

TM-MOTIF: an alignment viewer to annotate predicted transmembrane helices and conserved motifs in aligned set of sequences

Balasubramanian Nagarathnam¹, Kannan Sankar², Varadhan Dharnidharka³, Veluchamy Balakrishnan⁴, Govindaraju Archunan⁵, Ramanathan Sowdhamini^{1*}

¹National Center for Biological Sciences (TIFFR), UAS-GKVK Campus, Bellary Road, Bangalore 560 065, India; ²Birla Institute of Technology and Science, Pilani, India; Presently in: Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, USA; ³R.V. College of Engineering, Mysore Road, Bangalore, India; Presently in: Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, USA; ⁴Department of Biotechnology, K.S.Rangasamy College of Technology, KSR. Kalvi Nagar, Tiruchengode - 637215, Tamilnadu, India; ⁵Department of Animal Science, Bharathidasan University, Trichirapalli, Tamil Nadu, 620 024, India; Ramanathan Sowdhamini - Email: mini@ncbs.res.in; Phone: +91-080-23666001; Fax: +91-080-23636421; *Corresponding author

Received October 12, 2011; Accepted October 21, 2011; Published October 31, 2011

Abstract:

Multiple sequence alignments become biologically meaningful only if conserved and functionally important residues and secondary structural elements preserved can be identified at equivalent positions. This is particularly important for transmembrane proteins like G-protein coupled receptors (GPCRs) with seven transmembrane helices. TM-MOTIF is a software package and an effective alignment viewer to identify and display conserved motifs and amino acid substitutions (AAS) at each position of the aligned set of homologous sequences of GPCRs. The key feature of the package is to display the predicted membrane topology for seven transmembrane helices in seven colours (VIBGYOR colouring scheme) and to map the identified motifs on its respective helices /loop regions. It is an interactive package which provides options to the user to submit query or pre-aligned set of GPCR sequences to align with a reference sequence, like rhodopsin, whose structure has been solved experimentally. It also provides the possibility to identify the nearest homologue from the available inbuilt GPCR or Olfactory Receptor cluster dataset whose association is already known for its receptor type.

Availability: Package is available upon request to mini@ncbs.res.in

Key words: Transmembrane helices, Membrane topology, Amino acid conservation and substitutions, GPCR cluster association

Background:

Transmembrane proteins belong to the largest protein family and are highly important for various biological functions. Membrane proteins are implicated in various diseases and are related to more than 30 different human diseases such as cancer, diabetes, hyperthyroidism, ovarian hyper stimulation syndrome, and congenital stationary night blindness and in causing obesity [1]. Few contiguous residues, namely sequence

motifs, seem to be implicated in structural integrity and few other motifs are of relevance to diseases. LWYIK motif in HIV Type 1 Transmembrane gp41 protein for viral infection, permeabilizing motifs related to host membrane destabilization in alpha viruses, immunoreceptor tyrosine-based inhibitor motif (ITIM) in cell proliferations, motifs in protein-protein interfaces, various classical motifs in GPCRs and ORs are few examples to emphasize the importance of locating motifs in

transmembrane proteins to connect functional relevance from sequence studies. Employing the fine grained analytical approach on membrane proteins would be helpful in identifying receptor-specific motifs and to distinguish receptor subtypes of GPCR families and functional diversity among them. Several computational tools and databases have been developed in recent years to focus on issues to identify conserved residues in membrane proteins with varying principles [2, 3]. Aside from the convenience of automation due to large number of sequences, previous methods do not consider mapping the identified motifs on TM-helices. The main objective of TM-MOTIF is to map the identified motifs on the predicted membrane topology that will provide a platform for comparison of sequences across genomes (comparative genomics). The idea behind the current study is to employ computational tools effectively to identify sequence conservation in membrane proteins to connect it with functional, structural and evolutionary relationships.

Here, we implement the logic of mapping the detected conserved motifs on predicted membrane topology [4] by the proposed VIBGYOR colouring scheme which clearly distinguishes the seven helices in varying colours on MSA and helps to visualize predicted topology and conserved motifs simultaneously in the alignment window. As discussed in our recent publication [5], the package is very handy to document not only the conserved motifs, but also permitted amino acid exchanges within GPCR families at varying level of conservation at cross-genome level.

Methodology:

A flowchart (Figure 1) explains the pictorial representation of the Methodology and the step-wise procedure is described in detail below:

Inbuilt cross-genome GPCR and OR Cluster dataset

Cross-genome GPCR (Previous lab publication) [6] and OR clusters (data unpublished) were considered for the current study and are treated as in-built cluster dataset (please refer label 1 in Figure 2).

Human -Drosophila GPCR clusters: From our previous lab publication [6], phylogenetically established 32 human -Drosophila GPCR clusters of eight major receptor types such as peptide receptors (PR), chemokine receptors (CMK), nucleotide and lipid receptors (N&L), biogenic amine receptors (BGAR), secretin receptors (SEC), cell adhesion receptors (CAR), glutamate receptors (GLU) and frizzled /smoothened (FRZ) receptors were incorporated as an inbuilt GPCR cluster dataset in the package.

Human- C. elegans GPCR clusters: The selected human GPCR cluster association from previous lab publication [6] was retained where Drosophila GPCRs were replaced by introducing C. elegans GPCRs. Cluster association of C. elegans GPCRs was decided by RPS-BLAST technique, with varying E-value thresholds, and the resultant 32 human- C. elegans GPCR clusters were also considered and incorporated into the inbuilt GPCR cluster dataset (data unpublished).

Human-mouse OR clusters: Olfactory receptors (ORs) belong to Class-A type of GPCRs. 10 human-mouse OR clusters (data

unpublished) were established through neighbour-joining method and are added as “OR-sub clusters” in the package.

Alignment procedures for cross-genome GPCR/OR clusters:

Appropriate alignment tool is important in generating high-quality alignments. In our current study, the effective alignment procedures, like Clustal W [7], PRALINE™ [8] and MAFFT [9], were used to align sequences of human -Drosophila GPCR clusters, human- C. elegans GPCR clusters and human-mouse OR clusters, respectively. Different alignment methods were used since each cluster varies in parameters such as number of sequences, sequence length and percentage identity. But user-submitted queries are aligned by standalone Clustal W [7] incorporated in the package.

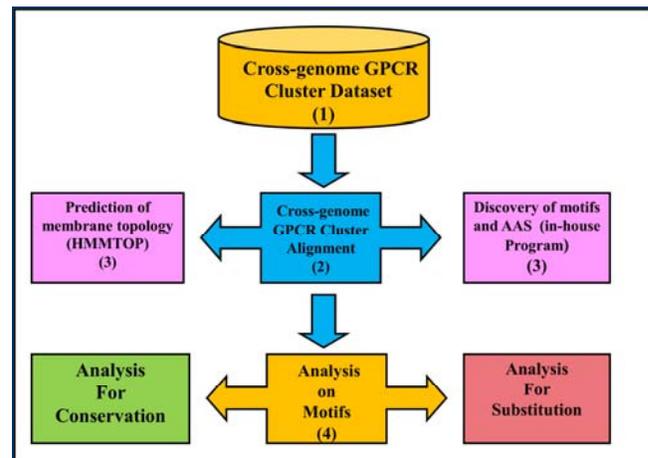


Figure 1: Flow-chart depicting the steps involved in methodology

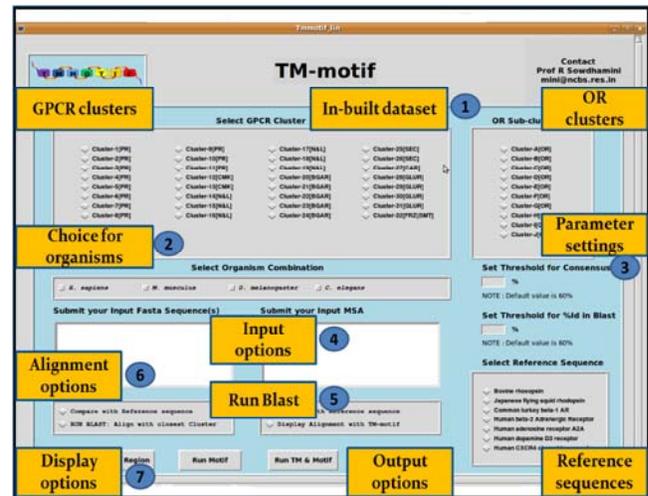


Figure 2: Snapshot for the available main menu of the front window of TM-MOTIF with user-interactive features. Note : Label 1 refers to cluster number, receptor type of the in-built dataset, Label 2 directs for select organism combinations, Label 3 denotes parameter settings to set threshold for consensus and %id in Blast, Label 4 provides input window to submit FASTA sequence or MSA, Label 5 provides option “Run Blast”, Label 6 refers option “compare with reference sequence”. Label 7 refers to the display options like “Run TM”, “Run Motif”, and “Run TM-Motif”.

Prediction of membrane topology for TM helices and loops

Each candidate receptor (GPCR or OR) from the cluster dataset was predicted for the membrane topology by using the standalone version of HMMTOP [4] added in the package. Care was taken to retain sequences having only 7(±2) TM helices for inbuilt GPCR/OR cluster dataset. Sequences which are predicted to have five to nine TM helices were alone considered for the current analysis. User-submitted query is also considered for the same cut-off and prediction procedure [4].

Detection of motifs and amino acid substitution (AAS) in the cross-genome GPCR/OR cluster alignment

Cross-genome GPCR/OR cluster alignments were taken as an input (test set) to our in-house MOTIFS program [5] to identify residue conservation and substitutions in each position of the alignment. Here, conservation simply refers to an average of all possible pairwise sequences and the score is consulted from a normalized AA exchange matrix. A motif is defined by at least three consecutive conserved AAs with amino acid conservation (more than 60% conservation score). The conservation of each residue in the set of aligned sequences was noted as 'consensus' and documented if the percentage conservation at an alignment position is from 60 to 100%. Once conserved amino acid patterns were recorded, the substituting AA residues were also identified and the properties for the AAS (like hydrophobic (@), aromatic (*), polar positive (+), polar negative (-) and polar uncharged (\$) were denoted by symbolic representation. The significant observation on preserved motifs and AAS in the helices and loop regions of the cross-genome GPCR cluster dataset was published recently [5] and the same principle was implemented in the package.

Mapping of identified motifs on TM-helices and Loops in MSA

The motifs discovered by the in-house program [5], was mapped on predicted membrane topology (refer Step2 in Methodology) of multiple sequence alignment (MSA) both for the in-built cluster alignment and user-submitted query. The predicted seven TM helices are displayed in seven varying colors such as Violet (V), Indigo (I), Blue (B), Green (G), yellow (Y), Orange (O) and Red (R) colors namely "VIBGYOR coloring scheme" along with the identified motifs, in self highlighting darker shades of VIBGYOR coloring scheme. Sequences with over predicted and under predicted TM-helices are denoted in 'pale cream color' on the MSA as we are unaware of positions for "over" and "under" predicted helices (refer option "Run TM-Motif").

Identification of homologous sequences by performing BLAST searches

User-submitted query is aligned with the nearest homologous sequence along with its pre-aligned set of sequences in the cluster by performing BLAST search (using the profile alignment option of ClustalW [7]) and the results are displayed in a new window according to the user specifications in setting parameters (refer parameters in TM-MOTIF package for more detail). The default threshold is set for the %identity is 60% to recognize hits. BLAST version 2.1.18 was incorporated to the package (refer option "Run -BLAST").

Pairwise alignment with reference Sequences

User also has an option to select any one of the listed reference sequences, whose structure has been solved experimentally and to align their sequence of interest to obtain a pairwise alignment

in the proposed VIBGYOR coloring scheme and aligned by CLUSTALW [7] (refer option "select reference sequence").

Requirements for TM-MOTIF package execution

This tool has been coded in Perl language (using Tk module for GUIs). The package can be executed in LINUX OS and requires the following backhand programs to be installed prior to use: PerlTk, BioPerl, FORTRAN compiler and standalone versions of ClustalW and BLAST2. TM-MOTIF package shall be made available to users for academic purposes upon request to mini@ncbs.res.in with immediate help for installation and use.

Software input and output options:

The main menu of the front window of the TM-MOTIF package describes available input and output options (also referred to as display options), parameter setting and choice of available organisms (refer to Label 1-7 in Figure 2 and Figure 4).

Input Options (refer Label 4 in Figure 2)

The user can submit sequence in FASTA format or MSA by using the available text box given as 'Submit your Fasta sequence(s),' 'Submit your Input MSA'. There are also choices given to select model organism, GPCR or OR cluster dataset. The cluster number, receptor type and organisms are mentioned in the main menu of the package (please refer to Labels 1 and 2 in Figure 2).

Output Options (refer Label 7 in Figure 2)

Display of predicted 7 TM- helices in VIBGYOR coloring scheme: (by using "Run TM" option)

The predicted TM helices 1-7 are highlighted in seven different colors (VIBGYOR coloring scheme) (Figure 5a). This facilitates the user to view the transmembrane proteins in long sequence alignments and to keep track of membrane topology along with consensus of residue conservation patterns at each alignment position.

Display of motifs and AAS in MSA: (by using "Run Motif" option)

If the user is interested to record only the identified motifs in the selected cluster or user-provided MSA, it is possible to display only the identified motifs and AAS in the MSA (using "Run motif" option). Here, observed motifs are highlighted in "grey color" on MSA (Figure 5b). The conserved amino acid residues are displayed below the alignment as "text" and referred as "consensus". Besides, mouse-over option is also provided on each residue in MSA to guide the user to document the observations like alignment position, percentage residue conservation, amino acid substitution (AAS), and property of amino acid substitutions in MSA (Figure 5b).

Display of motifs on TM-helices: (by using "Run TM-Motif" option)

It is also possible to display identified motifs and predicted membrane topology simultaneously on user-selected cluster or user-submitted MSA. This is probably the most crucial output delivered by the package since such an annotated alignment is biologically meaningful. As discussed by using "Run TM" option, the predicted membrane topology are displayed in VIBGYOR coloring scheme, and the embedded motifs in the alignment are denoted with the self-highlighting colors within the VIBGYOR coloring scheme (Figure 5c). Display of "consensus" and navigating mouse-over option at each

position of the alignment are also available to facilitate the observation on motif properties with corresponding output files for documentation are also generated at each level (please refer to output files for more details).

Display of motifs on intra- and extracellular loops

Predicted intra- and extracellular loop regions [4] are displayed in plain background between the VIBGYOR colored TM helices and the detected motifs in the loops are highlighted in “grey color” (refer to Figure 5b). Details about “consensus” are also given. This may facilitate understanding of sequence properties at the predicted loop regions for comparative sequence analysis and to generate loop libraries.

Alignment with reference sequence: (by using “compare with reference sequence” option)

User-submitted sequence can be aligned with any one of the selected reference sequences by using the option “Select reference sequence” (please refer to Label 6 in Figure 2). There are about seven reference sequences (such as bovine rhodopsin, Japanese flying squid rhodopsin, common turkey β -1 AR, human β -2 Adrenergic receptor, human adenosine receptor A2A, human dopamine D3 receptor and human CXCR4 chemokine receptor) whose crystal structures were solved experimentally and are available for the user to select as relevant reference sequence or as template. This particular option helps the user to prepare a pairwise alignment with their sequence of interest, which in future can guide to generate a reliable 3D-model by homology modeling (Figure 3 for the output of TM-MOTIF package).

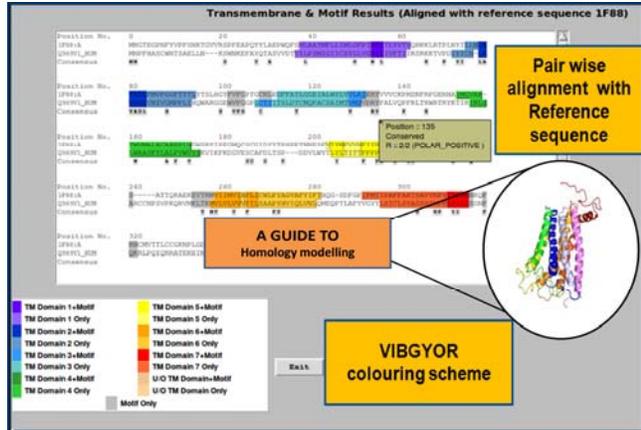


Figure 3: Snapshot for the pairwise alignment of user-submitted sequence with selected reference sequence (note: predicted topology is highlighted in VIBGYOR coloring scheme with highlighted motifs as consensus and navigating mouse -over options to display position ,conservation and AAS of the alignment).

Identifying closest homologues of user-submitted sequence in selected organisms: (by using “Run blast: Align with closest cluster” option)

When the user is interested to search the homologous sequence for their sequence of interest from the available GPCR and OR cluster dataset, “Run-blast” option is highly useful. The cluster from which maximum number of hits was obtained by BLAST search for the given query against the in-built GPCR and OR clusters is chosen for the resultant alignment and for

deciphering receptor type and functional relevance (please refer Label 5 in Figure 2).

Default settings of parameters:

The default cut-off for recognizing hits is 60% percentage identity in BLAST (please refer to Methodology) and the user has an opportunity to alter these values. The default threshold is fixed as 60% for recognizing conserved residues in “consensus” (please refer to Label 3 in Figure 2). By using “Select Organism Combination”, user can select any two organisms from the available dataset which includes *H. sapiens*, *M. Musculus*, *D. Melanogaster* and *C. elegans* to display cluster-specific or cross-genome cluster alignments. Options like “Select GPCR Cluster” and “OR-subclusters” (also refer Step 1-1.3 in Methodology) are helpful for the user to select their interested cluster type for intra- and inter-genomic cluster alignments. Clusters are noted with respective cluster number and receptor types in the main menu of the front window (Figure 2). User can select any one of the reference sequence in the given option “Select reference Sequence” or use “Run-blast” (also refer 5, 6 in Methodology) to recognize closest cluster for pairwise alignment or for building MSA.

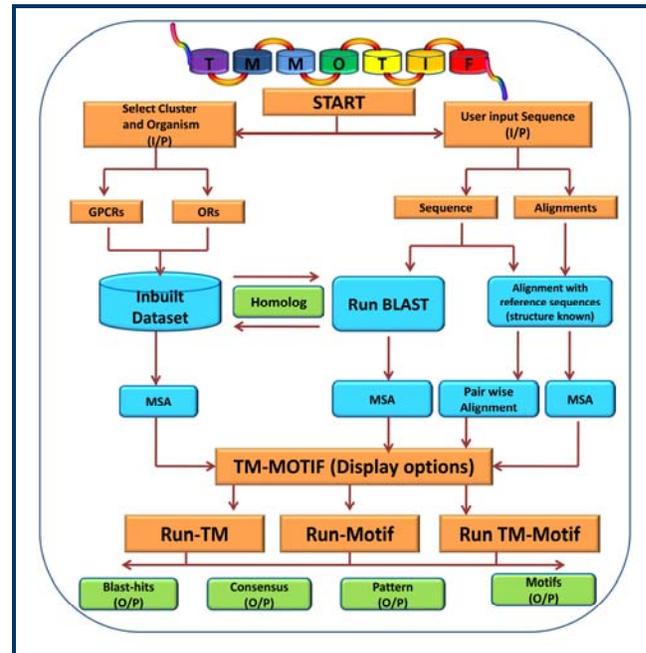


Figure 4: Pictorial representation for the tool guide of TM-MOTIF which depicts the available options: User can start with selecting the in-built cluster dataset, organisms (pipelines referred in Left side), and viewing the MSA for the available display options of the TM-MOTIF (refer a, b, c of Figure 5) i.e., “Run TM”, “Run Motif”, “Run TM-Motif” options. The User can also start by submitting their sequence of interest or alignment (pipelines referred in Right side), and can perform “Run-blast” (refer section 6 in Methodology), to search for homologues against in-built cluster dataset. By selecting “Alignment with reference sequence” (refer section 6 in Methodology) user get benefit by obtaining any one of the Display options -“Run TM”, “Run Motif”, “Run TM-Motif” (refer Figure 3 also). The four important output files are generated as given in the readme file of the package (also refer Output files section).

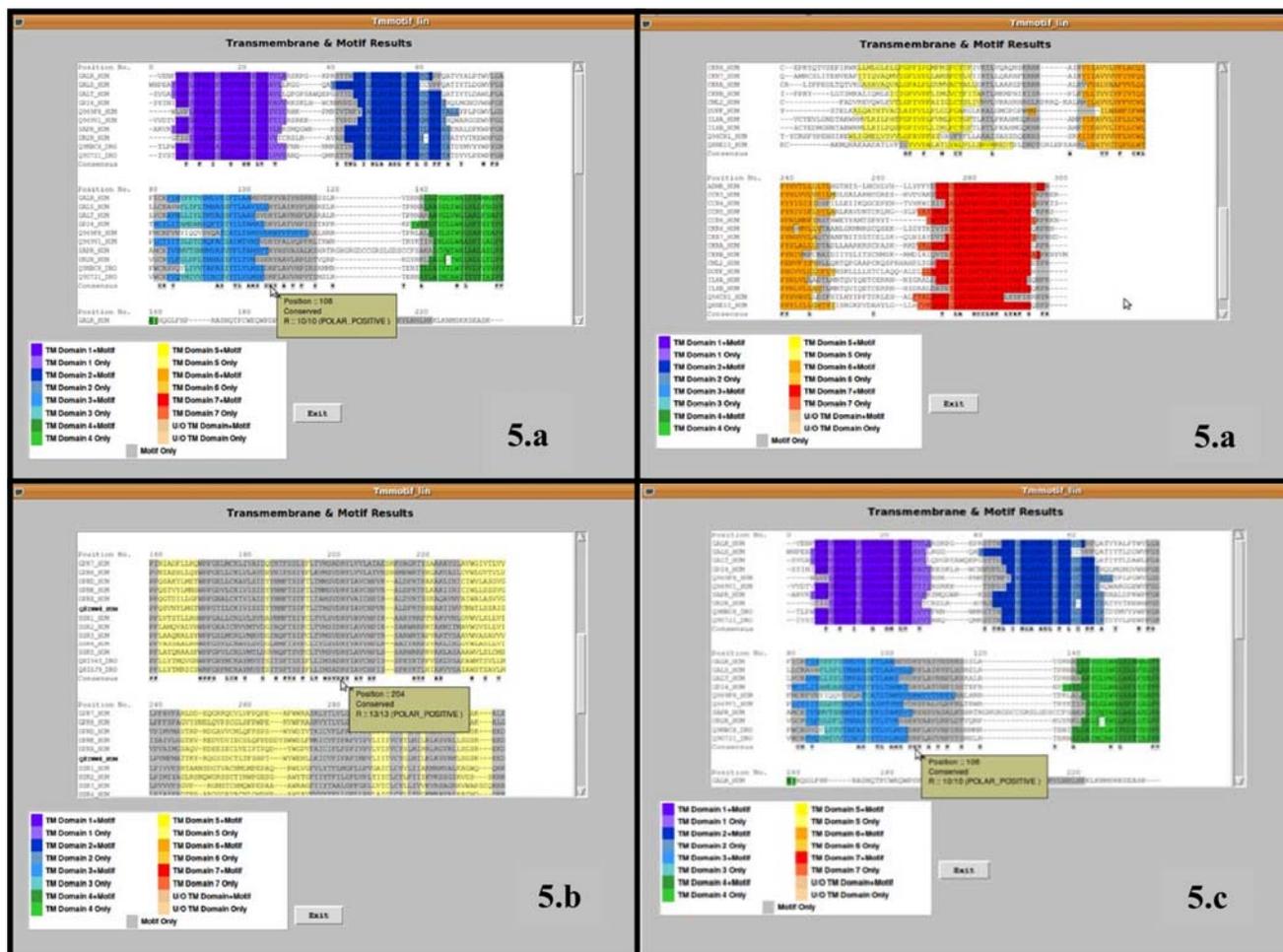


Figure 5: Snapshot of results after running TM-MOTIF package. Snapshot 5a depicts the display of predicted seven TM-helices in VIBGYOR coloring scheme. (Sample output for the option “RUN -TM”, given in two snapshots where first window for the display of alignment till TM IV (green color), next window with the rest of the three helices i.e., till TM VII (in red color)). Snapshot 5b depicts the display of identified motifs in MSA along with the mouse-over options (Sample out-put for the option “RUN - Motif”). Snapshot 5c depicts the display of identified motifs on predicted seven TM helices (Sample output for the option “RUN -TM-motif”, where alignment window is given till TM IV (green color)).

Output files:

Along with the previously discussed display options, three important output files are also generated. A detailed report on all the observed consensus residues with amino acid position for observed threshold values, including the percentage of conservation and percentage of substitutions with respective residue type is generated in output file called Zconsensus.txt. A list of all the conserved amino acid residues and substitutions observed at the motif sites of the alignment are given in Zpattern.txt. The consolidated result file, namely Zmotif.txt, provides the list of motifs with substitutions identified along

with their start and end positions of the alignment. The alignment of sequences submitted by the user can be obtained in ClustalW and PIR format from the files Zuser.aln and Zuser.pir respectively. For the option “Run BLAST”, the number of hits observed from in-built cluster can be obtained from the file called Zblast_sorted.txt (also refer Read Me file of the package).

Utility:

The result files, particularly Zmotif.txt file, helps the user to document the cluster-specific or receptor-specific motifs and

hence to discover novel motifs at inter- and intra-genomic level. The generated Zpattern.txt file is most helpful in preparing the pattern, and with minimal editing, can be used to define sequence pattern for interested receptor types and to supply these pattern files to perform PHI-blast. Resultant Zblast_sorted.txt is effective to identify nearest homologues for the query of interest to the user from the available in-built cluster dataset. In the interest of observing the influence of critical parameters such as average sequence length, percentage identity and conservation in the clusters, a pilot study on Human-only GPCR clusters, Human-*Drosophila* and Human-*C. elegans* GPCR cluster dataset was done (**Table S1, S2 in supplementary material**) and infer the dependency of conservation by the other mentioned factors crucially.

Conclusion:

TM-MOTIF, a software package and an alignment viewer, helps to map discovered motifs on predicted TM- helices and loop regions in MSA and providing effective understanding on residue conservation in the long sequence window with least computational time and strain. The applied VIBGYOR coloring scheme for TM-helices not only helps the user to keep track of the current location of membrane topology and relative positions of observed motifs in the large sequence alignment, but also to identify the influence of predicted secondary structures (helix or loop) or topology on conservation. By running a Blast search against the given GPCR and OR cluster dataset, the user can collect the nearest sequence homologue from the cluster dataset which further helps to correlate with functional similarities across genome. The pairwise alignment in VIBGYOR coloring scheme with selected reference sequence (structure known) could be a very useful starting step for the 3D modeling. By using the option “organisms combination”, opportunity is provided to the user to understand the sequence properties at cross-genome level and the package is highly suited to perform comparative genomics studies and to identify receptor-specific motifs. Also, the given navigating mouse-over option assists the user to obtain knowledge on position, amino acid conservation (motifs), amino acid substitutions (AAS) in each position of the alignment for better understanding of the

sequence properties to connect with evolutionary trends passed within and across genomes in the particular interest of transmembrane proteins.

Caveat & Future Development:

TM-MOTIF package could be extended to other genomes and membrane-bound helical proteins like ion channels and transporters in future. TM-MOTIF could evolve into a specialized alignment viewer for transmembrane helix-rich proteins with added features such a graphical display to provide a 2D cartoon representation of the helix topology embedded with identified motifs.

Acknowledgement:

This work is supported by Department of Biotechnology, India. We thank S. Kalaimathy and R. Archana for involvement during the clustering of GPCRs and the initial sequence analysis, respectively. We thank Harini and Aniket Sengupta for critical feedback about the package. We also thank NCBS (TIFR) for infrastructural facilities.

References:

- [1] Schoneberg T *et al. Pharmacol Ther* 2004 **104**: 173
- [2] Tusnady GE *et al. Bioinformatics* 2008 **24**: 1469 [PMID: 18434342]
- [3] Marsico A *et al. Nucleic Acids Res.* 2010 **38**: 141 [PMID: 19910368]
- [4] Tusnady GE & Simon I, *Bioinformatics* 2001 **17**: 849 [PMID: 11590105].
- [5] Nagarathnam *et al. Bioinformation* 2011 **7**: 15 [PMID: 21904433]
- [6] Raghu Prasad Rao Metpally & Ramanathan Sowdhamini, *BMC Genomics* 2005 **6**: 1
- [7] Thompson JD *et al. Nucleic Acids Res.* 1994 **22**: 4673 [PMID: 7984417]
- [8] Pirovano W *et al. Bioinformatics* 2008 **24**: 492 [PMID: 18174178]
- [9] Katoh K *et al. Nucleic Acids Res.* 2002 **30**: 3059 [PMID: 12136088]

Edited by P Kanguane

Citation: Nagarathnam *et al. Bioinformation* 7(5): 214-221 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Input Options [I/P] in TM-MOTIF package

Visualizing inbuilt data:

To visualize inbuilt data, users must (a) Select a Cluster, (b) Select (an) Organism(s) and (c) Select one of the three output options ('Run TM', 'Run Motif', or 'Run TM-Motif'). The user also has an option to set a cut-off value, a conservation above which would be considered as a consensus residue. If the user does not submit any value, a default value of 60% will be taken.

Submitting user's sequence:

Users can also submit their sequence(s) of interest in FASTA format or in MSA in the given text-box for the same by using the options "Submit your input FASTA sequence(s)" or "Submit your input MSA". There are two options available for users: (a) Align with the reference sequence or (b) Run BLAST search within the tool to search for the closest homolog from the inbuilt cluster dataset. After selecting anyone of the above options, users must select one of the three output options for display i.e., "Run TM" option or "Run Motif" option, or "Run TM-Motif" option and parameter settings as per requirement for their submitted sequence in the FASTA format or MSA. Further, the user has an option to specify the cut-off values (refer also Default settings of parameters) for checking conserved residues as well as a cut-off value for the percentage identity for a sequence in a cluster to be considered as a hit.

[O/P] in TM-MOTIF package

Annotation of Transmembrane helices on the alignment:

On selecting the "Run TM option", the sequence alignments will be displayed in a new window for the predicted transmembrane helices in "VIBGYOR coloring scheme". The sequences which are predicted for more than 7 TM helices (over-predicted) are displayed in a "pale cream color" to distinguish them from the others. The "consensus" (as per the cut-off given by the user) will also be displayed below the alignment as text and along with a mouse-over option by which the user can document conserved motifs and AAS at each position of the alignment.

Annotation of Motifs on the alignment:

On selecting the "Run Motif" option, discovered motifs from the user selected cluster, or user submitted sequence in FASTA or MSA will be mapped on the alignment and displayed in a new window. The identified motifs are highlighted in the "grey colour" and the rest of the residues in the alignment are noted in "pale yellow color". The display of consensus and mouse-over option is available for documenting motifs and AAS.

Table S1: Average length, identity and residue conservation observed in human-only GPCR clusters for 60% of conservation.

Cluster No	Human-only GPCR Clusters		
	Average length	Average identity	Average conservation
1	398	30	31
2	381	42	46
3	397	29	31
4	454	29	32
5	420	29	38
6	405	29	30
7	804	29	30
8	475	28	35
9	385	31	35
10	346	43	52
11	416	30	24
12	368	41	58
13	373	31	30
14	377	31	23
15	361	26	22
16	355	34	33
17	395	30	27
18	402	32	32
19	370	27	21
20	354	45	50
21	440	38	49
22	420	32	36
23	420	24	17
24	453	25	17
25	516	34	37

26	1425	18	17
27	989	49	55
28	935	29	48
29	397	39	40
30	915	23	53
31	640	42	57

Table S1 is listing out the available 31 human only GPCR clusters with observed sequence properties such as average sequence length, identity, residue conservation at 60% level of conservation.

Table S2: Average length, identity and motif conservation observed in Human-*Drosophila* and Human- *C. elegans* GPCR clusters

Cluster No	Human and <i>Drosophila</i> GPCR Clusters			Human and <i>C. elegans</i> GPCR Clusters		
	Average length	Average identity	Average conservation	Average length	Average identity	Average conservation
1	288	30	29	391	15	15
2	285	44	39	365	15	14
3	284	31	29	369	14	13
4	305	30	30	372	14	14
5	331	25	19	384	12	13
6	288	28	25	379	14	13
7	342	32	22	449	16	13
8	317	30	28	394	16	15
9	292	33	26	349	14	13
10	260	43	49	332	17	14
11	303	30	26	390	13	13
12	281	44	54	345	18	16
13	277	29	28	370	18	16
14	283	31	22	349	15	13
15	284	27	21	357	14	13
16	284	35	32	347	15	13
17	269	32	29	380	16	14
18	312	30	26	354	14	14
19	276	28	30	354	14	13
20	292	40	44	349	20	17
21	330	32	29	346	18	14
22	394	30	31	385	15	15
23	383	23	18	394	12	13
24	384	27	25	451	19	18
25	289	38	31	411	16	15
26	315	28	23	359	20	16
27	567	24	14	1068	14	22
28	492	52	50	591	19	18
29	274	32	47	611	17	17
30	260	31	44	348	17	14
31	439	22	21	542	16	16
32	362	40	45	421	18	16

Table S2 is listing out the available 32 Human-*Drosophila* and Human- *C. elegans* GPCR clusters with observed sequence properties such as average sequence length, identity, residue conservation at 60% and 30% of conservation respectively.