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.

© 2014 Elsevier B.V. All rights reserved.

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/© 2014 Elsevier B.V. All rights reserved.

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

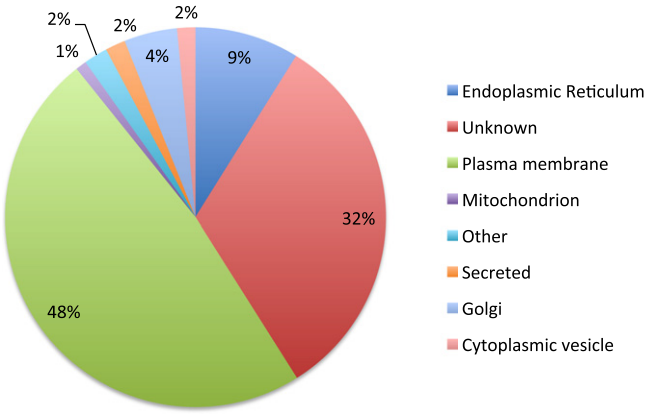

Fig. 1. Subcellular localization of candidate sphingolipid-binding proteins as annotated in UniProt.

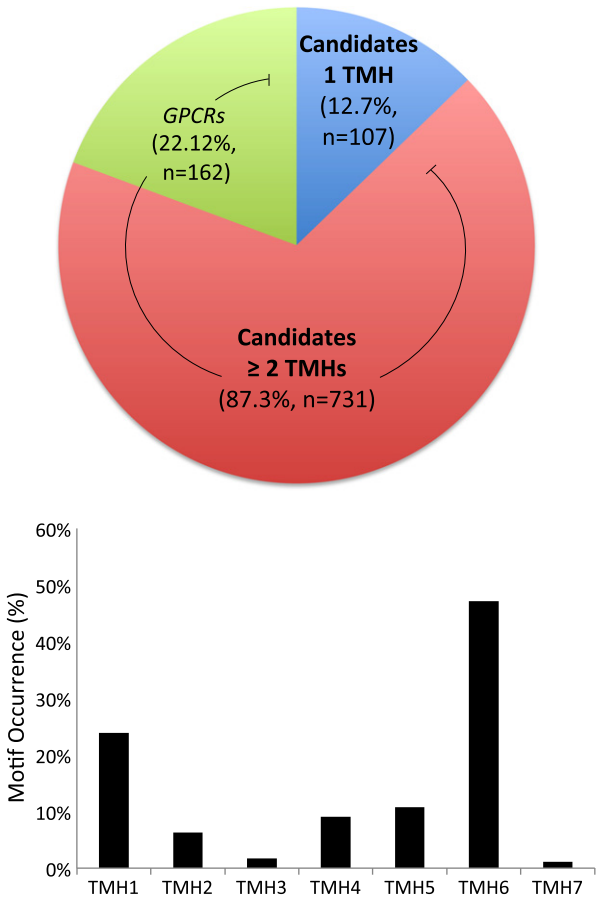

Fig. 2. (A) Percentages of candidate sphingolipid-biding proteins that are single-spanning, multi-spanning, and GPCRs. (B) Distribution of the occurrence of putative sphingolipid-binding motifs across the 7 TMHs of candidate sphingolipid-biding proteins in the GPCR superfamily.

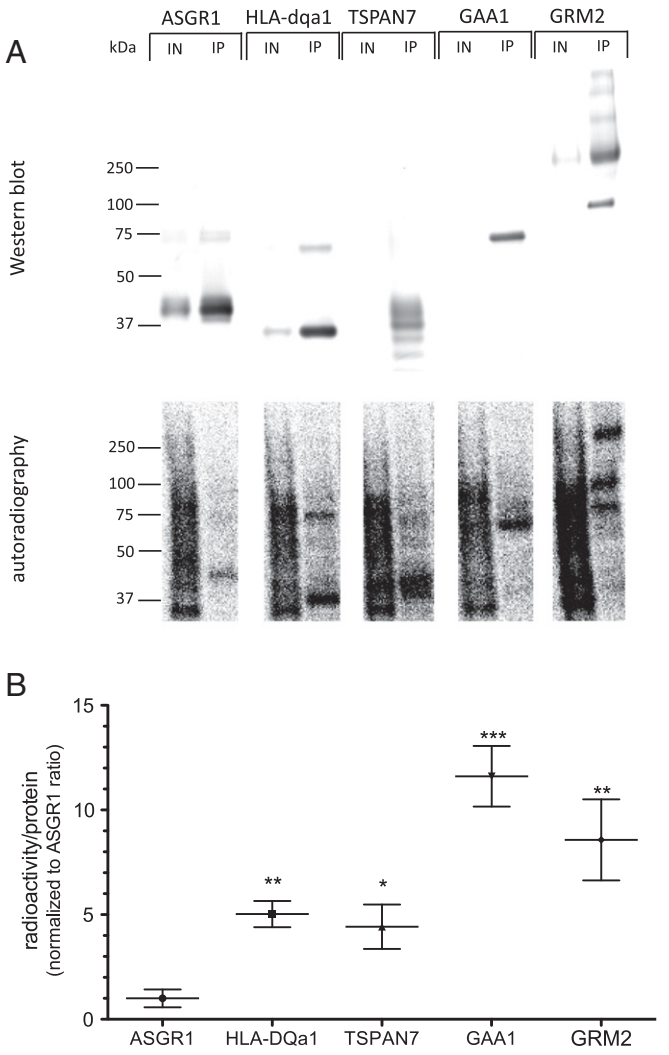

Fig. 3. Candidate sphingolipid-biding proteins significantly bind sphingolipids in vivo. (A) FLAG-tagged ASGR1, DQA1, TSPAN7, GAA1, and GRM2 were transiently expressed in HeLa cells for 18 h, and then incubated for additional 6 h with [^{3}H]-photosphingosine. The cells were then UV-irradiated, lysed, and analyzed by Western blot (upper panel) and digital autoradiography (lower panel). The order of the samples is the same in the upper and lower panels. (B) Radioactivity/protein was normalized to the non-candidate protein ASGR1, and the observed ratios from four independent experiments were statistically compared by two-tailed, unpaired t-tests. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

histocompatibility complex II Q α-chain subtype DQA1) and three multi-spanning proteins: GPI-attachment protein 1 (hGAA1, which contains the motif identified in p24), tetraspanin-7 (TSN7) and a GPCR, the metabotropic glutamate receptor 2 (GRM2) — were transiently expressed as FLAG-tagged fusion proteins in HeLa cells and probed for interaction with sphingolipids. Asialoglycoprotein receptor 1 (ASGR1) was used as a negative control, as it does not contain a putative sphingolipid-binding motif, yet localizes to the plasma membrane facing the bulk pool of cellular sphingolipids (Supplementary Fig. 1) [8]. Immunofluorescence analysis of cells expressing the FLAG-tagged candidate proteins showed that hGAA1 localized to the ER and DQA1 to the trans Golgi network, while TSN7 and GRM2 were found at the plasma membrane (Supplementary Fig. 1).

To probe for sphingolipid binding, cells were fed a tritiated and photoactivatable sphingolipid precursor, followed by UV irradiation, immunoprecipitation, Western blotting, autoradiography, and finally, normalization of labeling/protein to a non-motif containing plasma membrane resident protein (ASGR1) [5]. Strikingly, all four candidates exhibited a 5–12 fold higher extent of labeling over that observed for ASGR1, Fig. 3. The specific labeling of all four candidate proteins suggests that many more of the candidate proteins identified by MOPRO may bind sphingolipid.

In summary, our results suggest that p24-like sphingolipid-binding motifs are not restricted to single-spanning membrane proteins but are present also among multi-spanning membrane proteins. We find that a large number of GPCRs contain putative sphingolipid-binding motifs, with a remarkable overrepresentation of these motifs in TMH6, possibly implying a common regulation of these receptors by sphingolipids. Identification of DQA1 as sphingolipid-binding protein suggests the possibility of a novel kind of regulation of MHC class II molecule assembly and maturation. GPI-attachment protein 1 (GAA1), as an ER resident protein, may require sphingolipids as a co-factor for its enzymatic function. The specific interaction might additionally or alternatively direct the transamidase complex to subdomains that are enriched in sphingolipids and its substrate, the immature GPI anchor. Tetraspanins are well-described signaling-hubs that reside in glycosphingolipid-enriched membranes. Although ectopic domains were found responsible for carbohydrate-recognition of glycosphingolipids by these proteins [19], additional species specificity resulting from recognition within the membrane might alter subdomain targeting and function of these proteins that regulate adhesion molecule organization and signaling.

In conclusion, the novel putative sphingolipid-binding motifs identified by MOPRO serve as a starting point for more specialized functional studies of individual candidates. The MOPRO approach can be used in

very large datasets to evaluate complex motifs in TMHs that may mediate specific protein–lipid or protein–protein interactions.

3. Methods

3.1. Data sets

The mammalian data set was downloaded from the Uniprot resource webpage (www.uniprot.org) [15]. The query used to download the data set was "taxonomy:"Mammalia [40674]" AND "annotation:(type: transmem)". This query returns all annotated transmembrane proteins (known and predicted) of mammalian origin (18,116 sequences from 1698 mammalian species). The data set was homology reduced down to 30% identity using CD-HIT [10,20]. After homology reduction 2760 protein sequences from 50 mammalian species remained.

In the homology-reduced set, 10,506 TMHs were predicted by SCAMPI [2,6]. Since SCAMPI does not predict the precise ends of membrane-spanning segments with high precision, all sequences were extended by three residues in both the N- and C-terminal direction to ensure that the whole transmembrane domain is captured. For evaluation of motifs within multi-spanning transmembrane proteins, all entries containing one or two predicted TMHs were discarded (1187 proteins containing 8355 TMHs). In our earlier work [3] the opposite approach was used for evaluating motifs in single-spanning membrane proteins, i.e., all entries containing more than two predicted TMHs were discarded.

3.2. MOPRO

MOPRO processes a list of motifs to be evaluated over a homology-reduced sequence data set. The first step is to count the number of occurrences N_i of the given motif i in the non-randomized TMH data set. In the next step, the sequence of each TMH in the data set is randomized. The randomization is done using a swap algorithm, i.e., the positions of two randomly selected residues in each TMH are swapped. The number of swaps is set equal to the number of amino acids within each TMH. This ensures that on average each amino acid has swapped positions two times. The number of occurrences of motif i in the randomized TMH data set, R_i, is then counted. Finally, the randomization procedure is repeated 10,000 times to obtain the distribution of R_i values expected for motif i over a TMH data set with randomized sequences, and the p-value for motif i (i.e., the probability to observe $\geq N_i$ instances of the motif) is calculated from this distribution. This distribution also allows us to calculate the z-score, which is obtained by subtracting the average random occurrence of R_i from the true occurrence of the motif and then dividing this with standard deviation of R_i.

3.3. TMSTAT comparison

In the original TMSTAT publication [13], amino acid pair frequencies were calculated for all 4000 XYn pairs (X, Y run over all 20 natural amino acids and $n = 1$–10 is the separation between X and Y in the sequence), and compared to a random expectation. The data set used consisted of 13,606 non-homologous 18-residue long segments annotated as TMHs in the Swiss-Prot database. As a comparison, MOPRO statistics for the same 4000 XYn pairs were calculated for our mammalian data set. To make the comparison as fair as possible, only the 19 central residues in each TMH in the mammalian data set were included in the MOPRO run.

3.4. Candidate sphingolipid-binding proteins

The motifs that were found to be over-represented were used to screen for candidate sphingolipid-binding proteins. Candidates were mined from a 90% homology reduced mammalian data set, containing 9981 proteins. The homology reduction was done using cd-hit [10,20]. Both the previously identified motif and the 28 novel motifs were used. The TMHs were defined using SCAMPI, with the segments extended by three residues upstream and downstream of the predicted TMH.

3.5. The GPCR data set

All proteins that were annotated and reviewed as GPCRs were downloaded from Uniprot. The TMHs of these proteins were predicted using SCAMPI. The data set contains 3175 proteins. This set was homology-reduced to 90% sequence identity, which resulted in 2008 remaining sequences. The homology reduction was done using cd-hit [10,20].

3.6. Cell culture

HeLa cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS), 100 U/ml penicillin, 100 μg/ml streptomycin and 2 mM L-glutamine at 5% CO_2 and 37 °C.

3.7. In vivo photoaffinity labeling of HeLa cells transiently expressing selected candidate proteins

Four sphingolipid-binding motif-containing protein candidates and a control plasma membrane protein (ASGR1) not containing any of the 23 motifs (Table 1) were obtained as C-terminal FLAG-tag fusions from Origene. The candidate proteins were i) human HLA class II histocompatibility antigen, DQ alpha 1 chain, (DQA1), ii) human tetraspanin-7 (TSN7), iii) human glycosylphosphatidylinositol anchor attachment protein 1 (hGAA1), and iv) human metabotropic glutamate receptor 2 (GRM2). Photolabelling experiments were performed as described [3,5,17]. In brief, HeLa cells were grown in 6-well dishes and transfected using FuGENE HD as a transfection reagent (1.5 μg DNA + 4.5 μl transfection reagent in 100 μl Opti-MEM (Invitrogen)) at approximately 70% confluency. After 18 h of expression, cells were labeled with 12 μCi of the [^{3}H]-photo-sphingosine precursor in 3 ml of DMEM supplemented with 10% delipidated FCS for 6 h. Finally, the cells were UV-irradiated for 10 min on ice, and lysed for 1 h in 100 μl lysis buffer (50 mM HEPES-NaOH, pH 7.4, 100 mM NaCl, 5 mM EDTA, 1% Triton X-100 (v/v), 0.5% deoxycholate (w/v), and protease inhibitor cocktail). Post-nuclear supernatants were then subjected to immunoprecipitation using M2-(anti-FLAG)-affinity gel (Sigma-Aldrich, St Louis, USA) according to the manufacturer's instructions. Proteins were eluted in 4× SDS-PAGE sample buffer and subjected to SDS-PAGE (10–20% Tris/Tricine gradient gels, Invitrogen), Western blotting, and quantitative immunodetection of FLAG-tagged proteins using a Li-COR infrared imager. Radioactively labeled proteins were detected by digital autoradiography (beta-Imager 2000, Biospace) with acquisition times of 16 h. Each construct listed was probed in 4 independent experiments. In each experiment, the ratios of radioactivity/protein, resulting from quantitation of radioactivity and Western blot signal (capitalizing on the fact that all recombinant proteins share the same epitope), were determined for each of the transiently expressed proteins. The ratios were then normalized to the ratio observed for ASGR1, the control protein, and the independent experiments were averaged. Labeling ratios of candidates were then statistically compared by two-tailed, unpaired t-test with the average labeling ratio observed for ASGR1.

3.8. Subcellular localization of sphingolipid-binding motif-containing candidates in HeLa cells

HeLa cells were plated on 35 mm glass bottom dishes (MatTak Corporation, Ashland, USA) 24 h prior to transfection. 3.3 μg of the respective plasmids containing the FLAG-tagged cDNAs encoding for candidate proteins, were mixed with 10 μl of Fugene HD transfection reagent (Promega, Madison, USA) and added to the subconfluent cells

(following the protocol provided by Origene, Rockville, USA). After 24 h, the cells were washed three times with PBS (pH 7.4) and fixed with 3.6% PFA (Sigma Aldrich, St. Louis, USA) for 20 min on ice. After washing (3 times) and blocking for 1 h in 1% BSA in PBS, the cells were either directly incubated with Alexa Fluor 647-conjugated Wheat Germ Agglutinin (Life Technologies, Carlsbad, USA) at 5 mg/ml for 20 min at RT, or, for endomembrane stainings, permeabilized with 0.5% (w/v) Triton-X100 for 5 min at RT, followed by three washing-steps of 3 min each. Cells were then incubated with polyclonal sheep anti-TGN46 (TGOLN2, Serotec, 1:200), polyclonal rabbit anti-calnexin (Abcam, 1:200), or monoclonal mouse anti-FLAG-tag sequence DYDDDK (Sigma, 1:200), respectively, for 1 h at RT (Supplementary Fig. 1). After three washing steps with PBS of 3 min each, either donkey anti-sheep, Alexa 546, goat anti-rabbit, Alexa 546, goat anti-mouse, Alexa 488 were used as secondary antibodies (all from Invitrogen) at a 1:1000 dilution.

Cells were washed again and the samples were subjected to microscopy using a Zeiss LSM 510 confocal unit mounted on an Axiovert 200 inverted microscope, equipped with an argon laser beam. Image processing was performed using Zen 2011 and Image J software.

Acknowledgments

This work was supported by grants from the Swedish Cancer Foundation, the Swedish Research Council, the Swedish Foundation for Strategic Research, and the European Research Council (ERC-2008-AdG 232648) to GvH and from the German Research Foundation (DFG, TRR83) to BB and FW. AME was supported by the Peter and Traudl Engelhorn foundation. FW and BB are investigators of the CellNetworks Cluster of Excellence (EXC81).

Author contributions

PB, AME, FW, GvH, and BB conceived the work and wrote the paper. PB did the bioinformatics analysis and AME did the experiments together with MH.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbamem.2014.04.026.

References

[1] T.K. Attwood, A. Coletta, G. Muirhead, A. Pavlopoulou, P.B. Philippou, I. Popov, C. Roma-Mateo, A. Theodosiou, A. Mitchell, The PRINTS database: a fine-grained protein sequence annotation and analysis resource — its status in 2012, Database (2012), http://dx.doi.org/10.1093/database/base019.
[2] A. Bernsel, H. Viklund, J. Falk, E. Lindahl, G. von Heijne, A. Elofsson, Prediction of membrane-protein topology from first principles, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 7177–7181.
[3] F.X. Contreras, A.M. Ernst, P. Haberkant, P. Björkholm, E. Lindahl, B. Gönen, C. Tischer, A. Elofsson, G. von Heijne, C. Thiele, R. Pepperkok, F. Wieland, B. Brügger, Molecular recognition of a single sphingolipid species by a protein's transmembrane domain, Nature 481 (2012) 525–529.
[4] U. Coskun, K. Simons, Membranes: the lipid perspective, Struct. Cell 19 (2011) 1543–1548.
[5] P. Haberkant, O. Schmitt, F.X. Contreras, C. Thiele, K. Hanada, H. Sprong, C. Reinhard, F.T. Wieland, B. Brügger, Protein-sphingolipid interactions within cellular membranes, J. Lipid Res. 49 (2008) 251–262.
[6] T. Hessa, N. Meindl-Beinker, A. Bernsel, J. Kim, Y. Sato, M. Lerch, C. Lundin, I. Nilsson, S.H. White, G. von Heijne, Molecular code for transmembrane-helix recognition by the Sec61 translocon, Nature 450 (2007) 1026–1030.
[7] G.Z. Hertz, G.D. Stormo, Identifying DNA and protein patterns with statistically significant alignments of multiple sequences, Bioinformatics 15 (1990) 563–577.
[8] V. Katritch, V. Cherezov, R.C. Stevens, Structure–function of the G protein-coupled receptor superfamily, Annu. Rev. Pharmacol. Toxicol. 53 (2013) 531–556.
[9] C.E. Lawrence, S.F. Altschul, M.S. Boguski, J.S. Liu, A.F. Neuwald, J.C. Wootton, Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment, Science 262 (1993) 208–214.
[10] F. Limin, N. Beifang, Z. Zhengwei, W. Sitao, L. Weizhong, CD-HIT: accelerated for clustering the next generation sequencing data, Bioinformatics 28 (2012) 3150–3152.
[11] H.A. Overton, et al., Deorphanization of a G protein-coupled receptor for oleoylethanolamide and its use in the discovery of small-molecule hypophagic agents, Cell Metab. 3 (2006) 167–175.
[12] V. Popoff, F. Adolf, B. Brugger, F. Wieland, COPI budding within the Golgi stack, Cold Spring Harb. Perspect. Biol. 3 (2011) a005231.
[13] A. Senes, M. Gerstein, M. Donald, J. Engelman, Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with b-branched residues at neighboring positions, Mol. Biol. 296 (2000) 921–936.
[14] C.J.A. Sigrist, E. de Castro, L. Cerutti, B.A. Cuche, N. Hulo, A. Bridge, L. Bougueleret, I. Xenarios, New and continuing developments at PROSITE, Nucleic Acids Res. (2012), http://dx.doi.org/10.1093/nar/gks1067.
[15] The UniProt Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Res. 41 (2013) D43–D47.
[16] A.J. Venkatakrishnan, X. Deupi, G. Lebon, C.G. Tate, G.F. Schertler, M.M. Babu, Molecular signatures of G-protein-coupled receptors, Nature 494 (2013) 185–194.
[17] L. Wang, L. Liu, Y. Che, L. Wang, L. Jiang, C. Dong, Y. Zhang, Q. Li, Egress of HSV-1 capsid requires the interaction of VP26 and a cellular tetraspanin membrane protein, Virol. J. 7 (2010) 156.
[18] M. Yanagawa, T. Yamashita, Y. Shichida, Glutamate acts as a partial inverse agonist to metabotropic glutamate receptor with a single amino acid mutation in the transmembrane domain, J. Biol. Chem. 288 (2013) 9593–9601.
[19] M. Yanez-Mo, O. Barreiro, M. Gordon-Alonso, M. Sala-Valdés, F. Sanchez-Madrid, Tetraspanin-enriched microdomains: a functional unit in cell plasma membranes, Trends Cell Biol. 19 (2009) 434–446.
[20] H. Ying, N. Beifang, G. Ying, F. Limin, L. Weizhong, CD-HIT Suite: a web server for clustering and comparing biological sequences, Bioinformatics 26 (2010) 680.