EasyChair Preprint

№ 3726

Membrane Cholesterol Prediction from Cellular Receptor based on Spectral Clustering and Support Vector Machine

Rudra Kalyan Nayak, Ramamani Tripathy, Debahuti Mishra and Amiya Kumar Rath

July 3, 2020

Membrane Cholesterol Prediction from Cellular Receptor based on Spectral Clustering and Support Vector Machine

Rudra Kalyan Nayak[1], RamamaniTripathy[2], Debahuti Mishra[3], Amiya Kumar Rath[4]

[1]Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur, Andhra Pradesh, India.
rudrakalyannayak@gmail.com
[2]Department of Master of Computer Application,
United School of Business Management, Bhubaneswar, Odisha, India
ramatripathy1978@gmail.com
[3]Department of Computer Science and Engineering,
Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar, Odisha, India
mishradebahuti@gmail.com
[4]Department of Computer Science and Engineering,
Veer Surendra Sai University of Technology, Burla, Sambalpur, Odisha, India
amiyaamiya@rediffmail.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 (CRAC) or reverse orientation of CRAC (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. SVM concept is based on the decision planes so as to characterize the decision boundaries. From the analysis we found that our result shows better prediction accuracy with respect to time complexity.

Keywords: GPCR, TM, Membrane cholesterol, 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. G-protein couple receptor (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 [1-10]. 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 [1-20].

So our objectives have to predict cell membrane cholesterol with GPCR membrane proteins using algorithms. We know that as 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 etc. Here we have proposed a hybridized approach consisting of spectral clustering with support vector machine. Our rest 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 [21] 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 (**L/V**-X $_{(1-5)}$-**Y**-X $_{(1-5)}$-**R/K**) and CARC (**R/K**-X $_{(1-5)}$-**Y**-X$_{(1-5)}$-**L/V**). Table 1 denotes the possible cholesterol motif using forward and backward sequences with the presence of amino acid. Figure 1 shows one helical file snapshot which was retrieved from database.

Table1: 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 (L/V-*X*$_{(1-5)}$-Y-*X*$_{(1-5)}$-R/K)**	**BACKWARD MOTIF FORMULA (R/K-*X*$_{(1-5)}$-Y/F-*X*$_{(1-5)}$-L/V)**
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
```

Figure-1: 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 2.

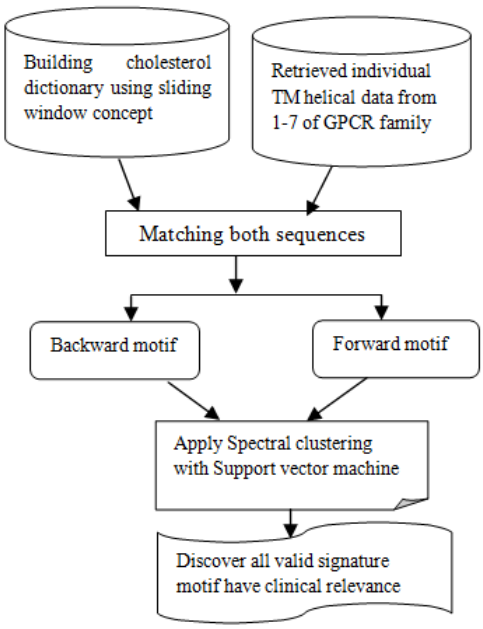

Figure 2: Proposed model of cholesterol with GPCR family

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 [22-24].

Algorithm: Spectral Clustering

Input:

Data set $Y = \{y_1, \dots, y_n\}$, Initailize σ $scaling$

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 $L^{norm} := I - P^{-\frac{1}{2}} A P^{-\frac{1}{2}}$

Step 4: Compute 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 $j$iffW_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 concept is based on the decision planes so as to characterize the decision boundaries [25-27].

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

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

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

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

Training vector w_i is mapped in to higher dimensional space with help of the kernel function ϕ . The separating surface depends on a subset of the original data known as a set of support vectors. $D > 0$ is known as penalty parameter in error case. The theory of SVM tells that, data points are classified on the basis of hyper planes although it becomes unfeasible for getting linear solution in two dimensional spaces. Hence we get rid of this problem by using kernel function $k(u_i, u_j) \equiv \Phi(u_i)^N \Phi(u_j)$for multidimensional data. Utilizing divergent kernels function, SVM algorithm is trained that is expressed in below equations (4), (5) plus in equation (6).

(i) *Linear kernel*:$(u_i, u_j) = u_i^N u_j$ (4)

(ii) *Polynomial kernel*: $f(u_i, u_j) = (\gamma u_i^N u_j + r)^e, \gamma > 0$, and (5)

(iii) *Radial Basis kernel (RBF):* $(u_i, u_j) = exp(-\gamma \parallel u_i - u_j \parallel^2), \gamma$ (6)

The entire kernel arguments such as $C, \gamma, r,$ and e are initialized by utilizing the dataset. All kernel parameters are affected based upon the size of training data [25-30].

3.3 Experimental Part Elaboration

The work flow of our manuscript is executed using Intel i5 processor with 8GB hard disk and windows 10 operating system for finishing the experiment and whole resultant code is written by Python 3. 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 first step, we extracted the helical protein data information of GPCR receptor from uniprot database like, protein Id, helix name (h1-h7) and length of the protein which is combination 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 (L/V-$X_{(1-5)}$-Y-$X_{(1-5)}$-R/K) and CARC(R/K-$X_{(1-5)}$-Y/F-$X_{(1-5)}$-L/V).

Step 3: After completion of step 2 work, we move to next step where we applied our proposed algorithm spectral clustering and support vector machine 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. Table 5 shows resultant prediction of cholesterol from GPCR receptor for backward motif. 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.

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

Protein Id	Helix Name	Motif Type	Sequence	Protein Name
P30550	5	55	LSIISVYYYFIAK	Gastrin-releasing peptide
Q13585	5	55	LIVGFCYVRIWTK	Melatonin-related
P28336	5	55	LAIISIYYYHIAK	Neuromedin-B
P49683	5	55	LVILLSYVRVSVK	Prolactin-releasing peptide receptor
P32745	5	55	LVICLCYLLIVVK	Somatostatin receptor(3)
P32248	5	54	LAMSFCYLVIIR	C-C chemokine
P25025	5	54	LIMLFCYGFTLR	CX3C chemokine
O43193	5	54	LCLSILYGLIGR	Motilin
P41146	5	54	LVISVCYSLMIR	Nociceptin
Q96G91	5	54	LLTLAAYGALGR	P2Y purinoceptor 11
P51582	5	54	LVTLVCYGLMAR	P2Y purinoceptor 4
Q9UKP6	5	54	LLIGLLYARLAR	Urotensin-2 receptor
P08908	5	54	LLMLVLYGRIFR	5-hydroxytryptamine
Q9NPB9	5	54	LIMGVCYFITAR	Atypical chemokine

O43603	5	44	VLGLTYARTLR	Galanin receptor type 2
P32239	5	44	VMAVAYGLISR	Gastrin/cholecystokinin type B receptor
P41146	5	44	VISVCYSLMIR	Nociceptin
P11229	5	44	VMCTLYWRIYR	Muscarinic acetylcholine receptor
P08912	5	44	VMTILYCRIYR	Muscarinic acetylcholine receptor
Q14833	5	44	VTCTVYAIKTR	Metabotropic glutamate receptor
P16473	5	42	VIVCCCYVK	Thyrotropin receptor
Q9UBY5	7	52	VVNPIIYSYK	Lysophosphatidic acid receptor
Q8NH63	7	52	VLNPIVYSVK	Putative olfactory receptor
Q14833	7	52	VSLGMLYMPK	Metabotropic glutamate receptor
O15303	7	52	VSLGMLYVPK	Metabotropic glutamate receptor
Q14831	7	52	VALGMLYMPK	Metabotropic glutamate receptor
O00222	7	52	VSLGMLYMPK	Metabotropic glutamate receptor
P29275	7	52	VVNPIVYAYR	Adenosine receptor
P41968	7	52	VIDPLIYAFR	Melanocortin receptor
P33032	7	52	VMDPLIYAFR	Melanocortin receptor
Q96R84	7	52	VMNPLIYSLR	Putative olfactory receptor
P46092	7	45	LNPVLYAFLGLR	C-C chemokine receptor(2)
P41231	7	45	LDPVLYFLAGQR	P2Y purinoceptor 2
P30411	7	45	LNPLVYVIVGKR	B2 bradykinin receptor

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

Protein Id	**Helix Name**	**Motif Type**	**Sequence**	**Protein Name**
P49238	2	55	KSVTDIYLLNLAL	CX3C chemokine
P41143	2	55	KTATNIYIFNLAL	Delta-type opioid receptor
P41145	2	55	KTATNIYIFNLAL	Kappa-type opioid receptor
P41146	2	55	KTATNIYIFNLAL	Nociceptin
P55085	2	55	KHPAVIYMANLAL	Proteinase-activated
Q99500	2	55	KFHNRMYFFIGNL	Sphingosine 1-phosphate receptor (2)

O00421	2	55	KRVENIYLLNLAV	C-C chemokine receptor(2)
Q9Y271	2	55	KSAFQVYMINLAV	Cysteinyl leukotriene receptor
Q969V1	2	55	KTVPDIYICNLAV	Melanin-concentrating hormone receptor
Q9GZQ4	2	55	KTPTNYYLFSLAV	Neuromedin-U receptor
P31391	2	55	KTATNIYLLNLAV	Somatostatin receptor(2)
P50052	2	55	KKVSSIYIFNLAV	Type-2 angiotensin II receptor
O43193	2	55	RTTTNLYLGSMAV	Motilin receptor
Q9HB89	2	55	RTPTNYYLFSLAV	Neuromedin-U receptor
P23945	2	54	KLTVPRFLMCNL	Follicle-stimulating hormone receptor
P22888	2	54	KLTVPRFLMCNL	Lutropin-choriogonadotropic hormone
P16473	2	54	KLNVPRFLMCNL	Thyrotropin receptor
Q6W5P4	2	54	KKSRMTFFVTQL	Neuropeptide S receptor
P30559	2	54	KHSRLFFFMKHL	Oxytocin receptor
Q9NYW0	2	54	KLSTIGFILTGL	Taste receptor
P21453	2	45	RPMYYFIGNLAL	Sphingosine 1-phosphate receptor
P43220	2	45	RALSVFIKDAAL	Glucagon-like peptide 1 receptor
P34969	2	42	RQPSNYLIV	5-hydroxytryptamine receptor
P29371	2	42	RTVTNYFLV	Neuromedin-K receptor
O43613	2	42	RTVTNYFIV	Orexin receptor
O43614	2	42	RTVTNYFIV	Orexin receptor
O95977	2	42	RRWVYYCLV	Sphingosine 1-phosphate receptor
P21452	2	42	RTVTNYFIV	Substance-K receptor
P25103	2	42	RTVTNYFLV	Substance-P receptor
P35368	2	42	RTPTNYFIV	Alpha-1B adrenergic receptor
P03999	2	42	RQPLNYILV	Short-wave-sensitive opsin 1
Q86VZ1	3	42	RHHWVFGVL	P2Y purinoceptor 8

Q9BZJ6	3	42	RVSAMFFWL	Probable G-protein coupled
P21453	3	42	REGSMFVAL	Sphingosine 1-phosphate receptor (5)
P21452	3	42	RAFCYFQNL	Substance-K receptor
Q9NYW4	3	42	RYLSIFWVL	Taste receptor
P32248	3	35	KMSFFSGMLLL	C-C chemokine
P25024	3	35	KEVNFYSGILL	CX3C chemokine receptor(2)
P47900	3	35	KLQRFIFHVNL	P2Y purinoceptor 1
P41231	3	35	KLVRFLFYTNL	P2Y purinoceptor 2
P51582	3	35	KFVRFLFYWNL	P2Y purinoceptor 4
Q9NYW0	3	35	KIANFSNYIFL	Taste receptor(6)
P59541	3	34	RITAYNVWAV	Taste receptor(4)
P41231	3	32	RFLFYTNL	P2Y purinoceptor 2
P46094	6	52	RTVKLIFAIV	Chemokine XC receptor
P25024	6	52	RAMRVIFAVV	CX3C chemokine receptor
P25025	6	52	RAMRVIFAVV	CX3C chemokine receptor
P30559	6	52	RTVKMTFIIV	Oxytocin receptor
P37288	6	52	RTVKMTFVIV	Vasopressin V1a receptor
P47901	6	52	RTVKMTFVIV	Vasopressin V1b
P49019	6	32	RIHIFWLL	Hydroxycarboxylic acid
P49683	6	32	RRRTFCLL	Prolactin-releasing peptide
O14514	6	32	RSALFQIL	Brain-specific angiogenesis inhibitor 1
Q9UP38	6	32	RIGVFSVL	Frizzled-1
Q9ULW2	6	32	RIGLFSVL	Frizzled-10
Q14332	6	32	RIGVFSVL	Frizzled-2
Q9NPG1	6	32	RIGVFSIL	Frizzled-3
Q13467	6	32	RIGIFTLL	Frizzled-5

O60353	6	32	RIGVFSGL	Frizzled-6
O75084	6	32	RIGVFSVL	Frizzled-7
Q9H461	6	32	RLGLFTVL	Frizzled-8

4 Conclusion

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. 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 have clinical relevance.

References

1. Tripathy, Ramamani, Debahuti Mishra, and V. Badireenath Konkimalla. "A novel fuzzy C-means approach for uncovering cholesterol consensus motif from human G-protein coupled receptors (GPCR)." *Karbala International Journal of Modern Science* 1, no. 4 (2015): 212-224.
2. K.L. Pierce, R.T. Premont, R.J. Lefkowitz, Seven transmembrane receptors, Nat. Rev. Mol. Cell Biol. 3 (9) (2002) 639e650.
3. Tripathy, Ramamani, Debahuti Mishra, V. Badireenath Konkimalla, and Rudra Kalyan Nayak. "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* Preprint (2018): 1-10.
4. O.G. Mouritsen, M.J. Zuckermann, What's so special about cholesterol? Lipids 39 (2004) 1101e1113.
5. Y. Zhang, M.E. Devries, J. Skolnick, Structure modeling of all identified G protein-coupled receptors in the human genome, PLoS Comput. Biol. 2 (2) (2006) article e13.
6. Y. Lange, B. Ramos, Analysis of the distribution of cholesterol in the intact cell, J. Biol. Chem. 258 (1983) 15130e15134.
7. K. Burger, G. Gimpl, F. Fahrenholz, Regulation of receptor function by cholesterol, Cell. Mol. Life Sci. 57 (2000) 1577e1592.
8. D. Lingwood, K. Simons, Lipid rafts as a membrane-organizing principle, Science 327 (2010) 46e50.
9. Jafurulla, Md, Shrish Tiwari, and Amitabha Chattopadhyay. "Identification of cholesterol recognition amino acid consensus (CRAC) motif in G-protein coupled receptors." *Biochemical and biophysical research communications* 404.1 (2011): 569-573.
10. L. Anson, Membrane protein biophysics, Nature 459 (7245) (2009) article 343.
11. T. Sch€oneberg, M. Hofreiter, A. Schulz, H. R€ompler, Learning from the past: evolution of GPCR functions, Trends Pharmacol. Sci. 28 (3) (2007) 117e121.
12. R. Fredriksson, H.B. Schi€oth, The repertoire of G-proteincoupled receptors in fully sequenced genomes, Mol. Pharmacol. 67 (2005) 1414e1425.
13. C. Ellis, The state of GPCR research in 2004, Nat. Rev. Drug Discov. 3 (7) (2004) 577e626.
14. C.J. Baier, J. Fantini, F.J. Barrantes, Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicoticin acetylcholine receptor, Sci. Rep. 1 (2011) 69.

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 139 (1998) 4991e4997.
16. S. Schlyer, R. Horuk, I want a new drug: G-protein-coupled receptors in drug development, Drug Discov. Today 11 (11e12) (2006) 481e493.
17. T.J. Pucadyil, A. Chattopadhyay, Role of cholesterol in the function and organization of G-protein coupled receptors, Prog. Lipid Res. 45 (2006) 295e333.
18. X. Sun, R.G. Whittaker, Role for influenza virus envelope cholesterol in virus entry and infection, J. Virol. 77 (23) (2003) 12543e12551.
19. R.M. Epand, A. Thomas, R. Brasseur, R.F. Epand, Cholesterol interaction with proteins that partition into membrane domains: an overview, Subcell. Biochem. 51 (2010) 253e278.
20. A.K. Hamouda, D.C. Chiara, D. Sauls, J.B. Cohen, M.P. Blanton, Cholesterol interacts with transmembrane alphahelices M1, M3, and M4 of the Torpedo nicotinic acetylcholine receptor: photolabeling studies using [3H]Azicholesterol, Biochemistry 45 (2006) 976e986.
21. UniProt Consortium. "The universal protein resource (UniProt)." *Nucleic acids research* 36.suppl_1 (2007): D190-D195.
22. Arias-Castro, Ery, Guangliang Chen, and GiladLerman. "Spectral clustering based on local linear approximations." *Electronic Journal of Statistics* 5 (2011): 1537-1587.
23. Zare, Habil, et al. "Data reduction for spectral clustering to analyze high throughput flow cytometry data." *BMC bioinformatics* 11.1 (2010): 403.
24. Tripathy, Ramamani, Nayak, Rudra Kalyan, Das, Priti and Mishra, Debahuti. "Cellular cholesterol prediction of mammalian ATP-binding cassette (ABC) proteins based on Fuzzy C-Means with Support Vector Machine algorithm." *Journal of Intelligent & Fuzzy Systems Pre-press, (2020):1-8*.
25. Nayak, RudraKalyan, Debahuti Mishra, and Amiya Kumar Rath. "A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices." *Applied Soft Computing* 35 (2015): 670-680.
26. Cortes, Corinna. "WSupport-vector network." *Machine learning* 20 (1995): 1-25.
27. Zhang, Xueying, and XianghanZheng. "Comparison of Text Sentiment Analysis Based on Machine Learning." In *2016 15th International Symposium on Parallel and Distributed Computing (ISPDC)*, pp. 230-233. IEEE, 2016.
28. Upadhyay, VedPrakash, SubhashPanwar, RamchanderMerugu, and RavindraPanchariya. "Forecasting stock market movements using various kernel functions in support vector machine."In *Proceedings of the International Conference on Advances in Information Communication Technology & Computing*, p. 107. ACM, 2016.
29. Nayak, Rudra Kalyan, Debahuti Mishra, and Amiya Kumar Rath. "An optimized SVM-k-NN currency exchange forecasting model for Indian currency market." *Neural Computing and Applications* 31, no. 7 (2019): 2995-3021.
30. Chandra, Nagasuma. "Support Vector Machine Classifier for Predicting Drug Binding to P-glycoprotein." (2009).