The N-terminal of NBPF15 causes multiple types of aggregates and mediates phase transition

Han Wu, Liu-Tao Zhai, Xue-Xue Guo, Stéphane Réty, Xu-Guang Xi

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

PORTLAND PRESS

Research Article

The N-terminal of NBPF15 causes multiple types of aggregates and mediates phase transition

Han Wu[1], Liu-Tao Zhai[1], Xue-Xue Guo[1], Stephane Rety[2] and Xu-Guang Xi[1,3]

[1]The State Key Laboratory of Crop Stress Biology in Arid Areas, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, PR China; [2]The University Lyon, ENS de Lyon, University Claude Bernard, CNRS UMR 5239, INSERM U1210, LBMC, 46 Allée d'Italie Site Jacques Monod, F-69007, Lyon, France; [3]The LBPA, Ecole Normale Supérieure Paris-Saclay, CNRS, Université Paris Saclay, 61 Avenue du Président Wilson, F-94235 Cachan, France

Correspondence: Xu-Guang Xi (xxi01@ens-cachan.fr)

The neuroblastoma breakpoint family (*NBPF*) consists of 24 members that play an important role in neuroblastoma and other cancers. *NBPF* is an evolutionarily recent gene family that encodes several repeats of Olduvai domain and an abundant N-terminal region. The function and biochemical properties of both Olduvai domain and the N-terminal region remain enigmatic. Human *NBPF15* encodes a 670 AA protein consisting of six clades of Olduvai domains. In this study, we synthesized and expressed full-length NBPF15, and purified a range of NBPF15 truncations which were analyzed using dynamic light scattering (DLS), superdex200 (S200), small-angle X-ray scattering (SAXS), far-UV circular dichroism (CD) spectroscopy, transmission electron microscope (TEM), and crystallography. We found that proteins containing both the N-terminal region and Olduvai domain are heterogeneous with multiple types of aggregates, and some of them underwent a liquid-to-solid phase transition, probably because of the entanglement within the N-terminal coiled-coil. Proteins that contain only the Olduvai domain are homogeneous extended monomers, and those with the conserved clade 1 (CON1) have manifested a tendency to crystallize. We suggest that the entanglements between the mosaic disorder-ordered segments in NBPF15 N terminus have triggered the multiple types of aggregates and phase transition of NBPF15 proteins, which could be associated with Olduvai-related cognitive dysfunction diseases.

Introduction

The *NBPF* gene family mostly located in three regions of chromosome 1, namely, 1p36, 1p12, and 1q21. The sequences of 1p36 and 1p12 were unchanged in the most recent human genome reference sequence, while numerous changes occurred on the long arm of chromosome 1 [1]. The 1q21 region encompasses the largest cluster of *NBPF* genes, containing 11 *NBPF* members (from *NBPF8* to *NBPF18p*) [2,3]. Only *NBPF* genes within this region had undergone evolutionary hyper-amplification [4].

Previous studies have shown that NBPF1 can interact with tumor suppressor, chibby and clusterin, to form a tri-molecular complex [5] and exerts different tumor-suppressive effects dependent on the cell line analysis in which it is expressed [6]. NBPF7 plays a key role in α-catenin signaling, which regulates keratinocyte proliferation through functional interaction with NF-κB [7]. *NBPF12* located in the proximal region of 1q21, and in a patient harboring an 889-kb microdeletion of this region, a delay in brain and body development was observed [8]. NBPF14 is highly expressed in both skin and ovary tissue and is a biomarker of whether women with bipolar disorders respond to lithium [9]. It is predicted that the majority of NBPF proteins contain similar N-terminal regions to transcription factor stat-1/DNA complex or stat3b/DNA complex, and NBPF may, therefore, function as a DNA-binding protein [10]. Despite the important roles of *NBPF* genes in physiological and pathological processes, the regulation of this gene family and their specific cellular functions and protein properties remain unknown.

Received: 7 August 2019
Revised: 2 December 2019
Accepted: 5 December 2019

Accepted Manuscript online: 6 December 2019
Version of Record published: 31 January 2020

Biochemical Journal (2020) 477 445–458
https://doi.org/10.1042/BCJ20190566

The *NBPF* gene family is predicted to encode a novel protein domain termed the NBPF repeat, which comprised a PFAM protein DUF1220, this repeat is referred as to NBPF/DUF1220 repeat. In 2017, NBPF/DUF1220 domain renamed to Olduvai domain [11]. This domain is built of two exons and is present in multiple copies of different NBPF proteins. Olduvai domain is ∼65 amino acids in length and expressed in several human tissues, including the brain, particularly in regions associated with higher cognitive function [12]. It can be subdivided into six subtypes or clades, based on sequence similarity [13,14]: three conserved clades (CON1–3) and three human lineages specific clades (HLS1–3). Each CON clade maintains a distinct position within NBPFs. CON1 is present at the N-terminus of all NBPFs, while CON2 is adjacent to CON1. CON3 resides at the C-terminus of the great majority of NBPFs. HLS clades typically fall between CON2 and CON3 in the order of HLS1–HLS2–HLS3. The highest copy number of Olduvai domain is found in humans: 302. The copy numbers in the other species include 97–138 in great apes, 48–75 in monkeys, 1–9 in all other mammals, and none outside of mammals, show signs of positive selection [13,15].

Specific Olduvai copy number variations correlate with specific cognitive phenotype. These include autism severity with CON1 [16], and schizophrenia severity with both CON1 and HLS1 [17], and cognitive aptitude improvement with CON2 [18]. Recurrent reciprocal deletions and duplications in the 1q21.1 region are associated with microcephaly or macrocephaly and developmental and behavioral abnormalities [19]. DUF1220 copy numbers are related to brain size and have implicated in normal and pathological variation in human brain gray matter volume and neuron number, functioning as drivers of neural stem cell proliferation across all primate lineages [20,21]. It has been proposed that a rapidly increasing Olduvai copy number results in high genomic instability in the 1q21 region, which in turn, precipitates in numerous recurrent human developmental and neuro-genetic diseases [22,23]. Copy number amplifications of Olduvai domain have been connected to a range of diseases that highlights its important role in human evolution and development. Naturally, the *NBPF* gene family, which encoded Olduvai, has received intense research attention.

To gain insight into the structure and the characteristic of NBPF or Olduvai, we have expressed and purified a series of truncated human NBPF15 proteins that contain different Olduvai domains and analyzed their biochemical properties. Besides, we have compared the characteristics of the NBPF15 N-terminal region with the Olduvai domain to examine their impacts.

Materials and methods

Plasmid construction, protein expression, and purification of NBPF15 proteins

We have searched the full-length human NBPF15 sequence in GeneBank (AAH23087.0), and synthesized its nucleotide sequence with restriction sites at both ends in General Biosystems. Full-length *NBPF15* was first cloned into the pET-15b-SUMO and the pET-21a-HMT vector between the *Eco*RI and *Xho*I sites, with a six his-tag followed by a SUMO and TEV cleavage site, respectively. Then, by taking the positive pET-15b-SUMO-NBPF15 as the template, the truncated genes were amplified by the corresponding forward and reverse primers and constructed into pET-15b-SUMO and pET-21a-HMT vector with conventional restriction digestion and ligation methods. NBPF15 truncations were constructed based on protein annotation from UniProt (https://www.uniprot.org/) and the predicted secondary and tertiary structures using NPS@ (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html), phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [24] and genome3D (http://www.genome3d.eu) predictors. The criterion for protein truncation is to preserve the integrity of the Pfam domain and retain more ordered regions. Different vectors were applied to find out the optimal expression vector for a higher expression quantity and purer protein. The pET-21a-HMT vector is suitable for protein with small molecular mass, and the pET-15b-SUMO vector is much recommendable for larger protein, but it is not conclusively.

Plasmids were transformed into *Escherichia coli* 2566 (DE3) cells, and a single transformant was amplified at 37°C in Luria-Bertani (LB) culture media containing 100 μg/ml ampicillin until the OD_{600} reached 0.6–0.8. Put the cultures in ice water for rapid cooling and added 0.3 mM IPTG into cooled cultures, which would cultivate for a further 15 h at 18°C. Cells were lysed using ultra-high pressure cell disrupter (JNBIO) and sonicated three times in 7 s interval −3 s ultrasonic −3 min total time mode in ice water in lysis buffer (300 mM NaCl, 20 mM Tris–HCl pH 7.5, 5% glycerol), followed by centrifugation at 12 000 rpm for 0.5–1 h and filtered the supernatant through a 0.45 μm filter before loading on the first column. For protein expressed in the pET-15b-SUMO vector, used the Ni-NTA column as the first column, and the binding buffer was composed of lysis buffer plus 5 mM imidazole. After protein was loaded, 10 column volumes (CVs) binding buffer were

applied to ensure the target molecules interact effectively with the ligand and are retained by the affinity medium as all other molecules wash through the column. Proteins were eluted from the Ni^{2+} affinity columns using Ni elution buffer (300 mM imidazole, 300 mM NaCl, 20 mM Tris–HCl pH 7.5, 5% glycerol) and digested with SUMO protease overnight at 4°C to remove the SUMO-tag. For protein expressed in the pET-21a-HMT vector, the amylose column was first needed. The component of the binding buffer was the same as the lysis buffer, and 20 CVs binding buffer was applied. Target protein was eluted with maltose buffer (0.5% w/v maltose, 300 mM NaCl, 20 mM Tris–HCl pH 7.5, 5% glycerol) and digested with TEV protease to remove MBP-tag. The SUMO- or TEV-digested proteins were reloaded on Hitrap-Q (GE Healthcare) and eluted in a 100 to 1000 mM NaCl gradient in the AKTA purifier in buffer Q (20 mM Tris–HCl pH 7.5, 5% glycerol, 2 mM EDTA, 2 mM DTT). Eluted fractions containing the target proteins were collected, concentrated and analyzed through SDS–PAGE electrophoresis to identify its purity. Concentrated protein was lastly loaded on S200 16/600 (GE Healthcare, 120 ml) and eluted with S200 buffer (500 mM NaCl, 20 mM Tris–HCl pH 7.5, 5% glycerol, 2 mM EDTA, 2 mM DTT) at a rate of 0.8 ml/min. Eluted proteins were concentrated through ultrafiltration in a centrifugal tube (MWCO 10 kDa, Millipore) and stored at −80°C until use.

Baculovirus/insect cell expression system is a eukaryotic expression system that thought could perform extensive post-translational modifications in a recombinant NBPF15 protein, which might helpful to the expression of full-length NBPF15 *in vitro*. The full-length NBPF15 sequence was also cloned into the pTrEx1.1 vector (Novagen, U.S.A.) and expressed in Professor Chen's laboratory according to the methods as described in [25]. Vector pTrEx1.1 is necessary for the expression of full-length NBPF15 in the baculovirus/insect cell expression system because it has a P10 eukaryotic promoter.

Site-directed mutagenesis

Deletions or site-directed point mutations of $NBPF15^{9-263}$ related to protein aggregation analysis were introduced into DNA fragments corresponding to $NBPF15^{9-263}$ through PCR using corresponded primers. Then, clone the fragments into the pET-21a-HMT vector. DNA sequencing verified the introduction of the desired mutations and confirmed that no undesired mutation was present. Expression and purification of the mutant $NBPF15^{9-263}$ proteins are similar to the wild type $NBPF15^{9-263}$.

Analytical size exclusion chromatography (SEC)

The SEC was performed on a 10/300 S200 analytical grade column using a fast protein liquid chromatography system at a constant temperature of 20°C. Equilibrated the S200 column with the S200 buffer at a flow rate of 0.3 ml/min. Approximately 100 μg of the purified protein was loaded onto the column and eluted in the same buffer and monitored the ultraviolet absorption at 280 nm. The S200 column was calibrated with chromatography profiles of standard proteins and fitted to standard curves. The log $(Mw) = -3.1929x + 7.5105$ ($R^2 = 0.9936$) was used to calculate the corresponding molecular mass of NBPF15 protein, where x represents V_e/V_0, and V_0 equals 8.3 ml (Supplementary Figure S6).

Dynamic light scattering (DLS)

DLS measurements were performed at 20°C using the DynaPro NanoStar instrument (Wyatt Technology Corporation) with a thermostat cell holder. DLS buffer (200 mM NaCl, 20 mM Tris–HCl pH 7.5, 5% glycerol, 2 mM DTT) and samples in a series of concentrations (each sample volume was 50 μl) were prepared and centrifuged for 30 min at 13 000×***g*** at 4°C. Collected the measurements for sample acquisition, typically for 5 s. A measurement period of 50 s — 10 acquisitions, 5 s each, were applied. Data were analyzed using Dynamics7.0 software with regularization methods. Regularization analysis produced an estimate of the radii and relative abundance of all species present in the solution without assuming an underlying distribution. According to the DYNAMICS user's guide [26], the Mw-R (molar mass estimated from the measured radii) was calculated using the empirical formula: $Mw = (1.68 \times R)^{2.34}$, where R represents the measured radii.

Far-UV CD spectroscopy

Far-UV CD measurements were performed on a MOS-450/AF-CD (Bio-logic, French) using a 0.1 cm path length cuvette. Diluted the proteins to 30 μM in 20 mM sodium phosphate buffer, pH 7.5. CD spectra were recorded from 260 to 190 nm with a scanning speed of 20 nm/min and 10 accumulations per scan at 20°C. Spectra were baseline corrected using the buffer.

Biochemical Journal (2020) 477 445–458
https://doi.org/10.1042/BCJ20190566

Analysis of the truncated NBPF15 proteins

NBPF15 truncations were analyzed on a segmented 4, 6, 8, 10, 12% acrylamide gradient TAE gel to observe the separation of different types of NBPF15 aggregates. The step gradient gel is a simplified version of a linear gradient gel that can be easily prepared without using additional instrumentation. First, 4% and 12% polyacrylamide resolution (19 : 1 acrylamide/ bis-acrylamide, 1× TAE, 0.1% w/v ammonium persulfate, 0.1% v/v tetramethylethylenediamine) were separately prepared. Then, mixed above two solutions in proportion to the final concentration of 4, 6, 8, 10, 12% in five EP tubes. Pipetted the same volume of the fully mixed solution one by one into the gaps between the glass plates, with the polyacrylamide concentration from largest to smallest, and insert the comb straight on down. The gel will polymerize in 30 min. The loading buffer consisted of 1× TAE and 50% glycerol. The gel is run at 4°C in 15 V/cm for 1 h with 1× TAE filled with the inner and outer chambers. 1× TAE contained 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA. After electrophoresis, stained the gels by coomassie brilliant blue and/or silver nitrate. Different stain methods were used to analyze the results more accurately. Short silver nitrate staining of proteins in both SDS–PAGE and TAE Native-PAGE gels were based on [27].

For protein depolymerization analysis, ∼10 μg NBPF15 protein was treated with indicated concentrations of DTT or urea. The reaction system consisted of a fixed amount of NBPF15 protein and varying volume of 1 M DTT or 8 M urea, fill the insufficient volume up to 10 μl with buffer (300 mM NaCl, 20 mM Tris–HCl pH 7.5, 5% glycerol) and performed SDS–PAGE and/or TAE-Native PAGE after incubation for 30 min at 4°C.

Small-angle X-ray scattering (SAXS)

SAXS experiment was carried out at 20°C with size-exclusion chromatography HPLC coupled to SAXS (SEC-SAXS) data collection at beamline BL19U2 of the National Center for Protein Sciences Shanghai (China). The samples of protein at a concentration of 10 mg/ml were injected on the BioSEC3 HPLC column and measured in SAXS buffer (20 mM Tris–HCl, pH 7.5, 300 mM NaCl, 5% glycerol). Scattering data collection and processing were performed with the program BL19U2 pipeline available on the beamlines. Additional processing of the SEC-SAXS HPLC profile was done with the specific module US-SOMO of U1-trascan2 software [28]. The radius of gyration (Rg) and maximum particle dimension (Dmax) were extracted using the PRIMUS and GNOM programs [29], and further analysis was carried out according to [30].

Transmission Electron Microscopy (TEM) and Differential Interferences Contrast Microscopy (DIC)

Samples were prepared in 100 mM NaCl, 20 mM Tris–HCl pH 7.5, 2 mM DTT buffer, and bound to the copper grid through treatment with glow discharge apparatus. Grids were washed twice with ddH_2O and negatively stained with 1% uranyl acetate. Specimens were evaluated by TEM (FEI, TECNAI G2 SPIRIT BIO).

Protein droplets were injected into a homemade flow chamber comprising a glass slide and sandwiched using a coverslip with one layer of double-sided tape as a spacer for DIC (LECIA DM6 B).

Protein crystallization

Round crystals of $NBPF15^{135–263}$ grew after 3 days in a well solution consisting of 2 M sodium formate and 0.1 M Bis-Tris propane pH 7.0 in SaltRx (Hampton Research). Microcrystals of $NBPF15^{167–263}$ were grown through the sitting drop method using commercial screening kits PACT with 0.2 M NaF, 20% (w/v) PEG3350 (Hampton Research). Aliquots of 20 mg/ml of both $NBPF15^{135–263}$ and $NBPF15^{167–263}$ in 100 mM NaCl, 20 mM Tris–HCl pH 7.5 were mixed with reservoir solution and gown at 20°C.

Results

NBPF15 N-terminal region mediates protein self-aggregation

At first, we aimed to obtain the structure of NBPF15 by crystallization. Initial efforts to purify full- length NBPF15 in both prokaryotic *Escherichia coli* and baculovirus/insect cell expression systems were unsuccessful (data not shown). Expression of NBPF15 truncated proteins containing more than one Olduvai domain in the *Escherichia coli* system was also a failed attempt (Supplementary Figure S1). The N-terminus of NBPF15 had a predicted coiled-coil in sequence 70 to 130 (Figure 1A), which was thought can increase the likelihood of protein crystallization. Therefore, a series of NBPF15 truncations with varying lengths of the N-terminal region (Figure 1B) were constructed based on protein annotation predictors, as described in the 'Method' part

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

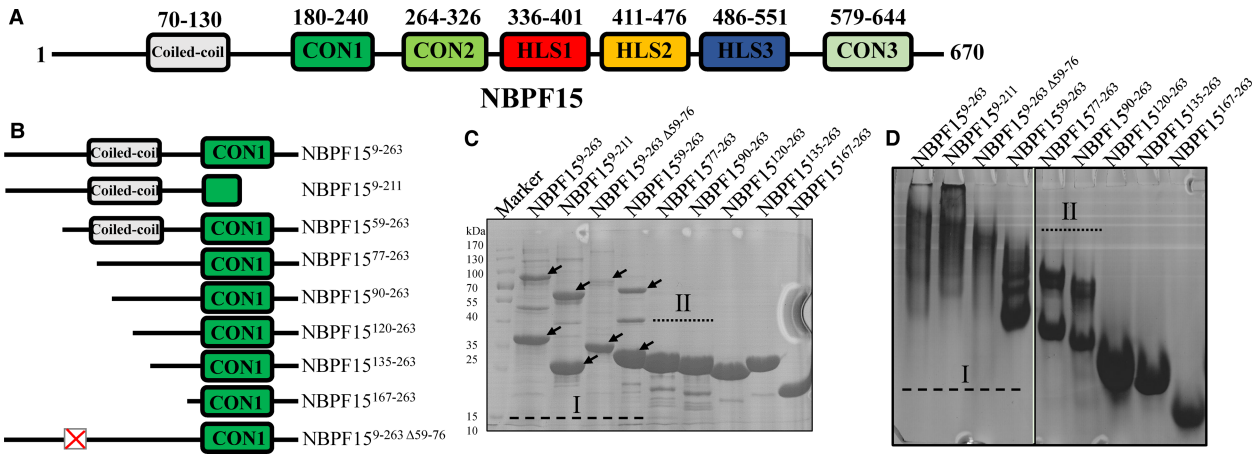

Figure 1. Different aggregated forms of NBPF15 truncations.

(**A** and **B**) Schematic representation of full-length NBPF15 and nine truncated proteins with different length N-terminal. (**C**) 12% SDS–PAGE gels of the nine truncations. The pre-stained protein marker is shown on the left lane, and different aggregated forms of category I proteins were marked with arrows. (**D**) 4%–12% gradient TAE Native PAGE gels of the nine truncations.

(Supplementary Figure S2A–C). The proteins that contained the CON1 domain were successfully purified, while proteins lacked the CON1 domain failed purification trials due to low expression and/or protein insolubility (Supplementary Figure S2D). These differences suggest that CON1 contributes to NBPF15 N-terminal protein expression and aids the folding of NBPF15 proteins. The theoretical molecular mass and purification process of purified proteins were summarized in Supplementary Table S1. We found that successfully purified N-terminal-included proteins had more than one aggregated form, and were subdivided based on their aggregation properties.

Category I proteins exist varieties of aggregated forms

$NBPF15^{9–263}$, $NBPF15^{9–211}$, and $NBPF15^{59–263}$ belong to category I. Proteins in the category I possess almost a complete N-terminal region and exist in a variety of aggregated forms on SDS–PAGE gel (Figure 1C and Supplementary Figure S3A) and on TAE Native PAGE (Figure 1D and Supplementary Figure S3A). The multiple types of aggregates appeared in the first step of the purification process (Supplementary Figure S2D).

When treated with reductant or denaturant, like DTT or urea, parts of the high-order aggregates of $NBPF15^{9–263}$ and $NBPF15^{9–211}$ could be depolymerized to low-order aggregates (Figure 2A,B), suggesting this part of the high-order aggregates were formed by disulfide bonds. The rest of the high-order aggregates could not be depolymerized even treated with high concentration DTT or urea, which means other molecular forces were involved. No transformation occurred on $NBPF15^{59–263}$ between the high-order aggregates and the low-order aggregates in response to DTT and urea (Figure 2C). Due to the formation of disulfide bonds, the cysteine location of three truncations was analyzed and found that the difference in cysteine location amongst the three proteins was at cys42, which was truncated off in $NBPF15^{59–263}$. The missing cys42 could explain the unresponsiveness of the $NBPF15^{59–263}$ to DTT. To verify this, cys42 was mutated to alanine (C42A), using the identical methods and conditions to the wild-type $NBPF15^{9–263}$. When the mutant protein was purified to near homogeneity and treated with DTT and urea, no depolymerization occurred (Supplementary Figure S3B). The site-mutation result confirmed that a certain percentage of the high-order aggregates of $NBPF15^{9–263}$ were formed due to disulfide bonds at cys42.

DLS was performed to investigate the protein species in solution and their respective sizes. Size distribution histogram of the category I proteins showed more than one peak, and the majority size distribution peak (high % mass peak) comprised of specific aggregates [26]. For instance, the specific aggregates of $NBPF15^{9–263}$ were the radii of 12.43 nm particles with 84.6% mass. The minority peaks (low% mass peak) were 11.7% mass with the radii of 46.12 nm, and 3.7% mass with the radii of 171.38 nm, corresponding to two large particles comprised of non-specific aggregates, so it produced a much stronger light signal (high% intensity peak): 74.1% intensity and 9.2% intensity (Figure 2D and Supplementary Table S2). $NBPF15^{9–211}$ had the radii of 9.99 nm

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

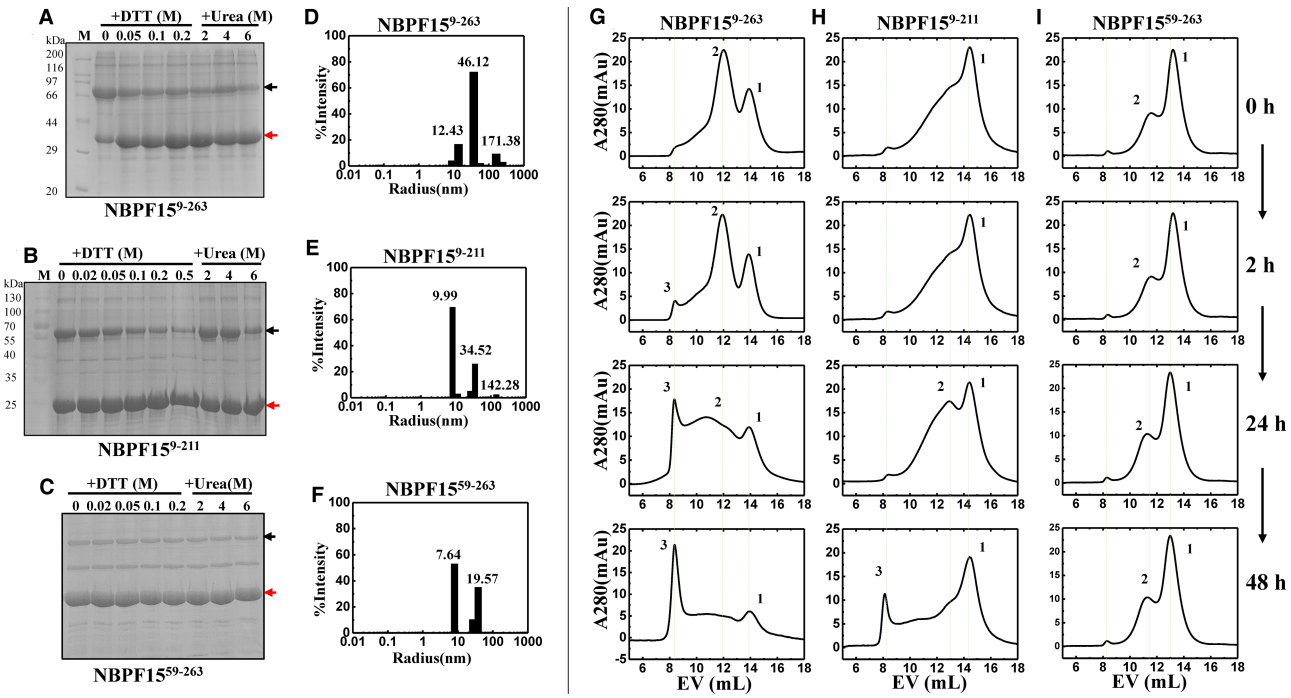

Figure 2. Category I proteins have multiple types of aggregated forms.

(**A–C**) 12% SDS–PAGE gels of DTT and urea treatment on category I proteins. High-order aggregates and low-order aggregates were indicated in the black and red arrows, respectively. (**D–F**) Size distribution histograms of category I proteins in DLS assays. (**G–I**) Self-aggregation of category I proteins in S200 10/300 GL over the placement time. From top to bottom, the elution curves of freshly purified proteins, and the curves of protein after placed on ice for 2, 24 and 48 h.

particles and possessed both high% mass and high% intensity (Figure 2E and Supplementary Table S2), suggesting these aggregates were highly specific and significant. NBPF15$^{59–263}$ existed two populations with mean radii of 7.64 and 19.57 nm (Figure 2F and Supplementary Table S2). The absence of a third population could be due to the low number of aggregates. Also, the peaks of specific aggregates of category I protein all <99.9% mass, suggesting they were unlikely to crystallize [26].

An interesting phenomenon was noted that NBPF15$^{9–263}$ and NBPF15$^{9–211}$ were under ceaselessly self-aggregation, which was monitored by gel filtration chromatography. Freshly purified NBPF15$^{9–263}$ eluted with two peaks in S200 (peak1 in 13.9 ml and peak 2 in 11.9 ml), and after 2 h, a third population (molecular mass >670 kDa) appeared around the bed volume (peak 3, 8.36 ml) (Figure 2G). The ratio of the third population increased over time, and after 48 h became predominant, which was accompanied by the complete disappearance of peak 2, highlighting the heterogeneity of NBPF15$^{9–263}$ (Figure 2G). Notably, peak 1 and 3 were not completely separated as the UV absorption curves were not horizontal, indicating a range of intermediate aggregates existed. A similar situation happened to NBPF15$^{9–211}$, whose high-order aggregates emerged after 24 h (Figure 2H). As for NBPF15$^{59–263}$, self-aggregation was less evident, though the corresponding high order aggregates did appeared around the bed volume yet did not increased after 48 h (Figure 2I), suggesting sequence 9 to 59 strongly promotes uninterrupted protein self-aggregation.

Category II proteins have two aggregated forms

Category II proteins contained NBPF15$^{77–263}$ and NBPF15$^{90–263}$, which had no high-order aggregates on SDS–PAGE during the initial step of the purification process (Supplementary Figure S2D). Though only one band was observed on SDS–PAGE, two individual bands appeared on TAE Native-PAGE (Figure 3A), and two peaks were observed on S200 (Figure 3B,C), suggesting the existence of two aggregated forms, which was consistent with the DLS histogram (Figure 3D,E). We found the additional high-order aggregates of NBPF15$^{77–263}$ on SDS–PAGE when free of DTT, which were also evident on TAE Native-PAGE (Figure 3A) revealing the existence of disulfide bonds.

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

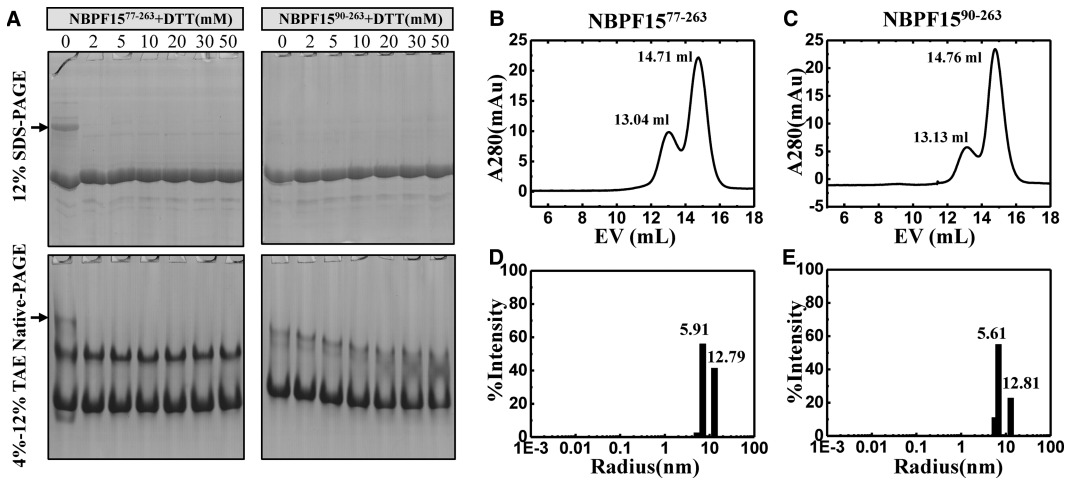

Figure 3. Basic properties of category II proteins NBPF15$^{77–263}$ and NBPF15$^{90–263}$.

(**A**) DTT treatment on category II proteins on 12% SDS–PAGE gels [top two] and on 4%–12% gradient TAE Native-PAGE gels [bottom two]. When buffer free of DTT, NBPF15$^{77–263}$ had a new population (arrows) that was visible in both SDS–PAGE gel and TAE Native-PAGE gel. (**B** and **C**) S200 10/300 GL elution curves and (**D** and **E**) DLS Size distribution histograms of NBPF15$^{77–263}$ and NBPF15$^{90–263}$, respectively.

Category II proteins did not undergo drastic changes in their S200 elution curves in the 48 h time interval. Their S200 elution curves vary with concentration. The concentration of 20 mg/ml NBPF15$^{77–263}$ leading to irregular, widely spanned S200 elution curves (Supplementary Figure S4A) and poor separation effects, as shown on the TAE Native PAGE (Supplementary Figure S4B). Following dilution to 5 mg/ml, the elution curves became regular with a shorter elution span, and the separation of the two aggregated forms became evident (Supplementary Figure S4C,D), indicating the presence of weak interaction between the molecules of NBPF15$^{77–263}$ proteins. The situation was the same for NBPF15$^{90–263}$ (data not shown). The initial elution volume of NBPF15$^{77–263}$ remains unchanged at high protein concentration, and no third population was observed on TAE Native PAGE, suggesting the cessation of self-aggregation.

The differences in the existence of high-order aggregates between NBPF15$^{59–263}$ and NBPF15$^{77–263}$ (Figure 1C,D and Supplementary Figure S3A) leading to the construction and purification of a mutant with deletion of amino acid 59–76, termed as NBPF15$^{9–263\Delta59–76}$, to investigate the function of sequence 59 to 76. The percentage of the high-order aggregates of NBPF15$^{9–263\Delta59–76}$ was indeed declined compared with NBPF15$^{9–263}$ but were not wholly absent following DTT or urea treatment (Supplementary Figure S3C), suggesting that sequence 9 to 76 contributed to the formation of high-order aggregates. The characteristic of N-terminal proteins being with multiple types of aggregates was gradually unapparent with longer N-terminal truncated, and was finally homogeneous until the truncation NBPF15$^{120–263}$, suggesting that sequence 9 to 119 mediates to the formation of multiple types of aggregates but was against to protein crystallization.

Proteins containing single Olduvai domain are monomers with high homogeneity

Three proteins, NBPF15$^{120–263}$, NBPF15$^{135–263}$, and NBPF15$^{167–263}$, were purified in high purities and homogeneities on both SDS–PAGE and TAE Native PAGE (Figure 1C,D and Supplementary Figure S3A). To determine their oligomeric state, we have performed the gel filtration chromatography, DLS, and SEC-SAXS. Their theoretical molecular mass were 16, 14, and 11 kDa, respectively, as calculated by Vector NTI (Supplementary Table S1). Their elution curves on S200 followed a Gaussian distribution with apparent molecular mass of 53, 47, and 40 kDa, respectively (Figure 4A and Supplementary Figure S5). The size distribution histograms of DLS were monomodal monodispersed (Figure 4B–D), with calculated radii of 4.15, 4.08, and 2.89 nm, indicating that these proteins might form trimers of high homogeneity. Two proteins containing HLS1 and HLS3 subtype were also purified, respectively, named NBPF15$^{334–414}$ and NBPF15$^{455–576}$. They shared the same characteristics of the proteins containing single CON1 and were like trimeric, based on the results of S200 and DLS

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

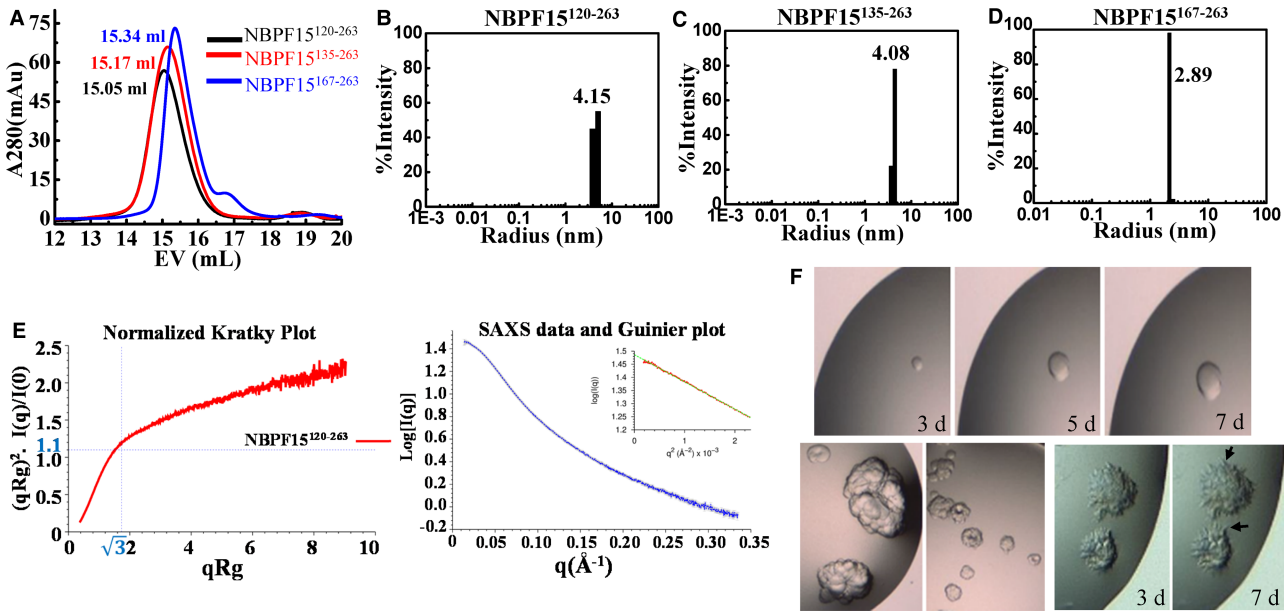

Figure 4. Basic properties of proteins that contained CON1 and barely any N terminus.

(**A**) SEC of NBPF15$^{120–263}$, NBPF15$^{135–263}$, and NBPF15$^{167–263}$. (**B–D**) DLS size distribution histograms of the three NBPF15 truncations. The name and mean radii size of the protein is labeled. (**E**) Normalized Kratky plot of NBPF15$^{120–263}$ [left], SAXS scattering profile of NBPF15$^{120–263}$ generated from the buffer-subtracted SEC-SAXS profile and a Guinier plot (inset) to calculate the radius of gyration [Right]. (**F**) Round crystals of NBPF15$^{135–263}$ grow larger over time [upper three]. The other two shapes of NBPF15$^{135–263}$ accumulated through crystallography [under two]. Deposits of NBPF15$^{167–263}$ [right two] with irregular edges acquired by initial screening after 3 days. Tentaculiform crystals (arrows) reached their maximum after 7 days.

(Supplementary Figure S6A–C). The oligomeric states of NBPF15$^{120–263}$, NBPF15$^{334–414}$, and NBPF15$^{455–576}$ which containing CON1, HLS1, and HLS3 subtype respectively, were further tested by SEC-SAXS. SAXS has long been used in protein folding studies and can decipher protein structural and dynamic features [31]. Take the data of NBPF15$^{120–263}$ as an example (Supplementary Table S3), the normalized Kratky plot clearly shows that the protein is unfolded (Figure 4E). For a globular folded protein, the curve is bell-shaped. The radius of gyration calculated with the Guinier plot is 39.6 Å, which is a high value for a very extended protein. The molecular mass calculated from the volume of the correlation method is 21 kDa. So, from SAXS, NBPF15$^{120–263}$ is a monomeric unfolded peptide, and the stokes radius can increase. This increase in hydrodynamic radius makes NBPF15$^{120–263}$ appear bigger with S200 and DLS techniques. The same results happened to NBPF15$^{334–414}$ and NBPF15$^{455–576}$ (Supplementary Figure S6D and Supplementary Table S3), suggesting that these proteins are all extended monomers.

We have screened all five proteins for crystallization considering their homogeneity. NBPF15$^{135–263}$ grew into the round, and the multi-layered crystals (Figure 4F) and NBPF15$^{167–263}$ grew tentaculiform crystals (Figure 4F), regrettably without diffraction points. Further optimization and protein modifications, including reductive methylation, ethylation, and isopropylation, were carried out but unsuccessful.

The N-terminal of NBPF15 was a mosaic of different (dis) ordered segments

Proteins with the less disordered secondary structures are conducive to crystallization. For NBPF15, longer assumptive helix-rich N-terminal region hindered protein crystallization because of the emergence of multiple types of aggregates and the tendency to self-aggregate. There are three cysteines in sequence 77 to 263, positioned at cys152, cys207, and cys211. Proteins including NBPF15$^{77–263}$, NBPF15$^{90–263}$, NBPF15$^{120–263}$, and NBPF15$^{135–263}$ have the same cysteine number in the same place, but their aggregated forms vary differently, suggesting that the dominating reason is not the specific cysteines but could be the whole conformational change. The specific mechanisms governing the self-aggregation of the N-terminal region remain undefined. Possible reasons include the intra- and/or inter-helical interactions within the coiled-coil (AA 70–130), or the sequence-intrinsic preference for conformational heterogeneity.

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

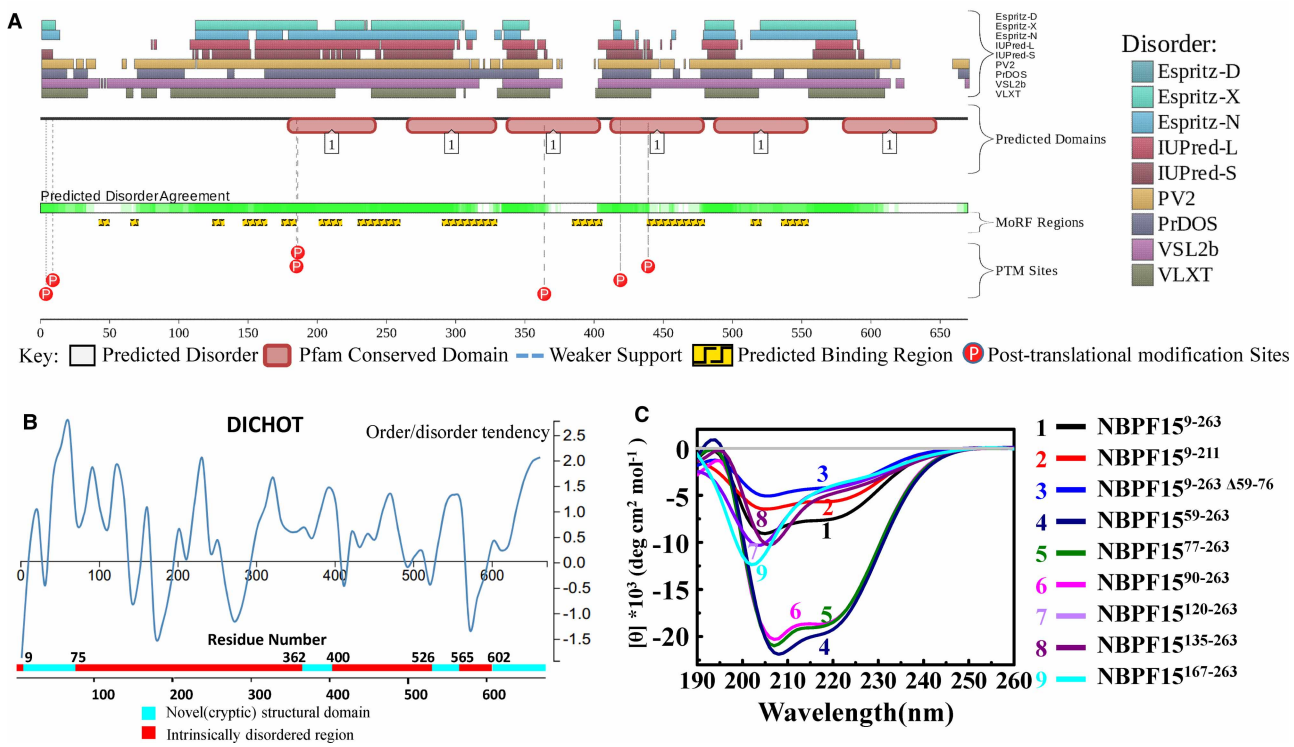

Figure 5. The N-terminal of NBPF was a mosaic of different (dis) ordered segments.

Prediction of the intrinsically disordered regions in NBPF15 by (**A**) D^2P^2, all disorder predictions (pastel-colored blocks) are stacked and aligned against the polypeptide chain in black. And by (**B**) DICHOT, the order/disorder probability for each of the amino acid sites was calculated using support vector machine (SVM). (**C**) Far-UV CD spectroscopy of nine NBPF15 truncations.

Altogether ten predictors from D^2P^2 (http://d2p2.pro/) [32] and DICHOT (http://www.ideal.force.cs.is.nagoyau.ac.jp/dichot/) [33] showed that NBPF15 is present in numerous disordered region (Figure 5A,B). In Figure 5A, below the disorder agreement line, some binding region predictions are displayed (yellow blocks with zigzag infill), yet NBPF15 shows no affinity to DNA or RNA in our experiment, and no research has ever reported the interaction between NBPF15 and other proteins.

To gain insight into the structural properties of the NBPF15 N-terminal region that caused multiple kinds of aggregates, we performed far UV-CD spectroscopy. The secondary structure of the truncations was estimated through online CD analysis and plotting tool (CAPITO, http://capito.nmr.fli-leibniz.de.) [34], and the calculated percentage of secondary structure was listed in Supplementary Table S4. The spectrums of the truncated proteins exhibit two different shapes (Figure 5C). CD spectra of proteins that contained N-terminus (which include all the category I and II proteins), recorded at 20°C, has a broad negative signal in 208 and 222 nm, were characteristic of α-helix behavior. Especially $NBPF15^{59–263}$, $NBPF15^{77–263}$, and $NBPF15^{90–263}$, the content of α-helix were over than 48% with the decreasing percentage of irregular structure. Upon the truncation of N-terminus, like $NBPF15^{120–263}$, the minimum of the CD signal became less pronounced and shifted to longer wavelength (Figure 5C). Besides spectral changes, the CD signal at ∼200 nm became evident, indicative of increased random coil. We conclude that the N-terminal was more like a mosaic of different (dis) ordered segments, consist of the first disordered residues, the middle ordered helical-rich region, and the last disordered residues. The helical-rich region matched the location of the coiled-coil, and the CON1 domain lacks a stable secondary structure.

TEM image illustrated the self-aggregation of NBPF15

To examine the morphological properties of the N-terminal truncations, we used TEM. Under the vision of TEM, $NBPF15^{9–263}$ was irregular and fibrillary, forming large strands of aggregates (Figure 6A). $NBPF15^{9–263Δ59–76}$ exhibited clear outlines of proteins that formed cluster aggregates (Figure 6B). For $NBPF15^{59–263}$, the large

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

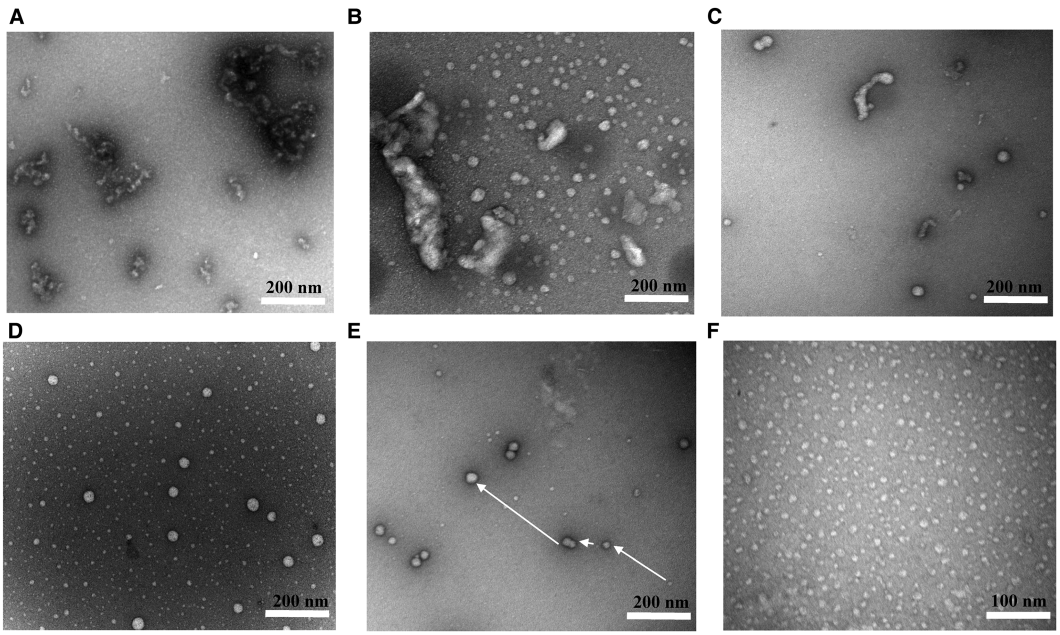

Figure 6. TEM images of NBPF15 truncated proteins following negative staining.
(**A**) NBPF15$^{9–263}$. (**B**) NBPF15$^{9–263\Delta59–76}$. (**C**) NBPF15$^{59–263}$. (**D** and **E**) NBPF15$^{77–263}$ (**F**) NBPF15$^{120–263}$.

clusters were replaced by smaller clusters with separated proteins (Figure 6C). Category II proteins NBPF15$^{77–263}$ had two distinct sizes, and it was reasonable to assume that the larger complexes were formed through the polymerization of smaller complexes (Figure 6D,E). When almost all N-terminal region was removed, the proteins were of uniform size and shape, as exhibited for NBPF15$^{120–263}$ (Figure 6F). TEM images highlighted the self-aggregation and multiple types aggregates of the N-terminal truncations.

Proteins containing the N-terminal region undergo a phase transition

In addition to being with multiple types of aggregates, the NBPF15 truncations that contained N-terminal were found to undergo a phase transition. When protein concentration reached a threshold, generally at 40 mg/ml, they gradually and spontaneously transited from a soluble state to a hydrogel state at 4°C. NBPF15$^{59–263}$ suffered the fastest phase transition and became gelatinous during the process of concentrate. The phase transition of both NBPF15$^{77–263}$ and NBPF15$^{90–263}$ occurred at ∼40 mg/ml in 1-day placement at 4°C, and NBPF15$^{9–263}$ became gelatinous at 40 mg/ml after 2 days (Figure 7A). This process can be reversed by diluting the protein. However, no phase transition was observed in NBPF15$^{9–211}$ and NBPF15$^{9–263\Delta59–76}$, even at concentrations as high as 110 mg/ml, we then added 7.5% PEG as molecular crowding agent into 40 mg/ml proteins and found that NBPF15$^{9–211}$ became irreversibly turbid (Figure 7B). After centrifugation, two separated phases were evident, and the droplets of PEG-added NBPF15$^{9–211}$ were solid particles in the vision of DIC, indicating the denaturation of NBPF15$^{9–211}$. Only NBPF15$^{9–211}$ precipitated at similar concentrations following PEG addition among all the assessed proteins, possibly due to the intrinsic instability of the incomplete CON1 domain. Considering the reduction in the α-helix content of NBPF15$^{9–263\Delta59–76}$ calculated by CD, we speculate that the interruption of the entanglement within the coiled-coil could also lead to the lack of phase transition. The speed of the phase transitions was in the order of NBPF15$^{59–263}$ > NBPF15$^{77–263}$ > NBPF15$^{90–263}$ > NBPF15$^{9–263}$, consist with the content of α-helix.

Discussion

In our experiment, phase transition occurred in proteins that contained the N-terminal region and occurred more rapidly for proteins possessed high percentage α-helix. Proteomics and genetic studies suggest that the presence of intrinsically disordered regions (IDRs) in proteins drives phase transitions, and coiled-coil is common contacts during phase transition [35–38]. According to the classification of Robin [37], the

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

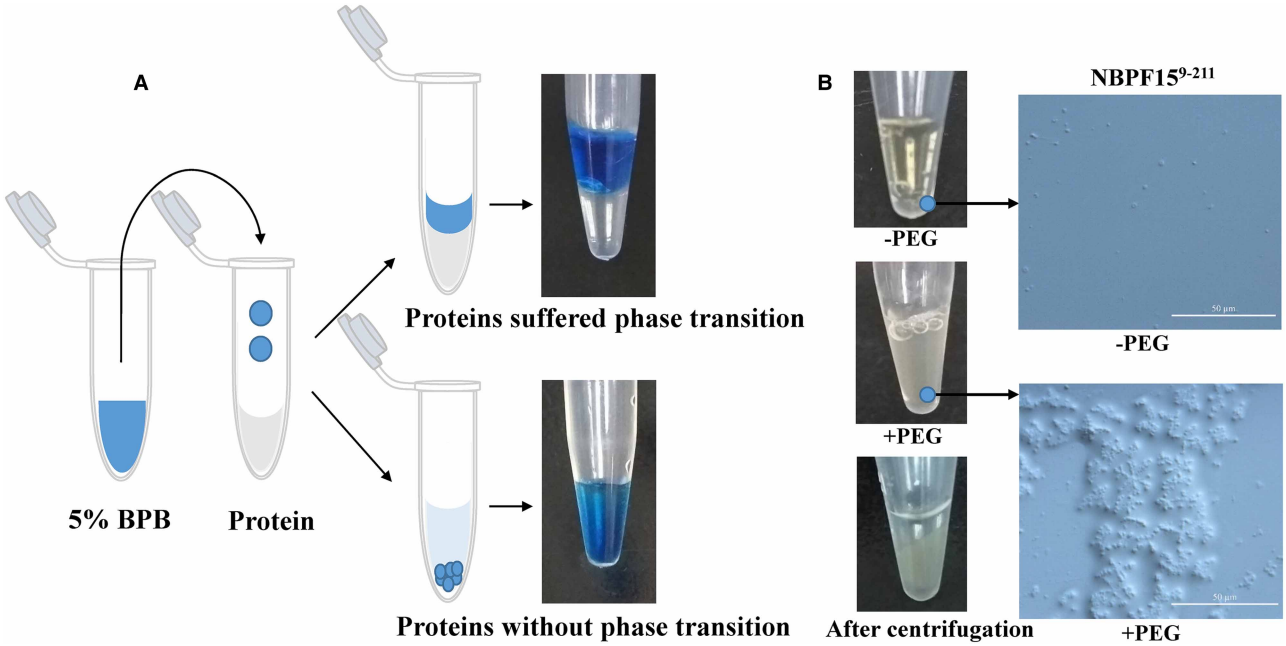

Figure 7. Observation of protein phase transition.

(**A**) Proteins that have undergone phase transition had form two insoluble layers following the addition of 5% bromophenol blue (BPB). Proteins that have not undergone the phase transition will dissolve with BPB. (**B**) 40 mg/ml NBPF15$^{9–211}$ became turbid following the addition of 7.5% PEG as a molecular crowding agent and formed two layers after centrifugation. Samples were visualized using LECIA DM6 DIC.

N-terminal of NBPF15 belongs to polyampholytes based on sequence-structural ensemble relationships, which have approximately equal numbers of positive and negative charges and are randomly distributed (Supplementary Figure S7A), the conformation of NBPF15 N-terminal was possibly like a pre-molten globule (Supplementary Figure S7B) with an ordered coiled-coil core and disordered wings, the intra-interaction between the helical region could be the reason that N-terminal proteins being with multiple types of aggregated forms.

NBPF15$^{59–263}$ received the most rapid phase transition, which could arise due to the alpha-helical entanglement within the coiled-coil and the formation of homogeneous higher-order aggregates. When the entanglement was disrupted, such as NBPF15$^{9–263\Delta59–76}$, no phase transition occurred. Supplementary Figure S2 shows that the lack of a DUF1220 domain reduced protein expression and decreased protein solubility, indicating its contribution to NBPF15 folding. Therefore, NBPF15$^{9–211}$, a protein with the full N-terminus but half the CON1 domain, probably fold incorrectly, so the presence of 7.5% PEG enhances the intrinsic self-aggregation of the proteins (reviewed in [39,40]), leading to the denaturation of NBPF15$^{9–211}$. Our current data support the hypothesis that N-terminal regions triggered the formation of multiple types of aggregates; the longer regions promoted more species of aggregated protein. The aggregates, particularly those aggregates contained the entire coiled-coil, promote phase transition.

The characteristics of self-aggregation and phase-transition of NBPF15 N-terminal proteins were consistent with neurodegenerative diseases (NDs)-related proteins. A related phase transition phenomenon has been reported for fused in sarcoma (FUS) protein, which contains the prion-like low complexity domain (LC), the specific subsets of IDRs. In cell-free systems, the incubation of purified FUS can lead to phase-separated droplets which alter maturation into the hydrogel state [41]. FUS was associated with amyotrophic lateral sclerosis (ALS), one of the neurodegenerative diseases (NDs) [42]. NDs, including Alzheimer's disease (AD) [43,44] and Parkinson's disease (PD) [45,46], each was related to specific protein aggregates. For example, the two main neuropathological lesions of AD are the β-amyloid peptide-containing plaques and the intra-neuronal neurofibrillary tangles of the microtubule-associated protein tau. The aggregation level of tau correlates with the degree of cognitive decline. These aberrant phase transitions underlie numerous neurogenic diseases, but the issue as to how and why aggregates are associated with the disease is largely unexplored.

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

Since Olduvai domains were less likely to be expressed without their promoter and the N-terminal regions [13], the expression of NBPF15 could lead to the emergence of multiple types of aggregates and the tangles between these aggregates. All NBPF proteins have abundant and conserved N-terminal sequence (Supplementary Figure S8), the biochemical characteristics of the NBPF15 N-terminus are likely to be similar to all NBPF proteins. Our current results could be studied *in vivo* by generating transgenic mouse models that overexpressed the full-length NBPF15. Monitor the behavior of mouse and using antibodies against NBPF15 and GFP to observe the situation of protein aggregation based on the method in reference [47]. Using double-label immunofluorescence to analyze protein aggregation differences in brain tissues between healthy adult and autism/macrocephaly patients with antibodies to a peptide derived from NBPF15 N terminus is also an alternative [12]. Olduvai has a known association with human cognitive diseases and shows neuron-specific expression in cell bodies and dendrites [12]. It is intriguing to speculate that the self-aggregation of NBPF15 and protein phase transitions have an impact on human cognitive function. Though no known association of aggregated NBPF proteins with neuro-genetic diseases has been confirmed *in vivo* and no studies of the self-aggregation of NBPF proteins have been reported *in vivo* or *in vitro*, this finding provides a new perspective for NBPF or Olduvai studies.

Abbreviations

AA, amino acids; AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; CD, circular dichroism; CON, conserved clade; CVs, column volumes; DIC, differential interferences contrast microscopy; DLS, dynamic light scattering; DTT, dithiothreitol; DUF, domains of unknown function; FUS, fused in sarcoma; HLS, human lineages specific clade; NBPF, neuroblastoma breakpoint family; NDs, neurodegenerative diseases; S200, superdex200; SAXS, small angle X-ray scattering; SEC, size exclusion chromatography; TEM, transmission electron microscopy.

Author Contribution

Xu-Guang Xi and Han Wu designed experiments. Han Wu performed experiments and analyzed data. Xue-Xue Guo purified the proteins. Stephane Rety processed and analyzed SAXS data. Han Wu and Liu-Tao Zhai wrote the manuscript. Xu-Guang Xi supervised the project, obtained grant support, and edited the paper before submission.

Funding

This work was supported by the National Natural Science Foundation of China [No. 31870788 and No. 11574252], Northwest A & F University Startup Funding [No. Z111021103, Z104021905].

Acknowledgements

We thank beamline scientists at BL19U2 of the National Center for Protein Sciences (China), for assistance with SEC-SAXS data collections.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

References

1 Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A. et al. (2010) Diversity of human copy number variation and multicopy genes. *Science* **330**, 641–646 https://doi.org/10.1126/science.1197005

2 Vandepoele, K., Van Roy, N., Staes, K., Speleman, F. and van Roy, F. (2005) A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol. Biol. Evol.* **22**, 2265–2274 https://doi.org/10.1093/molbev/msi222

3 Andries, V., Vandepoele, K. and van Roy, F. (2012) The NBPF Gene Family in *Neuroblastoma - Present and Future* Shimada, H. (ed.) IntechOpen, London, UK https://doi.org/10.5772/28470

4 O'Bleness, M., Searles, V.B., Dickens, C.M., Astling, D., Albracht, D., Mak, A.C. et al. (2014) Finished sequence and assembly of the DUF1220-rich 1q21 region using a haploid human genome. *BMC Genomics* **15**, 387 https://doi.org/10.1186/1471-2164-15-387

5 Vandepoele, K., Staes, K., Andries, V. and van Roy, F. (2010) Chibby interacts with NBPF1 and clusterin, two candidate tumor suppressors linked to neuroblastoma. *Exp. Cell Res.* **316**, 1225–1233 https://doi.org/10.1016/j.yexcr.2010.01.019

6 Andries, V., Vandepoele, K., Staes, K., Berx, G., Bogaert, P., Van Isterdael, G. et al. (2015) NBPF1, a tumor suppressor candidate in neuroblastoma, exerts growth inhibitory effects by inducing a G1 cell cycle arrest. *BMC Cancer* **15**, 391 https://doi.org/10.1186/s12885-015-1408-5

7 Zhu, H.Y., Bai, W.D., Li, C., Li, J. and Hu, D.H. (2017) NBPF7 promotes the proliferation of alpha-catenin-knockdown HaCaT cells via functional interaction with the NF-kappaB pathway. *Oncotarget* **8**, 65800–65808 https://doi.org/10.18632/oncotarget.19480

8 Ceylan, A.C., Sahin, I., Erdem, H.B., Kayhan, G., Simsek-Kiper, P.O., Utine, G.E. et al. (2019) An eight-case 1q21 region series: novel aberrations and clinical variability with new features. *J. Intellect. Disabil. Res.* **63**, 548–557 https://doi.org/10.1111/jir.12592
9 Eugene, A.R., Masiak, J. and Eugene, B. (2018) Predicting lithium treatment response in bipolar patients using gender-specific gene expression biomarkers and machine learning. *F1000Res.* **7**, 474 https://doi.org/10.12688/f1000research.14451.3
10 Zhou, F., Xing, Y., Xu, X., Yang, Y., Zhang, J., Ma, Z. et al. (2013) NBPF is a potential DNA-binding transcription factor that is directly regulated by NF-kappaB. *Int. J. Biochem. Cell Biol.* **45**, 2479–2490 https://doi.org/10.1016/j.biocel.2013.07.022
11 Sikela, J.M. and van Roy, F. (2017) Changing the name of the NBPF/DUF1220 domain to the Olduvai domain. *F1000Res.* **6**, 2185 https://doi.org/10.12688/f1000research.13586.1
12 Popesco, M.C., Maclaren, E.J., Hopkins, J., Dumas, L., Cox, M., Meltesen, L. et al. (2006) Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313**, 1304–1307 https://doi.org/10.1126/science.1127980
13 O'Bleness, M.S., Dickens, C.M., Dumas, L.J., Kehrer-Sawatzki, H., Wyckoff, G.J. and Sikela, J.M. (2012) Evolutionary history and genome organization of DUF1220 protein domains. *G3* **2**, 977–986 https://doi.org/10.1534/g3.112.003061
14 O'Bleness, M., Searles, V.B., Varki, A., Gagneux, P. and Sikela, J.M. (2012) Evolution of genetic and genomic features unique to the human lineage. *Nat. Rev. Genet.* **13**, 853–866 https://doi.org/10.1038/nrg3336
15 Zimmer, F. and Montgomery, S.H. (2015) Phylogenetic analysis supports a link between DUF1220 domain number and primate brain expansion. *Genome Biol. Evol.* **7**, 2083–2088 https://doi.org/10.1093/gbe/evv122
16 Davis, J.M., Searles Quick, V.B. and Sikela, J.M. (2015) Replicated linear association between DUF1220 copy number and severity of social impairment in autism. *Hum. Genet.* **134**, 569–575 https://doi.org/10.1007/s00439-015-1537-6
17 Searles Quick, V.B., Davis, J.M., Olincy, A. and Sikela, J.M. (2015) DUF1220 copy number is associated with schizophrenia risk and severity: implications for understanding autism and schizophrenia as related diseases. *Transl. Psychiatry* **5**, e697 https://doi.org/10.1038/tp.2015.192
18 Davis, J.M., Searles, V.B., Anderson, N., Keeney, J., Raznahan, A., Horwood, L.J. et al. (2015) DUF1220 copy number is linearly associated with increased cognitive function as measured by total IQ and mathematical aptitude scores. *Hum. Genet.* **134**, 67–75 https://doi.org/10.1007/s00439-014-1489-2
19 Brunetti-Pierri, N., Berg, J.S., Scaglia, F., Belmont, J., Bacino, C.A., Sahoo, T. et al. (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.* **40**, 1466–1471 https://doi.org/10.1038/ng.279
20 Keeney, J.G., Davis, J.M., Siegenthaler, J., Post, M.D., Nielsen, B.S., Hopkins, W.D. et al. (2015) DUF1220 protein domains drive proliferation in human neural stem cells and are associated with increased cortical volume in anthropoid primates. *Brain Struct. Funct.* **220**, 3053–3060 https://doi.org/10.1007/s00429-014-0814-9
21 Dumas, L.J., O'Bleness, M.S., Davis, J.M., Dickens, C.M., Anderson, N., Keeney, J.G. et al. (2012) DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am. J. Hum. Genet.* **91**, 444–454 https://doi.org/10.1016/j.ajhg.2012.07.016
22 Dumas, L. and Sikela, J.M. (2009) DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 375–382 https://doi.org/10.1101/sqb.2009.74.025
23 Sikela, J.M. and Searles Quick, V.B. (2018) Genomic trade-offs: are autism and schizophrenia the steep price of the human brain? *Hum. Genet.* **137**, 1–13 https://doi.org/10.1007/s00439-017-1865-9
24 Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 https://doi.org/10.1038/nprot.2015.053
25 Liu, B., Wang, R., Wu, F., Xu, X. and Chen, H. (2014) Rapid production of HIV-1 neutralizing antibodies in baculovirus infected insect cells. *Protein Expr. Purif.* **99**, 87–93 https://doi.org/10.1016/j.pep.2014.04.004
26 Wyatt Technology Corporation. (2010) Dynamics user's guide. https://physiology.case.edu/media/eq_manuals/eq_manual_Dynamics_Users_Guide_M1400_Rev_K.pdf
27 Chevallet, M., Luche, S. and Rabilloud, T. (2006) Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* **1**, 1852–1858 https://doi.org/10.1038/nprot.2006.288
28 Brookes, E., Vachette, P., Rocco, M. and Perez, J. (2016) US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *J. Appl. Crystallogr.* **49**, 1827–1841 https://doi.org/10.1107/S1600576716011201
29 Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M. et al. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 https://doi.org/10.1107/S0021889812007662
30 Lu, K.Y., Chen, W.F., Rety, S., Liu, N.N., Wu, W.Q., Dai, Y.X. et al. (2018) Insights into the structural and mechanistic basis of multifunctional *S. cerevisiae* Pif1p helicase. *Nucleic Acids Res.* **46**, 1486–1500 https://doi.org/10.1093/nar/gkx1217
31 Receveur-Brechot, V. and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13**, 55–75 https://doi.org/10.2174/138920312799277901
32 Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B. et al. (2013) D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–D516 https://doi.org/10.1093/nar/gks1226
33 Fukuchi, S., Hosoda, K., Homma, K., Gojobori, T. and Nishikawa, K. (2011) Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct. Biol.* **11**, 29 https://doi.org/10.1186/1472-6807-11-29
34 Wiedemann, C., Bellstedt, P. and Gorlach, M. (2013) CAPITO–a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics* **29**, 1750–1757 https://doi.org/10.1093/bioinformatics/btt278
35 Forman-Kay, J.D. and Mittag, T. (2013) From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* **21**, 1492–1499 https://doi.org/10.1016/j.str.2013.08.001
36 Tompa, P. (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **37**, 509–516 https://doi.org/10.1016/j.tibs.2012.08.004
37 van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K. et al. (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 https://doi.org/10.1021/cr400525m
38 Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F. et al. (2018) Protein phase separation: a new phase in cell biology. *Trend Cell Biol.* **28**, 420–435 https://doi.org/10.1016/j.tcb.2018.02.004
39 Fonin, A.V., Darling, A.L., Kuznetsova, I.M., Turoverov, K.K. and Uversky, V.N. (2018) Intrinsically disordered proteins in crowded milieu: when chaos prevails within the cellular gumbo. *Cell. Mol. Life Sci.* **75**, 3907–3929 https://doi.org/10.1007/s00018-018-2894-9

Biochemical Journal (2020) **477** 445–458
https://doi.org/10.1042/BCJ20190566

40 Kim, J.S. and Yethiraj, A. (2011) Crowding effects on protein association: effect of interactions between crowding agents. *J. Phys. Chem. B* **115**, 347–353 https://doi.org/10.1021/jp107123y

41 Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C. et al. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 https://doi.org/10.1016/j.cell.2012.04.017

42 Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y. et al. (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 https://doi.org/10.1016/j.cell.2015.07.047

43 Serrano-Pozo, A., Frosch, M.P., Masliah, E. and Hyman, B.T. (2011) Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **1**, a006189 https://doi.org/10.1101/cshperspect.a006189

44 Spillantini, M.G. and Goedert, M. (2013) Tau pathology and neurodegeneration. *Lancet Neurol.* **12**, 609–622 https://doi.org/10.1016/S1474-4422(13)70090-5

45 Spillantini, M.G., Schmidt, M.L., Lee, V.M., Trojanowski, J.Q., Jakes, R. and Goedert, M. (1997) Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840 https://doi.org/10.1038/42166

46 Taylor, J.P., Hardy, J. and Fischbeck, K.H. (2002) Toxic proteins in neurodegenerative disease. *Science* **296**, 1991–1995 https://doi.org/10.1126/science.1067122

47 Gordon, D., Dafinca, R., Scaber, J., Alegre-Abarrategui, J., Farrimond, L., Scott, C. et al. (2019) Single-copy expression of an amyotrophic lateral sclerosis-linked TDP-43 mutation (M337 V) in BAC transgenic mice leads to altered stress granule dynamics and progressive motor dysfunction. *Neurobiol. Dis.* **121**, 148–162 https://doi.org/10.1016/j.nbd.2018.09.024