Identification and *in silico* analysis of helical lipid binding regions in proteins belonging to the amphitropic protein family

ROB CA KELLER

Section Chemistry, Charlemagne College, Wilhelminastraat 13-15, Nijmegen 6524 AJ, The Netherlands

(Email, rcakeller@kpnmail.nl)

The role of protein–lipid interactions is increasingly recognized to be of importance in numerous biological processes. Bioinformatics is being increasingly used as a helpful tool in studying protein–lipid interactions. Especially recently developed approaches recognizing lipid binding regions in proteins can be implemented. In this study one of those bioinformatics approaches specialized in identifying lipid binding helical regions in proteins is expanded. The approach is explored further by features which can be easily obtained manually. Some interesting examples of members of the amphitropic protein family have been investigated in order to demonstrate the additional features of this bioinformatics approach. The results in this study seem to indicate interesting characteristics of amphitropic proteins and provide insight into the mechanistic functioning and overall understanding of this intriguing class of proteins. Additionally, the results demonstrate that the presented bioinformatics approach might be either an interesting starting point in protein–lipid interactions studies or a good tool for selecting new focus points for more detailed experimental research of proteins with known overall protein–lipid binding abilities.

[Keller RCA 2014 Identification and *in silico* analysis of helical lipid binding regions in proteins belonging to the amphitropic protein family. *J. Biosci.* **39** 771–783] **DOI** 10.1007/s12038-014-9479-z

1. Introduction

The role of lipids (Dowhan 1997; van Meer *et al.* 2008) and lipid–protein interactions has received considerable attention in life science through the years (Gennis 1977; Montecucco *et al.* 1988; Mouritsen and Bloom 1993; Cserhati 1993; Lee 2005; Smith 2012). Protein–lipid interactions have been studied in relation to numerous processes in biology such as lipid polymorphism (Epand 1998), membrane protein function and folding (Marsh *et al.* 2002; de Planque and Killian 2003; Hunte 2005; Dowhan and Bogdanov 2011), lipid domains (Tocanne *et al.* 1994; Marsh 1995; Fantini 2003) and lipid rafts (Lingwood *et al.* 2009; Simons and Sampaio 2011).

One interesting recent development is the bioinformatics approach which enables the identification of lipid binding helical regions in proteins (Gautier *et al.* 2008). A recent demonstration of this approach was shown for protein translocation motor proteins (Keller 2011a) where the prediction of multiple lipid binding regions indicated a possible general feature of this class of proteins. The bioinformatics approach utilizes Web-based software (Gautier *et al.* 2008) and includes a feature to identify (potential) lipid binding helix regions in proteins. Stretches of amino acids can be investigated and the use of a discrimination factor enables discrimination between lipid binding and non-lipid-binding regions of proteins and peptides. Additionally, based on this approach, the Heliquest-generated mean hydrophobicity

Keywords. Amphitropics; heliquest; hydrophobic moment plot; lipid binding regions; protein–lipid interactions

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/dec2014/supp/Keller.pdf*

Published online: 20 October 2014

(<H>) and hydrophobic moment (μH) was demonstrated to allow the development of a hydrophobic moment plot (Keller 2011b). This approach is essentially based on the 'classical' Eisenberg plot approach (Eisenberg *et al.* 1984) but now with Heliquest-generated data.

The combined use of the Heliquest lipid binding discrimination factor and the Heliquest-generated Eisenberg plot was successfully able to identify lipid binding regions in cytoplasmic and extracellular loops of protein translocation membrane proteins (Keller 2013). Furthermore, this approach indicated an interesting new feature of amphitropic proteins: multiple lipid regions which often seem to fall in the so-called globular protein region of the Eisenberg plot (Keller 2011b). Amphitropic proteins are defined as proteins with reversible membrane characteristics which are often related to the finding of populations of such a protein in both cytoplasmic as well as membrane-bound form; see elsewhere for an overview (Burn 1998; Johnson and Cornell 1999; Halskau et al. 2009).

This study explores the possibilities of bioinformatics further in an attempt to develop a total scan approach of all potential helical lipid binding regions in proteins based on their primary structure. For this purpose a substantial number of reported examples of proteins suggested to be members of the growing family of amphitropic proteins with demonstrated lipid binding were used. The relevance of the methodology and its findings was checked for a number of individual cases and the possible physiological role of lipid binding regions in the amphitropic protein family is discussed.

2. Materials and methods

2.1 *Primary and secondary structures identification*

The primary structure of the proteins was obtained either from the Swiss-Prot sequence database or the indicated references. The primary structures of those regions identified as lipid binding helix were collected and the included regions were checked for the extent of helicity either by crystallographic data if available and/or by secondary structure prediction using the program SOPMA (Combet *et al.* 2000) available at *http://npsa-pbil.ibcp.fr/*. Sequence alignment was performed using ClustalW (Larkin *et al.* 2007) available at *www.ebi.ac.uk/Tools/msa/Clustalw2/*.

2.2 *Determination lipid binding potential*

The mean hydrophobicity (<H>), the hydrophobic moment (μH) and the net charge (z) were calculated using the Heliquest software essentially as described before (Keller 2011a). In essence, for the analysis 18-residue windows were used and for each sequence under investigation the window with the highest discrimination factor was selected. The discrimination factor (D) is defined according to: D=0.944 (<μH>) + 0.33 (z). When this discrimination factor is above 0.68, the corresponding region can be considered to be a (potential) lipid-binding helix (Gautier *et al.* 2008). The Heliquest server and additional information are available at the website *http://heliquest.ipmc.cnrs.fr/*.

2.3 *Helical wheel plot*

The helical wheel representations were produced using the Heliquest software (Gautier *et al.* 2008). The obtained helical wheel plots were subsequently redrawn and customized.

2.4 *Eisenberg plot approach*

The Eisenberg plot approach was essentially performed as described in the original study (Keller 2011b). In short, both the mean hydrophobicity (<H>) and the hydrophobic moment (μH) were plotted using Heliquest-generated data. An 18-residue window is used unless stated otherwise and both the possible surface seeking and transmembrane (TM) helix properties were determined.

2.5 *Hydrophobic face and charge*

The possible presence of a hydrophobic face was extracted from the Heliquest data (Gautier *et al.* 2008) using the helical wheel plot. The net charge z and the possible presence of negative charged amino acids were automatically generated by the Heliquest program.

2.6 *Monte Carlo simulations*

Monte Carlo (MC) simulations of helical peptides in association with lipid membranes were performed using the MCPep server available at *http://bental.tau.ac.il/MCPep/* (Gofman *et al.* 2012). In essence, the program developed a model that performs Monte Carlo (MC) simulations of the interaction of helical peptides with lipid bilayers and attempts to discriminate between the configurations transmembrane (TM) and the surface orientation. A typical analysis included the input of the corresponding sequence in FASTA format, a membrane width of 30 Å, anionic lipid content as indicated in the results section and a RMSD cut-off for clustering of resultant conformations was 3 Å. The number of independent MC runs was set on 3 and number of MC cycles in each independent run was 500000. Molecular models were viewed and analyzed using Chimera (Pettersen *et al.* 2004), available at website *http://www.cgl.ucsf.edu/chimera/*. Final structure images were

customized using the free-available digital image software analysis program GIMP, available at the website *http://www.gimp.org*.

3. Results

3.1 *Illustrative results for some typical amphitropic proteins using Heliquest-generated data*

A number of amphitropic proteins as identified in the literature were investigated. For example, the complete sequence of α–Synuclein (AS) was run through the Heliquest program. Only the predominately helical regions were selected for further analysis. Based on the discrimination factor (D), which comprises both the hydrophobic moment (μH) and the net charge (z), a number of potential lipid-binding regions were identified (table 1). For a number of the identified regions, direct or indirect experimental or theoretical evidence already exists. It has been demonstrated experimentally that AS binds to lipids (Davidson *et al.* 1998) and it was proposed on theoretical grounds that several helices between AA 1-93 are the most likely candidates for this lipid binding (Davidson *et al.* 1998). These findings have been recently underlined by a study where labels were placed in one of the three AS domains: N-terminal, NAC, or C-terminal regions by labelling individual Ala-to-Cys mutants at positions 18, 90 and 140 and which concluded that amino acids on position 18 and 90 are in close contact to lipids (Shvadchak *et al.* 2011). The predictions of lipid binding regions AA 4–21 and AA 80–97 as summarized in table 1 are in full accordance with these experimental findings. Additionally, two lipid binding regions are predicted by the Heliquest approach, both in the earlier proposed AA 1–93 (Davidson *et al.* 1998).

Another example of an amphitropic protein is the adenylate cyclase toxin (CyaA), one of the virulence factors secreted by *Bordetella pertussis*, which is able to invade eukaryotic cells by translocation of its N-terminal catalytic domain directly across the plasma membrane of the target cells. It has been recently demonstrated that the region 375–485 is crucial for membrane insertion and translocation of the catalytic domain of CyaA (Karst *et al.* 2012). According to Heliquest-generated data three potential lipid binding regions could be predicted. As shown in table 1, one of these regions AA 457–474 is indeed located in the region 375–485. The possible role of the two novel lipid binding regions will be discussed later in this paper.

Additionally an example is selected on the basis of a recent study which introduced a useful system for measurement of the thermodynamics and kinetics of peptide insertion and folding across a lipid bilayer (Reshetnyak *et al.* 2008). This approach described the use of the so-called pH

Table 1. Data belonging to the lipid binding region search of typical examples of amphitropic proteins with experimentally demonstrated lipid binding

Sequence	H	μH	z	TM	LBR
Canary α–synuclein(Q91448):					
$_{4}$FMKGLSKAKEGVVAAAEK$_{21}$	0.173	0.442	2	-	Y
$_{21}$KTKQGVAEAAGKTKEGVL$_{38}$	0.007	0.258	2	-	Y
$_{43}$RTKEGVVHGVTTVAEKTK$_{60}$	0.061	0.287	2	-	Y
$_{80}$KTVEGAGNIAAATGLVKK$_{97}$	0.194	0.178	2	-	Y
B. pertussis CyaA toxin (P15318)					
$_{228}$YLARTRRAASEATGGLDR$_{235}$	0.035	0.396	2	-	Y
$_{238}$RIDLLWKIARAGARSAVG$_{255}$	0.382	0.315	3	-	Y
$_{457}$HWGQRALQGAQAVAAAQR$_{474}$	0.237	0.208	2	-	Y
pHLIP:					
$_{14}$WLFTTPLLLLDLALLVDA$_{31}$	1.066	0.123	-2	Y	N
Synechocystis HSP17(P72977):					
$_{10}$REMDNFQQQMNQLFEEVF$_{27}$	0.276	0.580	-3	-	N(S)
$_{89}$KDGVRRTEFRYGSFRRVI$_{106}$	0.086	0.345	4	-	Y[a]
E. coli enzyme IIAGlc (P69783)					
$_{1}$GLFDKLKSLVS$_{11}$	0.480	0.621	1	-	Y(S)[b]

The lipid binding region (LBR) is predicted by the Heliquest lipid binding discrimination factor (D). The presence of a possible surface seeking helix (S) or transmembrane helix (TM) is determined by the Heliquest-generated Eisenberg plot approach.

[a] According to SOPMA this region is slightly more than 50% helical, this region is therefore included for the sake of completeness.

[b] In this case a 11 AA-window was used, both SOPMA as well as reported data (Wang *et al.* 2003) indicated this region to be a short helix.

low-insertion peptide (pHLIP). This is an interesting member of a subclass of amphitropic proteins, since the subsequent insertion and folding across the bilayer of adsorbed peptide is a pH-driven process. Nicely in line with the experimentally proven lipid binding ability (Reshetnyak *et al.* 2008) the Heliquest-generated data predicted a possible helical lipid binding region. In this particular example the lipid binding discrimination factor is unable to identify a lipid binding region; however, the Heliquest-generated Eisenberg plot methodology (Keller 2011b) indicated a clear TM-helix type lipid binding region.

Another interesting example belongs to the protein family of heat shock proteins. For the amphitropic protein *Synechocystis* HSP17 it was concluded that it has not only an expected protein protective activity, but also an ability to stabilize lipid membranes (Török *et al.* 2001). This ability to bind to membranes is underlined by the finding that two potential lipid binding regions can be predicted, the AA 89–106 region by the Heliquest discrimination factor and the AA 10–27 region by the Eisenberg plot method using Heliquest-generated data, see table 1 for further details.

Finally in this part of the study, where well-described examples that can be found in the literature are selected, the signal transduction glucose-specific enzyme IIA (IIAGlc) is studied. It is another proposed amphitropic protein and structural studies demonstrated that the first seven residues including Lys 5 and Lys 7 of the N-terminal domain of IIAGlc are the most critical part of the membrane anchor (Wang *et al.* 2003). Indeed, the Heliquest lipid binding discrimination factor predicted a small lipid binding regions AA 1–11.

The described findings as summarized in table 1 seem to indicate that for members of the amphitropic protein family, similar to that previously shown (Keller 2011b) for other proteins, the combined Heliquest approach is able to predict one or more lipid binding regions in proteins with experimentally demonstrated lipid binding ability.

3.2 *Additional features of the Heliquest approach: Charge and hydrophobic face*

While looking for lipid binding regions in numerous examples of amphitropic proteins, a strong indication for the existence of additional features emerged. It was noted that while analysing some well-studied proteins and regions with demonstrated lipid binding activity that it seems as if the Heliquest approach was sometimes unable to find some of those lipid binding regions. In order to understand the possible reason for this, it is important to go into some details of the Heliquest approach itself.

The Heliquest program includes the conventional charges of the individual amino acids when it comes to calculating the net charge z. Clearly the aspartic acid and glutamic acid are known to be negatively charged. However, there are an increasing number of reports which seem to indicate that once such amino acids are positioned in the vicinity of the negatively charged head group region of acidic phospholipids they can become neutrally charged.

For example, with regard to the negatively charged glutamate, there is one study looking at the interfacial binding of bee venom secreted phospholipase A2 and predicted a significant upward shift in the pK_a of one (or more) of the glutamates when the enzyme binds to anionic phospholipid containing vesicles in such a way that the carboxylate of the glutamate side chain near the membrane surface undergoes protonation (Bollinger *et al.* 2004). Recent findings related to understanding the topology of membrane proteins seems to indicate an interesting role for phosphatidylethanolamine (PE) (Bogdanov *et al.* 2008). When a cytoplasmic domain contains a mixture of negative and positive amino acids, PE is proposed to suppress or neutralize the presence of negative residues (Bogdanov *et al.* 2008). In other words, the charge of amino acids while interacting in the context of a protein or peptide with phospholipids might differ significantly compared to the situation in solution. It has been hypothesized that for a tightly bound polypeptide adsorbed on the membrane, the pK values of the ionic residues are increased because of interaction with the negative charges at the membrane surface and as a consequence the pK shift would selectively neutralize aspartate and glutamate residues (Krishtalik and Cramer 1995). In addition, an interesting paper addressing a method to estimate the effective pKa of an arginine (Arg) side chain basically concluded that with additional stabilizations from negatively charged lipids, even a positively charged amino acid like Arg (once in the membrane) can adopt the protonated state (Yoo and Cui 2008).

With this in mind, the results of the analysis of a number of particular examples will be discussed here. In table 2 the results of the combined Heliquest approach are depicted for the amphitropic enzyme, CTP. All predicted lipid binding regions correspond nicely with the well-described M-domain which is believed to be important for lipid binding (see for example Johnson *et al.* 2003). One lipid binding regions (AA 273–290) is not identified as potential lipid binding by the Heliquest discrimination factor but is identified as a surface seeking region by the Heliquest based Eisenberg approach. This particular region is an interesting example of the above described matter of charge issues which are relevant in some particular cases. In a study addressing this issue it is proposed that particular glutamates in the protein become protonated in the low pH milieu at the surface of anionic membranes and it was suggested to be involved in the binding properties towards anionic phospholipids of the amphipathic helix in a phosphocholinecytidylyltransferase (CTP) (Johnson *et al.* 2003). The study gives, for example, detailed data about a peptide corresponding to region AA

256–288. The results are compared to a mutant analog where the interfacial glutamic acids 257, 268 and 279 were replaced with glutamines. It was demonstrated that the mutation results in a pH-independent binding to anionic phospholipid containing membranes, whereas the WT needs a lower pH. Interestingly a full sized Heliquest analysis (instead of the standard 18 AA) indicates a discrimination factor of 0.724 and −0,251 for the mutant and WT respectively. So the Heliquest approach predicts a lipid binding region for the mutant peptide and not for the (this) WT region. Considering the results as depicted in table 2, it is clear that a number of lipid binding regions are found, fully in agreement with what is know about this region. For region 240–257 a detailed helical wheel plot is depicted (see figure 1). The above discussion seems to indicate that in some cases neutralization of negatively charged AA could play a role and should be included in the analysis.

The complete sequence of MinD was run through the Heliquest program. For example, the MinD region AA 251–269 found with Heliquest corresponds well with the identified conserved sequence motif in *E.coli* 261–268 at the C-terminus (Szeto *et al.* 2002). Based on Heliquest-generated Eisenberg plot analysis this region is a potential amphipathic helix as was calculated and discussed before (Mileykovskaya *et al.* 2003). Two novel regions as depicted in table 2, were based on the prediction by the Heliquest lipid binding discrimination factor (region 16–33) and on the Heliquest-generated Eisenberg plot analysis (region 99–116). The latter region AA 99–116 (see figure 1 for a detailed depiction of the corresponding helical wheel plot) can be identified as surface seeking and although it is overall negatively charged it very well might be that as discussed above and indicated before (Mileykovskaya *et al.* 2003) a possible role of charge neutralisation plays a role here as well.

The diphtheria toxin fragment B is another interesting example. It is interesting to note that a role of a lower pH and the subsequent ability of certain regions to bind to lipids is well studied and discussed (Montagner *et al.* 2007). Three

Table 2. Data belonging to the lipid binding region search of a number of examples of amphitropic proteins where charge and hydrophobic face effect are involved

Sequence	H	μH	z	TM	LBR
CTP:Phosphocholine cytidylyl-transferase:					
$_{205}$IITRIVRDYDVYARRNLQ$_{222}$	0.313	0.423	2	-	Y
$_{240}$YHLQERVDKVKKKVKDVE$_{257}$	-0.074	0.421	2	-	Y
$_{248}$KVKKKVKDVEEKSKEFVQ$_{265}$	-0.178	0.490	3	-	Y
$_{273}$DLIQKWEEKSREFIGSFL$_{290}$	0.381	0.610	-1	-	N(S)
$_{292}$MFGPEGALKHMLKEGKGR$_{309}$	0.197	0.270	2	-	Y[b]
VEEKSKEFVQKVEEKSIDLIQKWEEKSREFIGS	0.095	0.433	-2	-	N[c]
VQEKSKEFVQKVEQKSIDLIQKWEQKSREFIGS	0.134	0.417	1	-	Y[c]
E. coli Min D:					
$_{16}$KTTSSAAIATGLAQKGKK$_{33}$	0.070	0.149	4	-	Y
$_{99}$ REGVAKVLDDLKAMDFEF $_{116}$	0.261	0.394	-2	-	N(S[a])
$_{251}$RPFRFIEEEKKGFLKRLF$_{268}$	0.287	0.498	3	-	Y(S[a])
Diphtheria toxin fragment B:					
$_{202}$INLDWDVIRDKTKTKIES$_{221}$	0.196	0.280	0	-	N(Y[e])
$_{226}$GPIKNKMSESWN$_{237}$	0.175	0.424	1	-	Y[d]
$_{239}$TVSEEKAKQWLEEFHQTA$_{256}$	0.178	0.340	-2	-	N(Y[e])
$_{260}$PELSELKTVTGTNW$_{277}$	0.390	0.221	-1	-	N(Y[e])
$_{335}$VAQSIALSSLMVAQAIPL$_{352}$	0.765	0.081	0	Y	N
$_{359}$IGFAAYNFVESIINLFQV$_{376}$	0.799	0.371	-1	Y	N
$_{1}$INLQWQVIRQKTKTKIES$_{18}$	0.287	0.241	3	-	Y

[a] According to the Heliquest-generated Eisenberg plot analysis this region is located in globular protein area but also close to the surface seeking area.

[b] According to SOPMA this region is slightly more than 50% helical, this region is therefore included for the sake of completeness.

[c] This corresponds to the 256-288 region as studied in Johnson *et al.* 2003.

[d] In this case a shorter window was used, SOPMA indicated this region to be a short helix.

[e] According to earlier indications in the literature this is a lipid binding region at low pH (see Montagner *et al.* 2007 for details).

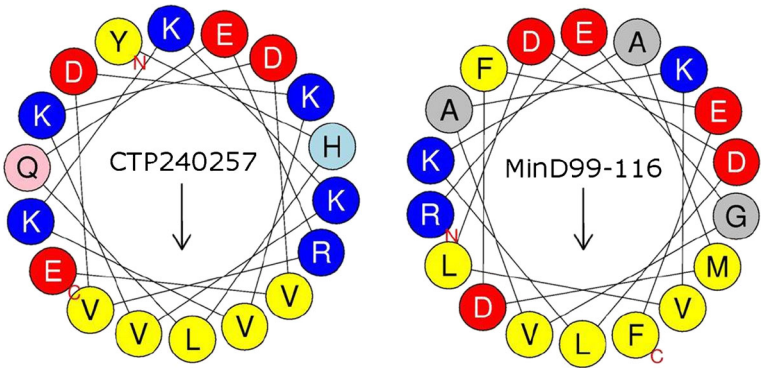


Figure 1. Helical wheel representations of CTP (AA 240-257) (left) and MinD (AA 99-116) (right). The best 18-residue result is depicted (see section 2 for details).

out of four described helical regions in this study are not identified by the Heliquest and/or Eisenberg plot approach. However when in for example the AA 202–221 region three negatively charged amino acids are replaced for neutral ones (to mimic the lower pH effect) this region is indeed identified by the Heliquest lipid binding discrimination factor, see table 2 for further details, again indicating that sometimes one has to include charge neutralization into the lipid binding region analysis. All other lipid binding regions as depicted in table 2 are nicely identified either by the Heliquest lipid binding discrimination factor (region 226–237) or the Heliquest-generated Eisenberg plot approach (regions 335–352 and 359–376).

3.3 *SecA: An interesting amphitropic protein*

SecA is a well-known and extensively studied motor protein involved in protein translocation. It was found that for *E.coli* SecA multiple lipid binding regions can be predicted (Keller 2011a). SecA can also be considered to be a member of the amphitropic protein family (Keller *et al.* 1995a; Keller 2011a). Inspired by an excellent sequence alignment report dealing with another Sec-protein SecDF (Eichler 2003), a similar attempt was performed for a number of SecA sequences from different origins. Figure 2 depicts some typical results for some of the regions; for a detailed picture of all data see supplementary figure 1 and supplementary table 1. It is interesting to see that some lipid binding regions are remarkably similar as might be expected for proteins with a reasonable degree of sequence homology. For example, region AA 108–125 in *E. coli* can be identified in almost all other organisms as well and all this with a high degree of sequence homology. However, some regions seem to be related to one specific organism and might indicate that certain lipid binding regions are organism specific. For example, region AA 323–340 in *B. subtilis* seems to be rather unique for this gram positive bacterium. The overall picture that some regions are well conserved while others seem unrelated to other organisms was also found for a similar type of alignment performed for SecD (Keller 2013).

3.4 *FtsY: A closer look*

Another interesting amphitropic protein is FtsY, the receptor of the signal recognition particle that mediates the targeting of integral membrane proteins in bacteria and thylakoids. A number of studies demonstrated protein–lipid interaction of FtsY with anionic phospholipids (see, for example, Braig *et al.* 2009; Parlitz *et al.* 2007). Interestingly, in relation to these observations for *E. Coli* FtsY is that indeed multiple lipid binding regions were predicted (Keller 2011a).

Table 3 summarizes the results of the lipid binding regions of FtsY as predicted for numerous species. For example, it has been demonstrated that deletion of the N-terminal residues 11–39 from the *S. coelicolor* FtsY (ScFtsY) drastically reduced its membrane-binding capability and that the N-terminus of ScFtsY alone was capable of targeting the soluble EGFP protein onto the membrane with high efficiency (Shen *et al.* 2012). Indeed the predicted region 9–26 fits perfectly into these experimental observations. The region 4–21 of the cpFtsY corresponds nicely with the extensively studied membrane binding motif that regulates GTPase activety (Marty *et al.* 2009). For *H. volcanii* FtsY one interesting region (226–243) is included, based on the arguments as presented in section 3.2. A sequence with a hydrophobic phase in the helical wheel plot and with a possible charge neutralisation is potentially a region capable of lipid binding. All other regions are predicted either based on the Heliquest lipid binding discrimination factor and/or the Heliquest-generated Eisenberg plot methodology. The closely related Ffh is included in this study as well. Compared to *E. coli* Ffh (Keller 2011b) the *B. subtilis* Ffh contains three

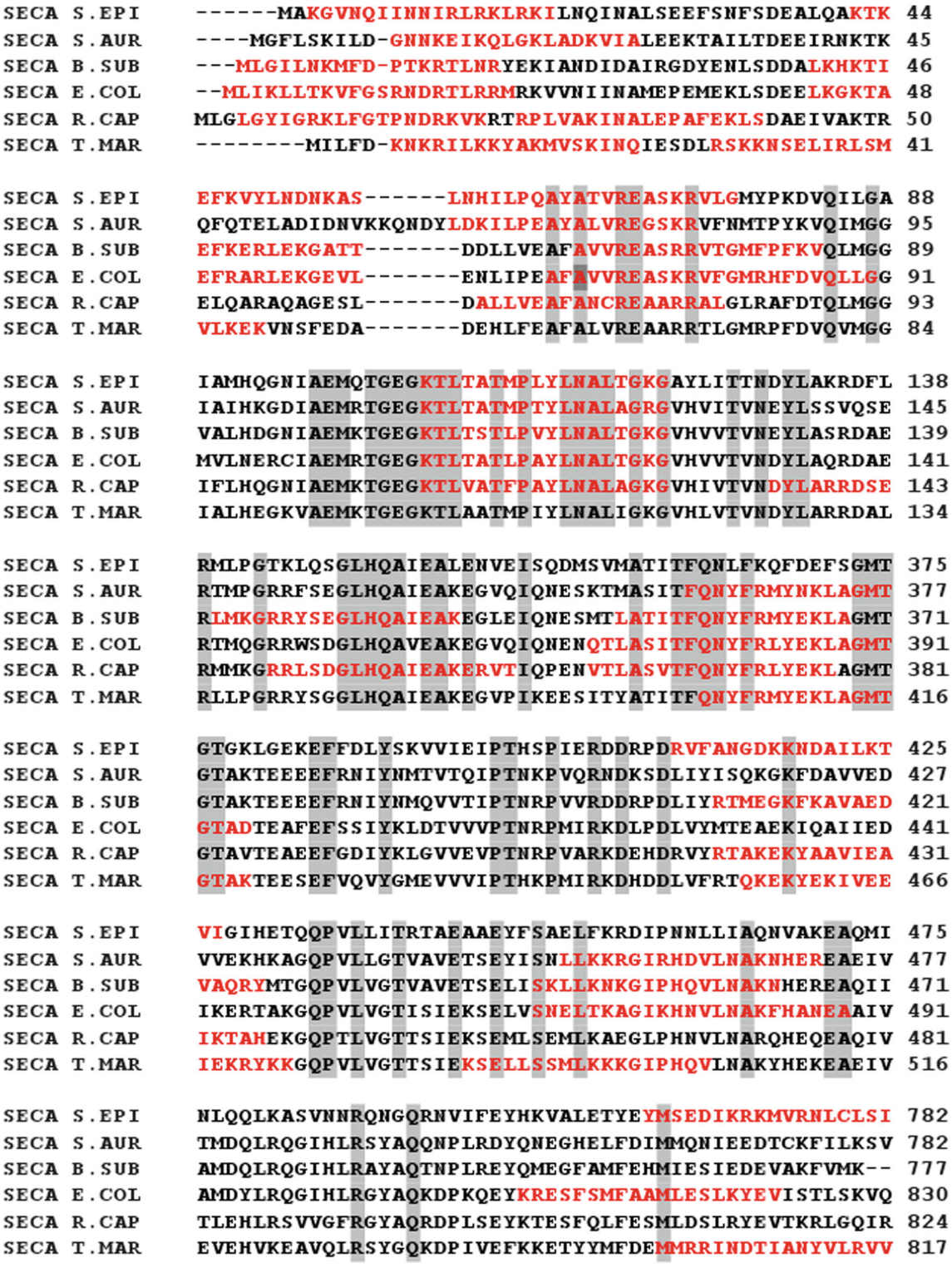


Figure 2. Sequence alignment of *S. epidermidis* SecA (Q8CMU9), *S.aureus* SecA (Q2YSH6), *B. subtilis* SecA (P28366), *E. coli* SecA (P10408), *R. capsulates* SecA (P52966) and *T. maritima* SecA (Q9X1R4), as obtained by ClustalW (Larkin *et al.* 2007). The indicated positions of the lipid binding regions (coloured) are based on the Heliquest-generated data. The residues with full sequence homology are indicated in grey.

Table 3. Data belonging to the lipid binding region search of a number of examples of amphitropic proteins involved in protein translocation

Sequence	H	μH	z	TM	LBR
B.subtilis FtsY(P51835)					
$_{1}$MSFFKKLKEKITKQTDSV$_{18}$	0.188	0.539	3	-	Y(S[a])
$_{30}$RNSFQNKVNDLVSRYRKV$_{47}$	0.013	0.483	4	-	Y
$_{67}$TTVMELIDELKKEVKRRN$_{84}$	0.062	0.451	1	-	Y
$_{133}$KTTTIGKLAHKMKQEGKS$_{150}$	0.006	0.277	4	-	Y[b]
$_{216}$NKVNLMKELEKVKRVIER$_{233}$	0.055	0.429	4	-	Y
H. volcanii FtsY(D4GYW6):					
$_{1}$MFDGLKKKLNRFRNDVEE$_{18}$	0.023	0.622	1	-	Y(S[a])
$_{158}$SSGPGRLRRAAAFATGKV$_{175}$	0.157	0.229	4	-	Y
$_{226}$GQLVSEALHDALYEVISV$_{243}$	0.551	0.301	-3	-	N[c]
$_{252}$IAEADKPVTLIFTGINGV$_{269}$	0.566	0.395	-1	-	N(S[a])[c]
$_{267}$NGVGKTTTIAKLAKYFEK$_{284}$	0.204	0.282	3	-	Y
S. coelicolor FtsY(Q9ZZP9):					
$_{9}$PVMEIVILAVVIAVVVIG$_{26}$	1.076	0.071	-1	Y	N
$_{111}$RLVRLRARLSRSQNALGK$_{128}$	0.094	0.431	6	-	Y
$_{161}$TQELVERLRERVKVLGTR$_{178}$	0.117	0.432	2	-	Y
$_{306}$RLHTKTGLMDELGKVKRV$_{323}$	0.168	0.209	3	-	Y
A. thaliana cpFtsY(O80842):					
$_{4}$GPSGFFTRLGRLIKEKAK$_{21}$	0.244	0.514	4	-	Y
$_{24}$VEKVFSGFSKTRENLAVI$_{41}$	0.353	0.434	1	-	Y
$_{72}$KITVRIVERLREDIMSGK$_{89}$	0.218	0.391	2	-	Y
$_{190}$AKAATVLSKAVKRGKEEG$_{207}$	0.036	0.329	3	-	Y
$_{224}$SLMEELIACKKAVGKIVS$_{241}$	0.472	0.467	1	-	Y
B. subtilis Ffh (P37105):					
$_{1}$MAFEGLADRLQQTISKIR$_{18}$	0.333	0.482	1	-	Y(S?)
$_{46}$FKVVKDFVKKVSERAVGQ$_{63}$	0.186	0.442	3	-	Y
$_{116}$TTSGKLANLLRKKHNRKP$_{133}$	0.024	0.268	6	-	Y
$_{413}$VNRLLKQFDEMKKMMKQM$_{430}$	0.217	0.607	3	-	Y(S[a])
$_{425}$KMMKQMTNMSKGKKKGFK$_{442}$	-0.046	0.475	7	-	Y

[a] According to the Heliquest-generated Eisenberg plot analysis this region is located in globular protein area but also close to the surface seeking area.

[b] According to SOPMA this region is slightly more than 50% helical, this region is therefore included for the sake of completeness.

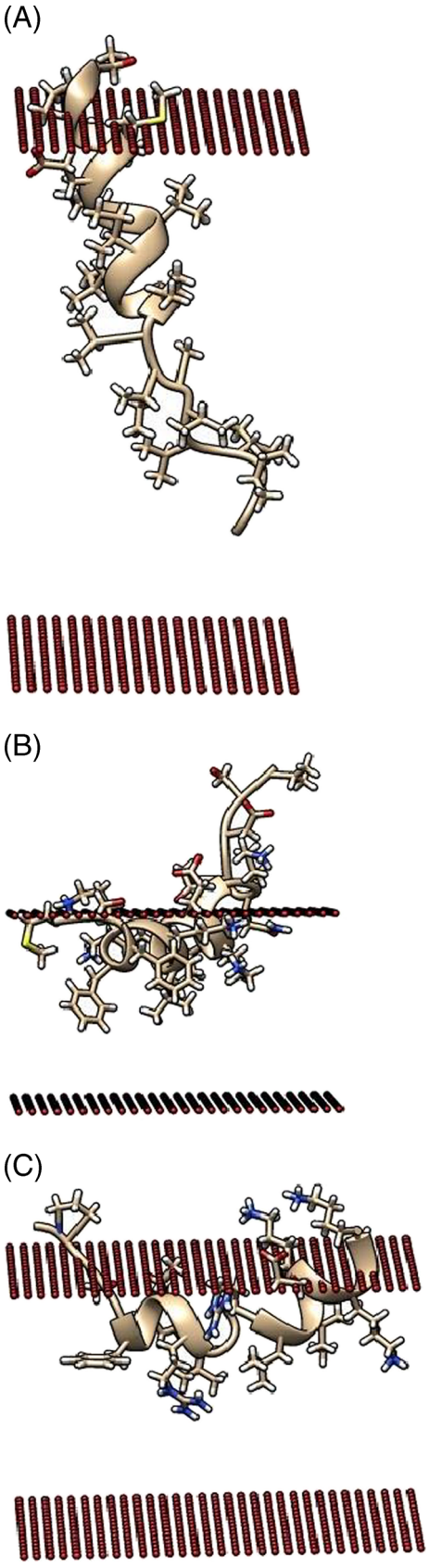
[c] This is not a lipid binding regions according to the lipid binding discrimination factor, however this region possesses a hydrophobic face.

[d] In this case a 17 AA window was used, see section 4 in text.

lipid binding region less. The percentages of Cardiolipin (CL), phosphatidylglycerol (PG) and lysylphosphatidylglycerol (lysPG) in *B. subtilis* is estimated to be 70–80% (Bishop *et al.* 1977; López *et al.* 2006) and therefor much higher than approximately 25% of negatively charged phospholipids in *E. coli* membranes. It is tempting to speculate that the lower amount of lipid binding regions in *B. subtilis* is related to the higher amount of negatively charged phospholipids in *B. subtilis* as compared to *E. coli* inner membranes.

3.5 *Further detailed characterization of lipid binding regions*

Once lipid binding regions have been identified by the in this report described prediction method it might be interesting to see what can be done to characterize such a lipid binding region further. First of all the helical wheel plot as depicted in figure 1 can provide some details about a particular region. Not only charge distribution but also the presence of a hydrophobic phase can be shown.



◀**Figure 3.** Monte Carlo simulations of three typical examples of helical as obtained by MCPep (Gofman *et al.* 2012). The representations belong to the region AA 9-26 of *S. coelicolor* FtsY (**a**), the region AA 1-18 of *B. subtilis* FtsY (**b**) and the region AA 4-21 of *A. thaliana*cpFtsY (**c**) respectively (see section 2 for details).

A recently developed program to perform Monte Carlo (MC) simulations of helical peptides in association with lipid membranes is checked for its use to characterize helical lipid binding regions in more detail. This program is developed to perform Monte Carlo (MC) simulations of the interaction of helical peptides with lipid bilayers and attempts to discriminate between the configurations transmembrane (TM) and the surface orientation (Gofman *et al.* 2012). In figure 3 a number of typical examples of the results are depicted. In all cases the test1, cluster 1 results are shown (see supplementary figure 2 for an additional example and more details).

The region AA 9–26 of *S. coelicolor* FtsY is according to the Heliquest-generated Eisenberg plot predicted as a typical transmembrane helix. Indeed figure 3A is depicting an image of a peptide oriented perpendicular on the plane of the membrane as expected for a transmembrane helix. The region AA 1–18 of *B. subtilis* FtsY is predicted to be a surface seeking helical lipid binding region. Indeed the helix in figure 3B depicts a surface bound helix parallel to the plane of the membrane as expected for an amphiphilic helix bound to the membrane surface. The region AA 4–21 of *A. thaliana* cpFtsY is predicted as a helical lipid binding region without typical transmembrane or surface seeking characteristics. Figure 3C depicts a peptide in a fashion with features of a surface and an oblique oriented helix. Although these simulations can only be used indicative it is clear that some preliminary ideas of the nature of the lipid binding can be obtained by using the MCPep server (Gofman *et al.* 2012).

4. Discussion

The availability of the Heliquest program allows prediction of helical lipid binding regions using the primary sequence of proteins (Gautier *et al.* 2008). Some interesting results have been already obtained using this approach (see, for example, Keller 2011a). The recently developed Eisenberg plot approach based on Heliquest-generated data (Keller 2011b) with or without the use of the Heliquest lipid binding discrimination factor has been proved to be a useful tool as well (Keller 2013; Lung and Chuong 2012; Phoenix *et al.* 2013). As discussed before (Keller 2011b) there is a necessity for a new terminology in the classification of proteins and protein regions. The protein classification needs, besides globular, surface-seeking and transmembrane, at least one new class: the amphitropic proteins.

In this study a number of amphitropic proteins with experimentally demonstrated lipid binding were analysed. In table 1 the results of this analysis are summarized and some of the identified regions are well described in the literature, and numerous novel lipid binding regions are found. Another set of typical amphitropic proteins were selected and summarized in table 2. A number of charge neutralization issues are described and discussed. Overall it is clear that multiple lipid binding regions can be found for each amphitropic protein. Based on the results of this study, a workflow chart can be built up as described in figure 4. In essence. as described in this paper. there are basically four factors determining a putative lipid binding region: First, the lipid binding discrimination factor (D) as determined by the Heliquest program; Second, if this factor fails to determine a lipid binding region, the Heliquest-generated Eisenberg plot approach might be able to identify a transmembrane or surface seeking region; Finally, as highlighted in this paper, an additional feature related to charge or charge neutralization could play a role, the helical wheel plot as generated by the Heliquest program might serve as a tool for this. In table 3 an extensive comparison of proteins from different origin are depicted and discussed. Although the lipid binding regions are often well conserved throughout the studied different species, it is not always simply related to a high sequence homology and some identified lipid binding regions seem to be specific for a particular organism. In relation to the amphipathic helix, a helix situated in the surface-seeking area of the Eisenberg plot, several attempts have been made to identify certain motifs as described in the literature. For example, the so-called ALPS (Amphipathic Lipid Packing Sensor) motif is excellently described and reviewed (Drin and Antonny 2010). The common motif of transmembrane helices, a helix that lies in the membrane area of the Eisenberg plot, is rather straightforward, a stretch of hydrophobic amino acids. However, the lipid binding regions identified in the so-called globular region solely depend on the physicochemical features hydrophobic moment (μH) and charge (z), and although no primary sequence-based motif is to be expected, it will be interesting to investigate this once more sequences have been identified.

In figure 5 all amphitropic helical lipid binding regions studied are depicted in a Heliquest-generated Eisenberg plot together with data obtained from an earlier study (Keller 2011b). The lipid binding regions in the surface and membrane region of the plot were identified using the Heliquest-generated Eisenberg plot approach: those located in the globular region are primarily based on the Heliquest lipid discrimination factor. From peptide studies it has been suggested that a particular region within the Eisenberg plot is

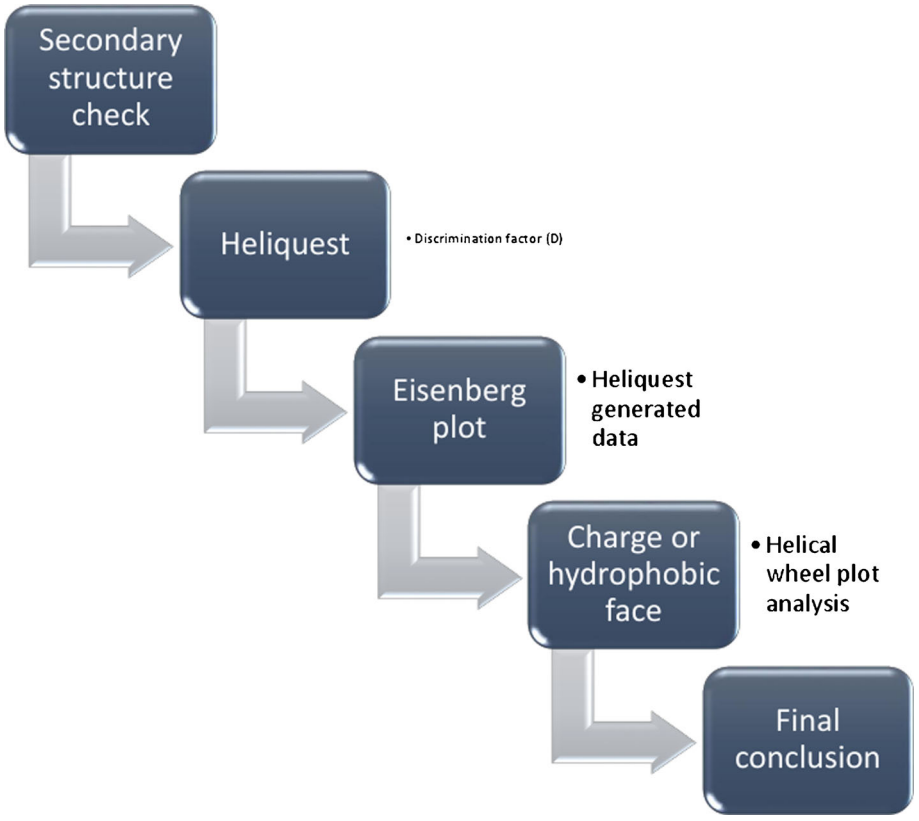


Figure 4. Workflow chart with the description of the consecutive steps necessarily in the search for lipid binding regions.

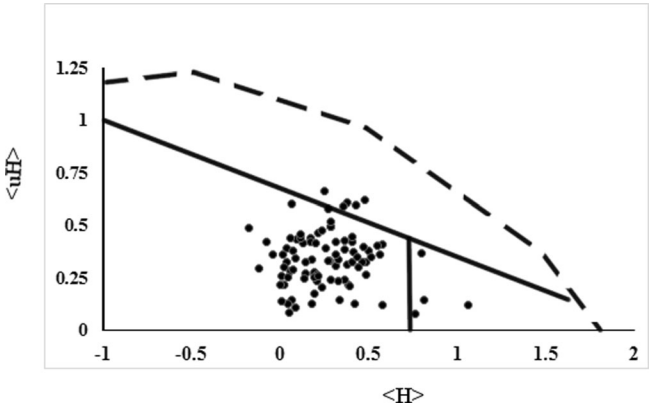


Figure 5. Eisenberg plot as obtained by Heliquest-generated data. The lipid binding regions as identified in this study and earlier (Keller 2011b) are depicted.

often indicative for oblique oriented α-helices and α-helical antimicrobial peptides (α-AMPs) (Phoenix *et al.* 2002; Dennison *et al.* 2005). This study seems to suggest a common feature of amphitropic proteins that the majority of all lipid binding regions (approximately 90%) are located in the so-called globular protein regions. Together with the general feature that multiple lipid binding regions can be identified this might indicate some of the features that proteins of the amphitropic protein family have in common.

Once a number of potential lipid binding regions in a protein are identified as here described, further characterization can be achieved by several possible experimental approaches. For example, FRET experiments can be performed by a suitable donor–acceptor couple specially equipped for lipid–protein interaction studies (Keller *et al.* 1995b), which has been successfully used to study both protein–lipid (Keller *et al.* 1996) and peptide–lipid (Caminati *et al.* 2004) interactions and is used for protein–protein interactions as well (Zhang *et al.* 2008). The total scan approach for the search of helical lipid binding regions as described in this study together with the one described elsewhere (Keller, 2011a, b) can therefore be considered to be a potentially interesting starting point for either a first characterization of those proteins with relatively yet unknown protein–lipid interaction features or for further in-depth characterization of proteins with already known overall protein–lipid interaction properties.

References

Bishop DG, Op den Kamp JA and van Deenen LL 1977 The distribution of lipids in the protoplast membranes of Bacillus subtilis. A study with phospholipase C and trinitrobenzenesulphonic acid. *Eur. J. Biochem.* **80** 381–391

Bogdanov M, Xie J, Heacock P and Dowhan W 2008 To flip or not to flip: lipid – protein charge interactions are a determinant of final membrane protein topology. *J. Cell Biol.* **182** 925–935

Bollinger JG, Diraviyam K, Ghomashchi F, Murray D and Gelb MH 2004 Interfacial binding of bee venom secreted Phospholipase A2 to membranes occurs predominantly by a nonelectrostatic mechanism. *Biochemistry* **43** 13293–13304

Braig D, Bär C, Thumfart J-O and Koch H-G 2009 Two cooperating helices constitute the lipid-binding domain of the bacterial SRP receptor. *J. Mol. Biol.* **390** 401–413

Burn P 1998 Amphitropic proteins: a new class of membrane proteins. *Trends Biochem. Sci.* **13** 79–83

Caminati G, Peroni E, Papini AM and Baglioni P 2004 Photophysical investigation of lipopeptides of myelin basic protein in phospholipid vesicles. *Progr. Colloid Polym. Sci.* **126** 163–168

Combet C, Blanchet C, Geourjon C and Deléage G 2000 NPS: network protein sequence analysis. *Trends Biochem. Sci.* **25** 147–150

Cserhati T 1993 Interaction of phospholipids with proteins and peptides. New advances III. *Int. J. Biochem.* **25** 123–131

Davidson WS, Jonas A, Clayton DF and George JM 1998 Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273** 9443–9449

Dennison SR, Harris F and Phoenix DA 2005 Are oblique orientated alpha-helices used by antimicrobial peptides for membrane invasion? *Protein Pept. Lett.* **12** 27–29

de Planque MR and Killian JA 2003 Protein–lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. *Mol. Membr. Biol.* **20** 271–284

Dowhan W 1997 Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annu. Rev. Biochem.* **66** 199–232

Dowhan W and Bogdanov M 2011 Molecular genetic and biochemical approaches for defining lipid-dependent membrane protein folding. *Biochim. Biophys. Acta* **1818** 1097–1107

Drin G and Antonny B 2010 Amphipathic helices and membrane curvature. *FEBS Lett.* **584** 1840–1847

Eichler J 2003 Evolution of the prokaryotic protein translocation complex: a comparison of archaeal and bacterial versions of SecDF. *Mol. Phylogenet. Evol.* **27** 504–509

Eisenberg D, Schwarz E, Komaromy M and Wall R 1984 Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179** 125–142

Epand RM 1998 Lipid polymorphism and protein–lipid interactions. *Biochim. Biophys. Acta* **1376** 353–368

Fantini J 2003 How sphingolipids bind and shape proteins: molecular basis of lipid-protein interactions in lipid shells, rafts and related biomembrane domains. *Cell. Mol. Life Sci.* **60** 1027–1032

Gautier R, Douguet D, Anthonny B and Drin G 2008 Heliquest: a web-server to screen sequences with specific α–helical properties. *Bioinformatics* **24** 2101–2102

Gennis RB 1977 Protein–lipid interactions. *Annu. Rev. Biophys. Bioeng.* **6** 195–238

Gofman Y, Haliloglu T and Ben-Tal N 2012 Monte Carlo simulations of peptide-membrane interactions with the MCPep web server. *Nucleic Acids Res.* **40** W358–W363

Halskau Ø, Muga A and Martínez A 2009 Linking new paradigms in protein chemistry to reversible membrane-protein interactions. *Curr. Protein Pept. Sci.* **10** 339–359

Hunte C 2005 Specific protein–lipid interactions in membrane proteins. *Biochem. Soc. Trans.* **33** 938–942

Johnson JE and Cornell RB 1999 Amphitropic proteins: Regulation by reversible membrane interactions (review). *Mol. Membr. Biol.* **16** 217–235

Johnson JE, Xie M, Singh LMR, Edge R and Cornell RB 2003 Both acidic and basic amino acids in an amphitropic enzyme, CTP: phosphocholine cytidylyltransferase, dictate its selectivity for anionic membranes. *J. Biol. Chem.* **278** 514–522

Karst JC, Barker R, Devi U, Swann MJ, Davi M, Roser SJ, Ladant D and Chenal A 2012 Identification of a region that assists membrane insertion and translocation of the catalytic domain of Bordetella pertussis CyaA toxin. *J. Biol. Chem.* **287** 9200–9212

Keller RCA, Snel MME, de Kruijff B and Marsh D 1995a SecA restricts in a nucleotide-dependent manner acyl chain mobility up to the center of a phospholipid bilayer. *FEBS Lett.* **358** 251–254

Keller RCA, Silvius JR and de Kruijff B 1995b Characterization of the novel resonance energy transfer couple coumarin-BODIPY and its possible applications. *Biochem. Biophys. Res. Commun.* **207** 508–514

Keller RCA, ten Berge D, Nouwen N, Snel MME, Tommassen J, Marsh D and de Kruijff B 1996 Mode of insertion of the signal sequence of a bacterial precursor protein into phospholipid bilayers as revealed by cysteine-based site directed spectroscopy. *Biochemistry* **35** 3063–3071

Keller RCA 2011a The prediction of novel multiple lipid-binding regions in protein translocation motor proteins: a possible general feature. *Cell. Mol. Biol. Lett.* **16** 40–54

Keller RCA 2011b New user-friendly approach to obtain an Eisenberg plot and its use as a practical tool in protein sequence analysis. *Int. J. Mol. Sci.* **12** 5577–5591

Keller RCA 2013 The prediction of lipid binding regions in cytoplasmic and extracellular loops of membrane proteins as exemplified by protein translocation membrane proteins. *J. Membr. Biol.* **246** 21–29

Krishtalik LI and Cramer WA 1995 On the physical basis for the cis-positive rule describing protein orientation in biological membranes. *FEBS Lett.* **369** 140–143

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM et al. 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* **23** 2947–2948

Lee AG 2005 How lipids and proteins interact in a membrane: a molecular approach. *Mol. BioSyst.* **1** 203–212

Lingwood D, Kaiser HJ, Levental I and Simons K 2009 Lipid rafts as functional heterogeneity in cell membranes. *Biochem. Soc. Trans.* **37** 955–960

López CS, Alice AF, Heras H, Rivas EA and Sánchez-Rivas C 2006 Role of anionic phospholipids in the adaptation of Bacillus subtilis to high salinity. *Microbiology* **152** 605–616

Lung SC and Chuong SD 2012 A transit peptide-like sorting signal at the C terminus directs the Bienertia sinuspersici preprotein receptor Toc159 to the chloroplast outer membrane. *Plant Cell* **24** 1560–1578

Marsh D 1995 Lipid-protein interactions and heterogeneous lipid distribution in membranes. *Mol. Membr. Biol.* **12** 59–64

Marsh D, Horváth LI, Swamy MJ, Mantripragada S and Kleinschmidt JH 2002 Interaction of membrane-spanning proteins with peripheral and lipid-anchored membrane proteins: perspectives from protein–lipid interactions (Review). *Mol. Membr. Biol.* **19** 247–255

Marty NJ, Rajalingam D, Kight AD, Lewis NE, Fologea D, Kumar TK, Henry RL and Goforth RL 2009 The membrane-binding motif of the chloroplast signal recognition particle receptor (cpFtsY) regulates GTPase activity. *J. Biol. Chem.* **284** 14891–14903

Mileykovskaya E, Fishov I, Fu X, Corbin BD, Margolin W and Dowhan W 2003 Effects of phospholipid composition on MinD-membrane interactions in vitro and in vivo. *J. Biol. Chem.* **278** 22193–22198

Montagner C, Perier A, Pichard S, Vernier G, Ménez A, Gillet D, Forge V and Chenal A 2007 Behavior of the N-terminal helices of the diphtheria toxin T domain during the successive steps of membrane interaction. *Biochemistry* **46** 1878–1887

Montecucco C, Papini E and Bisson R 1988 Protein–lipid interactions. *Ann. Ist. Super. Sanita* **24** 3–7

Mouritsen OG and Bloom M 1993 Models of lipid-protein interactions in membranes. *Annu. Rev. Biophys. Biomol. Struct.* **22** 145–171

Parlitz R, Eitan A, Stjepanovic G, Bahari L, Bange G, Bibi E and Sinning I 2007 Escherichia coli signal recognition particle receptor FtsY contains an essential and autonomous membrane-binding amphipathic helix. *J. Biol. Chem.* **44** 32176–32321

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE 2004 UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25** 1605–1612

Phoenix DA, Harris F, Daman OA and Wallace J 2002 The prediction of amphiphilic α-Helices. *Curr. Protein Pept. Sci.* **3** 201–221

Phoenix DA, Dennison SR and Harris F 2013 *Antimicrobial peptides* (Wiley-VCH Verlag GmbH & Co.: Weinheim)

Reshetnyak YK, Andreev OA, Segala M, Markin VS and Engelman DM 2008 Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane. *Proc. Natl. Acad. Sci. USA* **105** 15340–15345

Simons K and Sampaio JL 2011 Membrane organization and lipid rafts. *Cold Spring Harb. Perspect. Biol.* **3** a004697

Shen X, Li S, Du Y, Mao X and Li Y 2012 The N-terminal hydrophobic segment of Streptomyces coelicolor FtsY forms a transmembrane structure to stabilize its membrane localization. *FEMS Microbiol. Lett.* **327** 164–171

Shvadchak VV, Yushchenko DA, Pievo R and Jovin TM 2011 The mode of α-synuclein binding to membranes depends on lipid composition and lipid to protein ratio. *FEBS Lett.* **585** 3513–3519

Smith AW 2012 Lipid-protein interactions in biological membranes: a dynamic perspective. *Biochim. Biophys. Acta* **1818** 172–177

Szeto TH, Rowland SL, Rothfield LI and King GF 2002 Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. *Proc. Natl. Acad. Sci. USA* **99** 15693–15698

Tocanne JF, Cézanne L, Lopez A, Piknova B, Schram V, Tournier JF and Welby M 1994 Lipid domains and lipid/protein interactions in biological membranes. *Chem. Phys. Lipids* **73** 139–158

Török Z, Goloubinoff P, Horváth I, Tsvetkova NM, Glatz A, Balogh G, Varvasovszki V, Los DA, *et al*. 2001 Synechocystis HSP17 is an amphitropic protein that stabilizes heat-stressed membranes and binds denatured proteins for subsequent chaperone-mediated refolding. *Proc. Natl. Acad. Sci. USA* **98** 3098–3103

van Meer G, Voelker DR and Feigenson GW 2008 Membrane lipids; where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9** 112–124

Wang G, Keifer PA and Peterkofsky A 2003 Solution structure of the N-terminal amphitropic domain of Escherichia coli glucose-specific enzyme IIA in membrane-mimetic micelles. *Protein Sci.* **12** 1087–1096

Yoo J and Cui Q 2008 Does arginine remain protonated in the lipid membrane? Insights from microscopic pKa calculations. *Biophys. J.* L61–L63

Zhang X, Kung S and Shan S-O 2008 Demonstration of a multistep mechanism for assembly of the SRP·SRP receptor complex: Implications for the catalytic role of SRP RNA. *J. Mol. Biol.* **381** 581–593

MS received 09 July 2014; accepted 08 September 2014

Corresponding editor: AMIT CHATTOPADHYAY