Biochimica et Biophysica Acta 1838 (2014) 2066–2070

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Identification of novel sphingolipid-binding motifs in mammalian membrane proteins

Patrik Björkholm [a,b,1], Andreas M. Ernst [c,1], Moritz Hacke [c], Felix Wieland [c], Britta Brügger [c,*], Gunnar von Heijne [a,b,**]

[a] Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden
[b] Science for Life Laboratory, Stockholm University, SE-171 21 Solna, Sweden
[c] Heidelberg University Biochemistry Center, Im Neuenheimer Feld 328, 69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 15 March 2014
Received in revised form 18 April 2014
Accepted 24 April 2014
Available online 2 May 2014

Keywords:
Sphingolipid
Transmembrane helix
Membrane protein

ABSTRACT

Specific interactions between transmembrane proteins and sphingolipids is a poorly understood phenomenon, and only a couple of instances have been identified. The best characterized example is the sphingolipid-binding motif VXXTLXXIY found in the transmembrane helix of the vesicular transport protein p24. Here, we have used a simple motif-probability algorithm (MOPRO) to identify proteins that contain putative sphingolipid-binding motifs in a dataset comprising proteomes from mammalian organisms. From these motif-containing candidate proteins, four with different numbers of transmembrane helices were selected for experimental study: i) major histocompatibility complex II Q alpha chain subtype (DQA1), ii) GPI-attachment protein 1 (GAA1), iii) tetraspanin-7 TSN7, and iv), metabotropic glutamate receptor 2 (GRM2). These candidates were subjected to photo-affinity labeling using radiolabeled sphingolipids, confirming all four candidate proteins as sphingolipid-binding proteins. The sphingolipid-binding motifs are enriched in the 7TM family of G-protein coupled receptors, predominantly in transmembrane helix 6. The ability of the motif-containing candidate proteins to bind sphingolipids with high specificity opens new perspectives on their respective regulation and function.

1. Introduction

Motifs are linear amino acid sequence patterns that correspond to a specific structural or functional site in proteins. Motifs can indicate a common structure or functionality between proteins that do not share a high sequence similarity. Most motifs have been found using various forms of multiple sequence alignments [1,7,9]. However, there are only few databases, such as PROSITE and PRINT, that contain motifs, and few resources exist that enable the identification of novel motifs [1,14]. This is particularly the case for motifs located in transmembrane helices (TMHs) of membrane proteins.

Recently, we identified a sphingolipid-binding motif in the p24 protein, a type I membrane protein involved in vesicular transport in the early secretory pathway [3,12]. p24 specifically interacts with a single molecular species of sphingomyelin (SM), SM 18:0, and the specificity of the interaction was shown to be mediated by residues residing in the TMH. Residues critical for the interaction form a crevice at one end of the TMH, lined by rigid, β-branched amino acids and an aromatic residue located next to the crevice.

This opened the possibility that p24-like sphingolipid-binding motifs might allosterically regulate membrane protein function. Regulation of membrane proteins by distinct molecular sphingolipid species would contribute to the understanding of why such complexity in the lipidome is maintained and needed. Since our previous analysis was limited to single-spanning membrane proteins, we went on to explore the entire mammalian membrane proteome for the presence of p24-like sphingolipid-binding motifs. This demanded an algorithm capable of evaluating complex motifs in a large dataset.

Here, we describe a motif-finding algorithm that allowed us to test if this sphingolipid-binding motif was a unique occurrence within the p24 family, or rather represents a more widespread feature among transmembrane proteins. We first describe the new MOtif PRObability analysis tool (MOPRO; a downloadable version is included in the on-line Supplementary). MOPRO is similar to the TMSTAT method [13], but it allows more complex motifs to be analyzed within larger databases and uses the more reliable *z*-score rather than *p*-values. Using MOPRO, we identify 28 putative sphingolipid-binding motifs in a dataset comprising all predicted TMHs in membrane proteins found in entire proteomes from mammalian organisms; one additional sphingolipid-binding motif was identified in our earlier study. We find instances of these 28 motifs in 672 novel candidate sphingolipid-binding proteins. Among the new candidate proteins, four selected proteins tested positive for sphingolipid binding in a cellular context [4]. Notably, a high

* Corresponding author.
** Correspondence to: G. von Heijne, Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden.
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bbamem.2014.04.026
0005-2736/

number of the new candidate proteins are G protein-coupled receptors (GPCRs).

2. Results and discussion

2.1. Motif probability analysis allows proteome-wide identification of sequence motifs

Based on the biophysical properties of the residues that contribute to the sphingolipid-binding motif present in p24, a 'relaxed' motif was generated by allowing different permutations of the β-branched amino acids Val, Thr, Ile and the non-β-branched Leu in the lipid-binding crevice, and the aromatic residues Phe, Tyr or Trp in the interfacial position: [V/I/T/L]XX[V/I/T/L][V/I/T/L]XX[V/I/T/L][F/W/Y]. This relaxed motif corresponds to 768 unique motifs ($4 \times 4 \times 4 \times 4 \times 3$), generating an unmanageable amount of candidate proteins if used to screen a large protein database.

This problem was overcome by using the MOPRO algorithm to remove motifs that do not generate a statistically significant over-representation of hits in the database being screened. MOPRO compares the observed number of occurrences of a motif in the database with the expected number of occurrences of the same motif in a randomized data set. The randomized data set is generated by randomizing each TMH in the original data set by repeatedly swapping positions of randomly chosen amino acids. Hence, the random data set contains an identical number of TMHs with identical amino acid compositions, but with randomized sequences. Comparison of the frequency of a motif in the randomized and original data sets makes it possible to assess the statistical significance of the motif using either a *p*-value or *z*-score, and to identify motifs that are significantly over-represented in the original data set.

To validate MOPRO, we identified over-represented pairs of amino acids of varying sequence separation in the mammalian proteome data set, and compared the results to those presented in the original TMSTAT publication [13]. Out of the top-twenty over-represented motifs found by TMSTAT, only 4 were not found as significant hits with MOPRO, Table S1. Considering that the sequence data sets used are very different and were collected more than 10 years apart, there is a satisfying concordance between the MOPRO and TMSTAT results.

Applying MOPRO to search for instances of over-represented sphingolipid-binding motifs in multi-spanning proteins within a homology-reduced mammalian dataset, we identified 28 (26 novel motifs) putative sphingolipid-binding motifs that conform to the 'relaxed' motif defined above (Table 1). The occurrence of motifs in the original and randomized datasets were then used to derive *z*-scores, with motifs considered to be significantly over-represented if the *z*-score > 3.5. These motifs were then used to mine for candidates in a non-homology-reduced mammalian dataset.

A total of 672 novel candidate proteins were found in this way. We removed 57 redundant proteins among these protein candidates, leaving 615 unique candidate proteins (Supplementary Table I). The novel candidates predominantly localize to the plasma membrane, Fig. 1. This is consistent with our earlier results obtained for single-spanning membrane proteins, and correlates with the fact that the plasma membrane is highly enriched in sphingolipids [3].

The original motif (VXXTLXXIY) found in single-spanning membrane proteins [3] was not present above random expectation in the set of multi-spanning proteins, but we nevertheless added three multi-spanning membrane proteins that contain the original motif to our collection of candidates (Supplementary Table I) and tested one of them for sphingolipid binding (see below).

Notably, the number of candidate proteins that belong to the GPCR superfamily is four times higher than expected based on the number of GPCRs in the screening data set, Fig. 2A. Within the pool of these GPCR candidates, the motifs are over-represented in TMH 6, common in TM1, and under-represented in TMH 3 and TMH 7, Fig. 2B. The canonical model for activation of GPCRs involves a critical outward movement of TMH 6, needed to open up the G-protein binding crevice [16,18]. It is tempting to speculate that sphingolipid binding to TMH 6 may be involved in fine-tuning these conformational changes [11].

The under-representation of the motif in TMH 3 is consistent with the fact that TMH 3 forms the functional core of GPCRs and is located at the center of the protein, largely shielded from contact with membrane lipids [16].

Table 1
Putative sphingolipid-binding motifs identified by MOPRO. The last motif was identified in Ref. [3] and is included for completeness. The "Motif" column shows the motif analyzed. The two "*z*-value" columns contain results from two different runs of MOPRO, and the "<*z*-value>" column gives the average of these two runs. The "Tested candidate protein" column shows the four candidate proteins for which for sphingolipid-binding was tested experimentally in vivo.

Motif	*z*-value 1	*z*-value 2	<*z*-value>	Tested candidate protein
LXXILXXLF	8.09	8.02	8.06	
LXXLLXXTW	8.08	7.96	8.02	
LXXLLXXLY	5.45	5.44	5.45	
VXXVVXXLF	5.40	5.39	5.40	DQA1(TMH1)
LXXLLXXLF	5.13	5.17	5.15	
IXXVVXXIW	5.09	5.18	5.14	
TXXLIXXLF	5.00	4.98	4.99	
IXXLLXXLF	4.86	4.85	4.86	
TXXTIXXLF	4.78	4.83	4.81	
TXXVVXXLF	4.81	4.78	4.79	TSN7(TMH2)
VXXVTXXIF	4.76	4.77	4.77	
VXXVVXXVF	4.75	4.76	4.75	
LXXLLXXVF	4.39	4.39	4.39	
IXXLVXXVY	4.35	4.38	4.36	
VXXVTXXVF	4.32	4.34	4.33	
TXXVTXXVY	4.15	4.09	4.12	
IXXILXXIF	4.04	4.07	4.06	
LXXLLXXIF	4.04	4.07	4.06	
TXXLIXXVF	3.86	3.89	3.88	
IXXTLXXLW	3.84	3.79	3.81	
TXXTTXXIW	3.74	3.86	3.80	GRM2(TMH6)
VXXLIXXVW	3.73	3.75	3.74	
LXXLVXXIF	3.67	3.71	3.69	
LXXLLXXTY	3.69	3.66	3.67	
VXXLLXXIF	3.64	3.64	3.64	
LXXILXXVF	3.63	3.63	3.63	
TXXVIXXIF	3.51	3.52	3.51	
VXXTLXXIY [3]	0.97	1.00	0.99	GAA1(TMH4)

2.2. Experimental validation of candidate sphingolipid-binding proteins

As a first experimental test of the list of candidate sphingolipid-binding proteins, four candidate proteins containing different numbers of TMHs — one single-spanning transmembrane protein (the major

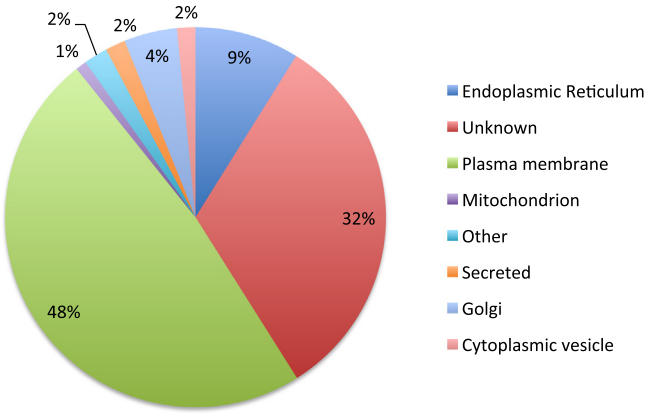

Fig. 1. Subcellular localization of candidate sphingolipid-binding proteins as annotated in UniProt.