J Membrane Biol (2013) 246:21–29
DOI 10.1007/s00232-012-9498-3

Prediction of Lipid-Binding Regions in Cytoplasmic and Extracellular Loops of Membrane Proteins as Exemplified by Protein Translocation Membrane Proteins

Rob C. A. Keller

Received: 16 June 2012 / Accepted: 18 August 2012 / Published online: 9 September 2012

Abstract The presence of possible lipid-binding regions in the cytoplasmic or extracellular loops of membrane proteins with an emphasis on protein translocation membrane proteins was investigated in this study using bioinformatics. Recent developments in approaches recognizing lipid-binding regions in proteins were found to be promising. In this study a total bioinformatics approach specialized in identifying lipid-binding helical regions in proteins was explored. Two features of the protein translocation membrane proteins, the position of the transmembrane regions and the identification of additional lipid-binding regions, were analyzed. A number of well-studied protein translocation membrane protein structures were checked in order to demonstrate the predictive value of the bioinformatics approach. Furthermore, the results demonstrated that lipid-binding regions in the cytoplasmic and extracellular loops in protein translocation membrane proteins can be predicted, and it is proposed that the interaction of these regions with phospholipids is important for proper functioning during protein translocation.

Keywords Eisenberg plot · Heliquest · Lipid-binding region · Protein–lipid interaction · SecYEG complex · Membrane protein

Electronic supplementary material The online version of this article (doi:10.1007/s00232-012-9498-3) contains supplementary material, which is available to authorized users.

R. C. A. Keller (✉)
Section of Chemistry, Charlemagne College,
Wilhelminastraat 13-15, 6524 AJ Nijmegen, The Netherlands
e-mail: rcakeller@kpnmail.nl

Introduction

It is nowadays well recognized that lipids play more roles in cells than just forming a lipid bilayer in order to create and maintain a membrane barrier. Over the past two or three decades it has been fascinating to see how the field of lipids and the notion of their multiple roles in biological processes fully emerged and ultimately even led to a completely new field in science, lipidomics (Wenk 2005).

A number of key findings have been important in this respect. First of all it is clear that on the level of the lipids themselves there is nowadays an awareness of the overwhelming diversity of lipids present in biomembranes (Dowhan 1997). Furthermore, on the level of the membrane, instead of the idea of homogenously distributed lipids in the membrane, a far more complex picture has become apparent than previously anticipated, with a variety of issues such as the heterogeneous distribution of lipids in lipid domains (Brown and London 1998) and the possible influence of the membrane curvature (Antonny 2011).

Our understanding of the role of lipids has evolved further by numerous protein–lipid interaction studies which indicated the influence of lipids on matters like protein and peptide secondary structure (Keller et al. 1992; Deber and Li 1995), transmembrane helix insertion (de Planque and Killian 2003) and membrane protein assembly (Stiegler et al. 2011). An important role of these studies has been the use and subsequent development of biophysical approaches suitable for protein–lipid interactions such as fluorescence spectroscopy (Keller et al. 1995; García-Sáez and Schwille 2008), ESR (Marsh 2010) and NMR (Ryba et al. 1986; Marsh and Páli 2004). Thanks to all of these research efforts, it is now believed that protein–lipid interactions are involved in various processes such as membrane protein functioning (Phillips et al. 2009), membrane protein

folding (Bogdanov and Dowhan 1999), antimicrobial peptide action (Haney et al. 2010) and protein translocation (van Klompenburg and de Kruijff 1998).

Last but not least the search for and characterization of lipid-binding domains in proteins have contributed to our current notion of the diverse roles of lipids (Stahelin 2009). The focus on these lipid-binding domains has led to an increase in our understanding of their possible role in diseases (Teasdale and Collins 2012) and as targets for drug development (Sudhahar et al. 2008). The attention to and ability to search even more specifically for lipid-binding regions in proteins and peptides can be an important step in understanding the role of lipids (Gautier et al. 2008; Brzeska et al. 2010). In this respect the recently developed Heliquest software, which enables identification of helical lipid-binding regions in proteins (Gautier et al. 2008), is a promising tool, as demonstrated recently by the predicted multiple lipid-binding regions in protein translocation motor proteins (Keller 2011a). Together with the hydrophobic moment plot methodology developed by Eisenberg et al. (1984), a promising tool to predict all possible helical lipid-binding regions in proteins is now available.

The presence of possible lipid-binding regions in the cytoplasmic and extracellular loops of membrane proteins is investigated in this study with an emphasis on protein translocation membrane proteins. Using the Heliquest software (Gautier et al. 2008), stretches of amino acids can be investigated with the use of a discrimination factor that enables discrimination between lipid-binding and non-lipid-binding regions of proteins. The Eisenberg plot approach was used to identify possible transmembrane helical segments and to characterize the possible lipid-binding regions. It will be demonstrated that in a number of the cytoplasmic and extracellular loops of the protein translocation membrane proteins investigated potential helical lipid-binding regions are predicted, and the possible implications will be discussed.

Methods

Primary and Secondary Structure Identification

The primary structure of the proteins was obtained from the Swiss-Prot sequence database, and the PBD codes used are indicated in the results as described (see the various tables). The included regions were checked for the extent of helicity using the available crystal structure data and/or data described in the corresponding sources. Additionally, secondary structure predictions were routinely checked using the program SOPMA (Combet et al. 2000; http://npsa-pbil.ibcp.fr/). Sequence alignment was performed using ClustalW, version 2.0 (Larkin et al. 2007; http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Determination of Lipid-Binding Potential

The mean hydrophobicity (<H>), the hydrophobic moment (μH) and the net charge (z) were calculated essentially as described (Keller 2011a). In essence, in the analysis mode 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 as $D = 0.944\ (<\mu H>) + 0.33\ (z)$. 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; see http://heliquest.ipmc.cnrs.fr/ for additional information).

Eisenberg Plot Approach

The Heliquest-generated Eisenberg plot approach was essentially performed as described elsewhere (Keller 2011b). In essence, both the mean hydrophobicity (<H>) and the hydrophobic moment (μH), as obtained by the Heliquest software (Gautier et al. 2008), can be plotted. The identification of possible transmembrane regions with <H> above 0.75 is the key parameter used in this study.

Helical Wheel Plot

The helical wheel representations were produced using the HeliQuest software (Gautier et al. 2008); in the analysis mode routinely 18-residue windows were used.

Structural Modeling

The 2D structural depiction of a number of studied proteins was displaced with the aid of the launched database of transmembrane proteins with known 3D structure called TOPDB (Tusnády et al. 2008; http://topdb.enzim.hu/). The 3D structures of SecG were generated by molecular modeling using I-Tasser (Zhang 2008; http://zhanglab.ccmb.med.umich.edu/I-TASSER/). Molecular models were viewed and analyzed using Chimera (Pettersen et al. 2004; http://www.cgl.ucsf.edu/chimera/). Final structural images were customized using the freely available digital image software analysis program GIMP (http://www.gimp.org).

Results

Rationale and Validation of the Lipid-Binding Region Search Approach

The rationale behind the approach utilized in this study was the use of the lipid-binding discrimination factor as determined by the Heliquest software (Gautier et al. 2008) in

order to identify possible helical lipid-binding regions. Additionally, the Eisenberg plot methodology was used with Heliquest-generated data (Keller 2011b) in order to identify possible lipid-binding regions missed by the lipid-binding discrimination factor. The Eisenberg plot approach was used to identify transmembrane (TM) helices and possible surface-seeking helices. The validation of this two-way approach has been demonstrated previously (Keller 2011b). In short, the Heliquest lipid discrimination factor was validated by the authors of the method report, which introduced the Web server Heliquest (Gautier et al. 2008). For the validation, multiple experimentally demonstrated lipid-binding proteins and peptides were used (see also the Help page of the Web server for additional information). The validation of the Eisenberg plot methodology is convincingly demonstrated in the original article (Eisenberg et al. 1984) and by the impressive number of reports using this approach. The use of Heliquest data to generate a hydrophobic moment plot is validated by the use of the original databases used by Eisenberg and coworkers and by additional examples of more recent reports with well-documented experimental proof for lipid binding (Keller 2011b).

The intrinsic nature of predictions is the uncertainty as to whether an approach which proved to render valuable data on one particular field (Keller 2011a) will work on another yet virtually unexplored area. The loop regions in membrane proteins were never systematically investigated in terms of protein–lipid interactions and lipid-binding ability. Two examples can be identified which approximate this issue (Lensink et al. 2010; Liu et al. 2012). How these examples fit into the results presented here will be elaborated further in Discussion.

The Protein Translocation Membrane Proteins Complex SecYEG

The membrane protein SecY is under physiological conditions an integrative part of the SecYEG complex (see for updated overviews Bondar et al. 2010; Mori et al. 2010). The search for lipid-binding regions with the Heliquest-generated data started with *Escherichia coli* SecY and subsequently with SecY as found in *Methanococcus jannaschii.* As depicted in Table 1, it is clear that in a number of loop structures potential lipid-binding helices can be predicted in SecY by the Heliquest lipid-binding discrimination factor. According to the Eisenberg plot approach, no surface-seeking helices could be identified. For the sake of completion also the TM helices were identified using the Eisenberg plot approach (see Supplementary Data, Table S1, for full details). In Fig. 1, a representation of the *M. jannaschii* SecY is depicted with the identified potential lipid-binding regions and TM helices. A number of interesting details are notable; for example, one region, AA 54–71, is predicted by the Heliquest-generated Eisenberg plot method to be a TM helix and appears in TOPDB (as well as in TOPCON) as a membrane loop. The positions of the identified TM helices are very much in line with what is described in the literature as well as with the results obtained by the state-of-the-art topology prediction method TOPCONS (Bernsel et al. 2009), especially if one takes the differences in the number of amino-residue windows investigated into account (see Supplementary Data, Table S2, for further details). With respect to the differences between the organisms studied, obviously the positions of the TM helices between *E. coli* SecY and *M. jannaschii* SecY differ due to differences in primary sequence. However, more interesting, as indicated in Table 1, the number of lipid-binding regions differs in the two organisms. One (and potentially two) lipid-binding region can be identified in *E. coli* SecY, while three (and potentially four) lipid-binding regions can be predicted in *M. jannaschii* SecY.

The membrane protein SecE is, like SecY, an integrative part of the SecYEG complex; and for both *E. coli* SecE and *M. jannaschii* SecE a lipid-binding region could be identified in one of the loop structures (see Table 1). The lipid-binding region in *E. coli* SecE is located between two TM helices (see Table S1 for further details), while in *M. jannaschii* SecE the predicted lipid-binding region is located at the terminal end of the sequence. Again, subtle but possibly relevant differences between the Sec proteins from these two different organisms were found.

The membrane protein SecG completes under physiological conditions the SecYEG complex. Throughout the years several functions for SecG have been postulated (Nishiyama et al. 1996; Bost and Belin 1997; Flower 2001). Recently, it has been convincingly demonstrated that SecG inversion is coupled to protein translocation activity (Morita et al. 2012). The search for lipid-binding regions in both *E. coli* SecG and *M. jannaschii* SecG rendered besides the well-known TM helices (see Table S1 for further details) also a helical lipid-binding region in one of the loop structures (see Table 1). Details of the two proteins are depicted in Fig. 2a. For example, the lipid-binding region in *E. coli* SecG can be characterized as a classical surface-seeking helix, as depicted in Fig. 2b. Using the molecular modeling program I-Tasser (Zhang 2008), indicative 3D structures were generated as well (see Supplementary Data, Fig. S1, for further details). Obviously, these structures are indicative since they are generated in the absence of membrane lipids and other membrane proteins like SecY and SecE; the positioning of the lipid-binding regions is, however, clearly indicated. It is again clear that for SecG as well subtle but possibly

Table 1 The identified lipid-binding regions of the protein translocation membranes of the SecYEG complex

Protein	Sequence	LBR	Surface-seeking
E. coli SecY	$_{9}$FQSAKGGLGELKRRLLFV$_{26}$	Y	N
	$_{101}$TLAEIKKEGESGRRKISQ$_{118}$	Y[a]	N
M. jannaschii SecY	$_{17}$PVKEITFKEKLKWTGIVL$_{34}$	Y[a]	N
(1 RHZ)	$_{100}$IPENRALFQGCQKLLSII$_{117}$	Y	N
	$_{357}$KGFRKSEKAIEHRLKRYI$_{374}$	Y	N
	$_{419}$LREKVSELHPAIAKLLNK$_{436}$	Y	N
E. coli SecE	$_{71}$FAREARTEVRKVIWPTRQ$_{88}$	Y	N
M. jannaschii SecE	$_{11}$QLKEFIEECRRVWLVLKK$_{28}$	Y	Y[b]
E. coli SecG	$_{46}$SGSGNFMTRMTALLATLF$_{63}$	Y	Y[b]
M. jannaschii SecG	$_{10}$ATSAGLIRYMDETFSKIR$_{27}$	Y	N

The lipid-binding region (LBR) is predicted by the Heliquest lipid-binding discrimination factor, and the surface-seeking potential is determined by the Heliquest-generated Eisenberg plot approach

[a] Less than 50 % helical according to SOPMA. For the sake of comparison, because of the high degree of sequence homology between the two SecY proteins, the regions are included in this overview

[b] Close to the border between globular and surface-seeking

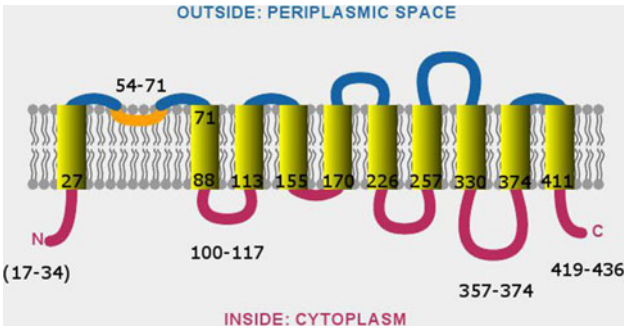


Fig. 1 Topology model of *M. jannaschii* SecY, as obtained by TOPDB (Tusnády et al. 2008). The indicated positions are based on the Heliquest-generated data. In *orange* a membrane loop, instead of a TM helix, is depicted (see "Results" section for further details) (Color figure online)

important differences are found for the protein translocation membrane proteins in *E. coli* and *M. jannaschii*.

It is tempting to speculate about the reported SecG inversion during protein translocation (Morita et al. 2012). A role for the predicted lipid-binding regions in such inversion of SecG might indicate one of the physiological functions of lipid-binding regions.

The Protein Translocation Membrane Proteins Complex SecDFyajC

The membrane protein SecD is under physiological conditions an integrative part of the SecDFyajC complex (see for additional information Sagara et al. 1994; Tsukazaki et al. 2011). It is clear that, besides the well-known and -described TM helices (see for full details Supplementary Data, Table S3), in a number of loop structures potential lipid-binding helices can be predicted in SecD (see Table 2 for summary). In Fig. 3a the sequence alignment of SecD in *E. coli* and *Thermus thermophilus* is depicted. The positions of the lipid-binding regions in these two organisms are completely different. However, the TM helices as predicted by the Heliquest-generated Eisenberg plot approach can be aligned nicely. In an interesting sequence alignment report (Eichler 2003) six conserved domains were identified along the various bacterial species. Figure 3a depicts these areas (in gray), and they were found to be aligned nicely as well (see Supplementary Data, Fig. S2, for full details). However, the positions of the lipid-binding regions seem less strictly conserved. Interestingly, one lipid-binding region in *E. coli* SecD is located at the periplasmic side, as depicted in Fig. 3b. Again, subtle but possibly important differences are found for the protein translocation membrane protein SecD in the different organisms. Also, in the SecF protein, multiple lipid-binding regions can be predicted. Interestingly enough, in contrast to SecD, for SecF in *T. thermophilus* more lipid-binding regions appear to be present than in *E. coli* SecF. Finally, as depicted in Table 2 also for the membrane protein yajC, part of the auxiliary SecDFyajC complex, a lipid-binding region can be predicted.

The Protein Translocation Membrane Protein YidC

E. coli YidC is involved in the integration of membrane proteins (Beck et al. 2001) and has been postulated to be a membrane insertase for Sec-independent proteins (Serek et al. 2004). It is clear that, besides the well-known and -described TM helices (see for full details Supplementary Data, Table S3), potential lipid-binding helices can be predicted in a number of loop structures of YidC (see Table 2 for summary). This result seems to underline that

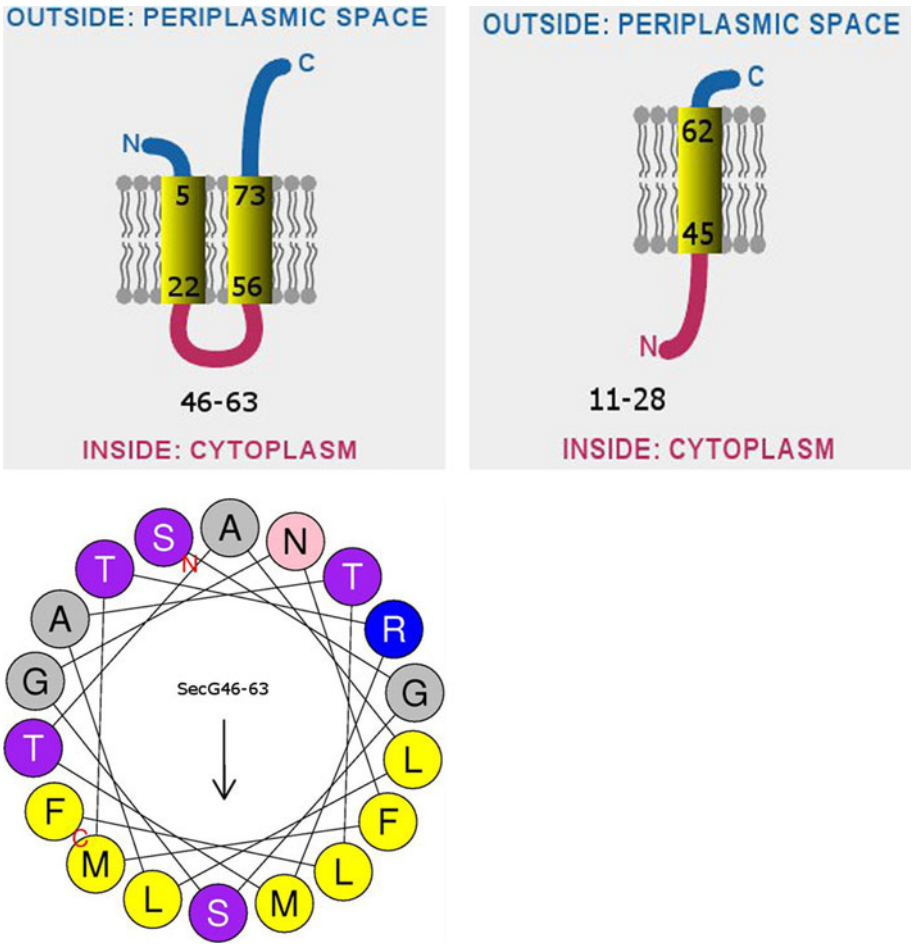
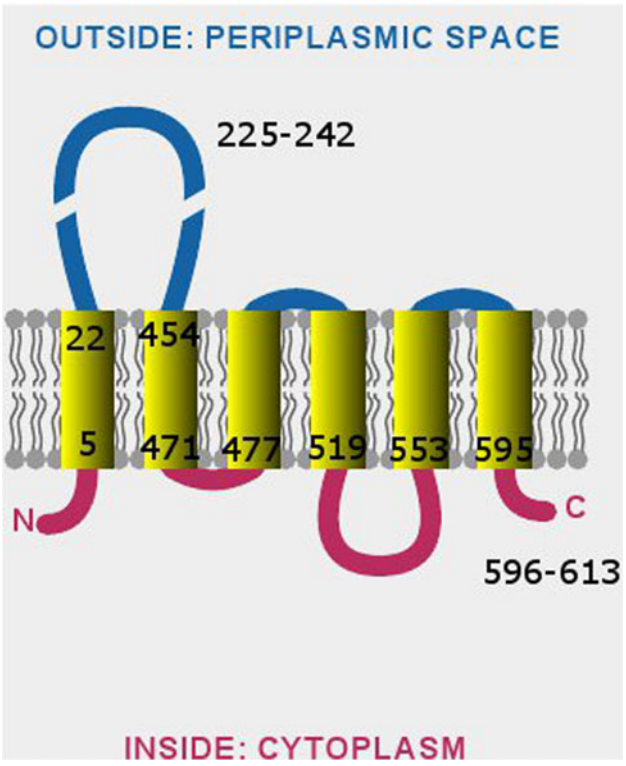


Fig. 2 Topology models of *E. coli* SecG (*left*) and *M. jannaschii* SecG (*right*), as obtained by TOPDB (Tusnády et al. 2008). The indicated positions are based on the Heliquest-generated data. Helical wheel representation of *E. coli* SecG(46–63). The best 18-residue window result is depicted (see "Methods" section for details) (Color figure online)

Table 2 The identified lipid-binding regions of the protein translocation membranes of the SecDFyajC complex, YidC and members of the Tat pathway

Protein	Sequence	LBR	Surface-seeking
E. coli SecD	$_{225}$YAVQQNINILRNRVNQLG$_{242}$	Y	N
	$_{596}$GTRAIVNLLYGGKRVKKL$_{613}$	Y	N
T. thermohilus SecD	$_{354}$RAGKKLRQAIPEGFRHST$_{371}$	Y	N
E. coli SecF	$_{217}$VSDRIRENFRKIRRGTPY$_{234}$	Y	N
T. thermophilus SecF	$_{207}$VSDRIRENQKLLRHLPYA$_{224}$	Y	N
	$_{227}$VNRSINQTLSRTVMTSLT$_{244}$	Y	N
	$_{278}$SIYVVSALVVAWKNRRKA$_{295}$	Y	N
E. coli YajC	$_{42}$QQKRTKEHKKLMDSIAKG$_{59}$	Y	N
E. coli YidC	$_{338}$QPLFKLLKWIHSFVGNWG$_{355}$	Y	Y
	$_{373}$TKAQYTSMAKMRMLQPKI$_{390}$	Y	N
	$_{524}$TIIQQQLIYRGLEKRGLH$_{541}$	Y	N
E. coli TatA	$_{23}$KKLGSIGSDLGASIKGFK$_{40}$	Y	N
E. coli TatB	$_{30}$KTVAGWIRALRSLATTVQ$_{47}$	Y	Y
	$_{62}$SLKKVEKASLTNLTPELK$_{79}$	Y	N
	$_{78}$LKASMDELRQAAESMKRS$_{95}$	Y	N
E. coli TatC	$_{7}$QPLITHLIELRKRLLNCI$_{24}$	Y	Y

```
SECD E.COL   MLNRYPLWKYVMLIVVIVIGLLYALPNLFGEDPAVQITGARGVAASEQTL 50
SECD H.INF   MLNRYPLWKNIMVIFIVAIGILYSLPNIYGEDPAVQISGTRGQEANTSVL 50
SECD T.THE   -MNRKNLTS-LFLLGVFLLALLFVWKPWAPEEPKVRLG------------ 36
SECD M.JAN   -MDISKLLKDRKILILIIFVTLSVFLIVFKG------------------- 30

SECD E.COL   VALNLAPATPRWLAAIHAEPMKLGLDLRGGVHFLMEVDMDTALGKLQEQN 150
SECD H.INF   TALNLAPATPAWLSMFGANPMKWGLDLRGGVRFLMEVDMNATLVKRQEQL 150
SECD T.THE   ------------------------LDLKGGLRIVLEADVEN--------- 53
SECD M.JAN   ----------------------LDFGIDLSGGTIIVLKAEKPMS-------- 52

SECD E.COL   VISSQGSNQLRAVMSDARLSEAREYAVQQNINILRNRVNQLGVAEPVVQR 250
SECD H.INF   DITEPDADSINLGLSTAALNEARDLAIEQNLTIIRKRVAELGVAEAVIQR 250
SECD T.THE   DD------------------------LEKARTVLENRINALGVAEPLIQI 82
SECD M.JAN   ----------------------DKEIEATIKIITERLNYNGLNDVVIYP 79

SECD E.COL   INIANIQSRLGNSFRITGINNPNEARQLSLLRAGALIAPIQIVEERTIG 442
SECD H.INF   INVATIQGRFGSNFQITGVDSIAEAHNLSTLLKSGALIAPIQIVEERTIG 443
SECD T.THE   IRQAITGG----QAVIEGLSSVEEASEIALVLRSGSLPVPLKVAEIRAIG 263
SECD M.JAN   ITVGAYPP---------TKEEIDEAMAIYSALKSGALPVKLDIEYISTIS 235

SECD E.COL   PTLGMQNIEQGLEACLAGLLVSILFMIIFY-KKFGLIATSALIANLILIV 491
SECD H.INF   PSLGAQNVEQGINASLWGLVAVIAFMLFYY-KMFGVIASFALVINIVLLV 492
SECD T.THE   PTLGQDAIQAGIRSALIGTLAIFLLIFAYYGPHLGLVASLGLLYTSALIL 313
SECD M.JAN   PEFGKEFLKGTAIALLLAFIAVGIIVSIRYKQPKIAIPILITCISEVIII 285

SECD E.COL   GIMSLLPGATLSMPGIAGIVLTLAVAVDANVLINERIKEELSNGRTVQQA 541
SECD H.INF   GIMSILPGATLSMPGIAGIVLTLGMSVDANVLIFERIKEEIRNGRSIQQA 542
SECD T.THE   GLLSGL-GATLTLPGIAGLVLTLGAAVDGNVLSFERIKEELRAGKKLRQA 362
SECD M.JAN   LGFASLIDWKLDLPSIAGIIAAVGTGVDNQIVITD---EALKRG---AGK 329

SECD E.COL   IDEGYRGAFSSIFDANITTLIKVIILYAVGTGAIKGFAITTGIGVATSMF 591
SECD H.INF   INEGYNGAFTSIFDANLTTILTAIILYAVGTGPIQGFAITLSLGVAISMF 592
SECD T.THE   IPEGFRHSTLTIMDVNIAHLLAAAALYQYATGPVRGFAVILAIGVVASVF 412
SECD M.JAN   IRASIKRAFFIIFASAATSIAAMLPLFVLGVGMLKGFAITTIAGVLIGIF 379

SECD E.COL   TAIVGTRAIVNLLYGGKRVKKLSI 615
SECD H.INF   TAITGTRALVNALYGGKQLKKLLI 616
SECD T.THE   SNLVFSRHLLERLADRGEIRPP-- 434
SECD M.JAN   ITRPAFARIIEEMFKKF------- 396
```



◀ **Fig. 3** Sequence alignment of *E. coli* SecD (P0AG90), *Haemophilus influenza* SecD (P44591), *T. thermophilus* SecD (Q5SKE6) and *M. jannaschii* SecD (Q57575), as obtained by ClustalW (Larkin et al. 2007). The indicated positions of the TM helices (*underlined*) and the lipid-binding regions (*colored*) are based on the Heliquest-generated data. The conserved domains as defined by Eichler (2003) are indicated in *gray*. The lines of amino acids containing TM helices are *underlined*, and the lipid-binding helices are indicated in *red*. In the *lower part* of the figure a topology model of *E. coli* SecD, as obtained by TOPDB (Tusnády et al. 2008) is depicted (Color figure online)

the presence of more lipid-binding regions is a general feature of all protein translocation membrane proteins.

The Protein Translocation Membrane Proteins TatA, TatB and TatC

The twin-arginine translocation pathway also known as the Tat pathway serves to actively translocate, in contrast to the Sec pathway, proteins in a folded manner across a lipid membrane (Shanmugham et al. 2006). In *E. coli* the Tat system is minimally composed of the three integral membrane proteins TatA, TatB and TatC. A number of Tat proteins have been well studied when it comes to the presence of lipid-binding regions and can therefore be used to check the validity of the lipid-binding regions approach as used in this study. For example, the protein translocation membrane protein TatA was investigated in this study. Previously, one transmembrane region (residues 4–21) and one amphipathic helix (residues 23–44) were found and thoroughly investigated (Greene et al. 2007). These findings correspond nicely with the results obtained with the approach used in this study, which also identifies one TM helix (residues 3–20) using the Heliquest-generated Eisenberg plot approach and one lipid-binding region (residues 23–40) using the Heliquest lipid-binding discrimination factor at virtually the same position (see Table 2 and Supplementary Data, Table S3, for full details).

The protein translocation membrane protein TatB was investigated in this study as well. One transmembrane region (residues 5–20) and one amphipathic helix (residues 23–53) were found previously (Lee et al. 2006). These findings correspond nicely with the results obtained with the approach used in this study, which also identifies one TM helix (residues 3–20) and one lipid-binding region (residues 30–47) at virtually the same position, both using the Heliquest-generated Eisenberg plot approach. Interestingly, two novel lipid-binding regions can be identified (regions 62–79 and 78–95) with the use of the Heliquest lipid-binding discrimination factor. The finding of these two additional lipid-binding regions in the N-terminal

region might explain why removal of the TM helix still renders peripheral interactions of TatB with the cytoplasmic membrane (de Leeuw et al. 2001).

The protein translocation membrane protein TatC was also investigated in this study. The approach used in this study identified one lipid-binding region (residues 7–24) and five TM regions (see Table 2 and Supplementary Data, Table S2, for full details). The lipid-binding region in TatC is, together with the first lipid-binding region (residues 30–47) of TatB, the only amphipathic helix which can be characterized as a classical surface-seeking helix as defined by Eisenberg plot analysis. All other lipid-binding regions in the Tat membrane proteins can only be identified by the Heliquest lipid-binding discrimination factor.

Discussion

Protein–lipid interactions play an unquestionably significant role in multiple cellular processes. The identification of lipid-binding regions in such proteins could contribute to a better mechanistic understanding of such protein–lipid interactions. In the protein translocation motor protein SecA, for example, two lipid-binding regions were postulated as a possible explanation of the experimental observations of phospholipid vesicle aggregation (Breukink et al. 1993). These postulated lipid-binding sites were recently predicted by the use of bioinformatics approaches along with a substantial number of novel lipid-binding regions (Keller 2011a). This is interesting since it has been proposed previously (Yang et al. 1997; Wang et al. 2003; Chen et al. 2007) and recently experimentally demonstrated that SecA alone can promote the translocation of a precursor protein (Hsieh et al. 2011). In other words, when SecA can facilitate the protein translocation of a precursor protein in a pure lipid system (Hsieh et al. 2011) this requires most likely multiple lipid-binding regions in the SecA protein, indicating the physiological relevance of lipid-binding region predictions.

Impressive progress has been made in our understanding of the functioning of membrane proteins and the process of membrane protein assembly and folding (von Heijne 1994; Vitrac et al. 2011). Although a role for the loop regions in membrane protein assembly has been postulated (von Heijne and Gavel 1988) and demonstrated (van Klompenburg et al. 1997), it is clear that the primary focus in membrane proteins, when it comes to protein–lipid interactions, has been on the TM helices. This study provides a first systematic attempt to focus on the loop regions in protein translocation membrane proteins in relation to possible involvement of protein–lipid interactions in such regions. Due to the fact that crystallization of proteins remains a difficult and even a hazardous process (Chayen and Saridakis 2008), for fragile proteins like SecDF and SecYEG it is often noticed that such proteins are isolated from more extremophile or thermophile origins with the idea that proteins isolated from such organisms are more firm and stable. For this reason, the sequences of SecYEG and SecD were investigated not only from the most studied organism, *E. coli*, but also from *M. jannaschii* and *T. thermophiles*, respectively.

The results clearly demonstrate the power of the approach based on Heliquest-generated data. First of all, in the loop structures of all investigated protein translocation membrane proteins one or more potential helical lipid-binding region can be predicted based on the Heliquest approach. In this approach the secondary prediction method SOPMA is routinely used along with secondary structural data obtained from other sources like the crystal structure. Despite its relatively straightforward approach, this method seems to be remarkably reliable since the predictions correspond very well with secondary predictions as given by, for example, a state-of-the-art program like I-Tasser (Zhang 2008) (see, e.g., the comparison for SecG in Supplementary Data, Table S4).

Furthermore, the TM helices are well depicted in the structure based on the Heliquest-generated Eisenberg plot approach. As discussed in the Results section, the positions of the TM regions correspond remarkably well with the state-of-the-art topology program TOPCONS (Bernsel et al. 2009) (see, e.g., the SecY comparison in Supplementary Data, Table S2). In this respect it is interesting to note that the TOPCONS results concerning the absolute value of the Z coordinate (as expressed in Å) seem to indicate that the lipid-binding regions are positioned either flat on the plane of the membrane or slightly tilted in the headgroup region of phospholipids. Using TOPCONS, no so-called reentrant regions were found for the protein translocation membrane proteins (data not shown).

The results for proteins like TatA and TatB seem to indicate the predictive power of the Heliquest-generated data approach since all TM helices and lipid-binding regions as described in the literature (Lee et al. 2006; Greene et al. 2007) are nicely identified. The possible lipid-binding ability of membrane loop regions has not been systematically investigated. Recently, however, two articles were published which approximate this issue. One molecular dynamic simulation study (Lensink et al. 2010) identified a specific protein–lipid interaction in LacY between Asp-68 and PE. Indeed, a Heliquest lipid-binding discrimination factor search renders a lipid-binding region, AA 65–82 ($H = 0.748$, $H = 0.165$, $z = 2$), which corresponds nicely. Another recent report (Liu et al. 2012) identified a lipid interaction at the C terminus in the ER protein atlastin. Indeed, a lipid-binding region, AA 1–18, could be identified ($H = 0.519$, $H = 0.480$, $z = 0$), which is missed by the Heliquest lipid-binding discrimination

factor but is nicely identified as a possible surface-seeking helix by the Eisenberg plot methodology. Together with the confirmatory results of the predictions of, for example, TatA, this seems to indicate that the predictions as determined in this study represent a reasonable predictive value, presumably higher than 80 %, as discussed and presented previously (Keller 2011b).

The physiological relevance of lipid-binding regions might be multifunctional. The positive inside rule as postulated by von Heijne and coworkers (von Heijne and Gavel 1988; von Heijne 1992) basically states that apolar regions in the TM helices are targets for membrane integration and that charged residues in the loop regions provide topological information. This obviously provides one possible role of the lipid-binding regions in loop structures of certain membrane proteins. The results as described in this study indicate that in protein translocation membrane proteins the lipid-binding regions in the loop structures might be actively involved in protein translocation. It is interesting to note that the lipid-binding regions are primarily located on the cytoplasmic side of the protein translocation membrane proteins, which is the initial side of action when it comes to protein translocation. Only in the case of SecDFyajC can a number of lipid-binding regions on the periplasmic side be identified. Indeed, this complex is believed to fulfill its auxiliary effect on the periplasmic side. Last but not least is the demonstrated SecG inversion during protein translocation (Morita et al. 2012), an example of a process where lipid-binding regions in loop structures of protein translocation membrane proteins might be involved during one or more stages of protein translocation.

References

Antonny B (2011) Mechanisms of membrane curvature sensing. Annu Rev Biochem 80:101–123

Beck K, Eisner G, Trescher D, Dalbey RE, Brunner J, Müller M (2001) YidC, an assembly site for polytopic *Escherichia coli* membrane proteins located in immediate proximity to the SecYE translocon and lipids. EMBO Rep 2:709–714

Bernsel A, Viklund H, Hennerdal A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res 37:W465–W468

Bogdanov M, Dowhan W (1999) Lipid-assisted protein folding. J Biol Chem 274:36827–36830

Bondar AN, del Val C, Freites JA, Tobias DJ, White SH (2010) Dynamics of SecY translocons with translocation-defective mutations. Structure 18:847–857

Bost S, Belin D (1997) prl mutations in the *Escherichia coli secG* gene. J Biol Chem 272:4087–4093

Breukink E, Keller RCA, de Kruijff B (1993) Nucleotide and negatively charged lipid-dependent vesicle aggregation caused by SecA. Evidence that SecA contains two lipid-binding sites. FEBS Lett 331:19–24

Brown DA, London E (1998) Structure and origin of ordered lipid domains in biological membranes. J Membr Biol 164:103–114

Brzeska H, Guag J, Remmert K, Chacko S, Korn ED (2010) An experimentally based computer search identifies unstructured membrane-binding sites in proteins: application to class I myosins, PAKS, and CARMIL. J Biol Chem 285:5738–5747

Chayen NE, Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. Nat Methods 5:147–153

Chen Y, Tai PC, Sui SF (2007) The active ring-like structure of SecA revealed by electron crystallography: conformational change upon interaction with SecB. J Struct Biol 159:149–153

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

de Leeuw E, Porcelli I, Sargent F, Palmer T, Berks BC (2001) Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway. FEBS Lett 506:143–148

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

Deber CM, Li SC (1995) Peptides in membranes: helicity and hydrophobicity. Biopolymers 37:295–318

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

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, Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179:125–142

Flower AM (2001) SecG function and phospholipid metabolism in *Escherichia coli*. J Bacteriol 183:2006–2012

García-Sáez AJ, Schwille P (2008) Fluorescence correlation spectroscopy for the study of membrane dynamics and protein/lipid interactions. Methods 46:116–122

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

Greene NP, Porcelli I, Buchanan G, Hicks MG, Schermann SM, Palmer T, Berks BC (2007) Cysteine scanning mutagenesis and disulfide mapping studies of the TatA component of the bacterial twin arginine translocase. J Biol Chem 282:23937–23945

Haney EF, Nathoo S, Vogel HJ, Prenner EJ (2010) Induction of non-lamellar lipid phases by antimicrobial peptides: a potential link to mode of action. Chem Phys Lipids 163:82–93

Hsieh YH, Zhang H, Lin BR, Cui N, Na B, Yang H, Jiang C, Sui SF, Tai PC (2011) SecA alone can promote protein translocation and ion channel activity: SecYEG increases efficiency and signal peptide specificity. J Biol Chem 286:44702–44709

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, Killian JA, de Kruijff B (1992) Anionic phospholipids are essential for α-helix formation of the signal peptide of prePhoE upon interaction with phospholipid vesicles. Biochemistry 31:1672–1677

Keller RCA, Silvius JR, De Kruijff B (1995) Characterization of the resonance energy transfer couple coumarin-BODIPY and its possible applications in protein–lipid research. Biochem Biophys Res Commun 207:508–514

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. Bioinformatics 23:2947–2948

Lee PA, Orriss GL, Buchanan G, Greene NP, Bond PJ, Punginelli C, Jack RL, Sansom MSP, Berks BC, Palmer T (2006) Cysteine-scanning mutagenesis and disulfide mapping studies of the conserved domain of the twin-arginine translocase TatB component. J Biol Chem 281:34072–34085

Lensink MF, Govaerts C, Ruysschaert JM (2010) Identification of specific lipid-binding sites in integral membrane proteins. J Biol Chem 285:10519–10526

Liu TY, Bian X, Sun S, Hu X, Klemm RW, Prinz WA, Rapoport TA, Hu J (2012) Lipid interaction of the C terminus and association of the transmembrane segments facilitate atlastin-mediated homotypic endoplasmic reticulum fusion. Proc Natl Acad Sci USA 109:E2146–E2154

Marsh D (2010) Electron spin resonance in membrane research: protein–lipid interactions from challenging beginnings to state of the art. Eur Biophys J 39:513–525

Marsh D, Páli T (2004) The protein–lipid interface: perspectives from magnetic resonance and crystal structures. Biochim Biophys Acta 1666:118–141

Mori T, Ishitani R, Tsukazaki T, Nureki O, Sugita Y (2010) Molecular mechanisms underlying the early stage of protein translocation through the Sec translocon. Biochemistry 49:945–950

Morita K, Tokuda H, Nishiyam KI (2012) Multiple SecA molecules drive protein translocation across a single translocon with SecG inversion. J Biol Chem 287:455–464

Nishiyama K, Suzuki T, Tokuda H (1996) Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. Cell 85:71–81

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

Phillips R, Ursell T, Wiggins P, Sens P (2009) Emerging roles for lipids in shaping membrane–protein function. Nature 459:379–385

Ryba NJ, Dempsey CE, Watts A (1986) Protein–lipid interactions at membrane surfaces: a deuterium and phosphorus nuclear magnetic resonance study of the interaction between bovine rhodopsin and the bilayer head groups of dimyristoylphosphatidylcholine. Biochemistry 25:4818–4825

Sagara K, Matsuyama S, Mizushima S (1994) SecF stabilizes SecD and SecY, components of the protein translocation machinery of the *Escherichia coli* cytoplasmic membrane. J Bacteriol 176:4111–4116

Serek J, Bauer-Manz G, Struhalla G, van den Berg L, Kiefer D, Dalbey R, Kuhn A (2004) *Escherichia coli* YidC is a membrane insertase for Sec-independent proteins. EMBO J 23:294–301

Shanmugham A, Wong Fong Sang HW, Bollen YJ, Lill H (2006) Membrane binding of twin arginine preproteins as an early step in translocation. Biochemistry 45:2243–2249

Stahelin RV (2009) Lipid-binding domains: more than simple lipid effectors. J Lipid Res 50:S299–S304

Stiegler N, Dalbey RE, Kuhn A (2011) M13 procoat protein insertion into YidC and SecYEG proteoliposomes and liposomes. J Mol Biol 406:362–370

Sudhahar CG, Haney RM, Xue Y, Stahelin RV (2008) Cellular membranes and lipid-binding domains as attractive targets for drug development. Curr Drug Targets 9:603–613

Teasdale RD, Collins BM (2012) Insights into the PX (phox-homology) domain and SNX (sorting nexin) protein families: structures, functions and roles in disease. Biochem J 441:39–59

Tsukazaki T, Mori H, Echizen Y, Ishitani R, Fukai S, Tanaka T, Perederina A, Vassylyev DG, Kohno T, Maturana AD, Ito K, Nureki O (2011) Structure and function of a membrane component SecDF that enhances protein export. Nature 474:235–238

Tusnády GE, Kalmár L, Simon I (2008) TOPDB: topology data bank of transmembrane proteins. Nucleic Acids Res 36:D234–D239

van Klompenburg W, de Kruijff B (1998) The role of anionic phospholipids in protein insertion and translocation in bacterial membranes. J Membr Biol 162:1–7

van Klompenburg W, Nilsson I, von Heijne G, de Kruijff B (1997) Anionic phospholipids are determinants of membrane protein topology. EMBO J 16:4261–4266

Vitrac H, Bogdanov M, Heacock P, Dowhan W (2011) Lipids and topological rules of membrane protein assembly: balance between long and short range lipid–protein interactions. J Biol Chem 286:15182–15194

von Heijne G (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol 225:487–494

von Heijne G (1994) Membrane proteins: from sequence to structure. Annu Rev Biophys Biomol Struct 23:167–192

von Heijne G, Gavel Y (1988) Topogenic signals in integral membrane proteins. Eur J Biochem 174:671–678

Wang HW, Chen Y, Yang H, Chen X, Duan MX, Tai PC, Sui SF (2003) Ring-like pore structures of SecA: implication for bacterial protein-conducting channels. Proc Natl Acad Sci USA 100:4221–4226

Wenk MR (2005) The emerging field of lipidomics. Nat Rev Drug Discov 4:594–610

Yang YB, Lian J, Tai PC (1997) Differential translocation of protein precursors across SecY-deficient membranes of *Escherichia coli*: SecY is not obligatorily required for translocation of certain secretory proteins in vitro. J Bacteriol 179:7386–7393

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40