Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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

Structural characterisation of amyloidogenic intrinsically disordered zinc finger protein isoforms DPF3b and DPF3a

Julien Mignon [a,b,c,*], Denis Mottet [d], Tanguy Leyder [a], Vladimir N. Uversky [e], Eric A. Perpète [a,c,f], Catherine Michaux [a,b,c]

[a] *Laboratoire de Chimie Physique des Biomolécules, UCPTS, University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium*
[b] *Namur Institute of Structured Matter (NISM), University of Namur, Namur, Belgium*
[c] *Namur Research Institute for Life Sciences (NARILIS), University of Namur, Namur, Belgium*
[d] *University of Liège, GIGA-Molecular Biology of Diseases, Gene Expression and Cancer Laboratory, B34, Avenue de l'Hôpital, 4000 Liège, Belgium*
[e] *Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, United States*
[f] *Institute of Life, Earth and Environment (ILEE), University of Namur, Namur, Belgium*

ARTICLE INFO

ABSTRACT

Double PHD fingers 3 (DPF3) is a zinc finger protein, found in the BAF chromatin remodelling complex, and is involved in the regulation of gene expression. Two DPF3 isoforms have been identified, respectively named DPF3b and DPF3a. Very limited structural information is available for these isoforms, and their specific functionality still remains poorly studied. In a previous work, we have demonstrated the first evidence of DPF3a being a disordered protein sensitive to amyloid fibrillation. Intrinsically disordered proteins (IDPs) lack a defined tertiary structure, existing as a dynamic conformational ensemble, allowing them to act as hubs in protein-protein interaction networks. In the present study, we have more thoroughly characterised DPF3a *in vitro* behaviour, as well as unravelled and compared the structural properties of the DPF3b isoform, using an array of predictors and biophysical techniques. Predictions, spectroscopy, and dynamic light scattering have revealed a high content in disorder: prevalence of random coil, aromatic residues partially to fully exposed to the solvent, and large hydrodynamic diameters. DPF3a appears to be more disordered than DPF3b, and exhibits more expanded conformations. Furthermore, we have shown that they both time-dependently aggregate into amyloid fibrils, as revealed by typical circular dichroism, deep-blue autofluorescence, and amyloid-dye binding assay fingerprints. Although spectroscopic and microscopic analyses have unveiled that they share a similar aggregation pathway, DPF3a fibrillates at a faster rate, likely through reordering of its C-terminal domain.

1. Introduction

Regulation of gene expression in eukaryotic organisms is made possible by the modification of chromatin structure. Indeed, the (de)condensed state of chromatin permits the accessibility of transcription factors and transcriptional machinery to selected promoters [1]. The balance between condensation/decondensation of chromatin is achieved not only by epigenetic modifications of histones, but also by disruptors of DNA-histone interactions, such as chromatin remodelling complexes [2]. In that respect, the double plant homeodomain (PHD)

fingers 3 (DPF3) protein, also named BAF45c or CERD4, is an epigenetic factor strongly expressed in all types of post mitotic neurons and found in the multiprotein BRM/BRG1-associated factors (BAF) chromatin remodelling complex [3–5]. DPF3 exists in two isoforms, sharing 83% of amino acid sequence identity. The canonical isoform is referred to as DPF3b, and the second one as DPF3a [6]. They differentiate themselves by their sequence length and composition at the C-terminal region (Fig. 1A). The two isoforms have common domain structure, which, starting from the N-terminus, includes: i) the 2/3 domain, ii) a Krüppel-like zinc finger (ZnF), known as the C_2H_2 domain, and iii) a partial PHD ZnF (PHD-1) up to residue 292. While DPF3b is characterised by a complete PHD tandem (PHD-1 and -2), DPF3a PHD-1 finger is truncated into a shorter one (PHD-1/2), which is followed by the C-terminal domain. Unlike DPF3a, DPF3b is able to recognise, through its double PHD ZnFs, acetylated or methylated lysine residues on histone tails, allowing the recruitment of BAF complex, and subsequently the transcriptional machinery complex for gene expression [6,7]. Besides DPF3b PHD tandem, the function of the other domains remains largely unknown.

Little is known about the respective roles of DPF3b and DPF3 in the cell. Only few distinct functions have already been reported. As mentioned above, DPF3b acts as a histone reader, but DPF3a is unable to bind modified histones due to its uncomplete PHD tandem. Nonetheless, DPF3a is involved in myogenic differentiation by interacting with the hepatoma-derived growth factor-related protein 2 (HRP2). Formation of DPF3a-HRP2 complex facilitates the association of HRP2 with BAF, leading to the transcription of myogenic gene [8]. Cardiac gene

transcription has also found to be regulated by DPF3a [9], especially through phosphorylation of its C-terminal domain, allowing interaction with the transcriptional repressors HEY (HES-related repressor protein). BAF is thereafter recruited by the DPF3a-HEY complex, triggering the transcription of fetal genes, which promote cardiac hypertrophy [10].

From a pathological point of view, DPF3a is indeed associated with human heart hypertrophies, such as the Tetralogy of Fallot (TOF), a congenital heart defect. In patients with TOF, DPF3a is found remarkably upregulated in the right ventricular myocardium [11,12]. Furthermore, DPF3a acts in maintaining the stemness of glioma initiating cells, which participate in glioblastoma propagation and drug resistance [13]. It has been recently reported that the two DPF3 isoforms act as oncogenic agents, contributing to breast cancer motility, as well as higher risks at developing renal carcinoma cancer [14–16]. In addition, clinical studies have found that single nucleotide polymorphism (SNP) in DPF3 is highly correlated with low sperm count associated with motility defects and male infertility [17–19].

Very limited structural information is available on DPF3 isoforms, being limited to X-ray and NMR-resolved structures of DPF3b double PHD finger complexed with modified histone peptides [6,7,20]. In agreement with predictions suggesting that most BAF complex subunits, including DPF3, are intrinsically disordered proteins (IDPs) [21], we have unveiled that DPF3a is highly disordered, and, indeed, exhibits typical spectroscopic footprints of an extended IDP [22]. IDPs are characterised by a dynamic ensemble of conformations, rather than a fixed tertiary structure, enabling them to interact with numerous partners [23]. Thanks to its ability to form diverse fuzzy complexes, a single
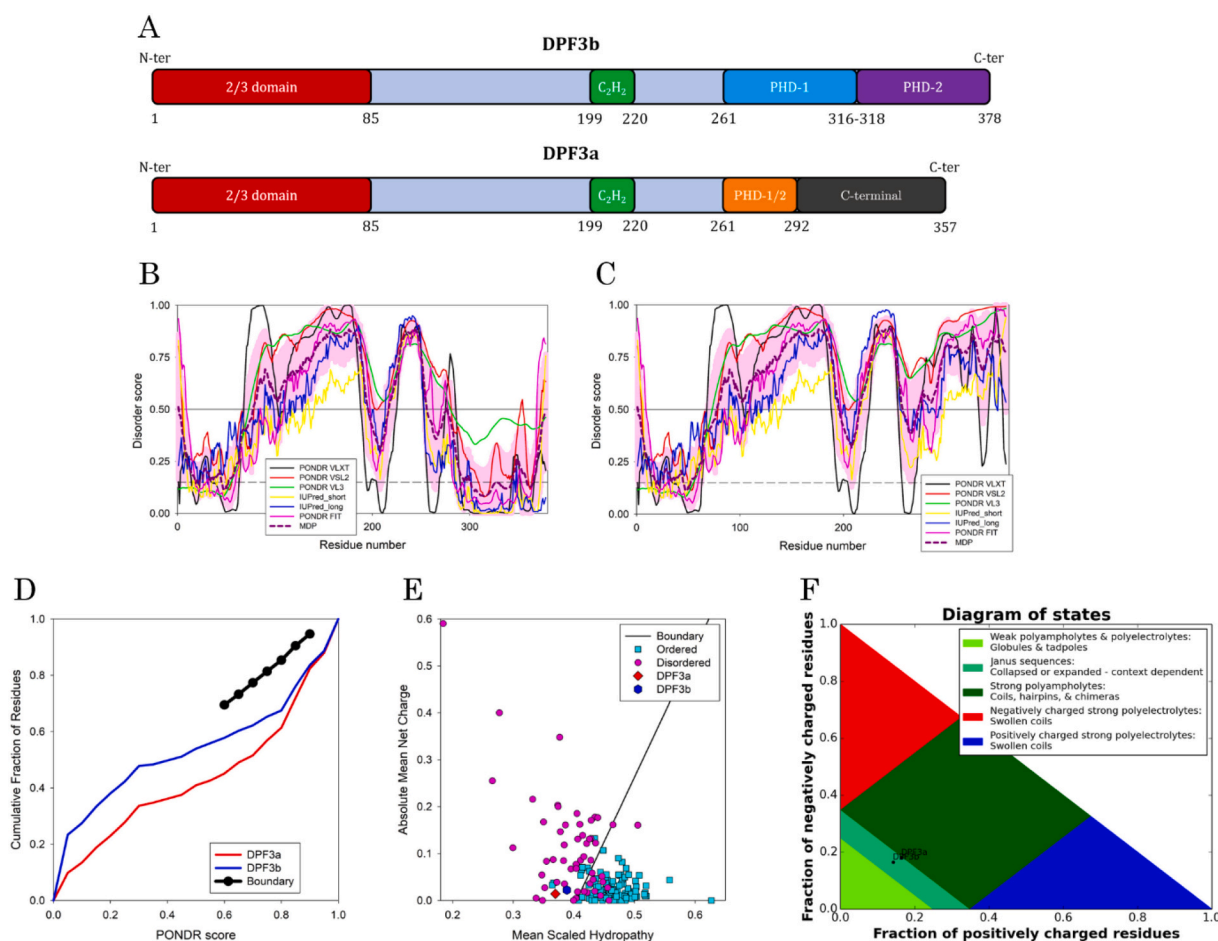
Fig. 1. Predicted disorder-associated properties of DPF3 isoforms. (A) Sequence and domains organisation of DPF3b and DPF3a. PONDR and IUPred disorder profile of (B) DPF3b and (C) DPF3a; MDP corresponds to the mean disorder profile. Every residue displaying a score above the 0.5 threshold is considered disordered. (D) CDF, (E) CH plots, and (F) Das-Pappu phase diagram of DPF3b and DPF3a.

IDP can be engaged in multiple cell functions, but is therefore more sensitive to misregulation, which might lead to severe diseases, such as cancer, cardiovascular and neurodegenerative disorders [24,25]. In this context, we recently demonstrated that DPF3a isoform is prone to aggregate into high-order oligomers and fibrous amyloid-like structures, suggesting that it may be involved in diseases called systemic amyloidosis [22].

Understanding the functional specificities and interaction networks of DPF3 isoforms requires the elucidation of their respective structural properties. Furthermore, their identification as new prone-to-aggregation IDPs can make them new interesting targets in anti-amyloid and anti-aggregation drug design. In this study, we report an in-depth comparative investigation of DPF3b and DPF3a intrinsically disordered nature and fibrillation pathways *in vitro*, by a combination of prediction methods, absorption and emission spectroscopy, light scattering technique, and electron microscopy.

2. Materials and methods

2.1. Overexpression and purification of DPF3 isoforms

DPF3 recombinant proteins were expressed with a GST tag at their N-terminus using a pET like vector in *E. coli* BL21 Rosetta (DE3) cells. Transformed bacterial strains were precultured in 20 g/L lysogeny Lennox broth (LB) with 0.36 mM ampicillin at 37 °C for 16 h. From 5.0 mL of preculture, strains were cultured in 20 g/L LB with 0.14 mM ampicillin at 37 °C until the optical density at 600 nm reached 0.5–0.8. Cultures were then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37 °C for 4 h, and centrifuged. Once the supernatants discarded, pellets were stored at −20 °C. For the lysis step, pellets were suspended in the lysis buffer (phosphate buffered saline (PBS) pH 7.3, 0.5% Triton X-100, 200 mM KCl, 200 μM phenylmethylsulfonyl fluoride), and sonicated in an ice-water bath. Supernatants were collected after centrifugation, and purified on an Äkta Purifier fast protein liquid chromatography. With the binding buffer (PBS pH 7.3, 200 mM KCl), DPF3-GST fusion proteins were bound on a 5 mL GSTrap FF pre-packed column (GE Healthcare). DPF3 proteins were cleaved on column at 30 °C for 2 h with 20 μL of TEV protease solution (Sigma) in Tris buffered saline (TBS) buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl). Cleaved proteins were gathered in TBS, and their purity was examined by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The presence of each DPF3 isoform was further confirmed by mass spectrometry analysis.

2.2. Protein concentration determination

Both DPF3 isoforms were concentrated in TBS through a 6–8 kDa cut-off dialysis membrane wrapped in PEG-20000, used as the water-absorbent. Final concentration was determined by measuring the absorbance at 214 nm with a UV-63000PC spectrophotometer (VWR), using a 10 mm pathlength quartz QS cell (Hellma). Knowing the molar extinction coefficient of each isoform at 214 nm ($\varepsilon_{DPF3b} = 763{,}132\ M^{-1}.cm^{-1}$, $\varepsilon_{DPF3a} = 622{,}915\ M^{-1}.cm^{-1}$), calculated from B. Kuipers and H. Gruppen's data [26], DPF3b and DPF3a working concentration amounted to ~0.2 mg/mL (~5 μM) in each experiment.

2.3. Structural properties prediction

All properties were predicted from the known sequence of full-length DPF3b (Uniprot ID: Q92784-1), and DPF3a (Uniprot ID: Q92784-2). The average percentage of per-residue predicted intrinsic disorder (PPID) was calculated on the basis of 19 online available disorder predictors: VL-XT [27], XL1-XT [28], CaN-XT [29], VL3 [28], and VSL2 [28] algorithms from Predictors of Natural Disordered Regions (PONDR); ESpritz NMR, X-ray, and Disprot-trained datasets [30]; Protein DisOrder prediction System (PrDOS) [31]; Prediction of Intrinsically Unstructured

Proteins for long and short disordered regions (IUPred3) [32]; meta-predict [33]; NORSnet, Ucon, and MetaDisorder MD algorithms from the PredictProtein server [34]; AUCpreD from RaptorX; NetSurfP [35]; Prediction of Order and Disorder by evaluation of NMR data with and without evolution (ODiNPred) [36]. Disorder plots along query sequences were generated with the PONDR and IUPred3 servers. Protein disorder relies on the nature of its constituting amino acids, which can be placed on a scale ranging from 0 (fully ordering residues) to 1 (fully disordering residues). Residues with a PPID above the threshold value of 0.5 participate in disorder.

Disorder propensity was also investigated by the cumulative distribution function (CDF) and charge-hydropathy (CH) plots, generated with the PONDR server [37]. Predicted conformational ensembles were examined on a Das-Pappu phase diagram, and its associated parameters, by using the Classification of Intrinsically Disordered Ensemble Regions (CIDER) [38]. Occurrence and localisation of low complexity domains (LCDs) and molecular recognition features (MoRFs) in DPF3 sequences were determined with the Simple modular architecture research tool (SMART) [39] and the PLAtform of Tools for Low Complexity (PlaTo-LoCo) [40] for LCDs, and with the MoRFchibi tool [41] for MoRFs.

Buried state of tryptophan and tyrosine residues was predicted for each isoform from their primary structure using the prediction of protein relative solvent accessibility web server (PaleAle 5.0) [42]. Residues are classified into four solvent accessibility classes: very buried (under 4% exposed), somewhat buried (between 4 and 25% exposed), somewhat exposed (between 25 and 50% exposed), and very exposed (over 50% exposed).

Aggregation and, more specifically, fibrillation hotspots were identified in each isoform by using CamSol [43] and AMYLPRED2 [44] web servers, respectively. CamSol provides an intrinsic solubility profile, where regions displaying scores above 1 are considered highly soluble, whereas those with scores below −1 are poorly soluble and aggregation-promoting. AMYLPRED2 allows the identification of amyloid-promoting regions in the query sequence on the basis of 11 independent fibrillation predictors. Hits are defined as a consensus of 5 out of 11 methods.

2.4. Far-UV circular dichroism spectroscopy (far-UV CD)

Far-UV CD spectra (190–260 nm) were recorded with a MOS-500 spectropolarimeter at 20 °C in TBS, using a 1 mm pathlength quartz Suprasil cell (Hellma). Four scans (15 nm/min, 2 nm bandwidth, 0.5 nm data pitch, and 2 s digital integration time) were averaged, baselines were subtracted, and corrected spectra were smoothed. Data are presented as the mean residue ellipticity ($[\Theta]_{MRE}$), calculated as follows: $[\Theta]_{MRE} = (M.\theta)/(n-1).(10\gamma.l)$, where M is the molecular mass (Da), θ the ellipticity (deg), n the sequence length, γ the protein concentration (mg/mL), and l is the cell pathlength (cm). Sample concentration for each isoform was ~0.2 mg/mL. In order to estimate the secondary structure content from each spectrum, the Beta Structure Selection online tool (BeStSel) [45] was used in the 200–250 nm range.

2.5. Fluorescence spectroscopy

Fluorescence procedures included intrinsic tryptophan fluorescence (ITF), intrinsic tyrosine fluorescence (ITyrF), deep-blue auto-fluorescence (dbAF), and Thioflavin T (ThT) binding assay. Every emission spectrum was recorded from the excitation wavelength up to 600 nm by 1.0 nm increment with an Agilent Cary Eclipse fluorescence spectrophotometer at ~20 °C in TBS, using a 10 mm pathlength quartz QS cell (Hellma), and an emission and excitation slit width (sw) of 10 nm. The following excitation wavelengths (λ_{ex}) were used: 295 nm for ITF, 275 nm for ITyrF, 400 nm for dbAF, and 440 nm for ThT. Regarding the extrinsic ThT fluorescence, 75 μL of protein sample were mixed with 75 μL of 20 μM ThT in TBS (final ThT concentration of 10 μM), beforehand filtered on polyether sulfone (PES) 0.2 μm. In addition, dbAF signatures were further investigated by scanning excitation

contributions from 200 nm to the emission wavelength ($\lambda_{em} = 456$ nm), using the same previously described acquisition parameters.

2.6. Dynamic light scattering (DLS)

DLS measurements were carried out at 20 °C with a Horiba Zetasizer SZ-100 nanoparticles analyser with the detector at 90°. The autocorrelation function was successfully fitted 20 times per analysis, and the results were expressed as the mean hydrodynamical diameter $\overline{D}_h$ (nm). Each protein sample was passed through a 0.2 μm PES filter before analysis.

2.7. Congo red (CR) binding assay

CR Absorption spectra were recorded in the range of 400–600 nm by 1.0 nm increment with a UV-63000PC spectrophotometer (VWR) at ~20 °C in TBS, using a 10 mm pathlength quartz QS cell (Hellma). Samples were prepared by mixing 75 μL of protein material with 75 μL of 20 μM CR in TBS (final CR concentration of 10 μM).

2.8. Transmission electron microscopy (TEM)

Protein aggregates were visualised by negative staining (NS) with a PHILIPS/FEI Tecnai 10 electron microscope operating at 100 kV. Formvar/carbon-coated copper grids were beforehand hydrophilised by glow discharge. A 5 μL droplet of a protein sample incubated for 7 days (168 h) at ~20 °C was left for 3 min onto the grid, and the excess was soaked with a blotting paper. The grid was put on a 5 μL droplet of 0.5% (w/v) uranyl acetate for 1 min, washed with Milli-Q water, and air-dried for 5 min before analysis.

3. Results and discussion

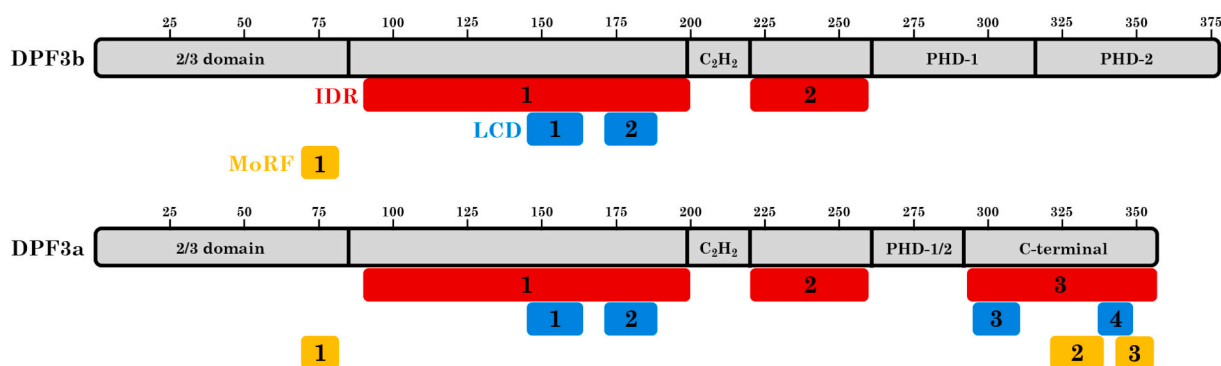### 3.1. Prediction of disorder-associated properties of human DPF3 isoforms

In order to determine the numerous roles assigned to DPF3 isoforms in the cell, acquiring information about their specific structural features might help to apprehend and discriminate their IDP-related multifunctionality. To this end, several structural predictors have been applied. Sequence composition differences between DPF3b and DPF3a are reflected in their disorder-associated properties (Fig. 1). Regarding the predicted per-residue percentage of intrinsic disorder (PPID), the resulting average of 19 disorder algorithms gives a PPID of 44% and 63% for DPF3b and DPF3a, respectively. Such scores indicate that both isoforms are highly disordered proteins, with a much higher disorder content being found in DPF3a, which is expected from the truncation of its PHD finger. Along the sequence, several features are common to the two isoforms, including characteristic intrinsically disordered regions (IDRs) found in both of them (Figs. 1B and C). From the N-terminus, the

disorder profile includes the 2/3 domain (and beyond), which is considered mostly ordered (residues 1 to 90), although disorder scores varying between 0.2 and 0.5 indicate some chain flexibility, especially from residues 60 to 90. This is followed by a 110-residue-long IDR (IDR-1) up to the C_2H_2 zinc finger (ZnF), where the disorder score drops below the 0.5 threshold (residues 199 to 220), though retaining some flexibility. Unsurprisingly, such decrease in disorder propensity is due to the C_2H_2 ZnF of the DPF family, a domain structured by coordination of Zn^{2+} [46]. A second shorter 40-residue-long IDR (IDR-2) is found between the C_2H_2 and PHD1 ZnFs (residues 221–260), followed by a gain in order at the level of the PHD1 finger (residues 261–292). Towards the C-terminus, where isoforms sequence composition differs, DPF3b remains fully ordered due to the PHD2 ZnF (residues 293–378, Fig. 1B), whilst DPF3a shows a third 60-residue-long IDR (IDR-3) with high disorder scores (residues 293–357, Fig. 2). The presence of this IDR-3 is consistent with the overall higher PPID score of DPF3a. Albeit DPF3b appears more ordered, it still meets the criteria for classification as an IDP, as it is > 30% disordered, and possesses at least one IDR longer than 30 residues [47].

The same DPF3b-DPF3a discrepancy can also be noticed by examining other disorder-related features. IDPs are known to be enriched in disorder-promoting residues (DPRs), compared to globular proteins, for which the proportion in DPRs and order-promoting residues (OPRs) is more balanced. The DPR group includes short non-polar (glycine and alanine), charged (arginine, lysine and glutamate), and polar (glutamine and serine) amino acids, as well as proline thanks to its secondary structure breaking properties [48]. According to its higher PPID, DPF3a is richer in DPRs (59%) than DPF3b (54%), and both isoforms are impoverished in OPRs (32 and 26% for DPF3b and DPF3a, respectively), as it is expected for an IDP.

In order to further investigate to which extent each DPF3 isoform is disordered, the cumulative distribution function (CDF), as well as at the charge-hydropathy plot (CH) have been generated. Considering the CDF plot, if the curve is located below the boundary, the protein is expected to be mainly disordered, which is the case for both DPF3 isoforms [49]. DPF3a curve being under the one of DPF3b suggests that DPF3a has a higher overall disorder propensity, which is in agreement with the corresponding PPID scores (Fig. 1D). The CH plot is split into two areas delimited by a boundary line of equation $<R> = 2.785<H> - 1.151$, where $<R>$ is the absolute mean net charge, and $<H>$ is the mean scaled hydropathy of a query protein [50]; with $<H>$ parameter being calculated according to the hydropathy scale defined by J. Kyte and R. F. Doolittle [51]. Proteins located below the boundary line in the CH plot are considered ordered or compact (ORDP area), whereas those located above the boundary are expected to be mostly disordered (IDP area). DPF3b and DPF3a isoforms are located in the IDP area, which, once again, argues towards their disordered behaviour (Fig. 1E).

Although DPF3 isoforms are displaying a low mean net charge, this is insufficient for favourable folding into compact globular conformers, as

Fig. 2. Distribution and location of predicted IDRs (red), LCDs (blue), and MoRFs (yellow) along DPF3b and DPF3a sequence. Each region is numbered in order of appearance from the N-terminus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

it follows from their conformational preferences. Extraction of κ and Ω patterning parameters from CIDER analysis allows a better depiction of DPF3 conformational state. The κ value reflects to which extent charged residues are distributed within a given sequence, with the κ value going from 0 to 1 as positively and negatively charged residues become segregated. High κ proteins tend to collapse due to the electrostatic attraction, while low κ ones assume more expanded conformations [52]. DPF3b and DPF3a have relatively similar κ values of 0.26 and 0.27, respectively, suggesting that expanded conformers might be favoured. Furthermore, the Ω parameter reports on the patterning of proline (Pro) and charged residues with respect to all other amino acids on a scale from 0 to 1. As Pro and charged residues are expansion-promoting, a uniformly mixed sequence yields low Ω value, resulting in extended protein conformers [53]. Low DPF3b and DPF3a Ω values of 0.17 and 0.15, respectively, are also indicative of more favourable expanded conformations.

Furthermore, the conformational ensemble preferentially adopted by DPF3 isoforms can be refined by generating a Das-Pappu phase diagram, which is based on the distribution of the fraction of positively against negatively charged residues [52]. The output shows that both isoforms belong to the R2 category, which is a transient region between weak (R1) and strong (R3) polyampholytes (Fig. 1F). Most IDPs (~40%) are found within the R2 region, comprising context-dependent coiled structures, called Janus sequences, either collapsed (similar to a molten globule state) by hydrophobic effect or expanded (similar to a random coil state) by a favourable chain solvation. R2 region is also populated by IDPs that fold upon binding. DPF3a slightly differs from DPF3b by tending towards the R3 region (~30% of IDPs), grouping stronger polyampholytes, presenting either coiled or hairpin conformations [54].

Due to their floppy structures with high solvent accessibility, low complexity domains (LCDs) are commonly found within IDPs. They are characterised by a reduced amino acid alphabet, generally occur in IDRs, are thought to be involved in IDP functionality, and are often related to increased aggregation propensity [55,56]. In order to pinpoint LCDs in DPF3 isoforms, SMART and PlaToLoCo servers were used (Fig. 2). The two predictors identify two LCDs in IDR-1: the first LCD is rich in glutamate (residues 145–164, LCD-1), and the second LCD is rich in arginine and glycine (residues 171–189, LCD-2). DPF3a is characterised by two additional LCDs overlapping its IDR-3: the third LCD is rich in alanine (residues 295–311, LCD-3), and the fourth one is enriched in arginine and glycine (residues 337–348, LCD-4). Interestingly, all LCD-comprised residues belong to the DPR class. Existence of the LCD-3 and LCD-4 is likely related to DPF3a distinctive functions, and may be involved in its proneness to fibrillate. Nevertheless, LCD-1 and LCD-2 might also be aggregation-promoting regions.

Molecular recognition features (MoRFs) represent short disordered segments that can fold upon binding to specific interaction partners. They are needed to mediate protein-protein interactions (PPIs) in transient complexes involving IDPs [57]. MoRFs were predicted with the MoRFchibi online tool (Fig. 2). The first MoRF (residues 69–82, MoRF-1) is located at the end of the 2/3 domain, considered partially disordered or at least flexible. As for LCDs, two other MoRFs are found in IDR-3 of DPF3a. The second MoRF stands between LCD-3 and LCD-4 (residues 321–339, MoRF-2), and the third MoRF overlaps with LCD-4 (residues 343–356, MoRF-3). These additional MoRFs may be specific to the DPF3a interactions network, comprising partners that cannot bind DPF3b.

Taken together, all these predictive data clearly indicate that DPF3 isoforms belong to the class of IDPs, and show that DPF3a is expected to be more disordered than DPF3b. The localisation and specificity of LCD-3 and LCD-4, as well as the presence of the MoRF-2 and MoRF-3 in DPF3a isoform further support the disorder character of its C-terminal domain, as well as the predominant role of this domain in DPF3a separate functionality.

3.2. Structural evidence of DPF3 isoforms disorder

In our previous study, DPF3a disorder has been partially characterised by far-UV circular dichroism (CD) and tryptophan fluorescence spectroscopy, revealing several IDP-specific fingerprints [22]. Here, structural properties and *in vitro* behaviour of DPF3a have been more thoroughly investigated and compared to the DPF3b isoform using light scattering, CD, emission, and absorption spectroscopy. Dynamic light scattering (DLS) measurements indicate that the DPF3 isoforms exhibit similar sizes, with a $\overline{D}_h$ of 24 ± 7 nm for DPF3b (Fig. S1A), and of 20 ± 4 nm for DPF3a (Fig. S1B). Although the hydrodynamic dimensions of the two isoforms seem huge for a protein of their sequence length (378 and 357 residues for DPF3b and DPF3a, respectively), they fall in the range expected for an unfolded protein, as anticipated for an IDP (from ~5 to ~22 nm) [58]. Due to their more extended conformational ensembles, IDPs have noticeably higher $\overline{D}_h$ than globular proteins of similar molecular mass (from ~3 to ~10 nm). Furthermore, the hydrodynamic diameter of IDPs has been correlated with their sequence length, revealing that disordered proteins in the range of DPF3 isoforms (300 to 400 residues long) can have a diameter ranging from 16 to 20 nm [59]. This is consistent with the measured $\overline{D}_h$ values, implying that the two DPF3 isoforms adopt relatively expanded conformations.

DPF3b CD spectrum shows a profile typical of a hybrid IDP (Fig. 3). Indeed, a predominant strong negative band towards 206 nm is indicative of disorder, while the positive band at 198 nm arises from ordered secondary structures. Indeed, a fully disordered protein is expected to have no positive signal in the 190–200 nm range. A broad negative shoulder from 220 to 225 nm is likely related to either α-helix, for which the second band at 208 nm is overlapped by the signal at 206 nm, or antiparallel β-sheets. DPF3a CD spectrum reveals similar features: a distinctive negative band at 206 nm, a positive one at 200 nm, and a broad shoulder centred at ~225 nm (Fig. 3). These results indicate that DPF3 isoforms have close CD footprints classifying them as highly disordered proteins.

Estimation of their secondary structure content with the BeStSel tool is in good agreement with the profile of CD spectra. At 0 h, outputs report very low percentages of α-helix and some β-sheets (Table 1). Disorder (addition of coil and turn) prevails in the two isoforms, and, consistent with predictions, DPF3a is more disordered than DPF3b. Furthermore, BeStSel identifies β-sheets as exclusively antiparallel in both cases. Coexistence of antiparallel β-sheets with small fractions of α-helix may explain the broad shoulder obtained between 220 and 225 nm. Higher β-sheet content can be due to the several ZnFs (C_2H_2 and PHD) present in the two isoforms.

Study of the intrinsic fluorescence properties of tryptophan (Trp) and tyrosine (Tyr) residues also allows to learn local conformational
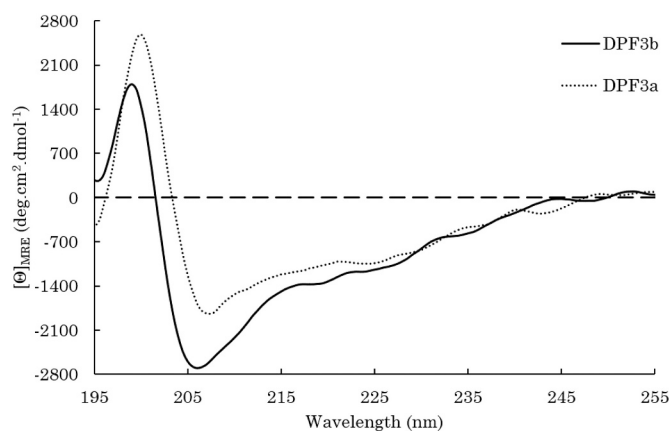

Fig. 3. Far-UV CD spectra of full-length DPF3b (solide line) and DPF3a (dotted line) in TBS at ~20 °C.

Table 1
Secondary structure content and fit-associated RMSD values of DPF3b and DPF3a in TBS at ~20 °C at 0 h, 24 h, 48 h, 72 h (for DPF3a) and 96 h (for DPF3b) incubation. Each estimated secondary structure percentage is extracted from BeStSel fit spectrum deconvolution.

| Protein (incubation time)	RMSD	Secondary structure content (%)				
		α-helix	β-sheet (antiparallel)	β-sheet (parallel)	Turn	Coil
DPF3b (0 h)	0.03	7	24	0	16	53
DPF3b (24 h)	0.03	2	29	0	16	53
DPF3b (48 h)	0.02	2	32	0	16	50
DPF3b (96 h)	0.03	2	36	0	16	46
DPF3a (0 h)	0.03	3	25	0	18	54
DPF3a (24 h)	0.02	6	31	0	16	47
DPF3a (48 h)	0.02	3	35	0	15	47
DPF3a (72 h)	0.03	0	39	0	16	45

information about each DPF3 isoform (Fig. 4). Tryptophanyl side chains display solvatochromism after excitation at 295 nm, resulting in characteristic emission wavelength (λ_{em}) shifts, ranging from 308 to 355 nm. Trp residues found in a highly non-polar or polar environment emit at shorter or longer λ_{em}, respectively [60,61]. DPF3b contains 5 Trp residues, and its intrinsic tryptophan fluorescence (ITF) spectrum shows an emission band at 335 nm, depicting partially exposed Trp residues to the solvent and/or polar amino acids (Fig. 4A). Such λ_{em} value is commonly found amongst collapsed IDPs, for which Trp residues are neither completely buried, nor fully exposed to the solvent. On the contrary, DPF3a only contains 2 Trp residues. The observed λ_{em} is red-shifted towards 340 nm and comes with a band broadening at higher wavelengths (Fig. 4A). This is highlighted by the difference spectrum between the two isoforms, where a contribution between 340 and 400 nm is noticeable. In that respect, DPF3a Trp residues appear more exposed to a polar environment, though neither fully exposed, which is not surprising regarding the localisation of these amino acids. Indeed, the two first Trp (Trp[56] and Trp[79]), common to both isoforms, are found close to the end of the 2/3 domain; a region predicted with some chain flexibility according to its disorder scores varying from 0.2 to 0.5. No Trp residue is found in predicted IDRs, which may explain why DPF3a emits closer to 340 than 355 nm. Considering DPF3b, the 3 additional Trp (Trp[311], Trp[358] and Trp[364]) are located in the double PHD ZnF, a well-folded domain, which is consistent with the observed blue-shift towards 335 nm.

Complementarily with ITF, intrinsic tyrosine fluorescence (ITyrF) is an efficient, though underused, tool to probe protein (un)folding and conformational changes, especially for IDPs. Albeit far less sensitive to polarity than tryptophanyl, tyrosyl moieties can exhibit, after excitation at 275 nm, a slight red-shift of their λ_{em} (from 303 to 305 nm), and an extinction of their quantum yield when displaced from a non-polar environment to a polar one [62]. In the case of folded proteins,

containing both Trp and Tyr residues, ITyrF signature is often undetected due to energy transfer and spectral overlap between tryptophanyl and tyrosyl side chains. Nonetheless, Trp-Tyr fluorescence resonance energy transfer (FRET), without introducing exogenous fluorophores, has recently proven to be highly reliable to describe unfolded and disordered systems [63]. Examination of DPF3b ITyrF spectrum reveals a dominant band at 336 nm, and a subtle shoulder at ~305 nm (Fig. 4B). While the latter relates to a small fraction of exposed Tyr residues, the emission at 336 nm can arise from tyrosinate formation. However, it has been reported that in commonly used biochemical buffers within the range of physiological pH, such as the TBS, tyrosinate fluorescence does not occur. Indeed, tyrosine deprotonation requires high acetate concentration or very alkaline medium (pH > 10) [64]. Realistically, this emission would rather come from Trp-Tyr FRET either due to amino acids promiscuity or spatial proximity through folding, meaning that most DPF3b Tyr residues are in the vicinity of Trp. Nonetheless, the coexisting weak fluorescence at 305 nm is indicative of disorder.

With respect to DPF3a ITyrF spectrum, the band at ~305 nm becomes much more intense, as evidenced by the difference spectrum with a remarkable contribution at 300–305 nm (Fig. 4B). DPF3a has therefore a larger fraction of exposed Tyr residues compared to DPF3b. The second emission band remains at 336 nm with a significant broadening to higher wavelengths between 340 and 400 nm. As it has been stated that Trp residues mainly participate in the emission at 336 nm, the observed broadening is consistent with the DPF3a ITF spectrum. Similar to DPF3b, some DPF3a Tyr residues are close enough to Trp ones to allow FRET, and Tyr emission being visible relates to a highly disordered character.

Furthermore, ITF and ITyrF results are supported by the predicted solvent accessibility of Trp and Tyr residues (Table 2, Fig. 5). The output from the PaleAle server (see Section 2.3) describes that the two Trp residues of DPF3a are either fully exposed or partially buried. Therefore,
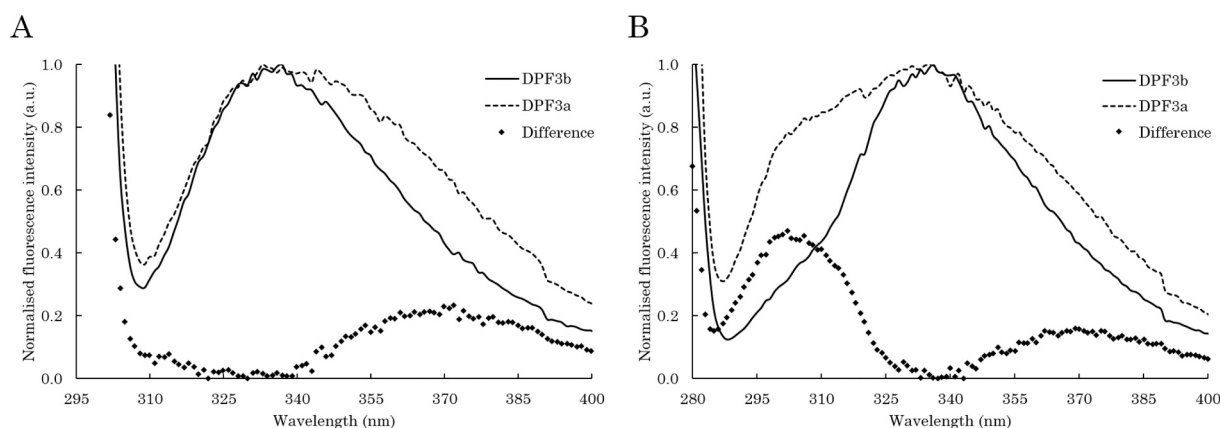
Fig. 4. Intrinsic fluorescence of full-length DPF3b and DPF3a in TBS at ~20 °C. (A) Normalised ITF spectra (λ_{ex} = 295 nm, sw = 10 nm) of DPF3b (solide line) and DPF3a (dashed line), and the associated difference spectrum (diamonds). (B) Normalised ITyrF (λ_{ex} = 275 nm, sw = 10 nm) of DPF3b (solide line) and DPF3a (dashed line), and the associated difference spectrum (diamonds).

Table 2
Solvent exposition class distribution of DPF3b and DPF3a Trp (W) and Tyr (Y) residues. Each per-residue solvent accessibility has been predicted using the PaleAle server.

	Residue position	
Solvent exposition class	DPF3b	DPF3a
Very buried ($< 4\%$)	Y[17], Y[341]	Y[17]
Somewhat buried (4–25%)	Y[27], Y[72], Y[74], W[79], Y[198], Y[207], Y[217], Y[261], W[311], Y[344], W[358], W[364]	Y[27], Y[72], Y[74], W[79], Y[207]
Somewhat exposed (25–50%)	Y[309]	Y[261]
Very exposed ($> 50\%$)	Y[54], W[56], Y[215]	Y[54], W[56], Y[198], Y[215], Y[217]

the sum of their contributions leads to a relative exposed state, while the consequent solvent accessibility of Trp[56] is responsible for the band broadening observed at higher wavelengths. Regarding DPF3b, its three additional Trp residues are all predicted partially buried, which significantly increases the proportion of less accessible Trp residues, thus causing the λ_{em} to be blue-shifted.

As for Trp residues, the per-Tyr solvent accessibility is also very informative with respect to ITyrF spectra. Essentially, DPF3b has ~77% and ~23% of buried and exposed Tyr residues, respectively. In the same way, DPF3a has a 50–50% ratio, suggesting why the contribution at 305 nm is more significant in DPF3a ITyrF spectrum. In more detail, Tyr[17] and Tyr[27] are considered buried, and thus should be involved in the 2/3 domain folding. Trp-Tyr FRET would take place for Tyr[54], Tyr[74], and Tyr[79] regarding their proximity to Trp[56] and Trp[79]. Although both isoforms share the 10 first Tyr residues in common, some discrepancy is observed (Fig. 5). Indeed, while most Tyr residues in the C_2H_2 domain are considered buried for DPF3b, those residues are predicted exposed for DPF3a. In the case of DPF3b, this could be indicative of local folding, whereas it still advocates for the higher emission intensity at 305 nm observed for DPF3a. Similarly, Trp-Tyr FRET conceivably occurs in the PHD tandem of DPF3b, arguing towards the reduction of Tyr fluorescence in DPF3b ITyrF spectrum. Also, as observed for Trp, no Tyr residue is located in IDRs. ITF and ITyrF results suggest that DPF3a has a more expanded conformational ensemble (random coil), while DPF3b has a more collapsed one (molten globule), especially towards the C-terminus. The same tendency is pinpointed by the Das-Pappu phase diagram (Fig. 1F).

UV absorption properties have recently been used to conformationally probe protein (un)folding. It relies on a foldedness index (FI), calculated by summing two absorbance ratios at 280, 275, and 258 nm: FI = $(A^{280}/A^{275}) + (A^{280}/A^{258})$. Most natively folded proteins have a FI > 2.5. The FI decreases upon denaturation through λ_{abs} shift from 280 to 258 nm, thus diminishing the contribution at 280 nm. As an example,

the FI value of pepsin decreases from 2.6 to 1.9 upon unfolding [65]. Both DPF3 isoforms display peculiar absorption properties, exhibiting a maximum absorption at ~258–260 nm (Fig. S2). This results in a low FI of ~1.7 and ~1.4 for DPF3b and DPF3a, respectively, meaning that they are highly disordered, as it is expected for IDPs. The lower FI of DPF3a suggests a higher content in disorder, which is consistent with previous observations.

3.3. Spontaneous aggregation propensity

We have recently established that DPF3a is prone to aggregation and is able to form amyloid-like fibrils [22]. In this study, understanding its fibrillation pathway has been further deepened and rationalised by combination of spectroscopic techniques, as well as compared to DPF3b aggregation propensity. DPF3 isoforms have been kept in the same buffer (TBS), temperature (~20 °C) and concentration (~5 μM) conditions over the course of 3–4 days in order to assess their ability to spontaneously aggregate *in vitro*. Far-UV CD spectroscopy was first used to follow over time changes in secondary structure content (Fig. 6). After 24 h, structural modifications can already be observed for DPF3b (Fig. 6A, Table 1). While the α-helix content significantly decreases, an enrichment in β-sheets is visible by the loss of signal at 206 nm. Turn and coil proportions remain unchanged. Additional conformational rearrangements proceed in the following 24 h, with a decrease of the random coil contribution, and further increase in β-sheets. Structural transformations are still noticed after 96 h incubation. Random coil contribution is further decreased, and the minimum takes shape at ~226 nm with a positive band still at 200 nm. Such a CD fingerprint is indicative of a protein predominantly folded into antiparallel and twisted β-sheets, which has already been reported for DPF3a [22]. Enrichment in β-sheet-type secondary structure, typically found amongst amyloidogenic proteins during their fibrillation process, highlights the amyloidogenic nature of DPF3b. In addition, versatile changes in its secondary structure
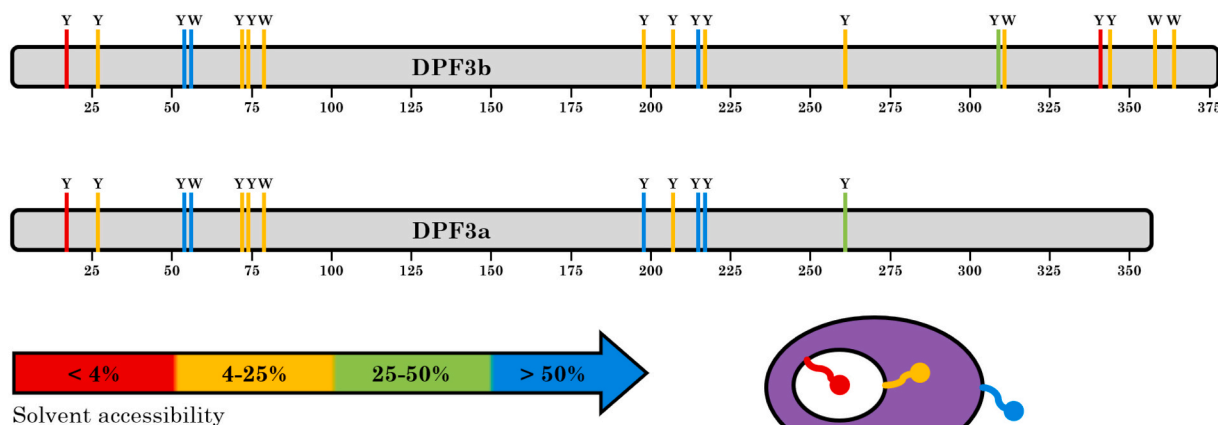
Fig. 5. Distribution and location of Trp (W) and Tyr (Y) residues of DPF3b and DPF3a with respect to their predicted solvent accessibility, ranging from very buried (red), somewhat buried (yellow), somewhat exposed (green), to very exposed (blue). Solvent exposition states of aromatic residues are schematically represented in a protein (purple) in the bottom right: aromatic residues in a hydrophobic pocket (red), buried near polar residues (yellow), one face (green) and two faces exposed to the solvent (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
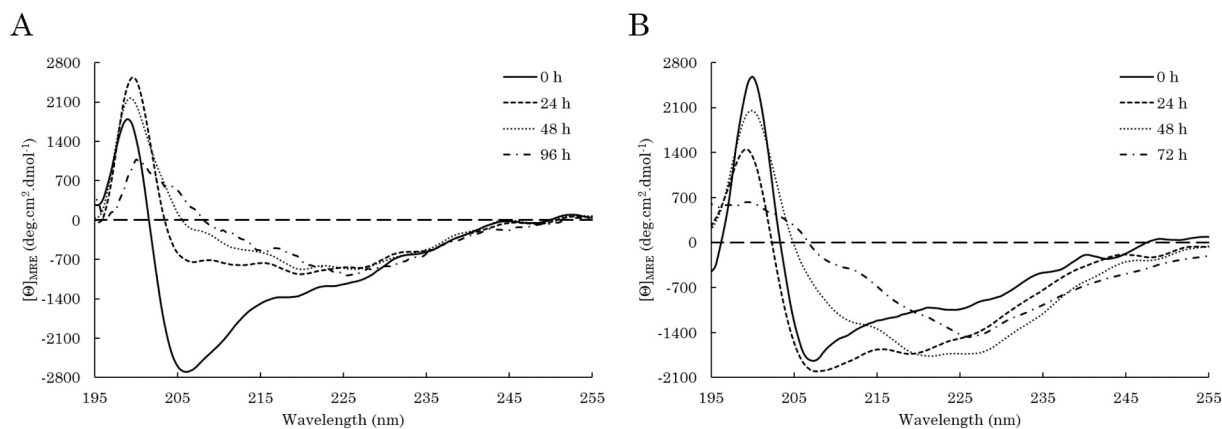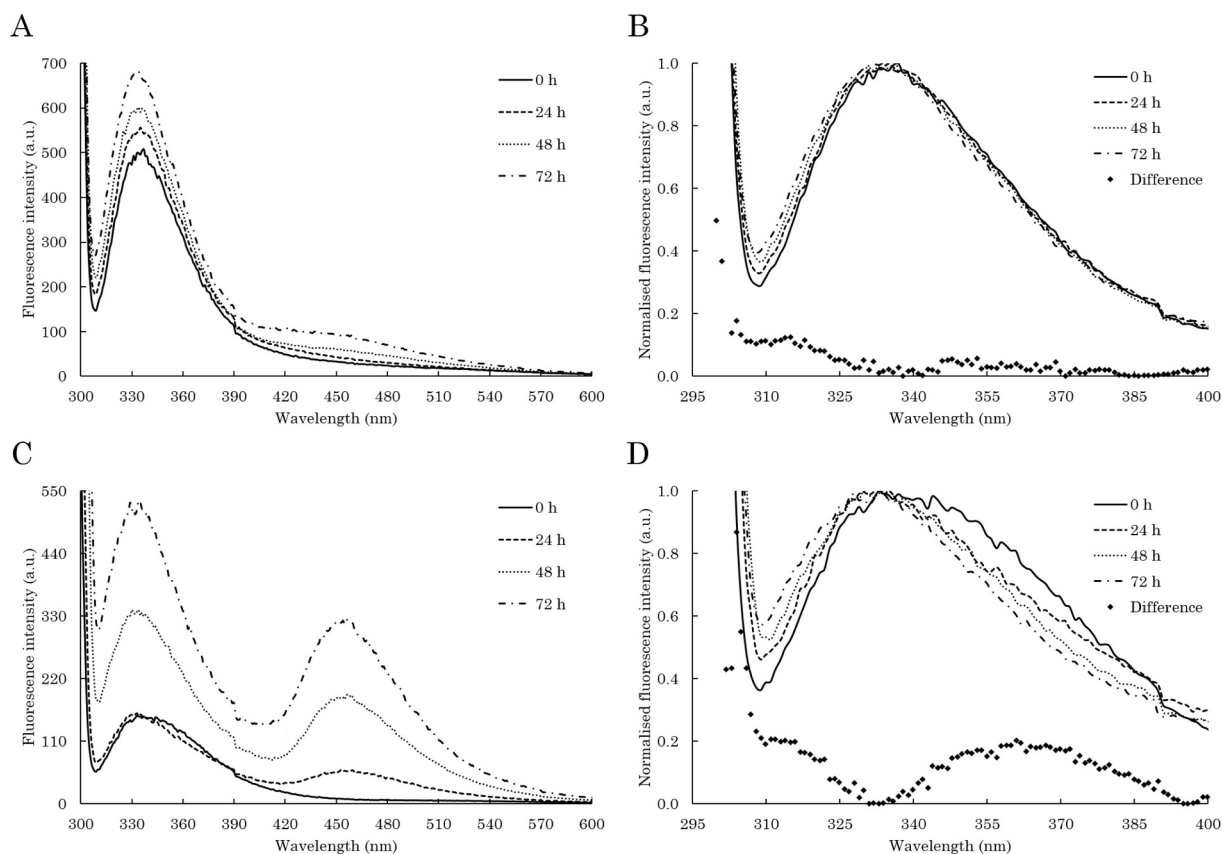
Fig. 6. Far-UV CD spectra of full-length (A) DPF3b and (B) DPF3a in TBS at ~20 °C at 0 h (solid line), 24 h (dashed line), 48 h (dotted line), 72 h for DPF3a and 96 h for DPF3b (dash-dotted line) incubation.

content can be due to the formation of subsequent transiently folded substructures, needed for the DPF3b fibrillation pathway.

Regarding DPF3a, comparable conformational transformations are detected by CD (Fig. 6B). After 24 h incubation, the main changes consist in a more distinctive α-helix contribution with more visible minima at 207 and 220 nm, and a loss in random coil, which is consistent with the BeStSel output. Unlike DPF3b, DPF3a observes an α-helix enrichment during the first steps of its aggregation, which may be indicative of distinctive structural rearrangement in the fibrillation pathway. The same behaviour has been reported for the amyloidogenic IDP, α-synuclein (α-syn), for which amyloid β-sheet cores are formed via

α-helix-rich intermediates [66]. The following 24 h are characterised by a significant loss in α-helical structure, resulting in a drastic shift of the minimum from 207 nm to an extensive negative band centred at ~224 nm. One can expect the formation of additional β-sheets, which is shown by secondary structure estimations. Such percentages are similar to the ones of DPF3b after 96 h incubation, suggesting that DPF3a structurally reorders faster into amyloid seeds, hence hastening its fibrillation. This is supported by the CD spectrum at 72 h, displaying a more defined negative band at 226 nm, characteristic of amyloid β-sheet-rich structures. In that respect, the spectrum deconvolution indicates further loss in α-helix and random coil in favour of antiparallel β-sheets. With

Fig. 7. (A, C) ITF spectra (λ_{ex} = 295 nm, sw = 10 nm) and (B, D) normalised first ITF band spectra of full-length (A, B) DPF3b and (C, D) DPF3a in TBS at ~20 °C at 0 h (solid line), 24 h (dashed line), 48 h (dotted line), and 72 h (dash-dotted line) incubation. On normalised spectra, the spectral difference between 0 and 72 h is represented by diamonds.
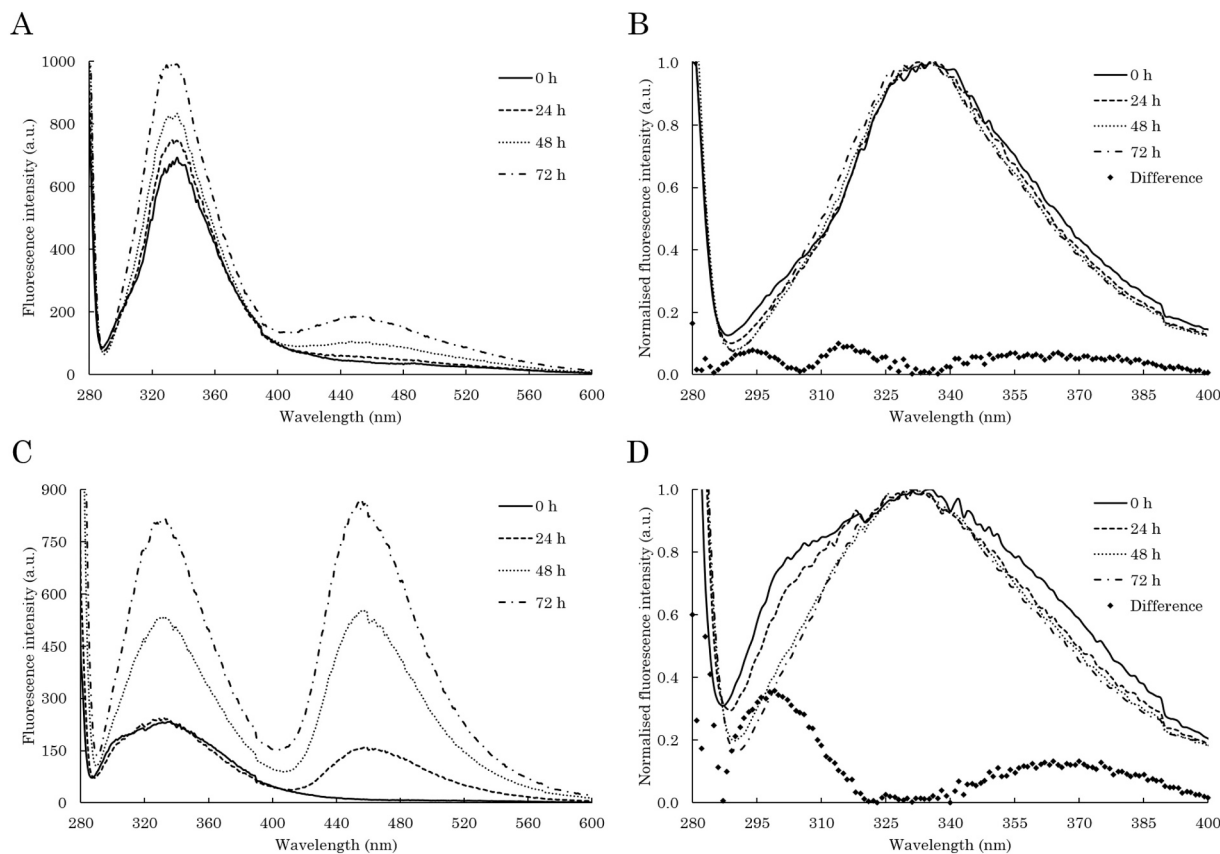
Fig. 8. (A, C) ITyrF spectra (λ_{ex} = 275 nm, sw = 10 nm) and (B, D) normalised Trp-Tyr emission band spectra of full-length (A, B) DPF3b and (C, D) DPF3a in TBS at ~20 °C at 0 h (solid line), 24 h (dashed line), 48 h (dotted line), and 72 h (dash-dotted line) incubation. On normalised spectra, the spectral difference between 0 and 72 h is represented by diamonds.

respect to the time scale, fewer transient conformers may be necessary to trigger DPF3a fibrillation thanks to its disordered C-terminal domain.

ITF (Fig. 7) and ITyrF (Fig. 8) were used to follow local conformational modifications upon aggregation. Over the course of 72 h, DPF3b Trp residues exhibit a gradual increase in their fluorescence intensity (Fig. 7A), while their λ_{em} is slightly blue-shifted to ~333 nm (Fig. 7B). Both phenomena are associated to more buried Trp residues in a hydrophobic core, which is coherent with DPF3b folding into residual ordered conformers in order to fibrillate. Such conformers may act as amyloid nucleation points. However, the most remarkable spectral feature is the emergence of a second emission band at ~456 nm, for

which the intensity increases as DPF3b proceeds towards aggregates (Fig. 7A). This new ITF signal has already been reported for DPF3a fibrils, but is yet to be rationalised [22]. Similarly, DPF3b Trp-Tyr emission band is increasing in intensity over time, indicating that Tyr residues are found in a less polar environment (Fig. 8A) [67]. The small shoulder at ~305 nm disappears, and the λ_{em} is shifted to ~333 nm (Fig. 8B). The loss of the specific Tyr fluorescence at 305 nm indicates an enhancement of Trp-Tyr FRET due to structural rearrangement of DPF3b aggregates. The band at 456 nm is also identified on ITyrF spectra, and appears more pronounced (Fig. 8A). Such ITyrF second emission has also been detected for transthyretin aggregates [68].
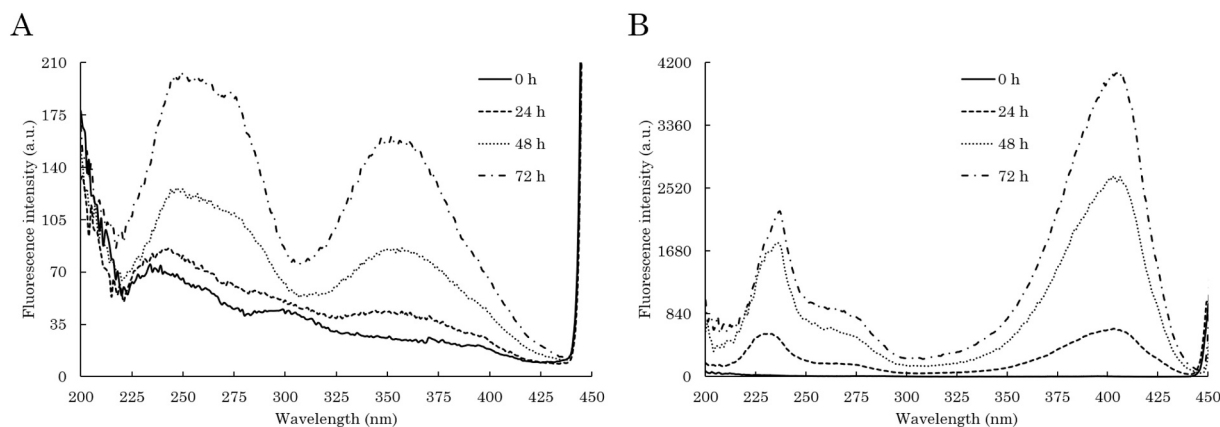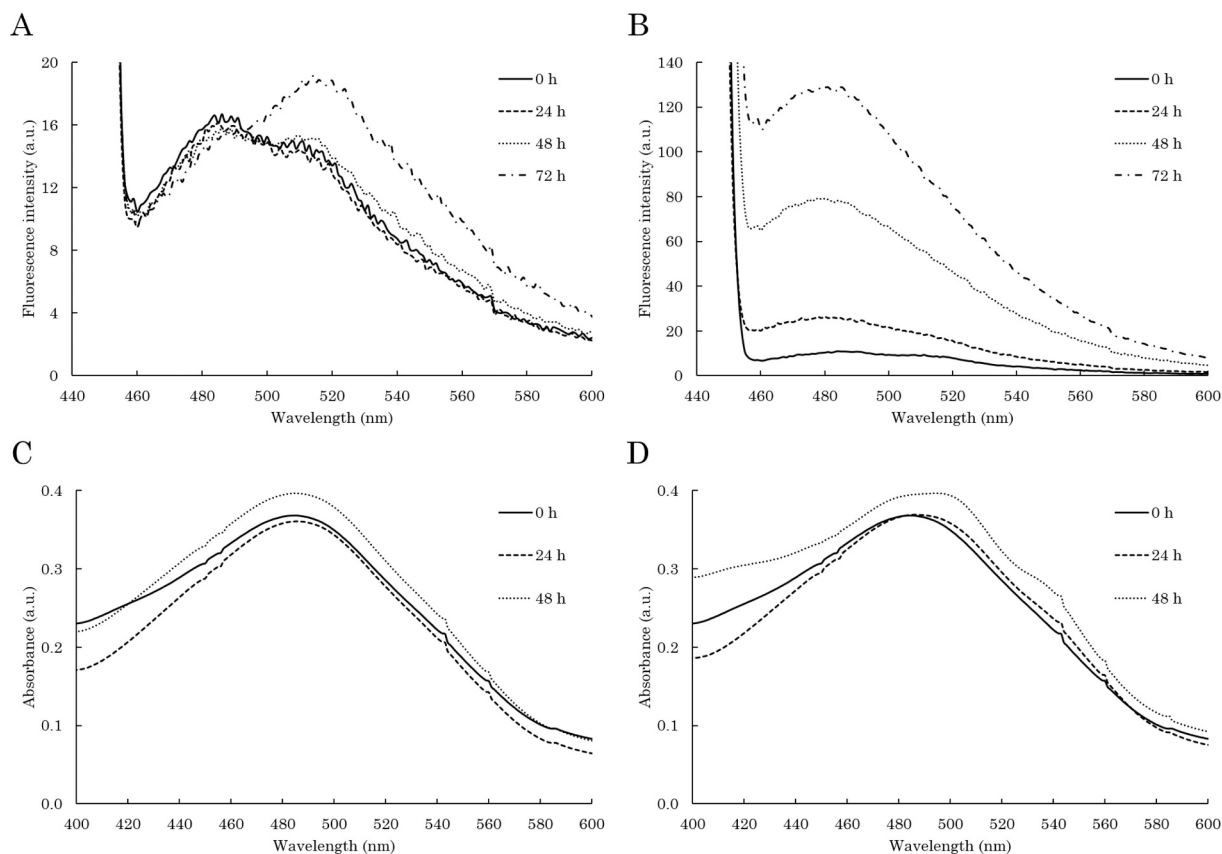

Fig. 9. dbAF excitation spectra (λ_{em} = 456 nm, sw = 10 nm) of full-length (A) DPF3b and (B) DPF3a in TBS at ~20 °C at 0 h (solid line), 24 h (dashed line), 48 h (dotted line), and 72 h (dash-dotted line) incubation.

A similar behaviour is observed for DPF3a, albeit of a different order of magnitude. In the first 24 h, Trp fluorescence intensity does not increase (Fig. 7C), while the λ_{em} starts to be blue-shifted with less broadening between 340 and 400 nm (Fig. 7D). In the following 48 h, the quantum yield of Trp residues is significantly increased, the λ_{em} is blue-shifted to 332 nm, and the broadening disappears, which unambiguously relates to Trp residues in a more non-polar environment. The second emission band at 456 nm is all the more noticeable, for which the intensity also drastically increases over aggregation (Fig. 7C). DPF3a spontaneous structural reorganisation is even more emphasised by ITyrF. No enhancement in emission intensity is observed within 24 h (Fig. 8C), but the contribution at ~305 nm decreases, and the Trp-Tyr band sharpens (Fig. 8D). From 24 to 72 h, the Tyr emission at 305 nm is definitively lost, as well as the broadening between 340 and 400 nm. By amyloid folding, previously isolated Tyr and Trp residues are brought together, resulting in Trp-Tyr FRET improvement. The fluorescence at 456 nm is once more very appreciable, as its intensity exceeds the Trp-Tyr one within 48 h (Fig. 8C).

In order to pave the way for understanding the origin of the 456 nm-emission detected by ITF and ITyrF, excitation spectra were recorded (Fig. 9). The spectrum associated to DPF3b reveals two main contributions: a broad one between 245 and 280 nm, and a second one centred at 355 nm (Fig. 9A). Consistently with ITF and ITyrF footprints, the intensity of these excitation bands increases over time. The first band explains why the 456 nm emission signal can arise from excitation of Trp and/or Tyr residues, the latter having a higher sensitivity. With respect to the second band, comparable emissions have previously been reported in the 350–400 nm excitation range for amyloidogenic proteins such as α-syn, tau, or human insulin [69,70]. This phenomenon is actually referred to as deep-blue autofluorescence (dbAF) or blue-green

autofluorescence. Though related to protein reordering, the precise origin of this autofluorescence remains debated, and its use as an intrinsic aggregation probe is still overlooked. It has been recently proposed that dbAF originates from numerous infrastructure interactions, such as aromatic and hydrogen bonds, due to the reorientation of residues side chains within aggregates. More specifically, it could arise from delocalisation of electrons through hydrogen bonds between cross β-sheets [71,72]. Nonetheless, dbAF emission in ITF and ITyrF spectra, as well as a second excitation band at 355 nm are indicative of DPF3b spontaneous aggregation.

The excitation profile of DPF3a dbAF completely differs (Fig. 9B). Three main components are identified: a first band at ~235 nm, a second one spread from 260 to 280 nm, and a third one, much more distinctive and centred at ~400–405 nm. In accordance with ITF and ITyrF, the intensity of each excitation band also increases over time. The band at 235 nm is attributed to the absorption of the peptide bond, at the level of carbonyl moieties [73,74]. The second between 260 and 280 nm is likely associated to aromatic residues, which accounts for dbAF emission in ITF and ITyrF spectra. Once again, Tyr residues are more sensitive to dbAF intensity changes. In comparison with DPF3b, the dominant one is completely shifted from 355 to 400–405 nm. As DPF3a seems to fibrillate faster than DPF3b, this dbAF excitation is considered typical of a straightforward amyloid pathway. On the contrary, DPF3b dbAF excitation signatures seem to describe more complex structural transitions, which are not yet amyloid-related within 72 h. This hypothesis is further supported by the DPF3b dbAF excitation spectra recorded after 7 days (168 h) of incubation (Fig. S3). Indeed, the resulting spectrum shows the same three excitation bands as for DPF3a, demonstrating that DPF3b can undergo fibrillation through additional conformational rearrangement steps. Therefore, shifting of the dbAF excitation maximum from 355 to

Fig. 10. Amyloid-specific dye binding assays. ThT emission spectra (λ_{ex} = 440 nm, sw = 10 nm, C_{ThT} = 10 μM) of full-length (A) DPF3b and (B) DPF3a in TBS at ~20 °C at 0 h (solid line), 24 h (dashed line), 48 h (dotted line), and 72 h (dash-dotted line) incubation. UV–visible CR absorption spectra (C_{CR} = 10 μM) of full-length (C) DPF3b and (D) DPF3a in TBS at ~20 °C at 0 h (solid line), 24 h (dashed line), and 48 h (dotted line) incubation.

400 nm would be an indicator of amyloid fibrillation. Consistent with excitation spectra, both isoforms are characterised by an overtime increase of the dbAF emission signal at 456 nm after excitation at 400 nm (Fig. S4A et B).

Evaluation of DPF3 aggregates amyloid character was confronted to amyloid-specific dye binding assay (Figs. 10A and B). Thioflavin T (ThT) is an extrinsic fluorophore known to exhibit a typical increase in its emission intensity at about 485 nm (after excitation at 440 nm) upon interacting with amyloid fibrils. ThT binding specificity arises from the fibril structure, presenting side-chain grooves at its surface [75,76]. DPF3b emission spectra in the presence of ThT support previous spectral observations (Fig. 10A). Indeed, no increase in ThT fluorescence intensity is detected within 72 h, which is in line with the formation of transient non-amyloid β-sheet-rich structures. In agreement with dbAF, ThT emission increases after 7 days (168 h) of incubation, which ascertains the formation of amyloid fibrils (Fig. S5). As expected from DPF3a higher fibrillation propensity, associated ThT spectra show a significant intensity enhancement up to a twelvefold increase after 72 h (Fig. 10B). Congruently with previous spectroscopy analyses, DPF3a ThT signatures ascertain the formation of amyloid seeds within 24–48 h.

In parallel with the ThT binding assay, the Congo red (CR) dye was also used, as it possesses comparable amyloid-specific spectroscopic properties (Fig. 10C and D). CR binding to amyloid fibrils is known to result in changes of its absorption spectrum, such as a shifted λ_{abs} from 490 to 500–510 nm, the emergence of a shoulder at 540 nm, and a general hyperchromic effect [77]. While a slight increase in absorbance intensity is observed for DPF3b after 48 h, there is no λ_{abs} shift, therefore confirming the absence of any amyloid-like aggregates (Fig. 10C). In contrast, DPF3a displays every characteristic amyloid signature upon CR addition after 48 h (Fig. 10D). The λ_{abs} is shifted towards 508 nm, the band presents a shoulder at 540 nm, and the absorption intensity is increased, advocating for the formation of an amyloid core within 48 h. Nonetheless, DPF3b also binds CR after 7 days (168 h) incubation, because it exhibits identical patterns (Fig. S6).

Along with experimental evidence, fibrillation propensity is another property that can be estimated by amyloid predictors (see Section 2.3). The AMYLPRED2 server, grouping 11 independent prediction algorithms, aims at finding aggregation hotspots in the query protein sequence on the basis of its primary structure. On a consensus of 5 different methods, outputs reveal that DPF3a has 2 amyloid hotspots in the 2/3 domain (Fig. S7A). DPF3b has an additional hotspot in the PHD-1 domain (Fig. S7B). In comparison, CamSol intrinsic solubility predictions assign 5 poor soluble segments in both isoforms, more precisely in the 2/3 domain, as well as in C_2H_2 and PHD-1 ZnFs (Fig. S8A). In addition, 6 more hotspots are found in DPF3b double PHD ZnF (Fig. S8B). While the hits overlapping with AMYLPRED2 are likely aggregation-promoting spots, the other highlighted by CamSol are realistically less hydrophilic due to their location in folded domains. One should not confuse aggregation potential with aggregation rate. Driven by intermolecular interactions, conformational dynamics of DPF3a C-terminal domain could more readily force it to adopt specific folds leading to aggregates. In other words, despite the absence of predicted hotspots at the C-terminus, its faster rearrangement into a typical amyloid β-sheet core probably hastens the fibrillation process. This hypothesis is in line with the DPF3b-DPF3a kinetics discrepancy as discussed in the spectroscopy results.

Metal ions are known to impact the aggregation pathways of proteins, especially IDPs. By binding to IDPs and reducing charge repulsion, they favour conformation ensembles that can promote self-assembly [78,79]. DPF3 isoforms possessing ZnF domains and zinc binding residues (Fig. S9), Zn^{2+} cations may modify their aggregation properties by affecting conformational changes and fibrillation kinetics. Thus, it could be informative to assess DPF3 sensitivity to Zn^{2+} and other divalent cations involved in neurodegenerative diseases, such as Ca^{2+}, Mg^{2+}, or Cu^{2+}.

3.4. Morphological features of DPF3 aggregates

Albeit spectroscopy techniques are powerful tools for monitoring conformational changes of proteins, they can be limited when it comes to discriminate different forms of aggregation. In that respect, transmission electron microscopy (TEM) enables to visualise aggregates and fibrils at the nanometre scale [80]. Previously, we have observed that DPF3a is able to form oligomers, protofibrils, and short needle-shape amyloid filaments [22]. Here, the morphological diversity of DPF3b aggregates has been unveiled and compared to that of DPF3a. Considering that amyloid spectroscopy footprints were seen for DPF3b after 7 days (168 h) of incubation, DPF3a samples were also incubated 7 days prior to TEM analysis. The presence of aggregated species was first assessed by DLS, which reveals a wide single population presenting a $\overline{D}_h$ ranging from ~200 to ~270 nm for both isoforms (Fig. S1C). TEM micrographs unveil several distinctive aggregate types, providing information on the assembly mechanisms followed by DPF3 isoforms (Fig. 11). For both isoforms, spherical nucleation units (SNUs), consisting of spheric-shaped high-order oligomers, are found in the vicinity of granular prefibrillar aggregates (also called protofibrils), which can elongate and branch out from SNUs clustering (Fig. 11A). SNUs diameter amounts to ~25–40 nm, which is in the range of reported tau SNUs [81].

Textbook single amyloid fibrils are difficult to find because both DPF3b (Fig. 11B) and DPF3a (Fig. 11C) tend to form large fibrillar clusters, taking the shape of striated twisted ribbon fibrils (STRFs). DPF3 STRFs present distinguishable morphologic features, though a noticeable variability in sizes is observed. They are characterised by a series of twists and untwists following a certain periodicity, giving STRFs a pseudo-helical appearance, which is an amyloid fingerprint [82]. This periodicity seems to arise from filaments assembly as it is not observed for isolated fibrils (Fig. 11B), which appear more like straight filaments (SFs). SFs are already mature fibrils that are ~18–25 nm wide, falling in the range observed for amyloid fibrils [83]. For DPF3b, the width of twists is ~100–110 nm, while untwists are ~200–230 nm wide (Fig. 11B). Regarding the representative DPF3a STRF (Fig. 11C), twists and untwists are ~30–40 nm and ~70–90 nm wide, respectively. Once again, the size discrepancy is not isoform-specific, and amyloid fibrils often display large variations in size. It appears that the width of untwists is twice larger than the twists. Also, the twist-to-twist distance in STRFs is approximately ten- to fifteenfold the twist width. Similar ratios have been described for tau twisted ribbon and paired helical filaments [81,84]. The striated aspect of DPF3b and DPF3a STRFs relates to the alignment of multiple stretched linear protofibrils and/or SFs (Fig. 11D). Each striation is ~3–4 nm wide, and corresponds to a single fibril. Furthermore, diversely sized STRFs are often encountered within densely entangled fibrous networks, where numerous small STRFs merge into large ones (Fig. 11E). Amyloidogenic proteins are known to bundle into neurofibrillary networks in neurodegenerative diseases [85]. Beside the formation of highly structured fibrils, amorphous aggregation can also occur during DPF3 isoforms assembly (Fig. 11F).

The presence of intermediary aggregated species suggests that DPF3 isoforms undergo a colloidal fibrillation pathway, rather than a nucleation-elongation process, in a comparable manner to the tau protein. In the colloidal model, SNUs are considered as colloidal units that interact with each other in order to form fibrillar structures. Conversely, in the nucleation-elongation model, fibrillation is carried out by addition of protein monomers [81]. Nevertheless, the colloidal pathway may not be exclusive, as both mechanisms could concomitantly occur and lead to morphologically similar fibrils. Because DPF3b share common aggregation features with DPF3a, we propose that it essentially follows the same colloidal fibrillation pathway (Fig. 12). In early aggregation stage, DPF3 assembles into SNUs, enriching in β-sheet secondary structure. Indeed, such oligomers have shown to bind ThT [86]. Pre-existence of smaller oligomers acting as a SNU initiator is very likely. SNUs serve as nucleation points for the growth of granular protofibrils, which
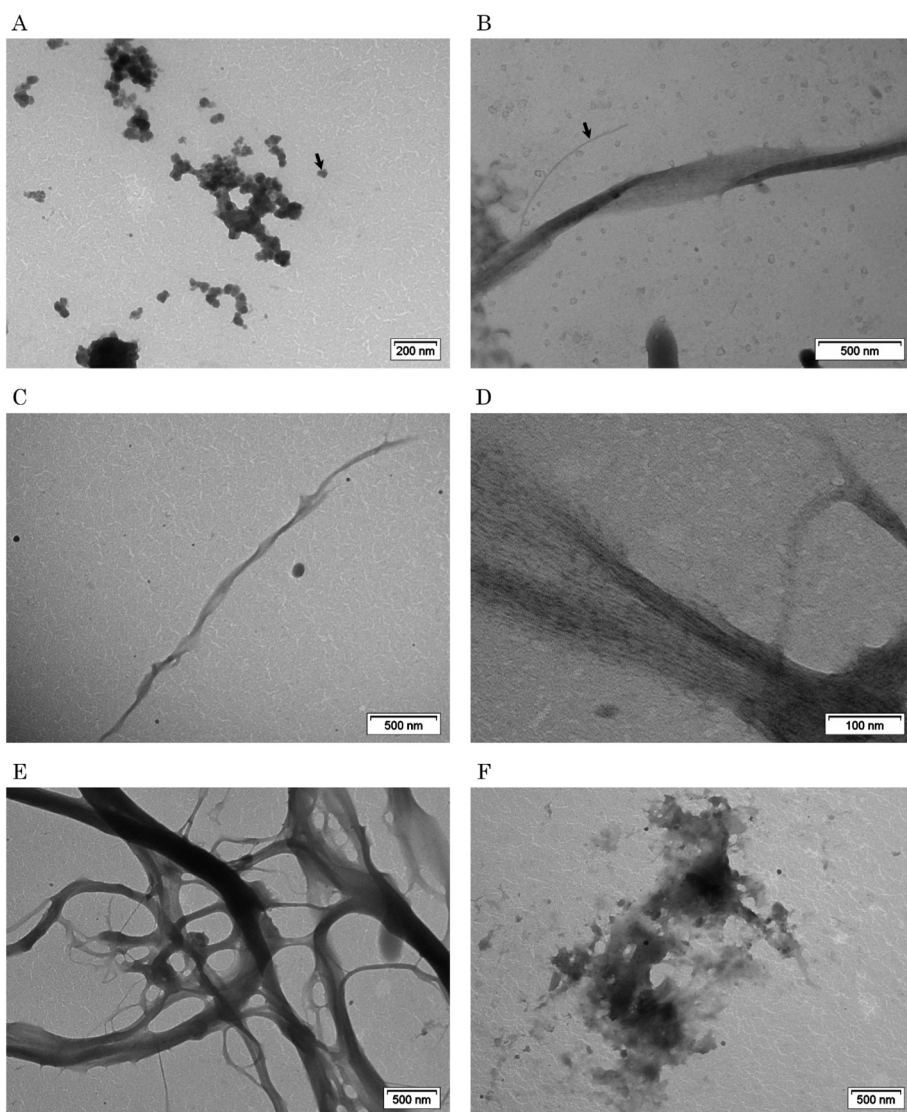
Fig. 11. NS TEM micrographs (voltage of 100 kV) of 7 days-incubated DPF3b and DPF3a in TBS at ~20 °C. (A) SNUs (an isolated one is indicated by the black arrow) clustering into granular protofibrils (DPF3a sample). (B) DPF3b and (C) DPF3a twisted ribbon-shaped mature amyloid assemblies (STRFs), made up of a multitude of smaller mature fibrils (an isolated one is indicated by the black arrow on panel B). (D) Close up on a large STRF with the typical striated pattern related to individual stretched fibrils (DPF3a sample). (E) Entangled network of fibrous amyloid aggregates (DPF3a sample). (F) Amorphous aggregated phase (DPF3a sample). On each micrograph, the scale bar is indicated at the bottom right.

resemble strings of beads. These prefibrillar structures then elongate, and can also ramify, by SNU packing. Linear protofibrils assemble and stretch into SFs. Such fibrils are then prone to intertwine and gather into large STRFs, forming tangled fibrillar amyloid networks.

4. Conclusions

In the present study, structural properties of DPF3b and DPF3a isoforms have been investigated *in vitro* by light scattering, spectroscopy, and microscopy techniques. Predictions and experimental evidence have been consistent with both isoforms being highly disordered proteins. CD spectroscopy has revealed similar footprints dominated by random coil, and measured hydrodynamic diameters by DLS indicate that they adopt expanded conformations. DPF3a has also exhibited more intrinsic disorder than DPF3b, as highlighted by structural predictors, UV–visible absorption spectroscopy, and the presence of additional IDR and MoRFs in its C-terminal domain. Supported by solvent accessibility predictions, ITF and ITyrF have highlighted that Trp and Tyr residues of DPF3b are less exposed to the solvent than DPF3a, suggesting that the latter possesses a more extended conformational ensemble.

Time-dependent biophysical characterisation has demonstrated that both DPF3 isoforms are prone to spontaneous aggregation, and are capable of assembling into amyloid fibrils. By confronting spectroscopy analyses, we found out that DPF3 isoforms have comparable, although kinetically different, fibrillation pathways. A gradual increase in anti-parallel β-sheet content has been observed by CD for the two isoforms with the formation of α-helix-enriched intermediates for DPF3a. ITF and ITyrF have also been efficient probes for local structural modifications, as well as for unravelling aggregation mechanisms. Trp and Tyr residues have shown to be found in a less polar environment upon aggregation. Complementarily, dbAF and the associated excitation spectra have proven to be suitable and sensitive techniques for monitoring the course of fibrillation, particularly useful in discriminating transient aggregated states.

ThT and CR binding assays have ascertained the amyloid nature of DPF3 aggregates, and supported that DPF3a fibrillates faster. Indeed, DPF3b aggregation appears slower due to successive order-to-disorder-to-order transitions. Indeed, the unfolding of its double PHD ZnF, followed by the refolding into an amyloid β-sheet core appears to elongate the fibrillation lag phase compared to DPF3a. Conversely, DPF3a rapidly folds and aggregates into amyloid seeds, which assemble into mature fibrils, thanks to its highly dynamic C-terminal domain, thus allowing quicker disorder-to-order transitions. DPF3 isoforms have displayed a similar aggregation polymorphism by TEM, indicative of their fibrillation mechanism, involving SNUs, "worm-like" protofibrils, SFs, STRFs, and neurofibrillary tangles. DPF3 isoforms being zinc finger proteins,
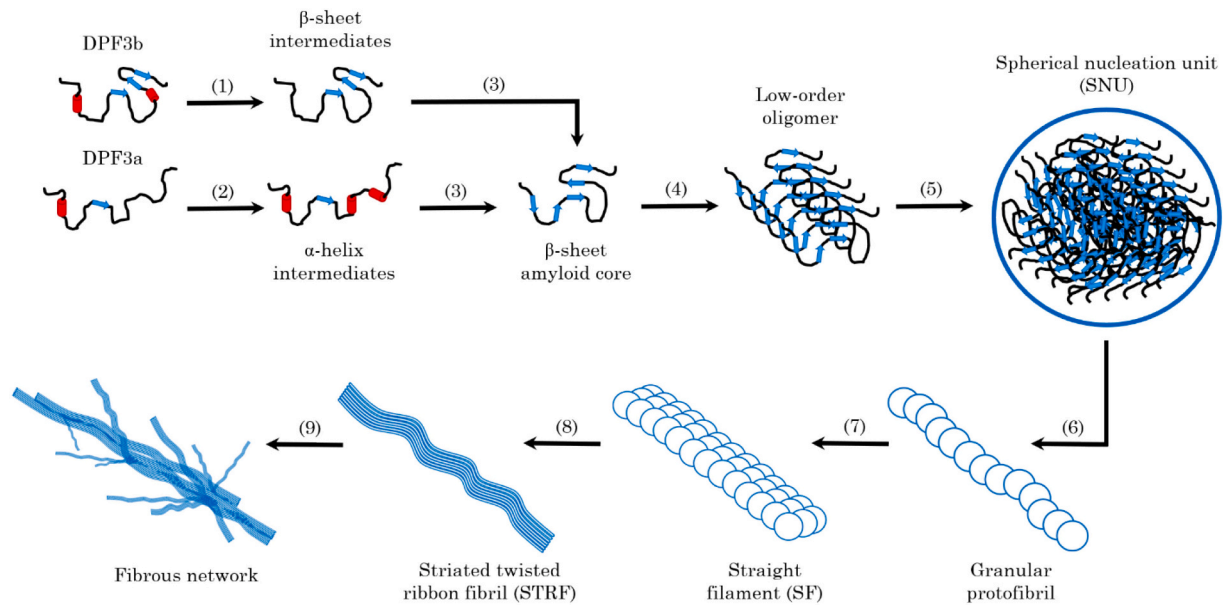
Fig. 12. Suggested fibrillation pathway of DPF3 isoforms. (1) DPF3b rearranges by losing α-helixes, while (2) DPF3a forms α-helix-enriched intermediates. (3) The β-sheet content of the two isoforms gradually increases in order to form an amyloid core, (4) driving the assembly of monomers. (5) Low-order oligomers gather into SNUs, acting as nucleation points for fibrils formation. (6) SNUs cluster into granular "worm-like" protobrils. (7) Protofibrils elongate, assemble, and stretch into SFs. (8) SFs bundle intro STRFs whose striations correspond to individual SFs. (9) STRFs of diverse sizes clump into densely entangled fibrous networks.

the influence of Zn^{2+} as well as other divalent metal cations on their aggregation properties will be investigated in future works. Additional physicochemical parameters, such as temperature, pH, or ionic strength, will also be considered.

At a physio-pathological point of view, our study opens new avenue of investigation for DPF3 isoforms. Like other proteins such as tau, DPF3 aggregation in neurons might be unequivocally associated with neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and amyotrophic lateral sclerosis. Because protein aggregation has been linked to TOF [87], it would deserve to determine whether DPF3 isoforms also tend to self-aggregate in these patients. Finally, mutation in CEP135 protein forms protein aggregates in flagella and induces multiple morphological abnormalities of the sperm flagella (MMAF), a rare disease associated with primary infertility [88]. As SNP in DPF3 is related to male infertility, it would be highly interesting to look whether the presence of SNP and/or mutations in DPF3 isoforms would further promote or accelerate protein aggregation, potentially affecting the normal function of spermatozoids.

In sum, this study has provided insights into the respective disorder-based functionality of DPF3 isoforms. DPF3b and DPF3a have been identified as two new amyloidogenic IDPs, comparatively to α-synuclein and the protein tau, making them new promising druggable targets in IDP- and aggregation-associated diseases, such as cancer, cardiovascular and neurodegenerative disorders.

CRediT authorship contribution statement

Julien Mignon: Conceptualisation; Data acquisition and interpretation; Writing; Reviewing; Editing.

Denis Mottet: Data interpretation; Reviewing; Editing.

Tanguy Leyder: Data acquisition and interpretation; Reviewing; Editing.

Vladimir N. Uversky: Bioinformatics data acquisition and analyses; Reviewing; Editing.

Eric A. Perpète: Supervision; Data acquisition; Reviewing; Editing.

Catherine Michaux: Conceptualisation; Supervision; Reviewing; Editing.

Declaration of competing interest

Authors declare that they do not have any conflict of interest.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2022.07.102.

References

[1] A.J. Morrison, Chromatin-remodeling links metabolic signaling to gene expression, Mol. Metab. 38 (2020), https://doi.org/10.1016/j.molmet.2020.100973.

[2] M. Tyagi, N. Imam, K. Verma, A.K. Patel, Chromatin remodelers: we are the drivers!!, Nucleus 7 (2016) 388–404, https://doi.org/10.1080/19491034.2016.1211217.

[3] M. Lange, B. Kaynak, U.B. Forster, M. Tönjes, J.J. Fischer, C. Grimm, J. Schlesinger, S. Just, I. Dunkel, T. Krueger, S. Mebus, H. Lehrach, R. Lurz, J. Gobom, W. Rottbauer, S. Abdelilah-Seyfried, S. Sperling, Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex, Genes Dev. 22 (2008) 2370–2384, https://doi.org/10.1101/gad.471408.

[4] A. Ishizaka, T. Mizutani, K. Kobayashi, T. Tando, K. Sakurai, T. Fujiwara, H. Iba, Double plant homeodomain (PHD) finger proteins DPF3a and -3b are required as transcriptional co-activators in SWI/SNF complex-dependent activation of NF-κB RelA/p50 heterodimer, J. Biol. Chem. 287 (2012) 11924–11933, https://doi.org/10.1074/jbc.M111.322792.

[5] J. Lessard, J.I. Wu, J.A. Ranish, M. Wan, M.M. Winslow, B.T. Staahl, H. Wu, R. Aebersold, I.A. Graef, G.R. Crabtree, An essential switch in subunit composition of a chromatin remodeling complex during neural development, Neuron 55 (2007) 201–215, https://doi.org/10.1016/j.neuron.2007.06.019.

[6] L. Zeng, Q. Zhang, S. Li, A.N. Plotnikov, M.J. Walsh, M.M. Zhou, Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b, Nature 466 (2010) 258–262, https://doi.org/10.1038/nature09139.

[7] W. Li, A. Zhao, W. Tempel, P. Loppnau, Y. Liu, Crystal structure of DPF3b in complex with an acetylated histone peptide, J. Struct. Biol. 195 (2016) 365–372, https://doi.org/10.1016/j.jsb.2016.07.001.

[8] X. Zhu, B. Lan, X. Yi, C. He, L. Dang, X. Zhou, Y. Lu, Y. Sun, Z. Liu, X. Bai, K. Zhang, B. Li, M.J. Li, Y. Chen, L. Zhang, HRP2-DPF3a-BAF complex coordinates histone modification and chromatin remodeling to regulate myogenic gene transcription, Nucleic Acids Res. 48 (2020) 6563–6582, https://doi.org/10.1093/nar/gkaa441.

[9] K. Fu, H. Nakano, M. Morselli, T. Chen, H. Pappoe, A. Nakano, M. Pellegrini, A temporal transcriptome and methylome in human embryonic stem cell-derived cardiomyocytes identifies novel regulators of early cardiac development, Epigenetics. 13 (2018) 1013–1026, https://doi.org/10.1080/15592294.2018.1526029.

[10] H. Cui, J. Schlesinger, S. Schoenhals, M. Tönjes, I. Dunkel, D. Meierhofer, E. Cano, K. Schulz, M.F. Berger, T. Haack, S. Abdelilah-Seyfried, M.L. Bulyk, S. Sauer, S. R. Sperling, Phosphorylation of the chromatin remodeling factor DPF3a induces cardiac hypertrophy through releasing HEY repressors from DNA, Nucleic Acids Res. 44 (2015) 2538–2553, https://doi.org/10.1093/nar/gkv1244.

[11] W. Guanglei, W. Bingbing, Y. Peixin, Epigenetics in congenital heart disease, J. Am. Heart Assoc. 11 (2022), https://doi.org/10.1161/JAHA.121.025163.

[12] H. Cui, J. Schlesinger, S. Schoenhals, M. Tönjes, I. Dunkel, D. Meierhofer, E. Cano, K. Schulz, M.F. Berger, T. Haack, S. Abdelilah-Seyfried, M.L. Bulyk, S. Sauer, S. R. Sperling, Phosphorylation of the chromatin remodeling factor DPF3a induces cardiac hypertrophy through releasing HEY repressors from DNA, Nucleic Acids Res. 44 (2015) 2538–2553, https://doi.org/10.1093/nar/gkv1244.

[13] H. Hiramatsu, K. Kobayashi, K. Kobayashi, T. Haraguchi, Y. Ino, T. Todo, H. Iba, The role of the SWI/SNF chromatin remodeling complex in maintaining the stemness of glioma initiating cells, Sci. Rep. 7 (2017), https://doi.org/10.1038/s41598-017-00982-3.

[14] W. Hao Lin, W. Gang Dai, X. Dong Xu, Q. Hua Yu, B. Zhang, J. Li, <check>H. </check>, Ping Li, Downregulation of DPF3 promotes the proliferation and motility of breast cancer cells through activating JAK2/STAT3 signaling, Biochem. Biophys. Res. Commun. 514 (2019) 639–644, https://doi.org/10.1016/j.bbrc.2019.04.170.

[15] L.M. Colli, L. Jessop, T.A. Myers, S.Y. Camp, M.J. Machiela, J. Choi, R. Cunha, O. Onabajo, G.C. Mills, V. Schmid, S.A. Brodie, O. Delattre, D.R. Mole, M. P. Purdue, K. Yu, K.M. Brown, S.J. Chanock, Altered regulation of DPF3, a member of the SWI/SNF complexes, underlies the 14q24 renal cancer susceptibility locus, Am. J. Hum. Genet. 108 (2021) 1590–1610, https://doi.org/10.1016/j.ajhg.2021.07.009.

[16] J. Protze, S. Naas, R. Krüger, C. Stöhr, A. Kraus, S. Grampp, M. Wiesener, M. Schiffer, A. Hartmann, B. Wullich, J. Schödel, The renal cancer risk allele at 14q24.2 activates a novel hypoxia-inducible transcription factor-binding enhancer of DPF3 expression, J. Biol. Chem. 298 (2022), https://doi.org/10.1016/j.jbc.2022.101699.

[17] G. Kosova, J.M. Hotaling, S. Ohlander, C. Niederberger, G.S. Prins, C. Ober, Variants in DPF3 and DSCAML1 are associated with sperm morphology, J. Assist. Reprod. Genet. 31 (2014) 131–137, https://doi.org/10.1007/s10815-013-0140-9.

[18] S.Y. Liu, C.J. Zhang, H.Y. Peng, H. Sun, K.Q. Lin, X.Q. Huang, K. Huang, J.Y. Chu, Z.Q. Yang, Strong association of SLC1A1 and DPF3 gene variants with idiopathic male infertility in Han Chinese, Asian J. Androl. 18 (2016) 486–492, https://doi.org/10.4103/1008-682X.178850.

[19] Y. Sato, C. Hasegawa, A. Tajima, S. Nozawa, M. Yoshiike, E. Koh, J. Kanaya, M. Namiki, K. Matsumiya, A. Tsujimura, K. Komatsu, N. Itoh, J. Eguchi, A. Yamauchi, T. Iwamoto, Association of TUSC1 and DPF3 gene polymorphisms with male infertility, J. Assist. Reprod. Genet. 35 (2018) 257–263, https://doi.org/10.1007/s10815-017-1052-x.

[20] A. Local, H. Huang, C.P. Albuquerque, N. Singh, A.Y. Lee, W. Wang, C. Wang, J. E. Hsia, A.K. Shiau, K. Ge, K.D. Corbett, D. Wang, H. Zhou, B. Ren, Identification of H3K4me1-associated proteins at mammalian enhancers, Nat. Genet. 50 (2018) 73–82, https://doi.org/10.1038/s41588-017-0015-6.

[21] N. El Hadidy, V.N. Uversky, Intrinsic disorder of the baf complex: roles in chromatin remodeling and disease development, Int. J. Mol. Sci. 20 (2019), https://doi.org/10.3390/ijms20215260.

[22] J. Mignon, D. Mottet, G. Verrillo, A. Matagne, E.A. Perpète, C. Michaux, Revealing intrinsic disorder and aggregation properties of the DPF3a zinc finger protein, ACS Omega. 6 (2021) 18793–18801, https://doi.org/10.1021/acsomega.1c01948.

[23] P. Kulkarni, S. Bhattacharya, S. Achuthan, A. Behal, M.K. Jolly, S. Kotnala, A. Mohanty, G. Rangarajan, R. Salgia, V. Uversky, Intrinsically disordered proteins: critical components of the wetware, Chem. Rev. 122 (2022) 6614–6633, https://doi.org/10.1021/acs.chemrev.1c00848.

[24] P. Kulkarni, V.N. Uversky, Intrinsically disordered proteins in chronic diseases, Biomolecules 9 (2019), https://doi.org/10.3390/biom9040147.

[25] V.N. Uversky, Intrinsically disordered proteins: targets for the future ? in: J.-P. Renaud (Ed.), Structural Biology in Drug Discovery: Methods, Techniques, and Practices, First Edit, John Wiley & Sons Inc, 2020, pp. 587–612.

[26] B.J.H. Kuipers, H. Gruppen, Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis, J. Agric. Food Chem. 55 (2007) 5445–5451, https://doi.org/10.1021/jf070337l.

[27] R.M. Williams, Z. Obradovic, V. Mathura, W. Braun, E.C. Garner, S. Takayama, C. J. Brown, A.K. Dunker, The protein non-folding problem: amino acid determinants of intrinsic order and disorder, Pac. Symp. Biocomp. (2001) 89–100.

[28] K. Peng, S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker, Z. Obradovic, Optimizing long intrinsic disorder predictors with protein evolutionary information, J. Bioinforma. Comput. Biol. 3 (2005) 35–60, https://doi.org/10.1142/s0219720005000886.

[29] Garner, Romero, Dunker, Brown, Obradovic, in: Predicting Binding Regions Within Disordered Proteins. Genome Informatics. Workshop on Genome Informatics 10, 1999, pp. 41–50.

[30] I. Walsh, A.J.M. Martin, S.C.E. Tosatto, T. Di domenico, Espritz: Accurate and fast prediction of protein disorder, Bioinformatics 28 (2012) 503–509, https://doi.org/10.1093/bioinformatics/btr682.

[31] T. Ishida, K. Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence, Nucleic Acids Res. 35 (2007) 460–464, https://doi.org/10.1093/nar/gkm363.

[32] G. Erdos, M. Pajkos, Z. Dosztányi, IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation, Nucleic Acids Res. 49 (2021) 297–303, https://doi.org/10.1093/nar/gkab408.

[33] R.J. Emenecker, D. Griffith, A.S. Holehouse, Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure, Biophys. J. 120 (2021) 4312–4319, https://doi.org/10.1016/j.bpj.2021.08.039.

[34] M. Bernhofer, C. Dallago, T. Karl, V. Satagopam, M. Heinzinger, M. Littmann, T. Olenyi, J. Qiu, K. Schütze, G. Yachdav, H. Ashkenazy, N. Ben-Tal, Y. Bromberg, T. Goldberg, L. Kajan, S. O'Donoghue, C. Sander, A. Schafferhans, A. Schlessinger, G. Vriend, M. Mirdita, P. Gawron, W. Gu, Y. Jarosz, C. Trefois, M. Steinegger, R. Schneider, B. Rost, PredictProtein - predicting protein structure and function for 29 years, Nucleic Acids Res. 49 (2021) 535–540, https://doi.org/10.1093/nar/gkab354.

[35] S. Wang, J. Ma, J. Xu, AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields, Bioinformatics 32 (2016) i672–i679, https://doi.org/10.1093/bioinformatics/btw446.

[36] R. Dass, F.A.A. Mulder, J.T. Nielsen, ODiNPred: comprehensive prediction of protein order and disorder, Sci. Rep. 10 (2020), https://doi.org/10.1038/s41598-020-71716-1.

[37] B. Xue, R.L. Dunbrack, R.W. Williams, A.K. Dunker, V.N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta, Proteins Proteomics 2010 (1804) 996–1010, https://doi.org/10.1016/j.bbapap.2010.01.011.

[38] A.S. Holehouse, R.K. Das, J.N. Ahad, M.O.G. Richardson, R.V. Pappu, CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins, Biophys. J. 112 (2017) 16–21, https://doi.org/10.1016/j.bpj.2016.11.3200.

[39] I. Letunic, S. Khedkar, P. Bork, SMART: recent updates, new developments and status in 2020, Nucleic Acids Res. 49 (2021) D458–D460, https://doi.org/10.1093/nar/gkaa937.

[40] P. Jarnot, J. Ziemska-Legiecka, L. Dobson, M. Merski, P. Mier, M.A. Andrade-Navarro, J.M. Hancock, Z. Dosztányi, L. Paladin, M. Necci, D. Piovesan, S.C. E. Tosatto, V.J. Promponas, M. Grynberg, A. Gruca, PlaToLoCo: the first web meta-server for visualization and annotation of low complexity regions in proteins, Nucleic Acids Res. 48 (2020) W77–W84, https://doi.org/10.1093/NAR/GKAA339.

[41] N. Malhis, M. Jacobson, J. Gsponer, MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences, Nucleic Acids Res. 44 (2016) W488–W493, https://doi.org/10.1093/NAR/GKW409.

[42] M. Kaleel, M. Torrisi, C. Mooney, G. Pollastri, PaleAle 5.0: prediction of protein relative solvent accessibility by deep learning, Amino Acids 51 (2019) 1289–1296, https://doi.org/10.1007/s00726-019-02767-6.

[43] P. Sormanni, M. Vendruscolo, Protein solubility predictions using the camsol method in the study of protein homeostasis, Cold Spring Harb. Perspect. Biol. 11 (2019), https://doi.org/10.1101/cshperspect.a033845.

[44] A.C. Tsolis, N.C. Papandreou, V.A. Iconomidou, S.J. Hamodrakas, A consensus method for the prediction of "aggregation-prone" peptides in globular proteins, PLoS ONE 8 (2013), https://doi.org/10.1371/journal.pone.0054175.

[45] A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, Nucleic Acids Res. 46 (2018) 315–322, https://doi.org/10.1093/nar/gky497.

[46] W. Zhang, C. Xu, C. Bian, W. Tempel, L. Crombet, F. MacKenzie, J. Min, Z. Liu, C. Qi, Crystal structure of the Cys2His2-type zinc finger domain of human DPF2, Biochem. Biophys. Res. Commun. 413 (2011) 58–61, https://doi.org/10.1016/j.bbrc.2011.08.043.

[47] R. Van Der Lee, M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D.T. Jones, P.M. Kim, R.W. Kriwacki, C. J. Oldfield, R.V. Pappu, P. Tompa, V.N. Uversky, P.E. Wright, M.M. Babu, Classification of intrinsically disordered regions and proteins, Chem. Rev. 114 (2014) 6589–6631, https://doi.org/10.1021/cr400525m.

[48] V.N. Uversky, Intrinsically disordered proteins and their "mysterious" (meta) physics, Front. Phys. 7 (2019) 8–23, https://doi.org/10.3389/fphy.2019.00010.

[49] B. Xue, C.J. Oldfield, A.K. Dunker, V.N. Uversky, CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions, FEBS Lett. 583 (2009) 1469–1474, https://doi.org/10.1016/j.febslet.2009.03.070.

[50] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41 (2000) 415–427, https://doi.org/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.

[51] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132, https://doi.org/10.1016/0022-2836(82)90515-0.

[52] R.K. Das, R.V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 13392–13397, https://doi.org/10.1073/pnas.1304749110.

[53] E.W. Martin, A.S. Holehouse, C.R. Grace, A. Hughes, R.V. Pappu, T. Mittag, Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation, J. Am. Chem. Soc. 138 (2016) 15323–15335, https://doi.org/10.1021/jacs.6b10272.

[54] R.K. Das, K.M. Ruff, R.V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins, Curr. Opin. Struct. Biol. 32 (2015) 102–112, https://doi.org/10.1016/j.sbi.2015.03.008.

[55] H.Y.J. Fung, M. Birol, E. Rhoades, IDPs in macromolecular complexes: the roles of multivalent interactions in diverse assemblies, Curr. Opin. Struct. Biol. 49 (2018) 36–43, https://doi.org/10.1016/j.sbi.2017.12.007.

[56] S. Das, U. Pal, S. Das, K. Bagga, A. Roy, A. Mrigwani, N.C. Maiti, Sequence complexity of amyloidogenic regions in intrinsically disordered human proteins, PLoS ONE 9 (2014), https://doi.org/10.1371/journal.pone.0089781.

[57] O.M. Morris, J.H. Torpey, R.L. Isaacson, Intrinsically disordered proteins: modes of binding with emphasis on disordered domains, Open Biol. 11 (2021), https://doi.org/10.1098/rsob.210222.

[58] O. Tcherkasskaya, V.N. Uversky, Denatured collapsed states in protein folding: Example of apomyoglobin, Proteins 44 (2001) 244–254, https://doi.org/10.1002/prot.1089.

[59] M. Nygaard, B.B. Kragelund, E. Papaleo, K. Lindorff-Larsen, An efficient method for estimating the hydrodynamic radius of disordered protein conformations, Biophys. J. 113 (2017) 550–557, https://doi.org/10.1016/j.bpj.2017.06.042.

[60] J.T. Vivian, P.R. Callis, Mechanisms of tryptophan fluorescence shifts in proteins, Biophys. J. 80 (2001) 2093–2109, https://doi.org/10.1016/S0006-3495(01)76183-8.

[61] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Third Edit, Springer Science, 2006.

[62] M. Noronha, J.C. Lima, P. Lamosa, H. Santos, C. Maycock, R. Ventura, A. L. Maçanita, Intramolecular fluorescence quenching of tyrosine by the peptide α-carbonyl group revisited, J. Phys. Chem. A 108 (2004) 2155–2166, https://doi.org/10.1021/jp037125l.

[63] K.B. Davis, Z. Zhang, E.A. Karpova, J. Zhang, Application of tyrosine-tryptophan fluorescence resonance energy transfer in monitoring protein size changes, Anal. Biochem. 557 (2018) 142–150, https://doi.org/10.1016/j.ab.2018.07.022.

[64] K.B. Davis, Z. Zhang, E.A. Karpova, J. Zhang, Application of tyrosine-tryptophan fluorescence resonance energy transfer in monitoring protein size changes, Anal. Biochem. 557 (2018) 142–150, https://doi.org/10.1016/j.ab.2018.07.022.

[65] A.B. Biter, J. Pollet, W.H. Chen, U. Strych, P.J. Hotez, M.E. Bottazzi, A method to probe protein structure from UV absorbance spectra, Anal. Biochem. 587 (2019), https://doi.org/10.1016/j.ab.2019.113450.

[66] D. Ghosh, P.K. Singh, S. Sahay, N.N. Jha, R.S. Jacob, S. Sen, A. Kumar, R. Riek, S. K. Maji, Structure based aggregation studies reveal the presence of helix-rich intermediate during α-synuclein aggregation, Sci. Rep. 5 (2015), https://doi.org/10.1038/srep09228.

[67] M. Noronha, J.C. Lima, P. Lamosa, H. Santos, C. Maycock, R. Ventura, A. L. Maçanita, Intramolecular fluorescence quenching of tyrosine by the peptide α-carbonyl group revisited, J. Phys. Chem. A 108 (2004) 2155–2166, https://doi.org/10.1021/jp037125l.

[68] E. Wieczorek, P. Bezara, A. Ożyhar, Deep blue autofluorescence reveals the instability of human transthyretin, Int. J. Biol. Macromol. 191 (2021) 492–499, https://doi.org/10.1016/j.ijbiomac.2021.09.102.

[69] M. Ziaunys, T. Sneideris, V. Smirnovas, Exploring the potential of deep-blue autofluorescence for monitoring amyloid fibril formation and dissociation, PeerJ (2019), https://doi.org/10.7717/peerj.7554.

[70] C.S.H. Jesus, H.T. Soares, A.P. Piedade, L. Cortes, C. Serpa, Using amyloid autofluorescence as a biomarker for lysozyme aggregation inhibition, Analyst 146 (2021) 2383–2391, https://doi.org/10.1039/d0an02260h.

[71] C. Niyangoda, T. Miti, L. Breydo, V. Uversky, M. Muschol, Carbonyl-based blue autofluorescence of proteins and amino acids, PLoS ONE. 12 (2017), https://doi.org/10.1371/journal.pone.0176983.

[72] T.N. Tikhonova, N.R. Rovnyagina, A.Y. Zherebker, N.N. Sluchanko, A.A. Rubekina, A.S. Orekhov, E.N. Nikolaev, V.V. Fadeev, V.N. Uversky, E.A. Shirshin, Dissection of the deep-blue autofluorescence changes accompanying amyloid fibrillation, Arch. Biochem. Biophys. 651 (2018) 13–20, https://doi.org/10.1016/j.abb.2018.05.019.

[73] M.A. Saraiva, Interpretation of α-synuclein UV absorption spectra in the peptide bond and the aromatic regions, J. Photochem. Photobiol. B Biol. 212 (2020), https://doi.org/10.1016/j.jphotobiol.2020.112022.

[74] C. Niyangoda, T. Miti, L. Breydo, V. Uversky, M. Muschol, Carbonyl-based blue autofluorescence of proteins and amino acids, PLoS ONE 12 (2017), https://doi.org/10.1371/journal.pone.0176983.

[75] M. Biancalana, S. Koide, Molecular mechanism of Thioflavin-T binding to amyloid fibrils, Biochim. Biophys. Acta, Proteins Proteomics 1804 (2010) 1405–1412, https://doi.org/10.1016/j.bbapap.2010.04.001.

[76] C. Xue, T.Y. Lin, D. Chang, Z. Guo, Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation, R. Soc. Open Sci. 4 (2017), https://doi.org/10.1098/rsos.160696.

[77] E.I. Yakupova, L.G. Bobyleva, I.M. Vikhlyantsev, A.G. Bobylev, Congo red and amyloids: history and relationship, Biosci. Rep. 39 (2019), https://doi.org/10.1042/BSR20181415.

[78] A.I. Bush, Copper, zinc, and the metallobiology of Alzheimer disease, Alzheimer Dis. Assoc. Disord. 17 (2003) 147–150.

[79] L. Breydo, V.N. Uversky, Role of metal ions in aggregation of intrinsically disordered proteins in neurodegenerative diseases, Metallomics 3 (2011) 1163–1180, https://doi.org/10.1039/c1mt00106j.

[80] S.L. Gras, L.J. Waddington, K.N. Goldie, Transmission electron microscopy of amyloid fibrils, Methods Mol. Biol. 752 (2011) 197–214, https://doi.org/10.1007/978-1-60327-223-0_13.

[81] S. Xu, K.R. Brunden, J.Q. Trojanowski, V.M.Y. Lee, Characterization of tau fibrillization in vitro, Alzheimers Dement. 6 (2010) 110–117, https://doi.org/10.1016/j.jalz.2009.06.002.

[82] D. Willbold, B. Strodel, G.F. Schröder, W. Hoyer, H. Heise, Amyloid-type protein aggregation and prion-like properties of amyloids, Chem. Rev. 121 (2021) 8285–8307, https://doi.org/10.1021/acs.chemrev.1c00196.

[83] W. Close, M. Neumann, A. Schmidt, M. Hora, K. Annamalai, M. Schmidt, B. Reif, V. Schmidt, N. Grigorieff, M. Fändrich, Physical basis of amyloid fibril polymorphism, Nat. Commun. 9 (2018), https://doi.org/10.1038/s41467-018-03164-5.

[84] S. Wegmann, J.J. Yu, S. Chinnathambi, E.M. Mandelkow, E. Mandelkow, D. J. Muller, Human tau isoforms assemble into ribbon-like fibrils that display polymorphic structure and stability, J. Biol. Chem. 285 (2010) 27302–27313, https://doi.org/10.1074/jbc.M110.145318.

[85] D. Kurouski, Supramolecular organization of amyloid fibrils, in: Exploring New Findings on Amyloidosis, 2016, pp. 73–98, https://doi.org/10.5772/62672.

[86] S. Maeda, N. Sahara, Y. Saito, M. Murayama, Y. Yoshiike, H. Kim, T. Miyasaka, S. Murayama, A. Ikai, A. Takashima, Granular tau oligomers as intermediates of tau filaments, Biochemistry 46 (2007) 3856–3861, https://doi.org/10.1021/bi061359o.

[87] R. Rauch, M. Hofbeck, C. Zweier, A. Koch, S. Zink, U. Trautmann, J. Hoyer, R. Kaulitz, H. Singer, A. Rauch, Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of fallot, J. Med. Genet. 47 (2010) 321–331, https://doi.org/10.1136/jmg.2009.070391.

[88] Y.W. Sha, X. Xu, L. Bin Mei, P. Li, Z.Y. Su, X.Q. He, L. Li, A homozygous CEP135 mutation is associated with multiple morphological abnormalities of the sperm flagella (MMAF), Gene 633 (2017) 48–53, https://doi.org/10.1016/j.gene.2017.08.033.