bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MBPpred: Proteome-wide detection of membrane lipid-binding proteins using profile Hidden Markov Models

Katerina C. Nastou, Georgios N. Tsaousis, Nikos C. Papandreou and

Stavros J. Hamodrakas*

Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 15701, Greece

*To whom correspondence should be addressed

Emeritus Prof. Stavros J. Hamodrakas

Department of Cell Biology and Biophysics, Faculty of Biology,

University of Athens, Panepistimiopolis,

Athens 15701, Greece

Phone: +30 210 727 4931

Fax: +30 210 7274254

e-mail: shamodr@biol.uoa.gr

http://biophysics.biol.uoa.gr

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

A large number of modular domains that exhibit specific lipid binding properties are present in many membrane proteins involved in trafficking and signal transduction. These domains are present in either eukaryotic peripheral membrane or transmembrane proteins and are responsible for the non-covalent interactions of these proteins with membrane lipids. Here we report a profile Hidden Markov Model based method capable of detecting Membrane Binding Proteins (MBPs) from information encoded in their amino acid sequence, called MBPpred. The method identifies MBPs that contain one or more of the Membrane Binding Domains (MBDs) that have been described to date, and further classifies these proteins based on their position in respect to the membrane, either as peripheral or transmembrane. MBPpred is available online at http://bioinformatics.biol.uoa.gr/MBPpred. This method was applied in selected eukaryotic proteomes, in order to examine the characteristics they exhibit in various eukaryotic kingdoms and phylums.

KEYWORDS

membrane binding domains, membrane binding proteins, peripheral membrane proteins, profile Hidden Markov Models, membrane lipids

ABBREVIATIONS

MBD(s): Membrane Binding Domain(s)

MBP(s): Membrane Binding Protein(s)

pHMM(s): profile Hidden Markov Model(s)

MCC: Matthew's Correlation Coefficient

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. INTRODUCTION

A cell's structure and functions rely significantly on membranes, since they are responsible for its compartmentalization and are associated with nearly half of all its proteins [1]. Membrane proteins are of central importance as they take part in a large variety of cellular functions such as macromolecular transport and signal transduction [2]. These proteins can either be embedded directly within the lipid bilayer (transmembrane proteins), or can be associated with the membrane indirectly via interactions with membrane proteins or lipids (peripheral membrane and lipid-anchored proteins).

Peripheral membrane proteins interact non-covalently with the membrane, either directly via membrane lipids or indirectly with transmembrane proteins. Directly interacting membrane proteins usually have domains that allow for the specific or non-specific interaction with membrane lipids. Besides peripheral membrane proteins, these domains are also present in extramembranous regions of transmembrane proteins [3] – either intracellular or extracellular – and are known as Membrane Binding Domains (MBDs). MBDs are of great importance to the cell, since proteins that contain such domains take part in a variety of cellular processes such as cell signaling and membrane trafficking, vital for the cell's survival and growth. Although domains with membrane lipid binding properties have recently been reported to exist in prokaryotes too [4], the main focus of experimental studies is on eukaryotic membrane binding proteins.

MBDs are extremely diverse and their only common characteristic is their non-covalent interaction with membrane lipids, with different affinities. A significant number of MBDs have been identified to date. While some of them, like C2, and BAR [5] have been extensively studied in the last decades, mainly with experimental methods, there is a growing number of recently identified MBDs for which very little is known, such as IMD and GOLPH3 [6]. Structural studies have aided in the elucidation of the interactions of MBDs with the membrane. However, the search of new membrane binding domains with experimental methods would be immensely time-consuming and expensive. Thus, the development of genome-wide prediction methods for the detection of membrane binding proteins is necessary.

A large number of Membrane Binding Proteins (MBPs) act as enzymes by recognizing specific lipid head groups. Mutations of these proteins affect their molecular function, and a number of diseases have been described, that are attributed to the malfunction of these proteins [7]. Despite their importance, MBPs have not been studied comprehensively with computational methods. Only two methods that allow for the detection of peripheral proteins from the existence of such domains have been reported to date. The first method, developed in 2006 [8], was based on structural characteristics of these proteins and the second, developed in 2010 [9], on information encoded in amino acid sequence. However, neither one of these methods is currently available online.

The comprehension of the molecular mechanisms that Membrane Binding Proteins use to perform their functions will be extremely significant for the unraveling of their

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

activity inside cells. The augmentation of large scale proteomic and computational studies of Membrane Binding Domains and proteins harboring them, will aid immensely towards achieving this goal in the next few years.

We report here the design and development of a sequence-based method that identifies Membrane Binding Proteins in eukaryotic proteomes with the use of profile Hidden Markov Models (pHMMs), specific to membrane binding domains (MBDs). The method also classifies Membrane Binding Proteins (MBPs) according to their relationship with the membrane, and thus allows for the detection of peripheral membrane proteins.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. METHODS

After an extensive literature search 18 domains were identified (Annexin, ANTH, BAR, C1, C2, ENTH, Discoidin, FERM, FYVE, Gla, GOLPH3, GRAM, IMD, KA1, PH, PX, PTB, Tubby) for which well-established biochemical and crystallographic experimental data for the interaction with membrane lipids exist. Each of these domains was mapped to at least one characteristic pHMM from the Pfam database [10], since in our case the majority of these profiles are well defined. Subsequently, a pHMM library (MBDslib) containing 40 pHMMs that were derived from Pfam was created. The mapping between the different pHMMs and the 18 MBDs is shown in Table 1.

Table 1. Mapping between the pHMMs of MBDslib and known MBDs. In the first column of this table the Membrane Binding Domains that were isolated from literature are shown and in the second column the unique pHMM identifier from the Pfam database.

MBD	pHMMs
Annexin	PF00191
ANTH	PF07651
BAR	PF03114, PF10455, PF16746
C1	PF00130, PF03107, PF07649
C2	PF00168, PF14429, PF00792, PF10409
Discoidin	PF00754
ENTH	PF01417
FYVE	PF01363, PF02318
Gla	PF00594
IMD	PF08397
KA1	PF02149
PH	PF00169, PF08458, PF14593, PF15404, PF15405, PF15406, PF15409, PF15410, PF15411, PF15413, PF16457, PF16652, PF14844
PTB	PF08416
GOLPH3	PF05719
PX	PF00787
Tubby	PF01167
GRAM	PF02893
FERM	PF00373, PF09380, PF09379

The MBPpred algorithm consists of two levels: a detection and a classification level. To develop the detection level of MBPpred, the HMMER package was utilized in order to search the pHMM library MBDslib, and detect Membrane Binding Proteins (MBPs) from a set of protein sequences. Using the hmmsearch program of HMMER, one can "search" the aforementioned library, and thus identify proteins which belong to the families used to create the library and, subsequently, find MBPs from a set of protein sequences. The classification level of MBPpred was created, in order to distinguish MBPs into transmembrane and peripheral membrane proteins with the use of the PredClass algorithm [11]. PredClass classifies proteins into four distinct

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

classes, namely membrane, globular, mixed and fibrous. Proteins, in the first class are actually only transmembrane proteins, while MBPs in the second and third class are considered peripheral MBPs.

MBPpred was evaluated using two datasets that were assembled from PDB [12] and was also compared with the predictor, which was developed by Bhardwaj et al. [8], in 2006.

In order to create non-redundant sequence datasets BLASTClust [13] was used, both for the creation of the positive and the negative datasets. Protein sequences with less than 30% sequence identity with each other, in a sequence length coverage of 90%, were retrieved using this program. The positive dataset consists of known MBPs, which target the membrane via MBDs. Initially it consisted of 202 proteins, 71 of which were non-redundant. After the removal of proteins present in the seed sequence sets of the pHMMs used to create MBDslib, the final positive dataset was assembled, which consists of 49 non-redundant proteins. The negative dataset was retrieved from a PDB search for eukaryotic proteins that do not have membrane or lipid binding properties, as described in their PDB files and contained 9057 non-redundant sequences. 500 sequences were randomly chosen from this dataset, in an attempt to balance the negative and positive datasets, while maintaining the information needed to evaluate our method (Table S1). In order to compare MBPpred with the predictor developed by Bhardwaj et al. [8] more precisely, the two datasets introduced in that study were used, one of membrane and one of non-membrane binding proteins, both with known three-dimensional structures. The negative dataset consists of 225 proteins and the positive of 35 proteins (Table S1). It should be noted, that this method used only 9 (ANTH, C1, C2, ENTH, FYVE, PH, PX, Tubby, BAR) out of the 18 MBDs incorporated in MBPpred.

For the prediction performance of MBPpred five measures were used, namely Accuracy, Sensitivity, Specificity, Balanced Accuracy and Matthew's Correlation Coefficient. True/false positives (TP, FP) and true/false negatives (TN, FN) were counted on a per protein basis.

Accuracy is the proximity of measurement results to the true value and is calculated as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1).$$

Sensitivity, or true positive rate is:

$$Sn = \frac{TP}{(TP + FN)} \quad (2.2),$$

and Specificity, or true negative rate is:

$$Sp = \frac{TN}{(TN + FP)} \quad (2.3).$$

Besides these measures, the balanced accuracy and Matthew's Correlation Coefficient (MCC) were used to appraise the performance of MBPpred. Balanced accuracy is the average of sensitivity and specificity and, together with MCC, is considered a better

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

measure [14] when the data sizes of the positive and negative datasets are not balanced. MCC is calculated as:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{((TN + FN) \cdot (TN + FP) \cdot (TP + FN) \cdot (TP + FP))}} \quad (2.4).$$

Moreover, MBPpred was applied to 30 selected (Table S2) and all 407 reference eukaryotic proteomes (Table S3) retrieved from UniprotKB [15] (release: 2015_04) in order to identify potential membrane binding proteins that interact with lipids non-covalently and to perform a quantification analysis regarding these proteins.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. RESULTS AND DISCUSSION

3.1. The MBPpred Algorithm

The detection level of MBPpred uses a library of 40 pHMMs, which correspond to 18 Membrane Binding Domains (MBDs) that were identified from literature. This library is used for the detection of Membrane Binding Proteins (MBPs). If, during a search of the library with HMMER, the score of an alignment between a query protein and at least one of the profiles is higher than the gathering threshold of each pHMM (as reported in Pfam), then the protein is characterized as a MBP. An analysis was performed, where different scoring thresholds than those defined by Pfam, were used. This analysis showed that, when tested against the proteins of the evaluation dataset, best results were retrieved with the use of the gathering thresholds and not with other more or less strict cut-offs (Table S4). Proteins that score higher than the threshold for at least one of the domains, in the library, are characterized as possible membrane binding.

The classification level of MBPpred uses the PredClass algorithm in order to classify MBPs, in respect to their interaction with the membrane, into peripheral or transmembrane. PredClass's speed and the use of information solely encoded by amino acid sequences makes this algorithm suitable for the implementation of the classification level of our algorithm.

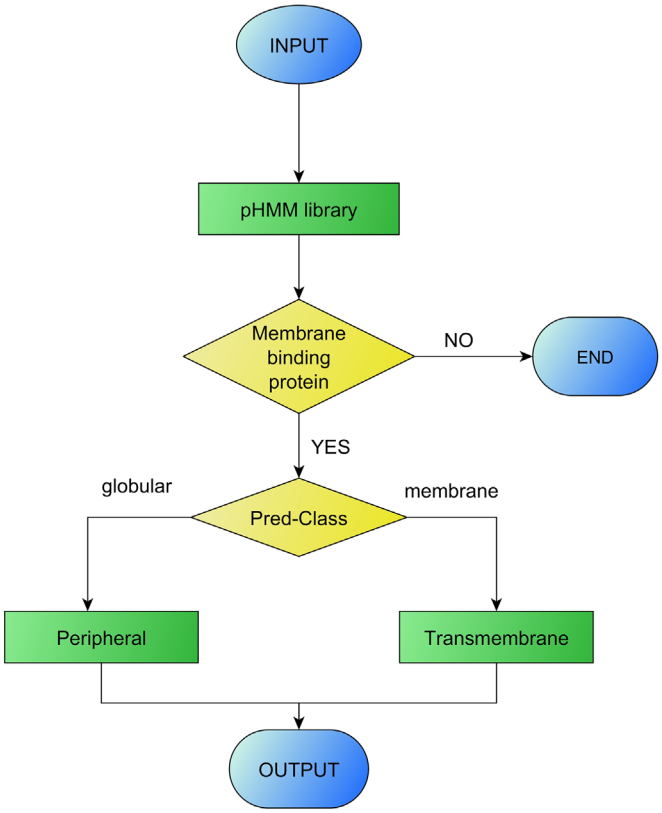

Figure 1. Flowchart of the MBPpred algorithm

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.2. Web interface of MBPpred

A web interface has been developed for MBP-Pred and the method is publicly available through http://bioinformatics.biol.uoa.gr/MBPpred . Through the main page, the user can access the query submission page, the manual and contact pages. Query submission can be performed by either pasting a single or a set of protein sequences in the textbox provided, or by uploading a file with fasta formatted sequences. Even though the method is meant to be used and has been tested with eukaryotic proteins, a user cannot be prevented of using the method on prokaryotic sequences; unfortunately the results from such submissions cannot be interpreted correctly.

After a successful query, submission users are transferred to the results page where they can gather information about their submission, as well as extensive information in downloadable files about MBPs (if any). The final results files contain a protein identifier, the position and score of the domain(s) present in the protein and the type of membrane protein (peripheral or transmembrane) along with its sequence and length. The output files produced by MBPpred and their contents are shown in Fig. 2. MBPpred is fast, since for a query length the size of the human proteome the algorithm produces results in ca. five minutes, which makes it sufficient for proteomic scale applications.

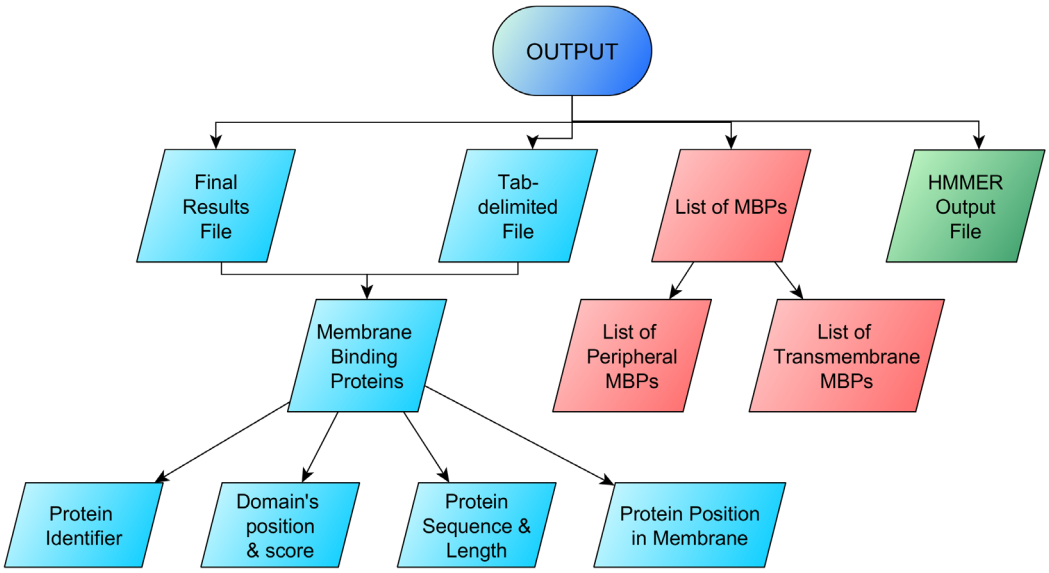

Figure 2. Output files produced by MBPpred and their contents

3.3. Evaluation of MBPpred

Our method was evaluated as a means to measure its performance against a non-redundant dataset of 49 membrane and 500 non-membrane binding proteins. Our method is accurate since it can detect all the proteins from the positive dataset as such (Sensitivity = 100%), while it falsely detects a very small percentage of non-MBPs as MBPs (Specificity = 97.0%) as shown in Table 2.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Results from the evaluation of MBPpred against the datasets assembled from PDB

Method	*Accuracy*	*Sensitivity*	*Specificity*	*Balanced Accuracy*	*MCC*
MBPpred	97.3%	100%	97.0%	98.5%	0.86

In addition, MBPpred was compared with the predictor developed by Bhardwaj et al. [8].

Table 3. Comparison of MBPpred with the SVM method developed by Bhardwaj et al. (2006)

Method	*Accuracy*	*Sensitivity*	*Specificity*	*Balanced Accuracy*	*MCC*
MBPpred	**97.7%**	88.2%	**99.1%**	**94.4%**	**0.91**
Bhardwaj et al.	90.1%	**90.6%**	92.4%	91.5%	**-**[a]

[a] MCC could not be calculated for this method since it is no longer publicly available. The other measures of performance were retrieved from the respective paper.

MBPpred outperforms this method, as shown in Table 3. We should note here that, our method could not be evaluated against the more recent method developed by Bhardwaj et al. [9], because the datasets used are not provided and none of the aforementioned methods are available online.

3.4. Application of MBPpred in eukaryotic proteomes

The application of MBPpred in 30 eukaryotic reference proteomes showed that, ca. up to 6.0% of the proteins in these proteomes are possible MBPs (Fig. 3). The percentages vary based on the kingdom and phylum in which these organisms belong. In general, animals have more MBPs than fungi and plants, while other eukaryotes have a great divergence in the proportion of MBPs in their proteomes, whereas in general, organisms that are evolutionary closer to plants have less MBPs than organisms closer to animals and fungi.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

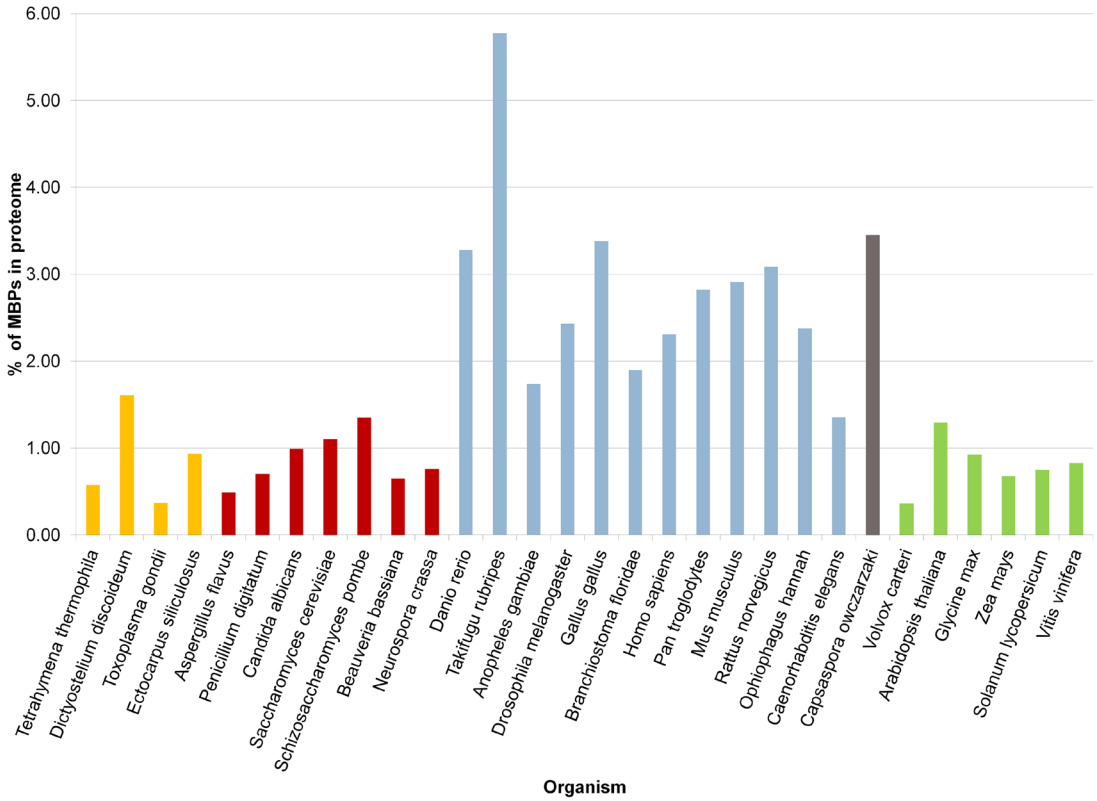

Figure 3. The percentage of MBPs in each of the 30 eukaryotic proteomes where MBPpred was applied. In general, Fungi (shown in red) and Plantae (shown in green) have less than 1.5% MBPs in their proteomes, while Animalia (light blue) have more than 1.5%. Organisms that belong to other eukaryotic kingdoms, like Amoebozoa, (orange and grey) have varying percentages of MBPs in their proteomes.

An enrichment analysis of the MBPs of 20 out of the 30 proteomes was performed (Table S5) using the DAVID functional annotation tool [16], in order to assess the functions of these proteins. Functional enrichment analysis could not be performed for MBPs from 10 proteomes (*Dictyostelium discoideum, Ectocarpus siliculosus, Candida albicans, Penicillium digitatum, Beauveria bassiana, Ophiophagus hannah, Capsaspora owczarzaki, Volvox carteri, Glycine max* and *Solanum lycopersicum*) because these proteomes have not been annotated with gene ontology terms. In all cases, terms related to lipid binding and certain membrane binding domains are overrepresented, as expected. Moreover, other terms associated with membrane trafficking and signal transduction are enriched, indicating the importance of MBPs in these cell processes. In particular, Gene Ontology (GO) [17] terms like regulation of Ras, Rho, small GTPase mediated and ARF protein signal transduction, protein kinase activity, endocytosis, cell junction and cytoskeleton organization are overrepresented in MBPs. There is no particular pattern in enriched terms – in any of the 3 major eukaryotic kingdoms – that would help us explain the differences in the percentages of MBPs between the studied organisms.

The classification of MBPs in peripheral membrane and transmembrane proteins, showed that in all cases peripheral MBPs are more than transmembrane. The deviation of the percentages in various kingdoms regarding MBPs can be attributed to

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the evolutionary diversity of these organisms (Fig. 4). Small differences in the membrane lipid and protein composition between these eukaryotes can be the cause of variability in the number of MBPs. Moreover, the big difference between animals and other eukaryotes can be attributed to the cell membrane differences of plants, fungi and animals, which consequently leads to differences in membrane protein composition [18]. Plants and fungi use different mechanisms to perform similar functions, like some of the functions in which MBPs take part in, e.g. endocytosis [19] and signal transduction [20-22]˙ MBDs have been mainly studied in animals, and in particular mammals, and so, it is expected that evolutionary distant organisms will (or at list seem to) have less MBPs than those mainly studied with experimental and computational methods.

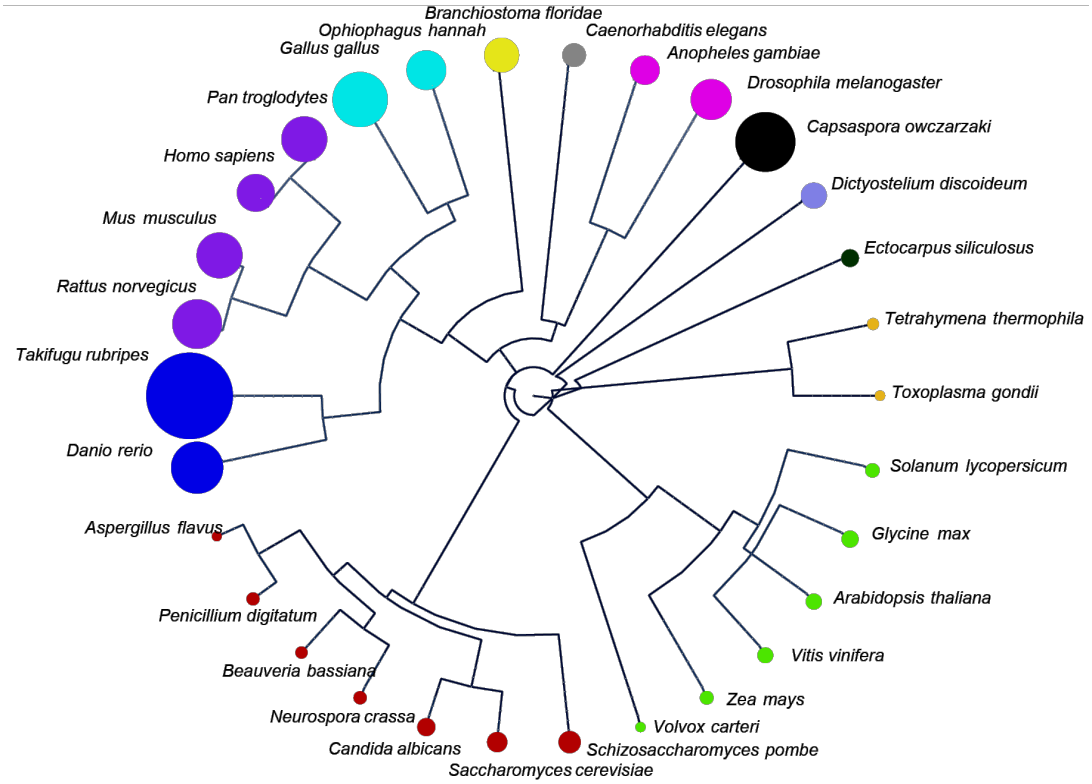

Figure 4. Cladogram for 30 selected proteomes. Each circle in the cladogram represents the percentage of MBPs in a proteome (from 0.3% in *Toxoplasma gondii* up to 5.7% in *Takifugu rubripes*). Each clade in the cladogram is color coded, where green represents plants, dark red represents fungi, dark blue is for fish, purple for mammals, light blue represents reptiles (including birds), yellow, grey and magenta other animals, orange amoebozoa and black, light purple and dark green represents organisms in other eukaryotic kingdoms. The cladogram was visualized with Cytoscape [23].

The classification of MBPs from the 30 eukaryotic proteomes in which MBPpred was applied showed that in peripheral membrane and transmembrane ca. 80% of MBPs are peripheral, while almost 20% are transmembrane (Table S6). The percentage of MBPs is similar within the different groups of eukaryotic proteomes. Interestingly, it was observed that almost 90% of all MBPs found in these 30

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proteomes are proteins of unknown function or subcellular location. This was also observed during the functional annotation of the proteomes, where the majority of proteins could not be connected with any term. These results show that the application of MBPpred can greatly contribute to the understanding of proteins with unknown function, and, additionally help in the annotation of current and newly sequenced proteomes.

MBPpred was also applied in all eukaryotic reference proteomes from UniProtKB (release: 2015_04) with similar results, which are included in Table S3.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. CONCLUSIONS

MBPpred is a relatively fast and accurate method, which can detect Membrane Binding Proteins from their sequence alone and is therefore applicable to entire proteomes. Our method is the first to include an extended list of MBDs, compiled after an extensive literature search, for the detection of MBPs. Moreover, MBPpred can distinguish between peripheral and transmembrane MBPs and thus can identify peripheral membrane proteins, a group of proteins extremely challenging to predict and study from sequence alone [9]. In addition, MBPpred is currently the only publicly available method which can detect MBPs.

Even though experimental studies have shown that the overwhelming majority of proteins with Membrane Binding Domains have the ability to bind to phosphoinositides or other membrane lipids [24], there have been reports of a small number of proteins with MBDs, that have lost their ability to bind to membranes during the course of evolution [25]. However, the lack of experimental information regarding these proteins does not allow their discrimination from MBPs with the same domains. Nevertheless, their identification is crucial for their further functional annotation.

Computational studies for membrane binding proteins in organisms other than mammals have not been performed to date, and information gathered from the application of MBPpred on novel proteomes, can be of great assistance towards their functional annotation. The use of MBPpred for the annotation of newly sequenced proteomes is very important, since it can provide novel candidates for biochemical and structural analysis. Lipidomic studies have shown that cell membranes contain over 1000 different lipids [26]. Several of these lipids act as targets for Membrane Binding Proteins, which are recruited during cell signaling and membrane trafficking to form various protein-protein and lipid-protein interactions [2, 27]. These interactions are vital for the conduction of other membrane protein functions, since other membrane proteins with which MBPs interact can act as receptors, transporters, enzymes, structural proteins and so on [28]. In addition, a previous study of the peripheral membrane protein interactome (peripherome) of the human plasma membrane [29] indicated the importance of MBPs, since these proteins can act as hubs and bottlenecks in the network, while they maintain connections with other membrane proteins in microdomains that are enriched in certain membrane lipids, called lipid rafts [30]. The application employed here in the 30 and 407 proteome datasets is the first large scale effort for the identification of MBPs and provides important information regarding the presence and types of MBPs in various eukaryotic proteomes.

As more experimental information about MBPs becomes available, more proteins with the ability to bind to membrane lipids non-covalently will be revealed in all eukaryotic kingdoms. Consequently, more detailed information about the mechanism these proteins use to bind lipids will be uncovered and thus we will be able to better comprehend the interactions of proteins in the membrane plane.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

We would like to thank the scientific and administrative staff of the "Bioinformatics" Master's Program at the Faculty of Biology of the University of Athens for generous support.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

[1] R.V. Stahelin, Lipid binding domains: more than simple lipid effectors, J. Lipid Res., 50 Suppl (2009) S299-304.

[2] W. Cho, R.V. Stahelin, Membrane-protein interactions in cell signaling and membrane trafficking, Annu. Rev. Biophys. Biomol. Struct., 34 (2005) 119-151.

[3] K. Moravcevic, C.L. Oxley, M.A. Lemmon, Conditional peripheral membrane proteins: facing up to limited specificity, Structure, 20 (2012) 15-27.

[4] I. Barak, K. Muchova, The role of lipid domains in bacterial cell processes, Int J Mol Sci, 14 (2013) 4050-4065.

[5] M.A. Lemmon, Membrane recognition by phospholipid-binding domains, Nat. Rev. Mol. Cell Biol., 9 (2008) 99-111.

[6] R.V. Stahelin, J.L. Scott, C.T. Frick, Cellular and molecular interactions of phosphoinositides and peripheral proteins, Chem Phys Lipids, 182 (2014) 3-18.

[7] J.D. Carpten, A.L. Faber, C. Horn, G.P. Donoho, S.L. Briggs, C.M. Robbins, G. Hostetter, S. Boguslawski, T.Y. Moses, S. Savage, M. Uhlik, A. Lin, J. Du, Y.W. Qian, D.J. Zeckner, G. Tucker-Kellogg, J. Touchman, K. Patel, S. Mousses, M. Bittner, R. Schevitz, M.H. Lai, K.L. Blanchard, J.E. Thomas, A transforming mutation in the pleckstrin homology domain of AKT1 in cancer, Nature, 448 (2007) 439-444.

[8] N. Bhardwaj, R.V. Stahelin, R.E. Langlois, W. Cho, H. Lu, Structural bioinformatics prediction of membrane-binding proteins, J. Mol. Biol., 359 (2006) 486-495.

[9] N. Bhardwaj, M. Gerstein, H. Lu, Genome-wide sequence-based prediction of peripheral proteins using a novel semi-supervised learning technique, BMC Bioinformatics, 11 Suppl 1 (2010) S6.

[10] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, Nucleic Acids Res., 42 (2014) D222-230.

[11] C. Pasquier, V.J. Promponas, S.J. Hamodrakas, PRED-CLASS: cascading neural networks for generalized protein classification and genome-wide applications, Proteins, 44 (2001) 361-369.

[12] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, C. Zardecki, The Protein Data Bank, Acta Crystallogr D Biol Crystallogr, 58 (2002) 899-907.

[13] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol., 215 (1990) 403-410.

[14] K.H. Brodersen, C.S. Ong, K.E. Stephan, J.M. Buhmann, The Balanced Accuracy and Its Posterior Distribution, in: Proceedings of the 2010 20th International Conference on Pattern Recognition, IEEE Computer Society, 2010, pp. 3121-3124.

[15] UniProt_Consortium, UniProt: a hub for protein information, Nucleic Acids Res., 43 (2015) D204-212.

[16] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat Protoc, 4 (2009) 44-57.

[17] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M.

bioRxiv preprint doi: https://doi.org/10.1101/034942; this version posted December 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet., 25 (2000) 25-29.

[18] E. Wallin, G. von Heijne, Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms, Protein Sci., 7 (1998) 1029-1038.

[19] M.J. Marcote, F. Gu, J. Gruenberg, F. Aniento, Membrane transport in the endocytic pathway: Animal versus plant cells, Protoplasma, 210 (2000) 123-132.

[20] R. Solano, J.R. Ecker, Ethylene gas: perception, signaling and response, Curr. Opin. Plant Biol., 1 (1998) 393-398.

[21] A.C. Harmon, Calcium-regulated protein kinases of plants, Gravit Space Biol Bull, 16 (2003) 83-90.

[22] T. Nagata, S. Iizumi, K. Satoh, H. Ooka, J. Kawai, P. Carninci, Y. Hayashizaki, Y. Otomo, K. Murakami, K. Matsubara, S. Kikuchi, Comparative analysis of plant and animal calcium signal transduction element using plant full-length cDNA data, Mol. Biol. Evol., 21 (2004) 1855-1870.

[23] G. Su, J.H. Morris, B. Demchak, G.D. Bader, Biological network exploration with cytoscape 3, Curr Protoc Bioinformatics, 47 (2014) 8 13 11-18 13 24.

[24] M.A. Lemmon, K.M. Ferguson, Signal-dependent membrane targeting by pleckstrin homology (PH) domains, Biochem. J., 350 Pt 1 (2000) 1-18.

[25] S.M. Singh, D. Murray, Molecular modeling of the membrane targeting of phospholipase C pleckstrin homology domains, Protein Sci., 12 (2003) 1934-1953.

[26] G. van Meer, D.R. Voelker, G.W. Feigenson, Membrane lipids: where they are and how they behave, Nat. Rev. Mol. Cell Biol., 9 (2008) 112-124.

[27] T.G. Kutateladze, Translation of the phosphoinositide code by PI effectors, Nat. Chem. Biol., 6 (2010) 507-513.

[28] G. von Heijne, Recent advances in the understanding of membrane protein assembly and structure, Q. Rev. Biophys., 32 (1999) 285-307.

[29] K.C. Nastou, G.N. Tsaousis, K.E. Kremizas, Z.I. Litou, S.J. Hamodrakas, The human plasma membrane peripherome: visualization and analysis of interactions, Biomed Res Int, 2014 (2014) 397145.

[30] B.P. Head, H.H. Patel, P.A. Insel, Interaction of membrane/lipid rafts with the cytoskeleton: impact on signaling and function: membrane/lipid rafts, mediators of cytoskeletal arrangement and cell signaling, Biochim. Biophys. Acta, 1838 (2014) 532-545.