL JOURNAL
## OF
## PHARMACEUTICAL SCIENCES
## AND
## RESEARCH

# QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP AND MOLECULAR DOCKING ANALYSIS OF CHOLESTEROL INHIBITORS AGAINST NIEMANN-PICK C2 TARGET GENE (NPC2)

Shilu Mathew [1] and Ishtiaq Qadri [*2]

Center of Excellence in Genomic Medicine [1], King Abdul Aziz University, P.O. Box 80216, Jeddah 21589, Saudi Arabia
Medical Biotechnology and Translational Medicine Research [2], King Fahd Medical Research Center, King Abdul Aziz University, PO Box 80216 Jeddah 21589, Saudi Arabia

**ABSTRACT:** In an effort to develop new targets with enhanced antihyperlipidemic activity, seven new inhibitors such as beta-sitosterol, cholesterol, cholesterol sulfate, desmosterol, lathosterol, stigmasterol and cholesterol acetate was targeted using in silico docking experiments with the modeled structure of the Niemann-Pick C2 target gene (NPC2). Quantitative structure–activity relationship models (QSAR models) was used to the predict the physico-chemical properties or pharmacology activity of the selected drugs and further antihyperlipdemic evaluation of NPC2 gene was studied by analyzing the interaction of hydrogen bonds within the active site of the modeled protein. The confirmation obtained after docking showed good energy binding and docking energy which is about -9.55 Kcal/mol and -11.3Kcal/mol, this shows the inhibitor demosterol showed good interactions towards the modeled protein.

**INTRODUCTION:** Presently with the advancement in molecular biology and the various high throughput equipment used for wet lab and *in silico* research in this field have shown way to rapidly sequence large portions of human genome [1]. The major study in the *in silico* field includes gene finding, genome assembly, sequence alignment, protein structure prediction, protein-protein interaction, modeling of the evolution as well as prediction of gene expression [2].

With the development of evolution, large amount of DNA sequences that encode proteins with unique function are known. With the help of computation biology, it is easy to predict the three dimensional structure of the unknown protein structure from the analysis of other proteins by comparing the similar amino acid sequences [3, 4].

Similarly, with the advancement of human genome sequencing, the use of bioinformatics tools is playing an essential role in analyzing multigene disorder. The diseases from more than two gene mutation results in gene hereditary and it can also occur under the influence of various environmental factors and known as many factor hereditary disease. Disorders that arise from various mutations in a single gene are best valuable for gene therapy. Unfortunately, the common disorders such as high

blood pressure, diabetes, heart disease, arthritis and alzheimer's disease are cause by the role of various mutations in many genes. Multifactorial or multigene disorders especially are difficult to be treated by using gene therapy [5]. For example, different genes that effect breast cancer liability have been recognized on chromosomes 6, 11, 13, 14, 15, 17, and 22.

Niemann-Pick type C (NPC) disease is a uncommon autosomal-recessive lysosomal storage disease naturally accompanied by progressive impairment of liver function and nervous system [6]. Biochemically, the disorder presents with an inhibited glycosphingolipids and degrees of cholesterol from lysosomal compartments and endosomal in non neuronal and neuronal cells.

In majority of NPC patients, mutated NPC1 gene is identified. Around 5% of patients indicate mutations in the NPC2 gene. Numerous different mutations can cause NPC disease and multiple variants not linked with the disease are known in both genes [7, 8]. However, clinical information and genotype data from various NPC patients provide information on the impact of functional variants. NPC-database addresses both professionals and nonprofessionals dealing with NPC disease on a diagnostic, clinical, research, or personal basis [9].

In our computational study, the protein structure of the NPC2 gene was modeled by using homology modelling technique followed by analyzing the physio-chemical property of the selected drugs by using QSAR (*QSAR* models) analysis. The target model was further docked with the recruited drugs by selecting the active site region of the modeled protein to study the interactions.

## MATERIALS AND METHODS:
**Protein Blast:** Protein Blast tool is used for both recognizing a query amino acid sequence as well as for identifying similar amino acid sequences in protein databases. Protein Blast provides a report on global alignment of similar amino acid sequence, which is the required for protein identification purposes[10, 11]. This tool was used to screen for the template structure with maximum sequence identity.

**PFAM-protein families database:** Pfam is a collection of multiple sequence alignments based on the hidden Markov models, covering common protein families and domains. Pfam was used to interpret the domain organization of proteins[12].

**Protein Data Bank:**
The PDB file format obtained was converted into MDL mol file (*.mol) format, because the E-Dragon software can accept this format .The conversion was carried by molecular structure file converter software, which is online software. The input file was taken from the saved PDB file and then pressed the convert button to convert the conversion. The results was copied and saved in mol format.

**Modeller:**
MODELLER was used for homology modeling of the template protein three-dimensional structures. The align2dcommand in MODELLER written in Python script is used to align the unknown protein sequence with the structure of the template in order to generate a model for the unknown protein. [13, 14].

**Procheck:**
This tool was used to check the stereo chemical quality of a protein structure by producing a number of Post Script plots and analyzing its overall geometry and residue-by-residue geometry [15]. The output of the Procheck tool is the Ramachandra plot, which denote the number of residues in allowed and disallowed region.

**ACD-chemsketch:**
The molecules were built using atoms like C, H, O and N present in the ACD-chemsketch. After drawing the molecule, the H bonds were added and optimized the molecule using optimization option in the ACD-chemsketch. This program integrates various advanced structures such as the ability to view colored, rotating molecules in space-filling mode [16].

**QSAR:**
The QSAR analysis was done using the Build QSAR software. The percentage inhibition concentration value was converted to log (1/IC50).These values were used for further calculations. The E-Dragon software was used to

calculate the descriptors for each molecule. Selected descriptors for each molecule were selected by the high correlation r-value. From the Heng-Ling Liou et.al., the seven derivatives which have a Log P value above 7.0 were taken and tabulated [17].

**QSAR Model Generation:**
The selected training set had selected descriptors, from these descriptors, five descriptors were chosen according to their high r-value. The predicted activity value is the IC50 value generated by the Build QSAR module; Residual activity is the difference between the experimental and predicted IC50 values. It was seen that the residual activity value does not deviate beyond +-0.5; hence the chosen descriptors best fit the structural features of the sterol and its derivatives [18]. The r-value for the training set was 0.994, which indicated good correlation between the selected descriptors and the biological activity. The predicted activity, observed activity and residues for the training set were tabulated in **Table 6.**

**Active Site Prediction:**
Active site of NPC2 was found with the help of Q-site finder [19]. They were GLU20,ASN38,ASN58, THR60,ILE87,SER108,LYS116,ALA6,THR7,PHE 8,THR15,ALA16,ALA17,GLN18,ALA19,GLU20, PRO21,VAL22,GLN23,CYS42,PRO43,THR44,and GLN45.

**Auto Dock:**
Auto Dock is a program for calculating the docking modes of small molecules into protein binding sites. In this work Auto Dock and its tools were used for docking process Auto Dock is very highly regarded within the molecular modeling community for its accuracy and reliability. Active atoms of the modeled protein were identified and listed in a word file and save as <file name >.pdb, the modeled protein file and the ligand file were also used in PDB format as input files. After running the Auto Dock, the best conformations having good docking energy and binding energy with minimum values were taken [20].

**Visualization of Auto Dock results:**
The docked results were analyzed by using Web lab viewer. From the Auto Dock results, the good

conformation which was build and pasted into the PDB of the macromolecule protein and saved in as best docked conformations. This file was viewed under the Web lab viewer. Web lab is a visualization tool which helps us to study the hydrogen bond interaction between the drugs i.e. ligand and the drug target i.e. receptor [21].

**RESULTS AND DISCUSSION:**
**BLAST result:**
The NPC2 gene sequence was collected from protein databases (http://www.ncbi.nlm.nih. gov/ entrez/), which has 151 amino acids and 16570 Da. So the target NPC2 gene protein sequence was BLAST for PDB sequence collection. The FASTA format of the NPC2 protein was collected and pasted in the BLAST-P. Then PDB BLAST program was used to search the identical structure in the PDB database. From the results of the BLAST program, the identical sequences were collected as the templates. The target sequence shows 85% identity with the template. **Fig.1**. shows the BLAST result for target sequence NPC2 with 1NEP.

**Homology Modelling:**
The template sequence was identified using BLAST tool against PDB.1HKA and 1NEP proteins which showed good identity compared to the target NPC2 gene, but the resolution of 1HKA protein was less compared to the other. Therefore the active site of the template protein was identified by PFAM tool. Template and target sequences were aligned using Clustal W. **Fig. 2.** denotes the ribbon-Like Structure of 1NEP Protein

**Domain Identification:**
Pfam was used for the domain residues identification for a protein. NPC2 binding domain MD-2-related lipid-recognition (ML) is implicated in lipid recognition, particularly in the recognition of pathogen related products. It has an immunoglobulin-like beta-sandwich fold similar to that of E-set Ig domains. The sequence which was identical towards this gene was 1NEP and 1HKA, but 1NEP was selected for its best homology because of less resolution of about 1.70 [22] whereas 1HKA showed a resolution about 1.81 [23]. **Fig.3** briefs the active site of 1NEP (red dots shows

ligand binding site, yellow box shows catalytic activity site).

## Structure Validation result:
The NPC2 gene which was modeled was evaluated using PROCHECK. The built model contains core region of 86.8%, additionally allowed region of 12.4% and 0.8% of generously allowed, 0% of disallowed region (**Fig. 4**).

## Inhibitors and its derivatives:
The activity data and the two-dimensional structures for seven cholesterol derivatives which were taken from the literature reported by Heng-Ling Liouet al., (2006). The log P value and the structure of each molecule is listed in **Table1**.

## QSAR interpretations:
## Set of Final descriptors:
The final set of descriptors for the series of sterol inhibitors along with their description was shown in **Table 2.**

## Model Evaluate For Test Set:
The seven descriptors were also checked for the observed activity, predicted activity and its residual characters were tabulated in **Table 2** and was done using the Build QSAR model evaluate panel. From results obtained for the test set it was seen that all the molecules in the test showed a positive correlation.

## Graphical View:
The graphical analysis has been performed and the graph is shown in Figure-5 for the training set. The graph had been plotted between the experimental and predicted IC 50 values.

## Docking analysis:
The protein ligand interactions were evaluated based on hydrogen bond distances. Both best-ranked conformations of high and low active compounds were used for measuring hydrogen bond interactions. Cholesterol and lithosterol inhibitors formed only one hydrogen interactions with active site of amino acids. The UNK-O of cholesterol formed hydrogen bond with GLN18. The other two inhibitors like sitosterol and stigmasterol showed no interactions with the amino acids of the protein. Except the demosterol which showed two hydrogen interactions with THR7 and GLN18 whereas cholesterol sulfate and cholesterol acetate showed one hydrogen bond interaction towards GLN45 from the UNK-O. **Fig. 6** denotes the Computational docking analysis. The confirmation obtained after docking showed good energy binding and docking energy which is about -9.55 Kcal/mol and -11.3 Kcal/mol, this shows the inhibitor demosterol showed good interactions towards the modeled protein (**Table 3).**

```
> □ pdb|1NEP|A S  Chain A, Crystal Structure Analysis Of The Bovine Npc2 (Niemann-Pick
C2) Protein
   pdb|2HKA|A S  Chain A, Crystal Structure Of Bovine Npc2 And Cholesterol Sulfate
Complex
   pdb|2HKA|B S  Chain B, Crystal Structure Of Bovine Npc2 And Cholesterol Sulfate
Complex
   pdb|2HKA|C S  Chain C, Crystal Structure Of Bovine Npc2 And Cholesterol Sulfate
Complex
Length=130

 Score =  225 bits (573),  Expect = 5e-60, Method: Compositional matrix adjust.
 Identities = 103/128 (80%), Positives = 119/128 (92%), Gaps = 0/128 (0%)

Query  20   EPVQFKDCGSVDGVIKEVNVSPCPTQPCQLSKGQSYSVNVTFTSNIQSKSSKAVVHGILM   79
            EPV+FKDCGS  GVIKEVNVSPCPTQPC+L +GQSYSVNVTFTSN QS+SSKAVVHGI+M
Sbjct  1    EPVKFKDCGSWVGVIKEVNVSPCPTQPCKLHRGQSYSVNVTFTSNTQSQSSKAVVHGIVM   60

Query  80   GVPVPFPIPEPDGCKSGINCPIQKDKTYSYLNKLPVKSEYPSIKLVVEWQLQDDKNQSLF   139
            G+PVPFPIPE DGCKSGI CPI+KDKTY+Y+NKLPVK+EYPSIK+VVEW+L DDKNQ  F
Sbjct  61   GIPVPFPIPESDGCKSGIRCPIEKDKTYNYVNKLPVKNEYPSIKVVVEWELTDDKNQRFF   120

Query  140  CWEIPVQI   147
            CW+IP+++
Sbjct  121  CWQIPIEV   128
```

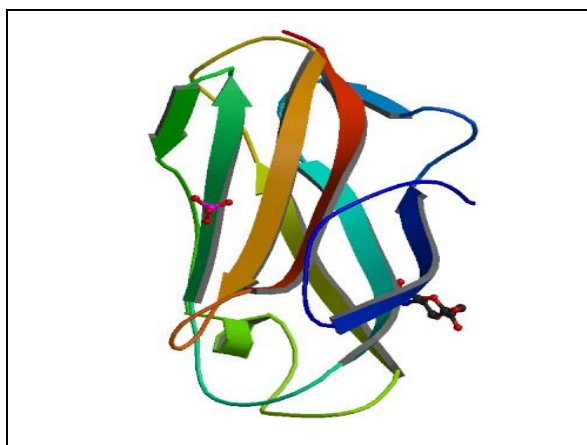**FIG.1: BLAST RESULT FOR TARGET SEQUENCE NPC2 WITH 1NEP**

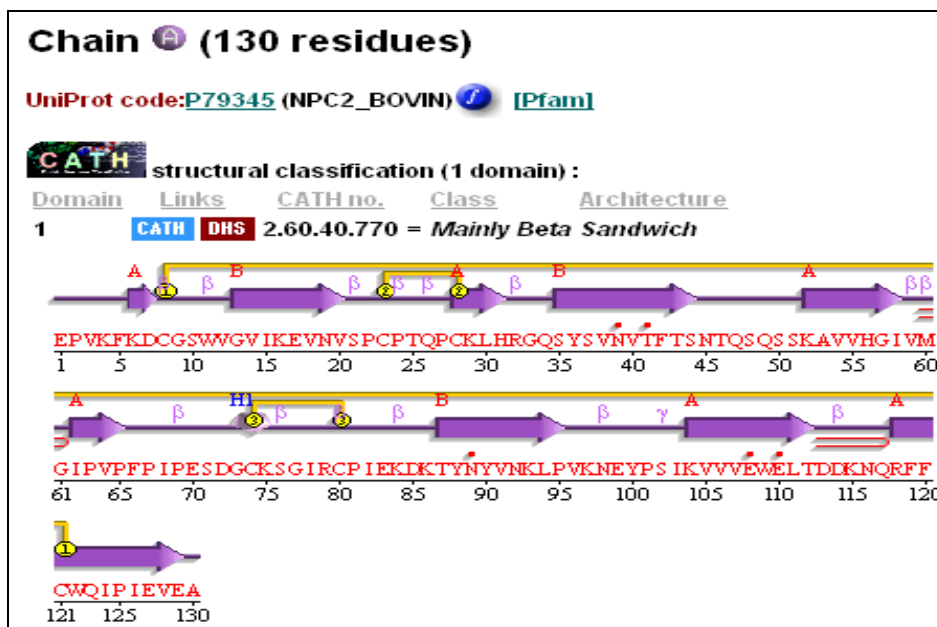**FIG. 2: RIBBON-LIKE STRUCTURE OF 1NEP PROTEIN**



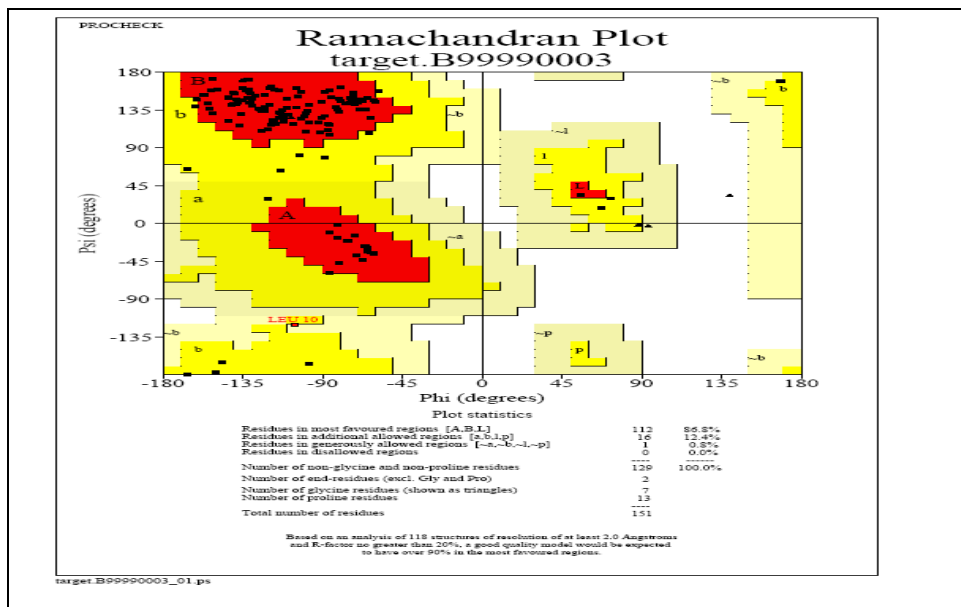**FIG.3: ACTIVE SITE OF 1NEP (RED DOTS SHOWS LIGAND BINDING SITE, YELLOW BOX SHOWS CATALYTIC ACTIVITY SITE)**
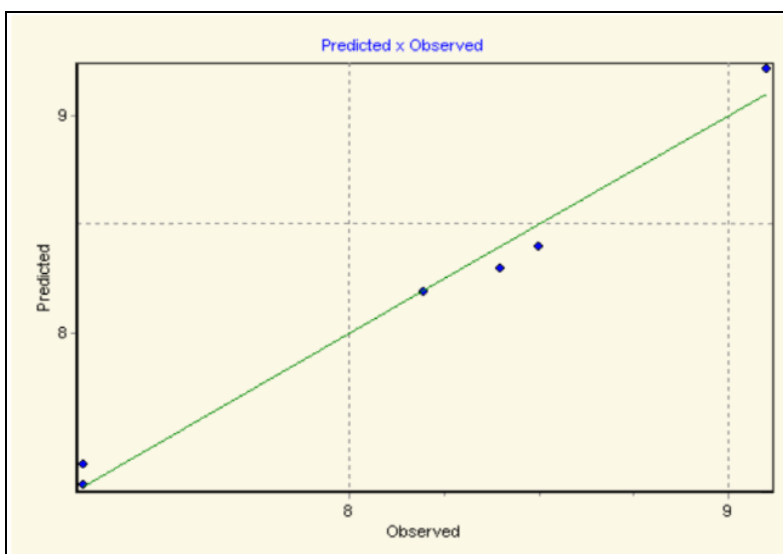


**FIG.4: RAMACHANDRAN PLOT**

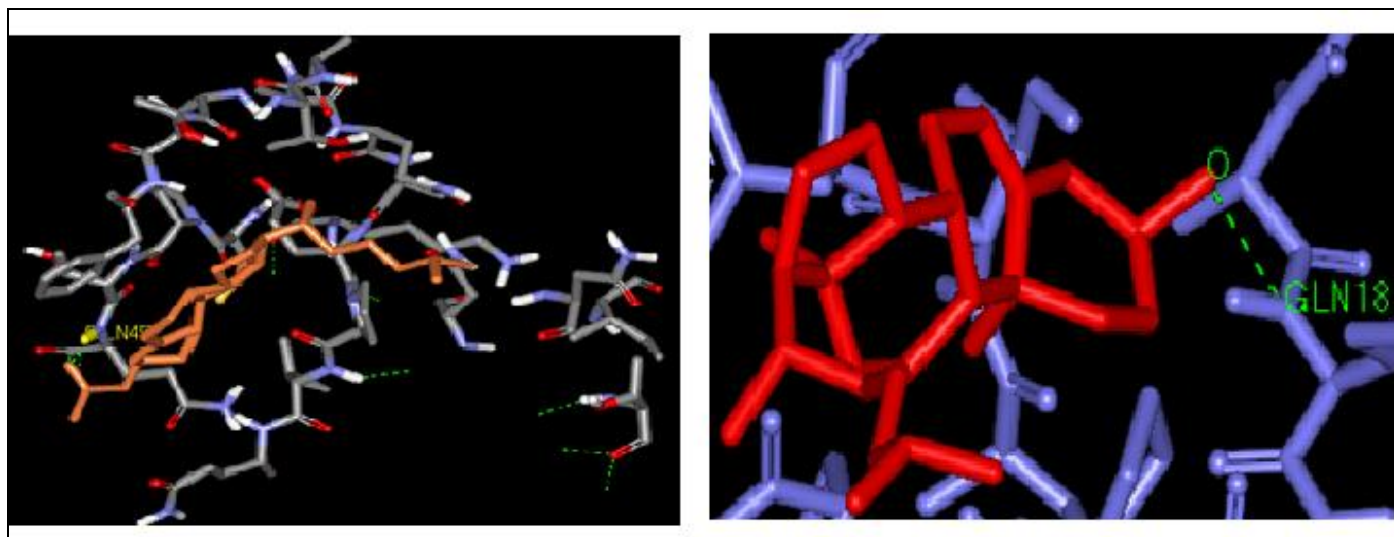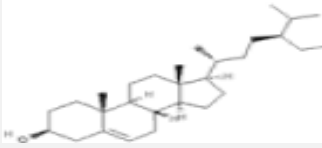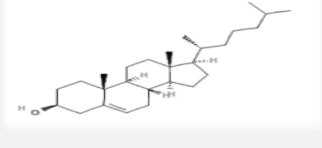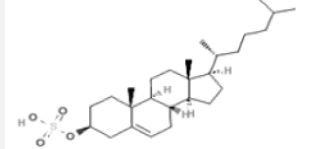**FIG. 5: PLOTS AGAINST PREDICTED Vs OBSERVED IC 50 VALUES OF TRAINING SET**



**FIG.6: COMPUTATIONAL DOCKING ANALYSISA) INTERACTIONS BETWEEN ATOMS b) DEMOSTEROL FORMING TWO HYDROGEN BONDS**

**TABLE 1: STRUCTURE OF STEROL DERIVATIVES AND THEIR LOG P VALUES**

| S.No | Name | Structure | Log P value |
|------|------|-----------|-------------|
| 1 | Beta-sitosterol |  | 9.1 |
| 2 | Cholesterol |  | **8.2** |
| 3 | Cholesterol sulfate |  | **7.3** |

| 4 | Desmosterol | | **7.3** |
| 5 | Lathosterol | | **8.4** |
| 6 | Stigmasterol | | **8.5** |
| 7 | Cholesterol acetate | | **8.9** |
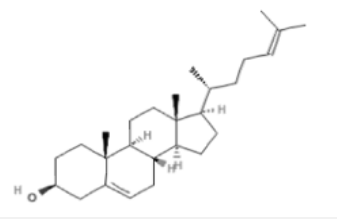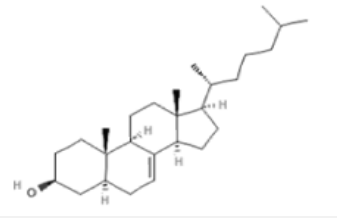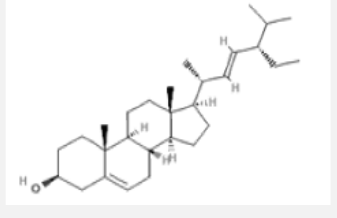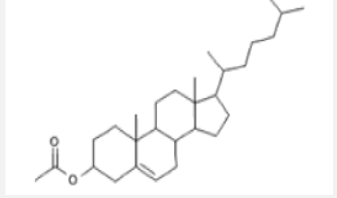
**TABLE 2: TEST SET QSAR RESULT IN DATA**

| Compounds | Observed activity | Predicted activity | Residual |
|---|---|---|---|
| Cholesterol | 8.2 | 8.197 | 0.003 |
| Demo sterol | 7.30 | 7.394 | -0.094 |
| Lathosterol | 8.400 | 8.306 | 0.094 |
| Beta-sitosterol | 9.100 | 9.196 | -0.096 |
| Stigmasterol | 8.5 | 8.400 | 0.100 |
| Cholesterol sulfate | 7.3 | 7.297 | 0.003 |
| Cholesterol acetate | 8.90 | 8.909 | -0.009 |

**TABLE 3: DOCKING ANALYSIS**

| S.no | Name Of the Derivative | Name Of the bond | No: of hydrogen bonds |
|---|---|---|---|
| 1 | Cholesterol | UNK-O | GLN18 |
| 2 | Cholesterol acetate | MOLD-O | GLN45 |
| 3 | Cholesterol sulfate | UNK-O | GLN45 |
| 4 | Demosterol | UNK-O, UNK-O | THR7, GLN18 |
| 5 | Sitosterol | No Interaction | - |
| 6 | Stigmasterol | No Interaction | - |
| 7 | Lithosterol | UNK-O | GLN18 |

**CONCLUSION:** In the current study, we employed computational approaches, such as molecular docking to estimate the binding free energy of 7 inhibitors with NPC2 model protein. Our results demonstrated the prediction of the important inhibitor targeted against NPC2 gene. The docking results will be useful for further research in *in vitro* and *in vivo* studies which can be achieved by employing structure-based drug design and the development of the pharmaceutical agents to treat against Niemann pick disease.

## REFERENCES:

1. www.wikieducator.org.
2. mathcswc.fontbonne.edu.
3. Whittaker PA: The role of bioinformatics in target validation. Drug Discov Today Technol 2004, 1(2):125-133.
4. Hogeweg P: The roots of bioinformatics in theoretical biology. PLoS Comput Biol 2011, 7(3):e1002021.
5. Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM: Gene therapy: therapeutic gene causing lymphoma. Nature 2006, 440(7088):1123.
6. http://www.ninds.nih.gov/disorders/niemann/niemann.htm.
7. Carstea ED, Morris JA, Coleman KG, Loftus SK, Zhang D, Cummings C, Gu J, Rosenfeld MA, Pavan WJ, Krizman DB et al: Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. Science 1997, 277(5323):228-231.
8. Millat G, Bailo N, Molinero S, Rodriguez C, Chikh K, Vanier MT: Niemann-Pick C disease: use of denaturing high performance liquid chromatography for the detection of NPC1 and NPC2 genetic variations and impact on management of patients and families. Mol Genet Metab 2005, 86(1-2):220-232.
9. Runz H, Dolle D, Schlitter AM, Zschocke J: NPC-db, a Niemann-Pick type C disease gene variation database. Hum Mutat 2008, 29(3):345-350.
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215(3):403-410.
11. Madden TL, Tatusov RL, Zhang J: Applications of network BLAST server. Methods Enzymol 1996, 266:131-141.
12. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL et al: The Pfam protein families database. Nucleic Acids Res 2008, 36(Database issue):D281-288.
13. Fiser A, Do RK, Sali A: Modeling of loops in protein structures. Protein Sci 2000, 9(9):1753-1773.
14. Sali A, Blundell TL: Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993, 234(3):779-815.
15. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM: AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996, 8(4):477-486.
16. http://www.acdlabs.com/company/reference.php.
17. Heng-Ling Liou SSD, Sujuan Xu, G. Stephen Tint, Ann M. Stock, and Peter Lobel: NPC2, the Protein Deficient in Niemann-Pick C2 Disease, Consists of Multiple Glycoforms That Bind a Variety of Sterols. THE JOURNAL OF BIOLOGICAL CHEMISTRY 2006, 281(48):36710-36723.
18. Pandian RS, Hemavathi K, Jayapradha R, Seenivasagam R: In-silico analysis and QSAR studies of tacrine hybrids with ubiquitin ligase on Alzheimer's disease. Int J Bioinform Res Appl 2010, 6(6):556-570.
19. Laurie AT, Jackson RM: Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 2005, 21(9):1908-1916.
20. Goodsell DS, Olson AJ: Automated docking of substrates to proteins by simulated annealing. Proteins 1990, 8(3):195-202.
21. http://bioinfo-out.curie.fr/projects/vimida/.
22. Friedland N, Liou HL, Lobel P, Stock AM: Structure of a cholesterol-binding protein deficient in Niemann-Pick type C2 disease. Proc Natl Acad Sci U S A 2003, 100(5):2512-2517.
23. Xiao B, Shi G, Chen X, Yan H, Ji X: Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. Structure 1999, 7(5):489-496.

---

**How to cite this article:**

Mathew S and Qadri I: Quantitative Structure Activity Relationship and Molecular Docking Analysis of Cholesterol Inhibitors against Niemann-Pick C2 Target Gene (NPC2). Int J Pharm Sci Res 2015; 6(9): 3788-95. doi: 10.13040/IJPSR.0975-8232.6(9).3788-95.

This article can be downloaded to **ANDROID OS** based mobile. Scan QR Code using Code/Bar Scanner from your mobile. (Scanners are available on Google Playstore*)*