Prediction of Plasma Membrane Cholesterol from 7-Transmembrane Receptor Using Hybrid Machine Learning Algorithm

Original Scientific Paper

Rudra Kalyan Nayak
School of CSE
VIT Bhopal University, Sehore, Madhya Pradesh, India
rudrakalyannayak@gmail.com

Ramamani Tripathy
Department of Master of Computer Application,
United School of Business Management,
Bhubaneswar, Odisha, India
ramatripathy1978@gmail.com

Hitesh Mohapatra
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, AP, India.
hiteshmahapatra@gmail.com

Amiya Kumar Rath
Department of Computer Science and Engineering,
Veer Surendra Sai University of Technology, Burla,
Sambalpur, Odisha, India
amiyaamiya@rediffmail.com

Debahuti Mishra
Department of Computer Science and Engineering,
Siksha 'O' Anusandhan (Deemed to be) University,
Bhubaneswar, Odisha, India
mishradebahuti@gmail.com

Abstract *– The researches have been made on G-protein coupled receptors (GPCRs) over the long-ago decades. GPCR is also named as 7-transmembrane (7TM) receptor. According to biological prospective GPCRs consist of large protein family with respective subfamilies and are mediated by different physiological phenomena like taste, smell, vision etc. The main functionality of these 7TM receptors is signal transduction among various cells. In human genome, cell membrane plays significant role. All cells are made up of trillion of cells and have dissimilar functionality. Cell membrane composed of different components. GPCRs are reported to be modulated by membrane cholesterol by interacting with cholesterol recognition amino acid consensus (**L/V**-X $_{(1-5)}$-**Y**-X $_{(1-5)}$-**R/K**) (CRAC) or reverse orientation of CRAC (**R/K**-X $_{(1-5)}$-**Y**-X $_{(1-5)}$-L/V) (CARC) motifs present in the TM helices. Among all, cholesterol is one who is regulated by membrane proteins. Here we took GPCR as membrane proteins and this protein modulates membrane cholesterol. According to cell biology, GPCR regulates a wide diversity of vital cellular processes and are targeted by a huge fraction of approved drugs. In this paper we have concentrated our investigation on membrane protein with membrane cholesterol. A hybrid algorithm consisting of spectral clustering and support vector machine is proposed for prediction of membrane cholesterol with GPCR. Spectral clustering uses graph nodes for calculating the cluster points and also it considers other concept such as similarity matrix, low-dimensional space for projecting the data points and upon this parameter at last construct the cluster centre. Supervised learning method is used for solving regression and classification problems. From the analysis we found that our result shows better prediction accuracy in terms of time complexity when compared with two existing models such as fuzzy c-means (FCM) and rough set with FCM model.*

Keywords*: GPCR, TM, Membrane cholesterol, FCM, Rough Set, Spectral clustering, SVM*

1. INTRODUCTION

In mammalian cells, so many important components are included with their diversified functionality. In recent decades, all researches have been going on cell biology. Because huge amount of unsolved issues are still there and varieties of challenges were emerged day by day. In this manuscript our focal point of research is on membrane cholesterol with plasma membrane protein. Plasma membrane is also known as biological membrane or cell membrane which surrounds every living cell to separate the internal stimuli from the outside stimuli and it is made up of bilayer, membrane proteins and carbohydrates in addition with phospholipids shown in figure 1 [1-7]. Basically, plasma membrane is semi-permeable in nature.

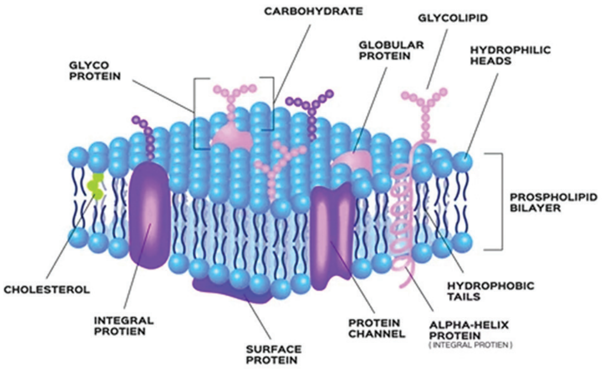


Fig. 1. Schematic representation of cell membrane [5]

By nature cholesterol is amphipathic and it has both hydrophilic and hydrophobic regions. Foremost work of cholesterol in plasma membrane is that it impacts the fluidness and facilitates to produce an effectual dispersal barrier. In the plasma membrane all gaps amongst phospholipids are filled up by cholesterol and also it forbids water soluble molecules from diffusive all around the plasma membrane which is shown in figure 2. A vital purpose of cholesterol is hormone production, Vitamin D production and bile production. Due to much deposit of lipoproteins (LDL) in cell wall, heart disease and other forms of cardiovascular diseases are basically shown in case of human body [8-13].

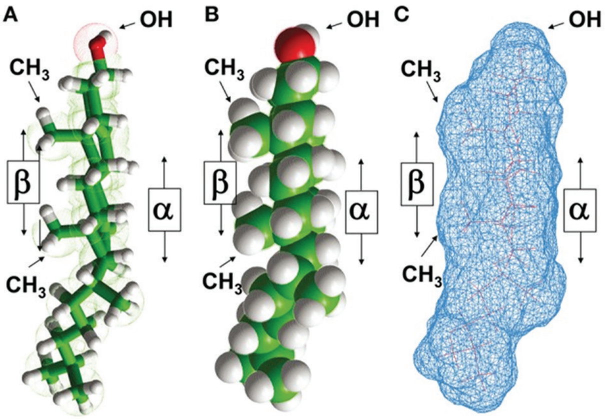


Fig. 2. Cholesterol Structural properties [13]

GPCRs are also called as 7 transmembrane (7-TM) receptors. They are treated as the most important diverse protein families in mammalian genomes.

GPCR is a bigger super family among all cell membrane proteins. It includes above 820 genes with their sub family and symbolized main targets in the development of novel drug candidates in all clinical areas. This family mostly known as larger receptor protein family and are involved in transmitting signals from a diversity of stimuli exterior part to its inside part of cells [13-18]. These families take part in a vital task in physiology by facilitating interaction among cell through recognition of dissimilar ligands, together with nucleosides, bioactive peptides, lipids and amines. Membrane Cholesterol is another imperative component of cellular membrane and has been reported to have a modulatory role in the function of a number of GPCRs. Due to novel functionality of GPCR protein with membrane lipids; it has come out as an exciting domain of research. Cholesterol is a waxy like substances and it is hydrophobic in nature. All cellular cholesterols are distributed unlikely inside the membrane from N-C terminus and to identify cholesterol binding sites of all motifs among seven helices named as helix 1 to helix 7 which is shown if figure 3 below [19-22].

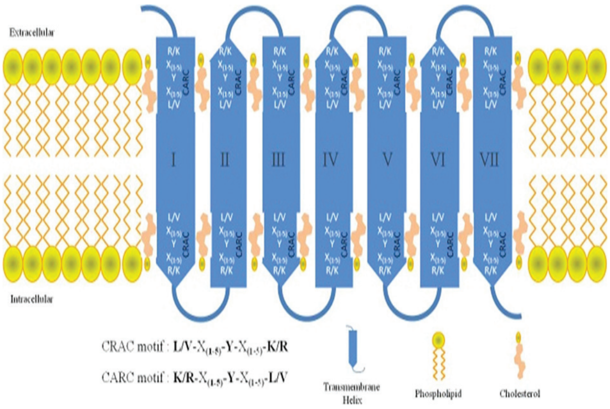


Fig. 3. The 7 helices with N-terminus and C-terminus [1]

Most of the researchers focused their work only on plasma membrane receptor like GPCR. Because GPCR is the largest super family among all the receptors in cell biology and has much functionality such as cell signaling, drug targeting etc. We know that it is an emerging area of research so many researchers have implemented varieties of algorithms upon it like support vector machine, naive Bayes, neural network, fuzzy c-means [21-30] etc. Therefore we have concentrated our experiment on GPCR along with membrane cholesterol which is an innovative idea. Here we proposed a hybrid spectral clustering based FCM model to predict the valid motif sequence(s) from different helices and also compared our proposed model with other two traditional models to showcase the efficacy of proposed model. The said model discovered improved result like more number of targeted promising motif types which could be used for drug design for patients. The rest part of the paper is ordered as follows. In part 2 data set of cholesterol with GPCR proteins and proposed model is discussed. Part 3 explains methodologies with experimental work and finally in part 4 we conclude our work with valid results.

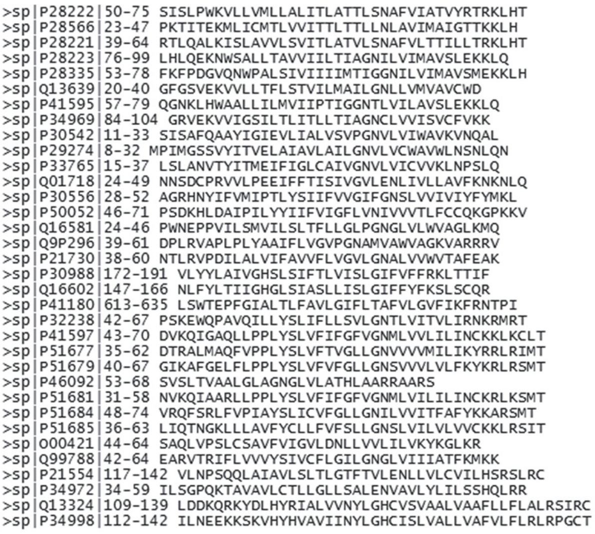
2. DATASET AND PROPOSED MODEL DESCRIPTION

2.1 DATASET DESCRIPTION

From uniprot database [31] we have collected the total helical sequences of each protein. Length of each helix may vary according with their gene ID. Total 820 known proteins with their amino acid sequences reside in database. All helices have individual transmembrane region which is the combination of different amino acids. All database genes contain 7 helices that means from helix 1 to helix 7. Another dataset is membrane cholesterol motif sequence. We prepared a cholesterol dictionary on basis of two algorithms that is CRAC for forward orientation and CARC for backward orientation. Table 1 denotes the possible cholesterol motif using forward and backward sequences with the presence of amino acid. Figure 4 shows one helical file snapshot which was retrieved from database.

Table 1. All probable motif type that is mixture of cholesterol plus X which signifies the arrangement of amino acid that can be different from (one to twenty).

Type of Motif	FORWARD MOTIF FORMULA	BACKWARD MOTIF FORMULA
11 to 15	L/V1Y1R/K,…., L/V1Y5R/K	K/R1Y/F1L/V, ….., K/R1Y/F5L/V
21 to 25	L/V2Y1R/K,……, L/V2Y5R/K	K/R2Y/F1L/V, ……, K/R2Y/F5L/V
31 to 35	L/V3Y1R/K, …, L/V3Y5R/K	K/R3Y/F1L/V, ……., K/R3Y/F5L/V
41 to 45	L/V4Y1R/K, ……, L/V4Y5R/K	K/R4Y/F1L/V, ……, K/R4Y/F5L/V
51 to 55	L/V5Y1R/K, ……, L/V5Y5R/K	K/R5Y/F1L/V, ..., K/R5Y/F5L/V



```
>sp|P28222|50-75 SISLPWKVLLVMLLALITLATTLSNAFVIATVYRTRKLHT
>sp|P28566|23-47 PKTITEKMLICMTLVVITTLTTLLNLAVIMAIGTTKKLH
>sp|P28221|39-64 RTLQALKISLAVVLSVITLATVLSNAFVLTTILLTRKLHT
>sp|P28223|76-99 LHLQEKNWSALLTAVVIILTIAGNILVIMAVSLEKKLQ
>sp|P28335|53-78 FKFPDGVQNWPALSIVIIIIMTIGGNILVIMAVSMEKKLH
>sp|Q13639|20-40 GFGSVEKVVLLTFLSTVILMAILGNLLVMVAVCWD
>sp|P41595|57-79 QGNKLHWAALLILMVIIPTIGGNTLVILAVSLEKKLQ
>sp|P34969|84-104 GRVEKVVIGSILTLITLLTIAGNCLVVISVCFVKK
>sp|P30542|11-33 SISAFQAAYIGIEVLIALVSVPGNVLVIWAVKVNQAL
>sp|P29274|8-32 MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQN
>sp|P33765|15-37 LSLANVTYITMEIFIGLCAIVGNVLVICVVKLNPSLQ
>sp|Q01718|24-49 NNSDCPRVVLPEEIFFTISIVGVLENLIVLLAVFKNKNLQ
>sp|P30556|28-52 AGRHNYIFVMIPTLYSIIFVVGIFGNSLVVIVIYFYMKL
>sp|P50052|46-71 PSDKHLDAIPILYYIIFVIGFLVNIVVVTLFCCQKGPKKV
>sp|Q16581|24-46 PWNEPPVILSMVILSLTFLLGLPGNGLVLWVAGLKMQ
>sp|Q9P296|39-61 DPLRVAPLPLYAAIFLVGVPGNAMVAWVAGKVARRRV
>sp|P21730|38-60 NTLRVPDILALVIFAVVFLVGVLGNALVVWVTAFEAK
>sp|P30988|172-191 VLYYLAIVGHSLSIFTLVISLGIFVFFRKLTTIF
>sp|Q16602|147-166 NLFYLTIIGHGLSIASLLISLGIFFYFKSLSCQR
>sp|P41180|613-635 LSWTEPFGIALTLFAVLGIFLTAFVLGVFIKFRNTPI
>sp|P32238|42-67 PSKEWQPAVQILLYSLIFLLSVLGNTLVITVLIRNKRMRT
>sp|P41597|43-70 DVKQIGAQLLPPLYSLVFIFGFVGNMLVVLILINCKKLKCLT
>sp|P51677|35-62 DTRALMAQFVPPLYSLVFTVGLLGNVVVVMILIKYRRLRIMT
>sp|P51679|40-67 GIKAFGELFLPPLYSLVFVFGLLGNSVVVLVLFKYKRLRSMT
>sp|P46092|53-68 SVSLTVAALGLAGNGLVLATHLAARRAARS
>sp|P51681|31-58 NVKQIAARLLPPLYSLVFIFGFVGNMLVILILINCKRLKSMT
>sp|P51684|48-74 VRQFSRLFVPIAYSLICVFGLLGNILVVITFAFYKKARSMT
>sp|P51685|36-63 LIQTNGKLLLAVFYCLLFVFSLLGNSLVILVLVVCKKLRSIT
>sp|O00421|44-64 SAQLVPSLCSAVFVIGVLDNLLVVLILVKYKGLKR
>sp|Q99788|42-64 EARVTRIFLVVVYSIVCFLGILGNGLVIIIATFKMKK
>sp|P21554|117-142 VLNPSQQLAIAVLSLTLGTFTVLENLLVLCVILHSRSLRC
>sp|P34972|34-59 ILSGPQKTAVAVLCTLLGLLSALENVAVLYLILSSHQLRR
>sp|Q13324|109-139 LDDKQRKYDLHYRIALVVNYLGHCVSVAALVAAFLLFLALRSIRC
>sp|P34998|112-142 ILNEEKKSKVHYHVAVIINYLGHCISLVALLVAFVLFLRLRPGCT
```

Fig. 4. Snapshot of one helical protein

2.2 PROPOSED MODEL

In eukaryotic membrane, cellular cholesterol plays a vital role and is modulated by GPCR which is renowned as cell signalling among intracellular with extracellular leaflets. So many receptors and transporters are available in mammalian cell. But our research focuses on superfamily GPCR receptor with membrane cholesterol. GPCR regulate a wide diversity of vital cellular processes and are targeted by a huge fraction of approved drugs. The workflow is explained in Figure 5 and step wise elaboration is mentioned below.

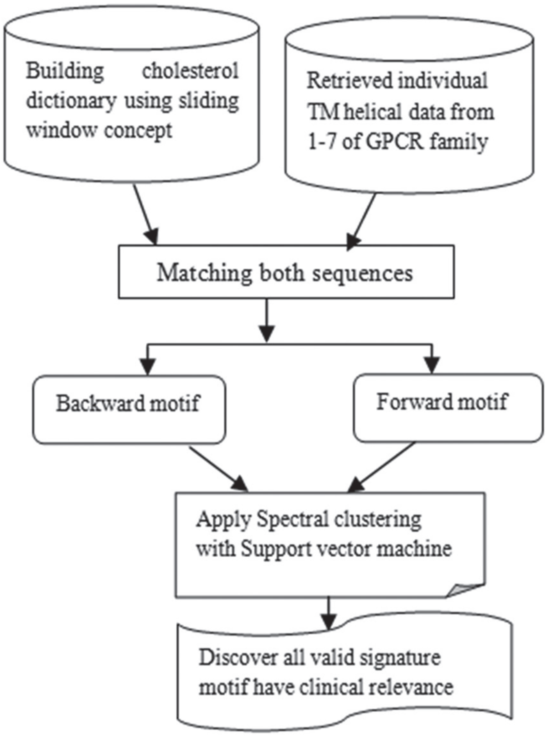

Building cholesterol dictionary using sliding window concept
Retrieved individual TM helical data from 1-7 of GPCR family
Matching both sequences
Backward motif
Forward motif
Apply Spectral clustering with Support vector machine
Discover all valid signature motif have clinical relevance

Fig. 5. Proposed model of cholesterol with GPCR family

Step-1: Firstly, we have collected GPCR proteins with their respective helices from uniprot database. And also have constructed cholesterol dictionary with the help of CRAC and CARC algorithm for both forward and backward direction.

Step 2: Then we took matching position of both dataset to find the backward and forward motif using Rabin-Karp string matching algorithm [1].

Step 3: In the next step we have applied hybrid spectral clustering with support vector machine method for both CRAC and CARC algorithm.

Step 4: Finally, we found our targeted valid motif which have clinical relevance.

3. METHODOLOGY WITH EXPERIMENTAL DISCUSSION

3.1 SPECTRAL CLUSTERING

In various fields like bioinformatics, image processing, networking, data mining etc. clustering approaches have been widely used for solving the numerous problems. In every aspect clusters are formed according with their similarity of data objects. As we know that clustering is an unsupervised machine learning-based algorithm that comprises a group of data points into clusters so that the objects belong to the same group. In our paper we used spectral clustering algorithm on forward and backward motif of membrane cholesterol to distinguish the motif sequences with the help of cluster. Spectral clustering uses graph nodes for calculating the cluster points and also it considers other concepts such as similarity matrix, low-dimensional space for project the data points and upon this parameter, at last constructs the cluster centre [32-34].

```
Algorithm: Spectral Clustering
Input:
        Data set Y={y_1,…,y_n}, Initailize $\sigma$ scaling
        start on:
Step 1: All data are preprocessing with the help of
        scaling method
Step 2: Make an Affinity matrix $A_{jk} = \exp\,(-\frac{||d_j - d_k||^2}{2\sigma^2})$
Step 3: Put up a Laplacian matrix $L^{norm} := I - P^{-\frac{1}{2}}AP^{-\frac{1}{2}}$
Step 4: Work out the k largest Eigen vectors $x_i,..,x_k$ of L
Step 5: matrix is X=$[u_i,..,u_k] \in R^{(m \times k)}$
Step 6: Outline a matrix $W$ from $X$ as $W_{ij} = \frac{x_{ij}}{(\sum_j x_{ij}^2)^{1/2}}$
Step 7: Cluster every one $W$ by $k$-means
Step 8: Allocate the $X_i$ to cluster $j$iff$W_i$ is assign to cluster $j$
```

3.2 SUPPORT VECTOR MACHINE (SVM)

In recent days, classification approach has been treated as one of the powerful tool for dissimilar applications like protein structure prediction, text categorization, face recognition, fingerprint recognition, speech recognition, data classification, micro-array gene expression, etc. In this paper we have applied a novel approach spectral with SVM algorithm on our dataset. Basically, this supervised learning method is used for solving regression and classification problems. SVM theory is always characterizing the decision boundaries using the decision planes concepts [35-37].

A training set that includes label pairs (w_i,v_i), i=1,...,n every w here $w_i \in R^n$ and $v \in \{profit, loss\}^i$, SVM algorithm wants outcome with the help of optimization problem that is stated below.

$$\min_{x,y,\xi} \frac{1}{2} U^M u + C \sum_{i=1}^{n} \xi_i \tag{1}$$

Subject to: $v_i(z^M \Phi(w) + y \geq 1 - \xi_i, \xi_i \geq 0.$ (2)

In equation (3) decision function is denoted as

$$P = pfn(\sum_{i=1}^{n} x_i \alpha_i B(x_i, x) + \rho) \tag{3}$$

With help of the kernel function ϕ, training vector w_i is mapped with their dimensional space. According to SVM concept all data points are classified using hyper planes even if it is impracticable for receiving linear solution in case of two dimensional spaces. For this reason we are using kernel function $k(u_i,u_j) \equiv \Phi(u_i)^N \Phi(u_j)$ for multidimensional data. Utilizing divergent kernels function, this algorithm is trained which is shown in below equations (4), (5) and (6).

(i) *Linear kernel:*

$$\left(u_i, u_j\right) = u_i^N u_j \tag{4}$$

(ii) Polynomial kernel:

$$f\left(u_i, u_j\right) = (\gamma u_i^N u_j + r)^e, \gamma > 0, \tag{5}$$

and

(iii) Radial Basis kernel

(RBF): $(u_i, u_j) = exp\left(-\gamma \parallel u_i - u_j \parallel^2\right), \gamma$ (6)

The entire kernel arguments such as C,γ,r, and e are initialized by utilizing the dataset. All kernel parameters areaffected based upon the size of training data [35-40].

3.3 EXPERIMENTAL PART ELABORATION

The work flow of our manuscript is finished using windows 10 operating system plus Intel i5 processor with hard disk of 8 GB for finishing the experimental part and here Python 3.7 is used for coding purpose. To compute the overall performance here we took helical data of GPCR receptor with dictionary of cholesterol. The work flow of our model is elaborated step wise manner.

Step 1: In very beginning step, first of all extracted the individual helix protein data of GPCR receptor from uniprot database like, protein Id, helix name (h1-h7) and length of the protein whichever is the mixture of different amino acid. Next dataset cholesterol is also computed based on CRAC/CARC approach.

Step 2: After collecting both dataset sequences we used sliding window concept on both to find out the separate motif sequences of backward and forward region. Window size is as *W* = {*w5, w6, w7, w8, w9, w10, w11, w12, w13*}. The formula for cholesterol dictionary is CRAC and CARC.

Step 3: After completion of step 2 work, we move to next step where we applied our proposed algorithm spectral clustering and SVM for prediction of membrane cholesterol with membrane receptor GPCR. Our proposed algorithm is well suited for both the datasets. The foremost objective of this paper is to find out valid signature motif from prediction.

Here in below, Table 4 shows resultant prediction of cholesterol from GPCR receptor for forward motif. In this table we have taken two columns which have represented as Id number proteins with their motif type, motif sequences with helix name. The entire predicted motifs are found using CRAC (L/V X_{1-5} Y X_{1-5} K/R) formula. Here we can explain one result. With protein id P30550 which is included in motif type 55 (*L/V XXXXX Y XXXXXK/R*) which means first position amino acid *L/V* is present and last position amino acid *K/R* is present. Expect 1st and last position another position is their where another amino acid Y is present. Therefore, our motif sequence is like LSIISVYYYFIAK. In this case first position is followed by L, after then 5 different amino acid are present such as SIISV and then middle position is Y and after that another dissimilar amino acid are there like YYFIA. Finally, in last position K is present. And all this sequence is present in helix 5 region of GPCR proteins. In this way all forward motif sequences are predicted using the proposed algorithm

Table 4. Resultant prediction of cholesterol from GPCR receptor for forward motif

Identification number of protein with Motif Type	Sequence of motif and helix name	Identification number of protein with Motif Type	Sequence of motif and helix name
P30550 (55)	LSIISVYYYFIAK (helix 5)	P11229(44)	VMCTLYWRIYR (helix 5)
Q13585 (55)	LIVGFCYVRIWTK (helix 5)	P08912(44)	VMTILYCRIYR (helix 5)
P28336 (55)	LAIISIYYYHIAK (helix 5)	Q14833(44)	VTCTVYAIKTR (helix 5)
P49683(55)	LVILLSYVRVSVK (helix 5)	P16473(42)	VIVCCCYVK (helix 5)
P32745(55)	LVICLCYLLIVVK (helix 5)	Q9UBY5(52)	VVNPIIYSYK (helix 7)
P32248(54)	LAMSFCYLVIIR (helix 5)	Q8NH63(52)	VLNPIVYSVK (helix 7)
P25025(54)	LIMLFCYGFTLR (helix 5)	Q14833(52)	VSLGMLYMPK (helix 7)
O43193(54)	LCLSILYGLIGR (helix 5)	O15303(52)	VSLGMLYVPK (helix 7)
P41146(54)	LVISVCYSLMIR (helix 5)	Q14831(52)	VALGMLYMPK (helix 7)
Q96G91(54)	LLTLAAYGALGR (helix 5)	O00222(52)	VSLGMLYMPK (helix 7)
P51582 (54)	LVTLVCYGLMAR (helix 5)	P29275(52)	VVNPIVYAYR (helix 7)
Q9UKP6 (54)	LLIGLLYARLAR (helix 5)	P41968(52)	VIDPLIYAFR (helix 7)
P08908(54)	LLMLVLYGRIFR (helix 5)	P33032(52)	VMDPLIYAFR (helix 7)
Q9NPB9 (54)	LIMGVCYFITAR (helix 5)	Q96R84(52)	VMNPLIYSLR (helix 7)
O43603(44)	VLGLTYARTLR (helix 5)	P46092(52)	LNPVLYAFLGLR (helix 7)
P32239(44)	VMAVAYGLISR (helix 5)	P41231(52)	LDPVLYFLAGQR (helix 7)
P41146(44)	VISVCYSLMIR (helix 5)	P30411(45)	LNPLVYVIVGKR (helix 7)

Table 5. Resultant prediction of cholesterol from GPCR receptor for backward motif

Identification number of protein with Motif Type	Sequence of motif and helix name	Identification number of protein with Motif Type	Sequence of motif and helix name
P49238(55)	KSVTDIYLLNLAL (helix 2)	P03999 (42)	RQPLNYILV (helix 2)
P41143(55)	KTATNIYIFNLAL (helix 2)	Q86VZ1 (42)	RHHWVFGVL (helix 3)
P41145(55)	KTATNIYIFNLAL (helix 2)	Q9BZJ6 (42)	RVSAMFFWL (helix 3)
P41146(55)	KTATNIYIFNLAL (helix 2)	P21453 (42)	REGSMFVAL (helix 3)
P55085(55)	KHPAVIYMANLAL (helix 2)	P21452 (42)	RAFCYFQNL (helix 3)
Q99500(55)	KFHNRMYFFIGNL (helix 2)	Q9NYW4 (42)	RYLSIFWVL (helix 3)
O00421(55)	KRVENIYLLNLAV (helix 2)	P32248 (35)	KMSFFSGMLLL (helix 3)
Q9Y271(55)	KSAFQVYMINLAV (helix 2)	P25024 (35)	KEVNFYSGILL (helix 3)
Q969V1(55)	KTVPDIYICNLAV (helix 2)	P47900 (35)	KLQRFIFHVNL (helix 3)
Q9GZQ4(55)	KTPTNYYLFSLAV (helix 2)	P41231 (35)	KLVRFLFYTNL (helix 3)
P31391(55)	KTATNIYLLNLAV (helix 2)	P51582 (35)	KFVRFLFYWNL (helix 3)
P50052(55)	KKVSSIYIFNLAV (helix 2)	Q9NYW0 (35)	KIANFSNYIFL (helix 3)
O43193(55)	RTTTNLYLGSMAV (helix 2)	P46094 (52)	RTVKLIFAIV (helix 6)
Q9HB89(55)	RTPTNYYLFSLAV (helix 2)	P25024 (52)	RAMRVIFAVV (helix 6)
P23945(54)	KLTVPRFLMCNL (helix 2)	P25025 (52)	RAMRVIFAVV (helix 6)
P22888(54)	KLTVPRFLMCNL (helix 2)	P30559 (52)	RTVKMTFIIV (helix 6)
P16473(54)	KLNVPRFLMCNL (helix 2)	P37288 (52)	RTVKMTFVIV (helix 6)
Q6W5P4(54)	KKSRMTFFVTQL (helix 2)	P47901 (52)	RTVKMTFVIV (helix 6)
P30559(54)	KHSRLFFFMKHL (helix 2)	P49019 (32)	RIHIFWLL (helix 6)
Q9NYW0(54)	KLSTIGFILTGL (helix 2)	P49683 (32)	RRRTFCLL (helix 6)
P21453(54)	RPMYYFIGNLAL (helix 2)	O14514 (32)	RSALFQIL (helix 6)
P43220(54)	RALSVFIKDAAL (helix 2)	Q9UP38 (32)	RIGVFSVL (helix 6)
P34969 (42)	RQPSNYLIV (helix 2)	Q9ULW2v (32)	RIGLFSVL (helix 6)
P29371(42)	RTVTNYFLV (helix 2)	Q14332 (32)	RIGVFSVL (helix 6)
O43613(42)	RTVTNYFIV (helix 2)	Q9NPG1 (32)	RIGVFSIL (helix 6)
O43614(42)	RTVTNYFIV (helix 2)	Q13467 (32)	RIGIFTLL (helix 6)
O95977(42)	RRWVYYCLV (helix 2)	O60353 (32)	RIGVFSGL (helix 6)
P21452(42)	RTVTNYFIV(helix 2)	O75084 (32)	RIGVFSVL (helix 6)
P25103(42)	RTVTNYFLV (helix 2)	Q9H461 (32)	RLGLFTVL (helix 6)
P35368(42)	RTPTNYFIV (helix 2)		

Table 5 shows resultant prediction of cholesterol from GPCR receptor for backward motif. All the motif sequences for forward motif was mentioned in table 4

Like table 4, all the motif sequences of table 5 have been predicted for backward direction. Here the CARC (K/RX1-5 YX1-5 L/V) formula has been implemented for prediction. Here we took one predicted motif sequence

RTVTNYFIV from table 5 for description. This sequence has protein id P21452 and it is included under helix2. First position of sequence motif is either K or R and middle part is constant I that is Y and the last position it contains either amino acid L or V. From this analysis we found backward motifs target the membrane proteins more in comparison to forward motifs. Most of the target sites are under higher motif 55, 52, 45, 42, 35 etc. and helix are 5, 2, 7, 3, 6.

3.4 COMPARATIVE ANALYSIS

In Table 6, we have compared FCM model, rough set with FCM model and our proposed model to showcase the efficacy of the proposed model in terms of discovering suitable types of motifs. Target site of helix by FCM model was found to be h2, h5 and h7 having motif type 11, 12, 21, 54, 34. Further, target site of helix by rough set with FCM model was found to be h2, h3, h5 and h7 having motif type 21, 51, 44, 54, 25, 53. Our proposed model discovers the target site of helix as h2, h3, h5 and h7 and extra helix in h6. Also their motif types are 55, 52, 45, 42, 35 which is higher than the existing models.

Table 6. Motif type comparison by different methods

Methods	Helix Name	Motif Type (Forward/ Backward)
FCM [1]	h2,h5,h7	11,12,21,54,34
Rough Set with FCM [28]	H2,h3,h5,h7	21,51,44,54,25
Spectral with SVM (Proposed)	H2,h3, h5,h7,h6	55,52,45,42,35

4. CONCLUSION

In this paper, we concentrated our work on prediction of uncovering membrane cholesterol from human GPCR super family. Frequently, such receptors are the most significant protein super family in biological membrane and play a substantial role in the transduction of signal across cell membranes. It also signifies as an essential drug target in all clinical fields. About 820 human proteins are included in this family. Membrane cholesterol has a modulatory role in the function of some GPCRs.

In our manuscript we have discussed about cellular receptor with membrane cholesterol. According to biological perspective GPCRs consist of large protein family in mammalian cells. The main functionality of these 7TM receptors is signal transduction among unlike cells. In human genome, cell membrane plays significant role. Cell membrane composed of different components. GPCRs are reported to be modulated by membrane cholesterol by interacting with these CRAC or CARC motifs present in the TM helices. Among all, cholesterol is one who is regulated by membrane proteins. From experiment, we found both forward and backward motif from different helices. Among all, we observed targeted sites are under higher motif 55, 52, 45, 42, 35 etc. and helix are 5, 2, 7, 3, 6.

Here, our experimental analyses conclude that prediction of membrane cholesterol with GPCR receptor using spectral and SVM performs well. Backward motif sequences target the protein sites greater than forward motif that means CARC algorithm has higher valid signature motifs which has clinical relevance.

5. REFERENCES

[1] R. Tripathy, M. Debahuti, V. B. Konkimalla, "A novel fuzzy C-meansapproach for uncovering cholesterol consensus motif from human G-protein coupled receptors (GPCR)", Karbala International Journal of Modern Science, Vol. 1, No. 4, 2015, pp. 212-224.

[2] R. D. DiMarchi, P. Mitra, "Gas regulates Glucagon-Like Peptide 1 Receptor-mediated cyclic AMP generation at Rab5 endosomal compartment", Molecular Metabolism, Vol. 6, No. 10, 2017.

[3] R. Tripathy, M. Debahuti, V. B. Konkimalla, "A computational approach for mining cholesterol and their potential target against GPCR seven helices based on spectral clustering and fuzzy c-means algorithms", Journal of Intelligent & Fuzzy Systems, Vol. 35, No. 1, 2018, pp. 305-314.

[4] O. Mouritsen, M. Zuckermann. "What's so special about cholesterol?", Lipids, Vol. 39, No.11, 2004, pp. 1101-1113.

[5] G. Karp, "The structure and function of the plasma membrane", Cell and Molecular Biology: Concepts and Experiments, 3rd edition, John Wiley and Sons, 2002, pp. 122-82.

[6] Y. Lange, B. V. Ramos, "Analysis of the distribution of cholesterol in the intact cell", Journal of Biological Chemistry, Vol. 258, No. 24, 1983, pp. 15130-15134.

[7] K. G. Burger, G. Gimpl, F. Fahrenholz, "Regulation of receptor function by cholesterol", Cellular and Molecular Life Sciences CMLS, Vol. 57, No. 11, 2000, pp. 1577-1592.

[8] D. Lingwood, K. Simons, "Lipid rafts as a membrane-organizing principle", Science, Vol. 327, No. 5961, 2010, pp. 46-50.

[9] Md, Jafurulla, S. Tiwari, A. Chattopadhya,. "Identification of cholesterol recognition amino acid consensus (CRAC) motif in G-protein coupled recep-

tors", Biochemical and biophysical research communications, Vol. 404, No. 1, 2011, pp. 569-573.

[10] H. Nakashima, Y. Kuroda, "Differences in dinucleotide frequencies of thermophilic genes encoding water soluble and membrane proteins", Journal of Zhejiang University SCIENCE B, Vol. 12, No.6, 2011, p. 419.

[11] T. Schöneberg et al. "Learning from the past: evolution of GPCR functions", Trends in pharmacological sciences, Vol. 28, No. 3, 2007, pp. 117-121.

[12] R. Fredriksson, H. B. Schiöth, "The repertoire of G-protein–coupled receptors in fully sequenced genomes", Molecular pharmacology, Vol. 67, No. 5, 2005, pp. 1414-1425.

[13] C. J. Baier, J. Fantini, F. J. Barrantes. "Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor", Scientific reports, Vol. 1, 2011, p. 69.

[14] C. Ellis, "The state of GPCR research in 2004", Nature Reviews Drug Discovery, Vol. 3, No. 7, 2004, pp. 577-626.

[15] H. Li, V. Papadopoulos, "Peripheral-type benzodiazepine receptor function in cholesterol transport. Identification of a putative cholesterol recognition/interaction amino acid sequence and consensus pattern", Endocrinology, Vol. 139, No. 12, 1998, pp. 4991-4997.

[16] S. Schlyer, R. Horuk, "I want a new drug: G-protein-coupled receptors in drug development", Drug discovery today, Vol. 11, No. 11-12, 2006, pp. 481-493.

[17] T. J. Pucadyil, A. Chattopadhyay, "Role of cholesterol in the function and organization of G-protein coupled receptors", Progress in lipid research, Vol. 45, No. 4, 2006, pp. 295-333.

[18] X. Sun, G. R. Whittaker. "Role for influenza virus envelope cholesterol in virus entry and infection", Journal of virology, Vol. 77, No. 23, 2003, pp. 12543-12551.

[19] R. M. Epand et al. "Cholesterol interaction with proteins that partition into membrane domains: an overview", Cholesterol Binding and Cholesterol Transport Proteins, Springer, Dordrecht, 2010, pp. 253-278.

[20] A. K. Hamouda et al. "Cholesterol interacts with transmembrane α-helices M1, M3, and M4 of the Torpedo nicotinic acetylcholine receptor: photolabeling studies using [3H] azicholesterol", Biochemistry, Vol. 45, No. 3, 2006, pp. 976-986.

[21] C. Banchhor, N. Srinivasu, "FCNB: Fuzzy correlative naïve bayes classifier with MapReduce framework for big data classification", Journal of Intelligent Systems, Vol. 29, No. 1, 2018, pp. 994-1006.

[22] R. Tripathy, R. K. Nayak, P. Das, D. Mishra, "Cellular cholesterol prediction of mammalian ATP-binding cassette (ABC) proteins based on Fuzzy C-Means with support vector machine algorithms", Journal of Intelligent & Fuzzy Systems, 2020. (Preprint).

[23] C. Subbalakshmi, G. Ramakrishna, S. Rao, "Evaluation of data mining strategies using fuzzy clustering in dynamic environment", Proceedings of the 3rd International Conference on Advanced Computing, Networking and Informatics, New Delhi, 2016, pp. 529-536.

[24] S. C. Gopi, K. Kiran, M. Veerabrahmam, Y. Ayyappa, "Fuzzy Based Classification of X-Ray Images with Convolution Neural Network", International Journal of Emerging Trends in Engineering Research, 2020, pp.4433-4436.

[25] S. P. Potharaju, M. Sreedevi, "A novel subset feature selection framework for increasing the classification performance of SONAR targets", Paper presented at the Procedia Computer Science, Vol. 125, 2018, pp. 902-909.

[26] S. Razia, M. R. Narasingarao, "A neuro computing framework for thyroid disease diagnosis using machine learning techniques", Journal of Theoretical and Applied Information Technology, Vol. 95, No. 9, 2018, pp. 1996-2005.

[27] A. Ayushree; S. Kumar, "Comparative Analysis Of Aodv And Dsdv Using Machine Learning Approach In Manet", Journal Of Engineering Science And Technology, Vol. 12, No. 12, 2017, pp. 3315-3328.

[28] R. K. Nayak, R. Tripathy, V. Saravanan, P. Das, D. K. Anguraj, "A Novel Strategy for Prediction of Cellular Cholesterol Signature Motif from G Protein-Coupled Receptors based on Rough Set and FCM Algorithm", Proceedings of the Fourth Interna-

tional Conference on Computing Methodologies and Communication, Erode, India, 11-13 March 2020, pp. 285-289.

[29] R. K. Nayak, R. Tripathy, D. Mishra, D., V. K. Burugari, P. Selvaraj, A. Sethy, A., B. Jena, “Indian Stock Market Prediction Based on Rough Set and Support Vector Machine Approach”, Intelligent and Cloud Computing, 2020, pp. 345-355.

[30] R. Tripathy, R. K. Nayak, V. Saravanan, D. Mishra, G. Parasa, K. Das, P. Das, “Spectral Clustering Based Fuzzy C-Means Algorithm for Prediction of Membrane Cholesterol from ATP-Binding Cassette Transporters” Intelligent and Cloud Computing, 2019, pp. 439-448.

[31] UniProt Consortium, "The universal protein resource (UniProt)”, Nucleic acids research, Vol. 36, 2007, pp. D190-D195.

[32] E. Arias-Castro, Ery, C. Guangliang G. Lerman, "Spectral clustering based on local linear approximations”, Electronic Journal of Statistics, Vol. 5, 2011, pp. 1537-1587.

[33] H. Zare et al. "Data reduction for spectral clustering to analyze high throughput flow cytometrydata”, BMC bioinformatics, Vol. 11, No. 1, 2010, p. 403.

[34] R. K. Nayak, D. Mishra, K. Shaw, S. Mishra, “Rough set based attribute clustering for sample classification of gene expression data”, Procedia engineering, Vol. 38, 2012, pp. 1788-1792.

[35] R. K. Nayak, D. Mishra, A. K. Rath, "A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices”, Applied Soft Computing, Vol. 35, 2015, pp. 670-680.

[36] C. Cortes, "WSupport-vector network”, Machine learning, Vol. 20, 1995, pp. 1-25.

[37] X. Zhang, X. Zheng, "Comparison of text sentiment analysis based on machine learning”, Proceedings of the 15th International Symposium on Parallel and Distributed Computing, Fuzhou, China, 8-10 July 2016.

[38] V. P. Upadhyay et al. "Forecasting stock market movements using various kernel functions in support vector machine”, Proceedings of the International Conference on Advances in Information Communication Technology & Computing, Bikaner, India, 12-13 August 2016.

[39] R. K. Nayak, D. Mishra, A. K. Rath, "An optimized SVM-k-NN currency exchange forecasting model for Indian currency market”, Neural Computing and Applications, Vol. 31, No. 7, 2019, pp. 2995-3021.

[40] N. Chandra, "Support Vector Machine Classifier for Predicting Drug Binding to P-glycoprotein”, Journal of Proteomics & Bioinformatics, Vol. 2, No. 4, 2009.