The role of 3′UTR-protein complexes in the regulation of protein multifunctionality and subcellular localization

Diogo M. Ribeiro[1], Alexis Prod'homme[1], Adrien Teixeira[1], Andreas Zanzoni[1,*,†] and Christine Brun [1,2,*,†]

[1]Aix Marseille Univ, Inserm, TAGC, UMR_S1090, Marseille, France and [2]CNRS, Marseille, France

ABSTRACT

Multifunctional proteins often perform their different functions when localized in different subcellular compartments. However, the mechanisms leading to their localization are largely unknown. Recently, 3′UTRs were found to regulate the cellular localization of newly synthesized proteins through the formation of 3′UTR-protein complexes. Here, we investigate the formation of 3′UTR-protein complexes involving multifunctional proteins by exploiting large-scale protein-protein and protein-RNA interaction networks. Focusing on 238 human 'extreme multifunctional' (EMF) proteins, we predicted 1411 3′UTR-protein complexes involving 54% of those proteins and evaluated their role in regulating protein cellular localization and multifunctionality. We find that EMF proteins lacking localization addressing signals, yet present at both the nucleus and cell surface, often form 3′UTR-protein complexes, and that the formation of these complexes could provide EMF proteins with the diversity of interaction partners necessary to their multifunctionality. Our findings are reinforced by archetypal moonlighting proteins predicted to form 3′UTR-protein complexes. Finally, the formation of 3′UTR-protein complexes that involves up to 17% of the proteins in the human protein-protein interaction network, may be a common and yet underestimated protein trafficking mechanism, particularly suited to regulate the localization of multifunctional proteins.

INTRODUCTION

Protein multifunctionality, like alternative splicing, allows cells to make more with less. Among multifunctional proteins, 'moonlighting' proteins form a particular subset that performs multiple *unrelated* functions (1,2) such as the human aconitase, an enzyme of the tricarboxylic acid (TCA) cycle that also functions as a translation regulator, upon an iron-dependent conformational change (3). However, for most of the moonlighting proteins, the manner with which their distinct functions can be performed, coordinated and regulated is largely unknown. In some cases, the different functions are associated with a change in their (i) structural conformation or oligomeric states (4), (ii) interaction partners, (iii) location in tissues or cellular compartments. Indeed, in several cases, the presence of a moonlighting protein in different cellular compartments has been found to be responsible for its change in function (5,6). For instance, several intracellular chaperones, cytosolic enzymes involved in glycolysis, enzymes of the TCA cycle, and other housekeeping proteins, also function as cell surface receptors. However, in contrast to many other cell-surface receptors, none of a compilation of 30 multi-species moonlighting proteins with different functions intracellularly and on the cell surface contains any signal or motif for membrane or cell surface targeting (6,7). This suggests that the subcellular localization of multifunctional proteins that participates in their functional diversity may be regulated by a yet unknown mechanism.

In 2015, a breakthrough work (8) described a novel mechanism directing protein translocation to the plasma membrane. This mechanism involves the interaction between 3′ untranslated mRNA regions (3′UTRs) and RNA-binding proteins (RBPs) during translation, promoting the formation of a protein complex that interacts with the nascent peptide chain (8–10). In the case of CD47—a cell-surface protein involved in cellular processes including apoptosis,

adhesion, migration, and phagocytosis ([11])—the relationships between alternative 3′UTRs, protein complex formation and subcellular localization have been deciphered in detail ([8]). When translated from an mRNA with a short 3′UTR, the CD47 protein is retained in the endoplasmic reticulum, whereas the protein translated from the mRNA with a long 3′UTR localizes to the plasma membrane, thereby affecting its function. Notably, in this manner, alternative 3′UTRs can affect the function of their cognate proteins without recurring to amino acid changes. In the CD47 case, this is achieved upon the formation of a 3′UTR-protein complex mediated by the ELAVL1 RBP, which recognizes a binding site on the long 3′-UTR that is absent from the short one ([8]). The complex also contains a specific CD47 protein partner, SET, responsible for addressing the cognate nascent CD47 protein to the plasma membrane. Finally, this 3′UTR-protein complex forms within a newly described membraneless subcellular compartment, the TIGER domain, constituted by TIS granules—accumulating another RBP named TIS11B—located at the Endoplasmic Reticulum surface ([12]).

The translocation mechanism involving 3′UTR-protein complex formation has the potential to be a widespread trafficking mechanism for proteins located at the membrane ([8]). However, its prevalence is not documented. Moreover, alternative 3′UTRs could play a role in mediating the multifunctionality of proteins, as recently shown for BIRC3 ([13]), by promoting the formation of different complexes containing distinct interaction partners. There is thus a need to determine whether the formation of 3′UTR-protein complexes is a major contributor to the diversification of protein function.

As the moonlighting functions of proteins have usually been discovered by serendipity, we have previously proposed a computational approach combining protein-protein interaction (PPI) network and Gene Ontology (GO) annotation analyses to identify at large-scale moonlighting candidates that we termed 'extreme multifunctional' (EMF) proteins ([14]). These proteins are characterized as a group by particular features, constituting a signature of extreme multifunctionality ([14]). Notably, EMF proteins contain more short linear motifs than other proteins ([14]), that can mediate transient interactions and be used as molecular switches between functions ([15,16]).

Here, we aim to determine the role of the 3′UTR-protein complexes in regulating protein cellular localization and multifunctionality. We first established EMF proteins as a model to investigate the regulation of multifunctionality mediated by 3′UTRs and then predicted all 3′UTR-protein complexes plausible to be formed with EMF proteins, using large-scale protein-protein and RBP-3′UTR interaction networks. With this approach, we identified more than a thousand possible 3′UTR-protein complexes, comprising 128 out of 238 EMF proteins. Comprehensive computational analysis of the composition of the predicted 3′UTR-complexes led us to propose that the translocation of proteins between subcellular compartments—particularly to the plasma membrane—which is often associated with the functional change of multifunctional proteins could be mediated by their 3′UTRs. Notably, our hypothesis is largely supported by the finding of numerous well-known moonlighting proteins among our predicted 3′UTR-protein complexes. In addition, we extend the current knowledge on the few experimentally described 3′UTR-protein complexes ([8,13,17]) by predicting that as much as 17% of proteins in the human PPI network (*i.e.* the interactome) are able to form such complexes. The formation of 3′UTR-protein complexes could, therefore, represent a common protein trafficking mechanism that has been so far largely overlooked and underestimated.

MATERIALS AND METHODS

Protein–protein interaction network, EMF proteins and protein groups

Predicted human extreme multifunctional (EMF) proteins (238 proteins) and a human PPI network (14046 proteins, 92348 interactions) were downloaded from MoonDB 2.0 ([18]) (http://moondb.hb.univ-amu.fr/). The human PPI network was constructed with interactions retrieved from the PSICQUIC web service ([19]) on January 2018, as described in ([14]). Network modules were extracted from the PPI network using OCG ([20]), a clustering algorithm that allows proteins to belong to more than one cluster. These network modules are available on MoonDB 2.0. Briefly, EMF proteins (238 proteins) are proteins that belong to two or more network modules whose Gene Ontology (GO) term annotations ('Biological Process') contain at least two terms that are dissimilar to each other according to PrOnto ([21]). GO term annotations and ontologies were collected from the Gene Ontology Consortium ([22]) on December 2017. Analysis involving the 'proteome' protein group used a human proteome (20349 proteins) retrieved from UniProt ('reviewed' proteins only) on June 2018 ([23]).

Datasets of 3′UTRs and polyadenylation sites

Ensembl v90 spliced 3′UTR sequences for all human transcripts were downloaded from the Ensembl BioMart service ([24]). The maximum 3′UTR length was calculated for each protein in the human proteome (UniProt AC) by selecting the longest 3′UTR among all transcripts encoding for a certain protein. Genome-wide polyadenylation sites for human were downloaded from APADB v2 ([25]) as well as PolyASite version r1.0 ([26]), on December 2017. APADB polyadenylation sites per kb were calculated for proteins produced from transcripts with 3′UTRs longer than 1000 nt, considering the length of the longest 3′UTR. For PolyASite, polyadenylation sites on the terminal-exon 'TE' category were considered. Gene names and Ensembl transcript IDs were converted to UniProt AC using the UniProt ID mapping tool ([23]).

RBP–3′UTR interaction network

Interactions between RBPs and 3′UTRs were retrieved from the Atlas of UTR Regulatory Activity (AURA) v2.4.3 database (AURAlight dataset) on January 2018 ([27]). The AURA database contains interactions between 3′UTRs and

RBPs collected and mapped from various experiments, including several types of cross-linking and immunoprecipitation (CLIP) methods. Gene and coding-transcript identifiers were mapped to reviewed UniProt ACs using UniProt ID cross-referencing files (HUMAN_9606_idmapping.dat) (23). Only interactions involving proteins present in the PPI network were used. The network contains 163,490 interactions between 201 RBPs and the 3′UTR of the mRNAs of 10893 protein-coding genes.

Prediction of 3′UTR-protein complexes

3′UTR-protein complexes were predicted with in-house Python v2.7 scripts using the protein-protein and RBP-3′UTR interaction networks described above. Each 3′UTR-protein complex includes: (i) an interaction between the RBP and the 3′UTR, (ii) an interaction between the intermediate protein (*i.e.* the protein which interacts with both the RBP and the nascent protein) and the nascent protein, (iii) an interaction between the intermediate protein and the RBP. We only considered the presence of one intermediate protein, and complexes formed without any intermediate protein were not examined (*i.e.* RBP interacting directly with the nascent protein). Since the PPI network used does not contain self-interactions, the intermediate protein must be different from the nascent and the RBP. 3′UTR-protein complexes were detected for EMF as well as all proteins in the human PPI network. Only proteins with (i) PPIs, (ii) 3′UTR-RBP interactions and (iii) presence in at least one HPA tissue (see below) were liable to be assessed for 3′UTR–protein complexes as 'nascent' proteins. 3′UTR-protein complexes were further filtered according to protein tissue presence, as described below.

Protein tissue presence filter

Tissue protein presence from Human Protein Atlas version 18 (January 2018) (28) was used to filter 3′UTR-protein complexes. This dataset contains data on 58 normal tissues. Information on cell type associated with tissue names was not used in this study. Proteins with reliability score (level of reliability of the protein expression pattern) indicated as 'uncertain' and proteins with presence level 'not detected' were excluded. We only considered 3′UTR-protein complexes where all proteins of the complex are present in at least one of the 58 tissues. Gene names and Ensembl Gene IDs were converted to 13044 reviewed UniProt AC using the UniProt ID mapping tool (23).

3′UTR usage in predicted complexes

Sequences and hg19 genomic coordinates of 3′UTRs interacting with RBPs were taken from the AURA database (file: UTR_hg19.fasta, 65 285 sequences) on December 2018. As AURA does not provide cross-references to Ensembl identifiers, we used BEDTools (29) intersect with parameters '-s -r -f 1.0' (to ensure strendness and a perfect match) to map AURA UTR coordinates on GENCODE release 19 (GRCh37.p13) gene annotation file. In doing so, we were able to associate 62,898 UTR sequences (96% of the total) to the corresponding protein-coding genes. UTR sequence

redundancy was reduced at 100% identity using the CD-HIT algorithm (30).

Proteins localized in the plasma membrane

Plasma membrane proteins were retrieved from two datasets: (i) UniProt (23), querying reviewed *Homo sapiens* proteins with the GO term 'plasma membrane' (GO:0005886) annotated by 'any manual assertion' method (4602 proteins) and (ii) plasma membrane proteins experimentally detected by HPA version 18 (31), querying for the subcellular locations 'plasma membrane' and 'cell junctions' (1734 genes mapped to 1776 UniProt ACs using the UniProt ID mapping tool). Note that both datasets include proteins that are integral to the plasma membrane (*e.g.* receptors) as well as peripheral membrane proteins that may attach to integral membrane proteins or penetrate the peripheral regions of the membrane (*e.g.* receptor-interacting proteins). Information on the presence or absence of signal peptide, lipidation sites, and transmembrane domains was obtained from UniProt on January 2019 (23). The set of 7025 nascent proteins liable to form 3′UTR-protein complexes (*i.e.* having protein-protein and protein–RNA interactions, as well as present in HPA), even though not enriched in plasma membrane proteins, contain a higher proportion of proteins localized in plasma membrane without a signal peptide or transmembrane domains than the proteome (52.4% versus 31.4%, using UniProt data). Thus, to avoid potential biases, statistical comparisons were done against this set of proteins instead of the proteome in plasma membrane-related analysis.

Dissimilar cellular component GO term analysis

The PrOnto method (21) was used to identify 'cellular component' (CC) GO term pairs that are unlikely to occur in the same protein (association probability) or in interacting proteins (interaction probability). PrOnto probabilities were recalculated for CC GO term annotations collected from the Gene Ontology Consortium (22) on December 2017, including terms 'Inferred from Electronic Annotation' (IEA). GO term pairs with PrOnto association and interaction probabilities <0.05 were considered dissimilar. EMF protein dissimilar GO terms can be consulted on the related MoonDB 2.0 database. We tested whether certain protein groups have more proteins with dissimilar terms than expected. Since the ability to find dissimilar GO term pairs depends on the number of GO terms annotated to a protein, we control this by sampling (10,000 times) sets of proteins from the human proteome of the same size of the protein group in question and with the same distribution of numbers of CC GO terms annotated.

Gene Ontology enrichment analysis

The Gene Ontology (GO) term enrichment analysis was performed using the g:Profiler R package (32). We only considered enrichments where the minimum query/term intersection size was = >5, and where the *P*-value was <0.05 after correction using the Benjamini-Hochberg procedure. In all g:Profiler analysis, electronic GO annotations (*i.e.* with the evidence code IEA) were considered.

Modules with dissimilar annotations

We retrieved from MoonDB 2.0 the list of network modules and the pairs of (i) dissimilar functions (*i.e.* GO Biological Process terms, BP) annotating the modules used for the identification of the EMF proteins, and (ii) dissimilar GO Cellular Component (CC) annotating the modules. We selected EMF proteins present in at least two predicted 3′UTR-protein complexes. Subsequently, for each selected EMF protein, we assessed whether its number of EMF intermediate protein pairs present in network modules with dissimilar annotations (BP, CC or both) can be obtained by chance, by comparing it to the number of interactor pairs of the same EMF protein present in network modules with dissimilar annotations. This approach takes into account both network topology and functional annotations. For this, we performed a Fisher's exact test (two-sided, alpha = 0.05) in which the total number of possible pairs between EMF interaction partners in modules with dissimilar annotations in the human interactome is used as background.

RESULTS

EMF proteins as a model to study the regulation of protein localization and multifunctionality by alternative 3′UTRs

The usage of alternative 3′UTRs has been found to regulate the subcellular localization of proteins (8,33), which in turn could regulate their functions. Therefore, we hypothesize that 3′UTRs may play a role in moonlighting protein function regulation. To investigate this possibility, we explore a comprehensive set of human moonlighting protein candidates, *i.e.* 238 'extreme multifunctional' (EMF) proteins (Supplementary File 1: Table S1) that we identified using our MoonGO approach (14) and that are available in our recently updated MoonDB 2.0 database (18) (see Methods).

Numerous moonlighting proteins have been found to perform different functions when localized in different cellular compartments. Coherently, we found that EMF proteins are annotated with significantly more 'Cellular Component' (CC) GO terms than other groups of proteins (see Materials and Methods, Figure 1A). Indeed, on average an EMF protein is associated with 7.8 CC GO terms, whereas the average in the PPI network—from which the EMF proteins are identified—is 4.6 CC GO terms (Mann–Whitney U test P-value = 1.1×10^{-22}). We also observed that 111 EMF proteins (46.6% of the total) belong to the multilocalizing proteome, defined by immunofluorescence-based approaches in the Human Protein Atlas (HPA) (31), a significantly higher fraction than in the PPI network (36.7%, two-sided Fisher's Exact Test, odds ratio (OR) = 1.52, P-value = 1.7×10^{-3}) (Figure 1B). Moreover, using a dataset of 1233 known and machine learning-predicted translocating proteins from the Translocatome database (34), we found that 41.2% EMF proteins (98 out of 238) could change their subcellular localization upon a regulatory event, representing, again, a higher fraction than in the PPI network (9%, two-sided Fisher's Exact Test, OR = 7.8, P-value<2.2×10^{-16}) (Figure 1B). Overall, this indicates that EMF proteins are more often localized in different cellular districts than expected, a feature that may be regulated by alternative 3′UTRs.

Next, given the fact that many proteins perform their moonlighting functions when localized at the plasma membrane, we sought for EMF proteins associated with the plasma membrane. Indeed, we found that 65 out of 238 (27.3%) EMF proteins have been observed at the plasma membrane according to GO term annotations ('plasma membrane', GO:0005886, see Materials and Methods). This is significantly more than expected when considering the PPI network as the background (24.1%, two-sided Fisher's Exact Test, OR = 1.46, P-value = 1.22×10^{-2}; Figure 1b). Coherently, the same trend is observed when considering the proteins localized at the 'plasma membrane' according to HPA (see Materials and Methods). However, although enriched in proteins localized at the plasma membrane, very few EMF proteins (31, *i.e.* 13%) contain a signal peptide or a transmembrane domain or a post-translational modification site (*i.e.* amino acid lipidation site) that could explain their subcellular location. Strikingly, these features are depleted or not significantly enriched in EMF proteins when compared to the PPI network considered as a group (two-sided Fisher's Exact Test, Signal peptides: 5.4%, OR = 0.32, P-value = 4×10^{-6}; Transmembrane domains: 5.8%, OR = 0.23, P-value = 1×10^{-10}; Lipidation sites: 5.8%, OR = 1.46, P-value = 0.18, Figure 1b). Altogether, this suggests that part of the EMF proteins is not localized at the membrane *via* the classical protein transport pathway.

In order to estimate the potential of EMF proteins to be regulated by the 3′UTR mechanism, we investigated the features of their 3′UTR sequences. Using 3′UTR models from the Ensembl database (24), we found that mRNAs encoding EMF proteins have significantly longer 3′UTRs than mRNAs encoding all other human proteins of the proteome, and more stringently, of the interactome (see Materials and Methods, mean 'EMF' (1987 nt) versus 'Proteome' (1643 nt), Mann-Whitney U test P-value = 1.5×10^{-3}; mean EMF (1987 nt) versus 'Interactome' (1739 nt), Mann-Whitney U test P-value = 3.0×10^{-2}) (Figure 1C).

In addition, using polyadenylation sites from APADB (25) and PolyASite databases (26), we showed that mRNAs encoding EMF proteins bear a higher number of alternative polyadenylation (APA) sites in 3′UTRs than mRNAs encoding the other groups of proteins (Figure 1D; mean EMF (2) versus 'Interactome' (1.6), Mann–Whitney U test P-value = 2.8×10^{-4}; Supplementary File 2: Supplementary Figure S1). Consistent with these findings, EMF proteins have significantly more 3′UTR isoforms than the other protein groups (Supplementary File 2: Supplementary Figure S2). Together, these results suggest that mRNAs encoding EMF proteins are more likely to be regulated by their 3′UTRs than those encoding other proteins.

Overall, the functions of EMF proteins could be affected by their cellular localization, particularly at the plasma membrane. Moreover, with longer and more variable 3′UTRs, the EMF proteins thus have the potential to be regulated by a mechanism involving their 3′UTR. EMF proteins, therefore, constitute a suitable model to investigate the role of 3′UTR–protein complex formation in the localization and the multifunctionality of proteins.
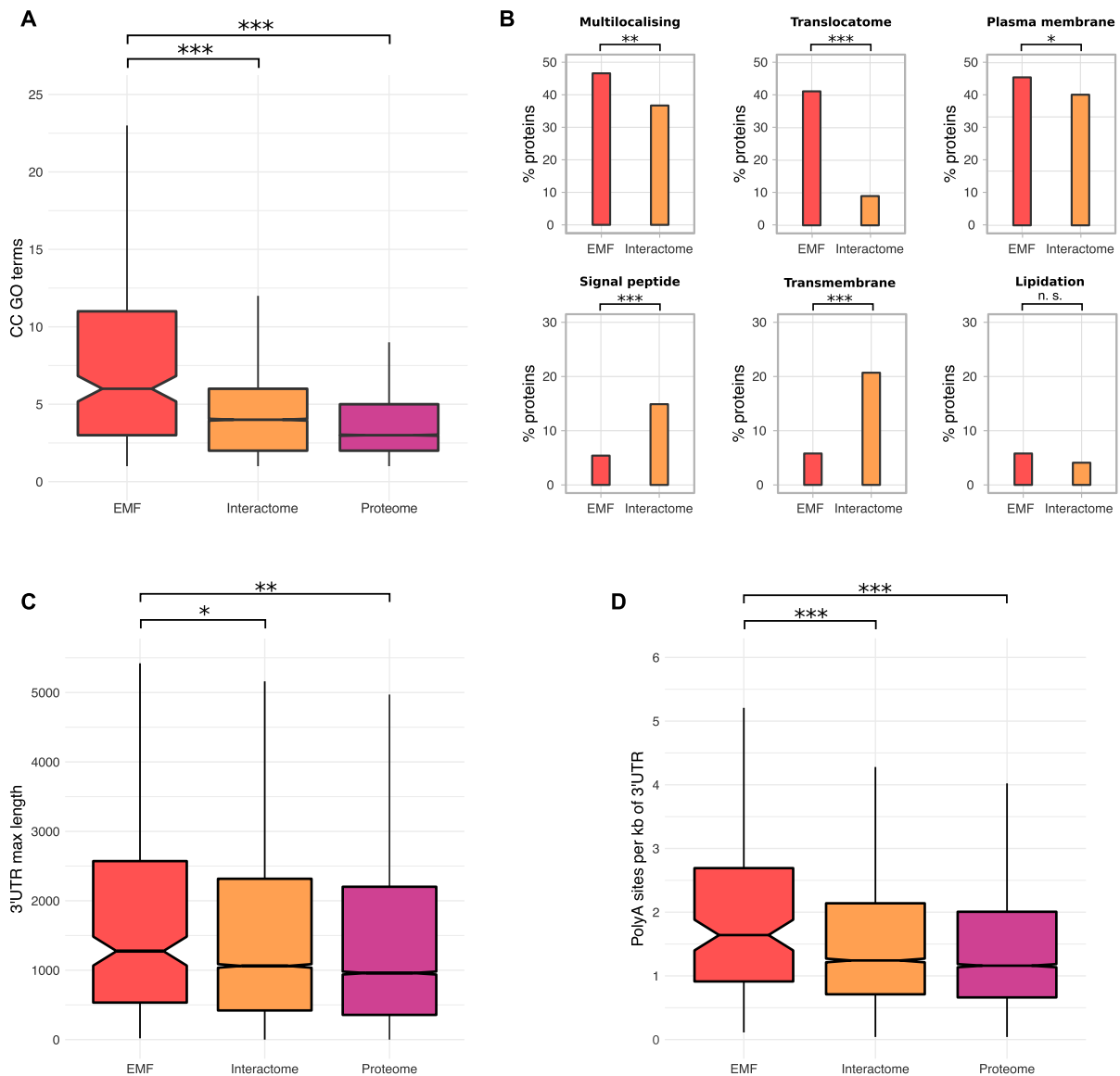
Figure 1. Cellular localization of EMF proteins and 3′UTR-related features of EMF proteins. (**A**) Comparison of number of Cellular Component (CC) GO term annotations between protein groups (Kruskal–Wallis rank sum test, P-value $< 2 \times 10^{-16}$): EMF ($n = 235$), interactome ($n = 12\,794$), and proteome ($n = 18\,298$). Mann–Whitney U tests were performed to assess pairwise statistical significance. The Benjamini-Hochberg procedure was applied for multiple test corrections. Significance: '***' indicates a FDR < 0.001. (**B**) Enrichment analysis of cellular localization annotations and membrane-targeting signals. Significance: '*' indicates a P-value < 0.05; '**' indicates a P-value < 0.01; '***' indicates a P-value < 0.001. (**C**) Comparison of maximum 3′UTR lengths. Kruskal-Wallis rank sum test, P-value $= 3 \times 10^{-14}$. (**D**) Number of polyadenylation (polyA) sites per kb of 3′UTR, using APADB data (22). Kruskal–Wallis rank sum test, P-value $= 5 \times 10^{-8}$. Only transcripts with 3′UTRs longer than 1000 nucleotides were considered. Mann–Whitney U tests were performed to test for pairwise statistical significance. The Benjamini-Hochberg procedure was applied for multiple test corrections. Significance: '*' indicates a FDR < 0.05; '**' indicates a FDR < 0.01; '***' indicates a FDR < 0.001.

Prediction of 3′UTR–protein complexes

In order to assess the possible involvement of the 3′UTR in the regulation of the localization and the function of EMF proteins, we predicted the potential formation of 3′UTR-protein complexes containing human EMF proteins. As the mechanism involves the recruitment of RBPs to the site of translation by 3′UTRs, which in turn may promote the co-translational formation of protein complexes that interact with the nascent peptide chain (8–10), the 3′UTR-protein complex formation conceptually involves the following components: (i) an mRNA with a 3′UTR; (ii) the

cognate protein being translated (hereby termed 'nascent' protein); (iii) an RBP able to bind the 3′UTR; (iv) one other protein (hereby termed 'intermediate' protein) that interacts with the RBP and the nascent protein. By searching for sets of co-interacting 3′UTRs, RBPs, nascent and intermediate proteins (Figure 2; see Materials and Methods) from two large-scale experimental datasets forming a RBP-3′UTR and a protein-protein interaction networks (from the AURA (27) and the MoonDB 2.0 (18) databases, respectively; see Materials and Methods), we identified possible 3′UTR-protein complexes. To increase the biological
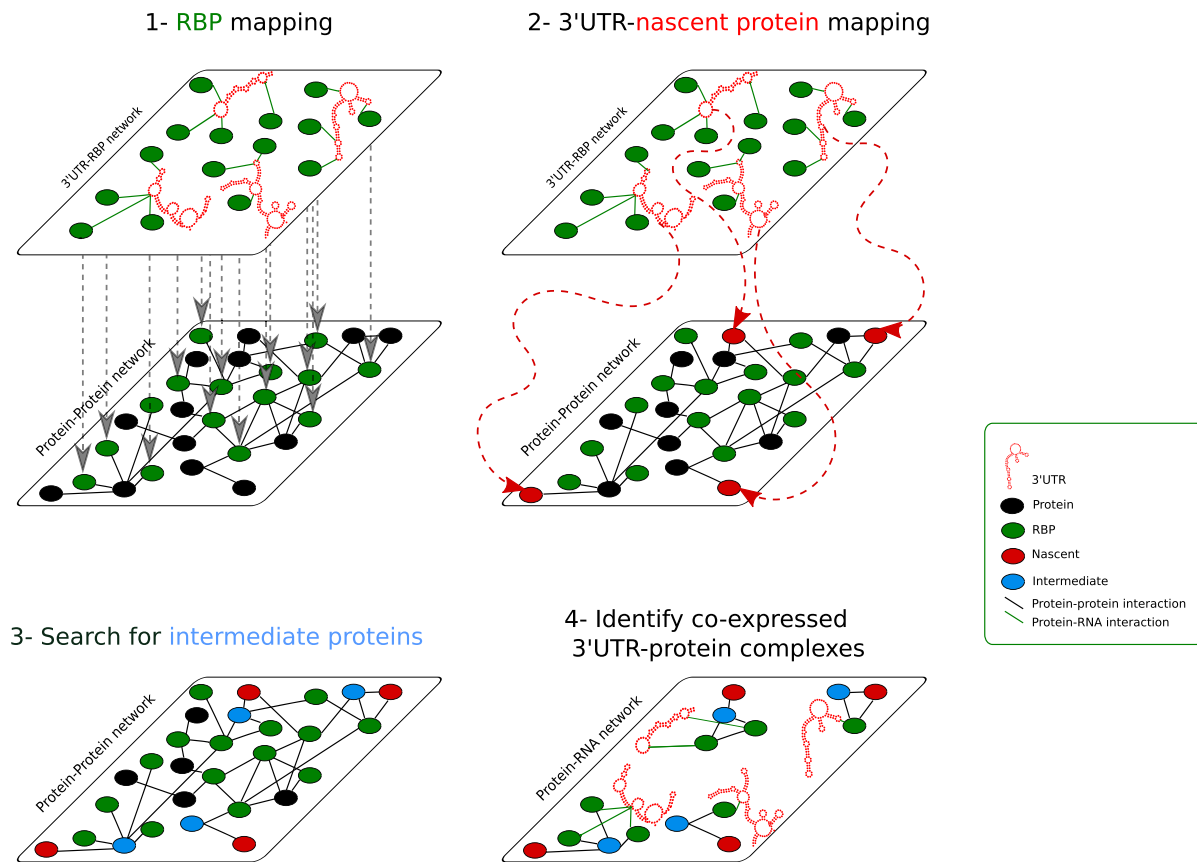
Figure 2. Workflow of the 3′UTR-protein complex prediction. Experimental RBP interactions with 3′UTRs of mRNA are retrieved from AURA v2 database A large-scale PPI network is retrieved from MoonDB 2.0 (see Materials and Methods). 3′UTR-protein complexes are predicted by (1) mapping the RBP (green nodes) interacting with 3′UTRs onto the PPI network, (2) finding cases in which the 3′UTR of a 'nascent' protein (protein under synthesis, red node) interacts with RBP, which in turn (3) interacts with an interaction partner of the nascent protein ('intermediate', blue node). Finally, (4) only 3′UTR complexes where the nascent, RBP and intermediate proteins are present in at least one same tissue are kept (Human Protein Atlas (HPA), 58 normal tissues).

relevance of the predictions, complexes should contain only proteins co-detected in at least one of the 58 HPA normal tissues (28).

A set of predicted 3′UTR-complexes containing EMF proteins

Using our novel approach, starting from 238 EMF proteins, we predicted a total of 1411 distinct 3′UTR-protein complexes comprising 128 EMF proteins and a combination of 87 RBPs and 440 interacting intermediate proteins (Table 1, Supplementary File 1: Table S2 and S3). Notably, 53.8% of the EMF proteins (128 out of 238) may form at least one 3′UTR-protein complex whereas a much lower percentage of the proteins of the interactome do so. Indeed, 16.9% of the proteins of the interactome are found in 9657 3′UTR-complexes (Fisher's Exact test, two-sided, OR = 6.46, P-value = 2.44×10^{-43}) (Table 1, Supplementary File 3).

To confirm the plausibility of our predictions, we estimate the number of 3′UTR-protein complexes formed by chance while shuffling 10 000 times all proteins in the PPI network. The numbers of 3′UTR-protein complexes predicted are 2.8–12 times higher than expected by chance for interac-

tome and EMF proteins, respectively (Supplementary File 2: Table S6), therefore supporting our predictions.

Only 42% of the RBPs known to interact with a 3′UTR in AURA database, and present in the PPI network (87 out of 204, Table 1) are found in EMF-containing complexes, despite the general propensity of RBPs to bind a large number of RNAs (35). This, therefore, increases our confidence in the specificity of the predicted 3′UTR complexes. EMF proteins are often highly connected in the PPI network (14). Indeed, 80% of them are hubs, defined here as nodes whose degree is at least twice the network average (≥ 32). They are thus potentially more likely to form 3′UTR-protein complexes, as the number of partners of a protein in the network could impact the number of 3′UTR complexes that could form. However, we found that these are not strongly correlated (Spearman correlation, rho = 0.326, P-value = 2.66×10^{-7}, Supplementary File 2: Supplementary Figure S3), indicating that the number of interaction partners of EMF proteins does not fully account for the number of observed 3′UTR complex formation. This suggests that other EMF protein features could influence the number of 3′UTR-complexes they are involved in.

Alternative 3′UTRs were found to regulate the localization and/or function of CD47 and BIRC3 (8,13). Notably,

Table 1. Summary of the 3′UTR-protein complex prediction. Percentages in the 'Nascent proteins in complex' column are relative to the initial number of proteins in the protein group (238 for EMF, 14046 for interactome)

Protein group	Nascent proteins in complex	RBPs	Intermediate proteins	3′UTR-protein complexes
EMF	128 (53.8%)	87	440	1411
Interactome	2354 (16.8%)	113	789	9657

we found that the vast majority of the predicted complexes, i.e. 1020 out of 1411 (72,3%), contains EMF proteins with at least two alternative 3′UTR isoforms (80 EMF proteins out of 128). Moreover, 60% of these complexes containing EMF proteins with several 3′UTR isoforms (612 out of 1020), are predicted to form exclusively by the binding of RBPs to the long 3′UTR isoforms (Figure 3A), as in the known cases. Interestingly, similar results are obtained when we extend the analysis to the 3′UTR-protein complexes predicted from the whole interactome (Supplementary File 2: Supplementary Figure S4). Altogether, these results underline the specificity of the 3′UTR-complexes formed and correctly recapitulate the knowledge obtained on the few complexes previously investigated experimentally (8,13). This, therefore, encouraged us to pursue our analysis and test the possibility that 3′UTR-protein complex formation may contribute to the regulation of EMF –and other protein– functions.

The role of 3′UTR–protein complexes in the subcellular localization of EMF proteins

The 128 EMF proteins predicted to belong to at least one 3′UTR-protein complex are enriched in plasma membrane annotations. Indeed, according to GO annotations, 45 out of the 128 EMF proteins in complexes (35%, Supplementary File 1: Table S4) have been associated to the plasma membrane, a significant enrichment when compared to the human interactome (OR = 1.5, two-sided Fisher's Exact Test, P-value = 3.4×10^{-2}, Table 2). Remarkably, 36 out of the 45 EMF proteins in complexes and annotated to the plasma membrane (80%, Supplementary File 1: Table S4) do not display any membrane addressing signal, such as signal peptides, transmembrane or intramembrane domains, or lipid anchors, all features known to be critical for localizing proteins to the membrane (36,37). This proportion is remarkably higher than in the PPI network, where it drops to 47.5% (OR = 4.56, two-sided Fisher's Exact Test, P-value = 1.2×10^{-5}) (see Materials and Methods). Furthermore, we observed a similar trend using an independent set of plasma membrane proteins from the HPA database (31) (see Methods; Supplementary File 2: Table S7). As the lack of addressing signal has been remarkably observed for dozens of cell surface moonlighting proteins (7) performing alternate functions in other subcellular locations, we speculate that the formation of 3′UTR-protein complex could play a role in the unconventional plasma membrane translocation of the EMF proteins and more generally in the regulation of the localization of the EMF proteins.

To investigate the potential role of 3′UTR–protein complex formation in protein localization, we assessed whether EMF proteins found at the plasma membrane have been also observed in other 'unexpected' cellular locations. For this, we employed the PrOnto method (21) which identifies dissimilar GO term pairs unlikely to occur among the annotations of the same protein, or among the annotations of

interacting proteins (see Materials and Methods). Doing so, we discovered that more EMF proteins than expected are annotated to at least one pair of dissimilar CC GO terms (Figure 3B; 46.6% observed versus 41.1% expected; empirical P-value = 0.022, 10 000 randomizations). Remarkably, the rarity of the observed proportion compared to expected (measured by the empirical P-value) increases when considering EMF proteins in 3′UTR-protein complexes (Figure 3B; 58.14% observed versus 48.72% expected; empirical P-value = 9×10^{-3}). This is particularly true for the subset of 36 EMF proteins found at the plasma membrane (Figure 3B; 92.3% observed versus 60.07% expected; empirical P-value = 1×10^{-4}). Indeed, this latter result postulates that 33 out of the 36 EMF proteins, besides being present at the plasma membrane, are also present at another 'unexpected' cellular location, demonstrating the great cellular localization versatility and fine-tune regulation of these proteins. Interestingly, 31 out of 33 of these proteins localize to the nucleus in addition to their plasma membrane localization, whereas only 5 of them have a nucleus addressing sequence (NES/NLS) (38) (Supplementary File 1: Table S5).

Overall, the vast majority of the EMF proteins found in 3′UTR–protein complexes and associated with the plasma membrane lack conventional membrane addressing signals and can also be found in the nucleus despite the absence of NES/NLS signals. This leads us to propose that 3′UTR-protein complexes could play a key role in the translocation of these extreme multifunctional proteins, not only by participating in protein transport to the plasma membrane as described for CD47 (8) but more generally, in protein trafficking between different subcellular compartments.

Predicted 3′UTR–protein complexes could explain protein multifunctionality

If 3′UTR-complexes formation contributes to protein multifunctionality, distinct 3′UTR-complexes assembled with several RBPs and different intermediates proteins, could provide the molecular environment for the EMF protein to perform different functions by interacting with different protein partners. In order to test this hypothesis, our reasoning was the following.

The EMF proteins used in this study were identified at the intersection of at least two network modules (20) involved in dissimilar cellular processes, according to GO Biological Process annotations (14). If the formation of a 3′UTR–protein complex contributes to the multifunctionality of a given EMF protein as recently shown for BIRC3 (13), we expect that the following conditions are met: (i) the EMF protein participates to different 3′UTR complexes (at least two); (ii) the recruited protein partners, *i.e.* the different intermediate proteins found in these complexes should belong to different network modules involved in dissimilar biological processes (Figure 4A). First, we found that 80% of the EMF proteins present in 3′UTR complexes fulfill
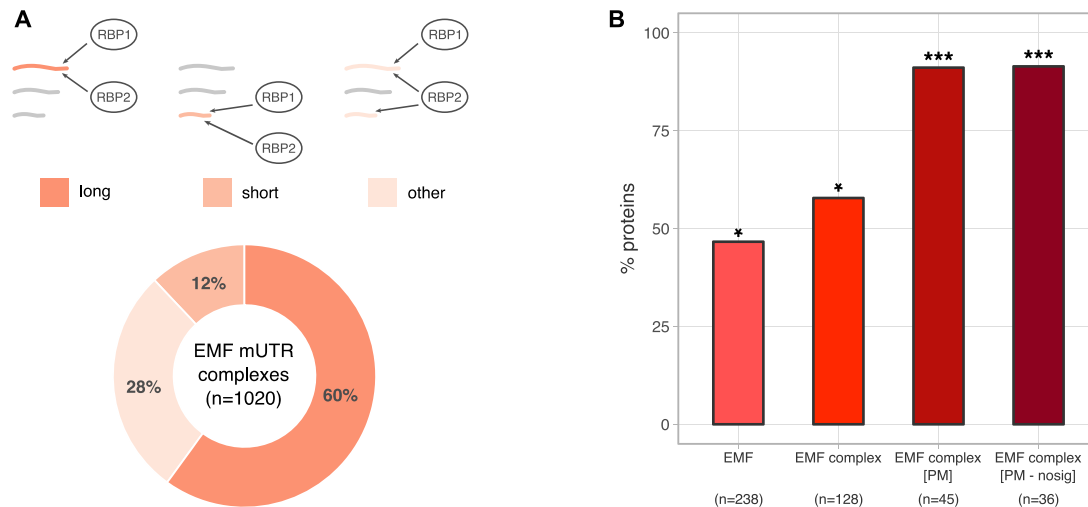
Figure 3. Preferential isoform usage in multiple UTR ('mUTR') EMF-containing predicted complexes and percentage of proteins with GO term annotations to dissimilar cellular components. (**A**) Nascent mUTR EMF proteins were classified according to the length of the 3′UTR(s) bound to RBPs: the longest or the longer isoforms are bound to RBPs ("long"); the shortest or shorter isoforms are bound to RBPs ("short"); RBPs bind all isoforms, or a substantial fraction of them, without a clear preference for long or short isoforms ("other"). (**B**) Dissimilar CC GO terms were determined with the PrOnto method. Each bar represents the percentage of proteins in the protein group annotated to at least one dissimilar GO term pair. We tested whether the observed proportion was higher than the expected values by sampling (10 000 times) sets of proteins from the human proteome of the same size than the considered protein group, and with identical distribution of annotated CC GO terms. 'EMF' refers to the 238 EMF proteins. 'EMF complex' represents the 128 EMF proteins in a 3′UTR-protein complex. 'EMF complex [PM]' refers to the subset of 45 EMF proteins which are found at the plasma membrane. 'EMF complex [PM-nosig]' refers to the subset of 436 EMF proteins which are found at the plasma membrane and lack addressing signal. Significance: '*' indicates an empirical *P*-value < 0.05; '***' indicates an empirical *P*-value < 0.001.

Table 2. Numbers of nascent proteins in 3′UTR-protein complexes localized at the plasma membrane and without conventional addressing signals. Percentages denote the proteins retained compared to the previous column. Where indicated, Fisher's exact tests were performed to test for statistical significance using as background the set of 7025 nascent proteins liable to be assessed for 3′UTR-protein complexes. In this background, 1877 proteins are associated with the plasma membrane, of which 892 lack a membrane addressing signal. The Benjamini-Hochberg procedure was applied for multiple test corrections. Significance: '*' indicates a *P*-value <0.05; '**' indicates a *P*-value <0.01; '***' indicates a *P*-value <0.001

| Protein group	Nascent proteins in complex	Nascent proteins in complex, localized at the plasma membrane	Nascent proteins in complex, localized at the plasma membrane, no addressing signal
EMF	128	45 (35%) *	36 (80%) ***
Background	7025	1877 (26.7%)	892 (47.5%)

the first condition (102 out of 128). Second, among these, 23,5% (24 out of 102) meet the second one (measured as the number of EMF proteins that has significantly more intermediate protein pairs in distinct network modules involved in dissimilar biological processes than expected by chance, with a *P*-value < 0.05) (see Materials and Methods, Figure 4B, Supplementary File 2: Table S8), thus proposing a molecular scenario for the multifunctionality of a quarter of the EMFs involved in 3′UTR-protein complexes. Similarly, since a change of function is often associated with a change in cellular localization, we checked how many of the 102 EMF proteins present in two or more 3′UTR complexes, have at least two intermediate proteins belonging to different network modules annotated with dissimilar GO cellular localizations (see Materials and Methods, Figure 4A). This was the case for 25.4% of these EMF proteins (26 out of 102) (Figure 4B, Supplementary File 2: Table S9). Altogether, a total of 35% of these EMF proteins (36 out of 102) (Figure 4B, Supplementary File 2: Table S9 and S10) have more intermediate protein pairs partitioning either in network modules annotated with dissimilar BP (24) or CC (26) GO terms or both (36) than expected by chance, with a *P*-value <0.05. Overall, the formation of 3′UTR-protein

complexes seems to provide two important prerequisites for multifunctionality, based on the topology and the annotations of network modules: diversity in interaction partners and distinct subcellular locations.

3′UTR–protein complexes could mediate multifunctional protein trafficking

If the 3′UTR-protein complexes participate in protein transport between subcellular compartments, EMF protein functional annotations should reveal and highlight the biological processes requiring their formation. For this, we performed a GO term analysis on the 128 EMF and the 440 intermediate proteins contained in the 1411 3′UTR-protein complexes. Significant enrichments in membrane-related cellular components, nucleoplasm and cytosol evoking cellular trafficking, and in functions related to signaling processes are observed in both categories (Supplementary Figure S5). In addition, conversely to EMF proteins, intermediate proteins are enriched in proteins containing potential NES/NLS signals according to NLSdb (38) (OR = 1.6, two-sided Fisher's Exact Test, *P*-value = 3.70 × 10^{-5}), sug-
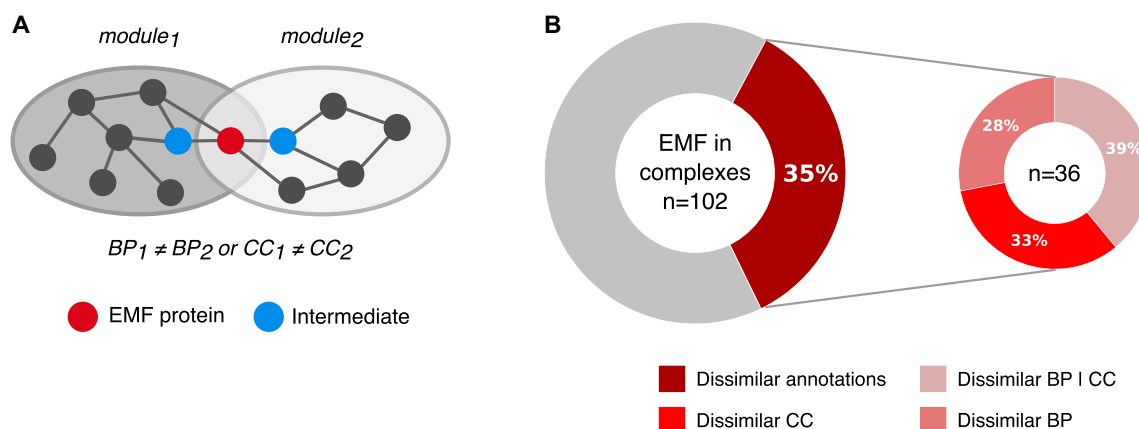
Figure 4. Network module topology and annotations support a 3′UTR-complex role in multifunctionality. (**A**) Network modules containing EMF proteins were annotated with GO biological process (BP) and cellular component (CC) terms. In a pairwise fashion, for each EMF protein, we check if any of its intermediate proteins were present in distinct network modules with dissimilar BP or CC annotations. (**B**) Proportion of EMF proteins in at least two complexes that have more intermediate protein pairs belonging to network modules with dissimilar annotations than expected by chance, with a *P*-value <0.05 (left). Split proportions of EMF according to the type of dissimilar annotations (BP, CC or both) (right).

gesting their possible role in translocating EMF proteins between compartments.

Archetypal moonlighting proteins are predicted to form 3′UTR-protein complexes

As our analysis has been performed on a dataset of candidate proteins to extreme multifunctionality, we aimed at confirming that experimentally verified moonlighting proteins could participate in 3′UTR-protein complexes. For this, we first verified that proteins of the reference dataset available in the MoonDB 2.0 (18) and the MoonProt (39) databases (47 and 65 literature-curated moonlighting proteins, respectively), form such complexes. Indeed, up to 52% of the known moonlighting proteins do so, a proportion that is similar to the one obtained for the 238 predicted EMF proteins (54%, Table 1). Moreover, the number of predicted complexes is significantly higher than expected by chance for both moonlighting proteins datasets (Supplementary File 2: Table S8).

Interestingly, our set of EMF proteins in 3′UTR-protein complexes includes the alpha-enolase (ENO1), a well-known cytosolic glycolysis enzyme that moonlights at the cell surface as a receptor and activator of plasminogen, being thereby involved in cell migration, tissue remodeling, apoptosis and tumorigenesis (40). Notably, ENO1 lacks a signal sequence and the mechanism leading to its translocation between subcellular compartments is yet unknown (40). We found that ENO1 forms 8 different 3′UTR-protein complexes with different RBPs and intermediate proteins with different tissue expression. Interestingly, one of them is only predicted in muscles (heart, skeletal and smooth muscles). It contains DDX6, an RNA helicase involved in muscular dystrophy (41) as RBP and Desmin (DES), forming myofibrils and linking them to the cytoskeleton, the nucleus, mitochondria, and the plasma membrane, as intermediate. It is thus tempting to speculate that this complex participates or mediates the plasminogen receptor function of ENO1 in muscle regeneration and muscle injury recovery (40) by localizing it at the muscle cell surface. Finally,

the fact that 4 out of 5 RBPs involved in the 3′UTR-ENO1 complexes bind only the longer ENO1 3′UTRs, as described for CD47 (8) and BIRC3 (13) and that 4 out of 7 intermediates proteins are annotated to both intracellular and extracellular locations, increases our confidence in the predictions.

Likewise, we also predicted a 3′UTR-protein complex formed with the nascent protein RHAMM/HMMR, a regulator of the stability of the mitotic spindle in normal cells that acts as an extracellular CD44-ligand promoting cell motility and invasion in cancer cells (42), IGF2BP2, involved in mRNA storage and transport, and the Dynactin subunit 1 (DCTN1) as intermediate, a protein present at the 'spindle' as well as at the 'membrane' and notably involved in the transport of organelles and vesicles by tethering the dynein cargo to the microtubule (43). We can thus hypothesize that the 3′UTR-RHAMM complex plays a role in the molecular mechanism allowing the extracellular localization of RHAMM in tumor cells.

Finally, in addition to enolase, almost all other glycolytic enzymes have been shown to display moonlighting activities unrelated to glycolysis. These additional functions often explain the phenotypes observed in the metabolic disorders caused by the dysregulation of these enzymes (44). Very interestingly, we found that not only proteins involved in the 'glycolytic process' are enriched among the nascent proteins of the interactome present in the 3′UTR-protein complexes (*P*-value = 3.0 × 10^{-3}, Supplementary File 3), but that all the enzymes of the glycolysis pathway are involved in 3′UTR-protein complexes, further pushing the idea that 3′UTR-protein complex formation plays a role in protein multifunctionality.

The formation of 3′UTR-protein complexes is a prevalent mechanism at the interactome-scale

What is the prevalence of the 3′UTR-protein complexes formation? Is it a widespread mechanism in the cell and not limited to known moonlighting and EMF proteins? As shown in Table 1, our predictions revealed that 17%

of the proteins of the interactome have the potential to be involved in at least one 3′UTR–protein complex, i.e., ∼3 times more than expected randomly (empirical *P*-value = 0.0002, Table S5). Interestingly, we found that nascent proteins involved in 3′UTR–protein complexes show many of the features previously described for the EMF proteins: they are enriched in translocating and multilocalizing proteins, although largely devoid/depleted in membrane addressing features and signal peptides. Additionally, the enriched GO annotations among the 2354 nascent proteins and their intermediate proteins recapitulate the biological processes and the subcellular compartments found as enriched for EMF proteins and their intermediates including cell-cell signaling, extrinsic and cytoplasmic membrane parts, coated membrane, extracellular exosome and as well as nuclear part (Supplementary File 3).

Overall, our results suggest that the formation of 3′UTR-protein complexes could be a prevalent mechanism in the cell that could promote protein multifunctionality through protein trafficking and/or subcellular localization change for 17% of the proteins of the interactome.

DISCUSSION

What are the molecular mechanisms that enable the functional changes of moonlighting proteins? Could the formation of 3′UTR-protein complexes explain the unexpected membrane localization of some multifunctional/moonlighting proteins in the absence of membrane addressing signals? Could these ribonucleoprotein complexes promote the multifunctionality of some proteins by recruiting different interaction partners that are necessary to their switch of function?

The regulation of protein localization and function by the formation of ribonucleoprotein complexes mediated by 3′UTRs, have been proposed for two proteins, namely CD47 (8) and BIRC3 (13). However, the prevalence of the formation of such complexes is still unknown. In order to circumvent the current scarcity of 3′UTR-complexes that hinders any large-scale analysis, we propose a computational mapping strategy that allows inferring, for the first time, those ribonucleoprotein complexes. For this, we took advantage of the availability and the combination of two network data types—3′UTR–protein and PPI networks —, covering a large part of the human transcriptome and proteome. Our inferences allowed us to both provide a detailed analysis of the characteristics of the 3′UTR-protein complexes as a group and evaluate the prevalence of their formation at the interactome level. We indeed discovered that as much as 54% of the extreme multifunctional proteins and up to 17% of the whole interactome proteins could be involved in such complexes, thereby underlining their potential high extent and cellular importance. However, our inferences are notably sensitive to the comprehensiveness of the analyzed data since cellular 3′UTR-protein complexes that are formed by molecular interactions not present in the analyzed datasets cannot be found. Indeed, while experimental human PPI networks cover most interacting proteins (13), public interaction databases may not include all the interactions known in the literature or recently discovered. In addition, when building our PPI network to be analyzed, we select binary interactions to increase the accuracy and the confidence of our predictions. For instance, the ELAVL1-SET-CD47 protein complex is found by our approach only when we extend our large-scale PPI network with interactions identified by co-immunoprecipitation techniques (data not shown). Moreover, given that current RBP-3′UTR interaction datasets contain data for only a subset (less than 400 RBPs (27)) of an increasingly growing number of proteins interacting with RNAs (which may amount to as many as 2000 RBPs (45,46)), we may have underestimated the number of existing 3′UTR-protein complexes in human cells. In addition, the contribution of ribonucleoprotein complexes to a number of cellular processes may have been largely overlooked so far, due to the fact that usual experimental methods to identify cellular macromolecular complexes routinely use an RNA nuclease step before protein purification, thereby hindering the possible detection of RNA components in protein complexes (47). Computational approaches can help to overcome this drawback, as suggested by our results. Nevertheless, this study provides for the first time an extensive overview of the 3′UTR-protein complex formation for a subset of human proteins, predicting that a sizeable amount of such cellular complexes can be formed, employing a large variety of RBPs and intermediate protein components. As studying the formation of 3′UTR complexes in the context of particular biological processes (e.g. development, cell differentiation), cell types or diseases would greatly inform us on the functional role of these complexes, this should be object of further analyses.

The diversity in subcellular localization and protein partners are two main protein multifunctionality determinants possibly influenced and driven by 3′UTRs. 3′UTRs are generally described as responsible for the localization of their cognate mRNAs in asymmetrical and polarized cells such as neurons, where localized translation modifies the nearby proteome in response to external cues (9,48). This is mediated by RBPs that interact with motor proteins, allowing the transport of mRNA along actin cables. On the other hand, 3′UTRs also mediate PPIs, as in the case of the 3′UTR-protein complexes described in our study. Notably, the enrichments in the GO terms 'nucleic acid transportation', 'synapse', 'dendrite', coupled to the function 'learning or memory' among EMF and intermediate (Supplementary Figure S5A) proteins suggest that these two 3′UTR functions, mRNA transport and protein complex scaffolding, are related and intermingled.

What could be the roles of the 3′UTR-protein complexes in protein localization change? In the CD47 case, protein localization to the plasma membrane depends on the intermediate protein SET and its protein partner, active RAC1 (8). Very interestingly, on a global scale, the enrichment in 'Rac protein signal transduction' annotation among intermediate proteins that we observe (Supplementary Figure S5b) could extend this dependency to a subset of predicted 3′UTR-protein complexes containing EMF proteins. This would suggest that small GTPases signaling generally contribute to the protein transport mediated by the formation of the 3′UTR-complexes. In addition, given the role of RAC signaling in actin remodeling and cell migration (49), it is interesting to note that other proteins involved in the regu-

lation of cell migration are also enriched among intermediate proteins (Supplementary Figure S5B). This underlines a possible involvement of the trafficking mechanism we describe in cell migration processes.

Most multifunctional proteins predicted to be translocated by the 3′UTR-protein complexes do not contain the corresponding conventional addressing signals in their sequence. However, most of the multifunctional proteins participating in 3′UTR-protein complexes are located in the nucleus and at the membrane. The 3′UTR-complex and more particularly, the intermediate protein could help to address properly the nascent protein. Indeed, the fact that intermediate proteins are enriched in NLS/NES signals suggest that they may contribute to the nuclear localization of their protein partners without NLS, which would then behave as cargos. Moreover, according to GO term enrichment analysis, some intermediate proteins bind NLS sequences. These are importins and transportins able to bind the nuclear pore complex and involved in nuclear protein import and export. Interestingly, these intermediates could contribute to the nuclear localization of their protein partners devoid of NLS, as the importance of the three-dimensional context rather than just the presence of the NLS has been recently illustrated for the recognition of cargos by importins (50). Together with the fact that intermediates are also enriched for endocytosis, receptor binding, and internalization, these results confirm the role of the 3′UTR-protein complexes in protein transport between the subcellular compartments at the cell scale.

Protein multifunctionality can be controlled by the formation of 3′UTR-protein complexes. Depending on the cell type and cellular conditions, alternative 3′UTR isoforms may bind different RBPs, that may interact with different intermediates, ultimately leading to different 3′UTR-protein complexes containing the same nascent protein but serving different functions. We showed this is the case for half of the EMF proteins contained in 3′UTR complexes. Notably, it also has been shown for BIRC3, originally known as an E3 ubiquitin ligase, but which has been recently discovered to be involved in protein trafficking, chromatin regulation, and mitochondrial processes as well, through the 3′UTR-protein complexes it belongs to (13).

The assembly of 3′UTR-complexes may provide the molecular context and proximity necessary for protein moonlighting functions. Indeed, a large number of intermediate proteins are kinases that can modify nascent proteins, illustrating the fact that moonlighting functions may be partly commanded by short linear motifs (14,51). Recently, it has been shown that 4 out of 9 *S. cerevisiae* protein complexes that form co-translationally (52), possibly according to a 3′UTR-complex formation model (53), involve 6 known multifunctional and moonlighting yeast proteins (52). Interestingly, two of those, MetRS and GluRS, possess nuclear and mitochondrial localization signals, which are only revealed when the proteins are not in complex with each other (54), indicating that their co-translational complex assembly –and possibly 3′UTR–protein complex formation– may regulate their moonlighting function.

Overall, our data bring the first large-scale prediction and analysis of 3′UTR-protein complexes. We propose that the formation of 3′UTR-protein complexes is a widespread phenomenon allowing (i) the transport of proteins between subcellular compartments in the absence of conventional addressing signals, and (ii) the assembly of multiple complexes that sustain the multifunctionality of proteins, thus representing a plausible molecular mechanism for the regulation of protein moonlighting functions.

DATA AVAILABILITY

All data generated are included in this article (and its Supplementary Files). All other data are available from the corresponding authors upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Jeffery,C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.*, **24**, 8–11.
2. Piatigorsky,J. and Wistow,G.J. (1989) Enzyme/crystallins: gene sharing as an evolutionary strategy. *Cell*, **57**, 197–199.
3. Volz,K. (2008) The functional duality of iron regulatory protein 1. *Curr. Opin. Struct. Biol.*, **18**, 106–111.
4. Jeffery,C.J. (2014) An introduction to protein moonlighting. *Biochem. Soc. Trans.*, **42**, 1679–1683.
5. Ostrovsky de Spicer,P. and Maloy,S. (1993) PutA protein, a membrane-associated flavin dehydrogenase, acts as a redox-dependent transcriptional regulator. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 4295–4298.
6. Jeffery,C.J. (2018) Protein moonlighting: what is it, and why is it important? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **373**, 20160523.
7. Amblee,V. and Jeffery,C.J. (2015) Physical features of intracellular proteins that moonlight on the cell surface. *PLoS ONE*, **10**, e0130575.
8. Berkovits,B.D. and Mayr,C. (2015) Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature*, **522**, 363–367.
9. Mayr,C. (2016) Evolution and biological roles of alternative 3′UTRs. *Trends Cell Biol.*, **26**, 227–237.

10. Mayr,C. (2017) Regulation by 3′-untranslated regions. *Annu. Rev. Genet.*, **51**, 171–194.

11. Soto-Pantoja,D.R., Kaur,S. and Roberts,D.D. (2015) CD47 signaling pathways controlling cellular differentiation and responses to stress. *Crit. Rev. Biochem. Mol. Biol.*, **50**, 212–230.

12. Ma,W. and Mayr,C. (2018) A Membraneless organelle associated with the endoplasmic reticulum enables 3′UTR-mediated protein-protein interactions. *Cell*, **175**, 1492–1506.

13. Lee,S.-H. and Mayr,C. (2019) Gain of additional BIRC3 protein functions through 3′-UTR-mediated protein complex formation. *Mol. Cell*, **74**, 701–712.

14. Chapple,C.E., Robisson,B., Spinelli,L., Guien,C., Becker,E. and Brun,C. (2015) Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.*, **6**, 7412.

15. Perkins,J.R., Diboun,I., Dessailly,B.H., Lees,J.G. and Orengo,C. (2010) Transient protein-protein interactions: structural, functional, and network properties. *Structure*, **18**, 1233–1243.

16. Van Roey,K., Gibson,T.J. and Davey,N.E. (2012) Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.*, **22**, 378–385.

17. Mayr,C. (2019) What are 3′ UTRs doing? *Cold Spring Harb Perspect Biol.*, **11**, a034728.

18. Ribeiro,D.M., Briere,G., Bely,B., Spinelli,L. and Brun,C. (2019) MoonDB 2.0: an updated database of extreme multifunctional and moonlighting proteins. *Nucleic Acids Res.*, **47**, D398–D402.

19. Aranda,B., Blankenburg,H., Kerrien,S., Brinkman,F.S.L., Ceol,A., Chautard,E., Dana,J.M., De Las Rivas,J., Dumousseau,M., Galeota,E. *et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods*, **8**, 528–529.

20. Becker,E., Robisson,B., Chapple,C.E., Guénoche,A. and Brun,C. (2012) Multifunctional proteins revealed by overlapping clustering in protein interaction network. *Bioinformatics*, **28**, 84–90.

21. Chapple,C.E., Herrmann,C. and Brun,C. (2015) PrOnto database: GO term functional dissimilarity inferred from biological data. *Front. Genet.*, **6**, 200.

22. The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.*, **45**, D331–D338.

23. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.

24. Kinsella,R.J., Kähäri,A., Haider,S., Zamora,J., Proctor,G., Spudich,G., Almeida-King,J., Staines,D., Derwent,P., Kerhornou,A. *et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*, **2011**, bar030.

25. Müller,S., Rycak,L., Afonso-Grunz,F., Winter,P., Zawada,A.M., Damrath,E., Scheider,J., Schmäh,J., Koch,I., Kahl,G. *et al.* (2014) APADB: a database for alternative polyadenylation and microRNA regulation events. *Database (Oxford)*, **2014**, bau076.

26. Gruber,A.J., Schmidt,R., Gruber,A.R., Martin,G., Ghosh,S., Belmadani,M., Keller,W. and Zavolan,M. (2016) A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.*, **26**, 1145–1159.

27. Dassi,E., Re,A., Leo,S., Tebaldi,T., Pasini,L., Peroni,D. and Quattrone,A. (2014) AURA 2: Empowering discovery of post-transcriptional networks. *Translation (Austin)*, **2**, e27738.

28. Uhlén,M., Fagerberg,L., Hallström,B.M., Lindskog,C., Oksvold,P., Mardinoglu,A., Sivertsson,Å., Kampf,C., Sjöstedt,E., Asplund,A. *et al.* (2015) Tissue-based map of the human proteome. *Science*, **347**, 1260419.

29. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

30. Fu,L., Niu,B., Zhu,Z., Wu,S. and Li,W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.

31. Thul,P.J., Åkesson,L., Wiking,M., Mahdessian,D., Geladaki,A., Ait Blal,H., Alm,T., Asplund,A., Björk,L., Breckels,L.M. *et al.* (2017) A subcellular map of the human proteome. *Science*, **356**, eaal3321.

32. Reimand,J., Arak,T., Adler,P., Kolberg,L., Reisberg,S., Peterson,H. and Vilo,J. (2016) g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.*, **44**, W83–W89.

33. Lianoglou,S., Garg,V., Yang,J.L., Leslie,C.S. and Mayr,C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.*, **27**, 2380–2396.

34. Mendik,P., Dobronyi,L., Hári,F., Kerepesi,C., Maia-Moço,L., Buszlai,D., Csermely,P. and Veres,D.V. (2019) Translocatome: a novel resource for the analysis of protein translocation between cellular organelles. *Nucleic Acids Res.*, **47**, D495–D505.

35. Zanzoni,A., Spinelli,L., Ribeiro,D.M., Tartaglia,G.G. and Brun,C. (2019) Post-transcriptional regulatory patterns revealed by protein-RNA interactions. *Sci. Rep.*, **9**, 4302.

36. von Heijne,G. (2006) Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.*, **7**, 909–918.

37. Zimmermann,R., Eyrisch,S., Ahmad,M. and Helms,V. (2011) Protein translocation across the ER membrane. *Biochim. Biophys. Acta*, **1808**, 912–924.

38. Bernhofer,M., Goldberg,T., Wolf,S., Ahmed,M., Zaugg,J., Boden,M. and Rost,B. (2018) NLSdb-major update for database of nuclear localization signals and nuclear export signals. *Nucleic Acids Res.*, **46**, D503–D508.

39. Chen,C., Zabad,S., Liu,H., Wang,W. and Jeffery,C. (2018) MoonProt 2.0: an expansion and update of the moonlighting proteins database. *Nucleic Acids Res.*, **46**, D640–D644.

40. Díaz-Ramos,A., Roig-Borrellas,A., García-Melero,A. and López-Alemany,R. (2012) α-Enolase, a multifunctional protein: its role on pathophysiological situations. *J. Biomed. Biotechnol.*, **2012**, 156795.

41. Pettersson,O.J., Aagaard,L., Andrejeva,D., Thomsen,R., Jensen,T.G. and Damgaard,C.K. (2014) DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1. *Nucleic Acids Res.*, **42**, 7186–7200.

42. Maxwell,C.A., McCarthy,J. and Turley,E. (2008) Cell-surface and mitotic-spindle RHAMM: moonlighting or dual oncogenic functions? *J. Cell. Sci.*, **121**, 925–932.

43. Sharma,S., Bartholdson,S.J., Couch,A.C.M., Yusa,K. and Wright,G.J. (2018) Genome-scale identification of cellular pathways required for cell surface recognition. *Genome Res.*, **28**, 1372–1382.

44. Sriram,G., Martinez,J.A., McCabe,E.R.B., Liao,J.C. and Dipple,K.M. (2005) Single-gene disorders: what role could moonlighting enzymes play? *Am. J. Hum. Genet.*, **76**, 911–924.

45. Hentze,M.W., Castello,A., Schwarzl,T. and Preiss,T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, **19**, 327–341.

46. Gerstberger,S., Hafner,M. and Tuschl,T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 829–845.

47. Ribeiro,D.M., Zanzoni,A., Cipriano,A., Delli Ponti,R., Spinelli,L., Ballarino,M., Bozzoni,I., Tartaglia,G.G. and Brun,C. (2018) Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Res.*, **46**, 917–928.

48. Cioni,J.-M., Koppers,M. and Holt,C.E. (2018) Molecular control of local translation in axon development and maintenance. *Curr. Opin. Neurobiol.*, **51**, 86–94.

49. Schmitz,A.A., Govek,E.E., Böttner,B. and Van Aelst,L. (2000) Rho GTPases: signaling, migration, and invasion. *Exp. Cell Res.*, **261**, 1–12.

50. Sankhala,R.S., Lokareddy,R.K., Begum,S., Pumroy,R.A., Gillilan,R.E. and Cingolani,G. (2017) Three-dimensional context rather than NLS amino acid sequence determines importin α subtype specificity for RCC1. *Nat. Commun.*, **8**, 979.

51. Zanzoni,A., Ribeiro,D.M. and Brun,C. (2019) Understanding protein multifunctionality: from short linear motifs to cellular functions. *Cell. Mol. Life Sci.*, **76**, 4407–4412.

52. Shiber,A., Döring,K., Friedrich,U., Klann,K., Merker,D., Zedan,M., Tippmann,F., Kramer,G. and Bukau,B. (2018) Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*, **561**, 268–272.

53. Mayr,C. (2018) Protein complexes assemble as they are being made. *Nature*, **561**, 186–187.

54. Frechin,M., Enkler,L., Tetaud,E., Laporte,D., Senger,B., Blancard,C., Hammann,P., Bader,G., Clauder-Münster,S., Steinmetz,L.M. *et al.* (2014) Expression of nuclear and mitochondrial genes encoding ATP synthase is synchronized by disassembly of a multisynthetase complex. *Mol. Cell*, **56**, 763–776.