


Systematic Prediction of FFAT Motifs Across Eukaryote Proteomes Identifies Nucleolar and Eisosome Proteins With the Predicted Capacity to Form Bridges to the Endoplasmic Reticulum

Contact
Volume 2: 1–21
© The Author(s) 2019
DOI: 10.1177/2515256419883136
journals.sagepub.com/home/ctc


John A. Slee¹ and Timothy P. Levine² 

Abstract

The endoplasmic reticulum (ER), the most pervasive organelle, exchanges information and material with many other organelles, but the extent of its interorganelle connections and the proteins that form bridges are not well known. The integral ER membrane protein vesicle-associated membrane protein-associated protein (VAP) is found in multiple bridges, interacting with many proteins that contain a short linear motif consisting of “two phenylalanines in an acidic tract” (FFAT). The VAP-FFAT interaction is the most common mechanism by which cytoplasmic proteins, particularly interorganelle bridges, target the ER. Therefore, predicting new FFAT motifs may both find new individual peripheral ER proteins and identify new routes of communication involving the ER. Here, we searched for FFAT motifs across whole proteomes. The excess of eukaryotic proteins with FFAT motifs over background was $\geq 0.8\%$, suggesting that this is the minimum number of peripheral ER proteins. In yeast, where VAP was previously known to bind 4 proteins with FFAT motifs, a detailed analysis of a subset of proteins predicted 20 FFAT motifs. Extrapolating these findings to the whole proteome estimated the number of FFAT motifs in yeast at approximately 50 to 55 (0.9% of proteome). Among these previously unstudied FFAT motifs, most have known functions outside the ER, so could be involved in interorganelle communication. Many of these can target well-characterized membrane contact sites; however, some are in nucleoli and eisosomes, organelles previously unknown to have molecular bridges to the ER. We speculate that the nucleolar and eisosomal proteins with predicted motifs may function while bridging to the ER, indicating novel ER–nucleolus and ER–eisosome routes of interorganelle communication.

Keywords

endoplasmic reticulum, lipid transfer protein, membrane contact sites, nucleolus, short linear motif, eisosome, VPS13, NOP2/Nsun6, BOPI, UBR4

Introduction

The endoplasmic reticulum (ER) is the most widely distributed organelle in eukaryotic cells. Its functions are created in part by peripheral membrane proteins that reversibly bind its cytoplasmic surface. The sole widely documented mechanism by which cytoplasmic proteins target the ER is through a short linear motif (SLiM) called “two phenylalanines in an acidic tract” (FFAT; Loewen et al., 2003). Proteins with FFAT motifs communicate between the ER and many other compartments, including plasma membrane, Golgi apparatus, peroxisomes, and mitochondria (Hanada et al., 2003; Hynynen et al., 2005; Saita et al., 2009; De Vos et al.,

2011; Murphy and Levine, 2016; Costello et al., 2017; Hua et al., 2017; Yadav et al., 2018). This fits the idea that SLiM interactions allow links to arise between

¹The Limes, Sawbridgeworth, UK

²UCL Institute of Ophthalmology, London, UK

Received February 25, 2019. Revised September 16, 2019. Accepted September 17, 2019.

Corresponding Author:

Timothy P. Levine, UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK.

Email: tim.levine@ucl.ac.uk



seemingly distinct subcellular processes (Kim et al., 2014). Significantly, all proteins in a non-ER location with FFAT motifs have since been found to target both that location and the ER at the same time. They bridge between the two compartments at membrane contact sites (Murphy and Levine, 2016). Thus, finding a FFAT motif is significant because it suggests direct bridging to the ER, rather than the existence of a long-distance intracellular shuttle between two compartments (Kumagai and Hanada, 2019).

FFAT motifs bind to the major sperm protein (MSP, 120 residues) domain of vesicle-associated membrane protein-associated protein (VAP), a protein found in almost all eukaryotes. The MSP domain is linked by a flexible region to the cytoplasmic face of the ER (Skehel et al., 2000). The core of FFAT motifs is an extended peptide of seven residues, two of which bind into pockets in VAP (Figure 1(a); Kaiser et al., 2005). Prior to the engagement of the core, an acidic tract N-terminal to the core has a low-affinity charge-based interaction with the MSP domain (Furuita et al., 2010). All FFAT motifs that fitted the early, strict definition, have since been validated as interacting with VAP, except for one that is located in the lysosomal lumen. Subsequently, predicting new FFAT motifs has been made possible because the definition of FFAT motifs has been expanded through the discovery of new variants (Saita et al., 2009; Mikitova and Levine, 2012; Baron et al., 2014). Further sequence variation has been shown with substitution of anionic positions with serine or threonine (S/T in one letter notation), which allows reversible phosphomimicking of motifs (Alpy et al., 2013; Kumagai et al., 2014; McCune et al., 2017; Johnson et al., 2018; Kirmiz et al., 2018).

Domain-SLiM interactions are hard to capture by standard protein-protein interaction techniques, and their discovery rate in databases is estimated at only $\leq 4\%$ (Davey et al., 2017). Bioinformatics can predict new instances (Krystkowiak and Davey, 2017). Widening the definition of FFAT's pattern allowed the prediction of many new human motifs, which were assessed against three criteria to reduce the number of false positives: (a) cytoplasmic location; (b) structure allowing formation of an extended loop, also described as an intrinsically disordered region (IDR); and (c) local conservation (Mikitova and Levine, 2012). Many of the FFAT motifs identified by bioinformatics have been verified in high-throughput protein interactomes (Hein et al., 2015; Huttlin et al., 2015), and some have been verified by specific mutagenesis (Costello et al., 2017; Kumar et al., 2018). To identify motifs that vary at multiple positions, a position weight matrix (PWM) can be used. The PWM for FFAT motifs produces a FFAT score for runs of 13 amino acids (7 core and 6 acidic tract preceding; Murphy and Levine, 2016). FFAT

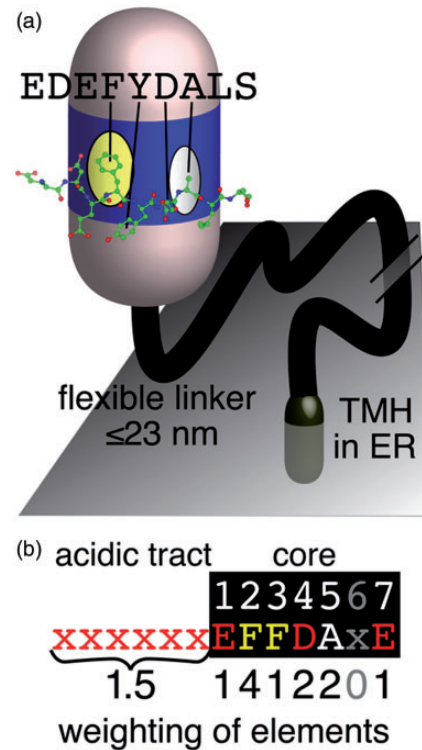


Figure 1. The conserved VAP-FFAT interaction. (a) A FFAT motif forms an extended loop that binds to a conserved basic face of the MSP domain of VAP. VAP also has a linker of approximately 100 residues that includes a coiled-coil section in some species and can maximally extend to $\geq 20\text{ nm}$ (not shown), and a tail-anchoring TMH in the ER membrane (Murphy and Levine, 2016). The basic face of VAP's MSP domain contains two pockets: one of which binds an aromatic residue (F/Y) and the other binds a small non-charged residue (A, C, G, S, and T). These are the second and fifth residues, respectively, of the seven-residue core of the FFAT motif. Peptide from ORP1L FFAT peptide crystallized in complex with VAP in PDB:1Z9O (Kaiser et al., 2005). (b) The weights of the seven elements scored by the FFAT motif. An ideal motif (six upstream = acids, then seven core residues = E F F D A X E) has FFAT score = 0. Suboptimal substitutions at each position in the motif are assessed by the PWM: six elements are all the positions in the core except one, the seventh element is the overall acidity of the upstream residues (Murphy and Levine, 2016). TMH = transmembrane helix; ER = endoplasmic reticulum.

score varies from “perfect” motifs with FFAT score = 0, increasing with each suboptimal substitution (Figure 1 (b); Murphy and Levine, 2016). Three sources of information contributed to the PWM: (a) residues in VAP interactors and their homologues, (b) the effect of substitutions on affinity for VAP, and (c) VAP-FFAT structures (Mikitova and Levine, 2012). A cutoff FFAT score ≤ 2.5 was chosen to reduce false positives, initially estimated at approximately 1% for this cutoff (Murphy and Levine, 2016). This PWM has been used to scan interactors of the two major human VAPs: VAPA and VAPB. These are 77% similar by sequence and share

80% of interactors (Huttlin et al., 2015), so they can be considered together. Forty percent interactors contained a motif with FFAT score ≤ 2.5 , and 10% formed complexes with these, so 50% of VAP interactions are attributable to FFAT motifs (Murphy and Levine, 2016).

Here, we applied the PWM for FFAT motifs systematically across eukaryote proteomes using negative controls to estimate background. We predict that at least 0.8% of the proteome (50 proteins in budding yeast and 170 in humans) bind VAP. This estimate was confirmed by detailed consideration of a subset of yeast motifs. Some of the FFAT-positive proteins we predict act in compartments that have no known molecular link to the ER. As most known FFAT-positive proteins carry out their function while binding VAP, this indicates novel pathways of interorganelle communication.

Results

Proteins With Highly Optimal FFAT Motifs Are Already Known to Bind VAP

Every residue of every protein in six model eukaryotes (human, *Drosophila melanogaster*, *Caenorhabditis elegans*, budding yeast *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, and *Plasmodium falciparum*) was scanned with the PWM we developed previously to identify FFAT scores for every sequence of 13 residues. The position with the lowest FFAT score in each protein was recorded as the FFAT score for that protein (Murphy and Levine, 2016). The overall distribution of FFAT scores was non-normal, and was similar across eukaryotes (Figure 2(a), and data not shown). Looking at proteins with extremely low FFAT scores, where VAP interactors are expected, the proteomes showed some variation, for example, *Arabidopsis* had fewer proteins with FFAT scores ≤ 1 (Figure 2(b)).

We next determined the correlation between low FFAT score and documented physical interaction with VAP, combining data from the two species with the best documented protein-protein interaction networks: human and yeast. For the human VAP interactome, we included VAPA, VAPB, and also MOSPD2. The latter is a newly described variant of VAP known to bind only a minority of the interactors of VAPA/B, possibly because it has not been studied in as much detail, or because the region of its MSP domain that binds FFATs is divergent from VAPA/B (Di Mattia et al., 2018). The three VAPs have 341 documented physical interactors (1.7% of proteome; Chatr-Aryamontri et al., 2017). Yeast has two VAPs, the major form Scs2p and a minor homologue Scs22p, with 74 physical interactors between them (1.2% of proteome). All of the proteins with FFAT score = 0.0 (two in yeast, six in human) are already known to bind VAP. This is to be expected

because the sequences of proteins with the lowest FFAT scores were used to create our PWM, so by definition their FFAT scores are low. The two yeast proteins in this group are the transcriptional regulator Opilp and the oxysterol binding protein-related protein (ORP) homologue Osh1p (Table 1(a) and (b)). For FFAT scores marginally higher (range 0.5–1.0), 2 of the 6 yeast proteins (Table 1(a) and (b)) and 13 of the 17 human proteins are documented to bind VAP (65% overall, Figure 2(c)). The motifs in the four missing human proteins (FRPD1, S22AF, TSYL2, and ORC2) were all identified previously by SCAN-PROSITE (Mikitova and Levine, 2012). The four missing yeast proteins (Bbc1p, Kri1p, Seg2p, and Ypr097wp; Table 1) are considered later.

Few of the Many Proteins With Suboptimal FFAT Motifs Are Known to Bind VAP

The number of proteins containing FFAT motifs that score 1.5, 2.0, or 2.5 rose steeply, so that across the six model eukaryotes 3.5% of proteins have suboptimal FFAT motifs with such scores (Figure 2(b) and Supplementary Tables 1A, 2, and 3). With higher FFAT scores, the proportion already identified as interacting with VAP declined: 11%, 6.7%, and 2.7% for FFAT score = 1.5, 2.0, and 2.5, respectively (Figure 2(d)). This could reflect a genuine decrease in the proportion of proteins with these motifs that ever binds VAP. Alternately, the copurification and coprecipitation methods of major high-throughput studies (Hein et al., 2015; Huttlin et al., 2015) could be too harsh to identify interactions between VAP and suboptimal motifs (Davey et al., 2017). Interactions between domains and SLiMs tend to have lower affinity than the interactions between domain pairs (Kim et al., 2014).

Estimating the Proportion of Proteins With Suboptimal FFAT Motifs That Interact With VAP

Among the eukaryotic proteins that have suboptimal FFAT motifs, we used two approaches to estimate the proportion that are irrelevant, equivalent to random background. In the first approach, we analyzed species without VAP, which have no selective pressure for FFAT motifs, so they may act as negative controls. Organisms from the two noneukaryotic kingdoms (bacteria and archaea) have been used before for this purpose (Meszaros et al., 2012). In addition, we identified a small group of variant unicellular eukaryotes that lack VAP: *Euglena gracilis*, *Spiroplasma salmonicida*, and *Paratrypanosoma confusum*. The levels of suboptimal FFAT motifs were lowest in bacteria and similar in archaea and VAP-negative eukaryotes. For FFAT scores ≤ 1.0 , the controls had approximately 10%

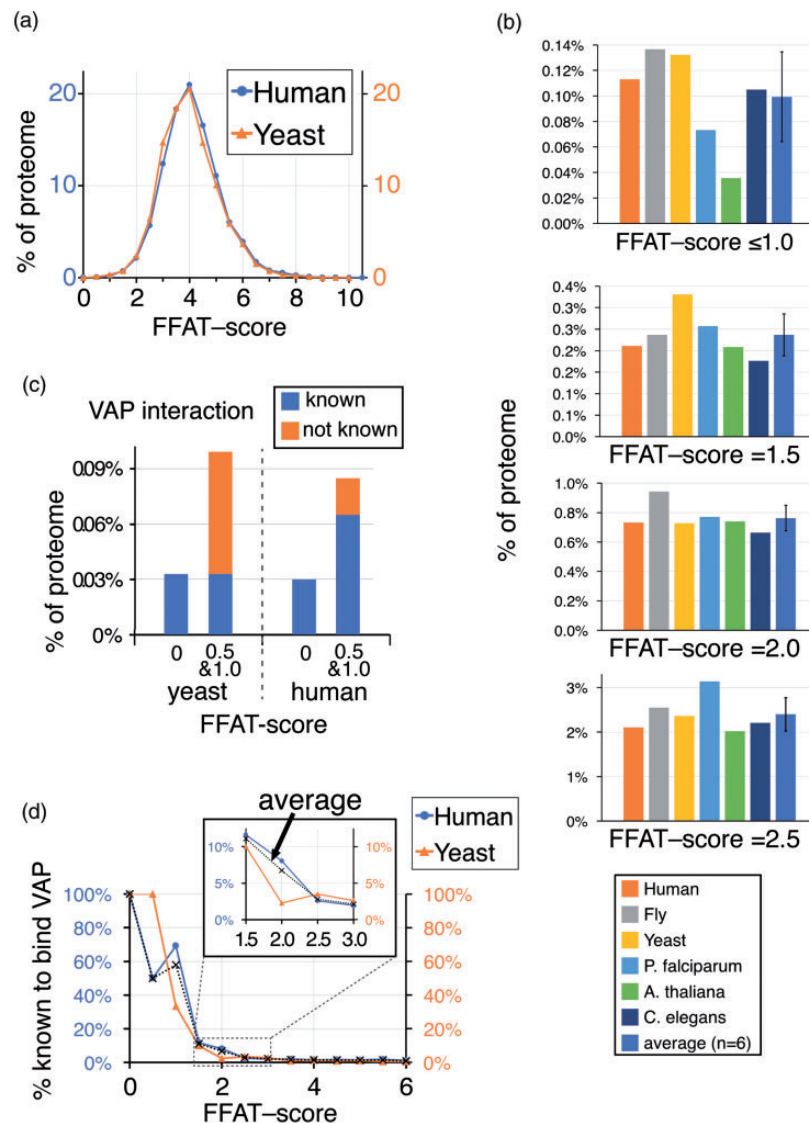


Figure 2. Three percent of eukaryotic proteins have FFAT scores ≤ 2.5 . (a) Distributions of FFAT scores in human and yeast (*Saccharomyces cerevisiae*). FFAT scores are assigned as described in “Methods” section. The means and standard deviations are yeast: 4.6 ± 1.1 , human 4.5 ± 1.1 ; the distributions are statistically similar (χ^2 0.998). (b) Proportion of six eukaryotic proteomes at different levels of low FFAT score, from highly optimal (top, range 0.0–1.0) to less optimal (bottom, =2.5), including the average, unweighted for proportion of proteins in each of the 6 proteomes, and standard deviation. (c) Proportions of yeast and human proteins with motifs with highly optimal FFAT scores (range 0.0–1.0) showing if they are known interactors of VAP. (d) Proportion of proteins identified as binding VAP in yeast and human according to their FFAT scores (range 0.0–6.0). Inset shows detail in the range FFAT scores 1.5 to 3.0. Dotted line indicates the average across the two species. Identification of proteins as VAP interactors declines rapidly as FFAT score increases: 59% for FFAT score ≤ 1.0 (human 11/17, yeast 2/5), 8% for $1.0 < \text{FFAT score} \leq 2.0$ (human 17/180, yeast 3/62), and 3% for $2.0 < \text{FFAT score} \leq 3.0$ (human 11/440, yeast 5/146). FFAT = two phenylalanines in an acidic tract; VAP = vesicle-associated membrane protein-associated protein.

observed motifs as VAP-positive eukaryotes. This proportion rose to 40% to 50% for motifs with FFAT score = 2.5 (Figure 3(a)).

The second estimation of the background level of low FFAT scores was achieved by randomizing VAP-negative proteomes. To check this randomization acted as expected, we compared the numbers of motifs before and after randomization in VAP-negative proteomes,

with the expectation that there would be no difference. Four of the 13 proteomes had more motifs after randomization. Among the nine others, we tested for an excess of motifs over background using the “ $N-1$ ” χ^2 test (Campbell, 2007) to compare total number of motifs at different levels of FFAT score: There was no significant excess of observed motifs over expected (lowest p value = 0.08, Supplementary Table 1B). By

Table 1. Yeast Proteins With Motifs With FFAT Scores ≤ 1.5 .

Name	FFAT score	Start	Flank	Core	End
(a) Description of motifs					
OPI1 ^a	0	194	EDDDDE	EFFDASE	206
OSHI ^a	0	710	EDSDAD	EFFDAEE	722
BBC1	1	547	EDTDDH	EFEDAND	559
KRI1	1	243	NEEDDE	EFEDAAE	255
OSH2 ^a	1	739	EASDAD	EFYDAAE	751
OSH3 ^a	1	508	YLEND	EFFDAEE	520
SEG2	1.2	637	DDDDDD	EYHDSYD	649
YPR097W	1	449	EESDFD	EYKDASD	461
ASG7	1.5	121	SDSNSE	EYYESKD	133
EDE1	1.5	1,130	SSSDDD	EFEDTRE	1,142
ERB1	1.5	46	SDEDDD	EYESAVE	58
EXO1	1.5	541	DGDTSE	DYSETAE	553
MCM3	1.5	746	IDEEES	EYEEALS	758
NOP2	1.5	254	SPAEAM	EFFEANE	266
OST1	1.5	59	ASEPAT	EYFTAFE	71
PET123	1.5	224	FNDETE	EFTDAYD	236
RQC1	1.5	149	DDTNEE	GFFTASE	161
SFB3	1.5	747	LEETDL	TFYDAND	759
SKP2	1.5	458	SLDTEd	DFDDCNS	470
STE13	1.5	554	EYETDT	IFFTANE	566
TUB1	1.5	412	EGMEEG	EFTEARE	424
TUB3	1.5	412	EGMEEG	EFTEARE	424
UBP10	1.5	139	EEEEGE	IFHEARD	151
UBR1	1.5	1,608	EVEEEL	EFEDTAE	1,620
VID27	1.5	193	LDSSSD	DFQDAKD	205
WSC2	1.5	484	VSDGDD	DYDDAKD	496
YLR149C	1.5	605	DVFEDD	EYYEAYN	617
YOR238W	1.5	280	GIEDDE	EYFETKI	292

(b) Descriptions of proteins

FFAT score 0.0

^b*Opi1p*: transcription factor that senses phosphatidic acid in ER to regulate phospholipid pathways

^b*Osh1p*: lipid transfer protein in the ORP family found in all eukaryotes; 64% similar to paralogue *Osh2p* (below).

Other name: *Swh1p*

FFAT score 1.0

Bbc1p: assembles actin patches in clathrin-mediated endocytosis; only few homologues

Kri1p: nucleolus; synthesis of 40S ribosomal subunit; widespread homologues incl. human.

^b*Osh2p*: ORP lipid transfer protein; 64% similar to paralogue *Osh1p* (above)

^b*Osh3p*: ORP lipid transfer protein with GOLD domain, a combination only found in fungi

Seg2p: stabilizes eisosomes, which are fungal and algal protein-rich assemblies that deform the plasma membrane into furrows (Moreira et al., 2012); no OrthoDB family

Ypr097wp: PUF found in purified mitochondria with PX and PXB domains—the same domain structure as human sorting nexins SNX20/21, indicating possible endocytic sorting function (Clairfeuille et al., 2015; Danson et al., 2018).

FFAT score 1.5

Asg7p: regulates signaling from G beta subunit *Ste4p* in mating

Ede1p: scaffolds clathrin-mediated endocytosis, binding membrane proteins with a C-terminal UBA domain; homologue of human *Eps15*, which has a C-terminal UIM domain.

Erb1p: nucleolus; 60S ribosomal subunit maturation; homologues in all eukaryotes called *BOPI*

Exo1p: 5'-3' exonuclease and flap-endonuclease involved in many pathways

Mcm3p: subunit of the mini chromosome maintenance DNA helicase complex

Nop2p: nucleolus; rRNA-methyltransferase of 25S rRNA, required for ribosomal maturation

Ost1p: N-linked oligosaccharyltransferase complex in ER lumen

Pet123p: component of 37S mito-ribosomal small subunit

Rqc1p: component of ribosomal quality control pathway; functional homologues in all eukaryote kingdoms share domain called *Tcf25*; mammalian components bind *VAP* (Zuzow et al., 2018)

(continued)

Table 1. Continued

(b) Descriptions of proteins

Sfb3p: homologous to Sec24, forming an alternate coat for COPII vesicles in heterodimers with Sec23p

Skp2p: Skp, Cullin, F-box containing (SCF) ubiquitin ligase complex subunit

Ste13p: Golgi-localized peptidyl aminopeptidase

Tub1p: alpha-tubulin; 95% similar to paralogue Tub3p

Tub3p: alpha-tubulin; 95% similar to paralogue Tub1p

Ubp10p: nucleolus; Ub-specific peptidase in rRNA synthesis; human orthologue is USP36

Ubr1p: N-recogin E3 Ubiquitin ligase; degrades proteins by the N-end rule; homologous to UBR4 in most non-fungal eukaryotes

Vid27p: implicated in the VID pathway for lysosomal protein degradation; homologues in all eukaryote kingdoms except animals

Wsc2p: type I plasma membrane protein involved in sensing stress (Wilk et al., 2010)

Yrl149cp: cytosolic PUF consisting almost mostly of WD40 repeats

Yor238wp: cytosolic PUF; predicted by HHsearch^c to bind S-adenosyl methionine like YdcF.

Note. FFAT = two phenylalanines in an acidic tract; ER = endoplasmic reticulum; VAP = vesicle-associated membrane protein-associated protein; ORP = oxysterol binding protein-related protein; PUF = protein of unknown function; VID = vacuolar import and degradation.

^aFour motifs already known.

^bFour proteins already shown to bind VAP.

^cHHsearch was used to predict domains for all yeast proteins, using settings described previously (Fidler et al., 2016).

comparison, all VAP-positive genomes had excess motifs with p values varying from 10^{-4} to 10^{-20} (Supplementary Table 1B, bottom). This showed that the randomization is a useful way to estimate background.

The background from randomization of VAP-positive proteomes was 76% for all FFAT score ≤ 2.5 , mainly because it was 85% for FFAT score = 2.5 (Figure 3(b)). This background is higher than the background estimated from VAP-negative organisms, and even this more conservative prediction of FFAT motifs still produces a minimum estimate that approximately 0.8% of proteins are positively selected above background to have motifs with FFAT score ≤ 2.5 (50 in yeast and 170 in human). The two background estimations agreed that the proportion of VAP interactors is higher where the FFAT score is lower, with rates of false positives starting low but rising steeply as FFAT score increases from 1.5 to 2.5 (40%–85%, respectively, Figure 3(c)).

Using Families of Fungal Orthologues to Determine Whether the Motif in an Individual Protein Is Subject to Preferential Conservation

To study VAP interactors in one organism, we chose yeast because its proteome is small and better documented than other eukaryotes. Inclusion criteria for determining if motifs are functionally relevant have been developed previously: (a) location in the cytoplasm, not inside an organelle or secreted; (b) predicted to form an extended loop, not in a helix or sheet, particularly if

the secondary structural element is conserved and in a globular domain and if it has been confirmed experimentally (crystal structure, etc.); and (c) preferential conservation of motifs that are not excluded by the other two criteria among closely related homologues, ideally orthologues (Mikitova and Levine, 2012; Murphy and Levine, 2016).

This third criterion was previously carried out by manual examination of homologues, particularly of the key core residues of the motif (#2/4/5). Here, we developed the criterion to measure conservation statistically. Families of orthologues of each yeast protein have been created by tools such as OrthoDB (family size $n = 18$ –700, median 150, Table 2(a); Kriventseva et al., 2015). We tested if the rate of occurrence of motifs among orthologues was significantly above background, which was calculated from randomized sequences, as done for proteomes above. Based on the statistical robustness of randomizing VAP-negative control proteomes (Supplementary Table 1B), our cutoff for statistical significance of excess of motifs with FFAT at different scores ($\leq 2.5/2.0/1.5/1.0$) was $p \leq .01$, with borderline excess where p values lie between .001 and .01, and confirmed excess for $p \leq .001$.

We first tested this approach on the four yeast proteins with known FFAT motifs: Osh1/2/3p and Opi1p. As these sequences were used to create our PWM, by definition, their FFAT scores are low, nevertheless they illustrate the use of families of orthologues. For Osh1p (FFAT score = 0.0), which includes Osh2p (FFAT score = 1.0) as a paralogue in the same family, 99% of orthologues had motifs with FFAT score ≤ 2.5 , indicating

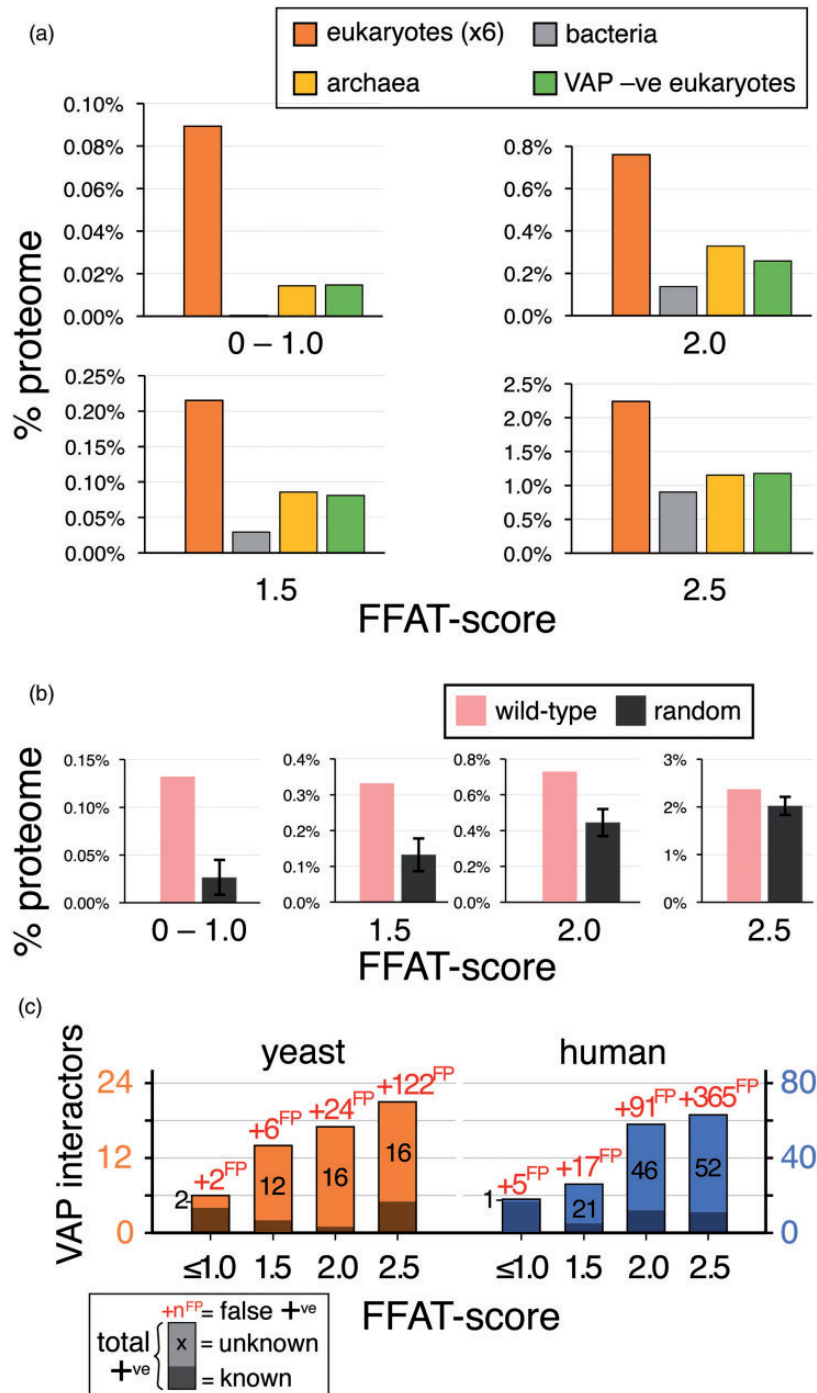


Figure 3. Comparison of the observed level of motifs with low FFAT scores with expected levels. (a) Levels of motifs with low FFAT scores in groups of bacteria and archaea, and VAP-negative eukaryotes across a range of low FFAT scores, from highly optimal (0.0–1.0, top left) to less optimal (=2.5, bottom right). In each case, average levels in eukaryotes are included for comparison. For species names, see “Methods” section. All data points are the average across all proteins in that class. (b) Levels of motifs with low FFAT scores in randomized yeast proteins (10 randomizations, showing means and standard deviations) compared with observed levels in yeast, from highly optimal (0.0–1.0, left) to less optimal (=2.5, right). (c) Predicted total number of FFAT motifs in yeast and human proteomes, based on the background of false positives (numbers in red as “xyz^{FP}”) calculated from randomization data. Motifs are subdivided into those already known (dark shading) and those predicted (light shading, black numbers). FFAT = two phenylalanines in an acidic tract; VAP = vesicle-associated membrane protein-associated protein.

uniform conservation of the VAP-Osh1/2 interaction. By comparison, only 74% of Osh3p orthologues (FFAT score = 1.0) had motifs with FFAT score ≤ 2.5 . Another 15% were short sequences representing fragments of longer Osh3p sequences that were readily identifiable in sequence databases. All of those contained motifs with low FFAT scores (data not shown). The remaining 10% Osh3 orthologues all had degenerate FFAT-like motifs at the same position as the motif in Osh3p. For example, *Meyerozyma guilliermondii* has GFVDAQD without acids upstream. Motifs with tryptophan substituted at Position 2 (which cannot arise from a point mutation of F or Y) were found both in Taphrinomycotina, a branch of the Ascomycota (*Pneumocystis*, *Schizosaccharomyces pombe*, etc.) and in Agaricomycetes, a branch of the Basidiomycota (*Fomitopsis*, *Coprinopsis*, etc.), but not in the closely related Ustilaginomycotina. As these Position 2 = tryptophan FFAT-like motifs are evolutionarily separated (Stajich et al., 2009), this substitution appears to have evolved in Osh3 orthologues on multiple occasions.

In the family of orthologues of Opi1p (FFAT score = 0.0), only 13% of proteins had a motif with FFAT score ≤ 2.5 . This created a statistical excess ($p = 10^{-4}$ for motifs with scores ≤ 2.0) of far less statistical significance than for Osh3 ($p = 10^{-26}$) or Osh1 ($p = 10^{-44}$). Many of the Opi1 orthologues have an alternate name: “related to clock-controlled protein 8 (ccg-8).” ccg-8 is a *Neurospora crassa* transcription factor that regulates an aspect of lipid metabolism different from that regulated by Opi1p in *S. cerevisiae* (Xue et al., 2019). FFAT motifs are absent from all proteins named ccg-8. This indicates that the fungal orthologue families have subfamilies with diverged function. To address this, we first looked at a narrower orthology family from the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology database (Kanehisa et al., 2004), excluding all ccg-8s. This had a higher incidence of FFAT motifs (28%), but the smaller size of the family meant that there was no greater statistical significance (Supplementary Table 4). We then created a custom-made family using BLAST to identify proteins closer to the budding yeast Opi1p, but with only 15% to 20% overall sequence identity, and no selection for the FFAT motif in the BLAST search. These proteins had a 48% incidence of motifs with FFAT scores ≤ 2.5 (most FFAT scores = 0.0) and a much higher statistical significance ($p = 10^{-10}$, Supplementary Table 4).

Overall, this indicates that low discovery rates of a motif among Opi1p orthologues did not result from a lack of sensitivity (i.e., not false negatives). Instead, there appears to have been evolutionary pressure to interact with VAP only in some ecological niches, reducing the statistical significance of excess motifs in a family that covers a wide range of organisms.

Ten New Predicted VAP Interactors in 23 Yeast Proteins Identified With FFAT Scores 1.0 or 1.5

To predict new FFAT motifs, first we considered proteins with FFAT score = 1.0 ($n = 4$) and = 1.5 ($n = 20$; Table 1 (a) and (b)). Among these 24 proteins, Erb1p had previously been identified as an interactor of yeast VAP (Ho et al., 2002). This protein contained a motif that met all our criteria ($p = 10^{-7}$), strongly suggesting a VAP-FFAT interaction in this family. Among the newly identified 23 proteins, our criteria excluded 13 of the candidate proteins with motifs with FFAT scores = 1.5 (Table 2(a)). Seven proteins were excluded by location or structure of the region containing the motif. Five of these had weak (or no) conservation (Asg7p, Exo1p, Ost1p, Pet123p, and Stel3p; Table 2(a)). In contrast, two highly homologous proteins (Tub1p and Tub3p) were excluded because the motifs are in a helix even though the motifs were universal among their orthologues. All α -tubulins, including in humans, apparently have the same inaccessible FFAT-like sequence (Murphy and Levine, 2016).

Six proteins were excluded because they failed the conservation criterion. Two of these had fewer motifs with low FFAT scores in the orthologue family than expected: Mcm3p and Yor238wp (Table 2(a)). Three had low numbers of motifs similar to the expected numbers (Skp2p, Wsc2p, and Ylr149cp). The final protein, Sfb3p, had a nonsignificant excess of motifs with FFAT score 1.5 to 2.5 within a large family of orthologues (7% of 274, $p = 0.02 = 10^{-1.7}$, Table 2(a)). Limiting Sfb3 orthologues to the narrower group of Ascomycota increased the rate of occurrence of motifs with low FFAT scores but reduced the significance (16% of 43, $p = 0.06 = 10^{-1.2}$, Supplementary Table 4), so there is a statistical trend for an excess of motifs in Sfb3 orthologues, but they fail to meet the statistical criterion.

After these exclusions, 10 new proteins with FFAT score = 1.0 or 1.5 met all the criteria for new predictions of FFAT motifs: Bbc1p, Ede1p, Kri1p, Nop2p, Rqc1p, Seg2p, Ubp10p, Ubr1p, Vid17p, and Ypr097wp (Table 2 (a) and described later). The positive prediction rates for motifs with FFAT scores = 1.0 and 1.5 were 100% and 32%, respectively. The latter is lower than the frequency of motifs we had predicted (60%), but with the low numbers involved the difference is not statistically significant.

Three of the predictions only reached borderline statistical significance (p value between .001 and .01): Seg2p, Ubp10p, and Nop2p. For Seg2p, the low significance was most likely because the protein family is small ($n = 21$). Ubp10p orthologues showed a small excess of motifs over background with FFAT scores 1.5 to 2.0 in a large orthologue family. Like Opi1p, the motif showed more significant conservation in a subfamily of Ubp10 orthologues from Ascomycota only (Supplementary Table 4). Nop2 was unique among the proteins we

studied in having two types of motifs with low FFAT scores in different orthologues: The vast majority (97%), including in *S. cerevisiae*, had a motif in a helix near the start of the RNA methyltransferase domain (Yakubovskaya et al., 2012). A minority (12%) had a second motif with FFAT score as low as 1.0 in a highly anionic IDR, and one orthologue had a motif after the methyltransferase domain. After excluding motifs with helical structure, the minority in IDRs were still significantly in excess over background (Table 2(a)).

This predicts that Nop2 orthologues functionally interact with VAP. This implies that the potential motif in Nop2p is conserved to interact with VAP (see “Discussion” section).

These 10 proteins with putative FFAT motifs, together with Erb1p, are implicated in diverse cellular activities (Figure 4). Two (Rqc1p and Ubr1p) have diverse cytoplasmic functions that are already linked to the ER. Four are involved in processes that have previously been linked to VAP-FFAT interactions: (a) Bbc1p and

Table 2. Application of Exclusion Criteria to Potential FFAT Motifs in Yeast.

(a) Motifs with FFAT-scores 0.0 to 1.5

Name	FFAT-score	Exclusion criteria															
		1. loc if not cyto.	2. struct if not IDR	FFAT-score ≤2.5	size of fam.	3. Conservation (family of orthologues)											
						observed motifs				expected background (random)				p-value (-log ₁₀)			
						2.5	2.0	1.5	0.0-1.0	2.5	2.0	1.5	0.0-1.0	≤2.5	≤2.0	≤1.5	0.0-1.0
Op1 ^{er}	0		H ^b	13%	151	1	3	2	13	2.2	0.7	0.0	0.0	3.2	4.2	4.0	3.5
Osh1 ^{er}	0			99%	154	55	26	3	149	9.2	2.7	0.5	0.1	44	38	34	34
Bbc1	1		H ^b	20%	168	15	8	4	7	7.0	1.6	0.4	0.0	3.9	3.7	2.8	2.1
Kri1	1		H ^b	23%	154	14	3	5	19	12	5.0	1.1	0.4	2.4	3.4	5	4.6
Osh2 ^{er}	1		H ^b			as Osh1											
Osh3 ^{er}	1			74%	156	13	31	23	62	6.6	1.2	0.0	0.0	24	26	20	15
Seg2	1.2			33%	21	3	1	1	2	0.1	0.0	0.0	0.0	2.0	1.3	1.1	0.8
Ypr097w	1			14%	168	7	6	4	7	9.0	1.7	0.3	0.1	1.5	3.2	2.8	2.0
Asg7	1.5	lumen		11%	18	0	1	1	0	0.2	0.0	0.0	0.0	0.6	0.8	/	/
Ede1	1.5			18%	221	16	7	11	5	12	2.6	0.3	0.0	3.0	4.1	4.0	1.6
‡ Erb1 ^{er}	1.5			31%	125	25	10	11	0	5.3	2.3	0.5	0.0	7	3.7	2.7	/
Exo1	1.5		E	9.9%	171	17	0	1	0	4.7	0.7	0.2	0.0	2.0	/	/	/
Mcm3	1.5		H ^b	3.1%	732	18	4	1	0	26	6.0	2.0	0.2	/	/	/	/
Nop2	1.5		H	85%	145	62	18	46	16	5.9	1.3	0.8	0.2	27	17	14	4.1
				12%	17	4	9	0	4					1.1	2.2	0.7	1.2
‡ Ost1 ^{er}	1.5	lumen	E	2.0%	148	2	0	1	0	1.9	0.6	0.2	0.0	0.0	/	/	/
Pet123	1.5	mito.		5.7%	35	0	1	1	0	0.4	0.0	0.0	0.0	0.5	0.8	/	/
Rqc1	1.5			20%	151	7	9	6	8	6.9	1.7	0.1	0.1	3.2	4.6	3.6	2.4
Sfb3 ^{er}	1.5			7.4%	274	10	4	6	0	6.8	1.9	0.1	0.1	1.4	1.6	1.7	/
Skp2	1.5			5.4%	37	1	0	1	0	0.6	0.4	0.0	0.0	0.2	/	/	/
Ste13	1.5	lumen	E	4.8%	230	8	2	2	0	10	2.4	0.7	0.1	/	0.1	0.3	/
Tub1	1.5		H	100%	74	2	0	71	0	3.0	0.4	0.0	0.0	15	16	17	/
Tub3	1.5		H			as Tub1											
Ubp10	1.5		H ^b	8.1%	533	20	14	8	4	24	6.2	1.0	0.1	1	2.9	2.6	1.3
Ubr1	1.5		H ^b	18%	186	18	13	9	3	19	3.6	1.1	0.1	1.7	3.7	2.5	1.0
Vid27	1.5			39%	155	17	25	17	15	12	1.5	0.7	0.0	10	12	7	4.0
Wsc2	1.5			1.9%	105	1	0	1	0	1.6	0.2	0.0	0.0	0.0	/	/	/
Ylr149c	1.5		H ^b	5.1%	99	3	1	1	0	2.9	0.6	0.2	0.1	0.2	0.3	/	/
Yor238w	1.5		H ^b	0.9%	117	1	0	0	0	1.3	0.1	0.1	0.0	/	/	/	/

Table 2. Continued.

(b) Samples of motifs with FFAT scores 2.0 and 2.5

Name	FFAT-score	Exclusion criteria															
		1. loc if not cyto.	2. struct if not IDR	FFAT-score ≤2.5	size of fam.	3. Conservation (family of orthologues)								p-value (-log ₁₀)			
						observed motifs				expected back-ground (random)							
						2.5	2.0	1.5	0.0-1.0	2.5	2.0	1.5	0.0-1.0	≤2.5	≤2.0	≤1.5	0.0-1.0
Acs1	2.0		E	10%	240	12	11	1	0	5.1	1.3	0.2	0.0	2.8	2.4	/	/
Ecm13	2.0			27%	123	24	12	0	2	5.7	2.7	0.7	0.1	4.6	1.9	0.3	0.7
Efr3	2.0			8.2%	159	4	2	6	1	4.3	0.6	0.1	0.0	1.2	2.1	2.0	/
Ent3	2.0			71%	139	85	19	0	0	3.4	0.6	0.2	0.0	21	4.4	/	/
Irr1	2.0			9.9%	302	23	8	0	0	24.3	5.9	2.0	0.6	0.1	0.0	/	/
Kar2 ^{er}	2.0	lumen		3.1%	413	6	13	0	0	13.1	3.5	1.3	0.0	0.1	1.3	/	/
Met4	2.0			5.0%	20	1	0	0	0	0.4	0.0	0.1	0.0	/	/	/	/
Mgr3	2.0	mito.		33%	43	2	12	0	0	0.7	0.0	0.1	0.1	3.2	3.1	/	/
Msh6	2.0			13%	563	57	20	1	1	26.7	5.6	1.3	0.3	4.7	2.2	0.1	/
Rkr1	2.0			4.5%	154	5	2	0	0	9.9	1.9	0.5	0.0	/	/	/	/
Sxm1	2.0			19%	158	27	4	1	4	12.8	3.2	0.3	0.1	2.2	0.9	1.3	1.3
‡ Utp9	2.0			38%	13	1	4	0	0	0.2	0	0	0	1.5	1.3	/	/
Yhl026c	2.0			3.7%	54	0	2	0	0	0.7	0.0	0.1	0.0	0.3	0.7	/	/
Aim44	2.5			35%	20	5	2	0	0	0.8	0.1	0.1	0.0	1.5	0.6	/	/
Chm7 ^{er}	2.5			7.4%	149	11	0	0	0	3.0	0.7	0.1	0.1	1.2	/	/	/
Dyn1	2.5			2.3%	175	4	0	0	0	23.8	4.7	0.9	0.4	/	/	/	/
‡ Epo1 ^{er}	2.5			24%	38	8	2	0	0	0.6	0.1	0.0	0.0	2.4	0.7	/	/
Kxd1 (x2)	2.5		H ^b	6.0%	83	6	1	0	0	0.5	0.8	0.0	0.0	1.3	/	/	/
Mdv1	2.5		E	4.1%	765	25	5	0	0	16.2	3.2	0.6	0.2	0.8	0.1	/	/
Mga2 ^{er}	2.5			7.1%	311	13	5	3	0	7.2	2.1	0.3	0.1	1.4	1.0	0.8	/
‡ Num1 ^{er}	2.5			52%	42	15	7	4	1	6.9	2.3	0.0	0.0	2.5	2.0	1.6	/
Pbi1 ‡	2.5		H	4.2%	24	1	0	0	0	0.5	0.3	0.0	0.0	0.1	/	/	/
Sec2	2.5			20%	163	24	7	9	0	4.6	1.7	0.2	0.1	6	3.0	2.4	/
Sec27	2.5			3.2%	1157	26	10	2	0	35.8	7.6	1.1	0.3	0.3	0.3	0.1	/
Sec63	2.5	lumen		6.4%	141	4	4	1	0	6.9	1.2	0.3	0.0	0.1	0.8	/	/
Sec66	2.5	lumen		2.2%	136	2	0	1	0	1.2	0.5	0.0	0.0	0.3	/	/	/
Tor1 (x2)	2.5		H (x2)	75%	158	133	1	0	0	9.3	1.2	0.1	0.0	24	/	/	/
Ubp1 ^{er}	2.5			5.2%	154	3	4	1	0	5.5	1.0	0.3	0.0	0.1	0.9	/	/
Ubp2	2.5		H	11%	154	18	1	0	0	13.9	2.7	0.8	0.3	0.1	/	/	/
Vps13 ^{er}	2.5			48%	142	27	24	10	14	23.0	4.7	1.2	0.3	5	8	5	3.5
Yml002w	2.5		H	11%	113	11	0	1	0	5.2	1.1	0.3	0.0	0.7	/	/	/
Ypt7	2.5			11%	38	3	1	0	0	0.9	0.2	0.0	0.0	0.7	/	/	/

Note. Sequences in order of increasing FFAT score. Red text, yellow fill indicates motif excluded and exclusion criterion: Location: lumen = translocated; mito. = mitochondrion; empty: cytoplasm. Structure: H helix, E sheet (red confirmed, black predicted, H^b indicates helix between domains); empty: loop. Conservation: exclusion either if observed motifs ≤ expected (numbers of motifs in red) or if all p values for excess motifs with FFAT score ≤2.5/2.0/1.5/1.0 ≥ 0.01 (i.e., -log₁₀ p value < 2.0). Display of p value: nonsignificant (p > 10⁻²): yellow; borderline (10⁻³ ≤ p < 10⁻²): pink; confirmed excess (p ≤ 10⁻³): blue. The most significant χ² is used to color the protein name. χ² is not valid if all expected values = 0, and not given if only one motif was observed (""). NOP2 has motifs of 2 main types: one in a confirmed helix and others in a loop. Kxd1 and Tor1 have two motifs with FFAT score = 2.5. Superscript^{er} = known on ER; FFAT = two phenylalanines in an acidic tract; IDR = intrinsically disordered region.

^aErb1 and Ost1 have been previously identified as binding Scs2, yeast vesicle-associated membrane protein-associated protein (VAP) (Chatr-Aryamontri et al., 2017).

^bEpo1, Num1, and Pbi1 have been reported to bind Scs2p, but potential FFAT motifs have not been tested (Chao et al., 2014; Riekhof et al., 2014).

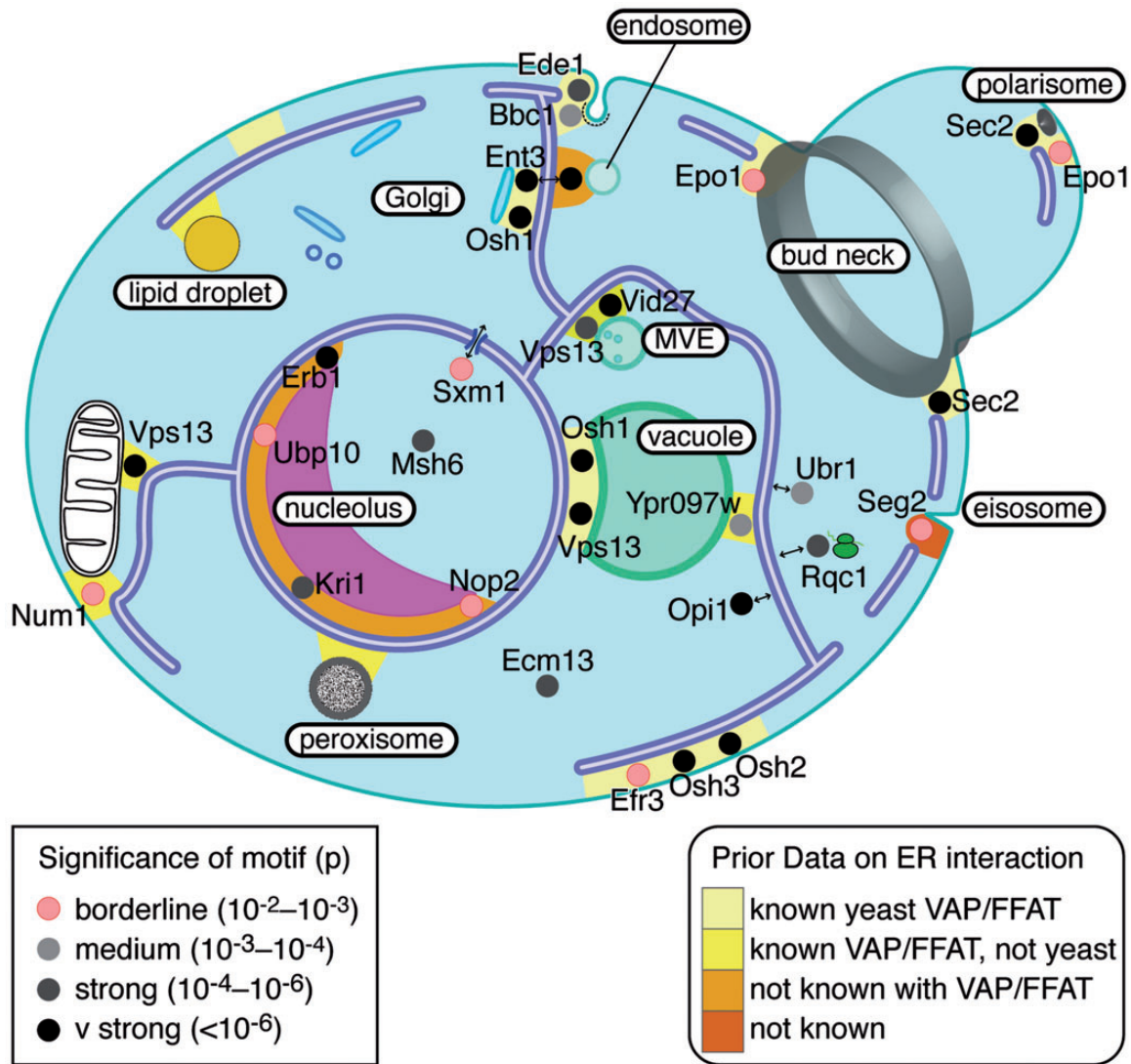


Figure 4. Cellular distribution of proteins with FFAT motifs. Likely location of 24 yeast proteins with FFAT motifs, including 4 known, all 11 with FFAT scores ≤ 1.5 , and a sample of 9 others with FFAT scores 2.0 or 2.5. Each protein is colored by the statistical significance of the χ^2 test of the number of motifs in its orthologue family over background (Table 2 and see left hand Key): pink = borderline significance, increasingly dark gray = increasing significance. To reflect known localization to multiple contact sites, Ent3, Epo1, Osh1, Sec2, and Vps13 appear more than once. Contacts between the ER and other cellular components are colored according to the level of prior knowledge (see right hand Key): light yellow (VAP-FFAT bridge known in yeast), dark yellow (VAP-FFAT bridge known, but not in yeast), orange (contact known, but no VAP-FFAT bridge in any cell type), and red (contact unknown). Note that no yeast FFAT proteins have been predicted yet that target lipid droplets or peroxisomes. MVE = multivesicular endosome; FFAT = two phenylalanines in an acidic tract; VAP = vesicle-associated membrane protein-associated protein; ER = endoplasmic reticulum.

Ede1p act in clathrin-coated endocytosis, which is also linked to VAP and ORPs via Myo2p (Encinar Del Dedo et al., 2017); (b) Ypr097wp localizes to the vacuole (the yeast equivalent of the lysosome) and resembles a sorting nexin, so it may parallel the binding of human SNX2 to VAP (Dong et al., 2016); and (c) Vid27p localizes to puncta resembling late (or multivesicular) endosomes and it has an ill-defined function in endocytosis, similar to human proteins at ER–endosome contacts (Rocha et al., 2009; Saita et al., 2009). The five other hits are involved with cellular components not previously known

to be linked to VAP: Erb1p, Kri1p, Nop2p, and Ubp10p assemble ribosomes in nucleoli; Seg2p stabilizes eisosomes.

Seven New Predicted VAP Interactors in a Sample of 28 Yeast Proteins Identified With FFAT Scores 2.0 and 2.5

Even though yeast has a relatively small proteome for a eukaryote, there are too many proteins with FFAT scores 2.0 and 2.5 ($n=44$ and 143, respectively,

Supplementary Table 1A) to consider them all here in detail. Therefore, we sampled these groups (Supplementary Table 1C). The basis for this sampling was firstly to include all proteins previously identified as interactors of yeast VAP and secondly to randomly sample the remaining proteins.

Among 13 proteins we examined with FFAT score = 2.0, 1 had previously been shown to bind yeast VAP: Utp9p. This protein contained a motif that just failed the statistical significance criterion possibly because the family of orthologues is small ($n = 13$, $p = 0.04 = 10^{-1.5}$) suggesting that the VAP-FFAT interaction is conserved at a rate below the sensitivity of our approach. Among the other 12 proteins in this group, 3 were excluded by structure or location and 4 by lack of conservation. Five met all criteria for predicting an active motif ($p < 0.01$), two with borderline statistical significance (Efr3p and Sxm1p), and three with the stronger significance (Ecm13p, Ent3p and Msh6p, Table 2(b), top). Changing the orthologue families for Efr3p and Sxm1p to Ascomycota increased the proportion of proteins positive for motifs but did not increase their statistical significance (Supplementary Table 4). For the sample of proteins with motifs with FFAT score = 2.0, the positive prediction rate 5/12 (42%) was close to the predicted rate (40%, Figure 3(b) and (c)). This implies a further 13 motifs would be found among the 31 proteins with FFAT score = 2.0 we did not study.

Among 19 proteins with FFAT score = 2.5 we examined, 3 (Epo1p, Num1p, and Pbi1p) have been shown to bind yeast VAP previously, though their predicted FFAT motifs have not been tested (Chao et al., 2014; Riekhof et al., 2014). Our analysis predicted that both Epo1p and Num1p have motifs that meet all the criteria for being an active FFAT motif, albeit with borderline significance in terms of conservation (Table 2(b), bottom). However, the motif in Pbi1p is in a known helix and did not meet the conservation criterion (Table 2(b), bottom). Among the other 16 proteins, 6 proteins were excluded by structure or location criteria and 8 by lack of sequence conservation. Four of these just missed our significance criterion ($p \leq 0.05$ but $p > 0.01$, like Utp9p see earlier), including integral ER proteins (Mga2p and Ubp1p), a trafficking protein (Kxd1p) and a polarity protein (Aim44p; Table 2(b)). A subfamily of sequences more closely related to Mga2p showed no greater statistical significance (Supplementary Table 4). This left two proteins that met all the criteria for predicting a FFAT motif: Sec2p and Vps13p (Table 2(b), bottom). The positive prediction rate among motifs with FFAT score = 2.5 (2/16, 13%) is similar to our proteome-wide estimate (15%, Figure 3(b)). Applying this rate to all other 124 yeast proteins with FFAT scores = 2.5 would produce a further 16 motifs.

The nine predicted motifs with FFAT scores 2.0 or 2.5 are in proteins localized to multiple cellular components that communicate with the ER, functioning in membrane traffic (Ent3p, Sec2p, and Vps13p), cytoskeleton and polarity (Epo1p and Num1p), PI4P synthesis at the plasma membrane (Efr3p), nuclear components (Msh6p and Sxm1p), and a cytoplasmic protein of unknown function (Ecm13p; Figure 4).

These findings create an overall picture of the number of FFAT motifs in yeast: 4 previously known, 3 in 5 partially characterized VAP interactors, and 17 we predict by detailed study. Extrapolating from the sample of proteins with FFAT scores 2.0 to 2.5 indicates the remainder of the proteome contains a further 29 proteins with motifs that meet all our criteria. However, with 155 proteins yet to be studied in these groups, this number has considerable margin for error. Thus, the bioinformatics approach predicts that the number of yeast proteins with FFAT motifs is approximately 50 to 55 (0.9% of the whole proteome), >10-fold higher than currently tested and verified.

Comparison of These Predictions With the MEME Suite Automated Pipeline

We next asked if an automated pipeline for SLiM discovery would perform just as well as the custom-made PWM. For this, we used the MEME Suite, which creates ungapped SLiMs from sequences and then searches for other occurrences in unrelated sequences (Bailey et al., 2009), and has been used to identify motifs in contact site biology (Bean et al., 2018). Seeded with a family of FFAT-positive proteins (here we used ORPs), MEME identified their shared FFAT motifs as the seven-residue core, together with acidic flanking residues both upstream and downstream (Supplementary Table 5A). We then used the FIMO (*find motifs*) module in MEME to identify motifs that fit this statistical model. To understand the significance of hits at the whole proteome level, we searched both the actual yeast proteome and randomizations of the yeast proteome, as we had done for the PWM. Motifs that match the MEME model most closely (with p values as low as 2×10^{-12}) were only found in the actual proteome. With increasing p values, the background hit rate increased, eventually approaching that of the observed proteome (Supplementary Table 5B), which is the same as we had observed for our PWM. The excess number of strong hits to the matrix in the actual proteome compared with random background was 37 ± 16 (number of randomizations = 4), which is approximately 70% of the level estimated by the PWM.

At the level of individual proteins, there was variation after the sixth hit (Supplementary Table 5C), and the rank order of yeast hits correlated only partially with

Table 3. Using Fungal Hits to Search for FFAT Motifs in Nonfungal Homologues.

Name	FFAT-score	alternate name	Family of orthologues														
			species group	FFAT-scores ≤ 2.5	size of family	observed motifs				expected background (random)				p-value ($-\log_{10}$)			
						2.5	2.0	1.5	0.0 - 1.0	2.5	2.0	1.5	0.0 - 1.0	≤ 2.5	≤ 2.0	≤ 1.5	≤ 1.0
EDE1	1.5	EPS15	non-fungi	2.6%	391	9	1	0	0	17	4.6	0.5	0.1	/	/	/	/
EFR3	2.0		animals	3.0%	364	11	0	0	0	8.2	1.9	0.4	0.0	0.0	/	/	/
			plants	3.0%	560	11	6	0	0	16	2.7	0.8	0.1	0.2	0.4	0.5	/
ENT3	2.0	Epsin	animals	5.5%	218	11	1	0	0	2.6	0.8	0.1	0.0	1.5	0.0	/	/
			plants	5.9%	170	8	2	0	0	3.6	1.0	0.3	0.0	0.7	0.1	/	/
ERB1	1.5	BOP1	All	17%	904	91	41	31	2	36	7.9	2.5	0.3	15	11	6	0.6
KRI1	1.0		animals	14%	226	16	16	1	0	19	5.8	1.5	0.3	0.4	1.2	/	/
			plants	39%	79	17	14	5	0	6.7	2.5	0.5	0.1	4.0	3.1	1.2	/
MSH6	2.0		non-fungi	6.7%	794	44	11	0	0	30	7.0	0.6	0.1	1.1	0.4	/	/
NOP2	1.5		bacteria	1.0%	620	4	1	0	1	8.2	1.6	0.4	0.0	/	0.0	/	/
			animals	82%	56	48	1	0	0	2.5	0.8	0.0	0.0	10	/	/	/
			plants	9.8%	122	10	1	1	0	3.3	1.0	0.2	0.0	1.2	0.2	/	/
RQC1	1.5	TCF25	animals	6.2%	324	21	0	0	0	7.4	1.8	0.5	0.0	1.4	/	/	/
			plants	14%	107	15	0	0	0	4.1	0.9	0.0	0.0	1.6	/	/	/
SEC2	2.5		non-fungi	0.0%	76	1	0	0	0	1.1	0.2	0.0	0.0	/	/	/	/
SXM1	2.0	Cse1	all	9.9%	565	44	20	1	0	37	7.0	2.0	0.5	1.1	1.4	/	/
UBP10	1.5	USP36	animals	24%	306	1	0	0	0	1.2	0.8	0.0	0.0	/	/	/	/
UBR1	1.5	UBR4	non-fungi	32%	507	147	20	4	3	58	12	2.2	0.4	10	1.3	0.8	0.8
VID27	1.5		non-fungi	24%	304	47	26	6	7	10	3.0	0.8	0.0	12	7	3.0	2.1
VPS13	2.5		all	18%	3860	415	210	83	138	242	50	17	3.0	55	58	38	29

Note. Alternate names for human homologues and range of species searched are in column 3. For some families, different species groups were searched separately (indicated in column 4). Coloring, p values, and exclusions as in Table 2. Where different groups of species within an overall family have different significance levels, only the most significant group is colored (column 4). FFAT = two phenylalanines in an acidic tract. “/” indicates that χ^2 is not valid or only motif was observed.

that produced by FFAT scores ($r = 0.3$). Approximately 20% of the top 3% of MEME hits contained residues that would prevent productive interaction with VAP (e.g., L2, E5). Thus, while the automated tool MEME generated a list of hits with much less work than our approach, in detail its predictions are less informative.

Widespread Motifs Found in Nonfungal Homologues in 30% of Yeast Proteins With a Motif

Finally, we looked for conservation of yeast FFAT motifs in other species for the 20 proteins with newly predicted FFAT motifs; 14 have nonfungal families, of which 6 had a clear excess of FFAT motifs (Table 3). Excess FFAT motifs occur in all eukaryotic kingdoms

for Vps13, Erb1 (alternate name BOP1 in other species), and Ubr1 (orthologue called UBR4 in mammals; data not shown). For Nop2, motifs are found in animal homologues but not significantly in plants. For Kri1, motifs are found in plants not animals. Vid27 homologues are only found in plants and alveolates, both of which have more motifs than fungal homologues. Twelve of the 20 new yeast hits have human homologues, three of which have motifs with low FFAT scores: VPS13 = 0.0 and NOP2 and UBR4 = 2.5.

Discussion

We studied one SLiM in one organism in detail. The potential FFAT motifs we predicted need verification

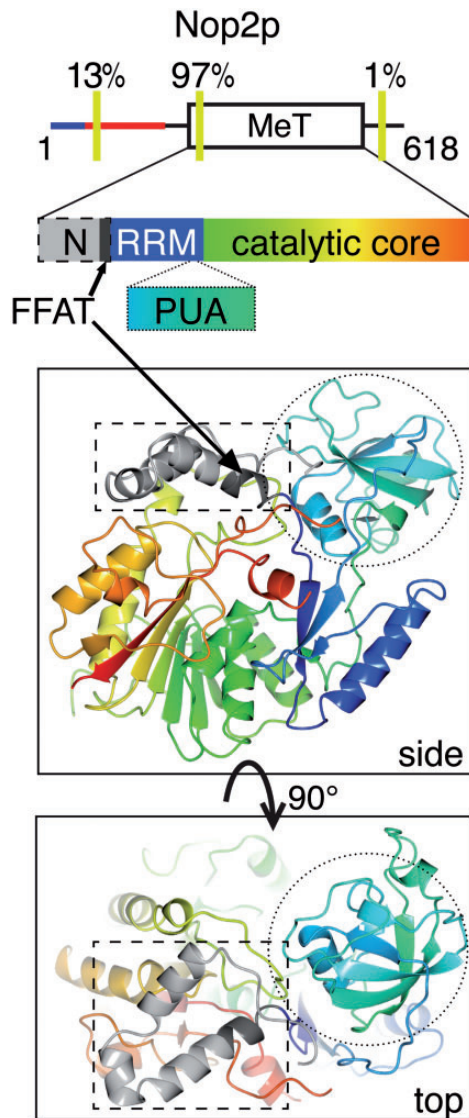


Figure 5. The FFAT motif in Nop2p is located in a helix. (a) FFAT motifs occur at three sites in the fungal family of Nop2 orthologues. Top: overall domain map, showing the position of motifs relative to the central methyltransferase domain and to the intrinsically disordered N-terminus, with its conserved basic (blue) and acidic (red) IDRs. The three locations for the 13 residues of motifs with FFAT scores ≤ 2.5 are indicated with yellow bars, showing the proportion of orthologues they are found there, including one motif found to the C-terminus of the folded domain. Bottom: the methyltransferase domain is divided into three parts: N-terminal extension (“N”, gray, dashed border) ending in the FFAT motif core (7 residues, dark gray), RNA Recognition Motif (RRM, blue), and catalytic core (rainbow green to red); the domain associates with a PseudoUridine synthase and Archaeosine transglycosylase (PUA) domain (rainbow blue to green, dotted border). In yeast, the PUA domain is in a separate protein, encoded by Nip7p, while animal homologues contain a PUA domain *in cis*. (b) Side and top views of methyltransferase domain of NSun6, a human homologue of Nop2, shown with colors and borders as in (a), from crystal structure 5WWQ (Liu et al., 2017). Four residues of the FFAT motif are folded into a single turn of a helix and then three in a linker to the RRM motif. FFAT = two phenylalanines in an acidic tract.

of ER targeting dependent on VAP. One question we cannot exclude is whether these sequences might be selected for another reason. Many eukaryotic SLiMs combine aromatic and charged residues. One overlaps with the FFAT motif: “FxDxF,” which sorts endocytic cargo (Brett et al., 2002; Dinkel et al., 2016). It is predicted that many more SLiMs will be discovered in the future (Tompa et al., 2014), and some may overlap FFAT more closely.

Maximizing Accurate Detection With a PWM

There are some issues with our PWM. First, it overvalues the known VAP interactors used to create it. Position 3 of FFAT motifs exemplifies this: while the original motifs varied little here, crystal structures showed little interaction with VAP (Kaiser et al., 2005), and divergent residues were later found (Saita et al., 2009; Baron et al., 2014). This raises a second problem: The PWM weights need verification by systematic measurement of how different substitutions affect affinity, especially combinations of multiple substitutions, which have only rarely been studied (Furuita et al., 2010; Mikitova and Levine, 2012). Phage display of peptides, particularly variants of FFAT, might address this in the future. For example, the finding of repeated substitution of F2 with tryptophan in Osh3 suggests the possibility that tryptophan at this position is not completely inhibitory to VAP interaction. This could be tested experimentally.

Searching for Motifs in Orthologues

Application of a PWM to large families of orthologues provided evidence that powerfully influenced the assessment of motifs. In some cases, the OrthoDB families were too inclusive: subfamilies of proteins more closely related to the Opi1p and Ubp10p produced more excess motifs that were more statistically significant. Therefore, improved accuracy in curating orthologue families would increase sensitivity of this approach.

Nop2p is unique because the yeast protein, like almost all fungal NOPS, has a motif that fails to meet our criteria because it occurs in a helix in the methyltransferase domain. However, other motifs in a minority of Nop2 orthologues cast doubt on the exclusion. These additional motifs in an anionic IDR (Figure 5(a)) were significantly in excess over the background of the whole family ($p = .006$, Table 2(a)). This has implications for the motifs in the methyltransferase domain both in fungi (consensus EFFEANE) and mammals (hNOP2 called NSun6, consensus EFLEANE), but not in plants (consensus ELIEAFE) and bacteria (consensus ALLAALN). One explanation for the motif in yeast and human is that its function is unrelated to VAP in the methyltransferase

domain. An alternative explanation is that the first helix of the domain identified in crystal structures is dynamic (Figure 5(b)) and can switch to an unstructured loop. This scenario speculates that sequences that match FFAT motifs have been positively selected in opisthokonts in a region with variable structure that is permissive for FFAT formation. This would explain the definite but low rate of motifs with low FFAT scores in bacterial NOP2 homologues (Table 3).

Estimates of the Background Rate of FFAT Motifs

Background levels of FFAT motifs were estimated from VAP-negative proteomes and sequence randomization. The main issue with VAP-negative proteomes is their variability. For example, 4.5% of proteins in the halophile archaeon *Haloterrigena turkmenica* have FFAT scores ≤ 2.5 . The level of low FFAT scores (more than eukaryotes) was not changed by randomizing the proteome (data not shown). This effect is associated with halophiles having high proportions of acidic residues and few bases to survive in their highly saline (2–4M salt) habitat (Lanyi, 1974). VAP-negative eukaryotes may represent better controls than prokaryotes, but they are separated from the bulk of eukaryotes by long phylogenetic branches, so their proteomes may vary for other unknown reasons.

For our randomizations, all residues in each protein were treated the same. In the future, we might attempt to improve this, randomizing residues in folded domains separately from those in IDRs, which are enriched for charged residues (Davey et al., 2012). A different way to avoid mixing IDRs with folded domains is to randomize within tiled windows longer than the motif. However, sequences rich in D/E with a minority of F/Y may produce a low FFAT score randomly. For example, the FFAT score of . . . DDDDDDD[F/Y]DD@DD . . . (where @ is A/C/S/T) is 1.0 to 1.5. Therefore, sliding windows would overestimate background levels of FFAT motifs because of their inherent repetitive nature.

Evolution of SLiMs

The ability of FFAT motifs to arise simply by the appearance in an acidic tract of F/Y and a small residue at +3 shows how FFAT motifs might evolve, especially as Y and A can both result from point mutations of D codons. We assume that motifs that occur in isolation have evolved relatively recently. Like any mutation, they may have no productive function initially and could be counterselected if deleterious. This may have occurred in the highly anionic motor protein Dyn1, orthologues of which have many fewer motifs than the same sequences randomized.

Among the eukaryotes we studied, *A. thaliana* lacks proteins with the lowest FFAT scores (Figure 2(b)). Could the plant VAP-FFAT interaction be marginally

different from yeast/human (opisthokonts)? Comparison of motifs in *A. thaliana* with those in yeast and human showed that the cores were similar, but upstream residues were less acidic in plants (data not shown). This suggests that plant VAP might have a compensatory difference from yeast and human VAP.

Can VAP Bind Approximately 0.9% of the Proteome?

There are pleiotropic phenotypes in cells lacking VAP which are consistent with dysfunction of many pathways in combination (Loewen and Levine, 2005; Di Mattia et al., 2018). However, the finding that cells still grow without VAP implies that loss of VAP-FFAT interactions mainly leads to reduced efficiency, which can be compensated by over-expression (Hanada et al., 2003). There are far less copies of VAP than the sum of its interactors, so only a fraction of FFAT motifs can be active at any one time. Part of the explanation might be that FFAT motifs interact only weakly (micromolar Kd), so they only bind VAP when multimerized. In addition, they may be masked by intramolecular interactions regulated by post-translational modifications (Alli-Balogun and Levine, 2019), as indicated by considerable diffuse background of green fluorescent protein-tagged FFAT-positive proteins (Breker et al., 2014). If the active FFAT motifs in total saturate VAP, then competitive binding will depend on both affinity (related to FFAT score) and avidity determined by local concentration. Thus, interaction may favor proteins concentrated near the ER, such as on an adjacent organelle.

FFAT Motifs at Contact Sites in Yeast

Most of the proteins with FFAT motifs newly predicted here are located on organelles that are already known to use VAP-FFAT interactions to communicate with the ER either in yeast or other cells (Figure 4) (Phillips and Voeltz, 2016). Some are involved in functions apparently outside the ER, but their location allows them to access it, for example, the intranuclear proteins Sxm1 and Msh6, which can access VAP that diffuses to the inner nuclear envelope (Beilharz et al., 2003; Brickner and Walter, 2004; James et al., 2019).

There are two yeast organelles that contain proteins with putative FFAT motifs where ER contact is not well-characterized: nucleolus and eisosome. For the nucleolus, we had four hits. Among these, Erb1p and Nop2p localize partly to the nuclear envelope, so may be constitutively associated with VAP, while Kri1p and Ubp10p are diffuse in the nucleolus (Breker et al., 2014). For the eisosome, Seg2 has a motif with a low FFAT score as does its paralogue Seg1p. Eisosomes are grooves 50 nm deep and 1 μ m long in the yeast plasma membrane that sense tension (Berchtold et al., 2012). The majority of eisosomes are well

Table 4. Suboptimal Scores for Substitutions in Core of FFAT Motif.

(a)	Score for core residues						
	1	2	3	4	5	6	7
Ideal	E	F	F	D	A	x	E
A	1	4	1	2	0	0	1
C	1	4	1	2	0	0	1
D	0	4	1	0	4	0	0
E	0	4	1	0	4	0	0
F	1	0	0	2	2	0	1
G	1	4	1	2	1	0	1
H	1	4	0.7	2	2	0	1.5
I	1	4	1	2	2	0	1
K	1.5	4	1	2	2.5	0	1.5
L	1	4	1	2	2	0	1
M	1	4	1	2	2	0	1
N	1	4	1	2	2	0	1
P	1	4	1	2	2	0	1
Q	1	4	1	2	2	0	1
R	2	4	1	2	3	0	1.5
S	0.5	4	1	0.5	0.5	0	0.5
T	0.5	4	1	0.5	0.5	0	0.5
V	1	4	1	2	2	0	1
W	1	2	0.7	2	2	0	1
Y	1	0	0.5	2	2	0	1
X	2	4	2	2	2	2	2
Acidic tract (N-terminal to core): charge							
(b)	1	2	3	4	5	6	
Ideal	D/E	D/E	D/E	D/E	D/E	D/E	D/E
D/E	+1	+1	+1	+1	+1	+1	+1
S/T	+0.5	+0.5	+0.5	+0.5	+0.5	+0.5	+0.5
K/R	-1	-1	-1	-1	-1	-1	-1
Others	0	0	0	0	0	0	0
Convert charge into score							
(c)	≥ 4	3–3.5	2–2.5	≤ 1.5			
Score	0	0.5	1	1.5			

Note. (a) Position weight matrix applied to core to convert sequence into scores and (b) and (c) estimation of charge in six residues in possible acidic tract upstream of core and conversion into scores.

separated from the nearest element of the cortical ER network (150–250 nm). This has been the basis for considering eisosomes to be unlinked to the ER (Stradalova et al., 2012). However, approximately 20% eisosomes are ≤ 50 nm distant from cortical ER (Stradalova et al., 2012), making it possible for Seg2p or Seg1p to bridge from eisosomes to ER. Mutating putative FFAT motifs in both Seg1/2p could probe this potential ER–eisosome bridging.

Contact Sites in Humans and Beyond

Conserved ER–nucleolar relationships are already known for the integral inner nuclear envelope protein RRS1 both in yeast (Horigome et al., 2011) and human cells, where it

mediates targeting of ER proteins to nucleoli (Sutherland et al., 2004; Carnemolla et al., 2009). We found an excess sequences with low FFAT scores in nucleolar proteins in different kingdoms: ERB1 in all eukaryotes, KRI1 in fungi and plants, and NOP2 (with the caveats above) in fungi and animals (Table 3). These should be the focus of future study into the relationship between the ER and nucleolus. Given that proteins with FFAT motifs typically engage VAP when they function to form bridges between the ER and other compartments, we speculate that there is ER–nucleolar bridging. ER–nucleolar bridging in yeast is easy to conceptualize because the nucleolus forms a crescent adherent to approximately 50% of the inner nuclear envelope (Thiry and Lafontaine, 2005). By contrast,

nucleoli in plants and animals are spherical structures deep in the nucleoplasm. Despite this, ER–nucleolar contact could be achieved by the tubular nucleoplasmic reticulum, which extends from the inner nuclear envelope deep into the nucleus of animal and plant cells (Malhas et al., 2011; Ohsaki et al., 2016). Our data suggest that nucleoplasmic reticulum may directly contact nucleoli. Alternatively, there may be no bridging, and the nucleolar proteins with FFAT motifs have inactive pools targeted to the nuclear periphery. Therefore, motifs with low FFAT scores in proteins with unexpected locations (nucleoli and likewise eisosomes) prompt further study of the relationship between the ER and other organelles.

Methods

Producing FFAT Scores

The scoring system is identical to that used previously (Murphy and Levine, 2016). In brief, to produce FFAT scores, suboptimal substitutions in any core of seven residues were allocated a score in a PWM (Table 4(a)). Some sequences in databases contain undetermined residues (identified as “X”). These were scored as inhibiting VAP binding. The six residues upstream were assessed as a single unit for charge (Table 4(b)), and the estimate of overall negative charge was scaled (Table 4(c)) and combined with the core to create the FFAT score. This represents the sum total of suboptimal scores across 13 continuous residues. FFAT scores were binned into 0.5 unit bins: “0” = 0 to 0.25, “0.5” = 0.251 to 0.75, “1.0” = 0.751 to 1.25, and so on. For proteomes and families of sequences, the scoring system was enacted in Python. The code for the scripts will be made available on GitHub or on request from the authors.

Sequences

Proteome-wide sequences were obtained from Uniprot. Control proteomes were as follows: bacteria—*Escherichia coli*, *Mycobacterium tuberculosis*, *Staphylococcus aureus*, and *Mycoplasma hominis*; archaea—*Pyrococcus horikoshii*, *Methanosarcina mazei*, *Thermoplasma acidophilum*, *Archaeoglobus fulgidus*, *Sulfolobus solfataricus*, and *Methanocaldococcus jannaschii*; and VAP-negative eukaryotes—*Euglena gracilis*, the diplomonad *Spironucleus salmonicida*, and the kinetoplastid *Paratrypanosoma confusum*. Randomization was carried out within each protein, repeated 10-fold to obtain a mean and standard deviation for FFAT scores. Families of proteins orthologous to yeast hits were obtained from Uniprot using the OrthoDB identifier of the yeast protein (Supplementary Table 1D). All lists were filtered to remove entries flagged as fragments and also with “identity: 0.9” filter, or “identity: 0.5” if

there were more than 1,000 sequences. Where no OrthoDB family was available (Ypt7 and Seg2), we searched for homologues using BLAST filtered to nr90 (Ypt7p), or for iterative searches we used HHblits three iterations (Seg2p) at the Tuebingen Toolkit (Zimmermann et al., 2018). Similarly, subfamilies made by KEGG and Orthologous Matrix (OMA) (Kanehisa et al., 2004; Schneider et al., 2007) used the KEGG/OMA identifiers. For oxysterol binding protein homologues to submit to MEME, nonredundant proteins containing the Pfam domain PF01237 (Oxysterol_BP) were downloaded from Uniprot and submitted (at meme-suite.org/tools/meme), with the resulting matrix forwarded to FIMO (<http://meme-suite.org/tools/fimo>; Bailey et al., 2009).

Analysis of Potential Motifs in Yeast

Criteria applied to each potential FFAT motif included two used previously: location and secondary structure (Murphy and Levine, 2016). Location was determined by combining two sources of data: (a) topology prediction from signal sequences (Signal4.0) and transmembrane domains (TMHMM2.0 and TOPCONS; Kahsay et al., 2005; Petersen et al., 2011; Tsirigos et al., 2015) and (b) reconsideration of topology predicted in (a) in the light of functional requirements implied by domain analysis in HHsearch (Soding et al., 2005), which is available online at the Tuebingen Toolkit as HHpred (Zimmermann et al., 2018). Predicted secondary structure was in three states (helix, sheet, and unstructured loop), which was produced by PSI-PRED 3.0 as part of HHsearch (Buchan et al., 2013).

For the third criterion, conservation of the motif in closely related species, excess motifs with low FFAT scores in families of orthologue were estimated by comparing numbers of observed motifs with FFAT scores ≤ 2.5 (and ≤ 2.0 , ≤ 1.5 , and ≤ 1.0 in parallel) with the average numbers of motifs in randomized proteins by the “ $N-1$ ” χ^2 test that compares the $(ad-bc)^2(N-1)/mnr$ s with the χ^2 distribution with one degree of freedom, as described (Campbell, 2007). Multiple incidences of motifs with FFAT scores ≤ 2.5 per protein were included, that is, statistics refer to every residue, with no reduction to the most optimal FFAT score per protein. The most significant of the four tests (excess ≤ 2.5 , ≤ 2.0 , ≤ 1.5 , or ≤ 1.0) was assumed to apply (Tables 2 and 3). A cutoff $p \leq 10^{-2}$ was used (Table 2, final four columns: cutoff ≥ 2 when expressed as $-\log_{10}$). The values of p between 10^{-2} and 10^{-3} were considered borderline. Any protein family with only one motif or fewer motifs than the background at a particular FFAT score was excluded as having too low numbers to assess, and with only two motifs it was considered of borderline significance. The level of motifs with low FFAT scores in all randomized families together was

similar to that in the randomized yeast proteome (data not shown).

For data mining, documented physical interactions of VAPA/VAPB/MOSPD2 and Scs2/Scs22 were obtained from BIOGRID. Where published literature is not informative on intracellular targeting, we used high-throughput localization studies at the LOQATE database (Breker et al., 2014), which includes genome-wide libraries of all yeast proteins tagged with fluorescent proteins at both N- and C-termini (Weill et al., 2018).

Analysis of Potential Motifs in Nonfungal Relatives of Yeast Hits

Twenty lists of nonfungal homologues of 14 yeast hits were constructed from UniProt Reference Clusters (UniRef, nonredundant as above) based either on OrthoDB families or on combinations of domains (Supplementary Table 1E). Fragments were removed, and sequences were filtered for nonredundancy as above. Observed motifs with low FFAT scores were compared with the number of motifs after sequence randomization ($n = 10$), including multiple incidences per protein as above.

Acknowledgments

The authors thank Martin Frisher (Keele University) and Matt Hayes (UCL Institute of Ophthalmology) for helpful discussions.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a grant from the Bioinformatics and Biological Resources Fund of the Biotechnology and Biological Sciences Research Council UK (BB/M011801/1).

ORCID iD

Timothy P. Levine  <https://orcid.org/0000-0002-7231-0775>

Supplemental material

Supplemental material for this article is available online.

References

- Alli-Balogun GO, Levine TP (2019). Regulation of targeting determinants in interorganelle communication. *Curr Opin Cell Biol* 57, 106–114. doi: 10.1016/j.ceb.2018.12.010
- Alpy F, Rousseau A, Schwab Y, Legueux F, Stoll I, Wendling C, Spiegelhalter C, Kessler P, Mathelin C, Rio MC, et al. (2013). STARD3 or STARD3NL and VAP form a novel molecular tether between late endosomes and the ER. *J Cell Sci* 126, 5500–5512. doi: 10.1242/jcs.139295
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37, W202–W208. doi: 10.1093/nar/gkp335
- Baron Y, Pedrioli PG, Tyagi K, Johnson C, Wood NT, Fountaine D, Wightman M, Alexandru G (2014). VAPB/ALS8 interacts with FFAT-like proteins including the p97 cofactor FAF1 and the ASNA1 ATPase. *BMC Biol* 12, 39. doi: 10.1186/1741-7007-12-39
- Bean BDM, Dziurdzik SK, Kolehmainen KL, Fowler CMS, Kwong WK, Grad LI, Davey M, Schluter C, Conibear E (2018). Competitive organelle-specific adaptors recruit Vps13 to membrane contact sites. *J Cell Biol* 217, 3593–3607. doi: 10.1083/jcb.201804111
- Beilharz T, Egan B, Silver PA, Hofmann K, Lithgow T (2003). Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*. *J Biol Chem* 278, 8219–8223. doi: 10.1074/jbc.M212725200
- Berchtold D, Piccolis M, Chiaruttini N, Riezman I, Riezman H, Roux A, Walther TC, Loewith R (2012). Plasma membrane stress induces relocalization of Slm proteins and activation of TORC2 to promote sphingolipid synthesis. *Nat Cell Biol* 14, 542–547. doi: 10.1038/ncb2480
- Breker M, Gymrek M, Moldavski O, Schuldiner M (2014). LoQAtE—Localization and Quantitation ATlas of the yeast proteome. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast. *Nucleic Acids Res* 42, D726–D730. doi: 10.1093/nar/gkt933
- Brett TJ, Traub LM, Fremont DH (2002). Accessory protein recruitment motifs in clathrin-mediated endocytosis. *Structure* 10, 797–809. doi: 10.1016/s0969-2126(02)00784-0
- Brickner JH, Walter P (2004). Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS Biol* 2, E342. doi: 10.1371/journal.pbio.0020342
- Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41, W349–W357. doi: 10.1093/nar/gkt381
- Campbell I (2007). Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Stat Med* 26, 3661–3675. doi: 10.1002/sim.2832
- Carnemolla A, Fossale E, Agostoni E, Michelazzi S, Calligaris R, De Maso L, Del Sal G, MacDonald ME, Persichetti F (2009). Rrs1 is involved in endoplasmic reticulum stress response in Huntington disease. *J Biol Chem* 284, 18167–18173. doi: 10.1074/jbc.M109.018325
- Chao JT, Wong AK, Tavassoli S, Young BP, Chruscicki A, Fang NN, Howe LJ, Mayor T, Foster LJ, Loewen CJ (2014). Polarization of the endoplasmic reticulum by ER-septin tethering. *Cell* 158, 620–632. doi: 10.1016/j.cell.2014.06.033
- Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A, et al. (2017). The BioGRID interaction database: 2017

- update. *Nucleic Acids Res* 45, D369–D379. doi: 10.1093/nar/gkw1102
- Clairfeuille T, Norwood SJ, Qi X, Teasdale RD, Collins BM (2015). Structure and membrane binding properties of the endosomal tetratricopeptide repeat (TPR) domain-containing sorting nexins SNX20 and SNX21. *J Biol Chem* 290, 14504–14517. doi: 10.1074/jbc.M115.650598
- Costello JL, Castro IG, Hacker C, Schrader TA, Metz J, Zeuschner D, Azadi AS, Godinho LF, Costina V, Findeisen P, et al. (2017). ACBD5 and VAPB mediate membrane associations between peroxisomes and the ER. *J Cell Biol* 216, 331–342. doi: 10.1083/jcb.201607055
- Danson CM, Pearson N, Heesom KJ, Cullen PJ (2018). Sorting nexin-21 is a scaffold for the endosomal recruitment of huntingtin. *J Cell Sci* 131, jcs211672. doi: 10.1242/jcs.211672
- Davey NE, Seo MH, Yadav VK, Jeon J, Nim S, Krystkowiak I, Blikstad C, Dong D, Markova N, Kim PM, Ivarsson Y (2017). Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *FEBS J* 284, 485–498. doi: 10.1111/febs.13995
- Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ (2012). Attributes of short linear motifs. *Mol Biosyst* 8, 268–281. doi: 10.1039/c1mb05231d
- De Vos KJ, Morotz GM, Stoica R, Tudor EL, Lau KF, Ackerley S, Warley A, Shaw CE, Miller CC (2011). VAPB interacts with the mitochondrial protein PTPIP51 to regulate calcium homeostasis. *Hum Mol Genet* 21, 1299–1311. doi: 10.1093/hmg/ddr559
- Di Mattia T, Wilhelm LP, Ikhlef S, Wendling C, Spohner D, Nomine Y, Giordano F, Mathelin C, Drin G, Tomasetto C, Alpy F (2018). Identification of MOSPD2, a novel scaffold for endoplasmic reticulum membrane contact sites. *EMBO Rep* 19 e45453. doi: 10.15252/embr.201745453
- Dinkel H, Van Roey K, Michael S, Kumar M, Uyar B, Altenberg B, Milchevskaya V, Schneider M, Kuhn H, Behrendt A, et al. (2016). ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res* 44, D294–D300. doi: 10.1093/nar/gkv1291
- Dong R, Saheki Y, Swarup S, Lucast L, Harper JW, De Camilli P (2016). Endosome-ER contacts control actin nucleation and retromer function through VAP-dependent regulation of PI4P. *Cell* 166, 408–423. doi: 10.1016/j.cell.2016.06.037
- Encinar Del Dedo J, Idrissi FZ, Fernandez-Golbano IM, Garcia P, Rebollo E, Krzyzanowski MK, Grottsch H, Geli MI (2017). ORP-mediated ER contact with endocytic sites facilitates actin polymerization. *Dev Cell* 43, 588–602 e586. doi: 10.1016/j.devcel.2017.10.031
- Fidler DR, Murphy SE, Curtis K, Antonoudiou P, El-Tohamy R, Ient J, Levine TP (2016). Using HHsearch to tackle proteins of unknown function: A pilot study with PH domains. *Traffic* 17, 1214–1226. doi: 10.1111/tra.12432
- Furuita K, Jee J, Fukada H, Mishima M, Kojima C (2010). Electrostatic interaction between oxysterol-binding protein and VAMP-associated protein A revealed by NMR and mutagenesis studies. *J Biol Chem* 285, 12961–12970. doi: 10.1074/jbc.M109.082602
- Hanada K, Kumagai K, Yasuda S, Miura Y, Kawano M, Fukasawa M, Nishijima M (2003). Molecular machinery for non-vesicular trafficking of ceramide. *Nature* 426, 803–809. doi: 10.1038/nature02188
- Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712–723. doi: 10.1016/j.cell.2015.09.053
- Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183. doi: 10.1038/415180a
- Horigome C, Okada T, Shimazu K, Gasser SM, Mizuta K (2011). Ribosome biogenesis factors bind a nuclear envelope SUN domain protein to cluster yeast telomeres. *EMBO J* 30, 3799–3811. doi: 10.1038/emboj.2011.267
- Hua R, Cheng D, Coyaud E, Freeman S, Di Pietro E, Wang Y, Vissa A, Yip CM, Fairn GD, Braverman N, et al. (2017). VAPs and ACBD5 tether peroxisomes to the ER for peroxisome maintenance and lipid homeostasis. *J Cell Biol* 216, 367–377. doi: 10.1083/jcb.201608128
- Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, et al. (2015). The BioPlex network: a systematic exploration of the human interactome. *Cell* 162, 425–440. doi: 10.1016/j.cell.2015.06.043
- Hynynen R, Laitinen S, Kakela R, Tanhuanpaa K, Lusa S, Ehnholm C, Somerharju P, Ikonen E, Olkkonen VM (2005). Overexpression of OSBP-related protein 2 (ORP2) induces changes in cellular cholesterol metabolism and enhances endocytosis. *Biochem J* 390, 273–283. doi: 10.1042/BJ20042082
- James C, Muller M, Goldberg MW, Lenz C, Urlaub H, Kehlenbach RH (2019). Proteomic mapping by rapamycin-dependent targeting of APEX2 identifies binding partners of VAPB at the inner nuclear membrane. *J Biol Chem* [Epub ahead of print]. doi: 10.1074/jbc.RA118.007283
- Johnson B, Leek AN, Sole L, Maverick EE, Levine TP, Tamkun MM (2018). Kv2 potassium channels form endoplasmic reticulum/plasma membrane junctions via interaction with VAPA and VAPB. *Proc Natl Acad Sci U S A* 115, E7331–E7340. doi: 10.1073/pnas.1805757115
- Kahsay RY, Gao G, Liao L (2005). An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. *Bioinformatics* 21, 1853–1858. doi: 10.1093/bioinformatics/bti303
- Kaiser SE, Brickner JH, Reilein AR, Fenn TD, Walter P, Brunger AT (2005). Structural basis of FFAT motif-mediated ER targeting. *Structure* 13, 1035–1045. doi: 10.1016/j.str.2005.04.010
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32, D277–D280. doi: 10.1093/nar/gkh063

- Kim I, Lee H, Han SK, Kim S (2014). Linear motif-mediated interactions have contributed to the evolution of modularity in complex protein interaction networks. *PLoS Comput Biol* 10, e1003881. doi: 10.1371/journal.pcbi.1003881
- Kirmiz M, Vierra NC, Palacio S, Trimmer JS (2018). Identification of VAPA and VAPB as Kv2 channel-interacting proteins defining endoplasmic reticulum-plasma membrane junctions in mammalian brain neurons. *J Neurosci* 38, 7562–7584. doi: 10.1523/JNEUROSCI.0893-18.2018
- Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM (2015). OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res* 43, D250–D256. doi: 10.1093/nar/gku1220
- Krystkowiak I, Davey NE (2017). SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res* 45, W464–W469. doi: 10.1093/nar/gkx238
- Kumagai K, Hanada K (2019). Structure, functions and regulation of CERT, a lipid-transfer protein for the delivery of ceramide at the ER-Golgi membrane contact sites. *FEBS Lett* 593, 2366–2377. doi: 10.1002/1873-3468.13511
- Kumagai K, Kawano-Kawada M, Hanada K (2014). Phosphoregulation of the ceramide transport protein CERT at serine 315 in the interaction with VAMP-associated protein (VAP) for inter-organelle trafficking of ceramide in mammalian cells. *J Biol Chem* 289, 10748–10760. doi: 10.1074/jbc.M113.528380
- Kumar N, Leonzino M, Hancock-Cerutti W, Horenkamp FA, Li P, Lees JA, Wheeler H, Reinisch KM, De Camilli P (2018). VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact sites. *J Cell Biol* 217, 3625–3639. doi: 10.1083/jcb.201807019
- Lanyi JK (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol Rev* 38, 272–290. PMID: PMC413857
- Liu RJ, Long T, Li J, Li H, Wang ED (2017). Structural basis for substrate binding and catalytic mechanism of a human RNA:m5C methyltransferase NSun6. *Nucleic Acids Res* 45, 6684–6697. doi: 10.1093/nar/gkx473
- Loewen CJ, Levine TP (2005). A highly conserved binding site in VAP for the FFAT motif of lipid binding proteins. *J Biol Chem* 280, 14097–14104. doi: 10.1074/jbc.M500147200
- Loewen CJ, Roy A, Levine TP (2003). A conserved ER targeting motif in three families of lipid binding proteins and in Opi1p binds VAP. *EMBO J* 22, 2025–2035. doi: 10.1093/emboj/cdg201
- Malhas A, Goulbourne C, Vaux DJ (2011). The nucleoplasmic reticulum: form and function. *Trends Cell Biol* 21, 362–373. doi: 10.1016/j.tcb.2011.03.008
- McCune BT, Tang W, Lu J, Eaglesham JB, Thorne L, Mayer AE, Condiff E, Nice TJ, Goodfellow I, Krezel AM, Virgin HW (2017). Noroviruses co-opt the function of host proteins VAPA and VAPB for replication via a phenylalanine-phenylalanine-acidic-tract-motif mimic in nonstructural viral protein NS1/2. *MBio* 8, e00668–e00717. doi: 10.1128/mBio.00668-17
- Meszaros B, Dosztanyi Z, Simon I (2012). Disordered binding regions and linear motifs—bridging the gap between two models of molecular recognition. *PLoS One* 7, e46829. doi: 10.1371/journal.pone.0046829
- Mikitova V, Levine TP (2012). Analysis of the key elements of FFAT-like motifs identifies new proteins that potentially bind VAP on the ER, including two AKAPs and FAPP2. *PLoS One* 7, e30455. doi: 10.1371/journal.pone.0030455
- Moreira KE, Schuck S, Schrul B, Frohlich F, Moseley JB, Walther TC, Walter P (2012). Seg1 controls eisosome assembly and shape. *J Cell Biol* 198, 405–420. doi: 10.1083/jcb.201202097
- Murphy SE, Levine TP (2016). VAP, a versatile access point for the endoplasmic reticulum: review and analysis of FFAT-like motifs in the VAPome. *Biochim Biophys Acta* 1861, 952–961. doi: 10.1016/j.bbali.2016.02.009
- Ohsaki Y, Kawai T, Yoshikawa Y, Cheng J, Jokitalo E, Fujimoto T (2016). PML isoform II plays a critical role in nuclear lipid droplet formation. *J Cell Biol* 212, 29–38. doi: 10.1083/jcb.201507122
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Phillips MJ, Voeltz GK (2016). Structure and function of ER membrane contact sites with other organelles. *Nat Rev Mol Cell Biol* 17, 69–82. doi: 10.1038/nrm.2015.8
- Riekhof WR, Wu WI, Jones JL, Nikrad M, Chan MM, Loewen CJ, Voelker DR (2014). An assembly of proteins and lipid domains regulates transport of phosphatidylserine to phosphatidylserine decarboxylase 2 in *Saccharomyces cerevisiae*. *J Biol Chem* 289, 5809–5819. doi: 10.1074/jbc.M113.518217
- Rocha N, Kuijl C, van der Kant R, Janssen L, Houben D, Janssen H, Zwart W, Neeffjes J (2009). Cholesterol sensor ORP1L contacts the ER protein VAP to control Rab7-RILP-p150 Glued and late endosome positioning. *J Cell Biol* 185, 1209–1225. doi: 10.1083/jcb.200811005
- Saita S, Shirane M, Natume T, Iemura S, Nakayama KI (2009). Promotion of neurite extension by protrudin requires its interaction with vesicle-associated membrane protein-associated protein. *J Biol Chem* 284, 13766–13777. doi: 10.1074/jbc.M807938200
- Schneider A, Dessimoz C, Gonnet GH (2007). OMA Browser—exploring orthologous relations across 352 complete genomes. *Bioinformatics* 23, 2180–2182. doi: 10.1093/bioinformatics/btm295
- Skehel PA, Fabian-Fine R, Kandel ER (2000). Mouse VAP33 is associated with the endoplasmic reticulum and microtubules. *Proc Natl Acad Sci U S A* 97, 1101–1106. doi: 10.1073/pnas.97.3.1101
- Soding J, Biegert A, Lupas AN (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33, W244–W248. doi: 10.1093/nar/gki408
- Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW (2009). The fungi. *Curr Biol* 19, R840–R845. doi: 10.1016/j.cub.2009.07.004

- Stradalova V, Blazikova M, Grossmann G, Opekarova M, Tanner W, Malinsky J (2012). Distribution of cortical endoplasmic reticulum determines positioning of endocytic events in yeast plasma membrane. *PLoS One* 7, e35132. doi: 10.1371/journal.pone.0035132
- Sutherland HG, Lam YW, Briers S, Lamond AI, Bickmore WA (2004). 3D3/lyric: a novel transmembrane protein of the endoplasmic reticulum and nuclear envelope, which is also present in the nucleolus. *Exp Cell Res* 294, 94–105. doi: 10.1016/j.yexcr.2003.11.020
- Thiry M, Lafontaine DL (2005). Birth of a nucleolus: the evolution of nucleolar compartments. *Trends Cell Biol* 15, 194–199. doi: 10.1016/j.tcb.2005.02.007
- Tompa P, Davey NE, Gibson TJ, Babu MM (2014). A million peptide motifs for the molecular biologist. *Mol Cell* 55, 161–169. doi: 10.1016/j.molcel.2014.05.032
- Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* 43, W401–W407. doi: 10.1093/nar/gkv485
- Weill U, Yofe I, Sass E, Stynen B, Davidi D, Natarajan J, Ben-Menachem R, Avihou Z, Goldman O, Harpaz N, et al. (2018). Genome-wide SWAp-Tag yeast libraries for proteome exploration. *Nat Methods* 15, 617–622. doi: 10.1038/s41592-018-0044-9
- Wilk S, Wittland J, Thywissen A, Schmitz HP, Heinisch JJ (2010). A block of endocytosis of the yeast cell wall integrity sensors Wsc1 and Wsc2 results in reduced fitness in vivo. *Mol Genet Genomics* 284, 217–229. doi: 10.1007/s00438-010-0563-2
- Xue W, Yin Y, Ismail F, Hu C, Zhou M, Cao X, Li S, Sun X (2019). Transcription factor CCG-8 plays a pivotal role in azole adaptive responses of *Neurospora crassa* by regulating intracellular azole accumulation. *Curr Genet* 65, 735–745. doi: 10.1007/s00294-018-0924-7
- Yadav S, Thakur R, Georgiev P, Deivasigamani S, Krishnan H, Ratnaparkhi G, Raghu P (2018). RDGBalpha localization and function at membrane contact sites is regulated by FFAT-VAP interactions. *J Cell Sci* 131. doi: 10.1242/jcs.207985
- Yakubovskaya E, Guja KE, Mejia E, Castano S, Hambardjjeva E, Choi WS, Garcia-Diaz M (2012). Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. *Structure* 20, 1940–1947. doi: 10.1016/j.str.2012.08.027
- Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007
- Zuzow N, Ghosh A, Leonard M, Liao J, Yang B, Bennett EJ (2018). Mapping the mammalian ribosome quality control complex interactome using proximity labeling approaches. *Mol Biol Cell* 29, 1258–1269. doi: 10.1091/mbc.E17-12-0714