A model for hydrophobic protrusions on peripheral membrane proteins

Edvin Fuglebakk[1,2†] **and Nathalie Reuter**[1,2,*]

[1]Computational Biology Unit, University of Bergen, University of Bergen, Pb7803, 5020 Bergen, Norway; [2]Department of Molecular Biology, University of Bergen, Pb7803, 5020 Bergen, Norway

***For correspondence:**
Nathalie.Reuter@cbu.uib.no (NR)

Present address: †Institute of Marine Research, Norway

Abstract With remarkable spatial and temporal specificities, peripheral membrane proteins bind to biological membranes. Prototypical peripheral membrane binding sites display a combination of patches of basic and hydrophobic amino acids that are also frequently present on other protein surfaces. The purpose of this contribution is to identify simple but essential components for membrane binding, through structural criteria that distinguish exposed hydrophobes at membrane binding sites from those that are frequently found on any protein surface. We formulate the concepts of *protruding hydrophobes* and *co-insertability* and have analysed more than 300 families of proteins that are classified as peripheral membrane binders. We find that this structural motif strongly discriminates the surfaces of binding and non-binding proteins. Our model constitute a novel formulation of a structural pattern for membrane recognition and emphasizes the importance of subtle structural properties of hydrophobic membrane binding sites.

Introduction

Biological membranes are ancient and crucial components in the organisation of life. Not only do they define the boundaries of cells and organelles, but they are central to a myriad of protein-protein and protein-lipid interactions. These encounters are instrumental for processes such as cell signalling (***Kutateladze, 2010***; ***Vögler et al., 2008***) and trafficking (***Cullen, 2008***), or regulation of cell structure and morphology (***Inaba et al., 2016***; ***Itoh et al., 2005***). Any attempt at understanding biological systems thus needs to incorporate protein-membrane interactions. A range of proteins has evolved to facilitate and regulate these processes. Besides the embedded transmembrane proteins and receptors, a number of soluble proteins interact transiently with the surface of cellular and organellar membranes achieving remarkable spatial and temporal specificities. These proteins are referred to as peripheral proteins and their membrane-binding site as interfacial binding site or IBS. Peripheral proteins include well-known lipid-binding domains that confer larger proteins the ability to bind membranes (***Lemmon, 2008***; ***Cho and Stahelin, 2005***). Other domains such as lipid-processing enzymes, endogenous or secreted by pathogens are also included in this definition. Advances in lipidomics that are now allowing large-scale mapping of protein-lipid interactions have already revealed novel lipid-interacting proteins (***Gallego et al., 2010***) suggesting that the current list of membrane-binding domains, and by extension of peripheral proteins, is not complete. An increased understanding and better characterization of membrane-protein interfaces is much needed for improved annotation of peripheral proteins as it would for example, ease the endeavor of lipidomics or transcriptomics initiatives. Efforts in drug development are also dependent on detailed structural characterization of such interfaces.

Unlike protein-protein or protein-ligand interactions, interfacial binding sites of peripheral proteins are poorly characterized in terms of amino acid composition and structural patterns.

Embedded and transmembrane proteins contain well defined regions of hydrophobic surface, clearly identifying their membrane interacting segments. This is seldomly the case for peripheral membrane proteins even though some have a fairly easily identifiable lipid binding pocket e.g. FYVE or some PH domains that bind preferentially phosphoinositides. Yet the majority of peripheral proteins do not belong to this category. Attempts to characterize the energetics of membrane binding has mostly focused on electrostatic complementarity with the head group charges of membrane lipids (***Mulgrew-Nesbitt et al., 2006***), rather than on the desolvation of hydrophobes which is more difficult to isolate in theoretical treatments. The preference of surface-exposed hydrophobic amino acids for the hydrophobic core of the membrane is indeed a result of their unfavorable interaction with solvent water, and is a consequence of the hydrophobic effect. The predictive power of implicit membrane models in the prediction of membrane binding sites has been a strong indication of the importance of the hydrophobic effect (***Lazaridis, 2003***). Lomize et al. could for example correctly predict membrane inserted residues of 53 peripheral proteins and peptides using a model that include only hydrophobic interactions, desolvation energy of polar groups and ionization energy (***Lomize et al., 2007***). In order to assert the generality of such binding mechanisms, it is however not only necessary to demonstrate the precense of the relevant amino acid types on known binding sites. It is also important to carefully analyse non-binding surfaces as well. Since they are soluble their interfacial binding site (IBS) is restricted in terms of the size of the hydrophobic patches they expose to their surface. The prototypical peripheral membrane binding sites display a combination of basic and hydrophobic amino acids. However, as both small hydrophobic patches and charged residues are frequently present on protein surfaces, it is challenging to distinguish membrane binding sites from the rest of the peripheral membrane proteins surface solely relying on amino acid composition.

For the hydrophobic component of binding sites, there is some evidence that structural considerations may allow signatures of membrane interacting hydrophobes to be defined. Terms like *hydrophobic spikes* (***Gilbert et al., 2002***; ***Gamsjaeger et al., 2005***) and *protruding loops* (***Lomize et al., 2007***) have been used to describe membrane binding sites, prompting the idea of hydrophobes protruding from the protein globule. A close look at amphipathic helices, also motivates the concept of protruding hydrophobes. Amphipathic helices are characteristic of membrane-binding peptides and proteins. When such membrane binding helices form, they are often found lining a protein, forming a cylindrical protrusion from the globule (e.g. ENTH domain of Epsin, PDBID: 1H0A (***Ford et al., 2002***)). Yet, no generalization of protruding membrane binding sites has been proposed for peripheral membrane proteins.

The purpose of this contribution is to identify structural criteria that distinguishes exposed hydrophobes at membrane binding sites from those that are frequently found on any protein surface. We propose a simple definition that formalizes the concept of protruding hydrophobes, and which can be easily computed from the protein structure. This definition allows us to systematically investigate to what extent protruding hydrophobes are found on both binding and non-binding surfaces, and to identify structural criteria for recognizing exposed hydrophobes that are likely to be important for membrane binding.

A major obstacle in developing general association models for peripheral membrane proteins is the scarcity of experimentally verified binding sites, and detailed descriptions of binding orientations. So far, computational studies on the role of hydrophobes on membrane binding sites have been based on relatively small sets of proteins with known binding sites (***Lomize et al., 2007***; ***Balali-Mood et al., 2009***; ***Lazaridis, 2003***). To get around this problem and to leverage the large number of proteins for which membrane binding has been identified without a detailed characterisation of the IBS, we perform a comparative statistical analysis of protein surfaces. Given a classification of proteins that separates membrane binders from non-binders, we compare peripheral membrane proteins with non-binding reference surfaces. With this we can extend our analysis to hundreds of protein families, rather than the few dozens for which binding sites have been partially identified by experiments.

With our simple definition of structural protrusions, we perform a statistical analysis of protruding hydrophobes in a large protein structure dataset and our results support their general role in membrane association. We find that protruding hydrophobes can be used to strongly discriminate protein surfaces invovled in membrane binding from those that are not. Hydrophobes are much more frequent on protruding sites of peripheral membrane proteins than in the reference dataset, and that they have a strong tendency to cluster on positions that can simultaneously interact with the membrane. We also derive membrane binding site predictors that are highly indicative of both experimentally identified membrane binding residues, and binding orientations predicted by other computational models. Even if we have delibaretely isolated the hydrophobic component of bindings sites, ignoring clearly important contributions from electorstatics and conformational flexibility, we find protruding hydrophobes to be a distinct signature of peripheral membrane proteins, and estimate that they are sufficient to identify binding sites in at least half of the 326 protein families we have analysed.

Results and Discussion

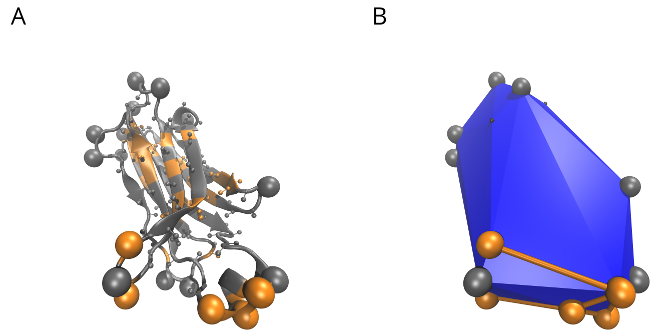


Figure 1. The definitions of *protrusions* and *co-insertable protruding hydrophobes*. Panel A show a cartoon representation of the C2 domain of human phospholipase A_2 (PDB ID: 1RLW), and panel B show the convex hull for the same protein. All C_α- and C_β-atoms are shown as spheres. Hydrophobes are coloured orange. The convex hull for the C_α- and C_β-atomic coordinates is shown in blue. All spheres visible on the convex-hull representation are vertex residues. *Protrusions* are defined as vertex residues with low local protein density, and shown as large spheres. *Co-insertable protruding hydrophobes* are protruding hydrophobes that are adjacent vertices of the convex hull, they are shown connected by orange lines. Small black spheres are vertex residues that have high local density, and do therefore not meet the criteria for protrusions.

Our formalisation of the concept of protruding amino acids is illustrated in Figure 1 and described in details in *Materials and Methods*. In short, it relies on firstly identifying the convex hull (in blue in Figure 1) of a coarse-grained protein model consisting of only C_α- and C_β-atoms. We then identify amino acids located at vertices of the convex hull which intuitively are good candidates to be inserted into a membrane without inserting other residues, and without deforming the protein backbone. The model thus implicitly assumes that (1) proteins interact with the membrane without appreciable conformational change, or prior to such change and (2) that the membrane is locally flat, which is a valid approximation in most cases. In order to single out the amino acids that are most exposed to solvent, we single out amino acids (vertices) in regions of low protein density, characterized as having a low number of neighboring atoms. Solvent accessibility is a necessary condition for the hydrophobic effect to contribute to binding. In addition, regions of low local protein density are also likely to cause less disruption of lipid packing upon membrane insertion.

In what follows, we present results of the application of this model to characterise hydrophobic properties of protrusions in peripheral membrane proteins. We do this by comparing a dataset of peripheral membrane proteins and a reference set of protein surfaces not interacting with the membrane, as described in *Materials and Methods*.

Protruding hydrophobes in a dataset of peripheral membrane proteins

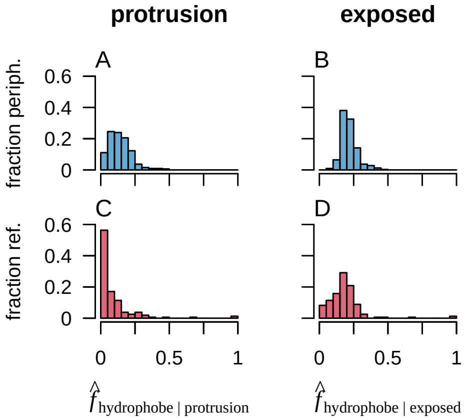


Figure 2. Hydrophobes are more common on protruding positions in peripheral proteins, than in the reference set. The plots show frequencies of hydrophobes on surface amino acids, both on protrusions (A and C) and among all solvent exposed amino acids (B and D). Compare peripheral proteins (blue) and the reference set (red). The horizontal axes show the mean fraction (Eq.1) of protrusions or solvent exposed amino-acids that are hydrophobic. The vertical axis shows the fraction of protein families.

First we calculated the frequency of hydrophobes on protrusions in peripheral proteins families and compared it to the reference dataset. In Figure 2, we observe a stark contrast between the set of peripheral proteins and the reference set (Figures 2 A and 2 C). Hydrophobes occur with high frequency and in almost all families on protrusions of peripheral proteins. In the reference set on the other hand, hydrophobes on protrusions are much less tolerated, reflected by a histogram mode of zero. This trend is specific for protruding positions, and does not reflect a general difference in composition of surface exposed amino-acids between the data sets as shown by plots in Figures 2 B and 2 D. Indeed if we consider the frequency of hydrophobes on all solvent exposed residues, the distributions look quite similar with both sets having histogram modes close to 0.2. This value is in agreement with the fraction of the surface of globular proteins typically reported to be hydrophobic (for instance 0.19 in Ref. (***Miller et al., 1987***)). The surfaces of the references set are in some cases very small, due to the way we ensure that these surfaces are not interacting with the membrane (see *Materials and Methods*). While these small surfaces are relevant samples for calculating average frequencies, the fraction of hydrophobes on such surfaces can take more extreme values (close to zero or 1). For this reason the tails of the histograms for the reference set are somewhat fatter than those for the peripheral membrane proteins.

Given the nature of our model the differences presented in Figure 2 are naturally ascribed to two factors; the accessibility of amino acids compared to other regions of the protein (they are vertices of the convex hull) and their low local protein density d defined as the number of neighbouring C_α- or C_β-atoms (Cf. definition in *Materials and Methods*). We here explore the dependence of this difference on d. In Figure 3 we show the difference between frequencies of hydrophobes in peripherals and reference data sets for different ranges of the local protein density d. The leftmost bar ($0 \leq d \leq 6$) corresponds to chain terminals. The other bars corresponding to ranges covered by our definition of protruding residues ($7 \leq d < 22$) show that hydrophobic residues are more frequently found at vertex residues with low local protein density in the peripheral proteins. This also serves as an *a posteriori* justification for constricting our definition of protrusions to amino-acids with $d < 22$.

Assuming that the over-representation of hydrophobes on protrusions in peripheral membrane proteins stems from actual membrane binding sites, one can expect those proteins to have more

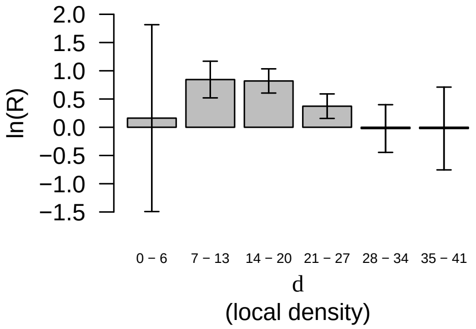


Figure 3. On peripheral proteins, protrusions in low density regions are more often hydrophobes, compared to the reference set. The plot shows the logarithm of the odds-ratio (Eq.10) comparing the frequency of hydrophobes on *vertex* residues in peripheral proteins and the reference set. Positive values reflect higher frequencies in the peripheral proteins. The horizontal axis shows the protein density d around the protrusion, measured as the number of C_α and C_β atoms within $1nm$. Vertex residues are all on the convex hull, but only the vertex residues with $d < 22$ are protrusions. The leftmost bar with $d < 7$ corresponds mostly to chain terminals. More precisely, the vertical axis shows $\mathrm{R}\left(A, B, \hat{F}_{\mathrm{hydrophobe|vertex}\cap l<d\leq u}\right)$ where A denotes the peripheral proteins, B the reference set, l and u denote the lower and upper limits of the ranges given on the vertical axis, and d is the local protein density defined in *Materials and Methods*. Error bars are 95% confidence intervals.

than one hydrophobic protrusion We estimated the tendency of each hydrophobic protrusion to be *co-insertable* by calculating the weighted frequency of co-insertion (Eq. 9) (Cf *Materials and Methods*) for both datasets (Figure 4). We note that peripheral membrane proteins do indeed tend to have hydrophobes on co-insertable protrusions to a significantly larger extent than what would be expected from randomly scattering hydrophobes among protruding positions. This tendency is much lower for the reference set, even when considering the extremities of the error bars, which are wide precisely because there are very few protruding hydrophobes in this set.

We further explore the degree of co-insertability of the hydrophobic protrusions present in our dataset of peripheral proteins and in the reference dataset. We seek to evaluate to what extent co-insertable hydrophobic protrusions can be used to discriminate likely peripheral membrane binders from other proteins. Figure 5 shows the fraction of proteins in each dataset that have at least one pair of co-insertable hydrophobic protrusions (labelled *Co-ins.*) and the fraction of proteins that have at least one *isolated* hydrophobic protrusion (i.e a protrusion that does not satisfy the criteria that define *co-insertability*). While we do see some discrimination between the data sets in the case of isolated protruding hydrophobes, the co-insertable ones prove to be very strong indicators of which proteins surfaces are membrane binding. As the coincidental occurrence of such properties increase with the size of the protein surface, we have grouped the proteins by total number of surface protrusions (regardless of hydropathic properties). We do however see no appreciable difference between the proteins of size $0 - 25$ and those of size $25 - 50$. We consider the fraction in the reference set to be a reasonable estimate of a false positive rate for predicting membrane binding function based on the presence of protruding hydrophobes. We find this false positive rate to be around 12% for co-insertable protrusions, in both the size ranges we have analysed (see Figure 5). For the peripheral membrane proteins, we estimate that 64% and 75% of the peripheral membrane proteins in the respective groups have co-insertable protruding hydrophobes. Assuming that such motifs occur by chance at a rate no higher than it does in the reference set, and that the over-representation is due to the membrane binding function defining the data sets, we conservatively estimate that co-insertable protruding hydrophobes are membrane-interacting motifs for more than half of the peripheral membrane proteins we have analysed.

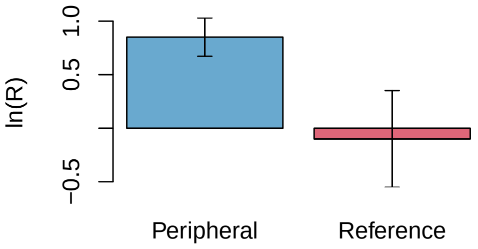


Figure 4. The *protruding hydrophobes* tend to be *co-insertable* in the peripheral proteins. The tendency for protrusions to be co-insertable is quantified by the weighted frequency of co-insertion (Eq. 9), and is compared between each data set and a null model using the odds ratio (Eq. 10). Positive values reflect higher frequencies of co-insertion than in the null model. More precisely, we show the comparisons $\mathrm{R}\left(set, null, \hat{F}^{\mathrm{pair}}_{\mathrm{one,both}}\right)$, where set represents the set of peripheral proteins (blue) and the reference set (red), and $null$ represent their respective null models where hydrophobes have been relocated randomly among protrusions as described in *Materials and Methods*. Error bars are 95% confidence intervals.

Protruding hydrophobes vs. experimentally verified membrane-binding sites

The analysis presented in Figures 3 and 5 suggests that the concepts of protruding hydrophobes and co-insertability can be used to identify membrane binding residues. Based on these results we seek to define a predictor of membrane binding sites. We define *the Likely Inserted Hydrophobe* as the protruding hydrophobe with the highest number of co-insertable protruding hydrophobes and lowest local protein density, as defined in *Materials and Methods*. Figure 6 illustrates that this simple definition is able to identify binding sites on modular membrane-binding domains: C1, C2, PX, ENTH, PLA2 and FYVE. For most of these cases, the Likely Inserted Hydrophobe has in fact been experimentally identified to contribute to membrane binding. For the other examples, it is clearly positioned close to the experimentally identified binding site. A more quantitative comparison between predicted and verified membrane interacting residues is complicated by the absence of negative assertions from either methods. Experiments aiming at identifying membrane-binding sites will usually only target some of the amino acids suspected to belong to the membrane binding residues, and usually not conclude on other amino acids. Similarly, the Likely Inserted Hydrophobe is by definition only one residue, and provide no negative prediction of which amino acids do not bind the membrane. We can however make a rough, but well defined, comparison by computing the angle between the vectors connecting the protein center with respectively; the mean position of the membrane interacting residues identified in experiments ($\mathbf{t}_{I_e}$), and the Likely Inserted Hydrophobe ($\mathbf{t}_{I_p}$, See Eq. 11). While this comparison does not provide a quantitative evaluation of whether experimentally determined IBS and predicted residues match exactly, it allows us to separate proteins where the predicted and verified residues are “on the same side” of the protein ($\angle \mathbf{t}_{I_e}\mathbf{t}_{I_p} < 90°$) from those where they are not. We show on Figure 7 such a comparison for proteins whose binding sites are experimentally determined. This is a coarse approximation to the protein orientation, which is sensitive to both protein shape, the selection of residues included in the partial biding sites, and any difference in backbone conformation between bound and unbound protein. Even so, we do expect that wrong binding site predictions should provide angles in the entire range from 0°to 180°with roughly uniform probability. But, we observe that almost all angles are sharper than 90°, indicating a reasonable agreement with experimental data. We also observe a similar range of angles for cases where the membrane interaction of the Likely Inserted Hydrophobe has been experimentally verified (marked with asterisks ($*$) in Figure 7) and the cases where it has not. We would like to emphasise at this point that the Likely Inserted Hydrophobes that are not yet found to be membrane interacting might very well never have been tested. We also calculated all angles between the set of experimentally identified residues and protruding amino acids of all

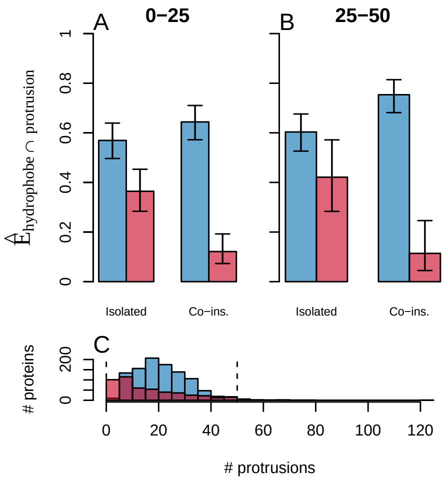


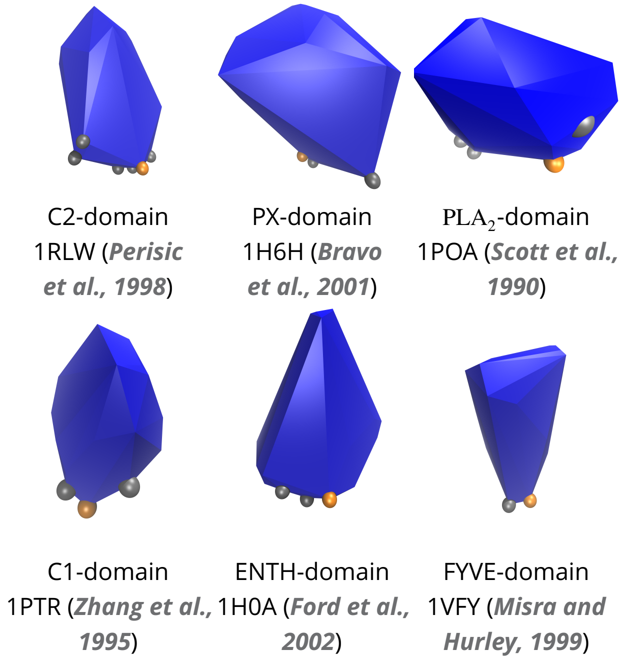
Figure 5. *Co-insertable protruding hydrophobes* are common in peripheral proteins and rare in the reference set. The plot shows the occurrence of *co-insertable protruding hydrophobes* on protein surfaces. Panels A and B show the weighted fraction (Eq. 5) of proteins that have protruding hydrophobes, in the peripheral proteins (blue) and the reference set (red). We have differentiated here between protrusions that have at least one co-insertable protruding hydrophobe (labeled "Co-ins."), and those that have not (labeled "isolated"). The analysis is done separately for two groups of proteins according to the total number of protrusions on the protein surface ($[0, 25\rangle$ in panel A, $[25, 50\rangle$ in panel B). Panel C shows the frequency distribution of the total number of protruding residues ("# protrusions") for all proteins. The selections analysed in panel A and B are found between the dashed lines in panel C. Error bars in panel A and B are 95% confidence intervals.

kinds. These results are displayed as box-plots in Figure 7. While they vary a bit between families, we note that all medians are close to $90°$, confirming that the statistical expectation for protrusions in general is to have roughly equally many observations larger than and smaller than $90°$.

We provide as Supporting Information the complete list of amino acids experimentally identified as being part of membrane binding sites (Table S2). It overlaps with the list provided by Lomize *et al.* (***Lomize et al., 2007***), but sometimes differ in exactly which amino acids are included, as we include membrane interacting residues even when they are not inserted in the hydrophobic core of the membrane.

Protruding hydrophobes on predicted membrane binding sites

The continuum-model presented by Lomize *et al.* (***Lomize et al., 2011a***) forms the basis for a systematic effort to predict binding orientations for peripheral membrane proteins. The OPM database (***Lomize et al., 2012***) provides prediction of spatial arrangements of membrane proteins with respect to the lipid bilayer for a selection of peripheral membrane proteins. We here investigate to what extent protruding hydrophobes are captured by the model proposed by Lomize *et al.*. We identify The Likely Inserted Hydrophobe for each of the proteins in our dataset, and extracts the OPM predicted insertion coordinate of its C_α-atom. The *insertion coordinate* of an atom measures its depth of insertion into the hydrocarbon region of the membrane model, and is thus positive for atoms located in the hydrocarbon core and negative for atoms located on either side of the membrane including the interfacial region (Cf. *Materials and Methods*). Figure 8 shows histograms of the median insertion coordinate of the Likely Inserted Hydrophobes identified in each family. A clear majority of those residues are located close to the interface of the membrane model in the OPM-predictions (Figure 8 A) and 75% of the families in the set of peripheral membrane proteins have the median insertion coordinate for the Likely Inserted Hydrophobe within a margin of 0.5 nm from the membrane. This fraction is similar to the estimated fraction of proteins that have co-insertable protruding hydrophobes (Figures 5 A and B). We allow this margin of 0.5 nm to compensate for the assumptions of rigid protein, flat membrane, and the distance between C_α-atoms and side-chain



C2-domain 1RLW (***Perisic et al., 1998***)

PX-domain 1H6H (***Bravo et al., 2001***)

PLA_2-domain 1POA (***Scott et al., 1990***)

C1-domain 1PTR (***Zhang et al., 1995***)

ENTH-domain 1H0A (***Ford et al., 2002***)

FYVE-domain 1VFY (***Misra and Hurley, 1999***)

Figure 6. Protruding hydrophobes are found on the membrane binding sites of well known membrane binding domains. The figure shows the convex hull (in blue) of the C_α and C_β-atoms of selected peripheral membrane binding domains. The C_β-atoms of *the Likely Inserted Hydrophobe* are shown as orange spheres and C_β-atoms of experimentally identified membrane-binding residues as gray spheres. The Likely Inserted Hydrophobe is an amino acid that has been experimentally verified to be a membrane binding residue for 1RLW, 1H6H, 1PTR and 1VFY. For 1H0A and 1POA the Likely Inserted Hydrophobe is located in the same area as the residues identified by experiments. **1RLW**: C2 domain of human phospholipase A2; **1H6H**: PX domain of P40PHOX ; **1POA**: snake phospholipase A2; **1PTR**: C1 domain of protein kinase C delta; **1H0A**: Epsin ENTH domain ;**1VFY**: FYVE domain of yeast vacuolar protein sorting-associated protein 27.

atoms. Fractions for other margins can be read from the cumulative histogram shown in Figure 8 C. By representing position with the insertion coordinate, we effectively project residue coordinates onto the membrane normal. We therefore do not expect surface amino acids to be uniformly distributed along the insertion coordinate axis and present control statistics for randomly chosen protruding amino acids of all hydropathic properties (Figure 8 B and 8 D). It appears clearly that the high number of Likely Inserted Hydrophobes close to the membrane model is not an effect of it simply being more protein there.

Structure and amino acid composition at hydrophobic protrusions

The analysis presented in Figure 3 indicates that the ability to discriminate the data sets based on the frequency of hydrophobes on protrusions gets lower as the local protein density gets higher. Local protein density of a protrusion is dependent on secondary structure elements with loops, turns and bends being those that intuitively favor low local protein density. These secondary structures typically mark a clear change in direction of the backbone trace, where the neighbouring residues ’make way’ for the protruding hydrophobe. Figure 9 A shows which secondary structure elements the protruding hydrophobes are associated to in the set of peripheral proteins. We note that loops, turns and bends are indeed abundant, but also helices and not beta-strands. Figure 9 B shows a comparison with the reference data set. We see that protruding hydrophobes on turns and bends are not only common in the peripheral membrane proteins as we saw in Figure 9 A, but they are also significantly more frequent than in the reference set. Interestingly, this is not the case for loops. A reason for this might be that turns and bends provide a rigid scaffold for exposing the hydrophobes, which would otherwise rearrange to desolvate when exposed to solvent, and thereby

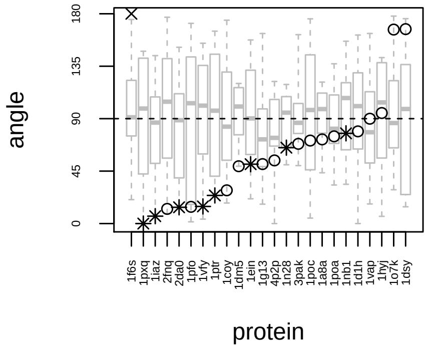


Figure 7. Protruding hydrophobes predict experimentally verified binding sites. The figure shows comparisons of predicted binding residues (*the Likely Inserted Hydrophobe*) with experimentally verified binding sites for a manually curated dataset of 24 proteins (listed in Table S2). The vertical axis corresponds to values of the angle (Eq.11) comparing the two vectors connecting the center of the protein with either the predicted or known binding sites. Smaller angles imply better agreement between prediction and experiment. Asterisks ($*$) mark proteins where the Likely Inserted Hydrophobe is an amino acid experimentally identified to be interacting with the membrane. The grey boxplots show the distribution of angles when the known binding site residues are compared to all protruding amino acids on the protein. 1iaz is analysed in its soluble monomeric state, while it forms a transmembrane pore upon oligomerisation. The structure of the C-type lysozyme (PDBID 1f6s) has no identified protruding hydrophobes and is marked with a cross at 180°. Interestingly, while our analysis is performed on its crystallised form, it is known to bind membranes in a molten globule state.

likely reduce the free-energy gain of membrane insertion. As the definition of *loop* here is simply absence of any of the other secondary structure definitions, we would expect this category to contain less regular, more flexible structures. We also expect this property of rigid scaffolding from amphipathic helices, which is an established motif for membrane association. Figure 9 illustrates however that protrusions are not dominantly helices, confirming that the concept of protruding hydrophobes provides a useful generalisation for the shapes of membrane-binding sites.

For purposes of isolating the structural component of hydrophobic membrane association, we have until now used a dichotomous definition of hydrophobicity based on the Wimley-White scale for interfacial insertion (***Wimley and White, 1996***). Yet, we do expect different amino acids to have varying contributions to the free energy of binding. We have therefore also assessed the relative importance of different amino acids for discriminating between our two data sets. Figure 10 B shows the comparison of the frequencies of different hydrophobic amino acids on protrusions in the two data sets. As expected, we find non-polar residues with large aliphatic or aromatic side chains to be much more frequent at the protrusions of peripheral proteins than in the reference data set. While the error bars in Figure 10 B are not corrected for multiple testing, the signal for the hydrophobes as a group is quite clear. They all occur as over-represented in the set of peripheral proteins and the odds-ratio is much larger for phenylalanine, leucine and tryptophan than for any of the amino-acids that are over-represented in the reference set. Recall that $\ln R$ (Eq.10) is symmetric around 0, so the magnitude of the bar representing phenylalanine on one end, can be directly compared to that of the bar representing threonine in the negative direction. Tyrosine on the other hand discriminates the sets poorly compared to its high hydrophobicity score in the Wimley-White scale. We consider this a possible consequence of the orientational restrictions on the binding sites of peripheral membrane proteins. The typical orientations consistent with shallow binding, has the residue anchored above the membrane. This probably allows less freedom for the hydroxyl group of tyrosine to orient towards regions of higher water density, than it has in the peptides used for the Wimley-White experiments or in transmembrane proteins. We also note

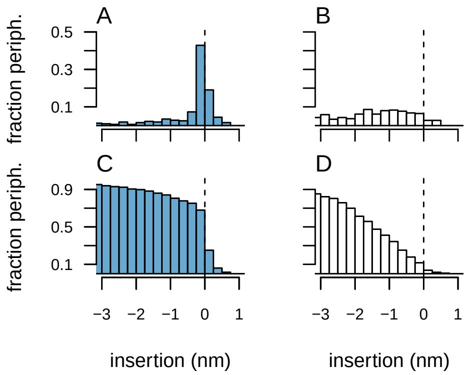


Figure 8. Comparing predictions based on protruding hydrophobes with the predicted IBS in the Orientation of Proteins in Membranes (OPM) database. The plots show the distributions of the median *insertion coordinate* from OPM for *the Likely Inserted Hydrophobe* in each family (measured at the C_α-atom). Values greater than or equal to zero correspond to atoms positioned in the hydrophobic core or at the boundary. Hence insertion coordinate values close to zero indicate agreement with OPM. Panel A (C) show data for the Likely Inserted Hydrophobes and panel B (D) for a null model of randomly selected *protruding* residues. Panel C and D show cumulative histograms (accumulated with decreasing insertion coordinates).

with interest that proline is among the residues that are somewhat over-represented in the set of peripheral proteins. In general, prolines are conformationally important protein components, that restricts the backbone with respect to its immediate neighbours along the peptide chain, and are therefore likely to promote local rigidity. They also serve to induce sharp changes in the backbone trace, which would facilitate solvent exposure of neighbouring side-chains, as discussed above. Specifically, they are in general frequently found on turns (***Wilmot and Thornton, 1988***).

Conclusion

Protein-membrane interactions are typically studied *in vitro* or *in silico* and inference to their biological context have to carry over from greatly simplified membrane models. To make sense of such experiments and simulations, it is essential to formulate general models that explain protein association in terms of factors that are present in both model systems and the relevant *in vivo* counterpart. In pursuit of such general models for membrane recognition, we have formulated the concepts of protruding hydrophobes and co-insertability. We have analysed more than 300 families of proteins that are classified as peripheral membrane binders and identified this model to be a good fit to more than half of them, after correcting for the small false positive rate estimated from the reference set (Figure 5). The generality of the model is corroborated by three important points. Hydrophobes are clearly over-represented on the protrusions of peripheral membrane proteins (compare Figure 2 A and 2 C, and see Figure 3), they tend to locate on co-insertable protrusions (see Figure 4 and Figure 5), and protruding hydrophobes are generally positioned consistent with experimentally identified binding sites (Figure 6 and Figure 7). Amphipathic helices are already well known membrane binding motifs which our definition of protrusion is well suited to capture, whenever these are stably folded and exposed. We do however find that the majority of identified protruding hydrophobes are not helices (Figure 9 A) and that hydrophobes are also highly over-represented on protruding turns and bends (Figure 9 B). We therefore propose the concept of protruding hydrophobes as a useful generalisation upon binding motifs that are identified in terms of secondary structure.

Both the choice of reference set, and the choice of quaternary structure modelling comes with some assumptions. We have elaborated on these in “materials and methods”. We have also

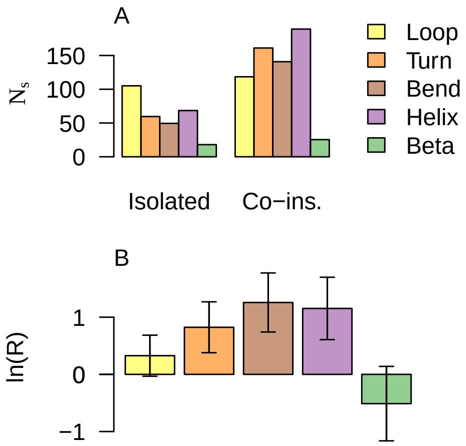


Figure 9. In peripheral proteins, hydrophobic protrusions are more frequent on turns, bends and α-helices, compared to the reference set. Panel A shows the weighted number (Eq. 2) of *protruding hydrophobes* associated with the different types of secondary structure elements. We have differentiated between protrusions that have at least one co-insertable protruding hydrophobe (right, labeled "Co-ins."), and those that have not (left, labeled "Isolated"). Panel B compares the weighted frequencies (Eq. 4) of hydrophobes on protruding secondary structures between the peripheral membrane proteins and the reference set, using the odds ratio (Eq. 10). Positive values reflect higher frequencies in the peripheral proteins. More precisely, panel A show the values $N_{\text{hydrophobe|protrusion}\cap sse}$, and panel B the comparisons $\mathrm{R}\left(A, B, \hat{F}_{\text{hydrophobe|protrusion}\cap sse}\right)$ where A denote the peripheral proteins, B the reference set, and sse specifies the secondary structures given in the color legend. Error bars in panel B are 95% confidence intervals.

performed some checks on how sensitive our analyses are to violations of these assumptions, and found that our conclusions are robust. We present details of these analyses as Supporting Information.

Investigation of the interfacial binding sites of numerous peripheral membrane proteins has revealed the presence of hydrophobic amino acids, and of basic amino acids such as arginines and lysines. This reflect the two almost universal traits of biological membranes; their hydrophobic core and anionic surface. Yet the focus on the electrostatic component of the free energy of transfer from water to membrane - often referred to as being long-range - has overshadowed the importance of hydrophobic contribution which is sometimes referred to as being short-range. The focus on electrostatic interaction is at least in part to be attributed to the difficulties in evaluating the hydrophobic contribution as opposed to for example, the computational tractability of continuum electrostatic models. In principle the contribution of hydrophobes to membrane binding can only be determined with a rigorous treatment of the hydrophobic effect, which requires very accurate treatment of large systems involving both protein, membrane and solvent. The mere presence of hydrophobes on the protein surface is to a large extent tolerated by non-binding soluble proteins as well, and for both hydrophobes and basic amino acids, it is challenging to determine when their presence on protein surfaces are coincidental, and when they are important for membrane binding. Moreover, amino acids on membrane binding sites are not typically strongly conserved (***Park et al., 2016***), so modeling their generic binding modes is important both for relating binding sites between homologs, and for understanding how additional factors determine differences in membrane specificities. Fortunately, as evident from the results presented in this contribution, the role of hydrophobes can often be understood in much simpler terms than what is required for an exact estimate of the energetics of the hydrophobic effect, and their importance for membrane-binding can be inferred from comparative statistical analyses. The subtle considerations of protein structure encoded in our definition of protrusions, strongly distinguishes the small hydrophobic patches on peripheral membrane proteins from those on other protein surfaces. This provides good reason

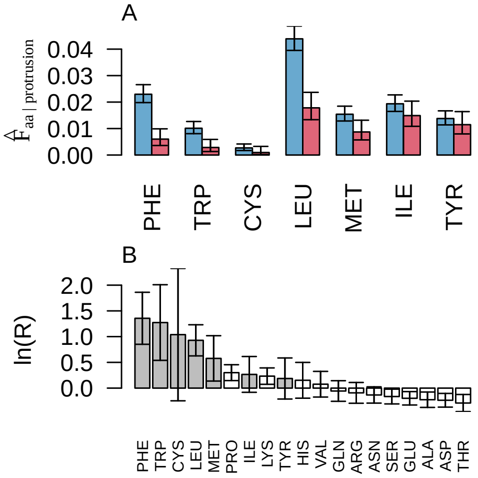


Figure 10. Large aliphatic and aromatic side chains are particularly over-represented on protrusion on peripheral proteins. Panel A shows the weighted fractions (Eq. 4) of hydrophobic amino acids on protrusions from peripheral proteins (blue) and from proteins in the reference set (red). In panel B, the contrast between the two sets is quantified by the odds ratio (Eq. 10), so that positive values reflect higher frequencies in the set of peripheral proteins than in the reference set. More precisely the vertical axis denote $\ln \mathrm{R}\left(\text{peripheral}, \text{reference}, \hat{F}_{aa,\text{protrusion}}\right)$, with *aa* representing each of the standard amino acids. Error bars are 95% confidence intervals.

to assume their importance for binding. Importantly, a minimalistic model such as the one we are proposing is an attempt at reducing membrane recognition to essential components. While a detailed understanding of the binding of individual proteins clearly requires treatment of both protein and membrane deformability, as well as the ever elusive solvent effects; our model assumes a rigid protein, a flat membrane and a dichotomous classification of hydrophobicity. It is therefore remarkable that in so many cases membrane recognition reduces to the simple idea of solvent-exposed hydrophobes protruding from the protein globule, ensuring that their desolvation will be energetically favorable upon transferring to a biological membrane.

Methods and Materials

Data sets

We obtained data sets from the collection of proteins in the OPM-database (***Lomize et al., 2012***). Our set of peripheral proteins are all proteins in OPM classified as *type*: *Monotopic/peripheral*. While the OPM has strict criteria for inclusion, membrane binding is not asserted by experiment in all cases, and the set might contain false positives.

The reference set consist of fragments of transmembrane complexes. We obtained these protein fragments from all proteins classified as *type*: *Transmembrane* in OPM. The fragments analysed are composed of all amino acids whose C_{α}-coordinates are at least 1.5 nm from the hydrocarbon region of the membrane model (The parameter Z_{HDC} in the OPM model (***Lomize et al., 2011b***)). We rely here on membrane models positioned by the OPM, which we deem very reliable for transmembrane proteins. While the entire protein complex was considered when calculating structural properties, only the fragments meeting this distance criteria were considered in the statistical analyses. When these proteins interact with secondary membranes or interact with membranes of extremely high curvature, it is not captured by the OPM model, and the assumption that these surfaces are not interacting with membrane may be violated. We have assumed that such issues are exceptional.

We do consider the assumptions mentioned above to be conservative. Inclusion of non-binding proteins in our set of peripheral membrane proteins would likely weaken any general signal from membrane binding proteins, and inclusion of secondary membrane interactions sites in the reference set would probably inflate the number of hydrophobes on protrusions in that set.

All protein structures are obtained by X-ray crystallography and NMR spectroscopy and we have assumed that at least the backbone coordinates are representative of the solvated state of the proteins. As the source of structural information for this database is the Protein Data Bank (PDB)(***Berman et al., 2000***) the relevant oligomeric state is not always determined, The curators of the OPM-database have decided on oligomer models, upon which we have relied. These are taken from PDBe (***Velankar et al., 2010***), generated by PISA (***Krissinel and Henrick, 2007***) or obtained from literature as described by Lomize *et al.* (***Lomize et al., 2012***). As weak protein-protein interaction interfaces may also contain exposed hydrophobic patches, we expect our analysis to be sensitive to how protein quaternary structure is modelled. As a quality control, we therefore also performed our analysis relying solely on computationally predicted quaternary structures, which we provide in the Supporting Information. This control reproduced qualitatively all observations that we have interpreted. In the Supporting Information we also report analysis on the sensitivity of the results to how the reference set is obtained, using a reference set based on the SCOPe-classification (***Fox et al., 2014***).

A few structures meeting the above criteria, were not included in the analysis for technical reasons, such as issues with formatting of PDB files. After exclusion of these cases, the final set of peripheral proteins contains 1012 protein structures classified into 326 families. The final reference set contains 495 protein structures classified into 158 families.

Based on experiments reported in available literature (***Hedin et al., 2002***; ***Grauffel et al., 2013***; ***Malmberg et al., 2003***; ***Stahelin et al., 2003a***, ***2002***; ***Wang et al., 2001***; ***Stahelin et al., 2003b***; ***Stahelin and Cho, 2001***; ***Frazier et al., 2003***; ***Gerber et al., 2002***; ***Rufener et al., 2005***; ***Corbalán-García et al., 2003***; ***Kim et al., 2000***; ***Gilbert et al., 2002***; ***Feng et al., 2002***, ***2003***; ***Grauffel et al., 2013***; ***Walther et al., 2004***; ***Kohout et al., 2003***; ***Goh et al., 2016***; ***B Campos et al., 1998***; ***Isas et al., 2004***; ***Kutateladze and Overduin, 2001***; ***Stahelin et al., 2002***; ***Wang et al., 2001***; ***Anderluh et al., 2005***; ***Shenkarev et al., 2006***; ***Lin et al., 1998***; ***Canaan et al., 2002***; ***Lathrop et al., 2001***; ***Chen et al., 2000***; ***Sekino-Suzuki et al., 1996***; ***Phillips et al., 2005***; ***Thennarasu et al., 2005***; ***Tatulian et al., 2005***; ***Oldham et al., 2005***; ***Mathias et al., 2009***; ***Agasøster et al., 2003***; ***Jian et al., 2015***), we made a data set of partially identified membrane binding sites on proteins with resolved structures. This set contains membrane interacting residues of 34 protein structures, classified into 22 families. A detailed description is provided in the Supporting Information (Table S2).

Definitions

Structural characteristics of protein surfaces

We characterise the surface of proteins with different criteria designed to capture solvent*exposed* residues, *protruding* residues and *co-insertable* protruding residues. The two latter are illustrated in Figure 1.

Exposed amino acids are defined as all amino acids that have a solvent accessible side-chain area greater than $0.2\ \mathrm{nm}^2$, as calculated with a probe with a radius of 0.14 nm, following the procedure described in Eisenhaber *et al.* (***Eisenhaber et al., 1995***) using van der Waals radi reported by Bondi (***Bondi, 1964***).

We identify a *protrusion* or a *protruding* residue via the calculation of the convex hull of the C_α- and C_β-coordinates of the protein. The convex hull of a set of points *S* is the smallest possible convex set containing *S*. We define *vertex* residues as residues whose C_β-atom is a vertex of this convex hull. A *protrusion* or a *protruding* residue, is defined as a *vertex* residue that also has low local protein density. For the purposes of this work, we will define the local protein density d of a residue, as the number of C_α- or C_β-atoms within a distance c of its C_β-atom. We will designate a local protein density as low, if $d < n$, with $n = 22$ and $c = 1\ \mathrm{nm}$. These parameters were manually chosen based on a set of six different families of peripheral membrane proteins (C2-domain, PX-domain, Discodin domain, ENTH domain, Lipoxygenases and a Bacterial Phospholipase C). A list of these proteins are provided as Supporting Information (Table S1).

We define two protrusions to be *co-insertable* or a *co-insertable pair*, if the straight line connecting them is an edge of the convex hull polygon.

Hydrophobic residues

An amino acid is defined to be *hydrophobic*, or a *hydrophobe*, if it contributes favourably to membrane interface partitioning of peptides, as determined in the Wimley-White scale for interfacial insertion (**Wimley and White, 1996**). These amino acids are: leucine, isoleucine, phenylalanine, tyrosine, tryptophan, cysteine and methionine.

Secondary structure

We use DSSP definitions (**Kabsch and Sander, 1983**) for protein secondary structure. DSSP codes H, G or I are reported as *helix*, DSSP codes B or E as β, DSSP code T as *bend* and DSSP code S as *turn*. All other residues are considered to be in *loops*.

Likely Inserted Hydrophobe

The *Likely Inserted Hydrophobe* is defined as the protruding hydrophobe with the largest number of co-insertable protruding hydrophobes in a protein. Ties are resolved by choosing the likely inserted hydrophobe with the smallest local protein density d. Further ties are resolved by random selection, so that each protein has exactly one Likely Inserted Hydrophobe, unless it has no protruding hydrophobes at all.

Insertion coordinate

For comparisons with OPM predictions, we define the *insertion coordinate* of atoms. This coordinate measures how deeply into the OPM membrane model an atom is inserted, and is therefore negative on the solvated side of the membrane. The membrane perimeter, where the insertion coordinate is 0, is the end of the hydrocarbon region. We identify this boundary as it is done in the model used to predict the OPM orientations, namely the planes where the volume fraction of total hydrocarbon is equal to 0.5. See eq. 2 in (**Lomize et al., 2011b**).

Measures

Averages of residues

We compare protein surfaces with respect to structural and hydropathic properties, reflected in different selection criteria and averaged over families or the entire data sets.

The mean fraction of residues having property s with respect to a reference property r in a family is:

$$\hat{f}_{s|r} = \frac{1}{|C|} \sum_{G \in C} \frac{|G_s \cap G_r|}{|G_r|} \tag{1}$$

where C is the set of proteins in a family, G is a protein, and, G_s is the set of residues on a protein meeting criteria s. Vertical bars denote size of sets. We will specify s and r according to the definitions above, using intersect notation to combine criteria when necessary. $\hat{f}_{\text{hydrophobe}|\text{protrusion}\cap\text{helix}}$, for instance, should be interpreted as the mean fraction of hydrophobes out of all protruding amino acids that are in helices.

We estimate weighted data set counts of amino acids with property s as:

$$\hat{N}_s = \sum_{C \in D} \left(\frac{1}{|C|} \sum_{G \in C} |G_s| \right) \tag{2}$$

where D is a data set, such as the set of peripheral proteins or the reference set. Similarly we quantify the weighted count of proteins that have at least one amino acid with property s as:

$$\hat{M}_s = \sum_{C \in D} \left(\frac{1}{|C|} \sum_{G \in C} \mathrm{H}\left(|G_s|\right) \right) \tag{3}$$

where H is the Heaviside step function. Given a property s and reference property r, we estimate the weighted fraction in a data set, $\hat{F}_{s|r}$:

$$\hat{F}_{s|r} = \frac{\hat{N}_{s\cap r}}{\hat{N}_r} \tag{4}$$

or the weighted fraction of proteins that have at least one residue with the given property s:

$$\hat{E}_s = \frac{\hat{M}_s}{|D|} \tag{5}$$

With $|D|$ being the number of families in the data set. When such fractions (Eqs. 4 or 5) are reported, we estimate 95%-confidence intervals using a normal approximation to the binomial distribution, with $|D|$ the total number of trials (Eq. 5), or $\hat{N}_r$ serving as a real-number analog to the total number of trials (Eq. 4).

Averages of co-insertable pairs

To analyse co-insertable residues, we estimate weighted data set counts of co-insertable pairs of residues with property s, as:

$$\hat{N}_s^{\text{pair}} = \sum_{C\in D}\left(\frac{1}{|C|}\sum_{G\in C}\left|G_s^{\text{pair}}\right|\right) \tag{6}$$

where $\left|G_s^{\text{pair}}\right|$ are the number of co-insertable amino acids pairs with property s. For quantification of the weighted count of proteins that have at least one co-insertable pair with property s, we calculate:

$$\hat{M}_s^{\text{pair}} = \sum_{C\in D}\left(\frac{1}{|C|}\sum_{G\in C} H\left(\left|G_s^{\text{pair}}\right|\right)\right) \tag{7}$$

Considering the set of co-insertable amino acid pairs in a protein, G^{pair}, we will denote the set of pairs where at least one of the amino acids is a protruding hydrophobe as $G_{\text{one}}^{\text{pair}}$, and the set where both are protruding hydrophobes as $G_{\text{both}}^{\text{pair}}$. We will report the weighted fraction of proteins that have co-insertable protruding hydrophobes as:

$$\hat{E}_{\text{both}}^{\text{pair}} = \frac{\hat{M}_{\text{both}}^{\text{pair}}}{|D|} \tag{8}$$

and the weighted frequency of co-insertion of protruding hydrophobes as:

$$\hat{F}_{\text{both}|\text{one}}^{\text{pair}} = \frac{\hat{N}_{\text{both}}^{\text{pair}}}{\hat{N}_{\text{one}}^{\text{pair}}} \tag{9}$$

Note that $\hat{F}_{\text{both}|\text{one}}^{\text{pair}}$ estimates the conditional probability that both amino acids of a co-insertable pair are protruding hydrophobes, given that one of them is. The tendency for protruding hydrophobes to be located at co-insertable positions can then be quantified by comparing with a null model for each set. We obtain these null models by randomly reassigning the hydrophobic amino acids to other protruding locations in the same protein.

Comparison between data sets

The frequency of properties in different data sets, are compared via weighted fractions. For two data sets, A and B, we compare a certain weighted fraction $\hat{F}$ using the odds ratio, $\mathrm{R}\left(A, B, \hat{F}\right)$:

$$\mathrm{R}\left(A, B, \hat{F}\right) = \frac{\hat{F}^A\left(1-\hat{F}^B\right)}{\hat{F}^B\left(1-\hat{F}^A\right)} \tag{10}$$

where $\hat{F}^A$ denotes the fraction $\hat{F}_{s|r}$ obtained for data set A. We will report $\ln\mathrm{R}$, which is symmetric around 0, so that $\ln\mathrm{R}\left(A, B, \hat{F}\right) = -\ln\mathrm{R}\left(B, A, \hat{F}\right)$. Wald 95%-confidence intervals for $\ln\mathrm{R}$ are calculated with $\hat{N}_{s\cap r}$ and $\left(\hat{N}_r - \hat{N}_{s\cap r}\right)$ serving as real number analogs for the count of successes and failures in the data sets compared. When $\hat{F}_{\text{both}|\text{one}}^{\text{pair}}$ is compared, the corresponding counts of successes and failures are $\hat{N}_{\text{both}}^{\text{pair}}$ and $\hat{N}_{\text{both}}^{\text{pair}} - \hat{N}_{\text{one}}^{\text{pair}}$, respectively.

Comparison of experimentally verified and predicted binding sites

We define two vectors which we then compare to evaluate the distance between experimentally verified and predicted membrane binding residues. The C_α-coordinate of experimentally verified membrane binding residues functions as a proxy for the membrane, and the vector defined by the latter residues and the center of mass (COM) of the protein is used as a reference to which we compare the vector defined by the protein COM and the Likely Inserted Hydrophobe. Given a set of identified or predicted membrane interacting resides, I, we compute the vector, $\mathbf{t}_I$:

$$\mathbf{t}_I = \frac{1}{|I|}\sum_{a\in I}\mathbf{V}_a - \frac{1}{|G_*|}\sum_{a\in G_*}\mathbf{V}_a \tag{11}$$

where $\mathbf{V}_a$ denotes the C_α-coordinates of residue a, and G_* is the set of all residues in the protein. We will denote vectors obtained for experimentally identified membrane binding residues as $\mathbf{t}_{I_e}$, and those obtained for a Likely Inserted Hydrophobe as $\mathbf{t}_{I_p}$. We then measure the angle $\angle\mathbf{t}_{I_e}\mathbf{t}_{I_p}$ between the two vectors for each protein in the dataset of known binding sites.

Implementation

The solvent accessible area was calculated with MMTK (***Hinsen, 2000***) (version 2.9.0), and the convex hull was calculated with Qhull (***Barber et al., 1996***) via scipy (***Jones et al., 2001***) (version 0.13.3). Proportion test confidence intervals were calculated with R (***Team, 2008***) (Version 2.12.0), odds ratios and corresponding confidence intervals were calculated with the R-package epitools (***Aragon, 2010***) (version 0.5-6). Secondary structure annotations were computed with the CMBI DSSP implementation (***Touw et al., 2015***) (version 2.0.4). Otherwise the analyses were implemented by us, using Python and R. Plots were produced with R, and other visualisations using VMD (Visual Molecular Dynamics) (***Humphrey et al., 1996***).

Acknowledgments

We thank Anne Gershenson at the University of Massachusetts Amherst, and Angèle Abboud at the University of Bergen for their valuable comments. This work was supported by grants from the Norwegian Research Council (FRIMEDBIO 214167 and FRIMEDBIO 251247).

References

Agasøster AV, Halskau Ø, Fuglebakk E, Frøystein NA, Muga A, Holmsen H, Martinez A. The interaction of peripheral proteins and membranes studied with alpha-lactalbumin and phospholipid bilayers of various compositions. J Biol Chem. 2003 Jun; 278(24):21790–21797.

Anderluh G, Razpotnik A, Podlesek Z, Macek P, Separovic F, Norton RS. Interaction of the eukaryotic pore-forming cytolysin equinatoxin II with model membranes: 19F NMR studies. J Mol Biol. 2005 Mar; 347(1):27–39.

Aragon T. epitools: Epidemiology Tools. R package. . 2010; .

B Campos, Y D Mo, T R Mealy, C W Li, M A Swairjo, C Balch, J F Head, G Retzinger, Dedman JR, B A Seaton. Mutational and crystallographic analyses of interfacial residues in annexin V suggest direct interactions with phospholipid membrane components. Biochemistry. 1998 Jun; 37(22):8004–8010.

Balali-Mood K, Bond PJ, Sansom MSP. Interaction of monotopic membrane enzymes with a lipid bilayer: A coarse-grained MD simulation study †. Biochemistry. 2009 Mar; 48(10):2135–2145.

Barber CB, Dobkin DP, Huhdanpaa H. The quickhull algorithm for convex hulls. ACM T Math Software. 1996 Dec; 22(4):469–483.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000 Jan; 28(1):235–242.

Bondi A. van der Waals volumes and radii. J Phys Chem. 1964; 68(3):441–451.

Bravo J, Karathanassis D, Pacold CM, Pacold ME, Ellson CD, Anderson KE, Butler PJ, Lavenir I, Perisic O, Hawkins PT, Stephens L, Williams RL. The crystal structure of the PX domain from p40(phox) bound to phosphatidylinositol 3-phosphate. Mol Cell. 2001 Oct; 8(4):829–839.

Canaan S, Nielsen R, Ghomashchi F, Robinson BH, Gelb MH. Unusual mode of binding of human group IIA secreted phospholipase A2 to anionic interfaces as studied by continuous wave and time domain electron paramagnetic resonance spectroscopy. J Biol Chem. 2002 Aug; 277(34):30984–30990.

Chen X, Wolfgang DE, Sampson NS. Use of the parallax-quench method to determine the position of the active-site loop of cholesterol oxidase in lipid bilayers. Biochemistry. 2000 Nov; 39(44):13383–13389.

Cho W, Stahelin RV. Membrane-protein interactions in cell signaling and membrane trafficking. Annu Rev Biophys Biomol Struct. 2005 Jun; 34(1):119–151.

Corbalán-García S, Sánchez-Carrillo S, García-García J, Gómez-Fernández JC. Characterization of the membrane binding mode of the C2 domain of PKCϵ †. Biochemistry. 2003 Oct; 42(40):11661–11668.

Cullen PJ. Endosomal sorting and signalling: an emerging role for sorting nexins. Nat Rev Mol Cell Biol. 2008 Jul; 9(7):574–582.

Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. J Comput Chem. 1995; 16(3):273–284.

Feng J, Bradley WD, Roberts MF. Optimizing the interfacial binding and activity of a bacterial phosphatidylinositol-specific phospholipase C. J Biol Chem. 2003 Jul; 278(27):24651–24657.

Feng J, Wehbi H, Roberts MF. Role of tryptophan residues in interfacial binding of phosphatidylinositol-specific phospholipase C. J Biol Chem. 2002 May; 277(22):19867–19875.

Ford MGJ, Mills IG, Peter BJ, Vallis Y, Praefcke GJK, Evans PR, McMahon HT. Curvature of clathrin-coated pits driven by epsin. Nature. 2002 Sep; 419(6905):361–366.

Fox NK, Brenner SE, Chandonia JM. SCOPe: Structural classification of proteins–extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res. 2014 Jan; 42(Database issue):D304–9.

Frazier AA, Roller CR, Havelka JJ, Hinderliter A, Cafisco DS. Membrane-bound orientation and position of the synaptotagmin I C2A domain by site-directed spin labeling. Biochemistry. 2003; 42:96–105.

Gallego O, Betts MJ, Gvozdenovic-Jeremic J, Maeda K, Matetzki C, Aguilar-Gurrieri C, Beltran-Alvarez P, Bonn S, Fernández-Tornero C, Jensen LJ, Kuhn M, Trott J, Rybin V, Müller CW, Bork P, Kaksonen M, Russell RB, Gavin AC. A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. Mol Syst Biol. 2010 Nov; 6:430.

Gamsjaeger R, Johs A, Gries A, Gruber HJ, Romanin C, Prassl R, Hinterdorfer P. Membrane binding of β 2-glycoprotein I can be described by a two-state reaction model: an atomic force microscopy and surface plasmon resonance study. Biochem J. 2005 Aug; 389(3):665–673.

Gerber SH, Rizo J, Sudhof TC. Role of electrostatic and hydrophobic interactions in Ca2+-dependent phospholipid binding by the C2A-domain from synaptotagmin I. Diabetes. 2002 Feb; 51(Supplement 1):S12–S18.

Gilbert GE, Kaufman RJ, Arena AA, Miao H, Pipe SW. Four hydrophobic amino acids of the factor VIII C2 domain are constituents of both the membrane-binding and von Willebrand factor-binding motifs. J Biol Chem. 2002 Feb; 277(8):6374–6381.

Goh BC, Wu H, Rynkiewicz MJ, Schulten K, Seaton BA, McCormack FX. Elucidation of lipid binding sites on lung surfactant protein A using X-ray crystallography, mutagenesis, and molecular dynamics simulations. Biochemistry. 2016 Jul; 55(26):3692–3701.

Grauffel C, Yang B, He T, Roberts MF, Gershenson A, Reuter N. Cation-π interactions as lipid-specific anchors for phosphatidylinositol-specific phospholipase C. J Am Chem Soc. 2013 Apr; 135(15):5740–5750.

Hedin EMK, Høyrup P, Patkar SA, Vind J, Svendsen A, Fransson L, Hult K. Interfacial orientation of thermomyces lanuginosaLipase on phospholipid vesicles investigated by electron spin resonance relaxation spectroscopy. Biochemistry. 2002 Dec; 41(48):14185–14196.

Hinsen K. The molecular modeling toolkit: A new approach to molecular simulations. J Comput Chem. 2000; 21(2):79–85.

Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996 Feb; 14(1):33–38.

Inaba T, Kishimoto T, Murate M, Tajima T, Sakai S, Abe M, Makino A, Tomishige N, Ishitsuka R, Ikeda Y, Takeoka S, Kobayashi T. Phospholipase Cβ1 induces membrane tubulation and is involved in caveolae formation. Proc Natl Acad Sci USA. 2016 Jul; 113(28):7834–7839.

Isas JM, Langen R, Hubbell WL, Haigler HT. Structure and dynamics of a helical hairpin that mediates calcium-dependent membrane binding of annexin B12. J Biol Chem. 2004 Jul; 279(31):32492–32498.

Itoh T, Erdmann KS, Roux A, Habermann B, Werner H, De Camilli P. Dynamin and the actin cytoskeleton cooperatively regulate plasma membrane invagination by BAR and F-BAR proteins. Dev Cell. 2005 Dec; 9(6):791–804.

Jian X, Tang WK, Zhai P, Roy NS, Luo R, Gruschus JM, Yohe ME, Chen PW, Li Y, Byrd RA, Xia D, Randazzo PA. Molecular basis for cooperative binding of anionic phospholipids to the PH domain of the Arf GAP ASAP1. Structure. 2015 Nov; 23(11):1977–1988.

Jones E, Oliphant E, Peterson P. SciPy: Open Source Scientific Tools for Python. . 2001; .

Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec; 22(12):2577–2637.

Kim SW, Quinn-Allen MA, Camp JT, Macedo-Ribeiro S, Fuentes-Prior P, Bode W, Kane WH. Identification of functionally important amino acid residues within the C2-domain of human Factor V using alanine-scanning mutagenesis . Biochemistry. 2000 Feb; 39(8):1951–1958.

Kohout SC, Corbalán-García S, Gómez-Fernández JC, Falke JJ. C2 domain of protein kinase Cα: Elucidation of the membrane docking surface by site-directed fluorescence and spin labeling. Biochemistry. 2003 Feb; 42(5):1254–1265.

Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007 Sep; 372(3):774–797.

Kutateladze T, Overduin M. Structural mechanism of endosome docking by the FYVE domain. Science. 2001 Mar; 291(5509):1793–1796.

Kutateladze TG. Translation of the phosphoinositide code by PI effectors. Nat Chem Biol. 2010 Jul; 6(7):507–513.

Lathrop B, Gadd M, Biltonen RL, Rule GS. Changes in Ca2+ affinity upon activation of Agkistrodon piscivorus piscivorus phospholipase A2. Biochemistry. 2001 Mar; 40(11):3264–3272.

Lazaridis T. Effective energy function for proteins in lipid membranes. Proteins: Struct, Funct, Bioinf. 2003; 52:176–192.

Lemmon MA. Membrane recognition by phospholipid-binding domains. Nat Rev Mol Cell Biol. 2008 Feb; 9(2):99–111.

Lin Y, Nielsen R, Murray D, Hubbell WL, Mailer C, Robinson BH, Gelb MH. Docking phospholipase A2 on membranes using electrostatic potential-modulated spin relaxation magnetic resonance. Science. 1998 Mar; 279(5358):1925–1929.

Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. BMC Structural Biology. 2007; 7:44.

Lomize AL, Pogozheva ID, Mosberg HI. Anisotropic solvent model of the lipid bilayer. 1. Parameterization of long-range electrostatics and first solvation shell effects. J Chem Inf Model. 2011 Apr; 51(4):918–929.

Lomize AL, Pogozheva ID, Mosberg HI. Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. J Chem Inf Model. 2011 Apr; 51(4):930–946.

Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2012 Jan; 40(Database issue):D370–6.

Malmberg NJ, Van Buskirk DR, Falke JJ. Membrane-docking loops of the cPLA2 C2 domain: detailed structural analysis of the protein-membrane interface via site-directed spin-labeling. Biochemistry. 2003 Nov; 42(45):13227–13240.

Mathias JD, Ran Y, Carter JD, Fanucci GE. Interactions of the GM2 activator protein with phosphatidylcholine bilayers: A site-directed spin-labeling power saturation study. Biophys J. 2009 Sep; 97(5):1436–1444.

Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol. 1987 Aug; 196(3):641–656.

Misra S, Hurley JH. Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. Cell. 1999 May; 97(5):657–666.

Mulgrew-Nesbitt A, Diraviyam K, Wang J, Singh S, Murray P, Li Z, Rogers L, Mirkovic N, Murray D. The role of electrostatics in protein-membrane interactions. Biochim Biophys Acta. 2006 Aug; 1761(8):812–826.

Oldham ML, Brash AR, Newcomer ME. Insights from the X-ray crystal structure of coral 8R-lipoxygenase: calcium activation via a C2-like domain and a structural basis of product chirality. J Biol Chem. 2005 Nov; 280(47):39545–39552.

Park MJ, Sheng R, Silkov A, Jung DJ, Wang ZG, Xin Y, Kim H, Thiagarajan-Rosenkranz P, Song S, Yoon Y, Nam W, Kim I, Kim E, Lee DG, Chen Y, Singaram I, Wang L, Jang MH, Hwang CS, Honig B, et al. SH2 domains serve as lipid-binding modules for pTyr-signaling proteins. Mol Cell. 2016 Apr; 62(1):7–20.

Perisic O, Fong S, Lynch DE, Bycroft M, Williams RL. Crystal structure of a calcium-phospholipid binding domain from cytosolic phospholipase A2. J Biol Chem. 1998 Jan; 273(3):1596–1604.

Phillips LR, Milescu M, Li-Smerin Y, Mindell JA, Kim JI, Swartz KJ. Voltage-sensor activation with a tarantula toxin as cargo. Nature. 2005 Aug; 436(7052):857–860.

Rufener E, Frazier AA, Wieser CM, Hinderliter A, Cafiso DS. Membrane-bound orientation and position of the synaptotagmin C2B domain determined by site-directed spin labeling. Biochemistry. 2005 Jan; 44(1):18–28.

Scott DL, White SP, Otwinowski Z, Yuan W, Gelb MH, Sigler PB. Interfacial catalysis: the mechanism of phospholipase A2. Science. 1990 Dec; 250(4987):1541–1546.

Sekino-Suzuki N, Nakamura M, Mitsui KI, Ohno-Iwashita Y. Contribution of individual tryptophan residues to the structure and activity of theta-toxin (perfringolysin O), a cholesterol-binding cytolysin. Eur J Biochem. 1996 Nov; 241(3):941–947.

Shenkarev ZO, Nadezhdin KD, Sobol VA, Sobol AG, Skjeldal L, Arseniev AS. Conformation and mode of membrane interaction in cyclotides. Spatial structure of kalata B1 bound to a dodecylphosphocholine micelle. Febs J. 2006 Jun; 273(12):2658–2672.

Stahelin RV, Burian A, Bruzik KS, Murray D, Cho W. Membrane binding mechanisms of the PX domains of NADPH oxidase p40phox and p47phox. J Biol Chem. 2003 Apr; 278(16):14469–14479.

Stahelin RV, Cho W. Differential Roles of Ionic, Aliphatic, and Aromatic Residues in Membrane-Protein Interactions: A surface plasmon resonance study on phospholipases A2. Biochemistry. 2001 Apr; 40(15):4672–4678.

Stahelin RV, Long F, Diraviyam K, Bruzik KS, Murray D, Cho W. Phosphatidylinositol 3-phosphate induces the membrane penetration of the FYVE domains of Vps27p and Hrs. J Biol Chem. 2002 Jul; 277(29):26379–26388.

Stahelin RV, Long F, Peter BJ, Murray D, De Camilli P, McMahon HT, Cho W. Contrasting membrane interaction mechanisms of AP180 N-terminal homology (ANTH) and epsin N-terminal homology (ENTH) Domains. J Biol Chem. 2003 Jul; 278(31):28993–28999.

Tatulian SA, Qin S, Pande AH, He X. Positioning membrane proteins by novel protein engineering and biophysical approaches. J Mol Biol. 2005 Sep; 351(5):939–947.

Team RDC. *R: A language and environment for statistical computing*. . 2008; .

Thennarasu S, Lee DK, Poon A, Kawulka KE, Vederas JC, Ramamoorthy A. Membrane permeabilization, orientation, and antimicrobial mechanism of subtilosin A. Chem Phys Lipids. 2005 Oct; 137(1-2):38–51.

Touw WG, Baakman C, Black J, te Beek TAH, Krieger E, Joosten RP, Vriend G. A series of PDB-related databanks for everyday needs. Nucleic Acids Res. 2015 Jan; 43(D1):D364–D368.

Velankar S, Best C, Beuth B, Boutselakis CH, Cobley N, Sousa Da Silva AW, Dimitropoulos D, Golovin A, Hirshberg M, John M, Krissinel EB, Newman R, Oldfield T, Pajon A, Penkett CJ, Pineda-Castillo J, Sahni G, Sen S, Slowley R, Suarez-Uruena A, et al. PDBe: Protein Data Bank in Europe. Nucleic Acids Res. 2010 Jan; 38(Database issue):D308–17.

Vögler O, Barceló JM, Ribas C, Escribá PV. Membrane interactions of G proteins and other related proteins. Biochim Biophys Acta. 2008 Jul; 1778(7-8):1640–1652.

Walther M, Wiesner R, Kuhn H. Investigations into calcium-dependent membrane association of 15-lipoxygenase-1. Mechanistic roles of surface-exposed hydrophobic amino acids and calcium. J Biol Chem. 2004 Jan; 279(5):3717–3725.

Wang QJ, Fang TW, Nacro K, Marquez VE, Wang S, Blumberg PM. Role of hydrophobic residues in the C1b domain of protein kinase C delta on ligand and phospholipid interactions. J Biol Chem. 2001 Jun; 276(22):19580–19587.

Wilmot CM, Thornton JM. Analysis and prediction of the different types of beta-turn in proteins. J Mol Biol. 1988 Sep; 203(1):221–232.

Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol. 1996 Oct; 3(10):842–848.

Zhang G, Kazanietz MG, Blumberg PM, Hurley JH. Crystal structure of the cys2 activator-binding domain of protein kinase C delta in complex with phorbol ester. Cell. 1995 Jun; 81(6):917–924.