



HAL
open science

Condensation of the N-terminal domain of human topoisomerase 1 is driven by electrostatic interactions and tuned by its charge distribution

Greta Bianchi, Marco Mangiagalli, Diletta Ami, Junaid Ahmed, Silvia Lombardi, Sonia Longhi, Antonino Natalello, Peter Tompa, Stefania Brocca

► To cite this version:

Greta Bianchi, Marco Mangiagalli, Diletta Ami, Junaid Ahmed, Silvia Lombardi, et al.. Condensation of the N-terminal domain of human topoisomerase 1 is driven by electrostatic interactions and tuned by its charge distribution. *International Journal of Biological Macromolecules*, 2023, 254, pp.127754. 10.1016/j.ijbiomac.2023.127754 . hal-04277952

HAL Id: hal-04277952

<https://hal.science/hal-04277952v1>

Submitted on 9 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Condensation of the N-terminal domain of human topoisomerase 1 is driven by electrostatic interactions and tuned by its charge distribution

Greta Bianchi¹, Marco Mangiagalli¹, Diletta Ami¹, Junaid Ahmed², Silvia Lombardi¹, Sonia Longhi³, Antonino Natalello¹, Peter Tompa^{3*}, Stefania Brocca^{1*}

¹Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126, Milano, Italy.

²VIB-VUB Center for Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium.

³Lab. Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Aix-Marseille University and CNRS, 13288, Marseille, France.

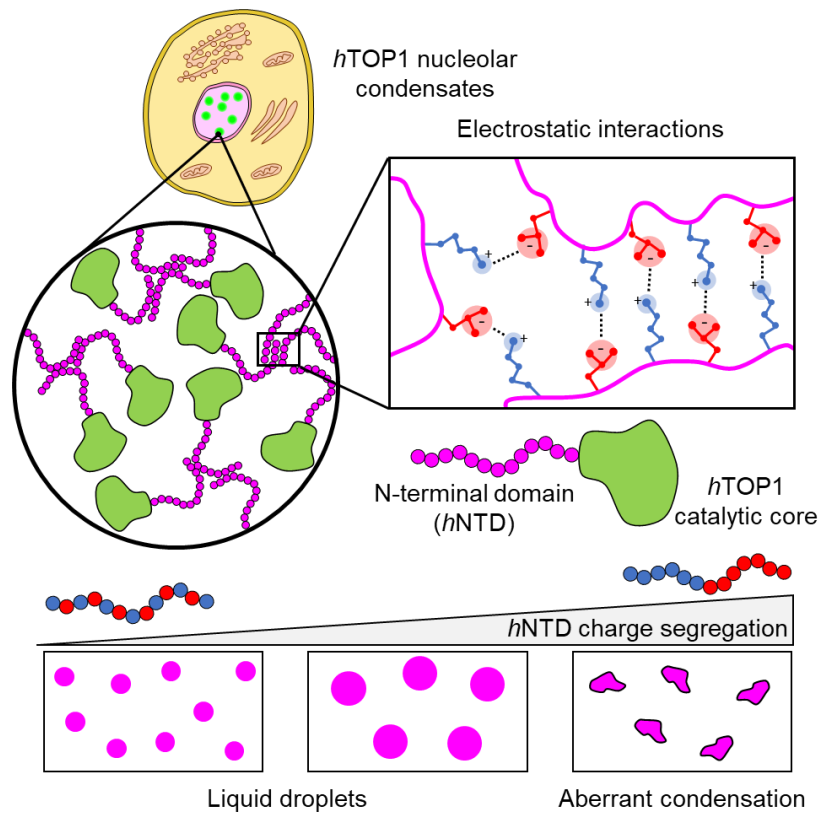
* Corresponding authors:

Prof. Stefania Brocca, E-mail: stefania.brocca@unimib.it;

Prof. Peter Tompa, E-mail: peter.tompa@vub.be

Highlights

- Human topoisomerase 1 (*hTOP1*) can form condensates in the nucleolus and nucleoplasm;
- Human topoisomerase I contains a condensation-prone, N-terminal domain (*hNTD*);
- *hNTD* is highly charged and its *in-vitro* condensation is driven by electrostatic forces;
- There is a critical charge clustering threshold affecting LLPS efficiency;
- This research extends the applicability of “stickers-and-spacers” theory to supercharged IDPs.



1. Introduction

Liquid-liquid phase separation (LLPS) plays a key role in the formation of biomolecular condensates, *i.e.* liquid assemblies of proteins and other macromolecules, such as nucleic acids [1]. These condensates dynamically emerge and dissolve in response to various stimuli, including temperature, protein concentration, pH, ionic strength, and cellular stresses [2]. They wield profound influence in diverse, biological processes encompassing chromatin reorganization [3], DNA replication [4], and the formation of membrane-less organelles (MLOs), exemplified by the nucleolus and the Cajal bodies, shaping the intracellular milieu [5]. Condensates typically harbor intrinsically disordered proteins (IDPs) or proteins with intrinsically disordered regions (IDRs), *i.e.* proteins or regions lacking a well-defined tertiary structure. IDPs/IDRs exhibit sequences enriched in polar and charged residues and depleted in hydrophobic and aromatic residues, often featuring low-complexity regions [6]. These attributes endow IDPs/IDRs with greater flexibility and lower chain compactness, ideally suiting them for the network of intermolecular interactions crucial for LLPS [7]. At the molecular level, LLPS is driven by both intra- and intermolecular interactions [8], including π - π stacking, cation- π , and electrostatic interactions, among residues termed “stickers” [9]. Conversely, “spacers” elements, such as Gly and polar amino acids, serve to separate sticker residues, facilitating the formation of optimal LLPS interactions [10]. Collectively, sequence-specific motifs and chemical context synergistically govern LLPS [11].

Positively charged (Lys and Arg) and negatively charged (Glu and Asp) amino acid residues act as sticker elements, with interaction strength depending on various factors including post-translational modifications [12], net charge, and distribution of charged residues along the protein primary structure [13]. To describe charge distribution, the κ parameter was introduced, where a κ value of 0 signifies opposite charges alternating uniformly along the sequence, and a κ value of 1 indicates segregation into two clusters [13]. A few computational [14-16] and experimental studies on proteins like Laf-1 N-terminal domain (Laf-1 RGG [17]), Ddx4 N-terminal domain (Ddx4^{N1}[18]), and Nephrin intracellular domain (NICD [19]) have underscored the significance of charge patterning, showing in particular that charge segregation promotes LLPS. Nevertheless, the role of charge distribution in highly charged IDPs remains enigmatic.

In this study, we investigated the impact of charge distribution within the N-terminal domain (*h*NTD, residues 1-214) of human topoisomerase 1 (*h*TOP1) on its propensity for LLPS, triggered by pH fluctuations or RNA interactions. DNA topoisomerases (EC 5.6.2.1) are ubiquitous nuclear enzymes active in relaxing DNA supercoils arising during DNA replication and transcription [20]. *h*NTD, endowed with nuclear and nucleolar localization sequences (NLSs), is highly charged and exerts influence over numerous physiological functions [21]. Our data indicate that *h*TOP1 can form biomolecular condensates *in vivo*, and that *h*NTD alone is capable of LLPS driven by electrostatic interactions *in vitro*. To delve further into the influence of charge distribution, we designed two synthetic variants featuring moderate ($\kappa = 0.224$) or marked ($\kappa = 0.298$) charge segregation, mirroring patterns found in *h*TOP1 orthologs. Overall, our results support the conclusion that moderate charge segregation promotes LLPS, while marked charge segregation leads to abnormal condensates. Despite using a limited number of charge permutants, our study supports the existence of an optimal distribution of charged stickers that enhances LLPS, consistent with prior research on π - π stacking and cation- π interactions [22].

2. Materials and Methods

2.1. Sequence analyses of *h*TOP1

Sequence analyses of *hTOP1* (UniProt code: P11387) were carried out using CIDER [23], IUPred [24] and FuzDrop [25] softwares. Nuclear localization sites (NLSs) were manually identified based on literature data [26].

hTOP1 orthologs were retrieved using OrthoDB v10 ([https:// www.orthodb.org](https://www.orthodb.org)), a database of putative orthologous gene sequences from genomes of eukaryotes, prokaryotes and viruses [27], and the sequence of *hTOP1* gene (NP_003277.1) as a query. OrthoDB identifies 1649 putative orthologs of the *hTOP1* gene, distributed among metazoa (676 genes), fungi (544 genes), viridiplantae (216 genes) and protista (144 genes). To narrow the dataset, evolutionary representative taxa are depicted by one or two sequences with the highest similarity to *hTOP1*. The representative sequences were manually selected from a protein BLAST (blastp) analysis [28] using the *hTOP1* sequence as a query. Short sequences (< 600 amino acids) and those annotated as “putative” or mitochondrial topoisomerases were excluded. The final data set includes 55 sequences of TOP1 (**Table S1**). The NTD of TOP1s were identified through Pfam [29]. For each NTD sequence, FCR, NCPR and κ value were calculated using CIDER [23]. The phylogenetic tree of full-length TOP1s was built by the Neighbour Joining algorithm [30] using MEGAX [31].

2.2. *In-vivo* assessment of the ability of *hTOP1* to form biomolecular condensates

To evaluate the ability of *hTOP1* to undergo LLPS *in vivo*, we monitored the formation of biomolecular condensates and their material properties *via* fluorescence recovery after photobleaching (FRAP) in HeLa cells that were transiently transfected with the pEGFP-*hTOP1* plasmid. This plasmid, which encodes the full-length *hTOP1* fused to the C-terminal end of EGFP, was generated in T. Honjo laboratory at Kyoto University [32] and kindly provided by S. Barabino (University of Milano-Bicocca). pEGFP-*hTOP1* was amplified in *Escherichia coli* strain DH5 α , extracted and purified with Plasmid Mini-Prep Kit (Jena Biosciences, Jena, Germany) and used to transfected HeLa cells using LipofectamineTM 2000 (Invitrogen, Waltham, Massachusetts, US).

HeLa cells (obtained from the European Collection of Cell Cultures, ECACC) were cultured in high glucose Dulbecco’s Modified Eagle Medium (DMEM), supplemented with 10% fetal bovine serum, 100 IU/mL penicillin, 0.1 mg/mL streptomycin and 2 mM L-glutamine (all from Sigma-Aldrich). Cells were maintained at 37°C and 5% CO₂ in a humidified incubator. The following day, transfected cells were seeded on glass-bottom 35-mm dishes and allowed to attach overnight. Before FRAP analysis, the cell medium was replaced by phenol red-free DMEM, with or without the nuclear-specific dye Hoechst 33342 (Sigma-Aldrich) at 0.5 μ g/mL. FRAP experiments were performed 48 hours post transfection using the 100x oil immersion objective and 5x optical zoom of the inverted confocal microscope Nikon Eclipse Ti, equipped with a humidified chamber set to 5% CO₂ and 37°C. Bleaching was performed with a 488 nm argon laser (100% intensity, scan speed 0.5, 8 iterations [3.786 s], bleaching area 20- pixel in diameter). For FRAP analysis, we assessed the fluorescence intensity of a nucleoplasmatic droplet (bleached region of interest, ROI), of two non-bleached nuclear regions (NB1 and NB2) and one background area (BG). Three images of each region were acquired before the bleaching, and then the fluorescence was monitored for 120 s at 2-s intervals. To study condensate fluorescence recovery, data were normalized as it follows. First, we calculated the relative fluorescence of ROI, NB1 and NB2 using the Eq.1:

$$\text{Eq. 1} \quad \text{Relative Fluorescence (ReflFluo)} = \frac{F_{BG}}{\text{Mean}(F_{BG}^{\text{BeforeBleaching}})}$$

Where F_{BG} corresponds to the fluorescence intensity of the ROI, NB1 and NB2 subtracted for the background contribution; $\text{Mean}(F_{BG}^{\text{BeforeBleaching}})$ represents the average F_{BG} before the bleaching.

Then, the normalized recovery was calculated using the Eq. 2:

$$\text{Eq. 2} \quad \text{Normalized recovery} = \frac{\text{RelFluo}_{ROI}}{\text{Mean}(\text{RelFluo}_{NB1}; \text{RelFluo}_{NB2})}$$

Where RelFluo_{ROI} corresponds to the relative fluorescence of the ROI; $\text{Mean}(\text{RelFluo}_{NB1}; \text{RelFluo}_{NB2})$ represents the average of the relative fluorescence of the two non-bleached regions NB1 and NB2.

2.3. Cloning and mutagenesis

The sequence encoding the N-terminal domain of *hTOP1* (*hNTD*, residues 1-214) was optimized for expression in *E. coli* cells, chemically synthesized (Genscript, Piscataway, NJ, USA) and cloned into the pET-21a vector (EMD, Millipore, Billerica, MA, USA) between the *NdeI* and *XhoI* sites (Jena Biosciences, Jena, Germany). We designed two variants of *hNTD*, namely medium κ ($M\kappa$ -NTD) and high κ ($H\kappa$ -NTD), by modifying the distribution of charged residues only within the first 116 amino acid residues ($hNTD^{1-116}$), while maintaining unchanged the NLSs-containing region ($hNTD^{117-214}$). The overall strategy of permutation of $hNTD^{1-116}$ was previously described in [33]; briefly the distribution of charged residues was modified to reach a κ value of 0.226 and 0.373, while the position of non-charged residues was left unchanged (**Fig. S1**). The sequences encoding $M\kappa$ -NTD and $H\kappa$ -NTD variants, optimized for expression in *E. coli*, were chemically synthesized and cloned into the pET-21a vector as described above. The *E. coli* strain DH5 α TM (Invitrogen, Waltham, USA) was used for DNA amplification.

The gene coding for the green fluorescent protein (GFP) was cloned at the 3' end of the NTD sequences using the *SpeI* site, yielding the following constructs: pET-21a[*hNTD*-GFP], pET-21a[$M\kappa$ -NTD-GFP] and pET-21a[$H\kappa$ -NTD-GFP]. The resulting constructs were checked by restriction analysis and bidirectional DNA sequencing (GATC-Biotech, Koeln, Germany).

The *E. coli* BL21(DE3) strain (EMD Millipore, Billerica, USA) was used for the expression of the recombinant proteins. *E. coli* cells were transformed with pET-21 vectors carrying the gene for the expression of each model protein with and without GFP. Transformant cells were selected on Lennox medium (LM, yeast extract 5 g/L, tryptone 10 g/L, NaCl 5 g/L) agar plates supplemented with ampicillin (100 mg/L). GFP was produced by BL21(DE3) cells transformed with pET21a[GFP] as previously described in [34].

2.4. Production and purification of *hNTD* and its variants

BL21(DE3) cell culture were grown at 37°C in LM to OD₆₀₀ 0.8-1, then diluted 1:20 in production medium (yeast extract 5 g/L, triptone 10 g/L, glycerol 2 g/L) with ampicillin (100 mg/L) and incubated for 2 h at 37°C to reach OD₆₀₀ 0.6-0.8 before the induction with 1% lactose or 0.1 mM isopropyl- β -D-thiogalactopyranoside (IPTG). The cultures were incubated at 25°C for 16 hours before the cells were harvested and lysed as previously described in Tedeschi et al. [35]. Recombinant proteins were purified from soluble fractions by immobilized-metal affinity chromatography (IMAC) on nickel-nitrilotriacetic acid agarose resin. High-concentrated fractions were pooled and the buffer was exchanged to an alkaline phosphate buffer (PB, 50 mM sodium phosphate, 50-300 mM NaCl, pH 10.5) or a neutral saline phosphate buffer (PBS, 50 mM sodium phosphate, 100 mM NaCl, pH 7.0) by gel filtration on PD-10 columns (GE Healthcare, Little Chalfont, UK). Protein concentration was determined by BCA assay (Novagen-Merck, Darmstadt, Germany) using bovine serum albumin as a standard.

2.5. pH-induced condensation

For pH jump experiments, recombinant proteins were dissolved in an alkaline phosphate buffer at pH 10.5. Protein condensation was induced by adding a small amount (typically 15% v/v) of 0.1 M HCl, which shifted the pH to 7.5. Condensation was assayed at different concentrations of protein, i.e. 25, 50 and 100 μ M, unless otherwise stated. The effects of NaCl on pH-induced condensation was tested at

different concentrations of NaCl, i.e. 50, 150 and 300 mM, with salt being added to the sample before the addition of HCl. Turbidimetry experiments were carried out on *h*NTD, Mκ-NTD and Hκ-NTD by monitoring the turbidity of the samples at 600 nm with a Jasco V-770 UV/NIR spectrophotometer (JASCO, Europe, Lecco, Italy), at 25°C for 1 hour. The experiments were conducted in triplicate and means and standard deviations were calculated.

Confocal microscopy analyses were carried out on *h*NTD-GFP, Mκ-NTD-GFP and Hκ-NTD-GFP with a Nikon Eclipse Ti inverted microscope. After the pH jump, samples were incubated for 5 minutes and then observed on a glass slide with a 100x oil immersion objective (numerical aperture 1.45), exciting the GFP with a 488 nm argon laser. Images were acquired with Nikon Elements Imaging Software at the indicated protein concentrations. Experiments were performed in triplicate. Image analysis was used to measure the extent of droplet formation, droplet size and fusion events using FIJI ImageJ software (v2.0). FRAP analyses were carried out on stationary droplets. Bleaching was performed with a 488 nm argon laser (100% intensity, scan speed 0.125 sec/frame, 1 iteration [121.09 ms], optical zoom 5x) and observed every second for one minute. To study fluorescence droplet recovery, data were normalized employing Eq. 3:

$$\text{Eq. 3} \quad \text{Normalized Recovery} = \frac{(FROI_{BG})}{(FNB_{BG})} * \frac{\text{Mean}(FNB_{BG}^{\text{BeforeBleaching}})}{\text{Mean}(FROI_{BG}^{\text{BeforeBleaching}})}$$

where $FROI_{BG}$ corresponds to the fluorescence intensity of the bleached region of interest (ROI) subtracted for the background contribution; FNB_{BG} represents the fluorescence intensity of a non-bleached area within the droplet, subtracted for the background contribution; $\text{Mean}(FROI_{BG}^{\text{BeforeBleaching}})$ corresponds to the average $FROI_{BG}$ before the bleaching; $\text{Mean}(FNB_{BG}^{\text{BeforeBleaching}})$ represents the average FNB_{BG} before the bleaching.

2.6. RNA-induced LLPS

For the RNA experiments, recombinant proteins were dissolved in PBS. Confocal microscopy analyses were carried out on GFP-fused proteins as previously described. Poly(A) RNA (P9403, Sigma Aldrich) was added at various concentrations (0, 0.5, 2, and 5 mg/mL) to *h*NTD-GFP, Mκ-NTD-GFP, and Hκ-NTD-GFP (final protein concentration: 20 μM). Samples were incubated for 3 minutes and then observed on a glass slide with a 100x oil immersion objective (numerical aperture 1.45), by exciting GFP at 488 nm with an argon laser. Morphological and FRAP analyses were conducted as previously described. As a control, GFP alone was incubated with 0.5 mg/mL of poly(A) RNA applying the same protocol (**Fig. S2B**), so as to rule out any possible GFP condensation induced by poly(A).

2.7. Determination of protein secondary structure

The secondary structure of *h*NTD, Mκ-NTD, and Hκ-NTD was monitored by Mid-Fourier transform infrared spectroscopy (FTIR) and far UV-circular dichroism (Far-UV CD). FTIR measurements were performed in attenuated total reflection (ATR) mode (Quest, Specac, USA). Spectra were collected on protein samples resuspended in alkaline PB at pH 10.5, or in PBS at pH 7.5. The formation of protein condensates was induced by either a pH jump or the addition of RNA (0.5 mg/mL), as described in the previous paragraphs. The pH-jump experiments were carried out at 50 mM NaCl for the *h*NTD and Mκ-NTD variants (protein concentration: 50 μM), and at 150 mM NaCl for the Hκ-NTD variant (protein concentration: 20 μM). After the pH jump, samples were incubated at room temperature for two hours and centrifuged at 20,800 g at 20°C for 10 minutes, to separate the sedimenting component, recovered in the pellet, from the soluble component in the supernatant (spin assay). For RNA experiments, after addition of 0.5 mg/mL poly(A) RNA, samples were incubated for 3 minutes at room temperature and centrifuged at 20,800 g at 4°C for 10 minutes, to perform the spin assay. The supernatant was collected and analyzed by FTIR, while the pellet was washed with PBS prior to measurement. Three microliters

of each sample were deposited on the single reflection diamond element of the ATR device (Ami *et al.*, 2022). ATR/FTIR spectra were collected after mild-solvent evaporation process (20-30 s of a gentle flow of dry air) [36] at room temperature using a Varian 670-IR spectrometer (Varian Australia Pty Ltd., Mulgrave, Victoria, Australia), equipped with a nitrogen-cooled mercury-cadmium-telluride detector, under the following conditions: 2 cm⁻¹ resolution, 25 kHz scan speed, 1024 scan co-additions, triangular apodization, and purging with dry air.

The ATR-FTIR absorption spectra were corrected for the buffer contribution and, when necessary, for water vapor contribution, normalized with respect to the area of the Amide I band (1700–1600 cm⁻¹) and smoothed using the Savitsky-Golay method (25 points) before calculation of the second derivative. Spectral analyses were performed with Resolutions-Pro software (Varian Australia Pty Ltd., Mulgrave, Victoria, Australia). At least three independent measurements were performed for each condition unless otherwise indicated.

Far-UV CD spectra were recorded in alkaline phosphate buffer using a Jasco J-815 spectropolarimeter (Jasco Europe, Lecco, Italy) in a 1-mm path-length quartz cuvette. Measurements were recorded at variable wavelengths (190–260 nm) with a scanning velocity of 20 nm/min and a data pitch of 0.2 nm. All spectra were corrected for buffer contribution, averaged from three independent acquisitions, and smoothed by the Means-Movement algorithm implemented in the Spectra Manager package (Jasco Europe, Lecco, Italy). The deconvolution of CD spectra to assess secondary structure content was performed using the BestSel program [37]. Experiments were performed in triplicate.

2.8 Statistical analysis

Statistical analysis was performed using OriginLab software (OriginLab Corporation, Northampton, USA). The p-values were determined by using an unpaired two-tailed t-test. No statistical methods or criteria were used to estimate size or to include or exclude samples.

3. Results

3.1. *h*TOP1 forms liquid condensates *in vivo*

*h*TOP1 is a nuclear enzyme that carries out pivotal functions in both nucleolus and nucleoplasm [38]. A seminal work carried out prior to the discovery of the biological relevance of LLPS suggests that *h*TOP1 is able to form nucleolar and/or nucleoplasmatic condensates, but their liquid behavior has not been assessed yet [21]. Here, we investigated the ability of *h*TOP1 to form liquid condensates by transiently expressing exogenous GFP-tagged *h*TOP1 in HeLa cells. In our cellular model, *h*TOP1 indeed forms nucleoplasmatic and nucleolar condensates, giving rise to a sparse punctuate pattern (**Fig. 1A**). Notably, these condensates exhibit a liquid-like behavior, as indicated by FRAP analyses, with a full fluorescence recovery within 80 seconds (**Fig. 1B**). In light of the well-established role of IDRs in LLPS, we reasoned that the formation of *h*TOP1 condensates might be primarily driven by its *h*NTD (residues 1-214), which has been identified as intrinsically disordered [39]. FuzDrop analysis predicted this domain as the sole droplet-promoting region (**Fig. 1C**). A detailed sequence analysis further revealed that *h*NTD is a highly charged domain (Glu + Asp: 27.6%; Lys + Arg: 32.7%), characterized by a scarcity of hydrophobic residues (Gly + Ala + Val + Leu + Ile + Pro: 17.3%) and aromatic residues (Tyr + Trp + Phe: 4.2%). Remarkably, the proportion of charged residues within *h*NTD is twice as high as that observed in other IDRs where charge distribution has been shown to play a role in LLPS, such as Ddx4^{N1} [18], Laf1 RGG [17] and NICD [19].

Collectively, these findings indicate that *h*TOP1 forms LLPS condensates *in cellula*, underscoring the prominent role of *h*NTD and electrostatic interactions in driving the LLPS phenomenon.

3.2. Three evolutionary conserved charge patterns are present in TOP1 NTD sequences

The *hNTD* exhibits a relatively low κ value of 0.128. To determine whether this charge distribution is unique to *hNTD* or a prevailing feature among *hTOP1* orthologs, we analyzed NTD sequences from *TOP1* across 55 different eukaryotic taxa, sampling the broad evolutionary pathway taken by this protein. Despite the notable variability in the length and sequence composition of the sampled NTDs, all these domains are highly charged, characterized by a balanced presence of oppositely charged residues (net charge per residue: 0.036 ± 0.035 , **Table S1**). These regions are particularly enriched in Ser and Ala while depleted in Cys and aromatic residues (**Fig. 2A**). The examination of κ values enabled us to categorize these sequences into three distinct groups: group 1, represented by vertebrates, with κ values < 0.15 (in blue in **Fig. 2B**); group 2, featuring κ values ranging from 0.15 to 0.3 and comprising protozoa, insects and certain plants, (in gray in **Fig. 2B**); and finally, group 3, encompassing fungi along with the rest of plants and protozoa, characterized by κ values > 0.3 (in red in **Fig. 2B**).

Overall, our sequence analyses indicate that the NTD is a highly-charged, disordered domain, with its charge distribution appearing to correlate with the evolutionary trajectory of *TOP1*. The most marked charge clustering ($\kappa > 0.3$) is prevalent in ancestral eukaryotes (such as fungi and unicellular eukaryotes), whereas higher eukaryotes exhibit a more regular alternation of positive and negative charges. Therefore, we resorted to assess whether *hNTD*, characterized by a low κ value, can indeed undergo LLPS and to elucidate the role of electrostatic interactions in driving this phenomenon.

3.3. *hNTD* undergoes electrostatically driven LLPS *in vitro*

The experimental plan used to investigate the *in-vitro* LLPS of *hNTD* and its variants is summarized in **Fig. S3**. To assess the ability of *hNTD* on its own to undergo LLPS, we produced it recombinantly and triggered phase separation through an abrupt change of pH (pH jump), under controlled salt concentration. Drastic pH changes are an efficient method for inducing electrostatic-driven protein condensation and for studying its kinetics *in vitro* [40].

To evaluate its propensity to condense, *hNTD* was prepared in an alkaline buffer at pH 10.5 with 50 mM of NaCl. By subsequently adding a small quantity of HCl, the solution pH was lowered to 7.5. Upon the pH jump, turbidity developed in the samples, which was monitored at 600 nm over time. We observed that turbidity of the samples increased to a maximum within the initial 10 minutes, and then either stabilized or slowly decreased (**Fig. 3A**). We found that protein concentration influenced the maximum turbidimetry value but not its kinetics (**Fig. 3A**).

To investigate whether electrostatic charges drove the protein condensation, we performed the pH jump in the presence of increasing salt concentrations, which can inhibit electrostatic interactions. We observed a significant decrease in turbidity at 150 mM and 300 mM of NaCl (**Fig. 3A**), indicating the crucial role of electrostatic interactions in *hNTD* condensation.

Since turbidimetry alone does not specifically confirm LLPS and could be ascribable to the formation of amorphous/fibrillar aggregates, confocal microscopic analyses were conducted to identify LLPS hallmarks: spherical shape, the ability to undergo fusion events, and liquid-like behavior [41]. The morphology and material properties of pH-induced condensates were investigated via confocal fluorescence microscopy and FRAP analysis on the *hNTD*-GFP construct. The fusion of *hNTD* with GFP did not affect its condensation propensity and kinetics (**Fig. S2A**), and it was confirmed that GFP does not undergo LLPS under these experimental conditions (**Fig. S2B**).

Condensates of *hNTD*-GFP obtained at low salt concentration (50 mM NaCl) displayed an average area of $0.65 \pm 0.35 \mu\text{m}^2$ (**Fig. 3B**) and a liquid-like behavior, coalescing readily (**Fig. 3C**) and recovering fluorescence efficiently after photobleaching ($\sim 80\%$ within 1 min) (**Fig. 3D**). Droplet size decreased with increasing salt concentration: their areas halved from 50 to 150 mM NaCl and were reduced ten times from 50 to 300 mM NaCl (**Fig. 3B**).

Overall, our results demonstrate that ionic strength counteracts *hNTD* condensation, underscoring the significant role of electrostatic charges in its LLPS. To monitor the secondary structure content of the

condensates, we utilized Fourier transform infrared (FTIR) spectroscopy directly applied to protein condensates harvested by centrifugating samples after the pH jump (spin assay, **Fig. 3E**). Before the pH jump, the second derivative spectrum of *hNTD* showed a main component at $\sim 1644\text{ cm}^{-1}$, indicative of disordered structures [42]. Additionally, low-intensity peaks at ~ 1677 and $\sim 1688\text{ cm}^{-1}$, within the spectral regions of turns and β -sheets (black line in **Fig. 3E**), suggested the existence of some residual secondary structure [43]. These elements of secondary structure were also observed in the far-UV CD spectrum of *hNTD* recorded in alkaline phosphate buffer (**Fig. S4A**).

After the pH jump, the FTIR spectrum of the supernatant (red line in **Fig. 3E**) resembled that observed before the pH jump, with the main peak shifted from ~ 1644 to $\sim 1652\text{ cm}^{-1}$, suggesting that the pH jump induced only minor changes in the secondary structure of unassembled *hNTD*. The FTIR spectrum of the pellet (blue line in **Fig. 3E**) collected after the pH jump displayed the main peak at $\sim 1652\text{ cm}^{-1}$ and a new component at low intensity around 1630 cm^{-1} , implying the formation of intermolecular β -sheets [42].

Overall, our data suggest that pH-induced condensation of *hNTD* does not induce drastic conformational changes within its disordered structure, and the involvement of intermolecular β -sheet structures in stabilizing protein condensates appears to be minimal.

In summary, we have observed that pH jump can initiate the liquid condensation of *hNTD*. In light of the relatively low content of aromatic and hydrophobic residues in *hNTD*, and the influence of salt, we postulate that a network of electrostatic interactions may be the primary driving force of *hNTD* LLPS. Having documented the propensity of *hNTD* for LLPS, our subsequent inquiry delved into the potential role of charge distribution in this phenomenon.

3.4. Charge segregation strongly influences *hNTD* condensation

A closer look inside *hNTD* reveals that the region encompassing residues 117-214 houses both the nuclear and nucleolar localization sequences (NLSs), which play a key role in nuclear transport [26]. These NLSs are notably enriched in basic amino acids, a property likely imposed by evolutionary constraints that impact the charge distribution of this specific region ($\kappa=0.224$, **Fig. 1D**). In order to explore the effects of charge distribution on LLPS, while keeping strictly unchanged the *hNTD* region containing the NLS intact, we modified the charge pattern of *hNTD*¹⁻¹¹⁶ ($\kappa=0.039$). Two synthetic sequences of *hNTD*¹⁻¹¹⁶, designated as “medium κ ” ($\kappa=0.226$) and “high κ ” ($\kappa=0.373$), were generated applying a permutation scheme as previously described in [33]. Briefly, these permutants share with *hNTD*¹⁻¹¹⁶ the same positions of non-charged residues and the number of charged ones, differing solely in the distribution of positive and negative residues. Each synthetic *hNTD*¹⁻¹¹⁶ sequence was then fused to the wild-type *hNTD*¹¹⁷⁻²¹⁴, which includes the nuclear localization signals. These resulting constructs, encompassing the NLS along with either the “medium κ ” or “high κ ” version of *hNTD*¹⁻¹¹⁶, are referred to as *M κ -NTD* ($\kappa=0.224$) and *H κ -NTD* ($\kappa=0.298$), respectively (**Fig. S1**). Importantly, the permutation of charged residues has no discernible impact on the degree of disorder or the predicted LLPS propensity of the protein (**Fig. S5**). Both *M κ -NTD* and *H κ -NTD* variants were recombinantly produced in *E. coli* cells, and their propensity to form protein condensates was assessed as previously described for wild-type *hNTD*.

Experimental conditions used to trigger *hNTD* condensation were equally applied to *M κ -NTD* and *H κ -NTD*. Notably, *M κ -NTD* displays a higher propensity for pH-induced LLPS when compared to *hNTD*, as indicated by both turbidimetry and fluorescence confocal microscopy analyses (**Fig. 4**). Turbidimetry experiments reveal that, at identical protein concentrations, *M κ -NTD* reaches optical densities 2-3 times greater than *hNTD*. Additionally, *M κ -NTD* exhibits notably faster condensation kinetics; for instance, at a concentration of $100\text{ }\mu\text{M}$, *M κ -NTD* reaches its peak optical density within just 5 minutes, in contrast to the ~ 30 minutes required by *hNTD* (**Fig. 4A**). While both *M κ -NTD-GFP* and *hNTD-GFP* exhibit similar liquid-like behavior concerning coalescence and FRAP, condensates formed by *M κ -NTD-GFP*

are ~10 times larger than those observed for *h*NTD-GFP (average droplet area of $6.95 \pm 2.67 \mu\text{m}^2$ at 50 mM NaCl) (**Fig. 4B-D**). Akin to *h*NTD-GFP, the condensation of *Mκ*-NTD-GFP diminishes at higher salt concentrations (**Fig. 4A**), resulting in smaller droplets (**Fig. 4B**), a phenomenon particularly pronounced at 300 mM NaCl (**Fig. 4A, B**). Analysis of the FTIR spectrum of bulk *Mκ*-NTD reveals striking similarity to that of *h*NTD, characterized by a main peak at $\sim 1644 \text{ cm}^{-1}$ (black line in **Fig. 4E**), suggesting that charge segregation *per se* does not affect the secondary structure of the protein. Remarkably, the FTIR spectrum of the supernatant collected after the pH jump closely resembles that obtained of *h*NTD (**Fig. 4E** and **3E**, respectively). Interestingly, the pellet spectrum of *Mκ*-NTD (blue line in **Fig. 4E**) exhibits a pronounced component at $\sim 1630 \text{ cm}^{-1}$, indicative of intermolecular β -sheet structures. This observation likely underpins the enhanced LLPS propensity of *Mκ*-NTD when compared to *h*NTD (**Fig. 3E**).

Overall, it appears that the charge distribution of *Mκ*-NTD optimizes intermolecular electrostatic interactions, facilitating the LLPS of this IDR.

Investigating the LLPS propensity of *Hκ*-NTD proved to be very challenging due to its low yield in recombinant production within *Escherichia coli* cells and its tendency to spontaneously condense or aggregate under conditions where other NTD variants remain soluble. For this reason, the highest concentration used for *Hκ*-NTD in this study was limited to 25 μM . Despite this constraint, we were able to obtain a reliable picture of the pH-induced behavior of LLPS and of the impact of salts. An atypical scenario emerged during the pH-jump experiments: at low salt concentrations, the turbidity of *Hκ*-NTD sample remained considerably low compared to *h*NTD and *Mκ*-NTD, and, contrary to the behavior observed in these latter variants, turbidity unexpectedly increased with higher salt concentrations (**Fig. 5A**). Confocal microscopy analysis indicated that pH-induced condensates of *Hκ*-NTD-GFP have distinctive shapes and are unable to coalesce (**Fig. 5B, C**). Nevertheless, FRAP experiments revealed that these condensates maintained a liquid-like nature, with approximately 70% recovery within one minute, signifying their amorphous condensates behavior (**Fig. 5D**).

In the FTIR spectrum of *Hκ*-NTD (black line in **Fig. 5E**), two main components were present: a peak at $\sim 1654 \text{ cm}^{-1}$, which can be assigned to α -helical structures and/or random coils, and another peak at $\sim 1639 \text{ cm}^{-1}$, suggestive of β -sheet structures with potential contributions from random coils. Consequently, upon band narrowing by second derivative calculations, *h*NTD and *Mκ*-NTD showed a featureless band in the 1660-1620 cm^{-1} region, while *Hκ*-NTD exhibited two well-resolved components, as observed in folded proteins [44]. This collective evidence from FTIR and CD spectroscopies suggests that the *Hκ*-NTD variant possesses a higher degree of ordered secondary structure compared to *h*NTD and *Mκ*-NTD (black line in **Fig. 5E** and **Fig. S4**). Therefore, it can be inferred that charge segregation induces the formation of intra-chain electrostatic interactions, thereby stabilizing more ordered conformations, as already observed in other high- κ permutants [33]. Following the spin assay, the FTIR spectrum of the supernatant mirrored the components detected in the bulk solution prior to the pH jump (red line in **Fig. 5E**), albeit with a higher intensity of the 1654 cm^{-1} peak. In contrast, the pellet exhibited a peak for intermolecular β -sheets shifted to $\sim 1626 \text{ cm}^{-1}$, indicating that the *Hκ*-NTD condensates adopt a unique β -sheet structure distinct from that observed in *h*NTD and *Mκ*-NTD (blue line in **Fig. 5E**).

In summary, our data reveal that charge clustering confers to *Hκ*-NTD a stronger polyampholyte trait when compared to *Mκ*-NTD. These interactions in *Hκ*-NTD are likely to be less multivalent and fluctuating. Notably, intramolecular interactions in *Hκ*-NTD may dominate to such an extent that condensation becomes feasible only at higher salt concentrations. This, in turn, results in chain swelling and the promotion of intermolecular interactions, albeit yielding condensates with aberrant characteristics.

3.5. Phase separation of *h*NTD is modulated by addition of RNA

RNA, a highly negatively charged molecule, can influence condensations processes driven by electrostatic interactions, in a manner analogous to the effect of salt adding [45]. Lower RNA concentrations are expected to enhance electrostatic interactions and promote LLPS, while higher concentrations may disrupt these interactions, reducing condensation propensity [46]. The interaction between RNA and proteins can also depend on the amino acid composition, with proteins rich in lysine or arginine potentially undergoing complex coacervation phenomena that have been well described [47]. To explore how RNA modulates LLPS in *h*NTD-GFP, we conducted experiments on each NTD variant in PBS, in the presence of various poly(A) RNA concentrations ranging from 0.5 to 5 mg/mL.

At lower RNA concentration (0.5 mg/mL), both *h*NTD-GFP and $M\kappa$ -NTD-GFP undergo LLPS, forming biomolecular condensates that were similar in terms of size and liquid nature (**Fig. 6** and **S6**). In contrast, $H\kappa$ -NTD-GFP did not respond to RNA (**Fig. 6E**), unless incubated for 90 minutes at room temperature with 0.5 mg/mL of RNA, a condition under which we observed the formation of amorphous condensates (**Fig. 6F**), with ~60% fluorescence recovery after photobleaching (**Fig. S6**).

To analyze the composition and secondary structure of these biomolecular condensates, we subjected them to spin assay and performed FTIR analysis on both fractions. In the Amide I region, the second derivative spectrum of RNA (black dotted line in **Fig. 6B, D, G**) showed a broad absorption at ~1649 cm^{-1} and a well-resolved component at ~1602 cm^{-1} , serving as an RNA marker band since proteins lack peaks at this wavenumber. In the case of *h*NTD, the ~1602 cm^{-1} component was undetectable in the supernatant (red line in **Fig. 6B**) but well evident in the pellet spectrum (blue line in **Fig. 6B**), which also featured a broad component at ~1641 cm^{-1} . After RNA absorption subtraction (**Fig. S6B**), the pellet showed a broad main component at ~1630 cm^{-1} , indicative of β -sheets, while the supernatant exhibited a main component at ~1652 cm^{-1} , suggesting α -helical and random coil structures.

Similar results were observed for the $M\kappa$ -NTD samples, although the RNA marker band at ~1602 cm^{-1} was detected in both supernatant and pellet (**Fig. 6D** and **Fig. S6D**). In contrast, $H\kappa$ -NTD exhibited a different behavior, with the RNA-related band solely present in the supernatant. In particular, the spectrum of the $H\kappa$ -NTD supernatant displayed two main components at ~1654-1652 cm^{-1} and ~1639 cm^{-1} (**Fig. 6G** and **Fig. S6F**), assigned as described earlier. In the pellet, the main component at ~1626 cm^{-1} (**Fig. 6G** and **Fig. S6F**) fell within the spectral region typical of β -sheet structures in protein aggregates.

Each variant exhibited a distinct and dose-dependent response to RNA. In the case of *h*NTD-GFP, 2 mg/mL of RNA inhibited the formation of biomolecular condensates (**Fig. 6A**). In contrast, at this RNA concentration, $M\kappa$ -NTD still underwent LLPS, producing condensates very similar to those observed at lower RNA concentration (**Fig. 6C**). Overall, these results support our previous hypothesis that intermolecular interactions are stronger in $M\kappa$ -NTD compared to the wild-type protein. A highly anomalous behavior was observed with $H\kappa$ -NTD, as RNA did not appear to induce condensation at all (**Fig. 6E**).

Discussion

Understanding the chemical forces that govern LLPS is crucial because of its relevance in numerous biological and pathological processes [48]. Our results, together with the seminal work of Christensen and co-workers [21], indicate the capacity of *h*TOP1 to form liquid condensates in both the nuclei and nucleoli of transfected human cells. While the mechanisms underlying LLPS of human topoisomerases remain elusive, it is conceivable that long IDRs may play a key role in inducing condensation and regulating the activity of these enzymes. This aligns with existing data for proteins involved in DNA replication [49] and regulation of chromosome recruitment [4].

Within topoisomerases, long IDRs are present in the N-terminal domain of TOP1B, the protein class of *h*TOP1, or in the C-terminal domain of TOPII [49]. *In-vitro* findings suggest that these regions promote LLPS in both *h*TOP1 (this study) and in human and yeast TOPII [49]. Sequence analysis of *h*NTD

highlights both the high density and uniform distribution of oppositely charged residues ($\kappa = 0.128$). This charge decoration underscores the significance of electrostatics in its intermolecular interactions, as confirmed by *h*NTD sensitivity to ionic strength, where higher salt concentrations prevent LLPS [50]. Of note, compared to other proteins undergoing LLPS [15-17], *h*NTD lower content of aromatic and hydrophobic residues implies weaker π - π stacking and cation- π interactions

In light of the “stickers and spacers” model, we reasoned that charged residues, acting as “stickers”, might profoundly influence LLPS propensity through their distribution [51]. As a proof of concept, we designed two synthetic variants (*i.e.* $M\kappa$ -NTD and $H\kappa$ -NTD) by permutating the positions of stickers while keeping spacers unchanged. Our observations indicate that moderate charge clustering, as in $M\kappa$ -NTD, triggers faster LLPS, with larger and more stable droplets, aligning with computational predictions [14].

In contrast, $H\kappa$ -NTD requires higher salt concentrations to initiate condensation and form condensates of abnormal shape and consistency, with no propensity to coalesce. This anomalous condensation cannot be explained by hydrophobic interactions, as in the case of FUS, TDP-43, Brd4, Sox2, and A11 [52]; rather $H\kappa$ -NTD may require a higher salt concentration to break stronger intramolecular interactions and allow intermolecular contacts. This interpretation is consistent with $H\kappa$ -NTD higher content of secondary structure, as evidenced by CD and FTIR analyses. Furthermore, we note that charge distribution also impacts the kinetics of pH-induced condensation. The nucleation kinetics of *h*NTD, not dependent on concentration, implies an initial arrangement limiting further growth. In contrast, $M\kappa$ -NTD kinetics are concentration-dependent, suggesting that the addition of protein molecules to form a critical size nucleus is the rate-limiting step. In $H\kappa$ -NTD, the situation is yet again different, as kinetics does not depend on protein concentration but is influenced by ionic strength. Here, the rate-limiting step probably lies in the disruption of intramolecular interactions.

RNA concentration, a known factor affecting LLPS [45], affects *h*NTD and $M\kappa$ -NTD condensation differently. While lower RNA concentrations promote condensation by enhancing interactions between positive protein charges and RNA negative charges, higher RNA concentrations abrogate condensation due to charge repulsions. Charge clustering, as seen in $M\kappa$ -NTD fosters more stable interactions with RNA, which persist over a wider RNA concentration range. In contrast, $H\kappa$ -NTD, with greater charge clustering, does not condense across tested RNA concentrations, potentially due to strong protein intramolecular interactions or repulsive forces between RNA negative charges and the protein acidic residues.

Overall, we propose that regions enriched in charged residues, like *h*NTD, could act as 'antennae' sensing the RNA (or salt and pH) in their environment. The configuration of their charges appears to modulate LLPS responses, adapting to specific cellular or subcellular contexts. Charge clustering emerges as a crucial functional feature for proteins undergoing electrostatically-driven LLPS.

Interestingly, our findings suggest that an optimal value of κ may exist, challenging previous LLPS simulations [14], and aligning with behavior predicted for increasing clustering of sticker elements based on π - π stacking and cation- π interactions [22].

5. Conclusions

Our findings reveal that *h*TOP1 is able to form condensates within both the nucleolar and nucleoplasmic compartments. While the mechanisms governing *h*TOP1 condensation remain unknown, our *in-vitro* results underscore the pivotal role of *h*NTD in this phenomenon, which deserves further investigation.

Despite the considerable sequence diversity observed among eukaryotic TOP1 orthologs, a strong evolutionary trend emerges in charge clustering within NTDs, resulting in a weaker clustering of charged residues from yeast to vertebrates. Such a distinctive charge distribution pattern implies a potential regulatory role in the LLPS of TOP1. This raises the question as to whether LLPS mechanisms may vary

along the evolutionary continuum, mirroring distinct charge patterns. This fascinating hypothesis merits rigorous exploration through *in vitro* and *in vivo* experiments.

Furthermore, we observed the existence of a charge clustering threshold, beyond which LLPS propensity is reduced or lost. In other words, maximizing charge separation does not necessarily optimize LLPS.

In summary, this work expands our understanding of the “stickers-and-spacers” model to biomolecular condensates, transcending the specific nature of chemical interactions. We hope that our experimental data will help to develop more reliable models of the behavior of supercharged polypeptides.

Acknowledgments

The authors thank Brunno Rocha Levonne, for his precious assistance in confocal microscopy experiments, Joris van Lindt for his support in RNA-mediated LLPS experiments and Prof. Silvia Barabino for providing EGFP-*h*TOP1 plasmid. The authors also thank Alessia Lambiase for her assistance in CD experiments and Prof. Domenico Salerno for his assistance in critical reading of the manuscript.

Funding

This work was supported by University of Milano-Bicocca Fondo di Ateneo Quota Competitiva [grant 2021-ATEQC-0048] (SB); Research Foundation–Flanders (FWO) [grant FWOSB77] (JA).

CRedit authorship contribution statement

Greta Bianchi: Investigation, Data curation, Visualization, Writing - original draft preparation, reviewing & editing; **Marco Mangiagalli:** Investigation, Data curation, Visualization, Writing - original draft preparation, reviewing & editing; **Diletta Ami:** Investigation, Data curation, Visualization, Writing - reviewing & editing; **Junaid Ahmed:** Investigation, Reviewing & editing; **Silvia Lombardi:** Investigation, Data curation, Visualization, Writing - reviewing & editing; **Antonino Natalello:** Methodology, Investigation, Data curation, Visualization, Writing - reviewing & editing; **Sonia Longhi:** Conceptualization, Supervision, Writing - reviewing & editing. **Peter Tompa:** Conceptualization, Methodology, Supervision, Writing - reviewing & editing. **Stefania Brocca:** Conceptualization, Methodology, Supervision, Writing - original draft preparation, reviewing & editing.

Declaration of competing interest

The authors declare that they have no competing interests.

References

- [1] S.F. Banani, H.O. Lee, A.A. Hyman, M.K. Rosen, Biomolecular condensates: organizers of cellular biochemistry, *Nature reviews Molecular cell biology* 18(5) (2017) 285-298.
- [2] Y. Gao, X. Li, P. Li, Y. Lin, A brief guideline for studies of phase-separated biomolecular condensates, *Nature Chemical Biology* 18(12) (2022) 1307-1318.
- [3] X. Li, Z. An, W. Zhang, F. Li, Phase Separation: Direct and Indirect Driving Force for High-Order Chromatin Organization, *Genes* 14(2) (2023) 499.
- [4] M.W. Parker, M. Bell, M. Mir, J.A. Kao, X. Darzacq, M.R. Botchan, J.M. Berger, A new class of disordered elements controls DNA replication through initiator self-assembly, *Elife* 8 (2019) e48562.
- [5] T. Hirose, K. Ninomiya, S. Nakagawa, T. Yamazaki, A guide to membraneless organelles and their various roles in gene regulation, *Nature Reviews Molecular Cell Biology* 24(4) (2023) 288-304.
- [6] W. Borchers, A. Bremer, M.B. Borgia, T. Mittag, How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation?, *Current opinion in structural biology* 67 (2021) 41-50.

- [7] S. Wu, J. Wen, S. Perrett, Unravelling the microscopic characteristics of intrinsically disordered proteins upon liquid–liquid phase separation, *Essays in Biochemistry* 66(7) (2022) 891-900.
- [8] J.A. Villegas, M. Heidenreich, E.D. Levy, Molecular and environmental determinants of biomolecular condensate formation, *Nature Chemical Biology* 18(12) (2022) 1319-1329.
- [9] G.M. Ginell, A.S. Holehouse, An Introduction to the Stickers-and-Spacers Framework as Applied to Biomolecular Condensates, *Phase-Separated Biomolecular Condensates*, Springer2023, pp. 95-116.
- [10] C.-G. Kim, D.-E. Hwang, R. Kumar, M. Chung, Y.-G. Eom, H. Kim, D.-H. Koo, J.-M. Choi, Recent trends in studies of biomolecular phase separation, *BMB reports* 55(8) (2022) 363.
- [11] E. Spruijt, Open questions on liquid–liquid phase separation, *Communications Chemistry* 6(1) (2023) 23.
- [12] J. Li, M. Zhang, W. Ma, B. Yang, H. Lu, F. Zhou, L. Zhang, Post-translational modifications in liquid-liquid phase separation: a comprehensive review, *Molecular biomedicine* 3(1) (2022) 13.
- [13] R.K. Das, R.V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, *Proceedings of the National Academy of Sciences* 110(33) (2013) 13392-13397.
- [14] Y.-H. Lin, H.S. Chan, Phase separation and single-chain compactness of charged disordered proteins are strongly correlated, *Biophysical journal* 112(10) (2017) 2043-2046.
- [15] S. Das, A.N. Amin, Y.-H. Lin, H.S. Chan, Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters, *Physical Chemistry Chemical Physics* 20(45) (2018) 28558-28574.
- [16] Y.-H. Lin, J.D. Forman-Kay, H.S. Chan, Theories for sequence-dependent phase behaviors of biomolecular condensates, *Biochemistry* 57(17) (2018) 2499-2508.
- [17] B.S. Schuster, G.L. Dignon, W.S. Tang, F.M. Kelley, A.K. Ranganath, C.N. Jahnke, A.G. Simpkins, R.M. Regy, D.A. Hammer, M.C. Good, Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior, *Proceedings of the National Academy of Sciences* 117(21) (2020) 11421-11431.
- [18] T.J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowitz, T.D. Craggs, D.P. Bazett-Jones, T. Pawson, J.D. Forman-Kay, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles, *Molecular cell* 57(5) (2015) 936-947.
- [19] C.W. Pak, M. Kosno, A.S. Holehouse, S.B. Padrick, A. Mittal, R. Ali, A.A. Yunus, D.R. Liu, R.V. Pappu, M.K. Rosen, Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein, *Molecular cell* 63(1) (2016) 72-85.
- [20] K.D. Corbett, J.M. Berger, Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases, *Annu. Rev. Biophys. Biomol. Struct.* 33 (2004) 95-118.
- [21] M.O. Christensen, H.U. Barthelmes, F. Boege, C. Mielke, The N-terminal domain anchors human topoisomerase I at fibrillar centers of nucleoli and nucleolar organizer regions of mitotic chromosomes, *Journal of Biological Chemistry* 277(39) (2002) 35932-35938.
- [22] E.W. Martin, A.S. Holehouse, I. Peran, M. Farag, J.J. Incicco, A. Bremer, C.R. Grace, A. Soranno, R.V. Pappu, T. Mittag, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains, *Science* 367(6478) (2020) 694-699.
- [23] 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, *Biophysical journal* 112(1) (2017) 16-21.
- [24] G. Erdős, Z. Dosztányi, Analyzing protein disorder with IUPred2A, *Current Protocols in Bioinformatics* 70(1) (2020) e99.
- [25] A. Hatos, S.C.E. Tosatto, M. Vendruscolo, M. Fuxreiter, FuzDrop on AlphaFold: visualizing the sequence-dependent propensity of liquid–liquid phase separation and aggregation of proteins, *Nucleic Acids Research* (2022).
- [26] Y.-Y. Mo, C. Wang, W.T. Beck, A novel nuclear localization signal in human DNA topoisomerase I, *Journal of Biological Chemistry* 275(52) (2000) 41107-41113.
- [27] E.V. Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F.A. Simão, E.M. Zdobnov, OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs, *Nucleic acids research* 47(D1) (2019) D807-D811.
- [28] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *Journal of molecular biology* 215(3) (1990) 403-410.
- [29] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, Pfam: The protein families database in 2021, *Nucleic acids research* 49(D1) (2021) D412-D419.
- [30] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular biology and evolution* 4(4) (1987) 406-425.
- [31] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, *Molecular biology and evolution* 35(6) (2018) 1547.

- [32] A. Husain, N.A. Begum, T. Taniguchi, H. Taniguchi, M. Kobayashi, T. Honjo, Chromatin remodeller SMARCA4 recruits topoisomerase 1 and suppresses transcription-associated genomic instability, *Nature communications* 7(1) (2016) 1-15.
- [33] G. Bianchi, M. Mangiagalli, A. Barbiroli, S. Longhi, R. Grandori, C. Santambrogio, S. Brocca, Distribution of Charged Residues Affects the Average Size and Shape of Intrinsically Disordered Proteins, *Biomolecules* 12(4) (2022) 561.
- [34] G. Tedeschi, M. Mangiagalli, S. Chmielewska, M. Lotti, A. Natalello, S. Brocca, Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue, *Biochimica et Biophysica Acta (BBA)-General Subjects* 1861(11) (2017) 2543-2550.
- [35] G. Tedeschi, E. Salladini, C. Santambrogio, R. Grandori, S. Longhi, S. Brocca, Conformational response to charge clustering in synthetic intrinsically disordered proteins, *Biochimica et Biophysica Acta (BBA)-General Subjects* 1862(10) (2018) 2204-2214.
- [36] E. Goormaghtigh, V. Raussens, J.-M. Ruysschaert, Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes, *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* 1422(2) (1999) 105-185.
- [37] 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 research* 46(W1) (2018) W315-W322.
- [38] A. Girstun, T. Ishikawa, B. Kowalska-Loth, A. Czuby, K. Staron, Subnuclear localization of human topoisomerase I, *Journal of cellular biochemistry* 118(2) (2017) 407-419.
- [39] O. Vassallo, S. Castelli, I. D'Annessa, B.M. Della Rocca, L. Stella, B.R. Knudsen, A. Desideri, Evidences of a natively unfolded state for the human topoisomerase IB N-terminal domain, *Amino Acids* 41(4) (2011) 945-953.
- [40] J. Van Lindt, A. Bratek-Skicki, P.N. Nguyen, D. Pakravan, L.F. Durán-Armenta, A. Tantos, R. Pancsa, L. Van Den Bosch, D. Maes, P. Tompa, A generic approach to study the kinetics of liquid-liquid phase separation under near-native conditions, *Communications biology* 4(1) (2021) 1-8.
- [41] S. Alberti, A. Gladfelder, T. Mittag, Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates, *Cell* 176(3) (2019) 419-434.
- [42] A. Natalello, D. Ami, S.M. Doglia, Fourier transform infrared spectroscopy of intrinsically disordered proteins: measurement procedures and data analyses, *Intrinsically disordered protein analysis*, Springer2012, pp. 229-244.
- [43] D. Ami, A. Natalello, Characterization of the Conformational Properties of Soluble and Insoluble Proteins by Fourier Transform Infrared Spectroscopy, *Insoluble Proteins*, Springer2022, pp. 439-454.
- [44] A. Barth, Infrared spectroscopy of proteins, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1767(9) (2007) 1073-1101.
- [45] Q. Guo, X. Shi, X. Wang, RNA and liquid-liquid phase separation, *Non-coding RNA Research* 6(2) (2021) 92-99.
- [46] S. Maharana, J. Wang, D.K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillén-Boixet, T.M. Franzmann, RNA buffers the phase separation behavior of prion-like RNA binding proteins, *Science* 360(6391) (2018) 918-921.
- [47] T. Ukmar-Godec, S. Hutten, M.P. Grieshop, N. Rezaei-Ghaleh, M.-S. Cima-Omori, J. Biernat, E. Mandelkow, J. Söding, D. Dormann, M. Zweckstetter, Lysine/RNA-interactions drive and regulate biomolecular condensation, *Nature communications* 10(1) (2019) 1-15.
- [48] S. Mehta, J. Zhang, Liquid-liquid phase separation drives cellular function and dysfunction in cancer, *Nature Reviews Cancer* 22(4) (2022) 239-252.
- [49] J. Jeong, J.H. Lee, C.C. Carcamo, M.W. Parker, J.M. Berger, DNA-Stimulated Liquid-Liquid Phase Separation by Eukaryotic Topoisomerase II Modulates Catalytic Function, *bioRxiv* (2022).
- [50] R. Vancraenenbroeck, Y.S. Harel, W. Zheng, H. Hofmann, Polymer effects modulate binding affinities in disordered proteins, *Proceedings of the National Academy of Sciences* 116(39) (2019) 19506-19512.
- [51] G. Bianchi, S. Longhi, R. Grandori, S. Brocca, Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins, *International Journal of Molecular Sciences* 21(17) (2020) 6208.
- [52] G. Krainer, T.J. Welsh, J.A. Joseph, J.R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M.A. Czekalska, Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions, *Nature communications* 12(1) (2021) 1-14.

Figure captions

Figure 1. *h*TOP1 forms biomolecular condensates *in vivo*. **A)** Representative confocal microscopy image of GFP-tagged *h*TOP1 condensates in the nucleus and nucleolus of HeLa cells transiently over expressing the protein. **B)** FRAP analysis of nuclear and nucleolar condensates of GFP-tagged *h*TOP1 (n = 26). **C)** Architecture, disorder and LLPS predictions of *h*TOP1. Top panel. The enzyme is composed of N-terminal domain (NTD), core domain, linker (L) and catalytic (C) domains. Middle panel. Disorder propensities per residue as predicted by IUPRED2A. The black line indicates the cutoff value of 0.5, residues with a score above this cutoff are predicted to be disordered. Bottom panel. LLPS propensity of each residue as predicted by the *FuzDrop* software. The black line indicates the cutoff value of 0.6, residues with a score above this cutoff are predicted to undergo LLPS. **D)** Plot of the net charge per residue (NCPR) of *h*NTD (residues 1-214) consisting of *h*NTD¹⁻¹¹⁶ (light yellow) and the NLS region (residues 117-214, dark yellow). Data were obtained using CIDER with a sliding window of 5 residues.

Figure 2. The charge distribution of the N-terminal disordered domain of TOP1 is not conserved. **A)** Heat map of the amino acid composition of the NTDs of TOP1 orthologous sequences. The boundaries of the different NTDs were obtained by Pfam analysis. **B)** Correlation between κ values of TOP1 NTDs and the phylogeny of TOP1 orthologous sequences.

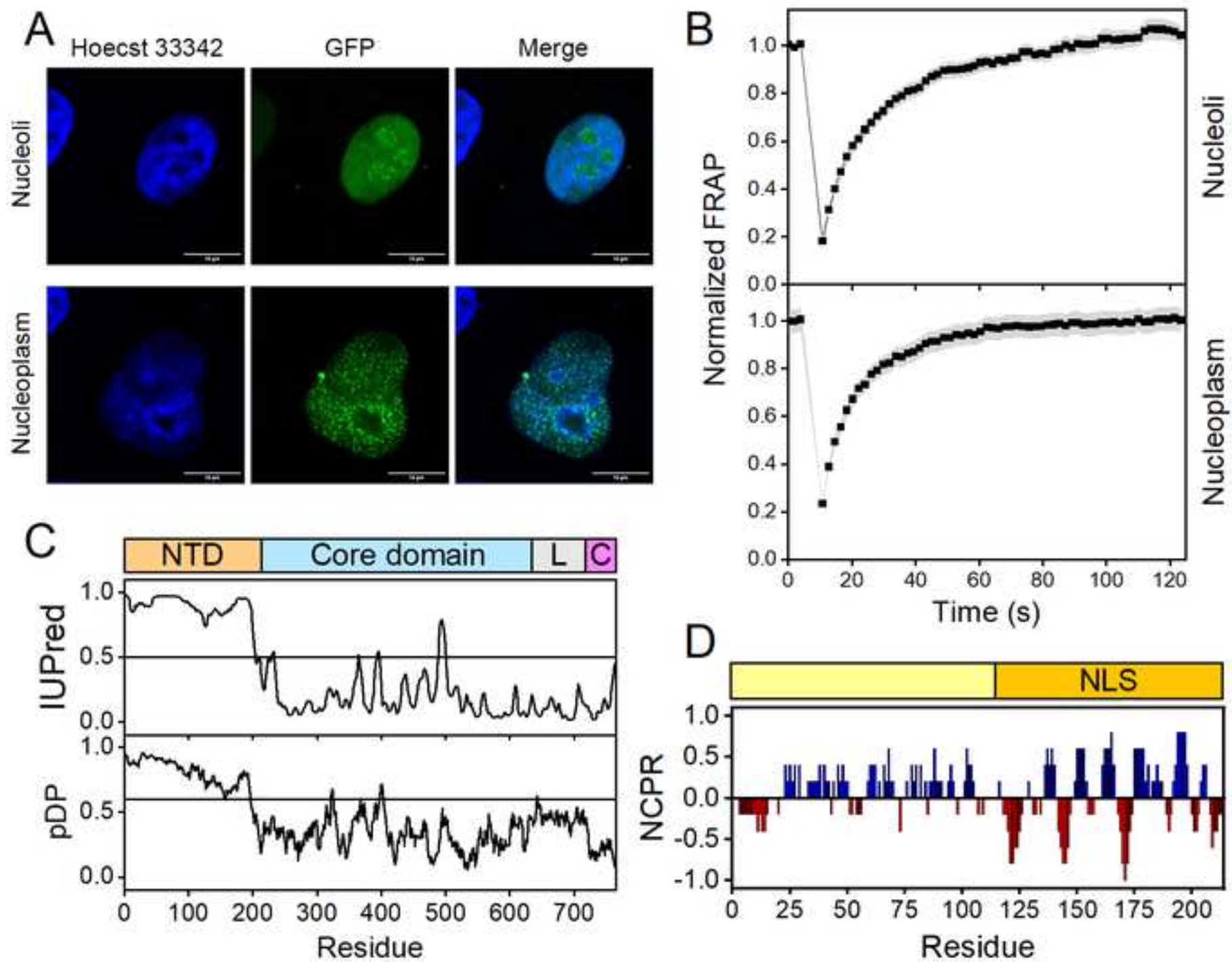
Figure 3. The LLPS of *h*NTD is driven by electrostatic interactions. **A)** Turbidity plots of *h*NTD after pH jump. The experiments were carried out with three different protein concentrations at three different salt concentrations. **B)** Fluorescence microscopy analysis of *h*NTD-GFP condensates. The confocal microscopy images of *h*NTD-GFP (50 μ M) condensates were taken after the pH jump at different salt concentrations. Scale bar: 10 μ m. Each image is representative of three independent measurements. The area of *h*NTD-GFP droplets was calculated using the FIJI ImageJ software v2.0 (n = 150 droplets). **C)** Fusion events of *h*NTD-GFP pH-induced droplets. Scale bar, 2 μ m. **D)** FRAP analysis of *h*NTD-GFP pH-induced droplets (50 μ M, n = 23). Confocal microscopy images show droplets before, after bleaching and at the final phase of recovery (1 min). The fluorescence signal has been normalized from 0 to 1. **E)** Secondary structure analysis of *h*NTD condensates. LLPS was triggered by pH jump and protein condensates were centrifuged to separate the pellet from the supernatant (spin assay). FTIR second derivative spectra of *h*NTD were collected before the pH jump on the *h*NTD solution and, after the jump, on both fractions obtained by centrifugation. The mean second derivatives of the absorption spectra from three experiments were reported and the shadowed area refers to the standard deviation of the data.

Figure 4. $M\kappa$ -NTD has a strong propensity to condense. **A)** Turbidity plots of $M\kappa$ -NTD after pH jump. The experiments were carried out with three different protein concentrations at three different salt concentrations. **B)** Morphological analysis of $M\kappa$ -NTD-GFP condensates. The confocal microscopy images of $M\kappa$ -NTD-GFP (50 μ M) were taken after the pH jump at different salt concentrations. Scale bar: 10 μ m. Each image is representative of three independent measurements. The area of $M\kappa$ -NTD-GFP droplets was calculated using the FIJI ImageJ software v2.0 (n = 150 droplets). **C)** Fusion events of $M\kappa$ -NTD-GFP droplets induced by pH jump. Scale bar, 5 μ m. **D)** FRAP analysis of $M\kappa$ -NTD-GFP pH-induced droplets (50 μ M, n=75). Confocal microscopy images show droplets before, after bleaching and at the end of recovery (1 min). The fluorescence signal has been normalized from 0 to 1. **E)** Secondary structure analysis of $M\kappa$ -NTD condensates. LLPS was triggered by pH jump and protein condensates were centrifuged to separate the pellet from the supernatant. FTIR spectra were collected before the pH jump on the $M\kappa$ -NTD solution and, after the jump, on both fractions obtained by centrifugation. The mean second derivatives of the absorption spectra from three experiments are reported and the shaded area refers to the standard deviation of the data.

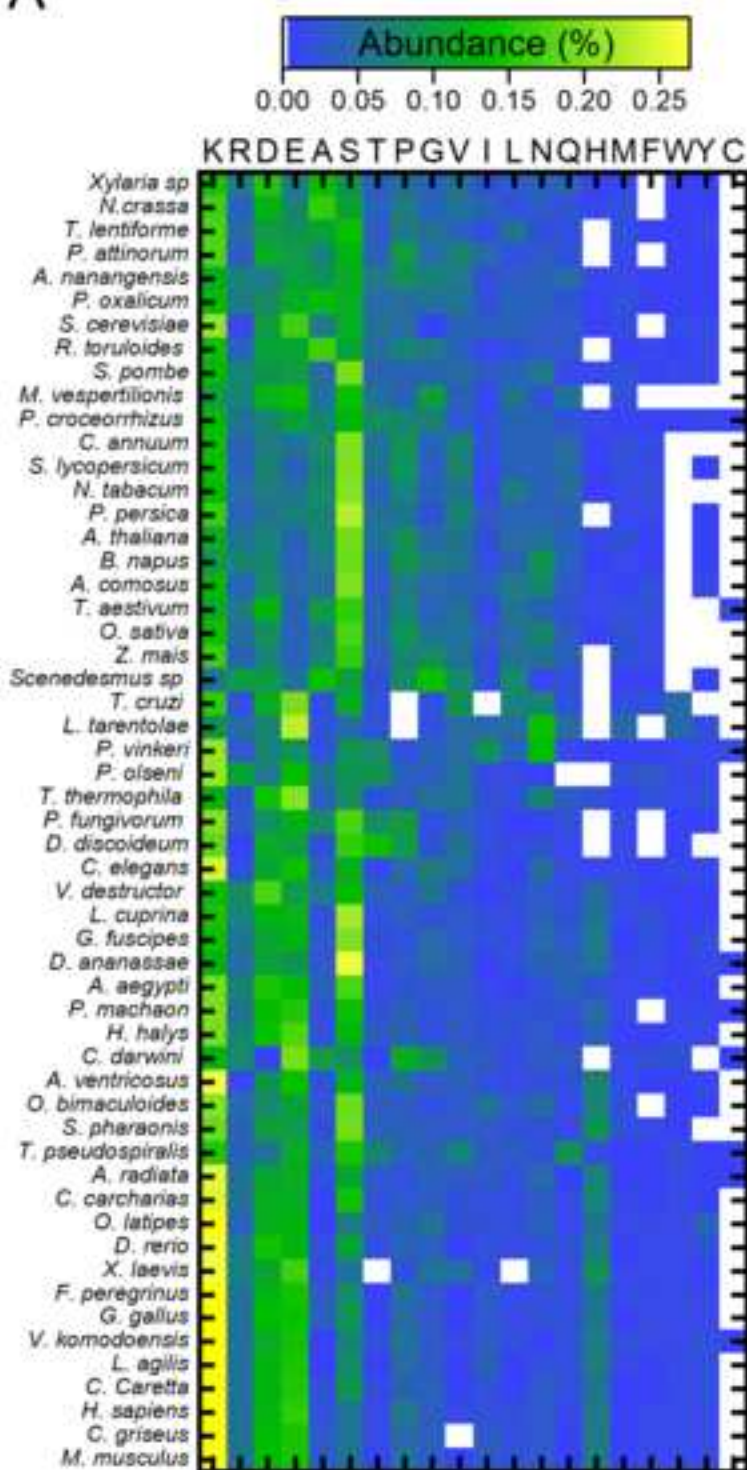
Figure 5. $H\kappa$ -NTD forms anomalous protein condensates. **A)** Turbidity plots of $H\kappa$ -NTD after pH jump. The experiments were carried out with three protein concentrations, at three different salt concentrations. **B)** Morphological analysis of $H\kappa$ -NTD-GFP condensates. The confocal microscopy images of $H\kappa$ -NTD-GFP (25 μ M) were taken after the pH jump at different salt concentrations. Scale bar: 10 μ m. Each image is representative of three independent measurements. The area of $H\kappa$ -NTD-GFP condensates was calculated using the FIJI ImageJ software v2.0 (n = 150 droplets). **C)** Absence of fusion events of $H\kappa$ -NTD-GFP condensates. Scale bar, 5 μ m. **D)** FRAP analysis of $H\kappa$ -NTD-GFP condensates induced by pH jump (25 μ M, n = 28). Confocal microscopy images

show droplets before, after bleaching and at the end of recovery (1 min). The fluorescence signal was normalized from 0 to 1. **E**) Secondary structure analysis of H κ -NTD condensates. The LLPS was triggered by a pH jump and the protein condensates were centrifuged to separate the pellet from the supernatant. FTIR spectra were collected on the H κ -NTD solution before the pH jump, and on both fractions obtained by centrifugation after the jump. The mean second derivatives of the absorption spectra from three experiments were reported and the shaded area refers to the standard deviation of the data.

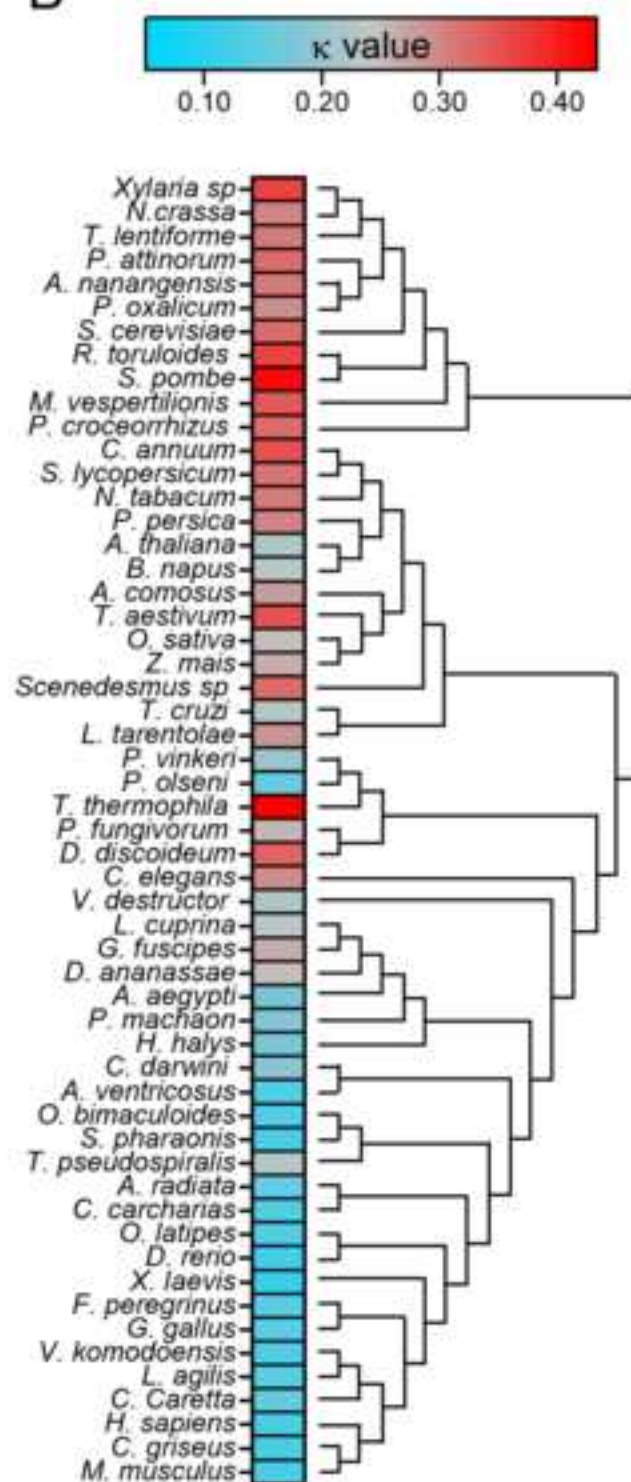
Figure 6. Phase separation of *h*NTD is modulated by addition of RNA. (**A, C, E**) Morphological analysis of *h*NTD-GFP (**A**), M κ -NTD-GFP (**C**), H κ -NTD-GFP (**E**) condensates induced by RNA addition. Confocal microscopy images of the three constructs (20 μ M) were taken after 3 min incubation with different poly(A) concentrations (0.5, 2, 5 mg/mL). Scale bar: 10 μ m. Areas of *h*NTD-GFP and M κ -NTD-GFP condensates were calculated using the FIJI ImageJ software v2.0 (n = 150 droplets). (**B, D, G**) Secondary structure analysis of *h*NTD, M κ -NTD, and H κ -NTD condensates. LLPS was primed by the addition of 0.5 mg/mL poly(A) and the protein condensates were separated from the supernatant by centrifugation (spin assay). Each panel shows the FTIR spectra of a solution of RNA alone, the supernatant and the pellet obtained from the condensates after the spin assay. The mean second derivatives of the absorption spectra of three experiments are shown and the shaded area refers to the standard deviation of the data. The dashed black line corresponds to free RNA (0.5 mg/mL). All spectra are corrected for the contribution of PBS. **F**) Amorphous condensates of H κ -NTD-GFP appear 90 minutes after the addition of RNA (0.5 mg/mL). Scale bar: 10 μ m.

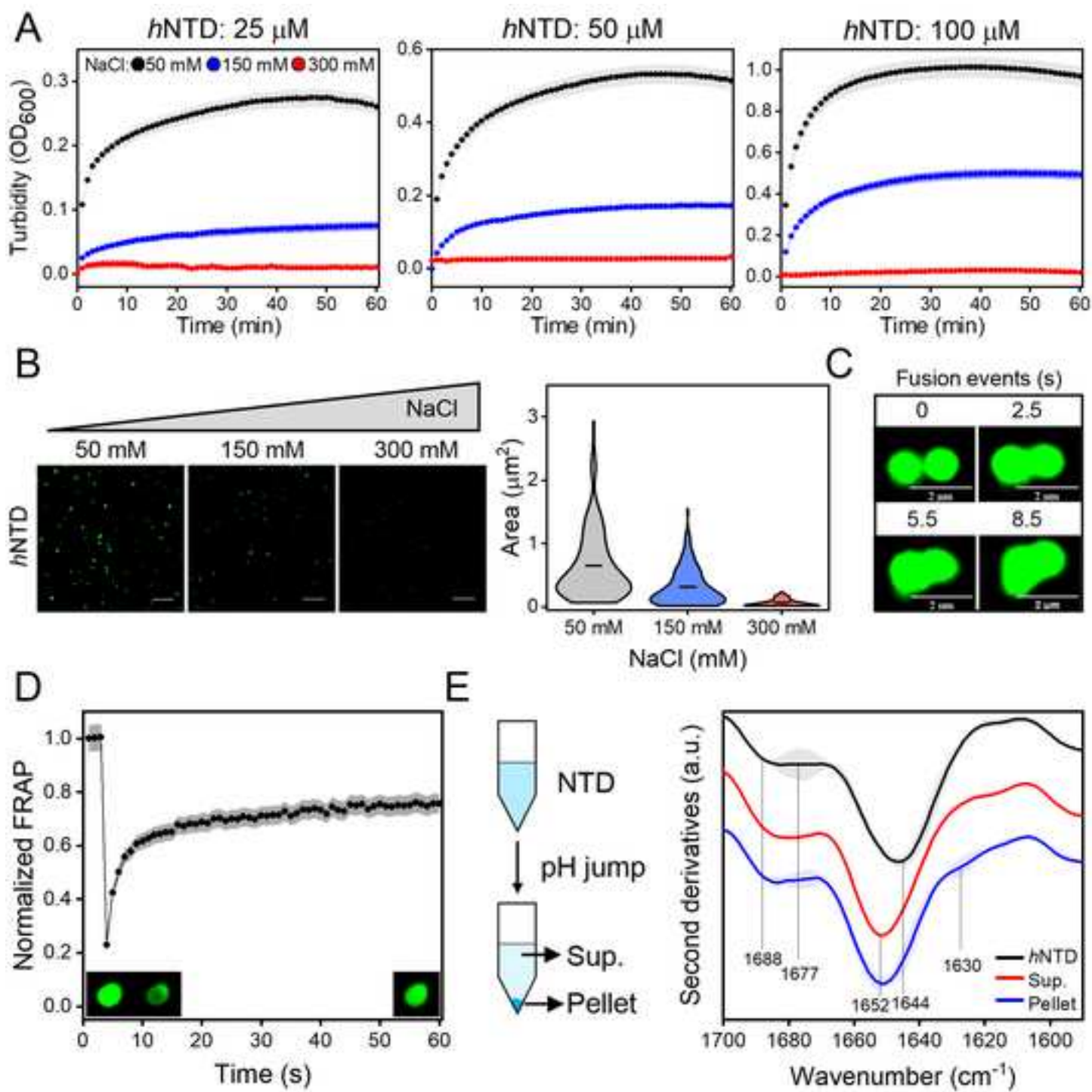


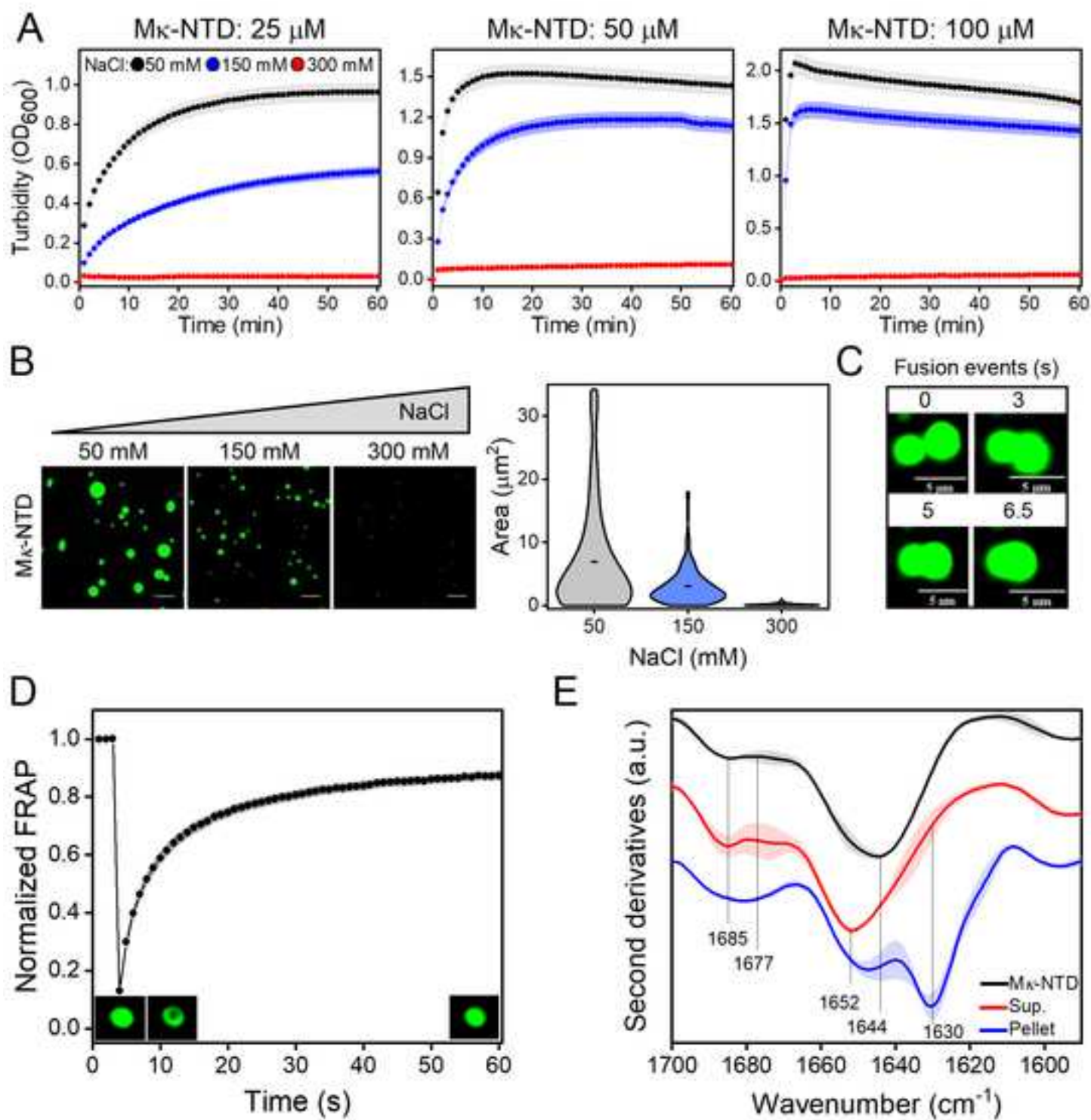
A

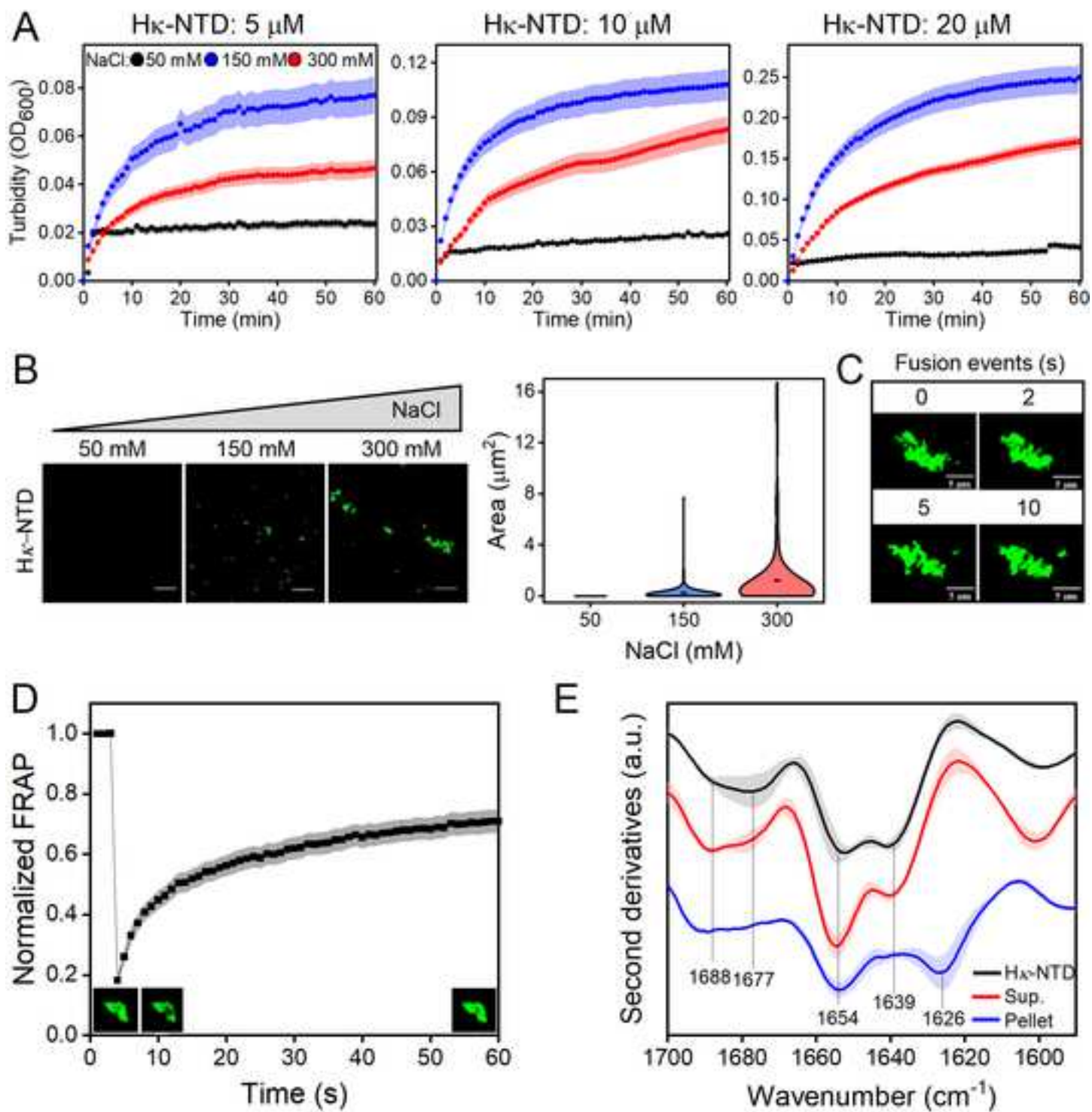


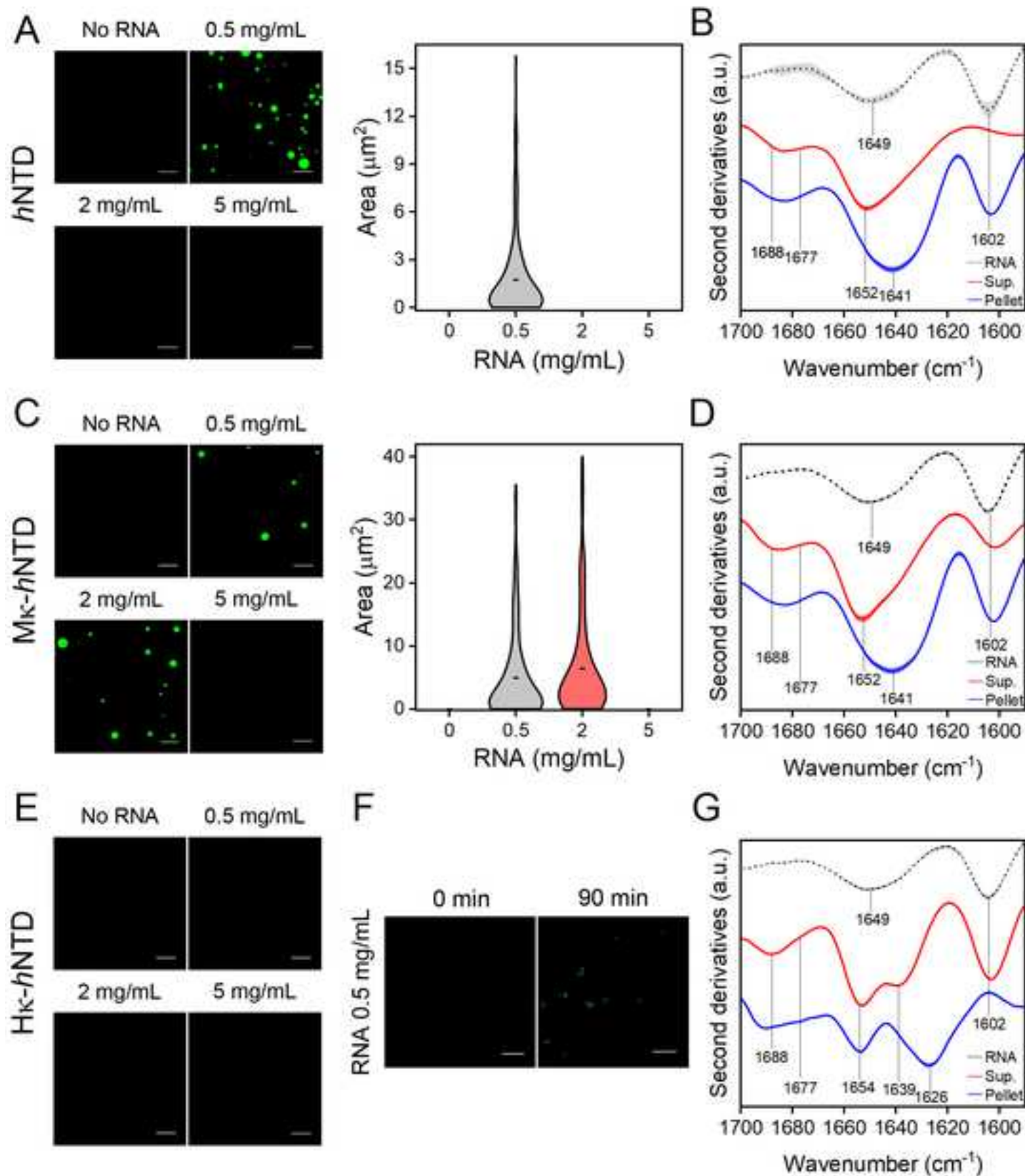
B











Supplementary Materials for

Condensation of the N-terminal domain of human topoisomerase 1 is driven by electrostatic interactions and tuned by its charge distribution

Greta Bianchi¹, Marco Mangiagalli¹, Diletta Ami¹, Junaid Ahmed², Silvia Lombardi¹, Sonia Longhi³, Antonino Natalello¹, Peter Tompa^{3*}, Stefania Brocca^{1*}

* Corresponding authors:

Prof. Stefania Brocca, E-mail: stefania.brocca@unimib.it;

Prof. Peter Tompa, E-mail: peter.tompa@vub.be

This PDF file includes:

Figs. S1 to S6

Table S1

<i>h</i> NTD	MHHHHHHSGLHLHND SQIEAD WRLND SHKHKDKHKDRE HRHKE HKKKEDREKSKHSNSE H	60
<i>Mκ</i> -NTD	MHHHHHHSGLHLHND SQIEAK WRLNK SHDH DDEHEDEEH DHED HKKRKREEKSKHSNSKH	60
<i>Hκ</i> -NTD	MHHHHHHSGLHLHND SQIEAD WELN ESH DDEHEDEEH DHED HKKRKKEEKSKHSNSKH	60
<i>h</i> NTD	KDSEK KKHKEKE KT KHKD GSS EKHKDKHKDRD KEKRKEEKV VRASGVD ASKIK KEKE ENGFSS	120
<i>Mκ</i> -NTD	RDSK KKHKE DEKTK HRKG GSS KKHRKD HKKKDEKRKKEEKV KASGVD ASKIK KEKE ENGFSS	120
<i>Hκ</i> -NTD	RDSK KRHKKR KKTK HRKG GSS KKHRKH HKKKRDEKDK EEKV KASGVDASKIK KEKE ENGFSS	120
<i>h</i> NTD	PPQIK DE PE DD GYFVPP KED IK PLKRE DE DD ADYK PKKIK TE D TK KEK KRK LE EE ED GK	180
<i>Mκ</i> -NTD	PPQIK DE PE DD GYFVPP KED IK PLKRE DE DD ADYK PKKIK TE D TK KEK KRK LE EE ED GK	180
<i>Hκ</i> -NTD	PPQIK DE PE DD GYFVPP KED IK PLKRE DE DD ADYK PKKIK TE D TK KEK KRK LE EE ED GK	180
<i>h</i> NTD	<u>LKKPKNKDKDKKVP</u> <u>EPD</u> NKKKK KEKE EEOK WKW EEER	218
<i>Mκ</i> -NTD	<u>LKKPKNKDKDKKVP</u> <u>EPD</u> NKKKK KEKE EEOK WKW EEER	218
<i>Hκ</i> -NTD	<u>LKKPKNKDKDKKVP</u> <u>EPD</u> NKKKK KEKE EEOK WKW EEER	218

Fig. S1. Amino acid sequences of the synthetic NTDs used in this study. The amino acid sequences of *h*NTD, and the *Mκ*- and *Hκ*-NTD permutants have been aligned to clearly convey the rationale used for the design of the permutants. Positively charged residues are shown in blue, negatively charged residues in red, and underlined residues are those belonging to the NLS region.

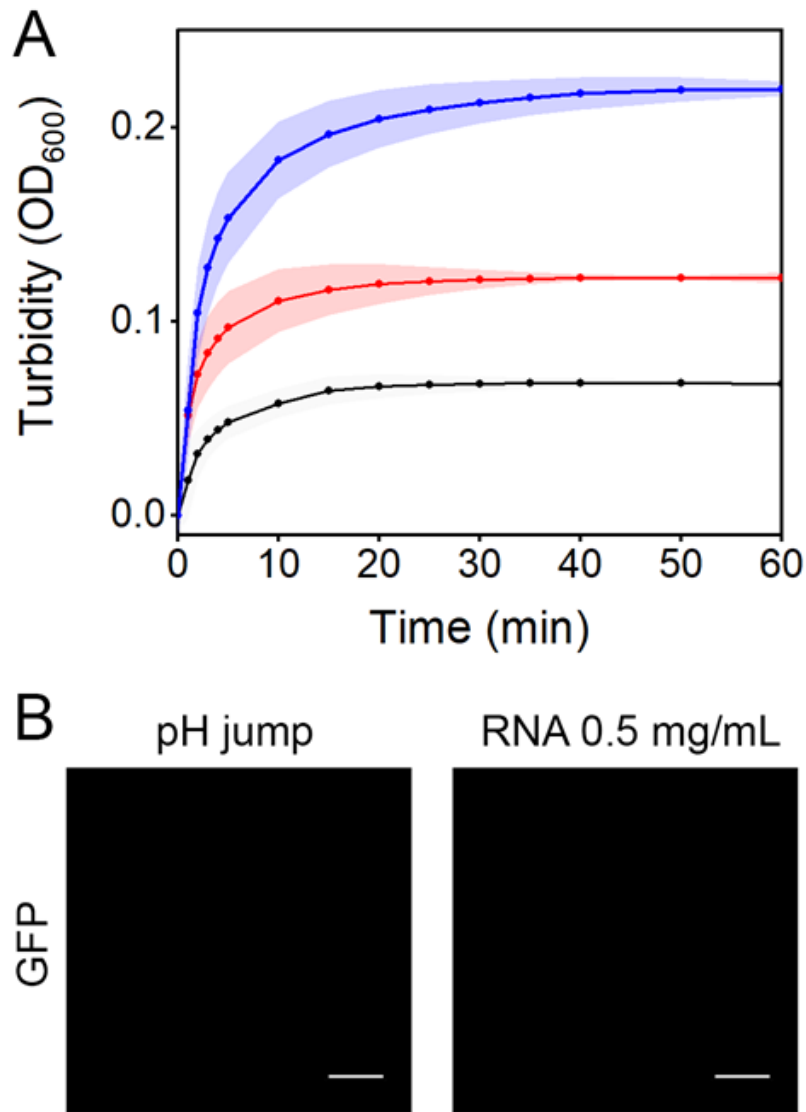


Fig. S2. The fusion of *h*NTD with GFP does not affect its condensation propensity and condensation kinetics. A) Turbidity plots of *h*NTD-GFP after pH jump. Experiments were carried out at three different protein concentrations (25, 50 and 100 μ M) at 150 mM NaCl. B) Confocal microscopy analysis of GFP subjected to either pH jump or addition of poly(A) (0.5 mg/mL). Scale bar 10 μ m.

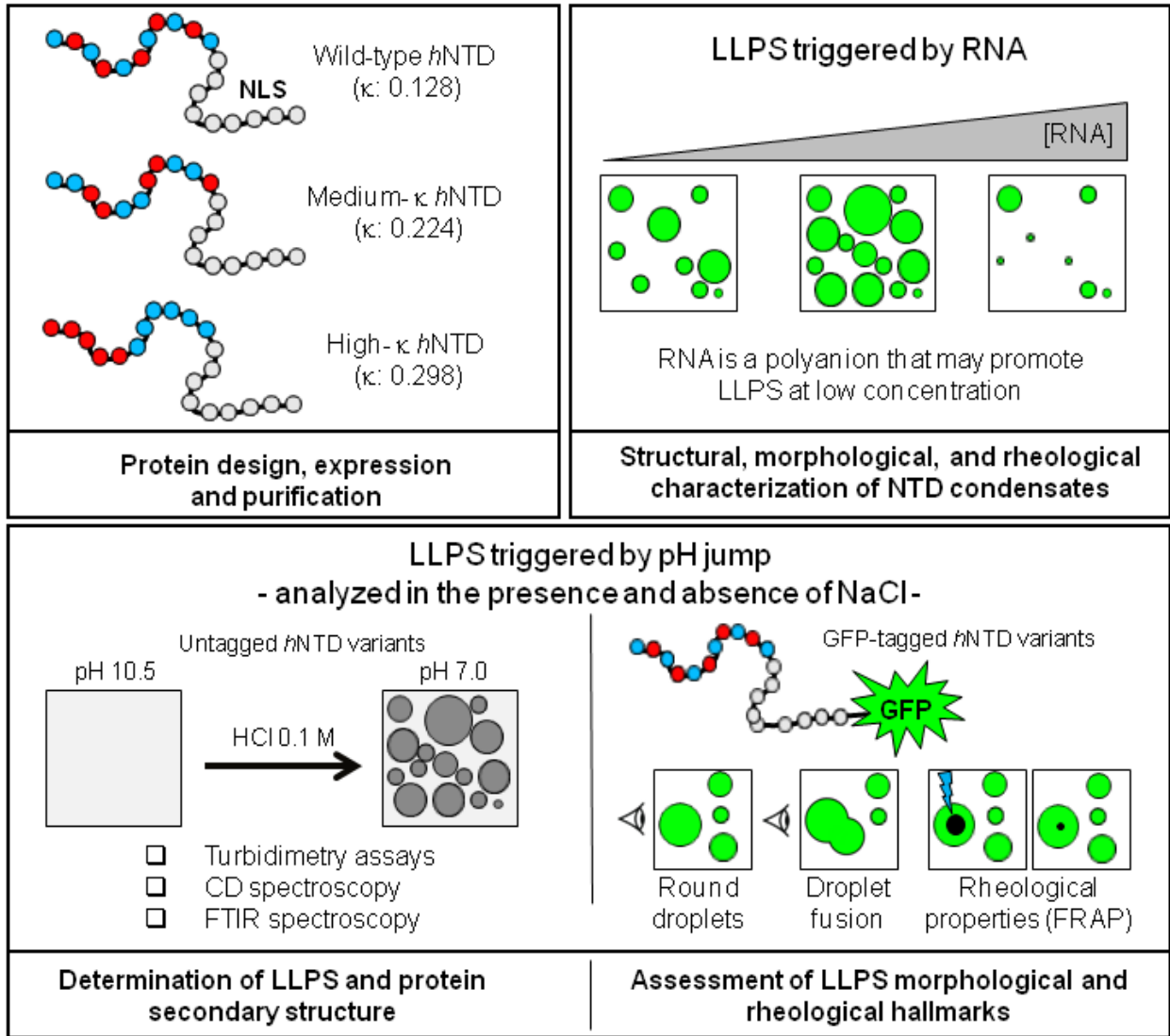


Fig. S3: Schematic of the experimental workflow followed for the induction and characterization of *in vitro* LLPS of *h*NTD and its variants.

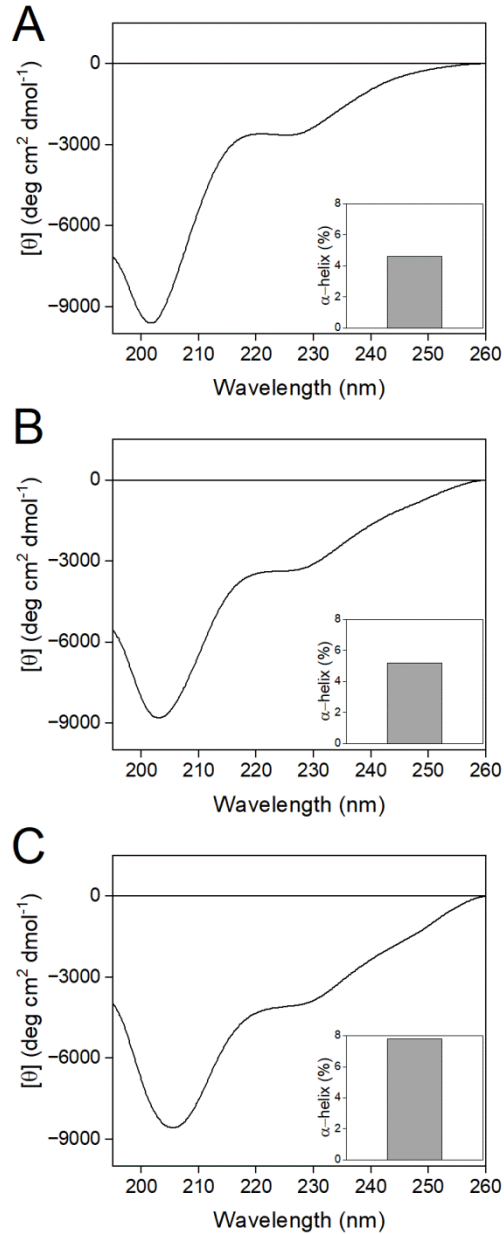


Fig. S4. Far-UV CD spectra of *h*NTD variants. CD spectra of *h*NTD (A), M κ -NTD (B) and H κ -NTD (C) variants. Insets show the α -helical content obtained by spectral deconvolution with BestSel program. CD analyses were performed in alkaline phosphate buffer before the pH jump. The CD spectra of *h*NTD (Panel A) and M κ -NTD (Panel B) are typical of disordered proteins with a minimum at \sim 203 nm. The ellipticity values observed in the range 215-222 nm are consistent with the existence of some secondary structure elements. The CD spectrum of H κ -NTD (C) displays a minimum at c.a. 205 nm and a slightly more pronounced shoulder in the range 215-222 nm, suggesting that this variant is more ordered than the others. Overall, the CD spectra are in agreement with the ATR-FTIR spectra reported in Fig. 3-5.

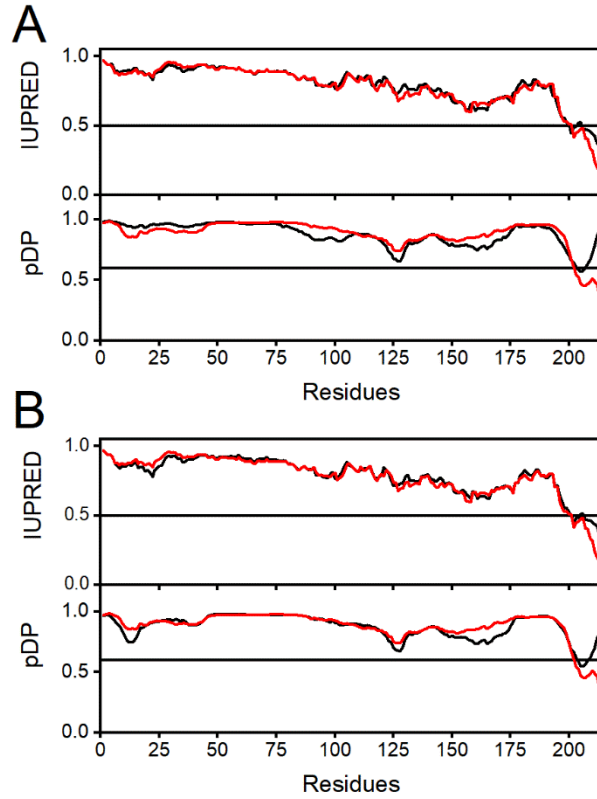


Fig. S5. Predictions of disorder and LLPS propensity for M κ -NTD (A, black line) and H κ -NTD (B, black line) showing that permutation of charged residues does not affect the degree of disorder predicted by IUPred2A and the propensity to undergo LLPS, expressed as liquid condensation probability per residue (pDP) by the FuzDrop server ([https://fuzdrop.bio.unipd.it/predictor](https://fuzdrop.bio.unipd.it/predictor;); (54)) with respect to hNTD (red line).

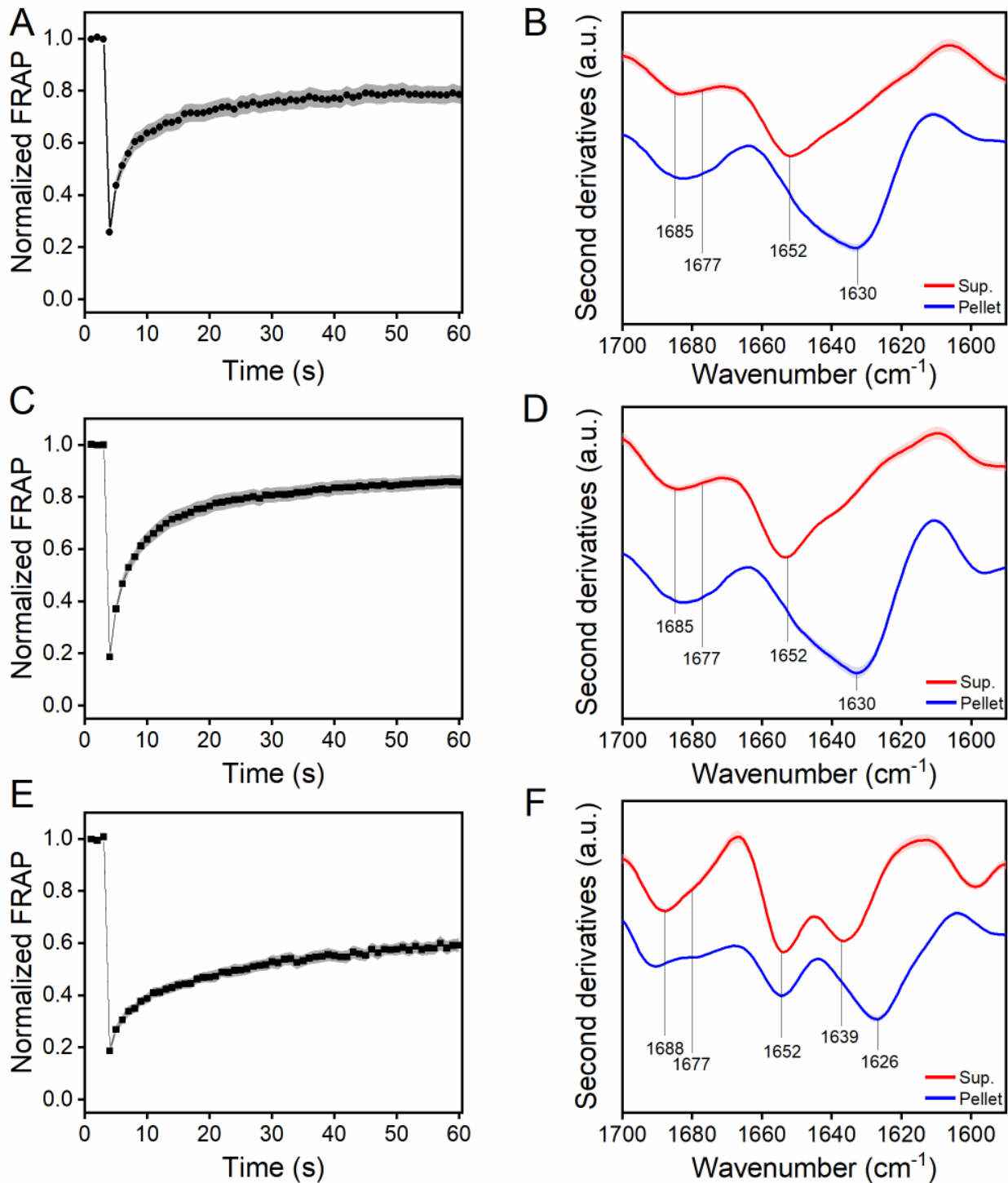


Fig. S6. Fluorescence recovery and secondary structure analysis of condensates obtained in the presence of RNA. **A, C, E**) FRAP analysis of *h*NTD-GFP (**A**), *Mκ*-NTD-GFP (**C**), and *Hκ*-NTD-GFP (**E**) condensates induced by poly(A) at 0.5 mg/mL (protein concentration 20 μ M; $n = 25$). The fluorescence signal was normalized from 0 to 1. **B, D, F**) Second derivatives of FTIR absorption spectra of *h*NTD (**B**), *Mκ*-NTD (**D**), and *Hκ*-NTD (**F**) condensates obtained in the presence of 0.5 mg/mL poly(A), subtracted for RNA and PBS contribution. Red line: supernatant; blue line: pellet.

Group	Subgroup	Species	NTD		
			κ value	NCPR	Length (aa)
Metazoa	Vertebrata	<i>Homo sapiens</i> (P11387)	0.128	0.051	214
		<i>Mus musculus</i> (NP_033434.2)	0.126	0.056	216
		<i>Cricetulus griseus</i> (CAA79748.1)	0.126	0.056	216
		<i>Gallus gallus</i> (NP_990441.1)	0.139	0.060	215
		<i>Falco peregrinus</i> (XP_005235944.1)	0.139	0.060	215
		<i>Danio rerio</i> (NP_001037789.1)	0.122	0.058	207
		<i>Oryzias latipes</i> (XP_023810923.1)	0.134	0.070	199
		<i>Carcharodon carcharias</i> (XP_041059975.1)	0.132	0.082	220
		<i>Amblyraja radiata</i> (XP_032897763.1)	0.145	0.091	220
		<i>Lacerta agilis</i> (XP_033009285.1)	0.140	0.060	215
		<i>Varanus komodoensis</i> (KAF7240897.1)	0.139	0.079	214
		<i>Caretta caretta</i> (XP_048675811.1)	0.149	0.056	215
		<i>Xenopus laevis</i> (XP_018089694.1)	0.114	0.063	270
		Lophotrochozoan	<i>Sepia pharaonis</i> (CAE1163795.1)	0.126	0.037
	<i>Octopus bimaculoides</i> (OX=37653)		0.133	0.074	326
	Nematoda	<i>Trichinella pseudospiralis</i> (KRX92265.1)	0.230	0.028	216
		<i>Caenorhabditis elegans</i> (NP_001379240.1)	0.294	0.016	184
	Arthropoda	<i>Araneus ventricosus</i> (GBM09797.1)	0.134	0.047	387
		<i>Caerostris darwini</i> (GIY45528.1)	0.190	-0.008	127
		<i>Varroa destructor</i> (XP_022671948.1)	0.220	-0.049	676
		<i>Papilio machaon</i> (KPJ12715.1)	0.180	0.013	374
<i>Halyomorpha halys</i> (XP_024217857.1)		0.174	-0.015	395	
<i>Aedes aegypti</i> (EAT46253.1)		0.169	-0.012	421	
<i>Drosophila ananassae</i> (XP_001963839.2)		0.248	-0.020	455	
<i>Glossina fuscipes</i> (XP_037889119.1)		0.265	-0.034	475	
<i>Lucilia cuprina</i> (XP_023291221.2)		0.236	-0.029	476	
<i>Perkinsus olseni</i> (KAF4667565.1)		0.154	0.131	130	
Protista	Alveolata	<i>Tetrahymena thermophila</i> (XP_001025304.2)	0.437	-0.253	133
		<i>Plasmodium vinckei petteri</i> (EUD73215.1)	0.204	0.084	179
		<i>Dictyostelium discoideum</i> (XP_639222.1)	0.337	-0.019	309
	Amoebozoa	<i>Planoprotostelium fungivorum</i> (PRP82603.1)	0.256	0.037	217
		<i>Leishmania tarentolae</i> (GET92410.1)	0.289	-0.163	43
	Euglenozoa	<i>Trypanosoma cruzi</i> (ESS68442.1)	0.225	-0.133	46
Fungi	Ascomycota	<i>Penicillium oxalicum</i> (XP_049968292.1)	0.296	0.009	228
		<i>Phialophora attinorum</i> (XP_018005181.1)	0.324	0.000	242
		<i>Aspergillus nanangensis</i> (KAF988681.1)	0.309	0.020	246
		<i>Neurospora crassa</i> (XP_011393242.1)	0.303	0.004	286
		<i>Xylaria sp. FL0933</i> (KAI1279976.1)	0.366	0.000	240
		<i>Trichoderma lentiforme</i> (KAF3074930.1)	0.315	0.037	273
		<i>Saccharomyces cerevisiae</i> (P04786)	0.324	-0.042	142
		<i>Schizosaccharomyces pombe</i> (NP_596209.1)	0.415	0.010	192
		<i>Pisolithus croceorhizus</i> (KAI6131672.1)	0.326	0.014	212
		<i>Rhodotorula toruloides</i> (XP_016275408.1)	0.361	-0.028	287
	<i>Malassezia vespertilionis</i> (PKI83396.1)	0.342	-0.089	101	
	Basidiomycota	<i>Brassica napus</i> (XP_013726218.2)	0.234	0.056	357
		<i>Arabidopsis thaliana</i> (OAO95951.1)	0.221	0.063	366
	Viridiplantae	Streptophyta	<i>Zea mays</i> (AQK50797.1)	0.263	0.104
<i>Oryza sativa</i> (XP_015650112.1)			0.252	0.070	371
<i>Triticum aestivum</i> (XP_044326396.1)			0.352	0.011	266
<i>Ananas comosus</i> (XP_020110486.1)			0.279	0.045	333
<i>Prunus persica</i> (XP_020425762.1)			0.302	0.070	201
<i>Nicotiana tabacum</i> (AAD08711.1)			0.312	0.057	332
<i>Solanum lycopersicum</i> (XP_010323847.2)			0.319	0.065	336
<i>Capsicum annuum</i> (XP_016565649.1)			0.346	0.061	343
Chlorophyta			<i>Scenedesmus sp</i> (KAF8067313.1)	0.328	0.004

Table S1: List of species from which the TOP1 orthologues considered in this study originate. The NCPR and κ -value of the NTDs, as identified by Pfam(57) (<https://www.ebi.ac.uk/interpro/search/text/398365647/#table>), were calculated using CIDER(52) (<http://pappulab.wustl.edu/CIDER/analysis/>)