Biochimica et Biophysica Acta 1834 (2013) 798–807

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

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

Analysis of Molecular Recognition Features (MoRFs) in membrane proteins

Ioly Kotta-Loizou, Georgios N. Tsaousis, Stavros J. Hamodrakas *

Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 15701, Greece

ARTICLE INFO

Article history:
Received 27 October 2012
Received in revised form 18 December 2012
Accepted 8 January 2013
Available online 15 January 2013

Keywords:
MoRFs
Intrinsic disorder
Protein–protein interactions
Transmembrane proteins
Peripheral membrane proteins

ABSTRACT

Molecular Recognition Features (MoRFs) are defined as short, intrinsically disordered regions in proteins that undergo disorder-to-order transition upon binding to their partners. As their name suggests, they are implicated in molecular recognition, which serves as the initial step for protein–protein interactions. Membrane proteins constitute approximately 30% of fully sequenced proteomes and are responsible for a wide variety of cellular functions. The aim of the current study was to identify and analyze MoRFs in membrane proteins. Two datasets of MoRFs, transmembrane and peripheral membrane protein MoRFs, were constructed from the Protein Data Bank, and sequence, structural and functional analysis was performed. Characterization of our datasets revealed their unique compositional biases and membrane protein MoRFs were categorized depending on their secondary structure after the interaction with their partners. Moreover, the position of transmembrane protein MoRFs in relation with the protein's topology was determined. Further studies were focused on functional analyses of MoRF-containing proteins and MoRFs' partners, associating them with protein binding, regulation and cell signaling, indicating half of them as putative hubs in protein–protein interaction networks. In conclusion, we provide insights into the disorder-based protein–protein interactions involving membrane proteins.

1. Introduction

During the past decade or so there has been a growing interest in the field of protein intrinsic disorder-related research. Intrinsically disordered proteins (IDPs) possess no rigid 3D structure under physiological conditions, yet they are functionally active [1]. A protein may be fully or partly disordered, containing long or short intrinsically disordered regions (IDRs) [2,3]. Genome-wide prediction-based assessments revealed that the natural abundance of IDRs significantly increases from Prokaryota to Eukaryota [4]. Specifically, in Eukaryota, 20–30% of proteins are IDPs and more than 50% of proteins contain long IDRs [5], while much lower percentages in Eubacteria and Archaea were noted [6]. Due to their biological abundance and functional importance, a number of specialized databases are dedicated to IDPs and IDRs, such as DisProt [7,8], MobiDB [9] and IDEAL [10].

As it is widely accepted, the amino acid sequence determines a protein's ability to fold or not to fold under given environmental conditions and comparison of IDPs with ordered proteins revealed their distinct amino acid composition [11–14]. IDPs and IDRs have been found to possess little or no secondary structure elements and lack a stable tertiary structure [15]. The biological functions of IDPs often depend on the absence of a specific 3D structure [16].

IDPs are classified into seven functional classes: molecular recognition effectors, molecular assemblers, chaperones, metal sponges, molecular recognition scavengers, modification sites and entropic chains [17]. It has been found that many IDPs are involved in regulation, control, molecular recognition and cell signaling. Moreover, IDPs are able to interact with multiple partners. They frequently serve as the popular nodes, or hubs, in protein interaction networks, where one IDP binds to many partners or many IDPs bind to one partner [18,19] (one-to-many signaling and many-to-one signaling [3], respectively). It should be noted that many IDPs are able to undergo a disorder-to-order transition upon binding to a specific partner [13,19–25]. These short binding regions were later named Molecular Recognition Features (MoRFs) or Molecular Recognition Elements (MoREs) [5]. In addition, Eukaryotic Linear Motifs (ELMs) [26,27] and Short Linear Motifs (SLiMs) [28–30], although based on consensus sequence patterns rather than structural characteristics, are essentially describing the same binding elements as MoRFs [27]. Both linear motifs and MoRFs refer to short protein segments and linear motifs often reside within IDRs and adopt a well-defined structure upon binding, similarly to MoRFs [31,32]. MoRFs and linear motifs participate in peptide-mediated interactions, in which a short peptide stretch from one partner interacts with a large protein surface from the other partner [33]. Furthermore, a number of common examples establish a connection

Abbreviations: MoRFs, Molecular Recognition Features; mpMoRFs, membrane protein MoRFs; non-mpMoRFs, non-membrane protein MoRFs; tmpMoRFs, transmembrane protein MoRFs; pmpMoRFs, peripheral membrane protein MoRFs; lampMoRFs, lipid-anchor membrane protein MoRFs; IDPs, Intrinsically Disordered Proteins; IDRs, Intrinsically Disordered Regions; ELMs, Eukaryotic Linear Motifs; SLiMs, Short Linear Motifs

* Corresponding author at: Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 15701, Greece. Tel.: +30 210 727 4931; fax: +30 210 7274254.

E-mail addresses: ikottaloizou@biol.uoa.gr (I. Kotta-Loizou), gtsaousis@biol.uoa.gr (G.N. Tsaousis), shamodr@biol.uoa.gr (S.J. Hamodrakas).

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.01.006

between these two models of molecular recognition and MoRFs and linear motifs were shown to overlap significantly [34].

MoRFs may form upon binding alpha-helices (α-MoRFs), beta-strands (β-MoRFs), irregular structure (i-MoRFs) or combinations of these elements (complex MoRFs). As their name suggests, they are considered to be implicated in molecular recognition, which serves as the initial step for protein–protein interactions [35]. Further studies, focused on characterizing the interactions between MoRFs and their binding partners, revealed the unique amino acid composition and geometric and physicochemical properties of these interfaces [36]. Based on these findings, a few MoRF prediction algorithms have been proposed [5,35,37–39]. The first predictors were specifically developed for α-MoRFs [5,37]. It was also noted that MoRFs often mapped in ordered regions flanked by long predictions of disorder [35]. More recently, two other algorithms for the prediction of MoRFs were developed, ANCHOR [38] and MoRFpred [39] as well as an algorithm for the prediction of SLiMs, SLiMPred [40]. All are available online as web-servers.

A functional analysis of MoRFs revealed that approximately 20% of MoRF-containing proteins are transmembrane [35]. Membrane proteins constitute approximately 30% of fully sequenced proteomes and are responsible for a wide variety of cellular functions, including cell signaling [41]. Membrane proteins are grouped into 3 distinct classes: transmembrane proteins that span the lipid bilayer, peripheral membrane proteins that are non-covalently associated with transmembrane proteins or lipids and lipid-anchor proteins that are covalently anchored to lipids [42]. Few studies concerning intrinsic disorder and membrane proteins have been conducted. One of them demonstrated that IDRs of alpha-helical and beta-barrel transmembrane proteins exhibit statistically distinct amino acid compositional biases in comparison with those of globular proteins [43]. Other studies confirmed the abundance of IDRs in eukaryotic transmembrane proteins as opposed to the bacterial ones, and revealed that IDRs preferentially occur on the cytoplasmic side of human plasma transmembrane proteins [44,45].

The aim of the current study was to identify and analyze MoRFs in membrane proteins, designated as mpMoRFs. Two mpMoRFs datasets were created: tmpMoRFs and pmpMoRFs, consisting of MoRFs in transmembrane and peripheral membrane proteins, respectively. Subsequently, sequence, structural and functional analysis of the mpMoRFs datasets was conducted.

2. Methods

2.1. Assembly of mpMoRFs and control datasets

An initial dataset of MoRFs was constructed from the Protein Data Bank (April 2012), following the approach of Mohan et al., by selecting membrane protein sequences between 10 and 70 residues [46], which interact with proteins longer than 100 residues [35]. The assumption was made that such short amino acid sequences would be less likely to form a rigid 3D structure prior to interaction [35]. However, fragments shorter than 10 residues were not included in the dataset, mainly in order to avoid chameleon segments, identical but unrelated sequences up to 8 residues that may adopt different secondary structures [36]. Initially, we used the PDB advanced search in order to isolate entries containing more than 2 protein entities and at least one sequence between 10 and 70 residues. The corresponding 4697 PDB files were downloaded and both nucleotide sequences and protein sequences containing unspecified amino acids (designated as UNK) were discarded. From the remaining 4682 PDB entries, only the 3569 containing at least one protein sequence longer than 100 residues were used. The appropriate information (PDB ID, chain, sequence and the DBREF record that provides cross-reference between the PDB sequences and a corresponding public database) concerning the chains between 10 and 70 residues was retained. Using the DBREF record of each PDB entry, 1115 unique MoRF-containing proteins were collected from the Uniprot database [47]. The annotation of these 1115 proteins in Uniprot was used to identify 289 membrane proteins, which were further categorized in 191 transmembrane, 79 peripheral membrane and 19 lipid-anchor membrane proteins. Finally, non-redundant datasets for mpMoRFs were created by applying length-dependent thresholds, using UniqueProt [48], on the short sequences of MoRFs. The non-redundant tmpMoRFs, pmpMoRFs and lampMoRFs datasets consist of 101 transmembrane, 56 peripheral membrane and 9 lipid-anchor protein MoRFs, respectively (Tables S1, S4, S5, S6).

In addition to the mpMoRFs sets, four additional control datasets were constructed:

(a) The short sequences of MoRFs that correspond to the remaining 826 non-membrane proteins were redundancy reduced using UniqueProt and a non-redundant set of 936 non-mpMoRFs was created (Table S7). The non-mpMoRFs dataset was used in order to compare the amino acid composition of mpMoRFs to MoRFs derived from non-membrane proteins, as described in Section 3.2. (b) The ordered protein–protein complexes set of Gunasekaran et al. [49] was obtained, which consists of interacting proteins, known to be ordered when isolated. Only 10 out of the 26 complexes of the initial dataset were used, those present in solution and not only in the crystal (Table S8). This dataset was used in the structure-based criterion that allows the discrimination between ordered proteins' complexes and MoRFs' complexes, as described in Section 3.4. (c) Human, experimentally verified peripheral plasma membrane proteins were selected from Uniprot and redundancy reduced to 30% using CD-HIT [50], in order to obtain a dataset of 148 human peripheral plasma membrane proteins (Table S10). This dataset was used in order to compare the amino acid composition of pmpMoRFs to peripheral membrane proteins, as described in Section 3.2. (d) The alpha-helical transmembrane proteins dataset provided by PDBTM [51], a database that contains the transmembrane proteins of PDB, was redundancy reduced to 30% using CD-HIT, to obtain a dataset consisting of 286 alpha-helical proteins (Table S9). This dataset was used in order to compare the amino acid composition of tmpMoRFs to transmembrane proteins, as described in Section 3.2. (e) Human transmembrane proteins were selected from Uniprot by performing an advanced search in order to create a dataset of 5191 reviewed human transmembrane proteins (Table S11). This dataset was used in GO-Term enrichment analysis, in order to compare the molecular functions of our human transmembrane MoRF-containing proteins to those of human transmembrane proteins in general, as described in Section 3.7.

2.2. Sequence analysis

Compositional profiling of mpMoRFs was analyzed in comparison to a reference set using Composition Profiler, a tool assessing statistically significant enrichment and depletion of amino acids, either individually or grouped according to their physicochemical and structural properties [52]. The default settings of the Composition Profiler web server were used for the calculations.

PONDR-FIT [53], a sequence-based meta-predictor of intrinsically disordered regions, was used to assess the presence of intrinsic disorder in mpMoRFs and MoRF-containing proteins.

2.3. Structural analysis

MoRFs were categorized into four different types, α, β, i and complex, based on the largest percentage value of their secondary structure types formed upon binding, as inferred by DSSP [54,55]. If a MoRF has no clear preponderance of any secondary structure type (at least 1% greater than the other two types), then it is designated as a complex-MoRF [27,39]. Predisposition of each mpMoRF to form a specific secondary structure type upon binding was assessed using the secondary structure predictor Porter [56].

Solvent accessible surface area (ASA) was calculated using DSSP [54,55]. The interface area (Å^2) buried by a complex was calculated as the difference between the sum of the surface areas of the separate monomers and the surface area of the complex. Per residue surface area and per residue interface area were calculated by dividing surface area and interface area with the number of MoRF residues present in the solved structure [49].

2.4. Functional analysis

Functional analyses of MoRF-containing proteins and MoRFs' partners were carried out using the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) [57] and the Generic Gene-Ontology Term Finder (GOTermFinder) [58]. Protein–protein interactions of MoRF-containing proteins and MoRFs' partners were inferred from the IntAct molecular interaction database [59]. Only binary interactions that did not originate from spoke-expanded co-complexes were used, in order to increase the accuracy of the analysis.

3. Results and discussion

3.1. Membrane protein MoRFs datasets

Three non-redundant datasets of membrane MoRFs were constructed: tmpMoRFs, pmpMoRFs and lampMoRFs (Tables 1, S4, S5, S6). The tmpMoRFs dataset consists of 101 MoRFs derived from 94 different transmembrane proteins. Specifically, there are 5 proteins containing 2 MoRFs each and 1 containing 3. Similarly, the pmpMoRFs dataset contains 56 MoRFs from 50 different proteins, with 5 proteins containing 2 MoRFs each and 1 containing 3. Finally, there are only 9 lampMoRFs from 9 different lipid-anchor proteins. This dataset was very small and therefore was not used in further statistical analyses. In all datasets, the PDB structures were mostly determined using X-ray crystallography (94, 48 and 8 in the tmpMoRFs, pmpMoRFs and lampMoRFs dataset, respectively) and, to a lesser extent, solution NMR. All tmpMoRFs and pmpMoRFs have an interacting interface at least 300 Å^2, and for more than 95% of them the interface is at least 400 Å^2 [39].

The vast majority of MoRF-containing proteins were found to be eukaryotic (Fig. S1), as expected, since MoRFs are by definition disordered when unbound and IDRs are more abundant in Eukaryota. Specifically, most MoRF-containing proteins in our datasets were human. The abundance of IDRs in Eukaryota has been attributed to their increased complexity in comparison to Prokaryota, which leads to a greater need for signaling and regulation [4,21,23]. Length distribution analysis of both non-mpMoRFs and mpMoRFs (Fig. S2) revealed that approximately half of the tmpMoRFs are very short, between 10 and 20 residues, as observed previously for MoRFs in general [35] and similarly to our non-mpMoRFs control dataset. On the other hand, length distribution of the pmpMoRFs dataset is more uniform.

3.2. Amino acid composition of mpMoRFs

Previous studies regarding amino acid preferences of IDRs led to the identification of disorder- and order-promoting residues (A, R, G, Q, S, P, E, K and W, C, F, I, Y, V, L, N, respectively) [11,13,14]. Amino acid compositional bias of MoRFs is expected to be between those of ordered and disordered proteins, since MoRFs are supposed to be disordered when unbound but, in parallel, must retain the ability to mediate protein–protein interactions through a disorder-to-order transition upon binding [35]. In order to examine the compositional biases of the mpMoRFs datasets, we compared them to a number of control datasets, using the Composition Profiler web server [52]. The amino acids are arranged from the least to the most hydrophobic according to the Kyte–Doolittle hydrophobicity scale [60] (Figs. 1,2).

Table 1
Number of protein MoRFs, MoRF-containing proteins, MoRF-containing organisms and total MoRFs' amino acids in all datasets.

Dataset	MoRFs	Amino acids	Proteins	Organisms
Non-membrane	936	27660	716	151
Transmembrane	101	2633	94	19
Peripheral membrane	56	1943	50	13
Lipid-anchor membrane	9	424	9	5

Initially, both tmpMoRFs and pmpMoRFs datasets were compared to the overall amino acid composition of proteins, more specifically the SwissProt background dataset provided by Composition Profiler [52]. A statistically significant preference of mpMoRFs for charged residues (D, E, K and R) was observed. It should be noted that D, K and R, are considered as disorder-promoting residues [11,13,14]. Furthermore, mpMoRFs are significantly depleted in the most hydrophobic, order-promoting residues, such as I, V and L. Therefore, the compositional biases of mpMoRFs are a possible indicator of their disordered nature. Interestingly, tmpMoRFs show also a statistically significant preference for C, as observed previously for MoRFs in general [35], in contrast to the pmpMoRFs. However, pmpMoRFs possess biases closer to those of IDRs than to other MoRFs: they are depleted in aromatic residues and C.

Similarly, the amino acid composition of tmpMoRFs and pmpMoRFs was assessed in comparison to non-mpMoRFs, and it was revealed that mpMoRFs have their own compositional biases. TmpMoRFs are significantly enriched in aromatic residues (W and Y). Notably, W is abundant in IDRs of transmembrane proteins [43]. Interestingly, tmpMoRFs are also enriched in C, while MoRFs in general already possess a high content of C, attributed to the formation of disulfide bonds [35] and, in contrast, IDRs of transmembrane proteins are depleted in C [43].

The compositional biases of MoRFs appear enhanced when compared to the amino acid sequence surrounding a given MoRF, which is less likely to contain unannotated MoRFs [39]. We defined the flanking regions of membrane MoRFs as the surrounding sequences, upstream and downstream of each MoRF, equal to the average length of MoRFs in each dataset: 27 residues for tmpMoRFs and 35 residues for pmpMoRFs. These flanking regions were assessed using PONDR-FIT, and approximately 80% and 89% of tmpMoRFs and pmpMoRFs, respectively, were found to reside in at least partly disordered regions. The comparison of mpMoRFs with their flanking regions revealed that mpMoRFs are significantly enriched in aromatic residues. It has been suggested previously that aromatic residues are strategically positioned within IDRs in order to ensure proper function [2], which, in the case of MoRFs, is the ability to contribute to protein–protein interactions. Furthermore, mpMoRFs have a tendency to be more hydrophobic than their flanking regions. This observation is in agreement with previous data [39] and with the notion that a local increase of hydrophobicity in IDRs marks the existence of a putative binding site [36,61]. A statistically significant enrichment of mpMoRFs in positively charged residues (K and R) was also noted. In the case of tmpMoRFs, this can be attributed to the fact that cytoplasmic regions of transmembrane proteins have a net positive charge (positive inside rule) [62,63], and IDRs, and consequently MoRFs, as noted in Section 3.5, are also located mostly in the cytoplasmic regions [44].

Finally, we wanted to compare the amino acid preferences of mpMoRFs to those of membrane proteins. To this end, we used as a background set 286 alpha-helical transmembrane sequences from PDBTM. Their transmembrane segments were removed, because their strong compositional biases for hydrophobic residues would cloud those of MoRFs. A statistically significant enrichment of tmpMoRFs in charged residues and depletion in hydrophobic residues was noted. The increased aromatic content of tmpMoRFs, evident previously in their comparison to non-mpMoRFs but not in their current comparison to non-transmembrane sequences of transmembrane proteins, can be attributed to the transmembrane alpha-helical proteins' compositional bias. Similarly, our pmpMoRFs were compared to the human plasma peripheral membrane proteins dataset. Again, pmpMoRFs were observed to be

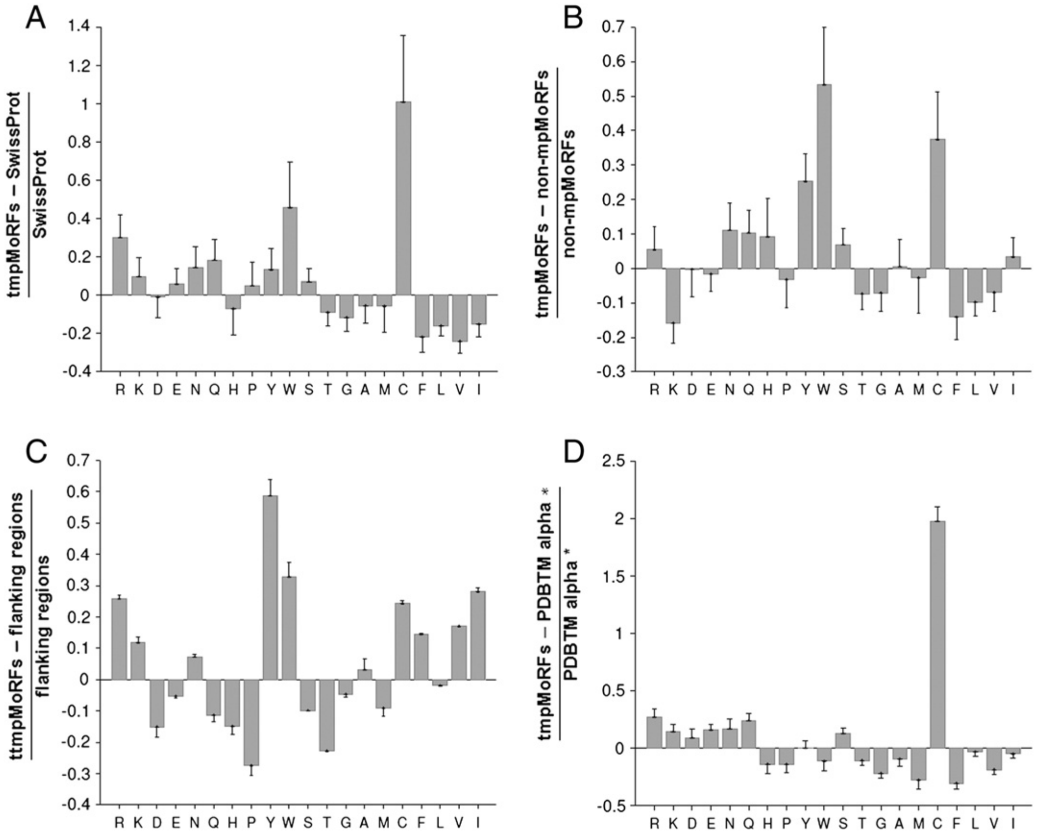

Fig. 1. Compositional profiling of tmpMoRFs. Amino acid propensity of tmpMoRFs is compared to: (A) SwissProt [47], (B) the non-mpMoRFs dataset, (C) the flanking regions of tmpMoRFs and (D) the PDBTM [51] alpha-helical proteins dataset (*non-transmembrane segments only).

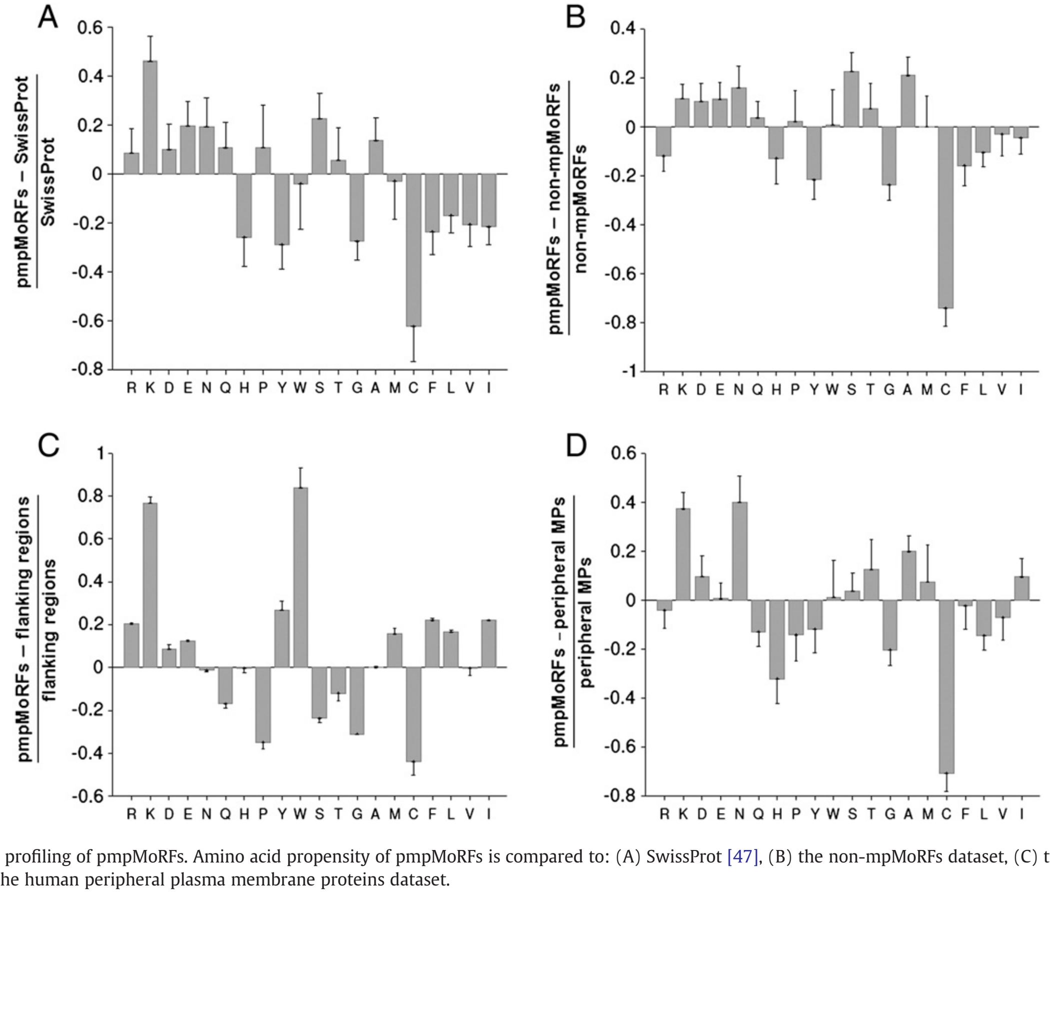

Fig. 2. Compositional profiling of pmpMoRFs. Amino acid propensity of pmpMoRFs is compared to: (A) SwissProt [47], (B) the non-mpMoRFs dataset, (C) the flanking regions of pmpMoRFs and (D) the human peripheral plasma membrane proteins dataset.

significantly enriched in positively charged residues and depleted in hydrophobic residues.

3.3. Secondary structure of mpMoRFs

MoRFs are categorized as α-MoRFs, β-MoRFs and i-MoRFs, based on their secondary structure upon binding: alpha-helix, beta-strand and irregular, respectively. Furthermore, complex-MoRFs exist, arising from combinations of α, β and i-MoRFs [5,35]. In both datasets, approximately half of the MoRFs are irregular, while β-MoRFs and complex-MoRFs are rare (Fig. S3A). In the case of tmpMoRFs, lack of beta-structure is possibly due to the prevalence of alpha-helical transmembrane proteins in our dataset, as will be discussed shortly in Section 3.5.

The secondary structure distribution of mpMoRFs' residues, as inferred by DSSP, was assessed and the percentages of residues in alpha-helical, beta-strand and irregular conformations were calculated (Fig. S3B). In both tmpMoRFs and pmpMoRFs datasets, approximately 60% of residues were either found in irregular conformation or with missing PDB file coordinates. Missing density residues are often associated with the presence of disorder. The percentage of residues in alpha-helical conformation is much higher than those in beta-strand in both datasets. Overall, these results are in agreement with those observed previously for MoRFs in general [35].

In the same study, it was noted that MoRFs may possess secondary structure preferences that determine their actual structure upon binding [35]. We wanted to examine whether protein structure formation influences not only mpMoRFs' disorder, but also their secondary structure. Therefore, we compared the actual and the predicted percentage value of residues in alpha-helical, beta-strand and irregular conformation in each type of mpMoRFs, using secondary structure predictor Porter (Tables S2, S3). In both mpMoRFs datasets, Porter predictions concerning MoRFs' secondary structure are similar to DSSP assessments, indicating that the secondary structural predisposition of mpMoRFs is mostly responsible for their structure upon binding, as has been shown in the case of MoRFs in general [35].

3.4. MpMoRFs and intrinsic disorder

IDPs have been considered to be functionally ergonomic because of their ability to provide extended intermolecular interfaces, without being exceptionally large themselves [64]. A structured-based criterion for distinguishing disordered from ordered protein complexes [49], which is based on the above idea, has been previously used for evaluation of intrinsic disorder in MoRF datasets [35]. The per-residue surface area versus the per-residue interface area plot (Fig. 3) demonstrated clearly the distinction between both mpMoRFs' complexes and a set of ordered protein–protein complexes. As expected, there is no possible distinction between the tmpMoRFs and the pmpMoRFs datasets.

Furthermore, we used a sequence-based meta-predictor, PONDR-FIT [53], to assess the presence of disorder in our MoRFs. Approximately 34% and 51% of tmpMoRFs' and pmpMoRFs' residues, respectively, were found to be disordered. The decreased percentage of predicted disorder in tmpMoRFs compared to pmpMoRFs may be attributed to the fact that PONDR-FIT was trained with globular proteins. It should be noted that PONDR-FIT and other intrinsic disorder predictors are not capable of recognizing MoRFs themselves, due to the less disordered, more stable and more hydrophobic nature of MoRFs [39]. Actually, it was observed that when using a predictor for long IDRs, such as PONDR VL-XT [65,66], binding sites were often predicted to be ordered short regions located within a long IDR [35].

3.5. Topology of transmembrane MoRF-containing proteins

MoRF-containing proteins in the tmpMoRFs dataset were classified according to the number of their membrane-spanning regions (Fig. S4). 63 out of 94 transmembrane MoRF-containing proteins were found to be single-spanning and classified into types I, II, III or IV (Fig. S5), with 47 characterized as type I. There are 4 proteins with 7 membrane-spanning regions and 3 of them are G-protein coupled receptors. It should be noted that all proteins but one possess alpha-helical transmembrane segments. The sole beta-barrel in the dataset has 12 transmembrane beta-strands, belongs to the phospholipase A1 family and is one of the very few enzymes located at the outer membrane of the bacterium *Escherichia coli*. Transmembrane beta-barrel proteins, particularly porins, have a negative association with disorder [67], thus possibly explaining their scarcity in our dataset.

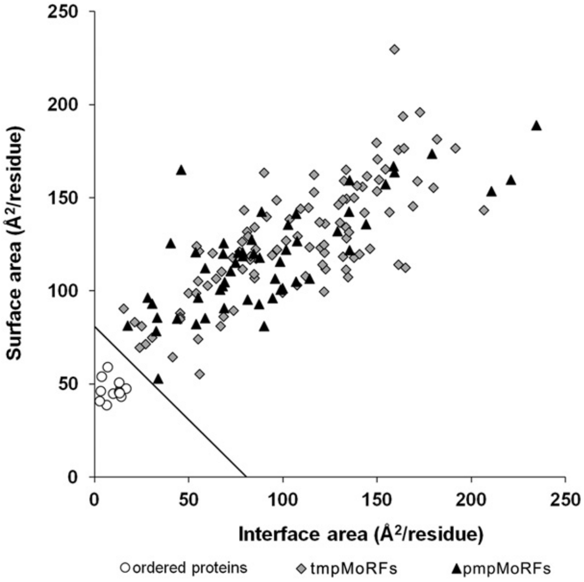

Fig. 3. Intrinsic disorder in mpMoRFs. A structure-based criterion [49] was used to evaluate the presence of intrinsic disorder in mpMoRFs datasets. The per-residue interface area (x-axis) and the per-residue surface area (y-axis) of MoRFs are larger than those of the known ordered proteins used for comparison.

We then determined the position of MoRFs in relation to the transmembrane proteins' topology (Fig. S6), as well as the distribution of MoRFs' residues in the cytoplasmic, extracellular, transmembrane and intramembrane segments of the MoRF-containing proteins. Approximately 70% of MoRFs and 60% of the MoRFs' residues were found to be cytoplasmic. Apparently, cytoplasmic MoRFs are shorter than extracellular ones. Interestingly, 1 MoRF, belonging to a bacterial protein, contains a re-entrant loop, a structure that goes halfway into the lipid bilayer and comes out again on the same side. In addition, the secondary structure distribution of membrane MoRFs' residues, as inferred by DSSP, was assessed once more, this time for cytoplasmic and extracellular MoRFs separately (Fig. S7). An increase in the sum of missing density and irregular structure residues was observed in cytoplasmic MoRFs (64%) compared to extracellular ones (54%). In summary, MoRFs are located mostly in the cytoplasmic region of the transmembrane proteins and cytoplasmic MoRFs are less likely to consist of well-defined secondary structural elements. These findings are in agreement with previous observations that IDRs preferentially occur on the cytoplasmic side [44,45,68], since MoRFs by definition are considered to be disordered when unbound. MoRFs present in the extracellular side, where disorder is relatively limited and a more rigid structure is favored, seem to comply with this tendency upon binding. A recent study reports the abundance of SLiMs on the cytoplasmic side of single-spanning proteins, co-localizing with a peak of disorder approximately 30 residues from the membrane [45]. Single-spanning proteins, which probably function as receptors [69,70], are approximately 2/3 of our dataset too. Notably, 68% of MoRFs located in single-spanning proteins were found to be cytoplasmic and their median distance from the membrane is 24 residues. Both MoRFs and SLiMs are considered to mediate molecular recognition and their localization probably reflects the increased need for protein–protein interactions within the cell in comparison to the extracellular environment.

3.6. MpMoRFs' partners

Apart from MoRF-containing proteins, mpMoRFs' partners in the initial complexes derived from PDB were determined and analyzed. It was observed that a MoRF could interact with more than one partner, although these cases are a minority. Specifically, in the tmpMoRFs dataset there are 18 MoRFs interacting with 2 partners each, 1 MoRF with 3 and 1 MoRF with 4. Similarly, in the pmpMoRFs dataset there are 3 MoRFs interacting with 2 partners each, 3 MoRFs with 3 and 3 MoRFs with 4. In some cases it was also observed that 1 protein could interact with more than 1 MoRFs. For example, human calmodulin was identified as the partner of 8 different MoRFs (7 tmpMoRFs and 1 pmpMoRF). In total, the tmpMoRFs' and pmpMoRFs' partner datasets consist of 84 and 60 different proteins, respectively.

The next step was to determine whether MoRFs' partners are transmembrane, peripheral membrane or lipid-anchor membrane proteins. Otherwise, if no reported association with cellular membranes was noted, they were designated as globular proteins (Fig. S8). Approximately half of tmpMoRFs' partners are globular proteins. Of these globular partners, 16 are antibodies, 7 secreted and 23 located in the cytoplasm. The vast majority of peripheral membrane partners in the tmpMoRFs dataset are also cytoplasmic. Regarding pmpMoRFs, approximately 40% of their partners are cytoplasmic peripheral membrane proteins, and their globular ones are also mostly cytoplasmic. The sole exception is the exotoxin type C of the bacterium *Streptococcus pyrogenes*. This is not surprising, since the vast majority of peripheral membrane MoRF-containing proteins are located intracellularly. Transmembrane partners were also noted in both datasets, while lipid-anchor membrane partners are a minority.

3.7. Functional analysis of membrane MoRF-containing proteins and MoRFs' partners

Traditionally, IDPs and proteins that contain IDRs are classified according to their function into a small number of categories [17], while, more recently [67,71], an extensive analysis revealed specific functions positively or negatively associated with the presence of intrinsic disorder. Functional analysis of membrane MoRF-containing proteins was restricted to human ones, due to their abundance in our datasets. More specifically, there are 49 transmembrane and 22 peripheral membrane MoRF-containing proteins.

A functional enrichment analysis of all 49 transmembrane and 22 peripheral membrane human MoRF-containing proteins was conducted, using GOTermFinder for finding significant over-represented GO molecular function terms. Initially, human genome was used as a background population. Protein binding, binding and regulation appear to be the most important molecular functions of the MoRF-containing proteins in both datasets. Alternatively, for human transmembrane MoRF-containing proteins, human transmembrane proteins dataset was used as background population, instead of the human genome. This modification allowed the comparison, in terms of molecular function, of the transmembrane MoRF-containing proteins specifically with other transmembrane proteins. Fewer GO terms were considered significant, all concerning binding and, to a lesser extent, cell-signaling. This is not unexpected, since 42 out of 49 human proteins are single-spanning, and probably function as receptors [69,70]. In addition, a GO Slim analysis of the membrane MoRF-containing proteins, using WebGestalt, supported the previous observations (Fig. S9).

Functional analysis of membrane MoRFs' partners was also restricted to the human ones. There are 29 and 21 human proteins in the tmpMoRFs' and pmpMoRFs' partner datasets, respectively. GO-Term enrichment analysis, carried out using GOTermFinder, revealed that the molecular functions of MoRFs' partners are very similar to those of MoRF-containing proteins, with protein binding being the most prominent one. GO Slim analysis supported this observation (Fig. S10). It should be noted that molecular recognition, process regulation and cell signaling are functions primarily associated with IDRs [19], supporting the connection between MoRFs and intrinsic disorder. Concerning molecular recognition, IDRs are considered to have a kinetic advantage in comparison to ordered proteins, since they specialize in high specificity - low affinity interactions, display increased interaction speed and possess larger interacting surfaces [13]. Furthermore, it has been observed that MoRFs have the ability to fold differently depending on their partner [35]. Proteins with IDRs play an important role in regulation as they can be more tightly regulated themselves: they are more vulnerable to proteolytic degradation [20,72] and more susceptible to enzymatic post-translational modifications [25]. Membrane proteins are heavily implicated in signal transduction and it is probable that our mpMoRFs mediate the interactions necessary for cell signaling.

It has been proposed that hubs, proteins with many connections in a protein–protein interaction network, use IDRs for binding to multiple partners [18], the abundance of intrinsic disorder in hubs when compared to non-hubs has been verified [73,74] and specific examples, such as p53, have been previously illustrated [75]. We examined whether the human membrane MoRF-containing proteins are implicated in an increased number of interactions and could possibly serve as hubs. The number of interactions for each MoRF-containing protein individually was inferred from the IntAct molecular interaction database. In general, proteins with more than 5 interactions are considered hubs [73,76–78]. 44% and 45% of the transmembrane and the peripheral membrane MoRF-containing proteins, respectively, have more than 5 interactions, and there are a few implicated in more than 30, such as single-pass type I amyloid beta A4 protein and human peripheral estrogen receptor. It is possible that the MoRF-containing hubs would be date hubs, proteins that interact with their numerous partners sequentially and not simultaneously, in contrast to party hubs. Date hubs are associated with cell signaling [79] and are more disordered that party hubs [80,81]. It should be mentioned that approximately 25% and 43% of residues in transmembrane and peripheral membrane MoRF-containing proteins, respectively, were predicted to be disordered using PONDR-FIT. In transmembrane proteins, the percentage of intrinsically disorder amino acids is increased in the cytoplasmic loops (43%), where the majority of protein–protein interactions would occur, in agreement with previous observations [44].

A similar analysis was carried out for the membrane MoRFs' partners. Again, some proteins are putative hubs: single-pass type I epidermal growth factor receptor, single-pass transmembrane Bcl-2-like protein 1, membrane-associated cell division control protein 42 homolog and globular calmodulin. Interestingly, the percentage of the partners implicated in an increased number of interactions is higher than that of MoRF-containing proteins (59% and 57% for the transmembrane and the peripheral membrane dataset, respectively). These hubs could be ordered proteins interacting with many disordered partners [18]. Concerning the other half of the datasets, the non-hub MoRF-containing proteins and MoRFs' partners, it has been reported that, in the case of non-hubs, proteins with IDRs preferentially interact with each other [82]. Therefore, it would be possible for partners to contain their own MoRFs.

3.8. Sequence-based prediction of MoRFs: selected examples

Two different algorithms designed to predict MoRFs, ANCHOR and MoRFpred, were used to assess our transmembrane and peripheral membrane MoRF-containing proteins, in order to evaluate their performance against mpMoRFs.

ANCHOR [38] is a tool built to predict disordered protein-binding regions that undergo a disorder-to-order transition upon binding. The algorithm identifies these segments within an amino acid sequence based on three criteria: (1) they must reside in regions predicted as disordered using IUPred [83], (2) they cannot form enough interchain interactions to fold on their own and (3) they are likely to energetically gain and be stabilized by interacting with a globular protein partner [38,61]. At least 1 MoRF was predicted in 54 out of 94 transmembrane and in 47 out of 50 peripheral membrane proteins. 20 out of 101 (20%) tmpMoRFs and 38 out of 56 (68%) pmpMoRFs were correctly predicted by ANCHOR. The increased success in the prediction of pmpMoRFs, which is equivalent to the almost 70% accuracy reported previously [38], is probably due to the greater similarity of peripheral membrane proteins with the globular ones that were used to build the algorithm. It should be taken into account that ANCHOR demonstrates increased success in predicting the presence, if not the exact position, of MoRFs in peripheral membrane proteins.

While this manuscript was in preparation, a new algorithm for the prediction of MoRFs, called MoRFpred, became available. MoRFpred utilizes both sequence alignments and support vector machines to predict all types of MoRFs and is claimed to outerperform ANCHOR [39]. In our datasets, 35 out of 101 (35%) tmpMoRFs and 33 out of 56 (59%) pmpMoRFs were predicted correctly. This algorithm seems to be more effective than ANCHOR at recognizing tmpMoRFs, possibly because they were not excluded from its training dataset. Again, an increased rate of success is observed for pmpMoRFs, although ANCHOR performed better on this dataset.

SLiMpred [40] predicts SLiMs using machine-learning techniques and structural, biophysical and biochemical data. We assessed our transmembrane and peripheral membrane MoRF-containing proteins using SLiMpred in order to examine whether predicted SLiMs overlap with mpMoRFs. Interestingly, 81 out of 101 (80%) tmpMoRFs and 43 out of 56 (77%) pmpMoRFs were found to contain at least one residue labeled as SLiM. This evidence supports the strong correlation between MoRFs and linear motifs [34]. Given that SLiMpred demonstrates increased efficiency in recognizing MoRFs in our datasets, compared to ANCHOR and MoRFpred, we propose that these algorithms, based on different concepts, could be used in a complementary way in MoRFs detection in order to achieve increased success rate. A reverse approach, using ANCHOR to discriminate between linear motifs real instances and random pattern matching hits, has already been used [34].

Selected examples from both the transmembrane and the peripheral membrane MoRF-containing protein datasets are illustrated below.

SED5 of *Saccharomyces cerevisiae* is a single-spanning type IV transmembrane protein, belonging to the syntaxin family. It plays a crucial role in vesicular transport between ER and the Golgi complex. SED5 protein is recognized by the SEC23/24 subcomplex of yeast COPII and mutagenesis of residue 206 (asparagine to alanine) abolishes the interaction between SED5 and SEC24 [84]. Furthermore, SM protein SLY1 interacts with SED5 protein in order to mediate intracellular membrane fusion [85]. Both MoRFs, an i-MoRF and an α-MoRF respectively, were identified by ANCHOR (Fig. 4 [86]).

Human ALK tyrosine kinase receptor (ALK) is a single-spanning type I transmembrane protein with constitutive tyrosine kinase activity that interacts with many signaling molecules. It was shown to

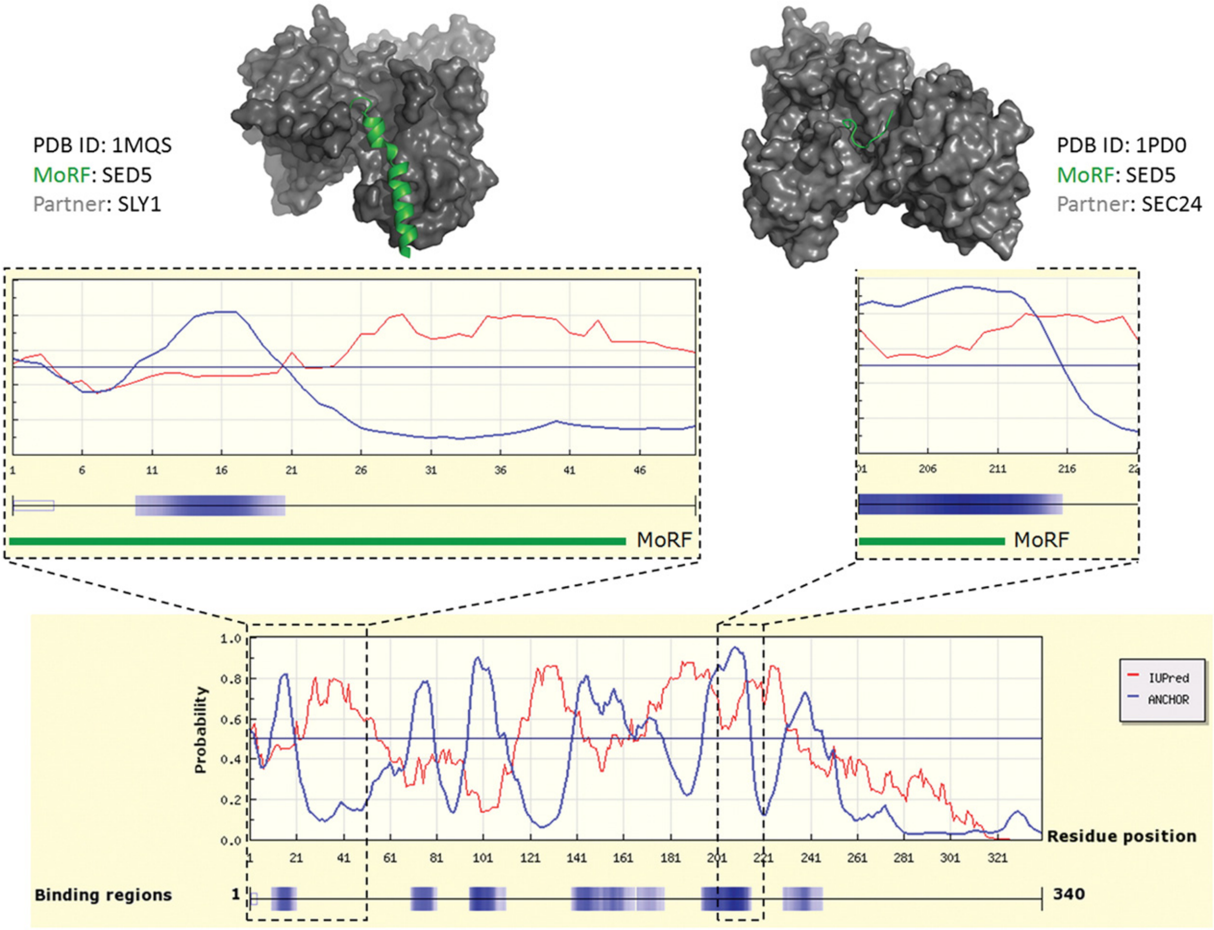

Fig. 4. Selected examples of mpMoRFs. ANCHOR [38] prediction of the yeast transmembrane MoRF-containing protein SED5 and the complexes between MoRFs (green) and their binding partners (grey) SLY1 and SEC24. PDB structures were visualized using PyMOL [86].

interact with fibroblast growth factor receptor substrate 3 (FRS3), a lipid-anchor membrane protein implicated in various cell-signaling pathways, through a phosphotyrosine binding domain [87]. The i-MoRF of ALK was predicted by ANCHOR (Fig. S11).

Human peripheral membrane ADP-ribosylation factor-binding protein GGA1 is implicated in protein sorting and trafficking between the trans-Golgi network and endosomes, in collaboration with clathrin and ADP-ribosylation factor. The interaction between GGA1 unstructured hinge region, an i-MoRF, and GGA1 gamma-adaptin ear is involved in the auto-regulation of GGA1 clathrin-mediated trafficking [88]. In the second complex [89], GGA1's partner is the mouse ADP-ribosylation factor 2 (ARF2), which has a 96% identity to human ADP-ribosylation factor 1 (ARF1). Multiple mutations within the α-MoRF (L182A, N194A, I197A, K198A, M200A and D204A) have been shown to abolish the interaction between GGA1 and ARF1 [90]. In this case too, both MoRFs were correctly predicted by ANCHOR (Fig. S12).

Human peripheral membrane charged multivesicular body protein 4a (CHMP4A) is a key component of transport complex III (ESCRT-III), which is involved in the sorting of endosomal proteins and the formation of multivesicular bodies. Programmed cell death 6-interacting protein (PDCD6IP) is responsible for the recruitment of CHMP4A to the ESCRT-III and a number of mutations within the α-MoRF (L214A, L217A and W220A) have been shown to abolish the interaction between CHMP4A and PDCD6IP [91]. ANCHOR correctly identified the MoRF (Fig. S13).

4. Conclusions

TmpMoRFs and pmpMoRFs share many common features with each other and with globular protein MoRFs, however they retain their own special characteristics, for example in terms of amino acid composition. Compositional biases of tmpMoRFs are influenced by IDRs', transmembrane proteins' and binding sites' preferences. There are many indications that mpMoRFs are likewise implicated in molecular recognition procedures, probably through a disorder-to-order transition upon binding to their partner. The distinction between mpMoRFs' complexes and ordered proteins' complexes, using a structure-based criterion, suggests that MoRFs are indeed disordered before the binding event. Especially in transmembrane proteins, which often function as the link between the cellular unit and its environment, high specificity, MoRF-mediated protein–protein interactions appear to be crucial. This notion is supported by the presence of MoRFs in the cytoplasmic segments of single-spanning proteins, which possibly function as receptors and are involved in cell signaling. In addition, peripheral membrane and globular partners of tmpMoRFs and pmpMoRFs, as well as peripheral membrane MoRF-containing proteins, are mostly located intracellularly, suggesting that they may be implicated in intracellular regulation procedures. The abundance of cytoplasmic MoRFs in both datasets can be attributed to the greater need for protein–protein interactions within the cell in comparison to the extracellular environment. It was also noted that available MoRF predictors do not perform very well in the case or tmpMoRFs. Therefore, we hope that, in the future, this information will facilitate identification, and possibly more accurate prediction, of tmpMoRFs and mpMoRFs in general.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and constructive criticism. This research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: Heracleitus II. "Investing in knowledge society through the European Social Fund". We also thank the University of Athens for support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2013.01.006.

References

[1] H.J. Dyson, Expanding the proteome: disordered and alternatively folded proteins, Q. Rev. Biophys. 44 (2011) 467–518.
[2] V.N. Uversky, Intrinsically disordered proteins from A to Z, Int. J. Biochem. Cell Biol. 43 (2011) 1090–1103.
[3] P. Romero, Z. Obradovic, C.R. Kissinger, J.E. Villafranca, E. Garner, S. Guilliot, A.K. Dunker, Thousands of proteins likely to have long disordered regions, Pac. Symp. Biocomput. (1998) 437–448.
[4] E. Schad, P. Tompa, H. Hegyi, The relationship between proteome size, structural disorder and organism complexity, Genome Biol. 12 (2011) R120.
[5] C.J. Oldfield, Y. Cheng, M.S. Cortese, P. Romero, V.N. Uversky, A.K. Dunker, Coupled folding and binding with alpha-helix-forming molecular recognition elements, Biochemistry 44 (2005) 12454–12470.
[6] A.K. Dunker, Z. Obradovic, P. Romero, E.C. Garner, C.J. Brown, Intrinsic protein disorder in complete genomes, Genome Inform. Ser. Workshop Genome Inform. 11 (2000) 161–171.
[7] S. Vucetic, Z. Obradovic, V. Vacic, P. Radivojac, K. Peng, L.M. Iakoucheva, M.S. Cortese, J.D. Lawson, C.J. Brown, J.G. Sikes, C.D. Newton, A.K. Dunker, DisProt: a database of protein disorder, Bioinformatics (Oxford, England) 21 (2005) 137–140.
[8] M. Sickmeier, J.A. Hamilton, T. LeGall, V. Vacic, M.S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V.N. Uversky, Z. Obradovic, A.K. Dunker, DisProt: the database of disordered proteins, Nucleic Acids Res. 35 (2007) D786–D793.
[9] T. Di Domenico, I. Walsh, A.J. Martin, S.C. Tosatto, MobiDB: a comprehensive database of intrinsic protein disorder annotations, Bioinformatics (Oxford, England) 28 (2012) 2080–2081.
[10] S. Fukuchi, S. Sakamoto, Y. Nobe, S.D. Murakami, T. Amemiya, K. Hosoda, R. Koike, H. Hiroaki, M. Ota, IDEAL: intrinsically disordered proteins with extensive annotations and literature, Nucleic Acids Res. 40 (2011) D507–D511.
[11] R.M. Williams, Z. Obradovi, V. Mathura, W. Braun, E.C. Garner, J. Young, S. Takayama, C.J. Brown, A.K. Dunker, The protein non-folding problem: amino acid determinants of intrinsic order and disorder, Pac. Symp. Biocomput. (2001) 89–100.
[12] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41 (2000) 415–427.
[13] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic, Intrinsically disordered protein, J. Mol. Graph. Model. 19 (2001) 26–59.
[14] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, Proteins 42 (2001) 38–48.
[15] V.N. Uversky, A.K. Dunker, Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes, Anal. Chem. 84 (2012) 2096–2104.
[16] V.N. Uversky, A.K. Dunker, Understanding protein non-folding, Biochim. Biophys. Acta 1804 (2010) 1231–1264.
[17] P. Tompa, Intrinsically unstructured proteins, Trends Biochem. Sci. 27 (2002) 527–533.
[18] A.K. Dunker, M.S. Cortese, P. Romero, L.M. Iakoucheva, V.N. Uversky, Flexible nets. The roles of intrinsic disorder in protein interaction networks, FEBS J. 272 (2005) 5129–5148.
[19] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, J. Mol. Recognit. 18 (2005) 343–384.
[20] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm, J. Mol. Biol. 293 (1999) 321–331.
[21] H.J. Dyson, P.E. Wright, Coupling of folding and binding for unstructured proteins, Curr. Opin. Struct. Biol. 12 (2002) 54–60.
[22] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat. Rev. Mol. Cell Biol. 6 (2005) 197–208.
[23] A.K. Dunker, Z. Obradovic, The protein trinity—linking function and disorder, Nat. Biotechnol. 19 (2001) 805–806.
[24] A.K. Dunker, C.J. Brown, Z. Obradovic, Identification and functions of usefully disordered proteins, Adv. Protein Chem. 62 (2002) 25–49.
[25] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradovic, Intrinsic disorder and protein function, Biochemistry 41 (2002) 6573–6582.
[26] H. Dinkel, S. Michael, R.J. Weatheritt, N.E. Davey, K. Van Roey, B. Altenberg, G. Toedt, B. Uyar, M. Seiler, A. Budd, L. Jodicke, M.A. Dammert, C. Schroeter, M. Hammer, T. Schmidt, P. Jehl, C. McGuigan, M. Dymecka, C. Chica, K. Luck, A. Via, A. Chatr-Aryamontri, N. Haslam, G. Grebnev, R.J. Edwards, M.O. Steinmetz, H. Meiselbach, F. Diella, T.J. Gibson, ELM—the database of eukaryotic linear motifs, Nucleic Acids Res. 40 (2012) D242–D251.

[27] W.L. Hsu, C. Oldfield, J. Meng, F. Huang, B. Xue, V.N. Uversky, P. Romero, A.K. Dunker, Intrinsic protein disorder and protein–protein interactions, Pac. Symp. Biocomput. (2012) 116–127.
[28] N.E. Davey, N.J. Haslam, D.C. Shields, R.J. Edwards, SLiMSearch 2.0: biological context for short linear motifs in proteins, Nucleic Acids Res. 39 (2011) W56–W60.
[29] N.E. Davey, N.J. Haslam, D.C. Shields, R.J. Edwards, SLiMFinder: a web server to find novel, significantly over-represented, short protein motifs, Nucleic Acids Res. 38 (2010) W534–W539.
[30] N.E. Davey, J.L. Cowan, D.C. Shields, T.J. Gibson, M.J. Coldwell, R.J. Edwards, SLiMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions, Nucleic Acids Res. 40 (2012) 10628–10641.
[31] A. Stein, P. Aloy, Contextual specificity in peptide-mediated protein interactions, PLoS One 3 (2008) e2524.
[32] N.E. Davey, K. Van Roey, R.J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, T.J. Gibson, Attributes of short linear motifs, Mol. Biosyst. 8 (2012) 268–281.
[33] L.G. Trabuco, S. Lise, E. Petsalaki, R.B. Russell, PepSite: prediction of peptide-binding sites from protein surfaces, Nucleic Acids Res. 40 (2012) W423–W427.
[34] B. Meszaros, Z. Dosztanyi, I. Simon, Disordered binding regions and linear motifs—bridging the gap between two models of molecular recognition, PLoS One 7 (2012) e46829.
[35] A. Mohan, C.J. Oldfield, P. Radivojac, V. Vacic, M.S. Cortese, A.K. Dunker, V.N. Uversky, Analysis of molecular recognition features (MoRFs), J. Mol. Biol. 362 (2006) 1043–1059.
[36] V. Vacic, C.J. Oldfield, A. Mohan, P. Radivojac, M.S. Cortese, V.N. Uversky, A.K. Dunker, Characterization of molecular recognition features, MoRFs, and their binding partners, J. Proteome Res. 6 (2007) 2351–2366.
[37] Y. Cheng, C.J. Oldfield, J. Meng, P. Romero, V.N. Uversky, A.K. Dunker, Mining alpha-helix-forming molecular recognition features with cross species sequence alignments, Biochemistry 46 (2007) 13468–13477.
[38] Z. Dosztanyi, B. Meszaros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins, Bioinformatics (Oxford, England) 25 (2009) 2745–2746.
[39] F.M. Disfani, W.L. Hsu, M.J. Mizianty, C.J. Oldfield, B. Xue, A.K. Dunker, V.N. Uversky, L. Kurgan, MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins, Bioinformatics (Oxford, England) 28 (2012) 75–83.
[40] C. Mooney, G. Pollastri, D.C. Shields, N.J. Haslam, Prediction of short linear protein binding regions, J. Mol. Biol. 415 (2012) 193–204.
[41] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (2001) 567–580.
[42] J.M. Berg, J.L. Tymoczko, L. Stryer, Biochemistry, 5th ed. W H Freeman, New York, 2002.
[43] B. Xue, L. Li, S.O. Meroueh, V.N. Uversky, A.K. Dunker, Analysis of structured and intrinsically disordered regions of transmembrane proteins, Mol. Biosyst. 5 (2009) 1688–1702.
[44] Y. Minezaki, K. Homma, K. Nishikawa, Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment, J. Mol. Biol. 368 (2007) 902–913.
[45] I. Stavropoulos, N. Khaldi, N.E. Davey, K. O'Brien, F. Martin, D.C. Shields, Protein disorder and short conserved motifs in disordered regions are enriched near the cytoplasmic side of single-pass transmembrane proteins, PLoS One 7 (2012) e44389.
[46] S. Jones, M. Stewart, A. Michie, M.B. Swindells, C. Orengo, J.M. Thornton, Domain assignment for protein structures using a consensus approach: characterization and analysis, Protein Sci. 7 (1998) 233–242.
[47] Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Res. 41 (2013) D43–D47.
[48] S. Mika, B. Rost, UniqueProt: creating representative protein sequence sets, Nucleic Acids Res. 31 (2003) 3789–3791.
[49] K. Gunasekaran, C.J. Tsai, R. Nussinov, Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers, J. Mol. Biol. 341 (2004) 1327–1341.
[50] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data, Bioinformatics (Oxford, England) 28 (2012) 3150–3152.
[51] D. Kozma, I. Simon, G.E. Tusnady, PDBTM: protein data bank of transmembrane proteins after 8 years, Nucleic Acids Res. 41 (2013) D524–D529.
[52] V. Vacic, V.N. Uversky, A.K. Dunker, S. Lonardi, Composition profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinforma. 8 (2007) 211.
[53] B. Xue, R.L. Dunbrack, R.W. Williams, A.K. Dunker, V.N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta 1804 (2010) 996–1010.
[54] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[55] R.P. Joosten, T.A. te Beek, E. Krieger, M.L. Hekkelman, R.W. Hooft, R. Schneider, C. Sander, G. Vriend, A series of PDB related databases for everyday needs, Nucleic Acids Res. 39 (2010) D411–D419.
[56] G. Pollastri, A. McLysaght, Porter: a new, accurate server for protein secondary structure prediction, Bioinformatics (Oxford, England) 21 (2005) 1719–1720.
[57] Y.P. Huang, J. Cheng, S.L. Zhang, L. Wang, J. Guo, Y. Liu, Y. Yang, L.Y. Zhang, G.Q. Bai, X.S. Gao, D. Ji, S.M. Lin, Q. Shao, Screening of hepatocyte proteins binding to F protein of hepatitis C virus by yeast two-hybrid system, World J. Gastroenterol. 11 (2005) 5659–5665.
[58] E.I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, G. Sherlock, GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes, Bioinformatics (Oxford, England) 20 (2004) 3710–3715.
[59] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R.C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeiffenberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard, H. Hermjakob, The IntAct molecular interaction database in 2012, Nucleic Acids Res. 40 (2012) D841–D846.
[60] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[61] B. Meszaros, I. Simon, Z. Dosztanyi, Prediction of protein binding regions in disordered proteins, PLoS Comput. Biol. 5 (2009) e1000376.
[62] G. Heijne, The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology, EMBO J. 5 (1986) 3021–3027.
[63] A. Elofsson, G. von Heijne, Membrane protein structure: prediction versus reality, Annu. Rev. Biochem. 76 (2007) 125–140.
[64] K. Gunasekaran, C.J. Tsai, S. Kumar, D. Zanuy, R. Nussinov, Extended disordered proteins: targeting function with less scaffold, Trends Biochem. Sci. 28 (2003) 81–85.
[65] Obradovic Romero, K. Dunker, Sequence data analysis for long disordered regions prediction in the calcineurin family, Genome Inform. Ser. Workshop Genome Inform. 8 (1997) 110–124.
[66] X. Li, P. Romero, M. Rani, A.K. Dunker, Z. Obradovic, Predicting protein disorder for N-, C-, and internal regions, Genome Inform. Ser. Workshop Genome Inform. 10 (1999) 30–40.
[67] H. Xie, S. Vucetic, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, V.N. Uversky, Z. Obradovic, Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions, J. Proteome Res. 6 (2007) 1882–1898.
[68] A. De Biasio, C. Guarnaccia, M. Popovic, V.N. Uversky, A. Pintar, S. Pongor, Prevalence of intrinsic disorder in the intracellular region of human single-pass type I proteins: the case of the notch ligand Delta-4, J. Proteome Res. 7 (2008) 2496–2506.
[69] M.S. Almen, K.J. Nordstrom, R. Fredriksson, H.B. Schioth, Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin, BMC Biol. 7 (2009) 50.
[70] P. Hubert, P. Sawma, J.P. Duneau, J. Khao, J. Henin, D. Bagnard, J. Sturgis, Single-spanning transmembrane domains in cell growth and cell-cell interactions: more than meets the eye? Cell Adhes. Migr. 4 (2010) 313–324.
[71] S. Vucetic, H. Xie, L.M. Iakoucheva, C.J. Oldfield, A.K. Dunker, Z. Obradovic, V.N. Uversky, Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions, J. Proteome Res. 6 (2007) 1899–1916.
[72] J. Gsponer, M.E. Futschik, S.A. Teichmann, M.M. Babu, Tight regulation of unstructured proteins: from transcript synthesis to protein degradation, Science (New York, N.Y.) 322 (2008) 1365–1368.
[73] A. Patil, H. Nakamura, Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks, FEBS Lett. 580 (2006) 2041–2045.
[74] A. Patil, K. Kinoshita, H. Nakamura, Domain distribution and intrinsic disorder in hubs in the human protein–protein interaction network, Protein Sci. 19 (2010) 1461–1468.
[75] C.J. Oldfield, J. Meng, J.Y. Yang, M.Q. Yang, V.N. Uversky, A.K. Dunker, Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, BMC Genomics 9 (Suppl. 1) (2008) S1.
[76] H. Jeong, S.P. Mason, A.L. Barabasi, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.
[77] Z. Dosztanyi, J. Chen, A.K. Dunker, I. Simon, P. Tompa, Disorder and sequence repeats in hub proteins and their implications for network evolution, J. Proteome Res. 5 (2006) 2985–2995.
[78] R.R. Vallabhajosyula, D. Chakravarti, S. Lutfeali, A. Ray, A. Raval, Identifying hubs in protein interaction networks, PLoS One 4 (2009) e5344.
[79] J.D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J. Walhout, M.E. Cusick, F.P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein–protein interaction network, Nature 430 (2004) 88–93.
[80] G.P. Singh, M. Ganapathi, D. Dash, Role of intrinsic disorder in transient interactions of hub proteins, Proteins 66 (2007) 761–765.
[81] D. Ekman, S. Light, A.K. Bjorklund, A. Elofsson, What properties characterize the hub proteins of the protein–protein interaction network of *Saccharomyces cerevisiae*? Genome Biol. 7 (2006) R45.
[82] K. Shimizu, H. Toh, Interaction between intrinsically disordered proteins frequently occurs in a human protein–protein interaction network, J. Mol. Biol. 392 (2009) 1253–1265.
[83] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics (Oxford, England) 21 (2005) 3433–3434.
[84] E. Mossessova, L.C. Bickford, J. Goldberg, SNARE selectivity of the COPII coat, Cell 114 (2003) 483–495.
[85] A. Bracher, W. Weissenhorn, Structural basis for the Golgi membrane recruitment of Sly1p by Sed5p, EMBO J. 21 (2002) 6114–6124.
[86] W.L. DeLano, The Pymol Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA, 2002.

[87] S. Koshiba, H. Li, Y. Motoda, T. Tomizawa, T. Kasai, N. Tochio, T. Yabuki, T. Harada, S. Watanabe, A. Tanaka, M. Shirouzu, T. Kigawa, T. Yamamoto, S. Yokoyama, Structural basis for the recognition of nucleophosmin-anaplastic lymphoma kinase oncoprotein by the phosphotyrosine binding domain of Suc1-associated neurotrophic factor-induced tyrosine-phosphorylated target-2, J. Struct. Func. Genom. 11 (2010) 125–141.

[88] M. Inoue, T. Shiba, K. Ihara, Y. Yamada, S. Hirano, H. Kamikubo, M. Kataoka, M. Kawasaki, R. Kato, K. Nakayama, S. Wakatsuki, Molecular basis for autoregulatory interaction between GAE domain and hinge region of GGA1, Traffic (Copenhagen, Denmark) 8 (2007) 904–913.

[89] T. Shiba, M. Kawasaki, H. Takatsu, T. Nogi, N. Matsugaki, N. Igarashi, M. Suzuki, R. Kato, K. Nakayama, S. Wakatsuki, Molecular mechanism of membrane recruitment of GGA by ARF in lysosomal protein transport, Nat. Struct. Biol. 10 (2003) 386–393.

[90] R. Mattera, R. Puertollano, W.J. Smith, J.S. Bonifacino, The trihelical bundle subdomain of the GGA proteins interacts with multiple partners through overlapping but distinct sites, J. Biol. Chem. 279 (2004) 31409–31418.

[91] J. McCullough, R.D. Fisher, F.G. Whitby, W.I. Sundquist, C.P. Hill, ALIX-CHMP4 interactions in the human ESCRT pathway, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 7687–7691.